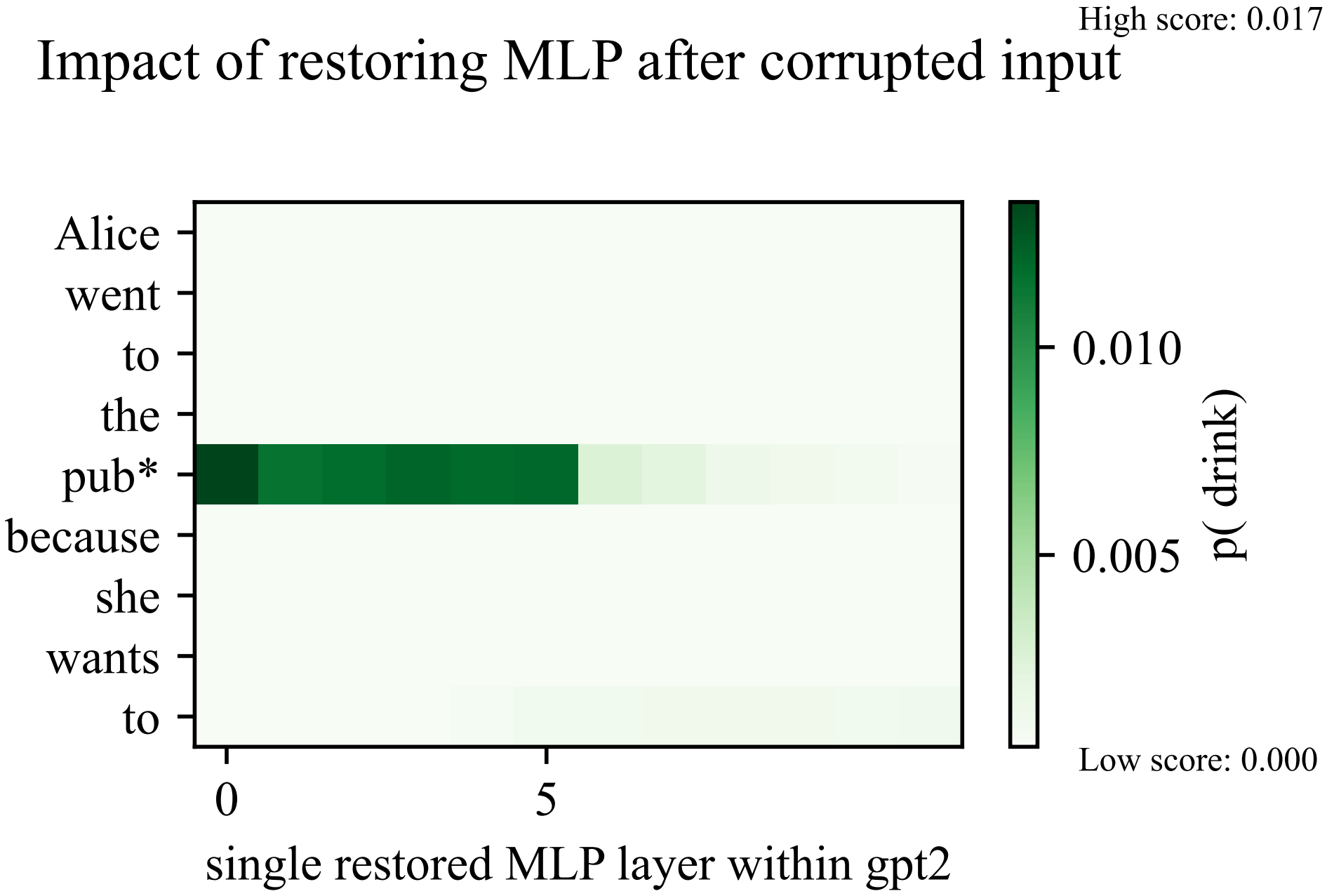
High score: 0.017
Impact of restoring MLP after corrupted input
Alice
went
to
the
pub*
because
she
wants
to
0
5
single restored MLP layer within gpt2
0.010
0.005
p(drink)
Low score: 0.000