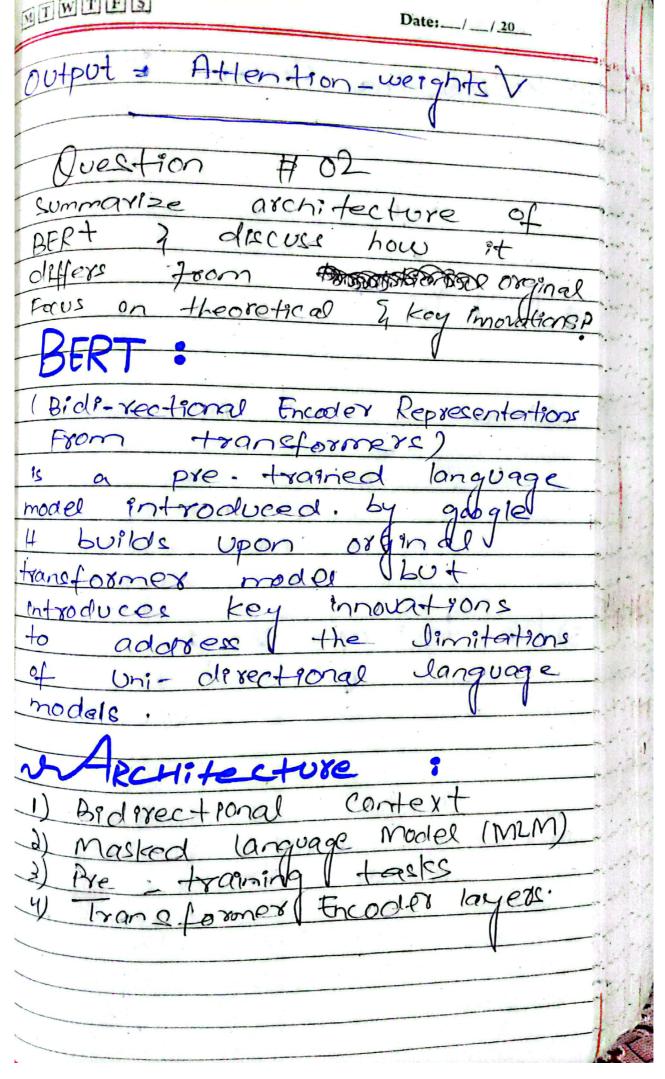
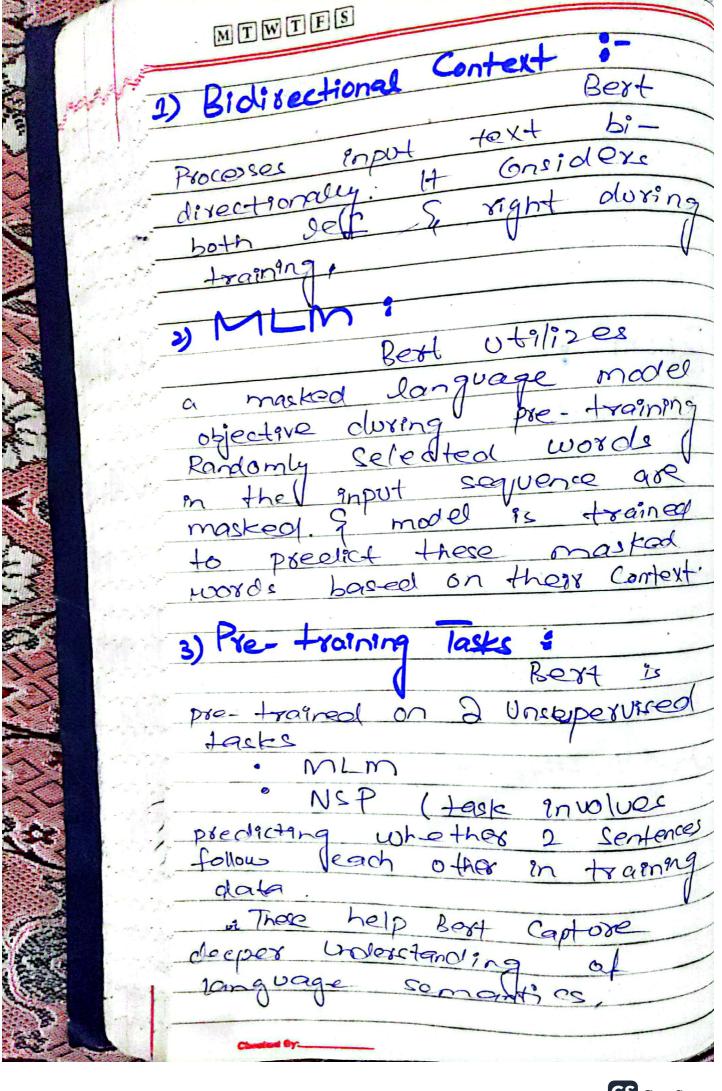
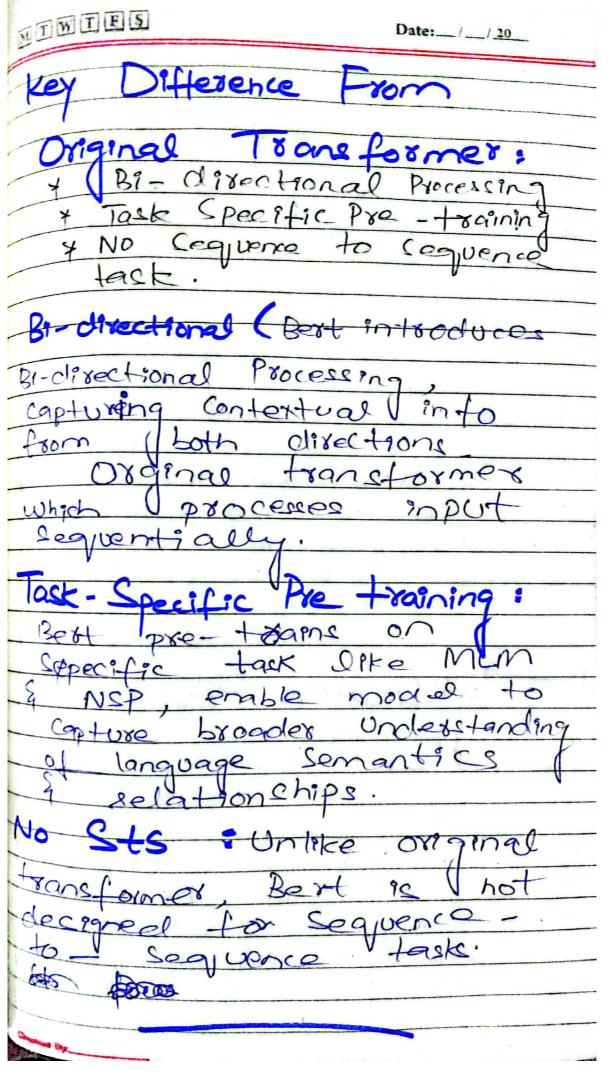
Date: _/_/20 MTWTFS Hssignment # in Concept of Self-attention Transformer mode: How differ in mathematical Quest-800 Explain HHention Self - Attention component of the transformer was introduced by Vasuumi et al - (I'm Paper Attention is All you Need? differ from traditional attention mechanism in its ability to capture relationships different positions between sequence, allowing in a the model to focus on sequence when procluding sequence tasks Scopuenco - Ito. like machine translation. Traditional Vs Self
Attention Meetanism

Date:_/_/20 MTWTFS athmati coulle cequence in put Give x22000 20 Query compute me matrices value IK X WO learnable X WK weights matrice celf- attention Scores are Product dot computed USPAG Overy (0, K)Attention dK Vector dimension Scores gre Scaled Saffmax Attention (OK weights = Saftmax final output 2 a weighted 2000 Values ver tors







MTWTES Scherio Consider a where tognstormer with ctronggling wo Sequence 36500 address this 400 X a explain why could approach effective. To address the issue of long-range dependencies in a transformer model, one effective, modification 19 to incorporate addition positional information in the 'c Relative Posytional enco ding 9 lodification Proposel In the original transformed model, Positional encoding are added to input embeddings to Provide intornation about the positions of each tokens in the sequence

MIWIES Date: __/__/20__ Howevor, these 950010 te positional encoding may effectively not relative distances the tokens b/w 2 especialu 1000 91 The proposed modification involves introducing relative poci tronal emkodings explicitly model the " relative distande blu +DKens. Thic can be achieved h corporating trainable pavame for s for encoding Yelative POST TONS, allowing model to eger the importance of different Postional during yelationships +809n9n9 Could be improved capture of lar Dependencies Holaptabality to sequence lengths Enhanced training Relative Positional encodings mable the +0 model Capture between the relationships toens distances at

enformation. added This Understand especially model effectivele nose where Sotuations are depon dencies posotion a absolute 2) Unlike 980 which encoclings Seglience regardibes encodinge positional nelative dynamically adapt difference tragnable antroduction relative parameters 108 introduces positional encoding duren additional flex9b9 19H Tragallo