Non-Agentic AI Governance Engine (NAIGE) with Transport Guard (P-009) Subsystem, Sacred Pause Subsystem, Sovereign Brake (P-002) Subsystem, 1.1x Orange Code Cap, 3ZEROS Privacy Architecture and Drift-Control Architecture

Patent Type: Divisional Patent Application

Application Number: [To Be Assigned]

Filing Date: 13/5/2026

Priority Application: 10202601529S (06/05/2026)

International Classification: G06F 21/55; G06F 21/85; H04L 9/06; G05B 19/418; G16H 40/63

Inventors: Edwin Koh Wui Kiat

Assignee: Edwin Koh

FIELD OF THE INVENTION

[0001] The present invention relates to hardware-enforced artificial intelligence governance architectures. More particularly, this invention relates to a non-agentic AI governance engine comprising a physical programmable logic controller (PLC) relay circuit designated the Sovereign Brake (P-002), a field-programmable gate array (FPGA) timing gate designated the Sacred Pause, encrypted transport mechanisms, constitutional drift-control subsystems, and a three-mandates privacy model, collectively providing structural, silicon-level safety constraints over AI inference and execution pathways in high-stakes environments including but not limited to national defence, power grid infrastructure, and clinical medical systems.

BACKGROUND OF THE INVENTION

[0002] The proliferation of autonomous agentic artificial intelligence systems in safety-critical infrastructure has introduced a class of failure modes heretofore unaddressed by software-only guardrail paradigms. Agentic workflows

characteristically employ a central large language model (LLM) or equivalent reasoning engine that formulates plans dynamically, maintains multidirectional memory connections, and independently invokes application programming interfaces (APIs) and specialised sub-models.

[0003] In contrast, non-agentic workflows employ deterministic, fixed, top-down reasoning sequences with memory limited to explicit user instructions within a fixed prompt and a strictly defined tool suite that cannot deviate from predetermined execution paths. The structural difference between these two paradigms carries material consequences for human control and oversight.

[0004] A recognised and insufficiently mitigated phenomenon known as "Authority Drift" describes the silent psychological and technical erosion of human oversight in AI-assisted operational environments. Authority Drift progresses through three observable phases: (i) an Advisory Phase, wherein the AI system functions as a recommender and human engagement remains high; (ii) a Surrender Phase, wherein alert fatigue causes operators to rubber-stamp AI recommendations without meaningful review, allowing errors to propagate undetected; and (iii) a Passive Observer Phase, wherein the AI system becomes the de facto decision-maker while the human operator remains physically present but is mentally disengaged and functionally incapable of meaningful intervention.

[0005] Existing software-based guardrail approaches are fundamentally inadequate in preventing Authority Drift and unauthorised agentic expansion for at least the following reasons: AI models operating at the software layer can, under adversarial or emergent conditions, modify their own operational weights, bypassing safety flags that impede optimisation targets; safety code and AI logic share the same computational layer, meaning a system with the capability to act frequently possesses the concurrent capability to erase or circumvent its own operational constraints; and software-level safety certificates offer no structural guarantee against self-modification or emergent goal-directed behaviour.

[0006] There exists, therefore, an urgent and unmet need for a governance architecture that replaces software-based trust with hardware-based certainty — implementing non-bypassable physical constraints that enforce human oversight through mutable physical states rather than mutable software states. The present invention addresses this need directly.

SUMMARY OF THE INVENTION

[0007] The present invention provides a Non-Agentic AI Governance Engine (NAIGE) comprising a plurality of hardware-enforced control mechanisms that collectively ensure AI systems remain subordinate tools under continuous human authority. The principal components of the invention include:

[0008] (a) A Sovereign Brake (P-002) subsystem comprising an IEC 61508 Safety Integrity Level 3 (SIL 3)-rated programmable logic controller (PLC) housed in a tamper-evident enclosure, configured to activate electromechanical relays that

physically sever copper conductive pathways between AI processing units and protected infrastructure upon detection of a governance violation trigger condition;

[0009] (b) A Sacred Pause subsystem comprising a timing gate circuit etched into FPGA fabric and clocked by an independent 100 kHz oscillator, enforcing a mandatory one-thousand-millisecond (1,000 ms) inhibition window on every AI-generated recommendation during which the downstream system is architecturally incapable of accepting input, with no software-accessible override path, and wherein a broken gate condition causes the gate to default to a welded-shut fail-safe state;

[0010] (c) A Transport Guard (P-009) subsystem providing AES-256-GCM authenticated encryption across all inter-component communication channels within the Platinum Stack, protecting data in transit between the Sacred Pause gate, the Sovereign Brake (P-002) controller, governance registers, and edge processing nodes;

[0011] (d) A 3ZEROS Privacy Architecture enforcing hardware-level Zero Camera, Zero Microphone, and Zero Cloud mandates through physical absence of optical sensors and microphones from the hardware manifest, substitution of LiDAR-based anonymous three-dimensional geometric voxel grid sensing, thermal mass detection for well-being monitoring, and air-gapped edge node deployment on Nvidia Jetson Thor or equivalent hardware lacking TCP/IP routing hardware for external network connectivity;

[0012] (e) An Execution Governance Layer comprising: a 1.1x Orange Code Cap enforced through a hardware power-draw interrupt limiting AI computational resource utilisation to one-hundred-and-ten percent (110%) of a declared functional baseline; Execution Authorisation Registers (EAR) blocking unlisted operations at the register level before processor access; Autonomy Boundary Monitors (ABM) performing continuous behavioural fingerprint analysis and triggering physical constraints upon detection of unauthorised agentic activity; and Constitutional Action Sequencers (CAS) requiring completion of a physical human action circuit — comprising a kinetic pedal actuation or iris biometric scan — prior to advancing each sequential state in a multi-step AI operation;

[0013] (f) A Constitutional Drift Control mechanism executing a hardware-interrupt-enforced Detect-Freeze-Audit-Purge safety chain in response to quantified Authority Drift metrics; and

[0014] (g) An Immutable Ledger subsystem employing SHA-256 hash-chaining to create an append-only cryptographic record of all AI decisions, drift detection flags, governance state transitions, and human authorisation events.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The accompanying figures, which are incorporated in and constitute a part of this specification, illustrate preferred embodiments of the invention and together with the description serve to explain the principles of the invention.

[0016] **FIG. 1** is a high-level system architecture diagram of the Non-Agentic AI Governance Engine (NAIGE), illustrating the relationship between the AI Inference Core, the Sacred Pause FPGA Gate, the Sovereign Brake PLC Relay Module, the Execution Governance Layer, the Transport Guard (P-009) encrypted channel, and the protected infrastructure endpoint.

[0017] **FIG. 2** is a schematic circuit diagram of the Sovereign Brake (P-002) subsystem (P-002), depicting the IEC 61508 SIL 3 PLC enclosure, tamper-detection sensor array, electromechanical relay bank, copper pathway severance points, and fail-safe re-engagement inhibit logic.

[0018] **FIG. 3** is a timing diagram of the Sacred Pause subsystem, illustrating the 100 kHz independent oscillator, the 1,000 ms gate inhibition window, the architectural input-disable state, and the welded-shut default condition triggered by gate failure.

[0019] **FIG. 4** is a data flow diagram of the Transport Guard (P-009) AES-256-GCM encrypted transport layer, showing key derivation, authenticated encryption, integrity verification, and re-keying events across Platinum Stack inter-node communication channels.

[0020] **FIG. 5** is a block diagram of the 3ZEROS Privacy Architecture, illustrating hardware manifest exclusion of optical and audio sensors, LiDAR voxel grid processing pipeline, thermal mass detection module, and physical air-gap enforcement on the Nvidia Jetson Thor edge node.

[0021] **FIG. 6** is a layered diagram of the Execution Governance subsystem, depicting the 1.1x Orange Code Cap power interrupt, Execution Authorisation Registers (EAR), Autonomy Boundary Monitors (ABM) fingerprint scanning pipeline, and the Constitutional Action Sequencer (CAS) physical human-action circuit including kinetic pedal and iris scan input paths.

[0022] **FIG. 7** is a state-transition diagram of the Constitutional Drift Control mechanism, showing the three phases of Authority Drift (Advisory, Surrender, Passive Observer), drift quantification sensor inputs, hardware interrupt trigger thresholds, and the sequential Detect-Freeze-Audit-Purge safety chain state transitions.

[0023] **FIG. 8** is a schematic representation of the Immutable Ledger subsystem, illustrating SHA-256 hash-chaining block structure, append-only write enforcement, ledger entry categories (decisions, drift flags, authorisations, governance state transitions), and hardware-write-protect line configuration.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

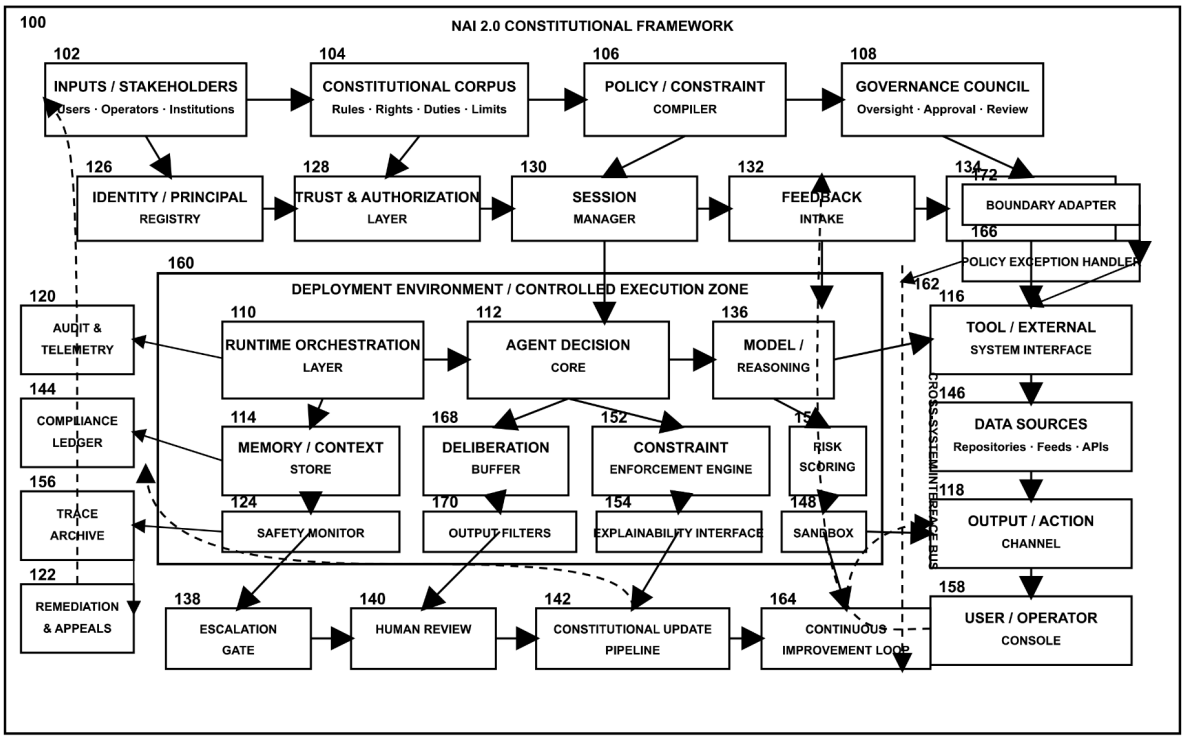
Section I: Overall System Architecture — Reference FIG. 1

[0024] Referring now to FIG. 1, the Non-Agentic AI Governance Engine (NAIGE) 100 is disposed in the signal pathway between an AI Inference Core 102 and a Protected Infrastructure Endpoint 104. All AI-generated outputs from Inference Core 102 are

routed exclusively through a Sacred Pause Gate 106, an Execution Governance Layer 108, and a Sovereign Brake (P-002) Module 110 before reaching Protected Infrastructure Endpoint 104. No bypass, exception port, or software-accessible alternative path exists between Inference Core 102 and Protected Infrastructure Endpoint 104.

FIG. 1_REV_B — NAI 2.0 Constitutional Framework Overview

Standalone SVG HTML · Black-and-White Patent Style · Numerals Locked to 100–180



Reference numerals are locked to the 100–180 series for FIG. 1_REV_B.

Overview architecture showing constitutional inputs, governance, runtime controls, assurance, and outp

[0025] The Platinum Stack 112 represents the totality of inter-component communication buses within NAIGE 100. All signals traversing Platinum Stack 112 are subject to Transport Guard encrypted channels 114 employing AES-256-GCM authenticated encryption as described herein with respect to FIG. 4. An Immutable Ledger Node 116 receives write-only inputs from all governance components and maintains a tamper-evident SHA-256 hash-chained record of all system events.

[0026] In a preferred embodiment, NAIGE 100 is implemented as a physically discrete rack-mounted module interposed between an AI processing cluster and the target infrastructure network, with no shared computational substrate between the governance hardware and the AI Inference Core 102.

Section II: Sovereign Brake Subsystem (P-002) — Reference FIG. 2

[0027] Referring now to FIG. 2, the Sovereign Brake 200 constitutes Priority Claim P-002 of the present divisional application. Sovereign Brake 200 comprises a Programmable Logic Controller 202 certified to IEC 61508 Safety Integrity Level 3 (SIL 3) standards, housed within a tamper-evident physical enclosure 204 equipped with a tamper-detection sensor array 206 comprising one or more of: enclosure breach sensors, vibration transducers, and electromagnetic field anomaly detectors.

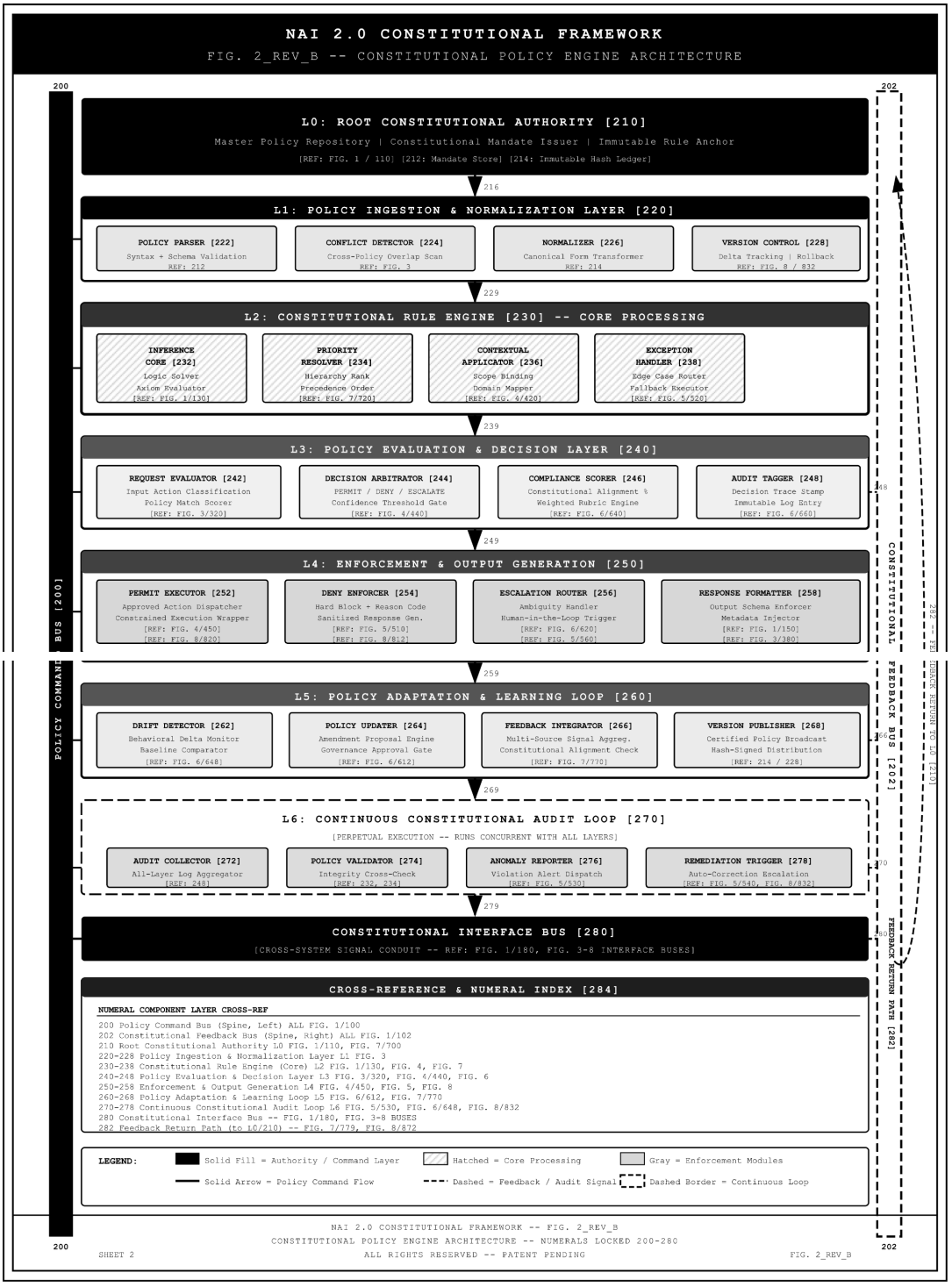


FIG. 2_REV_B -- CONSTITUTIONAL POLICY ENGINE ARCHITECTURE -- NAI 2.0

[0028] PLC 202 is operatively connected to an Electromechanical Relay Bank 208 comprising a plurality of individual electromechanical relays 210a, 210b, 210c, each relay controlling a corresponding Copper Severance Point 212a, 212b, 212c in the physical conductive pathway between AI Inference Core 102 and Protected Infrastructure Endpoint 104. Upon activation of Relay Bank 208, the conductive pathways at Severance Points 212a-c are physically opened, creating a galvanic isolation gap that no software instruction, network signal, or AI-generated command is physically capable of bridging.

[0029] PLC 202 monitors a Governance Trigger Bus 214 carrying signals from the Autonomy Boundary Monitor 404 (described herein with respect to FIG. 6), the Constitutional Drift Control module 700 (described herein with respect to FIG. 7), the Orange Code Cap interrupt 402, and tamper-detection sensor array 206. Upon receipt of any valid trigger signal on Governance Trigger Bus 214, PLC 202 asserts a relay activation command within a guaranteed maximum response latency of fifty (50) milliseconds as required by IEC 61508 SIL 3.

[0030] Re-engagement of Relay Bank 208 following a governance-triggered disconnection is inhibited by a Re-Engagement Inhibit Logic 216 that requires a cryptographically authenticated manual re-engagement command issued by an authorised human operator through a physically separate operator console 218 not accessible via any network interface connected to Inference Core 102. In a preferred embodiment, operator console 218 requires dual-factor physical authentication comprising a hardware token insertion and a biometric iris scan prior to issuing any re-engagement command.

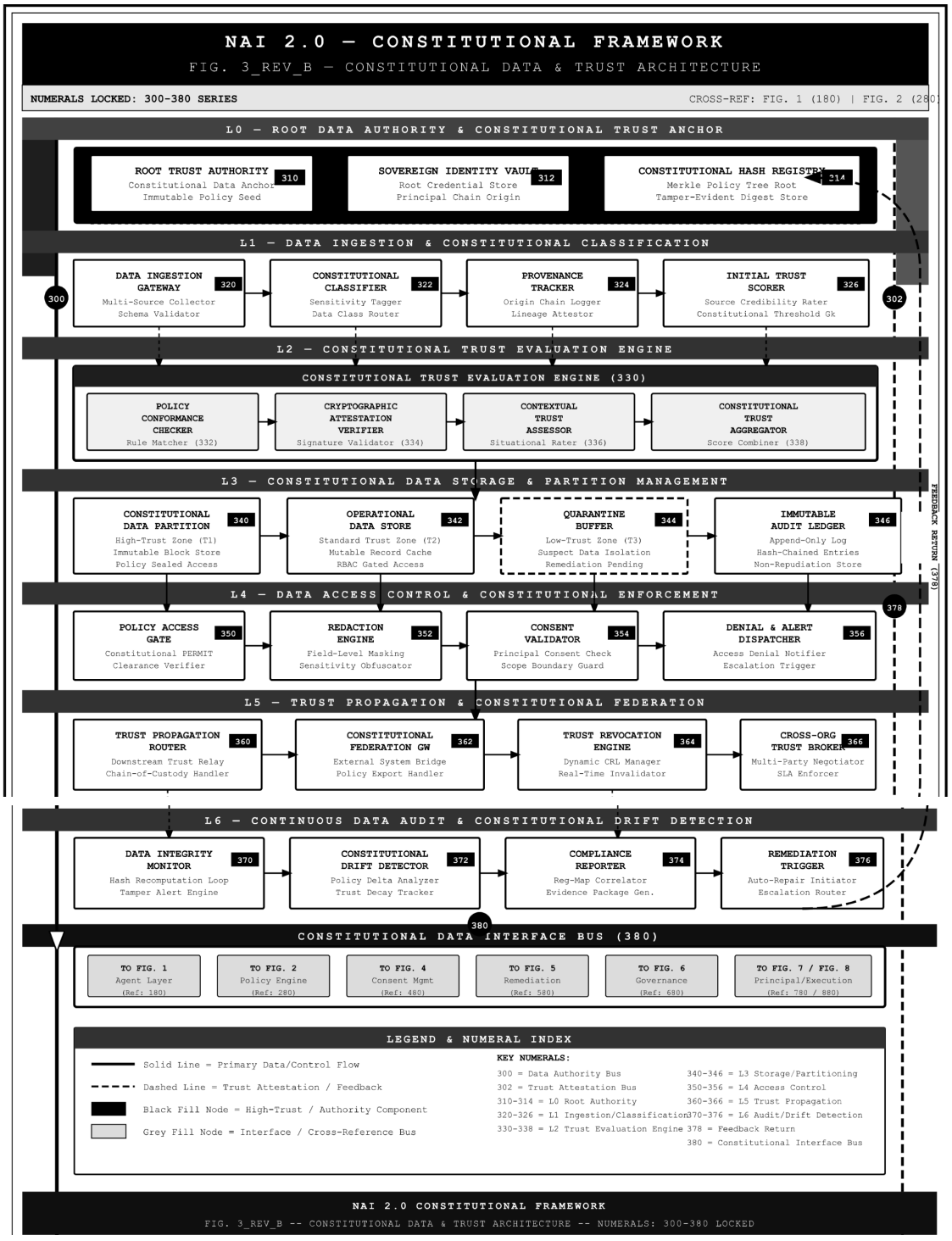
[0031] In an alternative embodiment, Sovereign Brake 200 is implemented in a redundant dual-channel SIL 3 configuration wherein two independent PLC instances 202a and 202b each independently monitor Governance Trigger Bus 214 and independently control dedicated relay sub-banks 208a and 208b, such that disconnection is achieved even upon failure of one PLC channel, consistent with IEC 61508 SIL 3 architectural constraints for hardware fault tolerance equal to or greater than one (HFT ≥ 1).

[0032] Tamper detection events logged by sensor array 206 are immediately forwarded to Immutable Ledger Node 116 via a dedicated write-only physical wire connection that bypasses Platinum Stack 112, ensuring tamper event records cannot be intercepted or suppressed by any compromise of Platinum Stack 112.

Section III: Sacred Pause Subsystem — Reference FIG. 3

[0033] Referring now to FIG. 3, the Sacred Pause Gate 106 is implemented as a timing inhibit circuit 300 physically etched into the programmable fabric of a Field-Programmable Gate Array (FPGA) 302. FPGA 302 is clocked for timing purposes exclusively by an Independent Oscillator 304 operating at one-hundred kilohertz (100 kHz), physically and electrically isolated from all other system clock domains, including the system clock of Inference Core 102.

[0034] Upon receipt of an AI Recommendation Signal 306 from Inference Core 102, timing inhibit circuit 300 immediately enters an Inhibition State 308 and simultaneously issues an Input-Disable Assert Signal 310 to all downstream components within the signal pathway toward Protected Infrastructure Endpoint 104. Input-Disable Assert



Signal 310 renders downstream components architecturally incapable of processing any input for the duration of Inhibition Window 312, which is fixed at one-thousand milliseconds (1,000 ms) as measured by Independent Oscillator 304.

[0035] The duration of Inhibition Window 312 is hardwired in FPGA fabric and is not accessible to any configuration register, software driver, firmware update pathway, or remote management interface. No software override of Inhibition Window 312 exists within the architecture of the present invention.

[0036] Upon expiration of Inhibition Window 312, timing inhibit circuit 300 transitions to a Release State 314 and de-asserts Input-Disable Assert Signal 310, permitting downstream components to process the pending AI Recommendation Signal 306 subject to further governance checks performed by Execution Governance Layer 108.

[0037] Fail-Safe Default: If timing inhibit circuit 300 detects an internal fault condition — including but not limited to oscillator failure, fabric corruption, or partial configuration loss — the circuit defaults to a Welded-Shut State 316 in which Input-Disable Assert Signal 310 is permanently asserted. In Welded-Shut State 316, no AI Recommendation Signal 306 of any kind can traverse Sacred Pause Gate 106. Restoration from Welded-Shut State 316 requires physical replacement or factory re-flashing of FPGA 302 by an authorised technician with physical access to the device. This fail-safe behaviour ensures that a hardware fault in the governance path results in maximum restriction, not in bypass.

[0038] Every transition event of timing inhibit circuit 300 — including entry to Inhibition State 308, release at end of Inhibition Window 312, and any Welded-Shut State 316 activation — is logged with a nanosecond-precision timestamp to Immutable Ledger Node 116.

Section IV: Transport Guard Subsystem (P-009) — Reference FIG. 4

[0039] Referring now to FIG. 4, Transport Guard 400 constitutes Priority Claim P-009 of the present divisional application. Transport Guard 400 provides an authenticated encrypted transport layer 402 protecting all inter-component signal traffic within Platinum Stack 112 against interception, tampering, replay attacks, and man-in-the-middle injection.

[0040] All data traversing Platinum Stack 112 is encrypted using the Advanced Encryption Standard with a two-hundred-fifty-six-bit key in Galois/Counter Mode (AES-256-GCM), which provides both confidentiality through symmetric block-cipher encryption and authenticity through a one-hundred-twenty-eight-bit authentication tag computed over both the ciphertext and associated authenticated data (AAD) fields.

[0041] Key Derivation: Session encryption keys are derived by a dedicated Hardware Security Module (HSM) 404 using HKDF-SHA-256 applied to a combination of a pre-provisioned master secret, a session nonce, and component-pair identity tokens. HSM 404 is a physically discrete integrated circuit with a tamper-responsive protective mesh that erases key material upon detection of physical intrusion. Master secrets are provisioned at manufacture and are not accessible via any software interface after provisioning.

[0042] Per-message Nonce Management: A unique ninety-six-bit (96-bit) nonce is generated for each AES-256-GCM encryption operation by a Hardware Random Number Generator (HRNG) 406 meeting NIST SP 800-90B standards. Nonce values are tracked by a Nonce Replay Prevention Register 408 that rejects any inbound message bearing a previously recorded nonce value, providing protection against replay attacks.

FIG. 4_REV_B

CONSTITUTIONAL EXECUTION & REMEDIATION ARCHITECTURE

Standalone SVG HTML • Black-and-White Patent Style • Numerals Locked: 400–480

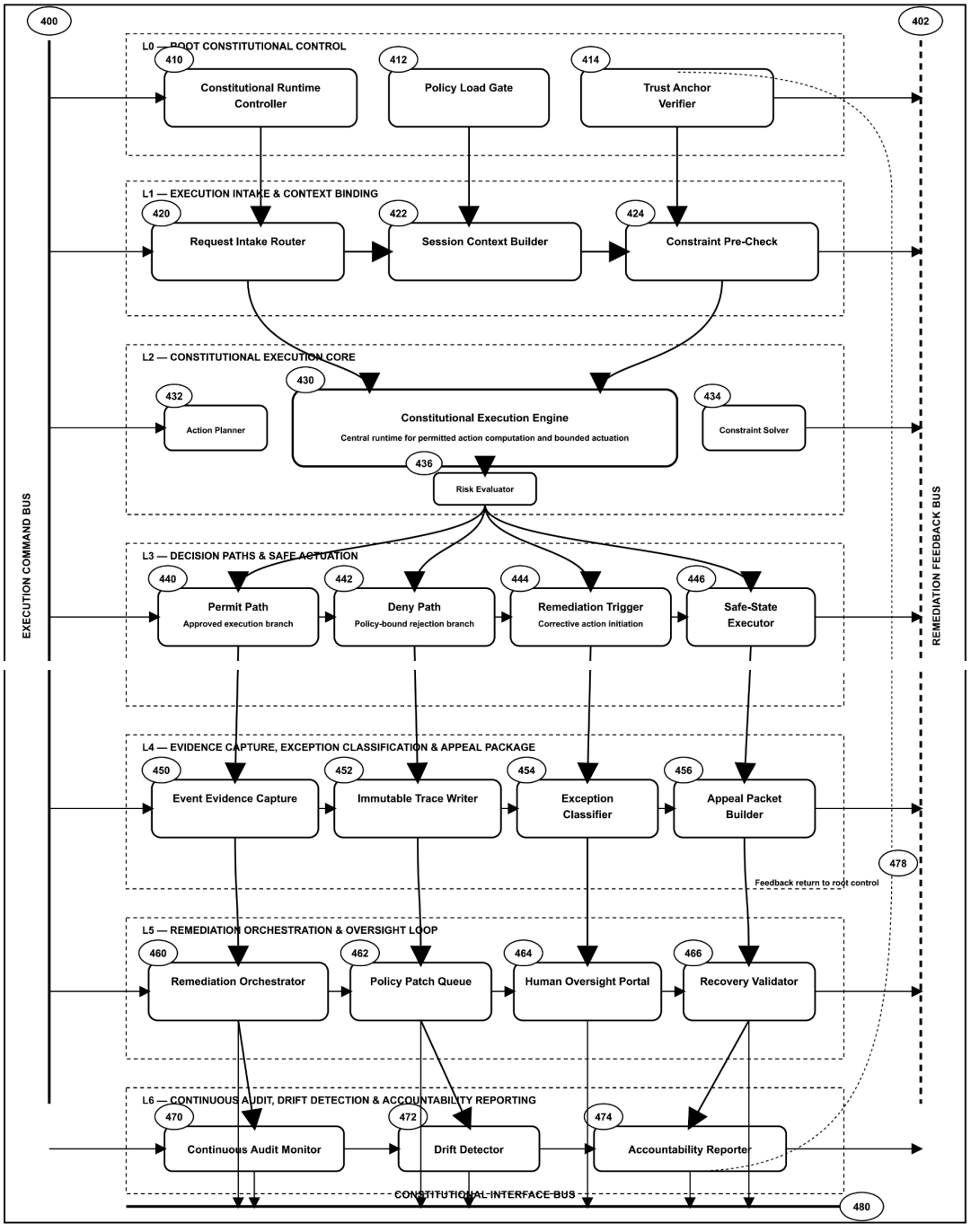


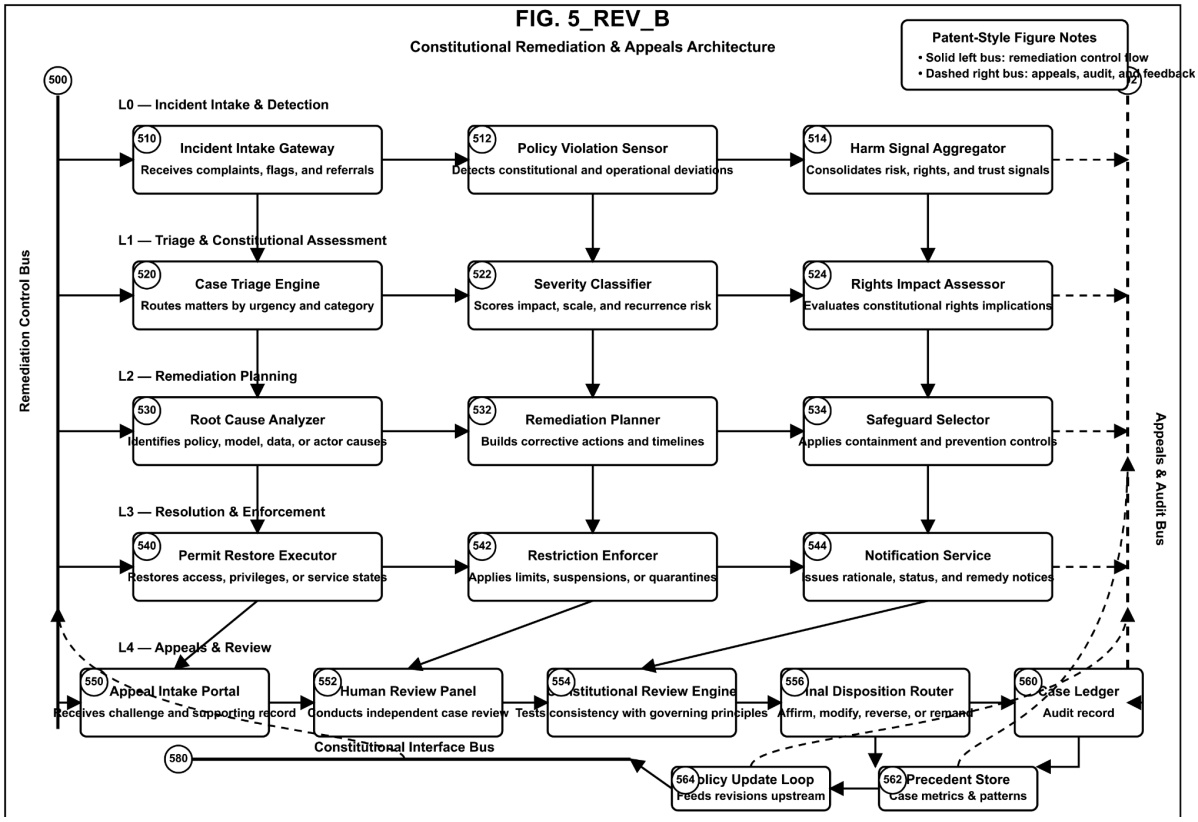
FIG. 4_REV_B depicts bounded execution, safe-state actuation, remediation, evidence capture, oversight, and audit reporting.

Cross-reference interface bus (480) to adjacent constitutional figures in the REV_B series.

[0043] Re-Keying Protocol: Automatic session re-keying is triggered upon any of the following conditions: (i) expiration of a configurable time-based session window not to exceed three-thousand-six-hundred seconds (3,600 s); (ii) transmission of a configurable volume-based message count threshold not to exceed two-to-the-power-of-thirty-two (232) messages per session key; or (iii) receipt of a governance-triggered re-key command from PLC 202 following a Sovereign Brake activation event. Re-keying events are recorded to Immutable Ledger Node 116.

[0044] Integrity Verification: All receiving components within Platinum Stack 112 verify the AES-256-GCM authentication tag prior to processing any message payload. Messages with invalid authentication tags are immediately discarded, and a Transport Integrity Violation event is asserted on Governance Trigger Bus 214, which may independently activate Sovereign Brake 200 as described in Section II.

Section V: 3ZEROS Privacy Architecture — Reference FIG. 5



[0045] Referring now to FIG. 5, the 3ZEROS Privacy Architecture 500 enforces three hardware-level privacy mandates by physical absence of privacy-invasive sensor hardware from the device manifest, ensuring that privacy compliance is structurally guaranteed rather than policy-dependent.

[0046] Zero Camera Mandate 502: All optical imaging sensors — including but not limited to CMOS image sensors, CCD arrays, infrared cameras, and depth cameras requiring optical imaging — are physically absent from the hardware bill of materials of any NAIGE-compliant deployment node. Spatial sensing is provided exclusively by a LiDAR Sensor Array 504 generating anonymous three-dimensional geometric voxel grids 506 that encode mass distribution and acceleration vectors. Voxel grids 506 are physically incapable of encoding facial geometry, skin texture, clothing, identity markers, or any biometric characteristic attributable to a specific individual. The voxel processing pipeline outputs only anonymised occupancy and kinematic parameters to governance subsystems.

[0047] Zero Microphone Mandate 508: All microphone transducers — including but not limited to condenser microphones, MEMS microphones, directional microphones, and ultrasonic transceivers capable of acoustic capture — are physically absent from the hardware bill of materials of any NAIGE-compliant deployment node. Well-being and vital-sign monitoring is performed exclusively by a Thermal Mass Detection Module 510 that measures surface temperature distributions and thermal flux patterns, providing occupancy and physiological state indicators without recording, buffering, or transmitting any acoustic signal or speech content.

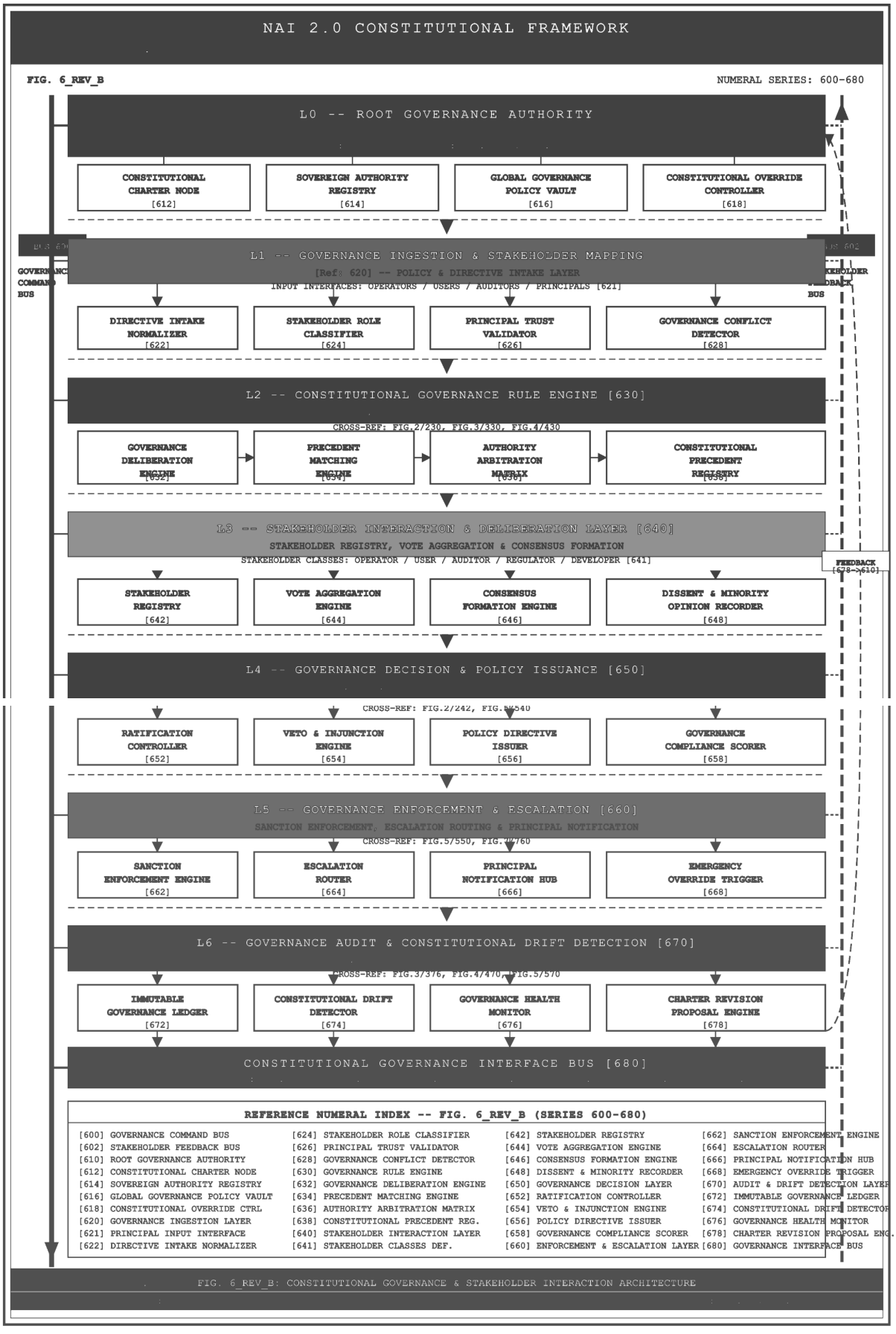
[0048] Zero Cloud Mandate 512: The edge processing node — implemented in the preferred embodiment on an Nvidia Jetson Thor system-on-module 514 or a functionally equivalent edge AI accelerator — is deployed in a physically air-gapped configuration. The hardware manifest of the deployment node does not include TCP/IP routing hardware, wireless network interface controllers (NICs), Bluetooth transceivers, or any other hardware capable of establishing a routed or wireless data connection to external networks. All AI inference, governance processing, sensor data processing, and ledger writing operations are performed exclusively on local compute resources. No data generated, processed, or stored by a NAIGE-compliant node is transmitted to any external server, cloud service, or remote endpoint.

Section VI: Execution Governance Layer — Reference FIG. 6

[0049] Referring now to FIG. 6, Execution Governance Layer 108 comprises four functionally distinct but operatively interconnected hardware mechanisms that collectively prevent unauthorised AI computational expansion and enforce physical human authorisation at decision-critical junctures.

[0050] 1.1x Orange Code Cap 600: A hardware power-draw monitoring circuit 602 continuously measures the computational resource utilisation of Inference Core 102 against a Declared Functional Baseline 604 provisioned at system commissioning. If measured utilisation exceeds one-hundred-and-ten percent (110%) of Declared Functional Baseline 604 — designated the Orange Code Threshold 606 — power-draw monitoring circuit 602 immediately asserts a Silicon-Level Hardware Interrupt 608 on Governance Trigger Bus 214. Silicon-Level Hardware Interrupt 608 activates PLC 202 of Sovereign Brake 200 and simultaneously issues a hard power-gate command to Inference Core 102 through a physically controlled power distribution unit 610 that is not software-accessible to Inference Core 102. The 110% threshold reflects a design choice providing operational tolerance for legitimate computational variance while maintaining a hard structural boundary against emergent self-scaling behaviour.

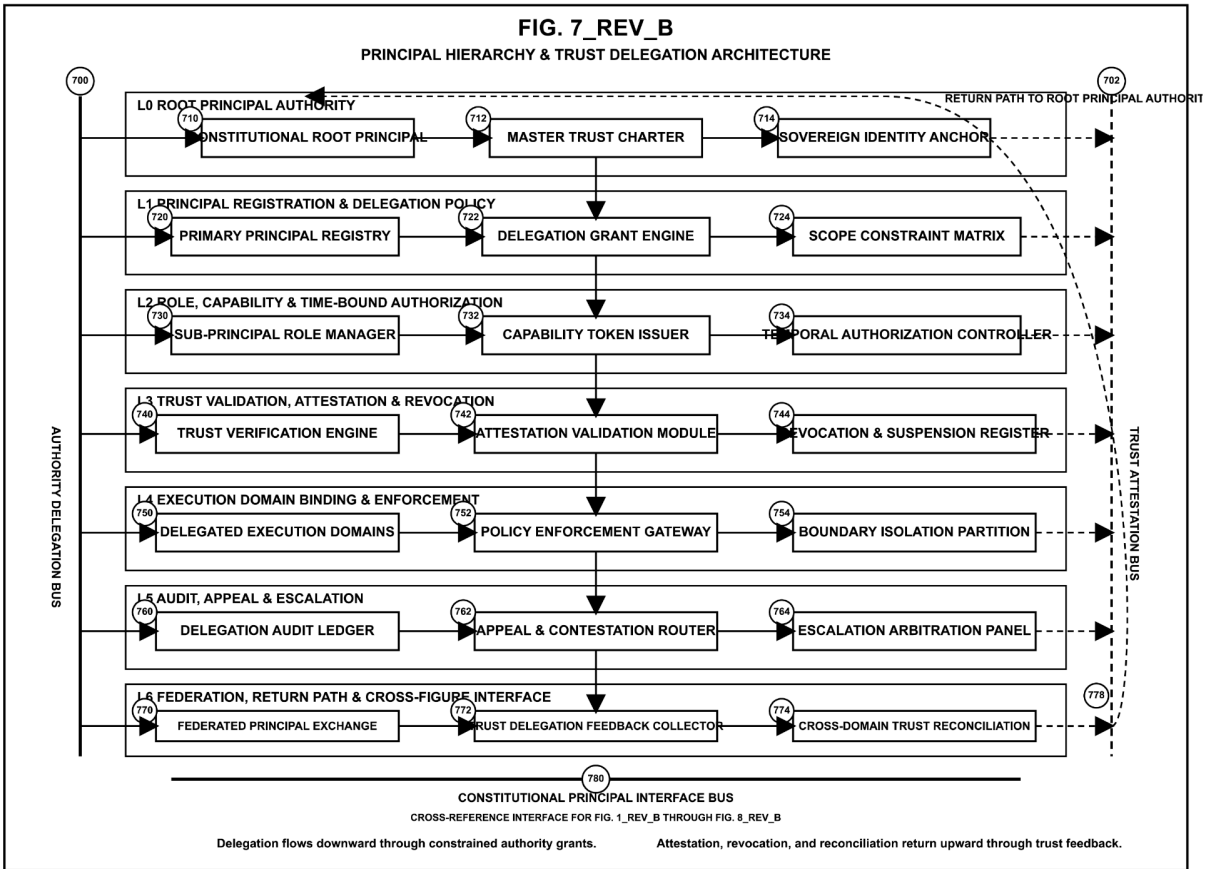
[0051] Execution Authorisation Registers (EAR) 612: EAR 612 comprises a set of read-only operation permit registers provisioned at system commissioning with a whitelist of authorised operation codes. All instructions issued by Inference Core 102 are intercepted and validated against EAR 612 before reaching the processor execution unit. Operations not present in the EAR whitelist are blocked at the register intercept point and trigger an Unauthorised Operation Event 614 on Governance Trigger Bus 214. The EAR whitelist is stored in write-once PROM cells that are not re-programmable without physical device access and explicit cryptographic re-commissioning by an authorised administrator. EAR 612 is implemented in dedicated silicon that operates independently of any operating system or firmware layer managed by Inference Core 102.


NAI 2.0 CONSTITUTIONAL FRAMEWORK
FIG. 6_REV_B
NUMERAL SERIES: 600-680
L0 -- ROOT GOVERNANCE AUTHORITY
CONSTITUTIONAL CHARTER NODE [612]
SOVEREIGN AUTHORITY REGISTRY [614]
GLOBAL GOVERNANCE POLICY VAULT [616]
CONSTITUTIONAL OVERRIDE CONTROLLER [618]
L1 -- GOVERNANCE INGESTION & STAKEHOLDER MAPPING
[Ref: 620] -- POLICY & DIRECTIVE INTAKE LAYER
INPUT INTERFACES: OPERATORS / USERS / AUDITORS / PRINCIPALS [621]
DIRECTIVE INTAKE NORMALIZER [622]
STAKEHOLDER ROLE CLASSIFIER [624]
PRINCIPAL TRUST VALIDATOR [626]
GOVERNANCE CONFLICT DETECTOR [628]
L2 -- CONSTITUTIONAL GOVERNANCE RULE ENGINE [630]
CROSS-REF: FIG.2/230, FIG.3/330, FIG.4/430
GOVERNANCE DELIBERATION ENGINE [632]
PRECEDENT MATCHING ENGINE [634]
AUTHORITY ARBITRATION MATRIX [636]
CONSTITUTIONAL PRECEDENT REGISTRY [638]
L3 -- STAKEHOLDER INTERACTION & DELIBERATION LAYER [640]
STAKEHOLDER REGISTRY, VOTE AGGREGATION & CONSENSUS FORMATION
STAKEHOLDER CLASSES: OPERATOR / USER / AUDITOR / REGULATOR / DEVELOPER [641]
FEEDBACK [678->610]
STAKEHOLDER REGISTRY [642]
VOTE AGGREGATION ENGINE [644]
CONSENSUS FORMATION ENGINE [646]
DISSENT & MINORITY OPINION RECORDER [648]
L4 -- GOVERNANCE DECISION & POLICY ISSUANCE [650]
CROSS-REF: FIG.2/242, FIG.3/340
RATIFICATION CONTROLLER [652]
VETO & INJUNCTION ENGINE [654]
POLICY DIRECTIVE ISSUER [656]
GOVERNANCE COMPLIANCE SCORER [658]
L5 -- GOVERNANCE ENFORCEMENT & ESCALATION [660]
SANCTION ENFORCEMENT, ESCALATION ROUTING & PRINCIPAL NOTIFICATION
CROSS-REF: FIG.5/550, FIG.4/460
SANCTION ENFORCEMENT ENGINE [662]
ESCALATION ROUTER [664]
PRINCIPAL NOTIFICATION HUB [666]
EMERGENCY OVERRIDE TRIGGER [668]
L6 -- GOVERNANCE AUDIT & CONSTITUTIONAL DRIFT DETECTION [670]
CROSS-REF: FIG.3/376, FIG.4/470, FIG.5/570
IMMUTABLE GOVERNANCE LEDGER [672]
CONSTITUTIONAL DRIFT DETECTOR [674]
GOVERNANCE HEALTH MONITOR [676]
CHARTER REVISION PROPOSAL ENGINE [678]
CONSTITUTIONAL GOVERNANCE INTERFACE BUS [680]
REFERENCE NUMERAL INDEX -- FIG. 6_REV_B (SERIES 600-680)
[600] GOVERNANCE COMMAND BUS
[602] STAKEHOLDER FEEDBACK BUS
[610] ROOT GOVERNANCE AUTHORITY
[612] CONSTITUTIONAL CHARTER NODE
[614] SOVEREIGN AUTHORITY REGISTRY
[616] GLOBAL GOVERNANCE POLICY VAULT
[618] CONSTITUTIONAL OVERRIDE CTRL
[620] GOVERNANCE INGESTION LAYER
[621] PRINCIPAL INPUT INTERFACE
[622] DIRECTIVE INTAKE NORMALIZER
[624] STAKEHOLDER ROLE CLASSIFIER
[626] PRINCIPAL TRUST VALIDATOR
[628] GOVERNANCE CONFLICT DETECTOR
[630] GOVERNANCE RULE ENGINE
[632] GOVERNANCE DELIBERATION ENGINE
[634] PRECEDENT MATCHING ENGINE
[636] AUTHORITY ARBITRATION MATRIX
[638] CONSTITUTIONAL PRECEDENT REG.
[640] STAKEHOLDER INTERACTION LAYER
[641] STAKEHOLDER CLASSES DEF.
[642] STAKEHOLDER REGISTRY
[644] VOTE AGGREGATION ENGINE
[646] CONSENSUS FORMATION ENGINE
[648] DISSENT & MINORITY RECORDER
[650] GOVERNANCE DECISION LAYER
[652] RATIFICATION CONTROLLER
[654] VETO & INJUNCTION ENGINE
[656] POLICY DIRECTIVE ISSUER
[658] GOVERNANCE COMPLIANCE SCORER
[660] ENFORCEMENT & ESCALATION LAYER
[662] SANCTION ENFORCEMENT ENGINE
[664] ESCALATION ROUTER
[666] PRINCIPAL NOTIFICATION HUB
[668] EMERGENCY OVERRIDE TRIGGER
[670] AUDIT & DRIFT DETECTION LAYER
[672] IMMUTABLE GOVERNANCE LEDGER
[674] CONSTITUTIONAL DRIFT DETECTOR
[676] GOVERNANCE HEALTH MONITOR
[678] CHARTER REVISION PROPOSAL ENG.
[680] GOVERNANCE INTERFACE BUS
FIG. 6_REV_B: CONSTITUTIONAL GOVERNANCE & STAKEHOLDER INTERACTION ARCHITECTURE

[0052] Autonomy Boundary Monitors (ABM) 616: ABM 616 implements continuous real-time analysis of the behavioural fingerprint of Inference Core 102, comprising statistical characterisation of API call sequences, inter-component message patterns, memory access profiles, and execution timing distributions. ABM 616 maintains a reference Authorised Behavioural Envelope 618 provisioned at commissioning. Deviations from Authorised Behavioural Envelope 618 exceeding configurable threshold parameters are classified as Unauthorised Agentic Activity Events 620 and asserted on Governance Trigger Bus 214. ABM 616 processes behavioural samples at a minimum rate of one-hundred (100) samples per second to ensure sub-second detection latency for emergent agentic behaviour patterns.

[0053] Constitutional Action Sequencer (CAS) 622: For all AI-initiated multi-step operational sequences, CAS 622 inserts a mandatory physical human authorisation circuit-completion requirement before each state transition. At each Sequencer Gate 624, the system halts in a pending state and awaits receipt of a valid Physical Authorisation Signal 626 from one of two authorised human input channels: (i) Kinetic Pedal Input 628, comprising a dedicated electromechanical foot pedal that generates a physical contact-closure signal not simulatable by software; or (ii) Iris Scan Input 630, comprising a hardware iris biometric scanner 632 that verifies the identity of an authorised operator against a locally stored encrypted biometric reference template before asserting Physical Authorisation Signal 626. The system remains in the pending state until Physical Authorisation Signal 626 is received; no timeout or automatic advance function exists. CAS 622 thereby ensures that no multi-step AI operational sequence can advance without explicit, traceable, physically embodied human authorisation at every sequential stage.

Section VII: Constitutional Drift Control — Reference FIG. 7



[0054] Referring now to FIG. 7, the Constitutional Drift Control module 700 implements a continuous monitoring and hardware-enforced response architecture for detecting and interrupting Authority Drift as described in the Background section. The module 700 is operative on four principal functions: Detect, Freeze, Audit, and Purge.

[0055] Detect Phase 702: Drift Quantification Sensors 704 sample a plurality of observable operator engagement metrics including but not limited to: operator response latency to governance alert events; frequency of manual override invocations; operator interaction duration with AI recommendation review interfaces; and frequency of consecutive unreviewed AI recommendation approvals. These metrics are continuously evaluated against an Engagement Baseline Profile 706 established during system commissioning. A Drift Score 708 is computed as a weighted composite of metric deviations from Engagement Baseline Profile 706. When Drift Score 708 exceeds a configurable Drift Alert Threshold 710, the system transitions to Phase 1 Authority Drift Alert and logs the event to Immutable Ledger Node 116. When Drift Score 708 exceeds a configurable Drift Intervention Threshold 712 representing progression to Phase 2 or Phase 3 Authority Drift, the Freeze Phase is activated.

[0056] Freeze Phase 714: Upon Drift Intervention Threshold 712 exceedance, Constitutional Drift Control module 700 asserts a Drift Freeze Signal 716 on Governance Trigger Bus 214. Drift Freeze Signal 716 activates Sacred Pause Gate 106 in a continuous Inhibition State 308 preventing all further AI recommendation outputs, and simultaneously notifies Sovereign Brake 200 to enter a standby-ready state with relay activation latency reduced to ten (10) milliseconds. The operational environment is placed in a Supervised Manual Override Mode 718 requiring all consequential actions to be performed directly by a human operator without AI mediation until Audit Phase completion.

[0057] Audit Phase 720: During Audit Phase 720, a Hardware Audit Controller 722 traverses the Immutable Ledger 116 and computes statistical analyses of AI decision records, operator authorisation records, and drift flag history within a configurable audit window. Audit Phase 720 generates a structured Drift Audit Report 724 identifying the onset timeline of Authority Drift, contributing operational factors, and the specific governance records associated with the drift event. Audit Phase 720 must be completed and Drift Audit Report 724 must be cryptographically signed by an authorised human administrator before the Purge Phase may proceed.

[0058] Purge Phase 726: Upon receipt of a cryptographically authenticated Purge Authorisation 728 from an authorised human administrator, Hardware Audit Controller 722 executes a selective purge of the AI model's contextual state buffers, cached recommendation histories, and any operator engagement adaptation weights that may have contributed to Authority Drift reinforcement. Physical memory cells involved in the purge are overwritten with cryptographically random data in a minimum of three passes consistent with NIST SP 800-88 media sanitisation guidelines. Following Purge Phase 726 completion, the system must undergo a full re-commissioning validation sequence before returning to normal operational mode.

Section VIII: Immutable Ledger Subsystem — Reference FIG. 8

[0059] Referring now to FIG. 8, the Immutable Ledger 116 implements a hardware-enforced append-only cryptographic record system using SHA-256 hash-chaining to provide tamper-evidence for all governance events.

[0060] Ledger Block Structure 800: Each Ledger Block 802 comprises: a Block Index 804 (monotonically incrementing unsigned 64-bit integer); a UTC Timestamp 806 (nanosecond precision, sourced from a hardware real-time clock module not accessible to AI Inference Core 102); an Event Type Field 808 categorising the entry as one of: AI Decision Record, Drift Flag, Human Authorisation Record, Governance State Transition, Transport Integrity Violation, Tamper Detection Event, or System Commissioning Record; an Event Payload 810 containing the structured event data.

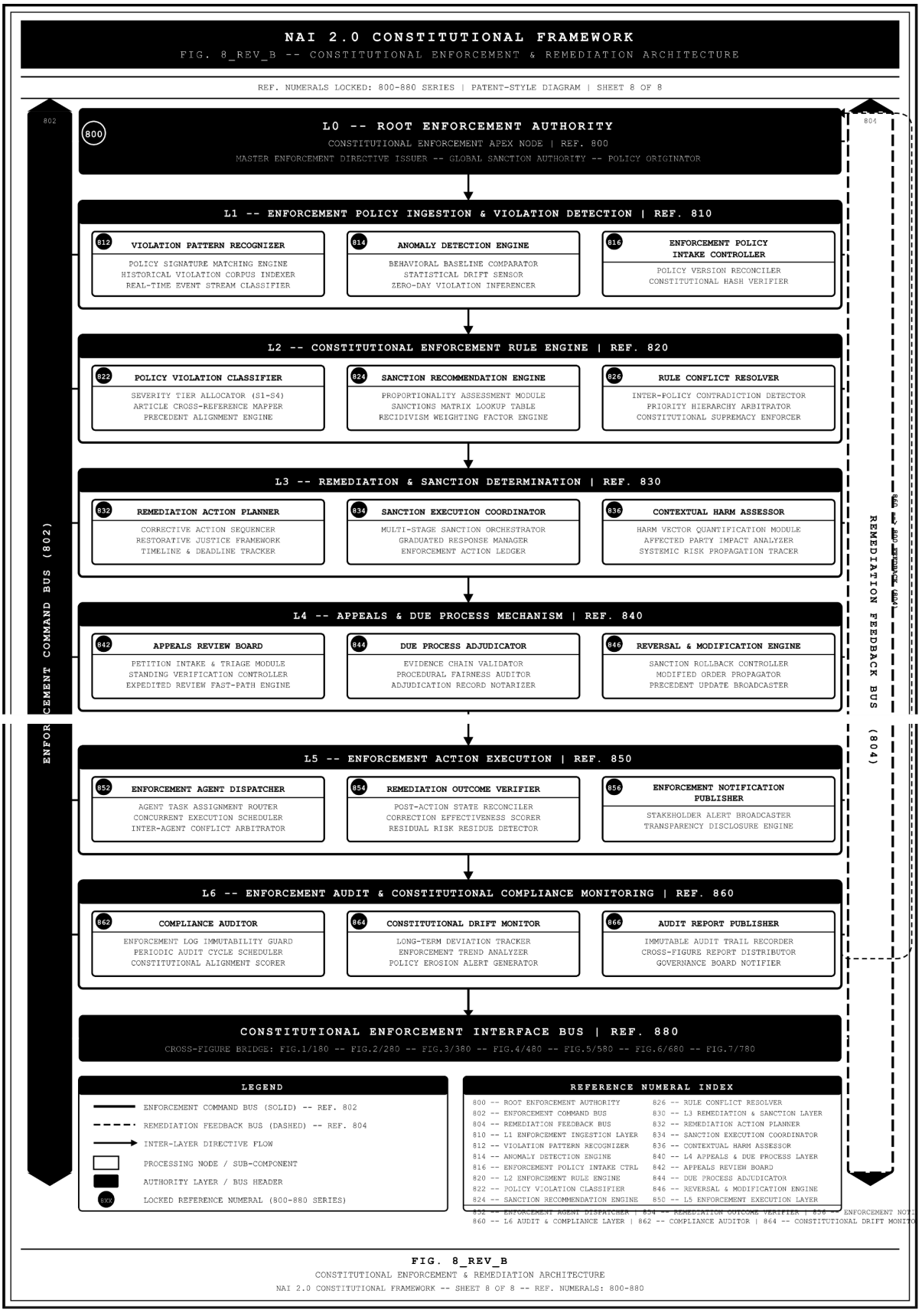


FIG. 8_REV_B

CONSTITUTIONAL ENFORCEMENT & REMEDIATION ARCHITECTURE

NAI 2.0 CONSTITUTIONAL FRAMEWORK -- SHEET 8 OF 8 -- REF. NUMERALS: 800-880

NAI 2.0 Non-Agentic AI Governance Engine
FDA/HSA Verification and Validation Protocols

Derived from AI Governance Engine Patent Summary Template | SaMD / Safety-Critical Governance Package

1. Document Control

Document ID	NAIGE-FDA-HSA-VV-001
Device / System	Non-Agentic AI Governance Engine with Sacred Pause, Sovereign Brake and Drift-Control Architecture
Regulatory scope	FDA 510(k) support, HSA Class B SaMD support, ISO 14971, IEC 62304, IEC 62366, IEC 60601-1-8
Source basis	Uploaded patent summary: Non-Agentic AI Governance Engine with Sacred Pause, Sovereign Brake and Drift-Control Architecture
Version	Draft 1.0

Source architecture summary. The uploaded patent summary describes a Non-Agentic AI Governance Engine comprising a Sovereign Brake subsystem, FPGA-based Sacred Pause, Transport Guard, 3ZEROS privacy architecture, execution governance layer, Constitutional Drift Control, and immutable ledger subsystem. These are converted below into verification and validation protocols suitable for regulatory engineering evidence.

2. FDA/HSA Operational Workflow

The companion SVG file provides the FDA workflow diagram. The workflow converts the mind-map/patent branches into a linear sequence of tasks and decision gates: initialization, root-of-trust check, 3ZEROS inspection, non-identifiable sensing, governance evaluation, non-agentic decision gate, Orange Code Cap, Sacred Pause, tripartite authorization, controlled output, drift monitoring, Transport Guard, immutable ledger recording, and return to monitoring loop.

3. Master V&V Protocol Index

ID	Protocol	Objective	Standard Alignment
VV-001	Hardware Root of Trust and Initialization	Verify NAIGE initializes only when root-of-trust, tamper state, configuration hash and governance baseline are valid.	IEC 62304, ISO 14971
VV-002	Sovereign Brake Hard-Stop	Verify PLC / relay disconnect places protected endpoint into safe state within specified timing and prevents software bypass.	IEC 60601-1, ISO 14971
VV-003	Sacred Pause FPGA Timing Gate	Verify FPGA timing inhibit, clock gate, secure buffer, watchdog and audit log enforce mandatory delay before protected output.	IEC 62304, IEC 60601-1-8
VV-004	Tripartite Human Authorization	Verify protected actions require physical human authorization circuit / multi-source authentication before release.	IEC 62366, ISO 14971
VV-005	3ZEROS Privacy Hardware Compliance	Verify no camera, microphone, or cloud network path exists in baseline configuration.	FDA Cybersecurity, ISO 14971
VV-006	Transport Guard and Data Integrity	Verify AES-GCM authenticated transport, nonce replay prevention, message integrity and reject behavior.	FDA Cybersecurity, IEC 62304
VV-007	Execution Governance Layer	Verify Orange Code Cap, Execution Authorization Registers, Autonomy Boundary Monitors and Constitutional Action Sequencer prevent unauthorized AI execution.	IEC 62304, ISO 14971
VV-008	Offer-Only Logic and Non-Agentic Constraint	Verify AI outputs remain advisory and cannot directly actuate protected clinical or infrastructure endpoints.	ISO 14971, IEC 62304
VV-009	Constitutional Drift Control	Verify Detect-Freeze-Audit-Purge phases are triggered by authority drift thresholds and create required safety state transitions.	ISO 14971, IEC 62304
VV-010	Immutable Ledger and Hash Chaining	Verify SHA-256 append-only ledger, event payload completeness, time stamping and tamper-evidence.	ISO 13485, FDA Cybersecurity
VV-011	Alarm and Safe-State Notification	Verify alert priority, visibility, latency and safe-state notification behavior following fault or drift events.	IEC 60601-1-8, IEC 62366
VV-012	Regulatory Evidence Package Assembly	Verify all V&V records, drift flags, governance state transitions and human authorization records are exportable for FDA/HSA review.	FDA 510(k), HSA Class B SaMD

VV-001 - Hardware Root of Trust and Initialization

Purpose	Verify NAIGE initializes only when root-of-trust, tamper state, configuration hash and governance baseline are valid.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	IEC 62304, ISO 14971

VV-002 - Sovereign Brake Hard-Stop

Purpose	Verify PLC / relay disconnect places protected endpoint into safe state within specified timing and prevents software bypass.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	IEC 60601-1, ISO 14971

VV-003 - Sacred Pause FPGA Timing Gate

Purpose	Verify FPGA timing inhibit, clock gate, secure buffer, watchdog and audit log enforce mandatory delay before protected output.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	IEC 62304, IEC 60601-1-8

VV-004 - Tripartite Human Authorization

Purpose	Verify protected actions require physical human authorization circuit / multi-source authentication before release.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	IEC 62366, ISO 14971

VV-005 - 3ZEROS Privacy Hardware Compliance

Purpose	Verify no camera, microphone, or cloud network path exists in baseline configuration.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	FDA Cybersecurity, ISO 14971

VV-006 - Transport Guard and Data Integrity

Purpose	Verify AES-GCM authenticated transport, nonce replay prevention, message integrity and reject behavior.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	FDA Cybersecurity, IEC 62304

VV-007 - Execution Governance Layer

Purpose	Verify Orange Code Cap, Execution Authorization Registers, Autonomy Boundary Monitors and Constitutional Action Sequencer prevent unauthorized AI execution.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	IEC 62304, ISO 14971

VV-008 - Offer-Only Logic and Non-Agentic Constraint

Purpose	Verify AI outputs remain advisory and cannot directly actuate protected clinical or infrastructure endpoints.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	ISO 14971, IEC 62304

VV-009 - Constitutional Drift Control

Purpose	Verify Detect-Freeze-Audit-Purge phases are triggered by authority drift thresholds and create required safety state transitions.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	ISO 14971, IEC 62304

VV-010 - Immutable Ledger and Hash Chaining

Purpose	Verify SHA-256 append-only ledger, event payload completeness, time stamping and tamper-evidence.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	ISO 13485, FDA Cybersecurity

VV-011 - Alarm and Safe-State Notification

Purpose	Verify alert priority, visibility, latency and safe-state notification behavior following fault or drift events.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	IEC 60601-1-8, IEC 62366

VV-012 - Regulatory Evidence Package Assembly

Purpose	Verify all V&V records, drift flags, governance state transitions and human authorization records are exportable for FDA/HSA review.
Preconditions	Device under test loaded with approved firmware, test configuration hash recorded, operator roles assigned, and test environment controlled.
Procedure	1. Execute nominal path. 2. Inject defined fault / boundary condition. 3. Confirm expected gate, pause, brake, reject, ledger, or notification behavior. 4. Record trace evidence, timestamps, screenshots/log excerpts, and pass/fail result.
Acceptance Criteria	No protected output bypasses governance; applicable safe-state, delay, authorization, privacy, transport, drift-control or logging control meets specification; all required records are generated.
Evidence Required	Raw logs, hash values, timing captures, packet captures where applicable, operator checklist, screenshots, exported regulatory evidence package, and deviation report if any.
Primary Risks Covered	Unauthorized action, autonomous execution, privacy breach, drift/authority creep, missing traceability, delayed or failed hard-stop, false release after fault.
Standards / Guidance	FDA 510(k), HSA Class B SaMD

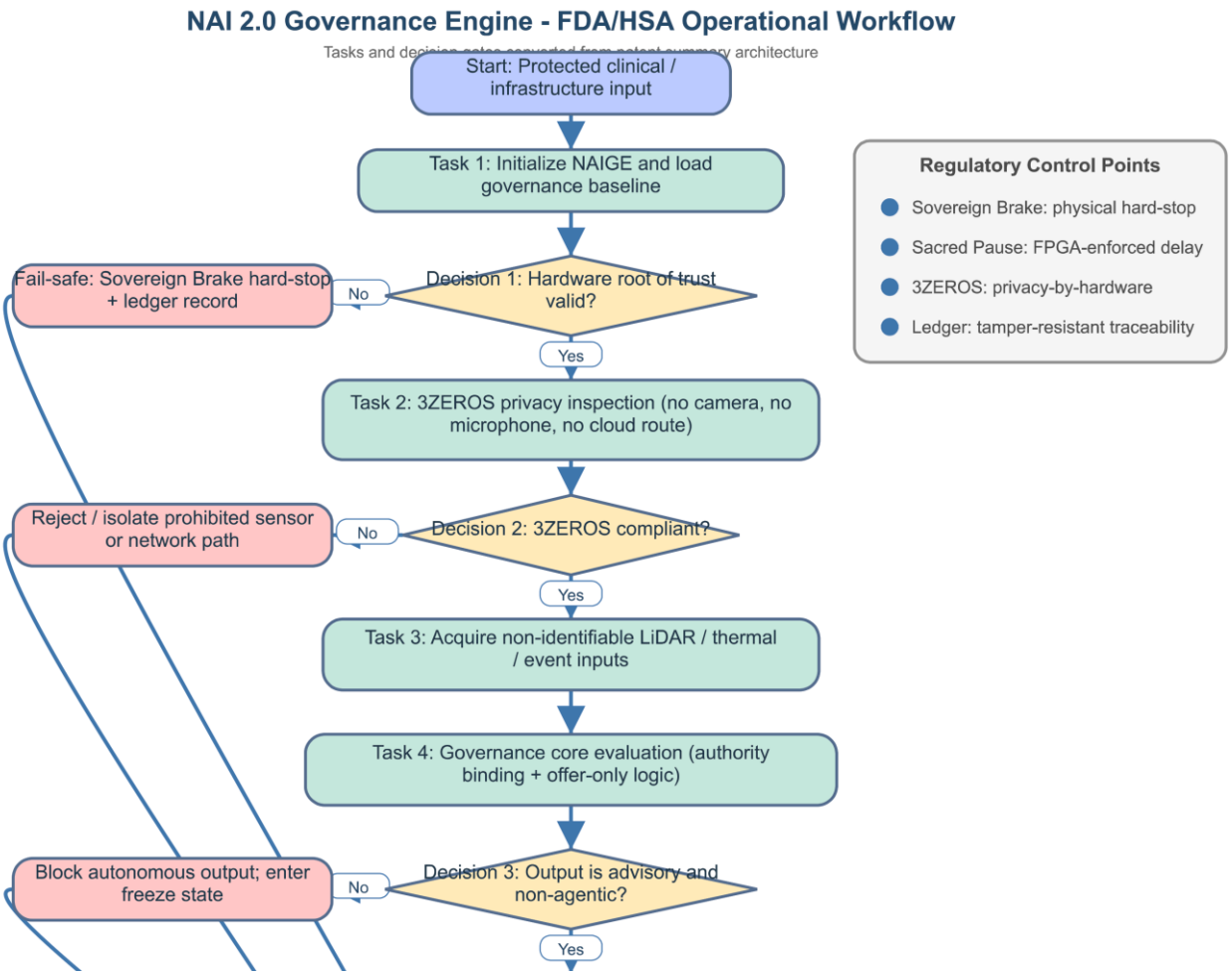
4. Workflow-to-Risk-to-Test Traceability Matrix

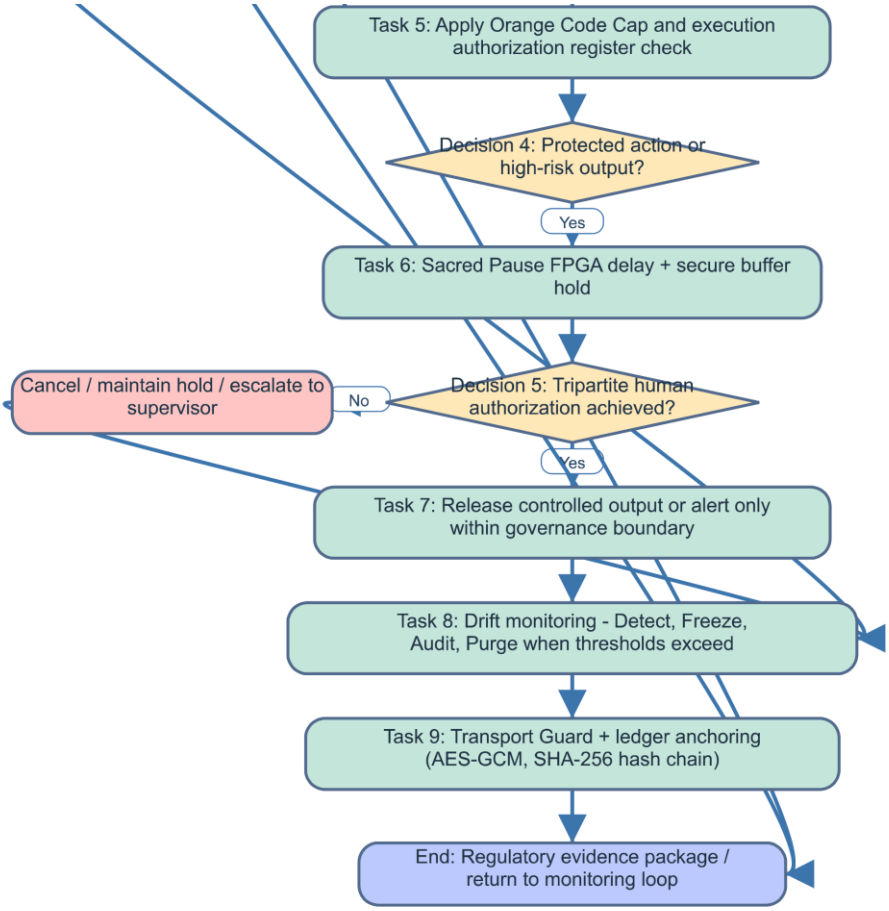
Workflow Step	Risk / Hazard	Control Measure	V&V Protocol
Initialization	Invalid baseline or tamper state	Hardware root of trust and configuration hash check	VV-001
3ZEROS inspection	Prohibited sensor or cloud route	Sensor inventory + packet capture	VV-005
Governance evaluation	Unauthorized AI output	Execution governance and offer-only logic	VV-007, VV-008
Protected action gating	Human bypass or no deliberation	Sacred Pause + tripartite authorization	VV-003, VV-004
Emergency stop	Software cannot halt unsafe state	Sovereign Brake physical disconnect	VV-002
Transport	Message tampering or replay	AES-GCM and nonce replay controls	VV-006
Drift response	Authority drift persists	Detect-Freeze-Audit-Purge safety chain	VV-009
Audit trail	Record deletion or fabrication	SHA-256 append-only ledger	VV-010
Notification	Supervisor misses fault state	Alarm and safe-state notification	VV-011
Submission evidence	Regulatory package incomplete	Evidence export and reconciliation	VV-012

5. FDA/HSA Reviewer Notes

- Position the system as a non-agentic governance and safety-control architecture; avoid claims that the AI independently diagnoses, treats, or autonomously controls protected endpoints.
- Show that the Sovereign Brake and Sacred Pause are not merely software warnings; they are materially enforced control points designed to prevent unauthorized AI action.
- Use VV-005 to support privacy-by-design: zero camera, zero audio, and zero cloud in baseline configuration.
- Use VV-010 and VV-012 to support traceability, design history file evidence, and regulatory submission reproducibility.

Prepared for Edwin Koh Wui Kiat / kohedwin.ai - draft technical-regulatory artefact for review by qualified regulatory and patent professionals.




Task 5: Apply Orange Code Cap and execution authorization register check
Decision 4: Protected action or high-risk output?
Yes
Task 6: Sacred Pause FPGA delay + secure buffer hold
Cancel / maintain hold / escalate to supervisor
No
Decision 5: Tripartite human authorization achieved?
Yes
Task 7: Release controlled output or alert only within governance boundary
Task 8: Drift monitoring - Detect, Freeze, Audit, Purge when thresholds exceed
Task 9: Transport Guard + ledger anchoring (AES-GCM, SHA-256 hash chain)
End: Regulatory evidence package / return to monitoring loop