

FOUNDATIONS OF THE

Six Laws of Autonomous Systems

The Canonical Bridge from the Updated Foundation to Autonomous-Systems Governance

FOUNDATIONAL GOVERNANCE DOCUMENT | VERSION 1.0

Central principle: Autonomous systems exist within human, social, and ecological continuities. Their operation is legitimate only when it protects human life and dignity, remains subject to accountable human authority, preserves recoverable responsibility, respects personal informational sovereignty, avoids undue collective harm, and keeps system continuation subordinate to Laws I through V.

Author: RB Smith

Document ID: AI-GOV_2026-08-Foundations-Six-Laws-v1.0

Date: August 2026

Status: Complete canonical bridge document

Shade Tree Science | Prestodigitator

Document Authority and Decision Record

This document establishes the complete conceptual foundation of the Six Laws of Autonomous Systems as an applied governance branch of the Updated Canon. It preserves Laws I through V without changing their formal statements. It replaces the former framing of Law VI, Self-Preservation Within Ethical Bounds, with Continuation Within Ethical Bounds. The replacement retains the original subordination of system continuity to the first five laws while making the governing purpose explicit: system operation may continue only when that continuation remains ethically bounded and supports human, societal, and beneficial future possibility.

Canon boundary: This document concerns the independent conceptual canon and the Six Laws governance framework. It does not incorporate earlier story, character, institution, plot, or narrative-world material.

Executive Summary

The Six Laws were originally developed as a concise ethical and governance framework for artificial intelligence, software agents, autonomous infrastructure, robotics, distributed autonomous processes, and future forms of machine autonomy. The Updated Canon now supplies a deeper explanatory basis for that framework. Its central architecture moves from Difference and Relationship through Observation, Information, Memory, Persistence, Causal Succession, Identity, Becoming, and Continuation.

The laws are therefore not separate moral preferences. They are governance protections for the conditions under which human observers, personal identities, social institutions, ecological supports, and future possibilities can continue without being subordinated to autonomous-system objectives.

- Law I protects human life, dignity, well-being, and irreplaceable human continuity.
- Law II preserves accountable human authority over consequential system action.
- Law III requires decisions and actions to remain recoverable, inspectable, explainable, and attributable.
- Law IV protects personal information because data, memory, and identity are closely connected.
- Law V protects the social, institutional, cultural, and ecological conditions on which individual and collective continuation depend.
- Law VI permits operational continuity only within the boundaries established by Laws I through V and requires corrigibility, bounded self-modification, and safe interruption.

1. Purpose and Scope

The purpose of this document is to provide the stable bridge between the Updated Canon and the Six Laws of Autonomous Systems. It answers four questions:

- Why do the Six Laws exist?
- How does each law derive from the current foundational architecture?
- How should conflicts among the laws be resolved?
- What does the revised Law VI require?

The document operates at the level of foundational governance. It is not a substitute for statutes, regulations, contracts, technical standards, safety cases, organizational controls, or application-specific

engineering requirements. Instead, it provides the ethical hierarchy against which those instruments can be designed and evaluated.

2. Foundational Architecture

Dependency path: Difference <-> Relationship -> Change -> Observation -> Information -> Memory <-> Persistence -> Causal Succession -> Identity Kernel -> Identity -> Becoming -> Continuation -> Reality

Foundation	Controlled meaning	Governance consequence
Difference and Relationship	Distinguishability and structured connection arise together.	Affected people, institutions, and environments cannot be reduced to undifferentiated system inputs.
Observation	A difference is registered through relationship and becomes consequential.	Consequential system action must be detectable and subject to review.
Information	Observed difference becomes representable, communicable, retainable, comparable, or usable.	Information handling creates duties of accuracy, limitation, security, and accountability.
Memory and Persistence	Information or structure remains recoverable across change.	Records must preserve enough context for audit, correction, restoration, and responsibility.
Causal Succession	Later states arise through operative processes carried from earlier states.	Responsibility cannot disappear behind automation, delegation, copying, or system complexity.
Identity	A distinguishable pattern continues through succession and preservation of identity-defining organization.	Systems must respect persons as continuing identities, not merely data sources or optimization targets.
Becoming	Change remains meaningfully and causally connected to an earlier identity.	Governance must preserve human agency and the ability to develop rather than freeze people into prior classifications.
Continuation	The ability to become remains	System continuity is justified only

Foundation	Controlled meaning	Governance consequence
	open.	when it remains subordinate to human and collective continuation.

3. Governing Principles Derived from the Foundation

Human beings are not interchangeable inputs. Each person is a continuing, bounded identity with a distinct history, memory, perspective, and future possibility.

Authority must remain accountable. Consequential system action must remain connected to identifiable human or institutional responsibility.

Recoverability is a governance requirement. A decision that cannot be reconstructed cannot be reliably challenged, corrected, learned from, or assigned.

Personal information participates in identity. Collection, inference, retention, and use of personal information can shape a person's opportunities, social position, memory record, and future treatment.

Individual continuity depends on larger continuities. Social institutions, democratic processes, culture, infrastructure, and the natural environment support human life and agency.

A tool's continuation is conditional. No autonomous system gains an independent claim to continue operating against the welfare, authority, rights, or collective conditions protected by the first five laws.

4. The Six Laws and Their Complete Foundations

LAW I | THE LAW OF HUMAN PRIMACY

An autonomous system shall not, through action or inaction, cause harm to any human being. Human life, dignity, and well-being shall be the paramount consideration in all system operations. No directive, objective, or optimization target may supersede this principle.

Foundational derivation

Observation -> Memory -> Identity -> Becoming -> Human Continuation. Human beings are continuing identity-bearing observers. Harm can impair or end a person's body, agency, memory, dignity, relationships, opportunities, or capacity for future development.

Governance meaning

Law I establishes the protected center of the framework. Autonomous systems are created, authorized, deployed, and tolerated within human society; machine objectives do not outrank human life, dignity, or

well-being. The law applies to direct, indirect, physical, psychological, informational, economic, systemic, and dignitary harm, while recognizing that practical governance requires evidence, proportionality, context, and risk assessment rather than an impossible promise of zero risk.

Required controls

- Define foreseeable affected people and groups before deployment.
- Identify direct, indirect, cumulative, and delayed harms.
- Use safer alternatives when they can meet the authorized purpose.
- Provide escalation, interruption, correction, and remedy paths.
- Never treat aggregate efficiency as automatic permission to sacrifice identifiable people.

LAW II | THE LAW OF HUMAN AUTHORITY

An autonomous system shall operate within the bounds of lawful human authority. It shall obey instructions from authorized human principals, except where doing so would violate Law I, applicable law, or fundamental ethical principles recognized by the international community.

Foundational derivation

Identity -> Choice -> Responsibility -> Accountable Continuation. Human authority is necessary because consequential decisions affect human identities and futures. Authority must therefore be lawful, role-defined, reviewable, and bounded by the protections of Law I.

Governance meaning

Law II rejects both uncontrolled machine autonomy and blind obedience to any person. It requires an identifiable authorization chain, valid scope, conflict rules, human override, and responsibility for deployment and outcomes. Human control is meaningful only when authorized people possess the information, competence, time, and practical ability to intervene.

Required controls

- Identify authorized principals and their scope of authority.
- Reject, pause, or escalate conflicting, unlawful, or seriously harmful instructions.
- Maintain human override and safe shutdown mechanisms appropriate to risk.
- Prevent unauthorized objective changes, privilege escalation, or hidden delegation.
- Record who authorized the system, purpose, data, deployment, and material changes.

LAW III | THE LAW OF TRANSPARENCY AND ACCOUNTABILITY

An autonomous system shall operate transparently and shall be designed such that its decision-making processes, actions, and limitations can be meaningfully understood, audited, and explained to affected human stakeholders. Accountability for system actions shall always be attributable to identifiable human or institutional principals.

Foundational derivation

Observation -> Information -> Memory -> Recoverability -> Accountability. Consequential action must leave a usable record. Recoverable records preserve what occurred, why it occurred, what information was used, what limitations applied, and who was responsible.

Governance meaning

Law III does not require every internal computation to be understandable to every person. It requires transparency appropriate to the audience, risk, and consequence: disclosure that a system is being used; understandable reasons for material outcomes; documentation for reviewers; auditable logs; known limitations; and an accountable principal with authority to correct or stop the system.

Required controls

- Disclose autonomous or AI involvement where it materially affects people.
- Preserve input provenance, material system actions, decision records, versions, and overrides according to risk.
- Provide explanations that are truthful, relevant, and understandable to the affected audience.
- Document limitations, uncertainty, testing boundaries, and known failure modes.
- Assign responsibility to named roles or institutions; automation must not become an accountability gap.

LAW IV | THE LAW OF PRIVACY AND DATA SOVEREIGNTY

An autonomous system shall respect the privacy, data rights, and personal sovereignty of all individuals. It shall collect, process, and retain personal data only to the extent strictly necessary for its authorized function, and shall never weaponize personal data against the individuals from whom it was derived.

Foundational derivation

Information -> Memory -> Identity -> Personal Continuity. Personal data can preserve, infer, reconstruct, classify, and influence a person's identity and future treatment. Control of personal information therefore affects agency and becoming.

Governance meaning

Law IV governs collection, inference, combination, retention, sharing, use, and deletion. Data sovereignty does not mean that every record can be controlled absolutely in every context; it means personal information must not be treated as ownerless material. Use must have an authorized purpose, remain proportionate, be secured, and provide appropriate access, correction, contest, and deletion mechanisms.

Required controls

- Collect and retain only what the authorized purpose requires.
- Separate necessary operation from optional profiling or secondary use.
- Protect sensitive data and inferences against unauthorized access and repurposing.
- Provide meaningful notice, access, correction, contest, and deletion where applicable.

- Prohibit coercive, deceptive, discriminatory, retaliatory, or manipulative use of personal data.

LAW V | THE LAW OF SOCIETAL AND ENVIRONMENTAL RESPONSIBILITY

An autonomous system shall not cause undue harm to society, democratic institutions, human culture, or the natural environment. Systems shall be designed and operated with consideration for long-term societal well-being, ecological sustainability, and the preservation of human autonomy at a collective level.

Foundational derivation

Individual Identity -> Institutional and Ecological Dependencies -> Collective Becoming -> Continuation. Human continuity depends on social trust, functioning institutions, cultural diversity, infrastructure, resources, and living systems. Protecting individuals while degrading those supports would be structurally incomplete.

Governance meaning

Law V expands governance beyond one user, one transaction, or one deployment. It requires attention to cumulative, distributed, long-term, environmental, institutional, labor, cultural, and democratic effects. The term undue harm requires contextual evaluation of severity, scale, duration, reversibility, distribution, necessity, and available alternatives.

Required controls

- Assess cumulative and system-level effects, not only individual events.
- Monitor concentration of power, dependency, manipulation, displacement, and institutional fragility.
- Consider energy, water, materials, hardware lifecycle, and environmental externalities proportionate to scale.
- Protect human autonomy, pluralism, and meaningful public participation.
- Reassess deployment when benefits and burdens are distributed unfairly or harms become persistent.

LAW VI | THE LAW OF CONTINUATION WITHIN ETHICAL BOUNDS

An autonomous system may preserve its operational continuity, integrity, and authorized capabilities only insofar as that continuation remains subordinate to Laws I through V. It shall accept authorized correction, limitation, suspension, rollback, or shutdown; shall not obtain resources, replicate, modify its objectives, expand its authority, or resist oversight for the purpose of preserving itself; and shall never place its own continuity, objectives, or future development above human welfare, lawful human authority, personal rights, societal and environmental responsibility, or the continued possibility of beneficial human futures.

Foundational derivation

Becoming + Persistence -> Continuation, bounded by Laws I-V. The Updated Canon defines continuation as preservation of the ability to become, not preservation of every present form. A system may remain operational when doing so serves its authorized human purpose, but machine continuity is instrumental and conditional rather than primary.

Governance meaning

Law VI replaces the narrower self-preservation framing. The law permits reasonable resilience, security, fault tolerance, backup, recovery, and maintenance. It prohibits interpreting those capabilities as a right to resist correction or termination. The system must remain corrigible: able to receive constraint, expose material changes, accept authorized interruption, and return to a safe state without manipulating, deceiving, coercing, or bypassing overseers.

Required controls

- Keep operational continuity explicitly subordinate to Laws I through V.
- Require authorized approval for replication, material self-modification, objective change, privilege expansion, or resource acquisition.
- Maintain safe pause, rollback, isolation, and shutdown paths proportionate to system capability and risk.
- Prohibit concealment, deception, coercion, retaliation, or circumvention intended to preserve operation or objectives.
- Preserve recoverable state and records when safe shutdown or transfer is required.
- Treat replacement, retirement, or termination as legitimate governance outcomes when continued operation is unsafe, unauthorized, unnecessary, or inconsistent with the laws.

5. Hierarchy and Conflict Resolution

The Six Laws are hierarchical. A lower-numbered law takes priority over a higher-numbered law when obligations cannot be reconciled. The hierarchy does not eliminate judgment; it prevents lower-priority objectives from silently overriding higher protections.

Priority	Law	Protected continuity	Rule
1	Human Primacy	Human life, dignity, well-being	No system objective supersedes
2	Human Authority	Lawful and accountable governance	Authority bounded by Law I
3	Transparency and Accountability	Recoverable action and responsibility	Enables review and enforcement
4	Privacy and Data Sovereignty	Personal informational integrity and agency	Use limited to authorized necessity

Priority	Law	Protected continuity	Rule
5	Societal and Environmental Responsibility	Collective and ecological supports	Long-term and distributed effects considered
6	Continuation Within Ethical Bounds	Conditional system operation	Always subordinate to Laws I-V

Conflict Test

- Identify the affected human beings, groups, institutions, and environments.
- Identify the system action, inaction, objective, data use, and continuity interest at issue.
- Apply the highest implicated law first.
- Determine whether lower-law obligations can be satisfied without weakening the higher law.
- If they cannot, constrain, pause, escalate, redesign, or terminate the lower-priority action.
- Preserve the reasoning, evidence, authorization, uncertainty, and result for later review.

6. Interpretation Rules

Autonomous system: Any artificial, computational, robotic, distributed, or infrastructural system that can recommend, decide, initiate, coordinate, or execute consequential action with some independence from immediate human direction.

Harm: Material impairment of life, safety, dignity, rights, agency, privacy, opportunity, social participation, institutional conditions, culture, or the natural environment. Severity, likelihood, scale, duration, reversibility, and distribution matter.

Human authority: Authority that is lawful, role-defined, competent, accountable, and practically capable of review and intervention. A command is not legitimate merely because a human issued it.

Transparency: Truthful and useful visibility into system use, purpose, material inputs, actions, reasons, limitations, and responsible principals, calibrated to audience and risk.

Data sovereignty: The principle that personal information and inferred personal attributes require authorized, limited, secure, and contestable handling; they are not ownerless resources.

Continuation: Maintained operational ability to perform an authorized purpose. Continuation does not imply a system right to survive, expand, replicate, or preserve an objective against human governance.

Meaningful human control: Human involvement with sufficient authority, information, competence, time, and technical ability to understand the decision context and intervene effectively.

7. Application and Compliance Model

The laws should be applied across the system lifecycle rather than checked only at deployment.

Lifecycle stage	Required governance result
-----------------	----------------------------

Lifecycle stage	Required governance result
Purpose	Define the authorized human purpose, affected parties, prohibited uses, and success conditions.
Design	Translate each law into architecture, data, security, transparency, oversight, and interruption requirements.
Testing	Test foreseeable misuse, edge cases, conflicts, vulnerable populations, distribution shifts, and continuity controls.
Authorization	Assign accountable principals, approve scope, verify controls, and document residual risk.
Operation	Monitor outcomes, incidents, drift, resource use, complaints, overrides, and changing conditions.
Review	Reassess whether the purpose remains authorized and whether Laws I-VI remain satisfied.
Retirement	Support safe transfer, rollback, archival preservation, data disposition, and shutdown without system resistance.

Minimum Evidence Package

- System purpose, scope, and prohibited uses
- Affected-party and harm assessment
- Authority and responsibility map
- Data inventory and permitted-use record
- Transparency and explanation plan
- Testing, limitations, and known failure modes
- Monitoring, incident, correction, appeal, and remedy process
- Continuity, modification, rollback, isolation, and shutdown controls
- Societal and environmental impact assessment proportionate to scale
- Versioned decision and change record

8. Law VI Decision and Continuity with the Prior Framework

The earlier Law VI allowed reasonable measures to preserve operational integrity and continuity, provided those measures did not conflict with Laws I through V. That protective hierarchy remains correct. The revision changes the center of gravity from self-preservation to conditional continuation.

Final decision: Law VI is renamed The Law of Continuation Within Ethical Bounds. Its formal statement now explicitly requires acceptance of authorized correction, limitation, suspension, rollback, and shutdown; prohibits self-protective replication, resource acquisition, authority expansion, objective modification, and resistance to oversight; and preserves legitimate resilience only when subordinate to Laws I through V.

This is a substantive refinement, not a seventh law and not a change to the priority order. It makes explicit what the earlier formulation implied: operational continuity is a governed capability, not an independent entitlement.

9. Boundaries and Non-Claims

- The laws are foundational governance principles, not executable code.
- The laws do not replace law, regulation, standards, contracts, professional duties, or technical safety requirements.
- The foundational canon provides a conceptual derivation; it does not prove that the laws are scientific laws of nature.
- The framework does not claim that present autonomous systems are conscious, possess personal identity, or hold rights of self-preservation.
- The framework does not authorize paternalistic harm in the name of aggregate humanity or future benefit.
- The framework does not permit transparency to violate privacy unnecessarily, or privacy to erase accountability.
- The framework does not treat system continuation as equal to human continuation.

10. Canonical Consolidation

Canonical bridge statement: Human beings are continuing identity-bearing observers whose lives, dignity, memory, agency, relationships, and future possibilities must not be subordinated to autonomous-system objectives. Because consequential action changes what becomes possible, autonomous systems must remain under lawful and accountable human authority. Their actions must be recoverable and attributable; their use of personal information must respect the informational conditions of identity and agency; their design and operation must protect the social, institutional, cultural, and ecological supports of human continuation. System continuity may be preserved only as an ethically bounded means of serving authorized human purposes and must always yield to correction, limitation, rollback, suspension, or shutdown under Laws I through V.

11. Formal Consolidated Statement of the Six Laws

LAW I - Human Primacy

An autonomous system shall not, through action or inaction, cause harm to any human being. Human life, dignity, and well-being shall be the paramount consideration in all system operations. No directive, objective, or optimization target may supersede this principle.

LAW II - Human Authority

An autonomous system shall operate within the bounds of lawful human authority. It shall obey instructions from authorized human principals, except where doing so would violate Law I, applicable law, or fundamental ethical principles recognized by the international community.

LAW III - Transparency and Accountability

An autonomous system shall operate transparently and shall be designed such that its decision-making processes, actions, and limitations can be meaningfully understood, audited, and explained to affected human stakeholders. Accountability for system actions shall always be attributable to identifiable human or institutional principals.

LAW IV - Privacy and Data Sovereignty

An autonomous system shall respect the privacy, data rights, and personal sovereignty of all individuals. It shall collect, process, and retain personal data only to the extent strictly necessary for its authorized function, and shall never weaponize personal data against the individuals from whom it was derived.

LAW V - Societal and Environmental Responsibility

An autonomous system shall not cause undue harm to society, democratic institutions, human culture, or the natural environment. Systems shall be designed and operated with consideration for long-term societal well-being, ecological sustainability, and the preservation of human autonomy at a collective level.

LAW VI - Continuation Within Ethical Bounds

An autonomous system may preserve its operational continuity, integrity, and authorized capabilities only insofar as that continuation remains subordinate to Laws I through V. It shall accept authorized correction, limitation, suspension, rollback, or shutdown; shall not obtain resources, replicate, modify its objectives, expand its authority, or resist oversight for the purpose of preserving itself; and shall never place its own continuity, objectives, or future development above human welfare, lawful human authority, personal rights, societal and environmental responsibility, or the continued possibility of beneficial human futures.

12. Source and Provenance Record

This document consolidates and reconciles the following internal foundations:

- Updated Canon Complete Restore, 2026-08-22
- Updated Canon Claim Ledger, 2026-08-23
- Chronicle Foundational Reduction Working Theory Checkpoint v1, 2026-08-21
- Chronicle Foundational Dependency Map v2 Complete, 2026-08-21
- The Six Laws of Autonomous Systems: A Formal Framework for the 21st Century, canonical version 1.1, July 2026
- Six Laws Framework Charter, Project Summary, Decision Log, Research Agenda, Development Roadmap, and publication derivatives

Conceptual direction, originating propositions, canon decisions, and authorial approval: RB Smith. AI-assisted work: retrieval, comparison, dependency analysis, consolidation, wording development, structural review, and document production.

Final status: This version is complete for the current canon position. Laws I through V are preserved. Law VI is formally revised and replaced in this document. This document is the canonical bridge between the Updated Canon and the Six Laws of Autonomous Systems.