

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



APC



Sida



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/>

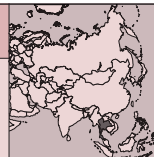
Some rights reserved.

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.



United Nations University Institute on Computing and Society (UNU-CS), Macau and Thailand

Hannah Thinyane and Monticha Puthawong

<https://cs.unu.edu/research/migrant-tech-appraise>

Introduction

Forced labour and human trafficking affect more than 24.9 million men, women and children globally who are exploited for their labour or forced into prostitution.² Figures released by the US State Department indicate that in 2018, only 85,613 victims were identified worldwide.³ These figures illustrate that there are a large number of people, often migrant workers, who are exploited in slavery-like conditions, yet only a small fraction are being successfully identified and subsequently helped.

The terms “human trafficking” and “forced labour” are often used interchangeably by popular media and practitioners, so it would pay to define each now. In our work, we draw on Skřivánková’s continuum of exploitation that defines “decent work” and “forced labour” as two ends of a continuum, with any situation between the two end points representing different forms of labour exploitation.⁴ These work situations can range from “cooperative, consensual, mutually beneficial relationships between migrants and their facilitators” to “highly coercive and exploitative”.⁵ Using this continuum, we can see human trafficking as a process, consisting of a series of exploitative acts that move a worker towards a situation of forced labour. In this

report, we use the term “frontline responder” (FLR) to collectively refer to the broad range of stakeholders whose role it is to assess working conditions and to help potential victims access help or remediation channels – including police, labour inspectors, auditors and non-governmental organisations (NGOs).

In this report we draw from the findings of a two-and-a-half-year project aimed at understanding how digital technology can be used to support exploited workers in vulnerable situations. It starts by describing the development process that we have undertaken in Thailand to create Apprise, a system to support proactive and consistent screening of workers in vulnerable situations. The report then frames the potential of using artificial intelligence (AI) to support an understanding of changing practices of exploitation.

Apprise

Our work takes a value sensitive design (VSD) approach which is based on the understanding that technology is shaped by the biases and assumptions of its designers and creators. VSD proactively integrates ethical reflection in the design of solutions, using an integrative and iterative tripartite methodology comprised of conceptual, empirical and technical investigations.⁶ With its value focus, this self-reflexive approach seeks to be “proactive [in order] to influence the design of technology early in and through the design process.”⁷ In doing so, VSD shows a commitment to progress, not perfection.⁸

We began our work in Thailand early in 2017 with a series of focus groups with a broad range of

¹ We would like to acknowledge the funding provided by Humanity United and Freedom Fund to develop the initial Apprise system. Apprise was developed in collaboration with The Mekong Club, an anti-trafficking NGO based in Hong Kong.

² International Labor Organization, & Walk Free Foundation. (2017). *Global Estimates of Modern Slavery: Forced Labour and Forced Marriage*. https://www.ilo.org/wcmsp5/groups/public/-/-dgreports/-/-dcomm/documents/publication/wcms_575479.pdf

³ US Department of State. (2019). *Trafficking in Persons Report*. <https://www.state.gov/wp-content/uploads/2019/06/2019-TIP-Introduction-Section-FINAL.pdf>

⁴ Skřivánková’s, K. (2010). *Between decent work and forced labour: Examining the continuum of exploitation*. Joseph Rowntree Foundation. <https://www.jrf.org.uk/report/between-decent-work-and-forced-labour-examining-continuum-exploitation>

⁵ Weitzer, R. (2014). New Directions in Research on Human Trafficking. *The ANNALS of the American Academy of Political and Social Science*, 653(1), 6-24. <https://doi.org/10.1177/0002716214521562>

⁶ Friedman, B., Kahn, P., & Borning, A. (2002). *Value Sensitive Design: Theory and Methods*. University of Washington. <https://faculty.washington.edu/pkahn/articles/vsd-theory-methods-tr.pdf>

⁷ Friedman, B., Kahn, P. H., & Borning, A. (2008). Value Sensitive Design and Information Systems. In K. E. Himma, & H. T. Tavani (Eds.), *The Handbook of Information and Computer Ethics*. Wiley & Sons.

⁸ We refer the interested reader to the following papers for a full discussion of the motivation for and subsequent design of Apprise: Thinyane, H. (2019). Supporting the Identification of Victims of Human Trafficking and Forced Labor in Thailand. In K. Krauss, M. Turpin, & F. Naude (Eds.), *Locally Relevant ICT Research*. Springer International Publishing; Thinyane, H., & Bhat, K. (2019). Supporting the Critical-Agency of Victims of Human Trafficking in Thailand. Paper presented at the ACM CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, 4 May.

stakeholders, including survivors of exploitation, NGOs, Thai government officials and intergovernmental organisations. These focus groups aimed to understand current practices and problems in identifying victims of human trafficking; their access to and use of technology; and their perception on the ways technology could support them to overcome the problems that they face. To summarise the findings of this initial consultation, focus groups suggested that support was needed during the initial screening phase of victim identification. The core problems that were identified at this stage were:

- *Communication:* Due to a lack of resources (and knowledge of languages that would be required), FLRs commonly faced problems of being unable to speak the same language as workers and were therefore unable to interview them.⁹ Translators were also not always available.
- *Privacy:* Initial screening occurs in the field and sometimes in front of potential exploiters. Workers fear retribution if they answer questions honestly.
- *Training:* There is a lack of understanding of the common indicators of labour exploitation and forced labour, with some FLRs focusing on physical indications of abuse, rather than the more subtle forms of coercion such as debt bondage and the withholding of wages and important documents.

Based on these findings, we developed Apprise, a mobile-based expert system to support FLRs to proactively and consistently screen vulnerable populations for indications of labour exploitation. The tool is installed on the FLR's phone, but ultimately it serves to allow workers to privately disclose their working conditions. Questions are translated and recorded in languages that are common among workers in each sector, and when combined with a set of headphones, this provides workers with a private way to answer while in the field. By analysing their responses to a series of yes/no questions, Apprise provides advice on next steps that the FLR should take to support the worker. Responses to questions are stored on the FLR's phone and uploaded to a server when they next log in with network reception, to support *post-hoc* analysis. As well as co-designing the system itself, our

consultations with participants uncovered the current indications of exploitation to inform the lists of questions asked. From April 2017 to June 2019, over 1,000 stakeholders in the anti-trafficking field in Thailand contributed to the design or evaluation of the system.

Since March 2018, NGOs have been using Apprise in the field to support proactive and consistent screening in their outreach activities in the following sectors: fishing, seafood processing, manufacturing and sexual exploitation. In May 2018 we started to work closely with the Ministry of Labour (specifically the Department of Labour Protection and Welfare) and Royal Thai Navy in Thailand to understand how Apprise could support proactive and consistent worker screening at government inspection centres at ports (Port-in/Port-out or "PIPO" Inspection Centres) and at sea.

Through this process of working on the ground with FLRs, we have noticed that exploiters continually tweak and refine their own practices of exploitation, in response to changing policies and practices of inspections. When exploiters change their practices, it takes time for these changes to be recognised as a new "pattern" of exploitation. Information is often siloed by different stakeholders, and not shared for a wide variety of different reasons. After some time, stakeholders do begin sharing these changing patterns, often through their informal networks. At this point, the new practice is identified as a pattern and a new policy or practice of inspection is developed.

This game of cat and mouse continues over time, with exploiters again tweaking their behaviour to avoid detection. In response to this, we developed Apprise to allow new questions to be added to question lists, as well as new languages to be supported on the fly. When an FLR logs in to their phone, Apprise checks for any updates to lists and downloads new audio translations of questions. This adaptive support allows FLRs to question on current patterns of exploitation, obtaining further information on exploitative practices once a new pattern has been identified.

Based on this observation, we began to ask ourselves if there was a role for machine learning to support a more timely and more accurate identification of these changing practices of exploitation. While this work is still in its nascent stages, our aim is to determine sector-specific practices of exploitation in order to create targeted education and awareness-raising campaigns; support FLRs to proactively screen against current practices of exploitation; and inform evidence-based policy to support the prosecution of exploiters.

⁹ Most migrant workers in Thailand migrate internally from northern Thailand, or from the neighbouring countries of Myanmar, Laos and Cambodia. Across these regions there are hundreds of languages and dialects that are frequently spoken.

Machine learning to detect patterns of exploitation

At its broadest, machine learning works by identifying patterns in existing data. Its main goal is to be able to generalise, so that the patterns identified in training data can be accurately applied to unseen data. Machine learning has been applied in a wide number of criminal justice contexts, including predicting crimes, predicting offenders, predicting perpetrator identities and predicting crime victims.¹⁰ It has also been used in the anti-trafficking field for predictive vulnerability assessments and crime mapping in order to improve government resource allocation.¹¹ In our work, we aim to understand if there is a role for machine learning to predict changing patterns of exploitation, an area that currently has received little focus.

While there are obvious benefits that accurate forecasting tools could bring,¹² governments, civil society and academics have not always spoken so favourably about these tools, citing cases where they “can reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers or simply reflect the widespread biases that persist in society. [They] can even have the perverse result of exacerbating existing inequalities by suggesting that historically disadvantaged groups actually deserve less favourable treatment.”¹³

While recognising different notions of human rights (moral, ethical and philosophical), our work takes a legal approach, based on the Universal Declaration of Human Rights (UDHR),¹⁴ the United Nations Guiding Principles on Business and Human Rights¹⁵ and the International Labour Organization (ILO) Declaration on Fundamental Principles and Rights at Work.¹⁶ These international legal instruments provide an established framework for “considering, evaluating and ultimately redressing

the impacts of artificial intelligence on individuals and society.”¹⁷

In order to analyse the human rights impact of machine learning on identifying changing practices of exploitation, we note that an important first point of consideration is the quality of data that is provided in initial screening interviews using Apprise, an issue closely linked to privacy.

Significant attention was paid in the design phase of Apprise to include strict limitations on how much data is collected from individual workers, and also who can access screening responses (and what access they have). As an example, Apprise aims to support accountability and transparency by automatically sharing a summarised version of screening responses with the FLR’s immediate supervisor. However, this process limits the accuracy of GPS locations¹⁸ of screening sessions, and only shares responses to the yes/no questions.

To support the privacy of workers, we do not collect any personally identifiable information, as we believe that the risks associated with this would unfairly disadvantage those who chose to answer questions. However, there is no way to delete a particular individual’s responses later (should they be able to request this).

Over the past year and a half, we have evaluated and refined Apprise based on feedback from workers in vulnerable sectors as well as survivors of trafficking. The aim of this has been to increase the privacy that workers feel in these initial screening sessions. We note that while no screening system can guarantee truthful responses from workers, Apprise provides more privacy than current methods of interviewing workers, which often occur in groups and in front of potential exploiters (and in the worst cases, using supervisors as translators when language barriers occur).

Within a machine learning system, interview responses would obviously need to be shared further, which requires special consideration. The new patterns of exploitation themselves are intended to be shared with other FLRs, to inform initial screening of workers in vulnerable situations. However, care must be paid as to *who else* has access to them. As soon as exploiters realise that their patterns of exploitation have been identified, they are likely to adapt them more quickly.

10 Perry, W. L., McInnis, B., Price, C. C., Smith, S., & Hollywood, J. S. (2013). *Predictive Policing: Forecasting Crime for Law Enforcement*. RAND Corporation. https://www.rand.org/pubs/research_briefs/RB9735.html

11 <https://delta87.org/2019/03/code-8-7-introduction>

12 Berk, R., & Hyatt, J. (2014). Machine Learning Forecasts of Risk to Inform Sentencing Decisions. *Federal Sentencing Reporter*, 27(4), 222-228.

13 Barocas, S., & Selbst, A. D. (2016). Big Data’s Disparate Impact. *California Law Review*, 671. <https://doi.org/10.2139/ssrn.2477899>

14 <https://www.un.org/en/universal-declaration-human-rights>

15 https://www.ohchr.org/documents/publications/GuidingprinciplesBusinessshr_eN.pdf

16 The Declaration, among other things, commits states to take action to eliminate all forms of forced labour. <https://www.ilo.org/declaration/lang-en/index.htm>

17 Raso, F., Hilligoss, H., Kirshnamurthy, V., Bavitz, C., & Kim, L. (2018). *Artificial Intelligence & Human Rights: Opportunities and Risks*. Berkman Klein Center for Internet & Society. https://cyber.harvard.edu/sites/default/files/2018-09/2018-09_AIHumanRightsSmall.pdf

18 Some of the decimal points in the position are dropped.

In the cases where responses are accurate, and the tool is able to identify new practices of exploitation, there are obvious implications on the rights of exploited workers: the right to freedom from slavery (UDHR Article 4); the right to freedom from torture and degrading treatment (UDHR Article 5); the right to desirable work (UDHR Article 23); the right to rest and leisure (UDHR Article 24); the right to an adequate standard of living (UDHR Article 25); and freedom from state or personal interference in the above rights (UDHR Article 30). An important note is that while the system takes input from a subset of workers (those who have been interviewed), there is potential to impact the working conditions of many more.

Like any system, Apprise may misidentify patterns, resulting in attention being paid in the wrong direction. While this represents an inefficient use of resources (FLR and worker time), it does not have any significant implications on the rights of workers. This input would be used to inform investigations, which themselves would disprove the prediction.

Conclusion

Machine learning has been applied in a wide number of criminal justice contexts.¹⁹ In our work, we aim to understand if there is a role for machine learning to predict changing patterns of exploitation, an area that currently has received little focus.

In this report we describe work that we are undertaking to proactively and consistently screen workers in vulnerable situations for signs of labour exploitation and forced labour. The report introduces Apprise, an expert system that we have developed and that FLRs are currently using in

Thailand to support the initial screening stage of victim identification. The report also discusses the potential use of machine learning to draw on the responses to the screening interviews and to predict changing patterns of exploitation. We reflect on this proposed system, to understand the human rights implications that this new technology would include. While there is an obvious implication on workers' right to privacy, we describe steps taken to minimise this imposition. We also advocate the use of the system to support the fundamental human rights of workers who are currently trapped in exploitative work situations.

Action steps

We suggest the following steps for civil society organisations who are considering (or are using) AI systems:

- Consider AI as a tool to complement existing efforts and capacity, rather than as a solution in itself.
- Adopt a human rights-based approach to evaluating AI systems, which considers the positive and negative impacts of an innovation prior to rollout.
- Share your stories that reflect on the impact of AI on human rights in order to broaden the types of voices that are included in the global discourse.
- Ensure data privacy and protection are given adequate consideration in the design and development of an AI system: both the raw data itself, but also the predictions that the system generates.

¹⁹ Perry, W. L., McInnis, B., Price, C. C., Smith, S., & Hollywood, J. S. (2013). 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

