# 1 Advanced Improvements

## 1.1 Compare the results of checkpoint averaging and ensembling and fill the results in the table:

| Model | BLEU Score | Time [s] |
|---|---|---|
| Only checkpoint25.pt | 28.4 | 26.2 |
| Only checkpoint26.pt | 29.0 | 24.5 |
| Only checkpoint27.pt | 28.7 | 25.5 |
| Averaged checkpoint ckpt_avg3.pt | 29.8 | 24.1 |
| Only ckpt_model2.pt | 28.6 | 25.2 |
| Ensemble checkpoint2{5,6,7}.pt | 29.8 | 68.7 |
| Ensemble ckpt_avg3.pt and ckpt_model2.pt | 30.1 | 46.8 |

## 1.2 What are your observations on the decoding speed?

**Checkpoint Averaging** maintains the decoding speed similar to that of a single model since the averaged model is still a single model.

**Ensembling** significantly slows down the decoding process because it requires running multiple models for each input. With Ensembling the runtime has increased linearly in the number of models. The more models in the ensemble, the slower the decoding process, as the computational overhead increases linearly with the number of models.

## 1.3 What are potential advantages and disadvantages of doing so?

The Dutch texts are more aggresively split into subwords, because we reuse the BPE learned on the German-English data.

| Advantages | Disadvantages |
|---|---|
| **Resource Efficiency:** Reusing BPE models means not having to train a new subword model specifically for Dutch | **Suboptimal Subword Segmentation for Dutch:** Dutch has its unique linguistic features and vocabulary that may not be optimally captured by a BPE model trained on German-English data |
| **Consistency Across Languages:** Applying the same BPE model across different languages may help maintain consistency in the handling of subwords | **Potential for Increased Out-of-Vocabulary (OOV) Rates:** If the BPE model does not segment Dutch words in a way that aligns with their actual usage or morphological structure, it could increase the OOV rate. |
| **Better Handling of Loanwords and Named Entities:** Dutch and German share a considerable amount of vocabulary, including loanwords and named entities | **Impacts on Model Training and Adaptation:** Training a model with suboptimal segmentation could also affect the efficiency of the model's learning process |

## 1.4 What would happen if we leave out –decoder-langtok?

Omitting –decoder-langtok in a multilingual translation task affects the model's performance by introducing ambiguity about the target language, potentially leading to decreased translation quality and reduced operational flexibility. Therefore the Bleu score drops sharply.