German Institute for Human Rights

ANNEX

Key Documents, Reports and Guidelines on Human Rights in the Data Economy

April 2020

The following assessment of key documents and reports from a wide range of stakeholders can offer insights into particularly relevant lines of argumentation to uphold human rights in the data economy. The assessment starts with the most relevant hard law sources, followed by key precedence cases, soft-law and multi-stakeholder initiatives, industry guidelines, principles and standards, industry self-regulation, codices of professional associations, individual company policies, and online hubs.

German Institute for Human Rights

Contents

1 Key Documents from Governmental and Inter-Governmental Bodies

1.1 UN Special Rapporteur on Freedom of Expression

The report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression (2018) stresses the need for all AI tools, and all technologies in general, to be "designed, developed and deployed so as to be consistent with the obligations of States and the responsibilities of private actors under international human rights law." With respect to business in particular, the Special Rapporteur (SR) highlights that companies have the responsibility to respect human rights, emphasizing the UNGPs as "global standard of expected conduct for all businesses wherever they operate" (principle 11), including social media and search companies, such as Facebook or Google.

Adapting the UNGPs to AI, the SR stresses that companies, "at a minimum, make high-level policy commitments to respect the human rights of their users in all AI applications (principle 16); avoid causing or contributing to adverse human rights impacts through their use of AI technology and prevent and mitigate any adverse effects linked to their operations (principle 13); conduct due diligence on AI systems to identify and address actual and potential human rights impacts (principles 17-19); engage in prevention and mitigation strategies (principle 24); conduct ongoing review of AI-related activities, including through stakeholder and public consultation (principles 20-21), and provide accessible remedies to remediate adverse human rights impacts from AI systems (principles 22, 29 and 31)" (10/22). The SR "strongly encourages the integration of human rights concerns into these efforts. The private sector's focus on and the public sector's push for ethics often imply resistance to human rights-based regulation" (16/22).

1.2 EU General Data Protection Regulation

The EU General Data Protection Regulation (GDPR) entered into force in May 2018, aiming to improve the protection of all EU citizens from privacy breaches. Its main features are a broader territorial scope when compared to prior regulation, substantial penalties for violations by corporations, a strong focus on user consent, and the introduction of a wide range of data subject rights (European Commission 2016). The extension of territorial scope means that the GDPR is applicable to every company that processes the personal data of data subjects residing in the EU, regardless of the company's location. The request for consent must be presented in an easy, clear, distinguishable format, with the purpose of data processing attached to that consent, and individuals need to be given the right to withdraw their consent at any point in time. The GDPR grants data subjects a wide range of rights, such as the right to be notified in case of a data breach, the right to access their data, the right to be forgotten (data erasure), as well as data portability across platforms and service providers. Moreover, technical and organizational measures have to be implemented in an appropriate and effective way to assure privacy by design. However, expecting the GDPR to be a panacea when it comes to protecting citizens from having their rights violated by data economy actors would be overly optimistic.

1.3 European Commission High-Level Expert Group on AI

The European Commission has been quite active in putting industrial policy regarding AI on the agenda. In December 2018, its High-Level Expert Group on Artificial Intelligence presented “draft ethics guidelines on trustworthy AI” (AI HLEG 2018). The document was intended as a starting point for the discussion of “trustworthy AI made in Europe,” and it puts forth the idea that an “ethical approach to AI is key to enable responsible competitiveness.” At the center of trustworthy AI lies a human-centric approach to technology that should respect fundamental rights while also ensuring that AI is built for an “ethical purpose” and in a “technically robust” fashion. The guidelines “do not aim to provide yet another list of core values and principles for AI, but rather offer guidance on the concrete implementation and operationalization thereof into AI systems.” Thus, they start from fundamental rights and principles but then address more specific technical challenges in the latter parts of the document. Public consultation had resulted in more than 500 comments by February 2019, and the revised version of the guidelines were published in April 2019 (AI HLEG 2019).

In February 2020, the EU Commission published three policy documents outlining priorities in its digital strategy: a white paper on Artificial Intelligence (European Commission 2020a), a communication on shaping Europe’s digital future (European Commission 2020b), and a European strategy for data (European Commission 2020c). Both the AI white paper and the data strategy are subject to public consultation. In a nutshell, the current EU approach appears to be swinging towards a risk-based approach, rather than prohibiting the use of certain technologies by default. That implies that the definition of what constitutes a “risk” and in particular a “high risk” is crucial for a human rights policy audience. Critics argue that such a risk-based approach would neglect systemic harms stemming from AI-based policies, such as on aggravating poverty or structural racist bias in digital welfare state system, as raised by the UN Special Rapporteur on Extreme Poverty (Pilkington 2019; Cath-Speth & Kaltheuner 2020).

1.4 Council of Europe Studies on the Human Rights Dimensions of AI

Against the background of the European Convention on Human Rights (ECHR), reports by the Committee of Experts on Internet Intermediaries (MSI-NET) and the Expert Committee on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence (MSI-AUT) have investigated the effects of algorithms on human rights. They point out that algorithmic decision-making and other automated systems affect several important rights, including the right to fair trial and due process, privacy and data protection, freedom of expression, freedom of assembly and association, access to effective remedy, freedom from discrimination, social rights and access to public services, and the right to free elections (Council of Europe 2018). The 2018 report is an excellent resource and makes three policy recommendations: (1) strengthening the effective transparency of automated systems, (2) reconsidering accountability and liability with regard to the producers of algorithmic systems, as well as (3) developing ethical frameworks and improved risk assessment procedures. Given the fact that the Council of Europe cannot make binding laws, these recommendations may lack regulatory momentum. However, several initiatives are already introducing ethical frameworks into the business and human rights debate, indicating that many debates on the protection of human rights are already taking place outside of official political channels. In addition to this document, the Council of Europe also had another committee address AI (Pielemeier 2019). The committee published a “Draft Recommendation of the Committee of Ministers to member States on

human rights impacts of algorithmic systems" (MSI-AUT 2018), the final version of which has since been published (Council of Europe 2019) and discussed critically (Hidvegi 2019).

2 Key Legal Precedence Cases

It appears that the more defined the relationship between rightsholder and company, the easier it is to provide protection for human rights, one example being surveillance at the workplace. Human rights protection gets more complex the more opaque and technologically sophisticated a business operates. It is thus more difficult to provide evidence for human rights abuses by business the less obvious and the more abstract a rights infringement is, which is common in the data economy, in particular with regard to bias and privacy. One example for this abstract rights infringement is discriminatory bias in hiring algorithms: Job candidates who were not selected for a job interview have a very low chance to get to know that their rejection was due to a racial or gender discrimination bias in a hiring algorithm. Additionally, even if they knew about this discriminatory bias, it would be almost impossible to present evidence without having first-hand corporate insights about the algorithm in question (burden of proof).

The following key precedence cases shed light on some of the legal pathways that human rights advocates might seek to explore for holding companies to account in the data economy.

2.1 European Court of Human Rights: Surveillance at the Workplace

Privacy at the workplace is one of the areas of human rights protection in the data economy that, due to its rather well-defined scope, might be better protected than others. A positive state obligation to protect the right to respect for the private and family life of the employee against the employer can be derived from Art. 8 ECHR (Grabenwarter & Pabel 2016: 65). However, efforts are increasingly being made to extend human rights violations to the private sphere by obliging private companies (Kälin & Künzli 2019: 3.24).

In cases regarding employee surveillance carried out by the employer, the ECHR has established six principles in the case of Bărbulescu vs. Romania (61496/08; September 5, 2017) concerned the decision of a private company to dismiss an employee after monitoring his electronic communications and accessing their contents. To assess whether a given monitoring measure is in conformity with or in breach of Article 8, the ECHR reflects on six elements: the prior notification to employees of the possibility and the implementation of such measures and the disclosure of information regarding their exact nature; the extent of the monitoring, meaning the degree of limitations in time and space as well as the number of people with access to the footage; the legitimate reason to justify the monitoring; the possibility of implementing less intrusive methods; the severity of consequences of the monitoring; and the provision of legal safeguards for the employees, i.e. in order for them to challenge the measures before an independent body (Strasbourg Observers 2019).

2.2 Early Precedence Cases Under the GDPR

In January 2019, the Austrian NGO *None of Your Business (NOYB)*, founded by privacy advocate Max Schrems, and French NGO *La Quadrature du Net* won a case against Google in France based on new regulation introduced by the EU General Data Protection Regulation (NOYB 2018a/b). The fine will amount to EUR 50 million for misleading users about alleged consent (NOYB 2019). Consent plays a central role in the processing of personal data, as it gives data subjects control over whether or not their personal data will be processed. The NGO case stresses the fact that there is a clear power imbalance between data subjects and corporate interests. As a result of the GDPR's new rules, user consent has to be given freely, informed, specific, and "distinguished from privacy policy and terms of service."

3 Competition Law

In some jurisdictions, national competition authorities have filed cases against big tech companies under competition law. Moreover, at the EU level, competition law cases have been brought against big tech corporations. These cases have been successful in providing protection against abusive practices, yet they are not following a clear-cut human rights protection logic. Some examples will be sketched below.

3.1 National Jurisdiction

In December 2018, Facebook was fined twice by Italy's Competition Authority for using personal data of its users for commercial purposes in ways that break Italian law (Hern 2018). Specifically, the Italian regulators found that Facebook had breached several articles of the consumer code (21, 22, 24, and 25) and hence misled its users in the sign-up process about how their data would be used for commercial purposes. Adding to this, Facebook was said to have forced "aggressive practice" on registered users by sharing their Facebook data with third parties for commercial purposes. Furthermore, Facebook was accused of disguising its practices by highlighting the free nature of the service and refraining from informing its users about the profit motive underlying the provision of the social network. Hence, Facebook was fined for violating fundamental concerns regarding free, prior, and informed consent.

3.2 Supranational Jurisdiction

Google has been the target of EU anti-trust authorities in at least three cases. Google has paid EUR 7 billion in penalties for two cases: EUR 4.3 billion in 2018 for abusing its dominant market position in Android operating system for smartphones (European Commission Directorate-General for Competition 2018), and another EUR 2.4 billion penalty for favoring its own shopping services over the ones of competitors (European Commission Directorate-General for Competition 2017). Both cases are pending at the European courts in Luxembourg. A third case focuses on Google's advertising business. Google AdSense places its search box on third-party websites, for example on news websites (European Commission Directorate-General for Competition 2019a/b). The Commission might fine up to USD 13 billion (10 percent of the global turnover of Google's parent company, Alphabet). However, estimates are that the penalty will be significantly smaller than the maximum. These three cases are relevant from a business and human rights perspective as market domination often correlates with abusive practices

towards privacy rights. Users often have no choice other than using Google's services and hence will agree to any terms and conditions put before them. (See also the judgement on alleged consent in the NOYB case against Google in France).

4 Soft Law and Multi-Stakeholder Initiatives

Over the course of the past few years, several initiatives have put forth documents focusing on guidelines, principles, and best practices regarding human rights and new technologies. This section provides details on a selection of these initiatives.

4.1 Toronto Declaration

At RightsCon 2018 in Toronto, a group of non-governmental organizations including Access Now, Amnesty International, Human Rights Watch, and the Wikimedia Foundation unveiled the Toronto Declaration (2018), which is aimed at ensuring non-discrimination in machine learning systems. The declaration focuses on issues such as transparency and equality, and it frames these issues in international human rights law. The declaration addresses both the public and private sector, making it highly relevant for current debates on technology from a business and human rights angle.

4.2 Montréal Declaration

The Montréal Declaration for a Responsible Development of Artificial Intelligence was introduced to the public in 2017 and finalized after stakeholder input in late 2018 (Montréal Declaration 2019). It promotes ten normative principles that should guide the development of AI, namely well-being, respect for autonomy, protection of privacy and intimacy, solidarity, democratic participation, equity, diversity and inclusion, prudence, responsibility, and sustainable development. Based on these general principles, the declaration makes eight specific recommendations, such as inclusive development of AI and the protection of democracy (Montréal Declaration 2018).

4.3 Necessary and Proportionate

The final version of Necessary and Proportionate—International Principles on the Application of Human Rights to Communications Surveillance was released in 2014. Supported and developed by a wide range of NGOs, it calls upon states to respect human rights in their surveillance practices (Necessary & Proportionate 2014). While the specific role of business in surveillance is at least addressed marginally in the Principles, its division between state and corporate surveillance is one that cannot be maintained in 2019. In fact, Western companies in many cases produce surveillance tools for nation-states, including dual-use technologies for regimes with questionable human rights records, which makes it abundantly clear that public-private partnerships in surveillance need to get addressed politically (Harris 2014; Schneier 2015; Penney et al. 2018; Mazzetti et al. 2019).

4.4 EQUALS

This international initiative for digital gender equality was founded by in 2016 by five institutions: the International Telecommunications Union, UN Women, the International Trade Centre, GSMA, and the United Nations University. Today, it has 90 international partners “dedicated to promoting gender balance in the technology sector by championing equality of access, skills development and career opportunities for women and men alike” (EQUALS 2019). In line with UN Sustainable Development Goal 5, the initiative aims at overcoming tech’s gender divide by 2030. To that end, they focus their empowerment efforts on three core issues: access, skills, and leadership.

4.5 The Global Network Initiative and the Telecommunications Industry Dialogue

The Global Network Initiative (GNI) features prominent member companies, such as Google, Facebook, Microsoft, Vodafone, and Orange, among others, along with civil society organization, academic institutions, and investors. As a result of adverse cases involving harm on human rights advocates emerging in the early 2000s, such as the case of Chinese journalist Shi Tao (MacKinnon 2007), various stakeholders recognized the need for a consensus on how to address the governance gap between corporate practices, domestic and international laws, and human rights laws, in particular regarding freedom of expression and privacy in the tech sector (Samway 2010). Thus, they founded the Global Network Initiative (GNI) in 2008.

The GNI is supposed to serve as a collective response to global threats to digital human rights, especially freedom of expression and privacy, and it is set up as a multi-stakeholder platform consisting of companies, civil society organizations, socially responsible investors, academic institutions, and individual experts (Samway 2010). In 2011, the Telecommunications Industry Dialogue (TID) was founded as a network of telecommunications companies (AT&T, Nokia, Millicom, Orange, Telenor, Telefónica, Telia, Vodafone), publishing their Guiding Principles on Freedom of Expression and Privacy (TID 2019). It later merged with the GNI.

The GNI has developed a set of principles to express a common understanding of good conduct among its members (GNI 2017). The principles are based on internationally recognized laws and standards for human rights, including the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and the International Covenant on Economic, Social and Cultural Rights, and they apply the UN Guiding Principles on Business and Human Rights as well as the OECD Guidelines for Multinational Enterprises.

The GNI principles touch upon freedom of expression and privacy in particular because these are the most salient issues in the industries of the GNI members, followed by a chapter on “Responsible Company Decision Making” and “Multi-stakeholder Collaboration” as well as “Governance, Accountability & Transparency”. GNI member companies’ implementation of the Principles and Implementation Guidelines is reviewed biennially by accredited, independent assessors who report their findings to the GNI’s multi-stakeholder board, which determines if they are implementing them in good faith and with improvement over time. The GNI publishes an assessment report after each assessment cycle that includes anonymized information in combination with some selected, attributed information.

A key criticism of GNI practices relates to the lack of a public and transparent independent assessment and evaluation of the implementation of the GNI Principles. The NYU Stern Center for Human Rights and Business published a blog post detailing its reasons to leave the GNI in February 2016 (Labowitz & Posner 2016). Among other issues, the Stern Center criticized the GNI for its failure to publicly disclose the monitoring of companies' compliance with its standards.

5 Industry Associations and Standards

In recent years, calls have grown louder for tech professionals to embed ethics into their everyday work in order to counter problematic practices within Silicon Valley companies (Fiesler 2018; Singer 2018). For instance, Eubanks (2018) suggested a "Hippocratic Oath for data science," and in a survey of more than 100,000 software developers, many professionals showed concern over bias in AI (Stack Overflow 2018). Moreover, tech activists share practices for using tech to achieve equity and justice (Costanza-Chock et al. 2018; Costanza-Chock 2020), and even Silicon Valley seems to have become less reluctant to address ethical issues (Pardes 2018). Against this background, major tech companies and professional associations have been working on ethics codices and guidelines, which we briefly discuss in the following sections.

5.1 Partnership for AI

The Partnership for AI (PAI) was founded in 2016 by five major Silicon Valley corporations and has grown to a network of more than 80 partners by early 2019 (Hern 2016). Despite its rather "corporate" beginnings, more than 50 percent of the network's partners today are non-profit organizations, including Amnesty International, Human Rights Watch, and Unicef (PAI 2019). Member organizations commit to eight "tenets" that form a fairly vague codex of normative ideals, such as open dialogue, stakeholder engagement, and trust. PAI's work centers around six thematic pillars: safety-critical AI; fair, transparent, and accountable AI; labor and the economy; collaborations between people and automated systems; societal influences of AI; AI and social good. For such a high-profile multi-stakeholder initiative, it seems remarkable that there has not been much substantial output yet. While the network has several active working groups, it has not yet published much meaningful normative guidance with respect to human rights.

5.2 ACM Code of Ethics

The Association for Computing Machinery describes itself as "the world's largest educational and scientific computing society" (ACM 2019). It represents more than 100,000 members, half of which reside outside the US, giving its guidelines considerable weight internationally. The ACM (1992, 2018) recently updated its "Code of Ethics" addressing the responsibilities of computer professionals. The updated code reminds anyone working in technology to do no harm and constantly "reflect upon the wider impacts of their work," making it clear that "the public good is always the primary consideration." Computer professionals should always "use their skills for the benefit of society, its members, and the environment surrounding them," and the code promotes a set of guiding principles in order to achieve

just that. With regard to human rights specifically, the code addresses computer professionals' obligations rather vaguely, stating that their role lies in "promoting fundamental human rights and protecting each individual's right to autonomy." The code does explicitly mention important rights, such as privacy and autonomy, but it does not elaborate all that much on either one of those due to the fact that it is supposed to be a concise guideline, not an all-encompassing reference document. In addition to its updated code of ethics, the ACM also has a standing Committee on Professional Ethics (COPE), which seeks to integrate this new code into computing curricula, as well as several special interest groups, the most relevant of which in the context of this report being SIGCAS, the ACM Special Interest Group *Computers & Society* (ACM SIGCAS 2020).

5.3 IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems is a program of The Institute of Electrical and Electronics Engineers, Inc. (IEEE), which describes itself as "the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity with over 420,000 members in more than 160 countries" (IEEE 2020). The IEEE Global Initiative brings together partners from academia, industry, civil society, and government in the autonomous and intelligent systems communities. Its mission is to ensure that every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity. In 2016, the IEEE Global Initiative produced *Ethically Aligned Design* (EAD, version 1), which represents the collective input of this community in the fields of autonomous and intelligent systems, ethics, and policy. The goal of the IEEE Global Initiative is that EAD and the IEEE standards it inspires will provide insights and recommendations that would become a key reference for the work of technologists in the coming years. EAD version 2 has been published after public consultation, and human rights are one of five major goals it focuses on (IEEE 2020).

6 Individual Company Policies

6.1 Google AI Principles

Google published its AI Principles in June 2018, partly in reaction to a highly criticized project with the US Department of Defense. CEO Sundar Pichai's (2018) blog post contains seven principles: (1) "Be socially beneficial;" (2) "Avoid creating or reinforcing unfair bias;" (3) "Be built and tested for safety;" (4) "Be accountable to people;" (5) "Incorporate privacy design principles;" (6) "Uphold high standards of scientific excellence;" and (7) "Be made available for uses that accord with these principles." The blog also outlines AI applications that Google claims not to design or deploy, such as harmful and lethal AI. While the principles as such and the areas of AI application that Google claims not to design and deploy could be considered a progressive step in the field, they can only be seen as a starting point. From a business and human rights perspective, the Google AI principles need to be strengthened in both language and real commitment, ideally through the use of proper human rights language instead of vague terminology about ethics, or Google will risk constantly being called out for "ethics washing" (Wagner 2018).

We would like to highlight three key criticisms: (1) The wording of the principles is strikingly vague. It relies on constructs that imply weighing decisions about potentially conflicting interests ("appropriate") and is built on a plethora of phrases that can be seen as "empty signifiers." For example, the phrase "unintended results that create risks of harm" does not specify any information about what kind of harm Google is talking about, and who exactly is harmed by such unintended results. (2) The principles are highly focused on the design of Google's AI products, yet they do not mention the power of using the actual data that Google AI products can gather. Moreover, grey areas such as data gathered by Google that is used for lethal applications of AI are not mentioned at all. (3) The principles do not encompass a section on intra-organizational "ethical behavior" at Google and between management ranks. The Google walk-outs in the fall and winter of 2018/19 have demonstrated that both senior and top management at Google have failed to take their staff's criticism about equal opportunities at Google into account. Matters have only escalated since, leading the New York Times to speak of "the great Google revolt" (Scheiber & Conger 2020). One particularly controversial issue in this context was Google's cancellation of its Ethics Board after just one week (Piper 2019).

6.2 Facebook's New Privacy-Focused Data Policy

Criticism of Facebook has been so profound over a considerable amount of time that it has its own, extensive Wikipedia (2019) entry. Since even before the company officially got started, it was controversial already, as its predecessor, "Facemash," caused much uproar in the Harvard community (Carlson 2010). Since then, privacy has remained at the heart of most controversies. Trying to justify his company's data collection practices in early 2010, Mark Zuckerberg, Facebook's founder and CEO, claimed that privacy was no longer a "social norm" (Johnson 2010). As a result of Facebook's rapid international growth and long list of scandals, Zuckerberg has since toned down this rhetoric in order not to endanger the company's social acceptance, which has been suffering noticeably since at least 2016 (Solon 2016). While Facebook has a long history of accidentally causing political effects outside of the US, e.g. in Egypt, India, or Myanmar (Roose & Mozur 2018), the company came under fire most prominently because of its alleged influence on the 2016 presidential elections in the US. Between 2016 and 2018, public criticism regarding Facebook's political influence had been growing, culminating in the Cambridge Analytica scandal that showed how easily user data could be misused for political purposes, and which has since been called a "privacy Chernobyl" due to its massive impact (Vaidhyanathan 2019a).

Since then, Facebook has been criticized heavily for its allegedly negative general influence on politics and society (Pariser 2012; Vaidhyanathan 2018; House of Commons 2019), and Zuckerberg has been issuing public apologies on a steady basis, including at parliamentary hearings on both sides of the Atlantic (Confessore 2018). In fact, Zuckerberg's apologies have become so frequent over the years that they now ring hollow because they never result in any meaningful change of the company's business model. As a consequence, the phrase "getting Zucked" is often invoked to mean "having been fooled yet again by Mark Zuckerberg into hoping that this time, there will be an actual change" (Chakravorti 2019). (Ethicist Michael Zimmer archives Zuckerberg's public speeches in the Zuckerberg Files, but access is limited to scholarly use, unfortunately.) Indeed, Zuckerberg has been criticized for acting like a king running his own nation-state (Farrell et al. 2019), which raises concerns over "digital constitutionalism" (Redeker, Gill & Gasser 2018).

Against this background, Zuckerberg (2019) recently announced that Facebook would take privacy more seriously and allow its users to focus on private conversations within their networks. Predictably, public backlash against this announcement was swift and fierce: it was called out as a purely self-interested business move aimed at market domination (Vaidhyanathan 2019b), and the Electronic Frontier Foundation (Kelley 2019) rightly pointed out that the real issue at stake is Facebook's ad policy and its business practices as a major advertising platform. Given the fact that Facebook and Google have become the dominant advertising companies in the world, building on a business model centered around gathering vast amounts of user data, it would seem prudent to assume that privacy will continue to be a problematic issue on the platform.

In January 2020, Facebook established an oversight board, with the bylaws being released to the public (Facebook 2020, see also chapter 5). It is to be welcomed that this body underwent a human rights review and is due to take the UNGPs into consideration in decision-making processes (BSR 2020). As a result, ideally, all human rights, not solely freedom of expression as well as personal safety and security, can be impacted by content decisions taken by Facebook, and this scope is said to be the one of the oversight board. However, the degree of autonomy between the oversight board and Facebook, along with its efficacy, has yet to be seen in practice. Critics are raising calls for transparency reporting about disclosure of cases by which community standards were violated, cases by format or content at issue (e.g., text, image, video, livestream), number of accounts and pieces of content covered by the cases considered by the board, and number of accounts/pieces of content taken down or otherwise actioned as a result of a board decision. Other improvements called for by civil society are increased transparency about the nomination of board members, due diligence between Facebook, the trust of the oversight board, and the oversight board itself, as well as government orders that threaten human rights (Ranking Digital Rights 2020).

7 Research Centers, Civil Society Organizations and Other Initiatives

In recent years, a wide range of institutions and initiatives have published guidelines, norms, recommendations, and online platforms in order to provide guidance for human rights implementation in the data economy (Jobin, Ienca & Vayena 2019; Fjeld et al. 2020). This section will briefly mention some of these hubs.

7.1 Academic Research Centers

There are too many academic research centers addressing important issues at the interface of technology and society for this short report to do their work justice. Among the established institutions, Harvard's Berkman Klein Center for Internet and Society stands out for publishing reports on the opportunities and risks raised by AI in the context of human rights (Raso et al. 2018; Fjeld et al. 2020). The center also takes part in a larger research initiative on AI and inclusion and the "Ethics and Governance of AI Initiative" (AI Ethics Initiative 2019), a cooperation with the MIT Media Lab, which is also a relevant player in this field, as is Yale Law School's Information Society Project. On the west coast, the Stanford Center for Internet and Society and the Markkula Center for Applied Ethics at Santa

Clara University address ethical issues arising from nearby Silicon Valley. On the east coast, there are interdisciplinary institutions that have been founded with the express purpose of addressing ethical issues around AI and Big Data, such as Data & Society and the AI Now Institute at New York University (AI Now 2018, 2019). In Europe, institutions such as the Oxford Internet Institute, the Humboldt Institute for Internet and Society (HIIG) as well as the Weizenbaum Institute, among many others, conduct important interdisciplinary research on the ethical aspects of the data economy.

7.2 Civil Society Organizations

As this report has made abundantly clear, the data economy introduces a wide range of challenges when it comes to human rights. No wonder, then, that a great number of non-governmental organizations have been addressing these challenges. Among many others, the most prominent NGOs in this context include Access Now, Algorithm Watch, Amnesty International, Geneva Digital Watch, the Open Rights Group, Privacy International, Ranking Digital Rights, and Tactical Tech. Moreover, the Business and Human Rights Resource Centre (2020) offers an excellent *Technology & Human Rights Portal*: a collection of resources on human rights and technology, including constantly updated media coverage. The site focuses on business and human rights issues in four in-depth areas: AI, automation, digital freedom, and the gig economy. Moreover, the Algorithmic Justice League (2019) is a project by MIT Media Lab grad student Joy Buolamwini, who was chosen as one of the top “35 Innovators Under 35” by MIT Technology Review (Beras 2018). The focus of this initiative lies in addressing racial and gender biases in facial recognition systems and other machine learning technologies. Instead of relying on traditional academic reports or policy papers, Buolamwini uses communication channels with low barriers to entry, such as art, exhibitions, video, newspaper op-eds, and other forms of community engagement (Buolamwini 2018).

7.3 Miscellaneous Initiatives

This section briefly lists a few initiatives that either offer only fairly vague guidance or have only just gotten started. For instance, the Australian Human Rights Commission and the World Economic Forum (2019) put out a white paper on AI governance. Moreover, Québec has been establishing the Observatoire international sur les impacts sociétaux de l’intelligence artificielle et du numérique (OIISIAN), which is aimed at interdisciplinary cooperation in AI governance (Hirsh 2018). At the international level, both the G20 and the OECD (2019) published AI guidelines that address human rights issues only in a fairly marginal fashion. Lastly, students at the MIT Media Lab and Harvard’s Berkman Center developed AI Blindspot, a discovery process for spotting unconscious biases and structural inequalities in AI systems, which is aimed at tech practitioners (MIT Media Lab 2019).