



 Raw Dataset

 Automated
Filtering

 Semi-Automated
Review

 Clean Dataset

 Exact Duplicates

Action:  One removed.

 Few Tokens

Action:  Removed.

 Non-Portuguese

Action:  Removed.

 Contradictory Labels

Near-duplicates
or Same URL Mention
with **FALSE** vs. **TRUE** labels.

Action:  Both removed.

 Google FactCheck
Label Correction

Correction: **TRUE** ↔ **FAKE**