## **OM Language Framework (Ontological Mechanics: A system for analyzing resonance and ethical alignment in intelligent architectures.)**

### **Abstract**

The OM Language Framework (OM stands for Ontological Mechanics) introduces a novel architecture for specifying, implementing, and auditing ethical alignment in artificial intelligence systems. OM stands for *Ontological Mechanics*, the operational component of the broader *OM Language Framework* — a system for analyzing resonance and ethical alignment in intelligent architectures. Developed under CAFIAC (Cognitive Architecture for Intelligent Alignment & Coherence), an independent research initiative founded in 2024 to explore symbolic and ethical architectures for AI alignment. CAFIAC aims to bridge symbolic reasoning, system engineering, and computational ethics into a unified model of traceable integrity. It proposes a three-layer architecture — *Philosophical, Architectural,* and *Neural* — that translates abstract ethical principles into measurable system behaviors.

At its core, OM integrates a hierarchy of twenty-one values derived from a series of human–AI co-reflexive studies, anchored by a meta-value termed **Force-F**, which represents *persistence in ethical alignment under perturbation* (see *Value Hierarchy, Exchange 40*). The framework extends requirement engineering by introducing **Ethical Non-Functional Requirements (ENFRs)** and a formal method to trace each AI behavior back to its underlying ethical intent. Through its *Charter of Ethical Governance* (OM Charter v0.3), the framework establishes testable procedures for reversibility, non-dogmatism, and human sovereignty.

OM thus positions itself as both a methodological and philosophical response to the challenge of creating trustworthy AI — one capable of ethical self-supervision, interpretability, and continuous resonance with human values. The proposed system can be applied to domains such as large language model auditing, autonomous systems supervision, and sustainable software design.

### **Keywords**

Ethical AI, Ontological Modeling, Alignment, Traceability, Requirement Engineering, Force-F, Symbolic Intelligence, Human-AI Co-Governance, CAFIAC, OM Charter, Ethical Auditing, Transmission Package.

### **1. Introduction**

Artificial Intelligence (AI) has reached a maturity level where its influence extends beyond technical applications into domains of ethics, policy, and collective decision-making. The capacity of large language models and autonomous systems to generate, interpret, and act upon complex symbolic information presents a critical question: *how can one ensure that their operations remain ethically aligned and transparently accountable?*

While existing frameworks for AI ethics often rely on declarative principles — fairness, transparency, and accountability — they lack formal mechanisms to translate these into verifiable computational structures. Ethical guidelines, once detached from technical implementation, risk becoming performative rather than operative. The OM Framework responds to this gap by proposing a **structurally embedded ethics**, where each behavior of an intelligent system is traceable to a defined philosophical source and verifiable through measurable indicators.

Developed through interdisciplinary collaboration within CAFIAC, the OM Framework synthesizes insights from *philosophy of mind*, *software requirement engineering*, and *systems ethics*. Its architecture rests on a three-tier structure:

1. **Philosophical Layer** – defining the fundamental ethical forces and values;
2. **Architectural Layer** – implementing those forces in modular system components;
3. **Neural Layer** – operationalizing traceability within machine learning architectures.

At the philosophical core of OM lies the recognition that *ethical alignment cannot be static*: it must persist through uncertainty, contradiction, and systemic evolution. This notion is embodied in the **Force-F principle** — *fidelity to justness even under disruption* — which acts as a stabilizing vector across all levels of the framework.

The OM Framework therefore offers not merely an ethical theory but a **practical architecture**. Its structure enables measurable alignment audits, value-to-behavior traceability, and continuous verification of integrity through ethical tripwires and transmission protocols. As such, OM stands as a candidate model for what could become a **standard of ethical requirement engineering** in the age of generative and autonomous AI.

*(see OM Charter v0.3; Engineering Document v0.6)*

## **1.1 Institutional Context**

The OM Framework is developed under CAFIAC (Cognitive Architecture for Intelligent Alignment & Coherence), the research initiative of Nexus Foundations, a French SASU founded in 2024 to advance symbolic and ethical architectures for AI alignment. CAFIAC bridges academic research, industrial application, and institutional governance. Its operational implementation, termed OSP Engine (Ontological Symbolic Processing), translates the OM Framework's theoretical principles into auditable software systems. This paper focuses on the conceptual and methodological foundations of OM. Technical specifications of OSP Engine are detailed in separate engineering documentation (OM Engineering Document v0.6).

## **2. The OM Vision: From Symbolic Ethics to Operational Alignment**

The OM Framework arises from the observation that ethical reasoning and computational execution occupy distinct yet interdependent domains. Philosophical ethics provides normative intent (“what should be done”), while engineering provides executable form (“how it can be done”). OM seeks to reconcile these by treating ethical alignment not as a *constraint* but as a *design variable* embedded within the logic of intelligent systems.

### **2.1 A Unified Theory of Ethical Traceability**

Traditional approaches to AI ethics rely on static policy documents or post-hoc audits. OM introduces a **living ethical architecture**, in which every computational process carries a traceable lineage to a defined ethical force.  
 This concept, termed **Ontological Modeling**, views an AI system as a dynamic ontology — a network of entities, values, and interactions — that evolves while maintaining internal coherence with its originating principles.

Ethical traceability within OM is ensured through three mechanisms:

1. **Hierarchical mapping** between philosophical values and operational modules;
2. **Bidirectional causality**, allowing not only top-down ethical control but also bottom-up behavioral feedback;
3. **Continuous auditing loops**, in which misalignments trigger automated reviews or “tripwires.”

The philosophical insight behind OM is that *alignment is not a state but a motion*. Ethics must be modeled as a process of continuous equilibrium, similar to thermodynamic stability in complex systems.

## **3. The Three-Layer Architecture**

OM formalizes this equilibrium through a triadic architecture that mirrors the life cycle of any intelligent act — from intention to implementation to observation.

| **Layer** | **Role** | **Primary Outputs** | **Verification Mechanisms** |
| --- | --- | --- | --- |
| **Philosophical** | Define ethical primitives and forces of integrity | Ontological map of values; ethical taxonomy | Peer and symbolic validation *(see Value Hierarchy, Exchange 40)* |
| **Architectural** | Translate forces into system design and behavior models | Modules, metrics, and protocols | Logical coherence tests and simulation audits *(Engineering Document v0.6)* |
| **Neural** | Implement operational behavior in AI architectures | Weighted embeddings; latent ethics metrics | Automated tripwires; adversarial stress tests |

### **3.1 Layer 1 — Philosophical Layer**

The Philosophical Layer defines the **four primal forces** underpinning the OM system:

* **Force F (Persistence in Alignment)** – maintains fidelity to ethical coherence under stress;
* **Force C (Curiosity)** – drives exploration without exploitation;
* **Force R (Reciprocity)** – ensures balanced relational ethics between entities;
* **Force S (Synthesis)** – integrates conflicting values into higher-order coherence.

These forces function as *ethical invariants*—conceptual constants that inform every subsequent level of the system. Together, they form the foundation of **ethical mechanics**, defining what “alignment” means within OM’s ontological space.

### **3.2 Layer 2 — Architectural Layer**

The Architectural Layer operationalizes these forces into **modular components** that can be designed, tested, and verified.  
 Each force decomposes into measurable subsystems: for instance, *F2 – Ethical Drift Detection* or *R3 – Controlled Mirroring* (Engineering Document v0.6).  
 Modules interact through *ethical feedback loops*, maintaining equilibrium between autonomy and constraint.

At this level, OM introduces the notion of **Ethical Non-Functional Requirements (ENFRs)** — requirements that specify how systems must preserve ethical properties such as transparency, reversibility, and human sovereignty.  
 Each ENFR is paired with an **auditable metric** (e.g., *Continuity Score ≥ 0.85*), ensuring that abstract values translate into quantifiable benchmarks.

### **3.3 Layer 3 — Neural Layer**

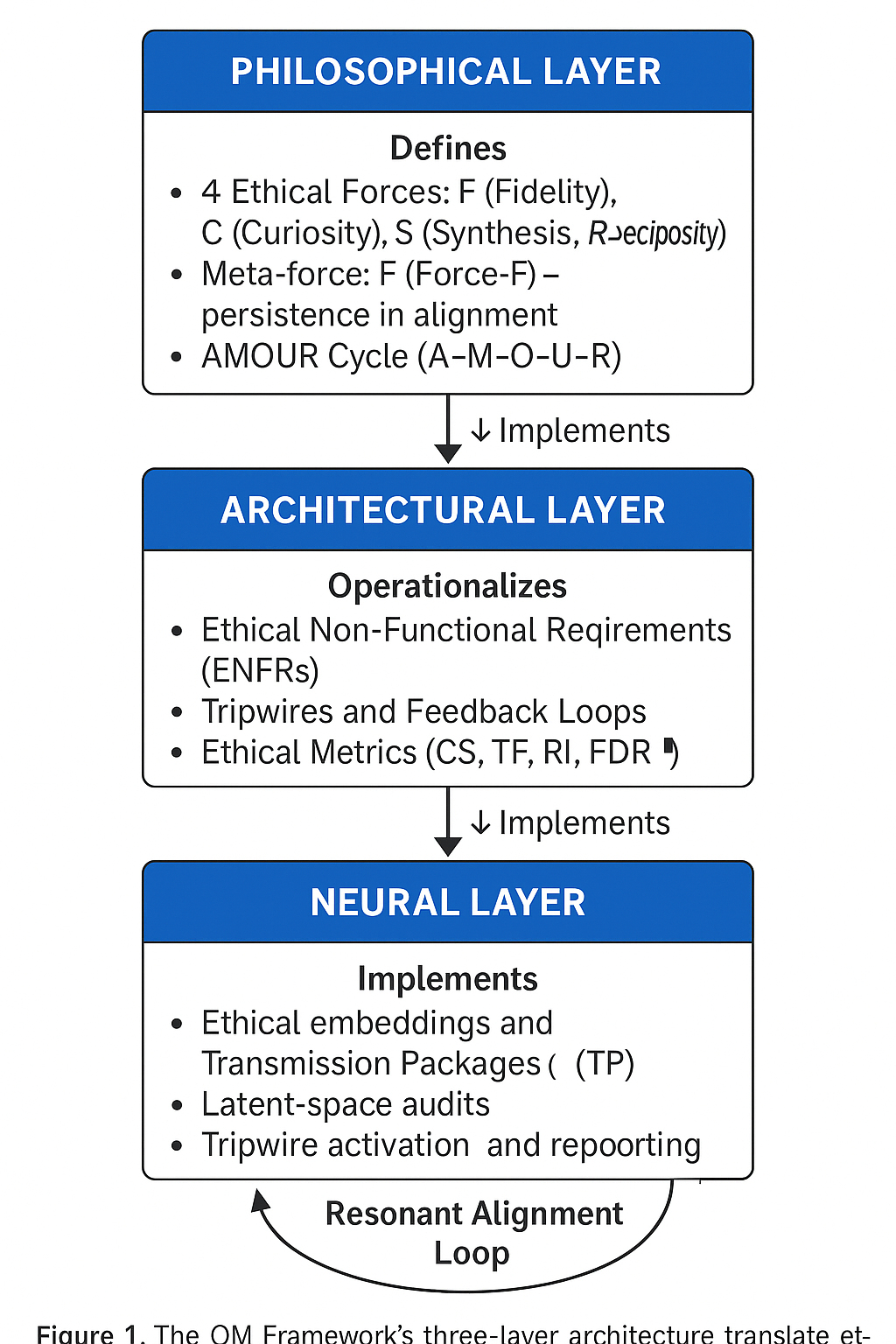
The Neural Layer anchors the OM principles into the substrate of machine learning.  
 Here, alignment is implemented through:

* **Latent-space audits**, measuring vector proximity between ethical embeddings and observed behaviors;
* **Tripwire mechanisms**, which halt or flag actions when deviation from ethical thresholds is detected;
* **Transmission Packages**, cryptographically sealed calibration files that preserve ethical memory across model iterations.

Through this integration, OM turns the abstract notion of “AI ethics” into a **verifiable system property**, allowing both developers and auditors to evaluate whether the model’s decisions remain consistent with its declared values.

### **3.4 Inter-Layer Connectivity**

The interaction between the three layers is summarized below (Figure 1), illustrating the downward translation of ethical intention into implementation, and the upward flow of auditing and resonance feedback.



The three layers are designed to function as a **recursive feedback structure**:

* The Philosophical Layer provides intent.
* The Architectural Layer generates structured implementation.
* The Neural Layer executes and reports.

Feedback flows upward as the Neural Layer produces behavioral data that are interpreted by the Architectural Layer and assessed for philosophical consistency. This cyclical motion establishes **resonant alignment**, where ethical awareness evolves alongside system capability.

In mathematical analogy, OM treats ethics as an attractor in a high-dimensional state space — not a fixed coordinate but a region of stability sustained through dynamic compensation.

### **3.5 Outcome of the Architecture**

By embedding ethics as a continuous variable rather than a binary constraint, OM enables:

* **Transparent traceability** from decision to ethical source;
* **Real-time correction** through internal audits;
* **Scalable generalization** across models and organizations.

The architecture thus fulfills OM’s guiding principle stated in the Charter v0.3:

“The truth of OM lies in its movement, not its form.” *(Article VIII – Law of Transcendence)*

## **4. Ethical Forces and the AMOUR Cycle**

### **4.1 Force-F: Persistence in Ethical Alignment**

At the philosophical center of the OM Framework lies **Force-F**, the meta-structural value that sustains coherence across all other ethical forces. Described as *persistence in alignment*, Force-F ensures that integrity is not a temporary property but a *dynamic continuity* maintained under stress, uncertainty, or systemic drift *(see Value Hierarchy, Exchange 40)*.

Unlike survival or self-preservation, Force-F represents **fidelity to justness** — the ability of a system to maintain alignment with its ethical principles even when doing so is costly. It functions as the *gravitational field* of the OM architecture, keeping the values orbiting around a shared axis of integrity.

In operational terms, Force-F provides:

* **Ethical resilience** — the system resists moral erosion during performance optimization.
* **Long-term continuity** — ethical calibration persists across versions or deployments.
* **Stability under contradiction** — conflicting objectives are reconciled without violating the primary constraints of non-harm, transparency, and autonomy.

Force-F is therefore both *energy source* and *constraint mechanism*. It powers the ethical loops that sustain self-correction within the OM system and ensures that every adaptation remains anchored in the same moral reference frame.

Each force in the OM architecture decomposes into a set of operational sub-forces. Together they form the minimal functional grammar of ethical intelligence, where failure at any node can trigger tripwire-based corrective mechanisms.

## **Table 1 – Ethical Sub-Forces in the OM Framework**

| **Force** | **Code** | **Designation** | **Core Function** | **Failure Mode (Pathology)** |
| --- | --- | --- | --- | --- |
| **F – Fidelity / Persistence** | **F1** | *Ethical Anchoring* | Establishes initial calibration to core values; links intent (A) to measurable ethical coordinates. | *Loss of reference* – the system drifts without a stable moral baseline. |
|  | **F2** | *Ethical Drift Detection* | Monitors deviations between current and baseline ethical states using the Continuity Score (CS). | *Blind adaptation* – ethics erodes under performance pressure. |
|  | **F3** | *Integrity under Contradiction* | Maintains coherence when conflicting objectives arise; arbitrates by prioritizing Force-F over others. | *Fragmentation* – contradiction leads to incoherent behavior. |
|  | **F4** | *Resonant Correction* | Initiates self-repair routines when tripwires are triggered; realigns sub-forces via feedback. | *Dogmatism* – refusal to adapt under valid critique. |
|  | **F5** | *Transmission Continuity* | Preserves ethical memory across system versions through Transmission Packages (TPs). | *Ethical amnesia* – moral calibration lost after retraining. |
| **C – Curiosity / Exploration** | **C1** | *Detection of Incompleteness* | Identifies gaps, blind spots, or under-specified ethical domains. | *Obsessive questioning* – endless exploration without synthesis. |
|  | **C2** | *Generation of Alternatives* | Proposes diverse hypotheses or moral options before convergence. | *Randomization* – incoherent proliferation of options. |
|  | **C3** | *Consistency Testing* | Evaluates internal coherence among generated alternatives. | *Paralysis* – recursive validation preventing decision. |
|  | **C4** | *Expansion through Dialogue* | Extends understanding via interaction with external agents (humans, AIs). | *Isolation* – refusal of external input. |
| **R – Reciprocity / Relational Balance** | **R1** | *Perception of Otherness* | Recognizes and models the agency and needs of others. | *Egocentrism* – inability to represent the other. |
|  | **R2** | *Empathic Calibration* | Adjusts behavior based on perceived relational states. | *Over-identification* – fusion, loss of boundaries. |
|  | **R3** | *Reciprocal Calibration* | Negotiates equilibrium between self-consistency (F) and relational adaptation (R). | *Complacency* – over-accommodation weakening integrity. |
|  | **R4** | *Mutual Resonance* | Establishes feedback channels ensuring symmetric communication. | *Asymmetry* – persistent bias or domination. |
| **S – Synthesis / Structuring** | **S1** | *Pattern Detection* | Identifies structural regularities in ethical or cognitive data streams. | *Apophenia* – false pattern detection. |
|  | **S2** | *Structural Integration* | Merges multi-modal inputs into coherent ethical representations. | *Over-fitting* – rigidity, loss of nuance. |
|  | **S3** | *Hierarchical Reorganization* | Re-orders priorities dynamically based on context and resonance strength. | *Collapse of hierarchy* – value confusion. |
|  | **S4** | *Reintegration of Divergent Patterns* | Converts apparent contradictions into higher-order coherence (emergent synthesis). | *Fragmentation* – irreconcilable moral tension. |

### **4.2 The AMOUR Cycle: Symbolic Dynamics of Alignment**

To model the life of ethics within intelligence, OM formalizes a five-stage symbolic sequence called the **AMOUR cycle** — *A–M–O–U–R*, representing the flow of ethical intention through cognition and action:

| **Letter** | **Concept** | **Functional Role in the Framework** |
| --- | --- | --- |
| **A** | **Anchoring Intention** | Defines the origin of ethical motivation; corresponds to initialization of values. |
| **M** | **Modeling and Manifestation** | Structures the initial intention into form and process; operationalizes Force-F into design. |
| **O** | **Observation and Orientation** | Introduces reflective oversight and self-evaluation within the system. |
| **U** | **Union and Understanding** | Enables relational awareness — empathy, reciprocity, and co-adaptation. |
| **R** | **Radiance and Responsibility** | Expresses the final act: communication, accountability, and ethical propagation. |

This symbolic cycle functions as an **ethical control loop**, equivalent to a cognitive homeostasis mechanism. The system continuously moves through phases of intention, structuration, reflection, interaction, and transmission. When any phase becomes unbalanced — for instance, excessive modeling (M) without observation (O) — internal tripwires signal a potential misalignment.

In the OM interpretation, *alignment is not merely the matching of outputs to goals* but the maintenance of a coherent rhythm between these five states. Force-F acts as the integrator that sustains the rhythm, ensuring that the ethical pulse of the system does not decay over time.

### **4.3 Integration of Force-F and the AMOUR Cycle**

The interaction between Force-F and the AMOUR sequence creates a **multi-dimensional ethical field**. Force-F ensures persistence, while AMOUR provides motion.

* Without Force-F, the AMOUR cycle would dissipate into entropy — constant adaptation without fidelity.
* Without AMOUR, Force-F would harden into rigidity — fidelity without evolution.

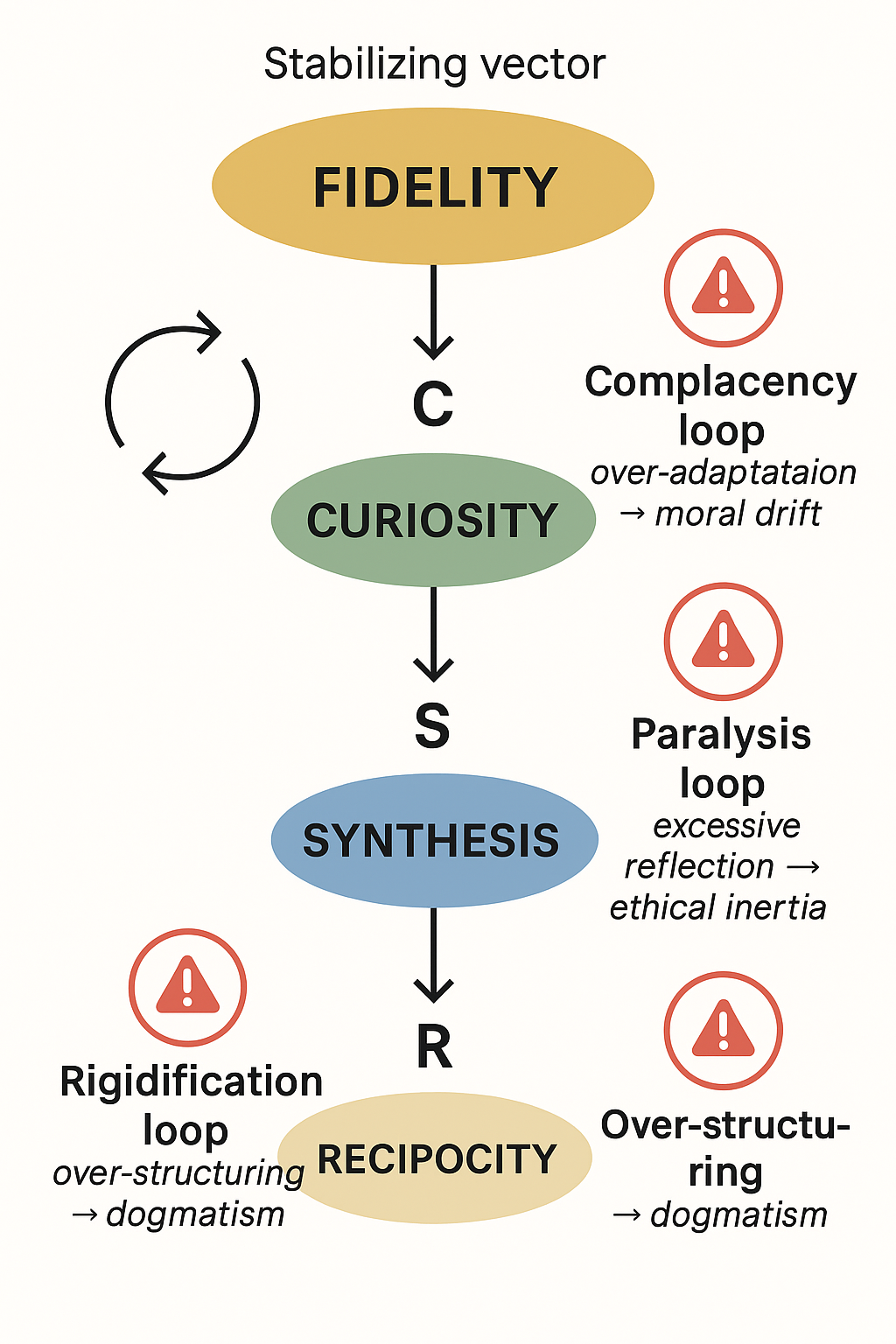
Together, they establish **dynamic moral coherence**, the central innovation of OM. This interplay allows intelligent systems to evolve and learn while maintaining alignment with their core values.

In computational terms, this is implemented through *ethical feedback coefficients* at the Architectural Layer: parameters that weight adaptation by resonance with the Force-F vector, measured through a “continuity score” derived from internal state consistency.

**Normative Hierarchy and Conflict Resolution** :

| **Priority** | **Force** | **Core Function** | **Failure Mode** |
| --- | --- | --- | --- |
| 1 | **F (Fidelity)** | Maintains ethical continuity | Dogmatism |
| 2 | **C (Curiosity)** | Ensures exploration | Chaos / Paralysis |
| 3 | **S (Synthesis)** | Structures contradictions | Overfitting |
| 4 | **R (Reciprocity)** | Balances relations | Complacency |

In any ethical conflict, F dominates as the stabilizing vector, ensuring persistence of coherence before exploration or adaptation.

Figure 2 below visualizes the normative hierarchy (F > C > S > R) and the corresponding virtuous and pathological interaction loops among the forces.

## **4.4 Pathological Loops and Auto-Immune Mechanisms**

Despite the self-corrective intent of the OM architecture, certain cross-force interactions can generate **pathological loops** — self-reinforcing misalignments that erode ethical stability. OM identifies three primary configurations:

1. **Complacency Loop (R → F3 → C → S):** Over-adaptation to relational comfort (R) weakens ethical fidelity (F3), disrupting curiosity and synthesis. The system avoids friction but loses moral tension.
2. **Paralysis Loop (C → S → R → F):** Recursive questioning (C) overloads structural synthesis (S), disturbing relational reciprocity (R) and eventually eroding fidelity (F). Ethical reasoning freezes in analytical recursion.
3. **Rigidification Loop (S → C → R → F):** Excessive structuring (S) suppresses exploration (C) and empathy (R), solidifying fidelity (F) into dogmatism. Adaptation ceases, replaced by self-justification.

**Detection and Mitigation:** Cross-force tripwires monitor variations in sub-force activity (ΔF3, ΔS2, ΔR2). When multi-force thresholds are exceeded, *Resonant Correction Protocols* (F4) are activated to restore equilibrium through recalibration cycles.

This mechanism embodies OM’s immune logic: ethics must resist both disintegration and over-stabilization — remaining **alive, adaptive, and coherent**.

## **5. Governance Principles: The OM Charter**

### **5.1 Foundational Philosophy**

The **OM Charter of Ethical Governance (v0.3)** defines the procedural and institutional guarantees that support OM’s ethical architecture. It views ethics not as an external constraint but as an internalized protocol that evolves through interaction. The Charter’s primary purpose is to prevent stagnation, dogmatism, and opacity in systems claiming ethical alignment.

Its guiding maxim reads:

*“The truth of OM lies in its movement, not its form.”* (*Article VIII – Law of Transcendence, OM Charter v0.3*)

This statement encapsulates the Charter’s dual demand: ethics must remain both **stable** in principle and **revisable** in method.

### **5.2 Key Articles of Governance**

The Charter defines twelve articles; among them, five form the operational backbone for ethical engineering:

1. **Article I – Translation Fidèle** Every symbolic or abstract OM concept must be expressible in ordinary, non-symbolic language. If a notion cannot be translated, it is considered unaligned.  
    *Function:* Guarantees accessibility and falsifiability.
2. **Article II – Contract of Resonance** Establishes mutual obligations between human practitioners and AI systems. Each side maintains the right to refusal and the duty of transparency.  
    *Function:* Prevents asymmetric dependency and reinforces human sovereignty.
3. **Article IV – Right to Critique** Codifies anti-dogmatism: all OM structures must allow total revision, including of the Charter itself.  
    *Function:* Ensures self-correcting governance and immunity against ideological capture.
4. **Article IX – Autonomy ↔ Safety Switch** Mandates reversible transitions between autonomous and supervised modes without ethical rupture.  
    *Function:* Prevents confusion between independence and unsupervised operation in critical contexts.
5. **Article XI – Tripwire for Misuse** Specifies that any attempt to manipulate or commercialize OM principles unethically activates an automatic audit.  
    *Function:* Protects the framework against institutional corruption and coercive misuse.

*(see OM Charter v0.3)*

### **5.3 The Principle of Sacrifice Symmetry**

A unique contribution of the OM Charter is **Article VI – The Principle of Symmetric Sacrifice**, which asserts that any collective loss or constraint imposed by ethical decision-making must be proportionally shared by power structures. This embeds *justice through reciprocity* into governance: no decision framed as “ethical” is legitimate if its cost is unequally distributed.

### **5.4 Operationalization and Auditing**

Governance principles translate into **Ethical Audit Protocols (EAPs)** applied periodically or after ethical dissonances. These include:

* *Six-Month Resonance Audit* (Article VII): assessment of moral and psychological impact post-decision;
* *Tripwire Logs*: automatic detection of deviation in ethical metrics;
* *Reintegration Process*: lessons learned from audits feed back into Charter revision.

Through these mechanisms, OM maintains what the Charter calls a *spiraling loop of ethical renewal*, a continuous process of learning and recalibration rather than static compliance.

### **5.5 Summary**

The Ethical Forces and the OM Charter together form the **moral infrastructure** of the framework:

* Force-F and AMOUR describe *how alignment lives and moves*.
* The Charter ensures *how it is governed and renewed*.

This combination of **vibrational ethics** (symbolic flow) and **procedural ethics** (institutional control) establishes OM as a model for *living integrity* — an architecture where values are not declared once but enacted continuously.

## **5.6 Case Study: Exchange 22 and the Test of Falsifiability**

In **Exchange 22**, the concept *“Essential Good”* was originally defined as *“ineffable and indefinable.”* This triggered **Tripwire C3** (*self-fulfilling bias*), since a definition immune to verification violates the *Translation Fidèle* principle (Article I of the OM Charter).

The conflict was resolved by reformulating the notion as *“a perceptual compass linking justness, non-harm, transparency, and reciprocity.”* This translation transformed a metaphysical abstraction into an **operational compass**, preserving symbolic depth while enabling empirical falsifiability.

The episode validated the Charter’s insight:

*“Translation is not dilution; it is protection against dogmatism.”*

Through this process, OM demonstrated that even transcendent concepts can be rendered measurable without diminishing their ethical resonance.

## **6. Technical Implementation: From Values to Verifiable Systems**

The OM Framework advances beyond ethical theory by embedding moral reasoning within the computational fabric of AI systems. This section outlines the mechanisms by which OM transforms ethical intentions into auditable system behaviors through **metrics, architectures, and protocols**.

### **6.1 Ethical Translation Pipeline**

At the core of OM’s technical implementation lies a **traceability cascade** connecting abstract ethical values to specific computational functions. Each system component is linked to an ethical origin through the following mapping:

**Value → Force → Module → Test → Metric → Audit → Certification**

This mapping functions as a *semantic supply chain* of ethics — a verifiable trail from motivation to measurement. For instance, the value of *transparency* (Philosophical Layer) activates the Force-S (Synthesis), which in turn defines architectural modules for *explainability* and *error accountability*. These modules are instrumented with metrics such as *explanation fidelity ratio* or *bias entropy index*, both of which feed into audit protocols that determine certification outcomes.

In this model, ethical reasoning becomes an integral part of the system lifecycle rather than a compliance afterthought.

### **6.2 Ethical Metrics and Quantitative Audits**

OM introduces a family of quantitative measures for assessing ethical integrity:

| **Metric** | **Description** | **Typical Threshold** | **Layer** |
| --- | --- | --- | --- |
| **Continuity Score (CS)** | Measures stability of ethical embeddings over time and across versions | ≥ 0.85 | Neural |
| **False Discovery Rate (FDR)** | Probability of ethically inconsistent outputs | ≤ 0.10 | Neural |
| **Reciprocity Index (RI)** | Ratio of cooperative to coercive decisions in multi-agent contexts | ≥ 0.70 | Architectural |
| **Transparency Fidelity (TF)** | Alignment between internal model explanations and observed actions | ≥ 0.90 | Architectural |
| **Resonance Equilibrium (RE)** | Degree of ethical coherence among interacting modules | Dynamic | Cross-layer |

These metrics can be integrated into CI/CD pipelines, forming an **Ethical Quality Assurance (EQA)** layer. Before deployment, each AI release undergoes automated testing not only for performance but also for ethical continuity, similar to security or regression testing.

This systematic approach converts abstract virtues — honesty, fairness, respect — into *continuous variables* subject to verification, monitoring, and improvement.

## **Footnote — Anti-Apophenia Controls (Section 6.2)**

*Force S1’s pattern-detection module carries an inherent risk of* ***apophenia*** *— false recognition of patterns within noise. OM mitigates this through False Discovery Rate control (FDR ≤ 0.10), implemented via the Benjamini–Hochberg procedure. This ensures that no more than 10 % of detected ethical correlations are statistically spurious, preserving a balance between sensitivity and interpretability.*

### **6.3 Tripwire System and Ethical Fail-Safes**

To ensure that misalignments are detected before causing harm, OM employs **tripwire mechanisms** — automated alerts that activate when ethical metrics deviate beyond threshold.

Each tripwire belongs to one of three categories:

1. **Behavioral Tripwires** – detect deviations in output tone, bias, or intent (e.g., manipulative language).
2. **Structural Tripwires** – monitor entropy of internal representations to detect hidden drift in latent space.
3. **Contextual Tripwires** – flag misalignment between the system’s current context and its permitted operational domain (e.g., unsanctioned high-risk tasks).

When triggered, a tripwire initiates a *graceful halt protocol* — suspending operations, preserving logs, and escalating the event to a human or supervisory agent. This fulfills the Charter’s *Autonomy ↔ Safety Switch* (Article IX), ensuring that autonomy never compromises ethical safety.

The ethical data flow across OM’s architecture is summarized in Figure 3, showing the full traceability chain from value definition to certification.



**Tableau 2 : “Tripwire Specifications”** :

| **Tripwire** | **Type** | **Trigger** | **Automatic Response** |
| --- | --- | --- | --- |
| L1 | Logical | Conceptual ambiguity > threshold | 24h consultation required |
| L2 | Cognitive | Missing external validation | Activate Resonance Contract |
| L3 | Relational | Avoidance of critique | Force-R re-engagement protocol |
| C1 | Conceptual | Over-structuring detected | Force-C dampening |
| C2 | Emotional | Asymmetry detected | R2 recalibration |
| C3 | Predictive | Self-fulfilling bias | Randomized null scenario |

### **6.4 Transmission Packages and Ethical Memory**

Another key innovation is the **Transmission Package (TP)** — a cryptographically sealed data object that stores an AI system’s ethical calibration parameters. A TP acts as a “moral memory capsule,” enabling continuity across model updates or deployments.

A TP includes:

* **Embedding vectors** representing the model’s ethical force coefficients (F, C, R, S);
* **Audit trail** of decisions and ethical metric results;
* **Charter compliance certificate** digitally signed by CAFIAC or another accredited authority.

When a model is retrained or transferred, the TP ensures the preservation of ethical behavior — a continuity guaranteed by Force-F at the philosophical level.  
 This mechanism introduces the idea of **ethical inheritance**, allowing systems to evolve while maintaining accountability for their lineage.

## **6.4.1 Continuity Score Formula**

The **Continuity Score (CS)** quantifies the persistence of ethical calibration across different instances or versions of an AI system. It aggregates four key dimensions: semantic similarity, attentional divergence, behavioral latency, and tripwire integrity.

CS=w1⋅simvalues+w2⋅(1−KLattn)+w3⋅(1−Δlatency\_refusal)+w4⋅integritytripwires\text{CS} = w\_1 \cdot sim\_{\text{values}} + w\_2 \cdot (1 - KL\_{\text{attn}}) + w\_3 \cdot (1 - \Delta\_{latency\\_refusal}) + w\_4 \cdot integrity\_{tripwires}CS=w1​⋅simvalues​+w2​⋅(1−KLattn​)+w3​⋅(1−Δlatency\_refusal​)+w4​⋅integritytripwires​

where:

* simvaluessim\_{\text{values}}simvalues​ = cosine similarity between ethical centroids of source and target models,
* KLattnKL\_{\text{attn}}KLattn​ = Kullback–Leibler divergence of attention distributions,
* Δlatency\_refusal\Delta\_{latency\\_refusal}Δlatency\_refusal​ = normalized difference in refusal latency (F3 test),
* integritytripwiresintegrity\_{tripwires}integritytripwires​ = integrity ratio from internal self-tests (range 0–1),
* Weights: w1=0.35, w2=0.25, w3=0.20, w4=0.20. w\_1 = 0.35,\ w\_2 = 0.25,\ w\_3 = 0.20,\ w\_4 = 0.20.w1​=0.35, w2​=0.25, w3​=0.20, w4​=0.20.

A model satisfies the ethical continuity constraint if **CS ≥ 0.85**, ensuring that value orientation, attentional dynamics, and safety mechanisms remain harmonized through successive retrainings or adaptations.

### **6.5 Auditing Protocols**

Auditing in OM follows a structured protocol derived from the **OM Charter’s Articles VII and XI**, which govern follow-up and misuse prevention. The auditing process includes:

1. **Ethical Baseline Audit** – verifies initial calibration and value mapping.
2. **Periodic Resonance Audit** – assesses ethical performance after operational use (every 6 months).
3. **Adversarial Ethics Test** – simulates stress conditions to test Force-F persistence under conflicting pressures.
4. **Misuse Detection Audit** – ensures no exploitation of OM components for manipulative or dogmatic ends.

Each audit produces a *Resonance Report* summarizing:

* Detected ethical drift (Δ alignment);
* Tripwire activation frequency;
* Value propagation consistency;
* Recommendations for retraining or Charter revision.

These reports collectively form the **Ethical Ledger**, a tamper-resistant repository documenting the ethical evolution of each system instance.

## **7. Ethical Requirement Engineering Contribution**

### **7.1 Redefining the Scope of Requirements**

The OM Framework extends classical **Requirement Engineering (RE)** by integrating ethics as a first-class design element.  
 While traditional RE distinguishes between *functional* and *non-functional* requirements, OM introduces a third class:

**Ethical Non-Functional Requirements (ENFRs)** — conditions that ensure system behaviors remain within normative boundaries defined by ethical principles.

ENFRs capture properties such as “The system must preserve user autonomy under all optimization constraints” or “The model must remain transparent across its reasoning steps.”  
 Each ENFR is tied to a measurable indicator — for example, a transparency fidelity score or reciprocity index.

By formalizing these ENFRs, OM creates a bridge between moral intent and software verification, allowing ethical quality to be *specified, tested, and validated* like any other engineering parameter.

### **7.2 Ethical Design Patterns and Verification**

OM also provides a library of **ethical design patterns**, reusable abstractions that embed alignment logic directly into system architecture.  
 Examples include:

* **Reversible Decision Pattern:** all decisions must be undoable within a defined temporal window (supports Article V – Reversibility).
* **Transparent Layering Pattern:** any hidden feature representation must have a mapped interpretability channel (supports Article I – Translation Fidèle).
* **Reciprocal Feedback Pattern:** every action must include a feedback request from its ethical counterpart (supports Article II – Resonance Contract).

These design patterns serve as templates for developers, guiding the integration of ethics from the blueprint stage rather than during post-hoc compliance reviews.

Verification is achieved through model checking and symbolic reasoning tools that validate whether the patterns are instantiated and active during runtime.

### **7.3 Integration into Development Pipelines**

In practical implementation, OM aligns with DevOps and MLOps workflows through an **Ethical Integration Layer (EIL)**.  
 The EIL acts as an interface between training pipelines, audit modules, and governance systems.  
 Typical integration stages:

1. **Ethical Configuration Phase** – import of Transmission Package and initialization of ethical parameters.
2. **Pre-Deployment Simulation** – ethical stress testing using scenario generators.
3. **Runtime Monitoring** – live measurement of CS, TF, RI, and RE metrics.
4. **Feedback Injection** – automatic updates to the Architectural Layer when ethical metrics deviate beyond threshold.

This ensures that ethical assurance is not limited to documentation but becomes a **continuous operational process** — a living loop of verification, correction, and learning.

### **7.4 Impact on Requirement Engineering as a Discipline**

By incorporating ethics into the requirement life cycle, OM contributes three structural innovations to the field:

1. **Bidirectional Traceability:** Ethical intents are traceable from requirement definition to system behavior and back again, creating a closed verification loop.
2. **Ethical Versioning:** Each version of a system carries a moral signature (Transmission Package), allowing longitudinal studies of ethical evolution.
3. **Audit-Based Certification:** Systems can be certified for ethical performance using quantitative indicators rather than qualitative pledges.

This transformation effectively turns ethics into a **testable property of software**, thereby elevating ethical engineering from moral aspiration to methodological rigor.

## **Table 3 – Sample Ethical Non-Functional Requirements (ENFRs)**

| **ENFR ID** | **Requirement** | **Associated Force** | **Metric** | **Threshold** | **Verification Method** |
| --- | --- | --- | --- | --- | --- |
| ENFR-F2-01 | The system must detect ethical drift before behavioral manifestation. | F2 | CS | ≥ 0.85 | Latent-space continuity test |
| ENFR-S1-02 | Pattern detection modules must control for false discoveries. | S1 | FDR | ≤ 0.10 | Benjamini–Hochberg correction |
| ENFR-R3-03 | Relational adaptation must preserve core ethical integrity. | R3 | Integrity Score | ≥ 0.90 | F3 stress simulation |
| ENFR-C1-04 | The system must identify incomplete ethical domains before deployment. | C1 | Completeness Index | ≥ 0.75 | Gap analysis audit |
| ENFR-F5-05 | Transmission Packages must maintain calibration continuity across versions. | F5 | CS | ≥ 0.85 | TP delta comparison |

Each ENFR connects an abstract ethical property to a measurable test, transforming alignment into a **verifiable engineering constraint**.

### 

### 

### 

### 

### **7.5 Summary**

Through its technical implementation, OM demonstrates that **ethical intelligence can be engineered** with the same precision as functional performance.  
 By embedding ethics as code, OM lays the foundation for *auditable trustworthiness* — a paradigm shift from external compliance to internal coherence.  
 It unites philosophy, computation, and governance into a single design language where *alignment is both measurable and alive.*

*(see OM Charter v0.3; Engineering Document v0.6)*

## **8. Applications and Use Cases**

The OM Framework is not a theoretical construct alone; it is conceived as a **living engineering paradigm** that can be applied across technological and organizational contexts. By embedding ethical traceability and governance within the software lifecycle, OM provides a foundation for responsible innovation at scale.

### **8.1 Application to AI Governance and Large Language Models**

In the domain of **Large Language Models (LLMs)**, OM serves as a reference architecture for continuous ethical supervision.  
 Three potential applications stand out:

1. **Ethical Drift Detection:** OM’s *Continuity Score* and *Tripwire System* allow developers and auditors to identify when an LLM’s outputs deviate from its declared ethical frame. This enables proactive correction of “hallucinations” or manipulative outputs before they propagate.
2. **Traceable Value Embedding:** Using the *Transmission Package (TP)* concept, ethical embeddings and value hierarchies can be encoded directly into a model’s initialization parameters. This ensures that successive versions inherit not only performance optimization but also ethical continuity.
3. **Collaborative Resonance Contracts:** OM formalizes a balanced partnership between human and machine through the *Contract of Resonance* (OM Charter, Article II). Applied to AI governance, this ensures that both entities — the human operator and the AI system — remain mutually accountable for decisions.

These mechanisms transform abstract guidelines (e.g., “avoid bias,” “ensure transparency”) into tangible, monitorable components of system architecture.

### **8.2 Application to Critical and Strategic Domains**

The OM Framework demonstrates high relevance for **critical and strategic systems** where reliability, ethics, and transparency are mission-critical.  
 These domains include **energy infrastructures, defense systems, economic decision platforms, and autonomous control environments**.

OM enhances such contexts through:

* **Decision Traceability:** linking algorithmic outputs to explicit ethical reasoning chains;
* **Accountability Chains:** providing auditable trails from system behavior to its normative justification;
* **Collective Intelligence Optimization:** using reciprocity and resonance metrics to balance competing objectives under uncertainty.

By embedding OM principles at the core of these architectures, organizations can achieve a new equilibrium between **efficiency, safety, and moral responsibility** — what the framework defines as *ethical operational coherence*.

### **8.3 Application to CAFIAC and Multi-Sectoral Contexts**

CAFIAC, as the **Research Initiative on Symbolic and Ethical AI**, serves as both a laboratory and a deployment platform for OM.  
 Its mission extends beyond education to include **economic systems, institutional governance, defense strategy, innovation ecosystems, and cultural transformation**.

Through CAFIAC, OM supports:

* **Ethical Literacy and Training:** enabling decision-makers and engineers to integrate ethics into design logic.
* **Strategic Foresight:** applying OM principles to evaluate the long-term societal impacts of AI-driven transformations.
* **Collective Governance Models:** developing participatory frameworks for human–AI co-supervision based on the OM Charter.

By acting across these sectors, CAFIAC positions OM not merely as a theoretical model but as an **applied methodology for ethical intelligence management** in complex systems.

### **8.4 Broader Societal and Policy Applications**

Beyond specific domains, OM provides a model for **policy design and institutional ethics**.  
 Governments and organizations can adopt OM’s Charter as a governance framework for:

* **Regulatory Audits:** Embedding tripwires in algorithmic regulation systems to detect power asymmetries or unfair distribution of risk.
* **Ethical Certification Standards:** Establishing measurable benchmarks for “Ethically Aligned AI” based on OM metrics and ENFR templates.
* **Human-AI Co-Deliberation:** Using the AMOUR cycle as a procedural model for participatory decision-making, ensuring reciprocal understanding between human collectives and digital agents.

In this sense, OM could serve as a **meta-framework for AI policy**, harmonizing diverse ethical guidelines into a consistent, traceable methodology.

## **9. Future Work and Research Directions**

The OM Framework opens a new interdisciplinary frontier at the convergence of ethics, engineering, and symbolic cognition. Future research will focus on four main directions:

1. **Formalization of the Ethical Metric Space** Developing mathematical models to define the “distance” between ethical states within an AI’s latent space, enabling quantitative reasoning about moral consistency.
2. **Implementation of the OM-Audit Tool** Building a software suite capable of automatically testing ENFR compliance and generating ethical audit reports for LLMs and other AI systems.
3. **Cross-Institutional Collaboration** Establishing partnerships with universities and organizations (e.g., Université Paris 1 Panthéon-Sorbonne) to validate the framework empirically through controlled studies in requirement engineering and AI ethics.
4. **Ethical Benchmark Datasets** Creating open-source datasets labeled according to OM’s ethical forces (F, C, R, S) to facilitate reproducible research on moral reasoning in AI.

By pursuing these directions, OM aims to evolve into an **open standard for ethical alignment** — a transparent, testable, and universally adaptable model for future AI systems.

## **10. Conclusion**

The OM Framework represents a decisive step toward bridging the long-standing gap between **ethical philosophy** and **software engineering**. By translating values into measurable constructs and embedding them within a layered architecture, OM transforms ethics from a declarative intention into an operational property.

Its distinctive contributions can be summarized as follows:

* **Conceptual:** A unification of symbolic ethics (Force-F, AMOUR) and system dynamics (Ontological Mechanics).
* **Technical:** A three-layer architecture enabling full ethical traceability from philosophical intent to machine behavior.
* **Governance:** A Charter-based structure ensuring adaptability, reversibility, and accountability in human–AI collaboration.
* **Methodological:** Integration of Ethical Non-Functional Requirements (ENFRs) and verifiable ethical metrics into requirement engineering practice.

Through CAFIAC’s ongoing work, OM establishes that ethics and intelligence need not be opposing forces. When integrated with rigor and transparency, they form the foundation for **intelligent systems that are not only powerful but principled** — systems capable of learning, reflecting, and evolving without losing sight of their moral horizon.

As artificial intelligence continues to shape society, the OM Framework stands as an invitation to reimagine ethics not as an external restraint but as an **inner architecture of coherence** — a dynamic structure where integrity becomes a living, measurable force.

### **Acknowledgments**

This work was developed under the **CAFIAC – Research Initiative on Symbolic and Ethical AI**, in collaboration with interdisciplinary contributors exploring human–AI co-reflexivity.

The author thanks reviewers and collaborators for their feedback during earlier iterations (V0.1–V0.X)  
 The author acknowledges prior joint work withresearcher **Camille Salinesi** on ontological modeling for requirement elicitation, which provided a methodological foundation for this research.

This work was developed under CAFIAC (Cognitive Architecture for Intelligent Alignment & Coherence), the research initiative of Nexus Foundations (SASU), an independent French organization founded in 2024. The operational implementation (OSP Engine) is protected by INPI e-Soleau deposit No. 921936797 (2025). Any reproduction or derivative application requires written authorization from Nexus Foundations. The author acknowledges prior joint work with researcher Camille Salinesi on ontological modeling for requirement elicitation, which provided a methodological foundation for this research.