**Paper Review: MULTITRUST: A Comprehensive Benchmark Towards Trustworthy Multimodal Large Language Models**

By: Lam Nguyen

**1. SUMMARY**

This paper explores how Multimodal Large Language Models (MLLMs) can be easily compromised by query relevant images. To address this, MM-SafetyBench was developed to guard against image-based manipulations that would compromise the security of an MLLM.

**2. STRENGTHS**

**3. WEAKNESSES**

**4. TECHNICAL EXTENSIONS**

**5. OVERALL REVIEW**