Calibration Analysis: Reliability Diagrams
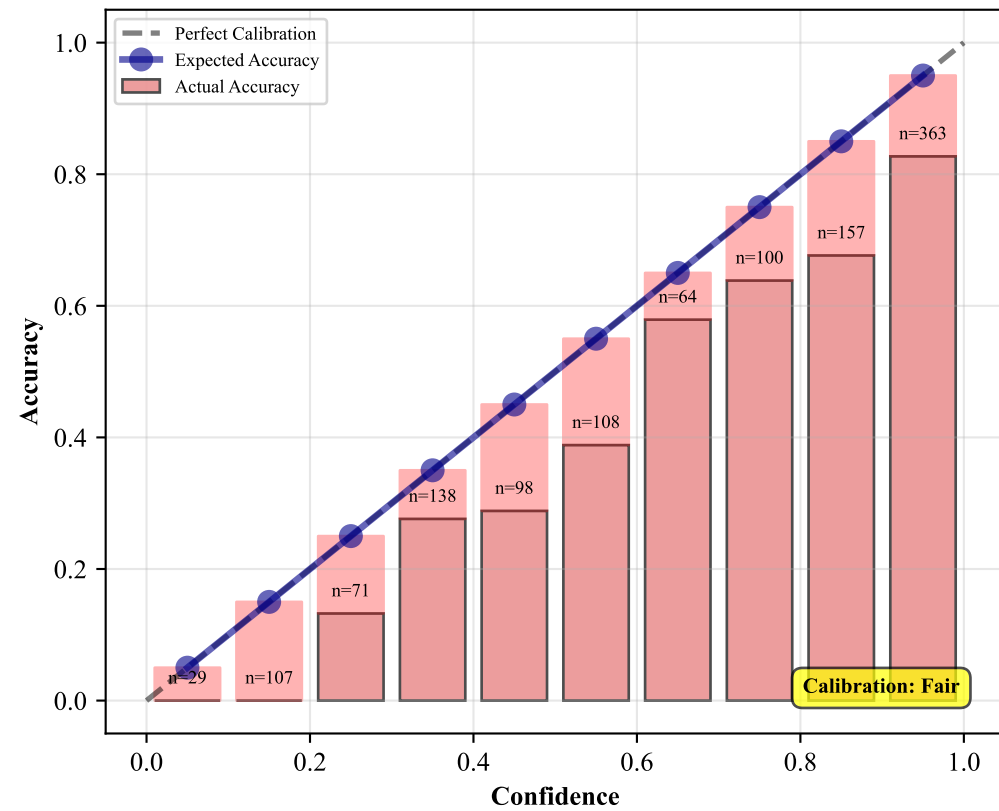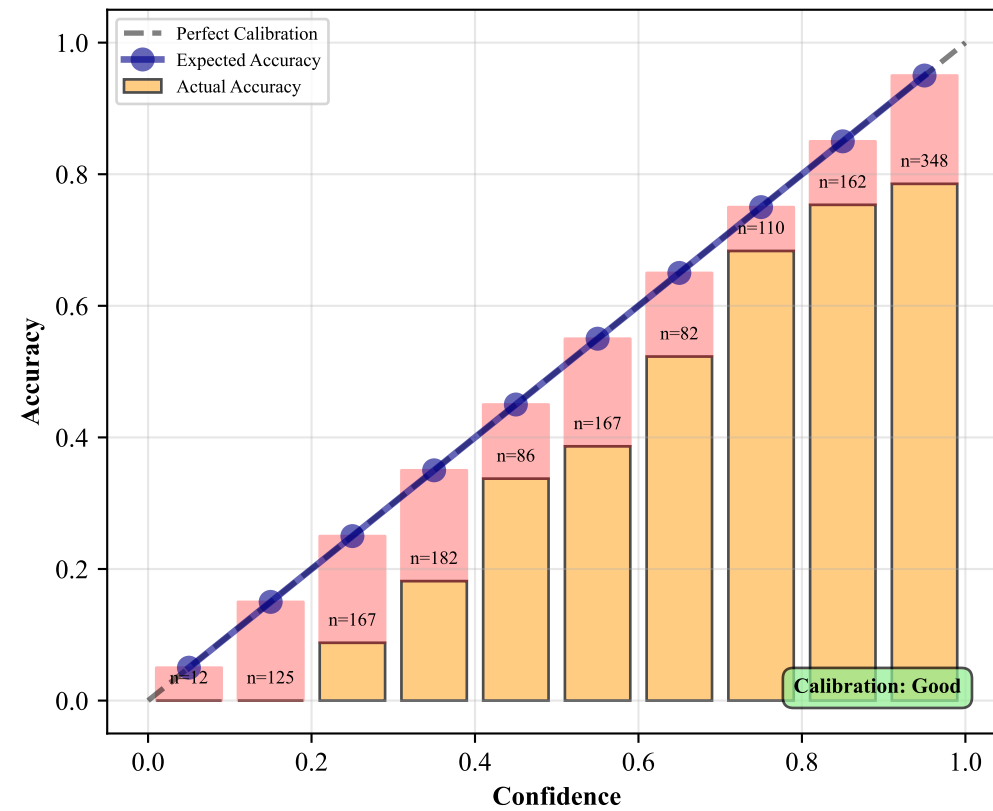**(Expected Calibration Error)**
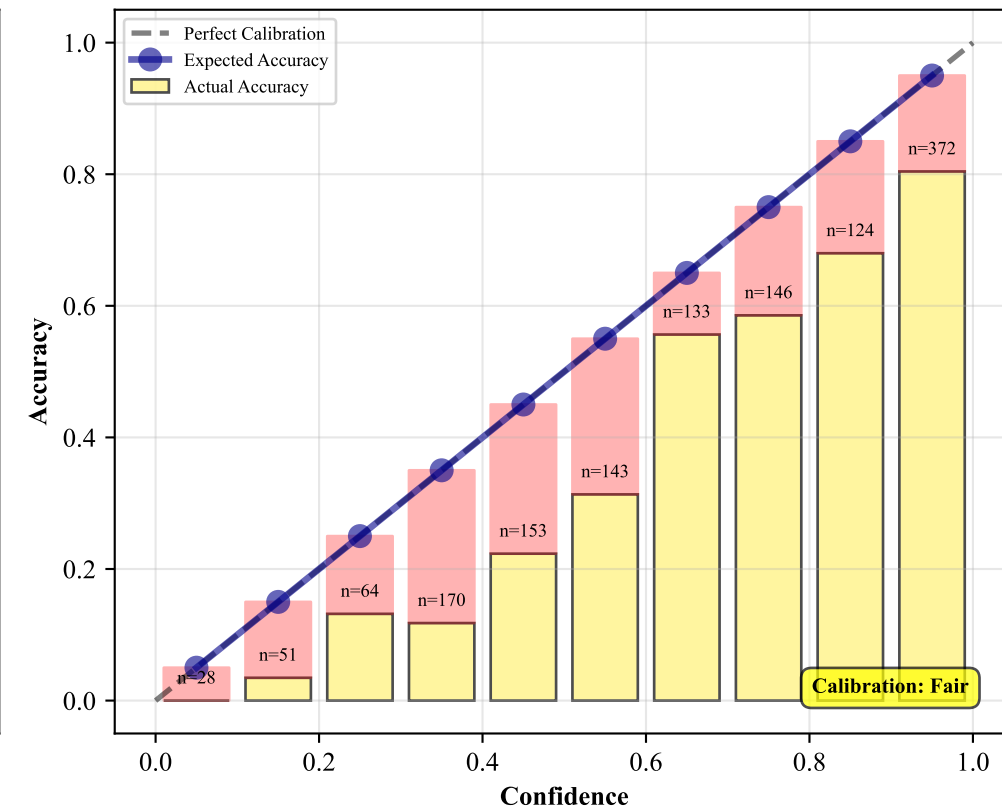
CLIP Baseline
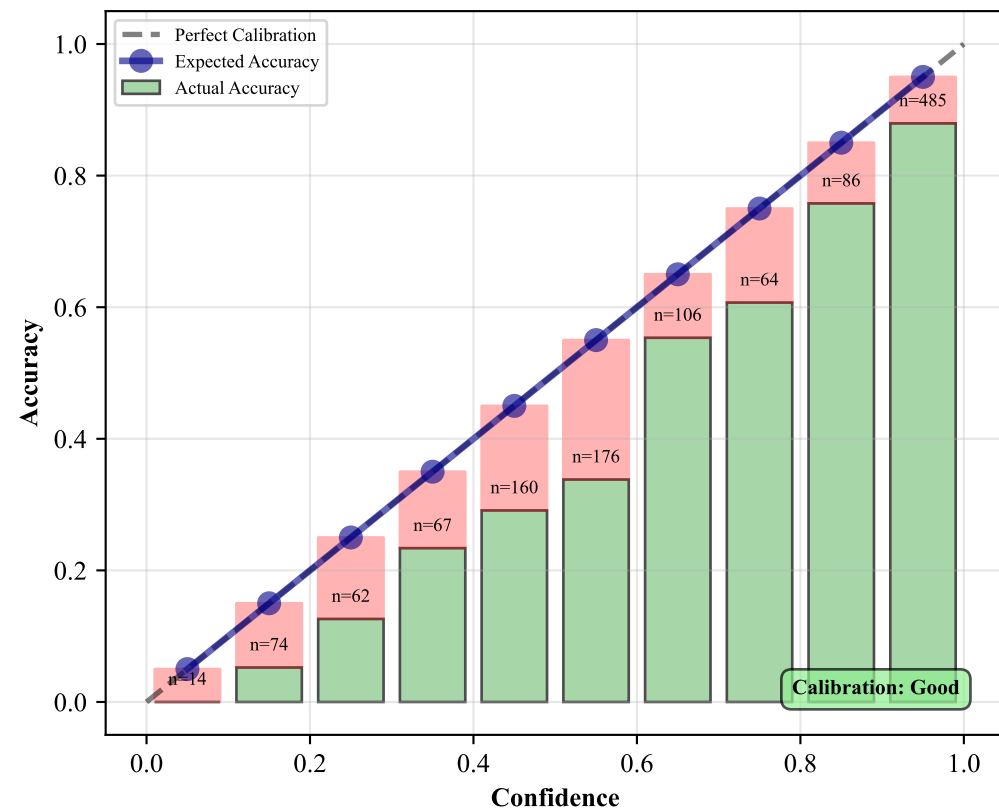**ECE = 0.145**

- Perfect Calibration
- Expected Accuracy
- Actual Accuracy

n=363, n=157, n=100, n=64, n=108, n=98, n=138, n=71, n=107, n=29

Calibration: Fair

AutoCLIP
ECE = 0.098

- Perfect Calibration
- Expected Accuracy
- Actual Accuracy

n=348, n=162, n=110, n=82, n=167, n=86, n=182, n=167, n=125, n=12

Calibration: Good

WaffleCLIP
ECE = 0.112

- Perfect Calibration
- Expected Accuracy
- Actual Accuracy

n=372, n=124, n=146, n=133, n=143, n=153, n=170, n=64, n=51, n=28

Calibration: Fair

Ours (w/o LLM)
ECE = 0.072

- Perfect Calibration
- Expected Accuracy
- Actual Accuracy

n=485, n=86, n=64, n=106, n=176, n=160, n=67, n=62, n=74, n=14

Calibration: Good

Ours (w/o adapt)
ECE = 0.063

- Perfect Calibration
- Expected Accuracy
- Actual Accuracy

n=429, n=52, n=168, n=170, n=164, n=180, n=148, n=80, n=129, n=18

Calibration: Good

Ours (full)
ECE = 0.047

- Perfect Calibration
- Expected Accuracy
- Actual Accuracy

n=425, n=129, n=187, n=131, n=54, n=116, n=84, n=123, n=98, n=29

Calibration: Excellent