

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 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

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

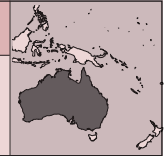
ISBN 978-92-95113-13-8

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.

AUSTRALIA

SOCIAL (IN)SECURITY AND SOCIAL (IN)JUSTICE: AUTOMATION IN THE AUSTRALIAN WELFARE SYSTEM



Deakin University School of Humanities and Social Sciences; Australian Privacy Foundation

Monique Mann

<https://www.deakin.edu.au/humanities-social-sciences>

<https://privacy.org.au>

Introduction

This report examines automation in social security settings in Australia using the case study of the “Better Management of the Social Welfare System” initiative,¹ also known as the “Online Compliance Intervention” or colloquially as “RoboDebt”. Three years ago, the Australian Department of Human Services (DHS), via Centrelink, the agency responsible for administration of social security benefits, launched the automated debt identification programme to detect income reporting discrepancies and the “overpayment” of social security benefits. In practice, the programme aims to achieve savings on social security via the raising of government debts for suspected “overpayment” of welfare benefits, while simultaneously introducing obstacles to contest them.

This programme clearly demonstrates the potential for abuse when big data and automation are deployed against vulnerable individuals. In a wider context of government austerity and cost cutting in both social security and social security services, it reveals the social justice impacts of automated systems that are explicitly designed to target vulnerable individuals. Yet this case study also offers a glimmer of hope regarding the central role that grassroots activism and community-led campaigns can play in countering unjust automated technological systems with human voices. This specifically relates to how conversations about automated technologies can be community focused and inclusive.

Background

In July 2016, the DHS launched an automated debt-raising programme with an algorithm identifying the suspected “overpayment” of government welfare benefits. Request for information letters were sent to welfare recipients requiring them to prove that they did not have a debt, with

unconfirmed or unpaid debts subject to an automated recovery process, which included withholding social security payments or outsourcing to private debt collectors.² A 10% debt recovery fee was added to the alleged debt.³

Between November 2016 and March 2017, 230,000 letters were sent to welfare recipients directing them to an online portal requiring them to prove they did not have a debt. The debt notices were sent out at a rate of approximately 20,000 each week. It was estimated that 20-40% of the debt letters were false positives due to errors in the data used in the system, and the processes that involved averaging annual income and matching it to fortnightly income periods.⁴

Two subsequent inquiries were held: one referred by the Senate Community Affairs References Committee in response to a grassroots community campaign (#NotMyDebt, discussed below), and one initiated by the Commonwealth Ombudsman⁵ in response to an increasing number of complaints. The Senate inquiry recommended suspension of the system until issues of procedural fairness were addressed.⁶ The Commonwealth Ombudsman recommended assistance and support be provided to vulnerable people, and consultation with stakeholders about the difficulties that vulnerable groups face in interacting with the system.⁷

1 <https://www.humanservices.gov.au/individuals/subjects/compliance-program>

2 Knaus, C. (2017, 11 April). Almost half of all Centrelink robo-debt cases sent to private debt collectors. *The Guardian*. <https://www.theguardian.com/australia-news/2017/apr/12/almost-half-of-all-centrelink-robo-debt-notices-sent-to-private-debt-collectors>

3 Commonwealth Ombudsman. (2017). *Centrelink’s automated debt raising and recovery system: A report about the Department of Human Services’ online compliance intervention system for debt raising and recovery*. https://www.ombudsman.gov.au/__data/assets/pdf_file/0022/43528/Report-Centrelinks-automated-debt-raising-and-recovery-system-April-2017.pdf

4 Community Affairs References Committee. (2017). Senate inquiry into the design, scope, cost-benefit analysis, contracts awarded and implementation associated with the Better Management of the Social Welfare System initiative. Canberra: Commonwealth of Australia. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/SocialWelfareSystem

5 Commonwealth Ombudsman. (2017). Op. cit.; see also: Commonwealth Ombudsman. (2019). *Centrelink’s Automated Debt Raising and Recovery System: Implementation Report*. https://www.ombudsman.gov.au/__data/assets/pdf_file/0025/98314/April-2019-Centrelinks-Automated-Debt-Raising-and-Recovery-System.pdf

6 Community Affairs References Committee. (2017). Op. cit.

7 Commonwealth Ombudsman. (2017). Op. cit.

The fiscal context in which the programme was implemented is of central relevance. In 2015-2016, the Australian government forecast it would save AUD 1.7 billion over five years via the identification of welfare overpayments. In 2016-2017, the government then indicated it would achieve AUD 3.7 billion in savings over four years. Within the first six months of the 2016-2017 financial year, the DHS had attempted to recover AUD 300 million of social security “overpayments”, and had secured AUD 24 million in repayments.⁸ Given the small amounts that the DHS had recovered, it has been questioned whether the government will achieve the forecasted savings via this programme.⁹ Casting further doubt on whether the system will achieve any savings, in June 2018 it was reported that the DHS had spent AUD 375 million on the automated debt recovery programme.¹⁰ Despite this, at the time of writing, the system continues to send out “RoboDebts” to meet the financial performance targets set by government.¹¹

The “RoboDebt” disaster

This case study has clear implications for inclusivity and social justice involving the explicit targeting of automated technology on vulnerable populations.¹² There are lessons for the design of automated systems to ensure that they produce accurate findings, with the need for clear avenues for review of automated decisions, especially when vulnerable populations are concerned. Importantly, the case study highlights the role that grassroots activism and community-led initiatives can play in foregrounding human impacts of automated systems. There is a need to engage with the community to evaluate social justice and human rights impacts of automated technology deployed in public settings for public service provision *prior* to their implementation. This should involve consultation about whether these types of programmes should be implemented at all, and if so, whether appropriate checks and balances are introduced such as risk, impact, appeal and accountability processes.

8 Ibid.

9 Ibid.

10 Barbaschow, A. (2019, 13 February). Human Services has spent AU\$375m on ‘robo-debt’. *ZDNet*. <https://www.zdnet.com/article/human-services-has-spent-au375m-on-robo-debt/>

11 Henriques-Gomes, L. (2019, 29 May). Centrelink still issuing incorrect robodebts to meet targets, staff claim. *The Guardian*. https://www.theguardian.com/australia-news/2019/may/29/centrelink-still-issuing-incorrect-robodebts-to-meet-targets-staff-claim?CMP=Share_iOSApp_Other

12 Mann, M., & Daly, A. (2019). (Big) Data and the North-in-South: Australia’s Informational Imperialism and Digital Colonialism. *Television and New Media*, 20(4), 379-395.

Targeting automated technology at vulnerable populations

This is an example of welfare surveillance to further marginalise people who have received social security benefits. It demonstrates that the use of automated technology can serve to perpetuate structural and administrative violence against those who are socially excluded and financially disenfranchised.¹³ There are significant social justice issues associated with explicit design of a programme involving automated technologies that target vulnerable individuals facing financial hardship. The automated debt-raising system began sending out debt notices seven weeks before Christmas, already a time where financial pressure is high, especially for those receiving welfare benefits.¹⁴

The design of the system created issues of administrative justice, procedural fairness, and the rule of law.¹⁵ Numerous challenges were experienced by individuals who attempted to challenge debts, with the onus placed on them to prove they did not have a debt. This reverses the onus of proof onto vulnerable people (and thus overturns the presumption of innocence), which requires them to navigate complex bureaucratic and technical systems to contest alleged debts. According to law, the onus to prove the debt existed technically remained with the DHS: “absent sufficient evidence of an actual debt based on the proper fortnightly data, there can be no *legally sustainable* decision to raise and recover the debt as speculated from averaging.”¹⁶ Despite this, the “scheme targets and raises debts in every case where the person cannot *disprove* the possible overpayment.”¹⁷ Further, the system is computerised and presumably “objective” so it is harder to argue that it is wrong.

Given that debt notices are being sent out by a government agency, extending six years into the past, when individuals are told by that same agency

13 Ibid.

14 Community Affairs References Committee. (2017). Op. cit.

15 See for example: Carney, T. (2018). The New Digital Future for Welfare: Debts Without Legal Proofs or Moral Authority? *UNSW Law Journal Forum*, March, 1-16; Galloway, K. (2017). Big data: A case study of disruption and government power. *Alternative Law Journal*, 42(2), 89-95; Hogan-Doran, D. (2017). Computer says “no”: Automation, algorithms and artificial intelligence in Government decision-making. *The Judicial Review*, 13, 1-39; Zalnieriute, M., Bennet-Moses, L., & Williams, G. (2019). The rule of law and automation of government decision-making. *Modern Law Review*, 82(3), 425-455.

16 Carney, T. (2018). The New Digital Future for Welfare: Debts Without Legal Proofs or Moral Authority? *UNSW Law Journal Forum*, March. www.unswlawjournal.unsw.edu.au/forum_article/new-digital-future-welfare-debts-without-proofs-authority (Emphasis in original.)

17 Ibid. (Emphasis in original.)

that they are only required to maintain records for six months, those who are “the least literate, least powerful, and most vulnerable” may accept the debt is true and seek to pay it off, even when there is a high probability it is erroneous.¹⁸ This places additional financial burdens on those in receipt of welfare benefits, and “[i]n light of the likely vulnerability of so many Centrelink clients, this is a heavy burden indeed.”¹⁹

Design of the debt-raising system

The debt-raising programme has issues in design and technical considerations, including data matching on the basis of inaccurate data, mismatching data, and the averaging of annual income data that is subsequently matched to fortnightly reporting periods to determine if overpayment has occurred during that period (due to under-reporting of fortnightly income data by the welfare recipient). At the initial design stages, a comprehensive risk assessment and consultation with experts and community groups should have identified such issues. However, no consultation occurred.

The averaging of annual income and matching it to fortnightly income periods involved the extrapolation and creation of an assumed fortnightly income average, when actual fortnightly income may fluctuate.²⁰ The Senate inquiry recommended that “the department resume full responsibility for calculating verifiable debts (including manual checking) relating to income support overpayments, which are based on actual fortnightly earnings and not an assumed average.”²¹ This is significant, as 20-40% of the automated debt notices raised by the system were estimated to be erroneous.²² It also raises the question of whether the DHS took reasonable steps in ensuring the accuracy of the information used for the purposes of debt recovery, whether individuals were able to correct their personal information, or even whether they knew they had a right to do so under Australian privacy law.²³

Individuals experienced numerous obstacles in speaking to a human in order to correct their personal information, or contest the accuracy of the automatically identified debts. In the two years

following the introduction of the programme, millions of calls to Centrelink went unanswered. In 2016-2017, the year immediately following the introduction of RoboDebt, there were 55 million unanswered calls to Centrelink. In 2017-2018, 48 million calls went unanswered.²⁴ Many individuals waited for hours on hold.²⁵ The inability to speak with a human undoubtedly created challenges for having false debts resolved. The DHS had a clear conflict of interest as “the harder it is for people to navigate this system and prove their correct income data, the more money the department recoups.”²⁶ So, individuals paid debts, even though they did not believe they owed a debt, because it was “too difficult or too stressful to challenge the purported debt raised against them. Others simply paid the purported debt because they thought the government wouldn’t make a mistake.”²⁷

Fairness and transparency

The use of big data and automated processes allowed for the government to implement a system at national scale with many inherent flaws. There was an absence of fairness in the entire operation of the system including: lack of consultation with stakeholders; no testing or risk assessment processes; the process of automated averaging and matching of data; millions of unanswered calls to DHS; no information being provided to individuals when they sought to challenge a debt; the imposition of a fee for debt recovery; and the fact that the programme extended six years into the past, when individuals are advised they only need to keep records for six months.²⁸ Accordingly, one of the recommendations arising from the Senate inquiry was that clear and comprehensive advice on reassessment, review rights and processes should be made available to impacted individuals.²⁹

There are ongoing issues of explainability and transparency. The debt-raising letters contained no information about how debts were calculated, nor were individuals informed that the debt could be false, or how to contest the calculations.³⁰ The DHS

18 Ibid.

19 Galloway, K. (2017). Big data: A case study of disruption and government power. *Alternative Law Journal*, 42(2), 89-95. <https://journals.sagepub.com/doi/full/10.1177/1037969X17710612>

20 Carney, T. (2018). Op. cit.

21 Community Affairs References Committee. (2017). Op. cit.

22 Ibid.

23 Hutchens, G. (2017, 18 May). New privacy code for public servants after Centrelink ‘robo-debt’ debacle. *The Guardian*. <https://www.theguardian.com/australia-news/2017/may/18/new-privacy-code-for-public-servants-after-centrelink-robo-debt-debacle>

24 Dingwall, D. (2018, 30 October). ‘No party poppers’ for Centrelink’s 48 million unanswered calls. *The Sydney Morning Herald*. <https://www.smh.com.au/politics/federal/no-party-poppers-for-centrelink-s-48-million-unanswered-calls-20181029-p5oco2.html>

25 Community Affairs References Committee. (2017). Op. cit.

26 Ibid.

27 Ibid.

28 Ibid.

29 Ibid.

30 Henman, P. (2017, 4 September). The computer says ‘DEBT’: Towards a critical sociology of algorithms and algorithmic governance. *Zenodo*. <https://doi.org/10.5281/zenodo.884117>

has refused to release documents that relate to the operation of the system, including risk assessment processes, issues papers, and ICT system reports. In June 2019, the Australian Information Commissioner ruled the DHS must release documents following freedom of information requests first lodged in 2017, yet the DHS later appealed this decision to the Administrative Appeals Tribunal. The DHS challenged the release of these documents claiming that publication may pose security risks and external threats, or as one media article stated: “Human Services claimed people wouldn’t pay debts if informed about its IT systems.”³¹

Doxing dissenters, chilling critics

In reaction to an individual complaining about the system online, the DHS publicly released their personal information including welfare history to the media.³² This is an outrageous and reprehensible breach of individual privacy in an attempt to suppress public criticism. In response, a new privacy code is being developed for Australian public servants³³ and it was acknowledged by the Senate inquiry that the system “disempowered people, causing emotional trauma, stress and shame. This was intensified when the Government subsequently publicly released personal information about people who spoke out about the process.”³⁴

#NotMyDebt: Human voices and community initiatives

Notwithstanding the department’s attempts to chill critics, a grassroots campaign known as #NotMyDebt was launched by volunteers led by Lyndsey Jackson, chair of Electronic Frontiers Australia, and united by their “deep concern about the injustice of Centrelink’s robo-debt fiasco and the impact it’s having on the lives of ordinary citizens.”³⁵ The #NotMyDebt campaign demonstrates avenues for making conversations about artificial intelligence more inclusive and human-centred. The campaign collects, houses and disseminates individual stories of false debts, and provides information and advice to individuals who have received a RoboDebt

notice. To date, the #NotMyDebt campaign has collected almost 900 anonymous stories that foreground individual human voices and show the human and social impact of the automated system.

Conclusion

This case study casts light on the failings of big data and automated technology when deployed in public settings for cost-cutting purposes. The RoboDebt programme encapsulates “an error-riddled, unaccountable and politically-driven process.”³⁶ It clearly demonstrates the potential for abuse when big data and automated technology are deployed against vulnerable individuals. Rather than advocating for technical fixes to perfect and optimise the operation of these types of automated systems and technologies, there is a greater need to question whether they should be developed and deployed for certain objectives at all.³⁷ This involves assessment of the desired aims of the system, and consideration of specific social contexts for their use. In the present case, the aim of the system was to explicitly target vulnerable individuals with debts in an attempt to achieve savings on social security. Given this, it is perhaps unsurprising that the system was designed and operated as intended: it raised large numbers of (erroneous) debts, simultaneously introduced insurmountable obstacles to understand and contest them, and publicly targeted those who spoke out against it.

This case study also demonstrates that automated technology is not merely a technical fix deployed to increase administrative efficiency, but is socially and politically embedded.³⁸ Automated verdicts have human victims. Recognition that automated technology and systems are socially contingent³⁹ necessitates proper *a priori* evaluation of the possible social justice and human rights consequences. It requires direct consultation with the community, and involves questioning the type of society that we wish to create, and the role of automated technology within it. As the #NotMyDebt campaign launched in response to RoboDebt has shown, putting human voices back into focus can be an effective strategy of demonstrating not only individual but also wider societal impacts of automated technology.

31 Stilgherrian. (2019, 7 June). Human Services claimed people wouldn't pay debts if informed about its IT systems. *ZDNet*. <https://www.zdnet.com/article/human-services-claimed-people-wouldnt-pay-debts-if-informed-about-its-it-systems>

32 Sadler, D. (2018, 30 May). A 'chilling effect' on free speech. *InnovationAus.com*. <https://www.innovationaus.com/2018/05/A-chilling-effect-on-free-speech>

33 Hutchens, G. (2017, 18 May). Op. cit.

34 Community Affairs References Committee. (2017). Op. cit.

35 <https://www.notmydebt.com.au/about-site>

36 Henman, P. (2017, 4 September). Op. cit.

37 See for example: Powles, J., & Nissenbaum, H. (2018, 7 December). The Seductive Diversion of 'Solving' Bias in Artificial Intelligence. *Medium*. <https://medium.com/s/story/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53>

38 Henman, P. (2017, 4 September). Op. cit.

39 Ibid.

Action steps

The following lessons are suggested by the debt-raising programme:

- Consult the community about the type of society that we wish to create, and the role of automated technology within it.
- Conduct and release risk and impact assessments, including assessment of the impacts for vulnerable individuals and groups, prior to the deployment of automated technology, and periodically during deployment.
- Ensure there are appropriate constraints, checks and balances, and mechanisms for review. This may include approaches such as Article 22 of the EU General Data Protection Regulation⁴⁰ that grants a right not to be subject to automated decisions, or transparency provisions such as a right to understand the basis of decisions, and availability of non-automated remedies, for example, the ability to speak to humans during reviews and appeals.
- Support community initiatives and grassroots campaigns that foreground human voices and the human impacts of automated systems.

⁴⁰ <https://gdpr.algolia.com/gdpr-article-22>

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

