

# Model Capability vs Adversarial Robustness

Higher capability models show better resistance to adversarial prompts (less negative degradation)

