CS685 Quiz 4: *document-level machine translation*
Released 5/5, due 5/12 on Gradescope (please upload a PDF!)
Please answer each question in 2-4 sentences.

1. Marzena's lecture on Wednesday described many of the benefits of document-level translation over sentence-level translation. However, most prior machine translation research has focused on sentence-level translation. Explain why that has historically been the case.

Answer: The field of machine translation can be described as being in a **"sentence-level rut."** It is obvious that translation is not a sentence-level task, but the paradigm is so deeply ingrained into both research and production environments that it is hard to break out of, despite the clear and well-understood benefits. The situation is also increasingly at odds with developments in the field, where large language models trained with standard architectures on long documents have completely overtaken nearly every leaderboard. Prior to when the AI winter was over, and neural networks resurfaced as the new approach to solving natural language processing (NLP) problems, it was seen as the next logical step to use them for machine translation as well, which started from sentence-level translation as a foundational step.

- **Computational complexity:** Document-level translation is more computationally complex than sentence-level translation. This is because document-level translation needs to take into account the context of the entire document, while sentence-level translation can only consider the context of the current sentence. Sentence-level translations are fast by providing almost instant translations and are cheaper in computational complexity.
- **Data availability:** It is much easier to collect parallel data for sentence-level translation than for document-level translation. For sentence-level translation, you only need to find parallel sentences with less effort by leveraging existing sentence-aligned bilingual resources such as parallel books, news articles, or movie subtitles. For document-level translation, you need to find parallel documents, which are relatively low-resourced and vary in size, domain, and language, which is much more difficult. Thus, an additional manual effort is needed to identify matching documents in both languages.
- **Evaluation:** Sentence-level translation is easier to evaluate than document-level translation, which requires costly and difficult human evaluation. It is much easier to create a dataset of sentence pairs and evaluate the translation quality on a sentence-by-sentence basis using established metrics like BLEU, ROUGE, or METEOR.
- **Model:** Sentence-level translation is simpler to model than document-level translation, which requires custom architectures and decoding algorithms to incorporate document context

These reasons have historically made sentence-level translation more appealing and feasible for machine translation research. However, with the recent advances in large language models and document-level datasets, document-level translation is becoming more accessible and promising for future research.

Document-level translation has many benefits over sentence-level translation, such as preserving consistency, avoiding ambiguity, and capturing discourse phenomena. It is especially useful for tasks such as document summarization, information retrieval, and machine translation for longer-form text. As a result, there has been a growing interest in developing models that can handle document-level translation, including approaches based on hierarchical models, context-aware models, and models that incorporate global coherence constraints.

2. Please provide any general comments about the guest lecture in terms of its content / pacing / overall presentation. Also include any other information that you think could help Marzena improve her teaching!
   Answer:

   Answer: I enjoyed the guest lecture on **Evaluating LLMs on machine translation**. The content was very interesting and relevant to the course topic. The slides were well-designed and clear, with helpful examples and diagrams. I also thought the pacing of the lecture was excellent. She covered a lot of material but didn't rush through it. She gave the students plenty of time to absorb the information and ask questions. Overall, the presentation was engaging and informative, with a good balance of technical details and high-level insights.

   Some suggestions for improvement are:

1. I particularly appreciated the way Marzena used real-world examples of classical literature translation from French to 20 novels in English to illustrate her points. It would be helpful if she cited more real-world examples to which a diverse audience from various backgrounds could relate—for example, the case of Google Translate's failed attempt to translate the Bible.
2. It would be nice to have more comparisons or contrasts between different approaches or models, such as how LLMs differ from standard NMT models or how different LLMs perform on different tasks or languages for sentence level and document level translations.

3. It would be interesting to hear more about the challenges or limitations of LLMs for machine translation, such as domain generalization, causal validity, model explainability, fairness, representativeness, transparency., data quality, model robustness, etc.

Overall, I think Marzena did a great job in delivering the guest lecture explaining the complex concepts involved in LLM-MT in a clear and concise way and I learned a lot from it. I appreciate her effort and enthusiasm in teaching us about this exciting topic.

AI Disclosure

AI1: Did you use any AI assistance to complete this quiz? If so, please also specify what AI you used.
No

(only complete the below questions if you answered yes above)

AI2: If you used a large language model to assist you, please paste *all* of the prompts that you used below. Add a separate bullet for each prompt, and specify which problem is associated with which prompt.

- *Your response here*

AI3: (*Free response)* For each problem for which you used assistance, describe your overall experience with the AI. How helpful was it? Did it just directly give you a good answer, or did you have to edit it? Was its output ever obviously wrong or irrelevant? Did you use it to get the answer or check your own answer?

- *Your response here*