Foreword by

Mr. Rajiv Ranjan

(Advocate General, Jharkhand High Court)

THE HUMAN RIGHTS TAPESTRY

A Multifaceted Perspective

Edited by

Tuhina Sinha
Aditya Nidhi

WALNUT PUBLICATION
INDIA • UK • USA

The Human Rights Tapestry

A Multifaceted Perspective

Foreword by

Mr. Rajiv Ranjan

Advocate General, Jharkhand High Court

Edited by

Tuhina Sinha

Aditya Nidhi

walnutpublication.com

INDIA • UK • USA

CHAPTER 17

AI-Powered Censorship and the Right to Information

AUTHOR: Shataxi Uma Jaiswal

Abstract

After the emergence of social media, the rapidly evolving Artificial Intelligence (hereinafter referred to as AI) is taking a centre stage in the digital world and is becoming a central force across all the sectors. AI has not only become an essential component of search engines but is becoming increasingly significant to governments, private sectors, corporations, security agencies, educational institutions and individuals. AI is said to have the potential to stimulate human intelligence to such an extent that it could replace humans in a variety of fields. AI can not only generate information as part of automated responsive technology mechanism but can also be used as a tool to filter and monitor information, and censor the dissemination of information as part of regressive or preventive approach. AI is being used by governments to monitor the digital space, and detect and censor the information or content which they consider harmful or inappropriate to be consumed by the general public. This AI-powered censorship of dissemination of information directly attacks on the twin ideals of constitutional-democratic societies and human rights which are freedom of expression and the right to information. While there are positive impacts of content moderation, the negative impacts such as violation of freedom of expression or the right to information cannot be overlooked. Democratic societies provide space for dissent, to raise voice against violations of human rights, and to speak against any acts which, either directly or indirectly, impacts the society. Therefore, it raises flags over the role of AI in content moderation or censorship of dissemination of information.

This chapter explores the relationship between AI-powered censorship and the right to information as to how the right gets undermined due to AI-based systems and algorithms. The chapter first explains the role of AI in censorship, on what basis content moderation done, what role the governments have played in content moderation, and on what tools does AI rely in content moderation. Further, the chapter discusses the legal implications of AI-powered censorship as to how it infringes upon the right to information and freedom of speech and expression. This segment also highlights the evolving nature of AI and that why is there a need to put into practice a regulatory mechanism at first place to oversee the practice of content moderation by the governments. Furthermore, the chapter discusses about the nature of the RTI as a human right and where has it been enumerated as a human right. Also, what are some of the major countries which have enacted a law to guarantee the right to information. In the last segment of the chapter, incidents of AI-powered censorship and infringement of the right to Information have been dealt with. This segment also lists down some of the instances where the governments used AI-based systems to censor information, potentially leading to the infringement of the right to information.

Introduction

It has not been easy, if not impossible, to define AI because of the dynamics attached with it. Perhaps this is because why there is no one common accepted definition of AI. In current times, there is hardly any field or area in which AI has not immersed itself. From education to finance, from social media to cybersecurity, AI

has gone to revolutionize almost every aspect of modern life. AI has been defined as a technology that enables computers and machines to stimulate human learning, comprehension, problem solving, decision making, creativity, and autonomy.[223] All the activities mentioned to have been performed by AI are also performed by we, humans. AI, therefore, has become a rapidly evolving transformative force shaping our digital and non-digital actions. Even though AI has assumed a significant role, there are serious concerns in terms of infringement of human rights, in general, and right to information, in particular. AI's role is increasingly getting enlarged to censor information and contents in digital space.

The Right to Information is not just a right but an essential component of democratic ethos which permits the citizens to access information which are with public authorities. With the exercise of the right to information, one gets closer to transparency and accountability. This is the extent of importance RTI holds. So, if such an essential right gets on the verge of getting infringed on the account of AI-based censorship, it warrants regulatory mechanism over AI-based systems. Because it has become impossible for humans to actively and efficiently monitor and moderate the high amount of data or information which is getting generated every second, AI is being deployed by social media platforms and governments to monitor and moderate content/information. In furtherance of this, there are increasing cases of AI-powered censorship and content moderation and its infringement upon the right to information. There needs to be counterbalance between the two, only then can an efficient AI based system function and the right to information be exercised.

AI's Role in Censorship

With the internet becoming part of everyone's lives and vast amount of information getting generated every minute or hour, it has become challenging for humans to monitor and regulate the information which is getting generated and circulated in vast amount every minute. AI has come to the rescue to search engines and social media platforms as they have started using AI based systems or mechanism to moderate and censor contents or information. As part of AI, various tools such as algorithms, machine learning, and natural language processing are deployed to moderate contents generated or processed on internet; they are moderated in such a manner that content which violates laws, rules, regulations, or community guidelines are filtered out and censored. No doubt, this mechanism has positive impacts in terms of efficiency but given the tools which are involved in content moderation to detect and filter out contents raises serious eyebrows over the category of content which may get filtered out and subsequently gets censored. Machine learning models or algorithms uses datasets which trains it to label content as harmful, insensitive, misinformation, hate speech, etc. so that the AI system, using the past learnings, can act on similar patterns and label contents. Algorithms play an essential role in moderating content on social media platforms as they can be used to identify material that has already been banned, as well as detect previously unseen forms of misinformation, by identifying signals that are indicative of malicious content and these tools, some of which are based on machine learning, provide the capacity to manage content at a speed and scale that would not be possible for human moderators operating alone.[224]

Governments have been using AI based systems to censor information but the grounds on which information are censored may vary from country to country. What may be harmful or insensitive information in one

223 Cole Stryker, et al., What is Artificial Intelligence, IBM (Aug. 16, 2024), https://www.ibm.com/topics/artificial-intelligence

224 Tom Lynn, et al., The Use of Algorithms in the Content Moderation Process, Responsible Technology Adoption Blog (Aug. 5, 2021), https://rtau.blog.gov.uk/2021/08/05/the-use-of-algorithms-in-the-content-moderation-process/

country may not be in another country. Countries like China, Russia, Iran, Venezuela have been actively using AI to censor online content. Michael J. Abramowitz, the president of Freedom House, remarked that "AI can be used to supercharge censorship. Surveillance, and the creation and spread of disinformation. Advances in AI are amplifying a crisis for human rights online."[225] Such infringement of human rights is not something which can be visibly identified as in the case of custodial torture, denial of refugee rights, so on and so forth, rather in the case of violation of human rights in the digital space, more dynamics are involved, be it the government, the individual, the internet platforms or the AI based system. In light of this, it has been observed that algorithms are generally poor at contextual interpretation because it is easy for a human to understand whether a post is satirical or not but algorithms can possibly struggle with this nuance.[226] Until and unless, algorithms or machine learning understands such nuances, there always remains a scope for inadvertent risk to the right to information. In the UK, there are currently no legal requirements for transparency measures in content moderation and design choices, as is so in most of the countries and therefore, high quality datasets are required to train AI based systems to moderate contents efficiently.[227]

Legal Implications of AI-Powered Censorship

AI-powered censorship based on content moderation, except in the cases of genuine censorship, generally infringes upon the right to information and the freedom of speech and expression. It is true that there is a need of AI based content moderation to filter and censor content which violates the laws, rules, regulations, or community guidelines but potential violations of human rights arising out of overreach of content moderation and censorship, cannot be overlooked, certainly not in countries governed by constitutional ideals. AI, as a field, is still evolving and in that case if the transparency and accountability mechanism are put into practice, governments will actively use AI based systems to censor information on the internet potentially leading to violation of human rights. Among other concerns associated with AI-powered censorship, the RTI, having thus been regarded as an essential component of governance, is potentially abused. The RTI which is guaranteed under International human rights laws as well as municipal laws cannot be infringed on the account of AI-powered censorship. There are ethical considerations attached with AI-powered censorship as it is devoid of transparency as to what datasets or criteria has it followed to censor an online information.

It needs to be understood that information which floats on digital space is an exercise of an individual's right to freedom of speech and expression and whenever AI based systems censor content or information, it may result in inadvertent suppression of freedom of speech and access to information. AI-related technologies have properties that can give rise to ethical and human rights concerns, such as privacy, bias and discrimination, safety and security, economic distribution, political participation or the changing nature of warfare.[228] It would, therefore, be unjustified to say that because positive impacts outweighs the negative impacts of the AI, the negative impacts can be overlooked. It is preposterous to even think of that societal benefit outweighs individual interests. The debate between individualism v. collectivism is totally unwarranted in this context. To what extent can they be counterbalanced despite the benefits and risks

225 Rina Chandra, AI ?Supercharges? Online Disinformation and Censorship, Thomson Reuters Foundation (Oct. 4, 2023), https://www.reuters.com/article/business/ai-supercharges-online-disinformation-and-censorship-report-warns-idUSL8N3B53P6/

226 Lynn, supra note 224.

227 Lynn, supra note 224.

228 Bernd Carsten Stahl, et al., Exploring Ethics and Human Rights in Artificial Intelligence, 191 Technological F.S.C. (2023).

involved would define the approach one adopts to look at the AI. Therefore, lack of access to and freedom of information

The RTI as a Human Right

With the human rights guaranteed under the constitutional-democratic set-up, what follows as a precondition is that there should be transparency and accountability. What becomes most essential, then, for ensuring transparency and accountability is the RTI. Except the fact that the sense of generality is often attached to the RTI, it is no less than a non-violent weapon. The world today is more interconnected than ever, information being the essential component of it. Not just information but the right information is what sustains the constitutional-democratic set-up in any country. RTI, which is a human right on the global plane and enumerated under International Human Rights Laws, is guaranteed in the form of legal right to access information under Municipal laws. Article 19 of the UDHR (the Universal Declaration of Human Rights), 1948 states that "Everyone has the right to freedom of opinion and expression. This right includes freedom to hold opinions without interference and to seek, receive, and impart information and ideas through any media and regardless of frontiers." Clause 2 of Article 19 of ICCPR (the International Covenant on Civil and Political Rights), 1966 lays down that "Everyone shall have the right to freedom of expression. This right shall include freedom to seek, receive, and impart information of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art or through any media of his choice." As per the resolution 59 of the UN General Assembly adopted in 1946, Freedom of information is an integral part of the fundamental right of freedom of expression. Sweden was the first country to adopt a law with respect to the RTI and Finland followed it. Later, Denmark, Colombia, Chile, Mexico, Uruguay, USA, Canada, Pakistan, Sri Lanka, Nepal, India, Bhutan, etc. provided for the right to information. In India, RTI Act was enacted in 2005 to give statutory right to access information from public authorities.

Incidents of AI-Powered Censorship

AI, driven by algorithms and machine learning, is being deployed by governments as a censorship mechanism to block access to information. Although it has been also potentially used as a censorship mechanism to censor hateful or insensitive contents, but it has also been deployed to be misused so as to curtail the RTI and freedom of speech and expression. Freedom House, in its annual report, has flagged that allows governments to enhance and refine online censorship.[229] Apart from its positive use, AI has become a regressive tool to enhance political agendas and suppress dissent by political activists, voices of civil societies, and individual freedom of expression. At a time when the whole world is getting rapidly digitalised and interconnected, and the world is entering into narrative war era, AI can cause severe detriment to democratic ideals. There is not one but several incidents of AI-powered censorship across the world which serves as an example as to what the future holds for us if this continues. As per the Freedom House researchers, there are as many as 22 countries that has enacted laws requiring or incentivizing internet platforms to use machine learning (ML) to remove unfavourable online speech.[230] As the world may be aware about the Tiananmen Square Protests and Massacre in 1989, the authoritarian regime in China does not want its people to know about the Tiananmen Square Protests and Massacre, and therefore, uses Chatbots which have been

229 Chandra, supra note 225.

230 Tate Ryan-Mosley, How Generative AI is Boosting the Spread of Disinformation and Propaganda, MIT Technology Review (Oct. 4, 2023), https://www.technologyreview.com/2023/10/04/1080801/generative-ai-boosting-disinformation-and-propaganda-freedom-house/

programmed not to answer any question about the Tiananmen Square. Governments have used regulatory mechanism over internet platforms to enhance the online censorship mechanism through AI Chatbot technology. As per the Freedom House, 'the world's most technically advanced authoritarian governments have responded to innovations in AI chatbot technology, attempting to ensure that the applications comply with or strengthen their censorship systems. Legal frameworks in at least 21 countries mandate or incentivize digital platforms to deploy machine learning to remove disfavoured political, social, and religious speech.'[231] Further, Freedom House also flagged concerns over violation of international human rights law. As per the report, of the Freedom House, 'a record 41 governments blocked websites with content that should be protected under free expression standards within international human rights law. Even in more democratic settings, including the United States of America and Europe, governments considered or actually imposed restrictions on access to prominent websites and social media platforms, an unproductive approach to concerns about foreign interference, disinformation, and online safety.'[232] Very recently Brazil banned X, a social media platform, but the decision have been less seen as a regulatory step over big tech giants and more as an attempt to curtail people's right to receive information. The increasing use of AI-powered censorship by governments warrants a transparent and accountable mechanism.

Conclusion

AI-powered censorship presents before the world a complex challenge to make a balance between an effective moderation of content and upholding human rights. As highlighted in the chapter, AI-powered censorship advertently encroaches upon the right to information and freedom of speech and expression. The positive impacts of AI must be carefully studied and adopted as practices in different domains but the worrying factor i.e., misuse and overreach of AI-based content moderation must be stifled. Currently, AI algorithms functions without any transparency, and therefore, it seriously undermines the character of digital space which is susceptible to various nuances. There should be a policy discussion, both at national and international level to protect internet freedom. There is an urgent need of regulatory framework which can effectively protect digital rights in the rapidly changing digital world. AI-powered censorship on the behest of governments must be transparent and any violation of human rights arising out of it must be dealt by an oversight body. It is between the RTI and right information is where the difference must subsist. There is a fear that AI, the idea of which was to create a digital transformation, may turn into a menace which would be tough to suppress if the strict regulatory mechanism are not put into practice and AI-powered censorship is not strictly dealt with by any oversight mechanism.

231 Allie Funk, et al., The Repressive Power of Artificial Intelligence, Freedom House, https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence

232 Funk, supra note 231.