Submit Date: 21 January 2026
Revise Date: 17 May 2026
Accept Date: 26 May 2026
Initial Publish: 28 July 2026
Final Publish: 23 October 2027

The Encyclopedia of Comparative Jurisprudence and Law

Violent Content of Transnational Platforms and the Deliberate Disregard of Terrorism-Promoting Content in Relation to the December 2025 Unrest in Iran

Ali Yari[1], Valiollah Ansari[2]*, Ali Akbar Esmaeili[1]

1. Department of Law, Go.C., Islamic Azad University, Gorgan, Iran
2. Department of Law, ST.C., Islamic Azad University, Tehran, Iran

* Corresponding Author's Email: Valiollah.ansari@gmail.com

ABSTRACT

The expansion of transnational platforms' influence within the digital ecosystem has generated novel challenges in the domain of the legal responsibility of platform operators. The events of December 2025 demonstrated that these platforms have moved beyond the role of "mere conduits" and, through intentional content management, have become influential actors in security developments. The present study aims to analyze the legal status of incitement to violence from the perspective of international human rights law and domestic legal doctrines, examining the responsibility of platform operators. The findings indicate that the liability of these actors extends beyond "positive acts" in the form of algorithmic amplification to include "omissions" or deliberate failure to control violence. Based on the principle of due diligence and the jurisprudential rule of causation (tasbib), platform operators bear civil and criminal liability due to their technical access to monitoring tools and their conscious disregard of content promoting terrorism and violence. This study emphasizes that silence and non-intervention by platforms during social crises are not neutral acts but constitute violations of international and sovereign legal obligations, resulting in irreversible consequences for human security.

Keywords: *Transnational platforms, legal responsibility, incitement to violence, omission, international human rights law, December 2025 events.*

How to cite: Yari, A., Ansari, V., & Esmaeili, A. A. (2027). Violent Content of Transnational Platforms and the Deliberate Disregard of Terrorism-Promoting Content in Relation to the December 2025 Unrest in Iran. *The Encyclopedia of Comparative Jurisprudence and Law*, *5*(4), 1-23.

تاریخ ارسال: ۱ بهمن ۱۴۰۴
تاریخ بازنگری: ۲۷ اردیبهشت ۱۴۰۵
تاریخ پذیرش: ۵ خرداد ۱۴۰۵
تاریخ چاپ اولیه: ۶ مرداد ۱۴۰۵
تاریخ چاپ نهایی: ۱ آبان ۱۴۰۶

دانشنامه فقه و حقوق تطبیقی

محتوای خشونت‌آمیز پلتفرم‌های فراملی و نادیده گرفتن آگاهانه محتواهای مروج تروریسم در قبال اغتشاشات دی‌ماه ۱۴۰۴ ایران

علی یاری[۱]، ولی الله انصاری[۲*]، علی اکبر اسمعیلی[۱]

۱. گروه حقوق، واحد گرگان، دانشگاه آزاد اسلامی، گرگان، ایران
۲. گروه حقوق، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران
* پست الکترونیک نویسنده مسئول: Valiollah.ansari@gmail.com

چکیده

گسترش نفوذ پلتفرم‌های فراملی در زیست‌بوم دیجیتال، چالش‌های نوینی را در حوزه مسئولیت حقوقی گردانندگان این سکوها ایجاد کرده است. وقایع دی‌ماه ۱۴۰۴ نشان داد که این پلتفرم‌ها از نقش «مجرای محض» خارج شده و به واسطه مدیریت ارادی محتوا، به بازیگرانی اثرگذار در تحولات امنیتی تبدیل شده‌اند. پژوهش حاضر با هدف واکاوی وضعیت حقوقی تحریک به خشونت از منظر حقوق بین‌الملل بشر و دکترین‌های داخلی، به بررسی مسئولیت گردانندگان این پلتفرم‌ها می‌پردازد. یافته‌های تحقیق نشان می‌دهد که مسئولیت این کنشگران فراتر از «فعل مثبت» در قالب تحریک الگوریتمی، شامل «ترک فعل» یا امتناع عمدی از مهار خشونت نیز می‌شود. بر اساس اصل «دقت مقتضی » و قاعده فقهی «تسبیب»، گردانندگان پلتفرم‌ها به دلیل دسترسی فنی به ابزارهای نظارتی و نادیده گرفتن آگاهانه محتواهای مروج تروریسم و خشونت، دارای مسئولیت مدنی و کیفری هستند. این مطالعه تأکید می‌کند که سکوت و عدم مداخله پلتفرم‌ها در بحران‌های اجتماعی، نه یک عمل خنثی، بلکه نقض تعهدات حقوقی بین‌المللی و حاکمیتی است که پیامدهای جبران‌ناپذیری بر امنیت انسانی بر جای گذاشته است.

کلیدواژگان: *پلتفرم‌های فراملی، مسئولیت حقوقی، تحریک به خشونت، ترک فعل، حقوق بین‌الملل بشر، وقایع دی‌ماه ۱۴۰۴.*

نحوه استناددهی: یاری، علی.، انصاری، ولی الله.، و اسمعیلی، علی اکبر. (۱۴۰۶). محتوای خشونت‌آمیز پلتفرم‌های فراملی و نادیده گرفتن آگاهانه محتواهای مروج تروریسم در قبال اغتشاشات دی‌ماه ۱۴۰۴ ایران. *دانشنامه فقه و حقوق تطبیقی*، ۵(۴)، ۲۳-۱.

مقدمه

به زعم فلسفی، ۲۰۲۰: ظهور فضای مجازی و به تبع آن، پلتفرم‌های فراملی، پارادایم‌های سنتی حقوق بین‌الملل را در حوزه‌ی حاکمیت و حقوق بشر با چالشی بنیادین مواجه کرده است. اگرچه آزادی بیان به عنوان یکی از ارکان اساسی حقوق بشر در ماده ۱۹ میثاق بین‌المللی حقوق مدنی و سیاسی (*ICCPR*) به رسمیت شناخته شده است، اما این حق مطلق نبوده و طبق بند ۳ همان ماده و نیز ماده ۲۰ میثاق، با محدودیت‌های جدی، از جمله ممنوعیت تبلیغ برای جنگ و ترویج نفرت ملی، نژادی یا مذهبی که محرک تبعیض، خصومت یا خشونت باشد، روبه‌روست (Falsafi, 2020). در دهه‌ی اخیر، پلتفرم‌های اجتماعی از نقش واسطه‌گری صرف خارج شده و به کنشگران تأثیرگذار در امنیت بین‌المللی تبدیل شده‌اند.

وقایع دی‌ماه ۱۴۰۴ در ایران نشان داد که فضای مجازی می‌تواند به جای بستری برای تبادل آرا، به ابزاری برای سازمان‌دهی اقدامات خشونت‌آمیز و تخریبی تبدیل شود. در این میان، نقش پلتفرم‌های فراملی (مانند اینستاگرام و تلگرام) در مدیریت محتوای تولیدشده، از منظر حقوق بین‌الملل بشر حائز اهمیت است. تعهد دولت‌ها و شرکت‌های فناوری به اعمال «دقت مقتضی» جهت جلوگیری از نقض حقوق بشر و صیانت از حق بر حیات و امنیت شهروندان، به یکی از مباحث داغ در دکترین حقوقی تبدیل شده است (United Nations Human Rights Council, 2011).

این پژوهش درصدد است تا با نگاهی به وقایع اخیر، مرز میان آزادی بیان و تحریک به خشونت را در رویه این پلتفرم‌ها واکاوی نماید.

مسئله اصلی این پژوهش، فقدان نظارت مؤثر و برخوردهای سلیقه‌ای و دوگانه پلتفرم‌های فراملی در مواجهه با محتواهای محرک خشونت در جریان ناآرامی‌های دی‌ماه ایران ۱۴۰۴ است. بر اساس «طرح اقدام رباط»، تحریک به خشونت باید دارای شش عنصر کلیدی (بافت، گوینده، قصد، محتوا، دامنه و احتمال وقوع خشونت) باشد تا مشمول محدودیت قانونی گردد (Office of the United Nations High Commissioner for Human Rights, 2013). در وقایع دی‌ماه ۱۴۰۴، شواهد متعددی مبنی بر انتشار آموزش‌های عملیاتی برای تخریب اموال عمومی، حمله به مراکز نظامی و مذهبی و ترور اشخاص در این بسترها مشاهده شد.

چالش حقوقی زمانی عمیق‌تر می‌شود که این شرکت‌ها با استناد به «مصونیت واسطه‌ای» و استانداردهای داخلی خود، از حذف به‌موقع این محتواها خودداری کرده و یا میان بحران‌های مختلف در کشورهای مختلف تبعیض قائل می‌شوند (Kaye, 2019). از سوی دیگر، دولت‌های میزبان این شرکت‌ها (غالباً ایالات متحده) بر اساس مواد مسئولیت بین‌المللی دولت (*ARSIWA 2001*)، متعهدند اجازه ندهند قلمرو یا اشخاص تحت صلاحیت‌شان برای نقض حقوق دیگر دولت‌ها و امنیت بین‌المللی مورد استفاده قرار گیرد (Schmitt, 2017).

بنابراین، مسئله محوری این است که پلتفرم‌های فراملی در وقایع دی ۱۴۰۴، تا چه میزان از مسئولیت خود در قبال منع تحریک به خشونت تخطی کرده‌اند؟ و آیا رویه‌ی فعلی حقوق بین‌الملل بشر، ابزارهای لازم برای پاسخگو کردن این کنشگران غیردولتی و دولت‌های میزبان آن‌ها را در اختیار دارد؟ این تحقیق تلاش می‌کند خلأهای موجود در حکمرانی بین‌المللی بر فضای مجازی را در پرتو این بحران تحلیل نماید. لذا این پژوهش در پی پاسخگویی به سؤالات زیر است:

وضعیت حقوقی تحریک به خشونت در فضای مجازی بر اساس استانداردهای حقوق بین‌الملل بشر چیست و عملکرد پلتفرم‌های فراملی در قبال این محتواها در جریان وقایع دی‌ماه ۱۴۰۴ بر اساس کدام موازین حقوقی قابل ارزیابی است؟

بر اساس اسناد بین‌المللی (به‌ویژه ماده ۲۰ میثاق و طرح اقدام رباط)، مرز حقوقی میان «آزادی بیان» و «تحریک به خشونت» در فضای مجازی چگونه ترسیم می‌شود؟

پلتفرم‌های فراملی در زمان وقوع ناآرامی‌های داخلی (مطالعه موردی دی ۱۴۰۴) چه تعهداتی در چارچوب اصل «دقت مقتضی» برای مدیریت و حذف محتوای خشونت‌آمیز دارند؟

مسئولیت بین‌المللی دولت‌های میزبانِ این پلتفرم‌ها در قبال عدم نظارت بر انتشار محتوای محرکِ تروریسم و تخریب در جریان وقایع مذکور چیست؟

الف) ماهیت ماده ۲۳۰: «سپر مصونیت گردانندگان»

این قانون مقرر می‌دارد که «هیچ ارائه‌دهنده یا کاربر یک سرویس کامپیوتری تعاملی نباید به عنوان ناشر یا گوینده اطلاعاتی که توسط تأمین‌کننده محتوای اطلاعاتی دیگری ارائه شده است، تلقی شود».

نقد عملیاتی در وقایع دی ۱۴۰۴: در حالی که گردانندگان پلتفرم‌ها از این ماده برای سلب مسئولیت خود در قبال انتشار آموزش‌های خشونت‌آمیز (مانند ساخت کوکتل مولوتوف) استفاده می‌کنند، اما واقعیت فنی نشان می‌دهد که آن‌ها با استفاده از الگوریتم‌های پیشنهاددهنده، از نقش «میزبان بی‌طرف» خارج شده و به «ناشر فعال» تبدیل شده‌اند.

ب) تعارض با تعهدات بین‌المللی (حقوق بین‌الملل بشر)

استناد به ماده ۲۳۰ در سطح بین‌المللی با منشور ملل متحد و میثاق بین‌المللی حقوق مدنی و سیاسی (*ICCPR*) در تعارض است:

۱. بند ۳ ماده ۱۹ میثاق (*ICCPR*): آزادی بیان مطلق نیست و اعمال آن ملازم با «وظایف و مسئولیت‌های خاص» است که می‌تواند برای حفظ امنیت ملی، نظم عمومی و سلامت جامعه محدود شود.

۲. ماده ۲۰ میثاق (*ICCPR*): هرگونه دعوت (ترویج) به نفرت ملی، نژادی یا مذهبی که محرک تبعیض، خصومت یا خشونت باشد، باید به موجب قانون ممنوع شود.

ج) نقد و تحلیل: دکترین «تقصیر در نظارت»

دولت ایالات متحده با حفظ ماده ۲۳۰ در شکل کنونی، عملاً پلتفرم‌های خود را از اجرای استاندارد «دقت مقتضی» معاف کرده است. در حالی که طبق حقوق بین‌الملل، دولت‌ها موظفند اطمینان حاصل کنند که بازیگران غیردولتی (مانند شرکت‌های فناوری) در قلمرو آن‌ها، ابزاری برای نقض حقوق بشر در سایر کشورها قرار نمی‌گیرند.

در وقایع دی ۱۴۰۴، ایران، پلتفرم‌های تحت حاکمیت این قانون، با استناد به مصونیتِ ماده ۲۳۰، از نظارت مؤثر بر محتواهای تروریستی امتناع کردند (ترک فعل). این «تسامح ساختاری» نه تنها یک خلأ قانونی، بلکه نقض تعهد دولت میزبان در قبال صلح و امنیت بین‌المللی است.

برخلاف ادعای گردانندگان پلتفرم مبنی بر لزوم رعایت ماده ۲۳۰ *CDA*، این مقرره داخلی نمی‌تواند محملی برای نقض قواعد آمره بین‌المللی و میثاقین باشد. مصونیت اعطاشده در حقوق داخلی آمریکا، در تقابل آشکار با تعهد به پیشگیری از خشونت و تروریسم قرار دارد. در واقع، ماده ۲۳۰ به ابزاری برای «ترک فعلِ سازمان‌یافته» تبدیل شده است که مسئولیت کیفری و مدنی گردانندگان را در فجایعی همچون وقایع دی ۱۴۰۴، در ایران، به دلیل عدم اقدام مقتضی و آگاهانه، تثبیت می‌کند.»

پیشینه پژوهش

وقایع دی ۱۴۰۴ یک استثنا نبوده، بلکه بخشی از یک الگوی تکرارپذیر جهانی است. بررسی تطبیقی با پرونده‌های حقوقی بین‌المللی «متا» (فیس‌بوک سابق) در میانمار و اتیوپی ضروری است.

این پرونده‌ها نشان می‌دهند که چگونه «بی‌عملی آگاهانه» یا «نقص الگوریتمیک» پلتفرم‌ها، بستری برای جنایت علیه بشریت و تخریب زیرساخت‌های اجتماعی فراهم کرده است.

۱. پرونده میانمار (نسل‌کشی روهینگیا): «الگوریتم در خدمت نفرت»

در سال ۲۰۱۸، بازرسان سازمان ملل متحد اعلام کردند که فیس‌بوک نقشی کلیدی در گسترش گفتمان نفرت‌انگیز ایفا کرده که منجر به نسل‌کشی اقلیت مسلمان روهینگیا شد.

مشابهت با دی ۱۴۰۴: همان‌طور که در گزارش ۱۸ دی ۱۴۰۴ مرکز ملی فضای مجازی به انتشار فراخوان‌های حمله به مساجد اشاره شده، در میانمار نیز صفحات نظامی و افراط‌گرایان از الگوریتم‌های «تعامل‌محور» برای وایرال کردن محتواهای ضداسلامی و تحریک به قتل‌عام استفاده کردند.

نقد حقوقی: متا اعتراف کرد که در نظارت بر محتوا به زبان بومی (برمه‌ای) کوتاهی کرده است. این همان «شکاف نظارتی» است که در دی ۱۴۰۴ شاهد بودیم؛ جایی که پلتفرم‌ها با ادعای عدم توانایی در تشخیص محتوای فارسی یا «سیاسی» بودن آن، از حذف فراخوان‌های تخریب فیزیکی سرباز زدند.

۲. پرونده اتیوپی (جنگ تیگرای): «سودآوری از خون»

در سال ۲۰۲۲، شکایتی علیه متا در دادگاه عالی کنیا به ثبت رسید که این شرکت را متهم می‌کرد با ترویج پست‌های خشونت‌آمیز، به جنگ داخلی اتیوپی دامن زده است.

مصداق عینی: در اتیوپی، فراخوان‌هایی برای قتل اساتید دانشگاه و چهره‌های مذهبی منتشر شد که با وجود گزارش ریپورت توسط کاربران، ساعت‌ها حذف نشدند.

ارتباط با ماده ۲۳۰ *CDA*: شاکیان اتیوپیایی استدلال کردند که مصونیت ماده ۲۳۰ نباید شامل پلتفرمی شود که با «تقویت الگوریتمیک»، عامدانه محتوای خطرناک را به کاربران بیشتری نشان می‌دهد. این دقیقاً همان نقدی است که به عملکرد پلتفرم‌ها در ایران وارد شده است: پلتفرم صرفاً یک «میزبان» نیست، بلکه یک «توزیع‌کننده فعال» خشونت است.

۳. تحلیل تطبیقی و استنتاج پژوهشی

با قرار دادن وقایع دی ۱۴۰۴ در کنار این دو پرونده، (میانمار / اتیوپی | وقایع دی ۱۴۰۴ ایران) می‌توان به نتایج زیر رسید:

جدول ۱. شاخص‌های مقایسه تطبیقی

شاخص مقایسه	پرونده میانمار و اتیوپی	وقایع دی ۱۴۰۴ ایران
نوع محتوا	تحریک به قتل و پاکسازی قومی	فراخوان حمله به مراکز انتظامی و مساجد
واکنش پلتفرم	تأخیر معنادار در حذف (به بهانه نقص فنی/زبان)	عدم واکنش (به بهانه آزادی بیان سیاسی)
نقض قوانین	دستورالعمل‌های حقوق بشر یونسکو	ماده ۲۳۰ *CDA* و دستورالعمل یونسکو
نتیجه عملیاتی	کشتار جمعی و آوارگی میلیون‌ها نفر	تخریب اموال عمومی و ایجاد رعب و وحشت

همان‌طور که در گزارش تحلیلی ۱۸ دی ۱۴۰۴ تأکید شده، پلتفرم‌های بین‌المللی از یک «استاندارد دوگانه» پیروی می‌کنند. در حالی که در کشورهای غربی به سرعت تحت فشار قوانین محلی (مانند قانون خدمات دیجیتال اتحادیه اروپا – *DSA*) عمل می‌کنند، در کشورهای در حال توسعه یا کشورهایی که با آن‌ها زاویه سیاسی دارند، از خلأهای قانونی مانند ماده ۲۳۰ *CDA* برای سلب مسئولیت از خود استفاده می‌کنند.

این پرونده‌های مشابه ثابت می‌کنند که «اراده انسانی برای تخریب» تنها زمانی به فاجعه منجر می‌شود که پلتفرم‌ها با هدف کسب سود بیشتر از «تنش و درگیری»، چشم خود را بر نقض آشکار حقوق بنیادین بشر و امنیت ملی کشورها می‌بندند.

چارچوب نظری و مبانی حقوقی منع تحریک به خشونت در حقوق بین‌الملل

این بخش به واکاوی مفاهیم بنیادین و اسناد بین‌المللی می‌پردازد که مرز میان آزادی‌های فردی و الزامات امنیت عمومی را ترسیم می‌کنند.

واکاوی مفهوم «تحریک به خشونت» در اسناد بین‌المللی

در حقوق بین‌الملل، آزادی بیان حقی مطلق نیست. ماده ۱۹ میثاق بین‌المللی حقوق مدنی و سیاسی (*ICCPR*)، ضمن شناسایی این حق، محدودیت‌هایی را برای صیانت از نظم و امنیت عمومی پیش‌بینی کرده است. اما ماده ۲۰ همین میثاق، فراتر رفته و دولت‌ها را «مکلف» به ممنوعیت قانونی هرگونه تبلیغ برای جنگ و ترویج نفرت ملی، نژادی یا مذهبی می‌کند که محرک خشونت باشد (United Nations, 1966).

نقطه عطف تفسیر این ماده، «طرح اقدام رباط» (۲۰۱۳) است. این سند برای تفکیک اعتراض مدنی از تحریک مجرمانه، یک آزمون شش‌مرحله‌ای ارائه می‌دهد:

۱. بافت (*Context*): بررسی اینکه آیا محتوا در شرایط بحرانی و ملتهب (مانند دی ۱۴۰۴) منتشر شده است یا خیر.

۲. گوینده (*Speaker*): جایگاه و میزان نفوذ فرد منتشرکننده محتوا.

۳. قصد (*Intent*): وجود اراده برای برانگیختن مخاطبان به اقدامات خشونت‌آمیز.

۴. محتوا (*Content*): میزان صراحت و تندی کلام در دعوت به تخریب یا ترور.

۵. دامنه (*Extent*): میزان فراگیری و دسترسی مخاطبان به آن مطلب.

۶. احتمال وقوع (*Likelihood*): وجود رابطه علیت میان انتشار محتوا و وقوع عینی خشونت (Office of the United Nations High Commissioner for Human Rights, 2013).

تطبیق این معیارها بر محتوای تولیدشده در پلتفرم‌های فراملی طی وقایع دی ۱۴۰۴، کلید درک مسئولیت حقوقی این پلتفرم‌هاست.

تعهد به «دقت مقتضی» در فضای مجازی

در حقوق بین‌الملل کلاسیک، تعهدات عمدتاً متوجه دولت‌ها بود؛ اما با ظهور پلتفرم‌های فراملی، مفهوم «مسئولیت اجتماعی و حقوقی شرکت‌ها» (*CSR*) اهمیت یافت. بر اساس «اصول راهنمای سازمان ملل درباره کسب‌وکار و حقوق بشر» (اصول راگی)، شرکت‌های فناوری موظفند از ایجاد یا مشارکت در پیامدهای سوء حقوق بشری خودداری کنند (United Nations Human Rights Council, 2011).

تعهد به «دقت مقتضی» در فضای مجازی به این معناست که پلتفرم‌ها (مانند اینستاگرام و تلگرام) باید:

ریسک‌های ناشی از الگوریتم‌های خود در گسترش محتوای خشن را پیش‌بینی کنند.

سازوکارهای نظارتی مؤثری برای حذف سریع محتوای تروریستی و محرک خشونت ایجاد نمایند.

در زمان بحران‌های ملی، با دولت‌های ذی‌نفع برای جلوگیری از نقض حق بر حیات شهروندان همکاری کنند (Kaye, 2019).

به‌عنوان نویسنده این پژوهش، بر این باورم که شرکت‌های فناوری موظفند از ایجاد یا مشارکت در پیامدهای سوء حقوق بشری خودداری کنند، و این مطلب صرفاً یک توصیه اخلاقی نیست، بلکه یک تکلیف حقوقی بنیادین در عرصه بین‌المللی است. در واکاوی وقایع دی‌ماه ۱۴۰۴ ایران، آنچه بیش از هر چیز به چشم می‌خورد، اعمال «استاندارد دوگانه» توسط پلتفرم‌های فراملی در مدیریت بحران است.

این شرکت‌ها در حالی که در جوامع توسعه‌یافته با استناد به «امنیت عمومی»، کوچک‌ترین جرقه‌های تحریض به خشونت را مهار می‌کنند، در مواجهه با ناآرامی‌های ایران با نوعی انفعالِ سازمان‌یافته، به بستری برای ترویج تخریب و نقض حقوق شهروندی بدل شدند. این رفتار، مصداق بارز نقض اصل «مراقبت مقتضی» (دای دای لیجینسی) در اسناد حقوق بشری ملل متحد است.

مسئولیت حقوقی این گردانندگان زمانی محرز می‌شود که آن‌ها با وجود آگاهی از ظرفیت پلتفرم خود در گسترش آشوب، از اجرای

سازوکارهای کنترلی امتناع ورزیده‌اند. پلتفرم‌های فراملی نباید با سوءاستفاده از وضعیت «فراسرزمینی» خود، از پاسخگویی در قبال نقض حق بر امنیت و حیات شهروندان بگریزند. مشارکت در پیامدهای سوء حقوق بشری، چه از طریق الگوریتم‌های محرک و چه از طریق سکوت در برابر خشونت، مستلزم مسئولیت کیفری و مدنی بین‌المللی است که نباید زیر سایه ملاحظات سیاسی نادیده گرفته شود.

چالش صلاحیت و حاکمیت دیجیتال

یکی از پیچیدگی‌های حقوقی، تداخل صلاحیت‌هاست. پلتفرم‌های فراملی غالباً در قلمرو ایالات متحده مستقر هستند و از قوانین داخلی آن کشور (مانند ماده ۲۳۰ قانون اخلاق ارتباطات) تبعیت می‌کنند که مصونیت گسترده‌ای به آن‌ها در برابر محتوای کاربران می‌دهد. اما از منظر حقوق بین‌الملل، این مصونیت نمی‌تواند توجیه‌گر نقض تعهدات بنیادین مانند منع تحریک به تروریسم باشد.

دولت‌ها بر اساس اصل «صلاحیت سرزمینی»، حق دارند فعالیت‌های دیجیتالی که آثار مخرب عینی در قلمرو آن‌ها دارد را تنظیم‌گری کنند. وقایع دی ۱۴۰۴ تعارض میان «حاکمیت دیجیتال دولت‌ها» و «یک‌جانبه‌گرایی پلتفرم‌های فراملی» را به شکلی بی‌سابقه آشکار کرد (Kadkhodaei, 2022).

تحول در پارادایم مسئولیت پلتفرم‌ها؛ از «بی‌طرفی فنی» تا «مدیریت ریسک سیستماتیک»

در ادبیات حقوقی سال‌های ۲۰۲۲ و ۲۰۲۳، رویکرد کلاسیک که پلتفرم‌ها را صرفاً «لوله‌های انتقال داده» «دامپ پیپز» می‌دانست، به طور کامل متروک شده است. پژوهش‌های نوین، مانند مطالعات ۲۰۲۲ و ۲۰۲۳، تأکید دارند که پلتفرم‌های فراملی به دلیل استفاده از الگوریتم‌های «بیشینه‌سازی تعامل»، از حالت واسطه بی‌طرف خارج شده و به «تولیدکنندگان محیطی» تبدیل شده‌اند.

نظریه «تقصیر الگوریتمی» و بازتعریف عنصر مادی تحریک

نوشتارهای حقوقی سال ۲۰۲۳ مفهوم جدیدی تحت عنوان «تقصیر در طراحی» را وارد حوزه مسئولیت پلتفرم‌ها کرده‌اند. بر اساس این نظریه، وقتی الگوریتم یک پلتفرم در جریان وقایعی مانند دی‌ماه ۱۴۰۴، محتوای خشونت‌آمیز را به دلیل «نرخ کلیک بالا» به کاربران بیشتری پیشنهاد می‌دهد، پلتفرم دیگر صرفاً میزبان نیست، بلکه در «رکن مادی» تحریک، مشارکت فعال دارد. در واقع، «تحریک الگوریتمی» شکلی نوین از تحریک غیرمستقیم است که در آن اراده انسانیِ طراح در قالب کدهای نرم‌افزاری، پیامدهای مخرب عینی ایجاد می‌کند (Mohammadi & Sotoudeh, 2023).

استانداردهای نوین در «قانون خدمات دیجیتال» (*DSA 2022*) به عنوان عرف در حال ظهور

یکی از مهم‌ترین تحولات، تصویب «قانون خدمات دیجیتال اتحادیه اروپا» (*DSA*) در اواخر سال ۲۰۲۲ است که اکنون به عنوان یک «مدل معیار» در حقوق بین‌الملل فضای مجازی شناخته می‌شود. طبق مواد ۳۴ و ۳۵ این قانون، پلتفرم‌های بسیار بزرگ (*VLOPs*) ملزم به انجام «ارزیابی ریسک سیستماتیک» هستند.

این استاندارد جدید، پلتفرم‌ها را موظف می‌کند پیش‌بینی کنند که چگونه خدمات آن‌ها ممکن است علیه «امنیت عمومی» یا «سلامت روان جامعه» به کار گرفته شود. در وقایع دی ۱۴۰۴، عدم اجرای این ارزیابی‌های پیش‌دستانه توسط پلتفرم‌های فراملی، بر اساس رویه‌های نوین ۲۰۲۳، مصداق بارز «نقض دقت مقتضی» تلقی می‌گردد (European Union, 2022).

حاکمیت شرکتی در برابر حاکمیت ملی؛ چالش «صلاحیت فراسرزمینی»

ادبیات حقوقی ۲۰۲۲ به پدیده‌ای اشاره دارد که آن را «حاکمیت پلتفرمی» می‌نامد. پلتفرم‌ها با وضع قوانین داخلی که فراتر از قوانین ملی دولت‌ها عمل می‌کند، نوعی نظام حقوقی موازی ایجاد

کرده‌اند. اما در سال ۲۰۲۳، گرایش شدیدی در حقوق بین‌الملل بشر (به‌ویژه در گزارش‌های گزارشگر ویژه آزادی بیان سازمان ملل) پدید آمده که تأکید می‌کند «استانداردهای جامعه» پلتفرم‌ها نباید ناقض تعهدات دولت‌ها در قبال حفظ نظم عمومی باشد. لذا، در وقایع دی ۱۴۰۴، استناد پلتفرم‌ها به قوانین داخلی خود برای توجیه عدم حذف محتوای محرک خشونت، از منظر دکترین‌های نوین ۲۰۲۳، فاقد وجاهت قانونی در برابر قواعد آمره بین‌المللی است (Kaye, 2019).

اسناد حقوقی و راهبردی یونسکو در زمینه حکمرانی بر پلتفرم‌ها

مهم‌ترین و جدیدترین سند این سازمان یعنی «دستورالعمل حکمرانی بر پلتفرم‌های دیجیتال» است که در نوامبر ۲۰۲۳ (آبان ۱۴۰۲) منتشر شد.

یونسکو به عنوان نهاد متولی آزادی بیان و دسترسی به اطلاعات در سازمان ملل، تلاش کرده است تا میان «حفاظت از آزادی بیان» و «مقابله با محتوای آسیب‌رسان» تعادلی حقوقی برقرار کند.

در ادامه، محورهای کلیدی این اسناد و استانداردهای یونسکو را ملاحظه می‌کنید:

۱. دستورالعمل حکمرانی بر پلتفرم‌های دیجیتال (نقطه عطف اسناد یونسکو)

این سند محصول یک فرآیند مشورتی جهانی با حضور هزاران کارشناس است و هدف آن ایجاد یک سیستم حکمرانی چندجانبه است. محورهای اصلی آن عبارتند از:

مسئولیت‌پذیری پلتفرم‌ها: پلتفرم‌ها باید در قبال تأثیرات الگوریتم‌های خود بر حقوق بشر پاسخگو باشند. یونسکو تأکید می‌کند که پلتفرم‌ها نباید صرفاً یک «توزیع‌کننده بی‌طرف» تلقی شوند، بلکه مسئولیتِ نحوه انتشار محتوا بر عهده آن‌هاست.

شفافیت الگوریتمیک: پلتفرم‌ها موظفند درباره نحوه رتبه‌بندی، توصیه و حذف محتوا شفاف‌سازی کنند. این دقیقاً با چالش‌های مطرح‌شده در مورد «ماده ۲۳۰ *CDA*» و قدرت نامحدود پلتفرم‌ها در تعدیل محتوا همخوانی دارد.

ارزیابی ریسک حقوق بشری: پلتفرم‌های بزرگ موظفند پیش از اعمال تغییرات در سیاست‌های خود یا در زمان بحران‌ها، تأثیر اقداماتشان را بر صلح و حقوق بشر ارزیابی کنند.

۲. توصیه در مورد اخلاق هوش مصنوعی (۲۰۲۱)

از آنجا که اکثر پلتفرم‌ها از هوش مصنوعی برای مدیریت محتوا استفاده می‌کنند، این سندِ یونسکو بر موارد زیر تأکید دارد:

حاکمیت قانون در فضای مجازی: پلتفرم‌ها نباید قوانین داخلی خود (*Terms of Service*) را فراتر از قوانین بین‌المللی حقوق بشر قرار دهند.

منع سوگیری: الگوریتم‌ها نباید به نفع یا علیه یک گروه سیاسی، مذهبی یا ملیتی خاص (بیزاری‌جویی یا تبعیض) عمل کنند.

۳. استانداردهای یونسکو در مواجهه با «محتوای غیرقانونی و آسیب‌رسان»

یونسکو بین «اطلاعات نادرست» (*Misinformation*)، «اطلاعات مخرب» (*Disinformation*) و «نفرت‌پراکنی» (*Hate Speech*) تفکیک قائل می‌شود:

در زمان بحران و ناآرامی‌ها (مانند وقایع دی ۱۴۰۴): طبق اسناد یونسکو، دولت‌ها حق دارند از پلتفرم‌ها بخواهند محتوایی را که مستقیماً به «خشونت قریب‌الوقوع» تحریک می‌کند، محدود کنند. اما این محدودیت باید طبق سه اصل باشد: ۱. قانونی بودن ۲. ضرورت ۳. تناسب.

دخالت در انتخابات و حاکمیت: یونسکو بر این باور است که پلتفرم‌ها نباید اجازه دهند از ابزارهای آن‌ها برای دخالت سازمان‌یافته در امور داخلی کشورها یا انتشار گسترده اخبار جعلی که منجر به بی‌ثباتی می‌شود، استفاده شود.

۴. چالش «تعدیل محتوا» از منظر یونسکو

یونسکو، برخلاف دیدگاه‌های افراطی که خواهان حذف کامل محتوا یا آزادی مطلق هستند، مدل «تعدیل محتوای مبتنی بر حقوق بشر» را پیشنهاد می‌دهد:

پلتفرم‌ها باید دارای فرآیند «استیناف» باشند؛ یعنی اگر محتوای کاربری به اشتباه حذف شد، او باید بتواند اعتراض کند.

پلتفرم‌ها باید در کشورهای مختلف، تیم‌های محلی آشنا به زبان و فرهنگ داشته باشند تا از «سانسور اشتباه» یا «رها کردن محتوای خشونت‌آمیز» به دلیل عدم درک زبان جلوگیری شود.

اگر بخواهیم وقایعی مانند دی ۱۴۰۴ یا رفتارهای پلتفرم‌ها را با اسناد یونسکو بسنجیم، نتیجه این است:

یونسکو از یک سو «قطع کامل اینترنت» توسط دولت‌ها را نفی می‌کند (به دلیل نقض حق دسترسی به اطلاعات)، اما از سوی دیگر، «ولنگاری پلتفرم‌ها در عدم مدیریت محتوای خشونت‌آمیز و مخرب» را نیز نمی‌پذیرد. از نظر یونسکو، حکمرانی موفق یعنی پلتفرم‌ها تحت نظارت نهادهای مستقل ملی و بین‌المللی باشند تا نتوانند با الگوریتم‌های خود به بحران‌ها دامن بزنند.

این اسناد راه را برای کشورها باز می‌گذارد تا پلتفرم‌های خارجی را به داشتن «نمایندگی پاسخگو» و «شفافیت در الگوریتم» ملزم کنند.

اما برخلاف همه قوانین بین‌المللی، ما در وقایع دی‌ماه ۱۴۰۴ ایران رفتارهای دوگانه این پلتفرم‌ها را دیدیم و کاملاً بدیهی است که کوتاهی شرکت‌های متصدی می‌تواند تحت شرایطی به دولت‌های میزبان (مانند آمریکا) منتسب شود.

نمونه‌هایی از تحریک مردم در دی ۱۴۰۴ توسط پلتفرم‌های فراملی برای ایجاد اغتشاش

تحلیل وقایع دی ۱۴۰۴ نشان می‌دهد که پلتفرم‌های فراملی از حالت «بستر ارتباطی» خارج شده و به عنوان «کنشگران میدانی» عمل کردند. بر اساس گزارش‌های فنی و تحلیل‌های محتوایی، نمونه‌های مشخصی از این تحریکات که منجر به تشدید فضای امنیتی و اغتشاشات شد، به شرح زیر است:

۱. استفاده از «تبلیغات هدفمند» برای فراخوان

یکی از موارد بی‌سابقه در این دوره، استفاده از سیستم تبلیغاتی پلتفرم‌هایی مانند تلگرام و گوگل بود. در حالی که این پلتفرم‌ها مدعی بی‌طرفی هستند، اما:

نمونه: تبلیغات رسمی تلگرام در کانال‌های پرمخاطب ایرانی، پیام‌هایی حاوی زمان و مکان تجمعات را نمایش می‌داد. این اقدام عملاً پلتفرم را از یک واسطه به یک «بسیج‌کننده سیاسی» تبدیل کرد که برخلاف استانداردهای یونسکو در مورد عدم دخالت در امور داخلی کشورهاست.

۲. آموزش مستقیم اقدامات پارتیزانی و تخریبی

بسیاری از کانال‌های تحت حمایت در تلگرام و صفحاتی در اینستاگرام، به جای اطلاع‌رسانی، به «آموزشکده خشونت» تبدیل شدند:

نمونه: انتشار ویدئوهای گرافیکی و ساده برای ساخت «کوکتل مولوتوف»، نحوه از کار انداختن دوربین‌های مداربسته شهری، و شیوه‌های درگیری فیزیکی با نیروهای حافظ امنیت.

نقض قوانین خود: علی‌رغم اینکه قوانین داخلی این پلتفرم‌ها ترویج خشونت را ممنوع کرده است، اما گزارش‌ها نشان می‌دهد که گزارش‌های مردمی برای بستن این کانال‌های خشونت‌طلب با بی‌توجهی پلتفرم‌ها مواجه می‌شد.

۳. الگوریتم‌های «بیش‌نمایی» و اخبار جعلی

الگوریتم‌های اینستاگرام و ایکس (توییتر سابق) به گونه‌ای تنظیم شده بودند که محتوای دارای بارِ خشونت و هیجان زیاد را در صدر فید کاربران قرار می‌دادند:

نمونه (کشته‌سازی): انتشار تصاویر قدیمی یا تصاویر مربوط به کشورهای دیگر و انتساب آن‌ها به وقایع دی ۱۴۰۴. الگوریتم‌ها با

ترند کردن این اخبار جعلی (بدون بررسی درستی‌آزمایی)، باعث ایجاد «خشم آنی» در جامعه می‌شدند.

ربات‌های ترویجی: فعال شدن هزاران حساب کاربری رباتیک در ایکس که با هشتگ‌های هماهنگ، فضای کشور را «جنگی» و «در حال فروپاشی» نشان می‌دادند تا مردم مردد را به خیابان‌ها بکشانند.

۴. تغییر در سیاست‌های نظارتی

در جریان وقایع دی ۱۴۰۴، برخی پلتفرم‌ها به طور موقت سیاست‌های منع خشونت خود را برای کاربران ایرانی تغییر دادند یا نادیده گرفتند:

نمونه: اجازه انتشار پیام‌هایی که به طور صریح خواستار ترور مقامات یا حمله به مراکز حساس دولتی و نظامی بودند. این در حالی است که مشابه چنین پیام‌هایی در مورد کشورهای غربی بلافاصله منجر به مسدودسازی حساب کاربری می‌شد.

۵. تحریک از طریق «اینفلوئنسرهای اجاره‌ای»

پلتفرم‌ها به عنوان بستر اصلی، به چهره‌های مشهور (سلبریتی‌ها و اینفلوئنسرها) که خارج از ایران حضور داشتند، ضریب نفوذ غیرعادی دادند:

نمونه: هدایت عملیاتی خیابان‌ها توسط افرادی که هیچ شناختی از وضعیت میدانی نداشتند، اما از طریق «لایوهای اینستاگرامی» به صورت لحظه‌ای دستوراتی صادر می‌کردند که مستقیماً امنیت جانی شهروندان را به خطر می‌انداخت.

۶. تلاش برای ایجاد «اینترنت موازی» و غیرقانونی

در اوج ناآرامی‌ها، تبلیغ گسترده برای استفاده از تجهیزات غیرقانونی مانند استارلینک در پلتفرم‌های خارجی انجام شد. این اقدام از نظر حقوقی، تلاشی برای دور زدن حاکمیت ملی و ایجاد یک زیرساخت ارتباطی خارج از کنترل دولت برای تداوم هماهنگی‌های اغتشاشات تلقی شد.

این نمونه‌ها نشان می‌دهد که پلتفرم‌های فراملی در دی ۱۴۰۴، فراتر از ماده ۲۳۰ *CDA* (که آن‌ها را از مسئولیت محتوای کاربران تبرئه می‌کند) عمل کردند. طبق اسناد یونسکو، زمانی که یک پلتفرم با الگوریتم‌های خود به خشونت دامن می‌زند یا فرآیند نظارت بر محتوا را به نفع یک جریان سیاسی خاص به حالت تعلیق درمی‌آورد، مسئولیت مستقیم «نقض صلح و امنیت اجتماعی» بر عهده مدیریت آن پلتفرم است.

مصادیقی از عملکرد پلتفرم‌های فراملی در جریان ناآرامی‌های دی ۱۴۰۴ ایران

ترویج آموزش سلاح‌های نامتعارف و تخریبی

در جریان ناآرامی‌های دی ۱۴۰۴، گزارش‌های فنی از پلتفرم تروریستی مربوطه و برخی صفحات اینستاگرامی نشان داد که الگوریتم‌ها نه تنها مانع انتشار محتواهای خطرناک نشدند، بلکه ویدئوهای آموزشی تحت عناوین زیر به سرعت وایرال شدند:

۱- آموزش ساخت کوکتل مولوتوف ترکیبی: در این گزارش‌ها به انتشار ویدئوهایی اشاره شده که در آن‌ها روش افزودن موادی مانند «صابون رنده‌شده» یا «کائوچو» به بنزین برای افزایش چسبندگی و قدرت سوزانندگی مایع در کوکتل مولوتوف آموزش داده می‌شد. این محتوا مستقیماً با هدف وارد آوردن آسیب شدید به نیروهای حافظ امنیت و اماکن عمومی تولید شده بود.

۲- تله‌های انفجاری و میخ‌های جاده‌ای: رصدها نشان‌دهنده انتشار نقشه‌های ساخت «سه‌پایه‌های آهنی نوک‌تیز» و تله‌های انفجاری ساده با استفاده از کپسول‌های گاز و مواد مشتعله خانگی بود که برای از کار انداختن خودروهای امدادی و انتظامی در مسیرها تعبیه می‌شدند.

مصادیق «ترک فعل» ساختاری در مدیریت محتوا

بر اساس گزارش‌های نهادهای نظارتی در دی ۱۴۰۴، موارد زیر به عنوان سند قصور مدیران پلتفرم قابل استناد است:

۱- نادیده گرفتن گزارش‌های مردمی: علیرغم اینکه هزاران کاربر محتویات مربوط به آموزش ساخت بمب‌های دست‌ساز را به عنوان «محتوای تروریستی و خطرناک» گزارش کرده بودند،

سیستم‌های نظارتی پلتفرم‌های فراملی از حذف آن‌ها خودداری کردند. این عمل در حقوق بین‌الملل به معنای (علم فعلی) پلتفرم به وجود محتوای مجرمانه و امتناع از اقدام تلقی می‌شود.

۲- ارائه نقشه‌های عملیاتی: برخی پلتفرم‌ها امکان تعیین نقاط تجمع و اشتراک‌گذاری لحظه‌ای موقعیت نیروهای انتظامی را فراهم آوردند که در کنار آموزش‌های ساخت سلاح، محیط فضای مجازی را به یک «اطاق جنگ» عملیاتی تبدیل کرد.

استناد حقوقی در مقاله برای این بخش

بر اساس گزارش‌های مستند رصد فضای مجازی در وقایع دی ۱۴۰۴، پلتفرم‌های فراملی با میزبانی و شتاب‌دهی به محتواهای آموزشی نظیر «چگونگی ساخت کوکتل مولوتوف با ماندگاری حریق بالا» و «تعبیه تله‌های انفجاری خانگی»، از نقش واسطه بی‌طرف خارج شدند. امتناع این پلتفرم‌ها از حذف ویدئوهای آموزشی که به وضوح نقض‌کننده قوانین ضدتروریسم خود بود، مصداق بارز ترک فعل و تسبیب در ایجاد خسارات جانی و مالی است.» (مرکز ملی فضای مجازی، دی ۱۴۰۴.)

مسئولیت بین‌المللی کنشگران در قبال تحریک به خشونت؛ از الگوریتم تا اراده انسانی

در این بخش، مسئولیت پلتفرم‌های فراملی و دولت‌های متبوع آن‌ها نه به عنوان ماشین‌های بی‌طرف، بلکه به عنوان کنشگرانی با «اراده و سوگیری انسانی» تحلیل می‌شود.

۱. تعهد به «دقت مقتضی» و نقش اراده انسانی در پالایش محتوا

اگرچه پلتفرم‌ها ادعا می‌کنند که مدیریت محتوا توسط هوش مصنوعی انجام می‌شود، اما «عنصر انسانی» در دو سطح نقشی حیاتی دارد: نخست در «طراحی الگوریتم‌ها» و دوم در «سیاست‌گذاری نظارتی».

بر اساس اصول راهنمای سازمان ملل درباره کسب‌وکار و حقوق بشر، شرکت‌های فراملی مکلف به اعمال دقت مقتضی هستند. در وقایع دی ۱۴۰۴، مشاهده شد که پلتفرم‌ها علیرغم دریافت گزارش‌های متعدد از سوی کاربران و نهادهای قانونی مبنی بر وجود محتوای محرک خشونت (مانند آموزش ساخت تله‌های انفجاری و تعیین مختصات برای حملات)، از حذف آن‌ها خودداری کردند. این «عدم اقدام» یک تصمیم فنی نیست، بلکه برآمده از یک اراده انسانی و مدیریتی است که میان بحران‌های مختلف (مثلاً بحران‌های اروپا در برابر بحران‌های خاورمیانه) تمایز قائل می‌شود (Kaye, 2019). این استاندارد دوگانه، نقض مستقیم تعهدات بین‌المللی حقوق بشر در زمینه برابری و عدم تبعیض است.

۲. عاملیت انسانی در «تحریک غیرمستقیم» و مهندسی افکار عمومی

عنصر انسانی در اینجا به معنای نیت و قصد سازمان‌دهندگان پشت پرده است. در حقوق بین‌الملل کیفری و بر اساس رویه دادگاه‌های بین‌المللی (مانند دادگاه روآندا در پرونده رسانه هزار تپه)، تحریک لازم نیست حتماً صریح باشد؛ بلکه تحریک «ضمنی» و «غیرمستقیم» نیز اگر در شرایط خاصی صورت گیرد که منجر به خشونت شود، واجد مسئولیت است (Ghari Seyed Fatemi, 2019).

در وقایع دی ۱۴۰۴، مدیران پلتفرم‌ها با اجازه دادن به فعالیت حساب‌های کاربری که به طور سیستماتیک به «نفرت‌پراکنی» و «انسان‌زدایی» از طیف خاصی از شهروندان می‌پرداختند، در واقع «بستر انسانی» لازم برای وقوع خشونت فیزیکی را فراهم کردند. اینجاست که «اراده انسانیِ» ناظران پلتفرم در «نادیده گرفتن عمدی» (*Willful Blindness*) محتواهای محرک، مسئولیت حقوقی آن‌ها را از حالت مدنی به حوزه‌ی همدستی در نقض حقوق بشر نزدیک می‌کند.

۳. انتساب رفتار پلتفرم‌ها به دولت‌های میزبان (مسئولیت بین‌المللی دولت)

یکی از چالش‌برانگیزترین مباحث حقوقی، انتساب رفتار این شرکت‌های خصوصی به دولت متبوع (مثلاً ایالات متحده) است. طبق مواد مسئولیت بین‌المللی دولت (*ARSIWA 2001*):

ماده ۸ (کنترل و هدایت): اگر ثابت شود دولت میزبان از این پلتفرم‌ها به عنوان ابزاری برای پیشبرد اهداف سیاست خارجی خود در جهت ناامن‌سازی استفاده کرده است، رفتار پلتفرم به دولت منتسب می‌شود.

تعهد به پیشگیری: حتی اگر انتساب مستقیم طبق ماده ۸ دشوار باشد، دولت‌ها بر اساس حقوق بین‌الملل عرفی متعهدند اجازه ندهند از قلمرو یا امکانات فنی تحت صلاحیت‌شان برای آسیب رساندن به حقوق سایر دولت‌ها استفاده شود (اصول حاکم بر قضیه کانال کورفو).

قصور انسانیِ مقامات نظارتی در دولت‌های میزبان برای ملزم کردن این شرکت‌ها به رعایت قوانین بین‌المللی در جریان وقایع دی ۱۴۰۴، مصداق «ترک فعل» است که مسئولیت بین‌المللی دولت را به همراه دارد (Schmitt, 2017).

۴. تحلیل رفتار پلتفرم‌ها در پرتو «آزمون شش‌مرحله‌ای رباط» با تأکید بر عنصر انسانی

در تحلیل مسئولیت پلتفرم‌های فراملی، ادبیات حقوقی ایران در سال‌های اخیر (۱۴۰۱ تا ۱۴۰۳) از رویکردهای سنتی فاصله گرفته و به سمت تبیین «مسئولیت مبتنی بر مدیریت خطر» و «تسبیب هوشمند» حرکت کرده است. در این میان، انطباق این مفاهیم با قواعد فقهی، بستری بومی برای پاسخگو کردن این کنشگران فراهم آورده است.

۵. بازخوانی قاعده «تسبیب» در فضای سایبر و نقش اراده انسانی

پژوهش‌های انجام‌شده در سال‌های ۱۴۰۱ و ۱۴۰۲ تأکید دارند که در وقایع بحرانی نظیر دی ۱۴۰۴، پلتفرم‌ها از نقش «مجرای محض» خارج شده و در مقام «مسبب» قرار می‌گیرند. در فقه اسلامی، «تسبیب» زمانی منجر به ضمان می‌شود که مسبب، زمینه‌ساز وقوع فعل توسط فاعل مباشر گردد، به‌گونه‌ای که عرفاً فعل به او استناد یابد (Tofighi, 2023).

نکته کلیدی در تحلیل‌های ۲۰۲۳ این است که الگوریتم‌های پلتفرم‌ها، «اراده‌ی برنامه‌ریزی‌شده» مدیران آن‌ها هستند. لذا وقتی پلتفرم با آگاهی از شرایط ملتهب جامعه (عنصر آگاهی انسانی)، الگوریتم را برای بازنشر بیشتر محتوای خشن تنظیم می‌کند، مصداق «تسبیب اقوی از مباشر» پیدا می‌کند. در اینجا، مباشر (کاربر خشونت‌طلب) تنها ابزاری در دستِ مهندسی اجتماعی پلتفرم است. پژوهش‌های نوین داخلی استدلال می‌کنند که «قصدِ نتیجه» در مدیران پلتفرم، از طریق «علم به پیامد» (معادل نیت در فقه) قابل اثبات است (Kadkhodaei, 2022).

۶. قاعده «اعانت بر اثم» و ترک فعل آگاهانه پلتفرم‌ها

یکی از مبانی استوار در منابع اسلامی که در مقالات پژوهشی سال ۱۴۰۲ (به‌ویژه در حوزه حقوق جزا و جرم‌شناسی) مورد تأکید قرار گرفته، قاعده «حرمت اعانت بر اثم و عدوان» است. طبق این قاعده، فراهم آوردن مقدمات معصیت و جرم، به ویژه زمانی که مقدمه منحصراً در اختیار یک نهاد باشد، موجب مسئولیت است.

در تحلیل مسئولیت پلتفرم‌های فراملی، ادبیات حقوقی ایران در سال‌های اخیر (۱۴۰۱ تا ۱۴۰۳) از رویکردهای سنتی فاصله گرفته و به سمت تبیین «مسئولیت مبتنی بر مدیریت خطر» و «تسبیب هوشمند» حرکت کرده است. در این میان، انطباق این مفاهیم با قواعد فقهی، بستری بومی برای پاسخگو کردن این کنشگران فراهم آورده است.

بازخوانی قاعده «تسبیب» در فضای سایبر و نقش اراده انسانی

پژوهش‌های انجام‌شده در سال‌های ۱۴۰۱ و ۱۴۰۲ تأکید دارند که در وقایع بحرانی نظیر دی ۱۴۰۴، پلتفرم‌ها از نقش «مجرای محض» خارج شده و در مقام «مسبب» قرار می‌گیرند. در فقه اسلامی، «تسبیب» زمانی منجر به ضمان می‌شود که مسبب، زمینه‌ساز وقوع

فعل توسط فاعل مباشر گردد، به‌گونه‌ای که عرفاً فعل به او استناد یابد (Tofighi, 2023).

نکته کلیدی در تحلیل‌های ۲۰۲۳ این است که الگوریتم‌های پلتفرم‌ها، «اراده‌ی برنامه‌ریزی‌شده» مدیران آن‌ها هستند. لذا وقتی پلتفرم با آگاهی از شرایط ملتهب جامعه (عنصر آگاهی انسانی)، الگوریتم را برای بازنشر بیشتر محتوای خشن تنظیم می‌کند، مصداق «تسبیب اقوی از مباشر» پیدا می‌کند. در اینجا، مباشر (کاربر خشونت‌طلب) تنها ابزاری در دستِ مهندسی اجتماعی پلتفرم است. پژوهش‌های نوین داخلی استدلال می‌کنند که «قصدِ نتیجه» در مدیران پلتفرم، از طریق «علم به پیامد» (معادل نیت در فقه) قابل اثبات است (Kadkhodaei, 2022).

قاعده «اعانت بر اثم» و ترک فعل آگاهانه پلتفرم‌ها

یکی از مبانی استوار در منابع اسلامی که در مقالات پژوهشی سال ۱۴۰۲ (به‌ویژه در حوزه حقوق جزا و جرم‌شناسی) مورد تأکید قرار گرفته، قاعده «حرمت اعانت بر اثم و عدوان» است. طبق این قاعده، فراهم آوردن مقدمات معصیت و جرم، به ویژه زمانی که مقدمه منحصراً در اختیار یک نهاد باشد، موجب مسئولیت است.

در ادامه، تحلیل حقوقی این بخش از مقاله برای اثبات پیوند میان «ترک فعل پلتفرم» و «مسئولیت دولت میزبان» ارائه می‌شود:

انتساب مسئولیت بین‌المللی به دولت میزبان ناشی از کوتاهی پلتفرم‌های فراملی

در حقوق بین‌الملل کلاسیک، رفتار شرکت‌های خصوصی مستقیماً به دولت منتسب نمی‌شود؛ اما در مورد پلتفرم‌های فراملی در وقایع دی ۱۴۰۴، می‌توان از سه مجرای حقوقی، مسئولیت را متوجه دولت میزبان (آمریکا) دانست:

۱. نقض تعهد به «دقت مقتضی (*Due Diligence*)»

این مهم‌ترین مبنای مسئولیت دولت میزبان است. طبق رویه قضایی بین‌المللی (مانند رأی کانال کورفو)، دولت‌ها متعهدند اجازه ندهند از قلمرو آن‌ها برای انجام اعمالی که به حقوق سایر دولت‌ها لطمه می‌زند، استفاده شود.

تحلیل: دولت آمریکا با حفظ «ماده ۲۳۰ قانون اخلاق ارتباطات»، عملاً یک «منطقه امن حقوقی» برای پلتفرم‌ها ایجاد کرده است. این قانون باعث شده دولت از انجام وظیفه نظارتی خود در قبال پلتفرم‌هایی که مرکز سرور و مدیریت آن‌ها در خاک اوست، امتناع کند. این «ترک فعلِ قانون‌گذاری»، نقض مستقیم تعهد به پیشگیری از وقوع جرایم بین‌المللی (مانند تحریک به تروریسم و خشونت) است.

۲. دکترین «هدایت و کنترل» (ماده ۸ طرح *ILC*)

اگر ثابت شود که دولت میزبان از این پلتفرم‌ها به عنوان ابزاری برای پیشبرد سیاست‌های مداخله‌جویانه خود استفاده کرده است، رفتار پلتفرم به دولت منتسب می‌شود.

مصداق در دی ۱۴۰۴: در مواردی که مقامات رسمی دولت میزبان (مانند سخنگویان وزارت خارجه یا کاخ سفید) به طور علنی از پلتفرم‌ها می‌خواهند که از حذف محتواهای خاص (حتی خشونت‌آمیز) خودداری کنند یا زمان‌بندی تعمیرات فنی خود را برای تداوم دسترسی آشوب‌گران تغییر دهند، پلتفرم از یک بازیگر خصوصی به «عامل تحت کنترل» تغییر وضعیت می‌دهد. در این حالت، طبق ماده ۸ طرح مسئولیت دولت‌ها، فعل و ترک فعل پلتفرم، عملِ خودِ دولت محسوب می‌شود.

۳. دکترین «پذیرش و تنفیذ» (ماده ۱۱ طرح *ILC*)

حتی اگر در ابتدا رفتار پلتفرم به دولت منتسب نباشد، چنانچه دولت میزبان پس از اطلاع از اقدامات مجرمانه (مانند آموزش ساخت سلاح)، آن رفتار را تأیید کرده یا با حمایت‌های سیاسی و قانونی از گردانندگان پلتفرم در برابر تعقیب قضایی بین‌المللی دفاع کند، آن رفتار را «از آنِ خود» کرده است.

نقد: حمایت سیاسی مطلق دولت آمریکا از عملکرد پلتفرم‌ها در قبال ایران و امتناع از همکاری قضایی برای استرداد یا مجازات عوامل تحریک، مصداق بارز تنفیذ رفتارهای متخلفانه است.

مسئولیت بین‌المللی دولت میزبان (ایالات متحده) در وقایع دی ۱۴۰۴، ناشی از یک «هم‌افزایی متخلفانه» است. از یک سو، ماده ۲۳۰ *CDA* مانع از نظارت قانونی بر پلتفرم‌ها شده و از سوی دیگر، حمایت‌های سیاسی دولت میزبان، پلتفرم‌ها را به «قلمروهای فراقانونی» تبدیل کرده است. بنابراین، دولت میزبان به دلیل عدم رعایت اصل دقت مقتضی در کنترل فعالیت‌های سایبری که از قلمرو حاکمیتش سرچشمه می‌گیرد و به نظم عمومی ایران آسیب جدی زده، دارای مسئولیت بین‌المللی است و موظف به جبران خسارت و توقف این رویه می‌باشد.»

واژگان کلیدی برای این بخش:

Due Diligence: دقت مقتضی.

Attribution of Conduct: انتساب رفتار.

Safe Harbor Policies: سیاست‌های پناهگاه امن (مانند ماده ۲۳۰).

State Responsibility: مسئولیت دولت.

واکاوی عملکرد پلتفرم‌های فراملی در وقایع دی‌ماه ۱۴۰۴؛ از نقض استانداردهای داخلی تا نادیده انگاشتن قواعد بین‌المللی

این بخش به بررسی عینی رفتار پلتفرم‌های فراملی (به‌طور خاص اینستاگرام، تلگرام و ایکس) در جریان ناآرامی‌های دی‌ماه ۱۴۰۴ ایران و تطبیق آن با موازین حقوق بین‌الملل بشر می‌پردازد.

گونه‌شناسی محتوای محرک خشونت در بستر پلتفرم‌ها (دی ۱۴۰۴)

در جریان وقایع دی‌ماه ۱۴۰۴، محتواهای منتشرشده در پلتفرم‌های فراملی از حالت «اعتراض مدنی» خارج و به حوزه‌ی «تحریک مستقیم به خشونت» وارد شد. بر اساس رصدهای انجام‌شده، این محتواها در سه دسته اصلی قابل بازیابی هستند:

۱. آموزش‌های عملیاتی تخریب: انتشار گسترده ویدئوهای آموزشی ساخت «کوکتل مولوتوف»، «تله‌های انفجاری دست‌ساز» و روش‌های از کار انداختن خودروهای امدادی و انتظامی.

۲. تعیین اهداف (*Targeting*): افشای اطلاعات شخصی و انتشار آدرس منازل و محل کار نیروهای حافظ امنیت، مسئولان محلی و حتی شهروندان عادی مخالف با اغتشاشات، با هدف ترور یا ضرب‌وشتم.

۳. تخریب زیرساخت‌های حیاتی: فراخوان‌های سازمان‌یافته برای حمله به مراکز حساس نظیر تصفیه‌خانه‌ها، ایستگاه‌های برق و انبارهای کالاهای اساسی که مستقیماً با «حق بر حیات» و «حق بر رفاه» کل جمعیت در ارتباط است (مرکز مطالعات فضای مجازی، ۱۴۰۴).

تحلیل عملکرد پلتفرم‌ها در پرتو «آزمون شش‌مرحله‌ای رباط»

برای ارزیابی حقوقی، باید وقایع دی ۱۴۰۴ را با معیارهای «طرح اقدام رباط» تطبیق داد:

بافت (*Context*): محتواها در زمانی منتشر شد که جامعه در وضعیت التهاب شدید قرار داشت؛ لذا احتمال تبدیل «کلام» به «عمل خشونت‌آمیز» بسیار بالا بود.

قصد (*Intent*): بسیاری از حساب‌های کاربری با هویت‌های جعلی یا هدایت‌شده از خارج از کشور، تعمدی روشن در برانگیختن توده‌ها به اعمال مسلحانه داشتند.

احتمال وقوع (*Likelihood*): رابطه علیت مستقیم میان فراخوان‌های منتشرشده در تلگرام و اینستاگرام با آتش‌سوزی‌های زنجیره‌ای در مراکز دولتی و مذهبی در دی ۱۴۰۴ به اثبات رسیده است.

بر این اساس، طبق ماده ۲۰ میثاق، پلتفرم‌ها موظف به مداخله فوری و حذف این محتواها بودند، اما رویه عملی آن‌ها نشان‌دهنده «تسامح عمدی» بود (Kadkhodaei, 2022).

استانداردهای دوگانه و نقض اصل «بی‌طرفی حقوقی»

یکی از جدی‌ترین نقدهای حقوقی به عملکرد پلتفرم‌هایی مانند «ایکس» و «متا» در وقایع دی ۱۴۰۴، اعمال استانداردهای دوگانه است.

*مقایسه تطبیقی: در حالی که این پلتفرم‌ها در وقایعی مانند حمله به ساختمان کنگره آمریکا (۶ ژانویه) یا شورش‌های فرانسه، به سرعت حساب‌های محرک خشونت را مسدود و دسترسی‌ها را محدود کردند، در وقایع دی ۱۴۰۴ ایران، نه تنها از مسدودسازی حساب‌های مروج تروریسم خودداری کردند، بلکه بر اساس گزارش‌های فنی، الگوریتم‌های «پیشنهاددهنده» به‌گونه‌ای عمل کردند که محتوای خشونت‌آمیز و دوقطبی‌ساز، ضریب انتشار بیشتری پیدا کند (Ghaemi, 2024).

این رفتار، نقض صریح تعهد به «دقت مقتضی» و نشان‌دهنده دخالت عنصر اراده انسانی در جهت‌دهی به بحران‌های امنیتی است.

مسئولیت پلتفرم‌ها در قبال «تروریسم سایبری» و امنیت انسانی

از منظر حقوق بین‌الملل بشر، امنیت تنها به معنای امنیت دولت نیست، بلکه «امنیت انسانی» اولویت دارد. وقتی پلتفرمی اجازه می‌دهد از زیرساخت‌هایش برای سازماندهی حمله به بیمارستان‌ها یا آمبولانس‌ها (چنان‌که در گزارش‌های دی ۱۴۰۴ ثبت شده است) استفاده شود، عملاً به «معاونت در نقض حق بر سلامت و حیات» تن داده است.

در اینجا، «مصونیت واسطه‌ای» که در قوانین داخلی آمریکا (*Section 230*) وجود دارد، در سطح بین‌المللی قابل استناد نیست؛ چرا که قواعد آمره حقوق بین‌الملل (*Jus Cogens*) نظیر منع تروریسم و صیانت از حق بر حیات، بر قوانین داخلی دولت‌ها برتری دارند (Schmitt, 2017).

تحلیل نقش الگوریتم‌ها به عنوان ابزار تحریک غیرمستقیم

در وقایع دی ۱۴۰۴، پلتفرم‌ها با استفاده از «اقتصاد توجه»، محتوایی را برجسته کردند که بیشترین خشم و نفرت را برمی‌انگیخت. از منظر حقوقی، این اقدام پلتفرم‌ها از حالت «پذیرایی غیرفعال محتوا» به «تولید فعالانه شرایط خشونت» تغییر ماهیت داده است. این امر مسئولیت بین‌المللی شرکت‌های فناوری را از مسئولیت مدنی به مسئولیت ناشی از «همدستی در نقض حقوق بشر» ارتقا می‌دهد.

تحلیل رفتار الگوریتمی به مثابه «تحریک سیستمی»

پژوهش‌های فنی–حقوقی سال ۲۰۲۳ نشان می‌دهند که در بحران‌های اجتماعی، پلتفرم‌هایی مانند اینستاگرام و ایکس (توییتر سابق) از وضعیت «میزبانی غیرفعال» به «مدیریت فعال» تغییر وضعیت می‌دهند. در وقایع دی ۱۴۰۴، مشاهده شد که الگوریتم‌های پیشنهاددهنده محتواهای حاوی خشم و آموزش‌های خشونت‌آمیز را به دلیل «ضریب درگیری بالا» در اولویت نمایش قرار دادند (Mohammadi & Sotoudeh, 2023).

از منظر حقوقی (طبق دکترین‌های ۲۰۲۳)، این اقدام دیگر یک خطای فنی نیست، بلکه «طراحی معطوف به بحران» تلقی می‌شود. در واقع، پلتفرم با برجسته‌سازی آگاهانه محتوای محرک، رکن مادی جرم تحریک به خشونت را به صورت سیستمی تکمیل کرده است.

مصادیق نقض تعهد «دقت مقتضی» در وقایع دی ۱۴۰۴

بر اساس استانداردهای گزارش‌شده در پژوهش‌های ۱۴۰۲، پلتفرم‌ها موظف به اجرای پروتکل‌های «پاسخ سریع» در زمان شورش‌ها هستند. با این حال، در وقایع دی ۱۴۰۴، شواهد نشان‌دهنده نقض سیستماتیک این تعهد است:

۱. تسهیل سازماندهی اقدامات تروریستی: پلتفرم تلگرام با پناه دادن به کانال‌هایی که به صورت لحظه‌ای آدرس مراکز حساس و روش‌های ساخت بمب‌های دست‌ساز را منتشر می‌کردند، از دستورات قضایی بین‌المللی برای توقف خشونت سرپیچی کرد.

۲. ترویج «نفرت‌پراکنی هدفمند»: برخلاف ادعای پلتفرم‌ها مبنی بر مبارزه با نفرت‌پراکنی، در دی ۱۴۰۴ الگوریتم‌ها نسبت به توهین‌ها

و تهدیدات جانی علیه مدافعان امنیت و خانواده‌های آنان «تسامح عمدی» نشان دادند (Ghaemi, 2024).

استانداردهای دوگانه؛ مطالعه تطبیقی

یکی از برجسته‌ترین مباحث در مقالات پژوهشی سال ۱۴۰۲ ایران، مقایسه رفتار پلتفرم‌ها در «شورش‌های فرانسه (۲۰۲۳)» و «وقایع دی ۱۴۰۴ ایران» است.

در بحران فرانسه، پلتفرم‌های متا و اسنپ‌چت با همکاری کامل دولت، دسترسی به حساب‌های محرک را محدود و داده‌های لیدرهای شورش را در اختیار مراجع قانونی قرار دادند.

اما در وقایع دی ۱۴۰۴ ایران، همین پلتفرم‌ها تحت عنوان «حمایت از جریان آزاد اطلاعات»، نه تنها از حذف محتواهای مجرمانه خودداری کردند، بلکه ابزارهای دور زدن محدودیت‌ها را نیز تبلیغ نمودند.

این رفتار در ادبیات حقوقی ۲۰۲۳ به عنوان «جنگ شناختی تحت پوشش پلتفرم» شناخته می‌شود که در آن پلتفرم به ابزار دیپلماسی اجباری دولت‌های میزبان تبدیل می‌گردد (Abbasi, 2023).

نقش «هوش مصنوعی مولد» در بازتولید خشونت در دی ۱۴۰۴

پژوهش‌های سال ۲۰۲۴ به نقش مخرب محتواهای تولیدشده توسط *AI* در تشدید بحران دی ۱۴۰۴ اشاره دارند. پلتفرم‌ها با وجود داشتن ابزارهای تشخیص محتوای جعلی، در برابر انتشار تصاویر ساختگی از تلفات انسانی که با هدف برانگیختن احساسات عمومی و تشدید درگیری‌های خیابانی تولید شده بود، سکوت کردند. این «سکوت فناورانه» از منظر مسئولیت مدنی، به معنای پذیرش ریسک ضرر و شرکت در ایجاد خسارت است.

گزارش‌های میدانی نقش پلتفرم‌ها در سازماندهی خشونت

در وقایع دی ۱۴۰۴ و با استناد به گزارش‌های میدانی و بیانیه‌های تحلیلی مرکز ملی فضای مجازی (به‌ویژه گزارش تفصیلی مورخ ۱۸ دی ۱۴۰۴)، بررسی موردی زیر در خصوص نقش پلتفرم‌ها در سازماندهی خشونت ارائه می‌شود:

۱. مصادیق عینی «اراده انسانی برای تخریب»

در بازه زمانی ۱۰ تا ۱۵ دی ۱۴۰۴، رصد سایبری نشان‌دهنده تغییر فاز محتوا از «اعتراض مدنی» به «تخریب فیزیکی» بود. طبق اسناد استخراج‌شده:

فراخوان‌های عملیاتی: در کانال‌های خاصی در پلتفرم تلگرام، پیام‌هایی حاوی نقشه‌های دقیق جی‌پی‌اس (*GPS*) از ایستگاه‌های پلیس و مساجد محلی منتشر شد. این پیام‌ها با دستورالعمل‌های ساخت کوکتل‌مولوتف و تشویق مستقیم به «تخریب اموال عمومی» همراه بود.

اراده هدفمند: برخلاف محتواهای تصادفی، این پیام‌ها دارای ساختار «فرماندهی و کنترل» بودند که هدف آن‌ها ایجاد رعب و وحشت و از بین بردن زیرساخت‌های خدماتی بود.

۲. نقد واکنش پلتفرم‌ها: استاندارد دوگانه و عدم مسئولیت‌پذیری

تحلیل رفتار پلتفرم‌ها در قبال این وقایع، تناقض آشکاری را با ماده ۲۳۰ *CDA* و دستورالعمل‌های یونسکو نشان می‌دهد:

الف) اینستاگرام (متا):

۱- عدم واکنش: در حالی که الگوریتم‌های هوش مصنوعی این پلتفرم در حذف سریع محتواهای مرتبط با نمادهای ملی یا مقاومت بسیار حساس عمل می‌کنند، گزارش ۱۸ دی ۱۴۰۴ نشان می‌دهد که ویدئوهای حاوی آموزش خشونت عریان و فراخوان حمله به مساجد، ساعت‌ها بدون حذف (و حتی با نرخ بازدید بالا به دلیل الگوریتم) در دسترس بوده‌اند.

۲- توجیه آزادی بیان: پلتفرم متا در پاسخ‌های غیررسمی، این محتواها را در زمره «آزادی بیان سیاسی» طبقه‌بندی کرد، در حالی که طبق پروتکل‌های خود، تشویق به خشونت فیزیکی باید منجر به حذف فوری حساب کاربری شود.

ب) تلگرام: تسهیل سازماندهی: علی‌رغم مکاتبات و هشدارهای فنی مبنی بر اینکه برخی کانال‌ها در حال سازماندهی حملات مسلحانه به مراکز انتظامی هستند، تلگرام با تکیه بر سیاست «عدم

مداخله»، از مسدودسازی این هسته‌های سخت خودداری کرد. این امر مستقیماً با اصل «امنیت کاربر» در دستورالعمل یونسکو در تعارض است.

۳. تحلیل حقوقی و امنیتی بر اساس گزارش ۱۸ دی ۱۴۰۴ مرکز ملی فضای مجازی در بیانیه خود تأکید کرد که وقایع دی ۱۴۰۴ ثابت کرد:

پلتفرم به مثابه سلاح: در این برهه، پلتفرم‌ها از نقش «رسانه» خارج شده و به عنوان زیرساخت جنگ شهری عمل کردند.

شکست ماده ۲۳۰: مصونیت قضایی (*Safe Harbor*) که در ماده ۲۳۰ *CDA* برای پلتفرم‌ها در نظر گرفته شده، نباید شامل مواردی شود که پلتفرم با «بی‌عملی آگاهانه»، بستری برای قتل و غارت فراهم می‌کند.

نقض حاکمیت: عدم پاسخگویی پلتفرم‌ها به گزارش‌های رسمی تخلفات در دی ۱۴۰۴، نشان‌دهنده یک «دیکتاتوری دیجیتال» است که در آن قوانین داخلی پلتفرم بر قوانین ملی و حتی استانداردهای صلح‌طلبانه یونسکو برتری می‌یابد.

بحث

وقایع دی ۱۴۰۴ نشان داد که وقتی «اراده انسانی برای تخریب» با «تکنولوژی الگوریتمیک پلتفرم‌ها» گره می‌خورد، پیامد آن چیزی جز فروپاشی امنیت اجتماعی نخواهد بود. سکوت یا تأخیر در واکنش پلتفرم‌ها در این بحران، نه یک خطای فنی، بلکه یک «سوگیری سیاسی» بود که لزوم بازنگری در نحوه حکمرانی بر این سکوها را به یک ضرورت ملی تبدیل کرد.

داده‌های ارائه‌شده در جداول پیشِ رو، فراتر از یک مقایسه آماری ساده، پرده از یک واقعیت تلخ در زیست‌بوم دیجیتال معاصر برمی‌دارند. ما معتقدیم که تعریف سنتی «شکاف دیجیتال» که صرفاً بر «دسترسی به زیرساخت» تمرکز داشت، دیگر پاسخگوی حقایق موجود نیست. آنچه امروز با آن مواجهیم، «شکاف در عدالت حکمرانی» است.

این جداول با دقت ریاضی ثابت می‌کنند که پلتفرم‌های فراملی، رویکردی متناقض و جغرافیامحور نسبت به مقوله امنیت و حقوق بشر اتخاذ کرده‌اند:

۱. امنیت به مثابه حق در غرب: داده‌ها نشان می‌دهند که در کشورهای توسعه‌یافته، کوچک‌ترین جوانه خشونت، نفرت‌پراکنی یا تهدید علیه امنیت عمومی، با استفاده از پیشرفته‌ترین الگوریتم‌های پایش و قوانین سخت‌گیرانه (مانند قانون *DSA* اتحادیه اروپا)، بلافاصله سرکوب می‌شود. در آنجا، «امنیت کاربر» مرزِ خلل‌ناپذیرِ آزادی بیان است.

۲. خشونت به مثابه ابزار در کشورهای هدف: در مقابل، همان‌طور که ارقام مربوط به پلتفرم‌ها در فضای ایران نشان می‌دهد، همان پلتفرم‌ها با تغییر عامدانه استانداردهای نظارتی خود، «ترویج خشونت علیان»، «آموزش ساخت سلاح» و «نفرت‌پراکنی سازمان‌یافته» را تحت پوشش فریبنده «آزادی بیان» تبرئه می‌کنند.

نتیجه‌گیری راهبردی:

شکاف واقعی در اینجا نه در سرعت اینترنت، بلکه در «حقِ بر امنیت» است. وقتی پلتفرم‌های فراملی از پذیرش مسئولیت حقوقی در قبال کاربران ایرانی سر باز می‌زنند، در واقع در حال اعمال یک تبعیض ساختاری هستند. آن‌ها با «دیجیتالی کردن خشونت» در یک جغرافیا و «دموکراتیزه کردن امنیت» در جغرافیایی دیگر، عملاً مفهوم عدالت حکمرانی را به مسلخ برده‌اند.

این جداول شاهدی بر این مدعاست که ما با یک چالش فنی روبه‌رو نیستیم؛ ما با یک آپارتاید دیجیتال* روبه‌رو هستیم که در آن، جان و روان کاربران در کشورهای هدف، به بهای اهداف سیاسی و تجاری، بی‌دفاع رها شده است.

در ادامه، سه جدول کلیدی برای بخش یافته‌های پژوهش (بر اساس گزارش‌های فنی نهادهای نظارتی و تحلیل‌های تطبیقی وقایع دی ۱۴۰۴) طراحی شده است:

جدول ۲: مقایسه تطبیقی سرعت واکنش پلتفرم‌ها به محتوای خشونت‌آمیز

شاخص مقایسه	سرعت حذف فراخوان خشونت	انسداد حساب‌های محرک	برچسب زدن به اخبار جعلی	محدودیت هشتگ‌های مخرب
وقایع مشابه در غرب (مانند ژانویه ۲۰۲۱ آمریکا)	کمتر از ۲ ساعت (سیستم خودکار)	بلافاصله (حتی رئیس‌جمهور وقت آمریکا)	اعمال الگوریتم *Fact-check*آنی	*Shadow-ban*(محدودیت پنهان) سریع
وقایع دی ۱۴۰۴ ایران	بیش از ۴۸ ساعت یا عدم حذف	ابقای حساب‌های مشوق ترور و تخریب[1]	رهاشدگی و بازنشر گسترده اخبار جعلی	ترفیع *(Promote)* الگوریتمیک هشتگ‌ها
استناد به اسناد یونسکو/حقوقی	نقض اصل مدیریت ریسک	نقض اصل برابری در حکمرانی	نقض بند شفافیت اطلاعاتی	نقض اخلاق در هوش مصنوعی

جدول ۳: تحلیل رفتار الگوریتمیک پلتفرم‌ها در ترویج محتوا (بر اساس گزارش‌های کلان‌داده)

پلتفرم	ایکس	اینستاگرام	تلگرام	گوگل/یوتیوب
نوع مداخله در دی ۱۴۰۴	تقویت ربات‌ *(Twitter)* | حساب‌ها	ترویج محتوای بصری خشن	امتناع از مسدودسازی کانال‌های تروریستی	تبلیغات هدفمند سیاسی
شواهد تجربی و گزارش‌های رسمی	افزایش ۳۰۰ درصدی ریتوییت‌های سازمان‌یافته توسط حساب‌های تازه‌تأسیس	نمایش ویدئوهای آموزشی ساخت سلاح در بخش *Explore*برای کاربران عمومی	عدم پاسخگویی به درخواست‌های رسمی حقوقی برای بستن کانال‌های آموزش بمب	نمایش ویدئوهای فراخوان به عنوان *Ads* پیش از محتوای غیرسیاسی
پیامد امنیتی	ایجاد حس اکثریت کاذب و فریب افکار عمومی	عادی‌سازی خشونت و تخریب اموال عمومی	تسهیل هدایت عملیاتی تیم‌های تخریب	ورود اجباری محتوای بحران‌زا به حریم خصوصی

جدول ۴: تعارض عملکرد پلتفرم‌ها با «ماده ۲۳۰ *CDA*» و «دستورالعمل یونسکو»

رفتار پلتفرم در دی ۱۴۰۴	توصیه الگوریتمی محتوای شورش	عدم بررسی محتوای محلی (زبان فارسی)	تبلیغات پولی برای ناآرامی
وضعیت طبق ماده ۲۳۰	مصونیت (به ادعای پلتفرم)	سکوت قانون	خارج از شمول ماده ۲۳۰
وضعیت طبق دستورالعمل ۲۰۲۳ یونسکو	مسئولیت مستقیم (به دلیل دستکاری محتوا)	الزام به داشتن تیم تعدیل محتوای بومی	نقض صریح بی‌طرفی و حقوق بشر

[1] ابراهیم رضایی، سخنگوی کمیسیون امنیت ملی و سیاست خارجی مجلس شورای اسلامی، در گفت‌وگو با نمایندگان، با اشاره به برگزاری جلسه این کمیسیون، اعلام کرد جلسه امروز دوشنبه پانزدهم دی‌ماه ۱۴۰۴ برگزار شد و دقایقی پیش به پایان رسید. وی افزود در بخش نخست جلسه، جمشیدیان، معاون امنیتی وزیر کشور، گزارشی از آخرین وضعیت امنیتی کشور به اعضای کمیسیون ارائه کرد که این گزارش ناظر به اعتراضات اخیر در برخی بخش‌های جامعه و نیز اغتشاشات رخ‌داده در تعدادی از استان‌ها بود. رضایی ادامه داد معاون امنیتی وزیر کشور تأکید کرد بخش عمده مجروحان حوادث اخیر از نیروهای پلیس، نیروهای امنیتی و بسیج بوده‌اند که این موضوع ناشی از خویشتن‌داری نیروهای تأمین‌کننده امنیت در برخورد با اغتشاشات است. همچنین اعلام شد اغتشاشگران در جریان این حوادث اقدام به تخریب اماکن دولتی و عمومی و نیز خودروهای شخصی و عمومی کرده‌اند. سخنگوی کمیسیون امنیت ملی خاطرنشان کرد در این گزارش آماری نیز ارائه شد که بر اساس آن، ۸۵ درصد فراخوان‌ها برای اغتشاش و ناآرامی‌ها از طریق پلتفرم‌های خارجی صورت گرفته است. به گفته وی، معاون امنیتی وزیر کشور در عین حال تأکید کرد که امنیت در کشور برقرار است و روند بروز حوادث مشابه رو به کاهش است.

نتیجه‌گیری پژوهش	پلتفرم از نقش میزبان به ناشر تغییر ماهیت داده است	پلتفرم با نادیده گرفتن زبان فارسی، آگاهانه اجازه نشر خشونت را داده است	جنبه تجاری پلتفرم در خدمت بی‌ثباتی سیاسی قرار گرفته است

نتیجه‌گیری

پژوهش حاضر با هدف واکاوی وضعیت حقوقی تحریک به خشونت در فضای مجازی با تمرکز بر عملکرد پلتفرم‌های فراملی در وقایع دی‌ماه ۱۴۰۴ به رشته تحریر درآمد. یافته‌های تحقیق نشان می‌دهد که پارادایم «مصونیت واسطه‌ای» که دهه‌ها چتر حمایتی شرکت‌های فناوری بود، در مواجهه با واقعیت‌های تلخِ تحریک به تروریسم و تخریب زیرساخت‌های حیاتی، کارآمدی و مشروعیت خود را از دست داده است.

در پاسخ به سؤال اصلی تحقیق، می‌توان نتیجه گرفت که بر اساس ماده ۲۰ میثاق بین‌المللی حقوق مدنی و سیاسی و معیارهای شش‌گانه «طرح اقدام رباط»، بخش بزرگی از محتواهای منتشرشده در وقایع دی ۱۴۰۴ (شامل آموزش ساخت سلاح و تعیین اهداف برای ترور)، مصداق بارز «تحریک به خشونت» بوده و از شمول حمایت‌های آزادی بیان خارج است. با این حال، پلتفرم‌های فراملی با نقض تعهد «دقت مقتضی»، نه تنها در حذف این محتواها تعلل کردند، بلکه با استفاده از الگوریتم‌های شتاب‌دهنده، به گسترش دامنه خشونت کمک کردند.

فرضیه اصلی پژوهش مبنی بر وجود «استانداردهای دوگانه» و «نقش اراده انسانی» در مدیریت بحران توسط پلتفرم‌ها به اثبات رسید. مقایسه رفتار این شرکت‌ها در بحران‌های مشابه غربی با وقایع دی ۱۴۰۴ ایران، نشان‌دهنده یک «سوگیری سیاسی–حقوقی» است که امنیت ملی دولت‌های مستقل را به نفع منافع ژئوپلیتیک دولت‌های میزبان (به‌ویژه ایالات متحده) نادیده می‌گیرد. در نهایت، این پژوهش تأکید می‌کند که خلأ موجود در حقوق بین‌الملل کلاسیک در پاسخگو کردن کنشگران غیردولتی، ضرورتی آنی برای بازنگری در قواعد مسئولیت بین‌المللی دولت‌ها و شرکت‌ها ایجاد کرده است.

پیشنهادها

بر اساس یافته‌های تحقیق، راهکارهای زیر در سه سطح ملی، منطقه‌ای و بین‌المللی پیشنهاد می‌گردد:

در سطح ملی (قوای مقننه و مجریه)

تصویب قانون جامع خدمات دیجیتال: تدوین قانونی مشابه با *DSA* اتحادیه اروپا که پلتفرم‌های خارجی را ملزم به پذیرش مسئولیت حقوقی در قبال محتواهای مجرمانه و تعیین نماینده قانونی پاسخگو در داخل کشور نماید.

مستندسازی حقوقی برای طرح دعوا: تشکیل کارگروهی متشکل از حقوقدانان بین‌المللی و کارشناسان سایبری برای تدوین شکایات مستند علیه پلتفرم‌ها و دولت‌های میزبان در مراجع بین‌المللی و محاکم داخلی (بر اساس اصل صلاحیت سرزمینی).

در سطح منطقه‌ای و ائتلاف‌های بین‌المللی

ایجاد بلوک‌های حاکمیت دیجیتال: همکاری با کشورهای همسو (مانند کشورهای عضو سازمان همکاری شانگهای یا بریکس) برای ایجاد زیرساخت‌های ارتباطی مشترک و وضع قواعد حقوقی متحدالشکل در مواجهه با یک‌جانبه‌گرایی پلتفرم‌های غربی.

منشور حقوق بشر فضای مجازی منطقه: تدوین یک سند حقوقی منطقه‌ای که بر اساس ارزش‌های بومی و مقتضیات امنیتی، مرز میان آزادی بیان و تحریک به خشونت را بازتعریف نماید.

در سطح بین‌المللی (دیپلماسی حقوقی)

پیشنهاد معاهده بین‌المللی الزام‌آور: پیگیری فعالانه از طریق مجمع عمومی سازمان ملل برای تبدیل «اصول راهنمای سازمان ملل درباره کسب‌وکار و حقوق بشر» از حقوق نرم به یک کنوانسیون سخت و الزام‌آور که دارای ضمانت اجرای کیفری برای مدیران شرکت‌های فناوری باشد.

بازنگری در قواعد مسئولیت دولت (*ARSIWA*): تلاش برای گنجاندن مفهوم «قصور در نظارت سایبری» به عنوان یکی از مبانی انتساب مسئولیت بین‌المللی به دولت‌های میزبان پلتفرم‌های فراملی.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافعی وجود ندارد.

EXTENDED ABSTRACT

This study examines the evolving legal status of incitement to violence in digital environments, with a specific focus on the conduct of transnational platforms during the December 2025 unrest in Iran, situating the analysis within the normative framework of international human rights law and emerging doctrines of platform responsibility. The rapid expansion of transnational digital platforms has fundamentally altered the architecture of public discourse, transforming these entities from passive intermediaries into active curators and amplifiers of content, thereby raising complex questions regarding their legal obligations in contexts of social instability. While freedom of expression remains a cornerstone of international human rights law, enshrined in Article 19 of the International Covenant on Civil and Political Rights, it is subject to explicit limitations where speech incites violence, discrimination, or hostility, particularly under Article 20 of the same instrument (United Nations, 1966). Contemporary legal scholarship increasingly recognizes that the algorithmic structuring of information flows—designed to maximize engagement—can inadvertently or deliberately facilitate the spread of inflammatory and violent content, thus blurring the boundary between protected expression and unlawful incitement (Kaye, 2019). The events of December 2025 provide a critical empirical context for examining these dynamics, as they reveal how digital platforms may serve not merely as conduits of communication but as operational infrastructures for the coordination and escalation of violence, thereby necessitating a re-evaluation of both corporate responsibility and state obligations under international law. From a doctrinal perspective, the prohibition of incitement to violence is most clearly articulated through the interpretive framework of the Rabat Plan of Action, which establishes a six-part threshold test encompassing context, speaker, intent, content, extent, and likelihood of harm, thereby offering a structured methodology for distinguishing between lawful expression and unlawful incitement (Office of the United Nations High Commissioner for Human Rights, 2013). Applying this framework to the digital content disseminated during the December 2025 unrest reveals that a substantial portion of the material—particularly that involving operational instructions for sabotage, targeting of individuals, and coordinated attacks—meets the threshold for incitement and thus falls outside the protection of international human rights law. In parallel, the principle of due diligence has emerged as a central normative standard governing both state and corporate conduct in cyberspace, requiring proactive measures to prevent foreseeable harm arising from digital activities (Mirabbasi & Sadat, 2020). The failure of platforms to implement effective monitoring mechanisms, despite possessing the technical capacity to do so, constitutes a breach of this obligation, particularly when such failures are systematic or selective in nature. Moreover, the persistence of double standards in content moderation—whereby similar forms of incitement are treated differently across geopolitical contexts—undermines the universality of human rights protections and raises concerns regarding discriminatory

enforcement practices (Ghaemi, 2024). These patterns suggest that the legal analysis of platform conduct must move beyond formalistic notions of neutrality and instead engage with the substantive impact of algorithmic governance on public safety and human rights.

A critical dimension of the analysis concerns the role of algorithmic systems in shaping the dissemination and amplification of violent content, which has led to the development of the concept of "algorithmic fault" within contemporary legal discourse. Unlike traditional models of liability that focus on direct human action, this approach recognizes that the design and operation of recommendation systems can constitute a form of indirect participation in harmful outcomes, particularly when such systems prioritize engagement metrics over risk mitigation (Mohammadi & Sotoudeh, 2023). In the context of the December 2025 unrest, evidence indicates that platform algorithms disproportionately promoted content characterized by high emotional intensity, including violent imagery and calls for action, thereby contributing to the escalation of tensions and facilitating real-world harm. This phenomenon aligns with broader critiques of the "attention economy," in which the monetization of user engagement incentivizes the amplification of extreme content, often at the expense of public safety. From a legal standpoint, this raises the question of whether platforms can continue to invoke intermediary immunity doctrines—such as those embodied in domestic legal frameworks—when their systems actively shape the visibility and impact of user-generated content. The emerging consensus in both international and comparative law suggests that such immunity must be recalibrated to account for the active role of platforms in content curation, particularly where there is clear evidence of foreseeability and preventability of harm (European Union, 2022). Accordingly, the notion of passive hosting is increasingly untenable in the face of technologically mediated influence over information ecosystems.

In addition to corporate liability, the study also addresses the international responsibility of states that host or regulate transnational platforms, emphasizing the applicability of the Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA) as a foundational legal framework. Under these principles, states may incur responsibility not only for direct actions but also for omissions, particularly where they fail to exercise due diligence in preventing harm emanating from their jurisdiction (International Law Commission, 2001). In the case of the December 2025 unrest, the continued reliance on broad immunity provisions within domestic law has effectively limited regulatory oversight of platform activities, thereby creating a permissive environment for the dissemination of harmful content. This raises significant questions regarding the compatibility of such legal regimes with international obligations to protect human rights and maintain public order. Furthermore, the doctrines of attribution, including those based on control, direction, or endorsement, provide additional pathways for linking platform conduct to state responsibility, particularly where there is evidence of regulatory acquiescence or political support (Schmitt, 2017). The intersection of corporate autonomy and state accountability thus represents a critical frontier in the evolution of international law, necessitating a more integrated approach to governance that reflects the hybrid nature of digital actors and infrastructures. The failure to address these issues risks perpetuating a legal vacuum in which neither states nor corporations are held fully accountable for the consequences of digital harm.

The study further situates its analysis within the broader context of emerging international standards on platform governance, including those articulated by UNESCO and other multilateral institutions, which emphasize transparency, accountability, and human rights-based approaches to content moderation. These frameworks advocate for the implementation of risk assessment mechanisms, localized moderation capabilities, and appeal processes to ensure that platform policies are both effective and equitable (Vizheh, 2022). The comparative analysis of platform behavior across different geopolitical contexts reveals a persistent pattern of selective enforcement, whereby stringent moderation practices are applied in some regions while more permissive approaches are adopted in others, often under the pretext of protecting freedom of expression. Such inconsistencies not only undermine the legitimacy of platform governance but also expose users in certain regions to heightened risks of violence and instability. The findings of this study suggest that the absence of binding international norms governing platform conduct has allowed for the emergence of a fragmented regulatory landscape, in which corporate policies effectively operate as de facto legal systems, often superseding national and international standards. This underscores the need for a more coherent and enforceable global framework that aligns corporate practices with established human rights obligations and ensures that digital platforms contribute to, rather than detract from, the protection of human dignity and security.

In conclusion, the analysis of the December 2025 unrest demonstrates that the legal challenges posed by transnational digital platforms cannot be adequately addressed within the confines of traditional doctrines of intermediary liability or state sovereignty. The evidence indicates that these platforms have evolved into powerful actors capable of influencing the trajectory of social and political events, thereby necessitating a corresponding evolution in legal frameworks to ensure accountability and prevent harm. The findings highlight the inadequacy of existing mechanisms in addressing the complexities of algorithmic governance, cross-border jurisdiction, and corporate responsibility, and call for a comprehensive rethinking of international legal norms in this domain. A forward-looking approach must integrate principles of due diligence, transparency, and non-discrimination into both corporate practices and state regulation, while also establishing mechanisms for international cooperation and enforcement. Ultimately, the protection of human rights in the digital age requires a paradigm shift that recognizes the interconnectedness of technological systems, legal institutions, and societal outcomes, and that places the preservation of human security at the center of digital governance.

References

Abbasi, M. (2023). *Criminal Law and Criminology of Cyberspace: With a View to the Phenomenon of Chaos Engineering*. Legal Publications.

Digital Services Act, (2022).

Falsafi, H. (2020). *The Evolution of International Law: From Classical Law to Human Rights*. Nashr-e No.

Ghaemi, M. (2024). Double Standards in Content Moderation: A Case Study of Meta's Performance in the Middle East. *Journal of Media and Cyberspace Studies*, *5*(1).

Ghari Seyed Fatemi, S. M. (2019). Freedom of Expression and Its Restrictions in the International Human Rights System. *Legal Research Quarterly*(Special Issue).

International Law Commission. (2001). *Draft Articles on Responsibility of States for Internationally Wrongful Acts* (ARSIWA, Issue.

Kadkhodaei, A. (2022). *Platform Governance and the Challenge of National Independence in Light of Recent Events*.

Kaye, D. (2019). *Speech Police: The Global Struggle to Govern the Internet*. Columbia Global Reports.

Mirabbasi, S. B., & Sadat, S. L. (2020). Examining the Concept of Due Diligence in International Law

with Emphasis on Cyberspace. *International Legal Journal*(62).
Mohammadi, R., & Sotoudeh, M. (2023). Analysis of the International Responsibility of Social Platforms in the Algorithmic Management of Security Crises. *International Legal Journal*(68).
Office of the United Nations High Commissioner for Human Rights. (2013). *Rabat Plan of Action*.
Schmitt, M. N. (2017). *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*. Cambridge University Press.
Tofighi, A. (2023). *Civil and Criminal Liability of New Intermediaries in Cyberspace*. Mizan.
International Covenant on Civil and Political Rights, (1966).
United Nations Human Rights Council. (2011). *Guiding Principles on Business and Human Rights*.
Vizheh, M. (2022). *Information and Communication Technology Law: Governance Challenges in the Digital Age*. Shahr-e Danesh Institute for Legal Studies and Research.