



RIZIKÁ A MOŽNOSTI VYUŽÍVANIA UMELEJ INTELIGENCIE

Peter Šantavý





ZJAVENIE, MAGISTÉRIUM A UMELÁ INTELIGENCIA

AI – artificial intelligence, umelá inteligencia



PREČO AI?

Lebo nám klasické algoritmy nestačia...

AI – artificial intelligence

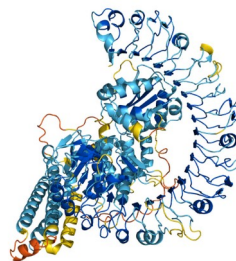
Prečo AI?

Lebo nám klasické algoritmy nestačia...

– úlohy, ktoré **nevieme riešiť**



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



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

Ak je to ťažké naprogramovať,
nech sa to stroje naučia!



P R E Ā O A I A Ź T E R A Z ?

Lebo uŹ na to máme schopnosti a vyuŹitie...

28

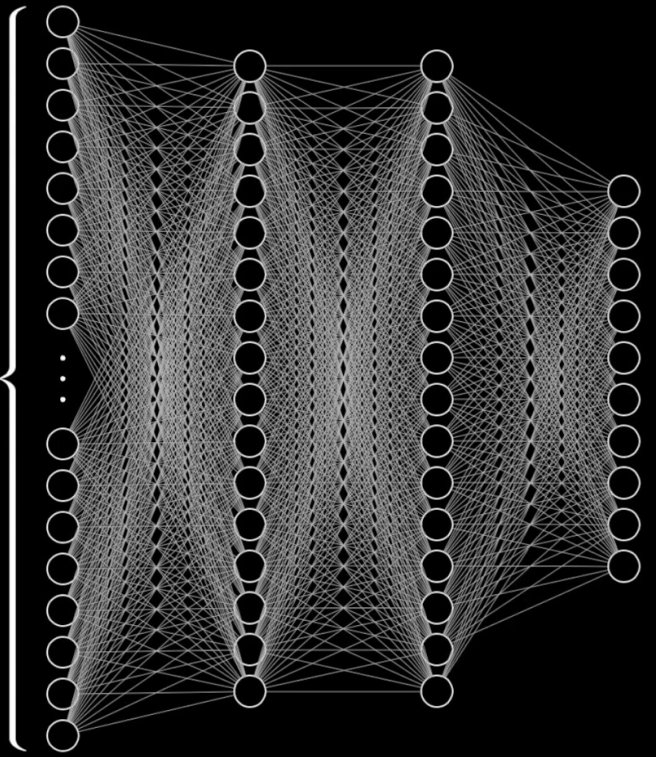
$$28 \times 28 = 784$$

13 002 !!!

28



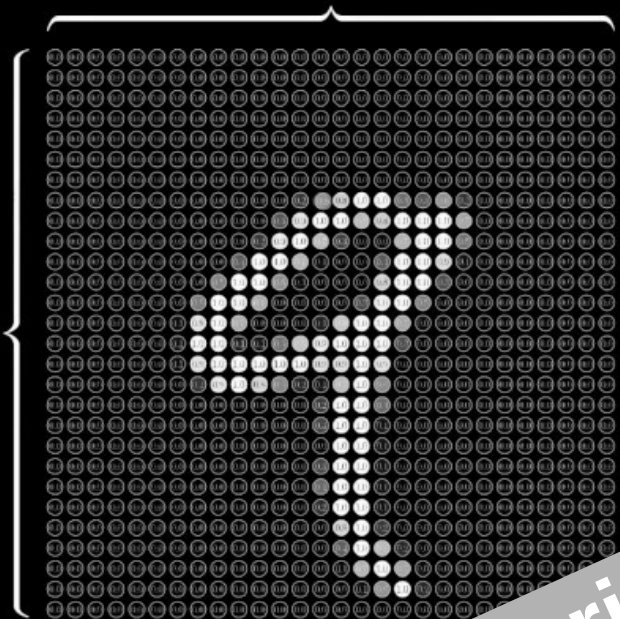
784



28

$$28 \times 28 = 784$$

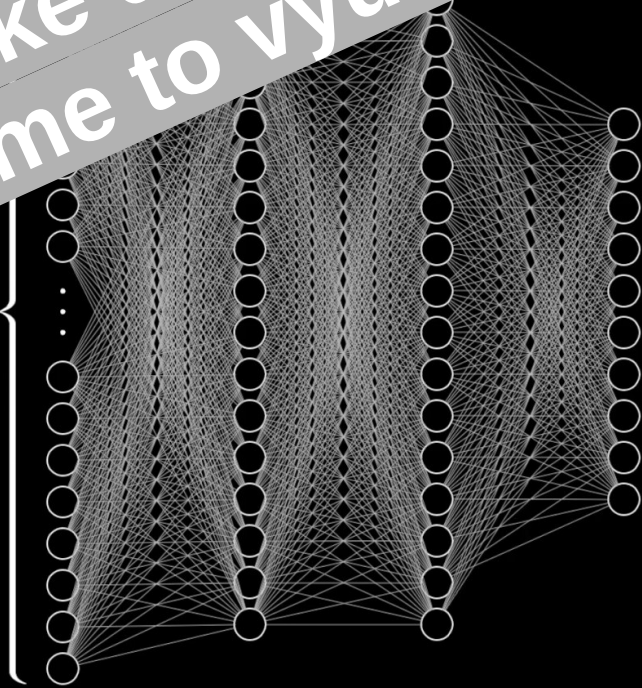
28

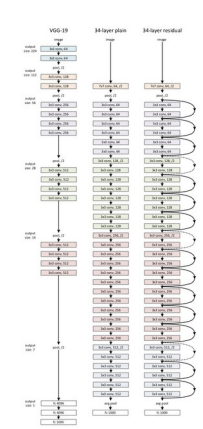
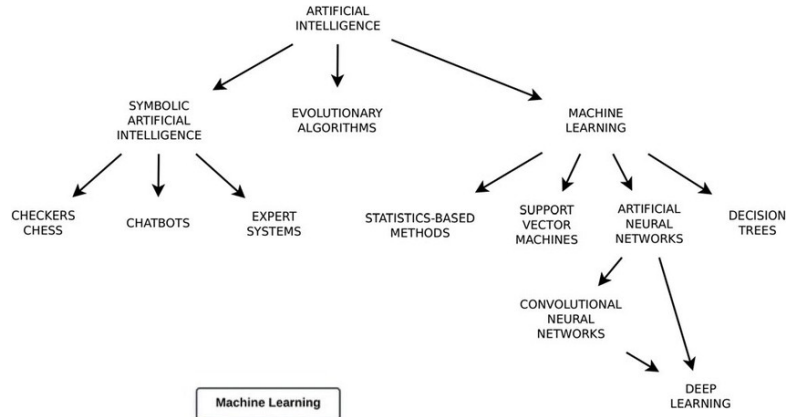


13 000

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

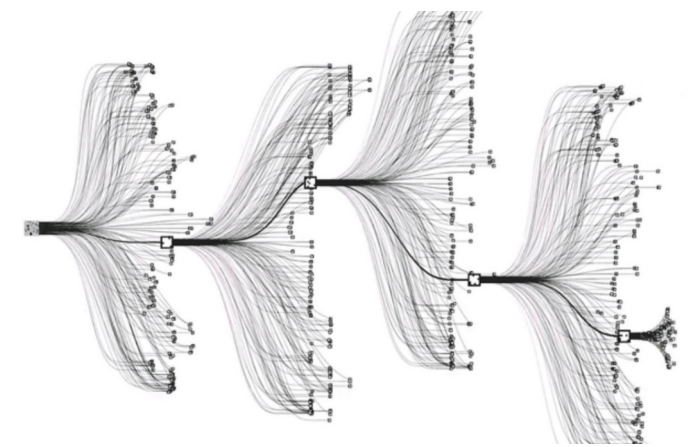
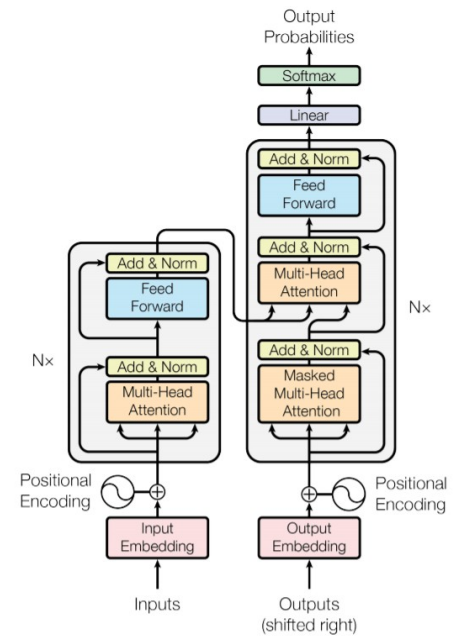
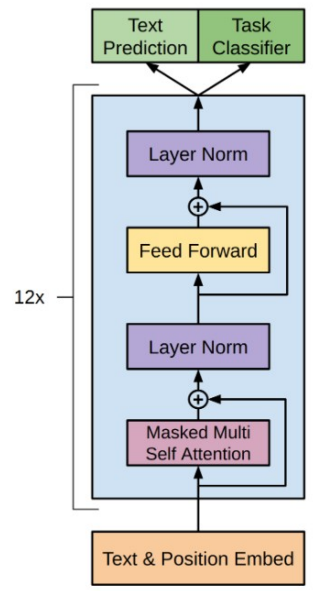
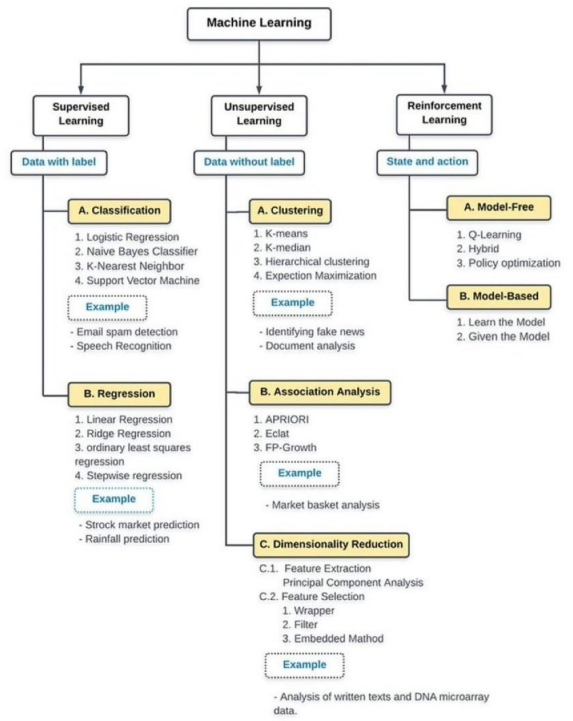
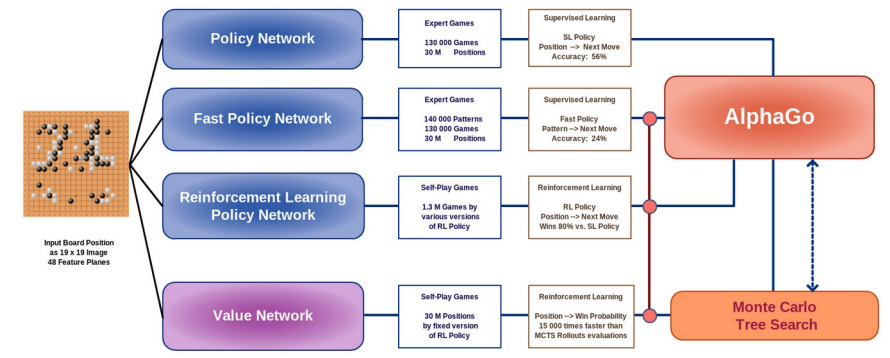
784





AlphaGo Overview

based on: Silver, D. et al. Nature Vol 529, 2016
copyright: Bob van den Hoek, 2016





PREČO AI?

Existujú ešte nejaké iné dôvody, prečo AI?

...finančné, mocenské, vojenské, ideologické,
vedecké, spoločenské, technologické,...

...bytostné

Definícia umelej inteligencie

Inteligenciu ľudského bytia vnímame ako schopnosť vnímať, chápať a spracovávať informácie, učiť sa, odôvodňovať, plánovať, riešiť komplexné problémy a rozhodovať.

(James Flynn: Súbor zručností, medzi ktoré patrí abstraktné a logické myslenie, ďalej predstavivosť, a teda aj schopnosť uvažovať o hypotetických možnostiach, a tiež aj jazykový cit.)

Umelá inteligencia je o vytváraní systému, ktorý vykazuje také správanie, o ktorom si myslíme, že vyžaduje inteligenciu (AI100).

Na prasknutie naplnený *kufor* rôznych technológií (Marvin Minsky)...

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 **intelligentné správanie na základe modelov a aplikovaných metód i tréningových dát.**

Hovoríme teda o systémoch, ktoré 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.**

Riziká a limity súčasných systémov AI

Technologické riziká a limity

Dôsledky – sociologické, psychologické, civilizačné

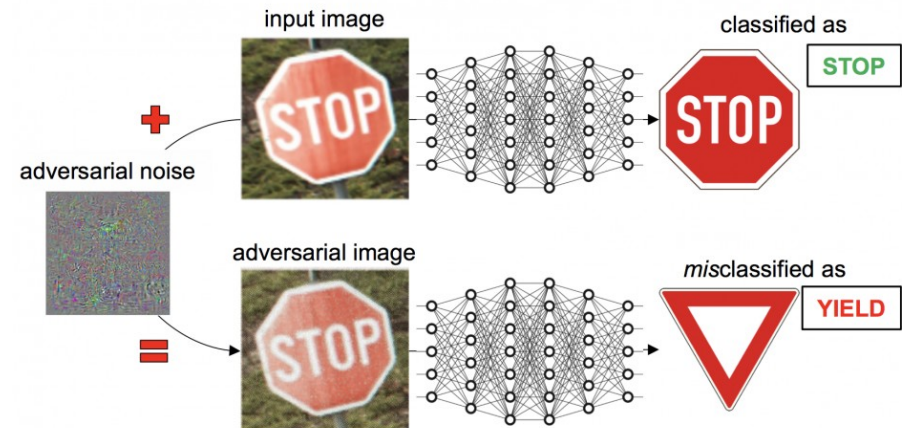
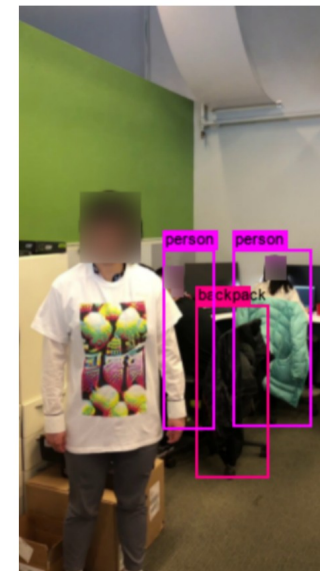
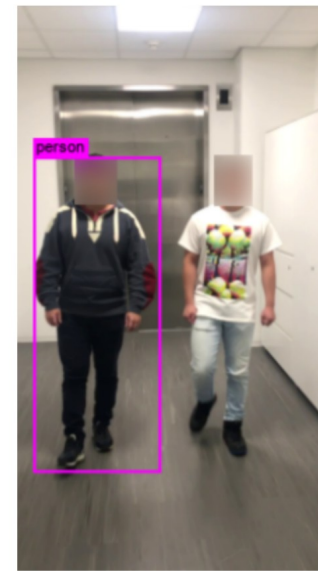
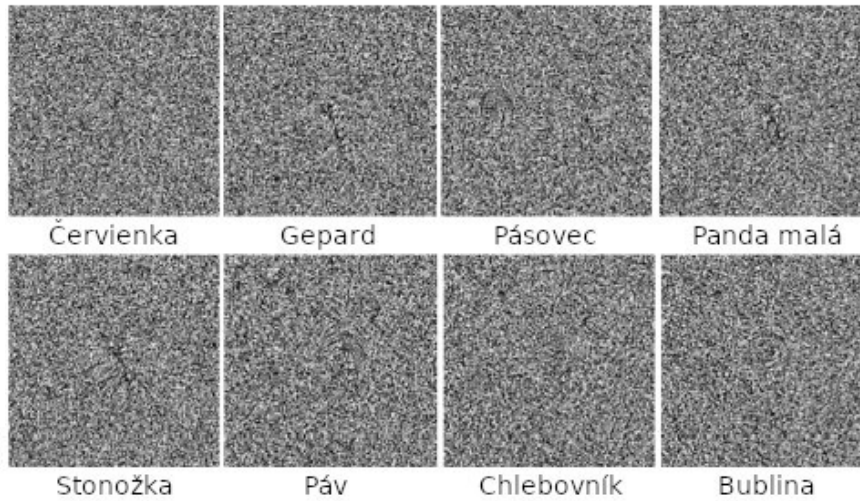
- závislosť
- digitálna demencia
- neschopnosť robiť vlastné zodpovedné rozhodnutia
- extrémne zásahy do spoločnosti - algokracia
- nástroj moci a ideológie
- cítiť sa ako boh – opantaný možnosťami zneužiť na manipuláciu a ovládanie, monitoring, obohatenie, zneužívanie,...



Riziká a limity súčasných systémov AI

Technologické riziká a limity

AI – artificial intelligence



Technologické riziká a limity ANI

Zraniteľnosti, slabiny a klamanie systémov strojového učenia

- malá množina tréningových dát (training dataset)
- nesprávne zvolená/nekvalitná množina tréningových dát a predsudky (biases)
- nadmerné prispôsobovanie sa tréningovým údajom (overfitting to training data)
- efekt dlhého chvosta (long-tail effect)
- klamanie hlbokých sietí a ich zraniteľnosti (adversarial examples, hacking,...)
- povery (superstition)
- špecifické zlyhania generatívnych systémov (jailbreak: roleplay, base64, UTS, prompt injections, data poisoning / backdoor attacks,...)

Technologické riziká a limity ANI

Zraniteľnosti, slabiny a klamanie systémov

- malá množina trénovacích dát (training data)
- nesprávne zvolená/nekvalitná tréningová množina a predsudky (biases)
- nadmerné prispôbenie na tréningovú množinu (overfitting to training data)
- efekt dlhého času tréningu
- klamanie zraniteľnosti (adversarial examples, hacking,...)
- klamanie generatívnych systémov (jailbreak: roleplay, base64, UTS, prompt injections, data poisoning / backdoor attacks,...)

Zlyhania sú odlišné od ľudských, ťažko predvídateľné, niekedy ľahko vykonateľné a mnohokrát prekvapivo robustné...

Procesné riziká a limity ANI

Procesné riziká v reálnom nasadení

- **útoky na dôvernost'** (confidentiality attacks)

Útoky zamerané na dáta uložené v rámci modelov systémov AI.

- **útoky na zraniteľnosti** (evasion attacks)

Odhaľovanie a zneužitie existujúcich zraniteľností v modeloch za účelom zmanipulovania výsledkov systémov AI.

- **útoky s cieľom ovplyvniť model** (poisoning attacks)

Ovplyvňovanie modelu, tréningového procesu a tým aj výslednej činnosti.

Etické výzvy, doteraz rozoberané ako súčasť návrhu a realizácie systémov AI, sa rozširujú aj o oblasť etiky použitia, resp. riziká zneužitia.

Sekundárne technologické rizikové faktory

Kybernetická bezpečnosť – je to napadnuteľné

- vektor útoku na zraniteľnosti AI
- **neexistuje dokonale bezpečný a spoľahlivý systém**

Infraštruktúra a komplexnosť – je to náročné

- s rozvojom algoritmov strojového učenia sa požadovaný výpočtový výkon v poslednej dekáde zvyšuje **exponenciálne**
- **extrémna komplexnosť**, je rizikovým faktorom bezpečnosti a stability fungovania systémov

Sekundárne technologické rizikové faktory

Kybernetická bezpečnosť – je to n

– vektor útoku na zraniteľnosť

– neexistuje dokon

Infraštr

ročné

účenia sa požadovaný výpočtový výkon

dekáde zvyšuje **exponenciálne**

– **extrémna komplexnosť**, je rizikovým faktorom bezpečnosti a stability fungovania systémov

**Problémom nie je len zlyhanie AI, ale i človek:
zodpovednosť vývoja, etika použitia,
riziká zneužitia...**

Vybrané riziká generatívnych systémov AI

Extrémna šírka využitia generatívnych systémov (LLM/MLM/GPT/...)

- LLM sa dokáže učiť vzory, kontext aj sémantiku jazyka, porozumieť dlhým konverzáciám, poskytovať súvislé a kontextovo relevantné odpovede

Halucinovanie

- vymýšľanie si odpovedí, ktoré systém predkladá ako relevantné a správne

Predsudky a neobjektívne výstupy

- riziko poloprávd a nesprávnych odpovedí

Nejasný spôsob narábania s údajmi

- dôsledkom môže byť únik dôverných dát, prehrešky voči ochrane osobných údajov i problémy s autorskými právami

Vybrané riziká generatívnych systémov AI

Extrémna šírka využitia generatívnych systémov (LLM)

- LLM sa dokáže učiť vzory, kontext aj sémantiku z veľkým množstvom dát, ktoré sa používajú v konverzáciám, poskytovať súvislé a kontextuálne relevantné odpovede

Halucinovanie

- vymýšľanie si odpovedí, ktoré sa neopierajú o faktografické údaje, ale sa zakladajú ako relevantné a správne

Predsudky a stereotypy

- riziko reprodukovania existujúcich predsudkov a stereotypov v generovaných odpovediach

Bezpečnosť a ochrana údajov

- pri používaní generatívnych systémov môže byť únik dôverných dát, prehrešky voči ochrane osobných údajov i problémy s autorskými právami

Generatívne nástroje AI nemôžeme vnímať ako faktografické, spoľahlivé a etické zdroje!

Etické riziká generatívnych systémov

Neschopnosť rozlišovať morálne dobré a zlé – systémy AI skutočnú inteligenciu nemajú, len ju simulujú. Nechápu zmysel a dôsledky.

Riziko digitálnej demencie – prichádza k degradácii intelektuálnych schopností a k dopamínovej závislosti na základe ponorenia sa do virtuálneho sveta, v ktorom systémy umelej inteligencie v čoraz väčšej miere supľujú kognitívne činnosti človeka, a to spôsobom, na ktorý nie sme evolučne vôbec pripravení.

Ďalšie riziká: **erózia identity, algoritmické modelovanie histórie** a pod...

Etické riziká generatívnych systémov

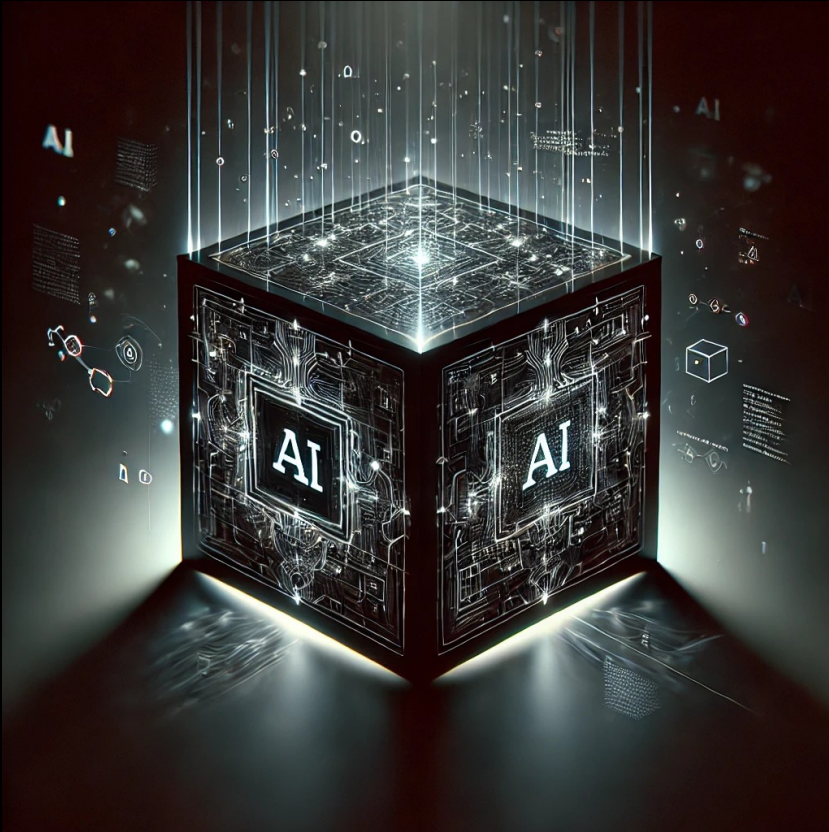
Neschopnosť rozlišovať morálne dobré a zlé – generatívne AI systémy, ktoré nemajú inteligenciu, len ju simulujú. Nemajú morálne hodnoty a emócie.

Riziko digitálnej demencie – závislosť na gradácii intelektuálnych schopností a kognitívnych funkcií na základe ponorenia sa do virtuálnych svetov. Generatívne AI systémy umelej inteligencie v čoraz väčšej miere nahradia ľudskú činnosť človeka, a to spôsobom, na ktorý nie sme vôbec pripravení.

Ďalšie riziká: **erózia identity, algoritmické modelovanie histórie** a pod...

Vďaka systémom AI môžeme postupne strácať časť našich schopností a psychickej odolnosti.

BLACK BOX



Prakticky nevieme, na základe čoho robia hlboké neurónové siete svoje rozhodnutia.

V zásade nevieme, čo presne sa neurónová sieť naučila a ako spoľahlivo to dokáže aplikovať nielen v bežnej prevádzke, ale osobitne v hraničných situáciách za extrémnych podmienok na vstupe, či pri činnosti systému.

BLACK BOX



Pracujeme na vysvetliteľných a interpretovateľných systémoch AI ;-)

Prakticky nevieme, ako to funguje, pretože to robia hlboké učenie. Zatiaľ čoho robia hlboké učenie, ktoré sa naučilo svoje rozhodnutia.

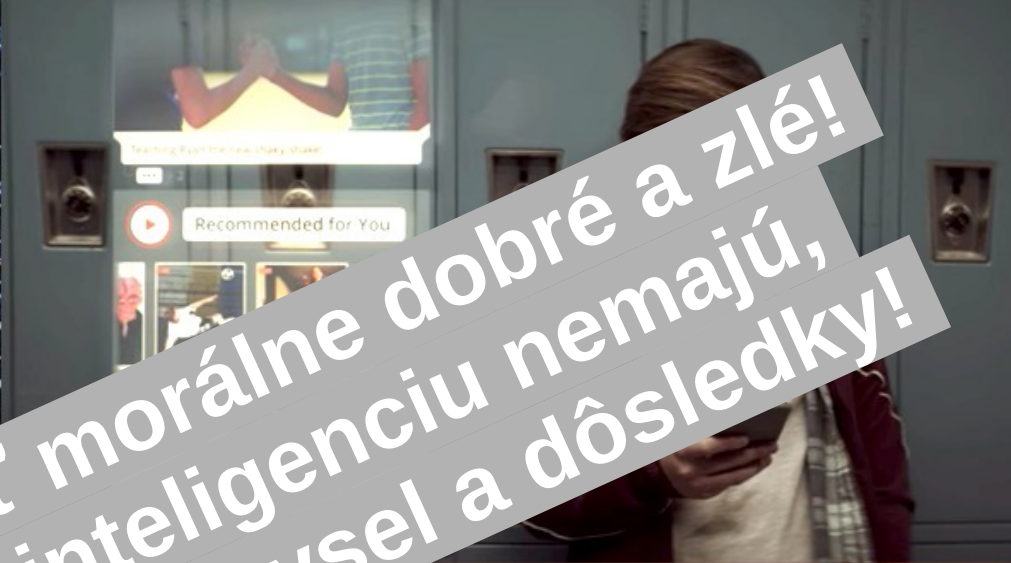
Zasade nevieme, ako to funguje, pretože to robia hlboké učenie, ktoré sa naučilo svoje rozhodnutia. Presne sa neurónová sieť naučila a ako spoľahlivo to dokáže aplikovať nielen v bežnej prevádzke, ale osobitne v hraničných situáciách za extrémnych podmienok na vstupe, či pri činnosti systému.

Vybrané dôsledky a riziká pre človeka

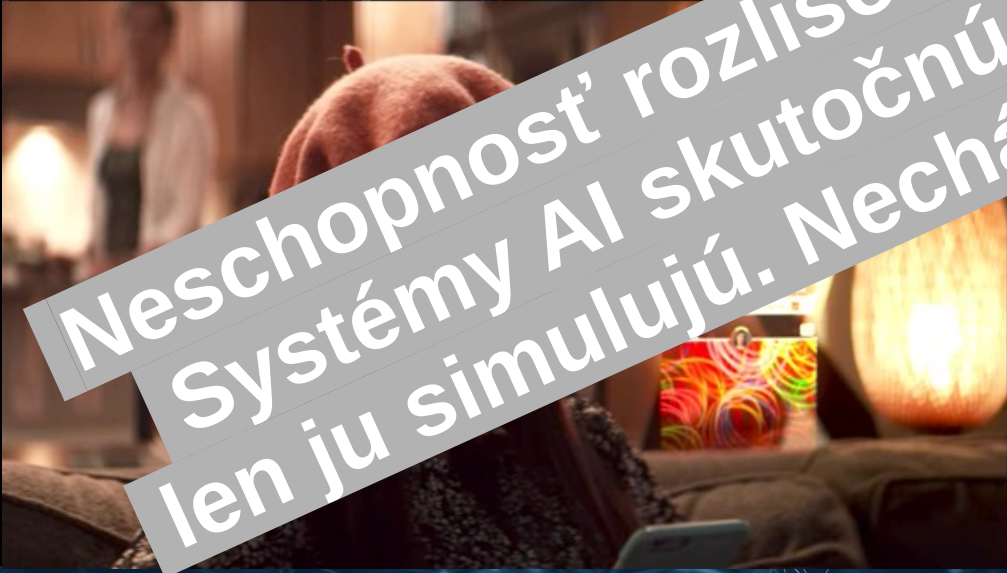
- extrémne **zhromažďovanie dát**
- neustály **dohľad a sledovanie** bez kontroly
- **ovplyvňovanie** nášho vnímania a konania
- modely **predikcie nášho správania**
- technológie AI vedú **k závislosti a manipulácii**
- **diskutabilná relevantnosť** činnosti a výsledkov systémov AI
- **digitálna demencia**, psychologické a sociologické dôsledky
- **algoritmické riadenie štátu**
- **vojenské nasadenie**

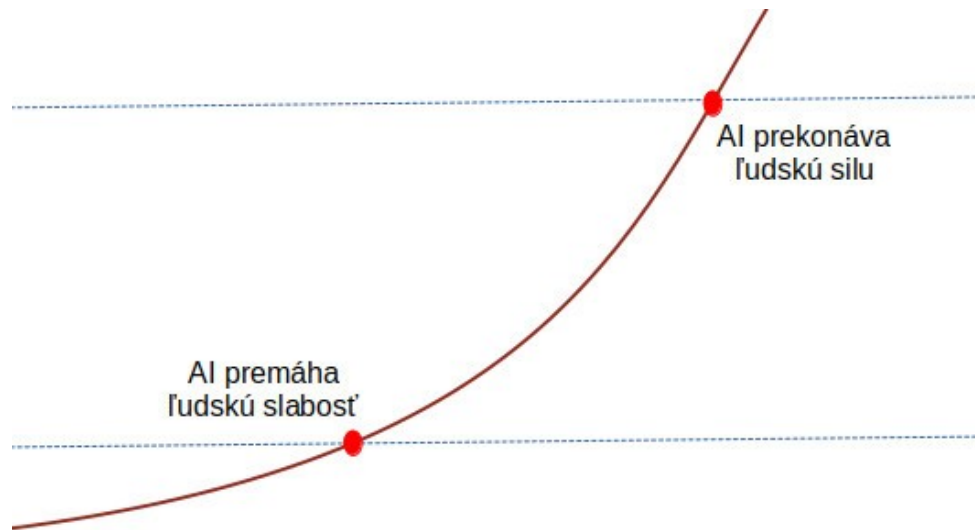
N

/the social dilemma



Neschopnosť rozlišovať morálne dobré a zlé!
Systémy AI skutočnú inteligenciu nemajú,
len ju simulujú. Nechápu zmysel a dôsledky!





Míľnikom, ktorého by sme sa mali obávať, nie je budúca technologická singularita v oblasti umelej inteligencie, v ktorej AI prevýši náš intelekt, ale oveľa skôr moment, keď **technológia ovládne a prekoná naše slabosti...** už vtedy prichádza víťazstvo umelej inteligencie a porážka ľudstva.



AI NEPOZNÁ ZL'UTOVANIE

AI môže ubližovať



Mocenské, vojenské, ideologické zneužitie AI!

CHINA'S SOCIAL CREDIT SYSTEM

It's been dubbed the most ambitious experiment in digital social control ever undertaken. The Chinese government plans to launch its Social Credit System nationally by 2020.

WHAT'S THE AIM?

The system intends to monitor, rate and regulate the financial, social, moral and, possibly, political behavior of China's citizens — and the country's companies — on a scale of purpose.

It will be used to monitor and regulate the behavior of citizens and companies in pilot schemes and to set expectations.

HOW DOES IT WORK?

Each citizen is expected to be given a social credit score that will increase or decrease depending on whether the subject's social behavior is acceptable.

The system is expected to draw on huge amounts of data about each and every individual, gathered from traditional sources — such as financial, criminal and government records and existing data from registries offices or school officials — along with digital sources. The latter include data collected on the Internet, such as the subject's search history, shopping preferences on e-commerce sites and interactions on social media.

Moreover, the system could also rely on information obtained through video surveillance systems with help from facial recognition technology.



REWARDS AND PUNISHMENTS

Citizens with high scores get to enjoy special 'privileges' while those with low scores ultimately get treated as second-class

HIGH SCORES CAN LEAD TO

- ★ Priority for school admissions.
- 🚗 Easier access to cash loans and consumer credit.
- 🚲 Deposit-free bicycle and car hire.
- 🏊 Free gym facilities.
- 🚇 Cheaper public transport.
- 🏥 Shorter wait times in hospitals.
- 🏢 Fast-track promotion at work.
- 🏠 Jumping the queue for public housing.
- 💰 Tax breaks.

PUNISHMENTS CAN LEAD TO

- 🚫 Denial of licenses, permits and access to some social services.
- 🚫 Exclusion from booking flights or high-speed train tickets.
- 🚫 Less access to credit.
- 🚫 Restricted access to public services.
- 🚫 Ineligibility for government jobs.
- 🚫 No access to private schools.
- 🚫 Public shaming: exposure either online or on TV screens in public spaces of the names, photos and ID numbers of blacklisted citizens, phone dial tones mandated by authorities that inform people that they are calling a 'dishonest debtor'.

BertelsmannStiftung



**AI JE ZABÁVAČ, KTORÉMU
NESMIETE VERIŤ**

This image was AI-generated



Eyeglasses appear distorted and fused to his cheek, eye area and the shadow

Crucifix is only hanging by one half of the chain, the other half is missing



Kredit: AP News

This image was AI-generated



Eyeglasses appear distorted and fused to his cheek, eye area and the shadow

half is
missing

Deepfake: klamanie, nekalá konkurencia,
pornografia, kriminalita...





Nekonečné možnosti tvorby audiovizuálneho obsahu, ktorý vytvára virtuálny svet a neskutočnú realitu...

Strata zmyslu pre realitu a skutočné hodnoty!



Nekonečné možnosti tvorby audiovizuálneho obsahu, ktorý vytvára virtuálnu a neskutočnú realitu.

Strata zmyslu pre

Matrix: Neo bojuje za niečo, čo je skutočné...
“A koho to zaujíma, či je to skutočné?”



**AI JE RIZIKOM A VÝZVOU
PRE NAŠU INTELIGENCIU**

RIZIKO DIGITÁLNEJ DEMENCIE

Dlhodobé a nesprávne využívanie systémov AI, osobitne generatívnych AI a algoritmov sociálnych sietí, prináša **intenzívny útok na naše kognitívne schopnosti a psychiku.**

Môže prichádzať k degradácii intelektuálnych schopností a k dopamínovej závislosti na základe ponorenia sa do virtuálneho sveta, v ktorom systémy AI v čoraz väčšej miere supľujú kognitívne činnosti človeka, a to spôsobom, na ktorý nie sme evolučne vôbec pripravení. Mozog sa mení...





RIZIKO DIGITÁLNEJ DEMENCIE

Dlhodobé a nesprávne využívanie digitálnych technológií, osobitne generatívnych AI a sociálnych sietí, prináša **intenzívne kognitívne vyčerpanie** a **postupnú strátu kognitívnej odolnosti**.

Vďaka systémom AI môžeme postupne strácať časť našich schopností a psychickej odolnosti. To vedie k degradácii intelektuálnych schopností a k dopamínovej závislosti na základe virtuálnych odmen. Človek sa do virtuálneho sveta, v ktorom systémy AI v čoraz väčšej miere supľujú kognitívne činnosti človeka, a to spôsobom, na ktorý nie sme evolučne vôbec pripravení. Mozog sa mení...

Riziko digitálnej demencie

Narušenie psychického vývoja priskorým, resp. nezrelým používaním digitálnych technológií a technológií umelej inteligencie.

Obmedzenie kognitívnych schopností a znižovanie inteligencie suplovaním činností mozgu systémami generatívnej AI.

Riziko dopamínovej závislosti na základe neustálych podnetov, ktoré súvisia s uvoľňovaním dopamínu a pozitívnou reakciou našej mysle.

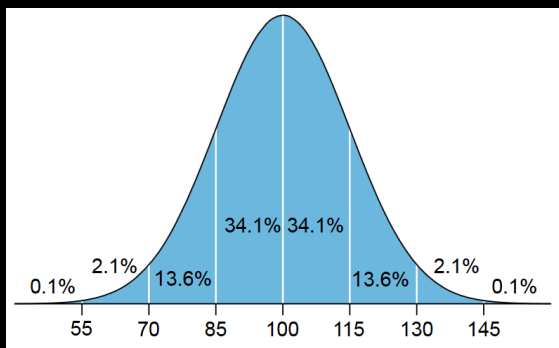
Schopnosť a možnosť manipulácie vlastníckmi systémov AI.

Strata kritického myslenia a nekritické preberanie výsledkov systémov AI spojené s relativizáciou pravdy, vytváraním sociálnych bublín a pod.



INTELIGENČNÉ NOŽNICE, KTORÉ SA OTVÁRAJÚ

Riziko znižovania inteligencie väčšiny populácie a súčasne intelligenčný rast menšiny, ktorá bude odolná voči digitálnej demencii.



Riziko digitálneho rozdelenia (digital divide) i v oblasti umelej inteligencie.

Výzva a varovanie

Vytvorenie nesprávnych návykov a spôsobov využívania AI môže mať negatívne dôsledky na celý život.

Ak škola nenaučí študentov kognitívne náročnejšie schopnosti ako to, čo dokáže AI, budú na trhu práce i v reálnom živote znevýhodnení.

Naviac, AI dokáže aj existujúce kognitívne schopnosti zmenšiť.

Výzva a varovanie

Vytvorenie nesprávnych návykov a spôsobov práce môže mať negatívne dôsledky na celý život.

Ak škola nenaučí žiakov kriticky hodnotiť a vyvíjať kriticky najmä tie najviac náročnejšie schopnosti ako to, čo dokáže AI, môže to byť pre nich veľkým znevýhodnením. Práca i v reálnom živote znevýhodnení.

AI dokáže aj existujúce kognitívne schopnosti zmenšiť.

Dôveryhodné systémy AI a ich etické využívanie sú nutnou podmienkou budúcnosti!



AI – DOBRÝ SLUHA, ZLÝ PÁN

Ak nezdoláme úskalia a výzvy, AI bude zlým pánom...

Na dobro človeka zameraná umelá inteligencia

UMELÁ INTELIGENCIA ZAMERANÁ NA DOBRO ČLOVEKA

(human-centered and beneficial artificial intelligence)

Známy a všeobecne prijímaný princíp, ktorý by však mal:

- byť chápaný v duchu kresťanskej antropológie, resp. klasickej filozofickej antropológie (predovšetkým biologickej a kultúrnej)
- zahŕňať každú ľudskú bytosť a nikoho nediskriminovať
- mať na zreteli dobro ľudstva a spoločnosti, chrániac pri tom a rešpektujúc dobro každej ľudskej bytosti
- sa vyznačovať starostlivosťou o náš „spoločný a zdieľaný domov“, teda o celý stvorený svet

Riziká: transhumanistická redefinícia človeka, digital divide,...

Na dobro človeka zameraná umelá inteligencia =

= dôveryhodné systémy AI, ktoré musia byť

- **funkčné a užitočné** – navrhnuté a realizované tak, aby vykonávali požadovanú činnosť
- **legálne** – musia vyhovovať požadovaným normám, zákonom i reguláciám a spĺňať všetky platné zákony a predpisy
- **etické** – musia rešpektovať etické zásady a hodnoty
- **odolné, resp. robustné*** – dosahujúce potrebné štandardy bezpečnosti a spoľahlivosti¹ nielen z technologického hľadiska, ale zohľadňujúce aj sociálne prostredie a dopady na spoločnosť

* schopnosť bezpečne a spoľahlivo pracovať, resp. fungovať za akýchkoľvek podmienok

Na dobro človeka zameraná umelá inteligencia

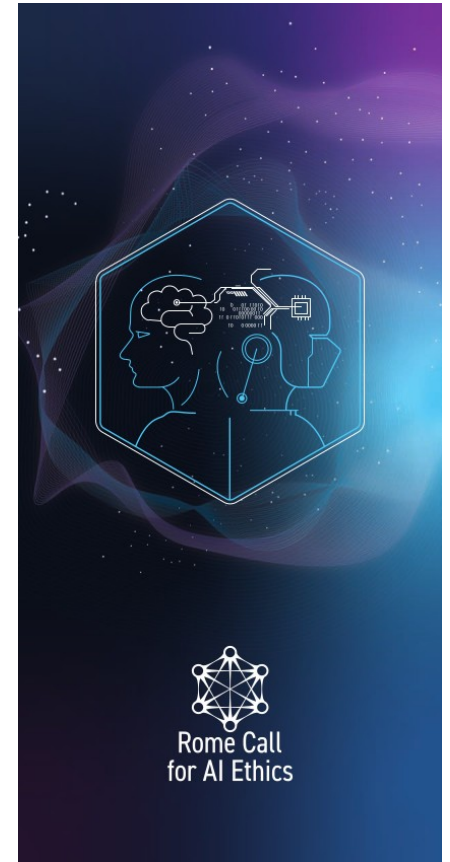
Súčasný rámec riešenia etických problémov ANI

- **Artificial Intelligence Act**

Nariadenie Európskeho parlamentu a Rady, ktorým sa stanovujú harmonizované pravidlá v oblasti umelej inteligencie (**Akt o umelej inteligencii**) a menia niektoré legislatívne akty únie
[<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>]

- **Rome Call for AI Ethics**

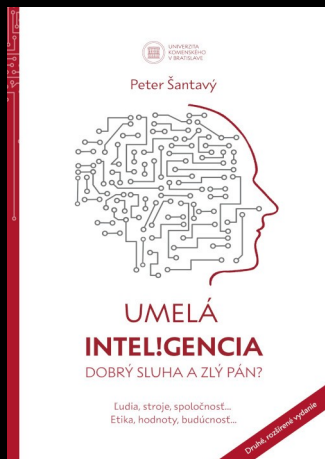
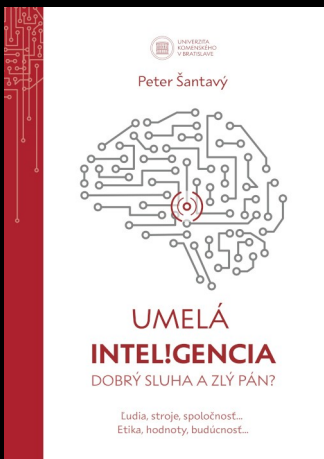
Etické odporúčania vatikánskej konferencie **renAIssance 2020**
[<https://www.romecall.org/>]





Niečo na čítanie...

**Systemy umelej inteligencie skutočnú
inteligenciu nemajú, len ju simulujú**
Rozhovor pre NM



**Nástroje a informácie k AI v rámci
vzdelávacieho procesu na UK
RKCMBF UK - Umelá inteligencia**
Moje knihy z oblasti AI

Moje knihy z oblasti AI



Ďakujem za pozornosť

ThLic. Ing. Peter Šantavý, PhD.
peter@santavy.sk





PRÍLOHA

Základné pojmy v oblasti umelej inteligencie

Definícia umelej inteligencie

Inteligenciu ľudského bytia vnímame ako schopnosť vnímať, chápať a spracovávať informácie, učiť sa, odôvodňovať, plánovať, riešiť komplexné problémy a rozhodovať.

(James Flynn: Súbor zručností, medzi ktoré patrí abstraktné a logické myslenie, ďalej predstavivosť, a teda aj schopnosť uvažovať o hypotetických možnostiach, a tiež aj jazykový cit.)

Umelá inteligencia je o vytváraní systému, ktorý vykazuje také správanie, o ktorom si myslíme, že vyžaduje inteligenciu (definícia podľa AI100).

Na prasknutie naplnený kufor rôznych technológií (Marvin Minsky)...

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 tréningových dát.**

Hovoríme teda o systémoch, ktoré 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.**

Základné delenie – symbolická a subsymbolická 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 – sa pomocou matematickej logiky snaží emulovať procesy myslenia.

Čisto symbolické systémy zlyhávali a zlyhávajú pri riešení problémov, ktoré sa nedajú exaktne popísať a v reálnych prostrediach, ktoré nie je možné deterministicky uchopiť a sú plné nejasných informácií.

Základné delenie – symbolická a 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 – veľmi zjednodušene povedané – sa 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.*

Riziká súčasných systémov ANI

Vybrané dôsledky a riziká

- **extrémne zhromažďovanie dát**

Sofistikované a plošné **zhromažďovanie a spracúvanie dát** a metadát o každom človeku, aktívne, či pasívne prítomnom vo virtuálnom svete.

Aby využívanie systémov umelej inteligencie bolo v jednotlivých oblastiach nasadenia úspešné, potrebné **dáta musia byť neustále zhromažďované z reálneho sveta** a priamo z ľudského prostredia.

Uspešná činnosť väčšiny pokročilých systémov AI je **závislá na extrémnom množstve** relevantných dát.

Riziká súčasných systémov ANI

Vybrané dôsledky a riziká

- **neustály dohľad a sledovanie bez kontroly**

Umožňuje to implementácia algoritmov AI, postavených na extrémnom zhromažďovaní dát a schopnosti tieto dáta relevantne spracúvať.

Vyžaduje to potreba algoritmov AI, ktoré bez neustáleho dohľadu nie sú schopné vytvárať extrémne presné predikcie.

Diskutabilné využitie v algoritmickom riadení štátu, dohľadových systémoch, tajných službách, zdravotníctve, finančnom sektore, sociálnom sektore, sociálnych sieťach,...

Riziká súčasných systémov ANI

Vybrané dôsledky a riziká

- **modely predikcie nášho správania**

Schopnosť systémov AI vytvárať veľmi presné psychologické profily, odhaľovať akékoľvek väzby a mnohé osobné informácie, predpovedať a ovplyvňovať naše konanie.

- **ovplyvňovanie nášho vnímania a konania**

Využívanie aspektov virtuálneho sveta systémami AI na ovplyvňovanie nášho vnímania, rozhodnutí a konania na základe presných profilov a predikcie nášho správania.

Systémy AI sú schopné vytvárať ekosystém závislosti a manipulácií pre dosahovanie stanovených cieľov.

Riziká súčasných systémov ANI

Vybrané dôsledky a riziká

- **technológie AI vedú k závislosti, manipulácii a relativizácii pravdy**
 - závislosť na svojej viditeľnej dokonalosti a neustálom prísune krátkodobých signálov odmeňovania až do tej miery, že to spájame s hodnotami a s pravdou
 - psychologické mantinely v kontexte vnemov virtuálneho sveta
 - algoritmami AI vytvárané a upevňované sociálne bubliny
 - extrémne rozdelenie spoločnosti a narušenie vzťahov
 - nárast zatvrdilej nevedomosti a ignorancie faktov

Pri riešení dôsledkov činnosti systémov AI je treba stále pamätať na to, **aké ciele sú zadané pre činnosť týchto systémov.**