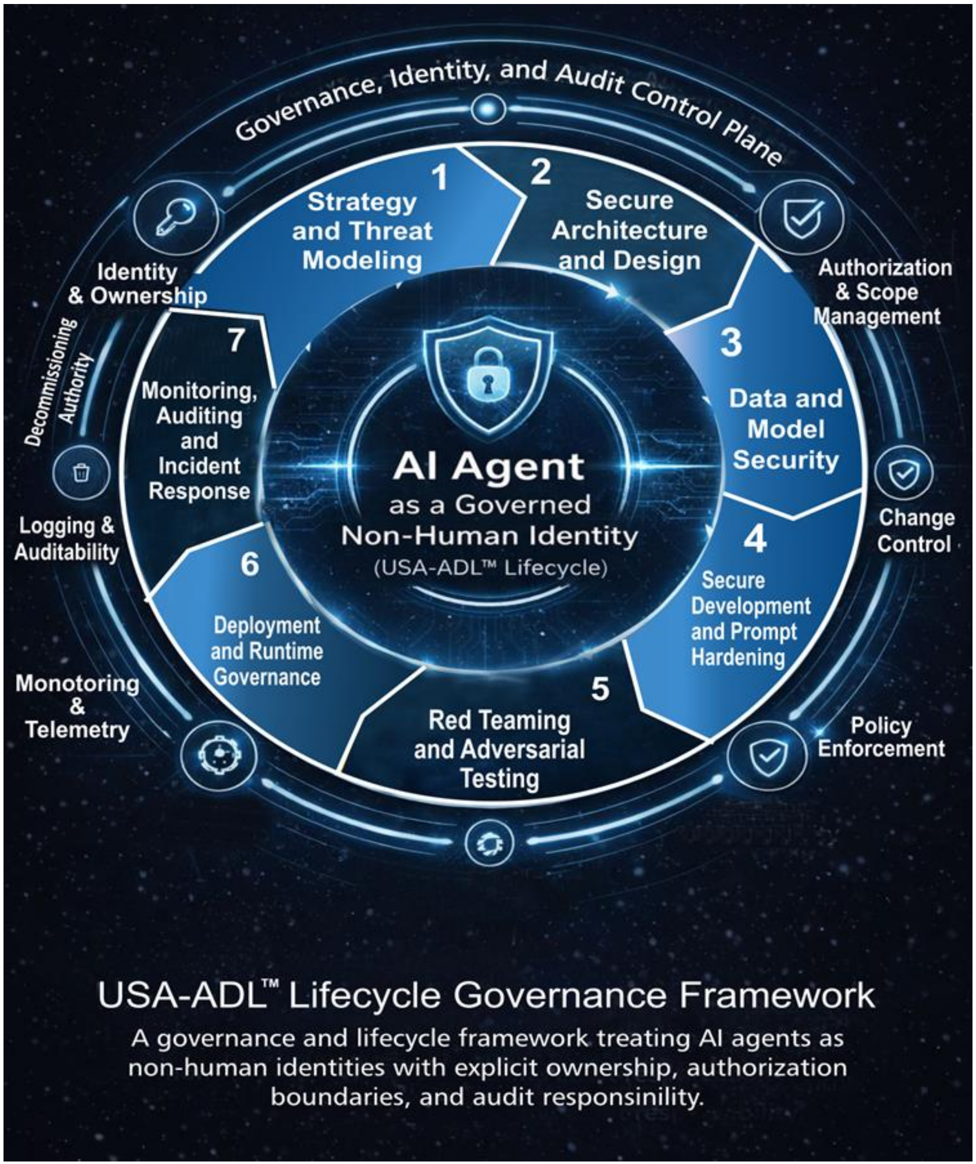
Uranusys

Your Security, Our Mission


Governance, Identity, and Audit Control Plane
1
Strategy and Threat Modeling
2
Secure Architecture and Design
3
Data and Model Security
4
Secure Development and Prompt Hardening
5
Red Teaming and Adversarial Testing
6
Deployment and Runtime Governance
7
Monitoring, Auditing and Incident Response
AI Agent
as a Governed Non-Human Identity
(USA-ADL™ Lifecycle)
Identity & Ownership
Decommissioning Authority
Logging & Auditability
Monotoring & Telemetry
Authorization & Scope Management
Change Control
Policy Enforcement

USA-ADL™ Lifecycle Governance Framework

A governance and lifecycle framework treating AI agents as non-human identities with explicit ownership, authorization boundaries, and audit responsinility.

Uranusys Secure AI Agent Development Lifecycle (USA-ADL) v1.0

Uranusys Secure AI Agent Development Lifecycle USA-ADL v1.0

Published: 2025-12-15
Version: 1.0
Status: Official Uranusys internal standard
Owner: Uranusys LLC

SUMMARY

1. Abstract

The Uranusys Secure AI Agent Development Lifecycle (USA-ADL™) provides a practical, security-first lifecycle for building, deploying, and governing AI Agents that organizations can trust. It defines mandatory governance, security, and operational controls for every AI Agent developed, deployed, or operated by Uranusys, across experimental and production environments.

Under USA-ADL, AI Agents are treated as non-human identities with explicit ownership, clearly defined operational scope, authorization boundaries, and full audit responsibility across their lifecycle. The framework ensures that AI Agents are secure by design, designed to resist misuse, governed end to end, and aligned with international AI governance standards, including ISO/IEC 42001.

2. Keywords

AI Agents; secure automation; Uranusys; USA-ADL; AI governance; AI risk management; non-human identity; lifecycle controls; ISO/IEC 42001 alignment; threat modeling; red teaming; Zero Trust; monitoring and audit.

3. Audience

USA-ADL v1.0 is intended for:

- **Security and AI leadership**
  CISOs, Heads of AI, product leaders, and executives responsible for AI adoption, risk, and compliance.
- **Engineering and product teams**
  AI engineers, MLOps teams, software developers, solution architects, and platform owners building or integrating AI Agents.
- **Governance, risk, and compliance stakeholders**
  Risk managers, internal auditors, privacy and legal teams, and stakeholders preparing for ISO/IEC 42001 or similar AI governance requirements.
- **Client and partner stakeholders**
  Organizations evaluating Uranusys AI Agents, due-diligence teams, and technical partners who need to understand how Uranusys governs AI Agents across their lifecycle.

4. Note to Readers

USA-ADL v1.0 is the official Uranusys Secure AI Agent Development Lifecycle and is considered a **locked baseline** for all Uranusys AI Agents. Future versions (for example v1.1 or v2.0) may extend or refine controls, but USA-ADL v1.0 remains the reference point for traceability, evidence, and historical audits.

Unless otherwise noted, external standards and frameworks cited in this document (such as ISO/IEC 42001 or NIST publications) are **referenced** for alignment and context. They are not reproduced, and their full content should be consulted directly from their official publishers.

USA-ADL governs how AI Agents are designed, built, and operated, not which specific business domains or workflows clients choose to automate. The examples given (vulnerability management, sports analytics, legal workflows, and others) are illustrative and may evolve with Uranusys products and client needs.

5. Openness & Interoperability Statement

USA-ADL v1.0 was designed and authored by Uranusys LLC, building on practical experience in cybersecurity, AI security, and secure architecture across multiple industries.

Uranusys acknowledges the value of international and national standards that inform AI and cybersecurity governance, including ISO/IEC 42001 for AI Management Systems and the NIST Cybersecurity Framework 2.0 for risk-based governance and lifecycle thinking. These references have influenced the structure and intent of USA-ADL while remaining independent and proprietary to Uranusys.

6. Acknowledgments

USA-ADL v1.0 was designed and authored by Uranusys LLC, informed by practical experience in cybersecurity, AI security, and secure architecture across multiple industries.

Uranusys acknowledges the value of international and national standards that inform AI and cybersecurity governance, including ISO/IEC 42001 for AI Management Systems and the NIST Cybersecurity Framework 2.0 for risk-based governance and lifecycle thinking. These references have informed the structure and intent of USA-ADL while remaining independent of, and not derived from, any external standard. USA-ADL is developed and governed by Uranusys LLC.

7. Relationship to External Standards and Frameworks

USA-ADL v1.0 is an independently developed and governed framework created by Uranusys LLC to define lifecycle-based governance for the secure design, development, deployment, and operation of AI agents.

While USA-ADL is not derived from any single external standard, it is intentionally designed to align with the intent and expectations of leading international AI and cybersecurity governance frameworks.

USA-ADL complements, but does not replace, organizational AI Management Systems such as ISO/IEC 42001. It provides technical, architectural, and lifecycle governance constructs that support the operationalization of AI governance in real-world AI agent deployments. Organizational policies, management system responsibilities, and certification activities remain the responsibility of the adopting organization.

Key external references include:

- **ISO/IEC 42001 (AI Management Systems)**

USA-ADL supplies lifecycle-level governance and technical control expectations that support ISO/IEC 42001 objectives related to AI risk management, human oversight, data governance, monitoring, and continual improvement. Formal compliance, certification, and management system accountability remain outside the scope of the framework.

- **NIST Cybersecurity Framework (CSF) 2.0 and related NIST publications**

USA-ADL reflects NIST's risk-based governance principles and lifecycle orientation, extending these concepts to AI agents through controls addressing agent autonomy, misuse resistance, prompt security, and auditability.

- **Zero Trust and identity-centric security models**

USA-ADL treats AI agents as non-human identities with explicit ownership, defined scope, authorization boundaries, and audit responsibility, aligning with modern identity-first security practices.

USA-ADL is designed to be framework-agnostic and forward-compatible, enabling organizations to map its governance constructs to evolving regulatory, industry, and certification requirements without requiring re-architecture of AI systems.

USA-ADL does not reproduce, incorporate, or redistribute copyrighted material from external standards. All referenced frameworks are cited for contextual alignment only and remain the intellectual property of their respective publishers.

8. Copyright & Usage Notice

USA-ADL™ (Uranusys Secure AI Agent Development Lifecycle) v1.0 is an original governance and lifecycle framework developed and first published by Uranusys LLC.

Permission is granted to reference, cite, reproduce, and map to the USA-ADL™ framework specification for informational, governance, due-diligence, and alignment purposes, provided appropriate attribution is maintained.

Commercial implementation methodologies, assessment models, certification programs, enforcement tooling, and derivative commercial offerings remain the exclusive rights of Uranusys LLC unless expressly authorized in writing.

USA-ADL™ v1.0 is provided for governance and informational purposes only and does not constitute legal advice, regulatory guidance, or certification by itself.

9. Purpose

The **Uranusys Secure AI Agent Development Lifecycle (USA-ADL™)** defines the mandatory governance, security, and operational controls applied to all AI Agents developed, deployed, or operated by Uranusys.

Under USA-ADL, every AI Agent is treated as a non-human identity with explicit ownership, defined operational scope, authorization boundaries, and full audit responsibility across its lifecycle.

USA-ADL ensures AI Agents are:

- Secure-by-design
- Resistant to misuse and manipulation
- Governed through their entire lifecycle
- Auditable and accountable
- Aligned with international AI governance standards, including ISO/IEC 42001

10. Scope

USA-ADL applies to:

- Client-facing AI Agents (e.g., Vulnerability AI Agent – VAA)
- Internal Uranusys AI systems
- Experimental and production AI Agent workloads
- AI Agents operating in cybersecurity, sports analytics, legal, compliance, and other domains

USA-ADL governs **how secure AI Agents are designed, built, deployed, and operated** across their lifecycle, including security controls, governance mechanisms, and operational safeguards.

11. Out of Scope

USA-ADL does not define business objectives, ethical policies, regulatory interpretations, or organizational decision-making processes.

USA-ADL does not determine **whether an AI Agent use case should be pursued, approved, or deployed**, nor does it replace enterprise AI governance, legal review, risk acceptance, or compliance obligations.

Decisions regarding **use case approval, organizational risk appetite, regulatory interpretation, and ethical policy enforcement** remain the sole responsibility of the client organization.

12. Definitions

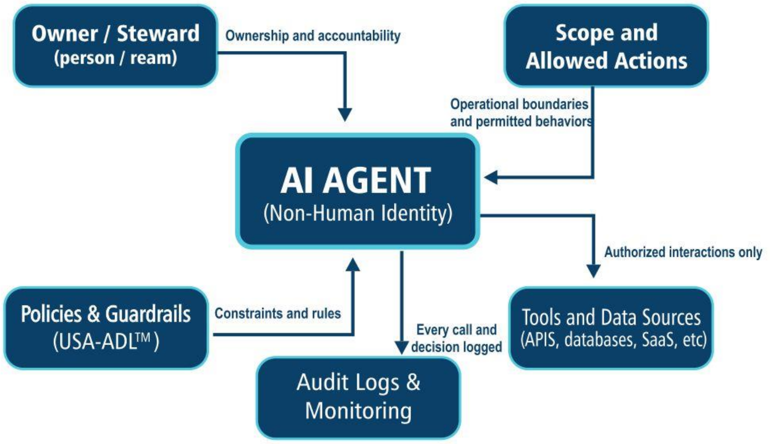
12.1 AI Agent

A software system that uses artificial intelligence to analyze inputs, apply reasoning or decision logic, and produce outputs or actions within a defined operational scope. Under USA-ADL™, an AI Agent may interact with data sources, tools, systems, or workflows but operates only within explicitly authorized boundaries and controls.

12.2 AI Agents as Non-Human Identities

A digital identity assigned to an AI Agent that is distinct from human users and service accounts. A non-human identity has defined ownership, authentication credentials, authorization boundaries, and auditability. Treating AI Agents as non-human identities enables identity-centric security, least privilege enforcement, traceability, and accountability throughout the AI Agent lifecycle.

AI Agent as a Non-Human Identity



12.3 Autonomous Action

Any action executed by an AI Agent that directly affects systems, data, workflows, or users without immediate human approval. Under USA-ADL, autonomous actions are prohibited by default and may only be enabled when explicitly approved, governed, monitored, and documented as part of the AI Agent's authorized scope.

12.4 Advisory Mode

An operating mode in which an AI Agent provides analysis, recommendations, insights, or suggested actions to human users without executing changes independently. Advisory mode is the default operating state under USA-ADL and ensures that humans retain decision authority and accountability.

12.5 Lifecycle Owner

A named human individual responsible for an AI Agent across its entire lifecycle. The lifecycle owner is accountable for approval, scope definition, risk acceptance, oversight, incident coordination, and ensuring compliance with USA-ADL controls from design through decommissioning.

12.6 Runtime Governance

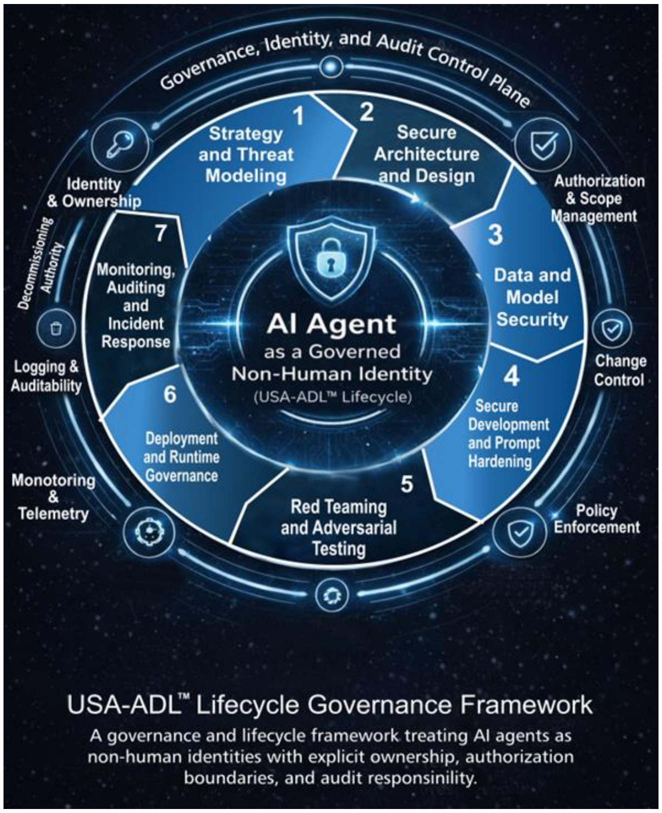
The set of technical and operational controls applied to an AI Agent during production operation. Runtime governance includes authentication and authorization, rate limiting, monitoring, logging, incident detection, kill-switch mechanisms, and enforcement of approved behaviors to ensure continued compliance with USA-ADL and organizational policies.

13. Core Principles

All Uranusys AI Agents must adhere to the following principles:

1. **Defined Purpose & Boundary**
    - Every AI Agent has a clearly documented intended use.
    - Actions outside this scope are explicitly prohibited.
2. **Least Privilege**
    - AI Agents have only the minimum data, tools, and permissions required.
3. **Advisory by Default**
    - AI Agents provide analysis and recommendations.
    - Autonomous execution is prohibited unless explicitly approved and governed.
4. **Human Accountability**
    - Humans remain accountable for decisions and actions.
    - AI outputs support, but do not replace, human judgment.
5. **Security Before Capability**
    - New capabilities are added only after security and misuse risks are assessed.

14. USA-ADL Lifecycle Overview


Governance, Identity, and Audit Control Plane
1
Strategy and Threat Modeling
2
Secure Architecture and Design
3
Data and Model Security
4
Secure Development and Prompt Hardening
5
Red Teaming and Adversarial Testing
6
Deployment and Runtime Governance
7
Monitoring, Auditing and Incident Response
AI Agent
as a Governed
Non-Human Identity
(USA-ADL™ Lifecycle)
Identity & Ownership
Authorization & Scope Management
Change Control
Policy Enforcement
Monotoring & Telemetry
Logging & Auditability
Decommissioning Authority
USA-ADL™ Lifecycle Governance Framework
A governance and lifecycle framework treating AI agents as non-human identities with explicit ownership, authorization boundaries, and audit responsinility.

15. USA-ADL Lifecycle Phases (v1.0)

15.1 Phase 1 – Strategy & Threat Modeling

Objective

Define purpose, scope, explicit out-of-scope behaviors, and potential misuse scenarios before building.

Required Controls

- Clear definition of intended use
- Explicit function and action boundaries
- AI-specific threat modeling (prompt abuse, data leakage, misuse)
- Definition of prohibited behaviors

Artifacts

- Purpose & boundary statement
- Threat model / misuse scenarios

Each AI Agent must have a named human owner responsible for approval, oversight, incident response coordination, and lifecycle accountability.

15.2 Phase 2 – Secure Architecture & Design

Objective

Ensure AI Agents are architected with isolation and least privilege.

Required Controls

- Separation between:
    - AI reasoning
    - Policy enforcement
    - Output generation
- Zero Trust assumptions
- No implicit trust in data sources or LLM outputs
- Constrained tool and API access

Artifacts

- Architecture diagrams
- IAM and trust boundary documentation

15.3 Phase 3 – Data & Model Security

Objective

Protect data handling and prevent leakage or misuse.

Required Controls

- Data minimization
- Ephemeral processing unless retention is approved
- Input classification (trusted vs untrusted)
- Output sanitization and validation
- No training on client data without authorization

Artifacts

- Data handling rules
- Sanitization logic
- Logging and retention policy

Where applicable, data handling controls may include geographic, jurisdictional, or contractual constraints defined outside USA-ADL.

15.4 Phase 4 – Secure Development & Prompt Hardening

Objective

Prevent prompt injection, instruction override, and unsafe outputs.

Required Controls

- System prompts defining:
  - Role
  - Prohibited actions
  - Refusal behavior
- Explicit treatment of user and external data as untrusted
- Output inspection for disallowed content
- No agent autonomy beyond approved scope

Artifacts

- Prompt templates
- Guardrail logic
- Output validation rules

15.5 Phase 5 – Red Teaming & Adversarial Testing

Objective

Validate resistance to misuse before and after deployment.

Required Controls

- Prompt injection testing
- Boundary violation attempts
- Oversized or malformed input testing
- Misuse scenario testing

Artifacts

- Red team test cases
- Test results
- Mitigation evidence

15.6 Phase 6 – Deployment & Runtime Governance

Objective

Control how AI Agents operate in production.

Required Controls

- Authentication and authorization
- Rate limiting and abuse prevention
- Kill switch or emergency disablement
- Explicit prohibition of unauthorized autonomous actions

Artifacts

- Runtime configuration
- Access control policies
- Emergency procedures

15.7 Phase 7 — Monitoring, Auditing & Incident Response

Objective

Detect misuse, maintain accountability, and support audits.

Required Controls

- Structured logging (metadata, not sensitive payloads)
- Monitoring and alerting
- AI Agent-specific incident categorization
- Incident response and rollback procedures

- Severity classification for AI Agent-related incidents to support prioritization and escalation
- Continuous improvement loop back to Phase 1**.**

Artifacts

- Logs and alerts
- Incident records
- Post-incident reviews

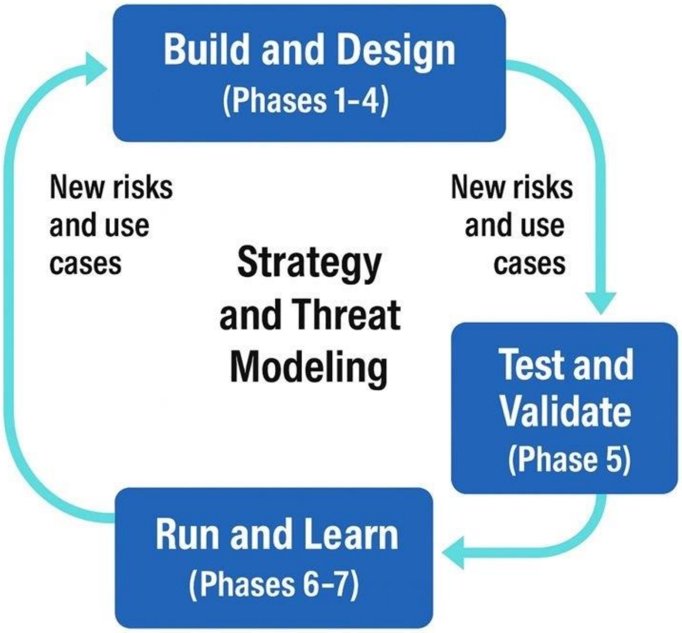


Figure 3 – USA-ADL™ Feedback Loop

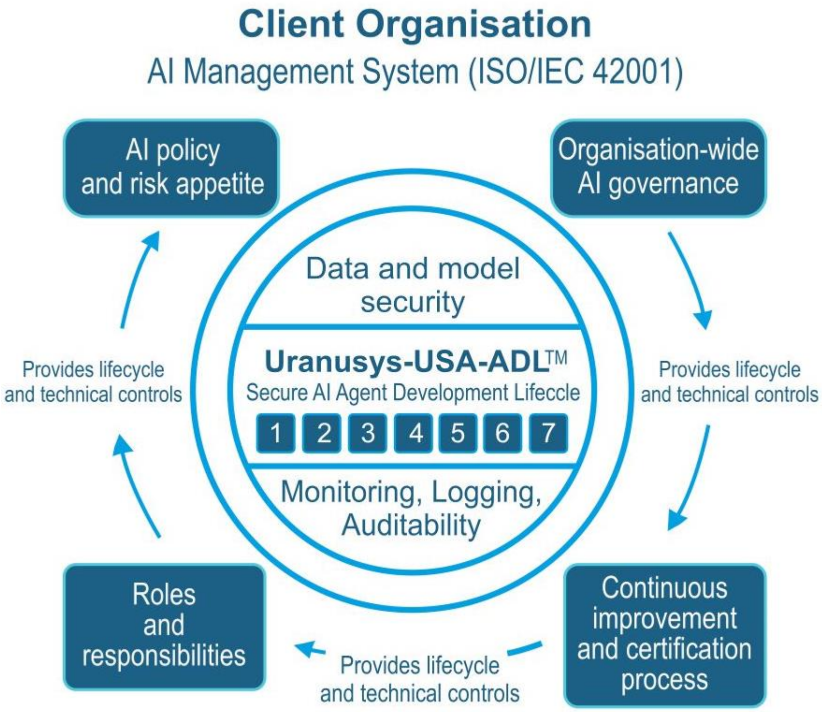
Phase 7 also governs AI Agent decommissioning, including permanent disablement, access revocation, data handling verification, and archival of audit evidence when an AI Agent is retired or withdrawn from service.

16. Evidence and Auditability

Each USA-ADL™ phase produces defined artifacts that serve as auditable evidence of control implementation. These artifacts support internal reviews, client due diligence, third-party assessments, and external audits without exposing sensitive data or proprietary logic.

Evidence artifacts may be reviewed internally, shared under NDA with clients, or presented to auditors in accordance with contractual and legal requirements.

17. Relationship to ISO/IEC 42001


Client Organisation
AI Management System (ISO/IEC 42001)
AI policy and risk appetite
Organisation-wide AI governance
Provides lifecycle and technical controls
Data and model security
Uranusys-USA-ADL™
Secure AI Agent Development Lifeccle
1
2
3
4
5
6
7
Monitoring, Logging, Auditability
Provides lifecycle and technical controls
Roles and responsibilities
Provides lifecycle and technical controls
Continuous improvement and certification process

USA-ADL v1.0 supports organizations seeking alignment with **ISO/IEC 42001 AI Management Systems** by providing a structured lifecycle and technical controls across:

- AI Agent lifecycle governance
- AI Agent risk management and control design
- Data and model governance
- Human oversight and accountability
- Monitoring, logging, and continual improvement

USA-ADL v1.0 satisfies the **technical and lifecycle control expectations** of ISO/IEC 42001. Remaining certification steps are predominantly **organizational and procedural**, not architectural.

Note: For a detailed mapping to ISO/IEC 42001 clauses and control areas, see the separate document "USA-ADL v1.0 – ISO/IEC 42001 Alignment & FAQ".

18. Versioning & Change Control

- USA-ADL v1.0 is locked.
- Changes must be versioned (e.g., v1.1, v2.0).
- Updates must be additive and documented.
- Existing AI Agents are assessed for impact before lifecycle changes apply.

19. Official Status Statement

USA-ADL v1.0 is formally approved as the Secure AI Agent Development Lifecycle governing all Uranusys AI Agents.