Impact of restoring MLP after corrupted input

High score: 0.002

Low score: 0.001

single restored MLP layer within gpt2