



Umělá inteligence pro každého

Jiří Materna



O mně

- Ph.D. ze zpracování přirozeného jazyka a umělé inteligence na Masarykově univerzitě
- 10 let v Seznam.cz (posledních 8 let vedoucí výzkumného oddělení)
- Zakladatel a lektor v Machine Learning College
- Zakladatel a spoluorganizátor Machine Learning Prague
- ML freelancer a konzultант

Osnova

- co je a není umělá inteligence
- slabá a silná umělá inteligence
- úlohy umělé inteligence
- historie umělé inteligence
- úvod do strojového učení
- aplikace strojového učení
- filosofie umělé inteligence
- etika umělé inteligence

Co je umělá inteligence?

umělá - nevyskytující se přirozeně, vytvořená člověkem nebo strojem

inteligence - schopnost uvažování a řešení problémů; schopnost učit se; schopnost přizpůsobit se novému prostředí

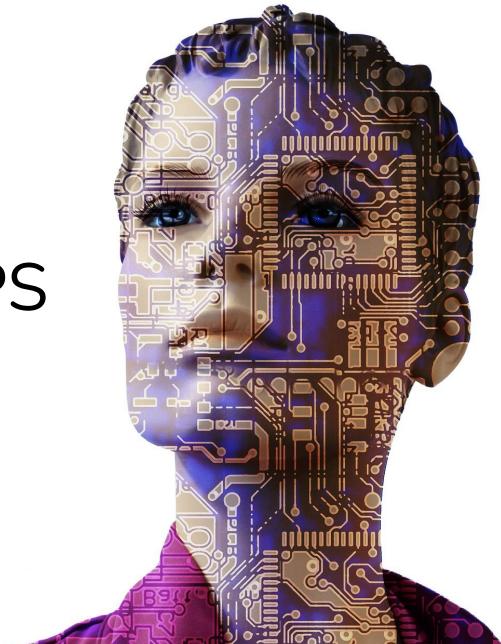
příklady:

- algoritmus protivníka ve hře
- vyhledávač Google
- robotický vysavač
- řazení příspěvků na Facebooku



Co není umělá inteligence?

- tabulka v databázi
- internet
- bitcoin/blockchain
- určení zeměpisné šířky a délky pomocí GPS
- výpočet největšího společného dělitele



Slabá a silná umělá inteligence

Umělá inteligence

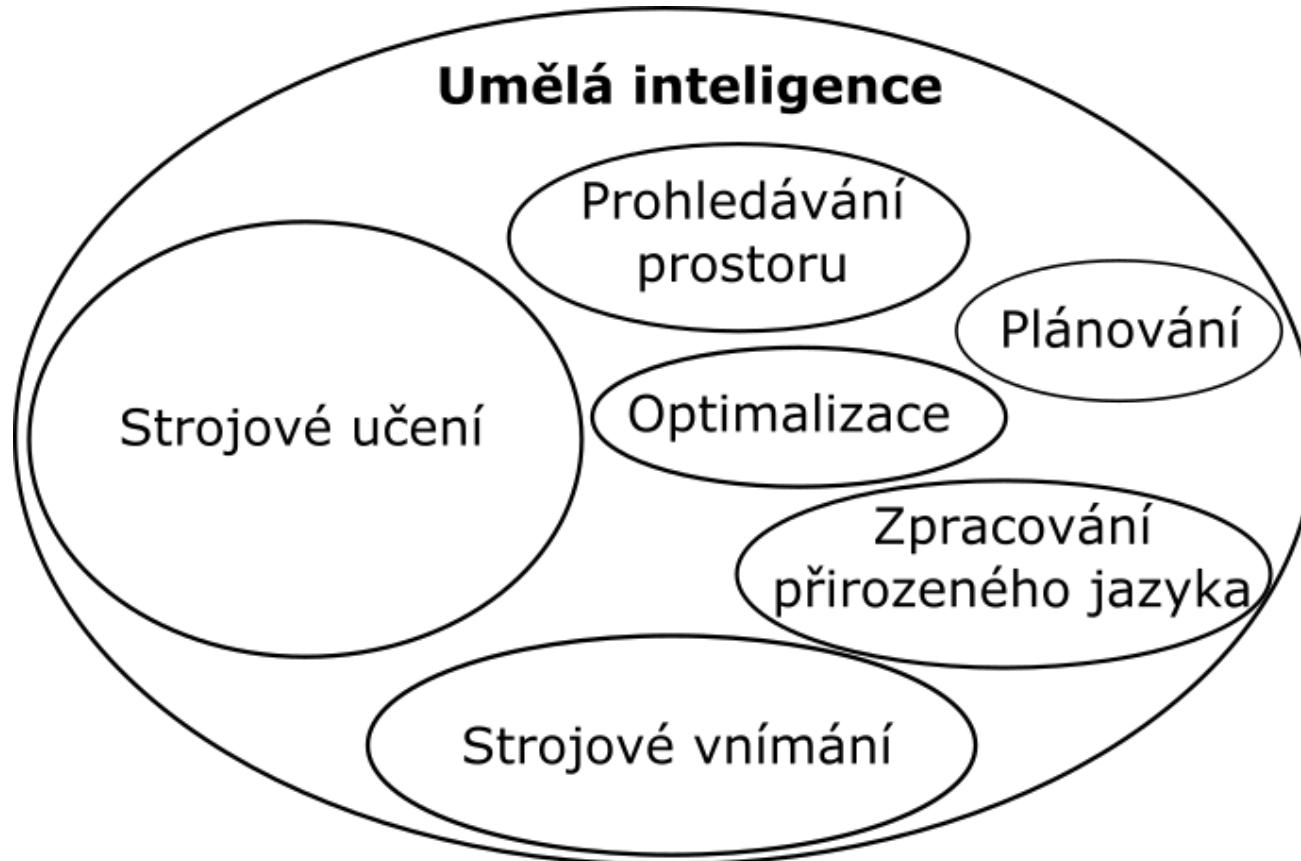
Slabá UI

- vždy řeší jeden konkrétní problém, neumí se adaptovat na nové problémy
- všechna existující UI jsou slabou umělou inteligencí

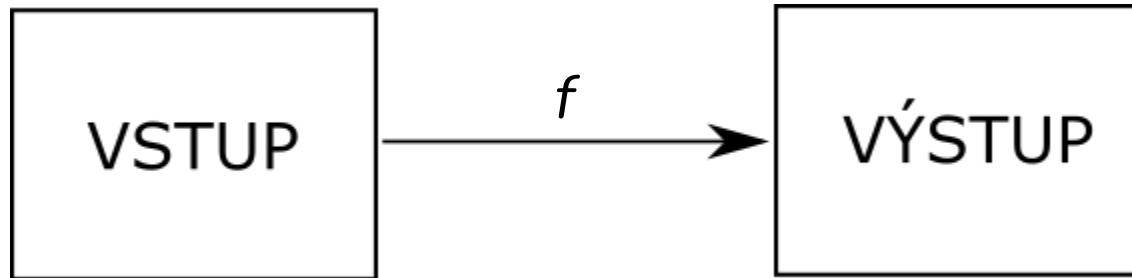
Silná (obecná) UI

- umí cokoliv, co dokáže člověk, nebo dokonce více
- zatím neexistuje

Úlohy umělé inteligence



Strojové učení



Úlohou strojového učení je na základě příkladů vstupů a výstupů nalézt funkci f , která pro nový vstup určí odpovídající výstup.

Příklady dvojic vstupů a výstupů nazýváme *trénovací data*.

V současnosti nejrozšířenější metoda umělé inteligence s největšími dopady.

Strojové učení - příklady



f → pes

klasifikace obrázků

hello f → ahoj

strojový překlad AJ -> ČJ

90 km/h f → 5,1 l

Predikce spotřeby auta pod
průměrné rychlosti

Prohledávání prostoru - hraní her



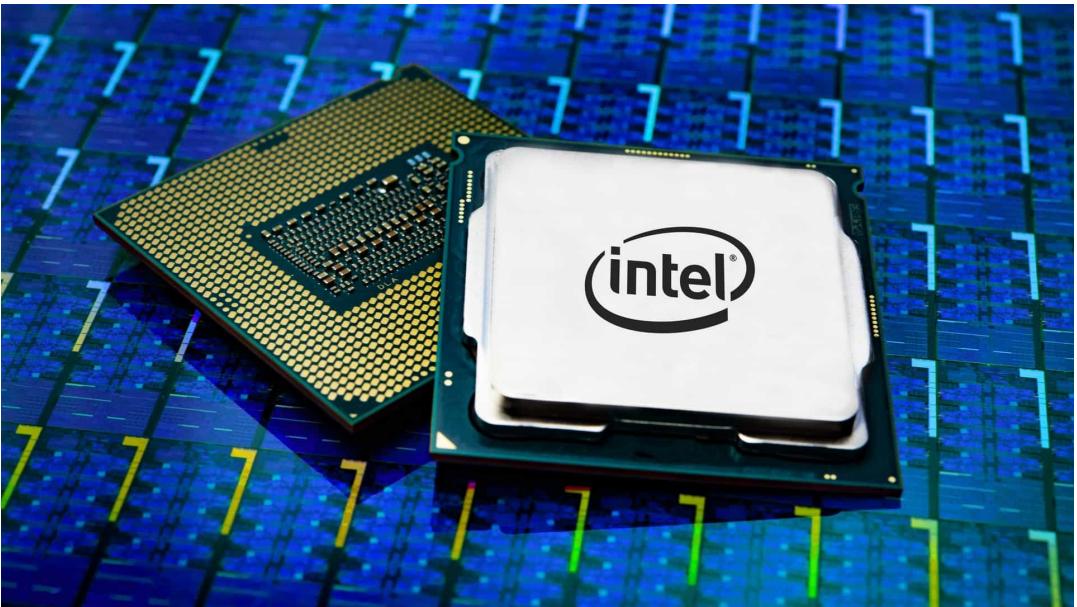
Cílem je prohledat všechny možnosti svých a soupeřových tahů a vybrat takový tah, který bude mít největší šanci na vítězství.

Prohledávání prostoru - nalezení nejkratší cesty



Cílem je najít nejkratší cestu mezi dvěma místy. Kromě vzdálenosti může být zohledněna i délka cesty nebo nějaká omezení.

Plánování



Příkladem automatického plánování je rozdělení výpočetních úloh mezi jádra procesoru tak, aby výpočet všech skončil co nejdříve. To řeší operační systém automaticky.

Jiným příkladem plánování je sestavování školního rozvrhu.

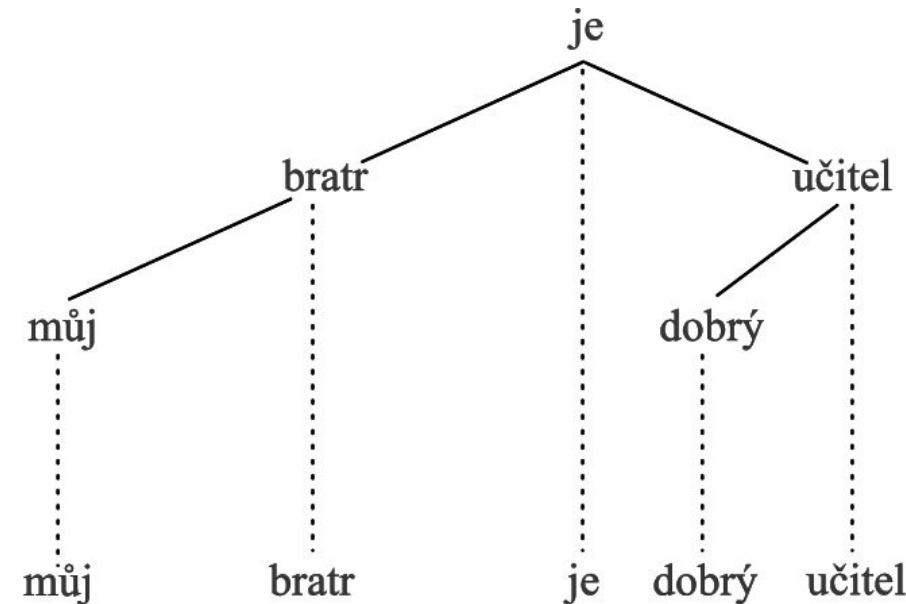
Optimalizace



Cílem je najít kombinaci vstupů, která bude minimalizovat nebo maximalizovat výstupní hodnotu nějaké funkce.

Příkladem může být problém rozmístění dílů uvnitř mobilního telefonu tak, aby jako celek zabíraly co nejmenší objem.

Zpracování přirozeného jazyka



Zpracováním přirozených jazyků (např. češtiny nebo angličtiny) rozumíme především jejich částečné strojové porozumění v různých aplikacích.

Příkladem aplikace může být automatické sestavení větného rozboru pro kontrolu pravopisu, strojový překlad jazyků, nebo třeba vyhledávač webových stránek na internetu.

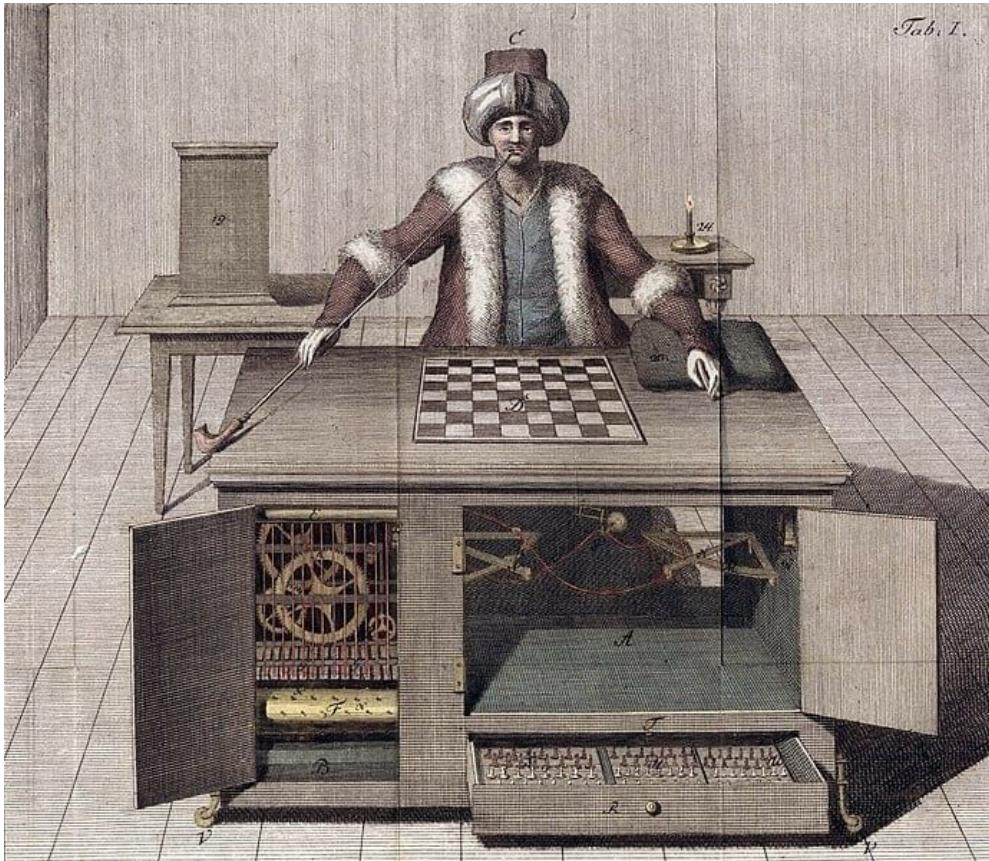
Strojové vnímání



Strojovým vnímáním rozumíme především zpracováním vizuálních a akustických vstupů.

Příkladem aplikací může být rozpoznání objektu na obrázku, identifikace reklamy ve videu nebo převod mluvené řeči na text.

Historie umělé inteligence



Historie umělé inteligence: 6. stol. př. n. l.



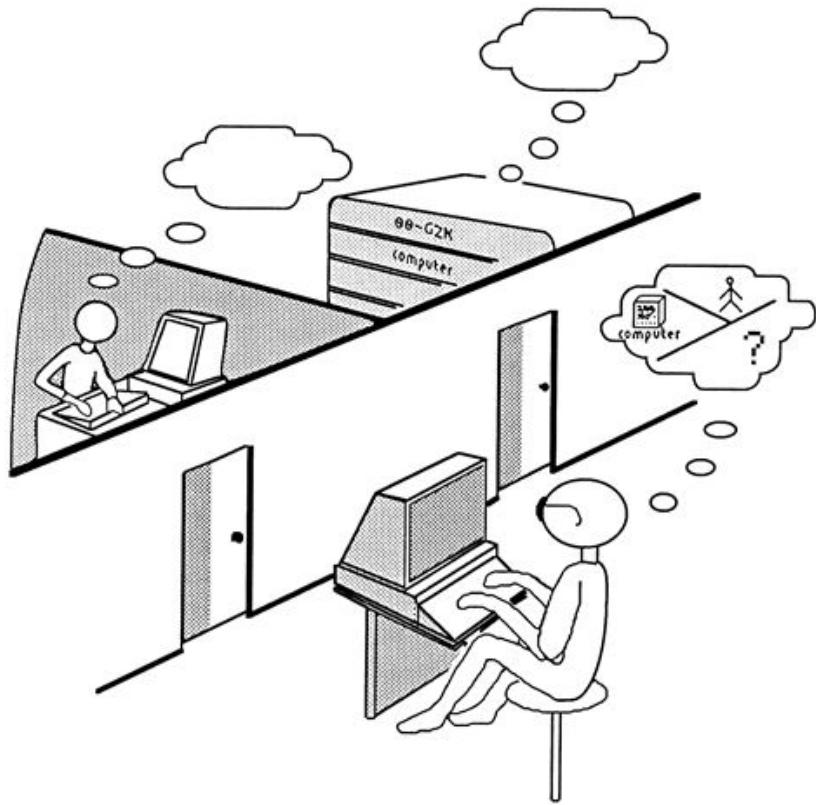
Legenda o Golemovi, hliněné soše, která ožije a je schopná vykonávat lidskou práci, pochází již ze starověkého Egypta.

Historie umělé inteligence: 1921



Karel Čapek dokončuje divadelní hru R.U.R., kde je poprvé použito slovo *robot*.

Historie umělé inteligence: 1950



Alan Turing navrhuje test strojové inteligence, který je později pojmenován *Turingův test*.

Historie umělé inteligence: 1965

Welcome to

EEEEEE	LL	IIII	ZZZZZ	AAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LL	II	ZZZ	AAAAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LLLLL	IIII	ZZZZZ	AA AA

Eliza is a mock Rogerian psychotherapist.

The original program was described by Joseph Weizenbaum in 1966.

This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?

YOU: Men are all alike.

ELIZA: What is the connection, do you suppose ?

YOU: They're always bugging us about something or other.

ELIZA: Can you think of a specific example ?

YOU: Well, my boyfriend made me come here.

ELIZA: Is it important to you that your boyfriend made you come here ?

YOU: He says I'm depressed much of the time.

ELIZA: I am sorry to hear that you are depressed.

YOU: It's true. I am unhappy.

ELIZA: Can you explain what made you unhappy ?

YOU:

Německý informatik
Joseph
Weizenbaum
vytváří počítačový
program ELIZA,
který simuluje
psychoterapeuta.

Historie umělé inteligence: ~ 1970 - 1990



Období skepse, řady neúspěchů a ztráty víry v umělou inteligenci, známé jako "AI Winter".

Historie umělé inteligence: 1997



Superpočítač Deep Blue poráží tehdejšího světového šampiona v šachu Garryho Kasparova.

Historie umělé inteligence: 2002



Vzniká první sériově vyráběný robotický vysavač Roomba, který je schopen adaptace na neznámé prostředí.

Historie umělé inteligence: 2011



Superpočítač IBM Watson vyhrává populární televizní vědomostní soutěž Jeopardy! (podobná české soutěži Riskuj!).

Historie umělé inteligence: 2014



Vzniká hlasem ovládaný virtuální asistent Alexa, kterým mohou uživatelé ovládat chytrou domácnost, nakupovat, nebo přehrávat hudbu.

Historie umělé inteligence: ~ 2015



Po silnicích v USA začínají jezdit v běžném provozu elektromobily s autonomním řízením (na obrázku elektromobil společnosti Google).

Historie umělé inteligence: 2016



Počítačový program AlphaGo firmy DeepMind poráží světového šampiona ve hře Go. Ta je považována za jednu z nejtěžších deskových her vůbec.

Historie umělé inteligence: 2019



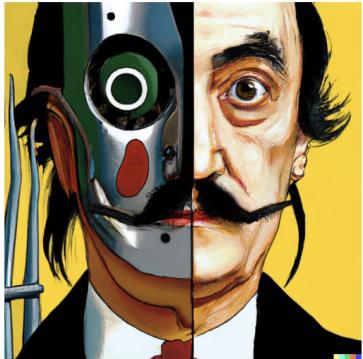
Počítačový program AlphaStar vítězí nad nejlepšími profesionálními hráči v počítačové hře Starcraft II.

Historie umělé inteligence: 2020



Začíná éra takzvaných “deep fakes”. Audiovizuálních a textových podvrhů, generovaných umělou inteligencí, které jsou nerozlišitelné od skutečnosti.

Historie umělé inteligence: 2022



vibrant portrait painting of Salvador Dalí with a robotic half face



a shiba inu wearing a beret and black turtleneck



a close up of a handpalm with leaves growing from it



an espresso machine that makes coffee from human souls, artstation



panda mad scientist mixing sparkling chemicals, artstation



a corgi's head depicted as an explosion of a nebula

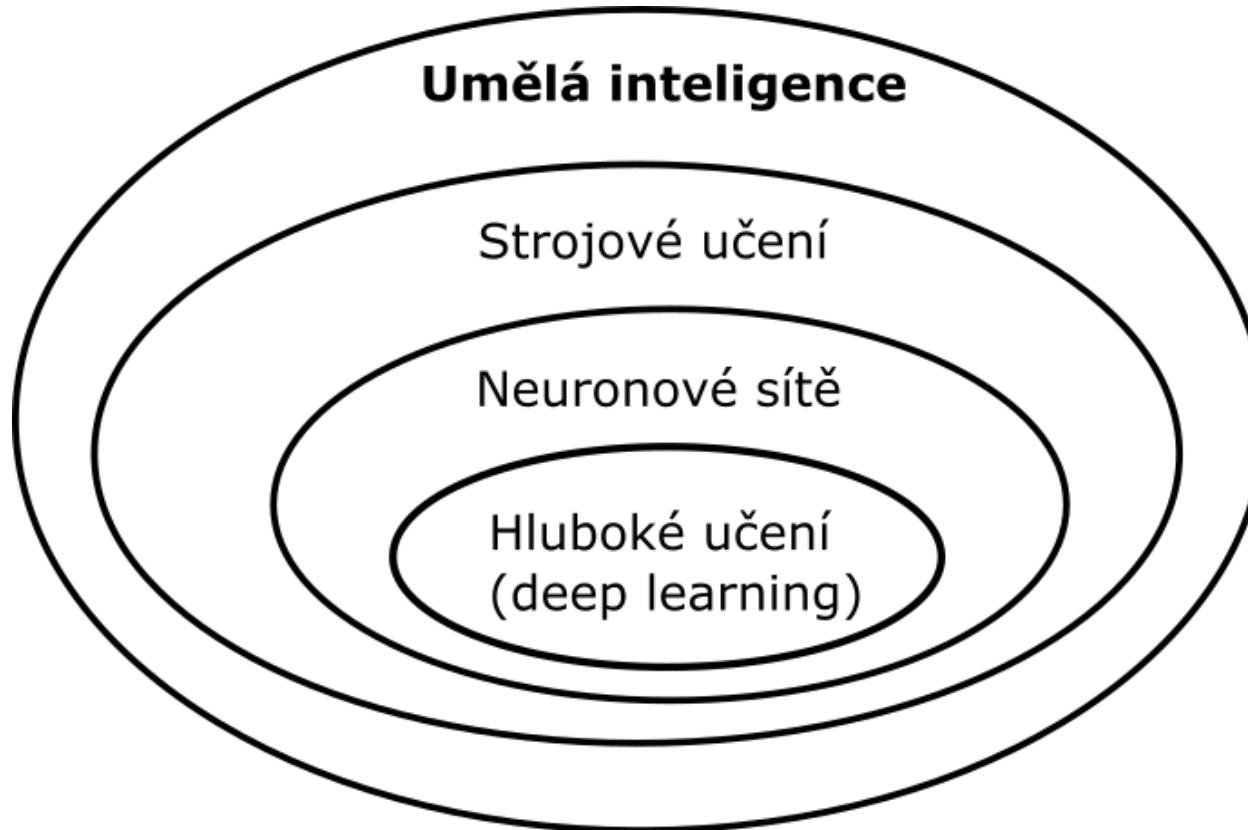
Ve společnosti OpenAI byl vytvořen model strojového učení DALL·E 2, který dokáže generovat fotorealistické obrázky podle textového zadání.

Historie umělé inteligence: 2022



Ve společnosti OpenAI byl vytvořen model upravený jazykový model ChatGPT, schopný vedení velmi věrohodného dialogu.

Strojové učení a umělá inteligence



Strojové učení je jedním z nástrojů umělé inteligence.

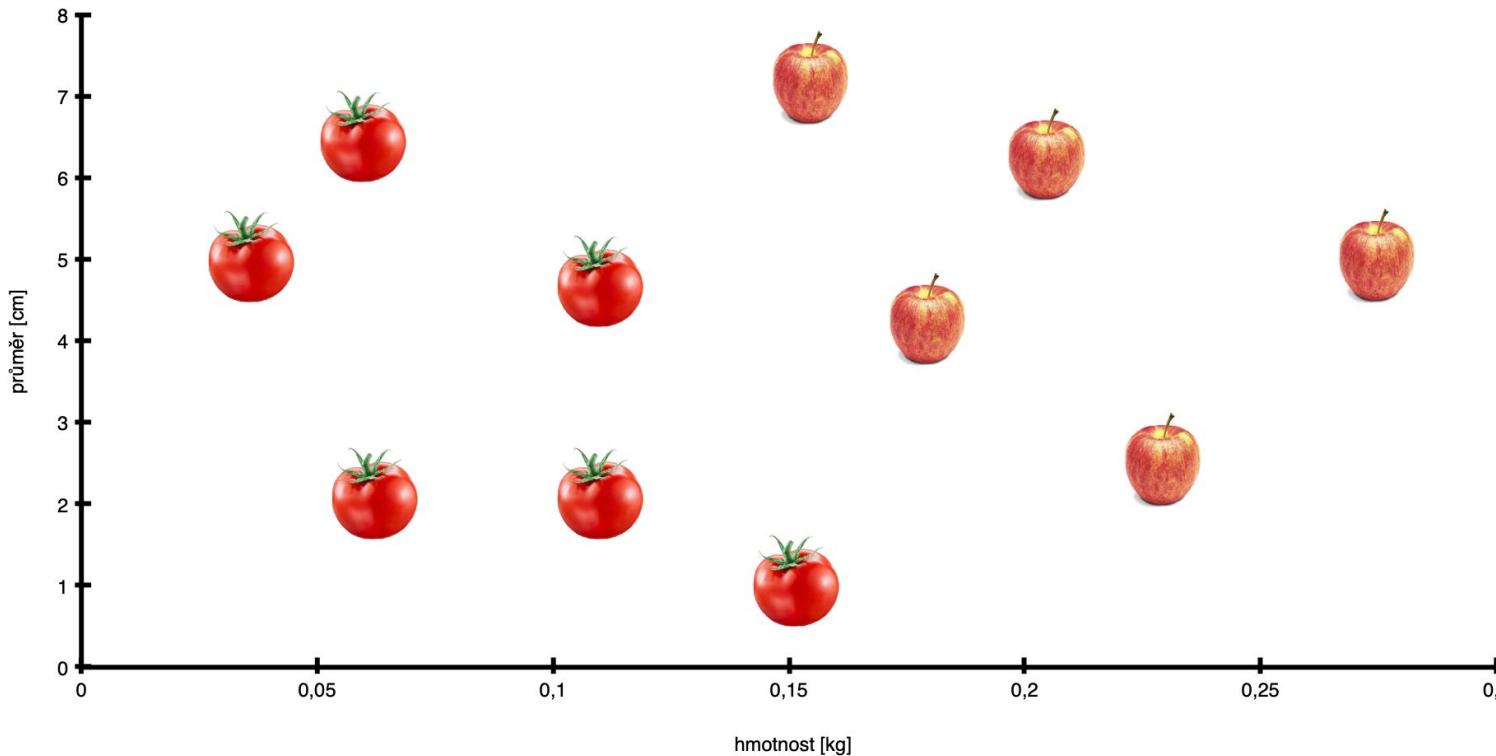
V současnosti je jednoznačně dominantní.

Co je učení

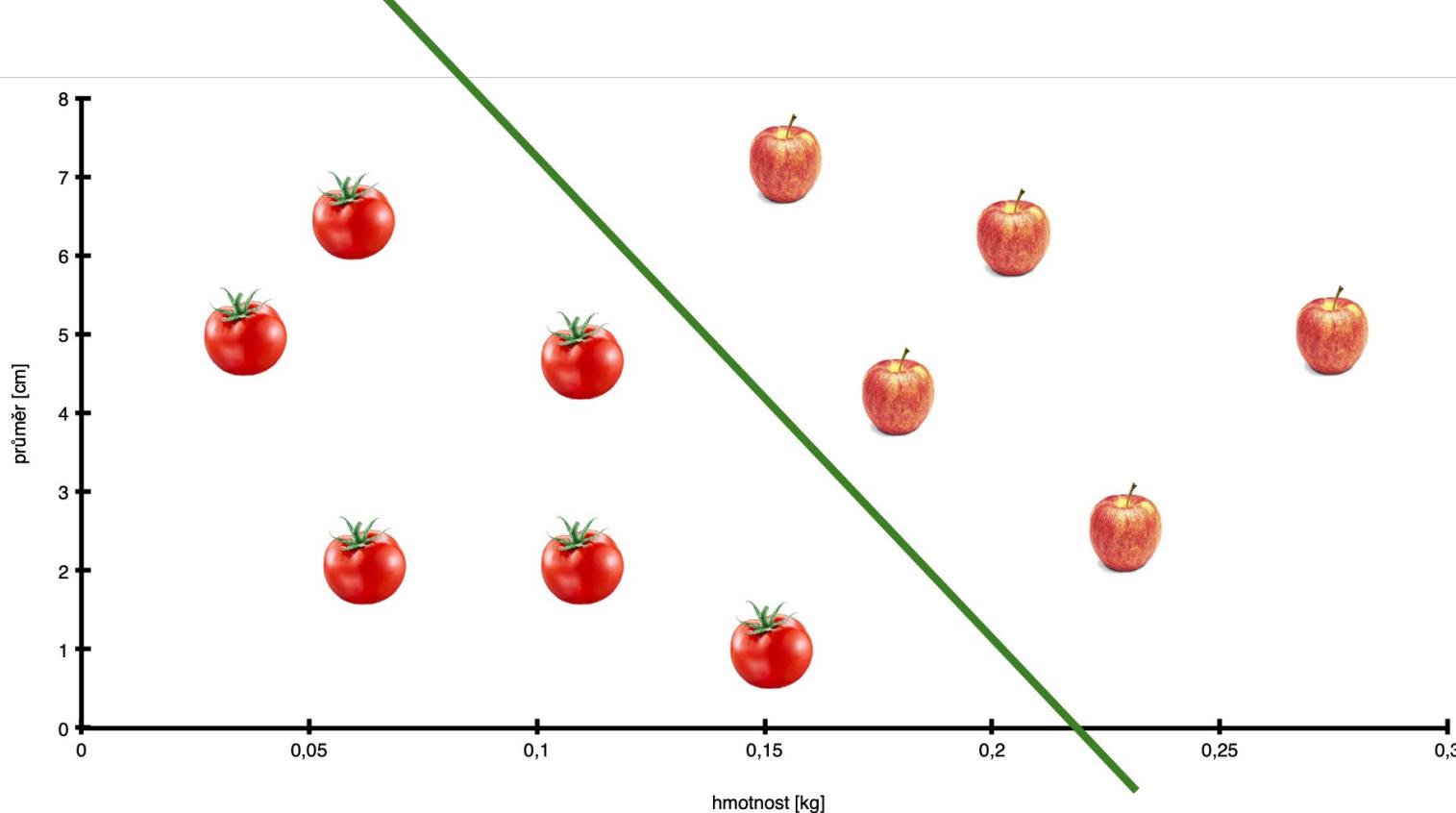
- Dítě se naučí rozlišovat jablka a hrušky tak, že mu rodiče ukáží příklady jablek a příklady hrušek.
- Dítě si podvědomě všimne různých charakteristik, které jablka a hrušky rozlišují (jablka jsou typicky kulatější, hrušky zelenější apod.)
- Poté je dítě schopno určit druh ovoce samo.



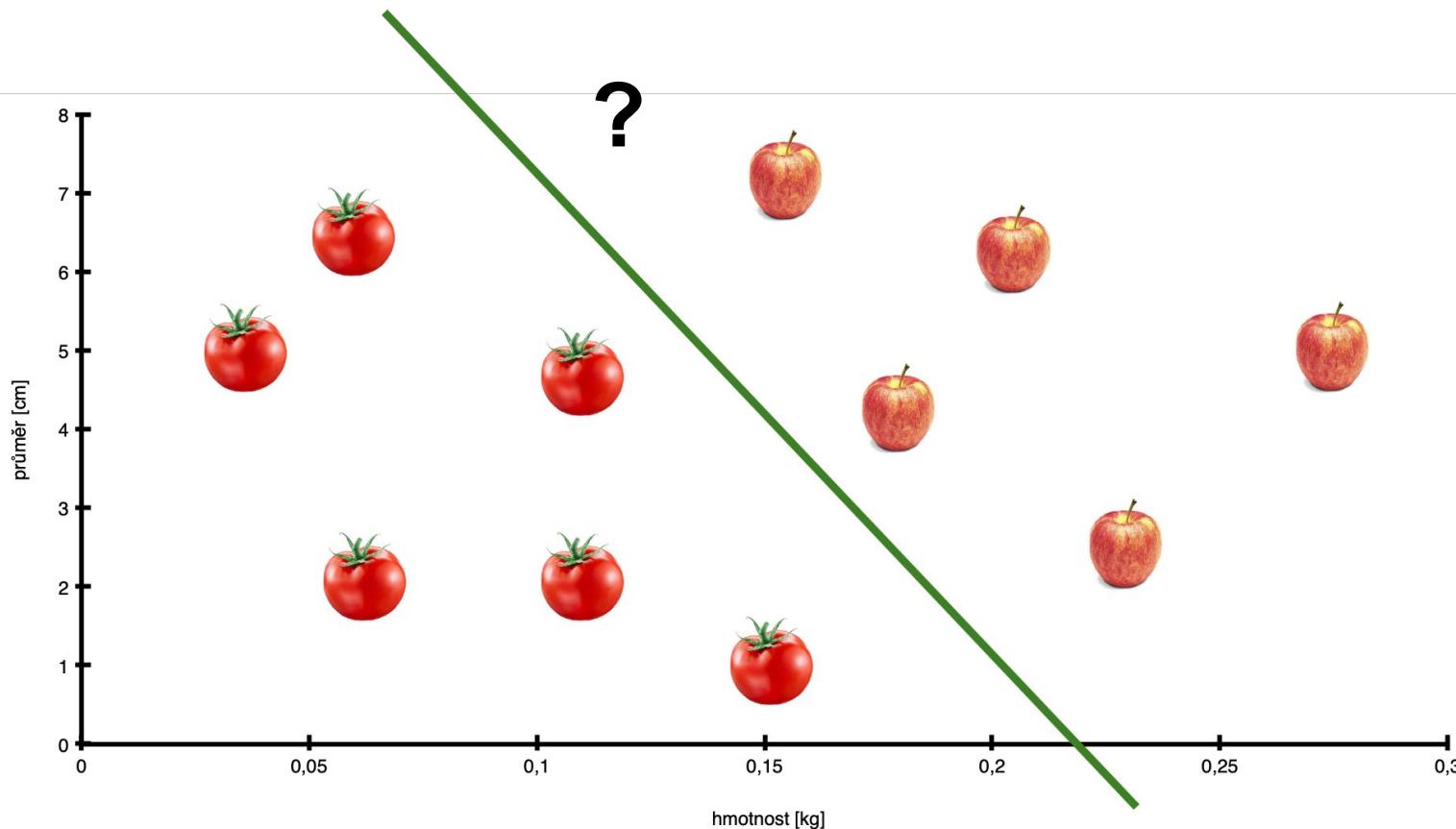
Strojové učení



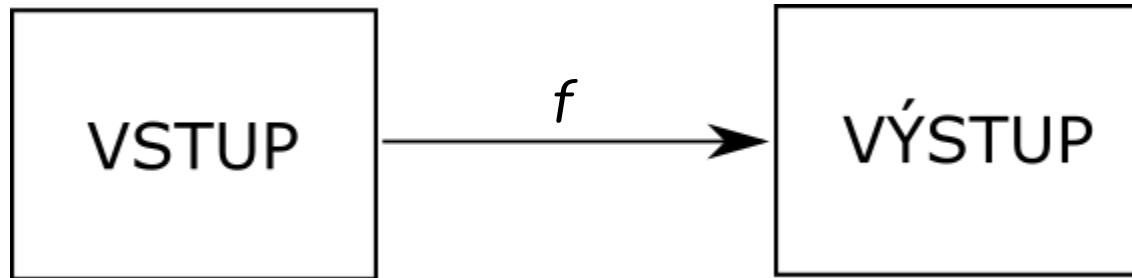
Strojové učení



Strojové učení



Strojové učení



Úlohou strojového učení je na základě příkladů vstupů a výstupů nalézt funkci f , která pro nový vstup určí odpovídající výstup.

Příklady dvojic vstupů a výstupů nazýváme trénovací data.

V současnosti je to nejrozšířenější metoda umělé inteligence s největšími dopady.

Strojové učení - příklady



f → pes

klasifikace obrázků

hello f → ahoj

strojový překlad AJ -> ČJ

90 km/h f → 5,1 l

Predikce spotřeby auta pod
průměrné rychlosti

Generalizace versus memorování

Pokud se člověk nebo stroj na základě konečného počtu trénovacích příkladů naučí rozeznat *libovolnou* hrušku nebo jablko, jedná se o **generalizaci (zobecnění)**.

Pokud si pouze zapamatuje trénovací příklady, ale není schopný znalost zobecnit pro nové příklady, jedná se o **memorování**.

Z pohledu strojového učení je mnohem užitečnější generalizace. K memorování nám postačí uložení příkladů do paměti počítače.

Proces strojového učení

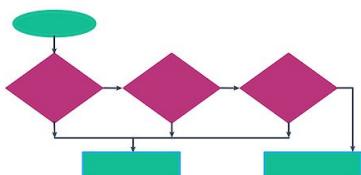
Trénovací data



Model



Algoritmus



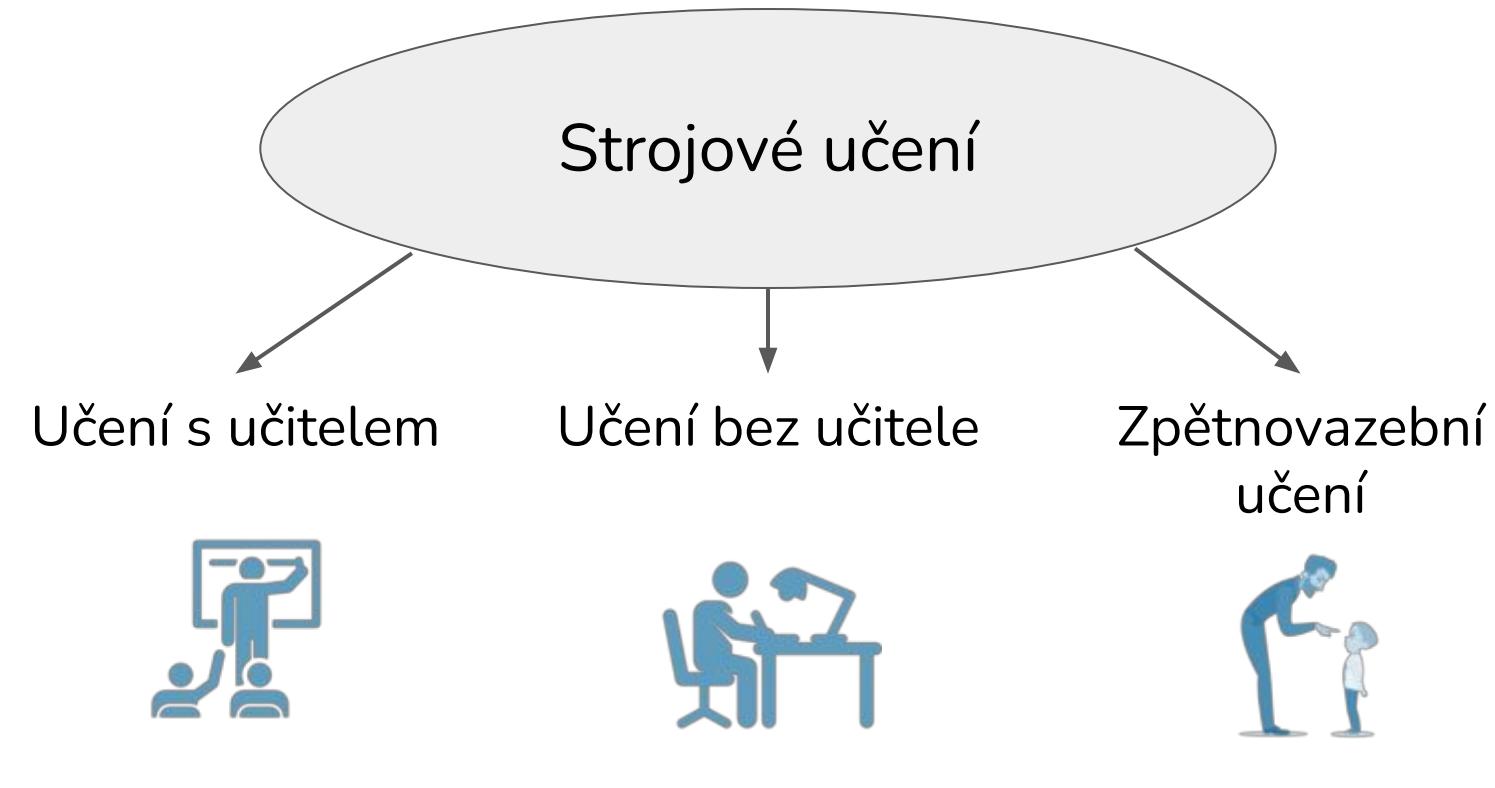
Testovací data



Predikce



Základní typy strojového učení



Učení s učitelem (supervised learning)

Součástí trénovacích dat učení s učitelem je požadovaný výstup (predikce).

Klasifikace - výstupem je nějaká kategorie (třída). Například barva, binární hodnota (*ano, ne*), den v týdnu, typ auta apod.

Regresi -výstupem je číselná hodnota. Například cena, teplota, počet lidí v místnosti apod.

Klasifikace - příklad

Rozlišení jablek a hrušek

vstup

tvar	barva	hmotnost (g)
kulatý	červená	146
šíšatý	žlutá	120
šíšatý	zelená	187
kulatý	červená	155

výstup



druh ovoce
jablko
hruška
hruška
jablko

Klasifikace - příklady aplikací

Klasifikace obrázků



Detekce spamu



Autentizace



Strojový překlad

Diagnostika onemocnění COVID-19
podle zvukového záznamu kašle

ČJ -> AJ



Regresce - příklad

Predikce ceny bytu

vstup

plocha	patro	počet místností
42	7	2
105	3	3
67	1	2
224	3	4

výstup

cena (mil. Kč)
3,2
6,8
4,1
13,9

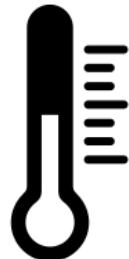


Regresy - příklady aplikací

predikce ceny Bitcoinu



předpověď teploty



určení vzdálenosti
samoříditeľného auta
od krajnice



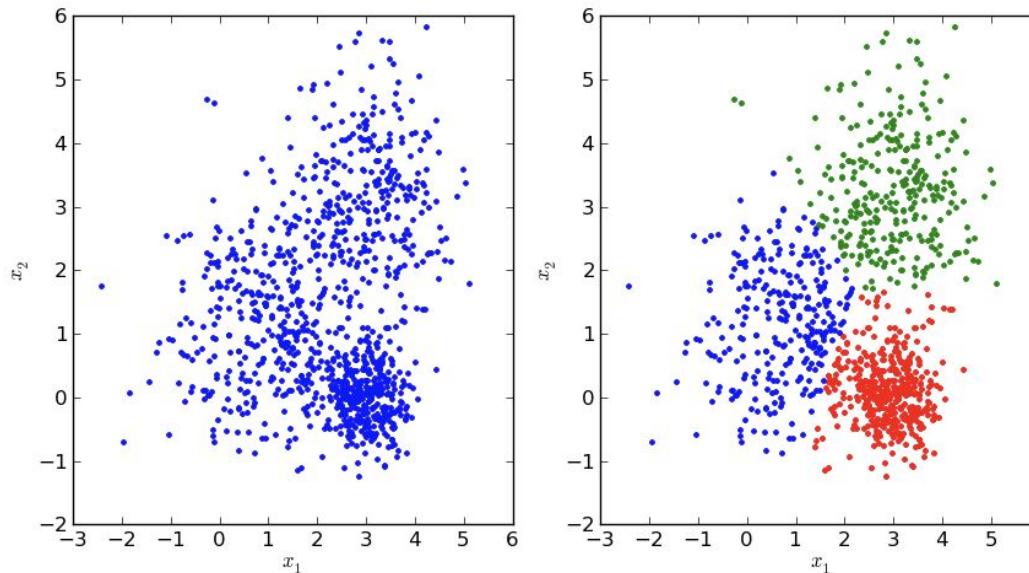
určení počtu
lidí v místnosti



Učení bez učitele (unsupervised learning)

Součástí trénovacích dat učení bez učitele jsou pouze vstupy.
Výstupní hodnoty nejsou předem známé.

Nejvýznamnější úlohou učení bez učitele je **shlukování (clustering)**.



Shlukování - příklady aplikací

Hledání genových rodin
analýzou DNA.



Identifikace komunit (podobných uživatelů) na sociálních sítích.

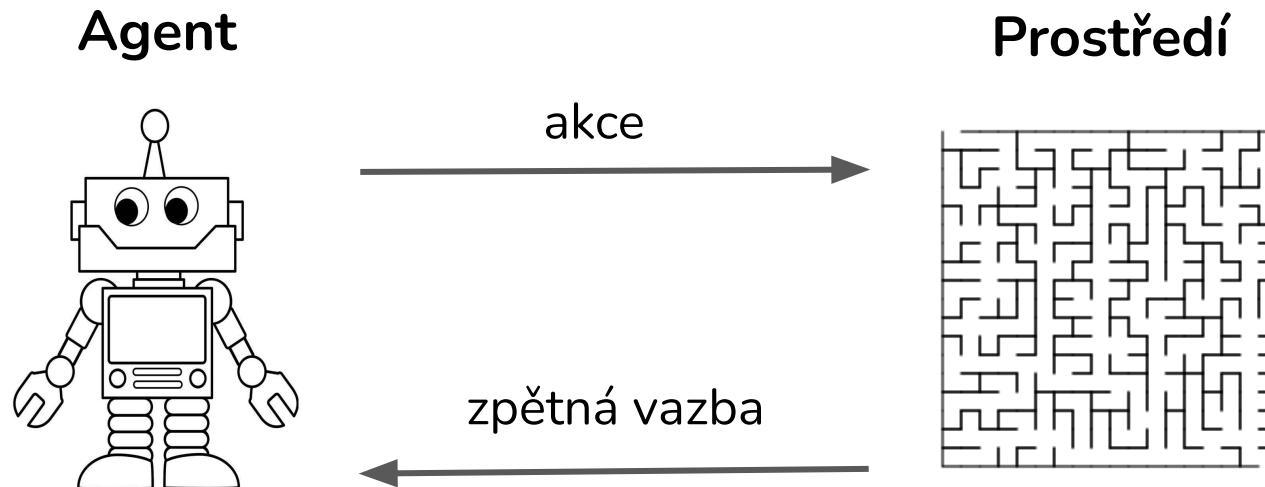


Doporučování filmů podle preferencí podobných uživatelů.



Zpětnovazební učení (reinforcement learning)

Učení funguje na principu agenta, který interahuje s prostředím a dostává zpětnou vazbu na své akce (odměnu nebo trest). Postupně se učí maximalizovat odměnu.



Zpětnovazební učení - příklady aplikací

samořiditelná auta,
drony



umělá inteligence
ve hrách



automatické
obchodování



regulace
komplexních
systémů



Nejvýznamnější oblasti využití strojového učení

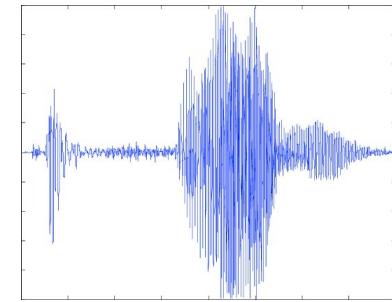
zpracování obrazu
a videa



zpracování
přirozeného jazyka



zpracování signálů
a časových řad



doporučování

robotika



amazon

Slabá a silná umělá inteligence

Umělá inteligence

Slabá UI

- vždy řeší jeden konkrétní problém, neumí se adaptovat na nové problémy
- všechna existující UI jsou slabou umělou inteligencí

Silná (obecná) UI

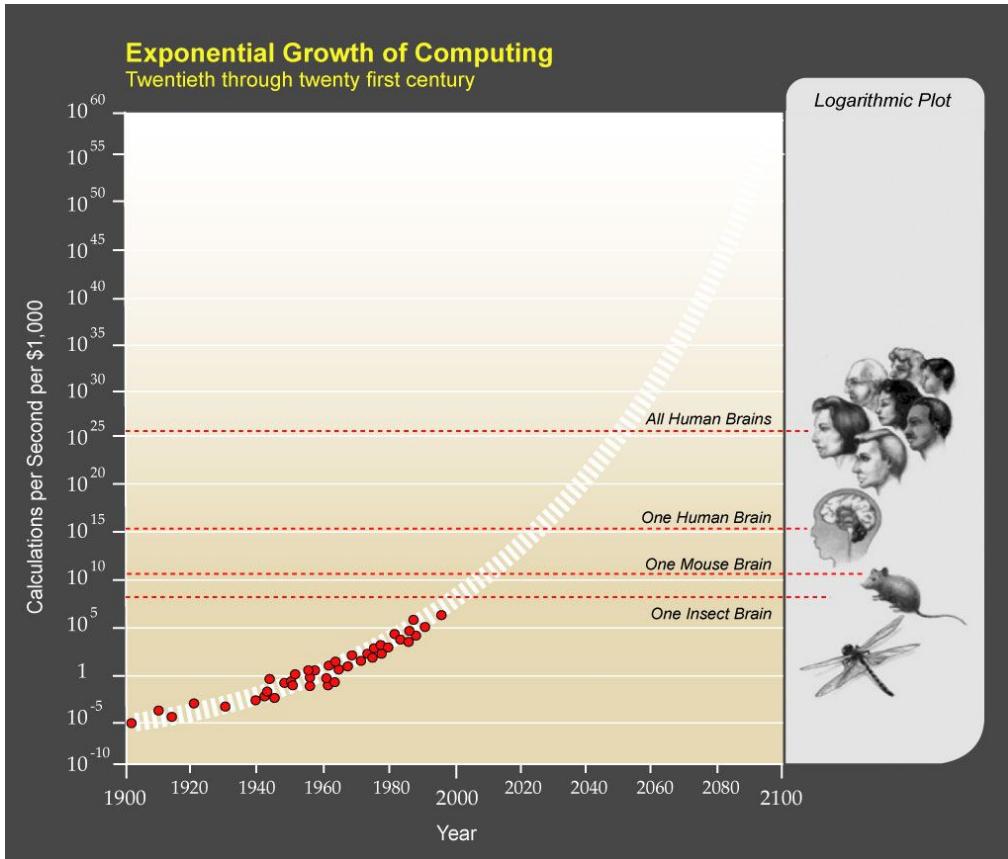
- umí cokoliv, co dokáže člověk, nebo dokonce více
- zatím neexistuje

Technologická singularita



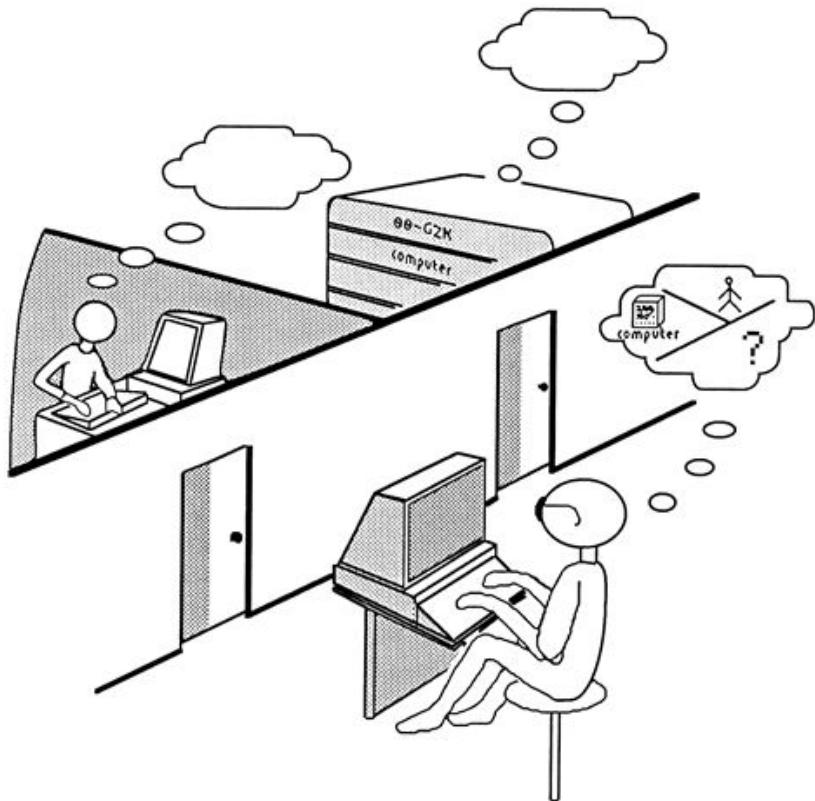
- hypotetický stav, kdy umělá inteligence dosáhne schopností člověka
- taková umělá inteligence by byla schopna sebe sama vylepšovat a dosáhnout tak *superinteligence*

Ray Kurzweil - Singularita je blízko



- Kurzweil argumentuje, že při současném exponenciálním růstu výkonu dosáhneme singularity kolem roku 2030.
- Odpůrci namítají, že dosažení stejného výpočetního výkonu jako lidský mozek ještě neznamená dosažení stejné inteligence.

Turingův test umělé inteligence

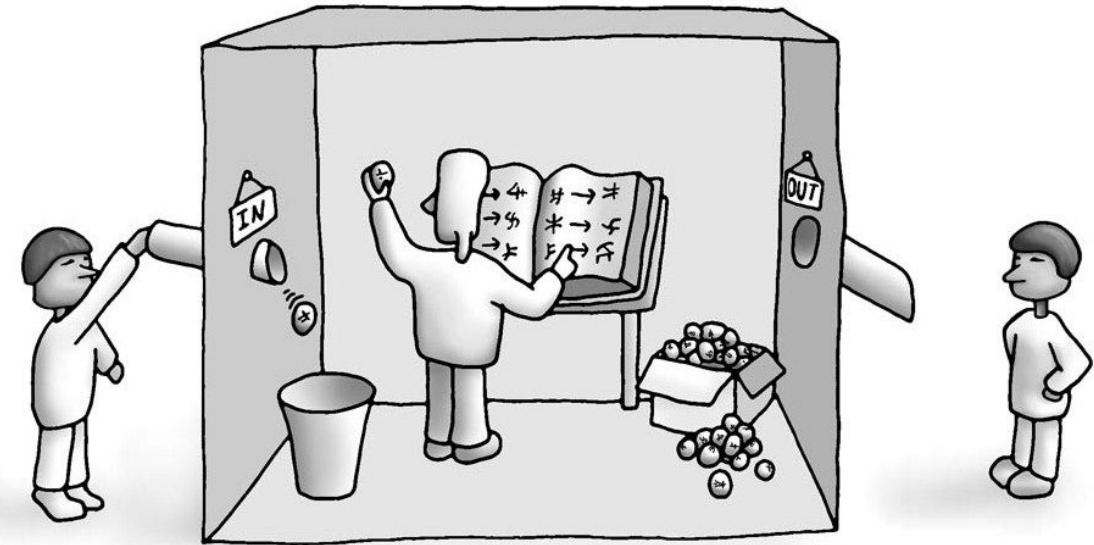


- test navržený pro ověření silné umělé inteligence
- podle A. Turinga je postačující, nikoliv nutnou podmínkou

Turingův test - vlastnosti

- Navržený Alanem Turingem v roce 1950, stále uznávaný, i když existuje mnoho kritiků.
- Zkoumá externí projevy inteligentního chování
- Je antropocentrický, ale nevylučuje existenci umělé inteligence, která test nesplní:
 - splnění Turingova testu \Rightarrow silnou AI
 - silná AI $\not\Rightarrow$ splnění Turingova testu

Kritika T. testu - Argument čínského pokoje



- Člověk v místnosti pouze manipuluje symboly na základě pravidel, ale čínsky vůbec nemusí rozumět.
- Autorem je filosof John Searle.
- Nejvýznamnější, ale zdaleka ne jediná kritika Turingova testu.

Loebnerova cena



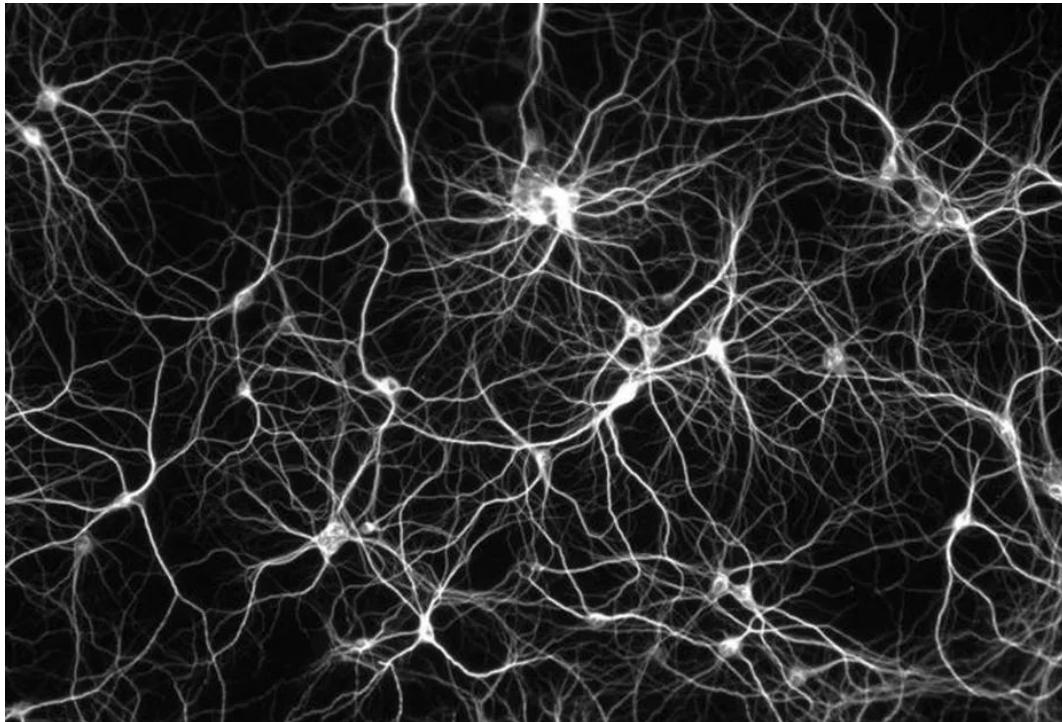
- V roce 1990 založil Hugh Loebner každoroční soutěž inteligenčních systémů s cílem splnit Turingův test.
- Cena za splnění testu je \$100.000. Zatím nebyla udělena.

Vyzkoušejte si ChatGPT



<https://chat.openai.com/chat>

Rizika umělé inteligence - vysvětlitelnost



Moderní systémy umělé inteligence u některých úloh v průměru překonávají schopnosti lidí (například hraní her, klasifikace obrázků apod.). Je však obtížné nebo nemožné jejich chování zdůvodnit pro člověka srozumitelným způsobem.

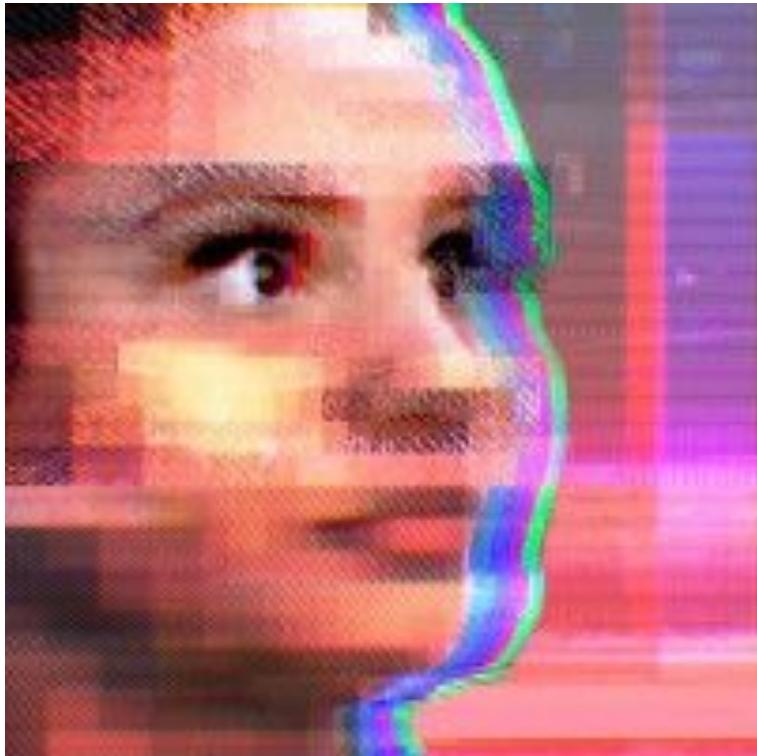
Není tedy zatím možné použít umělou inteligenci v kritických oblastech jako např. k plně automatické diagnóze a léčení pacientů.

Rizika umělé inteligence - sociální bubliny



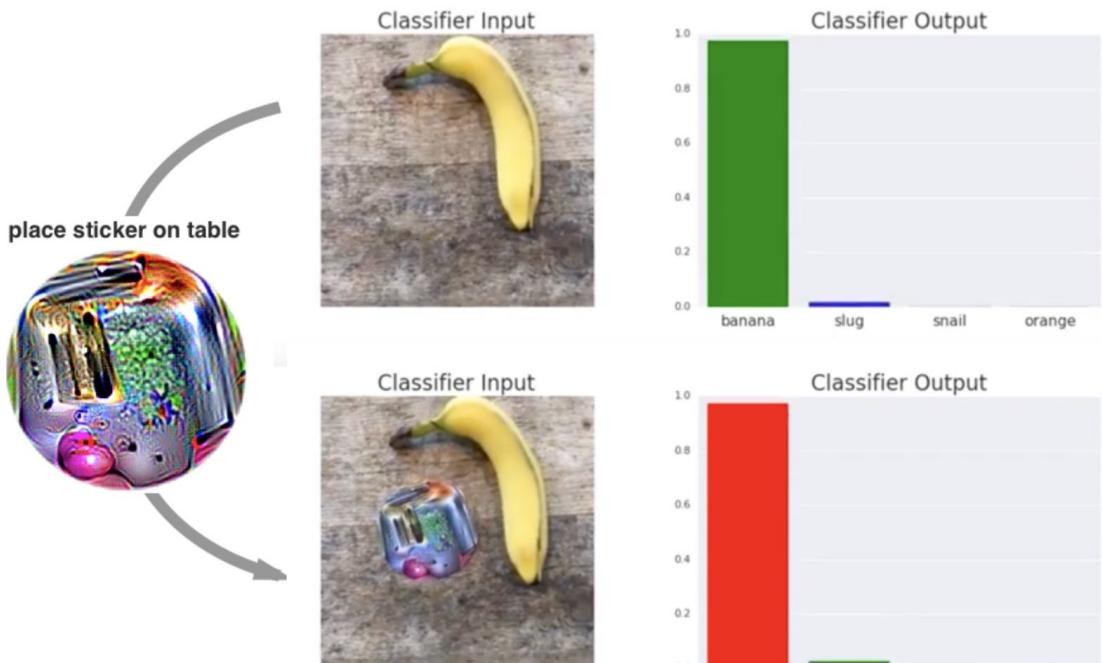
- Čím dál sofistikovanější doporučovací a personalizační systémy nás uzavírají do sociálních a informačních bublin.
- Je třeba si toho být vědomi a chovat se podle toho.

Rizika umělé inteligence - férovost a bias



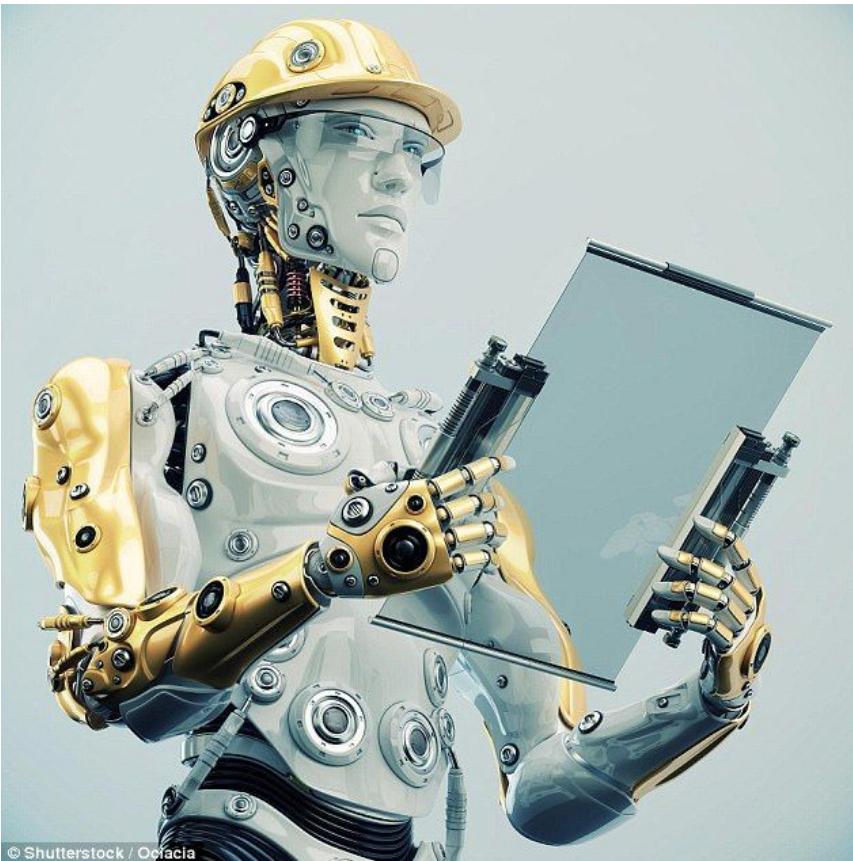
- Systémy umělé inteligence jsou do velké míry odrazem použitých trénovacích dat.
- Velmi snadno mohou např. diskriminovat skupiny lidí, které jsou v datech zastoupeny minimálně.
- V roce 2016 vytvořil Microsoft tweetujícího chatbota, kterého ale musel brzy po zveřejnění zastavit, protože se ze zpětné vazby od lidí naučil být rasistický a vulgární.

Rizika umělé inteligence - hacknutelnost



- Stejně jako u jiných počítačových technologií hrozí i u umělé inteligence možnost napadení nebo hacknutí.
- Pokud útočník dostane přístup k modelu umělé inteligence, je možné jej relativně snadno oklamat.
- Video zachycuje tzv. *adversarial patch*.

Rizika umělé inteligence - sebere nám práci?



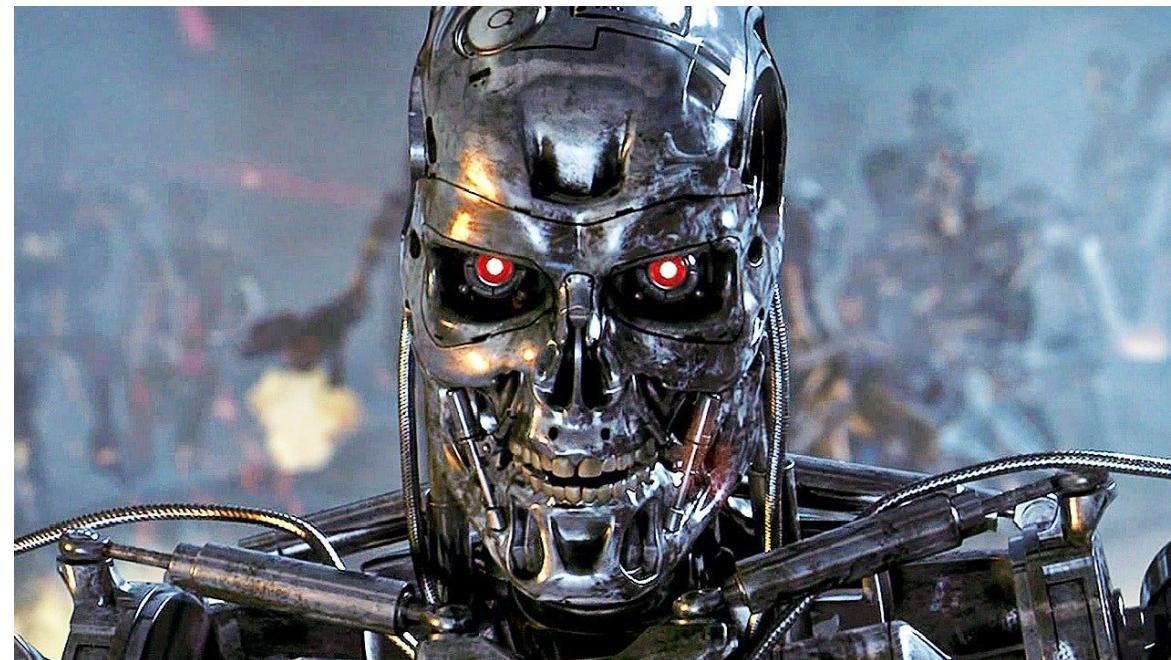
- Zřejmě nastane podobná situace jako na začátku průmyslové revoluce.
- Řada nekvalifikovaných pracovních pozic zanikne, ale spousta nových vznikne.
- Klíčem k úspěchu ve společnosti budoucnosti je vzdělání a schopnost adaptace.

Rizika umělé inteligence - autonomní zbraně



- Autonomní zbraně jsou potenciálně velmi nebezpečné.
- Nepotřebují dosažení silné umělé inteligence.
- Existují celosvětové snahy o jejich zákaz nebo regulaci jako např. u chemických zbraní.

Rizika umělé inteligence - existenční riziko



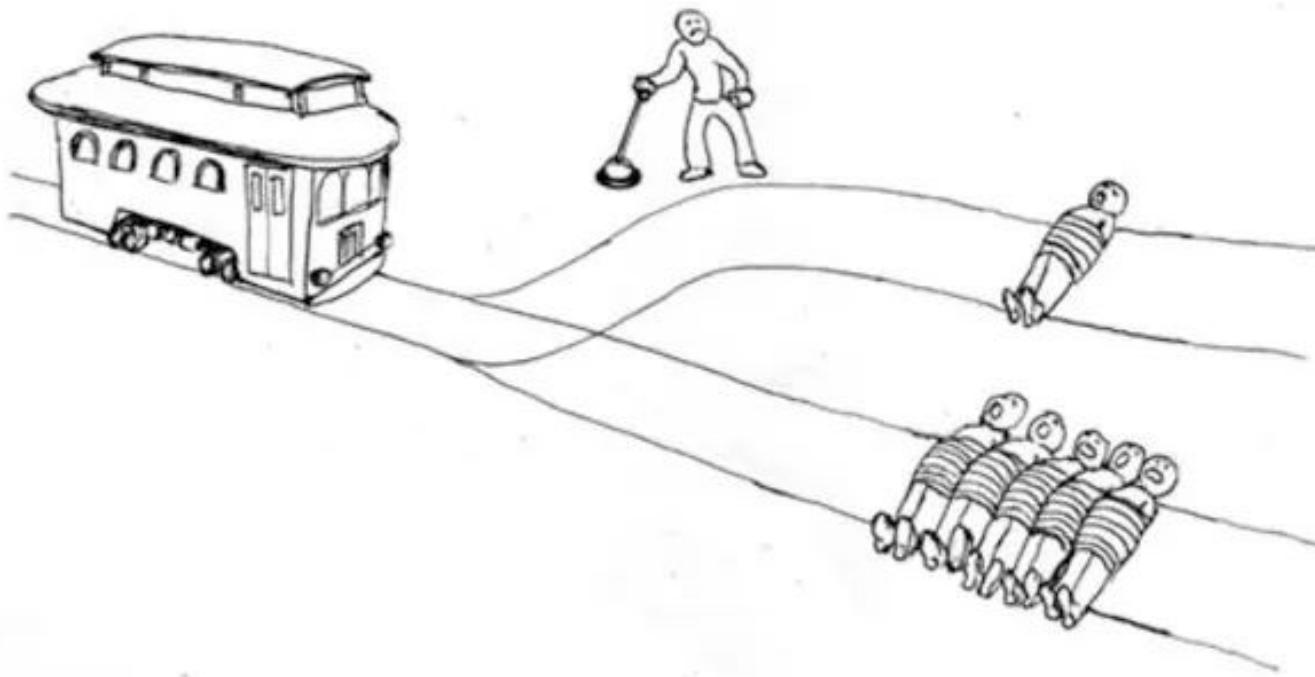
- Obava o zachování lidského druhu v případě vytvoření silné umělé inteligence.
- V blízkém časovém horizontu nepravděpodobné, ale přesto seriózně zkoumané.

Umělá inteligence - přínosy převažují nad riziky



- Umělá inteligence je jako oheň - dobrý sluha, ale zlý pán.
- Nemá smysl se jí bránit, ale je dobré znát rizika.
- Je třeba v ní vidět velkou příležitost.
- Přínosy jednoznačně převažují nad riziky.

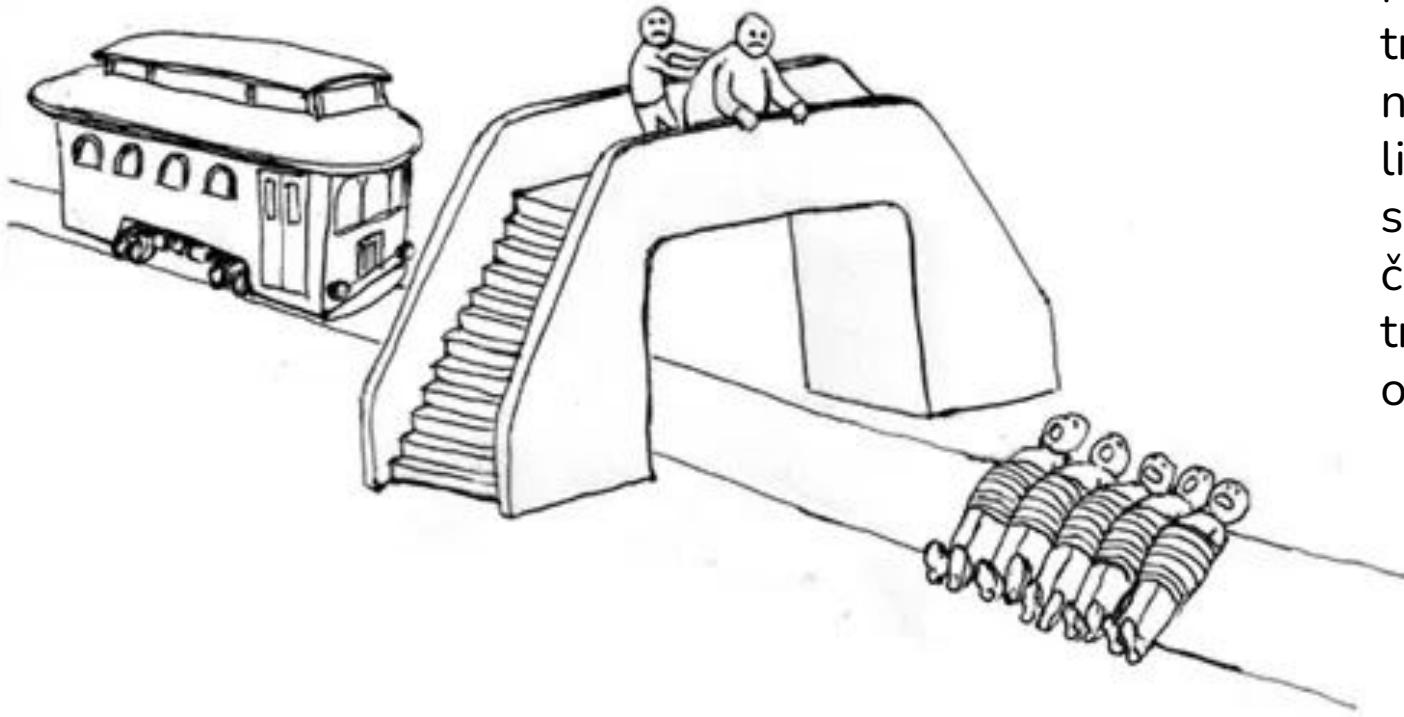
Morální otázky - Tramvajové dilema I



Po kolejích se řítí tramvaj a pokud nic neuděláte, zabije 5 lidí. Máte možnost přehodit výhybku tak, že tramvaj zabije pouze jednoho člověka.

Co uděláte?

Morální otázky - Tramvajové dilema II



Po kolejích se řítí tramvaj a pokud nic neuděláte, zabije 5 lidí. Máte možnost shodit z mostu člověka, jehož tělo tramvaj vykolejí, ale on zemře.

Co uděláte?

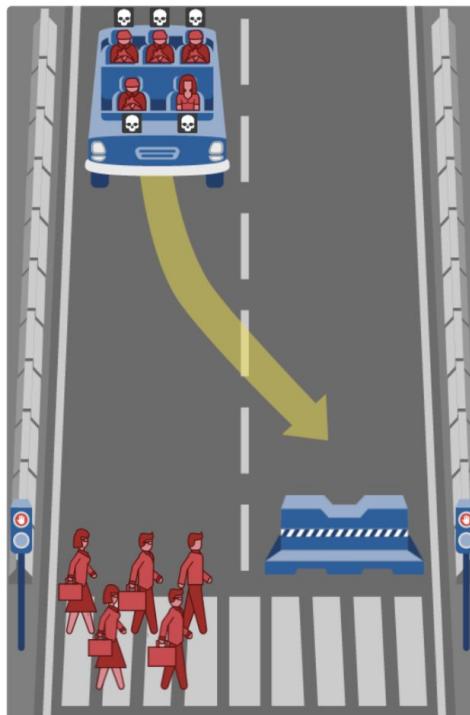
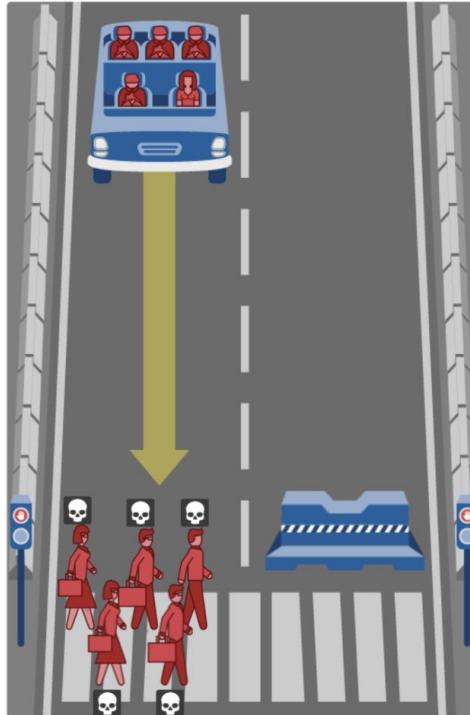
Morální otázky - MoralMachine.net



Home Judge Classic Design Browse About Feedback En

What should the self-driving car do?

1 / 13



Děkuji za pozornost

Jiří Materna

e-mail: jiri@mlcollege.com
web: www.mlcollege.com
twitter: @JiriMaterna
LinkedIn: <https://www.linkedin.com/in/jirimaterna/>