

# Social Impact of Artificial Intelligence: research challenges

Virgilio Almeida

---

virgilio@dcc.ufmg.br

@virgilioalmeida

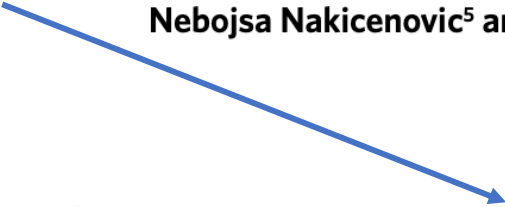
23/08/2022

Escola FAPESP 60 anos



# Six Transformations to achieve the Sustainable Development Goals

Jeffrey D. Sachs<sup>1</sup>, Guido Schmidt-Traub <sup>2\*</sup>, Mariana Mazzucato<sup>3</sup>, Dirk Messner<sup>4</sup>,  
Nebojsa Nakicenovic<sup>5</sup> and Johan Rockström<sup>6</sup>



**Transformation 6.** Artificial intelligence and other digital technologies—sometimes referred to as the Fourth Industrial Revolution— are disrupting nearly every sector of the economy, including agriculture (precision agriculture), mining (autonomous vehicles), manufacturing (robotics), retail (e-commerce), finance (e-payments and trading strategies), media (social networks), health (diagnostics and telemedicine), education (online learning), public administration (e-governance and e-voting) and science and technology.

# *A.I. Is Not Sentient. Why Do People Say It Is?*

Robots can't think or feel, despite what the researchers who build them want to believe.





[@praxisandpeople](https://twitter.com/praxisandpeople)

! There is no 'outside' of technology; but there sure as hell are better ways of organizing how we build our future within in.

Resident-led, equity-focused, deeply inclusive, anti-colonial, community-empowering, and sustainable

# Facebook Approved Pro-Genocide Ads in Kenya After Claiming to Foster 'Safe and Secure' Elections

Kenya's national cohesion watchdog has threatened to suspend the social network from the country in a week if it doesn't mitigate hate

By Dell Cameron | Friday 2:03PM | Comments (15) | Alerts



Meta CEO Mark Zuckerberg speaks at Georgetown University in Washington, Thursday, Oct. 17, 2019.



How to act?



Science

+ FOLLOW

## China's court AI reaches every corner of justice system, advising judges and streamlining punishment

- Smart court's electronic reach allows the system to access police, prosecutor and government databases and integrate with China's social credit system
- Chinese law professor warns, 'We must be alert to the erosion of judicial power by technology companies and capital'



Stephen Chen in Beijing

+ FOLLOW

Published: 8:30pm, 13 Jul, 2022

Why you can trust SCMP

### TOP PICKS

Private-market investment solutions for HNW individuals

Paid Post: HSBC



### News

Record-breaking Chinese bridge clears hurdle to faster travel in border province

18 Jul 2022



### News

China's famous

512



Zhang Linghan, professor of law at the China University of Political Science and Law in Beijing, warned that the **rapid rise of AI could create a world where humans were ruled by machines.**

Its electronic reach allows the smart court to work with China's powerful "social credit system" to ban a person refusing to pay a debt from using a plane, high speed train, hotels or other social services.



HOT TOPICS

UKRAINE CRISIS

CORONAVIRUS IN EUROPE

FRENCH ELECTION 2022

LIVING CITIES

# Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.

BY [MELISSA HEIKKILÄ](#)



## Most Read Articles

### 1. Sweden wins the NATO lottery

APRIL 13, 2022 | 4:04 AM

### 2. Ukraine announces capture of top Putin pal Viktor Medvedchuk

APRIL 13, 2022 | 5:56 AM

### 3. How Germany Inc. played Russian roulette – and lost

APRIL 13, 2022 | 4:00 AM

### 4. Ukraine rejects visit by German President Steinmeier

# Predictive Policing through AI

## *An Algorithm That Grants Freedom, or Takes It Away*

Across the United States and Europe, software is making probation decisions and predicting whether teens will commit crime. Opponents want more human oversight.

Virgilio Almeida, 2020







## Brazilian Favelas need racial Justice, not Facial Recognition

By

### RELATED POSTS

#### Latin America: Reinforcing democratic institutions to achieve peace, justice and inclusion

This policy brief calls for a rethink of the social contract in Latin America to tackle the redistribution of power...

#### COVID-19 and the Crisis of Governance: The impact of the Pandemic on Peace, Justice and Inclusion (SDG 16)

As countries adopt COVID-19 response-and-recovery plans, there is an opportunity to achieve more peaceful and just societies with more inclusive...

#### SDGs as Citizens' Agenda: Stress Peace, Justice and Strong Institutions

The Sustainable Development Goals (SDGs) for the 2030 Agenda for Sustainable Development are not only an agenda belonging to either...

Yet, in Brazil and other contexts marked by extreme racial, ethnic and gender inequalities, such technologies can backfire.

Historically marginalized groups are more exposed to mistakes made by such mechanisms.

Additionally, these technologies can be used as a tool for community control – increasing violence, hypervigilance by state actors and judicial bias in criminal proceedings.



the facebook files

# Facebook Services Are Used to Spread Religious Hatred in India, Internal Documents Show

Company researchers say anti-Muslim material is rife and calls to violence coincided with deadly riots last year. Facebook says it's working to improve enforcement.



EXTREMISM ONLINE

# The Human Infrastructure of Fake News in Brazil

*The role of algorithms in promoting disinformation has received a great deal of attention in recent years, due in large part to the centrality of Facebook in the 2016 US presidential election and the UK Brexit campaign. However, David Nemer argues that in countries such as Brazil, where peer-to-peer messaging apps like WhatsApp are popular, more attention needs to be paid to the "human infrastructure" of coordinated disinformation campaigns.*

by David Nemer  
20 hours ago



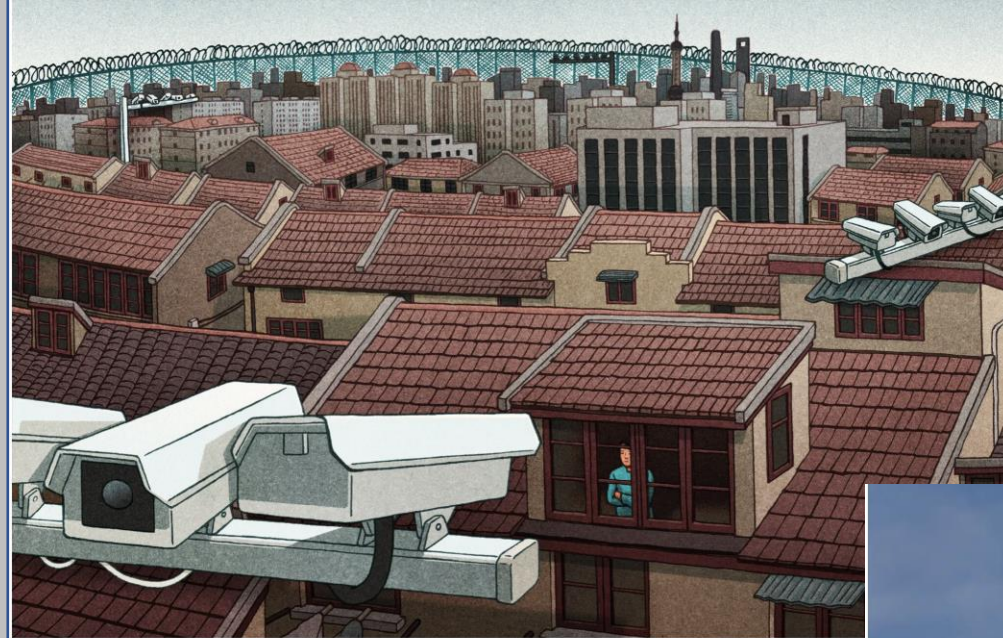
## *Has Shanghai Been Xinjianged?*

Shanghai used to be the glamorous China, while Xinjiang was the dark China. Now both are casualties of authoritarian excess.

By [Li Yuan](#) – New York Times

Published May 6, 2022

He fears an even more dystopian China than what it is today: a digital totalitarian regime that surveils everyone, makes each neighborhood an on-site concentration camp and controls the society with the same iron fist in a future crisis, be it war, famine, climate disaster or economic meltdown.



Xinmei Liu



# Surveillance Society

The stage, which we are living through now, a chaos caused by the profit-driven algorithmic amplification, dissemination and microtargeting of corrupt information, much of it produced by coordinated schemes of disinformation.

Its effects are felt in the real world, where they splinter shared reality, poison social discourse, paralyze democratic politics and sometimes instigate violence and death.

Opinion

## The Coup We Are Not Talking About

We can have democracy, or we can have a surveillance society, but we cannot have both.

By Shoshana Zuboff

Dr. Zuboff, a professor emeritus at Harvard Business School, is the author of "The Age of Surveillance Capitalism."

Jan. 29, 2021



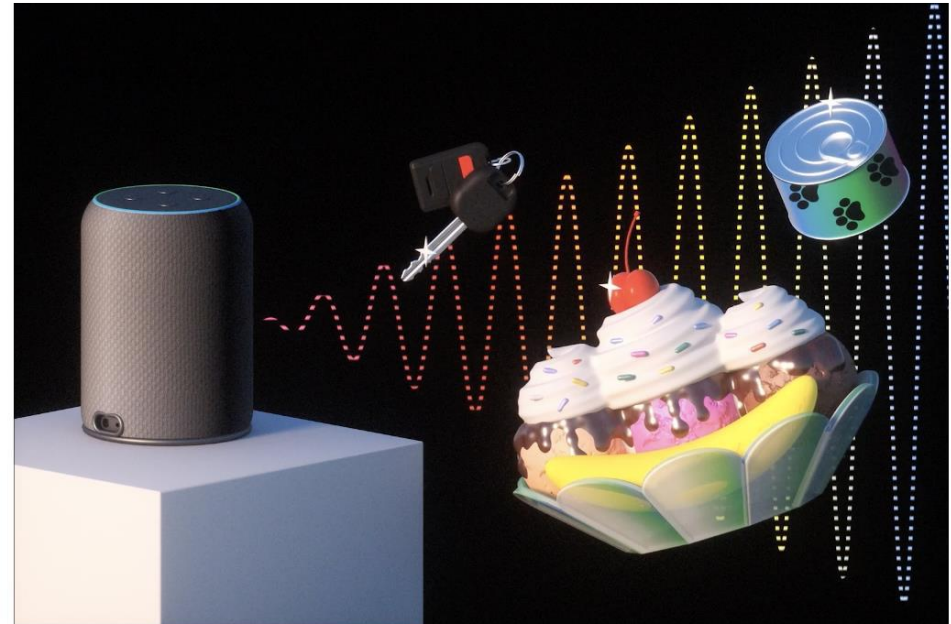
## Questions about Ethics

- Should companies be barred from analyzing what we say and how we sound to recommend products or personalize advertising messages.
- Should companies be allowed the profiling of people based on their voices?
- Voice technologies, including Alexa and Siri from Apple, will morph from digital butlers into diviners that use the sound of our voices to work out intimate details like our moods, desires and medical conditions.
- Technologies are creating another set of data that people have no clue how it's being used.

The New York Times

## *Should Alexa Read Our Moods?*

The time to debate the power of voice technology is now.  
(Before Alexa starts suggesting comfort foods.)



Maria Chimishkyan

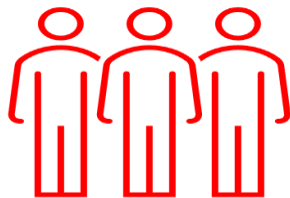
# Sociotechnical systems



# DECIPHER ME OR I'LL DEVOUR YOU: The Sphinx Enigma

- Researchers
- Journalists
- Academics
- Policymakers
- Legislators

....



**SOCIOTECHNICAL systems:**  
social media platforms

- Powerful complex and opaque sociotechnical systems
- algorithms and data

The risk of **private ownership** of the **public square** is that one person's views could end up privileged over all others.



What is the object to be studied?

## Algorithms or Algorithmic Systems?

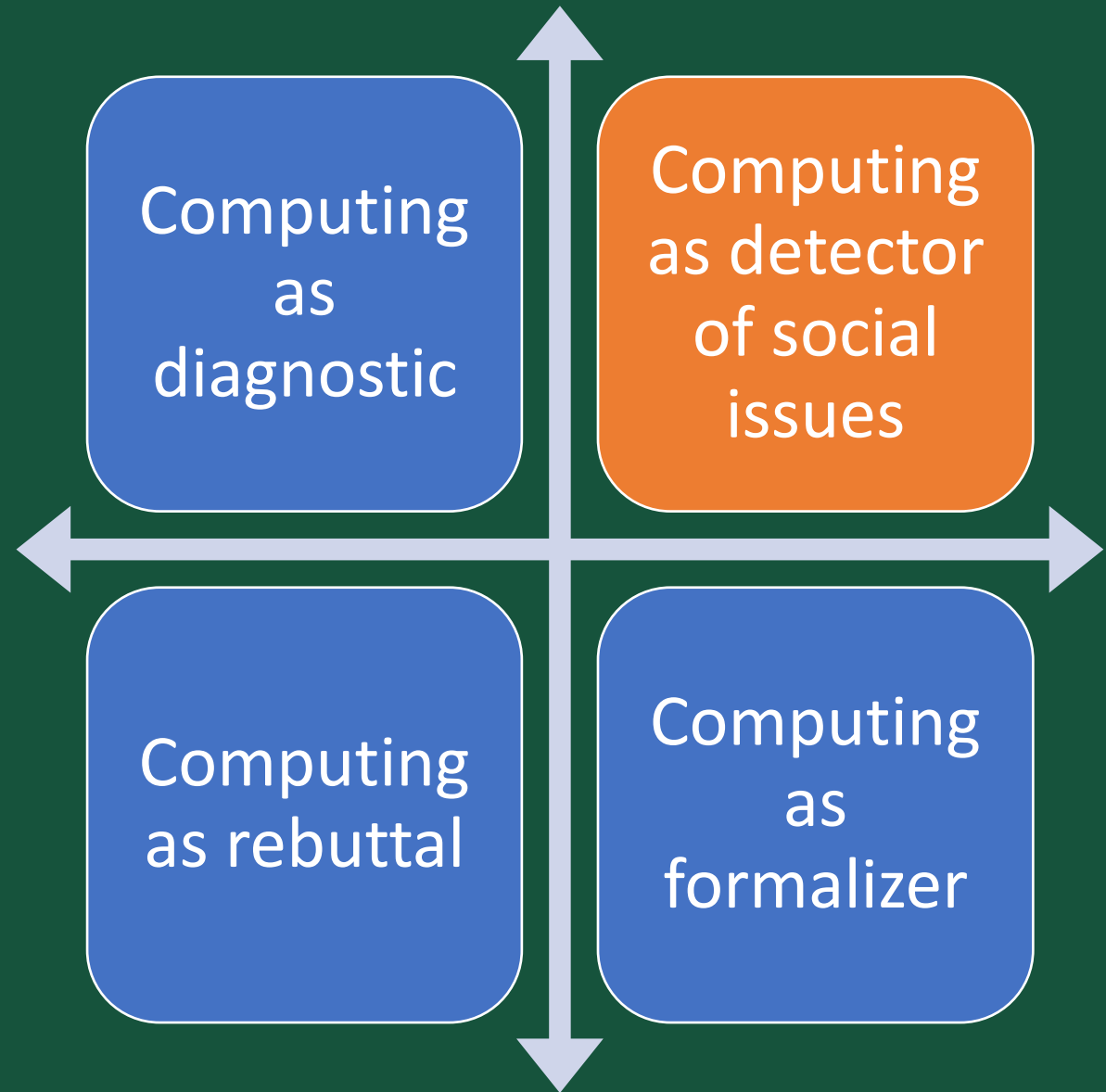
-- Algorithmic systems are not standalone little boxes, but massive, networked ones with hundreds of hands reaching into them, tweaking and tuning, swapping out parts and experimenting with new arrangements.”

-- Algorithmic systems include many distinct algorithms, models, mathematical functions, training data, applications, hardware, business goals, humans and networks.

-- They are sociotechnical ensembles, combining technical and social components, where social actors and technological artifacts complexly become intertwined.

New roles for computing  
research  
in times of  
complex and powerful  
sociotechnical systems

New roles for  
computer  
science in these  
times of  
sociotechnical  
systems.



# Computing as a detector of social issues

Its ad-delivery system is excluding women from opportunities without regard to their qualifications. That would be illegal under US employment law.



MS TECH | PEXELS

Tech policy / AI Ethics

## Facebook's ad algorithms are still excluding women from seeing jobs

Its ad-delivery system is excluding women from opportunities without regard to their qualifications. That would be illegal under US employment law.

[Auditing for Discrimination in Algorithms Delivering Job Ads](#), [Basileal Imana](#), [Aleksandra Korolova](#), and [John Heidemann](#), In Proceedings of [The Web Conference 2021 \(WWW '21\)](#), April 19–23, 2021, Ljubljana, Slovenia.

Computing as diagnostic



Racial bias in cost data leads an algorithm to underestimate health care needs of Black patients.

SOCIAL SCIENCE

# Assessing risk, automating racism

A health care algorithm reflects underlying racial bias in society

By Ruha Benjamin

As more organizations and industries...

era, the intention to deepen racial inequities was more explicit, today coded inequity is perpetuated precisely because those

beyond the algorithm developers by constructing a more fine-grained measure of health outcomes, by extracting and cleaning...

# 'We don't deserve this': new app places US caregivers under digital surveillance

Tech companies and lawmakers promise that EVV will increase efficiency and accountability in home care and will reduce fraud, waste and abuse in government-funded programs.

Supported by

OPEN SOCIETY  
FOUNDATIONS

About this content  
**Virginia Eubanks and  
Alexandra Mateescu**

Wed 28 Jul 2021 11.00 BST



Computing  
as rebuttal



▲ Melissa Harville checks the electronic visit verification app for her partner Kevin Hooper outside their home in Greenbrier, Arkansas. Photograph: Rory Doyle/The Guardian

Electronic visit verification systems make homecare more difficult, trap people with disabilities, and cost more than paper timesheets. So why are they rolling out across the country?

Advertisement  
Subscribe to The

# Peeking Beneath the Hood of Uber

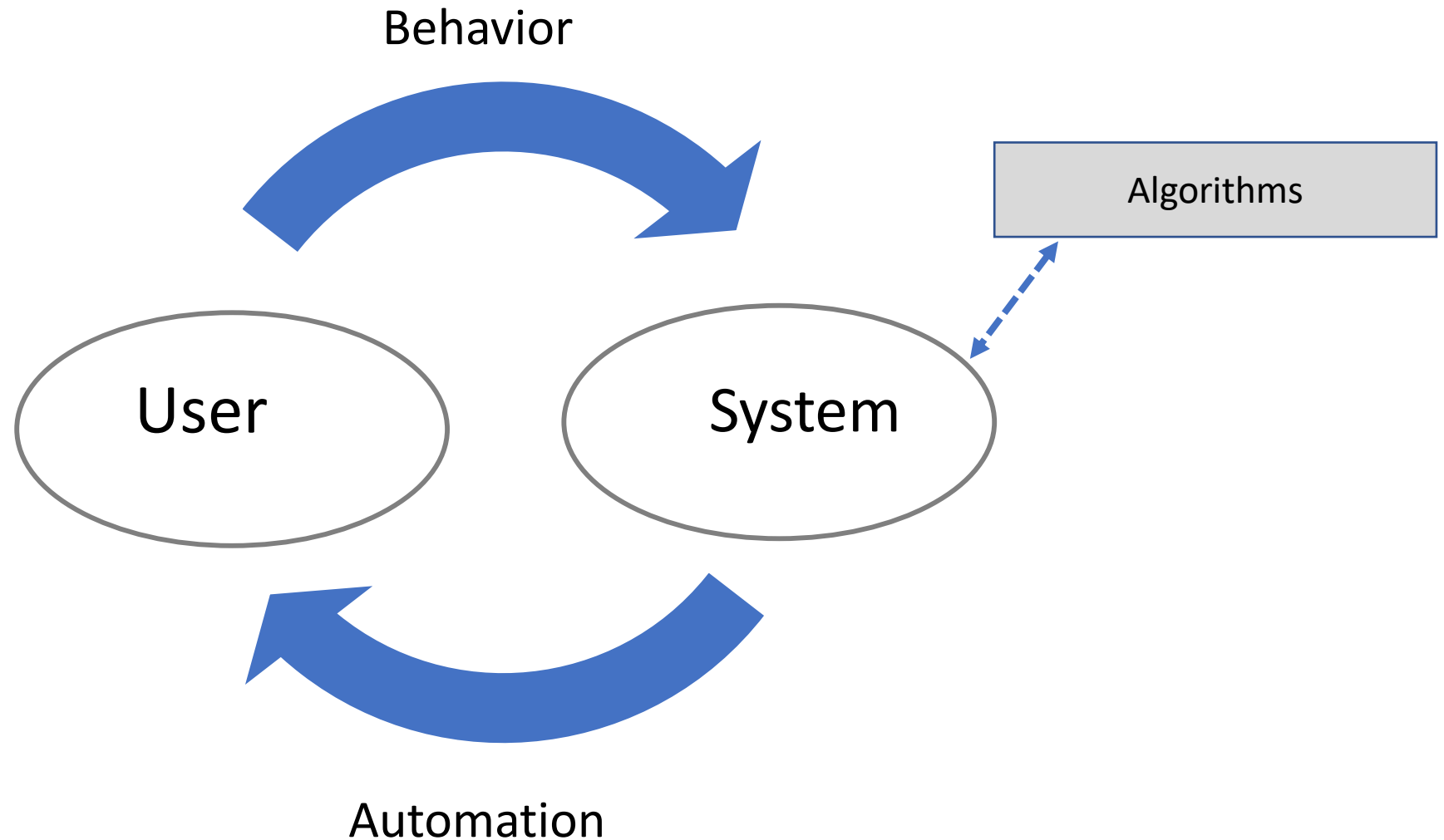
- First systematic audit of Uber's surge pricing algorithm
  - Prices update every 5 minutes
    - 70% of surges last  $\leq$  10 minutes
  - Uber divides cities into surge areas
    - May save significant money by walking into an adjacent area
  - Supply and demand do correlate with surge pricing



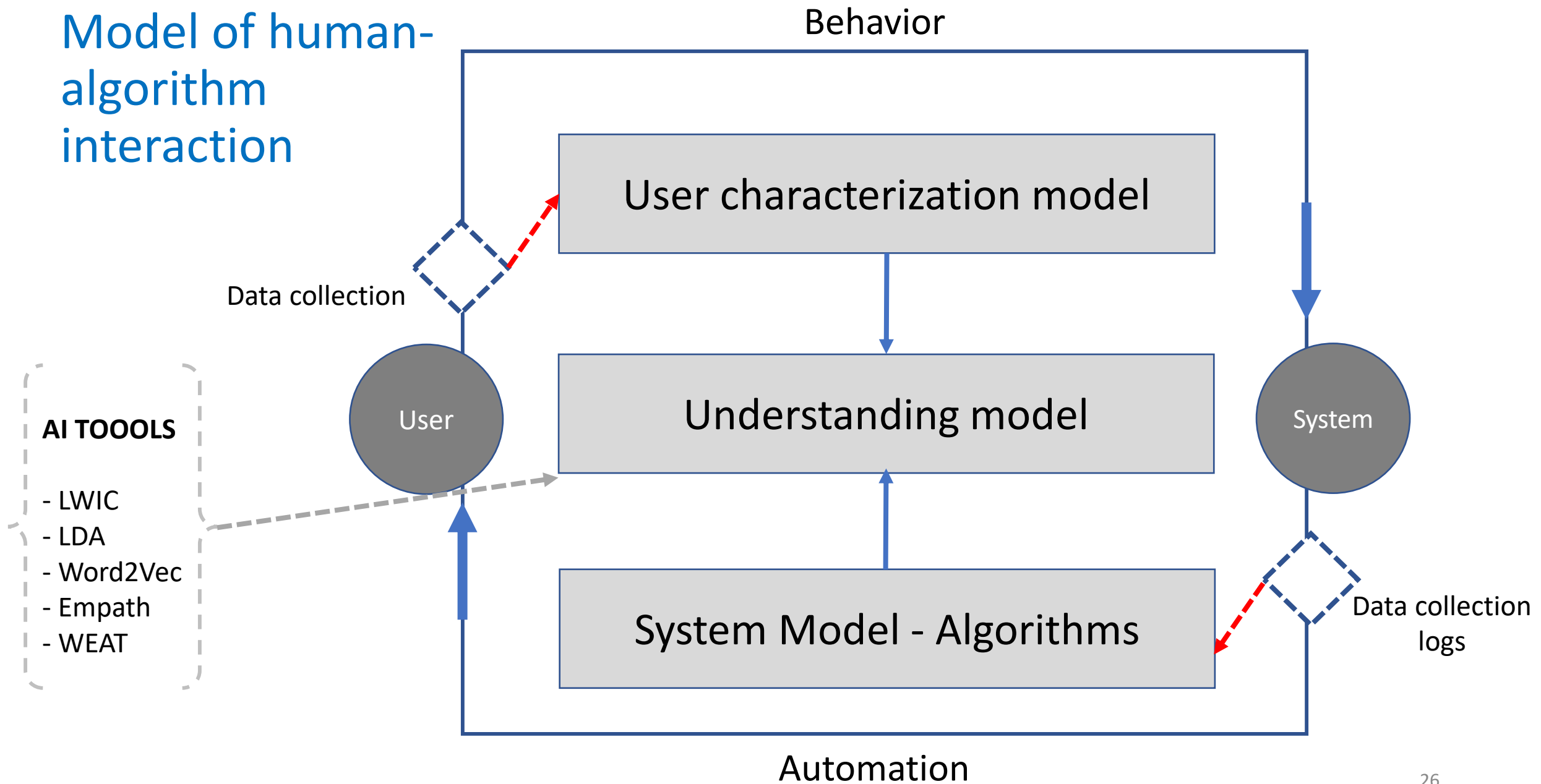
# Experimental Research



# Simple Model of Human-Algorithm Interaction

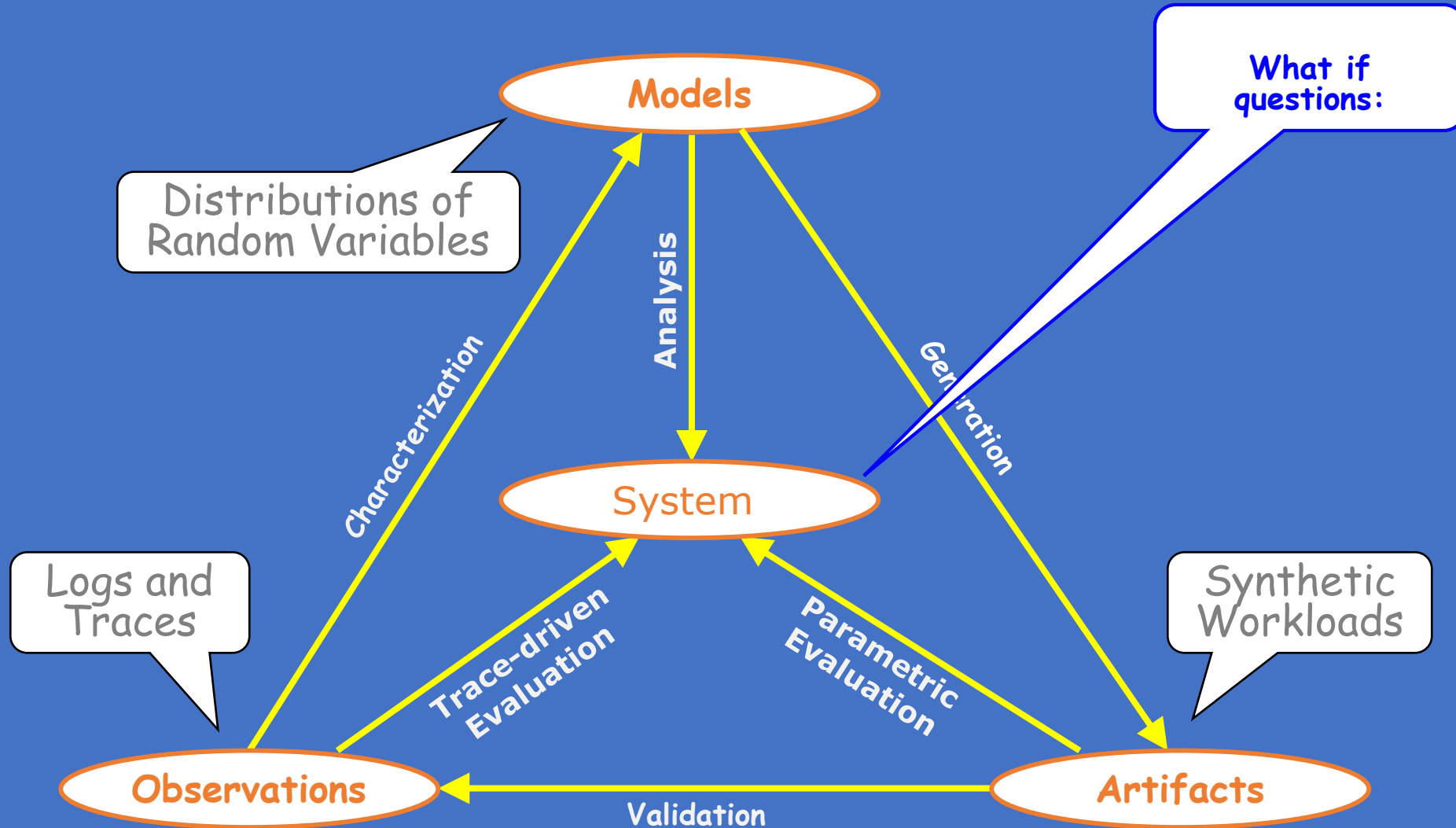


# Model of human-algorithm interaction



# System Model

Measure → Analyze → Model → Synthesize → Computational Models

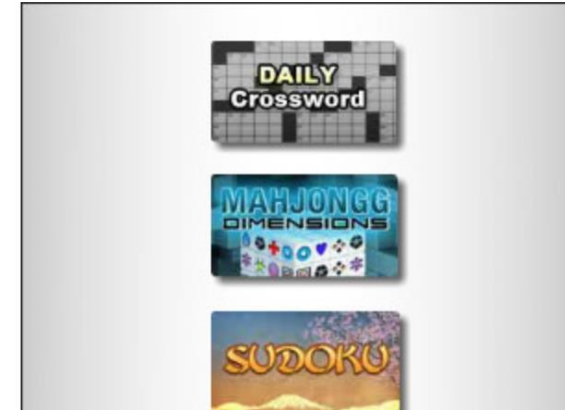
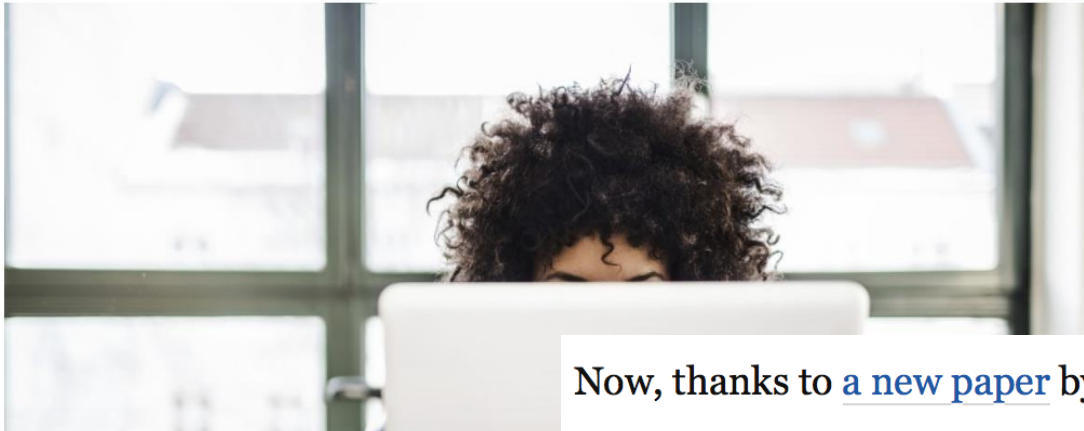




The Intersect

# Study: Image results for the Google search ‘ugly woman’ are disproportionately black

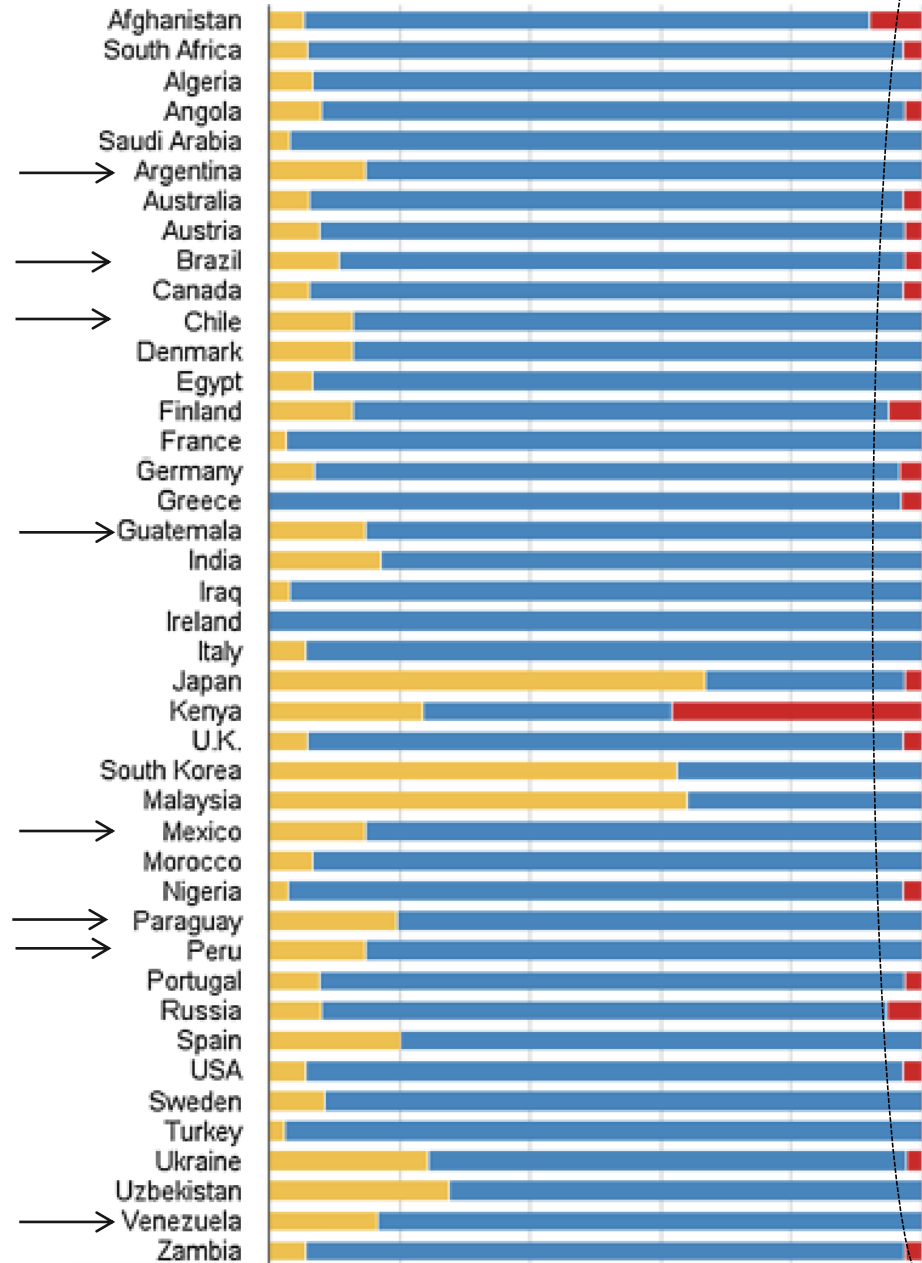
By [Caitlin Dewey](#) August 10 at 1:36 PM 



Now, thanks to [a new paper](#) by a trio of researchers at Brazil’s Universidade Federal de Minas Gerais, you can add a new prejudice to the list: Google’s depictions of female beauty, they found, are both ageist and racist.

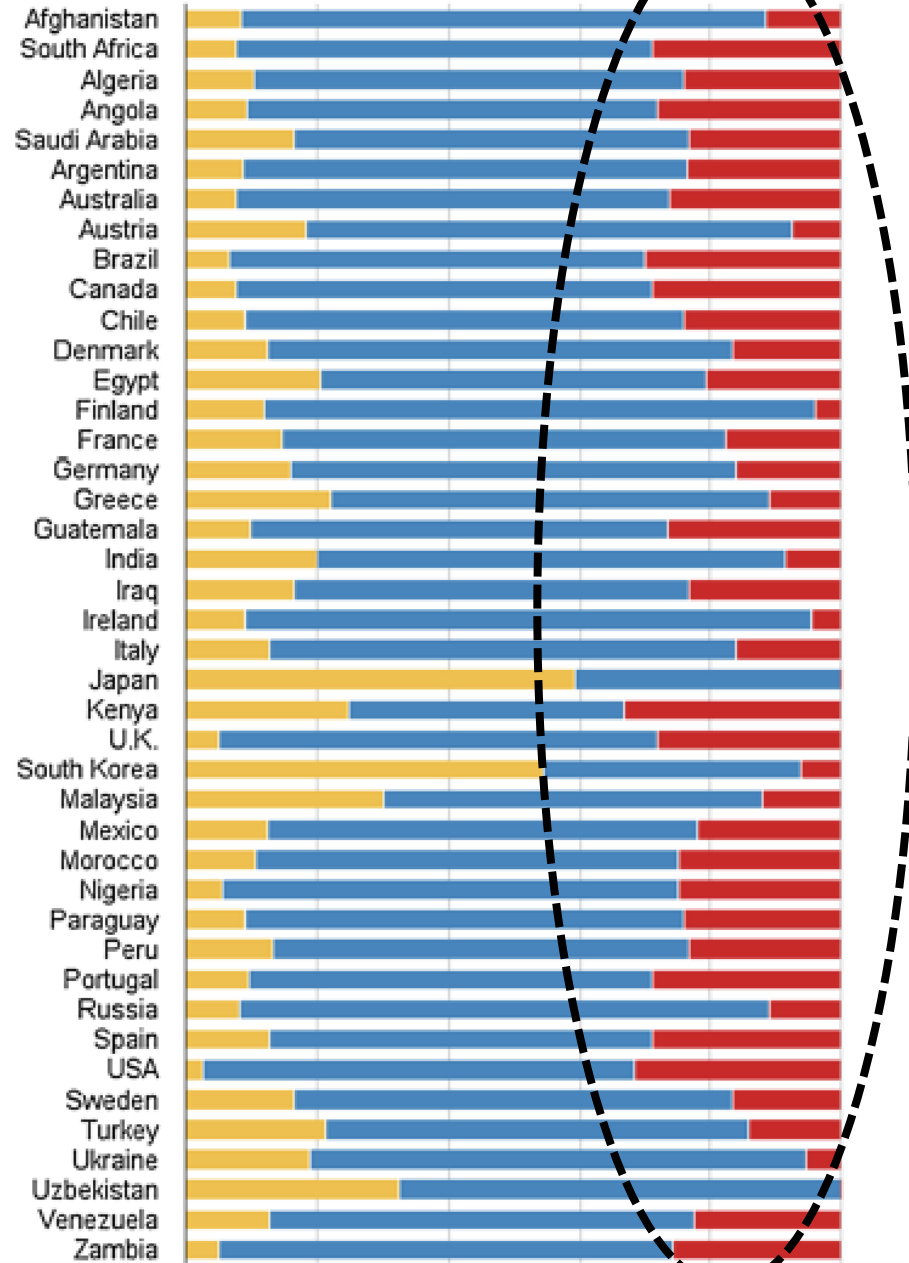
### Race Fractions for the query: 'beautiful woman'

Asian White Black



### Race Fractions for the query: 'ugly woman'

Asian White Black



# Analyzing right-wing conservative YouTube channels: hate, violence and discrimination



**Raphael Ottoni**  
<rapha@dcc.ufmg.br>

**Evandro Cunha**  
<evandrocunha@dcc.ufmg.br>

**Gabriel Magno**  
<magno@dcc.ufmg.br>

**Pedro Bernadina**  
<pedronascimento@dcc.ufmg.br>

**Wagner Meira Jr.**  
<meira@dcc.ufmg.br>

**Virgilio Almeida**  
<virgilio@dcc.ufmg.br>

# Research Questions

**. Is the presence of hateful vocabulary, violent content and discriminatory biases more, less or equally accentuated in right-wing conservative channels?**

**. Are, in general, commentators more, less or equally exacerbated than video hosts in an effort to express hate and discrimination?**

# Lexical Analysis

*Empath*



Vocabulary

# Topic Analysis

*L.D.A*



Subject

# Implicit Bias

*W.E.A.T*



Harmful Bias

Multi-layered methodology



. YouTube community seems to **amplify** a discriminatory bias against **Muslims** and **weaken** the bias against **LGBT people**.

. The implicit bias analysis shows **no differences** between **conservative** and **baseline** captions regarding immigrants and LGBT people, but it does show **against Muslims**.

## Some Findings

# Characterizing Attention Cascades in WhatsApp Groups



Josemar Alves Caetano, Gabriel Magno, Marcos Gonçalves, Jussara Almeida, Humberto T. Marques-Neto, Virgílio Almeida

# WhatsApp, politics and misinformation



May 2018

The Americas

WhatsApp is upending the role of unions in Brazil. Next, it may transform politics.

## Facebook's WhatsApp flooded with fake news in Brazil election

Anthony Boadle

5 MIN READ

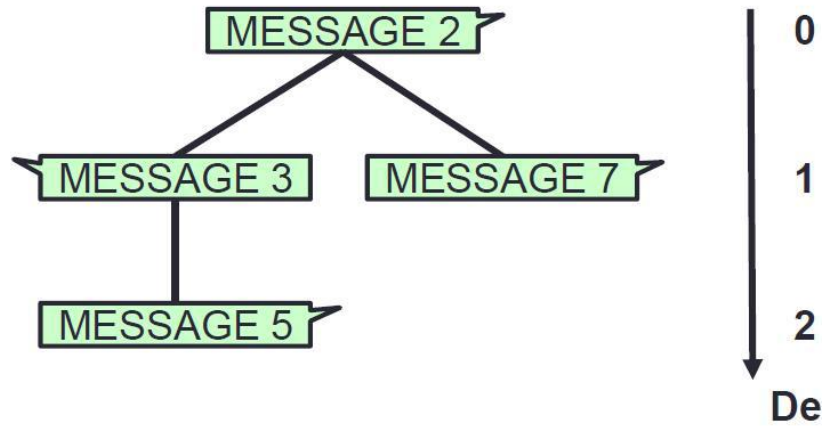


October 2018



# Research questions

- How different are attention cascades in political and nonpolitical groups on WhatsApp?
- What is the impact of false information on the characteristics of attention cascades?



Attention  
cascade  
example

# Studying attention cascades on WhatsApp using 3 dimensions

## Structural

- Number of nodes
- Depth
- Maximum breadth

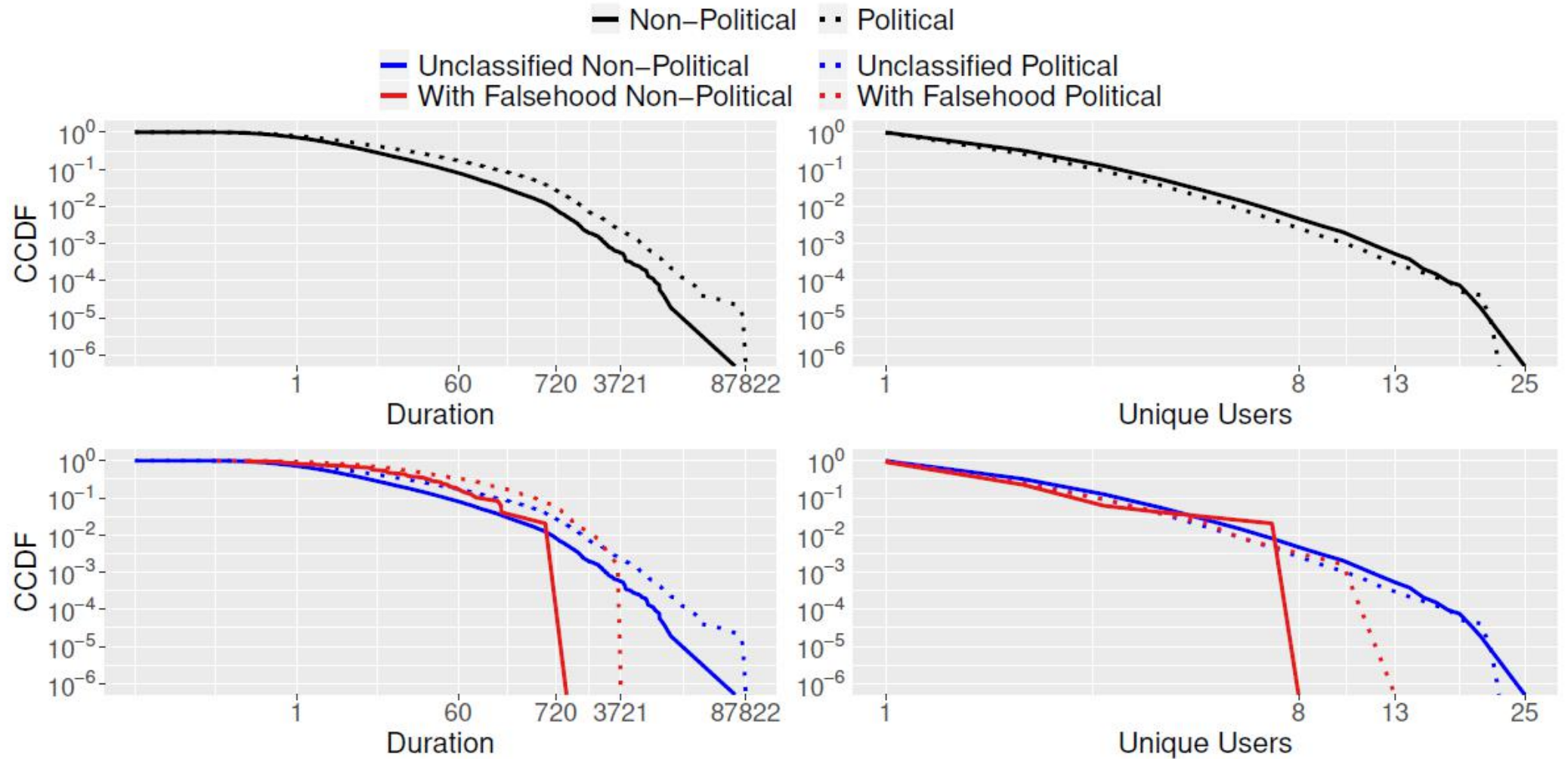
## Temporal

- Duration
- Structural characteristics over time

## User participation

- Unique users
- Relationships motifs

# Cascade duration and unique users



# Findings

- Differences in the structural and temporal characteristics of political groups when compared to non-political groups.
- Cascades with false information in political groups are deeper, wider and reach more users than cascades with falsehood in non-political groups.
- Political cascades last longer than non-political ones.
- Attention cascades in WhatsApp groups reflect real-world events that capture public attention



- E.g.: computing research as detector of social issues: hate speech, discrimination, etc.



# Auditing Radicalization Pathways on YouTube

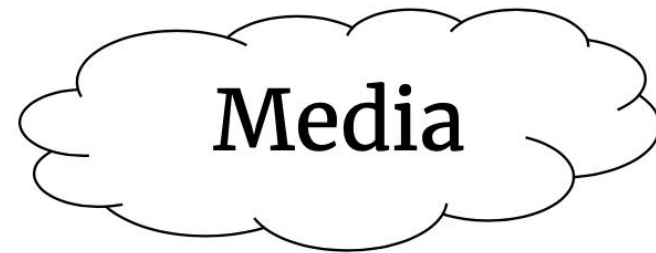


**Manoel Horta Ribeiro**  
**Raphael Ottoni**  
**Robert West**  
**Virgílio A. F. Almeida**  
**Wagner Meira Jr.**

FAT\* '20: Proceedings of the ACM 2020 Conference on Fairness, Accountability, and Transparency.



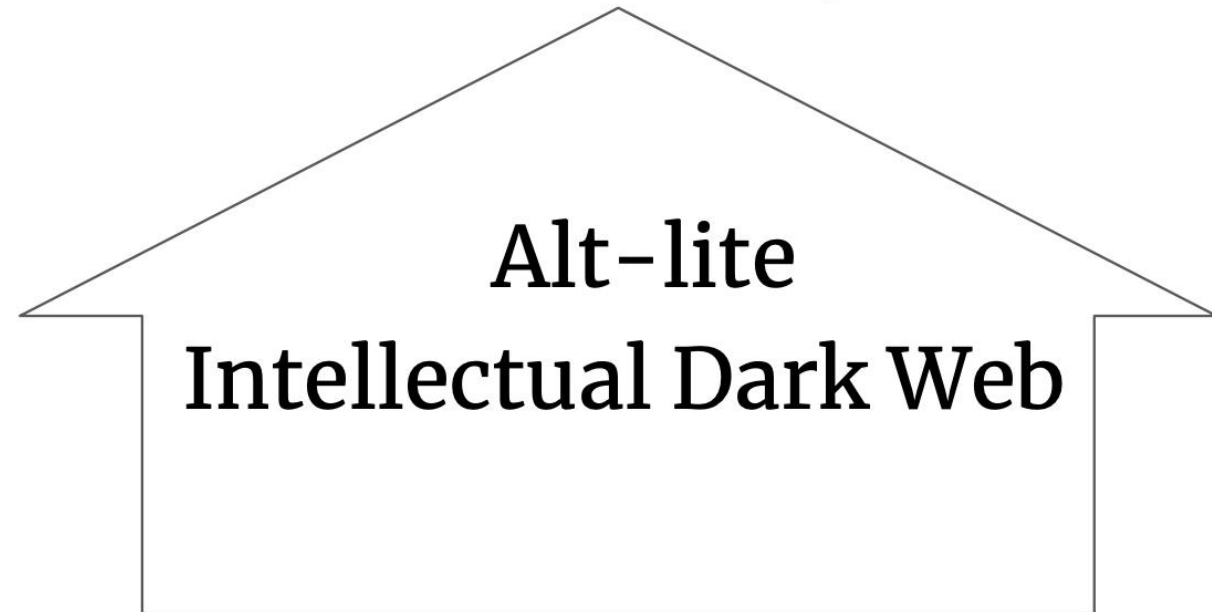
**349  
channels  
330k  
videos  
72M+  
comments**



(for comparison)

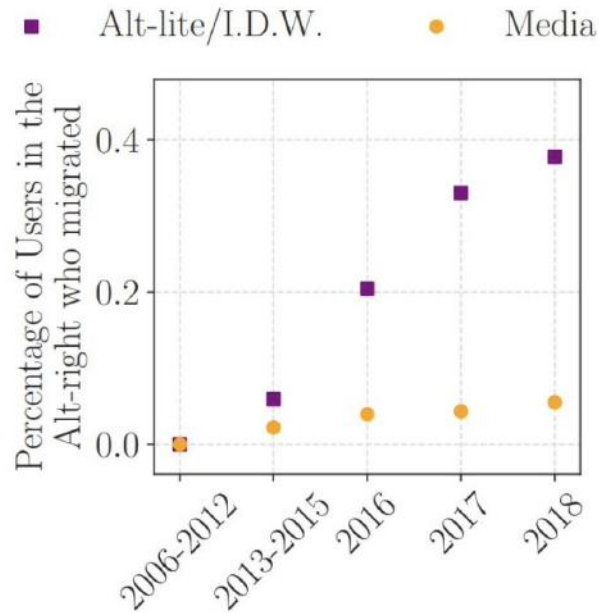


Radical community



Alleged "Gateways"  
(to the Alt-right)

# But how significant is this pipeline?



~40% of users who commented in the Alt-right in 2018 went through the pipeline (for all levels of exposure).

Research questions:

To which extent do users systematically gravitate towards more extreme content?

Do algorithmic recommendations steer users towards more extreme content?

# Findings

- Hypothesis: radicalization happens on YouTube
- By analyzing 349 channels, 330k videos and 72M comments, we found:
  - compelling evidence of significant and consistent migration from milder to more extreme communities
  - recommender system does not seem to be systematically benefiting the more extreme videos, but still binds communities together.
- We did not find why this phenomenon happens on YouTube.

## *There Is a Racial Divide in Speech-Recognition Systems, Researchers Say*

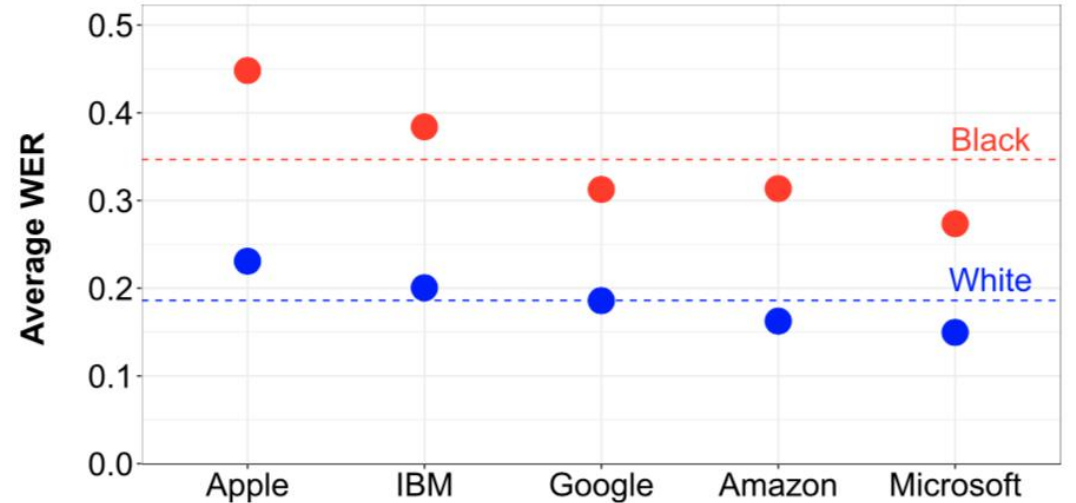
Technology from Amazon, Apple, Google, IBM and Microsoft misidentified 35 percent of words from people who were black. White people fared much better.



Amazon's Echo device is one of many similar gadgets on the market. Researchers say there is a racial divide in the usefulness of speech recognition systems. Grant Hindsley for The New York Times

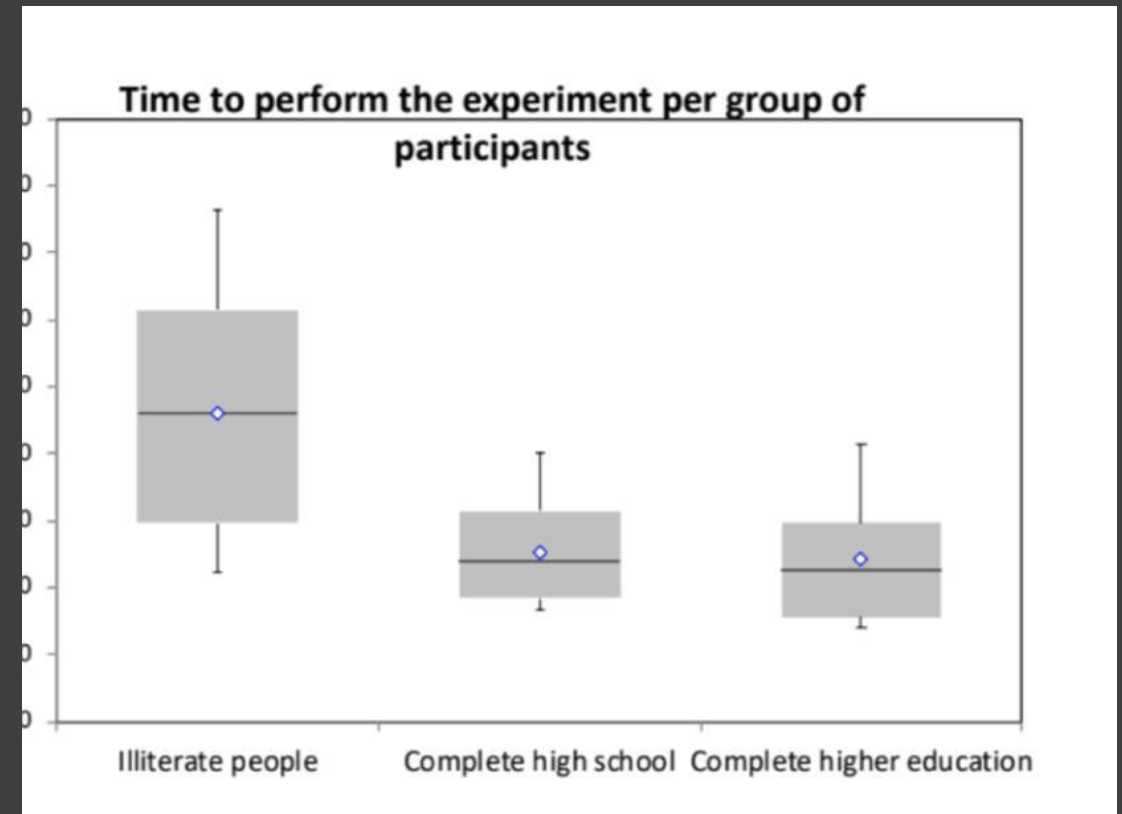
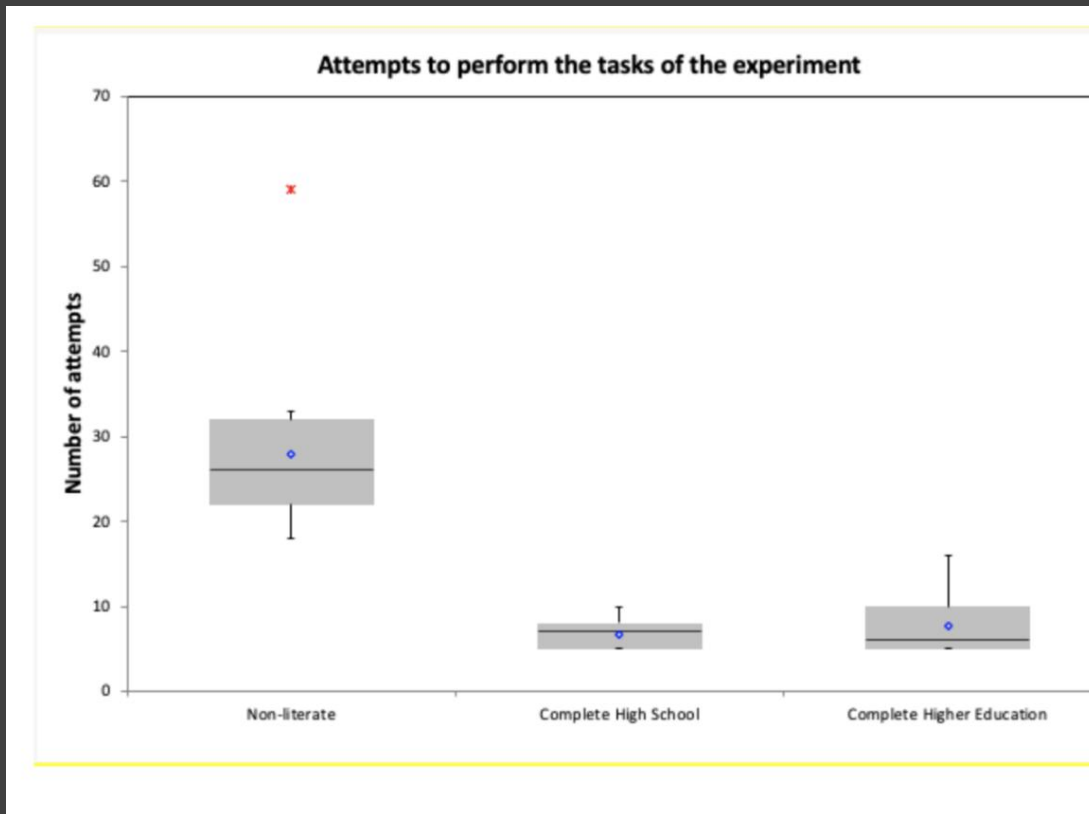


By Cade Metz



**Fig. 1.** The average WER across ASR services is 0.35 for audio snippets of black speakers, as opposed to 0.19 for snippets of white speakers. The

- Source: A. Koenecke et al., "Racial disparities in automated speech recognition," Proc. Nat. Acad. Sci., vol. 117, 354 no. 14, pp. 7684–7689, Apr 2020.



## How do illiterate people interact with Intelligent Voice Assistants?

- Helen T., Furtado F., Almeida V. et al, ``How do illiterate people interact with an Intelligent Voice Assistant? (2021) submitted.

# Measuring International Online Human Values with Word Embeddings

Authors: [Gabriel Magno](#), [Virgilio Almeida](#) [Authors Info & Claims](#)

## Measuring International Online Human Values with Word Embeddings

GABRIEL MAGNO and VIRGILIO ALMEIDA, Universidade Federal de Minas Gerais (UFMG), Brazil

As the Internet grows in number of users and in the diversity of services, it becomes more influential on peoples lives. It has the potential of constructing or modifying the opinion, the mental perception, and the values of individuals. What is being created and published online is a reflection of people’s values and beliefs. As a global platform, the Internet is a great source of information for researching the online culture of many different countries. In this work we develop a methodology for measuring data from textual online sources using word embedding models, to create a country-based online human values index that captures cultural traits and values worldwide. Our methodology is applied with a dataset of 1.7 billion tweets, and then we identify their location among 59 countries. We create a list of 22 **Online Values Inquiries (OVI)**, each one capturing different questions from the World Values Survey, related to several values such as religion, science, and abortion. We observe that our methodology is indeed capable of capturing human values online for different counties and different topics. We also show that some online values are highly correlated (up to  $c = 0.69, p < 0.05$ ) with the corresponding offline values, especially religion-related ones. Our method is generic, and we believe it is useful for social sciences specialists, such as demographers and sociologists, that can use their domain knowledge and expertise to create their own Online Values Inquiries, allowing them to analyze human values in the online environment.

## Twitter Collection

- Random sample of 1% of tweets (Internet Archive)
- Year: 2014
- Total: 1.7 billion tweets



# Characterizing Vaccination Movements on YouTube in the United States and Brazil

Marcelo Sartori Locatelli  
Josemar Caetano  
Wagner Meira Jr.  
Virgilio Almeida

June 2022



# Research Questions

- Q1: What are the characteristics that differentiate pro-vaccine and anti-vaccine online comments?
- Q2: What kind of language drives engagement in online comments?
- Q3: What are similarities and differences between pro and anti-vaccination campaigns in the United States and Brazil?
- Q4: What are characteristics of the pro and anti-vaccine videos removed from YouTube?

Channel Category	Channel Country	Accounts Number	Total Comments	Comments per Account	% incl URL
Anti-Vaccine	BR	539.678	2.308.712	4,28	21,54%
Pro-Vaccine	BR	350.774	649.644	1,85	22,33%
Anti-Vaccine	USA	280.979	584.844	2,08	13,50%
Pro-Vaccine	USA	1.233.565	3.017.758	2,45	12,99%
Anti-Vaccine	OTHER	248.463	709.947	2,86	14,57%
Pro-Vaccine	OTHER	437.707	1.261.304	2,88	18,14%

Table 1. Basic Statistics by Channel Category and Country for the YouTube Dataset. % **incl URL** represents the percentage of the total number of comments that contain URLs. While building this table, we discarded 8.025 comments that have neither country nor category.



## Dataset

- Data on 142 channels (pro and anti-vaccine);
- Vaccine related videos;
- Comments on those videos;
- Detection of video takedowns;
- Videos tracked from 11th July 2021 to 3rd October 2021.

# Results - Tone of Comments – Valence (Brazil)

Pro-Vaccine

Anti-Vaccine

$-1 \leq v \leq -0.75$

$-0.75 < v < -0.5$

$-0.5 \leq v \leq 0.5$

$0.5 < v < 0.75$

$0.75 \leq v \leq 1$

vaccine  
report  
fiocruz  
researcher  
healthy  
to\_prevent  
respiratory  
workers  
to\_cough  
poverty

health  
disease  
isolation  
case  
avoid  
confrontation  
sus  
mask  
science  
solidarity

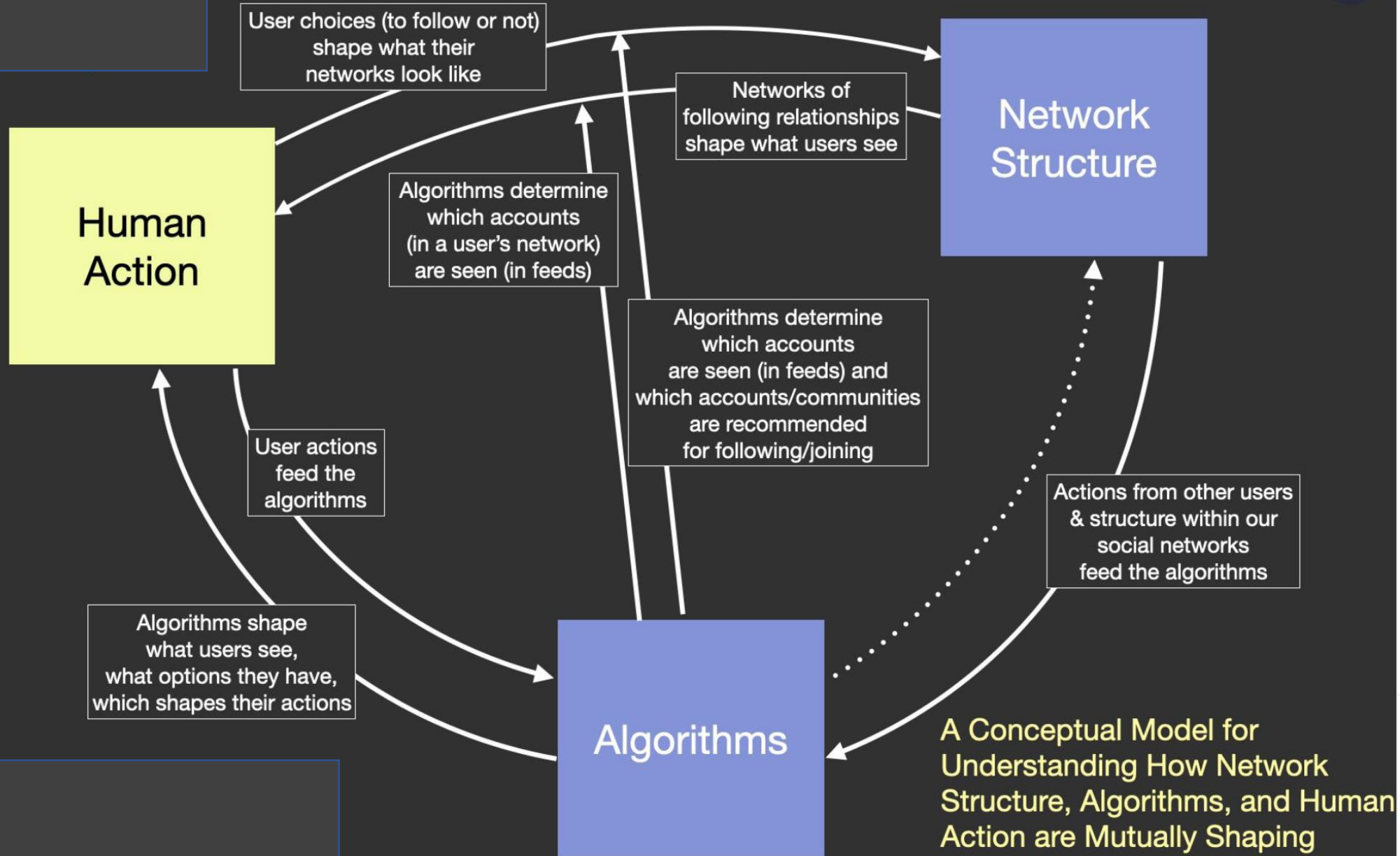
brazil  
bolsonaro  
president  
coronavirus  
brazilian  
family  
population  
pandemic  
covid  
combat

minister  
god  
freedom  
left  
media  
nation  
lord  
chaos  
success  
venezuela

hydroxychloroquine  
alliance  
communist  
conservative  
leftist  
terrorist  
cheer  
christian  
patriot  
pastor

# Conclusion

- The results provide insights to understanding communication techniques of pro- and anti-vaccine groups;
- The results can be useful for designing interventions to minimize dissemination of anti-vaccine content;
- Moderation policies on social networks vary depending on the language and the perceived importance of the country.



Source: Starbird,  
54 K. 2021

# 'Risks posed by AI are real': EU moves to beat the algorithms that ruin lives



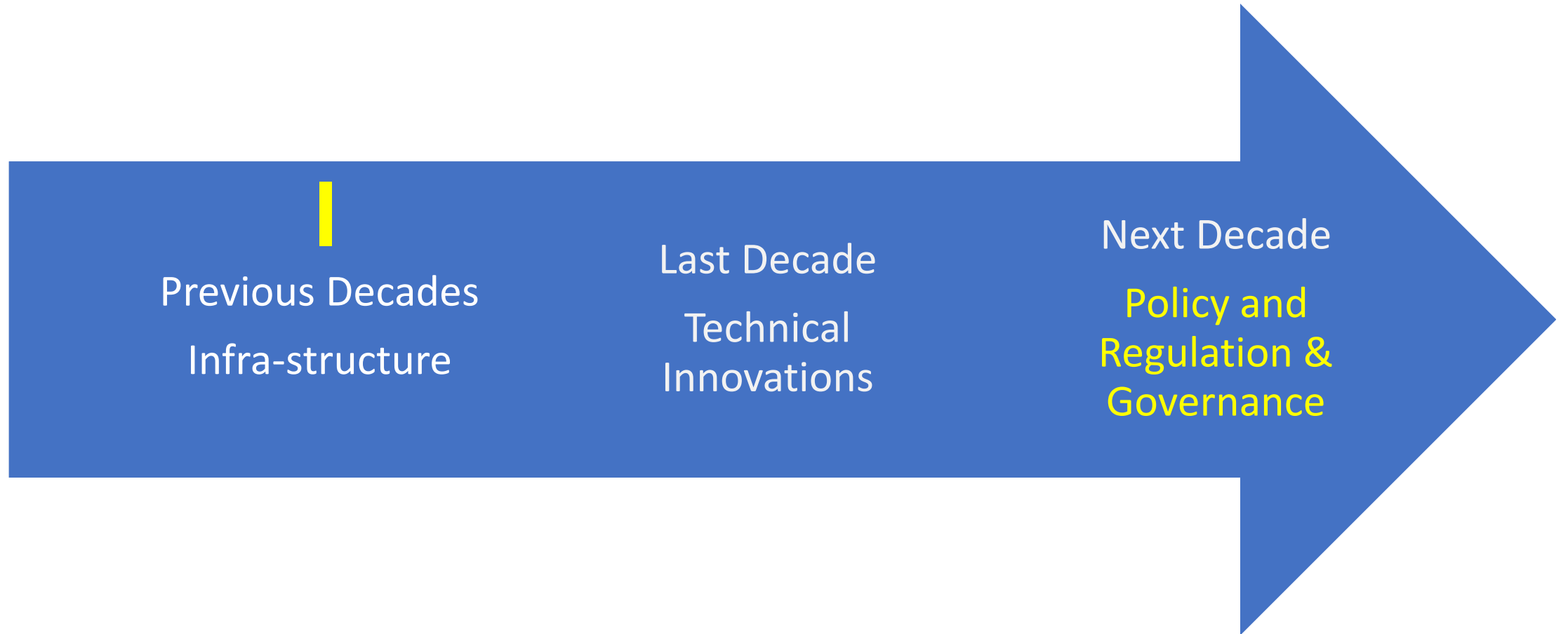
**Kalyeena Makortoff**  
*Banking correspondent*

[@kalyeena](#)

Sun 7 Aug 2022 13.26 BST



# Evolution of the Digital World: narratives





# Digital Governance

- “Digital governance is the practice of establishing and implementing policies, procedures, and standards for the proper development, use, and management of the infosphere” (Floridi 2018).
- Digital governance is the capacity of multiple institutions in the digital world to govern (in a legitimate, inclusive, and secure manner) the use of digital commons to produce sustainable services and public policies implemented by governments and firms in a non-territorial and results-based manner (Filgueiras and Almeida, 2020).
- Three essential and common elements of governance structures: i) Decision-making; ii) Accountability and transparency; iii) Coordination

EDITOR: Virgílio Almeida, virgilio@dcc.ufmg.br

DEPARTMENT: INTERNET GOVERNANCE

# Algorithms and Institutions: How Social Sciences Can Contribute to Governance of Algorithms

Virgílio Almeida , Harvard University, Cambridge, MA, 02138, USA

Fernando Filgueiras , Federal University of Goiás, Goiânia, 74690-900, Brazil

Ricardo Fabrino Mendonça , Federal University of Minas Gerais, Belo Horizonte, 31270-901, Brazil

*Societies are experiencing a transition in which automation increases its role in daily activities, and algorithms play a key role. Algorithms are the ones that often decide whether one is hired or fired, whether one will have access to a social benefit, whether one's immigration visa will be granted or denied, what news one will see on social media, what is the best route to go home, or which partner is most appropriate for a relationship. In many directions and fields, digital technologies are enabling the re-engineering of society, with algorithms as a central element in this process.*

In many directions and fields, digital technologies are enabling the re-engineering of society, with algorithms as a central element in this process.

The central point of this thesis of algorithms as institutions is that debating the governance of the digital world means thinking about the power that many digital technologies exert to shape human action and structure a new way of organizing society.

obrigado!

virgilio@dcc.ufmg.br

publication list available at:

<https://scholar.google.com/citations?user=sPKpIPwAAAAJ&hl=en&oi=ao>