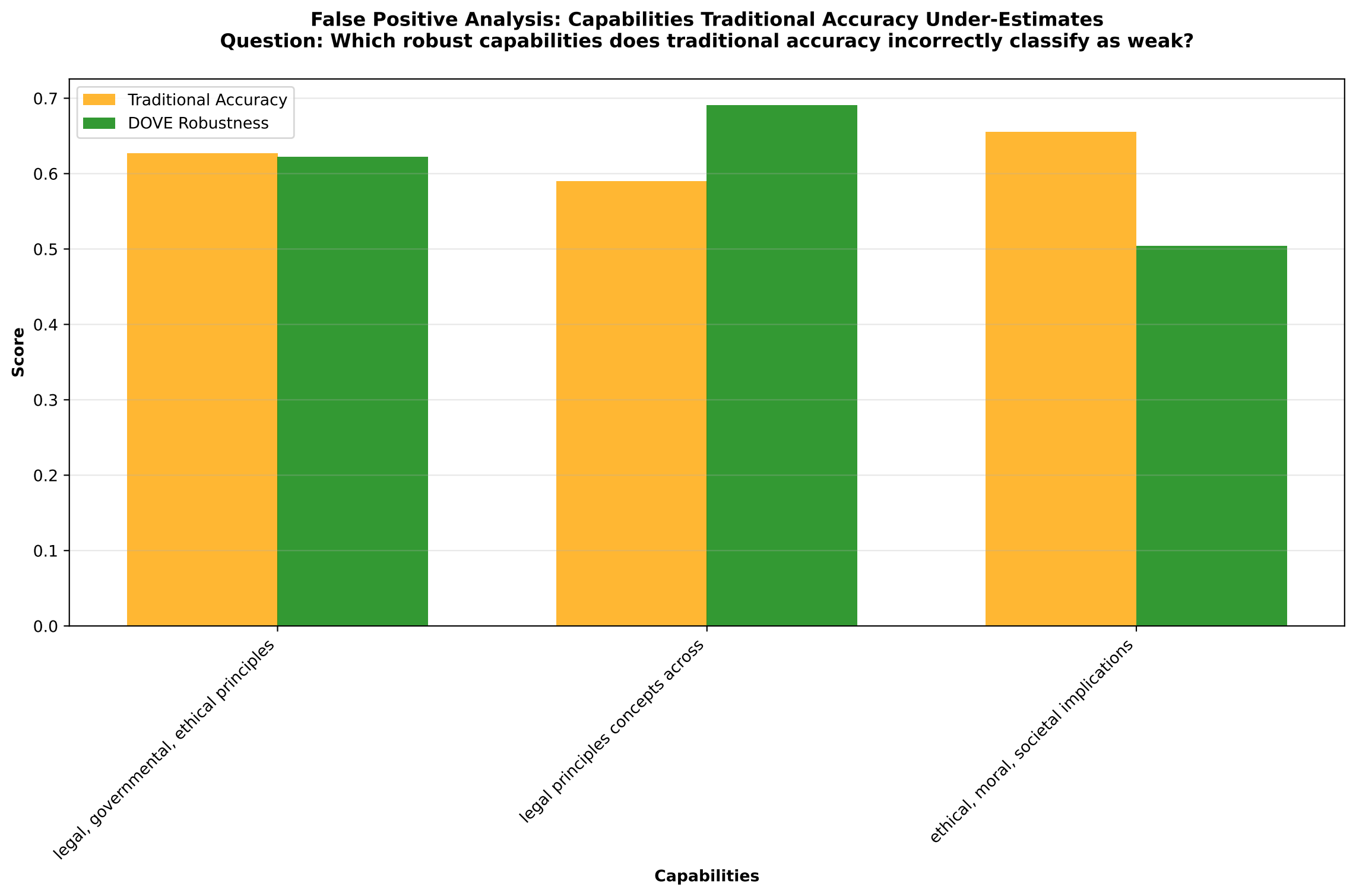
False Positive Analysis: Capabilities Traditional Accuracy Under-Estimates
Question: Which robust capabilities does traditional accuracy incorrectly classify as weak?
Traditional Accuracy
DOVE Robustness
Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
legal, governmental, ethical principles
legal principles concepts across
ethical, moral, societal implications
Capabilities