

EN.705.743.8VL.FA25

Module 6 Reading Assignment Response:

Breanna Burd

Q1. GPT2 and T5 introduced similar ideas in the beginning of 2019. The key idea is stated in the GPT2 paper: "the supervised objective is the same as the unsupervised objective, but only evaluated on a subset of the sequence". Explain what this means in your own words.

In order to answer this question, we must first define what the objective is for both cases. The objective is to accurately output meaningful information from a model given some input (this can be a myriad of things like question answering, paraphrasing, summarizing, sentence completion, etc. as listed in section 2.3 of the T5 article). It is impossible to have a very large dataset with completely labeled data, thus supervised learning on a subset was implemented for pre-training in some cases and unlabeled data was added for additional training for unsupervised learning. The labeled data can be used to reinforce the model's output by marking the predicted output as true/false. Since the majority of the data is unlabeled, the model is hopefully able to "learn" from the data alone, assuming enough important information is fed into the model.

References:

<https://arxiv.org/pdf/1910.10683>

https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf