

Legal Challenges of Artificial Intelligence Tools in Regulating Freedom of Expression in Iran: A Comparative Perspective with International Human Rights Instruments

1. Mansour Malekpour Bahabadi¹: PhD Student, Department of Law, Y.C., Islamic Azad University, Yazd, Iran

2. Seyed Mohsen Mirhosaini^{2*}: Assistant Professor, Department of Political Science, Faculty of Law and Political Science, University of Yazd, Yazd, Iran

3. Ghazaleh Kabirabadi³: Assistant Professor, Department of Law, Y.C., Islamic Azad University, Yazd, Iran

*Correspondence: e-mail: mmirhosaini@yazd.ac.ir

Abstract

With the growing application of artificial intelligence (AI) in governance processes including in content regulation and the monitoring of freedom of expression emerging challenges have arisen in the domain of public law and human rights. In the Islamic Republic of Iran, the increasing deployment of intelligent technologies in cyberspace governance, particularly through algorithmic filtering, automated identification of "harmful" content, and surveillance of social media platforms, has raised fundamental questions regarding the compatibility of these tools with the principle of freedom of expression and the legal obligations arising therefrom. The principal research question addressed in this study is: What legal challenges does the use of AI technologies in the regulation of freedom of expression pose in Iran, and to what extent are such practices consistent with international human rights standards. The findings indicate that AI implementation in Iran primarily serves a control-oriented function. The absence of transparent legal frameworks, fair trial guarantees, and effective judicial oversight has resulted in serious threats to the right to freedom of expression. Moreover, the lack of clear criteria for the design and deployment of algorithms has undermined public trust and led to violations of human rights standards. This study adopts a qualitative methodology based on documentary and comparative legal analysis. It examines the jurisprudence of the European Court of Human Rights, interpretations by the United Nations Human Rights Committee, domestic Iranian regulations, and technology-related policies governing cyberspace. Additionally, by utilizing AI-based tools such as network analysis and natural language processing models, the functionality and behavior of content-restricting algorithms are critically assessed.

Keywords: Freedom of Expression, International Human Rights Law, Iranian Legal Framework, Artificial Intelligence, Content Regulation

Received: date: 12 February 2025

Revised: date: 19 May 2025

Accepted: date: 06 June 2025

Published: date: 01 July 2025



Copyright: © 2025 by the authors. Published under the terms and conditions of Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

Citation: Malekpour Bahabadi, M., Mirhosaini, S. M., & Kabirabadi, G. (2025). Legal Challenges of Artificial Intelligence Tools in Regulating Freedom of Expression in Iran: A Comparative Perspective with International Human Rights Instruments. *Legal Studies in Digital Age*, 4(3), 1-15.

1. Introduction

In the contemporary era, emerging technologies particularly artificial intelligence (AI) have profoundly transformed the social, economic, and political structures of societies and have been applied in various fields (Gorgbandi & Nazari, 2025; Shahbazbeigi, 2024). One of the domains most affected by technological developments is freedom of expression and the regulation of digital content. AI, through tools such as machine learning, natural language processing, and big data analytics, has enabled governments and tech companies to increasingly monitor and control information flows a process often framed as "content moderation" (Flew, 2021).

Freedom of expression, as a fundamental human right and a cornerstone of democratic systems, has historically been in tension with the concepts of "public order" and "national security." With the integration of AI into governance frameworks, these tensions have taken on more complex dimensions, as algorithms are now capable of evaluating, filtering, and even censoring content without human intervention (Kaye, 2019). In the Islamic Republic of Iran where constitutional freedoms are balanced against religious, cultural, and security considerations the use of intelligent technologies in content control has raised numerous legal and ethical concerns.

The rationale for selecting this research topic lies in the growing application of AI in content regulation, particularly in closed legal systems such as Iran's, where such practices may pose serious challenges to the realization of fundamental rights, including freedom of expression. The lack of clear and explicit regulations governing the use of intelligent technologies in digital governance, combined with the absence of effective oversight and institutional accountability, raises the critical question of whether these mechanisms are compatible with international human rights norms.

On the other hand, the international human rights legal order particularly through the instruments and jurisprudence of bodies such as the European Court of Human Rights and the United Nations Human Rights Committee has established relatively clear standards regarding the scope, quality, and requirements of permissible restrictions on freedom of expression. A comparative analysis of these international approaches and the current legal situation in Iran can provide a valuable perspective for understanding existing challenges and proposing legal reforms (G. Smith, 2020).

The main research questions of this study are:

- Does Iran possess a defined legal framework for regulating the use of AI in cyberspace governance?
- Are the AI tools currently deployed in Iran compatible with principles of transparency, accountability, and judicial oversight?
- And finally, are Iran's content-restricting algorithms designed in accordance with international human rights standards?

The research problem is therefore articulated as follows: In the absence of a comprehensive and specific regulatory framework in Iran concerning the use of AI in content moderation, the existing trajectory may result in violations of fundamental rights such as freedom of expression. In particular, these tools often operate as "black boxes" lacking operational transparency, meaning that citizens are not even informed of the reasons behind content filtering or account removals. Such opacity not only undermines individual rights but also jeopardizes the legal legitimacy of regulatory institutions (Citron, 2018).

The necessity of this research can be justified on several grounds:

First, the human rights dimensions of AI-driven content regulation have not yet been substantially addressed in Iranian legal literature. Second, the growing deployment of smart filtering tools highlights the urgent need to revise and reform current legal frameworks. Third, by drawing on models from more developed legal systems and international institutions, it may be possible to formulate localized yet rights-oriented regulatory schemes. Accordingly, the central research question can be formulated as follows:

What are the legal challenges of using AI tools in the regulation of freedom of expression in the Islamic Republic of Iran, and to what extent are these practices consistent with international human rights standards?

The working hypothesis of this study is that: In the Iranian legal system, the use of AI in regulating freedom of expression faces serious challenges due to a lack of transparency, ineffective oversight, and contradictions with international human rights principles. The research proceeds by first outlining the conceptual and legal framework of the topic, and then, through an analytical-comparative approach, juxtaposes international standards with the current situation in Iran.

2. Literature Review

In recent years, the intersection of artificial intelligence and freedom of expression has attracted growing attention from legal scholars, theorists, and policymakers. A review of the literature indicates that international academic discourse largely emphasizes the need for rights-based regulation of intelligent technologies and critiques the opacity of algorithmic systems. Among these, critics of AI-driven content moderation argue that such practices are inherently dangerous and call for stringent restrictions on the automation of decision-making processes. Conversely, proponents assert that, with appropriate legal safeguards, AI can serve to support public discourse and reduce verbal violence. In Iran, there is a significant gap in the production of scientific and theoretical literature in this area, and much of the current content moderation is conducted using opaque technologies without a clear legal foundation. This underscores the need for comparative and analytical research. The existing body of research can be divided into three main categories:

2.1. *Macro-level Theories: The Relationship Between Technology and Freedom*

At a macro level, two dominant theoretical perspectives address the relationship between technology (particularly AI) and civil liberties such as freedom of expression: the institutionalist approach and the critical-technological approach. The institutionalist approach, rooted in legal liberalism, views technology as inherently neutral, arguing that it is the way governments and private entities use technology that determines whether it supports or threatens individual freedoms (Flew, 2021: 118). From this perspective, AI-based content regulation is legitimate only if it adheres to principles such as the rule of law, accountability, transparency, and judicial oversight. By contrast, the critical-technological approach advanced in the works of scholars such as Shoshana Zuboff and Julie Cohen argues that intelligent technologies are not merely tools but powerful structures that shape the public sphere. Algorithms and data, from this viewpoint, are integral to mechanisms of social and economic discipline and can systematically restrict or distort freedom of expression (Cohen, 2019: 79). Supporters of this critical perspective maintain that traditional legal systems are ill-equipped to address the complexities of "algorithmic decision-making," as many content moderation decisions are made without human involvement, through autonomous learning systems (Citron, 2018: 130). This process referred to as "automated censorship" can limit expression invisibly and without accountability. However, critics of this view argue that properly designed and regulated AI systems can in fact help safeguard freedom of expression, for instance by identifying hate speech and disinformation, which themselves pose threats to public discourse (Kaye, 2019: 55). From this angle, the emphasis is placed on democratic regulation of technology, rather than its rejection.

2.2. *Specific Theories on Algorithmic Content Moderation*

At the specific level, much of the research focuses on two issues: a) How content-moderating algorithms function, and b) The legal conflicts arising from their use. Notably, Citron (2018) and Gillespie (2018) have argued that content moderation algorithms often lack reliable criteria to distinguish between permissible and impermissible speech. For instance, algorithms may conflate sarcastic, humorous, or critical content with threatening or terrorist material. These shortcomings are largely due to limitations in natural language processing and insufficient sensitivity to cultural contexts (Citron, 2018; Gillespie, 2018).

In some jurisdictions particularly in the United States this issue has been debated under the concept of platform liability. Under Section 230 of the U.S. Communications Decency Act, platforms are not held liable for user-generated content; however, automated systems that delete content can, in effect, make legal decisions without being held accountable (R. K. M. Smith, 2020).

From the standpoint of international human rights law, the UN Human Rights Committee, in General Comment No. 34, emphasizes that any restriction on freedom of expression must be "provided by law, necessary, and proportionate". Automated content removal, without human review and without notifying the user, is inconsistent with these standards (United Nations Human Rights Committee, 2011).

2.3. *Comparative and Human Rights-Based Perspectives*

In the comparative domain, numerous studies have examined how the case law of the European Court of Human Rights aligns or clashes with national policies. In most of these studies, the Court has recognized freedom of expression as a fundamental pillar of a “democratic society,” allowing its restriction only in severe and well-substantiated cases. In cases such as *Delfi AS v. Estonia* and *Magyar Helsinki Bizottság v. Hungary*, the Court has stressed the importance of clear standards, adequate user notification, and effective appeals mechanisms for content removal decisions (G. Smith, 2020).

In Iran, however, such mechanisms are either absent or lack sufficient clarity and judicial oversight. Bodies such as the Supreme Council of Cyberspace, the Committee for Determining Instances of Criminal Content, and security agencies are the principal actors in content regulation, yet the legal transparency and accountability of these institutions remain highly questionable. Domestic studies in Iran remain limited, though some research such as that by Jafari and Soleimani (2022) and Karimi and Hatami (2021) has pointed out that smart filtering practices in Iran are often implemented without adequate assessment of their human rights implications, and existing laws are outdated and insufficient.

2.4. *Freedom of Expression*

Freedom of expression, in its simplest definition, refers to the right of individuals to express thoughts, opinions, criticisms, and information without fear of punishment or governmental censorship. This concept has been described in various texts using terms such as “freedom of opinion,” “freedom of information,” and “freedom of the press,” and is subject to both broad and narrow interpretations. Some approaches restrict it solely to oral or written expression, while others extend it to domains such as art, clothing, body language, and even silence (Merry, 2006). Freedom of expression is one of the most fundamental concepts within the traditions of political thought, moral philosophy, human rights, and communication theory. Although it may initially appear to be a straightforward notion, upon deeper examination, it reveals a complex, multi-layered, and context-sensitive structure. Freedom of expression is not merely an individual right to articulate thoughts and emotions; it serves as the cornerstone of democratic participation, the free flow of information, intellectual development of societies, and the safeguarding of human dignity (Putnam, 2000).

Freedom of expression is recognized in international instruments as one of the core human rights. Within the framework of international human rights law, freedom of expression holds a distinguished and foundational status not only as an individual right, but also as an essential pillar for the protection of other democratic rights and values (United Nations Human Rights Committee, 2011). Indeed, it constitutes the shared platform upon which dialogue, critique of power, intellectual progress, realization of civil rights, and public participation become possible. Without the guarantee of this right, many other rights including press freedom, the right to access information, freedom of assembly, and the right to vote would be rendered meaningless and hollow. The international human rights system, through the adoption of various global, regional, and even thematic instruments, has sought to define and protect this right comprehensively. The following section examines the most important legal instruments and provisions in this regard, followed by an analysis of their functions, scope, and limitations. Some of the key definitions include:

A) Article 19 of the Universal Declaration of Human Rights (UDHR)

The formal starting point for the international recognition of freedom of expression can be found in the UDHR, adopted in the aftermath of the atrocities of World War II to safeguard the inherent dignity of all human beings. Article 19 states (Berlin, 1958): “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”. This article, for the first time in human history, framed freedom of expression not merely as speech, but as a broad communicative right. Noteworthy elements of this article include:

- The threefold dimensions of freedom of expression:
 1. Freedom to hold opinions without interference;
 2. Freedom to seek and receive information;
 3. Freedom to impart and disseminate ideas freely.

- Media universality: The article imposes no restriction on the means of expression be it speech, writing, print, imagery, or digital media.
- Transnational nature: The phrase “regardless of frontiers” explicitly affirms the global entitlement of individuals to transmit information and ideas across national boundaries a point of significant importance in the era of globalization and transnational communication.

B) International Covenant on Civil and Political Rights (ICCPR)

Within the ICCPR, freedom of expression is recognized not only as an individual right, but also in relation to its function in popular sovereignty and civil society. In contrast to the UDHR, which is not legally binding, the ICCPR is a binding treaty for its state parties. Article 19 of the Covenant thoroughly addresses this right and outlines both its scope and permissible limitations. The article is divided into two main parts:

- Paragraphs 1 and 2:

These paragraphs affirm the right of all individuals to “hold opinions without interference” and to “seek, receive, and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing, in print, in the form of art, or through any other media of their choice.” ([United Nations Human Rights Committee, 2011](#))

- Paragraph 3:

This paragraph acknowledges the possibility of restrictions on this right, provided that such limitations:

- Are provided by law;
- Are necessary for respect of the rights or reputations of others, or for the protection of national security, public order, public health, or morals.

The importance of Article 19 of the ICCPR lies in its explicit recognition of the right while simultaneously delineating the legitimate framework for its restriction, thus minimizing the potential for arbitrary limitation ([Omidi, 2021](#)).

C) European Convention on Human Rights (ECHR)

The Convention upholds freedom of expression while permitting legitimate restrictions to safeguard national security, public order, morality, and the rights of others. These definitions affirm that freedom of expression is not confined to speech alone, but extends to all forms of expression written, visual, artistic, symbolic, media-based, and digital. The ECHR, regarded as one of the most advanced regional human rights instruments, defines freedom of expression in Article 10 as follows: “Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers” ([Mahmoudi, 2019](#)). The article further clarifies that the exercise of these freedoms carries with it duties and responsibilities, and may be subject to restrictions that are necessary in a democratic society for:

- National security
- Territorial integrity
- Prevention of disorder or crime
- Protection of health or morals
- Protection of the reputation or rights of others
- Preventing the disclosure of confidential information
- Maintaining the authority and impartiality of the judiciary

The ECHR not only affirms the right to freedom of expression but also underscores the principles of proportionality, necessity, and legitimacy in imposing restrictions. The case law of the European Court of Human Rights has shown it to be a strong guardian of this right within contemporary judicial systems.

D) Other Regional and International Instruments

In addition to the above, freedom of expression has been recognized in several other legal instruments, including:

- Charter of Fundamental Rights of the European Union (2000), Article 11;
- American Convention on Human Rights, Article 13;
- African Charter on Human and Peoples' Rights, Article 9;
- Arab Charter on Human Rights.

All these documents guarantee freedom of expression in varying forms. Some, such as the American Convention, place greater emphasis on political and media-related immunity of expression (Zamani, 2021).

3. Theoretical and Conceptual Framework of the Research

The theoretical framework of the present research, emphasizing the intersection of three key domains human rights, freedom of expression, and artificial intelligence technologies demonstrates that content regulation in digital spaces is increasingly facing profound challenges. Traditional legal systems, particularly in countries such as Iran, remain largely unprepared to adapt to this fundamental transformation. Therefore, revisiting domestic laws, clarifying content regulation standards, and harmonizing them with international norms appear to be urgent imperatives.

In the legal and philosophical analysis of the relationship between technology and freedom of expression, the central question is whether a sustainable balance can be achieved between technological advancement and the preservation of fundamental human freedoms. The theoretical foundation of this research rests on the premise that freedom of expression is not a political privilege, but rather a fundamental human right, which under the current conditions of the digital world is profoundly influenced by algorithmic decision-making and artificial intelligence. Human rights constitute a trans-legal and transnational concept grounded in human dignity. Among these fundamental rights, freedom of expression occupies a prominent place, as explicitly affirmed in numerous international legal instruments, including (but not limited to):

Table 1. Legal Provisions and Articles on Human Rights and the Right to Freedom of Expression

Instrument	Article	Provision
Universal Declaration of Human Rights	Article 19	Everyone has the right to freedom of opinion and expression.
International Covenant on Civil and Political Rights	Article 19, para. 2	Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information.

The theoretical framework of this research emphasizes the intersection of three domains: human rights, freedom of expression, and artificial intelligence technologies. It illustrates that content regulation in digital spaces is facing escalating challenges. Traditional legal systems especially in countries such as Iran are not yet equipped to adapt to this fundamental transformation. Thus, revising domestic laws, clarifying content regulation standards, and aligning them with international norms is deemed an urgent necessity.

From a legal and philosophical perspective, the core question is whether a sustainable balance can be established between technological advancement and the preservation of fundamental human freedoms. The present study rests on the premise that freedom of expression is not a mere political privilege but a fundamental human right that, in today's digital age, is increasingly influenced by algorithmic decision-making and artificial intelligence.

Human rights are a trans-legal and transnational concept grounded in human dignity. Among these fundamental rights is freedom of expression, as articulated in numerous international instruments, including:

The Principle of Freedom of Expression in Human Rights Law

According to the UN Human Rights Committee (2011: para. 11), the principle of freedom of expression possesses three core characteristics:

- **Universality:** This right is recognized for all human beings without distinction
- **Tolerance of Diverse Content:** Freedom of expression encompasses even unpleasant, critical, or offensive content, provided it does not fall within legally defined exceptions
- **Interrelation with Other Rights:** Freedom of expression is intrinsically linked to other rights, such as freedom of assembly, political participation, and the right to access information

From a theoretical perspective, the views of John Stuart Mill and Isaiah Berlin are essential to the philosophical understanding of this right. Mill underscores the role of freedom of expression in the "pursuit of truth" (Mill, 1859), whereas Berlin analyzes it within the framework of "negative liberty," i.e., the absence of state interference in individual expression (Berlin, 1958).

In the past, states were the sole actors in content regulation. However, in the contemporary era, platform companies and artificial intelligence technologies have emerged as dominant players. Algorithms, through big data analytics and automated

decision-making, have become modern instruments of power. Technological regulators refer to tools that filter, delete, or amplify digital content via algorithms, machine learning, and automated systems. These technologies operate on four levels:

Table 2. Technological Regulators

Level	Function	Example
Content Identification	Text, audio, and image analysis	Hate speech detection via NLP
Content Prioritization	Ranking and ordering of content display	Facebook's News Feed algorithm
Content Removal	Filtering, blocking, automatic reporting	Automatic deletion of violent videos on YouTube
Behavior Prediction	User behavior analysis for anticipatory censorship	Probable flagging on X/Twitter

These technologies generally operate within a “black box,” meaning that their input data, criteria, and decision-making processes are not transparent or accessible even to oversight bodies (Pasquale, 2015). As a result, users are neither informed of the grounds for content removal nor provided with avenues for objection or appeal. The interaction between technology and human rights creates a conceptual intersection, where traditional human rights must be translated into the language of the digital age. Three theoretical approaches are noteworthy:

➤ The “Rights by Design” Theory

This theory argues that human rights should be embedded in the initial design of algorithms. Legal foresight must be integrated into platform architecture, and algorithms should be developed in a way that avoids excessive restriction (Binns, 2018: 289). This approach aspires to the realization of ethical technology.

➤ The “Algorithmic Authority” Theory

This perspective holds that algorithms have acquired a form of legal authority by making decisions with legal implications (Gillespie, 2018). However, this authority lacks legal and democratic legitimacy, as it is neither transparent, nor accountable, nor based on informed consent.

➤ The “Post-Digital Rights” Theory

This critical theory argues that traditional human rights frameworks have become ineffective in the face of AI and must be redefined. AI has blurred the lines between lawmaker and subject, creating scenarios that traditional legal systems are ill-equipped to address (Cohen, 2019).

In the legal system of the Islamic Republic of Iran, the general principles of freedom of expression are articulated in Article 24 of the Constitution, and in select provisions of the Press Law and the Computer Crimes Law. However, these frameworks:

- Were developed prior to the emergence of AI and digital platforms;
- Lack specific regulations for algorithmic content governance;
- Do not provide for transparency or accountability of filtering bodies.

As a result, AI-based content regulation is applied without clear standards, leading to a restriction of freedom of expression without citizens being informed of removal processes, underlying criteria, or opportunities for appeal. This situation conflicts with Article 19 of the International Covenant on Civil and Political Rights and General Comment No. 34 of the UN Human Rights Committee (United Nations Human Rights Committee, 2011). The following section presents a conceptual model illustrating the relationship among the three core components of this study.

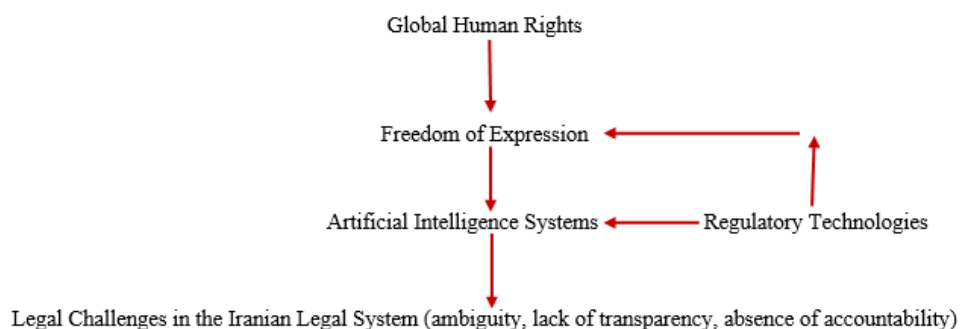


Figure 1. A Conceptual Model of the Relationship Between the Three Core Components of the Article

4. Discussion and Analysis

The use of artificial intelligence in content regulation is rapidly expanding, particularly in countries like Iran, while simultaneously raising significant legal and ethical challenges. This technology especially through machine learning algorithms and natural language processing plays a prominent role in the automatic filtering and removal of content. However, the lack of transparency in algorithmic operations (Pasquale, 2015) and the absence of judicial oversight and user notification may result in the violation of fundamental rights, including the right to freedom of expression. From a legal standpoint, adherence to four fundamental principles is essential: legality, necessity, proportionality, and transparency. The International Covenant on Civil and Political Rights (ICCPR), in General Comment No. 34, identifies these criteria as the preconditions for the legitimacy of any restriction on freedom of expression (United Nations Human Rights Committee, 2011). However, in Iran, the domestic legal framework not only lacks codified regulations for AI-driven content regulation, but also pays insufficient attention to key principles such as the right to appeal or continued access to information. In practice, bodies such as the Supreme Council of Cyberspace and the Working Group for Determining Criminal Content issue decisions that, from an international human rights perspective, require transparency and reviewability. Moreover, from ethical and societal perspectives, content removal without human judgment can result in the erosion of public discourse, self-censorship, and even a retreat from democratic society. Automated algorithms sometimes misclassify satirical or critical content as harmful illustrating cognitive errors in NLP systems (Gillespie, 2018).

The central question, therefore, is whether these challenges can be addressed solely through redesigning domestic legal frameworks or implementing transparent and accountable technologies, or whether Iran's legal system must be fundamentally restructured to require governmental institutions to ensure algorithmic systems are explainable, testable, and subject to oversight, with legal remedies available in the event of violations.

4.1. Comparative Analysis: Iran's Legal Context vs. International Instruments and Human Rights Case Law

In an era where emerging technologies especially AI are rapidly transforming digital governance, the regulation of freedom of expression has become a central and contentious issue in public law. Artificial intelligence, with its ability to process massive datasets, recognize patterns, and analyze content, has become a powerful tool in the hands of governments, tech corporations, and even security agencies to manage information flow.

However, without transparent, accountable, and human rights-based legal frameworks, such technical power becomes a serious threat to fundamental freedoms, especially the right to freedom of expression. In the Islamic Republic of Iran, while mechanisms for restricting digital content are vigorously enforced, the use of advanced technologies, including machine learning algorithms and AI systems, in filtering online content is expanding without dedicated legislation or independent oversight bodies. This approach is inconsistent with international human rights law and the jurisprudence of international human rights courts, particularly the ICCPR, which not only recognizes freedom of expression but also establishes stringent conditions for permissible restrictions. In what follows, we conduct a comparative legal analysis between the AI-driven regulation of freedom of expression in Iran and the standards set by international instruments. This comparison will clarify the gaps in Iran's legal framework and offer a basis for reform proposals drawn from global best practices. First, the principles and requirements of international instruments governing AI-based content regulation will be presented. Then, Iran's legal approach will be examined alongside precedents from international human rights tribunals, including the European Court of Human Rights (ECtHR). The objective is to demonstrate how internationally accepted legal norms can guide Iran in harnessing the potential of new technologies while avoiding excessive restrictions on freedom of expression, paving the way for responsible and transparent digital governance.

- International Standards for Freedom of Expression

According to Article 19 of the ICCPR and General Comment No. 34, freedom of expression is grounded in three key principles: legality, necessity, and proportionality (United Nations Human Rights Committee, 2011). In addition, transparency, accountability, and access to review mechanisms are essential. These principles are also affirmed in the jurisprudence of the ECtHR, especially in landmark cases such as *Delfi* and *Magyar Helsinki Bizottság*, where the court emphasized that restrictions must be clearly defined and reviewable (G. Smith, 2020).

- Comparative Assessment of Iran

A) Legality

While the Iranian Constitution (Article 24) and certain criminal and press laws limit freedom of expression based on Islamic criteria and public order, these laws predate the emergence of AI technologies and lack a clear definition of algorithmic regulators. Moreover, the use of AI tools by state entities without explicit legislative approval or parliamentary oversight presents a significant legal concern.

B) Necessity and Proportionality

Content-filtering algorithms, despite their propensity for error, often operate without human evaluation, leading to possible misclassification of satirical or critical material. These systems frequently fail to offer users avenues for review or correction. In contrast, international standards require that even in cases of filtering, mechanisms for objection must be available ([United Nations Human Rights Committee, 2011](#)).

C) Transparency and Accountability

Transparency in Iran remains severely limited. Users and media professionals do not know the criteria by which content is algorithmically filtered, nor how channels, keywords, or specific topics are assessed. However, courts such as the ECtHR in *Delfi* have ruled that no decision regarding content should be made without transparent standards ([G. Smith, 2020](#)).

D) Oversight and Right to Appeal

Under European and UN human rights systems, even private platforms must provide independent internal and external review mechanisms. In Iran, the formal legal system does not provide for independent judicial review of content removal. The current model is hierarchical and lacks redress avenues, violating the right to access courts under international law.

- Comparative Case Studies

A) ECtHR – *Delfi AS v. Estonia*

In this case, a news platform was held liable for offensive user comments. The ECtHR ruled that users must have clear access to complaint mechanisms, and stringent standards for accountability and transparency must be met ([R. K. M. Smith, 2020](#)). The ruling demonstrates that even for user-generated content, both state and platform regulations must ensure human oversight, clarity of rules, and appeal opportunities.

B) Iran's Practice

In Iran, decisions to block websites or posts are typically made without public announcements or accessible complaint procedures. Users often learn of such actions only through informal or state-controlled media, a practice that contradicts international due process standards. Key Findings of the Comparative Analysis:

- 1) Institutional Differentiation: Platforms, the government, and oversight bodies must have distinct yet coordinated roles, including the establishment of an independent judiciary for complaints.
- 2) Transparency of Criteria: It is essential that content filtering and removal standards be published and accessible.
- 3) Traceability and Reviewability: Users must have access to appeals mechanisms and review procedures.
- 4) Independent Oversight: Entities such as human rights institutions or technical auditors should monitor algorithmic performance and safeguard rights.

Given the above standards, it is strongly recommended that Iran pursue comprehensive legal reform and implement transparent and reliable technologies to move toward compliance with international human rights obligations.

3-2) Proposed AI-Based Legislative Model for the Regulation of Freedom of Expression

The following is a proposed AI-driven legislative model for regulating freedom of expression in the Islamic Republic of Iran. While recognizing the necessity of utilizing modern technologies such as artificial intelligence in digital governance, this model emphasizes compliance with fundamental human rights principles and international standards on freedom of expression. It is structured around five core layers, drawing upon comparative legal experiences, human rights principles, responsible technology, and the specific legal context of Iran.

Layer I: Legal Foundations and Guiding Principles

- ❖ Principle of Legality

Any deployment of AI tools for analyzing, filtering, or removing content must be grounded in legislation passed by Parliament not merely in executive decisions or administrative directives. The legal framework must include:

- a) Clear definitions of terms such as “automated filtering”, “inappropriate content detection”, and “algorithmic removal”
- b) A narrowly drawn boundary between prohibited content based on limited, clear, and predictable concepts

- ❖ Principle of Necessity and Proportionality

Any restriction imposed through AI must:

- a) Pursue a legitimate aim (e.g., national security or public order)
- b) Be strictly confined to necessary situations
- c) Utilize tools that impose the least possible infringement on freedom of expression

- ❖ Principle of Transparency and Accountability

AI systems must:

- a) Be explainable
- b) Be subject to periodic auditing
- c) Provide public reports on their operations under the oversight of an independent body

Layer II: Institutional Framework and Oversight Mechanisms

- ❖ Establishment of an Independent Digital Regulator

A legally established independent body composed of representatives from:

- a) Legal experts in human rights and technology
- b) Computer scientists and AI ethics specialists
- c) Civil society actors and user representatives

Responsibilities include:

- a) Oversight of public and private AI systems operating in digital spaces
- b) Addressing user complaints related to automated content moderation
- c) Issuing specialized regulations for domains such as domestic platforms

- ❖ Legal Remedies and Redress Mechanisms

Guaranteeing users’ rights to:

- a) Be informed of the reasons for content removal
- b) Request a review of the algorithmic decision
- c) Seek recourse through an independent judicial authority in case of rights violations

Layer III: Technical Requirements and Human-Rights-Oriented Algorithm Design

- ❖ Human-Rights-Compliant Algorithm Design

AI algorithms must:

- a) Be trained on local, non-discriminatory datasets
- b) Be conservative in moderating critical, political, or artistic content (avoiding absolute deletion)
- c) Be capable of distinguishing between illegal and lawful content with a low margin of error

- ❖ Independent Auditing of AI Systems

All AI systems must:

- a) Undergo regular technical testing by a neutral third-party entity
- b) Publicly disclose performance results, error rates, and filtering patterns
- c) Be suspended in case of proven discriminatory outcomes

Layer IV: International Safeguards and Human Rights Alignment

- ❖ Compatibility with the ICCPR

Domestic laws must:

- a) Align with interpretations of the UN Human Rights Committee (e.g., General Comment No. 34)
- b) Include structured provisions on “access to information,” “digital privacy,” and “right to appeal”

- ❖ Recognition of International Human Rights Jurisdiction

Providing avenues for submitting digital rights complaints to international bodies such as the ICCPR Committee or UN Human Rights Council mechanisms would enhance the credibility of domestic policies and prevent rights violations.

Layer V: Public Education, Transparency, and Collective Responsibility

- ❖ Digital Literacy and User Awareness Campaigns

Implementation of public campaigns to educate users on:

- a) Responsible use of digital spaces
- b) Awareness of digital rights and content moderation appeal mechanisms
- c) Distinction between legal censorship, automated filtering, and voluntary self-censorship

❖ Transparency in Governmental Operations

Publication of quarterly reports including:

- a) Volume of content removed by algorithms
- b) Reasons for page or keyword filtering
- c) Registered user complaints and resolution processes

This proposed model seeks to balance two core objectives:

- 1) Intelligent utilization of emerging technologies for managing the digital environment
- 2) Protection and enhancement of freedom of expression, as a pillar of human rights and democratic governance

By emphasizing the creation of independent regulatory institutions, legal reform, ethically-aligned algorithm design, alignment with international human rights standards, and the enhancement of public awareness, this model can provide a solid foundation for developing a modern, humane, and transparent legal framework for the Islamic Republic of Iran.

Table 3. Summary of the Proposed AI-Based Legislative Model for Regulating Freedom of Expression

Layer	Title	Key Themes	Proposed Tools or Institutions
1	Guiding Principles	Legality, Proportionality, Transparency	Reform of Press Law, Enactment of AI-Specific Legislation
2	Institutional Structure	Independent Regulator, Judicial Review	Digital Rights Commission
3	Technical Requirements	Rights-Based Algorithms, Independent Auditing	Ethical AI Auditing Systems
4	Global Alignment	ICCPR, International Arbitration	Acceptance of Reporting to the Human Rights Committee
5	Public Education	Digital Literacy, Institutional Transparency	Educational Campaigns, Public Reports

5. Conclusion

At the conclusion of this study focused on the challenges and policy solutions related to the use of artificial intelligence (AI) in regulating freedom of expression in the Islamic Republic of Iran, and its comparison with international human rights instruments a set of critical findings, analyses, and recommendations emerge, addressing both the theoretical and legislative-operational dimensions of the issue.

First, freedom of expression, as one of the fundamental human rights, has entered a stage where traditional legal and political methods are no longer sufficient for its protection or restriction, particularly in the face of new technologies. AI systems, with their immense capacities for processing, classifying, and removing digital content, offer both regulatory opportunities in cyberspace and considerable risks to fundamental rights especially when deployed in legal vacuums, without oversight, and for political ends. A comparative analysis between Iran and international human rights frameworks reveals that:

- There is no clear legal framework for the use of AI in Iran;
- Institutional transparency and accountability in automated content filtering are extremely limited;
- There is no independent oversight body or effective grievance mechanism to address violations of freedom of expression via technology;
- AI-based filtering in Iran is predominantly driven by security-centered and unilateral approaches, which contrary to the principles of necessity and proportionality lead to broad, unpredictable restrictions on free expression.

In contrast, international human rights case law and recommendations from bodies such as the UN Human Rights Committee, the Council of Europe, and UNESCO emphasize that any deployment of automated technologies in content regulation must be:

- Legal, transparent, challengeable, proportionate, and institutionally safeguarded;
- While also preserving the public interest and users' right to know.

One of the central theoretical concepts in this study is the "paradox of order and freedom" in digital space. That is, while governments seek to maintain security and social cohesion through content governance, when such regulation is driven by powerful AI tools in the absence of a human rights framework, the outcome may be not greater order, but a decline in legal legitimacy and public trust.

In Iran's legal system, the main challenge lies in the fact that technological tools are used preemptively and primarily as blunt instruments of censorship, not as tools for balancing competing rights. The result is an unregulated curtailment of freedom of expression and a lack of technical and legal accountability contrary to accepted principles in global human rights regimes. In light of the study's findings, the following actions are proposed to bring Iran's legal system in line with international human rights standards while utilizing smart technologies for rights preservation:

- **a) Enactment of a Comprehensive Human Rights-Based Digital Regulation Law**

Such a law should:

- Clearly define the scope and limits of algorithmic content intervention;
- Guarantee institutional oversight and access to effective complaint mechanisms;
- Subject algorithms to the rule of law, transparency, and the right to defense.

- **b) Establishment of an Independent Digital Rights Oversight Body**

This institution comprising legal, technological, civil society experts and user representatives should:

- Conduct ethical and technical assessments of AI-based filtering systems;
- Publish transparent annual reports;
- Offer remedies for users whose rights have been violated.

- **c) Public Education and Empowerment**

Efforts should focus on enhancing digital literacy among citizens and empowering them to identify, critique, and challenge unlawful interventions in cyberspace.

- **d) Harmonization with International Instruments**

This includes:

- Amending domestic laws in accordance with human rights interpretations of the International Covenant on Civil and Political Rights (ICCPR);
- Utilizing the jurisprudence of courts such as the European Court of Human Rights to identify permissible limitations.

The future of regulating freedom of expression in Iran is inevitably entangled with technologies such as AI. If this engagement lacks alignment with human rights principles and international standards, AI risks becoming a tool of repression. However, if developed with a rights-based, human-centered approach, it can serve as an effective model for maintaining digital order while upholding human dignity. Iran's legal system currently stands at a crossroads: it can either continue its current trajectory widening its gap with global digital governance or reform its approach and adopt a hybrid domestic-international model for AI-based regulation of free expression. The latter, though difficult, is achievable requiring simultaneous political will, scientific capacity, and institutional transparency. The main research question can be addressed through two axes: domestic legal challenges and comparative evaluation with international human rights standards:

- 1) Domestic Legal Challenges in the Use of AI to Regulate Freedom of Expression in Iran:
 - Lack of Specific and Preemptive Legislation

Currently, there is no dedicated legislation in Iran directly governing AI usage in content moderation or free speech regulation. Most decisions are made by executive bodies (e.g., the Committee for Determining Instances of Criminal Content) under vague and broad mandates violating the legality principle for restricting rights.

- Conflict with Transparency and the Right to Information

Automated filtering systems in Iran, including locally developed algorithms and digital monitoring tools, operate without notifying users or publishing performance reports. Users remain unaware of why their content is removed or filtered clearly violating transparency and accountability norms in human rights law.

- Lack of Redress Mechanisms

There is no effective official mechanism allowing users to contest AI-based filtering decisions, nor is there access to an independent judiciary for redress. This undermines fair trial rights and the right to remedy.

- Technological and Content-Based Discrimination

Many filtering systems, particularly those using machine learning, are trained on biased data and fail to distinguish legitimate critical content from illegal material. As a result, cultural, political, human rights, and even unofficial religious content is indiscriminately removed representing algorithmic discrimination.

- 2) Compatibility of Iran's Current Practices with International Human Rights Standards

According to key international instruments and jurisprudence especially:

- The ICCPR
- General Comment No. 34 of the UN Human Rights Committee
- Case law of the European Court of Human Rights
- UNESCO's 2021 Recommendation on the Ethics of Artificial Intelligence

Any restriction on freedom of expression, including via advanced tools, must meet five fundamental criteria, as shown in Table 4:

Table 4. The Five Fundamental Principles of Modern Tools for Freedom of Expression and a Comparison with the Situation in Iran

International Principle	Requirements	Iran's Status
Legality	Restrictions must be based on clear, enacted, and predictable law	Vague laws; no AI-specific legal framework (non-compliant)
Legitimate Aim	Restriction must serve a legitimate aim (e.g., public order)	General claim of preserving order exists, but lacks clarity
Necessity	Restriction must be genuinely necessary	Often excessive; no assessment of necessity (non-compliant)
Proportionality	Measures must be proportionate and minimally harmful	AI filtering often excessive and unmeasured (non-compliant)
Right to Remedy	Users must have access to appeal and remedy mechanisms	No effective or neutral complaint mechanism (non-compliant)

Iran's use of AI for online content regulation faces significant legal challenges in terms of legitimacy, transparency, accountability, and remedies. The current framework not only curtails free expression but also conflicts with Iran's international human rights obligations. Without legal, institutional, and technological reforms, the continued deployment of AI in this manner may gravely undermine the legitimacy of Iran's digital policies on both legal and international fronts. AI-based tools for regulating freedom of expression in Iran face multiple legal obstacles including the absence of specific legislation, lack of transparency, violations of proportionality, and lack of redress mechanisms and are, in their current form, substantially inconsistent with international human rights standards. To test the study's hypothesis, it was evaluated based on empirical data, legal documents, domestic and international jurisprudence, and comparative analysis, across three core components:

❖ 1) Transparency

Available evidence includes:

- The absence of official disclosure regarding the types of algorithms or AI-based systems employed for content filtering in Iran (e.g., intelligent filtering systems operated by the Ministry of ICT or the Cyber Police).
- Lack of user notification about the reasons behind content removal or filtering, the identity of the decision-making entity, and whether the action is subject to appeal.
- No public reporting by responsible institutions, such as the Committee to Determine Criminal Content, on the functioning of algorithms or their error rates.

Transparency is a foundational human rights condition for the legitimacy of any restriction on freedom of expression. The lack of transparency in the design, implementation, and oversight of AI-based systems in Iran constitutes a clear violation of this principle.

Conclusion: The hypothesis component of "lack of transparency" is confirmed.

❖ 2) Lack of Effective Oversight

Available evidence includes:

- The Committee to Determine Instances of Criminal Content, which serves as the main decision-making body for filtering, is a non-independent institution with a predominantly political-security composition and operates without transparency.
- The judiciary and other supervisory or judicial bodies lack effective mechanisms for handling user complaints regarding automated filtering.
- There is no independent regulatory body for emerging technologies, such as a "Digital Rights Oversight Authority," as seen in leading countries.

In international human rights law, the principle of effective institutional oversight is essential to prevent the abuse of technology in the restriction of rights. The absence of such mechanisms in Iran has increased decision-making monopolies and raised the risk of violations of the right to freedom of expression.

Conclusion: The hypothesis component of “lack of effective oversight” is confirmed.

3) Incompatibility with International Human Rights Principles

Available evidence includes:

- AI is predominantly used in Iran for large-scale content filtering and removal with political-security motivations, rather than for balancing conflicting rights.
- Core principles such as proportionality, necessity, and the right to appeal, as enshrined in Article 19 of the International Covenant on Civil and Political Rights (ICCPR), are not observed in the processes of intelligent filtering.
- International instruments such as UNESCO’s Recommendation on the Ethics of Artificial Intelligence emphasize human-centered design, yet Iranian systems lack even basic human rights impact assessments.

Comparing Iran’s performance with human rights treaties and the jurisprudence of courts like the European Court of Human Rights reveals a substantial gap between Iran’s implementation policies and international standards in the field of digital freedom of expression.

Thus, the hypothesis component of “incompatibility with international human rights principles” is confirmed.

Table 5. Hypothesis Validation Table

Hypothesis Component	Analytical Findings	Conclusion
Lack of transparency	Multiple pieces of evidence indicating lack of user notification, reporting, and transparency in operations	Fully confirmed
Lack of effective oversight	Absence of an independent body, appeal mechanisms, and external monitoring	Fully confirmed
Incompatibility with international human rights	Violation of principles of proportionality, necessity, and redress; contradiction with international instruments	Fully confirmed

The research hypothesis is fully confirmed. Documented evidence regarding Iran’s domestic practices in the use of artificial intelligence for regulating freedom of expression indicates that this approach not only faces serious legal challenges, but in many cases also contradicts fundamental principles of international human rights. Therefore, urgent reforms are necessary at the levels of legislation, oversight, and technological design.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

Acknowledgments

Authors thank all who helped us through this study.

Conflict of Interest

The authors report no conflict of interest.

Funding/Financial Support

According to the authors, this article has no financial support.

References

- Berlin, I. (1958). *Two Concepts of Liberty*. Oxford University Press.
- Citron, D. K. (2018). *Hate Crimes in Cyberspace*. Harvard University Press.
- Cohen, J. E. (2019). *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford University Press. <https://doi.org/10.1093/oso/9780190246693.001.0001>
- Flew, T. (2021). *Regulating Platforms: Communication Policy and the Rise of Platform Governance*. Palgrave Macmillan.

- Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press. <https://doi.org/10.12987/9780300235029>
- Gorgbandi, S. G., & Nazari, S. (2025). Medical Image Processing of Patients for Skin Cancer Diagnosis Using Artificial Intelligence. *Transactions on Machine Intelligence*, 8(1), 38-46. <https://doi.org/10.47176/TMI.2025.38>
- Kaye, D. (2019). *Speech Police: The Global Struggle to Govern the Internet*. Columbia Global Reports. <https://doi.org/10.2307/j.ctv1fx4h8v>
- Mahmoudi, B. (2019). *Theories of Freedom of Expression in Modern Thought*. Kavir Publishing.
- Merry, S. E. (2006). *Human Rights and Gender Violence: Translating International Law into Local Justice*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226520759.001.0001>
- Mill, J. S. (1859). *On Liberty*. J.W. Parker & Son.
- Omidi, J. (2021). Limitations of Freedom of Expression: Justifications and Legal Foundations. *Journal of Islamic Jurisprudence and Principles of Islamic Law*, 54(1), 29-52.
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674736061>
- Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster. <https://doi.org/10.1145/358916.361990>
- Shahbazbeigi, A. (2024). Application of Artificial Intelligence and Machine Learning in Accounting and Auditing. *Transactions on Data Analysis in Social Science*, 6(2), 21-30.
- Smith, G. (2020). *Internet Law and Regulation*. Sweet & Maxwell.
- Smith, R. K. M. (2020). *Textbook on International Human Rights*. Oxford University Press.
- United Nations Human Rights Committee. (2011). *General Comment No. 34 on Article 19: Freedoms of opinion and expressing*.
- Zamani, S. (2021). *Theoretical Foundations of Freedom of Expression in Public Law*. Institute for Humanities and Cultural Studies.