Privacy and Artificial Intelligence: Ranking Digital Rights Submission to Thematic Report on "the Right to Privacy in the Digital Age" from UN Human Rights

Ranking Digital Rights (RDR) welcomes this opportunity to provide input to UN Human Rights on the preparation of the thematic report on artificial intelligence (AI) and the right to privacy, as requested in HRC resolution 42/15 on "The right to privacy in the digital age," and informed by the 27-28 May 2020 expert seminar.[1] RDR is submitting a set of materials *(provided below*) in support of the present response. In particular, this submission highlights the centrality of privacy violations to the targeted advertising business model's reliance on AI, and provides RDR's transparency standards regarding the development and use of algorithmic systems for content moderation and governance. Our research has found that across the board, **companies disclose almost nothing about how they use AI, and even less about how they develop and train these systems**.[2] More recently, a report from the credit-rating agency FICO and Corinium, a consultancy, found that business leaders are putting little effort into ensuring that the AI systems they use are both fair and safe for widespread use.[3]

RDR works to promote freedom of expression and privacy on the internet by creating global transparency and accountability standards for companies to respect and protect human rights. We do this by producing the Ranking Digital Rights Corporate Accountability Index, which ranks the world's most powerful digital platforms and telecommunications companies on their policies and practices affecting fundamental freedom of expression, information, and privacy rights.[4] The RDR Index methodology is based on international human rights standards and frameworks, including the Universal Declaration of Human Rights,[5] the International Covenant on Civil and Political Rights,[6] the European Convention on Human Rights, and the U.N. Guiding Principles on Business and Human Rights.[7] To date, RDR has produced five RDR Indexes (2015, 2017, 2018, 2019 and 2020) and offers the only year-on-year ranking of these platforms. The 2020

[1] https://www.ohchr.org/EN/Issues/DigitalAge/Pages/SeminarArtificialIntelligence.aspx
[2] https://rankingdigitalrights.org/index2020/key-findings
[3] https://www.fico.com/en/solutions/fico-responsible-ai; see also https://www.zdnet.com/article/fico-report-finds-startling-disinterest-in-ethical-responsible-use-of-ai-among-business-leaders/
[4] https://rankingdigitalrights.org/index2020/
[5] "Universal Declaration of Human Rights," https://www.un.org/en/universal-declaration-human-rights/
[6] "International Covenant on Civil and Political Rights," *UN Human Rights Office of the High Commissioner*, https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx.
[7] "Guiding Principles on Business and Human Rights," *UN Human Rights Office of the High Commissioner*, https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

RDR Index was the first to include indicators focused on targeted advertising and algorithmic systems, the term we use for what many others call "AI."

Moreover, our 2020 two-part report series, "It's the Business Model," examines how **the targeted advertising business model—itself rooted in pervasive privacy violations at a global scale—drives business decisions about the design and deployment of AI systems** that undeniably cause and contribute to a range of human rights harms beyond the right to privacy. Chief among these harms are the right to freedom of expression and information and the right to non-discrimination, both of which can be uniquely threatened by artificial intelligence. Our recommendations highlight the importance of privacy and data protection, along with transparency and corporate governance, for promoting corporate respect for human rights.

As the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, noted in her April 2021 report on "Disinformation and freedom of opinion and expression," social media platforms and adtech firms (which are often one and the same) use opaque and unaccountable AI systems to curate, recommend and target both user-generated content and advertisements.[8] Though empirical evidence is all but impossible to obtain in light of companies' control of the required data, the consensus among outside experts is that these AI systems are designed to increase engagement with the platforms in question, which in turn maximizes the revenue that these firms can generate through targeted advertising.

These various AI systems rely on massive amounts of demographic and behavioral data collected within the platforms themselves, elsewhere on the internet, and increasingly in the physical world as well. In other words, many AI systems are designed for the purpose of increasing advertising revenue and corporate profits, and are made possible through indiscriminate mass surveillance at scale. Targeted advertising's value proposition to brands and advertisers is the ability to discriminate as precisely as possible among audiences and individuals in order to influence their behavior. While there are empirical questions about how well various companies fulfill their promises to their customers that cannot be answered without access to data that only social media and adtech firms possess, the value proposition itself infringes on the rights to privacy and to non-discrimination. Moreover, as highlighted by Ms. Khan: "content curation through powerful recommendations or microtargeting" is the "digital equivalent" of brainwashing and other coercive techniques for influencing people without their knowledge or consent. Such uses of artificial intelligence may therefore contravene the right to freedom of opinion.[9]

RDR's expertise as it relates to AI is its impact on privacy, expression, non-discrimination, and other human rights in the context of ICT platforms, notably social media sites and telecommunications operators. The basic point that many, perhaps most, AI systems rely on privacy violations at a massive scale is important to underscore. In order to mitigate some of the harms associated with algorithmic systems that curate, recommend and target content, digital platforms use a separate set of AI tools to identify content that violates either the platforms' own

[8] https://undocs.org/A/HRC/47/25
[9] https://undocs.org/A/HRC/47/25

rules or national law and remove it from the platform, in concert with human reviewers.[10] While such content moderation is necessary for these platforms to function and may be compatible with expression rights, depending on implementation details, AI enforcement of platform rules and national law is highly error prone, as existing tools cannot determine context or nuance. Moreover, AI systems for moderating content in English and other privileged languages are much more advanced, leading to uneven enforcement across languages. There are human rights risks when content is over-moderated (ie, false positive) and when it is under-moderated (ie, false negative) though different rights are implicated in each case. An in-depth analysis of these risks is beyond the scope of this submission: the point is that content moderation AI systems pose human rights risks, and that these risks are greater for speakers of languages that are not prioritized for AI development by companies.

The centrality of privacy violations to AI development is true in other domains as well, such as facial recognition and other types of remote physical surveillance. Seldom can the developers of AI systems vouch that their systems were built and trained using data whose subjects have given their free, informed consent to the specific use of their information. As participants in the May 2020 expert seminar noted, privacy plays a key role in safeguarding other rights affected by AI.[11] Ranking Digital Rights urges OHCHR and other stakeholders to focus on the right to privacy as an essential safeguard for all other rights impacted by artificial intelligence, across all sectors of human activity.

Materials for consideration

- "**It's the Business Model: How Big Tech's Profit Machine is Distorting the Public Sphere and Threatening Democracy**," a two-part series looking at the root cause of the proliferation of disinformation and other types of harmful content online: the targeted advertising business model.
    - **Executive Summary and Recommendations**
    - Part I - "**It's Not Just the Content, It's the Business Model: Democracy's Online Speech Challenge**"
    - Part II - "**Getting to the Source of Infodemics: It's the Business Model**"
- The **2020 RDR Corporate Accountability Index**, which evaluates 26 companies according to 58 indicators evaluating company disclosure of policies and practices affecting freedom of expression and privacy.
- The **2020 RDR Corporate Accountability Index methodology**, which includes new indicators on algorithmic systems and targeted advertising. The methodology reflects more than a year of stakeholder engagement, research, and pilot testing to develop

[10] The compatibility of these rules and laws with international human rights standards is a separate albeit enormously consequential question.
[11] https://www.ohchr.org/EN/Issues/DigitalAge/Pages/SeminarArtificialIntelligence.aspx

global accountability and transparency standards for how tech companies can demonstrate respect for human rights as they develop and deploy these new technologies.

- **Human rights risk scenarios:** a compilation of risk scenarios describing possible violations to human rights—as enumerated in the Universal Declaration of Human Rights (UDHR)—that could occur as a direct or indirect result of companies' targeted advertising policies and practices, and as a result of companies' development and use of algorithmic systems, respectively:

    - **"**Human rights risk scenarios: Targeted advertising," *Ranking Digital Rights* (2019): Scenarios illustrating the human rights harms related to privacy and expression that can result from targeted advertising business models and the company practices they incentivize.

    - "Human rights risk scenarios: Algorithms, machine learning and automated decision-making," *Ranking Digital Rights* (2019): Scenarios illustrating the human rights harms related to privacy and expression that can result from companies' use of algorithms, machine learning, and automated decision-making.

Transparency regarding the development and use of algorithmic systems for content moderation and governance: Using algorithmic systems to moderate and govern the dissemination of user content can have adverse effects on fundamental human rights, specifically, the rights to free expression, access to information, privacy, and non-discrimination.[12] Algorithmic content curation, recommendation, and ranking systems play a critical role in shaping what types of content and information users can see and access online. In addition, systems that are optimized for user engagement can have the effect of prioritizing controversial and inflammatory content, including content that is not protected under international human rights law. Over time, reliance on algorithmic curation and recommendation systems that are optimized for engagement can alter the news and information ecosystems of entire communities or countries. These systems can be manipulated to spread disinformation and otherwise distort the information ecosystem, which can in turn fuel human rights abuses. The development and testing of algorithmic systems can also pose significant risks to privacy, particularly when companies then use the information collected about users to develop, train, and test these systems without the data subject's informed consent.[13]

---

[12] "Consultation draft: Human Rights Risk Scenarios: Algorithms, Machine Learning and Automated Decision-Making," *Ranking Digital Rights* (2019), https://rankingdigitalrights.org/wp-content/uploads/2019/07/Human-Rights-Risk-Scenarios_-algorithms-machine-learning-automated-decision-making.pdf.

[13] Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York, NY, USA: PublicAffairs; Nathalie Maréchal. Targeted Advertising Is Ruining the Internet and Breaking the World, https://www.vice.com/en_us/article/xwjden/targeted-advertising-is-ruining-the-internet-and-breaking-the-world, *Vice Motherboard,* November 16, 2018; "Human Rights Risk Scenarios: Algorithms, machine learning and automated decision-making," *Ranking Digital Rights*, July 2019,

Online platforms that develop and deploy algorithmic systems therefore should:

- Disclose a clear commitment to uphold international human rights standards in their development and deployment of algorithmic systems (Indicator G1, Element 3), in line with the Council of Europe's Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems.

- Disclose evidence that they conduct regular, comprehensive, and credible due diligence, such as through robust human rights impact assessments, to identify how all aspects of its policies and practices related to the development and use of algorithmic systems affect users' fundamental rights to freedom of expression and information, to privacy, and to non-discrimination, and to mitigate any risks posed by those impacts (Indicator G4d).

- Publish policies that clearly describe the terms for how they use algorithmic systems across their services and platforms (Indicator F1d). Companies that use algorithmic systems with the potential to cause human rights harms should publish a clear and accessible policy stating the nature and functions of these systems.[14] This policy should be easy to find, presented in plain language, and contain options for users to manage settings.

- Publish information about whether they use algorithmic systems to curate, recommend, and rank content (Indicator F12). They should disclose how these systems work, what options users have to control how their information is used by these systems, and whether such systems are automatically on by default or users can opt-in to have their content automatically curated by the algorithmic system.

- Clearly disclose algorithmic system development policies in a way that users can easily access and understand, so that users can make informed decisions about whether to use a company's products and services (Indicator P1b).
- Clearly disclose that they provide users with options to control how their data is used for the development of algorithmic systems (Indicator P7, Element 7).
- Clearly disclose whether they use user data to develop algorithmic systems by default, or if users must affirmatively consent to such use of their data (Indicator P7, Element 8).

https://rankingdigitalrights.org/wp-content/uploads/2019/07/Human-Rights-Risk-Scenarios_-algorithms-machine-learning-automated-decision-making.pdf.

[14] "Addressing the impacts of Algorithms on Human Rights: Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems," Council of Europe, Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence (2019), https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf.

For more information

- Find out more about RDR and the RDR Corporate Accountability Index: www.rankingdigitalrights.org.
- View and download the 2020 RDR Index methodology: https://rankingdigitalrights.org/2020-indicators.
- Read a summary of our research process: https://rankingdigitalrights.org/2020-research-process.
- Read about how we develop the RDR Index methodology: https://rankingdigitalrights.org/methodology-development/.

Thank you for your consideration of these comments and the attached resources. We look forward to the opportunity to further discuss these resources and recommendations at your convenience.

Best regards,

Jessica Dheere, Director, Ranking Digital Rights, dheere@rankingdigitalrights.org

Nathalie Maréchal, PhD, Senior Policy Analyst, Ranking Digital Rights, marechal@rankingdigitalrights.org