Universitat Pompeu Fabra Barcelona

HARVARD HUMANITARIAN INITIATIVE

DATA-POP ALLIANCE

MIT Connection Science
the technology of innovation

(About) The Promises and Perils of AI for Statistics

Emmanuel Letouzé, PhD
eletouze@datapopalliance.org

18TH WORLD TELECOMMUNICATION/ICT INDICATORS SYMPOSIUM

ITUWTIS GENEVA2023

Advancing the measurement agenda to achieve universal and meaningful connectivity

3-4 July 2023
Geneva, Switzerland

#ITUWTIS
www.itu.int/wtis2023

Outline

1. Introduction(s) and Thanks (2mns)
2. Context, Concepts and Question(s) (6 mns)
3. Promise and Opportunities (6 mns)
4. Hurdles and Risks (6 mns)
5. The Way Forward: Requirements and Roadmap (5 mns)

Outline

1. **Introduction(s) and Thanks**
2. Context, Concepts and Question(s)
3. Promise and Opportunities
4. Hurdles and Risks
5. The Way Forward: Requirements and Roadmap

Many thanks to ITU, the organizing team, Esperanza, Thierry, Fredrik, Alexandre, Anne-Laure, Roseline…..

18TH WORLD TELECOMMUNICATION/ICT INDICATORS SYMPOSIUM

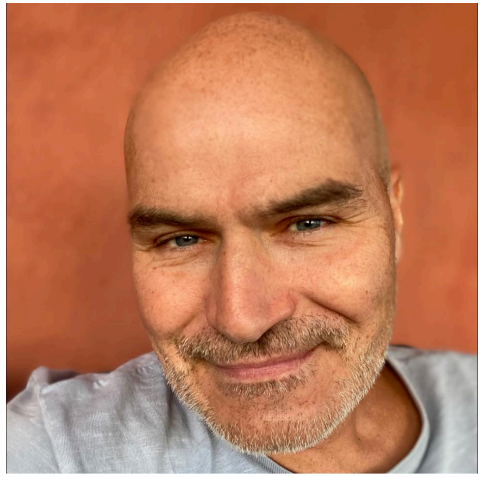
ITUWTIS GENEVA2023

Advancing the measurement agenda to achieve universal and meaningful connectivity

3-4 July 2023
Geneva, Switzerland

#ITUWTIS
www.itu.int/wtis2023

A bit about my background and trajectory



(Let's connect!)

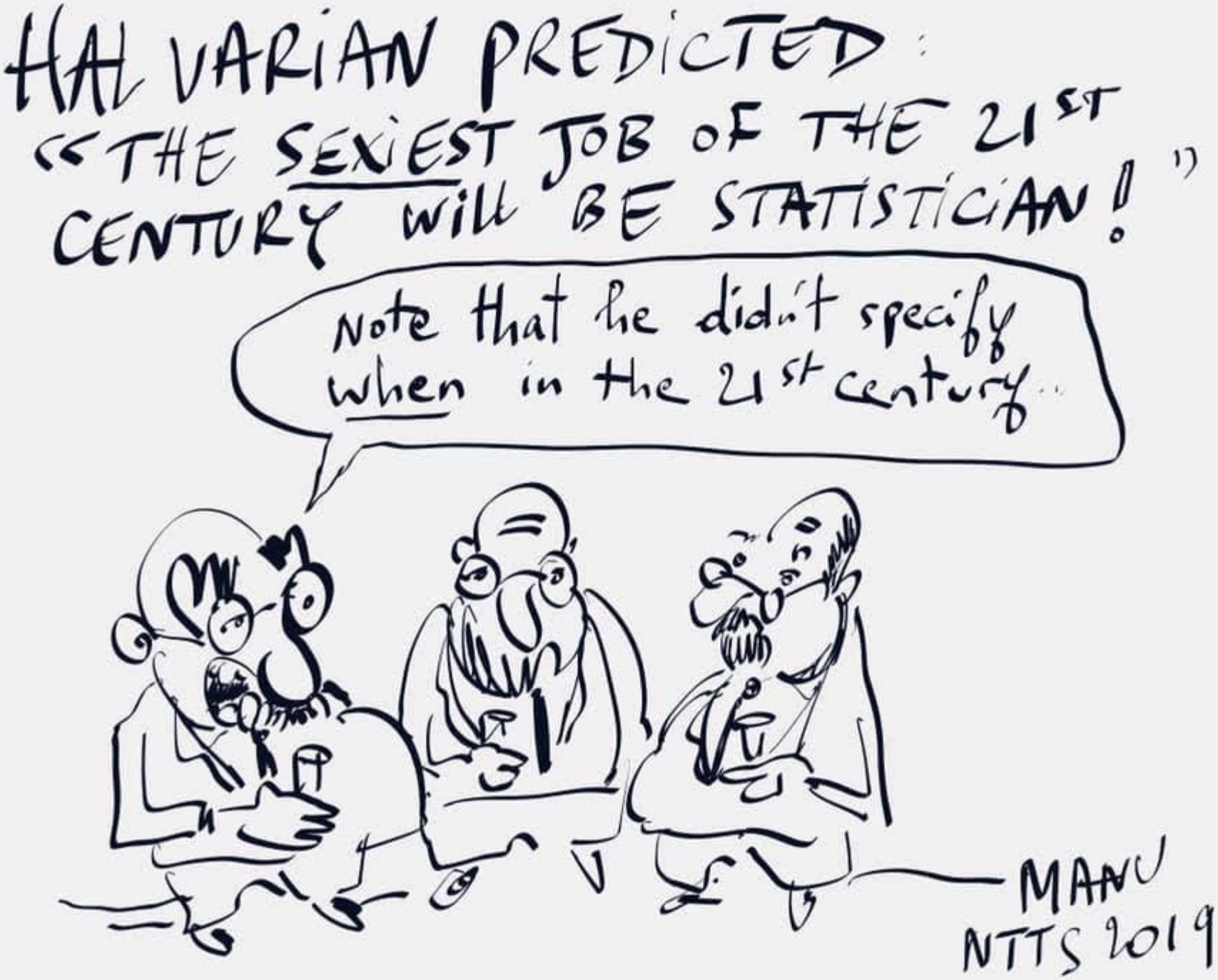
- **Director and Co-Founder of Data-Pop Alliance (2013—present)**
- Marie Curie Fellow at Universitat Pompeu Fabra (2021—2023)
- Founder and Executive Director of the OPAL project (2016-2020)
- **UN Global Pulse Senior Development Economist (2011-2012)**
- UNDP Economist (2006-2009)
- Technical Assistant for French Ministry of Finance in Vietnam (2000-04)
- (Semi-Failed) Political Cartoonist since 1975

- Post-Doctoral Fellow at MIT Media Lab (2016-17)
- **PhD in Demography from University of California, Berkeley (2009-2016)**
- MA in International Affairs and Development, Columbia U. (2004-2005)
- MA in Economic Demography, Sciences Po Paris (2000)
- BA in Political Science, Sciences Po Paris (1999)

- Technical Advisory Board Member, GPSDD (2016-present)
- **Member of the European Commission Expert Group on "Facilitating the Use of Non-Traditional Data for Official Statistics" (2021-2022)**
- Member of the International Union for the Scientific Study of Population (IUSSP)'s Scientific Panel on Big Data and Population Processes (2016-18)
- Program Committee Member UN World Data Forums I and II (2017-2018)

Outline

1. Introduction(s) and Thanks
2. **Context, Concepts and Question(s)**
3. Promise and Opportunities
4. Hurdles and Risks
5. The Way Forward: Requirements and Roadmap



Now….will GAI take over?

Will AI lead to a ‘Meiji’ moment?

” Big Data / AI and Official Statistics”: The topic is not very new, there’s been some progress, but limited and slow…

Big Data and Official Statistics: From Opportunities to Strategies

Emmanuel Letouzé*
OECD-Paris21, Paris, France, and University of California, Berkeley, USA
eletouze@berkeley.edu

Eric Bensel
OECD-Paris21, Paris, France Eric.BENSEL@oecd.org

Johannes Jütting
OECD-Paris21, Paris, France Johannes.JUTTING@oecd.org

60th Anniversary of the National Department of Statistics of Colombia (DANE)

Download

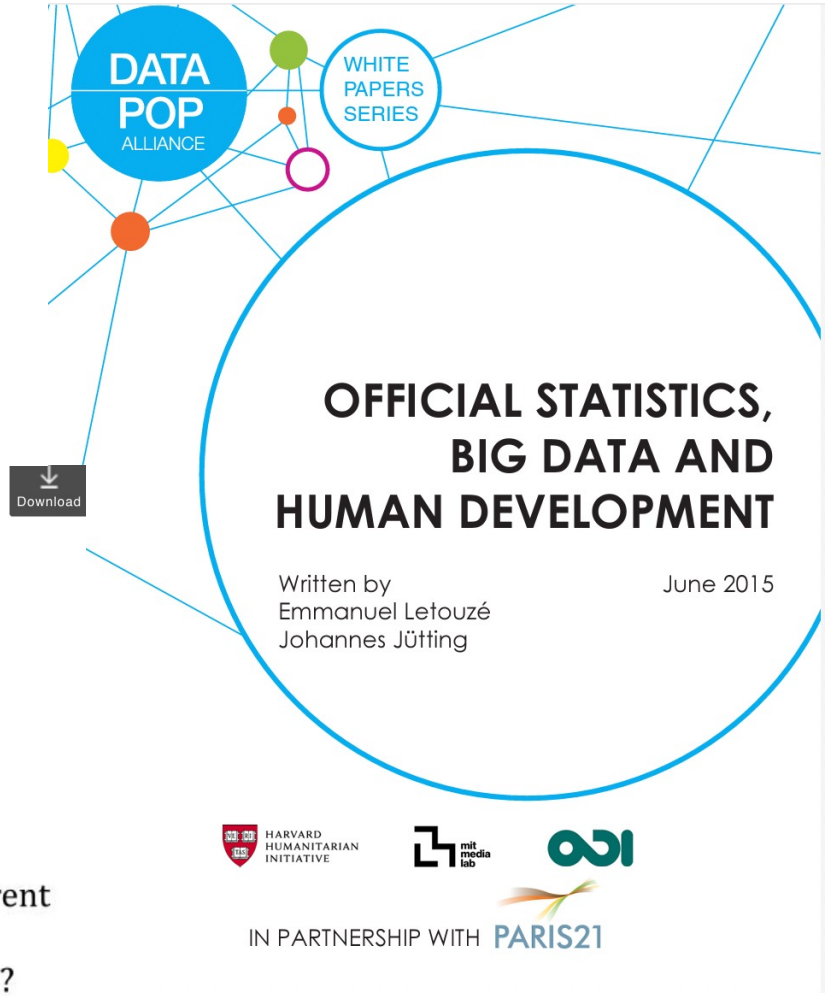
5 Questions on Official Statistics in the Big Data Era

Emmanuel Letouzé
PhD Candidate, UC Berkeley
Fellow, Harvard Humanitarian Initiative
Non-Resident Adviser, International Peace Institute
eletouze@berkeley.edu

Biblioteca Luis Angel Arango
Bogotá
October 28th, 2013

Outline: 5 questions

1. Why & what is the question? The “Statistical Tragedy” vs. the “Data Revolution”
2. What is ‘Big Data’ about—and not about?
3. What is/are Official Statistics and why does it matter?
4. Where do we stand? Early applications and current limitations
5. Where and how do we move forward and ahead?



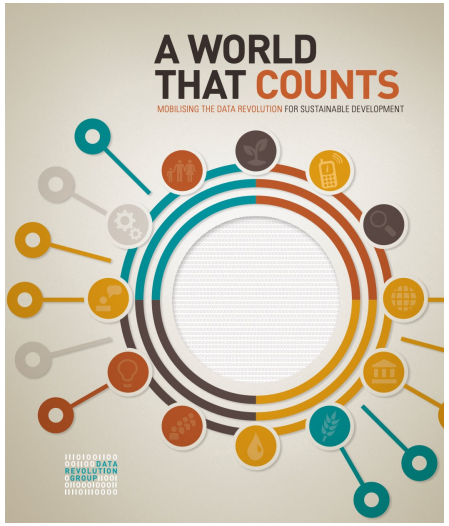
Download

DATA-POP ALLIANCE
Inputs to the Big Data and SDGs Chapter of the 2015 Global Sustainable Development Report

Reflections on Big Data and the Sustainable Development Goals: Measuring & Achieving Development Progress in the Big Data Era

February 2015

Let's rewind: the "Data Revolution" raised many expectations



*"We are at the beginning of what I call The **Industrial Revolution of Data.**"* Joe Hellerstein, Nov. 2008

THE DATA REVOLUTION FOR HUMAN DEVELOPMENT

Arab Region NSOs Call For Data Revolution

The world's most valuable resource is no longer oil, but data

The Economist

La révolution des données est-elle en marche ?
Implications pour la statistique publique et la démocratie
Thomas Roca, Emmanuel Letouzé
DANS AFRIQUE CONTEMPORAINE 2016/2 (n° 258), PAGES 95 À 111

WIRED

How Valencia crushed Covid with AI
By leveraging algorithms and unorthodox data sources, an MIT researcher has made Valencia a Covid-19 data pioneer

It has fallen short in many respects. Can we course-correct?

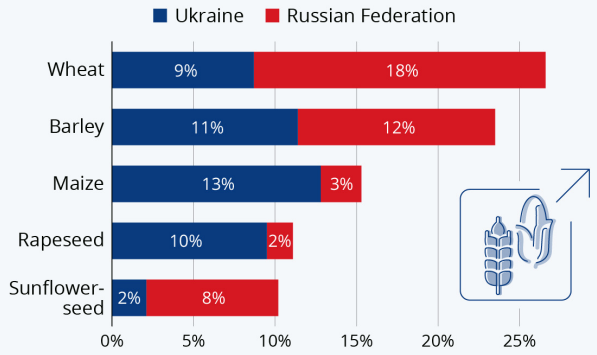
THE WORLD'S RICHEST 1% HAVE MORE THAN TWICE AS MUCH WEALTH AS 6.9 BILLION PEOPLE.

Source: Oxfam 2020

U.S. Billionaire Wealth Surges During Pandemic

Why the War in Ukraine Threatens Global Food Security

Ukraine's and Russia's share in global exports of selected crops (2016-2020 average)



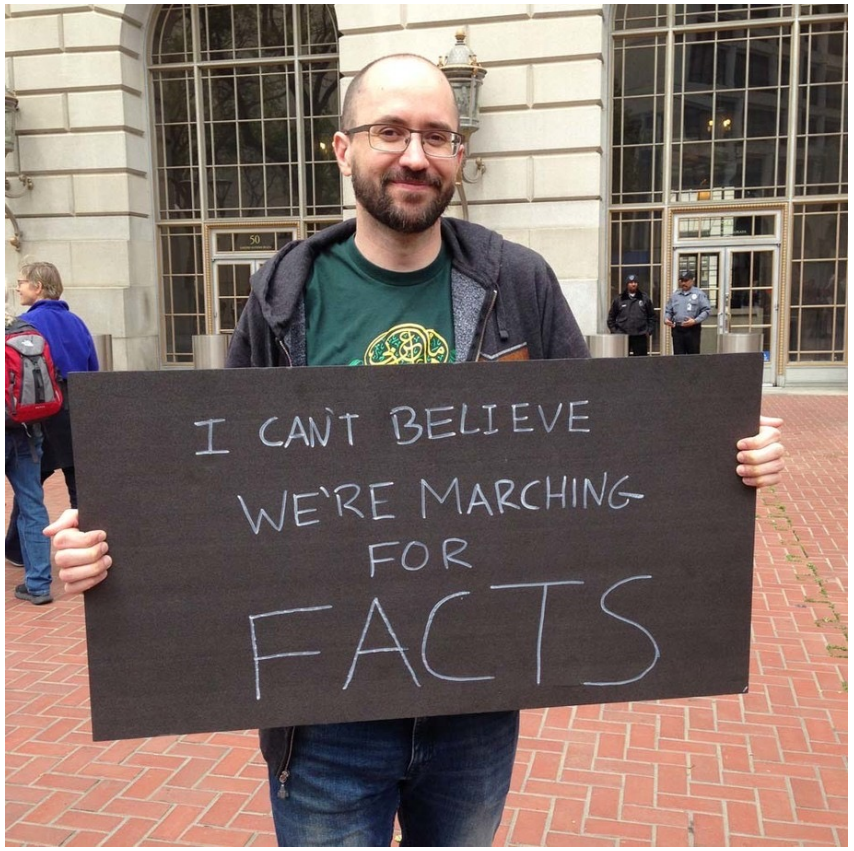
ista

IS DEMOCRACY DYING?

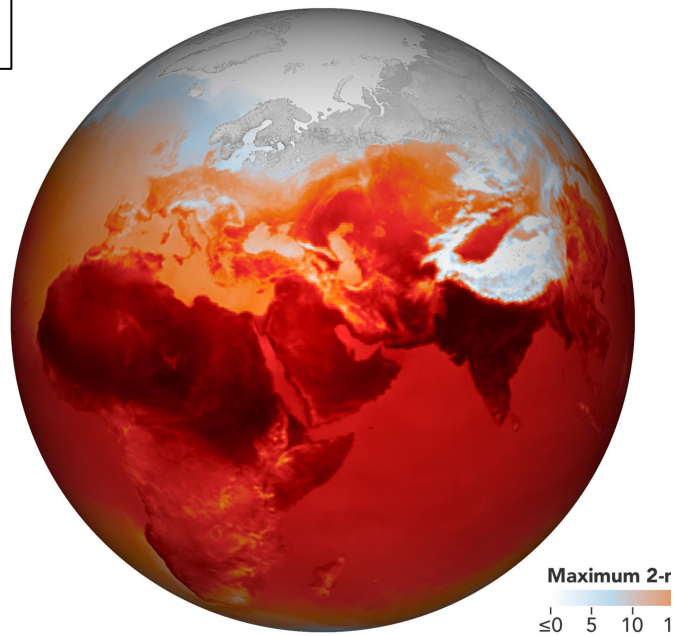
A WARNING FROM EUROPE: The Worst Is Yet to Come BY ANNE APPLEBAUM

HOW AI COULD GIVE RISE TO TYRANNY by Yuval Noah Harari

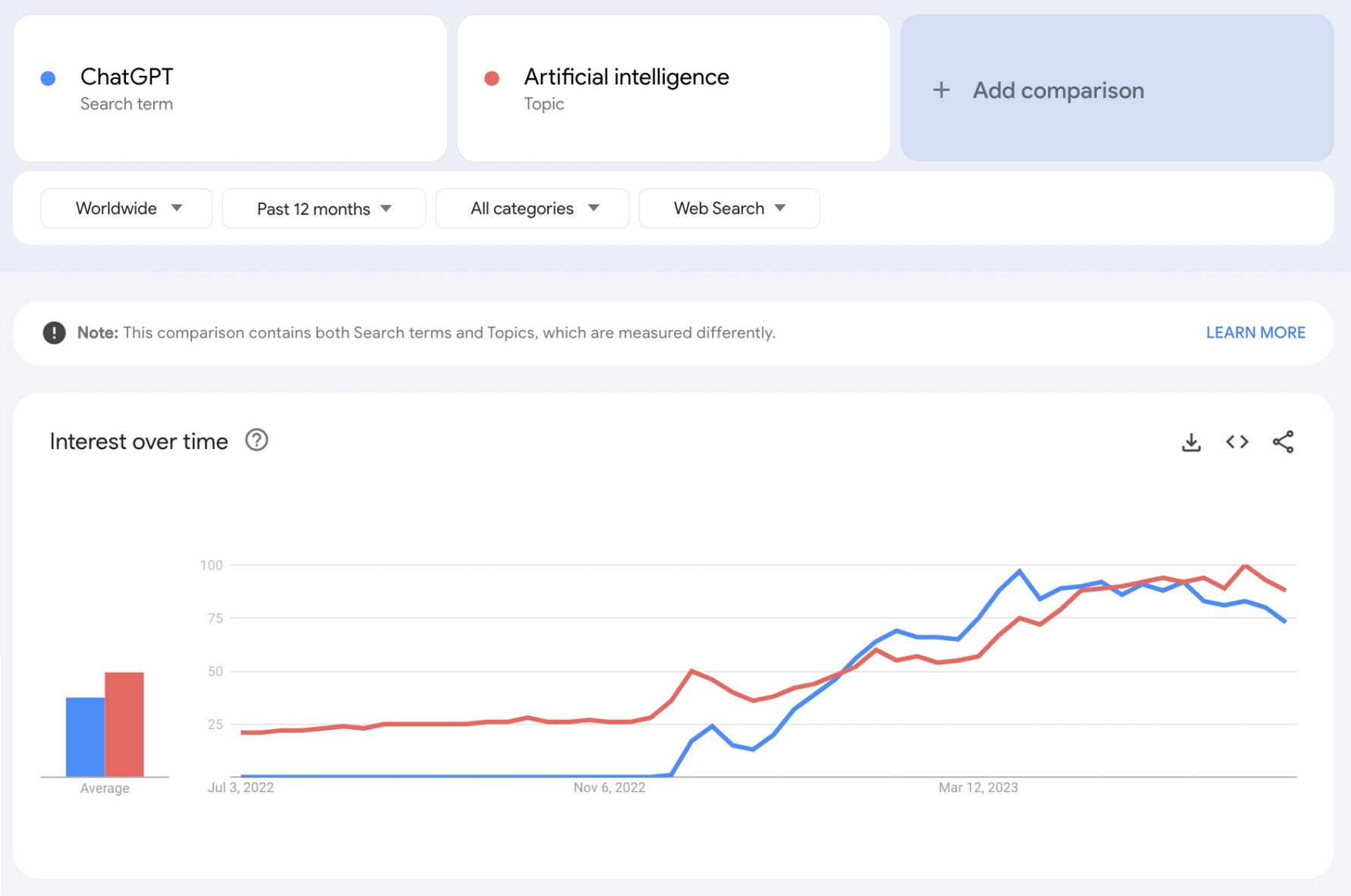
A new low for global democracy



MLTalks: How Data Killed Facts
Jill Lepore in conversation with
Andrew Lippman



‘New’ AI including LLMs (Large Language Models) creates new promises, risks, concerns, and confusion



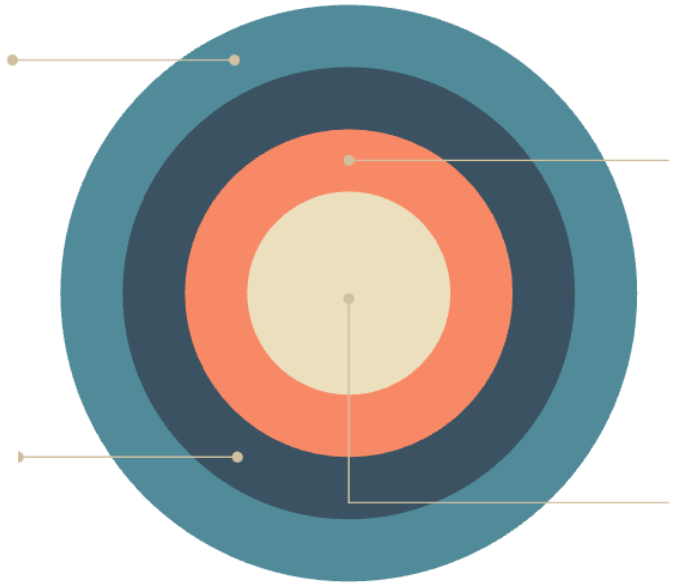
First, we need clarity and to use concepts and frameworks that account for real world complexities: mine is the 4Cs

| FIGURE 1 |

The four Cs of AI as a socio-technological phenomenon, based on Letouzé (2015).

AI culture
Incentives, expectations and norms that arise from and shape the use of AI systems. This can be thought of as the superstructure of AI.

AI communities
Contributors, users and developers of AI systems operating and interacting under specific arrangements and regulations (potentially the whole population). This can be thought of as the macrostructure of AI.

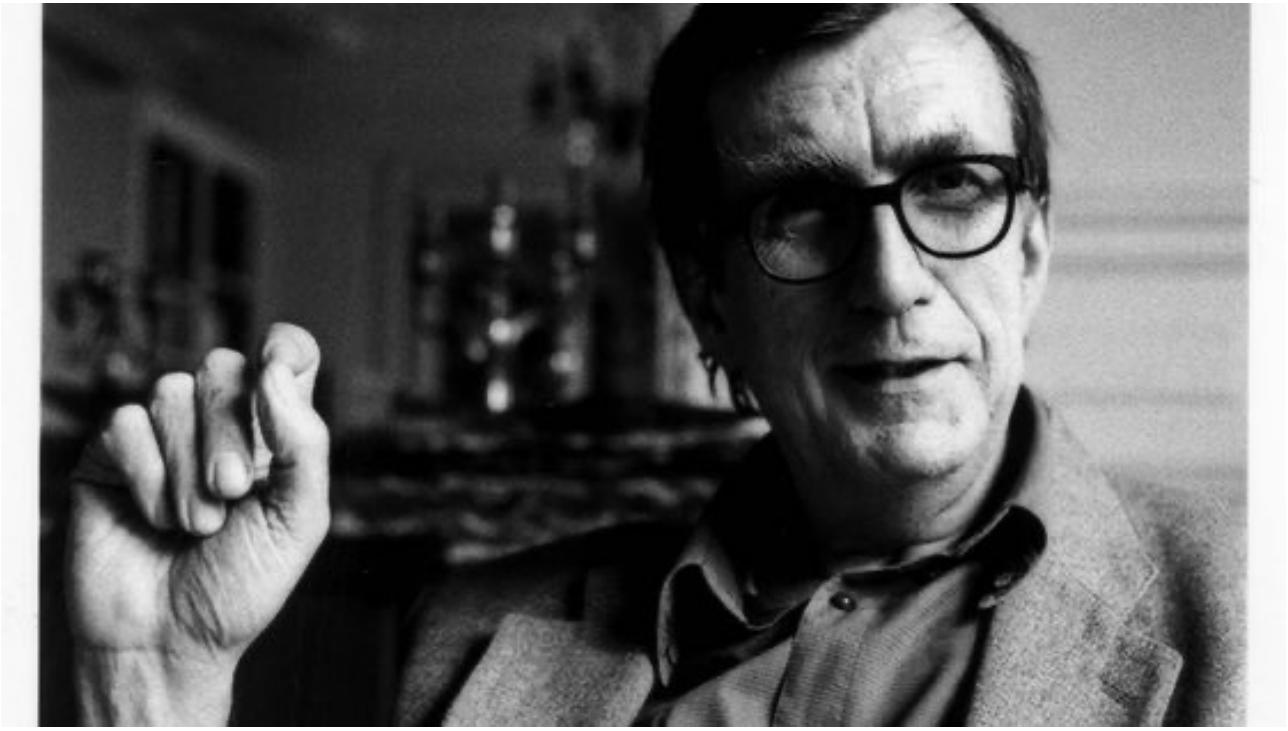


AI capacities
The tools and methods, hardware and software, know-how and skills necessary to process and analyze these new kinds of data, as well as to discuss and regulate them. This can be thought of as the infrastructure of AI.

AI crumbs
Pieces of digital data that humans leave behind passively as by-products of actions and interactions involving digital devices and services. This can be thought of as the fuel of AI.

"Better Data, Better Decisions, Better Lives" is a (good) slogan BUT the relationship between data, stats, facts, development, democracy, trust and culture is very complex

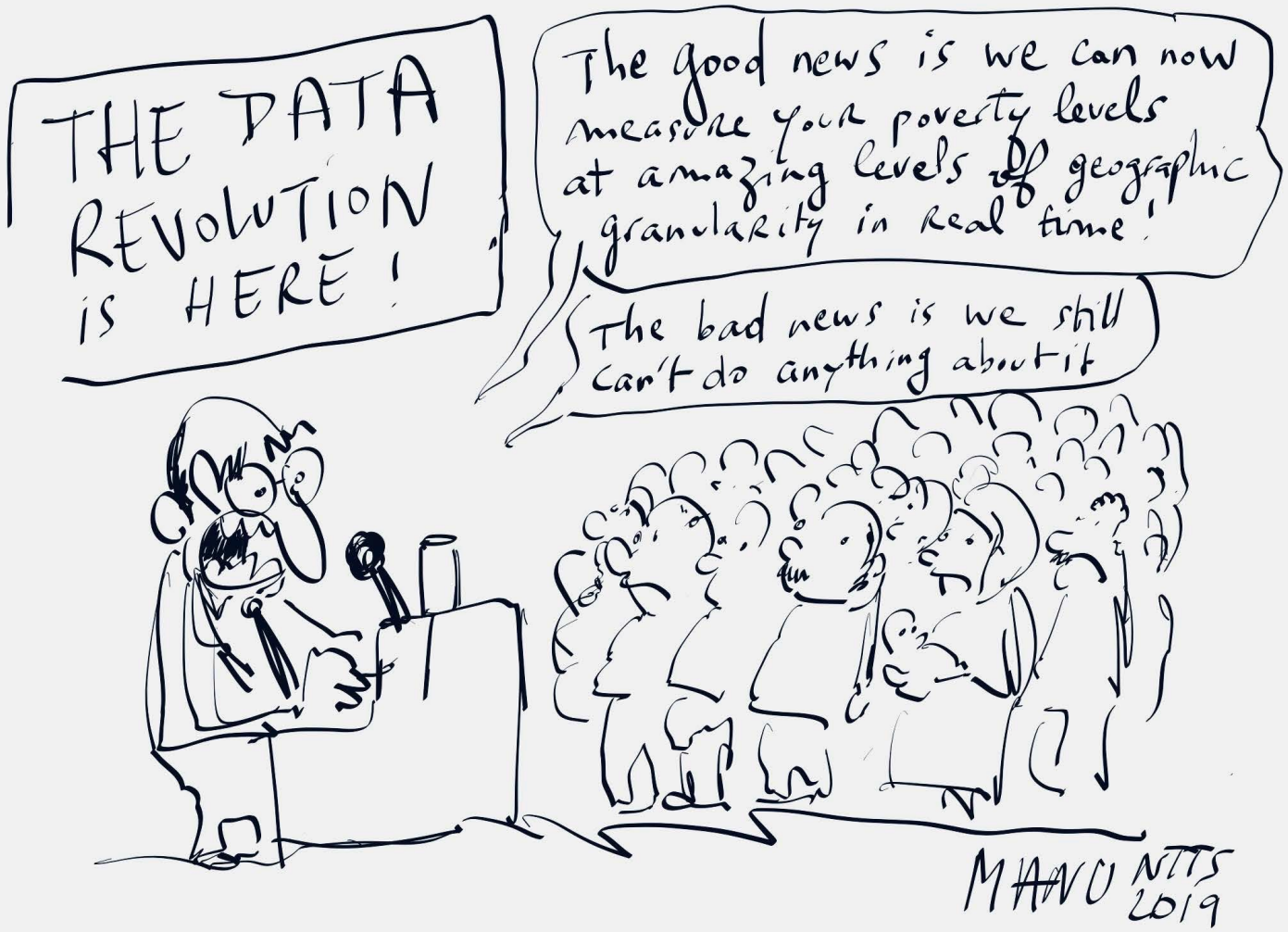
Bruno Latour* (1947-2022)

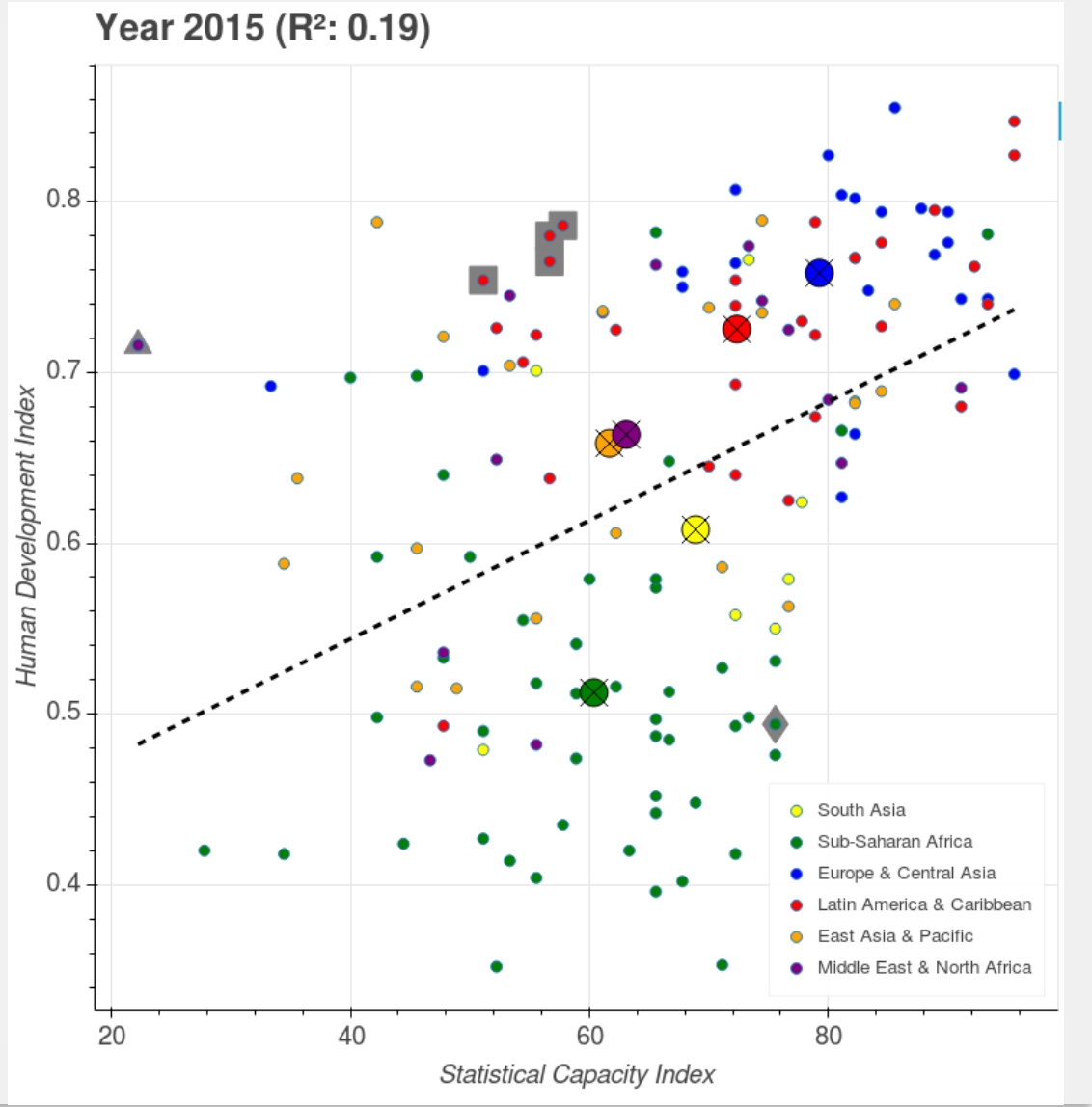


« What journalists, scientists and other experts fail to grasp, Latour argues, is that ***facts remain robust only when they are supported by a common culture, by institutions that can be trusted, by a more or less decent public life, by more or less reliable media*** ».

* https://en.wikipedia.org/wiki/Bruno_Latour

Statistiks ('Science of the State') is a pretty recent industry and often it seems statistics don't *matter* very much …





Statistics is actually 2 things, with a dual nature and role

Enrico Giovannini's "Statistics 2.0" on the role of Statistics (2010)

1) As *data* ("official statistics"), to produce knowledge responding to needs and standards

"The essential role of modern statistics, referred to as "statistics 2.0", is to provide society with "knowledge of itself, on which to base its own choices and evaluate the effects of political decisions.

2) As a *public institution* ("Official Statistics"), to provide a deliberative space

".. to freely and openly debate what is worth measuring, how it is measured and why it is measured" – to act as "a debated public institution

That has implications for our question

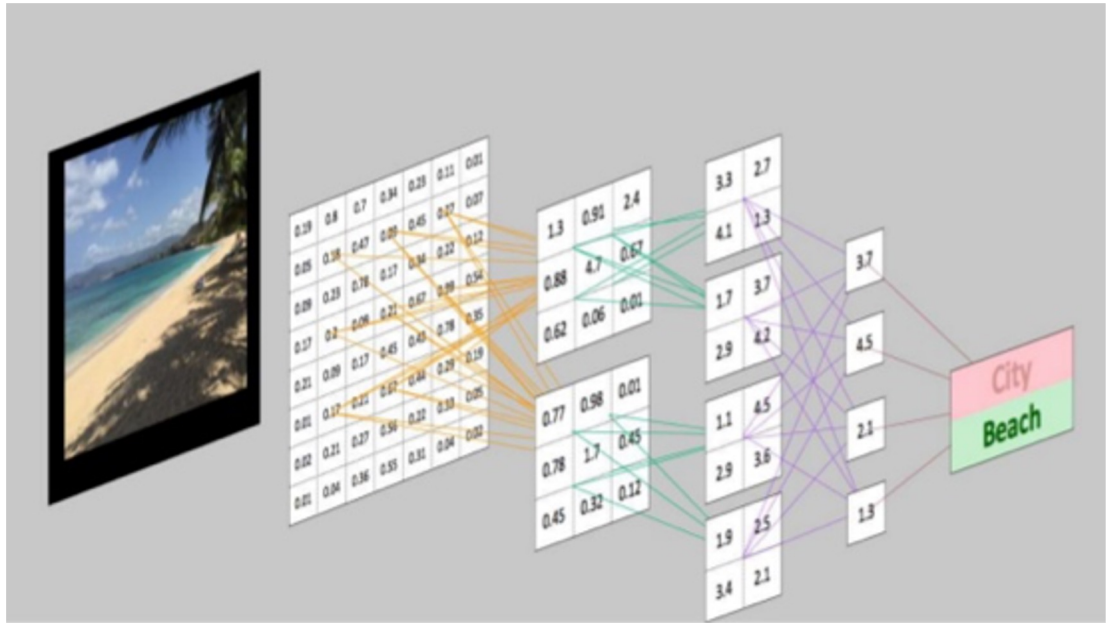
In my view the key question is / remains :

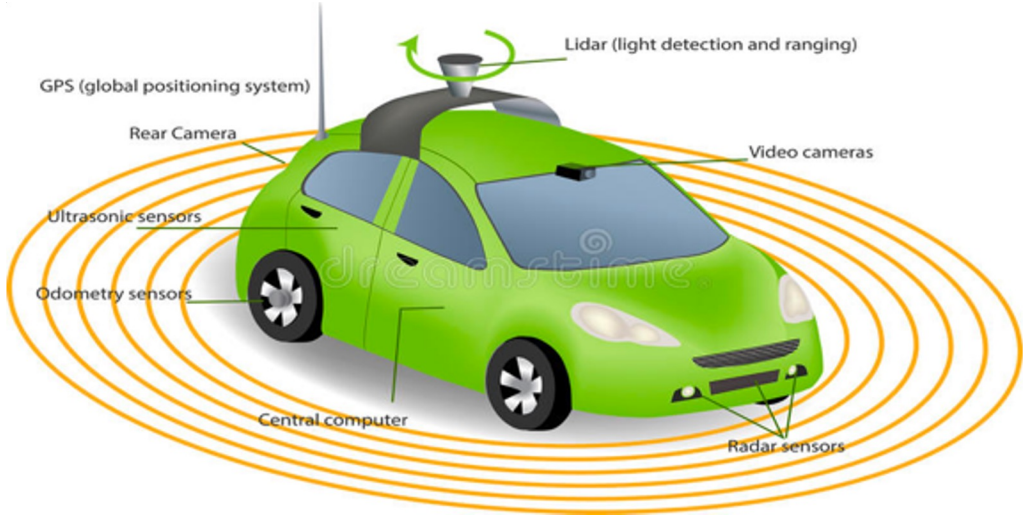
How can Official Statistics *as an industry producing official statistics* leverage AI *as conceptualized by the 4Cs* to retain, regain, or revive and reinforce its key role for development and democracy?

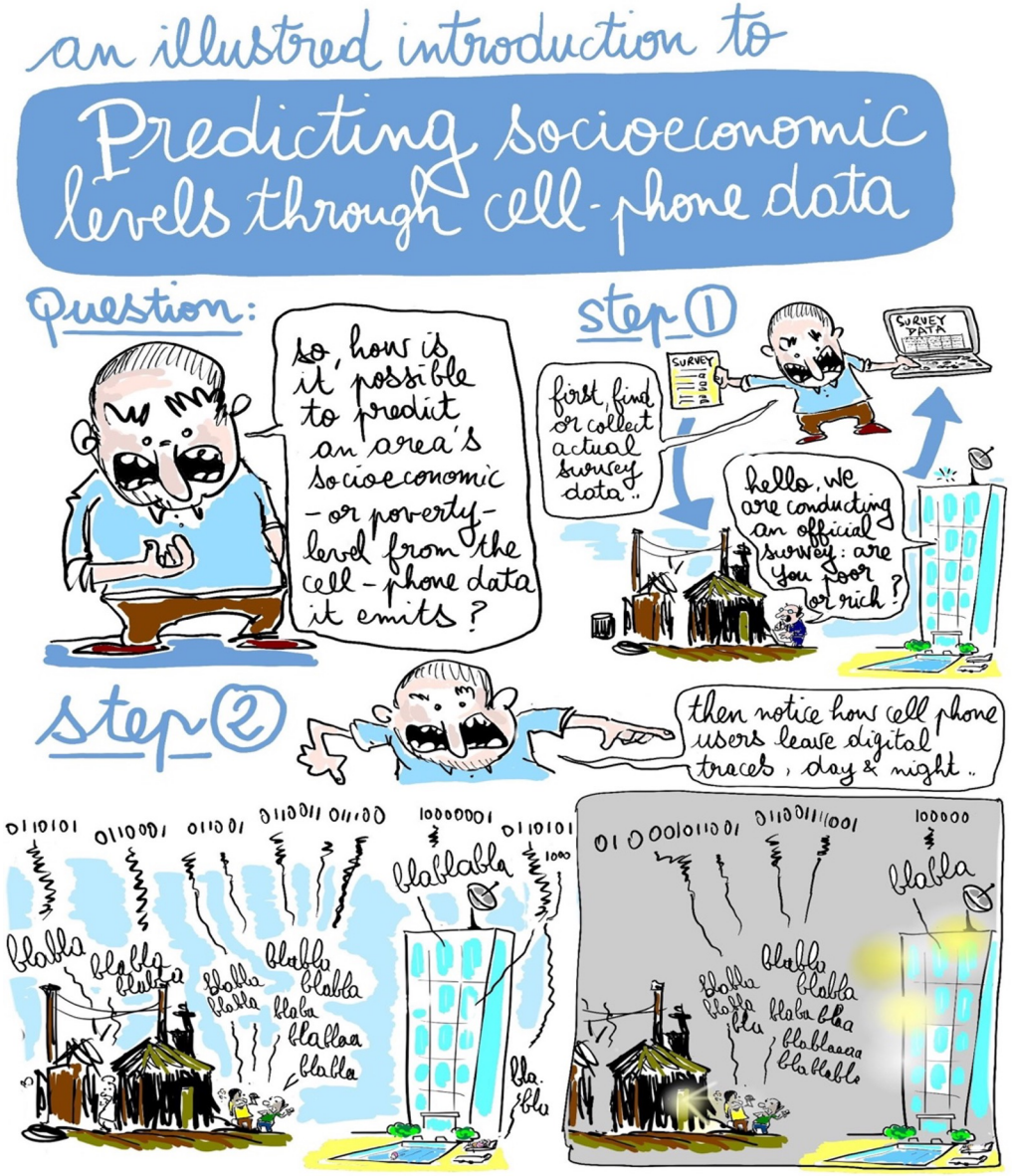
Outline

1. Introduction(s) and Thanks
2. Context, Concepts and Question(s)
3. **Promise and Opportunities**
4. Hurdles and Risks
5. The Way Forward: Requirements and Roadmap

Let's go 1 step further: "New" AI is based on machine learning from data (and rewards) through feedback loops

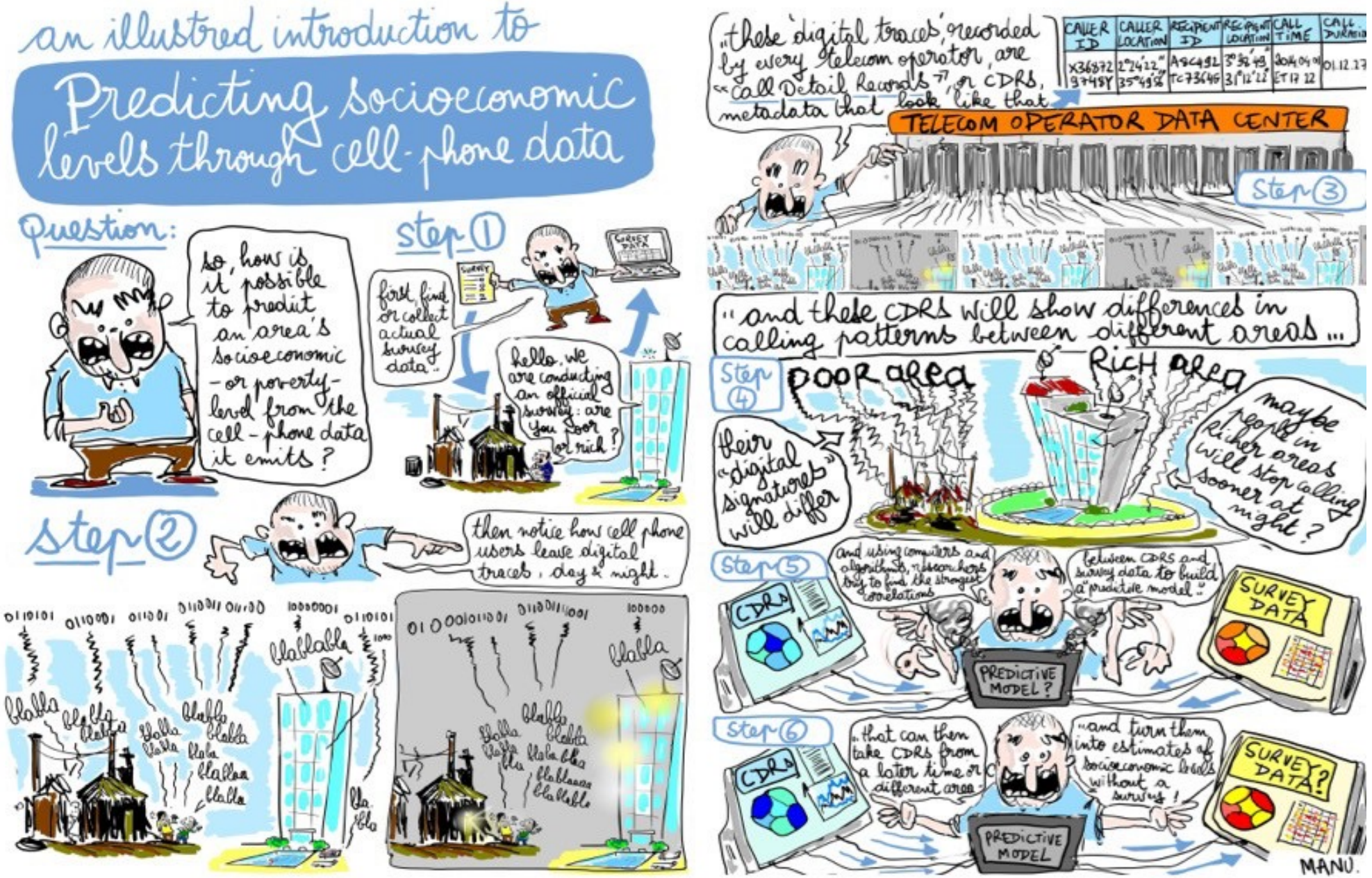




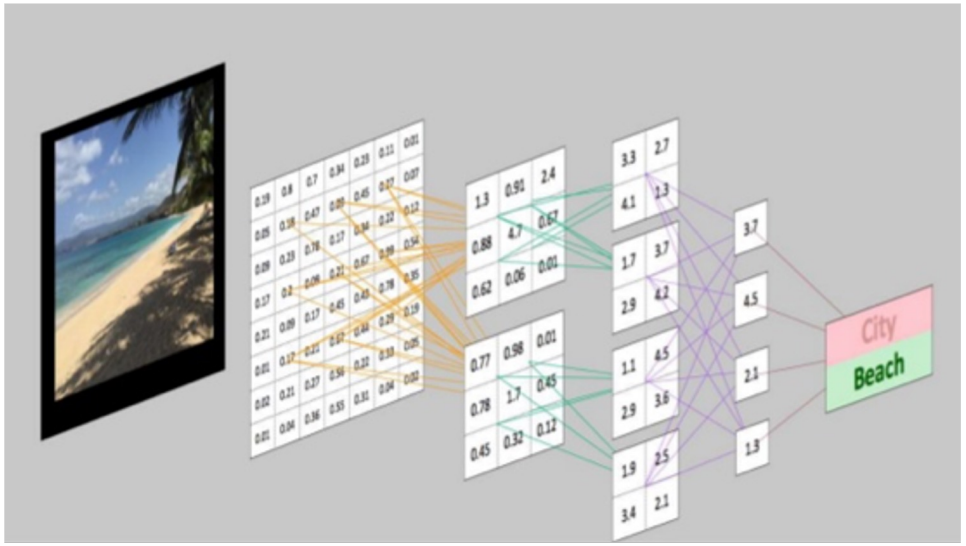


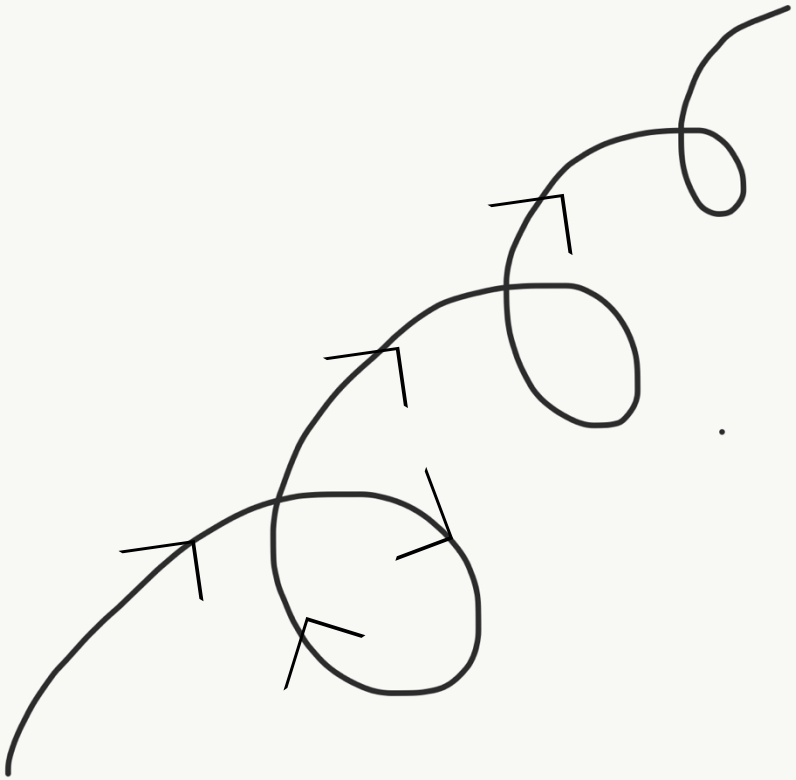
This is using AI as an *instrument* to do X, Y, Z better

(2013)



Another opportunity (an obligation) is a deeper cultural shift towards a "*Human AI*" where AI is used as an *inspiration*





Machine learning and human learning facilitated by data-infused feedback loops

Human AI is a vision where AI is used as an instrument ('Humane AI') + as an inspiration for learning societies

ITU Journal: *ICT Discoveries*, Special Issue No. 2, 6 Dec. 2018

TOWARDS A HUMAN ARTICIAL INTELLIGENCE FOR HUMAN DEVELOPMENT

Emmanuel Letouzé[1], Alex Pentland[2]
[1]Data-Pop Alliance, MIT Media Lab, and OPAL, [2]MIT and Data-Pop Alliance, and OPAL

Abstract *– This paper discusses the possibility of applying the key principles and tools of current artificial intelligence (AI) to design future human systems in ways that could make them more efficient, fair, responsive, and inclusive.*

3. Retrouver le sens de la mesure pour de meilleures décisions collectives : vers une intelligence artificielle humaine à l'ère des données

Emmanuel Letouzé

DANS REGARDS CROISÉS SUR L'ÉCONOMIE 2018/2 (n° 23), PAGES 47 À 55

CITER OU EXPORTER
AJOUTER À UNE LISTE
SUIVRE CETTE REVUE

ARTICLE | RÉSUMÉ | PLAN | AUTEUR | SUR UN SUJET PROCHE | ACHETER

Français Les données personnelles et institutionnelles récoltées de plus en plus massivement à l'échelle mondiale ouvrent la voie à de nouvelles façons de penser la prise de décision collective. Cet article présente les conditions d'un traitement de ces données tourné vers le développement humain, au travers du concept d'Intelligence Artificielle Humaine (IAH). Nous présentons à cet égard un exemple de projet destiné à appliquer ce concept dans plusieurs pays en développement.

unesco | Mila

AI FOR THE SDGS—AND BEYOND? TOWARDS A HUMAN AI CULTURE FOR DEVELOPMENT AND DEMOCRACY

EMMANUEL LETOUZÉ
Marie-Curie Fellow at Universitat Pompeu Fabra and Director and Co-Founder of the Data-Pop Alliance

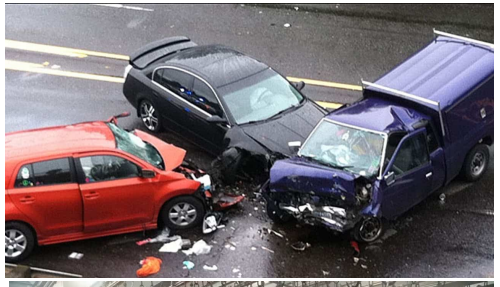
NURIA OLIVER
Scientific Director and Co-Founder of the ELLIS Unit Alicante Foundation and Chief Data Scientist at the Data-Pop Alliance

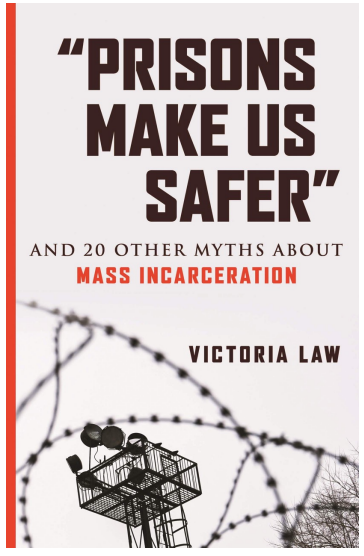
BRUNO LEPRI
Senior Researcher at Fondazione Bruno Kessler and Senior Research Affiliate at the Data-Pop Alliance

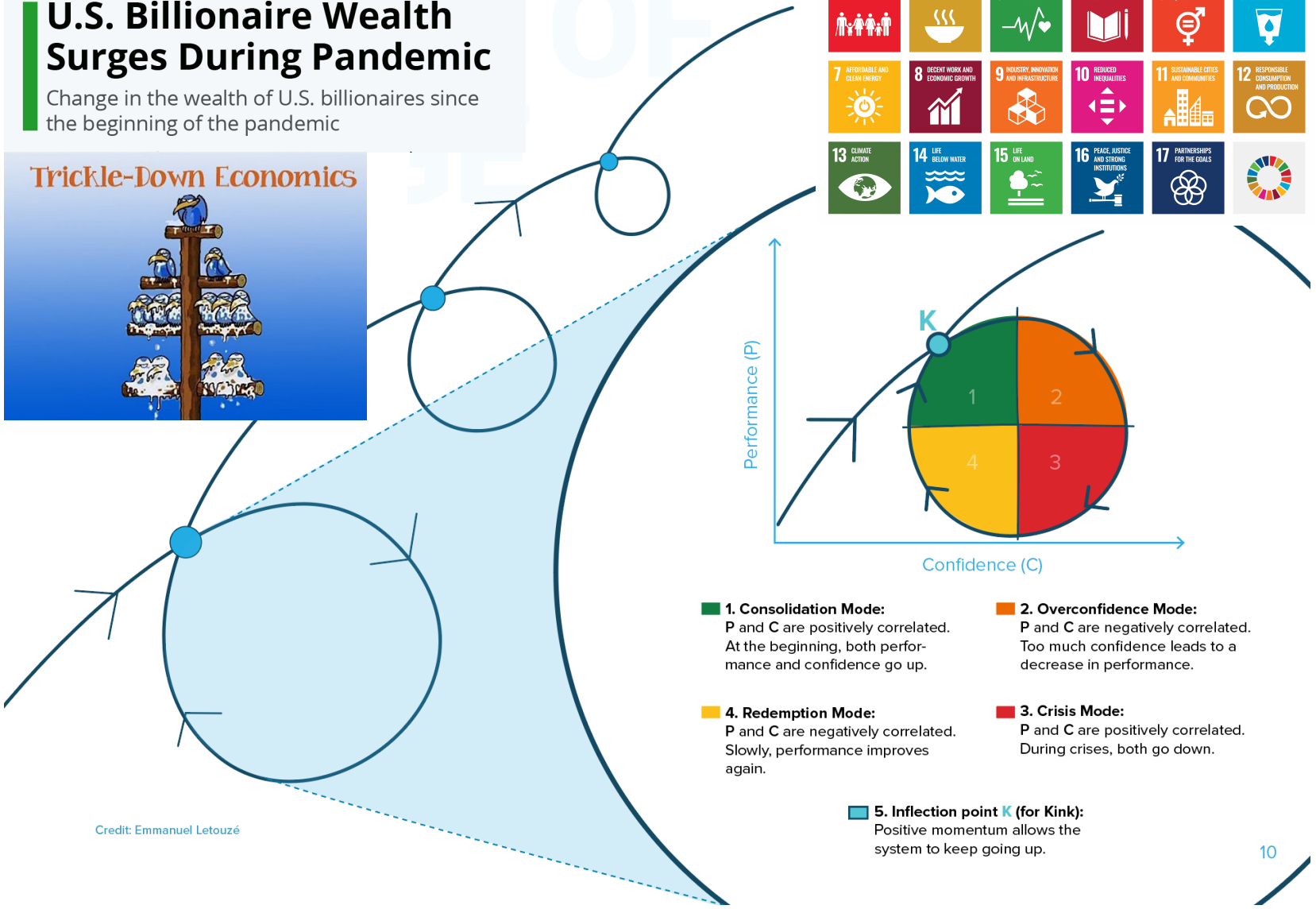
PATRICK VINCK
Assistant Professor at the Harvard Medical School and Harvard T.H. Chan School of Public Health, Co-Founder and Co-Director of the Data-Pop Alliance

SDG1 - No Poverty
SDG2 - Zero Hunger
SDG3 - Good Health and Wellbeing
SDG5 - Gender Equality
SDG8 - Decent Work and Economic Growth
SDG9 - Industry, Innovation and Infrastructure
SDG10 - Reduced Inequalities
SDG11 - Sustainable Cities and Communities
SDG12 - Responsible Consumption and Production
SDG16 - Peace, Justice and Strong Institutions
SDG17 - Partnerships for the Goals

Central to *Human AI* is the willingness and ability to learn and change from feedback provided by others and stats







So, what are the key ways in which Official Statistics can use AI to achieve its role for development and democracy?

1. Improve measurement of phenomenon that matter to societies

Better picture of human reality societies want to have; more timely, more granular, more disaggregated, more *relevant* too….== new indicators! Beyond GDP, social cohesion, happiness, corruption,

2. Improve and influence knowledge and trust among the the public

Better communicate, facilitate access (ChatGPT and other Chatbots)
Fight dis/misinformation to protect societies

3. Improve operations

More efficient / smarter official statistics; better allocation of resources during surveys for example; analyze systems performance

4. Upgrade culture, mindset and incentives for change among systems and staff

Change mindset towards innovation and impact not statistical production
Develop innovative projects and partnerships to retain high-skilled staff

Early works focused on CDRS have showed the promise

Scientific Prize and Ethics Mention: Construction of socio-demographic indicators with digital breadcrumbs

F. Bruckschen [1], T. Schmid [2], T. Zbiranski [1]

We show that socio-demographic indicators such as population, age, literacy, poverty, religion, ethnicity, electricity supply and others can be estimated in unprecedented detail and virtually ad-hoc using antenna-to antenna traffic data only. We offer a uniform approach that can be easily extended to other variables. Results are tested for spatio-temporal robustness and visualized as heat maps.

(1) Humboldt Universität Berlin, Germany - (2) Freie Universität Berlin, Germany

(2013)

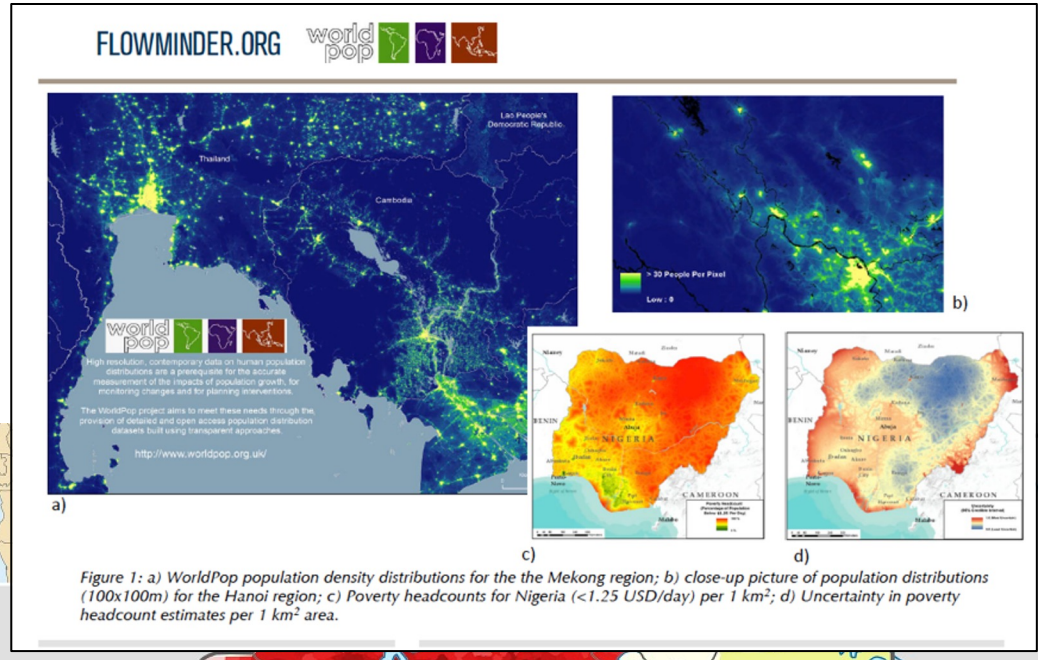


Figure 1: a) WorldPop population density distributions for the the Mekong region; b) close-up picture of population distributions (100x100m) for the Hanoi region; c) Poverty headcounts for Nigeria (<1.25 USD/day) per 1 km²; d) Uncertainty in poverty headcount estimates per 1 km² area.

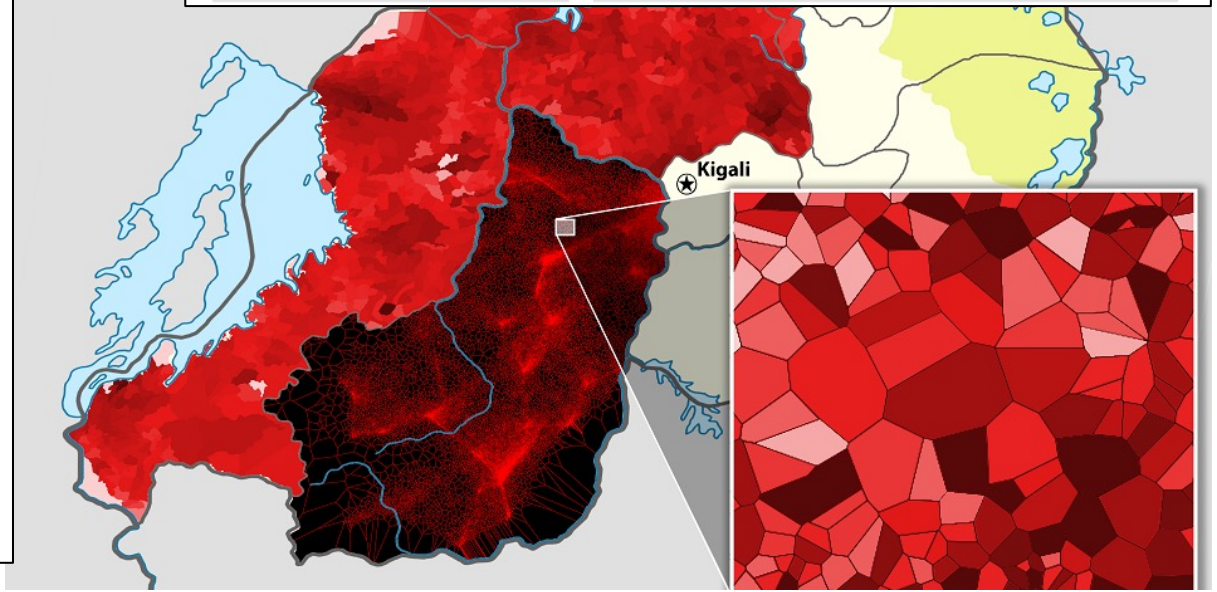
ECONOMICS

Predicting poverty and wealth from mobile phone metadata

(2014)

Joshua Blumenstock,[1]* Gabriel Cadamuro,[2] Robert On[3]

Accurate and timely estimates of population characteristics are a critical input to social and economic research and policy. In industrialized economies, novel sources of data are enabling new approaches to demographic profiling, but in developing countries, fewer sources of big data exist. We show that an individual's past history of mobile phone use can be used to infer his or her socioeconomic status. Furthermore, we demonstrate that the predicted attributes of millions of individuals can, in turn, accurately reconstruct the distribution of wealth of an entire nation or to infer the asset distribution of microregions composed of just a few households. In resource-constrained environments where censuses and household surveys are rare, this approach creates an option for gathering localized and timely information at a fraction of the cost of traditional methods.



EO data have also showed promise and are underused

Pilots in Brazil

Two Favelas:

- Favela dos Sonhos (Ferraz de Vasconcelos, São Paulo)
- Morro da Providência (Rio de Janeiro, Rio de Janeiro)

DATA-POP ALLIANCE
Changing the world with data

GERANDO FALCÕES

MIT MANAGEMENT
LATIN AMERICA OFFICE

Ferraz de Vasconcelos

Rio de Janeiro

10 REDUCED INEQUALITIES

11 SUSTAINABLE CITIES AND COMMUNITIES

New / sensitive questions should be (re)explored with AI

The Under-Representation of Women in a Disney Book from 1983

Using old data to understand how gender and racial norms have changed

Robert McKeon Aloe Nov 10, 2020 · 4 min read

5 GENDER EQUALITY

GENDER EQUALITY AND BIG DATA

GLOBAL PULSE UN WOMEN

ROYAUME DU MAROC
MINISTÈRE DE L'ECONOMIE ET DES FINANCES
DIRECTION DES ETUDES ET DES PREVISIONS FINANCIERES

DATA-POP ALLIANCE

Etude d'exploitation des données de Google Trends pour le suivi de l'indicateur 3.4.2 relatif au taux de mortalité par suicide au Maroc

3 BONNE SANTÉ ET BIEN-ÊTRE

Beyond GDP Sprint 2023

Updating Social Perspectives of Wellbeing and the Relevance of Past and Present UN Work for Today

1 June 2023, Virtual Meeting
7:00 a.m. – 9:30 a.m. New York Time

United Nations
DESA
Statistics Division

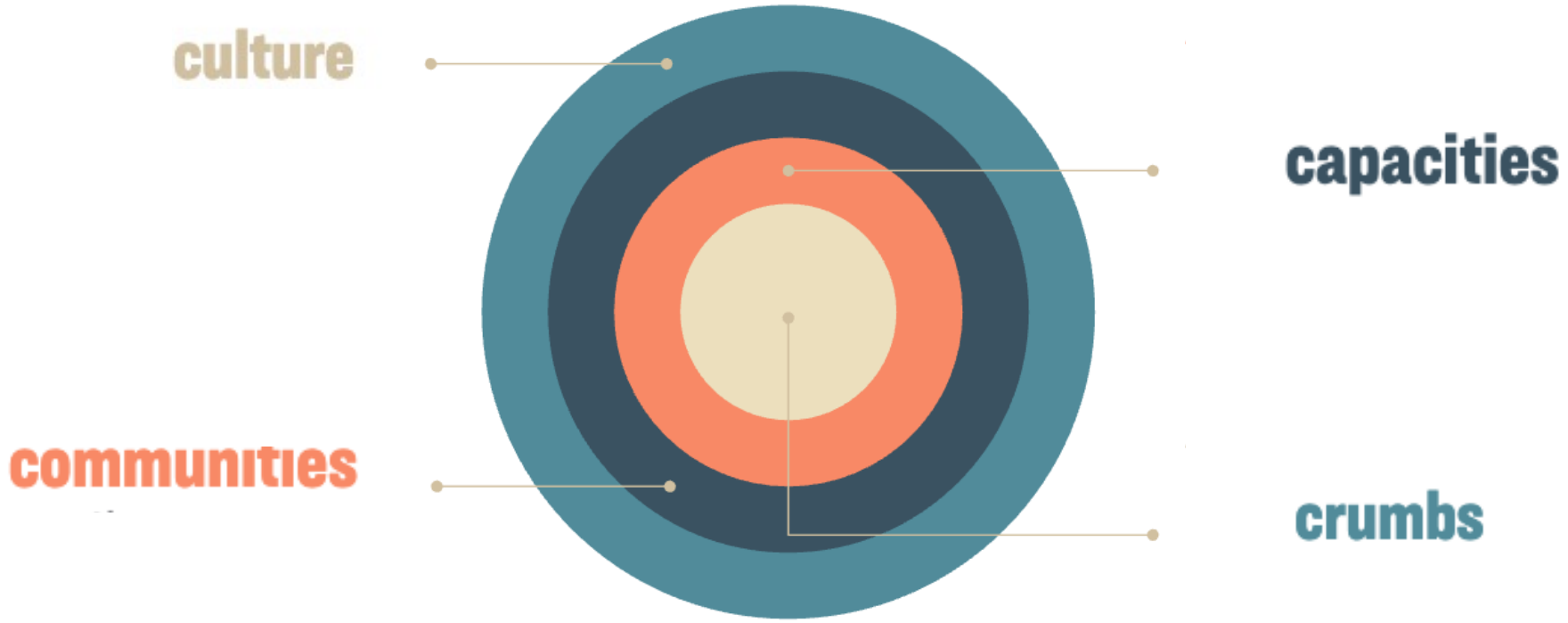
But we are stuck in an *MVP-PoC-Pilots-Promise trap*. Why?

Outline

1. Introduction(s) and Thanks
2. Context, Concepts and Question(s)
3. Promise and Opportunities
4. **Hurdles and Risks**
5. The Way Forward: Requirements and Roadmap

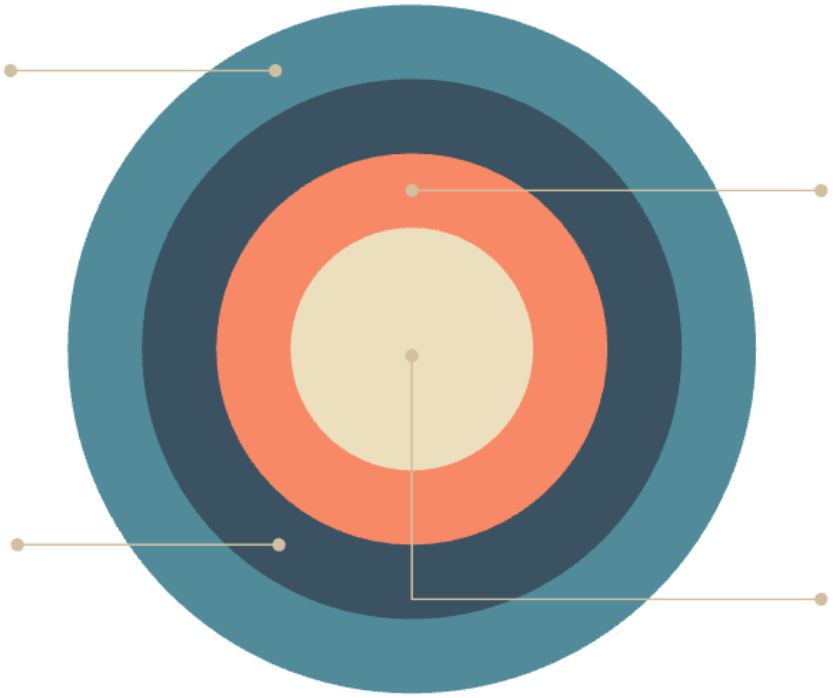
Hurdles and risks according to the 4 Cs framework

(N.B: classes of hurdles and risks for leveraging AI as an instrument and an inspiration)



Hurdles and risks according to the 4 Cs framework

(N.B: classes of hurdles and risks for leveraging AI as an instrument and an inspiration)



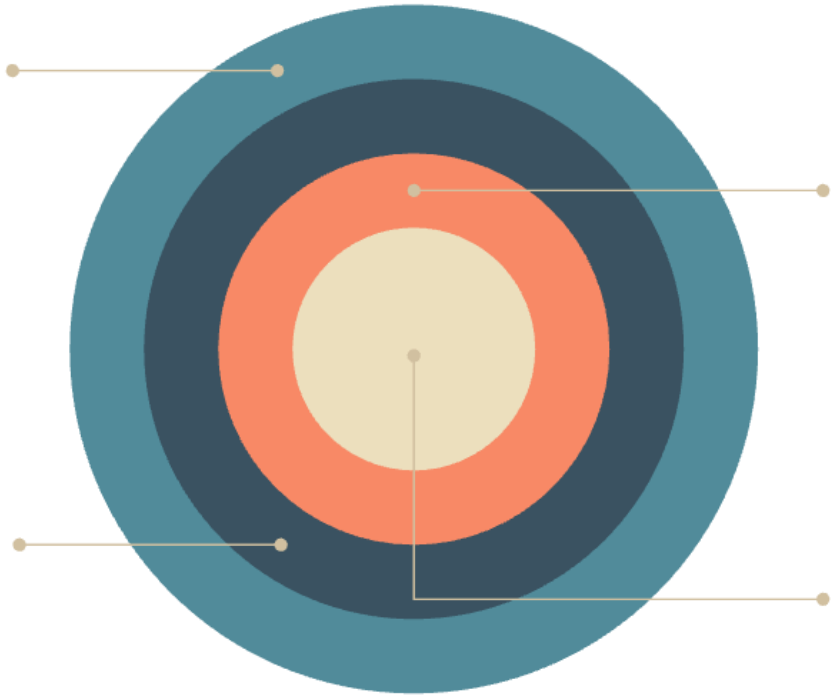
crumbs

Include:

Locked, messy, non-interoperable, hard to access;
plus dis-misinformation and deep fakes….

Hurdles and risks according to the 4 Cs framework

(N.B: classes of hurdles and risks for leveraging AI as an instrument and an inspiration)



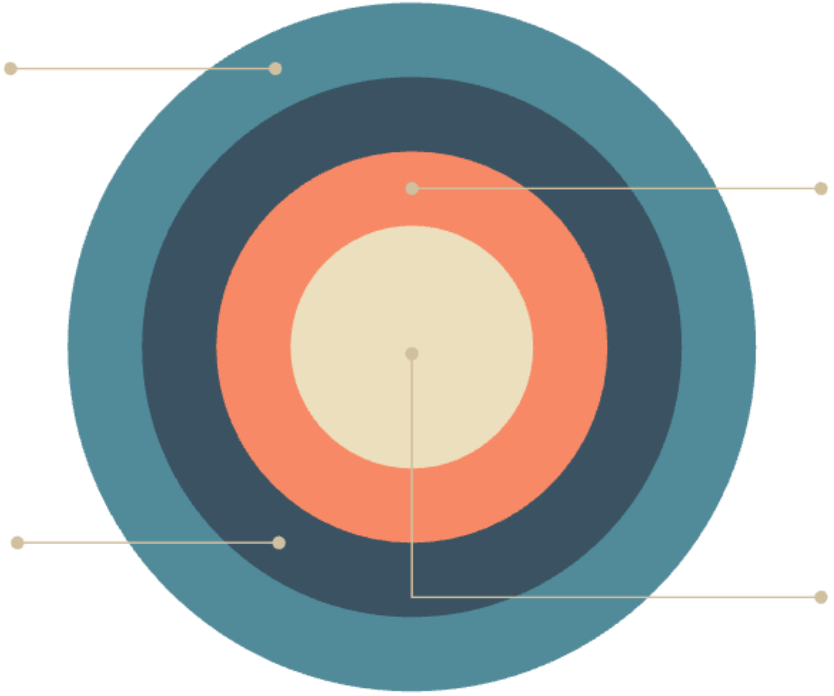
capacities

Include:

Major capacity and financial gaps and disparities; talent attraction and retention; infrastructure limitations…

Hurdles and risks according to the 4 Cs framework

(N.B: classes of hurdles and risks for leveraging AI as an instrument and an inspiration)



communities

Include:

Poor connections with industry and academia; Limited trust and interactions with users; Bad actors; Inadequate regulatory frameworks, no susainable business models….

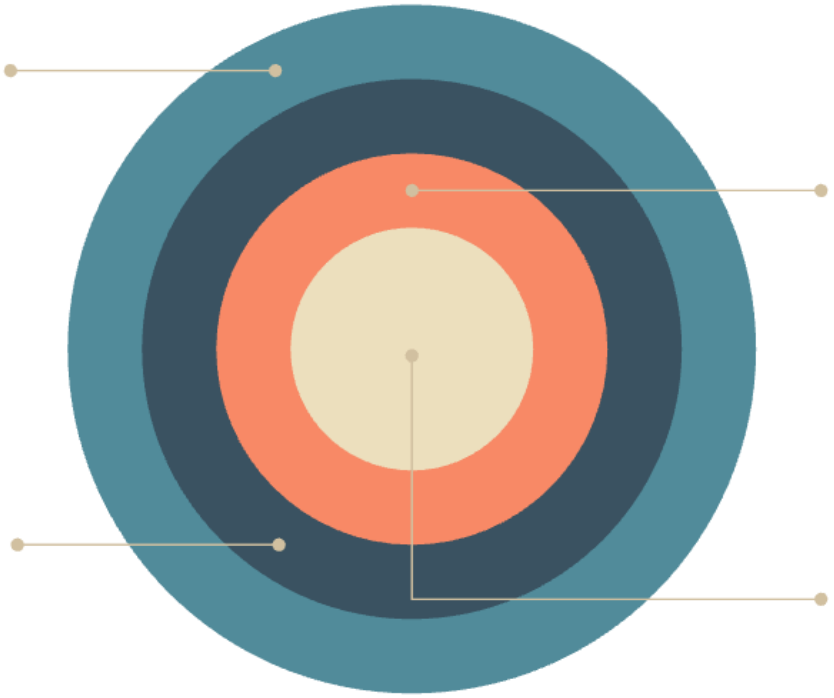
Hurdles and risks according to the 4 Cs framework

(N.B: classes of hurdles and risks for leveraging AI as an instrument and an inspiration)

culture

Include:

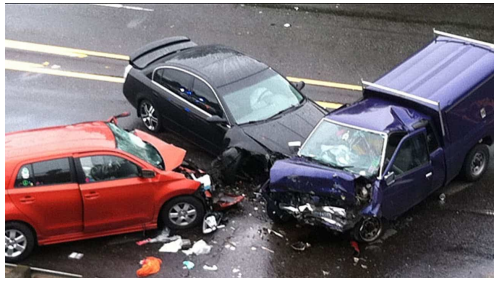
Focus on “statistical production” mindset, little to no incentive, time and funding think and do things differently at scale; culture of distruct for expertise…

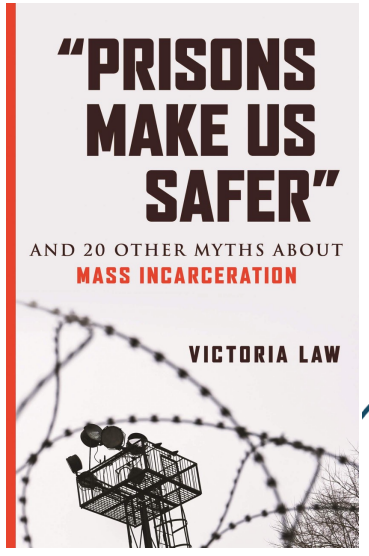


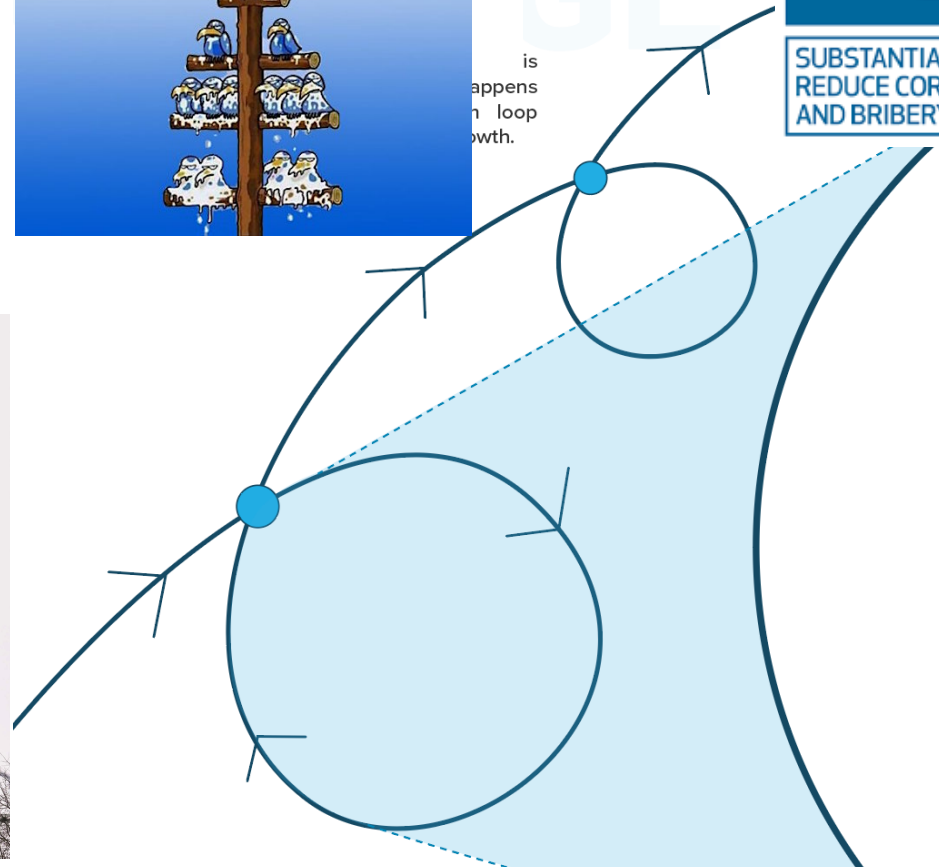
Outline

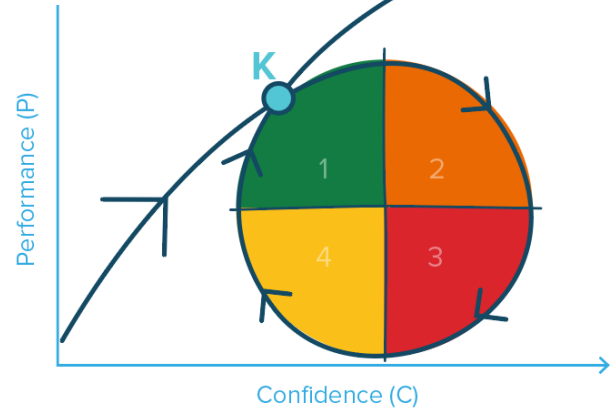
1. Introduction(s) and Thanks
2. Context, Concepts and Question(s)
3. Promise and Opportunities
4. Hurdles and Risks
5. **The Way Forward: Requirements and Roadmap**

Recall: key to *Human AI* is the willingness and ability to learn and change from feedback provided by statistics









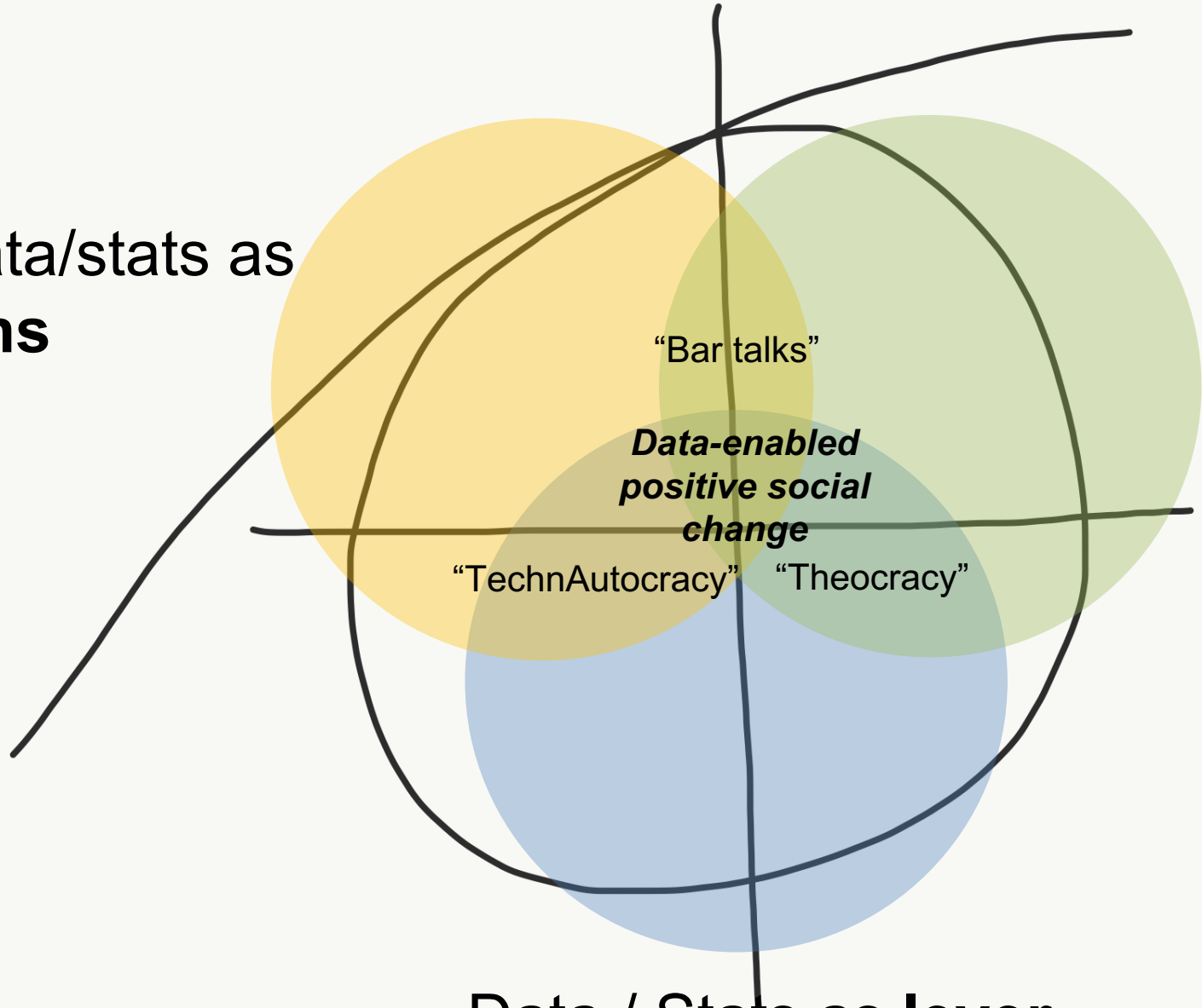
1. Consolidation Mode:
P and C are positively correlated. At the beginning, both performance and confidence go up.

2. Overconfidence Mode:
P and C are negatively correlated. Too much confidence leads to a decrease in performance.

4. Redemption Mode:
P and C are negatively correlated. Slowly, performance improves again.

3. Crisis Mode:
P and C are positively correlated. During crises, both go down.

3 functions of data and statistics enable this process



For this we need *socio-political innovations* for capacities, communities and culture to use AI towards humanistic values

DATA-POP ALLIANCE
WHITE PAPER SERIES

Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data

October 2015

Empowering society by reusing privately held data for official statistics — A European approach

FINAL REPORT PREPARED BY THE HIGH-LEVEL EXPERT GROUP ON FACILITATING THE USE OF NEW DATA SOURCES FOR OFFICIAL STATISTICS

2022 edition

BY STUART E. MIDDLETON, EMMANUEL LETOUZÉ, ALI HOSSAINI, AND ADRIANE CHAPMAN

Trust, Regulation, and Human-in-the-Loop AI within the European Region

Article | Published: 24 October 2018

The Moral Machine experiment

Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon & Iyad Rahwan

Nature **563**, 59–64 (2018) | Cite this article

Global Partnership for Sustainable Development Data

About | Countries | Partners | Resources | Initiatives | SDGs | News | Data Values

The Data Values Project

Data Values Project Blog →

Data as a route to inclusion and equity

The CODE for building participatory and ethical data projects

OpenAI

Research | Product | Developers | Safety

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.

OECD OPSI

Impact & Innovation Unit

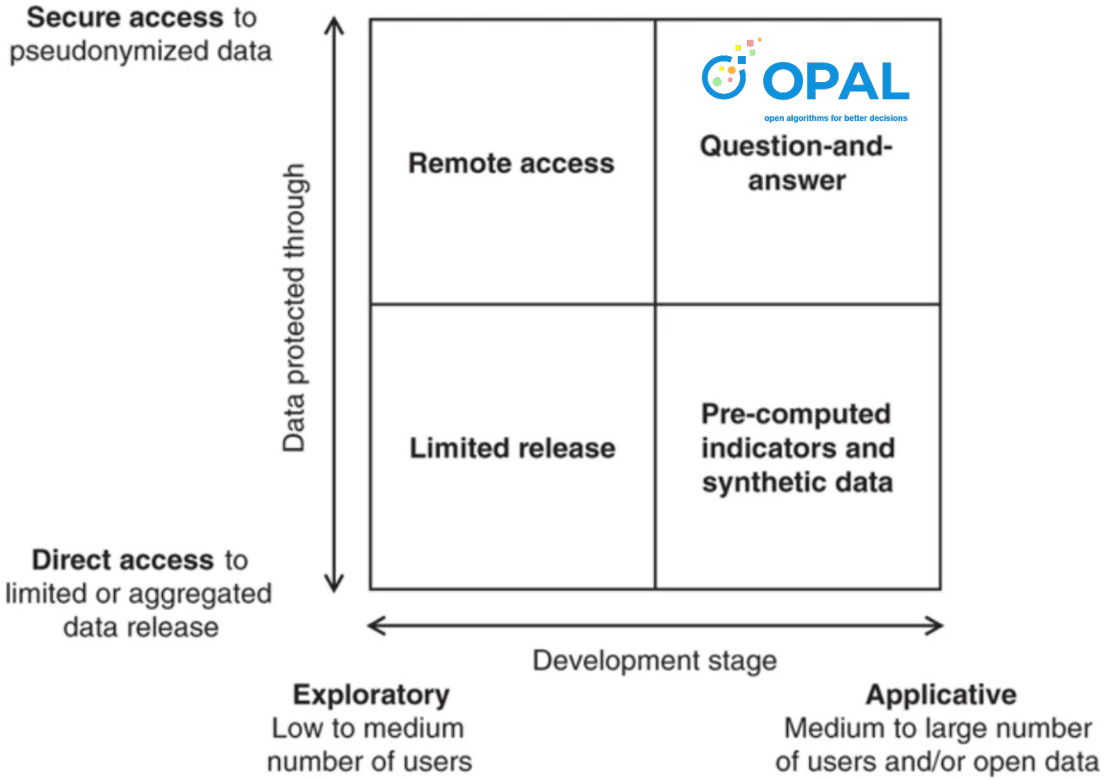
MINISTÈRE DE LA TRANSFORMATION ET DE LA FONCTION PUBLIQUES

Direction interministérielle de la transformation publique

We need to allow analysis of sensitive ‘Crumbs’ (both private, and public) in privacy-preserving and *participatory* ways at scale

Open algorithms: A new paradigm for using private data for social good

By Thomas Roca, Emmanuel Letouzé | 18 July 2016



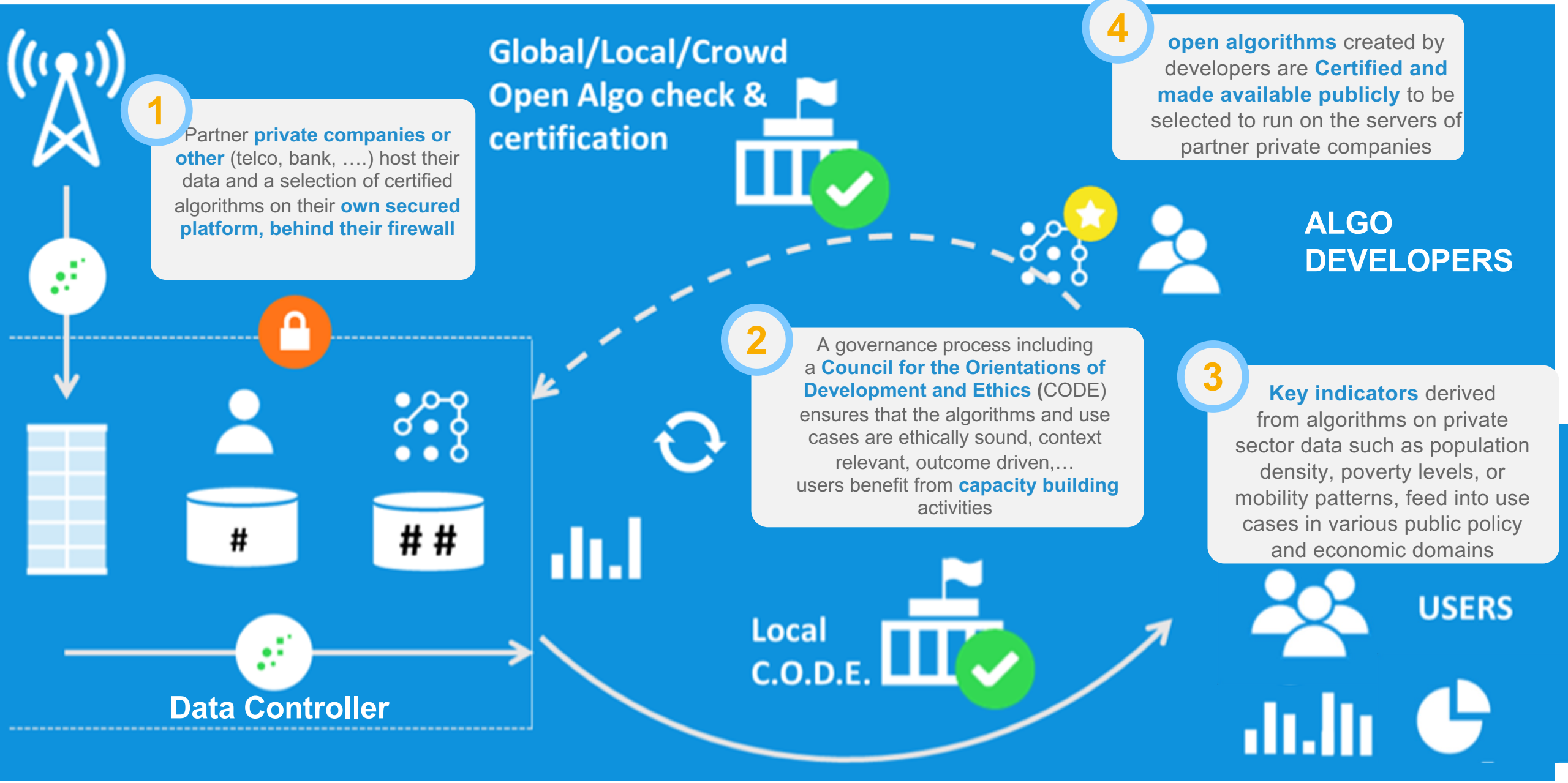
de Montjoye, Y., Gambs, S., Blondel, V. *et al.* On the privacy-conscientious use of mobile phone data. *Sci Data* 5, 180286 (2018). https://doi.org/10.1038/sdata.2018.286

Science & technology | Data privacy

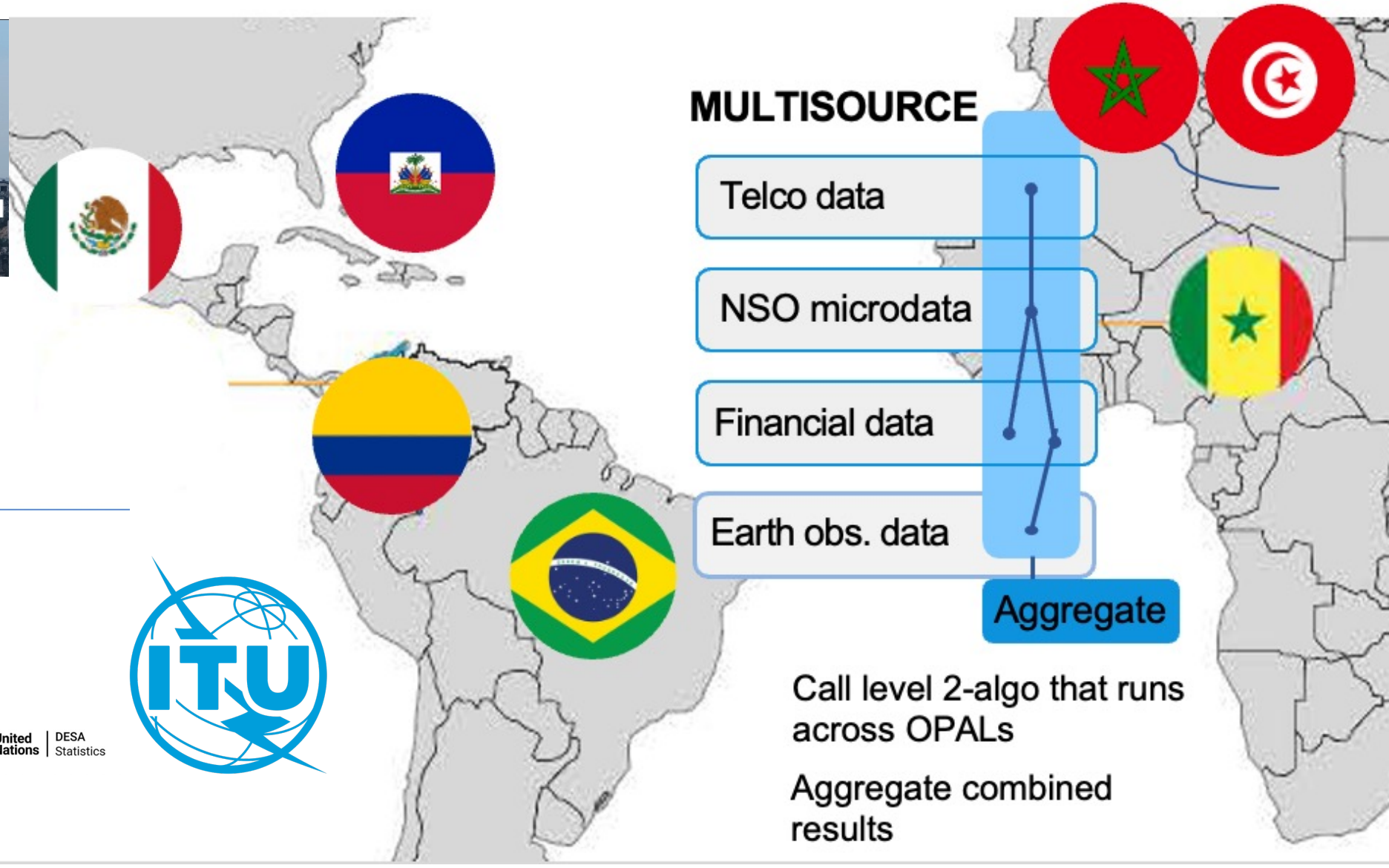
The UN is testing technology that processes data confidentially

How to analyse data without revealing their secrets

The Open Algorithms (OPAL) vision is 1 kind of *poly-innovation*



On DPA's roadmap: Multisource "OPAL for the SDGs" in Senegal, Morocco, Tunisia, Brazil, Mexico, Colombia, Haiti...



Other efforts too..

Data For Now › Data For Now in Colombia and Senegal

DATA FOR NOW

SUSTAINABLE DEVELOPMENT SOLUTIONS NETWORK
Global Partnership for Sustainable Development Data
THE WORLD BANK IBRD • IDA | WORLD BANK GROUP
United Nations | DESA Statistics

ITU

Conclusions

1. **The promises of the Data Revolution have not been met**; the world is not significantly better; Big Data and AI pilots, MVPs etc., have not scaled…this could go on with AI and backlash seriously
2. **That must change; AI offers a historical opportunity and a political-moral obligation for Official Statistics** as an industry to retain, regain and actually revive and reinforce its role as an enabler of societal progress through feedback loops
3. **If not, there are real political risks bad actors, bad statistics and dis-misinformation** will pollute societal debates and further fuel chaos
4. **In an "Human AI" vision, AI should serve as an instrument to improve statistical products and services, and as an inspiration for statistical systems and societies at large**, in which stats are key
5. **Achieving that vision requires massive investments, innovations and collaborations**—orders of magnitude more ambitious than what we are seeing and is projected. **Let's make it happen?**

Universitat Pompeu Fabra Barcelona

HARVARD HUMANITARIAN INITIATIVE

DATA-POP ALLIANCE

MIT Connection Science
the technology of innovation

Thank you

eletouze@datapopalliance.org