

### **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.

These are the following reasons why prior machine translation focused on sentence level tokenization

- Earlier machine learning techniques were prefix-based. They auto-regressively generate the tokens, and based on the previously generated tokens, they generate the next set of tokens. Also, earlier the recurrent neural networks processed generation sequentially, so the text generation didn't pay attention to the document level representation
- But with the advent of transformers, we can encode the entire document in a representation, and via cross-attention, we can enforce the text generation to pay attention to the entire document. So with the encoder-decoder model, we are able to enabling the document-level generation.

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!

## AI Disclosure

**AI1:** Did you use any AI assistance to complete this quiz? If so, please also specify what AI you used.

*Your answer here*

---

*(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*