

????-????????

????????

1.
 1.
 1. dropout epoch LLM
 2.
 1.

????

1.
2.

????????

1.
 1. RoPE
 2.
 3.
 1.
 1. GELU Relu Swish
 2. GQA / MOE / GLU Gated Linear Unit
 2.
 1. ResNet
 2. MultiHEAD
 3. vs
 1. / RMSNorm /
 2.
 4. vs
 1.
 4.
 1.
 1.
 2.
 3.
 2.

