

Velké jazykové modely v právu I.

Tereza Novotná, tereza.novotna@law.muni.cz

Osnova

1. LLMs a jak fungují
 1. Algoritmy
 2. Data
2. Seznam modelů a seznam témat
3. LLMs v právu a jak je používat
4. Regulace a metodiky

Large Language Models

Large Language Models

- **LLMs (velké jazykové modely)** jsou modely **strojového učení**, které jsou navrženy tak, aby se naučily statistické vlastnosti obsahu textu a generovaly nový text, který napodobuje styl původního vstupního textu.
- **Modely LLM** jsou vytvářeny pomocí techniky strojového učení známé jako Deep Learning. **Hluboké učení** je podmnožinou umělé inteligence, která je schopna učit se složité vzory v datech. Hluboké učení je realizováno pomocí **neuronových sítí**, což jsou výpočetní systémy, které jsou inspirovány schopností mozku učit se ze zkušeností. Algoritmy hlubokého učení lze zpracovávat velké datasety a mohou se učit z dat, která nejsou strukturovaná nebo označená. Díky tomu jsou vhodné pro **zpracování přirozeného jazyka (NLP)**.

Large Language Models

- LLM používají neuronové sítě založené na **transformátoru** (model strojového učení), který si poradí s velkým množstvím dat, protože může paralelizovat trénování, a vytvářet tak opravdu velké modely.
- Přístup, který LLM používají, se nazývá **autoregresní model**, což je feed-forward model, který předpovídá další slovo z množiny slov dané kontextem.
- Výkonnost LLM je založena na **množství dat** a **počtu parametrů**. Parametry jsou synonymem pro váhy, což je termín používaný pro parametry neuronových sítí. Váhy jsou to, co se učící algoritmus naučí během trénování.
- **Data** pro trénování modelů se používají z internetu, knih a dalších zdrojů, aby se natrénovalo hluboké porozumění lidskému jazyku.

Large Language Models

- **LLM** jsou dnes známé především pro generování textů (chatboty), lze je však využít i pro různé úlohy související s textem (vyhledávání informací, sumarizace apod.).
- **Chatboty** založené na LLM jsou aplikace postavené na LLM, které se dokážou přizpůsobit různým vstupům od uživatelů, chápat nuance a poskytovat relevantní odpovědi.
- **Generativní umělá inteligence** jsou modely, které mohou zahrnovat celou řadu úloh nad rámec generování jazyka, včetně generování obrázků a videí, skládání hudby a dalších.
- **Fine-tuning LLM** je proces přetrénování předem natrénovaného jazykového modelu na konkrétní úlohu nebo soubor dat s cílem přizpůsobit jej pro konkrétní aplikaci.

Chatboty

1. [ChatGPT](#): Konverzační platforma, která využívá modely GPT a kterou vyvinula společnost OpenAI, dokáže generovat text pro různé úkoly, jako jsou odpovědi na otázky, hledání tvůrčí inspirace a učení se něčemu novému.
2. [Bing Chat](#): Nová funkce služby Microsoft Bing založená na modelech GPT, která vám umožní komunikovat s chatbotem s umělou inteligencí, který vám odpoví na vaše dotazy, pomůže vám s úkoly a vygeneruje pro vás kreativní obsah.
3. [Google Bard](#): Využívá PaLM 2, odlehčenou a optimalizovanou verzi LaMDA. Na rozdíl od jiných chatbotů využívajících LLM, jako je ChatGPT, může Google Bard přistupovat k internetu a získávat informace z různých zdrojů, jako je například vyhledávač Google.
4. [Perplexity](#): Využívá možnosti modelů GPT společnosti OpenAI. Dokáže vygenerovat obecnou odpověď na váš dotaz, po které následuje řada odkazů na webové stránky, které AI považuje za relevantní k vašemu dotazu. Můžete také položit doplňující otázky nebo upřesnit svůj dotaz.
5. [LLaMa via Perplexity](#): Model LLaMa je rodina modelů LLM, které vydala společnost Meta AI; dodává se ve čtyřech velikostech, v těchto verzích jsou k dispozici čtyři varianty: 7B, 13B, 33B a 65B parametrů. Model LLaMa lze použít jako základ pro doladění nebo přizpůsobení konkrétním aplikacím nebo případům použití, například chatbotům, vyhledávačům, sumarizátorům, překladačům a dalším.
6. [SciSpace](#): Jedná se o nástroj k rešerši odborných zdrojů a vědeckých článků i se zdroji.

LLMs v právu a jak je používat

- LLM generuje odpověď pouze v případě, že informace (data) byly poskytnuty během tréninkové fáze.
- Velký problém u právních otázek týkajících se národních právních systémů (jako je CZ nebo jakýkoli jiný "menší" a neanglický národní právní systém) - **nedostatek údajů** nebo velmi omezené množství údajů a z irelevantních zdrojů.
- Problém související s daty - **nedostupnost dat a datových souborů** (korpusů) právních textů (soudních rozhodnutí, právních předpisů).

LLMs v právu a jak je používat

Kdy může pronajímatel zvednout nájemné dle české právní úpravy?

LLMs v právu a jak je používat

- Právní rešerše
- Sumarizace
- Anotace, shrnutí případů – FIRAC, extrakce argumentů
- Editace
- Psaní smluv

Limity LLMs v právu

- Nikdy nevkládejte osobní údaje ani neveřejné informace!
- Ptejte se na zdroje a argumentaci.
- Vždy kontrolujte výsledek, odpovědnost je vždy na uživateli.

Tipy pro používání

- Poskytni co nejvíc detailů.
- Iteruj otázku.
- Požaduj uvažování, které vedlo k odpovědi, krok za krokem.

Metodiky

- Stanovisko MUNI k užití AI ve výuce
- Stanovisko ČAKu

Regulace

- AI Act
- Záruky ochrany práv [čl. 22 GDPR](#)
- Autorskoprávní otázky – práva k datům a algoritmům

Díky za pozornost!