

Mistral-PRCT



Funded by
the European Union



UK Research
and Innovation

Usage/License



The model is available under the MIT License - Free for research and educational purposes via its Hugging Face repository.

Download



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

Researchers



- Erik Bran Marino
- Renata Vieira

These researchers are affiliated with Universidade de Évora.

Summary

Mistral-PRCT is a LoRA fine-tuned version of Mistral-7B-Instruct-v0.3, specifically adapted for detecting Population Replacement Conspiracy Theory (PRCT) content. Trained exclusively on informal social media discourse from Portuguese Telegram messages, the model demonstrates strong cross-domain generalization, achieving competitive performance on Italian news headlines. It offers an optimal balance between accuracy and inference speed for social media monitoring.

Key Features

- High In-Domain Performance: 89.6% accuracy on Portuguese Telegram data.
- Multilingual Capability: Best on Portuguese, good on Italian, handling both implicit and explicit PRCT framing.
- Resource Efficient: Lower memory requirements compared to larger LLMs while maintaining high accuracy.

What We Offer

- Efficient Detection Model: A 7B parameter model suitable for faster processing (~4.6s per sample).
- Robust Social Media Focus: Highly effective for informal discourse analysis (0.819 F1-Macro on Telegram PT).
- Cross-Domain Generalization: Effective transfer to formal news contexts.

Collaboration Objectives

- Content Moderators: Serve as an efficient AI-assisted tool for flagging harmful content on social media platforms.
- Social Scientists: Support large-scale analysis of informal conspiracy discourse online.

