

```
rag.question_encoder.question_encoder.bert_model.embeddings.word_embeddings.weight, torch.Size([30522, 768])
rag.question_encoder.question_encoder.bert_model.embeddings.position_embeddings.weight, torch.Size([512, 768])
rag.question_encoder.question_encoder.bert_model.embeddings.token_type_embeddings.weight, torch.Size([2, 768])
rag.question_encoder.question_encoder.bert_model.embeddings.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.embeddings.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.attention.self.query.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.attention.self.query.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.attention.self.key.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.attention.self.key.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.attention.self.value.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.attention.self.value.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.attention.output.dense.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.attention.output.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.attention.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.attention.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.intermediate.dense.weight, torch.Size([3072, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.intermediate.dense.bias, torch.Size([3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.output.dense.weight, torch.Size([768, 3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.output.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.0.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.attention.self.query.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.attention.self.query.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.attention.self.key.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.attention.self.key.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.attention.self.value.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.attention.self.value.bias, torch.Size([768])
```

```
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.attention.output.dense.weight, torch.Size([768, 768]))
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.attention.output.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.attention.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.attention.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.intermediate.dense.weight, torch.Size([3072, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.intermediate.dense.bias, torch.Size([3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.output.dense.weight, torch.Size([768, 3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.output.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.1.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.attention.self.query.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.attention.self.query.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.attention.self.key.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.attention.self.key.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.attention.self.value.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.attention.self.value.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.attention.output.dense.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.attention.output.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.attention.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.attention.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.intermediate.dense.weight, torch.Size([3072, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.intermediate.dense.bias, torch.Size([3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.output.dense.weight, torch.Size([768, 3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.output.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.2.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.3.attention.self.query.weight, torch.Size([768, 768])
```

[illegible]

[illegible]

```
rag.question_encoder.question_encoder.bert_model.encoder.layer.6.attention.output.dense.bias, torch.Size([768]))
rag.question_encoder.question_encoder.bert_model.encoder.layer.6.attention.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.6.attention.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.6.intermediate.dense.weight, torch.Size([3072, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.6.intermediate.dense.bias, torch.Size([3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.6.output.dense.weight, torch.Size([768, 3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.6.output.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.6.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.6.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.attention.self.query.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.attention.self.query.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.attention.self.key.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.attention.self.key.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.attention.self.value.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.attention.self.value.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.attention.output.dense.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.attention.output.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.attention.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.attention.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.intermediate.dense.weight, torch.Size([3072, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.intermediate.dense.bias, torch.Size([3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.output.dense.weight, torch.Size([768, 3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.output.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.7.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.8.attention.self.query.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.8.attention.self.query.bias, torch.Size([768])
```

[illegible]

```
rag.question_encoder.question_encoder.bert_model.encoder.layer.9.out
put.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.9.out
put.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.9.out
put.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.at
tention.self.query.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.at
tention.self.query.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.at
tention.self.key.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.at
tention.self.key.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.at
tention.self.value.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.at
tention.self.value.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.at
tention.output.dense.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.at
tention.output.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.at
tention.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.at
tention.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.in
termediate.dense.weight, torch.Size([3072, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.in
termediate.dense.bias, torch.Size([3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.ou
tput.dense.weight, torch.Size([768, 3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.ou
tput.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.ou
tput.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.10.ou
tput.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.at
tention.self.query.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.at
tention.self.query.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.at
tention.self.key.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.at
tention.self.key.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.at
tention.self.value.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.at
tention.self.value.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.at
tention.output.dense.weight, torch.Size([768, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.at
tention.output.dense.bias, torch.Size([768])
```

```
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.attention.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.attention.output.LayerNorm.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.intermediate.dense.weight, torch.Size([3072, 768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.intermediate.dense.bias, torch.Size([3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.output.dense.weight, torch.Size([768, 3072])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.output.dense.bias, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.output.LayerNorm.weight, torch.Size([768])
rag.question_encoder.question_encoder.bert_model.encoder.layer.11.output.LayerNorm.bias, torch.Size([768])
rag.generator.model.shared.weight, torch.Size([50265, 1024])
rag.generator.model.encoder.embed_positions.weight, torch.Size([1026, 1024])
rag.generator.model.encoder.layers.0.self_attn.k_proj.weight, torch.Size([1024, 1024])
rag.generator.model.encoder.layers.0.self_attn.k_proj.bias, torch.Size([1024])
rag.generator.model.encoder.layers.0.self_attn.v_proj.weight, torch.Size([1024, 1024])
rag.generator.model.encoder.layers.0.self_attn.v_proj.bias, torch.Size([1024])
rag.generator.model.encoder.layers.0.self_attn.q_proj.weight, torch.Size([1024, 1024])
rag.generator.model.encoder.layers.0.self_attn.q_proj.bias, torch.Size([1024])
rag.generator.model.encoder.layers.0.self_attn.out_proj.weight, torch.Size([1024, 1024])
rag.generator.model.encoder.layers.0.self_attn.out_proj.bias, torch.Size([1024])
rag.generator.model.encoder.layers.0.self_attn_layer_norm.weight, torch.Size([1024])
rag.generator.model.encoder.layers.0.self_attn_layer_norm.bias, torch.Size([1024])
rag.generator.model.encoder.layers.0.fc1.weight, torch.Size([4096, 1024])
rag.generator.model.encoder.layers.0.fc1.bias, torch.Size([4096])
rag.generator.model.encoder.layers.0.fc2.weight, torch.Size([1024, 4096])
rag.generator.model.encoder.layers.0.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.0.final_layer_norm.weight, torch.Size([1024])
rag.generator.model.encoder.layers.0.final_layer_norm.bias, torch.Size([1024])
rag.generator.model.encoder.layers.1.self_attn.k_proj.weight, torch.Size([1024, 1024])
rag.generator.model.encoder.layers.1.self_attn.k_proj.bias, torch.Size([1024])
rag.generator.model.encoder.layers.1.self_attn.v_proj.weight,
```



```
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.1.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.1.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.1.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.1.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.1.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.1.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.1.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.1.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.encoder.layers.1.fc1.bias, torch.Size([4096])
rag.generator.model.encoder.layers.1.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.encoder.layers.1.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.1.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.1.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.2.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.2.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.2.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.2.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.2.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.2.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.2.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.2.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.2.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.2.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.2.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.encoder.layers.2.fc1.bias, torch.Size([4096])
rag.generator.model.encoder.layers.2.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.encoder.layers.2.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.2.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.2.final_layer_norm.bias,
```

```
torch.Size([1024])
rag.generator.model.encoder.layers.3.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.3.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.3.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.3.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.3.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.3.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.3.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.3.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.3.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.3.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.3.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.encoder.layers.3.fc1.bias, torch.Size([4096])
rag.generator.model.encoder.layers.3.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.encoder.layers.3.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.3.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.3.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.4.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.4.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.4.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.4.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.4.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.4.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.4.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.4.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.4.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.4.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.4.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.encoder.layers.4.fc1.bias, torch.Size([4096])
```

```
rag.generator.model.encoder.layers.4.fc2.weight, torch.Size([1024,
4096]))
rag.generator.model.encoder.layers.4.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.4.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.4.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.5.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.5.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.5.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.5.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.5.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.5.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.5.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.5.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.5.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.5.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.5.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.encoder.layers.5.fc1.bias, torch.Size([4096])
rag.generator.model.encoder.layers.5.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.encoder.layers.5.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.5.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.5.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.6.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.6.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.6.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.6.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.6.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.6.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.6.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.6.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.6.self_attn_layer_norm.weight,
```

```
torch.Size([1024])
rag.generator.model.encoder.layers.6.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.6.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.encoder.layers.6.fc1.bias, torch.Size([4096])
rag.generator.model.encoder.layers.6.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.encoder.layers.6.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.6.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.6.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.7.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.7.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.7.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.7.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.7.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.7.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.7.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.7.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.7.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.7.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.7.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.encoder.layers.7.fc1.bias, torch.Size([4096])
rag.generator.model.encoder.layers.7.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.encoder.layers.7.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.7.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.7.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.8.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.8.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.8.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.8.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.8.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.8.self_attn.q_proj.bias,
```

```
torch.Size([1024])
rag.generator.model.encoder.layers.8.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.8.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.8.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.8.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.8.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.encoder.layers.8.fc1.bias, torch.Size([4096])
rag.generator.model.encoder.layers.8.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.encoder.layers.8.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.8.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.8.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.9.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.9.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.9.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.9.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.9.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.9.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.9.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.9.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.9.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.9.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.9.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.encoder.layers.9.fc1.bias, torch.Size([4096])
rag.generator.model.encoder.layers.9.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.encoder.layers.9.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.9.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.9.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.10.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.10.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.10.self_attn.v_proj.weight,
```

```
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.10.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.10.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.10.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.10.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.10.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.10.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.10.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.10.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.encoder.layers.10.fc1.bias, torch.Size([4096])
rag.generator.model.encoder.layers.10.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.encoder.layers.10.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.10.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.10.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.11.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.11.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.11.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.11.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.11.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.11.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.11.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.encoder.layers.11.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.11.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.11.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.encoder.layers.11.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.encoder.layers.11.fc1.bias, torch.Size([4096])
rag.generator.model.encoder.layers.11.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.encoder.layers.11.fc2.bias, torch.Size([1024])
rag.generator.model.encoder.layers.11.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.encoder.layers.11.final_layer_norm.bias,
```

```
torch.Size([1024])
rag.generator.model.encoder.layernorm_embedding.weight,
torch.Size([1024])
rag.generator.model.encoder.layernorm_embedding.bias,
torch.Size([1024])
rag.generator.model.decoder.embed_positions.weight,
torch.Size([1026, 1024])
rag.generator.model.decoder.layers.0.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.0.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.0.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.0.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.0.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.0.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.0.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.0.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.0.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.0.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.0.encoder_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.0.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.0.encoder_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.0.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.0.encoder_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.0.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.0.encoder_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.0.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.0.encoder_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.0.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.0.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.0.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.0.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.0.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.0.final_layer_norm.weight,
```

```
torch.Size([1024])
rag.generator.model.decoder.layers.0.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.1.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.1.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.1.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.1.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.1.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.1.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.1.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.1.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.1.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.1.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.1.encoder_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.1.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.1.encoder_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.1.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.1.encoder_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.1.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.1.encoder_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.1.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.1.encoder_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.1.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.1.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.1.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.1.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.1.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.1.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.1.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.2.self_attn.k_proj.weight,
```



```
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.2.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.2.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.2.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.2.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.2.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.2.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.2.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.2.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.2.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.2.encoder_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.2.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.2.encoder_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.2.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.2.encoder_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.2.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.2.encoder_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.2.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.2.encoder_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.2.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.2.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.2.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.2.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.2.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.2.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.2.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.3.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.3.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.3.self_attn.v_proj.weight,
```

```
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.3.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.3.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.3.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.3.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.3.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.3.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.3.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.3.encoder_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.3.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.3.encoder_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.3.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.3.encoder_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.3.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.3.encoder_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.3.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.3.encoder_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.3.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.3.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.3.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.3.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.3.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.3.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.3.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.4.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.4.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.4.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.4.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.4.self_attn.q_proj.weight,
```

```
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.4.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.4.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.4.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.4.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.4.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.4.encoder_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.4.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.4.encoder_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.4.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.4.encoder_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.4.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.4.encoder_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.4.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.4.encoder_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.4.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.4.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.4.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.4.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.4.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.4.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.4.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.5.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.5.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.5.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.5.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.5.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.5.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.5.self_attn.out_proj.weight,
```

```
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.5.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.5.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.5.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.5.encoder_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.5.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.5.encoder_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.5.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.5.encoder_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.5.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.5.encoder_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.5.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.5.encoder_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.5.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.5.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.5.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.5.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.5.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.5.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.5.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.6.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.6.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.6.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.6.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.6.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.6.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.6.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.6.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.6.self_attn_layer_norm.weight,
```

```
torch.Size([1024])
rag.generator.model.decoder.layers.6.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.6.encoder_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.6.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.6.encoder_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.6.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.6.encoder_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.6.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.6.encoder_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.6.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.6.encoder_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.6.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.6.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.6.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.6.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.6.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.6.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.6.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.7.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.7.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.7.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.7.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.7.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.7.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.7.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.7.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.7.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.7.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.7.encoder_attn.k_proj.weight,
```

```
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.7.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.7.encoder_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.7.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.7.encoder_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.7.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.7.encoder_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.7.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.7.encoder_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.7.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.7.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.7.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.7.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.7.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.7.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.7.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.8.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.8.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.8.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.8.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.8.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.8.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.8.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.8.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.8.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.8.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.8.encoder_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.8.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.8.encoder_attn.v_proj.weight,
```

```
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.8.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.8.encoder_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.8.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.8.encoder_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.8.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.8.encoder_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.8.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.8.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.8.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.8.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.8.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.8.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.8.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.9.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.9.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.9.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.9.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.9.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.9.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.9.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.9.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.9.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.9.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.9.encoder_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.9.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.9.encoder_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.9.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.9.encoder_attn.q_proj.weight,
```

```
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.9.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.9.encoder_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.9.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.9.encoder_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.9.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.9.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.9.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.9.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.9.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.9.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.9.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.10.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.10.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.10.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.10.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.10.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.10.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.10.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.10.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.10.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.10.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.10.encoder_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.10.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.10.encoder_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.10.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.10.encoder_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.10.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.10.encoder_attn.out_proj.weight,
```



```
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.10.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.10.encoder_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.10.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.10.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.10.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.10.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.10.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.10.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.10.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.11.self_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.11.self_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.11.self_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.11.self_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.11.self_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.11.self_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.11.self_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.11.self_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.11.self_attn_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.11.self_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.11.encoder_attn.k_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.11.encoder_attn.k_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.11.encoder_attn.v_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.11.encoder_attn.v_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.11.encoder_attn.q_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.11.encoder_attn.q_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.11.encoder_attn.out_proj.weight,
torch.Size([1024, 1024])
rag.generator.model.decoder.layers.11.encoder_attn.out_proj.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.11.encoder_attn_layer_norm.weight
```

```
, torch.Size([1024])
rag.generator.model.decoder.layers.11.encoder_attn_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layers.11.fc1.weight, torch.Size([4096,
1024])
rag.generator.model.decoder.layers.11.fc1.bias, torch.Size([4096])
rag.generator.model.decoder.layers.11.fc2.weight, torch.Size([1024,
4096])
rag.generator.model.decoder.layers.11.fc2.bias, torch.Size([1024])
rag.generator.model.decoder.layers.11.final_layer_norm.weight,
torch.Size([1024])
rag.generator.model.decoder.layers.11.final_layer_norm.bias,
torch.Size([1024])
rag.generator.model.decoder.layer_norm_embedding.weight,
torch.Size([1024])
rag.generator.model.decoder.layer_norm_embedding.bias,
torch.Size([1024])
```