

A Sample MR tags and References

YelpNLG
MR: food rice no_adj mention_1 food meat little mention_1, negative,len_medium,not_first_person,no_exclamation
Ref: the bowl was mostly rice and some veggies and little meat.
E2E NLG
MR: name[Blue Spice], eatType[pub], customer rating[5 out of 5], near[Crowne Plaza Hotel]
Ref: The Blue Spice pub, near Crowne Plaza Hotel, has a customer rating of 5 out of 5.
ED
MR: lonely (girlfriend years empty)
Ref: I just broke up with my girlfriend, we were together for 8 years, i feel so empty.

Table 9: Sample MR & reference for different datasets.

B Pre-Trained models

In this section we provide a brief introduction of the pre-trained models that have been used in this work.

B.1 BART

BART is a pre-trained model designed by combining Bidirectional and Auto-Regressive Transformers. The pre-training involves corrupting text with an arbitrary noise function, and learning to reconstruct the original text. The downstream tasks for BART includes sequence and token classification, sequence generation and NMT. We investigate BART generation capability from MR to text task by finetuning on task specific datasets using BART implementation from (Wolf et al., 2019). The input MR varies among dataset consisting only content slot types and values with varied granularity of style tags, style and content values.

B.2 GPT-2

GPT-2 is a large transformer-based language model with 1.5 billion parameters, trained on 8 million web pages for a total of 40GB of text. GPT-2 is trained with a simple objective³: predict the next word, given all of the previous words within some text.

³<https://openai.com/blog/better-language-models/>

B.3 GPT-Neo

GPT-Neo 125M is a transformer model with 125 million parameters, designed using EleutherAI’s replication⁴ of the GPT-3 architecture. GPT-Neo 125M was trained on the Pile, a large scale curated dataset created by EleutherAI for the purpose of training this model. This model was trained on the Pile for 300 billion tokens over 572,300 steps. It was trained as a masked autoregressive language model, using cross-entropy loss.

C Attention maps for YelpNLG examples

In this section, we provide a visualization of the attention maps as discussed in Section 5.1.

⁴<https://huggingface.co/EleutherAI/gpt-neo-125M>

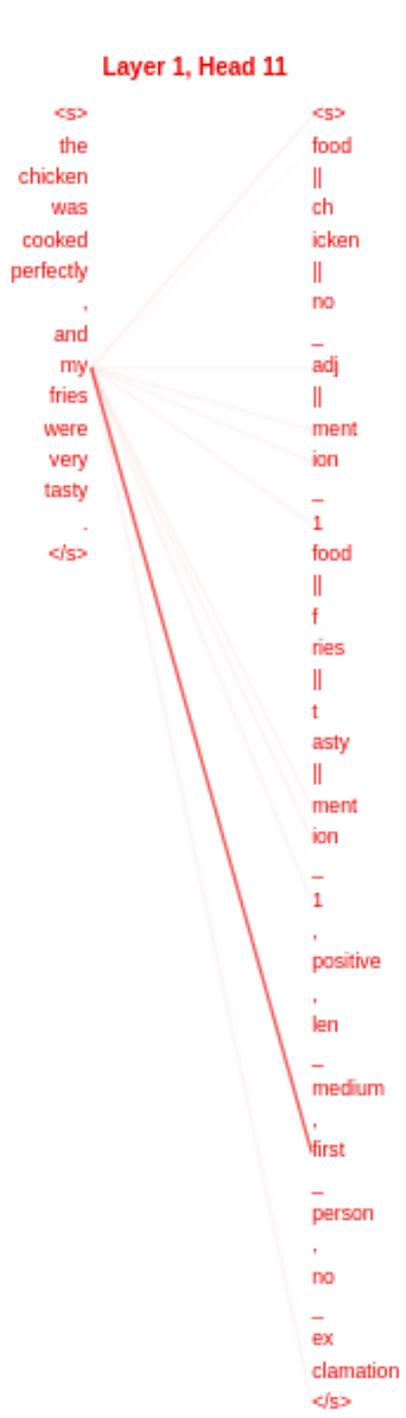


Figure 1: The first person reference 'my' in the output sentence is attending to 'first' token in the input in L1H11.

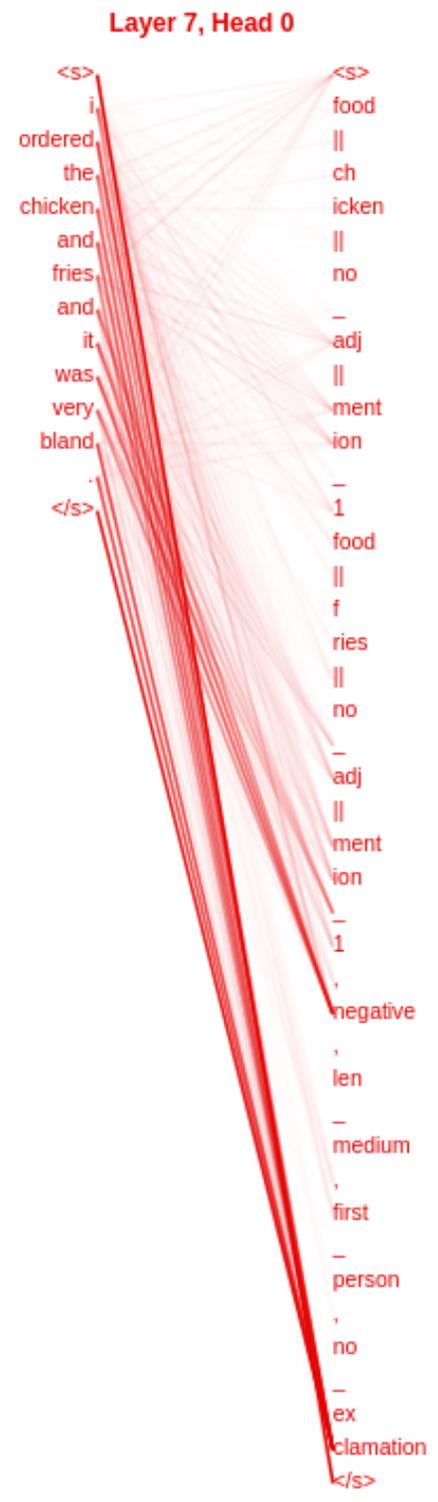


Figure 2: The sentiment tag 'negative' has been attended by most of the output tokens in L7H0.

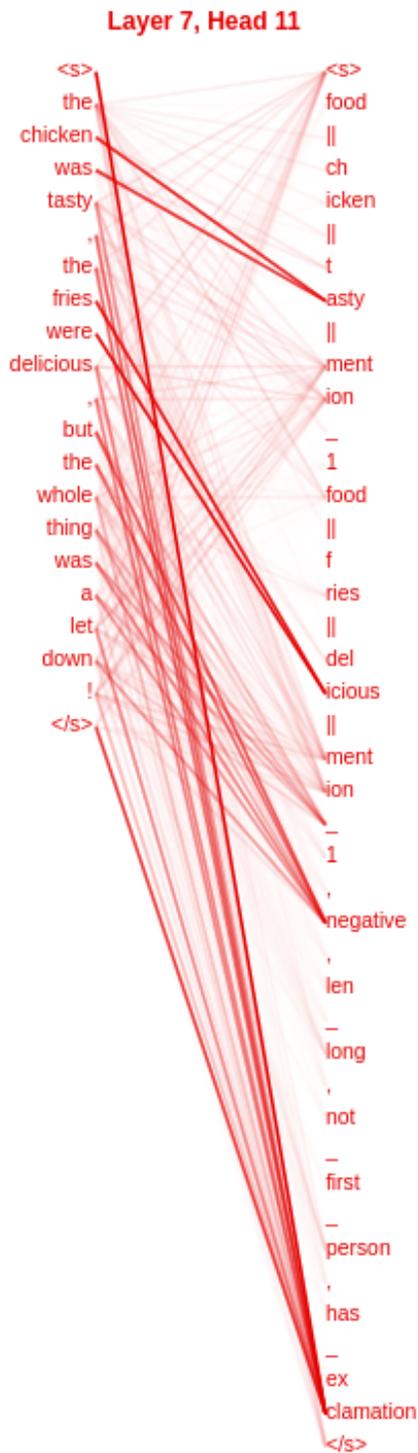


Figure 3: The food items 'chicken' and 'fries' in the output sentence are attending to their associated adjectives tags 'tasty' and 'delicious' in the input in L7H11.

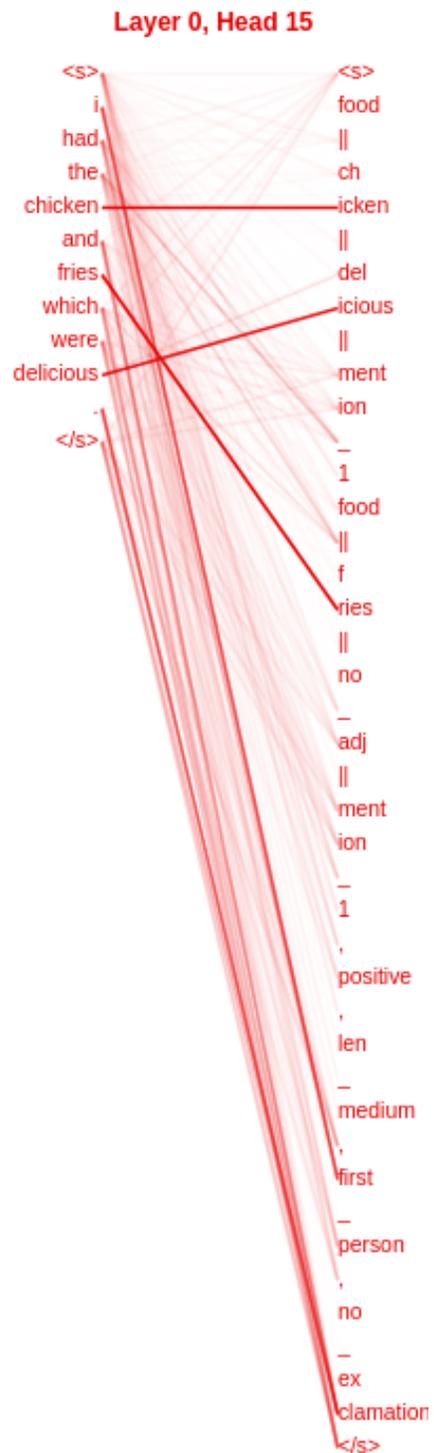


Figure 4: The food items 'chicken' and 'fries' in the output sentence are attending to the food item tags 'chicken' and 'fries' in the input in L0H15.