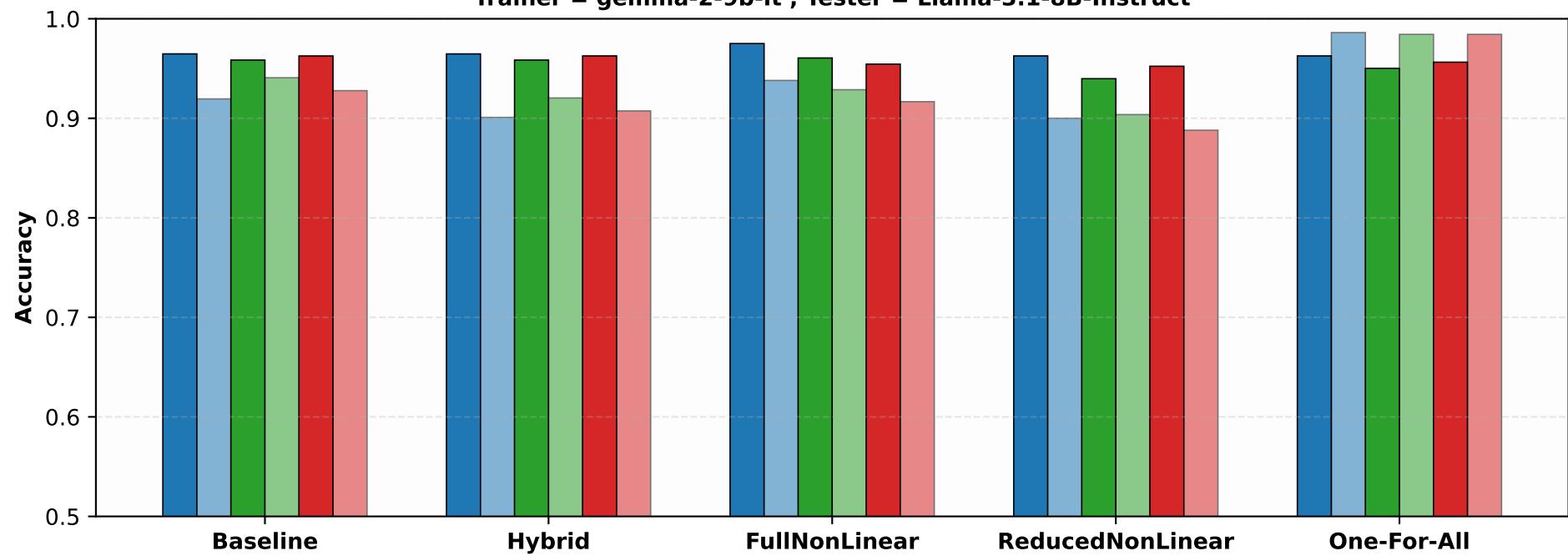
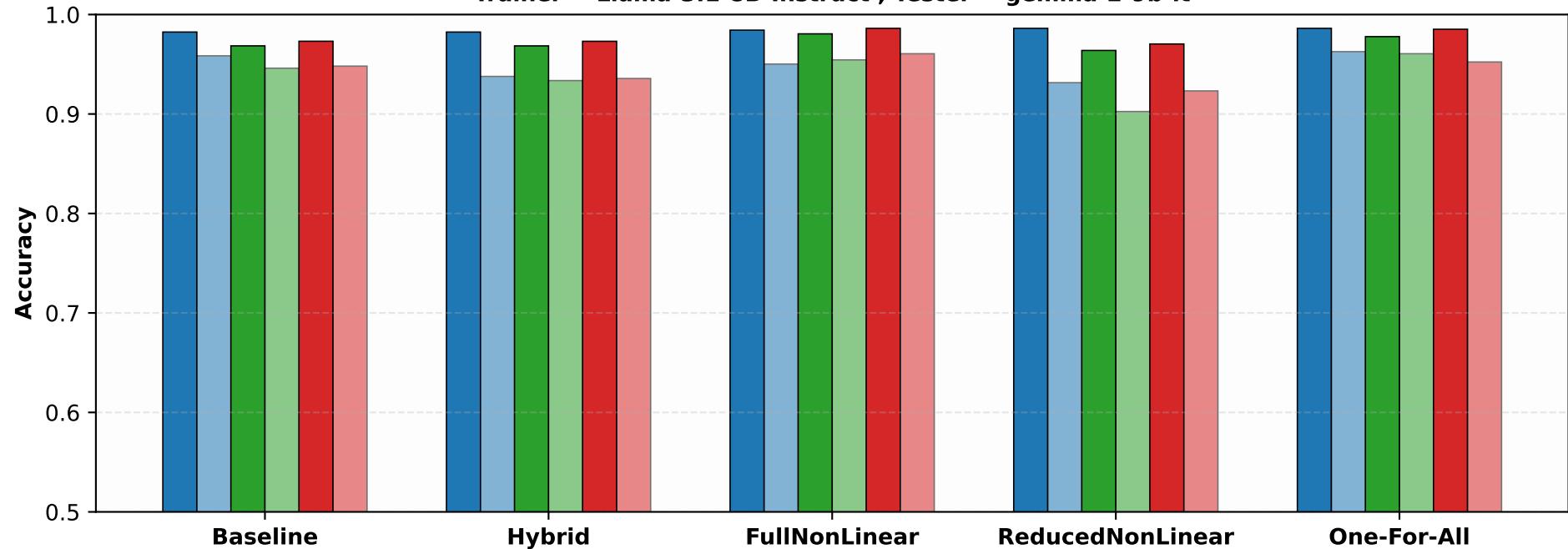


Trainer = gemma-2-9b-it , Tester = Llama-3.1-8B-Instruct



Trainer = Llama-3.1-8B-Instruct , Tester = gemma-2-9b-it



Layer (Model)

■ attn (Trainer) ■ attn (Tester) ■ hidden (Trainer) ■ hidden (Tester) ■ mlp (Trainer) ■ mlp (Tester)