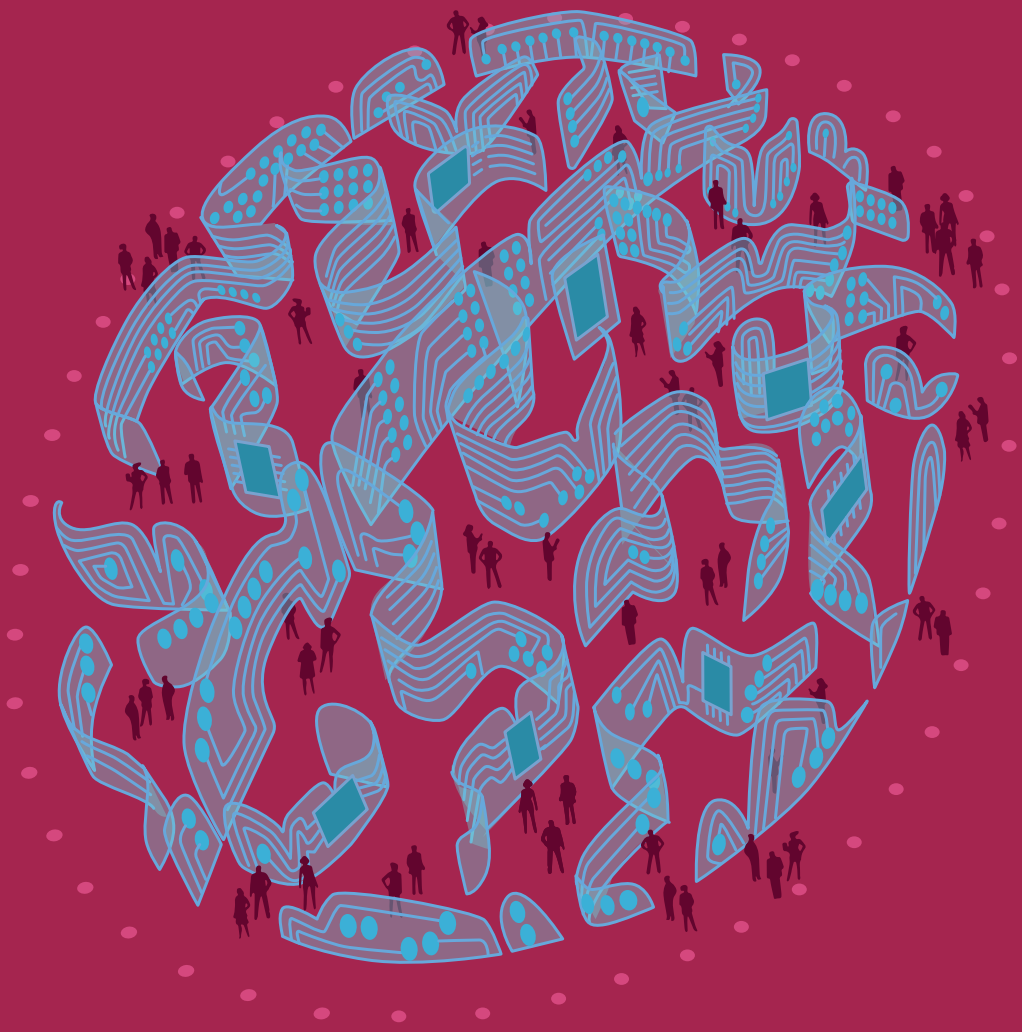
GLOBAL INFORMATION SOCIETY WATCH 2019

Artificial intelligence:
Human rights, social justice and development



Association for Progressive Communications (APC),
Article 19, and Swedish International Development Cooperation Agency (Sida)

Global Information Society Watch

2019

Global Information Society Watch 2019
Artificial intelligence: Human rights, social justice and development

Operational team
Valeria Betancourt (APC)
Alan Finlay (APC)
Mallory Knodel (ARTICLE 19)
Vidushi Marda (ARTICLE 19)
Maja Romano (APC)

Project coordination team
Valeria Betancourt (APC)
Cathy Chen (APC)
Flavia Fascendini (APC)
Alan Finlay (APC)
Mallory Knodel (ARTICLE 19)
Vidushi Marda (ARTICLE 19)
Leila Nachawati (APC)
Lori Nordstrom (APC)
Maja Romano (APC)

GISWatch 2019 advisory committee
Namita Aavriti (APC)
Rasha Abdul Rahim (Amnesty International)
Alex Comninos (Research ICT Africa)
Malavika Jayaram (Digital Asia Hub)
J. Carlos Lara (Derechos Digitales - América Latina)
Joy Liddicoat (Centre for Law and Emerging Technologies, University of Otago)
Andrew Lowenthal (EngageMedia)
Micaela Mantegna (Geekylegal/Machine Intelligence Lab, Center for Technology and Society, San Andres University)
Valeria Milanes (Asociación por los Derechos Civiles)

Project coordinator
Maja Romano (APC)

Editor
Alan Finlay (APC)

Assistant editor and proofreading
Lori Nordstrom (APC)

Publication production support
Cathy Chen (APC)

Graphic design
Monocromo

Cover illustration
Matías Bervejillo

We would like to extend a special note of thanks to a number of authors who have made ad honorem contributions to this edition of GISWatch.
We gratefully acknowledge the following:

Philip Dawson and Grace Abuhamad (Element AI)
Anita Gurumurthy and Nandini Chami (IT for Change)
Rasha Abdul Rahim (Amnesty International)

APC would like to thank the Swedish International Development Cooperation Agency (Sida) and ARTICLE 19
for their support for Global Information Society Watch 2019.

Published by APC
2019
Printed in USA

Global Information Society Watch 2019 web and e-book
ISBN 978-92-95113-13-8
APC Serial: APC-201910-CIPP-R-EN-DIGITAL-302

AI policing of people, streets and speech

Luis Fernando García Muñoz
R3D: Red en Defensa de los Derechos Digitales
www.r3d.mx

Introduction

An area in which artificial intelligence (AI) systems are producing a direct effect on the enjoyment of human rights today is the use of these systems for the alleged intention of protecting public safety and making justice systems more efficient and objective.

The rapid proliferation of these systems has, however, not only lacked a robust public discussion, but many of its impacts have not been evaluated before implementation. Therefore, it is crucial to review and examine the actual impacts of AI systems used for policing, surveillance and other forms of social control.

Predictive policing

Increasingly, law enforcement agencies have announced the use of AI with the purpose of predicting areas that are more prone to crime or even predicting which persons are more likely to be involved in a crime, both as perpetrators and as victims.[1] These predictions play an important role in decisions such as the deployment of police officers in those areas or a determination on the pre-trial detention of a suspect.

These tools rely on multiple sources of data such as criminal records, crime statistics, the demographics of people or neighbourhoods, and even information obtained from social media.[2]

As numerous reports have demonstrated, many of these data sets are flawed and biased in ways which can reinforce racial and other types of discrimination.[3] Moreover, predictions made by AI systems trained with skewed data are often seen as "neutral" or "objective", further ingraining discriminatory and abusive practices.

Often, predictive policing programmes are implemented without transparency, accountability or community participation in the decisions around their implementation[4] or in the evaluation and oversight of their impacts, further limiting the detection and remedy of undesired outcomes.

Social ranking

Some applications of AI systems are more straightforward in their repressiveness and authoritarianism. Take, for example, China's "social credit system", by which every person receives a score that factors in everyday behaviours, such as shopping habits or online opinions.[5] The score given is then used to determine access to services and jobs or may even prompt questioning or arrest by the police, thus influencing behaviour and social docility.

In some parts of China, massive amounts of data on each person, such as location data, data from ID cards, CCTV footage and even electricity consumption, are gathered, aggregated and processed to identify behaviour and characteristics deemed as suspicious by the state. This may also result in interrogation by the police, and even prolonged detention, often without any explanation given.[6]

1 Southerland, V. (2018, 9 April). With AI and criminal justice, the devil is in the data. *ACLU*. https://www.aclu.org/issues/privacy-technology/surveillance-technologies/ai-and-criminal-justice-devil-data

2 Buntin, J. (2013, October). Social Media Transforms the Way Chicago Fights Gang Violence. *Governing*. https://www.governing.com/topics/urban/gov-social-media-transforms-chicago-policing.html

3 See, for example, Richardson, R., Schultz, J., & Crawford, K. (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review Online, Forthcoming*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3333423 and Babuta, A., Oswald, M., & Rinik, C. (2018). *Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges*. London: Royal United Services Institute for Defence and Security Studies. https://rusi.org/sites/default/files/201809_whr_3-18_machine_learning_algorithms.pdf.pdf

4 Stanley, J. (2018, 15 March). New Orleans Program Offers Lessons In Pitfalls Of Predictive Policing. *ACLU*. https://www.aclu.org/blog/privacy-technology/new-orleans-program-offers-lessons-pitfalls-predictive-policing

5 Wang, M. (2017, 12 December). China's Chilling 'Social Credit' Blacklist. *Human Rights Watch*. https://www.hrw.org/news/2017/12/12/chinas-chilling-social-credit-blacklist

6 Wang, M. (2019, 1 May). China's Algorithms of Repression. *Human Rights Watch*. https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass-surveillance

Facial recognition surveillance

One of the most widespread and fast-growing applications of AI systems for policing is the use of facial recognition software for the surveillance of public spaces. The main capability of these systems is the identification of a person by comparing video images with existing databases, for example, mug shot, driver's licence or ID card databases. In the absence of clear video footage, even sketches or photos of celebrities described as having a resemblance to a suspect have been entered into the databases.[7]

Facial recognition software is usually used to analyse live video feeds captured by CCTV cameras, but it has been found to also be used to analyse recorded video footage. Some systems produce logs that register the historic detection of a person throughout a surveillance system, usually recording the location, time, date and relationships associated with each detection, and some systems claim to be able to even detect emotions such as happy, sad, calm, angry or surprised.[8]

The scale of this surveillance is unprecedented. For example, in the United States (US) it is estimated that approximately half of all residents are captured in the law enforcement facial recognition network. Also, the fact that this surveillance is difficult to escape, since it occupies public spaces, results in a particularly invasive tool with far-reaching consequences for participation in public life.

Facial recognition surveillance often lacks specific and robust regulation detailing the process and requirements to conduct a search through the system or establishing rules with regard to which individuals' faces can be included in the databases used and for how long, among other aspects. This has often led to serious abuse. For example, in the US county of Maricopa, Arizona, the complete driver's licence and mug shot databases of the country of Honduras were included in the database, which clearly indicates an intention to target a group of people with certain ethnic or national characteristics.

However, the potential for abuse is not limited to the arbitrary or discriminatory inclusion of databases in the system. There is a real risk that these tools are used by law enforcement to spy on people for reasons that have nothing to do with public safety. It has been reported that several law enforcement databases have been inappropriately accessed to spy on romantic partners, family members and journalists.[9]

The vulnerability of databases used by these systems adds an important layer of risk, particularly when the data that could be stolen is biometric. Differently from other types of data, like passwords, which can be modified if compromised, the effects of stolen biometric data are far more difficult to remediate. This risk has already materialised on multiple occasions. For example, in 2019, it was reported that the database of a contractor for the US Customs and Border Protection agency was breached, compromising photographs of travellers and licence plates.[10] Also in 2019, the fingerprints of over a million people, as well as facial recognition information, unencrypted usernames and passwords, and personal information of employees were discovered on a publicly accessible database for a company used by the Metropolitan Police, defence contractors and banks in the United Kingdom (UK).[11]

Additionally, facial recognition surveillance has been shown to be highly inaccurate. In the UK, an investigation revealed that implementations of the technology for certain events resulted in more than 90% of the matches being wrong.[12] The proneness of facial recognition surveillance to the misidentification of individuals has already resulted in the detention of innocent people[13] and produced a waste of law enforcement resources that could be allocated to more useful and adequate policing activities.

This technology has been shown to be particularly prone to misidentifying people of colour, women and non-binary individuals. For example, a study of three different kinds of facial analysis software demonstrated that while the error rate in determining the gender of light-skinned men was 0.8%, the error rate for darker-skinned women reached up to 34% in

7 Garvie, C. (2019, 16 May). Garbage in, garbage out: Face recognition on flawed data. *Georgetown Law Center on Privacy & Technology.* https://www.flawedfacedata.com

8 Dahua Technology. (2018). *AI Creates Value: Dahua AI Product & Solution Introduction.* https://www.dahuasecurity.com/asset/upload/download/20180327/2018_V1_Artificial-Intelligence%2820P%29.pdf

9 Gurman, S. (2016, 28 September). AP: Across US, police officers abuse confidential databases. *AP News.* https://apnews.com/699236946e3140659fff8a2362e16f43

10 Harwell, D., & Fowler, G. A. (2019, 10 June). U.S. Customs and Border Protection says photos of travelers were taken in a data breach. *The Washington Post*. https://www.washingtonpost.com/technology/2019/06/10/us-customs-border-protection-says-photos-travelers-into-out-country-were-recently-taken-data-breach

11 Taylor, J. (2019, 14 August). Major breach found in biometrics system used by banks, UK police and defence firms. *The Guardian.* https://www.theguardian.com/technology/2019/aug/14/major-breach-found-in-biometrics-system-used-by-banks-uk-police-and-defence-firms

12 Big Brother Watch. (2019, May). Face Off. https://bigbrotherwatch.org.uk/all-campaigns/face-off-campaign

13 Todo Noticias. (2019, 31 July). De un DNI mal cargado a una cara parecida: las víctimas del sistema de reconocimiento facial en Buenos Aires. *TN.* https://tn.com.ar/policiales/de-un-dni-mal-cargado-una-cara-parecida-las-victimas-del-sistema-de-reconocimiento-facial-en-buenos_980528

some cases.[14] This gender and racial bias creates an aggravated risk of perpetuating the discriminatory effects that policing and the criminal justice system have been found to be responsible for.

Despite the flaws and risks that facial recognition surveillance poses for the exercise of human rights, this technology is aggressively being pushed around the globe, including in countries with poor human rights records and a lack of robust institutional counterweights, which exacerbates the risk of abuse.

For example, facial recognition surveillance has been introduced or is already operating in Latin American countries like Argentina,[15] Brasil,[16] Chile,[17] Paraguay[18] and México[19] and African countries like Uganda, Kenya and Zimbabwe.[20] Besides the UK, facial recognition applications have been reported in Denmark[21] and Germany.[22]

Some jurisdictions are responding with regulations to limit the rapid proliferation of this technology. For example, it has been reported that the European Commission is preparing regulation[23] and the US cities of San Francisco,[24] Oakland[25] and Somerville[26] have all banned the police from using the technology. However, the vast majority of facial recognition systems remain unregulated and lack meaningful transparency and accountability mechanisms.

Impact on public protest

One strong concern about the use of AI for policing and surveillance of the public space is its impact on the exercise of the right to protest. This impact has recently become more evident, for example, in Hong Kong, where frequent protesting has encountered heavy resistance by the police. One of the tools that the Hong Kong police have used to try to thwart the protests has been the use of facial recognition cameras to attempt to identify the participants.[27]

Protesters in Hong Kong have resorted to multiple tactics to try to resist the heavy surveillance imposed on them – from using masks, certain kinds of makeup and umbrellas to try to cover their faces, to laser pointers aimed at obfuscating the operation of surveillance cameras, to even taking them down and destroying them.[28] The tension has prompted the Hong Kong government to use emergency powers to ban the use of masks[29] so facial recognition surveillance cameras are able to identify and track people participating in the protests. It is quite extraordinary that regulation on what people can wear is so strongly aimed at making an AI system work properly.

While often dismissed, privacy in public spaces is rapidly becoming more recognised as an essential value for the exercise of public protest. For example, the United Nations Human Rights Committee's (HRC) draft general comment on article 21 of the International Covenant on Civil and Political Rights (ICCPR) regarding the right of peaceful assembly[30] makes mention of the importance of the right to express your opinions anonymously, including in public spaces. It points out that even when "anonymous

14 Hardesty, L. (2018, 11 February). Study finds gender and skin-type bias in commercial artificial-intelligence systems. *MIT News.* https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212

15 Ucciferri, L. (2019, 23 May). #ConMiCaraNo: Reconocimiento facial en la Ciudad de Buenos Aires. *Asociación por los Derechos Civiles.* https://adc.org.ar/2019/05/23/con-mi-cara-no-reconocimiento-facial-en-la-ciudad-de-buenos-aires

16 Xinhua (2019, 2 March). Brasil estrena cámaras de reconocimiento facial coincidiendo con inicio del Carnaval. *Xinhua.* spanish.xinhuanet.com/2019-03/02/c_137862459.htm

17 Garay, V. (2018, 16 November). Sobre la ilegalidad de la implementación de un sistema de reconocimiento facial en Mall Plaza. *Derechos Digitales.* https://www.derechosdigitales.org/12623/sobre-la-ilegalidad-de-la-implementacion-de-un-sistema-de-reconocimiento-facial-en-mall-plaza

18 ABC Color. (2019, 11 July). Reconocimiento facial: nueva estrategia para combatir la delincuencia. *ABC Color.* https://www.abc.com.py/nacionales/2019/07/11/reconocimiento-facial-nueva-estrategia-para-combatir-la-delincuencia

19 R3D. (2019, 22 April). Gobierno de Coahuila anuncia compra de cámaras con reconocimiento facial. *Red en Defensa de los Derechos Digitales.* https://r3d.mx/2019/04/22/gobierno-de-coahuila-anuncia-compra-de-camaras-con-reconocimiento-facial

20 Wilson, T., & Murguía, M. (2019, 20 August). Uganda confirms use of Huawei facial recognition cameras. *Financial Times.* https://www.ft.com/content/e2058ode-c35f-11e9-a8e9-296ca66511c9

21 Mayhew, S. (2010, 1 July). Danish football stadium deploys Panasonic facial recognition to improve fan safety. *Biometric Update.* https://www.biometricupdate.com/201907/danish-football-stadium-deploys-panasonic-facial-recognition-to-improve-fan-safety

22 Delcker, J. (2018, 13 September). Big Brother in Berlin. *Politico.* https://www.politico.eu/article/berlin-big-brother-state-surveillance-facial-recognition-technology

23 Khan, M. (2019, 22 August). EU plans sweeping regulation of facial recognition. *Financial Times.* https://www.ft.com/content/90ce2dce-c413-11e9-a8e9-296ca66511c9

24 Conger, K., Fausset, R., & Kovaleski, S. (2019, 14 May). San Francisco Bans Facial Recognition Technology. *The New York Times.* https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html

25 Ravani, S. (2019, 17 July). Oakland bans use of facial recognition technology, citing bias concerns. *San Francisco Chronicle.* https://www.sfchronicle.com/bayarea/article/Oakland-bans-use-of-facial-recognition-14101253.php

26 Wu, S. (2019, 27 June). Somerville City Council passes facial recognition ban. *Boston Globe.* https://www.bostonglobe.com/metro/2019/06/27/somerville-city-council-passes-facial-recognition-ban/SfaqQ7mG3DGulXonBHSCYK/story.html

27 Mozur, P. (2019, 26 July). In Hong Kong Protests, Faces Become Weapons. *The New York Times.* https://www.nytimes.com/2019/07/26/technology/hong-kong-protests-facial-recognition-surveillance.html

28 Abacus. (2019, 30 August). Why are Hong Kong protesters targeting lamp posts? *South China Morning Post.* https://www.scmp.com/tech/big-tech/article/3024997/why-are-hong-kong-protesters-targeting-lamp-posts

29 Liu, N., Woodhouse, A., Hammond, G., & Meixler, E. (2019, 4 October). Hong Kong invokes emergency powers to ban face masks. *Financial Times.* https://www.ft.com/content/845056ca-e66a-11e9-9743-db5a370481bc

30 UN Human Rights Committee. (2019). *Draft General Comment No. 37 on Article 21 (Right of Peaceful Assembly) of the International Covenant on Civil and Political Rights.* https://www.ohchr.org/EN/HRBodies/CCPR/Pages/GCArticle21.aspx

participation and the wearing of face masks may present challenges to law enforcement agencies, for example by limiting their ability to identify those who engage in violence," masks or other mechanisms to hide the identity of participants in a protest "should not be the subject of a general ban."

The HRC further justifies the protection of anonymity in the context of a protest by noting that "concerns about identification may deter people with peaceful intentions from participation in demonstrations, or face masks could be part of the chosen form of expression."

It is in this context that the HRC recognises the importance of the protection of privacy in public places from technologies like facial recognition by stating that "the mere fact that participants in assemblies are out in public does not mean that their privacy cannot be infringed, for example, by facial recognition and other technologies that can identify individual participants in mass assemblies."

Content moderation

As online spaces increasingly become essential for deliberation and the formation of public opinion, the power wielded by the biggest internet platforms on deciding what can and cannot be expressed by the users of their services has become more and more relevant.

Increasing pressure for stricter content moderation, for example, with the aim of curbing copyright infringement, child pornography, incitement to violence and other categories of speech, has produced surging investment in the development of AI tools capable of detecting and removing infringing content.

While AI has been touted as a solution to the serious harms that content moderation produces for workers entrusted to carry out this task,[31] the risk of false positives and the increased obstacles for transparency and accountability pose a serious risk for freedom of expression online.

As UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression David Kaye mentioned in a report on the implications of AI technologies for human rights in the information environment, "AI-driven content moderation has several limitations, including the challenge of assessing context and taking into account widespread variation of language cues, meaning and linguistic and cultural particularities."[32] As a result, the use of AI for content moderation is susceptible to making many mistakes when removing content.

Increasing threats of regulation and sanctions for platforms that underperform in removing content deemed as infringing by regulators in different jurisdictions can also lead to incentives for overblocking as a means of protection against liability.

These risks become exacerbated by the difficulties in detecting the false positives that automated content removals create. As the special rapporteur points out, "AI makes it difficult to scrutinize the logic behind content actions." This is even more so the case when AI is expected to be used to moderate content as it is uploaded to the platforms,[33] without even allowing the content to be published, thus creating less awareness of the removal of content and adding even more opacity and difficulty to remediate errors or abuse caused by the content moderation systems.

The path forward

While AI should not be demonised as a technology, and many applications can contribute to social good, it is important to recognise the impacts that some applications can have on the exercise of human rights.

Policing, criminal justice systems and information flows are already flawed in complex ways, often reproducing systemic injustice against vulnerable groups.

Therefore, it is essential that AI is not deployed without regard of the context, the risks and the ways in which it can not only worsen the discrimination and violence against certain groups, but make these considerably more difficult to reverse.

Until the applications of AI for the attainment of security are informed by evidence, properly designed for human rights compliance and have multiple mechanisms to guarantee transparency and independent oversight, they should not be deployed at the accelerated pace that we see today.

Responsibility must prevail against the politically convenient idea of treating AI as a magical recourse to solve all problems real, perceived or artificially manufactured.

31 Newton, C. (2019, 25 February). The Trauma Floor: The secret lives of Facebook moderators in America. *The Verge*. https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona

32 United Nations General Assembly. (2018). *Report prepared by the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, on implications of artificial intelligence technologies for human rights in the information environment, focusing in particular on rights to freedom of opinion and expression, privacy and non-discrimination.* A/73/348. https://undocs.org/A/73/348

33 Porter, J. (2019, 21 March). Upload filters and one-hour takedowns: The EU's latest fight against terrorism online, explained. *The Verge*. https://www.theverge.com/2019/3/21/18274201/european-terrorist-content-regulation-extremist-terreg-upload-filter-one-hour-takedown-eu

GLOBAL INFORMATION SOCIETY WATCH 2019

Artificial intelligence:
Human rights, social justice and development

Artificial intelligence (AI) is now receiving unprecedented global attention as it finds widespread practical application in multiple spheres of activity. But what are the human rights, social justice and development implications of AI when used in areas such as health, education and social services, or in building "smart cities"? How does algorithmic decision making impact on marginalised people and the poor?

This edition of Global Information Society Watch (GISWatch) provides a perspective from the global South on the application of AI to our everyday lives. It includes 40 country reports from countries as diverse as Benin, Argentina, India, Russia and Ukraine, as well as three regional reports. These are framed by eight thematic reports dealing with topics such as data governance, food sovereignty, AI in the workplace, and so-called "killer robots".

While pointing to the positive use of AI to enable rights in ways that were not easily possible before, this edition of GISWatch highlights the real threats that we need to pay attention to if we are going to build an AI-embedded future that enables human dignity.

GLOBAL INFORMATION SOCIETY WATCH
2019 Report
www.GISWatch.org

APC
ARTICLE 19
Sida