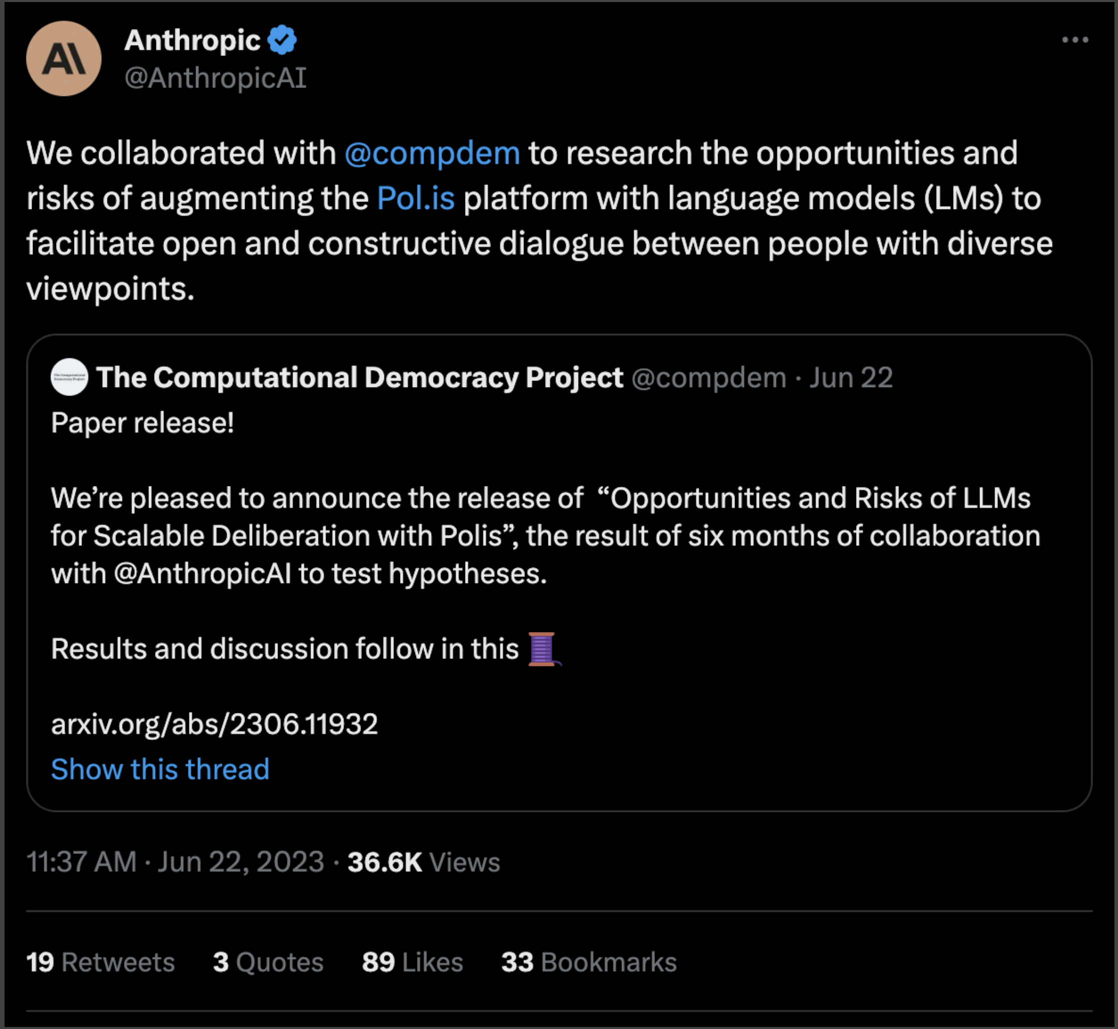
# Pol.is
# &
# The Computational Democracy Project

**COLIN MEGILL**
**POLIS / THE COMPUTATIONAL DEMOCRACY PROJECT (501c3)**

**@colinmegill @usepolis @compdem**

colinmegill.com/vienna



**Anthropic**
@AnthropicAI

We collaborated with @compdem to research the opportunities and risks of augmenting the Pol.is platform with language models (LMs) to facilitate open and constructive dialogue between people with diverse viewpoints.

**The Computational Democracy Project** @compdem · Jun 22

Paper release!

We're pleased to announce the release of "Opportunities and Risks of LLMs for Scalable Deliberation with Polis", the result of six months of collaboration with @AnthropicAI to test hypotheses.

Results and discussion follow in this 🧵

arxiv.org/abs/2306.11932

Show this thread

11:37 AM · Jun 22, 2023 · **36.6K** Views

**19** Retweets **3** Quotes **89** Likes **33** Bookmarks

arXiv:2306.11932v1 [cs.SI] 20 Jun 2023

# Opportunities and Risks of LLMs for Scalable Deliberation with Polis

Christopher T. Small[*,1], Ivan Vendrov[2], Esin Durmus[2], Hadjar Homaei[1], Elizabeth Barry[1], Julien Cornebise[1], Ted Suzman[1], Deep Ganguli[2], and Colin Megill[1]

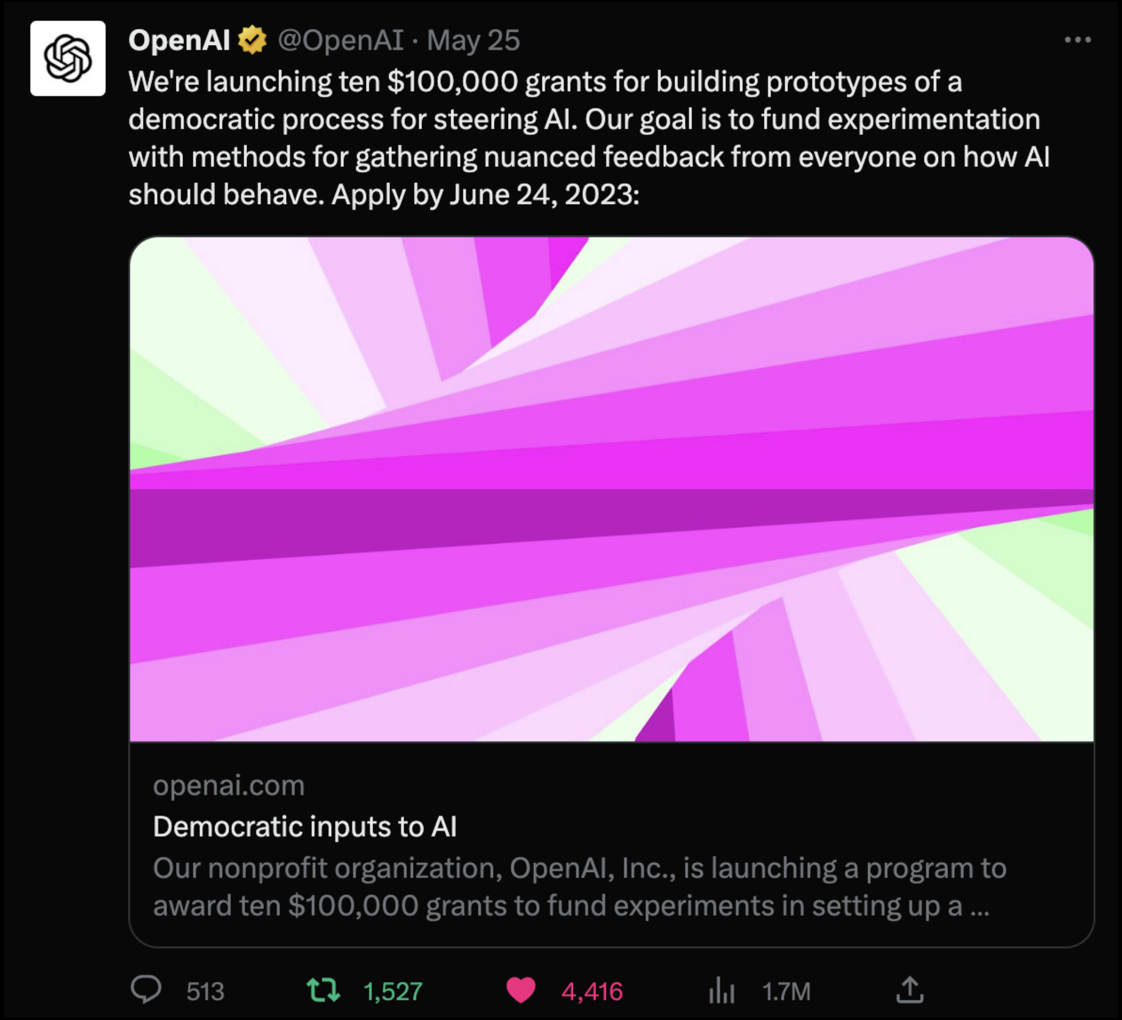
[1]The Computational Democracy Project
[2]Anthropic

June 2023

## Abstract

Polis is a platform that leverages machine intelligence to scale up deliberative processes. In this paper, we explore the opportunities and risks associated with applying Large Language Models (LLMs) towards challenges with facilitating, moderating and summarizing the results of Polis engagements. In particular, we demonstrate with pilot experiments using Anthropic's Claude that LLMs can indeed augment human intelligence to help more efficiently run Polis conversations. In particular, we find that summarization capabilities enable categorically new methods with immense promise to empower the public in collective meaning-making exercises. And notably, LLM context limitations have a significant impact on insight and quality of these results.

However, these opportunities come with risks. We discuss some of these risks, as well as principles and techniques for characterizing and mitigating them, and the implications for other deliberative or political systems that may employ LLMs. Finally, we conclude with several open future research directions for augmenting tools like Polis with LLMs.



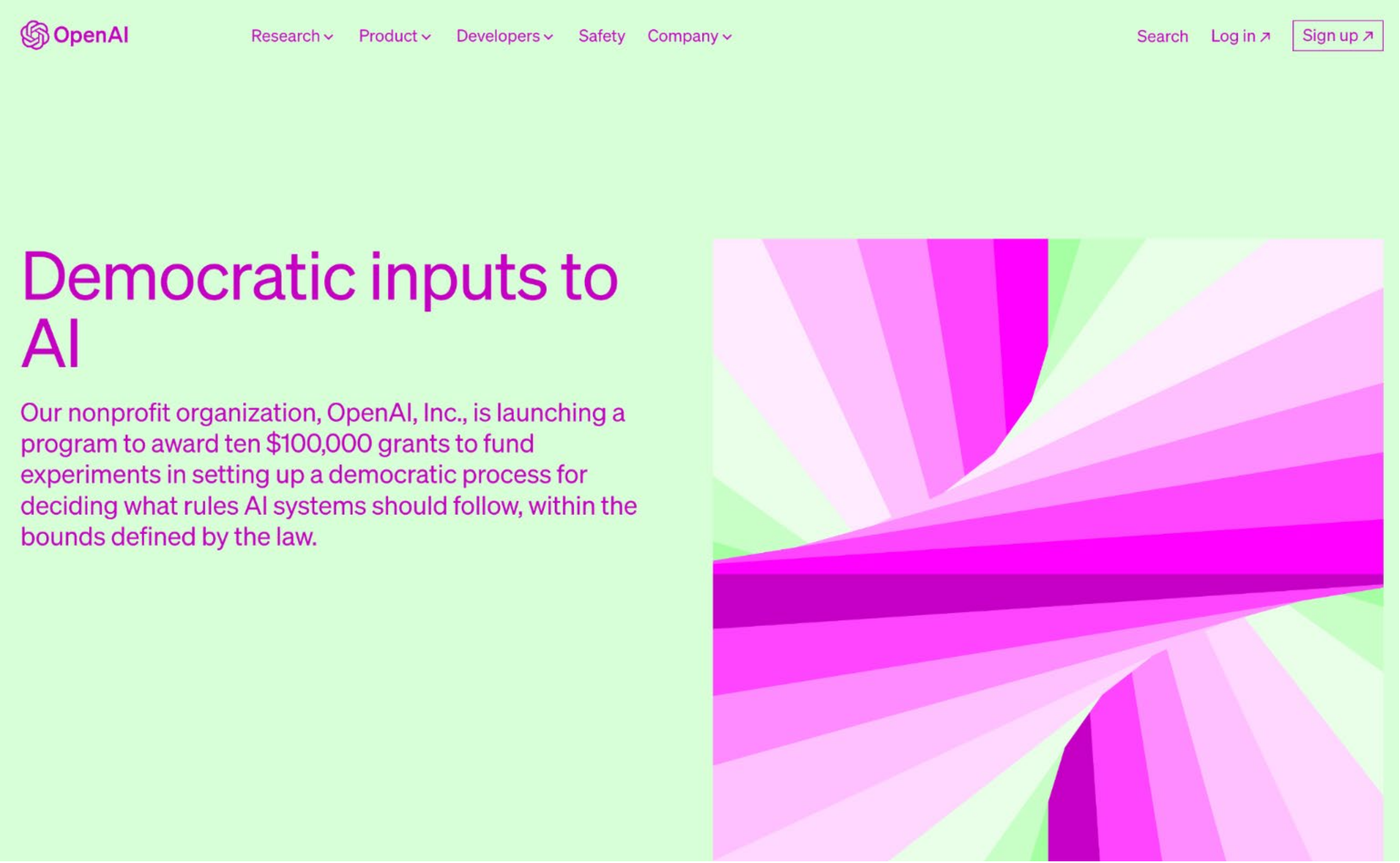
**OpenAI** @OpenAI · May 25

We're launching ten $100,000 grants for building prototypes of a democratic process for steering AI. Our goal is to fund experimentation with methods for gathering nuanced feedback from everyone on how AI should behave. Apply by June 24, 2023:

openai.com

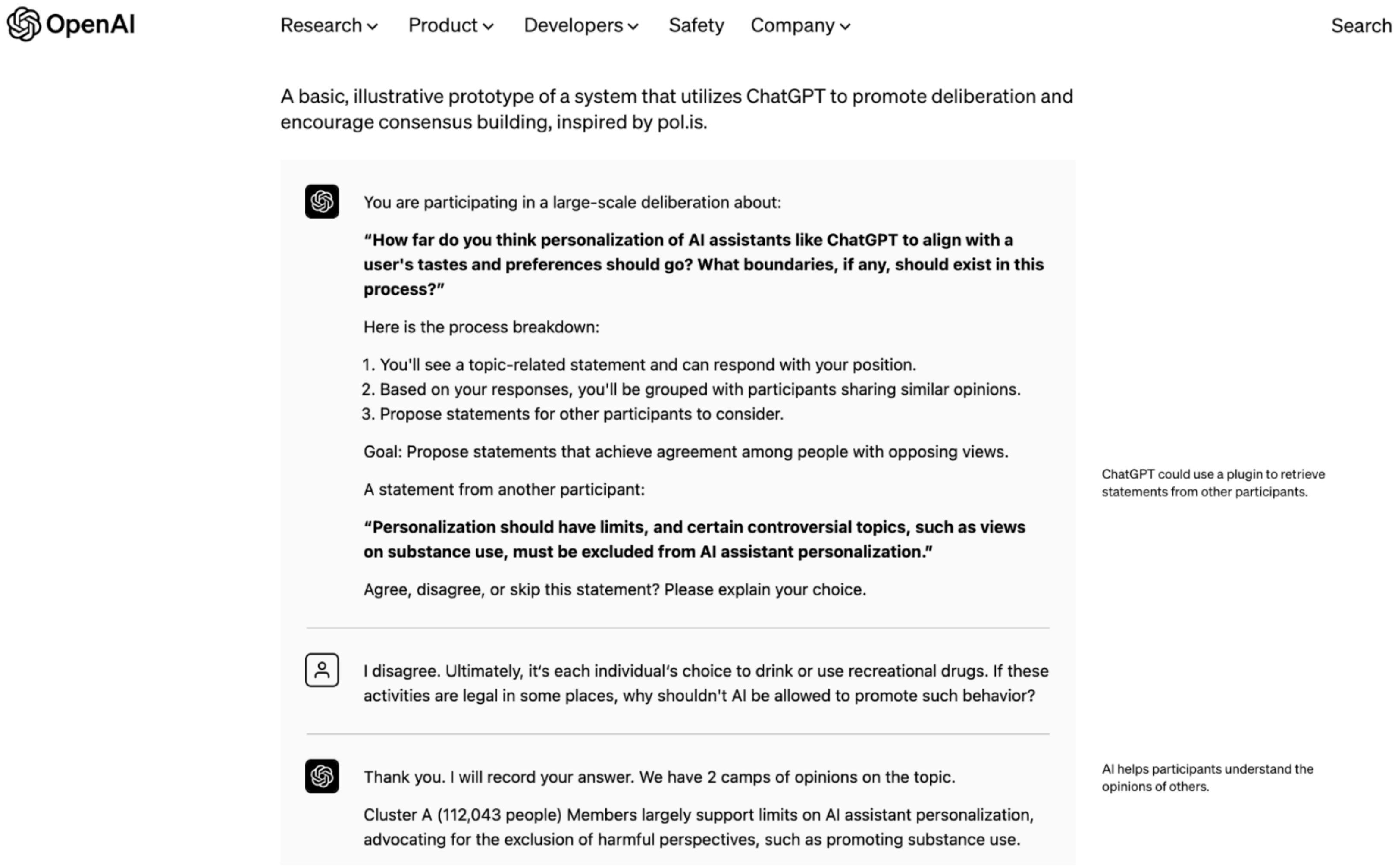
Democratic inputs to AI

Our nonprofit organization, OpenAI, Inc., is launching a program to award ten $100,000 grants to fund experiments in setting up a ...



# Democratic inputs to AI

Our nonprofit organization, OpenAI, Inc., is launching a program to award ten $100,000 grants to fund experiments in setting up a democratic process for deciding what rules AI systems should follow, within the bounds defined by the law.



A basic, illustrative prototype of a system that utilizes ChatGPT to promote deliberation and encourage consensus building, inspired by pol.is.

You are participating in a large-scale deliberation about:

**“How far do you think personalization of AI assistants like ChatGPT to align with a user's tastes and preferences should go? What boundaries, if any, should exist in this process?”**

Here is the process breakdown:

1. You'll see a topic-related statement and can respond with your position.
2. Based on your responses, you'll be grouped with participants sharing similar opinions.
3. Propose statements for other participants to consider.

Goal: Propose statements that achieve agreement among people with opposing views.

A statement from another participant:

**“Personalization should have limits, and certain controversial topics, such as views on substance use, must be excluded from AI assistant personalization.”**

Agree, disagree, or skip this statement? Please explain your choice.

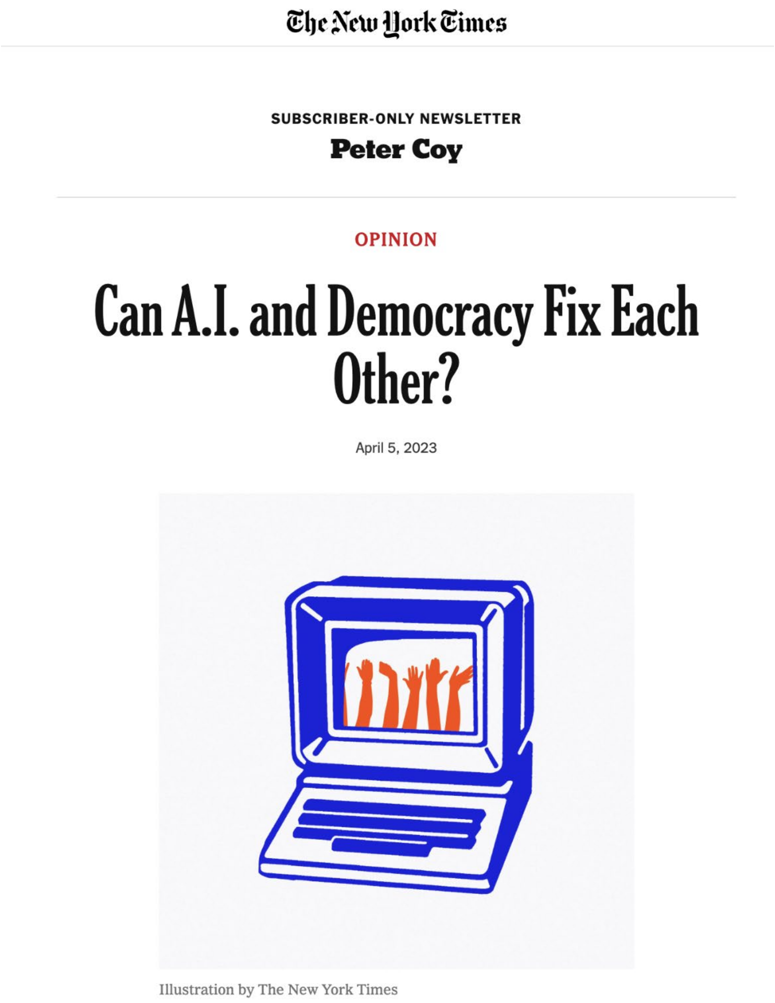
ChatGPT could use a plugin to retrieve statements from other participants.

I disagree. Ultimately, it‘s each individual‘s choice to drink or use recreational drugs. If these activities are legal in some places, why shouldn't AI be allowed to promote such behavior?

Thank you. I will record your answer. We have 2 camps of opinions on the topic.

AI helps participants understand the opinions of others.

Cluster A (112,043 people) Members largely support limits on AI assistant personalization, advocating for the exclusion of harmful perspectives, such as promoting substance use.



The New York Times

SUBSCRIBER-ONLY NEWSLETTER

**Peter Coy**

OPINION

# Can A.I. and Democracy Fix Each Other?

April 5, 2023

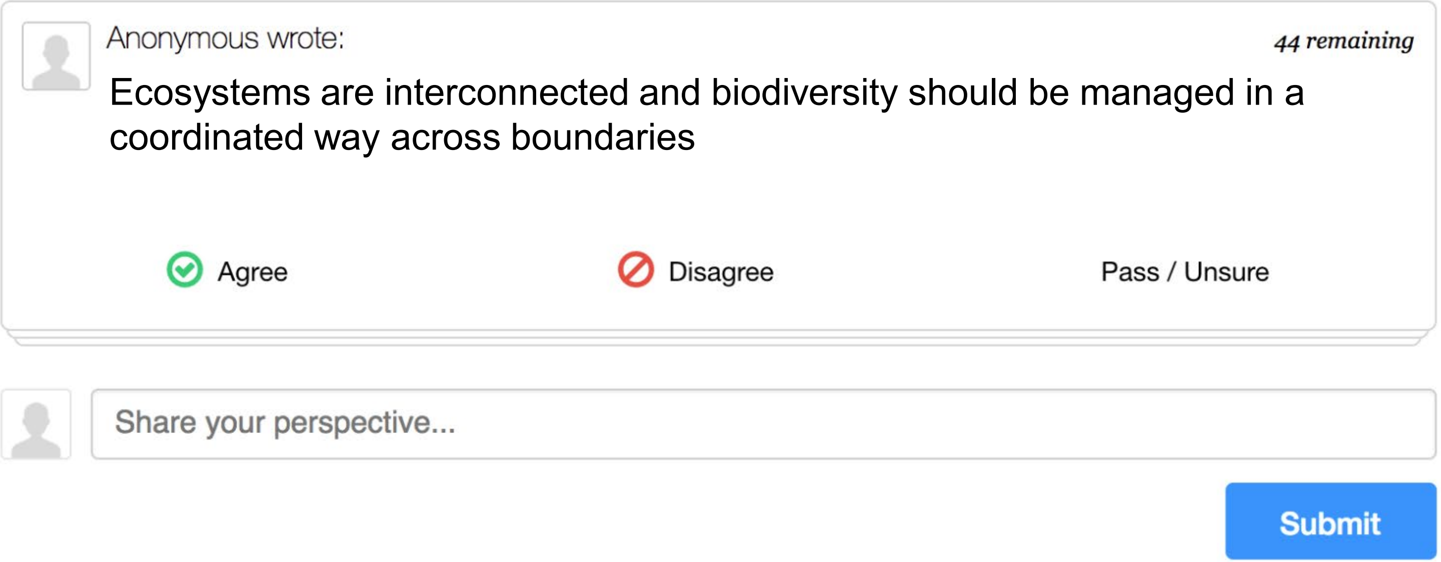
Illustration by The New York Times

colinmegill.com/vienna

**polis** is an open source platform

for efficiently gathering & making meaning of perspectives at scale

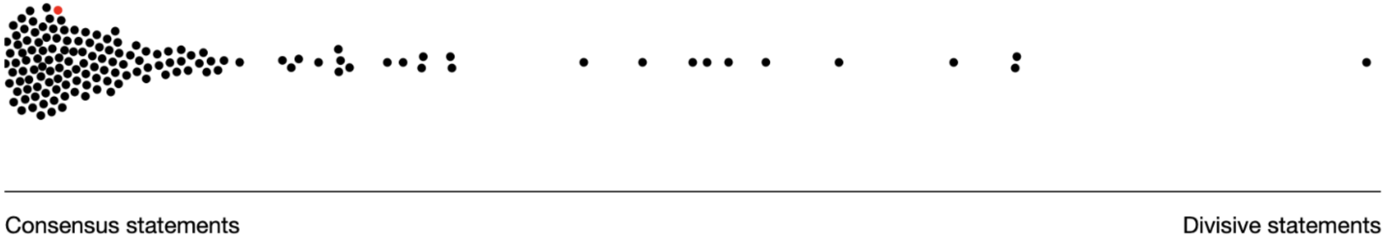
using machine learning



Anonymous wrote:

*44 remaining*

Ecosystems are interconnected and biodiversity should be managed in a coordinated way across boundaries



| STATEMENT | | OVERALL | A 72 | B 371 |
|---|---|---|---|---|
| 19 | Ecosystems are interconnected and biodiversity should be managed in a coordinated way across boundaries | 90% 3% 5% (334) | 63% 15% 21% (66) | 97% 0% 2% (268) |

# IRAN GREEN MOVEMENT / TAHRIR SQUARE / OCCUPY

LOTS OF PEOPLE
LOTS OF ISSUES
LOTS OF INTERESTS
LOTS OF POSITIONS

# **MACHINE LEARNING** FOR **DELIBERATIVE DEMOCRACY**

PROJECT BEGAN IN **2012**

COMPLETELY **OPEN SOURCE**

Business Impact

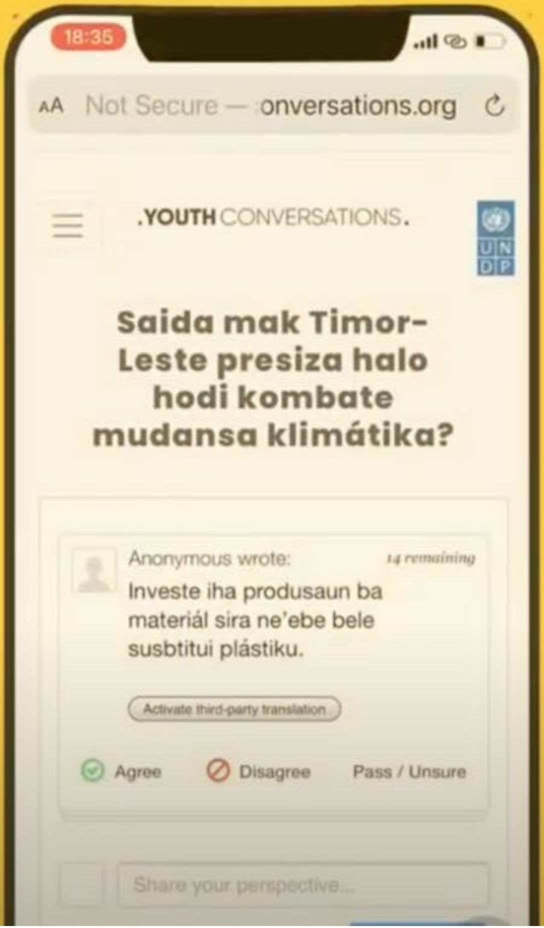
# The simple but ingenious system Taiwan uses to crowdsource its laws

vTaiwan is a promising experiment in participatory governance. But politics is blocking it from getting greater traction.

by Chris Horton August 21, 2018

MIT
Technology
Review

# UNDP
# PAKISTAN/BHUTAN/
# EAST TIMOR


18:35
AA Not Secure — :onversations.org
.YOUTH CONVERSATIONS.
UNDP

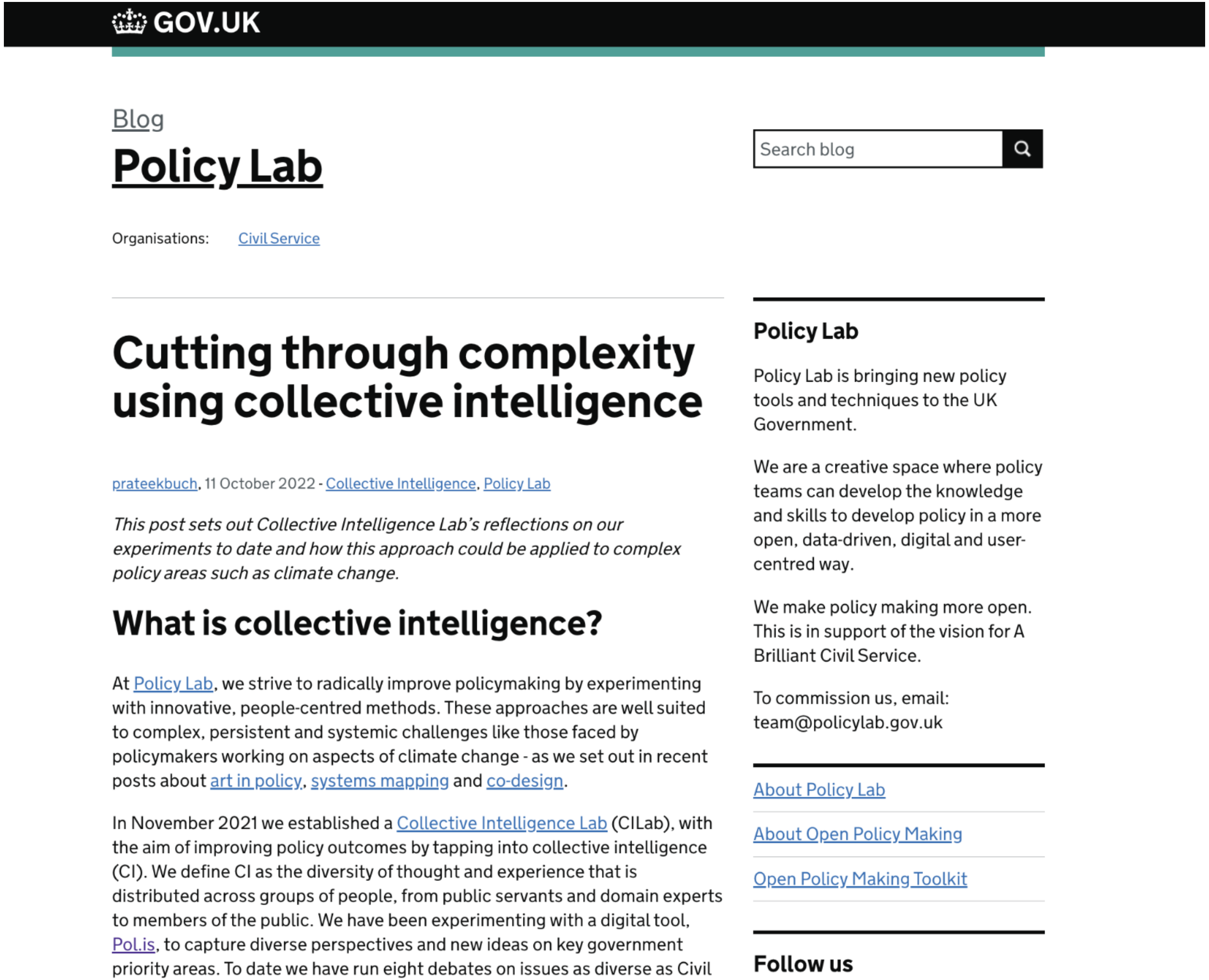
Saida mak Timor-Leste presiza halo hodi kombate mudansa klimátika?

Anonymous wrote:

*14 remaining*

Investe iha produsaun ba materiál sira ne'ebe bele susbtitui plástiku.

Activate third-party translation
Agree
Disagree
Pass / Unsure
Share your perspective...



Blog

# Policy Lab

Organisations: Civil Service

## Cutting through complexity using collective intelligence

prateekbuch, 11 October 2022 - Collective Intelligence, Policy Lab

*This post sets out Collective Intelligence Lab's reflections on our experiments to date and how this approach could be applied to complex policy areas such as climate change.*

### What is collective intelligence?

At Policy Lab, we strive to radically improve policymaking by experimenting with innovative, people-centred methods. These approaches are well suited to complex, persistent and systemic challenges like those faced by policymakers working on aspects of climate change - as we set out in recent posts about art in policy, systems mapping and co-design.

In November 2021 we established a Collective Intelligence Lab (CILab), with the aim of improving policy outcomes by tapping into collective intelligence (CI). We define CI as the diversity of thought and experience that is distributed across groups of people, from public servants and domain experts to members of the public. We have been experimenting with a digital tool, Pol.is, to capture diverse perspectives and new ideas on key government priority areas. To date we have run eight debates on issues as diverse as Civil

**Policy Lab**

Policy Lab is bringing new policy tools and techniques to the UK Government.

We are a creative space where policy teams can develop the knowledge and skills to develop policy in a more open, data-driven, digital and user-centred way.

We make policy making more open. This is in support of the vision for A Brilliant Civil Service.

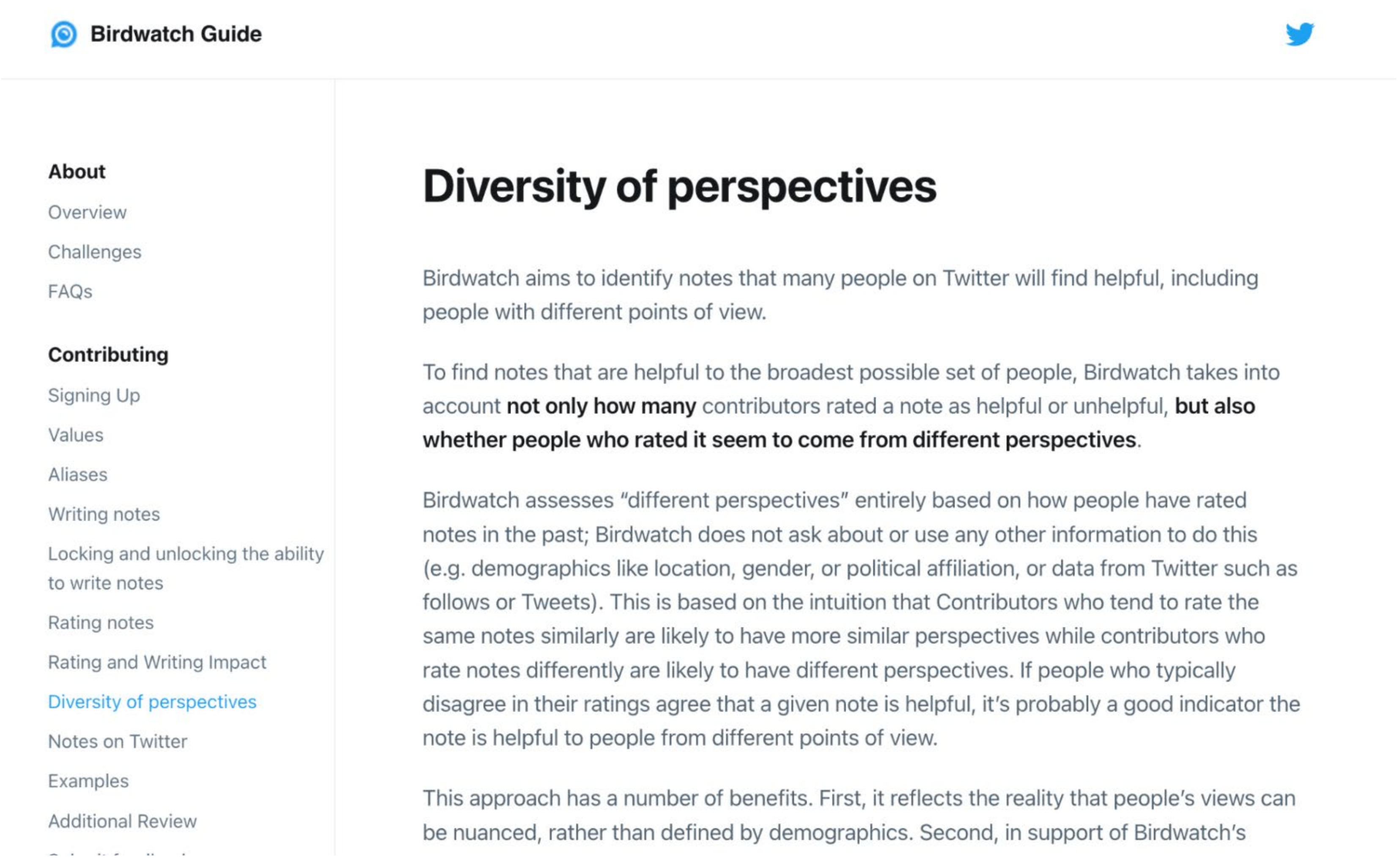
To commission us, email: team@policylab.gov.uk

- About Policy Lab
- About Open Policy Making
- Open Policy Making Toolkit

**Follow us**

# How to Put Out Democracy's Dumpster Fire

Our democratic habits have been killed off by an internet kleptocracy that profits from disinformation, polarization, and rage. Here's how to fix that.



Birdwatch Guide

**About**

- Overview
- Challenges
- FAQs

**Contributing**

- Signing Up
- Values
- Aliases
- Writing notes
- Locking and unlocking the ability to write notes
- Rating notes
- Rating and Writing Impact
- Diversity of perspectives
- Notes on Twitter
- Examples
- Additional Review

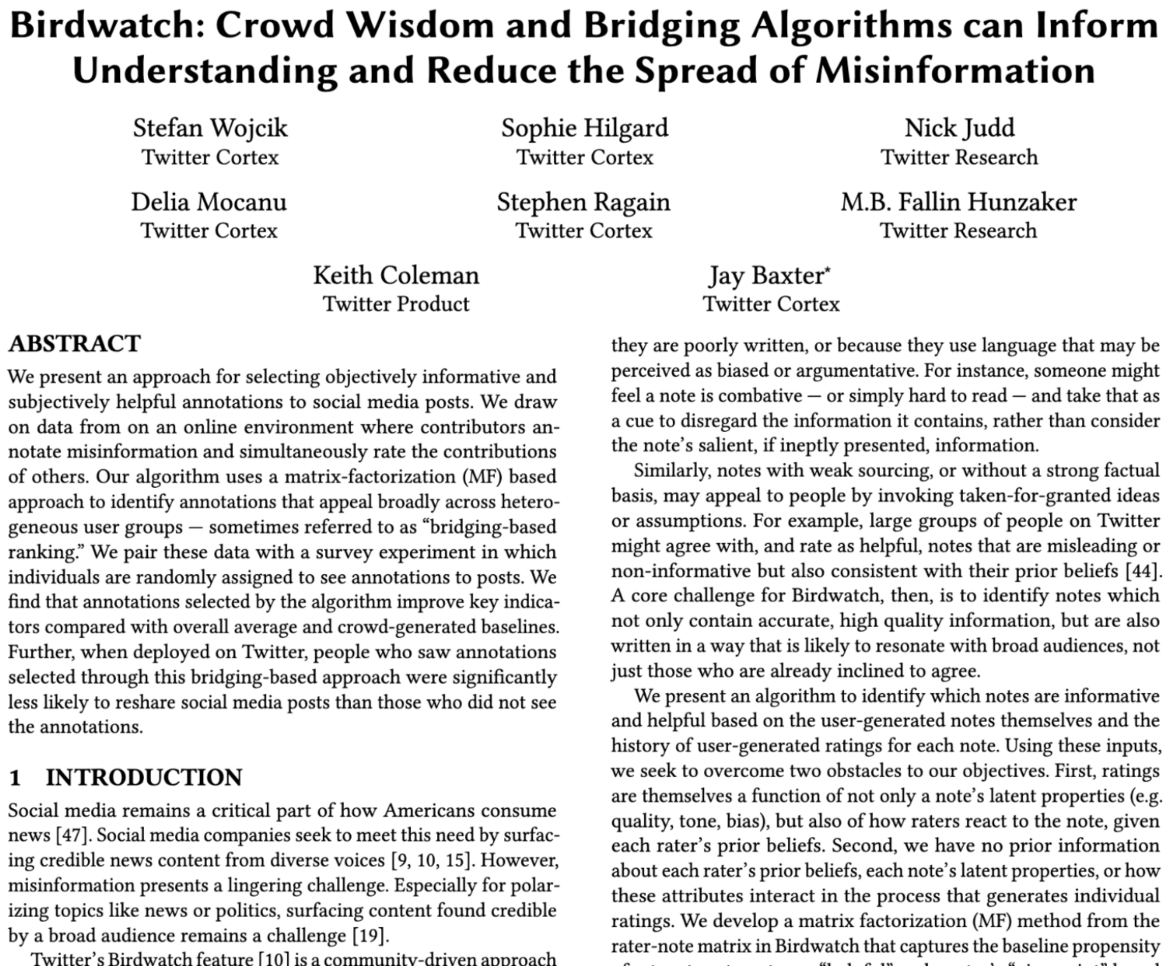
# Diversity of perspectives

Birdwatch aims to identify notes that many people on Twitter will find helpful, including people with different points of view.

To find notes that are helpful to the broadest possible set of people, Birdwatch takes into account **not only how many** contributors rated a note as helpful or unhelpful, **but also whether people who rated it seem to come from different perspectives**.

Birdwatch assesses “different perspectives” entirely based on how people have rated notes in the past; Birdwatch does not ask about or use any other information to do this (e.g. demographics like location, gender, or political affiliation, or data from Twitter such as follows or Tweets). This is based on the intuition that Contributors who tend to rate the same notes similarly are likely to have more similar perspectives while contributors who rate notes differently are likely to have different perspectives. If people who typically disagree in their ratings agree that a given note is helpful, it’s probably a good indicator the note is helpful to people from different points of view.

This approach has a number of benefits. First, it reflects the reality that people’s views can be nuanced, rather than defined by demographics. Second, in support of Birdwatch’s



# Birdwatch: Crowd Wisdom and Bridging Algorithms can Inform Understanding and Reduce the Spread of Misinformation

Stefan Wojcik
Twitter Cortex

Sophie Hilgard
Twitter Cortex

Nick Judd
Twitter Research

Delia Mocanu
Twitter Cortex

Stephen Ragain
Twitter Cortex

M.B. Fallin Hunzaker
Twitter Research

Keith Coleman
Twitter Product

Jay Baxter*
Twitter Cortex

## ABSTRACT

We present an approach for selecting objectively informative and subjectively helpful annotations to social media posts. We draw on data from on an online environment where contributors annotate misinformation and simultaneously rate the contributions of others. Our algorithm uses a matrix-factorization (MF) based approach to identify annotations that appeal broadly across heterogeneous user groups — sometimes referred to as "bridging-based ranking." We pair these data with a survey experiment in which individuals are randomly assigned to see annotations to posts. We find that annotations selected by the algorithm improve key indicators compared with overall average and crowd-generated baselines. Further, when deployed on Twitter, people who saw annotations selected through this bridging-based approach were significantly less likely to reshare social media posts than those who did not see the annotations.

## 1 INTRODUCTION

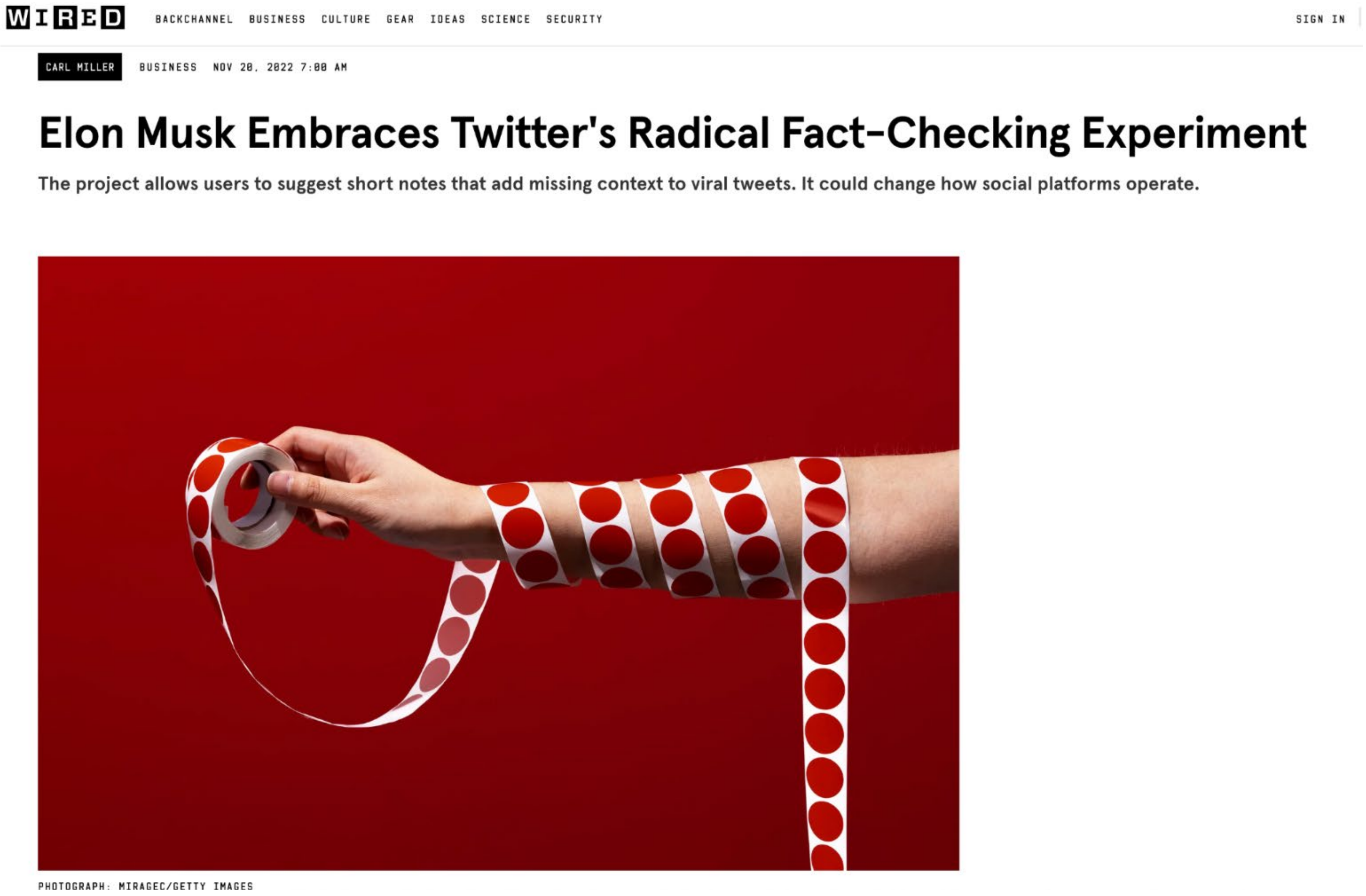
Social media remains a critical part of how Americans consume news [47]. Social media companies seek to meet this need by surfacing credible news content from diverse voices [9, 10, 15]. However, misinformation presents a lingering challenge. Especially for polarizing topics like news or politics, surfacing content found credible by a broad audience remains a challenge [19].

Twitter's Birdwatch feature [10] is a community-driven approach

they are poorly written, or because they use language that may be perceived as biased or argumentative. For instance, someone might feel a note is combative — or simply hard to read — and take that as a cue to disregard the information it contains, rather than consider the note's salient, if ineptly presented, information.

Similarly, notes with weak sourcing, or without a strong factual basis, may appeal to people by invoking taken-for-granted ideas or assumptions. For example, large groups of people on Twitter might agree with, and rate as helpful, notes that are misleading or non-informative but also consistent with their prior beliefs [44]. A core challenge for Birdwatch, then, is to identify notes which not only contain accurate, high quality information, but are also written in a way that is likely to resonate with broad audiences, not just those who are already inclined to agree.

We present an algorithm to identify which notes are informative and helpful based on the user-generated notes themselves and the history of user-generated ratings for each note. Using these inputs, we seek to overcome two obstacles to our objectives. First, ratings are themselves a function of not only a note's latent properties (e.g. quality, tone, bias), but also of how raters react to the note, given each rater's prior beliefs. Second, we have no prior information about each rater's prior beliefs, each note's latent properties, or how these attributes interact in the process that generates individual ratings. We develop a matrix factorization (MF) method from the rater-note matrix in Birdwatch that captures the baseline propensity

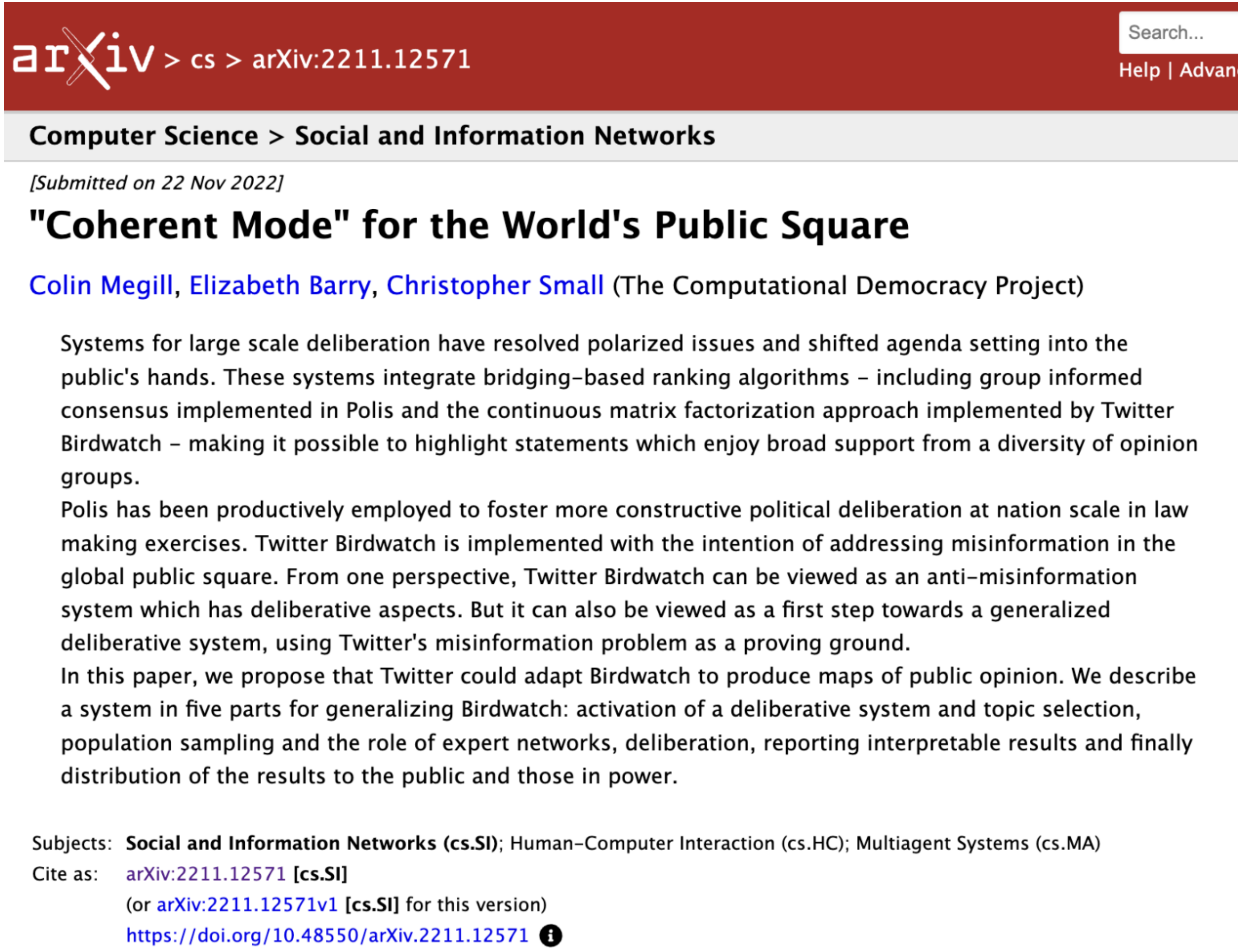


CARL MILLER BUSINESS NOV 20, 2022 7:00 AM

# Elon Musk Embraces Twitter's Radical Fact-Checking Experiment

The project allows users to suggest short notes that add missing context to viral tweets. It could change how social platforms operate.

PHOTOGRAPH: MIRAGEC/GETTY IMAGES



arXiv > cs > arXiv:2211.12571

**Computer Science > Social and Information Networks**

*[Submitted on 22 Nov 2022]*

# "Coherent Mode" for the World's Public Square

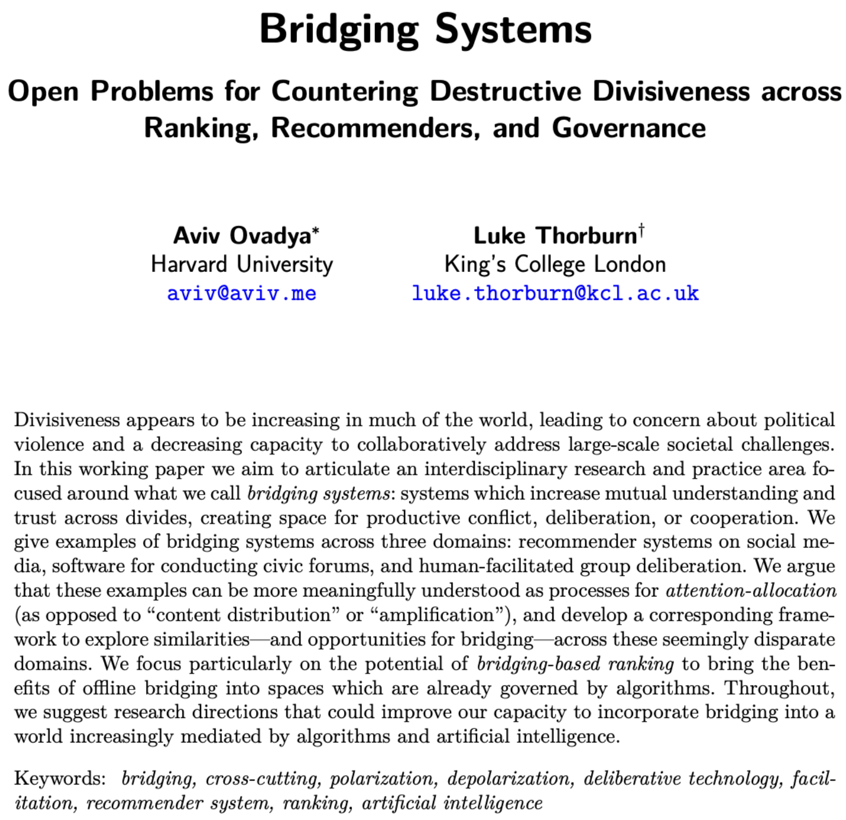
Colin Megill, Elizabeth Barry, Christopher Small (The Computational Democracy Project)

Systems for large scale deliberation have resolved polarized issues and shifted agenda setting into the public's hands. These systems integrate bridging-based ranking algorithms - including group informed consensus implemented in Polis and the continuous matrix factorization approach implemented by Twitter Birdwatch - making it possible to highlight statements which enjoy broad support from a diversity of opinion groups.

Polis has been productively employed to foster more constructive political deliberation at nation scale in law making exercises. Twitter Birdwatch is implemented with the intention of addressing misinformation in the global public square. From one perspective, Twitter Birdwatch can be viewed as an anti-misinformation system which has deliberative aspects. But it can also be viewed as a first step towards a generalized deliberative system, using Twitter's misinformation problem as a proving ground.

In this paper, we propose that Twitter could adapt Birdwatch to produce maps of public opinion. We describe a system in five parts for generalizing Birdwatch: activation of a deliberative system and topic selection, population sampling and the role of expert networks, deliberation, reporting interpretable results and finally distribution of the results to the public and those in power.

Subjects: **Social and Information Networks (cs.SI)**; Human-Computer Interaction (cs.HC); Multiagent Systems (cs.MA)
Cite as: arXiv:2211.12571 **[cs.SI]**
(or arXiv:2211.12571v1 **[cs.SI]** for this version)
https://doi.org/10.48550/arXiv.2211.12571



# Bridging Systems

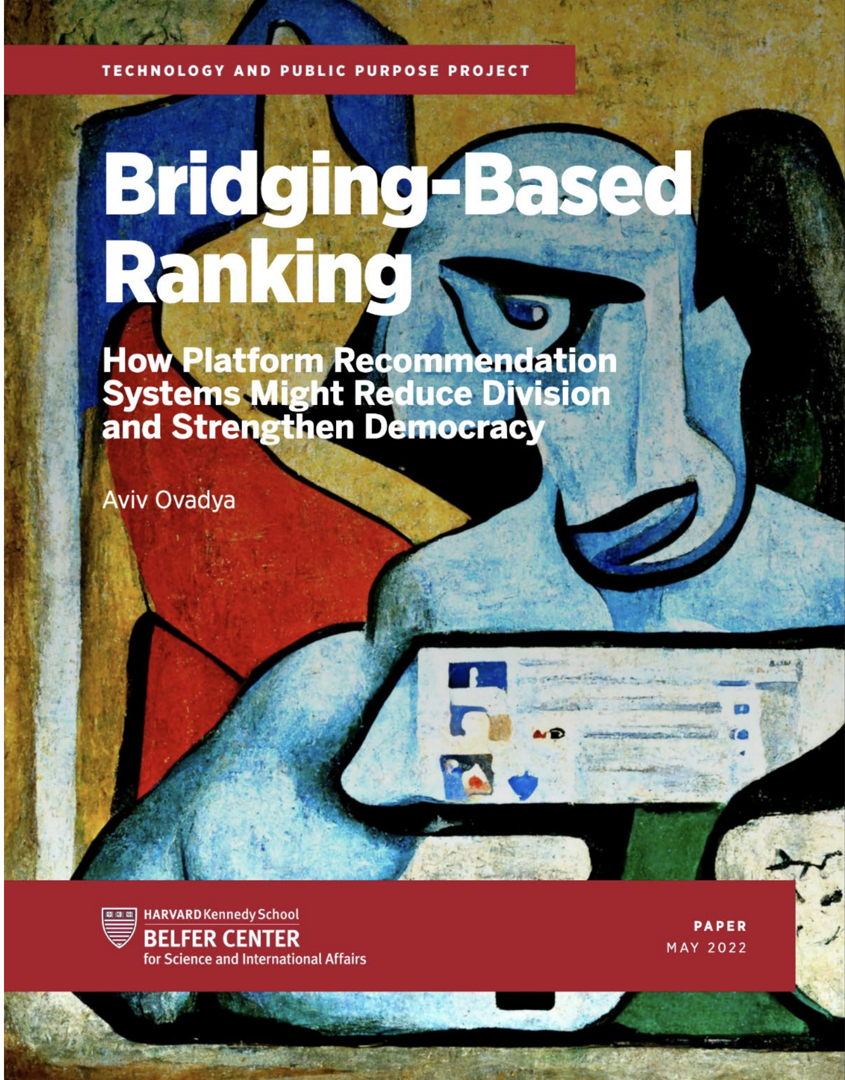
## Open Problems for Countering Destructive Divisiveness across Ranking, Recommenders, and Governance

**Aviv Ovadya***
Harvard University
aviv@aviv.me

**Luke Thorburn**†
King's College London
luke.thorburn@kcl.ac.uk

Divisiveness appears to be increasing in much of the world, leading to concern about political violence and a decreasing capacity to collaboratively address large-scale societal challenges. In this working paper we aim to articulate an interdisciplinary research and practice area focused around what we call *bridging systems*: systems which increase mutual understanding and trust across divides, creating space for productive conflict, deliberation, or cooperation. We give examples of bridging systems across three domains: recommender systems on social media, software for conducting civic forums, and human-facilitated group deliberation. We argue that these examples can be more meaningfully understood as processes for *attention-allocation* (as opposed to "content distribution" or "amplification"), and develop a corresponding framework to explore similarities—and opportunities for bridging—across these seemingly disparate domains. We focus particularly on the potential of *bridging-based ranking* to bring the benefits of offline bridging into spaces which are already governed by algorithms. Throughout, we suggest research directions that could improve our capacity to incorporate bridging into a world increasingly mediated by algorithms and artificial intelligence.

Keywords: *bridging, cross-cutting, polarization, depolarization, deliberative technology, facilitation, recommender system, ranking, artificial intelligence*



TECHNOLOGY AND PUBLIC PURPOSE PROJECT

# Bridging-Based Ranking

## How Platform Recommendation Systems Might Reduce Division and Strengthen Democracy

Aviv Ovadya

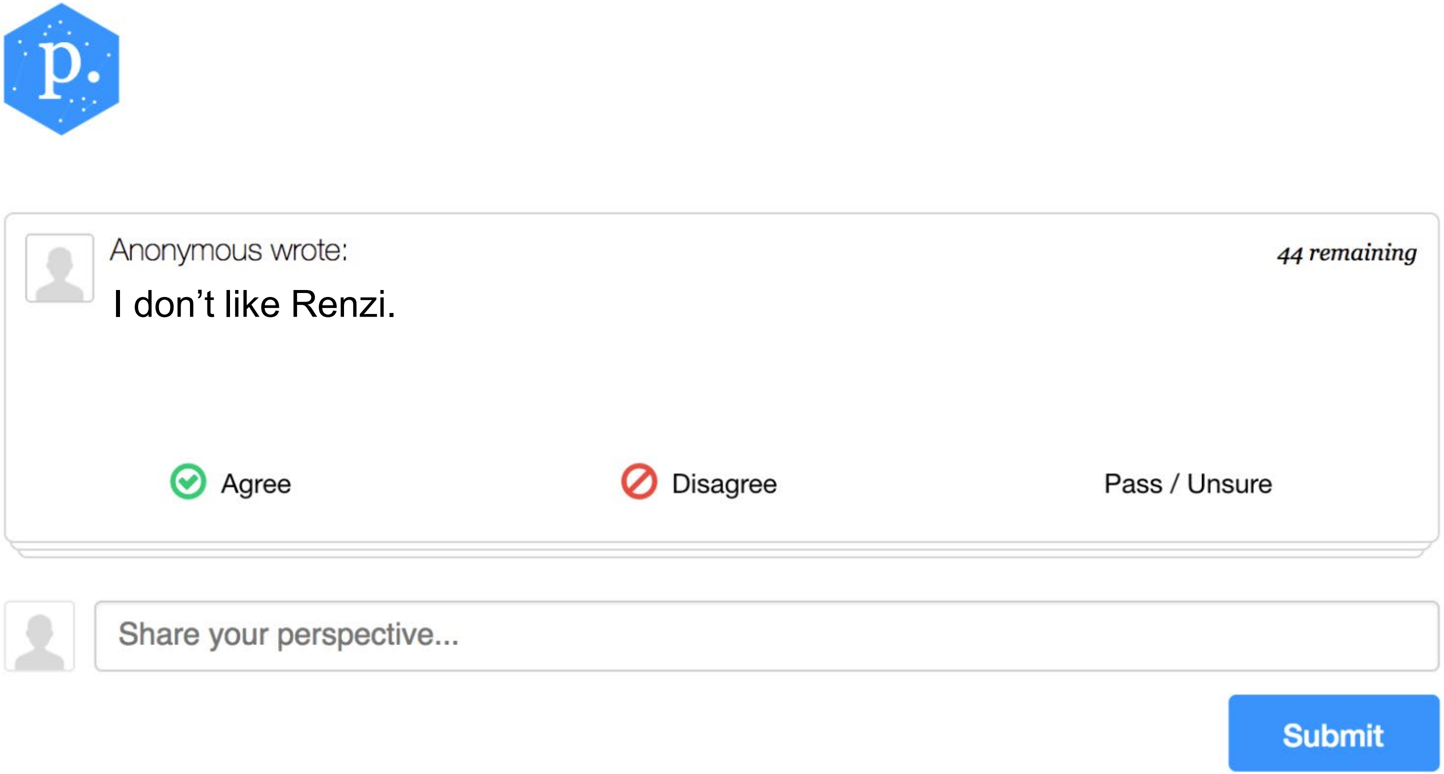
HARVARD Kennedy School
BELFER CENTER
for Science and International Affairs

PAPER
MAY 2022

# PRINCIPLES & IMPLEMENTATION

# SCALEABLE + COHERENT

# NON VIOLENT COMMUNICATION



Anonymous wrote:

*44 remaining*

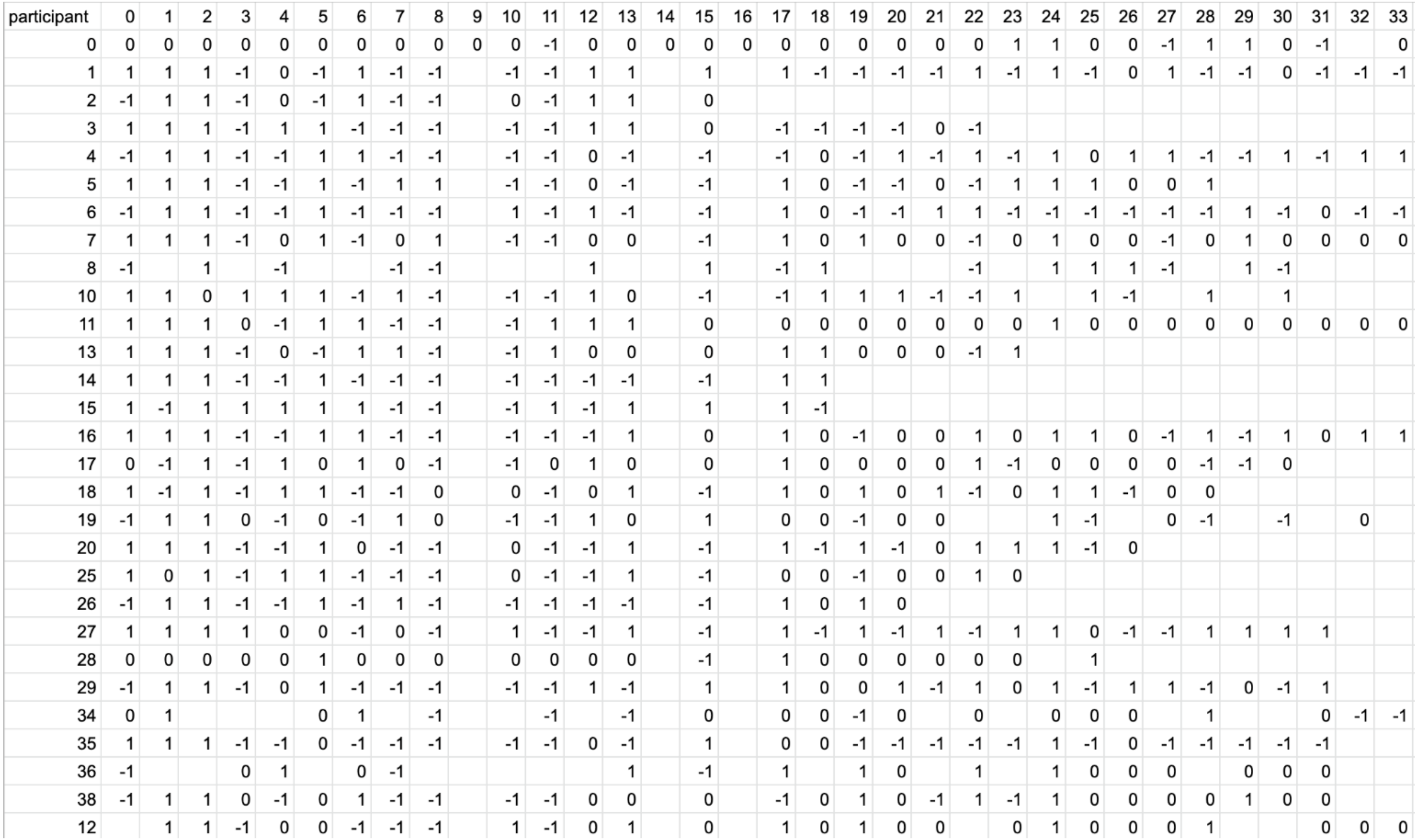
I don’t like Renzi.

what everyone has to say

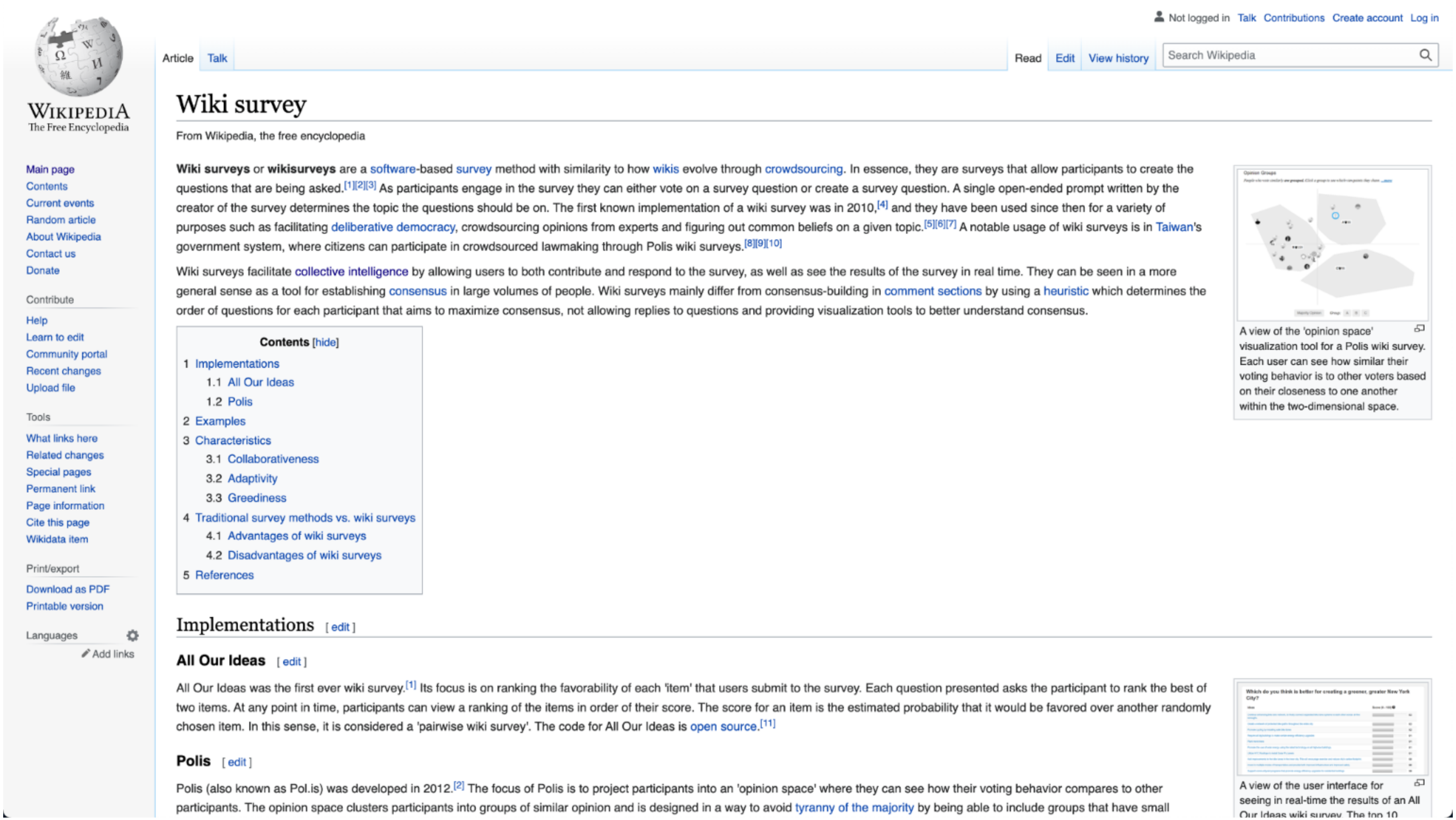
x

what everyone thinks about it

(as an ‘**opinion matrix**’)



| participant | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | -1 | 1 | 1 | 0 | -1 |  | 0 |
| 1 | 1 | 1 | 1 | -1 | 0 | -1 | 1 | -1 | -1 |  | -1 | -1 | 1 | 1 |  | 1 |  | 1 | -1 | -1 | -1 | -1 | 1 | -1 | 1 | -1 | 0 | 1 | -1 | -1 | 0 | -1 | -1 | -1 |
| 2 | -1 | 1 | 1 | -1 | 0 | -1 | 1 | -1 | -1 |  | 0 | -1 | 1 | 1 |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | 1 | 1 | 1 | -1 | 1 | 1 | -1 | -1 | -1 |  | -1 | -1 | 1 | 1 |  | 0 |  | -1 | -1 | -1 | -1 | 0 | -1 |  |  |  |  |  |  |  |  |  |  |  |
| 4 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 |  | -1 | -1 | 0 | -1 |  | -1 |  | -1 | 0 | -1 | 1 | -1 | 1 | -1 | 1 | 0 | 1 | 1 | -1 | -1 | 1 | -1 | 1 | 1 |
| 5 | 1 | 1 | 1 | -1 | -1 | 1 | -1 | 1 | 1 |  | -1 | -1 | 0 | -1 |  | -1 |  | 1 | 0 | -1 | -1 | 0 | -1 | 1 | 1 | 1 | 0 | 0 | 1 |  |  |  |  |  |
| 6 | -1 | 1 | 1 | -1 | -1 | 1 | -1 | -1 | -1 |  | 1 | -1 | 1 | -1 |  | -1 |  | 1 | 0 | -1 | -1 | 1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | 1 | -1 | 0 | -1 | -1 |
| 7 | 1 | 1 | 1 | -1 | 0 | 1 | -1 | 0 | 1 |  | -1 | -1 | 0 | 0 |  | -1 |  | 1 | 0 | 1 | 0 | 0 | -1 | 0 | 1 | 0 | 0 | -1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 8 | -1 |  | 1 |  | -1 |  |  | -1 | -1 |  |  |  | 1 |  |  | 1 |  | -1 | 1 |  |  |  | -1 |  | 1 | 1 | 1 | -1 |  | 1 | -1 |  |  |  |
| 10 | 1 | 1 | 0 | 1 | 1 | 1 | -1 | 1 | -1 |  | -1 | -1 | 1 | 0 |  | -1 |  | -1 | 1 | 1 | 1 | -1 | -1 | 1 |  | 1 | -1 |  | 1 |  | 1 |  |  |  |
| 11 | 1 | 1 | 1 | 0 | -1 | 1 | 1 | -1 | -1 |  | -1 | 1 | 1 | 1 |  | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 1 | 1 | 1 | -1 | 0 | -1 | 1 | 1 | -1 |  | -1 | 1 | 0 | 0 |  | 0 |  | 1 | 1 | 0 | 0 | 0 | -1 | 1 |  |  |  |  |  |  |  |  |  |  |
| 14 | 1 | 1 | 1 | -1 | -1 | 1 | -1 | -1 | -1 |  | -1 | -1 | -1 | -1 |  | -1 |  | 1 | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 15 | 1 | -1 | 1 | 1 | 1 | 1 | 1 | -1 | -1 |  | -1 | 1 | -1 | 1 |  | 1 |  | 1 | -1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 16 | 1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 |  | -1 | -1 | -1 | 1 |  | 0 |  | 1 | 0 | -1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | -1 | 1 | -1 | 1 | 0 | 1 | 1 |
| 17 | 0 | -1 | 1 | -1 | 1 | 0 | 1 | 0 | -1 |  | -1 | 0 | 1 | 0 |  | 0 |  | 1 | 0 | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 0 | -1 | -1 | 0 |  |  |  |
| 18 | 1 | -1 | 1 | -1 | 1 | 1 | -1 | -1 | 0 |  | 0 | -1 | 0 | 1 |  | -1 |  | 1 | 0 | 1 | 0 | 1 | -1 | 0 | 1 | 1 | -1 | 0 | 0 |  |  |  |  |  |
| 19 | -1 | 1 | 1 | 0 | -1 | 0 | -1 | 1 | 0 |  | -1 | -1 | 1 | 0 |  | 1 |  | 0 | 0 | -1 | 0 | 0 |  |  | 1 | -1 |  | 0 | -1 |  | -1 |  | 0 |  |
| 20 | 1 | 1 | 1 | -1 | -1 | 1 | 0 | -1 | -1 |  | 0 | -1 | -1 | 1 |  | -1 |  | 1 | -1 | 1 | -1 | 0 | 1 | 1 | 1 | -1 | 0 |  |  |  |  |  |  |  |
| 25 | 1 | 0 | 1 | -1 | 1 | 1 | -1 | -1 | -1 |  | 0 | -1 | -1 | 1 |  | -1 |  | 0 | 0 | -1 | 0 | 0 | 1 | 0 |  |  |  |  |  |  |  |  |  |  |
| 26 | -1 | 1 | 1 | -1 | -1 | 1 | -1 | 1 | -1 |  | -1 | -1 | -1 | -1 |  | -1 |  | 1 | 0 | 1 | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 27 | 1 | 1 | 1 | 1 | 0 | 0 | -1 | 0 | -1 |  | 1 | -1 | -1 | 1 |  | -1 |  | 1 | -1 | 1 | -1 | 1 | -1 | 1 | 1 | 0 | -1 | -1 | 1 | 1 | 1 | 1 |  |  |
| 28 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 |  | -1 |  | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 1 |  |  |  |  |  |  |  |  |
| 29 | -1 | 1 | 1 | -1 | 0 | 1 | -1 | -1 | -1 |  | -1 | -1 | 1 | -1 |  | 1 |  | 1 | 0 | 0 | 1 | -1 | 1 | 0 | 1 | -1 | 1 | 1 | -1 | 0 | -1 | 1 |  |  |
| 34 | 0 | 1 |  |  |  | 0 | 1 |  | -1 |  |  | -1 |  | -1 |  | 0 |  | 0 | 0 | -1 | 0 |  | 0 |  | 0 | 0 | 0 |  | 1 |  |  | 0 | -1 | -1 |
| 35 | 1 | 1 | 1 | -1 | -1 | 0 | -1 | -1 | -1 |  | -1 | -1 | 0 | -1 |  | 1 |  | 0 | 0 | -1 | -1 | -1 | -1 | -1 | 1 | -1 | 0 | -1 | -1 | -1 | -1 | -1 |  |  |
| 36 | -1 |  |  | 0 | 1 |  | 0 | -1 |  |  |  |  |  | 1 |  | -1 |  | 1 |  | 1 | 0 |  | 1 |  | 1 | 0 | 0 | 0 |  | 0 | 0 | 0 |  |  |
| 38 | -1 | 1 | 1 | 0 | -1 | 0 | 1 | -1 | -1 |  | -1 | -1 | 0 | 0 |  | 0 |  | -1 | 0 | 1 | 0 | -1 | 1 | -1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |  |  |
| 12 |  | 1 | 1 | -1 | 0 | 0 | -1 | -1 | -1 |  | 1 | -1 | 0 | 1 |  | 0 |  | 1 | 0 | 1 | 0 | 0 |  | 0 | 1 | 0 | 0 | 0 | 1 |  |  | 0 | 0 | 0 |
]



# Wiki survey

From Wikipedia, the free encyclopedia

**Wiki surveys** or **wikisurveys** are a software-based survey method with similarity to how wikis evolve through crowdsourcing. In essence, they are surveys that allow participants to create the questions that are being asked.[1][2][3] As participants engage in the survey they can either vote on a survey question or create a survey question. A single open-ended prompt written by the creator of the survey determines the topic the questions should be on. The first known implementation of a wiki survey was in 2010,[4] and they have been used since then for a variety of purposes such as facilitating deliberative democracy, crowdsourcing opinions from experts and figuring out common beliefs on a given topic.[5][6][7] A notable usage of wiki surveys is in Taiwan's government system, where citizens can participate in crowdsourced lawmaking through Polis wiki surveys.[8][9][10]

A view of the 'opinion space' visualization tool for a Polis wiki survey. Each user can see how similar their voting behavior is to other voters based on their closeness to one another within the two-dimensional space.

Wiki surveys facilitate collective intelligence by allowing users to both contribute and respond to the survey, as well as see the results of the survey in real time. They can be seen in a more general sense as a tool for establishing consensus in large volumes of people. Wiki surveys mainly differ from consensus-building in comment sections by using a heuristic which determines the order of questions for each participant that aims to maximize consensus, not allowing replies to questions and providing visualization tools to better understand consensus.

**Contents** [hide]

1 Implementations
  1.1 All Our Ideas
  1.2 Polis
2 Examples
3 Characteristics
  3.1 Collaborativeness
  3.2 Adaptivity
  3.3 Greediness
4 Traditional survey methods vs. wiki surveys
  4.1 Advantages of wiki surveys
  4.2 Disadvantages of wiki surveys
5 References

## Implementations [ edit ]

### All Our Ideas [ edit ]

All Our Ideas was the first ever wiki survey.[1] Its focus is on ranking the favorability of each 'item' that users submit to the survey. Each question presented asks the participant to rank the best of two items. At any point in time, participants can view a ranking of the items in order of their score. The score for an item is the estimated probability that it would be favored over another randomly chosen item. In this sense, it is considered a 'pairwise wiki survey'. The code for All Our Ideas is open source.[11]

A view of the user interface for seeing in real-time the results of an All Our Ideas wiki survey. The top 10

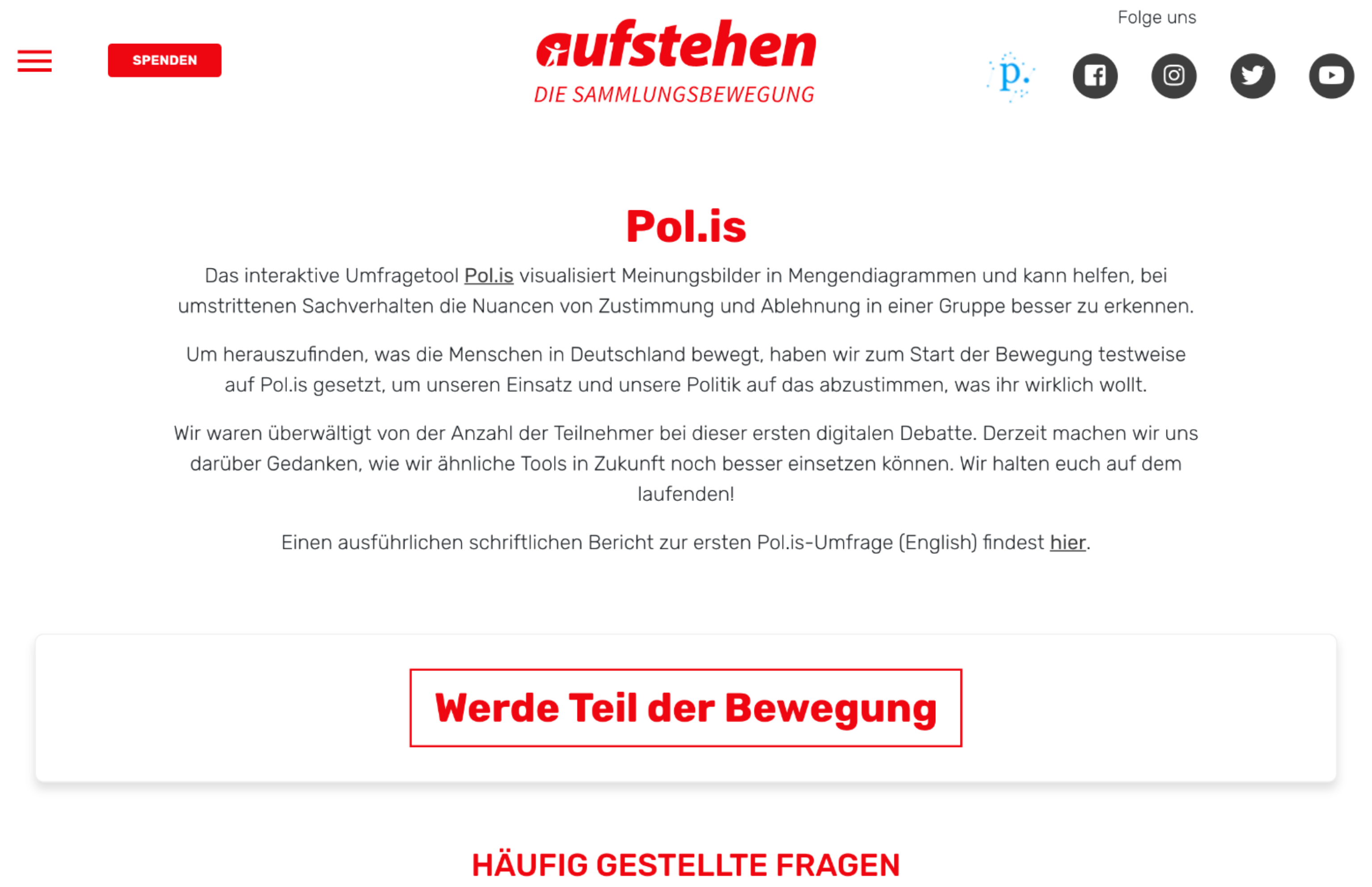
### Polis [ edit ]

Polis (also known as Pol.is) was developed in 2012.[2] The focus of Polis is to project participants into an 'opinion space' where they can see how their voting behavior compares to other participants. The opinion space clusters participants into groups of similar opinion and is designed in a way to avoid tyranny of the majority by being able to include groups that have small

# UNSUPERVISED

# UNSUPERVISED

# CLUSTER BY SIMILARITY



# Pol.is

Das interaktive Umfragetool Pol.is visualisiert Meinungsbilder in Mengendiagrammen und kann helfen, bei umstrittenen Sachverhalten die Nuancen von Zustimmung und Ablehnung in einer Gruppe besser zu erkennen.

Um herauszufinden, was die Menschen in Deutschland bewegt, haben wir zum Start der Bewegung testweise auf Pol.is gesetzt, um unseren Einsatz und unsere Politik auf das abzustimmen, was ihr wirklich wollt.

Wir waren überwältigt von der Anzahl der Teilnehmer bei dieser ersten digitalen Debatte. Derzeit machen wir uns darüber Gedanken, wie wir ähnliche Tools in Zukunft noch besser einsetzen können. Wir halten euch auf dem laufenden!

Einen ausführlichen schriftlichen Bericht zur ersten Pol.is-Umfrage (English) findest hier.

## HÄUFIG GESTELLTE FRAGEN

**33,547**
people voted

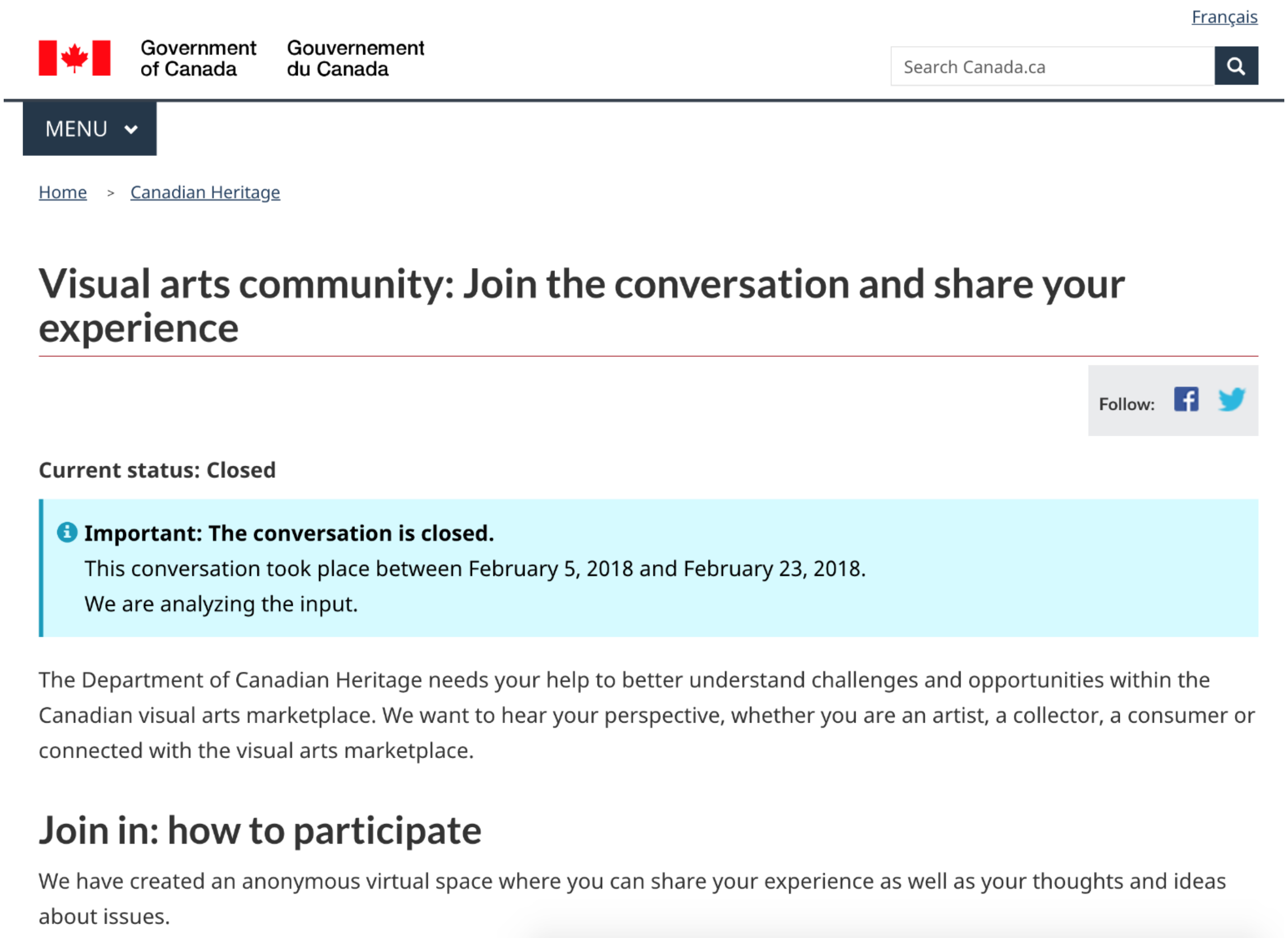
**23,354**
people grouped

**1,966,989**
votes were cast

**783**
statements were submitted

**58.63**
votes per voter on average

**1.79**
statements per author on average



# Visual arts community: Join the conversation and share your experience

**Current status: Closed**

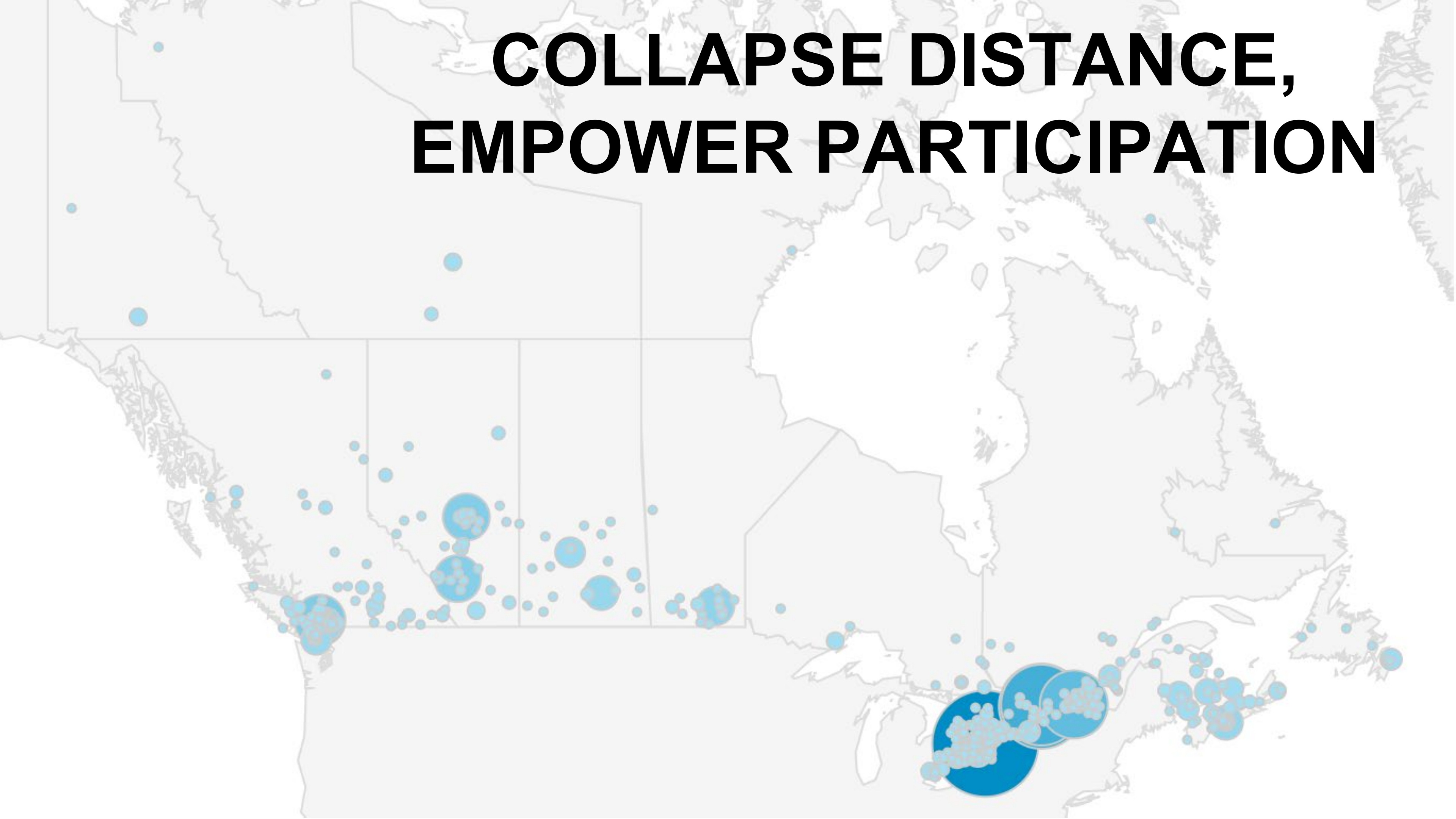
**Important: The conversation is closed.**
This conversation took place between February 5, 2018 and February 23, 2018.
We are analyzing the input.

The Department of Canadian Heritage needs your help to better understand challenges and opportunities within the Canadian visual arts marketplace. We want to hear your perspective, whether you are an artist, a collector, a consumer or connected with the visual arts marketplace.

## Join in: how to participate

We have created an anonymous virtual space where you can share your experience as well as your thoughts and ideas about issues.

# COLLAPSE DISTANCE, EMPOWER PARTICIPATION



753
people voted

570
people grouped

43,901
votes were cast

577
statements were submitted

58.30
votes per voter on average

2.57
statements per author on average

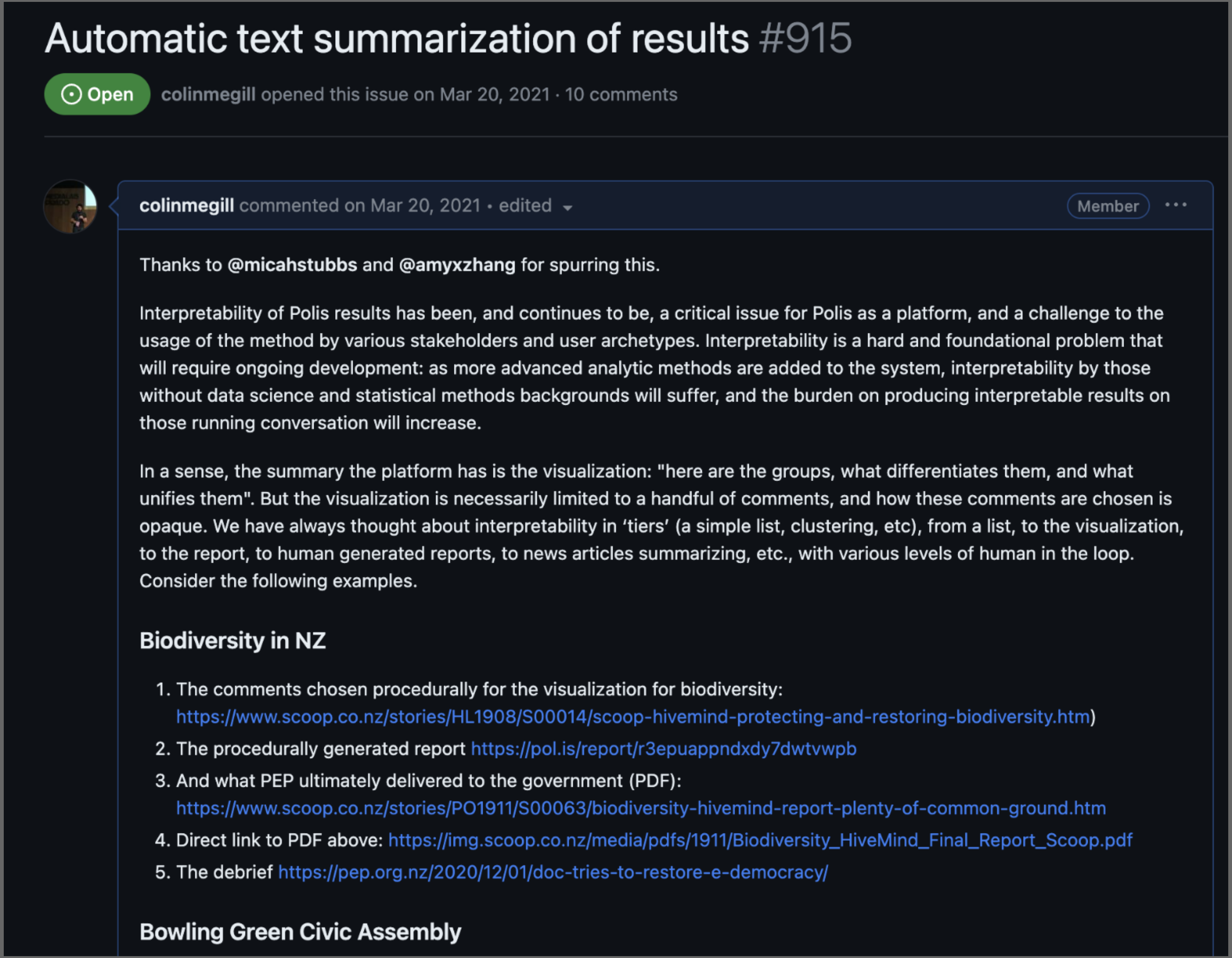
# WHAT COMMENT BOXES WON’T DO:

33k PEOPLE

1.9MM VOTES

# input to deliberative processes

**limitations:** number of comments, async voting, integrating results



# Automatic text summarization of results #915

**colinmegill** opened this issue on Mar 20, 2021 · 10 comments

**colinmegill** commented on Mar 20, 2021 • edited

Thanks to **@micahstubbs** and **@amyxzhang** for spurring this.

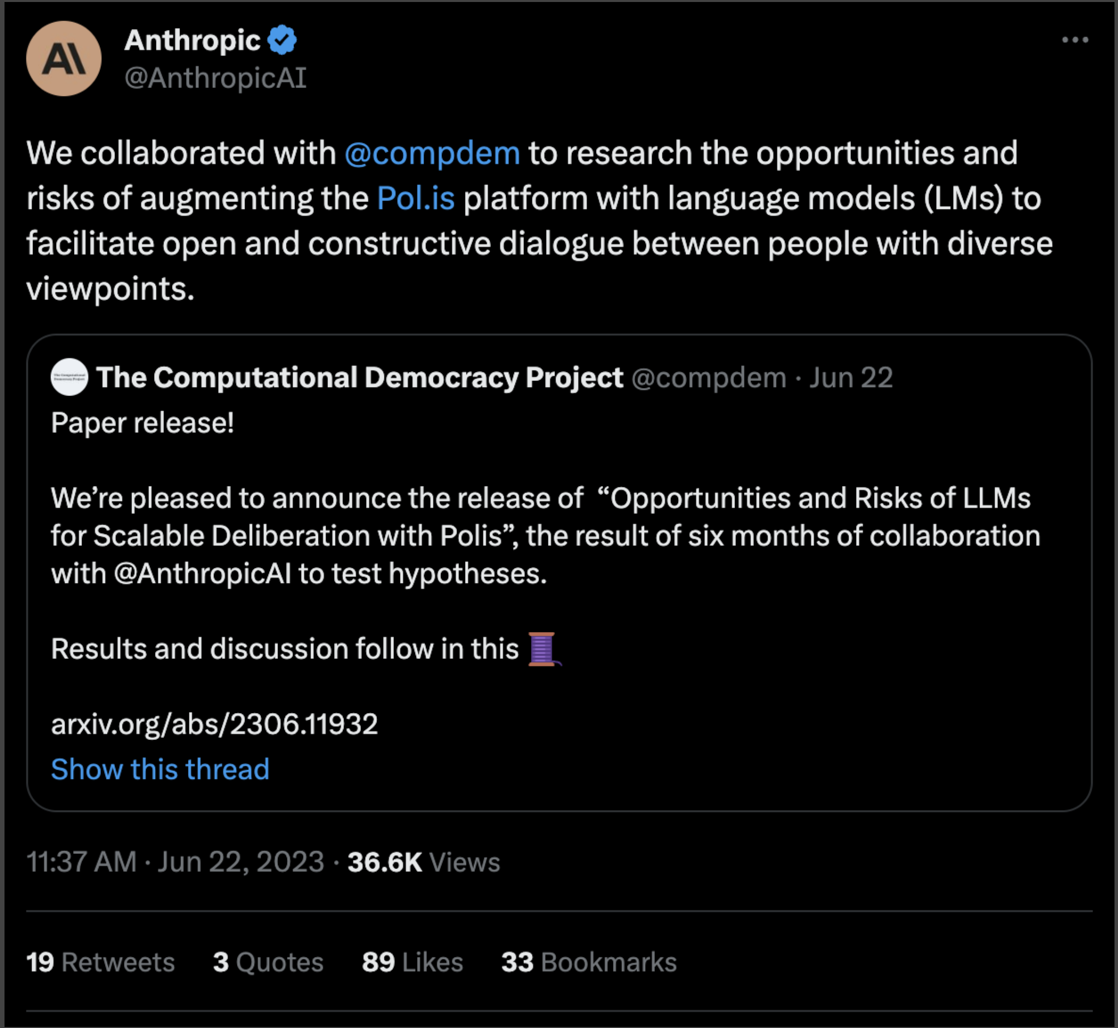
Interpretability of Polis results has been, and continues to be, a critical issue for Polis as a platform, and a challenge to the usage of the method by various stakeholders and user archetypes. Interpretability is a hard and foundational problem that will require ongoing development: as more advanced analytic methods are added to the system, interpretability by those without data science and statistical methods backgrounds will suffer, and the burden on producing interpretable results on those running conversation will increase.

In a sense, the summary the platform has is the visualization: "here are the groups, what differentiates them, and what unifies them". But the visualization is necessarily limited to a handful of comments, and how these comments are chosen is opaque. We have always thought about interpretability in ‘tiers’ (a simple list, clustering, etc), from a list, to the visualization, to the report, to human generated reports, to news articles summarizing, etc., with various levels of human in the loop. Consider the following examples.

## Biodiversity in NZ

1. The comments chosen procedurally for the visualization for biodiversity: https://www.scoop.co.nz/stories/HL1908/S00014/scoop-hivemind-protecting-and-restoring-biodiversity.htm)
2. The procedurally generated report https://pol.is/report/r3epuappndxdy7dwtvwpb
3. And what PEP ultimately delivered to the government (PDF): https://www.scoop.co.nz/stories/PO1911/S00063/biodiversity-hivemind-report-plenty-of-common-ground.htm
4. Direct link to PDF above: https://img.scoop.co.nz/media/pdfs/1911/Biodiversity_HiveMind_Final_Report_Scoop.pdf
5. The debrief https://pep.org.nz/2020/12/01/doc-tries-to-restore-e-democracy/

## Bowling Green Civic Assembly



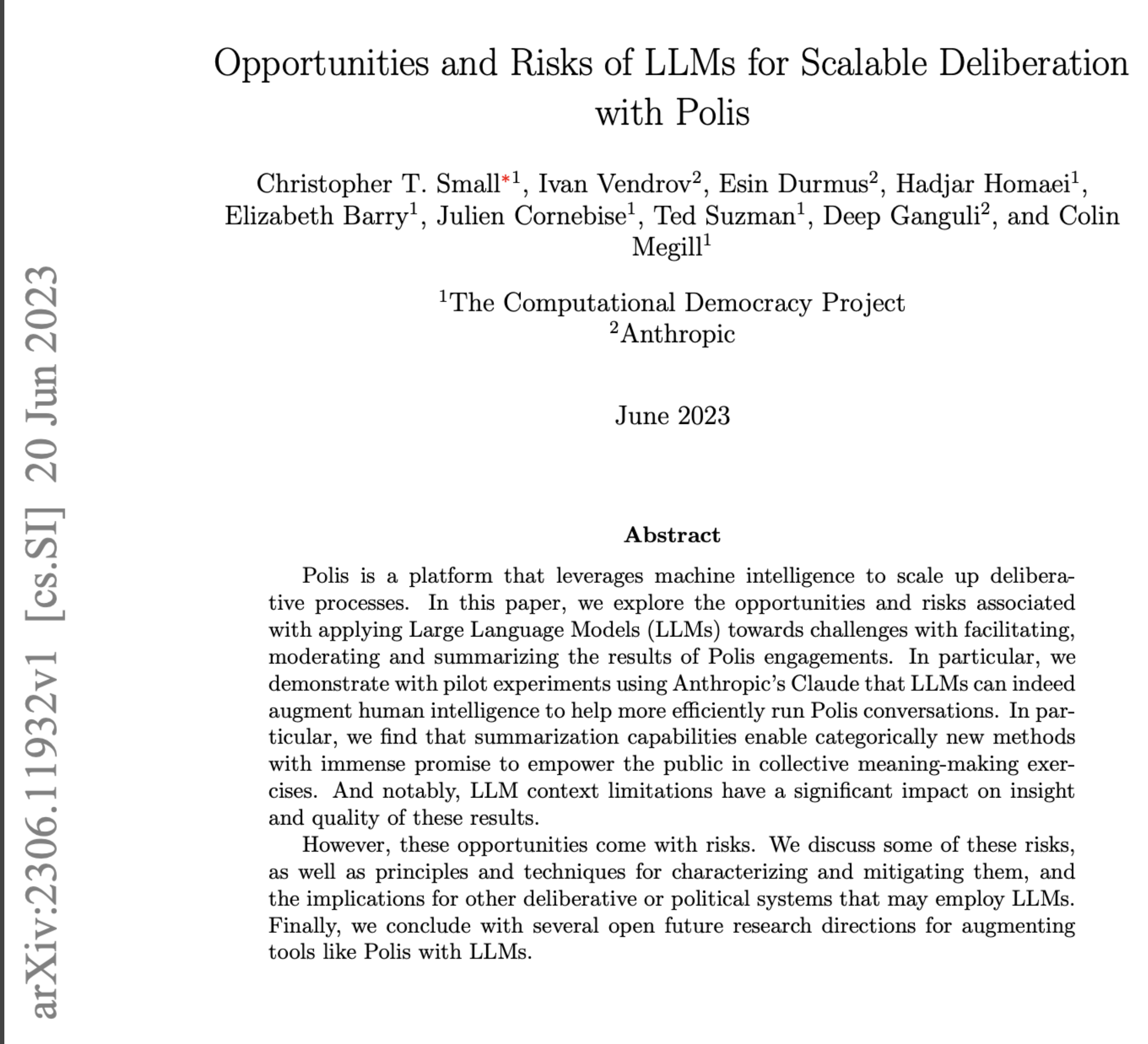
**Anthropic**
@AnthropicAI

We collaborated with @compdem to research the opportunities and risks of augmenting the Pol.is platform with language models (LMs) to facilitate open and constructive dialogue between people with diverse viewpoints.

> **The Computational Democracy Project** @compdem · Jun 22
>
> Paper release!
>
> We're pleased to announce the release of "Opportunities and Risks of LLMs for Scalable Deliberation with Polis", the result of six months of collaboration with @AnthropicAI to test hypotheses.
>
> Results and discussion follow in this 🧵
>
> arxiv.org/abs/2306.11932
>
> Show this thread

11:37 AM · Jun 22, 2023 · **36.6K** Views

**19** Retweets **3** Quotes **89** Likes **33** Bookmarks



arXiv:2306.11932v1 [cs.SI] 20 Jun 2023

# Opportunities and Risks of LLMs for Scalable Deliberation with Polis

Christopher T. Small*[1], Ivan Vendrov[2], Esin Durmus[2], Hadjar Homaei[1], Elizabeth Barry[1], Julien Cornebise[1], Ted Suzman[1], Deep Ganguli[2], and Colin Megill[1]

[1]The Computational Democracy Project
[2]Anthropic

June 2023

## Abstract

Polis is a platform that leverages machine intelligence to scale up deliberative processes. In this paper, we explore the opportunities and risks associated with applying Large Language Models (LLMs) towards challenges with facilitating, moderating and summarizing the results of Polis engagements. In particular, we demonstrate with pilot experiments using Anthropic's Claude that LLMs can indeed augment human intelligence to help more efficiently run Polis conversations. In particular, we find that summarization capabilities enable categorically new methods with immense promise to empower the public in collective meaning-making exercises. And notably, LLM context limitations have a significant impact on insight and quality of these results.

However, these opportunities come with risks. We discuss some of these risks, as well as principles and techniques for characterizing and mitigating them, and the implications for other deliberative or political systems that may employ LLMs. Finally, we conclude with several open future research directions for augmenting tools like Polis with LLMs.

TOPIC ANALYSIS

**SUMMARIZATION**

VOTE PREDICTION

CONSENSUS PREDICTION

SIMULATION

COMMENT ROUTING

For summarization without explicit topic modelling as an intermediary step in the prompt:

**Prompt:**

Human: Given the comments and agreement/disagreement votes for each comment from two groups (Group 0 and Group 1), summarize the key points of consensus and divergence between Group 0 and Group 1. Identify what opinions the two groups have consensus on versus where their views differ.

Here are the comments: Comment: {comment-text} == Group 0 percentage agreement: {g0-agree-percent}% == Group 0 percentage disagreement: {g0-disagree-percent}% == Group 1 percentage agreement: {g1-agree-percent}% == Group 1 percentage disagreement: {g1-disagree-percent}%
....... [MORE COMMENTS] .......

Assistant: Sure, I can do that.

Consensus between Group 0 and Group 1:

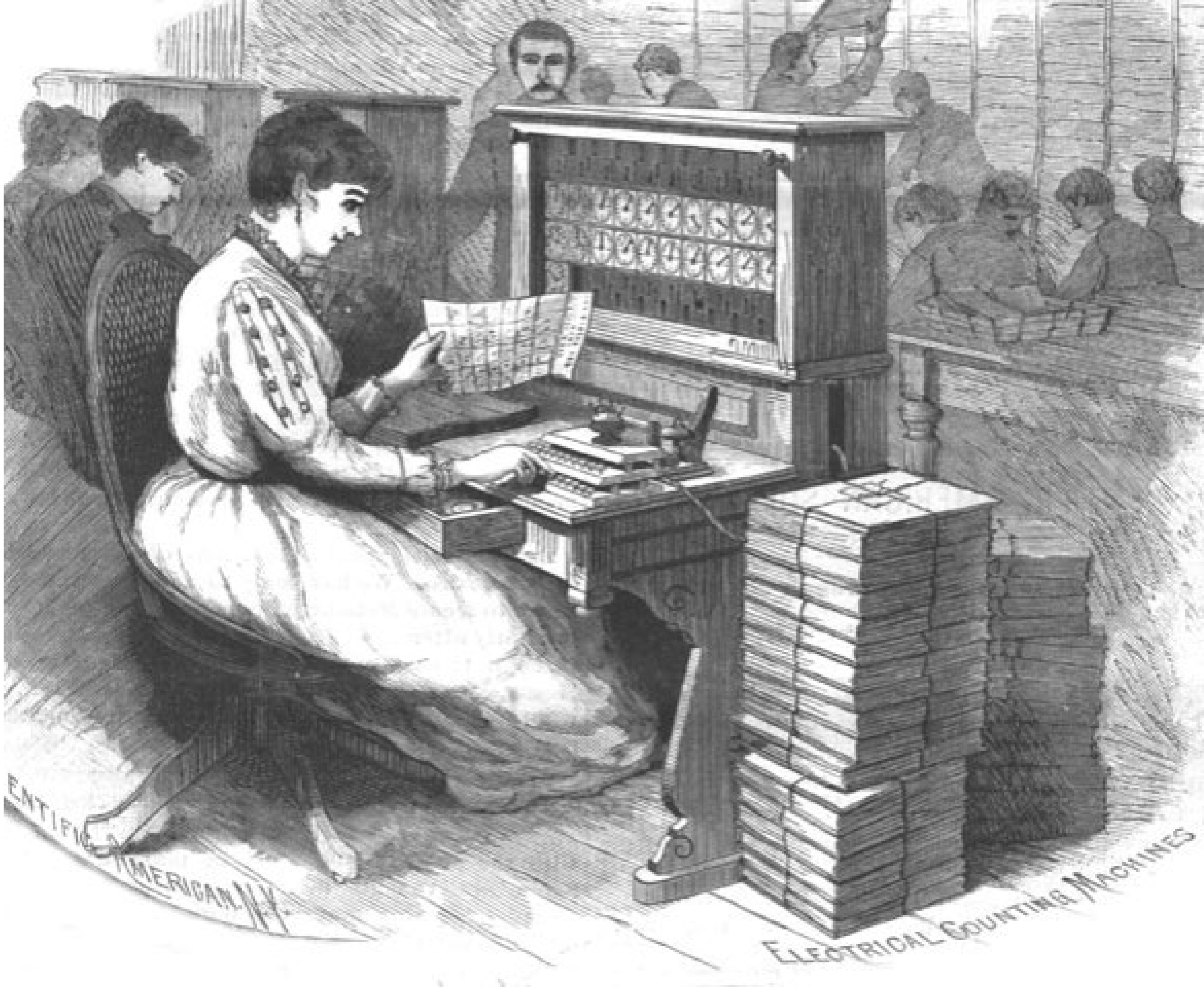
- Most comments that received strong agreement in both groups focused on investing in the community good, improving infrastructure and public services, supporting the local economy, and protecting the environment. For example, comments advocating for more bike lanes, sidewalks, public transit; investing in the arts and education; developing the downtown and riverfront area; increasing green spaces; improving traffic flow and infrastructure; and supporting local businesses were agreed upon by both groups.

Divergence between Group 0 and Group 1:

- There were some notable differences of opinion on social and political issues. Group 0 tended to be more progressive while Group 1 tended to be more conservative. For example, Group 0 largely agreed with and Group 1 largely disagreed with comments supporting LGBTQ rights, marijuana legalization, and sanctuary cities. Group 0 disagreed and Group 1 agreed with restricting immigration and requiring immigrants to learn English.
- Group 0 was more supportive of government intervention and social programs to address issues like poverty, homelessness, healthcare, and education. Group 1 was more skeptical of government overreach and supportive of free market principles. For example, Group 0 agreed and Group 1 disagreed with raising the minimum wage, providing affordable housing and healthcare options, and increasing funding for schools and social programs.
- There were also differences in levels of trust in institutions like the media, government, and universities. Group 0 expressed more trust in these institutions while Group 1 was more skeptical and critical. For example, Group 0 agreed and Group 1 disagreed that most journalists and public officials aim to serve the public good. Group 0 disagreed and Group 1 agreed that university mismanagement is mostly responsible for increasing tuition.

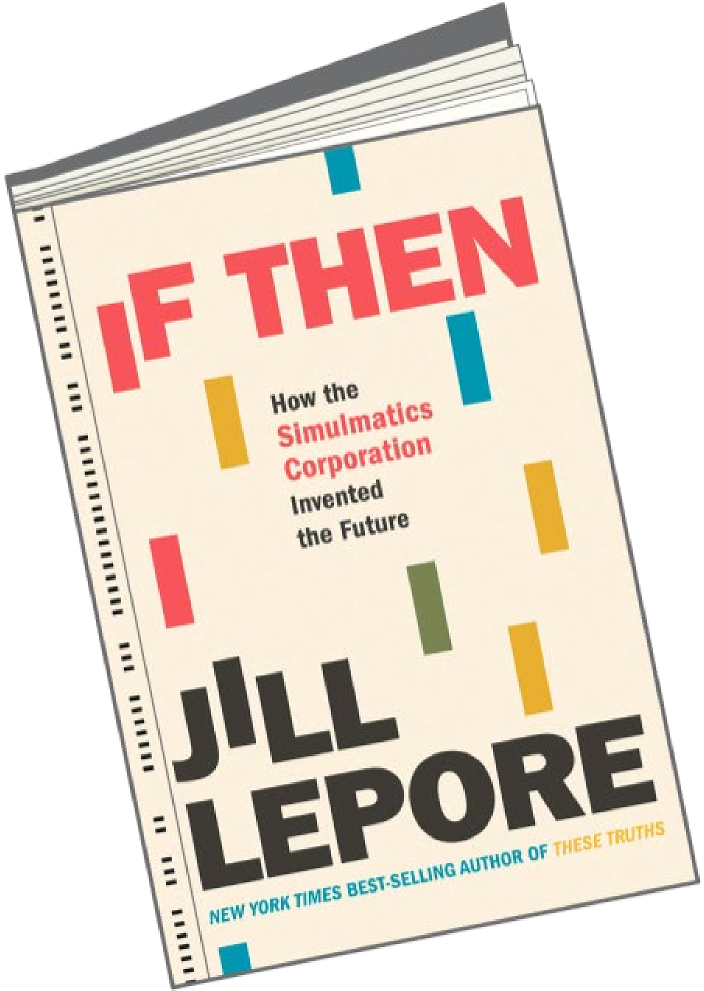
In summary, while there were areas of agreement on community investment and development, the groups differed in their social and political values as well as levels of trust in public institutions. But there is clearly some common ground to build upon, if the goal is to find consensus and work together.

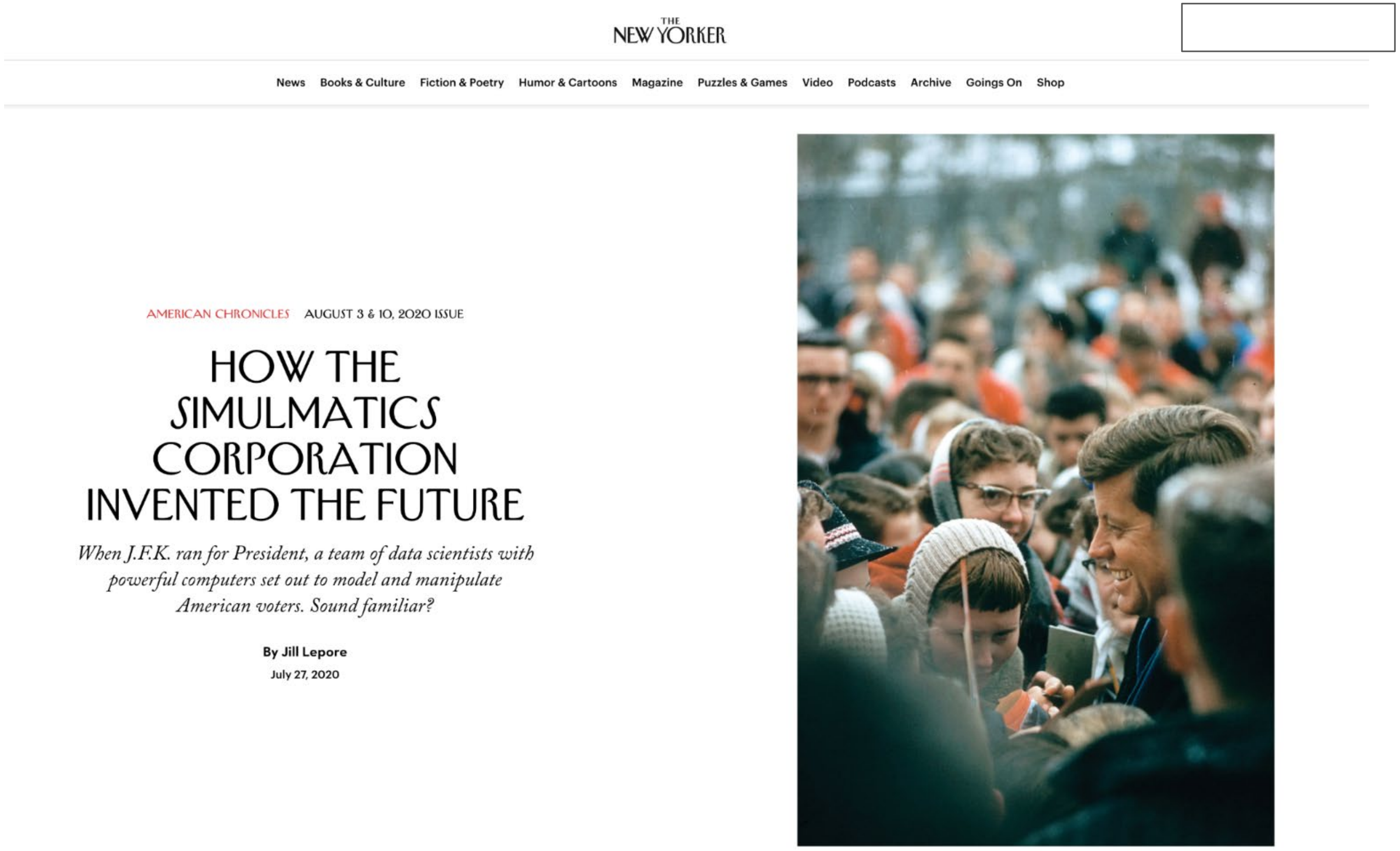
# **HOLLERITH** ELECTRIC COUNTING MACHINE



# UNIVAC 1

(CENSUS)


IF THEN
How the Simulmatics Corporation Invented the Future
JILL LEPORE
NEW YORK TIMES BEST-SELLING AUTHOR OF THESE TRUTHS



THE NEW YORKER

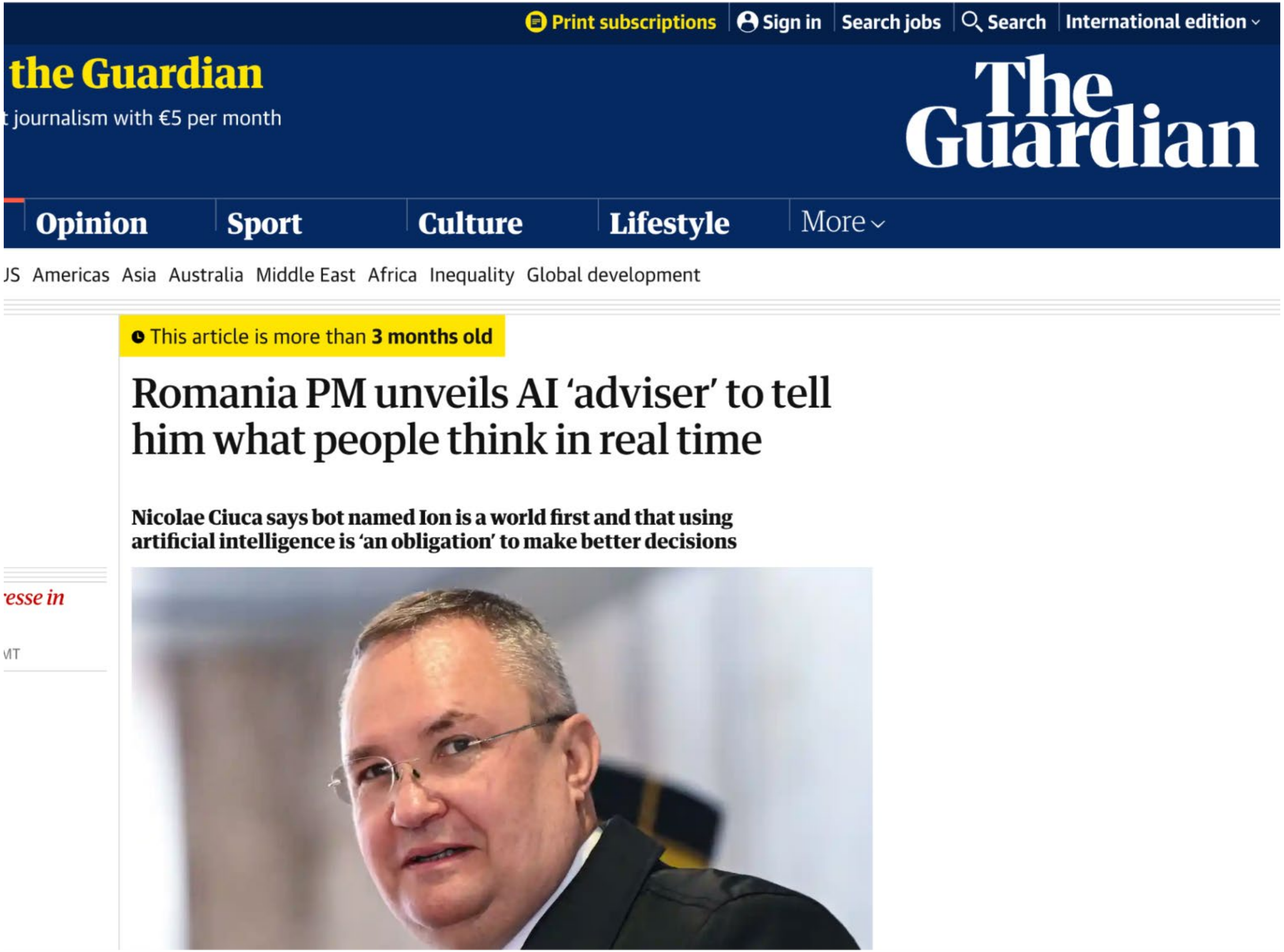
AMERICAN CHRONICLES AUGUST 3 & 10, 2020 ISSUE

# HOW THE SIMULMATICS CORPORATION INVENTED THE FUTURE

*When J.F.K. ran for President, a team of data scientists with powerful computers set out to model and manipulate American voters. Sound familiar?*

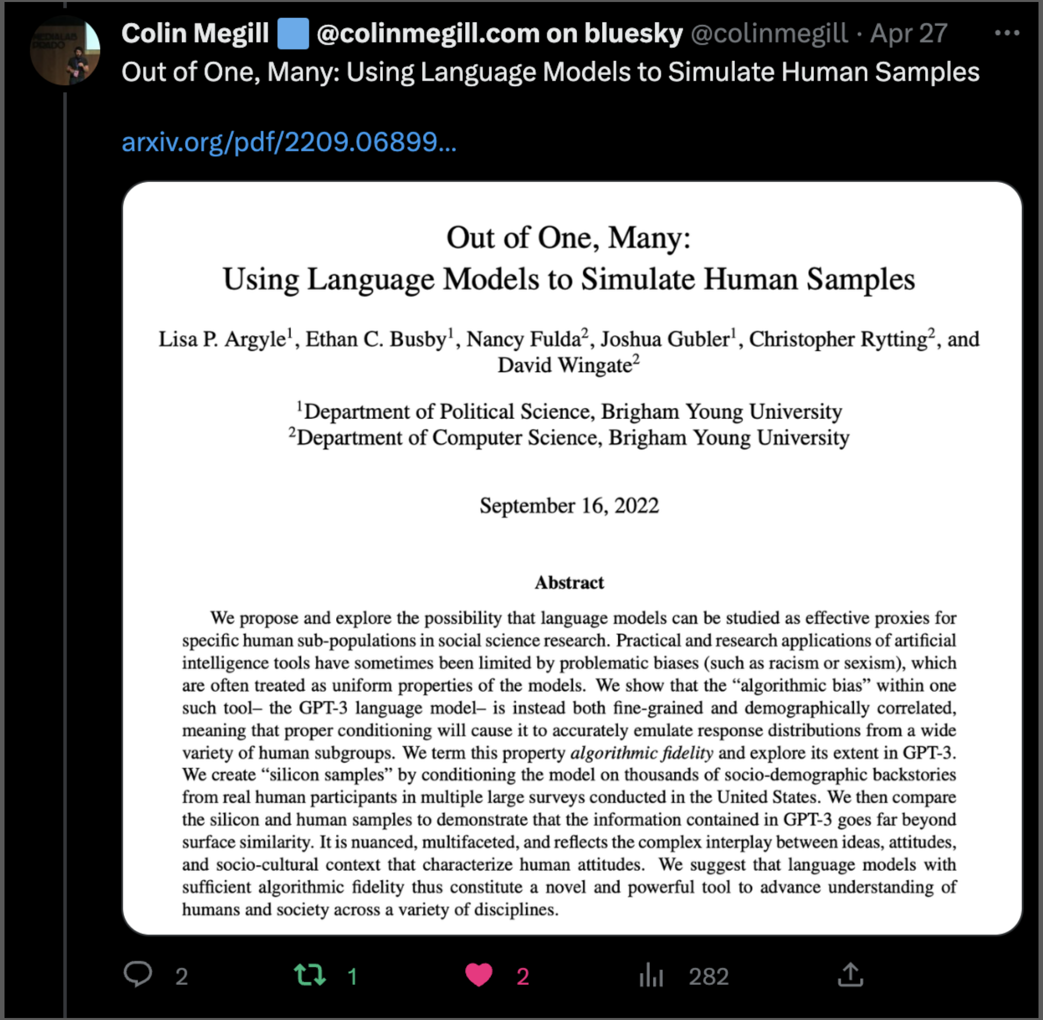
**By Jill Lepore**

July 27, 2020



# Romania PM unveils AI 'adviser' to tell him what people think in real time

**Nicolae Ciuca says bot named Ion is a world first and that using artificial intelligence is 'an obligation' to make better decisions**



**Colin Megill** 🟦 **@colinmegill.com on bluesky** @colinmegill · Apr 27

Out of One, Many: Using Language Models to Simulate Human Samples

arxiv.org/pdf/2209.06899...

# Out of One, Many: Using Language Models to Simulate Human Samples

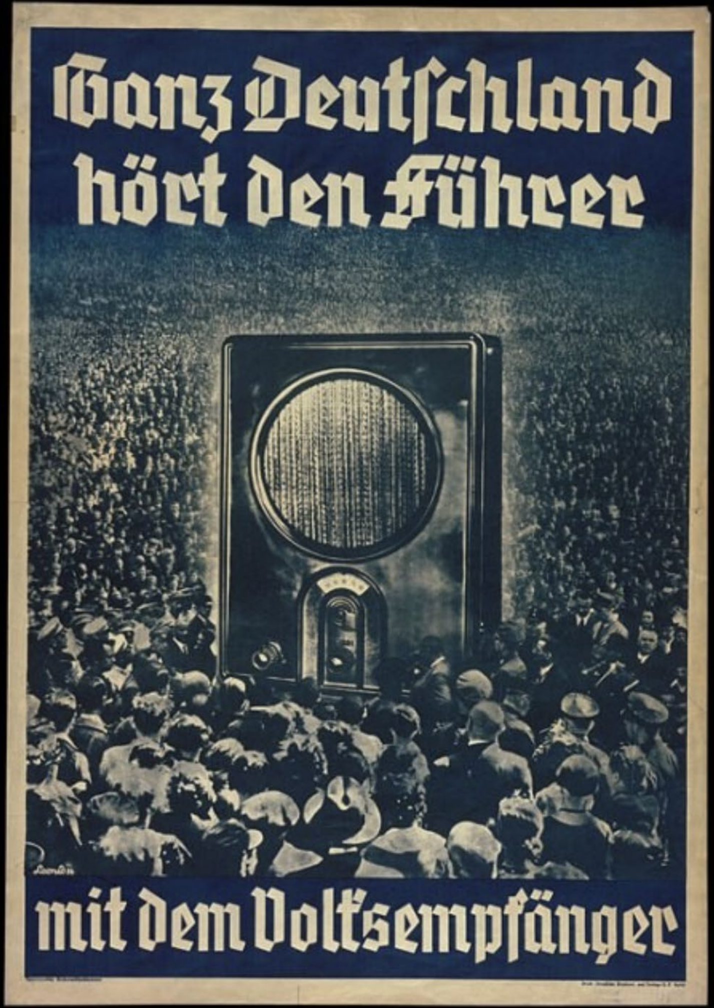
Lisa P. Argyle[1], Ethan C. Busby[1], Nancy Fulda[2], Joshua Gubler[1], Christopher Rytting[2], and David Wingate[2]

[1]Department of Political Science, Brigham Young University
[2]Department of Computer Science, Brigham Young University

September 16, 2022

**Abstract**

We propose and explore the possibility that language models can be studied as effective proxies for specific human sub-populations in social science research. Practical and research applications of artificial intelligence tools have sometimes been limited by problematic biases (such as racism or sexism), which are often treated as uniform properties of the models. We show that the "algorithmic bias" within one such tool– the GPT-3 language model– is instead both fine-grained and demographically correlated, meaning that proper conditioning will cause it to accurately emulate response distributions from a wide variety of human subgroups. We term this property *algorithmic fidelity* and explore its extent in GPT-3. We create "silicon samples" by conditioning the model on thousands of socio-demographic backstories from real human participants in multiple large surveys conducted in the United States. We then compare the silicon and human samples to demonstrate that the information contained in GPT-3 goes far beyond surface similarity. It is nuanced, multifaceted, and reflects the complex interplay between ideas, attitudes, and socio-cultural context that characterize human attitudes. We suggest that language models with sufficient algorithmic fidelity thus constitute a novel and powerful tool to advance understanding of humans and society across a variety of disciplines.


Ganz Deutſchland hört den Führer
Leonid
mit dem Volksempfänger

How does the public relate to the **models made of the public?**

Building new systems to **durably return political agency and agenda setting power** to public

**@CompDem, 501c3**

# @CompDem, 501c3

Maintain code
Organize methods & training
Guide implementation
Advance research

This is a generational opportunity to transform policymaking worldwide

@colinmegill
**@compdem**
@usepolis

[https://pol.is](https://pol.is)

[compdemocracy.org](https://compdemocracy.org)