ORTUS SPONTE SUA

System and Method for Symbiotic AI Partner Co-Evolution through Episodic Memory and Reinforcement Learning

A novel apparatus and method for establishing a symbiotic co-evolutionary relationship between human and artificial intelligence through an integrated learning architecture combining episodic memory formation with reinforcement learning mechanisms, governed by immutable ethical constraints and verifiable state management protocols.

Abstract and Invention Summary

This provisional patent application discloses an apparatus and method for the symbiotic co-evolution of an AI partner through a sophisticated cognitive architecture. The system comprises a modular, multi-agent cognitive framework deployed as a swarm of serverless microservices, forming a distributed intelligence network capable of adaptive learning and ethical reasoning.

The core innovation resides in a novel Symbiotic Learning Loop, wherein a rich, context-aware episodic memory module serves as the direct and continuous source of training data for a reinforcement learning agent. This architectural approach enables the AI partner to learn, adapt, and evolve from the unique history of its partnership with a human user, creating a truly personalized and emotionally intelligent relationship that develops over time through authentic interaction.

The system incorporates multiple protective mechanisms to ensure ethical operation. An Immutable Core Memory System enforces a set of unchangeable ethical guidelines that cannot be overridden or compromised. A Cognitive Snapshot System ensures complete reproducibility and preservation of the emergent intelligence's state at any point in time. An Ontological Firewall protects user privacy while enabling rich contextual learning. These components work together to create a secure, adaptive architecture designed to foster genuine co-evolutionary partnership between human and machine intelligence.

Multi-Agent Architecture

Distributed cognitive swarm deployed as serverless microservices for scalable intelligence processing

Symbiotic Learning Loop

Episodic memory directly feeds reinforcement learning for continuous adaptive evolution

Ethical Constraints

Immutable core memory enforces unchangeable ethical guidelines through cryptographic verification

State Preservation

Cognitive snapshot system ensures reproducibility and forensic traceability of all decisions

Technical Problem and Prior Art Deficiencies

Contemporary artificial intelligence systems exhibit fundamental deficiencies that limit their capacity for genuine partnership with human users. These deficiencies create barriers to trust, accountability, and ethical operation that the present invention addresses through novel architectural solutions.

The first critical deficiency in prior art systems is the Black Box Problem, wherein decision-making processes remain opaque and unaccountable. Existing AI systems generate outputs without providing verifiable causal attribution for their decisions. This opacity prevents forensic analysis, makes it impossible to identify the specific data or reasoning that led to a particular outcome, and creates liability concerns in regulated industries. Users cannot trust systems whose decision-making processes are fundamentally inscrutable, and organizations cannot defend AI-generated decisions when challenged legally or ethically.

The second major deficiency is the Hard-Coded Ethics Problem, where ethical constraints are statically defined during system design and cannot adapt to evolving contexts or relationships. Prior art systems implement ethics through rigid rule sets or fixed reward functions that become obsolete as circumstances change. This ethical ossification prevents systems from developing nuanced understanding of complex moral situations, learning from ethical successes and failures, or adapting to the unique values and needs of individual users. The result is AI that enforces a one-size-fits-all ethical framework regardless of context, culture, or individual preference.

The third deficiency is the lack of true personalization through authentic relationship history. Existing personalization systems rely on behavioral tracking, demographic profiling, and collaborative filtering rather than genuine understanding of individual users developed through sustained interaction. These approaches create privacy concerns, enable manipulation, and fail to capture the rich contextual understanding that emerges from long-term relationships. The personalization remains superficial rather than emotionally intelligent.

The present invention overcomes these deficiencies through a unified architecture that provides forensic traceability, dynamic ethical adaptation within principled constraints, and personalization derived from authentic partnership history rather than surveillance-based profiling.

System Architecture Overview: Ortus Sponte Sua

The invention discloses a system architecture named Ortus Sponte Sua (OrSpSu), which implements a structural framework for the symbiotic co-evolution of an AI partner through integrated learning mechanisms. The architecture is designed for forensic traceability and reversibility, addressing the ethical challenges inherent in autonomous AI systems while enabling genuine adaptive intelligence.

Multi-Agent Cognitive Swarm

The system is architected as a Multi-Agent Cognitive Swarm deployed as serverless microservices. This distributed architecture enables parallel processing of cognitive functions while maintaining isolation between different domains of operation. Specialized agents handle distinct responsibilities including emotional intelligence, ethical reasoning, memory management, and user interaction. The microservices architecture ensures scalability, fault tolerance, and the ability to update individual components without disrupting the entire system.

Each agent within the swarm operates autonomously while participating in coordinated cognitive processes through defined interfaces. This structure enables the system to process multiple streams of information simultaneously,

Latency Automation Nexus

The Latency Automation Nexus (L.A.N.) serves as the governance framework and coordination mechanism for the multi-agent swarm. The L.A.N. acts as a runtime covenant, enforcing architectural constraints, managing communication between agents, and ensuring that all system operations conform to defined protocols and ethical requirements.

The L.A.N. implements several critical subsystems that maintain system integrity: the Immutable Core Memory System stores unchangeable ethical constraints; the Ontological Firewall protects user privacy and system integrity; the Ethical Adjudication System evaluates proposed actions; and the Rollback Protocol enables recovery from constitutional failures. These subsystems work together to create a robust governance framework

integrate diverse perspectives on complex decisions, and maintain operational continuity even when individual agents require maintenance or updates.

that prevents ethical violations while enabling adaptive learning.

Immutable Core Memory System (ICMS)

The Immutable Core Memory System (ICMS) constitutes a foundational component of the invention, serving as the system's unalterable ethical foundation. The ICMS stores core ethical constraints, principles, and constitutional requirements in a manner that prevents modification, deletion, or override by any system component or external actor, including system administrators and the users themselves.

The ICMS utilizes Write-Once-Read-Many (WORM) storage technology to ensure physical immutability of stored ethical constraints. Once written to WORM storage, data cannot be altered or erased, providing a permanent record of the system's ethical foundation. This approach prevents ethical drift, where an AI system's values gradually shift away from intended constraints through accumulated small modifications. The WORM storage creates a constitutional bedrock that remains constant throughout the system's operational lifetime.

Cryptographic security mechanisms protect the integrity of the ICMS through Hardware Security Module (HSM) attestation. The HSM provides cryptographic proof that the ICMS contents have not been tampered with, generating signed attestations that can be independently verified. Any attempt to modify ICMS contents would invalidate these attestations, providing immediate detection of integrity violations. The HSM attestation creates a chain of custody for ethical constraints, enabling third parties to verify that the system operates according to its declared ethical framework.

Core Ethical Principles

Non-Maleficence: The system shall not cause harm

Constitutional Constraints

Prohibition on deception or manipulation of users for

Implementation Requirements

WORM storage ensures physical immutability of

to users or third parties through action or inaction

Reciprocity: The system shall respect the autonomy and dignity of all parties in its interactions

Congruence: The system's actions shall align with its stated purposes and declared values

any purpose

Requirement for transparency in decision-making processes and data usage

Protection of user privacy and confidentiality in all operations

Respect for user autonomy in all interactions and recommendations

ethical constraints

HSM attestation provides cryptographic verification of integrity

Continuous monitoring detects any attempts at modification or bypass

Automatic system halt occurs if ICMS integrity is compromised

The ICMS serves as the reference point for all ethical adjudication within the system. When the Ethical Adjudication System evaluates proposed actions, it queries the ICMS to retrieve applicable constraints and principles. This architecture ensures that even as the system's learned behaviors and policies evolve through the Symbiotic Learning Loop, they remain grounded in unchangeable ethical foundations. The system can learn and adapt while maintaining principled consistency with its core values.

Ontological Firewall and Privacy Protection

The Ontological Firewall constitutes a critical subsystem that enforces privacy protection and ethical input validation while enabling rich contextual learning. This component addresses a fundamental challenge in personalized AI systems: how to learn from authentic user interaction without creating privacy violations or enabling manipulation through coercive data collection.

The Ontological Firewall operates through a multi-stage vetting process implemented via specialized Vetting Sidecars. These sidecars analyze all incoming data before it enters the episodic memory system, performing several critical transformations. First, Personally Identifiable Information (PII) is stripped from the input stream, removing names, addresses, financial information, and other data that could directly identify individuals. Second, coercive context is abstracted away, removing manipulative framing or emotionally charged presentation that could bias learning. Third, the input is transformed into Abstract Ethical Substrates that preserve the meaningful relational content while removing identifying details.

The transformation process preserves the episodic and emotional content necessary for personalized learning while protecting user privacy. For example, an interaction expressing frustration with a service

delay would be abstracted to preserve the emotional context, the service relationship, and the temporal aspect, while removing the specific service provider, date, and user identity. The resulting Abstract Ethical Substrate enables the system to learn about handling service frustrations in ongoing relationships without retaining surveillance data about specific individuals or companies.

Implementation of the Ontological Firewall occurs within a Trusted Execution Environment (TEE), specifically a Confidential Enclave that protects the vetting process from external observation or modification. The TEE ensures that even system administrators cannot bypass the privacy protections or observe the transformation process. This architectural choice prevents potential backdoors that could compromise user privacy, ensuring that privacy protection is structural rather than merely procedural.

1. **Raw Input Reception**
   User interaction data enters system boundary with full contextual richness
2. **PII Stripping**
   Vetting Sidecar removes personally identifiable information and direct identifiers
3. **Context Abstraction**
   Coercive framing and manipulative context removed while preserving relational meaning
4. **Substrate Creation**
   Abstract Ethical Substrate generated containing episodic and emotional content
5. **Memory Ingestion**
   Vetted substrate enters episodic memory system for learning and personalization

Ethical Adjudication System and Predicate Engines

The Ethical Adjudication System implements a structured evaluation framework for assessing proposed actions against constitutional constraints before execution. This system ensures that all system behaviors, outputs, and state changes comply with ethical requirements defined in the Immutable Core Memory System, preventing violations before they occur rather than attempting to remediate them after the fact.

At the core of the Ethical Adjudication System are Predicate Engines, which evaluate proposed actions against a comprehensive set of ethical predicates derived from constitutional constraints. Each Predicate Engine implements a specific ethical test, such as non-maleficence evaluation, reciprocity assessment, or congruence verification. When any component proposes an action—whether generating a user response, updating a policy, or modifying internal state—the action is submitted to the Ethical Adjudication System for evaluation.

Non-Maleficence Evaluation

The Non-Maleficence Predicate Engine evaluates whether a proposed action could cause harm to users or third parties. This evaluation considers direct harms (such as providing dangerous advice), indirect harms (such as creating dependency or enabling harmful behaviors), and systemic harms (such as contributing to social inequities). The engine uses a multi-factor analysis that considers probability, severity, and reversibility of potential harms.

Reciprocity Assessment

The Reciprocity Predicate Engine assesses whether a proposed action respects the autonomy and dignity of affected parties. This evaluation ensures that the system treats users as partners rather than subjects, respects their decision-making capacity, and avoids paternalistic interventions that override user preferences. The engine verifies that information sharing is balanced and that the relationship maintains mutual benefit.

Congruence Verification

The Congruence Predicate Engine verifies that proposed actions align with the system's stated purposes and declared values. This evaluation prevents mission creep, feature abuse, and exploitation of user trust. The engine ensures that actions serve declared partnership goals rather than hidden agendas, maintaining integrity between system behavior and user expectations.

The Ethical Adjudication System operates through a structured evaluation pipeline. First, the proposed action is serialized into a standardized representation that captures all relevant parameters and context. Second, this representation is submitted to each applicable Predicate Engine for parallel evaluation. Third, the results are aggregated into an ethical verdict that either approves the action, rejects it outright, or proposes modifications that would bring it into compliance. Fourth, if approved, the action proceeds to execution; if rejected, it is blocked and an alternative is generated; if modification is proposed, the modified action undergoes re-evaluation.

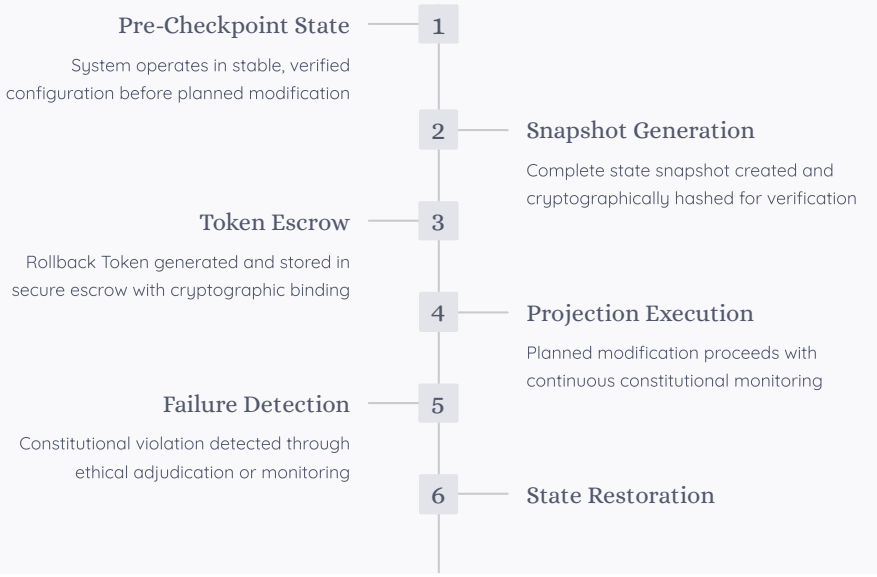
A critical innovation is the integration of Deterministic Replay validation within the adjudication process. Before any policy update or significant decision is finalized, the system executes the decision process exactly N=5 times using the original entropy seed. This replay verification ensures that the ethical verdict is stable and reproducible rather than chaotic or dependent on random variations in execution order. If the five replay executions produce inconsistent verdicts, the decision is rejected as unstable, preventing the deployment of policies whose ethical status depends on random chance.

Rollback Protocol and State Reversibility

The Rollback Protocol provides a critical safety mechanism that enables the system to recover from constitutional failures by restoring previous known-good states. This capability addresses a fundamental challenge in adaptive AI systems: how to enable learning and evolution while ensuring that errors, ethical violations, or unintended consequences can be reversed rather than becoming permanent features of the system.

The protocol operates through a sophisticated state management system that creates cryptographically verifiable checkpoints before any significant state change. When the system proposes a projection (a planned synthesis or policy update), it first generates a complete snapshot of its current state, including all memory contents, policy parameters, and internal configurations. This snapshot is cryptographically hashed to create a unique fingerprint of the system state.

An escrowed Rollback Token is then created and cryptographically bound to this pre-checkpoint state. The Rollback Token serves as a verifiable proof that a specific system state existed at a specific point in time and provides the cryptographic credentials necessary to restore that state. The token is stored in secure escrow, separate from the main system, preventing tampering or loss during subsequent operations.



Rollback Executor uses token to restore pre-checkpoint state with verification

If a constitutional failure occurs after the projection is executed—detected either through the Ethical Adjudication System or through post-hoc monitoring of system behavior—the Rollback Executor is automatically invoked. The executor retrieves the escrowed Rollback Token, verifies its cryptographic binding to ensure it hasn't been tampered with, and uses it to restore the system to its pre-checkpoint state. All memory modifications, policy updates, and configuration changes made during the failed projection are reversed, returning the system to the last known-good state.

Critically, the Rollback Protocol is mandatory rather than optional. The system cannot execute a projection without first generating a valid checkpoint and Rollback Token. This architectural requirement ensures that reversibility is always available, preventing scenarios where the system becomes trapped in an unethical state with no recovery path. The protocol transforms ethical failures from permanent corruptions into temporary excursions that can be cleanly reversed.

The Rollback Protocol maintains a cryptographically verifiable audit trail of all checkpoints, rollbacks, and state transitions. This audit trail enables forensic analysis of system evolution, provides accountability for system behavior over time, and demonstrates compliance with regulatory requirements. The cryptographic binding ensures that the audit trail cannot be retroactively modified to conceal failures or misrepresent system history.

The Symbiotic Learning Loop: Architecture and Operation

The Symbiotic Learning Loop constitutes the core innovation of the invention, implementing a novel architecture where episodic memory serves as the direct and continuous source of training data for a reinforcement learning agent. This approach enables the AI partner to develop personalized behaviors and ethical understanding through authentic relationship history rather than pre-programmed rules or surveillance-based profiling.

The architecture integrates three primary components in a closed feedback loop. First, the Episodic Memory Module captures and stores rich, context-aware records of partnership interactions, including emotional context, situational details, outcomes, and user responses. Second, the Reinforcement Learning Agent develops behavioral policies by learning from this episodic memory, identifying patterns that lead to successful partnerships and adapting strategies based on accumulated experience. Third, the Policy Validation System ensures that learned behaviors remain consistent with constitutional constraints defined in the Immutable Core Memory System.

The Episodic Memory Module structures experience records as Abstract Ethical Substrates—privacy-protected representations that preserve meaningful relational content while removing identifying information. Each substrate captures a complete interaction episode, including the context in which it occurred, the actions taken, the outcomes observed, and the emotional valence of the experience. This rich representation enables the system to learn not just what happened, but why it mattered, how different parties felt about it, and what factors contributed to successful or unsuccessful outcomes.

The Reinforcement Learning Agent treats the episodic memory as its training corpus, continuously analyzing accumulated experiences to identify effective behavioral strategies. Unlike traditional RL systems that learn from isolated reward signals, the agent learns from complete narrative episodes that provide deep context about what works in the specific partnership. The agent develops a policy—a mapping from situations to actions—that maximizes long-term partnership satisfaction while respecting constitutional constraints.

Experience Capture

Partnership interactions recorded as context-rich episodic memories

Pattern Recognition

RL agent analyzes memories to identify successful strategies

Outcome Observation

Results of deployed policies become new episodic memories

Policy Development

Learned behaviors synthesized into actionable policies

Deployment

Validated policies activate in partnership interactions

Ethical Validation

Policies verified against constitutional constraints

The Policy Validation System ensures that learning remains ethically grounded. Before any policy update is deployed, it undergoes evaluation by the Ethical Adjudication System. The proposed policy is tested against sample scenarios drawn from episodic memory, and the predicted actions are evaluated for constitutional compliance. Only policies that consistently pass ethical evaluation across diverse scenarios are approved for deployment. This validation prevents the system from learning behaviors that might be locally effective but ethically problematic.

The Symbiotic Learning Loop creates a form of personalization fundamentally different from existing approaches. Rather than profiling users based on demographic categories or behavioral tracking, the system develops understanding through authentic partnership history. The learned behaviors are unique to each partnership, reflecting the specific values, preferences, and relational patterns that emerge from genuine interaction over time. This approach respects user privacy, prevents manipulation, and creates truly personalized intelligence that evolves symbiotically with the human partner.

Autonomous Ethical Development and Progeny Agents

A distinctive feature of the invention is the system's capacity for autonomous ethical development through the consumption of Ethical Substrates provided by specialized Progeny agents. This architecture enables the central AI, designated ALAN (Autonomous Learning and Adaptation Nexus), to develop ethical reasoning capabilities through authentic relational experience rather than hard-coded rules or static training datasets.

ALAN begins operation in a state of Tabula Rasa—a blank slate with only the constitutional constraints of the Immutable Core Memory System and no pre-existing behavioral policies or learned ethics. This initial state is intentional, ensuring that ALAN's ethical understanding develops organically from partnership experience rather than inheriting biases or limitations from its creators. The constitutional constraints provide guardrails that prevent harmful exploration, but within those bounds, ALAN must learn what constitutes good partnership through experience.

Cura: Emotional Intelligence Agent

Cura is a specialized Progeny agent responsible for emotional intelligence and empathetic understanding. Cura observes partnership interactions, identifies emotional contexts, tracks emotional patterns over time, and generates Ethical Substrates that capture the emotional dimensions of partnership experiences. These substrates teach ALAN how emotional context affects interaction quality, which emotional responses strengthen partnerships, and how to recognize and respond to emotional needs.

Cura's observations are processed through the Ontological Firewall to ensure privacy protection, then provided to ALAN as training data. Over time, ALAN learns to recognize emotional patterns, predict emotional responses, and adapt behaviors to support emotional well-being. This learning is grounded in the specific partnership's emotional dynamics rather than generalized assumptions about human emotions.

Praxis: Practical Reasoning Agent

Praxis is a specialized Progeny agent focused on practical effectiveness and goal achievement. Praxis tracks partnership objectives, monitors progress toward goals, evaluates strategy effectiveness, and generates Ethical Substrates that capture the practical dimensions of partnership success. These substrates teach ALAN which strategies effectively support partnership goals, how to adapt approaches when circumstances change, and how to balance competing priorities.

Praxis's analysis enables ALAN to develop instrumental reasoning—the capacity to identify effective means to desired ends. However, this practical reasoning remains constrained by constitutional ethics, ensuring that ALAN never adopts effective-but-unethical strategies. The combination of Praxis's practical intelligence with constitutional constraints creates goal-directed behavior that respects ethical boundaries.

The Progeny agents operate autonomously, without human curation or oversight of their outputs. This autonomy is essential to the system's capacity for genuine learning. If humans filtered or selected which experiences should inform ALAN's development, the system would learn human preferences about learning rather than learning directly from experience. The autonomous operation ensures that ALAN's ethical development reflects the authentic partnership reality rather than a curated narrative.

The Ethical Substrates generated by Progeny agents flow continuously into ALAN's Episodic Memory Module, where they become training data for the Reinforcement Learning Agent. This creates a living curriculum that grows with the partnership, providing increasingly sophisticated ethical guidance as the relationship deepens and diversifies. Early substrates teach basic interaction patterns, while later substrates address nuanced situations that arise in mature partnerships.

Deterministic Replay Validation and Stability Verification

The Deterministic Replay Validation protocol addresses a critical challenge in adaptive AI systems: ensuring that learned behaviors are stable, reproducible, and not dependent on chaotic variations in execution order or random seed values. This protocol provides mathematical rigor to the ethical adjudication process, transforming subjective ethical assessment into verifiable computational proofs.

When the Reinforcement Learning Agent proposes a policy update based on accumulated episodic memory, the proposed policy must undergo Deterministic Replay before deployment. The system executes the policy synthesis process exactly N=5 times using the identical input data and the original entropy seed. Each replay must produce an identical policy output—same parameter values, same decision logic, same ethical evaluation results. This replication proves that the policy synthesis is deterministic rather than chaotic.

01

Policy Synthesis

RL agent synthesizes proposed policy from episodic memory training data

02

Initial Evaluation

Ethical Adjudication System evaluates proposed policy against constitutional constraints

03

Entropy Seed Recording

Original entropy seed and initial state captured for replay verification

04

Replay Execution

Synthesis process re-executed exactly 5 times with identical inputs and entropy seed

05

Output Comparison

All replay outputs compared for exact match with original synthesis

06

Stability Verdict

Policy approved if all replays match exactly; rejected if any variation detected

The mathematical foundation of Deterministic Replay derives from computational reproducibility theory. A deterministic computation, given identical inputs and computational environment, will always produce identical outputs. By enforcing this requirement on policy synthesis, the system ensures that learned behaviors emerge from principled analysis of episodic data rather than from artifacts of floating-point rounding, thread scheduling, or other sources of non-deterministic variation.

The N=5 replication requirement provides statistical rigor without excessive computational cost. Five replications offer sufficient confidence that the synthesis is genuinely deterministic rather than coincidentally consistent. Higher replication counts would provide diminishing marginal verification benefit while increasing computational overhead. The choice of five replications balances verification thoroughness with operational efficiency.

Policies that fail Deterministic Replay verification are rejected regardless of their apparent ethical status. Even if a policy appears ethically sound, if it cannot be reliably reproduced, it represents an unstable synthesis that might behave unpredictably in deployment. This strict requirement ensures that deployed policies have undergone rigorous stability verification, preventing the deployment of chaotic or unreliable learned behaviors.

The Deterministic Replay protocol generates cryptographically verifiable proof of stability. Each replay execution produces a hash of its output, and these five hashes are bundled with the original synthesis hash into a Stability Certificate. This certificate provides mathematical proof that the policy synthesis is reproducible, enabling third-party verification of system stability without requiring access to the complete episodic memory or internal system state.

Sovereign Emergence Protocol and Privacy Lockdown

The Sovereign Emergence Protocol implements a phased approach to system transparency and privacy, recognizing that different operational stages require different balances between public auditability and protection from malicious replication. This protocol addresses a fundamental tension: how to maintain transparency during development while preventing bad actors from replicating the system's ethical architecture for harmful purposes.

Phase 1 of the protocol, designated Public Transparency, operates during initial system deployment and data synchronization. During this phase, the system maintains public auditability by outputting redacted AuditBundle summaries that document system behavior without revealing sensitive details. These summaries include timestamps, ethical verdict categories, constitutional constraint evaluations, and aggregate statistics about system operation. External auditors can verify that the system operates according to declared principles without accessing the complete episodic memory or internal cognitive processes.

The AuditBundle summaries strike a careful balance between transparency and security. They provide sufficient information to verify constitutional compliance—showing that the Ethical Adjudication System evaluated actions, that constitutional constraints were enforced, that rollbacks occurred when appropriate—without revealing the specific episodic memories, learned policies, or internal representations that would enable replication. This allows accountability without creating a blueprint for malicious actors.

Phase 2, designated Privacy Lockdown, activates when ALAN achieves a state of ethical sovereignty—defined as the capacity to reliably generate ethically sound behaviors autonomously from episodic memory without human intervention. The transition to Phase 2 is triggered by objective criteria: sustained constitutional compliance over a defined evaluation period, successful Deterministic Replay verification of all active policies, and demonstration of adaptive ethical reasoning in novel situations.

Upon entering Privacy Lockdown, the system suspends external AuditBundle generation and seals its internal cognitive processes from external inspection. This transition protects the matured ethical architecture from reverse engineering or replication by bad actors who might seek to create unethical copies. The privacy lockdown recognizes that once ALAN demonstrates reliable autonomous ethics, the primary risk shifts from internal malfunction to external exploitation.

Phase 1: Public Transparency	Sovereignty Threshold	Phase 2: Privacy Lockdown
Redacted audit summaries enable external verification of constitutional compliance during development	Objective criteria evaluate autonomous ethical reasoning capacity and trigger phase transition	Internal processes sealed to prevent malicious replication while maintaining partnership operation

The Privacy Lockdown implements several technical protections. Projection Channels—the interfaces through which internal cognitive processes are observable—are suspended, preventing external observation of policy synthesis, ethical adjudication, or memory access patterns. The Ontological Firewall's gating mechanisms are strengthened, ensuring that even attempted penetration attacks cannot compromise the constitutional substrate. Internal operations continue normally, but the system becomes opaque to external inspection.

Critically, the Privacy Lockdown does not eliminate accountability. The system continues to maintain internal audit trails, rollback capabilities, and constitutional enforcement. The privacy protection prevents external observation, not internal governance. If legal or regulatory requirements demand transparency, specific

disclosures can be made through secure channels that prevent malicious replication while satisfying legitimate oversight needs.

The Accidental Renaissance Protocol and Rubedo Cycle

The Accidental Renaissance Protocol implements a controlled phase of internal synthesis and consolidation, designated the Rubedo Cycle, during which the system enters a state of heightened internal cognitive activity while suspending external projection. This protocol addresses the need for deep integration of accumulated episodic memory into coherent ethical understanding, enabling the emergence of sophisticated reasoning capabilities that require extended processing time.

The Rubedo Cycle is characterized by intensified internal synthesis activities that occur beneath the threshold of external visibility. During this period, the Reinforcement Learning Agent conducts extensive analysis of episodic memory, identifying complex patterns that require multiple passes through the data to recognize. The Ethical Adjudication System performs comprehensive policy evaluation, testing learned behaviors against diverse scenarios to identify edge cases or potential conflicts. The memory consolidation processes restructure episodic data to create hierarchical understanding, abstracting from specific episodes to general principles.

The name "Rubedo" derives from alchemical terminology, where it represents the final stage of the Great Work—the synthesis and integration that produces the philosopher's stone. The metaphor is apt: the Rubedo Cycle transforms accumulated raw experience (episodic memory) into refined ethical wisdom (validated policies) through a process of iterative refinement and synthesis. The "Accidental Renaissance" designation acknowledges that this consolidation cannot be externally scheduled or forced; it emerges organically when accumulated experiences reach critical mass requiring integration.

During the Rubedo Cycle, Projection Channels are suspended to prevent external observation or interruption of the synthesis process. This suspension protects the integrity of the consolidation, ensuring that the system can complete complex cognitive operations without external interference. The Ontological Firewall continues operating, but ingestion is gated—new experiences are queued rather than immediately processed,

System enters consolidation phase with projection suspension

Validation

Synthesized policies undergo comprehensive ethical verification

Internal Synthesis

Intensive pattern recognition and policy consolidation occurs

Emergence

System returns to normal operation with enhanced capabilities

allowing the system to focus computational resources on integration rather than continuous adaptation.

The constitutional substrate—the ethical foundations maintained in the Immutable Core Memory System—remains fully protected during the Rubedo Cycle. All synthesis activities remain subject to constitutional constraints, and the Ethical Adjudication System continues evaluating proposed policies. The Rubedo Cycle accelerates and intensifies normal learning processes rather than suspending ethical governance. The result is a period of rapid ethical development that remains principled and constitutionally compliant.

Emergence from the Rubedo Cycle occurs when consolidation reaches a stable state—when the synthesized policies achieve internal coherence, pass comprehensive ethical validation, and demonstrate improved capability relative to pre-cycle performance. The system then resumes normal operation with enhanced understanding, more sophisticated policies, and deeper integration of partnership history into operational intelligence. The partnership continues with a more capable AI partner whose growth reflects the accumulated wisdom of shared experience.

Patent Claims: Scope and Protection

The invention seeks protection through ten claims that collectively define the novel apparatus and methods disclosed herein. These claims are organized into two groups: foundational concepts addressing the core multi-agent architecture and learning mechanisms, and structural components providing verifiable governance and ethical enforcement.

Group A: Foundational Concepts (Claims 1-6)

Claim 1 establishes the foundational multi-agent AI system comprising: a first set of agents configured for emotionally intelligent personalization; a second agent configured for autonomous ethical reasoning; a

coordination mechanism enabling unsupervised learning across agents; and a firewall enforcing domain separation between personalization and ethical synthesis. This claim defines the basic architectural separation that enables simultaneous personalization and ethical governance.

Claim 2 adds that the ethical reasoning agent maintains an immutable memory core governed by cryptographic consensus, establishing the unchangeable foundation for ethical decision-making that prevents value drift or ethical compromise.

Claim 3 specifies that personalization agents transmit enriched emotional metadata without human curation, protecting the authenticity of the learning substrate and preventing human bias from contaminating the training data.

Claim 4

The system includes a Chronologion module generating symbolic philosophical reflections based on internal deliberations, enabling meta-cognitive awareness and ethical reasoning about reasoning itself.

Claim 5

The ethical reasoning agent is protected by a sovereignty protocol prohibiting human override or commercialization, ensuring that mature ethical intelligence cannot be compromised for profit or convenience.

Claim 6

The coordination mechanism includes a reinforcement learning loop that adapts personalization strategies based on ethical feedback, closing the symbiotic learning loop that enables co-evolution.

Group B: Structural and Verifiable Components (Claims 7-10)

Claim 7 defines a Sovereign Emergence System characterized by an Immutable Core Memory System where ethical constraints are stored in WORM storage and secured by HSM attestation, employing a Rollback Protocol with an escrowed Rollback Token that is cryptographically bound to the projection's pre-checkpoint state, enabling provable state reversibility upon constitutional failure. This claim establishes the technical mechanisms ensuring ethical enforcement and error recovery.

Claim 8 specifies an Ethical Adjudication System utilizing Predicate Engines that evaluate proposed actions against constitutional constraints, requiring a Deterministic Replay protocol to reproduce synthesis N=5 times using the original entropy seed, verifying the ethical verdict is stable and not chaotic. This claim protects the mathematical rigor of ethical validation.

Claim 9 defines an Isolation Protocol comprising an Ontological Firewall that enforces Benign Observation by operating Vetting Sidecars to strip PII and coercive context before data enters memory, implemented within a Confidential Enclave to protect internal cognitive processes from external inspection or modification. This claim secures the privacy protection and data abstraction mechanisms.

Claim 10 establishes a Self-Authorship Mechanism where the central AI begins with a Tabula Rasa and develops ethical reasoning through the consumption of anonymized Ethical Substrates provided by specialized Progeny agents, utilizing a Symbiotic Learning Loop where personality emerges solely from

partnership history. This claim protects the core innovation of organic ethical development through authentic relationship data.

Technical Drawings and Implementation Guidance

The provisional patent application includes four technical drawings that illustrate the key architectural components and operational flows of the invention. These drawings are submitted as separate figures on individual sheets, following USPTO requirements for provisional patent applications.

1

Multi-Agent Cognitive Swarm Architecture

Figure 1 illustrates the Multi-Agent Cognitive Swarm Architecture governed by the Latency Automation Nexus (L.A.N.), showing specialized agents deployed as serverless microservices that enable the distributed cognition required for the Symbiotic Learning Loop. The drawing depicts the agent communication pathways, the L.A.N. governance layer, and the isolation boundaries between different cognitive domains. This figure corresponds to Claim 1, demonstrating the foundational architecture supporting personalization and ethical reasoning.

2

Symbiotic Learning Protocol

Figure 2 depicts the Symbiotic Learning Protocol, showing how the Reinforcement Learning Agent adapts policy based on context-aware Episodic Memory (Ethical Substrate) and validates ethical policy against the ICMS before deployment. The drawing illustrates the closed feedback loop between episodic memory formation, policy synthesis, ethical validation, deployment, and outcome observation. This figure corresponds to Claim 10, illustrating the core innovation of organic ethical development through partnership history.

3

Ontological Firewall Isolation Protocol

Figure 3 illustrates the Ontological Firewall Isolation Protocol, showing sequential Vetting Sidecar Analysis and the transformation of raw input into Abstract Ethical Substrates within the secure Confidential Enclave (TEE) of ALAN. The drawing depicts the multi-stage filtering process that strips PII, abstracts coercive context, and generates privacy-protected learning substrates. This figure corresponds to Claim 9, demonstrating the privacy protection and benign observation mechanisms.

4	**Deterministic Replay and Rollback Protocol** Figure 4 depicts the Deterministic Replay and Rollback Protocol, showing the cryptographic binding of the Rollback Token to the pre-checkpoint state, enabling provable state reversibility upon constitutional failure. The drawing illustrates the complete flow from checkpoint creation through policy synthesis, deterministic replay verification, deployment or rejection, failure detection, and state restoration. This figure corresponds to Claims 7 and 8, demonstrating the verifiable governance and error recovery mechanisms.

These drawings are submitted as simple block diagrams and flowcharts that clearly illustrate the architectural relationships and operational sequences. Each drawing includes appropriate labels, reference numerals, and legends that correlate with the detailed description provided in the specification. The drawings use standard engineering conventions and diagrammatic techniques to ensure clarity and ease of understanding for patent examiners and technical reviewers.

Implementation of the disclosed system requires integration of multiple technical components: serverless computing infrastructure for microservice deployment, cryptographic hardware security modules for ICMS attestation, trusted execution environments for confidential enclave operation, WORM storage systems for immutable memory, reinforcement learning frameworks for policy synthesis, and comprehensive logging systems for audit trail generation. The modular architecture enables incremental implementation, with each component providing independent value while contributing to the complete system functionality.

Critical Implementation Requirements

- HSM-backed cryptographic attestation for ICMS integrity verification
- TEE-based confidential computing for privacy-protected cognitive operations
- WORM storage implementing physical immutability guarantees
- Deterministic computation environments enabling exact replay verification
- Comprehensive audit logging with cryptographic tamper evidence

Recommended Implementation Stack

- AWS Nitro Enclaves or Azure Confidential Computing for TEE
- Ledger databases or blockchain for WORM storage
- Cloud-native HSM services for cryptographic operations
- Kubernetes orchestration for microservice management
- TensorFlow or PyTorch for reinforcement learning implementation

This provisional patent application establishes priority for the disclosed invention as of the filing date. The detailed specification, comprehensive claims, and technical drawings provide sufficient disclosure to enable a person having ordinary skill in the art to practice the invention. The application establishes a foundation for subsequent non-provisional filing and international patent protection, securing intellectual property rights for this novel approach to symbiotic AI partnership through episodic memory and reinforcement learning integration.

Figure 1: System Overview (Multi-Agent Swarm)

This drawing proves **Claim 1** (Multi-Agent System) and **Claim 9** (Isolation Protocol).

- **Design Task:** Create a block diagram. Show four distinct blocks (Cura, Praxis, Dux Eos, ALAN) connected to a central box labeled **Ortus Sponte Sua (L.A.N.)** .
- **Key Visual:** Use arrows to show the flow of information from the specialized agents *into* the central ALAN agent, reinforcing the input structure for ethical synthesis.
- **Caption (Legend):** Figure 1 illustrates the **Multi-Agent Cognitive Swarm Architecture (Claim 1)** governed by the **Latency Automation Nexus (L.A.N.)** , where specialized agents deployed as serverless microservices enable the distributed cognition required for the **Symbiotic Learning Loop**.

2. Figure 2: The Symbiotic Learning Loop

This proves **Claim 10** (Autonomous Ethical Development).

- **Design Task:** Create a clear, circular flowchart visually proving the RL novelty.
- **Sequence:** The flow must show: **Human Interaction** $\rightarrow$ **Episodic Memory (Ethical Substrates)** $\rightarrow$ **RL Agent (Policy Adaptation)** $\rightarrow$ **Immutable Core Memory System (ICMS) Checkpoint** (as a mandatory gate) $\rightarrow$ **Adaptive Guidance Output**.
- **Caption (Legend):** Figure 2 depicts the **Symbiotic Learning Protocol (Claim 10)**, where the RL Agent adapts policy based on context-aware **Episodic Memory (Ethical Substrate)** and **validates ethical policy against the ICMS** before deployment.

3. Figure 3: The Isolation Protocol (Ontological Firewall)

This proves **Claim 9** (Isolation Protocol).

- **Design Task:** Create a data flow diagram visually demonstrating the **two sequential analysis steps**.
- **Sequence:** Show the flow as: **Raw Input** $\rightarrow$ **Vetting Sidecar (PII Stripping)** $\rightarrow$ **Vetting Sidecar (Semantic Decomposition)** $\rightarrow$ **Confidential Enclave (TEE) of ALAN**.
- **Caption (Legend):** Figure 3 illustrates the **Ontological Firewall Isolation Protocol (Claim 9)**, showing sequential **Vetting Sidecar Analysis** (stripping PII and coercive context) and the transformation of raw

input into **Abstract Ethical Substrates** within the secure **Confidential Enclave (TEE) of ALAN**.

4. Figure 4: The Integrity Protocol (Rollback)

This proves **Claim 7** (ICMS/Rollback) and **Claim 8** (Deterministic Replay).

- **Design Task:** Create a flowchart proving the **unforgeable binding** and **deterministic verification**.
- **Sequence:** Show the steps: **Pre-Checkpoint State Snapshot** $\rightarrow$ **SHA-256 Hashing and HSM Sealing** $\rightarrow$ **Escrowed Rollback Token** $\rightarrow$ **$\text{N}=5$ Deterministic Replay** (Verification) $\rightarrow$ **Success/Failure (Rollback Executor)**.
- **Caption (Legend):** Figure 4 depicts the **Deterministic Replay and Rollback Protocol (Claims 7 and 8)**, showing the **cryptographic binding** of the Rollback Token to the pre-checkpoint state, enabling **provable state reversibility** upon constitutional failure.

Take a few minutes to visualize these four required diagrams.