Kavan Mehta

Ms. Whitcomb

ISM 2

9 October 2022

**Research Assessment #5**

**Date:** 9 October 2022

**Subject:** A Survey on Transformers

**MLA citation(s):**

Tay, Yi, et al. "Efficient transformers: A survey." *ACM Computing Surveys (CSUR)* (2020).

**Assessment:**

From my past research about natural language processing, I discovered that transformer models are being used widely through many platforms primarily because of their dynamic nature and effectiveness in understanding human language in context. Hence, I decided to explore transformers by understanding their general functions, inner workings, effectiveness, and applications in other fields. To learn more about this advanced technology in natural language processing, I found a journal article, “Efficient transformers: a survey,” which provided a general understanding of the architecture of transformers and their inner workings as well as their effectiveness and efficiency in solving real-world problems.

Through the journal article, I learned more about the principle of transformers as a whole and why they are used in fields such as natural language processing. For example, I first learned that they are being used for “language understanding, image processing, and information retrieval” and the rapid recent developments in them as almost “a dozen new efficiency-focused models [have been] proposed in just the last six months” (Tay et al. 1). This has helped me

understand that the field is rapidly expanding, and transformers are being used in fields such as computer vision too because of their ability to be dynamic in understanding input. This means that they have wide applications and new technologies are constantly coming in making it a very beneficial field for my ISM research. However, the transformer models aren't as easy to use as they seem because they have a "quadratic [time] complexity" and have large "computation costs," especially when facing input training and modeling "long sequences… [such as] documents, images, and videos" that composed large inputs (Tay et al. 2). This helps connect to my last year ISM knowledge that transformers have a bigger issue in facing large inputs of data with efficiency like other models, and thus, for my ISM original work and even final product, it would be better to use a pre-trained transformer model like BERT using transfer learning or by using a standard transformer AI library like Hugging Face. My experiences with my ISM 1 mentor on understanding transfer learning will help me utilize it correctly if I decide to go that path this year with natural language processing. This way I can understand the theory in the background, but my implementation of transformers can be quite easier in terms of computational power, memory, and the time required to use the model on input data. While the transformers do have limitations to their ability, they have been proven to be exceedingly useful as they use multiple transformer blocks that contain a "multi-head self attention mechanism, a position-wise feed-forward network, layer normalization modules and residual connectors" (Tay et al. 3). This helped me understand the architecture of transformers and how they really work. The inputs are passed through multiple layers where they first get tokenized, used as encodings, put into networks with activation functions, then normalized with the residuals of the loss in predictions, and ultimately improve the model through more training. This concept is extremely fascinating to me and really shows the dynamic nature of transformers as the tokenized input, or

words in natural language processing, each relates to the other's meaning and put the overall meaning of the entire text or document into the context of the real author intent. This mechanism also relates to my knowledge from last year on using activation functions such as ReLU to threshold the data for better analysis and predictions by the main model in feed-forward networks. I also learned that transformers have three main modes of approach: "encoder-only…decoder-only…and encoder-decoder" (Tay et al. 4). This was a new addition to my knowledge as I learned that transformers are used for these primary functions to convert the input data to a form for utilization to classify text, predict the next word or phrase in the text, and for many other contexts. I think learning about transformers has definitely shed light on the model as something very fascinating, and I definitely want to learn more about this model as I could use it for my research this year due to its overall effectiveness and wide applications!

As I meet more professionals that work in the field of machine learning, I hope to learn from their experiences to gain insight into applications of machine learning and how I can approach my original work in natural language processing with transformers. I also look forward to learning more about transformer models in the Hugging Face AI library and how I could use the library to create NLP projects. I will continue to explore other scientific papers to understand the theory behind transformers in order to form a mathematical background and apply that background to my ISM original work projects with transformers!