

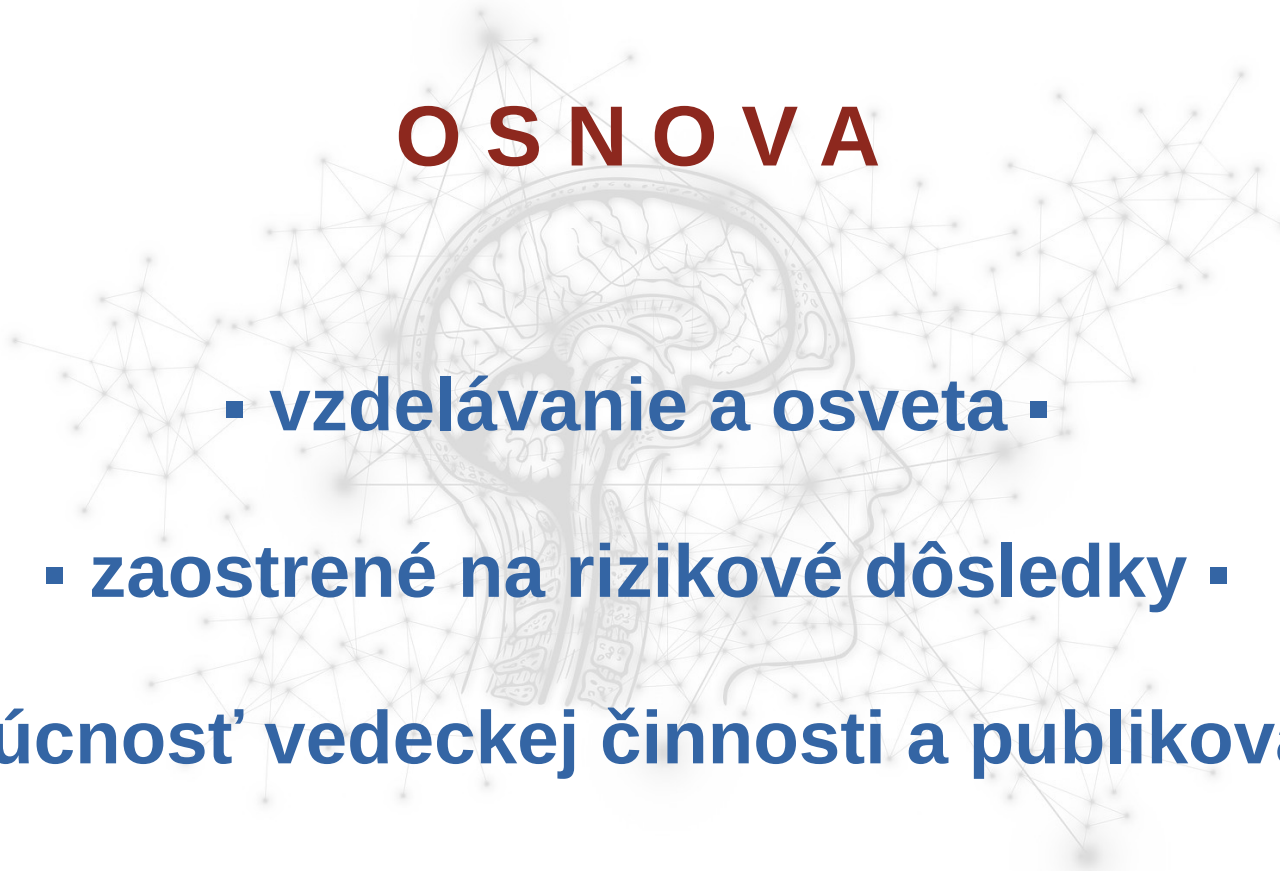


SYSTÉMY AI VO VZDELÁVANÍ: ETICKÉ ASPEKTY A POTENCIÁLNE RIZIKÁ

Peter Šantavý

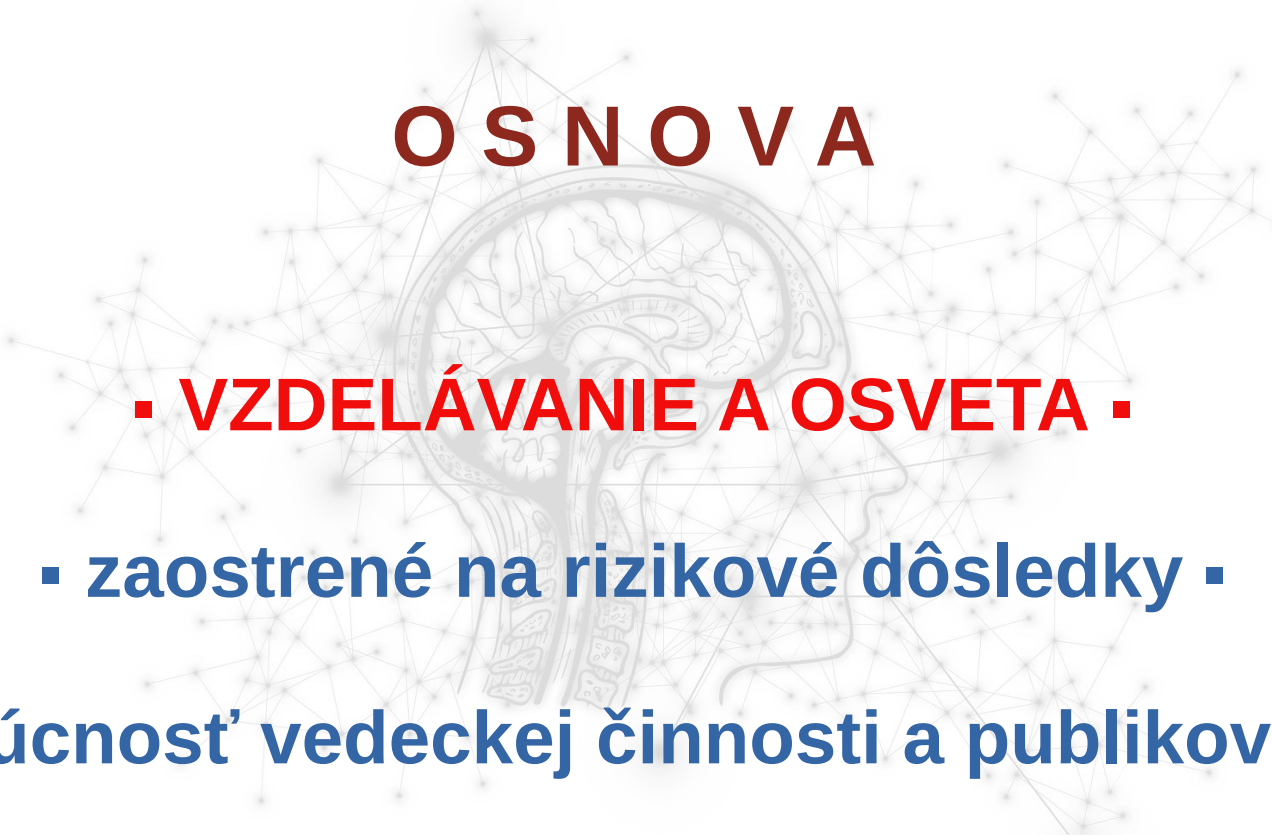


OSNOVA



- **vzdelávanie a osveta ▪**
- **zaostrené na rizikové dôsledky ▪**
- **budúcnosť vedeckej činnosti a publikovania ▪**

OSNOVA



▪ VZDELÁVANIE A OSVETA ▪

▪ zaostrené na rizikové dôsledky ▪

▪ budúcnosť vedeckej činnosti a publikovania ▪

OSNOVA

▪ VZDELÁVANIE

▪ ETA

Každá dostatočne pokročilá technológia
je na nerozoznanie od mágie :-)

(Arthur C. Clarke)

▪ Prírodné dôsledky

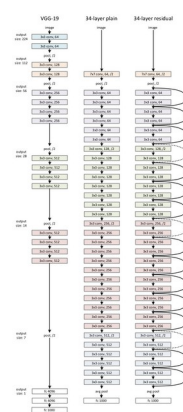
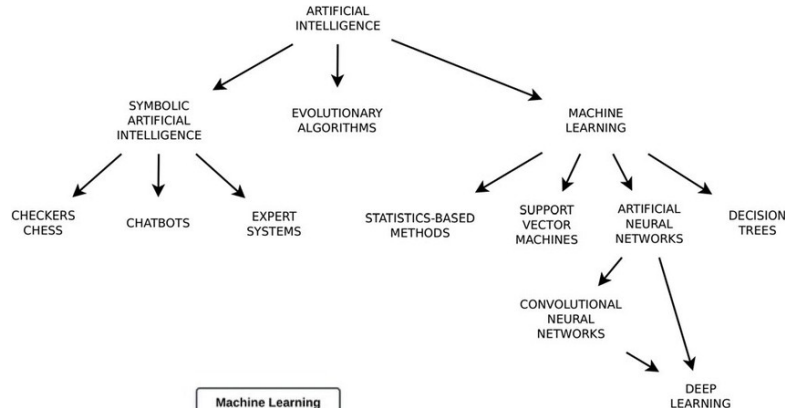
▪ Podpora vedeckej činnosti a publikovania

Primerané poznanie fungovania systémov AI

Dôležité pre:

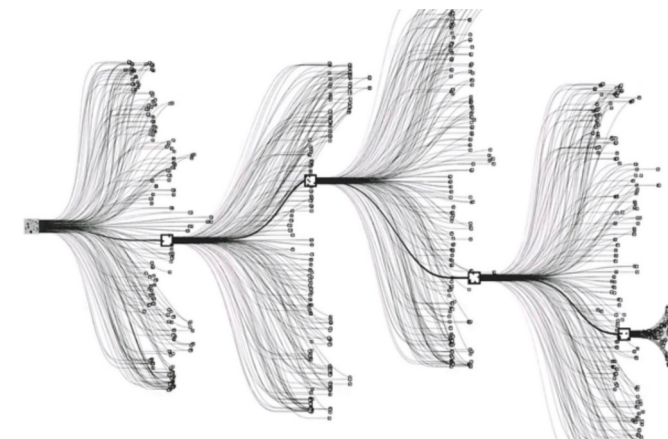
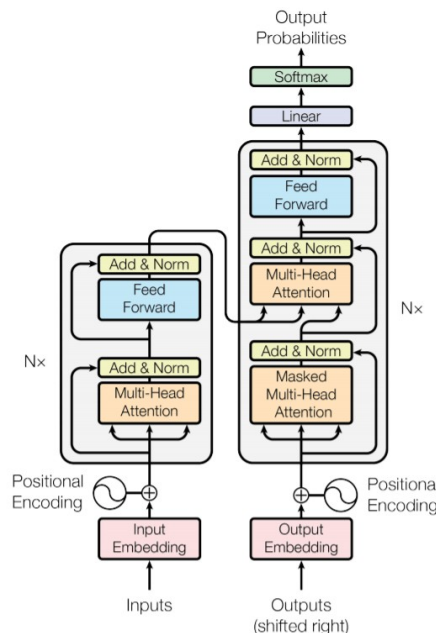
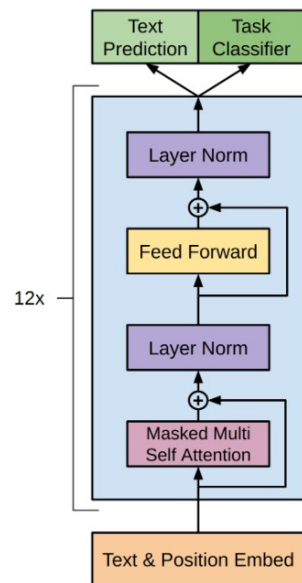
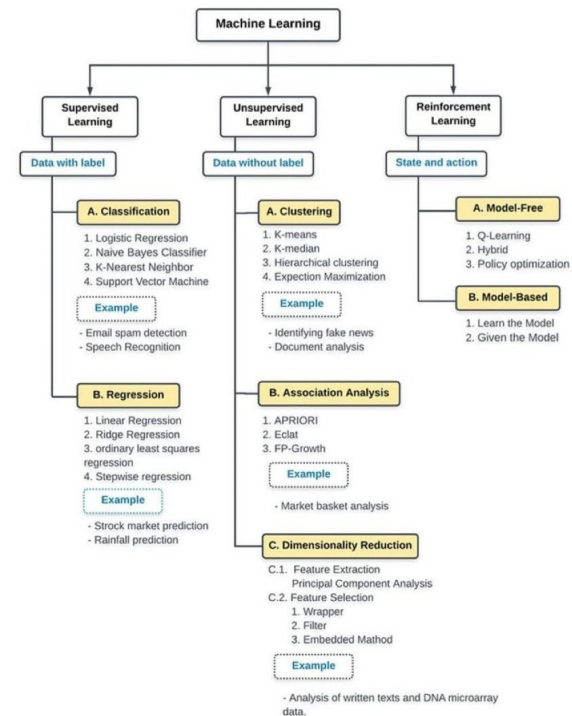
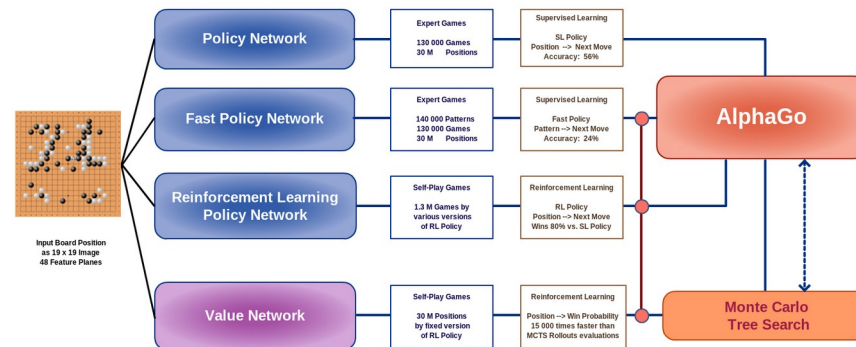
- » efektívne využívanie
- » vyvarovanie sa chybám
- » nemat' nereálne očakávania

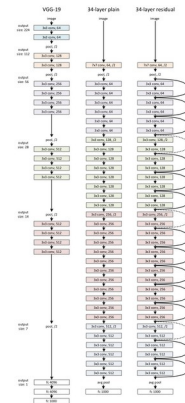
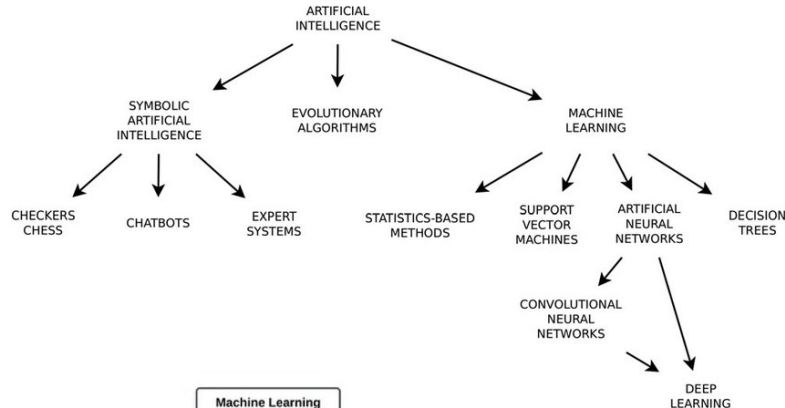
Ide o komplexné systémy, ktoré treba vedieť správne využívať...



AlphaGo Overview

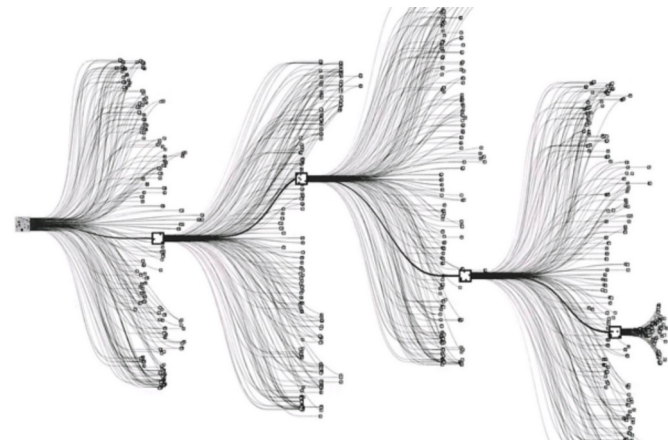
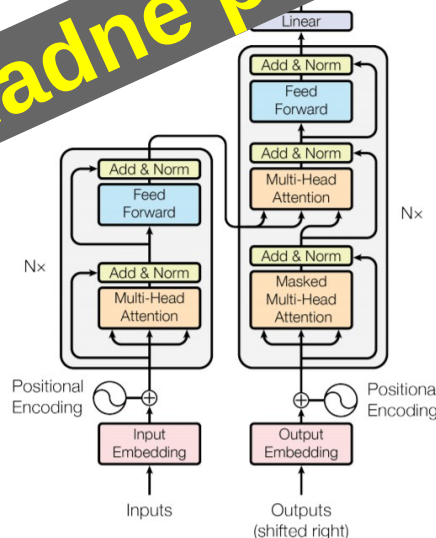
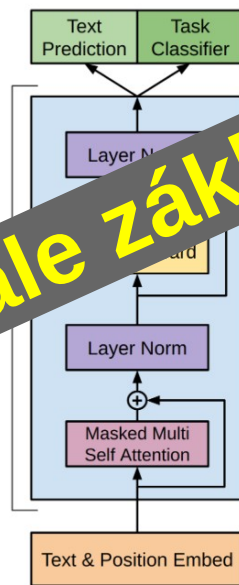
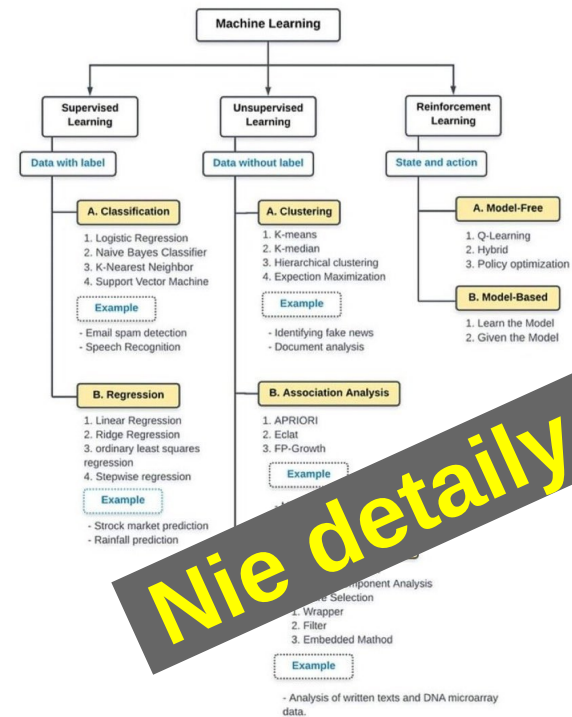
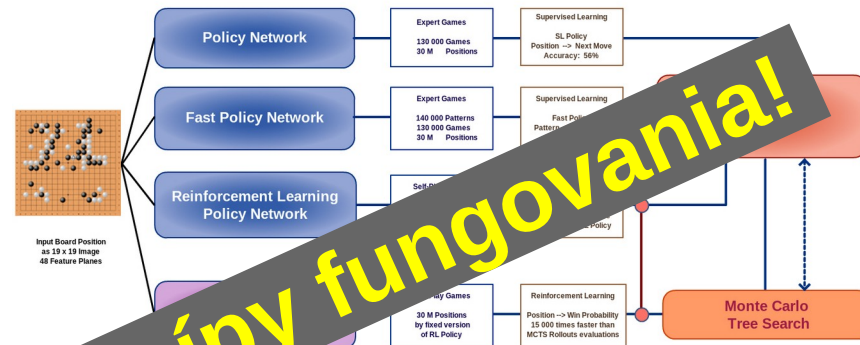
based on: Silver, D. et al. Nature Vol 529, 2016
copyright: Bob van den Hoek, 2018





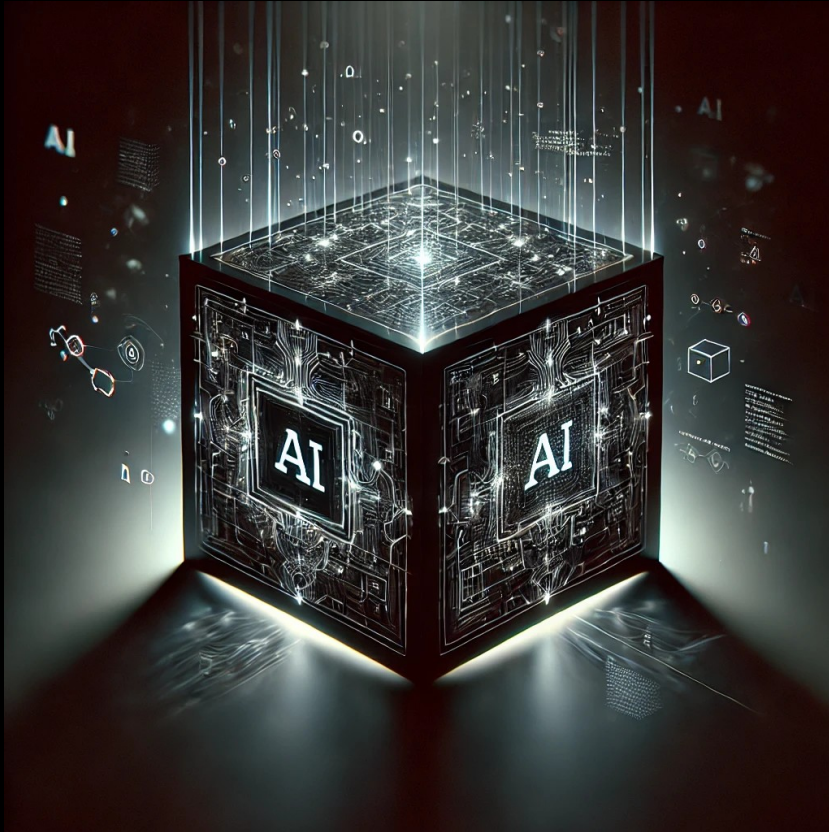
AlphaGo Overview

based on: Silver, D. et al. Nature Vol 529, 2016
copyright: Bob van den Hoek, 2018



Nie detaily, ale základné princípy fungovania!

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



Prakticky nevieme, ako to robia hlboké učenie. Pracujeme na vysvetliteľných a interpretovateľných systémoch AI ;-)

Prakticky nevieme, ako to robia hlboké učenie. Zásade nevieme, či 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.

S čím sa potýkame? (problémy s AI)

- » technologické riziká a limity
- » nesprávne používanie a zneužitie systémov AI
- » dôsledky pre človeka a spoločnosť
sociologické, psychologické, civilizačné,...

S čím sa potýkame? (problémy s AI)

» technologické riziká

» nesprávne používanie

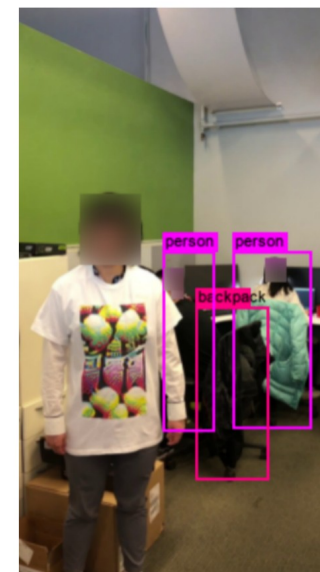
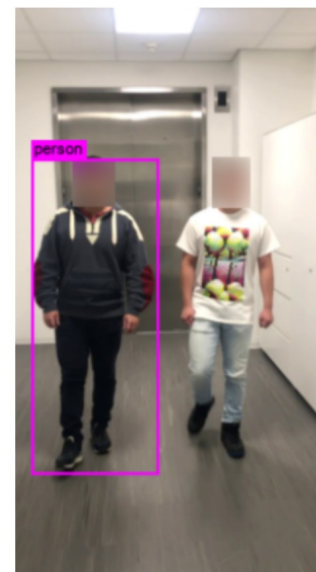
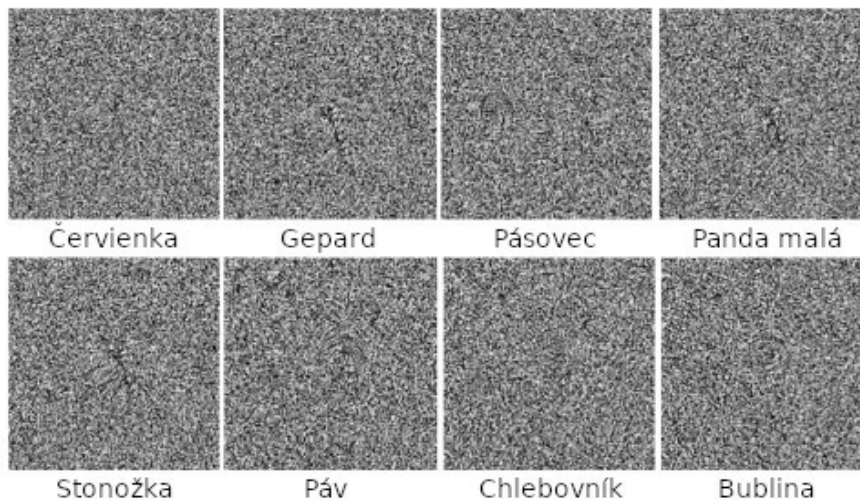
» dôsledky pre prácu a spoločnosť

» etické, psychologické, civilizačné,...

AI systémy myslenie len napodobňujú!

AI systémy "rozmyšľajú" diametrálne odlišne, inak než ľudia!

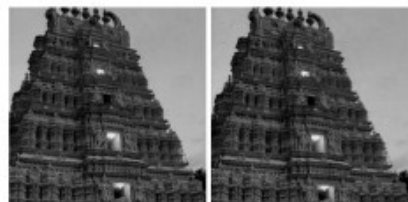
systémov AI



Školský autobus Pštros



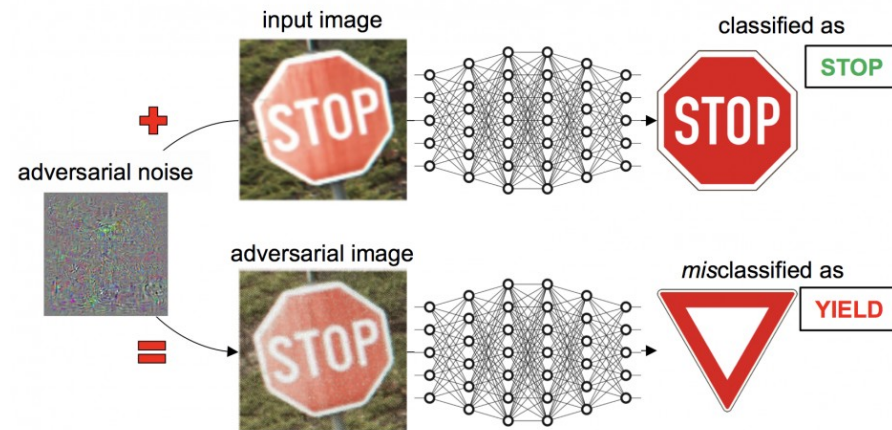
Kudlanka nábožná Pštros

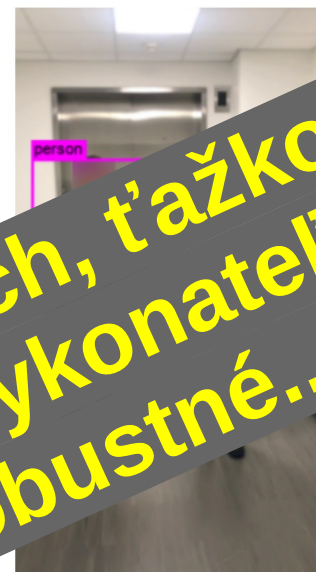


Chrám Pštros

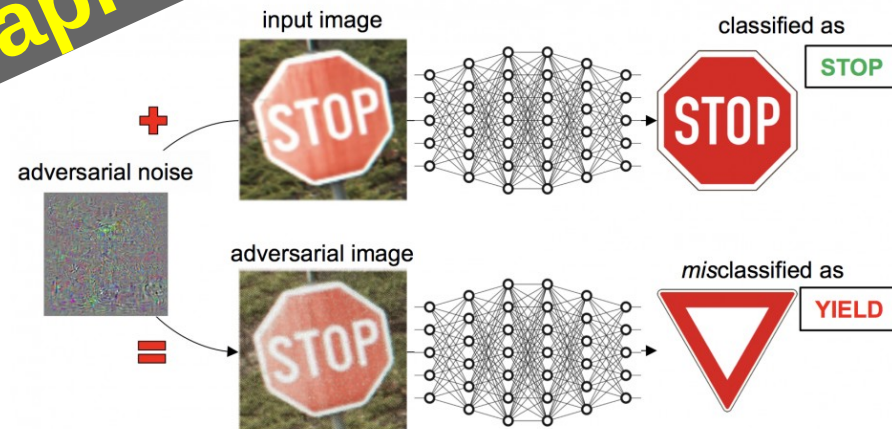


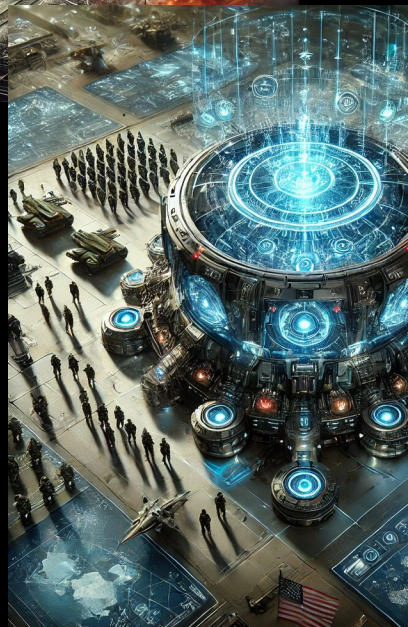
Shih tzu (plemeno) Pštros





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





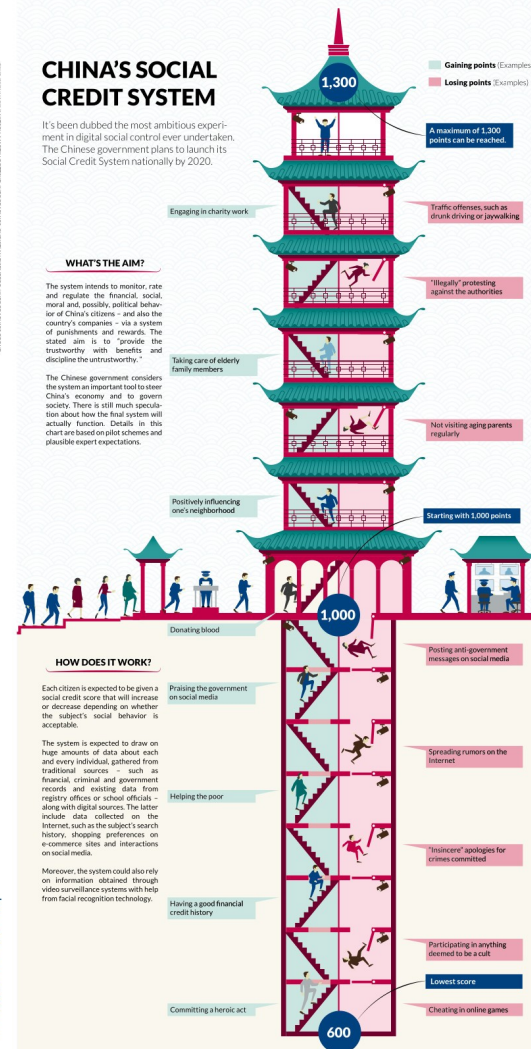
CHINA'S SOCIAL CREDIT SYSTEM

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 also the country's companies – via a system of punishments and rewards. The stated aim is to "provide the trustworthy with benefits and discipline the untrustworthy."

The Chinese government considers the system an important tool to steer China's economy and to govern society. There is still much speculation about how the final system will actually function. Details in this chart are based on pilot schemes and plausible expert expectations.



BertelsmannStiftung

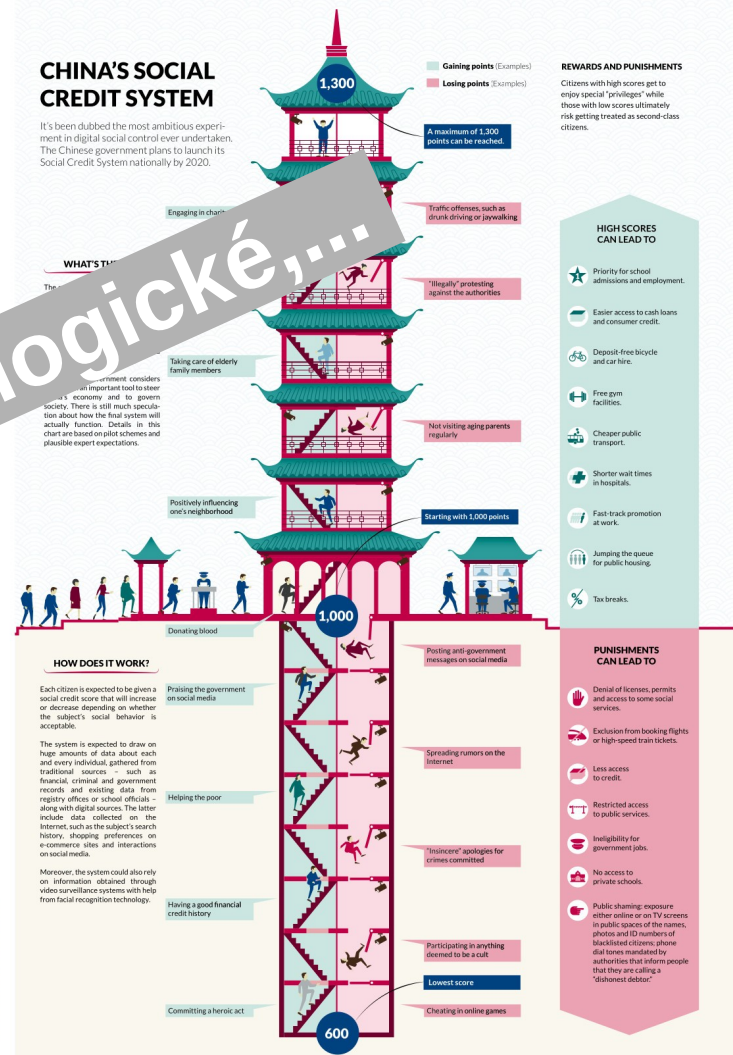


Mocenské, vojenské, ideologické, zneužitie AI!

BertelsmannStiftung

CHINA'S SOCIAL CREDIT SYSTEM

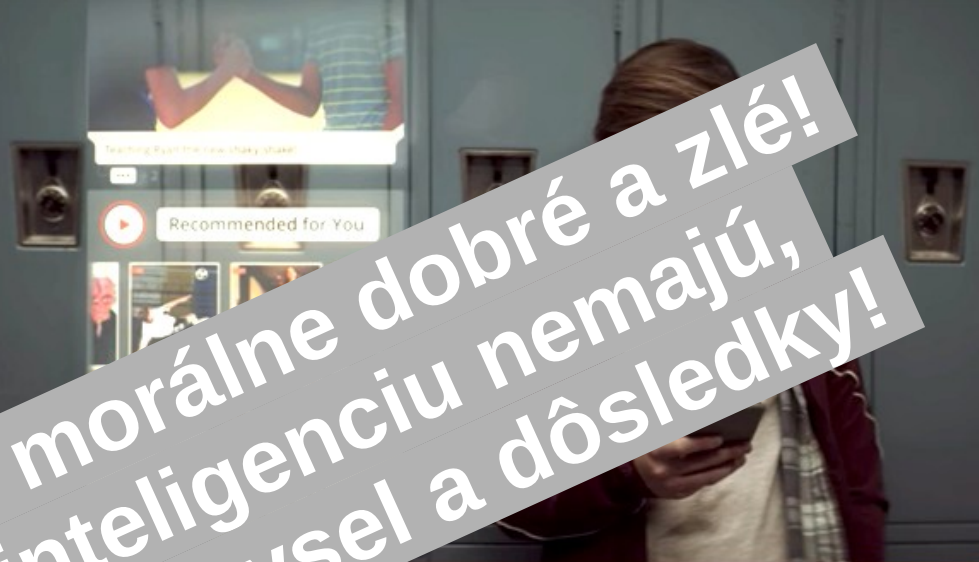
It's been dubbed the most ambitious experiment in digital social control ever undertaken. The Chinese government plans to launch its Social Credit System nationally by 2020.



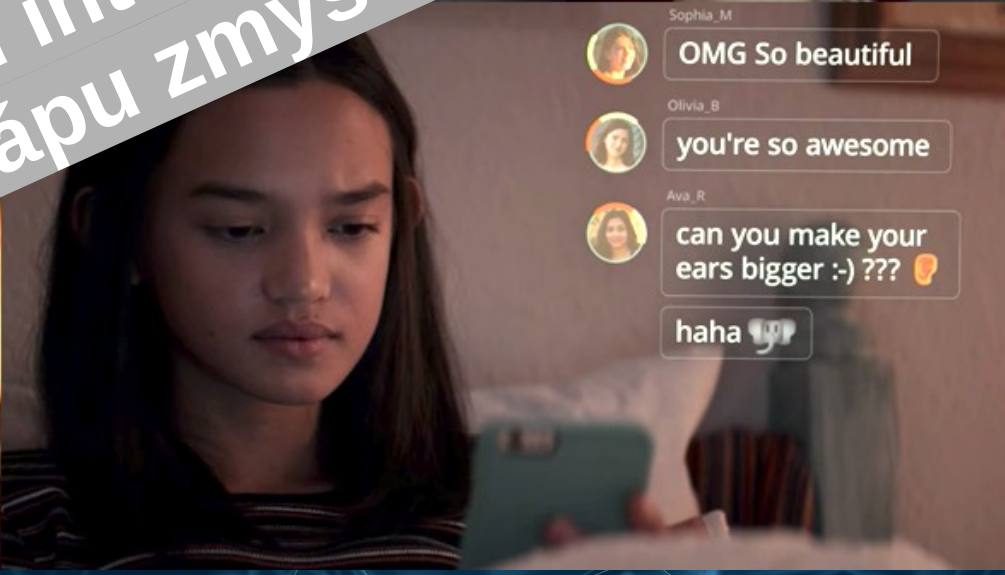
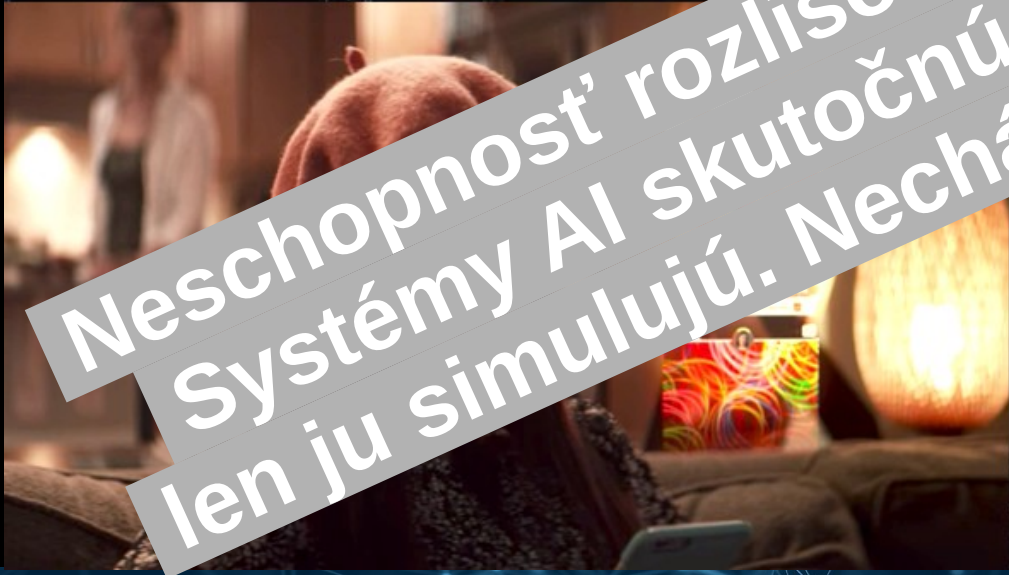


N

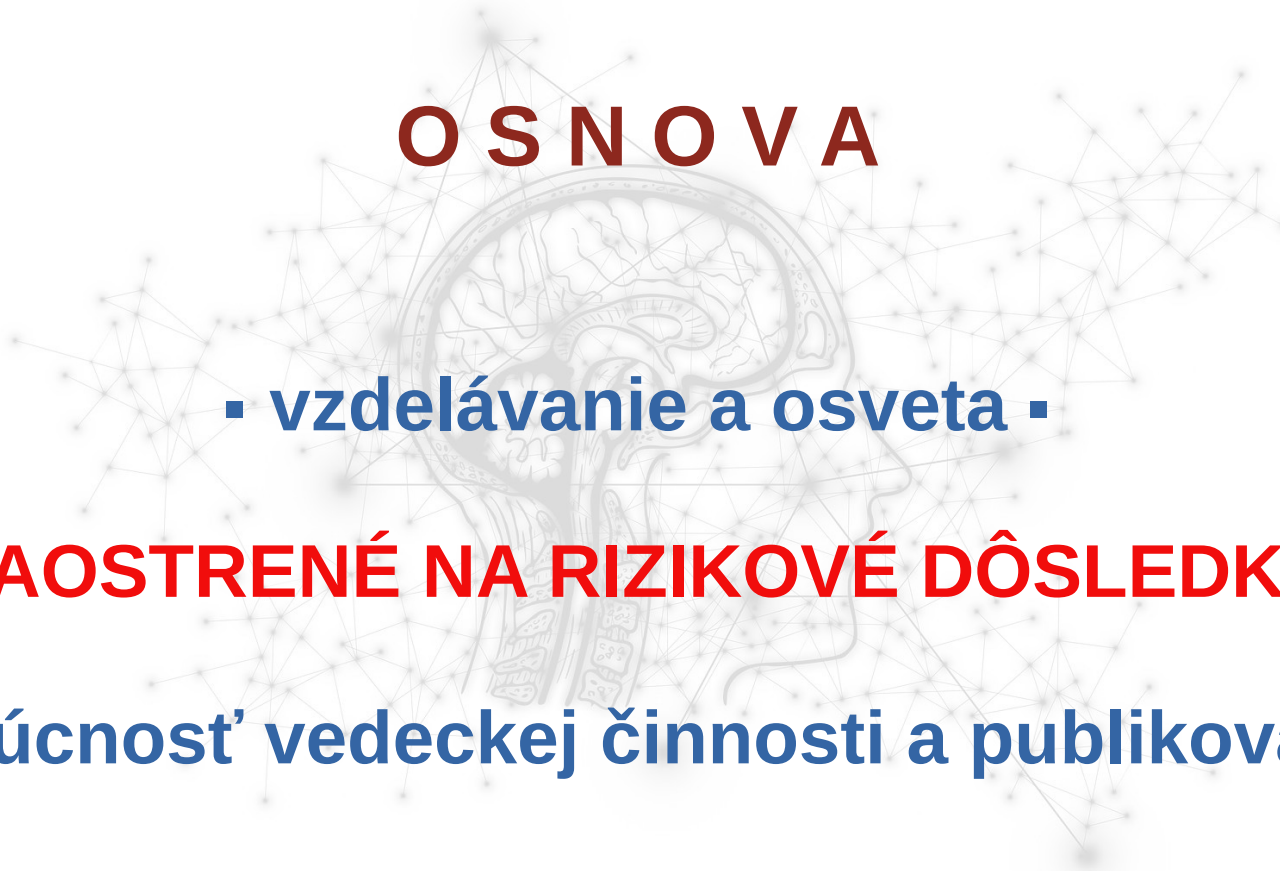
/the social dilemma



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



OSNOVA



- vzdelávanie a osveta ▪
- **ZAOSTRENÉ NA RIZIKOVÉ DÔSLEDKY** ▪
- budúcnosť vedeckej činnosti a publikovania ▪

O S N O V A

▪ vzdelávanie študentov

▪ ZAO

bez prekliatia.
(Sofokles)

PRÁVNÉ DÔSLEDKY

▪ "neúčnosť" vedeckej činnosti a publikovania

Vybrané riziká generatívnych systémov AI

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

- únik dôverných dát, problémy s ochranou osobných údajov a autorskými právami

Vybrané riziká generatívnych systémov

Halucinovanie

- vymýšľanie si odpovedí, ktoré sú nepravdivé a správne

Predsudky

- únik dôverných dát, problémy s ochranou osobných údajov a autorskými právami

Generatívne nástroje AI nemôžeme vnímať ako faktografické, spoľahlivé a etické zdroje! Potrebujú človeka, ktorý ich vie správne používať...

Etické problémy (nielen) 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 **suplujú** kognitívne činnosti človeka, a to spôsobom, na ktorý nie sme evolučne vôbec pripravení.

Digitálne rozdelenie – riziko digitálneho rozdelenia (digital divide) i v oblasti umelej inteligencie (tí, ktorí nepoužívajú / sú manipulovaní / využívajú AI,...).



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 DEMENCE

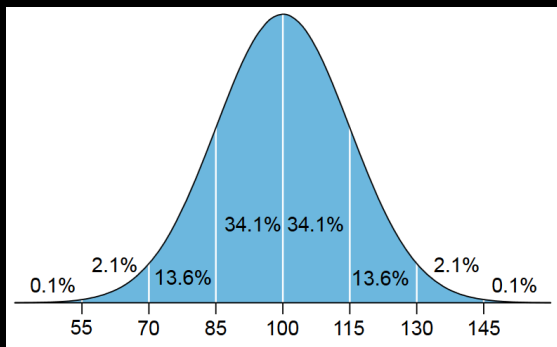
Dlhodobé a nesprávne využívanie osobitne generatívnych AI a sociálnych sietí, prináša **intenzívne kognitívne vyčerpanie a kognitívne vyčerpávanie.**

Vďaka systémom AI môžeme postupne strácať časť našich schopností a psychickej odolnosti! To vedie k degradácii intelektuálnych schopností a k dopamínovej závislosti na základe toho, že sa dostávame 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...



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.

Etické problémy (nielen) generatívnych systémov

Mozgová hniloba (brainrot) – postupná strata schopnosti sústrediť sa a vnímať hlbší kontext. Vytvárajú sa dopamínové dráhy závislosti a návyk na také hodnoty adrenalínu a dopamínu, na ktoré nie sme evolučne pripravení.

Deepfake – vytváranie dôveryhodných virtuálnych identít ľudí a podvrhnutie ich falošnej komunikácie a konania (klamanie, nekalá konkurencia, pornografia, kriminalita, propaganda,...). **Stieranie rozdielov medzi pravdou a lžou.**

Strata zmyslu pre realitu – únik do virtuálneho sveta a do vzťahov s AI. Neschopnosť fungovať v reálnom svete a rozvíjať skutočné vzťahy, hodnotové posuny, závislosti, zlyhávajúce v reálnom živote (kritické myslenie, sexualita,...).

Etické problémy (nielen) generatívnych AI

Mozgová hniloba (brainrot) – postupná strata schopnosti vnímať hlbší kontext. Vytvárajú sa na také hodnoty adrenalínu a dopamínu, ktoré sa nedajú nájsť v reálnom živote.

Deepfake

falošné video, ktoré môže viesť k ľudí a podvrhnutie ich dôvery. Je to konkurencia, pornografia, dezinformácie, rozdielov medzi pravdou a lžou.

Stíhajúci únik – únik do virtuálneho sveta a do vzťahov s AI.

Nesú schopní fungovať v reálnom svete a rozvíjať skutočné vzťahy, hodnotové postoje, závislosti, zlyhávajú v reálnom živote (kritické myslenie, sexualita,...).



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



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 skutočné hodnoty!



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

Ďalšie etické problémy generatívnych systémov

Neschopnosť robiť vlastné zodpovedné rozhodnutia – delegujeme to na stroje.

Narušenie psychického vývoja priskorým, resp. nezrelým používaním digitálnych a AI technológií (od synaptického plasticity až po napr. tzv. *virtuálny autizmus*...).

Riziko erózie identity – preceňovanie AI a "splošťovanie" ľudskej komplexnej inteligencie, racionality a celkovej ľudskosti na úroveň technickej inteligencie.

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

Schopnosť a možnosť manipulácie vlastníckmi systémov AI.

Propaganda a manipulácia, nástroj moci a ideológie, algoritmické modelovanie histórie,...

Ďalšie etické problémy generatívnych systémov

Neschopnosť robiť vlastné zodpovedné rozhodnutia – delirium

Narušenie psychického vývoja priskorým, resp. preexponovaním detských mozgov digitálnych a AI technológií (od synaptickej plasticity až po autizmus...).

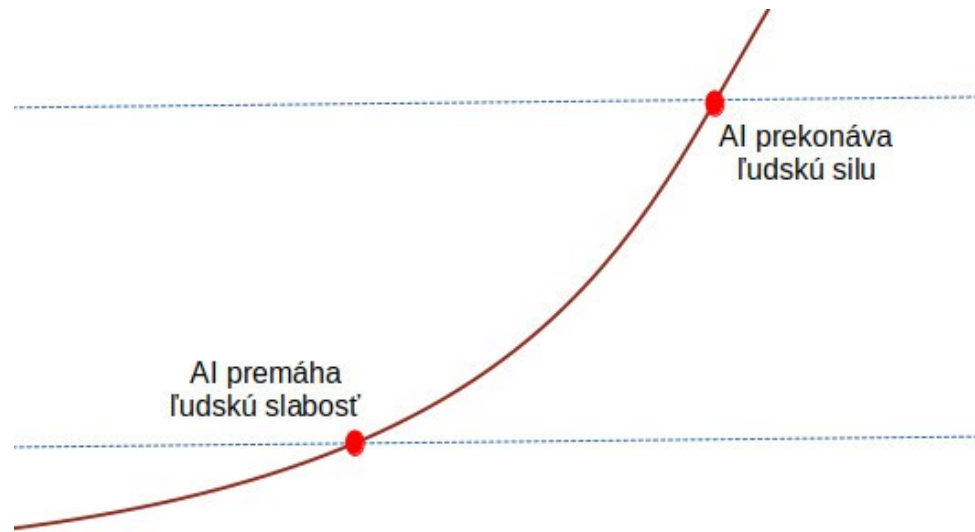
Riziko erózie identity – preexponovanie "človeka" ľudskej komplexnej inteligencie, racionality a morálke na úroveň technickej inteligencie.

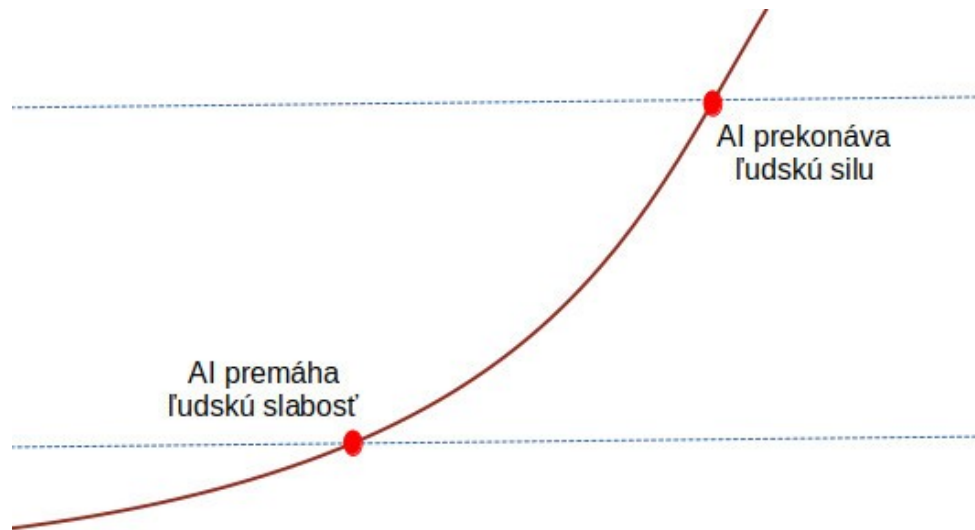
Strata kritického myslenia – nekritické preberanie výsledkov systémov AI spojené s dôverou, vytváraním sociálnych bublín a pod.

Možnosť manipulácie vlastníckmi systémov AI.

Propaganda a manipulácia, nástroj moci a ideológie, algoritmické modelovanie histórie,...

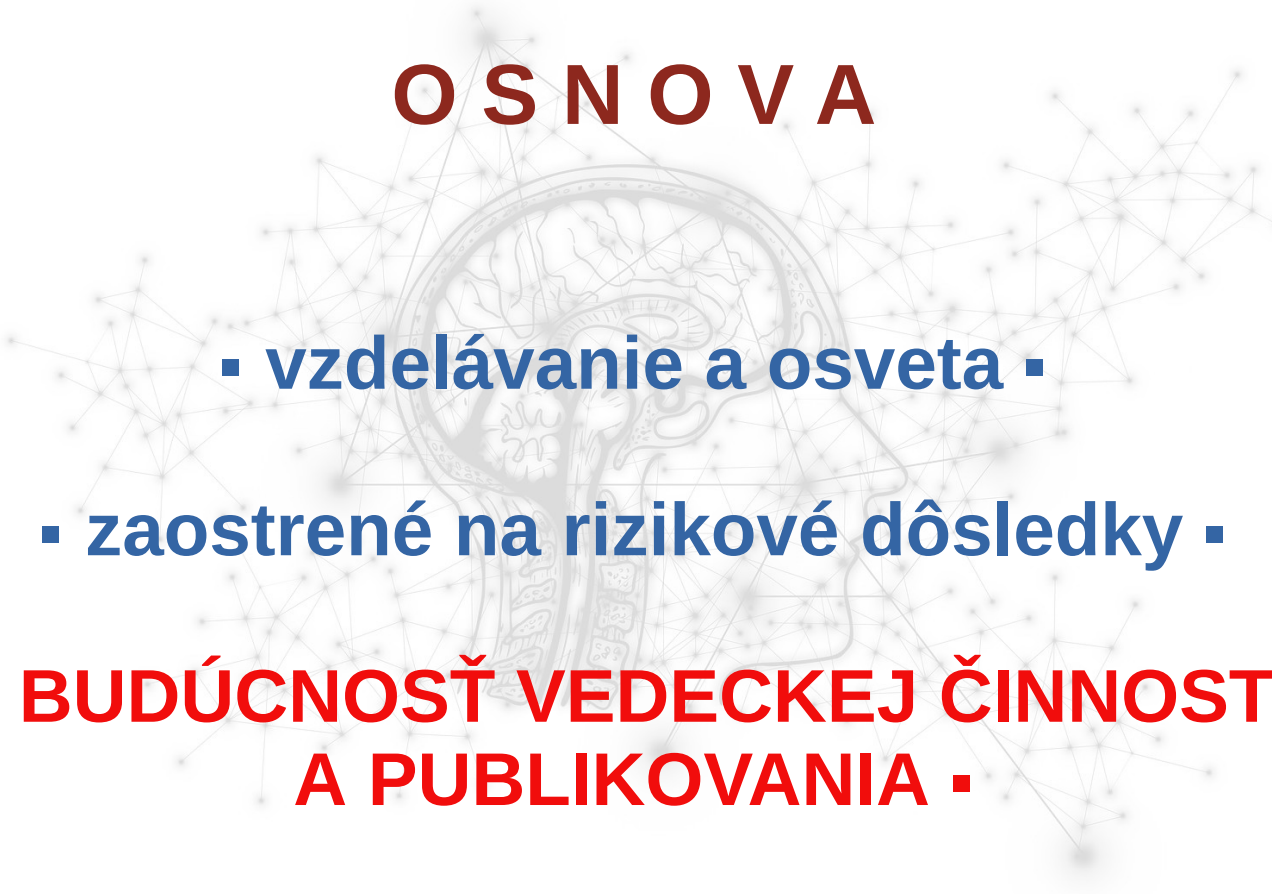
Vďaka systému AI môžeme postupne strácať časť našich schopností a psychickej odolnosti.





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.

OSNOVA



- vzdelávanie a osveta ▪
- zaostrené na rizikové dôsledky ▪
- **BUDÚCNOSŤ VEDECKEJ ČINNOSTI
A PUBLIKOVANIA ▪**

Problémy vo vedeckej činnosti a publikovaní

Niektoré problémy

- požiadavky na výkon v publikovaní (*publish or perish*)
- problematické metriky (IF, H-index,... → môžu byť princípy DORA/CoARA riešením?)
- recenzné konanie (peer review, fake peer review → open peer review riešením?)
- prepojenie vedy, politiky a peňazí (financovanie vs. priority a sloboda výskumu)
- replikačná kríza (reprodukovateľnosť a replikovateľnosť výsledkov; Open Science?)
- kríza integrity, narastajúci počet retrakcií kvôli masívnym publikačným podvodom...

Dopad AI

- fenomén AI tieto problémy dokáže prehĺbovať i akcelerovať a komplikuje snahu o zaistenie integrity vedeckého záznamu a publikovania
- na druhej strane AI môže byť aj pomôckou pre riešenie niektorých problémov

Poznámky k univerzitnému využívaniu nástrojov AI

Základný postoj – pozitívny!

Technológie AI dokážu byť pri správnom nasadení neskutočne nápomocné v rámci vzdelávacej, vedecko-výskumnej 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 vedecké prostredie evokuje oveľa vyššie požiadavky na kritické myslenie tvorcov i prijímateľov vedeckej práce.

AI vo vzdelávaní

AI ako neriadená strela

- mnohoraké využívanie (i zneužívanie) nástrojov AI študentami
- otvorený príbeh s rôznymi možnosťami využívania a ovocím, čo to prinesie

Príklad pravidiel využívania nástrojov AI na UK (smernica + komisia)

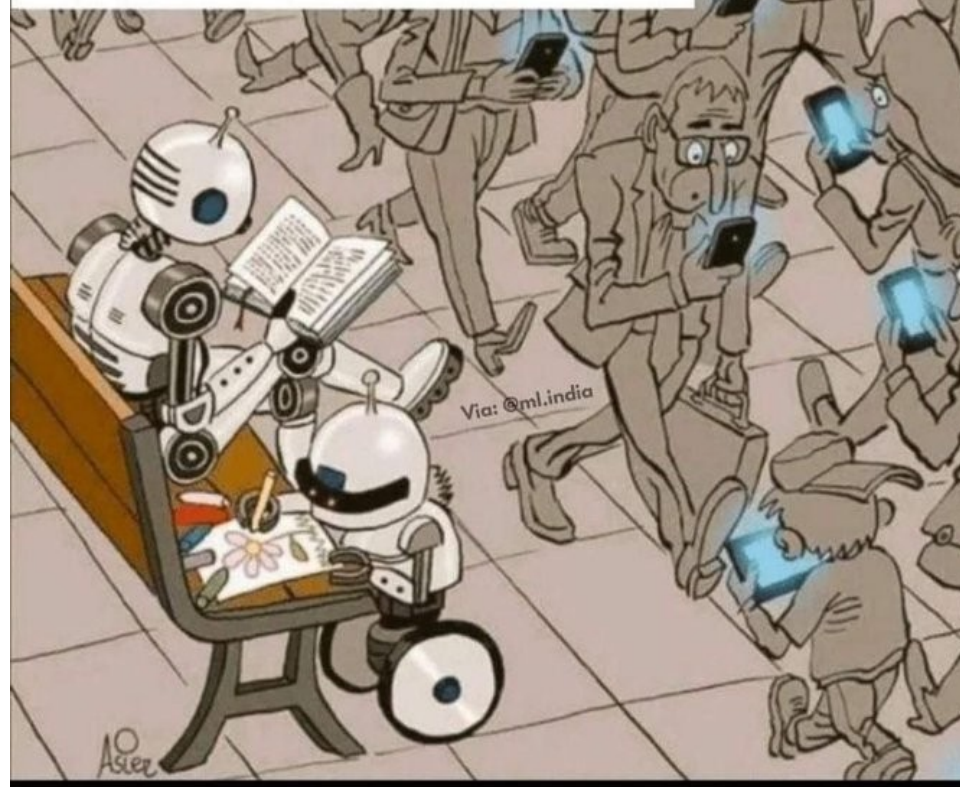
- UK **podporuje** aktívne využívanie AI a získavanie potrebných zručností
- **cieľ**: tvorivé, bezpečné, zodpovedné, etické a transparentné používanie
- nástroje AI majú byť **podporou a pomocou pri štúdiu** (aj rôzne kurzy...)
- **priame vypracovanie celých textov** pomocou nástrojov AI je **plagiarizmus**
- používanie nástrojov AI musí byť **riadne citované**
- pravidlá stanovujú **vyučujúci** (+ výzva prečo pedagóg a nie AI tútor)
- **vízia**: personalizácia vzdelávania; hĺbková analýza dát a pohľad na štúdium

Smernica rektora UK | Dodatok č.1 (citovanie) | Dodatok č. 2 (využívanie)

Nástroje a využívanie AI na RKCMBF UK

Humans are hooked.

Machines are learning.



AI premáha
ľudskú slabosť

AI prekonáva
ľudskú silu

Kredit: autor, AI4D



Ďakujem za pozornosť



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

Moderné systémy AI – dôsledky...

Súčasť paradigmatickej zmeny informačnej a znalostnej spoločnosti

Využívanie nástrojov genAI smeruje k prirodzenému ľudskému spôsobu komunikácie a k radikálnej zmene obsahu i spôsobu práce.

Dôsledky nielen na rozvoj znalostnej spoločnosti, ale – predovšetkým – na ľudskú psychiku, smerovanie a rozvoj modernej spoločnosti.

Strategické technológie

Najlepšie a najsofistikovanejšie systémy AI s očakávaným potenciálom rozvoja môžeme považovať za **strategické technológie**, ktoré **majú potenciál meniť svet – k lepšiemu, ale i k horšiemu...**

Čo by som chcel záverom akcentovať

Potenciál fenoménu umelej inteligencie

Ide o emergentnú a disruptívnu technológiu so schopnosťou zásadne meniť spôsob, akým spoločnosť funguje a s potenciálom posunúť civilizáciu ďalej...

Skutočnosti, ktoré si (ne)uvedomujeme – riziká, zneužitie a dopady

- riziká a limity fungovania i problematika zneužitia
- psychologické a spoločenské dopady

Odvahu, snahu a schopnosť mať “opraty pevne v rukách”

- etické princípy a regulácie
- osobný postoj a spôsob vývoja, nasadenia, prevádzky a využívania
- vzdelávanie a osвета

UMELÁ INTELIENCIA

V texte sú pre umelú inteligencia použité skratky UI (umelá inteligencia) a AI (artificial intelligence).

Univerzita Komenského **podporuje využívanie** nástrojov umelej inteligencie študujúcimi, ako aj vyučujúcimi, no **s transparentne nastavenými pravidlami** obzvlášť v otázke uvádzania zdrojov. Nástroje UI by sa mali používať najmä ako podpora a pomoc pri štúdiu a mali by slúžiť na rozvoj a efektívne využívanie nadobudnutých vedomostí, schopností a zručností študentov. Vyučujúcim sa odporúča priebežné oboznamovanie sa so systémami UI a používanie jej nástrojov pri pedagogickej činnosti.

Univerzita vo svojich nariadeniach a usmerneniach reaguje na vplyv UI na akademické prostredie a formuluje princípy pre **tvorivé, bezpečné, zodpovedné, etické a transparentné používanie** jej nástrojov v prostredí univerzity. Nástroje UI sú určené na používanie pri príprave a úpravách textov, možno ich však primerane aplikovať aj na iné formy generatívnej UI. Univerzita pripravila a priebežne aktualizuje aj **odporúčané spôsoby citovania** nástrojov UI.

Univerzita Komenského dbá na **dodržiavanie akademickkej integrity, transparentnosti a spravodlivosti**. Rozvoj a využívanie nástrojov UI vníma ako podnet, ktorý by mohol byť spúšťačom dôležitých zmien vo vzdelávaní – za predpokladu, ak tento fenomén dokážeme správne uchopiť a prispôsobiť sa novým podmienkam.

Smernica rektora UK pre využívanie AI

Smernica rektora Univerzity Komenského v Bratislave k používaniu nástrojov umelej

Nástroje a informácie k AI v rámci vzdelávacieho procesu na UK RKCMBF UK - Umelá inteligencia

Niečo na čítanie...

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

Moje knihy z oblasti AI

