



C.A.S.C.A

**CENTRE FOR ADVANCED STUDIES
IN CYBER LAW AND ARTIFICIAL
INTELLIGENCE**

at

**RAJIV GANDHI NATIONAL UNIVERSITY OF
LAW, PUNJAB**

**RESPONSE TO CALL FOR INPUTS:
THE USE OF ARTIFICIAL
INTELLIGENCE**

&

**THE UN GUIDING PRINCIPLES ON
BUSINESS AND HUMAN RIGHTS**

Call For Inputs By The United Nations

The Use of Artificial Intelligence & The UN
Guiding Principles on Business and
Human Rights

Authors: Tanmay Durani, R. Dayasakthi, Kunaal Hemnani, Vishwaroop Chatterjee

Research Consultants:

Prof. (Dr.) Jai Shankar Singh, Prof. (Dr.) Yogesh Pratap Singh, Prof. (Dr.) Naresh Vats,
Dr. Ivneet Walia, Ms. Medha Garg



CENTRE FOR ADVANCED STUDIES IN CYBER LAW AND ARTIFICIAL INTELLIGENCE [CASCA] is a research-driven centre at RGNUL dedicated to advancing scholarly research and discourse in the field of Technology Law and Regulation. As a research centre of a leading institution in India, we are committed to promoting interdisciplinary research, fostering collaboration, and driving innovation in the fields of cyber law, artificial intelligence, and other allied areas.

For more information

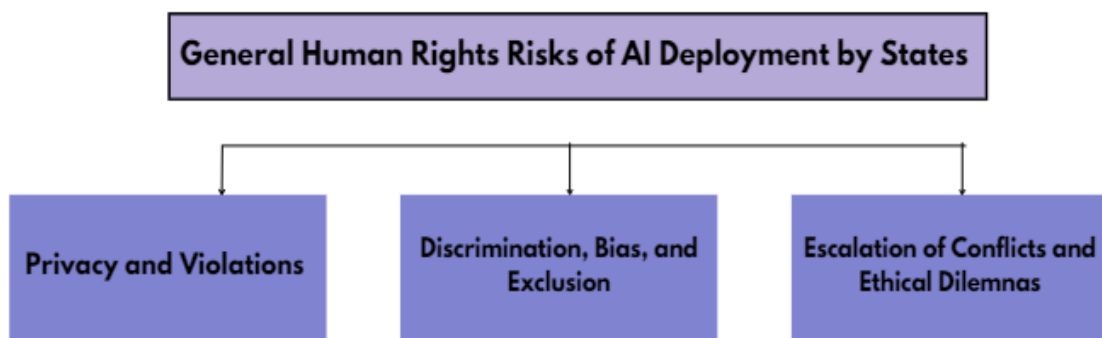
Visit cascargnul.com

Disclaimer

The facts and information in this report may be reproduced only after giving due attribution to CASCA.

Executive Summary

The submission explores how the use of Artificial Intelligence (AI) by states and businesses affects human rights in areas like public health, policing, and welfare. While technologies such as facial recognition and predictive policing enhance service efficiency, they also pose risks like privacy breaches, bias, and due process violations due to biased data. To address these concerns, it is vital to implement Human Rights Impact Assessments (HRIAs) and Algorithmic Impact Assessments (AIAs), enforce human rights due diligence, and ensure strict regulation and accountability. This will help align AI applications with ethical principles and protect against their negative impacts on human rights.



1. State AI Applications & Human Rights Impact

While these risks are pervasive, they are particularly pronounced in few areas of AI deployment by States. Three critical use cases are discussed below illustrating the manifestation of human rights risks in specific contexts:

a. Public Health

During the pandemic, India saw the rollout of novel thermal imaging technologies combined with FRT. In May 2020, Kerala introduced its first thermal and optical imaging camera equipped with AI-enhanced facial recognition capabilities, designed to detect fevers from a distance while maintaining social distancing.¹ The integration of FRT with thermal imaging for monitoring mask usage and body temperatures became a widespread practice at the entry points of public venues like malls and hospitals.

¹ Kerala Gets First Thermal Screening Camera with Face Detection Tech to Ensure Social Distancing (*The New Indian Express*, 2 May 2020) <<https://www.newindianexpress.com/states/kerala/2020/May/02/kerala-gets-first-thermal-screening-camera-with-face-detection-tech-to-ensure-social-distancing-2138272.html>> accessed 13 January 2025.

b. Policing and due-process of law

AI revolutionizes policing and criminal investigation through advanced tools such as predictive analytics, facial recognition, etc.² Predictive policing systems like CompStat analyze historical crime data to forecast crime hotspots, enabling law enforcement agencies to allocate resources effectively and preempt criminal activities. Facial recognition technologies assist in identifying suspects by matching surveillance footage with extensive databases. These systems are instrumental in crime prevention, targeting "risky individuals" or locations through tools like heat lists, and supporting investigations by uncovering patterns in communication and behavior.³

Automation in policing poses significant human rights risks. Clearview AI, used by NYPD, relies on a 50-billion-image database scraped from social media without consent. This breaches privacy, erodes trust in online platforms, and heightens the risk of data misuse.⁴

Predictive policing systems perpetuate systemic biases. For instance, Chicago's "heat list" program identified individuals based on predictive models, leading to increased police scrutiny predominantly in minority communities.⁵ Since these systems are built on historical crime data, they reflect decades of unequal enforcement practices. Thus, neighborhoods that have historically faced over-policing continue to be flagged as high-risk, while others remain overlooked. The focus on certain types of offenses, like street-level crimes, and the underrepresentation of crimes in affluent areas amplify these disparities, reinforcing cycles of inequality and mistrust in the system.

In India, from 2019 to 2023, various state and city-level law enforcement projects involving FRT have been implemented in multiple locations across India, including Hyderabad, Chennai, Chandigarh, Uttar Pradesh, Uttarakhand, Bihar, Rourkela, Delhi, Jammu and Kashmir, Dharamsala, Odisha, and Haryana.

The "black box" nature of AI decision-making makes these systems opaque, even to their creators, raising concerns about accountability and fairness.⁶ In *Loomis v. Wisconsin*⁷, the COMPAS

² Marcus Smith and Seumas Miller, 'The Ethical Application of Biometric Facial Recognition Technology' (2021) 37 AI & SOCIETY 167 <<https://link.springer.com/article/10.1007/S00146-021-01199-9>> accessed 29 November 2024.

³ Thaddeus L Johnson and others, 'Facial Recognition Systems in Policing and Racial Disparities in Arrests' (2022) 39 Government Information Quarterly 753 <<https://www.sciencedirect.com/science/article/pii/S0740624X22000892>> accessed 29 November 2024.

⁴ Robert Hart, 'Clearview AI—Controversial Facial Recognition Firm—Fined \$33 Million for "Illegal Database"' (*Forbes* 4 September 2024) <<https://www.forbes.com/sites/roberthart/2024/09/03/clearview-ai-controversial-facial-recognition-firm-fined-33-million-for-illegal-database/>> accessed 30 November 2024.

⁵ Matt Stroud, 'An Automated Policing Program Got This Man Shot Twice' (*The Verge* 24 May 2021) <<https://www.theverge.com/c/22444020/chicago-pd-predictive-policing-heat-list>> accessed 30 November 2024.

⁶ Holland Michel A, "The Black Box, Unlocked: Predictability and Understandability in Military AI" (United Nations Institute for Disarmament Research 2020) <<https://doi.org/10.37559/sectec/20/ai1>> accessed January 3, 2025.

⁷ *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

algorithm, developed by Equivant, was used to predict the defendant's likelihood of reoffending, influencing sentencing. Classified as high-risk, Loomis could not challenge the assessment due to the proprietary secrecy of COMPAS's methodology. This undermined right to due process, as the data and logic behind the decision remained inaccessible. Such opacity risks embedding hidden biases and eroding fairness in judicial proceedings.

c. Welfare Programs

During the pandemic, the use of Facial Recognition Technology (FRT) expanded to include authentication for welfare programs, promoting social distancing. One of the earliest examples occurred in 2021 when the government of Meghalaya introduced an app⁸ to replace the traditional method of pensioner verification, which required periodic visits to the Treasury Officer or Pension Disbursing Authority.

That same year, the Union Government's Department of Pension & Pensioners' Welfare (DoPPW) introduced an FRT-based application that verifies pensioners and confirms if they are dead/alive.⁹ This app uses the UIDAI AadhaarFaceRD mobile application along with a facial recognition database to issue life certificates.¹⁰ Advancing further, the Telangana Government's Department of Consumer Affairs, Food, and Civil Supplies issued a tender for the installation and maintenance of 17,500 electronic point of sale (ePoS) devices designed to support photo matching and facial recognition features.

d. Use of Automated Weapons System (AWS)

AWSs leverage technologies like machine learning, computer vision etc to execute target identification, decision-making, and lethal action etc. Operating on a "sense-decide-act" framework, AWS processes sensor data to predict and recommend actions at speeds surpassing human cognition. The U.S. Department of Defense's Project Maven¹¹ and Advanced Targeting

⁸ Jain A, 'Meghalaya Government, stop the use of FRT for verification of pensioner's identities' (*Internet Freedom Foundation*, 16 August 2021) <<https://internetfreedom.in/we-wrote-to-the-govt-of-meghalaya/>> accessed 13 January 2025.

⁹ Union Minister Dr Jitendra launches unique Face Recognition Technology for Pensioners, says it will bring Ease of Living for the retired and elderly citizens (PIB Delhi 2021) <<https://pib.gov.in/PressReleasePage.aspx?PRID=1776146>> accessed 13 January 2025.

¹⁰ Jain A, 'Facial recognition app to verify pensioners rolled out by Indian govt amid risks' (*MEDIANAMA*, 3 December 2021) <www.medianama.com/2021/12/223-facial-recognition-app-pensioners-uidai/> accessed 13 January 2025.

¹¹ Saleha Mohsin, 'Inside Project Maven' (*Bloomberg*, 29 February 2024) <<https://www.bloomberg.com/news/newsletters/2024-02-29/inside-project-maven-the-us-military-s-ai-project>> accessed 3 January 2025.

and Lethality Automated System¹² aim to enhance precision and reduce human cognitive load, limiting the operator's role to confirming pre-selected targets within seconds.

Delegating critical decisions to AI raises concerns due to the opaque nature of AWS, referred to as "black box" systems¹³. AlphaGo¹⁴, an AI for the board game Go, made decisions its programmers couldn't comprehend. While this unpredictability is benign in games, it becomes dangerous in warfare, where incomprehensible AI logic has catastrophic outcomes. Despite progress, such disconnects remain problematic in complex, dynamic environments.

Biases in AWS due to skewed datasets exacerbate risks in combat. Studies highlight inaccuracies in identifying darker-skinned individuals, raising concerns. For instance, a neural network misclassified wolves and dogs based on irrelevant cues like snowy or grassy backgrounds. Similarly, military AI failed to detect camouflaged tanks because it associated tanks with cloudy skies. These reveal AI's reliance on flawed correlations, which falter in diverse environments. In warfare, AI's probabilistic reasoning lacking human contextual judgment could prove catastrophic.

e. Education

Another development in India across various industries was the application of FRT for attendance verification. Initially implemented in educational settings¹⁵ to support social distancing measures¹⁶ during the pandemic and to monitor temperatures, this technology quickly spread to other areas, including national transportation systems like highways and railways. Within the education sector, FRT has also been used to verify identities during entrance exams¹⁷ and to deter cheating.¹⁸ In a

¹² 'Advanced Targeting and Lethality Automated System Archives' (*Breaking Defense*) <<https://breakingdefense.com/tag/advanced-targeting-and-lethality-automated-system/>> accessed 3 January 2025.

¹³ Holland Michel A, "The Black Box, Unlocked: Predictability and Understandability in Military AI" (United Nations Institute for Disarmament Research 2020) <<https://doi.org/10.37559/sectec/20/ai1>> accessed January 3, 2025.

¹⁴ 'AlphaGo Algorithm in Artificial Intelligence' (*GeeksforGeeks* 25 June 2024) <<https://www.geeksforgeeks.org/alphago-algorithm-in-artificial-intelligence/>> accessed 3 January 2025.

¹⁵ Bhatnagar G, "'Pandora's Box of Privacy Issues': Experts on Delhi Govt Schools' Use of Facial Recognition Tech" (*The Wire*, February 24, 2021) <<https://thewire.in/education/delhi-government-schools-facial-recognition-cctv-cameras>> accessed January 3, 2025.

¹⁶ NT S, "UP: Teachers Oppose Facial Recognition Based Attendance Rule" (*MEDIANAMA*, December 5, 2023) <<https://www.medianama.com/2023/12/223-up-school-teachers-facial-recognition-attendance/>> accessed January 3, 2025.

¹⁷ Sur A, "National Testing Agency Expands Surveillance Cover for Entrance Exams like JEE, NEET" (*MEDIANAMA*, August 11, 2021) <<https://www.medianama.com/2021/08/223-facial-recognition-surveillance-exams-2/>> accessed January 3, 2025.

¹⁸ Sharma N, "In a 1st, UP Uses AI against Paper Leaks & Cheating in Recruitment Exam for Police Constables" (*theprint*, September 25, 2024) <<https://theprint.in/india/in-a-1st-up-uses-ai-against-paper-leaks-cheating-in-recruitment-exam-for-police-constables/2283399/>> accessed January 3, 2025.

unique case in Maharashtra, the technology was even utilized to maintain the quality of food distributed to students.¹⁹

2. Human Rights Risks of AI in Non-Tech Businesses

Businesses across sectors have been deploying AI to replace or supplement existing manpower. While the technology sector has substantial control over how the algorithms are programmed, non-technological businesses procuring AI from outside have encountered human rights violations due to such automations.

a. Banking

AI-based credit scoring uses diverse datasets ranging from social media profiles to bill payments to profile consumers as credit-worthy or not²⁰. The results determine consumers' loan eligibility. In practice, AI tends to discriminate against minorities by flagging them down with lower credit scores than their white counterparts. Biased AI coupled with insufficient training data on minorities results in inaccurate credit scores.

The problem with the noisy data set extends to the infringement of privacy stemming from capricious data collection and interpretation by AI²¹. The vast dataset includes non-financial personal information like location, and eating habits which are subject to AI processing. AI draws inferences on the consumers' sexuality, affiliations, etc resulting in higher discrimination absent amidst traditional lenders²². This simultaneously infringes upon the consumer's privacy and leads to network discrimination where individuals are impacted by their personal social circles.

b. Workforce

¹⁹ NT S, "Indian School Installs AI-Based Machine to Assess Food Quality for Students" (*MEDIANAMA*, April 24, 2023) <<https://www.medianama.com/2023/04/223-ai-based-machine-to-assess-food-quality-concerning/>> accessed January 3, 2025.

²⁰ Ashraf Faheem M, Speridian Technologies, and Lahore Leads University, "AI-Driven Risk Assessment Models: Credit Scoring and Default Prediction" (2021) 5 *IRE Journals* 177 <<https://www.irejournals.com/formatedpaper/1702907.pdf>> accessed November 29, 2024.

²¹ Richardson R, "Racial Segregation and the Data-Driven Society: How Our Failure to Reckon with Root Causes Perpetuates Separate and Unequal Realities" [2021] *SSRN Electronic Journal* <<https://ssrn.com/abstract=3850317>> accessed November 30, 2024.

²² The World Bank Group and others, *CREDIT SCORING APPROACHES GUIDELINES* (2019) <<https://thedocs.worldbank.org/en/doc/935891585869698451-0130022020/original/CREDITSCORINGAPPROACHESGUIDELINESFINALWEB.pdf>> accessed January 2, 2025.

Here, AI has been extensively deployed for workforce management which has led to significant human rights concerns regarding discrimination and the erosion of worker autonomy. These issues infringe upon the right to just and favorable working conditions given in Article 7 of ICESCR.²³

AI's reliance on skewed and incomplete datasets perpetuates institutional biases against workers on religious, communal grounds. Algorithms trained on historical employment data often replicate existing biases, like favoring men over women while hiring. Additionally, worker autonomy is compromised when AI penalizes employees in rigid and context-insensitive ways. Example, algorithms fail to account for unique circumstances, like a worker arriving late due to an accident, resulting in undue penalties under automated systems.

AI's inability to distinguish between serious concerns and harmless sarcasm infringes upon workers' freedom of expression by compelling candidates to sanitize their online presence²⁴. Good Egg's Algorithm, designed to screen candidates' social media activity, penalized individuals for comical interactions. Such surveillance forces workers to maintain a "polished" online profile, compromising their freedom of expression.

c. Healthcare

Currently, AI in healthcare operates on a restricted dataset giving rise to skewed results and misdiagnosis. The lack of data stems from structural issues of digital inaccessibility. Smartphone data comes primarily from men earning above-average income and if reliance is placed upon this, it distorts our understanding of the health needs of disadvantaged communities. Instances of this can be seen with AI following men's heart attack symptoms in the diagnosis of a female patient, repetition of such occurrences eventually leads to under-diagnosis²⁵.

Patient autonomy constitutes a fundamental part of medical ethics. AI Deployment has often been threatened when patients are deprived of their say in choosing whether AI could be used in caregiving pathways. Simultaneously, deployment of AI implies continuous behavioral monitoring of people's online presence which restricts and interferes with their personal lives.

3. Possible Approaches for Regulation

Effective AI governance must integrate proactive measures (ex-ante assessments) and ongoing evaluations (ex-post assessments) to mitigate risks throughout an AI system's life cycle. These

²³International Covenant on Economic, Social and Cultural Rights (adopted 16 Dec 1966, entered into force 3 January 1976) UNTS 999, Art 7.

²⁴Filippo Raso and others, "Artificial Intelligence & Human Rights: Opportunities & Risks" (September 25, 2018) <<http://nrs.harvard.edu/urn-3:HUL.InstRepos:38021439>>accessed November 28, 2024.

²⁵United Nations Development Programme, *Report: Artificial Intelligence & Potential Impacts on Human Rights in India* (July 2022)[Report_Artificial Intelligence & Potential Impacts on Human Rights in India \(2\) \(1\) 0.pdf](#) accessed December 3, 2024.

evaluations should be informed by inputs from individuals vulnerable to discrimination, enhancing fairness, transparency, and accountability.

Human Rights Impact Assessments (HRIAs) and Algorithmic Impact Assessments (AIAs) should be mandatory for high-risk AI systems in areas such as predictive policing, and social welfare ensuring AI systems align with ethical and legal principles.

Businesses must comply with enforceable human rights due-diligence obligations, as outlined in Principle 15 of the UNGP on Business and Human Rights²⁶, requiring them to identify, prevent, and address human rights impacts. States should enforce this through legislation. H&M's Responsible AI Framework exemplifies good practice, collaborating with workers and experts to address biases and protect vulnerable groups.²⁷ H&M Group has taken significant steps to ensure ethical and sustainable use of AI through its Responsible AI Framework, initiated in 2018. The framework comprises nine key principles, developed to guide the design, deployment, and utilization of AI within the organization in an ethical manner. To implement these principles effectively, H&M established a Responsible AI Team, which collaborates closely with internal and external stakeholders to address potential biases and unintended consequences of AI applications.

The Responsible AI Team at H&M uses a detailed Checklist as part of a rigorous review process for all AI projects. This checklist, consisting of 30 questions, helps assess AI applications against the Responsible AI Principles to identify and mitigate risks, especially those that could affect vulnerable groups, including children. For ongoing oversight, each AI product is regularly reviewed post-deployment to ensure it continues to meet ethical standards. If issues are identified, mitigation measures are implemented to align with responsible design practices.

Right to Privacy

In India, the Supreme Court recognized the right to privacy as a fundamental right under the Constitution through the *Puttaswamy* judgment. This right can be limited by laws that meet specific criteria: they must exist formally as law, serve a legitimate state aim, and be proportional to their purpose. FRT by the state or private entities lacks legislative backing in India. Although the National Crime Records Bureau (NCRB) claims that FRT has cabinet approval,²⁸ this does not equate to legal authority, since cabinet approval is not a legislative measure.

²⁶ UN Guiding Principles on Business and Human Rights, [2011] UN Office of the High Commissioner for Human Rights <<https://www.undp.org/sites/g/files/zskgke326/files/migration/in/UNGP-Brochure.pdf>> accessed November 29, 2024.

²⁷ "Responsible AI, Is Better AI" (*H&M Group*, August 4, 2023) <<https://hmgroupp.com/our-stories/responsible-ai-is-better-ai/>> accessed December 1, 2024.

²⁸ Soumyarendra Barik, "'Automatic Facial Recognition System Is Made Legal by a 2009 Cabinet Note,' Says NCRB" (*MEDIANAMA*, November 14, 2019) <<https://www.medianama.com/2019/11/223-ncrb-afrs-legality/>> accessed November 30, 2024.

In contrast, the European Union has specific legislative measures regulating FRT use under the GDPR, which prohibits processing biometric data for unique identification unless it meets conditions of substantial public interest or involves data made public by the individual (Art 9).²⁹

Usage Limitations

Facial recognition technology (FRT) can be used by governments and law enforcement to identify and track individuals involved in protests or other forms of social or political dissent, which can lead to increased surveillance. This surveillance has the potential to suppress free speech and assembly, especially when biases are reinforced by the technology. The concept of proportionality, as defined in the Puttaswamy judgment, is essential when evaluating the use of FRT. This principle includes several criteria: a legitimate goal, suitability of the measures to achieve this goal, necessity (where no less restrictive but equally effective alternatives are available), and balancing (to avoid a disproportionate impact on the rights of individuals).

In scenarios like attendance monitoring and identity verification, alternative methods that use less sensitive data could be employed, suggesting that FRT may fail the necessity criterion for proportionality. The blanket application of FRT for street surveillance essentially constitutes mass surveillance, which lacks the specificity of targeted surveillance.

In Europe, legislative and judicial actions offer insights into possible regulatory frameworks for FRT in law enforcement. The EU's Law Enforcement Directive limits the use of biometric data to circumstances where it is strictly necessary.³⁰ Although initial drafts of the AI Act in the EU proposed bans on remote biometric surveillance, the final version includes broader exemptions for law enforcement. In the UK, the Court of Appeals ruled that the South Wales Police's deployment of FRT was illegal under the European Convention on Human Rights and the Data Protection Act because it did not meet necessary legal standards and failed to respect public sector equality obligations.³¹ The ruling highlighted that FRT allows for the collection of facial biometrics en masse without the subject's cooperation or awareness, emphasizing its covert and extensive nature.

Addressing Bias and Discrimination

In the UK, the House of Commons' Science and Technology Committee has raised concerns over potential abuses of facial recognition technology and has advised delaying its widespread use until its efficacy and bias issues are resolved.³² Similarly, the European Union's AI Act does not explicitly ban discrimination related to facial recognition technology, but it includes clauses that

²⁹ Council Regulation 2016/679 of 27 April 2016 (General Data Protection Regulation) [2016] OJ L119/38, Art 9.

³⁰ Directive (EU) 2016/680 of the European Parliament and of the Council [2016] OJ L119/1.

³¹ *R (Bridges) v CC South Wales & Ors* [2020] EWCA Civ 1058.

³² Science and Technology Committee, *The work of the Biometrics Commissioner and the Forensic Science Regulator: nineteenth report* (HC 2017-19).

encourage the incorporation of more stringent safety measures.³³ Notably, Article 10 mandates that high-risk AI systems adopt robust data governance and management protocols. This includes a requirement to assess potential biases that could adversely affect personal safety, infringe on fundamental rights, or result in discrimination as defined by EU law, particularly when the outputs from data are used as inputs in future operations.

Conclusion

In essence, the deployment of AI across various business sectors may amount to severe human rights infringement in the form of privacy violations or systematic biases. To minimise such negative consequences, there needs to be adequate accountability, especially in domains such as policing, welfare & warfare. Regulatory frameworks are imperative to implement principles of fairness, justice, and transparency. This can be achieved through international cooperation and domestic frameworks that are aligned with businesses to uphold ethical principles to facilitate the development of AI to promote secure and inclusive progress.

³³ Presidency of the Council of the European Union, "Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts", 2024, <<https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf>>