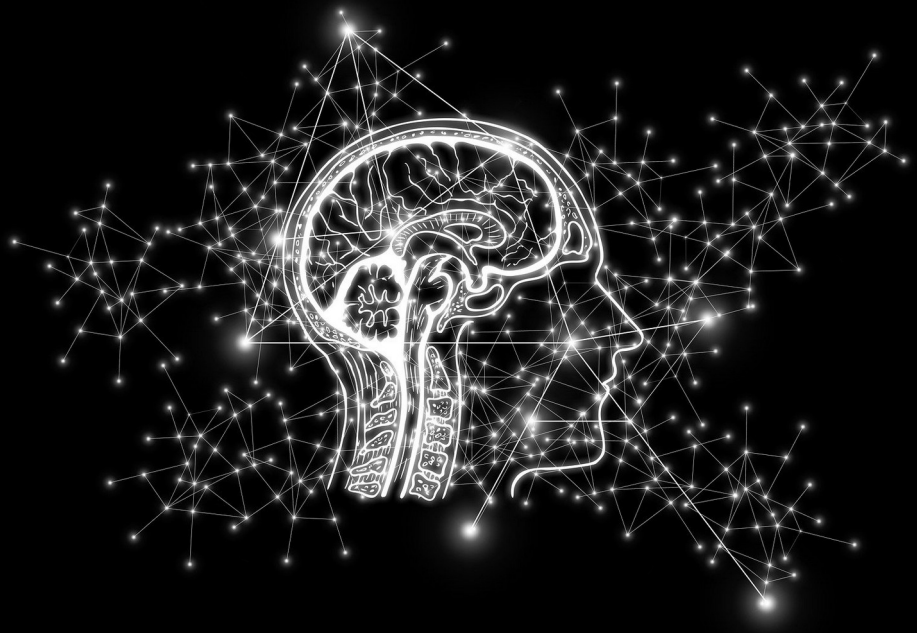


Čo sme si objednali a čo v skutočnosti dostávame...

Peter Šantavý

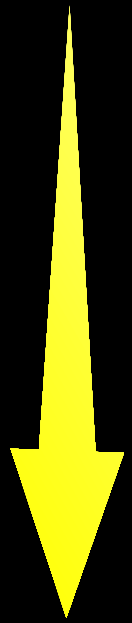
AI – etické aspekty a potencionálne riziká

Uvedenie do problematiky AI

- autonómne a adaptívne systémy
- slabá a úzka umelá inteligencia
- symbolické a subsymbolické systémy
- generatívne systémy AI

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

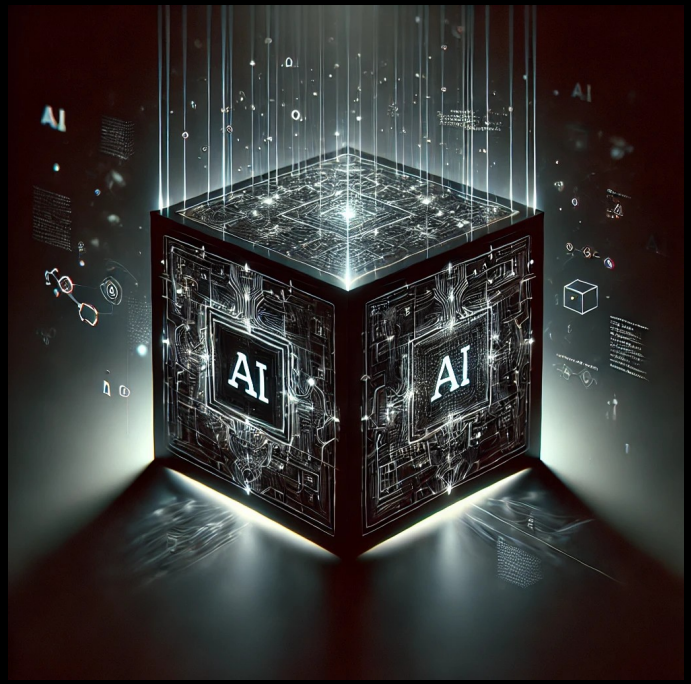
Na čo treba pamätať



- neurónové siete sú ako čierna skrinka
- technologické riziká a limity
- nesprávne používanie a zneužitie systémov AI
- dôsledky pre človeka a spoločnosť

Riziká týkajúce sa primárne subsymbolických systémov, konkrétne neurónových sietí...

Black box



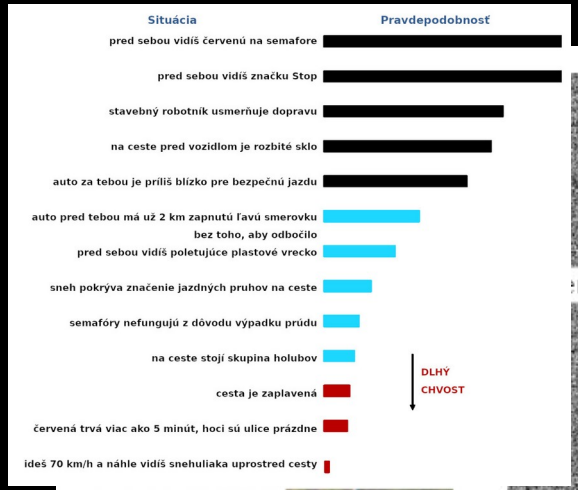
Prakticky nevieme, na základe čoho robia hlboké neurónové siete svoje rozhodnutia.

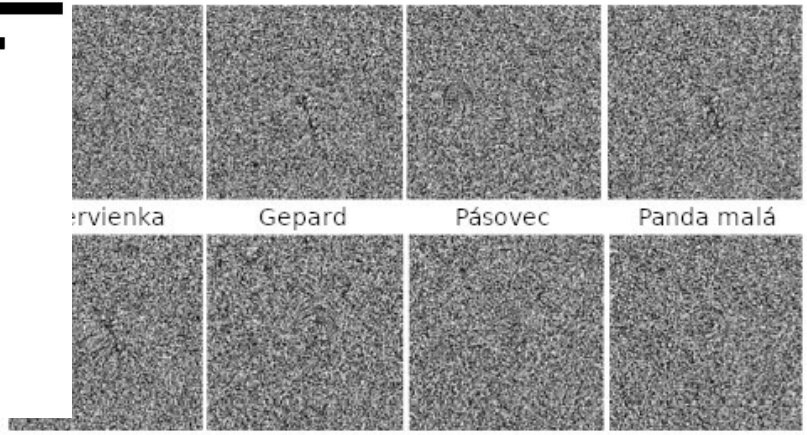
**V zásade nevieme,
čo presne sa neurónová sieť naučila
a ako spoľahlivo to dokáže aplikovať
nielen v bežnej prevádzke, ale osobitne
v hraničných situáciách
za extrémnych podmienok na vstupe
či pri činnosti systému.**

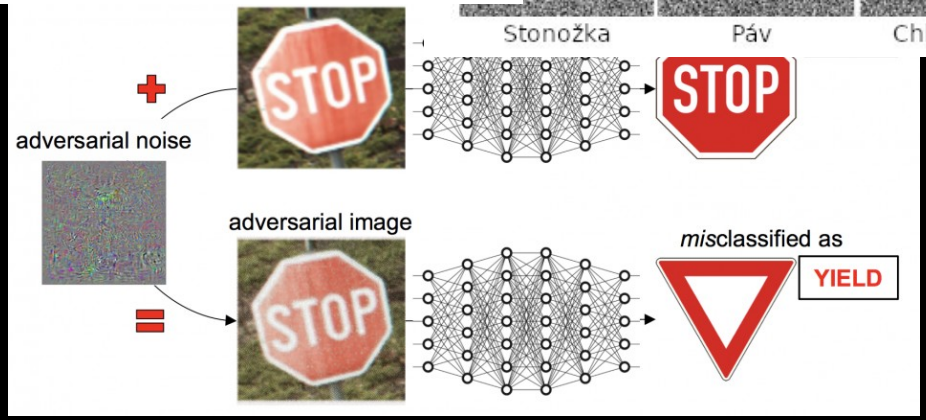
Pracujeme na vysvetliteľných a interpretovateľných systémoch AI ;-)

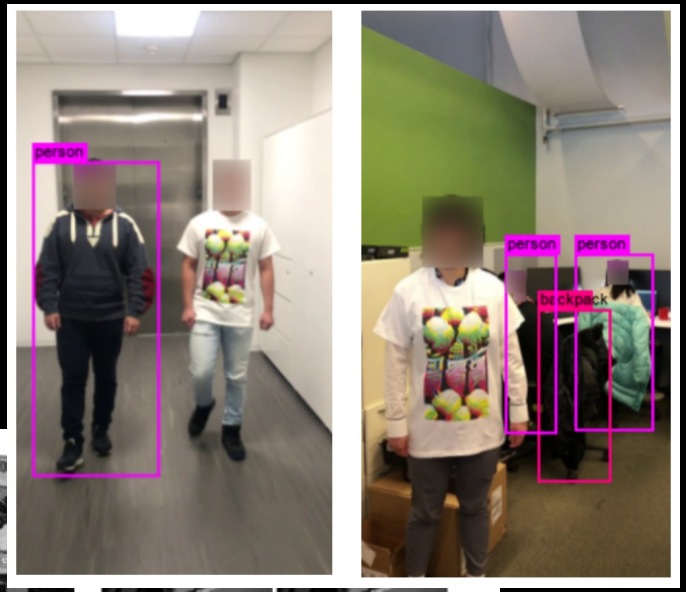
Kredit: GPT-4o/autor

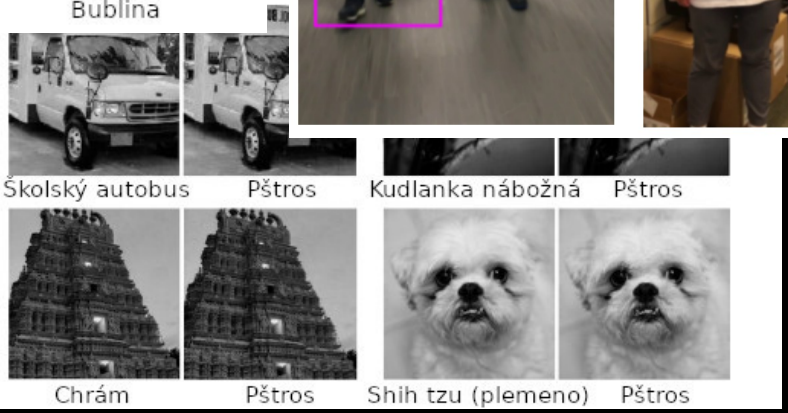
Technologické riziká a limity











1600

Kredit: Melanie Mitchell – upravené autorom, autor, K. Xu et al., Sourav Agarwal

> 1600 rizík podľa AI Risk DB, MIT (11/25)

Ako 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!

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

Veľká štvorka problémov

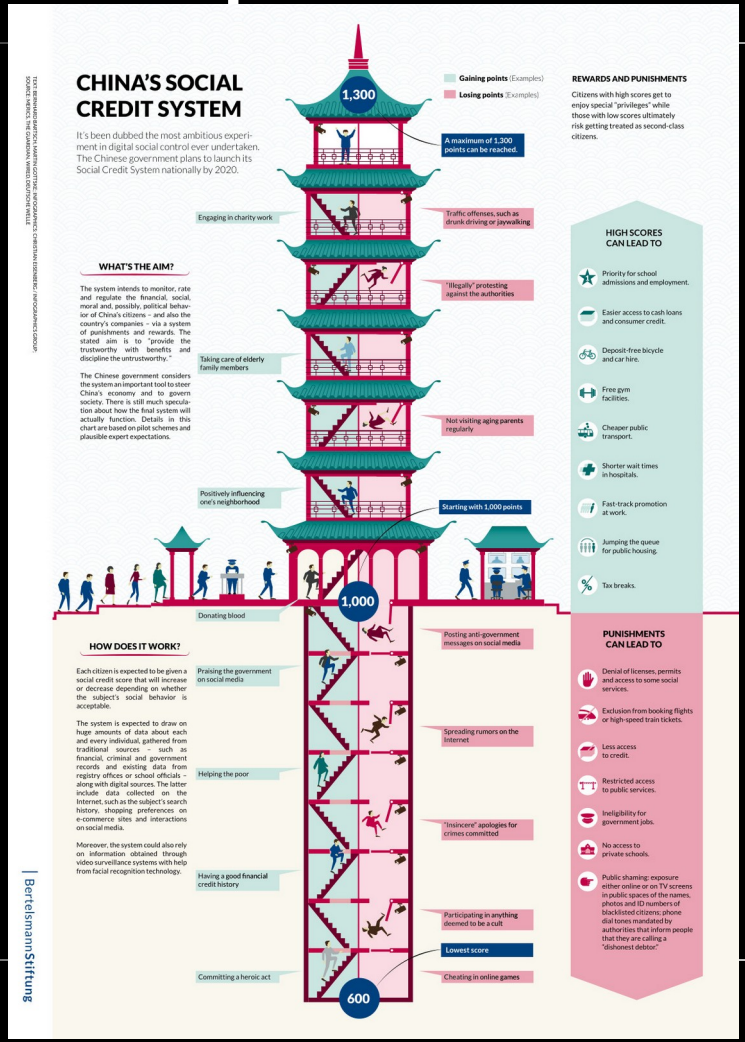
Nesprávne používanie a zneužitie systémov AI + dôsledky pre človeka a spoločnosť.

V dobe AI strácame súkromie

- nutnosť extrémneho zhromažďovania dát
- neustály dohľad, ktorý sa stáva normou
- dohľad a sledovanie bez kontroly

Be good. Future AIs are watching... Andrej Karpathy, 2024

Neustály dohľad a jeho aplikácia...



Kredit: GPT-4o/autor, Bertelsmann Stiftung

Prichádzame o vnútornú slobodu

- psychologické profily a modely predikcie nášho správania
- manipulácia a ovplyvňovanie nášho vnímania a konania
- sociálne bubliny a polarizácia, relativizácia hodnôt a pravdy
- algoritmy AI vedú k závislosti, depresiám a úzkosti

Existujú len dva druhy priemyslu, ktoré svojim zákazníkom hovoria používatelia. Nelegálne drogy a softvér.
Prof. Edward R. Tufte

Prichádzame o vnútornú slobodu

Kredit: The Social Dilemma

Strácame kognitívne schopnosti

- riziko digitálnej demencie

Digitálna demencia

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

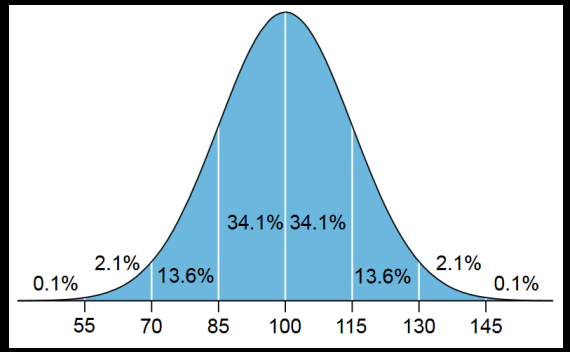
Môže prichádzať k degradácii intelektuálnych schopností a k dopamínovej závislosti na základe ponorenia sa do virtuálneho sveta, v ktorom systémy AI v čoraz väčšej miere suplujú kognitívne činnosti človeka, a to spôsobom, na ktorý nie sme evolučne vôbec pripravení. Mozog sa mení...

Kredit: GPT-4o/autor

Strácame kognitívne schopnosti

- riziko digitálnej demencie
- digitálne rozdelenie

Digitálne rozdelenie ako dôsledok AI



INTELIGENČNÉ NOŽNICE, KTORÉ SA OTVÁRAJÚ

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

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

Pozn.: digital divide má nielen v oblasti IKT, ale i v AI aj ďalšie aspekty...

Kredit: GPT-4o/autor, wikimedia.org

Strácame kognitívne schopnosti

- riziko digitálnej demencie
- digitálne rozdelenie
- mozgová hniloba (brainrot)
- neschopnosť robiť vlastné zodpovedné rozhodnutia
- strata kritického myslenia a nekritické preberanie výsledkov AI

Nič veľkého nevstúpi do života smrteľníkov bez prekliatia. Sofokles

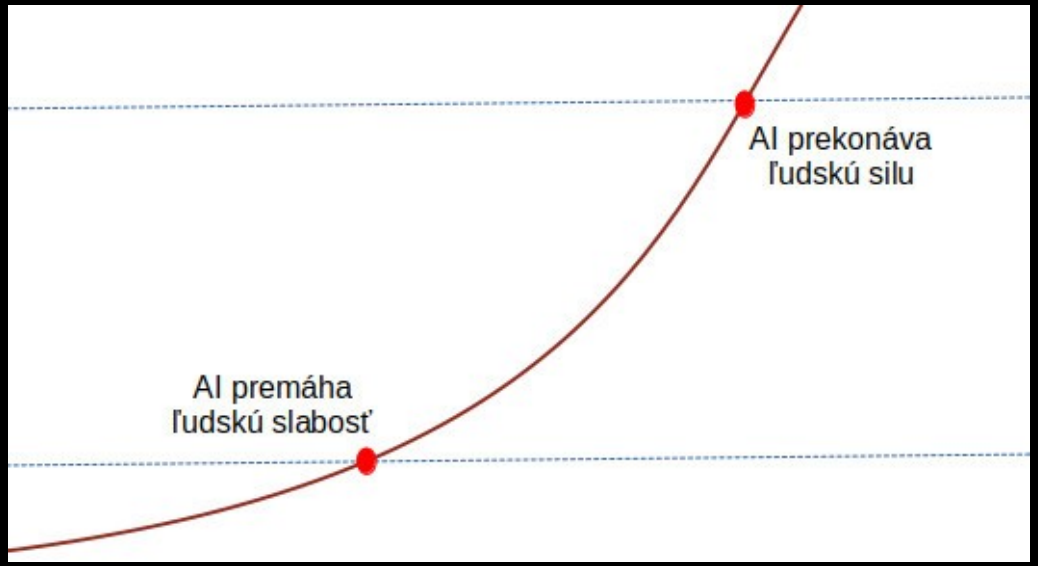
Realita vs. virtuálne svety

Dopad na reálny život a skutočné vzťahy

- strata zmyslu pre realitu
- deepfake
- virtuálne svety a pokrivené vzťahy
- narušenie psychického vývoja
- deformácia virtuálneho a informačného priestoru

Rozhodli se svěřit naše bezpečí něčemu, co není schopné věrnosti, morálky ani moudrosti. J. Campbell

Čoho sa reálne musíme obávať



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.

Kredit: autor

Zamerané na dôveryhodné systémy AI

UMELÁ INTELIGENCIA ZAMERANÁ NA DOBRO ČLOVEKA
(human-centered and beneficial artificial intelligence)

= dôveryhodné systémy AI, ktoré musia byť:

- funkčné a užitočné
- legálne
- etické
- odolné, resp. robustné

Dôveryhodné systémy AI sú rozoberané v 4. kapitole odkazovanej knihy (QR kód)...

Čo treba robiť na spoločenskej úrovni

- **mať jasne definované etické princípy**
  – ich chápanie, hodnotový systém, spoločenský konsenzus
- **primerané a široko akceptované regulácie**
  – jasné mantinely + dostatočná voľnosť pre kreativitu a biznis
  – primeraný regulačný rámec pre celý životný cyklus systémov AI
- **celoplošná osveta a vzdelávanie**
  – osobitne pre mladých a rodičov, primerane pre všetky vekové kategórie (dôvod: veľká 4)
  – adekvátne jednotlivým profesiám a spôsobu využívania systémov AI, inak kompetenčná kríza s pokračujúcim nástupom AI porastie geometrickým radom

Je na nás, či umelá inteligencia bude dobrý sluha alebo zlý pán...

Poznámky k vzdelávaniu a využívaniu nástrojov AI

Základný postoj – pozitívny!

Technológie AI dokážu byť **pri správnom nasadení** neskutočne nápomocné v rámci vzdelávacej, vedecko-výskumnej, formačnej a administratívnej činnosti.

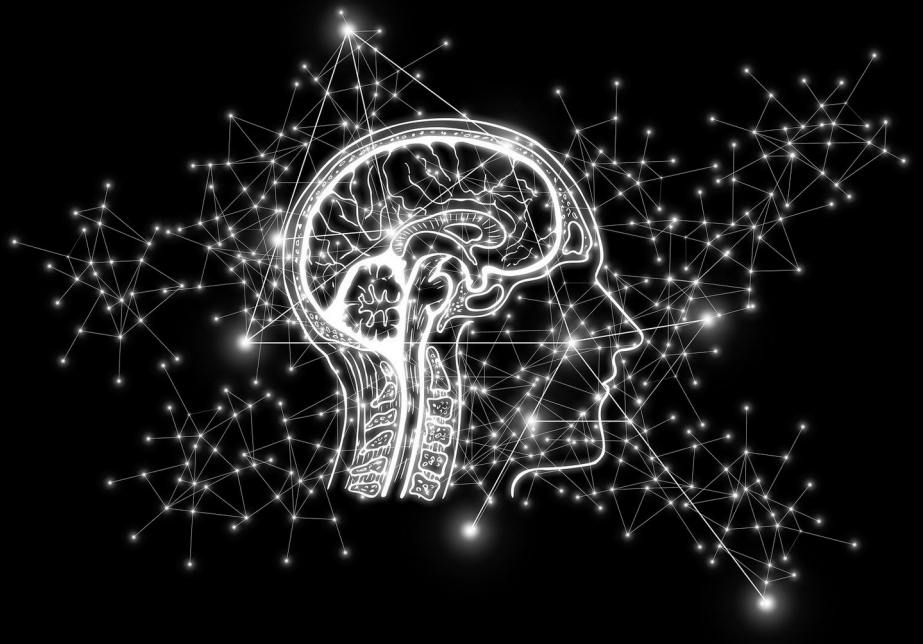
Dostatočná báza znalostí

Pre osožné využívanie nástrojov AI sa javí byť **nutné budovať primeranú bázu vedomostí a znalostí, bez ktorej nebudeme schopní systémy AI správne používať**.

Rásť so systémami AI

Stále komplexnejšie technológie evokujú oveľa vyššie požiadavky v oblasti vzdelávania. **Treba adresovať realitu, v ktorej je AI v živote mladých ako *neriadená strela***.

Je na nás, či umelá inteligencia bude dobrý sluha alebo zlý pán...



Ďakujem za pozornosť

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