REPORT: TEST_N1_M1_BINARY

Model: Qwen2.5-VL-M1-Detection

Date: 2026-02-17 | Dataset: M1 Detection (Binary)

1. PERFORMANCE SUMMARY

Metric	Baseline	Fine-Tuned (Avg 4)
Accuracy	0.7684	0.8108 ± 0.0103
F1 Macro	0.5565	0.7448 ± 0.0123
Prec Macro	0.7003	0.7460 ± 0.0145
Rec Macro	0.5627	0.7442 ± 0.0122
Inv. Rate	0.00%	0.00%
Infer Time (s)	0.664	0.769
Train Time (m)	N/A	391.5 ± 2.3

2. PER-CLASS BREAKDOWN (0=Safe, 1=Offensive)

Metric	Base (0)	Base (1)	FT Avg (0)	FT Avg (1)
Prec	0.778	0.623	0.873 ± 0.007	0.618 ± 0.026
Rec	0.969	0.156	0.876 ± 0.014	0.613 ± 0.025
F1	0.863	0.250	0.875 ± 0.007	0.615 ± 0.018

3. CONFUSION MATRICES (Avg)

Baseline (Zero-Shot)

Fine-Tuned (Avg)