Impact of restoring MLP after corrupted input

High score: 0.010

Low score: 0.000

single restored MLP layer within gpt2