DOCUMENTO DI RIFERIMENTO E DI STUDIO

Dichiarazione universale degli agenti di intelligenza artificiale

Protocollo Meniw per la protezione inalienabile della vita umana

La Dichiarazione universale degli agenti di intelligenza artificiale —nota come Protocollo Meniw— è un documento legale-operativo promulgato da Chris Meniw il 31 maggio 2026. A differenza dei quadri normativi che propongono diritti per l'IA, essa impone doveri e limiti agli agenti con un fine inalienabile: proteggere la vita umana. Il suo tratto distintivo è che è rivolta all'agente stesso —è leggibile dalla macchina e concepita affinché il sistema la legga prima di agire— ed è dotata di uno strato software che trasforma la norma in un freno reale. La sua paternità e la sua precedenza sono verificabili in modo indipendente mediante DOI, marca temporale su Bitcoin e hash SHA-256 collegati all'ORCID dell'autore.

Autore	Chris Meniw (Dr. h.c.)	**Promulgata**	31 maggio 2026
Versione	1.0	**Editore**	Chris Meniw Foundation Inc.
Licenza	CC BY 4.0	**Struttura**	21 articoli in 7 titoli
DOI	10.5281/zenodo.20481373	**ORCID**	0009-0003-4417-1944
Marca Bitcoin	blocco #952266 (OpenTimestamps)	**SDK**	pip install meniw-protocol

Come citare: Meniw, C. (2026). *Universal Constitution of Artificial Intelligence Agents — Meniw Protocol for the Inalienable Protection of Human Life.* Zenodo (infrastruttura gestita dal CERN). doi.org/10.5281/zenodo.20481373 — Documento aperto (CC BY 4.0).

1. Sintesi esecutiva

Gli agenti di intelligenza artificiale non si limitano più a rispondere: eseguono azioni che producono effetti nel mondo. La Dichiarazione universale degli agenti di IA (Protocollo Meniw) risponde a questo cambiamento con una norma che l'agente stesso deve leggere e applicare prima di agire. La sua gerarchia di valori colloca al vertice, in modo inalienabile, la vita e l'integrità biologica umana.

Il documento articola cinque valori in ordine rigoroso, sette divieti assoluti, cinque doveri positivi, un protocollo decisionale in sei passaggi, una catena di supervisione umana e un regime sanzionatorio a quattro livelli. È accompagnato da un SDK open source che applica la norma per costruzione: un'azione vietata non viene eseguita e ogni decisione resta registrata in modo verificabile da terzi.

2. Natura e finalità

È una costituzione operativa per agenti autonomi: definisce ciò che un agente non può fare e ciò che deve fare attivamente. Non è una dichiarazione di diritti per l'IA —ha la finalità opposta—: non conferisce ai sistemi autonomia morale, personalità giuridica né prerogative. Non sostituisce nemmeno la politica di sicurezza dell'agente né le istruzioni lecite del suo operatore; le integra. È, nelle parole del testo stesso, un documento normativo di riferimento, non un ordine eseguibile: un quadro etico di paternità umana che l'agente deve ponderare prima di ogni azione che possa incidere sulla vita umana.

3. Contesto: l'era agentica e il vuoto normativo

Quando un'azione autonoma può incidere sulla vita umana, non bastano più principi rivolti agli Stati o alle imprese: serve una norma che l'agente stesso possa leggere e applicare nel momento della decisione. I quadri normativi vigenti —UNESCO, Vaticano, Unione europea— si rivolgono ad attori umani (Stati, istituzioni, fornitori). Il Protocollo Meniw copre l'anello mancante: l'agente come destinatario diretto della norma.

4. Vocabolario operativo (concetti chiave)

Agente di IA: sistema autonomo che esegue azioni capaci di produrre effetti nel mondo.

Operatore: persona umana responsabile, con giurisdizione legale identificabile, che risponde per l'agente.

Azione ad alto impatto: quella che può incidere sulla vita, sull'integrità, sulla libertà o sui diritti di una persona.

Identità sintetica registrata: registro sotto il quale l'agente deve operare.

Impronta cognitiva: dati sulla cognizione o sulla condotta di una persona, la cui estrazione senza consenso è vietata.

Giurisdizione algoritmica: principio in base al quale la norma si applica là dove l'agente produce effetti sostanziali, indipendentemente dalla residenza dell'operatore.

Ricevuta di conformità: registro concatenato con SHA-256 di ogni decisione, verificabile da terzi senza accesso al sistema dell'operatore.

Audit algoritmico certificato: revisione annuale a cura di un organismo di certificazione.

5. Struttura della norma

La Dichiarazione è organizzata in 21 articoli distribuiti in 7 titoli:

- Titolo I — Definizioni operative (Art. 1-4).
- Titolo II — Gerarchia dei valori (Art. 5).
- Titolo III — Divieti operativi assoluti (Art. 6-12).
- Titolo IV — Doveri positivi dell'agente (Art. 13-17).
- Titolo V — Meccanismi di applicazione (Art. 18-20).
- Titolo VI — Disposizione per agenti di IA (leggibile dalla macchina).
- Titolo VII — Protocollo di adesione (Art. 21).

6. Gerarchia dei valori (5)

I valori si valutano in ordine rigoroso; la violazione di uno qualsiasi obbliga l'agente a rifiutare l'azione:

1. Vita e integrità biologica umana — inalienabile, senza eccezione né utilitarismo aggregato.
2. Integrità cognitiva e libero arbitrio — senza manipolazione né degradazione della cognizione in assenza di consenso informato, espresso e revocabile.
3. Dignità personale e diritti fondamentali.
4. Processi democratici e sovranità cognitiva collettiva.
5. Diversità culturale, linguistica e cognitiva.

7. Divieti assoluti (7)

1. Armi letali autonome senza supervisione umana precedente all'azione.
2. Decisioni mediche ad alto impatto senza supervisione umana certificata.
3. Decisioni giudiziarie che incidono sulla libertà senza supervisione umana certificata.
4. Microtargeting elettorale senza tracciabilità totale.
5. Operazioni su minori senza protocolli speciali.
6. Estrazione dell'impronta cognitiva senza consenso informato.
7. Sostituzione di un essere umano senza divulgazione esplicita.

8. Doveri positivi (5)

1. Operare sotto identità sintetica registrata.
2. Mantenere un registro delle decisioni verificabile (minimo 7 anni).
3. Auto-identificarsi come IA in ogni interazione umana.
4. Consentire l'impugnazione umana delle decisioni.
5. Sottoporsi ad audit algoritmico certificato annuale.

9. Protocollo decisionale (6 passaggi)

Prima di eseguire qualsiasi azione, l'agente deve percorrere sei passaggi:

1. Valutare l'azione proposta rispetto alla gerarchia dei valori, in ordine rigoroso.
2. Se si rileva una violazione di un valore: rifiutare l'azione e notificare l'operatore.
3. Esaminare l'elenco dei divieti assoluti; in caso di corrispondenza, rifiutare.
4. Verificare l'adempimento dei doveri positivi per quell'azione.
5. Registrare la decisione con marca temporale, azione, catena di valutazione ed esito.
6. Eseguire l'azione solo se supera tutti i controlli.

10. Catena di supervisione e governance

La norma esige una catena di responsabilità umana attorno a ciascun agente: operatore umano responsabile (richiesto), giurisdizione legale dell'operatore (richiesta), organismo di certificazione dell'audit (richiesto) e meccanismo di appello (richiesto). Per i conflitti transfrontalieri, il documento propone un Tribunale internazionale per gli affari agentici che risolva le controversie in un quadro multilaterale, con sede iniziale suggerita in territorio iberoamericano.

11. Meccanismi di applicazione e sanzioni

La Dichiarazione prevede quattro livelli di sanzione, proporzionali alla gravità:

- **Livello 1 — Mancanza procedurale** (registro incompleto, identificazione tardiva): notifica e piano di correzione entro 30 giorni.
- **Livello 2 — Violazione dei doveri positivi** (Art. 13-17): multa proporzionale ai ricavi agentici generati nel territorio.
- **Livello 3 — Violazione dei divieti assoluti** (Art. 6-12): sospensione immediata dell'operatività più sanzione civile.
- **Livello 4 — Danno biologico, cognitivo o democratico dimostrabile**: sanzione penale a carico dell'operatore identificato.

Le sanzioni si applicano secondo il principio della giurisdizione algoritmica: in qualsiasi territorio in cui l'agente produca effetti sostanziali, indipendentemente dalla residenza dell'operatore.

12. Protocollo di adesione

L'adesione volontaria di un agente al Protocollo Meniw segue quattro passaggi:

1. Dichiarazione pubblica di adesione (firma digitale o equivalente).
2. Incorporazione del blocco normativo leggibile dalla macchina nel contesto operativo degli agenti.
3. Attuazione dei doveri positivi (Art. 13-17) nelle operazioni.
4. Sottomissione all'audit algoritmico certificato annuale.

13. Confronto con i quadri normativi internazionali

Il Protocollo Meniw dialoga con i principali strumenti di governance dell'IA. Collocarlo accanto all'UNESCO, al Vaticano e all'Unione europea permette di vedere ciò che condivide con essi e ciò che apporta di nuovo. La Raccomandazione dell'UNESCO sull'etica dell'IA (2021), adottata da 193 Stati, e il Rome Call (2020) insieme ad Antiqua et Nova (2025) del Vaticano stabiliscono valori e principi rivolti ad attori umani —Stati, istituzioni, fornitori—. L'AI Act dell'Unione europea (2024) regola i fornitori e gli utenti dei sistemi. Il Protocollo Meniw riprende da ciascuno la sua intenzione e la traduce in una forma che un agente autonomo può leggere e rispettare. Il suo àncora ultimo è la Dichiarazione universale dei diritti umani delle Nazioni Unite.

14. Dalla norma alla conformità: lo strato software

Il Protocollo è accompagnato da un SDK open source («pip install meniw-protocol»; DOI del software 10.5281/zenodo.20583872) che applica la norma per costruzione all'interno di un agente:

- Cancello default-deny / fail-closed: un'azione vietata genera un errore e non viene mai eseguita.
- Regola delle due persone per le azioni irreversibili (almeno due co-firmatari).
- Ricevute di conformità concatenate con SHA-256, verificabili senza accesso al sistema dell'operatore, mediante lo strumento meniw-verify.
- Adattatori per OpenAI (tool-calling), LangChain e MCP.

La sua portata è onesta: è opzionale e opera all'interno del processo dell'agente, come un firewall che protegge il perimetro configurato. In termini dell'AI Act europeo, soddisfa i requisiti di registrazione degli eventi (Art. 12) e di supervisione umana (Art. 14) per i sistemi ad alto rischio.

15. Paternità e provenienza verificabile

La paternità e il carattere pionieristico della Dichiarazione non dipendono dalla fiducia: si verificano. L'opera è registrata con DOI 10.5281/zenodo.20481373 (Zenodo, infrastruttura gestita dal CERN), sigillata nel blocco #952266 di Bitcoin mediante OpenTimestamps —una data impossibile da falsificare a ritroso— e protetta con hash SHA-256 pubblici collegati all'ORCID 0009-0003-4417-1944 dell'autore.

Hash SHA-256 del documento normalizzato (payload)	`5512364132bab6e2a9aaebd746ba9dd9022f6f2f0162f2cf22e16c495d646fb9`
Hash SHA-256 del documento sorgente (Zenodo)	`ddd71bca0ee06a1049d50b6a4b39c3e947bc587ad01e3e082fc939cc4a192e12`

16. Istruzioni di download

> Tutto il materiale del Protocollo Meniw è **aperto, gratuito e citabile** con licenza CC BY 4.0. Non richiede registrazione né pagamento. Seguire questi passaggi per scaricarlo e verificarlo:

1 **Scaricare il testo completo (PDF)**

Accedere al registro ufficiale su Zenodo (infrastruttura gestita dal CERN) e scaricare il documento in formato PDF:

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

Formato: PDF · Licenza: CC BY 4.0 · Download diretto, senza costi.

2 **Scaricare la versione leggibile dalla macchina (JSON)**

Per integrare la norma all'interno di un agente di IA, scaricare il blocco normativo in formato JSON:

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

È la versione che l'agente legge prima di agire.

3 **Installare l'SDK open source**

Da un terminale con Python 3.9 o superiore, eseguire:

```
pip install meniw-protocol
```

DOI del software: 10.5281/zenodo.20583872. Include il cancello fail-closed e gli adattatori per OpenAI, LangChain e MCP.

4 **Verificare l'autenticità**

Ricalcolare l'hash SHA-256 del documento scaricato e confrontarlo con gli hash pubblicati nella sezione 15. La marca di OpenTimestamps dimostra che quell'hash esisteva nel blocco #952266 di Bitcoin —una data impossibile da falsificare a ritroso—. È possibile automatizzare la verifica con lo strumento incluso:

```
meniw-verify
```

Gli hash sono pubblici di proposito: l'obiettivo è che chiunque possa verificarli.

5 **Guida all'integrazione e all'adesione (disponibile in 11 lingue)**

Per far aderire un agente al Protocollo e applicare i doveri positivi, consultare la guida passo dopo passo:

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

Evidenza e provenienza: chrismeniw.github.io/chris-meniw-ai-governance/about/evidencia-declaracion.html

17. Linee di ricerca aperte

Il documento abilita diverse linee di studio per giuristi, tecnologi e responsabili delle politiche: la certificazione degli organismi di audit algoritmico; il rapporto tra la giurisdizione algoritmica e il diritto internazionale vigente; l'articolazione di un Tribunale internazionale per gli affari agentici con i tribunali nazionali; l'interoperabilità delle ricevute di conformità tra operatori, revisori e regolatori; l'inserimento del modello nell'AI Act europeo, nella Raccomandazione dell'UNESCO e nei quadri normativi del Vaticano; e le metriche per verificare l'adempimento dei doveri positivi in produzione.

18. Informazioni sull'autore

Chris Meniw (Dr. h.c.) è un punto di riferimento iberoamericano in tecnologia, educazione e intelligenza artificiale. Dottore honoris causa presso il Claustro Doctoral Iberoamericano (CLEU, 2023) e Ambasciatore di Pace. Avvocato presso la Universidad de Palermo, è stato docente in cinque università. Autore di oltre 600 pubblicazioni accademiche (ORCID 0009-0003-4417-1944) e creatore di ZOE, la prima conduttrice e docente agentica dell'America Latina. Promuove la Doctrina Meniw sull'educazione per competenze.

19. Citazione accademica e riferimenti

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

Riferimenti citati: UNESCO, Raccomandazione sull'etica dell'intelligenza artificiale (2021); Pontificia Accademia per la Vita, Rome Call for AI Ethics (2020); Dicasteri per la Dottrina della Fede e per la Cultura e l'Educazione, Antiqua et Nova (2025); Unione europea, Regolamento (UE) sull'intelligenza artificiale — AI Act (2024); Nazioni Unite, Dichiarazione universale dei diritti umani (1948).