Intelética Vol.3(6) (2026) 27-39
Revista de Inteligencia Artificial, Ética y Sociedad

INTELETICA

https://inteletica.iberamia.org/

Incertidumbre algorítmica y agencia moral en la UCI: comunicar los límites de la predicción clínica.

Alfonso Canabal-Berlanga [1,2,3], Beatriz Abad [2], Juan Antonio Sánchez-Giralt [1,2], Claudia Alvargonzález[2].

[1] Universidad Autónoma de Madrid.

[2] Hospital Universitario de la Princesa. Madrid

[3] Instituto Bioética de la Universidad Francisco de Vitoria, Madrid

Abstract Introduction: Artificial intelligence (AI) systems can support data interpretation and risk estimation in intensive care units (ICUs), but a prediction does not determine a clinical decision. Objective: To propose an analytical framework connecting data uncertainty, model uncertainty, and axiological uncertainty—concerning values and preferences—with concrete clinical responses. Method: Conceptual analysis and a selective narrative review of clinical, technical, bioethical, and regulatory literature, using exploratory searches in PubMed/MEDLINE, and Google Scholar, this is not a systematic review. Discussion: The proposal is compared with prior taxonomies and applied to a hypothetical extubation case. A table links sources of uncertainty to assessment, communication, and clinical response. Conclusion: The framework is neither validated taxonomy nor evidence of efficacy; it is a heuristic tool that requires empirical evaluation to help interpret predictions without displacing clinical deliberation.

Resumen Introducción: Los sistemas de inteligencia artificial (IA) pueden apoyar la interpretación de datos y la estimación de riesgos en las unidades de cuidados intensivos (UCI), pero una predicción no determina por sí sola una decisión clínica. Objetivo: Proponer un marco analítico que conecte incertidumbre de los datos, incertidumbre del modelo e incertidumbre axiológica —relativa a valores y preferencias— con respuestas clínicas concretas. Método: Análisis conceptual y revisión narrativa selectiva de literatura clínica, técnica, bioética y normativa, con búsqueda exploratoria en PubMed/MEDLINE y Google Scholar, no es una revisión sistemática. Desarrollo: Se contrasta la propuesta con taxonomías previas y se aplica a un caso hipotético de extubación. Se presenta una tabla que vincula fuentes de incertidumbre, evaluación, comunicación y respuesta clínica. Conclusión: El marco no constituye una taxonomía validada ni demuestra eficacia; ofrece una herramienta heurística que debe evaluarse empíricamente para ayudar a interpretar las predicciones sin desplazar la deliberación clínica.

Palabras clave: inteligencia artificial; incertidumbre clínica; cuidados intensivos; calibración; predicción conformal; prudencia clínica; agencia moral.

Keywords: artificial intelligence; clinical uncertainty; intensive care; calibration; conformal prediction; practical wisdom; moral agency.

ISSN: 3020-7444
© Los Autores. Open Access, bajo Licencia Creative Commons (CC BY-NC).

1 Introducción: una predicción no es una decisión

Imaginemos un caso hipotético. Un paciente de 72 años ingresa en la UCI por shock séptico secundario a neumonía. Tras varios días de ventilación mecánica y soporte vasopresor, mejora de forma gradual. Un sistema de apoyo a la decisión estima un riesgo del 87 % de fracaso de extubación durante las próximas 12 horas. La cifra se presenta en una interfaz clara y destacada. No se trata de un resultado procedente de un modelo real ni de una cifra tomada de un estudio: es un ejemplo deliberadamente ilustrativo.

La estimación suscita preguntas clínicas relevantes: ¿con qué población se entrenó el modelo?, ¿qué desenlace define como fracaso?, ¿cómo se comportó en una validación externa?, ¿está calibrado en pacientes comparables?, ¿faltan datos o hay señales de que este caso se aparta de los datos de entrenamiento? Pero hay otras preguntas que una probabilidad, por sí sola, no responde. ¿Qué riesgos son aceptables para este paciente? ¿Qué importancia tienen sus preferencias? ¿Qué significa para él una recuperación satisfactoria? La hija recuerda que su padre no quería vivir conectado a máquinas sin una posibilidad razonable de recuperar una vida que reconociera como propia, aunque tampoco quería que se abandonara una opción clínicamente proporcionada. La familia no coincide en cómo interpretar ese deseo.

El ejemplo no plantea que la decisión correcta sea ignorar la predicción, ni seguirla de manera automática. El equipo debe valorar el estado clínico, los criterios vigentes para iniciar una prueba de respiración espontánea, los riesgos previsibles y los valores del paciente. La predicción puede aportar información; no sustituye esa deliberación. El problema aparece cuando la aparente precisión de una cifra oculta la incertidumbre del modelo y hace que una decisión discutible parezca una consecuencia obligada de los datos.

La medicina siempre ha tenido que decidir con conocimiento incompleto. Tonelli ha defendido el papel del juicio experto en una práctica clínica que no puede reducirse a la aplicación de resultados poblacionales a cada individuo [1]. Upshur, desde la filosofía de la medicina basada en la evidencia, también ha advertido sobre la dificultad de convertir reglas generales en respuestas suficientes para situaciones clínicas singulares [2]. La incorporación de IA no suprime este problema: añade nuevas fuentes de incertidumbre y nuevas formas de autoridad técnica. Los desafíos éticos del aprendizaje automático en salud y la necesidad de orientación profesional para sus usuarios han sido señalados, entre otros, por Char y colaboradores y por Smith y colaboradores [3,4].

La tesis de este artículo es que una IA clínicamente responsable no debe limitarse a ofrecer predicciones: debe comunicar sus límites de una manera que ayude al profesional a interpretar el resultado y a deliberar. La frase “comunicar lo que no sabe” es aquí una metáfora funcional, no una atribución de conciencia o autoconocimiento al sistema. Se refiere a que el sistema muestre de forma fiable las condiciones de uso, la incertidumbre estimada, la falta de evidencia y los escenarios en los que no debería formular una recomendación.

1.1 Enfoque y alcance de la propuesta

Se realizó una búsqueda narrativa exploratoria, no sistemática, en PubMed/MEDLINE y Google Scholar, complementada con búsquedas dirigidas de referencias citadas y consulta de fuentes primarias regulatorias (EUR-Lex, FDA y BOE). Se priorizaron (no de forma exclusiva) publicaciones sobre incertidumbre clínica, calibración, cambio de distribución, comunicación de incertidumbre en aprendizaje automático médico, interacción humano-IA, ética clínica y regulación. Se incluyeron trabajos revisados por pares, artículos metodológicos fundacionales, estudios clínicos ilustrativos y normas pertinentes; se excluyeron afirmaciones promocionales sin evaluación independiente y fuentes sin datos bibliográficos comprobables. La selección fue intencional y orientada por la pregunta conceptual, no exhaustiva; no se aplicó un proceso de cribado reproducible ni se pretende ofrecer una síntesis sistemática. La construcción del marco fue iterativa: se compararon las fuentes de incertidumbre, su reducibilidad, el tipo de evaluación posible y las decisiones que pueden orientar. Como criterio de transparencia inspirado en SANRA, se explicitan aquí la justificación, las fuentes consultadas, el razonamiento conceptual y los límites de la selección [5].

La aportación no consiste en descubrir tres clases universales ni mutuamente excluyentes. Propone una matriz pragmática para enlazar (a) problemas de medición y datos, (b) problemas de validez y fiabilidad del modelo, y (c) preguntas axiológicas sobre qué desenlaces importan al paciente. Las dos primeras se relacionan con distinciones

técnicas como incertidumbre aleatoria y epistémica; la tercera no es una incertidumbre estadística adicional, sino una cuestión normativa que condiciona la interpretación y el uso de las predicciones. El valor añadido es operativo: ante una misma decisión, la matriz obliga a preguntar qué puede comprobarse técnicamente, qué sigue sin saberse y qué requiere deliberación con el paciente o su representante.

2 Tres dimensiones de la incertidumbre en medicina crítica

La incertidumbre clínica cuenta con taxonomías previas. Han, Klein y Arora distinguen variedades de incertidumbre relevantes para la comunicación y la toma de decisiones sanitarias [6]. En aprendizaje automático, la distinción entre incertidumbre aleatoria y epistémica describe, respectivamente, variabilidad del fenómeno y limitaciones del conocimiento o del modelo [7]. Los trabajos sobre calibración, cambio de distribución y comunicación de incertidumbre muestran que una estimación puede ser útil en una muestra y perder fiabilidad al trasladarse a otro contexto [8–10]. Esta literatura impide presentar la tripartición de este artículo como una clasificación completamente nueva. Su contribución más modesta es conectar esas discusiones con la dimensión axiológica del cuidado y con una respuesta clínica concreta.

La relación entre categorías es deliberadamente no excluyente. Un dato ausente puede originar una predicción poco fiable; esa misma limitación puede aparecer como incertidumbre del modelo cuando afecta a la validez externa. La pregunta axiológica, por su parte, puede persistir incluso si el pronóstico estuviera bien calibrado: conocer mejor la probabilidad de supervivencia no decide qué desenlace es aceptable para esa persona. Esta distinción entre fuente técnica y pregunta normativa es el criterio organizador de la propuesta.

2.1 Incertidumbre de los datos: medir no es observarlo todo

Esta primera dimensión comprende datos ausentes, mediciones imprecisas, señales artefactadas, desfases temporales y variables que no representan adecuadamente el fenómeno clínico. En una UCI puede haber presión arterial invasiva amortiguada, muestras de laboratorio que no reflejan el estado actual o registros incompletos durante un traslado. La incertidumbre puede disminuir mediante una medición mejor, la revisión de la señal, la integración contextual de datos o la repetición de una prueba cuando esté clínicamente indicada.

La incertidumbre de los datos comprende ausencia, retraso, imprecisión, artefactos y falta de representatividad de las mediciones disponibles para el caso. Se solapa con la incertidumbre del modelo cuando la selección o calidad de los datos de entrenamiento limita la generalización; por ello, las dimensiones no son compartimentos estancos. En la matriz propuesta, se clasifica el problema según el punto donde puede intervenirse primero: revisar la medición y el contexto antes de atribuir el fallo al algoritmo.

2.2 Incertidumbre algorítmica: una cifra puede ser precisa y estar mal fundamentada

La incertidumbre del modelo incluye la especificación, el ajuste, la calibración, la validación externa, el rendimiento por subgrupos y la estabilidad ante cambios de distribución. No equivale a opacidad ni coincide exactamente con la incertidumbre epistémica: esta última es una distinción técnica sobre fuentes de incertidumbre, mientras que “incertidumbre del modelo” se usa aquí como categoría práctica que también abarca evidencia de validación insuficiente. La calibración debe evaluarse junto con discriminación y utilidad clínica, y no presumirse a partir de un AUROC elevado [11,8].

Conviene diferenciar discriminación y calibración. La discriminación indica hasta qué punto un modelo ordena a los pacientes según la probabilidad del desenlace; la calibración examina si las probabilidades predichas se corresponden, en promedio, con las frecuencias observadas en el contexto evaluado. Un AUROC elevado no garantiza una calibración adecuada, y la calibración observada en una cohorte no garantiza que se mantenga en otro hospital, en otro periodo o tras un cambio en la población. La literatura sobre calibración de redes neuronales documenta que los modelos modernos pueden producir probabilidades sobreconfiadas si no se evalúan y corrigen de forma apropiada [11].

La evidencia sobre predicción de extubación ilustra tanto el potencial como el límite de los resultados retrospectivos. Otaguro y colaboradores compararon modelos de aprendizaje automático en 117 pacientes; LightGBM alcanzó un AUC de 0,950 en aquella muestra [12]. Ese resultado justifica investigar modelos predictivos,

pero el tamaño de la cohorte, el diseño retrospectivo y el contexto concreto impiden convertirlo en una garantía de rendimiento en cualquier UCI. No demuestra que los modelos de extubación utilizados en la práctica tengan sistemáticamente ese rendimiento, ni que una predicción de riesgo baste para decidir si se inicia una prueba de respiración espontánea.

La distinción entre incertidumbre aleatoria y epistémica ayuda a describir mecanismos, pero no se corresponde uno a uno con las categorías prácticas de este artículo. La variabilidad aleatoria puede aparecer en mediciones y desenlaces; la incertidumbre epistémica puede afectar tanto a los datos como al modelo. El cambio de distribución puede degradar calibración y estimaciones de incertidumbre, por lo que estas también necesitan evaluación externa [7,9]. La matriz que proponemos no reemplaza esas distinciones: las complementa al preguntar qué respuesta clínica es apropiada.

La explicabilidad requiere la misma cautela. Herramientas de Inteligencia Artificial Explicable como LIME (*"Local Interpretable Model-agnostic Explanations"*) o SHAP *("SHapley Additive exPlanations")* pueden aportar información sobre la contribución local de variables a una salida, pero una explicación post hoc no equivale a una explicación causal ni demuestra que el modelo haya utilizado una razón clínicamente válida. En medicina, el riesgo es confundir una explicación visualmente convincente con una justificación suficiente para actuar. Ghassemi y colaboradores han advertido sobre las limitaciones de los enfoques actuales de IA explicable en salud [13]. Una explicación puede complementar la evaluación, nunca sustituir la validación del modelo.

2.3 ¿Qué puede aportar la cuantificación de incertidumbre?

Existen métodos técnicos que intentan representar algo más que una predicción puntual. Sin embargo, es importante describir correctamente qué ofrecen y qué no. La calibración evalúa la correspondencia entre probabilidades y frecuencias observadas en un conjunto de evaluación. Los intervalos de confianza, los intervalos de predicción y los conjuntos conformales responden a preguntas diferentes; no deben intercambiarse como si fueran expresiones equivalentes de "confianza" individual.

La predicción conformal puede producir conjuntos de etiquetas o intervalos para una variable de resultado con garantías de cobertura marginal bajo supuestos como la intercambiabilidad entre datos de calibración y nuevos casos. Una cobertura del 90 %, por ejemplo, significa —bajo esas condiciones y con la construcción correspondiente— que el procedimiento pretende incluir el resultado verdadero en aproximadamente el 90 % de los casos de la población considerada. No significa que la probabilidad real de un paciente concreto se encuentre dentro de un intervalo del tipo "entre el 65 % y el 92 %", ni que el método garantice la cobertura en cualquier hospital o ante cualquier cambio de distribución [14].

Este matiz no resta valor a la predicción conformal. Su interés es que puede hacer explícita la incertidumbre de ciertas predicciones y permitir al sistema abstenerse o devolver un conjunto de resultados cuando la información no basta para una respuesta única. Se han propuesto marcos de predicción conformal para decisiones relacionadas con la ventilación mecánica [15]. Pero la evaluación de un método en un conjunto de datos, simulación o análisis retrospectivo no equivale a demostrar beneficio clínico en el trabajo cotidiano de una UCI. El despliegue exige validar tanto el rendimiento estadístico como la seguridad, la utilidad y los efectos sobre las decisiones humanas.

Las redes neuronales bayesianas y otros métodos basados en conjuntos de modelos intentan representar la incertidumbre derivada de parámetros o especificaciones alternativas. Su desempeño depende de los supuestos, de la arquitectura, del entrenamiento y del modo en que se evalúan las estimaciones. La comparación de métodos bayesianos en apoyo a la decisión clínica muestra que la elección de priors y procedimientos de estimación importa; no es correcto presentar todas las aproximaciones como intercambiables o como una solución ya resuelta [16]. Por tanto, el sistema debe comunicar qué método se ha usado, cómo se ha validado y qué significado tiene la señal de incertidumbre que muestra.

Un estudio publicado en 2026 aplicó predicción conformal a estimaciones de GPT-4o sobre mortalidad intrahospitalaria usando notas clínicas de pacientes con insuficiencia renal aguda en MIMIC-III. Los autores observaron una cobertura global empírica del 90,4 % para un objetivo del 90 %, pero también una cobertura desigual entre las clases de muerte y supervivencia, además de sobreconfianza en predicciones erróneas [17]. El hallazgo es informativo y limitado: muestra una vía metodológica prometedora, no valida el uso general de un LLM como pronosticador clínico ni demuestra que la incertidumbre esté ya comunicada de forma adecuada en los productos comerciales.

En consecuencia, la pregunta no es si una arquitectura puede “conocer su propia ignorancia” en sentido humano. La pregunta técnica y evaluable es si el sistema detecta condiciones bajo las que sus salidas son menos fiables, si esa detección se mantiene en datos externos y si la interfaz presenta la advertencia de forma comprensible y accionable. Las estimaciones de incertidumbre necesitan su propia validación: no basta con añadir una barra, un color o un porcentaje de confianza a la interfaz.

2.4 Sesgo, distribución de los datos y justicia

La incertidumbre algorítmica también tiene una dimensión distributiva. Un modelo puede funcionar de manera distinta entre grupos demográficos o entre centros con poblaciones, recursos y prácticas clínicas diferentes. El cambio de distribución puede afectar tanto a la precisión como a la calibración. Por ello, la validez global de un modelo no permite concluir que sea igualmente fiable para cada subgrupo.

La validación externa del Epic Sepsis Model ofrece un ejemplo clínico de por qué no basta con su implantación o con el rendimiento comunicado por el desarrollador: en una cohorte retrospectiva de 27.697 pacientes y 38.455 hospitalizaciones, Wong y colaboradores observaron un AUROC de 0,63, inferior al descrito previamente por el desarrollador [18]. El estudio no prueba que todos los modelos de sepsis rindan mal; muestra que la evaluación independiente en el contexto de uso es necesaria y que el rendimiento publicado no debe darse por transferible sin comprobación.

Un estudio publicado en 2026 sobre predicción de mortalidad en UCI mediante modelos de lenguaje evaluó el uso de casos históricos similares como contexto para la predicción y analizó diferencias entre grupos. Sus resultados apuntan a mejoras de rendimiento y a una reducción de algunas disparidades, sin eliminar todas las diferencias, especialmente las relacionadas con la edad [19]. Debe leerse como evidencia de un estudio concreto, no como prueba de que el prompting por casos resuelva el sesgo algorítmico. Estos resultados muestran por qué el rendimiento agregado no elimina la necesidad de evaluar las diferencias entre grupos; además, una explicación post hoc no sustituye la validación del sistema [19,13].

Una interfaz clínicamente responsable debería indicar la población en la que el modelo fue desarrollado y validado, el desempeño conocido en subgrupos relevantes, los límites de los datos y los procedimientos de vigilancia tras la implantación. Cuando no existan datos suficientes, la ausencia de evidencia debe comunicarse como tal; no debe transformarse en una presunción de equivalencia entre grupos.

2.5 Incertidumbre axiológica: valores, preferencias y significado de los desenlaces

La tercera dimensión es diferente. La cuestión central no es una incertidumbre sobre el ser en sentido metafísico, sino la indeterminación normativa acerca de qué resultados considera aceptables una persona y cómo pondera supervivencia, autonomía, comunicación, dependencia y calidad de vida. Esta dimensión es cualitativamente distinta de la incertidumbre estadística; no se agrega a ella ni se estima con el mismo tipo de método. Se relaciona con la incertidumbre clínica descrita por Han y colaboradores y con la toma de decisiones compartida [6].

En España, esta deliberación debe respetar la autonomía del paciente y el marco de consentimiento por representación e instrucciones previas de la Ley 41/2002, especialmente sus artículos 8, 9 y 11 [20]. En la Comunidad de Madrid, la Ley 3/2005 regula el ejercicio del derecho a formular instrucciones previas [21]. Cuando el paciente no puede expresar su voluntad, la familia no sustituye automáticamente sus valores por los propios: debe ayudar a reconstruir, en la medida posible, la voluntad y preferencias del paciente. La limitación del esfuerzo terapéutico exige una valoración clínica individualizada, proporcionalidad y diálogo; no puede derivarse de una cifra algorítmica aislada.

La enfermedad grave afecta al cuerpo vivido, a la identidad y a las relaciones. La fenomenología de la enfermedad ha subrayado que la experiencia del paciente y la comprensión clínica no siempre se sitúan en el mismo horizonte de significado [22,23]. En la UCI, el delirio, la dependencia temporal de otras personas y la posible aparición del síndrome post-cuidados intensivos pueden transformar la experiencia del paciente y de su familia [24]. Un modelo puede estimar probabilidades de supervivencia, ventilación prolongada o discapacidad según definiciones específicas; no puede deducir solo de esas probabilidades qué vida considerará valiosa una persona determinada.

La incertidumbre biográfica no es un argumento contra la predicción. Es un recordatorio de su alcance. Los datos pueden informar sobre la posibilidad de ciertos resultados, pero la valoración de esos resultados requiere conocer, hasta donde sea posible, los valores, las preferencias, las relaciones y los proyectos del paciente. Cuando estos no están expresados o documentados, el equipo debe deliberar con la familia y aplicar los marcos clínicos y jurídicos pertinentes, distinguiendo con cuidado la voluntad conocida del paciente de las preferencias de sus allegados.

2.6. Aplicación del marco al caso hipotético

En el caso inicial, la estimación del 87 % debe abrir una comprobación, no cerrar la decisión. La matriz distingue tres preguntas que pueden coexistir: (1) ¿es fiable la información clínica introducida?; (2) ¿se ha validado el modelo para este paciente, horizonte temporal y centro?; y (3) ¿qué resultados consideraría aceptables el paciente, a la luz de sus preferencias conocidas? Cada pregunta requiere una respuesta distinta. La tabla 1 resume cómo traducirlas en una conducta clínica prudente.

Tabla 1. Matriz heurística para analizar la estimación hipotética del 87 % de fracaso de extubación. Las categorías pueden solaparse; la tabla no constituye un instrumento validado ni sustituye los protocolos clínicos.

Dimensión	Indicadores y evaluación	Comunicación	Respuesta clínica
Datos y medición	Datos ausentes/desactualizados, artefactos o desfase temporal; revisar señal, historia y procedencia.	Advertir de datos incompletos o no fiables; no presentar la cifra como plenamente aplicable.	Corregir o repetir la medición cuando proceda; no interpretar automáticamente si faltan datos críticos.
Modelo y evidencia	Calibración local, validación externa, rendimiento por subgrupos, cambio de distribución y representatividad.	Indicar población, horizonte temporal, evidencia externa y límites conocidos.	Activar revisión humana o abstención si no se ha demostrado fiabilidad en ese contexto.
Valores y preferencias	Objetivos del paciente, tolerancia al riesgo, autonomía y significado de supervivencia/dependencia; revisar instrucciones previas.	Explicar que el pronóstico no determina por sí solo el objetivo asistencial; explicitar preferencias conocidas y dudas.	Deliberar con el paciente o representante sobre proporcionalidad y objetivos, conforme al marco jurídico.

La matriz aporta una utilidad analítica concreta: impide tratar todas las dudas como un único “porcentaje de incertidumbre”. Si la señal es defectuosa, se revisa la medición; si el modelo no se ha validado en el contexto, se limita su uso y se activa revisión; si lo incierto es qué resultado desea el paciente, se delibera sobre valores y proporcionalidad. En este último supuesto, añadir datos o mejorar la calibración no resuelve por sí solo la pregunta normativa.

3 Agencia moral: de la predicción a la deliberación

La prudencia práctica —phronesis, en el vocabulario aristotélico— no es una intuición opuesta a la ciencia. Es la capacidad de deliberar bien sobre lo que conviene hacer en circunstancias concretas, integrando conocimiento, experiencia, fines y valores [25]. En medicina, esta idea se ha desarrollado desde una ética de la virtud centrada en la finalidad de la práctica clínica y en la responsabilidad del profesional hacia la persona enferma [26]. La competencia técnica es indispensable; la prudencia ayuda a decidir cómo emplearla y cuándo la evidencia disponible no determina por sí sola la acción.

La IA puede fortalecer esa deliberación si sintetiza información difícil de manejar, detecta patrones ignorados o alerta sobre riesgos que merecen evaluación. También puede debilitarla si la salida algorítmica se interpreta como una orden, si el clínico siente que debe justificar cualquier discrepancia con la máquina o si la organización presenta una predicción como sustituto de la responsabilidad. Este riesgo no implica que todos los profesionales obedezcan automáticamente a la IA, ni que la responsabilidad pueda transferirse literalmente al software. Es un riesgo de diseño, de contexto institucional y de interacción humana.

La literatura sobre sesgo de automatización identifica el riesgo de que los usuarios acepten recomendaciones automatizadas o no detecten errores, aunque su frecuencia y sus factores moduladores dependen de la tarea y del diseño [27]. La idea de "moral crumple zone" advierte, además, que la responsabilidad puede concentrarse injustamente en el profesional situado al final de una cadena sociotécnica, aun cuando su control real sea limitado [28]. Estos marcos permiten formular el riesgo como hipótesis fundada, no afirmar que la IA cause necesariamente obediencia automática o malestar moral en la UCI. La evidencia directa sobre ese efecto en intensivistas sigue siendo insuficiente.

La agencia moral clínica no consiste en decidir al margen de los datos, sino en conservar la capacidad de examinarlos, ponderarlos y responder por la decisión. El profesional debe poder preguntar por la población de validación, detectar una discordancia con la evolución observada, contrastar la recomendación con otros datos y considerar los valores del paciente. Para que esa revisión sea real, la interfaz no debe ocultar los límites del sistema ni presentar la recomendación como inequívoca cuando no lo es.

3.1 Malestar moral y residuo moral

El malestar moral puede surgir cuando un profesional considera que una decisión sería más adecuada, pero encuentra obstáculos clínicos, organizativos, jerárquicos o normativos para llevarla a cabo. La literatura ha descrito el malestar moral y el residuo moral en contextos sanitarios [29,30], pero la evidencia citada no permite afirmar que la IA los intensifique de manera general ni demostrar ese efecto específicamente en UCI. En este artículo, esa posibilidad se presenta como una hipótesis que merece evaluación: podría aparecer si una recomendación algorítmica se usa como justificación automática, si no hay tiempo para deliberar o si el profesional carece de autoridad real para cuestionarla.

Proponemos considerar el residuo moral como una señal para revisar el proceso de decisión, no como prueba inequívoca de que el juicio del profesional era correcto o de que la tecnología causó el daño. El malestar también puede reflejar sufrimiento, desacuerdos razonables, incertidumbre persistente o condiciones de trabajo inadecuadas. Su valor está en abrir una revisión: ¿se entendía el alcance de la predicción?, ¿se expresaron las preferencias del paciente?, ¿hubo oportunidad para una segunda opinión?, ¿la institución ofreció un marco claro de responsabilidad? Las sesiones de deliberación y debriefing ético pueden contribuir a esa revisión sin patologizar la duda ni idealizarla.

La responsabilidad tampoco desaparece por el hecho de que el modelo influya en una decisión. El profesional sigue teniendo deberes de evaluación y de documentación, mientras que desarrolladores, proveedores, gestores y organizaciones sanitarias tienen responsabilidades relacionadas con validación, adquisición, formación, vigilancia y condiciones de uso. Van Baalen y Boon han defendido la importancia de la responsabilidad epistemológica en la práctica basada en evidencia [31]; la implantación de IA amplía esa responsabilidad al exigir que los usuarios comprendan los límites relevantes del sistema, sin que ello convierta a cada clínico en experto en aprendizaje automático.

4 Comunicar la incertidumbre de forma clínicamente útil

Mostrar más información no siempre mejora una decisión. La comunicación debe ser correcta desde el punto de vista estadístico, comprensible para el profesional y compatible con el flujo real de trabajo. En un ensayo aleatorizado con 30 farmacéuticos, la presentación de ayuda de IA y de información sobre su incertidumbre modificó los patrones de interacción y aumentó en algunas condiciones el tiempo dedicado a revisar las imágenes originales; los autores señalaron que la transparencia puede añadir carga cognitiva si no se diseña adecuadamente [32]. El estudio se realizó en verificación de medicación, no en UCI, por lo que sus resultados informan sobre diseño humano-IA, pero no demuestran efectos idénticos en intensivistas.

Una comunicación responsable debería distinguir al menos cinco elementos: (1) la estimación del riesgo y el desenlace al que se refiere; (2) la calidad de la calibración en la población evaluada; (3) los límites de aplicabilidad, incluidos datos ausentes o señales de distribución atípica; (4) el rendimiento externo y por subgrupos, cuando exista evidencia suficiente; y (5) la acción recomendada, sus fundamentos y la opción de abstención cuando la fiabilidad sea insuficiente. No es necesario mostrar todos los indicadores con idéntico nivel de detalle en cada pantalla; el diseño debería priorizar información accionable, con acceso a la documentación ampliada.

Una probabilidad puntual del 87 % no debe etiquetarse como “87 % de confianza de la IA” si lo que representa es una estimación de riesgo. Tampoco debe añadirse un intervalo numérico sin explicar cómo se calcula y qué cubre. En una herramienta de predicción de resultados, podría ser más útil indicar que el riesgo estimado procede de una cohorte retrospectiva, que la calibración no se ha evaluado en el centro del usuario o que el perfil está insuficientemente representado en los datos. Si el modelo no puede evaluar una situación de forma fiable, una salida del tipo “evidencia insuficiente para generar una recomendación” puede resultar más segura que una cifra engañosamente precisa.

La explicación de variables debe presentarse con límites claros. Decir que determinados factores contribuyeron a la salida no demuestra que sean causas del desenlace ni que modificar uno de ellos cambie el resultado. De igual modo, un indicador “fuera de distribución” no identifica siempre todos los casos no fiables; es una señal de alerta que necesita validación. La interfaz debe evitar una falsa sensación de seguridad tanto por exceso de detalle técnico como por simplificación excesiva.

4.1 El marco regulatorio: requisitos relevantes y una cautela

El Reglamento europeo de inteligencia artificial establece obligaciones diferenciadas según la categoría del sistema y su contexto de uso, e incluye requisitos de gestión de riesgos, gobernanza de datos, información para quienes despliegan el sistema, supervisión humana, precisión, robustez y ciberseguridad para los sistemas de alto riesgo que entran en su ámbito [33]. En el sector sanitario, la clasificación depende de la finalidad prevista y del encaje en el marco aplicable a los productos sanitarios; no es correcto afirmar que cualquier herramienta de apoyo a la decisión en una UCI sea automáticamente un sistema de alto riesgo o tenga una clasificación regulatoria idéntica.

La regulación europea no convierte por sí sola los intervalos conformales, la descomposición bayesiana u otra representación numérica concreta de la incertidumbre en un requisito universal de interfaz. La transparencia legal es importante, pero no determina una única forma de hacer comprensible el riesgo al profesional. En Estados Unidos, la FDA publica recomendaciones para dispositivos médicos habilitados por IA, incluidos planes predeterminados de control de cambios y gestión de modificaciones durante su ciclo de vida [34]. Esas orientaciones son relevantes para la seguridad y el mantenimiento del rendimiento; no deben interpretarse como un aval automático a cada aplicación clínica ni como sustituto de la evaluación local.

Las guías de reporte también ofrecen un marco útil. TRIPOD+AI orienta la comunicación de estudios que desarrollan o evalúan modelos de predicción clínica [35], mientras que DECIDE-AI se centra en la evaluación clínica temprana, incluidos factores humanos, seguridad y utilidad real de sistemas de apoyo basados en IA [36]. Son guías de reporte y evaluación, no pruebas de eficacia de un sistema concreto. Su aportación para nuestro argumento es práctica: la comunicación de resultados, límites y condiciones de uso debe formar parte del proceso de evaluación desde el inicio.

4.2 Un diseño que apoye el juicio clínico

A partir de la matriz de la tabla 1, proponemos un único conjunto de requisitos para la interfaz y su gobernanza: identificar desenlace y horizonte temporal; describir la población y la validación disponible; mostrar limitaciones relevantes, datos ausentes y alertas de cambio de distribución; definir umbrales de revisión o abstención antes del despliegue; y evaluar comprensión, carga cognitiva y consecuencias clínicas con usuarios reales. Estos requisitos reúnen los elementos de comunicación y los principios de diseño expuestos anteriormente, evitando tratarlos como dos listas independientes. No se propone un umbral universal de abstención: debe preespecificarse para cada tarea y evaluarse prospectivamente.

Una evaluación inicial podría comparar, en simulación clínica, una interfaz con probabilidad puntual frente a otra que añade población de validación, calibración, limitaciones y opción de abstención. Variables medibles serían: exactitud al interpretar el riesgo, calibración de la confianza del profesional, proporción de recomendaciones aceptadas y rechazadas, detección de casos no fiables, tiempo de decisión, carga cognitiva, concordancia con protocolos y calidad de la justificación documentada. En un estudio posterior, la evaluación prospectiva debería añadir decisiones modificadas, eventos adversos, diferencias entre subgrupos y resultados centrados en pacientes. Estas variables son una agenda de investigación propuesta, no resultados obtenidos en este artículo.

La decisión de usar una probabilidad, un rango, un conjunto de predicciones o una indicación cualitativa debe depender del método validado y de la tarea clínica. Una interfaz no debería utilizar la etiqueta “intervalo de confianza” como término genérico. Tampoco debe dar por hecho que la exposición a más información genera automáticamente una confianza mejor calibrada. El estudio de Tsai y colaboradores muestra que los efectos de comunicar incertidumbre dependen también de cómo se integra esa información en la tarea [32].

5 Implicaciones para la práctica, la formación y la gobernanza

Las consecuencias de esta propuesta son operativas. La comunicación de incertidumbre no se resuelve con una nota legal al pie de la pantalla ni con una sesión aislada sobre IA. Requiere vincular formación, diseño, evaluación clínica y responsabilidad institucional.

5.1 Formación clínica

La formación de residentes y especialistas debería incluir alfabetización algorítmica proporcionada a la tarea: diferencia entre discriminación y calibración, lectura crítica de validaciones, sesgo de selección, cambio de distribución, límites de las explicaciones post hoc y significado de los indicadores de incertidumbre. El objetivo no es convertir a cada intensivista en ingeniero de aprendizaje automático, sino prepararlo para reconocer qué preguntas debe formular al proveedor y cuándo una salida no basta para fundamentar una decisión.

La formación filosófica y bioética puede ayudar a interpretar de forma explícita la diferencia entre estimar un resultado y decidir qué resultado es aceptable. La *phronesis* no se enseña solo como una definición teórica: se cultiva mediante deliberación supervisada sobre casos, revisión de decisiones inciertas, comunicación con familias y contraste de perspectivas clínicas. Los escenarios de simulación pueden incluir recomendaciones algorítmicas discordantes, datos incompletos, modelos aplicados fuera de su ámbito y casos en los que las preferencias del paciente sean conocidas de forma parcial.

5.2 Práctica y deliberación a pie de cama

Antes de utilizar una predicción que pueda modificar la conducta clínica, el equipo debería poder responder a tres preguntas: ¿qué predice exactamente el sistema?, ¿qué evidencia sostiene su uso en este paciente y este entorno?, y ¿cómo se integra el resultado con la exploración, la evolución y las preferencias del paciente? El registro clínico debería dejar constancia de la información relevante, las discrepancias sustantivas y las razones de la decisión. No se trata de documentar cada cálculo, sino de hacer auditable la deliberación en decisiones de consecuencias importantes.

La supervisión humana debe ser real y contar con tiempo, información y autoridad suficientes. Una firma humana al final de una recomendación no constituye supervisión efectiva si el flujo de trabajo hace inviable revisarla. Las instituciones deberían disponer de procesos para solicitar revisión experta, reportar comportamientos inesperados, reevaluar modelos y suspender su uso cuando se detecte deterioro relevante. La posibilidad de apartarse de una recomendación debe acompañarse de procedimientos transparentes de documentación y revisión; no sería prudente prometer inmunidad legal genérica a quien la anule, sino establecer una gobernanza justa que valore el razonamiento clínico documentado y la actuación de buena fe.

En el caso hipotético del inicio, el 87 % no determina la conducta. El equipo debe comprobar la validez de la predicción, aplicar los criterios clínicos pertinentes y conversar sobre los objetivos de la atención. La IA puede ayudar a formular preguntas o reconocer factores de riesgo; no tiene autoridad para resolver por sí sola el significado que el paciente atribuye a la supervivencia.

5.3 Gobierno institucional y evaluación

La adquisición de un sistema de IA debería involucrar a profesionales clínicos, especialistas en informática médica, bioética, seguridad, protección de datos y representantes de pacientes cuando sea apropiado. Antes de la implantación deben definirse la finalidad prevista, los usuarios, las acciones que puede activar, los criterios de abstención, el procedimiento de actualización y las condiciones para retirar el sistema. La validación debe contemplar el rendimiento en el centro de uso, la calibración, el desempeño por grupos relevantes y las consecuencias de falsos positivos y falsos negativos.

Tras el despliegue, es necesario vigilar cambios en la población, en las prácticas terapéuticas, en los equipos de monitorización y en los datos disponibles. El rendimiento inicial no garantiza estabilidad indefinida. Los incidentes deben poder comunicarse y analizarse sin atribuirlos automáticamente a un error individual; en ocasiones revelarán una interfaz poco clara, una formación insuficiente, una validación inadecuada o una descoordinación organizativa.

La revista INTELETICA ha publicado análisis de los retos epistemológicos y éticos de la IA generativa en salud mental, incluido el riesgo de simplificar la complejidad subjetiva y desplazar el juicio clínico [37,38]. En el contexto de la UCI, esta perspectiva debe conectarse con controles técnicos verificables: no basta con declarar que el humano es central si las herramientas y los procesos no permiten comprender los límites del modelo, expresar desacuerdo y revisar sus consecuencias.

6. Límites de la propuesta y agenda de investigación

Este trabajo es un análisis conceptual basado en una revisión narrativa selectiva. No permite estimar la frecuencia con que los sistemas de IA en UCI comunican incertidumbre ni demostrar efectos sobre decisiones, resultados clínicos o malestar moral. La búsqueda no fue sistemática y puede haber omitido literatura relevante; la selección estuvo orientada por la pregunta conceptual y no se sometió a evaluación formal por duplicado. La matriz propuesta es heurística y no ha sido validada por expertos, pacientes ni pruebas de usabilidad. La denominación “incertidumbre del modelo” no es una métrica única, y la dimensión axiológica no es cuantificable en los mismos términos que la incertidumbre estadística.

La agenda de evaluación debe comenzar por estudiar si profesionales y pacientes distinguen correctamente los tres tipos de preguntas y si la matriz mejora la identificación de acciones apropiadas. Después pueden compararse interfaces con diferentes formas de comunicar riesgo, incertidumbre y abstención, midiendo comprensión, calibración de la confianza, carga cognitiva y conducta clínica. Finalmente, estudios prospectivos deberían evaluar utilidad, seguridad, equidad y resultados centrados en el paciente en contextos reales. La hipótesis de que una comunicación explícita de incertidumbre favorezca la deliberación es plausible y está apoyada por literatura previa, pero su beneficio en UCI requiere demostración.

7. Conclusión

La IA puede ayudar a los equipos de cuidados intensivos a manejar grandes volúmenes de información y a estimar riesgos difíciles de calcular mediante la valoración aislada de cada variable. Pero el rendimiento predictivo no elimina la incertidumbre, y una cifra precisa no equivale a una decisión justificada. La incertidumbre de los datos, la incertidumbre algorítmica y la incertidumbre biográfica describen problemas relacionados, pero no intercambiables: cada uno exige una respuesta distinta.

Un sistema clínico de IA responsable debe comunicar el significado y los límites de sus resultados, describir la evidencia que respalda su uso y permitir que el profesional reconozca cuándo la predicción no basta. A su vez, el clínico debe conservar la capacidad y la responsabilidad de integrar esa información con la evolución, el contexto y los valores del paciente. Esta relación no enfrenta filosofía y tecnología: exige que el rigor técnico y la prudencia clínica se corrijan y se sostengan mutuamente.

La finalidad no es que el algoritmo sustituya la incertidumbre por una falsa certeza, ni que el juicio humano quede exento de examen. Es que cada uno contribuya donde resulta pertinente: la IA, aportando estimaciones evaluables y límites explícitos; el equipo clínico, deliberando sobre decisiones concretas y haciéndose responsable de ellas; y la institución, garantizando formación, validación, revisión y condiciones de trabajo seguras. Comunicar

los límites de la predicción es así una condición importante —aunque no suficiente por sí sola— para que la IA apoye una práctica clínica técnicamente rigurosa y moralmente responsable.

Declaraciones

Financiación: Este trabajo no ha recibido financiación específica.

Conflictos de interés: Los autores declaran no tener conflictos de interés relacionados con este trabajo.

Uso de IA generativa: Se ha utilizado Grammarly para la revisión gramatical y de expresión. Elicit para la organización y gestión de referencias bibliográficas. Los autores han realizado la verificación del contenido y son responsables del contenido del manuscrito.

Contribución de los autores:

Alfonso Canabal Berlanga: conceptualización, metodología, redacción del borrador original y revisión y edición del manuscrito en su versión final

Beatriz Abad Santamaría: Organización y gestión de la información. Aprobación de la versión final

Juan Antonio Sanchez Giralt. Elaboración de la tabla, análisis de la información. Aprobación de la versión final

Claudia Alvargonzález Riera: Resumen del contenido de las referencias y traducción de los contenidos clave. Aprobación de la versión final.

Referencias

[1] Tonelli MR. In defense of expert opinion. Acad Med. 1999;74(11):1187-1192. doi:10.1097/00001888-199911000-00010.

[2] Upshur REG. Looking for rules in a world of exceptions: reflections on evidence-based practice. Perspect Biol Med. 2005;48(4):477-489. doi:10.1353/pbm.2005.0098.

[3] Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. N Engl J Med. 2018;378(11):981-983. doi:10.1056/NEJMp1714229.

[4] Smith H, Downer J, Ives J. Clinicians and AI use: where is the professional guidance? J Med Ethics. 2024;50(7):437-441. doi:10.1136/jme-2022-108831.

[5] Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. Res Integr Peer Rev. 2019;4:5. doi:10.1186/s41073-019-0064-8.

[6] Han PKJ, Klein WMP, Arora NK. Varieties of uncertainty in health care: a conceptual taxonomy. Med Decis Making. 2011;31(6):828-838. doi:10.1177/0272989X10393976.

[7] Hüllermeier E, Waegeman W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. Mach Learn. 2021;110:457-506. doi:10.1007/s10994-021-05946-3.

[8] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW; Topic Group ‘Evaluating diagnostic tests and prediction models’ of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17:230. doi:10.1186/s12916-019-1466-7.

[9] Ovadia Y, Fertig E, Ren J, Nado Z, Sculley D, Nowozin S, et al. Can you trust your model’s uncertainty? Evaluating predictive uncertainty under dataset shift. Adv Neural Inf Process Syst. 2019;32:13991-14002.

[10] Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digit Med. 2021;4:4. doi:10.1038/s41746-020-00367-3.

[11] Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research. 2017;70:1321-1330.

[12] Otaguro T, Tanaka H, Igarashi Y, Tagami T, Masuno T, Yokobori S, et al. Machine learning for prediction of successful extubation of mechanical ventilated patients in an intensive care unit: a retrospective observational study. J Nippon Med Sch. 2021;88(5):408-417. doi:10.1272/jnms.JNMS.2021_88-508.

[13] Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11):e745-e750. doi:10.1016/S2589-7500(21)00208-9.

[14] Angelopoulos AN, Bates S. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. Found Trends Mach Learn. 2023;16(4):494-591. doi:10.1561/2200000101.

[15] Eghbali N, Alhanai T, Ghassemi MM. Distribution-free uncertainty quantification in mechanical ventilation treatment: a conformal deep Q-learning framework. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2025;39(27):27960-27968. doi:10.1609/aaai.v39i27.35013.

[16] Blattmann M, Lindenmeyer A, Franke S, Neumuth T, Schneider D. Implicit versus explicit Bayesian priors for epistemic uncertainty estimation in clinical decision support. PLOS Digit Health. 2025;4(7):e0000801. doi:10.1371/journal.pdig.0000801.

[17] Shah-Mohammadi F, Millar A, Facelli JC, Gouripeddi R. Understanding uncertainty in large language model predictions of early death in critically ill patients: a conformal prediction approach. JAMIA Open. 2026;9(3):ooag108. doi:10.1093/jamiaopen/ooag108.

[18] Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med. 2021;181(8):1065-1070. doi:10.1001/jamainternmed.2021.2626.

[19] Zhang G, Long Y, Zhou Y, Zhang Y, Hong S. Improving fairness of large language model-based ICU mortality prediction via case-based prompting. Artif Intell Med. 2026;183:103541. doi:10.1016/j.artmed.2026.103541.

[20] Jefatura del Estado. Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica. BOE. 2002;274. Available from: https://www.boe.es/eli/es/l/2002/11/14/41/con.

[21] Comunidad de Madrid. Ley 3/2005, de 23 de mayo, por la que se regula el ejercicio del derecho a formular instrucciones previas en el ámbito sanitario y se crea el registro correspondiente. BOE. 2005;269. Available from: https://www.boe.es/eli/es-md/l/2005/05/23/3/con.

[22] Toombs SK. The meaning of illness: a phenomenological approach to the patient-physician relationship. J Med Philos. 1987;12(3):219-240. doi:10.1093/jmp/12.3.219.

[23] Svenaeus F. The body uncanny—further steps towards a phenomenology of illness. Med Health Care Philos. 2000;3(2):125-137. doi:10.1023/A:1009920011164.

[24] Lobo-Valbuena B, Sánchez Roca MD, Regalón Martín MP, et al. Post-intensive care syndrome: ample room for improvement. Med Intensiva (Engl Ed). 2021;45(8):e43-e46. doi:10.1016/j.medine.2021.08.001.

[25] Aristotle. Nicomachean Ethics. 2nd ed. Translated by Irwin T. Indianapolis: Hackett Publishing; 1999.

[26] Pellegrino ED, Thomasma DC. The Virtues in Medical Practice. New York: Oxford University Press; 1993.

[27] Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012;19(1):121-127. doi:10.1136/amiajnl-2011-000089.

[28] Elish MC. Moral crumple zones: cautionary tales in human-robot interaction. Engag Sci Technol Soc. 2019;5:40-60. doi:10.17351/ests2019.260.

[29] Epstein EG, Hamric AB. Moral distress, moral residue, and the crescendo effect. J Clin Ethics. 2009;20(4):330-342.

[30] Epstein EG, Delgado S. Understanding and addressing moral distress. Online J Issues Nurs. 2010;15(3):Manuscript 1. doi:10.3912/OJIN.Vol15No03Man01.

[31] van Baalen S, Boon M. An epistemological shift: from evidence-based medicine to epistemological responsibility. J Eval Clin Pract. 2015;21(3):433-439. doi:10.1111/jep.12282.

[32] Tsai CC, Kim JY, Chen Q, Rowell B, Yang XJ, Kontar R, et al. Effect of artificial intelligence helpfulness and uncertainty on cognitive interactions with pharmacists: randomized controlled trial. J Med Internet Res. 2025;27:e59946. doi:10.2196/59946.

[33] European Parliament and Council of the European Union. Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Off J Eur Union. 2024;L 2024/1689. Available from: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

[34] U.S. Food and Drug Administration. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions: Guidance for Industry and Food and Drug Administration Staff. August 2025. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence.

[35] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378.

[36] Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al; DECIDE-AI Expert Group. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022;377:e070904. doi:10.1136/bmj-2022-070904.

[37] Canabal A, Keough Delgado EA, Alonso González M. Inteligencia artificial en la salud mental: oportunidades, dificultades y cuestionamientos. INTELETICA. 2024;1(2):16-26. doi:10.4114/INTELETICA.vol1iss21pp16-26.

[38] Sotelo Martín JA. La inteligencia artificial generativa en los trastornos de personalidad: potencialidades, riesgos clínicos, epistemológicos y éticos. INTELETICA. 2026;3(5):1-12. doi:10.4114/inteletica.vol3iss5pp1-12.