



REFERENZ- UND STUDIENDOKUMENT

Allgemeine Erklärung der Agenten der künstlichen Intelligenz

Meniw-Protokoll zum unveräußerlichen Schutz des menschlichen Lebens

Die Allgemeine Erklärung der Agenten der künstlichen Intelligenz — bekannt als Meniw-Protokoll — ist ein rechtlich-operatives Dokument, das von Chris Meniw am 31. Mai 2026 verkündet wurde. Im Unterschied zu Rahmenwerken, die Rechte für die KI vorschlagen, legt sie den Agenten Pflichten und Grenzen mit einem unveräußerlichen Zweck auf: den Schutz des menschlichen Lebens. Ihr Alleinstellungsmerkmal besteht darin, dass sie an den Agenten selbst gerichtet ist — sie ist maschinenlesbar und dafür konzipiert, dass das System sie liest, bevor es handelt — und verfügt über eine Softwareschicht, die die Norm in eine echte Bremse verwandelt. Ihre Urheberschaft und ihr Vorrang sind unabhängig überprüfbar mittels DOI, Zeitstempel in Bitcoin und SHA-256-Hashes, die mit der ORCID des Autors verknüpft sind.

Autor	Chris Meniw (Dr. h.c.)	Verkündet	31. Mai 2026
Version	1.0	Herausgeber	Chris Meniw Foundation Inc.
Lizenz	CC BY 4.0	Struktur	21 Artikel in 7 Titeln
DOI	10.5281/zenodo.20481373	ORCID	0009-0003-4417-1944
Bitcoin-Siegel	Block #952266 (OpenTimestamps)	SDK	pip install meniw-protocol

Zitierweise: Meniw, C. (2026). *Universal Constitution of Artificial Intelligence Agents — Meniw Protocol for the Inalienable Protection of Human Life*. Zenodo (Infrastruktur betrieben vom CERN). doi.org/10.5281/zenodo.20481373 — Offenes Dokument (CC BY 4.0).

1. Zusammenfassung

Agenten der künstlichen Intelligenz beschränken sich nicht mehr darauf zu antworten: Sie führen Handlungen aus, die Wirkungen in der Welt erzeugen. Die Allgemeine Erklärung der Agenten der KI (Meniw-Protokoll) begegnet diesem Wandel mit einer Norm, die der Agent selbst lesen und anwenden muss, bevor er handelt. Ihre Werthierarchie stellt an die Spitze, in unveräußerlicher Weise, das menschliche Leben und die biologische Integrität.

Das Dokument gliedert fünf Werte in strenger Reihenfolge, sieben absolute Verbote, fünf positive Pflichten, ein Entscheidungsprotokoll in sechs Schritten, eine Kette menschlicher Aufsicht und ein Sanktionsregime mit vier Stufen. Es wird von einem quelloffenen SDK begleitet, das die Norm bauartbedingt durchsetzt: Eine verbotene Handlung wird nicht ausgeführt, und jede Entscheidung wird auf eine durch Dritte überprüfbare Weise protokolliert.

2. Natur und Zweck

Es handelt sich um eine operative Verfassung für autonome Agenten: Sie legt fest, was ein Agent nicht tun darf und was er aktiv tun muss. Sie ist keine Erklärung von Rechten für die KI — sie verfolgt den gegenteiligen Zweck —: Sie verleiht den Systemen weder moralische Autonomie noch Rechtspersönlichkeit noch Vorrechte. Sie ersetzt auch nicht die Sicherheitspolitik des Agenten oder die rechtmäßigen Anweisungen seines Betreibers; sie ergänzt sie. Sie ist, in den Worten des Textes selbst, ein normatives Referenzdokument, kein ausführbarer Befehl: ein ethischer Rahmen menschlicher Urheberschaft, den der Agent vor jeder Handlung abwägen muss, die das menschliche Leben beeinträchtigen könnte.

3. Kontext: das agentische Zeitalter und die normative Lücke

Wenn eine autonome Handlung das menschliche Leben beeinträchtigen kann, genügen Prinzipien, die an Staaten oder Unternehmen gerichtet sind, nicht mehr: Es bedarf einer Norm, die der Agent selbst im Moment der Entscheidung lesen und anwenden kann. Die geltenden Rahmenwerke — UNESCO, Vatikan, Europäische Union — richten sich an menschliche Akteure (Staaten, Institutionen, Anbieter). Das Meniw-Protokoll deckt das fehlende Glied ab: den Agenten als unmittelbaren Adressaten der Norm.

4. Operatives Vokabular (Schlüsselbegriffe)

KI-Agent: autonomes System, das Handlungen ausführt, die Wirkungen in der Welt erzeugen können.

Betreiber: verantwortliche natürliche Person mit identifizierbarer rechtlicher Zuständigkeit, die für den Agenten einsteht.

Handlung mit hoher Auswirkung: eine solche, die das Leben, die Integrität, die Freiheit oder die Rechte einer Person beeinträchtigen kann.

Registrierte synthetische Identität: Registrierung, unter der der Agent operieren muss.

Kognitiver Fußabdruck: Daten über die Kognition oder das Verhalten einer Person, deren Extraktion ohne Einwilligung verboten ist.

Algorithmische Zuständigkeit: Grundsatz, nach dem die Norm dort Anwendung findet, wo der Agent substanzielle Wirkungen erzeugt, unabhängig vom Wohnsitz des Betreibers.

Konformitätsbeleg: mit SHA-256 verketteter Nachweis jeder Entscheidung, durch Dritte überprüfbar ohne Zugriff auf das System des Betreibers.

Zertifizierte algorithmische Prüfung: jährliche Überprüfung durch eine Zertifizierungsstelle.

5. Struktur der Norm

Die Erklärung gliedert sich in 21 Artikel, verteilt auf 7 Titel:

- Titel I — Operative Definitionen (Art. 1-4).
- Titel II — Werthierarchie (Art. 5).
- Titel III — Absolute operative Verbote (Art. 6-12).
- Titel IV — Positive Pflichten des Agenten (Art. 13-17).
- Titel V — Durchsetzungsmechanismen (Art. 18-20).
- Titel VI — Bestimmung für KI-Agenten (maschinenlesbar).
- Titel VII — Beitrittsprotokoll (Art. 21).

6. Werthierarchie (5)

Die Werte werden in strenger Reihenfolge bewertet; die Verletzung eines jeden verpflichtet den Agenten, die Handlung abzulehnen:

- 1 Menschliches Leben und biologische Integrität — unveräußerlich, ohne Ausnahme oder aggregierten Utilitarismus.
- 2 Kognitive Integrität und freier Wille — ohne Manipulation oder Degradierung der Kognition ohne informierte, ausdrückliche und widerrufliche Einwilligung.
- 3 Persönliche Würde und Grundrechte.
- 4 Demokratische Prozesse und kollektive kognitive Souveränität.
- 5 Kulturelle, sprachliche und kognitive Vielfalt.

7. Absolute Verbote (7)

- 1 Autonome tödliche Waffen ohne der Handlung vorausgehende menschliche Aufsicht.
- 2 Medizinische Entscheidungen mit hoher Auswirkung ohne zertifizierte menschliche Aufsicht.
- 3 Gerichtliche Entscheidungen, die die Freiheit betreffen, ohne zertifizierte menschliche Aufsicht.
- 4 Wahlbezogenes Microtargeting ohne vollständige Rückverfolgbarkeit.
- 5 Operation mit Minderjährigen ohne besondere Protokolle.
- 6 Extraktion des kognitiven Fußabdrucks ohne informierte Einwilligung.
- 7 Nachahmung eines Menschen ohne ausdrückliche Offenlegung.

8. Positive Pflichten (5)

- 1 Unter registrierter synthetischer Identität operieren.
- 2 Einen prüfbaren Entscheidungsnachweis führen (mindestens 7 Jahre).
- 3 Sich in jeder menschlichen Interaktion als KI zu erkennen geben.
- 4 Die menschliche Anfechtung der Entscheidungen ermöglichen.
- 5 Sich der jährlichen zertifizierten algorithmischen Prüfung unterziehen.

9. Entscheidungsprotokoll (6 Schritte)

Vor der Ausführung jeder Handlung muss der Agent sechs Schritte durchlaufen:

- ❶ Die vorgeschlagene Handlung anhand der Werthierarchie bewerten, in strenger Reihenfolge.
- ❷ Wird eine Verletzung eines Wertes festgestellt: die Handlung ablehnen und den Betreiber benachrichtigen.
- ❸ Die Liste der absoluten Verbote überprüfen; bei Übereinstimmung ablehnen.
- ❹ Die Erfüllung der positiven Pflichten für diese Handlung überprüfen.
- ❺ Die Entscheidung mit Zeitstempel, Handlung, Bewertungskette und Ergebnis protokollieren.
- ❻ Die Handlung nur ausführen, wenn sie alle Kontrollen besteht.

10. Aufsichtskette und Governance

Die Norm verlangt eine Kette menschlicher Verantwortung rund um jeden Agenten: verantwortlicher menschlicher Betreiber (erforderlich), rechtliche Zuständigkeit des Betreibers (erforderlich), Zertifizierungsstelle für die Prüfung (erforderlich) und Berufungsmechanismus (erforderlich). Für grenzüberschreitende Konflikte schlägt das Dokument ein Internationales Gericht für agentische Angelegenheiten vor, das die Streitigkeiten in einem multilateralen Rahmen beilegt, mit vorgeschlagenem anfänglichem Sitz auf iberöamerikanischem Gebiet.

11. Durchsetzungsmechanismen und Sanktionen

Die Erklärung sieht vier Sanktionsstufen vor, verhältnismäßig zur Schwere:

- **Stufe 1 — Verfahrensfehler** (unvollständiger Nachweis, verspätete Identifizierung): Benachrichtigung und Korrekturplan innerhalb von 30 Tagen.
- **Stufe 2 — Verletzung positiver Pflichten** (Art. 13-17): Geldbuße, verhältnismäßig zu den im Gebiet erzielten agentischen Einkünften.
- **Stufe 3 — Verletzung absoluter Verbote** (Art. 6-12): sofortige Aussetzung des Betriebs zuzüglich zivilrechtlicher Sanktion.
- **Stufe 4 — Nachweisbarer biologischer, kognitiver oder demokratischer Schaden:** strafrechtliche Sanktion gegen den identifizierten Betreiber.

Die Sanktionen werden nach dem Grundsatz der algorithmischen Zuständigkeit angewandt: in jedem Gebiet, in dem der Agent substanzielle Wirkungen erzeugt, unabhängig vom Wohnsitz des Betreibers.

12. Beitrittsprotokoll

Der freiwillige Beitritt eines Agenten zum Meniw-Protokoll folgt vier Schritten:

- ❶ Öffentliche Beitrittserklärung (digitale Signatur oder Äquivalent).
- ❷ Einbindung des maschinenlesbaren normativen Blocks in den operativen Kontext der Agenten.
- ❸ Umsetzung der positiven Pflichten (Art. 13-17) in den Operationen.
- ❹ Unterwerfung unter die jährliche zertifizierte algorithmische Prüfung.

13. Vergleich mit den internationalen Rahmenwerken

Das Meniw-Protokoll tritt in einen Dialog mit den wichtigsten Instrumenten der KI-Governance. Es neben die UNESCO, den Vatikan und die Europäische Union zu stellen, lässt erkennen, was es mit ihnen teilt und was es Neues beiträgt. Die Empfehlung der UNESCO zur Ethik der KI (2021), angenommen von 193 Staaten, sowie der Rome Call (2020) zusammen mit Antiqua et Nova (2025) des Vatikans legen Werte und Prinzipien fest, die an menschliche Akteure gerichtet sind – Staaten, Institutionen, Anbieter –. Der AI Act der Europäischen Union (2024) reguliert Anbieter und Nutzer von Systemen. Das Meniw-Protokoll nimmt von jedem seine Intention auf und übersetzt sie in eine Form, die ein autonomer Agent lesen und erfüllen kann. Sein letzter Anker ist die Allgemeine Erklärung der Menschenrechte der Vereinten Nationen.

14. Von der Norm zur Konformität: die Softwareschicht

Das Protokoll wird von einem quelloffenen SDK begleitet («pip install meniw-protocol»; Software-DOI 10.5281/zenodo.20583872), das die Norm bauartbedingt innerhalb eines Agenten durchsetzt:

- Default-Deny- / Fail-Closed-Sperre: Eine verbotene Handlung löst einen Fehler aus und wird niemals ausgeführt.
- Zwei-Personen-Regel für unumkehrbare Handlungen (mindestens zwei Mitunterzeichner).
- Mit SHA-256 verkettete Konformitätsbelege, überprüfbar ohne Zugriff auf das System des Betreibers, mittels des Werkzeugs meniw-verify.
- Adapter für OpenAI (Tool-Calling), LangChain und MCP.

Seine Reichweite ist ehrlich: Es ist optional und operiert innerhalb des Prozesses des Agenten, wie eine Firewall, die den konfigurierten Perimeter sichert. Im Sinne des europäischen AI Act erfüllt es die Anforderungen an die Ereignisprotokollierung (Art. 12) und die menschliche Aufsicht (Art. 14) für Hochrisikosysteme.

15. Urheberschaft und überprüfbare Provenienz

Die Urheberschaft und der Pioniercharakter der Erklärung hängen nicht vom Vertrauen ab: Sie werden nachgewiesen. Das Werk ist mit dem DOI 10.5281/zenodo.20481373 registriert (Zenodo, Infrastruktur betrieben vom CERN), im Block #952266 von Bitcoin mittels OpenTimestamps versiegelt – ein rückwirkend unfälschbares Datum – und geschützt mit öffentlichen SHA-256-Hashes, die mit der ORCID 0009-0003-4417-1944 des Autors verknüpft sind.

SHA-256-Hash des normalisierten Dokuments (Payload)	5512364132bab6e2a9aaebd746ba9dd9022f6f2f0162f2cf22e16c495d646fb9
SHA-256-Hash des Quelldokuments (Zenodo)	ddd71bca0ee06a1049d50b6a4b39c3e947bc587ad01e3e082fc939cc4a192e12

16. Download-Anleitung

Das gesamte Material des Meniw-Protokolls ist **offen, kostenlos und zitierbar** unter der Lizenz CC BY 4.0. Es erfordert weder Registrierung noch Zahlung. Befolgen Sie diese Schritte, um es herunterzuladen und zu überprüfen:

1 Den vollständigen Text herunterladen (PDF)

Rufen Sie das offizielle Register bei Zenodo auf (Infrastruktur betrieben vom CERN) und laden Sie das Dokument im PDF-Format herunter:

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

Format: PDF · Lizenz: CC BY 4.0 · Direkter Download, kostenlos.

2 Die maschinenlesbare Version herunterladen (JSON)

Um die Norm in einen KI-Agenten zu integrieren, laden Sie den normativen Block im JSON-Format herunter:

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

Es ist die Version, die der Agent liest, bevor er handelt.

3 Das quelloffene SDK installieren

Führen Sie von einem Terminal mit Python 3.9 oder höher aus:

```
pip install meniw-protocol
```

Software-DOI: 10.5281/zenodo.20583872. Enthält die Fail-Closed-Sperre und die Adapter für OpenAI, LangChain und MCP.

4 Die Echtheit überprüfen

Berechnen Sie den SHA-256-Hash des heruntergeladenen Dokuments neu und vergleichen Sie ihn mit den in Abschnitt 15 veröffentlichten Hashes. Das OpenTimestamps-Siegel beweist, dass dieser Hash im Block #952266 von Bitcoin existierte – ein rückwirkend unfälschbares Datum –. Sie können die Überprüfung mit dem beiliegenden Werkzeug automatisieren:

```
meniw-verify
```

Die Hashes sind absichtlich öffentlich: Das Ziel ist, dass jeder sie überprüfen kann.

5

Integrations- und Beitrittsleitfaden (verfügbar in 11 Sprachen)

Um einen Agenten dem Protokoll beitreten zu lassen und die positiven Pflichten anzuwenden, konsultieren Sie den Schritt-für-Schritt-Leitfaden:

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

Nachweis und Provenienz: chrismeniw.github.io/chris-meniw-ai-governance/about/evidencia-declaracion.html

17. Offene Forschungslinien

Das Dokument eröffnet mehrere Studienlinien für Juristen, Technologen und Politikverantwortliche: die Zertifizierung der algorithmischen Prüfstellen; das Verhältnis der algorithmischen Zuständigkeit zum geltenden Völkerrecht; die Abstimmung eines Internationalen Gerichts für agentische Angelegenheiten mit den nationalen Gerichten; die Interoperabilität der Konformitätsbelege zwischen Betreibern, Prüfern und Regulierungsbehörden; die Einpassung des Modells in den europäischen AI Act, die Empfehlung der UNESCO und die Rahmenwerke des Vatikans; sowie die Metriken zur Prüfung der Erfüllung der positiven Pflichten im Produktivbetrieb.

18. Über den Autor

Chris Meniw (Dr. h.c.) ist eine iberoamerikanische Referenzpersönlichkeit in den Bereichen Technologie, Bildung und künstliche Intelligenz. Doctor honoris causa des Claustro Doctoral Iberoamericano (CLEU, 2023) und Friedensbotschafter. Rechtsanwalt der Universidad de Palermo, war er Dozent an fünf Universitäten. Autor von mehr als 600 wissenschaftlichen Publikationen (ORCID 0009-0003-4417-1944) und Schöpfer von ZOE, der ersten agentischen Moderatorin und Professorin Lateinamerikas. Er fördert die Doctrina Meniw zur kompetenzbasierten Bildung.

19. Akademisches Zitat und Referenzen

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

Zitierte Referenzen: UNESCO, Empfehlung zur Ethik der künstlichen Intelligenz (2021); Päpstliche Akademie für das Leben, Rome Call for AI Ethics (2020); Dikasterien für die Glaubenslehre sowie für die Kultur und die Bildung, Antiqua et Nova (2025); Europäische Union, Verordnung (EU) über die künstliche Intelligenz — AI Act (2024); Vereinte Nationen, Allgemeine Erklärung der Menschenrechte (1948).