PHRG
PEACE HUMAN
RIGHTS GOVERNANCE

Volume 5, Issue 1, March 2021

Building Global Algorithmic Accountability Regimes: A Future-focused Human Rights Agenda Beyond Measurement

Raenette Gottardo

Research Articles*

DOI:

10.14658/pupj-phrg-2021-1-3

How to cite:

Gottardo, R. (2021) 'Building Global Algorithmic Accountability Regimes: A Future-focused Human Rights Agenda Beyond Measurement', *Peace Human Rights Governance*, 5(1), 65-96.

Article first published online

March 2021

*All research articles published in PHRG undergo a rigorous double-blind review process by at least two independent, anonymous expert reviewers

PADOVA UP

PHRG 5(1), March 2021

Building Global Algorithmic Accountability Regimes: A Future-focused Human Rights Agenda Beyond Measurement

*Raenette Gottardo**

Abstract
This paper focuses its research attention on global algorithmic accountability and on governance for algorithms that is embedded in a human rights and International Human Rights Law (IHRL) framework. It looks at how algorithmic accountability can be affected across the 'life-cycle' of an algorithm by ensuring an adherence to core tenets of IHRL, norms and obligations. It looks at how core areas of Human Rights are affected by algorithms and argues that the profound intersection of public and private power, which contributes to the rise of algorithms in public services and business processes, requires a web of interactive legal frameworks that embrace administrative law, anti-discrimination law, data protection law, IHRL and the Ruggie Principles as well as core tenets of corporate law and corporate risk accountability in order to ensure a deep global accountability regime. Only such an interlocked and complimentary web of 'laws' enhanced for the digital age, with IHRL as its anchor, can ensure that algorithms serve humans and not the other way around. Additionally, the paper argues for a new permanent structure to be created within the United Nations Human Rights Council – an Algorithmic Accountability High Level Expert Network (AAHLEN) – that brings together technology skills and developers, human rights lawyers, and public sector senior civil servants to ensure ongoing engagement with the evolution of algorithmic systems and AI and compliance with IHRL. Such an AAHLEN must become a new permanent feature of the United Nations Human Rights Council and must ensure this cross-section of skills builds a long-term Global Algorithmic Accountability system based on IHRL. This paper proposes that a collective meeting be convened of the various UN Special Rapporteurs, that have thus far dealt with the impact of algorithms on their mandate areas, and key algorithm developers and transnational civil society to explore the formation of an AAHLEN within the UN HRC. There is a clear need to ensure that human rights epistemic communities focus on the digital age's permanent impact on respecting and protecting Human Rights in an era of algorithms and AI.

Keywords: *Algorithms, Code, Human Rights, Governance, Regulation*

* University of Stellenbosch, email: Raenettegottardo@gmail.com

Introduction: Governance 'of' and Governance 'by' Algorithm – a Key Distinction

This paper seeks to probe what challenge algorithms and coding poses to the field of human rights and explores what can be done to bring greater human rights to bear on public and private sector algorithmically assisted decision-making to ensure a coherent human rights ethos across both spheres. It argues that anchoring algorithms in an International Human Rights Law ethos framework will suffice and adds an argument that only an interlocked system of different legal regimes, with some key modifications in corporate law and intellectual property law, will assist. It further argues that institutional innovations are necessary at the level of global human rights bodies to ensure the proper construction of a regulatory regime to craft proper governance of algorithms. The research focus is exploratory, and the research methods consists of a qualitative analysis of secondary source material. The exploratory research paper proposes two key regulatory innovations that can be made at the level of the UN Human Rights Council that could aid algorithmic accountability.

Latzer and Just (2020) have compiled a fascinating research approach that makes a clear distinction between 'governance by' and 'governance of' algorithms in order to clearly demarcate different research approaches and analytical approaches to the phenomenon of algorithms and their consequences. There is an effort to bring some analytical order to a growing and significant yet fragmented body of research and emerging scholarship. But how does this distinction work and help us? What is 'governance by' algorithm? How do we distinguish it from 'governance of' algorithms?

According to Latzer and Just (2020, 2):

> *'Governance by algorithms directs the attention towards the steering mechanisms by specific software systems and consequently towards the economic and social effects of algorithms on individuals and the society, that is, on all the opportunities and risks involved. Governance of algorithms builds on these results and focuses on the need, options, and actual policy reactions to shape and control algorithms and their use'.*

Studies on algorithms also differ depending on their units of analysis. Some probe single algorithms *per se* and others probe the socio-technical context or environment of applications that run on algorithmic selection. Both these units of analysis are of interest to scholars of Human Rights that are seeking to build bridges between human rights frameworks in law and algorithm-assisted decision-making processes that put challenges to human rights in play.

This research paper is focused on 'governance of' algorithms and algorithmic accountability and it is interested in both units of analysis cited above as they both impact human rights frameworks.

Lazar and Just (2020) point out that there are at least four different approaches to probing governance of algorithms: (1) risk-based approaches, (2) human rights-based approaches, (3) ethics-based approaches and (4) principled-based approaches. Risk-based approaches are interested in controlling risks from a public interest perspective. Human rights-based approaches probe how algorithms have the power to impact specific human rights (e.g. privacy, freedom of expression and opinion etc.) that could suffer negative consequences from their wide-spread use. The individual is the core focus for human rights-based approaches as is redress for any breeches and core questions of remedies for abuses and violations. Ethics-based approaches have a two-fold focus namely epistemic concerns (quality of evidence generated by algorithms that can be inconclusive, inscrutable and/or opaque, or misguided in their outputs) and normative concerns (engaging with the actual action of the algorithm and unfair outcomes or discrimination and/or transformative effects). The core ethical question of traceability touches on both cause and responsibility for harm as an overreaching concern in an interdisciplinary approach and a specialised new field of 'big data' ethics. The same can be said of law as well as corporate risk models and intellectual property innovations that may be required.

On the 'governance of' algorithms side, Lazar and Just (2020, 10) point out that some core progress has already been made in crafting regulatory responses although much more will be required. In this regard, according to Lazar and Just (2020, 10):

> *'Alongside various regulatory responses, policymakers are also developing national strategies to cope with AI, for example developing guidelines such as the Ethics Guidelines for Trustworthy AI of the European Commission's High Level Group on Artificial Intelligence, or establishing Committees and Centres, including the United Kingdom Centre for Data Ethics and Innovation and the German Inquiry Committee on AI'.*

These structures are good first steps in building global accountability systems as they are tasked with informing the policy processes of challenges emerging from AI, algorithms, and Big Data. They also aid in public awareness raising of the human rights consequences of these new systems.

As far as principles-based approaches are concerned, the most frequently cited are accountability, transparency, fairness, non-discrimination, liability and justification or remedies. These principles are at issue in policy discourses

as well as existing regulations such as the European Union's General Data Protection Regulation (GDPR) or Regulation (EU) 2016/679; GDPR.

Some of the principles-based approaches include:

- Legal protections of accountability and transparency,
- The contribution of transparency to accountability,
- The addresses of accountability,
- The technical solutions and tools available to enhance principles of fairness, accountability, and transparency, or
- Methods of algorithmic accountability reporting that probe algorithms and their powers, biases, and mistakes via reverse engineering.

Having established that this research paper is aimed at probing 'governance of' algorithms it will touch on varied aspects of the highlighted approaches at different times during the course of its analysis.

1. Algorithms, Human Rights and IHRL: Do we Need a Framework for Algorithmic Accountability and a New Algorithmic Accountability High Level Expert Network (AAHLEN) at UN level?

Given that data rights are human rights, this paper views algorithmically assisted decision-making processes in the public sector and private sector as a fundamental challenge to human rights – a challenge that must be confronted by bringing algorithms under the banner and control of IHRL as a moral and practical policy imperative. This is so as every aspect of human rights such as requirements for non-discrimination, privacy, fairness and many other first and second-generation rights are impacted by algorithms and their opaque 'black box' nature. The pervasiveness of algorithms represents a fundamental human rights challenge as algorithms have a darker side relevant to the law due to three main characteristics they have that are interrelated. According to Gerards (2019, 205) these are*: "…algorithms are non-transparent, non-neutral, human constructs".*

Algorithms are human constructs as they are essentially built and trained by humans (potentially replicating our biases), but they can also replicate our bodies and develop as AI and machine learning does. Taken together these three characteristics have an impact on an entire array of human rights including but not limited to the right to privacy, the right to freedom of expression and opinion, the right to access information and they may have discriminatory effects that fly in the face of core non-discrimination provisions of IHRL. Procedural fundamental rights with respect to information available to courts may also suffer negative consequences due to the 'black box' effect of

algorithms where a key party (the designer) inevitably knows more about how the algorithm works compromising the principle of equality of arms before courts. These are challenges to human rights that policymakers, legislators, and courts must confront.

The need for human rights protection in a world of algorithms arises in the public and private sectors alike and imply vertical and horizontal human rights application as a modern necessity despite the challenges of placing the private sector under a human rights rubric that is enforceable and the corporate consequences this may entail.

1.1. Algorithms, Transparency, and International Human Rights Law: Core Building Blocks that Curtail the 'Black Box' Effect:

Clearly this bold new algorithmic world requires adaptations to our laws and legal machinery but also to the way people (lawyers and policymakers adapt accordingly. Changes may include that our laws and legal mechanisms take cognisance of the increasingly intertwined nature of public and private power and sectors that bring conflict between human rights and other public values and interests. This will require court teams of legal, ethical, socio-psychological, and technical experts working side-by-side to make judicial review not only more feasible in the Fourth Industrial Revolution but also more sensitive to human rights. Training and education must become interdisciplinary to create bridges between lawyers and technical experts that both grasp how technology works and legal intricacies and their impact on algorithms. Some of these suggestions made by Gerards (2019, 209) clearly imply that there is an urgent need for a new generation of human rights scholars interested in the effects of Big Data, AI, and algorithms on securing human rights in the 21st century. Such proposals will become indispensable parts of the building blocks required to build a global algorithmic accountability system that ensures 'governance of' algorithms in the interest of human beings.

A convincing argument can be made that IHRL is a perfect framework for algorithmic accountability. In fact, some authors, like McGregor et al. (2019) argue that whilst transparency may be an adequate baseline for algorithmic accountability, only IHRL provides a solid framework that conceptualises 'harm' and means to assess 'harm', that can deal with a multitude of actors (both public and private sector actors) and their varied responsibilities and that can apply to the entire life cycle of an algorithm from its development to its deployment and evolution through feedbacks. IHRL already has such a consistent framework and it can guide decisions to employ algorithmically assisted decision-making tools to ensure safety rails accompany their use. Their work illustrates how transparency is a key bulwark to counter algorithmic

'black boxes' and show that despite this some cases may in fact call for algorithms to be restricted as appropriate forms of checks and balances in a human rights-friendly oversight form. IHRL is therefore a crucial framework to safeguard society against human rights violations both before an age of algorithms and now during one with deeper pressures for transparency to counter the opaque nature of algorithms.

Algorithms therefore raise a number of human rights concerns including discrimination, social rights, the right to health, the right to life, the right to liberty. There is a strong literature on algorithmic accountability that focuses on different parts of the algorithmic process and on how only the framework of IHRL can provide algorithmic accountability across the entire life cycle of an algorithm. IHRL is therefore the best tool to anchor a global algorithmic accountability regime.

According to McGregor et al. (2019, 310):

> *'Instead, the complex nature of algorithmic decision-making necessitates that accountability proposals be set within a wider framework, addressing the overall algorithmic life cycle, from the conception and design phase, to actual deployment and use of algorithms in decision-making'.*

Despite these proposals, their work acknowledges that the system of IHRL crucially places a positive duty on states to ensure that they put a framework in place to prevent human rights violations, create monitoring mechanisms and oversight tools as safeguards to violations, provide for accountability for those responsible for any rights breaches and crafts a remedy for those who have had rights violated in any form. These positive obligations apply to all state actions or omissions and have a bearing on state decisions to implement algorithmically assisted decision-making in public services that are by their very definition the area of core human rights. Whilst IHRL is an attractive framework for algorithmic accountability, some challenges concerning algorithms in the private sector sphere remain as the framework of IHRL only establishes 'expectations' as to how business should operate and not direct obligations under international law. Holding business accountable and ensuring effective remedies under IHRL remains a challenge that can be addressed, in part, by the Ruggie principles and in part by the positive obligations alluded to earlier that apply to state that can imply an obligations to regulate private sector conduct against human rights compliant standards and values. Using IHRL as a source of algorithmic accountability means that it creates an organizing framework for the process of design, development, and deployment of algorithms. This is crucial as automated decision-making in either the public or private sector is far from a neutral tool as every step in an algorithm can reflect a moral choice made by a coder.

1.2. Algorithms and the Fight Against Discrimination: A Human Rights Imperative

Discrimination can both be introduced and amplified by algorithmic-assisted decision-making and that this is a core focus area of debates on algorithmic accountability. Ng (2017) shows that the ideal is to prevent such discrimination via algorithms in the first place and takes issue with the assumptions that are made that algorithmic systems are somehow entirely objective and neutral when the fact that algorithmic models clearly rest on assumptions and data inputs that are not bias-free. Ideas of bias-free algorithms are therefore untenable, and she identifies five stages of a 'life-cycle' of an algorithmic decision-making process where different challenges may arise that are relevant to discrimination: conceptualisation and initial analysis, design, testing, deployment and monitoring and evaluation. Very often the algorithmic models are kept in-house to protect proprietary information and are often confidential. The obvious danger is that this may serve to hide that they contain discriminatory elements due to design, differentiations, data -sourcing or other forms of indirect discrimination or due to biased data inputs or poor-quality data. It is for this very reason that algorithms' opaqueness is a challenge worth confronting despite the corporate risks. Transparency and accountability are absolutely crucial throughout all five stages of the algorithmic life cycle as well as in the implementation of algorithms in the field. Tension between transparency and proprietary privilege must be managed. The complex issue of human control over algorithms and the 'human in the loop' question requires great attention as artificial intelligence can far surpass the ability and capacity of humans to regulate algorithmic decision-making processes if they are not designed to assist such regulation. The moral hazard and risks of vagaries and undesired and unintended consequences pose very real risks to human rights.

In working on the impact of algorithmic decision-making on human rights and international human rights law, it is important to reflect on some examples of how this works and why it is of such significant importance in the 21st century to curtail discrimination and new and deeper forms of inequality. A host of authors have worked on topics ranging from the impact of bias, discrimination on the basis of race or gender or class, the issue of fair outcomes from algorithmic decisions and challenges that arise from the opaque nature of algorithms protected by trade secrets laws.

In her ground-breaking work *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor,* Eubanks (2018) warns us of the 21st century "digital poorhouse" built by algorithms, thereby invoking Dickensian visions

of the fate of the poor under artificial intelligence. Her work found that AI-powered public and private systems linked to health-benefits and policing can make decisions based on flawed or racially or gender-biased data. In doing so her work shows up moralistic and punitive management of the poor.

She uses personal narrative and qualitative methods to show how automated decision-making in social services programmes in the United States has built a "digital poorhouse" marred by ethical abandonments and instrumental efficiency where the data of the poor is manipulated to control them in Indiana, how algorithmic decisions impact the homeless in Los Angeles and their ability to secure shelter and how child welfare recipients in Pittsburgh are captive to the data mining of the poor and arbitrary data-point decisions. Eubanks (2018) leverages these examples to show how automated decision systems deepen social and economic inequality by design and how it undercuts public and private sector welfare efforts. Automated Decision Systems (ADS) opens up an entire new chapter in the global history of the politics of welfare. Her work is a clarion call to sensitise us to how modern governance will be enshrined in impenetrable legal and computer code and what the societal impact could be if steps are not taken pro-actively to minimise their opaque nature.

A similar thematic set of examples is developed and traced by O'Neill (2016) in *Weapons of Math Destruction: How Big Data Increases Inequality,* which looks at how mathematical models or algorithms already have far-reaching impact in assessing things as diverse as teacher quality, recidivism risks of offenders, creditworthiness of clients and a host of other policy where the world runs the risk of entrenching inequality digitally. Her work shows vividly how people are impacted by mathematical models that can encode biases, that can enable predatory corporate behaviour and how algorithms can perpetuate bias or injustice and not be the 'unmediated machine' often depicted for purposes of persuasion. Notably, the author started her own algorithmic auditing firm after probing how college rankings, employment application screening, policing, and sentencing algorithms, workplace wellness and monitoring and credit scores can entrench global inequality under the broad banner of digital change and transformation.

Equally crucially, Zuboff (2019) takes us into new ground on depicting modern capitalism in her work *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* and demonstrates how the development of digital companies like Google and Amazon have business models that represent a new form of capital accumulation. The title of her book neatly encapsulates the challenge of ensuring a human rights compliant framework in an era of a new style and mode of capitalism where

capital accumulation is hallmarked by surveillance and data. As she shows, surveillance capital extracts its profits capturing, rendering, and analysing behavioural data through instrumentalization methods where internally collected data now represents a possible opportunity i.e., the indiscriminate sale of data with profound implications for human rights in all aspects. Her arguments that there are new forms of economic oppression that is creeping into our lives are wholly apt.

Zuboff's (2019) work is deeply reminiscent of Hanna Arendt for solid reasons. Whereas Zuboff (2019) highlights the totalitarian traits of surveillance capital Arendt's body of work on the *Origin of Totalitarianism,* paved the way for us to observe these trends in a new guise in algorithmic form. The argument that surveillance capitalism's deep hunger for data and algorithmically aided decisions and exploitable data is an unprecedented form of capitalism is convincing. Combined with the work of Eubanks (2018) and O'Neill (2016) the contours of the moral risks are vividly visible for champions of human rights.

In *Race After Technology: Abolitionist Tools for the New Jim Code,* Benjamin (2018) shows us how emerging technologies can reinforce white supremacy and deepen social inequality. She argues that automation has the capacity to hide, speed up and deepen discrimination whilst appearing to be neutral or benevolent. Such automations will amplify racial hierarchies, ignore and so doing replicate social cleavages and aim to fix racial bias whilst achieving quite the opposite. She believes that this will create a "New Jim Crow" in reference to the reality of Jim Crow policy outcomes in the United States prison system and in policing. This view suggests and argues that race itself is a kind of technology that is designed to stratify and condone social inequity in the architecture of everyday life and in the 21st century in the digital architecture of everyday life. On the question of gender bias, discrimination, and algorithms, Wachter-Boettch (2018) highlights how chatbots are often expressly designed to harass women in her book *Technically Wrong: Sexist Applications, Biased Algorithms and Other Threats of Toxic Tech.* Keane (2020) shows us how The Pentagon's plans for Automated War will Fundamentally imperil the core human right to life as new Joint All-Domain Command and Control (JADC2) replaces not only soldiers and their weapons but also generals with robotic systems and the prospect of automated war in a post-Clausewitzian version of absolute war. This is of significant import as a new case before the Inter-American Commission on Human Rights witnesses relatives of 34 Yemenis allegedly killed in United States drone strikes seek damages for catastrophic drone strikes. This is important uncharted terrain for international human rights litigation and case law.

1.3. Can We Design Algorithms for a Human Rights Compliant Era?

A significantly important response to concerns about algorithms and human right is to ensure ethical algorithmic design. Indeed, in their report for the Brookings Institute and in their book *The Ethical Algorithm: The Science of Socially Aware Algorithm Design,* Kearns, and Roth (2019) argue that traditional fixes such as laws, regulations and watchdog groups may prove to be unsuitable to the task of one of the most pressing issues in the 21st Century. They argue instead for fixing the problem at source by designing in such a way that human principles (preferably international human rights law) are properly embedded in machine code. This may be one of the ways to square progress with principle in the 21st century.

The ideas advocated by Kearns and Roth (2019) could be adequately canvassed in a special hearing that could be convened by a new United Nations Human Rights structure (such as the AAHLEN proposed in this paper). At such a public hearing the floor could be given to the various UN Special Rapporteurs that have had their human rights focused mandates impacted by algorithms to share their perspectives on the human rights challenges they have observed in probing automated decision systems. This could then be followed by a session of experts from the technology sector that can speak to ethical design and embedding human rights law in machine code.

As Kearns and Roth (2019) have argued, we are not only seeking to prevent biased outcomes but trying to ensure that such concepts can best be captured and represented in the lines of code. They show how evolved this discourse is with respect to differential privacy, but also show how it is needed for fairness and explainability pending proper definitions.

As they argue with respect to their research:

> "At a high level, this research agenda proposes formalising the ethical and social values that we want our algorithms to maintain – values including privacy, fairness and explanation – and then to embed these social values directly into our algorithms as part of their design" (Kearns and Roth 2019, 31).

2. Algorithms and Fairness: Leveraging Administrative Law as an Important Ally

One of the options on the map of interlocking laws is the use of core tools of administrative law which has constrained discretionary power being exercised for centuries in the public sector. As Oswald (2018) has pointed

out, there is absolutely no reason why these rules cannot also be applied and leveraged as the public sector starts to use algorithmically assisted decision-making and predictive tools. This carries some promise that core features of just administrative action, due process rules, transparency requirements and the possibility of judicial review of administrative decisions can play a role in building algorithmic accountability. What Oswald (2018) calls 'old' law can be fashioned to fit a new context and can be used to guide lawyers, scientists and civil service practitioners to fashion these procedures to fit their steps to incorporate new algorithm-assisted decision-making tools in public services where these do not fundamentally offend human rights.

Focusing on COMPAS (a recidivism risk prediction tool used in the United States criminal justice system) and HART (a predictive policing tool used in the United Kingdom), Oswald (2018) showsus how the opacity of algorithms and their complexity (which contributes to opacity) or the 'black box' nature of algorithms can cause challenges that concepts of 'natural justice' and administrative law can assist us with. In the case of 'natural justice' its prescripts of procedural fairness and clear knowledge of the procedures by which public bodies make decisions or take administrative action is key as such principles are also enshrined in Article 6 of the European Convention on Human Rights.

It is the notion of the 'right to be heard'' that is at issue if an algorithm in a decision-process creates doubt as to the basis on which a decision is made and would also be true if a 'human in the loop' were using an algorithmically-informed decision. This is why UN Special Rapporteur on Poverty, Philip Alston, has already so strongly opposed the use of algorithmically-based public service platforms or AI for any social security services as such services ought to make it very clear whether AI is being used on their platforms and what lies at the basis of their decisions.

The administrative law duty to give reasons for decisions can also reveal flaws in the process. These are powerful tools to leverage in the service of human rights. In this regard the notion of the 'human in the loop' becomes important to proving that core administrative decisions were not based on automated processing alone as this would fall foul of EU data protection laws. The decision of all public bodies are susceptible to judicial review and there can be no cogent reason why a requirement that algorithmic systems should provide explanations for their recommendations (suitable to context) cannot be required as a positive contribution to maintaining the rule of law in the 21st century.

The secrecy around the COMPAS algorithm and appeals to corporate secrecy 'carve-outs', must be treated with requisite suspicion when they occur if the rule of law is at stake. This does not imply a higher standard

for algorithms, but as Oswald (2018, 6) has shown, a compliance with age-old constraints on discretionary power that apply to the public sector with or without algorithms being used. There is a clear need for intelligible algorithms and explanations for data-driven classifications.

As Oswald (2018, 7) aptly puts it:

> '*Developments in algorithmic intelligibility and explain-ability can improve 'techniques' of government, and administrative law principles can inform the requirements for such intelligibility, an approach with fairness as its goal'.*

Where the public sector uses algorithmically assisted decision-making processes, administrative law dictates that a public sector decision-maker keeps control of 'steering' the algorithm to ensure it operates in a lawful manner. The 'human in the loop' notion is crucial to ensure that people check the assumptions machines make to ensure compliance with just administrative action and human rights.

The administrative law 'duty to give reasons', would imply that algorithmically assisted decision-making in the public sector be reframed and proper data explanations be given with appropriate granularity depending on context. In terms of such data explanations these challenges will arise at both the level of data inputs/predictors as well as data outputs to check for relevance. As far as risk assessment and predictors are concerned, a core challenge remains that of the 'group-to-individual' problem where risk predictors are often functioning at 'group' levels allowing very little variability for individuals and individual factors raising core human rights and administrative law concerns.

Predictive accuracy and extrinsic factors that lie outside of the algorithmic tools' causal assumptions may also be a complicating issue. One of the key questions that has already arisen for some Human Rights Commissions in the United Kingdom and New Zealand is the question of whether to refrain from algorithmically assisted decision-making in the public sector due to the discrimination risks they carry. The manner in which fairness of state decisions has become a core feature of administrative law means that such norms are in Oswald's words (2018, 17) 'tech-agnostic'. Algorithms and their designers will have to adapt and not the other way around. This means extremely sensitive design of algorithms will be a moral imperative in an era where data rights are human rights.

In this regard, Edward and Veale (2018) argue that we need to move from a 'right to an explanation' to a 'right to better decisions' which implies better designs. We therefore need to in their words 'enslave' the algorithms. This 'enslavement' would be aimed at minimising concerns about unfairness and

a negative impact on the right not to be discriminated against that comes from the 'black box' nature of algorithms. The conventional response thus far to any concerns about negative rights impacts has been to assert a 'right to an explanation' as a route to redress. Highlighting the limited provisions that exist in Articles 22 and 15 and Recital 71 of the existing European Data Protection Regulation to address evolving complexity of algorithm use in the public sector, Edwards and Veale (2018) point to the strengths of existing French administrative law as a possible stronger route to protection and to the modernized Council of Europe Convention 108 as better routes for redress and protection. They argue that the 'right to an explanation' as a form of redress creates a 'transparency fallacy' as an average individual data subject will have considerable challenges to properly protect their rights to privacy. Like Oswald (2018), Edwards and Veale (2018) seek to draw on tools that lie beyond human rights-based approaches to impact assessments and administrative law and judicial review. They go further in their logic of 'enslaving' the algorithms by ensuring actual users control algorithmic system design. France's Digital Republic Act, law no. 2016 – 1321 seems to be a stronger codification of a 'right to an explanation' for administrative algorithmic decisions made about individuals. Where such decisions are made the rules that define that treatment as well as it 'principal characteristics' must be communicated if requested and upon request. The French approach allows for algorithmic decision support system to be made public. It seems as if vague descriptions of a complex model ('black box') will not suffice and needs to be a subject-based explanation. Whilst the French system seems imminently progressive it must be borne in mind that it only covers administrative decisions leaving the private commercial sector's profiling systems untouched.

They also mention the scope of Council of Europe Convention 108 for the Protection of Individuals with regard to Automatic Processing of Personal Data (COE108) which envisages a right to an explanation for all automated decisions that will cover both the public and private sector. Despite these advances they argue strongly that we need a 'right to better decisions' that can only come from system design and ensuring that such system design takes place with a human rights-centric frame of mind. According to Edwards and Veale (2018, 7): '*For this, we must consider the governance tools that have impacts upstream, while systems are being designed or at least before they are deployed'.*

In order to make this happen they argue for two crucial things to happen:

- Privacy by design, data protection by design and impact assessment, and
- Certification systems for ML systems.

They appear to be properly cognisant of the scope for injustices that would still exist from well-designed systems and the need for proper remedies that do not burden individuals and advocate for representation bodies for data subjects that can assist in enforcing rights and securing remedies for breaches. Such bodies would need to be effective and for proper judicial review courts would need to be far more willing to call for access to source-code for decision-making systems. What is desperately required is scrutiny while systems are being built and not only after they have caused human rights harms that could be irreversible and not susceptible to being remedied at all. This strongly echoes the call made by Kearns and Roth (2019) for ethical algorithmic design dealt with earlier. An administrative law approach and better design combined will also ensure that core rights and traditions of access to information, just administrative action and judicial review are maintained along with crucial oversight by legislators. This has implications to ensure proper relationships between algorithm-deploying executive authorities and agencies and parliaments (especially as procurement of such complex and controversial systems in transparent and competitive circumstances are concerned), parliamentary oversight that has necessary technical and information technology expertise to grasp the ethical and human rights and technical risks, and judges and courts with similar skills in cases of judicial review where source-code may require access and proper analysis, and information regulators in countries where access to such regulators offers a viable remedy for any transgressions of violations of privacy rights and data protection laws that may occur.

2.1. Building Stronger Anti-Discrimination Protections by Design not by Default

The analysis thus far has shown how a combination of more focused legal application and better algorithmic design can build a modern human rights compliant guardrail. But more would be needed.

There is a clear need to strengthen legal protection against discrimination by algorithms and artificial intelligence. Borgesius (2020) argues that we need to ensure that direct and indirect non-discrimination standards – contained in Article 14 of the European Convention on Human Rights - are respected as sacrosanct by ensuring that indirect discrimination concepts prohibit any type of algorithmic discrimination. Data protection laws and proper enforcement of non-discrimination laws could work in tandem as a protective barrier. He points out how the GDPR and the Data Protection Convention 108 both require that proper Data Protection Impact Assessments (DPIA) be done that can aid this effort. Noting some key challenges inherent

to enforcing data protection laws, Borgesius (2020, 11) proposes that other initiatives such as the OECD's recommendations on AI as well as the Council of Europe's recommendations on the human rights impact of algorithmic systems ought to be useful as they are instruments that advocate for fair, accountable and ethical AI. The proposals of the European Commission's High-Level Expert Group on AI that advocates for AI Ethics Guidelines in 2019 was a big step forward. Others that have joined in such self-regulatory principles and efforts on Ethics in AI include the likes of Google, UNI Global and the Future of Life Institute. Such voluntary codes, much like those that were adopted by private security and private military companies, often lack enforcement teeth which proves to be an Achilles heel for self-regulation. He concludes that new laws may be required as firms can see ethics as a 'soft' option which they can skip. Borgesius (2020, 12) highlights that the law requires public sector use of algorithmic systems only where such systems enable oversight and auditing and have been subjected to thorough risk assessments. He highlights the need for Equality Bodies and Data Protection Authorities to be given proper enforcement powers to act as real deterrents to violations. This implies proper funding as well as deep technical expertise. One could add that Intellectual Property registering authorities could also be required as a positive obligation in law to ensure that patents are only registered that are human rights compliant and that such certifications could be requested by public authorities and/or Human Rights Commissions were such patents to become algorithms in use in public sector decision-making. Sensible sector-specific algorithmic decision-making regulation is a moral imperative.

According to Borgesius (2020, 13):

> '*Additional regulation should be considered to defend human rights and fairness in the area of algorithmic decision-making'*. As he aptly points out the world did not adopt a single statute to regulate the complete industrial revolution end-to-end and it cannot adopt an all-encompassing statute to regulate the algorithmic decision-making of the Fourth Industrial Revolution as it has become known in World Economic Forum parlance.

One of the crucial challenges of regulation in the Fourth Industrial Revolution will be what legal identity to assign to AI as such artificial persons (a non-human deemed to be a natural person in law) will not only have this status in law, but will additionally also require formal training, testing, verification, certification, regulation, and insurance as the authors point out. Barnett et al.(2017) deal with this challenge extensively and remind us once more that new 'professional industry bodies' of practitioners may be

required to regulate professional standards and ethics for developers and for the evolving AI systems that they build. In fact, the complications arise across the institutional spectrum and will require not only regulation and innovation but also key adaptations of existing structures and our policy and legislative oversight processes themselves as accountability modalities and mechanisms as well as legal standards that govern decision processes hardly ever keep pace.

The tools used today by legislators, policymakers and judicial review processes in courts were built to oversee human decision-makers and not algorithm-assisted decision-makers where 'humans are in the loop' of technology. As Kroll et al. (2016) have argued, our existing frameworks are just not up to the task to act as a bulwark against potentially incorrect or unjust or unfair algorithmic outcomes that emanate form an algorithm running on a computer with little human intervention. If such new algorithmically assisted decision-making tools are to be useful and not harmful to humanity and human rights, we need to make them governable and bridge the gaps that currently exist between these new methods and citizens and society in the throes of the Fourth Industrial Revolution. The co-authors launch an impassioned plea for far greater collaboration across computer science, law and policy which will require developers and politicians/legislators to step out of their respective comfort zones and into one another's worlds. They urge developers to understand that they must design mindful of and with a view to 'After-the-fact-Oversight' whether by parliaments, data regulators or courts. This means that computer scientists and developers may sometimes have to design where there is lack of precision or a need to pursue flexible objectives. This is a tough hurdle for designing decision-making algorithms that may be in front of courts or other regulators.

As ambiguity is often part of the law and policy process as a deliberate objective this is a challenge for computer scientists. It could mean creating algorithms that are regularly reviewable. On the other side of the technology/political divide politicians/legislators need to grasp clearly the risks of ambiguous laws and policies and reduce such ambiguity as far as possible, reduce aspects of policy or decision-making processes that are secret and ensure greater accountability to citizens that will be the subjects of algorithmically-assisted decision-making. One should also realise that a choice of algorithm may entrench a policy choice encoded in software for a long time unless provision is made for sunset provisions in software code. This could be eased by what the authors call a 'general statement of purpose for the algorithm'. (Kroll et al.(2016:633)). Beyond this, the need for clear transparency to combat algorithmic opaqueness is a critical challenge that must be addressed by design.

2.2. Combating Algorithmic Opaqueness: A New Transparency Imperative to Protect Rights:

Clearly that one of the core issues we face in algorithmic accountability is the 'black box' effect – i.e., the sheer complexity of algorithms that complicate questions of transparency. As Desai and Kroll (2017) show, it is important that assertions that society cannot understand or govern these outcomes because the decision-making process is a 'black box' be challenged vigorously. Both by sheer political willpower and intelligent systems design it can be made possible to minimise the 'black box' effect in the interest of human rights, the rule of law and proper oversight and accountability in a new age of algorithmic accountability requirements.

What we have already seen in the field of national security and defense policy over a considerable period of time, is a significant fusion of public and private power. Information technology, defence, private security, private military and private intelligence often flows seamlessly into the public sector and these firms have created and co-created fusions that have global significance. The concomitant human rights challenges have already clearly emerged as Margulies (2016) shows us where surveillance by algorithm and computerised intelligence collection has seen an emergence of two camps of interest – a state-centric camp that asserts that the International Covenant on Civil and Political Rights cannot have extraterritorial application and that all and any actions they take are therefore permissible and surveillance critics that argue that human rights and civil and political rights are compromised as human access and machine access to data are equal clear invasions of the right to privacy. This question has already engaged the attention of the 41st Session of the Human Rights Council in 2019 and is set to remain a burning issue. As this fusion of public and private power potentially moves to other spheres of public policy and the public sector (e.g. significantly areas of the criminal justice, social security, and migration systems) these salient lessons ought to be borne in mind. This will have clear implications for the corporate sector.

3. Ethics, Corporate Law and Algorithms: The Private Sector's Duty in Governance 'of' Algorithms?

Questions about ethical approaches to algorithms on the part of the private sector and corporations also requires our attention as does the impact of algorithmically assisted decision-making as it has a human rights impact in the private sector as well. The core question is whether there is scope for a new ethos in corporate law favouring accountability for algorithms.

This has been particularly visible in the financial services sector with credit records and access to financing and loans that could be discriminatory or breach privacy such as a recent massive data leak affecting global credit risk assessment firm Experian. It has also been evident where private sector firms are bidding for government security contracts that may involve the use of or purchase of facial recognition systems for various uses. Martin (2019) shows how algorithms are silently structuring our lives from recruitment and promotions to loans and political news that impacts our civil and political rights. She argues that algorithms are not neutral but profoundly value-laden entities. They consequently create moral consequences, ethical concerns and impact delegations of roles and responsibilities within algorithmically assisted decision-making processes where 'humans in the loop', or outside of it, may also have an impact. She argues that firms cannot shirk their responsibilities in this regard for value-laden algorithms themselves or for design-based decisions that assign roles and responsibilities.

According to Martin (2019, 835) we simply cannot just absolve firms of responsibility for the development or use of algorithms:

> '*…firms developing algorithms are accountable for designing how large a role individuals will be permitted to take in the subsequent algorithmic decision'.*

The argument that if an algorithm is designed in such a manner that it precludes individuals from taking clear responsibility for a decision, the algorithm designer ought to be held to account for any ethical or other breaches that may occur when such an algorithm is in use as the design decisions about delegations reflect clear moral choices is an important one with wide-ranging implications. Clearly, Martin's (2019) theory of algorithmic accountability goes further than how we currently hold firms accountable for products. Instead of only focusing on responsibility when things go wrong, she proposes that we hold firms accountable for products that work as designed This takes us beyond transparency of designs via accountability attribution. It goes beyond the ''black box' arguments.

Her study on the ethical implications of accountability for algorithms dealt extensively with issues raised by the United States' COMPAS recidivism risk assessment algorithm and its risk ratings in a case of a prisoner (Rodriquez) who argued that the algorithm was discriminatory as it relegated his risk analysis to a group identity instead of weighing him on his individual merits as an individual not a member of a group which the algorithm's design was purported to have done.

Martin (2019, 839) reminds us that 'inscrutable' algorithms are not an opt-out route out of accountability:

> *'I find creating inscrutable algorithms may, in fact, necessitate greater accountability afforded to the algorithm and the developer rather than less – counter to prevailing arguments within computer science, public policy, and law'.*

While technology may be value-laden it remains subject to society's social controls and control mechanisms. For Martin (2019) developers that create algorithms are taking a stand or expressing a view that can be considered moral judgments that trigger accountability not only for how designs operate when in use but for the design itself. Where algorithms are deliberately kept secret, as was the case with COMPAS, which compromised due process rights of Rodriguez, such accountability should be present due to the design, and here intellectual property authorities and regulatory regimes may once more have a role to play. They could be engaged in co-operating with human rights bodies on design details that could be risky and could violate human rights. This could become an additional step for patent authorities.

Martin (2019, 844) argues persuasively that developers are accountable for ethical implications of algorithms in use as well as for role delegations that may leave humans in or out of the loop. According to Martin (2019, 844):

> '*…a firm's obligation for the ethical implications of an algorithm is created because the firm is knowledgeable as to the design decisions and is in a unique position to inscribe the algorithm with the value-laden biases as well as roles and responsibilities of the algorithmic decisions'.*

The firm's obligations grow when it sells the product in the market it does not diminish. One may even conceive of corporate responsibility more broadly to embrace algorithms and their consequences as part and parcel of corporate responsibility. The development and sale of the algorithmic product ought not to be any different to the corporate responsibility that applies to the development and sale of any other products that can have ethical, human rights violations or other risks.

Effectively, we are dealing with a classical principal-agent relationship where the agent (algorithm) acts for the principal (the design firm). Another area where algorithmic delegations have shown principal-agent problems and challenges is the area where drones have been used for targeted killing operations using 'pattern of life algorithms' to launch 'signature strikes' for kill decisions that compromised the right to life, IHRL and IHL as was pointed out by a succession of United Nations Special Rapporteurs. As Martin shows (2019, 845): '*Delegating decisions to drones in military situations takes on similar scrutiny where the developers (a contractor for the government or the military itself) remains responsible for the actions of the agent'.*

What is therefore foreseeable is an expansion of corporate responsibility that takes cognisance of the manner in which the firm designed a particular algorithm to adopt a role in decision-making which will imply responsibility for the decision itself. What is also foreseeable is that coders and firms will have specific fiduciary responsibilities as they will be structuring agents making crucial decisions about access rights for social goods.

What is also implied is that there may be a need to have a different approach to trade secrets law or intellectual property law if human rights bodies and firms can find common ground on designing ethical algorithms. This can be achieved by having a system where some design details are shared with human rights bodies and where human rights bodies set a minimum set of guidelines, akin to the ones developed by the EU Expert Group on AI, in order to ensure human rights language is given effect to in code. Whilst these collaborations will carry some hazards for private sector firms, such hazards by far outweigh the deeper moral hazard of human rights problems through algorithmic deployment which could, in turn, completely destroy a firm and its reputation. Whilst this may carry costs to ensure compliance, such costs can be borne by states that host the technology companies as part of their positive obligations under international law to ensure the realisation of rights. Those who work within human rights bodies will need to be held to high standards to ensure that they do not seek to personally profit from any proprietary information to which they may be privy in ensuring that source code is human rights compliant. This can be achieved through codes of conduct and extremely specific new insider trading offences that could be contemplated for such categories of workers of human rights bodies that have privileged access to proprietary coding information. There could therefore be a need for Big Data review boards or some level of professionalisation specifically for those who will be playing a role in algorithmically assisted decision-making as a new 'professional class' given the moral hazards their actions imply. This may even require re-tooling core curricula of data analysis degrees to include courses in ethics, public policy, and human rights.

Whilst debates on Lethal Autonomous Weapons Systems (LAWS) and 'humans in the loop' of lethal kill decisions leveraging algorithms is far advanced in the context of the Convention on Conventional Weapons (CCW) and Expert Groups, and within private sector firms that have developed these systems already, similar questions will arise about human/ robotic interfaces and delegation of roles in algorithms as these roll out to other areas of public policy, private sector development and our society.

According to Martin (2019, 847):

> *"While augmented labor with robots is regularly examined, we must next consider the ethics and accountability of algorithmic decisions and how individuals are impacted by being a part of the algorithmic decision-making process with non-human actors in the decisions"..*

Mittelstadt et al.(2016) show how we need to map the debate on the ethics of algorithms given how they mediate social processes, mediate across business transactions and mediate government decisions that use algorithmically assisted programs to make decisions. They show how gaps between design and ethical impact can adversely affect individuals, groups, and society as a whole. They raise six overall ethical concerns which they group into epistemic and normative concerns. The three epistemic concerns are: inconclusive evidence, inscrutable evidence, and misguided evidence. The three normative concerns are: unfair outcomes, transformation effects, and traceability. Having mapped these ethical concerns in the field they then identify how ethical issues are treated and ethical aspects of algorithms are addressed. These include:

1. Inconclusive evidence leading to unjustified actions,
2. Inscrutable evidence leading to opacity,
3. Misguided evidence leading to bias,
4. Unfair outcomes leading to discrimination,
5. Transformative effects leading to challenges for autonomy,
6. Transformative effects leading to challenges in informational privacy,
7. Traceability leading to moral responsibility.

The co-authors point out that the ethical concerns from (i) to (iv) have generally been traversed in research and practical terms, the issues raised in (v) to (vii) remain to be addressed.

Ajunwa (2018) reminds us that the rights-based challenge is not only applicable to algorithmically assisted decision-making in the public sector for policing, criminal justice, and social security services, but that it is also present in the world of work. The human rights challenges that arise in the private sector with respect to labor laws, anti-discrimination provisions and employee privacy rights are formidable.

In her study she focuses on the prevalence and increase in productivity monitoring applications and wearable technology. She argues strongly that a 'reasonable expectation of privacy' for employees ought to exist as legal questions arise over employee data collection and use. New monitoring tools that may become ever-more prevalent as COVID-19 alters the world of work with remote working will increase the need for such a properly codified 'reasonable expectations of privacy'.

Beyond privacy concerns, discrimination and workers' safety and workers' compensation are other burning issues that could fall into an algorithmically value-laden space.

According to Ajunwa (2018, 53):

> *"The introduction oof productivity applications and wearable technology in the workplace will create new opportunities to capture employee data. There will be legal controversies as to who should control the data, what data could be introduced in legal proceedings, and how they should be interpreted, et cetera. These issues may, unfortunately, overshadow the greater socio-legal question of whether employers should be able to collect such data in the first place."*

4. Quo Vadis Regulation: 'Governance of' Algorithms and Algorithmic Accountability

While research into fairness, accountability and transparency in machine learning are crucial debates, further debates are needed on proper regulatory bodies of different kinds and debates on what structure would ultimately be well-suited to host such regulatory bodies.

Whilst the cost of regulation and compliance is a perennial theme when these discussions come up, every professional industry has a regulatory body and this ought to also be true for the 'profession' and 'firms' that are developing such enormously powerful instruments that impact every area of human rights.

This research paper argues in addition that, given the human rights issues at stake, that a permanent body must be established within the context of the UN HRC's special procedures that could be called the Algorithmic Accountability High Level Expert Network (AAHLEN) with a clear focus area of anchoring global algorithmic accountability regimes on the guiding framework of IHRL.

Given how many UN Special Rapporteurs with varied mandates have thus far dealt with the impact of algorithmic decision-making on their mandate areas, it seems sensible for a high-level meeting of Special Rapporteurs to be convened with experts and information technology specialists that have designed some algorithms already in wide-spread use to focus on the challenges to human rights that have emerged and to pave the way for the creation of an AAHLEN.

Similarly, a host of national Human Rights Commissions and bodies have added their voices to Special Rapporteurs and their concerns about the operation of algorithmically assisted decision-making tools that have entered

the public sector services basket of public goods. These bodies have flagged specific concerns as far as algorithms and human rights are concerned.

A summit, such as the one envisaged, ought to also hear evidence from these national Human Rights Commissions about their experiences in trying to protect and promote human rights in a world of algorithmic decision-making. This would allow for the issues and legal modalities at stake to be canvassed properly and thoroughly as part of charting a course for a proper more permanent structure to be created as it is warranted given the profound challenges to human rights that are emerging.

Transnational global civil society organisations that have worked on algorithmic accountability ought to also be given an opportunity to present evidence on developments since the adoption of the Toronto Declaration. What should be encouraged is a broad interpretation of the algorithmic accountability issues at stake for the broader international norms' edifice of IHRL as well as ICCPR.

4.1. Regulatory Responses: United Nations Special Rapporteur Reports and National Human Rights Bodies

Given what our analysis has shown about the crucial importance of ensuring data rights are properly seen as human rights and treated accordingly, what are some of the steps we have already seen Special Rapporteurs of the United Nations take? What steps have national Human Rights Commissions taken?

United Nations Special Rapporteur on racism, Tendayi Achiume (2020) raised concern that emerging digital technologies such as Big Data and AI ought to be subjected to far greater scrutiny as they might uphold racial inequality, discrimination, and intolerance. She has gone further into the fabric of IHRL by making her views clear that anyone affected should receive justice and reparations. She argues that in some cases, such as facial recognition, the discriminatory effect of digital technologies may require outright prohibition.

In a Report (2018) the United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression has stated that States should ensure that human rights are at the core of private sector design of AI and machine learning technologies including updating data protection provisions to take cognisance of AI and machine learning. The report envisages that private companies be required to create and apply guidelines for the deployment of AI that are grounded in human rights principles and allow themselves to be transparent and open for auditing of their use of AI. They are also to be required to prevent discrimination actively at both input and output levels of AI.

In another Report (2019) the United Nations Special Rapporteur on extreme poverty and human rights, Philip Alston, condemns in strong terms the way governments are privatising and automating welfare management despite the potential promise that digitization could, in theory, hold for welfare recipients. He essentially stated that he denounced the lack of regulation of private sector companies (Big Tech expressly) and the fact that they operate in a 'human rights free zone' where their tools (i.e. algorithms) eliminate the all-important human component. For Special Rapporteur Alston, a technology-driven future devoid of human rights will be a calamity and a digital dystopia.

A further Report (2018) of the United Nations Special Rapporteur for the promotion and protection of the right to freedom of opinion and expression, David Kaye, also focused on discrimination. The report emphasised that AI technologies have clear implications for human right and gave an overview of all potential human rights affected and a framework for a human rights-based approach to these new technologies. The report highlights that AI is still dependent on human intervention and cannot be said to be 'neutral' considering data inputs. AI's automated decisions may result in discrimination effects and impact key rights such as the right to freedom of opinion and expression (where information and 'micro-targeting' can reinforce biases), the right to privacy (microtargeting arises again as a problem here) and the right to non-discrimination (moderation and filtering hat occurs online due to identity).

The Special Rapporteur proposed a set of tools to oversee the development of AI:

- Human rights impact assessments performed prior, during and after the use of an AI system,
- External audits and consultations with human rights organisations,
- Enabled individual choice thanks to notice and consent, and
- Effective remedy processes to end human rights violations.

The Special Rapporteur made it clear that state policy or regulation in AI must ensure human rights consideration and human rights must guide the development of business practices, AI design and deployment and comply with enhanced transparency, disclosure obligations and robust data protection laws. An important provision suggested is that online providers must indicate which decisions are made with human review and which with AI and data must be kept on complaints about AI and remedies provided.

On 4 July 2018 the United Nations Human Rights Council adopted a crucial resolution (A/HRC/RES/38/7) on the promotion, protection and enjoyment of human rights on the internet that affirmed that people have the exact

same rights offline and online that has clear implications for algorithms and algorithmic accountability and that clearly makes IHRL the anchor for any global algorithmic accountability regime.

Equality of human rights online and offline is particularly the case as far as freedom of opinion and expression was concerned but also applies to all other rights. In its specific set of issues, it urges states to do, the UN HRC's resolution called on states to ensure effective remedies for human rights violations, including those relating to the internet. In accordance with their international obligations. It called on States to address security concerns on the internet in accordance with their international human rights obligations to ensure the protection of all human rights online. The resolution furthermore called on States to consider formulating and adopting national internet-related policies that have at their core the objective of universal access, and the enjoyment of human rights.

It seems clear that the debate cries out for greater coherence and a more systematic approach at United Nations level which can draw on some of the experiences of sub-regional bodies such as the experience of the European Union with its High Level Exert Network on AI (HLEGAI) dealt with below.

At the level of national Human Rights bodies, the United Kingdom and New Zealand have been at the forefront of some of the questions raised. The United Kingdom's Equality and Human Rights Commission submitted evidence to the United Nations that the use of automated facial recognition and predictive policing falls short of its obligations to respect privacy rights under the provisions of the ICCPR. The Commission made it clear that it believed that such practices places discrimination laws on the backfoot and that such advance facial recognition and predictive policing ought to be suspended pending assurances that it can comply with rights regimes.

In New Zealand, the Human Rights Commission worried as early as 2018 in reports that public sector use of algorithms for predictive purposes could lead to unfair treatment of individual groups and that steps ought to be taken to ensure compliance with ethical and human rights standards. Its Ministry for Social Development set out to develop a privacy, human rights, and ethics framework for predictive modelling initiatives in social services.

4.2. Regional Regulatory Responses: the EU Commission's High Level Expert Group on AI as a Template for a UN Algorithmic Accountability High Level Expert Network?

At the level of regional organisations, the European Union has also taken some proactive steps. In a ground-breaking Declaration on AI and Personal Autonomy the Council of Europe's Committee of Minister's dealt with the

impact of the use of algorithms on democracy, human rights, and the rule of law and warned that AI and machine-learning technology ought not to be used to unduly influence or manipulate people's thought and behaviour. This was adopted on 3 February 2019 (Decl. 13/02/2019) Declaration by Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes. This Declaration added its voice to concerns already being raised by various UN Special Rapporteurs as we have seen.

But why do we need a new AAHLEN at UN level? Are we not just replicating what already exists? The proposal in this paper for a AAHLEN is made being mindful of the existence and work of the European Union's High Level Expert Group on AI and the ongoing activities of the EU's legislative agenda and the work of the European AI Alliance. Whilst there are key similarities, for example with the Guidelines for Trustworthy AI (i.e., that AI is lawful, ethical, and robust) and the seven key requirements that AI should meet to be regarded as trustworthy (i.e. human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental wellbeing, and accountability) there would be scope to expand on some of the EU initiatives at UN HRC level which could ensure better governance standards for the adoption of AI in UN Member States. In the context of the new Biden administration in the United States' return to the UN HRC and given the significance scope and norm-setting importance of the United States tech sector, locating ethical and accountable AI efforts properly within the remit of the UN's human rights machinery makes imminent sense.

The EU's focus on specific areas such as Public Sector Services, healthcare and manufacturing and the internet of things could be useful guides as could the Assessment List for Trustworthy AI (ALTAI) which effectively translates the ethical guidelines into an accessible and dynamic checklist of self-assessment for self-regulatory purposes. However, hard law will still be necessary to secure human rights beyond regulatory self-assessment it seems.

In fact, the High Level Exert Group on AI set up by the EU Commission makes the link between their work and the edifice of human rights and human rights law explicit in their report, *The Assessment List For Trustworthy AI (ALTAI) for Self-assessment.* It states (HLEGAI 2020, 3): "This assessment list (ALTAI) is firmly grounded in the protection of people's fundamental rights enshrined in the EU Treaties, the Charter of Fundamental Rights (the Charter, and international human rights law".

However, the voluntary nature of the use of ALTAI and its essential use as a self-regulatory tool potentially points to the need for hard law that can trigger greater obligations on states to ensure AI complies with international

human rights law. Such a discourse, and indeed, such debates on the need for hard law at the global level would be best positioned within the context of the UN HRC in the form of an AAHLEN as proposed in this paper. Such a body can draw on all state and sub-regional experiences and systems and structures, such as, for example, the European Union and the United States and elsewhere aimed at ensuring trustworthy AI to broaden the discourse about the need for hard law to secure rights. Indeed, the ALTAI document itself states that prior to self-assessing an AI system with the Assessment List, it is crucial to conduct a fundamental rights impact assessment (FRIA) to firstly determine whether the AI system negatively impacts any rights. It stands to reason that the body best placed to advise on such FRIA and Algorithmic Accountability and Ethics in the first instance would be the UN HRC and a specialist structure for this purpose, such as the AAHLEN, proposed here, that could draw on extremely specific experts on AI to sit side-by-side with their expert international human rights law counterparts to ensure hard regulatory law to protect human rights can be embedded in machine code in effective ways. Core questions, identified by the HLEGAI report, that must form part of a FRIA for AI include but are not limited to the following:

- Does the AI system potentially negatively discriminate against people on the basis of any of the following grounds, e.g., sex, race, colour, ethnic and social origin, genetic features, language, religion, or belief, political or any other opinion, membership of a national minority, property, births, disability age or sexual orientation?
- Does the AI system respect the rights of the child?
- Does the AI system protect personal data relating to individuals in line with privacy laws (e.g., the GDPR)?
- Does the AI system respect the freedom of expression and information and/or assembly and association?

Most notably, perhaps, with respect to modern weapon systems and so-called 'pattern-of-life' targeting tools, does the AI system respect the right to life?

These questions are being raised in some draft forms of hard law in the United States where H.R. 2231 – the Algorithmic Accountability Act of 2019 was introduced into the 116the Congress, with an equivalent Bill in the Senate. The Act, in essence, required the Federal Trade Commission to require entities that use, store, or share personal information to conduct automated decision system impact assessments and data protection impact assessments. These impact assessments cover evaluation of algorithms in terms of their accuracy, fairness, bias, discrimination, privacy, and security,

use of personal data, security and information systems and stores. Whilst the EU and Hong Kong have adopted guidelines that are voluntary and these issues are being debated in the UN and OECD, the AAA is an example of what hard law that respects rights could look like.

The AAA is seeking to restore the balance between the digital AI-powered corporations and the data subjects – the citizens. For example, the AAA defines high risk Automated Decisions Systems as a system that 'systematically monitors a large, publicly accessibly physical place. Under such a definition an ADS that uses CCTV footage of a bus or rail station or airport for law enforcement could be legally challenged as could the use of CCTV footage and facial recognition technology for predictive policing programmes that have already been challenged in the United Kingdom and New Zealand due to human rights concern.

The AAA has not been passed in the United States as yet and an interesting discourse is presently whether it would be subsumed into a broader overall new United States privacy law that could resemble the EU's GDPR.

Whilst it seems possible that it could pass the House and Senate in some form, following the control Democrats established over both chambers in November 2020's election outcomes, it will be a crucial evolution to watch given the size, scale and scope of the United States' technology industry and AI and ADS systems debates globally and the role it could play in the UN context and discourses about human rights and algorithms. As has been argued, an AALEN could provide a forum for such discourse on new policy and hard law. As Borgesius (2020, 1590) has shown laws (even ones on privacy and data protection and non-discrimination) have deep weaknesses in the field of AI: "We probably need additional regulation to protect fairness and human rights in the area of algorithmic decision-making". He concludes his arguments by calling for sector-specific rules and it can be envisaged that this is exactly the kind of issues of both policy and law to be debated at an AALEN with a view to crafting hard law embedded in machine code that can prevent human rights abuses in the digital age through smart design techniques that are mindful of laws.

4.3. Beyond Regulation: Ethical Employees of Tech Giants: A Vanguard?

At the time of writing the quest for ethical algorithms and ethical AI is growing in importance and in public visibility. Ironically it is not being driven by pro-active regulatory debate or by new hard law discourses, but by the actions of ethical employees of tech giants such as Google that have raised the alarm on ethical questions as diverse at The Pentagon's Project

Maven – which saw some resignations and significant pushback from Google employees - as well as the high-profile firing of AI Ethicist researcher Timnit Cebru from Google's AI Lab Google Brain and AI Ethics Founder Margaret Mitchell following the sacking of Cebru for raising concerns about AI systems that process text, use facial recognition and language processing.

The backlash both from inside and outside Google has seen the ITC giant change its research process in February 2021 attesting to the complexity of ensuring AI is rights compliant. Whilst the employees of tech giants are now unionising as a form of protection for when or if they get isolated when asking moral or ethical questions about AI, these events also speak loudly to the risk of leaving such crucial rights-based questions to self-assessment tools in the absence of hard law.

It is perhaps a reflection of our times that the important work of Eubanks (2018) on the "digital poorhouse" did not focus the mind on the human rights challenges ahead but that work on AI bias in research circles at Google has served to thrust AI bias and international human rights law back into the spotlight. What will be needed is ethical algorithms, ethical employees of big technology, self-regulatory tools supplemented by unambiguous hard law. In a world of facial recognition, surveillance, speech, and language tools that seek to categorise people automated weapons and alternative sets of reality (fake news, deep-fakes, vices, and images), the impact on international human rights law is to be expected. But a robust response is required. Only proper systems and institutional design can address the challenge. This paper's suggestion for the formation of an AALEN within the UN in an institutional design proposal that could contribute to such a robust response to secure rights in the digital transformation age.

Conclusion

This research paper has sought to show that IHRL acts as an important bedrock of global algorithmic accountability regimes. It has shown that an effective web of legal collaboration can be crafted between IHRL, anti-discrimination provisions in particular, administrative law, data protection law the Ruggie principles and areas of expanded concepts of corporate responsibility in corporate law and IP law that can be combined to favour a robust global algorithmic accountability regime.

In addition the research paper has shown clear concerns across domestic human rights bodies, different UN Special Rapporteur reports, EU Declarations and UN HRC Resolutions of algorithms, AI, Big Data and algorithmically assisted decision-making that there is a clear need for some

form of a regulatory node that can bring all these issues under a proper guiding IHRL framework.

The research paper proposed two such initiatives namely, firstly, the creation of an Algorithmic Accountability High Level Expert Network (AAHLEN) housed in the UN HRC to deal with these issues on a more systematic basis and in order to anchor algorithms in IHRL.

Secondly, a Summit that brings together varied UN Special Rapporteurs who have mandates that have been affected by the rise of algorithmically assisted decision-making, transnational civil society groupings and information, communications and technology firms to flag the most crucial issues for the human rights agenda in an era of the Fourth Industrial Revolution and newly emerging human rights challenges associated with it that an AAHLEN would be expected to deal with on a consistent and persistent basis if rights are to be secured in an era of AI and machine learning.

References

'Human Rights Commission warns on discrimination by algorithm', COMPUTERWORLD, 11 May 2018.

Ajunwa, I. (2018) 'Algorithms at Work: Productivity Monitoring Applications and Wearable Technology as the New Data-Centric Research Agenda for Employment and Labor Law', *St Louis University Law Journal* 63(1), 21-53.

Barnett, J. Soares Koshiyama, A. and Treleaven, P. (2017) Algorithms and the Law, Legal Futures Blog. http://www.legalfutures.co.uk/blog/algorithms-and-the-law.

Borgesius, F.J.Z. (2020) 'Strengthening legal protection against discrimination by algorithms and artificial intelligence', *The Journal of Human Rights* http:///www.doi.org/10.1080/13642987.2020.1743976.

Council of Europe Committee of Ministers (2019) Declaration on AI and Personal Autonomy Decl. 13/02/2019, 13 February 2019.

Desai, P.R. and Kroll, J.A. (2017) 'Trust but Verify: A guide to algorithms and the law', *Harvard Journal of Law and technology*, 31(1), 2017.

Edwards, L. and Veale, M. (2018) 'Enslaving the algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' *IEEE Security and Privacy*, 16(3), 46-54.

EU Commission (2020) High Level Expert Group on AI (AIHLEG) The Assessment List for Trustworthy AI (ALTAI) for Self-assessment, Brussels: EU Commission.

Eubanks, V. (2018) *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*, London: St. Martin's Press.

Gerards, J. (2019) 'The fundamental rights challenge of algorithms', *Netherlands Quarterly of Human Rights*, 37(3), 205-209.

Tett, G. (2021) 'After Google Drama, Big Tech must fight against AI bias'. *Financial Times,* 25 February, retrieved from: http://www.ft.com/content/ef0c61ab-240d-42b1-af3c-aca2e4896bd2, (accessed: 10/03/2021).

Hutson, M. (2021) 'Who should Stop Unethical A.I.?', *The New Yorker,* 15 February, retrieved from: http://www.newyorker.com/tech/annals-of-technology/who=shuld-stop-unethical-ai/amp, (accessed: 10/03/2021).

Kearns, M. and Roth, A. (2019) 'Ethical Algorithm Design' *ACM SIGecom Exchanges*, 18(1), 31-36.

Kearns, M. and Roth, A. (2019) *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*, Oxford: Oxford University Press.

Kroll, J., Boracas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G. and Yu, H. (2016) *Accountable Algorithms*, *University of Pennsylvania Law Review*, 165(3), 633.

Latzer, M. and Just, N. (2020) 'Governance by and of Algorithms on the Internet: Impact and Consequences', in Oxford Research Encyclopaedia, Communication and Technology, Communications Theory, Communication and Social Change, Media, and Communications policy, http://www.doi.org/10.1093/acrefore/9780190228613.013.904.

Marguiles, P. (2016) 'Surveillance by Algorithm: The NSA, Computerized Intelligence Collection and Human Rights", *Florida Law Review*, 68(4), 1045.

Martin, K. (2019) 'Ethical Implications and Accountability of Algorithms', *Journal of Business Ethics*, 160, 835-850.

McGregor, L. Murray, D. and Ng, V. (2019) 'International human rights law as a framework for algorithmic accountability' *International and Comparative Law Quarterly* 68(2), 309-343.

Klare, M.T. (2020) 'The Pentagon's Next Project: Automated War', *The Nation,* 27 August, retrieved from: http://www.thenation.com/article/world/trump-pentagon-jadc2/tnamp/, (accessed: 10/03/2021).

Ryan, M. and Mekhennet, S. (2021) 'In a first, Yemenis seek redress for U.S. drone strikes at Inter-American rights body', *The Washington Post,* 27 January.

Mittelstandt, B.D., Allo, P., Taddeo, M., Wachter, S. and Fioridi, L. (2016) The Ethics of Algorithms? Mapping the debate. Big Data and Society. July/December 206, 1-21. http://www.doi.org/10.1177/2053951716679679.

Ng, V. (2017) 'Algorithmic Decision-Making and Human Rights', 21 April 2017, Human Rights Centre Blog, University of Essex, retrieved from: http://www.hrcessex.wordpress.com/2017/04/21/algorithmic-decision-making-and-human-rights/amp/ (accessed: 10/03/2021).

Oswald, M. (2018) 'Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. Philosophical Transactions of the Royal Society' Mathematical, Physical and Engineering Sciences 376 (2128), 20170359, retrieved from: http://www.royalsocietypublishing.org (accessed: 10/03/2021).

Shrimsley, R. (2021) 'Politicians are using yesterday's tools for today's tech challenges', *Financial Times, 24 February*, retrieved from: http://www.ft.content/bd8accda-fb12-4664-a8c9-182e80e8000d, (accessed: 10/03/2021).

Spagnolo, A. (2017) 'Human Rights Implications of autonomous weapon systems in domestic law enforcement: sci-fi reflections on a lo-fi reality', QIL, Zoom-in 43, 35-58.

UN HRC Resolution, UN Doc. A/HRC/RES/38/7, 4 July 2018.

UN News, 'Independent rights expert says emerging technologies entrenching racism, discrimination', 15 July 2020, retrieved from: http://www.news.un.org/en/story/2020/07/1068441, (accessed: 10/03/2021).

United Kingdom Equality and Human Rights Commission, 'Facial Recognition technology and predictive policing algorithms out-pacing the law', 12 March 2020.

United Nations (2018) Report of the UN Special Rapporteur for the protection and promotion of the right to freedom of opinion and expression. UN Doc. A/73/348, 29 August 2018.

United Nations (2018) Report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. UN Doc. A/73/348, 28 August 2018.

United Nations (2019) Report of the UN Special Rapporteur on digital technology and social protection. UN Doc. A/74/48037, 11 October 2019.