



INTERNATIONAL LAW  
JOURNAL

---

**WHITE BLACK  
LEGAL LAW  
JOURNAL  
ISSN: 2581-  
8503**

*Peer - Reviewed & Refereed Journal*

The Law Journal strives to provide a platform for discussion of International as well as National Developments in the Field of Law.

[WWW.WHITEBLACKLEGAL.CO.IN](http://WWW.WHITEBLACKLEGAL.CO.IN)

## **DISCLAIMER**

No part of this publication may be reproduced or copied in any form by any means without prior written permission of Editor-in-chief of White Black Legal – The Law Journal. The Editorial Team of White Black Legal holds the copyright to all articles contributed to this publication. The views expressed in this publication are purely personal opinions of the authors and do not reflect the views of the Editorial Team of White Black Legal. Though all efforts are made to ensure the accuracy and correctness of the information published, White Black Legal shall not be responsible for any errors caused due to oversight or otherwise.

WHITE BLACK  
LEGAL

## **EDITORIAL TEAM**

### **Raju Narayana Swamy (IAS ) Indian Administrative Service officer**



Dr. Raju Narayana Swamy popularly known as Kerala's Anti Corruption Crusader is the All India Topper of the 1991 batch of the IAS and is currently posted as Principal Secretary to the Government of Kerala . He has earned many accolades as he hit against the political-bureaucrat corruption nexus in India. Dr Swamy holds a B.Tech in Computer Science and Engineering from the IIT Madras and a Ph. D. in Cyber Law from Gujarat National Law University . He also has an LLM (Pro) ( with specialization in IPR) as well as three PG Diplomas from the National Law University, Delhi- one in Urban Environmental Management and Law, another in Environmental Law and Policy and a third one in Tourism and Environmental Law. He also holds a post-graduate diploma in IPR from the National Law School, Bengaluru and diploma in Public

a professional Procurement from the World Bank.

### **Dr. R. K. Upadhyay**

Dr. R. K. Upadhyay is Registrar, University of Kota (Raj.), Dr Upadhyay obtained LLB , LLM-degrees from Banaras Hindu University & Phd from university of Kota.He has succesfully completed UGC sponsored M.R.P for the work in the ares of the various prisoners reforms in the state of the Rajasthan.



## **Senior Editor**

### **Dr. Neha Mishra**



Dr. Neha Mishra is Associate Professor & Associate Dean (Scholarships) in Jindal Global Law School, OP Jindal Global University. She was awarded both her PhD degree and Associate Professor & Associate Dean M.A.; LL.B. (University of Delhi); LL.M.; Ph.D. (NLSIU, Bangalore) LLM from National Law School of India University, Bengaluru; she did her LL.B. from Faculty of Law, Delhi University as well as M.A. and B.A. from Hindu College and DCAC from DU respectively. Neha has been a Visiting Fellow, School of Social Work, Michigan State University, 2016 and invited speaker Panelist at Global Conference, Whitney R. Harris World Law Institute, Washington University in St.Louis, 2015.

### **Ms. Sumiti Ahuja**

Ms. Sumiti Ahuja, Assistant Professor, Faculty of Law, University of Delhi,

Ms. Sumiti Ahuja completed her LL.M. from the Indian Law Institute with specialization in Criminal Law and Corporate Law, and has over nine years of teaching experience. She has done her LL.B. from the Faculty of Law, University of Delhi. She is currently pursuing Ph.D. in the area of Forensics and Law. Prior to joining the teaching profession, she has worked as Research Assistant for projects funded by different agencies of Govt. of India. She has developed various audio-video teaching modules under UGC e-PG Pathshala programme in the area of Criminology, under the aegis of an MHRD Project. Her areas of interest are Criminal Law, Law of Evidence, Interpretation of Statutes, and Clinical Legal Education.



### **Dr. Navtika Singh Nautiyal**

Dr. Navtika Singh Nautiyal presently working as an Assistant Professor in School of law, Forensic Justice and Policy studies at National Forensic Sciences University, Gandhinagar, Gujarat. She has 9 years of Teaching and Research Experience. She has completed her Philosophy of Doctorate in 'Intercountry adoption laws from Uttranchal University, Dehradun' and LLM from Indian Law Institute, New Delhi.



### **Dr. Rinu Saraswat**

Associate Professor at School of Law, Apex University, Jaipur,  
M.A, LL.M, Ph.D,

Dr. Rinu have 5 yrs of teaching experience in renowned institutions like Jagannath University and Apex University. Participated in more than 20 national and international seminars and conferences and 5 workshops and training programmes.

### **Dr. Nitesh Saraswat**

E.MBA, LL.M, Ph.D, PGDSAPM

Currently working as Assistant Professor at Law Centre II, Faculty of Law, University of Delhi. Dr. Nitesh have 14 years of Teaching, Administrative and research experience in Renowned Institutions like Amity University, Tata Institute of Social Sciences, Jai Narain Vyas University Jodhpur, Jagannath University and Nirma University.

More than 25 Publications in renowned National and International Journals and has authored a Text book on Cr.P.C and Juvenile Delinquency law.



### **Subhrajit Chanda**

BBA. LL.B. (Hons.) (Amity University, Rajasthan); LL. M. (UPES, Dehradun) (Nottingham Trent University, UK); Ph.D. Candidate (G.D. Goenka University)

Subhrajit did his LL.M. in Sports Law, from Nottingham Trent University of United Kingdoms, with international scholarship provided by university; he has also completed another LL.M. in Energy Law from University of Petroleum and Energy Studies, India. He did his B.B.A.LL.B. (Hons.) focussing on International Trade Law.

## ***ABOUT US***



WHITE BLACK LEGAL is an open access, peer-reviewed and refereed journal providededicated to express views on topical legal issues, thereby generating a cross current of ideas on emerging matters. This platform shall also ignite the initiative and desire of young law students to contribute in the field of law. The erudite response of legal luminaries shall be solicited to enable readers to explore challenges that lie before law makers, lawyers and the society at large, in the event of the ever changing social, economic and technological scenario.

With this thought, we hereby present to you

W H I T E   B L A C K  
L E G A L

# **IMPECTS OF ARTIFICIAL INTELLIGENCE ON HUMAN RIGHTS: A COMPREHENSIVE STUDY**

AUTHORED BY - BILAL KHAN

## **ABSTRACT**

The world is evolving in front of our eyes thanks to artificial intelligence. AI has a huge potential to make our lives better. When it comes to identifying specific diseases, AI-based systems are already beating medical personnel, and their application in the financial sector is opening up credit to previously unbankable individuals. However, AI also has drawbacks that reduce its enormous potential. Since AI-based systems rely on the gathering and use of massive amounts of data to generate predictions, they influence the right to privacy. These predictions have frequently been used to uphold preexisting social patterns of bias and discrimination. A movement to incorporate ethical issues into AI development and use has emerged in response to these unsettling potentials. However, this study shows how valuable it is to assess and deal with the multifaceted effects of AI on society using human rights law. Human rights legislation offers a common language, established norms, and institutional framework to assist in guaranteeing that AI delivers on its promises while avoiding its biggest risks.

By analysing the effects of six existing applications of AI on human rights, our study aims to further the growing discussion on AI and human rights. Our theory acknowledges that social contexts with intricate pre-existing effects on human rights constitute the context in which AI systems are being implemented, rather than a blank canvas. We may observe how contemporary AI implementations affect all of the human rights protected by international law, privacy first, by delving deeply into them. Additionally, we learn about the uneven distribution of AI's benefits and drawbacks for human rights across society and investigate how the human rights framework might be used to reconcile these divergent effects.

## **INTRODUCTION**

The world is evolving in front of our eyes through the use of artificial intelligence, or "AI." We now wear and carry AI-powered devices on our wrists and carry them in our pockets, formerly the domain of science fiction. Automobiles on Today's markets are self-driving, diagnostic tools identify our illnesses, and risk assessment algorithms increasingly determine our fate after being accused of a crime: imprisoned or released. AI has a huge potential to make our lives better. When

identifying specific diseases, AI-based systems are already beating medical personnel, and their application in the financial sector is opening up credit to previously unbankable individuals. Automated hiring systems make the promise to assess job applicants based on their genuine qualifications rather than attributes like age or looks that frequently sway human decision-makers. AI has the potential to increase institutional productivity while lowering costs, which would improve the availability and accessibility of a wide range of services.

However, AI also has drawbacks that reduce its enormous potential. The most important of them is that AI systems rely on the creation, gathering, storing, processing, and utilization of enormous amounts of data, which has an influence on people's right to privacy. Through the application of AI algorithms, seemingly innocent bits of data can reveal some of our most personal secrets by revealing remarkable correlations. The most popular method for implementing AI systems is to "train" them to reproduce existing bias and discriminatory practices, which makes it easy for these systems to continue the results of decisions made by people. Even worse, the "vener of objectivity" that surrounds high-tech systems, in general, can mask the reality that the outcomes they generate are sometimes just as bad as those crafted from the "crooked timber of humanity," if not worse.

However, while this effort was underway, a number of powerful individuals realized how important it was to look at the issues surrounding AI from a human rights standpoint. Two important documents have already come out of this early discussion on AI and human rights.<sup>1</sup> One is the Toronto Declaration, which was made available for signature on May 16, 2018, and it protects the rights to equality and non-discrimination in machine learning systems.<sup>2</sup>

The Toronto Declaration, as its full name implies, emphasizes the possible negative consequences of machine learning on the rights to equality and nondiscrimination and asks for the creation of efficient corrective methods for everyone who is harmed by these systems. The other is Global Affairs Canada's Draft Strategy Paper on the Implications of AI for Foreign Policy and Human Rights, which looks at how AI may affect people's rights to equality, privacy, free speech,

---

<sup>1</sup> For example, Amnesty International launched a structured initiative on Artificial Intelligence and Human Rights in 2017, while the New York-based Data & Society Research Institute hosted a workshop on Artificial Intelligence and Human Rights in April, 2018. See Sherif Elsayed-Ali, "Artificial Intelligence and the Future of Human Rights," Nov. 19, 2023. <https://medium.com/amnesty-insights/artificial-intelligence-and-the-future-of-human-rights-b58996964df5>. Mark Latonero, "Artificial Intelligence & Human Rights: A Workshop at Data & Society." Dec 11, 2023. <https://points.datasociety.net/artificial-intelligence-human-rights-a-workshop-at-data-society-fd6358d72149>.

<sup>2</sup> Toronto Declaration on Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems, Nov. 12, 2023. <https://www.accessnow.org/cms/assets/uploads/2018/05/Toronto-Declaration-D0V2.pdf>.



association, and assembly and offers solutions for these effects.<sup>3</sup>

Although there are numerous interpretations of human rights, ranging from philosophical to moral, our proposal adopts a legal perspective. We understand human rights in terms of the legally enforceable declarations made by the global community. Throughout the three historic documents that comprise the International Bill of Rights.<sup>4</sup> The ratification of new treaties, the release of General Comments, which provide authoritative interpretations of their provisions, and the work of domestic and international courts and tribunals, which have applied the terms of these treaties to particular cases, have all contributed to the development of this body of law over time. By charting the effects of the current use of AI systems in six distinct sectors of endeavor on human rights, our study aims to further the rapidly developing discussion on AI and human rights. We strive to go beyond the prevalent emphasis on how AI affects certain civil and political rights and instead take into account how these technologies are affecting other rights protected by international law, including rights to economic, social, and cultural expression.

### **WHAT IS ARTIFICIAL INTELLIGENCE**

The term "artificial intelligence" lacks a commonly acknowledged definition, despite its increasing prevalence in numerous facets of our existence.<sup>5</sup> Rather, it is an umbrella word that encompasses a range of computing methods and related procedures aimed at enhancing machines' capacity to do intelligent tasks including language processing, pattern recognition, and computer vision. Given the loose definition and the quick development of technology, it is unsurprising that the definition of artificial intelligence evolves over the time. This phenomenon is referred to as the "AI effect" or the "odd paradox": once-state-of-the-art inventions become commonplace and normal and lose their eligibility to be classified as AI, while newer technologies with more remarkable capabilities get designated as AI instead.

It is possible to classify the obscenely vast array of technology, methods, and applications that are included under the AI umbrella into two categories. Knowledge-based systems, "committed to the notion of generating behavior by means of deduction from a set of axioms," make up the first

---

<sup>3</sup> Digital Inclusion Lab, Global Affairs Canada, "Artificial Intelligence: Human Rights & Foreign Policy Implications." Accessed Nov. 14, 2023. [https://docs.google.com/document/d/1fhIJYznWSI7oD3TVJ5CgLGHJMJ2H0uEZiQ9a\\_qKbLG0/edit](https://docs.google.com/document/d/1fhIJYznWSI7oD3TVJ5CgLGHJMJ2H0uEZiQ9a_qKbLG0/edit) ("GAC Strategy Paper").

<sup>4</sup> The "International Bill of Rights" is a term to describe the three most important international human rights instruments, namely the Universal Declaration of Human Rights ("UDHR"), the International Covenant on Civil and Political Rights ("ICCPR"), and the International Covenant on Economic, Social, and Cultural Rights ("ICESCR").

<sup>5</sup> National Science and Technology Council: Committee on Technology, "Preparing for the Future of Artificial Intelligence," Government Report (Washington, D.C.: Executive Office of the President, October 2016).

category.<sup>6</sup> Among these are "expert systems," which are capable of reasoning using codified rules and formal logic. These systems, which are also referred to as "closed-rule algorithms," range from first-generation healthcare diagnostic decision support algorithms to commercial tax preparation software. These systems excel in taking concrete circumstances and deriving the best possible judgments within a given domain by applying predefined rules. However, unless they are combined with some of the strategies outlined below, they are unable to automatically learn new information or make better use of the information they have gathered over time.

The second set of technologies continuously enhances its decision-making capabilities through statistical learning. This new wave of technology—which includes the much-discussed methods known as "deep learning" and "machine learning"—is the result of the acceleration of data collection efforts, the exponential growth of computer processing power, and the dramatic drop in the cost of digital storage.<sup>7</sup> This category of systems includes self-driving cars, facial recognition technology used in law enforcement, natural language processing methods for automated content moderation and translation, and algorithms that suggest videos for you to watch on video streaming services. Although the combined capabilities of these systems are outstanding, their individual dependability can be questionable due to their probabilistic nature. Deep learning computer vision systems, for instance, are nearly as accurate at classifying images as humans are. Nevertheless, sometimes these algorithms make mistakes that a human would never make, such as mistaking a picture of a turtle for a rifle.<sup>8</sup> Additionally, "adversarial examples," or manipulated inputs that cause an algorithm to confidently produce an inaccurate response, have the potential to deceive them.

## WHAT ARE THE HUMAN RIGHTS

Human rights are inherent to our existence as human beings and are not bestowed upon us by any political entity. All people have these fundamental rights, regardless of their gender, race, nationality, ethnicity, color, religion, language, or any other characteristic. They include the most basic, the right to life, as well as those that are essential to a fulfilling life, like the rights to food, shelter, employment, health care, and liberty.<sup>9</sup> The first official text outlining fundamental human rights to be universally protected was the 1948 UN General Assembly adoption of the Universal Declaration of Human Rights (UDHR). On December 10, 2023, the UDHR will celebrate its 75th

---

<sup>6</sup> Nello Cristianini, "On the Current Paradigm in Artificial Intelligence," *AI Communications* 27, no. 1 (January 1, 2024): 37–43, <https://doi.org/10.3233/AIC-130582>.

<sup>7</sup> Gheorghe Tecuci, "Artificial Intelligence," *Wiley Interdisciplinary Reviews: Computational Statistics* 4, no. 2 (2012): 168–80, <https://doi.org/10.1002/wics.200>.

<sup>8</sup> Adam Conner-Simons, "Fooling Neural Networks w/3D-Printed Objects," *MIT Computer Science & Artificial Intelligence Lab (blog)*, November 28, 2023, <https://www.csail.mit.edu/news/fooling-neural-networks-w3d-pri>

<sup>9</sup> What are human rights, United nations, (December 2, 2024), [What are human rights? | OHCHR](#)

anniversary as the cornerstone of all international human rights law.

As mentioned in the introduction, we take a legal interpretation of human rights in this paper. The Universal Declaration of Human Rights ("UDHR"), the International Covenant on Civil and Political Rights ("ICCPR"), and the International Covenant on Economic, Social, and Cultural Rights ("ICESCR") further elaborate on the individual and collective rights that we refer to as human rights. The primary declaration of the rights that each and every person has by virtue of their birth is the UDHR.

Despite the fact that the UDHR was adopted by a non-binding U.N. General Assembly resolution, Canada and many other states have long held the view that states have a binding international law obligation to uphold the human rights and fundamental freedoms enshrined in the [UDHR] as a result of their adherence to the UN Charter.<sup>10</sup>

As a result, international treaties like the ICCPR and ICESCR have legal force behind them for the states who have ratified them. These accords define the obligations of governments with regard to two types of rights and expand upon the international human rights that were initially stated by the UDHR. Although the ICESCR requires states to take steps to gradually realize the economic, social, and cultural rights it protects, taking into account the state's economic condition and resources,<sup>11</sup> the protections of the ICCPR's civil and political rights take effect immediately upon ratification.<sup>12</sup> According to international law, states are obligated to defend human rights. They have an obligation to uphold human rights in their own behavior as well as to stop violations of those rights by a natural or legal person under their control, including companies. Even after services that may have an impact on human rights are privatized, these obligations still exist.

Businesses are now seen as having their own obligations under international law to uphold human rights, particularly after the end of the Cold War.<sup>13</sup> The United Nations Guiding Principles on Business and Human Rights (also known as the "UNGP" or "Guiding Principles") are the most authoritative source for defining the nature and extent of these obligations. In particular, businesses that uphold human rights have an obligation to refrain from directly producing or exacerbating negative effects on human rights via their operations, and to actively work to prevent or lessen such

---

<sup>10</sup> Universal Declaration of Human Rights (10 Dec. 1948), U.N.G.A. Res. 217 A (III) (1948) [hereinafter "UDHR"]

<sup>11</sup> International Covenant on Economic, Social and Cultural Rights (New York, 16 Dec. 1966) 993 U.N.T.S. 3, entered into force 3 Jan. 1976, art. 2(1) [hereinafter "ICESCR"].

<sup>12</sup> International Covenant on Civil and Political Rights (New York, 16 Dec. 1966) 999 U.N.T.S. 171 and 1057 U.N.T.S. 407, entered into force 23 Mar. 1976, art. 2 [hereinafter "ICCPR"].

<sup>13</sup> Guiding Principles, Principle 11.

effects when the company has a direct commercial relationship with the affected parties.<sup>14</sup> As a result, businesses must continue to conduct due diligence procedures in order to recognize, avoid, and lessen significant threats to human rights.

If businesses have a negative impact on human rights, they should address it through legal channels. However, it is the state's responsibility to ensure that individuals who have been subjected to business-related violations of their rights receive adequate compensation through legal and other channels. The Guiding Principles illustrate how current international human rights norms apply to business operations, even if they lack legal power. They also offer helpful suggestions on how enterprises should operate in a way that respects human rights. In any case, the Guiding Principles are crucial to ensuring that the potent new technologies have a good impact on human rights because corporations have created and implemented AI.

As a result, the Guiding Principles will be thoroughly discussed in the next section along with our recommendations on how to handle the human rights implications of AI systems that are now in use.

## **IDENTIFYING THE EFFECTS OF ARTIFICIAL INTELLIGENCE ON HUMAN RIGHTS**

AI is not being created in a vacuum or applied to undefined scenarios. Instead, certain social actors are using AI to make decisions in specialized domains of expertise automatically. In the context of social institutions with preexisting implications for human rights, they are acting to get outcomes that they consider desirable. We can only comprehend the effects of integrating AI into the criminal justice system—or any other human institution on human rights if we adopt a comparative approach that takes into consideration pre-AI world background conditions. The human rights effects of introducing AI will be confused with the continued effects of whatever was there previously unless the human rights implications, both positive and negative, of pre-existing institutional frameworks are discovered and taken into account. Here, we offer a two-step process for averting such problems.

### **FIRST STEP- SETUP THE BASELINE**

As mentioned, the first step in implementing AI in any field of endeavor is to simply think about the consequences for human rights that already exist, both positively and negatively. The first step is to assess the human rights implications of the pre-AI status quo when human decision-making in

---

<sup>14</sup> *ibid.*, Principle 13.

the relevant field has already been replaced by a first-generation automated decision-making technology, such as a closed-rule diagnostic algorithm.

## **SECOND STEP- IDENTIFY THE EFFECTS OF AI**

The second phase is figuring out how the introduction of AI alters the effects on human rights in the industry that uses the technology. Artificial Intelligence can be considered to have a beneficial impact on human rights if it helps the field perform better when it comes to human rights. In contrast, it will be evident that artificial intelligence has a negative impact on human rights if the field of study's performance in this area declines after AI is introduced.

## **THE SEVERAL EFFECTS OF AI ON HUMAN RIGHTS**

Using the two-step paradigm from the previous section, we now investigate the wide-ranging effects on human rights that AI decision-making has in six different fields, this paper focuses on risk assessment of Criminal Justice.

- Criminal justice (Risk Assessment)
- financial (credit ratings)
- Medical care (diagnostics)
- Content Moderation (forcing of norms)
- Human Resources (Employment and Recruiting)
- Education (grading of essays)

## **CRIMINAL JUSTICE (RISK ASSESSMENT)**

The most powerful and terrifying instrument that democratic countries may use to limit a person's ability to exercise their fundamental human rights is the criminal justice system. Due to the serious effects on human rights, society has developed a framework of procedural rights to shield criminal defendants and prisoners from the bias of human judgment, including deliberate abuse of authority and unintentional influences like weariness and bigotry.<sup>15</sup>

Justice systems are increasingly using automated decision-making tools at every stage of the procedure in an effort to be both efficient and fair. This is particularly true for risk evaluations, which are utilized to guide choices regarding parole, sentence, and pretrial custody. Risk assessment

---

<sup>15</sup> Millicent H. Abel and Heather Watters, "Attributions of Guilt and Punishment as Functions of Physical Attractiveness and Smiling," *The Journal of Social Psychology* 145, no. 6 (December 2005): 687–702, <https://doi.org/10.3200/SOCP.145.6.687-703>; Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso, "Extraneous Factors in Judicial Decisions," *Proceedings of the National Academy of Sciences of the United States of America* 108, no. 17 (April 26, 2011): 6889–92, <https://doi.org/10.1073/pnas.1018033108>.

tools can significantly improve the rights of those who are accused and found guilty of crimes, provided that they are fair and accurate. The implication, on the other hand, is that a variety of rights may be negatively impacted by defects or unidentified restrictions in the way such systems function. Based on past data, criminal histories, and other risk variables, the traditional method to risk assessment in the criminal justice system evaluates and forecasts the possibility of a person's future criminal activity. Typically, this approach uses standardized instruments and systematic assessments to classify people into various risk categories, which helps with decision-making procedures like bail, parole, and punishment. The intention is to educate criminal justice professionals about the possible threats that offenders may represent and to assist in directing actions or choices that will improve public safety.<sup>16</sup> Whereas, Artificial intelligence algorithms are used in the criminal justice system to develop risk assessments that estimate the probability of an individual committing crimes or reoffending based on a variety of parameters. These evaluations often generate a risk score by taking into account factors including demographics, criminal histories, and other pertinent data. But issues with algorithmic transparency, potential biases in the data, and the moral ramifications of using AI to make such important judgments surface. Opponents contend that because these systems have the potential to reinforce current injustices and discrimination, their introduction requires close examination and monitoring.<sup>17</sup> Complete automation eliminates the need for manual data collection, entry, and scoring, which carries with it the potential to increase the accuracy of these systems by, for example, allowing additional variables to be taken into account. By interpreting every piece of data consistently, the software improves the predictive accuracy and validity of risk assessment tools.<sup>18</sup>

The rights to life, liberty, and personal security of "low-risk" criminal defendants and offenders may be positively impacted by the current generation of automated risk-assessment technologies.<sup>19</sup> If these instruments prove to be more precise than human beings in assessing the likelihood of recidivism, the incarceration rate and duration of low-risk offenders will be reduced relative to the current system.<sup>20</sup> If fewer crimes occur as a result of these technologies, everyone in society will

---

<sup>16</sup> Thomas H. Cohen, "Automating Risk Assessment Instruments and Reliability: Examining an Important but Neglected Area in Risk Assessment Research," *Criminology & Public Policy* 16, no. 1 (January 1, 2024): 271–79, <https://doi.org/10.1111/1745-9133.12272.5>

<sup>17</sup> The Minnesota Screening Tool Assessing Recidivism Risk 2.0 ("MnSTARR 2.0") under development by the government of the U.S. State of Minnesota is a leading example of a fully-automated risk assessment tool in the criminal justice context. Kenneth C. Land, "Automating Recidivism Risk Assessment: Should We Stay or Should We Go?," *Criminology & Public Policy* 16, no. 1 (January 1, 2024): 231–33, <https://doi.org/10.1111/1745-9133.12271>.

<sup>18</sup> According to Barocas and Selbst, one source of bias is inaccuracies in the selected features. Additional features should, in theory, allow for more accurate generalizations to be developed. Barocas and Selbst, "Big Data's Disparate Impact."

<sup>19</sup> UDHR art. 2.

<sup>20</sup> This assumption has been questioned. Julia Dressel and Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism," *Science Advances* 4, no. 1 (January 1, 2018), <https://doi.org/10.1126/sciadv.aao5580>.

feel more secure in the exercise of their right to personal security.

It is difficult to determine, though, whether the current wave of automated risk assessment tools is improving or worsening the equality and non-discrimination rights of criminal defendants who belong to historically marginalized groups, like the mentally ill and members of ethnic minorities.<sup>21</sup> The findings of the New York City proof-of-concept study indicate that, despite the possibility of systemic biases in the training data automating preexisting social biases against members of these groups, these systems might still be able to lessen the disproportionate representation of members of these groups in the jail and prison populations.

Ultimately, we believe that the most recent generation of automated risk assessment tools, with their opaque nature and the secrecy surrounding their development by the private sector, will likely negatively affect criminal defendants' rights to a fair and public trial before an impartial tribunal<sup>22</sup> and to all the guarantees required for their defense.<sup>23</sup>

## **ADDRESSING THE EFFECTS OF AI ON HUMAN RIGHTS:**

### ***POSITIVE IMPACTS OF AI ON HUMAN RIGHTS IN RISK EVALUATION***

*Efficiency and Fairness:* By swiftly evaluating enormous volumes of data, artificial intelligence (AI) algorithms can improve the efficiency of risk assessment procedures. By lessening human biases, these systems can support consistent and equitable decision-making when they are properly developed and put into place. AI systems have the ability to make objective decisions by using pre-established rules and facts, which can reduce the influence of subjective opinions. Risk assessments may produce more impartial and consistent results as a result of this objectivity.

*Resource Optimization:* AI can assist in the criminal justice system's more efficient resource allocation. Through the identification of persons who pose a greater risk of reoffending, authorities can direct interventions and resources towards the most vulnerable.

### ***AI'S NEGATIVE IMPACT ON HUMAN RIGHTS IN RISK ASSESSMENT***

*Discrimination & Bias:* Biases found in the training data can be inherited by AI models, producing discriminating results. The AI risk assessment tools have the potential to reinforce and magnify current disparities if past data exhibits systematic biases.

---

<sup>21</sup> UDHR art. 3.

<sup>22</sup> UDHR art. 10.

<sup>23</sup> UDHR art. 11(1). Relatedly, the right of criminal convicts under ICCPR art. 14(5) is similarly impacted.

*Lack of Transparency:* A lot of AI algorithms function as "black boxes," making it difficult to comprehend how they choose particular outcomes. Concerns of accountability and the capacity to contest or appeal judgments made by these systems are brought up by this lack of transparency.<sup>24</sup>

*Privacy Concerns:* Sensitive personal data is frequently gathered and analyzed as part of the usage of AI in risk assessment. Because people might not have control over how their data is used or shared, this could violate their right to privacy and give rise to privacy issues.<sup>25</sup>

*Over-reliance on Technology:* Using artificial intelligence (AI) too much in decision-making could reduce the importance of human judgment and supervision. This over-reliance may result in AI systems' decisions being held less accountable.

*Exacerbation of Social Injustice:* AI systems have the potential to reinforce and magnify current social inequities if they are not properly constructed. Biased algorithms have the potential to disproportionately affect underprivileged communities, hence exacerbating existing imbalances within the criminal justice system.

In order to meet these issues, fairness and transparency principles must be incorporated into the development and application of AI in risk assessment both inside and outside of the criminal justice system, along with continuous monitoring and careful consideration of ethical rules. The key to reducing possible adverse effects on human rights is to include stakeholders in the design process, conduct regular audits, and use diverse and representative training data.

## CONCLUSION

It should now be evident that there is a complicated link between human rights and artificial intelligence. Many civil, political, economic, social, and cultural rights can be impacted by a single AI application, and different people may have good and negative effects on the same right at the same time. The extent to which artificial intelligence (AI) is already being used or will soon be

---

<sup>24</sup> In this vein, New York University's AI Now Institute has developed a framework for public-sector entities in the United States to use in carrying out "algorithmic impact assessments" prior to purchasing or deploying automated decision systems. Dillon Reisman et al., "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability" (New York University AI Now Institute, April 2018), <https://ainowinstitute.org/aiareport2018.pdf>.

<sup>25</sup> For example, the eleven member companies of the Global Network Initiative ("GNI"), which include some of the biggest players in the AI space, commit to "carry out human rights due diligence to identify, prevent, evaluate, mitigate and account for risks to the freedom of expression and privacy rights that are implicated by the company's products, services, activities and operations." GNI member companies are independently assessed every two years to evaluate their compliance with this and other commitments. "Implementation Guidelines" (Global Network Initiative), accessed June 21, 2018, <https://globalnetworkinitiative.org/implementation-guidelines/>.



widely used can be shown by multiplying these effects by the entire spectrum of scenarios. In the past, when society has faced dramatic technological change, we have always managed to find a new balance.

However, because the Universal Declaration of Human Rights was adopted this December 70 years ago, we are in a better position than our ancestors to deal with the change that is upon us. The UDHR provides us with a strong and widely accepted foundation for not only recognizing and righting historical and contemporary wrongs, but also for creating a future that upholds the rights of all people. But for this to happen, we must continue to be aware of how our actions affect other people's rights. Because of this, the Guiding Principles emphasize the need of conducting due diligence both before and during the lifecycle of these potent new technologies.

The increasing focus on human rights-based methods for evaluating and mitigating the social effects of artificial intelligence gives us hope. The fact that so many of the commercial companies leading the AI revolution are realizing their obligation to behave in a way that respects human rights is encouraging to us. But governments have a vital role to play, both in their capacities as the creators and implementers of this technology and as the guarantors of human rights under international law. The private sector cannot, nor should, do it alone. It is impossible to overestimate the essential role that the government plays in identifying and providing remedies for abuses of human rights.

However, assessing and addressing the distributive effects of AI is a political role that is just as important, if not more. It is time for democratic government institutions and processes to accept their responsibility in guiding society through the impending changes, as they are the only ones with the legal authority to decide how to fairly distribute benefits and liabilities throughout society.

