

HYBRIDS

OVERCOMING THE AI BLACK BOX IN POLITICAL ANALYSIS: A HYBRID APPROACH

SUSO BALEATO, ERIK MARINO, RENATA VIEIRA

THE CHALLENGE OF AI IN SOCIAL SCIENCE

AI's Strengths: AI is great in processing large volumes of data and detecting patterns.

The Problem: AI models, particularly those based on deep learning, often fail to distinguish between correlation and causation.

Social Science Needs: In social sciences, we need to explain why something happens, not just observe that it happens.

Main Question: How can we make AI tools useful for deeper, causal analysis in fields like political science?

Scientific Research	AI Shortcomings
Lack of Bias	AI models can inherit and even amplify biases present in training data, leading to skewed or unethical research outcomes.
Reliability of Techniques	AI models can produce inconsistent results due to variations in input data, model parameters, or random initialization.
Reproducibility	Many AI models operate as "black boxes," making it difficult to understand how outputs are generated, hindering reproducibility.
Contextual Understanding	AI struggles with understanding context, nuance, and ambiguity, often leading to oversimplified or inaccurate interpretations.
Causal Inference	AI lacks inferential capacity and does not inherently understand causal relationships, which is crucial in many scientific fields.
Ethical Standards	AI use raises ethical concerns, particularly related to privacy, consent, and potential misuse of AI-generated data.
Generalization from Data	AI models can overfit to training data, capturing noise rather than generalizable patterns, limiting their application to new data.

COMPUTER SCIENCE HAS PROPOSED ALTERNATIVES

- ▶ World models (Yann LeCun): build internal world representations to plan and predict
 - ▶ Bayesian Program Learning (Joshua Tenenbaum): learn from few examples instead of big data
 - ▶ Behavior-Based (Rodney Brooks): simple reactive behavior instead of complex internal architectures
 - ▶ OpenCog (Ben Goertzel): combine symbolic, probabilistic logic, and deep learning
 - ▶ Hybrid Intelligence (Gary Marcus): human/ AI interaction to combine deep learning strengths for data pattern recognition with rule-based symbolic reasoning
- HYBRIDS research project - ~ 2.85 million USD - European Union
Marie Skłodowska-Curie Actions (MSCA) 101073351

A HYBRID APPROACH TO OVERCOME LIMITATIONS

Core Idea: Ensure that the research question, theoretical framework, hypothesis, and operationalization of variables are carefully defined before any data is retrieved.

In this hybrid approach, AI models are not generic. They are tailored to specific research questions by combining AI techniques with human expertise from the outset.

HOW HYBRID AI WORKS

Define the research question and theoretical framework before using AI.

STEP 1

AI techniques are then applied, but they are tailored to the specific research problem.

STEP 2

Social scientists validate the model's performance to ensure it aligns with theory

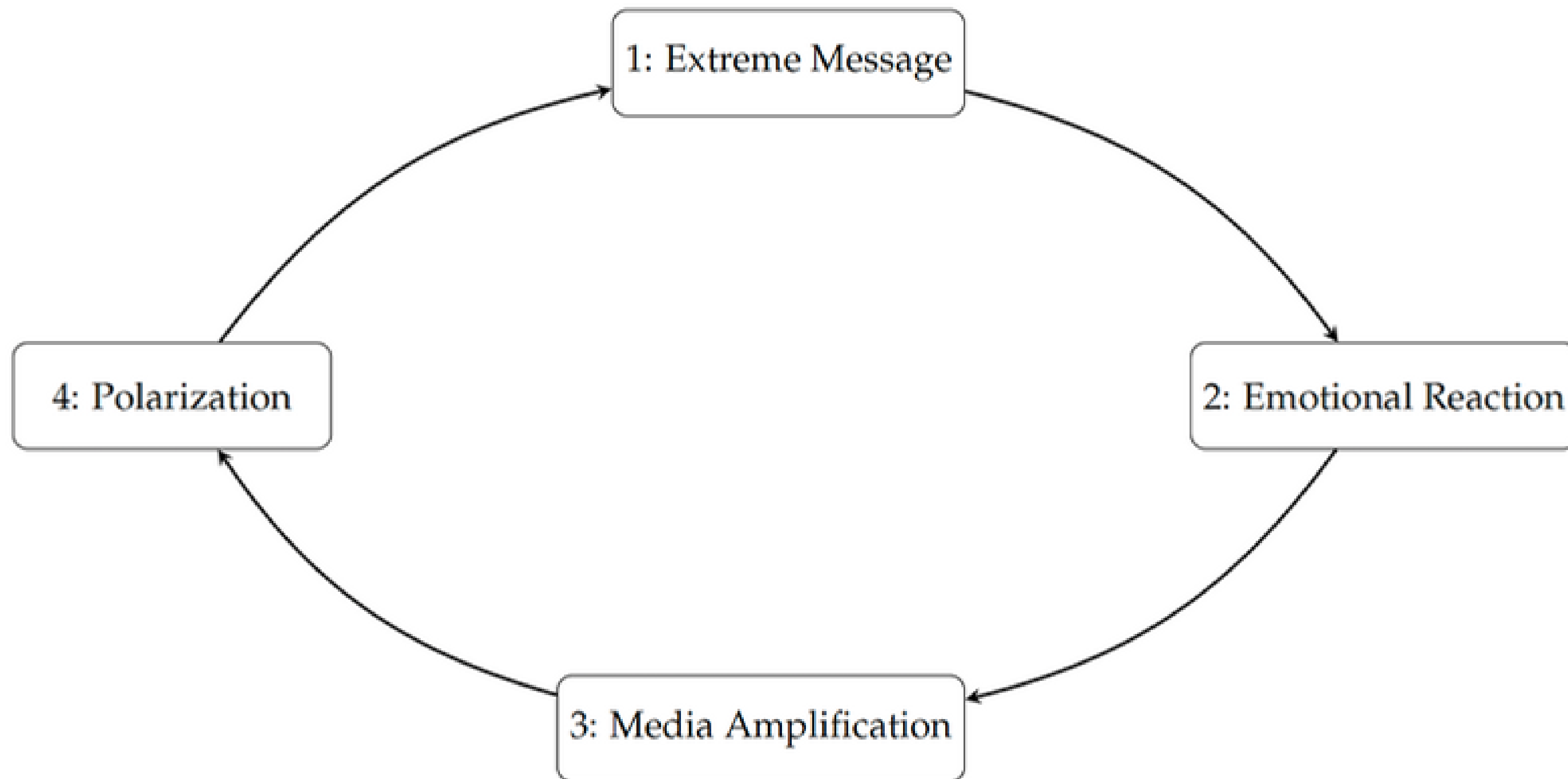
STEP 3

CASE STUDY: DISINFORMATION AND POLARIZATION

Our theoretical framework proposes a “polarization loop,” where disinformation about conspiracy theories intensifies societal polarization.

To test this theory, we analyze the conspiratorial discourse on media, social networks, parliamentary speeches, and other sources.

THE POLARIZATION LOOP





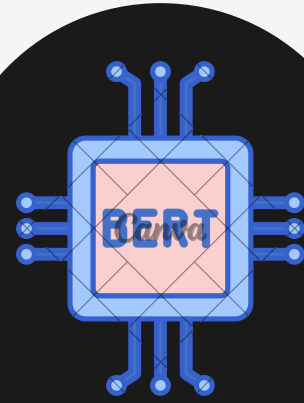
DEVELOPING A CONSPIRACIST DATASET

Telegram messages known for conspiratorial content and hate speech.

Our annotation process involved tagging messages for key indicators such as:

- Conspiracist tone
- Population replacement conspiracy theories
- Vaccine-related conspiracies
- Hate speech

FINE-TUNING BERT FOR POLITICAL DISCOURSE



Multilingual
BERT model

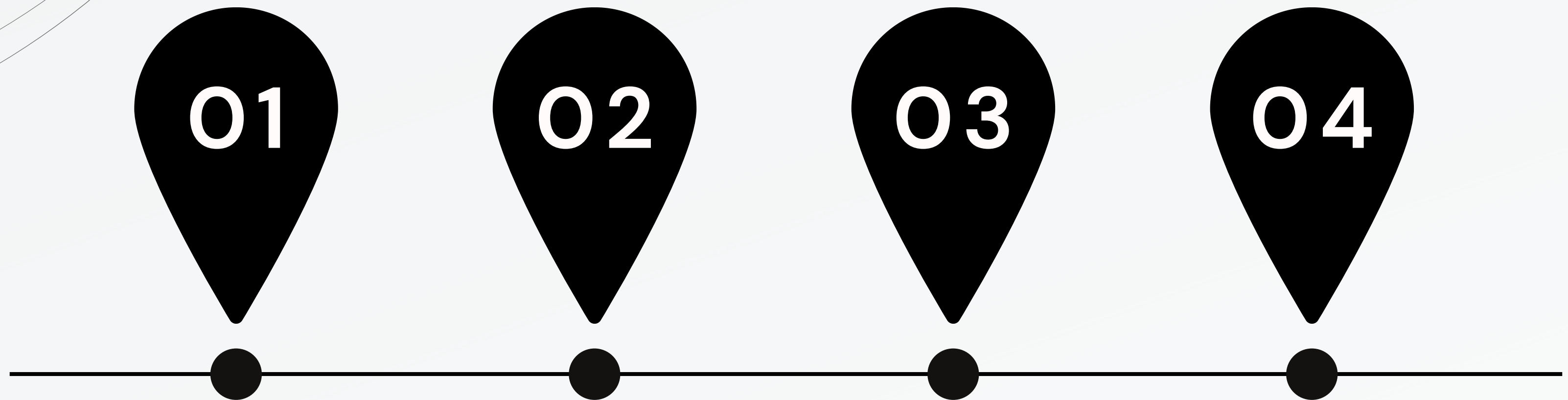


Manually
annotated
corpus



Classify texts
within
theoretical
framework

EVALUATING THE CAUSAL LINK



CONCLUSIONS

The hybrid AI approach can provide a pathway to overcoming the limitations of traditional AI models in social sciences.

By grounding AI models in domain-specific theories and validating them with expert input, we can tackle complex societal questions with greater transparency, explainability, and scientific rigor.



THANK YOU

[HTTPS://HYBRIDSPROJECT.EU](https://hybridsproject.eu)



Centro Interdisciplinar de História, Culturas e Sociedades da Universidade de Évora

CIDEHUS