Comparative Studies in Jurisprudence, Law, and Politics

Preventive Criminal Policy and the Challenges of Attributing Criminal Liability in the Production and Dissemination of Sexual Deepfakes

1. Mahboube Sadat Razavi Nejad: PhD Student, Department of Law, Se.C., Islamic Azad University, Semnan, Iran
2. Hossein Haji Hosseini*: Department of Law, Se.C., Islamic Azad University, Semnan, Iran. Email: h.hosiny64@gmail.com (Corresponding Author)
3. Hossein Haji Hosseini*: Department of Law, Se.C., Islamic Azad University, Semnan, Iran

ABSTRACT

The emergence of novel technologies, particularly generative artificial intelligence and the phenomenon of "deepfakes," has created fundamental and unprecedented challenges for criminal justice systems. Among these challenges, the production and dissemination of fabricated pornographic content have increasingly targeted individuals' reputation, privacy, human dignity, and psychological security on a large scale. In confronting this phenomenon, the traditional architecture of criminal law and existing legislation, including the Computer Crimes Act of 2009, have become structurally inadequate due to their predominantly ex post approach and the absence of differentiated criminalization tailored to the specific characteristics of deepfake-related conduct. The present study seeks to identify legislative gaps and propose an appropriate regulatory and preventive model for addressing this phenomenon. Using a descriptive-analytical method and drawing upon library-based sources, the study examines the legal and criminological dimensions of sexual deepfakes. The findings indicate that, within the decentralized and networked environment of cyberspace, an exclusive judicial focus on punishing end users—whether producers or disseminators—is inconsistent with the logic of criminal justice economy, particularly given the extensive and diffuse nature of potential victimization. Overcoming this crisis requires a paradigm shift from a "repressive criminal policy" toward a "participatory and preventive regulatory model." Accordingly, criminal liability should be fairly and purposefully distributed across three levels: users, dissemination platforms, and technology developers. Abandoning a passive notice-based approach and imposing "enhanced corporate liability" on hosting platforms for proactive and algorithmic content filtering, together with establishing a regime of "liability arising from the design of high-risk digital products" for developers—including obligations to embed covert watermarking mechanisms and technological safeguards—constitute key strategies of situational and technical prevention. Ultimately, this article argues that effectively controlling deepfake-related offending requires the enactment of comprehensive artificial intelligence legislation, the adoption of active legal diplomacy at the international level, and the reciprocal use of defensive artificial intelligence in the scientific detection and investigation of crimes.

Keywords: *preventive criminal policy, deepfake, criminal liability, generative artificial intelligence, synthetic media, participatory regulation, situational prevention.*

How to cite: Razavi Nejad, M. S., Haji Hosseini, H., & Salehi, M. (2027). Preventive Criminal Policy and the Challenges of Attributing Criminal Liability in the Production and Dissemination of Sexual Deepfakes. *Comparative Studies in Jurisprudence, Law, and Politics, 9*(6), 1-17.

Submit Date: 26 May 2026
Revise Date: 14 September 2026
Accept Date: 23 September 2026
Initial Publish Date: 27 September 2026
Final Publish Date: 20 February 2028

پژوهش‌های تطبیقی فقه، حقوق و سیاست

سیاست جنایی پیشگیرانه و چالش‌های انتساب مسئولیت کیفری در تولید و انتشار دیپ‌فیک جنسی

۱. محبوبه سادات رضوی نژاد: دانشجوی دکتری، گروه حقوق، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران
۲. حسین حاجی حسینی*: گروه حقوق، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران. پست الکترونیک: h.hosiny64@gmail.com (نویسنده مسئول)
۳. محمود صالحی: گروه حقوق، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران

چکیده

ظهور فناوری‌های نوپدید، به ویژه هوش مصنوعی مولد و پدیده «جعل عمیق» چالش‌های بنیادین و بی‌سابقه‌ای را برای نظام‌های عدالت کیفری ایجاد کرده است. در این میان، تولید و انتشار محتوای مستهجن جعلی با هدف هتک حیثیت، حریم خصوصی، کرامت انسانی و امنیت روانی شهروندان را در مقیاسی گسترده هدف قرار داده است. در مواجهه با این پدیده، معماری سنتی حقوق کیفری و قوانین موجود (از جمله قانون جرایم رایانه‌ای مصوب ۱۳۸۸) به دلیل تمرکز بر رویکردهای پسینی و فقدان جرم‌انگاری افتراقی، دچار ناکارآمدی ساختاری شده‌اند. پژوهش حاضر با هدف تبیین خلأهای تقنینی و ارائه یک الگوی مطلوب مقابله‌ای، با استفاده از روش توصیفی-تحلیلی و بهره‌گیری از منابع کتابخانه‌ای به بررسی این موضوع پرداخته است. یافته‌های این تحقیق حاکی از آن است که در فضای شبکه‌ای و غیرمتمرکز سایبر، تمرکز انحصاری نظام قضایی بر مجازات کاربر نهایی (تولیدکننده یا منتشرکننده) به دلیل گستردگی آماج جرم، با منطق اقتصاد کیفری در تضاد است. غلبه بر این بحران نیازمند تغییر پارادایم از «سیاست جنایی سرکوب‌گر» به یک «الگوی تنظیم‌گری مشارکتی و پیشگیرانه» است. در این راستا، توزیع عادلانه و هدفمند مسئولیت کیفری در سه لایه کاربران، سکوهای انتشار (پلتفرم‌ها) و توسعه‌دهندگان فناوری ضرورت دارد. لغو رویکرد منفعلانه و مبتنی بر اخطار، و تحمیل «مسئولیت تشدیدیافته شرکتی» بر سکوهای میزبان جهت پالایش پیش‌دستانه و الگوریتمیک محتوا، در کنار اعمال رژیم «مسئولیت ناشی از طراحی محصولات پرخطر دیجیتال» برای توسعه‌دهندگان (الزام به تعبیه نشانه‌گذاری پنهان و سپرهای امنیتی)، از مهم‌ترین راهبردهای پیشگیری وضعی-فنی به شمار می‌روند. در نهایت، مقاله حاضر پیشنهاد می‌کند که مهار بزهکاری جعل عمیق تنها از طریق تدوین قانون جامع هوش مصنوعی، اتخاذ دیپلماسی فعال حقوقی در سطح بین‌المللی و استفاده متقابل از هوش مصنوعی تدافعی در فرآیند کشف علمی جرایم امکان‌پذیر خواهد بود.

واژگان کلیدی: *سیاست جنایی پیشگیرانه، جعل عمیق (دیپ‌فیک)، مسئولیت کیفری، هوش مصنوعی مولد، رسانه‌های ترکیبی، تنظیم‌گری مشارکتی، پیشگیری وضعی.*

نحوه استناددهی: رضوی نژاد، محبوبه سادات.، حاجی حسینی، حسین.، و صالحی، محمود. (۱۴۰۶). سیاست جنایی پیشگیرانه و چالش‌های انتساب مسئولیت کیفری در تولید و انتشار دیپ‌فیک جنسی. *پژوهش‌های تطبیقی فقه، حقوق و سیاست*، ۹(۶)، ۱۷-۱.

تاریخ ارسال: ۵ خرداد ۱۴۰۵
تاریخ بازنگری: ۲۳ شهریور ۱۴۰۵
تاریخ پذیرش: ۱ مهر ۱۴۰۵
تاریخ چاپ اولیه: ۵ مهر ۱۴۰۵
تاریخ چاپ نهایی: ۱ اسفند ۱۴۰۶

مقدمه

توسعه پرشتاب و تصاعدی هوش مصنوعی مولد در سال‌های اخیر، پارادایم‌های سنتی ارتباطات، تولید محتوا و حریم خصوصی را دچار دگرگونی بنیادین کرده است. در این میان، فناوری «جعل عمیق» یا دیپ‌فیک به‌عنوان یکی از پیچیده‌ترین دستاوردهای یادگیری ماشین، توانایی خلق محتوای دیداری و شنیداری فوق‌واقع‌گرایانه را به سطحی رسانده است که تمایز میان واقعیت و جعل برای چشم غیرمسلح تقریباً غیرممکن شده است (Westerlund, 2019).

اگرچه این فناوری در صنایع سرگرمی، پزشکی و آموزش کاربردهای مفیدی دارد، اما «دموکراتیزه شدن» و دسترسی آسان عمومی به آن از طریق برنامه‌های کاربردی متن‌باز و رابط‌های کاربری ساده، بستر مناسبی را برای سوءاستفاده‌های مجرمانه فراهم آورده است؛ به‌طوری‌که امروزه تولید محتوای مستهجن جعل به یکی از بحرانی‌ترین اشکال خشونت دیجیتال مبتنی بر جنسیت تبدیل شده است (Harris, 2019). مهم‌ترین چالش در مواجهه با این پدیده نوظهور، پدیده‌ای است که در جامعه‌شناسی حقوقی از آن تحت عنوان «مشکل گام‌برداری» یاد می‌شود؛ وضعیتی که در آن سرعت تکامل تکنولوژی به‌مراتب بیشتر از سرعت واکنش سیستم‌های قانون‌گذاری است (رابطه نامتقارن VTech≫VLaw). سیستم‌های حقوقی ذاتاً محافظه‌کار، گذشته‌نگر و نیازمند طی کردن فرایندهای طولانی تقنینی هستند، درحالی‌که الگوریتم‌های هوش مصنوعی با سرعتی نمایی خود را ارتقا می‌دهند (Floridi, 2018). این شکاف عمیق باعث شده است تا حقوق کیفری در برابر جرایم سایبری نوپدید، همواره در موضع انفعال و واکنش دیرهنگام قرار گیرد و نتواند حمایت‌های پیشگیرانه و کیفری مؤثری را در زمان طلایی به بزه‌دیدگان ارائه دهد (Elahimanesh, 2020).

در خصوص دیپ‌فیک جنسی، این کندی حقوق فاجعه‌بارتر است. زمانی که یک تصویر یا ویدئوی جعلی از یک فرد (غالباً زنان) در فضای سایبر منتشر می‌شود، به دلیل معماری غیرمتمرکز اینترنت و ویژگی تکثیرپذیری نامحدود داده‌ها، در کسری از ثانیه در سراسر جهان بازنشر می‌شود (Chesney & Citron, 2019). در این شرایط، حتی اگر قانون‌گذار پس از ماه‌ها بحث، عمل را جرم‌انگاری کرده و دستگاه قضایی مجرم را مجازات کند، آسیب واردشده به «حق بر تمامیت دیجیتال» و آبروی بزه‌دیده غیرقابل جبران خواهد بود.

ازاین‌رو، اکتفا به رویکرد سنتی و پسینی حقوق کیفری، عملاً به معنای رها کردن بزه‌دیده در برابر هجوم ویرانگر اطلاعات جعلی است (Niazpour, 2016). علاوه بر این، معماری فضای سایبر بر اساس تئوری انتقال فضایی جرایم، مختصاتی نظیر گمنامی، فقدان مرزهای فیزیکی و تأخیر در کشف را به مجرمان ارائه می‌دهد (Yar & Steinmetz, 2019).

مجرمان با استفاده از شبکه‌های خصوصی مجازی و دارک‌وب، هویت خود را پنهان کرده و از حاشیه امنی برای تولید و توزیع دیپ‌فیک‌های جنسی برخوردار می‌شوند (Jalali, 2022). در چنین محیطی، تکیه صرف بر سیاست جنایی سرکوب‌گر و مجازات‌محور کارایی خود را از دست می‌دهد، زیرا بازدارندگی کیفری مستلزم قطعیت در کشف جرم و حتمیت در اجرای مجازات است؛ مؤلفه‌هایی که در فضای بی‌کران و رمزنگاری‌شده اینترنت به‌شدت متزلزل شده‌اند (Yar & Steinmetz, 2019).

بنابراین، ضرورت دارد که نظام عدالت کیفری از یک رویکرد منفعلانه و جبرانی، به سمت یک «سیاست جنایی پیشگیرانه، کنش‌گر و توزیع‌شده» تغییر مسیر دهد. این سیاست جنایی نوین باید بتواند با بهره‌گیری از ظرفیت‌های خود فناوری (مانند استفاده از هوش مصنوعی برای مقابله با هوش مصنوعی)، اعمال الزامات فنی بر توسعه‌دهندگان (نظیر واترمارکینگ) و بازتعریف قواعد مسئولیت کیفری برای تمامی کنشگران فضای سایبر (از پلتفرم‌ها تا طراحان الگوریتم)، حلقه محاصره را بر تولیدکنندگان محتوای مجرمانه تنگ‌تر کند (Chesney & Citron, 2019).

مقاله حاضر بر آن است تا با واکاوی این شکاف تقنینی، الگوهای نوین سیاست جنایی پیشگیرانه و چالش‌های پیچیده انتساب مسئولیت در جرایم مرتبط با دیپ‌فیک جنسی را مورد تحلیل و بررسی قرار دهد.

پدیده دیپ‌فیک جنسی به‌عنوان یکی از مخرب‌ترین مصادیق سوءاستفاده از هوش مصنوعی، نظام‌های حقوقی معاصر را با بحرانی چندوجهی در حوزه جرم‌انگاری، پیشگیری و انتساب مسئولیت مواجه ساخته است. مسئله اصلی این پژوهش از آنجا نشئت می‌گیرد که مفاهیم و عناوین مجرمانه سنتی در قانون مجازات اسلامی و حتی قانون جرایم رایانه‌ای (نظیر جعل، هتک حیثیت، نشر اکاذیب و انتشار محتوای مستهجن)، ظرفیت و کشش لازم برای پوشش دادن ابعاد پیچیده و ماهیت الگوریتمی دیپ‌فیک را ندارند (Rostami, 2020). در جرایم سنتی، رابطه علیت و هویت مرتکب غالباً خطی و مشخص است، اما در فرایند تولید و انتشار دیپ‌فیک، با زنجیره‌ای پیچیده از بازیگران انسانی و غیرانسانی روبه‌رو هستیم که هر یک نقش متفاوتی در وقوع بزه ایفا می‌کنند و همین امر، دکترین سنتی «مسئولیت کیفری فردی» را با بن‌بست مواجه کرده است (Jalali, 2022).

در این اکوسیستم پیچیده، بحران انتساب مسئولیت کیفری خود را در سه سطح مجزا نشان می‌دهد: نخست، «کاربران» که به دو دسته تولیدکننده (با انگیزه انتقام، اخاذی یا سرگرمی) و منتشرکننده (بازنشرکنندگان در شبکه‌های اجتماعی) تقسیم می‌شوند. دوم، «پلتفرم‌های میزبان و واسطه‌های اینترنتی» که بستر انتشار و وایرال شدن این محتوا را فراهم می‌کنند و غالباً با استناد به اصل عدم مسئولیت واسطه‌ها (مانند سپر دفاعی بخش ۲۳۰ قانون نزاکت ارتباطات آمریکا)، از زیر بار مسئولیت شانه خالی می‌کنند. سوم و از همه چالش‌برانگیزتر، «توسعه‌دهندگان نرم‌افزار و مدل‌های هوش مصنوعی» هستند که ابزار ارتکاب جرم را به‌صورت متن‌باز و بدون تعبیه سپرهای ایمنی در اختیار عموم قرار می‌دهند (King et al., 2020). پرسش اساسی این است که آیا در نظام حقوقی کنونی می‌توان یک برنامه‌نویس یا شرکت فناوری را به دلیل فراهم کردن پلتفرمی که خروجی آن به‌طور بالقوه مجرمانه است، تحت عنوان معاونت در جرم یا تقصیر کیفری مسئول دانست؟

از سوی دیگر، انفعال سیاست جنایی در حوزه پیشگیری، مسئله را بغرنج‌تر کرده است. رویکرد فعلی قانون‌گذار عمدتاً بر «پیشگیری کیفری» استوار است که بر اساس فرمول بازدارندگی Deterrence=f(Severity,Certainty,Celerity) عمل می‌کند؛ اما در جرم دیپ‌فیک، به دلیل استفاده از فناوری‌های پنهان‌ساز هویت و فقدان مرزهای جغرافیایی، متغیرهای قطعیت و سرعت در مجازات به‌شدت میل به صفر دارند (Najafi Abrandabadi, 2019). در نتیجه، اتکای صرف به تهدید کیفری نمی‌تواند مانع از ارتکاب این جرایم شود. فقدان الزام قانونی برای شرکت‌های فناوری جهت اجرای تدابیر «پیشگیری وضعی» — مانند الزام به تعبیه اثر انگشت دیجیتال، واترمارک‌های پنهان، یا توسعه الگوریتم‌های فیلترینگ خودکار یک خلأ استراتژیک در سیاست جنایی تقنینی محسوب می‌شود (King et al., 2020).

علاوه بر چالش‌های داخلی، ماهیت فراملی فضای سایبر مسئله «صلاحیت و حاکمیت سایبری» را به‌شدت پررنگ می‌کند. یک فایل دیپ‌فیک جنسی ممکن است توسط کاربری در یک کشور تولید، در سرورهایی در کشور دوم میزبانی، و علیه بزه‌دیده‌ای در کشور سوم منتشر شود (Fazaeli, 2021). تضاد قوانین ملی، فقدان معاهدات جامع بین‌المللی در خصوص هوش مصنوعی و کندی فرایند معاضدت قضایی بین‌المللی، باعث شده است تا مرتکبان این جرایم از پناهگاه‌های امن سایبری برخوردار شوند و عملاً پیگرد قانونی آن‌ها ناممکن گردد. عدم وجود یک رژیم حقوقی یکپارچه جهانی و تشتت در رویه‌های تقنینی دولت‌ها، اثربخشی مبارزه با این پدیده را به حداقل رسانده است (Fazaeli, 2021).

با توجه به این مقدمات، مسئله بنیادین مقاله حاضر پاسخ به این چالش‌های درهم‌تنیده است. این پژوهش در پی آن است تا روشن سازد چگونه می‌توان با عبور از پارادایم سنتی حقوق کیفری، یک مدل «مسئولیت توزیع‌شده» را در نظام حقوقی پایه‌گذاری کرد که در آن بار پیشگیری و پاسخ‌گویی به‌صورت عادلانه میان تولیدکنندگان، پلتفرم‌ها و توسعه‌دهندگان تقسیم شود. همچنین، این مقاله واکاوی خواهد کرد که سیاست جنایی پیشگیرانه ایران نیازمند چه اصلاحات ساختاری و الزامات فنی–وضعی است تا بتواند همگام با سرعت فزاینده فناوری یادگیری عمیق، از حق بر تمامیت دیجیتال شهروندان محافظت نماید.

سبب‌شناسی جرم‌شناختی تولید دیپ‌فیک جنسی

برای اتخاذ یک سیاست جنایی پیشگیرانه کارآمد، پیش از هرگونه مداخله تقنینی یا فنی، درک چرایی و چگونگی وقوع بزه ضروری است. در جرم‌شناسی سایبری، وقوع جرایم پیچیده‌ای نظیر تولید و انتشار دیپ‌فیک جنسی، صرفاً ناشی از انحرافات فردی نیست، بلکه محصول تقاطع عواملی است که محیط دیجیتال در اختیار بزهکار قرار می‌دهد. بر اساس «نظریه فعالیت‌های روزمره»، وقوع جرم نیازمند تقارن سه عنصر بزهکار باانگیزه، آماج مناسب و فقدان مراقبین کارآمد است (Najafi Abrandabadi, 2019). در اکوسیستم دیپ‌فیک، تکنولوژی نه‌تنها آماج‌های جرم (تصاویر بزه‌دیدگان) را به‌وفور فراهم کرده، بلکه با توانمندسازی فنی افراد، خیل عظیمی از بزهکاران بالقوه را به عرصه کشانده است. در ادامه، دو عامل کلیدی در سبب‌شناسی این جرم بررسی می‌شود.

دسترسی‌پذیری ابزار (دموکراتیزه شدن هوش مصنوعی)

یکی از مهم‌ترین محرک‌های جرم‌شناختی در گسترش تصاعدی دیپ‌فیک‌های جنسی، پدیده‌ای است که در ادبیات فناوری از آن به‌عنوان «دموکراتیزه شدن هوش مصنوعی» یاد می‌شود. تا چند سال پیش، تولید یک ویدیوی جعل عمیق نیازمند دانش پیشرفته برنامه‌نویسی، دسترسی به ابررایانه‌ها و صرف صدها ساعت زمان برای آموزش مدل‌های یادگیری ماشین بود. این موانع فنی، به‌عنوان یک عامل بازدارنده طبیعی یا «سپر وضعی» عمل می‌کردند.

اما امروزه، با ظهور پلتفرم‌های متن‌باز، ربات‌های تلگرامی و نرم‌افزارهای کاربرپسند، هزینه و تلاش لازم برای ارتکاب این جرم به‌شدت کاهش یافته است. از منظر تئوری «انتخاب عقلانی»، زمانی که تلاش لازم برای انجام یک بزه کاهش یابد، تمایل به ارتکاب آن به‌صورت نمایی افزایش می‌یابد. اگر دسترسی‌پذیری را با متغیر A و تلاش لازم را با متغیر E نشان دهیم، رابطه آن‌ها به‌صورت معکوس $A \propto 1/E$ تعریف می‌شود. در حال حاضر، با وجود الگوریتم‌های آماده و سرویس‌های موسوم به «هوش مصنوعی به‌عنوان خدمت»، میزان تلاش به سمت صفر میل می‌کند $\lim_{E \to 0} A = \infty$. این سهولت دسترسی باعث شده است تا افراد فاقد هرگونه تخصص فنی تنها با فشردن چند دکمه و بارگذاری یک عکس از چهره بزه‌دیده، بتوانند محتوای مستهجن بسیار واقع‌گرایانه‌ای تولید کنند (Chesney & Citron, 2019). این دموکراتیزه شدن بی‌ضابطه، عملاً ابزار تولید سلاح‌های روانی–سایبری را در جیب هر شهروندی قرار داده است.

گمنامی و کاهش بازدارندگی در فضای سایبر

عامل دوم در گسترش این بزه، مختصات ذاتی فضای اینترنت و به‌طور خاص، ویژگی «گمنامی» است. بر اساس «نظریه انتقال فضایی»، افراد زمانی که از فضای فیزیکی به فضای سایبر منتقل می‌شوند، به دلیل پنهان ماندن هویتشان، تمایل بیشتری به رفتارهای ناهنجار و مجرمانه از خود نشان می‌دهند (Yar & Steinmetz, 2019).

فضای سایبر با ارائه ابزارهایی نظیر شبکه‌های خصوصی مجازی، شبکه‌های تاریک و حساب‌های کاربری جعلی، نقابی دیجیتال بر چهره بزهکار می‌کشد. این نقاب، یک اثر روان‌شناختی عمیق به نام «فردیت‌زدایی» ایجاد می‌کند که طی آن، فرد احساس مسئولیت اخلاقی و ترس از عقوبت را از دست می‌دهد. از منظر اقتصاد جرم، بازدارندگی کیفری (D) تابعی از قطعیت دستگیری (C)، شدت مجازات (S) و سرعت اعمال آن (T) است، که می‌توان آن را به‌صورت رابطه $D=f(C,S,1/T)$ مدل‌سازی کرد.

در جرم دیپ‌فیک، گمنامی باعث می‌شود متغیر «قطعیت کشف و دستگیری» (C) به‌شدت کاهش یافته و به صفر نزدیک شود ($C\approx0$). در نتیجه، مهم نیست قانون‌گذار چقدر شدت مجازات (S) را افزایش دهد؛ وقتی احتمال شناسایی و دستگیری مجرم ناچیز باشد، فرمول بازدارندگی عملاً فرو می‌پاشد و بازده انتظاررفته از جرم برای بزهکار همیشه مثبت خواهد بود. مجرم سایبری می‌داند که می‌تواند پشت لایه‌های رمزنگاری پنهان شود و یک ویدیوی مخرب را بدون اینکه ردپای فیزیکی از خود به جای بگذارد، در سطح جهانی منتشر کند. این گمنامی، دیوار دفاعی سیستم عدالت کیفری را دور زده و ضرورت اتخاذ راهکارهای پیشگیرانه وضعی–الگوریتمی را پیش از وقوع جرم، اجتناب‌ناپذیر می‌سازد (Yar & Steinmetz, 2019).

سیاست جنایی پیشگیرانه (وضعی و فنی)

با توجه به ناکارآمدی رویکردهای صرفاً کیفری و پسینی در فضای سایبر، سیاست جنایی نوین باید به سمت «پیشگیری وضعی» تغییر مسیر دهد. پیشگیری وضعی با تغییر محیط ارتکاب جرم، فرصت‌های مجرمانه را کاهش می‌دهد. در نظام حقوقی ایران، اگرچه فصل پیشگیری از جرایم رایانه‌ای در قانون جرایم رایانه‌ای (مصوب ۱۳۸۸) مورد توجه قرار گرفته، اما فقدان الزامات مشخص برای شرکت‌های فناوری به‌شدت احساس می‌شود. امروزه قانون‌گذاران پیشرو در جهان تلاش می‌کنند تا با وضع مقررات الزام‌آور، بار پیشگیری را بر دوش طراحان و پلتفرم‌ها قرار دهند (Citron, 2014).

نقش هوش مصنوعی در تشخیص و حذف خودکار محتوا

یکی از مؤثرترین راهکارهای پیشگیری وضعی، استفاده از خود فناوری برای مقابله با سوءاستفاده‌های فناورانه است؛ رویکردی که تحت عنوان تقابل AIDefense در برابر AIOffense شناخته می‌شود. از منظر حقوقی، پلتفرم‌های میزبان نقش کلیدی در توقف انتشار دیپ‌فیک‌های جنسی دارند. مطابق با ماده ۱۴ قانون جرایم رایانه‌ای ایران، ارائه‌دهندگان خدمات میزبانی در صورت اطلاع از وجود محتوای مجرمانه مکلف به حذف آن هستند. بااین‌حال، این ماده نیازمند یک رویکرد «کنش‌گرایانه» است.

در اسناد حقوقی جدید بین‌المللی مانند «قانون خدمات دیجیتال اتحادیه اروپا»، پلتفرم‌های بزرگ موظف شده‌اند تا با استفاده از الگوریتم‌های هوش مصنوعی، محتوای غیرقانونی را پیش از وایرال شدن تشخیص داده و مسدود کنند (Floridi, 2018). سیاست جنایی ایران نیز باید با اصلاح قوانین موجود، پلتفرم‌ها را ملزم سازد تا سیستم‌های فیلترینگ خودکار و تشخیص دیپ‌فیک را به‌عنوان یک «تکلیف قانونی و شرط تداوم فعالیت» پیاده‌سازی کنند تا از انتشار اولیه محتوای مستهجن جعلی جلوگیری شود.

الزامات فنی برای توسعه‌دهندگان (واترمارکینگ و اثرانگشت دیجیتال)

بخش عمده‌ای از مسئولیت پیشگیری باید متوجه توسعه‌دهندگان سیستم‌های هوش مصنوعی باشد. پیشگیری فنی ایجاب می‌کند که ابزارهای خلق محتوا از همان ابتدا با در نظر گرفتن اصل «امنیت در طراحی» کدنویسی شوند. یکی از راهکارهای الزام‌آور حقوقی، اجبار توسعه‌دهندگان به استفاده از تکنیک‌های «واترمارکینگ پنهان» و «اثرانگشت دیجیتال» است (Floridi, 2018).

از منظر حقوق تطبیقی، «قانون هوش مصنوعی اتحادیه اروپا» مصوب ۲۰۲۴، یک استاندارد طلایی تقنینی در این زمینه محسوب می‌شود. بر اساس این قانون، توسعه‌دهندگان هوش مصنوعی مولد قانوناً مکلف‌اند خروجی‌های سیستم خود را به‌گونه‌ای برچسب‌گذاری کنند که ماشینی بودن آن‌ها قابل ردیابی باشد. در سیاست جنایی ایران، جای خالی یک «قانون جامع هوش مصنوعی» مشهود است که بتواند شرکت‌های دانش‌بنیان و توسعه‌دهندگان نرم‌افزار را ملزم به تعبیه این سپرهای ایمنی نرم‌افزاری کند و در صورت تخلف، آن‌ها را با ضمانت‌اجراهای اداری یا حتی معاونت در بزه روبه‌رو سازد (Jalali, 2022).

پیشگیری از طریق آموزش و ارتقای سواد رسانه‌ای

در کنار پیشگیری‌های فنی و قانونی، بُعد اجتماعی سیاست جنایی پیشگیرانه بر توانمندسازی آماج‌های جرم استوار است. در دنیایی که تمایز میان واقعیت و جعل به حداقل رسیده است، «سواد رسانه‌ای انتقادی» به‌عنوان یک سپر دفاعی روان‌شناختی عمل می‌کند. قانون‌گذار می‌تواند از طریق الزام نهادهای آموزشی و رسانه ملی (در چارچوب قوانین بودجه و تکالیف دستگاه‌های فرهنگی)، برنامه‌های مدونی برای آگاهی‌بخشی عمومی اجرا کند (Jalali, 2022).

ارتقای سواد رسانه‌ای باعث می‌شود تا کاربران شبکه‌های اجتماعی با مشاهده یک ویدیوی جنجالی، بلافاصله آن را باور نکرده و از بازنشر آن خودداری کنند. این امر به‌طور مستقیم به قطع زنجیره انتشار کمک کرده و با کاهش تقاضا و توجه عمومی برای تماشای این محتوا، انگیزه مجرمان برای تولید دیپ‌فیک جنسی را کاهش می‌دهد.

چالش‌های انتساب مسئولیت کیفری

یکی از بنیادین‌ترین بحران‌هایی که ظهور هوش مصنوعی مولد و فناوری جعل عمیق برای نظام‌های عدالت کیفری ایجاد کرده است، فروپاشی دکترین سنتی «مسئولیت کیفری فردی» است. در حقوق کیفری کلاسیک، زنجیره علیت غالباً خطی، مستقیم و قابل درک است؛ فردی با سوءنیت مشخص، فعل مادی مجرمانه‌ای را مرتکب می‌شود و نتیجه زیان‌بار مستقیماً به او منتسب می‌گردد.

اما در اکوسیستم پیچیده جرایم سایبری مبتنی بر الگوریتم، ما با پدیده‌ای روبه‌رو هستیم که می‌توان آن را «مسئولیت توزیع‌شده» یا اشتراکی نامید. اگر کل ضرر واردشده به بزه‌دیده را با متغیر H (Harm) نشان دهیم، این ضرر تابعی از فعل چندین کنشگر مستقل است:

$$H=f(Adev,Aprod,Apub,Aplat)$$

در این معادله حقوقی، توسعه‌دهنده ابزار Adev، تولیدکننده محتوا Aprod، منتشرکننده Apub و پلتفرم میزبان Aplat، هر یک نقشی ایفا می‌کنند، اما انطباق فعل هر یک از آن‌ها با عناصر سه‌گانه جرم (قانونی، مادی و روانی) در چارچوب قوانین فعلی با چالش‌های جدی و تضادهای تفسیری مواجه است (Jalali, 2022).

مسئولیت کاربر تولیدکننده در مقابل کاربر منتشرکننده

در زنجیره ارتکاب جرم دیپ‌فیک جنسی، اولین حلقه مربوط به کاربران است که باید میان «تولیدکننده اولیه» و «منتشرکنندگان ثانویه» تفکیک قائل شد. کاربر تولیدکننده: از منظر حقوق کیفری ایران، عمل فردی که با استفاده از نرم‌افزارهای هوش مصنوعی، تصویر چهره یک شخص را بر روی بدن برهنه فرد دیگری در یک ویدیوی مستهجن قرار می‌دهد، انطباق نسبتاً روشنی با برخی عناوین مجرمانه دارد.

ماده ۱۶ قانون جرایم رایانه‌ای (موضوع ماده ۷۴۴ بخش تعزیرات قانون مجازات اسلامی) صراحتاً بیان می‌دارد: «هرکس به‌وسیله سیستم‌های رایانه‌ای یا مخابراتی، فیلم یا صوت یا تصویر دیگری را تغییر دهد یا تحریف کند و آن را منتشر یا با علم به تغییر یا تحریف منتشر کند، به نحوی که عرفاً موجب هتک حیثیت او شود...» مستوجب مجازات است.

همچنین در صورت صدق عنوان «محتوای مستهجن»، ماده ۱۴ همین قانون (ماده ۷۴۲ ق.م.ا) ناظر بر تولید و انتشار محتوای مستهجن قابل استناد است. در اینجا، عنصر مادی (تغییر و تحریف داده‌ها) و عنصر روانی (سوءنیت عام در استفاده از ابزار و سوءنیت خاص در قصد هتک حیثیت یا انتقام) محرز است و انتساب مسئولیت به تولیدکننده با مانع حقوقی چندانی مواجه نیست (Rostami, 2020).

کاربر منتشرکننده (بازنشرکننده): چالش اصلی در خصوص هزاران یا میلیون‌ها کاربری است که این محتوای مجرمانه را در شبکه‌های اجتماعی (نظیر تلگرام، ایکس یا اینستاگرام) بازنشر می‌کنند. آیا می‌توان تمام این افراد را به‌عنوان مباشر یا شریک در جرم هتک حیثیت یا اشاعه فحشا مجازات کرد؟ از یک سو، نص ماده ۱۴ قانون جرایم رایانه‌ای، عمل «ارسال» یا «انتشار» را مستقلاً جرم‌انگاری کرده است. اما از سوی دیگر، اثبات «علم و آگاهی» در میان توده کاربران بسیار دشوار است. بسیاری از منتشرکنندگان ممکن است اصلاً از جعلی بودن ویدیو مطلع نباشند و با تصور واقعی بودن آن، اقدام به بازنشر کنند (که در این صورت ممکن است جرم آن‌ها به نشر اکاذیب موضوع ماده ۱۸ تغییر یابد).

گستردگی بیش از حد آماج مجرمان و حجم عظیم منتشرکنندگان، اصل «فردی کردن مجازات‌ها» و «تناسب جرم و مجازات» را زیر سؤال می‌برد. برخورد کیفری با میلیون‌ها کاربر عملاً برای دستگاه قضایی غیرممکن است Noffenders→∞ و همین امر باعث شکل‌گیری نوعی «مصونیت در پناه جمعیت» برای منتشرکنندگان می‌شود که بازدارندگی قانون را خنثی می‌سازد (Najafi Abrandabadi, 2019).

چالش مسئولیت کیفری پلتفرم‌های میزبان و واسطه‌های اینترنتی

حلقه دوم در این اکوسیستم، پلتفرم‌های ارائه‌دهنده خدمات میزبانی و شبکه‌های اجتماعی هستند که زیرساخت لازم برای وایرال شدن و توزیع جهانی دیپ‌فیک‌ها را فراهم می‌کنند. رویکرد سنتی حقوق اینترنت در سطح بین‌المللی، مبتنی بر «اصل عدم مسئولیت واسطه‌ها» یا پناهگاه امن بوده است؛ مشابه آنچه در ماده ۲۳۰ قانون نزاکت ارتباطات ایالات متحده آمده است که پلتفرم‌ها را ناشر محتوای کاربران نمی‌داند.

در نظام حقوقی ایران، قانون‌گذار در ماده ۲۳ قانون جرایم رایانه‌ای (ماده ۷۵۱ ق.م.ا) مسئولیت ارائه‌دهندگان خدمات میزبانی را مشروط به «علم و آگاهی» کرده است. به این معنا که اگر پلتفرم از وجود محتوای مجرمانه مطلع شود (مثلاً از طریق دستور قضایی یا گزارش کاربران) و از حذف آن امتناع ورزد، از نظر کیفری مسئول خواهد بود.

چالش بنیادین این رویکرد در مواجهه با دیپ‌فیک جنسی، «مسئله زمان و سرعت تصاعدی انتشار» است. الگوریتم‌های شبکه‌های اجتماعی برای حداکثرسازی تعامل کاربران، محتواهای جنجالی و شوکه‌کننده را با سرعت نمایی ترویج می‌کنند. تا زمانی که بزه‌دیده متوجه انتشار ویدیو شود و فرایند کند و بوروکراتیک قضایی برای صدور دستور حذف طی شود، ویدیو میلیون‌ها بار دیده شده و در سرورهای مختلف در سراسر جهان ذخیره شده است. در اینجا، رویکرد منفعلانه «مسئولیت مبتنی بر اخطار»، عملاً دردی از بزه‌دیده دوا نمی‌کند. حقوق کیفری نوین نیازمند گذار از مسئولیت مبتنی بر تقصیر پسینی به سمت «مسئولیت مبتنی بر انفعال در پیشگیری فنی» است؛ یعنی پلتفرم‌ها باید به دلیل نقص در الگوریتم‌های نظارتی و عدم فیلترینگ پیش‌دستانه محتوای مستهجن جعلی، با جریمه‌های سنگین و مسئولیت کیفری شرکتی مواجه شوند (Yazdian Jafari, 2018).

آیا توسعه‌دهنده نرم‌افزار مسئول است؟ (تحلیل معاونت در جرم)

پیچیده‌ترین گره حقوقی در مبحث دیپ‌فیک، انتساب مسئولیت به برنامه‌نویسان، شرکت‌های فناوری و توسعه‌دهندگان مدل‌های متن‌باز هوش مصنوعی است که ابزار ارتکاب این جنایت دیجیتال را در اختیار عموم قرار می‌دهند. آیا می‌توان سازنده یک الگوریتم تعویض چهره را تحت عنوان «معاونت در جرم» تحت تعقیب قرار داد؟

برای پاسخ به این پرسش، باید به شرایط تحقق معاونت در ماده ۱۲۶ قانون مجازات اسلامی ایران رجوع کرد. تحقق معاونت نیازمند دو شرط اساسی است: نخست، «وحدت قصد» میان معاون و مباشر (الزام به وجود سوءنیت مشترک) و دوم، «تقدم یا اقتران زمانی» بین عمل معاون و مباشر. در تحلیل حقوقی ابزارهای هوش مصنوعی، دکترین «فناوری‌های دوگانه» مطرح می‌شود. بسیاری از مدل‌های یادگیری ماشین کاربردهای مشروع و قانونی فراوانی در صنعت سینما، آموزش و سرگرمی دارند. زمانی که یک شرکت، الگوریتمی را برای استفاده‌های عمومی توسعه می‌دهد و کاربری از آن سوءاستفاده کرده و دیپ‌فیک جنسی می‌سازد، شرط «وحدت قصد» مفقود است. برنامه‌نویس قصد اشاعه فحشا یا هتک حیثیت بزه‌دیده خاصی را نداشته است، بنابراین انتساب عنوان معاونت در جرم به وی از منظر حقوق کیفری کلاسیک مردود است (Jalali, 2022).

بااین‌وجود، استثنائاتی وجود دارد. اگر نرم‌افزاری ذاتاً و منحصراً برای اهداف مجرمانه طراحی شده باشد (مانند برنامه‌های موسوم به دیپ‌نود که صراحتاً برای برهنه‌سازی تصاویر زنان تبلیغ و فروخته می‌شوند)، در اینجا سوءنیت توسعه‌دهنده با نتیجه مجرمانه هم‌راستا است و تسهیل وقوع جرم (موضوع بند پ ماده ۱۲۶) محرز می‌گردد؛ لذا می‌توان توسعه‌دهنده را به‌عنوان معاون جرم مجازات کرد.

اما خلأ قانونی بزرگ در جایی است که توسعه‌دهندگانِ مدل‌های عمومی، با بی‌مبالاتی و بدون تعبیه پروتکل‌های ایمنی، الگوریتم‌های قدرتمند خود را به‌صورت متن‌باز منتشر می‌کنند. حقوق کیفری فعلی ظرفیت مجازات افراد به دلیل «بی‌احتیاطی و بی‌مبالاتی در طراحی کد» را در جرایم عمدی ندارد. ازاین‌رو، ضرورت دارد قانون‌گذار با الهام از مقررات نوین بین‌المللی، یک رژیم حقوقی مستقل برای «مسئولیت مدنی و کیفری ناشی از تولید محصولات فناورانه پرخطر» ایجاد کند تا توسعه‌دهندگان مجبور شوند پیش از عرضه عمومی الگوریتم‌های خود، ارزیابی‌های دقیق خطرات سایبری را انجام دهند (King et al., 2020).

همکاری‌های بین‌المللی و حاکمیت سایبری در مهار دیپ‌فیک

ماهیت فضای سایبر، مرزهای جغرافیایی و فیزیکی را بی‌معنا ساخته است. یک ویدیوی دیپ‌فیک جنسی می‌تواند توسط فردی در قاره آسیا تولید، بر روی سرورهایی در اروپا میزبانی و در شبکه‌های اجتماعی مستقر در آمریکا منتشر شود، درحالی‌که بزه‌دیده در کشور دیگری سکونت دارد. این پراکندگی جغرافیایی، مفهوم سنتی «حاکمیت ملی» را با چالشی بنیادین مواجه می‌کند و نیازمند سازوکارهای نوین بین‌المللی است. اگر سرعت انتشار محتوا را با متغیر Vspread و سرعت پیگیری‌های حقوقی بین‌المللی را با Vlegal نشان دهیم، در حال حاضر همواره با نامعادله قطعی Vspread≫Vlegal روبه‌رو هستیم که منجر به شکست تلاش‌های محلی در مهار این جرایم می‌شود (Fazaeli, 2021).

تقابل حاکمیت سایبری و ماهیت فرامرزی پلتفرم‌ها

کشورها در راستای حفظ حاکمیت سایبری خود تلاش می‌کنند قوانین داخلی سفت‌وسختی برای کنترل محتوا وضع کنند. بااین‌حال، پلتفرم‌های بزرگ فناوری غالباً تابع قوانین کشور متبوع خود هستند. به‌عنوان مثال، زمانی که یک مقام قضایی در ایران دستور حذف یک دیپ‌فیک جنسی را صادر می‌کند، الزام پلتفرم‌های خارجی به اجرای فوری این دستور، به دلیل فقدان دفاتر نمایندگی رسمی و تفاوت در رژیم‌های حقوقی، عملاً با بن‌بست مواجه می‌شود. غلبه بر این چالش نیازمند گذار از حاکمیت محلی به سمت «حاکمیت سایبری اشتراکی» است که در آن پلتفرم‌ها ملزم به پذیرش حداقل استانداردهای اخلاقی و حقوقی بین‌المللی در حوزه کرامت انسانی باشند.

ضرورت توسعه معاهدات معاضدت قضایی متقابل

در غیاب یک پلیس بین‌الملل سایبری با اختیارات اجرایی مستقیم، پیگرد مجرمان سایبری فراملی متکی به «معاهدات معاضدت قضایی متقابل» است. بااین‌حال، رویه‌های سنتی بسیار کند و بوروکراتیک هستند و ماه‌ها به طول می‌انجامند. در جرایم مرتبط با دیپ‌فیک که ثانیه‌ها در گسترش آسیب نقش دارند، جامعه جهانی نیازمند پروتکل‌های اورژانسی برای مسدودسازی آنی محتوا و تبادل سریع ادله الکترونیکی میان ضابطان قضایی کشورها است (Fazaeli, 2021). پیوستن به اسناد بین‌المللی نظیر کنوانسیون جرایم سایبری (بوداپست) یا مشارکت فعال در تدوین کنوانسیون‌های جدید سازمان ملل در حوزه جرایم سایبری، پیش‌نیاز این همکاری است.

چالش صلاحیت محلی در جرایم فراملی دیپ‌فیک

از منظر آیین دادرسی کیفری، نخستین گام در رسیدگی به یک جرم، احراز «صلاحیت دادگاه» است. در جرایم فیزیکی، محل وقوع جرم تکلیف صلاحیت را روشن می‌کند، اما در جرایم سایبری، مکان وقوع جرم مفهومی سیال و چندگانه دارد. پیچیدگی این امر را می‌توان به‌صورت تابعی از متغیرهای مختلف زیر نشان داد که در آن صلاحیت (J) به مکان بزهکار، مکان سرور، مکان بزه‌دیده و تابعیت طرفین بستگی دارد.

$$J=f(Lperpetrator,Lserver,Lvictim,Nparties)$$

تنوع اصول صلاحیت کیفری در فضای سایبر

قانون مجازات اسلامی ایران (مصوب ۱۳۹۲) در مواد ۳ تا ۸، اصول مختلفی را برای اعمال صلاحیت در نظر گرفته است:

☐ صلاحیت سرزمینی (مکان‌محور): بر اساس ماده ۴، اگر قسمتی از جرم در ایران واقع شود یا «نتیجه» آن در ایران حاصل گردد، دادگاه‌های ایران صلاحیت دارند. در مورد دیپ‌فیک، از آنجا که هتک حیثیت و آسیب روانی به بزه‌دیده (نتیجه جرم) در محل سکونت او رخ می‌دهد، محاکم محلی صلاحیت رسیدگی دارند، حتی اگر سرور و مجرم در خارج از کشور باشند.

☐ صلاحیت شخصی (مبتنی بر تابعیت): اگر سازنده دیپ‌فیک در خارج از کشور ایرانی باشد (صلاحیت شخصی فعال – ماده ۷) و یا بزه‌دیده ایرانی باشد (صلاحیت شخصی منفعل ماده ۸)، قانون ایران تحت شرایطی قابل اعمال است.

با وجود این گستردگی در ادعای صلاحیت تقنینی، اعمال «صلاحیت اجرایی» (دسترسی فیزیکی به متهم برای محاکمه) چالش اصلی است (Fazaeli, 2021).

معضل استرداد مجرمین و شرط مجرمیت متقابل

حتی اگر دادگاه محلی صلاحیت خود را احراز کرده و مجرم سایبری را در کشور دیگری شناسایی کند، دستگیری و محاکمه وی منوط به فرایند «استرداد مجرمین» است. یکی از شروط بنیادین در حقوق بین‌الملل عرفی برای استرداد، شرط «مجرمیت متقابل» است؛ یعنی عمل ارتکابی باید هم در کشور متقاضی (ایران) و هم در کشور متوقف‌فیه (کشور میزبان مجرم) جرم‌انگاری شده باشد.

چالش دیپ‌فیک در اینجا خودنمایی می‌کند: درحالی‌که در ایران تولید این محتوا به‌وضوح تحت عناوین مستهجن و هتک حیثیت جرم است، ممکن است در برخی کشورها (به‌ویژه کشورهایی با تفسیر موسع از آزادی بیان)، تولید دیپ‌فیک تا زمانی که به کلاهبرداری مالی یا پورنوگرافی کودکان ختم نشود، صرفاً یک تخلف مدنی یا نقض کپی‌رایت محسوب شود. فقدان هماهنگی در جرم‌انگاری جهانی، باعث می‌شود شرط مجرمیت متقابل محقق نشده و مجرمان دیپ‌فیک در بهشت‌های امن سایبری پناه گیرند (Fazaeli, 2021).

نتیجه‌گیری و ارائه الگوی مطلوب سیاست جنایی

ظهور فناوری هوش مصنوعی مولد و پدیده جعل عمیق، به‌ویژه در قالب تولید محتوای مستهجن با هدف هتک حیثیت، نمایانگر یکی از پیچیده‌ترین بحران‌های تقنینی در تاریخ حقوق کیفری مدرن است. این پژوهش نشان داد که معماری سنتی حقوق کیفری که بر پایه‌های مفاهیمی چون مسئولیت فردی، رابطه علیت خطی و تقصیر بنا شده است، در مواجهه با ماهیت غیرمتمرکز، تصاعدی و الگوریتمی جرایم سایبری نوین دچار ناکارآمدی ساختاری شده است. شکاف عمیق میان سرعت سرسام‌آور تکامل فناوری و کندی مفرط در تطبیق قوانین کیفری، به‌وضوح خودنمایی می‌کند؛ شکافی که نتیجه قطعی آن، شکل‌گیری حاشیه امن برای بزهکاران و بی‌دفاع ماندن بزه‌دیدگان در فضای مجازی است.

نخستین و مهم‌ترین چالش استنتاج‌شده در این پژوهش، فقدان جرم‌انگاری افتراقی و صریح در قوانین موضوعه ایران است. قانون جرایم رایانه‌ای مصوب سال یکهزار و سیصد و هشتاد و هشت، با وجود پیش‌بینی موادی نظیر ماده شانزده در خصوص تغییر و تحریف داده‌ها و ماده هفده با محوریت هتک حیثیت، متناسب با مقتضیات دهه‌های پیشین تدوین شده است و ظرفیت پاسخ‌گویی به پیچیدگی‌های رسانه‌های ترکیبی و تولیدات ماشینی را ندارد. محتوای تولیدشده توسط هوش مصنوعی، صرفاً تغییر یک تصویر از پیش موجود نیست، بلکه خلق یک واقعیت مجازی از پایه است. ازاین‌رو، قانون‌گذار به‌شدت نیازمند وضع مقرراتی است که نفسِ «تولید محتوای مستهجن جعلی با استفاده از هوش مصنوعی»، مستقل از قصد انتشار یا تحقق نتیجه مجرمانه، به‌عنوان یک بزه مستقل و از باب جرایم مانع شناسایی شود.

در حوزه انتساب مسئولیت کیفری، نظام حقوقی فعلی در دام فردگرایی کلاسیک گرفتار مانده است. توزیع ضرر در فضای شبکه‌ای نیازمند تحلیلی فراتر از کنش یک فرد واحد است. در این زنجیره پیچیده، تمرکز صرف نظام قضایی بر مجازات کاربر نهایی اعم از تولیدکننده یا منتشرکننده، به معنای نادیده گرفتن نقش بنیادین سکوهای انتشار و توسعه‌دهندگان فناوری است. اعمال مجازات بر میلیون‌ها کاربری که در بازنشر یک ویدیوی جعل عمیق مشارکت دارند، عملاً اصل فردی کردن مجازات‌ها و منطق اقتصاد کیفری را مختل می‌سازد. قانون باید به‌جای تمرکز انحصاری بر انتهای زنجیره بزهکاری، رویکرد خود را به سمت حلقه‌های اولیه، ثروت‌آفرین و قدرتمندتر معطوف نماید.

در خصوص سکوهای ارائه‌دهنده خدمات میزبان یا همان شبکه‌های اجتماعی، رویکرد منفعلانه «مسئولیت مبتنی بر اخطار» که در قانون جرایم رایانه‌ای و ماده هفتصد و پنجاه و یک قانون مجازات اسلامی متبلور است، در برابر سرعت نمایی انتشار محتوای جعلی کاملاً فلج و ناکارآمد است. قانون‌گذار باید با الگوبرداری از اسناد نوین بین‌المللی، اصل حاشیه امن نرم‌افزاری را لغو کرده و مدیران این سکوها را قانوناً مکلف به اجرای پالایش پیش‌دستانه و خودکار نماید. در این ساختار تقنینی جدید، انفعال سکوهای مجازی در مسدودسازی فوری جعل عمیق جنسی، باید مستوجب مسئولیت کیفری شخص حقوقی و جرایم نقدی بسیار سنگین بر مبنای درصدی از درآمد سالانه آن‌ها باشد تا بازدارندگی اقتصادی محقق گردد.

مسئولیت توسعه‌دهندگان نرم‌افزار، حلقه مفقوده و به‌شدت چالش‌برانگیز دیگر در نظام قانونی ماست. با توجه به دکترین حقوقی ناظر بر فناوری‌های دارای کاربرد دوگانه، امکان انتساب عنوان معاونت در جرم، موضوع ماده صد و بیست و شش قانون مجازات اسلامی، به برنامه‌نویسانِ مدل‌های متن‌باز به دلیل فقدان وحدت قصد میان سازنده نرم‌افزار و کاربر بزهکار، از منظر تحلیلی بسیار دشوار است. بنابراین، نهاد قانون‌گذار ملزم است رژیم حقوقی کاملاً جدیدی تحت عنوان «مسئولیت ناشی از طراحی محصولات پرخطر دیجیتال» پایه‌گذاری کند.

بر اساس این رژیم، الزام به رعایت اصل امنیت در مرحله طراحی و تعبیه اجباری سازوکارهای نشانه‌گذاری پنهان دیجیتال، باید به یک تکلیف قانونی غیرقابل چشم‌پوشی و شرط اصلی دریافت مجوز فعالیت برای شرکت‌های دانش‌بنیان تبدیل شود.

از منظر آیین دادرسی کیفری و ادله اثبات دعوا، فناوری جعل عمیق اعتبار سنتی ادله الکترونیکی را در محاکم دادگستری به‌شدت متزلزل کرده است. هنگامی که ماشین قادر است مستنداتی دیداری و شنیداری تولید کند که حتی برای کارشناسان خبره انسانی غیرقابل تشخیص باشد، قضات در استناد به فیلم‌ها و تصاویر دچار تردیدهای جدی و شبهه در ادله می‌شوند که این امر به‌سادگی می‌تواند قاعده درأ را به نفع متهم فعال سازد. لذا، اصلاح بنیادین فصل ادله الکترونیکی در قانون آیین دادرسی کیفری و الزام قانونی محاکم به استفاده از روش‌های نوین جرم‌یابی دیجیتال و بهره‌گیری از ابزارهای اعتبارسنجی ماشینی در فرایند کشف حقیقت، امری حیاتی و غیرقابل اجتناب است.

در سطح فراملی نیز چالش صلاحیت محلی و اصل مجرمیت متقابل در فرایند استرداد مجرمین، اثربخشی قوانین داخلی را به‌شدت تقلیل می‌دهد. مواد سه الی هشت قانون مجازات اسلامی، اگرچه صلاحیت گسترده‌ای برای محاکم ایران در رسیدگی به جرایم سایبری در نظر گرفته‌اند، اما در فاز اجرایی و بدون همکاری سکوهای بین‌المللی عملاً عقیم می‌مانند. سیاست جنایی ایران باید از طریق فعال‌سازی دیپلماسی حقوقی، پیوستن به معاهدات نوین معاضدت قضایی در فضای مجازی و تدوین تشریفات فوری برای مسدودسازی محتوای مستهجن در سطح جهانی را با جدیت در مجامع بین‌المللی پیگیری کند.

با عنایت به تمامی چالش‌های مطروحه در این مقال، الگوی مطلوب سیاست جنایی در تقابل با بزهکاری جعل عمیق، باید از مدل سنتی و ناکارآمدِ سیاست جنایی سرکوب‌گر و پسینی، به مدل تنظیم‌گری مشارکتی و پیشگیرانه تغییر ماهیت دهد. در این الگوی مطلوب، بار اصلی کنترل جرم نباید منحصراً بر دوش سیستم قضایی، دادسراها و نهادهای انتظامی باشد، بلکه این مسئولیت خطیر باید به‌صورت تکالیف قانونی سفت‌وسخت بر عهده بخش خصوصی و شرکت‌های بزرگ فناوری توزیع شود و ضمانت‌اجراهای تخلف از این تکالیف نیز به‌روشنی در قانون تصریح گردد.

این الگوی مطلوب تقنینی بر سه پایه حقوقی استوار است: پایه نخست، «جرم‌انگاری پیش‌دستانه» است که تولید نرم‌افزارهای فاقد سپرهای ایمنی را مستقلاً جرم‌انگاری می‌کند؛ پایه دوم، «مسئولیت تشدیدیافته شرکتی» است که سکوهای انتشار را به لحاظ قانونی ضامن امنیت روانی کاربران در برابر محتوای جعلی می‌داند؛ و پایه سوم، «پیشگیری وضعی و فنی اجباری» است که استفاده از ابزارهای رمزنگاری و تشخیص ماشین‌پایه را در تمامی زیرساخت‌های ارتباطی کشور الزامی می‌سازد. تنها در سایه چنین معماری حقوقی جامع‌نگری است که می‌توان از حقوق بنیادین شهروندان، به‌ویژه کرامت انسانی و حریم خصوصی اشخاص، در عصر حکمرانی الگوریتم‌ها دفاع نمود.

در نهایت، چنین استنتاج می‌شود که قانون کیفری با تکیه بر ابزارهای سنتی نظیر مجازات حبس و جزای نقدی خرد، به هیچ وجه یارای مقابله با پدیده مخرب جعل عمیق را ندارد. حقوق کیفری فردا، حقوقی است که کدهای برنامه‌نویسی را به‌عنوان بخشی از هنجارهای قانونی به رسمیت بشناسد و مجازات را نه صرفاً بر جسم بزهکار خرد، بلکه بر ساختار اقتصادی و الگوریتمی شرکت‌های متخلف اعمال نماید. تدوین یک قانون جامع، مستقل و روزآمد با محوریت هوش مصنوعی و رسانه‌های ترکیبی که واجد ضمانت‌اجراهای فنی، مدنی و کیفریِ درهم‌تنیده باشد، فوری‌ترین و حیاتی‌ترین گام نهاد قانون‌گذاری کشور در مسیر تحقق این الگوی مطلوب و صیانت از امنیت اخلاقی و روانی جامعه خواهد بود.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافعی وجود ندارد.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

حامی مالی

این پژوهش حامی مالی نداشته است.

EXTENDED SUMMARY

The rapid and exponential development of generative artificial intelligence has fundamentally transformed conventional paradigms of communication, content production, privacy, and the evidentiary reliability of digital media. Among the most disruptive manifestations of this transformation is deepfake technology, which enables the production of highly realistic synthetic audiovisual content through advanced machine-learning techniques. The progressive accessibility of such technologies has significantly reduced the technical, financial, and computational barriers that previously limited their use to highly specialized actors. Consequently, deepfake tools have increasingly become available through open-source software, user-friendly interfaces, mobile applications, and automated online services. Although deepfake technology has legitimate applications in entertainment, education, medicine, accessibility, and creative industries, its misuse has generated substantial criminal-law and criminological concerns, particularly in relation to the production and dissemination of non-consensual sexual and pornographic content. Sexual deepfakes may seriously undermine reputation, privacy, human dignity, psychological security, and personal autonomy, while their rapid circulation through social media platforms can exponentially intensify victimization (Westerlund, 2019). The phenomenon has therefore become an important manifestation of technology-facilitated abuse in contemporary cyberspace (Harris, 2019). The principal difficulty for criminal justice systems arises from the structural asymmetry between the speed of technological innovation and the comparatively slow process of legal adaptation. Criminal legislation is generally reactive, institutionally cautious, and dependent on formal legislative procedures, whereas artificial intelligence technologies evolve through rapid and continuous technical development. This regulatory gap places criminal law in a persistent position of delayed intervention and weakens its capacity to provide timely preventive protection against emerging forms of cybercrime (Rostami, 2020). In the specific context of sexual deepfakes, delayed intervention is particularly problematic because digital content can be copied, downloaded, modified, and redistributed across multiple platforms within seconds, making complete removal practically impossible once dissemination has begun (Chesney & Citron, 2019). Accordingly, reliance exclusively on conventional punitive mechanisms is insufficient, and criminal policy must increasingly incorporate preventive, regulatory, technological, and participatory mechanisms capable of intervening before irreversible digital harm occurs (Elahimanesh, 2020).

The central problem addressed in this study concerns the inadequacy of traditional criminal-law concepts for responding to the complex and distributed architecture of sexual deepfake production and dissemination. Existing criminal offences such as defamation, dissemination of obscene content, falsification, publication of false information, and violation of privacy may partially cover particular manifestations of deepfake-related conduct, but they were generally formulated before the emergence of contemporary generative artificial intelligence and therefore do not fully reflect the algorithmic and synthetic nature of these offences (Rostami, 2020). This difficulty is compounded by the multiplicity of actors involved in the production and circulation of deepfake content. Unlike

conventional offences in which the relationship between the perpetrator, criminal act, and harmful result is often relatively direct, deepfake-related harm may arise from the combined conduct of software developers, content producers, secondary disseminators, hosting platforms, social media companies, and algorithmic recommendation systems. This fragmented structure challenges the classical doctrine of individualized criminal responsibility, which traditionally assumes a clearly identifiable human offender whose intentional conduct directly produces the prohibited result (Jalali, 2022). The study therefore examines whether criminal responsibility should remain concentrated on the end user or whether a broader model of distributed responsibility should be developed. In such a model, responsibility would be differentiated according to the functional role, degree of control, knowledge, capacity for prevention, and contribution of each actor within the technological ecosystem. Particular attention is paid to three principal categories: users who initially create or intentionally redistribute sexual deepfakes; intermediary platforms that facilitate rapid dissemination and algorithmic amplification; and developers who design, release, or commercialize artificial intelligence systems capable of being used for harmful synthetic content. The study further considers how the safe-harbor approach traditionally applied to internet intermediaries may become increasingly difficult to justify when platforms possess sophisticated automated detection capabilities and exercise substantial algorithmic control over the visibility and distribution of user-generated content (Chesney & Citron, 2019). The broader objective is therefore to conceptualize a preventive criminal-policy framework in which responsibility is allocated proportionately across the digital ecosystem rather than being imposed exclusively on individual users after harmful content has already circulated widely.

From a criminological perspective, the expansion of sexual deepfake offending can be explained through the interaction of technological accessibility, suitable targets, anonymity, weak guardianship, and declining perceived risks of detection. The development of user-friendly artificial intelligence tools has effectively democratized capabilities that previously required advanced programming knowledge, expensive computing infrastructure, and extensive technical expertise. This reduction in the effort required to produce manipulated audiovisual content substantially expands the pool of potential offenders. Under a rational-choice perspective, when the practical costs and technical obstacles associated with offending decrease while perceived benefits remain stable or increase, criminal opportunities become more attractive. Deepfake generators, automated face-swapping applications, downloadable models, and artificial-intelligence-as-a-service platforms have significantly lowered the effort necessary to create realistic synthetic sexual material (King et al., 2020). At the same time, the architecture of cyberspace provides offenders with anonymity-enhancing mechanisms, including virtual private networks, encrypted communication channels, anonymous accounts, offshore hosting services, and dark-web infrastructures. These characteristics weaken the probability of identification and therefore reduce the deterrent effect of criminal punishment (Yar & Steinmetz, 2019). Deterrence depends not merely on the severity of punishment, but also on the perceived certainty and celerity of detection and sanction. When offenders reasonably believe that attribution is difficult and transnational investigation is slow, increasing statutory penalties alone may have limited preventive value (Najafi Abrandabadi, 2019). This criminological analysis demonstrates why a preventive criminal policy should focus on modifying the opportunity structure of offending rather than relying exclusively on punishment after the offence has occurred. Situational prevention can reduce criminal opportunities by increasing technical barriers, strengthening guardianship, limiting anonymous abuse, and introducing automated detection mechanisms. Such an approach is particularly relevant to sexual deepfakes because the principal

harm often arises from the speed and scale of dissemination rather than merely from the initial act of production. Accordingly, preventive measures must operate at the technological and infrastructural levels, where harmful content can be identified and interrupted before large-scale circulation occurs.

A central component of the proposed preventive framework is the development of technologically informed obligations for digital platforms and artificial intelligence developers. Hosting services and social media platforms occupy a strategic position in the deepfake ecosystem because they control the infrastructure through which synthetic content is uploaded, recommended, replicated, and made visible to large audiences. Traditional intermediary-liability regimes have frequently been based on a reactive notice-and-takedown model, under which a platform becomes responsible only after obtaining knowledge of unlawful content and failing to remove it. However, this approach is poorly adapted to sexual deepfakes because harmful material may achieve viral distribution before the victim can identify the content, submit a complaint, or obtain a judicial removal order. A preventive model therefore requires a transition from passive notice-based liability toward enhanced duties of algorithmic monitoring, rapid detection, and risk-based content governance. Such obligations should be proportionate to the size, technological capacity, and systemic influence of the platform. The use of artificial intelligence for detecting synthetic media can form part of this preventive architecture, allowing platforms to employ defensive AI against malicious AI-generated content (Miró-Llinares, 2020). In parallel, developers of generative artificial intelligence systems should be subject to security-by-design obligations intended to reduce foreseeable misuse. Such obligations may include embedded watermarking, digital fingerprinting, traceability mechanisms, abuse-reporting systems, restrictions on high-risk functions, and technical safeguards against non-consensual sexual manipulation. The ethical governance of digital technologies increasingly requires responsibility to be integrated into design decisions rather than treated solely as an external legal response imposed after harm has occurred (Floridi, 2018). From the perspective of criminal policy, these obligations can function as forms of technical and situational prevention. Where developers knowingly design or market software specifically for criminal or abusive purposes, conventional concepts of complicity may become applicable. However, where general-purpose technologies are involved, criminal responsibility is more difficult because the requirement of shared intent between the developer and the principal offender is often absent (Jalali, 2022). For this reason, the study supports the development of a distinct regulatory regime for high-risk digital products in which failure to implement reasonable safety measures may trigger administrative, civil, or, in aggravated circumstances, corporate criminal consequences.

The challenges posed by sexual deepfakes are further intensified by the transnational nature of cyberspace and the fragmentation of criminal jurisdiction. A deepfake may be created in one state, processed through servers located in another jurisdiction, disseminated through a platform incorporated in a third country, and cause reputational and psychological harm to a victim residing elsewhere. Such scenarios complicate the determination of the locus delicti, the applicable criminal law, access to digital evidence, extradition procedures, and enforcement of judicial removal orders. Traditional territorial approaches to criminal jurisdiction are therefore increasingly supplemented by principles based on nationality, victim location, harmful effects, data sovereignty, and cross-border investigative cooperation (Fazaeli, 2021). Nevertheless, the practical exercise of jurisdiction remains dependent on international cooperation, particularly where the suspect, relevant data, platform infrastructure, or evidence is located abroad. Mutual legal assistance procedures are frequently too slow for offences in which harmful content can spread globally within minutes.

Accordingly, effective prevention and prosecution require accelerated mechanisms for preserving electronic evidence, identifying offenders, suspending access to unlawful content, and facilitating direct cooperation between competent authorities and major technology companies. The absence of harmonized criminalization also creates problems for extradition because many systems require dual criminality as a condition for surrendering an accused person. If sexual deepfake production constitutes a criminal offence in the requesting state but is treated merely as a civil wrong or privacy violation in the requested state, extradition may be unavailable. This regulatory fragmentation creates digital safe havens and undermines the effectiveness of domestic criminal policy. International legal cooperation should therefore move toward greater convergence in defining harmful synthetic sexual content, protecting victims of non-consensual digital manipulation, establishing minimum duties for platforms, and developing emergency procedures for cross-border content restriction and evidence preservation. At the domestic level, these developments should be accompanied by clearer rules governing the admissibility and authentication of electronic evidence. Deepfakes also create an evidentiary challenge because the increasing sophistication of synthetic media may weaken judicial confidence in audiovisual evidence. Criminal procedure must therefore incorporate advanced forensic verification tools and scientifically validated methods capable of distinguishing authentic from artificially generated materials (Schick, 2020).

In conclusion, the study demonstrates that sexual deepfakes cannot be effectively addressed through a conventional criminal policy centered exclusively on ex post punishment of individual offenders. The structure of contemporary digital harm is networked, technologically mediated, rapidly scalable, and distributed among multiple actors whose decisions collectively shape the opportunity for offending and the magnitude of resulting harm. An effective legal response therefore requires a transition toward a preventive and participatory model of criminal policy in which individual users, digital platforms, artificial intelligence developers, regulatory institutions, and international enforcement bodies assume differentiated but interconnected responsibilities. The proposed model is founded on three principal components. First, legislation should establish explicit and technologically neutral offences capable of addressing the intentional production and dissemination of non-consensual synthetic sexual content, while distinguishing such conduct from legitimate uses of generative technologies. Second, digital platforms should be subject to enhanced corporate duties requiring proactive detection, rapid restriction, transparent reporting, and effective preservation of evidence in relation to high-risk synthetic content. Third, developers of powerful generative systems should be required to implement safety-by-design mechanisms, including traceability, watermarking, digital fingerprinting, and safeguards against foreseeable criminal misuse. These measures should be complemented by media-literacy programs, improved digital-forensic capacity, stronger mechanisms for protecting victims, rapid international judicial cooperation, and comprehensive legislation governing artificial intelligence and synthetic media. The broader implication is that the future effectiveness of criminal law will increasingly depend on its ability to regulate technological architectures rather than merely sanction human conduct after harm has occurred. A comprehensive legal framework capable of integrating technical, administrative, civil, and criminal mechanisms is therefore necessary to protect human dignity, privacy, reputation, and psychological security in an era in which algorithmic systems can manufacture highly persuasive but entirely artificial realities.

References

Chesney, R., & Citron, D. K. (2019). Deepfakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, *107*, 1753-1820.

Citron, D. K. (2014). *Hate Crimes in Cyberspace*. Harvard University Press.
Elahimanesh, M. R. (2020). *Criminal Policy in Cyberspace: Approaches and Challenges*. Mizan Publishing.
Fazaeli, M. (2021). Criminal Jurisdiction in Cyberspace: From the Territoriality Principle to Global Data Sovereignty. *Criminal Law Research Journal*, *12*(1).
Floridi, L. (2018). Soft Ethics, the Governance of the Digital, and the General Data Protection Regulation. *Philosophy & Technology*, *31*(4), 543-547.
Harris, D. (2019). Deepfakes: False Pornography Is Here and the Law Cannot Protect You. *Duke Law & Technology Review*, *17*, 99-127.
Jalali, M. (2022). Criminal Liability in Artificial Intelligence-Related Crimes and Its Legislative Challenges. *Judiciary Law Journal*, *86*(118).
King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2020). Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*, *26*(1), 89-120.
Miró-Llinares, F. (2020). Artificial Intelligence and the Criminal Justice System: A Challenging Integration. *European Journal on Criminal Policy and Research*, *26*(3), 285-290.
Najafi Abrandabadi, A. H. (2019). Lecture Notes on Preventive Criminology and Criminal Policy. Shahid Beheshti University.
Niazpour, A. H. (2016). *Crime Prevention in Iranian Criminal Policy*. Mizan Publishing.
Rostami, H. (2020). Challenges of Substantive Criminal Law in Confronting Emerging Technologies: With Emphasis on Artificial Intelligence. *Criminal Law Research Quarterly*, *8*(31).
Schick, N. (2020). *Deepfakes: The Coming Infocalypse*. Hachette UK.
Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, *9*(11), 39-52.
Yar, M., & Steinmetz, K. F. (2019). *Cybercrime and Society*. SAGE Publications.
Yazdian Jafari, J. (2018). Analysis of the Foundations of Criminal Liability of Legal Persons in Iranian Law. *Judiciary Law Journal*, *82*(103).