



The International Federation of Library Associations and Institutions welcomes the Special Rapporteur's initiative to examine the impacts of new information technologies on racial equality and equal enjoyment of human rights. We would like to thank the Rapporteur for the opportunity to contribute to this call for submissions and offer a perspective from libraries' experience and field of work and interest.

Numerous cases of biased or discriminatory outcomes stemming from the applications of new technologies, including Artificial Intelligence (AI) and algorithmic decision-making, have been documented and widely discussed in the recent few years. Prominent examples from high-stakes areas include criminal justice (e.g. predictive policing or recidivism prediction), risk assessment and resource allocation in healthcare, credit scoring or other forms of 'social ranking', allocation of social benefits, and more.

### **New communication technologies, access to information and non-discrimination**

Of particular importance for the library sector, due to its focus on access to information, are the possible discriminatory impacts on the information environment. The impacts of AI, algorithmic systems and other new information technologies on freedom of expression and intellectual freedom have been explored in several recent reports, particularly in the context of search engines and social media – e.g. the 2018 UN OHCHR Report of the Special Rapporteur to the General Assembly on AI and its impact on freedom of opinion and expression<sup>1</sup> and the 2018 report "Privacy and Freedom of Expression In the Age of Artificial Intelligence" by Privacy International and Article 19.<sup>2</sup>

Naturally, there are ways in which AI, Machine Learning (ML) and new technologies can play a positive role in improving access to information for linguistic and ethnic minorities. For example, both the Special Rapporteur's report and an IFLA trend report update<sup>3</sup> note that AI-powered translation tools can play an important role in helping individuals access information in more languages.

#### *Targeting and profiling*

However, some concerns have been expressed about the possible negative impacts of AI, ML and other new technologies on the information environment. The 2018 Special Rapporteur's Report notes that AI-powered targeting can perpetuate discrimination by excluding ethnic minorities (among other vulnerable groups) from accessing information or opportunities.

A recent experiment, for example, showed differences in the delivery of employment and housing advertisements – even when there were no differences in how the advertiser chose to target them. An experimental set of advertisements for janitorial and taxi driver positions, for example, were delivered to black users more than white ones (e.g. the estimated total fraction of white audiences

---

<sup>1</sup> "Report of the United Nations Special Rapporteur to the General Assembly on AI and its impact on freedom of opinion and expression", 2018, <https://undocs.org/A/73/348>.

<sup>2</sup> Privacy International and Article 19, "Privacy and Freedom of Expression In the Age of Artificial Intelligence", 2018, <https://privacyinternational.org/report/1752/privacy-and-freedom-expression-age-artificial-intelligence>.

<sup>3</sup> IFLA, "Technological Trends: Advances in Artificial Intelligence", <https://trends.ifla.org/literature-review/advances-in-artificial-intelligence>.

for five experimental janitorial ads was 0.36). The experiment also saw differences in racial mark-ups of target audiences for several house renting and purchasing advertisements it ran.<sup>4</sup>

#### *Algorithms and the online information environment: risks of bias in inputs and outputs*

A 2019 study “Discrimination, Artificial Intelligence and Algorithmic Decision-Making” prepared for the Anti-discrimination Department of the Council of Europe notes other areas where AI application can carry discrimination risks: these can include online price discrimination that may lead to certain ethnic groups paying more, as well as racially biased results returned by image search engines.<sup>5</sup>

The latter point is extensively illustrated in the 2018 work of Dr. Safyia Umoja Noble “Algorithms of Oppression”, documenting how search engines exhibit racial bias (e.g. in autocomplete suggestion or image search functions). As such, it may be worthwhile to further examine the impact of new technologies including AI and ML on the racial biases and related intolerances within the online information environment.

This also points to the possibilities of machine learning reinforcing existing biases in the online information environment. A 2017 study by Caliskan, Bryson and Narayanan detailed an experiment in which a Natural Language Processing model was trained on a large corpus of texts available online and then demonstrated ‘learned’ word associations that exhibited racial (and gender) biases: e.g. where typical African-American names were seen as less “pleasant”, assessed to have less semantic nearness to prompts meant to evoke pleasant stimuli than typical European names.<sup>6</sup>

A report published by the Brookings Institution points out that, should such learned associations inform any auto-complete or ranking functions of a search engine tool, they could reinforce existing biases.<sup>7</sup> The biases in texts available online can also have further impacts when such texts are used, for instance, to train conversational AI. For example, a small experiment described in *Towards Data Science* in 2019 illustrated a racial audit of a Neural Language Model trained on a set of 255 publicly available news articles that contain the word ‘crime’ in the title. The model then showed a significantly different probability of producing a sentence that contains a negative association for different racial prompts.<sup>8</sup>

#### *Library collections as data*

The question of biased training data is also relevant for libraries as the prominence of Big Data prompted the rise of the concept of “collections as data”. Cultural heritage collections can be used for text and data mining and other methods of computational research, including machine learning.

However, as the 2019 OCLC position paper on Data Science, Machine Learning and Artificial Intelligence in libraries points out, such collections can see marginalised communities underrepresented and/or presented within the dominant cultural narratives. To avoid reinforcing the existing biases when using cultural collections as data, the position paper suggests broadening machine-actionable collections to remedy such underrepresentation and ensure linguistic

<sup>4</sup> Ali Muhammad et al., 2019, “Discrimination through optimization: How Facebook’s ad delivery can lead to skewed outcomes”, <https://arxiv.org/pdf/1904.02095.pdf>.

<sup>5</sup> Frederik Zuiderveen Borgesius, 2018, “Discrimination, artificial intelligence, and algorithmic decision-making”, Council of Europe Directorate General of Democracy, <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>.

<sup>6</sup> Aylin Caliskan, Joanna J. Bryson, Arvind Narayanan, 2017, “Semantics derived automatically from language corpora contain human-like biases”, *Science*, <https://science.sciencemag.org/content/356/6334/183.full>.

<sup>7</sup> Nicol Turner Lee, Paul Resnick, Genie Barton, 2019, “Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms”, Brookings, <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>.

<sup>8</sup> Urwa Muaz, 2019, “Racial Bias in Conversational Artificial Intelligence”, *Towards Data Science*, <https://towardsdatascience.com/racial-bias-in-conversational-artificial-intelligence-be0d3abd9b39>.

diversity.<sup>9</sup> In doing so, it is key to work with the underrepresented communities and follow clear ethical commitments to respect their rights and needs.<sup>10</sup>

### *Algorithmic literacy*

Finally, one of the key considerations for libraries within the discussions on discrimination risks associated with the use of AI, ML and other new information technologies is algorithmic (or AI) literacy. Awareness and understanding of the fundamental operating principles of AI and new technologies among the general public is an important step towards facilitating engaged public discourse and dialogue on how such technologies should be regulated.<sup>11</sup> Such literacy, while not sufficient to ensure algorithmic accountability on its own, could nevertheless help people critically evaluate decisions made by algorithmic systems and know when they can exercise their rights to seek remedies.<sup>12</sup>

The importance of raising AI literacy and awareness is pointed out several AI governance and policy documents, from the Council of Europe’s recommendations in ‘Unboxing Artificial Intelligence: 10 steps to protecting Human Rights’,<sup>13</sup> the G7 Innovation Ministers’ Statement on Artificial Intelligence,<sup>14</sup> to “Ethics Guidelines for Trustworthy AI” by the AI High Level Expert Group set up by the European Commission.<sup>15</sup>

A related document prepared by the High Level Expert Group, “Policy and Investment Recommendations for Trustworthy AI”, emphasises that AI literacy courses need to be accessible to all, accommodating the needs of lower-skilled people and disadvantaged groups.<sup>16</sup> To this end, several libraries and library associations have introduced AI and algorithmic literacy initiatives; as well as AI learning kits or courses in an effort to “democratise” AI learning and make it available for different publics.<sup>17</sup>

<sup>9</sup> Thomas Padilla, 2019, “Responsible Operations: Data Science, Machine Learning, and AI in Libraries”, Dublin, OH: OCLC Research, <https://doi.org/10.25333/xk7z-9g97>.

<sup>10</sup> Ibid.; see also “The Santa Barbara Statement on Collections as Data”, <https://collectionsasdata.github.io/statement/>

<sup>11</sup> IEEE, 2019, “IEEE Position Statement: Artificial Intelligence”, <https://globalpolicy.ieee.org/wp-content/uploads/2019/06/IEEE18029.pdf>

<sup>12</sup> Ansgar Koene et al., 2019, “A governance framework for algorithmic accountability and transparency”, Panel for the Future of Science and Technology - European Parliamentary Research Service, [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS\\_STU\(2019\)624262\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf). Notably, the study points out that broad algorithmic literacy needs to be combined with specific information on any given algorithmic system for a public to be able to assess its decision.

<sup>13</sup> The Council of Europe Commissioner for Human Rights, 21019, “Unboxing Artificial Intelligence: 10 steps to protect Human Rights”, <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64>.

<sup>14</sup> G7 Innovation Ministers, 2018, “G7 Innovation Ministers’ Statement on Artificial Intelligence”, <http://www.g8.utoronto.ca/employment/2018-labour-annex-b-en.html>.

<sup>15</sup> “Ethics Guidelines for Trustworthy AI”, 2019, Independent High-Level Expert Group On Artificial Intelligence Set Up by The European Commission, <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

<sup>16</sup> “Policy and Investment Recommendations for Trustworthy Artificial Intelligence”, 2019, Independent High-Level Expert Group On Artificial Intelligence Set Up by The European Commission, <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>.

<sup>17</sup> For example, a 2019 brief by the Urban Libraries Council ([https://www.urbanlibraries.org/files/AI\\_Leadership-Brief.pdf](https://www.urbanlibraries.org/files/AI_Leadership-Brief.pdf)) lists different learning initiatives focused on AI in libraries throughout North America; ‘AI for All Canada’ is a recently launched project aiming to develop educational AI literacy programs that can be delivered in public libraries in Canada (<https://aiforall.ca/>); and the Frisco Public Library in the US offers ‘AI maker kits’ that patrons can take home to learn to code a program for a ‘smart’ speaker that can recognise voice commands (<https://ejournals.bc.edu/index.php/ital/article/view/10974>).