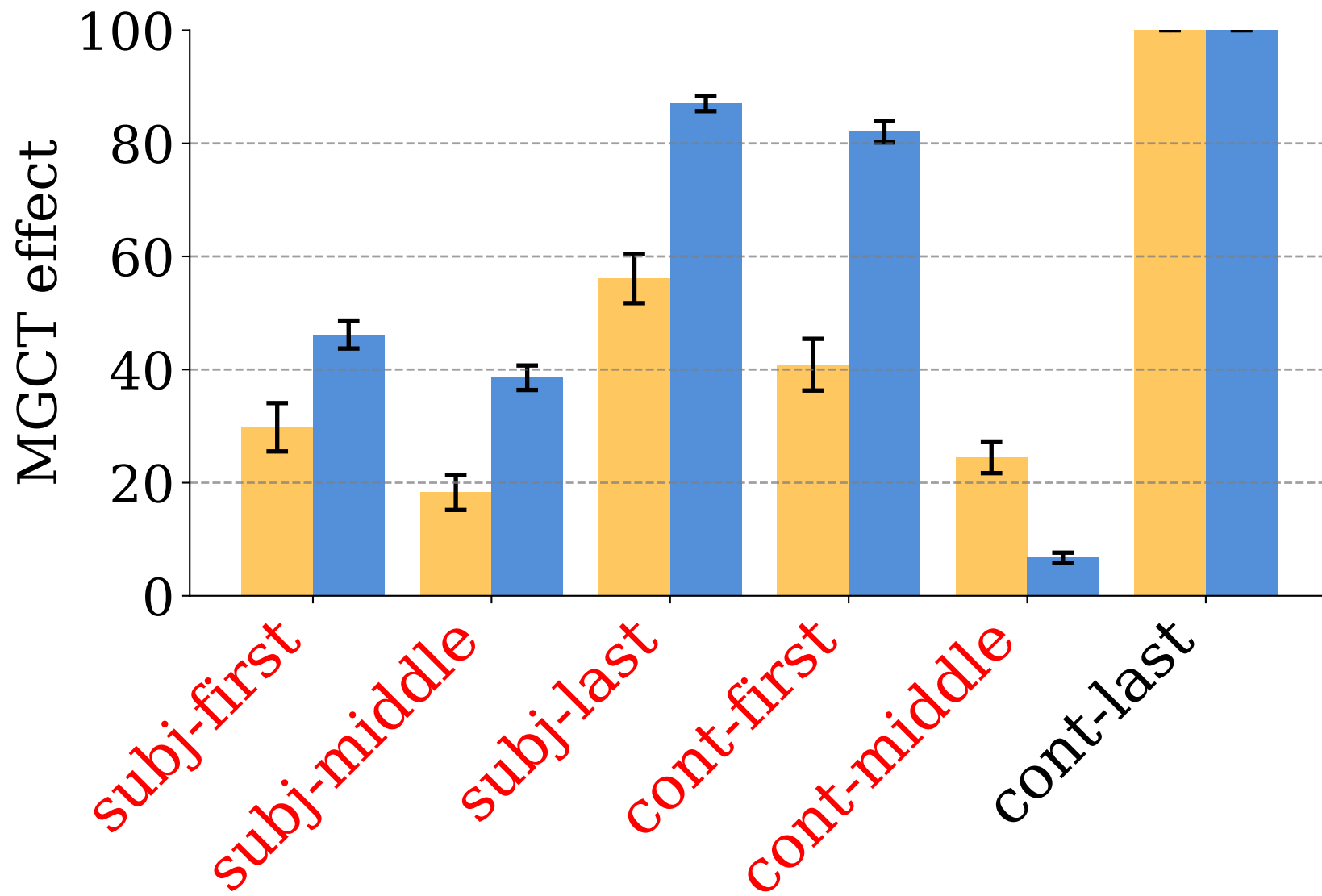
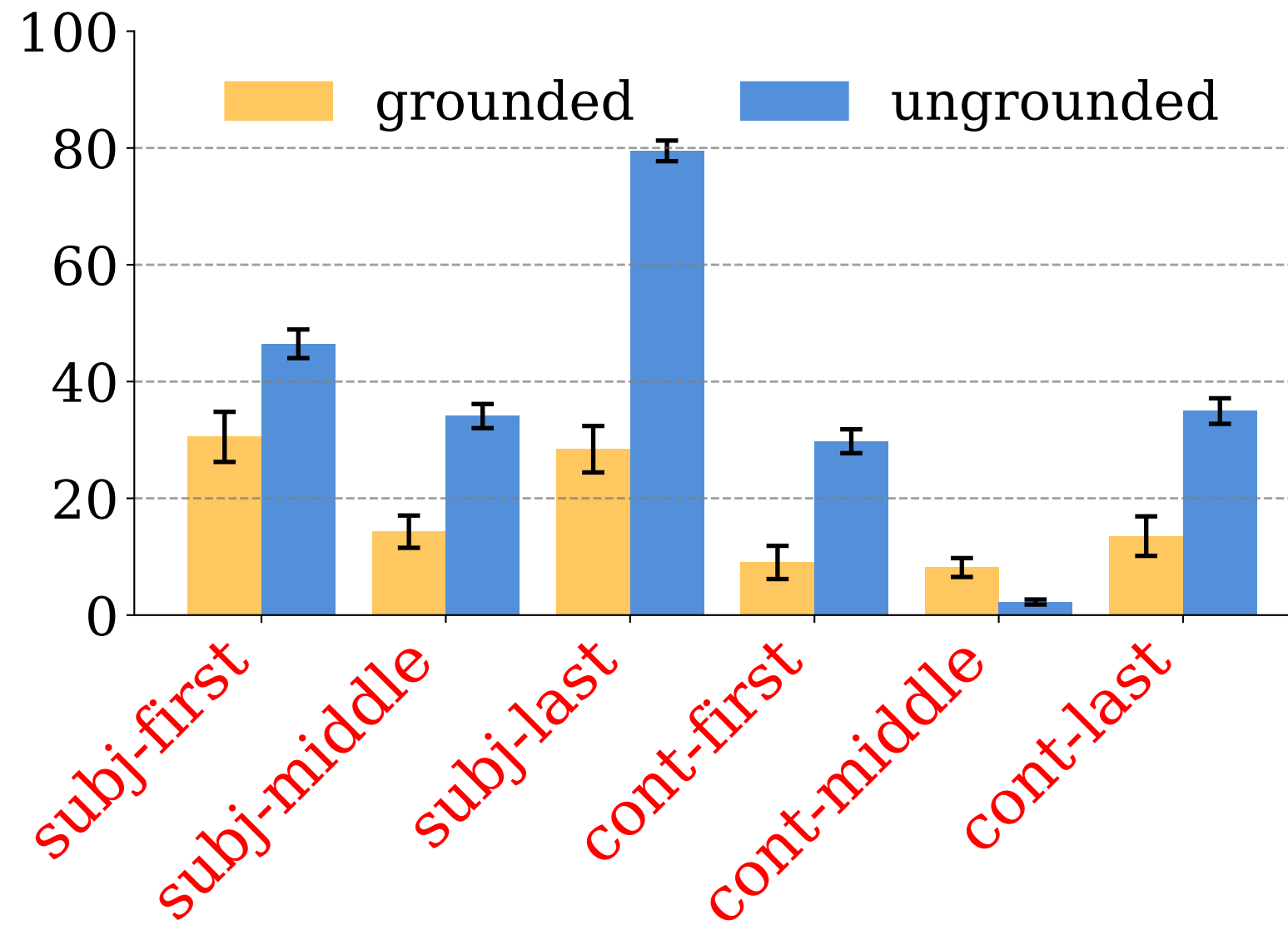


LLaMA-7B (FakePedia-MH)

Hidden activations



MLPs



Attention heads

