

Umělá inteligence

Doc.Ing. Jiří Krejsa, PhD

krejsa@fme.vutbr.cz, A2/720

Prerekvizity

O čem to bude

K čemu vám to bude

Požadavky k zápočtu / zkoušce

Nástroje

Literatura

Úvod do problematiky

Prerekvizity

Nějaká ta matematika – operace s maticemi, derivace, konvoluce

Základy programování – cyklus, podmínka, funkce, předávání parametrů

Matlab – m-files, functions, toolboxes

Python, NumPy, TensorFlow

O čem to bude

Prohledávání stavového prostoru (do hloubky, do šířky, Dijkstra, A*)

Teorie her (min-max)

Stochastická optimalizace (genetický algoritmus)

Neuronové sítě (klasické vrstvené dopředné)

Konvoluční neuronové sítě (klasifikační úlohy)

Rekurentní neuronové sítě

Zpětnovazební učení

Extra přednášky (SVM,...)

K čemu vám to bude

Budete rozumět základům v praxi reálně využívaných algoritmů

Rozšíření obzorů o metody, které mohou fungovat tam, kde ostatní selhávají

A.I. není samospasitelná – nenechte se ovládnout myšlenkou, že zvládne všechno

Požadavky k zápočtu / zkoušce

Zápočet

- odevzdání povinných úloh
- *vlastní algoritmus na hru Reverzi, účast na celokruhovém turnaji, poražení algoritmu vašeho vyučujícího (tohle ještě uvidíme, záleží na cvičícím)*

Zkouška

- projekt
- podle složitosti může být i skupinový
- schválení zadání vyučujícím
- report PDF
- zdrojáky emailem nebo odkaz na úložiště s daty
- prezentace před ostatními
- *ústní zkoušení teorie (vylosování otázky)*

Nástroje

Matlab

- základní Matlab
- Neural Network Toolbox
- Deep learning Toolbox

Python

- základy

Numpy

- základy

TensorFlow

- základy, rozchození výpočtu na grafické kartě

Literatura

Mařík a kolektiv: Umělá inteligence 1-5

Matlab manuály (online)

Matlab Central file Exchange (online)

Python manuály (online)

TensorFlow manuály (online)

BigBook (neoficiální)

Úvod do problematiky

Co to je inteligence?

- Umělá inteligence je „něco“, co vnímá okolní prostředí a dokáže „konat“, aby maximalizovalo šanci na dosažení cíle. Což souvisí s termíny jako je **učení** a **řešení úloh**.
- Co všechno do oboru spadá? To co ještě nebylo vyřešeno (některé úlohy se za AI považovaly (OCR, Face detection), ale nyní se berou jako hotové a vyřešené).
- Dlouhodobý cíl: obecná inteligence

Moravcův paradox

- Co je pro lidi relativně těžké (rozumové uvažování) je výpočetně nenáročné, co je pro lidi snadné (vnímání, motorika) je výpočetně velmi náročné.
- Obtížné problémy jsou snadné, snadné problémy jsou obtížné (řešit technologicky)
- Čtyřleté dítě: rozpozná obličej, zvedne tužku, naučí se samo chodit



Vysvětlení (jedno z možných): **Hans Moravec**: „Ve velké, vysoce vyvinuté smyslové a motorické části lidského mozku je zakódována miliarda let zkušeností o povaze světa a o tom, jak v něm přežít. Vědomý proces, který nazýváme uvažováním, je, myslím, nejtenčí povrchová vrstva lidského myšlení, účinná pouze proto, že je podporována touto mnohem starší a mnohem silnější, i když obvykle nevědomou, senzomotorickou znalostí. Všichni jsme hvězdní olympionici v percepční a motorické oblasti, tak dobří, že obtížné u nás vypadá jako snadné. Abstraktní myšlení je však nový trik, možná méně než 100 tisíc let starý. Ještě jsme ho nezvládli. Není sám o sobě tak složitý; jenom tak vypadá, když ho děláme.“

Aktuální obory:

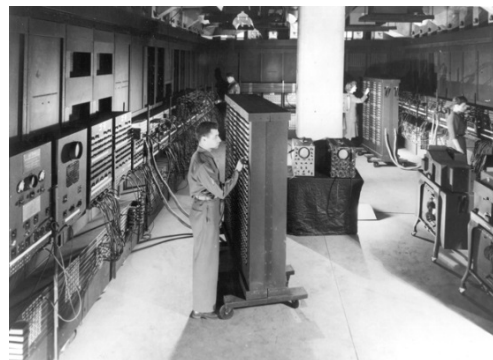
- reprezentace znalostí (knowledge representation)
- strojové učení (machine learning)
- zpracování přirozeného jazyka (natural language processing)
- plánování (planning)
- získávání znalostí (data mining)
- prohledávání stavového prostoru (state space search)
- stochastická optimalizace (stochastic optimization)
- řešení úloh s neúplným zadáním (problem solving)

Historie

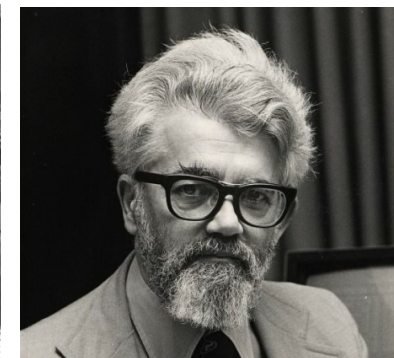
- Legendy, Talos, sci-fi (tam je všechno)
- 1956 workshop, John McCarthy – termín A.I. (aby se obor odlišil od kybernetiky)
- Počáteční optimismus (během 20 let budou stroje schopné dělat to co lidi)
- A.I. winter (vystřízlivění, cyklické opakování)
- 1965 expertní systémy, LISP (80-tá léta up, 90 léta down)



Talos © John Schroeter



ENIAC



John McCarthy (1927-2011)

- 1980 neuronové sítě – strojové učení s učitelem a bez učitele (supervised, unsupervised learning), optimismus, vystřízlivění
- 1996 počátky zpětnovazebního učení (reinforcement learning)
- 1997 DeepBlue poráží Gary Kasparova
- 2004 Darpa Grand Challenge (240 km trať, nikdo nedojel, rekord necelých 12km)
- 2005 Darpa Grand Challenge (212 km, 5 vozů dojelo, vítěz Stanley (Sebastian Thrun))



IBM 702



Stanley (2005)

- 2007 Darpa Urban Challenge (96 km v běžném provozu)
- 2010+ výrazný nárůst výpočetního výkonu a objemu zpracovaných dat, hluboké učení, konvoluční neuronové sítě (na GPU,TPU)
- 2011 IBM Watson vítězí ve vědomostní soutěži Jeopardy!
- 2012-2015 Darpa Robotics Challenge (robot zachránář)
- 2017 AlphaGo poráží Ke Jie, AlphaGo Zero poráží AlphaGo 100:0
- 2017 GPT (117M param.)



HUBO (Humanoid roBOt)



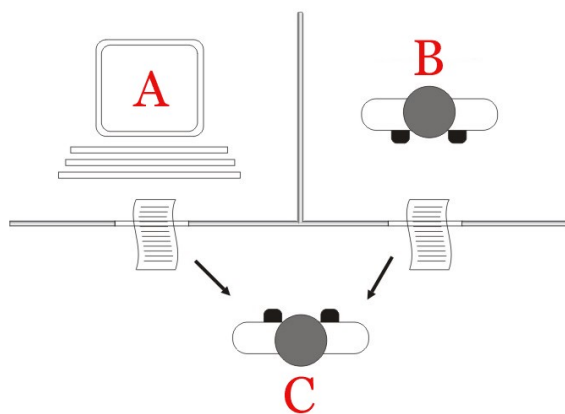
AlphaGo HW - TPU

Filozofie a A.I.

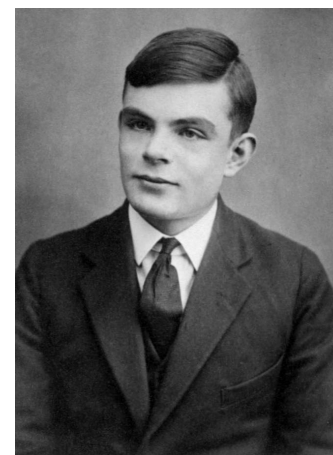
Základní otázky:

- Jaká je podstata inteligence? Může stroj plně nahradit uvažování lidské mysli?
- Je podstata počítače a lidského mozku stejná?
- Může mít stroj mysl, duševní stavy a vědomí, podobné těm lidským? Může stroj cítit?
Pokud ano, má mít práva?

Turingův test (Alan Turing, 1950)



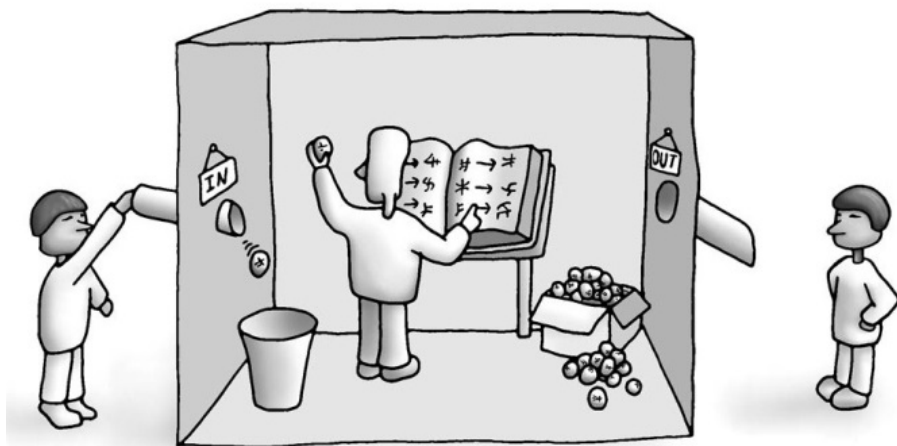
© obr: Juan Alberto Sánchez Margallo



Alan Turing 1912-1954

Pokud se stroj chová stejně inteligentně jako lidská bytost, pak je stejně inteligentní jako lidské bytosti.

Kritika: argument čínského pokoje (John Searl 1980)



John Searl (1932-)

Samotná schopnost smysluplně odpovídat na položené otázky není dostatečná pro prokázání schopnosti porozumění.

Turingův test netestuje, zda je stroj inteligentní. Testuje, zda se chová jako člověk. Což není totéž, člověk se může (a často to dělá) chovat neinteligentně.

Loebnerova cena (1991 – dosud)

- Platforma pro absolvování Turingova testu, postupné uvolňování podmínek (omezení tématu, délka konverzace).
- Mitsuku (vítěz 2016- 2019, Steve Worswick), dostupná na FB
- Vliv komplice (confederate effect) – rozhodčí označují skutečné lidi jako AI. Opak ELIZA jevu, kdy lidé přisuzují počítačům chování analogické lidem (antropomorfismus)
- Zneužití: CyberLover malware chatbot¹

¹ Pan Juin Yang Jonathan and Chun Che Fung: Artificial Intelligence in Malware - Cop or Culprit?
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.666.3540&rep=rep1&type=pdf>

Přátelská / nepřátelská AI

- Tři zákony robotiky (Isaac Asimov)
- Termín „přátelská AI“ (friendly AI) - Eliezer Yudkowsky, přátelská ve smyslu „bezpečná“ a „užitečná“
- Problém: neexistuje důvod, proč by měla umělá inteligence sdílet cíle a hodnoty lidí
- Problém 2: řízení fyzických zařízení pomocí počítačů (nejen autonomní auta, drony, ...)



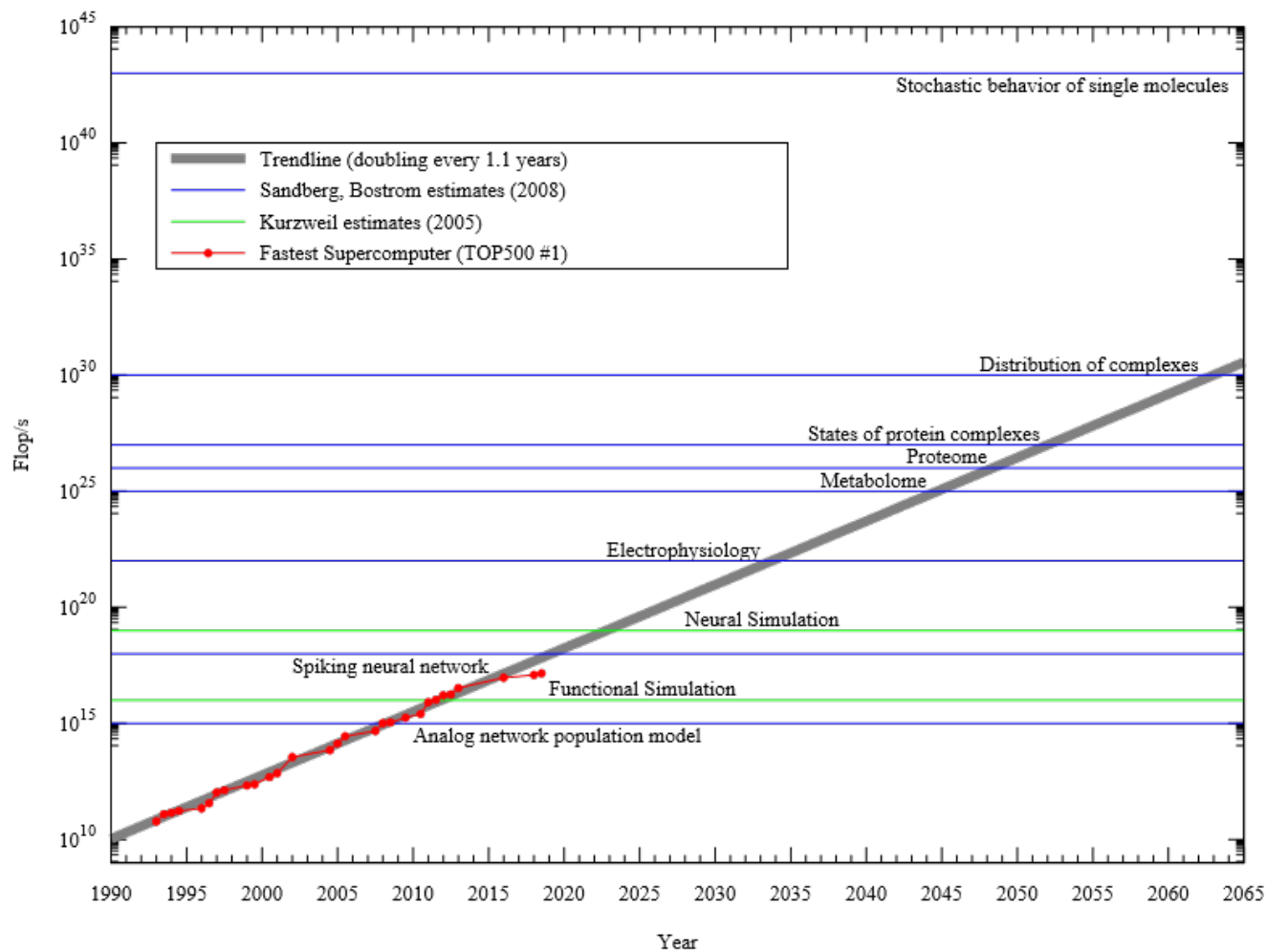
Elizer Yudkowsky

Doporučená literatura: Stuart J. Russell: Human Compatible: Artificial Intelligence and the Problem of Control (2019)

Výhled: obecná AI je daleko daleko daleko

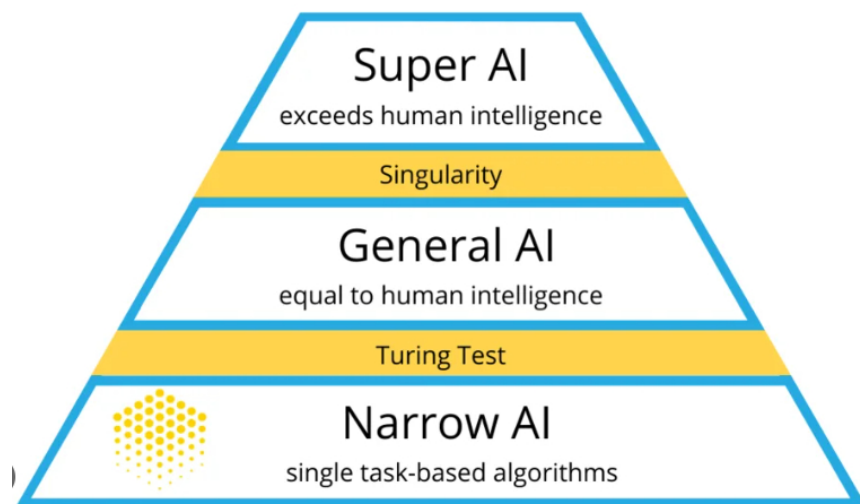
Obečná AI (AGI – Artificial General Intelligence)

- Simulace mozku: kdy bude na úrovni, která umožní vznik vědomí?
- Lidský mozek: 10^{11} neuronů, každý neuron průměrně 7000 synapsí, mozek má řádově 10^{15} synapsí.
- Pro jednoduchý model neuronu je již k dispozici odpovídající výpočetní síla (odhad 10^{14} změn synapsí / s, což je cca 10^{16} FLOPs, dosaženo 2012).
- ALE! OpenWorm projekt – komplexní model neuronu, 302 neuronů, výsledky žádná sláva (openworm.org).
- Souvisí s oblastí „mind uploading“, virtuální digitální nesmrtelností



Odhad kolik výpočetní síly je potřeba na emulaci lidského mozku na různé rozlišovací úrovni (Ray Kurzweil, Anders Sandberg, Nick Bostrom). Pozor, na ose Y je logaritmické měřítko.

Obecná AI



Specializovaná (narrow) AI

- Dobrá na specifické úlohy
- Nemá obecné kognitivní schopnosti

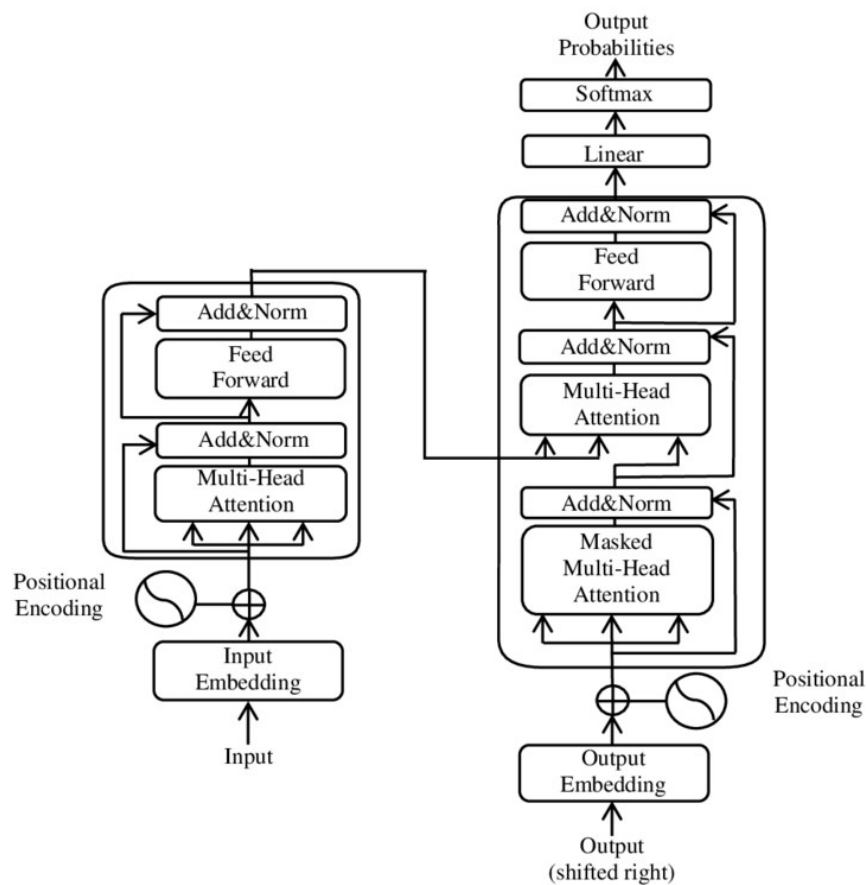
Obecná (strong, general) AI

- Zvládne libovolnou úlohu, kterou zvládne člověk
- Nemá vědomí

Sebastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, Yi Zhang: **Sparks of Artificial General Intelligence: Early experiments with GPT-4**, Microsoft Research, 2023

Velké jazykové modely (Large Language Models – LLM)

- Jazykový model založený na transformerech
- Vytváří další slova (sekvence) v závislosti na vstupní sekvenci
- Učení s učitelem
- Miliardy trénovatelných parametrů
- Aktuální modely: **GPT-4**, LLaMa, PaLM, BLOOM, Ernie 3.0 Titan, ...



Příklady kdy to funguje

Selský rozum (common sense)

- Netriviální úlohy – skládání objektů

Teorie mysli (porozumění lidem)

- Odraz mentálního stavu někoho jiného
- Odraz odrazu mentálního stavu...

Usuzování (reasoning)

- Hádanka s lovcem

GPT-4 vs ChatGPT

Prompt: Here we have a book, 9 eggs, a laptop, a bottle and a nail. Please tell me how to stack them onto each other in a stable manner.

GPT-4:

One possible way to stack the objects onto each other in a stable manner is:

- Place the book flat on a level surface, such as a table or a floor. The book will serve as the base of the stack and provide a large and sturdy support.
- Arrange the 9 eggs in a 3 by 3 square on top of the book, leaving some space between them. The eggs will form a second layer and distribute the weight evenly. Make sure the eggs are not cracked or broken, and handle them gently to avoid dropping them.
- Place the laptop on top of the eggs, with the screen facing down and the keyboard facing up. The laptop will fit snugly within the boundaries of the book and the eggs, and its flat and rigid surface will provide a stable platform for the next layer. Make sure the laptop is closed and turned off, and do not press any keys or buttons.
- Place the bottle on top of the laptop, with the cap facing up and the bottom facing down. The bottle will add some height and balance to the stack, and its cylindrical shape will reduce the contact area with the laptop. Make sure the bottle is empty and closed, and do not shake or tilt it.
- Place the nail on top of the bottle cap, with the pointy end facing up and the flat end facing down. The nail will be the final and smallest object in the stack, and its sharp tip will prevent it from rolling or sliding off the bottle. Make sure the nail is clean and straight, and do not touch or move it.

ChatGPT:

It would be difficult to stack all of these objects on top of each other in a stable manner without using any extra objects. The book and laptop are likely too large and heavy to stack on top of the eggs without damaging them. The bottle and nail may be too small to provide a stable base for the other objects.

Here is one possible way to stack the objects:

- Place the bottle on a flat surface.
- Carefully balance the nail on top of the bottle.
- Place the eggs on top of the nail, making sure they are balanced and not tilting to one side.
- Place the laptop on top of the eggs, again making sure it is balanced and not tilting to one side.
- Finally, place the book on top of the laptop.

This stack may not be very stable, so it is important to be careful when handling it. Alternatively, you could try stacking the objects in a different order or using some of the objects as supports for others to create a more stable stack.

Příklady kdy to funguje

Model jedné domény (jen text) – vygeneruje TikZ kód který zobrazí určitý objekt

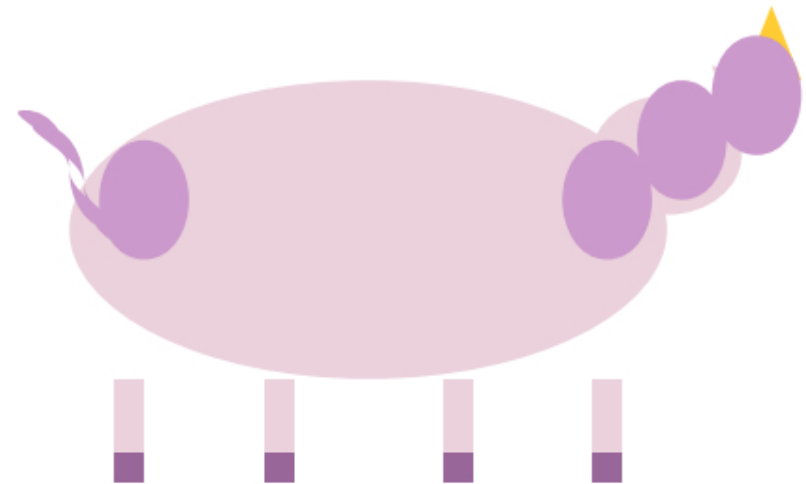
Prompt: Draw a unicorn in TikZ.

ChatGPT: [Produces \LaTeX compiling to following picture.]



Prompt: Draw a unicorn in TikZ.

GPT-4: [Produces \LaTeX compiling to following picture.]



Příklady kdy to nefunguje

Omezení daná autoregresivní architekturou (jednosměrnou)

Q: How many prime numbers are there between 150 and 250?

A: 13 (WRONG)

Q: First list the prime numbers between 150 and 250 and then output the size of this list

A: 151, ..., 241; the size of the list is 18 (CORRECT)

Q: $7 * 4 + 8 * 8 = ?$

A: 88 (WRONG)

Tak jak jsme blízko?

Znepokojivé klady

- Obrovský posun u „selského rozumu“
- Velmi pokročilá teorie mysli
- Generování vizuálních znaků popsaných pouze textově

Zápory plynoucí z daného paradigmatu

- Nedostatky v plánování
- Nedostatek pracovní paměti
- Neschopnost vracet se při řešení zpět
- Nízké uvažovací schopnosti

Problematické jsou úlohy, které nelze řešit sekvenčním způsobem

Rychlé myšlení – automatické, intuitivní, bez námahy

Pomalé myšlení – řízené, racionální, namáhavé

Lidské myšlení: směs rychlého a pomalého myšlení, pomalé myšlení „dozoruje“ používání rychlého myšlení jak subrutin.