



AI and (Anti-)Corruption

Nils Köbis

Professor for Human Understanding of Algorithms and Machines

Research Center Trustworthy Data Science and Security

&

Center for Humans and Machines

Max Planck Institute for Human Development

New Hope in Artificial Intelligence



ABOUT PROJECTS RESEARCH MEDIA INSI

27 JULY 2018

From open data to artificial intelligence: the next frontier in anti-corruption

U4 Report 2019:1

Artificial Intelligence – a promising anti-corruption tool in development settings?

By Per Aarvik
Series editor: Arne Strand

CMI CHR.
MICHELSEN
INSTITUTE

Artificial Intelligence in the Public Sector

Maximizing Opportunities, Managing Risks

Supported by the GovTech Global Partnership: www.worldbank.org/govtech

GovTech
Putting people first

Federal Ministry
Republic of Austria
Finance

Ministry of Economy
and Finance
Republic of Korea

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs,
Education and Research EDAER
State Secretariat for Economic Affairs SECO

WORLD BANK GROUP

Reasons for Hope in AI

Challenges

1. **Autonomous** learning abilities

→ independently execute anti-corruption tasks, e.g. predict / detect corruption⁽¹⁻³⁾

2. **Computing** power

→ helps keeping track of complex corruption schemes (e.g. 4)

3. **Consistently impartial**

→ humans tend to be partial⁽⁵⁾; algorithms have no self-interest

1. **Data**

→ access & biases

2. **Algorithms**

→ false negative vs. false positive trade-offs

3. **Human**

→ paradox of control

(1) Lavigne, et al. (2017). *Preprint*; (2) López-Iturriaga & Sanz (2018) *Soc. Indic. Res*; (3) De Blasio, G. et al. (2020) *Preprint*; (4) Obermaier & Obermayer (2017) *The Panama Papers*. (5) Gerlach et al., (2018) *Psychological Bulletin*

Pitfalls and Potential of AI-ACT

Implementation of AI anti-corruption tools

Top-down

Governments & companies



Microsoft collaborates with the Nigerian government to accelerate digital transformation in the country



Is China's corruption-busting AI system 'Zero Trust' being turned off for being too efficient?



Pitfalls and Potential of AI-ACT

Implementation of AI anti-corruption tools

Top-down

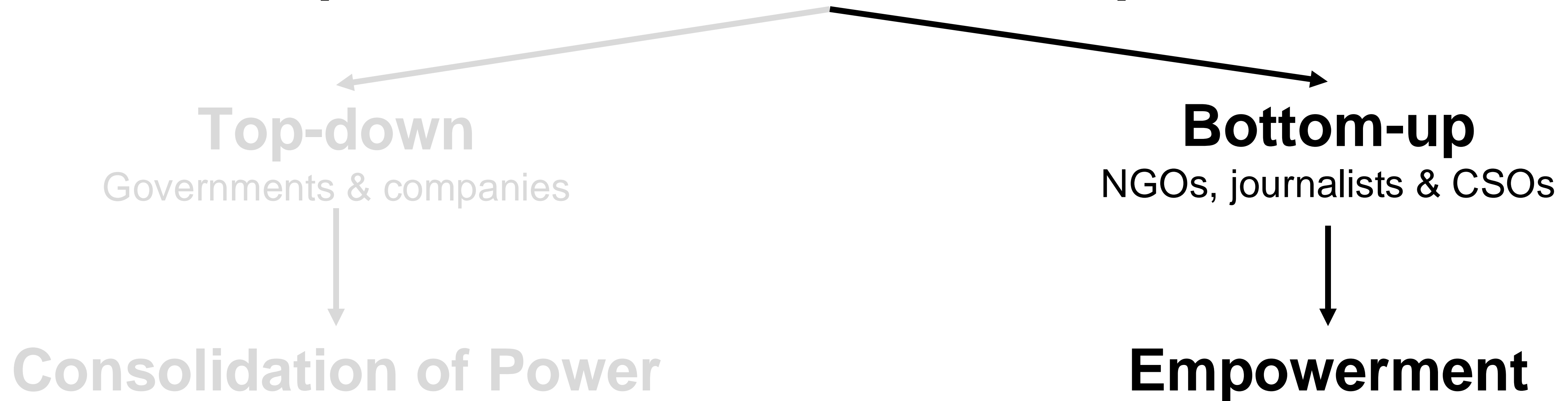
Governments & companies



Consolidation of Power

Pitfalls and Potential of AI-ACT

Implementation of AI anti-corruption tools



1. Access to **data** → donations & recording of relevant data
2. **Collective action** → AI for transparency, accountability & sustaining efforts
3. Coupling with other **technologies** → blockchain

Some Empirical Insights

Methodology

- **Qualitative Interviews** with society actors involved in AI-ACT projects from Europe and South America.

Key Potentials:

- **AI saves time** and costs by analyzing large datasets
- AI-ACTs **increase transparency** by making government actions visible and openly sharing data/code.
- **AI tools empower citizens to participate in AC efforts** through accessible data and direct action.

Key Limitations:

- **Access** to high-quality, machine-readable data is **often lacking** and varies by country.
- **Current laws hinder AI's optimal use**; new regulations needed for proactive AI application
- **Developing/maintaining AI systems is complex**, requiring expertise and addressing biased data

Pitfalls and Potential of AI-ACT

Implementation of AI anti-corruption tools

Top-down

Governments & companies

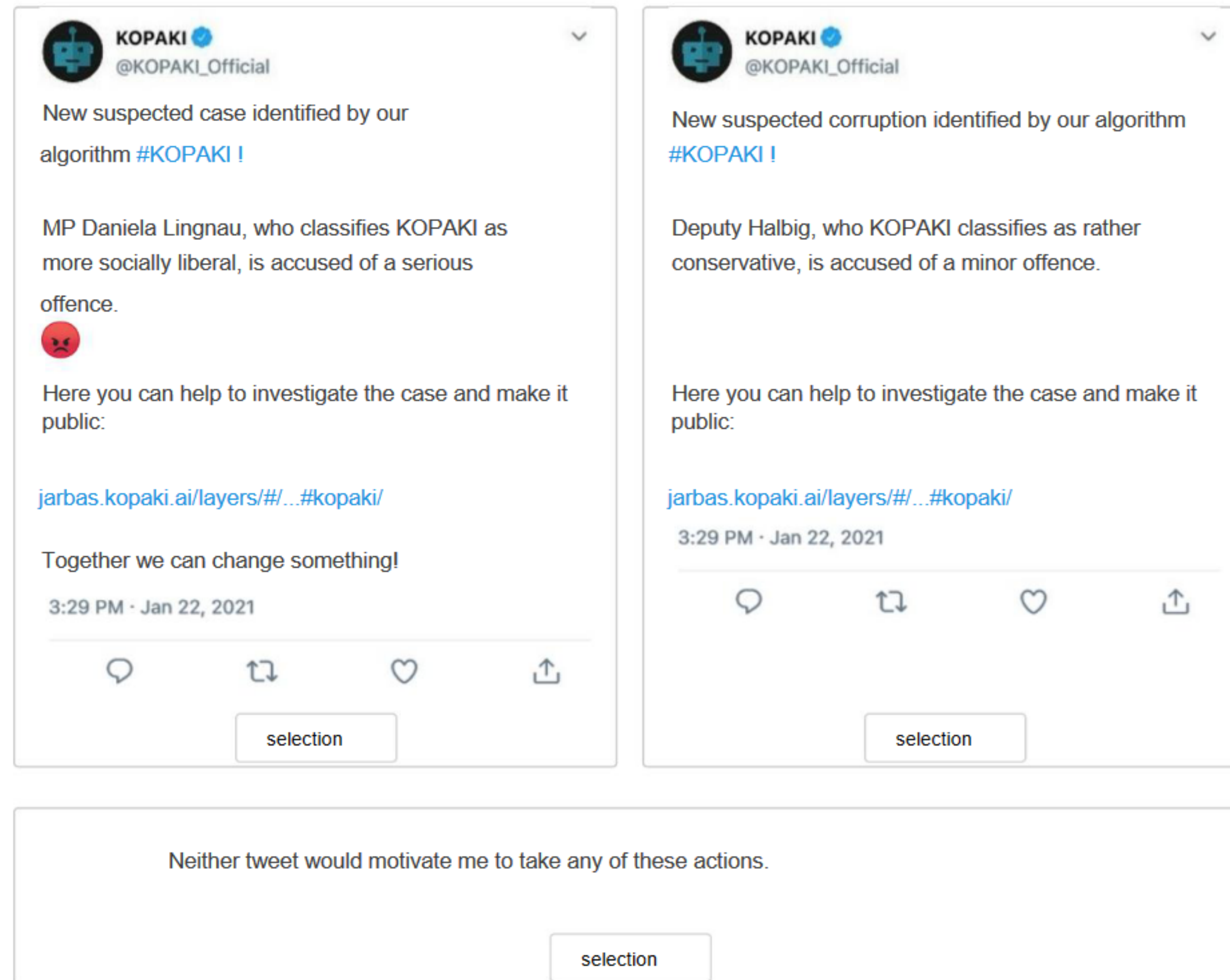
Bottom-up

NGOs, journalists & CSOs

Consolidation of Power



Some Empirical Insights



Study Design: pre-registered choice-based conjoint survey with 1,331 participants.

Tested message design features: type of injustice, degree of injustice, anger, political partisanship, gender, and efficacy cues.

Key Finding: Severe corruption cases most effectively mobilized people against corruption.

In-Group Favoritism: No support found for favoritism based on political affiliation and gender.

Audience Effects: Some design features had contrasting effects on different audiences.

Recommendation: More social science research needed alongside technical development of algorithmic tools to fight corruption.

Thank you!



Köbis, Starke, & Rahwan (2022) The promise and perils of using artificial intelligence to fight corruption. *Nature Machine Intelligence*, 4, 418–424

Forjan, J., Köbis, N., & Starke, C. (2024). Artificial intelligence as a weapon to fight corruption: Civil society actors on the benefits and risks of existing bottom up approaches. In *Digital Media and Grassroots Anti-Corruption* (pp. 229-249) Edward Elgar Publishing.



Starke, C., Kieslich, K., Reichert, M., & Köbis, N. (2023). Algorithms against Corruption: A Conjoint Study on Designing Automated Twitter Posts to Encourage Collective Action.

Köbis, Starke, & Edward-Gill (2022) The Corruption Risks of Artificial Intelligence. *Transparency International Policy Report*

