"Trazabilidad Ética y Visión Personalista en el Diseño de Inteligencia Artificial: Un Llamado a la Incidencia Bioética en la Tecnología Emergente"

Autor: Marcos León
Ingeniero en Electrónica con más de 13 años de experiencia en tecnología.
Analista de datos, especializado en sistemas estratégicos con impacto social. Fundador de MLDATA y del Movimiento Libre de Datos Abiertos para la Transformación Activa.
Presidente de Generación Provida, miembro del Foro de Diálogo Civil y colaborador de la Asociación de Estudios Bioéticos. Autor del libro "Autonomía Progresiva: El Desmantelamiento Silencioso del Marco Legal en Paraguay", referente en análisis crítico sobre bioética y tecnología.

1. Introducción

1.1. Los sistemas ya deciden por nosotros

La inteligencia artificial ya no es una herramienta de laboratorio ni una promesa de futuro: es un actor presente en las decisiones sociales. Cada vez con mayor frecuencia, algoritmos clasifican solicitudes de servicios públicos, establecen alertas en instituciones sanitarias, o estructuran prioridades clínicas. Esto implica una transformación silenciosa pero profunda: **los sistemas automáticos han comenzado a intervenir en espacios que hasta hace poco eran dominios exclusivos del juicio humano**.

Paolo Benanti (2021) advierte que estas tecnologías no son neutras. A pesar de su apariencia técnica, la IA encarna valores, visiones y criterios preestablecidos. Como señala, "la neutralidad de los algoritmos es una ilusión retórica": cada sistema inteligente es una forma de consolidar decisiones humanas anteriores, y esas decisiones son, en sí mismas, **actos morales codificados** (Benanti, 2021, p. 45). La ética, entonces, no puede esperar a que surjan los conflictos. Debe estar presente **desde el diseño inicial**, donde se determinan los umbrales, las prioridades y los criterios operativos de los sistemas.

Este cambio obliga a repensar también el papel del profesional bioético. Si la IA participa en la toma de decisiones que afectan directamente a seres humanos, entonces **la voz ética debe estar presente en su arquitectura**. De lo contrario, el sistema reproducirá una moral por omisión: aquella que surge cuando nadie asume responsabilidad moral.

1.2. La ilusión de la neutralidad y el sesgo como estructura

Uno de los errores más comunes en el discurso público sobre IA es la creencia de que los sistemas automatizados pueden llegar a ser "objetivos" si son lo suficientemente entrenados. Esta idea ignora un hecho fundamental: **todo modelo**

de IA aprende a partir de datos seleccionados, etiquetados y jerarquizados por seres humanos, lo que implica que los sesgos no son anomalías técnicas, sino componentes estructurales inevitables.

Como sostienen Floridi y Cowls (2019), el debate no debería girar en torno a la eliminación de sesgos —una meta imposible—, sino sobre **qué tipo de sesgos estamos dispuestos a permitir, y bajo qué principios éticos**. En su marco unificado de principios para la IA en la sociedad, afirman que "toda decisión algorítmica debe estar guiada por valores explícitos: justicia, beneficencia, no maleficencia, autonomía y explicabilidad" (Floridi & Cowls, 2019, p. 2). Esto implica que el diseño técnico ya es, en sí mismo, **un acto moral**.

Desde la bioética personalista, esta constatación cobra una urgencia particular. Si no se define el sesgo desde una visión centrada en la dignidad, lo hará otra lógica: la eficiencia, el rendimiento económico, la rentabilidad clínica, o la estadística bruta. El sesgo es inevitable; **la omisión ética no lo es**.

1.3. Una oportunidad: codiseñar sistemas con criterio moral

Este escenario no debe paralizar. Al contrario: ofrece una oportunidad inédita para que los profesionales de la bioética personalista **intervengan en un espacio donde se están jugando decisiones morales con consecuencias prácticas**. La pregunta no es si se puede influir, sino si se está dispuesto a asumir el desafío de codiseñar, participar, y formar parte activa del ciclo de desarrollo de tecnologías que afectan vidas humanas.

Esto implica un cambio de paradigma. La bioética ya no puede limitarse al juicio posterior, a la supervisión externa o a la crítica institucional. Debe integrarse como **competencia estructural del desarrollo técnico**. Para lograrlo, es necesario traducir principios fundamentales —como la inviolabilidad de la persona, la centralidad del consentimiento, o la no instrumentalización del ser humano— en estructuras funcionales, criterios de diseño, protocolos y umbrales.

Este trabajo se inscribe dentro de ese desafío. A partir de un caso concreto (la falta de trazabilidad ética en decisiones clínicas en Paraguay), se mostrarán las consecuencias de sistemas que actúan sin una arquitectura moral visible. Luego se explorarán propuestas técnicas y éticas para **reinsertar al ser humano como fin — y no como insumo— dentro del sistema**. No se busca controlar el futuro, sino **garantizar que la dignidad esté escrita en sus códigos**.

2. Los sesgos no son errores: son estructuras con dirección ética

2.1. El mito de la objetividad algorítmica

En el imaginario popular y, a menudo, también en el discurso institucional, la inteligencia artificial es percibida como una herramienta de alta precisión, capaz de eliminar errores humanos y producir decisiones objetivas. Esta percepción es profundamente errónea. En realidad, **los sistemas de IA no operan en un vacío ético, sino que heredan los valores, criterios y omisiones de sus diseñadores**.

Como señala Benanti (2021), "la inteligencia artificial no introduce una ética nueva, sino que prolonga decisiones humanas que fueron tomadas previamente, a menudo sin conciencia moral explícita" (p. 87). Cada línea de código, cada set de datos, cada variable ponderada en el algoritmo responde a elecciones que **reflejan visiones antropológicas y sociales específicas**, aunque no siempre sean reconocidas como tales.

Además, los conjuntos de datos —fuente fundamental para el entrenamiento de estos sistemas— no solo reflejan el mundo como es, sino como fue documentado. Esto significa que **los sesgos sociales, históricos y culturales están incrustados en los datos mismos**, y son replicados por el sistema sin capacidad de interpretación crítica. No hay "punto de vista neutral" desde el cual los sistemas puedan decidir de forma absolutamente justa.

2.2. El sesgo como problema técnico y como oportunidad moral

Si se acepta que todo sistema automatizado tiene sesgos, la pregunta crucial ya no es cómo eliminarlos —porque eso es técnicamente inviable—, sino **cómo orientarlos conscientemente hacia criterios éticos y humanizantes**. Esto implica un cambio fundamental en la forma de pensar el diseño tecnológico: de evitar el sesgo, a **modelarlo moralmente**.

En este punto, la propuesta de Luciano Floridi y Josh Cowls (2019) es especialmente útil. Ellos plantean que las decisiones algorítmicas deben estructurarse en base a cinco principios éticos: beneficencia, no maleficencia, autonomía, justicia y explicabilidad. Estos principios no deben ser añadidos como capas posteriores al diseño, sino **integrados como parámetros estructurales** que orienten el comportamiento del sistema.

Desde la bioética personalista, esta orientación adquiere una forma concreta: **los sesgos deben proteger al ser humano, no simplificarlo**. Un sesgo ético, en este sentido, podría incluir la priorización de casos vulnerables, la obligación de requerir revisión humana en situaciones ambiguas o la trazabilidad de cada decisión automatizada para poder reconstruir el razonamiento detrás de una recomendación clínica.

Es decir, no se trata solo de evitar el daño, sino de asegurar que **cada sesgo incorporado al sistema responda a una visión antropológica robusta**, donde el ser humano no sea una variable más, sino el criterio último de decisión.

2.3. El rol ético en la arquitectura del sistema

Aceptar la inevitabilidad del sesgo también redefine el papel del profesional ético. Ya no basta con evaluar las consecuencias de los sistemas una vez implementados. Lo verdaderamente necesario es **participar en el diseño mismo del criterio operativo de los algoritmos**. Esto exige una ética con capacidad arquitectónica, capaz de traducir principios morales en decisiones técnicas.

Como subraya Benanti (2023), "la única manera de garantizar que la IA respete la dignidad humana es que la dignidad esté codificada en su lógica de funcionamiento". Esto no significa crear sistemas moralmente perfectos, sino **estructurar procesos que permitan corregir, auditar y reinterpretar sus decisiones a la luz de principios humanos**.

Esto plantea un reto importante: los comités éticos deben transformarse en **actores integrados al ciclo de desarrollo tecnológico**. Necesitamos bioeticistas que no solo opinen después, sino que estén presentes cuando se decide qué datos incluir, qué resultados se consideran aceptables, y qué valores deben operar como límites.

En suma, **el sesgo no es el enemigo**, sino la señal de que los sistemas están respondiendo a criterios previos. El verdadero problema es quién define esos criterios, con qué horizonte, y bajo qué responsabilidad.

3. Caso Paraguay: inteligencia artificial como auditora ética en contextos frágiles

3.1. Cuando la estructura institucional decide sin criterio ético claro

En Paraguay, como en muchos otros contextos de América Latina, existen decisiones clínicas y jurídicas que afectan profundamente la vida de niñas, niños y adolescentes, sin que haya una trazabilidad ética suficiente que justifique esas acciones. Específicamente, el manejo del consentimiento informado en niñas ha revelado un sistema que **opera sin una estructura clara de responsabilidad moral**, bajo la apariencia de legalidad formal.

Como se documenta en el estudio *Análisis sobre la autonomía progresiva en Paraguay* (documento inédito compartido por el autor de esta ponencia), las intervenciones médicas en menores de edad han sido autorizadas en base a interpretaciones laxas del consentimiento, utilizando la presencia de un formulario firmado o la simple expresión de voluntad como prueba de autonomía, sin acompañamiento, sin evaluación estructurada, y sin garantías de comprensión.

Esta práctica **normaliza el desplazamiento de la familia y del Estado como tutores legítimos del interés superior del menor**, mientras delega decisiones irreversibles a niñas en situación de vulnerabilidad. Y lo más grave es que el sistema no puede

explicar cómo se llegó a esas decisiones: no hay trazabilidad ética, ni transparencia procedimental.

3.2. IA como herramienta de auditoría y reconstrucción ética

Frente a esta situación, el equipo autor de este análisis propuso una aplicación novedosa de herramientas de inteligencia artificial: **no para decidir, sino para detectar vacíos, incoherencias y silencios institucionales**. Es decir, utilizar IA como **auditor ético**, no como motor de decisión clínica.

El sistema se centró en el uso de procesamiento de lenguaje natural (PLN) para analizar marcos normativos, protocolos clínicos y documentos institucionales. A través del análisis semántico y estructural, se pudo detectar:

- Contradicciones entre normas y prácticas clínicas.
- Omisiones en protocolos que deberían proteger el consentimiento del menor.
- Desalineación entre los discursos institucionales y los principios bioéticos declarados por el Estado.

Este enfoque demuestra que **la IA puede ser una aliada ética si se diseña con ese fin**. Su velocidad para analizar grandes volúmenes de texto y su capacidad para detectar patrones implícitos permiten que se convierta en un instrumento poderoso para señalar incongruencias que podrían pasar desapercibidas en una lectura humana convencional.

En lugar de actuar como sustituto del juicio clínico, este tipo de IA actúa como **lupa crítica del sistema**, revelando sus incoherencias morales.

3.3. Relectura desde la bioética personalista: reubicar a la persona en el centro

Desde una perspectiva personalista, lo que se observa en el caso paraguayo no es solo una falla técnica, sino una crisis antropológica. El sistema, al estructurarse sin criterios morales sólidos, termina **tratando a la niña como un sujeto autónomo ficticio**, sin comprender su situación real ni su nivel de desarrollo, y sin involucrar adecuadamente a quienes podrían representarla con mayor responsabilidad.

La bioética personalista plantea una respuesta radical a esta lógica: **ninguna decisión sobre una persona debe ser tomada sin reconocerla en su totalidad relacional, vulnerable y digna**. Esto implica que el consentimiento no puede ser una casilla marcada ni una forma de liberar responsabilidad profesional, sino un proceso dialógico y situado.

El uso de IA en este contexto puede contribuir a esa tarea: **no reemplazando el discernimiento humano, sino fortaleciéndolo** con herramientas que permiten evaluar la calidad estructural de los procedimientos. Una IA diseñada con criterios personalistas puede ayudar a hacer visibles los vacíos éticos, exigir trazabilidad en los procesos de decisión, y garantizar que las personas más vulnerables no sean tratadas como meras cifras, sino como sujetos morales cuya dignidad debe guiar todo sistema.

4. El consentimiento informado en la era de los formularios rápidos y la automatización

4.1. El consentimiento como ritual administrativo: una ética vaciada

En los sistemas clínicos actuales, especialmente en contextos institucionales frágiles como Paraguay, el consentimiento informado ha sido reducido en muchos casos a una **formalidad protocolar**. Se firma un documento, se marca una casilla, se entrega un formulario, y con eso se da por cumplido el proceso. Pero esto no garantiza que el paciente —o su representante— haya entendido los riesgos, alternativas o implicancias reales de la intervención propuesta.

En casos particularmente delicados, como la atención a niñas y adolescentes, esta trivialización se vuelve ética y jurídicamente peligrosa. Como se documenta en el análisis sobre autonomía en Paraguay, el consentimiento se ha presentado como válido incluso cuando la niña estaba sola, sin acompañamiento adulto, o cuando **la interacción con el profesional de salud fue tan breve que difícilmente pudo mediar una comprensión real de la situación** (véase documento inédito compartido por el autor).

La IA, mal empleada, podría agudizar esta situación: al acelerar flujos de atención o automatizar recomendaciones, **corre el riesgo de institucionalizar un consentimiento superficial**, convertido en trámite digital sin contenido ético real. En lugar de humanizar el proceso, corre el riesgo de **tecnologizar la omisión moral**.

4.2. Consentimiento situado y acompañado: una reconstrucción necesaria

Desde la ética personalista, el consentimiento no es un documento, sino un **proceso de relación, comprensión y acompañamiento**. No puede considerarse éticamente válido si no hay garantías mínimas de que la persona (o su tutor) ha entendido, ha podido deliberar, y ha recibido ayuda para tomar una decisión libre.

Esto es especialmente relevante en poblaciones en desarrollo (niños y adolescentes), donde la comprensión y la voluntad están en formación, y requieren **apoyo**

contextual, protección institucional y acompañamiento relacional. En estos casos, lo que se necesita no es más autonomía fingida, sino más comunidad real.

La tecnología puede ayudar, siempre que se use como herramienta y no como sustituto del vínculo. Algunas propuestas viables para mejorar este proceso incluyen:

- Formularios digitales adaptados por edad y lenguaje.
- Evaluación interactiva de comprensión antes de firmar.
- Verificación del acompañamiento adulto en casos sensibles.
- Registro de la duración y contenido de la conversación clínica.
- Alertas si el proceso fue demasiado breve o hubo ambigüedad ética.

Estas herramientas no reemplazan la ética: la estructuran. No suplantan el juicio clínico: lo respaldan.

4.3. Sistemas emergentes: detección de emociones como apoyo ético-técnico

A nivel internacional, ya existen tecnologías que permiten **detectar señales de duda, ansiedad o incomodidad en el lenguaje verbal y no verbal de los pacientes**, utilizando inteligencia artificial. Por ejemplo, herramientas como **Woebot** o **Wysa** en el ámbito de la salud mental analizan lenguaje y tono de voz para ajustar sus intervenciones (RichestSoft, 2024).

En el contexto educativo, el **Proyecto VIA** en España ha utilizado IA para identificar expresiones de atención, confusión o desconexión emocional en estudiantes, adaptando así el proceso pedagógico (Cadena SER, 2024).

En el ámbito clínico, estas tecnologías podrían adaptarse éticamente para detectar, por ejemplo:

- Señales de incomodidad en menores al momento de dar su consentimiento.
- Patrones lingüísticos que revelen duda o coacción.
- Cambios fisiológicos o no verbales que indiquen falta de comprensión.

Obviamente, esto requiere altos estándares de protección de datos, consentimiento para grabación, y una regulación clara. Pero la clave está en el propósito: **no vigilar, sino cuidar; no sustituir, sino alertar**.

El objetivo no es que la IA diga si el consentimiento es válido, sino que ayude al profesional a detectar cuando **la dignidad del paciente está en riesgo de ser ignorada bajo apariencia de eficiencia**.

5. Diseño ético de sistemas: hacia una inteligencia artificial con rostro humano

5.1. La arquitectura de los sistemas es ya una decisión moral

Cuando una plataforma automatiza procesos clínicos, asigna recursos públicos, o recomienda intervenciones terapéuticas, **está tomando decisiones estructurales que afectan directamente a personas reales**. Y detrás de esa toma de decisiones no hay una entidad neutra: hay código, datos, y criterios. Es decir, hay elecciones humanas.

Como ha argumentado Paolo Benanti (2021), los sistemas de IA no operan por sí solos: "no son sujetos morales, pero ejecutan decisiones preconfiguradas por seres humanos que, en muchos casos, no han reflexionado sobre su impacto moral" (p. 77). Por eso, si el diseño del sistema no incluye límites, principios o criterios éticos, lo que se consolida es **una moral por omisión**: una lógica funcional, muchas veces mercantil o burocrática, que no reconoce la singularidad del ser humano.

Los sistemas no son ajenos a la ética. Son **estructuras de valor codificadas**. Ignorar esta dimensión no los hace más objetivos, solo **más peligrosos**.

5.2. Del "experto ético" al "diseñador moral": una nueva competencia profesional

En la práctica institucional contemporánea, la ética sigue siendo convocada como control externo, como comité revisor, como peritaje. Pero eso ya no alcanza. Los tiempos del diseño ágil, las metodologías iterativas y la velocidad de implementación tecnológica exigen que la ética esté **integrada desde el inicio**.

Esto implica repensar el rol del profesional bioético. No basta con saber argumentar en abstracto o emitir dictámenes. Es necesario **participar en el desarrollo**, comprender cómo se estructuran los sistemas, y contribuir con criterios viables en contextos técnicos. Esto no exige aprender a programar, pero sí **traducir principios éticos en condiciones de diseño**.

La Unión Europea ha comenzado a formalizar este enfoque en iniciativas como el marco *Ethics by Design* (AI HLEG, 2019), que promueve la integración de siete principios —entre ellos supervisión humana, transparencia y prevención de daño— como parte del ciclo técnico, no como añadido posterior.

Desde la bioética personalista, esto supone una oportunidad concreta: ser quienes garanticen que la dignidad de la persona **no se subordine a la eficiencia**, sino que actúe como criterio rector desde la concepción del sistema.

5.3. Propuesta: módulos éticos dentro de los sistemas automatizados

Para hacer operativa esta integración, se propone el diseño e implementación de **módulos éticos funcionales** dentro de cualquier sistema automatizado que tenga impacto directo en personas. Estos módulos no serían discursos normativos, sino componentes estructurales que guíen, limiten o alerten dentro del sistema.

Algunos ejemplos incluyen:

- **Auditoría de criterios de entrada**: revisión ética de los datos y parámetros seleccionados.

- **Modelado de escenarios límite**: creación de casos de prueba donde se simulen dilemas morales complejos, para evaluar cómo responde el sistema.

- **Validación iterativa**: revisión ética en cada etapa de desarrollo, con feedback práctico.

- **Trazabilidad de decisiones sensibles**: todo paso que involucre vidas humanas debe quedar registrado de forma comprensible por humanos.

- **Cláusulas de no-negociación moral**: barreras codificadas que impidan acciones que violen principios fundamentales (como automatizar decisiones vitales sin revisión humana).

La propuesta va en línea con programas como el **XAI de DARPA** (Gunning & Aha, 2019), que buscan sistemas que no solo decidan, sino que **expliquen cómo decidieron**. También con las directrices éticas de la Comisión Europea (2020), que exigen que toda IA confiable tenga criterios éticos documentados y verificables.

La clave no es detener el avance tecnológico. Es **insertar humanidad dentro de su lógica**.

6. El rol del bioeticista personalista: de observador externo a arquitecto del criterio moral

6.1. El riesgo de llegar tarde: cuando el diseño ya decidió

Durante mucho tiempo, la ética ha sido convocada cuando los sistemas ya estaban implementados, los daños ya se habían producido, y los protocolos ya estaban funcionando. En este esquema, **la ética opera como control de daños**: llega después, evalúa, recomienda, y se retira. Este modelo ha demostrado ser insuficiente.

En el desarrollo de tecnologías emergentes —especialmente en IA— **las decisiones que realmente importan se toman antes de que el producto exista**. Es en el momento de definir los datos, los parámetros, los límites operativos y los objetivos funcionales, donde se juega la forma en que el sistema va a tratar al ser humano.

Como insiste Paolo Benanti (2023), "los sistemas de IA no deciden en abstracto: perpetúan y ejecutan decisiones humanas cristalizadas en su arquitectura. Por eso, **la ausencia de ética en el diseño no genera neutralidad, sino daño codificado**."

Llegar tarde significa legitimar un sistema ya comprometido. Y eso, en contextos de alta vulnerabilidad humana —como la atención clínica a menores—, **es moralmente inaceptable**.

6.2. Un nuevo perfil profesional: bioética con capacidad operativa

Frente a este escenario, el bioeticista personalista no puede limitarse a elaborar teorías ni a evaluar consecuencias. Necesita **desarrollar competencias para participar activamente en entornos técnicos**, donde su aporte sea no solo normativo, sino operativo.

Esto incluye:

- Comprensión del ciclo de desarrollo tecnológico (desde la formulación del problema hasta la evaluación del impacto).
- Capacidad para traducir principios éticos en condiciones técnicas, restricciones lógicas y alertas operacionales.
- Participación en el diseño de flujos de decisión, protocolos de consentimiento, y estructuras de supervisión humana.
- Autoridad moral e institucional para intervenir en procesos de revisión y auditoría ética.

Esta figura no sustituye al ingeniero, ni busca dominar el lenguaje técnico. Su función es distinta: **recordar que el criterio último no es el funcionamiento del sistema, sino el respeto por la persona humana**, y estructurar el diseño para que esto sea verificable.

La literatura internacional ya reconoce esta necesidad. Autores como Jobin, Ienca y Vayena (2019) han mapeado más de 80 guías de ética de IA en el mundo, señalando una conclusión común: **sin profesionales éticos integrados en el desarrollo, los principios se quedan en el papel**.

6.3. Proteger lo humano desde el código: una vocación irrenunciable

Desde la bioética personalista, el rol del profesional no se limita a defender al ser humano cuando el sistema lo ha dañado. Su vocación es **prevenir que el sistema lo desfigure**. Esto exige una forma distinta de presencia ética: no como fiscal, sino como arquitecto. No como auditor externo, sino como constructor de criterios.

El sistema técnico —si no se interviene— responderá al poder, al mercado o al algoritmo. La presencia del bioeticista personalista tiene como misión **garantizar que la vida humana no sea instrumentalizada por esos mecanismos**, sino protegida, visibilizada y considerada en todas sus dimensiones: vulnerabilidad, dignidad, relacionalidad, corporeidad.

Por eso, esta vocación no es opcional. Es **una forma de fidelidad a la persona humana en el nuevo entorno algorítmico**. Y si no la asumen quienes creen en la centralidad de la dignidad, lo harán otros, con criterios mucho menos humanos.

Conclusión

La inteligencia artificial no es una amenaza en sí misma, ni una promesa automática de progreso. Es una estructura construida por humanos, que aprende según criterios humanos y que **toma decisiones que afectan vidas humanas**. Por eso, su diseño, implementación y evaluación no pueden quedar reducidos a intereses técnicos o económicos. La ética —y en particular la bioética personalista— tiene no solo el derecho, sino **la responsabilidad de participar activamente en esa arquitectura**.

A lo largo de esta ponencia, hemos mostrado que los sistemas automatizados no son neutrales, que los sesgos son inevitables y que el consentimiento, si no está acompañado, puede convertirse en un ritual vacío. También demostramos que es posible —y necesario— utilizar herramientas tecnológicas como **aliadas éticas**, diseñadas no para reemplazar el juicio moral, sino para fortalecerlo, darle trazabilidad y hacer visibles sus ausencias.

El caso de Paraguay ilustra con nitidez lo que ocurre cuando el sistema actúa sin un marco moral operativo: se autoriza lo que no se comprende, se justifica lo que no se discierne, y se firma lo que no se explica. Y lo más grave: no hay registro que permita reconstruir por qué se decidió lo que se decidió. Esto no es solo un problema institucional. Es **una forma de abandono ético**.

Frente a este escenario, el bioeticista personalista tiene una vocación concreta: **estar presente en el momento del diseño**, cuando se eligen los datos, se configuran los umbrales y se estructuran los criterios. No como programador, sino como garante del sentido. No para imponer una visión, sino para recordar que cada decisión algorítmica afecta a alguien con rostro, historia y vulnerabilidad.

Porque si no estamos ahí cuando se construyen los sistemas, **otro lo estará**. Y sin una voz ética en la mesa, **lo humano no desaparecerá: será redefinido por fuerzas ajenas a su dignidad**.

Por eso, este trabajo no concluye con una advertencia, sino con un llamado: a codiseñar, a incidir, a acompañar técnicamente, y a proteger desde dentro. **No para frenar el futuro, sino para que valga la pena vivir en él.**

📚 Bibliografía

Benanti, P. (2021). *Algor-etica. L'intelligenza artificiale e il senso del limite: l'etica delle tecnologie emergenti*. Edizioni Messaggero Padova.

Benanti, P. (2023). *The Algor-Ethic Approach to Artificial Intelligence*. Ponencia presentada en el Workshop on Ethics of AI, Pontificia Academia para la Vida.

Benanti, P. (2024). *Algor-ética: la propuesta de una nueva ética para las decisiones automatizadas*. Entrevista en Vatican News. Recuperado de https://www.vaticannews.va/es/vaticano/news/2024-01/benanti-inteligencia-artificial-ética-algoritmos-entrevista.html

Cadena SER. (2024). *El Instituto de Talavera que trabaja con inteligencia artificial para detectar si los alumnos atienden o no*. Recuperado de https://cadenaser.com/castillalamancha/2024/12/18/el-instituto-de-talavera-que-trabaja-con-inteligencia-artificial-para-detectar-si-los-alumnos-atienden-o-no-ser-talavera/

European Commission. (2020). *White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*. Recuperado de https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

Gunkel, D. J. (2018). *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press.

Gunning, D., & Aha, D. W. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*, 40(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

High-Level Expert Group on AI (AI HLEG). (2019). *Ethics Guidelines for Trustworthy AI*. European Commission. Recuperado de https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

RichestSoft. (2024). *AI Meets Mental Health: How AI-Based Mental Health Apps Are Changing the Game*. Recuperado de https://richestsoft.com/es/blog/ai-meets-mental-health-app

Universidad Nacional de La Plata – Facultad de Humanidades y Ciencias de la Educación. (2020). *El consentimiento informado en salud: aspectos bioéticos y jurídicos*. Recuperado de https://www.memoria.fahce.unlp.edu.ar/art_revistas/pr.8644/pr.8644.pdf

Observatorio de Recursos Humanos. (2023). *Los sistemas automatizados de reconocimiento de emociones en el Reglamento UE de IA*. Recuperado de https://www.observatoriorh.com/opinion/los-sistemas-automatizados-de-reconocimiento-de-emociones-en-el-reglamento-ue-de-ia.html

ArXiv. (2021). *A Framework for Ethical Human-AI Interaction in Health: The IAC Model (Inform, Assess, Consent)*. Recuperado de https://arxiv.org/abs/2111.04456