Kryteria i aspekty filozoficzne sztucznej inteligencji

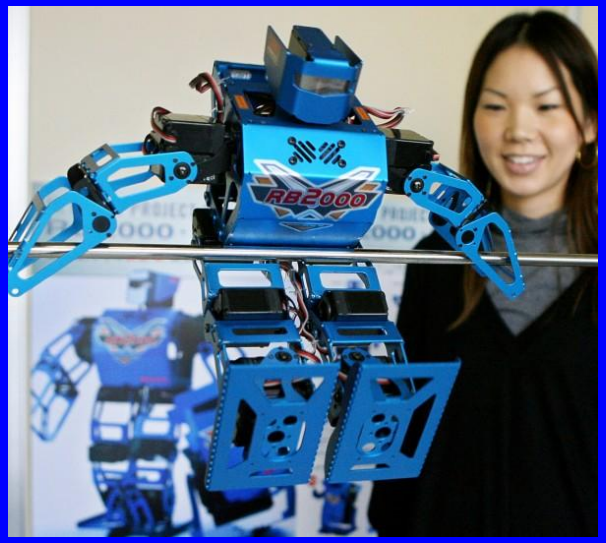
dr hab. inż. Jerzy Balicki, prof. nadzw.

Kryteria sztucznej inteligencji

- **Badania nad naturalną inteligencją człowieka były początkowo inspirowane tym, że ludzie różnią się pod względem zdolności umysłowych.**
- **Psychologiczne teorie inteligencji powstały jako próba zrozumienia istoty tych różnic oraz ich prawidłowego opisu.**

Trzy kryteria sztucznej inteligencji

- Symulacja procesów naturalnych (z użyciem testu Turinga)
- Inteligentne czynności
- Racjonalne sprawstwo



Test Turinga (1)

Test Turinga (1)

- **Alain Turing (1950), *Computing machinery and intelligence*, Mind, 59, 433-460**
- **Zaproponował przeformułowanie pytania: “czy maszyny (komputery) mogą myśleć?”, wprowadzając koncepcję “gry naśladowczej” (imitacji, symulacji).**

Test Turinga (2)

B: człowiek lub komputer

C: człowiek lub komputer

A: człowiek lub komputer

Trzy osoby bawiące się w grę ustalania tożsamości. Nie mogą się one widzieć, są w oddzielnych pokojach, a porozumiewają się za pomocą pisemnych protokołów. Zasadniczym elementem gry jest pytanie Turinga:
"Co się stanie, jeśli komputer zajmie miejsce któregoś z uczestników, a zadaniem będzie ustalenie, kto jest człowiekiem, a kto komputerem?"

Test Turinga (3)

- **Pisanie programów komputerowych zdolnych zdać test Turinga ma głównie cel poznawczy.**
- **Celem poznania może być inteligencja ludzka, a nie maszynowa.**
- **Występują dwa systemy inteligencji:**
  - **naturalny (czyli umysł ludzki),**
  - **sztuczny (program komputerowy).**

Test Turinga (4)

- **System naturalny można poznać tylko pośrednio, wnioskując o zachodzących w nim procesach wyłącznie na podstawie zewnętrznego zachowania.**
- **System sztuczny nie wymaga zabiegów poznawczych, ponieważ został zaprojektowany.**

Test Turinga (5)

- **Jeżeli system sztuczny zda test Turinga, będzie można na tej podstawie wnioskować o możliwym sposobie funkcjonowania systemu naturalnego.**
- **A jeśli system sztuczny nie zda testu, można go wtedy zmieniać i poprawiać aż do momentu, kiedy bezstronny obserwator nie będzie w stanie odróżnić, czy ma do czynienia z maszyną, czy też z człowiekiem.**
- **Tak opracowany program komputerowy można uznać za dobrą symulację naturalnych procesów poznawczych, charakterystycznych dla inteligencji ludzkiej.**

Test Turinga (6)

- **Program komputerowy może być modelem procesów poznawczych człowieka.**
- **Dysponując programem dobrze naśladującym człowieka i znając doskonale jego przebieg i strukturę, uzyskuje się wiedzę o możliwym przebiegu i strukturze ludzkiego procesu poznawczego.**

Inteligentne czynności (1)

- **Czy maszyny zdolne są do wykonywania czynności uznanych przez badacza za inteligentne?**

Inteligentne czynności (1)

- **Badacz może się kierować intuicją lub powszechnie żywionymi przekonaniami.**
- **Gra w szachy, prowadzenie sensownej rozmowy, uczenie się oraz dowodzenie twierdzeń matematycznych to czynności niewątpliwie inteligentne.**

Inteligentne czynności (2)

- **Jedną z pierwszych zaawansowanych prób w zakresie AI był *Teoretyk Logiki*, zaprojektowany przez Newella i Simona - program wyspecjalizowany w dowodzeniu twierdzeń Whiteheada i Russella (*Principia Mathematica*).**
- **Maszyna nie przeszukiwała wyczerpująco zbioru potencjalnie dostępnych sposobów rozwiązania, lecz kierowała się zasadami ograniczającymi zakres przeszukiwania.**
- **Zasady takie nazywa się *heurystykami*. *Heurystyka to reguła pozwalająca ograniczyć obszar przeszukiwania, a tym samym skrócić czas rozwiązywania problemu*.**

Inteligentne czynności (3)

- **Heurystyki znacząco skróciły czas pracy maszyny, ale także sprawiły, że jej zachowanie nie było w stu procentach przewidywalne.**
- **Maszyna wykorzystywała wyniki swoich wcześniejszych działań, aby zwiększyć skuteczność czynności bieżących - przejawiała zdolność do uczenia się na podstawie własnych doświadczeń, a nie jedynie wykonywała zadane jej rozkazy.**
- **Inteligentne czynności to także:**

  (i) użycie heurystyk,

  (ii) uczenie się.

Racjonalne Sprawstwo (1)

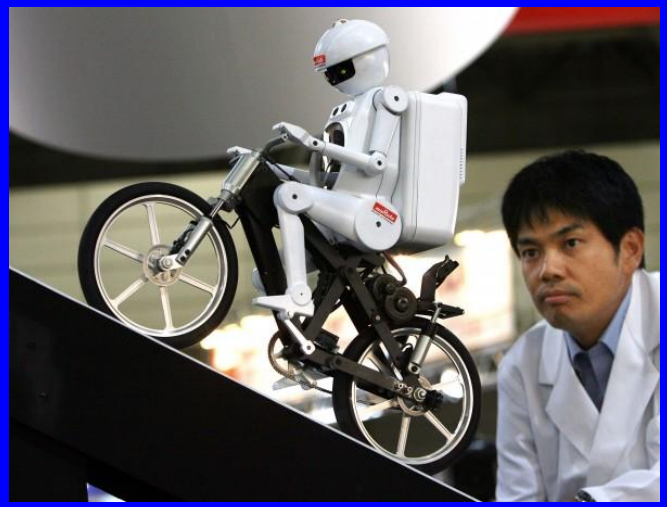
- **RS - zdolność systemu komputerowego do *inicjowania* działań, które są sensowne w określonym środowisku, a następnie do skutecznego *kierowania* tymi działaniami.**
- **System nazywamy inteligentnym wtedy, gdy jest on sprawcą, a nie tylko wykonawcą poleceń, zgodnie z zadanym algorytmem.**
- **Musi być przy tym sprawcą racjonalnym, tzn. dostosowującym swe działania do wymagań bieżącej sytuacji i "naturalnego" dla siebie środowiska.**
- **System spełniający te kryteria należy uznać za podmiot własnych działań.**

Racjonalne Sprawstwo (2)

- **System musi być wyposażony nie tylko w procedury umożliwiające *wykonywanie określonych czynności*, lecz również w *system motywacji*.**
- **Racjonalny sprawca musi posiadać *wiedzę o otoczeniu*.**
- **Może to być wiedza niezwykłe *uproszczona*, zredukowana do najbardziej niezbędnych informacji, ale musi istnieć.**
- ***Ludzka wiedza* też nie jest doskonała ani pełna.**
- **Stanowi zawsze model rzeczywistości, tworzony w ściśle utylitarnym celu: aby nam ułatwić skuteczne poruszanie się w rzeczywistości i rozwiązywanie problemów.**

Racjonalne Sprawstwo (3)

- **Racjonalne sprawstwo jest najtrudniejszym do spełnienia kryterium sztucznej inteligencji.**
- **Nie ma w tej chwili systemów, o których można by z przekonaniem powiedzieć, że to kryterium spełniają.**



Filozofia AI

- **Czym jest myślenie, czucie i świadomość?**
- **Jaki jest związek umysłu z ciałem?**
- **Czy możliwa jest budowa urządzenia, które by posiadało cechy umysłu człowieka?**

Filozofia AI

- **Kartezjusz** w XVII wieku uważał, że funkcjonowanie ciała człowieka powinno dać się wyjaśnić **mechanistycznie**. Człowiek jest złożoną maszyną, której działanie daje się opisać przy pomocy zasad fizyki i chemii.
- Z drugiej strony zjawiska psychiczne są innej natury i nie dają się wyprowadzić ze zjawisk cielesnych. (**doktryna dualizmu** - istnienia dwóch niesprowadzalnych do siebie czynników: ciała i ducha.

Filozofia AI

- Analogia między **zachowaniem człowieka** a **wykonywaniem programu komputerowego**.
- Myślenie ludzkie opiera się na takich samych mechanizmach, jakie występują w maszynach cyfrowych, albo ostrożniej – że *wszelkie ludzkie działanie da się opisać i sformalizować przy pomocy reguł mechanistycznych* (Gurba 1997).

Hipoteza silnej AI

- Newell i Simon z Carnegie Mellon University w Pittsburghu (1976) twierdzą, że wszystko co wykonuje człowiek fizycznie i umysłowo jest wynikiem przetwarzania informacji analogicznego do ***przetwarzania w maszynie cyfrowej przez nadzwyczaj złożony program***.
- **Program** przejawiający działanie analogiczne do działania mózgu ludzkiego będzie posiadał cechy umysłu człowieka łącznie z **myśleniem, czuciem, inteligencją, rozumieniem i świadomością**.
- Atrybuty te są tylko cechami ***algorytmu wykonywanego przez mózg.***

Chiński pokój

- Searle (Searle 1991, Churchland 1991, Penrose 1995, Lem 1996) przedstawił hipotetyczne przykłady działania komputerów prowadzących dialog z użytkownikiem w uproszczonym języku naturalnym (uproszczony **test Turinga**)
- Wyobraźmy siebie w roli **procesora** znajdującego się w zamkniętym pokoju. Komunikacja ze światem zewnętrznym odbywa się przez wąską szczelinę, przez którą wprowadzane i wyprowadzane są ***zapytania i odpowiedzi w języku naturalnym***.

Chiński pokój

- Po otrzymaniu zapytania ***procesor*** wykonuje program, pobierając *instrukcje znajdujące się w regałach wewnątrz pokoju* i wytwarza odpowiedź.
- Jeżeli językiem naturalnym jest np. ***język chiński***, którego nie znamy, to ***wykonując instrukcje*** będziemy prowadzili ***poprawny dialog*** w języku chińskim kompletnie go nie rozumiejąc.
- Zatem ***wykonanie algorytmu*** nie oznacza, że mamy do czynienia ***z prawdziwym zrozumieniem***.

Teleportacja

- Rozważa się przykład „podroży” poprzez ***przesyłanie informacji o strukturze człowieka***, a następnie w miejscu docelowym ***odtworzenie go na podstawie tej informacji*** (teleportacja).
- Czy nowo utworzony „osobnik” będzie miał tą samą ***świadomość***?
- A co się stanie, jeżeli nie zniszczono jednocześnie oryginału?

Umysł a komputer

- **Intencjonalność** umysłu to ***cecha zjawisk psychicznych*** polegająca na kierowaniu się ich ku jakiemuś przedmiotowi lub ***cecha świadomości*** kierująca się na przedmiot, ***konstytuującej sens przez „zdawanie sobie z czegoś sprawy”*** (NEP 1995, Brit 1998).
- ***Myślenie, wiara, pożądanie i inne tego rodzaju odczucia*** uważa się za ***podobne do siebie*** w tym znaczeniu, że ***obejmują obiekt lub kierują się ku obiektowi***, w sposób zasadniczo ***rożny od oddziaływań spotykanych w świecie nieożywionym***.

Umysł a komputer

- Obiekty zainteresowania intencjonalnego mogą nawet nie istnieć.
- ***Przeciwnicy*** poglądu o możliwości budowy sztucznego umysłu uważają, że ***intencjonalność nie jest możliwa do sztucznego odtworzenia***.
- Ponadto wciąż ***trudności napotyka*** komputerowa realizacja takich atrybutów inteligencji jak ***adaptowanie się do nowych warunków, rozwiązywanie nowych zadań czy działalność twórcza.***

Modele komputerowe umysłu

Roger Penrose (1995, 2000):

1. ***Myślenie polega na obliczeniach, a świadome doznania powstają wskutek realizacji odpowiedniego procesu obliczeniowego*** (***silna AI***).
2. ***Świadomość jest cechą fizyczną biologicznego mózgu; wprawdzie wszystkie fizyczne procesy można symulować, ale nie towarzyszy temu świadomość (chiński pokój).***
3. ***Procesy fizyczne w mózgu powodują powstanie świadomości, ale tych procesów nie można symulować obliczeniowo (Penrose).***
4. ***Świadomości nie można wyjaśnić w żaden fizyczny, obliczeniowy czy inny naukowy sposób.***

Modele komputerowe umysłu

- W przyrodzie, a tym samym w mózgu, występują procesy, które ***nie dają się modelować*** przy pomocy uniwersalnej maszyny Turinga (komputera).
- Jest to pogląd sprzeczny z uznawaną ***hipotezą Churcha-Turinga*** mówiącą, że ***procesy fizyczne są obliczalne, a zatem dają się modelować za pomocą maszyny Turinga***.
- Hipotezy o „***silnej AI***” i „***chińskim pokoju***” implikują możliwość posiadania przez maszyny inteligencji. Biorąc pod uwagę systematycznie ***rosnące moce obliczeniowe komputerów***, można spodziewać się ***powstania maszyn o inteligencji przewyższającej inteligencję człowieka i w konsekwencji zmarginalizowanie roli człowieka***.

Modele komputerowe umysłu

- Paradoksalnie koncepcja typu „chiński pokój” jest bardziej pesymistyczna niż koncepcja „silnej AI”, bo wynika z niej, że ***światem rządzić będą automaty nieposiadające umysłu (świadomości).***
- Z hipotezy silnej AI wynikałoby, że ***automaty te posiadałyby umysł***, a tym samym można by powiedzieć, że w pewnym sensie dziedziczyły by cechy człowieka.