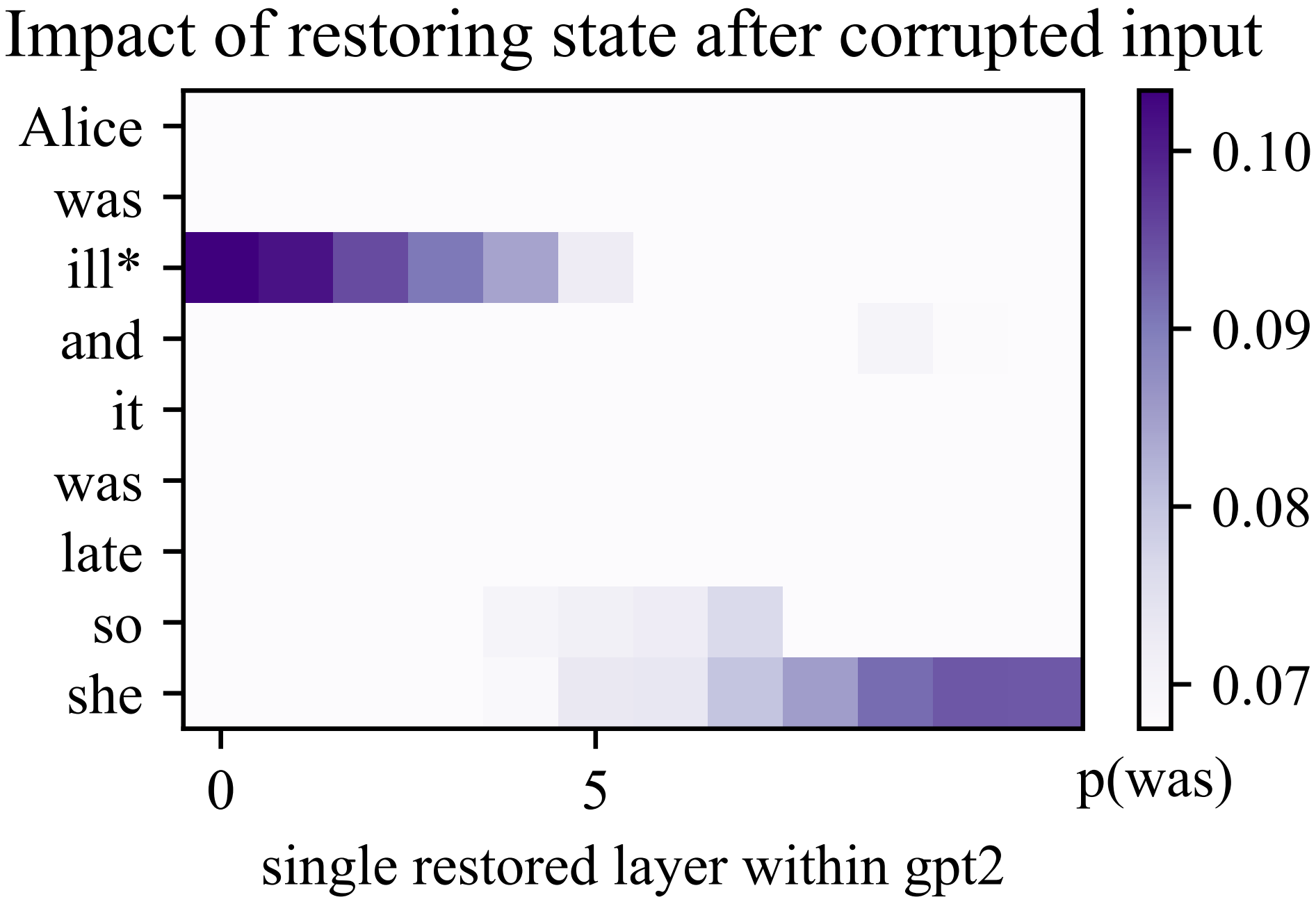
Impact of restoring state after corrupted input
Alice
was
ill*
and
it
was
late
so
she
0
5
single restored layer within gpt2
0.10
0.09
0.08
0.07
p(was)