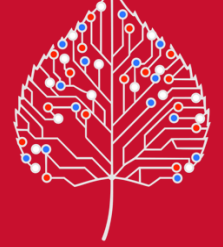


bohem.ai

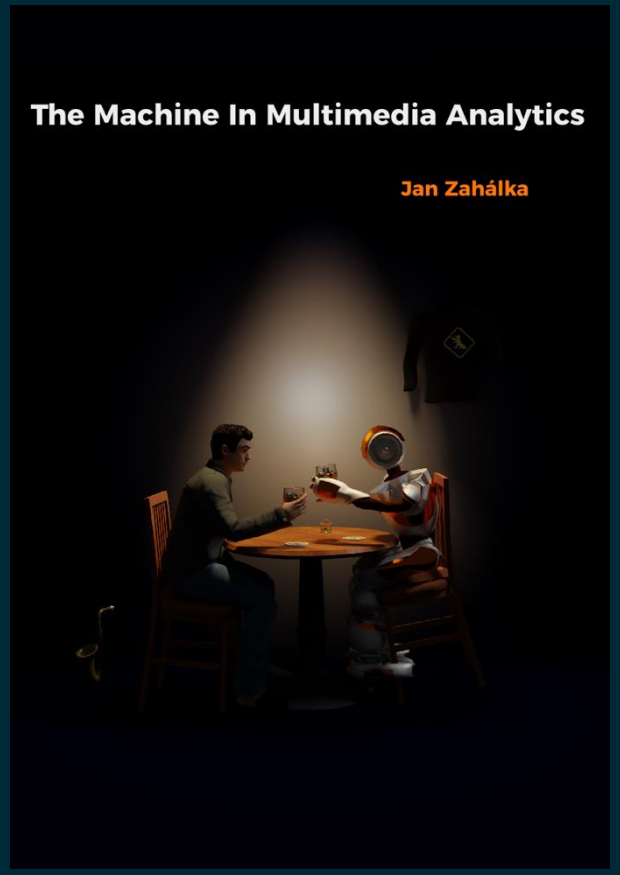
GAPP
SYSTEM

AI/ML
a jejich reálné využití

Jan Zahálka

Představení

- **AI/ML expert s 15+ lety praxe**
- Vzdělání:
  - **dr.** v oboru počítačových věd, University of Amsterdam
  - **Ing.** v oboru umělá inteligence, FEL ČVUT
- S vědou a výzkumem špičkových AI technologií jsem spjat dodnes
  - Působím na **Českém institutu informatiky, robotiky a kybernetiky** (CIIRC) ČVUT



Představení

- 2017 – založení firmy **bohem.ai**
  - AI/ML konzultantství & vývoj
- **Realizované projekty**:
  - AI/ML strategie firmy
  - Doporučovací systémy
  - Odborný chatbot
  - Extrakce strukturovaných dat z dokumentů
  - Zajištění vzdělávání

Klienti bohem.ai:

TOLION Health

CEPS

CASEC

UNIVERSITEIT VAN AMSTERDAM

Představení

- Jaro 2024 – **GAPP + bohem.ai**
- **GAPP** – 30 let na trhu, spolehlivá velká infrastrukturní a cloudová řešení, silné business týmy
- **bohem.ai** – 7 let na trhu, AI konzultanství a vývoj
- **Společně** jsme schopni zajistit špičkovou produkční business AI

Osnova

1. Co je AI/ML?
2. Proč by nás AI a ML měly v businessu zajímat?
3. MLOps: Jak AI/ML do businessu integrovat a na co myslet?

bohem.ai

GAPP SYSTEM

Co je AI/ML?

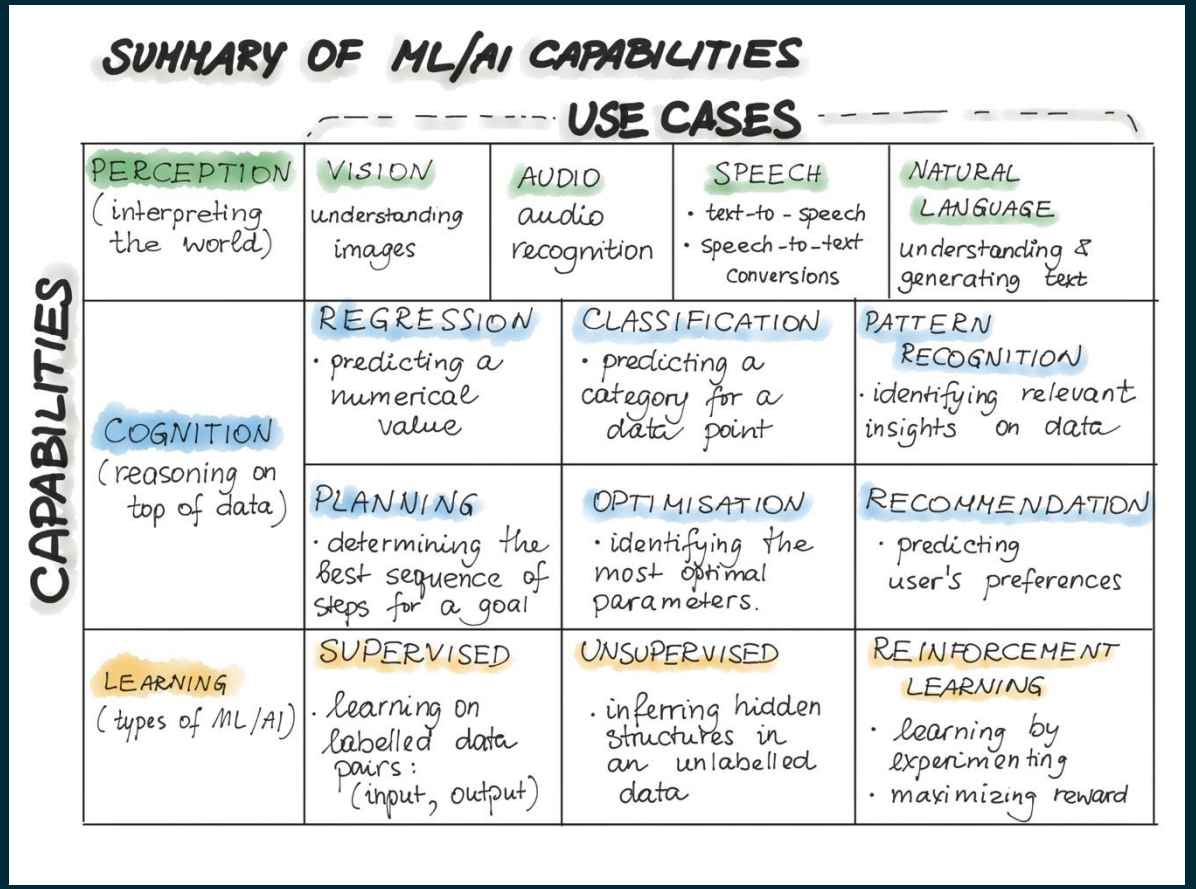
Umělá inteligence (AI)

- **AI = umělá inteligence** (artificial intelligence)
  - Široký obor zahrnující obecně *"inteligenci strojů"*
  - Schopnost řešit úlohy vlastním "mozkem", nikoliv jen dopředu přesně naprogramovaným předpisem
  - Člověk tedy programuje "předpis pro učení", ne řešení problému

Strojové učení (ML)

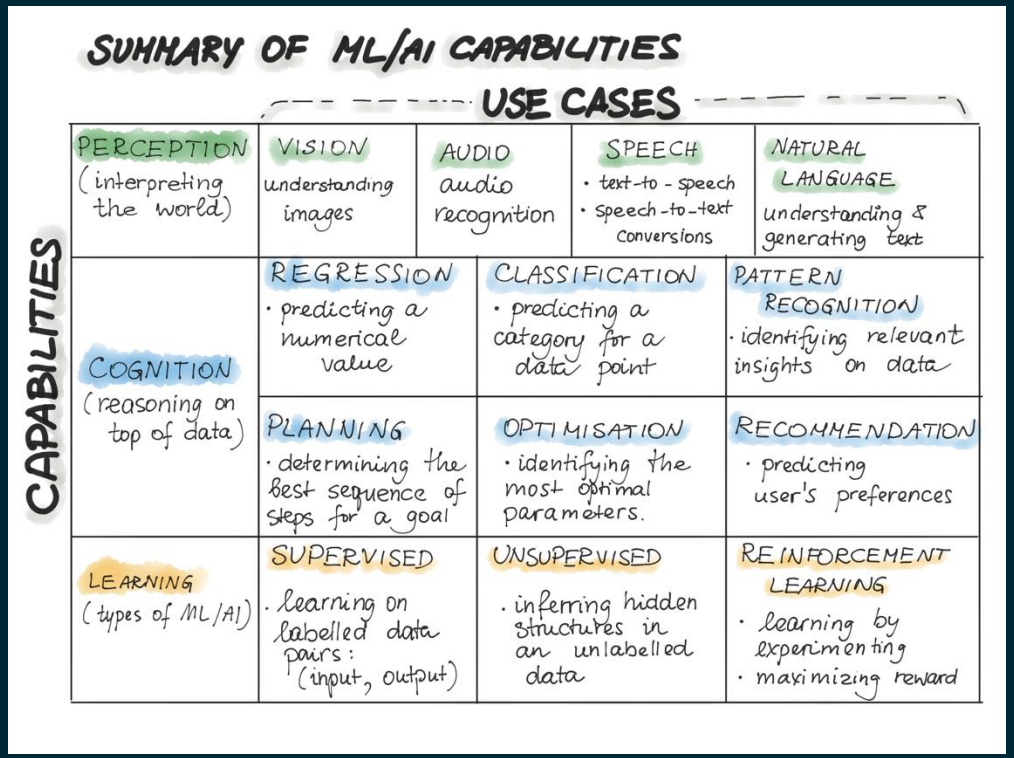
- **ML = strojové učení** (machine learning), poddisciplína AI
  - Schopnost stroje naučit se úlohu na základě **poskytnutých dat**
  - Světem businessu hýbe v rámci AI převážně ML
- Na tomto workshopu se zaměřujeme na **ML**, ale v rámci AI jsou i jiné přístupy
  - Symbolic AI, knowledge representation, velká část robotiky…

Úlohy strojového učení (ML)



Úlohy strojového učení (ML)

- **Mapování na business úlohy**
  - Automatická klasifikace, kategorizace
  - Doporučení
  - Predikce hodnoty/trendu
  - Extrakce strukturovaných dat
  - Sumarizace
  - Tvorba obsahu
  - Odpovídání na otázky, konverzace
  - Detekce anomálií/podvodů
  - …



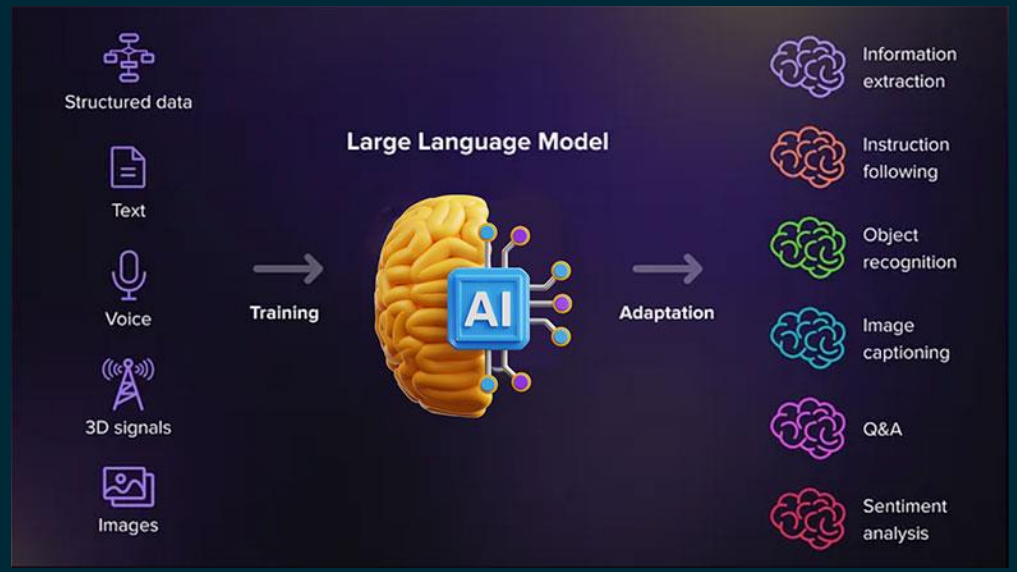
Příklady metod strojového učení (ML)

- **Velké jazykové modely (LLM)**, slavným příkladem ChatGPT
- **Deep learning** – hluboké neuronové sítě
- **XGBoost** – predikce časových řad

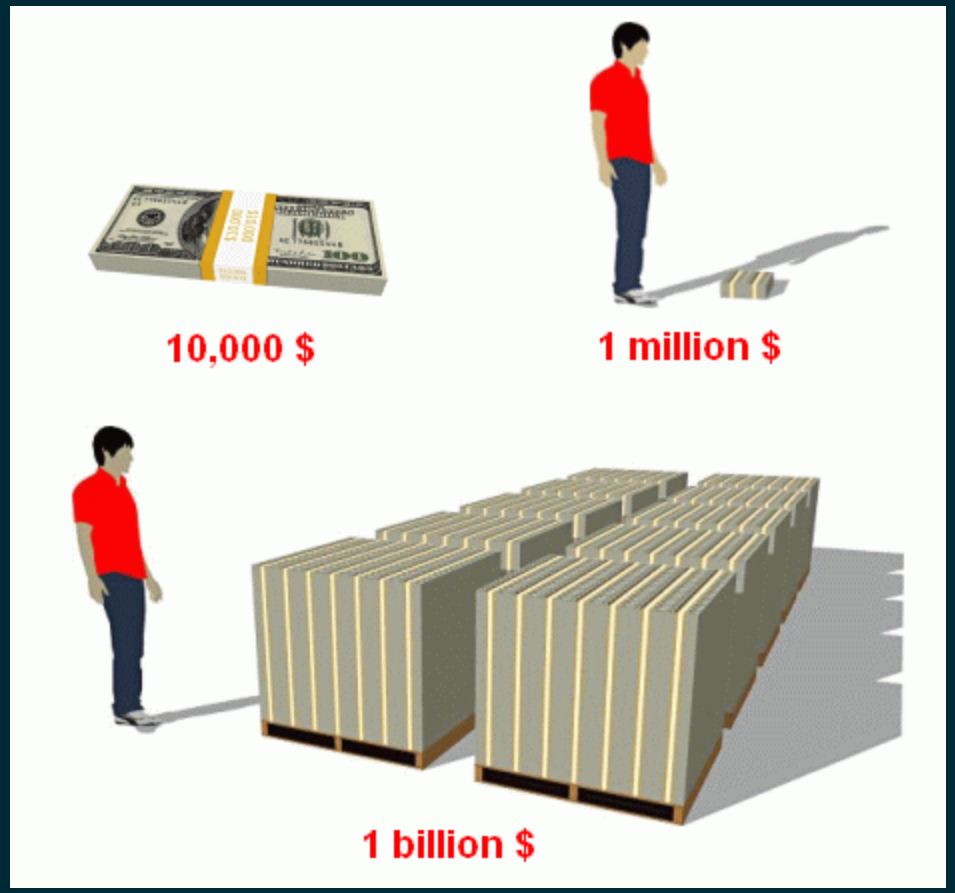
- … a mnoho, mnoho dalších, toto jsou jen reprezentativní příklady

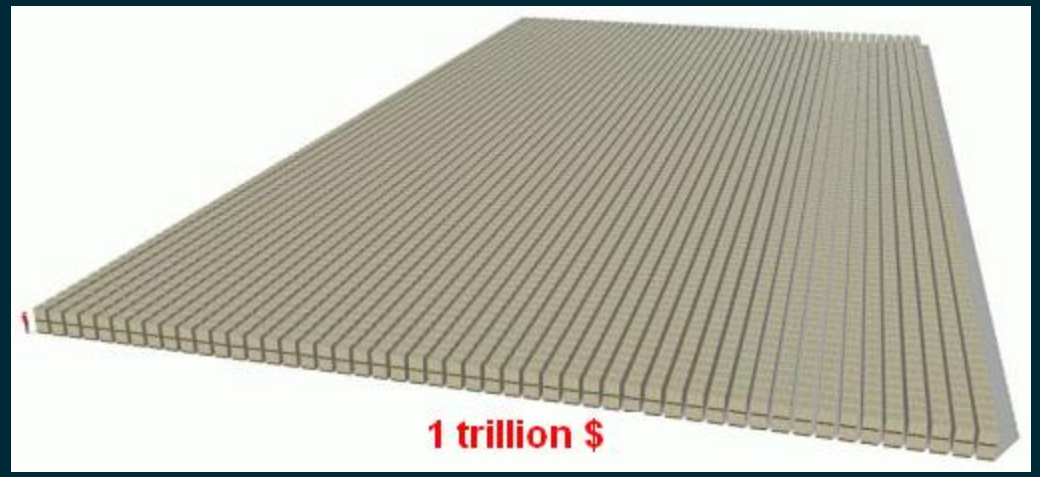
Velké jazykové modely (LLM)

- Zjednodušeně: **“chatbot”**, který s námi srozumitelně komunikuje
  - Přirozeným jazykem
  - Strukturovaným formátem (např. JSON)
- Trénován na **obrovských datech**, “zná všechno”
- Sám **model je obrovský**
  - GPT-4: 1,8 biliónu parametrů



Malá odbočka: vizualizace biliónu





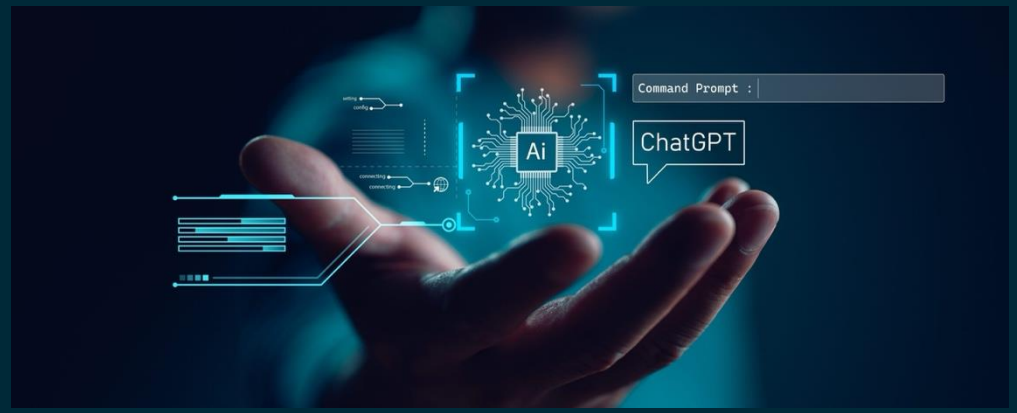
(billion = miliarda, trillion = bilión)

Velké jazykové modely (LLM)

- Skvěle řeší
  - **Extrakci strukturovaných dat** z dokumentů v libovolném výstupním formátu
  - **Tvorbu obsahu**
  - **Odpovídání na otázky**
  - **Sémantickou klasifikaci** na základě významu textu
  - **Sumarizaci** textu
  - … a mnoho dalších úloh

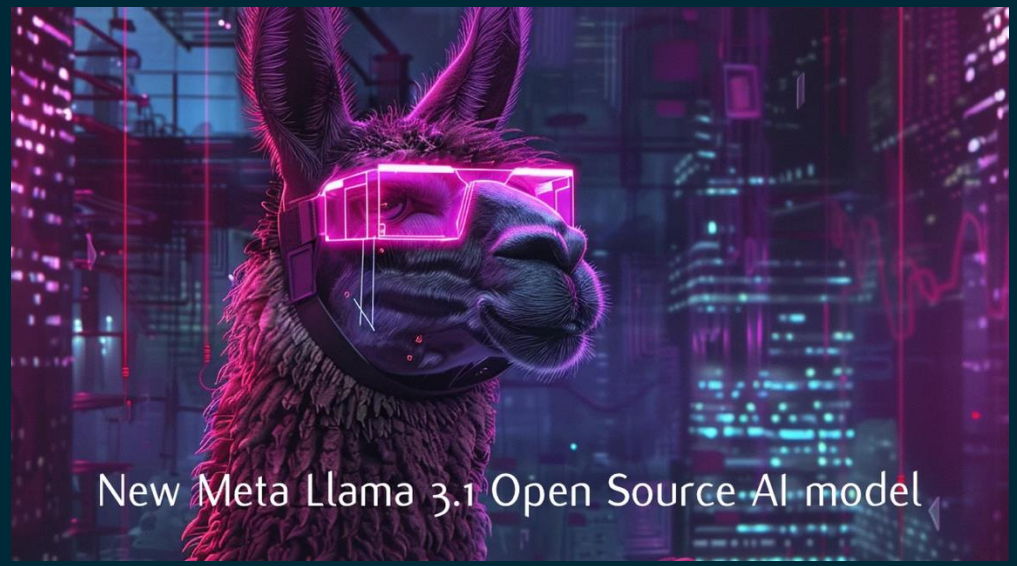
Velké jazykové modely (LLM)

- Nejslavnějším zástupcem **ChatGPT**
- **Špičkový výkon**, neustálé další rapidní zlepšování
- **Dostupné přes API** pro business účely
  - 1 dokument o 3000 slovech vyjde na cca 0.01 USD
- Skvělá volba pro **PoC většiny vhodných business ML úloh**



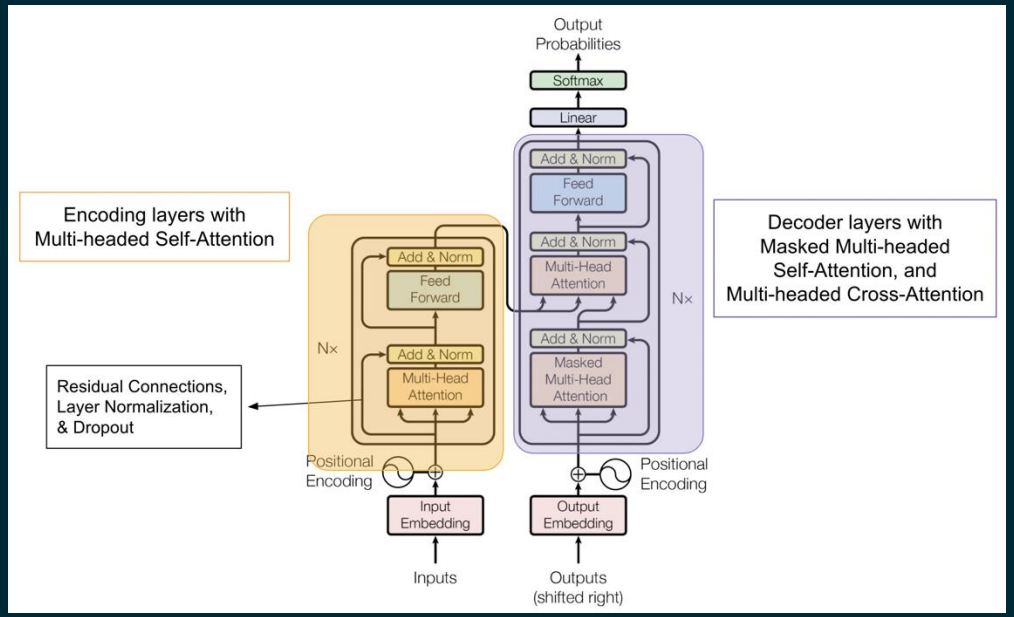
Velké jazykové modely (LLM)

- On-prem produkční řešení: **proprietární fine-tuned LLM**
- Zvolíme **dobrý open-source LLM**, např. Llama 3.1 405B
- **Dotrénujeme jej** na našich datech na konkrétní úlohu
- Výhody:
  - **100% vlastnictví** modelu
  - **Maximální bezpečnost** – žádná data nejdou OpenAI
- Nutno **pořídit odpovídající serverovou infrastrukturu**
  - „405B" = 405 miliard parametrů



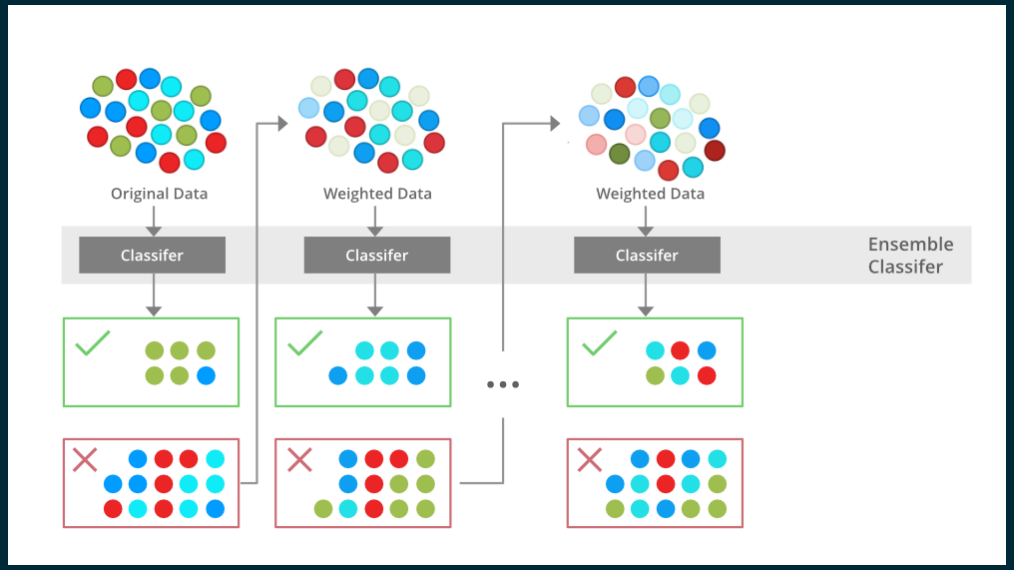
Deep learning

- **Hluboké neuronové sítě**, které inteligentně mapují datové vstupy na požadované výstupy
- **Hluboké** -> dostatečně inteligentní a přesné, ale zároveň hladové po trénovacích datech
- Z médií známy spíše **hi-tech aplikace** jako autonomní vozidla nebo medicínské rozpoznávání, vhodně řeší i „běžné" business potřeby, např.:
  - Klasifikaci
  - Predikci
  - Detekci anomálií/podvodů
  - … a mnoho dalších úloh

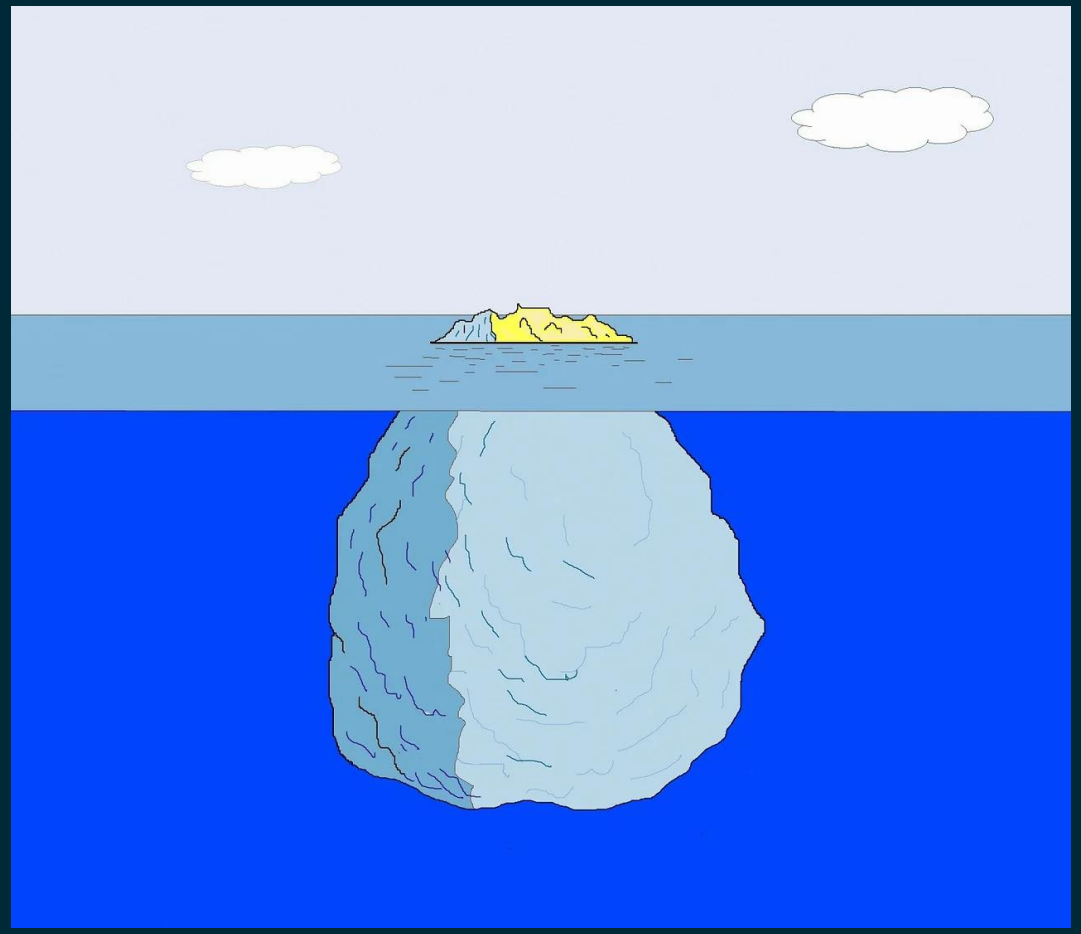


XGBoost

- E**X**treme **G**radient **Boost**ing - místo jednoho obřího modelu kolekce menších jednoduchých rozhodovacích stromů
- Příklad staršího ML přístupu, který ale na řadu úloh funguje skvěle s minimálními HW nároky
  - LLM: síť high-end serverů
  - XGBoost: obyčejný počítač
- Ne vždy je nejdražší řešení nejlepší
- Dodnes skvěle řeší např. **predikci časových řad**



...a to není v ML ani zdaleka vše



bohem.ai

GAPP SYSTEM

Proč by nás AI a ML měly zajímat?

Proč AI/ML?

- Z předchozích slides víme přehledově, co to je… ale **k čemu nám to je?**
- O AI/ML se **hodně mluví**, ale **co reálně znamená pro business?**
- **A je to vůbec k něčemu?** V médiích jdou pořád zprávy, jak to dělá chyby…
- Není to jen **hype**, módní vlna, která přejde?

Výhody AI/ML

- Dobře navržené AI řešení je schopno **kvalitně řešit určité business úlohy na úrovni juniora**
- Rapidní rozvoj: před rokem bychom řekli spíše **šikovného gymnazisty**

Výhody AI/ML

- Umí nad daty **uvažovat**, ne z nich jen počítat či je jen přeformulovávat
- AI řešení nejen spočítá statistické veličiny, ale také **připraví analytické podklady** či i **samo rozhodne**
- Komunikuje srozumitelně **s lidmi i stroji**
  - S lidmi přirozeným jazykem
  - Se stroji ve validovaných strukturovaných formátech, které každá IT infrastruktura preferuje, dle libosti

Výhody AI/ML

- **Neunaví se**
- Miliontý dokument bude zpracován stejně kvalitně, jako ten první
- K dispozici 24/7
- Na AI/ML nejsou zaměstnanecké náklady

Výhody AI/ML

- Toto vše a další **ušetří hodně peněz**
- **Autonomii AI/ML** lze nastavit podle našich potřeb
  - AI/ML může člověka u určitých úloh nahradit…
  - … a nebo mu významně **pomáhat** tím, že udělá pořádný kus práce za něj
- Příklad: **vyčítání dat z faktur**
  - **Výchozí stav**: 2 zaměstnanci na plný úvazek, kteří udělají 250 faktur denně
  - **AI-optimizovaný stav**: AI udělá 500 faktur denně, připraví přehled pro kontrolu, kterou udělá jeden zaměstnanec v pátek odpoledne
  - Z **2 FTE na 0,1 FTE**

AI/ML: chyby a halucinace

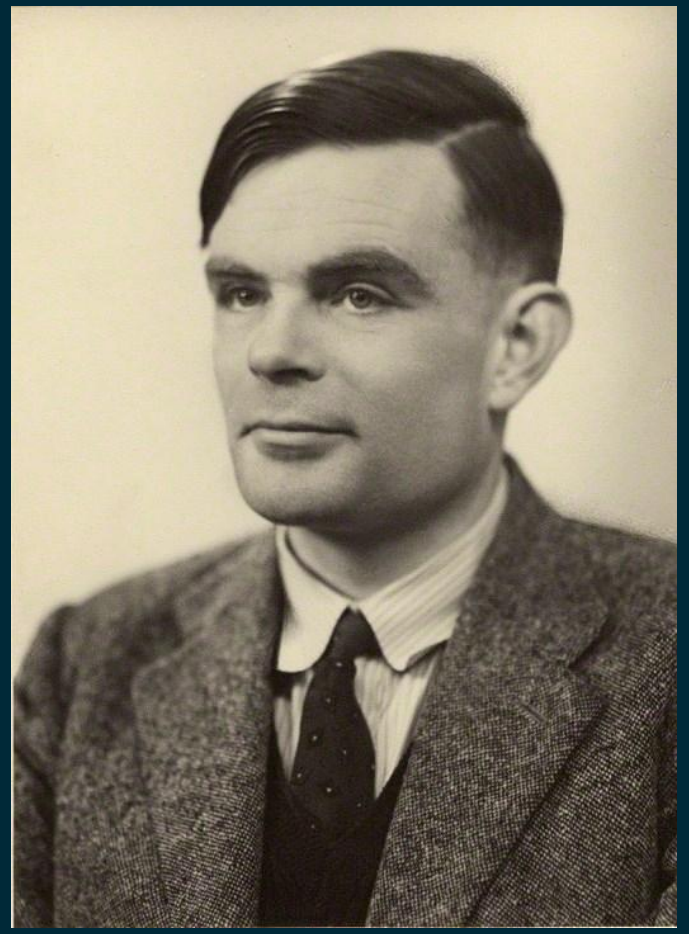
- Médii jdou zprávy, že AI/ML jsou zatím ještě **moc hloupé a chybující**
  - "Právník v USA dal LLM zpracovat podklady a následně prohrál případ, protože citoval neexistující precedens..."
  - "Na tomto příkladu vidíte, jak lze LLM donutit k celkem libovolnému blábolu..."
  - ... a mnoho jiných

AI/ML: chyby a halucinace

"If a machine is expected to be infallible, it cannot also be intelligent."

- Alan Turing, britský matematik a počítačový vědec, zakladatel AI

Neplatí to náhodou i pro lidi?



AI/ML: chyby a halucinace

- **Jak lidé, tak stroje jsou chybující** a je to tak z podstaty inteligence
  - Lidé dospívají 18+ let
- Moderní AI/ML modely dosahují pro rozličné úlohy přesnosti **přes 90%**, je-li na to budget, dá se jít stále výše...
- **...podobně jako lidi!**
  - Zkušenost z praxe: ještě se mi nestalo, aby klient dodal 100% přesná vzorová řešení úloh, které má řešit AI/ML
  - A je to tak v pořádku: chyby se u netriviálních úloh prostě dějí

AI/ML: chyby a halucinace

- Stejně jako lze **nacenit chybu člověka**, lze nacenit také **chybu AI**
- Chybám AI se dá předcházet **vícevrstevnou automatickou kontrolou**
  - Klasická **validace výstupů**
  - **AI validátory** – "jiné AI kontroluje, jestli jsou výstupy správné"
- Bonus: **AI se neunaví** a **nedělá překlepy**

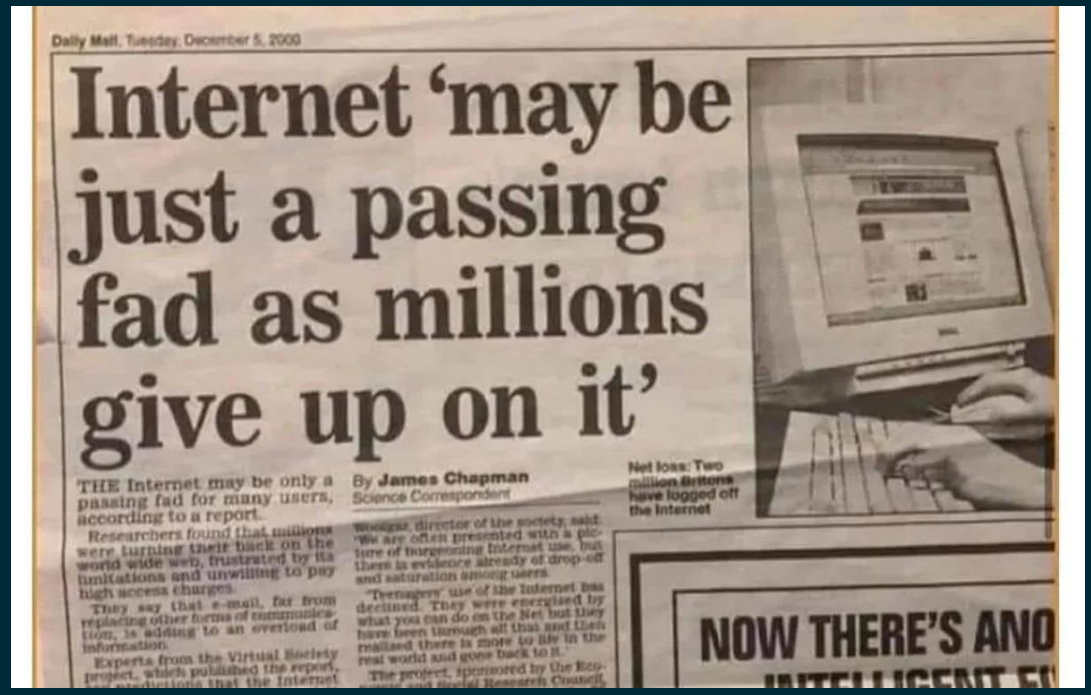
AI/ML: chyby, halucinace

- Příklady mediálních zpráv...
  - “Právník v USA dal LLM zpracovat podklady a následně prohrál případ, protože citoval neexistující precedens...”
  - “Na tomto příkladu vidíte, jak lze LLM donutit k celkem libovolnému blábolu...”
  - ... a mnoho jiných

- **Chyba AI**, nebo **špatné použití**?
  - Právník byl lenoch, LLM netrénované na právničinu nechal, ať ho nahradí, místo aby ho využil jako asistenta
  - Ano, pokud chcete AI rozbít, tak ji rozbijete – v první řadě plní vaše přání, chcete-li chybné výstupy...

AI/ML: jen módní vlna?

- O AI/ML mainstream hovoří jen posledních pár let…
  - … ale vznikly v **50. letech 20. stol.**
- Desítky let samozřejmě strávily ve fázi prototypů…
  - … ale už **minimálně 20 let** jsou v praxi úspěšně používány
  - AI/ML přežily několik "zim", ba "smrtí"



Daily Mail, Tuesday, December 5, 2000

Internet 'may be just a passing fad as millions give up on it'

By James Chapman
Science Correspondent

THE Internet may be only a passing fad for many users, according to a report.

Researchers found that millions were turning their back on the world wide web, frustrated by its limitations and unwilling to pay high access charges.

They say that e-mail, far from replacing other forms of communication, is adding to an overload of information.

Experts from the Virtual Society project, which published the report,

Net loss: Two million Britons have logged off the Internet

NOW THERE'S ANO

Daily Mail, 2000

Pár příkladů z mé dřívější kariéry

- 2013 – nalezení nových forem **genetické exprese MDS** (formy **leukémie**)
- 2014 – **New Yorker Melange**, systém doporučující návštěvníkům NY zajímavá místa
  - Vítěz **ACM Multimedia Grand Challenge** financované IBM
- 2018 – **VOXPol Video Analytics**, videoanalytický systém pro potlačování násilné propagandy online politického extremismu

AI/ML: jen módní vlna?

- **Ne.**
- Moderní AI/ML modely už jsou **příliš dobré na to**, aby jen tak zmizely
- Stále jsme v relativně **rané fázi adopce**
- AI/ML určitě čekají výzvy – např. spotřeba elektrické energie – ale je nepravděpodobné, že by na ně zahynula
  - Nutnost sofistikovaných, efektivních řešení

AI/ML v businessu: sečteno podtrženo

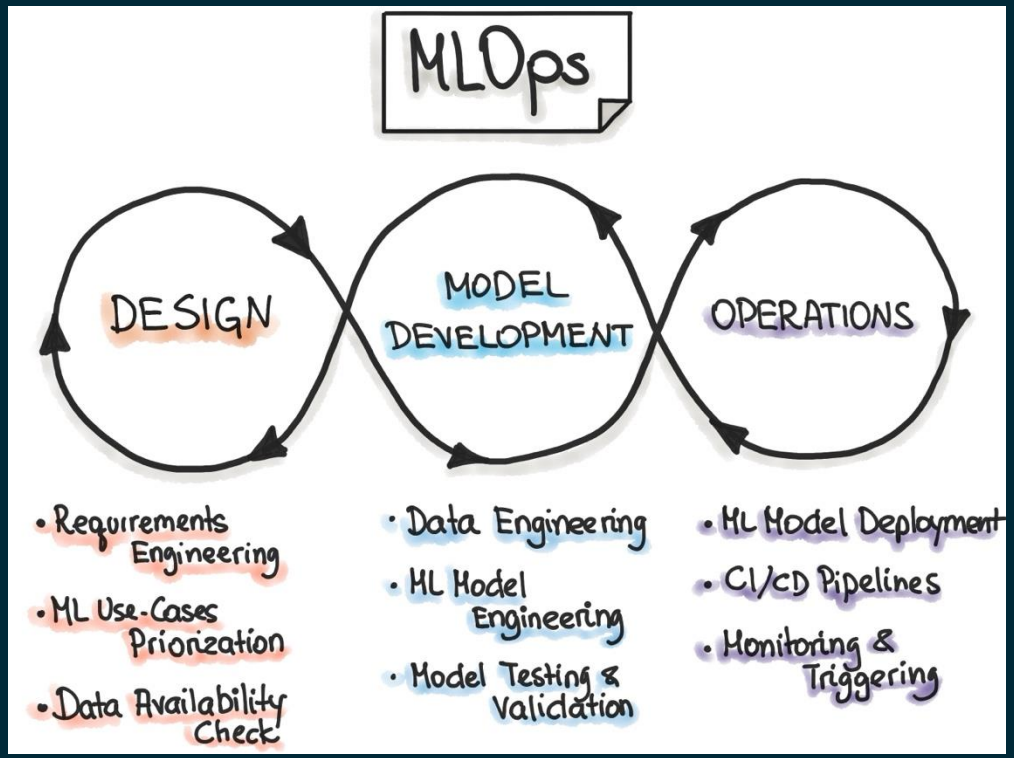
- AI má spoustu výhod, pro business ta snad nejdůležitější: **šetří peníze**
  - Nemusíme platit lidi na repetitivní, mechanickou práci
  - Z odborníků děláme díky inteligentní asistenci “superodborníky”
- Dobrá AI nedělá víc chyb než člověk a jejich dopadům se dá **předcházet**
- AI tu **zůstane**, není to jen hype

- V čím dál větším množství businessů **AI úspora > náklady na chyby**
- Už to nejsou jen prototypy užitečné úzké množině high-tech institucí


bohem.ai
GAPP
SYSTEM
MLOps: Integrace
AI/ML do businessu

MLOps: Motivace

- Dosud jsme se bavili pouze o schopnostech AI/ML řešit zadané úlohy
- Mít jen **samostatně stojící ML model** ale ještě nestačí - byť by byl **sebelepší**
- V této části přednášky se podíváme na **MLOps**: integraci celého lifecycle ML s business procesy firmy
  - Název vychází ze zavedeného **DevOps**: metodologii integrace vývoje SW (“Dev”) s procesy firmy (“Ops”)
  - **MLOps = klíčová ingredience ML úspěchu!**



MLOps: Kroky

1. AI strategie
2. Zajištění hardwarové infrastruktury
3. Návrh & učení ML modelů
4. Návrh & implementace SW řešení
5. Integrace řešení do business procesů
6. Dlouhodobá podpora

MLOps: Kroky

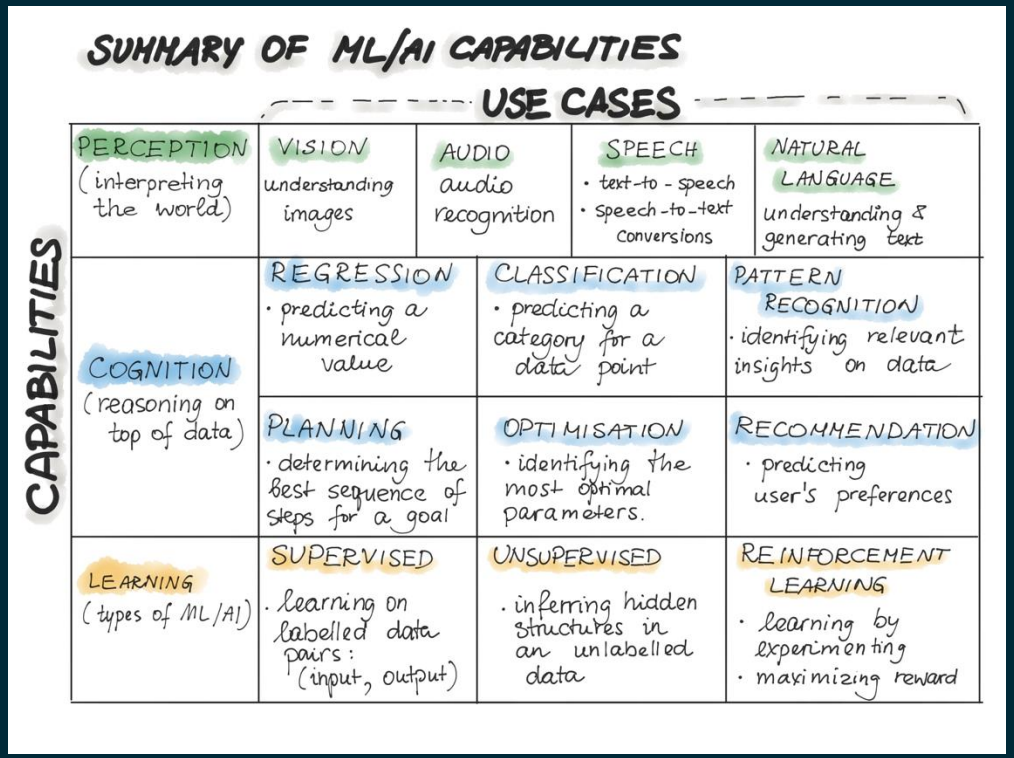
1. **AI strategie**
2. Zajištění hardwarové infrastruktury
3. Návrh & učení ML modelů
4. Návrh & implementace SW řešení
5. Integrace řešení do business procesů
6. Dlouhodobá podpora

AI strategie

- Specifikace řešených AI úloh
- Datový návrh
- Návrh HW infrastruktury
- Návrh SW tech stacku
- Strategie vývoje a podpory

AI strategie: Specifikace řešených úloh

- Které business procesy je **možné** a **vhodné** podporovat pomocí AI?
- O jaký **druh úlohy** se jedná?
- Jaká **AI metoda** tuto úlohu řeší?
- Co je **datovým vstupem** úlohy?
- Jaká **výstupní data** požadujeme?
- Jaké mohou **vzniknout chyby**, jakou k nim **máme toleranci** a **kolik taková chyba stojí?**
- Jakým způsobem **ověříme kvalitu** výstupu?
- Jaký je **rozpočet** pro řešení úloh?

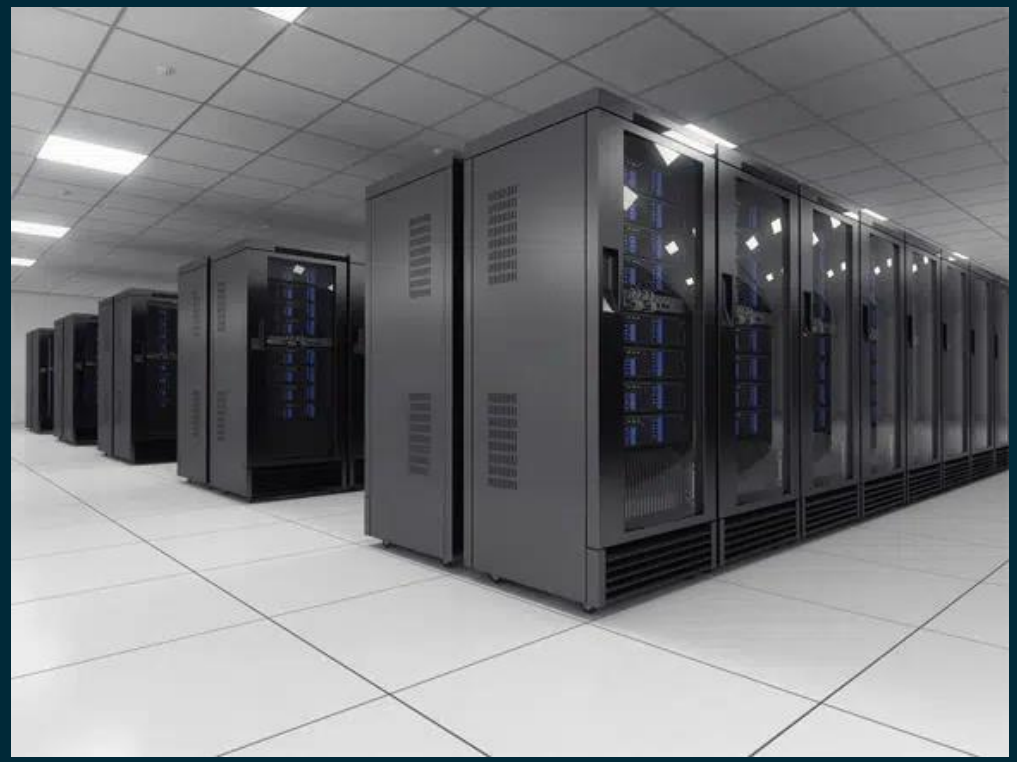


AI strategie: Datový návrh

- Jak má vypadat **datové úložiště**, odkud si AI bude brát data?
- Jakým způsobem se budou data **sbírat**?
- Jakým způsobem se budou data **validovat**?
- Jak ze surových dat uděláme **datasety pro učení**?
- **Kam a jak** se budou posílat **výstupy**?
- Jaká je **periodicita** datových operací?

AI strategie: Návrh HW infrastruktury

- Jaké **počítače/servery** potřebujeme pro řešení AI úloh?
- Jaké jsou požadavky na **síťovou konektivitu**?
- Jak **škálovat** výpočetní výkon?



AI strategie: Návrh SW tech stacku

- **API**
- **Web** nebo grafické uživatelské rozhraní
- **CI/CD pipeline** (automatická integrace a nasazování)
- **Sledování výkonu**
- **AutoML** – automatické ladění natrénovaných ML modelů
- Datové **triggery**
- **Messaging**
- Automatické **testování**
- …

AI strategie: Vývoj & podpora

- Jak zajistit další on-prem **vývoj**?
- Jaké jsou postupy pro správnou **správu a údržbu** AI systémů?
- Jak zajistit **aktuálnost modelů**?
- Jak řešit **chyby**?
- Kdo je za co **zodpovědný**?

MLOps: Kroky

1. AI strategie
2. **Zajištění hardwarové infrastruktury**
3. Návrh & učení ML modelů
4. Návrh & implementace SW řešení
5. Integrace řešení do business procesů
6. Dlouhodobá podpora

Zajištění HW infrastruktury

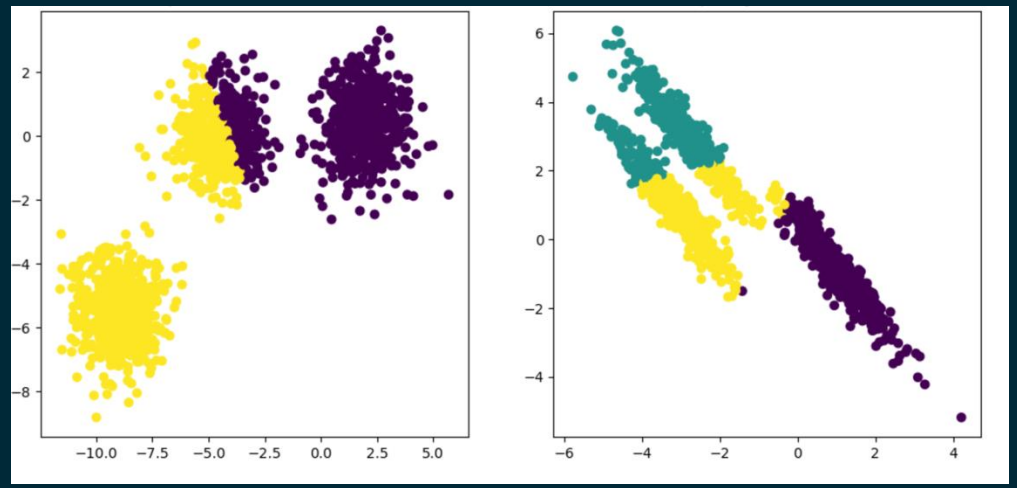
- … o tom uslyšíte ve **druhé části workshopu**

MLOps: Kroky

1. AI strategie
2. Zajištění hardwarové infrastruktury
3. **Návrh & učení ML modelů**
4. Návrh & implementace SW řešení
5. Integrace řešení do business procesů
6. Dlouhodobá podpora

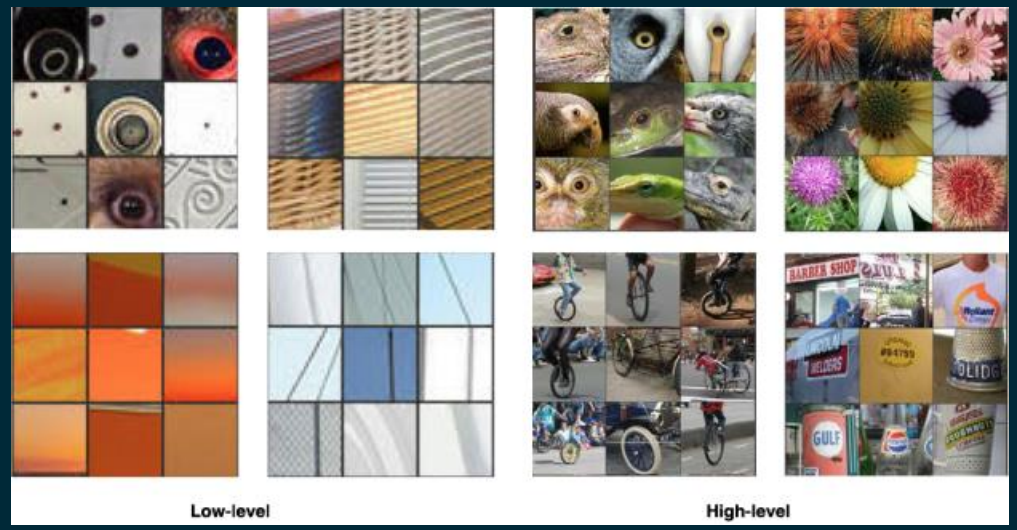
Návrh & učení ML modelů

- Výběr **správné ML metody** pro řešení úlohy
- Vytvoření **trénovacího datasetu**, včetně:
  - **Čištění & procesování** surových vstupních dat
  - **Feature engineering** – vytvoření nových datových atributů, tzv. příznaků, které pomáhají v řešení úlohy
  - **Feature representation** – reprezentace vstupních dat vektory příznaků
  - Zajištění **reprezentativnosti** trénovacího datasetu, aby postihoval všechny parametry úlohy



Návrh & učení ML modelů

- Samotné **učení a ladění**
  - 1) **Nastavíme parametry** modelu co nejvhodněji, jak naše aktuální znalost umožňuje
  - 2) **Naučíme model** na trénovacích datech
  - 3) **Vyhodnotíme přesnost** modelu na testovacích datech
  - 4) Pokud je nižší než požadovaná, provedeme **analýzu,** proč je nižší
  - 5) Zformulujeme nové **nastavení parametrů**
  - goto 1)
- Návrh & učení = **iterativní process**
  - Často se také od učení a ladění musíme vrátit ke sběru dat či data processingu



MLOps: Kroky

1. AI strategie
2. Zajištění hardwarové infrastruktury
3. Návrh & učení ML modelů
4. **Návrh & implementace SW řešení**
5. Integrace řešení do business procesů
6. Dlouhodobá podpora

Návrh & implementace SW řešení

- Model samotný je potřeba **zabalit do SW řešení**, aby byl schopen požadovaným způsobem **komunikovat** se zbytkem podniku
- Typické řešení: **API service**, která se nasadí v rámci IT infrastruktury
- Běžné je využití **batch processingu**, kdy se požadavky na zpracování vyhodnocují po várkách

MLOps: Kroky

1. AI strategie
2. Zajištění hardwarové infrastruktury
3. Návrh & učení ML modelů
4. Návrh & implementace SW řešení
5. **Integrace řešení do business procesů**
6. Dlouhodobá podpora

Integrace řešení do business procesů

- **Produkční nasazení řešení**… a vše, co to obnáší i u klasického SW
  - Ověření, že vše **funguje**
  - **Odladění** provozních chyb
  - Sběr požadavků na případné **updaty**

MLOps: Kroky

1. AI strategie
2. Zajištění hardwarové infrastruktury
3. Návrh & učení ML modelů
4. Návrh & implementace SW řešení
5. Integrace řešení do business procesů
6. **Dlouhodobá podpora**

Dlouhodobá podpora

- **HW/SW podpora** v klasickém smyslu
  - Servery běží
  - SW nepadá a je schopen vypořádávat požadavky
  - …
- ML s sebou z podstaty nese další **specifické požadavky na podporu**
  - ML modely jsou **“zmrazeny”**, jakmile jsou naučeny; samy o sobě se nedoučují
  - Existují i modely s **kontinuálním učením**, jsou ale v současnosti spíše na okraji zájmu

Dlouhodobá podpora

- Ideálně bychom chtěli **“nadčasové” modely**, které budou skvěle fungovat kontinuálně
- Řešením jsou **správně nastavené procesy** pro trénink updatovaných modelů

Dlouhodobá podpora: Spouštěče updatů

- **Data drift** - úloha, kterou řešíme, zůstala stejná, ale zásadním způsobem se nám časem změnila *data*
  - Např. řešíme predikci prodejů na základě údajů sesbíraných od zákazníků, ale začali nám ve velkém nakupovat výrazně mladší zákazníci
- **Concept drift** - změnila se *podstata* celé úlohy, tj. na základě vstupních dat už nejsme schopni kvalitních výstupů
  - Např. náš model detekuje emailový spam, indikátorem nám je četnost slov jako “free”, “win”, nebo “sex”. Spammeři ale přestanou tato slova používat.
- Zapracování **zkušeností získaných používáním**
- Nová **business realita**

bohem.ai

GAPP SYSTEM

Závěr

Závěr

- **AI & ML** jsou již dnes schopny **plně profesionální podpory** business procesů
  - Šetří peníze
  - Umí nad daty uvažovat a úvahy komunikovat s člověkem i stroji
  - “Neunaví se”, konzistentní kvalita výstupů
- **MLOps** - rozvinutá metodologie integrace AI/ML do businessu
  - Nutné pro maximalizaci přidané hodnoty AI
- **GAPP + bohem.ai** nabízí kompletní AI, ML a MLOps služby