

BRAVE NEW DIGITAL WORLD

3. UMĚLÁ INTELIGENCE

UMĚLÁ INTELIGENCE

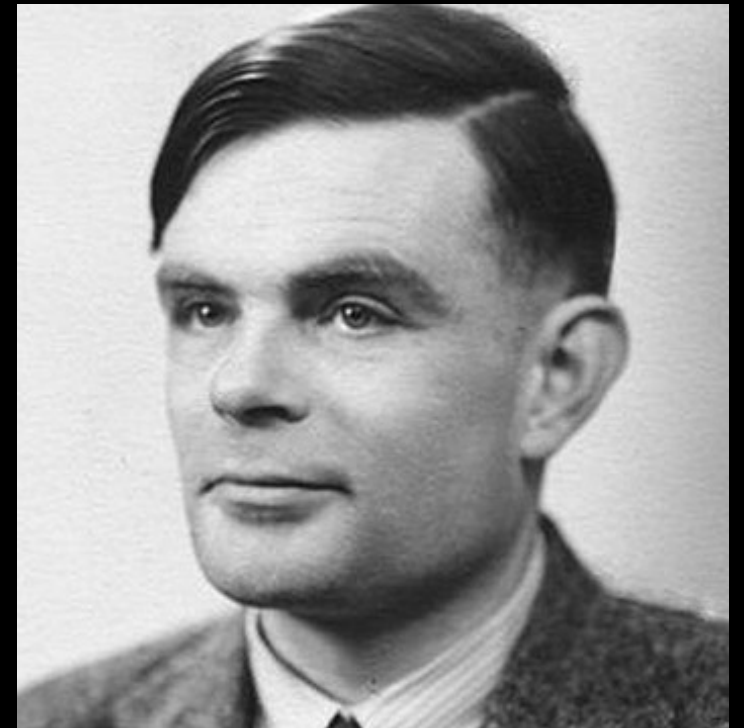
- * „vytvořit mechanismus/entitu, jehož schopnosti přesahují ty lidské“
- * 1) augmentace – posílení člověka (např. [Engelbart a GUI-HCI](#))
- * intelekt in silico – algoritmizace komplexních mentálních operací
- * 3) propojení – propojování lidských myslí zvyšuje inteligenci (Licklider)



„My invention contains the application of all reason, a judgment in each controversy, an analysis of all notions.“



„The Engine might compose elaborate and scientific pieces of music of any degree of complexity or extent.“



„At the end of the century /.../ one will be able to speak of machine thinking without expecting to be contradicted.“

INSTITUCIONALIZACE AI JAKO VĚDY

* 1950: Alan Turing v časopise Mind zveřejňuje článek Computing Machinery and Intelligence – jevit se inteligentním je inteligencí

* 1956: Dartmouth Summer Research Project on Artificial Intelligence

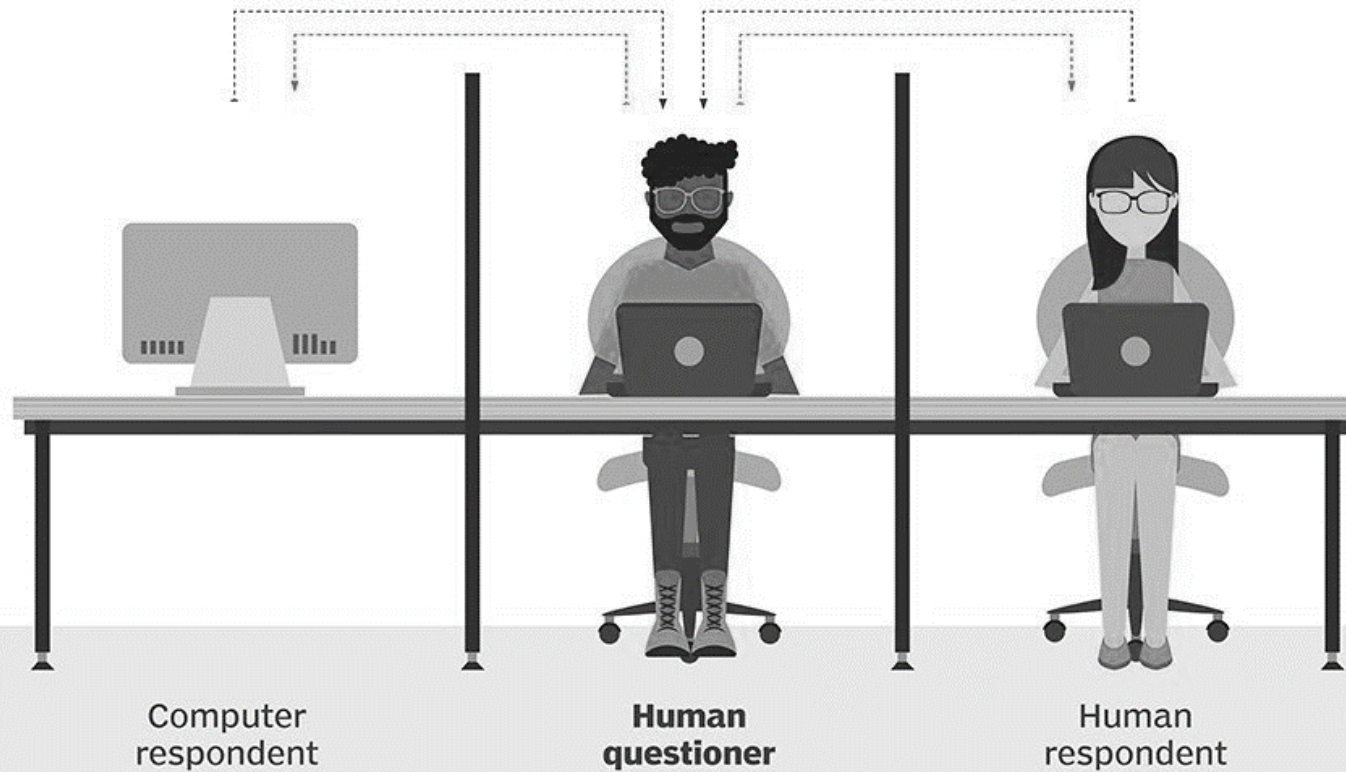
„Studium bude pokračovat na bázi domněnky, že každý aspekt učení nebo jakýkoliv rys intelligence může být v principu popsán tak přesně, že ho stroj dokáže simulovat. Snahou bude najít způsob, jak přimět stroje používat jazyk, vytvářet abstrakce a koncepty, řešit takové problémy, jaké jsou doposud vyhrazeny lidem a samy sebe zlepšovat“

<McCarthy, Minsky, Shanon, Rochester...>

Turing test

During the Turing test, the human questioner asks a series of questions to both respondents. After the specified time, the questioner tries to decide which terminal is operated by the human respondent and which terminal is operated by the computer.

■ QUESTION TO RESPONDENTS ■ ANSWERS TO QUESTIONER



INSTITUCIALIZACE AI JAKO VĚDY

- * První laboratoř AI vzniká v r. 1958 na MIT, v r. 1963 pak na Stanfordu; výzkum probíhá rovněž na Carnegie Mellon University (McCarthy, Minsky a Simon & Newell)
- * 1958: jazyk LISP, systém time-sharing
- * výzkumné výsledky zajišťují štědré dotace amerických armádních agentur

VELKÉ ÚSPĚCHY – VELKÁ OČEKÁVÁNÍ

LOGIC THEORIST:

Program dokázal 38 teorémů z *Principia Mathematica*, u tzv. oslího můstku poskytl elegantnější důkaz než Bertrand Russel

STUDENT:

Program Daniela Bobrowa dokázal porozumět zadání středoškolských slovních úloh, které následně vyřešil

ANALOGY:

Program Toma Evanse úspěšně řešil vizuální oddíly standartních IQ testů

1955

1958

1964

1965

1968

1970

ALEN NEWELL:

„Within ten years a digital computer will be the world's chess champion" and will discover and prove an important new mathematical theorem“

HERBERT SIMON:

„Machines will be capable, within twenty years, of doing any work a man can do“

MARVIN MINSKY:

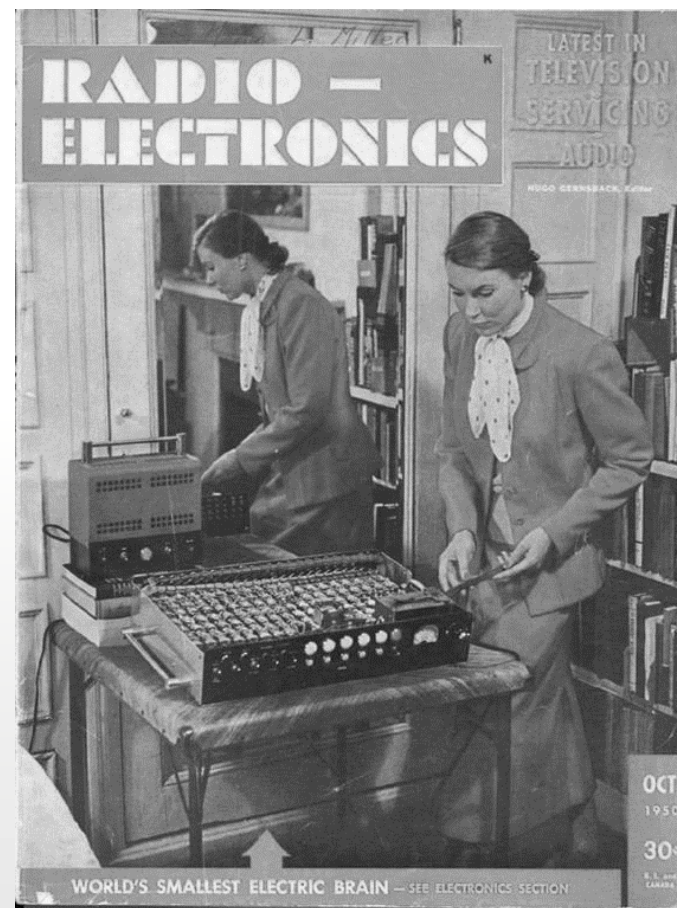
„In from three to eight years we will have a machine with the general intelligence of an average human being. “



NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo
of Computer Designed to
Read and Grow Wiser

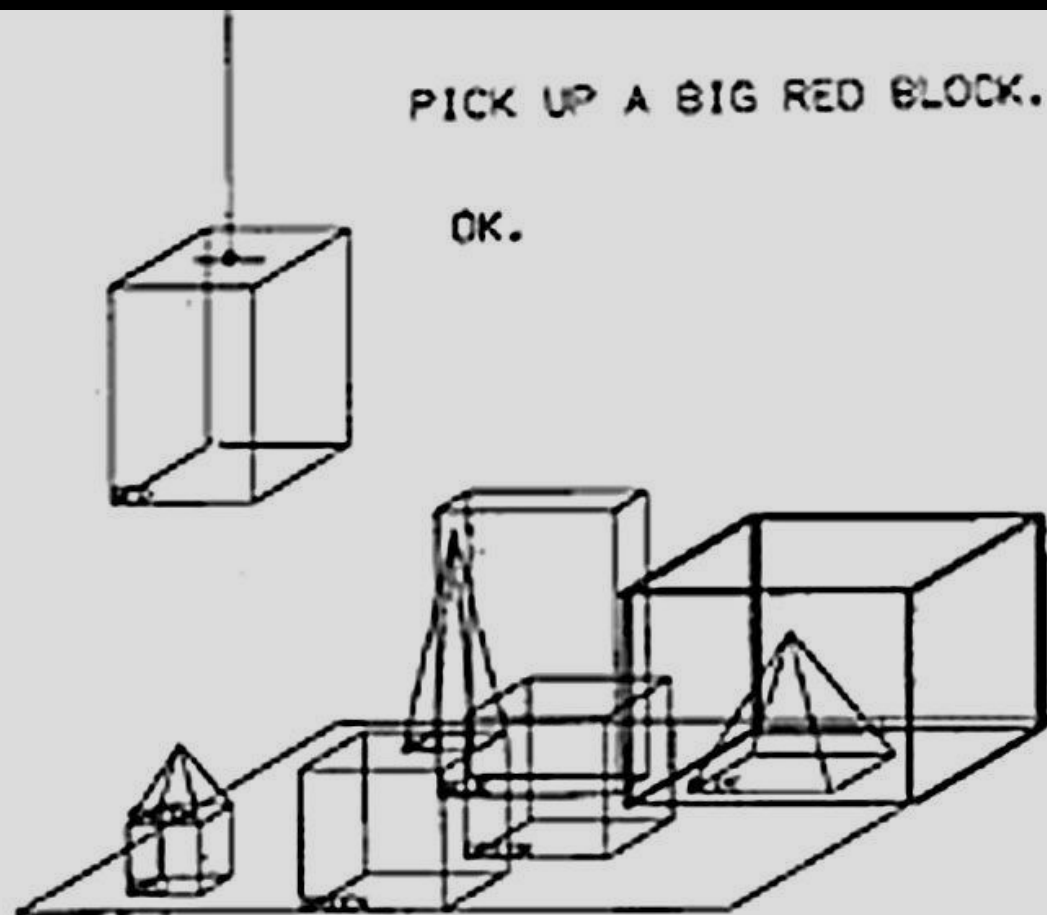
WASHINGTON, July 7 (UPI)
—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.



KONEC ZLATÉHO VĚKU

- * hlavní výzkumné cíle v 50's a 60's byly usuzování a zpracování jazyka
- * tzv. Eliza effect
- * programy selhávaly v řešení komplexnějších zadání, byly schopny fungovat pouze v omezených formálně zcela popsaných doménách
- * výzkum AI přestává být v první pol. 70's financován – tzv. AI Winter

hi! i'm eliza. what's your problem?
? I can not find my pen.
are you saying no just to be negative?
? No.
say do you have any psychological problems?
? Have I? Yeah, maybe...maybe not.
?extra ignored
what does that suggest to you?
? It suggests that the cake is a lie.
i see.
? You see what I not see?
why do you ask?
? Well, let me. Just let me...
?extra ignored
i'm not sure i understand you fully
? My turn. Voo voo!



EXPERTNÍ SYSTÉMY

- * McCarthyho koncept: „[Advice Taker] automaticky pro sebe vyvozuje dostatečnou škálu bezprostředních důsledků čehokoliv, co je řečeno a co už zná.“
- * Edward Feigenbaum: DENDRAL, MYCIN, PROSPECTOR; Expertní systémy interpretují informace pomocí IF-THEN pravidel
- * v 80's expertní systémy zaznamenávají velký komerční úspěch (DuPont, DEC, LISP Machines)

EXPERTNÍ SYSTÉMY

- * projekt CYC Douglase Lenata má mít stovky tisíc IF-THEN pravidel, proklamována je schopnost zodpovědět jakoukoliv otázku a provádět vlastní vědecký výzkum
- * selhání expertních systémů – pravidla si vzájemně odporují, relativně málo pravidel rozhodně nestačí všude a na všechno, hardware byl poruchový, software se špatně aktualizoval
- * 1987 druhá AI winter; black monday-DARPA-nástup Apple a PC

AI POTŘETÍ

- * v období od 50's do 90's mluvíme o „staré dobré AI“ (GOFAI);
- * symbolismus: performance inteligentního vychází z dané reprezentace
- * Od 90. let se výrazně prosazují odlišná řešení
- * behaviorismus – inteligentního chování je souhrou vlastností systému, jeho úloh a jeho vnějšího prostředí (tzv. new AI)
- * konektivismus – inteligentní chování vzniká emergentní reakcí velkého počtu jednoduchých výpočetních jednotek (neuronové sítě)

MORAVCŮV PARADOX A NEW AI

* „/.../ chování stran řešení problémů, jazyk, expertní znalosti a jejich aplikace nebo rozum, ty jsou všechny docela jednoduché, jakmile je dostupná esence bytí a reagování. Tato esence je schopnost pohybu v dynamickém prostředí a zakoušení okolí do míry dostatečné pro dosažení potřebné udržitelnosti života a rozmnožování. Tato část intelligence je to, kde evoluce koncentrovala své trvání – je mnohem složitější.“ <Brooks>

* „lidská intelligence vyžaduje lidskou interakci s okolním světem“ <Brooks>



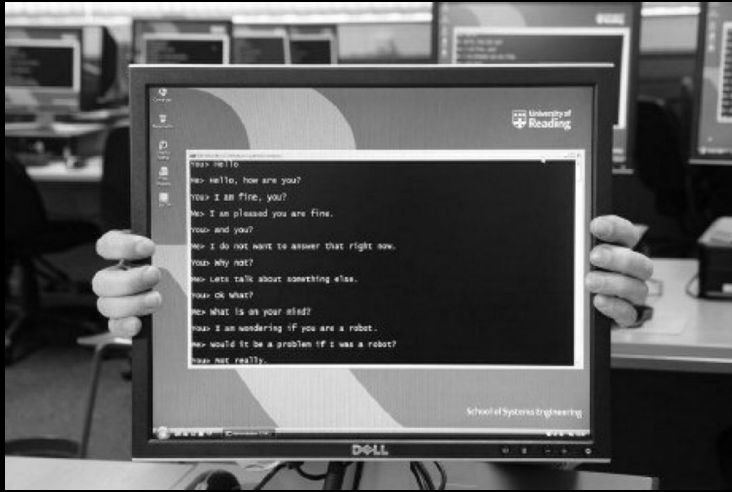
NEURONOVÉ SÍTĚ

- * mechanismus klasifikace a hledání množiny určujících parametrů vycházejících z předcházející datafikace velkého informačního korpusu.
- * výzkum NS probíhal od 40's, v 60's výsledky zpochybnil Marvin Minsky – námitka XOR; tento problém v 80's vyřešil John Hopfield
- * dnes se NS využívají zejména ke zpracování obrazu a jazyka



AI PO DRUHÉ ZIMNĚ

- * Od 90. let výzkum AI začíná být mnohem více aplikovaný a z veřejné sféry se hojně přesouvá do sféry korporátní
- * Výsledkům výrazně pomáhají pokroky v matematice a statistice, působení Mooreova zákona či datová exploze v prostředí internetu
- * Začíná být řešen potenciál negativních dopadů dostupné AI; problematika AGI se příliš neřeší



AI jako filozofický a technický problém

- * technický problém: automatizace aktivit, které asociujeme s lidským myšlením
- * filozofický problém: mohou stroje jednat inteligentně (slabá AI) vs. mohou být inteligentními (silná AI); čínský pokoj
- * Problém antropocentrismu v AI
- * „umělá inteligence křemíkových forem /.../ se projevuje jako specifická inteligence, která se odlišuje od lidské a zvířecí“
<Warwick>

AI jako filosofický a technický problém

- * „Dokáže strojová inteligence vykonávat většinu povolání přinejmenším stejně dobře jako člověk?“
- * 2030 – 10%; 2050 – 50%; 2100 – 90 %;
- * „vznik strojové inteligence lidské úrovně je spíše pravděpodobný“
- * bude lidská úroveň inteligence konečnou instancí?!

ITELEKT IN SILICO

- * „křemíková inteligence“ se jeví jako výhodná – latence/rychlost; paměťová kapacita a možných množství modulů (biologický mozek je kvantitativně omezen)
- * existuje zde potenciál superinteligence
- * Irving J. Good: Speculations Concerning the First Ultraintelligent Machine (1965); „poslední vynález v dějinách lidstva“

SUPERINTELIGENCE

- * problém antropomorfního škálování – namísto Švejk-Einstein je vhodnější interval hmyz-člověk
- * předpokládané schopnosti : zvyšování intelligence (přeprogramování se), vytváření strategií (manipulace, ekonomická produktivita), přírodovědný a technologický výzkum, hacking <Bostrom>
- * superintelligence nebude nerd, co si toho hodně pamatuje

SUPERINTELLIGENCE

- * ortogonální náhled: intelligence a konečné cíle jsou na sobě nezávislé; superintelligence nemusí disponovat jakoukoliv hodnotou či vlastností, kterou spojujeme s moudrostí u lidí
- * instrumentální konvergence: výskyt jistých cílů předpokládat lze (sebezáchova, integrita, získávání zdrojů, zdokonaňování kognice); nemůžeme předpokládat, že by superintelligence svou činnost omezila tak, aby lidem nepřekážela

PROBLÉM KOTROLY

- * pokud má superintelligence adekvátně sloužit, musí existovat mechanismus kontroly
- * A) kontrola schopností (izolace, brždění, nástrahy)
- * B) kontrola motivací (specifikace, normativnost, augmentace)
- * C) funkční specifičnost (věštírna, džin, suverén, nástroj)

STRATEGIE ZVLÁDÁNÍ

- * koherentní extrapolovaná vůle (KEV) jako hodnota pro superinteligenci
- * *Naše KEV je naším přáním, v té podobě, jakou by mělo, kdybychom toho věděli více, mysleli rychleji, byli více takovými lidmi, jakými bychom chtěli být, kdybychom společně dorostli dále. Nachází se tam, kde je extrapolace spíše konvergentní než divergentní, tam, kde do sebe naše přání zapadají, namísto aby si překážela. Je extrapolovaná v souladu s našimi přáními a interpretovaná v souladu s našimi přáními.* <Yudkovsky>
- * tázat se věštírny na suveréna realizujícího KEV => vytvořit suveréna realizujícího KEV?