

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

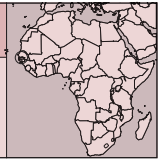
ISBN 978-92-95113-13-8

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.

THE GAMBIA

THE NEED TO USE AI TO DEAL WITH ONLINE HARASSMENT IN THE GAMBIA



SHOAW Gambia

Rameesha Qazi and Anna Anet Sambou
www.shoawgambia.org

Introduction

This report argues for the use of natural language processing to protect women and girls from being harassed online. By tracking how people communicate with each other, we will be able to limit the ways in which messages of harassment are sent and who they reach. To protect the greatest number of at-risk users of social media, civil society organisations and companies like Facebook (which also owns WhatsApp and Instagram) need to work together to protect human rights that are currently being violated.

Context

Women and girls all over the world are targets of online harassment, stalking, and so-called “revenge porn”. In extreme cases, harassment that started online has turned into murder. These kinds of behaviour need to be stopped in the quickest and most straightforward way possible.

The Gambia is no exception – young girls are falling victim to many forms of online harassment and abuse. Recently, a video of a high school girl wearing her school uniform and dancing “inappropriately” with a boy went viral. Subsequently the girl was expelled from school. When human rights activists reached out to the school to review their decision, one of their best teachers said if the girl came back to school he would quit. The Ministry of Education failed to address this. This young girl had a bright future ahead of her and now she will carry around the shame and trauma of this incident for the rest of her life, effectively limiting her future opportunities. We believe that if used in the right way, artificial intelligence (AI) has the power to limit these stories.

Social media in The Gambia is not as widely used as it is in the West, but almost everyone uses WhatsApp (which was acquired by Facebook in 2014). WhatsApp will be the focus of this report, but the impacts of online harassment can be far reaching across all platforms around the world.

Politically, the time is right in The Gambia for something like this to be taken seriously and pursued. Currently the country has no policies or regulations in place dealing with online harassment, nor does it have any legal consequences for perpetrators. This is something that was brought up at the national Internet Governance Forum (IGF) and West African IGF that happened locally in July 2019.

Gender bias in AI: How women’s needs are neglected

The most basic indicator of diversity is gender, and AI is a male-dominated field. According to the World Economic Forum’s latest Global Gender Gap Report, only 22% of AI professionals globally are female compared to 78% who are male.¹ The biggest problem with this is that when male developers create their systems, they incorporate, often in an unconscious way, their own biases in the different stages of their creation.²

There are already many examples of biases in AI that have been found to be impeding the success of women in a variety of fields:

- Several reports have found that voice and speech recognition systems performed worse for women than for men.
- Face recognition systems have been found to make more errors with female faces.
- Recruiting tools based on text mining can inherit gender bias from the data they are trained on.³

AI is impacting on the lives of millions of people around the world, from Netflix predicting what movies people should watch to corporations, governments and law enforcement deciding who gets a loan, a job or immigration status. When AI systems make biased, unjust decisions, it has real-world consequences for people – very often impacting

1 Teigland, J. L. (2019, 2 April). Why we need to solve the issue of gender bias before AI makes it worse. *EY.com*. https://www.ey.com/en_gl/wef/why-we-need-to-solve-the-issue-of-gender-bias-before-ai-makes-it

2 Gomez, E. (2019, 11 March). Women in Artificial Intelligence: mitigating the gender bias. *JRC Science Hub Communities*. <https://ec.europa.eu/jrc/communities/en/community/humaint/news/women-artificial-intelligence-mitigating-gender-bias>

3 Ibid.

on women and people of colour.⁴ Researchers have found that AI systems will spit out biased decisions when they have “learned” how to solve problems using data that is exclusive and homogeneous – and those mistakes disproportionately affect women, people of colour, and low-income communities.⁵

There are also many layers of biases in AI. One is the “unknown unknowns”⁶ which appear after the AI system is complete. Karen Hao writes:

The introduction of bias isn’t always obvious during a model’s construction because you may not realize the downstream impacts of your data and choices until much later. Once you do, it’s hard to retroactively identify where that bias came from and then figure out how to get rid of it. In Amazon’s case, when the engineers initially discovered that its tool was penalizing female candidates, they reprogrammed it to ignore explicitly gendered words like “women’s”. They soon discovered that the revised system was still picking up on implicitly gendered words – verbs that were highly correlated with men over women, such as “executed” and “captured” – and using that to make its decisions.⁷

Amazon’s system taught itself that male candidates were preferable for many jobs based on the data it had been built on.

Dealing with the “unknown unknown” – or predicting online harassment

For the preparation of this report, meetings were arranged with victims of online abuse and harassment. These meetings were with students from age 13 to 16. The aim was to understand how they are being targeted and what kinds of things are happening to them. These students were representative of various socioeconomic backgrounds and ranged from public, private and international schools.

All the students concurred that if they had a circle of 10 friends, more than half would either have been or are being targeted online by men, many of whom they do not know, and some who have positions of power in the victim’s life. All the girls said that they do not know how to address these kinds of problems as there is a culture of silence in The

Gambia, and when girls do speak out the blame falls on their shoulders. All the girls who were present at these meetings said that they were targeted on WhatsApp and Facebook – both platforms that use AI to manage messages⁸ and track behaviour.

The issue is not that AI is being used; the issue is how the AI is built and the gaps that exist in its learning. As suggested, the risk exists in the fact that the learning mimics that of its creator, which can make the problem of gender inequality worse by not addressing it consciously when the AI is developed. So, if a male developer does not see online harassment as a key problem with the internet, it is unlikely to receive much attention in the development of the AI system.

The problems presented here can be easily rectified if civil society organisations and governments come together to gather stories from victims that can be used in the design of AI systems. For example, the creators of the AI can use the messages and behaviours from these interactions to teach the AI which messages to block and how to ensure users who are attempting to harass women and girls are tracked and managed properly.

Julie Teigland writes:

Women need to be builders and end users of the AI-enabled products and services of the future. By shifting the perception, and role, of women within society, we can correct the digital bugs that perpetuate existing bias and make the AI lifecycle more trustworthy. Technology can do many great things, but it cannot solve all our problems for us. If we are not careful, it could end up making our problems worse – by institutionalizing bias and exacerbating inequality.⁹

The problem of online harassment is clearly manageable and solvable. Monitoring text communication using AI can be done, and natural language processing can be introduced to AI systems that already exist to filter voice messages. Natural language refers to language that is spoken and written by people, and natural language processing attempts to extract information from the spoken and written word using algorithms.¹⁰

A typical human-computer interaction based on natural language processing happens in this order: (1) the human says something to the computer,

4 Gullo, K. (2019, 28 March). Meet the Bay Area Women in Tech Fighting Bias in AI. *Seismic Sisters*. <https://www.seismsisters.com/newsletter/women-in-tech-fighting-bias-in-ai>

5 Ibid.

6 Hao, K. (2019, 4 February). This is how AI bias really happens – and why it’s so hard to fix. *MIT Technology Review*. <https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix>

7 Ibid.

8 For example, certain messages cannot be sent on Facebook Messenger, such as links where you can download music or movies illegally.

9 Teigland, J. L. (2019, 2 April). Op. cit.

10 Nicholson, C. (n.d.). A Beginner’s Guide to Natural Language Processing (NLP). *SkyMind*. <https://skymind.ai/wiki/natural-language-processing-nlp>

(2) the computer captures the audio, (3) the captured audio is converted to text, (4) the text's data is processed, (5) the processed data is converted to audio, and (6) the computer plays an audio file in response to the human. But there are many other uses for natural language processing:

- Chatbots use natural language processing to understand human queries and respond.
- Google's search tries to interpret your question or statement.
- Auto-correct is a lot less frustrating these days thanks to deep learning and natural language processing.¹¹

By applying the existing text and voice management technology to the serious human rights problem of online harassment, we will be able to protect every vulnerable user from being harassed, abused, and even murdered.

Conclusion

At the Gambian IGF and the West African IGF, both held in The Gambia in July 2019, these topics were addressed. It was concluded that the internet space needed to become safer for all users and this issue will be raised again at the global IGF.

By implementing natural language processing in AI systems used around the world, one glaring problem arises, which is that the kinds of messages sent to victims could be similar to messages between consenting adults – for example, through sexually explicit flirting, or “sexting”. If the AI has learned how to protect victims of harassment and abuse, then it will block consensual messages as well. But, in our view, the well-being of millions of young girls and women should be protected over the ability of some to be able to send explicit messages.

Civil society organisations and governments need to come together to ensure that proper policies on online safety exist, proper consequences for perpetrators of harassment are in place, and that stories of harassment are collected as they happen so that social media platforms can better teach their AI systems, and install proper reporting tools. AI is a new domain that needs new policies and procedures, but it is critical that all countries have them in place – and soon, as these issues are happening every day to society's most vulnerable groups, and they need to be protected.

Everyone has conscious and unconscious biases about a variety of things, and AI has the potential to overcome but also to inherit/perpetuate biases. We need more women in AI to make sure AI systems are developed *by* women and *for* women's welfare.¹² A gender-responsive approach to innovation will help to rectify the bias that is already built into the system; however, it requires thinking about how we can better leverage AI to protect the most vulnerable users globally.¹³ Once we can secure the most basic human rights online, women and girls around the world can use the power of the internet and social media to empower themselves and their communities beyond what can be imagined today.

Action steps

The following action steps are necessary in The Gambia:

- Collect stories and messages from victims of online harassment to feed into the design of AI systems. While text- or image-based harassment can be easily monitored, AI can also learn speech patterns to block voice messages.
- Civil society organisations need to pressure social media platforms to adapt the AI used in their systems so that they block attempts at harassment of women and girls online. Reporting tools need to be implemented that monitor online harassment more strictly, with real consequences for the perpetrators of these behaviours.
- Civil society organisations need to work together to ensure that victims of online harassment have a safe place to report incidents and get the help they need.

¹¹ Greene, T. (2018, 25 July). A beginner's guide to AI: Natural language processing. *TNW*. <https://thenextweb.com/artificial-intelligence/2018/07/25/a-beginners-guide-to-ai-natural-language-processing>

¹² Gomez, E. (2019, 11 March). Op. cit.

¹³ Teigland, J. L. (2019, 2 April). Op. cit.

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

