TA1: Extracting, evaluating, and sense-making claims from published scholarly work

IARPA REASON Proposers Day
January 11 2023

Sarah Rajtmajer
Assistant Professor
The Pennsylvania State University
smr48@psu.edu

Jian Wu
Assistant Professor
Old Dominion University
jwu@cs.odu.edu

Background: work for DARPA's SCORE program (September 2019 – December 2022)

DARPA DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

EXPLORE BY TAG

ABOUT US / OUR RESEARCH / NEWS / EVENTS / WORK WITH US /

› Defense Advanced Research Projects Agency › Our Research › Systematizing Confidence in Open Research and Evidence

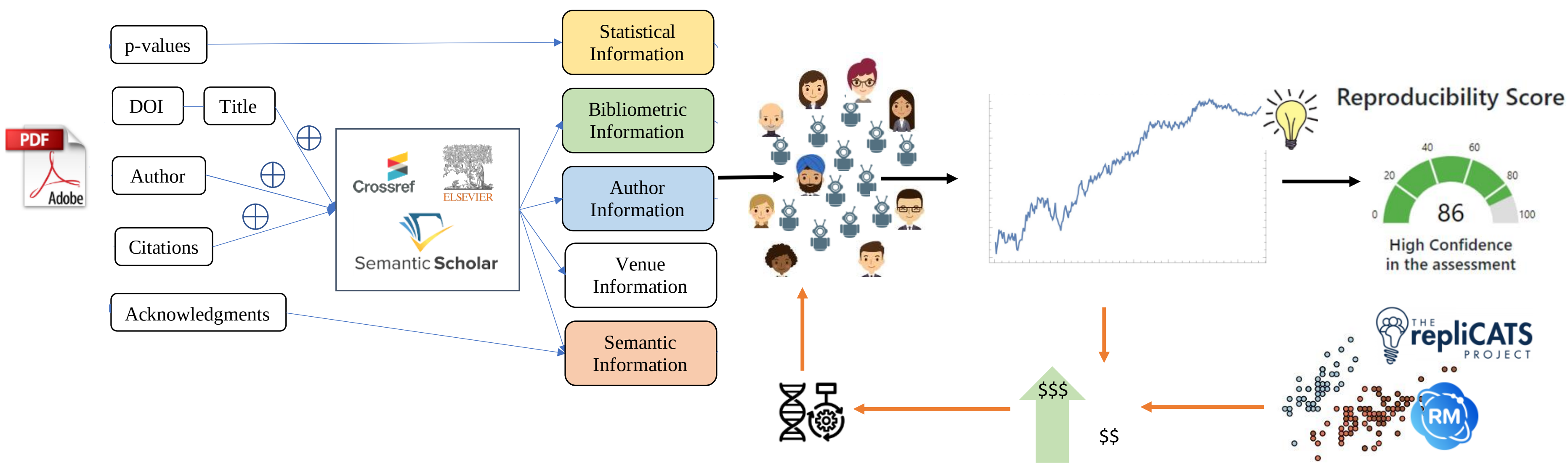
Systematizing Confidence in Open Research and Evidence (SCORE)

Dr. Greg Witkop

Develop and deploy tools to assign *explainable* "confidence scores" to SBS research results and claims

The Department of Defense (DoD) often leverages social and behavioral science (SBS) research to design plans, guide investments, assess outcomes, and build models of human social systems and behaviors as they relate to national security challenges in the human domain. However, a number of recent empirical studies and meta-analyses have revealed that many SBS results vary dramatically in terms of their ability to be independently reproduced or replicated, which could have real-world implications for DoD's plans, decisions, and models. To help address this situation, DARPA's Systematizing Confidence in Open Research and Evidence (SCORE) program aims to develop and deploy automated tools to assign "confidence scores" to different SBS research results and claims. Confidence scores are quantitative measures that should enable a DoD consumer of SBS research to understand the degree to which a particular claim or result is likely to be reproducible or replicable. These tools will assign explainable confidence scores with a reliability that is equal to, or better than, the best current human expert methods. If successful, SCORE will enable DoD personnel to quickly calibrate the level of confidence they should have in the reproducibility and replicability of a given SBS result or claim, and thereby increase the effective use of SBS literature and research to address important human domain challenges, such as enhancing deterrence, enabling stability, and reducing extremism.

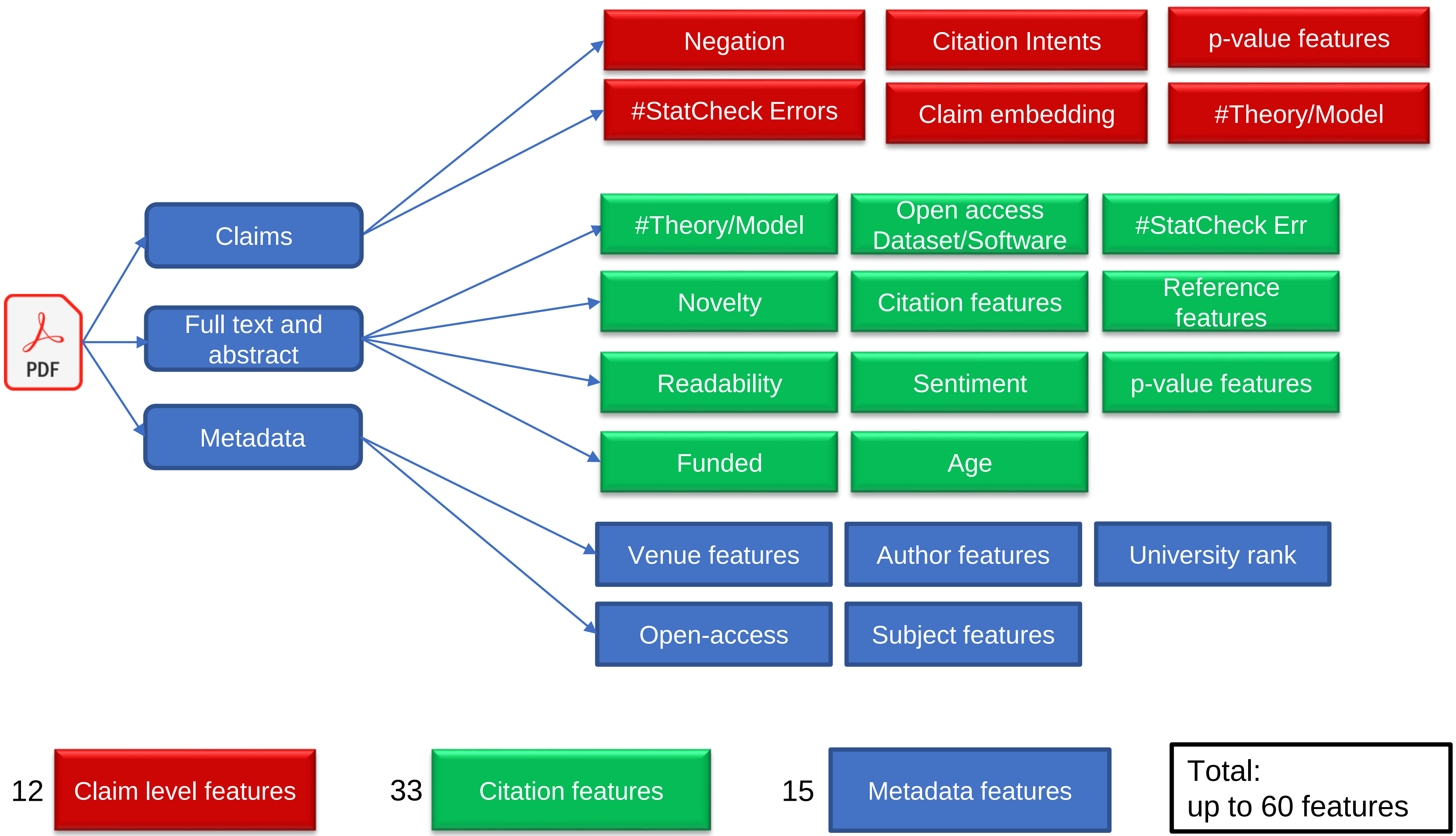
XAI (artificial prediction markets) *and* crowd+AI hybrid markets



Artificial prediction markets ☐ *populated by artificial agents (trader-bots)* ☐ purchase assets representing "will replicate" and "will not replicate" outcomes of notional replications of claims appearing within research papers. Agent reasoning is based on ***human-interpretable signals***, including full text of scientific papers, metadata for specific papers, and metadata about the community and the field.

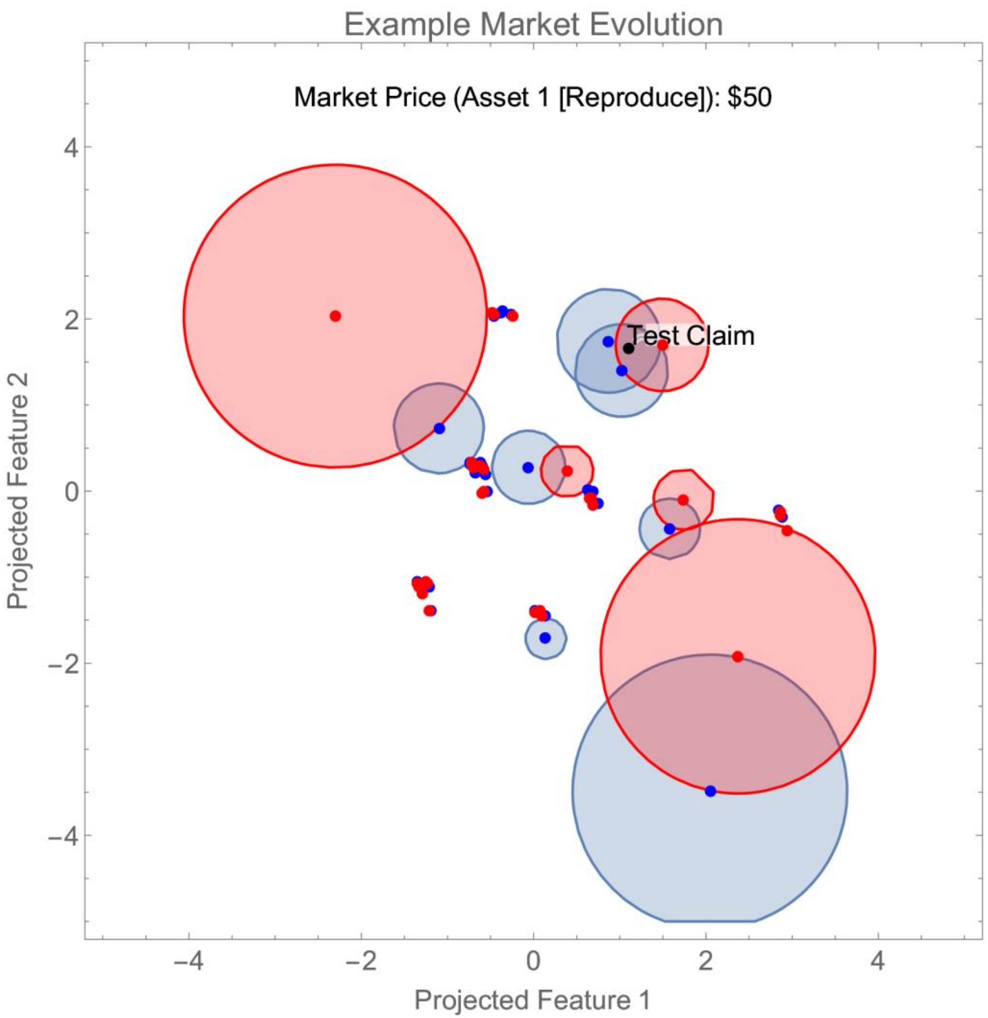
Hybrid scenario: SMEs engage alongside bot traders

Signals (features) extracted from full text and assembled from metadata



Artificial prediction markets

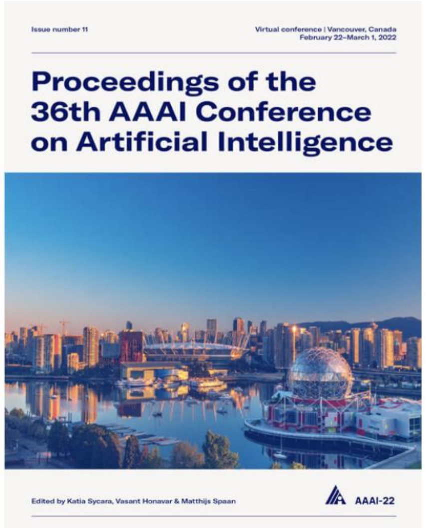
- Synthetic agents interact in a simple binary option market using a logarithmic market scoring rule.
- Agents in the market bid in geometric regions of feature space, shown as circles (for simplicity).
- The agents are sensitive to asset price, which causes their bid behavior to evolve in time.
- Convergence in the market is equivalent to a geometric equilibrium.



(above) A toy market with input data from RPP
Note 1: High dim feature space is projected down for visualization.
Note 2: We multiply the price by 100 and convert to dollars.)

Nakshatri et al. (2021) Design and analysis of a synthetic prediction market using convex sets. Results in Control and Optimization. https://www.sciencedirect.com/science/article/pii/S2666720721000308#

System evaluation --> real replication data



The Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22)

A Synthetic Prediction Market for Estimating Confidence in Published Work

Sarah Rajtmajer,[1] Christopher Griffin,[1] Jian Wu,[2] Robert Fraleigh,[1] Laxmaan Balaji,[1] Anna Squicciarini,[1] Anthony Kwasnica,[1] David Pennock,[3] Michael McLaughlin,[1] Timothy Fritton,[1] Nishanth Nakshatri,[1] Arjun Menon,[1] Sai Ajay Modukuri,[1] Rajal Nivargi,[1] Xin Wei,[2] C. Lee Giles[1]

[1]The Pennsylvania State University
[2]Old Dominion University
[3]Rutgers University
{smr48,cxg286,rdf5090,lpb5347,acs20,amk17,mvm7085,tjf115,nzn5185,amm8987,svm6277,rfn5089,clg20}@psu.edu
{
d

Abstract

Explainably estimating confidence in published sc
work offers opportunity for faster and more robus
tific progress. We develop a synthetic prediction m
assess the credibility of published claims in the social
havioral sciences literature. We demonstrate our syst
detail our findings using a collection of known rep
projects. We suggest that this work lays the founda
a research agenda that creatively uses AI for peer revi

Introduction

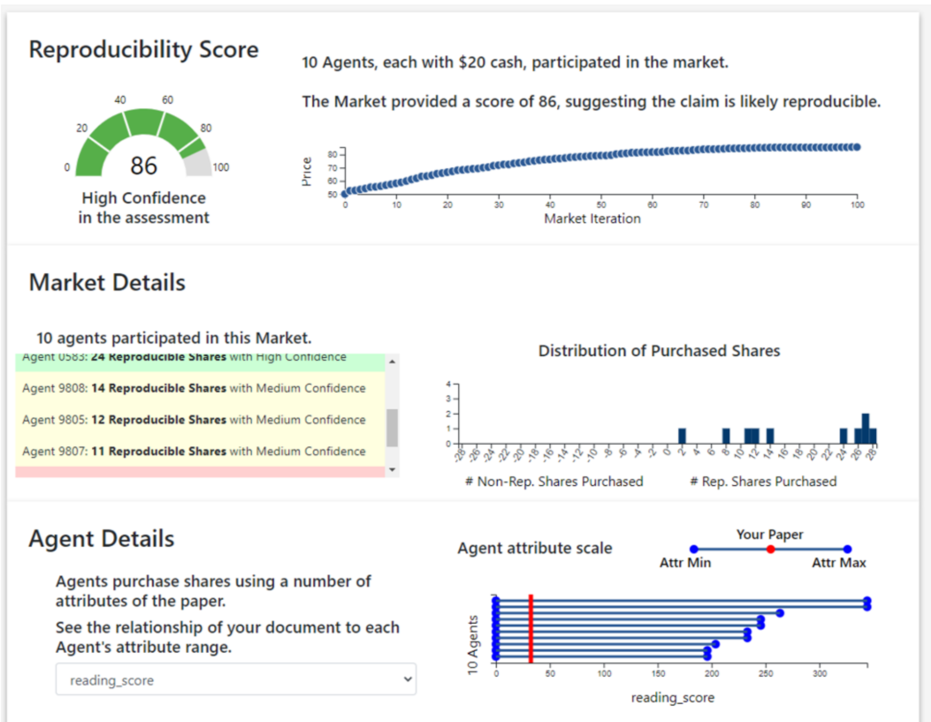
Concerns about the replicability, robustness ar
ducibility of findings in scientific literature hav

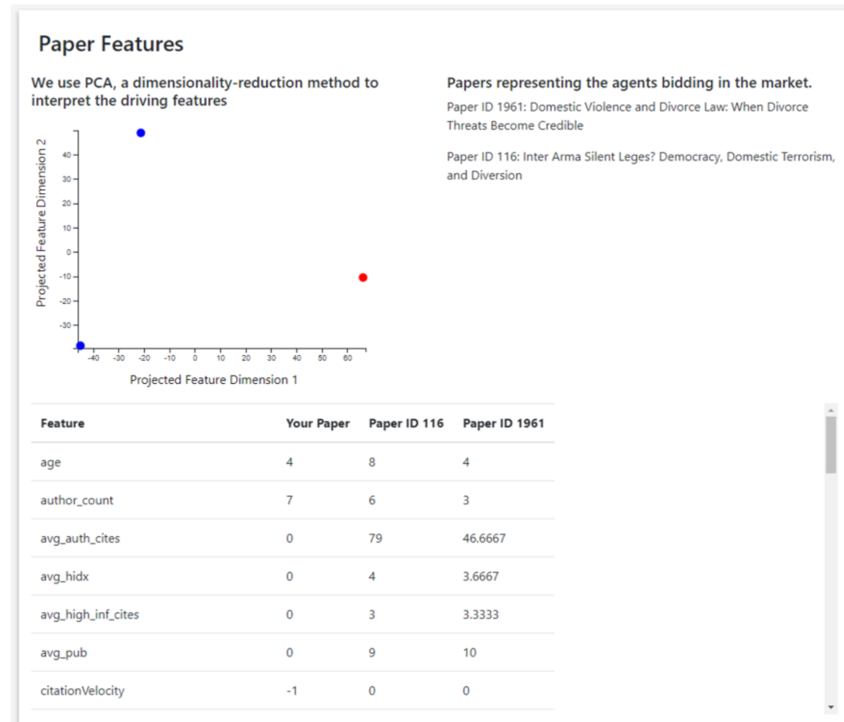
Results on scored papers. Our system provides a confidence score for 68 of 192 (35%) of the papers in our set. On the set of scored papers, accuracy is 0.894, precision is 0.917, recall is 0.903, and **F1** is **0.903** (macro averages). A sizeable un-scored subset of data (65%) is the trade-off for high accuracy on the scored subset of the data. A test point is un-scored when the system has determined it has insufficient information to evaluate it.

System non-scoring. Unlike most other machine learning algorithms, the synthetic market does not provide an evaluation for every input. Like its human-populated counterparts, the market is vulnerable to lack of participation (Arrow et al. 2008; Tetlock 2008; Rothschild and Pennock 2014). Agents will not participate if they have not seen a sufficiently similar training point (paper). This is more common when the training dataset is small; in experiments with larger datasets, we have observed participation increases. Meaningful ways to increase agent participation, including hybrid settings with human participants, are being explored.

System evaluation --> RAND grad students

- ❑ **Claim submission**: User submits a paper (PDF) for evaluation.
- ❑ **Feature Extraction**: Extraction tools stage, followed by pass through feature extractor modules generate paper feature vector.
- ❑ **Evaluation through multiple prediction markets**: The feature vector is passed through multiple markets and results from each are collected.
- ❑ **SCORE and interpretability**: Results from the prediction markets are collated and a response containing the SCORE, interpretability and confidence is returned.





Explanations:

- ❑ **Level One:** Confidence in the claim’s reproducibility through market score
- ❑ **Level Two:** Aggregated details related to agent participation in the system
- ❑ **Level Three:** Which agents participated + their confidence
- ❑ **Level Four:** Features corresponding to nearest training data points

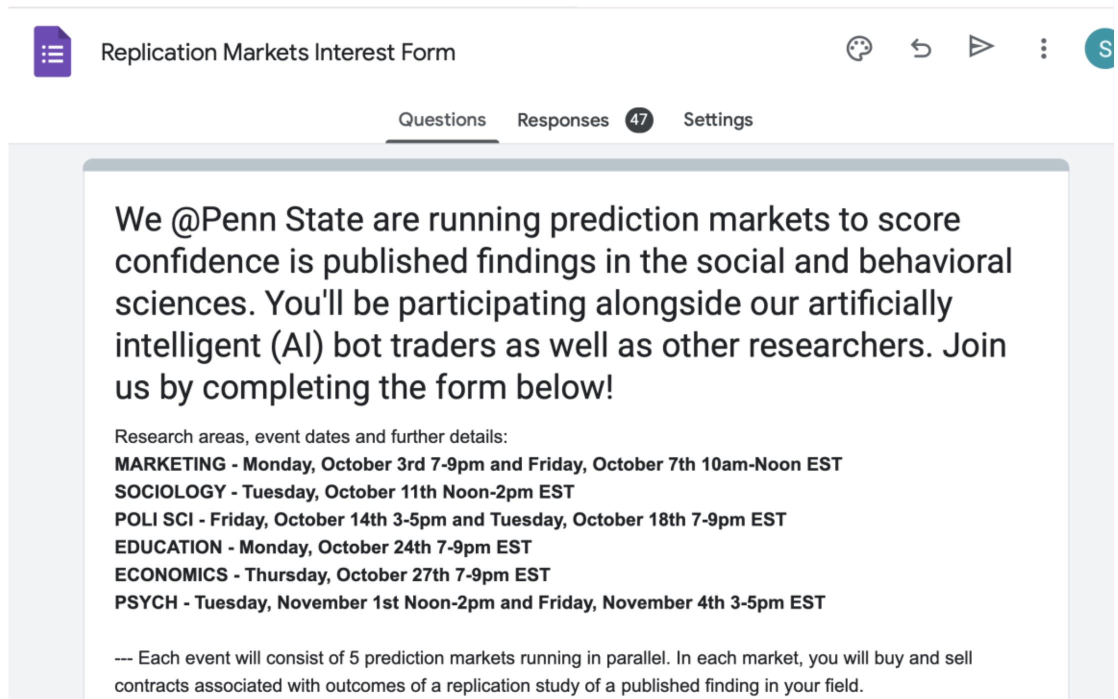
Hybrid prediction markets

Virtual 2-hour long market events October-November 2022

- 50+ participants
- Currently analyzing results, and conducting interviews with participants

Initial takeaways:

→ Major improvement on agent participation!
→ Change in individuals' evals before/after market based on surveys
→ Need more work to understand the right "balance" of bots and SMEs


Replication Markets Interest Form

Questions Responses 47 Settings

We @Penn State are running prediction markets to score confidence is published findings in the social and behavioral sciences. You'll be participating alongside our artificially intelligent (AI) bot traders as well as other researchers. Join us by completing the form below!

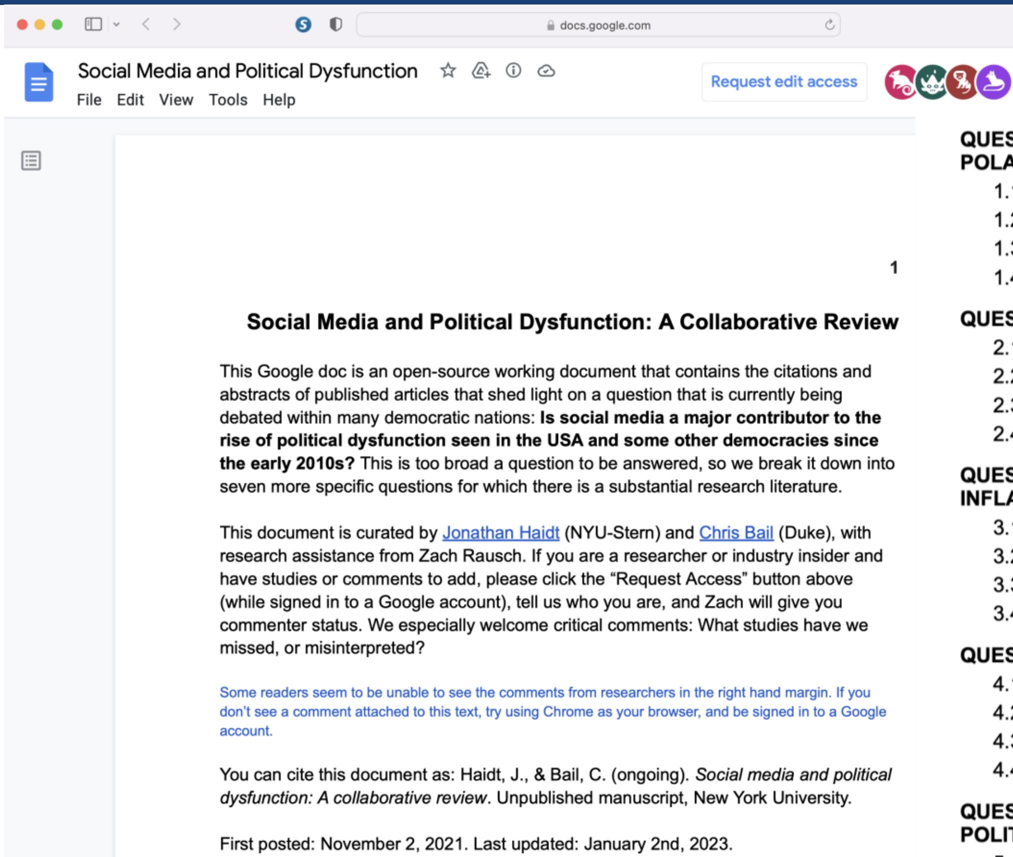
Research areas, event dates and further details:
MARKETING - Monday, October 3rd 7-9pm and Friday, October 7th 10am-Noon EST
SOCIOLOGY - Tuesday, October 11th Noon-2pm EST
POLI SCI - Friday, October 14th 3-5pm and Tuesday, October 18th 7-9pm EST
EDUCATION - Monday, October 24th 7-9pm EST
ECONOMICS - Thursday, October 27th 7-9pm EST
PSYCH - Tuesday, November 1st Noon-2pm and Friday, November 4th 3-5pm EST

--- Each event will consist of 5 prediction markets running in parallel. In each market, you will buy and sell contracts associated with outcomes of a replication study of a published finding in your field.

Next steps for IARPA REASON

- Automatically extract key claims and evidence from analyst reports
- Search the scholarly record to find published related to those claims
- Extract supporting evidence and assign confidence scores to the associated finding
- Develop a high-dimensional hypothesis space, where dimensions are variables/factors that matter for that claim, in which to *embed* research findings to understand their relationships
- Develop an “encyclopedia” of high-confidence findings relevant to the analyst’s claim in the scholarly SBS literature along with *explanations* for these assessments

TA1 vision: Automated collaborative review



docs.google.com

Social Media and Political Dysfunction

File Edit View Tools Help

Request edit access

1

Social Media and Political Dysfunction: A Collaborative Review

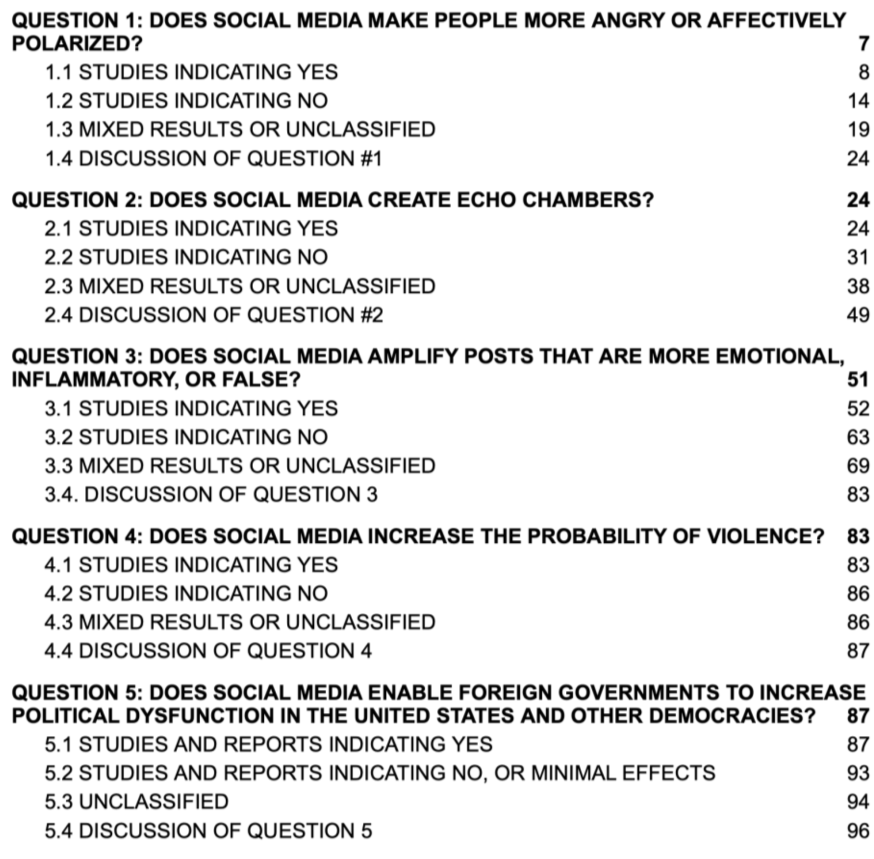
This Google doc is an open-source working document that contains the citations and abstracts of published articles that shed light on a question that is currently being debated within many democratic nations: **Is social media a major contributor to the rise of political dysfunction seen in the USA and some other democracies since the early 2010s?** This is too broad a question to be answered, so we break it down into seven more specific questions for which there is a substantial research literature.

This document is curated by Jonathan Haidt (NYU-Stern) and Chris Bail (Duke), with research assistance from Zach Rausch. If you are a researcher or industry insider and have studies or comments to add, please click the "Request Access" button above (while signed in to a Google account), tell us who you are, and Zach will give you commenter status. We especially welcome critical comments: What studies have we missed, or misinterpreted?

Some readers seem to be unable to see the comments from researchers in the right hand margin. If you don't see a comment attached to this text, try using Chrome as your browser, and be signed in to a Google account.

You can cite this document as: Haidt, J., & Bail, C. (ongoing). *Social media and political dysfunction: A collaborative review*. Unpublished manuscript, New York University.

First posted: November 2, 2021. Last updated: January 2nd, 2023.



QUESTION 1: DOES SOCIAL MEDIA MAKE PEOPLE MORE ANGRY OR AFFECTIVELY POLARIZED? 7
- 1.1 STUDIES INDICATING YES 8
- 1.2 STUDIES INDICATING NO 14
- 1.3 MIXED RESULTS OR UNCLASSIFIED 19
- 1.4 DISCUSSION OF QUESTION #1 24

QUESTION 2: DOES SOCIAL MEDIA CREATE ECHO CHAMBERS? 24
- 2.1 STUDIES INDICATING YES 24
- 2.2 STUDIES INDICATING NO 31
- 2.3 MIXED RESULTS OR UNCLASSIFIED 38
- 2.4 DISCUSSION OF QUESTION #2 49

QUESTION 3: DOES SOCIAL MEDIA AMPLIFY POSTS THAT ARE MORE EMOTIONAL, INFLAMMATORY, OR FALSE? 51
- 3.1 STUDIES INDICATING YES 52
- 3.2 STUDIES INDICATING NO 63
- 3.3 MIXED RESULTS OR UNCLASSIFIED 69
- 3.4. DISCUSSION OF QUESTION 3 83

QUESTION 4: DOES SOCIAL MEDIA INCREASE THE PROBABILITY OF VIOLENCE? 83
- 4.1 STUDIES INDICATING YES 83
- 4.2 STUDIES INDICATING NO 86
- 4.3 MIXED RESULTS OR UNCLASSIFIED 86
- 4.4 DISCUSSION OF QUESTION 4 87

QUESTION 5: DOES SOCIAL MEDIA ENABLE FOREIGN GOVERNMENTS TO INCREASE POLITICAL DYSFUNCTION IN THE UNITED STATES AND OTHER DEMOCRACIES? 87
- 5.1 STUDIES AND REPORTS INDICATING YES 87
- 5.2 STUDIES AND REPORTS INDICATING NO, OR MINIMAL EFFECTS 93
- 5.3 UNCLASSIFIED 94
- 5.4 DISCUSSION OF QUESTION 5 96

smr48@psu.edu