

Examining the Liability of Platforms for the Dissemination of Criminal or Misleading Visual Representations in the Digital Environment

Alireza Abousaeidi^{1*}

1*- Master of Arts in Illustration, Islamic Azad University, Central Tehran Branch, Tehran, Iran

ABSTRACT

Rapid developments in digital technologies and the expansion of content creation and sharing platforms have led to the emergence of new environments for the production and dissemination of various forms of visual representations in cyberspace. In some cases, such representations may have a criminal or misleading nature and may cause harmful effects on public security, individual reputation, and social trust. This study aims to examine and clarify the scope and foundations of the legal liability of platforms regarding the dissemination of such content, analyzing the civil and criminal dimensions of their responsibility within the framework of cyber law. In this regard, the concept and manifestations of criminal and misleading visual representations in the digital context are first explained, followed by an analysis of the role of platforms as information intermediaries in the production, processing, and dissemination of content. Furthermore, the study examines the impact of factors such as algorithmic design, content moderation and filtering policies, the platform's awareness of harmful content, and their level of economic benefit in determining the scope of liability. The findings indicate that in most legal systems, platform liability is not absolute but rather depends on criteria such as knowledge and awareness, effective control, and the active or passive role in content dissemination. Accordingly, approaches such as conditional immunity, notice-and-takedown regimes, and fault-based liability have been adopted in this area. At the same time, the emergence of new technologies such as generative artificial intelligence and advanced image-generation tools has significantly increased the complexity of determining liability and highlighted the need to reconsider traditional legal rules. Finally, this research emphasizes the necessity of balancing freedom of expression, digital innovation, and the protection of individual rights, and proposes measures aimed at developing legislative frameworks, enhancing platform transparency, and strengthening regulatory mechanisms.

Keywords:

Platform liability, criminal visual representations, misleading content, digital environment, content moderation, civil liability

Article Type: Research Article

How to Cite: Abousaeidi, A. (2026). Examining the Liability of Platforms for the Dissemination of Criminal or Misleading Visual Representations in the Digital Environment. *Journal of Cyber Law (JOCL)*, 2(4), 59-70. doi:10.22054/jocl.2025.8563.7894.

Journal of Cyber Law in Development and Evolution is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.

© Authors



¹Corresponding Author: alireza.abousaeidi@gmail.com

بررسی مسئولیت پلتفرم‌ها در انتشار تصویرسازی‌های مجرمانه یا گمراه‌کننده در فضای دیجیتال

علیرضا ابوسعیدی*

۱- کارشناسی ارشد تصویرسازی، دانشگاه آزاد اسلامی، واحد تهران مرکز، تهران، ایران

چکیده

تحولات شتابان فناوری‌های دیجیتال و گسترش پلتفرم‌های مبتنی بر تولید و اشتراک‌گذاری محتوا، موجب شکل‌گیری بسترهای نوینی برای خلق و انتشار انواع تصویرسازی‌ها در فضای مجازی شده است؛ تصویرسازی‌هایی که در برخی موارد می‌توانند ماهیتی مجرمانه یا گمراه‌کننده داشته و آثار زیان‌باری بر امنیت عمومی، حیثیت اشخاص و اعتماد اجتماعی برجای گذارند. این پژوهش با هدف بررسی و تبیین حدود و مبانی مسئولیت حقوقی پلتفرم‌ها در قبال انتشار چنین محتواهایی، به تحلیل ابعاد مختلف مسئولیت مدنی و کیفری آن‌ها در چارچوب حقوق سایبری می‌پردازد. در این راستا، ابتدا مفهوم و مصادیق تصویرسازی‌های مجرمانه و گمراه‌کننده در بستر دیجیتال تبیین شده و سپس نقش پلتفرم‌ها به‌عنوان واسطه‌های اطلاعاتی در چرخه تولید، پردازش و انتشار محتوا مورد واکاوی قرار می‌گیرد. همچنین، تأثیر عواملی نظیر طراحی الگوریتمی، سیاست‌های تعدیل و پالایش محتوا، سطح آگاهی پلتفرم از محتوای زیان‌بار و میزان انتفاع اقتصادی آن‌ها در تعیین حدود مسئولیت مورد بررسی قرار گرفته است. یافته‌های پژوهش نشان می‌دهد که در اغلب نظام‌های حقوقی، مسئولیت پلتفرم‌ها به‌صورت مطلق نبوده بلکه تابعی از معیارهایی همچون علم و اطلاع، کنترل مؤثر و نقش فعال یا منفعل در انتشار محتوا است و رویکردهایی نظیر مصونیت مشروط، مسئولیت مبتنی بر اخطار و حذف، و یا مسئولیت تقصیرمحور در این زمینه اتخاذ شده‌اند. در عین حال، ظهور فناوری‌های نوین مانند هوش مصنوعی مولد و ابزارهای پیشرفته تصویرسازی، پیچیدگی تعیین مسئولیت را دوچندان کرده و ضرورت بازنگری در قواعد سنتی را برجسته ساخته است. در پایان، این پژوهش با تأکید بر لزوم ایجاد توازن میان آزادی بیان، نوآوری دیجیتال و حمایت از حقوق اشخاص، پیشنهادهایی در جهت توسعه چارچوب‌های تقنینی، ارتقای شفافیت عملکرد پلتفرم‌ها و تقویت سازوکارهای نظارتی ارائه می‌دهد.

کلیدواژه‌ها:

مسئولیت پلتفرم‌ها، تصویرسازی مجرمانه، محتوای گمراه‌کننده، فضای دیجیتال، تعدیل محتوا، مسئولیت مدنی.

نوع مقاله: پژوهشی

نحوه استناد:

ابوسعیدی، علیرضا. (۱۴۰۴). بررسی مسئولیت پلتفرم‌ها در انتشار تصویرسازی‌های مجرمانه یا گمراه‌کننده در فضای دیجیتال. (۴)۲، ۵۹-۷۰.

نشریه حقوق سایبری در توسعه و تکامل تحت مجوز کرییتیو کامنز انتساب - غیرتجاری ۴٫۰ بین‌المللی منتشر شده است.

©نویسنده‌گان



ایمیل نویسنده مسئول: alireza.abousaidi@gmail.com

۱. مقدمه

فضای دیجیتال در دهه‌های اخیر به یکی از مهم‌ترین بسترهای شکل‌گیری، بازنمایی و انتقال واقعیت‌های اجتماعی تبدیل شده است؛ فضایی که در آن، تصویر نه تنها ابزار بیان، بلکه سازنده ادراک و حتی جایگزین واقعیت تلقی می‌شود (Leone, 2023)، گسترش پلتفرم‌های مبتنی بر تولید و اشتراک‌گذاری محتوا، در کنار پیشرفت فناوری‌های پردازش تصویر و هوش مصنوعی، امکان خلق انواع تصویرسازی‌ها را به گونه‌ای فراهم ساخته که مرز میان واقعیت و جعل، بیش از هر زمان دیگری مبهم شده است (Mirsky & Lee, 2020). در چنین شرایطی، تصویرسازی دیگر صرفاً بازنمایی واقعیت نیست، بلکه می‌تواند خود به ابزاری برای تحریف، فریب یا حتی ارتکاب جرم تبدیل شود. تصویرسازی‌های مجرمانه یا گمراه‌کننده، به‌ویژه در قالب‌هایی نظیر تصاویر دستکاری‌شده، دیپ‌فیک‌ها و بازنمایی‌های هدفمند، می‌توانند آثار زیانباری بر حقوق فردی و نظم عمومی برجای گذارند. این نوع محتواها، از یک سو ممکن است موجب نقض حیثیت، حریم خصوصی یا امنیت روانی اشخاص شوند و از سوی دیگر، با ایجاد ادراک نادرست در افکار عمومی، اعتماد اجتماعی و سلامت اطلاعاتی جامعه را تضعیف کنند. در این میان، پلتفرم‌های دیجیتال به‌عنوان بازیگران اصلی در چرخه تولید، انتشار و توزیع محتوا، نقشی تعیین‌کننده در گسترش یا کنترل این پدیده ایفا می‌کنند. (Cochran & Napshin, 2021)

با این حال، پرسش اساسی این است که حدود مسئولیت این پلتفرم‌ها در قبال چنین تصویرسازی‌هایی چیست؟ آیا صرف میزبانی محتوا، مسئولیت‌آور است یا تحقق مسئولیت نیازمند احراز شرایطی نظیر علم، کنترل یا مداخله فعال است؟ تجربه نظام‌های حقوقی مختلف نشان می‌دهد که پاسخ به این پرسش، نه تنها به ملاحظات حقوقی، بلکه به موازنه‌ای میان آزادی بیان، نوآوری فناوری و حمایت از حقوق اشخاص وابسته است. در نظام حقوقی ایران نیز، اگرچه قواعد کلی مسئولیت مدنی و کیفری قابلیت تسری به فضای دیجیتال را دارند، اما در خصوص مسئولیت پلتفرم‌ها در قبال تصویرسازی‌های مجرمانه یا گمراه‌کننده، ابهامات و خلأهای قابل توجهی وجود دارد. به‌ویژه آنکه ظهور فناوری‌های نوینی همچون هوش مصنوعی مولد، پیچیدگی این مسئله را دوچندان کرده و کاربست قواعد سنتی را با چالش مواجه ساخته است. (حسینی، ۱۴۰۲).

بر این اساس، مقاله حاضر با تمرکز بر تصویرسازی‌های مجرمانه یا گمراه‌کننده در فضای دیجیتال و مسئولیت پلتفرم‌ها، درصدد است ضمن تبیین مفهومی این پدیده، به بررسی معیارهای انتساب مسئولیت به پلتفرم‌ها پرداخته و کارآمدی رویکردهای موجود را مورد ارزیابی قرار دهد. در این مسیر، بهره‌گیری از رویکرد تحلیلی-انتقادی و توجه به حقوق تطبیقی، زمینه‌ای برای ارائه راهکارهای اصلاحی در جهت ایجاد توازن میان آزادی بیان و حمایت از حقوق اشخاص فراهم می‌سازد.

مبحث اول: مفهوم‌شناسی تصویرسازی‌های مجرمانه و گمراه‌کننده در فضای دیجیتال

تحلیل مسئولیت پلتفرم‌ها در قبال تصویرسازی‌های زیانبار، مستلزم درک دقیق از ماهیت و حدود این نوع محتواهاست. بدون تبیین روشن مفاهیم تصویرسازی مجرمانه و تصویرسازی گمراه‌کننده، تعیین قلمرو مسئولیت و تفکیک موارد مشمول یا غیرمشمول دشوار خواهد بود. (Kietzmann et al., 2020) از این رو، این مبحث به بررسی مفهومی و

تحلیلی این دو دسته از تصویرسازی‌ها می‌پردازد. در فضای دیجیتال، تصویرسازی‌ها دیگر محدود به بازنمایی‌های ساده نیستند، بلکه می‌توانند ترکیبی از داده‌های واقعی و مصنوعی باشند که با اهداف مختلف از سرگرمی گرفته تا فریب یا تخریب تولید می‌شوند. همین ویژگی، موجب شده است که مرزبندی میان مشروع و نامشروع بودن این محتواها، نیازمند تحلیل دقیق حقوقی باشد (Leone, 2023).

گفتار اول: تصویرسازی مجرمانه و ویژگی‌های آن

تصویرسازی مجرمانه را می‌توان به عنوان هر گونه محتوی بصری تعریف کرد که یا ذاتاً ناقص قواعد آمره حقوق کیفری است یا در بستر خاص، منجر به تحقق یک رفتار مجرمانه می‌شود (Citron & Chesney, 2019). این نوع تصویرسازی‌ها، بسته به ماهیت و کارکرد خود، ممکن است در قالب‌های مختلفی ظاهر شوند؛ از جمله نشر تصاویر خصوصی بدون رضایت، تولید تصاویر جعلی با هدف انتساب اعمال مجرمانه به اشخاص، یا استفاده از تصاویر برای فریب عمومی. آنچه تصویرسازی مجرمانه را از سایر انواع محتوا متمایز می‌کند، وجود عنصر ضرر حقوقی قابل شناسایی است؛ به این معنا که این تصویرسازی‌ها نه تنها از منظر اخلاقی یا اجتماعی ناپسند هستند، بلکه به‌طور مشخص یکی از حقوق مورد حمایت قانون را نقض می‌کنند. این نقض می‌تواند متوجه اشخاص (مانند حیثیت و حریم خصوصی) یا متوجه جامعه (مانند نظم عمومی) باشد (Yadlin-Segal & Oppenheim, 2020).

بند اول: تمایز تصویرسازی مجرمانه از تصویرسازی صرفاً گمراه‌کننده

یکی از چالش‌های اساسی در این حوزه، تمایز میان تصویرسازی مجرمانه و تصویرسازی گمراه‌کننده است. در حالی که اولی به‌طور مستقیم مشمول ضمانت اجرای کیفری قرار می‌گیرد، دومی لزوماً واجد وصف مجرمانه نیست، اما می‌تواند آثار زیانباری به همراه داشته باشد (Citron & Chesney, 2019). تصویرسازی گمراه‌کننده معمولاً در مواردی مطرح می‌شود که محتوا، بدون نقض صریح قانون، موجب ایجاد برداشت نادرست در مخاطب می‌شود. برای مثال، استفاده از تصاویر قدیمی در زمینه‌ای جدید، یا دستکاری جزئی تصاویر به گونه‌ای که معنای آن تغییر کند، می‌تواند مصداقی از این نوع تصویرسازی باشد (Wardle & Derakhshan, 2017). با این حال، در صورتی که این گمراه‌سازی به سطحی برسد که موجب ورود ضرر قابل توجه یا اختلال در نظم عمومی شود، ممکن است به حوزه مسئولیت حقوقی اعم از مدنی یا کیفری وارد گردد. در این میان، معیار تشخیص، نه صرفاً ماهیت تصویر، بلکه زمینه انتشار، قصد تولیدکننده و اثرگذاری بر مخاطب است. به بیان دیگر، یک تصویر ممکن است در یک بستر کاملاً بی‌ضرر تلقی شود، اما در بستر دیگر، آثار گمراه‌کننده یا حتی مجرمانه داشته باشد. این نسبی بودن، کار تحلیل حقوقی را پیچیده‌تر کرده و ضرورت توجه به معیارهای تکمیلی را افزایش می‌دهد. از منظر حقوقی، اهمیت این تفکیک در آن است که مسئولیت پلتفرم‌ها نیز بر اساس همین تمایز شکل می‌گیرد؛ به گونه‌ای که در مواجهه با محتوای مجرمانه، انتظار مداخله سریع و مؤثر بیشتر است، در حالی که در مورد محتوای گمراه‌کننده، ملاحظات آزادی بیان نقش پررنگ‌تری ایفا می‌کند (Khan, 2024). به همین دلیل، تعیین دقیق مرز میان این دو، پیش‌شرط هرگونه تحلیل درباره مسئولیت پلتفرم‌ها محسوب می‌شود.

گفتار دوم: تصویرسازی‌های گمراه‌کننده و جایگاه آن در مسئولیت حقوقی

تصویرسازی‌های گمراه‌کننده، اگرچه در بسیاری از موارد فاقد وصف کیفی صریح هستند، اما از حیث آثار اجتماعی و حقوقی، می‌توانند به مراتب پیچیده‌تر از تصویرسازی‌های مجرمانه باشند (Wardle & Derakhshan, 2017). این نوع تصویرسازی‌ها معمولاً در قالب بازنمایی‌های ناقص، تحریف‌شده یا خارج از زمینه واقعی شکل می‌گیرند و بدون آنکه لزوماً عنصر قانونی جرم را محقق سازند، موجب ایجاد برداشت نادرست در مخاطب می‌شوند. در حقوق ایران، اگرچه عنوان مستقلی تحت عنوان تصویرسازی گمراه‌کننده پیش‌بینی نشده است، اما می‌توان مصادیق آن را ذیل برخی عناوین کلی‌تر مورد بررسی قرار داد. برای مثال، در مواردی که تصویرسازی منجر به فریب عمومی یا اخلال در نظم اجتماعی شود، می‌توان آن را با مقررات مربوط به نشر اکاذیب یا تشویش اذهان عمومی مرتبط دانست. همچنین، در حوزه مسئولیت مدنی، هرگاه این نوع محتوا موجب ورود ضرر به اشخاص شود، قواعد عام مسئولیت مدنی قابل اعمال خواهد بود (کاتوزیان، ۱۳۸۶:۴۵). همچنی نظام‌های حقوقی مختلف رویکردهای متفاوتی در قبال این نوع محتوا اتخاذ کرده‌اند. برای مثال، در حقوق اتحادیه اروپا، مقررات مربوط به خدمات دیجیتال^۱ بر ضرورت شفافیت پلتفرم‌ها و مقابله با اطلاعات گمراه‌کننده تأکید دارد. در این چارچوب، حتی در مواردی که محتوا ماهیت مجرمانه ندارد، پلتفرم‌ها موظف به اتخاذ تدابیر لازم برای کاهش آثار زیانبار آن هستند (Khan, 2024).

مبحث دوم: مبانی و حدود مسئولیت پلتفرم‌ها در قبال تصویرسازی‌های دیجیتال

با گسترش نقش پلتفرم‌های دیجیتال در مدیریت جریان اطلاعات، تعیین حدود مسئولیت آنها به یکی از چالش‌های اساسی حقوق سایبری تبدیل شده است (Suzor, 2019). این پلتفرم‌ها، برخلاف رسانه‌های سنتی، نه تنها بستر انتشار محتوا را فراهم می‌کنند، بلکه از طریق الگوریتم‌ها، در نحوه توزیع و دیده‌شدن آن نیز نقش فعال دارند. همین ویژگی، تمایز اساسی میان واسطه منفعل و کنشگر فعال را در تحلیل مسئولیت آنها ایجاد می‌کند. در نظام حقوقی ایران، مقررات صریح و جامعی درباره مسئولیت پلتفرم‌ها وجود ندارد، اما می‌توان با استناد به قواعد عام، چارچوبی برای تحلیل این مسئولیت ترسیم کرد. برای مثال، بر اساس قانون مسئولیت مدنی مصوب ۱۳۳۹، هر شخصی که بدون مجوز قانونی به دیگری ضرر وارد کند، مسئول جبران آن است (ماده ۱). این قاعده کلی، در مورد پلتفرم‌ها نیز قابل اعمال است، مشروط بر آنکه رابطه سببیت و تقصیر اثبات شود (کاتوزیان، ۱۳۸۶: ۷۸).

گفتار اول: معیار علم و اطلاع در انتساب مسئولیت

یکی از مهم‌ترین معیارهای تعیین مسئولیت پلتفرم‌ها، علم و اطلاع آنها نسبت به محتوای زیانبار است. در بسیاری از نظام‌های حقوقی، از جمله ایالات متحده، اصل بر آن است که پلتفرم‌ها تا زمانی که از محتوای غیرقانونی آگاه نشده‌اند، مسئولیتی ندارند. این رویکرد که در قالب مصونیت مشروط شناخته می‌شود، به‌ویژه در بخش ۲۳۰ قانون ارتباطات آمریکا مورد تأکید قرار گرفته است. اندیشمندانی همچون پریت (۱۹۹۶)^۲ بر این باورند که مسئولیت واسطه‌ها باید مبتنی بر سطح آگاهی و امکان مداخله آنها باشد. به بیان دیگر، اگر پلتفرم از وجود محتوای زیانبار مطلع باشد و با وجود امکان

^۱ Digital Services Act

^۲ Perritt (1996)

حذف آن، اقدامی نکند، می‌توان مسئولیت را متوجه آن دانست. در حقوق ایران نیز، اگرچه چنین قاعده‌ای به صورت صریح بیان نشده، اما می‌توان از اصول کلی، به ویژه قواعد مربوط به تقصیر، چنین برداشتی را استنباط کرد. به عبارت دیگر، علم پلتفرم به محتوای زیانبار و عدم اقدام برای حذف آن، می‌تواند مصداقی از تقصیر تلقی شود.

بند اول: جایگاه اخطار و حذف در نظام مسئولیت

یکی از مهم‌ترین سازوکارهای عملی در تعیین مسئولیت پلتفرم‌ها، نظام اخطار و حذف^۱ است. بر اساس این مدل، پلتفرم‌ها در صورت دریافت اخطار درباره محتوای غیرقانونی، موظف به بررسی و در صورت لزوم، حذف آن هستند. عدم انجام این تکلیف، می‌تواند منجر به مسئولیت حقوقی شود.

این رویکرد، تلاشی برای ایجاد تعادل میان دو ارزش اساسی است:

- آزادی بیان

- حمایت از حقوق اشخاص

در حقوق تطبیقی، به ویژه در اتحادیه اروپا، این مدل به طور گسترده مورد پذیرش قرار گرفته است. در مقابل، برخی منتقدان بر این باورند که اتکای بیش از حد به این سازوکار، می‌تواند منجر به حذف بیش از حد محتوا و محدودسازی آزادی بیان شود (گلد اسمیت و وو، ۲۰۰۶).

گفتار دوم: نقش الگوریتم‌ها در تشدید مسئولیت پلتفرم‌ها

یکی از تحولات مهم در سال‌های اخیر، افزایش نقش الگوریتم‌ها در مدیریت و توزیع محتواست. پلتفرم‌ها با استفاده از الگوریتم‌های پیشنهاددهنده، برخی محتواها را برجسته تر کرده و در معرض دید کاربران بیشتری قرار می‌دهند. این امر، پرسش مهمی را مطرح می‌کند: آیا این مداخله الگوریتمی، پلتفرم را از یک واسطه منفعل به یک ناشر فعال تبدیل می‌کند؟ برخی اندیشمندان، از جمله ترلتون گیلسپی^۳، معتقدند که پلتفرم‌ها دیگر نمی‌توانند خود را صرفاً میزبان بی طرف معرفی کنند، زیرا طراحی الگوریتمی آنها به طور مستقیم بر شکل‌گیری جریان اطلاعات تأثیر می‌گذارد. در این چارچوب، اگر الگوریتم‌ها به گونه‌ای طراحی شوند که محتوای گمراه‌کننده یا زیانبار را تقویت کنند، می‌توان این امر را به عنوان یکی از عوامل تشدیدکننده مسئولیت پلتفرم در نظر گرفت.

گفتار سوم: معیار کنترل مؤثر و نقش فعال پلتفرم در انتشار محتوا

یکی دیگر از معیارهای کلیدی در تعیین مسئولیت پلتفرم‌ها، میزان کنترل مؤثر آنها بر محتوای منتشر شده است. در ادبیات حقوقی، تمایز میان واسطه منفعل و واسطه فعال از همین معیار نشأت می‌گیرد. واسطه منفعل صرفاً بستر فنی برای انتشار محتوا فراهم می‌کند، در حالی که واسطه فعال، به نحوی در انتخاب، سازمان‌دهی یا برجسته‌سازی محتوا مداخله دارد (Gillespie, 2018). در حقوق تطبیقی، به ویژه در نظام حقوقی اتحادیه اروپا، این تمایز نقش تعیین‌کننده‌ای در انتساب مسئولیت دارد. بر اساس رویه قضایی، هرگاه پلتفرم از طریق ابزارهایی مانند الگوریتم‌های پیشنهاددهنده یا سیاست‌های هدفمند، در جهت‌دهی به محتوا نقش ایفا کند، دیگر نمی‌توان آن را صرفاً یک میزبان بی طرف تلقی کرد.

¹ Notice and Takedown

² Goldsmith & Wu, 2006

³ Tarleton Gillespie

در چنین شرایطی، سطح مسئولیت افزایش می‌یابد. اندیشمندانی مانند ون ایکه (۲۰۱۱)^۱ بر این باورند که معیار کنترل، باید نه صرفاً بر اساس ون ایکه امکان فنی، بلکه بر مبنای کنترل واقعی و عملی ارزیابی شود. به بیان دیگر، صرف توانایی بالقوه برای حذف محتوا کافی نیست، بلکه باید بررسی شود که آیا پلتفرم در عمل، سازوکارهای مؤثری برای نظارت و مداخله دارد یا خیر. در حقوق ایران، اگرچه چنین تفکیکی به صورت صریح در قوانین بیان نشده، اما می‌توان آن را از اصول کلی استنباط کرد. برای مثال، در چارچوب قواعد مسئولیت مدنی، هرگاه شخصی نقشی فراتر از یک واسطه ساده ایفا کند و در ایجاد یا تشدید ضرر دخالت داشته باشد، مسئولیت او تقویت می‌شود (کاتوزیان، ۱۳۸۶:۱۱۲). بر این اساس، پلتفرم‌هایی که از طریق طراحی الگوریتمی، به انتشار گسترده‌تر محتوای زیانبار کمک می‌کنند، ممکن است در معرض مسئولیت بیشتری قرار گیرند.

بند اول: تحلیل نقش سیاست‌های تعدیل محتوا

سیاست‌های تعدیل^۲ محتوا، یکی از ابزارهای اصلی پلتفرم‌ها برای مدیریت محتوای منتشر شده هستند. این سیاست‌ها شامل مجموعه‌ای از قواعد و رویه‌هاست که به منظور شناسایی، ارزیابی و حذف محتوای نامناسب یا غیرقانونی به کار گرفته می‌شوند (Gillespie, 2018). از منظر حقوقی، وجود چنین سیاست‌هایی می‌تواند دو اثر متفاوت داشته باشد:

نخست، اثر کاهشنده مسئولیت: در صورتی که پلتفرم بتواند نشان دهد که سازوکارهای معقول و مؤثری برای شناسایی و حذف محتوای زیانبار دارد، این امر می‌تواند به عنوان نشانه‌ای از فقدان تقصیر تلقی شود (Kuczerawy, 2022).

دوم، اثر تشدیدکننده مسئولیت: در مقابل، اگر ثابت شود که پلتفرم از وجود محتوای زیانبار آگاه بوده اما به‌رغم داشتن ابزارهای لازم، اقدامی انجام نداده، این امر می‌تواند دلالت بر تقصیر و حتی سوءنیت داشته باشد (Citron & Wittes, 2017).

در این زمینه، گلیسپی معتقد است که تعدیل محتوا، پلتفرم‌ها را از حالت بی‌طرفی خارج کرده و آنها را به تنظیم‌گران خصوصی فضای عمومی تبدیل می‌کند. این دیدگاه، پیامدهای مهمی برای تحلیل مسئولیت دارد، زیرا نشان می‌دهد که پلتفرم‌ها نه تنها نقش فنی، بلکه نقش هنجاری نیز ایفا می‌کنند (کاتوزیان، ۱۳۸۶: ۹۲).

مبحث سوم: مسئولیت مدنی و کیفری پلتفرم‌ها در حقوق ایران

تحلیل مسئولیت پلتفرم‌ها در نظام حقوقی ایران، مستلزم بررسی هم‌زمان دو حوزه مسئولیت مدنی و کیفری است. این دو حوزه، اگرچه از حیث مبانی و اهداف متفاوت هستند، اما در بسیاری از موارد، هم‌پوشانی دارند و می‌توانند به‌طور هم‌زمان اعمال شوند.

گفتار اول: مسئولیت مدنی پلتفرم‌ها

در حقوق ایران، مسئولیت مدنی بر مبنای ورود ضرر، وجود فعل زیانبار و رابطه سببیت میان آن دو استوار است. (کاتوزیان، ۱۳۸۶: ۳۵). مطابق ماده ۱ قانون مسئولیت مدنی:

¹ Van Eecke (2011)

² Content Moderation

هر کس بدون مجوز قانونی عمداً یا در نتیجه بی احتیاطی، به جان یا مال یا آزادی یا حیثیت یا شهرت تجارتي یا هر حق ديگر که به موجب قانون برای افراد ایجاد شده لطمه‌ای وارد نماید، مسئول جبران خسارت ناشی از عمل خود می‌باشد. با استناد به این ماده، می‌توان استدلال کرد که پلتفرم‌ها نیز در صورتی که از طریق فعل یا ترک فعل خود، موجب ورود ضرر شوند، مسئول خواهند بود. برای مثال، اگر پلتفرمی از انتشار یک تصویرسازی گمراه‌کننده که به حیثیت فردی لطمه زده آگاه باشد و با این حال از حذف آن خودداری کند، می‌توان رابطه سببیت را میان رفتار پلتفرم و ضرر وارده احراز کرد (Suzor, 2019).

در دکترین حقوقی، برخی نویسندگان مانند دکتر ناصر کاتوزیان، بر این باورند که در مواردی که چند عامل در ایجاد ضرر نقش دارند، باید به نقش مؤثرتر توجه کرد (کاتوزیان، ۱۴۰:۱۳۸۶). این دیدگاه می‌تواند در تحلیل مسئولیت پلتفرم‌ها نیز کاربرد داشته باشد، به ویژه در مواردی که تولیدکننده محتوا و پلتفرم، هر دو در ایجاد ضرر دخیل هستند.

بند اول: رابطه سببیت در فضای دیجیتال

یکی از چالش‌های اساسی در انتساب مسئولیت به پلتفرم‌ها، اثبات رابطه سببیت است. در فضای دیجیتال، انتشار محتوا معمولاً نتیجه تعامل چندین عامل است: تولیدکننده، پلتفرم، الگوریتم و حتی کاربران دیگر. این پیچیدگی، تعیین اینکه کدام عامل نقش اصلی در ورود ضرر داشته، را دشوار می‌سازد. در این زمینه، برخی حقوقدانان مانند گلد اسمیت و وو بر این باورند که باید به نقش تسهیل‌گر پلتفرم توجه کرد. به بیان دیگر، اگر پلتفرم شرایطی را فراهم کرده باشد که بدون آن، وقوع ضرر ممکن نبود، می‌توان رابطه سببیت را احراز کرد (Goldsmith & Wu, 2006).

گفتار دوم: مسئولیت کیفری پلتفرم‌ها

در حوزه مسئولیت کیفری، اصل شخصی بودن مسئولیت و ضرورت وجود عنصر معنوی (سوءنیت) از اهمیت ویژه‌ای برخوردار است (اردبیلی، ۱۳۹۹). در حقوق ایران، بر اساس اصول کلی، انتساب مسئولیت کیفری به اشخاص حقوقی، از جمله پلتفرم‌ها، در موارد خاص امکان‌پذیر است. بر اساس ماده ۱۴۳ قانون مجازات اسلامی: در جرائم ارتكابی توسط اشخاص حقوقی، هرگاه جرم در راستای منافع شخص حقوقی و به دستور یا با اطلاع مدیران آن ارتكاب یابد، شخص حقوقی نیز مسئول خواهد بود. با توجه به این ماده، در صورتی که پلتفرم به‌عنوان یک شخص حقوقی، از انتشار محتوای مجرمانه منتفع شود یا مدیران آن از وجود چنین محتوایی آگاه باشند و اقدامی نکنند، امکان انتساب مسئولیت کیفری وجود خواهد داشت.

مبحث چهارم: تحلیل تطبیقی مسئولیت پلتفرم‌ها در نظام‌های حقوقی منتخب

بررسی تطبیقی نشان می‌دهد که نظام‌های حقوقی مختلف، با توجه به ملاحظات خاص خود، رویکردهای متفاوتی در قبال مسئولیت پلتفرم‌ها اتخاذ کرده‌اند. این تفاوت‌ها عمدتاً ناشی از نحوه موازنه میان آزادی بیان، حمایت از حقوق اشخاص و ضرورت تنظیم‌گری فضای دیجیتال است. مطالعه این نظام‌ها می‌تواند در شناسایی نقاط قوت و ضعف حقوق ایران و ارائه راهکارهای اصلاحی مؤثر باشد. در اتحادیه اروپا، تنظیم‌گری پلتفرم‌ها به‌طور فزاینده‌ای به سمت مسئولیت فعال حرکت کرده است. مقررات جدید، به ویژه در قالب قانون خدمات دیجیتال، بر افزایش شفافیت، پاسخ‌گویی و تعهد پلتفرم‌ها به مدیریت ریسک‌های ناشی از انتشار محتوا تأکید دارد (Kuczerawy, 2022). در این نظام، پلتفرم‌ها در

صورت دریافت اخطار معتبر درباره محتوای غیرقانونی، موظف به حذف یا محدودسازی آن هستند. با این حال، تفاوت مهم رویکرد جدید با مدل‌های سنتی، در تأکید بر اقدامات پیشگیرانه است. به بیان دیگر، پلتفرم‌ها صرفاً پس از اخطار مسئول نیستند، بلکه باید با طراحی سازوکارهای داخلی، از گسترش محتوای زیانبار جلوگیری کنند. در حقوق آمریکا، اصل بر حمایت گسترده از پلتفرم‌ها در راستای تضمین آزادی بیان است. مهم‌ترین مبنای این رویکرد، بخش ۲۳۰ قانون ارتباطات^۱ است که پلتفرم‌ها را در برابر مسئولیت ناشی از محتوای تولیدشده توسط کاربران، تا حد زیادی مصون می‌دارد (Klonick, 2017). اگرچه این مصونیت نقش مهمی در رشد اقتصاد دیجیتال داشته، اما در سال‌های اخیر با انتقادات گسترده‌ای مواجه شده است. منتقدان بر این باورند که این رویکرد، موجب کاهش انگیزه پلتفرم‌ها برای کنترل محتوای زیانبار شده است. در مقابل، برخی دیگر از نظریه پردازان، این مصونیت را برای حفظ آزادی بیان ضروری می‌دانند و معتقدند که هرگونه محدودسازی آن می‌تواند به سانسور گسترده منجر شود. در نظام حقوقی فرانسه، که تأثیر زیادی بر حقوق ایران داشته، تلاش شده است تعادلی میان دو رویکرد فوق ایجاد شود. در این نظام، پلتفرم‌ها در صورتی که از محتوای غیرقانونی آگاه شوند و اقدام لازم را انجام ندهند، مسئول شناخته می‌شوند (Gillespie, 2018). این رویکرد رد، که مبتنی بر علم و عدم اقدام است، شباهت زیادی به تحلیل‌های قابل استنباط از حقوق ایران دارد و می‌تواند به عنوان الگویی مناسب برای توسعه مقررات داخلی مورد توجه قرار گیرد.

مبحث پنجم: چالش‌های نوین؛ هوش مصنوعی و تصویرسازی‌های پیشرفته (دیپ‌فیک)

ظهور فناوری‌های نوین، به‌ویژه هوش مصنوعی مولد، تحولی بنیادین در حوزه تولید محتوا ایجاد کرده است. این فناوری‌ها امکان خلق تصاویر کاملاً واقع‌نما اما جعلی را فراهم کرده‌اند که تشخیص آنها حتی برای کارشناسان نیز دشوار است (Chesney & Citron, 2019). دیپ‌فیک‌ها، به‌عنوان یکی از پیشرفته‌ترین اشکال تصویرسازی دیجیتال، می‌توانند در مواردی مصداق جرائم مختلفی باشند؛ از جمله:

- نشر اکاذیب
- هتک حیثیت
- نقض حریم خصوصی

در حقوق ایران، اگرچه عنوان خاصی برای دیپ‌فیک پیش‌بینی نشده، اما می‌توان با تفسیر موسع از قوانین موجود، این پدیده را ذیل مقررات مربوط به جرایم رایانه‌ای تحلیل کرد (اردبیلی، ۱۳۹۹).

بند اول: استناد به قانون جرایم رایانه‌ای

بر اساس ماده ۱۶ قانون جرایم رایانه‌ای: هر کس به‌وسیله سامانه‌های رایانه‌ای یا مخابراتی، فیلم یا صوت یا تصویر دیگری را تغییر دهد یا تحریف کند و آن را منتشر نماید به‌نحوی که عرفاً موجب هتک حیثیت او شود، به حبس یا جزای نقدی محکوم خواهد شد. این ماده، یکی از مهم‌ترین مبانی قانونی برای مقابله با تصویرسازی‌های جعلی است و

¹ Communications Decency Act

به‌طور خاص، می‌تواند شامل دیپ‌فیک‌ها نیز شود. همچنین ماده ۱۷ این قانون، انتشار تصاویر خصوصی بدون رضایت را جرم‌انگاری کرده است.

در این راستا یکی از چالش‌های اساسی، تعیین مسئولیت در مواردی است که محتوای زیانبار توسط الگوریتم‌ها یا ابزارهای هوش مصنوعی تولید شده است. در این موارد، پرسش این است که: آیا پلتفرم صرفاً ابزار است یا می‌توان آن را به‌عنوان یک عامل مسئول در نظر گرفت؟ (Gillespie, 2018) **برخی** اندیشمندان مانند بالکین^۱ معتقدند که در چنین شرایطی، باید مفهوم مسئولیت توزیعی را به کار گرفت؛ به این معنا که مسئولیت میان بازیگران مختلف (طراحان، پلتفرم‌ها و کاربران) تقسیم می‌شود (Balkin, 2018).

نتیجه‌گیری و پیشنهادها

تحلیل انجام‌شده در این پژوهش نشان می‌دهد که مسئولیت پلتفرم‌ها در قبال تصویرسازی‌های مجرمانه یا گمراه‌کننده، نه مطلق است و نه قابل نفی کامل؛ بلکه تابعی از مجموعه‌ای از معیارهای حقوقی و فنی است. مهم‌ترین این معیارها عبارتند از: علم و اطلاع، کنترل مؤثر، نقش فعال در انتشار محتوا و میزان انتفاع اقتصادی. در نظام حقوقی ایران، اگرچه قواعد کلی برای تحلیل این مسئولیت وجود دارد، اما فقدان مقررات خاص و شفاف در حوزه پلتفرم‌ها، موجب بروز ابهام و عدم انسجام در رویه قضایی شده است. این در حالی است که در نظام‌های حقوقی پیشرو، تلاش شده است با تدوین مقررات اختصاصی، این خلأها برطرف شود.

بر این اساس، پیشنهادهای زیر ارائه می‌شود:

۱. تدوین قانون جامع مسئولیت پلتفرم‌ها با الهام از تجربیات تطبیقی

۲. تصریح معیارهای مسئولیت (علم، کنترل، انتفاع) در قوانین داخلی

۳. الزام پلتفرم‌ها به شفافیت الگوریتمی

۴. تقویت سازوکارهای اخطار و حذف

۵. پیش‌بینی ضمانت‌اجراهای مؤثر در صورت نقض حقوق اشخاص

در نهایت، باید تأکید کرد که هرگونه تنظیم‌گری در این حوزه، باید با رعایت توازن میان سه ارزش اساسی صورت گیرد: **آزادی بیان، نوآوری فناوری و حمایت از حقوق اشخاص**. عدم توجه به هر یک از این عناصر، می‌تواند پیامدهای نامطلوبی برای نظام حقوقی و اجتماعی به همراه داشته باشد.

¹ Balkin (2015)

منابع

۱. منابع فارسی

کتاب‌ها

- اردبیلی، محمدعلی. (۱۳۸۵). *حقوق جزای عمومی*. تهران: میزان.
- کاتوزیان، ناصر. (۱۳۸۶). *مسئولیت مدنی*. تهران: شرکت سهامی انتشار.
- گلدوزیان، ایرج. (۱۳۹۴). *حقوق جزای اختصاصی*. تهران: مجد.
- آخوندی، محمود. (۱۳۹۱). *آیین دادرسی کیفری*. تهران: سمت.
- آشوری، محمد. (۱۳۹۷). *آیین دادرسی کیفری*. تهران: سمت.
- زراعت، عباس. (۱۳۹۸). *حقوق جزای عمومی*. تهران: ققنوس.
- رحمدل، محمد. (۱۳۹۸). *حقوق سایبری و جرایم رایانه‌ای*. تهران: جنگل.

مقالات

- فرح‌بخش، حسین. (۱۳۸۵). مفهوم و ابعاد مختلف حقوق دفاعی متهم. *مجله دانشکده حقوق و علوم سیاسی دانشگاه تهران*، ۳۶ (۷۴)، ۲۱-۴۵.
- حسینی، محمد. (۱۴۰۲). «چالش‌های حقوقی هوش مصنوعی و مسئولیت پلتفرم‌های دیجیتال». *فصلنامه مطالعات حقوق سایبری*، ۵ (۲)، ۱۰۵-۸۱.
- یوسفی، احسان. (۱۴۰۲). تحلیل تطبیقی مرحله تحقیقات مقدماتی در نظام حقوقی ایران و فرانسه. *فصلنامه مطالعات تطبیقی حقوق کیفری*، دانشگاه شهید بهشتی، ۱۲ (۴)، ۷۰-۴۵.
- نجفی ابرندآبادی، علی حسین. (۱۳۹۶). جرایم رایانه‌ای و چالش‌های نوین حقوق کیفری. *مجله تحقیقات حقوقی*.
- صادقی نشاط، امیر (۱۳۸۰)، «مصاحبه در خصوص چگونگی رسیدگی به جرایم ارتكابی در فضای سایبر». *نشریه شبکه*.

۲. منابع انگلیسی

Books.

- Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.
- Goldsmith, J., & Wu, T. (2006). *Who Controls the Internet? Illusions of a Borderless World*. Oxford University Press.
- Kuczerawy, A. (2022). *The Digital Services Act and Platform Liability*. Verfassungsblog.
- Lessig, L. (2006). *Code: Version 2.0*. Basic Books.
- Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Council of Europe Report.
- Suzor, N. (2019). *Lawless: The Secret Rules That Govern Our Digital Lives*. Cambridge University Press.
- Yadlin-Segal, A., & Oppenheim, Y. (2020). Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence*.
- Zittrain, J. (2008). *The Future of the Internet and How to Stop It*. Yale University Press.
- Kuner, C. (2017). *Transborder Data Flows and Data Privacy Law*. Oxford University Press.

Articles

- Balkin, J. M. (2015). Information fiduciaries and the First Amendment. *UC Davis Law Review*, 49, 1183-1234.
- Citron, D. K., & Chesney, R. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, 107(6), 1753-1820.
- Khan, F. (2024). Does the Digital Services Act achieve a balance between regulating deepfakes and protecting the fundamental right to freedom of expression? *SSRN Electronic Journal*.
- Leone, M. (2023). The Spiral of Digital Falsehood in Deepfakes. *International Journal for the Semiotics of Law*, 36, 385-405.
- Perritt, H. H. (1996). Tort liability and the Internet. *Harvard Journal of Law & Technology*, 9(2), 1-45.

- Van Eecke, P. (2011). Online service providers and liability: A European perspective. *Computer Law & Security Review*, 27(1), 5–18.
- Keller, D. (2019). Internet platforms and law. *Stanford Law Review*, 71(4), 1–36.
- Cochran, J. D., & Napshin, S. A. (2021). Deepfakes: Awareness, Concerns, and Platform Accountability. *Cyberpsychology, Behavior, and Social Networking*, 24(3), 164–172.
- Solove, D. (2007). The future of reputation. *Yale Law Journal*, 117, 116–170.
- Leigh Gray, T. "Minimum Conducts in Cyberspace: the Classic Jurisdiction Analysis in a New Setting". *Journal of High Technology Law*. (2002): 3, accessed December 17, 2008, <http://www.law.suffolk.edu/highlights/stuorgs/jht/docs/pdf/tgrayvinin.pdf>

Documents

- Universal Declaration of Human Rights (1948)
- International Covenant on Civil and Political Rights (1966)
- Directive 2000/31/EC (EU E-Commerce Directive)
- Digital Services Act (EU, 2022)
- Communications Decency Act, Section 230 (USA)