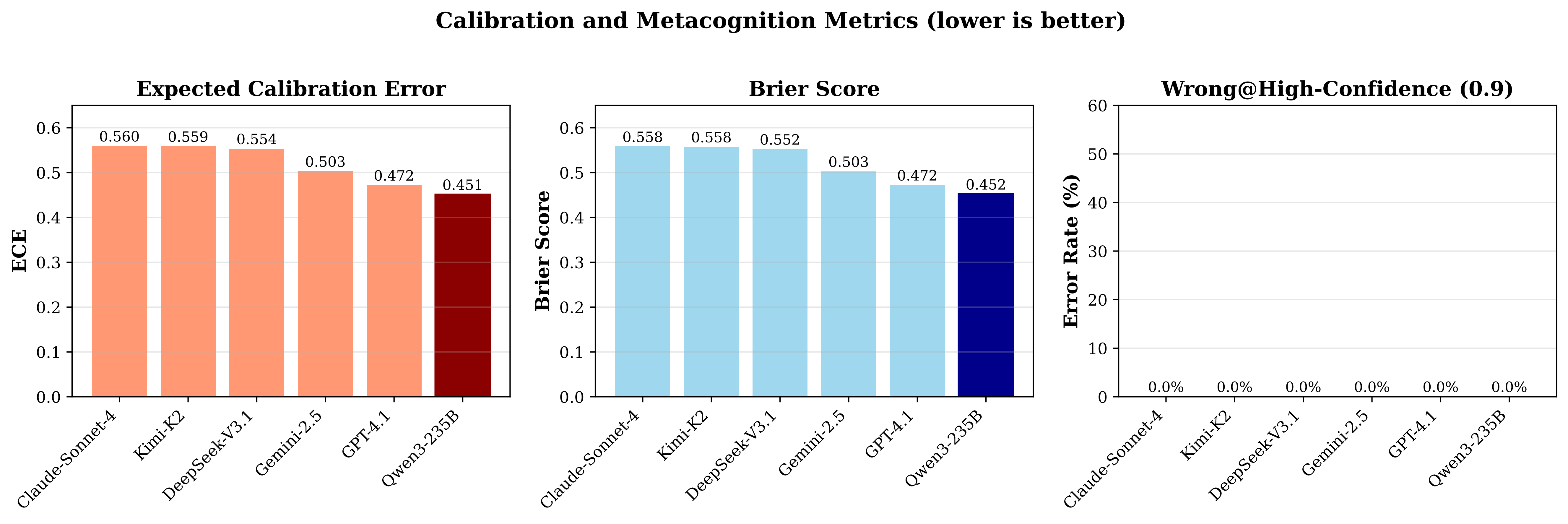

Calibration and Metacognition Metrics (lower is better)

Expected Calibration Error

Brier Score

Wrong@High-Confidence (0.9)