

 [Topics](#) [My posts](#) [More](#)▼ [CATEGORIES](#)[General](#)[Site Feedback](#) [All categories](#)[Skip to main content](#)

2025 USA-NA-AIO Round 2, Problem 2, Part 7

USAAIO 

May 2025

Part 7 (10 points, coding task)

In this part, please build your own GQA module called `MyGQA`.

- The requirement is pretty much the same as Part 5.
- Do NOT create H/G copies of key-projection and value-projection matrices. Otherwise, you will use too much unnecessary memory.
- No loop is allowed.

USAAIO 

May 2025

```
### WRITE YOUR SOLUTION HERE ###
```

```
class MyGQA(nn.Module):
    def __init__(self, D_1, D_2, D_qk, D_v, H, G):
        super().__init__()
        self.D_1 = D_1
        self.D_2 = D_2
        self.D_qk = D_qk
        self.D_v = D_v
        self.H = H
        self.G = G

        self.W_Q = nn.Linear(in_features=D_1, out_features=H*D_qk, bias=False)
        self.W_K = nn.Linear(in_features=D_2, out_features=G*D_qk, bias=False)
        self.W_V = nn.Linear(in_features=D_2, out_features=G*D_v, bias=False)
        self.W_O = nn.Linear(in_features=H*D_v, out_features=D_1, bias=False)
```

```

def forward(self, x, y):
    B = x.shape[0] # batch size
    L_1 = x.shape[1] # the length of sequence x
    L_2 = y.shape[1] # the length of sequence y
    num_copies = self.H // self.G

    Q = self.W_Q(x) # shape: (B,L_1,H*D_qk)
    K = self.W_K(y) # shape: (B,L_2,G*D_qk)
    V = self.W_V(y) # shape: (B,L_2,G*D_v)

    Q = Q.reshape(B,L_1,num_copies,self.G,self.D_qk) # shape: (B,L_1,num_copies,G,D_qk)
    K = K.reshape(B,L_2,1,self.G,self.D_qk) # shape: (B,L_2,1,G,D_qk)
    V = V.reshape(B,L_2,1,self.G,self.D_v) # shape: (B,L_2,1,G,D_v)

    Q = Q.permute(0,2,3,1,4) # shape: (B,num_copies,G,L_1,D_qk)
    K = K.permute(0,2,3,1,4) # shape: (B,1,G,L_2,D_qk)
    V = V.permute(0,2,3,1,4) # shape: (B,1,G,L_2,D_v)

    logits = Q @ K.transpose(-2,-1) / (self.D_qk**0.5) # shape: (B,num_copies,G,L_1,L_2)
    alpha = torch.softmax(logits, dim=-1) # shape: (B,num_copies,G,L_1,L_2)

    O = alpha @ V # shape: (B,num_copies,G,L_1,D_v)

    O = O.permute(0,3,1,2,4) # shape: (B,L_1,num_copies,G,D_v)
    O = O.reshape(B,L_1,-1) # shape: (B,L_1,H*D_v)
    return self.W_O(O) # shape: (B,L_1,D_1)

""" END OF THIS PART """

```

[Skip to main content](#)
[✦ Related topics](#)

Topic	Replies	Activity
2025 USA-NA-AIO Round 2, Problem 2, Part 5	1	May 2025
2025 USA-NA-AIO Round 2, Problem 2, Part 6	2	May 2025
2025 USA-NA-AIO Round 2, Problem 2, Part 10	1	May 2025
2025 USA-NA-AIO Round 2, Problem 2, Part 2	2	Dec 2025
2025 USA-NA-AIO Round 2, Problem 2, Part 1	2	Dec 2025

 Powered by Discourse