

Introduction

Portuguese corpora lack annotations for evidence.

Claim: "We've had patients who took COVID-19 tests [...] that came back negative, but when repeated [...] showed positive results. Remember that nasal swab tests may yield false negatives..."

Label: TRUE

Evidence: **NOT PROVIDED**

Figure 1: The evidence gap in the COVID19.BR Corpus [2].

Our Contributions

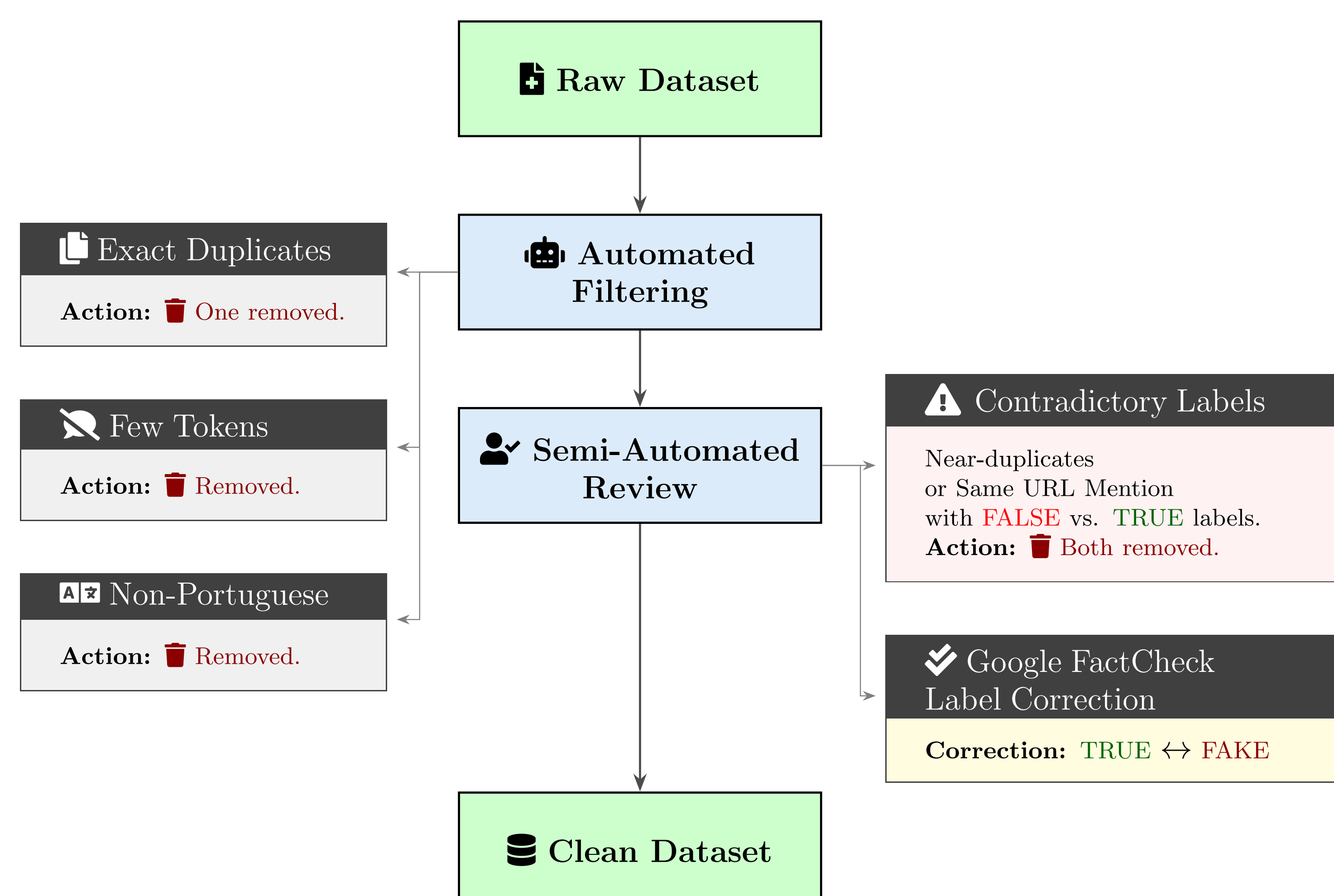
- (i) **Semi-automatic cleaning process**
- (ii) **Enriching existant datasets:** Claim Extraction + Web Search

Corpora Analysis

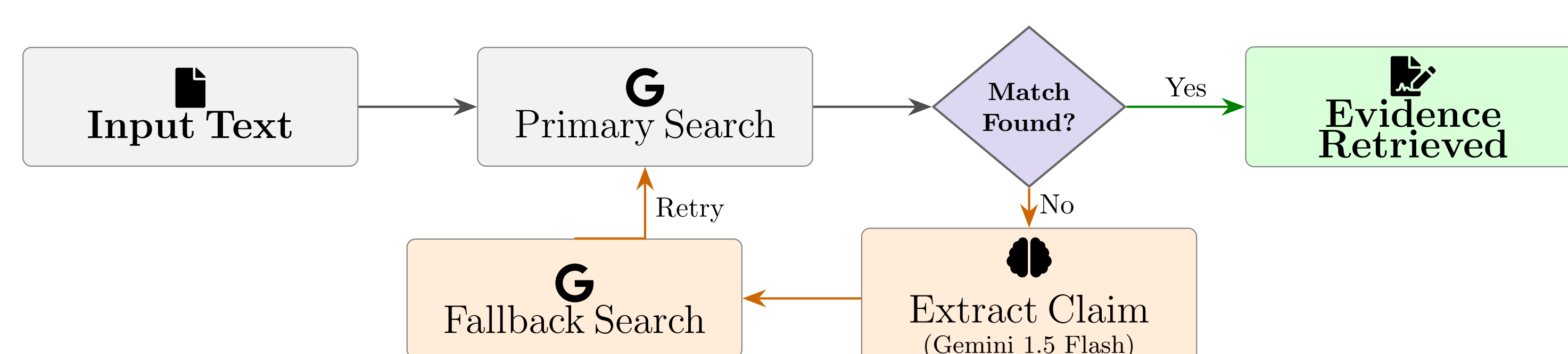
Info (fake/true)	Fake.BR [3]	COVID19.BR [2]	MuMiN-PT [4]
Domain	General	Health	"General" (Health)
Year	2016–2018	2020	2020–2022
Approach [1]	bottom-up	bottom-up	top-down
Size	3580/3580	848/ 1139	1339 /65
% URL	1.0%/0.7%	28.9/ 56.9%	0.3%/0.0%
Avg. # words	181.4/ 183.1	167.7/111.1	18.9/16.9

Table 1: Corpora characteristics after cleaning. **Top-down** starts with fact-checked claims; **bottom-up** seeks for new misinformation in posts. MuMiN-PT excluded from classification due to severe class imbalance.

Cleaning Process



Enrichment Pipeline



Retrieved Evidence Patterns

For TRUE Claims

T1: Corroboration

Claim: COVID-19 tests may yield false negatives.

Evidence (Reliable News): "...false negatives may occur due to low specificity... no laboratory test is perfect..."

T2: Explicit Confirmation

Fact-checkers primarily debunk false claims. Explicit "TRUE" verdicts from them are rare (only ~1% of our retrieved API results).

For FALSE Claims

F1: Direct Debunk (Ideal)

Claim: Coronavirus transmitted via parcels.

Evidence (Fact-Checker): "...coronavirus cannot survive in parcels..." – A direct refutation from a reliable source.

F2: Misinfo. Reinforcement

Claim: Government will end the popular pharmacy program.

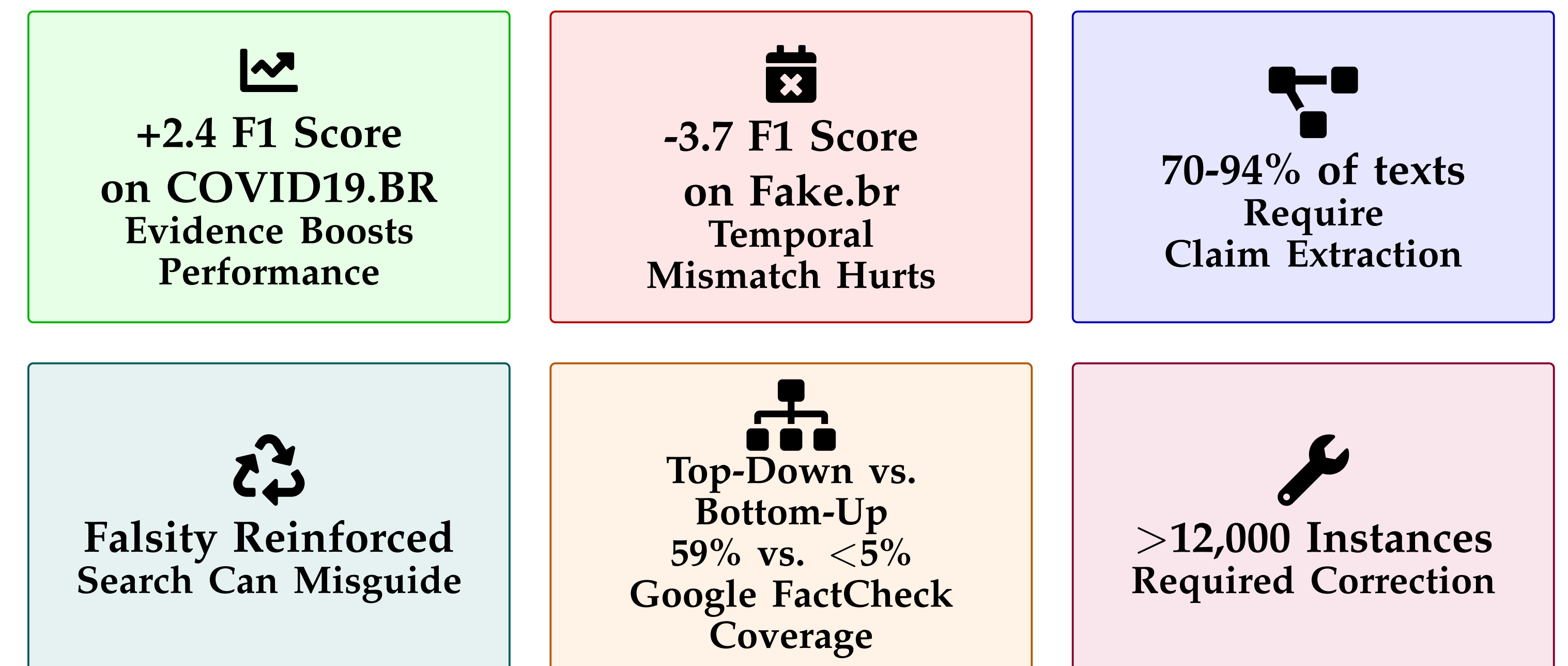
Evidence (Unreliable Blog): "...Government closes Popular Pharmacy..."

Experimental Results

Model	Input	COVID19.BR	Fake.br
BERTimbau [6] (Fine-tuned SLM)	+ Evidence	81.4 82.4 (+1.0)	98.9 99.2 (+0.3)
Gemini 1.5 Flash [5] (15-shot LLM)	+ Evidence	76.9 79.3 (+2.4)	80.4 76.7 (-3.7)

Table 2: F1-Macro scores on the test sets.

Key Findings



Limitations & Future Work



Resources & Access



References

- [1] S. Hangloo and B. Arora. Combating multimodal fake news on social media: methods, datasets, and future perspective. *Multimedia Systems*, 28(6):2391–2422, 2022.
- [2] A. D. F. Martins, L. Cabral, P. J. C. Mourao, I. C. de Sá, J. M. Monteiro, and J. Machado. Covid19.br: A dataset of misinformation about covid-19 in brazilian portuguese whatsapp messages. In *Anais do III Dataset Showcase Workshop*, pages 138–147. SBC, 2021.
- [3] R. A. Monteiro, R. L. S. Santos, T. A. S. Pardo, T. A. de Almeida, E. E. S. Ruiz, and O. A. Vale. Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In *Computational Processing of the Portuguese Language*, pages 324–334. Springer, 2018.
- [4] D. S. Nielsen and R. McConville. Mumin: A large-scale multilingual multimodal fact-checked misinformation social network dataset. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2022.
- [5] M. Reid, N. Savinov, D. Teplyashin, et al. Gemini 1.5: Unlocking multimodal understanding across millions oftokens of context. *arXiv preprint arXiv:2403.05530*, 2024.
- [6] F. Souza, R. Nogueira, and R. Lotufo. Bertimbau: Pretrained bert models for brazilian portuguese. In *Intelligent Systems*, pages 403–417. Springer, 2020.