

# 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 Comminos (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 Milanés (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

Creative Commons Attribution 4.0 International (CC BY 4.0)

<https://creativecommons.org/licenses/by/4.0/>

Some rights reserved.

Global Information Society Watch 2019 web and e-book

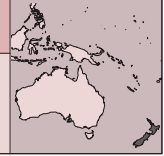
ISBN 978-92-95113-13-8

APC Serial: APC-201910-CIPP-R-EN-DIGITAL-302

Disclaimer: The views expressed herein do not necessarily represent those of Sida, ARTICLE 19, APC or its members.

# NEW ZEALAND

## ALGORITHMS AND SOCIAL MEDIA: A NEED FOR REGULATIONS TO CONTROL HARMFUL CONTENT?



**New Zealand Law Foundation Artificial Intelligence and Law Project, Centre for Law and Policy in Emerging Technologies, University of Otago**

Joy Liddicoat

<https://www.cs.otago.ac.nz/research/ai/AI-Law/index.html>

### Introduction

On 15 March 2019, a white supremacist committed a terrorist attack on two mosques in Christchurch, murdering 51 people as they were peacefully worshipping, injuring many others and live streaming the attack on Facebook. The attack was the worst of its kind in New Zealand's history and prompted an emotional nationwide outpouring of solidarity with Muslim communities. Our prime minister, Jacinda Ardern, moved quickly, travelling immediately to the Muslim communities affected, framing the attack as one on all New Zealanders, vowing compassion, refusing to ever say the name of the attacker, issuing a pledge to ban semi-automatic weapons of the kind used in the attack, and steering her people through a difficult emotional time of grief, anger and shock. The global response led Ardern and French President Emmanuel Macron to issue the #Christchurch Call,<sup>1</sup> calling for, among other things, an examination of the use of algorithms by social media platforms to identify and interfere with terrorist extremist online content. This country report critically examines the events, including discussion of technical measures to find and moderate the objectionable content. In doing so, it asks whether the multistakeholder model is failing in the sphere of social media internet-related policy.

### Country context

New Zealand is a small-islands country with a well-functioning democracy and stable economy. However, significant social inequalities exist among indigenous and some migrant populations, with children, young people and the elderly in particular economically, culturally and socially vulnerable in areas such as housing, health, education and income. Levels of internet access are generally high, there is good telecommunications infrastructure and an active information and communications technology sector.

The public discourse on artificial intelligence (AI) in New Zealand largely reflects that elsewhere, with an either utopian or dystopian view of how AI will affect humans, especially their employment. The positive benefits of using AI for service improvements is helping to change this, but most people still associate AI with robots and automated processing, rather than the machine learning or predictive algorithmic tools which are increasingly used in everyday life. More nuanced voices are emerging, but remain largely confined to a small range of actors mainly in the business, academic and government sectors.

Unlike other countries, there is no national AI strategy or research and development plan. However, while development, implementation and uptake of AI is patchy, it is growing. In government, for example, many departments are developing and implementing their own algorithms for a variety of service improvement purposes, rather than buying off-the-shelf products from third party foreign service providers.<sup>2</sup> In the private sector, a diverse range of businesses are developing AI-related products and services. There is one non-government AI organisation, the AI Forum, which was established in 2017 to bring together researchers, entrepreneurs, business and others to promote discussion and uptake of AI technologies.<sup>3</sup> In 2018, an inter-disciplinary initiative, the Centre for AI and Public Policy, was also established at the University of Otago.<sup>4</sup>

Few civil society voices offer critical analyses of AI issues, although there are pockets of activity such as in relation to harmful content, civic participation and social media use. Most of these voices appear to reflect the views of the small group of civil society actors commenting on internet policy and internet governance more generally. A small but growing number of community groups are developing their own machine-learning tools to deliver public information services, such as CitizenAI, which has developed the tools and is making them available on

<sup>1</sup> <https://www.christchurchcall.com/christchurch-call.pdf>

<sup>2</sup> Gavaghan, C., Knott, A., Maclaurin, J., Zerilli, J., & Liddicoat, J. (2019). *Government Use of Artificial Intelligence in New Zealand*. University of Otago and New Zealand Law Foundation. <https://www.cs.otago.ac.nz/research/ai/AI-Law/NZLF%20report.pdf>

<sup>3</sup> The AI Forum is also a member of the Partnership on AI. <https://aiforum.org.nz>

<sup>4</sup> <https://www.otago.ac.nz/caipp/index.html>

third-party platforms such as Facebook Messenger.<sup>5</sup> It was within this context that the #Christchurch Call was made, including the call for consideration of the use of algorithms by social media platforms.

## What can we learn from the #Christchurch Call?

### *Defining harmful content*

The prime minister was shocked that the attacker broadcasted the attack via Facebook's live-streaming service. The original footage was viewed 4,000 times before being removed by Facebook (the video was removed within 27 minutes of being uploaded). However, even within that short space of time the video had already been uploaded to 4chan, 8chan and other platforms. Within 24 hours the video had spread widely and 1.5 million copies of it had already been removed. There was one upload per second to YouTube alone within the first 24 hours. Like me, many people saw all or part of the video inadvertently, for example, as we followed news of the events on Twitter and had copies of the video posted in our feeds. This live streaming and sharing caused widespread public disgust and distress and there were demands that online platforms find and remove all copies of the recording.

A key question was the legal status of the video. The Films, Videos and Publications Classification Act 1993<sup>6</sup> regulates the distribution and sale of films, videos and other publications. The Act has a legal test relating to the harm that content might cause and established a chief censor with powers to give legal ratings to material (such as an age restriction, a parental guidance recommendation or a content warning).

The chief censor, David Shanks, viewed the attacker's video and ruled it was objectionable, within the legal meaning of the Act, considering its content was harmful if viewed by the public. The result was that distribution and possession of the video was a criminal offence. It is important to know that the chief censor did not ban the video. Instead, he ruled it could be made available to certain groups of people including experts, reporters and academics. Shanks considered that restricting distribution was a justifiable limitation pursuant to law under the New Zealand Bill of Rights Act 1990.<sup>7</sup> In doing so, he emphasised the video was not only a depiction of the attacks, but went further and promoted terrorism and murder:

There is an important distinction to be made between "hate speech", which may be rejected by many right-thinking people but which is legal to express, and this type of publication, which is deliberately constructed to inspire further murder and terrorism. It crosses the line.

Shanks considered the material was of a similar nature to ISIS terrorist promotion material that had also previously been ruled objectionable. However, he said it was in the public interest for some people to have access to the video for legitimate purposes, including education, analysis and in-depth reporting, and advised that reporters, researchers and academics could apply for an exemption to access and hold a copy. For similar reasons, he also ruled the attacker's manifesto was objectionable, finding that it was likely to be persuasive to its intended audience and promoted terrorism, including mass murder.<sup>8</sup>

The decision was upheld on review.<sup>9</sup> The effect of the ruling was retrospective (the publications were deemed objectionable from the date created) with the result that it is a criminal offence to possess them; people who had already downloaded the video, for example, were advised they should "destroy it."<sup>10</sup>

### *Dancing to the algorithms*

Having lawfully deemed the video to be objectionable, public debate turned to how the spread of the video could be halted. Some New Zealand internet service providers (ISPs) took their own initiative to block sites that were attempting to distribute the video on the day of the attacks.<sup>11</sup> These actions were criticised on the grounds that the blocking was not authorised according to law – the censor's ruling on the video was made three days after the attack – and that ISPs were wrong to take down the content before it had been declared objectionable.<sup>12</sup> The ISPs were certainly taking a risk in blocking content that had not been ruled unlawful. However, the ef-

5 <https://citizenai.nz>

6 [legislation.govt.nz/act/public/1993/0094/latest/DLM312895.html?src=qs](https://legislation.govt.nz/act/public/1993/0094/latest/DLM312895.html?src=qs)

7 Office of Film and Literature Classification. (2019, 23 March). Christchurch attacks classification information. <https://www.classificationoffice.govt.nz/news/latest-news/christchurch-attacks-press-releases>

8 The Censor did not impose tailored restrictions to allow journalists or researchers to access the manifesto. <https://www.classificationoffice.govt.nz/news/featured-classification-decisions/the-great-replacement>

9 *Johnson v Office of Film and Literature Classification*, Film and Literature Review Board, Wellington, 14 April 2019.

10 See Office of Film and Literature Classification. (2019, 23 March). Op. cit. Several people have already been convicted of possessing the objectionable material.

11 <https://twitter.com/simonmoutter/status/1106418640167952385>

12 See, for example, Free Speech Coalition. (2019, 25 March). Christchurch and Free Speech. [https://www.freespeechcoalition.nz/christchurch\\_and\\_free\\_speech](https://www.freespeechcoalition.nz/christchurch_and_free_speech) and Chen, C. (2019, 18 March). ISPs in AU and NZ censoring content without legal precedent. *Privacy News Online*. <https://www.privateinternetaccess.com/blog/2019/03/isps-in-au-and-nz-start-censoring-the-internet-without-legal-precedent>

fect of the censor's ruling was that the video was deemed objectionable *at the time it was made* and having retrospective effect, so that ISPs were not acting unlawfully in blocking access to the content on the day of the attacks. Had the censor ruled the video was not objectionable, the criticisms might have been more valid.

At the domain name service (DNS) level, the Domain Name Commissioner issued a statement saying that if necessary to protect the security of the .nz ccTLD and the DNS, the commissioner may suspend a registered domain name on the request of the government computer emergency response team or the Department of Internal Affairs.<sup>13</sup>

The prime minister sought to steer a balanced path between affirming internet openness on the one hand and human rights – including religious freedom – on the other. In her address to parliament four days after the attack, Ardern said:

There is no question that [the] ideas and language of division and hate have existed for decades, but their form of distribution, the tools of organisation, they are new. We cannot simply sit back and accept that these platforms just exist and that what is said on them is not the responsibility of the place where they are published. They are the publisher. Not just the postman. There cannot be a case of all profit, no responsibility.

In preparing for a subsequent Paris summit on the #Christchurch Call in May 2019, Ardern made clear that her focus was on the harm of online terrorist extremist content, saying the “task here is to find ways to protect the freedom of the internet and its power to do good, while working together to find ways to end its use for terrorism.”<sup>14</sup> In particular, she said: “We ask that you assess how your algorithms funnel people to extremist content and make transparent that work.”

The leader of the opposition political party also weighed in, saying: “It’s smart algorithms on the internet traffic into New Zealand that allow you to lawfully target, whether it’s white supremacists or whatever those extremists are. I think we were overly cautious, I think we need to revisit that.”<sup>15</sup>

Some commentators agreed with the prime minister that social media platforms can no longer say they are not publishers when their own algorithms enable and drive content sharing, while others said this was simply a new take on an old debate.<sup>16</sup> The risks of relying on algorithms to filter content are already known to be fraught. In New Zealand, for example, Good Bitches Baking is a not-for-profit network which shares home baking with people going through a difficult time.<sup>17</sup> However, whenever they attempt to post on Facebook they are blocked because of their name.

Many pointed to the futility of trying to chase copies of the terrorist attack video, likening this to playing “whack-a-mole” because the content would be constantly appearing elsewhere on the internet. The technical difficulty in identifying “copies” was highlighted, as small changes could be made to the video (such as editing to add material) making it difficult to clearly identify the relevant content. Concerns were raised about the collateral damage to legitimate and legal content on platforms. Some members of the Muslim community wanted to see the video (for example, to see if family members had survived) and some had in fact already watched it. There were concerns that prohibiting the video would drive it to the so-called dark web, thereby embedding the harm it was sought to avoid. Finally, some considered that there will “always be harmful content on the internet, outside of anyone’s control” and there was little point in trying to contain this particular video.<sup>18</sup>

However, others pointed out that these arguments were fallacious, since algorithms were already being used to curate and feed content and could therefore be redesigned: there may be ways to interfere with recommendation algorithms to prevent the development of filter bubbles that channelled users to extremist content. Others decried Facebook’s failure to implement its own community standards and its initial silence after the attack.<sup>19</sup> Another view was that algorithms feeding the objectionable video content to those who did not want to see it was an interference with their right to privacy (their right to be let alone and to decide for themselves what they wished to view).

13 Carey, B. (2019, 29 March). Emergency Response to the Christchurch Terrorist Attacks. *Domain Name Commission*. <https://www.dnc.org.nz/christchurchterroristattackresponse>

14 Ardern, J. (2019, 16 May). Christchurch Call opening statement. <https://www.beehive.govt.nz/speech/jacinda-ardern%E2%80%99s-christchurch-call-opening-statement>

15 Simon Bridges of the National Party. See: [https://www.tvnz.co.nz/one-news/new-zealand/simon-bridges-calls-tougher-cyber-security-laws-in-wake-christchurch-terror-attacks?variant=tb\\_v\\_1](https://www.tvnz.co.nz/one-news/new-zealand/simon-bridges-calls-tougher-cyber-security-laws-in-wake-christchurch-terror-attacks?variant=tb_v_1)

16 Brown, R. (2019, 12 April). This is not the internet you promised us. *The Spinoff*. <https://thespinoff.co.nz/partner/actionstation/12-04-2019/this-is-not-the-internet-you-promised-us>

17 <https://www.gbb.org.nz>

18 Moskovitz, D. (2019, 8 April). Publisher or postman? <https://dave.moskovitz.co.nz/tag/freedom-of-speech>

19 Manhire, T. (2019, 19 May). Mark Zuckerberg, four days on, your silence on Christchurch is deafening. *The Spinoff*. <https://thespinoff.co.nz/society/19-03-2019/mark-zuckerberg-four-days-on-your-silence-on-christchurch-is-deafening>

The outcome of the #Christchurch Call included a set of voluntary commitments by governments and online service providers “intended to address the issue of terrorist and violent extremist content online.”<sup>20</sup> Among the range of measures that online service providers agreed on was to:

Review the operation of algorithms and other processes that may drive users towards and/or amplify terrorist and violent extremist content to better understand possible intervention points and to implement changes where this occurs. This may include using algorithms and other processes to redirect users from such content or the promotion of credible, positive alternatives or counter-narratives. This may include building appropriate mechanisms for reporting, designed in a multi-stakeholder process and without compromising trade secrets or the effectiveness of service providers’ practices through unnecessary disclosure.

The topic of algorithms to promote and share content, as well as to find and limit its spread, was now squarely in the public domain – a considerable step forward in the discourse on AI. A small number of New Zealand groups have already responded to the outcome of the #Christchurch Call, urging more research with an interdisciplinary approach. The AI Forum’s Ethics, Law and Society Working Group, for example, noted the three ways content is disseminated on the internet (user upload, internet searches and social media feeds), pointing out that implementing the #Christchurch Call could involve filtering at some point in each of these processes. Because some of these processes would have to be automatic, the challenge would be to identify items to be filtered in an AI classification system which can determine whether each item would be allowed or blocked.<sup>21</sup>

The group identified technical challenges to accurately identifying the relevant content, such as choosing the best classifiers, getting classification consistently correct, how to deal with errors, and whether classifiers should err on the side of accepting or blocking content. This in turn, the group said, raised ethical questions about freedom of expression and censorship and economic questions about the cost of running different types of filtering

systems.<sup>22</sup> Despite these difficulties, the working group considered that “small changes to feed recommendation algorithms could potentially have large effects – not only in curbing the transmission of extremist material, but also in reducing the ‘filter bubbles’ that can channel users towards extremist political positions.”

### *Other community responses*

Communities responded to the attacks in a variety of ways. Many took up the prime minister’s call to deny the attacker the infamy he sought by refusing to use his name and by not sharing any photos of him. For example, telecommunications company Spark called on people to support “a #ShareNoEvil movement that could help deprive terrorists of the fame and oxygen their evil needs to survive” and with the explicit ambition of “making the act of sharing terrorist content culturally unacceptable in Aotearoa.” The campaign enabled supporters to download a Google Chrome extension that lets users block the attacker’s name and replace it with the words “Share no evil”.<sup>23</sup>

Muslim women leaders called out the community on the racism that they face, speaking of the efforts they made to alert the government, including police, to the harassment they were experiencing from white supremacists and other right wing groups. Islamic Women’s Council spokesperson Anjam Rahman cited numerous examples of anti-Muslim and racist incidents both before and after the terrorist attacks and said “this is New Zealand.”<sup>24</sup> Anti-racism activities sprang up throughout the country, even including a local TV series called “That’s a Bit Racist”.<sup>25</sup>

At the same time, the internet was used to spread misinformation about the attack, including to confuse or misinform the public about the proposed gun law reform, and a host of expressly racist and Islamophobic groups were set up on Facebook and other platforms.

The internet also enabled support for and sharing of the massive public outpourings of grief. Tens of thousands of people attended rallies throughout the country to decry the attack, holding public Muslim prayer vigils to show support for Muslim communities, to grieve, and to come together in

20 Office of the Prime Minister. (2019, 16 May). Christchurch Call to eliminate terrorist and violent extremist online content adopted (press release). <https://www.beehive.govt.nz/release/christchurch-call-eliminate-terrorist-and-violent-extremist-online-content-adopted>

21 AI Forum. (2019, 23 May). Reaction to the Christchurch Call from the AI Forum’s Ethics, Law and Society Working Group. <https://aiforum.org.nz/2019/05/23/reaction-to-the-christchurch-call-from-the-ai-forums-ethics-law-and-society-working-group>

22 Ibid.

23 <https://sharenoevil.co.nz>

24 Fitzgerald, K. (2019, 18 March). Christchurch terror attack: ‘This is New Zealand’ - Muslim woman reflects on past racist attacks. *Newshub*. <https://www.newshub.co.nz/home/new-zealand/2019/03/christchurch-terror-attack-this-is-new-zealand-muslim-woman-reflects-on-past-racist-attacks.html>

25 <https://www.tvnz.co.nz/shows/thats-a-bit-racist>

acts of democratic solidarity and call for deeper examination and honesty about racism, religious intolerance and hate speech.

The New Zealand Law Society joined the debate about hate speech, to inform and educate about this form of speech and what it means to different groups of people.<sup>26</sup> The Free Speech Coalition expressed horror at the attack, saying that the “principle of freedom of expression should be inseparable from non-violence,” but condemned the legal ruling on the status of the manifesto, saying New Zealanders “need to be able to understand the nature of evil and how it expresses itself.”<sup>27</sup> The Coalition has so far been silent on the ruling of the video.

### *Algorithms, social media platforms and internet regulation*

In this environment, the local internet community had to work hard to determine how best to engage with government and also create shared spaces for the community to discuss the issues. Much of this also involved educating about the nature of the internet, the various infrastructural layers and where content regulation fits within other areas of internet policy making. InternetNZ launched a forum for civil society and the technical community to participate in the lead-up to the meeting in Paris to support these discussions.<sup>28</sup>

Regulation of online content is fraught with problems including how to ensure lawful content (such as evidence of war crimes) is not affected by definitions of “terrorist” content. However, the international human rights standards provide a framework for balancing these different rights. At the same time, hard questions must be asked as to whether the multistakeholder cooperation processes which work to create agreed norms at the technical DNS layers of the internet are really working in the social media environment and whether, in the absence of an alternative, regulation has now arrived as the only realistic option. Jay Daley, for example, called for regulation of social media platforms primarily because these “are not the internet” and are not developed and coordinated in

the same ways as other cooperative multistakeholder processes, such as IEEE, the IETF and W3C. Daley argues this is totally unlike the processes used by social media platforms, which have proven incapable of enforcing even their own moral code of conduct standards.<sup>29</sup>

Jordan Carter, the chief executive of InternetNZ, echoed this view, saying that the principles of an open and free internet and content regulation have been “elided, sometimes by organisations that are parts of our constituency, into that sort of cyber-libertarian ethos of government is always bad, freedom of speech is always good, any moves to regulate content or services are always bad.” Given the market dominance of the social media platforms and their impact on public opinion, Carter argues, “just as the public square and the media were always regulated, it isn’t obvious to me that these platforms should be exempt just because they’re on the internet.”

The debate about regulation is continuing and it remains to be seen whether the outcome of the #Christchurch Call will have any significant, long-term impact. The role of social media in the terrorist attack has been expressly excluded from the terms of reference of the national inquiry into the attack.<sup>31</sup>

### **Conclusion**

The effects of the terrorist attacks are still being felt in Christchurch and throughout New Zealand. Public support for gun control laws and wider discussion about racism shows that most New Zealanders abhor the actions of the attacker and want to take some level of personal responsibility for addressing racism in their daily life. Four months on from the attacks, our experience was that prompt legal classification of the video enabled take-down of online content according to the rule of law, thereby upholding and affirming the centrality of human rights in the midst of a horrific terrorist attack. The use of algorithms by social media platforms received considerable attention, helping to inform the public and give more nuance and depth to discussions about AI. This bodes well for future discussion of AI. However, more research is needed to understand

26 Cormack, T. (2019, 4 April). Freedom of speech vs Hate speech. *New Zealand Law Society*. <https://www.lawsociety.org.nz/practice-resources/practice-areas/human-rights/freedom-of-speech-vs-hate-speech>

27 Freedom of Speech Coalition. (2019, 23 March). Banning of manifesto is a step too far. [https://www.freespeechcoalition.nz/banning\\_of\\_manifesto\\_a\\_step\\_too\\_far](https://www.freespeechcoalition.nz/banning_of_manifesto_a_step_too_far)

28 The forum was open to all interested civil society and technical community members in New Zealand and globally. <https://christchurchcallcoord.internetnz.nz>

29 Daley, J. (2019, 16 April). A case for regulating social media platforms. *LinkedIn*. <https://www.linkedin.com/pulse/case-regulating-social-media-platforms-jay-daley>

30 Brown, R. (2019, 12 April). Op. cit.

31 Ardern, J. (2019, 8 April). Supreme Court judge to lead terror attack Royal Commission (press release). <https://www.beehive.govt.nz/release/supreme-court-judge-lead-terror-attack-royal-commission>. Para 6(3)(b) of the terms of reference of the inquiry limits the matters that the Inquiry has power to consider, namely “activities by entities or organisations outside the State Sector, such as media platforms.” See: <https://christchurchattack.royalcommission.nz/about-the-inquiry/terms-of-reference>



the human rights and ethical implications of diverse algorithmic classifiers and the rules that might be created to identify and curate online content.

### Action steps

The following action steps can be suggested for civil society:

- Foster increased public discussion about AI and provide case studies to improve and build understanding of the human rights issues involved.
- Strengthen and develop an interdisciplinary approach to AI to ensure technical, philosophical, legal and other approaches are brought together to develop responses.
- Ensure civil society, academic and technical perspectives play an equal role with government and business perspectives in developing responses to the human rights implications of AI.
- Continue to deal with specific types of harmful content according to law, such as extremist terrorist online content, rather than rejecting regulation of content *per se*.
- Develop research strategies to support technical considerations of the human rights and ethical issues that arise in the development of AI tools (for example, identifying appropriate classifiers and the use of diverse data sets for machine-learning tools).

# 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](http://www.GISWatch.org)

