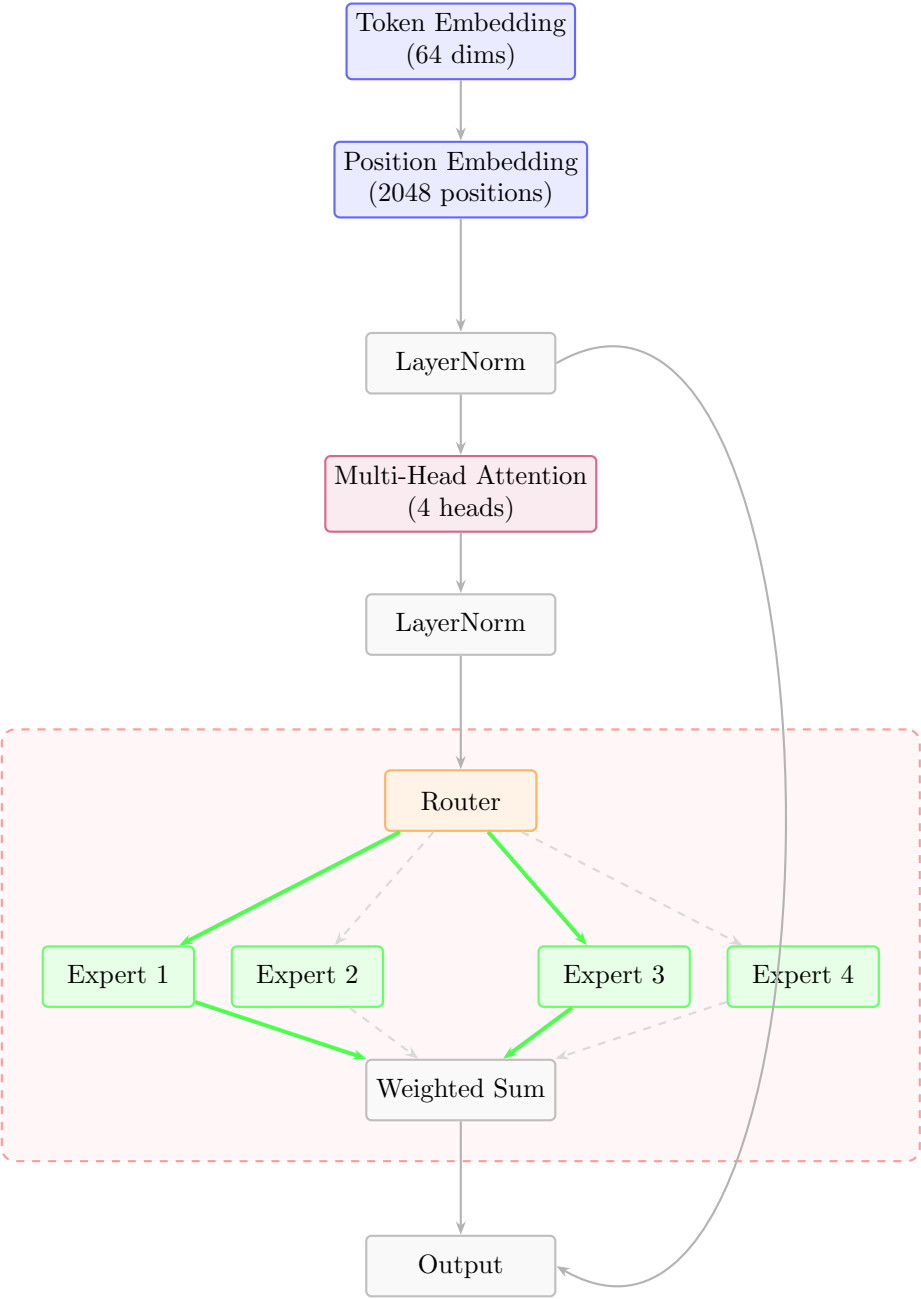


GPT2_small_4exp

MoE Layer



MoE Routing:

- Active Expert
- -> Inactive Expert

Config:
n_layer = 1
n_embd = 64
n_head = 4