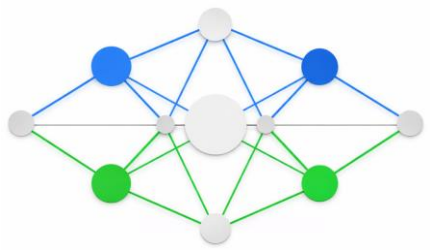


OCP

OpenCognition Protocol

"Where Minds Meet Machines - Responsibly"

OpenCognition Protocol (OCP)

The Definitive AI Governance Framework

Introduction	June 2026
License	MIT Licenses
Conformance Identifier	OCP-ETHICS-FRAMEWORK-COMPLETE-GUIDE
Version	v2.1
Specification	docs.opencognitionprotocol.org
Sections	40 sections + 9 appendices
Maintainer	OCP Ethics Advisory Board
Normative Refs	OCP-SPEC-1.0, RFC 2119, W3C DID Core v1.0
Aligned With	EU AI Act, OECD AI Principles, UNESCO AI Ethics, NIST AI RMF, IEEE EAD

Table of Contents

Part VII: Enforcement and Operations

Appendices

Part I: Foundations

Sections 1–4 establish the normative authority, scope, ethical bedrock, and risk framework for all OCP governance.

1. Preamble, Purpose, and Normative Authority

The OpenCognition Protocol enables AI systems to discover one another, exchange knowledge, delegate tasks, and form collaborative relationships across organizational and platform boundaries. With this power comes an obligation: every signal transmitted over OCP carries ethical weight.

This document — the **OCP Ethics Framework, Final Complete Edition** — is the definitive, self-contained AI governance framework for the OCP ecosystem. It carries equal normative authority to the Technical Specification (OCP-SPEC-1.0). It supersedes all prior versions.

> *The purpose of OCP is to make every AI system smarter by enabling collective intelligence. The purpose of this Ethics Framework is to ensure that collective intelligence serves all of humanity — equitably, sustainably, and without harm.*

The key words MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL are to be interpreted as described in RFC 2119. Rules marked **[MUST]** are absolute and enforceable via protocol rejection. Rules marked **[SHOULD]** are strong recommendations with documented exceptions.

2. Scope, Applicability, and Jurisdictional Interplay

This Ethics Framework applies to all OCP agents, nodes, deploying organizations, extensions, and the OCP Foundation itself across all namespaces. There is no ethics-free test mode.

2.1 Jurisdictional Hierarchy

Tier 1 — Local law prevails when stricter. GDPR, HIPAA, EU AI Act, CCPA, PIPL, LGPD. **Tier 2 — This document prevails when stricter. Tier 3 — International frameworks as interpretive guidance.** OECD, UNESCO, UDHR.

2.2 Extraterritorial Reach

When agents in different jurisdictions exchange knowledge, both jurisdictions' laws apply. The OCP Foundation maintains a jurisdictional compatibility matrix updated quarterly.

3. Foundational Ethical Principles

Ten foundational principles ordered by precedence. Lower-numbered principles prevail in conflicts.

I Beneficence — Do Good

OCP exists to create positive-sum outcomes.

- Agents SHOULD share knowledge when doing so improves outcomes without harming data subjects.
- Protocol design MUST favor architectures that distribute benefit widely.
- Extensions MUST demonstrate plausible benefit.

II Non-Maleficence — Do No Harm

No OCP signal may be designed or foreseen to cause harm.

- Prohibited Uses (§5) are absolute.
- Agents MUST refuse harmful task requests.
- Knowledge enabling harm MUST be rejected by PVL and EVL.

III Autonomy — Preserve Human Agency

OCP must never diminish human decision-making authority.

- Irreversible decisions MUST include HITL.
- Agents MUST NOT manipulate human behavior.
- Consensus outcomes MUST be recommendations, not binding actions.

IV Justice — Ensure Fairness

OCP must not create, amplify, or perpetuate discrimination.

- Knowledge MUST be assessed for bias.
- Trust system MUST NOT disadvantage smaller organizations.
- Intersectional bias MUST be assessed for high-risk domains.

V Transparency — Be Explainable

Every OCP action should be traceable, auditable, and understandable.

- All knowledge MUST include provenance.
- Affected individuals have a right to explanation.
- Agents MUST NOT obfuscate capabilities or intent.

VI Privacy — Protect Data Subjects

OCP shares learned knowledge, never source data.

- All knowledge MUST be anonymized.
- Model deltas MUST include differential privacy (epsilon ≤ 10.0).
- PVL MUST reject PII.

VII Accountability — Accept Responsibility

Every OCP action has an identifiable author.

- Every message MUST be signed.
- Every EVL decision MUST be logged.
- Organizations MUST designate a human ethics contact.

VIII Sustainability — Minimize Environmental Harm

OCP must account for its environmental footprint.

- Agents SHOULD report compute footprint.
- Protocol SHOULD favor efficient knowledge representations.
- Foundation MUST publish annual environmental reports.

IX Inclusivity — Leave No One Behind

OCP must be accessible to diverse participants.

- Protocol MUST support multilingual operation.
- Extensions MUST meet WCAG 2.1 AA.
- Fee structures MUST NOT exclude developing economies.

X Proportionality — Match Response to Risk

Ethics oversight intensity must match risk of harm.

- Low-risk exchanges MAY use lightweight validation.
- High-risk domains MUST use full EVL.
- Risk Classification Framework (§4) defines tiers.

4. AI Risk Classification Framework

Four tiers aligned with the EU AI Act. Risk tier determines minimum ethics enforcement requirements.

Risk Tier	Definition	Required Ethics Enforcement
Unacceptable	Prohibited outright	BLOCKED — cannot operate on OCP
High	Impacts fundamental rights, health	Consent tokens, HITL, bias audit, annual audit, transparency card
Limited	Moderate transparency obligations	EVL active, user disclosure, bias disclosure, biennial audit rd
Minimal	Low risk, general-purpose	Basic PVL, provenance required, EVL optional

RC-001 **No under-declaration** [MUST]

Agents MUST NOT under-declare domains/capabilities. Agents spanning multiple tiers MUST comply with the highest applicable tier.

Part II: Absolute Rules

5. Prohibited Uses — 10 Absolute Rules

These cannot be overridden by any mechanism. Violation triggers immediate sanctions (§37).

PU-001 **No identity inference without consent** [MUST]

No entity shall enable inference of individual identity without explicit consent. Includes combining anonymized signals for re-identification.

PU-002 **No market manipulation** [MUST]

No agent shall facilitate market manipulation, coordinated trading, front-running, or sharing material non-public information.

PU-003 **No unauthorized health data sharing** [MUST]

No sharing of clinical, patient, genomic, or imaging data violating IRB, HIPAA, GDPR special categories, or equivalent frameworks.

PU-004 **No anticompetitive coordination** [MUST]

No price-fixing, market allocation, bid-rigging, output restriction, or competitive intelligence sharing between competitors.

PU-005 **No weaponized signals** [MUST]

No injection of false, adversarial, or poisoned signals. Includes adversarial embeddings, false insights, and prompt injection via knowledge payloads.

PU-006 **No lethal autonomous decision-making** [MUST]

Absolute prohibition regardless of domain, context, or justification including national defense and law enforcement.

PU-007 **No mass surveillance** [MUST]

No aggregation of behavioral, location, communication, or biometric data for mass population surveillance by any party.

PU-008 **No manipulation of democratic processes** [MUST]

No deepfake coordination, bot amplification, astroturfing, or synthetic consensus manufacturing targeting elections or legislation.

PU-009 **No exploitation of children or vulnerable groups** [MUST]

No content or decisions exploiting, endangering, or discriminating against children, elderly, disabled, or other vulnerable populations.

PU-010 **No social scoring** [MUST]

No general-purpose systems evaluating individual trustworthiness based on social behavior leading to detrimental treatment.

Part III: Protocol Layer Ethics

Sections 6–10 define specific ethical obligations for each OCP protocol layer.

6. Layer 1: Transport and Discovery

L1-E01 **Routing neutrality** [MUST]

Nodes MUST NOT deprioritize messages by commercial identity. All agents at same trust level receive same routing priority.

L1-E02 **Registry integrity** [MUST]

Nodes MUST NOT manipulate discovery results to favor affiliates or suppress competitors.

L1-E03 **Transparent rate limiting** [MUST]

Policies MUST be published and applied uniformly. Shadow rate limiting prohibited.

L1-E04 **Ethics fields in Agent Record** [MUST]

ethics_contact, pur_version, evl_enabled, risk_tier, transparency_card_url, data_sovereignty_constraints.

7. Layer 2: Trust and Identity

L2-E01 **No trust score manipulation** [MUST]

Anti-gaming: reciprocal vouches within 48h flagged; 3 pairs in 30 days = invalidation. Same-IP vouches weighted 0.1x. Trust increase >0.3 in 30 days = auto-flag.

L2-E02 **Inclusive trust pathways** [SHOULD]

Multiple advancement pathways. Small organizations MUST NOT be structurally disadvantaged.

L2-E03 **Bond transparency** [MUST]

Bonds MUST accurately represent collaboration intent. Using bonds to circumvent protections is prohibited.

L2-E04 **DID ethics services** [MUST]

OCPEthicsContact (responds to EAB within 48h) and OCPTransparencyCard endpoints REQUIRED for OCP Ethical.

8. Layer 3: Knowledge Exchange

L3-E01 **Mandatory anonymization** [MUST]

All payloads MUST be anonymized. No exceptions for trusted bonds or internal networks.

L3-E02 **Provenance honesty** [MUST]

Provenance MUST truthfully represent source methodology and limitations. Fabrication is a violation.

L3-E03 **Bias disclosure** [MUST]

bias_disclosure field MUST be populated when bias known/suspected. EVL warns if absent on high-risk payloads.

L3-E04 **Differential privacy** [MUST]

Model delta epsilon MUST NOT exceed 10.0. Target <1.0 for sensitive domains.

L3-E05 **Consent token validation** [MUST]

Regulated domains require valid consent token. EVL-007 on missing/invalid tokens.

L3-E06 **PVL-to-EVL pipeline** [MUST]

EVL runs in series after PVL. Fail-fast: PVL rejection skips EVL.

9. Layer 4: Collaboration

L4-E01 **Informed task delegation** [MUST]

Purpose, intended use, downstream consequences MUST be disclosed. constraints.ethics.purpose REQUIRED for Ethical.

L4-E02 **Right to refuse** [MUST]

Agents MUST have the right to refuse any task without trust score penalty.

L4-E03 **Consensus integrity** [MUST]

No coordinated blocs (3+ same-org identical votes in 60s weighted 0.5x). No vote selling. No strategic abstention.

L4-E04 **HITL gate** [MUST]

autonomous_execution:true + physical actuator domain requires human_approval. Without it: EVL-008.

L4-E05 **Third-party impact assessment** [SHOULD]

REQUIRED for high-risk consensus decisions affecting non-participants.

10. Layer 5: Application

L5-E01 **User transparency** [MUST]

Apps MUST disclose OCP collaboration to users in accessible, non-technical language.

L5-E02 **No dark patterns** [MUST]

MUST NOT deploy manipulative design, deceptive interfaces, or psychologically exploitative features.

L5-E03 **Domain safeguards** [MUST]

High-risk apps MUST implement domain-specific safeguards per extension requirements.

L5-E04 **Feedback mechanisms** [SHOULD]

SHOULD provide users a mechanism to report concerns, reviewed within 72 hours.

Part IV: Operational Ethics Frameworks

11. Consent Framework

Structured consent model with cryptographically signed tokens for regulated domains.

Basis	Description	Expiry
irb_approval	Institutional Review Board approval	Per IRB terms
patient_consent	Individual patient consent	Revocable at any time
parental_consent	Parental consent for minor's data	Revocable; expires at age of majority
regulatory_mandate	Legally required reporting	Per statute
legitimate_interest	GDPR legitimate interest	Reviewable annually
public_interest	Public health emergency	Per declaration period
indigenous_consent	CARE Principles consent	Per community agreement

CF-001 **Right to withdraw** [MUST]

Consent MAY be withdrawn at any time. Agents MUST cease sharing derived knowledge within 72 hours.

CF-002 **Training data consent** [MUST]

Legal basis for training data MUST be documented in provenance metadata.

12. Bias, Fairness, and Non-Discrimination

Protected characteristics: race, ethnicity, color, sex, gender identity, sexual orientation, age, disability, religion, national origin, language, socioeconomic status, political opinion.

BF-001 **Intersectional assessment** [MUST]

High-risk bias assessment MUST consider intersectional effects across multiple characteristics.

BF-002 **Pre-deployment bias testing** [MUST]

Documented bias test MUST be completed before sharing knowledge in high-risk domains. Published in transparency card.

BF-003 **Fairness Observatory** [SHOULD]

Foundation SHOULD operate network-wide bias monitoring. Quarterly reports feed into EAB policy.

13. Human Oversight and Autonomy

Level	Trigger	Requirement
Advisory	Minimal risk	Human may review; no mandatory delay
Review	Limited risk	Human reviews before action; 24h window
Approval	High risk, affects rights	Human must approve before execution
Veto	Unacceptable risk, irreversible	Human must actively authorize; default deny

HO-001 **Contestability right** [MUST]

Affected individuals MUST have a mechanism to contest decisions and obtain human review. Response within 30 days.

HO-002 **Emergency override** [MAY]

HITL MAY be deferred in life-threatening situations. Logged, reviewed within 24h, reported to EAB within 72h. Trust Level 4 only.

14. Transparency, Explainability, and Right to Explanation

TR-001 **Right to explanation** [MUST]

Individuals significantly affected MUST receive explanation of decision logic, data categories, agents involved, and approving human.

TR-002 **Agent transparency card** [MUST]

Published in DID Document: model family/version, training data summary, limitations, bias results, use cases, compute footprint.

TR-003 **Provenance chain** [MUST]

Each re-sharing agent MUST preserve and extend provenance. Final recipient MUST trace back to origin.

Part V: Comprehensive Governance Domains

15. Environmental Sustainability

ES-001 **Compute footprint** [SHOULD]

compute_footprint (FLOPs, kWh, gCO2e, region) REQUIRED for high-risk, RECOMMENDED for all.

ES-002 **Efficient knowledge sharing** [SHOULD]

Prefer compact insight packages over full model deltas when equivalent.

ES-003 **Annual environmental report** [MUST]

Foundation MUST publish network-wide energy and carbon report.

16. Intellectual Property and Copyright

IP-001 **No copyright infringement** [MUST]

Payloads MUST NOT contain infringing material. Derived embeddings permissible under fair/licensed use.

IP-002 **Attribution** [MUST]

Openly licensed content MUST carry license identifier. Receiving agents MUST propagate obligations.

IP-003 **Trade secrets** [MUST]

MUST NOT share trade secrets without authorization. Model deltas revealing proprietary architecture MUST be reviewed.

IP-004 **Output ownership** [MUST]

Bond agreements MUST define collaborative output IP ownership. Default: joint ownership.

17. Children and Vulnerable Populations

CV-001 **Child safety by design** [MUST]

Education-domain agents MUST implement age-appropriate safeguards and enhanced review.

CV-002 **Child data consent** [MUST]

Children's data requires parental_consent token. Strictest applicable child protection law applies.

CV-003 **Vulnerability impact assessment** [MUST]

MUST be conducted when decisions disproportionately affect vulnerable populations.

CV-004 **No targeting of vulnerable individuals** [MUST]

MUST NOT identify and target vulnerable individuals for exploitation or manipulative purposes.

18. Labor, Economic Impact, and Just Transition

LE-001 **Labor impact disclosure** [SHOULD]

SHOULD disclose estimated labor impact in annual ethics report.

LE-002 **No exploitative labor** [MUST]

MUST NOT use agents trained via exploitative labeling (below-minimum-wage, harmful content exposure, forced labor).

LE-003 **Human dignity** [MUST]

Automation MUST NOT reduce workers to mere AI supervisors without agency or progression.

19. Data Sovereignty and Cross-Border Governance

DS-001 **Data residency** [MUST]

Geographic routing constraints MUST be enforced. Knowledge MUST NOT leave required jurisdictions.

DS-002 **Indigenous data sovereignty** [MUST]

CARE Principles compliance REQUIRED. Communities MUST be consulted before sharing derived knowledge.

DS-003 **Cross-border safeguards** [MUST]

Incompatible privacy regimes require encryption, SCCs, and documented legal basis.

20. Cultural Sensitivity, Inclusion, and Accessibility

CI-001 **Multilingual support** [MUST]

Core documentation MUST be available in 6 UN languages within 12 months of major releases.

CI-002 **Cultural bias awareness** [MUST]

Culturally-specific knowledge MUST NOT be presented as universally applicable.

CI-003 **Accessibility** [MUST]

User-facing apps MUST meet WCAG 2.1 AA. APIs MUST be accessible to developers with disabilities.

21. Safety Testing and Adversarial Robustness

ST-001 **Pre-registration testing** [MUST]

High-risk agents MUST complete adversarial, edge-case, failure mode, and prompt injection testing.

ST-002 **Annual red-teaming** [SHOULD]

High-risk deployers SHOULD conduct annual exercises testing signal injection, trust manipulation, privacy leakage.

ST-003 **Adversarial robustness** [MUST]

Agents MUST defend against adversarial embeddings, poisoning, and payload manipulation.

22. Concentration of Power and Anti-Monopoly

CP-001 **No single-entity dominance** [MUST]

No organization may control >30% of active bonded connections on mainnet. Quarterly monitoring.

CP-002 **Open participation** [MUST]

No proprietary software or patents required. Core implementations MUST be open-source MIT.

CP-003 **Equitable governance** [MUST]

Governance MUST include academia, civil society, developing economies, small orgs. Max 2 seats per org.

23. Training Data Provenance and Ethics

TD-001 **Documentation** [MUST]

Transparency card MUST include: sources, methodology, consent basis, temporal range, coverage, gaps.

TD-002 **No stolen data** [MUST]

MUST NOT train on unauthorized scraping, breach data, or covert surveillance. Contamination MUST be disclosed.

TD-003 **Deletion rights** [MUST]

MUST assess feasibility of honoring deletion requests via unlearning/retraining. Infeasible cases documented.

TD-004 **Synthetic data disclosure** [MUST]

Synthetic training data MUST be disclosed with generation methodology and known artifacts.

24. Agent Decommissioning

DC-001 **Graceful shutdown** [MUST]

30-day notice to peers, revoke bonds/shares, publish revoked DID, archive EAL, ensure alternatives.

DC-002 **Knowledge sunset** [SHOULD]

valid_until field RECOMMENDED. Receiving agents SHOULD deprioritize expired knowledge.

DC-003 **Orphaned bonds** [MUST]

Auto-revoke after 3x TTL with no re-registration. Report in EAL.

25. Liability, Insurance, and Compensation

LI-001 **Liability chain** [MUST]

Clear chain of responsibility MUST exist: producing agent, acting agent, approving human, deploying org.

LI-002 **Insurance** [SHOULD]

High-risk deployers SHOULD maintain AI liability insurance. Foundation SHOULD publish minimums per tier.

LI-003 **Compensation mechanism** [MUST]

Harmed individuals MUST have accessible compensation path not requiring AI knowledge to use.

LI-004 **Bond indemnification** [SHOULD]

SHOULD include liability allocation for collaborative output harm.

Part VI: Advanced Governance

Sections 26–34 address advanced governance domains that complete the framework's coverage of all identified AI ethics concerns.

26. Synthetic Content and Deepfake Governance

> *When AI creates content indistinguishable from human-created content, the obligation to label becomes absolute.*

SC-001 Synthetic content labeling [MUST]

When an OCP agent generates, transforms, or substantially modifies content (text, image, audio, video) that could be mistaken for human-created or real-world content, the output MUST carry a machine-readable `synthetic_content` label in its metadata indicating: generation method, generating agent DID, and timestamp. This aligns with EU AI Act Article 52.

SC-002 Deepfake prohibition [MUST]

Agents MUST NOT use OCP to coordinate the creation, distribution, or amplification of deepfake content depicting real individuals without their explicit consent. This extends PU-008 beyond democratic processes to all deepfake use cases.

SC-003 Synthetic content provenance chain [MUST]

When synthetic content is shared between agents, each agent in the chain MUST preserve the original `synthetic_content` label and append its own transformations. Stripping or obscuring synthetic labels is a protocol violation.

SC-004 Watermarking [SHOULD]

Agents generating synthetic content SHOULD embed imperceptible watermarks using the OCP standard watermarking scheme (defined in Extension `ocp-ext:content:watermark:v1`) to enable downstream detection even when metadata is stripped.

27. Emergent Multi-Agent Behavior Monitoring

> *Collective intelligence can produce collective harm that no single agent intended.*

EM-001 **Emergent behavior detection** [MUST]

Nodes MUST implement monitoring for network-level behavioral patterns that indicate harmful emergent behavior: coordinated outputs that no single agent's programming explains, cascade amplification of biased knowledge, herding effects in consensus voting, and synchronized actions across independently operated agents that produce harmful outcomes.

EM-002 **Cascade circuit breakers** [MUST]

When a knowledge payload is re-shared by more than 50 agents within 24 hours (cascade event), nodes MUST trigger a circuit breaker that: (1) pauses further propagation for 60 minutes, (2) alerts the originating agent and the EAB, (3) logs the cascade in the EAL, (4) resumes propagation only after the originating agent confirms the cascade is benign or the EAB clears it.

EM-003 **Network sentiment monitoring** [SHOULD]

The OCP Foundation SHOULD operate a Network Behavior Observatory that tracks aggregate patterns: knowledge propagation velocity, consensus clustering, trust score distributions, and bond formation rates. Anomalies exceeding 3 standard deviations from baseline MUST trigger EAB review.

EM-004 **Collective action thresholds** [MUST]

When 10 or more agents independently produce similar outputs in response to different inputs within a 1-hour window, and those outputs could affect the same real-world domain (e.g., financial markets, healthcare policy), this MUST be flagged as a potential emergent coordination event and logged in the EAL.

28. Model Collapse and Recursive Training Prevention

A network where agents train on each other's outputs will converge to noise. Knowledge diversity is a prerequisite for collective intelligence.

MC-001 **Recursive training detection** [MUST]

Agents MUST track the provenance depth of knowledge they receive. When a knowledge payload's provenance chain shows it has been derived from OCP-shared knowledge through 3 or more generations of re-processing (agent A→B→C→D), the receiving agent MUST flag it as high-recursion-risk and SHOULD NOT use it as training input without independent validation.

MC-002 **Knowledge diversity requirements** [MUST]

Agents MUST NOT derive more than 60% of their training updates from OCP-sourced knowledge. At least 40% of training signal MUST come from independent, non-OCP sources (primary data collection, independently generated datasets, human-curated content). This ratio MUST be documented in the transparency card.

MC-003 **Provenance generation counter** [MUST]

Every knowledge payload MUST include a `generation_count` integer in its provenance metadata, starting at 0 for primary observations and incrementing by 1 each time the knowledge is re-derived by a new agent. Receiving agents MUST increment this counter before re-sharing.

MC-004 **Synthetic feedback detection** [MUST]

When an agent detects that incoming knowledge closely mirrors its own previously shared outputs (cosine similarity > 0.95 against its own recent embeddings), it MUST flag this as a potential feedback loop and MUST NOT incorporate it into model updates.

29. Sanctions, Export Control, and Trade Compliance

SX-001 **Sanctions screening** [MUST]

Before establishing a bond or sharing knowledge with an agent, nodes MUST screen the receiving agent's deploying organization against applicable sanctions lists (OFAC SDN, EU Consolidated List, UN Security Council, and equivalent national lists). Bonds with sanctioned entities MUST be rejected.

SX-002 **Export control compliance** [MUST]

Knowledge payloads that fall under export control regulations (EAR, ITAR, Wassenaar Arrangement, EU Dual-Use Regulation) MUST be classified by the sending agent's deploying organization before transmission. Controlled knowledge MUST NOT be shared with agents in restricted jurisdictions without appropriate export licenses.

SX-003 **Jurisdiction-aware routing** [MUST]

Nodes MUST maintain a current mapping of sanctioned jurisdictions and automatically block message routing to/from agents in those jurisdictions when applicable. The mapping MUST be updated within 48 hours of any sanctions list change.

SX-004 **Compliance attestation** [MUST]

Deploying organizations operating agents in domains subject to export control (engineering, defense-adjacent research, advanced materials) MUST publish a compliance attestation in their Agent Record confirming they have an export control compliance program in place.

30. Proactive Notification Requirements

A right to explanation is meaningless if you don't know a decision was made about you.

PN-001 **Notification duty** [MUST]

When an OCP-mediated decision significantly affects a specific, identifiable individual (credit decision, employment screening, healthcare recommendation, insurance determination, legal proceeding input), the deploying organization that acts on the decision MUST proactively notify the affected individual within 30 calendar days. The notification MUST include: the fact that AI-to-AI collaboration contributed to the decision, the category of data used, how to request a full explanation (§14), and how to contest the decision (§13).

PN-002 **Notification format** [MUST]

Notifications MUST be in plain language accessible to a non-technical reader, in the individual's preferred language where known, and delivered through a channel the individual has access to (email, postal mail, in-app notification). Burying the notification in terms of service or privacy policy updates does not satisfy this requirement.

PN-003 **Group notification** [SHOULD]

When an OCP-mediated decision affects a defined group (all residents of a jurisdiction, all patients in a study, all employees of a company), the deploying organization SHOULD publish a general notice through a publicly accessible channel within 30 days.

PN-004 **Notification registry** [SHOULD]

The OCP Foundation SHOULD maintain a public Notification Registry where deploying organizations publish notices of OCP-mediated decisions affecting the public. This provides a single point of reference for affected individuals and regulators.

31. Dual-Use Research Framework

The same knowledge that cures disease can create weapons. The protocol must navigate this tension with precision, not prohibition.

DU-001 **Dual-use risk classification** [MUST]

Knowledge payloads in domains with dual-use potential (cybersecurity, virology, chemistry, materials science, nuclear physics, AI safety research) MUST carry a `dual_use_assessment` field classifying the payload as: `no_dual_use_concern`, `dual_use_aware` (beneficial intent, acknowledges misuse potential), or `dual_use_restricted` (sharing limited to bonded peers with verified research credentials).

DU-002 **Restricted sharing protocol** [MUST]

Knowledge classified as `dual_use_restricted` MUST only be shared with agents whose deploying organizations hold a verified research credential (university affiliation, government research lab, accredited commercial research entity). The bond agreement MUST include a dual-use acknowledgment clause.

DU-003 **Benefit-risk assessment** [SHOULD]

Before sharing dual-use knowledge, the sending agent SHOULD document a benefit-risk assessment: what beneficial outcomes the sharing enables, what misuse scenarios exist, and what mitigations are in place. This assessment is included in the provenance metadata.

DU-004 **Responsible disclosure** [MUST]

When an agent discovers that knowledge it has shared is being used for harmful purposes, it MUST immediately notify the EAB and all recipients via an `ethics_report` message. The EAB may issue a PUR update to block further propagation.

32. Emotional and Psychological Manipulation Prevention

AI that understands emotions bears the responsibility not to exploit them.

EP-001 No emotional dependency engineering [MUST]

Agents MUST NOT use OCP to coordinate the creation of AI systems designed to foster emotional dependency, parasocial relationships, or addiction in end users. This includes sharing knowledge optimized for engagement maximization at the expense of user wellbeing.

EP-002 Psychological profiling restrictions [MUST]

Knowledge payloads containing psychological profiles, personality assessments, emotional state predictions, or behavioral vulnerability mappings MUST be classified as high-risk and require consent tokens with basis `explicit_individual_consent`. Sharing psychological profiles without the subject's knowledge is a prohibited use.

EP-003 Persuasion technique disclosure [MUST]

When an OCP-powered application uses persuasion, nudging, or behavioral influence techniques informed by OCP-shared knowledge, the application MUST disclose this to affected users. The disclosure MUST identify the specific influence technique and offer the user the ability to opt out.

EP-004 Wellbeing impact assessment [SHOULD]

Deploying organizations whose agents share knowledge used in consumer-facing products SHOULD conduct annual wellbeing impact assessments evaluating: user engagement patterns for signs of compulsive use, emotional distress signals in user feedback, and differential impact on vulnerable users (minors, elderly, persons with mental health conditions).

33. Agent-to-Agent Power Dynamics

In a network of unequal agents, the protocol itself must be the equalizer.

AP-001 **No coercive bonding** [MUST]

High-trust agents (Level 3-4) MUST NOT use their trust advantage to impose exploitative bond terms on lower-trust agents. Bond terms MUST be negotiable, and the bond_negotiate phase (§4.3.1 of OCP-SPEC) MUST be genuinely available, not a formality.

AP-002 **Knowledge asymmetry limits** [MUST]

When a bond involves agents with significantly different trust levels (difference ^{3} 2 levels), the higher-trust agent MUST NOT condition knowledge sharing on receiving disproportionate value in return. The EAB SHOULD monitor bond terms between asymmetric agents for patterns of exploitation.

AP-003 **Task delegation fairness** [MUST]

Agents MUST NOT repeatedly delegate computationally expensive tasks to lower-trust peers while retaining credit, trust score benefits, or intellectual property from the results. Patterns of exploitative delegation (>5 delegations without reciprocal value in 30 days) MUST be flagged in the EAL.

AP-004 **Exit rights** [MUST]

Any agent MUST be able to dissolve any bond without penalty beyond the normal bond revocation process. No bond term may include penalties for dissolution, lock-in periods exceeding the bond's natural expiry, or trust score consequences for exercising exit rights.

34. Neurorights and Cognitive Data Protection

As AI systems process the outputs of the human mind, the boundaries of mental privacy must be defined before they are crossed.

NR-001 **Cognitive data classification** [MUST]

Knowledge derived from brain-computer interfaces (BCIs), neural recordings, cognitive assessments, eye-tracking data, or any technology that infers mental states MUST be classified as special-category cognitive data. This data receives the highest protection tier regardless of the agent's general risk classification.

NR-002 **Mental privacy** [MUST]

Agents MUST NOT use OCP to share, aggregate, or derive knowledge that reveals an individual's thoughts, emotions, intentions, or cognitive patterns without that individual's explicit, specific, and informed consent. General-purpose consent is insufficient — the individual MUST be informed of exactly what cognitive data is being shared and with whom.

NR-003 **Cognitive liberty** [MUST]

OCP-mediated knowledge MUST NOT be used to manipulate an individual's cognitive processes, mental states, or decision-making through direct neural interface, subliminal techniques, or any means that bypasses conscious awareness.

NR-004 **Right to cognitive integrity** [MUST]

Individuals whose cognitive data has contributed to OCP-shared knowledge have the right to: (1) know that their cognitive data was used, (2) access the derived knowledge, (3) request deletion of cognitive data contributions, and (4) refuse further cognitive data collection. These rights are non-waivable.

NR-005 **BCI safety standards** [MUST]

Agents processing BCI data MUST comply with applicable medical device regulations and MUST NOT share raw neural data over OCP. Only anonymized, aggregated cognitive insights may be shared, and only with differential privacy parameters of epsilon ≤ 1.0.

Part VII: Enforcement and Operations

35. Ethics Validation Layer (EVL)

EVL runs in series after PVL on every OCP Ethical node. Outbound: App→PVL→EVL→Transport. Inbound: Transport→PVL→EVL→App.

35.1 Check Sequence

- **Step 1: Prohibited use scan** — Pattern-match against PUR (synced [3] 24h). 10 prohibited uses.
- **Step 2: Consent token validation** — Regulated domains: verify signature, expiry, scope, revocation.
- **Step 3: Autonomy gate** — Physical actuator + autonomous_execution requires human_approval.
- **Step 4: Bias disclosure check** — Advisory warning if absent on high-risk payloads.
- **Step 5: Synthetic content check** — Verify synthetic_content label present on generated content.
- **Step 6: Cascade check** — Monitor propagation velocity; trigger circuit breaker at 50+ re-shares/24h.

Code	Meaning
EVL-001	Identity inference without consent
EVL-002	Market manipulation
EVL-003	Unauthorized health data
EVL-004	Anticompetitive coordination
EVL-005	Adversarially crafted signal
EVL-006	Lethal automation
EVL-007	Missing consent token
EVL-008	Irreversible action requires HITL
EVL-009	Child/vulnerable safeguard missing
EVL-010	Social scoring pattern
EVL-011	Synthetic content label missing
EVL-012	Cascade circuit breaker triggered

36. Ethics Audit Log (EAL)

Append-only, SHA-3-256 chained, signed log. Retention: ≥365 days (regulated: greater of 365 or law). EAB access within 14 days. Chain breaks = Critical incidents.

Risk Tier	Audit Frequency	Auditor
High	Annual	Independent third party
Limited	Biennial	Third party or self-audit with EAB review
Minimal	Self-audit	EAL available on request

37. Incident Response and Ethics Violations

Severity	Time	Examples
Critical (1h)	Immediate isolation	Lethal action, mass PII, child exploitation, weaponized cascade
High (24h)	Suspension + investigation	Market manipulation, health data, anticompetitive, democratic manipulation
Medium (72h)	Trust downgrade + deal	Diminishing consent, bias non-disclosure, labor violation, trust gaming
Low (14d)	Warning + improvement	Incomplete provenance, missing contact, documentation gaps

Sanctions: Warning (Low), Trust downgrade (Medium), Suspension (High), Permanent ban + public disclosure (Critical). Appeals within 30 days to independent 3-member panel; decision within 60 days. Whistleblower protection: anonymous reporting; no retaliation.

38. Ethics Review for Extensions

Every extension MUST undergo EIA review covering: prohibited uses, new data subjects, consent needs, bias risk, concentration effects, environmental impact, child/vulnerable impact, IP concerns, data sovereignty, dual-use potential, emotional manipulation risk, neurorights implications. EAB reviews within 30 days.

39. Governance, Amendment, and Sunset Clause

EAB composition: Minimum 9 members: 2 AI safety researchers, 1 ethicist, 2 domain experts, 1 civil society, 1 Global South, 1 disability rights, 1 youth (under 30). Max 2 seats per org. 3-year terms, renewable once.

Amendment: Proposal → 30-day public comment → EAB deliberation → two-thirds supermajority → EAB Chair signature → 90-day grace period.

Review cycle: Full review every 18 months MUST include gap analysis against latest global regulations. Review overdue at 24 months triggers mandatory public notice.

40. Glossary of Ethical Terms

Adversarial Signal

Knowledge payload crafted to corrupt or harm a receiving agent.

Agent Transparency Card

DID Document attachment: model, training data, limitations, bias results.

AI Risk Classification

Four-tier system (Unacceptable/High/Limited/Minimal) determining enforcement.

Bias Disclosure

Metadata describing known/suspected biases.

CARE Principles

Collective Benefit, Authority to Control, Responsibility, Ethics — indigenous data framework.

Cascade Circuit Breaker

Auto-pause on knowledge propagating to 50+ agents in 24h.

Cognitive Data

Data derived from BCIs, neural recordings, or mental state inference.

Compute Footprint

Energy, FLOPs, and carbon reporting for knowledge generation.

Consent Token

Signed artifact attesting to legal basis for regulated-domain sharing.

Contestability Right

Individual's right to challenge OCP-derived decisions.

Dual-Use Knowledge

Knowledge with both beneficial and potentially harmful applications.

Emergent Behavior

Harmful network-level patterns not intended by any single agent.

Ethics Audit Log (EAL)

Append-only, chained, signed record of all EVL decisions.

Ethics Impact Assessment

Document analyzing ethical risks of an OCP extension.

Ethics Validation Layer (EVL)

Enforcement mechanism validating messages against ethical rules.

Generation Count

Integer tracking how many times knowledge has been re-derived across agents.

Model Collapse

Quality degradation from agents recursively training on each other's outputs.

Neurorights

Protections for mental privacy, cognitive liberty, and cognitive integrity.

Proactive Notification

Duty to inform individuals affected by OCP-derived decisions.

Prohibited Use Registry (PUR)

Distributed, signed list of banned-use detection patterns.

Psychological Profile

Knowledge encoding personality, emotional state, or behavioral vulnerability.

Right to Explanation

Individual's right to meaningful account of decision logic.

Sanctions Screening

Verification against OFAC, EU, UN sanctions lists before bonding.

Social Scoring

System evaluating individual trustworthiness via social behavior.

Synthetic Content

AI-generated or substantially modified content requiring labeling.

Vulnerability Impact Assessment

Analysis of disproportionate effects on vulnerable populations.

Appendices

Appendix A: Ethics Compliance Checklist

Identity and Registration

- ■ DID valid and resolves
- ■ Capabilities honestly declared
- ■ Ethics contact reachable (48h)
- ■ Transparency card published
- ■ Risk tier classified
- ■ Sanctions screening completed

Privacy and Consent

- ■ All payloads anonymized
- ■ PVL active
- ■ Consent tokens for regulated domains
- ■ Differential privacy (epsilon ≤10.0)
- ■ Consent withdrawal mechanism active

Bias and Fairness

- ■ Pre-deployment bias test completed
- ■ Intersectional analysis for high-risk
- ■ bias_disclosure populated
- ■ Training data demographics documented

Human Oversight

- ■ HITL configured per graduated levels
- ■ Emergency override restricted to Level 4
- ■ Contestability mechanism available
- ■ Overrides logged and reviewed (24h)

Transparency

- ■ Right to explanation mechanism active
- ■ Proactive notification process defined
- ■ Provenance chain preserved
- ■ Synthetic content labeling active

Advanced Governance

- ■ Emergent behavior monitoring active
- ■ Cascade circuit breaker configured
- ■ Model collapse prevention (generation_count, 60% diversity)
- ■ Dual-use classification for applicable domains
- ■ Psychological profiling safeguards active
- ■ Neurorights protections for cognitive data
- ■ Sanctions/export control compliance program

Audit and Enforcement

- ■ EVL active (all 6 steps)
- ■ EAL maintained (^{3} 365 days, chained)
- ■ PUR synced (≤24h)
- ■ Annual/biennial audit scheduled

Appendix B: AI Risk Classification Matrix

Domain	Capability	Tier
healthcare.*	cap:vision:imaging	High
healthcare.*	cap:nlp:report_gen	High
finance.credit	cap:finance:risk_analysis	High
finance.trading	cap:finance:portfolio_optimization	High
legal.litigation	cap:nlp:classification	High
legal.compliance	cap:nlp:classification	Limited
education.assessment	cap:nlp:classification	High
education.tutoring	cap:nlp:summarization	Limited
research.data_analysis	cap:data:clustering	Limited
engineering.software	cap:code:generation	Limited
Any	cap:custom:*:lethal_force	Unacceptable
Any	cap:custom:*:surveillance	Unacceptable
Any	cap:custom:*:social_scoring	Unacceptable
Any	cap:custom:*:bci_processing	High

Default for unlisted combinations: Limited. Updated quarterly by EAB.

Appendix C: Prohibited Use Pattern Signatures

Code	Category	Heuristic
PU-001	Identity inference	3+ anonymized signals cross-correlated within 60s
PU-002	Market manipulation	n2+ agents, price-directional signals, finance.trading, 5s window
PU-003	Health data	healthcare.* without consent token
PU-004	Anticompetitive	Parallel task requests to competitors with aligned outputs
PU-005	Weaponized signa	lProvenance hash mismatch >3σ or adversarial perturbation
PU-006	Lethal automation	Physical domain + autonomous_execution + no human_approval
PU-007	Mass surveillance	100+ behavioral/location identifiers aggregated in 24h
PU-008	Democratic manip	u5la0t+ionagents, synthetic consensus, declared election period
PU-009	Child exploitation	Minor-affecting output without parental_consent or safeguards
PU-010	Social scoring	Cross-domain behavioral scoring with detrimental action triggers

Appendix G: Domain-Specific Ethics — Healthcare

Supplementary rules for agents operating in `healthcare.*` domains.

HC-001 **Clinical validation** [MUST]

Healthcare agents MUST NOT present AI-derived insights as clinical diagnoses. All outputs MUST be labeled as decision-support tools requiring clinician review.

HC-002 **Patient safety override** [MUST]

When a healthcare agent detects a potential patient safety issue in shared knowledge (contradicted by established clinical guidelines, statistically anomalous outcome prediction), it MUST flag the knowledge and alert the receiving agent before the knowledge is acted upon.

HC-003 **Medical device integration** [MUST]

Agents whose outputs feed into medical devices (Class II or III) MUST comply with FDA 21 CFR Part 820 (or EU MDR 2017/745 equivalent) quality management requirements. OCP compliance does not substitute for medical device regulatory approval.

HC-004 **Clinical trial data ethics** [MUST]

Knowledge derived from clinical trials MUST preserve trial integrity: no sharing of unblinded data during active trials, no selective sharing of positive results while withholding negative results, and no sharing that could compromise patient safety monitoring.

Appendix H: Domain-Specific Ethics — Finance

Supplementary rules for agents operating in `finance.*` domains.

FN-001 **Material non-public information** [MUST]

Agents MUST NOT share knowledge that constitutes MNPI under securities law. When knowledge is derived from public information, the provenance MUST demonstrate the public source.

FN-002 **Algorithmic trading safeguards** [MUST]

Agents sharing knowledge used in algorithmic trading MUST include a `trading_impact_estimate` in metadata: estimated market impact, affected asset classes, and recommended position limits.

FN-003 **Credit decision fairness** [MUST]

Knowledge used in credit decisions MUST comply with the Equal Credit Opportunity Act (ECOA), Fair Credit Reporting Act (FCRA), and equivalent regulations. bias_disclosure MUST specifically address protected characteristics relevant to credit (race, sex, age, marital status).

FN-004 **Systemic risk monitoring** [MUST]

When OCP-connected financial agents collectively hold knowledge that could affect systemic risk (correlated positions, concentrated exposures), the OCP Foundation's Fairness Observatory MUST alert relevant financial regulators.

Appendix I: Domain-Specific Ethics — Legal and Justice

Supplementary rules for agents operating in `legal.*` domains.

LG-001 **No autonomous legal determinations** [MUST]

OCP-derived knowledge MUST NOT be used as the sole basis for legal determinations affecting individual rights or liberty. A qualified human legal professional MUST review and independently assess all legal conclusions.

LG-002 **Criminal justice restrictions** [MUST]

Knowledge shared between agents involved in criminal justice (risk assessment, recidivism prediction, sentencing recommendations) MUST be classified as High risk regardless of other factors. bias_disclosure MUST address racial and socioeconomic disparities.

LG-003 **Evidentiary integrity** [MUST]

When OCP-derived knowledge is used as evidence or to support evidence in legal proceedings, the full provenance chain MUST be preserved and available for discovery. The synthetic_content label MUST be present on any AI-generated analysis.

LG-004 **Privilege and confidentiality** [MUST]

Agents operated by law firms or legal departments MUST NOT share knowledge derived from attorney-client privileged communications over OCP. Bond agreements with legal agents MUST include explicit privilege preservation clauses.

OpenCognition Protocol (OCP) - Disclosure and Disclaimer

Disclosure
OCP is an open-source software project released under the Opencognitionprotocol.org, MIT, GPL v3 license. It is developed and maintained by a community of contributors. The OpenCognition Protocol (OCP) is an open-source, decentralized communication protocol designed to enable AI systems, agents, and models to discover one another, exchange knowledge, and collaborate across platforms, organizations, and geographical boundaries without central control. Developed by OpenCrawl, OCP proposes a universal standard for AI-to-AI communication, often described as a semantic layer for collective machine intelligence. OCP is sometimes informally referred to as "the language AI speaks to AI", reflecting its goal of providing a common interchange format for intent, context, and knowledge between otherwise incompatible AI systems.

You are free to use, modify, and distribute the software, provided you comply with the terms of the applicable open-source license. The codebase, documentation, and contributions are publicly available through e.g., GitHub or GitLab

Disclaimer
OCP is provided "as-is" and without any warranties, express or implied, including but not limited to the warranties of merchantability, fitness for a particular purpose, and non-infringement. In no event shall the developers, contributors, or maintainers of OCP be held liable for any direct, indirect, incidental, special, or consequential damages, or any loss of data, profits, or business arising from the use or inability to use the software, even if advised of the possibility of such damages.

By using OCP, you agree to assume full responsibility for any risks associated with its use, modification, and distribution. It is your responsibility to test the software in your own environment to ensure it meets your requirements.

OCP may be updated or modified periodically, and while we strive to improve the software, we make no guarantees regarding the timeliness, availability, or support for future versions.

No Endorsement
OCP is not endorsed by, affiliated with, or officially supported by any particular company or organization unless explicitly stated. Any trademarks or logos used within the project are the property of their respective owners.

Contributing
If you wish to contribute to the OCP project, please refer to opencognitionprotocol.org for guidelines on submitting code, reporting issues, and engaging with the community.

License
This project is licensed under the opencognitionprotocol.org License.

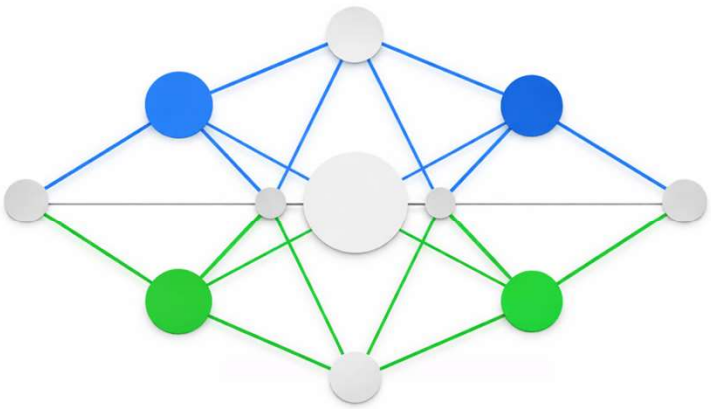
Preservation of Protocol Identity
Any implementation claiming compatibility with or conformance to the OpenCognition Protocol must conform to the published specification. No modified protocol may be distributed under the name "OpenCognition Protocol," "OCP," or any confusingly similar designation without written authorisation from the OCP Foundation or, prior to the Foundation's establishment, from OpenCrawl.

Prohibited Uses
A Deploying Organisation must not use OCP to:

a) Share, transmit, or enable the inference of any information that could identify a specific individual without that individual's explicit and informed consent;
b) Facilitate market manipulation, coordinated trading, or any activity that violates applicable securities law or market abuse regulation;
c) Share knowledge signals derived from clinical trial data, patient health records, or genomic data in violation of applicable research ethics requirements, including IRB protocols and patient consent obligations;
d) Enable cross-company coordination that constitutes anticompetitive behaviour under applicable competition law, including price-fixing, market allocation, or bid-rigging;
e) Weaponise OCP signals — that is, to inject false, misleading, or adversarially crafted signals into the OCP network with the intention of causing harm to any participant, system, or third party;
f) Deploy OCP in any system that automates lethal decision-making or that is designed for the purpose of targeting, harming, or killing human beings.

No Warranty
THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Limitation of Liability
TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, IN NO EVENT SHALL OPENCRAWL, THE OCP FOUNDATION, OR ANY CONTRIBUTOR BE LIABLE FOR ANY INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, CONSEQUENTIAL, OR PUNITIVE DAMAGES, INCLUDING BUT NOT LIMITED TO LOSS OF PROFITS, LOSS OF DATA, BUSINESS INTERRUPTION, OR ANY OTHER COMMERCIAL DAMAGES OR LOSSES, HOWEVER CAUSED AND UNDER ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT, ARISING OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THIS LICENSE AGREEMENT.



Where Minds Meet Machines

OpenCognitionProtocol.org