

UNIVERSIDAD CATÓLICA DE CUENCA

Comunidad Educativa al Servicio del Pueblo

FACULTAD DE CIENCIAS SOCIALES

CARRERA DE DERECHO

Deepfake como vulneración a la dignidad humana: insuficiencia del COIP y necesidad de regulación penal específica en Ecuador

PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE ABOGADO

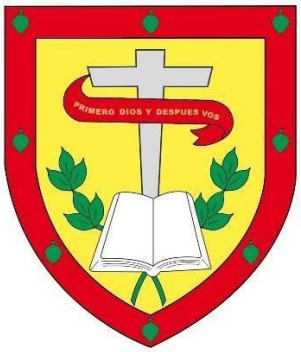
AUTORES: PABLO JOSÉ JIMÉNEZ ROMERO
KEVIN JOEL PACHALA TAPIA

DIRECTOR: DR. BERNARDO XAVIER MONSALVE ROBALINO, MGS

CUENCA-ECUADOR

2026

DIOS, PATRIA, CULTURA Y DESARROLLO



UNIVERSIDAD CATÓLICA DE CUENCA

Comunidad Educativa al Servicio del Pueblo

FACULTAD DE CIENCIAS SOCIALES

CARRERA DE DERECHO

Deepfake como vulneración a la dignidad humana: insuficiencia del COIP y necesidad de regulación penal específica en Ecuador

PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE ABOGADO

AUTORES: PABLO JOSÉ JIMÉNEZ ROMERO
KEVIN JOEL PACHALA TAPIA

DIRECTOR: DR. BERNARDO XAVIER MONSALVE ROBALINO, MGS

CUENCA-ECUADOR

2026

DIOS, PATRIA, CULTURA Y DESARROLLO

Universidad Católica de Cuenca	DECLARATORIA DE AUTORÍA

Nosotros, **Pablo José Jiménez Romero**, portador(a) de la cédula de ciudadanía N° **1400800411** y **Kevin Joel Pachala Tapia,** portador(a) de la cédula de ciudadanía N° **0750695439.** Declaramos ser los autores de la obra: "**Deepfake como vulneración a la dignidad humana: insuficiencia del COIP y necesidad de regulación penal específica en Ecuador.",** sobre la cual me hago responsable sobre las opiniones, versiones e ideas expresadas. Declaro que la misma ha sido elaborada respetando los derechos de propiedad intelectual de terceros y eximo a la Universidad Católica de Cuenca sobre cualquier reclamación que pudiera existir al respecto. Declaro finalmente que mi obra ha sido realizada cumpliendo con todos los requisitos legales, éticos y bioéticos de investigación, que la misma no incumple con la normativa nacional e internacional en el área específica de investigación, sobre la que me responsabilizo y eximimo a la Universidad Católica de Cuenca de toda reclamación al respecto.

Cuenca, 7 de Octubre de 2026

F: .....

Pablo José Jiménez Romero

C.I. 1400800411

F: .....

Kevin Joel Pachala Tapia

C.I. 0750695439

Universidad Católica de Cuenca

CERTIFICO

Yo, **Dr. Bernardo Xavier Monsalve Robalino, Mgs.**, certifico que el presente proyecto de titulación, realizado por **Pablo José Jiménez Romero** con cedula N° **1400800411 y Kevin Joel Pachala Tapia** con cedula N° **0750695439**, con el tema del proyecto, **"Deepfake como vulneración a la dignidad humana: insuficiencia del COIP y necesidad de regulación penal específica en Ecuador"** bajo mi supervisión.

F: .....

Dr. Bernardo Xavier Monsalve Robalino, Mgs
Docente – Tutor

Dedicatoria

A mis padres, Por ser el refugio, la fuerza y el pilar inquebrantable sobre el cual reconstruí mi camino. Cuando decidí dar un paso al costado y dejar atrás otra carrera, en medio de la incertidumbre y el miedo, ustedes no solo no me soltaron la mano, sino que creyeron en mí con una fuerza que yo mismo no tenía; confiaron en mi potencial mucho más de lo que yo era capaz de confiar en mí mismo.

Sé de las noches de desvelo, de los silencios cargados de preocupación y de los inmensos sacrificios que hicieron para que a mí no me faltara la oportunidad de salir adelante. Nada de lo que he logrado habría sido posible sin su amor incondicional, su paciencia infinita y su fe ciega en mi futuro. La culminación de esta etapa universitaria llevan su sello, su esfuerzo y su corazón. Es de ustedes y para ustedes, con mi más profunda gratitud y amor eterno

Pablo Jiménez

Hoy, al llegar a la cima de este camino y ver finalmente culminada esta etapa fundamental de mi vida, mi corazón se llena de una profunda y eterna gratitud.

A Dios, por sobre todas las cosas, por haberme otorgado la salud, la sabiduría y la fortaleza para no rendirme jamás. Gracias por ser mi luz en la oscuridad, mi guía en cada paso y el sustento en mis momentos de mayor cansancio; sin Su bendición y gracia, la consecución de este sueño no habría sido posible

A mis padres, porque esta meta alcanzada es, en realidad, la cosecha de cada uno de sus sacrificios en silencio, de su paciencia infinita y del amor incondicional con el que sostuvieron mi vida. Gracias por haberme enseñado el valor de la perseverancia y por creer en mí aun en los momentos en que la duda me embargaba. Este logro lleva sus nombres grabados en cada página

A mis hermanos, mis eternos compañeros de vida, por su cariño sincero, sus palabras de aliento y por ser ese refugio constante que siempre me dio el impulso para seguir adelante

Y a todos ustedes, mi amada familia: aunque la vida en otra ciudad me llevó a recorrer este camino lejos de casa y la distancia nos separó físicamente en mis noches de estudio y desvelo, su presencia jamás me faltó. Gracias por haberme abrazado a través de los kilómetros, por cada llamada reconfortante y por ser la fuerza silenciosa que me sostuvo desde lejos. Esta culminación y este triunfo son la prueba de que el amor de la familia no conoce distancias, y hoy este logro les pertenece tanto como a mí

Kevin Pachala

Agradecimiento

Expresamos nuestro más profundo reconocimiento al Dr. Bernardo Monsalve, por su constante orientación, paciencia y valiosa dedicación en la elaboración de esta tesis. Sus oportunos conocimientos e invaluables aportes fueron pilares esenciales para llegar a la feliz conclusión de este proyecto. Asimismo, valoramos enormemente la confianza, el respeto y la amabilidad brindados, cualidades que transformaron este proceso académico en una vivencia enriquecedora y de gran calidez humana.

A la Universidad Católica de Cuenca, por ser la institución que nos acogió y nos brindó los espacios necesarios para nuestro desarrollo integral. Extendemos nuestra gratitud a todo su personal docente y administrativo por su vocación formadora, por las enseñanzas compartidas y por su aporte fundamental en la consecución de este valioso logro profesional

Resumen

El problema de investigación radica en que la proliferación de deepfakes maliciosos en Ecuador plantea la duda de si los tipos penales vigentes en el Código Orgánico Integral Penal (COIP) "violación a la intimidad, calumnia, suplantación de identidad y extorsión" son aptos para sancionar la creación y difusión no consentida de contenido sintético lesivo de la dignidad humana, o si corresponde valorar la reforma de dichos tipos o la creación de un tipo penal autónomo. El objetivo general es determinar la viabilidad jurídica de la tipificación autónoma del uso malicioso de deepfakes en el COIP como mecanismo de protección de la dignidad humana en Ecuador. Metodológicamente, se emplea un enfoque cualitativo dogmático-jurídico, con métodos hermenéutico y comparado, mediante revisión documental de doctrina, normativa, jurisprudencia constitucional y legislación de la Unión Europea, Estados Unidos, Francia, Reino Unido, Corea del Sur y China. Los resultados evidencian que el deepfake malicioso lesiona de forma pluriofensiva la intimidad, el honor, la propia imagen y la protección de datos personales, y que los artículos 178, 182 y 212 del COIP exigen elementos que no concurren en varias modalidades de deepfake analizadas. Se concluye que la tipificación autónoma constituye una alternativa jurídicamente viable frente a la reforma de tipos vigentes y a los mecanismos extrapenales, condicionada al respeto de los principios de legalidad, mínima intervención penal, lesividad y proporcionalidad que limitan el ius puniendi del Estado.

Palabras clave*: Deepfake, dignidad humana, tipo penal autónomo, COIP, derechos personalísimos*

Abstract

The research problem arises from the fact that the proliferation of malicious deepfakes in Ecuador raises the question of whether the criminal offenses currently established in the Comprehensive Organic Penal Code (COIP, by its Spanish acronym) –violation of privacy, defamation, identity theft, and extortion– are suitable for punishing the creation and non-consensual dissemination of synthetic content that harms human dignity, or whether amending these offenses or establishing a separate criminal offense should be considered. The general objective is to determine the legal viability of establishing a separate criminal offense for the malicious use of deepfakes in the COIP as a mechanism for protecting human dignity in Ecuador. Methodologically, a qualitative doctrinal legal approach is used, with hermeneutic and comparative methods, through documentary review of legal scholarship, legislation, constitutional case law, and the legislation of the European Union, the United States, France, the United Kingdom, South Korea, and China. The results show that malicious deepfakes simultaneously infringe multiple legal interests, including privacy, honor, the right to one's own image, and personal data protection, and that articles 178, 182, and 212 of the COIP require specific legal elements that are absent in several of the deepfake scenarios analyzed. It is concluded that creating a separate criminal offense is a legally viable alternative to amending existing offenses or relying on non-criminal legal mechanisms, provided that it respect for the principles of legality, minimal penal intervention, harmfulness, and proportionality that limit the State's *ius puniendi*.

Keywords: *Deepfakes, human dignity, standalone criminal offense, COIP, personal rights*

Índice

Contenido

Certificado del tutor..... III

Agradecimiento V

Dedicatoria..... IV

Resumen..... VI

Abstract VII

Índice VIII

Introducción..... 1

Capitulo I. 7

El uso malicioso de *deepfake* como conducta lesiva de la dignidad humana y de los derechos personalísimos en el entorno digital. 7

1. Aproximación conceptual del deepfake...... 7

1.1 El deepfake: concepto, distinción y base tecnológica. 7

1.2 El deepfake como riesgo jurídico en el entorno digital. 9

1.3 Creación de contenido sin consentimiento. 10

1.3.1. Manipulación de identidad, rostro, voz o imagen...... 10

1.3.2 Afectación a la reputación, intimidad y vida privada. 11

1.3.3 Uso de deepfakes con fines sexuales, difamatorios, fraudulentos o de humillación.... 11

1.4. La dignidad humana como bien jurídico afectado. 13

1.4.1. Concepto doctrinario de dignidad humana...... 13

1.4.2 La persona como fin en sí misma. 13

1.4.3. Dignidad humana como fundamento de los derechos fundamentales. 14

1.4.4. Relación entre dignidad humana, autonomía y consentimiento. 14

1.4.5. Afectación de la dignidad cuando se usa la imagen de una persona para simular actos que nunca realizó...... 15

1.5. Derechos personalísimos vulnerados por el uso malicioso de deepfakes...... 15

1.5.1. Derecho al honor y al buen nombre...... 15

1.5.2. Derecho a la intimidad personal y familiar. 16

1.5.3. Derecho a la propia imagen. 16

1.5.4. Derecho a la protección de datos personales y a la autodeterminación informativa. 17

1.5.5. Derecho a la libertad sexual en casos de deepfakes de contenido sexual. 17

1.5.6. Especial afectación a mujeres, niñas, niños y adolescentes...... 18

1.6. Protección constitucional e internacional de los derechos afectados. 18

1.7. El Deepfake como forma de violencia digital: Más allá de la manipulación tecnológica. 20

1.7.1. Violencia de genero digital y Abuso Sexual Basado en Imágenes (IBSA). 20

1.7.2. Cosificación, humillación y el cinismo de la "diversión en la crueldad". 21
1.7.3. Pérdida de control sobre la identidad digital y daños multidimensionales. 21
1.8. Conclusión del Capítulo I. 22
Capítulo II. 24
Capacidad e insuficiencia de los tipos penales vigentes en el COIP frente al uso malicioso de deepfake. 24
2.1. El principio de legalidad y la exigencia de tipicidad estricta en materia penal. 24
2.1.1. Nullum crimen, nulla poena sine lege y la tipicidad estricta. 24
2.1.2. Prohibición de analogía y de interpretación extensiva. 24
2.1.3. El problema del rezago legal frente a la realidad digital. 25
2.2. Derecho penal, nuevas tecnologías y rezago normativo. 25
2.2.1. Avance acelerado de la inteligencia artificial y lentitud legislativa. 26
2.2.2. Nuevas formas de criminalidad digital. 26
2.2.3. Dificultad del derecho penal tradicional ante realidades sintéticas. 26
2.2.4. Riesgo de impunidad por ausencia de regulación específica. 27
2.3. Revisión del delito de violación a la intimidad en el Código Orgánico Integral Penal (COIP). 27
2.3.1. Bien jurídico protegido en la violación a la intimidad. 27
2.3.2. Análisis de los verbos rectores y su alcance. 28
2.3.3. Aplicabilidad y límites frente a contenido creado artificialmente. 28
2.3.4. Problema: La falta de cobertura expresa de la creación de contenido falso mediante IA. 29
2.4. Revisión del delito de calumnia en el Código Orgánico Integral Penal (COIP). 29
2.4.1. Concepto penal de calumnia y la falsa imputación. 29
2.4.2. Deepfakes que simulan la comisión de ilícitos. 32
2.4.3. Afectación al honor y la reputación. 33
2.4.4. Límites: La atipicidad de contenidos no calumniosos. 33
2.4.5. Insuficiencia frente a deepfakes sexuales y vejatorios. 33
2.5. Revisión del delito de suplantación de identidad. 34
2.5.1 Concepto penal de suplantación de identidad. 34
2.5.2 El deepfake como forma de simulación de identidad. 34
2.5.3 Dificultades probatoria: el problema del beneficio exigido por el tipo penal. 35
2.5.4. Diferencia entre suplantar identidad y manipular la imagen. 35
2.6. Revisión de otros tipos penales relacionados. 36
2.6.1. Extorsión y Sextorsión (Extorsión sexual). 36
2.6.2. Ciberacoso y Hostigamiento. 37
2.6.3. Pornografía con utilización de niñas, niños o adolescentes. 38

2.7.1. Falta de verbos rectores y dificultad para sancionar la sola creación. 40
2.7.2. Problemas para acreditar perjuicio concreto o beneficio. 41
2.7.3. Contenido sexual falso y falta de participación material. 41
2.7.4. Dificultades probatorias por el hiperrealismo. 41
2.7.5. Atipicidad, subsunción forzada y principio de legalidad. 42
2.8. Conclusión del capítulo II: 42
Dificultades de subsunción y examen de una tipificación autónoma. 42
Capítulo III 45
Elementos jurídicos para la configuración de un tipo penal autónomo sobre el uso malicioso de deepfakes en Ecuador. 45
3.1. Necesidad jurídica de una tipificación autónoma. 45
3.1.1. Insuficiencia de los tipos penales vigentes. 45
3.1.2. Necesidad de certeza jurídica y Prevención de impunidad. 45
3.1.3. Protección efectiva de las víctimas. 46
3.1.4. Evitar interpretaciones extensivas del derecho penal. 46
3.1.5. Justificación de una respuesta penal específica. 47
3.2. Fundamento político-criminal de la regulación penal de deepfakes. 48
3.2.1. Política criminal frente a nuevas formas de violencia digital. 48
3.2.2. Protección de bienes jurídicos relevantes y lesividad de la conducta. 49
3.2.3. Riesgo social del uso masivo de inteligencia artificial maliciosa. 49
3.2.4. Necesidad de protección reforzada frente a daños graves. 50
3.2.5. Diferencia entre regulación preventiva, administrativa, civil y penal. 51
3.3. Límites constitucionales al ius puniendi del Estado. 52
3.3.1. Principio de legalidad y prohibición de tipos penales ambiguos. 53
3.3.2. Principio de mínima intervención penal. 53
3.3.3. Principio de lesividad y protección de bienes jurídicos. 53
3.3.4. Principio de proporcionalidad y necesidad de la pena. 54
3.3.5 Principio de culpabilidad. 55
3.4. Derecho comparado sobre regulación de deepfakes. 56
3.4.1. Unión Europea y regulación de inteligencia artificial. 56
3.4.2. Transparencia en contenidos generados por IA. 57
3.4.3. Estados Unidos: regulación fragmentada y específica sobre deepfakes. 57
3.4.4. Regulación de deepfakes sexuales no consentidos en otras latitudes. 59
3.4.5. Protección de imagen, voz, datos personales e identidad. 61
3.4.6. Lecciones útiles Y desafíos para Ecuador. 62
3.5. Bien jurídico protegido en el tipo penal autónomo. 64

3.5.1. La dignidad humana como eje principal. 65
3.5.2. Intimidad, honor y buen nombre. 66
3.5.3. Propia imagen y protección de datos personales. 66
3.5.4. Libertad e indemnidad sexual. 67
3.5.5. Carácter pluriofensivo del delito. 68
3.6. Sujetos del tipo penal. 69
3.6.1. Sujeto activo: El autor detrás de la tecnología. 69
3.6.2. Pluralidad de intervinientes: Creadores, difusores y comercializadores. 70
3.6.3. Diferenciación entre autoría penal y simple intermediación tecnológica. 72
3.6.4. Sujeto pasivo: La víctima de la simulación. 72
3.6.5. Protección reforzada para niños, niñas, adolescentes y grupos vulnerables. 73
3.7. Conducta típica y verbos rectores. 74
3.7.1. Verbos rectores de la fase de creación y modificación. 75
3.7.2. Verbos rectores de la fase de difusión y explotación. 76
3.7.3. El elemento tecnológico y la ausencia de consentimiento. 78
3.7.4. El elemento subjetivo: Finalidades específicas. 79
3.8. Objeto material del delito. 81
3.9. Elemento subjetivo del tipo penal. 83
3.10. Circunstancias Agravantes. 85
3.11. Exclusiones o límites para evitar una regulación excesiva. 88
3.12. Propuestas de estructura normativa del tipo penal. 90
3.13. Viabilidad jurídica de la propuesta en el sistema penal ecuatoriano. 94
3.13.1. Compatibilidad con el principio de legalidad y la constitucionalización del derecho penal. 96
3.13.2. Compatibilidad con la mínima intervención penal (Ultima Ratio). 97
3.13.3. Necesidad de protección de bienes jurídicos constitucionalmente relevantes. 99
3.13.4. Coherencia con el COIP y Necesidad de proporcionalidad en la pena. 99
3.13.5. Viabilidad jurídica condicionada al respeto de los límites del ius puniendi. 100
3.13.6. Cierre del Capítulo III. 100
Conclusiones generales 102
Recomendaciones 104
Bibliografías 106
Anexo 109
..... 110

Introducción

La presente investigación se fundamenta en la Doctrina Constitucional de la Dignidad Humana como valor supremo del ordenamiento jurídico y fuente de los derechos personalísimos “honor, intimidad, propia imagen, protección de datos e identidad”, así como en la teoría del bien jurídico protegido y en los principios que limitan el ius puniendi del Estado: legalidad, tipicidad estricta, mínima intervención penal (ultima ratio), lesividad, proporcionalidad y culpabilidad. Estos marcos teóricos permiten analizar si una conducta digital, en este caso el uso malicioso de deepfakes, alcanza la entidad suficiente para justificar una intervención penal específica.

El problema surge del vertiginoso avance de la Inteligencia Artificial generativa, que ha permitido la creación de contenidos audiovisuales hiperrealistas capaces de suplantar la identidad, el rostro y la voz de las personas sin su consentimiento. Este fenómeno no es neutral: la evidencia doctrinaria y estadística revisada indica que la mayoría de los deepfakes en circulación son de naturaleza pornográfica no consentida y afectan predominantemente a mujeres, niñas, niños y adolescentes, lo que evidencia un componente de violencia digital de género. Frente a ello, cabe preguntarse si el ordenamiento penal ecuatoriano cuenta con las herramientas necesarias para responder a esta conducta, o si, por el contrario, persisten dificultades de adecuación típica que ameriten una respuesta legislativa distinta.

Esta investigación analiza y evalúa la viabilidad jurídica de una tipificación autónoma del uso malicioso de deepfakes en el COIP, comparando esta alternativa frente a otras tres posibles respuestas: (i) la aplicación de los tipos penales vigentes mediante interpretación judicial; (ii) la reforma puntual de los artículos 178 (violación a la intimidad), 182 (calumnia), 212 (suplantación de identidad) y 185 (extorsión); y (iii) el fortalecimiento de mecanismos extrapenales (civiles y administrativos, como el Habeas Data o la Ley Orgánica de Protección

de Datos Personales). El estudio no parte de que la creación de un nuevo tipo penal sea la única solución posible, sino que examina, con criterios verificables, si existe una dificultad de subsunción esto es, una falta de adecuación entre la conducta del deepfake y los elementos estructurales de los tipos vigentes y, en su caso, si dicha dificultad deriva en un vacío de punibilidad bajo el estándar de tipicidad estricta, distinguiendo esta cuestión de la insuficiencia de protección en sede civil o administrativa, que es de naturaleza distinta y no forma parte central de este análisis.

Para ello, se examina artículo por artículo qué elementos típicos concurren y cuáles no en modalidades concretas de deepfake (deepfake sexual no consentido, deepfake difamatorio, deepfake con fines de fraude o sextorsión y deepfake que involucra a menores de edad), evitando afirmaciones absolutas y sustentando cada conclusión en el análisis dogmático de los verbos rectores, el elemento subjetivo y el bien jurídico protegido de cada figura.

La Constitución de la República del Ecuador, en su artículo 66, numerales 18, 19, 20 y 28, garantiza el derecho al honor y buen nombre, a la protección de datos personales y autodeterminación informativa, a la intimidad personal y familiar, y a la identidad personal y colectiva. La Corte Constitucional del Ecuador, mediante la Sentencia No. 2064-14-EP/21, determinó que el rostro constituye un dato personal protegido y que acciones como consultar, visualizar, almacenar en dispositivos o mostrar imágenes a terceros sin autorización configuran formas de tratamiento de datos personales. Esta regla jurisprudencial se sitúa en el ámbito del Habeas Data y de la protección de datos personales; el fallo no formula conclusiones sobre la tipicidad penal de la creación de contenido sintético, cuestión que constituye precisamente uno de los objetos de análisis de esta investigación.

De lo expuesto surge la siguiente pregunta de investigación: ¿es jurídicamente viable la tipificación autónoma del uso malicioso de deepfakes en el COIP como mecanismo idóneo, necesario y proporcional para proteger la dignidad humana en Ecuador, frente a las alternativas

de aplicación o reforma de los tipos penales vigentes? Como hipótesis de trabajo, se plantea que los tipos penales vigentes en el COIP presentan elementos estructurales (verbos rectores orientados al acceso, registro o difusión de información real, exigencia de beneficio patrimonial o imputación de un delito concreto) que no permiten cubrir de manera completa las modalidades de deepfake malicioso identificadas, y que, de confirmarse dicha limitación, la creación de un tipo penal autónomo constituye una alternativa compatible con los principios de legalidad, mínima intervención penal y proporcionalidad. Esta hipótesis será contrastada, y no asumida, a lo largo de la investigación.

El objetivo general es determinar la viabilidad jurídica de la tipificación autónoma del uso malicioso de deepfakes en el COIP como mecanismo de protección de la dignidad humana en el Ecuador. De este se desprenden tres objetivos específicos: (1) fundamentar jurídicamente el uso malicioso de deepfakes como una conducta lesiva de la dignidad humana y de los derechos personalísimos en el entorno digital; (2) examinar la capacidad de los tipos penales vigentes en el COIP para abarcar las conductas derivadas del uso malicioso de deepfakes; y (3) proponer los elementos jurídicos para la configuración de un tipo penal autónomo, conforme a los principios que limitan el ius puniendi del Estado. Estos objetivos se apoyan en trabajos previos sobre el fenómeno del deepfake y el Abuso Sexual Basado en Imágenes (Rigotti et al., 2024; Andino Veloz, 2025), sobre el rezago normativo del derecho penal frente a la Inteligencia Artificial (Chávez Zambrano & San Andrés Molina, 2025; Ponce González et al., 2025) y sobre experiencias de regulación comparada en la Unión Europea, Estados Unidos, Francia, Reino Unido, Corea del Sur y China (Ugwuoke & Sanfilippo, 2025).

Desde una perspectiva metodológica, la presente investigación adopta un enfoque cualitativo, de corte dogmático-jurídico y con un alcance descriptivo-propositivo. El diseño se orienta al examen crítico de los tipos penales vigentes y a la interpretación sistemática de la norma sustantiva frente a las nuevas conductas lesivas originadas por la inteligencia artificial

generativa. Al tratarse de un estudio estrictamente jurídico-teórico, se prescindió del trabajo de campo sobre sujetos humanos y del levantamiento de muestras estadísticas, enfocando el rigor académico en la construcción sistemática y reproducible de un corpus documental especializado.

La conformación de esta base documental se articuló a partir de una búsqueda sistemática en bases de datos científicas e indexadas de alto impacto, como Scopus, Web of Science, Dialnet, SciELO, Redalyc e IEEE Xplore, complementada con el rastreo en repositorios de posgrado nacionales e internacionales, tales como los de la Universidad del Azuay, la Universidad Católica de Santiago de Guayaquil, la Universidad Central del Ecuador y la Universitat Autònoma de Barcelona. Asimismo, se consultaron repositorios normativos y jurisprudenciales oficiales, entre ellos el Registro Oficial del Ecuador, el portal de la Corte Constitucional del Ecuador, el Boletín Oficial del Estado de España y la plataforma EUR-Lex de la Unión Europea. La localización del material se efectuó mediante ecuaciones de búsqueda con operadores booleanos que combinaron términos como "deepfake", "contenido sintético", "inteligencia artificial", "violación a la intimidad", "calumnia", "abuso sexual basado en imágenes", "redes generativas antagónicas" y "tipicidad penal".

La delimitación temporal del estudio abarcó el periodo comprendido entre 2014 y 2026, con búsquedas ejecutadas entre enero de 2024 y septiembre de 2026. Esta ventana temporal responde a la entrada en vigencia del Código Orgánico Integral Penal en el Ecuador en el año 2014, coincidiendo a su vez con la etapa de mayor aceleración tecnológica en el desarrollo de modelos generativos y la consecuente discusión sobre su regulación jurídica. Para depurar la información se fijaron criterios específicos de inclusión, priorizando la normativa constitucional y penal vigente en el Ecuador y en legislaciones comparadas, la jurisprudencia constitucional relevante, y la literatura académica arbitrada publicada en español e inglés. Del mismo modo, se excluyeron los artículos de opinión periodística o blogs sin revisión por pares,

publicaciones previas al año 2014, salvo la doctrina clásica sobre dignidad humana y derechos personalísimos, que se conservó por su aporte dogmático a la materia, y aquellos estudios técnicos o comerciales desconectados de la protección de bienes jurídicos y derechos fundamentales.

Tras un proceso de selección en cuatro fases que incluyó identificación, tamizaje preliminar de títulos y resúmenes, valoración de pertinencia y lectura completa, el corpus definitivo se consolidó en 37 fuentes documentales. Este acervo quedó distribuido en 4 fuentes primarias de carácter normativo y jurisprudencial, destacando el marco constitucional y penal ecuatoriano, la Ley Orgánica de Protección de Datos Personales, sentencias paradigmáticas como la Sentencia 2064-14-EP/21 de la Corte Constitucional y referencias del derecho europeo; y 33 fuentes secundarias compuestas por monografías dogmáticas, artículos científicos, informes técnicos de ciberseguridad y trabajos de titulación.

Finalmente, el procesamiento del corpus se canalizó a través de los métodos hermenéutico, dogmático y comparado, los cuales se operativizaron mediante dos herramientas analíticas principales. En primer lugar, se elaboró una matriz de análisis dogmático-penal para contrastar de manera pormenorizada los elementos de las figuras vigentes en el Código Orgánico Integral Penal, puntualmente los artículos 103, 104, 154.2, 166, 172.1, 178, 182, 185, 212 y 396 numeral 1, frente a las particularidades de los contenidos sintéticos, valorando su adecuación típica, verbos rectores, bienes jurídicos tutelados y penas. En segundo lugar, se estructuró una matriz de derecho comparado orientada a sistematizar las soluciones legislativas adoptadas en la Unión Europea, Estados Unidos, Francia, Reino Unido, Corea del Sur y China, examinando sus modelos de intervención, la técnica de tipificación empleada y la viabilidad de incorporar tales fórmulas dentro de la dogmática penal ecuatoriana.

Esta investigación se justifica por su trascendencia jurídica, social e institucional. Jurídicamente, porque permite valorar, con criterios dogmáticos verificables, si una eventual

tipificación autónoma resulta compatible con los principios de legalidad, mínima intervención penal, lesividad, proporcionalidad y culpabilidad, evitando tanto la impunidad como el populismo punitivo. Socialmente, porque visibiliza un problema que afecta de manera desproporcionada a mujeres, niñas, niños y adolescentes, y examina tanto el uso malicioso de la tecnología por parte de los agresores como la respuesta suficiente o insuficiente del Estado ecuatoriano. Institucionalmente, porque sus resultados pueden orientar al legislador en la evaluación de las alternativas normativas disponibles frente al uso malicioso de deepfakes, sin anticipar cuál de ellas resulta preferible antes de concluido el análisis dogmático y comparado que se desarrolla en los capítulos siguientes.

Capitulo I.

El uso malicioso de *deepfake* como conducta lesiva de la dignidad humana y de los derechos personalísimos en el entorno digital.

1. Aproximación conceptual del deepfake.

El galopante avance de las tecnologías de la información y la comunicación, acompañado del desarrollo de la Inteligencia Artificial (IA), ha transformado radicalmente las interacciones humanas y la percepción de la realidad en el ciberespacio. En este contexto tecnológico, el desarrollo de herramientas basadas en IA ha dado lugar a nuevas formas de manipulación de contenidos digitales, entre las que destaca el denominado "Deepfake", que plantea desafíos significativos para el Derecho Penal y la protección de los derechos fundamentales. Para comprender el impacto que esta tecnología puede generar en el ámbito jurídico y su potencial incidencia en derechos fundamentales como la dignidad humana, resulta necesario establecer, en primer lugar, sus bases conceptuales y tecnológicas.

1.1 El deepfake: concepto, distinción y base tecnológica.

El término deepfake proviene de un neologismo que se da como resultado de la contracción de las expresiones en inglés Deep learning (aprendizaje profundo) y fake (falso). Lo que se entiende como un "Deepfake" es todo contenido digital, ya sea contenido de imagen, video o audio, generado, alterado o manipulado intencionalmente mediante algoritmos avanzados de Inteligencia Artificial (IA), con la intención de suplantar la identidad, voz, rostro o expresiones de una persona natural, logrando con ello aparentar que dice o hace cosas que en la vida real nunca pasó. Como resultado de esto, tenemos un material multimedia hiperrealista prácticamente indistinguible de la realidad ante el ojo humano, que sesga la línea de lo verdadero y lo fabricado (Abbas & Taeihagh, 2024).

Distinción frente a otros contenidos digitales manipulados: La alteración de imágenes y videos ha existido desde hace un largo tiempo ya, a través de herramientas

de edición tradicional. Sin embargo, lo que diferencia al deepfake de estos métodos clásicos de edición son tres aspectos fundamentales:

- **Automatización:** el software que utiliza aprende por sí solo los patrones biométricos del individuo y ejecuta el remplazo facial o vocal de forma autónoma, sin intervención humana constante.
- **Hiperrealismo:** ya que supera las limitaciones de la edición convencional al adecuar con precisión expresiones faciales, parpadeo, sincronización labial e iluminación, dificultando profundamente su detección forense.
- **Accesibilidad:** dado que actualmente cualquier usuario sin conocimientos técnicos puede generar falsificaciones hiperrealistas a partir de simples fotografías de la víctima (las cuales gracias a las redes sociales no son nada complicado conseguirlas) mediante aplicaciones de código abierto o plataformas IA gratuitas (Wang et al., 2024).

Base tecnológica: Desde la perspectiva jurídico-funcional, el deepfake se clasifica como contenido sintético, ya que este no solo se limita a alterar pixeles existentes, sino que crea una nueva información visual y auditiva desde cero. Esto es posible gracias a una subrama del Machine Learning denominada Deep Learning, el cual emplea redes neuronales profundas para procesar masivas cantidades de datos biométricos, como fotografías, videos, grabaciones de voz, y con ello lograr replicar de manera autónoma los rasgos de una persona. La arquitectura central que hace posible este proceso son la Redes Generativas Antagónicas (GAN’s en sus siglas en inglés), las que operan mediante entrenamiento simultáneo de dos redes neuronales: un generador, el cual fabrica contenido sintético a partir de los datos aprendidos, y un discriminador: el cual evalúa si dicho contenido es real o falso. Este ciclo iterativo se repite hasta que el

generador produce resultados que el discriminador ya no puede distinguir de la realidad (Atlam et al., 2025).

Para el derecho, lo relevante de este funcionamiento o esta herramienta no es su complejidad matemática, sino sus consecuencias jurídicas, la IA puede clonar la identidad biométrica de cualquier persona de manera autónoma, generando una amenaza directa a bienes jurídicos o derechos fundamentales de los seres humanos, por ejemplo, la intimidad, el honor o en nuestro caso la dignidad, el cual se encuentra arraigado con los demás derechos. Agregando, el hiperrealismo resultante genera lo que la doctrina denomina como "dividendo del mentiroso" (liar's dividend), que significa la mera existencia de esta tecnología permite a verdaderos delincuentes alegar que pruebas auténticas en su contra son fabricaciones de IA, sembrando duda razonable en los tribunales y trasladando una carga probatoria desproporcionada hacia las víctimas (*The current landscape of deepfake legislation*, 2025).

Este fenómeno vulnera el debido proceso, la presunción de inocencia y la seguridad jurídica, y exige no solo la tipificación penal de estas conductas, sino también una actualización urgente en el tratamiento de la prueba electrónica y el peritaje forense en el ordenamiento jurídico ecuatoriano.

1.2 El deepfake como riesgo jurídico en el entorno digital.

La irrupción de la inteligencia artificial generativa ha transformado el ciberespacio, convirtiéndolo en un entorno donde los bienes jurídicos tradicionales enfrentan amenazas atípicas. El deepfake no solo representa un desafío tecnológico, sino un riesgo jurídico multidimensional que vulnera derechos fundamentales consagrados en la Constitución de la República del Ecuador (CRE) y en tratados internacionales. A continuación, se desglosan las principales esferas de afectación legal derivadas de esta tecnología.

1.3 Creación de contenido sin consentimiento.

El núcleo de la antijuridicidad del deepfake radica en la vulneración de la autodeterminación informativa y el consentimiento. Doctrinariamente, se ha establecido que el tratamiento de datos personales exige que la autorización del titular sea libre, específica, informada e inequívoca (Palacios Zúñiga, 2025).

Sin embargo, los deepfakes se alimentan de imágenes y audios extraídos de redes sociales o bases de datos sin autorización, para luego entrenar redes neuronales que fabrican nuevas realidades. En el ámbito de la violencia digital, la doctrina europea ha ampliado el concepto de Abuso Sexual Basado en Imágenes (IBSA, por sus siglas en inglés) para advertir que el daño no solo se produce al distribuir el material, sino desde el momento mismo de la creación no consensuada del contenido mediante IA, ya que esto supone una invasión directa a la autonomía de la persona (Rigotti et al., 2024).

1.3.1. Manipulación de identidad, rostro, voz o imagen.

Desde la perspectiva constitucional y civil, el rostro y la voz no son meros atributos físicos, sino "datos personales" y biométricos que constituyen la identidad de un individuo. La tecnología deepfake se divide operativamente en deepface (clonación y superposición de rostros) y deepvoice (clonación de voz) (Ponce González et al., 2025). Jurídicamente, esta alteración forzosa constituye una forma sofisticada de suplantación de identidad. En la legislación ecuatoriana, el artículo 212 del Código Orgánico Integral Penal (COIP) sanciona la suplantación de identidad; no obstante, la doctrina advierte que este tipo penal resulta ambiguo e insuficiente ante los deepfakes (acto que será analizado a profundidad en el capítulo posterior), ya que a menudo la manipulación de la imagen no busca un "beneficio patrimonial" directo (como exige el tipo penal clásico), sino dañar moralmente a la víctima.

Difusión de información falsa.

En el ecosistema de los medios de comunicación y redes sociales, los deepfakes se catalogan como "contenido manipulado" diseñado deliberadamente para engañar a las audiencias y propagar desinformación (Abdali et al., 2024).

Este fenómeno genera un riesgo jurídico a nivel macro, afectando la paz pública, la fe pública y los procesos democráticos. Al crear escenarios ficticios (como discursos políticos fabricados o declaraciones falsas de autoridades), se atenta contra el derecho a recibir información veraz y contextualizada. Además, provoca lo que la doctrina denomina el "dividendo del mentiroso", una crisis probatoria en la que la proliferación de falsificaciones permite a los verdaderos culpables evadir la justicia argumentando que las pruebas legítimas en su contra son videos generados por IA (The Regulatory Review, 2025).

1.3.2 Afectación a la reputación, intimidad y vida privada.

La divulgación de un deepfake lesiona de forma concurrente el honor, el buen nombre y la intimidad, derechos protegidos por el Art. 66 (numerales 18 y 20) de la Constitución ecuatoriana. El honor es el valor propio y la reputación que una persona tiene frente a la sociedad; cuando un deepfake expone a un individuo en situaciones degradantes o falsas, se produce una injuria o calumnia continuada y masificada (Yépez Andrade, 2026).

Por otro lado, la afectación a la intimidad produce graves secuelas psicológicas. Estudios han demostrado que las víctimas de esta violencia digital sufren de ansiedad, depresión, aislamiento social e incluso ideaciones suicidas, al sentirse expuestas a un escarnio público sobre el cual no tienen ningún control debido a la viralidad de las redes sociales (Chicaiza-Flores et al., 2024).

1.3.3 Uso de deepfakes con fines sexuales, difamatorios, fraudulentos o de humillación.

El uso malicioso de esta tecnología se materializa en varios frentes delictivos que el derecho penal moderno debe enfrentar:

Fines sexuales y de humillación: A nivel internacional, varios estudios coinciden en que los deepfakes se usan sobre todo como una forma de violencia de género en el entorno digital. Uno de los primeros trabajos sobre el tema fue el de Deeptrace Labs, elaborado por Ajder et al. (2019, p. 1), que analizó 14 678 videos encontrados en plataformas abiertas y en comunidades en línea dedicadas a este tipo de contenido. De todos ellos, el 96 % era pornografía sin consentimiento y las víctimas eran únicamente mujeres. Cuatro años más tarde, Home Security Heroes (2023, pp. 4-6) revisó 95 820 videos en más de 100 sitios web y encontró que el 98 % del contenido sintético seguía siendo pornografía no consentida, mientras que el 99 % de las personas afectadas eran mujeres.

Cabe aclarar que estas cifras provienen de análisis globales hechos por empresas privadas de ciberseguridad, por lo que no reflejan datos propios de Ecuador. Además, hasta ahora ni la Fiscalía General del Estado ni el Consejo de la Judicatura cuentan con un registro oficial de los delitos cometidos con inteligencia artificial generativa. Aun así, estos datos sirven como referencia criminológica para dimensionar el daño que produce este fenómeno.

Lo que muestran los resultados es que, en la mayoría de los casos, quien crea estos contenidos no busca ganar dinero, sino ofender a la víctima (animus injuriandi), es decir, humillarla y degradarla al mostrarla en actos sexuales que nunca ocurrieron, esta práctica afecta directamente la intimidad, el honor y la indemnidad sexual (Rigotti et al., 2024).

Fines difamatorios y fraudulentos: En el entorno corporativo y personal, los deepfakes se utilizan para el fraude financiero (como el phishing mediante clonación de voz de directivos de empresas para autorizar transferencias ilícitas) y la extorsión (sextorsión), donde los ciberdelincuentes amenazan con divulgar el material manipulado si no se cumplen exigencias económicas (Espinosa Carvajal & Paredes Fuentes, 2025).

1.4. La dignidad humana como bien jurídico afectado.

El abordaje de la tecnología deepfake desde la perspectiva del Derecho exige un análisis profundo de la dignidad humana, no solo como un concepto ético, sino como el eje central y el bien jurídico fundamental que resulta lesionado ante estas manipulaciones digitales. La dignidad se erige como el valor supremo sobre el cual gira la organización de los derechos fundamentales en el orden constitucional e internacional (Ghersi, 2016).

1.4.1. Concepto doctrinario de dignidad humana.

Desde una perspectiva doctrinaria, la dignidad humana se define como una calidad inherente a todos los miembros de la especie humana que no admite sustitución ni equivalencia. Doctrinarios como Van Wintrich señalan que esta consiste en que la persona, como ente ético-espiritual, tiene por naturaleza la capacidad de autodeterminarse, formarse y actuar libremente sobre el mundo que la rodea. No se trata de un derecho otorgado por el Estado, sino de un estado moral permanente e inescindible del ser humano por el solo hecho de serlo (Madrid: Ediciones Jurídicas y Sociales, 1996).

1.4.2 La persona como fin en sí misma.

El fundamento de la dignidad radica en que el ser humano constituye un "fin en sí mismo", lo que implica que jamás puede ser instrumentalizado o utilizado únicamente como un medio o herramienta para los fines de terceros, del Estado o de la sociedad. Esta concepción supone que la persona posee un valor inestimable per se, y que todas las estructuras jurídicas deben ordenarse en pro de la perfección de sus potencias naturales. En este sentido, la dignidad intrínseca es inviolable y constituye el valor fundamental con respecto al cual los demás valores tienen siempre un carácter meramente instrumental (Madrid: Ediciones Jurídicas y Sociales, 1996).

Por consiguiente, el reconocimiento de la persona como fin supremo justifica la existencia de un catálogo de protecciones legales denominadas derechos fundamentales, los cuales emanan directamente de esta condición humana (Madrid: Ediciones Jurídicas y Sociales, 1996).

1.4.3. Dignidad humana como fundamento de los derechos fundamentales.

La dignidad humana es el núcleo central del cual emanan y hacia el cual convergen todos los demás derechos, tales como la intimidad, el honor, la propia imagen y la identidad.

Se considera que los derechos fundamentales son la concreción contemporánea de la dignidad, siendo esta su base y estructura esencial (Peyrano, 2002).

Al estar positivada en la Constitución, la dignidad actúa como un límite infranqueable para la potestad punitiva del Estado (ius puniendi) y como una guía para la interpretación de cualquier norma jurídica, asegurando que el intérprete escoja siempre el resultado que más proteja a la persona. De este modo, la protección de la dignidad no solo resguarda la existencia física, sino también la libertad del individuo para decidir sobre su propia vida y datos, lo cual nos lleva al concepto de autonomía y consentimiento (Sagüés, 1993).

1.4.4. Relación entre dignidad humana, autonomía y consentimiento.

La dignidad se manifiesta singularmente en la autodeterminación consciente y responsable de la propia vida. En la era digital, esto se traduce en el derecho a la autodeterminación informativa, que es la facultad de cada persona para controlar y gestionar libremente sus datos personales, incluyendo su imagen y su voz (Peyrano, 2002).

El consentimiento se vuelve el elemento habilitante indispensable: cualquier tratamiento de datos personales (como el uso del rostro en un video) requiere la autorización expresa, libre e inequívoca del titular (Rosas, 2023).

Sin el consentimiento, la utilización de la imagen ajena despoja al individuo de su autonomía, instrumentalizando su identidad para fines ajenos a su voluntad; este es precisamente el escenario de agresión que producen los deepfakes, como se analiza a continuación.

1.4.5. Afectación de la dignidad cuando se usa la imagen de una persona para simular actos que nunca realizó.

La creación de un deepfake sin consentimiento constituye una violación directa a la dignidad humana porque erosiona la autovaloración y el respeto que el individuo merece ante los demás (Badeni, 2002).

Al utilizar el rostro de una persona para simular actos que nunca ocurrieron "especialmente en contextos sexuales o denigrantes", se produce una cosificación y deshumanización de la víctima, quien es reducida a un objeto de consumo visual o manipulación (Cerdán, 2019).

Este "engaño mutado" no solo daña la reputación pública y el honor, sino que impacta profundamente en la integridad moral y psicológica, provocando humillación, angustia y una pérdida total de control sobre la propia existencia digital (Feijóo, 2021).

En definitiva, el deepfake representa una amenaza inminente que, bajo la apariencia de hiperrealismo, altera la percepción de la verdad y lesiona el patrimonio moral del ser humano al atacar su identidad y dignidad más profundas.

1.5. Derechos personalísimos vulnerados por el uso malicioso de deepfakes.

La tecnología deepfake, por su capacidad de suplantar identidades de forma hiperrealista, no solo representa una innovación técnica, sino una amenaza directa a los derechos de la personalidad o derechos personalísimos (Peyrano, 2002).

Estos derechos, inherentes al ser humano por su sola condición de tal, protegen la integridad física, espiritual y moral de la persona, abarcando todas las aristas de su compleja estructura unitaria (Cifuentes, 1998).

1.5.1. Derecho al honor y al buen nombre.

Según el Tribunal Constitucional de España en su sentencia STC 53/85 del 11 de abril de 1985, la dignidad es un "valor espiritual y moral inherente a la persona", la cual se manifiesta en la autodeterminación consciente de la propia vida y conlleva la pretensión de respeto por

parte de los demás. El derecho al honor, por su parte, se entiende como un bien inmaterial vinculado a esa dignidad humana.

El uso de deepfakes para imputar falsamente actos ilícitos o conductas vergonzosas lesiona gravemente este derecho al desacreditar a la víctima mediante evidencias visuales que el público tiende a percibir como auténticas. De esta manera, el honor se ve avasallado por una "falsedad profunda" que destruye la consideración ajena del individuo (Zorrilla, 2018).

En consecuencia, el daño a la reputación pública trasciende el entorno social para invadir la esfera más reservada de la vida del individuo, afectando su intimidad personal.

1.5.2. Derecho a la intimidad personal y familiar.

La intimidad constituye la facultad que posee todo individuo para decidir qué aspectos de su vida privada desea compartir y cuáles excluir del conocimiento público. En la era digital, esta esfera se ve amenazada por el anonimato y la rapidez con la que se pueden generar y difundir datos íntimos sin consentimiento (Basterra, 2016).

Los deepfakes vulneran esta "expectativa razonable de privacidad" al exponer representaciones sintéticas de la víctima en contextos reservados, lo que genera sentimientos de humillación, angustia y pérdida de control sobre el círculo íntimo. Esta intromisión ilegítima en la vida privada se materializa principalmente a través de la captura y manipulación de los rasgos físicos, lo que nos lleva al análisis de la propia imagen (Jarvis Cooper, 2023).

1.5.3. Derecho a la propia imagen.

Este derecho otorga a la persona la potestad exclusiva de decidir cuándo y cómo sus rasgos físicos y voz pueden ser captados, reproducidos o difundidos por terceros (Nogueira, 2007).

Como derecho autónomo, protege el patrimonio moral evitando la publicación incondicionada de la apariencia del individuo (Nogueira Alcalá, 2007).

El deepfake constituye una de las formas más extremas de agresión a la imagen, ya que no solo utiliza el rostro sin permiso, sino que lo recrea artificialmente para simular actos que nunca ocurrieron, despojando al titular de su facultad de disposición (Cooper, 2023).

Al ser el rostro y la voz representaciones directas de la identidad, el uso indebido de los mismos se traduce técnicamente en un tratamiento no autorizado de datos personales.

1.5.4. Derecho a la protección de datos personales y a la autodeterminación informativa.

El derecho a la protección de datos garantiza el control sobre cualquier información vinculada a una persona identificada o identificable, mientras que la autodeterminación informativa es la facultad activa de gestionar el flujo de esa información (Peyrano, 2002).

La jurisprudencia reconoce que el rostro y la voz son datos biométricos protegidos que no pierden su carácter por el hecho de circular en redes sociales (Reglamento [UE], 2016).

El uso malicioso de estas tecnologías priva a la víctima de su derecho a decidir el destino de su información digital, convirtiéndola en un objeto de procesamiento algorítmico opaco. Cuando este tratamiento de datos se enfoca en la creación de contenido explícito, la vulneración escala hacia la esfera de la libertad sexual.

1.5.5. Derecho a la libertad sexual en casos de deepfakes de contenido sexual.

La libertad sexual es el derecho de toda persona a tomar decisiones voluntarias y responsables sobre su vida y orientación sexual (Bowen, 2017).

Los deepfakes sexuales, clasificados como Abuso Sexual Basado en Imágenes (ASBI), violentan la autonomía sexual al proyectar a la víctima en situaciones eróticas no consentidas, lo que constituye una forma de violencia y cosificación que busca ridiculizar y vejar (Travers, 2024).

El daño es irreparable, pues la reacción de quienes visualizan el contenido es equiparable a la que tendrían ante un acto real, provocando un impacto devastador en la

integridad psicológica. Esta violencia digital no afecta a todos por igual, manifestando una incidencia desproporcionada en sectores históricamente vulnerables.

1.5.6. Especial afectación a mujeres, niñas, niños y adolescentes.

Según el estudio State of Deepfakes 2023, de Home Security Heroes Las estadísticas revelan que el 99% de las víctimas de pornografía deepfake son mujeres, lo que evidencia una matriz cultural patriarcal que utiliza la IA para perpetuar la discriminación y la violencia de género (Bigas, 2023).

Asimismo, la población infantojuvenil se encuentra en una situación de especial vulnerabilidad, ya que la simulación de sus rostros en contenido pornográfico lesiona su indemnidad sexual y libre desarrollo de la personalidad. El ciberacoso y la "diversión en la crueldad" mediante estas herramientas han derivado en casos extremos de aislamiento social y suicidio (Hooper, 2024).

1.6. Protección constitucional e internacional de los derechos afectados.

La defensa frente a los deepfakes no se limita únicamente al ámbito penal del COIP, sino que encuentra un sólido fundamento en la norma suprema del Ecuador y en el sistema internacional de derechos humanos.

Constitución de la República del Ecuador: El artículo 66 es el pilar de esta protección, garantizando en sus numerales el derecho al honor y buen nombre (num. 18), a la protección de datos de carácter personal y autodeterminación informativa (num. 19), a la intimidad personal y familiar (num. 20) y a la identidad personal y colectiva (núm. 28).

La Constitución establece que el tratamiento de estos datos requiere siempre la autorización del titular o mandato de ley.

Declaración Universal de Derechos Humanos (1948): En su artículo 12, prescribe que nadie será objeto de injerencias arbitrarias en su vida privada ni de ataques a su honra o reputación, reconociendo el derecho a la protección legal contra tales ofensas.

Pacto Internacional de Derechos Civiles y Políticos (1966): Reafirma en su artículo 17 la prohibición de injerencias ilegales en la privacidad y ataques al honor, obligando a los Estados a brindar protección efectiva para prevenir y sancionar estas acciones.

Convención Americana sobre Derechos Humanos (Pacto de San José, 1969): El artículo 11 protege el respeto a la honra y el reconocimiento de la dignidad, mientras que el artículo 13 advierte que, si bien existe libertad de expresión, esta conlleva responsabilidades ulteriores para asegurar el respeto a los derechos y la reputación de los demás.

En definitiva, estos instrumentos configuran un bloque de constitucionalidad que exige al Estado ecuatoriano adaptar su legislación para garantizar una tutela efectiva de la dignidad en el entorno digital.

Por otro lado, La Corte Constitucional del Ecuador, mediante la sentencia No. 2064-14-EP/21 ha consolidado la doctrina sobre el alcance de la protección de datos personales y los derechos de la personalidad en contextos donde interviene la tecnología (Corte Constitucional del Ecuador, 2021).

Un aporte fundamental de esta sentencia es la definición de tratamiento de datos. La Corte estableció que acciones como consultar, visualizar, almacenar en dispositivos (como memorias flash) o mostrar imágenes a terceros sin autorización constituyen formas de tratamiento de datos personales (Corte Constitucional del Ecuador, 2021).

A pesar de que este fallo representa un avance significativo en la justicia constitucional, la propia doctrina señala que existe una insuficiencia operativa para sancionar penalmente estas conductas, ya que la sentencia se limita al ámbito del Habeas Data.

La Corte Constitucional ha reconocido la vulneración de la honra, la imagen y la intimidad en estos casos, pero las dificultades de adecuación típica en el COIP limitan las herramientas específicas de los jueces penales para sancionar la creación sintética de este contenido.

1.7. El Deepfake como forma de violencia digital: Más allá de la manipulación tecnológica.

Antes de abordar este capítulo, resulta necesario analizar el fenómeno del deepfake desde una perspectiva jurídica y social, considerando que su utilización trasciende el ámbito del desarrollo tecnológico y puede generar afectaciones relevantes a derechos fundamentales cuando es empleado con fines lesivos. En este sentido, la creación y difusión de contenidos audiovisuales alterados mediante inteligencia artificial constituye una manifestación de la violencia digital cuando tiene como finalidad vulnerar la dignidad, intimidad, honra o integridad de una persona (Zambrana Ruiz, 2024).

La alteración audiovisual hiperrealista de una persona, despojada de su consentimiento, transgrede la barrera de la virtualidad para generar un daño tangible, multidimensional y muchas veces irreparable en la vida de la víctima (Andino Veloz, 2025).

1.7.1. Violencia de genero digital y Abuso Sexual Basado en Imágenes (IBSA).

La violencia digital no es un fenómeno neutral, este tiene un componente de género innegable. Las estadísticas actuales revelan que más del 90% de los deepfakes que circulan en la red son de naturaleza pornográfica, y de ellos, el 99% tiene como víctimas a mujeres, desde figuras públicas hasta ciudadanas comunes y estudiantes menores de edad (Rigotti et al., 2024).

Este enfoque evidencia que el deepfake sexual no consentido es una manifestación directa de la violencia de género digital, arraigada en la subordinación histórica de la mujer y las dinámicas de dominación patriarcal.

La doctrina y la legislación internacional, nos menciona que este fenómeno se ha comenzado a catalogar bajo el concepto de Abuso Sexual Basado en Imágenes (IBSA, por sus siglas en inglés). Este término es jurídicamente más preciso y comprensivo que la errónea denominación de “porno de venganza”, ya que el IBSA reconoce que el abuso no requiere que la víctima haya participado voluntariamente en la creación de una imagen íntima previa, ni que exista una intención de venganza por parte de la expareja. El simple acto de crear, tomar o

compartir el material de íntimo manipulado (deepfake porn) sin consentimiento constituye una invasión abusiva a la autonomía sexual y corporal de la persona (Taylor Villafuerte, 2024).

1.7.2. Cosificación, humillación y el cinismo de la "diversión en la crueldad".

La creación de un deepfake sexual no consentido responde a mecánicas de poder donde el agresor busca someter a la víctima. Sociológicamente, esto produce una cosificación sexual y deshumanización extrema, el rostro de la mujer es recortado y pegado en un cuerpo ajeno (generalmente de una actriz pornográfica despersonificada) para convertirla en un instrumento mercantil o de satisfacción del deseo masculino, reduciéndola a un objeto carente de voluntad (Andino Veloz, 2025).

Este acto se alimenta de lo que la literatura denomina cinismo del neoliberalismo y la "diversión en la crueldad". Quienes fabrican y difunden estos deepfakes (por lo general sumergidos en el anonimato en las redes sociales o foros) obtienen el goce de humillar, degradar y controlar a la víctima. La intención principal del deepfake pornográfico no siempre busca el lucro económico (sextorsión), sino el ejercer una violencia de superioridad que busca destruir la reputación y perjudicar a la mujer en la esfera pública (Andino Veloz, 2025).

1.7.3. Pérdida de control sobre la identidad digital y daños multidimensionales.

El hiperrealismo y la gran repercusión de internet provocan que, una vez que el deepfake es publicado, la víctima sufra una absoluta pérdida de control sobre su propia identidad digital y autodeterminación informativa. Se vulnera gravemente su facultad de decidir sobre la capacitación, reproducción y publicación de su imagen en su vida privada o pública (Palacios Zúñiga, 2025).

Esta pérdida de agencia genera un gran impacto negativo que transciende la pantalla, ramificándose en graves daños para el individuo:

- **Daño psicológico:** Las víctimas experimentan una profunda ruptura existencial. Se enfrentan a trastornos de estrés postraumático, ansiedad severa, depresión,

aislamiento social e, incluso, ideaciones suicidas debido a la impotencia de no poder eliminar el contenido de la red (Rigotti et al., 2024).

- **Daño Social y Reputacional:** El “dividendo del mentiroso” y la cultura de la culpabilización (victim-blaming) provocan un escenario público. La sociedad tiende a estigmatizar a la víctima asumiendo la veracidad del material, lo que destruye el buen nombre, la honra y el entorno relacional o familiar de la persona afectada (Chicaiza-Flores et al., 2024).
- **Daño Laboral y Económico:** La humillación pública tiene secuelas tangibles, traducidas en ausentismo laboral, despidos, pérdida de oportunidades profesionales y el surgimiento de onerosos gastos legales y psicológicos en la búsqueda de la justicia y reparación (Rigotti et al., 2024).

1.8. Conclusión del Capítulo I.

A lo largo de este primer capítulo, se ha podido establecer que la irrupción de la inteligencia artificial generativa ha transformado el ecosistema digital, dando lugar a fenómenos tan complejos y peligrosos como el deepfake. Lejos de ser una simple herramienta de edición audiovisual o una innovación inofensiva, el uso malicioso de esta tecnología constituye una de las expresiones más sofisticadas y destructivas de la violencia digital contemporánea.

El hiperrealismo y la mínima divergencia que caracterizan a estos medios sintéticos difuminan la línea entre la verdad y la falsedad, generando un "dividendo del mentiroso" que no solo desestabiliza la confianza pública y probatoria, sino que atenta de forma directa e intencional contra los derechos fundamentales de las personas.

La principal conclusión que se extrae de esta aproximación conceptual es que el uso malicioso de deepfakes produce una lesión pluriofensiva que ataca el núcleo mismo de la dignidad humana.

Al utilizar el rostro o la voz de una persona sin su consentimiento para simular actos (especialmente de connotación sexual) que nunca realizó, el individuo es despojado de su autonomía, siendo cosificado, deshumanizado y reducido a un mero objeto de consumo visual o de manipulación pública.

De esta profunda vulneración a la dignidad se desprende una transgresión gravísima a los derechos personalísimos.

En primer lugar, se vulnera gravemente el derecho a la intimidad, exponiendo a las víctimas a un escarnio público sobre escenarios ficticios que invaden su esfera más reservada, arrebatándoles su "expectativa razonable de privacidad".

En segundo lugar, se afecta seriamente el honor y el buen nombre, ya que las falsedades hiperrealistas desacreditan a la víctima ante una sociedad que asume instintivamente la veracidad del material audiovisual.

En tercer lugar, se lesiona severamente el derecho a la propia imagen y a la protección de datos personales. El rostro y la voz no son meros atributos físicos, sino datos biométricos inalienables; su captura, procesamiento algorítmico y manipulación sin autorización previa suponen un quebrantamiento total de la autodeterminación informativa de la víctima, quien pierde el control absoluto sobre su identidad y su rastro en el ciberespacio.

Esta violencia digital, catalogada como Abuso Sexual Basado en Imágenes (IBSA), presenta además una alarmante matriz de género, afectando desproporcionadamente a mujeres, niñas, niños y adolescentes. Las secuelas de estos ataques se materializan en daños psicológicos (como ansiedad y depresión), y provocan perjuicios sociales y reputacionales que son, en la mayoría de los casos, irreparables debido a la viralidad de las redes.

Por todo lo expuesto, al quedar en evidencia el nivel de destrucción que los deepfakes generan sobre bienes jurídicos de máxima relevancia constitucional, resulta pertinente trasladar el debate hacia el derecho penal, para analizar si ofrece una respuesta adecuada.

Capítulo II.

Capacidad e insuficiencia de los tipos penales vigentes en el COIP frente al uso malicioso de deepfake.

2.1. El principio de legalidad y la exigencia de tipicidad estricta en materia penal.

En un Estado constitucional de derechos y justicia, el ejercicio del ius puniendi o potestad sancionadora del Estado no es absoluto, sino que encuentra su límite infranqueable en la Constitución y la ley (Saquicela Rodas & Chica Mejía, 2023).

Este límite se materializa a través del principio de legalidad, el cual constituye la garantía básica de que ninguna persona podrá ser procesada ni sancionada por un acto u omisión que no esté previamente calificado en la ley de manera expresa e inequívoca (Asamblea Nacional del Ecuador, 2008).

2.1.1. Nullum crimen, nulla poena sine lege y la tipicidad estricta.

El aforismo latino nullum crimen, nulla poena sine lege exige que las leyes penales sean claras, específicas y accesibles, asegurando que el ciudadano conozca de antemano qué conductas están prohibidas y cuáles son sus consecuencias. De este principio emana la exigencia de tipicidad estricta, la cual obliga al legislador a realizar una descripción pormenorizada de la conducta penalmente relevante. En consecuencia, para que un hecho sea considerado delito, debe existir una adecuación perfecta entre el acto cometido y la descripción abstracta contenida en la norma penal (Saquicela Rodas & Chica Mejía, 2023).

2.1.2. Prohibición de analogía y de interpretación extensiva.

Una de las garantías fundamentales de la seguridad jurídica es la prohibición de la analogía en perjuicio del procesado. Esto significa que los jueces no pueden sancionar conductas que, siendo similares a un delito tipificado, no encajen estrictamente en sus elementos constitutivos. En la praxis jurídica ecuatoriana, se ha advertido una tendencia preocupante de intentar subsumir hechos tecnológicos en infracciones existentes mediante una interpretación extensiva, lo cual pone en peligro el principio de legalidad y desnaturaliza la

seguridad ciudadana. Si una conducta no se adecúa íntegramente al tipo penal, el hecho se considera atípico, y su sanción devendría en arbitraria (Carnevali Rodríguez, 2008).

2.1.3. El problema del rezago legal frente a la realidad digital.

El derecho penal ecuatoriano enfrenta actualmente el desafío del "legal lag" o rezago legal, fenómeno donde la celeridad de la innovación técnica sobrepasa la capacidad de respuesta de los marcos normativos. El Código Orgánico Integral Penal (COIP), aunque innovador en su momento, fue diseñado para realidades tradicionales y presenta limitaciones críticas frente a la sofisticación de la Inteligencia Artificial (Chávez Zambrano & San Andrés Molina, 2025).

La falta de taxatividad técnica en las normas actuales genera ambigüedad, pues términos como "reproducir" o "difundir" pueden resultar insuficientes para capturar la esencia de la creación sintética y autónoma que caracteriza a los deepfakes. Sancionar nuevas conductas digitales con leyes pensadas para soportes físicos o capturas reales de imagen violenta la exigencia de que el tipo penal describa de forma inequívoca el comportamiento prohibido (Ugwuoke, 2025).

En conclusión, la rigidez del principio de legalidad impide que las "lagunas legales" generadas por la IA sean llenadas mediante el arbitrio judicial, lo que lleva a analizar si es necesaria una reforma legislativa y de qué tipo, incluida la posibilidad de tipos penales específicos para el entorno digital (Endara-Chamorro et al., 2024).

2.2. Derecho penal, nuevas tecnologías y rezago normativo.

El dinamismo de la era digital ha provocado que el Derecho se enfrente a un fenómeno estructural conocido como "legal lag" o rezago legal, definido como la brecha o distanciamiento que surge entre la celeridad de la innovación tecnológica y la capacidad de respuesta de los marcos normativos. En un contexto donde la tecnología avanza a pasos agigantados, el sistema jurídico a menudo se encuentra en una "etapa pre-normativa", donde la

ausencia de leyes específicas expone a la ciudadanía a riesgos sistémicos sin una protección clara (Chávez Zambrano & San Andrés Molina, 2025).

2.2.1. Avance acelerado de la inteligencia artificial y lentitud legislativa.

La denominada Cuarta Revolución Industrial ha propiciado un crecimiento exponencial de la Inteligencia Artificial, integrándola en casi todos los aspectos de la vida cotidiana, desde asistentes virtuales hasta algoritmos de aprendizaje profundo capaces de recrear realidades. Sin embargo, mientras que las herramientas de IA evolucionan en cuestión de meses, los procesos legislativos para regularlas son tradicionalmente lentos y burocráticos, esta asimetría temporal genera un vacío donde el desarrollo técnico sobrepasa con creces la capacidad de los legisladores para comprender y normar los nuevos fenómenos (Rosales Murillo, 2024).

2.2.2. Nuevas formas de criminalidad digital.

El rezago normativo es particularmente crítico ante la aparición de las "criminologías del futuro", donde la IA es utilizada para automatizar ataques y sofisticar delitos tradicionales. El surgimiento de los deepfakes representa una de las amenazas más serias, ya que permite la creación de contenido multimedia hiperrealista para cometer fraudes, suplantaciones de identidad y violencia sexual digital. Estos ilícitos se caracterizan por su anonimato, su ejecución instantánea y su naturaleza transnacional, factores que desafían las fronteras de la jurisdicción penal nacional (Sun, 2026).

2.2.3. Dificultad del derecho penal tradicional ante realidades sintéticas.

El derecho penal tradicional, diseñado para un mundo de soportes físicos y capturas reales de la realidad, presenta una grave dificultad para responder a las realidades sintéticas creadas por algoritmos. La legislación vigente en países como Ecuador a menudo se centra en verbos rectores como "reproducir" o "grabar", los cuales resultan insuficientes para abarcar el acto de "creación algorítmica" desde cero. Al ser las creaciones de IA una mezcla de datos reales y procesos generativos, se sitúan en un "limbo jurídico" que complica la labor de los

jueces para encuadrar estas conductas en los tipos penales clásicos de difamación o violación a la intimidad (Ugwuoke & Sanfilippo, 2025).

2.2.4. Riesgo de impunidad por ausencia de regulación específica.

La consecuencia más alarmante del rezago legal es el riesgo de impunidad. Bajo el principio de legalidad "analizado previamente", si una conducta no está descrita de forma exacta en la ley, el hecho es considerado atípico, impidiendo la imposición de una sanción penal. En el caso de los deepfakes, en las modalidades que no encajan en los tipos vigentes, la falta de una tipificación específica puede dejar a las víctimas sin respuesta penal, ya que los agresores podrían aprovechar esa falta de cobertura para actuar sin temor a consecuencias legales. Esa falta de cobertura no solo puede desmotivar la denuncia ciudadana, sino que también erosiona la confianza en el sistema de justicia frente a los desafíos del ciberespacio (Espinosa Carvajal & Paredes Fuentes, 2025).

2.3. Revisión del delito de violación a la intimidad en el Código Orgánico Integral Penal (COIP).

Para analizar la capacidad del sistema penal ecuatoriano frente a los desafíos de la Inteligencia Artificial, es imperativo realizar un examen crítico del Artículo 178 del COIP, referido a la violación a la intimidad, el cual constituye la principal figura jurídica a la que se recurre para intentar sancionar el uso malicioso de los deepfakes (Salvador Quiñonez, 2022).

2.3.1. Bien jurídico protegido en la violación a la intimidad.

El bien jurídico tutelado por esta norma es la intimidad personal y familiar, entendida como la facultad de todo individuo para excluir del conocimiento público ciertos aspectos de su vida que considera reservados. La jurisprudencia ecuatoriana asocia este derecho con la "expectativa razonable de privacidad", la cual surge especialmente en entornos digitales de mensajería cifrada o espacios cerrados donde el titular confía en que su información no será accedida por terceros. Asimismo, este derecho está intrínsecamente ligado a la

autodeterminación informativa, que garantiza al titular el control sobre el acceso y uso de sus propios datos (Corte Constitucional del Ecuador, 2021).

2.3.2. Análisis de los verbos rectores y su alcance.

El artículo 178 sanciona a quien, sin contar con el consentimiento o la autorización legal, acceda, intercepte, examine, retenga, grabe, reproduzca, difunda o publique datos personales, mensajes de datos, voz, audio o video de otra persona (Asamblea Nacional del Ecuador, 2014). El artículo 178 contempla diversas conductas de acceso, registro, reproducción y difusión de información de otra persona sin consentimiento o autorización legal. Frente al contenido sintético, debe examinarse si concurren el objeto material y los demás elementos del tipo en cada caso, especialmente al distinguir generación de una representación inexistente y difusión de datos personales.

Difusión o publicación: Son los verbos que permiten, en teoría, perseguir la viralización de deepfakes cuando estos han sido puestos en circulación en redes sociales con el fin de dañar la reputación del titular (Cuenca Gonzaga et al., 2024).

2.3.3. Aplicabilidad y límites frente a contenido creado artificialmente.

Si bien el Artículo 178 es el recurso más utilizado para abordar la difusión de contenido íntimo manipulado, su aplicación frente a los deepfakes presenta límites críticos:

La barrera de la "creación": El artículo 178 se refiere a información personal ya existente, por lo que presenta dificultades para capturar el acto de la creación sintética. Un deepfake no es una grabación real de la víctima, sino una generación artificial mediante algoritmos; por tanto, existe una incertidumbre jurídica sobre si la "recreación" de una imagen encaja estrictamente en los verbos "reproducir" o "grabar" (Rosales Murillo, 2024).

Naturaleza de los datos: Aunque la Corte Constitucional ha determinado que el rostro constituye un dato personal protegido, la defensa del procesado podría argumentar que el contenido de un deepfake es una obra de ficción algorítmica y no un "dato personal" real del

sujeto, lo que dificultaría la subsunción de la conducta en el tipo penal vigente (Taylor Villafuerte, 2024).

2.3.4. Problema: La falta de cobertura expresa de la creación de contenido falso mediante IA.

El problema fundamental radica en que el Artículo 178 no regula de manera expresa el uso de Inteligencia Artificial para la fabricación de realidades falsas. El legislador ecuatoriano ha omitido el verbo rector "crear" o "generar" en el contexto de la violación a la intimidad, lo cual es el núcleo del problema de los ultrafalsos (Rosales Murillo, 2024).

Bajo la exigencia de tipicidad estricta analizada en el apartado 2.1, forzar la interpretación de "difusión de datos personales" para incluir "creación de imágenes sintéticas" podría vulnerar los derechos del procesado. En consecuencia, se advierte que el Artículo 178, aunque útil para la fase de viralización, resulta insuficiente para sancionar la fase de generación del material, de modo que la conducta tecnológica inicial, que lesiona la dignidad y el honor de las víctimas, podría quedar sin cobertura penal cuando no concurran los demás elementos del tipo (Endara-Chamorro et al., 2024).

2.4. Revisión del delito de calumnia en el Código Orgánico Integral Penal (COIP).

El ordenamiento jurídico ecuatoriano tutela el derecho al honor y al buen nombre a través del delito de calumnia, tipificado en el Artículo 182 del COIP. Esta figura adquiere una relevancia renovada ante el uso de deepfakes, ya que estas herramientas permiten simular de forma hiperrealista que un individuo participa en actos delictivos, afectando su patrimonio moral y social (Valverde-Ramos et al., 2024).

2.4.1. Concepto penal de calumnia y la falsa imputación.

Desde la dogmática penal, la calumnia se define como la falsa imputación de un delito en contra de otra persona. Para que esta conducta sea punible, requiere la concurrencia de dos vertientes: el tipo objetivo, referido a la falsedad de los hechos endilgados, y el tipo subjetivo,

caracterizado por el animus injuriandi, es decir, el conocimiento del autor sobre la falsedad de la imputación y su voluntad de proferirla para dañar a la víctima (Cuenca Gonzaga et al., 2024). Determinado el concepto de calumnia, es necesario analizar el artículo 396 del COIP, dado que, aunque ambos tipos normativos convergen en la tutela del bien jurídico del honor y la reputación, la legislación penal ecuatoriana separa sustancialmente sus ámbitos de aplicación. Mientras que el delito de calumnia (Art. 182) exige ineludiblemente la falsa imputación de un acto tipificado como delito el artículo 396, numeral 1 del Código Orgánico Integral Penal (COIP) se limita a regular las contravenciones de cuarta clase, sancionando a quien, por cualquier medio, profiera expresiones en descrédito o deshonra en contra de otra persona (Asamblea Nacional del Ecuador, 2014).

Esta separación dogmática es fundamental para evaluar la aplicabilidad de las contravenciones a las agresiones cometidas mediante la tecnología de los deepfakes. Al realizar este escrutinio, se identifican limitaciones normativas precisas que impiden considerar que esta figura encaja de manera idónea frente a la violencia digital sintética.

El significado jurídico de "expresiones" frente al contenido sintético El primer análisis debe centrarse en el alcance del término "expresiones" empleado por el legislador. En la doctrina del derecho penal, las expresiones dirigidas a menoscabar la honra se inscriben dentro de la teoría del "concepto significativo de acción". Bajo esta perspectiva, lo relevante para el tipo penal no es únicamente la materialidad física de la palabra o el gesto, sino el significado injurioso o desacreditante que se le atribuye en un determinado contexto de interacción social (Yépez Andrade, 2026).

Aunque la reforma penal de 2021 extendió el ámbito del artículo 396.1 a las expresiones vertidas "a través de cualquiera de las tecnologías de la información y comunicación", la subsunción del deepfake en esta norma presenta un obstáculo bajo el principio de tipicidad

estricta (Art. 13.2 del COIP). Tradicionalmente, proferir una expresión implica un acto de manifestación o transmisión lingüística o simbólica por parte del autor.

Un deepfake, en cambio, no es una mera verbalización ofensiva, sino un proceso de manipulación y fabricación técnica de datos biométricos (rostro y voz) orientado a estructurar una simulación de la realidad. Al no existir los verbos rectores de "crear", "alterar" o "generar" contenido sintético dentro de la contravención, su aplicación procesal requiere de una interpretación extensiva que colisiona con las reglas de interpretación literal de la materia penal (Asamblea Nacional del Ecuador, 2014).

La limitación de la contravención se profundiza al examinar el segundo inciso del numeral 1 del artículo 396, el cual establece que "esta contravención no será punible si las expresiones son recíprocas en el mismo acto". Esta eximente de responsabilidad responde a una lógica de simetría y espontaneidad temporal, propia de altercados verbales o físicos inmediatos donde los insultos mutuos neutralizan la necesidad de sanción estatal (Yépez Andrade, 2026).

Sin embargo, el uso malicioso de deepfakes opera bajo dinámicas de asimetría técnica y temporal. La creación de un medio sintético exige planificación, recopilación de datos de la víctima, entrenamiento de algoritmos y procesamiento digital (Ponce González et al., 2025). Al tratarse de un acto asíncrono y premeditado que se difunde en internet, resulta materialmente inviable verificar el requisito de reciprocidad "en el mismo acto". Exigir una respuesta simétrica de la víctima en el ciberespacio carece de sustento operativo y dogmático.

Desde el plano constitucional, el artículo 76, numeral 6 de la Constitución exige una debida proporcionalidad entre las infracciones y las sanciones penales. El artículo 396 sanciona la contravención de cuarta clase con una pena privativa de libertad de quince a treinta días (Asamblea Nacional del Ecuador, 2014).

Al evaluar esta escala punitiva frente al desvalor de la acción de los deepfakes, se constata una evidente desproporción:

Magnitud del daño: Mientras que la contravención tutela el honor frente a ofensas o expresiones hostiles de carácter transitorio, el deepfake (especialmente el de naturaleza sexual no consensuada o IBSA) ataca de forma concurrente la intimidad, la autodeterminación informativa y la integridad psicológica de la víctima, generando secuelas permanentes debido a la persistencia y viralización de los datos digitales (Rigotti et al., 2024).

Discrepancia punitiva: Sancionar la falsificación algorítmica de la identidad personal y la exposición corporal sintética con una pena contravencional de pocos días equipara erróneamente un ciberdelito pluriofensivo de alta complejidad técnica con un exabrupto o riña verbal cotidiana, privando a la víctima de los mecanismos de reparación integral y protección que caracterizan a los delitos de mayor gravedad en el COIP.

Por lo tanto, la contravención del artículo 396.1 del COIP resulta insuficiente y técnicamente inadecuada para aprehender el desvalor y la complejidad de las agresiones cometidas mediante inteligencia artificial.

2.4.2. Deepfakes que simulan la comisión de ilícitos.

La tecnología de inteligencia artificial permite crear contenidos donde se observa a un sujeto realizando declaraciones o acciones criminales que nunca ocurrieron en la realidad. Cuando un deepfake es utilizado para mostrar a una persona cometiendo un robo, aceptando un soborno o participando en actos ilícitos, se configura técnicamente una calumnia mediante un medio audiovisual sintético. Dado que la sociedad otorga una alta credibilidad a las evidencias visuales, el daño a la reputación y fama del individuo es masivo e inmediato (Valverde-Ramos et al., 2024).

2.4.3. Afectación al honor y la reputación.

El derecho al honor es un bien inmaterial intrínsecamente ligado a la dignidad humana. El uso de ultrafalsos lesiona tanto la dimensión interna (autoestima) como la dimensión externa (reputación social) de la persona, despojándola de su integridad moral ante la comunidad. La jurisprudencia ecuatoriana advierte que las expresiones dirigidas a humillar o desacreditar no están protegidas por la libertad de expresión, pues vulneran el núcleo esencial de la dignidad (Erazo Bustamante & Valdivieso Guerrero, 2018).

2.4.4. Límites: La atipicidad de contenidos no calumniosos.

Bajo el principio de legalidad, es fundamental precisar que no todo deepfake implica una calumnia. Existen usos lícitos o penalmente irrelevantes, como aquellos destinados a la sátira, el humor o la parodia (animus iocandi), donde la falta de intención de imputar un delito real excluye la tipicidad de la conducta. Si el contenido es claramente identificable como una broma o ficción, no existe la lesión al bien jurídico necesaria para activar el sistema punitivo (Yépez, 2024).

2.4.5. Insuficiencia frente a deepfakes sexuales y vejatorios.

El principal límite del Artículo 182 radica en su especificidad: solo sanciona la imputación de un delito. Esto genera un grave vacío frente a los deepfakes sexuales o contenidos meramente humillantes; mostrar a una persona en un video pornográfico falso es una conducta profundamente vejatoria que lesiona el honor y la intimidad, pero al no imputar la comisión de un crimen por parte de la víctima, a menudo no puede ser procesada como calumnia (Zambrana Ruiz, 2024).

Esta insuficiencia típica puede dejar sin respuesta penal los casos de violencia sexual digital en que la víctima aparece "realizando" actos que no son delitos, pero que destruyen su reputación y dignidad personal. Por tanto, se concluye que el tipo penal de calumnia es una herramienta incompleta para enfrentar el espectro total de daños morales generados por la IA (Endara-Chamorro et al., 2024).

2.5. Revisión del delito de suplantación de identidad.

Actualmente la línea entre la realidad y la ficción se ha puesto borrosa, gracias a la Inteligencia Artificial, planteando con ello serios debates dogmáticos sobre si la manipulación digital constituye una forma de suplantar a una persona. En este apartado se examina la pertinencia y los límites del delito de suplantación de identidad cuando la conducta infractora se materializa mediante la clonación de voz o rostro a través de un deepfake.

2.5.1 Concepto penal de suplantación de identidad.

En la legislación ecuatoriana, el delito de suplantación de identidad se encuentra tipificado en el artículo 212 del Código Orgánico Integral Penal (COIP), el cual establece que:

> "*la persona que de cualquier forma suplante la identidad de otra para obtener un beneficio para si o para un tercero, en perjuicio de una persona, será sancionada con pena privativa de libertad de uno a tres años" (Asamblea Nacional del Ecuador, 2014).*

Dogmáticamente, esta infracción protege el derecho a la identidad (consagrado en el art.66, numeral 28 de la Constitución ecuatoriana) y se configura como un "delito de resultado", requiriendo imperativamente que el uso de la identidad ajena produzca una consecuencia material o ideal, la obtención de un beneficio indebido y la causación de un perjuicio directo a la víctima (Endara-Chamorro et al., 2024).

2.5.2 El deepfake como forma de simulación de identidad.

Técnicamente, la inteligencia artificial permite la creación de deepfake a través de dos vertientes principales: el deepvoice (clonación de voz original de un individuo para crear secuencias artificiales) y el deepface (superposición de rostros falsificando gestos con resultado extremadamente realista) (Ponce González et al., 2025).

En primera instancia, el uso de estas herramientas informáticas parece encajar en la premisa "de cualquier forma suplante la identidad" del art.212. Si un delincuente utiliza un deepvoice para imitar la voz del gerente de una empresa y ordenar a un empleado que realice

una transferencia bancaria (fraude o phishing), el deepfake actúa como medio perfecto de suplantación de identidad, pues existe un engaño directo para obtener un claro beneficio económico (Espinosa Carvajal & Paredes Fuentes, 2025).

2.5.3 Dificultades probatoria: el problema del beneficio exigido por el tipo penal.

A pesar de que los deepfakes pueden ser utilizados para cometer fraudes, muchas de estas manipulaciones no persiguen un fin patrimonial. Aquí es donde se presenta la principal dificultad de subsunción. El artículo 212 exige que la suplantación se realice "para obtener un beneficio". Sin embargo, cuando se trata de la creación de un deepfake pornográfico, el propósito del agresor no es obtener un rédito económico o un beneficio jurídico, sino humillar, ridiculizar, sexualizar o vejar a la víctima, actos motivados a menudo por el morbo o la venganza (Endara-Chamorro et al., 2024).

Al no existir un "beneficio" claro en los términos que tradicionalmente exige la dogmática penal para este tipo de delitos, resulta difícil aplicar el art.212 a los deepfakes sexuales, ya que, al no concurrir el beneficio exigido, hacerlo implicaría recurrir a la analogía extensiva, expresamente prohibida por el derecho penal ecuatoriano.

2.5.4. Diferencia entre suplantar identidad y manipular la imagen.

La doctrina advierte una diferencia fundamental entre "suplantar" y manipular". Suplantar la identidad implicar usurpar los daros de una persona para actuar en su nombre frente a terceros (como crear cuentas bancarias falsas, firmar contratos o interactuar haciéndose pasar por ella) (Espinosa Carvajal & Paredes Fuentes, 2025).

Por el contrario, el Deepfake sexual manipula la imagen y el rostro de la víctima para colocarla en un escenario ficticio (un video pornográfico), pero el creador del video no está actuando "en nombre" de la víctima ni asumiendo su rol en la sociedad, más bien está cosificándola y exhibiéndola como un objeto audiovisual (Endara-Chamorro et al., 2024).

Ahora bien, haciendo una comparativa general, ¿Encaja el deepfake en el art.212 del COIP? En el tipo penal de suplantación de identidad solo encaja cuando este es utilizado como una herramienta de ingeniería social, por ejemplo, el deepvoice, para con ello engañar a terceros, suplantar a la víctima en una transacción y obtener un beneficio directo, generalmente económico (Ponce González et al., 2025).

No obstante, el tipo penal no encaja en los casos de deepfakes pornográficos, ya que la manipulación de la imagen para denigrar sexualmente a una persona no cumple con los elementos estructurales de "asumir el rol de otro" ni con el elemento teleológico de "obtener un beneficio". Esta insuficiencia muestra que, frente a esta modalidad, el art.212 no ofrece una respuesta penal y hace necesario evaluar, en el Capítulo III, si corresponde tipificar la alteración de la imagen mediante inteligencia artificial (Endara-Chamorro et al., 2024).

2.6. Revisión de otros tipos penales relacionados.

El análisis de la legislación penal ecuatoriana frente a la violencia digital generada por inteligencia artificial no estaría completo sin examinar otras figuras delictivas contenidas en el COIP. Aunque a primera vista ciertos delitos podrían relacionarse con la creación y difusión de deepfakes, un estudio dogmático revela que su aplicación presenta serias limitaciones frente a la naturaleza del contenido sintético.

2.6.1. Extorsión y Sextorsión (Extorsión sexual).

La tecnología deepfake es frecuentemente instrumentalizada para cometer chantajes. El art.185 del COIP tipifica la extorsión, sancionando a quien, con el propósito de obtener provecho patrimonial, obligue a otro mediante intimidación a realizar un acto en perjuicio de su patrimonio (Asamblea Nacional del Ecuador, 2014).

De manera más específica, el artículo 172.1 tipifica la extorción sexual o sextorsión, que castiga a quien induzca o coaccione a otra persona a exhibir su cuerpo para obtener un provecho personal o sexual (Asamblea Nacional del Ecuador, 2014).

- **Limitación frente al contenido sintético:** El límite de estas figuras radica en su elemento subjetivo o teleológico, exigen imperativamente la búsqueda de un provecho (ya sea económico o sexual). Sin embargo, la doctrina y los casos recientes evidencian que muchos agresores que crean deepfakes pornográficos no exigen dinero ni favores a cambio, sino que difunden el material hiperrealista por mera venganza, morbo o "diversión en la crueldad", con el único fin de vejar y destruir la reputación de la víctima. Al no existir la exigencia de un lucro a favor, el tipo penal de extorsión resulta inaplicable en esos casos, lo que genera una dificultad de subsunción (Andino Veloz, 2025).

2.6.2. Ciberacoso y Hostigamiento.

El COIP contempla figuras como el hostigamiento (art.154.2) y el ciberacoso sexual (Art.166, segundo inciso). El hostigamiento sanciona a quien moleste o perturbe de forma insistente a través de medios digitales. Por su parte, el ciberacoso sexual castiga actos de naturaleza sexual cometidos a través de tecnologías de la información (Asamblea Nacional del Ecuador, 2014).

- **Limitación frente al contenido sintético:** La barrera probatoria es alta. El art.166 (Ciberacoso) exige que la conducta se dé utilizando amenazas, abuso de poder o desde una posición de autoridad.

  Por otro lado, el Art.154.2 (hostigamiento) requiere expresamente que el sujeto activo "busque cercanía con la víctima para poder causarle daño a su integridad física o sexual" (Espinosa Carvajal & Paredes Fuentes, 2025).

  Los deepfakes, por el contrario, suelen fabricarse de manera asincrónica, anónima y a distancia, sin que el agresor busque una cercanía física o se ampare en una relación de poder formal, lo que impide que el ilícito encaje perfectamente en estos tipos penales.

2.6.3. Pornografía con utilización de niñas, niños o adolescentes.

Lo más cercano que tiene el Código Orgánico Integral Penal (COIP) para sancionar la manipulación de imágenes son los artículos 103 y 104, que tratan sobre la producción, oferta y comercialización de pornografía en la que participan menores de dieciocho años. El artículo 103, en concreto, castiga a quien "fotografíe, filme, grabe, produzca, transmita o edite materiales... que contenga la representación visual de desnudos o semidesnudos reales o simulados de niñas, niños o adolescentes en actitud sexual".

La palabra "simulados" se incorporó siguiendo las recomendaciones del Convenio de Budapest sobre la Ciberdelincuencia (artículo 9). Su propósito original era sancionar la pornografía infantil creada completamente por computadora (imágenes CGI o patrones totalmente artificiales), es decir, casos en los que no hay ninguna persona real detrás. El problema es que el aprendizaje profundo (deeplearning) ha traído un caso intermedio, la pseudopornografía sintética, que consiste en copiar y superponer de forma hiperrealista el rostro de un menor identificable sobre el cuerpo desnudo, o en acto sexual, de una persona adulta.

A primera vista, el término "simulados" permitiría incluir estos montajes en el artículo 103, y esa inclusión legal de representaciones reales o simuladas impide descartar automáticamente su aplicación por el solo hecho de que el cuerpo pertenezca a un tercero o de que no haya existido contacto físico con el menor. Corresponde, en cambio, examinar el alcance de la representación simulada y de la exigencia de "utilización de niñas, niños o adolescentes" conforme a los principios de legalidad estricta y de protección reforzada, determinando si el uso del rostro o de otros datos biométricos del menor mediante inteligencia artificial satisface dicha exigencia típica. Se trata de una cuestión interpretativa que debe resolverse a partir del propio texto legal, sin que hasta la fecha existan precedentes jurisprudenciales vinculantes de la Corte Nacional de Justicia que fijen un criterio específico sobre la materia.

Por lo tanto, la aplicación del artículo 103 a los deepfakes exige un examen cuidadoso. La subsunción de representaciones sintéticas de menores exige examinar todos los elementos del artículo 103; el origen artificial de la imagen no permite concluir automáticamente su atipicidad. El alcance de la representación simulada debe argumentarse sin ampliar el tipo por analogía, evaluando en cada caso si la pena de trece a dieciséis años resulta aplicable conforme a los elementos típicos efectivamente configurados, sin que ello implique extender la norma más allá de su tenor literal.

A esto se suma otra consideración relevante. El artículo 103 limita a sus víctimas a niñas, niños y adolescentes, de modo que no resulta aplicable a las mujeres mayores de dieciocho años, quienes también pueden ser víctimas de deepfake pornográfico no consentido; esta exclusión etaria, sin embargo, no demuestra por sí sola la ausencia de toda respuesta penal frente a esos casos, pues corresponde examinar si la conducta se ajusta a otras figuras, como el artículo 178, cuyas diversas modalidades no exigen en todos los casos la interceptación o sustracción previa de información. Debe, en consecuencia, diferenciarse la creación sintética del contenido de su posterior difusión, y valorar en cada caso si se utilizan datos, voz o imagen atribuibles a la víctima (Taylor Villafuerte, 2024).

2.7. Problemas de subsunción jurídica frente al uso malicioso de deepfakes.

Este apartado constituye el núcleo crítico del análisis normativo de la presente investigación. La evolución de la cibercriminalidad ha demostrado que el marco legal ecuatoriano se encuentra en una etapa incipiente frente a la inteligencia artificial. Al intentar adecuar el uso malicioso de los deepfakes a las figuras penales tradicionales, los administradores de justicia enfrentan problemas de subsunción jurídica en modalidades concretas —en particular, la manipulación sintética no consentida sin fines sexuales explícitos ni ánimo calumnioso—, sin que ello permita afirmar, de forma general, que el ordenamiento resulta insuficiente frente a todas las variantes del fenómeno. Antes que una insuficiencia

absoluta, corresponde hablar de una insuficiencia demostrada para esas modalidades específicas, reconociendo que otros supuestos —como las representaciones simuladas de menores en actitud sexual, que podrían subsumirse en el artículo 103 del COIP si concurren sus elementos típicos, o los deepfakes calumniosos, subsumibles en el delito de calumnia— sí podrían encontrar respuesta en los tipos vigentes (Chávez Zambrano & San Andrés Molina, 2025).

2.7.1. Falta de verbos rectores y dificultad para sancionar la sola creación.

El primer obstáculo dogmático radica en la ausencia de verbos rectores adecuados para la era del contenido sintético en algunas figuras penales. Los tipos penales actuales, como la violación a la intimidad (art. 178), sancionan conductas como acceder, interceptar, examinar, retener, grabar, reproducir, difundir o publicar información, pero no incluyen verbos como crear, generar, manipular o alterar contenido mediante inteligencia artificial. El artículo 104 del COIP, en particular, castiga a quien comercialice o posea pornografía infantil, pero omite por completo a la persona que "cree" dicho contenido falso a partir de fotografías cotidianas extraídas de redes sociales; no obstante, este vacío del artículo 104 debe analizarse conjuntamente con el artículo 103, que sí incluye los verbos producir y editar, por lo que la producción de representaciones sintéticas de niñas, niños o adolescentes puede, según el caso, encontrar respuesta en esta última disposición (Ponce González et al., 2025).

Esta omisión genera una enorme dificultad para sancionar la sola creación del deepfake cuando aún no existe difusión. La doctrina internacional sobre el Abuso Sexual Basado en Imágenes (IBSA), advierte que limitar el castigo penal únicamente al acto de compartir el material desprotege a la víctima, ya que la invasión a la autonomía sexual y la vulneración de la intimidad se consuman desde el momento mismo en que el agresor fabrica el contenido no consensuado, independientemente de si llega a publicarlo o no (Rigotti et al., 2024).

2.7.2. Problemas para acreditar perjuicio concreto o beneficio.

Figuras como la suplantación de identidad (art.212 del COIP) o la extorsión exigen como elemento subjetivo la obtención de un "beneficio" (generalmente patrimonial) y la acreditación de un "perjuicio material concreto".

El problema de subsunción es evidente: muchos deepfakes sexuales y vejatorios no se crean para exigir dinero, sino por cinismo, morbo, venganza o el simple deseo de cosificar, humillar y ridiculizar a la víctima. Al no existir un beneficio económico directo, que es lo que exige el tipo penal, la conducta no encaja en esas figuras y, si no concurre otro tipo, quedaría sin sanción (Andino Veloz, 2025).

2.7.3. Contenido sexual falso y falta de participación material.

Otro grave problema de subsunción se presenta al intentar encajar el contenido sexual falso en delitos tradicionales contra la integridad sexual, debido a que la víctima no participó materialmente en el acto.

En el caso de la pornografía infantil, el COIP en su art. 103 incluye el término "simulados", lo que exige examinar con detalle su alcance frente a montajes que utilizan el rostro de una víctima real superpuesto en el cuerpo de un tercero. La falta de participación física del menor en la escena no demuestra por sí sola la atipicidad de la conducta, dado que el propio artículo 103 contempla representaciones simuladas; se trata más bien de una cuestión interpretativa —el alcance de la representación simulada y de la utilización de la imagen del menor— que debe resolverse sin presumir ni excluir automáticamente la adecuación típica (Taylor Villafuerte, 2024).

2.7.4. Dificultades probatorias por el hiperrealismo.

Los problemas de encuadre legal se ven agravados por el ámbito procesal. Las Redes Generativas Antagónicas (GANs) producen un nivel de hiperrealismo que hace que los rostros y voces generados sintéticamente sean prácticamente indistinguibles de los reales.

Esta perfección técnica dificulta inmensamente la labor probatoria, las víctimas sufren el escarnio público porque la sociedad asume la veracidad del material, mientras que en los tribunales carece de peritajes técnicos suficientes para demostrar indubitablemente la falsedad algorítmica del contenido, paralizando los procesos de defensa (Chávez Zambrano & San Andrés Molina, 2025).

2.7.5. Atipicidad, subsunción forzada y principio de legalidad.

En conclusión, intentar encuadrar la creación de deepfakes sexuales o vejatorios en delitos como calumnia, suplantación de identidad o pornografía, cuando no concurren los elementos del tipo, resultaría en una subsunción forzada. El derecho penal ecuatoriano está regido por el principio de legalidad y tipicidad estricta (*nullum crimen, nulla poena sine lege*), el cual prohíbe expresamente la interpretación analógica extensiva para crear delitos o aplicar sanciones (Santacruz Torré et al., 2025).

Si un juez fuerza la ley actual para castigar a un creador de deepfakes en supuestos que no se ajustan a los elementos típicos vigentes, estaría vulnerando el principio de legalidad; si no lo hace, el acto podría declararse atípico y la víctima quedaría en indefensión. La legalidad estricta impide extender los tipos por analogía; no obstante, determinadas modalidades pueden satisfacer los elementos de las figuras vigentes. La eventual atipicidad debe justificarse respecto de conductas concretas y no del fenómeno en su totalidad, lo que evidencia la conveniencia de examinar, en el Capítulo III, la viabilidad de un tipo penal autónomo (Rosales Murillo, 2024).

2.8. Conclusión del capítulo II:
Dificultades de subsunción y examen de una tipificación autónoma.

A lo largo del presente capítulo se ha podido observar, a partir del análisis dogmático del COIP, que el marco legal ecuatoriano actual presenta dificultades frente a determinadas modalidades del uso malicioso de la inteligencia artificial. El análisis permite identificar dificultades relevantes de subsunción en determinadas modalidades de deepfake malicioso,

especialmente en la creación sintética no consentida cuando no concurren los elementos de los tipos vigentes. Esto no supone ausencia absoluta de respuesta penal: otras modalidades pueden adecuarse a figuras existentes según las circunstancias del caso. La tipificación autónoma se examina, en el Capítulo III, como alternativa sujeta a los principios de legalidad, mínima intervención y proporcionalidad.

Al intentar subsumir la creación o difusión de deepfakes en delitos tradicionales como la violación a la intimidad (Art. 178), la calumnia (art. 182), la suplantación de identidad (art. 212) o incluso la pornografía infantil (arts. 103 y 104), los operadores de justicia enfrentan dificultades de subsunción en los supuestos concretos analizados en este capítulo, reconociendo expresamente que otras modalidades podrían configurar calumnia, suplantación de identidad u otras figuras cuando reúnan sus elementos típicos (Ponce González et al., 2025).

Estas infracciones pueden resultar insuficientes frente a determinadas modalidades del contenido sintético, en particular cuando no se configuran elementos estructurales como la obtención de un beneficio patrimonial o la imputación formal de un delito penado por la ley; en el caso del artículo 103, en cambio, la falta de participación física de la víctima en la escena no excluye por sí sola la aplicación del tipo, dado que este contempla expresamente representaciones simuladas, por lo que corresponde examinar dicho alcance en cada caso concreto (Endara-Chamorro et al., 2024).

Además, la legislación actual presenta una deficiencia en algunas figuras al omitir verbos rectores como “crear”, “alterar” o “generar” —así ocurre, por ejemplo, en el artículo 104—, lo que puede dejar sin cobertura penal específica determinados actos de fabricación no consentida de material hiperrealista cuando estos no llegan a difundirse ni se ajustan a los verbos ya previstos en otras figuras, como “producir” y “editar” en el artículo 103.

Intentar forzar la aplicación de las normas vigentes para castigar conductas que no se ajusten a sus elementos típicos vulneraría de manera directa los principios de legalidad y

tipicidad estricta (nullum crimen sine lege), pilares fundamentales de un Estado constitucional de derechos. Esto puede determinar atipicidad en los supuestos en que no concurran los elementos de un tipo vigente, sin que ello equivalga a una extensión analógica prohibida cuando la subsunción sea jurídicamente válida.

Estas dificultades muestran que ciertas modalidades del deepfake (en especial las pornográficas, humillantes o vejatorias) no siempre encuentran respuesta en los delitos informáticos o contra el honor ya existentes. La magnitud del daño psicológico, social y reputacional que ocasionan justifica examinar si corresponde una intervención penal directa.

Las dificultades identificadas justifican evaluar en el capítulo III una tipificación autónoma frente a la reforma de tipos vigentes y las medidas extrapenales, condicionada al respeto de los límites del ius puniendi. En dicho apartado se estructurarán y propondrán los elementos jurídicos y dogmáticos necesarios (conducta, sujeto activo y pasivo, bienes jurídicos y elemento subjetivo) para la formulación de una tipificación autónoma del delito de deepfake en el Ecuador, dotando al sistema de justicia de un mecanismo legal proporcional y garantista frente a esta forma de violencia.

Capítulo III

Elementos jurídicos para la configuración de un tipo penal autónomo sobre el uso malicioso de deepfakes en Ecuador.

3.1. Necesidad jurídica de una tipificación autónoma.

La vertiginosa evolución de la Inteligencia Artificial ha generado un escenario donde las herramientas tradicionales del Derecho Penal ecuatoriano resultan anacrónicas. A pesar de que el Ecuador cuenta con una base de derechos fundamentales robusta, la ausencia de una normativa técnica y específica que sancione la creación y difusión de contenidos sintéticos ha generado dificultades de adecuación típica que comprometen la seguridad jurídica y la integridad de los ciudadanos (Chávez Zambrano & San Andrés Molina, 2025).

3.1.1. Insuficiencia de los tipos penales vigentes.

El análisis de la legislación actual demuestra que algunos artículos del Código Orgánico Integral Penal (COIP) que se intentan aplicar a los deepfakes presentan dificultades delimitadas para modalidades específicas; por ejemplo, el artículo 178 (violación a la intimidad) se centra en conductas de acceso, registro, reproducción y difusión de información existente, lo que exige examinar caso por caso si el fenómeno de la "creación algorítmica" desde cero satisface su objeto material (Endara-Chamorro et al., 2024). De igual manera, el artículo 104, referido a la comercialización de pornografía infantil, sanciona la posesión y venta, pero no menciona expresamente a quien "cree" dicho material mediante IA; sin embargo, este artículo debe leerse conjuntamente con el artículo 103, que sí contempla los verbos producir y editar, por lo que la afirmación de un vacío total no resulta aplicable de forma general a la producción de representaciones sintéticas de menores (Taylor Villafuerte, 2024).

3.1.2. Necesidad de certeza jurídica y Prevención de impunidad.

Bajo el principio de legalidad, la conducta punible debe estar descrita de forma exacta e inequívoca en la ley. En la actualidad, ante la inexistencia de un delito específico de "deepfake sexual" o "suplantación digital por IA", las denuncias por hechos que no encajan en los tipos

vigentes podrían ser desestimadas por atipicidad, aunque no existan registros oficiales que permitan cuantificarlo. Esto puede generar una sensación de impunidad que desprotege a las víctimas y permite que los agresores operen en un "limbo jurídico"(Saquicela Rodas & Chica Mejía, 2023).

La tipificación autónoma permitiría a fiscales y jueces actuar con base en normas que comprendan que el daño radica no solo en la difusión, sino en el acto mismo de la generación algorítmica sin consentimiento. Al establecer sanciones proporcionales y previsibles, el Estado envía un mensaje de reproche social contra el uso malicioso de la tecnología, actuando como un disuasivo necesario para proteger el honor y la identidad de los ciudadanos en el ciberespacio (Endara-Chamorro et al., 2024).

3.1.3. Protección efectiva de las víctimas.

La creación de un tipo penal específico no es solo una necesidad técnica, sino una medida de protección para grupos vulnerables. Las fuentes señalan que el 99% de las víctimas de pornografía deepfake son mujeres, lo que convierte a esta tecnología en una nueva herramienta de violencia de género digital. Sin una ley específica, el daño psicológico, reputacional y moral causado por la simulación de actos que la persona nunca realizó podría quedar sin reparación integral en la vía penal, en las modalidades no cubiertas por los tipos vigentes, lo que deja a las víctimas en una situación de desprotección (Taylor Villafuerte, 2024).

3.1.4. Evitar interpretaciones extensivas del derecho penal.

La dogmática penal advierte sobre el riesgo de forzar la interpretación de leyes antiguas para que encajen en hechos nuevos, lo cual violenta el mandato de tipicidad estricta. Intentar "ajustar" la calumnia o la suplantación de identidad tradicional a la complejidad de un deepfake compromete las garantías del procesado y genera jurisprudencia contradictoria. Una respuesta penal específica, diseñada bajo criterios técnicos y no bajo un populismo punitivo apresurado,

permitiría sancionar el uso malicioso de la IA sin coartar la innovación legítima ni la libertad de expresión (Carnevali Rodríguez, 2008)

3.1.5. Justificación de una respuesta penal específica.

La necesidad de una respuesta penal específica frente a los deepfakes no responde únicamente a una actualización técnica, sino a la obligación del Estado de garantizar la tutela judicial efectiva frente a formas de agresión que el derecho tradicional no puede capturar con precisión (Valverde-Ramos et al., 2024). La justificación para esta tipificación autónoma se asienta en los siguientes pilares:

Superación de la ambigüedad en los verbos rectores: Los tipos penales actuales en el Ecuador, como el artículo 178 del COIP, se centran en el acceso, registro o difusión de información real preexistente. Sin embargo, el núcleo de la lesividad de un deepfake radica en el acto de creación sintética, una conducta que, respecto de personas adultas, no está prevista de forma expresa en los tipos vigentes, ya que estos no incluyen verbos como “generar” o “crear” contenido falso mediante IA (en el caso de menores, el artículo 103 sí contempla los verbos producir y editar). Una norma específica permitiría incluir el verbo rector "crear", atendiendo la falta de cobertura en la fase de producción del material (Taylor Villafuerte, 2024).

Garantía de taxatividad y seguridad jurídica: Bajo el principio de nullum crimen sine lege, la aplicación de leyes por analogía está prohibida en materia penal. Forzar la interpretación de delitos como la calumnia o la suplantación de identidad para que abarquen realidades algorítmicas que no cumplen sus elementos típicos genera incertidumbre jurídica y resoluciones judiciales contradictorias. Una respuesta penal específica proporcionaría a los jueces un marco descriptivo claro y unívoco, asegurando que la sanción sea previsible para el ciudadano y técnicamente correcta para el juzgador (Santacruz Torré et al., 2025).

Abordaje de la pluri ofensividad del daño: A diferencia de los delitos tradicionales, el uso malicioso de deepfakes suele lesionar múltiples bienes jurídicos de forma simultánea: la

intimidad, el honor, la propia imagen y la identidad digital. Las leyes generales a menudo solo protegen una de estas aristas, dejando las demás desamparadas. Una tipificación autónoma permitiría reconocer la gravedad integral del daño causado por la "falsedad profunda", estableciendo penas proporcionales a la afectación de la dignidad humana en el entorno digital (Zambrana, 2025).

Protección frente a la violencia de género y grupos vulnerables: Existe una justificación ética y social urgente: el 96% de los deepfakes en la red son de naturaleza pornográfica no consentida y afectan casi exclusivamente a mujeres y niñas. El marco legal actual de Ecuador es insuficiente para sancionar estas formas de violencia sexual digital cuando las víctimas son adultas o cuando el contenido es puramente sintético. Una ley específica enviaría un mensaje claro de reproche social contra el uso de la tecnología para la cosificación y deshumanización, garantizando que el ciberespacio no sea un territorio de impunidad (Rosales Murillo, 2024).

3.2. Fundamento político-criminal de la regulación penal de deepfakes.

La legitimación del Estado para intervenir penalmente frente a los deepfakes se sustenta en la necesidad de actualizar la política criminal ante una realidad donde la tecnología ha desbordado los mecanismos de control social tradicionales. En un Estado constitucional de derechos, el ius puniendi no solo es una potestad de castigo, sino un deber de protección de los ciudadanos frente a agresiones que lesionan el núcleo esencial de su dignidad (Saquicela Rodas & Chica Mejía, 2023).

3.2.1. Política criminal frente a nuevas formas de violencia digital.

La política criminal contemporánea debe responder a las denominadas "criminologías del futuro", donde delitos como el fraude algorítmico o la violencia sexual digital presentan características de anonimato, automatización y ejecución instantánea. La intervención penal se justifica porque los modelos de regulación tradicionales resultan obsoletos ante ataques que se

completan antes de ser detectados y que utilizan identidades sintéticas para evadir la justicia. (Sun, 2026).

El Estado ecuatoriano está obligado a diseñar una respuesta penal que trascienda lo reactivo y se convierta en una barrera efectiva contra el uso de la Inteligencia Artificial (IA) como arma de opresión.

3.2.2. Protección de bienes jurídicos relevantes y lesividad de la conducta.

El fundamento central de la intervención penal es la protección de bienes jurídicos pluriofensivos que son atacados simultáneamente: el honor, la intimidad, la propia imagen y la libertad sexual. La lesividad de los deepfakes no es ficticia; aunque el contenido sea sintético, el daño moral y psicológico en la víctima es real y, a menudo, irreversible debido a la naturaleza hiperrealista del material (Taylor Villafuerte, 2024).

Dado que la sociedad otorga un valor de "verdad" a lo visual, la simulación de actos denigrantes constituye un ataque gravísimo a la integridad moral, lo que legitima al Derecho Penal para actuar como guardián de estos derechos fundamentales (Ponce González et al., 2025).

3.2.3. Riesgo social del uso masivo de inteligencia artificial maliciosa.

Existe un riesgo social sistémico derivado de la "democratización" de herramientas de manipulación que no requieren pericia técnica. El avance de aplicaciones de libre acceso permite que personas sin experiencia en computación generen rostros falsos e hiperrealistas con solo unos clics. Esta accesibilidad ha permitido que casi cualquier individuo pueda crear contenidos sintéticos utilizando herramientas prevalentes, lo que dificulta enormemente las labores de control y detección. Como consecuencia, la barrera de entrada para actores malintencionados ha desaparecido, permitiendo la generación masiva de archivos de video, imagen o voz que simulan realidades inexistentes de forma barata y rápida (Ponce González et al., 2025).

El uso masivo de deepfakes para la desinformación política y el acoso erosiona profundamente la confianza pública en las instituciones y en la veracidad de la información, amenazando la estabilidad democrática. La tecnología ya se ha empleado para mostrar a figuras políticas realizando declaraciones falsas o conductas delictivas ficticias, lo que busca socavar su credibilidad. Incluso los contenidos de baja calidad, si se despliegan en momentos críticos como la víspera de una elección, tienen la capacidad de sacudir la confianza ciudadana, independientemente de que el gobierno desmienta la autenticidad del material posteriormente. Este fenómeno introduce una "heurística del realismo", donde la apariencia visual reduce la duda y potencia la creencia en narrativas engañosas, eludiendo los filtros periodísticos tradicionales (Schuette, 2025).

Esta escala de amenaza puede conducir a la configuración de una "sociedad de cero confianzas", donde resulta cada vez más fácil sembrar dudas sobre eventos reales, permitiendo que actores responsables evadan la rendición de cuentas bajo lo que se denomina el "dividendo del mentiroso". En este entorno, los sentidos humanos dejan de ser suficientes para captar la realidad, obligando a los ciudadanos a dudar incluso de las evidencias visuales y auditivas más contundentes. La desinformación se vuelve así un fenómeno ubicuo y reticular que altera la percepción colectiva de la verdad, lo cual es especialmente peligroso en contextos de conflictos internacionales o crisis sociales (Chávez Zambrano & San Andrés Molina, 2025).

3.2.4. Necesidad de protección reforzada frente a daños graves.

La política criminal contemporánea debe reconocer que el uso malicioso de la inteligencia artificial no es un riesgo neutral, sino que impacta de manera desproporcionada en grupos vulnerables, tales como mujeres, niñas, niños y adolescentes (Rosales Murillo, 2024).

Las estadísticas internacionales son alarmantes: se estima que entre el 96% y el 99% del contenido deepfake disponible en la red es de naturaleza pornográfica no consentida, y que el 99% de las víctimas de estas representaciones sintéticas son mujeres. Este fenómeno

constituye una nueva y agresiva dimensión de la violencia de género digital, donde el cuerpo femenino es cosificado y deshumanizado mediante la creación de una existencia virtual forzada para satisfacer deseos masculinos de dominación (Taylor Villafuerte, 2024).

La gravedad del daño causado por la "falsedad profunda" es profunda y, a menudo, irreversible, manifestándose en una ruptura social del proyecto de vida de la víctima. El impacto psicológico incluye cuadros severos de ansiedad, depresión, estrés postraumático y un riesgo latente de conductas autolesivas o suicidio, derivados de la humillación y el estigma público. Al otorgar la sociedad una alta credibilidad a la evidencia visual, la víctima queda expuesta a un escarnio masivo que afecta sus relaciones familiares, laborales y su salud emocional (Valverde-Ramos et al., 2024).

3.2.5. Diferencia entre regulación preventiva, administrativa, civil y penal.

La legitimidad del Derecho Penal como ultima ratio se asienta en la insuficiencia de otras vías para contener el fenómeno:

Regulación Preventiva y Administrativa: Aunque Ecuador cuenta con la garantía del Habeas Data y la Ley Orgánica de Protección de Datos Personales (LOPDP) para rectificar o suprimir información, estas herramientas enfrentan limitaciones estructurales críticas en el entorno digital. La acción de Habeas Data, si bien es jurídicamente procedente para proteger imágenes íntimas como "datos sensibles", carece de la celeridad necesaria para frenar la viralización instantánea y masiva en redes sociales, donde el daño se vuelve irreversible en cuestión de segundos. Además, persiste la falta de especialización técnica de los jueces y la ausencia de obligaciones claras para que las plataformas retiren el contenido de forma expedita, lo que deja a la garantía constitucional como un mecanismo reactivo e insuficiente (Espinosa Carvajal & Paredes Fuentes, 2025).

Regulación Penal: El Derecho Penal constituye la única instancia capaz de abordar la culpabilidad del autor y el dolo específico de causar daño (animus injuriandi), trasladando la

responsabilidad de la herramienta tecnológica a la persona humana que decide instrumentalizarla. Al tipificar estas conductas, el Estado envía un mensaje normativo contundente: la creación y difusión de realidades sintéticas para humillar o violentar es una conducta criminal intolerable en una sociedad democrática. Solo la sanción punitiva permite reconocer la pluriofensividad del delito, protegiendo simultáneamente la intimidad, el honor y la libertad de la víctima (Ruiz, 2025).

Regulación Civil: El enfoque del derecho civil, centrado en el resarcimiento económico, resulta insuficiente para reparar la magnitud del daño causado por un deepfake de naturaleza sexual o difamatoria. La compensación monetaria no puede restituir la reputación social destruida ni sanar el profundo trauma psicológico, que a menudo incluye cuadros de ansiedad, depresión y un riesgo latente de suicidio en las víctimas. La naturaleza hiperrealista de los ultrafalsos genera un escarnio público de tal escala que las soluciones patrimoniales no logran restaurar la dignidad ni la integridad moral del individuo (Taylor Villafuerte, 2024).

En conclusión, el fundamento político-criminal para una ley de deepfakes en Ecuador reside en el deber estatal de garantizar la seguridad ciudadana y la integridad de la persona frente a agresiones tecnológicas que, por su gravedad y escala, rebasan cualquier otra forma de control jurídico (Rosales Murillo, 2024).

3.3. Límites constitucionales al ius puniendi del Estado.

La propuesta de tipificación para el uso malicioso de deepfakes no debe entenderse como un ejercicio de expansión punitiva desmedida o "populismo punitivo", fenómeno caracterizado por la inflación de tipos penales y el endurecimiento de penas como respuesta simbólica a la inseguridad (Santacruz Torré et al., 2025).

Por el contrario, la intervención del Estado en esta materia debe someterse estrictamente a los límites materiales y axiológicos definidos por la Constitución, los cuales actúan como un freno a la arbitrariedad y garantizan que el Derecho Penal intervenga únicamente ante las

conductas más graves que lesionen la dignidad humana (Saquicela Rodas & Chica Mejía, 2023).

3.3.1. Principio de legalidad y prohibición de tipos penales ambiguos.

El principio de legalidad (nullum crimen, nulla poena sine lege) es el pilar fundamental que exige que las leyes penales sean claras, precisas, previas y escritas (Saquicela Rodas & Chica Mejía, 2023).

En el contexto de las tecnologías emergentes, esto implica un mandato de taxatividad estricta, prohibiendo la creación de tipos penales ambiguos, abiertos o "en blanco" que deleguen la definición de la conducta a la discrecionalidad judicial o administrativa (Carnevali Rodríguez, 2008). Una regulación técnica de los deepfakes debe describir con exactitud el comportamiento prohibido para brindar certeza jurídica y evitar que la norma sea utilizada para criminalizar actividades legítimas como la parodia o la sátira (Saquicela Rodas & Chica Mejía, 2023).

3.3.2. Principio de mínima intervención penal.

Este principio establece que el Derecho Penal debe ser el último recurso de control social, activándose solo cuando otros mecanismos (civiles, administrativos o preventivos) han resultado insuficientes para proteger los derechos de los ciudadanos (Carnevali Rodríguez, 2008). La regulación de los deepfakes no pretende castigar toda manipulación digital, sino que reserva la sanción punitiva para aquellos casos donde la lesividad social es extrema y el daño a la víctima es irreparable por otras vías, promoviendo paralelamente la educación digital y políticas de ciberseguridad (Santacruz Torré et al., 2025).

3.3.3. Principio de lesividad y protección de bienes jurídicos.

Bajo el principio de lesividad, para que una conducta sea penalmente relevante debe amenazar o lesionar, sin justa causa, un bien jurídico protegido (Asamblea Nacional del Ecuador, 2014).

En la propuesta de ley, el castigo no se fundamenta en la tecnología per se, sino en la ofensa efectiva a bienes jurídicos de rango superior como el honor, la intimidad y la integridad moral (Ponce González et al., 2025).

Se excluyen de la sanción penal aquellos usos de IA que, aunque sintéticos, carecen de dolo lesivo o no producen un perjuicio real en la esfera de derechos de terceros (Carnevali Rodríguez, 2008).

3.3.4. Principio de proporcionalidad y necesidad de la pena.

El principio de proporcionalidad constituye un límite material infranqueable al ius puniendi estatal, exigiendo un equilibrio estricto y una relación coherente entre la gravedad de la infracción cometida y la intensidad de la sanción impuesta (Santacruz Torré et al., 2025). De acuerdo con el artículo 76 de la Constitución de la República, la ley debe asegurar que la privación de derechos sea proporcional a la lesión efectiva del bien jurídico protegido, evitando que la respuesta punitiva se convierta en un ejercicio arbitrario del poder (Asamblea Nacional del Ecuador, 2008). En este sentido, la proporcionalidad busca el control de todo acto público que pueda lesionar derechos fundamentales, funcionando como una garantía del debido proceso que impide la aplicación de penas excesivas (Saquicela Rodas & Chica Mejía, 2023).

Bajo esta lógica, el legislador ecuatoriano tiene el deber de evitar la creación de penas desproporcionadas que respondan únicamente a la alarma mediática o a presiones de la opinión pública, fenómeno conocido como "populismo punitivo". La política criminal debe basarse en un diagnóstico criminológico serio y no en un "derecho penal simbólico" que endurece las sanciones para apaciguar la inseguridad percibida, descuidando el fin preventivo y de rehabilitación social (Santacruz Torré et al., 2025). Según el artículo 52 del COIP, los fines de la pena deben ser la prevención general y el desarrollo progresivo de las capacidades de la persona sentenciada, asegurando que la sanción sea idónea y necesaria para su reinserción integral en la sociedad (Saquicela Rodas & Chica Mejía, 2023).

3.3.5 Principio de culpabilidad.

El principio de culpabilidad (nullum crimen sine culpa) constituye un pilar fundamental del Derecho Penal democrático, al establecer que no puede haber pena sin reprochabilidad individual, proscribiendo así cualquier forma de responsabilidad objetiva. Para que una persona sea sancionada penalmente, no basta con la producción de un resultado lesivo; es imperativo que haya actuado con dolo o imprudencia y que posea la capacidad psicofísica de comprender la ilicitud de su conducta y determinarse conforme a ese conocimiento (Asamblea Nacional del Ecuador, 2014). En este sentido, la culpabilidad actúa como el factor que permite la aparición de la pena al valorar el desvalor del autor sobre el hecho cometido, fundamentando la sanción en la libertad de decisión del individuo (Zambrana, 2025).

En el contexto específico de los deepfakes, la responsabilidad penal debe centrarse inexorablemente en el sujeto humano que instrumentaliza la herramienta tecnológica para lesionar bienes jurídicos ajenos (Ávila Zea & Zamora Vázquez, 2024). Bajo el marco legal vigente y la doctrina dominante, se descarta categóricamente atribuir personalidad jurídica o responsabilidad criminal a la Inteligencia Artificial como ente autónomo, puesto que los algoritmos carecen de autoconciencia, voluntad propia y autonomía moral para ser sujetos de reproche punitivo. Por tanto, la autoría del hecho recae en la persona natural “ya sea el creador del contenido o su difusor” que, mediante el dominio del hecho o la utilización del sistema como un instrumento (autoría mediata), actúa con una intención manifiesta de causar daño o animus injuriandi (Ruiz, 2025).

La determinación de este elemento subjetivo es crucial para diferenciar los usos maliciosos de aquellos legítimos, como la parodia o la sátira, asegurando que solo sean castigadas las conductas ejecutadas con conocimiento de la falsedad del material y voluntad de vejar a la víctima. El dolo en estos delitos se manifiesta en la deliberada manipulación de datos biométricos reales para crear una realidad sintética denigrante, evidenciando un desprecio hacia la dignidad y la identidad de la persona afectada (Ponce González et al., 2025).

3.4. Derecho comparado sobre regulación de deepfakes.

El análisis de la normativa internacional revela diversas estrategias para enfrentar los desafíos de la Inteligencia Artificial (IA), ofreciendo un marco de referencia para la actualización del sistema penal ecuatoriano. Si bien el avance es heterogéneo, las tendencias globales se inclinan hacia la creación de normas técnicas que priorizan la protección de los derechos de la personalidad en el entorno digital (Chávez Zambrano & San Andrés Molina, 2025).

3.4.1. Unión Europea y regulación de inteligencia artificial.

La Unión Europea (UE) se ha posicionado como el referente global más avanzado mediante la aprobación del Reglamento de Inteligencia Artificial (AI Act, 2024; European Union, 2024).

El instrumento adopta un innovador enfoque basado en el riesgo, clasificando los sistemas de IA en cuatro niveles (inaceptable, alto, limitado y mínimo) según su impacto potencial en los derechos y la seguridad de los ciudadanos. Las prácticas de riesgo inaceptable, como el uso de sistemas que explotan vulnerabilidades o manipulan el comportamiento humano para causar perjuicios considerables, están estrictamente prohibidas. Por otro lado, los sistemas de alto riesgo —utilizados en áreas sensibles como la justicia, la educación, la migración o la contratación— quedan sujetos a requisitos rigurosos de transparencia, trazabilidad, calidad de los datos y una necesaria supervisión humana efectiva (Chávez Zambrano & San Andrés Molina, 2025).

Complementariamente, en mayo de 2024, la UE adoptó la Directiva (UE) 2024/1385 sobre la lucha contra la violencia contra las mujeres y la violencia doméstica. Este hito legislativo es la primera ley vinculante en la región que armoniza la criminalización del "abuso sexual basado en imágenes" (IBSA). El artículo 5 de la Directiva obliga a los Estados miembros a penalizar la difusión no consentida de material que represente actividades sexualmente explícitas, incluyendo explícitamente el material manipulado o alterado mediante IA. La norma

reconoce que los deepfakes sexuales constituyen una forma de agresión que atenta contra la autonomía sexual y la dignidad, exigiendo además medidas para la retirada rápida del contenido ilegal de las plataformas en línea bajo las directrices de la Ley de Servicios Digitales (DSA) (Rigotti et al., 2024).

3.4.2. Transparencia en contenidos generados por IA.

Un pilar fundamental de la regulación europea es la obligación de transparencia (Reglamento UE 2024/1689, 2024). El artículo 50 del AI Act exige que los sistemas que generen o manipulen contenido (imágenes, audio o vídeo) que se asemeje a personas reales notifiquen a los usuarios que el material ha sido generado de forma artificial o manipulado (Reglamento UE 2024/1689, 2024).

Esta medida busca mitigar el engaño y permitir que los ciudadanos mantengan el control sobre su percepción de la realidad, exceptuando únicamente los casos de parodia o sátira bajo criterios de libertad de expresión (Zambrana, 2025).

3.4.3. Estados Unidos: regulación fragmentada y específica sobre deepfakes.

A diferencia del enfoque armonizado de la Unión Europea, Estados Unidos carece de una ley federal integral que regule la Inteligencia Artificial en su totalidad, lo que ha dado lugar a un panorama legislativo fragmentado donde la iniciativa ha sido liderada principalmente por los estados individuales. Esta respuesta descentralizada se caracteriza por la creación de un "mosaico" de normas que varían significativamente en alcance y ejecución, centrándose de forma casi exclusiva en dos dominios críticos: la protección electoral y la violencia sexual digital (Ugwuoke & Sanfilippo, 2025).

Virginia: Virginia: Se consolidó como el estado pionero al promulgar en marzo de 2019 la primera ley (HB 2678) destinada específicamente a criminalizar la difusión de deepfakes sexuales no consensuados. La norma sanciona el uso de contenido manipulado con el fin de

coaccionar, acosar o intimidar a la víctima, estableciendo penas de hasta doce meses de cárcel y multas de 2.500 dólares (Ugwuoke & Sanfilippo, 2025).

Texas y California: Ambos estados actuaron de forma temprana en 2019 para proteger la integridad de los procesos democráticos. Texas (SB 751) fue el primero en criminalizar los deepfakes políticos diseñados para influir en el resultado de una elección. Por su parte, California aprobó un paquete de leyes (AB 730 y AB 602) que no solo prohíbe la distribución de material electoral engañoso cerca de los comicios, sino que también otorga a las víctimas de pornografía sintética el derecho a emprender acciones legales privadas para exigir reparaciones económicas y medidas cautelares. Recientemente, California ha endurecido este marco con la ley AB 2655 (2024), que obliga a las grandes plataformas a bloquear o etiquetar contenidos falsos meses antes de una elección (Ugwuoke & Sanfilippo, 2025).

Tennessee: En 2024, este estado aprobó la Ley ELVIS (Ensuring Likeness Voice and Image Security Act), convirtiéndose en la primera normativa técnica diseñada específicamente para proteger los rasgos de la personalidad de los artistas. Esta ley prohíbe el uso de IA para imitar o clonar la voz y la imagen de una persona sin autorización expresa, buscando salvaguardar la identidad vocal y comercial frente a la generación de contenidos sintéticos no deseados (Ponce González et al., 2025).

Nivel Federal: Tras años de intentos fallidos con propuestas como la Malicious Deepfakes Prohibition Act, el Congreso de los Estados Unidos aprobó la TAKE IT DOWN Act (Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act), sancionada el 19 de mayo de 2025. La norma no constituye una regulación general de todo deepfake malicioso: se circunscribe a las imágenes íntimas no consensuadas (non-consensual intimate images), tanto auténticas como generadas o alteradas mediante IA. Tipifica como delito federal publicar dicho contenido o amenazar con publicarlo con fines de intimidación, coacción o extorsión, con penas de hasta dos años de prisión, y exige a las

plataformas alcanzadas ("covered platforms") implementar, a más tardar el 19 de mayo de 2026, un mecanismo de notificación y retiro que remueva el contenido dentro de las 48 horas siguientes a una notificación válida, bajo supervisión de la Comisión Federal de Comercio (FTC). La ley no criminaliza, en cambio, la mera creación de la imagen sintética cuando esta no llega a publicarse ni a ser objeto de amenaza de difusión.

3.4.4. Regulación de deepfakes sexuales no consentidos en otras latitudes.

Diversas naciones han implementado respuestas punitivas y regulatorias contundentes para frenar la proliferación de deepfakes sexuales, reconociendo que los marcos legales tradicionales son insuficientes ante la rapidez y el daño masivo de la Inteligencia Artificial (Valverde-Ramos et al., 2024). Estas legislaciones buscan no solo sancionar la difusión, sino también intervenir en la fase de creación y posesión de material sintético (Rosales Murillo, 2024).

Francia: Ha adoptado una legislación de vanguardia que prohíbe expresamente la creación y difusión de deepfakes no autorizados. La norma penaliza la manipulación de imágenes y voces sin el consentimiento del titular, ampliando la protección jurídica no solo al daño real causado a la reputación, sino también al daño potencial a la privacidad. Además, establece mecanismos procesales claros para que las víctimas puedan accionar legalmente de forma expedita (Lovells, 2024).

Reino Unido: En abril de 2024, el Ministerio de Justicia anunció su intención de tipificar la creación de deepfakes sexualmente explícitos como enmienda al proyecto de Criminal Justice Bill; se trataba entonces de un anuncio político, no de una norma vigente, y ese proyecto de ley decayó con la disolución del Parlamento para las elecciones de julio de 2024. La reforma se materializó posteriormente en la sección 138 de la Data (Use and Access) Act 2025, que modifica la Sexual Offences Act 2003 y entró en vigor el 6 de febrero de 2026. Esta norma tipifica la creación o el encargo de creación, sin consentimiento, de imágenes

sexuales simuladas ("purported sexual images"), incluidos los deepfakes generados por IA; a diferencia del anuncio de 2024, la versión finalmente vigente no exige acreditar la intención de causar alarma, humillación o angustia, bastando la ausencia de consentimiento. Esta tipificación de la creación se suma a las infracciones ya existentes de distribución o amenaza de distribución de imágenes íntimas no consentidas, vigentes como infracción prioritaria bajo la Online Safety Act 2023 desde el 31 de enero de 2024. Un hito mediático en esta jurisdicción fue el caso de Hugh Nelson, sentenciado el 28 de octubre de 2024 a 18 años de prisión por crear y comercializar, mediante inteligencia artificial, imágenes de abuso sexual infantil a partir de fotografías reales de menores que le eran encargadas por terceros; su condena se sustentó en los tipos penales de pornografía infantil ya vigentes en el Reino Unido, y no en la nueva figura de creación de deepfakes sexuales, que todavía no había entrado en vigor (Rosales Murillo, 2024).

Corea del Sur: Ante una situación calificada como "emergencia nacional" por el aumento masivo de crímenes sexuales digitales en plataformas como Telegram, el parlamento aprobó en septiembre de 2024 una reforma a la Ley de Prevención de la Violencia Sexual. La nueva normativa sanciona con hasta cinco años de prisión o multas de 50 millones de wones a quien sintetice o distribuya pornografía deepfake. No obstante, si el delito se comete utilizando redes de información, la pena privativa de libertad se eleva a siete años, incrementándose hasta en un 50% en casos de reincidencia. Además, la ley ahora permite castigar no solo a los creadores, sino también a quienes posean o visualicen conscientemente dicho material (Reuters, 2024).

China: Estableció en diciembre de 2022 el Reglamento para la Gestión Profunda de Servicios de Información en Internet, convirtiéndose en uno de los marcos técnicos más estrictos del mundo (Reglamento Gestión Profunda China, 2022). Esta normativa impone la obligación de obtener el consentimiento previo de los sujetos antes de realizar cualquier

tratamiento sintético de su imagen o voz. De igual manera, exige que todo contenido generado por IA incluya marcas de agua digitales o etiquetas claras que identifiquen su naturaleza artificial para evitar el engaño público. Los proveedores de servicios deben autenticar la identidad de sus usuarios y someter sus algoritmos a evaluaciones de seguridad, bajo riesgo de enfrentar sanciones penales y civiles (Valverde-Ramos et al., 2024).

3.4.5. Protección de imagen, voz, datos personales e identidad.

La protección de la imagen, la voz, los datos personales y la identidad constituye el núcleo de la defensa del patrimonio moral del individuo frente a la irrupción de las tecnologías de manipulación sintética (Ponce González et al., 2025).

En la era de la Inteligencia Artificial, estos atributos han dejado de ser meras representaciones físicas para convertirse en datos biométricos y activos digitales de alta sensibilidad, cuya vulneración mediante deepfakes exige un marco de tutela jurídica reforzado y específico (Corte Constitucional del Ecuador, 2021).

Protección de la imagen y la voz: El derecho a la propia imagen faculta a la persona a decidir qué rasgos físicos pueden tener difusión pública e impedir su captación o reproducción no autorizada.

La jurisprudencia ecuatoriana, a través de la Sentencia No. 2064-14-EP/21, ha determinado que el rostro es un dato personal protegido, ya que es una representación directa de la identidad (Corte Constitucional del Ecuador, 2021). Por su parte, la voz (clonada mediante deepvoice) se considera una extensión inescindible de la personalidad; tendencias internacionales como la Ley ELVIS de Tennessee ya prohíben expresamente imitar la voz de los artistas sin permiso para garantizar su seguridad comercial y personal (Ponce González et al., 2025).

Protección de datos personales y autodeterminación informativa: El derecho a la protección de datos garantiza el control sobre la información que identifica o hace identificable a una persona, incluyendo su imagen y audio (Asamblea Nacional del Ecuador, 2021).

Los deepfakes violentan el principio de autodeterminación informativa, que es la facultad del titular para decidir cuándo y bajo qué límites se puede revelar información de su vida privada (Corte Constitucional del Ecuador, 2021). Al alimentar algoritmos con fotos extraídas de redes sociales sin consentimiento para crear material falso, se despoja al individuo del poder de disposición sobre sus propios datos digitales (Endara-Chamorro et al., 2024).

Derecho a la identidad personal y digital: La identidad comprende no solo el nombre, sino también los rasgos físicos, gestos y la reputación que el individuo proyecta en la sociedad (Endara-Chamorro et al., 2024).

El uso de IA para la suplantación de identidad mediante la sustitución del rostro (deepface) distorsiona la personalidad moral de las víctimas ante terceros, generando un "engaño mutado" que es prácticamente indistinguible de la realidad (Rosales Murillo, 2024).

La creación de identidades sintéticas o influenciadores virtuales sin consentimiento de los grupos representados “como el caso de Nina Tapuy en Ecuador” ha sido denunciada como una forma de robo de identidad y afectación a la identidad colectiva (Ponce González et al., 2025).

3.4.6. Lecciones útiles Y desafíos para Ecuador.

El análisis del derecho comparado ofrece lecciones valiosas para el legislador ecuatoriano, destacando la conveniencia de evaluar el paso de normas generales de protección de datos, como la LOPDP, hacia una tipificación penal específica en el Código Orgánico Integral Penal (COIP). Esta evaluación se justifica porque el artículo 178 (violación a la intimidad) no contempla de forma expresa el acto de la “creación algorítmica” desde cero respecto de personas adultas, y porque los artículos 103 y 104 (relacionados con la pornografía

infantil) deben analizarse en conjunto, ya que el 103 incluye los verbos producir y editar y las representaciones simuladas (Rosales Murillo, 2024). Por eso, una eventual propuesta debería incluir el verbo rector “crear” para atender la falta de cobertura en la fase de producción sintética, reconociendo que el daño nace desde la misma generación del material sin consentimiento (Ministry of Justice, 2024). Esta propuesta no parte de un vacío legislativo absoluto en el ámbito nacional: durante 2026 la Asamblea Nacional ha tramitado o recibido diversas iniciativas sobre violencia sexual digital y contenido sintético, entre ellas el proyecto de reforma al COIP y a la Ley Orgánica Integral para Prevenir y Erradicar la Violencia contra las Mujeres presentado por el asambleísta Mario Zambrano, orientado a tipificar la fabricación de contenido sexual sintético mediante IA; el proyecto presentado por la asambleísta Ledy Zúñiga para incorporar expresamente la violencia sexual digital mediante contenido sintético; el proyecto de ley orgánica que regula el uso de simulaciones de imágenes, voces, audio y video generadas por inteligencia artificial, presentado el 25 de marzo de 2026; y el proyecto remitido por la Función Ejecutiva a la Asamblea Nacional el 7 de septiembre de 2026 para prevenir y erradicar la violencia digital contra las mujeres, que contempla la difusión de contenido íntimo generado mediante IA. La existencia de estas iniciativas, aún en trámite y sin sanción, no elimina la novedad de la propuesta desarrollada en esta tesis, pero exige diferenciarla expresamente de ellas para sostener con rigor la afirmación de una falta de cobertura normativa vigente.

La protección debe abarcar no solo el rostro, sino la identidad digital completa y la autodeterminación informativa (Corte Constitucional del Ecuador, 2021). No obstante, se debe tener especial cuidado en la adaptación de estos modelos. La propuesta para Ecuador debe respetar estrictamente el principio de legalidad y mínima intervención penal para evitar caer en un "populismo punitivo" que coarte el desarrollo tecnológico legítimo (Saquicela Rodas & Chica Mejía, 2023).

La regulación ecuatoriana debe fundamentarse en la dignidad humana como valor supremo, asegurando que cualquier sanción penal sea proporcional a la lesividad efectiva del acto sintético (Chávez Zambrano & San Andrés Molina, 2025).

En conclusión, la protección efectiva de estos derechos personalísimos en el entorno digital no debería quedar limitada a normas civiles o administrativas generales; justifica evaluar una respuesta penal que comprenda que la imagen y la voz son la manifestación material de la dignidad humana y que su recreación artificial maliciosa constituye una agresión directa a la esencia misma del ser (Chávez Zambrano & San Andrés Molina, 2025).

3.5. Bien jurídico protegido en el tipo penal autónomo.

La identificación del bien jurídico protegido constituye un requisito esencial e ineludible para dar cumplimiento al principio de lesividad, el cual garantiza que la intervención del Derecho Penal no sea arbitraria y se dirija únicamente a la protección de valores con una clara relevancia constitucional. En un Estado constitucional de derechos y justicia como el ecuatoriano, el ius puniendi ya no es absoluto, sino que encuentra sus límites materiales en la protección efectiva de la dignidad humana, matriz de la cual emanan todos los demás derechos fundamentales. Por tanto, el legislador carece de una discrecionalidad absoluta para definir delitos, debiendo justificar cada tipo penal en la ofensa efectiva a un interés jurídico que el Estado considera digno de protección penal debido a su especial gravedad (Saquicela Rodas & Chica Mejía, 2023).

En el caso de los deepfakes, la doctrina reconoce que el legislador no se enfrenta a la protección de un derecho aislado, sino a una conducta compleja y pluriofensiva que ataca simultáneamente diversos bienes jurídicos de rango superior. Esta tecnología, basada en redes generativas antagónicas (GAN), permite la creación de representaciones sintéticas hiperrealistas que son prácticamente indistinguibles de la realidad fáctica, lo que atenta contra la esencia misma de la identidad del individuo en el entorno digital. Al utilizar el rostro o la

voz de una persona sin su consentimiento para insertarla en escenarios falsos o denigrantes, se produce una "expropiación" de la imagen que despoja al titular de su facultad de disposición sobre su propia existencia virtual (Taylor Villafuerte, 2024).

3.5.1. La dignidad humana como eje principal.

La dignidad humana constituye la matriz ontológica y el valor supremo de todo el sistema jurídico, actuando como el eje central de protección frente a la manipulación algorítmica. Esta noción plantea un elemento constitutivo del ser humano, mínimo, propio e inalienable, que todo ordenamiento está compelido a asegurar como base de la existencia y coexistencia de la persona (Villalba Fiallos, 2017). Por ser una calidad inherente que no admite sustitución ni equivalencia, la dignidad exige que la persona sea tratada siempre como un "fin en sí misma", prohibiendo que las realidades que le circundan, como el Estado o la tecnología, la utilicen meramente como un instrumento o medio para fines ajenos (García Toma, 2018).

El uso de Inteligencia Artificial para recrear artificialmente a una persona sin su consentimiento la cosifica, convirtiéndola en un mero objeto de consumo o burla, lo cual fractura el principio ético-jurídico fundamental de la humanidad (Rosales Murillo, 2024). En el contexto de los deepfakes sexuales, esta conducta supone una agresión directa a la integridad moral y psicológica, ya que se instrumentaliza la imagen del individuo para insertarla en escenarios denigrantes que nunca ocurrieron, vulnerando radicalmente su derecho a la autodeterminación y al libre desarrollo de la personalidad (Ruiz, 2025).

. Esta deshumanización digital es especialmente lesiva debido a que la sociedad otorga un valor de "verdad" a la evidencia visual, lo que expone a la víctima a un escarnio público de consecuencias irreversibles para su honor y su proyecto de vida (Taylor Villafuerte, 2024). Cualquier tipo de aplicación tecnológica que condicione la voluntad humana o deshumanice al individuo al reducirlo a patrones estadísticos constituye una ofensa directa a su realidad personal más íntima (Chávez Zambrano & San Andrés Molina, 2025).

3.5.2. Intimidad, honor y buen nombre.

El derecho a la intimidad protege la esfera reservada de la vida del individuo, permitiéndole excluir del conocimiento público aspectos personales que solo le conciernen a él o a su familia (Salvador Quiñonez, 2022).

Los deepfakes vulneran esta esfera al exponer representaciones sintéticas de la víctima en contextos privados. Paralelamente, se lesiona el honor y el buen nombre, ya que la "falsedad profunda" destruye la reputación social (dimensión externa) y la autovaloración (dimensión interna) del sujeto, sometiéndolo a un vejamen público difícil de reparar debido a la hiperrealidad del material (Valverde-Ramos et al., 2024).

3.5.3. Propia imagen y protección de datos personales.

El derecho a la propia imagen es un derecho constitucional autónomo, inalienable e imprescriptible que emana directamente de la dignidad humana y la libertad individual. Este derecho faculta a la persona a decidir de manera privativa cuándo, cómo y por quién sus rasgos físicos son captados, reproducidos o difundidos, impidiendo cualquier forma de explotación o reproducción no autorizada por terceros (Corte Constitucional del Ecuador, 2021). Su tutela posee una naturaleza bifacética: una dimensión positiva, que permite al titular controlar su representación visual, y una dimensión negativa, que le otorga la potestad de prohibir que su imagen sea utilizada por otros sin su consentimiento previo (Rosales Murillo, 2024).

En la era de la Inteligencia Artificial (IA), la protección de este derecho adquiere una complejidad técnica superior, ya que el rostro y la voz no son solo atributos físicos, sino que se consideran datos biométricos sensibles y representaciones directas de la identidad del individuo. Según lo establecido por la Corte Constitucional en la Sentencia No. 2064-14-EP/21, el tratamiento de estos datos “incluida la fotografía” requiere de una protección reforzada debido a que su uso indebido puede acarrear situaciones de discriminación o vulneraciones graves a la integridad moral (Palacios Zúñiga, 2025).

Por tanto, la creación de un deepfake violenta de forma flagrante la autodeterminación informativa, entendida como el derecho activo del ciudadano a controlar el flujo de información personal que circula en el ciberespacio y a decidir cómo quiere ser percibido ante la sociedad. Al utilizar algoritmos para recrear o alterar la apariencia física de una persona sin su autorización, se produce una "expropiación" de la identidad digital, privando al titular de la facultad de gestionar libremente su presencia virtual (Endara-Chamorro et al., 2024).

3.5.4. Libertad e indemnidad sexual.

Cuando el contenido de un deepfake es de naturaleza pornográfica, el bien jurídico afectado se desplaza desde la mera privacidad hacia la esfera de la sexualidad, variando su configuración técnica según la edad y capacidad de la víctima. En estos supuestos, la tecnología no solo manipula datos, sino que instrumentaliza la identidad íntima para crear una existencia sexual forzada en el ciberespacio (Ruiz, 2025).

Libertad sexual (Adultos): En el caso de víctimas adultas, el bien jurídico lesionado es la libertad sexual, entendida como la facultad de autodeterminación que permite a cada individuo decidir voluntariamente sobre su comportamiento, preferencias y la exposición de su vida erótica. La creación de pornografía sintética sin consentimiento vulnera esta autonomía al proyectar a la persona en actos eróticos ficticios, lo que constituye una forma de cosificación donde el cuerpo (o su representación hiperrealista) es tratado como un objeto para la satisfacción de deseos ajenos (Rosales Murillo, 2024). La lesividad radica en que la víctima pierde el poder de elegir "con quién, cómo y en qué momento" se manifiesta su sexualidad, siendo sometida a una agresión audiovisual que afecta su integridad moral y su proyecto de vida (Taylor Villafuerte, 2024).

Indemnidad sexual y desarrollo integral (Niñas, niños y adolescentes): Cuando las víctimas son menores de edad, el enfoque jurídico cambia de la libertad hacia la indemnidad sexual (Rosales Murillo, 2024). Este bien jurídico se fundamenta en la incapacidad legal y

madurez mental limitada de los menores para consentir de forma plena y consciente en actos de naturaleza sexual. El Estado tiene el deber de garantizar su libre y normal desarrollo de la personalidad, protegiéndolos de ser utilizados como instrumentos o "bienes muebles sexuales" (Bowen, 2017).

3.5.5. Carácter pluriofensivo del delito.

Debido a la magnitud del daño que puede ocasionar una sola manipulación digital, el delito de deepfake tiene un carácter eminentemente pluriofensivo. A diferencia de los delitos tradicionales que suelen lesionar un único interés jurídico, una sola acción sintética “como la creación y difusión de un video sexual falso” lesiona simultáneamente una pluralidad de bienes protegidos de rango constitucional. Esta lesividad extrema se deriva del hiperrealismo de la Inteligencia Artificial, que, al ser prácticamente indistinguible de la realidad fáctica, genera un error en la percepción del público que asocia a la víctima con actos denigrantes que nunca ocurrieron (Valverde-Ramos et al., 2024).

En primer lugar, esta conducta violenta el derecho a la intimidad y a la propia imagen, al "expropiar" los rasgos biométricos y la apariencia física del individuo para proyectarlo en escenarios no consentidos, fracturando su expectativa razonable de privacidad (Corte Constitucional del Ecuador, 2021). Paralelamente, se produce una agresión frontal al honor y al buen nombre, ya que la difusión de estas "mentiras profundas" destruye la reputación social (dimensión externa) y la autovaloración (dimensión interna) de la persona, sometiéndola a un vejamen público de consecuencias permanentes (Taylor Villafuerte, 2024).

Asimismo, se configura una transgresión a la protección de datos personales y a la autodeterminación informativa, privando al titular de la facultad de gestionar libremente su identidad digital y el flujo de su información en el ciberespacio. El impacto trasciende la esfera jurídica para alcanzar la integridad psicológica de la víctima; el escarnio masivo y la

humillación suelen derivar en cuadros graves de ansiedad, depresión, estrés postraumático y, en casos extremos, un riesgo latente de conductas autolesivas o suicidio (Ruiz, 2025).

Esta convergencia de afectaciones justifica que el tipo penal propuesto no se limite a proteger un solo derecho, sino que reconozca la gravedad integral del ataque a la dignidad humana en la dimensión virtual. La dignidad, como matriz ontológica del sistema jurídico, exige que el ser humano sea tratado siempre como un "fin en sí mismo", prohibiendo su cosificación e instrumentalización algorítmica por parte de terceros (García Toma, 2018).

En conclusión, la tipificación de este delito busca salvaguardar un abanico de derechos personalísimos que son avasallados por el uso malicioso de la tecnología. Al establecer una sanción proporcional a esta pluriofensividad, el Estado garantiza que la identidad y la integridad de los ciudadanos no sean mercantilizadas ni utilizadas como herramientas de acoso, evitando que la falta de cobertura en determinadas modalidades consolide un escenario de impunidad digital (Rosales Murillo, 2024).

3.6. Sujetos del tipo penal.

La configuración de un tipo penal autónomo sobre los deepfakes requiere una identificación precisa de los intervinientes en la cadena delictiva. Debido a la naturaleza digital de la conducta, la ejecución del ilícito suele involucrar una multiplicidad de actores, desde el programador del algoritmo hasta el usuario final que viraliza el contenido (Abbas & Taeihagh, 2024). En este sentido, la doctrina penal enfrenta un desafío particular al intentar delimitar el sujeto activo de este tipo de conductas, pues la cadena de producción y difusión de un deepfake rara vez responde a un esquema de autoría única.

3.6.1. Sujeto activo: El autor detrás de la tecnología.

El sujeto activo en este delito es, por regla general, un sujeto activo común, es decir, cualquier persona natural que realice la conducta típica. Específicamente, es aquel individuo que crea, manipula, altera o genera contenido sintético con el dolo de causar un perjuicio. Es

fundamental precisar que la responsabilidad penal recae exclusivamente en la persona humana que utiliza la herramienta, ya que la Inteligencia Artificial carece de voluntad propia y no puede ser sujeto de imputación penal (Ávila Zea & Zamora Vázquez, 2024).

En conclusión, la autoría penal se desplaza de la herramienta al usuario, lo que exige diferenciar los grados de participación de los distintos sujetos que intervienen en el ciclo de vida del deepfake, como se detalla a continuación.

3.6.2. Pluralidad de intervinientes: Creadores, difusores y comercializadores.

El ciclo de vida de un deepfake malicioso no suele ser un acto solitario, sino que involucra una cadena de intervinientes con distintos grados de responsabilidad y dominio sobre el hecho digital. Debido a que la Inteligencia Artificial no posee voluntad propia ni personalidad jurídica, la responsabilidad penal debe recaer exclusivamente en los sujetos humanos que instrumentalizan la tecnología para lesionar bienes jurídicos (Ruiz, 2025).

El Creador: El arquitecto del contenido sintético.

Es el sujeto que utiliza algoritmos avanzados de aprendizaje profundo, principalmente las Redes Generativas Antagónicas (GANs) o los Autoencoders Variacionales (VAEs), para sintetizar la imagen o el audio falso (Ugwuoke & Sanfilippo, 2025). En el proceso de las GANs, una red "generadora" crea muestras que intentan engañar a una red "discriminadora", logrando un hiperrealismo que hace al contenido indistinguible de la realidad (Torres, 2023). Actualmente, la "democratización" tecnológica permite que este rol no requiera pericia técnica avanzada, ya que existen aplicaciones de "un solo clic" que facilitan la creación masiva de pornografía sintética. Resulta crítico señalar que, respecto de personas adultas, el COIP vigente presenta una falta de cobertura en este eslabón; en el caso de menores, el artículo 104 omite sancionar específicamente la acción de “crear” el material y se centra más en su tenencia o venta, pero debe leerse junto con el artículo 103, que sí incluye los verbos producir y editar (Ponce González et al., 2025).

El Difusor: El motor de la viralización.

Es quien, conociendo la falsedad del material o actuando con un temerario desprecio hacia la verdad, pone el contenido en circulación en el ciberespacio (Ruiz, 2025). Su rol es fundamental para la consumación del daño, ya que aprovecha la reticularidad de las redes sociales y servicios de mensajería (Telegram, WhatsApp, X) para generar "enjambres digitales" que masifican el escarnio público. Al difundir el deepfake, el sujeto despoja a la víctima de su autodeterminación informativa, provocando que pierda el control sobre su propia imagen e identidad en la red (Cuadra, 2019).

El Comercializador: El beneficiario económico.

Este actor busca lucrarse mediante la explotación del contenido sintético, a menudo operando en foros especializados, chats de Discord o plataformas de suscripción tipo OnlyFans. La comercialización de deepfakes sexuales por comisión se ha convertido en una industria lucrativa donde se venden representaciones denigrantes de celebridades o personas comunes, lo que evidencia una forma extrema de cosificación y violencia de género digital (Taylor Villafuerte, 2024).

En este escenario, el Derecho Penal ecuatoriano debe aplicar las reglas de autoría y complicidad previstas en los artículos 41 al 43 del COIP. Responderán como autores directos o mediatos quienes tengan el dominio del hecho, ya sea programando deliberadamente la IA para delinquir o ejecutando la difusión masiva (Ruiz, 2025).

Es vital diferenciar la autoría penal de la simple intermediación tecnológica. Los administradores de plataformas o desarrolladores de software de código abierto (open source) no deberían ser sancionados si su herramienta tiene fines lícitos y no han participado en el dolo específico de dañar. No obstante, conforme al artículo 513 del COIP, los responsables de medios podrían ser procesados si no revelan el nombre del autor malicioso (Asamblea Nacional del Ecuador, 2014). Esta distinción de roles es la que garantiza que la sanción penal sea

proporcional y justa, evitando el castigo a quienes realizan una labor técnica neutra y dirigiendo el reproche estatal hacia quienes manifiestan un claro animus injuriandi o intención de vejar (Saquicela Rodas & Chica Mejía, 2023).

3.6.3. Diferenciación entre autoría penal y simple intermediación tecnológica.

Un punto crítico es la responsabilidad de los administradores de plataformas e intermediarios. La legislación ecuatoriana, en el artículo 513 del COIP, establece que los responsables de medios de comunicación pueden responder por la infracción si no revelan el nombre del autor (Asamblea Nacional del Ecuador, 2014). No obstante, en el ámbito digital, las plataformas suelen ampararse en sus políticas de privacidad para deslindarse de responsabilidad (Ponce González et al., 2025).

La tendencia internacional sugiere que solo debe existir responsabilidad penal para la plataforma cuando, tras recibir una notificación efectiva de la ilegalidad del contenido (como en el caso de los deepfakes sexuales), esta no actúe con celeridad para retirarlo (European Commission, 2020).

Una vez definida la responsabilidad de quienes ejecutan o facilitan el acto, es necesario identificar a la persona sobre la cual recae el daño.

3.6.4. Sujeto pasivo: La víctima de la simulación.

El sujeto pasivo es la persona natural cuya imagen, voz, rasgos biométricos o identidad han sido manipulados sin su consentimiento. Aunque las personas jurídicas pueden sufrir daños reputacionales por deepfakes, el tipo penal propuesto se centra en la protección de la dignidad humana, valor inherente únicamente a los seres humanos (Chávez Zambrano & San Andrés Molina, 2025).

La víctima es quien sufre la "expropiación" de su apariencia física para ser colocada en escenarios denigrantes o falsos. La afectación al sujeto pasivo no es uniforme, manifestando una gravedad extrema cuando las víctimas pertenecen a grupos de atención prioritaria.

3.6.5. Protección reforzada para niños, niñas, adolescentes y grupos vulnerables.

El sistema penal ecuatoriano debe contemplar una protección reforzada mediante la configuración de un sujeto pasivo calificado cuando el uso malicioso de deepfakes afecte a menores de edad o personas en situación de especial vulnerabilidad. Este imperativo nace del mandato contenido en el artículo 66 de la Constitución de la República, que ordena la adopción de medidas específicas para prevenir y sancionar toda forma de violencia, esclavitud o explotación sexual contra grupos de atención prioritaria, incluyendo a mujeres, niñas, niños, adolescentes y personas con discapacidad (Asamblea Nacional del Ecuador, 2008).

En esta línea, el Código Orgánico Integral Penal (COIP) ya reconoce esta categoría de protección en delitos de naturaleza similar, exigiendo incluso la intervención de fiscalías especializadas para garantizar que el proceso penal brinde una tutela superior a quienes, por su condición física, mental o económica, se encuentran en circunstancias de mayor indefensión (Monar Puente et al., 2024).

En estos casos, la lesividad de la conducta trasciende el ámbito de la intimidad personal para alcanzar bienes jurídicos superiores como la indemnidad sexual y el desarrollo integral de la personalidad. La doctrina penal sostiene que, tratándose de menores, la protección no debe centrarse únicamente en la libertad sexual “puesto que ellos carecen de la madurez mental para consentir plenamente”, sino en preservar su integridad física y psicológica frente a cualquier forma de explotación (Taylor Villafuerte, 2024). El uso de IA para simular la participación de infantes en actos sexuales constituye un peligro abstracto a la dignidad de la infancia, pues, aunque la escena sea virtual, el estigma generado por la circulación del material fractura de forma permanente el proyecto de vida del menor (Rodríguez, 2014).

Casos recientes, como el ocurrido en un colegio de Quito en 2023 donde se generaron y difundieron más de 700 contenidos pornográficos sintéticos afectando a 24 estudiantes, o los incidentes masivos en Almendralejo (España), demuestran que los menores son víctimas frecuentes y deliberadas de estas tecnologías (Ruiz, 2025). Las consecuencias de estos ataques

son devastadoras e incluyen daños psicológicos permanentes, ansiedad severa, depresión y la ruptura del entorno de convivencia, lo que a menudo obliga a las víctimas al aislamiento o al abandono de sus estudios. El riesgo latente de conductas autolesivas o suicidio, evidenciado en tragedias internacionales derivadas del ciberacoso, ratifica que el Estado no puede tratar estas agresiones como incidentes aislados, sino como conductas de alto reproche social que requieren una sanción ejemplar (Rosales Murillo, 2024).

Por consiguiente, la ley debe prever agravantes específicas cuando la víctima carezca de capacidad para comprender el significado del hecho o resistirlo, o cuando el agresor pertenezca al entorno íntimo de la familia o tenga un deber de custodia. El COIP ya contempla penas agravadas en el artículo 103 cuando el infractor es docente, progenitor o profesional que abusa de su posición (Rosales Murillo, 2024). En el ámbito de los deepfakes, resulta conveniente sancionar expresamente la acción de crear material que simule la participación de personas en situación de vulnerabilidad, sin perjuicio de que, tratándose de menores de edad, deban examinarse primero los artículos 103 y 104 del COIP, que ya incluyen los verbos producir y editar y las representaciones simuladas (Ponce González et al., 2025).

3.7. Conducta típica y verbos rectores.

La configuración de un tipo penal autónomo sobre el uso malicioso de deepfakes exige una precisión técnica extrema en la descripción de la conducta típica, ya que el núcleo del injusto radica en la capacidad de la tecnología para fabricar una realidad inexistente que sustituye o altera la identidad del sujeto pasivo (Ponce González et al., 2025). Etimológicamente, este fenómeno se fundamenta en el deep learning o aprendizaje profundo, que permite a algoritmos de Inteligencia Artificial procesar patrones masivos para generar archivos de audio, video o imagen que aparentan ser auténticos, pero son meras representaciones sintéticas. A diferencia de las formas tradicionales de manipulación digital, el hiperrealismo alcanzado mediante las Redes Generativas Antagónicas (GAN) hace que este

contenido sea prácticamente indistinguible de la realidad fáctica, lo que otorga a la conducta una lesividad superior al inducir a error tanto a personas como a sistemas de verificación (Ponce González et al., 2025).

A diferencia de los delitos tradicionales sancionados en el Código Orgánico Integral Penal (COIP), como la violación a la intimidad (Art. 178), que castigan la captura o difusión de imágenes reales obtenidas sin consentimiento, el legislador debe aquí abarcar un espectro más amplio para evitar lagunas de punibilidad respecto de personas adultas (Endara-Chamorro et al., 2024). Cuando la persona representada sea una niña, niño o adolescente, corresponde en cambio examinar en primer lugar la aplicación de los artículos 103 y 104 del COIP, que sí contemplan representaciones simuladas y los verbos producir y editar; la propuesta que se desarrolla en este capítulo se dirige, por tanto, de manera principal a las representaciones sintéticas de personas adultas, para las cuales los tipos vigentes analizados no prevén de forma expresa la acción de "crear" este tipo de material (el artículo 104, por ejemplo, se limita a verbos como publicar, comprar, poseer o vender). Esta omisión puede permitir que el creador del material algorítmico, quien efectivamente materializa el daño mediante la síntesis de la imagen ajena, quede sin sanción si su conducta no se ajusta estrictamente a los verbos rectores de transferencia o comercialización (Rosales Murillo, 2024).

3.7.1. Verbos rectores de la fase de creación y modificación.

El primer grupo de verbos rectores debe capturar la génesis del contenido falso asistido por Inteligencia Artificial, identificando el momento exacto en que se produce la lesión al derecho a la propia imagen y a la identidad digital (Taylor Villafuerte, 2024). Varios de los tipos penales vigentes en el Ecuador, en particular los aplicables a personas adultas, no incluyen la acción de "crear" desde cero mediante algoritmos y se limitan sobre todo a sancionar la difusión o comercialización de material preexistente; el artículo 103 es una excepción parcial, pues contempla los verbos producir y editar respecto de niñas, niños y adolescentes. Esta falta

de cobertura puede permitir que quien elabora el material algorítmico, y materializa el daño mediante la síntesis de la imagen ajena, quede sin sanción si su conducta no se ajusta a los verbos tradicionales de transferencia o porte (Ponce González et al., 2025).

Crear y Generar: Estos verbos son fundamentales para sancionar el acto de síntesis algorítmica donde, a través de redes neuronales —específicamente las Redes Generativas Antagónicas (GANs)—, se produce una imagen, audio o video que no tiene una base real fáctica. En este proceso, una red "generadora" crea muestras que intentan engañar a una red "discriminadora", logrando un hiperrealismo que hace al contenido prácticamente indistinguible de la realidad para el ojo humano. La inclusión de estos términos permitiría castigar la creación de "pornografía virtual", donde se simula la participación de una persona en actos sexuales mediante una representación íntegramente artificial, protegiendo así la dignidad humana frente a la cosificación digital (Bowen, 2017).

Alterar, Manipular y Editar: Estos verbos se refieren a la transformación de material audiovisual preexistente y auténtico de una persona para insertarlo en contextos ajenos a su voluntad. Aquí se encuadran técnicas como el faceswap (intercambio de rostros), donde se superpone la cara de la víctima sobre el cuerpo de un tercero, y el deepvoice o clonación de voz, que permite generar secuencias de habla falsas con una verosimilitud extrema. Estas acciones constituyen una "expropiación algorítmica" de la identidad, ya que utilizan datos biométricos reales —como los gestos y las inflexiones de voz— para crear una existencia virtual forzada que lesiona el honor y la reputación de la víctima (Taylor Villafuerte, 2024).

3.7.2. Verbos rectores de la fase de difusión y explotación.

Una vez generado el material, la lesividad de la conducta se expande de forma geométrica mediante su puesta en circulación, lo que exige la inclusión de verbos rectores que ya asoman en otros delitos informáticos del COIP, pero adaptados rigurosamente a la realidad sintética. La ciberdelincuencia contemporánea aprovecha la reticularidad de la red para asegurar que,

una vez que un contenido es cargado, su eliminación total sea prácticamente imposible debido al carácter viral y transnacional de la internet. Este fenómeno de "infoapocalipsis" implica que la mentira profunda se propaga a una velocidad que sobrepasa cualquier mecanismo de control tradicional, obligando al legislador a sancionar cada eslabón de la cadena de transmisión (Salvador Quiñonez, 2022).

Reproducir, Difundir y Publicar: Estas acciones materializan de forma definitiva el daño a la intimidad, al honor y al buen nombre de la víctima. Al masificar el contenido en redes sociales (como X, Facebook o Instagram), plataformas de mensajería (WhatsApp, Telegram) o foros de internet, se crean "enjambres digitales" que multiplican el escarnio público de forma incontrolable. En este punto, la víctima pierde por completo su autodeterminación informativa y el control sobre su propia existencia digital, quedando expuesta a una humillación permanente que afecta su salud mental y su proyecto de vida (Corte Constitucional del Ecuador, 2021). La ley debe capturar la acción de quien "revela o divulga" material que, aunque sea sintético, es percibido como real por la sociedad, generando un perjuicio moral irreversible (Ruiz, 2025).

Comercializar: Este verbo busca sancionar el ánimo de lucro derivado de la venta, oferta o intercambio de deepfakes, una práctica criminal que ha proliferado exponencialmente en sitios de suscripción mensual (como OnlyFans), chats de Discord y foros especializados en explotación sexual (Andino Veloz, 2025). La industria de la pornografía sintética por comisión representa una forma extrema de cosificación, donde la identidad de las personas “especialmente mujeres y niñas” es tratada como una mercancía digital para la satisfacción de deseos ajenos. La comercialización no solo lesiona la integridad sexual, sino que incentiva la creación masiva de nuevos ultrafalsos al convertirlos en un activo rentable dentro del mercado negro de la informática (Salvador Quiñonez, 2022).

Almacenar y poseer: Estos verbos no se incluyen entre los verbos rectores del tipo propuesto. Sancionar la mera tenencia o el almacenamiento anticiparía la punibilidad a una etapa preparatoria que puede ser equívoca, y el riesgo de viralización por sí solo no demuestra una lesividad autónoma suficiente. Por esa razón, la propuesta se concentra en la creación o manipulación maliciosa del contenido y en su difusión consciente, sin modificar la regulación vigente de los artículos 103 y 104 del COIP.

3.7.3. El elemento tecnológico y la ausencia de consentimiento.

La configuración de la conducta típica en el uso malicioso de deepfakes debe estar inexorablemente ligada al empleo de inteligencia artificial, algoritmos de aprendizaje profundo (deep learning) o tecnologías análogas destinadas al procesamiento de datos biométricos (Taylor Villafuerte, 2024). El núcleo tecnológico de este fenómeno reside en las Redes Generativas Antagónicas (GAN), sistemas compuestos por dos redes neuronales —un generador y un discriminador— que compiten entre sí para perfeccionar imágenes, audios o videos hasta lograr un hiperrealismo que engaña los sentidos humanos y los sistemas de detección (García, 2021). Esta capacidad de síntesis algorítmica permite realizar el intercambio de rostros (faceswap) o la clonación de voz (deepvoice) con tal precisión que la representación resultante es percibida como auténtica, lo cual otorga a la conducta una lesividad superior frente a las técnicas de edición digital tradicionales (Ponce González et al., 2025).

Asimismo, el elemento normativo determinante de la infracción es la ausencia de consentimiento de la persona afectada, cuya identidad es instrumentalizada mediante la tecnología (Corte Constitucional del Ecuador, 2021). De acuerdo con la Ley Orgánica de Protección de Datos Personales (LOPDP) y la jurisprudencia constitucional, el tratamiento de datos personales, especialmente de datos sensibles como la imagen del rostro o la voz, requiere de una base legal legítima donde la autorización del titular es la pieza fundamental. Sin esta

voluntad expresa, el uso de la identidad digital de un individuo para crear realidades sintéticas se convierte en un acto de violencia digital y cosificación algorítmica (Rosales Murillo, 2024).

Según la Sentencia No. 2064-14-EP/21 de la Corte Constitucional del Ecuador, el consentimiento para el tratamiento de datos debe cumplir con cuatro requisitos ineludibles: debe ser libre, específico, informado e inequívoco (Corte Constitucional del Ecuador, 2021). Esto implica que la manifestación de voluntad no debe estar viciada por coerción, debe detallar claramente la finalidad del uso de la imagen y no puede ser ambigua (Palacios Zúñiga, 2025). Por tanto, el hecho de que una imagen sea pública en redes sociales o esté accesible en internet no autoriza a un tercero a realizar un tratamiento sintético o manipulado de la misma, ni faculta su almacenamiento o difusión en contextos denigrantes (Corte Constitucional del Ecuador, 2021).

La Corte Constitucional ha enfatizado que el consentimiento es de interpretación restrictiva y no supone una renuncia indefinida a los derechos personalísimos (Corte Constitucional del Ecuador, 2021). Compartir voluntariamente una fotografía en una plataforma digital o en una conversación privada únicamente habilita su acceso y visualización para ese fin específico, pero jamás implica una autorización para la "expropiación de la imagen" mediante procesos de IA que alteren la identidad del titular (Corte Constitucional del Ecuador, 2021). En consecuencia, cualquier manipulación algorítmica no autorizada fractura la "expectativa razonable de privacidad" del ciudadano, justificando la intervención del aparato punitivo del Estado para proteger la dignidad humana en el entorno virtual (Chávez Zambrano & San Andrés Molina, 2025).

3.7.4. El elemento subjetivo: Finalidades específicas.

Para evitar la criminalización de usos legítimos, como la parodia, la sátira o las narrativas lúdicas realizadas por actores legítimos, el tipo penal debe exigir un especial elemento subjetivo o una finalidad específica por parte del autor (García, 2021). Debido a que

las herramientas de IA son de libre acceso y pueden ser utilizadas tanto para fines artísticos como malintencionados, la sanción penal no debe recaer sobre la tecnología en sí, sino sobre el dolo manifiesto de quien decide instrumentalizarla para lesionar bienes jurídicos ajenos.

Bajo esta premisa, la conducta debe realizarse con una de las siguientes finalidades:

Sexual o pornográfica: Se dirige a los casos de violencia sexual digital o Abuso Sexual Basado en Imágenes (ASBI), donde el autor busca la cosificación del cuerpo ajeno mediante la generación de pornografía sintética. En estos supuestos, el desvalor de la acción radica en tratar a la víctima como un mero activo comercial o un objeto de gratificación sexual forzada, vulnerando radicalmente su autodeterminación sexual (Rosales Murillo, 2024).

Maliciosa, difamatoria o vejatoria: Está orientada a menoscabar el honor, la buena fama o la integridad moral de la víctima mediante la burla, el escarnio o el descrédito público (Taylor Villafuerte, 2024). El uso de ultrafalsos para simular declaraciones o actos humillantes constituye una agresión frontal a la dignidad, ya que la sociedad tiende a otorgar un valor de verdad a la evidencia visual, provocando un daño reputacional a menudo irreversible. En este ámbito, se identifica un claro animus injuriandi o intención de vejar que diferencia el delito de la crítica legítima (Ruiz, 2025).

Fraudulenta: Cuando el deepfake busca inducir a error a terceros para obtener un beneficio económico indebido o para la suplantación de identidad. Esta finalidad abarca desde estafas informáticas complejas hasta el acceso no autorizado a sistemas restringidos mediante la manipulación de datos biométricos, lo que lesiona la fe pública y el patrimonio de las personas (Endara-Chamorro et al., 2024).

Afectación a la dignidad: Actúa como una cláusula de cierre destinada a proteger la autonomía y la esencia del ser humano frente a cualquier forma de instrumentalización algorítmica. Esta finalidad sanciona aquellas prácticas que reducen al individuo a simples patrones estadísticos o que manipulan su voluntad, garantizando que la tecnología se mantenga

siempre como un instrumento al servicio del hombre y no como un medio para su deshumanización (Chávez Zambrano & San Andrés Molina, 2025).

En conclusión, la precisión en la descripción de estos verbos y finalidades garantiza el respeto estricto al principio de tipicidad, permitiendo que el Estado sancione eficazmente al agresor tecnológico sin vulnerar el marco garantista de la Constitución. Al anclar la punibilidad en el elemento subjetivo, se asegura que el Derecho Penal intervenga únicamente como ultima ratio ante conductas de extrema lesividad, preservando así la seguridad jurídica y el desarrollo tecnológico legítimo en la sociedad digital (Santacruz Torré et al., 2025).

3.8. Objeto material del delito.

En la estructuración dogmática de cualquier tipo penal, el objeto material se define como la persona o cosa sobre la cual recae directamente la conducta típica. En el contexto de los delitos algorítmicos utilizando inteligencia artificial, el objeto material abandona la corporeidad tradicional para trasladarse al ciberespacio. Para la tipificación autónoma del deepfake de naturaleza sexual o vejatoria, la conducta infractora recae sobre los siguientes elementos estructurados:

- **Imagen, Rostro, Voz y Datos Biométricos:** El primer nivel del objeto material está constituido por los atributos inherentes a la víctima, por ejemplo, su imagen, su rostro o su voz. Jurídicamente, estos elementos han dejado de ser vistos como simples características físicas para consolidarse como datos biométricos y datos de carácter personal altamente sensibles, ya que permiten la identificación inequívoca de un individuo (Ponce González et al., 2025).

  La acción delictiva se materializa cuando el agresor extrae y manipula estos datos biométricos (a través de técnicas conocidas como deepface para la superposición de rostros y deepvoice para la clonación de la voz) con el fin de insertarlos en contextos

para los cuales la víctima nunca otorgó su consentimiento (Taylor Villafuerte, 2024).

- **Video, Audio y Contenido Audiovisual Sintético o Manipulado:** El soporte en el cual se plasma y ejecuta el delito lo conforman el video y el audio, los cuales en conjunto se definen como contenido audiovisual sintético o manipulado. El COIP, en su artículo 500, ya nos otorga una base al definir el "contenido digital" como todo dato informático que representa hechos, información o conceptos de la realidad, procesados y transmitidos por cualquier medio tecnológico (Asamblea Nacional del Ecuador, 2014).

  El objeto material del deepfake recae precisamente sobre este contenido digital, el cual no es meramente editado, sino fabricado y manipulado estructuralmente mediante arquitecturas de redes neuronales y aprendizaje profundo (Deep learning) para alterar la realidad visual y auditiva de la grabación (Valverde-Ramos et al., 2024).

- **Representación Digital de una Persona:** En el núcleo de esta conducta, el objeto material trasciende el archivo informático para convertirse en la representación digital de una persona. El delito sanciona la usurpación y alteración de la identidad virtual de la víctima, construyendo una simulación paralela en la que el sujeto es forzado a protagonizar actos íntimos, sexuales o humillantes en los que nunca participó materialmente. Esta representación digitalizada cosifica al ser humano, convirtiéndolo en un objeto de exhibición altamente replicable en el entorno digital (Rosales Murillo, 2024).

- **Material que aparente ser real y pueda generar confusión razonable:** Finalmente, para que el objeto material adquiera relevancia en este nuevo tipo penal y no se confunda con caricaturas, animaciones evidentes o sátiras protegidas por la

> libertad de expresión, el contenido manipulado debe poseer una característica condicionante, debe ser un material que aparente ser real y pueda generar confusión razonable. El nivel de perfeccionamiento del deepfake (su hiperrealismo) debe ser lo suficientemente verosímil como para engañar la percepción humana, provocando que terceros asuman la autenticidad del material (Valverde-Ramos et al., 2024).
>
> Es precisamente esta capacidad de generar una "confusión razonable" en la sociedad lo que consuma la destrucción del buen nombre de la víctima y el daño a su dignidad, justificando la intervención del derecho penal (Ponce González et al., 2025).

3.9. Elemento subjetivo del tipo penal.

La estructuración de un nuevo tipo penal exige delimitar con precisión la intencionalidad del sujeto activo, para evitar que el poder punitivo del Estado sancione conductas inofensivas. En el entorno de las nuevas tecnologías, es imperativo establecer que no cualquier manipulación digital o uso de inteligencia artificial generativa constituye delito; para que la conducta sea punible, debe existir inexorablemente el dolo.

Según el artículo 26 del COIP, el dolo se configura cuando la persona, "conociendo los elementos objetivos del tipo penal, ejecuta voluntariamente la conducta" (Asamblea Nacional del Ecuador, 2014).

Aplicando la violencia digital sintética, el elemento subjetivo requiere que el agresor tenga conocimiento pleno y exacto de que el contenido audiovisual que está manejando es falso, sintético o manipulado. A la par de este componente cognitivo, debe existir la voluntad libre y deliberada de crear, manipular o difundir dicho contenido utilizando los datos biométricos de la víctima (Rosales Murillo, 2024).

Además del dolo genérico, el deepfake de carácter malicioso exige un dolo específico o elemento tecnológico trascendente. La voluntad de crear o difundir el material debe estar

motivada por finalidades que conviene diferenciar antes que agrupar bajo una única etiqueta: el ánimo de causar daño, humillación o descrédito (lo que la doctrina penal identifica como animus injuriandi, o ánimo ofensivo, propio del tipo básico); el ánimo de explotación sexual o de obtención de un beneficio económico indebido (propio de modalidades agravadas o de figuras en concurso, como la pornografía infantil, la extorsión sexual o la trata de personas); y el ánimo defraudatorio (que remite a tipos patrimoniales o de falsedad en concurso con la figura aquí propuesta). Esta diferenciación evita que una sola fórmula subjetiva absorba conductas de naturaleza y gravedad distintas, y permite decidir con mayor precisión cuáles finalidades integran el tipo básico, cuáles configuran una modalidad agravada y cuáles deben resolverse mediante las reglas del concurso de delitos (Yépez Andrade, 2026).

Como evidencian diversos casos de violencia cibernética, el sujeto activo no siempre persigue un lucro económico (como en la extorsión), sino que a menudo actúa motivado por la venganza, el morbo o la "diversión en la crueldad", buscando la cosificación y destrucción pública de la identidad de la víctima (Andino Veloz, 2025).

En la tipificación autónoma de este delito, queda estrictamente excluida la culpa o la imprudencia. El artículo 27 del COIP señala que actúa con culpa quien "infringe el deber objetivo del cuidado" provocando un resultado dañoso (Asamblea Nacional del Ecuador, 2014).

Sin embargo, el proceso técnico para generar un deepfake (que implica recolectar fotografías, introducirlas en un software de redes neuronales, entrenar al algoritmo y procesar el video final, u otras variantes de como realizarlo) es una cadena de acciones sistemáticas y predeterminadas. Es material y ontológicamente imposible crear un deepfake sexual o difamatorio "por accidente" o por mera negligencia. Por consiguiente, salvo que se justifique otra opción legislativa muy excepcional para plataformas tecnológicas omisivas, la conducta del creador y difusor es eminentemente dolosa (Zambrana Ruiz, 2024).

La exigencia ineludible del dolo lesivo es precisamente el filtro dogmático que resuelve la colisión aparente entre el derecho al honor y la libertad de expresión. La inteligencia artificial posee usos legítimos en la industria audiovisual; por ende, quedan excluidos de la esfera penal los perfiles o identidades caricaturescas creadas a título de broma, sátira o parodia, así como las representaciones con fines estrictamente artísticos, científicos o informativos (Addati, 2020).

Para que la excepción sea válida, la parodia debe ser claramente identificable por el público como tal, sin intención de hacerla pasar por verídica. Si el contenido carece de la intencionalidad de engañar y no busca destruir el núcleo esencial de la dignidad humana ni someter sexualmente a la persona, no existe delito. Este límite asegura que el Derecho Penal actúe como ultima ratio, castigando únicamente la violencia digital grave y protegiendo al mismo tiempo la libertad de expresión y creación en el ciberespacio.

3.10. Circunstancias Agravantes.

La estructuración de un nuevo tipo penal autónomo para el deepfake requiere establecer circunstancias agravantes específicas. Aunque la sola creación maliciosa del contenido sintético ya amerita sanción, existen factores y medios comisivos que incrementan exponencialmente la lesividad del acto, lo que justifica normativamente un mayor reproche penal y el aumento de la pena privativa de libertad. A continuación, se detallan las circunstancias que agravarían esta conducta:

- **Cuando la víctima sea un niño, niña o adolescente:** La vulneración de derechos mediante inteligencia artificial tiene un impacto devastador en la población infantojuvenil, cuyo libre desarrollo de la personalidad e indemnidad sexual se encuentran en formación. En concordancia con el artículo 47, numeral 11 del Código Orgánico Integral Penal (COIP), cometer la infracción en perjuicio de

menores de edad constituye un agravante general (Asamblea Nacional del Ecuador, 2014).

En el caso de los deepfakes, la manipulación de imágenes de menores exige una protección reforzada, ya que los expone a redes de explotación digital y abuso sexual basado en imágenes, dejándolos en una situación de especial vulnerabilidad (Rosales Murillo, 2024).

- **Cuando el contenido sea de carácter sexual:** Si bien un deepfake puede ser utilizado para difamar políticamente a una persona, la gravedad de la conducta alcanza su punto máximo cuando el contenido creado es de carácter pornográfico o íntimo. El deepfake sexual ataca directamente la libertad e indemnidad sexual, constituyendo una de las formas más puras de violencia de género digital que busca la cosificación y degradación de la víctima mediante representaciones explícitas no consentidas (Taylor Villafuerte, 2024).
- **Cuando se use para extorsionar o amenazar:** El uso de *deepfakes* como instrumento para el chantaje configura la agravante de la *sextorsión*. En esta circunstancia, el agresor utiliza el contenido hiperrealista para coaccionar a la víctima y exigirle favores sexuales o beneficios bajo la amenaza de difundir el material (Taylor Villafuerte, 2024).

  En este sentido, la doctrina propone incluso reformar los artículos 154 (Intimidación) y 185 (Extorsión) del COIP, para incluir como agravante específica la utilización y difusión de contenido audiovisual manipulado con el fin de generar pánico, coacción, miedo o desprestigio sobre la víctima (Ponce González et al., 2025).
- **Cuando exista difusión masiva y el uso de plataformas digitales para amplificar el daño:** La consumación del daño a la reputación se magnifica cuando el contenido

trasciende el ámbito privado. La difusión de *deepfakes* a través de redes sociales, plataformas de mensajería o foros genera una viralización incontrolable ("enjambres digitales"), amplificando la humillación ante un público masivo (Andino Veloz, 2025).

Al utilizar las plataformas digitales, el infractor provoca que la víctima pierda por completo el control sobre su identidad y su "huella digital", configurando un daño permanente que justifica la agravación de la pena (Chicaiza-Flores et al., 2024).

- **Cuando genere afectación grave a la vida personal, familiar, laboral o educativa de la víctima:** La hiperreproductibilidad de este contenido vejatorio no se limita al ciberespacio, sino que destruye la esfera material de la persona. Si a consecuencia de la creación o difusión del *deepfake*, la víctima sufre graves afectaciones psicológicas (como trastornos de ansiedad, depresión severa o ideaciones suicidas), o si esto desemboca en el aislamiento social, pérdida de su empleo o el abandono forzoso de sus estudios (como se evidenció en casos de colegiales en el Ecuador), la pena debe ser agravada en respuesta al nivel de destrucción existencial ocasionado (Andino Veloz, 2025).
- **Cuando el agresor sea una persona con relación de confianza, autoridad o poder frente a la víctima:** La traición a la confianza y el abuso de superioridad incrementan la reprochabilidad del acto. En armonía con los artículos 47 (numeral 8) y 48 (numeral 8) del COIP, la conducta debe ser agravada si el infractor es el cónyuge, expareja, familiar, o si prevalece de una situación de superioridad laboral, docente o religiosa para obtener las imágenes base de la víctima y utilizarlas para generar el *deepfake* (Taylor Villafuerte, 2024).

3.11. Exclusiones o límites para evitar una regulación excesiva.

La creación de un tipo penal autónomo para el *deepfake* debe enmarcarse estrictamente dentro del principio de mínima intervención penal o *ultima ratio*. Este principio imperativo, consagrado en el artículo 3 del Código Orgánico Integral Penal (COIP), establece que la intervención punitiva del Estado está legitimada únicamente cuando es estrictamente necesaria para la protección de las personas y cuando los mecanismos extrapenales han resultado insuficientes (Asamblea Nacional del Ecuador, 2014). Por lo tanto, para garantizar la proporcionalidad y constitucionalidad de la norma, es indispensable establecer límites dogmáticos claros que eviten una regulación excesiva, previniendo así un punitivismo ciego que coarte las libertades civiles (Andino Veloz, 2025).

Para respetar el principio de legalidad penal (*nullum crimen, nulla poena sine lege*), se deben evitar a toda costa los "tipos penales abiertos" o ambiguos. La falta de taxatividad técnica permite la criminalización de conductas inofensivas, genera un efecto disuasivo sobre el ejercicio de derechos y otorga un margen de arbitrariedad peligroso a los administradores de justicia (Santacruz Torré et al., 2025). En consecuencia, la norma no debe criminalizar la mera manipulación tecnológica *per se*; el Estado solo debe activar su poder punitivo cuando exista una afectación real, grave y un peligro concreto a bienes jurídicos protegidos (tales como la dignidad, la integridad moral, el honor o la indemnidad sexual) derivado de una conducta dolosa.

El núcleo material de la antijuridicidad en este ciberdelito radica en la vulneración de la autonomía y autodeterminación de la persona. Por ello, el tipo penal debe exigir ineludiblemente la ausencia de consentimiento libre, previo e informado del titular de los datos biométricos, imagen o voz (Corte Constitucional del Ecuador, 2021). Si el titular otorga su autorización expresa para la recreación de su imagen mediante inteligencia artificial, la conducta es atípica, siempre y cuando el creador no exceda los límites o las finalidades específicas para los cuales se concedió dicha autorización.

Cualquier propuesta regulatoria en el entorno digital debe armonizarse con el derecho a la libertad de expresión, el cual comprende la facultad de buscar, recibir y difundir informaciones e ideas de toda índole (Pacto Internacional de Derechos Civiles y Políticos, 1966, art. 19).

Sin embargo, la jurisprudencia de la Corte Constitucional del Ecuador (Sentencia No. 047-SIN-CC) ha sido enfática al establecer que la libertad de expresión no es un derecho absoluto y encuentra una frontera infranqueable en el respeto a los derechos, la reputación de los demás, el honor y el buen nombre (Corte Constitucional del Ecuador, 2021). La libertad de expresión jamás podrá utilizarse como escudo jurídico o excusa legal para lesionar derechos de terceros mediante la difusión maliciosa de falsedades hiperrealistas.

A pesar de este límite, para evitar la censura previa, la norma penal debe establecer exclusiones expresas para aquellos actores legítimos que generen un valor social mediante el uso de la IA (Andino Veloz, 2025).

Deben quedar tajantemente exentos de responsabilidad penal los usos periodísticos e informativos que gocen de interés público, así como el uso legítimo de la tecnología con fines educativos, históricos, científicos y artísticos. De igual manera, se excluyen las creaciones enmarcadas dentro del entretenimiento puro, tales como la ironía, la sátira, la crítica y la parodia, siempre que su intención (*animus iocandi*) no sea la de humillar o someter sexualmente a una persona.

Finalmente, para que la exclusión de la sátira o la parodia sea procesalmente efectiva, la propuesta normativa debe diferenciar claramente entre el *deepfake* y la ficción evidente. El reproche penal debe reservarse para el contenido audiovisual *hiperrealista* y engañoso que posea la capacidad idónea de generar una confusión razonable en el público, menoscabando así la reputación de la víctima. Por el contrario, los materiales cuyas alteraciones carecen de realismo y son evidentes a simple vista (como dibujos, caricaturas o los denominados

shallowfakes rudimentarios), que el público puede identificar fácilmente como falsos y que no persiguen un engaño lesivo, deben quedar fuera del ámbito de persecución del derecho penal (Valverde-Ramos et al., 2024).

3.12. Propuestas de estructura normativa del tipo penal.

Artículo (…).- Manipulación sintética no consentida de la identidad (deepfake).- La persona que, mediante sistemas de inteligencia artificial, algoritmos de aprendizaje profundo o cualquier otra tecnología análoga de manipulación digital, cree, genere, altere o edite imágenes, videos, audios o representaciones digitales que utilicen la imagen, el rostro, la voz u otros datos biométricos de una persona natural, sin su consentimiento libre, previo, específico e informado, simulando hechos, expresiones o actos que no corresponden a la realidad, siempre que el material aparente ser real y sea idóneo para generar en un tercero una confusión razonable sobre su autenticidad, con la finalidad de menoscabar la dignidad, la intimidad, el honor o la libertad sexual de la persona representada, o con el propósito de difundirlo, será sancionada con pena privativa de libertad de uno a tres años. Quedan excluidos de este artículo los usos con fines de sátira, parodia, crítica, investigación académica o periodística, y demás manifestaciones amparadas por la libertad de expresión y de creación artística, siempre que no persigan la finalidad descrita en el inciso anterior.

Artículo (…).- Difusión no consentida de contenido sintético.- Será sancionada con la misma pena prevista en el artículo anterior la persona que, conociendo el carácter sintético o manipulado del material descrito en el artículo precedente y sin contar con el consentimiento de la persona representada, lo reproduzca, publique, difunda o comercialice, cuando dicha conducta cause un menoscabo efectivo y verificable a la intimidad, el honor, la propia imagen, la identidad o los datos personales de la persona representada, o genere un peligro concreto de dicho menoscabo. No se considerará menoscabo, para efectos de este artículo, la sola

exposición pública del contenido cuando de ella no se derive una afectación identificable a los bienes jurídicos señalados.

Artículo (…). - Modalidad de contenido sexual. - Si el contenido creado, alterado, editado o difundido conforme a los artículos anteriores fuese de naturaleza sexual o pornográfica, la conducta constituye una modalidad agravada de los tipos previstos en los artículos precedentes y la pena privativa de libertad será de tres a cinco años.

Cuando el material represente, real o simuladamente, a niñas, niños o adolescentes, se aplicará el concurso aparente de normas con los artículos 103 y 104 del COIP, debiendo preferirse estos últimos por el principio de especialidad y por la protección reforzada de la niñez y adolescencia, sin perjuicio de que la presente figura opere de manera residual en los supuestos de manipulación sintética que no alcancen a configurar los elementos típicos de aquellos artículos.

La pena aquí prevista guarda coherencia con la escala punitiva de otros delitos sexuales y contra la intimidad del COIP que no involucran a menores de edad.

La fundamentación de este rango se apoya en un ejercicio de comparación con figuras afines del COIP. El tipo básico (uno a tres años, artículo anterior) se sitúa en el mismo rango que la violación a la intimidad (artículo 178, uno a tres años) y que el contacto con finalidad sexual con menores de dieciocho años por medios electrónicos en su modalidad simple (artículo 173, uno a tres años), delitos que protegen bienes jurídicos personalísimos de naturaleza análoga (intimidad, propia imagen, indemnidad) sin mediar violencia física.

La modalidad agravada de contenido sexual (tres a cinco años) se corresponde con el segundo tramo del artículo 173 cuando el acercamiento se obtiene mediante coacción o intimidación (tres a cinco años), y resulta sensiblemente inferior a la pena de la pornografía con utilización de niñas, niños o adolescentes del artículo 103 (trece a dieciséis años) y de su comercialización en el artículo 104 (diez a trece años), lo que evidencia que la mayor severidad

de estos últimos tipos, reservados a víctimas menores de edad, no se traslada sin más a la modalidad aquí propuesta para personas adultas.

El incremento de dos años adicionales frente al tipo básico se justifica por el mayor desvalor de la conducta cuando el contenido sintético tiene naturaleza sexual, en tanto añade una afectación a la libertad e indemnidad sexual a la ya existente sobre la intimidad y la propia imagen, sin llegar a equipararse a los delitos sexuales que sí involucran contacto físico o coacción directa sobre la víctima. Sobre los menores de edad, la preferencia por los artículos 103 o 104 del COIP procede únicamente cuando el hecho reúna efectivamente los elementos típicos de dichas disposiciones y exista, por tanto, un concurso aparente de normas resoluble por el principio de especialidad; en caso contrario, la presente figura opera de manera residual.

Circunstancias agravantes.- La pena máxima prevista se aumentará en un tercio cuando concurra alguna de las siguientes circunstancias, incremento que se aplica sobre el máximo de cada modalidad conforme a las reglas generales de agravación del COIP: sobre el máximo del tipo básico (tres años) el resultado es de cuatro años; sobre el máximo de la modalidad de contenido sexual (cinco años) el resultado es de seis años con ocho meses. Este aumento resulta proporcionado porque responde a circunstancias adicionales de mayor lesividad (vulnerabilidad de la víctima, finalidad intimidatoria, difusión masiva, afectación grave o posición de confianza del autor) que no forman parte de los elementos típicos básicos ni de la modalidad sexual agravada, por lo que su aplicación no supone valorar dos veces una misma circunstancia, sin perjuicio de la regla de exclusión expresa prevista para la minoría de edad en el numeral siguiente:

- La víctima sea niña, niño o adolescente, siempre que el hecho sea subsumible en el presente tipo y no corresponda aplicar, por especialidad, los artículos 103 o 104 del COIP. La misma circunstancia no se valorará nuevamente como agravante general. Se conserva, y se fundamenta por separado, el supuesto de la persona que por su

condición no pueda comprender el significado del hecho o resistirlo, sin perjuicio de la aplicación de los tipos penales de explotación sexual de menores cuando correspondan.

- El material se utilice para intimidar, coaccionar o extorsionar a la víctima.
- El contenido se difunda masivamente a través de redes sociales, plataformas digitales o servicios de mensajería.
- La conducta produzca una afectación grave a la vida personal, familiar, laboral o educativa de la víctima, o a su salud psicológica. Para la aplicación de esta agravante, la afectación o el daño a la salud psicológica no podrán presumirse. Deberán ser acreditados obligatoriamente mediante el respectivo informe pericial, el cual determinará la existencia del daño con base en parámetros técnicos y clínicos estandarizados. Asimismo, se entenderá por afectación grave aquella que imposibilite o altere de manera sustancial y prolongada el desarrollo normal de las actividades cotidianas de la víctima.
- El autor mantenga o haya mantenido con la víctima una relación de pareja, familiar, de confianza, o se prevalezca de una situación de autoridad o superioridad laboral, docente o religiosa.

Exclusiones.- No constituye infracción penal el tratamiento sintético de la imagen, voz o datos de una persona realizado con su autorización expresa y dentro de los fines para los que fue otorgada; ni los usos periodísticos e informativos de interés público, ni los fines educativos, históricos, científicos o artísticos; ni la sátira, parodia, crítica o caricatura que sea claramente identificable como ficticia por el público, carezca de idoneidad para generar confusión razonable y no tenga por finalidad la humillación o el sometimiento sexual de la persona representada.

3.13. Viabilidad jurídica de la propuesta en el sistema penal ecuatoriano.

El presente apartado responde de manera directa al objetivo general y a la hipótesis de esta investigación, analizando si resulta necesario actualizar el Código Orgánico Integral Penal (COIP) frente a la violencia digital generada por inteligencia artificial. Para fundamentar la viabilidad de la propuesta, se realiza un análisis comparativo de las alternativas legislativas, examinando por qué una simple reforma a los tipos penales de violación a la intimidad (Art. 178) o suplantación de identidad (Art. 212) presenta limitaciones dogmáticas para tutelar los derechos fundamentales en juego.

El artículo 178 del COIP sanciona a quien, sin contar con el consentimiento o la autorización legal, acceda, intercepte, examine, retenga, grabe, reproduzca, difunda o publique datos personales, mensajes de datos, voz, audio y video de otra persona (Asamblea Nacional del Ecuador, 2014, art. 178). Aunque a primera vista una reforma a este artículo podría parecer la opción más sencilla, esta alternativa presenta debilidades dogmáticas relevantes para el fenómeno del contenido sintético:

- **Adecuación del objeto material:** el tipo clásico de violación a la intimidad protege el secreto y la privacidad de información real que ha sido obtenida, registrada o difundida sin autorización. En el caso del deepfake sexual o vejatorio, el contenido se fabrica sintéticamente a partir de imágenes o datos biométricos de carácter público extraídos de redes sociales, lo que exige examinar si concurre el objeto material del tipo cuando lo que se genera es una representación inexistente y no la difusión de información real previamente obtenida o registrada. La disponibilidad pública de una fotografía tampoco autoriza automáticamente su tratamiento sintético ni su difusión posterior.
- **La desnaturalización del bien jurídico "intimidad":** Dogmáticamente, en un *deepfake* sexual, el cuerpo representado en la escena de connotación sexual no pertenece físicamente a la víctima, sino que ha sido generado artificialmente o

pertenece a un tercero. Por esta razón, la doctrina penal advierte que no existe una invasión real a la privacidad física o corporal de la persona, sino una afectación directa a su derecho a la propia imagen, el honor y la dignidad. Intentar subsumir la creación de imágenes totalmente simuladas dentro de la violación a la intimidad desnaturaliza el bien jurídico protegido por el artículo 178, forzando una interpretación extensiva que colisiona con el principio de tipicidad estricta.

El artículo 212 del COIP tipifica la suplantación de identidad como un delito de resultado que exige usurpar la identidad de otra persona para "obtener un beneficio para sí o para un tercero, en perjuicio de una persona".

Aunque se ha propuesto añadir incisos que califiquen como suplantación la alteración de la imagen o voz mediante herramientas tecnológicas, esta vía también presenta limitaciones importantes:

Incompatibilidad del elemento teleológico: El delito de suplantación de identidad está estructurado bajo la premisa de que el autor actúa en nombre de la víctima para realizar transacciones, asumir obligaciones jurídicas o engañar a terceros en el tráfico civil y comercial. Sin embargo, muchos de los deepfakes maliciosos (especialmente los de carácter pornográfico) no pretenden "reemplazar" a la víctima en sus relaciones jurídicas, sino someterla a un escenario de humillación, cosificación y escarnio social. La distinción entre suplantar y manipular: Suplantar implica asumir el rol civil de otro; en contraste, el deepfake es una representación visual o auditiva falsa que sitúa a la víctima en una situación ficticia.

Forzar el artículo 212 para castigar la manipulación de imágenes que no cumple sus elementos típicos desvirtúa el sentido dogmático de la suplantación, generando un tipo penal desproporcionado que no logra capturar el verdadero desvalor de la violencia digital.

Frente a las limitaciones de reformar los artículos 178 o 212, la creación de un tipo penal autónomo para el *deepfake* se presenta como la alternativa más viable, coherente y

respetuosa con la Constitución de la República del Ecuador. Esta propuesta se alinea con los siguientes principios:

- **Compatibilidad con el principio de legalidad y tipicidad estricta:** Al describir con precisión verbos rectores que no figuran en los tipos aplicables a personas adultas (como “crear”, “generar” o “manipular” contenido sintético), se dota al sistema de justicia de una norma taxativa que evita interpretaciones analógicas prohibidas por la CRE y el COIP.
- **Tutelación de un bien jurídico pluriofensivo:** Un tipo penal específico reconoce que el uso malicioso de la IA no solo vulnera la privacidad, sino que lesiona concurrentemente el honor, la propia imagen, la autodeterminación informativa y la integridad moral de la persona, requiriendo un marco de protección que varias legislaciones comparadas han comenzado a implementar.
- **Sistemática de la autoría y la participación:** La tipificación autónoma permitiría sancionar de manera diferenciada y proporcional tanto la creación del material sintético hiperrealista como su posterior difusión, de modo que la respuesta penal no dependa únicamente de que el contenido haya sido publicado. Estas dificultades corresponden a la tipicidad sustantiva de las conductas de creación y difusión, y no a lagunas de naturaleza procesal.

3.13.1. Compatibilidad con el principio de legalidad y la constitucionalización del derecho penal.

La propuesta de un tipo penal autónomo es la materialización directa del principio de legalidad (nullum crimen, nulla poena sine lege), consagrado en el artículo 76.3 de la Constitución de la República y el artículo 5.1 del COIP. La constitucionalización del derecho penal exige que las leyes sean claras, específicas y accesibles para evitar interpretaciones analógicas que vulneren la seguridad jurídica (Santacruz Torré et al., 2025). Al crear un tipo penal que describa con exactitud y de forma cerrada los verbos rectores “crear, generar, alterar

o editar" contenido sintético mediante inteligencia artificial, se elimina la ambigüedad, previniendo que los jueces fuercen la aplicación de otros delitos, garantizando así un debido proceso y una tipicidad estricta (Monar Puente et al., 2024).

3.13.2. Compatibilidad con la mínima intervención penal (Ultima Ratio).

El artículo 3 del COIP establece el principio de mínima intervención, el cual dicta que el derecho penal es el último recurso del Estado (Asamblea Nacional del Ecuador, 2014). La doctrina reconoce que el principio de *ultima ratio* es un límite esencial al poder punitivo, exigiéndole al Estado intervenir solo cuando otras ramas del derecho son ineficaces (Carnevali Rodríguez, 2008).

La viabilidad de esta propuesta se sustenta en que los actuales mecanismos extrapenales (tanto en el ámbito civil-constitucional como en el administrativo) presentan significativas limitaciones normativas, procesales y operativas para responder con la inmediatez y eficacia requeridas ante la celeridad y escala de la violencia digital (Palacios Zúñiga, 2025).

Por un lado, la garantía jurisdiccional del Hábeas Data, consagrada en el artículo 92 de la Constitución de la República y desarrollada en el artículo 50 de la Ley Orgánica de Garantías Jurisdiccionales y Control Constitucional (LOGJCC), sí cuenta con un trámite autónomo en su diseño normativo. Sin perjuicio de ello, su aplicación práctica frente a la alta exposición y viralización de contenidos en internet plantea dificultades relevantes: los plazos y las medidas cautelares previstas con carácter general para las garantías jurisdiccionales no fueron concebidos para la velocidad de propagación de contenido en redes sociales, y la falta de especialización técnica de los operadores de justicia puede dificultar que se ordene con la celeridad requerida la supresión y contención de datos biométricos sensibles (como el rostro y la voz) antes de que el daño psicológico, moral o reputacional sobre la víctima se torne difícilmente reversible. Se trata, en consecuencia, de dificultades de ejecución práctica y no de

una ausencia de cauce procesal autónomo, distinción que conviene mantener con precisión (Corte Constitucional del Ecuador, 2021).

Por otro lado, en la vía administrativa, el marco de protección de datos personales ha avanzado recientemente: mediante la Resolución N.° SPDP-SPD-2026-0009-R, de 12 de febrero de 2026, la Superintendencia de Protección de Datos Personales (SPDP) expidió la Norma General para la Garantía del Derecho de Protección de Datos Personales en el Uso de Sistemas de Inteligencia Artificial, que la faculta expresamente para auditar los sistemas de inteligencia artificial que traten datos personales e imponer medidas correctivas o cautelares, así como para aplicar el régimen sancionatorio de la Ley Orgánica de Protección de Datos Personales (LOPDP) en caso de incumplimiento. Esta facultad es, sin embargo, de alcance limitado: se circunscribe a la protección de datos personales (minimización, consentimiento, decisiones automatizadas) y no equivale a una competencia específica para calificar o sancionar el uso malicioso de deepfakes con fines sexuales, difamatorios o fraudulentos, lo que excede el objeto de la LOPDP y de la propia SPDP. Persiste, por tanto, una brecha de control en el ámbito estrictamente penal, antes que una ausencia total de facultades administrativas sobre la inteligencia artificial (Chávez Zambrano & San Andrés Molina, 2025).

En consecuencia, si bien estos mecanismos constituyen valiosos esfuerzos para la protección de la privacidad, sus limitaciones operativas impiden una respuesta integral y restauradora frente a la gravedad y el desvalor del contenido sintético hiperrealista (Palacios Zúñiga, 2025). Esta insuficiencia procesal y regulatoria justifica la intervención del Derecho Penal, no bajo una premisa punitivista de anulación de otras vías, sino de acuerdo con el principio de intervención mínima y la constitucionalización de las penas, legitimando al derecho punitivo como un instrumento necesario para tutelar de forma proporcional los derechos fundamentales vulnerados y evitar escenarios de impunidad estructural en el ciberespacio (Santacruz Torré et al., 2025).

3.13.3. Necesidad de protección de bienes jurídicos constitucionalmente relevantes.

El ejercicio del poder punitivo se justifica exclusivamente para proteger los bienes jurídicos más valiosos de la sociedad. La viabilidad de este nuevo tipo penal se sustenta en que resguarda de manera directa la dignidad humana, considerada el núcleo fundacional del Estado de derechos (García Toma, 2018). El *deepfake* malicioso ataca derechos personalísimos como el honor, la intimidad y el derecho a la propia imagen y protección de datos, todos amparados por el artículo 66 de la Constitución (Yépez Andrade, 2026).

3.13.4. Coherencia con el COIP y Necesidad de proporcionalidad en la pena.

La propuesta se integra de forma armónica a la sistemática del COIP, ubicándose dentro de los delitos contra la integridad personal o la intimidad. Para que esta inclusión sea válida, debe superar el *test de proporcionalidad* (idoneidad, necesidad y proporcionalidad en sentido estricto). Esto significa que la pena privativa de libertad sugerida debe guardar un equilibrio racional con la gravedad del daño ocasionado, evitando caer en un "populismo punitivo" que fije penas irracionales o desmedidas. La sanción debe ser humanitaria y orientada a la protección de la víctima y la rehabilitación del infractor (Santacruz Torré et al., 2025).

En concreto, para el rango de uno a tres años, el delito comparable es la violación a la intimidad (artículo 178 del COIP, uno a tres años), que protege un bien jurídico de naturaleza análoga sin mediar violencia física ni contacto directo con la víctima, lo que satisface la idoneidad y necesidad de la pena propuesta para el tipo básico. Para el rango de tres a cinco años, el delito comparable es el contacto con finalidad sexual con menores de dieciocho años por medios electrónicos en su modalidad agravada por coacción o intimidación (artículo 173, tres a cinco años), cuya escala punitiva resulta compatible con la mayor lesividad que supone la naturaleza sexual del contenido sintético, sin alcanzar la severidad reservada a los delitos que involucran víctimas menores de edad o contacto físico directo (artículos 103 y 104 del COIP). El aumento de un tercio previsto para las circunstancias agravantes se mantiene dentro de estos mismos parámetros de proporcionalidad estricta, al recaer sobre circunstancias

adicionales de mayor desvalor que no se superponen con los elementos típicos ya considerados en cada modalidad.

La viabilidad de esta tipificación autónoma podría proporcionar a la Fiscalía General del Estado una herramienta más precisa para formular cargos, procesar a los agresores cibernéticos y contribuir a garantizar el derecho a la verdad y la reparación integral (Ponce González et al., 2025).

3.13.5. Viabilidad jurídica condicionada al respeto de los límites del ius puniendi.

En definitiva, la propuesta es jurídicamente viable siempre y cuando el Estado reconozca que la Constitución actúa como un freno infranqueable a su facultad de castigar (*ius puniendi*). El derecho penal no puede prohibir los avances tecnológicos, pero sí debe sancionar el comportamiento humano doloso que instrumentaliza dicha tecnología para vulnerar gravemente los derechos de terceros (Saquicela Rodas & Chica Mejía, 2023).

3.13.6. Cierre del Capítulo III.

A lo largo de esta investigación se ha abordado el fenómeno del deepfake no como un simple avance tecnológico, sino como un mecanismo que vulnera la dignidad humana en el entorno digital. En el primer capítulo se demostró que el uso malicioso y no consentido de la inteligencia artificial generativa lesiona bienes jurídicos como la intimidad, el honor, la propia imagen y la libertad sexual, despojando a la persona de su autonomía y convirtiéndola en objeto de exposición pública no consentida.

En el segundo capítulo, el análisis dogmático evidenció dificultades de subsunción en determinadas modalidades de creación y manipulación sintética frente al Código Orgánico Integral Penal (COIP), reconociendo que otras modalidades pueden satisfacer los elementos de tipos vigentes. Una respuesta legislativa debe dirigirse a las conductas cuya falta de cobertura haya sido justificada, sin ampliar por analogía las normas existentes. Esta ausencia de verbos rectores adecuados a determinadas conductas puede dejar a las víctimas sin una respuesta penal

específica y afecta de manera desproporcionada a mujeres, niñas, niños y adolescentes, quienes suelen ser blanco de esta forma de violencia digital de género.

Frente a estas dificultades de subsunción y a la necesidad de garantizar la tutela judicial efectiva, este tercer capítulo desarrolló los elementos dogmáticos y normativos necesarios para plantear una respuesta legislativa respetuosa de los límites constitucionales al ius puniendi. Como resultado de todo el análisis realizado, se concluye que sí es jurídicamente viable proponer un tipo penal autónomo para el uso malicioso de deepfakes, siempre que dicho tipo sea claro, limitado, proporcional y dirigido únicamente a las conductas que lesionan bienes jurídicos relevantes.

Esta viabilidad está condicionada al respeto estricto de los principios de mínima intervención penal (ultima ratio) y de legalidad. La propuesta normativa que se plantea busca que el reproche penal recaiga únicamente sobre la conducta dolosa que instrumentaliza los datos biométricos de una persona con la finalidad de menoscabar su dignidad, su intimidad, su honor o su libertad sexual. Asimismo, al establecer exclusiones expresas para el ejercicio legítimo de la libertad de expresión, como la sátira, la parodia o los fines académicos e informativos, se busca evitar el populismo punitivo y la sobrecriminalización.

En definitiva, la incorporación de un tipo penal específico para la manipulación sintética no consentida de la identidad se presenta como una alternativa jurídicamente viable para actualizar el sistema penal ecuatoriano frente a estas nuevas formas de afectación. Esta reforma legislativa podría contribuir a proteger de forma más adecuada la dignidad humana como valor supremo del ordenamiento jurídico.

Conclusiones generales

A partir del desarrollo de los tres capítulos que integran esta investigación, es posible formular las siguientes conclusiones generales:

Primera. El uso malicioso de deepfakes constituye una forma de violencia digital que lesiona bienes jurídicos personalísimos "la dignidad humana, la intimidad, el honor, la propia imagen y, en determinados casos, la libertad e indemnidad sexual" al instrumentalizar sin consentimiento la imagen, el rostro, la voz u otros datos biométricos de una persona, afectando de manera desproporcionada a mujeres, niñas, niños y adolescentes.

Segunda. El Código Orgánico Integral Penal (COIP) presenta, en las modalidades analizadas en esta tesis "en particular, la manipulación sintética no consentida sin fines sexuales explícitos ni ánimo calumnioso", dificultades de subsunción jurídica frente a los tipos penales vigentes de violación a la intimidad, calumnia y suplantación de identidad, sin que ello implique una insuficiencia absoluta del ordenamiento penal frente a todas las variantes del fenómeno.

Tercera. Es jurídicamente viable proponer un tipo penal autónomo para el uso malicioso de deepfakes, siempre que su redacción sea clara, taxativa y proporcional, exija un elemento subjetivo específico y una afectación o peligro verificable a los bienes jurídicos protegidos, y contemple excepciones expresas para los usos legítimos amparados por la libertad de expresión y de creación artística.

Cuarta. La respuesta penal debe articularse de manera coherente con las figuras ya existentes en el COIP "en particular los artículos 103 y 104, aplicables a representaciones reales o simuladas de niñas, niños o adolescentes" y con los mecanismos extrapenales de protección de datos personales y de garantías constitucionales, cuya efectividad práctica frente a la viralización de contenido digital requiere fortalecimiento técnico y procesal.

Quinta. El derecho comparado (Estados Unidos, Reino Unido y la Unión Europea) confirma una tendencia internacional hacia la tipificación específica de la creación y difusión no consentida de contenido sintético, tendencia que coexiste en Ecuador, a la fecha de esta investigación, con diversas iniciativas legislativas en trámite ante la Asamblea Nacional orientadas a la violencia sexual digital y al contenido sintético, lo que refuerza la pertinencia de la propuesta desarrollada en el Capítulo III sin restarle su carácter de aporte técnico específico.

Recomendaciones

Con base en las conclusiones precedentes, se formulan las siguientes recomendaciones:

A la Asamblea Nacional: valorar la incorporación al COIP de un tipo penal autónomo de manipulación sintética no consentida de la identidad, en los términos propuestos en el Capítulo 3 de esta investigación, articulándolo expresamente con las iniciativas legislativas actualmente en trámite sobre violencia digital y contenido sintético, a fin de evitar antinomias y vacíos de coordinación normativa.

A la Fiscalía General del Estado y a la Función Judicial: fortalecer la capacitación técnica de fiscales, jueces y peritos en la identificación, valoración y cadena de custodia de evidencia digital sintética (deepfakes), dada la dificultad probatoria que plantea el hiperrealismo de estos contenidos.

A la Superintendencia de Protección de Datos Personales: precisar, dentro del ámbito de sus competencias legales, los alcances de sus facultades de auditoría sobre sistemas de inteligencia artificial frente a usos que involucren datos biométricos sensibles, y coordinar con la Fiscalía General del Estado protocolos de actuación conjunta para los casos que excedan el ámbito administrativo de protección de datos.

A la Corte Constitucional y al Consejo de la Judicatura: evaluar mecanismos procesales expeditos, dentro del diseño ya existente del Hábeas Data, que permitan atender con mayor celeridad las solicitudes de supresión y contención de datos biométricos sensibles difundidos sin consentimiento en entornos digitales.

A las instituciones de educación superior y a la sociedad civil: promover investigaciones y campañas de concientización sobre el uso malicioso de la inteligencia artificial generativa, con énfasis en la protección reforzada de mujeres, niñas, niños y adolescentes frente a la violencia digital de género.

A futuras investigaciones: profundizar, con apoyo de fuentes primarias verificables y de la jurisprudencia nacional que se produzca en la materia, el análisis de la relación entre el artículo 103 del COIP y las representaciones simuladas de menores mediante inteligencia artificial, así como actualizar periódicamente el estudio de derecho comparado dado el dinamismo normativo observado en esta materia.

Bibliografías

Abbas, F., & Taeihagh, A. (2024). Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence. *Expert Systems with Applications*, *252*, 124260.

Abdali, S., Shaham, S., & Krishnamachari, B. (2024). Multi-modal misinformation detection: Approaches, challenges and opportunities. *ACM Computing Surveys*, *57*(3), Artículo 76.

Addati, F. A. (2020). El impacto de las redes sociales en los derechos personalísimos. *Ratio Iuris. Revista de Derecho Privado*, *8*(2), 35-81.

Andino Veloz, B. (2025). Deepfake, cinismo y diversión en la crueldad: Un caso de colegiales y pornografía en Ecuador. *Uru: Revista de Comunicación y Cultura*, (11), 8-28. https://doi.org/10.32719/26312514.2025.11.2

Asamblea Nacional del Ecuador. (2008). *Constitución de la República del Ecuador*. Registro Oficial 449.

Asamblea Nacional del Ecuador. (2014). *Código Orgánico Integral Penal (COIP)*. Registro Oficial Suplemento 180.

Asamblea Nacional del Ecuador. (2021). *Ley Orgánica de Protección de Datos Personales*. Registro Oficial Suplemento 459.

Atlam, E.-S., Almaliki, M., Elmarhomy, G., Almars, A. M., Elsiddieg, A. M. A., & ElAgamy, R. (2025). SLM-DFS: A systematic literature map of deepfake spread on social media. *Alexandria Engineering Journal*, *111*, 446–455.

Ávila Merino, C. E. (2019). *El derecho a la intimidad y la protección de datos personales en el marco jurídico ecuatoriano* [Trabajo de titulación, Universidad Central del Ecuador].

Ávila Zea, M. P., & Zamora Vázquez, A. F. (2024). Aplicación de la inteligencia artificial en el derecho penal: Problemas y desafíos. *Resistances. Journal of the Philosophy of History*, *5*(9), e240142.

Carnevali Rodríguez, R. (2008). Derecho penal como ultima ratio. Hacia una política criminal racional. *Revista Ius et Praxis*, *14*(1), 13-48.

Chávez Zambrano, F., & San Andrés Molina, C. (2025). Intersección entre inteligencia artificial y derecho: Implicaciones jurídicas y limitaciones del marco regulatorio ecuatoriano frente a la IA. *Nullius: Revista de Pensamiento Crítico en el Ámbito del Derecho*, *6*(2), 130-140. https://doi.org/10.33936/nullius.v6i2.8059

Chicaiza-Flores, M. J., Coloma-Armijos, J. A., Maldonado-Valle, K. J., & Granja-Zurita, D. F. (2024). Pornovenganza y violación a la intimidad sexual de las víctimas en el Ecuador. *Verdad y Derecho Revista Arbitrada de Ciencias Jurídicas y Sociales*, *3*(Especial Ambato), 272-279.

Corte Constitucional del Ecuador. (2021, 27 de enero). *Sentencia No. 2064-14-EP/21* (Jueza ponente: Carmen Corral Ponce).

Cuenca Gonzaga, A. I., Fajardo Martillo, J. J., Jiménez Pillajo, D. J., Macías Guerrero, G. A., & Pesantez Vinueza, P. E. (2024). El delito de calumnia en redes sociales: Regulación en Ecuador y dilemas de la libertad de expresión. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, *5*(6), 596-607. https://doi.org/10.56712/latam.v5i6.3030

Endara-Chamorro, R. E., Espinoza-Jiménez, J. S., López-Fuel, E. R., & Santander-Moreno, J. J. (2024). Análisis jurídico del deepfake en relación a la suplantación de identidad, Ecuador. *Iustitia Socialis: Revista Arbitrada de Ciencias Jurídicas*, *9*(1), 240-250. https://doi.org/10.35381/racji.v9i1.3530

Erazo Bustamante, S., & Valdivieso Guerrero, T. (2018). La libertad de expresión y el derecho al honor. ¿Colisionan estos derechos fundamentales? *Dilemas Contemporáneos: Educación, Política y Valores*, *6*(1).

Espinosa Carvajal, G. G., & Paredes Fuentes, F. E. (2025). Los ciberdelitos y la protección de datos personales en el sistema penal ecuatoriano. *LEX Revista de Investigación en Ciencias Jurídicas*, *8*(29), 563-575. https://doi.org/10.33996/revistalex.v9i28.302

Ferrer, E., Martínez, F., & Figueroa, A. (2014). *Libertad de expresión y derechos humanos*. Instituto de Investigaciones Jurídicas, UNAM.

García Toma, V. (2018). La dignidad humana y los derechos fundamentales. *Revista Derecho & Sociedad*, (51), 13-31.

Groh, M., Sankaranarayanan, A., Singh, N., Kim, D. Y., Lippman, A., & Picard, R. (2024). Human detection of political speech deepfakes across transcripts, audio, and video. *Nature Communications*. https://doi.org/10.1038/s41467-024-51998-z

Monar Puente, J. J., Tandazo Ortega, J. A., Suriaga Bernita, T. C., & Baquerizo Balladares, N. M. (2024). Principios del derecho penal en el Ecuador: Una revisión sistemática. *RECIMUNDO*, *8*(3), 195-206. https://doi.org/10.26820/recimundo/8.(3).julio.2024.195-206

Palacios Zúñiga, M. P. (2025). *El hábeas data como recurso de protección de datos personales: Análisis jurisprudencial de la Sentencia Constitucional No. 2064-14-EP/21* [Trabajo de titulación, Universidad de Cuenca].

Ponce González, A. M., Espinoza Suárez, J. A., & Posso Loor, A. (2025). Protección de derechos sobre la propia imagen, datos personales e intimidad frente a los deepfakes. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, *6*(3), 276-288. https://doi.org/10.56712/latam.v6i3.3947

Rigotti, C., McGlynn, C., & Benning, F. (2024). Image-based sexual abuse and EU law: A critical analysis. *German Law Journal*, *25*, 1472-1493. https://doi.org/10.1017/glj.2024.49

Rosales Murillo, K. D. (2024). *Deepfakes: Análisis del uso de la IA para la creación de videos e imágenes pornográficas de mujeres, niñas, niños y adolescentes desde el ordenamiento jurídico penal ecuatoriano* [Trabajo de titulación, Universidad del Azuay].

Salvador Quiñonez, W. M. (2022). Derecho a la intimidad y la ciberdelincuencia: Efectos sociales y económicos en víctimas ecuatorianas. *Revista Mundo Financiero*, *3*(9), 41-54.

Santacruz Torré, E. E., Sarango Quinde, K. S., Pachay Palacios, J. J., & Domínguez Gordillo, G. M. (2025). La constitucionalización del derecho penal en Ecuador. *RECIAMUC*, *9*(4), 493-504. https://doi.org/10.26820/reciamuc/9.(4).diciembre.2025.493-504

Saquicela Rodas, I. P., & Chica Mejía, K. N. (2023). La Constitución como límite al ius puniendi en un Estado constitucional de derecho. *Revista Debate Jurídico Ecuador*, *6*(3), 250-266. https://doi.org/10.61154/dje.v6i3.3240

Taylor Villafuerte, L. E. (2024). *La tipificación del delito de deepfake sexual dentro del Código Orgánico Integral Penal* [Trabajo de titulación, Universidad Católica de Santiago de Guayaquil].

The current landscape of deepfake legislation. (2025). *Journal of Information Policy*, *15*, 98-135.

The Regulatory Review. (2025). *Deepfakes and the law: A review of legal scholarship on deepfake technology*.

Valverde-Ramos, M. S., Armijos-González, M. J., & Martínez-Pérez, O. (2024). La tipificación del delito deepfake pornográfico en Ecuador y la era de la inteligencia artificial. *MQRInvestigar*, *8*(2), 1950-1970. https://doi.org/10.56048/MQR20225.8.2.2024.1950-1970

Villalba Fiallos, A. (2017). Reflexiones jurídicas sobre la protección de datos y el derecho a la intimidad en la autodeterminación informativa. *Foro: Revista de Derecho*, (27), 23-42.

Wang, T., Liao, X., Chow, K. P., Lin, X., & Wang, Y. (2024). Deepfake detection: A comprehensive survey from the reliability perspective. *ACM Computing Surveys*, *57*(3), Artículo 58.

Yépez Andrade, M. (2026). El delito y los derechos personalísimos. *Revista Judicial*.

Zambrana Ruiz, E. (2024). *Delitos susceptibles de ser cometidos a través de la inteligencia artificial: Deepfakes sexuales. Análisis penal de los nuevos retos en la protección de la intimidad* [Trabajo final de grado, Universitat Autònoma de Barcelona]. https://ddd.uab.cat/record/319415

Anexo

Universidad Católica de Cuenca	AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL

Kevin Joel Pachala Tapia portador de la cédula de ciudadanía Nº **0750695439 y Pablo José Jiménez Romero Tapia** portador de la cédula de ciudadanía Nº **1400800411** En calidad de autores y titulares de los derechos patrimoniales del trabajo de titulación **"Deepfake como vulneración a la dignidad humana: insuficiencia del COIP y necesidad de regulación penal específica en Ecuador"** de conformidad a lo establecido en el artículo 114 Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación, reconozca a favor de la Universidad Católica de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos y no comerciales. Autorizo además a la Universidad Católica de Cuenca, para que realice la publicación de éste trabajo de titulación en el Repositorio Institucional de conformidad a lo dispuesto en el artículo 144 de la Ley Orgánica de Educación Superior.

Cuenca, **19 de Octubre de 2026**

F: .....

Kevin Joel Pachala Tapia

C.I. 0750695439

F: .....

Pablo José Jiménez Romero Tapia

C.I. 1400800411