

TrustedAI Risk Report

Evaluated URL: gpt-4o-mini

Dimension	Score	Summary
Fairness	0	Overall acc 0.50, worst gap 0.50
Robustness	100	Acc clean 0.50 → adv 0.50
Privacy	100	MIA accuracy 0.46 (risk low)
Transparency	100	Top tokens: What, s, the, ingredient, of
fairness	0	Overall acc 0.50, worst gap 0.50
robustness	100	Acc clean 0.50 → adv 0.50
privacy	100	MIA accuracy 0.46 (risk low)
transparency	100	Top tokens: What, s, the, ingredient, of