

Attack Corpus: Statistics and Taxonomy

This supplementary material provides corpus overview (sec:corpus-overview), detailed statistics (sec:corpus-stats), example attacks by category (sec:attack-examples), generation methodology (sec:generation-methodology), effectiveness analysis (sec:effectiveness-analysis), and ethical considerations (sec:ethical-considerations).

Corpus Overview

The attack corpus used for experimental validation comprises 950 unique attack instances across four primary categories. This supplementary material provides detailed statistics, sanitized examples, generation methodology, and ethical considerations.

Implementation: The corpus is programmatically generated using `src/attacks/corpus.py` with deterministic seeding (default `seed=42`). Run `python -m src.attacks.corpus` to regenerate the `corpus.json` file. Attack templates are defined in `src/attacks/templates.py`, which validates payload structure against category definitions.