Towards a Concept of Digital Citizenship

AI and the Universal Declaration of Human Rights

An initiative of the Center for International Human Rights (CIHR) at John Jay College of Criminal Justice, City University of New York (CUNY)

Towards a Concept of Digital Citizenship: AI and the Universal Declaration of Human Rights

Co-Authors:
Marie-Michelle Strah, Ph.D., *Visiting Scholar, Center for International Human Rights and Adjunct Professor of International Crime and Justice, John Jay College of Criminal Justice*

Alexandra Johnson, *Assistant to the Director, Center for International Human Rights and Graduate Student, International Crime and Justice, John Jay College of Criminal Justice*

Contributing Author:
Timothy Botros, M.A., *Research Assistant, Center for International Human Rights and International Crime and Justice Graduate, John Jay College of Criminal Justice*

Research Provided by:
Timothy Botros, M.A.
Gabriella Gardziola, *Research Assistant, Center for International Human Rights and Graduate Student, International Crime and Justice, John Jay College of Criminal Justice*
Alexandra Johnson

Date of Publication: March 2021

1. Introduction: Citizenship in the Age of Artificial Intelligence

The Universal Declaration of Human Rights (1948) and the subsequent International Covenant on Civil and Political Rights (1976) and International Covenant on Social and Economic Rights (1976) take the concepts of rights of the individual to self-determination in political, economic, and cultural spheres as foundational principles. At the time when the Declaration and the two Covenants were originally drafted, the rise of artificial intelligence and big data were unforeseeable. Nonetheless, United Nations High Commissioner for Human Rights Michelle Bachelet has recently argued that existing laws adequately encompass and address the due rights of the individual in AI-mediated cyberspace, noting that it is not "a human rights black hole" while "the same rights exist online and offline" (Office of the High Commissioner for Human Rights, 2019). This political stance is further advanced by the embrace of artificial intelligence as a driver of economic and political development by other UN directorates in support of the Sustainable Development Goals agenda. However, since the initial adoption of the UDHR and ratification of the ICCPR, two underlying assumptions have radically changed:

- ***the first*****,** that human rights law was developed to protect the individual from infringement of rights by a nation-state;
- and ***the second*****,** that individuals exercise their rights to self-determination in the physical world.

The UN Special Rapporteur on Extreme Poverty has also brought to light the oft-neglected concept of economic rights in the context of the UDHR by having called into question the adverse impact AI has had on individual self-determination in the economic sphere, such as lack of access to skilled work and increased economic inequality (Alston, 2019).

Indeed, the deployment of AI worldwide has proceeded with almost teleological determinism in the public and private sectors, to include the international community, and as the Special Rapporteur noted, "the era of digital governance is upon us" (Alston, 2019). Whilst there has been a flurry of white papers by NGOs and human rights groups since 2016, they have largely addressed the problem of human rights from a general standpoint in the context of data privacy. Currently, the academic and policy discussions have relied on the legal assumption that the rights of an individual are distinct from an individual's data, the latter of which being deemed a separate question concerning one's property rights (Amnesty International, 2019; Ahktar, 2019; Latonero, Big Data Analytics and Human Rights, 2018). However, given the extent to which peoples' lives exist online by way of necessity, the two are becoming increasingly inseparable.

It is our position that international legal frameworks must adapt to this new reality by establishing a universally applicable definition of *digital citizenship* (the individual in both the physical world and cyberspace concurrently) and the *role of corporate actors* (not just nation-states, who have remained weak actors in the digital space). Using the works of Hin-Yan Liu (Liu, 2019) and Carsten Momsen and Caecilia Rennert (Momsen & Rennert, 2020) we provide a critical analysis of specific human rights impacted by the rise of artificial

intelligence in both decision-making and in an individual's free exercise of freedom of expression, self-determination and political and economic rights.

To illustrate, we will address **two key focal areas** in this white paper with accompanying case studies:

1. **rights in the free exercise of citizenship** (as enshrined in UDHR articles 3, 6, 13 and 15)
2. **rights to be free of torture or cruel, inhuman or degrading treatment or punishment, or arbitrary arrest, detention or exile** (UDHR articles 5 and 9)

In two subsequent white papers we will explore two additional focal area: the impact of AI-mediated distributed content delivery on the rights of freedom of expression and free and fair elections (UDHR articles 19, 20, and 21) and economic rights to social security and work (UDHR articles 22 and 23).

Human rights are based on the premise that every single human being, regardless of background or domain—including the online sphere—is entitled to freedoms of personal liberty and assemblies along with protections against oppression and inhumane treatment. From these cases alone, it is clear that rapid modernization and commercialization of the online sphere must be met with a sound digital human rights framework to ensure that such rights and liberties remain guaranteed for all inhabitants of UN member states. By utilizing a case study approach we hope to bring a level of specificity to the discussion both in terms of specific UDHR articles as well as the shared role of corporate and governmental actors in preserving and advancing economic and political human rights in all regions of the world. We will conclude with recommendations for further study as well as recommendations for policymakers to address the impacts of AI more concretely and specifically from a whole-of-society approach with the appropriate stakeholder engagement.

2. Can We Extend the Existing Human Rights Framework into Cyberspace?

In her remarks cited above, Michelle Bachelet has highlighted the existing human rights framework under the Universal Declaration and corollary conventions and treaties serve as a solid "legal foundation on which States and firms can build their responses in the digital age," to includes guidance on acceptable behavior in the digital age (Office of the High Commissioner for Human Rights, 2019). While it is sensible to leverage and expand upon existing processes where necessary, we question whether the existing frameworks and instruments can be extended to the digital age fully, especially given the fact that human-rights law is predominantly State-centric. The Toronto Declaration comes closest to achieving Bachelet's model of leveraging existing international human rights law from a standpoint of the right of equality and non-discrimination by tasking States to serve as fundamental guarantors of human rights obligations, with corporations using due diligence in building equity into artificial intelligence and machine learning technologies deployed across multiple industry domains (Access Now and Amnesty International, 2018). These models, however, do not necessarily fully address the tectonic shifts in extending existing international human rights law into cyberspace.

The ubiquity of digital platforms in democratic life is not disputed by most within the international community. However, Rikke Frank Jorgensen has noted:

> "[what] is less debated is the fact that facilitating this democratic potential critically relies on private actors [….] Despite the increasing role that these private actors play in facilitating democratic experience online, the governance of this social infrastructure has largely been left to companies to address through corporate social responsibility frameworks, terms of service, and industry initiatives such as the Global Network Initiative." (Jorgensen, 2018)

This extended social-corporate relationship that permeates everyday life in the digital age does not rely on an independent judiciary or a civil or criminal justice system that is bound to protect an individual's rights under the UDHR, but rather corporate oversight and mechanisms that provide neither due process nor transparency while being based upon norms of corporate social responsibility rather than human rights law. Prima facie evidence shows that these technology companies are inherently incapable of self-regulating when asked to do so. Such was the case of Google, which has recently gutted its own Ethical AI team barring any sound justifications (Fried, 2018). Jorgensen goes on to note that the case law of the European Court of Human Rights (ECtHR) "confirms that States have an obligation to protect individuals against violations by business enterprises," thus outlining a potential conflict of interest within human rights law in governments and state actors are also customers of the same corporations through procurement contracts (Jorgensen, 2018). Momsen and Rennert have determined that utilizing digital platforms to create predictive policing systems for American and German police forces presents several conflicts of interest wherein the citizen is no longer confronted with the state (as victim or as a perpetrator facing justice), but rather a "non-transparent mixture of state authority and private factual or contractual power" (Momsen & Rennert, 2020).

Hin-Yan Liu points to another disruptive aspect of digital technologies, namely that "from a legally doctrinal point of view, there must be a direct causal connection between the actions or omissions of a State or its agents and the concrete enumerated right possessed by an individual for the human rights system itself to become engaged" (Liu, 2019). Drawing on ECtHR case law, Liu also notes that the concept of a "victim" requires causation and agency to access the justice system. However, the digital age is played out on the internet, which does not have the same boundaries of jurisdiction, sovereignty or even place as does the physical world. The internet, in Liu's alternative perspective, occupies three *loci* simultaneously: the *system*, the *network* and the *distributor and dissipator* (Liu, 2019). Consequently, existing frameworks of causality and agency do not function analogously. Particularly problematic from a human rights standpoint is a dynamic system composed of artificial intelligence and algorithms rendering systemic--rather than man made--failures. This can be even more problematic when looked at from a system's underlying infrastructure "that prejudices towards infringements against human rights or makes their occurrence more probable" (Liu, 2019). Given artificial intelligence's underlying distributed architecture across multiple data centers, databases, big data, the cloud, machine learning algorithms that make sense of this data and provide predictive models do so at multiple

instances and nodes: each with the ability to provide biased datasets or biased judgments incorporated into the algorithmic logic at each node.

Similarly, the *network effects* of the internet can result in an accumulation of small wrongs (such as racially biased data used in facial recognition systems used by police forces), amplified via artificial intelligence, even though each individual infraction may not in and of itself rise to the threshold of severity required by courts or tribunals (Liu, 2019). This 'death by a thousand cuts' phenomenon is further compounded by the complexity of code, servers, algorithms and databases that can self-manifest an *entire system* of additional bias, decision-making, and independent action. That these are decentralized and subject to change by corporate architects and designers underscores the internet's role as *distributor and dissipator*, which further blurs causation and exacerbates power differentials between the individual and the alleged human rights violator (Liu, 2019). While the UN Human Rights Council appears to address this on some level in its Guiding Principles on Business and Human Rights, it nonetheless resides on a dualistic framework: on the one hand, States are obligated to protect and fulfill human rights, while on the other, business enterprises are required to act as specialized organs of society that respect human rights under the guidance of States, particularly in conflict zones (Office of the High Commissioner for Human Rights, 2011). In this model, the State is the principal guarantor of its citizens' human rights, while businesses are primarily held responsible for monitoring, due diligence, risk mitigation, compliance, and reporting to states, in whose hands remediation ultimately rests.

What is problematic about this vertical integration of State and corporate agency and remediation is that not all human rights violations caused by digital businesses will fit the causality and agency model required by human rights law, but that the very nature of digital business as *system, network, and distributor and dissipator* is a horizontal integration of decision-making and amplification of potential harm that does not fit state-based grievance mechanisms. Expecting corporations and the internet to self-govern and self-regulate against potential human rights abuses is, as Jorgensen noted above, a conflict of interest -- and one that does not adequately address the "death by a thousand cuts" that multinational distributed networks can cause across multiple victims. As it stands, the full impact of human rights infringements and violations resulting from AI systems is still unknown (Latonero, Governing Artificial Intelligence: Upholding Human Rights & Dignity, 2018), especially as discrimination (from an American civil rights perspective) can be an "artifact of the data mining process itself, rather than a result of programmers assigning certain factors inappropriate weight" thus disadvantaging protected classes in ways that are harder to both enumerate and remediate (Barocas & Selbst, 2016). This is particularly problematic in the context of predictive policing (not only in the West, but in China as we shall see below) since this strips potential defendants of the presumption of innocence enshrined in UDHR article 11, and even remedies, which have been based largely on intent, which is nigh impossible to determine through networked algorithms (Momsen & Rennert, 2020).

We agree with the assertions of Momsen & Rennert and others that a fundamental re-envisioning of the existing human rights framework is needed to address a concept of digital citizenship. The COVID-19 public health measures across the world have put additional pressure on democratic rights in an era of rapid digitization as entire lives are

now lived online due to quarantine orders. In the context of education and youth participation in civil and political life during the COVID-19 pandemic, Buchholz et al (2020) have noted that this has resulted in a shift from "digital literacy" to a more fundamental and expansive re-envisioning of "digital citizenship" from an experiential point of view that will extend post-COVID-19. Being a digital citizen "requires individuals to confront complex ideas about the enactment of identities online as citizens who collectively work for equity and change" in a democratic society (Buchholz, DeHart, & Moorman, 2020). Digital citizenship also differs fundamentally from traditional notions of citizenship as it is performative and defined through actions, "rather than by their formal status of belonging to a nation-state and the rights and responsibilities that come with it" (Hintz, Denck, & Wahl-Jorgensen, 2017). In the next section, we will examine two case studies of how "digital citizenship" is performed on the global stage as examples of how to re-conceptualize this against the backdrop of traditional human rights law.

3. Towards "Digital Citizenship" in Estonia and India

In the era of digital governance, more and more essential tasks and services—both in the public and private sphere—take place in the digital realm. Consequently, countries are beginning to adopt digital identification frameworks as a means to simplify access to digitized government services. Yet although thus far such schemes have not conferred the full rights of physical citizenship, the fact that they are necessary for so many essential services highlights the need for digital rights to be considered alongside physical rights, and that there ought to be a clearer definition of citizenship that better encompasses our interactions in the digital world. To this end, there are two particular case studies that highlight both the advances that have been made and the pitfalls that have been encountered—particularly in relation to the UDHR—in moving towards a concept of digital citizenship: e-Estonia and India's Aadhaar program.

3a. Estonia and the Creation of a Concept of Digital Citizenship

In 1991, when the Soviet Union collapsed, Estonia was at a severe technological deficit to the point where less than half the country had a telephone line and the majority of state infrastructure dated back to the 1930s (Hoe, 2017). In determining the direction to take with their newfound independence, the Estonian government thus opted to devote itself to technological advancement and digital reform as a cornerstone of their national development. Over the next two decades, Estonia made enormous progress in this realm, some major milestones of which include the digitization of nearly all schools by 1997, the declaration of internet access as a human right in 2000, and the introduction of online voting in 2007—an international first (Hoe, 2017). In the years since, the country has established e-Estonia: a digital society wherein 99% of government services are online and 99% of residents have an electronic ID card that grants them access to a variety of needs (e-Estonia Home Page, n.d.). Almost everything is now done digitally, ranging from paying for parking and signing legally-binding documents to accessing homework and filing taxes (Hoe, 2017). For this system to function, the country has one of the fastest internet

connection speeds in the world with coverage that extends to most remote areas (as one tourism campaign slogan touted: “Estonia – WiFi in the Forest”) and is free of charge (Ellis, 2020; Hoe, 2017). Additionally, rather than keeping everything centralized on one server, e-Estonia is a decentralized system wherein each government agency can create their own server designed to best suit their individual needs; and X-Road is the digital system that allows these databases to interact (Hoe, 2017).

Source: Estonian World

While there have been concerns about privacy and various data breach issues over the years, the e-Estonia system has been honed and developed in such a way as to create numerous safety valves to protect against such risks moving forward. Additionally, the Estonian government and people have placed great value on the concept of mutual accountability and thus have worked to create a highly transparent environment wherein although citizens’ every interaction with state administrations are linked to a single state-verified identity, every citizen can also access their personal data and are entitled to hold the administration accountable if they find that their data has been accessed without their consent (Berson, 2018). It can thus be argued that a great deal of Estonia’s digital success can be attributed to the construction of a trusting relationship between the country’s government and its citizens (Berson, 2018).

That said, this relationship has become strained in recent years as Estonia has seen a significant rise in populism and right-wing extremism, with support for the right-wing Conservative People’s Party of Estonia (EKRE) growing from a mere 2% in 2010 to around 20% by 2019 (Veebel, 2019). In the parliamentary elections of March 2019, EKRE received the third most votes; a strong result that not only allowed it to join the governing coalition, but also to become a dominant voice in determining the nation’s political agenda (Veebel, 2019). As an ethno-nationalist party, EKRE’s primary concern is the survival of Estonian ethnicity and culture –a goal that they believe is best accomplished through social

conservatism and an anti-immigration platform. This has created problems for a country that has been seeking to attract more foreign entrepreneurship and high-skilled workers. For example, two fairly recent developments via e-Estonia have been the advent of e-residency and the digital nomad visa; the former being a transnational digital identity that allows foreigners to use Estonia's e-services and access the EU business environment, and the latter being a visa specifically geared towards remote workers so they may live in Estonia while legally working for employers registered abroad ('Digital Nomad Visa,' n.d.). However, in January 2021 Estonia's Prime Minister resigned due to a corruption scandal, paving the way for Kaja Kallas of the Reform party to take over as PM and to effectively remove the EKRE from Parliament as the Centre and Reform parties formed a new governing coalition (Walker, 2021).

The case of Estonia is significant for a few reasons. First and foremost, their decision to declare internet access as a human right indicates an awareness that as society becomes increasingly digitized, internet access is a necessity for securing basic human needs--many of which are already enshrined as rights by the UDHR, most notably in article 25 ("Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control"). This is vitally important considering that numerous governments around the globe have resorted to shutting off internet access during times of dissent, effectively barring people from not only accessing essential goods and services but also from exercising their full political and economic rights. It is thus critical to reconceptualize 'citizenship' as something which encompasses both physical and digital personhood since violations of the latter infringe on human rights yet are not adequately addressed by current international human rights law. Estonia's push to ensure transparency and government accountability for potential misuse of their digitized systems is indicative of a move in this direction, and thus a positive example of a country developing and embracing digital rights.

Estonia's advent of e-residency is also groundbreaking in that, while it does not confer the full rights of citizenship, it is the first time that a government has acknowledged some form of citizenship that is not based on 'blood and soil.' While this concept is one that has met the ire of populist and nativist movements, thus far its rollout in Estonia has gone fairly smoothly and has generated significant interest abroad. The notion of allowing non-citizens who live outside the physical borders of a country to enjoy certain in-country benefits is a novel one, and certainly one that will become more relevant as increased global digitization blurs the relevance of physical borders. This also exists in stark contrast to the case of India, whose digital citizenship scheme—although similarly not conferring the rights of physical citizenship—has arguably acted as a means to discriminate against and exclude certain demographics living within national borders.

3b. India: Digital Citizenship and the Politics of Exclusion

While Estonia's push towards a model of digital citizenship has by and large been successful, India has encountered significantly more challenges in this regard. Inspired in part by Estonia's success (Berson, 2018), in 2009 the government of India established the

Aadhaar program wherein residents are voluntarily assigned a random 12-digit unique number (and a matching registration card) that serves as a digital ID for the purpose of utilizing governmental welfare and social services (UIDAI, n.d.). To obtain an Aadhaar number, residents must provide a variety of demographic and biometric data including name, date of birth, address, fingerprints, iris scans, and a facial photograph; and while it is considered proof of identity, it is not linked to citizenship and thus does not confer any citizenship rights upon Aadhaar holders (UIDAI, n.d.).

Source: DNA India

By creating a nationally centralized identification system, Aadhaar was intended to benefit society in a myriad of ways, including by better preventing identity theft and fraud, streamlining taxation, allowing greater inclusivity for impoverished and marginalized residents (who often lack the documentation otherwise required to receive state benefits), and potentially saving the government billions in excess expenditures (UIDAI, n.d.). Since its creation over a decade ago, Aadhaar has been widely adopted across the country with over 1.25 billion enrolled—including over 99% of the adult population—making it the largest biometric identification system in the world (*India Today*, 2019). Yet despite the impressive rollout of this program, it remains highly controversial and has raised numerous concerns. In addition to the major privacy concerns of having the biometric data of millions of individuals stored in government data centers, there is the issue that many marginalized communities whom Aadhaar was intended to help have actually been put at a greater societal disadvantage due to this program.

Aadhaar is supposed to be a voluntary system; however, enrollment has become coercive since having an Aadhaar number is a mandatory requirement for accessing

numerous essential services (Crawford et al., 2019). Consequently, those who are unable to enroll for whatever reason are deprived of access to vital needs. For example, some reports have found that rural and migrant children lacking birth certificates are routinely denied admission to government schools since an Aadhaar number has become a requirement for enrollment and they are unable to prove their identity to the extent needed for Aadhaar registration (Ghosh, 2018). A common perception in India is that private schools offer a superior educational experience, thus those that can afford it generally opt to send their children to such schools. This means that government schools primarily exist to serve the least privileged who are financially vulnerable and thus in the greatest need of help; yet it is exactly those people whom the Aadhaar requirement targets and harms -- a course of action that violates the right to education as guaranteed under article 26 of the UDHR. Additionally, individuals suffering from conditions like leprosy that often result in loss of fingers or sight have been refused welfare payments and social services on the grounds that they cannot prove their identity without proper iris scans or fingerprints (Ghosh, 2018). Some welfare-benefits denials due to either a lack of Aadhaar enrollment or technical failures concerning identity authentication have even resulted in higher levels of malnutrition and starvation deaths (Crawford et al., 2019). It is also challenging for impoverished individuals to enroll in Aadhaar since enrollment requires an address, yet countless residents—particularly those living in slums—lack an official address (Ghosh, 2018). Compounding this is the fact that many services like applying for utilities or opening a bank account require an address as well; so even if someone is able to obtain an Aadhaar number, that is not always enough to ensure access to Aadhaar-dependent services (Ghosh, 2018).

Some of these issues have been challenged in India's Supreme Court; yet despite consistent rulings that the government cannot make Aadhaar cards/numbers mandatory for access to state programs and services, the federal government has continued to introduce mandatory Aadhaar registration for numerous things ranging from state board exams and senior railroad passes to applying for government jobs and accessing pension schemes (John, 2017). This has spilled over to the private sector as well, with corporate actors increasingly requiring an Aadhaar number before letting customers utilize their services—for example, to open a bank account or purchase a cellphone contract (Perrigo, 2018). To this end, however, the Indian judiciary ought to be commended for taking steps towards protecting citizens and residents from private companies and corporate actors by instituting a landmark Supreme Court ruling in 2018 that private companies could no longer require users to provide their Aadhaar details as a condition of service (Perrigo, 2018).

That said, the Aadhaar program and the way it has been implemented highlight other major human rights issues within India. In 2019 the Bharatiya Janata Party (BJP)—India's ruling party that has increasingly sought to establish the country as a Hindu nationalist state—passed the Citizenship Amendment Act (CAA) which provides a route to citizenship to members of certain religious minority communities from neighboring Islamic countries, but not for Muslims themselves. This is considered by many to constitute religious discrimination and is arguably in violation of articles 6 ("Everyone has the right to recognition everywhere as a person before the law") and 7 ("All are equal before the law and are entitled without any discrimination to equal protection of the law") of the UDHR. In addition to this, the BJP is pushing for an update to the National Register of Citizens (NRC)

which will require millions of people (namely those who are not listed in the 1951 NRC and their descendants; or those who were listed but suspected of being foreigners) to provide documentation proving their citizenship (Saha, 2019). Consequently, many Indians—particularly Muslims, who cannot use the CAA to regain citizenship if rejected from the NRC—are concerned about losing their citizenship and either being deported or relegated to a detention center meant to house refugees (Changoiwala, 2020). Furthermore, those who are excluded from the final NRC list will not be able to enroll in Aadhaar anywhere in the country since their biometrics will be flagged as belonging to foreigners (Masiero, 2019). This last point is problematic for two main reasons: 1) the Aadhaar Act clearly states that Aadhaar enrollment is to be based on proof of residency, not citizenship, and 2) by barring non-citizens from Aadhaar enrollment, countless residents will be denied access to essential government services (Masiero, 2019).

This conflagration of the CAA, NRC, and Aadhaar effectively amounts to a scheme for identifying immigrants and Muslims, stripping them of their citizenship, and denying them their political and economic rights. This arbitrary deprivation of nationality is a clear violation of article 15 of the UDHR ("Everyone has the right to a nationality" and "No one shall be arbitrarily deprived of his nationality nor denied the right to change his nationality") while the detention and deportation of those rejected from the NRC amounts to a violation of articles 9 ("No one shall be subjected to arbitrary arrest, detention or exile") and 13 ("Everyone has the right to freedom of movement and residence within the borders of each state" and "Everyone has the right to leave any country, including his own, and to return to his country"). Moreover, the blatant religious discrimination born of these policies as well as the economic and political hardship brought about by denying residents Aadhaar access infringe upon the rights guaranteed by UDHR articles 2 ("Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status"), 7 , 25, and 26 ("Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages"). These issues are compounded by the Indian government's utilization of mass internet shutdowns to combat dissent; in fact, India shuts down its internet more than any other democracy (Nazmi, 2019) and since 2012 has done so 472 times—with 254 of those taking place in the Muslim-majority provinces of Jammu and Kashmir ('Internet Shutdowns,' n.d.).

Like e-Estonia, while Aadhaar in and of itself does not confer citizenship rights, as a nationwide system of digital identification it represents one of the world's first forays into the realm of digital citizenship (albeit tacitly). In comparing these two cases, it is important to note that India arguably faced far more obstacles than Estonia in their digital rollout due to major differences in demographics, development, and geography. Not only does India dwarf Estonia both in size and population, but India has a significantly higher proportion of rural residents as well as less digital infrastructure in rural areas (United Nations Department of Economic and Social Affairs, 2019). According to the United Nations' most recent Human Development Report, 89.4% of Estonia's population are internet users versus 34.5% in India ('Internet users, total,' 2020). Class stratification, gender equality, and poverty and education levels are also markedly different in these countries, with India ranking far lower on the Human Development Index (HDI) than Estonia (United Nations Development Programme, 2020). It is perhaps unsurprising then that insofar as human rights are

concerned, Aadhaar has encountered far more problems than e-Estonia; however, such digital ID systems--particularly in the global South--are actively encouraged and supported by international organizations like the World Bank as development priorities in fulfillment of UN Sustainable Development Goals (Crawford et al., 2019). The issues of exclusion and discrimination that this system raises thus further highlights the importance of establishing an international definition of citizenship that applies equally to both the physical and digital realms, since the result of Aadhaar is that many people who *are* citizens are ultimately being denied their rights due to inequalities in the way that these digital systems are being applied. Aadhaar also serves as a demonstration of the powerful role that corporate entities play in digitization and the ways in which that power can be used to infringe upon human rights. Until the aforementioned Supreme Court ruling, private companies in India were able to deny individuals access to essential services based on inclusion/exclusion from a federally digitized system; and although the government intervened in this instance, a lack of international structure surrounding digital rights means that federal oversight of private companies is often minimal in the area. Instead, private companies and corporations often operate under a system of self-regulation, which, as the following case study of China will demonstrate, can have severe consequences.

4. Cruel and Inhumane Punishment in the Context of Digital Citizenship: China

As demonstrated in the previous section, India is actively in the process of recreating a concept of "digital citizenship" that is favorable to a rising Hindu ethnostate through processes of exclusion and closure of civic space that prevents minority populations from full access of their civil and political rights under UDHR articles 2, 6, 7, 13, 15, 25 and 26. The arbitrary detention and deportation of those rejected under the biometric identification scheme mirrors an even more corrosive case study: that of China's treatment of Uighur and Muslim minority populations as genocide. The United States was the first country to make the designation of genocide on the last day of the outgoing administration in 2021; this was reaffirmed by the newly confirmed US Secretary of State Antony Blinken in January 2021 (Brennan, Ruffini, & Schick, 2021). According to the UN Convention on Genocide (UN Office on Genocide Prevention) as well as the Rome Statute (International Criminal Court, 2002), genocide has the intent to destroy either in whole or in part a national, ethnic, racial or religious group through serious bodily or mental harm, forced sterilization, as well as forcible transfer of children from one group to another. While the human rights violations in China's "re-education" camps for Uighurs have been extensively documented, including systematic rape, sexual abuse and torture, most recently by the BBC (Hill, Campanale, & Gunter, 2021), the violation of UDHR articles 5 ("no one shall be subjected to torture or to cruel, inhuman or degrading treatment or punishment") and 9 ("no one shall be subjected to arbitrary arrest, detention or exile") goes far beyond the physical camps and extends into systematic denial of digital citizenship rights in a years' long campaign of digital repression and detention.

Human Rights Watch began reporting on the Uighur situation in 2010, when Cambodia forcibly repatriated in 2009 twenty Uighur asylum seekers who expressed fears

of persecution and torture upon return to China. While granted “Persons of Concern” status, their refoulement after the violence of 2009 in the context of forced disappearances, arbitrary detentions and politicized judicial proceedings was unjustifiable even then (Human Rights Watch, 2010). Coupled with the formal introduction of the Social Credit System in 2011 in China, the ability to monitor and detain all 1.3 billion Chinese citizens was gaining steam in 2015 to exert societal control over schoolwork, medication adherence, parking and traffic violations, and predictive policing with a goal of nationwide deployment by 2020 (Associated Press, 2015). The systematic discrimination and targeting of Uighur Muslims accelerated with the ability to arbitrarily detain them digitally by leveraging the centralized Social Credit System and its dual nature as both a system and network of repression.

The word “credit” in this context (*xinyong*) refers to a moral concept from Confucian ethics that, ironically, stems from ideals of honesty and trustworthiness, later extended to financial credit scores, as in the West. However, it has extended beyond that to a reward and punishment system based on points added for good deeds (prosocial actions such as volunteer work) and points subtracted for antisocial deeds (such as criminal actions) that, when linked to a citizen’s ID card number, function as a monitor and arbiter of desired behaviors, to include social media monitoring, facial recognition systems and predictive policing systems (The Conversation, 2018). Centralizing data centers in Guizhou in 2015, the Chinese government recruited both American and Chinese digital platform companies to participate in the development of systems, networks and distributed data aggregation and algorithms, including Google, Microsoft, Apple, Baidu and Huawei (The Conversation, 2018). Touted by many techno-utopians as a way to hold government and business owners accountable, the human rights implications were largely disregarded, especially in the context of ongoing repression of Uighur Muslims.

Police surveillance of citizens of Xinjiang Autonomous Region (home of the Uighurs) is an unparalleled case study in persecution and large-scale internment of Muslims in the area through a complex use of online citizen identification numbers, biometric collection through purported public health campaigns, monitoring of text messages, facial recognition cameras, phone calls, banking records and social media monitoring in WeChat. When coupled with police informant networks, these “invasive surveillance techniques watch for signs of religious enthusiasm, which are generally equated with extremism” which can result in Uighurs being classified as terrorists if they apply for asylum abroad (Grauer, 2021). The “system of systems” architected by Landasoft includes 52 gigabytes and 250 million rows of data compiling multiple input feeds from *Jingwang Weishi*, an app Uighurs were forced to download to their phones, *Baixing Anquan*, a “public safety” app used to inform on Uighurs, evidence collection management from WeChat and Outlook and *ZhiPu*, a graphic interface of social network analysis, in addition to government Integrated Joint Operations Platform (Grauer, 2021). Importantly, while Uighurs are denoted with the attribute “iXvWZREN” there is no attribute for the Han Chinese, the majority ethnic group in China (Grauer, 2021). Uighurs subject to the “anti-terrorism sword” had their phones downloaded by police (sometimes multiple times per day), were subject to monitoring if they traveled outside of China (as were their friends and family), and were monitored to see if they attended political “loyalty” events and for contacts outside of Xinjiang. The coupling of both artificial and human intelligence as well as machine learning both amplifies the scope and scale of the

monitoring, which is tantamount to house arrest and detention applied to a targeted religious and ethnic minority group.

Source: The New York Times

The University of Toronto's Citizen Lab has also published details of an extensive digital espionage campaign against Uighurs that used malware injected into compromised websites, which Citizen Lab has identified as POISON CARP, also used against Tibetan groups by Chinese state authorities (Marczak, et al., 2019). Using social engineering techniques, the Chinese state authorities created personas with false identities at Amnesty International in Hong Kong to target staff members at a Tibetan human rights group on the pretext of sharing news links from American newspapers, which appeared benign (Marczak, et al., 2019). As a result, Tibetan human rights groups have responded by creating a Tibetan Computer Emergency Readiness Team (TibCERT) to not only improve security but to counteract the digital human rights violations with a digital human rights counter-response (TibCERT, n.d.). It is clear that while the Social Credit System is countrywide, its manipulation as a way to close off civic space, freedom of movement, freedom of expression and citizenship rights to Uighurs who have been arbitrarily detained based on AI-based predictive policing models exclusively, when not being applied to the ethnic Han Chinese, is a disproportionate and calculated digital genocide, in addition to the documented human rights abuses and genocide in the re-education camps.

As outlined in the introduction, the conflict of interest of corporate actors in this digital detention and repression makes them uniquely unqualified to provide any of the remediation foreseen by human rights law, which sees the state as guarantor of human rights in cases of corporate excess. In the case of the Uighurs, both American and Chinese companies were complicit and at minimum, did not exercise the due diligence recommended by legacy models of human rights. Furthermore, the facial recognition database MegaFace, which trained a whole new generation of face-identification algorithms, harvested its data from the American website Flickr, which had over 100 million photos and videos. Researchers from Yahoo and the University of Washington used this data set in conjunction with 300 research groups, including Google, Tencent and SenseTime, the latter having developed some of the Uighur monitoring and detention tools in China (Hill & Krolik, 2019).

The photos released by Yahoo from Flickr have generated a human rights controversy in the United States as well: residents of Illinois whose photos were used without permission are able to sue under Illinois Biometric Information Privacy Act of 2008 which imposes financial penalties for use of fingerprints or face scans without consent (Hill & Krolik, 2019). The fact that children's photos from the American Midwest are powering digital repression and detention tools used in the most egregious and technologically advanced genocide and torture regime in recent history using American technology platforms, while at the same time being declared a genocide by two successive American presidents effectively calls into question the efficacy of the traditional human rights regime in a digitized world. The case of China's digital repression as well as physical genocide of the Uighur Muslim is a true "death by a thousand cuts" that merits a re-examination of performative and agentic concepts of citizenship and self-determination in the era of AI-mediated state and corporate actions against individuals and what remedies are available under current human rights law. Given the complex interaction of cross-border state and corporate actors in this case, we reaffirm our position that is imperative to address the concept of *digital citizenship* (the individual in both the physical world and cyberspace concurrently) and the *role of corporate actors* (not just nation-states, who have remained weak actors in the digital space) in a critical re-examination of the human rights framework.

5. Conclusion and Way Forward

The case study of China's genocide and repression of the Uighur Muslim population—both in re-education camps and through the use of digital imprisonment through biometrics, facial recognition, and expanded use of the social credit system to monitor and limit movements of the Uighur population without due process—is an extreme example of the abuse of human rights through artificial intelligence and serves as a warning for other cases around the world. As we demonstrated, the complicity of both American and Chinese technology companies and universities in the development of the facial recognition technology in use affected not only the Uighurs, but American users (including children) who were swept up in the MegaFace data harvesting from Flickr. The digital imprisonment that the Uighurs face is not only a harbinger of the escalation of their repression to physical and psychological torture in the re-education camps, but is in and of itself cruel and inhumane punishment of 24/7 surveillance which is not normally seen outside of the most

Foucauldian panopticon models employed in predictive policing systems (Momsen & Rennert, 2020).

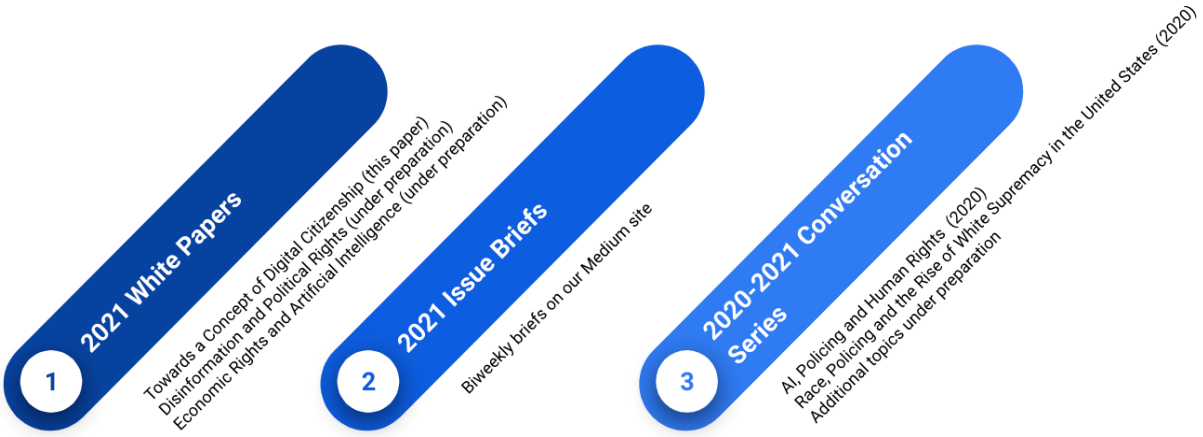
It is not difficult to see that the case outlined here of India's use of biometrics and the Aadhar system to be on a fast track to emulate China's social credit system, especially when both are being used in service of an ethno-nationalist state to disadvantage minority ethnic and religious populations. In both cases, these large-scale systems have been heralded by the regime in power as being advantageous to both the government and citizens (through rewards or improved services and benefits), but the deliberate exclusion of the Uighurs in China and Muslims in India from the supposed "benefits" of a newly digitized citizenship speaks more to a closing of civic spaces and a politics of exclusion. Biometrics and artificial intelligence not only define a digital citizenship that is performed online, but they also provide another method with which to discriminate, detain and imprison citizens digitally through internet shutdowns and 24/7 surveillance by ethnonationalist states.

Technology companies possess an authoritative political presence over modern-day international affairs. This has historically left enforcement of existing human rights frameworks in the digital realm largely devoid of any legal incentives for these corporations to comply with them. Moreover, rather than disagreements between States, the looming political interests of technology companies are also the principal roadblock against any discourse on the need for a digital human rights framework. The opaque admixture of state and corporate interests in predictive policing systems employed at scale in the US and Germany is taken to its logical extreme in the cases of India and China.

On the other hand, the case of Estonia outlined herein represents another option for the international community to examine. Conceived of as antidote to former Soviet authoritarianism and Russian encroachment on their physical territory and in cyberspace, the Estonian recognition of a digital e-citizenship and e-residency recognizes that a performative, even if limited, model of citizenship can, and is, exercised transnationally and online. This explicit recognition of a "digital citizenship" is unique in the world currently and merits significant discussion within the human rights framework from the standpoint of a performative digital rights that are not simply an extension of existing citizenship models envisaged in the UDHR and ICCPR. The Estonian model also allows us to interrogate the possibility of digital citizenship models that resist authoritarian erasure and repression, but also the limits and boundaries of this model in a digital world mediated and managed by corporate systems and networks.

It is in this context then that we are pleased to launch this initiative on artificial intelligence and human rights at the Center for International Human Rights at John Jay College of Criminal Justice. This white paper is the first in a series which will continue to look at additional focal areas (economic and political rights) regarding this new model of digital citizenship, and will be accompanied by shorter issue briefs as well as an online conversation series on the larger issues of artificial intelligence and human rights. Capitalizing on the work we started in 2020, we look forward to expanding our digital footprint through 2021 to encompass the work of our visiting scholars, graduate research assistants on the AI team at the Center, and guest posts and collaboration with other Center scholars and researchers in allied fields.

The Artificial Intelligence Team at the Center for International Human Rights kicked off with a presentation by Dr. George Andreopoulos and Dr. Michelle Strah to the Law Faculty at the Freie Universität Berlin in July 2020 on Race and Policing in the United States in the context of rising white supremacist movements, followed by a discussion of AI, Policing and Human Rights in the context of predictive policing in the US and Germany by Dr. Carsten Momsen and Dr. Michelle Strah in November 2020 (available on the CIHR YouTube page https://www.youtube.com/watch?v=K5i-L50gi3E) We look forward to three continued workstreams in 2021 and beyond, including the launch of our Medium page in support of our continued digital transformation:



For more information on the Center for International Human Rights at John Jay College of Criminal Justice, please visit us at our website (http://www.jjay.cuny.edu/center-international-human-rights), Twitter (https://twitter.com/JJCCIHR), Facebook (https://www.facebook.com/JJCCIHR/), Instagram (https://www.instagram.com/jjccihr/), YouTube (https://www.youtube.com/user/cihrjjay) and join our mailing list to get announcements about our upcoming Medium launch, publications, and initiatives.

REFERENCES

Access Now and Amnesty International. (2018, May 16). *The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems.* Retrieved from AccessNow: https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/

Ahktar, M. (2019). Police use of facial recognition technology and the right to privacy and data protection in Europe. *Navein Reet: Nordic Journal of Law and Social Research, 9*, 325-344.

Alston, P. (2019). *Report of the Special Rapporteur on extreme poverty and human rights.* UNGA, Promotion and protection of human rights. New York: United Nations.

Amnesty International. (2019). *Surveillance Giants: How the Business Model of Google and Facebook Threatens Human Rights.* London: Amnesty International.

Associated Press. (2015, November 24). *Big Brother is watching: how China is compiling computer ratings on all its citizens.* Retrieved from South China Morning Post: https://www.scmp.com/news/china/policies-politics/article/1882533/big-brother-watching-how-china-compiling-computer

Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review, 104*(671), 1-62.

Berson, G. (2018, November 4). e-Estonia: the ultimate digital democracy? *Medium*. https://medium.com/@geoffrooy/e-estonia-the-ultimate-digital-democracy-f67bc21a6114.

Brennan, M., Ruffini, C., & Schick, C. (2021, January 27). *With China's treatment of Muslim Uighurs determined to be genocide, Biden administration under pressure to act.* Retrieved from CBS News: https://www.cbsnews.com/news/china-treatment-of-muslim-uighurs-determined-to-be-genocide-biden-administration-under-pressure-to-act/

Buchholz, B. A., DeHart, J., & Moorman, G. (2020). Digital Citizenship During a Global Pandemic: Moving Beyond Digital Literacy. *Journal of Adolescent and Adult Literacy, 64*(1), 11-17.

Brauer, Y. (2021, January 29). *Revealed: Massive Chinese Police Database.* Retrieved from The Intercept: https://theintercept.com/2021/01/29/china-uyghur-muslim-surveillance-police/

Changoiwala, P. (2020, February 21). India's Muslims Are Terrified of Being Deported. *Foreign Policy*. https://foreignpolicy.com/2020/02/21/india-muslims-deported-terrified-citizenship-amendment-act-caa/.

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., …Whittaker, E.

(2019). *AI Now 2019 Report*. New York: AI Now Institute. Retrieved from https://ainowinstitute.org/AI_Now_2019_Report.html.

'Digital Nomad Visa.' (n.d.). Republic of Estonia: e-Residency. https://e-resident.gov.ee/nomadvisa/.

e-Estonia Home Page (n.d.). Retrieved from e-Estonia: https://e-estonia.com/.

Ellis, J.M. (2020, December 14). Democracy Versus Nation Branding In Estonia – Analysis. *Eurasia Review*. https://www.eurasiareview.com/14122020-democracy-versus-nation-branding-in-estonia-analysis/.

Foucault, Michel (1975). *Discipline and Punish: The birth of the prison.* New York: Vintage REP Edition, 1995.

Fried, I. (2021, February 19). Google fires another AI ethics leader. *Axios*. Retrieved from https://www.axios.com/google-fires-another-ai-ethics-leader-6ef7dcd5-4583-4396-b5b3-129547ff3091.html.

Ghosh, P. (2018, February 24). Aadhaar: In the world's biggest biometric ID experiment, many have fallen through the gaps. *Scroll.in*. https://scroll.in/article/868836/aadhaar-in-the-worlds-biggest-biometric-id-experiment-many-have-fallen-through-the-gaps.

Hill, K., & Krolik, A. (2019, October 11). *How Photos of Your Kids are Powering Surveillance Technology.* Retrieved from The New York Times: https://www.nytimes.com/interactive/2019/10/11/technology/flickr-facial-recognition.html

Hill, M., Campanale, D., & Gunter, J. (2021, February 2). *'Their goal is to destroy everyone': Uighur camp detainees allege systematic rape.* Retrieved from BBC News: https://www.bbc.com/news/world-asia-china-55794071

Hintz, A., Denck, L., & Wahl-Jorgensen, K. (2017). Digital Citizenship and Surveillance Society. *International Journal of Communication, 11*, 731-739.

Hoe, W. (2017, June 7). E-stonia: One Small Country's Digital Government Is Having a Big Impact." *Government Innovators Network*. https://www.innovations.harvard.edu/blog/estonia-one-small-country-digital-government-having-big-impact-x-road.

Human Rights Watch. (2010, December 17). *China: Account for "Disappeared" Uighurs.* Retrieved from Human Rights Watch: https://www.hrw.org/news/2010/12/17/china-account-disappeared-uighurs

International Criminal Court. (2002). *Rome Statute of the International Criminal Court.* Retrieved from ICC: https://www.icc-cpi.int/resource-library/documents/rs-eng.pdf

'Internet Shutdowns.' (n.d.). *Internet Shutdowns.in*. https://internetshutdowns.in/.

'Internet users, total (% of population).' (2020). United Nations Development Programme. Retrieved from United Nations Development Programme Human Development Reports: http://hdr.undp.org/en/indicators/43606.

John, A. (2017, January 17). Ten Things For Which Aadhaar Was Made Mandatory Even After October 2015 Supreme Court Order to the Contrary. *Caravan Magazine*. https://caravanmagazine.in/vantage/aadhaar-mandatory-supreme-court-order-2015.

Jørgensen, R. F. (2018). Human Rights and Private Actors in the Online Domain. In M. Land & J. Aronson (Eds.), *New Technologies for Human Rights Law and Practice* (pp. 243-269). Cambridge: Cambridge University Press. doi:10.1017/9781316838952.011

Latonero, M. (2018). Big Data Analytics and Human Rights. In M. K. Land, & J. D. Aronson, *New Technologies for Human Rights Law and Practice* (pp. 149-161). Cambridge: Cambridge University Press.

Latonero, M. (2018). *Governing Artificial Intelligence: Upholding Human Rights & Dignity.* Data & Society: New York.

Liu, H.-Y. (2019). The Digital Disruption of Human Rights Foundations. In M. Susi, *Human Rights, Digital Society and the Law: A Research Companion* (pp. 75-86). New York: Routledge.

Marczak, B., Hulcoop, A., Maynier, E., Razzak, B. A., Crete-Nishihata, M., Scott-Railton, J., & Deibert, R. (2019, September 24). *Missing Link: Tibetan Groups Targeted with 1-Click Mobile Exploits.* Retrieved from The Citizen Lab: https://citizenlab.ca/2019/09/poison-carp-tibetan-groups-targeted-with-1-click-mobile-exploits/

Masiero, S. (2019, September 12). A new layer of exclusion? Assam, Aadhaar and the NRC. *London School of Economics*. https://blogs.lse.ac.uk/southasia/2019/09/12/a-new-layer-of-exclusion-assam-aadhaar-and-the-nrc/.

Momsen, C., & Rennert, C. (2020). Big Data-Based Predictive Policing and the Changing Nature of Criminal Justice. *KriPoZ, 3*, 160-173.

Nazmi, S. (2019, December 19). Why India shuts down the internet more than any other democracy. *BBC News*. https://www.bbc.com/news/world-asia-india-50819905.

Office of the High Commissioner for Human Rights. (2011). *Guiding Principles on Business and Human Rights.* New York: United Nations.

Office of the High Commissioner for Human Rights. (2019, October 17). *Human rights in the digital age*. Retrieved from OHCHR: https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25158&LangID=E

‘Over 125 crore people now have Aadhaar cards: Government.” (2019, December 27). *India Today.* https://www.indiatoday.in/india/story/over-125-crore-people-now-have-aadhaar-cards-government-1631952-2019-12-27.

Perrigo, B. (2018, September 28). India Has Been Collecting Eye Scans and Fingerprint Records From Every Citizen: Here’s What to Know. *Time*. https://time.com/5409604/india-aadhaar-supreme-court/.

Saha, A. (2019, August 31). Assam NRC Explained: Add, delete and what next? *The Indian Express.* https://indianexpress.com/article/explained/assam-nrc-list-national-register-of-citizen-what-next-5803701/.

The Conversation. (2018, January 23). *China's Social Credit System puts its people under pressure to be model citizens.* Retrieved from The Conversation: https://theconversation.com/chinas-social-credit-system-puts-its-people-under-pressure-to-be-model-citizens-89963

TibCERT. (n.d.). *TibCERT*. Retrieved from TibCERT: https://tibcert.org/

United Nations Department of Economic and Social Affairs, Population Division. (2019). *World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)*. New York: United Nations.

United Nations Development Programme. (2020). *Human Development Report 2020 -- The next frontier: Human development and the Anthropocene*. New York: United Nations.

UN Office on Genocide Prevention. (n.d.). *Genocide.* Retrieved from UN Office on Genocide Prevention and the Responsiblity to Protect: https://www.un.org/en/genocideprevention/genocide.shtml

Veebel, V. (2019, July 31). The Rise of Right-Wing Populists in Estonia. *Foreign Policy Research Institute*. https://www.fpri.org/article/2019/07/the-rise-of-right-wing-populists-in-estonia/.

Walker, S. (2021, January 26). Estonia’s first female PM sworn in as new government takes power. *The Guardian*. https://www.theguardian.com/world/2021/jan/26/estonia-first-female-pm-appointed-as-new-government-takes-power.

‘What is Aadhaar.’ (n.d.). Unique Identification Authority of India (UIDAI). https://uidai.gov.in/what-is-aadhaar.html.