

CT-BERT-PRCT

CT-BERT fine-tuned for
Population Replacement
Conspiracy Theory
Detection

Usage/License



The model is available under the MIT License via its Hugging Face repository (per standard open-source practice described in the repo).

Download



The model can be accessed and downloaded from its Hugging Face repository: <https://huggingface.co/erikbranmarino/CT-BERT-PRCT>

Researchers



- Erik Bran Marino
- Renata Vieira
- Davide Bassi

These researchers are affiliated with Universidade de Évora and CiTIUS-Universidade de Santiago de Compostela

Summary

CT-BERT-PRCT is a fine-tuned version of CT-BERT specifically adapted for detecting Population Replacement Conspiracy Theory (PRCT) content across social media platforms. The model has been trained to identify both explicit and implicit PRCT narratives while maintaining robust cross-platform generalization capabilities. It demonstrates strong performance on its primary training domain (YouTube - English) while maintaining reasonable effectiveness in cross-platform and multilingual scenarios (Telegram - Portuguese and Spanish).

What We Offer

- Specialized Detection Model: A BERT model tailored for conspiracy theory detection.
- Multilingual Support: Evaluated on English, Spanish, and Portuguese content.
- Cross-Platform Capability: Designed for use across different social media environments (e.g., YouTube, Telegram).

Key Features

- High Accuracy: 83.8% accuracy on YouTube English content.
- Cross-Platform Generalization: 71.2% F1-score on Telegram across multiple languages.
- Nuanced Analysis: Capable of detecting both explicit and implicit PRCT narratives (e.g., The Great Replacement Theory, Eurabia).

Collaboration Objectives

- Content Moderation Teams: Integrate as an AI-assisted moderation tool (with humans in the loop) for social media platforms.
- Academic Researchers: Support studies on the propagation and evolution of conspiracy theories online.