doi:10.1017/aju.2018.86

SYMPOSIUM ON THE UNIVERSAL DECLARATION OF HUMAN RIGHTS AT SEVENTY

UNDERMINING HUMAN AGENCY AND DEMOCRATIC INFRASTRUCTURES? THE ALGORITHMIC CHALLENGE TO THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

*Helmut Philipp Aust**

"Digital technology is transforming what it means to be a subject."[1] The increase in the use of big data, self-learning algorithms, and fully automated decision-making processes calls into question the concept of human agency that is at the basis of much of modern human rights law. Already today, it is possible to imagine a form of "algorithmic authority," i.e., the exercise of authority over individuals based on the more or less automated use of algorithms.[2] What would this development mean for human rights law and its central categories? What does the Universal Declaration of Human Rights (UDHR), adopted seventy years ago as a founding document of the human rights movement at the international level, have to say about this?

It is not hard to think of conflicts between these technological developments and individual rights set forth by the UDHR.[3] The right to privacy (Article 12) springs to mind first. Other challenges include, most prominently, the impact of algorithmic decision-making processes on sentencing decisions (Article 10), the use of automatic weapons and the ensuing implications for the right to life (Article 3), and the impact that the manipulation of news and social media can have on the realization for the right to "take part in the government" in a given country (Article 21(1)).

Yet the more important challenges arguably go beyond clashes with specific rights of the UDHR. This essay identifies two foundational challenges. First, the essay argues that these technological developments pose a fundamental challenge to the very notion of human agency. Second, we need to inquire into the effects of these major technological changes on what I call here the infrastructure of democratic decision-making. There are of course other conceivable issues that merit discussion in this context. However, the two challenges just identified deserve to be singled out, as they illustrate fundamental concerns on two levels: the first is conceptual in nature, while the second is practical.

A few words about definitions are in order. At the very base of automated decision making are algorithms, which can be defined as a "sequence of computer code commands that tells a computer how to proceed through a series of instructions to arrive at a specified endpoint."[4] Big data, in turn, is a concept meant to capture the vast expansion of the quantity of available data. It is a "generalized, imprecise term that refers to the use of large data

* *Professor of Law, Freie Universität, Berlin; Senior Fellow, Melbourne Law School.*

[1] Andreas Reckwitz, Die Gesellschaft der Singularitäten 244 (2017).

[2] Deborah Lupton, Digital Sociology 49–50 (2015).

[3] For an overview in the context of the European Convention on Human Rrights, see Algorithms and Human Rights, (Council of Europe Study DGI(2017)12, Mar. 2018).

[4] Lupton, *supra* note 2, at 11.

The American Society of International Law and Helmut Philipp Aust © 2018. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/aju.2018.86 Published online by Cambridge University Press

sets in data science and predictive analysis."[5] Finally, automated decision-making implies that data processing takes place without human intervention. The crucial question is at which point it is suitable to speak of a fully automated decision. Until today, and probably for some time to come, the use of such processes will be connected to a decision by a human person to employ such techniques. Self-learning programmes can potentially sever this connection to the human origin of such devices, which brings us to the first major challenge for the UDHR and human rights law emanating from these technological developments.

Human Agency and Algorithmic Authority

The notions of human agency and responsibility are central to the conceptual arsenal of human rights law. Human rights depend on the notion of a clearly identifiable rights-holder, but also on the philosophical premise that these human rights are protected against the state as a collective, which consists of humans who hold duties towards others. This idea is encapsulated in UDHR Article 1, which stipulates that all human beings are "endowed with reason and conscience and should act towards one another in a spirit of brotherhood."

To the extent that decisions of the state are based on the exercise of algorithmic authority, this human bond between the rights-holder and the state (and its organs) starts to disappear. This follows from the potential dehumanization of decision-making in a world of "algorithmic authority." Regardless of the processes by which the state and its organs arrive at a decision, a human element has traditionally been involved. Human agency ensures that it is possible to give reasons for the conduct in question.[6] This does not mean that conduct violating human rights is reasonable—but it can be explained by recourse to the considerations that triggered the conduct. The reasons may be dubious to the point where we start to question the common bonds among humans—but to some human being it made sense to act that way. Human rights violations can be explained, even if not necessarily understood, let alone justified. This possibility to explain is coupled with the capability of humans for empathy.[7] And this capacity requires a sense for the possibility of suffering.[8]

These foundations can start to wear thin when automated decision-making becomes the norm. This can be illustrated with an example of current practices. In 2014, the Chinese government launched a plan to develop a "social credit system."[9] Although many uncertainties persist with respect to the concrete unfolding of this plan, its general idea is to generate a "social score" for all Chinese citizens that draws on data derived from a wide range of different activities. The plan merges features of well-known credit score systems with a more ambitious goal to provide for societal "sincerity" and cohesion. The only way to run and manage that plan—be it in a centralized or decentralized manner—is through the use of big data processes and self-learning algorithms that connect the various dots. Consequences of the social score might range from access to public services over the imposition of travel bans to the shaming of individuals on social media platforms. If implemented in this wide-ranging manner, this scheme clearly has the potential to exercise authority over individuals based on algorithmic

[5] Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. Rev. 93, 96 (2014).

[6] Mireille Hildebrandt, Smart Technologies and the End(s) of Law 73 (2015).

[7] *See* Lynn Hunt, Inventing Human Rights: A History 39–40 (2008) (describing the processes of learning empathy from an early age onwards).

[8] Hildebrandt, *supra* note 6, at 125; *see also* Susan Sontag, Regarding the Pain of Others (2003) (discussing the visual production of empathy); Itamar Mann, Humanity at Sea – Maritime Migration and the Foundations of International Law 12, 187 (2016) (identifying the human encounter as a formative basis for human rights law).

[9] *See China Invents the Digital Totalitarian State*, Economist (Dec. 17, 2016).

https://doi.org/10.1017/aju.2018.86 Published online by Cambridge University Press

decision-making.[10] Through the coupling of all conceivable forms of data, it will become increasingly difficult to justify or rationalize the consequences that are imposed on a person in light of his individual behaviour.

While there might not be a universal right to know why a state has taken a particular action, it is an essential element of both the rule of law and modern understandings of human rights law that the state must justify exercises of its power. Based on the information conveyed about state action, it becomes possible to argue about the compliance of a given act with human rights law.[11] This can become difficult when the public sector relies on powerful algorithms and big data operations, especially and even more so when this form of conduct involves cooperation with the private sector. The provision of public goods becomes privatized even when the state is still making the relevant decisions.

The conventional answer to this problem is a turn to transparency as a means to realising the conditions for the right to an effective remedy. Yet this turn is not without its own complications.[12] It appeals to a widely shared feeling that a loss of transparency is one of the main problems with respect to the data-driven processes analysed in this essay.[13] This lack of transparency stands in the way of giving reasons, to which we are accustomed with respect to state action triggered by human agency. After all, if it were possible to see through the processes underlying the exercise of algorithmic authority, it would be sufficient for a state organ to ratify the decision, to fully understand the rationale behind the data operation, and to translate these insights into the regular forms of communication between the state and its citizens. If it were so simple, this would indeed be the best way forward.

Yet it is very unclear whether the processes behind the exercise of algorithmic authority lend themselves to such an identification and translation process. The highly technical nature of the underlying data operations will potentially limit what transparency can achieve here. It is of course conceivable to disclose the code contained in an algorithm and thereby to help explain how a certain decision was reached. This would only be a very thin form of an explanation, however. The use of algorithms entails a fundamental shift in governance techniques, from a model that is centered on language and words to techniques that are premised on data-driven analytics. It is no longer words and their interpretation which would be able to rationalize conduct but a production of meaning through the language of programming.[14] What is more, due to the growing reliance by the public sector on software produced by private corporations, intellectual property concerns might stand in the way of disclosing the precise workings of algorithms, big data, and self-learning machines. Ultimately, transparency might thus in this context be a symbolic move expressive of the wish to "affirm human authority over data in principle" rather than exposing the underlying mechanism to effective scrutiny.[15] Such effective scrutiny, in turn, would depend on a technological infrastructure and savviness on the part of state powers and the wider public to truly engage with these complex features, which might be hard to realize.[16]

[10] Stefan Brehm & Nicholas Loubere, *China's dystopian social credit system is a harbinger of the global age of the algorithm*, Conversation (Jan. 15, 2018).

[11] Hildebrandt, *supra* note 6, at xi.

[12] *See generally* Andrea Bianchi, *On Power and Illusion: The Concept of Transparency in International Law*, *in* Transparency in International Law 1, 6–10 (Andrea Bianchi & Anne Peters eds., 2013).

[13] Fleur Johns, *Global Governance Through the Pairing of List and Algorithm*, 34 Env't & Plan. D: Soc'y & Space 126, 140 (2016).

[14] Larry Catá Backer, *And an Algorithm to Bind them All? Social Credit, Data Driven Governance, and the Emergence of an Operating System for Global Normative Orders* (Entangled Legalities Workshop think piece, May 24–25, 2018).

[15] Fleur Johns, *Data, Detection, and the Redistribution of the Sensible*, 111 AJIL 57, 84 (2017).

[16] *See* Bianchi, *supra* note 12, at 15–19.

https://doi.org/10.1017/aju.2018.86 Published online by Cambridge University Press

Algorithmic Authority and the Infrastructures of Democratic Decision-Making

Secondly, the growing importance of algorithmic authority also potentially undermines a crucial condition for the existence of democratic societies. A democratic exchange among equals presupposes an information infrastructure that is accessible to all members of the public. At first sight, modern technology seems to further that ideal through the enhanced access to information on the internet and through the various channels of social media. Much of the early enthusiasm about the "Arab Spring" in 2011 related to the spread of protest through various channels of social media, leading some to speak of a "Facebook revolution."[17] On closer look, however, information and debate through social media channels might come to dissolve "the public" as we know it. Going back to the quotation set out at the beginning, we are living in a "society of singularities," according to Reckwitz a defining feature of our time. Each individual receives customized information based on the preferences and previous online activities registered with the respective social media outlets—at the risk of creating the now already proverbial "filter bubbles." This has serious consequences for the public debate on which democracies necessarily rely. This has become painfully obvious in the revelations about Russian intervention in the 2016 U.S. presidential elections[18], about the role that Facebook played in the run-up to the Brexit referendum of the same year[19], and about how YouTube's use of algorithms pushes its users in Germany to ever more fringe and alt-right content instead of presenting a balanced set of news.[20]

The UDHR has its own troubled history when it comes to democracy (which relates mainly to the "no distinction" clause of Article 2).[21] It does not set forth a "right to democracy," and is no outlier in this regard when compared with later human rights treaties.[22] But it envisages that everyone has "the right to take part in the government of his country, directly or through freely chosen representatives." Important here is what "freely chosen" means. This can be given a narrow meaning that only relates to elections without any form of coercion vis-à-vis the citizens invited to express their vote. But one can also understand it in a broader sense that implies that the process of "taking part" is organized in a manner conducive to the realization of the underlying goals of the UDHR. If the public discourse leading up to elections becomes the subject of technological manipulation, this is a vivid illustration of democracy's capture by the most powerful forces of twenty-first century market capitalism.

A related sentiment was foreshadowed even before the UDHR was adopted in 1948. When UNESCO engaged in its 1947/1948 Human Rights Survey, the "technological society of the future" was a theme that several respondents identified in a way that resonates today. In his entry, novelist Aldous Huxley recognized that "applied science has in fact resulted in the creation of monopolistic industries, controlled by private capitalists or centralized national governments . . . . Applied science in the service, first, of big business and then of government has made possible the modern totalitarian state."[23] At the time of writing, that was a backward-looking comment as well as a worry about what would come.

[17] Jose Antonio Vargas, *Spring Awakening*, N.Y. Times (Feb. 17, 2012).

[18] Mike Isaac & Daisuke Wakabayashi, *Russian Influence Reached 126 Million Through Facebook Alone*, N.Y. Times (Oct. 30, 2017).

[19] Mark Scott, *Cambridge Analytica Helped 'Cheat' Brexit Vote and US election, Whistleblower Claims*, Politico (Mar. 27, 2018).

[20] Max Fisher & Katrin Bennhold, *As Germans Seek News, YouTube Delivers Far-Right Tirades*, N.Y. Times (Sept. 7, 2018).

[21] Sundyha Pahuja, Decolonising International Law: Development, Economic Growth and the Politics of Universality 64 (2011).

[22] On the "right to democracy," see Tom Ginsburg, *Introduction to the Symposium on Thomas Franck's "Emerging Right to Democratic Governance" at 25*, 112 AJIL Unbound 64 (2018) and the contributions to that symposium.

[23] Aldous Huxley, *The Rights of Man and the Facts of the Human Situation*, *in* Letters to the Contrary – A Curated History of the UNESCO Human Rights Survey 207, 211 (Mark Goodale ed., 2018).

https://doi.org/10.1017/aju.2018.86 Published online by Cambridge University Press

In the post-Cold War era of the 1990s and early 2000s, there seemed to have been reason to develop a more optimistic view about the potential contribution of technology for the common good. Did not the supposedly final triumph of a Western-led liberal world order coincide with the blossoming of the World Wide Web? And did the "New World Order" of the post-Cold War, with its emphasis on a "disaggregated statehood,"[24] not chime well with expressed hopes that "Code is Law"[25] and that cyberspace would be a state-free domain?

Reflecting on these questions in 2018 makes us reconsider Huxley's dark remarks. The euphoria of the 1990s and early 2000s is certainly gone. At the latest since the revelations of whistle-blower Edward Snowden, there is acute awareness of the growing ease with which (and the extent to which) states use the internet and related technological innovations for surveillance.[26] Today, the internet and the related technological complex of algorithms, big data, and self-learning machines may still seem to be very much dominated by private companies from the United States. However, the rise of China as a major power is a reality in this field. The beginning experiments with the above-described social credit system alert us to the potential impacts of algorithmic authority of the life of citizens in virtually all areas. It is not unlikely that competing visions about the digital life by U.S. actors and Chinese forces will dominate the debates in this field for decades to come. Hence, we are back to the alternatives that Huxley sketched with respect to the control over the driving forces of technological change in 1947—private capitalism or centralized national governments. Both options on offer today raise serious questions about the viability of democratic processes in the future—either for the reason of an excessive overreach by state authorities in the case of the Chinese social credit system or through the dominance of an oligopoly of private corporations controlling the flow of information in an unprecedented manner.

Conclusion

It is uncertain whether any of the powerful forces that seem to be on the verge of defining technological progress in the twenty-first century care much for either of the two challenges to human rights law outlined above. The operation of both Western-driven social media platforms and the Chinese social credit system depend on dehumanizing processes that imply the more or less automatic generation of content and attribution of consequences to individual behaviour. At the risk of some overgeneralization, neither of the "two systems" will be ready and willing to disclose their inner workings so as to enable effective ex post scrutiny. And neither of the two models seems to be interested in maintaining or creating an infrastructure for an open and pluralistic public debate. The open question is which model will win: a "singularistic" turn to customized information with a pseudo-liberal and individualistic touch or the "collectivist" utopia of the state as the ultimate harbinger of peacefulness and equity in private social relations. For the time being, it seems unlikely that the UDHR will be replaced with a new document that could serve as a new foundation for human rights law in the twenty-first century. This might also be unnecessary in the first place. Rather, a political debate on how to enable the effective realisation of the rights set forth in the UDHR in a changing technological environment is needed.

[24] Anne-Marie Slaughter, A New World Order (2005).

[25] Lawrence Lessig, Code and other Laws of Cyberspace (1999).

[26] On the Snowden revelations see Helmut Philipp Aust, *Spionage im Zeitalter von Big Data? Globale Überwachung und der Schutz der Privatsphäre im Völkerrecht*, 52 Archiv des Völkerrechts 375 (2014).

https://doi.org/10.1017/aju.2018.86 Published online by Cambridge University Press