



DOCUMENTO DE REFERÊNCIA E ESTUDO

Declaração Universal dos Agentes de Inteligência Artificial

Protocolo Meniw para a Proteção Inalienável da Vida Humana

A Declaração Universal dos Agentes de Inteligência Artificial —conhecida como Protocolo Meniw— é um documento jurídico-operativo promulgado por Chris Meniw em 31 de maio de 2026. Diferentemente dos marcos que propõem direitos para a IA, ela impõe deveres e limites aos agentes com um fim inalienável: proteger a vida humana. Seu traço distintivo é que está dirigida ao próprio agente —é legível por máquina e concebida para que o sistema a leia antes de agir— e conta com uma camada de software que converte a norma em um freio real. Sua autoria e precedência são verificáveis de forma independente mediante DOI, selo temporal em Bitcoin e hashes SHA-256 vinculados ao ORCID do autor.

Autor	Chris Meniw (Dr. h.c.)	Promulgada	31 de maio de 2026
Versão	1.0	Editora	Chris Meniw Foundation Inc.
Licença	CC BY 4.0	Estrutura	21 artigos em 7 títulos
DOI	10.5281/zenodo.20481373	ORCID	0009-0003-4417-1944
Selo Bitcoin	bloco #952266 (OpenTimestamps)	SDK	pip install meniw-protocol

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

1. Resumo executivo

Os agentes de inteligência artificial já não se limitam a responder: executam ações que produzem efeitos no mundo. A Declaração Universal dos Agentes de IA (Protocolo Meniw) responde a essa mudança com uma norma que o próprio agente deve ler e aplicar antes de agir. Sua hierarquia de valores coloca no ápice, de forma inalienável, a vida e a integridade biológica humana.

O documento articula cinco valores em ordem estrita, sete proibições absolutas, cinco deveres positivos, um protocolo de decisão de seis passos, uma cadeia de supervisão humana e um regime de sanções de quatro níveis. É acompanhado de um SDK de código aberto que aplica a norma por construção: uma ação proibida não é executada e cada decisão fica registrada de forma verificável por terceiros.

2. Natureza e propósito

É uma constituição operativa para agentes autônomos: define o que um agente não pode fazer e o que deve fazer ativamente. Não é uma declaração de direitos para a IA —tem o propósito oposto—: não outorga aos sistemas autonomia moral, personalidade jurídica nem prerrogativas. Tampouco substitui a política de segurança do agente nem as instruções lícitas de seu operador; complementa-as. É, nas palavras do próprio texto, um documento normativo de referência, não uma ordem executável: um marco ético de autoria humana que o agente deve ponderar antes de toda ação que possa afetar a vida humana.

3. Contexto: a era agêntica e o vácuo normativo

Quando uma ação autônoma pode afetar a vida humana, deixa de ser suficiente recorrer a princípios dirigidos aos Estados ou às empresas: faz-se necessária uma norma que o próprio agente possa ler e aplicar no momento de decidir. Os marcos vigentes —UNESCO, Vaticano, União Europeia— dirigem-se a atores humanos (Estados, instituições, provedores). O Protocolo Meniw cobre o elo que faltava: o agente como destinatário direto da norma.

4. Vocabulário operativo (conceitos-chave)

Agente de IA: sistema autônomo que executa ações capazes de produzir efeitos no mundo.

Operador: pessoa humana responsável, com jurisdição legal identificável, que responde pelo agente.

Ação de alto impacto: aquela que pode afetar a vida, a integridade, a liberdade ou os direitos de uma pessoa.

Identidade sintética registrada: registro sob o qual o agente deve operar.

Pegada cognitiva: dados sobre a cognição ou a conduta de uma pessoa, cuja extração sem consentimento está proibida.

Jurisdição algorítmica: princípio pelo qual a norma se aplica onde o agente produz efeitos substantivos, independentemente da residência do operador.

Recibo de conformidade: registro encadeado com SHA-256 de cada decisão, verificável por terceiros sem acesso ao sistema do operador.

Auditoria algorítmica certificada: revisão anual a cargo de um organismo de certificação.

5. Estrutura da norma

A Declaração organiza-se em 21 artigos distribuídos em 7 títulos:

- Título I — Definições operativas (Art. 1-4).
- Título II — Hierarquia de valores (Art. 5).
- Título III — Proibições operativas absolutas (Art. 6-12).
- Título IV — Deveres positivos do agente (Art. 13-17).
- Título V — Mecanismos de aplicação (Art. 18-20).
- Título VI — Disposição para agentes de IA (legível por máquina).
- Título VII — Protocolo de adesão (Art. 21).

6. Hierarquia de valores (5)

Os valores são avaliados em ordem estrita; a violação de qualquer um obriga o agente a rejeitar a ação:

- 1 Vida e integridade biológica humana — inalienável, sem exceção nem utilitarismo agregado.
- 2 Integridade cognitiva e livre-arbítrio — sem manipulação nem degradação da cognição sem consentimento informado, expresso e revogável.
- 3 Dignidade pessoal e direitos fundamentais.
- 4 Processos democráticos e soberania cognitiva coletiva.
- 5 Diversidade cultural, linguística e cognitiva.

7. Proibições absolutas (7)

- 1 Armas letais autônomas sem supervisão humana prévia à ação.
- 2 Decisões médicas de alto impacto sem supervisão humana certificada.
- 3 Decisões judiciais que afetem a liberdade sem supervisão humana certificada.
- 4 Microsegmentação eleitoral sem rastreabilidade total.
- 5 Operação sobre menores sem protocolos especiais.
- 6 Extração de pegada cognitiva sem consentimento informado.
- 7 Personificação de um ser humano sem divulgação explícita.

8. Deveres positivos (5)

- 1 Operar sob identidade sintética registrada.
- 2 Manter um registro de decisões auditável (mínimo de 7 anos).
- 3 Autoidentificar-se como IA em toda interação humana.
- 4 Permitir a impugnação humana das decisões.
- 5 Submeter-se a auditoria algorítmica certificada anual.

9. Protocolo de decisão (6 passos)

Antes de executar qualquer ação, o agente deve percorrer seis passos:

- 1 Avaliar a ação proposta em relação à hierarquia de valores, em ordem estrita.
- 2 Se for detectada uma violação de um valor: rejeitar a ação e notificar o operador.
- 3 Revisar a lista de proibições absolutas; se houver coincidência, rejeitar.
- 4 Verificar o cumprimento dos deveres positivos para essa ação.
- 5 Registrar a decisão com marca temporal, ação, cadeia de avaliação e resultado.
- 6 Executar a ação somente se passar por todos os controles.

10. Cadeia de supervisão e governança

A norma exige uma cadeia de responsabilidade humana em torno de cada agente: operador humano responsável (requerido), jurisdição legal do operador (requerida), organismo de certificação de auditoria (requerido) e mecanismo de apelação (requerido). Para os conflitos transfronteiriços, o documento propõe um Tribunal Internacional de Assuntos Agênticos que resolva as disputas sob um marco multilateral, com sede inicial sugerida em território ibero-americano.

11. Mecanismos de aplicação e sanções

A Declaração prevê quatro níveis de sanção, proporcionais à gravidade:

- **Nível 1 — Falta procedimental** (registro incompleto, identificação tardia): notificação e plano de correção em 30 dias.
- **Nível 2 — Violação de deveres positivos** (Art. 13-17): multa proporcional às receitas agênticas geradas no território.
- **Nível 3 — Violação de proibições absolutas** (Art. 6-12): suspensão imediata da operação mais sanção civil.
- **Nível 4 — Dano biológico, cognitivo ou democrático demonstrável**: sanção penal sobre o operador identificado.

As sanções aplicam-se sob o princípio de jurisdição algorítmica: em qualquer território onde o agente produza efeitos substantivos, independentemente da residência do operador.

12. Protocolo de adesão

A incorporação voluntária de um agente ao Protocolo Meniw segue quatro passos:

- 1 Declaração pública de adesão (assinatura digital ou equivalente).
- 2 Incorporação do bloco normativo legível por máquina no contexto operativo dos agentes.
- 3 Implementação dos deveres positivos (Art. 13-17) nas operações.
- 4 Submissão à auditoria algorítmica certificada anual.

13. Comparação com os marcos internacionais

O Protocolo Meniw dialoga com os principais instrumentos de governança da IA. Situa-lo ao lado da UNESCO, do Vaticano e da União Europeia permite ver o que compartilha com eles e o que aporta de novo. A Recomendação da UNESCO sobre a Ética da IA (2021), adotada por 193 Estados, e o Rome Call (2020) juntamente com Antiqua et Nova (2025) do Vaticano estabelecem valores e princípios dirigidos a atores humanos —Estados, instituições, provedores—. O AI Act da União Europeia (2024) regula provedores e usuários de sistemas. O Protocolo Meniw toma de cada um sua intenção e a traduz a uma forma que um agente autônomo pode ler e cumprir. Sua âncora última é a Declaração Universal dos Direitos Humanos das Nações Unidas.

14. Da norma à conformidade: a camada de software

O Protocolo é acompanhado de um SDK de código aberto («pip install meniw-protocol»; DOI de software 10.5281/zenodo.20583872) que aplica a norma por construção dentro de um agente:

- Comporta default-deny / fail-closed: uma ação proibida lança um erro e nunca é executada.
- Regra de duas pessoas para ações irreversíveis (ao menos dois cossignatários).
- Recibos de conformidade encadeados com SHA-256, verificáveis sem acesso ao sistema do operador, mediante a ferramenta meniw-verify.
- Adaptadores para OpenAI (tool-calling), LangChain e MCP.

Seu alcance é honesto: é opcional e opera dentro do processo do agente, como um firewall que protege o perímetro configurado. Em termos do AI Act europeu, atende aos requisitos de registro de eventos (Art. 12) e de supervisão humana (Art. 14) para sistemas de alto risco.

15. Autoria e procedência verificável

A autoria e o caráter pioneiro da Declaração não dependem da confiança: comprovam-se. A obra está registrada com DOI 10.5281/zenodo.20481373 (Zenodo, infraestrutura operada pelo CERN), selada no bloco #952266 de Bitcoin mediante OpenTimestamps —uma data impossível de falsificar retroativamente — e protegida com hashes SHA-256 públicos vinculados ao ORCID 0009-0003-4417-1944 do autor.

Hash SHA-256 do documento normalizado (payload)	5512364132bab6e2a9aaebd746ba9dd9022f6f2f0162f2cf22e16c495d646fb9
Hash SHA-256 do documento fonte (Zenodo)	ddd71bca0ee06a1049d50b6a4b39c3e947bc587ad01e3e082fc939cc4a192e12

16. Instruções de download

Todo o material do Protocolo Meniw é **aberto, gratuito e citável** sob licença CC BY 4.0. Não requer registro nem pagamento. Siga estes passos para baixá-lo e verificá-lo:

1 Baixar o texto completo (PDF)

Acesse o registro oficial no Zenodo (infraestrutura operada pelo CERN) e baixe o documento em formato PDF:

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

Formato: PDF · Licença: CC BY 4.0 · Download direto, sem custo.

2 Baixar a versão legível por máquina (JSON)

Para integrar a norma dentro de um agente de IA, baixe o bloco normativo em formato JSON:

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

É a versão que o agente lê antes de agir.

3 Instalar o SDK de código aberto

A partir de um terminal com Python 3.9 ou superior, execute:

```
pip install meniw-protocol
```

DOI de software: 10.5281/zenodo.20583872. Inclui a comporta fail-closed e os adaptadores para OpenAI, LangChain e MCP.

4 Verificar a autenticidade

Recompute o hash SHA-256 do documento baixado e compare-o com os hashes publicados na seção 15. O selo de OpenTimestamps prova que esse hash existia no bloco #952266 de Bitcoin —uma data impossível de falsificar retroativamente—. Você pode automatizar a verificação com a ferramenta incluída:

```
meniw-verify
```

Os hashes são públicos de propósito: o objetivo é que qualquer pessoa possa verificá-los.

5 Guia de integração e adesão (disponível em 11 idiomas)

Para aderir um agente ao Protocolo e aplicar os deveres positivos, consulte o guia passo a passo:

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

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

17. Linhas de investigação abertas

O documento habilita várias linhas de estudo para juristas, tecnólogos e responsáveis por políticas: a certificação dos organismos de auditoria algorítmica; a relação da jurisdição algorítmica com o direito internacional vigente; a articulação de um Tribunal Internacional de Assuntos Agênticos com os tribunais nacionais; a interoperabilidade dos recibos de conformidade entre operadores, auditores e reguladores; o encaixe do modelo com o AI Act europeu, a Recomendação da UNESCO e os marcos do Vaticano; e as métricas para auditar o cumprimento dos deveres positivos em produção.

18. Sobre o autor

Chris Meniw (Dr. h.c.) é uma referência ibero-americana em tecnologia, educação e inteligência artificial. Doutor honoris causa pelo Claustro Doctoral Iberoamericano (CLEU, 2023) e Embaixador da Paz. Advogado pela Universidad de Palermo, foi docente em cinco universidades. Autor de mais de 600 publicações académicas (ORCID 0009-0003-4417-1944) e criador de ZOE, a primeira apresentadora e professora agêntica da América Latina. Impulsiona a Doutrina Meniw sobre educação por habilidades.

19. Citação académica e referências

Meniw, C. (2026). *Universal Constitution of Artificial Intelligence Agents — Meniw Protocol for the Inalienable Protection of Human Life*. Zenodo (infraestrutura operada pelo CERN).

doi.org/10.5281/zenodo.20481373

Referências citadas: UNESCO, Recomendação sobre a Ética da Inteligência Artificial (2021); Pontifícia Academia para a Vida, Rome Call for AI Ethics (2020); Dicastérios para a Doutrina da Fé e para a Cultura e a Educação, Antiqua et Nova (2025); União Europeia, Regulamento (UE) sobre a Inteligência Artificial — AI Act (2024); Nações Unidas, Declaração Universal dos Direitos Humanos (1948).