

**M U N I
L A W**

Směrem k hybridní metodě evaluace velkých jazykových modelů v právu: kvalitativní hodnocení uživatelů a jeho implikace

Tereza Novotná, Ústav práva a technologií, Právnická fakulta MU

Marco Sanchi, University of Bologna

Osnova

1. Motivace
2. Metodologie a výzkumné otázky
3. Přehled literatury
4. Případová studie
5. Výsledky a diskuze

Motivace

- LLMs se v právu používají a jejich užití bude pravděpodobně čím dál častější.
- LLMs se v právu používají různě – k řadě různých úkolů, které jsou navíc velmi odlišné od „běžného“ užití, tzv. doménově specifické.
- **Je nutné systematicky a objektivně vyhodnocovat kvalitu jejich výstupů.**

Metodologie a výzkumné otázky

– Metodologie:

- Přehled literatury na evaluaci LLMs v právu a
- Případová studie

– Výzkumné otázky:

1. Jaké jsou nejvhodnější metody k evaluaci LLMs v právu podle state-of-the-art literatury?
2. Co právníci považují za důležité při práci s LLMs?
3. Jsou tyto nejvhodnější metody v souladu s tím, co právníci považují za důležité?

Přehled literatury

- Scoping review – seznam důležitých časopisů a konferencí v AI & Law a poté tzv. snowballing metoda
- 3 hlavní směry:
 - automatické – nejjednodušší, objektivní, porovnatelné, ne vždy odpovídá realitě
 - hodnocení lidmi – časově i finančně náročné, závisí na velikosti vzorku a metodologii, subjektivní, limity v porovnatelnosti
 - hybridní – kombinace, ne příliš časté, zdůrazňuje výhody obou

Případová studie - data

- předmět Umělá inteligence v právní praxi
- studenti pomocí dostupných LLMs vytvářeli právní dokumenty dle výběru
- součástí byl vznik tzv. reportu, kde popisovali svou práci s LLM
- v reportu hodnotili jak práci, tak výsledek

Případová studie – tematická analýza

- Reporty byly kvalitativně analyzovány tematickou analýzou (*Braun a Clarke: Using thematic analysis in psychology, 2006*):
 - označování částí textů z hlediska toho, co studenti nacházeli důležitým/hodnotili
 - přiřazování tzv. labelů těmto částem textu
 - propojování těchto labelů do meta-témat
 - všechny kroky se provádí iterativně

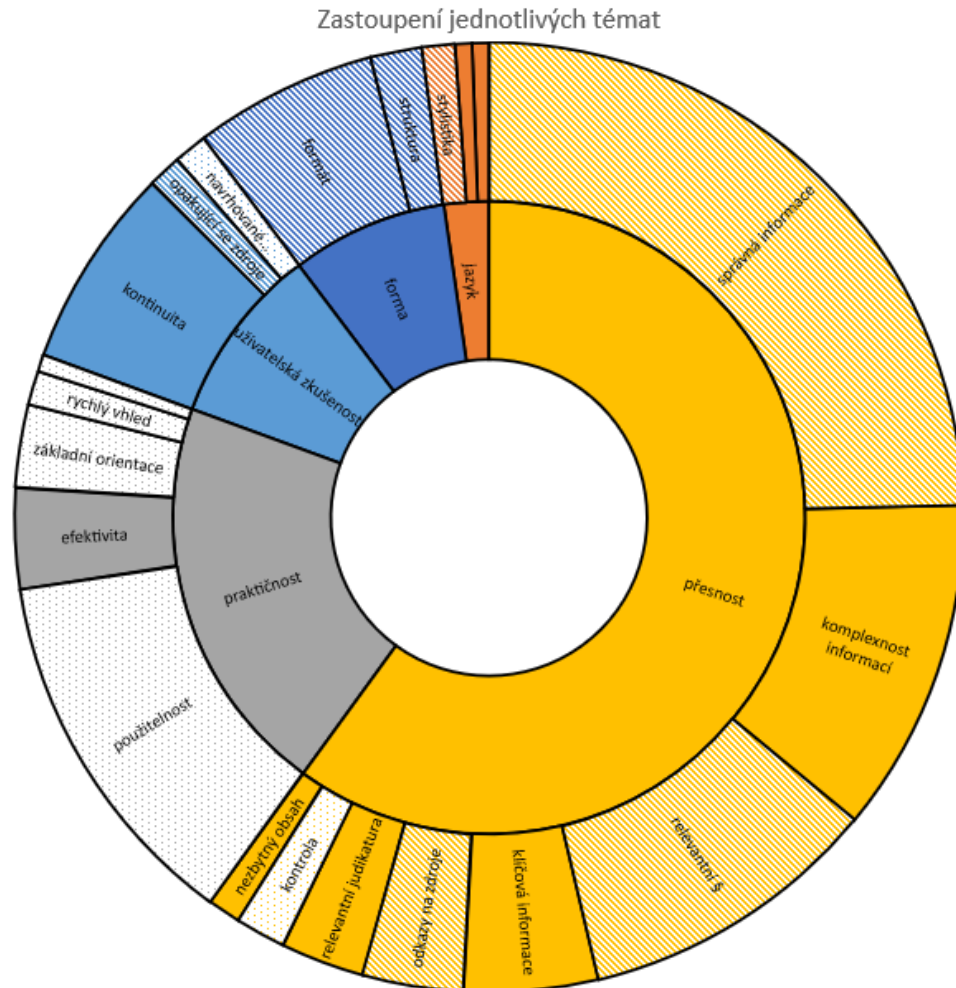
Případová studie - data

Typ	Nástroje	Text	Label
Nájemní smlouva	ChatGPT (GPT - 3.0)	Zajímavé je, že v dalších verzích naprosto rezignoval na generování podpisového řádku	nezbytný obsah
Zprostředkovatelská smlouva	ChatGPT (GPT - 3.0)	při pročtení textu smlouvy lze bezsrozumitelnost problému zjistit smysl a cíl právního jednání smluvních stran.	
Pracovní smlouva	Bard	. Z nějakého důvodu také slovo „nepracuje“ bral jako sprosté slovo, když jej cenzuroval (viz výše	správné informace
Žaloba	Copilot	Žalovaného označil jako „protižalovaného“ a petit napsal pouze ve větách.	správné informace

Případová studie - data

Téma	Label	Výskyt	Metodologie evaluace
přesnost	správná informace	43	automatická
praktičnost	použitelnost	22	lidská
přesnost	komplexnost informací	20	obojí
přesnost	relevantní §	18	automatická
uživatelská zkušenost	kontinuita	12	obojí
forma	formát	11	automatická
přesnost	klíčová informace	8	obojí
praktičnost	efektivita	6	obojí
přesnost	odkazy na zdroje	6	automatická
praktičnost	základní orientace	5	lidská
přesnost	relevantní judikatura	5	obojí
forma	struktura	3	automatická
přesnost	kontrola	3	lidská
jazyk	stylistika	2	automatická
praktičnost	rychlý vhled	2	lidská
přesnost	nezbytný obsah	2	obojí
uživatelská zkušenost	navrhované upřesňující dotazy	2	lidská
uživatelská zkušenost	opakující se zdroje	2	automatická
jazyk	srozumitelnost	1	obojí
jazyk	kvalita textu	1	obojí
praktičnost	praktické prameny	1	lidská

Výsledky a diskuze



- Tečkovaná: lidská
- Šrafovaná: automatická
- Plná: obojí

Závěry a limity

- Neoddělitelnost automatických metod a metod založených na hodnocení lidmi (ztráta kontextu) -> hybridní metody
- Definice 5 hlavních témat, která právní studenti považují za důležitá při práci s LLMs
- Důležitost detailnosti a granularity evaluační metodologie
- Práce s LLMs v právu je kombinací získávání informací a generování textu – musí být takto hodnocena.

Závěry a limity

- Limity: malý vzorek, homogenní skupina, studenti, práce ve dvojicích, velmi široké zadání
- Budoucí kroky: větší vzorek, hodnotit definovaná kritéria, navrhnout na jejich základě metodologii

**Díky za pozornost!
Máte dotazy?**