ScriptChain – Technical Assessment

1. *Suppose that we design a deep architecture to represent a sequence by stacking self-attention layers with positional encoding. What could be the issues? (paragraph format)*

- Stacking multiple self-attention layers increases the model's complexity and computational cost. Self-attention, especially in large sequences, can be computationally expensive due to the calculation of attention scores for each pair of positions in the input sequence.
- Deep models with a large number of parameters are prone to overfitting, especially when trained on smaller datasets.
- Deep architectures, especially with many layers, can suffer from vanishing gradients during training, making it difficult for the model to learn.
- Also, it makes it difficult to parallelize the computation, making GPUs run out of memory as well.

A study has been done to highlight that stacking of self- attention layers leads to exploding gradient effect and also potentially overfitting the model.

2. *Can you design a learnable positional encoding method using pytorch? (Create dummy dataset)*

   Designed a learnable positional encoding method using pytorch to classify texts by creating a dummy dataset and evaluated using accuracy – 57 percent.