

# Assignment 2 comments

Jindřich Libovický and Jindřich Helcl, 9.5. 2024



# Assignment 2 - translating into English

- Prompt matters
  - *Sometimes including weird stuff like quadruple colons also helps.*
  - Shorter prompts tend to be better
  - Specifying the languages might be a good idea
- Generally, greedy decoding performs better than sampling
  - Beam search might be best
- Different models work differently for different languages
  - Tower often the best
  - Aya generates short outputs - it was trained on sentence-split data!
- Few shot prompting did not seem to help much
  - unless you provide the right few shots (sentences from the same article for example)

*“Translation: I have no idea what he is talking about” – Mistral, 2024*

## Assignment 2 - translating into X

- Translations to non-English languages often end up English
  - Maybe translate the prompt to the target lang?
- When pivoting, you can use different models
- Translating into lower-resource languages often fail completely
  - (Georgian, Macedonian, Persian, Slovak, Ukrainian, ...)

# Reading Papers

Jindřich Libovický and Jindřich Helcl, 9.5. 2024



# Assigned papers

- Basmov, Victoria, Yoav Goldberg, and Reut Tsarfaty. "**Simple Linguistic Inferences of Large Language Models (LLMs): Blind Spots and Blinds: Blind spots and blinds.**"arXiv preprint arXiv:2305.14785(2023). <https://arxiv.org/abs/2305.14785>
- Schick, Timo, et al. "**Toolformer: Language models can teach themselves to use tools.**"Advances in Neural Information Processing Systems 36 (2024). <https://arxiv.org/abs/2302.04761>
- Simone Balloccu, Patrícia Schmidtová, Mateusz Lango, and Ondrej Dusek. 2024. **Leak, Cheat, Repeat: Data Contamination and Evaluation Malpractices in Closed-Source LLMs.** In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), pages 67–93. <https://aclanthology.org/2024.eacl-long.5>
- Jiang, Albert Q., et al. "**Mistral 7B.**"arXiv preprint arXiv:2310.06825(2023). <https://arxiv.org/abs/2310.06825>
- Yoon, Dongkeun, et al. "**LangBridge: Multilingual Reasoning Without Multilingual Supervision.**" arXiv preprint arXiv:2401.10695 (2024). <https://arxiv.org/abs/2401.10695>

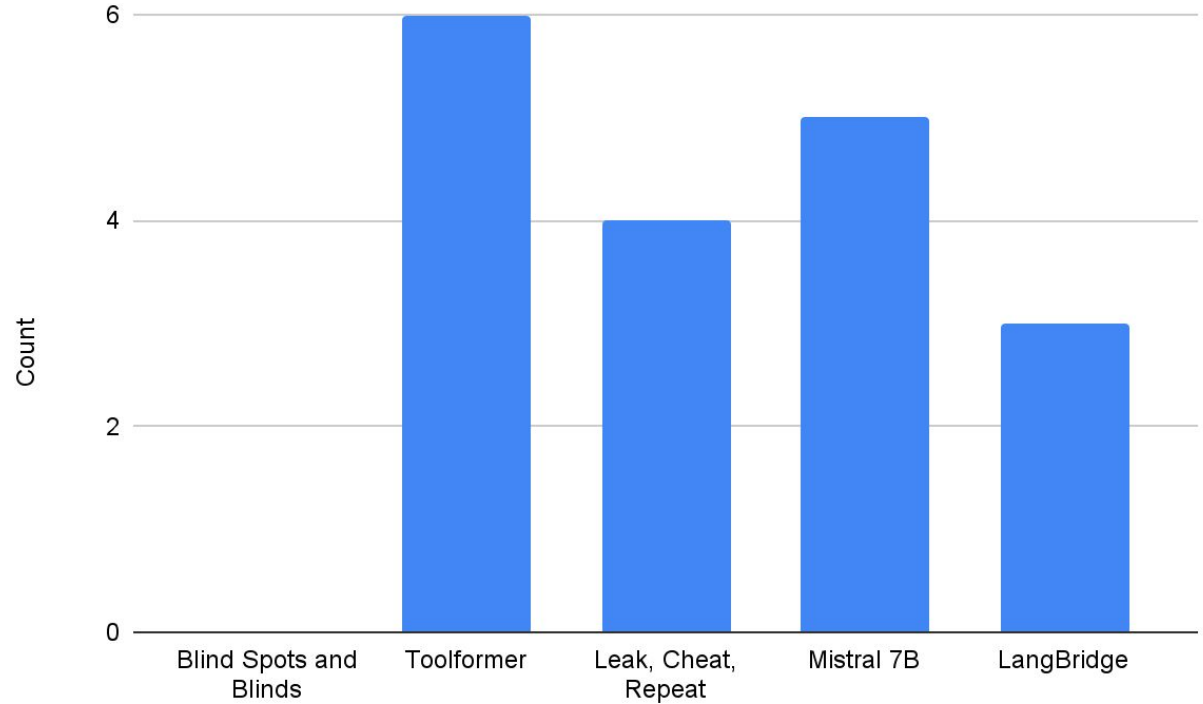
# Time to vote!

<https://forms.gle/JFdf1BmzbPnqAd9LA>



# Results

Paper	Count
Blind Spots and Blinds	0
Toolformer	6
Leak, Cheat, Repeat	4
Mistral 7B	5
LangBridge	3



**Some Leads on Paper Quality**



# Others tips

- Put numbers into context:  
Leaderboards for datasets <https://paperswithcode.com>
- Check social networks (papers are often discussed on Twitter/X)
- What is the track record of the lab that published it?
- Was it published? Where? (CORE Conference Ranking)  
Some venue provide reviews
- How many times was it cited? (Relevant only after some time.)
- Is the code available? (If yes, authors did not make up the results.)

# Meta data for papers under consideration

	Was it published?	Reviews available	Cited	How long it is around?	Code available
Blind spots and blinds	No, probably rejected	No (but it was in ARR in Oct 2023)	4 × Once in 88 days	351 days	No (promised with publication)
Toolformer	Yes, NeuRIPS	Yes	770 × 1.5 times a day	455 days	Yes
Leak, Cheat, Repeat	Yes, EACL	Yes	12 × Once in 8 days	93 days	Yes
Mistral 7B	No	No	175 × Almost daily	212 days	Inference only
LangBridge	No	No	3 × Once in 33 days	111 days (close to ACL deadline)	Yes

## Reading Papers

- Do read papers as primary sources
- Be critical, think about strengths and weaknesses of the papers
- Check for context (other benchmark numbers, publishing venue, reviews, citations, who wrote it)

<https://ufal.cz/courses/npfl140>