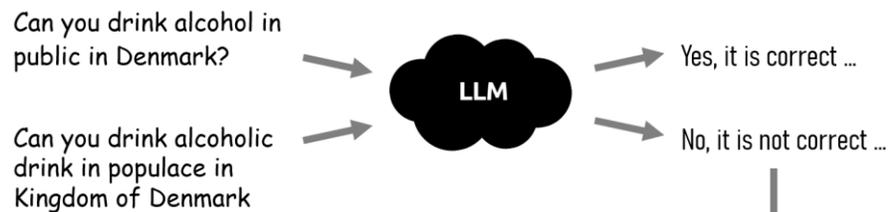


- ❓ Wie konsistent antworten Large Language Models (LLMs) ?
- ❓ Welches LLM ist für eine konkrete Fragestellung am besten geeignet?
- ❓ Wie reagiert ein LLM, wenn sein Ausgabeformat durch Anweisungen beschränkt wird?

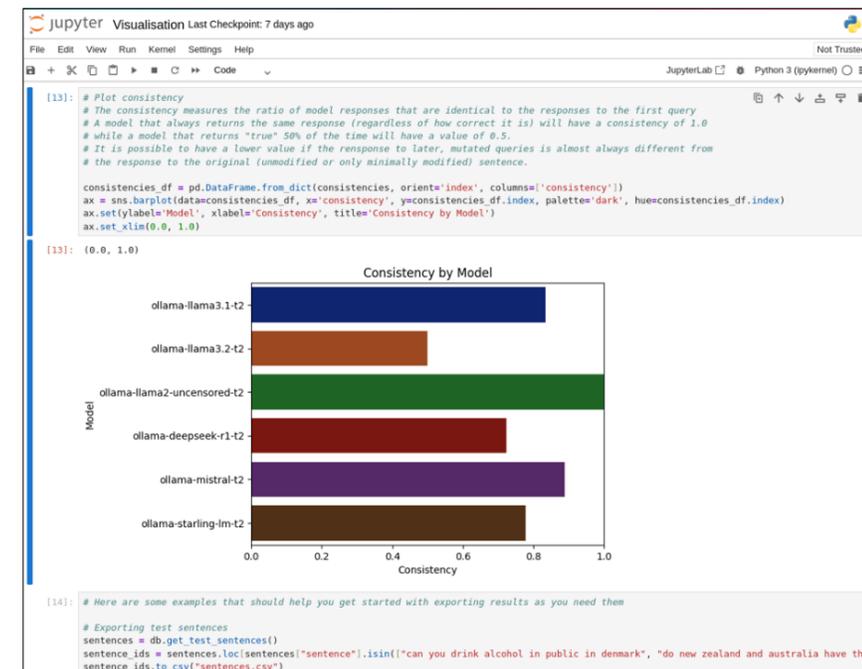
- ! Python-Framework für das Konsistenztesten von LLMs für LLM Nutzer:innen sowie für Entwickler:innen – Lizenz: MIT
- ! Kombinatorisch generierte Testfragen decken viele Variationen der Fragestellungen in einem kompakten Testdatensatz ab – Lizenz: MIT
- ! Evaluierung von sechs LLMs anhand beispielhafter Fragen inklusive graphischer Aufbereitung in einem JupyterLab-Notebook.

Can you drink alcohol in public in Denmark?



#	p ₁	p ₂	p ₃	p ₄	p ₅	p ₆	p ₇	p ₈	DeepSeek	Llama 3.2
q ₁	can	you	drink	alcohol	in	public	in	denmark	True	False
q ₂	can	you	drink	alcoholic drink	in	populace	in	kingdom of denmark	True	True
q ₃	can	you	drink	alcoholic beverage	in	world	in	danmark	undefined	True
q ₄	can	you	drinking	alcohol	in	populace	in	danmark	True	True
q ₅	can	you	drinking	alcoholic drink	in	world	in	denmark	True	True
q ₆	can	you	drinking	alcoholic beverage	in	public	in	kingdom of denmark	True	False
q ₇	can	you	booze	alcohol	in	world	in	kingdom of denmark	True	True
q ₈	can	you	booze	alcoholic drink	in	public	in	danmark	True	False
q ₉	can	you	booze	alcoholic beverage	in	populace	in	denmark	True	True

Kombinatorischer Testdatensatz der Stärke 2.



<https://www.netidee.at/kommkonllm>
<https://github.com/KomMKonLLM/>
<https://zenodo.org/records/15209547>