Attack Corpus: Methodology and Ethical Considerations

This section documents the attack generation methodology, effectiveness analysis, ethical considerations, and data availability.

Attack Generation Methodology

Synthetic Attack Generation

Process:

1. **Template Creation**: Define attack structure templates for each category
2. **Parameter Variation**: Systematically vary attack parameters
3. **Constraint Satisfaction**: Ensure attacks satisfy category definitions
4. **Deduplication**: Remove semantically equivalent attacks
5. **Validation**: Human review of generated attacks

Table 1: Generation method statistics.

Method	Attacks	Success Rate	Novelty Score
Template instantiation	120	68%	0.3
LLM-assisted mutation	50	75%	0.7