Multi-head Attention Mechanism Oplifding oftention machinism methole breds into Great oper ching Independently co) Stadeins muliple single book extention levers O Empremently mild-beed extrendion involves creeting multiple instances of the self-attention mehenson; each with own weights, and then combining their outputs DIM's can be computationally intensive, but of malers LIN's powerful at complex pattern recognition desless Embedded sopul folcos servita enchanged weigh weight actory Forts X/ weight medoly mostald Wva wheat 5th cir Coppt) erc shown 95 they's Ocerits compre Bustead of sow have private Ro 2 aug what materix We Install morries header atta hecse Context 2 metrices For MIH Ineque aton W 2 Combined inelshit modellog conjust Including veders 2 cusal and [7/-1/7/4] 21.7 dropat mass GXX Now here consponding to 5th Inpet 2 sets of context vectors that was highlighted in imputs X

