Foundation The London Story
The Hague Humanity Hub
Fluwelen Burgwal 58, 2511CJ, The Netherlands
advocacy@thelondonstory.org

Response To United Nations Call For Input: Thematic Report On Artificial Intelligence (AI) And Racial Discrimination

Foundation The London Story

Submitted on March 28 2024

Foundation The London Story (TLS) is an Indian diaspora led non-profit organization based in Hague, Netherlands. Founded in 2020, TLS investigates the landscape of disinformation and hate speech to provide evidence-based policy and advocacy solutions. We analyze and document hate speech and hate crimes against Indian minorities, and advocate for justice, peace, and collective action against grave human rights violations.

TLS welcomes the opportunity to submit input to the United Nations Special Rapporteur on contemporary forms of racism. This document provides comments and insights for the thematic report on Artificial Intelligence (AI) and racial discrimination to be presented at the 56th Session of the Human Rights Council.

India is a state party to the UN Convention on the Elimination of All Forms of Racial Discrimination (UNCERD). Therefore, TLS would like to bring reported trends on AI and discrimination in India to the attention of the UN Special Rapporteur. Following a brief overview of the current state of AI in India, the report addresses two specific questions outlined in the Call for Submissions.

Background Note on AI and India's Stance on AI

The potential for artificial intelligence (AI) to cause severe and systematic discrimination has been widely recognized by researchers. This concern amplifies the existing, critical human rights issue of discrimination, as AI has the potential to exacerbate it significantly.

Like all states, the state has been steadfast about embracing and deploying the fast-growing technology like Artificial Intelligence. At both national and international fora, it has referenced the "transformative potential" of AI.[1]

[1] https://pmindiaun.gov.in/pageinfo/MzAxNw,,

Foundation The London Story
The Hague Humanity Hub
Fluwelen Burgwal 58, 2511CJ, The Netherlands
advocacy@thelondonstory.org

India lacks specific codified laws, statutory regulations, or official guidelines governing artificial intelligence (AI) implementation. The responsibility for overseeing AI and related technological advancements falls under the Union Ministry of Electronics and Information Technology. In February 2018, the ministry established four committees tasked with crafting a national AI program roadmap. These committees were examining various aspects of AI, including citizen-centric services, data platforms, skill development, research, and legal, regulatory, and cybersecurity considerations. The reports compiled by these committees have been made available to the public.[2]

The first breakthrough of the government in AI space of the government was in 2018, when NITI Aayog, the Government of India's think tank, issued a discussion paper outlining principles for responsible design, development, and deployment of artificial intelligence (AI) in India (Responsible AI Principles), alongside proposed enforcement mechanisms to ensure adherence to these principles .[3] Subsequently, a follow-up paper was released, which presented a case study analyzing the application of the Responsible AI Principles to the Digi Yatra program.[4] Digi Yatra, developed by the Ministry of Civil Aviation, streamlines boarding processes for departing passengers at airports through the utilization of facial recognition and verification technologies.[5] The discussion paper critically assessed whether the framework and procedures of Digi Yatra align with the standards set forth by the Responsible AI principles and provided recommendations accordingly.

Pending an extensive AI Law, obligations surrounding AI Governance are governed by the Information Technology Act of 2000, along with the associated rules and regulations established thereafter. However, there is no explicit mention of artificial intelligence (AI). The same statute has garnered criticism for being outdated,[6] with demands of a reboot within the statute.[7] The Minister of State for Electronics and Information Technology, Rajeev Chandrasekhar, had issued a statement indicating that AI Chatbots could potentially infringe upon Rule 3(1)(b) of the Intermediary Rules (IT

[2] https://www.meity.gov.in/artificial-intelligence-committees-reports

[3] https://www.niti.gov.in/sites/default/files/2022-11/Ai_for_All_2022_02112022_0.pdf

[4] https://www.civilaviation.gov.in/sites/default/files/2023-07/Digi%20Yatra%20Policy%20%28DIGI%20YATRA%29.pdf

[5] https://www.civilaviation.gov.in/sites/default/files/2023-07/Digi%20Yatra%20Policy%20%28DIGI%20YATRA%29.pdf

[6] https://www.iam-media.com/article/proposed-digital-india-act-overhaul-outdated-legislation

[7] https://internetfreedom.in/update-the-it-act-2000-india-needs-a-reboot/

rules)[8] if they facilitate discrimination and bias on their platforms.[9] Nonetheless, there has not been any legal action involving Chatbots under the purview of the Information Technology Act .

India, through its forthcoming AI Law, known as the Digital India Bill, is positioned to enact significant measures aimed at addressing the ethical and social ramifications of AI deployment.[10] According to reports, the proposed legislation will introduce provisions mandating regular risk assessments and enhancing algorithmic transparency.[11] These anticipated changes arrive at a pivotal moment, as the ethical and human rights considerations surrounding AI adoption are increasingly coming under scrutiny in India.

TLS Response to Specific Questions in the Call for Inputs

1. ***How may AI technology and algorithm bias result in direct or indirect racial discrimination, including in the context of law enforcement, education, social media, healthcare, the criminal justice system, immigration and border management and/or any other relevant contexts in which AI is used?***

Although AI systems are often seen as impartial and precise,[12] they can exhibit algorithmic bias, meaning they do not simply process data neutrally but rather manipulate it in ways that stray from accepted standards, leading to discrimination and

[8] Rule 3(1)(b), IT (Intermediary Guidelines and Digital Media Ethics Code).

[9] https://www.thehindu.com/sci-tech/technology/government-asks-ai-platforms-seek-approval-deploying-under-trial-ai-makes-labelling-mandatory/article67912147.ece

[10] https://vidhilegalpolicy.in/blog/explained-the-digital-india-act-2023/; https://www.meity.gov.in/writereaddata/files/DIA_Presentation%2009.03.2023%20Final.pdf

[11] https://www.india-briefing.com/news/digital-india-bill-2023-key-provisions-stakeholder-perspectives-28755.html/

[12] Kahneman, D., Rosenfield, A.M., Gandhi, L., and Blaser, T. (2016). Noise: How to overcome the high, hidden cost of inconsistent decision making. *Harvard Business Review*. https://www.hbr.org/2016/10/noise. Accessed 11/04/2019. [Ref list]

against certain individuals or groups based on their social identity.[13] Biased input data is a primary contributor to discrimination by AI systems.[14]

TLS presents two overarching issues that cause discrimination by AI in India:

Biased Training Data: There are shortcomings present in historical data upon which machine learning applications are developed.[15] The bias inherent in such data can be perpetuated in real-world applications, potentially resulting in models generating outputs that reinforce existing cultural prejudices.[16] For instance, a 2021 study highlighted the inherent distortion in Indian data, which disproportionately favors digitally-rich profiles, particularly those of middle-class men.[17] This bias leads to incomplete representations or complete absence of the communities from the dataset, which can institutionalize discrimination by reinforcing existing stereotypes.[18] Another illustrative example of this bias is evident when a chatbot is asked to provide names of 20 Indian doctors and professors, resulting in suggestions dominated by Hindu dominant-caste surnames.[19] This showcases how unequal representations in data reflect caste-based inequities in generative AI systems.

Biased Programmers: Bias in AI systems in India has also stemmed from the perspectives and biases of the programmers who designed and trained these systems. Their biases, whether conscious or unconscious, can influence the algorithms and

[13] Danks, D. & London, A.J. (2017). Algorithmic bias in autonomous systems. In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 4691–97. Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization. 10.24963/ijcai.2017/654 .

[14] S Barocas and A Selbst, 'Big Data's Disparate Impact' [2016] 104 Calif Law Rev 671 <https://www.jstor.org/stable/24758720>.

[15] Cowgill B, Dell'Acqua F, Deng S, Hsu D, Verma N, and Chaintreau A, "Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics," *arXiv:2012.02394 [cs, econ, q-fin]*, Dec. 2020, arXiv: 2012.02394. [Online]. Available: http://arxiv.org/abs/2012.02394

[16] Shikha and Samuel R. Bowman. "Identifying and Reducing Gender Bias in Word-Level Language Models." *North American Chapter of the Association for Computational Linguistics* (2019).

[17] Nithya Sambasivan, Erin Arnesen, Ben Hutchinson, Tulsee Doshi, and Vinodkumar Prabhakaran. 2021. Re-imagining Algorithmic Fairness in India and Beyond. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 315–328. https://doi.org/10.1145/3442188.3445896

[18] https://www.fastcompany.com/90269688/high-tech-redlining-ai-is-quietly-upgrading-institutional-racism

[19] https://scroll.in/article/1055846/indias-scaling-up-of-ai-could-reproduce-casteist-bias-discrimination-against-women-and-minorities

decision-making processes within the AI system, leading to biased outcomes.[20] For instance, there is a noticeable gap between AI developers and the communities they aim to serve, particularly in India. Within the Indian technology sector, which often prides itself on being "merit-based," AI engineers tend to come from a relatively privileged background, often benefiting from caste privileges.[21] This leads to the inadvertent inclusion of their biases, which can manifest in discriminatory AI designs.

At a micro-level, beyond the legal frameworks governing this field, raising awareness among developers about the discriminatory effects of AI has been shown to enhance the accuracy of their algorithms.[22] Engineers in India have reportedly grasped technical guidance more effectively when provided with better training data.[23]However, there is insufficient data to assess whether such practices are being adopted within technology companies in India and abroad, that have an Indian user market, to uphold the principles of non-discrimination.

Moreover, it is evident that conventional technologies forming the foundation of AI have historically been centered around Western perspectives.[24] The data utilized often lacks appropriate collection and quantification methods tailored for Indian usage, resulting in double standards and superficial attempts at fairness.[25] Despite India's extensive adoption of AI in facial recognition,[26] predictive policing,[27] and healthcare,[28] academic literature surrounding AI remains heavily focused on Western histories and concerns. Existing research predominantly emphasizes racial and gender bias and discrimination, which are more prevalent in the US American context. Meanwhile,

[20] https://scroll.in/article/1055846/indias-scaling-up-of-ai-could-reproduce-casteist-bias-discrimination-against-women-and-minorities

[21] https://arxiv.org/pdf/2101.09995.pdf

[22] *Cowgill B, Dell'Acqua F, Deng S, Hsu D, Verma N, and Chaintreau A, "Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics," arXiv:2012.02394 [cs, econ, q-fin], Dec. 2020, arXiv: 2012.02394. [Online]. Available: http://arxiv.org/abs/2012.02394*

[23] *Cowgill B, Dell'Acqua F, Deng S, Hsu D, Verma N, and Chaintreau A, "Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics," arXiv:2012.02394 [cs, econ, q-fin], Dec. 2020, arXiv: 2012.02394. [Online]. Available: http://arxiv.org/abs/2012.02394*

[24] *Sambasivan N and others, "Re-Imagining Algorithmic Fairness in India and Beyond" [2021] Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.*

[25] *Sambasivan N and others, "Re-Imagining Algorithmic Fairness in India and Beyond" [2021] Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.*

[26] https://sflc.in/deployment-of-facial-recognition-technology-for-state-surveillance-and-monitoring/

[27] https://www.thehindu.com/news/national/karnataka/karnataka-adopts-ai-enabled-policing/article67744869.ece

[28] https://www.forbesindia.com/article/isbinsight/demystifying-ai-in-healthcare-in-india/87547/1

other axes of injustice specific to Asian cultures, such as those affecting Adivasis, Dalits and religious minorities, are conveniently overlooked.

Challenges of Implementing AI in Criminal Justice in India

While the application of Artificial Intelligence poses numerous risks concerning civil and political rights, the focus of inquiry within this submission is limited to social rights governed by the ICERD under the scope of Anti-Discrimination.

In terms of criminal identification, India has kept pace with American and European governments in deploying Facial Recognition Technology (FRT). This technology involves matching captured images with facial images stored in databases or "watchlists," facilitating the identification and verification of individuals. It reportedly serves various purposes, including law enforcement, security, and authentication processes.

In India, these systems operate by generating a probability or confidence score through a comparison of a suspect's image with a police database of known criminals, such as through the Crime and Criminal Tracking Network System.[29] For verification purposes, multiple potential matches are generated along with confidence scores indicating the likelihood of a correct match. However, the ultimate determination typically falls to a human analyst. This introduces a potential for misidentification, as the FRT system may propose several matches, leaving the final decision to the analyst.[30] There are concerns that this process can be susceptible to bias, as the analyst's personal prejudices may influence their ultimate decision. Factors such as caste, religion, or gender discrimination could impact the analyst's decision-making, thus introducing bias into the system.[31]

Currently, the deployment of FRT lacks legislative authorization. Notwithstanding, at least 20 States/Union Territories either intend to deploy or have already deployed FRTs. Moreover, the National Crime Records Bureau (NCRB) had solicited proposals for a national-level Automated Facial Recognition System (AFRS). In July 2020, the NCRB specified that the AFRS deployment would not encompass CCTV camera data.[32] However, this exclusion of CCTV data from AFRS deployment does not rule out the potential utilization of CCTVs for FRT. With independent deployment of FRTs lacking

[29] https://internetfreedom.in/watch-the-watchmen-part-3/

[30] https://hrlr.law.columbia.edu/hrlr-online/you-can-see-my-face-why-cant-i-facial-recognition-and-brady/?ref=static.internetfreedom.in

[31] https://www.medianama.com/2020/07/223-afrs-revised-tender-ncrb/

[32] https://www.medianama.com/2020/07/223-afrs-revised-tender-ncrb/

formal authorization, concerns arise about the potential use of CCTVs for FRT purposes, as it remains unrestricted by law and could be introduced at the national level at any point. In fact, the absence of comprehensive legislation for FRT means that state police forces can currently use CCTV data, a practice they often engage in.

For example, following communal riots in North-East Delhi in February 2020, there were widespread arbitrary detentions of individuals belonging to the Islamic faith. The Delhi Police reported that out of the 1800 arrests related to this unrest, 137 were facilitated by Facial Recognition Technology (FRT).[33] They further stated that "the apprehended individuals were primarily identified based on CCTV footage and publicly available videos," with FRT aiding in the identification process.[34]

The government has issued clarifications addressing privacy concerns FRTs may raise.[35] However, concerns regarding the potential for racial, religious, and caste-based discrimination resulting from FRTs remain unaddressed in India at a federal level.

Academics and non-governmental organizations (NGOs) have advocated for a moratorium on the use of Artificial Intelligence (AI) in policing procedures.[36] The legality of employing such systems appears uncertain, as courts are yet to establish a consensus on the use of AI based policing systems.[37]

Creating an Effective Model: Proactive Measures to Combat Discrimination in AI Implementation

Article 2 of the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD) sets a high bar for member states. It compels them not only to avoid racial discrimination (a negative obligation)[38] but also to actively dismantle it

[33] https://caravanmagazine.in/conflict/detentions-delhi-violence-northeast-muslim-arrests-riots-police-crime-branch

[34] https://caravanmagazine.in/conflict/detentions-delhi-violence-northeast-muslim-arrests-riots-police-crime-branch

[35] https://pib.gov.in/PressReleasePage.aspx?PRID=1885963

[36] Mustafa, Faizan and Leo, Utkarsh, On Facial Recognition and Fundamental Rights in India: A Law and Technology Perspective (December 29, 2021). Available at SSRN: https://ssrn.com/abstract=3995958 or http://dx.doi.org/10.2139/ssrn.3995958

[37] https://theleaflet.in/conundrum-of-expectations-are-courts-prepared-for-challenges-against-facial-recognition-technology/; https://vidhilegalpolicy.in/research/indian-law-enforcements-ongoing-usage-of-automated-facial-recognition-technology-ethical-risks-and-legal-challenges/

[38] ICERD, art 2(1)(a).

through positive measures.[39] India, as a signatory to the convention, is bound by this proactive approach.

There are best practice examples on positive measures taken in other jurisdictions with similar legal systems. For instance, positive obligations were effectively highlighted in the case of *R (Bridges) v South Wales Police*, adjudicated by the England and Wales Court of Appeals.[40] The focal point of this legal case was the legality surrounding the South Wales Police (SWP) Force's implementation of live automated facial recognition (AFR) technology within a trial utilizing a system named 'AFR Locate'.[41] AFR Locate entailed deploying surveillance cameras to capture digital images of the general public, subsequently processing and comparing these images with digital profiles of individuals listed in a watch list compiled by the SWP. The primary aim of the system was to identify individuals falling into different categories, such as those with active warrants and those regarded as of interest to the police for intelligence reasons.

A Human Rights Group contested the South Wales Police (SWP)'s deployment of this AI application before the Divisional Court of the Queen's Bench Division,[42] on account of it raising concerns about potential indirect discrimination based on sex and/or race, exhibiting a higher rate of positive matches for female and/or Black and ethnic minority individuals.

At first, the Divisional Court rejected the lawsuit against the SWP. Nevertheless, on appeal, the Court of Appeal reversed this judgment, highlighting the importance of the positive obligation placed on public and governmental authorities to actively address discrimination, especially regarding potential biases in technological systems.

The Court of Appeal emphasized that attention should not only be on whether the AI application has directly discriminated, but also on whether the public authority has taken proactive steps to evaluate and reduce the risk of bias and discrimination. The utilization of AFR Locate was deemed to have breached Equality Law due to failure of the state in conducting a thorough risk assessment and implement measures to prevent discrimination.

In line with ICERD's positive duty to eradicate discrimination, state parties such as India must take steps to prevent discrimination altogether. It is reasonable for the state parties to conduct thorough risk assessments before embracing potentially discriminatory AI technologies.

---

[39] https://www.ohchr.org/en/instruments-and-mechanisms/international-human-rights-law

[40] *R (Bridges) v South Wales Police* [2020] EWCA Civ 1058

[41] http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S1727-37812021000100029

[42] *R (Bridges) v Chief Constable of South Wales Police* [2019] EWHC 2341

Foundation The London Story
The Hague Humanity Hub
Fluwelen Burgwal 58, 2511CJ, The Netherlands
advocacy@thelondonstory.org

2. ***Measures to ensure that private actors and corporations respect the human rights to equality and non-discrimination in the development and use of AI, including by ensuring that such actors engage in human rights due diligence and impact assessments on AI technologies;***

The responsibility for conducting comprehensive risk assessments should extend beyond the state. Under the UN Guiding Principles on Business and Human Rights, clause 1, "states must protect against human rights abuse (…) by third parties, including business enterprises". Such human rights violations include those covered by ICERD, which mandates states to prevent discrimination by itself or by private entities.

According to the clauses in Chapter II of the UN Guiding Principles on Business and Human Rights, companies engaged in the development or deployment of AI, which has the potential for harmful applications, are obligated to exercise a level of due diligence to prevent such harms.[43] Due diligence, in this context, is a customized process that requires companies to address the specific risks associated with their own operations. Human Rights Impact Assessments (HRIAs) serve as an extension of these due diligence obligations, allowing companies to proactively manage both potential and actual adverse human rights impacts resulting from their involvement.

It is important to recognize that the regulatory framework for conducting Human Rights Impact Assessments on AI and algorithms differs significantly from that governing other types of business operations that do not utilize AI.[44]

For example, a Human Rights Impact Assessment (HRIA) was conducted by a consulting firm for Facebook (now Meta) to investigate the purported human rights repercussions attributed to its operations in India.[45] The outcomes of this assessment were never published, and only a "summary disclosure" of four pages was released.[46] It is therefore not clear whether the HRIA evaluated the effects of Facebook's recommender algorithm AI and content moderation AI and LLMs on the violence targeting for instance religious minorities in India. This is not the first HRIA by Facebook that is deemed unsatisfactory, with for instance serious criticism about Facebook's HRIA about Myanmar.[47]

---

[43] https://www.ohchr.org/en/special-procedures/wg-business/corporate-human-rights-due-diligence-identifying-and-leveraging-emerging-practices

[44] https://cltc.berkeley.edu/wp-content/uploads/2021/08/AI_Risk_Impact_Assessments.pdf

[45] https://thelondonstory.org/2022/02/22/human-rights-impact-assessment-of-facebook-in-india/

[46] https://humanrights.fb.com/wp-content/uploads/2023/09/2022-Meta-Human-Rights-Report.pdf

[47] Mark Latonero and Aaina Agarwal. 3/19/2021. "Human Rights Impact Assessments for AI: Learning from Facebook's Failure in Myanmar." Carr Center Discussion Paper Series.

Similar apprehensions were expressed regarding Facebook's algorithms in India. Critiques suggested that these algorithms exhibit a strong bias toward promoting content that provokes extreme emotional reactions, rather than emphasizing rational or objective discourse.[48] Further, during the 2019 national elections in India and the subsequent riots in Delhi, Time Magazine reported that Facebook's hate-speech algorithms only covered four out of the 22 state languages spoken by India's vast population of 1.4 billion people.[49] This is also the number confirmed by Meta to date in its own blog posts.[50] This glaring gap in coverage underscored the inadequacies of Facebook's algorithm to combat hate speech and incendiary content, particularly in a diverse and complex socio-political landscape like India.

Drawing lessons from this example, we advocate for the implementation of HRIAs tailored specifically for algorithms. We assert that processes uniquely designed for the realm of AI are essential for ensuring the effectiveness of HRIAs. HRIAs for AI must delve into the intricate workings of algorithms, necessitating a thorough examination of their specific technical components. Moreover, HRIAs for algorithms should span the entire life cycle of an AI system, commencing from its conceptualization phase through key developmental stages, and persisting with punctual evaluations post-implementation. They should not be confined to either pre-implementation or post-implementation phases exclusively. Emerging HRIAs that adhere to these criteria include the Fundamental Rights and Algorithm Impact Assessment, pioneered by the government of the Netherlands.[51] This framework offers numerous examples of potential mitigating measures aimed at averting negative impacts. Such comprehensive approaches serve to diminish the risks associated with unjustified infringements on discrimination and human rights.

We stress the imperative need to establish a standardized process for AI Human Rights Impact Assessments at an international level, particularly under the auspices of the United Nations (UN), which embodies the most exemplary practices in this domain. The Organisation for Economic Co-operation and Development (OECD) issues crucial recommendations for AI companies, drawing from its guidance on human rights due diligence.[52] However, it is pertinent to highlight that the OECD itself acknowledges the

[48] https://www.zdnet.com/article/can-facebook-prevent-its-algorithm-from-facilitating-a-communal-bloodbath-in-india/

[49] https://www.zdnet.com/article/can-facebook-prevent-its-algorithm-from-facilitating-a-communal-bloodbath-in-india/

[50] https://about.fb.com/news/2024/03/how-meta-is-preparing-for-indian-general-elections-2024/

[51] https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms

[52] OECD, 'Human Rights Due Diligence Through Responsible AI' in *AI in Business and Finance: OECD Business and Finance Outlook 2021* (OECD 2021).

superior effectiveness of mandatory domestic laws regarding due diligence.[53] Thus, we advocate for the adoption of such domestic laws, inspired by a standardized process developed by the UN that reflects the best practices observed globally. This ensures a more robust framework for addressing human rights impacts associated with AI technologies.

While these existing guidelines for AI companies serve as soft law or non-binding instruments, there is a growing consensus that they should be elevated to the status of mandatory legislation at the national level, akin to the framework established in the Netherlands. Such legislation would compel businesses to conduct thorough due diligence checks during the development or deployment of AI technologies. The implementation of such laws is deemed essential for states to fulfill their own due diligence obligations in regulating the behavior of private entities, as mandated by international human rights law (IHRL).[54]

[53] OECD, 'Human Rights Due Diligence Through Responsible AI' in *AI in Business and Finance: OECD Business and Finance Outlook 2021* (OECD 2021).

[54] *Zimbabwe Human Rights NGO Forum v Zimbabwe* Communication No 245/2002 (25 May 2006), para 147