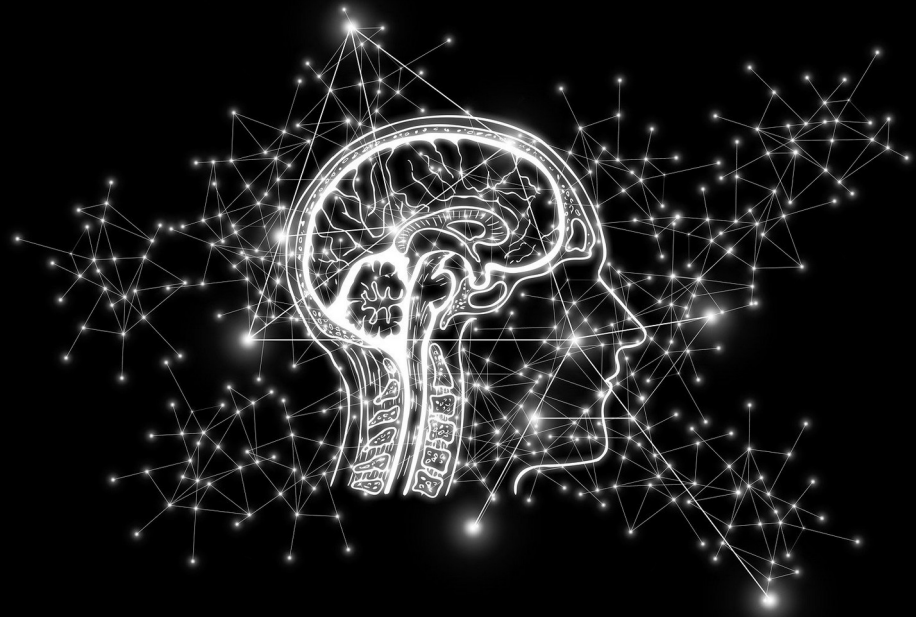




Umelá inteligencia – 70 rokov cesty

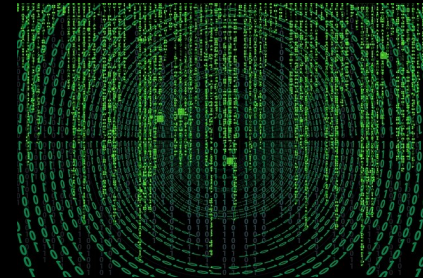
Peter Šantavý

| ...míľniky, úspechy, problémy a výzvy

Prečo umelá inteligencia?

Lebo nám klasické algoritmy nestačia...

- úlohy, ktoré algoritmicky **nevieme riešiť**
- úlohy, ktoré sú extrémne náročné a **nie je v ľudských silách ich zvládnuť** (časovo, organizačne, intelektuálne...)



„V súčasnosti každé dva dni vyprodukuje toľko informácií, koľko sme vytvorili od úsvitu civilizácie až do roku 2003.“ (E. Schmidt, 2010)

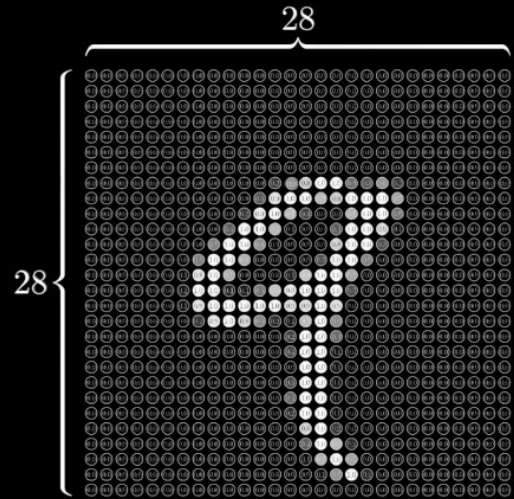
Ked' to algoritmicky nevieme riesit'...

(8, 9) (0, 0) (2, 2) (1, 1) (7, 7) (8, 8) (3, 3) (3, 3)
(0, 0) (4, 4) (6, 6) (7, 7) (4, 4) (6, 6) (1, 1) (0, 0)
(7, 7) (8, 8) (6, 6) (1, 1) (5, 5) (7, 7) (9, 9) (7, 7)
(1, 1) (1, 1) (4, 4) (3, 3) (2, 2) (2, 2) (9, 9) (3, 3)
(1, 1) (1, 1) (0, 0) (4, 4) (6, 6) (0, 0) (0, 0) (6, 6)
(1, 1) (0, 0) (2, 2) (9, 9) (1, 1) (8, 8) (8, 8) (4, 4)
(9, 9) (6, 6) (3, 3) (4, 4) (3, 5) (4, 4) (1, 1) (8, 8)
(3, 3) (8, 8) (5, 5) (4, 4) (2, 7) (7, 7) (4, 4) (2, 2)
(8, 8) (5, 5) (8, 8) (1, 1) (9, 7) (3, 3) (4, 4) (6, 6)
(1, 1) (9, 9) (9, 9) (6, 6) (0, 0) (1, 1) (1, 1) (2, 2)

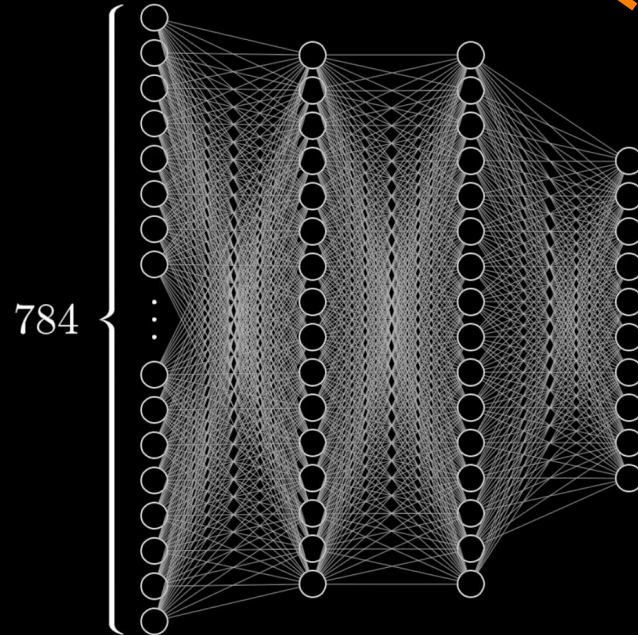


Ak je to t'azké naprogramovať, nech sa to stroje naučia!

Prečo veľký rozvoj AI až teraz?



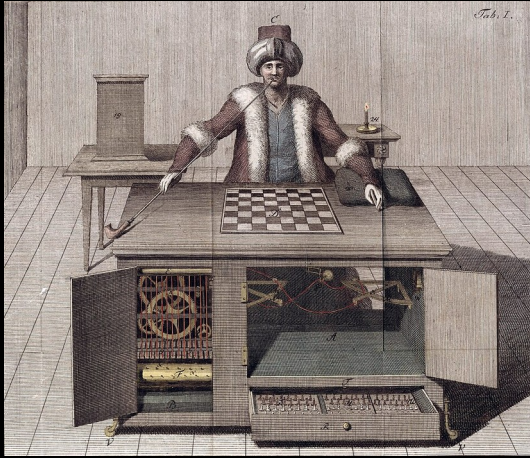
$$28 \times 28 = 784$$



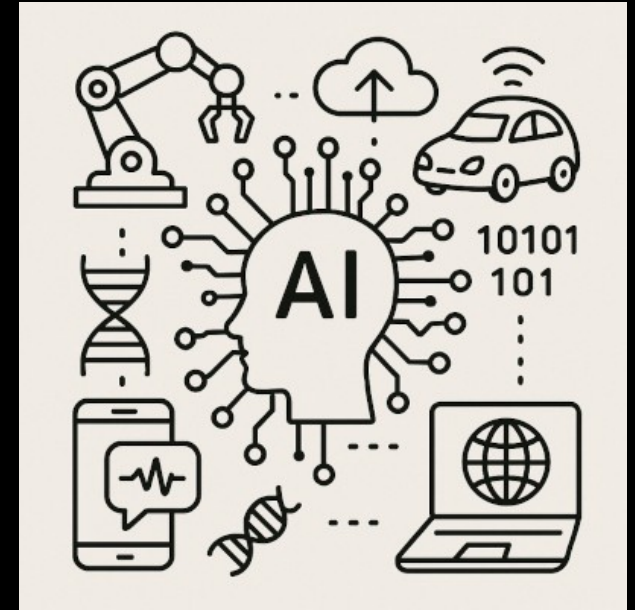
13 002 !!!

Máme algoritmy, máme veľké dáta, máme výkonnú techniku a vieme to využiť...

Túžba zostrojiť mysliace stroje...



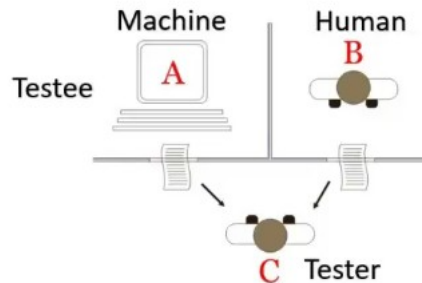
3. priemyselná revolúcia → Industry 4.0



Imitácia mysliaceho stroja → Algoritmy s hrubou silou miesto inteligencie → Systémy AI

Je to vôbec inteligentné?

Turing test of intelligence



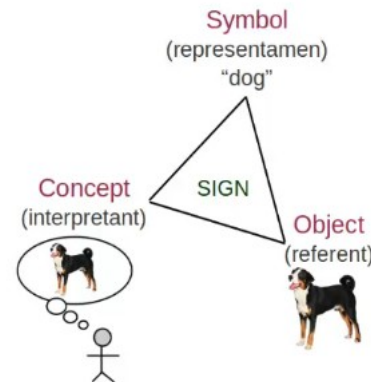
Turing (1950)

Chinese room problem



Searle (1980)

Symbol grounding problem



Harnad (1990)

Distributional models of word meaning →

Vector grounding problem

Mollo & Millière (2023)

Problematika ukotvenia symbolov s testovaním konceptuálneho myslenia a problematika ukotvenia vektorov pri distribučných modeloch významov slov.

I keď nejaký systém dokáže odpovedať tak, akoby konverzácii rozumel, neznamená to, že je inteligentný a že konverzácii aj skutočne rozumie

Dartmouthský seminár – zrod AI

- dvojmesačný výskumný projekt – pracovný seminár v lete 1956
- John McCarthy, Marvin Minsky, Claude Shannon, Nathaniel Rochester
- vynájdenie termínu **umelá inteligencia** (John McCarthy ;-)
- „všetky aspekty učenia, resp. akékoľvek iné črty inteligencie môžu byť v princípe tak detailne popísané, že **je možné vytvoriť stroje, ktoré by inteligenciu mohli simulovať**“
- **hlavné témy**: strojové spracovanie ľudskej reči, neurónové siete, strojové učenie, abstraktné koncepty a myslenie, kreativita...

Nielen biologicko-matematický **veľký problém**, ale i otázky, ktoré sa dotýkajú koexistencie AI a človeka, resp. spoločnosti. Ide teda o otázky psychologické, sociologické a etické...

Snaha definovať pojmy

- nejednoznačnosť pojmov a termínov – **suitcase word** (Minsky):
inteligencia, myslenie, poznanie, vedomie, city,...
- kategorizovanie inteligencie:
 - **binárne** (niečo je alebo nie je inteligentné)
 - **kontinuálne** (jeden objekt je inteligentnejší ako druhý)
 - **multidimenzionálne** (objekt má určitú úroveň inteligencie vo viacerých dimenziách)
- **absencia presnej definície môže byť výhodou...**

Čo to vlastne je inteligencia?

- **Inteligenciu ľudského bytia** vnímame ako:
schopnosť vnímať, chápať a spracovávať informácie, učiť sa, odôvodňovať, plánovať, riešiť komplexné problémy a rozhodovať. Schopnosť tiež používať abstraktné a logické myslenie, predstavivosť a uvažovať o hypotetických možnostiach, jazykový cit, atď.
Inteligencia stelesnená a vzťahová... Má (nielen) kognitívne a metakognitívne schopnosti...
- **Umelá inteligencia** je o vytváraní systému, ktorý vykazuje také správanie, o ktorom si myslíme, že vyžaduje inteligenciu (AI100, 2021).

Funkcionalistická umelá inteligencia vs. komplexná inteligencia ľudskej bytosti!

Základné vlastnosti

- **Autonómnosť** – schopnosť samostatne konať

Schopnosť systému vykonávať úlohy v komplexnom prostredí bez neustáleho vedenia používateľom.

- **Adaptabilita** – schopnosť sa prispôsobovať

Schopnosť zlepšovať svoj výkon (a schopnosti) učením sa (nielen) zo skúseností.

Základné delenie – ANI a AGI

- **Slabá umelá inteligencia (ANI)** – úzko špecializované systémy umelej inteligencie (**narrow** AI), ktoré sú optimalizované na zvládnutie konkrétnej úlohy, resp. množiny úloh.
Ide súčasne o systémy **slabej** umelej inteligencie (**weak** AI), ktoré vykazujú inteligentné správanie na základe modelov a aplikovaných metód i dát, na ktorých sa učí, resp. je natrénovaná.
- systémy ANI sú zamerané na riešenie **konkrétnych** úloh a sú **závislé** na ľudskom vstupe a konfigurácii.

Základné delenie – ANI a AGI

- **Všeobecná umelá inteligencia (AGI)** – tzv. silná (**strong**) a všeobecná (**general**) umelá inteligencia.

Všeobecná, lebo dokáže zvládnuť **akúkoľvek** intelektuálnu úlohu a má **schopnosť zovšeobecňovať** a prenášať, či adaptovať naučené schopnosti na iné úlohy.

Silná, pretože aj **skutočne rozumie** tomu, čo rieši a vykonáva.

- Rozlišujeme **rôzne stupne AGI**:

BAGI (below-human), HLAGI/HAGI (human-level AGI), MAGI (moderately-superhuman AGI), SAGI/ASI (superintelligent AGI)

Základné delenie – symbolická AI

- **Symbolická AI** – napodobňuje ľudské myslenie na úrovni pojmov, slov, fráz (= symboly) a vzťahov medzi nimi.
Na základe definovaných pravidiel a postupov („ak niečo, tak potom toto“) sú jednotlivé symboly spracovávané a vykonávajú sa priradené úlohy.
- Symbolická AI – veľmi zjednodušene povedané – **pomocou matematickej logiky sa snaží emulovať procesy myslenia.**
- Čisto symbolické systémy **zlyhávajú** pri riešení problémov, ktoré sa nedajú exaktne popísať a v reálnych prostrediach, ktoré nie je možné deterministicky uchopiť a sú plné nejasných informácií.

Symbolické systémy sa v prvý dekádach javili ako najschodnejšia cesta k AI.

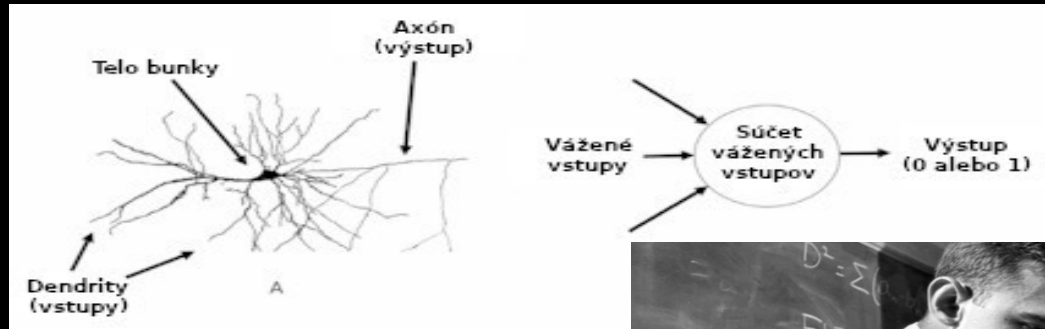
Základné delenie – subsymbolická AI

- **Subsymbolická AI** – napodobňuje myšlienkové procesy, ktoré by sme mohli nazvať niekedy nevedomými, či automatickými, a ktoré sú základom rýchleho vnímania (fast perception; rozpoznávanie tvárí, identifikácia hovorených slov,...).

Subsymbolická AI – tiež zjednodušene povedané – sa v mnohých prípadoch snaží emulovať činnosť mozgu na úrovni neurónov. **Subsymbolické systémy sú navrhnuté tak, aby sa učili vykonávať úlohy na základe dát.**

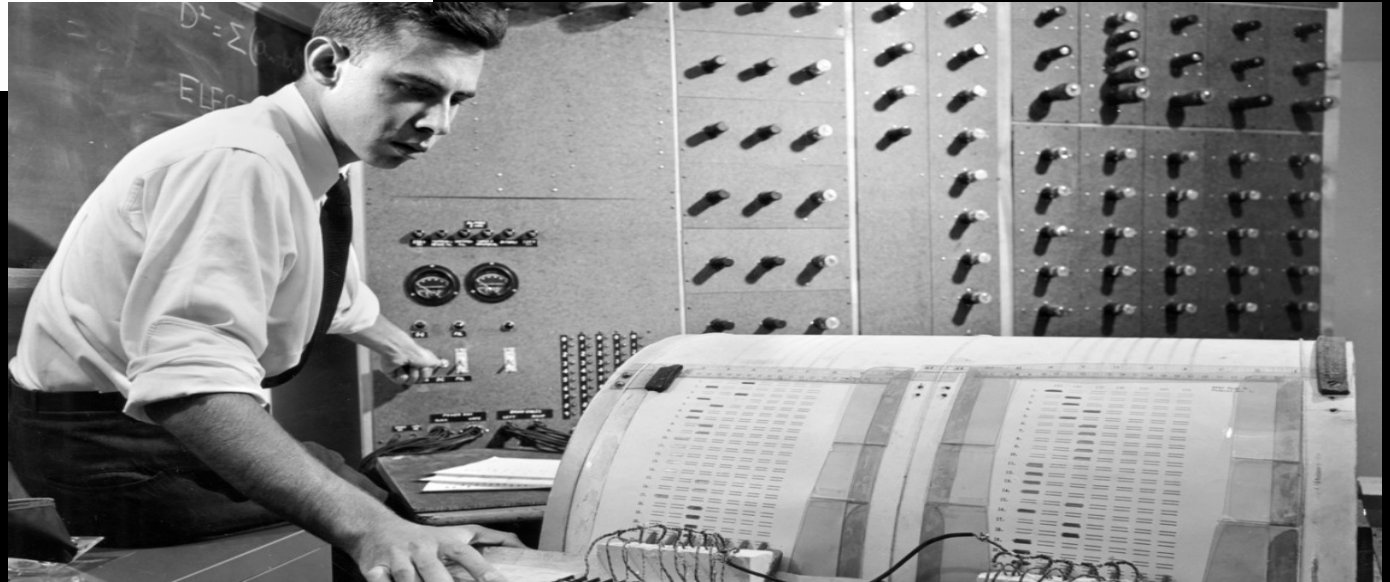
- Väčšina moderných implementácií systémov AI, medzi ktoré patria **systémy strojového učenia a neurónové siete**, vychádza zo subsymbolického prístupu, ktorý sa snaží reálne problémy uchopiť a v určitej miere ich aj úspešne riešiť.

Počiatky neurónových sietí...

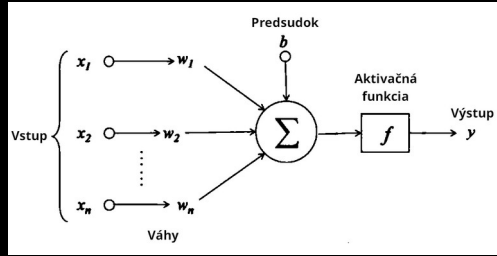
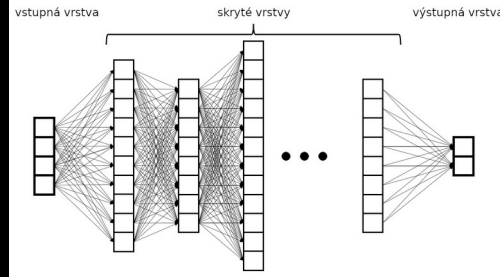
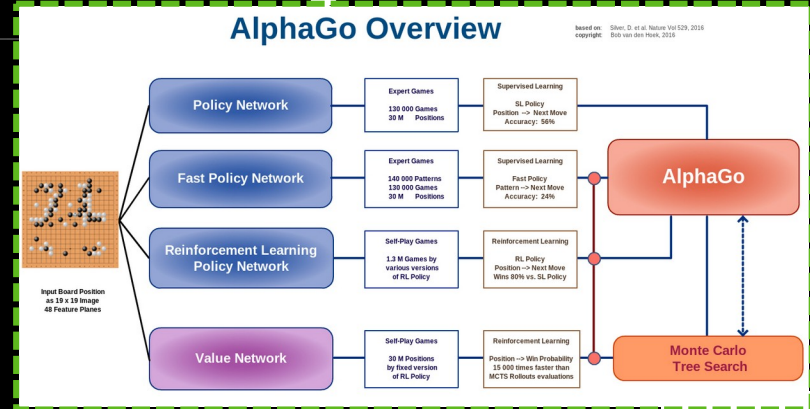
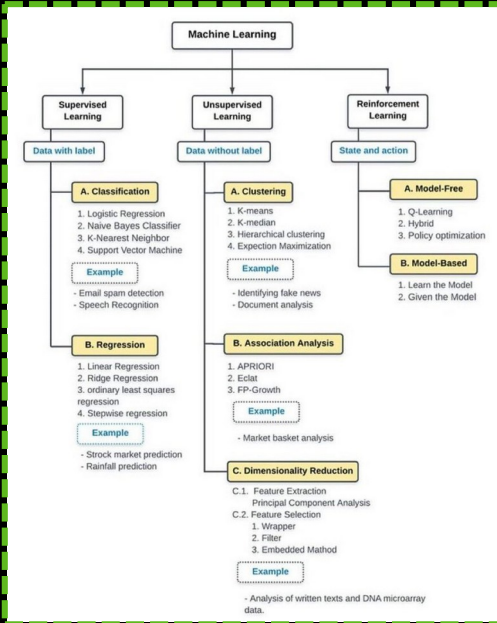
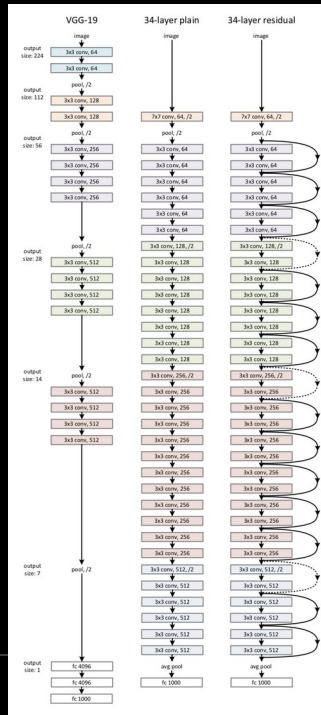
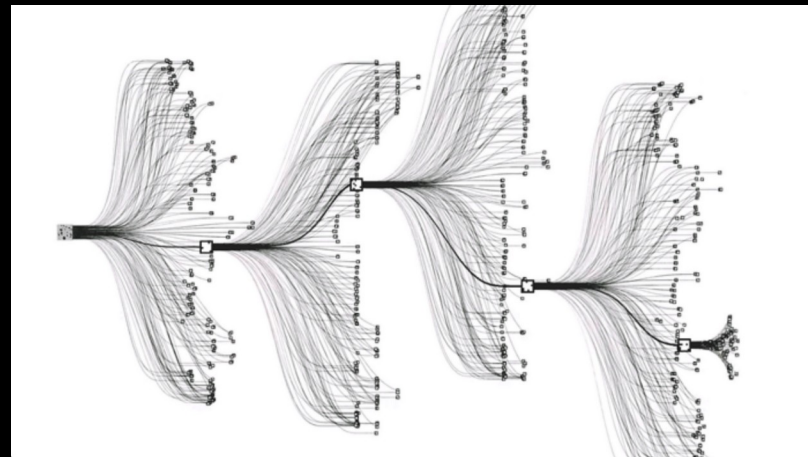
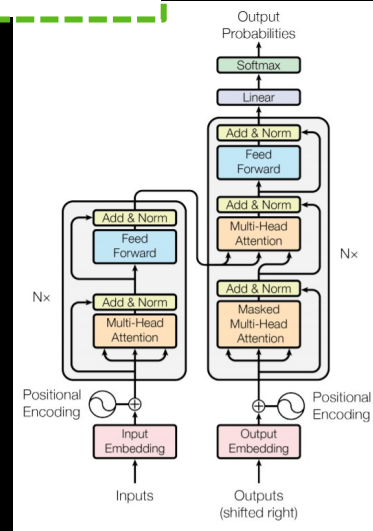


1943 – W. McCulloch a W. Pitts
Prvý matematický model NN ako ovocie
skúmania fungovania biologických neurónov.

1957 – Frank Rosenblatt
Perceptron ako počítačová
simulácia procesu spracovania
informácií v neurónoch.

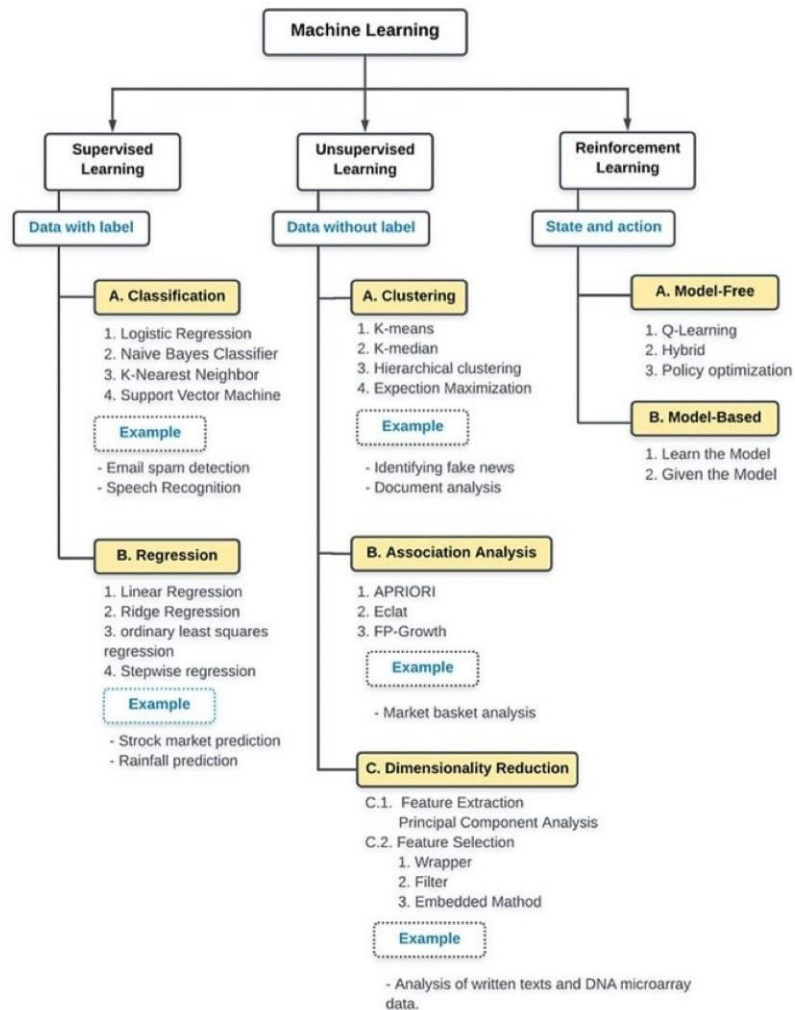
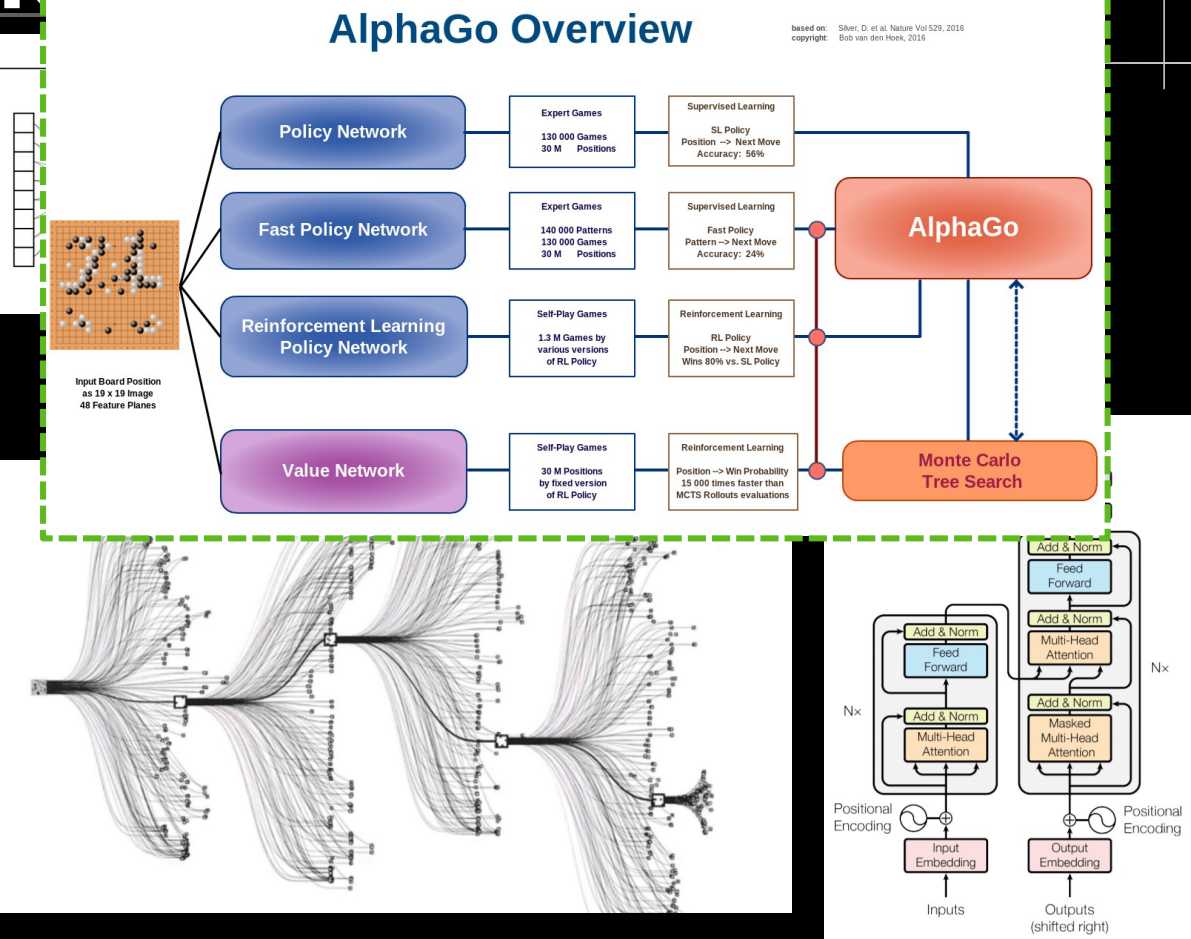


Rozvoj (nielen) neurónových sietí...



Kredit: autor, wikipedia, Fabio Galbusera, Eric Gaubert, Bob van den Hoek, Seth Weidma

neurónových sietí



Symbolická vs. subsymbolická AI?

- aktuálny výskum i reálne nasadenie **stierajú hranice** medzi symbolickým a subsymbolickým prístupom.
- klasické subsymbolické systémy na dosiahnutie skutočnej inteligencie **nestačia!** [fast perception vs. logika, umelé vs. biologické neuróny a siete]

„Na dosiahnutie skutočnej inteligencie nestačia súčasné systémy hlbokého učenia, ktoré sú ekvivalentami zmyslových centier v mozgu, napr. vizuálneho alebo sluchového. Skutočná inteligencia je však oveľa viac než to, keďže je treba skombinovať zmyslové vstupy do myslenia vyššej úrovne a symbolického uvažovania. Musíme preto tieto systémy budovať až do symbolickej úrovne uvažovania a myslenia - t. j. do úrovne matematiky, jazyka a logiky.“

Biologické vs. umelé neuróny a siete

- **rámcová podobnosť** medzi biologickými a umelými neurónmi i celými neurónovými sieťami (tak sme to navrhli;-)

biologický (reálny) neurón

synaptická účinnosť

séria akčných potenciálov

postsynaptický potenciál

sumárny synaptický potenciál

prah excitácie

synaptická plasticita

umelý neurón (uzol)

váha spojenia (w_i)

výstupná aktivita (y)

$w_i * x_i$

$\Sigma (w_i * x_i)$

σ

Δw_i (zmena synaptickej váhy)

Rozdielov je viac, než by sme si želali¹

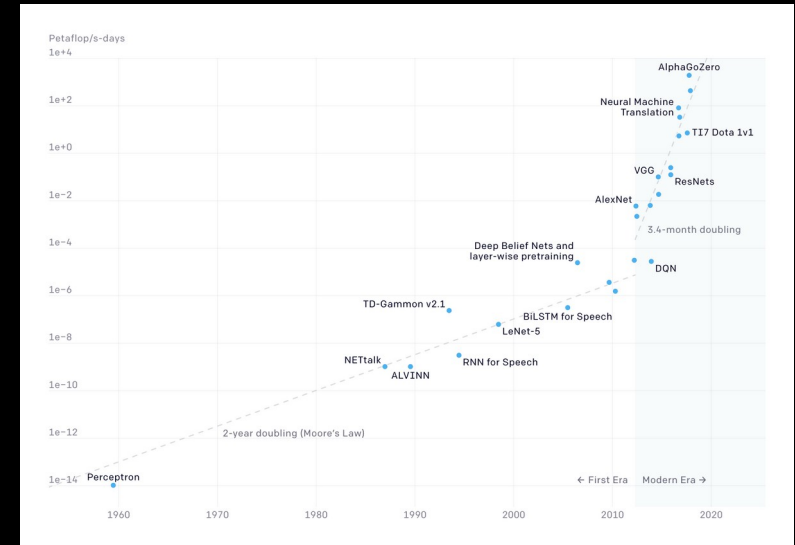
- binárny **impulz** $\longleftrightarrow$ membránový a akčný (spike) **potenciál** (pohyb iónov...)
- váhový **stav** „synapsy“ $\longleftrightarrow$ kontinuálna **aktivita** (synaptické potenciály,...)
 - zmeny synaptických váh $\longleftrightarrow$ synaptická plasticita
- **nelineárna vnútorná dynamika** biologických neurónov:
 - znižovanie membránového potenciálu (riedke impulzy = žiadny výstup)
 - oscilácie (vlny neuronálnej aktivity)
 - distribuované a paralelné aktivácie rôznych mozgových centier pri konkrétnej činnosti
- vážené **vstupy** $\longleftrightarrow$ zložité **mechanizmy** v dendritoch
 - lineárne/nelineárne operácie založené na iónových kanáloch (and/or,...)
- **synchrónna** NN $\longleftrightarrow$ **asynchrónna** komunikácia a spracovanie informácií

Rozdielov je viac, než by sme si želali²

- oveľa prepracovanejší systém **spätnej väzby** na úrovni neurónov
 - algoritmus spätného šírenia chyby (backpropagation) $\longleftrightarrow$ biologické algoritmy
- do konkrétnej činnosti sa dokážu zapájať **neuróny rôznych mozgových centier**
 - jednotlivé neuróny dokážu plniť rôzne úlohy pre rôzne činnosti
 - ukladanie informácií sa môže vykonávať naprieč rôznymi regiónmi v mozgu
- rozdielny **spôsob ukladania a vybavovania si spomienok**
 - pamäť uprednostňuje užitočné informácie, ktoré nám dovoľujú fungovať vo svete
 - genAI: komprimovaný obraz sveta a jeho informácií (strácanie znalostí v dotrénovaní..)
- **d'alšie typy prepojení** medzi biologickými neurónmi v mozgu
 - synaptické, medzibunkové (gap) spojenia, difúzne prenášače
 - **dendritické nanotrúbice**: vytváranie dynamických a dočasných komunikačných sietí
- činnosť neurónov a prenosu signálov ovplyvňovaná **na úrovni génov**

Energetická náročnosť

- súčasné sofistikované systémy AI majú **extrémne a neustále rastúce** výpočtové nároky a tým aj vysokú spotrebu energie
 - exponenciálny nárast výpočtového výkonu algoritmov strojového učenia
 - nové formy energetickej záťaže (genAI: vyťažovanie internetových zdrojov)
 - rastúca environmentálna záťaž
- prirodzená inteligencia **s nepatrnou spotrebou energie (cca. 20 wattov!)**
 - biologické systémy schopné kreativity, riešenia problémov a multitaskingu, ktoré si prirodzenou evolúciou osvojili spracovanie informácií



Hľadanie riešení...

- 3. generácia NN, predovšetkým **spiking neural networks** (SNNs)
- implementácia SNNs v rámci **neuromorfnej HW a SW architektúry**
 - neuromorfné siete/systémy sa snažia o pokročilejšie napodobnenie biologických systémov (štruktúry a fungovania ľudského mozgu, resp. mozgu cicavcov)
- **mechanistické modely** biologických neurónov ako inšpirácia pre vývoj NN
 - simulovanie anatomických, fyziologických, biofyzikálnych a biochemických mechanizmov biologickej nervovej bunky pomocou jednoduchého RC obvodu
- hľadanie HW a SW alternatív s využitím **biologického materiálu**
- vývoj a implementácia **nových algoritmov** v rámci strojového učenia
- rozvoj **lokálnych modelov** a HW implementácia parciálnych činností NN

Ako **súčasný** systémy AI „rozmýšľajú“

- **„rozmýšľajú“ úplne inak než ľudia**
Zlyhania sú odlišné od ľudských, ťažko predvídateľné, niekedy ľahko vykonateľné a mnohokrát prekvapivo robustné...
- **v skutočnosti nerozmýšľajú – inteligenciu len napodobňujú**
Systémy AI skutočnú inteligenciu nemajú, len ju simulujú.
- **s vážnymi etickými dôsledkami**
Nie sú schopné rozlišovať morálne dobré a zlé!
Nechápu zmysel a dôsledky!



Funkcionalistická umelá vs. komplexná ľudská inteligencia!

QR kód odkazuje na rozhovor, v ktorom *inteligenciu* AI rozoberám...

Striedanie ročných období

- **cyklické striedanie období** aktívneho rozvoja a útlmu v oblasti AI
 - **jar umelej inteligencie (AI spring)** – nové idey a pokrok vo výskume → predikcie prevratného pokroku vo vývoji systémov AI → mediálny boom a prílev štátnych dotácií i rizikového kapitálu do akademického bádania i komerčných startupov
 - **zima umelej inteligencie (AI winter)** – prevratný pokrok neprichádza, výsledky síce sú, no vôbec nespĺňajú extrémne očakávania → prílev financií a kapitálu ustáva, startupy krachujú a akademický výskum sa spomaľuje a obmedzuje
- **je striedanie ročných období aktuálne i v súčasnosti?**
 - **ANI** – široko plošná akceptácia a prelom vo využívaní AI, striedanie období asi končí
 - **AGI** – definitívna méta; nutnosť zvládnuť koncepčné prelomy + schopnosti súčasných top systémov AI → predpokladám určité pokračovanie striedania rozvoja a útlmu

Jednoduché veci sú ťažké

- **Moravcov paradox** (Moravec, Brooks a Minsky)
„Jednoduché veci sú ťažké“, keďže riešenie úloh, ktoré sa človeku zdajú byť jednoduché, je pre umelú inteligenciu veľmi náročné. A naopak to, čo je pre človeka extrémne náročné, dokážu systémy AI zvládať veľmi dobre.
- **problém podstaty ANI**
 - chýba schopnosť „**zdravého rozumu**“ (common sense is the dark matter of AI)
 - systémy, ktoré sa javia ako inteligentné → systémy, ktoré sú inteligentné

Steven Pinker: „The hard things are easy, but the easy things are hard.“

Marvin Minsky: „Vo všeobecnosti si najmenej uvedomujeme to, čo naša myseľ robí najlepšie.“

Gary Marcus: „Common sense is the precondition for general intelligence; until we get there we will be stuck with narrow AI that is rarely robust and never as flexible as human reasoners.“

Generatívne systémy umelej inteligencie (genAI)

- **stochastické systémy, ktoré sú schopné generovať rôznorodé výstupy**
 - **generatívne predtrénovanie**, ktoré učí model rozpoznávať vzory v neoznačených údajoch a následne tieto vzory aplikovať na nové vstupy
 - **architektúra transformerov**, ktorá umožňuje modelu paralelne spracovať všetky časti vstupnej sekvencie a generovať výstupy
- **„vnímanie“ syntaktickej a sémantickej analýzy** na základe pravdepodobnostného učenia prostredníctvom naučených vzorov [vnímanie – nepozná pravidlá a nechápe význam]
 - na základe štatistickej analýzy obrovského množstva textov sa modely učia **syntaktické vzory a vzťahy**, ktoré sú zachytené v neurónových váhach
 - emergentným dôsledkom sú tzv. **sémantické reprezentácie**, ktoré nie sú formálnou sémantickou analýzou tradičnej lingvistiky, ale naučením sa významov a vzťahov medzi slovami z kontextu

Emergentná a disruptívna technológia, ktorá má potenciál posunúť nás ďalej.

Ak ju nezvládneme, má potenciál nás priviesť ku skutočne veľkým problémom.

GenAI – ľahký pohľad pod kapotu...

- **prevádzka (beh) modelu – nič zložité** (v zásade dva súbory)
 - **súbor parametrov modelu** – natrénovaný model, resp. jeho parametre
 - **samotný počítačový program** – cca. 500 riadkov kódu (napr. v jazyku C), ktorý implementuje architektúru neurónovej siete a využíva súbor parametrov pre beh modelu
- **trénovanie a príprava modelu**
 - zhromažďovanie a príprava veľkého množstva dát (**data pre-processing**)
 - predtrénovanie, resp. vytvorenie základného modelu (**base model, foundation model**)
 - jemné ladenie (fine tuning), ktorého výsledkom je tzv. **assistant model**
 - finálne ladenie na základe porovnávania odpovedí (**RLHF / RLAF**)
- **technológie tvoriace generatívne systémy AI**
 - tokenization, embedding, encoders, transformers, decoders,...

Vylepšovanie systémov genAI a budúci vývoj¹

- škálovania generatívnych systémov AI (počet parametrov, veľkosť dát...)
- používanie externých nástrojov na vykonanie úlohy
- pokrok v rámci multimodálnych modelov LMM/VLM (bez potreby konverzie)
- komplexné rozhodovanie generatívnych systémov
 - system1 (fast), system2 (reasoning), system3 (metakognícia, epizodická pamäť,...)
 - hybridné modely uvažovania (hybrid reasoning models)
- prispôsobené (custom) generatívne modely (RAG,...)
- agenti generatívnych systémov umelej inteligencie (agentic AI)
 - aktívne plánovanie, uvažovanie, využívanie nástrojov na dosiahnutie zadaných cieľov
 - AI orchestrátor – generatívne systémy AI ako „operačné systémy“

Vylepšovanie systémov genAI a budúci vývoj²

- zväčšovanie **kontextového** okna
- **destilácia znalostí** (knowledge distillation) a vytváranie malých lokálnych modelov
- **sebazdokonaľovanie** podľa vzoru AlphaGo
- **pokroky v neurovede** ako základ pre nové modely
- **testovanie vlastností a riešenie dôveryhodnosti** generatívnych systémov AI
- „**stelesnenie**“ (embodiment) systémov AI (interakcia s reálnym svetom – senzory, robotika,...)

GenAI – ešte niečo na rozmýšľanie...

- modely genAI ako **zazipované obrazy sveta**
 - parametre veľkých modelov, ktoré dosahujú miliardy natrénovaných hodnôt, by sme v prenesenom zmysle slova mohli považovať za stratovú kompresiu sveta
- **dáta sú fosílnym palivom AI**
- **ľudská adaptácia, ktorá odomyká potenciál systémov genAI**

Súčasnosc' – umelá inteligencia „na koni“¹

Na čom stojí súčasný rozmach umelej inteligencie?

- **pokrok v AI**: vo vývoji metód a systémov umelej inteligencie
- **extrémne množstvo dát** (štruktúrovaných i neštruktúrovaných)
- rozvoj automatizácie smerujúci k **potrebe autonómnych systémov** schopných samostatnej adaptívnej činnosti s modernými prístupmi schopnými spracovávať big data
- využívanie nástrojov genAI smerujúce k prirodzenému **ľudskému spôsobu komunikácie** a k aplikácii **postupov meniacich obsah i spôsob našej práce**
- možnosť **integrovat' nástroje AI** do širokého spektra pracovných činností
- jednoduchá a lacná **dostupnosť nástrojov AI pre bežnú populáciu**
- **spoločenská akceptácia** fenoménu umelej inteligencie

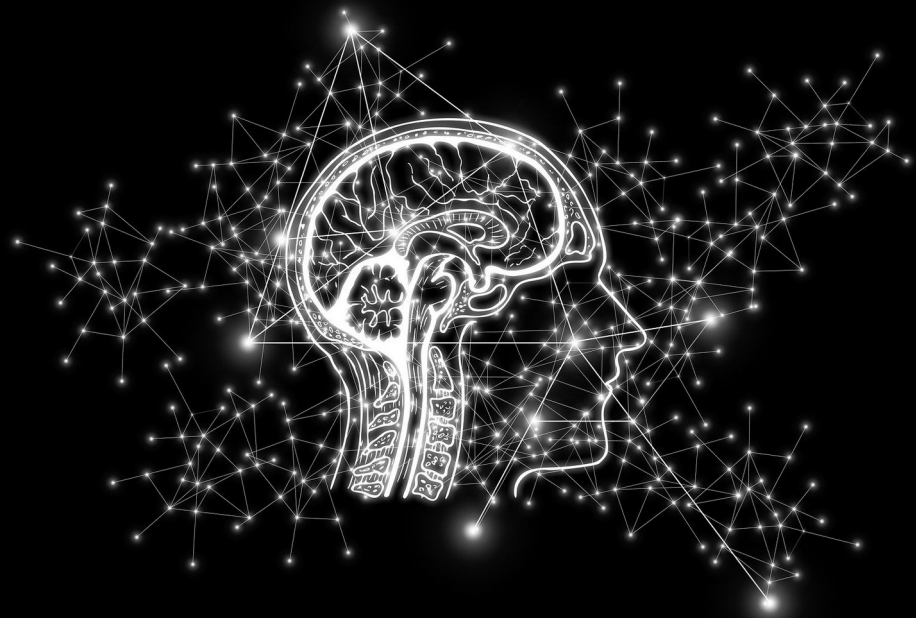
Súčasnost' – umelá inteligencia „na koni“²

AI ako bytostná súčasť paradigmatických zmien

- technológie a dáta od prostriedku ku kontextu: AI ako motor a konečný činiteľ
- AI ako základ novej paradigmatickej zmeny: technológia ako aktér, resp. aktívny činiteľ vo svete

Technológie a dáta sa z roviny prostriedku dostávajú do roviny kontextu, resp. z roviny kontextu do úlohy aktéra až do tej miery, že sa mení medziľudská komunikácia v esenciálnej rovine, menia sa vzťahy, mení sa spoločnosť a princípy jej fungovania.

Na môj pomerne rozsiahly popis systémov AI odkazuje zobrazený QR kód...



Ďakujem za pozornosť



ThLic. Ing. Peter Šantavý, PhD., UK v Bratislave
peter.santavy@uniba.sk