

Gradescope (4 pts)

1. Make a brief introduction about one of [the variant of Transformer](#), and use a image of the structure of the model to help explain (2 pts)
2. Briefly explain what's the advantages of this variant under certain situations. (2 pts)