Comparative Studies in Jurisprudence, Law, and Politics

An Optimal Criminal Policy toward Deepfakes: A Maximalist or Minimalist Response?

1. Hossein Sajedinia: Ph.D. Student, Department of Law and Islamic Jurisprudence, Payame Noor University, Tehran, Iran
2. Heydar Amirpour*: Associate Professor, Department of Law and Islamic Jurisprudence, Payame Noor University, Tehran, Iran. Email: dr.amirpour@pnu.ac.ir (Corresponding Author)
3. Mohammadsadegh Jamshidirad: Associate Professor, Department of Law and Islamic Jurisprudence, Payame Noor University, Tehran, Iran
4. Akram Abdolahpour: Assistant Professor, Department of Law and Islamic Jurisprudence, Payame Noor University, Tehran, Iran

ABSTRACT

With the emergence and rapid expansion of deepfake technology and the use of artificial intelligence to generate realistic fabricated content, societal psychological security, privacy, and the integrity of democratic institutions have been exposed to novel threats. By enabling identity impersonation and the manipulation of reality, this technology has created unprecedented challenges for criminal justice systems. This study aims to analyze and explain the optimal criminal policy for addressing this phenomenon and to determine whether a "maximalist response"—involving the use of severe punitive and coercive measures—or a "minimalist response"—relying on civil remedies, platform self-regulation, and non-criminal preventive mechanisms—can adequately respond to the requirements of this emerging threat. Using a descriptive-analytical method and drawing on library-based and documentary research, the study comparatively examines legislative and judicial approaches adopted in various legal systems and analyzes criminal policies concerning artificial intelligence–enabled cybercrime. The findings indicate that, owing to the technology-driven nature of deepfakes, their rapid dissemination, and the fluid boundary between "creativity" and "crime," the adoption of a maximalist penal policy, or absolute punitivism, may result in violations of civil liberties and prove ineffective in preventing harm. Conversely, a minimalist policy lacks sufficient deterrent capacity when deepfakes pose threats to national security or cause large-scale reputational harm. The findings therefore demonstrate the need to adopt a "participatory and stratified criminal policy." An optimal criminal policy toward deepfakes should involve neither an exclusively maximalist response nor the abandonment of responsibility under a minimalist approach. Instead, a "hybrid criminal policy model" is proposed. Under this model, precise and proportionate criminal sanctions should be applied to high-risk abuses, such as large-scale fraud and systematic reputational or institutional destruction, whereas the broad range of low-risk content should primarily be addressed through non-criminal policies, enhanced accountability of technology platforms, improved media literacy among citizens, and the development of digital authentication and content-provenance tools.

Keywords: *deepfake, criminal policy, artificial intelligence, maximalist response, minimalist response, cybercrime, non-criminal prevention.*

How to cite: Sajedinia, H., Amirpour, H., Jamshidirad, M., & Abdolahpour, A. (2026). An Optimal Criminal Policy toward Deepfakes: A Maximalist or Minimalist Response?. *Comparative Studies in Jurisprudence, Law, and Politics*, *8*(6), 1-17.

© 2026 the authors. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

Submit Date: 28 February 2026
Revise Date: 23 July 2026
Accept Date: 28 July 2026
Initial Publish Date: 04 November 2026
Final Publish Date: 20 February 2027

پژوهش‌های تطبیقی فقه،
حقوق و سیاست

سیاست جنایی مطلوب در قبال دیپ فیک‌ها: واکنش حداکثری یا حداقلی؟

۱. حسین ساجدی نیا: دانشجوی دکتری، گروه فقه و مبانی حقوق اسلامی، دانشگاه پیام نور، تهران، ایران
۲. حیدر امیر پور*: دانشیار، گروه فقه و مبانی حقوق اسلامی، دانشگاه پیام نور، تهران، ایران. پست الکترونیک: dr.amirpour@pnu.ac.ir (نویسنده مسئول)
۳. محمد صادق جمشیدی راد: دانشیار، گروه فقه و مبانی حقوق اسلامی، دانشگاه پیام نور، تهران، ایران
۴. اکرم عبدالله پور: استادیار، گروه فقه و مبانی حقوق اسلامی، دانشگاه پیام نور، تهران، ایران

چکیده

با ظهور و گسترش فناوری «دیپ‌فیک» و بهره‌گیری از هوش مصنوعی برای تولید محتوای جعلی واقع‌گرایانه، امنیت روانی جامعه، حریم خصوصی و تمامیت ارکان دموکراتیک با تهدیدات نوظهوری مواجه شده است. این فناوری، با ایجاد قابلیت جعل هویت و دستکاری واقعیت، چالش‌های بی‌سابقه‌ای را برای نظام عدالت کیفری ایجاد کرده است. هدف از این پژوهش، تحلیل و تبیین سیاست جنایی مطلوب در مواجهه با این پدیده است تا مشخص شود آیا رویکرد «واکنش حداکثری» (به‌کارگیری ابزارهای سرکوب‌گرانه و کیفری شدید) یا «واکنش حداقلی» (تکیه بر حقوق مدنی، خودانتظامی پلتفرم‌ها و ابزارهای پیشگیرانه غیرکیفری)، پاسخگوی مقتضیات این تهدید نوین است. این پژوهش با روش توصیفی-تحلیلی و با بهره‌گیری از مطالعات کتابخانه‌ای و اسنادی، به بررسی تطبیقی رویکردهای تقنینی و قضایی در نظام‌های حقوقی و تحلیل سیاست‌های جنایی در مواجهه با جرایم سایبریِ مبتنی بر هوش مصنوعی پرداخته است. یافته‌های تحقیق نشان می‌دهد که به دلیل ماهیت فناوری‌محور، سرعت بالای تکثیر و لغزندگیِ مرز میان «خلاقیت» و «جرم» در دیپ‌فیک، اتخاذ سیاست کیفری حداکثری (کیفرگرایی مطلق) منجر به نقض آزادی‌های مدنی و ناکارآمدی در پیشگیری می‌شود. از سوی دیگر، سیاست حداقلی نیز در برابر مخاطراتی که به امنیت ملی و هتک حیثیت‌های کلان منجر می‌شود، فاقد بازدارندگی لازم است. بنابراین، یافته‌ها حاکی از لزوم اتخاذ یک «سیاست جنایی مشارکتی و لایه‌بندی‌شده» است. سیاست جنایی مطلوب در قبال دیپ‌فیک، نه واکنش حداکثری صرف است و نه رهاسازیِ مسئولیت در قالب واکنش حداقلی؛ بلکه «مدل سیاست جنایی ترکیبی» پیشنهاد می‌شود. در این الگو، برای سوءاستفاده‌های خطرناک (مانند کلاهبرداری‌های کلان و تخریب سیستماتیک) واکنش کیفریِ دقیق و متناسب، و برای طیف گسترده‌ای از تولیداتِ با ریسک پایین، تمرکز بر سیاست‌های غیرکیفری، مسئولیت‌پذیری پلتفرم‌های تکنولوژیک، ارتقای سواد رسانه‌ای شهروندان و توسعه ابزارهای احراز اصالتِ دیجیتال، به عنوان راهکارهای اصلی تأکید می‌گردد.

واژگان کلیدی: *دیپ‌فیک، سیاست جنایی، هوش مصنوعی، واکنش حداکثری، واکنش حداقلی، جرایم سایبری، پیشگیری غیرکیفری.*

نحوه استناددهی: ساجدی نیا، حسین.، امیر پور، حیدر.، جمشیدی راد، محمد صادق.، و عبدالله پور، اکرم. (۱۴۰۵). سیاست جنایی مطلوب در قبال دیپ فیک‌ها: واکنش حداکثری یا حداقلی؟. *پژوهش‌های تطبیقی فقه، حقوق و سیاست*، ۸(۶)، ۱۷-۱.

© ۱۴۰۵ تمامی حقوق انتشار این مقاله متعلق به نویسنده است. انتشار این مقاله به‌صورت دسترسی آزاد مطابق با گواهی (CC BY-NC 4.0) صورت گرفته است.

تاریخ ارسال: ۹ اسفند ۱۴۰۴
تاریخ بازنگری: ۱ مرداد ۱۴۰۵
تاریخ پذیرش: ۶ مرداد ۱۴۰۵
تاریخ چاپ اولیه: ۱۳ آبان ۱۴۰۵
تاریخ چاپ نهایی: ۱ اسفند ۱۴۰۵

مقدمه

ظهور و گسترش هوش مصنوعی (*AI*) و فناوری‌های وابسته به آن، نقطه عطفی در تاریخ تحولات بشری محسوب می‌شود. چنان‌که در مطالعات مرتبط نیز تأکید شده است، این فناوری‌ها در کانون انقلابی جهانی قرار گرفته‌اند که رفاه، امنیت و حیات اجتماعی جوامع را در اقصی نقاط جهان تحت تأثیر قرار داده است (Yang Chik & Samsudin, 2023). ذات هوش مصنوعی که با مؤلفه‌هایی نظیر خودآموزی، تصادف‌انگاری، پیش‌بینی‌ناپذیری و خودمختاری گره خورده است (Zakerinia & Gholampour, 2024)، این فناوری را از ابزارهای سنتی متمایز ساخته و ظرفیت‌های خیره‌کننده‌ای را در اختیار بشر قرار داده است. با این‌همه، در کنار کارکردهای مثبت و تسهیل‌گر هوش مصنوعی، بروز چالش‌های نوپدید ناشی از آن انکارناپذیر است. در این میان، فناوری «دیپ‌فیک» (*Deepfake*)، به‌عنوان محصولی پیچیده از هوش مصنوعی، نمونه‌ای بارز از فناوری‌های «دولبه» است. اگرچه این فناوری در حوزه‌های هنری و فرهنگی کاربردهای مشروع و خلاقانه‌ای دارد، اما ظرفیت بالای آن در تولید محتوای فریبنده، به تهدیدی جدی علیه حریم خصوصی، امنیت فردی و انسجام اجتماعی تبدیل شده است. ماهیت دوگانه دیپ‌فیک‌ها، دستگاه‌های قانون‌گذار و سیاست‌گذاران جنایی را با این پرسش بنیادین مواجه ساخته است که چه نوع واکنش کیفری در قبال این پدیده، «مطلوب» و متناسب با اقتضائات نوین است؟ در تبیین راهبردهای تقنینی، تقابل میان دو رویکرد «واکنش حداکثری» و «واکنش حداقلی» خودنمایی می‌کند. سیاست جنایی حداکثری، مبتنی بر پارادایم سرکوب‌گرایانه، بر جرم‌انگاری موسع و تشدید مجازات‌ها تمرکز دارد؛ درحالی‌که سیاست جنایی حداقلی، با تأکید بر اصل تناسب، جرم‌انگاری محدود و مدارای کیفری، تلاش دارد از تورم کیفری جلوگیری کند. مقاله حاضر با هدف پاسخ به این مسئله اساسی تدوین شده است که با عنایت به چیستی، خصوصیات فنی و ابعاد مخرب دیپ‌فیک‌ها، کدام‌یک از دو رویکرد فوق، الگویی مناسب برای مواجهه با این پدیده محسوب می‌شود؟ بدین منظور، ضمن واکاوی ماهیت دیپ‌فیک‌ها و بررسی انتقادی استدلال‌های له و علیه واکنش‌های حداکثری و حداقلی، تلاش خواهد شد الگوی سیاست جنایی مطلوب در این حوزه ترسیم گردد.

شناخت دیپ‌فیک‌ها و فناوری‌های ساخت آن

این فناوری در سال ۲۰۱۴ توسط یان گودفلو خلق شد (Kulger & Pace, 2021) و در سال ۲۰۱۷، هنگامی‌که کاربران گروه ردیت یک ویدئوی ساختگی هرزه‌نگاری مشهور را با استفاده از فناوری هوش مصنوعی ایجاد کردند، متداول گردید (Gamage, 2022). به‌منظور اینکه مشخص شود چه سیاستی می‌تواند رویکردی مناسب برای برخورد با دیپ‌فیک‌ها تلقی شود، نخستین گام، دستیابی به درکی جامع از این پدیده است.

مفهوم‌شناسی دیپ‌فیک

اولین مسئله‌ای که باید به آن پرداخته شود، تعریف دیپ‌فیک است. واژه دیپ‌فیک در مقابل چیپ‌فیک یا جعل‌های ساده قرار داده می‌شود که اساساً شامل محتوای مجعولی است که با استفاده از فناوری‌هایی ساخته می‌شود که از قبل موجود بوده و به‌سادگی قابل استفاده‌اند و نیازمند تجربه و مهارت فنی بسیار اندکی هستند (Albahar & Almalki, 2019). آن‌ها به‌جای استفاده از هوش مصنوعی برای تولید محتوا، به دست انسان و از طریق قابلیت‌های به‌اصطلاح ویرایش یا فناوری‌های پایین‌تر از هوش مصنوعی ساخته می‌شوند. نمونه متداول امروزی آن، فتوشاپ (Gamage, 2022) یا فناوری‌های آهسته‌کردن سرعت فیلم است که از طریق یک برنامه پایه ویرایش ویدئو انجام می‌شود (Vizoso et al., 2021).

در تعاریفی که از این واژه ارائه شده، دیپ‌فیک‌ها هم برای فناوری ساخت و هم در مورد محصول آن به کار گرفته شده‌اند. به‌عنوان مثال، برخی در رابطه با این اصطلاح بیان داشته‌اند: «اصطلاح دیپ‌فیک ترکیبی از "آموزش عمیق" و "تولید جعلی" است» (Vizoso et al., 2021). فناوری دیپ‌فیک به‌معنای ابزاری با هدف تغییر و دست‌کاری تصاویر و ویدئوها تعریف شده است (Albahar & Almalki, 2019). به‌طور خاص، این واژه به‌معنای فناوری دست‌کاری ویدئویی است که اجازه می‌دهد چهره یک فرد در یک کلیپ ویدئویی یا تصویر، جایگزین چهره شخص دیگری شود (Maras & Alexandrou, 2018). فناوری دیپ‌فیک می‌تواند برای جایگزین‌کردن چهره‌ها در ویدئوها یا تصاویر، به‌منظور نشان‌دادن اینکه اشخاص چیزی گفته‌اند یا انجام داده‌اند که هرگز نگفته یا انجام نداده‌اند، به کار رود (Nguyen & Ademiluyi, 2023). دیپ‌فیک‌ها به‌شکل متداول در رسانه‌های اجتماعی ارسال شده و اخبار اصلی را پوشش می‌دهند (Vizoso et al., 2021).

بر اساس یک تعریف، دیپ‌فیک‌ها به «رسانه‌ای دست‌کاری‌شده یا مصنوعی اشاره دارند که توسط نوعی از هوش مصنوعی، به‌شکلی متقاعدکننده، شخصی جعلی یا نمایشی جعلی از یک شخص را به‌صورت دیجیتال دست‌کاری یا تغییر داده و عرضه می‌کنند». ویدئوهای دیپ‌فیک شامل فناوری‌های پیچیده‌ای از هوش مصنوعی هستند و انواع مختلفی از نرم‌افزارها برای آن‌ها وجود دارد (Albahar & Almalki, 2019). در دسته دیگری از تعاریف، این اصطلاح در مورد محصول فناوری اعمال شده است. از جمله، آن را محتوای دست‌کاری‌شده با استفاده از فناوری‌های هوش مصنوعی و یادگیری عمیق پیشرفته تعریف کرده‌اند (Gamage, 2022).

بنابراین، می‌توان گفت اصطلاح دیپ‌فیک به ویدئو، صوت، تصویر یا فناوری ساخت آن‌ها اشاره دارد که در آن، از طریق جابه‌جایی چهره، محتوایی قانع‌کننده، دست‌کاری‌شده و غیرواقعی از رفتار یا گفتار شخصی ارائه می‌شود که در واقع چنین امری رخ نداده است. این بدان معناست که محصولات ساخته‌شده با این فناوری‌ها، یک محتوای غیرواقعی را به‌شکلی باورپذیر به مخاطب ارائه می‌دهند.

فناوری‌های ساخت دیپ‌فیک

دومین رکنی که در شناخت دیپ‌فیک‌ها باید به آن پرداخته شود، نرم‌افزارهای ساخت آثار سمعی و بصری مجعول هستند. فارغ از اقدامات فنی و محاسباتی که این نرم‌افزارها بر اساس آن عمل می‌کنند و خارج از موضوع این نوشتار است، نرم‌افزارهای ساخت دیپ‌فیک‌ها به چند دسته تقسیم می‌شوند: ۱- نرم‌افزارهای تخصصی. این نرم‌افزارها یا محصول مشخصی را تولید می‌کنند یا بر بُعد خاصی متمرکز هستند. به‌عنوان مثال، برای ساخت دیپ‌فیک‌های رقص یا صرفاً حرکات اجزای صورت مورد استفاده قرار می‌گیرند. ۲- نرم‌افزارهایی که کاربرد آموزشی دارند. ۳- نرم‌افزارهایی که مختص کاربر فاقد دانش رایانه بوده و بدون نیاز به اقدام پیچیده‌ای، محصول را تهیه می‌کنند.

ساخت دیپ‌فیک‌ها توسط انواع مختلفی از نرم‌افزارها انجام می‌شود. فیس‌سوآپ، دیپ‌فیس‌لب، فیس‌اپ، ومبو، زائو، ری‌فیس، مورفین و جیگی برخی از شناخته‌شده‌ترین این برنامه‌ها هستند (Albahar & Almalki, 2019) که هریک کاربرد ویژه‌ای دارند. به‌عنوان مثال، دیپ‌فیس‌لب با هدف تحقیق ساخته شده و ادعا شده است که بیش از ۹۵ درصد ویدئوهای دیپ‌فیک با این برنامه ساخته می‌شوند. این برنامه به شما اجازه جایگزینی صورت، تمام سر، تغییر در حرکت لب‌ها و همچنین تغییر سن افراد را می‌دهد. فیس‌سوآپ یا مبادله صورت نیز همانند این برنامه برای اهداف آموزشی طراحی شده است (Nguyen & Ademiluyi, 2023). دیپ‌فیس‌لب امکان استفاده آسان را برای افرادی فراهم می‌کند که فاقد درکی جامع از یادگیری عمیق هستند. دیپ‌آرت‌افکت برنامه دیگری است که برای کارهای هنری بر روی عکس‌ها استفاده می‌شود (Cejudo, 2023). همچنین، فیس‌اپ برای ایجاد تغییرات بر روی تصاویر، پس‌زمینه‌ها و نیز اعمال انواع فیلترها روی عکس‌ها مورد استفاده قرار می‌گیرد.

هرچند بسیاری از این برنامه‌ها کاربردهای سرگرمی و تفریحی دارند، باید توجه داشت که در سیاست جنایی برای مقابله با دیپ‌فیک‌ها، یکی از ابعادی که می‌تواند مورد توجه قرار گیرد، نرم‌افزارها و برنامه‌های ساخت آن‌هاست.

کاربردهای دیپ‌فیک

مسئله دیگر در مورد دیپ‌فیک‌ها، کاربردهای آن‌هاست. استفاده از دیپ‌فیک‌ها به‌شکل قابل‌ملاحظه‌ای متنوع است. این استفاده‌ها می‌توانند مثبت یا منفی باشند. با تسامح، می‌توان آن‌ها را معادل کاربردهای غیرمجرمانه و مجرمانه طبقه‌بندی کرد.

کاربردهای مثبت

دیپ‌فیک‌ها می‌توانند برای انواعی از کاربردهای متداول مورد استفاده قرار گیرند. از دیپ‌فیک‌ها می‌توان برای صنایع هنری و حتی کارهای هنری استفاده کرد (Cejudo, 2023). از آن‌ها می‌توان برای خلق موسیقی‌های جدید با استفاده از نمونه آثار هنرمندان موجود (Cejudo, 2023) یا برای گنجاندن اجرای افراد متوفی در تولیدات سمعی و بصری، به‌منظور انتشار میراث فرهنگی و خلق تصاویر و اصوات به سبک بازیگران واقعی، استفاده کرد (Cejudo, 2023). دیپ‌فیک‌ها از طریق ارائه پاسخ‌های شخصی‌سازی‌شده به تماس‌گیرندگان، که شامل ارجاع تماس‌ها و سایر خدمات پاسخ‌گویی است، می‌توانند در خدمات تلفن گویا مورد استفاده قرار گیرند. پشتیبانی تلفنی مشتری و کاربردهای تفریحی، همچون شبیه‌سازی و دست‌کاری صدای بازیگران برای صحنه‌های خاص در فیلم‌های هالیوود و بازی‌های ویدئویی، از دیگر استفاده‌های مثبت دیپ‌فیک‌ها هستند (Cejudo, 2023).

کاربردهای منفی

کیفیت و اثربخشی ایجاد تصاویر جایگزینی چهره به‌وسیله یادگیری عمیق، به‌شکل قابل‌توجهی تقویت شده است. به‌عنوان مثال، دست‌کاری چهره با دیپ‌فیک‌ها به‌قدری واقعی شده است که به‌دشواری می‌توان از طریق تشخیص خودکار یا دستی، اصالت آن را تشخیص داد (Zhang et al., 2020). همین مسئله زمینه بسیاری از کارکردها را برای دیپ‌فیک‌ها فراهم ساخته است. علی‌رغم جنبه‌های مفید، از فناوری‌های دیپ‌فیک در اخبار جعلی، اطلاعات غلط، ضد اطلاعات و فریب، که می‌توانند جامعه را تهدید کنند، سوءاستفاده می‌شود. جایگزینی چهره می‌تواند قربانیان را به روشی منفی فریب دهد (Yang Chik & Samsudin, 2023). انتشار اطلاعات و جایگزینی چهره در آن‌ها بدون رضایت انجام می‌شود و می‌تواند ناقض اطلاعات خصوصی و شخصی افراد باشد (Lussier, 2022). آن‌ها همچنین برای ویدئوهای هرزه‌نگاری آنلاین بدون رضایت (Kulger & Pace, 2021)، از جمله برای شرمنده‌سازی و بی‌اعتبارکردن زنان، مورد استفاده قرار می‌گیرند. به‌عنوان مثال، می‌توان به بهره‌گیری از دیپ‌فیک در مورد روزنامه‌نگاران هندی منتقد دولت، مادری که می‌خواست هم‌تیمی‌های ژیمناستیک دخترش را حذف کند و سازندگان فیلتر برهنه‌کننده‌ای که لباس را از روی عکس شخص جدا می‌کند و به‌گونه‌ای طراحی شده است که فقط برای زنان کار کند، اشاره کرد (Lussier, 2022).

یکی دیگر از کاربردهای مجرمانه دیپ‌فیک‌ها، کلاهبرداری از طریق تقلید صداست. علی‌رغم تمام تلاش‌هایی که از سوی خزانه‌داری و مدیران مالی صورت گرفته است، گزارش‌ها نشان می‌دهد که بخش قابل‌توجهی از کلاهبرداری‌ها از طریق دیپ‌فیک‌ها انجام شده‌اند (Albahar & Almalki, 2019). دیپ‌فیک‌ها، غیر از زیان‌های فردی، می‌توانند زیان‌هایی برای جامعه نیز به همراه داشته باشند. در این رابطه بیان شده است که تبلیغات اطلاعات جعلی و اخبار جعلی این قابلیت را دارند که افراط‌گرایی خشونت‌آمیز و سخنرانی‌های تنفرآمیز را ترویج کنند و در نهایت، دموکراسی‌ها را تضعیف کرده و اعتماد به فرایندهای دموکراتیک را کاهش دهند (Vizoso et al., 2021).

همین امر نیز موجب نگرانی‌هایی در مورد استفاده از این فناوری به‌عنوان یک سلاح در جنگ‌ها و بحران‌ها شده است. انتشار ویدئوی جعلی از نخست‌وزیر بلژیک، سوفی ویلمز، توسط گروهی تحت عنوان فعالان شورش انقراض، در رابطه با کووید-۱۹ فضایی بحرانی ایجاد کرد. این ویدئو به‌عنوان بخشی از کمپین بلژیکی «حقیقت را بگو» ساخته شد. این گروه متعاقباً سخنرانی دیگری را از سوی برخی کارشناسان، با این ادعا منتشر کرد که بیماری‌هایی همچون کووید-۱۹، ابولا و سایر بیماری‌ها به‌واسطه سوءاستفاده و تخریب محیط‌زیست طبیعی توسط انسان‌ها رخ داده‌اند (Gamage, 2022). از دیگر مثال‌های چنین کارکردهایی می‌توان به افزایش وحشتناک پیامدهای ناشی از اطلاعات نادرست درباره خشونت علیه شهروندان و دست‌کاری انتخابات، از جمله در میانمار، اشاره کرد (Vizoso et al., 2021).

واکنش جنایی نسبت به تولیدات هوش مصنوعی

یکی از آثار بهره‌گیری از فناوری‌های هوش مصنوعی مولد این است که تشخیص محتوای ساخته‌شده به دست بشر از محتوای ساخته‌شده به‌وسیله هوش مصنوعی، به‌شکل فزاینده‌ای دشوارتر شده و به‌طور بالقوه، رفتارهای غیرقانونی و زیان‌بار را ممکن می‌سازد. ازاین‌رو، سیاست‌گذاران در اقصی نقاط جهان در حال تأمل درباره طراحی و اجرای تکنیک‌های داغ‌زنی به‌منظور تضمین صداقت محیط هوش مصنوعی هستند (Yang Chik & Samsudin, 2023). در همین راستا، می‌توان واکنش‌هایی را از سوی برخی کشورها در قبال تهدیدهای مذکور مشاهده کرد. به‌عنوان مثال، چین گام‌هایی را برای ممنوعیت تصاویر خلق‌شده با هوش مصنوعی اتخاذ کرده است. مقررات اجرایی پیشنهاد الگوریتم برای خدمات اطلاعات اینترنتی ۲۰۲۲، موسوم به دستورالعمل پیشنهاد الگوریتم، مقررات مدیریت سنتز عمیق اطلاعات اینترنتی ۲۰۲۳، موسوم به دستورالعمل سنتز عمیق، و مقررات موقت مدیریت خدمات هوش مصنوعی مولد ۲۰۲۳، موسوم به دستورالعمل هوش مصنوعی مولد، از اقدامات این کشور در این خصوص هستند (Yang Chik & Samsudin, 2023). ایالات متحده نیز سازوکارهای مؤثری برای ایجاد برچسب‌گذاری در پیش گرفته است؛ به‌گونه‌ای که کاربر نهایی قادر باشد تشخیص دهد چه زمانی یک محتوا با استفاده از هوش مصنوعی ساخته شده و چه زمانی چنین نیست. در قانون جدید اتحادیه اروپا در دسامبر ۲۰۲۳ نیز توافق شد تعهداتی برای تولیدکنندگان و کاربران سیستم‌های هوش مصنوعی مقرر شود تا تشخیص محتواهای خلق‌شده توسط هوش مصنوعی میسر گردد (Yang Chik & Samsudin, 2023). همچنین، دستورالعمل شماره ۲۰۲۴/۱۶۸۹ اتحادیه اروپا قواعد هماهنگی را در رابطه با هوش مصنوعی بیان کرده است که تولیدکنندگان و استقراردهندگان هوش مصنوعی را با الزامات و تعهدات مشخصی در رابطه با کاربردهای خاص هوش مصنوعی مواجه می‌سازد.

حال این پرسش مطرح می‌شود که چه سیاست جنایی برای مقابله با دیپ‌فیک‌ها مناسب است. برای این منظور، دو رویکرد حداکثری و حداقلی مورد بررسی قرار خواهد گرفت.

رویکرد حداکثری

برای اعمال سیاست جنایی حداکثری، لازم است ابتدا مفهوم، ریشه‌ها، ابعاد و علل این رویکرد مشخص شود و سپس علل موجه توسل به آن در رابطه با مقابله جنایی با دیپ‌فیک‌ها مورد بررسی قرار گیرد.

چارچوب رویکرد حداکثری

به عقیده برخی، «سیاست کیفری مجموعه‌ای منسجم و پیوسته از راهبردهای تقنینی، قضایی و اجرایی است که با به‌کارگیری ابزار کیفر و ضمانت‌اجراهای حقوقی، درصدد مبارزه با بزهکاران و واکنش متناسب پیرامون انواع جرایم است» (Sheikh et al., 2022). سیاست جنایی

حداکثری در دو زمینه جرم‌انگاری و مجازات متصور است. در فرایند جرم‌انگاری، با تکیه بر منافع و مصالح موردنظر قانون‌گذار، بر لزوم گزینشی تأکید می‌شود که می‌تواند در حیطه جرایم مختلف، جرم‌انگاری‌های حداقلی یا حداکثری را فراهم سازد (Majidi & Tajabadi, 2019).

رویکرد حداکثری، در مواردی که علت آن نه شیوه حکومت، بلکه خصوصیات پدیده تهدیدآمیز است، در صورتی ضرورت دارد که چند مؤلفه مورد توجه قرار گیرد. نخست، شناعت آن پدیده است. هر اندازه که یک رفتار یا وضعیت از شناعت بیشتری برخوردار باشد، اتخاذ رویکرد سخت‌گیرانه‌تر ضروری‌تر است. ۲- سهولت وقوع؛ هرچه ارتکاب یک رفتار یا خلق یک وضعیت آسان‌تر باشد، امکان مواجه‌شدن با آن بیشتر است و بنابراین، رویکرد سخت‌گیرانه‌تری را ضروری می‌کند. ۳- گستره بزه‌دیدگان و جامعه هدف؛ هرچه یک تهدید بتواند اشخاص بیشتری را در معرض تأثیر قرار دهد، آثار گسترده‌تری بر جای می‌گذارد. بنابراین، اجتناب از وقوع آن واکنش سخت‌گیرانه‌تری را ایجاب می‌کند. ۴- شدت تأثیرگذاری؛ آثار تهدید نیز نقش قابل‌ملاحظه‌ای در نوع برخورد با آن دارند. در مواردی که این آثار شدید و جبران‌ناپذیر باشند، واکنش سخت‌گیرانه‌تر ضروری‌تر است.

رویکرد حداکثری در جرم‌انگاری به دو شکل کمی و کیفی صورت می‌پذیرد. از منظر کمی، «جرم‌انگاری حداکثری به‌معنای وجود قوانین بیش‌ازاندازه در حیطه حقوق جزاست» (Mohammadnejad & Tajik, 2019). اما رویکرد حداکثری تنها محدود به تعدد کمی نیست، بلکه می‌تواند در قالب توسعه مفهومی جرایم نیز صورت پذیرد. در این حالت، قانون‌گذار با ارائه تعاریف یا معیارهای شناسایی موسع، رویکردی حداکثری را برای جرم‌انگاری اتخاذ می‌کند؛ به‌گونه‌ای که جرایم، مصادیق گسترده‌ای از رفتارها یا پدیده‌ها را تحت پوشش قرار دهند.

درحالی‌که جرم‌انگاری حداکثری کمی می‌تواند وابسته به ساختار حاکمیت و به‌عنوان یک سیاست کلی در پیش گرفته شود، جرم‌انگاری حداکثری کیفی می‌تواند به‌واسطه اهمیت امنیت ملی و خطر جرم یا جرایم مشخص نسبت به آن اتخاذ شود. تأمین نظم و امنیت عمومی و ایجاد ثبات دائمی در حاکمیت سیاسی، منفعتی اولویت‌دار و مهم است که دغدغه قانون‌گذاران برای پرهیز از جرم‌انگاری حداکثری را از یاد برده است (Majidi, 2009). به‌دنبال نقض گسترده امنیت و آرامش روانی شهروندان در پی گسترش ارتکاب جرم، اهداف سیاست جنایی از مجرم‌مداری به‌سمت جرم‌محوری تغییر مسیر داده است و تضمین امنیت به دغدغه اصلی سیاست جنایی بدل شده است. سیاست جنایی امنیت‌گرا به تشکیل نوعی حقوق کیفری دشمن‌مدار سوق یافته است (Ghannad & Akbari, 2017). سیاست جنایی امنیت‌مدار یا تسامح صفر، راهبردی است که به‌موجب آن متولیان سیاست‌گذاری کیفری، در جهت تأمین امنیت و مبارزه با جرایم و مجرمان، رهیافت‌های سخت‌گیرانه‌ای را تدوین و اجرا می‌کنند (Tajabadi & Meshkani, 2022).

رویکرد حداکثری ایجاب می‌کند تا ماهیت تهدیدبودن عمل یا وضعیت، معیار قرار گیرد. بنابراین، در جرم‌انگاری، اولاً توصیف موسعی از پدیده مجرمانه ارائه شود؛ به‌گونه‌ای که طیف وسیعی از اقدامات یا وضعیت‌ها را تحت پوشش قرار دهد. ثانیاً، از عبارات مبهم و تفسیرپذیر استفاده شود تا عنوان مجرمانه بتواند مصادیق جدیدی را که در زمان تدوین آن وجود نداشته یا به ذهن متبادر نشده‌اند نیز پوشش دهد تا به‌واسطه طولانی‌بودن زمان قانون‌گذاری، نظام حقوقی با خلأ مقررات و آسیب‌پذیری ناشی از رشد قارچ‌گونه یک تهدید جدید، دچار آسیب جبران‌ناپذیر نشود.

ضرورت رویکرد حداکثری نسبت به دیپ‌فیک‌ها

با توجه به توضیحات مذکور، این پرسش مطرح می‌شود که آیا می‌توان دیپ‌فیک‌ها را پدیده‌ای دانست که واکنش حداکثری نسبت به آن‌ها را ایجاب کند؟ رویکرد حداکثری نسبت به دیپ‌فیک‌ها شامل جرم‌انگاری حداکثری، هم نسبت به فناوری ساخت و هم نسبت به محصول آن است. مبتنی بر ماهیت تهدیدگونه، رویکرد حداکثری می‌تواند شامل ۱- جرم‌انگاری ساخت، دسترس‌پذیری و انتشار فناوری‌های ساخت دیپ‌فیک‌ها؛ ۲- جرم‌انگاری آموزش ساخت دیپ‌فیک‌ها؛ و ۳- جرم‌انگاری ساخت، استفاده و انتشار آثار سمعی و بصری مجعول مبتنی بر هوش مصنوعی باشد. جرم‌انگاری حداکثری به دو شکل متصور است. نخست، جرم‌انگاری مطلق: در این حالت، دیپ‌فیک‌ها به‌دلیل ذات دروغین خود و قابلیت ایجاد تهدیدی جدی و جبران‌ناپذیر، تهدیدی مطلق علیه اشخاص و جامعه قلمداد شده و مطلقاً جرم‌انگاری می‌شوند. صورت دوم آن است که اصل بر مجرمانه‌بودن آن‌ها همراه با استثنا باشد. صورت نخست را می‌توان با شکنجه مقایسه کرد که خصوصیات آن جرم‌انگاری مطلق آن را اقتضا می‌کند. صورت دوم نیز با سلب حیات قابل‌مقایسه است که اصل بر ممنوعیت آن، جز در موارد استثنایی و با توجه به جدیت و ضرورت استثناهاست. از این منظر، دیپ‌فیک‌ها اصولاً پدیده‌ای مجرمانه تلقی می‌شوند، مگر مواردی که غیرمجرمانه و مجاز باشند یا استانداردهای سخت‌گیرانه‌ای در مورد استثناها ارائه شود.

چنین رویکردی را تا حدودی می‌توان در پاره‌ای از سیاست‌های جنایی نیز مشاهده کرد. به‌عنوان مثال، اتحادیه اروپا دیپ‌فیک‌ها را به‌عنوان یکی از بدنام‌ترین چالش‌های ناشی از شکل‌گیری فناوری هوش مصنوعی معرفی کرده و بیان می‌دارد که آن‌ها به ابزاری برای انتشار اطلاعات غلط و ضد اطلاعات تبدیل شده‌اند که جامعه را با چالش‌های جدی مواجه کرده‌اند. آن‌ها مرزهای میان واقعیت و خیال را از طریق خلق رسانه‌های جعلی فوق‌العاده باورپذیر تیره کرده‌اند (Albahar & Almalki, 2019).

تعریف موسع در جرم‌انگاری را می‌توان در توصیف هوش مصنوعی مولد در اتحادیه اروپا مشاهده کرد. این نهاد، اصطلاح مذکور را شامل فناوری‌ای می‌داند که برای تولید انواع مختلفی از محتوای جدید در پاسخ به درخواست یک کاربر به کار می‌رود (Yang Chik & Samsudin, 2023) و می‌تواند طیف گسترده‌ای از فناوری‌های مبتنی بر هوش مصنوعی را شامل شود. بنابراین، هرگونه واکنش جزایی شامل توصیف موسعی از این فناوری‌ها خواهد بود.

رویکرد مشابهی را می‌توان در کشور چین نیز مشاهده کرد. در ۲۴ مه ۲۰۲۴، کمیته فنی استانداردسازی اطلاعات امنیت ملی چین، پیش‌نویس جدیدی با عنوان «فناوری امنیت سایبری ـ الزامات امنیتی اولیه برای خدمات هوش مصنوعی مولد» منتشر کرد. این پیش‌نویس شامل امنیت اطلاعات آموزشی، مدل امنیتی و اقدامات امنیتی برای تقلیل مؤثر خطرها و همچنین مقرراتی کلیدی برای تضمین شفافیت و تبعیت خدمات مولد هوش مصنوعی و ارائه‌دهندگان این خدمات بود. خدمات مولد هوش مصنوعی در این پیش‌نویس به خدماتی اشاره دارد که برای تولید انواع مختلف محتواها، همچون متن، تصویر، صدا و فیلم، برای مصرف‌کننده عام از فناوری‌های هوش مصنوعی مولد استفاده می‌کنند و تولیدکنندگان را شامل کلیه اشخاص و سازمان‌ها توصیف می‌کند.

در این سند، اطلاعات زیان‌بار باید شامل هر داده‌ای باشد که حاوی بیش از پنج درصد اطلاعات غیرقانونی و زیان‌بار است. پیش‌نویس مذکور تعریف موسعی از اطلاعات زیان‌بار ارائه داده و حوزه‌های خطر را شامل ۱- نقض ارزش‌های بنیادین سوسیالیستی؛ ۲- فحاشی و خشونت؛ ۳- محتوای غیرقانونی، یعنی هر محتوایی که توسط قانون و مقررات ممنوع شده و ذاتاً مضر است؛ ۴- محتوای تبعیض‌آمیز؛ ۵- نقض‌های

تجاری؛ و ۶- نقض حقوق قانونی می‌داند. این سند اقدامات امنیتی‌ای را پیشنهاد می‌کند؛ از جمله اینکه ارائه‌دهندگان باید ضرورت و ایمنی استقرار هوش مصنوعی مولد در حوزه خدمات خود را توجیه کنند (Yang Chik & Samsudin, 2023).

در ایالات متحده آمریکا، قانون «دفاع از هر فرد در مقابل ارائه نادرست از طریق بهره‌برداری موضوع قانون پاسخ‌گویی ۲۰۲۳» یا قانون پاسخ‌گویی دیپ‌فیک‌ها، در سپتامبر ۲۰۲۳ تصویب شد. ماده ۱۰۴۱ این قانون از عبارت «سابقه شخصیت کاذب با فناوری پیشرفته» استفاده کرده و بیان می‌دارد: «جز آن‌گونه که در بند (j) مقرر شده است، هر شخصی که با استفاده از هر ابزار یا تسهیلات تجارت بین‌ایالتی یا خارجی، سابقه شخصیت کاذب با فناوری پیشرفته را با هدف توزیع آن در اینترنت تولید می‌کند یا می‌داند که چنین ثبتی باید این‌گونه توزیع شود، باید تضمین کند که چنین ثبت‌هایی از موارد زیر پیروی می‌کنند: ...». بند دوم این ماده نیز بیان می‌دارد که ثبت‌های مذکور شامل عناصر تغییریافته صوتی و تصویری یا خلق‌شده توسط هوش مصنوعی مولد یا فناوری‌های مشابه است. از منظر این قانون، ثبت شخصیت کاذب با فناوری پیشرفته به‌معنای هرگونه دیپ‌فیکی است که:

الف- یک فرد متعارف، با درنظرگرفتن کیفیت سمعی و بصری آن و ماهیت کانال توزیعی که در آن ظاهر شده است، به‌درستی باور کند که:

۱- یک فعالیت عمده از شخص زنده‌ای را مشاهده می‌کند که آن شخص در حقیقت انجام نداده است؛ یا ۲- هر فعالیت عمده‌ای از فرد متوفی را مشاهده می‌کند که آن فرد در حقیقت انجام نداده و آن تصویر به‌شکل قابل‌ملاحظه‌ای شبیه اقدامی مجرمانه بوده یا منجر به مداخله نامناسبی در اقدامات عمومی، مباحث نظم عمومی یا انتخابات می‌شود.

ب- بدون رضایت آن فرد زنده یا، در صورت فوت، بدون رضایت متوفی یا وراث وی تولید شده باشد.

بند ششم این ماده نیز، طی مقررات مفصلی، هرکس را که آگاهانه از رعایت الزامات بند نخست قصور ورزد، مشمول مجازات کیفری قرار داده است. علاوه بر مجازات‌های کیفری، قانون مذکور مجازات مدنی بیش از یکصد و پنجاه هزار دلار برای هر ثبت یا تغییر و همچنین جبران تنبیهی مناسب را مقرر کرده است (Yang Chik & Samsudin, 2023).

بنابراین، ملاحظه می‌شود که قانون‌گذار نه‌تنها تعریف موسعی از دیپ‌فیک‌ها ارائه داده است، بلکه واکنش‌های گسترده‌ای را نیز در رابطه با آن‌ها اتخاذ کرده است. باید توجه داشت که موارد مذکور، غیر از جبران خساراتی است که بسته به محتوا و آثار دیپ‌فیک‌ها، به میزان پنجاه، یکصد و یکصد و پنجاه هزار دلار تعیین شده‌اند.

رویکرد حداقلی

در عصر حاضر، بسیاری از کشورها با پدیده جرم‌انگاری حداکثری روبه‌رو بوده‌اند و این امر به یکی از مهم‌ترین چالش‌های بسیاری از کشورها بدل شده است (Arabiyan & Akhtari, 2024). این مسئله، با توجه به توسعه فناوری‌های جدید و به‌طور اخص هوش مصنوعی و فضای سایبر، که جوامع را با طیف گسترده‌تری از تهدیدها مواجه می‌سازند، جدی‌تر نیز شده است. در چنین شرایطی، نظم اجتماعی جامعه نیز می‌تواند تحت تأثیر رویکرد حداکثری در واکنش‌های کیفری به پدیده‌ها قرار گیرد. ازاین‌رو، این پرسش مطرح می‌شود که آیا مناسب‌تر نیست با درنظرگرفتن مجموع اوضاع‌واحوال و تحولات جوامع، که ظهور فناوری‌های جدید نیز بخشی از آن است، رویکردی حداقلی برای مقابله با چنین تحولاتی اتخاذ شود؟ برای پاسخ به این پرسش، لازم است ابتدا حدود و ثغور رویکرد حداقلی مشخص شود. دوم، دلایلی که توجیه‌کننده رویکرد حداقلی هستند، مورد بررسی قرار گیرد.

چارچوب و الزامات رویکرد حداقلی

اولین مسئله در خصوص ارجحیت رویکرد حداقلی نسبت به رویکرد حداکثری این است که رویکرد حداقلی به چه معناست و چه حدود و ثغوری را مدنظر قرار می‌دهد. به‌عبارتی، هنگامی‌که بیان می‌شود واکنش کیفری نسبت به دیپ‌فیک‌ها باید حداقلی باشد، دقیقاً چه شکلی از واکنش موردنظر است؟

رویکرد حداقلی در حوزه جرم‌انگاری چند اصل را شامل می‌شود. نخست اینکه نباید دیپ‌فیک‌ها به‌لحاظ ذاتی به‌عنوان یک تهدید نگریسته شوند، بلکه دیپ‌فیک محصول یک فناوری است که دارای کارکردهای مثبتی نیز هست؛ بنابراین، رویکرد حقوق جزا به ذات آن باید بی‌طرف باشد. دوم، جرم‌انگاری باید بر اساس کارکرد دیپ‌فیک صورت پذیرد؛ بدین معنا که صرفاً کارکردهایی از دیپ‌فیک که تهدیدی علیه حقوق افراد و جامعه هستند، در گستره اقدامات مجرمانه توصیف شوند. بنابراین، اصل بر غیرمجرمانه‌بودن دیپ‌فیک‌ها و کاربردهای آن‌هاست، مگر مواردی که قانون‌گذار بر اساس هدف مذکور، تحت عنوان مجرمانه‌ای قرار داده باشد. سوم، جز در موارد تهدید جدی که واکنش جزایی را ایجاب می‌کند، لازم است از جرم‌انگاری یا تفسیر موسعی که قابلیت تسری یک عنوان مجرمانه به موارد غیرضروری را ایجاد می‌کند، اجتناب شود.

رویکرد حداقلی در حوزه مجازات، تناسب میان جرم و مجازات را ایجاب می‌کند. اصل تناسب جرم و مجازات یکی از اصول حاکم بر تعیین مجازات‌هاست (Yazdian Jafari, 2008). تناسب، هماهنگی و همسویی میان مجازات و جرم و برخورداری از مبنایی منطقی برای مجازات، از اصول اولیه و بنیادین نظام عدالت کیفری و عامل مشروعیت آن است (Sabzevarinejad, 2017). در تعیین مجازات، اولاً باید هدف کلان مجازات روشن شود (Rahmdel, 2023). ثانیاً، مجازات باید با توجه به ماهیت جرم و شناعت آن تعیین شود. این مسئله، به‌خصوص در حقوق بشر، مورد تأکید قرار گرفته است. از جمله، تأکید شده است که مجازات باید متناسب با شدت و جدیت جرم باشد (Rahmdel, 2023). در تعیین مجازات‌های اسلامی نیز به معیارهای تناسب میان جرم و مجازات، نوع جرم ارتکابی، خصوصیات شخصیتی مجرم و درجه و نوع تقصیر بزه‌دیده توجه کامل شده است. بنابراین، اهداف کیفرشناختی، مانند تنبیه و اصلاح مجرم، پیشگیری عمومی و خصوصی و دفاع اجتماعی، باید تأمین شود (Alian & Faker Meybodi, 2018). ازاین‌رو، لازم است در استفاده‌های منفی دیپ‌فیک‌ها نیز انواع استفاده‌ها به‌تفکیک مشخص شده و با درنظرگرفتن عوامل مذکور، نوع و شدت مجازات در هر مورد تعیین شود. علاوه بر آن، رویکرد حداقلی ایجاب می‌کند تا حد ممکن، ظرفیت‌های بهره‌گیری از تمهیداتی همچون تعلیق، تخفیف و تأخیر مجازات و همچنین جایگزین‌های مجازات فراهم شود.

دلایل موجه توسل به رویکرد حداقلی

دومین مسئله‌ای که باید به آن پرداخته شود، بررسی دلایلی است که توسل به رویکرد حداقلی را توجیه می‌کنند. این دلایل شامل انتقادهایی نسبت به اصل رویکرد حداکثری و الزامات رویکرد حداقلی است. یکی از انتقادهایی که نسبت به جرم‌انگاری حداکثری بیان شده، تعدد عناوین مجرمانه برای جرایم و، به‌تبع آن، ایجاد عنوان ثانویه در بسیاری از جرایم، تعدد معنوی گسترده جرایم و تشویش و سردرگمی قضات و افراد جامعه است (Ghodrati et al., 2023). این مسئله در مورد جرم‌انگاری دیپ‌فیک‌ها نیز می‌تواند رخ دهد. تنوع مضمون دیپ‌فیک‌ها باعث می‌شود تا طیف گسترده‌ای از جرایم را شامل شوند و بنابراین، می‌تواند با مصادیق زیادی از دیگر جرایم هم‌پوشانی پیدا کند. هرچه توصیف گسترده‌تری از جرایم مبتنی بر دیپ‌فیک‌ها به عمل آید، به همان میزان نیز تعدد معنوی بیشتری متصور خواهد بود.

علاوه بر آن، رویکرد حداقلی، چه به‌طور کلی در حقوق جزا و چه به‌طور خاص در مورد دیپ‌فیک‌ها، قابل‌دفاع است. اصل حداقل‌بودن حقوق جزا، با احترام به انسان و ارزش‌هایی چون آزادی، خودمختاری و حق متفاوت‌بودن او، از یک‌سو و بیان نقش حداقلی دولت و عدم حق آن بر مداخله در تمام شئون زندگی انسان‌ها، از سوی دیگر، نظام حقوقی کیفری‌ای را ترسیم می‌کند که گستره آن اندک و محدود به موارد ضروری است (Bakhshizadeh et al., 2023). جرم‌انگاری دیپ‌فیک‌ها نیز تحت تأثیر همین اصل قرار دارد. بنابراین، اصل مذکور ایجاب می‌کند فناوری‌های ساخت دیپ‌فیک‌ها از دایره جرم‌انگاری خارج شوند و در رابطه با محصولات نیز، جز در صورت ضرورت، جرم‌انگاری صورت نپذیرد.

سومین علت ارجحیت رویکرد حداقلی، تعارض جرم‌انگاری دیپ‌فیک‌ها با حق آزادی بیان است که یکی از حقوق بنیادین بشر و مورد حمایت اسناد بین‌المللی و منطقه‌ای حقوق بشر است. بند دوم ماده ۱۹ میثاق حقوق مدنی و سیاسی، که با عبارت‌پردازی مشابه در اسناد حقوق بشری منطقه‌ای، همچون ماده ۱۰ کنوانسیون اروپایی حقوق بشر، نیز تکرار شده است، ضمن تأکید بر حق آزادی بیان، این حق را شامل آزادی جست‌وجو، دریافت و مشارکت در اطلاعات و ایده‌ها از هر نوع، در قالب هنر یا از طریق هر رسانه دیگری به انتخاب شخص، می‌داند. بنابراین، دیپ‌فیک‌ها به‌عنوان صورتی از بیان باید مورد حمایت قرار گیرند.

البته، بند سوم این ماده به دولت‌ها اجازه می‌دهد تا به دلایلی، از جمله احترام به حقوق و اعتبار دیگران و حمایت از امنیت ملی، نظم عمومی یا اخلاق و سلامت عمومی، در صورت ضرورت محدودیت‌هایی را نسبت به آزادی بیان اعمال کنند. ماده ۲۰ میثاق نیز هرگونه تبلیغات جنگ یا دفاع از دشمنی‌های ملی، مذهبی و نژادی را که منجر به تحریک تبعیض، دشمنی و خشونت شود، منع کرده است. بااین‌حال، بر اقدامات حداقلی برای محدودکردن آزادی بیان تأکید شده است.

بر اساس این دیدگاه، «رابطه میان حقوق و محدودیت‌ها و میان هنجارها و استثناها نباید معکوس شود» (Sabzevarinejad, 2017). به‌عبارتی، بند دوم ماده ۱۹ باید اصل و بند سوم، که حاوی استثناهای بند دوم است، باید استثنایی در نظر گرفته شود. به همین جهت، مندرجات بند مذکور باید با آزمون دقیق ضرورت و تناسب منطبق باشند (Yazdian Jafari, 2008). از سوی دیگر، استثنایی‌بودن بند سوم حکم می‌کند که محدودیت‌هایی بر مبنایی که در این بند ذکر نشده است، مجاز نباشند، حتی اگر آن دلایل نسبت به سایر حقوق تحت پوشش میثاق توجیه شده باشند. محدودیت‌ها باید فقط برای اهداف توصیف‌شده اعمال شوند و باید مستقیماً با الزامات خاص پیش‌بینی‌شده مرتبط باشند (Sabzevarinejad, 2017).

همچنین بیان شده است که «تحت هیچ شرایطی، حمله به افراد به‌دلیل اجرای حق آزادی بیان یا عقیده آنان، از جمله بازداشت خودسرانه، شکنجه، تهدید جانی و مرگ، نمی‌تواند با ماده ۱۹ منطبق باشد» (Sabzevarinejad, 2017). محدودیت‌ها نباید گسترده باشند و باید برای دستیابی به کارکردهای حمایتی نیز مناسب باشند. هنگامی‌که محدودیتی اعمال می‌شود، لازم است دلایل خاص و منحصربه‌فرد مربوط به ماهیت دقیق تهدیدی که این اقدام برای مقابله با آن انجام می‌گیرد نیز بیان شود (Sabzevarinejad, 2017).

بنابراین، هرگونه جرم‌انگاری نسبت به دیپ‌فیک‌ها نیز باید با درنظرگرفتن استثنایی‌بودن آن و حمایت از حق آزادی بیان صورت پذیرد. به نظر می‌رسد در این‌گونه موارد باید میان دیپ‌فیک‌های متعارض با امنیت ملی و نظم عمومی و مواردی که متضمن تعرض به حقوق دیگران هستند، تفکیک قائل شد. عدم وجود سلسله‌مراتب میان مصادیق حقوق بشر و لزوم حمایت از حقوق دیگران ایجاب می‌کند رویکرد موسع‌تری در جرم‌انگاری اتخاذ شود. هرچند در این موارد نیز اعمال محدودیت بدون واکنش جزایی اولویت دارد.

چالش‌های توسل به رویکرد حداقلی

هرچند دلایل مذکور به‌عنوان توجیه لزوم واکنش حداقلی قابل‌تأمل هستند، نباید از ماهیت و آثار دیپ‌فیک‌ها غفلت کرد. ماهیت مجعول دیپ‌فیک‌ها، سهولت انتشار و پراکنده‌شدن آن‌ها و نبود محدودیت در محتوای آن‌ها، دیپ‌فیک‌ها را به پدیده‌ای خطرناک تبدیل می‌کند. زیرا هرچند آن‌ها می‌توانند استفاده‌های مثبتی را در اختیار جامعه قرار دهند، هزینه آن، به‌خطرافتادن اعتماد عمومی نسبت به صحت هرگونه اطلاعات منتشرشده در بستر رسانه‌های اجتماعی است. عدم اطمینان مذکور می‌تواند آثار جدی بر انسجام اجتماعی و نظم عمومی ایجاد کند. با درنظرگرفتن آثار جبران‌ناپذیر دیپ‌فیک‌ها، لزوم متناسب‌بودن جرم‌انگاری و مجازات با هدف آن، که در اینجا حفظ ساختار سازمان‌یافته جوامع است، پیروی از رویکرد حداقلی را با چالش مواجه می‌کند. به‌عبارتی، حداقلی‌ترین رویکردی که برای مقابله با دیپ‌فیک‌ها معقول به نظر می‌رسد، نسبت به دیگر جرایم سخت‌گیری بسیار بیشتری را ایجاب می‌کند.

نتیجه‌گیری

هوش مصنوعی انقلابی را در فناوری‌ها، شئون مختلف زندگی و موضوعاتی که اشخاص و جامعه با آن‌ها سروکار دارند، ایجاد کرده است. خلق تولیدات سمعی و بصری مبتنی بر هوش مصنوعی، امکان ارائه خدماتی بسیار متنوع و گسترده را در حوزه‌های مختلف، از جمله فیلم‌سازی، تبلیغات، سرگرمی و پاسخ‌گویی به مخاطبان، فراهم کرده است. اما با توجه به غیرواقعی‌بودن محتوای این خدمات، چنین کاربردهایی می‌توانند به ابزار استفاده‌های منفی نیز تبدیل شوند. نبود محدودیت در موضوعاتی که هوش مصنوعی مولد می‌تواند به مخاطب ارائه دهد، باعث می‌شود جامعه و اشخاص در معرض طیف گسترده‌ای از تهدیدها نیز قرار گیرند. دیپ‌فیک‌ها، به همان میزان که می‌توانند مفید باشند یا حتی شاید به میزان بیشتری، می‌توانند تهدیدی برای نظم اجتماعی و اعتماد عمومی قلمداد شوند. آن‌ها حتی اگر به‌صورت خاص زیان‌هایی به اشخاص و جامعه وارد نسازند، به‌لحاظ ماهیت غیرواقعی خود می‌توانند مرزهای میان واقعیت و خیال را درهم بریزند و از این طریق، اعتماد عمومی نسبت به صحت هرگونه اطلاعات منتشرشده و دریافت‌شده را خدشه‌دار کنند. پیشرفت این فناوری در گذر زمان، امکان تشخیص واقعی یا غیرواقعی‌بودن این اطلاعات را به متخصصان علوم رایانه محدود خواهد کرد و این به‌معنای آسیب‌پذیری بالقوه و بدون دفاع عموم جامعه در مقابل این فناوری خواهد بود. در چنین شرایطی، بنیان‌های حقوق کیفری مبتنی بر مداخله حداقلی و ضروری در جرم‌انگاری و مجازات، با تردیدهای جدی مواجه می‌شوند. عبور سیاست جنایی از مجرم به پدیده اجتماعی، چهره واقعی‌تری به خود می‌گیرد و ذات و آثار یک پدیده اجتماعی تهدیدگونه مشخص می‌کند که چه سیاست جنایی‌ای می‌تواند برای مقابله با آن مناسب باشد. دیپ‌فیک‌ها هم به‌لحاظ محتوا و آثار و هم به‌لحاظ سهولت و گستردگی انتشار، می‌توانند تهدیدی بسیار جدی تلقی شوند و بنابراین، آثاری جبران‌ناپذیر نه‌تنها برای افراد، بلکه برای جوامع به همراه داشته باشند. در چنین شرایطی، اتخاذ یک رویکرد جنایی حداکثری منطقی به نظر می‌رسد. اما درعین‌حال، اتخاذ چنین رویکردی می‌تواند آثار سوئی نیز به همراه داشته باشد. از سوی دیگر، درحالی‌که دلایلی منطقی اتخاذ رویکرد حداقلی را توجیه می‌کنند، باید توجه داشت که این رویکرد نباید به قیمت ناکارآمدی نظام پاسخ جزایی و ساده‌انگاری جدیت خطر دیپ‌فیک‌ها تمام شود. از سوی دیگر، جدیدبودن ظهور این پدیده باعث شده است واکنش‌های نظام‌های کیفری در این زمینه نیز تازگی داشته باشند و بنابراین، هنوز بازخوردهای آن‌ها و چگونگی لزوم تعدیل آن‌ها به‌سوی رویکرد حداکثری یا حداقلی مشخص نباشد. اما آنچه به نظر می‌رسد این است که هرچند رویکرد امنیت‌مدار و مبتنی بر تسامح صفر ممکن است با واقعیات جهان در حال تحول همخوانی نداشته باشد، به میزان زیادی توجه به این رویکرد، در مقابل رویکردی حداقلی، نسبت به دیپ‌فیک‌ها را ایجاب می‌کند.

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 emergence and rapid development of artificial intelligence have transformed the structure of social communication, information production, audiovisual representation, and criminal behavior. Among the most controversial products of this technological transformation is deepfake technology, which relies on advanced machine-learning and deep-learning techniques to generate highly realistic but fabricated images, videos, and audio recordings. Unlike conventional forms of manipulation, deepfakes can create persuasive representations of individuals saying or doing things that never occurred, thereby weakening the traditional distinction between authentic and fabricated content. The technical architecture of artificial intelligence, which is characterized by autonomy, learning capacity, unpredictability, and partially independent decision-making processes, has expanded the range of both legitimate and harmful applications of these technologies (Zakerinia & Gholampour, 2024). Deepfakes may be used in cinema, artistic production, entertainment, education, cultural preservation, and the recreation of historical or deceased figures; therefore, they cannot be treated solely as inherently criminal instruments (Cejudo, 2023). Nevertheless, their capacity to manipulate identity and reality has created serious risks to privacy, personal dignity, electoral integrity, public trust, national security, and the reliability of digital evidence. Nonconsensual sexual deepfakes, voice-cloning fraud, reputational destruction, political misinformation, and fabricated audiovisual evidence illustrate the broad range of harms that may result from their misuse (Kulger & Pace, 2021; Lussier, 2022; Maras & Alexandrou, 2018). Moreover, the growing technical quality of face-swapping and synthetic media has made both manual and automated detection increasingly difficult, particularly for ordinary users who lack specialized knowledge (Nguyen & Ademiluyi, 2023; Zhang et al., 2020). These developments have placed criminal justice systems in a difficult normative position: they must protect individuals and society against serious technology-enabled harms without suppressing lawful creativity, freedom of expression, technological innovation, and the legitimate circulation of information. Consequently, the central issue is not whether deepfakes require legal regulation, but what form of criminal policy can respond effectively, proportionately, and legitimately to the specific dangers posed by this dual-use technology.

This study seeks to identify the optimal criminal policy toward deepfakes by examining the competing rationales of maximalist and minimalist responses. A maximalist criminal policy is based on extensive criminalization, broad legal definitions, stringent sanctions, and a security-oriented conception of criminal law. Its principal assumption is that emerging and highly destructive threats require anticipatory intervention before their harms become uncontrollable. In this approach, the production, distribution, facilitation, training, accessibility, and use of deepfake technologies may all become subject to criminal regulation, especially when legislators consider the technology itself to possess a high potential for fraud, defamation, political destabilization, sexual abuse, or national-security threats. Maximal criminalization has historically been justified by the perceived severity, ease of commission, broad victim base, and potentially irreversible consequences of particular

offenses. Security-oriented criminal policy may consequently shift attention from the individual offender to the dangerous phenomenon itself and adopt zero-tolerance strategies that emphasize prevention through deterrence and repression (Ghannad & Akbari, 2017; Tajabadi & Meshkani, 2022). In quantitative terms, maximal criminalization may lead to a proliferation of criminal statutes, while in qualitative terms it may be reflected in broad and flexible definitions capable of encompassing future technological manifestations not foreseen at the time of legislation (Majidi & Tajabadi, 2019; Mohammadnejad & Tajik, 2019). Such an approach appears particularly attractive in relation to deepfakes because their harmful content can be replicated instantly, distributed globally, and directed toward numerous victims at minimal cost. Deepfake-related harms may also continue after the original content has been removed, since copies can remain permanently available across digital networks. However, extensive criminalization may also generate legal uncertainty, excessive prosecutorial discretion, overlapping offenses, disproportionate punishment, and restrictions on civil liberties. The study therefore examines whether the seriousness of deepfake harms is sufficient to justify a predominantly maximalist policy or whether such a policy would reproduce the well-known deficiencies of overcriminalization.

The research adopts a descriptive-analytical method based on library and documentary studies. It examines legal scholarship concerning artificial intelligence, deepfake technologies, criminalization, proportionality, privacy, freedom of expression, and security-oriented criminal policy. It also comparatively analyzes regulatory and legislative tendencies in selected legal systems, including the European Union, China, and the United States, to determine how different jurisdictions have responded to synthetic audiovisual content. The comparative evidence demonstrates that contemporary legal systems increasingly rely on disclosure duties, content labeling, platform obligations, authentication mechanisms, civil liability, and targeted criminal sanctions rather than a single universal model. Regulatory responses to deepfakes have particularly focused on transparency, consent, digital provenance, the responsibilities of technology providers, and the identification of artificial content. Media and internet platforms have also developed converging strategies to detect, label, restrict, or remove manipulated media, although their policies remain inconsistent and may be influenced by commercial interests and political pressures (Vizoso et al., 2021). Public attitudes toward deepfakes further reveal substantial concerns regarding privacy, deception, sexual exploitation, and the unauthorized use of personal identity (Kulger & Pace, 2021). In parallel, research on public discussions of deepfakes indicates that social concerns are not limited to individual injury but extend to misinformation, distrust, democratic instability, and the erosion of confidence in digital communication (Gamage, 2022). The analysis also distinguishes between the technology used to create deepfakes and the harmful products generated through it. This distinction is essential because general-purpose software may have lawful artistic, educational, and research applications, whereas particular uses may violate privacy, facilitate fraud, or threaten national security. Accordingly, the methodological framework evaluates criminal policy at three interconnected levels: regulation of technological tools, regulation of synthetic content, and regulation of harmful conduct committed through such content. It further assesses whether criminal law should intervene directly or whether civil, administrative, technological, and preventive mechanisms should be given priority.

The findings indicate that a purely maximalist criminal policy is neither normatively desirable nor practically effective. Although the broad social risks associated with deepfakes may support targeted and serious criminal intervention, treating all deepfake technologies or outputs as inherently criminal would disregard their legitimate artistic, educational, commercial, and cultural

applications. Excessive criminalization can create multiple overlapping offenses, confuse courts and citizens, expand prosecutorial power, and undermine the coherence of the criminal justice system (Ghodrati et al., 2023). The experience of maximum criminalization demonstrates that the accumulation of penal rules does not necessarily result in greater public safety and may instead weaken legal certainty and public confidence (Arabiyan & Akhtari, 2024). Furthermore, broad criminal prohibitions may interfere with freedom of expression and artistic creativity, particularly where synthetic content is used for satire, parody, performance, or experimentation. The ethical evaluation of deepfakes in art and culture shows that their legitimacy depends significantly on consent, transparency, context, and the likelihood of deception or harm rather than on the mere use of synthetic-media technology (Cejudo, 2023). A maximalist approach may also be ineffective because the global, decentralized, and rapidly changing nature of digital technologies makes comprehensive prohibition difficult to enforce. Developers and users may operate outside the territorial jurisdiction of the affected state, anonymous distribution may complicate attribution, and criminal investigations may depend on highly specialized technical expertise. Moreover, a security-oriented framework that employs vague terms such as harmful content, false identity, or public disorder may enable disproportionate state interference and may be used against journalists, artists, political critics, or minority groups. The principle of minimum criminal law therefore requires that criminalization be reserved for conduct causing or seriously threatening substantial harm and that noncriminal mechanisms be considered before penal intervention (Bakhshizadeh et al., 2023). From this perspective, the mere possession of deepfake software, access to technical knowledge, or creation of harmless synthetic content should not ordinarily attract criminal liability.

At the same time, the findings demonstrate that an exclusively minimalist response cannot adequately address the most serious manifestations of deepfake misuse. The speed, scale, anonymity, persuasive realism, and replicability of synthetic content can produce harms that civil remedies, voluntary platform regulation, or post-publication corrections are unable to repair. Nonconsensual intimate deepfakes may irreversibly damage dignity, psychological well-being, employment, and family relationships, while political deepfakes may influence electoral behavior before their falsity can be established (Lussier, 2022; Vizoso et al., 2021). Fraudulent voice and video impersonation may also enable financial crimes by exploiting trust in familiar identities, corporate authority, and audiovisual evidence. The capacity of deepfakes to undermine evidentiary authenticity creates an additional criminal-justice concern, because both fabricated evidence and false claims that authentic recordings are fabricated may obstruct investigations and judicial fact-finding (Maras & Alexandrou, 2018). These circumstances justify carefully defined criminal offenses directed at intentional and harmful misuse rather than at the technology itself. Such offenses should distinguish between levels of culpability, forms of harm, the presence or absence of consent, the scale of dissemination, the identity of the victim, the intention to deceive, and the threat posed to public interests. Proportionality must govern both criminalization and punishment, since legitimate penal policy requires a rational relationship between the seriousness of the offense and the severity of the sanction (Rahmdel, 2023; Sabzevarinejad, 2017; Yazdian Jafari, 2008). Accordingly, high-risk conduct such as large-scale fraud, nonconsensual sexual deepfakes, systematic reputational attacks, election interference, incitement to violence, fabrication of judicial evidence, and threats to national security may warrant explicit criminalization and substantial penalties. By contrast, low-risk uses should primarily be addressed through disclosure requirements, consent rules, rapid removal procedures, civil compensation, administrative enforcement, platform accountability, media literacy, and

technical authentication systems. This differentiated approach also reduces the danger that criminal law will be used as the first rather than the last instrument of social control.

In conclusion, the optimal criminal policy toward deepfakes should not be based on an absolute choice between maximalism and minimalism. The distinctive character of deepfake technology requires a hybrid, participatory, and layered model that combines proportionate criminal intervention with civil, administrative, technological, educational, and regulatory measures. Criminal law should be directed toward clearly defined high-risk abuses involving intentional deception, serious individual harm, large-scale financial loss, sexual exploitation, manipulation of democratic processes, destruction of evidentiary reliability, or threats to public and national security. At the same time, the law should preserve legitimate uses in art, education, research, entertainment, and cultural production and should avoid criminalizing the mere development, possession, or neutral use of synthetic-media technologies. Platform providers, software developers, public institutions, civil society, media organizations, and users should share responsibility for prevention. Mandatory labeling, digital watermarking, provenance verification, accessible complaint mechanisms, rapid content removal, identity protection, public education, and specialized investigative capacity should form essential components of the regulatory framework. Sanctions should be graded according to intent, harm, scale, consent, repetition, and the vulnerability of victims, while alternatives to imprisonment should remain available for less serious cases. Such a model recognizes that the dangers of deepfakes are real and potentially irreversible, but it also acknowledges that overbroad criminalization can threaten freedom, innovation, and legal certainty. A balanced policy must therefore be sufficiently strong to deter and punish serious misuse, yet sufficiently restrained to protect lawful expression and technological development. The most defensible response is consequently neither comprehensive repression nor regulatory passivity, but an adaptive criminal policy that reserves severe penal measures for the most harmful conduct and relies on preventive and noncriminal instruments for the broader range of lower-risk applications.

References

Albahar, M., & Almalki, J. (2019). Deepfakes: Threats and countermeasures systematic review. *Computer Science, Political Science, Law*, *97*(22), 1817-1852.

Alian, I., & Faker Meybodi, M. (2018). Proportionality of crime and punishment based on the verses of qisas. *Tafsir Studies*, *9*(36), 29-46.

Arabiyan, A., & Akhtari, S. (2024). A comparative study of maximum criminalization in Iran and the United States. *Interdisciplinary Studies of Jurisprudence*, *4*(16), 143-164.

Bakhshizadeh, T., Bashiriyeh, T., & Gholami, H. (2023). The principle of the minimum nature of criminal law in Iranian criminal law: Challenges and solutions. *Teachings of Criminal Jurisprudence and Law*(7), 65-84.

Cejudo, R. (2023). Ethical problems of the use of deepfakes in the arts and culture. In *Ethics of Artificial Intelligence*. Springer Nature Switzerland.

Gamage, D. (2022). Are deepfakes concerning? Analyzing conversations of deepfakes on Reddit and exploring societal implications. CHI Conference on Human Factors in Computing Systems,

Ghannad, F., & Akbari, M. (2017). Securitization of criminal policy. *Criminal Law Research Quarterly*, *5*(18), 39-67.

Ghodrati, N., Mirkhalili, S. M., & Haji Dehabadi, M. A. (2023). Legal consequences of pervasive criminalization in Iran's judicial system. *Islamic Law*, *20*(78), 7-34.

Kulger, M. B., & Pace, C. (2021). Deepfake privacy: Attitudes and regulation. *Northwestern University Law Review*, *116*(3), 611-680.

Lussier, N. (2022). Nonconsensual deepfakes: Detecting and regulating this rising threat to privacy. *Idaho Law Review*, *58*(2), 353-383.

Majidi, S. M. (2009). Manifestations of security-oriented criminal law in France. *Private Law Studies*, *39*(2), 327-344.

Majidi, S. M., & Tajabadi, F. (2019). Security-oriented criminalization techniques in Iranian criminal law. *Studies in Islamic Jurisprudence and Law*(21), 289-316.

Maras, M. A., & Alexandrou. (2018). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos. *International Journal of Evidence and Proof*, *23*(2), 255-262.
Mohammadnejad, H., & Tajik, R. (2019). Causes of maximum criminalization and the requirements for its structural transformation in Iran. *Private Law Research*(38), 131-152.
Nguyen, D. S. W., & Ademiluyi, D. T. (2023). Deepfake video detection employing human facial features. *Journal of Computer and Communications*, *11*(12), 1-13.
Rahmdel, M. (2023). *Proportionality of Crime and Punishment*. SAMT.
Sabzevarinejad, H. (2017). The position of the principle of proportionality between crime and punishment in Iranian and English criminal law. *Judicial Law Perspectives Quarterly*, *22*(77), 133-164.
Sheikh, S., Ghiyasi, J. a.-D., & Salehi, M. K. (2022). A reflection on the pathology of substantive legislative criminal policy concerning currency crimes against security. *Iranian Political Sociology*, *5*(10), 193-205.
Tajabadi, F., & Meshkani, A. H. (2022). Security-oriented criminal policy techniques (zero tolerance): Challenges of public law and human rights. National Conference on Professor Katouzian and Public Law Thought, Bushehr.
Vizoso, Á., Vaz-Álvarez, M., & García, X. L. (2021). Fighting deepfakes: Media and internet giants' converging and diverging strategies against hi-tech misinformation. *Media and Communication*, *9*(1), 291-300.
Yang Chik, A., & Samsudin, S. A. B. (2023). Proceedings of the 3rd International Conference on Law and Digitalization 2023 (ICLD 2023).
Yazdian Jafari, J. (2008). The principle of proportionality between crime and punishment: Its why and how. *Nameh-ye Mofid*, *14*(67), 139-156.
Zakerinia, H., & Gholampour, Z. (2024). Reasonable and customary algorithms and strengthening the theory of imputability of civil liability for artificial intelligence. *Journal of Contract Law and Emerging Technologies*, *5*(1), 155-168.
Zhang, W., Zhao, C., & Li, Y. (2020). A novel counterfeit feature extraction technique for exposing face-swap images based on deep learning and error level analysis. *Entropy*, *22*(2), 249-268.