COLLOQUIA

*Revista de Pensamiento y Cultura Vol. **12** (2025), págs. 71 - 105. ISSN: 1390-8731*
DOI 10.31207/colloquia.v12i1.188

IA generativa para la elaboración de dictámenes en psicología forense: Desafíos éticos, técnicos y jurídicos

Generative AI for the preparation of reports in forensic psychology: Ethical, technical, and legal challenges

Guillermo Ramírez Zavala
guillermormzzavala0@gmail.com
Red Internacional de Estudios de los Problemas Sociales Multidisciplinarios y Multisectoriales del Análisis de la Criminalidad (REPSMAC)
ORCID ID 0009-0003-4179-8303

Resumen: La introducción de la inteligencia artificial generativa (IA Gen) en la psicología forense promete agilizar la elaboración de dictámenes periciales y aportar nuevas herramientas de análisis en el sistema jurídico mexicano. Este artículo examina (1) la viabilidad de aplicar IA Gen en dictámenes psicológicos forenses –por ejemplo, como apoyo en la redacción de informes o en el procesamiento de grandes volúmenes de datos–, (2) los riesgos éticos asociados –como sesgos algorítmicos, opacidad, posibles vulneraciones a derechos fundamentales y dilemas bioéticos–, y (3) las implicaciones jurídicas y normativas –incluyendo el marco legal mexicano vigente (Ley General de Ciencia, Tecnología e Innovación 2023) y la necesidad de lineamientos específicos. Se integran hallazgos de literatura científica reciente (2023–2025) y experiencias comparadas en Colombia, España y EE. UU. para contextualizar el debate. A la luz de los principios bioéticos de autonomía, beneficencia, no maleficencia, justicia y conceptos de justicia algorítmica, se realiza un análisis crítico del uso de IA Gen en peritajes forenses. Finalmente, se propone un conjunto de lineamientos técnicos y éticos para un uso responsable de estas tecnologías en la prueba pericial psicológica. Las conclusiones subrayan la importancia de equilibrar innovación tecnológica con salvaguardas éticas y jurídicas,

Recepción: 07-09-2025 Aceptación: 08-12-2025 Publicación electrónica: 20-12-2025

para aprovechar las ventajas de la IA Gen sin comprometer derechos ni la integridad del proceso judicial.

Palabras clave: Inteligencia artificial generativa; psicología forense; ética; evidencia pericial; sistema jurídico mexicano.

Abstract: The introduction of generative artificial intelligence (GenAI) in forensic psychology holds the potential to streamline the preparation of expert reports and provide new analytical tools within the Mexican legal system. This article examines (1) the feasibility of applying GenAI in forensic psychological assessments—for instance, as support in report writing or processing large volumes of data; (2) the associated ethical risks—such as algorithmic bias, opacity, potential violations of fundamental rights, and bioethical dilemmas; and (3) the legal and regulatory implications—including the current Mexican legal framework (General Law of Science, Technology and Innovation 2023, CNPP) and the need for specific guidelines. Recent scientific literature (2023–2025) and comparative experiences in Colombia, Spain, and the United States are integrated to contextualize the debate. In light of the bioethical principles of autonomy, beneficence, non-maleficence, and justice, as well as concepts of algorithmic justice, a critical analysis of GenAI use in forensic assessments is carried out. Finally, a set of technical and ethical guidelines is proposed for the responsible use of these technologies in psychological expert evidence. The conclusions emphasize the importance of balancing technological innovation with ethical and legal safeguards in order to harness the benefits of GenAI without compromising rights or the integrity of the judicial process.

Keywords: Generative artificial intelligence; forensic psychology; ethics; expert evidence; Mexican legal system.

Introducción

En años recientes la inteligencia artificial generativa (IA Gen), ejemplificada por modelos de lenguaje como *ChatGPT*, ha irrumpido en numerosos campos

profesionales –incluido el ámbito legal– suscitando entusiasmo e inquietud a partes iguales. En el contexto de la psicología forense, que combina saber psicológico y jurídico, la IA Gen podría influir en la elaboración de dictámenes periciales presentados ante tribunales. Un dictamen de psicología forense es la opinión técnica que emite un perito psicólogo, tras evaluar a personas o analizar hechos relevantes en un proceso judicial (p. ej., valorar el estado mental de un acusado, la credibilidad de un testimonio o el daño psicológico en una víctima). Dicha prueba pericial suele ser compleja y determinante en procedimientos penales, familiares o laborales. Por ello, resulta crucial examinar cómo la utilización de IA generativa podría transformar la práctica pericial forense, tanto en sus potenciales beneficios como en sus riesgos.

La viabilidad tecnológica de la IA Gen para asistir en tareas periciales es cada vez mayor, estos modelos pueden *generar texto coherente, resumir documentos extensos, traducir idiomas, identificar patrones* en datos no estructurados, e incluso producir imágenes o voz sintética. En teoría, un sistema de IA Gen entrenado con información psicológica y legal podría ayudar a redactar informes periciales, analizar resultados de tests psicológicos o simular escenarios clínico-forenses. De hecho, ya se emplean algoritmos en la práctica forense: psicólogos y criminólogos utilizan instrumentos actuariales de evaluación de riesgo para estimar probabilidad de reincidencia u otros factores con peso probatorio (Baz, 2023).

La IA Gen representaría un paso adelante al procesar gran cantidad de datos heterogéneos (entrevistas, historial, pruebas psicométricas) y generar un borrador de dictamen o insights útiles para el perito. Un ejemplo ilustrativo es el concepto de *modelos multimodales generativos* capaces de integrar audio, video y texto de evaluaciones psiquiátricas para apoyar diagnósticos e incluso asistir en la producción de informes clínicos (Tortora, 2024). Estas posibilidades anuncian mayor eficiencia y profundidad en los peritajes.

Ahora bien, la adopción de IA Gen en la justicia plantea interrogantes éticas y jurídicas de primer orden. Un dictamen pericial incide en derechos fundamentales (vida, libertad, honra) y debe fundarse en metodología

científica sólida, transparente e imparcial. Surge así la preocupación de si una IA puede generar contenido fiable y exento de sesgos, y cómo garantizar responsabilidad y respeto al debido proceso. Casos recientes alimentan la cautela: en EE. UU., un abogado fue sancionado por presentar un escrito legal elaborado con ChatGPT que contenía citas jurisprudenciales inventadas por la herramienta (Cedillo Lazcano, 2020).

En Colombia, un juez reveló haber usado ChatGPT para redactar partes de una sentencia de tutela en 2023, lo que provocó debate público sobre la legitimidad de tal práctica; la Corte Constitucional colombiana, al revisar el caso en 2024, avaló un uso asistencial de la IA pero estableciendo criterios estrictos (transparencia, verificación, no delegar la función decisoria, etc.) (Gutiérrez, 2023).

En España, la Sala de lo Civil y Penal del TSJ de Navarra examinó en 2024 si un abogado incurrió en mala fe procesal por apoyar su escrito en ChatGPT sin verificar una referencia legislativa errónea; en su auto, el tribunal recalcó que el uso de estas tecnologías exige *especial diligencia y verificación adicional* por parte del profesional, siendo el abogado el último responsable de la exactitud y rigor del documento forense. Estos incidentes evidencian riesgos reales de la IA Gen (p. ej., "alucinaciones" o errores factuales, sesgos en la información) y han motivado respuestas regulatorias iniciales (Barrientos, 2025).

En México, el tema comienza a cobrar relevancia a través un hito jurisprudencial reciente: la tesis aislada II.2o.C.9 K (11a.) del Segundo Tribunal Colegiado en Materia Civil del Segundo Circuito (Queja 212/2025), publicada el 22 de agosto de 2025 en el Semanario Judicial, que por primera vez fija "elementos mínimos" para un uso ético y responsable de IA en sede jurisdiccional con perspectiva de derechos humanos—i) proporcionalidad e inocuidad (restringir la IA a apoyos técnicos, p. ej., cálculos numéricos, sin sustituir el razonamiento jurídico); ii) protección de datos personales; iii) transparencia y explicabilidad (declarar su uso, metodología, datos y trazabilidad para auditoría); y iv) supervisión y decisión humanas—parámetros que, por analogía, resultan transferibles al quehacer pericial: la IA

Gen puede fungir como auxiliar documentado, nunca como "perito" ni como fuente autónoma de juicio experto, y su empleo exige protocolizar insumos, salidas y validaciones, resguardando confidencialidad, cadena de custodia y control de sesgos conforme a los principios de la LGHCTI (ética, inclusión, no discriminación y responsabilidad social).

Esta línea, además, dialoga con referentes internacionales expresamente citados por el propio tribunal (UNESCO 2021; Directrices de IA Fiable de la UE; y el AI Act de la Unión Europea), ofreciendo un marco normativo blando para guiar lineamientos técnico-forenses internos mientras el legislador mexicano desarrolla reglas específicas (El Economista, 2025).

Este artículo aborda de forma integral la viabilidad, los riesgos éticos y las implicaciones jurídicas de la IA generativa en los dictámenes de psicología forense dentro del sistema mexicano. Se estructura en secciones que analizan: (1) el potencial y usos concretos en que la IA Gen podría apoyar al perito psicólogo (viabilidad), (2) los peligros y dilemas éticos que conlleva su empleo sin las debidas precauciones (riesgos éticos), (3) el estado del marco legal aplicable y comparado, incluyendo la situación en México y en otros países que sirven de referencia, (4) una discusión crítica que contrasta beneficios vs. riesgos a la luz de principios bioéticos y de justicia algorítmica, y (5) una propuesta de lineamientos técnicos y éticos para regular el uso responsable de IA Gen en peritajes psicológicos.

Metodología

El presente estudio se basó en una revisión documental y análisis crítico de fuentes especializadas. Se consultaron artículos científicos publicados entre 2019 y 2025 en los campos de psicología forense, inteligencia artificial, derecho y ética, obtenidos de bases de datos académicas (*Frontiers in Psychiatry*, *IDP*, *InDret*, etc.), así como informes de organismos internacionales (UNESCO, ABA) y legislación nacional pertinente. La selección se enfocó en investigaciones sobre IA generativa en contextos forense-legales, estudios empíricos de uso de IA en salud mental y justicia, y ensayos teóricos sobre ética de la IA.

Asimismo, se incluyeron casos jurisprudenciales recientes y pronunciamientos regulatorios de distintos países para un análisis comparado (p. ej., sentencia T-323/2024 de la Corte Constitucional de Colombia, caso del TSJ Navarra 2024 en España, Opinión Formal 512 de la American Bar Association).

Se procedió primero a una lectura exploratoria para mapear los subtemas clave: formas de uso potencial de IA Gen en peritajes, tipos de riesgos identificados (bias, opacidad, etc.), principios ético-jurídicos implicados y situación normativa. Con base en ello, se definió la estructura del artículo y se realizó una lectura analítica profunda de cada fuente relevante, extrayendo evidencias, datos y argumentos significativos.

La información recopilada se organizó temáticamente (viabilidad, riesgos, marco legal, etc.) y se contrastó de forma crítica, identificando convergencias y divergencias entre autores y entre distintas jurisdicciones. Por último, se aplicó un enfoque hermenéutico-jurídico para interpretar los hallazgos a la luz del ordenamiento mexicano vigente y los principios deontológicos de la práctica pericial, formulando propuestas normativas adaptadas a nuestra realidad.

Esta metodología combinó la perspectiva interdisciplinaria (integrando conocimientos de tecnología, psicología y derecho) con un rigor académico en la selección de fuentes (priorizando literatura revisada por pares y normativa oficial). Dicha triangulación garantiza que el análisis y las conclusiones presentadas se sustenten en evidencia actual y confiable. No obstante, dada la rapidez con que evoluciona la IA Gen, se reconoce que este estudio ofrece una fotografía temporal de un fenómeno en desarrollo; se espera que futuras investigaciones sigan enriqueciendo el debate conforme haya nuevas experiencias prácticas y regulaciones sobre la materia.

Resultados

La posibilidad de incorporar IA generativa en el trabajo pericial psicológico es técnicamente factible y presenta varias vías de aplicación. En esencia, la IA

Gen puede funcionar como una *herramienta de apoyo* que potencia las capacidades del perito, más que como un sustituto de su juicio profesional. A continuación, se examinan los usos viables más relevantes:

Asistencia en la redacción de informes periciales

Un dictamen forense suele ser extenso y demandar tiempo en su elaboración. Los modelos de lenguaje (LLMs) podrían ayudar a *esbozar partes del informe*, por ejemplo, redactando antecedentes del caso, explicaciones teóricas o incluso borradores de conclusiones basadas en datos ingresados. Esto podría agilizar la producción de dictámenes, permitiendo al perito concentrarse en el análisis sustantivo.

Estudios recientes señalan que los LLM avanzados pueden generar textos coherentes y estilísticamente adecuados en contextos clínicos, actuando como "copilotos" de escritura científica. En el ámbito psiquiátrico, ya se observó que IA multimodales pueden asistir en la elaboración de reportes clínicos integrando múltiples fuentes de datos del paciente. De manera análoga, un modelo entrenado con informes psicológicos forenses previos podría sugerir redacciones estándar para ciertos apartados (p. ej., descripción de pruebas psicométricas aplicadas, criterios diagnósticos), siempre bajo la revisión final del experto humano (Corvalán, 2021).

Análisis de grandes volúmenes de información

En casos complejos, el perito debe revisar expedientes voluminosos, historiales médicos, declaraciones, etc. La IA Gen combinada con técnicas de procesamiento de lenguaje natural puede *resumir documentos*, extraer menciones clave o detectar inconsistencias en testimonios.

Su capacidad de integrar datos no estructurados (texto libre de entrevistas, notas clínicas, publicaciones científicas) podría ayudar a formar una visión más completa del caso (Diez-Feijóo Varela, 2025). Por ejemplo, un modelo podría sintetizar cientos de páginas de actuaciones judiciales en unos cuantos párrafos resaltando información psicológicamente relevante para la pericia. También podría comparar rápidamente el relato de un

imputado en distintas declaraciones para ver si mantiene consistencia, lo cual es útil en evaluaciones de credibilidad. Esta automatización del "minado" de información permitiría al perito ahorrar tiempo y no pasar por alto detalles importantes.

Evaluación y predicción de riesgos mediante IA

Como se mencionó, los psicólogos forenses ya usan escalas e instrumentos actuariales para estimar riesgo de violencia o reincidencia (Herramientas como HCR-20, VRAG, etc.). La IA Gen, especialmente combinada con *machine learning* discriminativo, puede elevar esa práctica al analizar patrones complejos en bases de datos de delincuentes. En Estados Unidos, algoritmos como COMPAS han sido empleados para predecir riesgo de reincidencia en decisiones pre-sentenciales, si bien con polémicas por sesgos raciales (Muñoz Rodríguez, 2020).

En el entorno forense mexicano, podría entrenarse una IA con datos históricos (p. ej., perfiles psicológicos de agresores condenados, circunstancias del delito) para estimar la probabilidad de conductas futuras o peligrosidad de un evaluado. Este soporte cuantitativo *podría dotar de mayor objetividad* al dictamen en lo referente a riesgo, siempre y cuando se verifique la validez del modelo. Algunos trabajos reportan que los algoritmos predictivos alcanzan performances moderadas (AUC ~0.75) en predicción de reincidencia, útiles como complemento pero no como único criterio. Así, la IA sería viable como sistema de alerta o apoyo estadístico para las conclusiones del perito en materia de riesgo (American Board of Professional Psychology, 2025).

Detección de simulación o malingering

Otra aplicación promisoria es usar IA para identificar patrones sutiles que delaten simulación de trastornos. Investigaciones en psicología han aplicado *machine learning* a medidas comportamentales (como movimientos de ratón o tiempos de respuesta) para distinguir pacientes genuinos de simuladores con alta precisión (hasta 90%+).

Un modelo generativo entrenado con datos de evaluaciones podría, por ejemplo, elaborar hipótesis sobre la congruencia de los resultados de pruebas de personalidad o síntomas reportados, señalando al perito posibles indicadores de fingimiento. Nuevamente, no sería una "prueba" por sí sola, pero sí un *indicador de probabilidad* que oriente al experto a realizar exploraciones más dirigidas (Gordon & Turnbull, 2024).

Simulación de escenarios y formación pericial

Más allá de casos específicos, la IA Gen puede emplearse para *entrenamiento y educación* de peritos. Investigadores han propuesto generar "gemelos digitales" o simulaciones interactivas de pacientes y situaciones forenses, mediante modelos generativos, a fin de que profesionales en formación practiquen sus habilidades diagnósticas y decisorias. Por ejemplo, un entorno virtual podría simular la entrevista con un imputado con trastorno mental, con respuestas generadas por IA, de forma que el psicólogo practique cómo interrogar y detectar inconsistencias éticas. Asimismo, en juicios simulados, la IA podría asumir el rol de testigo o incluso de "juez virtual" para entrenar al perito en la defensa de su dictamen bajo contrainterrogatorio. Estos usos formativos incrementarían la preparación y reducirían errores en la práctica real (Siliceo Portugal, 2024).

La adopción de IA Gen entre profesionales de la salud mental ya es una tendencia emergente. Una encuesta internacional de 2024 encontró que 43% de los profesionales de salud mental encuestados utilizan alguna herramienta de IA, principalmente para labores de investigación y redacción de reportes . Esto indica que casi la mitad ha incorporado IA en tareas como revisar literatura científica, generar borradores o manejar información, lo que sugiere una creciente confianza en estas tecnologías como apoyo diario.

En el sector legal, otra encuesta citada por la ABA muestra un salto en uso de IA en despachos jurídicos de 19% a 79% en solo un año, reflejando la velocidad con que están permeando todas las profesiones vinculadas a la administración de justicia (American Board of Professional Psychology, 2025). Psicólogos forenses en México podrían sumarse a esta ola, aprovechando IA Gen para lidiar con sobrecarga de casos, uniformizar

criterios periciales y reducir la duración de los procesos. Además, dada la escasez de peritos en zonas apartadas, la IA podría *democratizar el acceso* a herramientas de apoyo forense, siempre que haya capacitación adecuada.

En suma, la viabilidad técnica existe y las aplicaciones potenciales son diversas: desde aceleración de tareas burocráticas hasta análisis sofisticados de conducta. La IA Gen bien empleada tiene el potencial de mejorar la calidad y eficiencia de los dictámenes psicológicos forenses, reforzando la base empírica de sus conclusiones. No obstante, cada una de estas promesas viene acompañada de importantes consideraciones éticas y limitaciones que se deben sopesar cuidadosamente antes de incorporar la IA en casos reales, como se detalla en la siguiente sección (Loaiza et al., 2024).

Riesgos éticos y desafíos de la IA generativa en la pericia forense

A pesar de sus ventajas, la IA generativa conlleva riesgos éticos sustanciales en el contexto forense, donde un error puede traducirse en injusticias (condenar a un inocente, revictimizar a una persona, etc.). Es imprescindible identificar y mitigar estos riesgos antes de utilizar IA Gen en la práctica pericial.

Entre los desafíos éticos y prácticos más significativos destacan:

Sesgos algorítmicos y discriminación. Los modelos de IA aprenden de datos históricos que pueden contener prejuicios. En el ámbito penal, esto es crítico: si los datos de entrenamiento reflejan disparidades (p. ej., perfiles de condenados donde cierto grupo étnico está sobrerrepresentado), la IA perpetuará o incluso exacerbará sesgos en sus resultados.

Estudios han demostrado que grandes modelos de lenguaje replican estereotipos de género y raza presentes en sus fuentes: por ejemplo, se ha visto que ChatGPT y GPT-3.5 asocian profesiones prestigiosas con el género masculino y generan con mayor frecuencia contenido violento al referirse a la religión musulmana. En modelos generativos de imágenes, como DALL-

E 2, se halló subrepresentación de mujeres en roles técnicos y sesgos occidentalocéntricos.

Trasladado a psicología forense, un sistema sesgado podría *evaluar como más peligrosa* a una persona solo por pertenecer a un grupo minoritario o atribuir rasgos de violencia basándose en estereotipos raciales, aunque el individuo en cuestión no presente tales indicios de riesgo (Ramirez, 2023).

Esto atentaría contra el principio de igualdad ante la ley y los derechos humanos. De hecho, en EE. UU. ya hubo controversia con el algoritmo COMPAS de riesgo de reincidencia: un análisis reveló que tenía tasa de falsos positivos mayor para personas negras (marcándolos erróneamente como de alto riesgo). Un dictamen pericial generado o apoyado por una IA con sesgo racial podría no solo ser científicamente inválido sino también inconstitucional por discriminatorio.

Por tanto, la presencia de sesgos es uno de los riesgos más graves, ya que minaría la imparcialidad de la justicia e incrementaría la criminalización desproporcionada de colectivos marginados. El perito tiene la obligación ética de asegurar que las herramientas que use no introduzcan discriminación, pero dada la complejidad de los modelos, detectar sesgos ocultos resulta difícil sin auditorías algorítmicas rigurosas (Roa Avella et al., 2022).

Falta de transparencia y explicabilidad (la "caja negra"). Los modelos generativos complejos, basados en redes neuronales profundas, operan como cajas negras cuyo razonamiento interno resulta opaco incluso para sus creadores. En un proceso judicial, esto choca con la exigencia de motivación y explicabilidad de las decisiones (Bejerano, 2024; Rodriguez-Sanchez, 2024). Todo dictamen pericial debe exponer claramente cómo llega a sus conclusiones (metodología, datos considerados, literatura respaldatoria) para que pueda ser controvertido y valorado por el juez. Si un informe se apoya en las salidas de una IA que no son explicables ("porque el algoritmo así lo indicó"), se compromete el derecho de defensa y al debido proceso.

La literatura advierte que muchos modelos actuales no permiten conocer el peso que cada factor tuvo en su output ni las inferencias realizadas.

Esto genera serias dificultades de accountability: ¿quién responde si la IA se equivocó? ¿Cómo refutar un resultado cuyo proceso subyacente no puede desentrañarse? La opacidad, sumada a la posible naturaleza propietaria del software (código cerrado de compañías), agrava el problema (Calderon & Cueto, 2022; Cantos Pardo, 2024).

La ausencia de explicaciones vulnera principios de justicia procedimental y puede volver inadmisible o poco fiable la prueba. En términos éticos, esto contraviene el deber de transparencia y la obligación del perito de sustentarlo todo en conocimientos comprensibles. Además, la "magia" de la IA puede impresionar indebidamente a jueces o jurados ("efecto aureola tecnológica"), llevándolos a otorgar un peso excesivo a lo que no entienden, lo cual es peligroso. En síntesis, mientras la IA siga siendo en gran medida inexplicable, no debería utilizarse para decisiones que requieren fundamentación clara, como los dictámenes que apoyan sentencias.

Posibilidad de errores factuales y "alucinaciones". La IA generativa es propensa a inventar datos plausibles pero falsos cuando no encuentra información suficiente, un fenómeno conocido como *alucinación*. Esto se evidenció dramáticamente en el caso del abogado neoyorquino sancionado: ChatGPT generó referencias a fallos judiciales totalmente ficticios pero con apariencia real En un informe pericial, una IA podría "rellenar" lagunas con contenido inexacto: por ejemplo, citar un estudio psicológico inexistente para respaldar una conclusión, o describir síntomas que el evaluado nunca reportó, si se le pidiera completar un perfil. Si el perito no detecta estas falsedades, el dictamen contendría información espuria que podría inducir a error al tribunal. La verificación de hechos es, pues, crítica. Otra faceta es la generación de *deepfakes*: IA que producen imágenes, videos o audios falsos pero creíbles.

En un contexto pericial, alguien podría presentar (maliciosamente) un audio "deepfake" simulando la voz de un acusado confesando, o una imagen adulterada de la escena del crimen. Si los peritos o jueces confían ciegamente en herramientas de IA para analizar evidencias multimedia, podrían ser engañados. Incluso hay proyectos como *Forensic Sketch AIrtist* que generan

retratos hablados a partir de descripciones, lo cual podría amplificar los sesgos o errores de los testigos originales. Imaginemos un peritaje donde se use tal herramienta: cualquier distorsión en el rostro generado podría llevar a identificar erróneamente a una persona inocente. Por ende, la IA puede introducir falsedades con apariencia de verdad, lo que plantea un riesgo ético de desinformación y requerirá que el perito mantenga una actitud escéptica y comprobatoria frente a las salidas de la IA (Spinak, 2023).

Excesiva dependencia y erosión del juicio profesional. Un peligro más sutil es que, al usar cada vez más la IA como muleta, el perito humano termine reduciendo su participación activa, aceptando acriticamente las sugerencias del modelo. Esta *automatización de la decisión* podría minar la autonomía y pericia del profesional, convirtiéndolo en un mero validador pasivo. Autores como Adela Cortina advierten que no debemos ceder nuestra capacidad crítica y deliberativa a sistemas automatizados. Si un perito delega demasiado en la IA (por ejemplo, permitiendo que redacte casi todo el informe), se corre el riesgo de una "sustitución de la racionalidad humana", lo cual la Corte colombiana explícitamente señaló que debe evitarse en la administración de justicia (Cortina Orts, 2024).

Además, la interacción constante con modelos generativos podría producir cierta *homogeneización de los informes*: todos con un estilo similar generado por IA, repitiendo los mismos clichés, perdiendo la riqueza del criterio individual y la adaptación a las particularidades del caso. Desde la bioética, esto afecta el principio de autonomía profesional y puede diluir la responsabilidad: el perito podría alegar "eso puso la IA" ante un fallo, intentando eludir culpa. Pero éticamente, el psicólogo forense no puede abdicar de su discernimiento; usar IA exige un equilibrio delicado para que siga siendo el humano quien toma las decisiones finales.

Las guías de la ABA para abogados sobre IA van en esa línea: el profesional debe entender la tecnología y supervisar sus resultados con rigor. Similarmente, los psicólogos deberán resistir la tentación de confiar ciegamente en la IA por comodidad, pues eso puede llevar a errores de juicio, pérdida de pericia con el tiempo (si uno deja de practicar análisis complejo

porque la máquina lo hace) e incluso problemas de ética profesional por negligencia (American Bar Association, 2024).

Problemas de privacidad y confidencialidad de los datos. Los peritajes psicológicos manejan datos personales sensibles (historial psiquiátrico, traumas, conductas delictivas, etc.). Ingresar esta información en una plataforma de IA Gen (muchas de las cuales operan en la nube) podría vulnerar la confidencialidad si no se toman precauciones. Por ejemplo, servicios como ChatGPT almacenan las consultas de los usuarios; cargar detalles de un caso podría implicar que esos datos residan en servidores externos, tal vez ubicados fuera de México, escapando al control del perito y potencialmente accesibles a terceros no autorizados.

Esto transgrede obligaciones de secreto profesional y podría violar la Ley Federal de Protección de Datos Personales. Incluso si la IA opera localmente, existe el riesgo de brechas de seguridad o de que el modelo revele información sensible *aprendida* a partir de casos previos (aunque en teoría los LLM no deberían divulgar datos específicos, ha habido demostraciones de que pueden filtrar partes de sus datos de entrenamiento).

Así, sin protocolos estrictos, el uso de IA puede comprometer la privacidad de evaluados y víctimas. Éticamente, esto contraviene el principio de confidencialidad inherente a la práctica psicológica forense. Además, podría invalidar un dictamen si la parte contraria argumenta que se divulgó indebidamente información o se violó la cadena de custodia de evidencias psicológicas. Por tanto, cualquier implementación tendría que asegurar un entorno seguro para los datos (idealmente sistemas offline o con encriptación robusta) y apego estricto a normas de protección de datos (Mendoza, 2021).

Impacto en la legitimidad y equidad del sistema de justicia. Sumando todos los puntos anteriores, existe un riesgo macro: que un uso imprudente de IA Gen erosione la confianza pública en la justicia. Si los ciudadanos perciben que las decisiones periciales –y por ende judiciales– las toma una máquina incomprensible, o que incorporan sesgos sistémicos, la legitimidad social del sistema penal puede verse afectada. Estudios

criminológicos señalan que la legitimidad influye en el cumplimiento voluntario de la ley y la cooperación ciudadana. Una justicia percibida como "deshumanizada" o "injusta algorítmicamente" podría generar rechazo y menor confianza en peritos y jueces (Moran, 2021).

En España, Baz Cores (2023) reflexiona que la IA debe manejarse con cuidado para no dañar la justicia procedimental, es decir, la impresión de las partes de haber sido tratadas con dignidad, respeto y con decisiones fundamentadas. Si un dictamen forense basado en IA no puede ser cuestionado adecuadamente porque es opaco, o si se sospecha que penaliza más a ciertos grupos por sesgo, las partes podrían sentir indefensión. En última instancia, la introducción de IA Gen sin suficientes garantías podría vulnerar el derecho a un juicio justo (Art. 20 constitucional mexicano) y llevar a más impugnaciones y apelaciones en tribunales superiores, lo que entorpecería el sistema.

En resumen, los riesgos éticos de la IA generativa en peritajes son reales y variados. No se trata de demonizar la tecnología, sino de reconocer sus límites actuales: reproduce prejuicios de su entrenamiento, carece de explicabilidad, puede inventar datos, requiere supervisión constante, y suscita interrogantes sobre responsabilidad y privacidad. Estos desafíos demandan respuestas antes de su aplicación práctica.

La ética profesional y los principios bioéticos son una guía indispensable: *beneficencia* (usar IA solo si trae beneficio claro a la justicia), *no maleficencia* (evitar dañar a individuos con diagnósticos o recomendaciones erróneas), *autonomía* (mantener el control humano y respetar los derechos de las personas evaluadas) y *justicia* (que la IA no introduzca inequidades, sino idealmente ayude a reducirlas). El siguiente apartado profundiza en este marco normativo-ético, incluyendo aportes de autores clave, para fundamentar lineamientos que respondan a tales riesgos (Gonzalez & Martinez, 2020).

Tabla 1

Tipos de uso de IA generativa en peritajes psicológicos y su nivel de aceptabilidad

Categoría de uso	Ejemplos de aplicación	Condiciones mínimas de aceptabilidad
I. Usos aceptables (bajo control humano directo)	- Redacción preliminar de apartados descriptivos del dictamen. - Síntesis y organización de expedientes voluminosos. - Ordenamiento de información clínica, documental o cronológica.	- Revisión exhaustiva por parte del perito. - Declaración explícita del uso de IA en el informe. - Sin procesamiento autónomo de datos sensibles. - No interviene en la interpretación clínica.
II. Usos riesgosos (solo con verificación reforzada)	- Análisis de credibilidad del testimonio. - Evaluación o estimación de riesgo de violencia o reincidencia. - Procesamiento de datos sensibles o confidenciales.	- Validación cruzada con métodos tradicionales. - Documentación de parámetros y outputs. - Verificación independiente de resultados. - No utilizarse como evidencia principal sin sustento adicional.
III. Usos inadecuados o incompatibles con el debido proceso	- Inferir diagnósticos sin evaluación humana. - Determinar imputabilidad, peligrosidad o psicopatología de forma autónoma. - Generar conclusiones periciales automáticas o redactadas íntegramente por IA. - Sustituir la entrevista, la valoración clínica o el contacto directo con las personas evaluadas.	- Prohibido su uso. - Inadmisible en procesos judiciales. - Contraviene principios de autonomía, no maleficencia y debido proceso. - Riesgo elevado de sesgo, opacidad y afectación de derechos.

Nota. Elaboración propia a partir de revisión documental.

Marco jurídico y análisis normativo en México e internacional

En México, el uso de IA generativa en la elaboración de dictámenes periciales aún no cuenta con una regulación específica; por ende, debemos remitirnos al marco general de la prueba pericial, la normativa en ciencia y tecnología, y principios legales aplicables. Destacan los siguientes referentes:

Ley General en materia de Humanidades, Ciencias, Tecnologías e Innovación (LGHCTI, 2023). Esta nueva ley, de alcance

general, incorpora principios rectores para el desarrollo científico-tecnológico en México. Su Artículo 6 enfatiza que el Estado promoverá la ciencia y tecnología velando por la responsabilidad ética, social y ambiental, así como la igualdad, no discriminación, inclusión y pluralidad en esas actividades. También alude al *principio precautorio* ante posibles daños. Si bien no menciona expresamente IA, estos principios son totalmente pertinentes: el desarrollo e implementación de IA (como podría ser entrenar modelos para uso forense) deben evitar causar daño social, discriminación o injusticia. La LGHCTI consagra el derecho humano a gozar de los beneficios de la ciencia sin menoscabo de otros derechos.

Esto sugiere que innovaciones como la IA Gen pueden adoptarse para mejorar procesos (beneficio de la ciencia), pero *no* sacrificando derechos fundamentales en el camino. De hecho, la ley prohíbe la discriminación por motivos como género, origen étnico, condición de salud, etc., al aplicar avances científicos. Una IA sesgada violaría este mandato. La LGHCTI podría servir de base para requerir que cualquier incorporación de IA en el sistema judicial pase por una evaluación ética y cuente con directrices que aseguren su uso responsable.

Además, esta ley faculta a las autoridades para expedir políticas y lineamientos en materia de innovación; un ejemplo podría ser que el Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT) –creado por esta ley– elabore, junto con el Poder Judicial, guías sobre IA en la administración de justicia. Cabe mencionar que en 2023 se planteó una posible reforma al CNPP para modernizar la prueba pericial, aunque sin enfoque específico en IA; no obstante, a futuro podrían incorporarse previsiones sobre transparencia de herramientas algorítmicas usadas por peritos.

Legislación en protección de datos personales y ciberseguridad. Aunque tangencial, es relevante. La Ley Federal de Protección de Datos Personales en Posesión de Particulares (LFPDPPP) y su regulación en posesión de sujetos obligados (LGPDPPSO) establecen que datos sensibles (como los de salud mental) requieren consentimiento expreso para su

tratamiento, y medidas de seguridad especiales. Un perito que suba información de un evaluado a un servicio de IA en la nube *podría estar vulnerando estas leyes* si no tiene autorización o si el proveedor no garantiza un nivel adecuado de protección.

Además, la Ley General de Archivos y lineamientos del Poder Judicial sobre manejo de expedientes podrían implicar restricciones a sacar información del expediente fuera de los sistemas institucionales. Todo esto sugiere que, legalmente, el uso de IA debe hacerse preservando estrictamente la confidencialidad, quizá mediante convenios específicos con los desarrolladores de IA o usando versiones locales aisladas de internet. De lo contrario, no solo hay un riesgo ético sino legal de sanciones o nulidad de actuaciones por violar la privacidad de la información forense.

Posibles reformas o iniciativas específicas. Hasta la fecha, no existe en México una "Ley de Inteligencia Artificial" o disposición semejante. Sin embargo, en foros académicos y legislativos se ha discutido la necesidad de un marco para IA. Por ejemplo, en 2024 el Senado mexicano organizó foros sobre IA y derecho, y algunos legisladores mencionaron explorar regulaciones inspiradas en el enfoque europeo (donde se clasifica el uso de IA por niveles de riesgo).

Una reforma plausible al mediano plazo podría ser agregar al CNPP (o en una ley de ciencia aplicada a justicia) la obligación de transparencia algorítmica: es decir, si un perito utilizó un algoritmo o IA para llegar a sus conclusiones, que deba revelarlo en su informe y aportar la información técnica necesaria para evaluarlo (parámetros, tasa de error conocida, etc.).

Algo análogo existe en países como EE. UU. bajo estándares tipo *Daubert*, donde el juez actúa como guardián de la fiabilidad científica de las pruebas. En México, los criterios de admisibilidad de la pericia no están tan desarrollados judicialmente, pero bien podría la Suprema Corte emitir criterios en caso de impugnaciones sobre el uso de IA en algún caso futuro. También podría contemplarse la figura de un perito en informática forense/IA que evalúe la corrección de la herramienta utilizada, actuando casi

como "perito de peritos" para dar fe de que la IA opera sin sesgos inaceptables. Estas ideas aún son especulativas, pero apuntan a la dirección de *adaptar el marco jurídico* para brindar certeza.

A nivel internacional, es instructivo revisar cómo otros países y organismos están abordando la IA en la justicia, pues ofrecen modelos a considerar: En Colombia, la Sentencia T-323/2024 de la Corte Constitucional estableció un precedente central al permitir el uso asistencial de IA, pero bajo 12 principios estrictos: transparencia, responsabilidad del usuario, privacidad, no sustitución de la racionalidad humana, verificación, prevención de riesgos, igualdad, control humano permanente, regulación ética, adecuación a buenas prácticas, seguimiento continuo e idoneidad técnica. Además, ordenó crear una guía oficial para la Rama Judicial. Este modelo muestra que es posible innovar sin sacrificar garantías, y aplicado a la pericia implicaría: informar explícitamente el uso de IA en el dictamen, revisar y corregir toda salida generada, proteger datos personales y no delegar nunca la apreciación probatoria a un sistema automatizado.

En España, aunque no existe una ley específica, el Auto 2/2024 del TSJ de Navarra analizó por primera vez el uso indebido de ChatGPT por un abogado y enfatizó que la IA no exime los deberes de diligencia y veracidad, sino que exige verificación adicional. El debate doctrinal, especialmente sobre la "prueba pericial informática" (Velasco, 2023, *Diario La Ley 95*), refuerza que la última palabra en la valoración probatoria debe permanecer en manos del juez. A nivel europeo, el AI Act —en fase final (2023)— clasifica los sistemas de IA usados en justicia como de alto riesgo, imponiendo transparencia, gestión de riesgos, supervisión humana estricta y prohibiciones como el reconocimiento emocional en entornos judiciales. Su entrada en vigor prevista para 2025 obligará a los Estados miembros a adecuar sus prácticas, ofreciendo un marco útil que México podría considerar para evitar rezagos regulatorios.

En Estados Unidos, pese a la ausencia de regulación federal unificada, la American Bar Association emitió la Opinión Formal 512 (2024), que extiende los deberes de competencia, confidencialidad, comunicación

informada y honorarios razonables al uso de IA. Esto implica que el profesional debe entender la tecnología, proteger datos sensibles, explicar riesgos y no atribuirse trabajo generado por IA. En el ámbito judicial, algunos jueces ya exigen declarar si un escrito fue elaborado con IA y presentar verificación de fuentes, como en el caso del juez federal de Texas (2023), para evitar errores como los ocurridos en el caso Avianca. Además, organismos como NIST han desarrollado marcos de gestión de riesgos aplicables al entorno forense, con énfasis en *fairness*, *accountability* y robustez.

Los organismos internacionales también ofrecen estándares relevantes. La UNESCO (2021) propone principios globales como proporcionalidad, seguridad, justicia, inclusión y supervisión humana. La CEPEJ del Consejo de Europa adoptó en 2018 la Carta Ética Europea sobre IA en justicia, que destaca cinco principios: derechos fundamentales, no discriminación, calidad y seguridad, transparencia e imparcialidad, y control humano. Estos instrumentos, aunque no vinculantes para México, trazan una orientación clara: la IA debe ser auditada, trazable, contestable y nunca desplazar la deliberación humana.

Tabla 2

Síntesis comparada de lineamientos, marcos regulatorios y estándares sobre IA en justicia y su relevancia para la pericia psicológica

Jurisdicción / Organismo	Hallazgos y lineamientos clave	Implicaciones para la pericia psicológica
México Marco jurídico general aplicable a IA en peritajes	Sin regulación específica sobre IA en dictámenes periciales. El uso debe remitirse al marco general de la prueba pericial, la normativa científico-tecnológica y principios constitucionales. LGHCTI (2023): Art. 6 establece principios de responsabilidad ética, igualdad, no discriminación, inclusión, pluralidad y principio precautorio. Derecho humano a beneficios de la ciencia sin vulnerar otros derechos. Prohíbe discriminación en la aplicación de tecnología. • Potencial para que CONAHCyT y el Poder Judicial desarrollen guías sobre IA en justicia.	• Todo uso de IA debe cumplir principios constitucionales: debido proceso, dignidad, igualdad y no discriminación. • Prohibición implícita de utilizar IA con sesgos que generen daño o discriminación (mandato de LGHCTI). • Exigencia de consentimiento informado, confidencialidad reforzada

Jurisdicción / Organismo	Hallazgos y lineamientos clave	Implicaciones para la pericia psicológica
		y uso de IA en entornos seguros.
	Protección de datos: LFPDPPP y LGPDPPSO exigen consentimiento expreso y medidas reforzadas para datos sensibles. Riesgo legal si un perito sube datos a IA en la nube sin garantías adecuadas. Ley General de Archivos y lineamientos judiciales podrían limitar extracción de información fuera de sistemas institucionales.	• Se prevé que la transparencia algorítmica y la trazabilidad sean requisitos futuros para la validez de la prueba pericial.
	Reformas futuras: en discusión nacional—posibles normas sobre transparencia algorítmica, criterios tipo Daubert, obligaciones de revelar el uso de IA y parámetros técnicos, y eventual figura de un perito en informática forense/IA para auditar herramientas utilizadas.	• Necesidad urgente de lineamientos técnicos-éticos para evitar nulidades por violar garantías procesales.
Colombia Corte Constitucional, Sent. T-323/2024	Uso permitido pero altamente regulado. Define 12 principios: transparencia, responsabilidad, privacidad, no sustitución de racionalidad humana, verificación, prevención de riesgos, igualdad, control humano permanente, regulación ética, adecuación a estándares, seguimiento continuo, idoneidad técnica. Ordena elaborar una guía judicial oficial sobre IA.	• Informar en el dictamen si se usó IA. • Verificar y corregir todos los outputs generados. • Proteger datos sensibles bajo estándares reforzados. • No delegar apreciación probatoria a IA. • Modelo útil para guiar lineamientos mexicanos.
España TSJ Navarra, Auto 2/2024	La IA no exime deberes profesionales; exige verificación adicional. Caso pionero sobre uso indebido de ChatGPT por un abogado. Discusión activa sobre fiabilidad de la "prueba pericial informática". Doctrina enfatiza que el juez debe conservar la última palabra (Velasco, 2023).	• Revisión exhaustiva y diligente de outputs de IA. • Perito sigue siendo responsable técnico integral del dictamen. • Necesidad de mantener control humano decisorio.
Unión Europea AI Act (2023–2025)	Clasifica IA en justicia como alto riesgo. Requiere transparencia, gestión de riesgos, supervisión humana y calidad de datos. Prohíbe puntuación social y limita reconocimiento emocional en ámbitos judiciales.	• La práctica pericial deberá adaptarse a estándares de auditabilidad, explicabilidad y trazabilidad. • Relevante para México como modelo regulatorio prospectivo.
Estados Unidos	Extiende deberes de competencia, confidencialidad, comunicación informada y honorarios razonables al uso de IA.	• Peritos deben acreditar competencia técnica antes de usar IA.

Jurisdicción / Organismo	Hallazgos y lineamientos clave	Implicaciones para la pericia psicológica
ABA, Opinión Formal 512 (2024)	Abogado debe entender tecnología y proteger datos. Jueces federales exigen declarar si un escrito usó IA y proporcionar verificación de fuentes (Texas 2023). NIST desarrolla marcos de riesgo en IA (fairness, accountability, robustez).	• Posible requerimiento de consentimiento informado para procesar datos. • Debe declararse el uso de IA en el dictamen. • Verificación y control documental como estándar de buena práctica pericial.
Organismos internacionales UNESCO (2021); CEPEJ (2018)	UNESCO: principios globales (proporcionalidad, seguridad, justicia, inclusión, supervisión humana). CEPEJ: 5 principios —derechos fundamentales, no discriminación, calidad-seguridad, transparencia-imparcialidad, control humano.	• Reforzamiento del principio de supervisión humana permanente. • Uso de IA debe ser auditado, trazable y contestable. • Estándares éticos útiles como guía base para el contexto mexicano.

Nota. Elaboración propia a partir de revisión documental.

Discusión: ponderando beneficios, riesgos y principios bioéticos

Tras examinar las promesas de la IA generativa y sus peligros, es preciso realizar una valoración crítica: ¿cómo encontrar un equilibrio que permita aprovechar la IA en la psicología forense sin socavar la ética ni la justicia? Esta discusión se centra en dos ejes: (a) la aplicación de principios bioéticos y de justicia algorítmica para evaluar moralmente el uso de IA Gen, y (b) la necesidad de una aproximación prudente, incremental y con controles humanos robustos.

La bioética clásica propone cuatro principios cardinales (Beauchamp & Childress) que aquí ofrecen un marco útil:

Autonomía

Implica respetar la capacidad de decisión de las personas involucradas y la independencia del profesional. En nuestro contexto, la autonomía se concreta en mantener al perito como agente decisor. La IA debe servirle de apoyo,

pero nunca reemplazar su juicio experto ni coartar su libertad de evaluar críticamente cada caso.

Asimismo, desde la perspectiva de las partes (evaluado, acusado, víctima), su autonomía y dignidad exigen que no sean tratados meramente como "datos" para un algoritmo, sino como sujetos cuyo contexto y singularidad se considera. Por ejemplo, un evaluado tiene derecho a que su evaluación psicológica no sea totalmente estandarizada por una máquina, sino que el perito escuche activamente su relato. La autora Adela Cortina subraya que no podemos delegar las decisiones morales a una supuesta racionalidad algorítmica: la deliberación ética y jurídica debe seguir en manos humanas. Solo así se garantiza respeto por la autonomía de la voluntad (p. ej., decidiendo imputabilidad, medidas terapéuticas, etc., caso por caso con sensibilidad) (Suarez-Munoz, 2024).

Beneficencia

Obliga a buscar el bienestar y beneficio, en este caso de la administración de justicia y de las personas involucradas. La IA Gen podría aumentar la beneficencia si acelera procesos (justicia más pronta) y mejora la calidad de las pericias (mayor base empírica, reducción de errores humanos). Una adopción responsable de IA Gen podría, por ejemplo, reducir la carga de trabajo de peritos escasos, permitiendo que los casos se atiendan con menos demoras –lo cual beneficia a víctimas y acusados al acortar la incertidumbre procesal.

También podría ayudar a estandarizar criterios periciales, reduciendo arbitrariedad. Sin embargo, para cumplir genuinamente la beneficencia, la IA debe demostrar que sus aportes efectivamente mejoran la precisión o equidad de las conclusiones. Introducir IA por moda o presión, sin evidencia de su utilidad, sería contrario a este principio. Por ende, habría que validar cada aplicación (quizá mediante estudios piloto controlados) antes de adoptarla formalmente, asegurando que añade valor (p. ej., detecta mentiras más eficazmente que el promedio humano, o identifica riesgo de violencia con mayor acierto) en lugar de entorpecer (Garcia, 2025; Paulo, 2024).

No maleficencia

"Primero, no hacer daño". Esto es crucial: no maleficencia implica que si la IA generativa conlleva riesgos graves no mitigables, es preferible no usarla. Por ejemplo, si un modelo no puede explicar su razonamiento y podría generar conclusiones sesgadas, entonces introducirlo en un caso penal podría causar un daño (condena injusta o absolución indebida). Hasta que no se minimice ese riesgo con salvaguardas, sería antiético emplearlo. Aquí aplica el principio precautorio de la LGHCTI: ante incertidumbre sobre un posible perjuicio, abstenerse o extremar precauciones.

La no maleficencia exige también *corregir activamente* los sesgos y errores para que la IA no perjudique a poblaciones vulnerables. Por ejemplo, sabiendo que hay sesgo racial en muchos datos criminales, cualquier sistema de riesgo debería calibrarse para eliminar ese sesgo (quizá mediante técnicas de re-muestreo, ajuste de umbrales por grupo, etc.). De lo contrario, su implementación sería maleficente al replicar discriminación. Además, la no maleficencia cubre daños psicológicos: una IA podría etiquetar a alguien erróneamente como psicópata o simulador; tal estigma es dañino. El perito tiene el deber de prevenir esas lesiones morales verificando cuidadosamente las salidas de la IA antes de plasmar nada lesivo en un dictamen (Peñas, 2024).

Justicia

En este contexto, se refiere a la justicia distributiva y procedimental. Supone equidad, trato igualitario y fundamentación consistente. La "justicia algorítmica" extiende este principio: que los algoritmos sean *justos* en sus resultados y *transparentes* en su lógica, para no minar la igualdad ante la ley. Una IA generativa en pericia solo sería aceptable si promueve la equidad o al menos no la socava. Por ejemplo, si ayuda a eliminar el *bias* personal de un perito (todos tenemos prejuicios conscientes o inconscientes), podría incrementar la justicia; pero si añade un nuevo sesgo sistémico, la reduce.

Asimismo, la justicia procedimental implica que las partes puedan entender y confrontar la evidencia; por tanto, un dictamen apoyado en IA debe ser inteligible para no lesionar ese derecho. La transparencia y

explicabilidad son, en el fondo, exigencias de justicia. Como afirma Cortina, una IA debe estar alineada con nuestros valores fundamentales de dignidad y no convertirse en un nuevo mecanismo de dominación acrítica. En suma, desde el prisma de la justicia, la IA Gen debe evaluarse preguntando: ¿Contribuye a decisiones más justas o podría introducir arbitrariedad invisible? Si es lo segundo, sería injusto adoptarla sin remedio (Fernandez, 2019).

Junto a estos principios, se añaden consideraciones de justicia algorítmica: *transparencia*, *rendición de cuentas* y *no discriminación*, que ya hemos abordado como ejes éticos pero que conviene reiterar. Una "IA ética" en el campo forense debe ser desarrollada y usada de manera que cumpla con transparencia (auditabilidad de sus datos y procesos), explicabilidad (poder dar razones comprensibles de sus recomendaciones), *accountability* (que siempre haya un responsable humano identificable por sus acciones) y mitigación de sesgos (pruebas y ajustes constantes para evitar resultados discriminatorios). Estos son justamente los principios que instituciones como UNESCO y la UE han puesto sobre la mesa. Incorporarlos al ejercicio pericial es complejo pero necesario para legitimar la IA ante la sociedad (Colomina Saló, Innerarity & Cantero Gamito, 2024).

Enfoque prudente e incremental

A partir de lo anterior, se aboga por un uso gradual y controlado de la IA generativa en psicología forense, nunca abrupto o generalizado sin preparación. Esto significa que, de introducirse, debería empezar en *proyectos piloto* bajo supervisión estricta, en casos de baja complejidad o como experimento paralelo (por ejemplo, el perito hace su dictamen normal y aparte corre la IA para comparar, sin que influya en el fallo, a fin de evaluar su desempeño).

Solo si los resultados muestran concordancia y utilidad, se pasaría a un uso oficial, y aun así limitado a funciones de apoyo bien definidas. En términos prácticos, la IA Gen podría destinarse inicialmente a tareas auxiliares de gabinete (organizar información, redactar borradores no definitivos) pero la evaluación psicológica cara a cara, la interpretación clínica y la emisión de

conclusiones sustantivas deben permanecer en la esfera humana. Al menos mientras la IA carezca de comprensión genuina y empatía, es impensable delegar en ella la apreciación fina de la conducta humana (Gordon, 2025; Ortega, 2025).

La prudencia en la adopción de IA generativa en la justicia se fundamenta en que sus errores pueden ser sutiles pero altamente impactantes, y en que los sistemas judiciales suelen ser conservadores al incorporar evidencia científica novedosa —como ocurrió con el ADN, que requirió largos procesos de estandarización antes de su aceptación rutinaria. En este sentido, conviene actuar con cautela. La comunidad pericial, idealmente a través de colegios profesionales, debería desarrollar códigos de conducta específicos que delimiten qué herramientas de IA son aceptables y bajo qué condiciones. Asimismo, los jueces necesitan formación para comprender las fortalezas y límites de estas tecnologías; sin dicha capacitación interdisciplinaria, valorar un dictamen que incorpora IA puede volverse problemático y generar litigios sobre admisibilidad (López de Mántaras, 2025).

Un enfoque prudente exige también planes de contingencia ante fallas técnicas o sesgos: si se descubre un error en el sistema utilizado por un perito, podrían cuestionarse la prueba o incluso el proceso mismo. Para evitarlo, todo resultado relevante obtenido mediante IA debería verificarse mediante métodos independientes. Así, si un algoritmo sugiere simulación de síntomas, el perito debería corroborarlo mediante pruebas tradicionales de validación o detección de engaño. Esta duplicación metodológica puede ser demandante, pero es necesaria en las primeras etapas de adopción para no depender ciegamente de la tecnología (López Martínez & García Peña, 2024; Vásquez Pérez, 2024).

Las autoridades judiciales, por su parte, tienen el deber de transparentar y justificar cualquier innovación, aclarando que la IA no sustituye a los expertos, sino que los auxilia, y demostrando que su uso incrementa objetividad y reduce errores humanos. La recomendación general es clara: IA generativa, sí, pero con cautela responsable.

La experiencia internacional proporciona referentes valiosos —los principios de Colombia, la ética regulatoria europea y las cautelas estadounidenses— que deben adaptarse al contexto nacional. Un error en su implementación podría afectar gravemente la justicia y la confianza pública, mientras que un acierto permitiría modernizar la labor pericial con mayor rigor y eficacia. Con esta perspectiva, el siguiente apartado desarrolla lineamientos concretos para una adopción responsable de la IA en la prueba pericial (Cuatrecasas, 2024).

Conclusiones

La incorporación de inteligencia artificial generativa (IA Gen) en el ámbito de la psicología forense plantea desafíos sustantivos en términos técnicos, éticos y jurídicos. Frente a este escenario, se propone un conjunto de lineamientos orientados a guiar su utilización de forma responsable, transparente y científicamente fundamentada. Estas directrices, dirigidas a peritos, autoridades judiciales y desarrolladores de tecnología, pueden fungir como base para futuras guías de buenas prácticas institucionales o para la eventual creación de marcos normativos vinculantes. Su estructura se fundamenta en principios de ética profesional, legalidad, debido proceso, y protección de derechos fundamentales, en concordancia con estándares internacionales y nacionales.

Rol asistencial, no decisorio

La IA generativa debe entenderse como herramienta de apoyo técnico y no como sustituto del juicio profesional humano. Toda información procesada por modelos de IA debe ser validada críticamente por el perito, quien conserva la responsabilidad exclusiva sobre las conclusiones del dictamen.

El principio de no delegación de la racionalidad humana, reconocido en decisiones judiciales como las de la Corte Constitucional colombiana, impone que ningún diagnóstico, estimación de riesgo, o inferencia sobre imputabilidad se derive automáticamente de la IA. Su uso puede incluir la elaboración de borradores o el procesamiento preliminar de documentos,

pero toda producción automatizada deberá ser corregida, contrastada y justificada por el experto ante instancias judiciales.

Transparencia y divulgación en el informe pericial

La utilización de IA debe ser expresamente declarada en el informe pericial. Esta declaración debe incluir el propósito del uso, la herramienta empleada, y el grado de intervención del perito en la edición del contenido. Ejemplo de ello sería: "Se utilizó la herramienta X de IA generativa para elaborar un resumen inicial de antecedentes, el cual fue revisado y corregido por el perito". Adicionalmente, deberá conservarse información técnica relevante (modelo, versión, parámetros) para permitir auditoría y contrainterrogatorio por parte de las partes procesales. La omisión de esta información podría constituir una falta ética y afectar la validez jurídica del dictamen.

Verificación humana y validación cruzada

Todo contenido generado por IA debe ser sometido a un proceso riguroso de verificación humana. Ello implica cotejar datos, identificar errores o inferencias infundadas, y asegurar que las conclusiones sean coherentes con la evidencia empírica y normativa. En casos sensibles, se recomienda la validación cruzada con instrumentos estandarizados. Este procedimiento debe registrarse como parte del expediente técnico del perito, y podrá ser requerido por el juez o las partes como muestra de diligencia profesional.

Capacitación y competencia técnica del perito

El uso de IA requiere habilidades especializadas. En consecuencia, solo los peritos capacitados técnica y éticamente en herramientas de IA generativa deben emplearlas. La capacitación debe abarcar el funcionamiento de los modelos, sus sesgos estructurales, interpretación de outputs probabilísticos, límites jurídicos y salvaguardas éticas. Esta formación es equiparable a la requerida para la aplicación de pruebas psicométricas. La carencia de competencias en este campo podría constituir negligencia profesional en caso de error o afectación al debido proceso.

Selección de herramientas confiables y auditables

Se deben emplear únicamente herramientas que ofrezcan garantías de precisión, protección de datos, capacidad de auditoría y posibilidad de revisión de logs. La validación previa por parte de instituciones judiciales, laboratorios independientes o universidades contribuiría a legitimar su uso. Debe evitarse el uso de modelos opacos o sin certificaciones mínimas de calidad. El criterio de selección deberá documentarse, y estar disponible ante requerimientos judiciales o periciales.

Mitigación activa de sesgos y promoción de la equidad

Antes de su implementación práctica, las herramientas de IA deben ser sometidas a pruebas de sesgo para evitar efectos discriminatorios por género, etnia, edad u otras condiciones. El perito debe asumir una postura activa en la revisión de outputs, identificando estereotipos, lenguaje sesgado o inferencias generalizadas que vulneren el principio de individualización. Las conclusiones no pueden basarse en perfiles poblacionales automatizados, sino en análisis del caso concreto. Este lineamiento se alinea con el principio de igualdad y no discriminación contenido en la Constitución y en tratados internacionales.

Protección de la confidencialidad y seguridad de la información

El tratamiento de datos personales mediante IA requiere un nivel máximo de confidencialidad. Se debe priorizar el uso de sistemas en entornos cerrados y protegidos. En caso de recurrir a plataformas externas, es obligatorio anonimizar los datos y verificar que los términos de uso no impliquen cesión o reentrenamiento con la información ingresada. Los peritos deben actuar conforme a la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP), solicitando consentimiento informado o autorización judicial previa. Los archivos generados por IA deben resguardarse con el mismo nivel de seguridad que el expediente clínico o pericial.

Rendición de cuentas y supervisión institucional

La responsabilidad del uso de IA recae exclusivamente en el perito, quien deberá asumir las consecuencias técnicas, éticas y legales de su implementación. Se propone establecer mecanismos de supervisión colegiada o comités de ética judicial para la revisión aleatoria o dirigida de dictámenes asistidos por IA. Asimismo, los peritos deberán conservar registros de las interacciones con las herramientas utilizadas (prompts, outputs, versiones), los cuales podrán ser auditados en instancias procesales. Esto garantiza trazabilidad y posibilidad de apelación fundamentada.

Limitaciones de uso en materias sensibles

Ciertos ámbitos del peritaje forense deben mantenerse, al menos temporalmente, al margen del uso de IA, en atención a su complejidad y sensibilidad. Casos que involucren evaluaciones de imputabilidad penal, infancia, pueblos indígenas o violencia de género, podrían representar contextos donde los riesgos de sesgo, malinterpretación o deshumanización superen los beneficios potenciales de la tecnología. La exclusión podrá definirse por criterios normativos emanados del Consejo de la Judicatura o jurisprudencia especializada.

Actualización y mejora continua de los lineamientos

La naturaleza dinámica de la IA exige que estos lineamientos sean revisados periódicamente. Se sugiere una revisión anual por un comité multidisciplinario que integre expertos en psicología forense, derecho, ética e inteligencia artificial. Este comité incorporará nueva evidencia científica, incidentes registrados y avances tecnológicos. México deberá integrarse activamente a foros internacionales sobre IA y justicia, a fin de armonizar estándares, intercambiar experiencias y anticipar escenarios futuros.

El sistema jurídico mexicano se encuentra en una fase inicial respecto a la regulación de la IA en la justicia, lo que brinda la oportunidad de construir un marco sólido inspirado en experiencias comparadas. Aunque no existe normatividad específica, principios constitucionales como la no

discriminación, el debido proceso y la ética en la ciencia deben orientar cualquier uso de IA generativa en peritajes, mientras que la ética profesional y la prudencia judicial llenan el vacío regulatorio.

En este contexto, se proponen lineamientos técnicos y éticos que preserven la centralidad del juicio humano —el perito como garante de calidad y el juez como filtro de admisibilidad— reforzados por obligaciones de transparencia, verificación, capacitación y salvaguardas contra sesgos y violaciones a la privacidad. La premisa es que la tecnología debe ajustarse a los valores jurídicos, lo cual requiere gobernanza activa y la capacidad de excluir usos que no cumplan estándares de equidad y fiabilidad. Frente a la reticencia inicial hacia estas tecnologías, un marco claro y formación adecuada pueden convertir la incertidumbre en un uso competente y consciente, favoreciendo una integración gradual que aproveche la IA en tareas que realmente aportan valor sin renunciar a la interpretación contextual, la empatía y el razonamiento ético-jurídico.

El objetivo es una justicia aumentada pero humanizada, donde la IA complemente sin sustituir, guiada por principios bioéticos y de justicia algorítmica; como advierte Cortina, "una tecnología sin ética es simplemente poder sin brújula" (2024). Alcanzar este equilibrio exige diálogo interdisciplinario y más investigación sobre sus impactos psicológicos, procesales y sociales, pero los hallazgos permiten afirmar que la IA generativa puede incorporarse sin sacrificar garantías ni principios, siempre que permanezca bajo el control reflexivo de profesionales responsables.

Tabla 3

Recomendaciones diferenciadas para actores clave en el uso de IA generativa en peritajes psicológicos

Actor	Responsabilidades y acciones recomendadas	Justificación ética y jurídica
Peritos psicólogos	● Declarar explícitamente el uso de IA en el dictamen. ● Utilizar IA únicamente para tareas auxiliares (borradores, síntesis documental).	● Garantiza transparencia y trazabilidad. ● Salvaguarda la autonomía profesional y el debido proceso.

Actor	Responsabilidades y acciones recomendadas	Justificación ética y jurídica
	● Realizar verificación cruzada de toda salida generada por IA. ● Mantener control humano absoluto en la interpretación clínica y en las conclusiones. ● Proteger datos sensibles mediante anonimización y entornos seguros.	● Reduce riesgo de sesgos, errores y afectaciones a derechos fundamentales. ● Cumple con LFPDPPP y principios bioéticos.
Tribunales y autoridades judiciales	● Exigir que todo dictamen indique si se utilizó IA y con qué propósito. ● Solicitar métricas de desempeño o tasa de error cuando la IA aporte elementos analíticos. ● Establecer criterios de admisibilidad y estándares de verificación. ● Capacitar jueces en limitaciones técnicas y éticas de IA generativa. ● Supervisar que la decisión final siempre sea humana.	● Asegura legalidad y debido proceso. ● Facilita el contrainterrogatorio y la defensa técnica adecuada. ● Impide la delegación indebida de funciones jurisdiccionales. ● Previene decisiones basadas en "caja negra" algorítmica.
Colegios profesionales y asociaciones periciales	● Emitir guías de buenas prácticas para uso responsable de IA en pericia psicológica. ● Desarrollar códigos deontológicos específicos para IA generativa. ● Implementar programas de capacitación continua en ética algorítmica. ● Crear comités de revisión técnica de dictámenes asistidos por IA.	● Uniformiza estándares profesionales. ● Minimiza variabilidad arbitraria entre peritos. ● Proporciona marcos de control ético reconocidos por tribunales. ● Promueve la confianza pública en la profesión.
Desarrolladores y proveedores de IA	● Diseñar modelos auditables y compatibles con revisión pericial y judicial. ● Incluir métricas de sesgo, desempeño y explicabilidad del modelo. ● Implementar salvaguardas para evitar uso indebido en contextos forenses de alto riesgo. ● Ofrecer versiones locales o seguras para manejo de datos sensibles.	● Permite trazabilidad algorítmica y auditoría técnica. ● Reduce discriminación algorítmica y riesgo de afectación de derechos humanos. ● Satisface obligaciones legales de protección de datos. ● Facilita la integración segura de IA en el sistema de justicia.

Nota. Elaboración propia a partir de revisión documental.

Referencias

American Bar Association, Standing Committee on Ethics and Professional Responsibility. (2024). *Formal Opinion 512: Conducting competent and ethical representation in the use of generative AI*. https://debevoisedatablog.com/wp-content/uploads/sites/2/2024/08/aba-formal-opinion-512.pdf

American Board of Professional Psychology. (2025, 24 de marzo). AI, mental health, and forensics: Is this the future? *On Board with Professional Psychology, 5*.

Baz, O., & Fernández Molina, E. (2022). La justicia procedimental y la legitimidad del sistema penal en España. *European Journal of Criminology, 19*(2), 228–247.

Baz Cores, O. (2023). Repercusiones de la inteligencia artificial en la legitimidad del sistema penal: Una reflexión desde la justicia procedimental. *InDret, 4*(2023), 291–316.

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics* (8th ed.). Oxford University Press.

Bejerano, P. G. (2024, 15 de octubre). La caja negra de la IA que se resiste a los investigadores. *El País*. https://elpais.com/tecnologia/2024-10-15/la-caja-negra-de-la-ia-que-se-resiste-a-los-investigadores.html

Calderón-Ortega, M. A., & Cueto Calderón, C. A. (2022). Prueba por inteligencia artificial: Una propuesta de producción probatoria desde el dictamen pericial científico en Colombia. *Civilizar, 22*(42), e20220106. https://doi.org/10.22518/jour.ccsh/20220106

Cantos Pardo, M. (2024). La algoritmización del dictamen pericial: ¿Puerta de entrada para la aparición del "perito-robot"? *Actualidad Jurídica Iberoamericana, 21*, 434–467.

Colomina Saló, C., Innerarity, D., & Cantero Gamito, M. (Coords.). (2024, diciembre). *Desigualdad algorítmica: Gobernanza, representación y derechos en la IA* (Núm. 138). CIDOB. https://www.cidob.org/sites/default/files/2025-02/AFERS%20138.pdf

Corte Constitucional de Colombia. (2024, 6 de junio). *Sentencia T-323 de 2024* (M. P. Diana Fajardo Rivera).

Cortina, A. (2024). *¿Ética o ideología de la inteligencia artificial? El eclipse de la razón comunicativa en una sociedad tecnologizada*. Ediciones Paidós.

Cross, S., Bell, I., Nicholas, J., Valentine, L., Mangelsdorf, S., Baker, S., Titov, N., & Alvarez-Jimenez, M. (2024). Use of AI in mental health care: Community and mental health professionals survey. *JMIR Mental Health, 11*(1), e45140.

Cuatrecasas Monforte, C. (2022). *La inteligencia artificial en el proceso penal de instrucción español: Posibles beneficios y potenciales riesgos* [Tesis doctoral, Universitat Ramon Llull]. TDX. http://hdl.handle.net/10803/675100

Diario Oficial de la Federación. (2023, 9 de mayo). *Ley General en materia de Humanidades, Ciencias, Tecnologías e Innovación*.

Fernández Delgado, M. Á. (2019). Interdisciplina y derecho: El ascenso de la tecnarquía y un llamado a deconstruir la Torre de Babel de las profesiones. *Revista de Investigaciones Jurídicas, 43*, 201–236.

García Garduño, D. J. M. (2025). Entre algoritmos, derechos humanos y principios éticos: El impacto de la inteligencia artificial en la atención médica. *Ecos Sociales, 13*(38), 46–66. https://doi.org/10.19136/es.v13n38.6609

González Arencibia, M., & Martínez Cardero, D. (2020). Dilemas éticos en el escenario de la inteligencia artificial. *Economía y Sociedad, 25*(57), 93–109. https://doi.org/10.15359/eys.25-57.5

Gordon, S. F., & Turnbull, B. (2024). Adopción de la inteligencia artificial en el campo de la psicología. *Psicología Iberoamericana, 31*(2). https://doi.org/10.48102/pi.v31i2.547

Granda Romero, B. M., Quezada Quezada, J. M., & Durán Ocampo, A. R. (2025). Aplicaciones de la inteligencia artificial en la Criminalística. *Didáctica y Educación, 16*(1), 420–449.

López de Mántaras, R. (2025, 9 de junio). IA y Justicia: Una combinación que exige prudencia y responsabilidad. *La Vanguardia.*

López Martínez, F., & García Peña, J. H. (2024). IA y sesgos: Una visión alternativa expresada desde la ética y el derecho. *Revista Iberoamericana de Derecho Informático*, 15(1), 109–121.

Mendoza Enríquez, O. A. (2021). El derecho de protección de datos personales en los sistemas de inteligencia artificial. *Revista IUS, 15*(48), 179–207. https://doi.org/10.35487/rius.v15i48.2021.743

Miró-Llinares, F. (2023). Digitalización y algoritmización de la justicia: Introducción a un monográfico en tiempos de regulación de la IA. *IDP. Revista de Internet, Derecho y Política, 39*, 1–6.

Morán Espinosa, A. (2021). Responsabilidad penal de la inteligencia artificial (IA). ¿La próxima frontera? *IUS: Revista del Instituto de Ciencias Jurídicas de Puebla, 15*(48), 289–323.

Muñoz Rodríguez, A. B. (2020). El impacto de la inteligencia artificial en el proceso penal. *Anuario de la Facultad de Derecho. Universidad de Extremadura, 36*, 695–728. https://doi.org/10.17398/2695-7728.36.695

Ortega Litardo, F. M. (2025). El potencial de la inteligencia artificial en la psicología clínica: Innovaciones que transforman la asistencia psicológica. *Multidisciplinary Journal Educational Regent, 2*(1), 1–12.

Parmigiani, G., Meynen, G., Mancini, T., & Ferracuti, S. (2024). Editorial: Applications of AI in forensic mental health – Opportunities and challenges. *Frontiers in Psychiatry, 15*, 1153764. https://doi.org/10.3389/fpsyt.2024.1153764

Paulo Neto, A. (2024). Bioética y democracia en la sociedad digital. *Revista de Bioética y Derecho, 60*, 19–34. https://doi.org/10.1344/rbd2024.60.42990

Peñas García, L. (2024, 28 de mayo). La ética en la inteligencia artificial generativa. *Revista Labor Hospitalaria, 338*.

Ramirez Autrán, R. (2023). Sesgos y discriminaciones sociales de los algoritmos en inteligencia artificial: Una revisión documental. *Entretextos, 15*(39). https://doi.org/10.59057/iberoleon.20075316.202339664

Roa Avella, M. del P., Sanabria-Moyano, J. E., & Dinas-Hurtado, K. (2022). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. *Revista Brasileira de Direito Processual Penal, 8*(1). https://doi.org/10.22197/rbdpp.v8i1.615

Rodríguez-Sánchez, A. E. (2024). La posibilidad de explicación científica a partir de modelos basados en redes neuronales artificiales. *Revista Colombiana de Filosofía de la Ciencia, 24*(48). https://doi.org/10.18270/rcfc.4288

Starke, G., & Ienca, M. (2023). Out of their minds? Externalist challenges for using AI in forensic psychiatry. *Frontiers in Psychiatry, 14*, Article 1134497. https://doi.org/10.3389/fpsyt.2023.1134497

Suárez-Muñoz, F., & Bautista Tejeda, A. A. (2024). Autonomía de la inteligencia artificial: Una reflexión crítica desde la filosofía kantiana. *SUMMA, 6*(1), 1–11. https://doi.org/10.47666/summa.6.1.9

Tiffon Nonis, B.-N. (2024). *La inteligencia artificial aplicada en la psicología criminal y forense: ¿Reto, realidad o ficción?* Real Academia Europea de Doctores.

Tortora, L. (2024). Beyond discrimination: Generative AI applications and ethical challenges in forensic psychiatry. *Frontiers in Psychiatry, 15*, Article 1346059. https://doi.org/10.3389/fpsyt.2024.1346059

Tribunal Superior de Justicia de Navarra. (2024, 4 de septiembre). *Auto 2/2024* (Recurso Núm. 17/2024, Sección Penal).

Vásquez Pérez, M. N. (2024). Ética y género en la IA: Identificar sesgos de género en IA mediante pensamiento complejo. *Revista de Investigación Científica y Crítica Educativa, 2*(2), Artículo 49. https://doi.org/10.48168/RICCE.v2n2p49