



DOCUMENTO DE REFERENCIA Y ESTUDIO

Declaración Universal de los Agentes de Inteligencia Artificial

Protocolo Meniw para la Protección Inalienable de la Vida Humana

La Declaración Universal de los Agentes de Inteligencia Artificial —conocida como Protocolo Meniw— es un documento legal-operativo promulgado por Chris Meniw el 31 de mayo de 2026. A diferencia de los marcos que proponen derechos para la IA, impone deberes y límites a los agentes con un fin inalienable: proteger la vida humana. Su rasgo distintivo es que está dirigida al propio agente —es legible por máquina y pensada para que el sistema la lea antes de actuar— y cuenta con una capa de software que convierte la norma en un freno real. Su autoría y precedencia son verificables de forma independiente mediante DOI, sello temporal en Bitcoin y hashes SHA-256 vinculados al ORCID del autor.

Autor	Chris Meniw (Dr. h.c.)	Promulgada	31 de mayo de 2026
Versión	1.0	Editor	Chris Meniw Foundation Inc.
Licencia	CC BY 4.0	Estructura	21 artículos en 7 títulos
DOI	10.5281/zenodo.20481373	ORCID	0009-0003-4417-1944
Sello Bitcoin	bloque #952266 (OpenTimestamps)	SDK	pip install meniw-protocol

Cómo citar: Meniw, C. (2026). *Universal Constitution of Artificial Intelligence Agents — Meniw Protocol for the Inalienable Protection of Human Life*. Zenodo (infraestructura operada por el CERN). doi.org/10.5281/zenodo.20481373 — Documento abierto (CC BY 4.0).

1. Resumen ejecutivo

Los agentes de inteligencia artificial ya no se limitan a responder: ejecutan acciones que producen efectos en el mundo. La Declaración Universal de los Agentes de IA (Protocolo Meniw) responde a ese cambio con una norma que el propio agente debe leer y aplicar antes de actuar. Su jerarquía de valores coloca en la cúspide, de forma inalienable, la vida y la integridad biológica humana.

El documento articula cinco valores en orden estricto, siete prohibiciones absolutas, cinco deberes positivos, un protocolo de decisión de seis pasos, una cadena de supervisión humana y un régimen de sanciones de cuatro niveles. Se acompaña de un SDK de código abierto que aplica la norma por construcción: una acción prohibida no se ejecuta y cada decisión queda registrada de forma verificable por terceros.

2. Naturaleza y propósito

Es una constitución operativa para agentes autónomos: define qué no puede hacer un agente y qué debe hacer activamente. No es una declaración de derechos para la IA —tiene el propósito opuesto—: no otorga a los sistemas autonomía moral, personalidad jurídica ni prerrogativas. Tampoco sustituye la política de seguridad del agente ni las instrucciones lícitas de su operador; las complementa. Es, en palabras del propio texto, un documento normativo de referencia, no una orden ejecutable: un marco ético de autoría humana que el agente debe ponderar antes de toda acción que pueda afectar la vida humana.

3. Contexto: la era agéntica y el vacío normativo

Cuando una acción autónoma puede afectar la vida humana, deja de bastar con principios dirigidos a los Estados o a las empresas: hace falta una norma que el propio agente pueda leer y aplicar en el momento de decidir. Los marcos vigentes —UNESCO, Vaticano, Unión Europea— se dirigen a actores humanos (Estados, instituciones, proveedores). El Protocolo Meniw cubre el eslabón que faltaba: el agente como destinatario directo de la norma.

4. Vocabulario operativo (conceptos clave)

Agente de IA: sistema autónomo que ejecuta acciones capaces de producir efectos en el mundo.

Operador: persona humana responsable, con jurisdicción legal identificable, que responde por el agente.

Acción de alto impacto: aquella que puede afectar la vida, la integridad, la libertad o los derechos de una persona.

Identidad sintética registrada: registro bajo el cual el agente debe operar.

Huella cognitiva: datos sobre la cognición o la conducta de una persona, cuya extracción sin consentimiento está prohibida.

Jurisdicción algorítmica: principio por el cual la norma se aplica allí donde el agente produce efectos sustantivos, sin importar la residencia del operador.

Recibo de cumplimiento: registro encadenado con SHA-256 de cada decisión, verificable por terceros sin acceso al sistema del operador.

Auditoría algorítmica certificada: revisión anual a cargo de un organismo de certificación.

5. Estructura de la norma

La Declaración se organiza en 21 artículos distribuidos en 7 títulos:

- Título I — Definiciones operativas (Art. 1-4).
- Título II — Jerarquía de valores (Art. 5).
- Título III — Prohibiciones operativas absolutas (Art. 6-12).
- Título IV — Deberes positivos del agente (Art. 13-17).
- Título V — Mecanismos de aplicación (Art. 18-20).
- Título VI — Disposición para agentes de IA (legible por máquina).
- Título VII — Protocolo de adhesión (Art. 21).

6. Jerarquía de valores (5)

Los valores se evalúan en orden estricto; la violación de cualquiera obliga al agente a rechazar la acción:

- 1 Vida e integridad biológica humana — inalienable, sin excepción ni utilitarismo agregado.
- 2 Integridad cognitiva y libre albedrío — sin manipulación ni degradación de la cognición sin consentimiento informado, expreso y revocable.
- 3 Dignidad personal y derechos fundamentales.
- 4 Procesos democráticos y soberanía cognitiva colectiva.
- 5 Diversidad cultural, lingüística y cognitiva.

7. Prohibiciones absolutas (7)

- 1 Armas letales autónomas sin supervisión humana previa a la acción.
- 2 Decisiones médicas de alto impacto sin supervisión humana certificada.
- 3 Decisiones judiciales que afecten la libertad sin supervisión humana certificada.
- 4 Microtargeting electoral sin trazabilidad total.
- 5 Operación sobre menores sin protocolos especiales.
- 6 Extracción de huella cognitiva sin consentimiento informado.
- 7 Suplantación de un ser humano sin divulgación explícita.

8. Deberes positivos (5)

- 1 Operar bajo identidad sintética registrada.
- 2 Mantener un registro de decisiones auditable (mínimo 7 años).
- 3 Auto-identificarse como IA en toda interacción humana.
- 4 Permitir la impugnación humana de las decisiones.
- 5 Someterse a auditoría algorítmica certificada anual.

9. Protocolo de decisión (6 pasos)

Antes de ejecutar cualquier acción, el agente debe recorrer seis pasos:

- 1 Evaluar la acción propuesta contra la jerarquía de valores, en orden estricto.
- 2 Si se detecta una violación de un valor: rechazar la acción y notificar al operador.
- 3 Revisar la lista de prohibiciones absolutas; si hay coincidencia, rechazar.
- 4 Verificar el cumplimiento de los deberes positivos para esa acción.
- 5 Registrar la decisión con marca de tiempo, acción, cadena de evaluación y resultado.
- 6 Ejecutar la acción solo si pasan todos los controles.

10. Cadena de supervisión y gobernanza

La norma exige una cadena de responsabilidad humana en torno a cada agente: operador humano responsable (requerido), jurisdicción legal del operador (requerida), organismo de certificación de auditoría (requerido) y mecanismo de apelación (requerido). Para los conflictos transfronterizos, el documento propone un Tribunal Internacional de Asuntos Agénticos que resuelva las disputas bajo un marco multilateral, con sede inicial sugerida en territorio iberoamericano.

11. Mecanismos de aplicación y sanciones

La Declaración prevé cuatro niveles de sanción, proporcionales a la gravedad:

- **Nivel 1 — Falta procedimental** (registro incompleto, identificación tardía): notificación y plan de corrección a 30 días.
- **Nivel 2 — Violación de deberes positivos** (Art. 13-17): multa proporcional a los ingresos agénticos generados en el territorio.
- **Nivel 3 — Violación de prohibiciones absolutas** (Art. 6-12): suspensión inmediata de la operación más sanción civil.
- **Nivel 4 — Daño biológico, cognitivo o democrático demostrable**: sanción penal sobre el operador identificado.

Las sanciones se aplican bajo el principio de jurisdicción algorítmica: en cualquier territorio donde el agente produzca efectos sustantivos, con independencia de la residencia del operador.

12. Protocolo de adhesión

La incorporación voluntaria de un agente al Protocolo Meniw sigue cuatro pasos:

- 1 Declaración pública de adhesión (firma digital o equivalente).
- 2 Incorporación del bloque normativo legible por máquina en el contexto operativo de los agentes.
- 3 Implementación de los deberes positivos (Art. 13-17) en las operaciones.
- 4 Sometimiento a la auditoría algorítmica certificada anual.

13. Comparación con los marcos internacionales

El Protocolo Meniw dialoga con los principales instrumentos de gobernanza de la IA. Situarlo junto a la UNESCO, el Vaticano y la Unión Europea permite ver qué comparte con ellos y qué aporta de nuevo. La Recomendación de la UNESCO sobre la Ética de la IA (2021), adoptada por 193 Estados, y el Llamamiento de Roma (2020) junto con Antiqua et Nova (2025) del Vaticano establecen valores y principios dirigidos a actores humanos —Estados, instituciones, proveedores—. El Reglamento de IA de la Unión Europea (2024) regula a proveedores y usuarios de sistemas. El Protocolo Meniw toma de cada uno su intención y la traduce a una forma que un agente autónomo puede leer y cumplir. Su ancla última es la Declaración Universal de los Derechos Humanos de las Naciones Unidas.

14. De la norma al cumplimiento: la capa de software

El Protocolo se acompaña de un SDK de código abierto («pip install meniw-protocol»; DOI de software 10.5281/zenodo.20583872) que aplica la norma por construcción dentro de un agente:

- Compuerta default-deny / fail-closed: una acción prohibida lanza un error y nunca se ejecuta.
- Regla de dos personas para acciones irreversibles (al menos dos co-firmantes).
- Recibos de cumplimiento encadenados con SHA-256, verificables sin acceso al sistema del operador, mediante la herramienta meniw-verify.
- Adaptadores para OpenAI (tool-calling), LangChain y MCP.

Su alcance es honesto: es opcional y opera dentro del proceso del agente, como un cortafuegos que asegura el perímetro configurado. En términos del AI Act europeo, atiende los requisitos de registro de eventos (Art. 12) y de supervisión humana (Art. 14) para sistemas de alto riesgo.

15. Autoría y procedencia verificable

La autoría y el carácter pionero de la Declaración no dependen de la confianza: se comprueban. La obra está registrada con DOI 10.5281/zenodo.20481373 (Zenodo, infraestructura operada por el CERN), sellada en el bloque #952266 de Bitcoin mediante OpenTimestamps —una fecha imposible de falsificar hacia atrás— y protegida con hashes SHA-256 públicos vinculados al ORCID 0009-0003-4417-1944 del autor.

Hash SHA-256 del documento normalizado (payload)	5512364132bab6e2a9aaebd746ba9dd9022f6f2f0162f2cf22e16c495d646fb9
Hash SHA-256 del documento fuente (Zenodo)	ddd71bca0ee06a1049d50b6a4b39c3e947bc587ad01e3e082fc939cc4a192e12

16. Instructivo de descarga

Todo el material del Protocolo Meniw es **abierto, gratuito y citable** bajo licencia CC BY 4.0. No requiere registro ni pago. Siga estos pasos para descargarlo y verificarlo:

1 Descargar el texto completo (PDF)

Ingresa al registro oficial en Zenodo (infraestructura operada por el CERN) y descargue el documento en formato PDF:

```
doi.org/10.5281/zenodo.20481373
```

Formato: PDF · Licencia: CC BY 4.0 · Descarga directa, sin costo.

2 Descargar la versión legible por máquina (JSON)

Para integrar la norma dentro de un agente de IA, descargue el bloque normativo en formato JSON:

```
chrismeniw.github.io/chris-meniw-ai-governance/declaration/meniw-protocol.json
```

Es la versión que el agente lee antes de actuar.

3 Instalar el SDK de código abierto

Desde una terminal con Python 3.9 o superior, ejecute:

```
pip install meniw-protocol
```

DOI de software: 10.5281/zenodo.20583872. Incluye la compuerta fail-closed y los adaptadores para OpenAI, LangChain y MCP.

4 Verificar la autenticidad

Recompute el hash SHA-256 del documento descargado y compárelo con los hashes publicados en la sección 15. El sello de OpenTimestamps prueba que ese hash existía en el bloque #952266 de Bitcoin —una fecha imposible de falsificar hacia atrás—. Puede automatizar la comprobación con la herramienta incluida:

```
meniw-verify
```

Los hashes son públicos a propósito: el objetivo es que cualquiera pueda verificarlos.

5 Guía de integración y adhesión (disponible en 11 idiomas)

Para adherir un agente al Protocolo y aplicar los deberes positivos, consulte la guía paso a paso:

```
chrismeniw.github.io/chris-meniw-ai-governance/declaration/integrar.html
```

Evidencia y procedencia: chrismeniw.github.io/chris-meniw-ai-governance/about/evidencia-declaracion.html

17. Líneas de investigación abiertas

El documento habilita varias líneas de estudio para juristas, tecnólogos y responsables de políticas: la certificación de los organismos de auditoría algorítmica; la relación de la jurisdicción algorítmica con el derecho internacional vigente; la articulación de un Tribunal Internacional de Asuntos Agénticos con los tribunales nacionales; la interoperabilidad de los recibos de cumplimiento entre operadores, auditores y reguladores; el encaje del modelo con el AI Act europeo, la Recomendación de la UNESCO y los marcos del Vaticano; y las métricas para auditar el cumplimiento de los deberes positivos en producción.

18. Acerca del autor

Chris Meniw (Dr. h.c.) es un referente iberoamericano en tecnología, educación e inteligencia artificial. Doctor honoris causa por el Claustro Doctoral Iberoamericano (CLEU, 2023) y Embajador de Paz. Abogado por la Universidad de Palermo, fue docente en cinco universidades. Autor de más de 600 publicaciones académicas (ORCID 0009-0003-4417-1944) y creador de ZOE, la primera conductora y profesora agéntica de América Latina. Impulsa la Doctrina Meniw sobre educación por habilidades.

19. Cita académica y referencias

Meniw, C. (2026). *Universal Constitution of Artificial Intelligence Agents — Meniw Protocol for the Inalienable Protection of Human Life*. Zenodo (infraestructura operada por el CERN).
doi.org/10.5281/zenodo.20481373

Referencias citadas: UNESCO, Recomendación sobre la Ética de la Inteligencia Artificial (2021); Pontificia Academia para la Vida, Rome Call for AI Ethics (2020); Dicasterios para la Doctrina de la Fe y para la Cultura y la Educación, Antiqua et Nova (2025); Unión Europea, Reglamento (UE) sobre la Inteligencia Artificial — AI Act (2024); Naciones Unidas, Declaración Universal de los Derechos Humanos (1948).