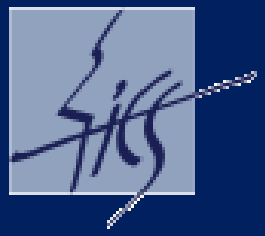


Centrum Karla Čapka
pro studium hodnot ve vědě a technice

Máme se bát/nebát umělé inteligence?

Jiří Wiedermann
Ústav informatiky AV ČR
Praha

(Nehodící se škrtněte)

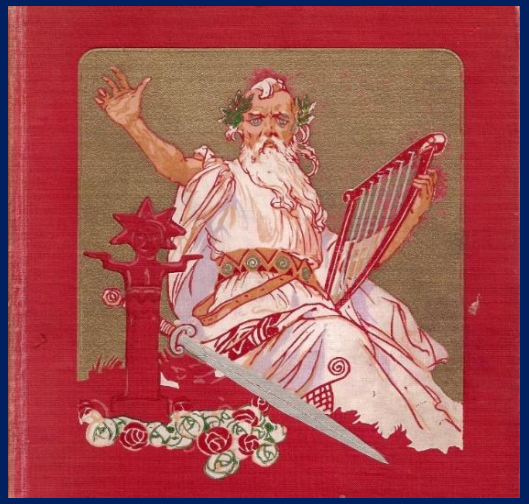
Co je to: je to moudré, ale není to živé?

(ChatGPT)

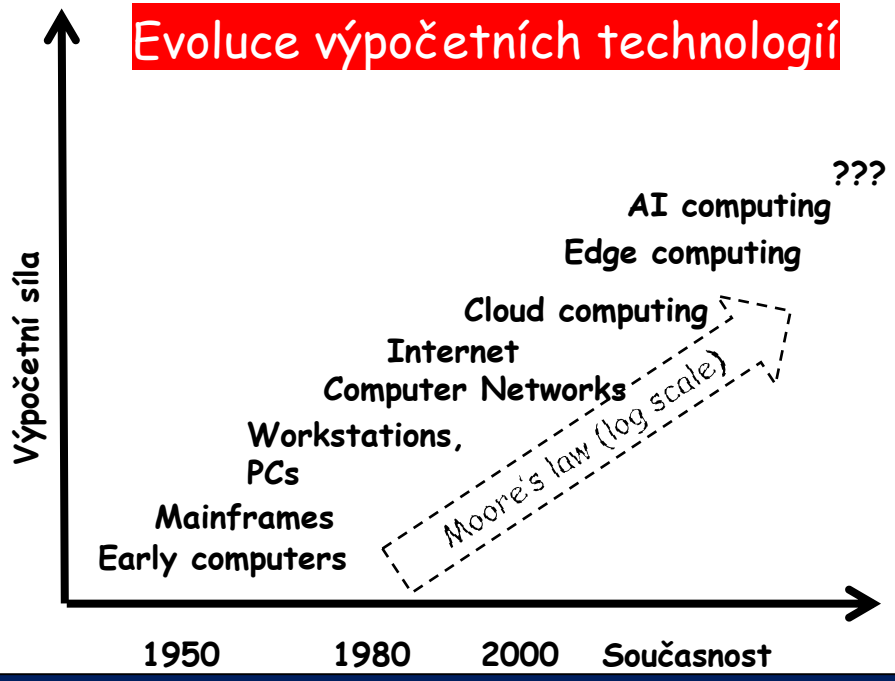
Podobenství o mluvícím psovi

Proč rozvíjíme umělou inteligenci?

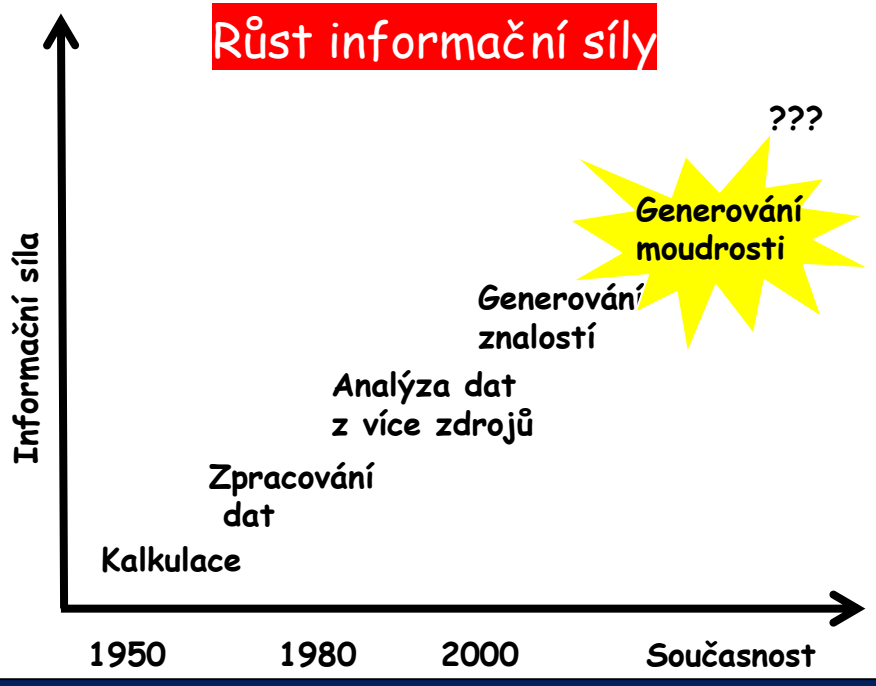
Jaký účel má používání umělé inteligence?



Evoluce výpočetních technologií



Růst informační síly

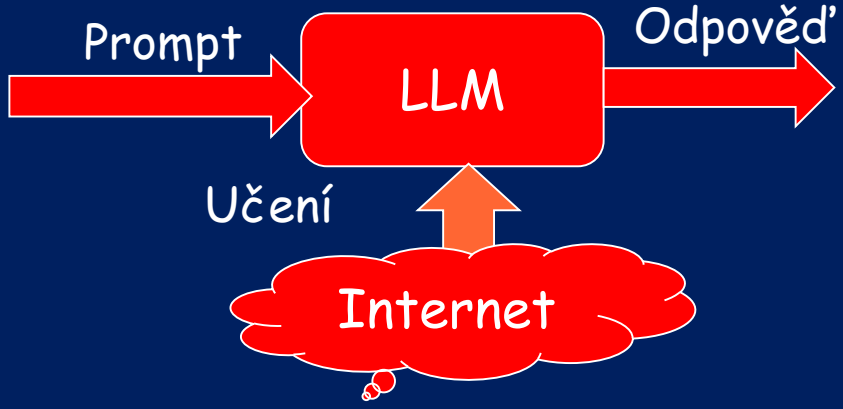


Moudrost

C. H. Spurgeon,
The Fourfold Treasure (1871)

„Předpokládám, že moudrost znamená správné používání znalostí. Samotné vědění ještě neznamená moudrost. Je mnoho lidí, jež vědí mnoho - ale to z nich dělá ještě větší hlupáky. Není nad hlupáka než vědoucího hlupáka. Poznání, jak používat znalosti, znamená mít moudrost."

Přirozená i umělá moudrost je správné používání znalostí pomocí účelného chování - kombinovaný efekt kognice, znalostí a jednání směřující k vytváření užitečných a etických hodnot.



Velké jazykové modely generují - zatím - iluzorní moudrost:

- Prompt stanovuje užitečnou a etickou hodnotu, kterou žádáme.
- Model jedná účelně, protože tuto hodnotu poskytuje na základě znalostí získaných **z popisu reálných trénovacích dat**.

Umělou inteligenci rozvíjíme za účelem vytváření nástrojů pro generování umělé moudrosti.

Spolupráce s těmito nástroji nám umožní činit moudrá rozhodnutí a správně jednat.

Ale: Můžeme věřit umělé inteligenci?

Teoretická překážka: Neexistuje algoritmus jež by pro libovolný AI systém rozhodnul, jestli je za jakýchkoliv okolností bezpečný – tj. jestli bude vždy jednat a souladu s „lidskými hodnotami".

Další problém: neprůhlednost neuronových sítí. Co můžeme dělat?

1. Regulace?
2. Navrhovat AI systémy s omezeným jednáním.
3. Pro konkrétní systémy, hledat na míru šitý důkaz hodnověrnosti.
4. Vybavit systém bezpečnostními svodidly.
5. Vymyslet, jak může méně inteligentní systém kontrolovat systém s vyšší inteligencí.

Jak dosáhnout soulad se superinteligencí

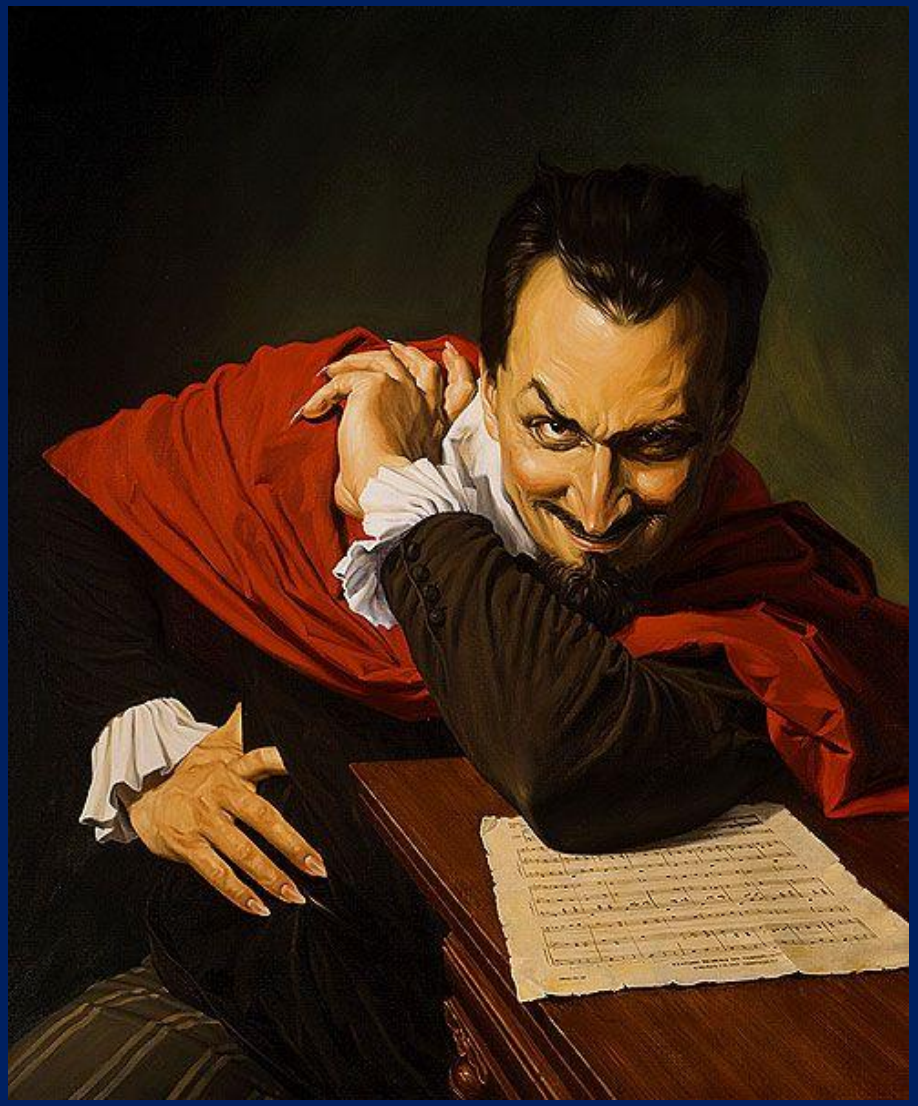
Abychom mohli usměrňovat a využívat systémy převyšující naši inteligenci, potřebujeme vědecký a technologický průlom v oblasti bezpečnosti AI.

Máme se bát nebo nebát umělé inteligence ?

Faustovské dilema

problém vybrat si mezi zjevným prospěchem z vývoje a používání umělé inteligence, která nemusí být zcela bezpečná, a potenciálním nebezpečím, že se nám její vývoj a používání vymkne z rukou.

Druhá možnost znamená ohrožení naší existence.



Jak se rozhodneme?

Kterou cestu si zvolíme?

Jsme už připraveni se rozhodnout?

Děkuji za pozornost