REVISTA INVECOM
ISSN en línea: 2739-0063

Difamación en redes sociales y derecho al honor: una revisión sistemática

Defamation on social media and the right to honor: a systematic review

Recibido: 28/07/2026 - **Aceptado:** 03/10/2026

Juan Carlos Alcántara Farfán
https://orcid.org/0000-0003-0221-9148
jalcantaraf@ucvvirtual.edu.pe
Universidad César Vallejo. Lima, Perú

Nelly Patricia Saravia Yataco
https://orcid.org/0000-0002-4998-4958
npsaraviay@ucvvirtual.edu.pe
Universidad César Vallejo. Lima, Perú

Joaquín Sebastián Alcántara Saravia
https://orcid.org/0000-0002-2570-3576
alcantara.saravia.sebas@gmail.com
Universidad César Vallejo. Lima, Perú

Resumen

Este estudio se propuso analizar la literatura sobre difamación en redes sociales e identificar criterios normativos, jurisprudenciales y remediales que contribuyan a la protección del honor en el Perú. Para ello, se llevó a cabo una revisión sistematizada de carácter jurídico-documental, basada en 30 artículos publicados entre 2012 y 2026 y recuperados de SciELO, Redalyc, Dialnet, DOAJ y otras plataformas que ofrecen metadatos verificables. La búsqueda combinó términos en español e inglés, mientras que la selección de fuentes se apoyó en criterios previamente definidos, una matriz de extracción de datos y una evaluación crítica *ad hoc* centrada en cinco dominios. Los hallazgos se organizaron en cinco ejes principales, a saber: la ponderación entre el derecho al honor y la libertad de expresión; la proporcionalidad de la respuesta penal; los remedios civiles, la rectificación y las medidas anti-SLAPP; la determinación de autoría, jurisdicción y responsabilidad de los intermediarios; y los desafíos derivados de la inteligencia artificial y los *deepfakes*. En el caso peruano, se advierte la ausencia de criterios uniformes para valorar la gravedad del ataque, el interés público, la diligencia, la difusión, la prueba digital y las formas de reparación. A partir de ello, se propone un modelo escalonado que priorice la rectificación, el retiro o la corrección del contenido y la reparación civil, reservando el derecho penal para conductas graves y dolosas. Asimismo, se recomienda fortalecer la autenticación de la evidencia digital y proteger el discurso sobre asuntos de interés público.

Palabras clave: difamación, redes sociales, derecho al honor.

Abstract

This study aimed to analyze the literature on defamation on social media and to identify normative, jurisprudential, and remedial criteria that could contribute to the protection of one's reputation in Peru. To this end, a systematized legal-documentary review was conducted, based on 30 articles published between 2012 and 2026, retrieved from SciELO, Redalyc, Dialnet, DOAJ, and other platforms providing verifiable metadata. The search combined terms in both Spanish and English, while the selection of sources relied on pre-defined criteria, a data extraction matrix, and an "*ad hoc*" critical assessment focused on five domains. The findings were organized into five main thematic areas: the balancing of the right to reputation against freedom of expression; the proportionality of criminal law responses; civil remedies, rectification, and anti-SLAPP measures; the determination of authorship, jurisdiction, and intermediary liability; and the challenges posed by artificial intelligence and deepfakes. In the Peruvian context, there is a notable lack of uniform criteria for assessing the severity of an attack, the degree of public interest, the

level of diligence, the extent of dissemination, the handling of digital evidence, and the appropriate forms of reparation. Consequently, a tiered model is proposed that prioritizes rectification, the removal or correction of content, and civil redress, while reserving criminal law for serious, intentional misconduct. Furthermore, it is recommended that the authentication of digital evidence be strengthened and that discourse concerning matters of public interest be protected.

Keywords: defamation, social media, right to reputation.

Introducción

Las redes sociales han transformado profundamente la manera en que las personas se comunican. Hoy, cualquier usuario puede crear, compartir y amplificar contenidos de forma inmediata, sin los filtros tradicionales de la prensa o la radiodifusión. Si bien esta apertura favorece la participación y el debate público, también facilita que afirmaciones falsas, calificativos vejatorios o informaciones descontextualizadas se difundan con gran rapidez. La posibilidad de réplica, la permanencia de las publicaciones, su indexación en buscadores y su viralización pueden intensificar el daño reputacional y obligan a repensar el equilibrio entre el derecho al honor y la libertad de expresión en los entornos digitales (Coe, 2015; Mills, 2015; Molina, 2022).

Ante este escenario, la literatura comparada no ofrece una respuesta única. Algunos autores consideran que las categorías tradicionales de ponderación siguen siendo suficientes, siempre que se adapten al contexto tecnológico, al alcance real de la publicación y a la condición pública o privada de las personas involucradas (Molina, 2022; Cabrera & Montenegro, 2024; Ortega, 2025). Otros, en cambio, sostienen que la transnacionalidad, el anonimato, la intermediación de plataformas y la facilidad de replicación desbordan las reglas concebidas para medios convencionales (Turner, 2014; Mills, 2015; Lewis, 2015). Existe, además, una corriente crítica frente a la expansión del derecho penal, pues advierte que esta puede inhibir la expresión legítima; por ello, propone privilegiar remedios civiles, mecanismos de rectificación, umbrales de daño serio y herramientas anti-SLAPP (Spaić et al., 2016; Hanna, 2024; Bogle & Lindsay, 2024; Manitra et al., 2026).

En América Latina, esta tensión adquiere una especial relevancia, ya que el honor y la reputación cuentan con protección constitucional, al tiempo que la libertad de expresión cumple una función esencial para la vida democrática. Las experiencias de Chile, Colombia, Ecuador, México y Perú evidencian criterios dispares para valorar el interés público, la condición de personaje público, la veracidad, la diligencia, el lenguaje empleado y el remedio más adecuado (Bonilla et al., 2020; Contreras & Lovera, 2021; Jiménez & Meneses, 2023; Cabrera & Montenegro, 2024). En el Perú, la Constitución Política reconoce tanto las libertades informativas y expresivas como el derecho al honor, la reputación, la intimidad, la imagen y la rectificación (Constitución Política del Perú, 1993, art. 2, incs. 4 y 7). A su vez, el Código Penal tipifica la difamación y agrava su sanción cuando se comete mediante libro, prensa u otro medio de comunicación social (Decreto Legislativo N.º 635, 1991, art. 132). Más recientemente, la Ley N.° 32314 incorporó la inteligencia artificial, los *deepfakes* y los contenidos generados artificialmente que difunden información falsa o denigrante, con capacidad de afectar la reputación o la imagen de una persona.

La jurisprudencia reciente confirma que el marco normativo tradicional puede dialogar con la realidad digital. Así, la Casación N.° 2476-2022/Selva Central analizó publicaciones realizadas en Facebook y exigió acreditar la verdad del hecho delictivo imputado para invocar la *exceptio veritatis*. Por su parte, la Sentencia 1063/2026 del Tribunal Constitucional, recaída en el Expediente N.° 01848-2024-PA/TC, vinculó el honor con la dignidad, reafirmó la importancia de la libertad de expresión en el debate político y subrayó la necesidad de verificar la autenticidad y pertinencia de la evidencia digital. Frente a una literatura dispersa entre el derecho constitucional, penal, civil, procesal y digital, esta investigación busca identificar criterios que permitan resolver los conflictos derivados de la difamación en redes sociales. Su objetivo es integrar dichos criterios y proponer una decisión jurídicamente compatible con la Constitución y con el principio de proporcionalidad.

Metodología

Para el desarrollo de la investigación, se realizó una revisión sistematizada de carácter jurídico-documental, complementada con una síntesis cualitativa. Siguiendo la propuesta de Grant y Booth (2009), el estudio incorporó etapas explícitas de búsqueda, evaluación y síntesis, sin pretender alcanzar la exhaustividad propia de una revisión sistemática completa. Esta decisión respondió a la naturaleza del objeto de estudio, pues la investigación requería articular doctrina, análisis jurisprudenciales y estudios comparados con transparencia y reproducibilidad, adaptadas a las particularidades de la investigación jurídica (Rodríguez et al., 2024; Masbernat et al., 2025; Pichlak et al., 2025). En este sentido, la guía PRISMA 2020 orientó el reporte del proceso, sin utilizarse

para reconstruir conteos que no podían verificarse de manera independiente (Page et al., 2021). Se emplearon, además, criterios claros de selección y una matriz estandarizada para la extracción de información.

El período de búsqueda comprendió los años 2012 a 2026, lo que permitió incorporar tanto el antecedente temprano sobre difamación penal y redes sociales de Selvadurai e Islam (2012) como la reforma peruana de 2025 y las contribuciones más recientes. La búsqueda se actualizó hasta el 26 de agosto de 2026. Para delimitar el problema, se utilizó el esquema PICo, que permitió precisar la población o problema de estudio —los conflictos relacionados con el honor, la reputación o el buen nombre derivados de publicaciones digitales—, el fenómeno de interés —su tratamiento penal, civil, constitucional y procesal, junto con la responsabilidad de los intermediarios— y el contexto, entendido como las redes sociales y los entornos digitales aplicables al Perú. Esta delimitación evitó restringir la búsqueda únicamente al término "difamación".

Se consultaron SciELO, Redalyc, Dialnet, DOAJ y plataformas editoriales o repositorios institucionales que ofrecieran metadatos verificables. Se priorizaron artículos revisados por pares, con DOI y publicados en revistas trazables. Las ecuaciones de búsqueda combinaron conceptos vinculados a la difamación, el honor, la reputación, las redes sociales, la libertad de expresión, la responsabilidad y la prueba, incorporando también términos relativos a *deepfakes*, demandas estratégicas contra la participación pública, responsabilidad de intermediarios y evidencia digital. Cuando existían discrepancias entre el año, el volumen, la paginación o el DOI, se privilegió la información proporcionada por el editor o el repositorio institucional.

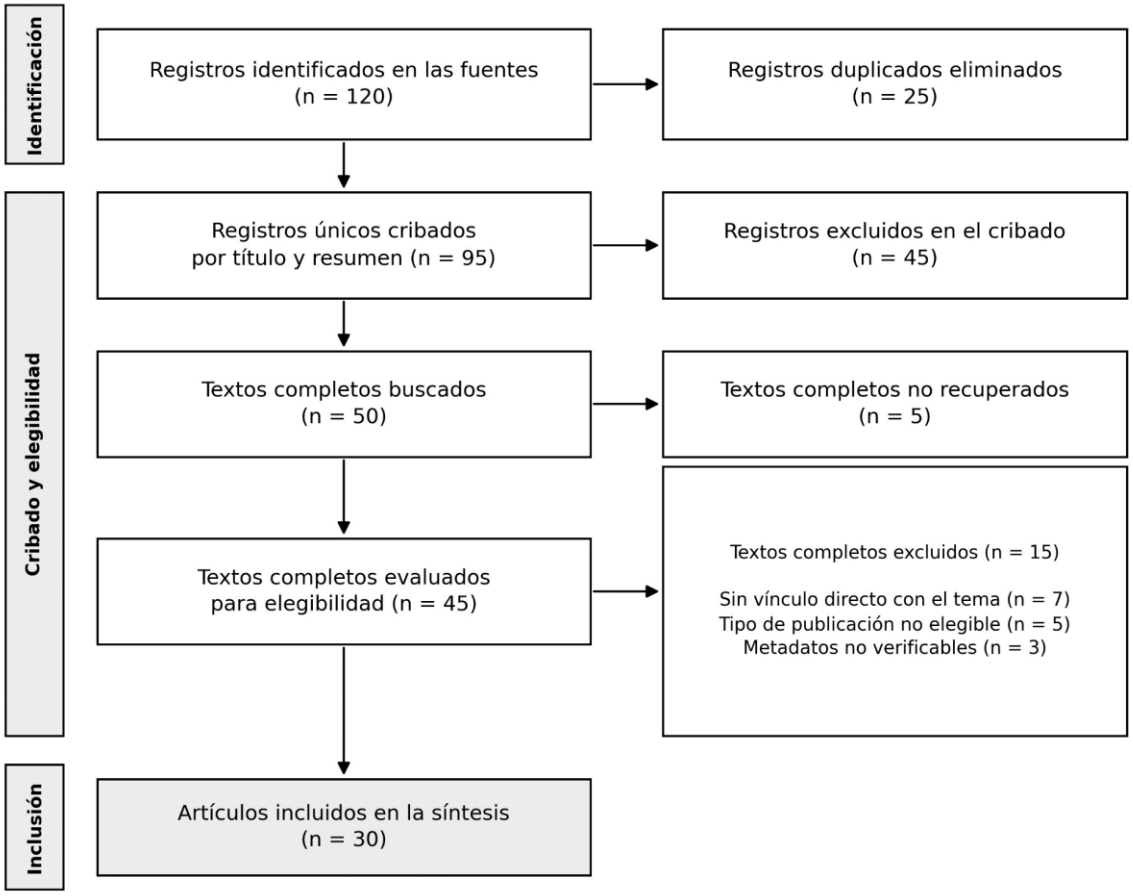
La búsqueda se realizó mediante dos ecuaciones equivalentes, en español e inglés, organizadas en tres bloques conceptuales: difamación y protección del honor, entornos digitales y mecanismos de tutela jurídica. En español, se empleó la siguiente ecuación: ("difamación" *OR* "derecho al honor" *OR* reputación) *AND* ("redes sociales" *OR* internet *OR* "plataformas digitales") *AND* ("libertad de expresión" *OR* "responsabilidad penal" *OR* "daño moral" *OR* rectificación). En inglés, se utilizó: (*defamation OR "right to reputation" OR "reputation protection"*) *AND ("social media" OR "social networking sites" OR internet*) *AND ("freedom of expression" OR "criminal liability" OR "civil remedies" OR rectification*).

Se incluyeron artículos publicados en español o inglés entre 2012 y 2026, con autoría y metadatos verificables, que abordaran la difamación, el honor, la reputación o su tensión con la libertad de expresión. Se excluyeron tesis, notas, blogs, reseñas, documentos sin trazabilidad y trabajos sin una relación directa con el tema. Aunque las normas y sentencias se utilizaron para interpretar el marco jurídico peruano, no formaron parte de los 30 artículos analizados. La unidad de análisis fue, por tanto, el artículo académico.

Para cada artículo se registraron el autor, el año, la jurisdicción, el problema abordado, las fuentes utilizadas, el criterio de ponderación, la prueba, el remedio propuesto, el interés público y el hallazgo principal. Posteriormente, se aplicó una matriz *ad hoc* de cinco dominios, que valoró el objetivo, la transparencia, la trazabilidad, la coherencia y la pertinencia o actualidad de cada estudio en una escala de 0 a 2 puntos. La síntesis fue narrativa y comparada, sin recurrir al metaanálisis, y se organizó en cinco ejes: ponderación entre derechos, proporcionalidad penal, remedios civiles y medidas anti-SLAPP, autoría e intermediarios con prueba digital, e inteligencia artificial y *deepfakes*. Los trabajos con menor robustez metodológica no fueron retenidos, y los criterios se aplicaron de manera uniforme durante todo el proceso.

En cuanto a la selección de fuentes, se identificaron inicialmente 120 registros en las bases y plataformas consultadas. Tras eliminar 25 duplicados, se cribaron 95 registros mediante la revisión de títulos y resúmenes, de los cuales se descartaron 45. Posteriormente, se buscaron 50 textos completos; sin embargo, cinco no pudieron recuperarse, por lo que se evaluaron 45 documentos para determinar su elegibilidad. En esta etapa se excluyeron 15 estudios: siete por carecer de vínculo directo con el tema, cinco por corresponder a tipos de publicación no elegibles y tres por presentar metadatos no verificables. Finalmente, se conservaron 30 artículos, que fueron sometidos a evaluación crítica, extracción de información y síntesis narrativa comparada.

Figura 1
Secuencia analítica de la revisión sistematizada



Fuente. Elaboración propia

Resultados

El corpus estuvo conformado por 30 artículos con metadatos verificables y una pertinencia jurídica directa con el objeto de estudio. Se observó un aumento progresivo de las publicaciones a partir de 2020, lo que puede explicarse por la creciente incidencia de los conflictos reputacionales en plataformas digitales y por la aparición de tecnologías capaces de generar contenidos sintéticos. De los estudios incluidos, veinte presentaron una robustez metodológica alta y diez una robustez moderada. La cobertura geográfica abarcó América Latina, Europa, África, Asia y Oceanía, lo que permitió contrastar cuestiones como la jurisdicción aplicable, la responsabilidad de los intermediarios, el estándar de daño serio, la despenalización y los mecanismos anti-SLAPP con los marcos jurídicos latinoamericanos.

En cuanto a la relación entre el honor y la libertad de expresión, los estudios coinciden en que ninguno de estos derechos es absoluto. La resolución de los conflictos exige considerar la naturaleza del discurso, el interés público involucrado, la condición de la persona afectada, el contexto de publicación y la distinción entre afirmaciones de hecho y valoraciones u opiniones. Bonilla et al. (2020), Contreras y Lovera (2021), Jiménez y Meneses (2023) y Cabrera y Montenegro (2024) coinciden en que la simple incomodidad provocada por una crítica no es suficiente para restringir la libertad de expresión. En el caso peruano, esto implica que las decisiones deben motivarse con suficiencia y evitar razonamientos automáticos; a mayor interés público, mayor debe ser la exigencia para justificar una limitación.

Tabla 1
Matriz sintética de los 30 estudios incluidos

Autor(es), año y ámbito	Enfoque	Aporte principal	Valoración
Selvadurai & Islam (2012) Australia / comparado	Doctrinal penal	Reexamina la difamación penal frente a comentarios difundidos en redes sociales y advierte la tensión entre el daño reputacional y el efecto inhibidor que puede generar la sanción penal.	Alta
Turner (2014) Australia / Reino Unido / NZ / Hong Kong	Análisis jurisprudencial comparado	Identifica la incertidumbre existente respecto de la publicación por omisión y la responsabilidad de los operadores de plataformas, proponiendo defensas legales para intermediarios que no son autores primarios del contenido.	Alta
Roos & Slabbert (2014) Sudáfrica	Comentario jurisprudencial	Analiza la primera indemnización concedida en Sudáfrica por comentarios difamatorios publicados en Facebook y resalta la importancia de etiquetar, retirar y rectificar contenidos lesivos.	Alta
Coe (2015) Reino Unido	Doctrinal	Expone la paradoja entre la expansión de la libertad de expresión en las redes sociales y el incremento de la exposición de los usuarios a sanciones penales, alertando sobre los riesgos de una sobrerregulación.	Alta
Mills (2015) Reino Unido / internacional	Derecho internacional privado	Demuestra que la difamación transfronteriza en redes sociales plantea problemas específicos relacionados con la ley aplicable, el foro competente y la delimitación del público afectado.	Alta
Lewis (2015) Jamaica / comparado	Análisis normativo	Examina cómo las características globales de las redes sociales intensifican el riesgo de responsabilidad y condicionan la aplicación de la Defamation Act de Jamaica.	Moderada
Spaić, Nolasco & Novovic (2016) Unión Europea / Balcanes	Comparado	Evalúa los alcances de la criminalización y la despenalización de la difamación, sosteniendo que la protección de la reputación puede garantizarse sin recurrir necesariamente a la privación de libertad.	Alta
Gil (2016) España	Doctrinal	Analiza el honor, la intimidad y la imagen en redes sociales, y subraya la necesidad de ponderar el contexto, el consentimiento y el alcance de la difusión digital.	Moderada
Cantoral & Jiménez (2018) México	Doctrinal y jurisprudencial	Contrasta el daño moral con el derecho de réplica y propone respuestas rápidas y específicas frente a contenidos lesivos difundidos en internet.	Moderada
Alarcón (2020) Perú	Estudio de caso jurisprudencial	Muestra que la resolución del conflicto entre honor y libertad de expresión exige un razonamiento judicial explícito sobre la veracidad, el interés público y la proporcionalidad de la medida restrictiva.	Moderada
Bonilla, Vergara & Santamaría (2020) Ecuador	Análisis jurídico	Propone un test tripartito y una perspectiva sistémica del entorno digital para determinar la antijuridicidad y el daño moral derivados de publicaciones en redes sociales.	Alta
Cantoral (2020) México / comparado	Revisión jurídica	Destaca la extraterritorialidad, la reparación integral y el daño moral como vías relevantes para enfrentar las afectaciones al honor en redes sociales.	Alta

Vásconez & López (2020) Ecuador	Doctrinal	Examina los límites civiles y penales de la libertad de expresión y vincula el principio de proporcionalidad con la consolidación de una democracia deliberativa.	Moderada
Contreras & Lovera (2021) Chile	Análisis jurisprudencial	Observa que la Corte Suprema suele ordenar el retiro de publicaciones, y critica que ello puede reflejar una protección insuficiente de la libertad de expresión, además de plantear dificultades vinculadas al *doxing*.	Alta
Rudas (2021) Perú	Doctrinal	Relaciona las redes sociales, la inteligencia artificial, el anonimato y la tutela del honor, proponiendo responsabilidades preventivas y límites frente a expresiones lesivas.	Moderada
Aditya & Al-Fatih (2021) Indonesia	Constitucional y jurisprudencial	Examina la criminalización de usuarios mediante normas sobre internet y advierte sus riesgos para la opinión y la expresión constitucionalmente protegidas.	Alta
Molina (2022) España	Constitucional	Concluye que las redes sociales requieren modular el análisis jurídico clásico, sin que ello implique necesariamente crear un derecho al honor enteramente nuevo.	Alta
Vidal (2022) España	Comentario constitucional	Analiza la primera respuesta del Tribunal Constitucional frente al conflicto entre honor y libertad de expresión en redes sociales, así como el debate sobre la necesidad de un canon digital específico.	Alta
Díaz (2023) Perú	Doctrinal constitucional	Sistematiza el contenido y los límites de la libertad de expresión en redes sociales, considerando fenómenos como la cultura de cancelación, los trolls y la intolerancia digital.	Moderada
Jiménez & Meneses (2023) Colombia	Revisión constitucional	Identifica criterios de prevalencia, colisión y ponderación entre la libertad de expresión, la intimidad y el buen nombre en internet y redes sociales.	Alta
Setiadi (2023) Indonesia	Investigación jurídica normativa	Señala problemas de vaguedad normativa, capacidad de ejecución y obtención de evidencia electrónica en los casos de difamación cometida a través de redes sociales.	Moderada
Hasbi et al. (2023) Indonesia	Jurídico-normativo	Contrasta la difamación digital con los derechos humanos y el derecho penal, cuestionando la adecuación de una respuesta penal amplia frente a este tipo de conductas.	Moderada
Bennett (2023) Inglaterra y Gales	Doctrinal y jurisprudencial	Cuestiona la distinción rígida entre hecho y opinión en materia de difamación y advierte que su aplicación mecánica puede inhibir el debate público.	Alta
Cabrera & Montenegro (2024) Colombia	Revisión documental y jurisprudencial	Sostiene que la libertad de expresión en redes sociales conserva una protección reforzada, por lo que la remoción de contenidos exige un análisis contextual y no puede resolverse de manera automática.	Alta
Miró (2024) Perú	Doctrinal y comparado	Examina las querellas por difamación presentadas contra denuncias públicas de violencia y sistematiza criterios para identificar el discurso protegido y las demandas estratégicas contra la participación pública.	Alta
Kumar (2024) India / internacional	Doctrinal	Expone los desafíos derivados de los *deepfakes*, los bots, la jurisdicción y la atribución de responsabilidad en la difamación digital contemporánea.	Moderada

Hanna (2024) Reino Unido / comparado	Análisis normativo comparado	Propone mecanismos de desestimación temprana para las demandas estratégicas contra la participación pública, especialmente en litigios de difamación que involucran asuntos de interés público.	Alta
Bogle & Lindsay (2024) Inglaterra y Gales	Análisis de 85 casos	Señala que el umbral de daño serio refuerza la materialidad del perjuicio, aunque su complejidad puede limitar su utilidad como herramienta procesal frente a las demandas estratégicas.	Alta
Ortega (2025) España	Análisis jurisprudencial constitucional	Sistematiza la doctrina reciente y los factores propios del entorno digital que deben considerarse para resolver los conflictos entre las libertades comunicativas y el derecho al honor.	Alta
Manitra et al. (2026) Indonesia / comparado	Doctrinal comparado	Propone priorizar los remedios civiles, establecer un umbral de daño serio, reconocer excepciones vinculadas al interés público y reservar la respuesta penal para los casos más extremos.	Alta

Fuente. Elaboración propia a partir del corpus revisado. La valoración corresponde a la matriz crítica de cinco dominios descrita en la metodología; los trabajos de baja robustez no fueron retenidos

Ponderación entre honor y libertad de expresión

El entorno digital no exige crear un canon jurídico completamente autónomo. Molina (2022) y Ortega (2025) proponen trasladar las reglas clásicas de ponderación al contexto digital, atendiendo a factores como la audiencia alcanzada, la permanencia del contenido, su viralización y la arquitectura de la plataforma. Una publicación en Facebook, por ejemplo, no puede asimilarse a un comentario privado, pero tampoco justifica por sí sola la aplicación de un régimen penal especial. Estos elementos inciden, además, en la determinación del alcance del daño y en la reparación correspondiente.

Proporcionalidad del derecho penal y efecto inhibidor

El análisis de la dimensión penal evidencia una tensión relevante. Selvadurai e Islam (2012) reconocen que la amplificación propia de las redes puede renovar la relevancia práctica de la difamación penal. No obstante, Coe (2015), Spaić et al. (2016), Aditya y Al-Fatih (2021), Hasbi et al. (2023) y Manitra et al. (2026) advierten que las normas vagas o las penas desproporcionadas pueden inhibir críticas legítimas y afectar el debate público. Por ello, la síntesis propone aplicar el principio de *ultima ratio*: el derecho penal debería reservarse para imputaciones fácticas gravemente lesivas, realizadas con conocimiento de su falsedad o con dolo, de alcance relevante y sin protección de interés público. En consecuencia, no se plantea una despenalización automática de la difamación.

Remedios civiles, rectificación y mecanismos anti-SLAPP

Los remedios civiles, la rectificación y las medidas correctivas pueden resultar más eficaces que un proceso penal prolongado para restaurar la reputación afectada (Cantoral & Jiménez, 2018; Cantoral, 2020; Bonilla et al., 2020). Hanna (2024), Miró (2024) y Bogle y Lindsay (2024) destacan, además, la importancia de la desestimación temprana de demandas infundadas, la protección frente a litigios estratégicos y la exigencia de un daño serio como umbral de tutela. Para el Perú, esto supone proteger a las personas frente al uso instrumental de los procesos judiciales, sin dejar de brindar protección efectiva ante daños graves. La rapidez de la respuesta, su visibilidad y su capacidad para alcanzar una audiencia equivalente a la de la publicación original resultan determinantes para su eficacia.

Autoría, jurisdicción, intermediarios y prueba digital

En relación con la autoría, la jurisdicción y la responsabilidad de los intermediarios, Turner (2014), Lewis (2015) y Mills (2015) recomiendan examinar la publicación primaria, los riesgos derivados de la transnacionalidad, la ley aplicable, el conocimiento de los contenidos, el grado de control ejercido y la capacidad de respuesta de cada actor. Asimismo, Setiadi (2023) y Kumar (2024) advierten que una captura de pantalla no acredita, por sí sola, la autoría, la integridad, la fecha, el alcance ni el contexto de una publicación. Por ello, resulta necesario documentar la URL, la fecha y hora de acceso, los metadatos disponibles y los archivos originales, recurriendo a

pericia técnica cuando corresponda. La atribución de responsabilidad exige distinguir entre la constatación del contenido y la determinación de su autor.

Inteligencia artificial, *deepfakes* y manipulación de contenido

La inteligencia artificial y los *deepfakes* permiten crear voces, imágenes o videos con una apariencia cada vez más auténtica, lo que incrementa los riesgos para el honor y la reputación de las personas. Rudas (2021), Kumar (2024) y Manitra et al. (2026) coinciden en señalar la gravedad de estos supuestos. En el Perú, la Ley N.° 32314, al modificar el Código Penal (Decreto Legislativo N.º 635, 1991, art. 132), exige acreditar la atribución, la autenticidad, el dolo, la causalidad y la proporcionalidad frente a contenidos sintéticos. Por ello, los mecanismos de autenticación de la evidencia digital deben reforzarse especialmente en estos casos.

Tabla 2

Síntesis temática y consecuencia para el ordenamiento peruano

Eje	Criterios recurrentes	Implicación para el Perú
Ponderación constitucional	Interés público; condición de la persona; hecho/opinión; contexto; necesidad de expresiones ofensivas	Evitar automatismos: ni toda crítica fuerte es difamación ni toda publicación digital queda amparada por libertad de expresión.
Intervención penal	Gravedad; falsedad/dolo; alcance; daño; disponibilidad de remedios menos restrictivos	Aplicar *ultima ratio* y motivar por qué una respuesta civil o rectificatoria sería insuficiente.
Reparación y proceso	Rectificación; retiro/corrección; daño moral; serious harm; anti-SLAPP	Priorizar medidas capaces de detener o reparar el daño sin generar un efecto inhibidor desproporcionado.
Prueba digital	Autoría; integridad; fecha; URL; metadatos; alcance; contexto; manipulación	Separar constatación del contenido de atribución al autor y usar pericia cuando autenticidad o edición sean controvertidas.
IA y plataformas	*Deepfakes*; bots; rol del intermediario; conocimiento/control; respuesta al aviso	Aplicar la Ley N.° 32314 sin presumir responsabilidad por mera intermediación y con especial cuidado en la autenticación del contenido.

Discusión

La respuesta jurídica frente a la difamación en redes sociales debe centrarse en la calidad de los criterios aplicados, más que en la mera ampliación de las sanciones disponibles. En el Perú, el Código Penal ya comprende los medios de comunicación social (Decreto Legislativo N.º 635, 1991, art. 132), la práctica judicial ha conocido casos derivados de publicaciones en Facebook y la Ley N.° 32314 ha incorporado la inteligencia artificial y los *deepfakes*. Sin embargo, persiste la necesidad de uniformar la tutela penal, la valoración del interés público, la prueba de la autoría y del alcance de la difusión, así como la adopción de remedios proporcionales. La aplicación del derecho debe evitar, por un lado, subestimar el daño que puede generar la viralización y, por otro, recurrir de manera automática a respuestas punitivas.

En esta línea, Molina (2022) y Ortega (2025) coinciden en que internet modifica el contexto en el que se producen los conflictos, pero no altera los principios fundamentales de ponderación. El número de seguidores, la configuración de la cuenta, las republicaciones, la permanencia del contenido, la posibilidad de corrección y la arquitectura de amplificación de las plataformas no convierten por sí mismas una expresión en ilícita; no obstante, sí inciden en la gravedad del daño, en la causalidad y en el remedio más adecuado. El modelo peruano debería explicitar estos factores para evitar tanto la trivialización del perjuicio como la criminalización indiscriminada de la expresión. Si bien la persistencia y la indexación de los contenidos pueden agravar el daño, no sustituyen el examen del contenido ni de las circunstancias en que este fue difundido.

La distinción entre hechos y valoraciones tampoco puede aplicarse de forma rígida. Bennett (2023) advierte que hacerlo puede terminar penalizando críticas legítimas sobre asuntos de interés público. En el ámbito peruano, la Casación N.° 2476-2022/Selva Central exigió sustento probatorio para imputar una estafa, mientras que la Sentencia 1063/2026 del Tribunal Constitucional, recaída en el Expediente N.° 01848-2024-PA/TC, reafirmó la

protección de la libertad de expresión en el debate político. Esta última decisión también llamó la atención sobre deficiencias en la autenticación de impresiones digitales, en la transcripción de contenidos almacenados en dispositivos USB y en la pertinencia de la evidencia ofrecida.

La intervención penal debe mantenerse como una medida excepcional, debido a su potencial efecto inhibidor sobre la expresión legítima (Coe, 2015; Spaić et al., 2016; Manitra et al., 2026). Aun así, las campañas dolosas, masivas o basadas en contenidos sintéticos pueden provocar daños especialmente intensos. Por ello, cuando se acuda a la vía penal, se requiere una motivación reforzada sobre la necesidad y proporcionalidad de la medida: el juez debe explicar por qué el caso supera lo que puede repararse mediante mecanismos menos gravosos.

La rectificación reconocida en la Constitución Política del Perú (1993, art. 2, incs. 4 y 7) permite corregir información inexacta, aunque, en el entorno digital, su eficacia depende de la visibilidad, el alcance y la oportunidad con que se difunda. En algunos supuestos, el retiro del contenido, una aclaración o la reparación civil pueden resultar suficientes; en otros, la gravedad del ataque, el dolo o la reiteración pueden justificar medidas adicionales. Hanna (2024) y Miró (2024) respaldan la incorporación de mecanismos anti-SLAPP, capaces de identificar demandas instrumentales sin bloquear reclamos legítimos. Una rectificación tardía o de escasa difusión puede ser formalmente correcta, pero materialmente ineficaz para reparar el daño.

Finalmente, la prueba digital no puede equipararse a una simple captura de pantalla. Su valoración exige acreditar el contenido, el momento y lugar de la publicación, el control de la cuenta y el alcance real de la difusión. Ante hipótesis de manipulación, suplantación o contenidos generados artificialmente, la pericia técnica resulta central, especialmente tras la entrada en vigencia de la Ley N.° 32314. No obstante, estas conclusiones deben leerse con cautela, pues el estudio se apoya en fuentes públicas y en un corpus heterogéneo. Se requieren investigaciones adicionales sobre decisiones judiciales peruanas, autenticación de evidencia digital, mecanismos anti-SLAPP, estándares de daño serio y rectificación digital. Cualquier mejora normativa o jurisprudencial debería, en todo caso, ser compatible con la participación pública y con las garantías del debido proceso.

Conclusiones

La revisión permite identificar un conjunto de criterios útiles para resolver los conflictos derivados de la difamación en redes sociales. Entre ellos destacan la naturaleza fáctica u opinativa del contenido, el interés público involucrado, la condición de la persona afectada, la diligencia empleada, la intención del autor, el contexto y alcance digital de la publicación, la gravedad del daño, la integridad de la prueba y la proporcionalidad del remedio adoptado. Aplicar estos elementos de manera conjunta ofrece mayor previsibilidad que decidir únicamente a partir de la existencia de una publicación en una red social.

El Perú no carece de regulación frente a estos conflictos. La Constitución Política protege tanto la libertad de expresión como el derecho al honor y la reputación (Constitución Política del Perú, 1993, art. 2, incs. 4 y 7). Asimismo, el Código Penal contempla la difamación cometida a través de medios de comunicación social (Decreto Legislativo N.º 635, 1991, art. 132), mientras que la Ley N.° 32314 incorporó la inteligencia artificial, los *deepfakes* y los contenidos generados mediante estas tecnologías. Aun así, la respuesta penal debe reservarse para casos graves, dolosos y debidamente acreditados. En tales supuestos, corresponde explicar por qué la rectificación, el retiro, la corrección o la reparación civil resultan insuficientes, especialmente cuando está en juego el discurso sobre asuntos de interés público, la denuncia ciudadana, el periodismo o la fiscalización social.

La principal contribución de esta investigación consiste en integrar criterios constitucionales, penales, civiles y probatorios que permitan diferenciar entre publicaciones ordinarias, hostigamiento reputacional y contenidos sintéticos. Para fortalecer esta tutela, se requiere desarrollar jurisprudencia empírica, establecer protocolos de preservación y autenticación de evidencia digital frente a *deepfakes* y suplantaciones, y evaluar la incorporación de mecanismos anti-SLAPP y umbrales de daño serio. Estas herramientas deben proteger simultáneamente la participación pública y el derecho al honor. Su eficacia dependerá, en última instancia, de criterios explícitos y de remedios escalonados, capaces de favorecer decisiones coherentes y reparaciones ajustadas a las circunstancias de cada caso.

Referencias

Aditya, Z. F., & Al-Fatih, S. (2021). Indonesian constitutional rights: Expressing and purposing opinions on the internet. *The International Journal of Human Rights, 25*(9), 1395–1419. https://doi.org/10.1080/13642987.2020.1826450

Alarcón Requejo, G. (2020). Libertad de expresión y derecho al honor en el razonamiento judicial. Análisis a partir de un estudio de caso. *Ius Inkarri, 9*(9), 89–114. https://revistas.urp.edu.pe/challenge?next=%2Findex.php%2FInkarri%2Farticle%2Fview%2F3707

Bennett, T. D. C. (2023). Interpretation is opinion: Realigning the fact/opinion distinction in English defamation law. *Journal of Media Law, 15*(1), 62–89. https://doi.org/10.1080/17577632.2023.2216523

Bogle, S., & Lindsay, B. (2024). Serious harm: Six lessons since Lachaux. *Journal of Media Law, 16*(2), 417–439. https://doi.org/10.1080/17577632.2024.2421061

Bonilla Manotoa, E. S., Vergara Caicedo, M. V., & Santamaría Viteri, C. (2020). La honra versus la libertad de expresión en redes sociales: Mecanismo de determinación de daño moral ante la colisión de derechos humanos. *USFQ Law Review, 7*(1), 183–201. https://doi.org/10.18272/ulr.v7i1.1680

Cabrera, K. I., & Montenegro, Y. A. (2024). Libertad de expresión en redes sociales: Estudio de sentencias de la Corte Constitucional colombiana. *Prolegómenos, 27*(54), 29–43. https://doi.org/10.18359/prole.6227

Cantoral Domínguez, K. (2020). Daño moral en redes sociales: Su tratamiento procesal en el derecho comparado. *Revista IUS, 14*(46), 163–182. https://doi.org/10.35487/rius.v14i46.2020.519

Cantoral Domínguez, K., & Jiménez Vázquez, C. K. (2018). Protección del honor frente a expresiones divulgadas en internet: ¿Daño moral o derecho de réplica? *Ecos Sociales, 6*(18). https://revistaecos.ujat.mx/ecosoc/es/article/view/2591

Coe, P. (2015). The social media paradox: An intersection with freedom of expression and the criminal law. *Information & Communications Technology Law, 24*(1), 16–40. https://doi.org/10.1080/13600834.2015.1004242

Constitución Política del Perú. (1993, 29 de diciembre). *Diario Oficial El Peruano*.

Contreras Vásquez, P., & Lovera Parmo, D. (2021). Redes sociales, funas, honor y libertad de expresión: Análisis crítico de los estándares de la jurisprudencia de la Corte Suprema chilena. *Derecho PUCP*, (87), 345–371. https://doi.org/10.18800/derechopucp.202102.010

Corte Suprema de Justicia de la República, Sala Penal Permanente. (2025, 24 de abril). *Casación N.º 2476-2022/Selva Central*. https://img.lpderecho.pe/wp-content/uploads/2025/05/Casacion-2476-2022-Selva-Central-LPDerecho-1.pdf

Decreto Legislativo N.º 635, Código Penal. (1991, 8 de abril). *Diario Oficial El Peruano*. Artículo 132, modificado por la Ley N.º 32314 (2025). https://www.refworld.org/es/leg/decre/pejec/1991/es/132326

Díaz Giunta, R. (2023). El derecho a la libertad de expresión y las redes sociales. *Athina*, (15), 91–114. https://alicia.concytec.gob.pe/vufind/Record/RULI_28e458bd9f2ae1ec84ef76d2afd445d2/Details

Gil Vallilengua, L. (2016). Los derechos al honor, a la intimidad y a la propia imagen en las redes sociales: La difusión no consentida de imágenes. *Revista Electrónica de Derecho de la Universidad de La Rioja*, (14), 161–190. https://doi.org/10.18172/redur.4153

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Hanna, M. (2024). SLAPPs: What are they? And how should defamation law be reformed to address them? *Journal of Media Law, 16*(1), 118–145. https://doi.org/10.1080/17577632.2024.2345982

Hasbi, Y., Adwani, Rizanizarli, & Rinaldi, Y. (2023). Defamation and insult via digital media in Indonesia: The Islamic law and human rights perspective. *Journal of Law and Sustainable Development, 11*(7), Artículo e1016. https://doi.org/10.55908/sdgs.v11i7.1016

Jiménez, W. G., & Meneses Quintana, O. (2023). Libertad de expresión en internet y redes sociales vs. derechos a la intimidad y el buen nombre. Prevalencia, colisión y ponderación en el derecho constitucional (1992–2019). *Revista Derecho del Estado*, (56), 275–304. https://doi.org/10.18601/01229893.n56.10

Kumar, S. (2024). Online defamation in the digital age: Issues and challenges with particular reference to deepfakes and malicious bots. *International Journal of Law and Policy, 2*(8), 32–41. https://doi.org/10.59022/ijlp.200

Lewis, C. (2015). Social media: Cyber trap door to defamation. *Masaryk University Journal of Law and Technology, 9*(1), 65–84. https://doi.org/10.5817/MUJLT2015-1-5

Ley N.º 32314, Ley que modifica el Código Penal, Decreto Legislativo 635, y la Ley 30096, Ley de Delitos Informáticos, para incluir el uso de la inteligencia artificial en la comisión de delitos. (2025, 29 de abril). *Diario Oficial El Peruano*. https://busquedas.elperuano.pe/dispositivo/NL/2394851-2

Manitra, R. R. M., Prabandari, A. P., Jibril, A. M., & Hossain, A. (2026). A proposal for decriminilisation of online defamation in Indonesia: Towards a human rights-based approach. *Cogent Social Sciences, 12*(1), Artículo 2613959. https://doi.org/10.1080/23311886.2026.2613959

Masbernat, P., Andrés-Aucejo, E., & Ramos-Fuentes, G. (2025). La revisión de literatura y las investigaciones en derecho. *Revista Pedagogía Universitaria y Didáctica del Derecho, 12*(1), 57–80. https://doi.org/10.5354/0719-5885.2025.77833

Mills, A. (2015). The law applicable to cross-border defamation on social media: Whose law governs free speech in 'Facebookistan'? *Journal of Media Law, 7*(1), 1–35. https://doi.org/10.1080/17577632.2015.1055942
Miró Quesada Gayoso, J. (2024). De víctimas a victimarias: El delito de difamación y las denuncias públicas por violencia de género. *IUS ET VERITAS*, (69), 168–191. https://doi.org/10.18800/iusetveritas.202402.010
Molina Martínez, L. (2022). Honor y libertad de expresión en las redes sociales. *Derecho Privado y Constitución*, (41), 227–276. https://doi.org/10.18042/cepc/dpc.41.01
Ortega Gutiérrez, D. (2025). Las libertades comunicativas del art. 20.1 CE frente al derecho al honor en las redes sociales. Nueva doctrina del Tribunal Constitucional. *Teoría y Realidad Constitucional*, (56), 419–461. https://doi.org/10.5944/trc.56.2025.47080
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Artículo n71. https://doi.org/10.1136/bmj.n71
Pichlak, M., Kaleta, K. J., & Nawacka, A. (2025). Systematic literature review method in legal research: Prospects and challenges. *Law and Method*, 1–23. https://doi.org/10.5553/REM/.000085
Rodríguez Cairo, V., Vilchez Olivares, P. A., & Obando Peralta, E. C. (2024). Revisión sistemática de literatura científica aplicada a la investigación jurídica. *Revista Pedagogía Universitaria y Didáctica del Derecho, 11*(1), 63–91. https://doi.org/10.5354/0719-5885.2024.70653
Roos, A., & Slabbert, M. (2014). Defamation on Facebook: Isparta V Richter 2013 6 SA 529 (GP). *Potchefstroom Electronic Law Journal, 17*(6), 2845–2868. https://doi.org/10.4314/pelj.v17i6.18
Rudas Murga, C. (2021). Redes sociales: Inteligencia artificial en el derecho al honor desde una perspectiva peruana. *Lucerna Iuris et Investigatio*, (1), 99–110. https://doi.org/10.15381/lucerna.v0i1.20137
Selvadurai, N., & Islam, M. R. (2012). The application of criminal defamation to inflammatory comments made on social networking sites: A new role for an old law? *Criminal Law Journal, 36,* 38–49. https://researchers.mq.edu.au/en/publications/the-application-of-criminal-defamation-to-inflammatory-comments-m/
Setiadi, A. S. (2023). The freedom of opinion expression through social media and the impact of acts of defamation to the perpetrator. *Global Legal Review, 3*(2), 127–140. https://doi.org/10.19166/glr.v3i2.6738
Spaić, A., Nolasco, C., & Novovic, M. (2016). Decriminalization of defamation – The Balkans case: A temporary remedy or a long term solution? *International Journal of Law, Crime and Justice, 47*, 21–30. https://doi.org/10.1016/j.ijlcj.2016.05.002
Tribunal Constitucional del Perú, Sala Segunda. (2026, 19 de junio). *Sentencia 1063/2026, Expediente N.º 01848-2024-PA/TC*. https://www.tc.gob.pe/jurisprudencia/2026/01848-2024-AA.html
Turner, R. J. (2014). Internet defamation law and publication by omission: A multi-jurisdictional analysis. *UNSW Law Journal, 37*(1), 34–62. https://classic.austlii.edu.au/au/journals/UNSWLawJl/2014/2.html
Vásconez, V., & López, E. (2020). Cimientos de la libertad de expresión en internet: Limitaciones desde la esfera civil y penal. *Revista de Estudios Jurídicos Cálamo*, (13), 98–110. https://doi.org/10.61243/calamo.13.164
Vidal Marín, T. (2022). El conflicto entre la libertad de expresión y el derecho al honor en las redes sociales. Vulneración del derecho al honor de un torero fallecido: El caso de Víctor Barrio Hernanz (STC 93/2021). *Revista Anuario Parlamento y Constitución*, (23), 469–485. https://doi.org/10.71206/rapc.30

CONTRIBUCIÓN DE LOS AUTORES:

- **Juan Carlos Alcántara Farfán:** conceptualización, metodología, investigación, supervisión y administración del proyecto.
- **Nelly Patricia Saravia Yataco:** investigación, curación de datos, validación y redacción del borrador original.
- **Joaquín Sebastián Alcántara Saravia:** curación de datos, análisis formal, visualización y redacción del borrador original. Los tres autores participaron en la redacción mediante revisión y edición.

Financiamiento: Los autores declaran que la investigación se desarrolló sin recibir apoyo financiero.

Conflicto de intereses: Los autores manifiestan que no tienen conflictos de intereses relacionados con la publicación de este artículo.