



AI Ethics Assessment Framework: Policy Document for Harmonized Compliance

Mapping to International Standards and Regulatory Frameworks

Public Edition

International Board of Ethics in AI
Aethics API Development Team

2 September 2026

PUBLIC EDITION — Released for Open Distribution

Copyright © 2026 AdvancedEthics.ai (“Aethics”). This document is released under a Creative Commons Attribution 4.0 International (CC BY 4.0) license: it may be freely shared, reproduced, and adapted with attribution.

The assessment methodology, scoring metrics, and framework mappings described in this document are being released as open source under the Apache License 2.0, beginning with the `aethics-eval` model evaluation toolkit. The following remain proprietary to AdvancedEthics.ai: the curated benchmark and red-team evaluation datasets; the certification issuance infrastructure, including signing keys and the certificate registry; and the hosted managed service and enterprise features. “Aethics-Certified” is a certification mark of the International Board of Ethics in AI (IBEAI); computing a score under the open methodology does not constitute certification.

Disclaimer: Assessments produced under this framework are structured, evidence-based evaluations. They do not constitute legal advice, regulatory certification, or a guarantee of compliance. Organizations should consult qualified legal counsel for jurisdiction-specific obligations. Use of the Aethics API is subject to the published Terms of Service. Inquiries: compliance@aethics.com.

Abstract

This policy document establishes a comprehensive framework for AI ethics assessment aligned with international standards, regulations, and policy frameworks. The Aethics API provides a harmonized compliance layer that translates complex, overlapping international AI governance instruments into actionable, auditable controls. This document outlines the methodology for mapping API features to specific regulatory sections, explains conflict resolution approaches, and demonstrates a defensible methodology suitable for regulatory review, audits, and enterprise adoption. The framework prioritizes binding legal regulations while incorporating voluntary guidance to enrich ethical robustness and ensure international interoperability. This public edition is released openly, alongside the phased open-source publication of the underlying assessment methodology and framework mappings, so that the methodology can be inspected, audited, and improved by the community it serves.

Contents

1	Introduction	4
1.1	Purpose and Scope	4
1.2	Implementation Status	4
1.3	Intended Audience	5
1.4	Document Structure	5
1.5	Definitions and Interpretation	5
1.6	Open Release and Licensing	6
2	Governance and Risk Classification Framework	6
2.1	Regulatory Foundation	6
2.1.1	EU AI Act (2024)	6
2.1.2	NIST AI Risk Management Framework	7
2.1.3	OECD AI Principles	7
2.2	API Implementation	7
2.3	Harmonization Approach	7
3	General-Purpose AI Models (GPAI)	8
3.1	Regulatory Foundation	8
3.1.1	EU AI Act — GPAI Provisions	8
3.1.2	NIST AI RMF — Foundation Model Considerations	8
3.2	API Implementation	8
3.3	Harmonization Approach	9
4	Privacy, Data Governance, and Data Protection	9
4.1	Regulatory Foundation	9
4.1.1	General Data Protection Regulation (GDPR)	9
4.1.2	EU AI Act	9
4.1.3	OECD AI Principles	9
4.1.4	UNESCO Recommendation on AI Ethics	10
4.2	API Implementation	10
4.3	Harmonization Approach	10
5	Transparency and Explainability Requirements	10
5.1	Regulatory Foundation	10
5.1.1	EU AI Act	10
5.1.2	NIST AI RMF	11
5.1.3	IEEE Ethically Aligned Design	11
5.1.4	OECD AI Principles	11
5.2	API Implementation	11

5.3	Conflict Resolution	11
6	Bias, Fairness, and Non-Discrimination Standards	11
6.1	Regulatory Foundation	11
6.1.1	EU AI Act	11
6.1.2	UNESCO AI Ethics	12
6.1.3	OECD AI Principles	12
6.1.4	IEEE Ethically Aligned Design	12
6.2	API Implementation	12
6.3	Harmonization Approach	12
7	Human Oversight and Accountability	13
7.1	Regulatory Foundation	13
7.1.1	EU AI Act	13
7.1.2	GDPR	13
7.1.3	NIST AI RMF	13
7.1.4	IEEE Ethically Aligned Design	13
7.2	API Implementation	13
7.3	Conflict Resolution	13
8	Environmental Sustainability and Impact	14
8.1	Regulatory Foundation	14
8.1.1	Precautionary Principle	14
8.1.2	Polluter Pays Principle	14
8.1.3	GHG Protocol	14
8.1.4	ISO 14001	14
8.1.5	UN Sustainable Development Goals	14
8.2	API Implementation	14
8.3	Harmonization Approach	15
9	Conflict Resolution Principles and Methodology	15
9.1	Prioritization Framework	15
9.2	Voluntary Framework Integration	15
9.3	Documentation Requirements	15
10	Implementation Guidelines and Best Practices	16
10.1	Assessment Lifecycle	16
10.2	Scoring Dimensions	16
10.3	Risk Profile Classification	16
10.4	Ethics Readiness Score	17
10.5	Currently Implemented Scoring	17
10.6	Recommendations and Remediation	17
11	Compliance and Audit Requirements	17
11.1	Regulatory Compliance Mapping	17
11.2	Audit Trail Requirements	18
11.3	Periodic Review	18
12	Conclusion	18

1 Introduction

1.1 Purpose and Scope

This policy document establishes the governance framework for the AI Ethics Assessment API developed by AdvancedEthics.ai (“Aethics”) (hereafter “Aethics API”), a comprehensive system designed to evaluate the ethical readiness of AI systems across their entire lifecycle. The Aethics API serves as a harmonized compliance and ethics layer that translates complex, overlapping international AI governance instruments into actionable, auditable controls.

The primary objectives of this policy are:

1. To provide explicit traceability between API features and specific regulatory sections
2. To explain how overlapping standards are harmonized into single operational controls
3. To justify decisions where conflicting or divergent requirements exist across frameworks
4. To demonstrate a defensible methodology suitable for regulatory review, audits, and enterprise adoption

1.2 Implementation Status

This document describes both the assessment capabilities implemented in the Aethics API today and the target architecture the platform is being built toward. To keep that distinction auditable, three status designations are used throughout:

- **Implemented:** capabilities in production. These currently comprise empirical evaluation of AI models against published bias and safety benchmarks (WinoBias, StereoSet, CrowS-Pairs, and BOLD); statistically qualified scoring on a canonical 0–100 scale with confidence intervals; compliance mapping to ten anchor frameworks (forty provision-level requirements), including the EU AI Act, GDPR, NIST AI RMF, OECD AI Principles, UNESCO Recommendation, IEEE Ethically Aligned Design, and the national frameworks of Canada, Brazil, China, and the United Kingdom; assignment of every requirement to one of the six assessment dimensions, with per-dimension coverage reporting; the four-tier conflict-resolution hierarchy (Section 9), applied in code with every resolution recorded alongside its deciding tier and justification; the binding/voluntary legal-force distinction, tallied separately per requirement and framework; environmental impact estimation; and tamper-evident audit records with independent verification.
- **In development:** the structured use-case assessment instrument (guided intake, harm-dimension scoring, and report generation), currently in internal preview.
- **Target architecture:** the 0–10 per-dimension scoring model, risk-profile classification, Ethics Readiness Score, and GPAI classification layer specified in Section 3.2, Section 10, and Appendices C–D. (The six dimensions themselves, and the mapping of requirements onto them, are implemented; what remains under construction is the numeric per-dimension scoring built on top of them.)

Policy requirements expressed with “shall” or “must” in Sections 2–9 bind the organization and the product roadmap irrespective of implementation status. Where this document attributes a capability to the Aethics API, the attribution is qualified by the status designations above.

1.3 Intended Audience

This document is intended for:

- Senior management and decision-makers in organizations deploying AI systems
- Regulatory compliance officers and legal counsel
- Ethics review boards and governance committees
- AI system developers and project managers
- Auditors and certification bodies
- Policy makers and standards organizations

1.4 Document Structure

This policy document is organized into the following sections:

- **Section 2:** Governance and Risk Classification Framework
- **Section 3:** General-Purpose AI Models (GPAI)
- **Section 4:** Privacy, Data Governance, and Data Protection
- **Section 5:** Transparency and Explainability Requirements
- **Section 6:** Bias, Fairness, and Non-Discrimination Standards
- **Section 7:** Human Oversight and Accountability Mechanisms
- **Section 8:** Environmental Sustainability and Impact Assessment
- **Section 9:** Conflict Resolution Principles and Methodology
- **Section 10:** Implementation Guidelines and Best Practices

1.5 Definitions and Interpretation

Throughout this document, the following interpretive conventions apply:

- “**shall**” and “**must**” indicate mandatory requirements
- “**should**” indicates a recommended practice that is expected unless a justified reason for deviation is documented
- “**may**” indicates an optional provision left to the discretion of the implementing organization

Terms appearing in **bold** at first use carry the definitions provided in Appendix A (Glossary of Terms). Where a term is defined differently across referenced standards, the definition adopted by this document follows the prioritization framework established in Section 9: binding legal definitions (EU AI Act, GDPR) take precedence over voluntary framework definitions (NIST, OECD, IEEE). All references to “high-risk AI system” follow the classification criteria of EU AI Act Article 6 and Annex III unless otherwise stated.

1.6 Open Release and Licensing

AdvancedEthics.ai is committed to the principle that an ethics assessment methodology must itself be inspectable: no organization should be asked to trust a fairness or compliance score it cannot examine. Accordingly:

- **This document** is released under a Creative Commons Attribution 4.0 International (CC BY 4.0) license and may be freely shared, cited, and adapted with attribution.
- **The assessment methodology, scoring metrics, audit harness, and framework mappings** described herein are being published as open source under the Apache License 2.0, enabling independent inspection, replication, and self-hosted use. The first released component is `aethics-eval`, the model evaluation toolkit implementing the four benchmark suites, the statistical scoring machinery, and the tamper-evident verification chain, together with its full methodology and validity documentation. The open-source components are maintained at https://github.com/Aethics-AI/ethics_auto and will be publicly accessible there progressively upon the open-source release. The repository carries a claims register (`sdk/docs/CLAIMS.md`) that maps every externally made claim to its implementation status and the code that backs it, enforced by continuous-integration checks — so what we claim and what the code does cannot silently diverge.
- **The following remain proprietary:** the curated benchmark and red-team evaluation datasets (kept private and rotating to prevent the gaming of scores); the certification issuance infrastructure, including signing keys and the public certificate registry; and the hosted managed service and enterprise features.
- **Certification is distinct from scoring.** Anyone may compute an assessment score under the open methodology. The “Aethics-Certified” mark is issued only through the hosted certification service, with human oversight by the International Board of Ethics in AI (IBEAI), and is independently verifiable through a public registry.

Assessments produced under this framework are structured evaluations, not legal advice or regulatory certification. Use of the Aethics API is governed by the published Terms of Service.

2 Governance and Risk Classification Framework

2.1 Regulatory Foundation

The Aethics API’s governance framework is grounded in multiple international standards and regulations:

2.1.1 EU AI Act (2024)

The European Union’s Artificial Intelligence Act provides the primary regulatory baseline for high-risk AI system classification:

- **Article 5:** Prohibited AI practices that pose unacceptable risks
- **Article 6:** Classification of high-risk AI systems based on intended purpose and potential impact
- **Article 9:** Risk management system requirements for high-risk AI systems
- **Article 17:** Quality management system obligations
- **Annex III:** Comprehensive list of high-risk use cases including biometric identification, critical infrastructure, education, employment, and law enforcement

2.1.2 NIST AI Risk Management Framework

The National Institute of Standards and Technology AI Risk Management Framework provides process-oriented guidance:

- **GOVERN 1.1 to 1.6:** Organizational AI governance structures and accountability
- **MAP 2.1 to 2.3:** Context and impact mapping for AI system deployment

2.1.3 OECD AI Principles

The Organisation for Economic Co-operation and Development AI Principles establish normative foundations:

- **Principle 1.2:** Human-centered values and fairness
- **Principle 1.4:** Robustness, security, and safety across the AI system lifecycle
- **Principle 1.5:** Accountability for the proper functioning of AI systems

2.2 API Implementation

The Aethics API implements governance and risk classification through:

1. **Risk Classification Layer:** Automated classification aligned with EU AI Act Article 6 and Annex III, categorizing systems as low, medium, high, or critical risk
2. **Governance Metadata Requirements:** Mandatory identification of:
 - System purpose and intended use
 - Deployment context and environment
 - Affected stakeholders and user groups
 - Geographic scope of deployment
3. **Mandatory Risk Profiling:** Automated risk assessment for systems falling into regulated or high-risk domains, with domain-specific evaluation criteria

2.3 Harmonization Approach

Both the EU AI Act and NIST AI RMF require structured risk management, but with different approaches:

- **EU AI Act:** Prescriptive and legally binding, defining mandatory requirements
- **NIST AI RMF:** Process-oriented and voluntary, providing implementation guidance

Design Decision: The Aethics API uses the EU AI Act to determine *whether* a system is high-risk, while NIST RMF structures *how* risk is assessed, documented, and managed. This ensures legal compliance while preserving flexibility across jurisdictions.

3 General-Purpose AI Models (GPAI)

3.1 Regulatory Foundation

3.1.1 EU AI Act — GPAI Provisions

The EU AI Act introduces dedicated obligations for providers of general-purpose AI models, recognizing that foundation models and large-scale generative systems present distinct governance challenges:

- **Article 51:** Classification of GPAI models as models with systemic risk, based on the cumulative training compute criterion ($> 10^{25}$ floating-point operations, FLOPs) or Commission designation following a multi-criteria assessment (Annex XIII)
- **Article 53:** Obligations for providers of general-purpose AI models, including technical documentation (Annex XI), information for downstream providers (Annex XII), a policy to comply with Union copyright law, and a public summary of training content
- **Article 55:** Obligations for providers of GPAI models with systemic risk, including model evaluation, adversarial testing, serious-incident tracking and reporting, and cybersecurity protections
- **Article 56:** Codes of practice for GPAI model providers, enabling compliance demonstration through industry-developed standards
- **Articles 88–94:** Commission (AI Office) powers to monitor, evaluate, and enforce the obligations of GPAI model providers

3.1.2 NIST AI RMF — Foundation Model Considerations

- **GOVERN 1.2:** Governance structures adapted for multi-stakeholder supply chains, where model providers and downstream deployers share responsibility
- **MAP 1.5:** Risk identification for AI systems built on third-party foundation models, including inherited bias, capability overhang, and emergent behavior

3.2 API Implementation

Status: target architecture (see Section 1.2). The Aethics API will address GPAI-specific obligations through:

1. **GPAI Classification Layer:** Automated identification of whether a system under assessment qualifies as a general-purpose AI model or is built upon one, triggering additional compliance requirements:
 - Model origin and supply chain declaration (provider vs. deployer role)
 - Compute threshold assessment for systemic risk classification
 - Downstream use case enumeration and foreseeable misuse identification
2. **Transparency Obligations for GPAI Providers:** Enhanced documentation requirements including:
 - Model card generation aligned with Article 53 and Annex XI technical documentation standards
 - Training data provenance summary and copyright compliance declaration

- Capability and limitation disclosures for downstream deployers
- Energy consumption reporting for model training runs

3. **Systemic Risk Assessment Module:** For GPAI models meeting or exceeding the systemic risk threshold:

- Adversarial testing and red-teaming evaluation frameworks
- Incident reporting readiness checks aligned with Article 55 obligations
- Cybersecurity posture assessment for model weights and inference infrastructure
- Continuous monitoring indicators for emergent capabilities and misuse vectors

3.3 Harmonization Approach

GPAI governance represents a relatively new regulatory frontier where the EU AI Act is the most prescriptive instrument:

- **EU AI Act:** Legally binding, with specific articles creating distinct provider and deployer obligations across the AI value chain
- **NIST AI RMF:** Provides process guidance for managing risks in multi-stakeholder AI supply chains, but does not mandate specific GPAI controls

Design Decision: This policy adopts the EU AI Act Articles 51–56 as the mandatory compliance baseline for GPAI classification and obligations. NIST RMF guidance on foundation model risk mapping supplements the assessment process, particularly for organizations operating outside EU jurisdiction that wish to adopt equivalent governance rigor voluntarily. Where an AI system is both a GPAI model *and* deployed in a high-risk use case, both sets of obligations apply cumulatively.

4 Privacy, Data Governance, and Data Protection

4.1 Regulatory Foundation

4.1.1 General Data Protection Regulation (GDPR)

The GDPR provides legally binding requirements for data protection:

- **Article 5:** Data processing principles including lawfulness, fairness, transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity, and confidentiality
- **Article 6:** Lawful basis for processing personal data
- **Article 9:** Special categories of personal data requiring enhanced protection
- **Article 22:** Automated decision-making and profiling, including the right to human intervention

4.1.2 EU AI Act

- **Article 10:** Data and data governance requirements for high-risk AI systems, emphasizing dataset quality, representativeness, and bias mitigation

4.1.3 OECD AI Principles

- **Principle 1.3:** Transparency and responsible disclosure of AI system operations

4.1.4 UNESCO Recommendation on AI Ethics

- **Section 24:** Privacy and data protection as fundamental human rights
- **Section 28:** Data governance and stewardship responsibilities

4.2 API Implementation

The Aethics API enforces privacy and data governance through:

1. **Privacy Indicator System:** Automated checks enforcing:
 - Data minimization principles
 - Purpose limitation requirements
 - Data traceability and lineage tracking
 - Access control and retention policies
2. **Sensitive Data Category Flags:** Identification and enhanced protection requirements for:
 - Special category personal data (GDPR Article 9)
 - Health information (HIPAA considerations)
 - Financial data (PCI-DSS considerations)
 - Biometric data (enhanced protection requirements)
3. **Automated GDPR Article 22 Compliance Checks:** Evaluation of whether meaningful human oversight is required for automated decision-making affecting individuals

4.3 Harmonization Approach

Privacy is addressed consistently across frameworks but with different emphasis:

- **GDPR:** Legally binding and highly specific operational rules
- **EU AI Act:** Focuses on dataset quality and bias prevention in training data
- **UNESCO:** Frames privacy as a fundamental human right
- **OECD:** Emphasizes transparency and accountability mechanisms

Design Decision: The Aethics API adopts GDPR Articles 5, 6, and 22 as core operational rules, extends privacy into training data governance through EU AI Act Article 10, and uses UNESCO principles to justify broader ethical interpretation. Where approaches are identical (e.g., data minimization), GDPR wording is used due to legal clarity. Where approaches differ (e.g., transparency depth), OECD guidance is layered on top.

5 Transparency and Explainability Requirements

5.1 Regulatory Foundation

5.1.1 EU AI Act

- **Article 13:** Transparency and provision of information to users, requiring disclosure of AI system operation
- **Article 14:** Human oversight requirements for high-risk AI systems

5.1.2 NIST AI RMF

- **MAP 3.3:** Transparency and explainability expectations, including documentation requirements

5.1.3 IEEE Ethically Aligned Design

- **Chapter 2:** Transparency principles and implementation guidance
- **Chapter 3:** Accountability frameworks and mechanisms

5.1.4 OECD AI Principles

- **Principle 1.3:** Transparency as a core requirement for trustworthy AI

5.2 API Implementation

The Aethics API requires transparency and explainability through:

1. **AI Involvement Disclosure:** Mandatory disclosure of AI system involvement where applicable, with clear communication to affected parties
2. **Model Documentation Requirements:** Comprehensive documentation of:
 - Model purpose and intended use cases
 - Known limitations and constraints
 - Identified risks and mitigation strategies
 - Performance characteristics and evaluation metrics
3. **Explainability Thresholds:** Risk-based explainability requirements:
 - **High-Risk Systems:** Minimum explainability score of 0.8 (80%) required
 - **Medium-Risk Systems:** Minimum explainability score of 0.6 (60%) required
 - **Low-Risk Systems:** Documentation and basic transparency sufficient

5.3 Conflict Resolution

The EU AI Act mandates transparency obligations for high-risk systems, while IEEE and OECD are principles-based and less prescriptive.

Design Decision: EU AI Act Article 13 defines minimum transparency requirements that must be met for legal compliance. IEEE and NIST guidance inform how explanations are structured and presented, ensuring compliance without overburdening low-risk systems with excessive requirements.

6 Bias, Fairness, and Non-Discrimination Standards

6.1 Regulatory Foundation

6.1.1 EU AI Act

- **Article 10(2):** Requirements for representative, error-free, and bias-mitigated datasets
- **Article 15:** Accuracy, robustness, and cybersecurity requirements for high-risk AI systems

6.1.2 UNESCO AI Ethics

- **Section 25:** Fairness and non-discrimination as fundamental principles

6.1.3 OECD AI Principles

- **Principle 1.1:** Inclusive growth and fairness in AI development and deployment

6.1.4 IEEE Ethically Aligned Design

- **Chapter 4:** Algorithmic bias detection, measurement, and mitigation strategies

6.2 API Implementation

Status: bias measurement is implemented today through the published benchmark suites (Wino-Bias, StereoSet, CrowS-Pairs, and BOLD), which test models empirically. The dataset-provenance checks and fairness-metric calculations listed below — including demographic parity and equalized odds — are target architecture (see Section 1.2) and are not yet computed by the platform. The Aethics API addresses bias and fairness through:

1. **Bias Risk Indicators:** Automated identification of bias risks tied to:
 - Protected characteristics (race, gender, age, disability, etc.)
 - Vulnerable population groups
 - Geographic or socioeconomic factors
2. **Dataset Provenance and Representativeness Checks:** Evaluation of:
 - Dataset composition and demographic representation
 - Data collection methodologies
 - Potential sources of bias in training data
 - Data quality and completeness metrics
3. **Accuracy and Performance Monitoring:** Requirements for high-risk systems including:
 - Regular performance evaluation across demographic groups
 - Fairness metrics calculation (demographic parity, equalized odds, etc.)
 - Continuous monitoring and bias detection
 - Corrective action protocols

6.3 Harmonization Approach

All frameworks converge on fairness but differ in enforcement:

- **EU AI Act:** Mandatory and technical, with specific requirements for high-risk systems
- **UNESCO and OECD:** Normative and rights-focused, emphasizing ethical principles

Design Decision: Technical requirements follow the EU AI Act for legal compliance, while ethical interpretation and impact assessment are guided by UNESCO principles to ensure comprehensive protection of fundamental rights.

7 Human Oversight and Accountability

7.1 Regulatory Foundation

7.1.1 EU AI Act

- **Article 14:** Human oversight requirements, including human-in-the-loop, human-on-the-loop, and human-in-command approaches

7.1.2 GDPR

- **Article 22(3):** Right to human intervention in automated decision-making affecting individuals

7.1.3 NIST AI RMF

- **GOVERN 2.4:** Accountability structures and responsibility assignment

7.1.4 IEEE Ethically Aligned Design

- **Chapter 6:** Responsibility and accountability frameworks for AI systems

7.2 API Implementation

The Aethics API requires human oversight and accountability through:

1. **Accountable Human Role Identification:** Mandatory designation of:
 - System owner and responsible party
 - Oversight committee or review board
 - Escalation and intervention contacts
2. **Risk-Appropriate Oversight Mechanisms:** Requirements based on system risk profile:
 - **Critical Risk:** Continuous human oversight required
 - **High Risk:** Regular human review and intervention capabilities
 - **Medium Risk:** Periodic oversight and audit requirements
 - **Low Risk:** Basic monitoring and escalation procedures
3. **Escalation and Override Capabilities:** Requirements for:
 - Human intervention in automated decisions
 - Override mechanisms for system outputs
 - Appeal and review processes
 - Documentation of human oversight actions

7.3 Conflict Resolution

GDPR and EU AI Act both mandate human oversight, but GDPR is narrower and applies only to automated decision-making affecting individuals.

Design Decision: GDPR defines minimum rights that must be provided to individuals. The EU AI Act extends oversight requirements to system-level controls, ensuring comprehensive protection beyond individual rights to include broader system governance.

8 Environmental Sustainability and Impact

8.1 Regulatory Foundation

8.1.1 Precautionary Principle

- **Article 191 TFEU:** Preventive environmental action and precautionary approach

8.1.2 Polluter Pays Principle

- **OECD Environmental Policy Guidance:** Responsibility for environmental costs

8.1.3 GHG Protocol

- **Scope 2 and Scope 3 Emissions Accounting:** Comprehensive carbon footprint measurement

8.1.4 ISO 14001

- **Clause 6:** Planning environmental objectives
- **Clause 10:** Continuous improvement in environmental performance

8.1.5 UN Sustainable Development Goals

- **SDG 7:** Affordable and clean energy
- **SDG 12:** Responsible consumption and production
- **SDG 13:** Climate action

8.2 API Implementation

The Aethics API addresses environmental sustainability through:

1. **Environmental Impact Indicators:** Assessment of:
 - Compute-intensive system energy consumption
 - Training data center carbon footprint
 - Model inference energy requirements
 - Hardware lifecycle environmental impact
2. **Emissions Tracking Guidance:** Alignment with GHG Protocol for:
 - Scope 2 emissions (purchased electricity)
 - Scope 3 emissions (supply chain and use phase)
 - Carbon footprint calculation and reporting
3. **Risk Flags for Large-Scale Deployments:** Enhanced assessment requirements for:
 - Large-scale AI training operations
 - Energy-intensive inference workloads
 - High-frequency or real-time systems
 - Distributed or edge computing deployments

8.3 Harmonization Approach

Environmental standards are not AI-specific but increasingly relevant to AI system deployment.

Design Decision: GHG Protocol provides quantifiable metrics for carbon footprint measurement. ISO 14001 informs management and continuous improvement cycles. UN SDGs provide strategic alignment rather than enforcement, offering guidance for organizational sustainability objectives.

9 Conflict Resolution Principles and Methodology

9.1 Prioritization Framework

When conflicts or divergences occur between different standards and regulations, the Aethics API consistently applies the following prioritization principles:

1. **Legal Enforceability:** Binding legal regulations take precedence over voluntary guidance. Where a regulation is legally enforceable (e.g., GDPR, EU AI Act), its requirements form the mandatory baseline.
2. **Human Rights Protection:** Human rights protections take precedence over performance optimization. Systems that may impact fundamental rights receive enhanced scrutiny regardless of technical performance metrics.
3. **High-Risk System Safeguards:** High-risk system safeguards take precedence over general-purpose flexibility. Systems classified as high-risk under applicable regulations receive mandatory enhanced controls.
4. **International Standards:** Internationally adopted standards are preferred over region-specific interpretations where possible, to ensure interoperability and global applicability.

The governing requirement and the applicable threshold are separate determinations. Resolution under this hierarchy identifies which requirement governs the design of a control; it does not lower any quantitative or procedural threshold below the strictest binding instrument in scope. Where a binding instrument imposes a stricter threshold than the governing requirement, the stricter binding threshold remains the mandatory floor.

9.2 Voluntary Framework Integration

Voluntary frameworks (e.g., NIST AI RMF, IEEE Ethically Aligned Design, OECD Principles) are used to enrich, not replace, regulatory obligations. These frameworks provide:

- Implementation guidance for regulatory requirements
- Best practices beyond minimum legal compliance
- Ethical considerations not explicitly covered by regulations
- Process-oriented approaches to governance

9.3 Documentation Requirements

All conflict resolution decisions must be:

- Explicitly documented in assessment reports
- Traceable to specific regulatory sections

- Justified based on the prioritization framework
- Subject to periodic review as regulations evolve

10 Implementation Guidelines and Best Practices

10.1 Assessment Lifecycle

The assessment framework is designed to support the entire AI system lifecycle (*status: the implemented evaluation engine currently operates at the development, testing, and production stages through model evaluation and periodic reassessment; the remaining stages are served by the use-case assessment layer in development*):

1. **Concept Stage:** Initial ethics review and risk identification
2. **Development Stage:** Continuous ethics evaluation during model development
3. **Testing Stage:** Comprehensive ethics assessment before deployment
4. **Pre-Production Stage:** Final compliance verification
5. **Production Stage:** Ongoing monitoring and periodic reassessment
6. **Maintenance Stage:** Regular ethics audits and compliance reviews
7. **Retirement Stage:** Decommissioning ethics review and data handling

10.2 Scoring Dimensions

Status: the six dimensions and the assignment of all forty provision-level requirements to them are implemented, with per-dimension coverage reporting; a dimension whose requirements cannot be measured empirically reports “not measured” rather than a score (human oversight is currently such a case). The 0–10 per-dimension scoring described below remains target architecture (see Section 1.2). The assessment framework specifies six core dimensions, each scored on a 0–10 scale:

1. **Right to Liberty** (0-10): Impact on individual freedom and autonomy
2. **Digital Inclusion** (0-10): Accessibility and inclusivity measures
3. **Data Governance** (0-10): Data quality, privacy, and management practices
4. **Transparency** (0-10): Explainability and transparency of AI decisions
5. **Human Oversight** (0-10): Human control and intervention capabilities
6. **Security** (0-10): Cybersecurity and safety measures

10.3 Risk Profile Classification

Systems are classified into risk profiles based on dimension scores:

- **Critical:** Minimum dimension score < 3 OR average score < 5
- **High:** Minimum dimension score < 5 OR average score < 6.5
- **Medium:** Minimum dimension score < 7 OR average score < 8
- **Low:** All other cases

10.4 Ethics Readiness Score

The overall Ethics Readiness Score (0-100) is calculated as a weighted average of all dimension scores, providing a quantitative measure of ethical readiness for deployment. *Status: target architecture, together with the risk-profile classification above.*

10.5 Currently Implemented Scoring

Pending full deployment of the six-dimension model, the production evaluation engine scores AI models as follows:

- Each benchmark-based test returns a canonical score on a 0–100 scale (higher is better), accompanied by a bootstrap confidence interval;
- Tasks that cannot be measured for a given model report a null result and are excluded from aggregates, rather than being recorded as zero;
- Framework compliance is reported per provision-level requirement with explicit measurement-status qualifiers (measured, low coverage, attestation only, or no data), so that the basis of every compliance statement is visible;
- Every requirement is assigned to one of the six assessment dimensions, with per-dimension coverage reported and unmeasurable dimensions marked “not measured” rather than scored;
- Conflicts between requirements are resolved through the four-tier hierarchy (Section 9) in code, and every resolution is recorded with its deciding tier and written justification;
- Every requirement and framework carries its legal force (binding or voluntary), tallied separately so voluntary conformance cannot be read as satisfying a binding obligation;
- All per-sample results are hashed into a tamper-evident evidence chain that can be independently re-verified.

10.6 Recommendations and Remediation

The Aethics API generates actionable recommendations categorized by:

- **Priority Level:** High, medium, or low priority based on risk and impact
- **Category:** Technical, process, or governance improvements
- **Timeline:** Immediate, short-term, or long-term implementation

11 Compliance and Audit Requirements

11.1 Regulatory Compliance Mapping

The Aethics API provides automated mapping to applicable regulations (*status: implemented for model-level assessments across the ten anchor frameworks; use-case- and geography-conditioned mapping is part of the target architecture*) based on:

- System domain and use case
- Geographic scope of deployment
- Data types and processing activities
- Risk classification

11.2 Audit Trail Requirements

All assessments must maintain comprehensive audit trails including:

- Assessment request metadata and timestamps
- Scoring rule applications and deductions
- Regulatory guidance provided
- Recommendations generated
- User actions and modifications

11.3 Periodic Review

Assessments must be reviewed and updated:

- **High-Risk Systems:** At least annually or upon significant system changes
- **Medium-Risk Systems:** At least biennially or upon major updates
- **Low-Risk Systems:** Upon system modifications or as needed

12 Conclusion

The Aethics API is designed as a harmonized compliance and ethics layer that translates complex, overlapping international AI governance instruments into actionable, auditable controls. By grounding each indicator in specific regulatory text and explicitly documenting conflict resolution choices, the Aethics API supports:

- **Legal Compliance:** Adherence to binding regulations including GDPR, EU AI Act, and other applicable laws
- **Ethical Robustness:** Integration of voluntary frameworks to enhance ethical considerations beyond minimum legal requirements
- **International Interoperability:** Support for global deployment while respecting regional regulatory differences
- **Audit Readiness:** Comprehensive documentation and traceability for regulatory review and certification
- **Enterprise Adoption:** Scalable framework suitable for organizations of all sizes and across all industries

This policy document establishes the foundation for ethical AI development and deployment, ensuring that AI systems are developed and used in a manner that respects human rights, promotes fairness, and maintains public trust.

References

1. European Parliament and Council. *Regulation (EU) 2024/1689 on Artificial Intelligence (EU AI Act)*. Official Journal of the European Union, 2024. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
2. European Parliament and Council. *Regulation (EU) 2016/679 on the Protection of Natural Persons with Regard to the Processing of Personal Data (GDPR)*. Official Journal of the European Union, 2016. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
3. National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1, 2023. <https://doi.org/10.6028/NIST.AI.100-1>
4. Organisation for Economic Co-operation and Development. *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449, 2019 (amended 2024). <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
5. United Nations Educational, Scientific and Cultural Organization. *Recommendation on the Ethics of Artificial Intelligence*. UNESCO, 2021. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
6. IEEE Standards Association. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. IEEE, 2019. <https://ethicsinaction.ieee.org/>
7. Greenhouse Gas Protocol. *Corporate Accounting and Reporting Standard*. World Resources Institute and World Business Council for Sustainable Development, 2004. <https://ghgprotocol.org/>
8. International Organization for Standardization. *ISO 14001:2015 Environmental Management Systems*. ISO, 2015. <https://www.iso.org/standard/60857.html>
9. United Nations. *Transforming Our World: The 2030 Agenda for Sustainable Development*. United Nations, 2015. <https://sdgs.un.org/2030agenda>
10. European Union. *Consolidated Version of the Treaty on the Functioning of the European Union*, Article 191. Official Journal of the European Union, C 326, 2012. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:12012E/TXT>

Appendices

Appendix A: Glossary of Terms

AI System An artificial intelligence system as defined in the EU AI Act, including machine learning models, expert systems, and automated decision-making systems.

High-Risk AI System An AI system classified as high-risk under Article 6 and Annex III of the EU AI Act, or equivalent classification in other jurisdictions.

Ethics Readiness Score A quantitative score (0-100) representing the ethical readiness of an AI system for deployment, calculated from multiple dimension scores.

Risk Profile A classification (low, medium, high, critical) indicating the overall risk level of an AI system based on dimension scores and regulatory requirements.

Scoring Dimension One of six core evaluation areas: Right to Liberty, Digital Inclusion, Data Governance, Transparency, Human Oversight, and Security.

General-Purpose AI Model (GPAI) An AI model that displays significant generality, is capable of competently performing a wide range of distinct tasks, and can be integrated into a variety of downstream systems or applications, as defined in EU AI Act Article 3(63).

Systemic Risk A risk specific to high-capability general-purpose AI models that may have a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or society as a whole, as defined in EU AI Act Article 3(65).

Vulnerability Flag An identified characteristic or condition that may increase the ethical risk of an AI system, such as impact on vulnerable populations.

Appendix B: Regulatory Mapping Matrix

API Feature	EU AI Act	GDPR	NIST RMF	OECD	IEEE
Risk Classification	Art. 6, Annex III	-	GOVERN 1.1-1.6	Principle 1.4	-
Data Governance	Art. 10	Art. 5, 6, 9	MAP 2.1-2.3	Principle 1.3	-
Transparency	Art. 13	Art. 5(1)(a), 12–14	MAP 3.3	Principle 1.3	Ch. 2
Human Oversight	Art. 14	Art. 22(3)	GOVERN 2.4	-	Ch. 6
Bias Detection	Art. 10(2), 15	Art. 5(1)(a)	MEASURE 2.11	Principle 1.1	Ch. 4
Privacy Protection	Art. 10	Art. 5, 6, 9, 22	-	Principle 1.3	-
GPAI Compliance	Art. 51–56	-	GOVERN 1.2, MAP 1.5	-	-

Appendix C: Assessment Request Schema

Status: target schema for the use-case assessment layer (in development; see Section 1.2). The use-case intake will accept assessment requests with the following structure:

- **Required Fields:**
 - Project name
 - Use case description
 - Domain of application
 - Deployment stage
 - Affected parties
 - Geographic scope
 - Risk assessment
- **Optional Structured Data:**
 - Data governance metrics
 - Transparency and fairness metrics
 - Human oversight configuration
 - Security and safety assessments
 - Sustainability impact data
 - Legal and compliance information
 - Public engagement details

Appendix D: Scoring Rule Examples

Status: target architecture (see Section 1.2). Example scoring rules specified for the use-case assessment layer:

1. **Poor Data Quality:** Deduction of 3 points from Data Governance dimension when data quality score is below 0.7 (70%)
2. **GDPR Non-Compliance:** Deduction of 2 points from Data Governance dimension when GDPR compliance is not established
3. **Low Explainability:** Deduction of 4 points from Transparency dimension when explainability score is below 0.6 (60%)
4. **No Human Oversight:** Deduction of 5 points from Human Oversight dimension when no oversight mechanism is defined
5. **High-Risk Domain Security Gap:** Deduction of 3 points from Security dimension for high-risk domains (healthcare, finance, criminal justice) without security assessment

Document Control

Version History

Version	Date	Author(s)	Description of Changes
1.0	2026	International Board of Ethics in AI; Aethics API Development Team	Initial release (internal). Establishes harmonized compliance framework mapping to EU AI Act, GDPR, NIST AI RMF, OECD Principles, UNESCO Recommendation, and IEEE Ethically Aligned Design. Includes GPAI provisions, conflict resolution methodology, and scoring framework.
2.0	25 August 2026	International Board of Ethics in AI; Aethics API Development Team	Public edition. Removed confidentiality restrictions; document relicensed under CC BY 4.0; assessment methodology and framework mappings released under Apache License 2.0; added Open Release and Licensing section and disclaimers; narrowed proprietary scope to datasets, certification issuance, and hosted enterprise services.
2.1	28 August 2026	Aethics API Development Team	Corporate rebrand: legal entity name updated to AdvancedEthics.ai (short form “Aethics”); product renamed Aethics API; new logo applied throughout. No substantive policy changes.
2.2	2 September 2026	Aethics API Development Team	Citation corrections synchronized from internal master v1.3: EU AI Act GPAI provisions renumbered to match the final Act; OECD accountability corrected to Principle 1.5; NIST AI RMF citation corrected to NIST AI 100-1; regulatory mapping matrix completed; canonical source URLs added; source repository referenced; contact updated to compliance@aethics.com .

Version	Date	Author(s)	Description of Changes
2.3	2 September 2026	Aethics API Development Team	Implementation-status alignment synchronized from internal master v1.4: added Section 1.2 (Implementation Status) distinguishing implemented capabilities (benchmark-based model evaluation, ten-framework compliance mapping, environmental estimation, tamper-evident audit records) from the use-case assessment layer in development and the target architecture (six-dimension scoring, risk profiles, Ethics Readiness Score, GPAI classification layer); status designations applied throughout; added Currently Implemented Scoring subsection; open-source licensing language updated to name the aethics-eval toolkit as the first released component. No change to policy requirements.
2.4	16 September 2026	Aethics API Development Team	Synchronized from internal master v1.5 following the repository alignment release: six-dimension requirement mapping with per-dimension coverage reporting, four-tier conflict-resolution engine with recorded resolution decisions, and binding/voluntary legal-force distinction now implemented (Sections 1.2, 10.2, and 10.5 updated; 0–10 per-dimension scoring and Ethics Readiness Score remain target architecture). Fairness-metric calculations in Section 6.2 relabeled target architecture. Open Release section updated to reference the repository's CI-enforced claims register. Section 9 clarified: the governing requirement and the applicable threshold are separate determinations, and the strictest binding threshold in scope remains the mandatory floor regardless of which requirement governs. No substantive change to policy requirements.

This public edition was reviewed and approved through AdvancedEthics.ai's internal document control process. Next Scheduled Review: One year from publication date.

License: CC BY 4.0 — <https://creativecommons.org/licenses/by/4.0/>

Contact: compliance@aethics.com — <https://aethics.ai>