Distributions for "attention_output" in gpt2

