AI v podnikání a praxi

Obsah

Souhrnné informace

Co je to silná umělá inteligence?

Silná umělá inteligence (AGI – Artificial General Intelligence) je hypotetický typ umělé inteligence, která by dokázala vykonávat jakýkoliv intelektuální úkol, který dokáže člověk, včetně vědomí, sebereflexe a generalizace mezi různými doménami.

- V současnosti neexistuje žádný funkční systém silné AI
- Zahrnuje schopnost skutečného porozumění, kreativitu a vědomí
- Dokáže řešit libovolné problémy bez specifického tréninku
- Může sama sebe vylepšovat (rekurzivní sebezdokonalování)
- Je předmětem filozofických debat i výzkumu
- Je spojována s konceptem technologické singularity
- Vyvolává etické a bezpečnostní otázky

Co je to slabá umělá inteligence?

Slabá umělá inteligence (ANI – Artificial Narrow Intelligence) je typ AI vyvinutý k plnění specifických úkolů v určité doméně bez skutečného porozumění nebo schopnosti generalizace mimo svůj účel.

- Všechny současné AI systémy jsou slabé AI
- Jsou specializované na konkrétní úlohy (např. hraní šachů, rozpoznávání obrazu)
- Nemohou přenášet znalosti mezi různými doménami
- Nemají vědomí ani skutečné porozumění
- Jsou efektivní v úzce vymezených oblastech
- Patří sem virtuální asistenti (Siri, Alexa), šachové programy, systémy pro rozpoznávání obrazu
- Jsou navrženy k řešení konkrétních problémů

Co je to strojové učení?

Strojové učení je podoblast umělé inteligence zaměřená na vývoj algoritmů a statistických modelů, které počítačovým systémům umožňují učit se z dat a zlepšovat svůj výkon při řešení úkolů bez explicitního programování.

- Identifikuje vzory a pravidla v datech automaticky
- Hlavní kategorie: učení s dozorem, bez dozoru a posilované učení
- Využívá statistiku, pravděpodobnost a optimalizační techniky
- Vyžaduje kvalitní a reprezentativní data
- Základem jsou algoritmy jako rozhodovací stromy, SVM, k-means clustering
- Aplikace: predikce, klasifikace, rozpoznávání vzorů, doporučující systémy
- Efektivita závisí na kvalitě dat a vhodnosti modelu

Co je to hluboké učení?

Hluboké učení je specializovaná podmnožina strojového učení využívající vícevrstvé (hluboké) neuronové sítě schopné automaticky extrahovat hierarchické reprezentace a abstraktní vzory z velkých objemů dat.

- Využívá neuronové sítě s mnoha vrstvami (hluboké architektury)
- Dokáže automaticky identifikovat komplexní vzory v datech
- Není nutné ruční vytváření příznaků (feature engineering)
- Vyžaduje velké množství dat a výpočetní síly
- Dosahuje mimořádných výsledků v obrazové a řečové analýze
- Základem velkých jazykových modelů (GPT, BERT)
- Klíčové architektury zahrnují CNN, RNN, LSTM, Transformery
- Umožnilo průlom v oblasti počítačového vidění a zpracování přirozeného jazyka

Co je to neuronová síť?

Neuronová síť je výpočetní model inspirovaný strukturou a funkcí biologických neuronových sítí v mozku, skládající se z propojených uzlů (neuronů) organizovaných do vrstev, které transformují vstupní data na výstupy pomocí vážených spojení a aktivačních funkcí.

- Skládá se ze vstupní vrstvy, skrytých vrstev a výstupní vrstvy
- Každý neuron přijímá vstupy od předchozích neuronů, zpracovává je a posílá výstupy dál
- Učí se úpravou vah spojení pomocí zpětné propagace chyby
- Používá aktivační funkce (ReLU, sigmoid, tanh) k zavedení nelinearity
- Umožňuje identifikaci komplexních vztahů v datech
- Vyžaduje optimalizaci hyperparametrů (počet vrstev, neuronů, learning rate)
- Existuje mnoho specializovaných architektur pro různé účely

Co je to generativní umělá inteligence?

Generativní umělá inteligence je kategorie AI systémů schopných vytvářet nový, originální obsah jako text, obrázky, hudbu nebo video na základě vzorů naučených z trénovacích dat, nikoliv pouze analyzovat nebo klasifikovat existující obsah.

- Zahrnuje modely jako GPT, DALL-E, Stable Diffusion, Midjourney
- Používá techniky jako variační autokodéry (VAE) a generativní adversariální sítě (GAN)
- Funguje na principu pravděpodobnostní predikce následujících prvků
- Dokáže tvořit obsah podobný lidské tvorbě, ale nemá skutečné porozumění
- Vyvolává otázky autorských práv a etiky
- Trpí problémy jako halucinace (vytváření nepravdivých informací)
- Transformuje způsob tvorby obsahu v mnoha oborech

Co je to prompt?

Prompt je textový vstup nebo pokyn zadaný uživatelem do systému umělé inteligence, který určuje, jaký typ výstupu má model generovat, a poskytuje kontext pro generování odpovědi.

- Funguje jako instrukce nebo dotaz pro AI systém
- Může obsahovat kontext, příklady nebo specifické požadavky
- Kvalita a přesnost promptu významně ovlivňuje výstup AI
- Prompt engineering je technika optimalizace promptů pro požadované výsledky
- Složitější modely dokážou reagovat na složitější a nuancovanější prompty

- Může zahrnovat texty v přirozeném jazyce i specifické formátovací instrukce
- V generativních modelech definuje parametry generovaného obsahu

Co je to selfprompting?

Selfprompting je technika, při které AI model generuje nebo upravuje vlastní prompty během generování odpovědi, což mu umožňuje samostatně strukturovat své myšlení, rozdělit komplexní problémy na menší kroky a zdokonalit svůj výstup bez dodatečných instrukcí od uživatele.

- Model si vytváří vlastní vnitřní instrukce nebo dílčí kroky
- Používá se k strukturování myšlenkových procesů AI
- Umožňuje modelu rozložit složité úlohy na jednodušší
- Zlepšuje kvalitu odpovědí u komplexních problémů
- Je podobné konceptu "řetězového myšlení" (chain-of-thought)
- Pomáhá modelu sledovat logickou linii uvažování
- Zvyšuje spolehlivost a konzistenci odpovědí

Co jsou to halucinace chatbotů?

Halucinace chatbotů jsou případy, kdy AI systémy generují nepřesné, vymyšlené nebo nepravdivé informace, které prezentují s důvěryhodností, ačkoliv tyto informace nejsou podloženy faktickými znalostmi nebo trénovacími daty.

- Vznikají z pravděpodobnostních předpovědí následujících tokenů v textu
- Projevují se generováním neexistujících citací, zdrojů nebo faktů
- Častější u témat, kde má model omezenou znalost nebo nejistotu
- Mohou vypadat velmi přesvědčivě díky sebejistému tónu
- Souvisí s tím, že modely optimalizují plynulost odpovědí, ne přesnost
- Problematické zejména v oblastech vyžadujících faktickou přesnost
- Omezují se pomocí technik jako RLHF (učení z lidské zpětné vazby)
- Představují jedno z hlavních omezení současných LLM

Co je to velký jazykový model?

Velký jazykový model (LLM – Large Language Model) je typ neuronové sítě s miliardami parametrů trénovaný na obrovském korpusu textových dat, který dokáže generovat, překládat a zpracovávat přirozený jazyk na základě statistických vzorů, které se naučil z trénovacích dat.

- Využívá architekturu Transformer s mechanismem pozornosti (attention)
- Obsahuje miliardy až biliony parametrů (vah)
- Trénuje se na masivních textových korpusech
- Dokáže generovat koherentní a kontextuálně relevantní text
- Může provádět širokou škálu jazykových úkolů bez specifického přetrénování
- Příklady zahrnují GPT-4, Claude, PaLM, Llama
- Porozumění je statistické, ne konceptuální
- Kvalita odpovědí závisí na kvalitě a rozsahu trénovacích dat

Jak se učí stroj při učení pod dohledem?

Učení pod dohledem (supervised learning) je metoda strojového učení, při které algoritmus analyzuje označená trénovací data obsahující vstupní-výstupní páry a vytváří model, jenž dokáže předpovídat výstupy pro nové, dříve neviděné vstupy.

- Vyžaduje označená data (vstupy s přiřazenými správnými výstupy)
- Algoritmus se snaží minimalizovat rozdíl mezi predikcí a skutečným výstupem
- Využívá ztrátovou funkci k měření chyby predikce
- Postupně optimalizuje parametry modelu k minimalizaci chyby
- Typické úlohy: klasifikace, regrese, rozpoznávání objektů
- Příklady algoritmů: rozhodovací stromy, neuronové sítě, SVM
- Kvalita modelu závisí na reprezentativnosti trénovacích dat
- Po natrénování je model evaluován na testovacích datech

Jak se učí stroj při učení bez dozoru?

Učení bez dozoru (unsupervised learning) je metoda strojového učení, při které algoritmus analyzuje neoznačená data a snaží se v nich samostatně identifikovat skryté struktury, vzory nebo skupiny bez jakýchkoliv předem poskytnutých správných odpovědí.

- Pracuje s neoznačenými daty (chybí informace o správném výstupu)
- Hledá přirozené vzory, struktury nebo shluky v datech
- Nevyžaduje lidské anotování dat (výhoda při velkých datových sadách)
- Hlavní typy: shlukování, dimenzionální redukce, detekce anomálií
- Příklady algoritmů: k-means, hierarchické shlukování, PCA, autokodéry
- Nemá jasně definovanou metriku úspěšnosti
- Používá se pro průzkumnou analýzu dat a identifikaci trendů
- Často předchází učení s dozorem jako přípravný krok

Co je posilované učení?

Posilované učení je metoda strojového učení, při které agent interaguje s prostředím, provádí akce a učí se optimální strategie na základě zpětné vazby ve formě odměn nebo penalizací, s cílem maximalizovat kumulativní odměnu.

- Agent se učí metodou pokus-omyl v interaktivním prostředí
- Rozhoduje o akcích na základě stavu prostředí
- Získává odměny nebo tresty za své akce
- Snaží se maximalizovat dlouhodobou kumulativní odměnu
- Využívá kompromis mezi průzkumem (exploration) a využíváním (exploitation)
- Klíčové algoritmy: Q-learning, DQN, SARSA, Policy Gradient
- Aplikace: robotika, autonomní řízení, herní AI, trading
- Kombinuje se s hlubokým učením v deep reinforcement learning

Co je výstupem tréninku neuronové sítě?

Výstupem tréninku neuronové sítě je naučený model s optimalizovanými hodnotami vah a biasů všech neuronů, který dokáže transformovat vstupní data na požadované výstupy na základě vzorů naučených z trénovacích dat.

- Optimalizované hodnoty vah spojení mezi neurony
- Hodnoty biasů jednotlivých neuronů
- Architektura sítě (pokud byla modifikována během tréninku)
- Hyperparametry modelu (často stanoveny před tréninkem)
- Transformační funkce pro převod vstupů na výstupy
- Naučené reprezentace dat ve vnitřních vrstvách (feature maps)
- Kvantifikovatelná přesnost na testovacích datech
- Někdy i informace o průběhu tréninku (learning curves)

Jak funguje zpětná propagace (backpropagation) v neuronové síti?

Zpětná propagace je algoritmus používaný k tréninku neuronových sítí, který počítá gradient ztrátové funkce vzhledem k vahám sítě a systematicky upravuje tyto váhy směrem k minimalizaci chyby, přičemž postupuje od výstupní vrstvy zpět ke vstupní.

- Využívá řetězové pravidlo diferenciálního počtu
- Probíhá ve dvou fázích: dopředný průchod a zpětný průchod
- V dopředném průchodu se vypočítá výstup sítě a chyba predikce
- Ve zpětném průchodu se vypočítá gradient chyby vzhledem k vahám
- Váhy se upravují v opačném směru gradientu (gradient descent)
- Velikost úprav řídí parametr learning rate
- Efektivně distribuuje "vinu" za chybu mezi jednotlivé neurony
- Umožňuje trénink hlubokých neuronových sítí

Kdo je autorem díla vytvořeného pomocí AI, pokud člověk AI jen zadal pokyn?

Autorem díla vytvořeného pomocí AI je podle současné právní úpravy ve většině jurisdikcí osoba, která poskytla tvůrčí vstup, přičemž pouhé zadání jednoduchého pokynu AI obvykle není považováno za dostatečný tvůrčí příspěvek pro přiznání autorství.

- Autorské právo vyžaduje lidského původce díla
- Míra lidského tvůrčího vkladu je klíčovým faktorem
- Komplexní a kreativní prompt může zakládat autorství
- V některých jurisdikcích může být dílo vytvořené AI nepřiznatelné autorství
- V EU se posuzuje prvek originality a tvůrčí vklad člověka
- V USA existuje požadavek "human authorship" (lidského autorství)
- Právní rámce se stále vyvíjejí s technologickým pokrokem
- Vývojáři AI systémů obvykle nemají nárok na autorství děl vytvořených uživateli

Může být AI považována za autora?

AI nemůže být v současném právním rámci většiny zemí považována za autora díla, protože autorské právo tradičně vyžaduje lidského tvůrce s právní subjektivitou a kreativním záměrem.

- AI nemá právní subjektivitu (není právní osobou)
- Nemá vědomí ani vlastní kreativní záměr
- Autorské právo bylo vytvořeno pro ochranu lidské tvořivosti
- V USA Copyright Office odmítá registrovat díla vytvořená AI bez lidského vstupu
- V EU směrnice o autorském právu nepočítají s AI jako autorem
- Díla vytvořená čistě AI mohou spadat do veřejné domény
- Některé jurisdikce zvažují speciální kategorie ochrany pro AI díla
- Téma zůstává předmětem právních a etických diskusí

Co je to zaměstnanecké dílo?

Zaměstnanecké dílo je autorské dílo vytvořené zaměstnancem v rámci plnění pracovních povinností, u kterého majetková autorská práva automaticky náleží zaměstnavateli, zatímco osobnostní práva zůstávají autorovi.

- Vytváří se v rámci pracovněprávního vztahu
- Zaměstnavatel vykonává majetková práva k dílu svým jménem
- Zaměstnanec zůstává autorem (osobnostní práva)
- Zaměstnavatel může dílo upravovat, spojovat s jinými díly
- Může vzniknout i na základě jiných smluvních vztahů (např. služební dílo)
- Specifická úprava se liší podle jurisdikce
- V ČR je upraveno autorským zákonem
- Nevzniká mimo rozsah pracovních povinností zaměstnance

Jaké je riziko, pokud AI systém diskriminuje při náboru?

Riziko diskriminace AI systému při náboru spočívá v porušení antidiskriminačních zákonů, poškození reputace společnosti, právních postizích a vytváření či posilování společenských nerovností prostřednictvím systematického znevýhodňování určitých skupin uchazečů na základě chráněných charakteristik.

- Právní riziko porušení antidiskriminačních zákonů
- Finanční náklady spojené s právními spory a odškodněním
- Reputační škody a ztráta důvěry veřejnosti
- Vytváření nebo posilování existujících nerovností na pracovišti
- Znemožnění přístupu kvalifikovaným kandidátům z důvodu předpojatosti
- Diskriminace může být skrytá a obtížně odhalitelná
- Odpovědnost nese organizace používající AI systém
- EU AI Act klasifikuje HR systémy jako vysoce rizikové

Co je to zakázaný AI systém podle AI Act?

Zakázaný AI systém podle AI Actu je aplikace umělé inteligence, která představuje nepřijatelné riziko pro základní práva a bezpečnost občanů EU, a proto je její používání na území Evropské unie zcela zakázáno.

- Systémy využívající manipulativní techniky k ovlivňování chování
- Systémy využívající zranitelnosti osob (věk, postižení)
- Systémy sociálního skórování jednotlivců veřejnými orgány
- Systémy pro biometrickou identifikaci v reálném čase na veřejných prostranstvích (s výjimkami)
- Systémy umožňující necílenou extrapolaci biometrických dat z internetu
- Systémy prediktivního policie používané pro předvídání trestných činů jednotlivce
- Systémy pro rozpoznávání emocí na pracovišti a vzdělávacích institucích
- Zakázání má na rozdíl od ostatních kategorií absolutní charakter