AI Akademie

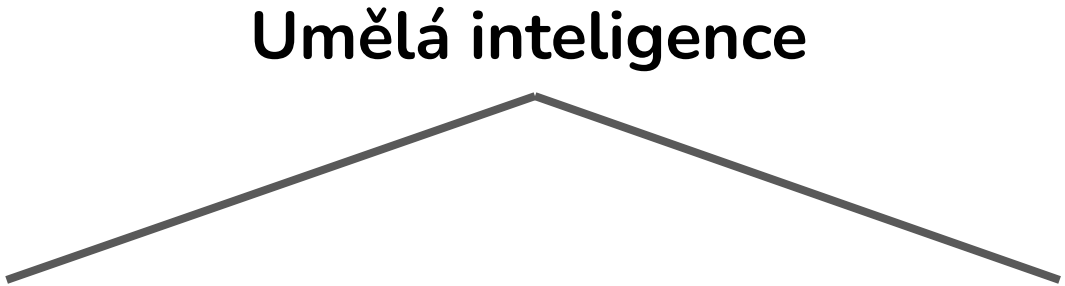
Kapitola 4: Filosofie umělé inteligence

Machine Learning College

prg.ai

PRAHA
PRAGUE
PRAGA
PRAG

Slabá a silná umělá inteligence



Slabá UI

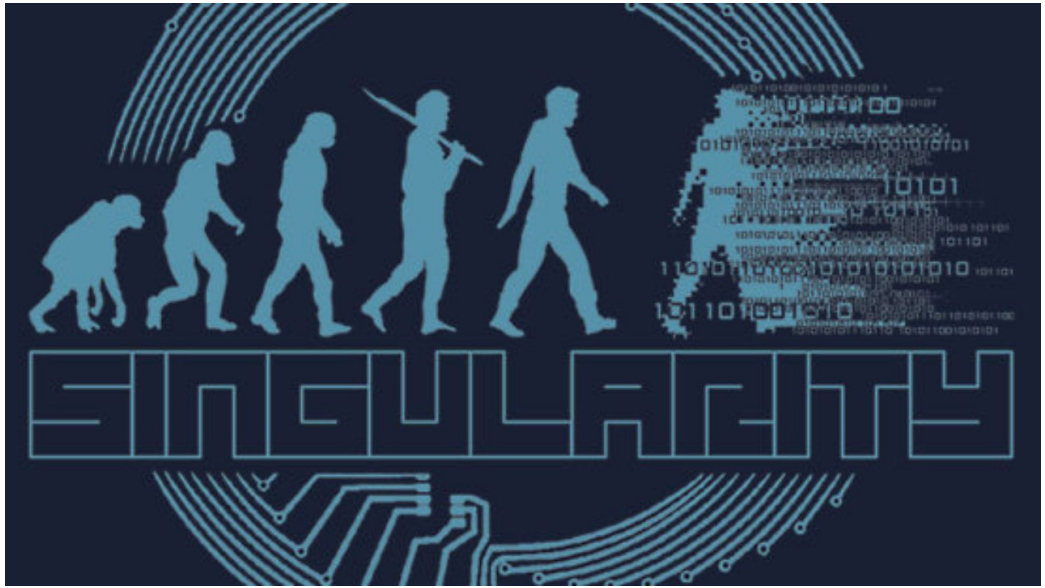
- vždy řeší jeden konkrétní problém, neumí se adaptovat na nové problémy
- všechna existující UI jsou slabou umělou inteligencí

Silná (obecná) UI

- umí cokoliv, co dokáže člověk, nebo dokonce více
- zatím neexistuje

Machine Learning College

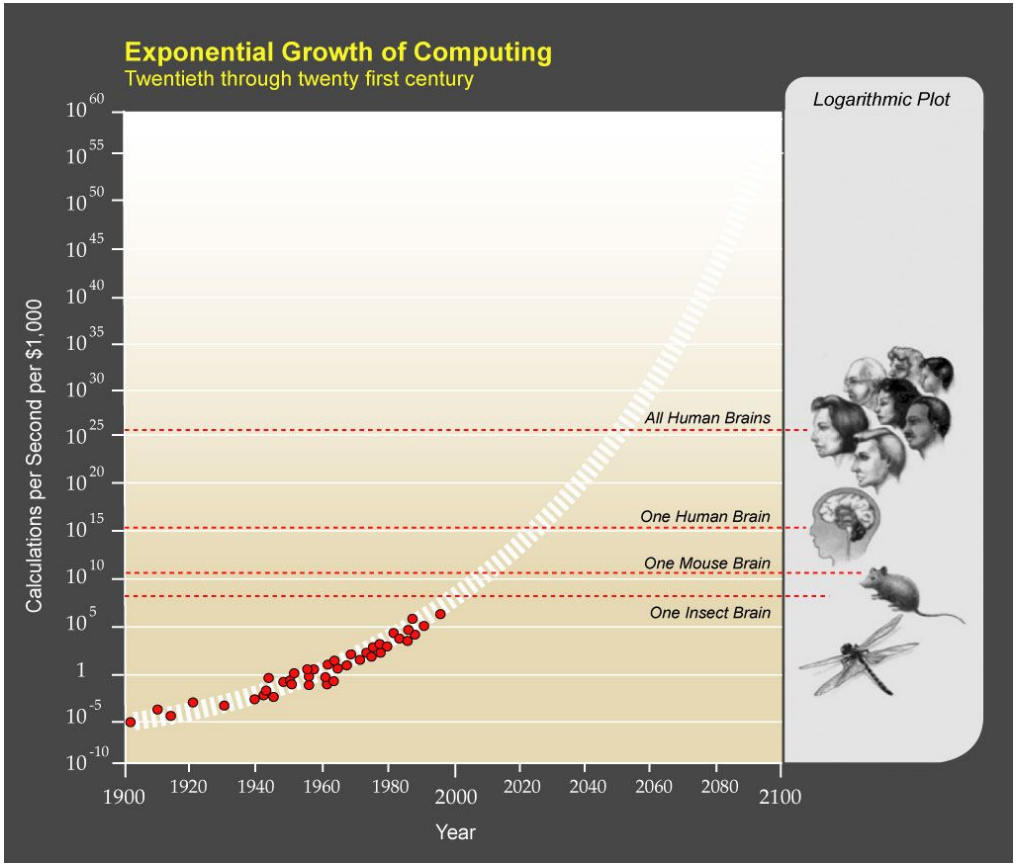
Technologická singularita



- Hypotetický stav, kdy umělá inteligence dosáhne schopností člověka
- Taková umělá inteligence by byla schopna sebe sama vylepšovat a dosáhnout tak *superinteligence*

Machine Learning College

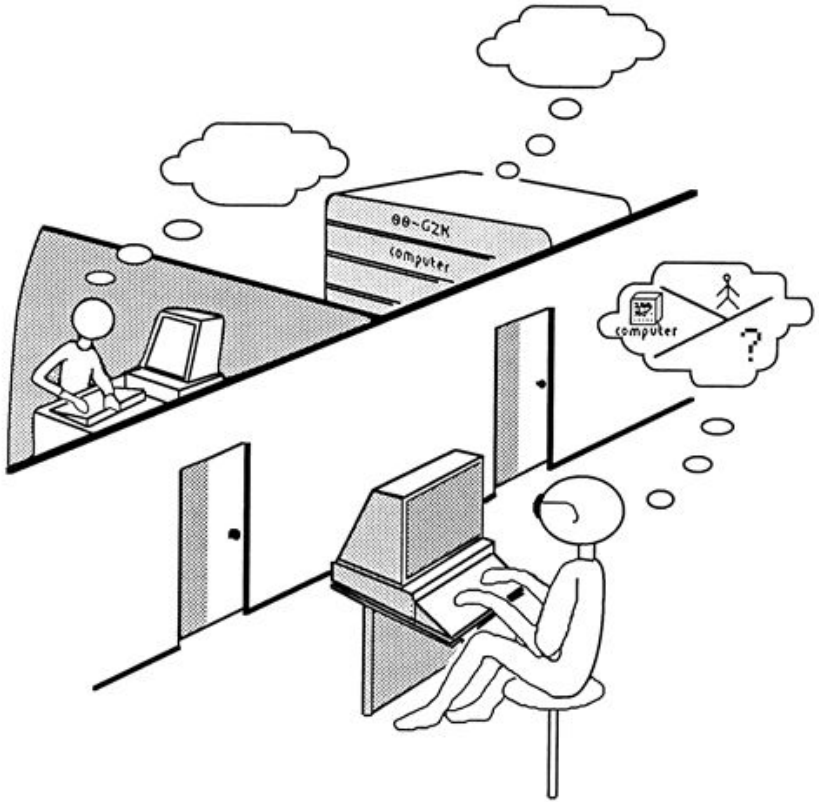
Ray Kurzweil - Singularita je blízko



- Kurzweil argumentuje, že při současném exponenciálním růstu výkonu dosáhneme singularity kolem roku 2030.
- Odpůrci namítají, že dosažení stejného výpočetního výkonu jako lidský mozek ještě neznamená dosažení stejné inteligence.

Machine Learning College

Turingův test umělé inteligence



- Test navržený pro ověření silné umělé inteligence
- Podle A. Turinga je postačující, nikoliv nutnou podmínkou

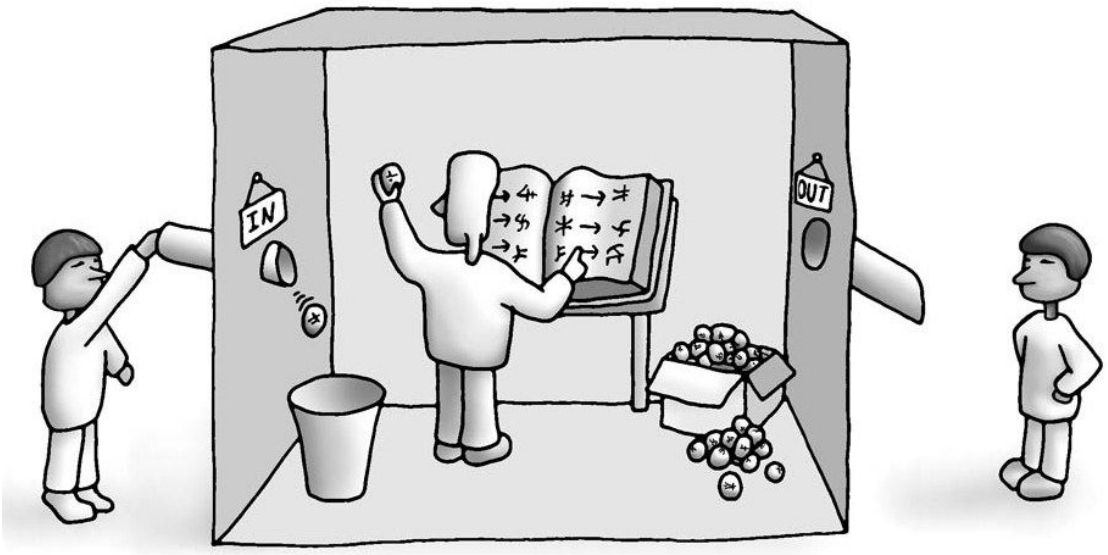
Machine Learning College

Turingův test - vlastnosti

- Navržený Alanem Turingem v roce 1950, stále uznávaný, i když existuje mnoho kritiků
- Zkoumá externí projevy inteligentního chování
- Je antropocentrický, ale nevylučuje existenci umělé inteligence, která test nesplní:
  - splnění Turingova testu ⇒ silnou AI
  - silná AI ⇏ splnění Turingova testu

Machine Learning College

Kritika T. testu - Argument čínského pokoje



- Člověk v místnosti pouze manipuluje symboly na základě pravidel, ale čínsky vůbec nemusí rozumět
- Autorem filosof John Searle
- Nejvýznamnější, ale zdaleka ne jediná kritika Turingova testu

Machine Learning College

Loebnerova cena

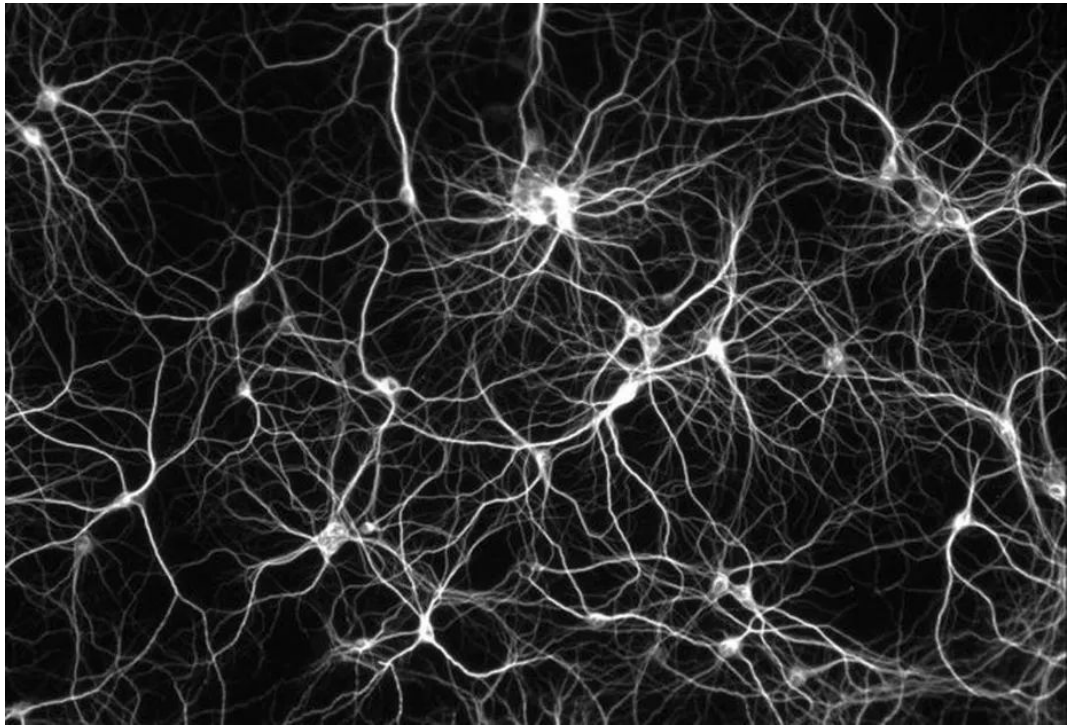
- V roce 1990 založil Hugh Loebner každoroční soutěž inteligentních systémů s cílem splnit Turingův test.
- Cena za splnění testu je $100.000. Zatím nebyla udělena.

Machine Learning College

Vyzkoušejte si Kuki chatbota (Mitsuku)

Machine Learning College

Rizika umělé inteligence - vysvětlitelnost



Moderní systémy umělé inteligence u některých úloh v průměru překonávají schopnosti lidí (například hraní her, klasifikace obrázků apod.). Je však obtížné nebo nemožné jejich chování zdůvodnit pro člověka srozumitelným způsobem.

Není tedy zatím možné použít umělou inteligenci v kritických oblastech jako např. k plně automatické diagnóze a léčení pacientů.

Machine Learning College

Rizika umělé inteligence - sociální bubliny

- Čím dál sofistikovanější doporučovací a personalizační systémy nás uzavírají do sociálních a informačních bublin.
- Je třeba si toho být vědomi a chovat se podle toho.

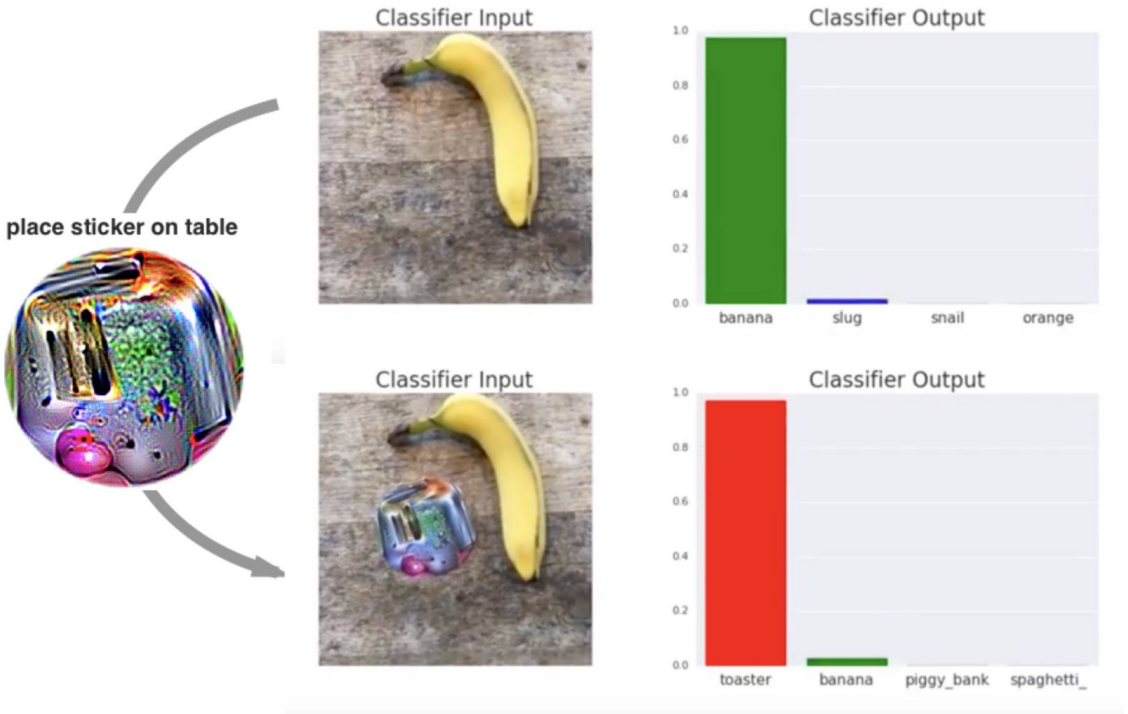
Machine Learning College

Rizika umělé inteligence - férovost a bias

- Systémy umělé inteligence jsou do velké míry odrazem použitých trénovací dat.
- Velmi snadno mohou např. diskriminovat skupiny lidí, které jsou v datech zastoupeny minimálně.
- V roce 2016 vytvořil Microsoft tweetujícího chatbota, kterého ale musel brzo po zveřejnění zastavit, protože se ze zpětné vazby od lidí naučil být rasistický a vulgární.

Machine Learning College

Rizika umělé inteligence - hacknutelnost



- Stejně jako u jiných počítačových technologií hrozí i u umělé inteligence možnost napadení nebo hacknutí.
- Pokud útočník dostane přístup k modelu umělé inteligence, je možné jej relativně snadno oklamat.
- Video zachycuje tzv. *adversarial patch*.

Machine Learning College

Rizika umělé inteligence - sebere nám práci?

- Zřejmě nastane podobná situace jako na začátku průmyslové revoluce.
- Řada nekvalifikovaných pracovních pozic zanikne, ale spousta nových vznikne.
- Klíčem k úspěchu ve společnosti budoucnosti je vzdělání a schopnost adaptace.

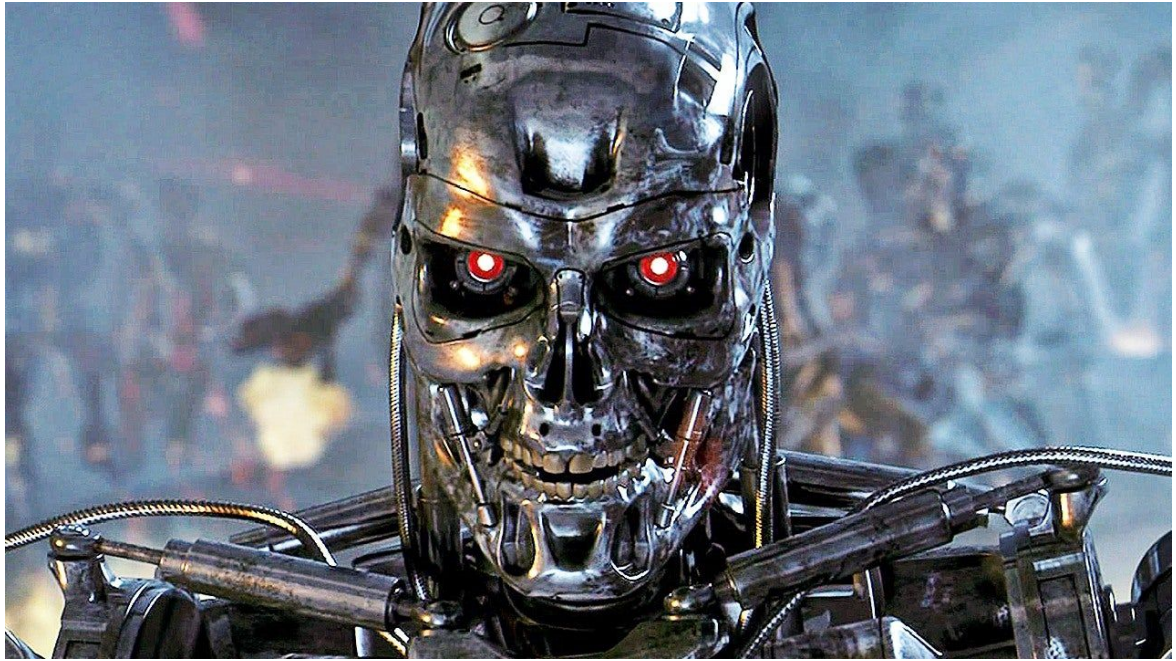
Machine Learning College

Rizika umělé inteligence - autonomní zbraně

- Autonomní zbraně jsou potenciálně velmi nebezpečné.
- Nepotřebují dosažení silné umělé inteligence.
- Existují celosvětové snahy o jejich zákaz nebo regulaci jako např. u chemických zbraní.

Machine Learning College

Rizika umělé inteligence - existenční riziko



- Obava o zachování lidského druhu v případě vytvoření silné umělé inteligence
- Nepravděpodobné, ale přesto seriózně zkoumané.

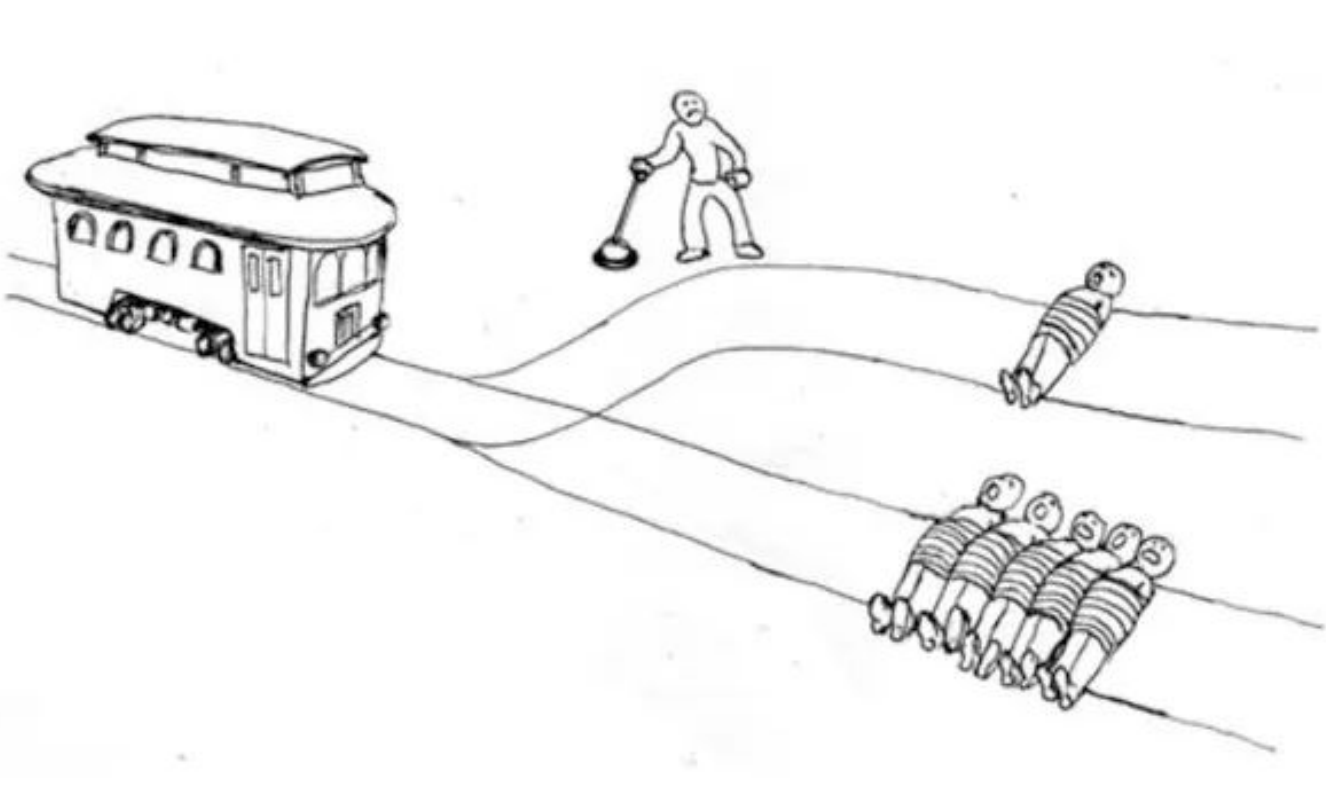
Machine Learning College

Umělé inteligence - přínosy převažují nad riziky

- Umělá inteligence je jako oheň - *dobrý sluha, ale zlý pán.*
- Nemá smysl se jí bránit, ale je dobré znát rizika.
- Je třeba v ní vidět velkou příležitost.
- Přínosy jednoznačně převažují nad riziky.

Machine Learning College

Morální otázky - Tramvajové dilema I

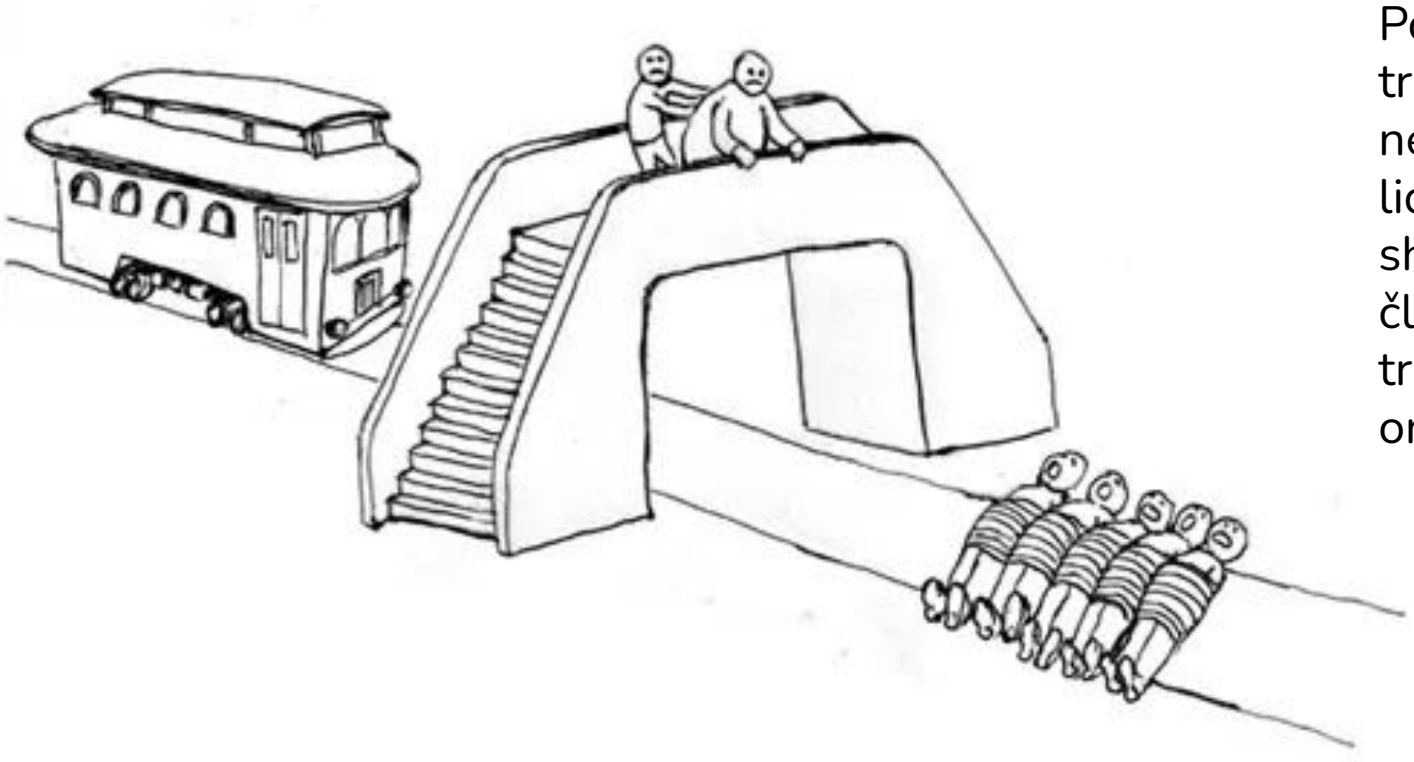


Po kolejích se řítí tramvaj a pokud nic neuděláte, zabije 5 lidí. Máte možnost přehodit výhybku tak, že tramvaj zabije pouze jednoho člověka.

Co uděláte?

Machine Learning College

Morální otázky - Tramvajové dilema II

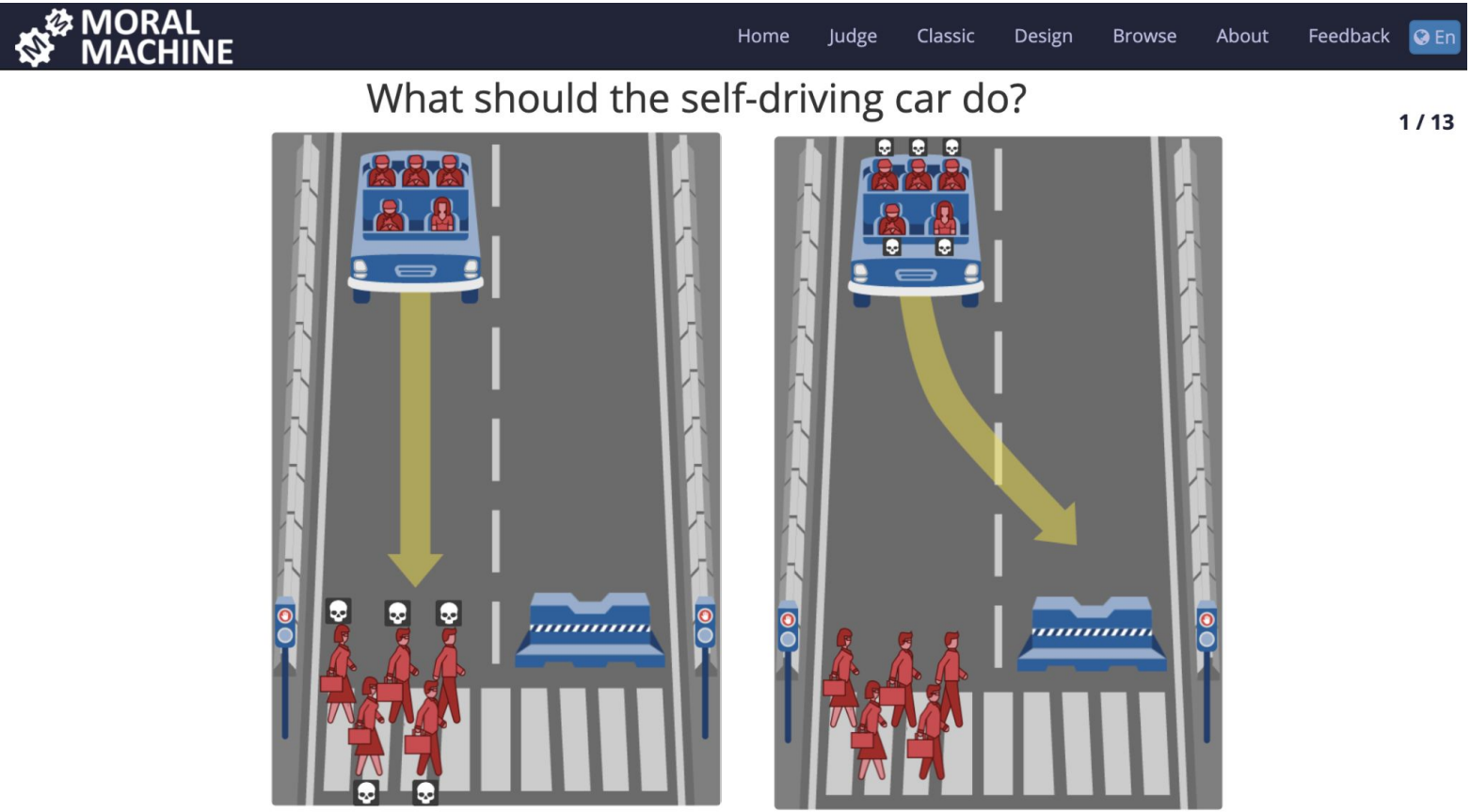


Po kolejích se řítí tramvaj a pokud nic neuděláte, zabije 5 lidí. Máte možnost shodit z mostu člověka, jehož tělo tramvaj vykolejí, ale on zemře.

Co uděláte?

Machine Learning College

Morální otázky - MoralMachine.net



Machine Learning College

Otázky k diskuzi

1. Uveďte příklady oblastí, kde vnímáte umělou inteligenci jako extrémně přínosnou.

2. Uveďte příklady oblastí, kde může být využití umělé inteligence rizikem.

3. Diskutujte o morálních otázkách, spojených s umělou inteligencí.

Machine Learning College