RANKING DIGITAL RIGHTS

Rationale for expanding the Ranking Digital Rights Corporate Accountability Index to address algorithms, machine learning and automated decision-making

This document provides an overview of why and how Ranking Digital Rights (RDR) plans to expand the RDR Corporate Accountability Index methodology to address the impact of algorithms, machine learning, and automated decision-making on freedom of expression and information and privacy. It also introduces the accompanying *Human Rights Risk Scenarios* and *Best Practices* documents, which will form the basis for further research and development of pilot indicators.

All documents can be downloaded from the RDR website at:
https://rankingdigitalrights.org/methodology-development/2021-revisions/#algorithms.

Background: In May 2019, RDR published its fourth Corporate Accountability Index, ranking some of the world's most powerful internet, mobile, and telecommunications companies on their disclosed commitments and policies affecting users' freedom of expression and privacy. The 2019 RDR Index ranked 24 companies on 35 indicators. The RDR Index provides an effective roadmap for companies to improve policies and disclosures in order to prevent and mitigate a range of threats to users' rights to privacy and to freedom of expression and information. Since its inaugural launch in 2015 the RDR Index has contributed to improved company disclosure of policy and practice across a number of areas, including transparency reporting, content removals, account restrictions and shutdowns, and of handling and securing user information.

However, given geopolitical and technological developments with clear human rights implications in the years since the RDR Index methodology was first developed, it has become clear that the methodology needs to be updated if companies are to be held fully accountable for the range of potential online threats to human rights. Earlier this year, we undertook a process to expand the RDR Index methodology to include benchmarks that hold companies accountable for their targeted advertising policies and practices. This document describes the rationale for developing indicators that address companies' use and development of algorithms, machine learning and automated decision-making, thus enabling greater accountability.

Our goal is to set global accountability and transparency standards for how major, publicly traded internet, mobile, and telecommunications companies can demonstrate respect for human rights online as they develop and deploy these new technologies. In the future, RDR's work in this area can inform the work of other stakeholders: investors conducting due diligence on portfolio risk, policymakers seeking to establish regulatory frameworks to protect the individual

and collective rights of internet users, and activists looking to encourage companies to improve their policies and disclosures pertaining to algorithms, machine learning and related technologies that enable the automation of decision-making.

Scope of work: The current RDR Index contains 35 indicators divided into three categories: Governance, Freedom of Expression, and Privacy. These indicators focus on identifying and mitigating human rights harms primarily to individuals—specifically infringement of internet users' freedom of expression and information and privacy—that directly result from using the products and services of internet, mobile, and telecommunications companies.

For the 2019 RDR Index methodology we took an initial step toward addressing algorithms, machine learning, and automated decision-making by adding a new element to one indicator in the Governance category (G4) that seeks company disclosure of whether it conducts human rights due diligence covering all aspects of a company's business related to freedom of expression and privacy to identify and mitigate any potential human rights risks. The new element (G4.5) asks whether the company assesses freedom of expression and privacy risks associated with its use of automated decision-making, such as through the use of algorithms and/or artificial intelligence. The 2019 RDR Index found that only three companies made such disclosures.

For future iterations of the RDR Index, we plan to include new indicators that address specific harms to privacy and expression, as well as other human rights to individuals as well as communities, caused or exacerbated by companies' use of algorithms, machine learning, and/or automated decision-making. The *Human Rights Risk Scenarios* and *Best Practices* documents that follow this overview were drafted as the initial steps of our indicator development process. The *Human Rights Risk Scenarios* document contains a list of scenarios, each describing human rights harms directly or indirectly related to privacy and expression that can result from the use of algorithms, machine learning, and/or automated decision-making. They are derived from news reports and published research. The *Best Practices* describe specific standards of disclosure that companies should implement in order to mitigate the human rights risks associated with their use of algorithms, machine learning and automated decision-making.

Stakeholder consultation: Feedback from a wide range of experts and stakeholders is essential to RDR's methodology development process. After receiving feedback from experts, human rights advocates, and companies on these documents, and conducting further research, the risk scenarios and best practices will be used to develop draft indicators. These will be tested in a pilot study, which will enable us to further refine the methodology. Please send all feedback by **September 13, 2019** to methodology@rankingdigitalrights.org.