

Sztuczna Inteligencja

Szukanie i ludzkie myślenie

Włodzisław Duch

Katedra Informatyki Stosowanej UMK

Google: Wlodzislaw Duch

Co będzie

Omówiliśmy procesy szukania, gry i różne zastosowania.

- Szukanie i ludzkie myślenie.
- Paradoksy kognitywne.
- Systemy oparte na wiedzy, SOW.
- Porównanie mózgi-AI.
- Definicje pojęć w obrębie teorii SOW.

Szukanie a ludzkie myślenie

Jak ludzie rozwiązują problemy wymagające wiedzy i rozumowania?

H.A. Simon: [Artificial intelligence: an empirical science](#) (1995)

Intuicja = ogromna pamięć, rozpoznawanie przestrzennych wzorców i skojarzenia – inne ograniczenia sprzętowe niż AI.

Szachy: człowiek stosuje „zmodyfikowane progresywne pogłębianie”, ograniczone poszukiwanie w głąb, ocena w oparciu o heurystyki związane z rozpoznanymi wzorcami na podstawie doświadczenia.

Liczba pamiętanych „prototypowych” wzorców ~ 50.000.

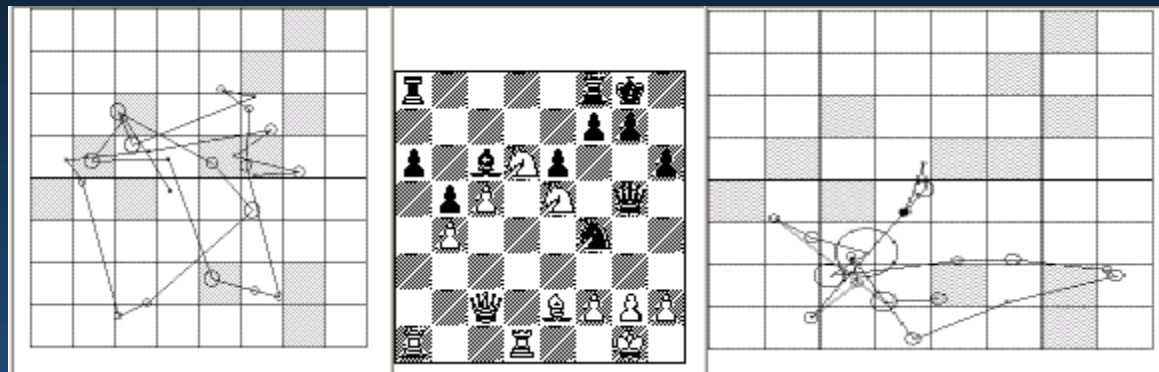
Mistrz szachowy po 5-sek. ekspozycji układa średnio 23 figury z 25 na pokazanych pozycjach, nowicjusz 3 lub 4 figury.

Figury ułożone przypadkowo → taka sama liczba błędów.

Szybki test poziomu gry w szachy.

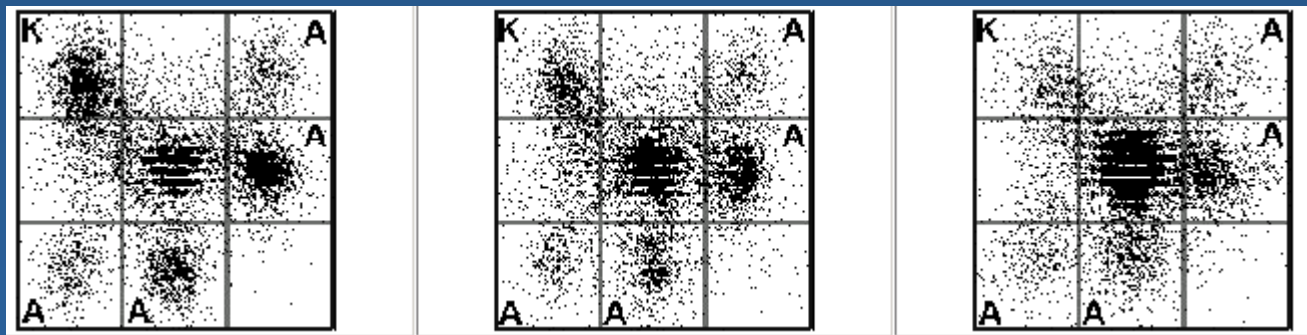
Badania psychologiczne.

Prezentacja 5 sekund, ruchy oczu mistrza i nowicjusza.



Prezentacja dłuższa, od nowicjusza do mistrza.

Największe różnice są w czasach reakcji i liczbie prawidłowo zapamiętanych figur.

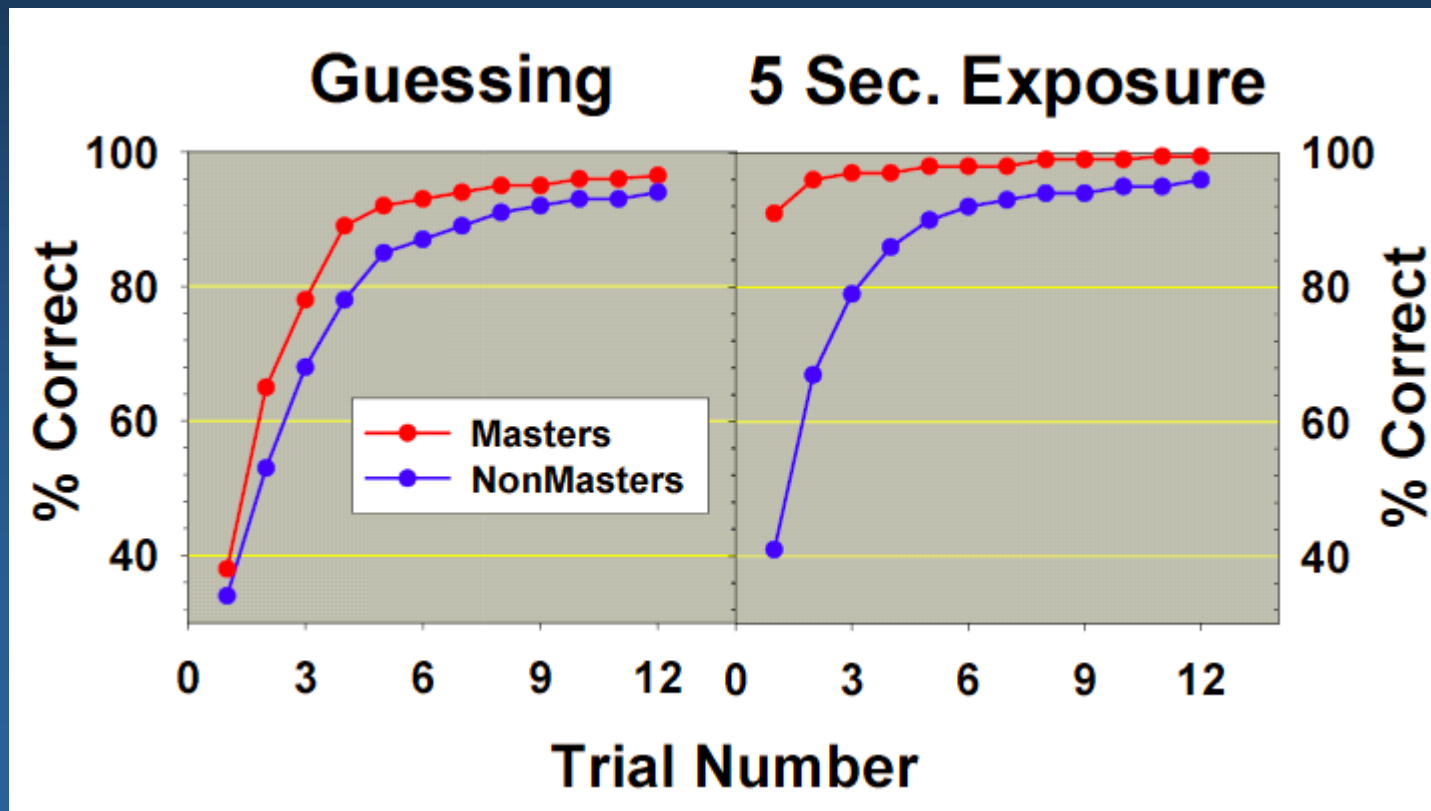


Grupowanie (chunking).

Prezentacje 5 sekund, przypadkowe ułożenie i rzeczywista partia.

Przypadkowe: nowicjusze i mistrzowie robią dużo błędów.

Rzeczywista partia: nowicjusze robią podobną liczbę błędów jak dla przypadkowego rozłożenia, mistrzowie tylko kilka mało istotnych.



Szukanie a ludzkie myślenie.

Mała pamięć robocza człowieka (ok. 7 lub tylko 4 obiektów) wymaga ciągłego „porcjowania”, (ang. chunking), czyli rozbijania złożonych struktur na mniejsze i łatwe do zapamiętania fragmenty.

Działa to trochę jak kompilacja przyrostowa, zastępujemy kawałek kodu skomilowanym modułem.

Szachy: uczenie się + pamięć i rozpoznawanie wzorców, ocena sytuacji oparta na pamięci ograniczająca wybór ruchów.

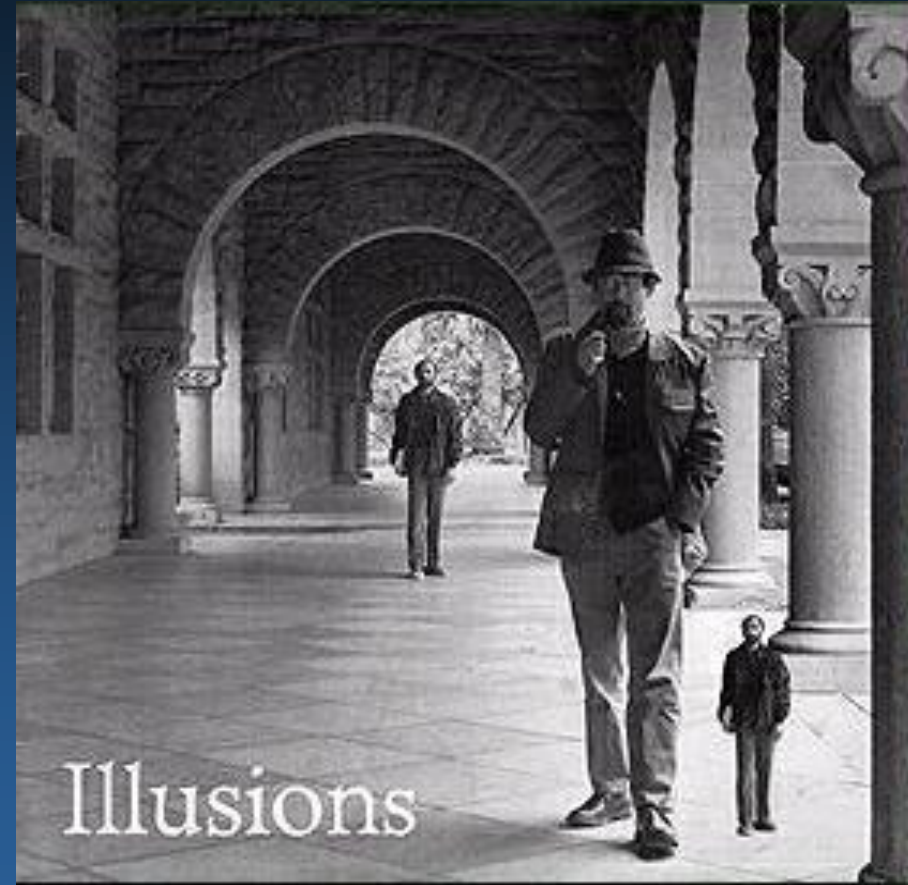
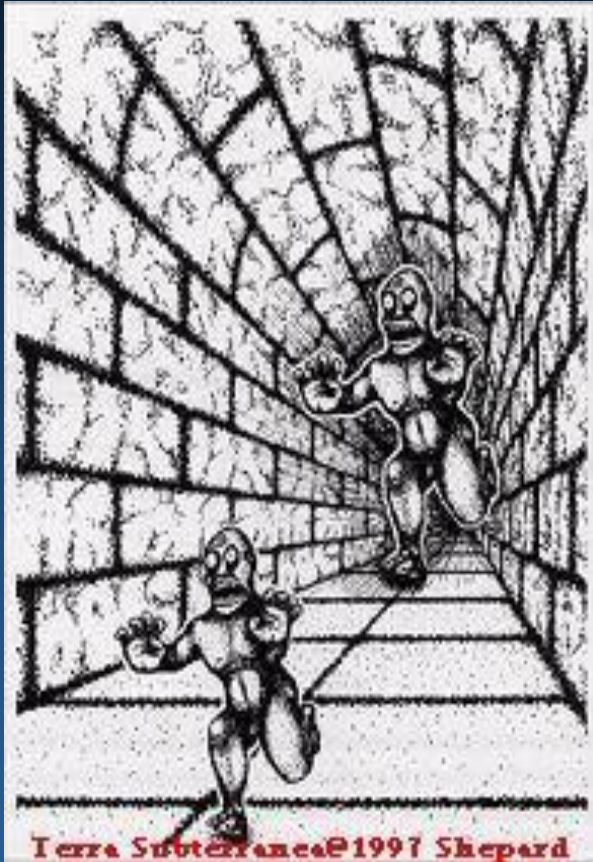
Uwaga skupiona jest (rozwijana jest gałąź grafu) na obszarach słabszych, ocena wzrokowa wzorców struktur.

Ekspert lepiej rozpoznaje i lepiej pamięta, ale tylko istotne struktury, znane z doświadczenia, a nie struktury przypadkowe.

Porcjowanie jest wykorzystywane w mnemotechnice.

Inteligencja jest ściśle związana z sytuacją, domeną wiedzy w której znane są nam wzorce. Ogólny czynnik inteligencji g jest wynikiem współdziałania pamięci, symbolizacji, oraz procesów szukania.

Złudzenia zmysłowe



Zmysły często nas zwodzą, np. na tych obrazkach.

Złudzenia kognitywne

Przeczytaj zdanie tylko raz i policz ile jest liter F w tym zdaniu.

Policz je TYLKO RAZ, nie cofaj się i nie powtarzaj liczenia.

Potem podnieś rękę i pokaż na pacach ile policzyłeś.

Prosty test

FINISHED FILES ARE THE RESULT OF YEARS OF SCIENTIFIC STUDY COMBINED WITH THE EXPERIENCE OF YEARS.

Paradoksy ludzkiego myślenia

Myślenie przebiega schematycznie, dlatego możliwe są złudzenia poznawcze.

Przykłady wnioskowań, łatwe, trudniejsze i b. trudne (niemożliwe?)

Każdy człowiek jest ssakiem.

Sokrates jest człowiekiem.

Wniosek: Sokrates jest ssakiem.

Schemat: $A \Rightarrow B$, $C \Rightarrow A$, więc $C \Rightarrow B$

Żaden rolnik nie jest żeglarzem.

Wszyscy Tybetańczycy to rolnicy.

Wniosek: żaden Tybetańczyk nie jest żeglarzem.

Schemat: $\sim \exists A \Rightarrow B$, $C \Rightarrow A$, więc $\sim \exists C \Rightarrow B$

Niemożliwy ?

Wszyscy członkowie gabinetu to złodzieje.

Żaden muzyk nie jest członkiem gabinetu.

Co można **jednoznacznie** powiedzieć o relacji muzycy $\Leftrightarrow$ złodzieje ?

Takie sylogizmy rozważano już w starożytności.

Jest 256 możliwości ale tylko 24 są poprawne.

<https://pl.wikipedia.org/wiki/Sylogizm>

Czemu jest tak trudno przeprowadzić pozornie proste rozumowanie?

To ważne pytanie w kontekście dyskusji o powielaniu ludzkich błędów w systemach AI, np. trzymaniu się stereotypów (algorithmic bias).

AI może się pozbyć uprzedzeń, ale czy ludzie to potrafią ?

Błędne modele mentalne trudno jest zmienić. Za badania nad formami irracjonalnego myślenia, ekonomią behawioralną, Daniel Kahneman dostał w 2002 roku nagrodę Nobla.

Test karciany (Wason 1960)

Na jednej stronie kart są cyfry, na drugiej litery.

Które karty należy obrócić, by sprawdzić prawdziwość reguły:

- jeśli jest samogłoska to cyfra jest parzysta



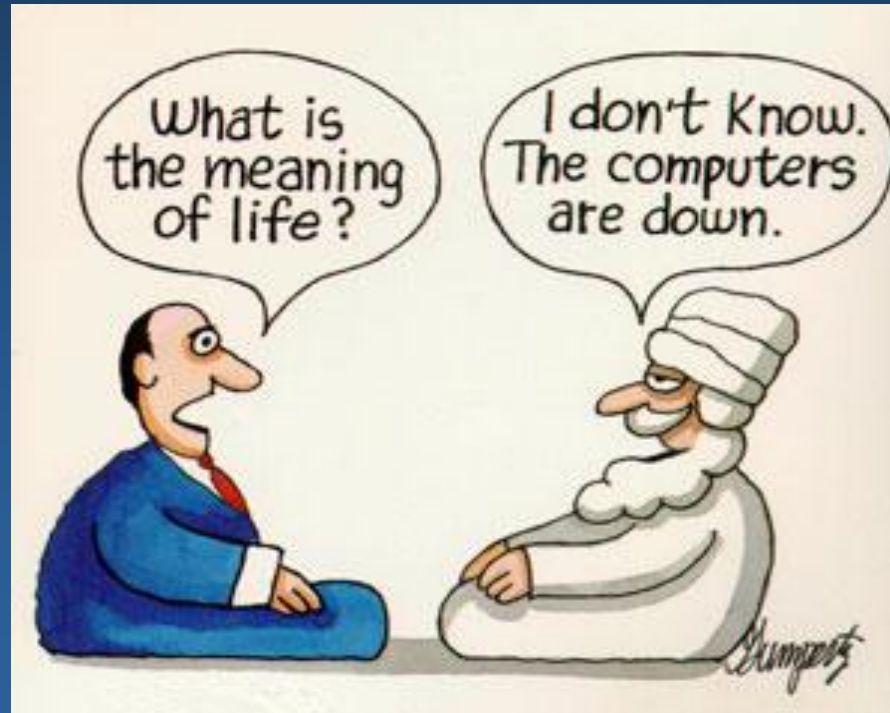
Ile minimalnie kart trzeba przewrócić?

Test Wasona na kilka sposobów.

Wnioski

Myślenie jest rzeczą trudną ...
prościej jest używać schematów.

Tylko w kontekście naturalnych sytuacji myślenie
przychodzi nam łatwo.



Behawioralny test Turinga

Behawioralny test Turinga: czy rozpoznamy zachowanie człowieka lub zwierzęcia od AI?

Wyzwanie: sztuczny szczur we wrogim środowisku.

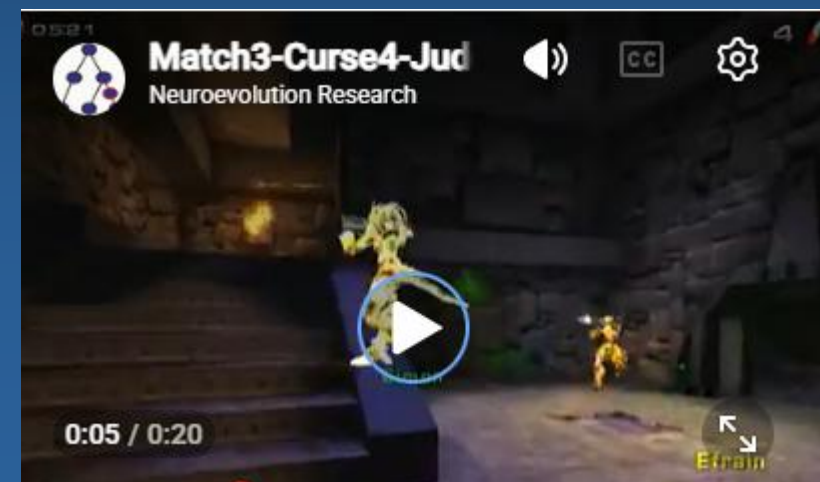
Inny rodzaj testu Turinga: czy walczę z człowiekiem czy z programem?

Botprize: w 2012 sterowanie postaciami z Unreal Tournament 2004 przez AI zostało uznane za sterowanie przez ludzi. Winner of the 2012 BotPrize: UT^2.

Po 5 latach od rozpoczęcia w 2012 roku; dwie drużyny (3 osoby z Univ. of Texas at Austin i Mihai Polceanu, rumunski student z Brest, Francja) przekonały sędziów, że ich bot jest sterowany przez człowieka.

Sterowanie dronami w czasie rzeczywistym – wygrywa AI (wideo, Nature, 30.08.23)

Our world in data on AI ma interesujące dane statystyczne o AI.



Teorie poznania



- Mózgi to jedyne systemy naprawdę inteligentne, AI powinno się więc wzorować na naturalnych systemach i z nimi współdziałać.
- Jak zrozumieć działanie mózgu/umysłu?
- Fizyka to modelowa teoria świata, ma różne gałęzie, stosuje różne przybliżenia, ale nadal brak jest jednej teorii wszystkiego (ciągle są marzenia o TOE).
- Potrzebujemy wielu teorii poznania, w zależności od zjawisk, które opisujemy i poziomu opisu, który nas zadowoli.
- Należy odpowiedzieć na konkretne pytania dotyczące rozwiązywania problemów, podejmowania decyzji, pamięci, uczenia się, sprawności motorycznej, percepcji, języka, motywacji, emocji, wyobraźni, śnienia, halucynacji ...

Teorie zunifikowane

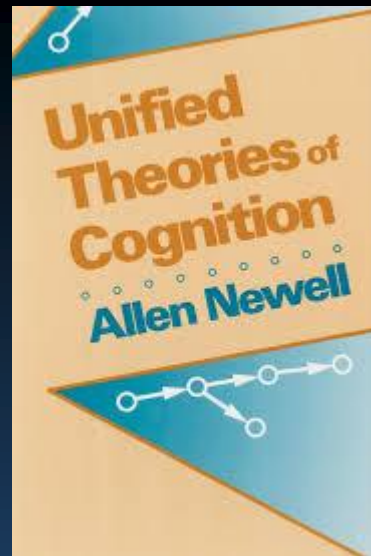
- Zunifikowana teoria poznania? Tak ale to naprawdę trudne ...
GUT czy OTW w fizyce to nadal kwestia przyszłości.

Modele zunifikowane powinny wykazywać:

- zdolność adaptacji do środowiska,
- racjonalne zachowanie celowe,
- działać w czasie rzeczywistym,
- skupiać i kontrolować uwagę (wybierać istotne informacje),
- używać obszernej wiedzy (rozumienie kontekstowe),
- kontrolować działania systemu, np. ruchy agenta,
- używać symboli i metasymboli, w miarę naturalnego języka,
- uczyć się spontanicznie, rozwijać nowe zdolności,
- działać autonomicznie, współdziałać z innymi,
- posiadać poczucie „ja” a może i coś więcej ?

Świadomość nie musi być w takim modelu istotnym problemem.

Na ten temat jest wiele bezsensownych teorii niczego nie wyjaśniających.



Modele umysłu - wymagania

W modelach realizowanych w praktyce możliwe jest uwzględnienie tylko niewielkiej części tych życzeń.

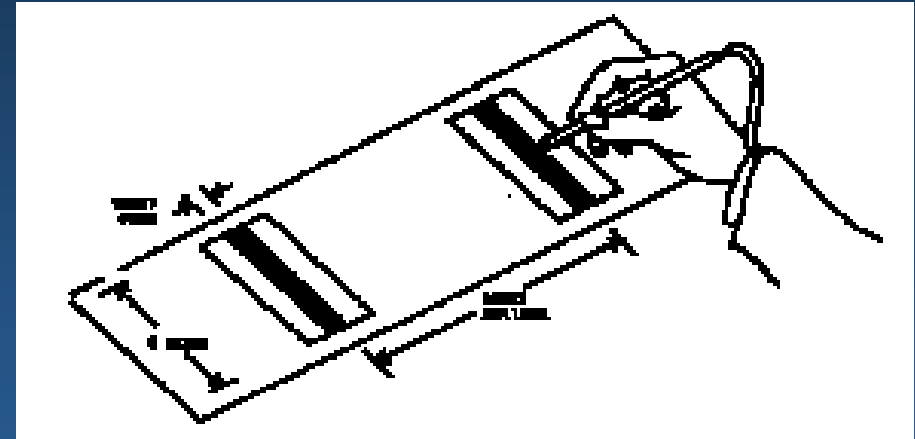
Modele kognitywne muszą spełniać wiele ograniczeń:

- Powinny być realizowalne w postaci sieci neuronowych (choć większość modeli psychologicznych jest tylko „w zasadzie” realizowalna).
- Dać się stworzyć na drodze (symulowanej) ewolucji, wzrostu embrionalnego i stopniowego rozwoju (evo-devo).
- Nie powinny być zbyt doskonałe! Umysł to nie maszyna Turinga, robi różne błędy (choć niektórzy sądzą, że jest doskonały).
- Modele powinny wyjaśniać liczne obserwacje psychologów poznawczych dotyczących zachowań ludzkich.

Prawo Fittsa

Przykład obserwacji psychologicznych: prawo Fittsa (1954).

Przesuwając palec lub wskaźnik myszy z jednego miejsca na drugie, o rozmiarach S odległe o D od punktu startu, człowiek potrzebuje czasu, który jest proporcjonalny do $t \sim \log(D/S)$.



Dokładność tej zależności to ok. 10%, niemal niezależnie od badanego. Ma to liczne konsekwencje dla projektowania interfejsów.

http://en.wikipedia.org/wiki/Fitts'_law

Prawo uczenia się

Potęgowa zależność czasu reakcji od liczby prób/powtórzeń.

Czas reakcji T po wykonaniu N prób daje się dopasować z dużą dokładnością, niezależnie od człowieka, do krzywej potęgowej

$$T = A * N^{-\alpha}.$$

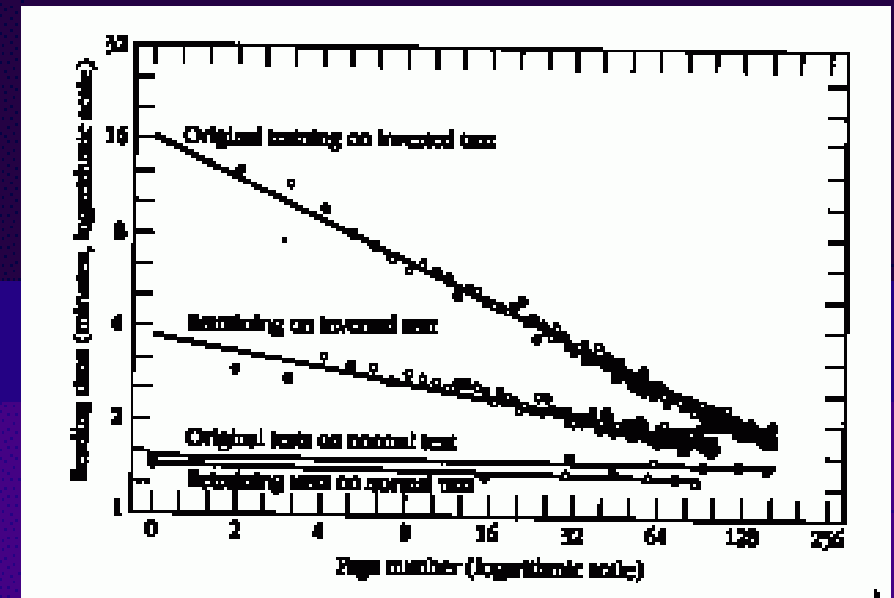
Przykłady:

- Badanie czasu reakcji na naciskanie 10 przycisków odpowiadających 10 lampkom, które zapalają się pokazując różne wzory.
- Zwijanie cygar i inne umiejętności manualne.
- Nauka czytania odwróconego tekstu.
- Wiele innych, sami poszukajcie.

[Wiki Power Law of Practice](#)

Examples: Reading Inverted Text (Kolars & Perkins, 1975)

- The Power Law of Practice Again
- Savings When Relearning a Skill



Umysty i wiedza

Symboliczne podejścia do modelowania umysłu
(Unified Theories of Cognition, A. Newell 1990).

Definicja:

Umyst jest systemem kontrolnym, określającym zachowanie się systemu przy oddziaływaniach ze złożonym, zmiennym w czasie środowiskiem.

Umyst = zbiór wielu współdziałających ze sobą systemów, ale definicja dotyczy nie samych systemów, tylko poziomu funkcjonalnego.

Zbiór reakcji (response functions), działanie kooperatywne.

Umyst działa w oparciu o zgromadzoną wiedzę.

Opis działania systemu w oparciu o zgromadzoną wiedzę jest użytecznym przybliżeniem do sposobu działania prawdziwego umysłu.

SOW

SOW, system oparty na wiedzy (knowledge-based system) $\approx$ umysł.

SOW: stawia sobie pewne cele i podejmuje działania korzystając z wiedzy i gromadząc nową.

System intencjonalny (Brentano) – to system, którego działania i symbole są „o świecie”, a nie „o systemie”, czyli formalnych relacjach wewnątrz systemu.

Symbol: klasa abstrakcji pozwalająca na sprecyzowanie wiedzy.

Znaki drogowe, diagramy, obrazy, słowa, symbole chemiczne odwołują się do wiedzy o świecie.

Reprezentujemy wiedzę za pomocą procesów obliczeniowych, reguł, ram, sieci semantycznych, sieci neuronowych umożliwiających skojarzenia.

Systemy symboliczne mają moc uniwersalnej maszyny Turinga, mogą realizować dowolne procesy obliczeniowe.

Mózgi posługują się znacznie bardziej abstrakcyjnymi strukturami reprezentując wiedzę o świecie w sposób rozproszony. Sztuczne sieci neuronowe są pod tym względem dalekie od biologicznych.

Poziomy realizacji modeli

Poziom:	SOW	Umysły
Substrat:	Wiedza	Świat wewnętrzny
Prawa:	Zasady racjonalnego działania	Prawa psychologii

Poziom:	Systemy oprogramowania	Zachowania wyuczone
Substrat:	Struktury danych i programy	Neurodynamika
Prawa:	Interpretacja syntaktyczna instrukcji	Dynamika złożonych układów

Poziom:	Uniwersalny komputer	Mózg
Substrat:	Ciągi bitów	Stany neuronów
Prawa:	Arytmetyka binarna	Reguła Hebba

Poziomy realizacji modeli 2

Poziom:	Architektura sprzętowa	Przetwarzanie sygnałów
Substrat:	Obwody logiczne	Moduły neuronów
Prawa:	Logika	Neurofizjologia

Poziom:	Obwody elektryczne	Neurony
Substrat:	U/I/zjawiska elektryczne	Zjawiska elektryczne
Prawa:	Ohma, Kirchhoffa, Faradaya	Ohma, Kirchhoffa, Faradaya

Poziom:	Obwody scalone	Biochemiczny
Substrat:	Atomy, elektrony, półprzewodniki	Neurochemia
Prawa:	Fizyka ciała stałego	Fizyka molekularna

SOW i symbole

SOW oddziałują z środowiskiem, wykonują akcje, kontrolują zachowania.

Wiedza – to substrat przetwarzany przez SOW, określa cele działania.

Zadanie SOW: podejmować działania by spełnić swoje cele korzystając przy tym w pełni z posiadanej wiedzy.

Wiedza zawarta jest w symbolach i wzajemnych relacjach, np. szyk słów.

Fizykalne systemy symboliczne zawierają:

Symbole, to powtarzające się wzorce jakiegoś substratu, wskazujące na elementy pamięci lub inne struktury.

Pamięć, to struktury złożone ze znaków symbolicznych.

Operacje, to procesy działające na strukturach symbolicznych i produkujące inne struktury symboliczne + procesy interpretujące struktury symboliczne z punktu widzenia zachowania się systemu, prowadzące od struktur do zachowania się systemu.

Symbole i inteligencja

Systemy symboliczne to praktyczna realizacja SOW.

Semantyczne znaczenie jest wynikiem oddziaływania systemu ze środowiskiem; symbole nabierają sensu poprzez działanie.

Zadaniem AI jest poszukiwanie przybliżeń do SOW w oparciu o systemy symboliczne. Symbole to nie tylko słowa, mogą to być dowolne wzorce.

Architektura = struktura całości, realizująca działanie systemu symbolicznego.

Nie wystarczy sama informacja np. zawarta w DNA, muzyce czy filmie, potrzebna jest jeszcze skomplikowana architektura systemu, który potrafi tę informację odczytać i odtworzyć.

Co jest potrzebne by odtworzyć informację z DNA, zbudować organizm? Złożone środowisko.

Wiedza w parametrach LLM wymaga też środowiska by ją wydobyć.

Inteligencja

Inteligencja: pojęcie kontrowersyjne, szczególnie IQ. Jest >70 definicji inteligencji.

Wszystkie zadania, których nie można efektywnie rozwiązać przy pomocy algorytmów, wymagając inteligencji.

System używający całej dostępnej mu wiedzy i wyciągający z niej wszystkie wnioski jest **doskonale inteligentny**.

Brak wiedzy to nie braku inteligencji; posiadanie wiedzy, niemożność jej użycia to wynik braków inteligencji.

Definicja inteligencji: inteligencja to zdolność do wykorzystania wiedzy do osiągnięcia stojących przed systemem celów.

Inteligencja zależy więc od wiedzy i celów: w osiągnięciu jakiegoś celu system może wykazywać doskonałą inteligencję a w osiągnięciu innych celów zerową.

Stąd mamy inteligencję wieloraką (M. Gardner, 8 rodzajów), inteligencję kolektywną grup ludzkich i systemów wieloagentowych.

Przestrzenie problemów

Podstawą inteligentnego zachowania są procesy poszukiwania rozwiązań (dotyczy to „wyższych czynności poznawczych”).

Inteligentne zachowanie wymaga rozważenia możliwych rozwiązań, strategii, oceny i wyboru najlepszego rozwiązania.

Jeśli nie wiadomo, w jaki sposób osiągnąć dany cel, utwórz przestrzeń podproblemów i przeszukuj ją w celu znalezienia drogi do celu.

Należy ustalić reprezentację problemów i celów, ustalić strategię poszukiwania.

LLM-reprezentacja wektorowa, uwzględniająca kontekst.

Komputery i umysły

Powszechne przekonanie: komputery liczą, przeszukują a ludzie dokonują świadomych wyborów optymalnych? Dzięki jakiejś magii?

Systemy inteligentne mają różne ograniczenia, zależnie od swojej konstrukcji.

Relacja między ilością wiedzy a szybkością przeszukiwań; krzywe stałej kompetencji (na następnej stronie). Komputery szybko szukają, mózgi znacznie wolniej.

Pamięć biologiczna uaktywnia tysiące reprezentacji złożonych pojęć i sytuacji jednocześnie, ale pamięć robocza jest niewielka.

Zgromadzenie obszernej wiedzy wymaga uczenia się na przykładach lub sposobów reprezentowania wiedzy w jakieś postaci, np. reguł lub grafów wiedzy.

Systemy symboliczne otrzymują gotową wiedzę.

Przygotowanie wiedzy nadającej się do procesów rozumowania, szukania rozwiązań, wymaga kompresji. Od percepcji, analizy sygnałów, do symboli dochodzimy przez uczenie maszynowe, sieci neuronowe, analizy probabilistyczne i statystyczne.

Wiedza i szukanie

Konieczność rozważenia wielkiej liczby wariantów by znaleźć optymalne rozwiązanie problemu, vs. wiedza, liczba reguł, które dany system zna.

Wiedza człowieka-eksperta to ok. 10^4 - 10^5 reguł/kroków.

Hitech - 64 proc, 175Kp/sek.

10 reguł - 1900 punktów;

100 reguł - 2360 punktów.

Technologia neurobiologiczna daje jeden typu umysłu, technologia półprzewodnikowa całkiem inny.

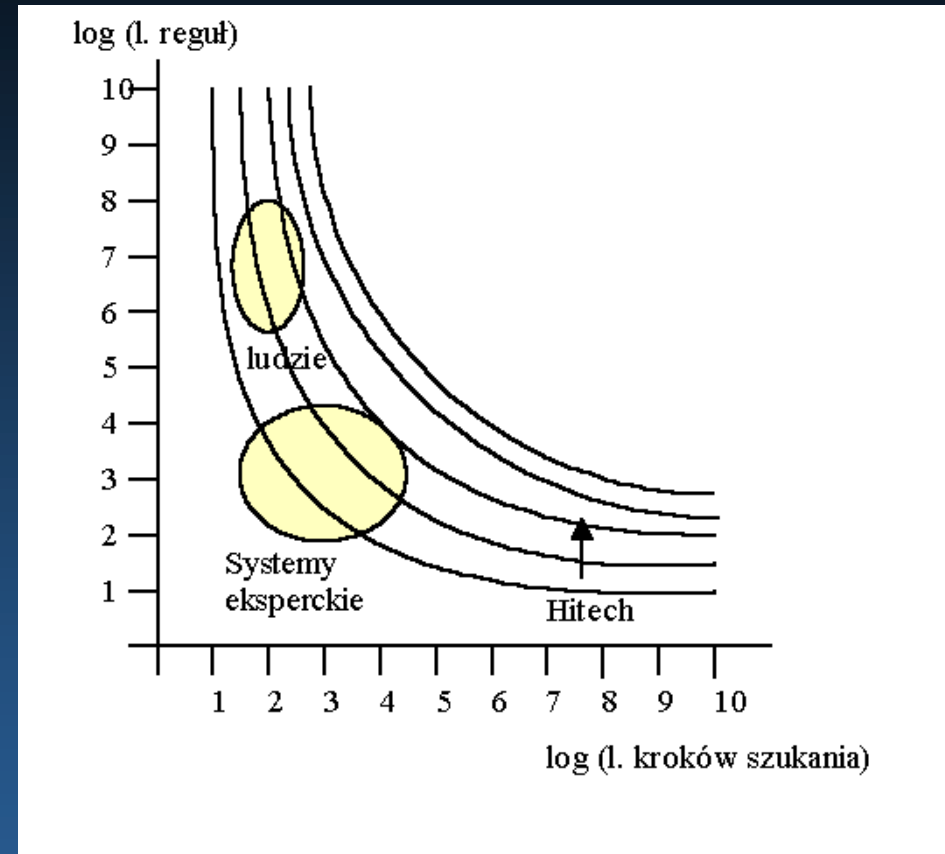
Jakie umysły mogą powstać w różnych substratach?

Co nazwiemy umysłem? Zgodnie z definicją Newella,

umysł to system kontrolny określający działanie w środowisku – tu bardzo prostym.

Jakie powinno być środowisko (kontekst) by można taki system nazwać umysłem?

Ile musi być drzew by nazwać je lasem?



Architektura umysłu

SOW, systemy symboliczne: przybliżenie systemów inteligentnych.

Podejście hierarchiczne: system jako zbiór współdziałających agentów, złożonych i będących częścią systemu wyższego rzędu.

Stabilność działania złożonych systemów wymaga hierarchii.

- **Skale czasowe:**

neurony: mikro/milisek, procesy kooperatywne > 10 milisek;

motoryka/percepcja 0.1-10 sek;

procesy racjonalne - minuty, godziny, dni ...

Komunikacja modułów neuronowych: 1 cm, 10 ms.

1 sek = 100 kroków układu nerwowego.

Procesy automatyczne – szybkie, współbieżne, System 1.

Procesy kontrolowane – wolne, mechanizm sekwencyjny, System 2.

Uczenie: złożone zadania stają się dzięki hierarchii prostymi.

Pytania: szukanie i myślenie

- Jaką strategię stosują ludzie w grze w szachy?
- Co umożliwia ludziom sprawne działanie pomimo niewielkiej pojemności pamięci roboczej? Jakie inspiracje dla AI z tego wynikają?
- Wszyscy A to B. Żaden C nie jest A. Jaki stąd wniosek?
- Oszacuj liczbę operacji wykonywanych przez mózg Kasparowa i wytłumacz, dlaczego przegrał z systemem Deep Blue.
- Jak zależy kompetencja od szybkości szukania i wielkości bazy wiedzy.
- Co powinny uwzględniać zunifikowane teorie poznania?
- Jakie wymagania stoją przed teorią umysłu?
- Podać przykład praw, które teorie umysłu powinny wyjaśniać.
- Definicja umysłu/inteligencji wg. Newella.
- Poziomy realizacji modeli: co można ze sobą porównać?
- Symbole, systemy symboliczne i systemy oparte na wiedzy.
- Architektura systemu symbolicznego.