GPT-3.5 Turbo Performance on Stepwise Dataset