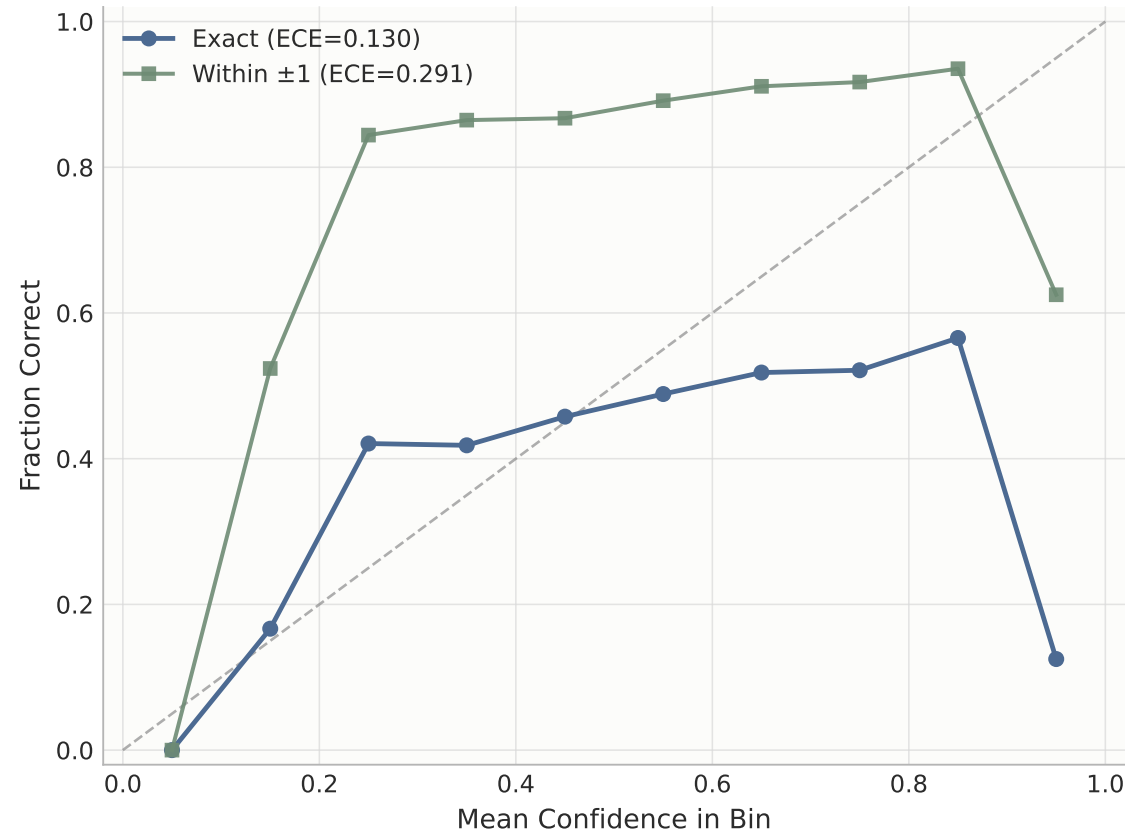


Calibration Analysis GPT 5.1 high reasoning (n=16107)

Reliability Diagram



Confidence Distribution Mean=61.4%

