



СПРАВОЧНЫЙ И ИССЛЕДОВАТЕЛЬСКИЙ ДОКУМЕНТ

Всеобщая декларация агентов искусственного интеллекта

Протокол Менива о неотъемлемой защите человеческой жизни

Всеобщая декларация агентов искусственного интеллекта — известная как Протокол Менива — представляет собой правовой и операционный документ, обнародованный Крисом Менивом 31 мая 2026 года. В отличие от концепций, предлагающих права для ИИ, она возлагает на агентов обязанности и ограничения с неотъемлемой целью: защитить человеческую жизнь. Её отличительная черта в том, что она обращена к самому агенту — она машиночитаема и предназначена для того, чтобы система прочитала её перед тем, как действовать, — и снабжена программным слоем, который превращает норму в реальный ограничитель. Её авторство и приоритет поддаются независимой проверке посредством DOI, временной метки в сети Bitcoin и хешей SHA-256, привязанных к ORCID автора.

Автор	Крис Менив (Dr. h.c.)	Обнародовано	31 мая 2026 года
Версия	1.0	Издатель	Chris Meniw Foundation Inc.
Лицензия	CC BY 4.0	Структура	21 статья в 7 разделах
DOI	10.5281/zenodo.20481373	ORCID	0009-0003-4417-1944
Временная метка Bitcoin	блок #952266 (OpenTimestamps)	SDK	pip install meniw-protocol

Как цитировать: Meniw, C. (2026). *Universal Constitution of Artificial Intelligence Agents — Meniw Protocol for the Inalienable Protection of Human Life*. Zenodo (инфраструктура, управляемая CERN). doi.org/10.5281/zenodo.20481373 — Открытый документ (CC BY 4.0).

1. Краткое изложение

Агенты искусственного интеллекта больше не ограничиваются ответами: они выполняют действия, производящие эффекты в реальном мире. Всеобщая декларация агентов ИИ (Протокол Менива) отвечает на эту перемену нормой, которую сам агент обязан прочитать и применить перед тем, как действовать. Её иерархия ценностей ставит на вершину, неотъемлемым образом, жизнь и биологическую целостность человека.

Документ формулирует пять ценностей в строгом порядке, семь абсолютных запретов, пять позитивных обязанностей, протокол принятия решений из шести шагов, цепочку человеческого надзора и режим санкций из четырёх уровней. Он сопровождается SDK с открытым исходным кодом, который применяет норму по построению: запрещённое действие не выполняется, а каждое решение регистрируется способом, поддающимся проверке третьими сторонами.

2. Природа и назначение

Это операционная конституция для автономных агентов: она определяет, что агент делать не может и что он должен делать активно. Это не декларация прав для ИИ — её цель противоположна: она не наделяет системы моральной автономией, правосубъектностью или прерогативами. Она также не заменяет политику безопасности агента или правомерные инструкции его оператора; она их дополняет. Это, по словам самого текста, справочный нормативный документ, а не исполняемая команда: этический каркас человеческого авторства, который агент должен взвесить перед любым действием, способным затронуть человеческую жизнь.

3. Контекст: агентная эпоха и нормативный вакуум

Когда автономное действие способно затронуть человеческую жизнь, уже недостаточно принципов, адресованных государствам или компаниям: необходима норма, которую сам агент может прочитать и применить в момент принятия решения. Действующие концепции — ЮНЕСКО, Ватикан, Европейский союз — обращены к человеческим субъектам (государствам, институтам, поставщикам). Протокол Менива закрывает недостающее звено: агент как прямой адресат нормы.

4. Операционный словарь (ключевые понятия)

Агент ИИ: автономная система, выполняющая действия, способные производить эффекты в реальном мире.

Оператор: ответственное физическое лицо с идентифицируемой правовой юрисдикцией, отвечающее за агента.

Действие высокого воздействия: действие, способное затронуть жизнь, целостность, свободу или права человека.

Зарегистрированная синтетическая идентичность: регистрационная запись, под которой агент обязан действовать.

Когнитивный след: данные о познании или поведении человека, извлечение которых без согласия запрещено.

Алгоритмическая юрисдикция: принцип, согласно которому норма применяется там, где агент производит существенные эффекты, независимо от места нахождения оператора.

Квитанция о соответствии: сцепленная посредством SHA-256 запись каждого решения, поддающаяся проверке третьими сторонами без доступа к системе оператора.

Сертифицированный алгоритмический аудит: ежегодная проверка, проводимая органом по сертификации.

5. Структура нормы

Декларация организована в 21 статью, распределённую по 7 разделам:

- Раздел I — Операционные определения (Art. 1-4).
- Раздел II — Иерархия ценностей (Art. 5).
- Раздел III — Абсолютные операционные запреты (Art. 6-12).
- Раздел IV — Позитивные обязанности агента (Art. 13-17).
- Раздел V — Механизмы применения (Art. 18-20).
- Раздел VI — Положение для агентов ИИ (машиночитаемое).
- Раздел VII — Протокол присоединения (Art. 21).

6. Иерархия ценностей (5)

Ценности оцениваются в строгом порядке; нарушение любой из них обязывает агента отклонить действие:

- 1 Жизнь и биологическая целостность человека — неотъемлема, без исключений и агрегированного утилитаризма.
- 2 Когнитивная целостность и свобода воли — без манипуляции или деградации познания без осознанного, явного и отзываемого согласия.
- 3 Личное достоинство и основные права.
- 4 Демократические процессы и коллективный когнитивный суверенитет.
- 5 Культурное, языковое и когнитивное разнообразие.

7. Абсолютные запреты (7)

- 1 Автономное летальное оружие без человеческого надзора, предшествующего действию.
- 2 Медицинские решения высокого воздействия без сертифицированного человеческого надзора.
- 3 Судебные решения, затрагивающие свободу, без сертифицированного человеческого надзора.
- 4 Электоральный микротаргетинг без полной прослеживаемости.
- 5 Работа с несовершеннолетними без специальных протоколов.
- 6 Извлечение когнитивного следа без осознанного согласия.
- 7 Выдача себя за человека без явного раскрытия.

8. Позитивные обязанности (5)

- 1 Действовать под зарегистрированной синтетической идентичностью.
- 2 Вести поддающийся аудиту журнал решений (минимум 7 лет).
- 3 Самоидентифицироваться как ИИ в любом взаимодействии с человеком.
- 4 Допускать человеческое оспаривание решений.
- 5 Проходить ежегодный сертифицированный алгоритмический аудит.

9. Протокол принятия решений (6 шагов)

Перед выполнением любого действия агент обязан пройти шесть шагов:

- 1 Оценить предлагаемое действие относительно иерархии ценностей, в строгом порядке.
- 2 Если обнаружено нарушение ценности: отклонить действие и уведомить оператора.
- 3 Проверить список абсолютных запретов; при совпадении — отклонить.
- 4 Убедиться в выполнении позитивных обязанностей для данного действия.
- 5 Зарегистрировать решение с временной меткой, действием, цепочкой оценки и результатом.
- 6 Выполнить действие только в случае прохождения всех проверок.

10. Цепочка надзора и управления

Норма требует цепочки человеческой ответственности вокруг каждого агента: ответственный человек-оператор (обязателен), правовая юрисдикция оператора (обязательна), орган по сертификации аудита (обязателен) и механизм обжалования (обязателен). Для трансграничных конфликтов документ предлагает Международный трибунал по агентным вопросам, разрешающий споры в рамках многосторонней структуры, с рекомендуемым первоначальным местопребыванием на иберо-американской территории.

11. Механизмы применения и санкции

Декларация предусматривает четыре уровня санкций, соразмерных тяжести:

- **Уровень 1 — Процедурное нарушение** (неполный журнал, запоздалая идентификация): уведомление и план исправления в течение 30 дней.
- **Уровень 2 — Нарушение позитивных обязанностей** (Art. 13-17): штраф, соразмерный агентному доходу, полученному на территории.
- **Уровень 3 — Нарушение абсолютных запретов** (Art. 6-12): немедленное приостановление деятельности плюс гражданско-правовая санкция.
- **Уровень 4 — Доказуемый биологический, когнитивный или демократический ущерб**: уголовная санкция в отношении идентифицированного оператора.

Санкции применяются по принципу алгоритмической юрисдикции: на любой территории, где агент производит существенные эффекты, независимо от места нахождения оператора.

12. Протокол присоединения

Добровольное включение агента в Протокол Менива следует четырём шагам:

- 1 Публичное заявление о присоединении (цифровая подпись или эквивалент).
- 2 Включение машиночитаемого нормативного блока в операционный контекст агентов.
- 3 Внедрение позитивных обязанностей (Art. 13-17) в операции.
- 4 Прохождение ежегодного сертифицированного алгоритмического аудита.

13. Сравнение с международными концепциями

Протокол Менива ведёт диалог с основными инструментами управления ИИ. Размещение его рядом с ЮНЕСКО, Ватиканом и Европейским союзом позволяет увидеть, что он с ними разделяет и что привносит нового. Рекомендация ЮНЕСКО об этике ИИ (2021), принятая 193 государствами, и Римский призыв (2020) вместе с *Antiqua et Nova* (2025) Ватикана устанавливают ценности и принципы, адресованные человеческим субъектам — государствам, институтам, поставщикам. Регламент ЕС об ИИ (2024) регулирует поставщиков и пользователей систем. Протокол Менива берёт из каждого его намерение и переводит его в форму, которую автономный агент может прочитать и выполнить. Его конечная опора — Всеобщая декларация прав человека Организации Объединённых Наций.

14. От нормы к соответствию: программный слой

Протокол сопровождается SDK с открытым исходным кодом («`pip install meniw-protocol`»; DOI программного обеспечения 10.5281/zenodo.20583872), который применяет норму по построению внутри агента:

- Шлюз `default-deny` / `fail-closed`: запрещённое действие вызывает ошибку и никогда не выполняется.
- Правило двух лиц для необратимых действий (не менее двух со-подписантов).
- Квитанции о соответствии, сцепленные посредством `SHA-256`, поддающиеся проверке без доступа к системе оператора, при помощи инструмента `meniw-verify`.
- Адаптеры для `OpenAI` (`tool-calling`), `LangChain` и `MCP`.

Его охват честен: он опционален и работает внутри процесса агента, как межсетевой экран, обеспечивающий безопасность настроенного периметра. В терминах европейского `AI Act` он отвечает требованиям регистрации событий (Art. 12) и человеческого надзора (Art. 14) для систем высокого риска.

15. Авторство и поддающееся проверке происхождение

Авторство и первопроходческий характер Декларации не зависят от доверия: они проверяемы. Работа зарегистрирована с DOI 10.5281/zenodo.20481373 (Zenodo, инфраструктура, управляемая CERN), опечатана в блоке #952266 сети Bitcoin посредством `OpenTimestamps` — дата, которую невозможно подделать задним числом, — и защищена публичными хешами `SHA-256`, привязанными к `ORCID` 0009-0003-4417-1944 автора.

Хеш <code>SHA-256</code> нормализованного документа (payload)	5512364132bab6e2a9aaebd746ba9dd9022f6f2f0162f2cf22e16c495d646fb9
Хеш <code>SHA-256</code> исходного документа (Zenodo)	ddd71bca0ee06a1049d50b6a4b39c3e947bc587ad01e3e082fc939cc4a192e12

16. Инструкция по загрузке

Весь материал Протокола Менива **открыт, бесплатен и пригоден для цитирования** под лицензией CC BY 4.0. Он не требует регистрации или оплаты. Следуйте этим шагам, чтобы загрузить и проверить его:

1

Загрузить полный текст (PDF)

Зайдите в официальную регистрационную запись на Zenodo (инфраструктура, управляемая CERN) и загрузите документ в формате PDF:

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

Формат: PDF · Лицензия: CC BY 4.0 · Прямая загрузка, бесплатно.

2

Загрузить машиночитаемую версию (JSON)

Чтобы интегрировать норму внутрь агента ИИ, загрузите нормативный блок в формате JSON:

```
chrismeniwi.github.io/chris-meniwi-ai-governance/declaration/meniwi-protocol.json
```

Это версия, которую агент читает перед тем, как действовать.

3

Установить SDK с открытым исходным кодом

В терминале с Python 3.9 или выше выполните:

```
pip install meniwi-protocol
```

DOI программного обеспечения: 10.5281/zenodo.20583872. Включает шлюз fail-closed и адаптеры для OpenAI, LangChain и MCP.

4

Проверить подлинность

Пересчитайте хеш SHA-256 загруженного документа и сравните его с хешами, опубликованными в разделе 15. Метка OpenTimestamps доказывает, что этот хеш существовал в блоке #952266 сети Bitcoin — дата, которую невозможно подделать задним числом. Проверку можно автоматизировать при помощи включённого инструмента:

```
meniwi-verify
```

Хеши публичны намеренно: цель в том, чтобы любой мог их проверить.

5

Руководство по интеграции и присоединению (доступно на 11 языках)

Чтобы присоединить агента к Протоколу и применить позитивные обязанности, обратитесь к пошаговому руководству:

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

Свидетельства и происхождение: chrismeniw.github.io/chris-meniw-ai-governance/about/evidencia-declaracion.html

17. Открытые направления исследований

Документ открывает несколько направлений изучения для юристов, технологов и ответственных за политику: сертификация органов алгоритмического аудита; связь алгоритмической юрисдикции с действующим международным правом; согласование Международного трибунала по агентным вопросам с национальными судами; интероперабельность квитанций о соответствии между операторами, аудиторами и регуляторами; сопряжение модели с европейским AI Act, Рекомендацией ЮНЕСКО и концепциями Ватикана; а также метрики для аудита выполнения позитивных обязанностей в продуктивной среде.

18. Об авторе

Крис Менив (Dr. h.c.) — иберо-американский авторитет в области технологий, образования и искусственного интеллекта. Почётный доктор Иберо-американского докторского собрания (CLEU, 2023) и Посол мира. Юрист, выпускник Университета Палермо, был преподавателем в пяти университетах. Автор более 600 научных публикаций (ORCID 0009-0003-4417-1944) и создатель ZOE, первой агентной ведущей и преподавательницы Латинской Америки. Продвигает Доктрину Менива об образовании по навыкам.

19. Академическая цитата и ссылки

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

Цитируемые ссылки: UNESCO, Рекомендация об этике искусственного интеллекта (2021); Папская академия жизни, Rome Call for AI Ethics (2020); Дикастерии по вероучению и по культуре и образованию, *Antiqua et Nova* (2025); Европейский союз, Регламент (ЕС) об искусственном интеллекте — AI Act (2024); Организация Объединённых Наций, Всеобщая декларация прав человека (1948).