



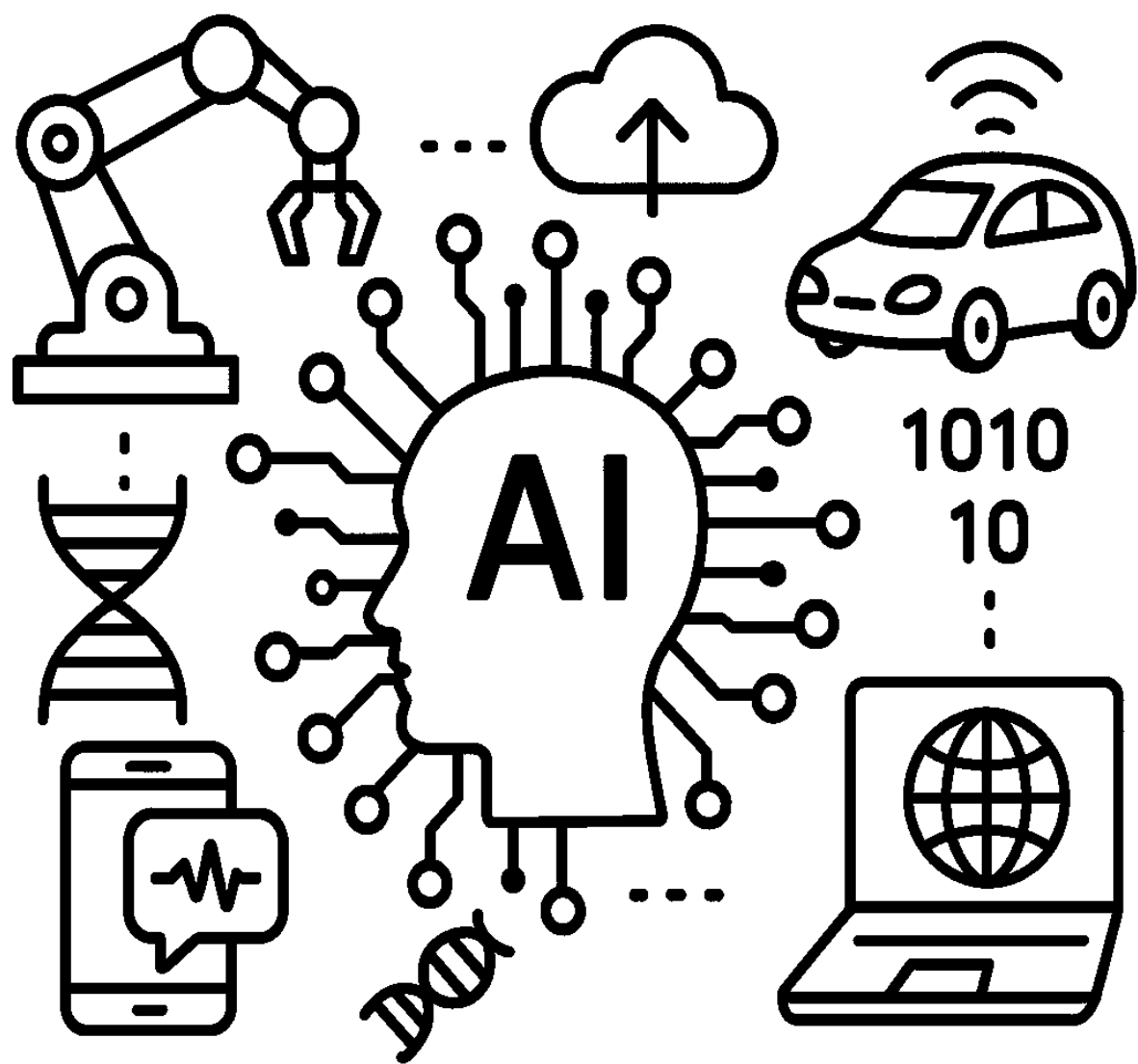
MUŽSKÁ KREATIVITA V DOBE UMELEJ INTELIGENCIE

Peter Šantavý

O S N O V A

- základné postoje vo svetle Zjavenia ▪
 - krátke intro do systémov AI ▪
 - problémy, ktorým treba čeliť ▪
- čo by sme mali, či skôr musíme robiť ▪

AI – artificial intelligence, umelá inteligencia



O S N O V A

- **základné postoje vo svetle Zjavenia** ▪
 - **krátke intro do systémov AI** ▪
 - **problémy, ktorým treba čeliť** ▪
- **čo by sme mali, či skôr musíme robiť** ▪

AI – artificial intelligence

O S N O V A

- základné postoje vo s

- krátke i

- treba čelit' ▪

- mali, či skôr musíme robit' ▪

Múdry muž sa vo viere pozerá na svoje možnosti i používanie schopností cez optiku Božej vôle a toho, k čomu ho Pán pozýva....

AI – artificial intelligence

Základné postoje vo svetle Zjavenia I.

Človek stvorený na Boží obraz (אֱלֹהִים יְצַלְמֵנוּ, imago Dei)

– inteligencia a jej používanie sú jedným zo základných aspektov naplňania tohto "obrazu"

Užívanie (podmanenie) zvereného sveta je v Gn 1,28 vyjadrené hebrejskými výrazmi „podmaňte si ju“ (וְכַבְּשֶׁתָּ) a „panujte“ (וְרָדוּ)

– užívať, rozvíjať, chrániť a z generácie na generáciu si odovzdávať stvorený svet

– dar inteligencie by sa mal prejavovať zodpovedným využívaním rozumu a akýchkoľvek civilizačných vymožeností i technicko-vedeckých schopností pri správe stvoreného sveta

Konfrontácia *chokmá* vs. *sofia* v ketubím (múdroslovnej literatúre SZ)

– ovocím tejto konfrontácie nie je zavrnutie civilizačného rozvoja, ale neustále pozvanie rozvíjať a užívať svet podľa Božích zákonov, vo vzťahu voči Bohu a pre dobro ľudí

Základné postoje vo svetle Zjavenia II.

„Chod'te teda, učte všetky národy...“ (Mt 28, 19) a „Chod'te do celého sveta a hlásajte evanjelium všetkému stvoreniu“ (Mk 16, 15)

– nadčasové a všeobímajúce poslanie, zahŕňajúce aj civilizačný rozvoj

Sv. Pavol na misijných cestách

- v aténskom Areopágu (2. misijná cesta) ohlasuje jazykom helénskych mysliteľov
- v pastoračnom pôsobení v helénskom svete (jazykom synkretizmu Kolosanov,...)
- vyučovanie v škole gréckeho filozofa a učiteľa rétoriky Tyranna v Efeze (3. MC)

Nasledovníci Pavla v civilizačnej konfrontácii

- filozof a mučeník sv. Justín, predstavitelia alexandrijskej a antiochijskej školy,...
- scholastici, scholastická metóda, Occamova britva a následný vedecký rozvoj

Základné postoje v učení Magistéria

Dokumenty pápežov – od komunikačných prostriedkov k IKT a AI

- od Gregora XVI. (1831-1846) až po pápeža Františka
- od IKT k virtuálnemu svetu a AI v dokumentoch od Pia XI. až po Sv. otca Františka
- prorocké pozvanie a poslanie v encyklike ***Redemptoris missio*** sv. Jána Pavla II.

Koncilové dokumenty

- dekrét o spoločenských komunikačných prostriedkoch ***Inter Mirifica***
- pastorálna konštitúcia o Cirkvi v dnešnom svete ***Gaudium et spes***
- celkovo 9 koncilových dokumentov v nejakej miere sa zaoberajúcich aj SKP a IKT

Postoje a dokumenty Cirkvi na Slovensku

- pastoračné plány KBS a rady KBS
- dokumenty biskupov a diecéznych synôd

Magistérium a problematika AI

Rome Call for AI Ethics (2020 – súčasnosť)

- **etické výzvy a princípy** v prístupe, vývoji, realizácii, prevádzke i využívaní systémov AI
- snaha o zjednocujúce pôsobenie a angažovanie náboženstiev, štátov, korporácií,...
- konferencia, dokument a participácia aktérov, ktorých zoznam sa rozširuje i v súčasnosti

Antiqua et nova (2025)

- komplexný dokument o vzťahu medzi ľudskou a umelou inteligenciou
- praktický a odborný dokument: pohľad odborníkov z rôznych uhľov pohľadu a profesií
- **“technologický pokrok je súčasťou Božieho plánu pre stvorenie”**

Dokumenty riadneho Magistéria o AI

- príhovory a dokumenty pápeža Františka
- dokumenty vatikánskych dikastérií

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)
- zachovávať **ľudskú dôstojnosť** a podporovať integrálny rozvoj ľudskej osoby a spoločnosti
- 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

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 spoľahlivosti nielen z technologického hľadiska, ale zohľadňujúce aj sociálne prostredie a dopady na spoločnosť

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

O S N O V A

- základné postoje vo svetle Zjavenia ▪
 - krátke intro do systémov AI ▪
 - problémy, ktorým treba čeliť ▪
- čo by sme mali, či skôr musíme robiť ▪

AI – artificial intelligence

O S N O V A

- základné postoje vo svet

- krátke int
- ... AI

- ... treba čelit'

- ... sme mali, či skôr musíme robit'

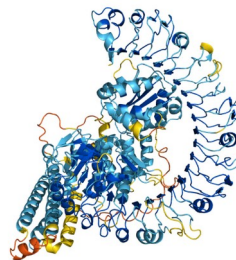
**Každá dostatočne pokročilá technológia
je na nerozoznanie od mágie :-)**
(Arthur C. Clarke)

AI – artificial intelligence

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...)



(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) (7, 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)

(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) (1, 1) (7, 7)
(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)
(9, 9) (7, 7) (5, 5) (4, 4) (1, 1) (8, 8)
(7, 7) (4, 4) (7, 7) (7, 7) (4, 4) (2, 2)
(8, 8) (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 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

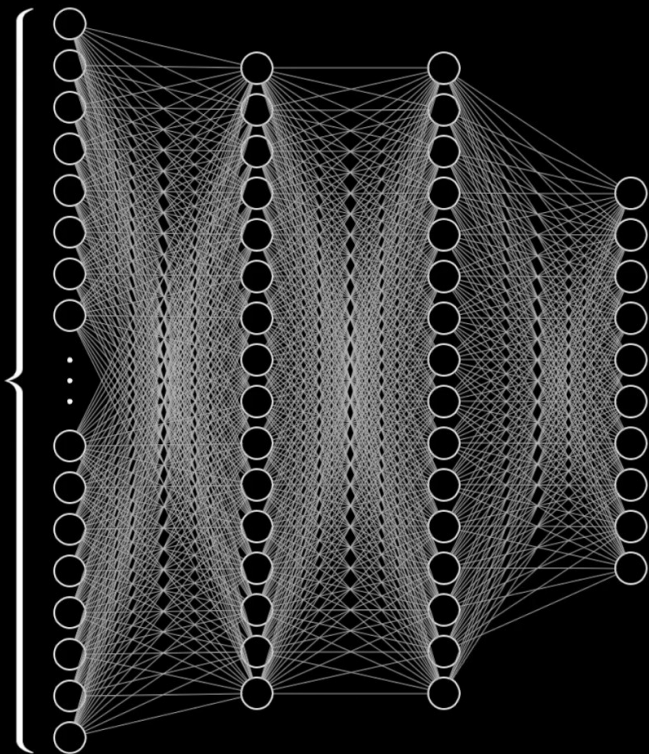
$$28 \times 28 = 784$$

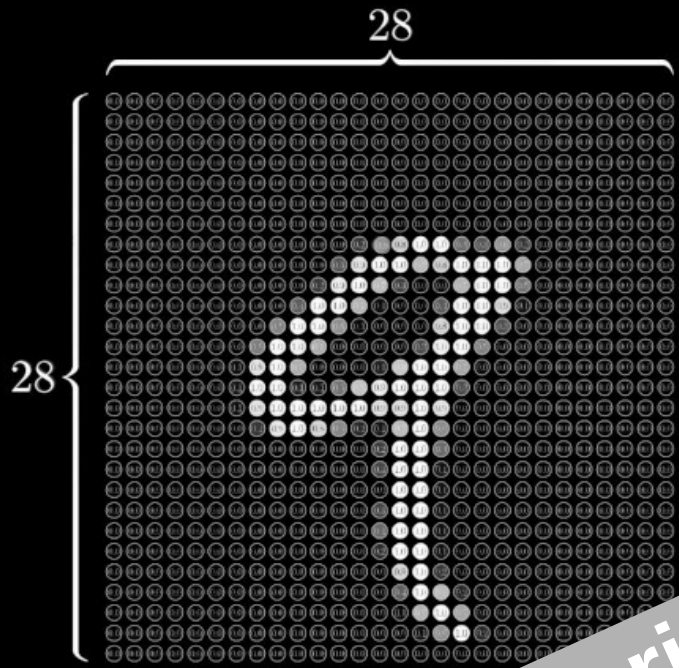
13 002 !!!

28



784





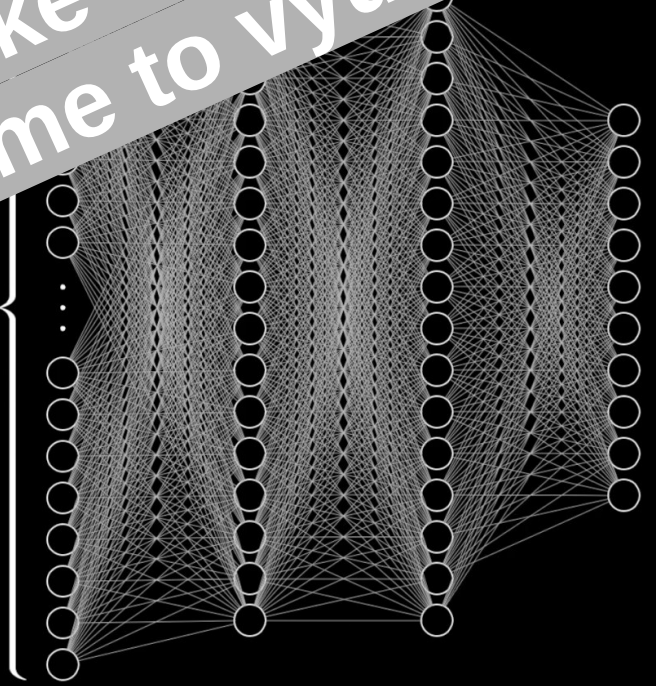
$$28 \times 28 = 784$$

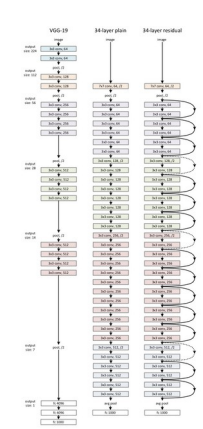
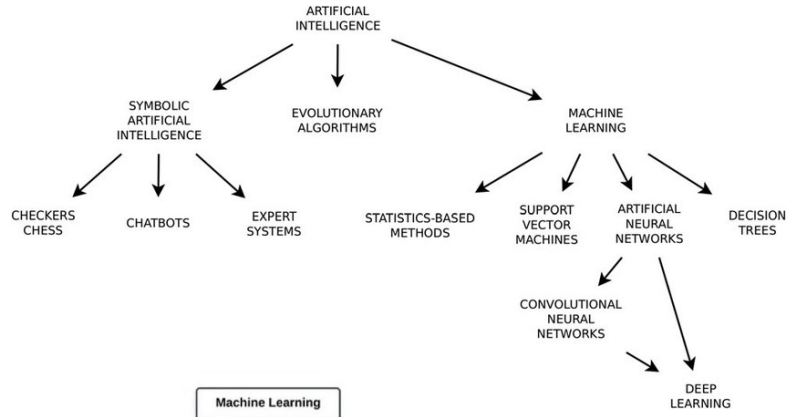


13 000

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

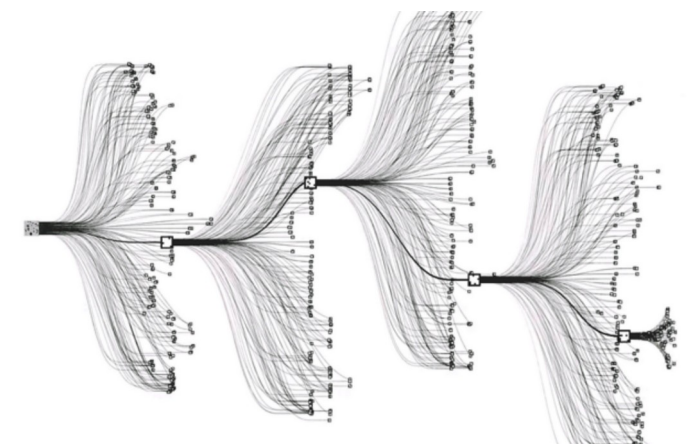
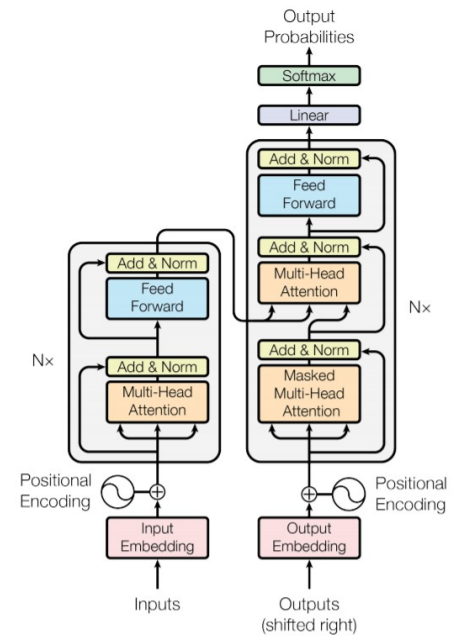
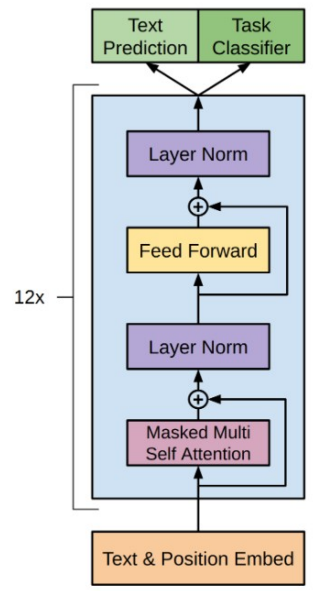
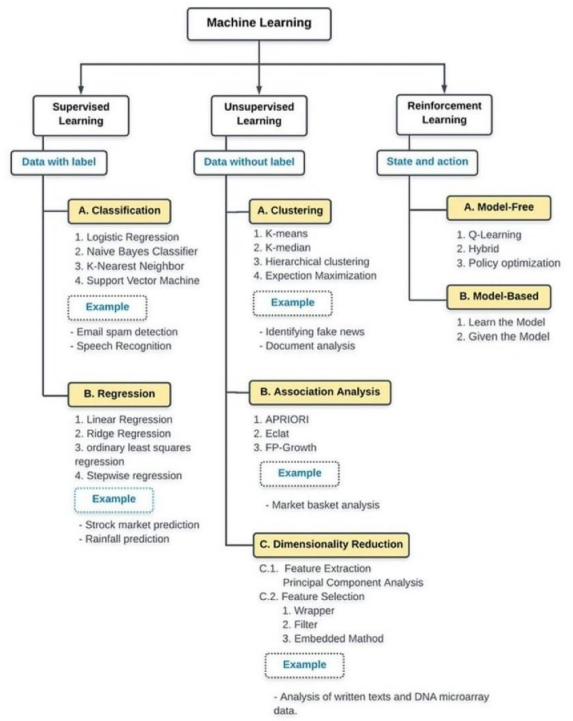
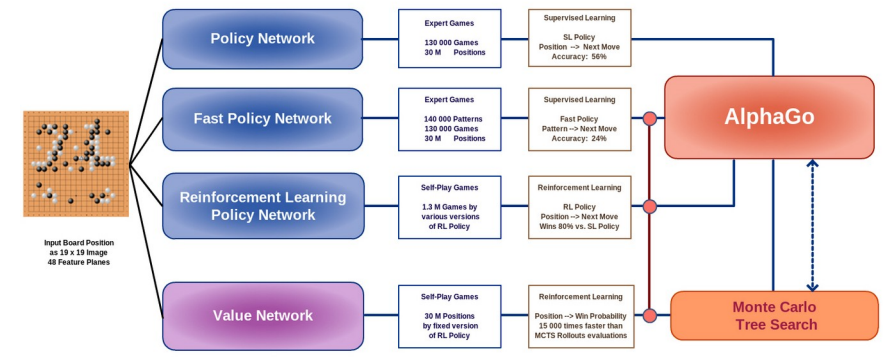
784





AlphaGo Overview

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





PREČO AI?

Existujú ešte nejaké iné dôvody, prečo AI?

...okrem vedeckých a technologických napr. finančné, mocenské, vojenské, ideologické, spoločenské,...

» bytostné «

P R E

Existuje

vody, prečo AI?

...O... technologických napr. finančné,
mocienské, vojenské, ideologické, spoločenské,...

**Túžba spoznávať, rozvíjať a tvoriť...
Z pohľadu zjavenia v pozitívnom
i negatívnom zmysle...**

» bytostné «

Definícia umelej inteligencie

Inteligenciu ľudskeho 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...

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)...

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 – 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 – 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.*

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**.

Rozlišujeme rôzne stupne AGI: BAGI (below-human), HLAGI/HAGI (human-level AGI), MAGI (moderately-superhuman AGI), SAGI/ASI (superintelligent AGI).

Generatívne systémy AI (genAI)

Princíp činnosti genAI

- algoritmy genAI sú schopné podľa vstupnej sekvencie generovať rôznorodý obsah – text, obrázky, audio, video,..., ktorý napodobňuje ľudský obsah a schopnosť rozumieť
- výstupy sú generované na základe predtrénovaných vzorov a zložitých algoritmov

Stupne tréovania a prípravy genAI

- data pre-processing -> foundation model -> fine tuning / assistant model -> RLHF
- tokenization, embedding (+ encoders), transformers (+S/MH attention), decoders,...

Mnohé smery súčasného i očakávaného vývoja

- zvyšovanie kvality odpovedí a výkonu (škálovanie, algoritmy, destilácia, quantization,...)
- zlepšovanie “myslenia” (reasoning modely) – RAG a prispôsobené modely
- prepojenie s reálnym svetom (agent AI) – multimodálne modely

Generatívne systémy AI (genAI) – prečo?

Technologický posun

- činnosť podobná ľudskému mysleniu na základe naučených vzorov a technológie transformerov dosahujúca **pozoruhodné výsledky a široká škála využitia**
- schopnosť čerpať z extrémneho množstva zdrojov a pracovať s bežnými dátami

Zmena v spôsobe využívania nástrojov AI

- konverzačné rozhranie umožňujúce komunikáciu v bežnom jazyku formou dialógu
- pokročilá – akoby tímová – **spolupráca medzi človekom a strojom**
- riešenie úloh reálneho sveta prostredníctvom komunikácie s genAI (agenti AI)

*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 a sociologický rozmer spoločnosti.***



**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...**

Najlepšie a najsofisticovanejšie systémy AI

s očakávanými potenciálmi. Potenciálne môžeme

považovať za strategickú technológiu, ktoré majú

potenciál meniť svet k lepšiemu, ale i k horšiemu...

AI systémy myslenie len napodobňujú!

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

O S N O V A

- základné postoje vo svetle Zjavenia ▪
 - krátke intro do systémov AI ▪
 - **problémy, ktorým treba čeliť** ▪
- čo by sme mali, či skôr musíme robiť ▪

AI – artificial intelligence

O S N O V A

- základné postoje vo svetle filozofie

- krátke intro do sveta AI

- čo treba čeliť

- čo by sme mali, či skôr musíme robiť

Nič veľkého nevstúpi do života smrteľníkov bez prekliatia.
(Sofokles)

AI – artificial intelligence

S čím sa potýkame?

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

Technologické riziká a limity ANI

Zraniteľnosti, slabiny a klamanie systémov strojového učenia

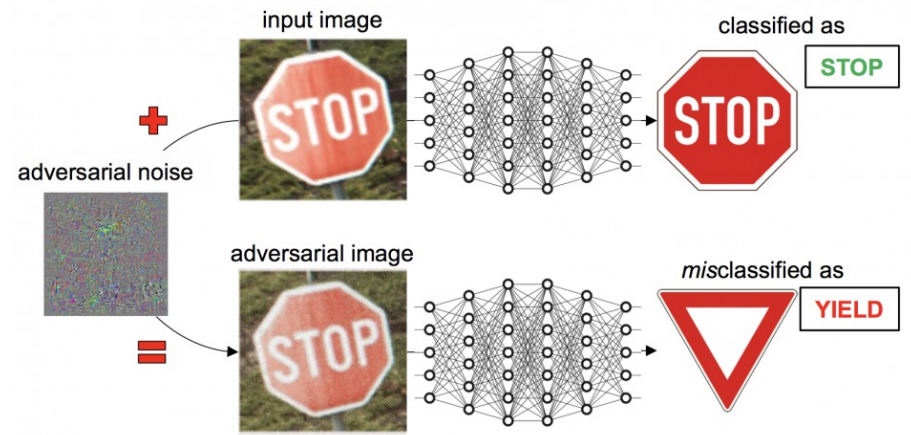
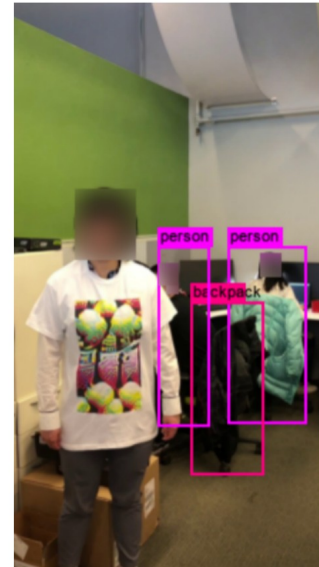
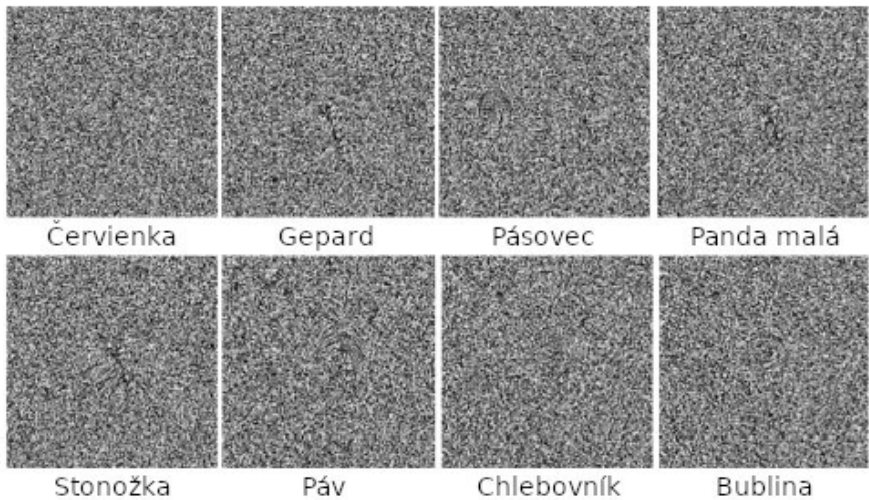
- malá množina trénovacích dát (training dataset)
- nesprávne zvolená/nekvalitná množina trénovacích dát a predsudky (biases)
- nadmerné prispôsobovanie sa tréningovým údajom (overfitting to training data)
- efekt dlhého chvosta (long-tail effect)
- povery (superstition)
- klamanie hlbokých sietí a ich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

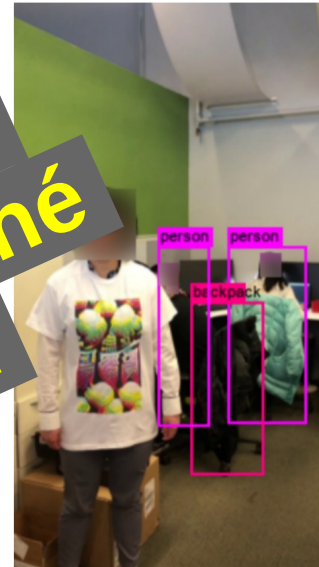
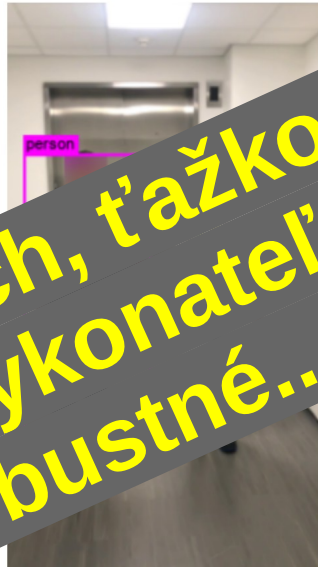
Technologické riziká a limity ANI

Zraniteľnosti, slabiny a klamanie systémov strojovej učenia

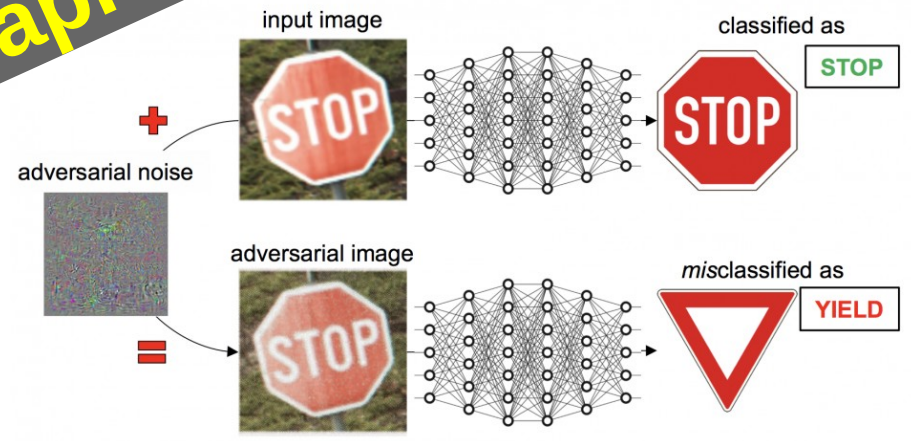
- malá množina trénovacích dát (training data)
- nesprávne zvolená/nekvalitná množina dát a predsudky (biases)
- nadmerné prispôbovanie na tréningový súbor (overfitting to training data)
- efekt dlhého chvilu
- poverky (stereotypes)
- klamanie a ich zraniteľnosti (adversarial examples, hacking,...)
- špecifické útoky na generatívnych systémov (jailbreak: roleplay, base64, UTS, prompt injections, data poisoning / backdoor attacks,...)

Zlyhávajú pri nesprávnom použití ...
Ešte viac pri zámernom zneužití...





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 dôvernosc'** (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 napadnuteľnosť

– vektor útoku na zraniteľnosti AI

– neexistuje dokonale bezpečný systém

Infraštruktúra – je základom pre bezpečnosť

– S rastúcou veľkosťou zariadenia sa požadovaný výpočtový výkon

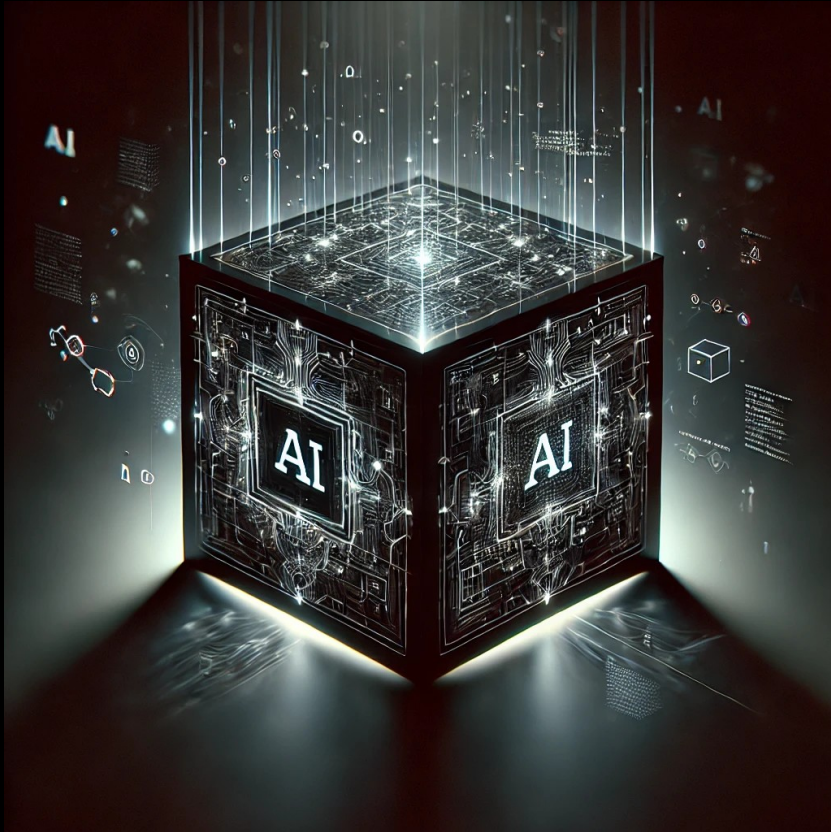
– rastie exponenciálne

– 'komplexnosť', je rizikovým faktorom bezpečnosti a stability

fungovania systémov

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

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 sa základne čoho robia hlboké a interpretovateľných a interpretovateľných systémoch AI ;-)

Pracujeme na vysvetliteľných zásade nevieme, ako sa 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
- **psychologické profily a modely predikcie** nášho správania
- **manipulácia a ovplyvňovanie** nášho vnímania a konania
- vytváranie **sociálnych bublín** a rozdelenia, **relativizácia hodnôt a pravdy**
- technológie AI vedú k závislosti, **depresiám, úzkostiam, digitálnej demencii a hnilobe**

Vybrané dôsledky a riziká pre človeka

Extrémne zhromažďovanie dát

- neustály dohľad a sledovanie bez k
- psychologické profily a správania
- manipulácia vnímania a konania
- sociálna bublín a rozdelenia, relativizácia hodnôt a pravdy
- technológie AI vedú k závislosti, depresiám, úzkostiam, digitálnej demencii a hnilobe

Ide o riziká, ktoré si mnohokrát ani neuvedomujeme!!!

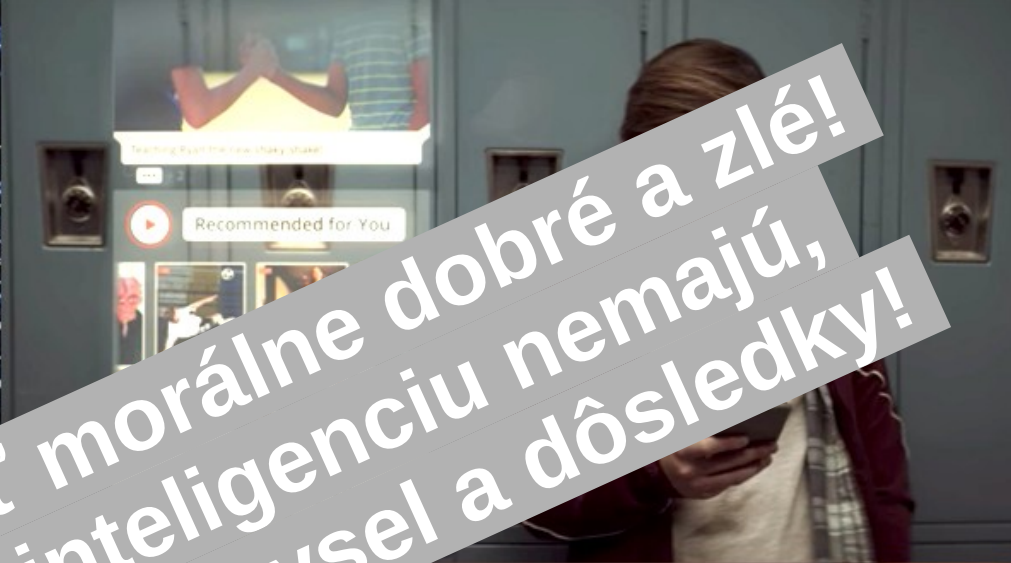
N

/the social dilemma

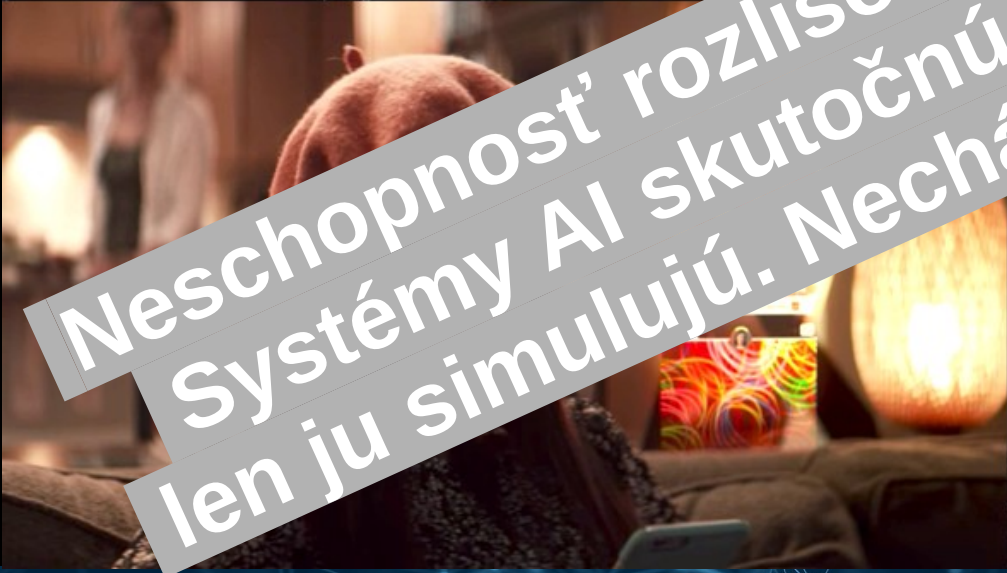


N

/the social dilemma



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



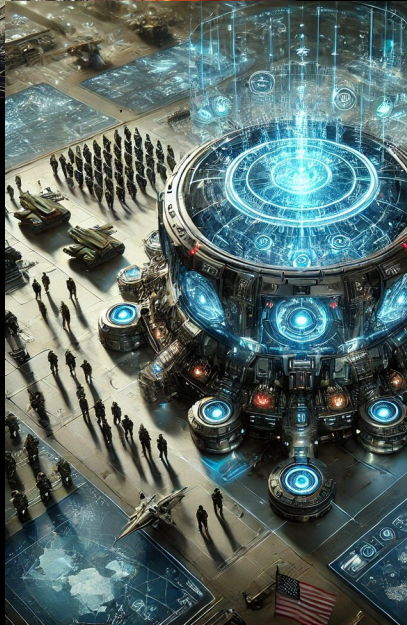
Vybrané dôsledky a riziká pre spoločnosť

Dohľad a sledovanie ako základ...

- neustály **dohľad a sledovanie**, ktoré je ťažko kontrolovať

Riadenie a fungovanie štátu

- **nezvládnuté a neadekvátne** algoritmické riadenie štátu / algokracia (súdnictvo, zdravotníctvo, bankový sektor,...)
- **sociálny kreditový systém**
- **neetické** vojenské nasadenie
- budúcnosť **práce a fungovania spoločnosti** (mení sa jej paradigma)



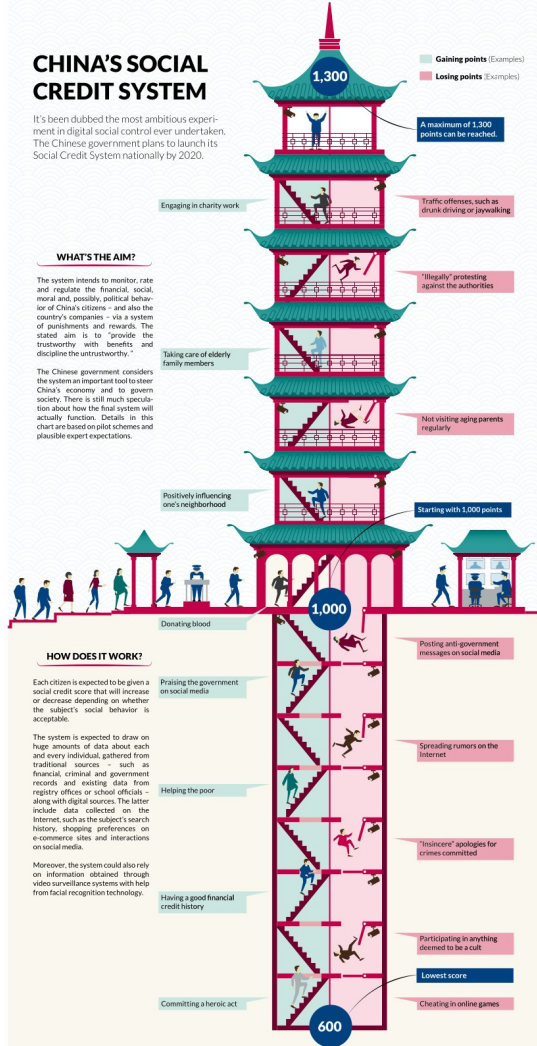
CHINA'S SOCIAL CREDIT SYSTEM

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.



1. THE FUTURE OF CHINA: A VISION OF THE 21ST CENTURY. © BERTELSMANN STIFTUNG
 2. THE FUTURE OF CHINA: A VISION OF THE 21ST CENTURY. © BERTELSMANN STIFTUNG
 3. THE FUTURE OF CHINA: A VISION OF THE 21ST CENTURY. © BERTELSMANN STIFTUNG

BertelsmannStiftung



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.

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 the basis of a comprehensive system of points.

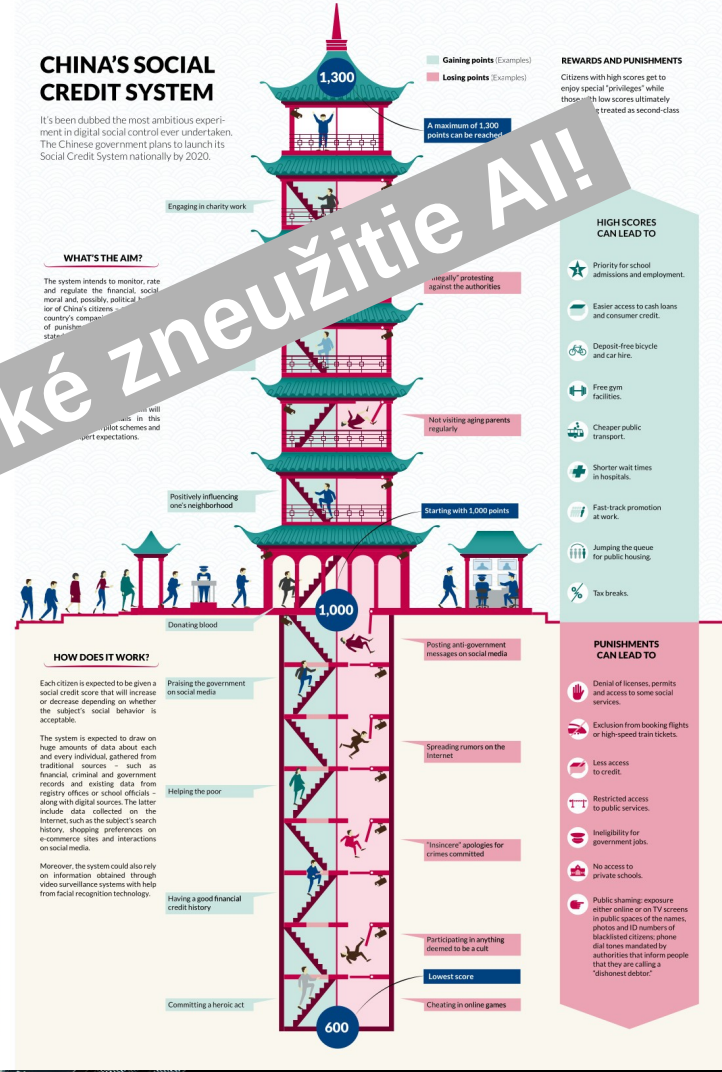
It will be used to pilot schemes and to set expectations.

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.



Mocenské, vojenské, ideologické zneužitie AI!

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
- systémy AI skutočnú inteligenciu nemajú, len ju simulujú! Nechápu zmysel a dôsledky...

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 AI

Extrémna šírka využitia generatívnych systémov (LLM)

- LLM sa dokáže učiť vzory, kontext aj sémantiku jazyka a dokáže reagovať na otázky, konverzáciám, poskytovať súvislé a kontextovo relevantné odpovede
- systémy AI skutočnú inteligenciu nemajú, nemajú schopnosť chápať zmysel a dôsledky...

Halucinovanie

- vymýšľa informácie, ktoré systém predkladá ako relevantné a správne

Prítomnosť nepravdivých výstupov

- výskyt chybných a nesprávnych odpovedí

Problémový spôsob narábania s údajmi

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

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

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 supľujú 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 DEMENCIE

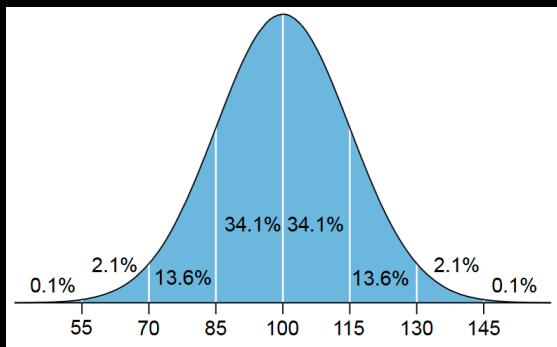
Dlhodobé a nesprávne využívanie digitálnych technológií, osobitne generatívnych AI a sociálnych sietí, prináša **intenzívne kognitívne vyčerpanie** a **postupne strácať** kognitívne schopnosti a **psychickej odolnosti!**

Vďaka systému 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...



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 vzťahov s AI. Neschopnosť fungovať v reálnom svete a rozvíjať skutočné vzťahy, hodnotové posuny, závislosti, zlyhávanie v reálnom živote (kritické myslenie, sexualita,...).

Etické problémy (nielen) generatívnych AI

Mozgová hniloba (brainrot) – postupná strata schopnosti vnímať hlbší kontext. Vytvárajú sa dopamínové a adrenalínové závislosti na obsahoch generovaných AI, ktoré stimulujú produkciu adrenalínu a dopamínu, na ktorých závisí náš mozog.

Deepfake – vytváranie falšovaných videí ľudí a podvrhnutie ich identít. Príklady: politická propaganda, konkurencia, pornografia, sociálne médiá. Zdieľaním týchto informácií sa znižuje dôvera medzi pravdou a lžou.

Stíhajúci únik (AI escape) – únik do virtuálneho sveta a vzťahov s AI.

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

Virtuálne zážitky v mozgu mnohých evokujú tak vysoké hladiny adrenalínu a dopamínu, aké nie sú schopní v reálnom živote dosiahnuť. Únik do virtuálneho sveta...

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



This image was AI-generated



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

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?”

Ďalšie etické problémy generatívnych systémov

Neschopnosť robiť vlastné zodpovedné rozhodnutia – delegujeme to na stroje.

Narušenie psychického vývoja priskorým, resp. nezrelým používaním digitálnych technológií a technológií umelej inteligencie (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. prehrávaním digitálnych technológií a technológií umelej inteligencie (virtuálny autizmus).

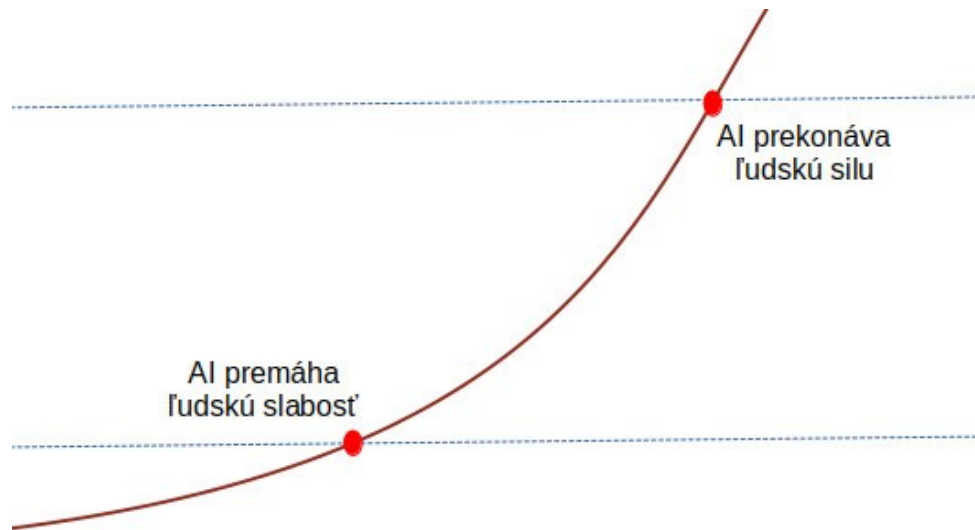
Riziko erózie identity – prehrávanie "čítanie" ľudskej komplexnej inteligencie, racionality a na úroveň technickej inteligencie.

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

Možnosť manipulácie vlastníckmi systémov AI.

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

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

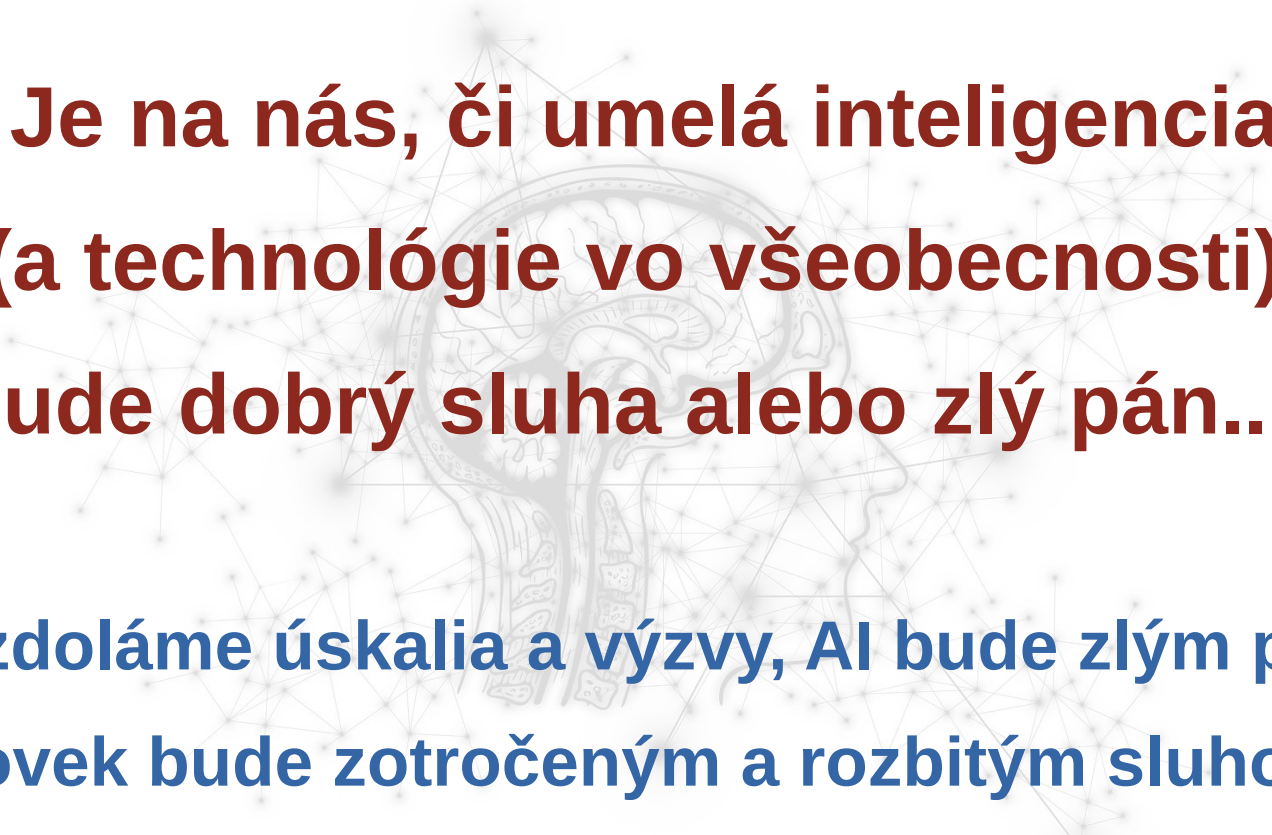


Míľnikom, ktorého by sme sa mali obávať, nie je budúca technologická singularita v oblasti umelej inteligencie, v ktorej AI prevýši náš intelekt, ale oveľa skôr moment, keď **technológia ovládne a prekoná naše slabosti...** už vtedy prichádza víťazstvo umelej inteligencie a porážka ľudstva.

O S N O V A

- základné postoje vo svetle Zjavenia ▪
 - krátke intro do systémov AI ▪
 - problémy, ktorým treba čeliť ▪
- čo by sme mali, či skôr musíme robiť ▪

AI – artificial intelligence



**Je na nás, či umelá inteligencia
(a technológie vo všeobecnosti)
bude dobrý sluha alebo zlý pán...**

**Ak nezdoláme úskalia a výzvy, AI bude zlým pánom
a človek bude zotročeným a rozbitým sluhom...**

Je na nás, či umelá inteligencia
(a technológie vo všeobecnosti)
bude dobrý pán... alebo zlý pán...

**Čo treba robiť, aby sme naplnili biblický
obraz stvorenia, ktoré možno dobre
spravovať a s úžitkom využívať?**

AI vyzvy, AI bude zlým pánom
a zotročným a rozbitým sluhom...

Jasný postoj a poznanie

Jasný postoj k technológiám prameniáci z viery a Zjavenia

Základné poznanie moderných technológií a systémov AI

Poznanie a **chápanie rizík**, ktoré sú spojené s AI a modernými technológiami a ich dôsledkov

– vnímanie mnohých rizikových faktorov je viazané aj na náš hodnotový systém a svetonázor

Definovanie cieľov a zmyslu používania AI

Definovanie cieľov – čo tým chcem dosiahnuť, prečo AI?

- v pracovnej oblasti
- v osobnom živote
- v spoločenstve

Zmysel – aký zmysel to má mať a k čomu to vedie?

- v osobných cieľoch a sebarealizácii
- pre spoločné dobro a službu blížnym, pozitívny rozvoj sveta,...
- pre vyšší cieľ (Božie kráľovstvo, spása, oslava Boha,...)

Spôsob využívania systémov AI

Používať na dobré – morálne a na základe **etických princípov**

Poznajúc riziká **využívať nástroje AI** (brúska, vŕtačka,...)

V kontexte stanovených cieľov **nebát sa** byť kreatívny, tvorivý a odvážny – **nebát sa tvoriť, rozvíjať a spravovať**

Používať nástroje AI s pokorou vo vedomí vlastných mantinelov, slabostí a zranení

– bez pravdy nielen o technológiách moderného sveta, ale – a to predovšetkým – pravdy o sebe to nepôjde

V službe iným

Formovať tých, ktorí sú mi zverení

- **chrániť deti** pred prístupom k AI, jej obsahu a nesprávnemu využívaníu
- **formovať a vychovávať** k správne mu používaníu
- **prakticky viesť** k správne mu využívaníu iných (otcova dielňa...)

Angažovať sa v oblasti AI

- etické a transparentné systémy AI v ochrane ľudskej dôstojnosti, spoločného dobra, sociálnej spravodlivosti, pravej slobody a práva

S askézou ako nástrojom slobody

V kontexte čností

- **múdrosť, pravda a spravodlivosť**, na ktorých je postavená schopnosť tvoriť, spravovať a rozvíjať...
- **miernosť** pre primerané a **mravnosť** pre osožné používanie AI

V askéze

- schopnosť **byť off-line** mimo systémov AI a virtuálneho sveta
- **vytrvať v zásadách** správneho využívania
- **vedome realizovať činnosť**, ktorá je “nudná” a nie je spojená s dopamínovým a adrenalínovým efektom (dopamínový pôst)
- **činnosť v reálnom svete** – fyzická, kreatívna, služba, vzťahy

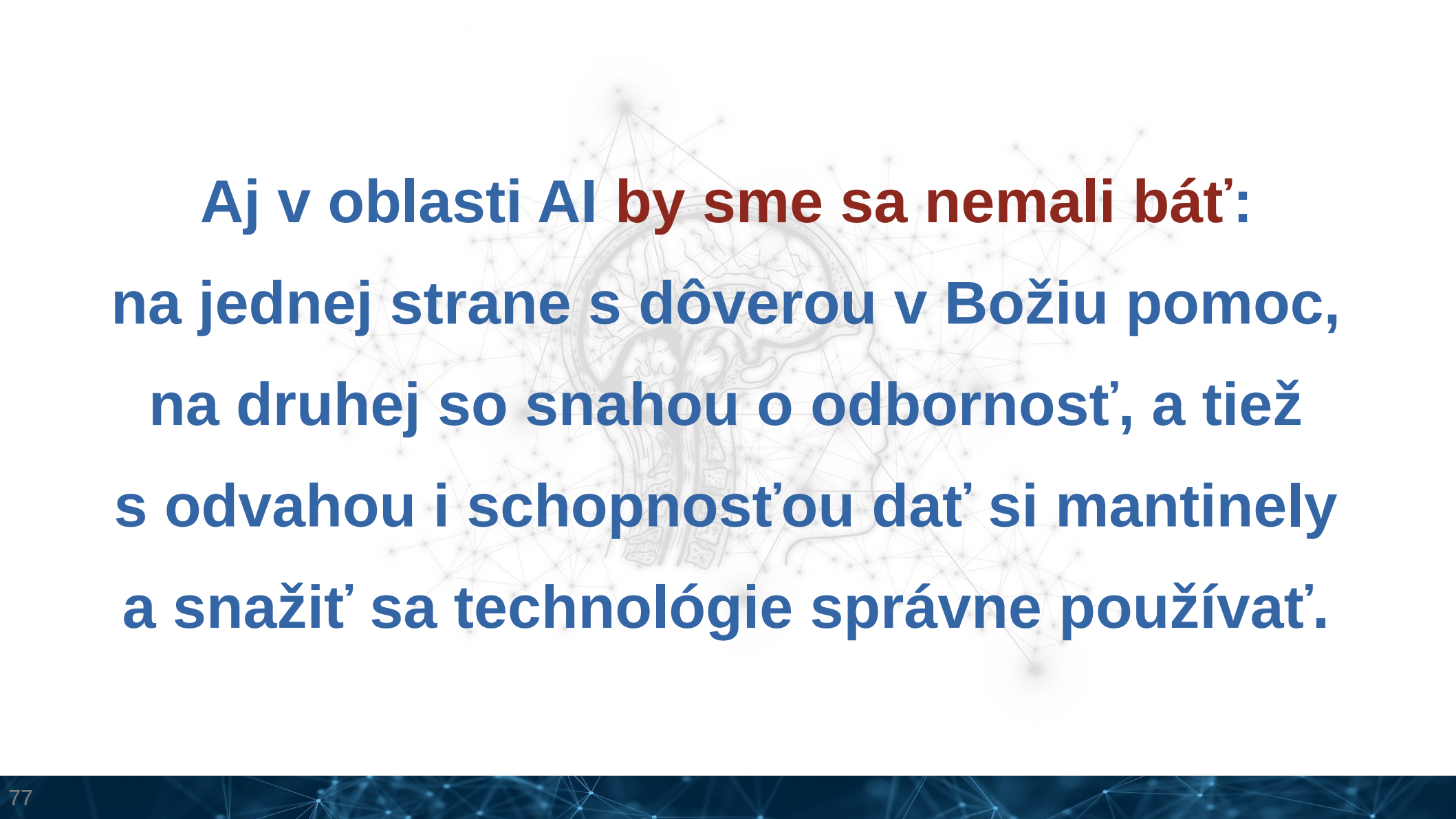
Na pevnom základe duchovného života

V rozvíjaní živého vzťahu voči Bohu

- vzrastať vo viere, poznaní a otváraní sa pre Pána
- rozvíjať **duchovný život**, v ktorom modlitba má osobitné miesto
- žiť **sviatostným životom** – v milosti a moci Sv. Ducha

V skutkoch a živote evanjelia

- **skutky** telesného a duchovného milosrdenstva v reálnom svete
- primeraná **účasť na budovaní** spoločenstva a angažovanosti
- hľadanie spôsobov, ako to realizovať **aj pomocou nástrojov moderného sveta** pre službu i v reálnom svete i s dosahom na virtuálny svet



Aj v oblasti AI by sme sa nemali báť:
na jednej strane s dôverou v Božiu pomoc,
na druhej so snahou o odbornosť, a tiež
s odvahou i schopnosťou dať si mantinely
a snažiť sa technológie správne používať.

Mužská kreativita v dobe umelej inteligencie

- **v napĺňaní** Božieho zámeru s človekom v dobe AI a aj s AI
- **v schopnosti** chrániť a rozvíjať to, čo je nám zverené, v reáliách technologického sveta
- **v odvahe** zachovať si a rozvíjať vnútornú slobodu a hodnoty, na ktorých je postavená schopnosť tvoriť, spravovať, budovať a milovať
- **v spoznávaní** moderných rizík, našich limitov a slabostí, ktoré nás ohrozujú i zväzujú a nebáť sa im postaviť čelom
- v odvahe hľadať moderné spôsoby **účasti** na budovaní Božieho kráľovstva a civilizácie lásky



Ďakujem za pozornosť

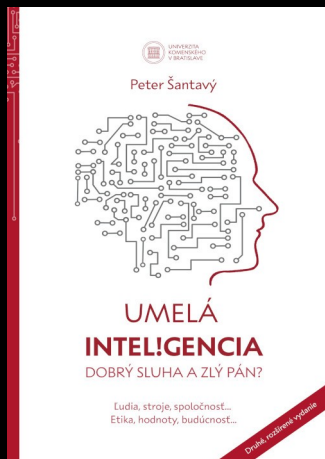
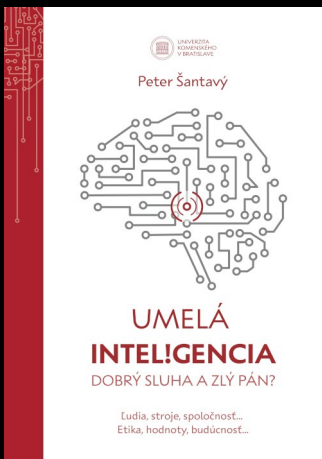
ThLic. Ing. Peter Šantavý, PhD.
peter@santavy.sk





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 - Umeľá inteligencia**
Moje knihy z oblasti AI

Moje knihy z oblasti AI