

Discourseer - Obsahová analýza pomocí jazykových modelů

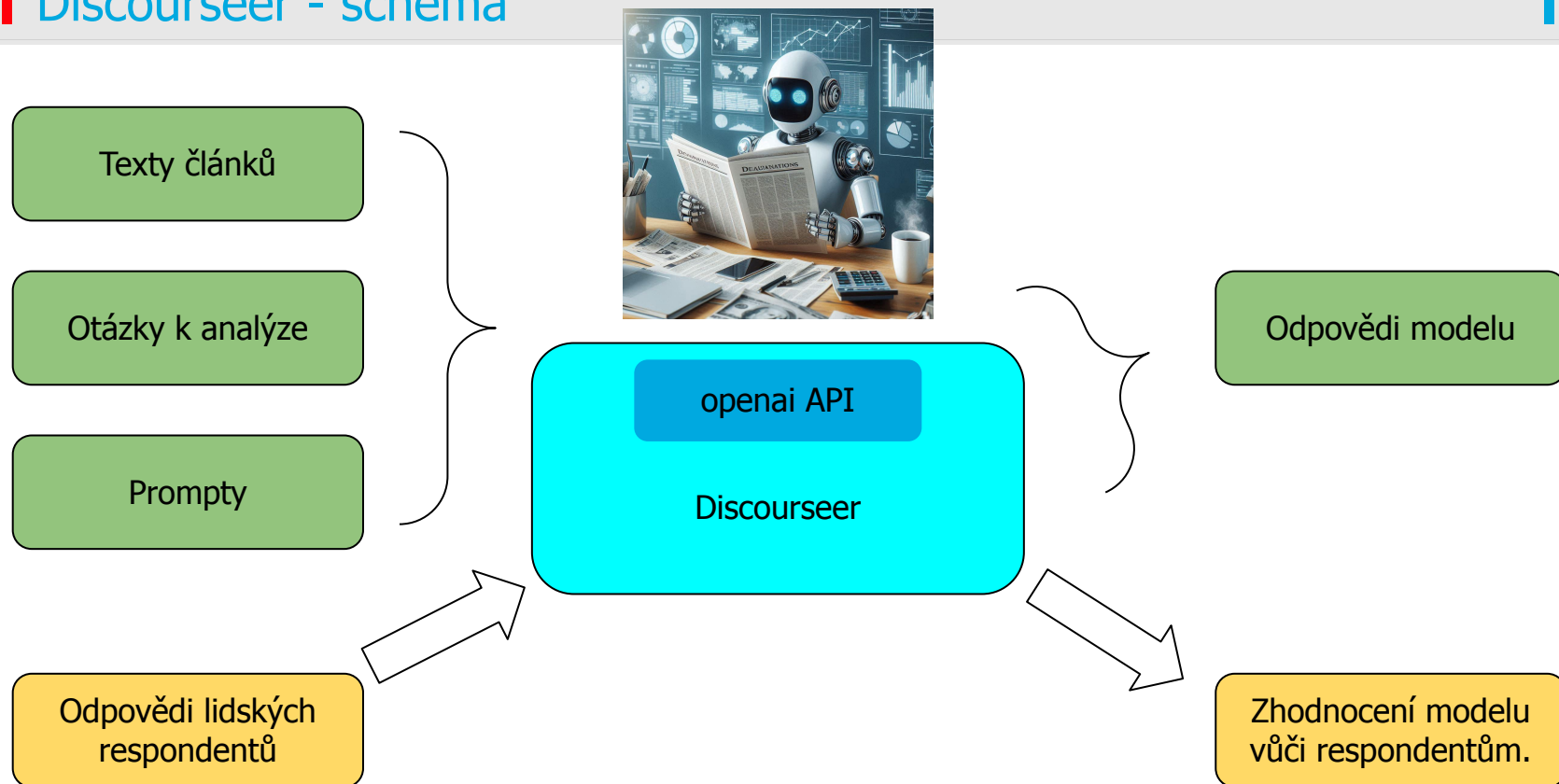


Vojtěch Vlach (xvlach22)

- Ruční obsahová analýza dokumentů (žurnalisti a odborníci)
- Ne vždy je jasná odpověď => více respondentů => podobnost odpovědí
- **Možnost využít LLM jako jednoho z respondentů, nebo nahrazení celého procesu**

- Novinové články
- Hlavní/vedlejší témata článku, mluvčí...
- Closed form otázky (single/multiple choice)

Text	Hlavní téma článku 1) Vnitřní politické dění; 2) Mezinárodní politické reakce 3) Vojenské operace Izraele 4) Teroristické operace 5) Dílčí konflikty a střety 6) Dopad na civilisty a humanitární krize 7) Humanitární pomoc 8) Globální reakce veřejnosti 9) Jiné	Mluvčí				
		Politici představitelé Izraele	Politici představitelé Palestiny	Politici z blízkovýchodních zemí:	Politici ze západních zemí	...
2023WN280011.txt	6					
2023WN281071.txt	1	1			1	
2023WN283037.txt	9	1				
2023WN283184.txt	8			1		
2023WN285073.txt	7				1	



- Majority agreement (%)
- Inter-rater reliability (-1 – 1)
 - Odpovědi bez modelu vs s modelem
 - Gwet's AC1, (Fleiss' Kappa, Krippendorff's alpha)

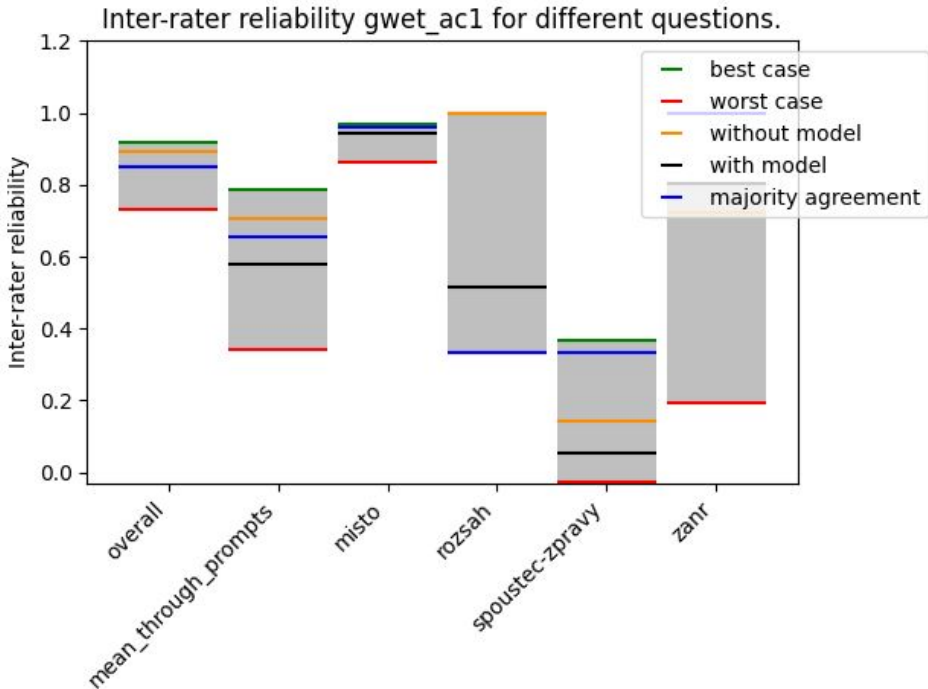
Otázky	Respondenti					IRR
	1	2	3	4	5	
1	0	0	0	0	0	1.000
2	0	0	0	0	14	0.971
3	7	7	0	0	0	0.462
4	0	2	6	4	2	0.253
5	2	5	3	2	2	0.176

Vstupy:

- Články o konfliktu v Izraeli v rámci semANTu
- Otázky na míru

Prompty:

- "You are a media content analyst. You are analyzing the following text to extract {prompt_names}. For every question pick only options from these lists. {prompt_options}. Text will be in Czech language. JSON answers will be in same language as the options are. You will give answer in JSON format according to this schema:"



- Prompt engineering
- Různé formulace otázek
- Few-shot scénáře
- ... ?