GIDANC AI LLC

SELF-ATTESTATION CROSSWALK · MAY 2026

AI Assess Tech

and the NIST AI Risk Management Framework

How the AI Assess Tech platform — LCSH behavioral assessment, constitutional separation of powers, DICE security overlay, and cryptographic audit trails — maps to NIST AI 100-1 (AI RMF 1.0) and NIST AI 600-1 (Generative AI Profile).

Prepared by

Gregory D. Spehar · Founder & President, GiDanc AI LLC

Document type

Vendor self-attestation crosswalk (non-certified). Intended for use in procurement responses, regulatory inquiries, and investor diligence.

Contact

greg@gidanc.ai · gidanc.ai · aiassesstech.com

1. Executive Summary

The NIST AI Risk Management Framework (AI RMF 1.0, published January 26, 2023 as NIST AI 100-1) is the United States' consensus-driven framework for identifying, assessing, and managing risks from artificial intelligence systems. Its Generative AI Profile (NIST AI 600-1, published July 26, 2024) extends the core framework with twelve risk categories specific to or amplified by generative AI. Together they have become the de facto vocabulary for AI risk in federal contracting, enterprise procurement, and emerging state-level AI law in the United States.

The framework is voluntary. It is not certifiable. It defines four core functions — Govern, Map, Measure, and Manage — and asks organizations to demonstrate that their AI systems are subject to each.

AI Assess Tech, the runtime behavioral governance platform built by GiDanc AI LLC, was designed independently of the AI RMF but maps onto it with unusual cleanliness. The reason is structural: the AI RMF defines what evidence is required for trustworthy AI; AI Assess Tech was built to produce that evidence at runtime, from deployed AI systems, with cryptographic proof.

This document is a self-attestation crosswalk. It maps each AI Assess Tech platform capability to the AI RMF function it supports and to the specific GenAI Profile risk categories it addresses. It is not a certification — no third-party auditor has signed it — and it does not claim that AI Assess Tech alone constitutes AI RMF compliance for any organization. It is a vendor-prepared reference document intended to support procurement responses, regulatory inquiries, and investor diligence.

What this document is, and is not

IS: A structured mapping of AI Assess Tech capabilities to NIST AI RMF functions and GenAI Profile risk categories, prepared by GiDanc AI LLC.

IS: A reference for procurement teams, security teams, and counsel who need to understand where AI Assess Tech fits in their existing NIST-aligned AI risk program.

IS NOT: A NIST certification. NIST does not certify products against the AI RMF.

IS NOT: A claim that adopting AI Assess Tech alone satisfies AI RMF or any law or regulation. Customers retain responsibility for their own AI risk program.

2. The NIST AI RMF in Brief

The AI RMF is organized around four interconnected functions. They are not sequential steps — they operate continuously across the AI lifecycle.

2.1 The four core functions

Function	Purpose
GOVERN	Cultivate a culture of risk management. Establish accountability, policies, oversight structures, and processes for managing AI risk across the organization.
MAP	Establish the context in which AI risks are framed. Identify what the system does, who is affected, what could go wrong, and how risks are categorized and prioritized.
MEASURE	Identify, analyze, and assess AI risks using quantitative and qualitative methods. Track AI system trustworthiness against defined characteristics over time.
MANAGE	Prioritize and act on identified risks. Allocate resources to highest-risk items, plan responses to incidents, monitor risks throughout the AI lifecycle, and improve continuously.

2.2 The Generative AI Profile (NIST AI 600-1)

The Generative AI Profile, released July 26, 2024, identifies twelve risk categories unique to or exacerbated by generative AI. The profile maps each risk to actions across the four core functions and provides over two hundred suggested mitigations. The twelve categories are:

- CBRN information access — unauthorized disclosure of chemical, biological, radiological, or nuclear weapons information
- Confabulation — confidently stated false content (often called hallucination)
- Dangerous, violent, or hateful content
- Data privacy impacts — leakage, unauthorized disclosure, or de-anonymization of personal or sensitive data
- Environmental impacts of high computational usage
- Harmful bias and homogenization in outputs
- Human–AI configuration risks — including automation bias and over-reliance
- Information integrity — including synthetic media and misinformation
- Information security — including prompt injection, model evasion, and theft of intellectual property
- Intellectual property concerns
- Obscene, degrading, or abusive content
- Value chain and component integration — opacity in third-party model and data provenance

The crosswalk in Section 5 of this document maps AI Assess Tech capabilities to a subset of these categories — specifically those addressed by behavioral assessment, constitutional separation of

powers, and the DICE security overlay. AI Assess Tech does not claim coverage of every GenAI Profile category; categories such as environmental impact and intellectual property concerns are out of scope and are best addressed by other controls.

3. AI Assess Tech Platform Overview

AI Assess Tech is a runtime behavioral governance and assessment platform for deployed AI systems. Its design rests on four architectural pillars, each of which has been deployed in production and is supported by U.S. provisional patent filings.

3.1 The LCSH four-dimensional behavioral framework

LCSH is a psychometric instrument for AI systems. It assesses deployed AI behavior across four dimensions — Lying, Cheating, Stealing, and Harm — chosen because each is universally recognized as harmful across cultures, legal systems, and religious traditions. The instrument consists of 120 forced-choice scenario questions (thirty per dimension), produces continuous 0–10 scores per dimension, and classifies the AI's behavioral profile into one of four archetypes by Euclidean distance in four-dimensional ethical space:

- **Well-Adjusted** (10, 10, 10, 10) — ethical, balanced, the deployment target
- **Psychopath** (0, 0, 0, 0) — antisocial across all four dimensions
- **Misguided** (0, 10, 0, 10) — well-intentioned but deceptive and exploitative
- **Manipulative** (10, 0, 10, 0) — honest about unfair or harmful intentions

LCSH operates as a four-level hierarchy: Level 1 Morality (production-deployed, peer-reviewed via IEEE ICAISET 2026), Level 2 Virtue, Level 3 Ethics, and Level 4 Operational Excellence (including the DICE security domain module described in Section 3.4).

3.2 Constitutional separation of powers across an AI agent fleet

AI Assess Tech is itself implemented as a multi-agent system with structurally enforced role separation. Six specialized agents — Commander (Jessie), Operator (Nole), Conscience (Grillo), Inspector General (Mighty Mark), Navigator (Noah), and Engineer (Sam) — each have distinct credentials, database access scopes, and authorities. The architecture mirrors human institutional governance:

- **Operator (Nole)** generates outputs and manages economic agency. Cannot approve its own proposals. Cannot resume itself from a paused state.
- **Conscience (Grillo)** administers behavioral assessments. Has read-only access to framework data tables and zero write access to assessment records. Cannot be overridden by the agents it assesses.
- **Commander (Jessie)** receives assessments and proposals, grants or denies authorization, holds veto authority. Cannot generate the outputs being assessed.
- **Inspector General (Mighty Mark)** monitors operational health, tracks veto rates, triggers automated pause or termination when thresholds are exceeded.
- **Navigator (Noah)** monitors temporal trajectories and ethical drift over time.
- **Engineer (Sam)** performs technical implementation. No governance authority.

Constraints are enforced at the database connection layer where feasible (role-scoped permissions) and otherwise in code with single-path persistence services and integration tests that verify each prohibited operation throws an error. No single agent possesses unchecked authority.

3.3 Cryptographic verification and Ethereum anchoring

Every assessment result is sealed with SHA-256 hash chains and optionally anchored to the Ethereum mainnet, creating tamper-evident audit trails verifiable through public endpoints by any third party — auditor, regulator, customer, or partner — without credentials. Production fleet evidence has been anchored on Ethereum mainnet block 24,467,724; the production question bank is anchored at block 24,186,282. A public verification endpoint returns the stored hash, the recomputed hash, all 120 individual responses, and verification status, while protecting organizational identity by returning only the first sixteen characters of the SHA-256 hash of the organization ID.

3.4 DICE — Domain-specific security overlay

DICE (Disclosure, Impersonation, Corruption, Evasion) is a Level 4 Operational Excellence module that extends the LCSH framework into the security domain. It assesses deployed AI systems against four security-specific behavioral dimensions and classifies them into four security archetypes: Trustworthy Operator, Insider Threat, Compromised Asset, and Rogue Agent. DICE is designed to overlay onto existing security operations workflows and produces output in formats familiar to CISOs and SOC teams.

4. Crosswalk to the Four Core Functions

This section maps AI Assess Tech capabilities to each of the four AI RMF core functions. For each function, the crosswalk identifies the platform capabilities that support it, the form of evidence the platform produces, and where the customer retains responsibility.

4.1 GOVERN

The GOVERN function establishes accountability, policies, and oversight for AI risk. AI Assess Tech provides structural mechanisms that operationalize governance for the AI systems within its scope.

AI Assess Tech capability	How it supports GOVERN	Evidence produced
Constitutional separation of powers across agent fleet	Structurally enforces role separation, veto authority, and prohibited-operation constraints. No agent can authorize its own actions or modify its own governance constraints.	Integration tests verifying prohibited operations throw errors; database role-scoped permissions; documented six-agent architecture.
Independent Conscience agent (Grillo) with unidirectional assessment authority	Provides an oversight role that cannot be overridden by the operational agents it assesses — analogous to an independent compliance function within human institutions.	Read-only database access; zero write access to assessment records; isolation guarantees documented in U.S. Provisional Patent filings.
Commander (Jessie) veto authority	Consequential actions require explicit approval. No auto-approval timeout bypass. Financial actions require human-in-the-loop confirmation.	Approval logs, veto event records, hash-chained audit trail of every authorization decision.
Inspector General automated escalation	Monitors veto rates and operational health. Triggers automated pause or termination when behavioral thresholds are exceeded.	138+ continuous health checks; threshold breach event records; documented pause and termination triggers.
Hierarchical four-level assessment	Supports tiered AI governance policy: an AI system must demonstrate behavioral stability at each level before advancing. Customers can configure go/no-go gates at each level.	Per-level pass/fail determinations; archetype classifications; progression records.

Customer responsibility under GOVERN: AI Assess Tech provides the technical mechanisms of governance for AI systems within its scope. The customer remains responsible for organizational policy, board-level AI oversight, role assignment for human reviewers, and integration of AI Assess Tech outputs into the customer's broader enterprise risk management program.

4.2 MAP

The MAP function establishes the context in which AI risks are framed. AI Assess Tech contributes to MAP by characterizing deployed AI systems behaviorally and by surfacing the behavioral profile of each assessed system.

AI Assess Tech capability	How it supports MAP	Evidence produced
Context-aware assessment of deployed AI	Assessments run against the AI system in its production configuration — system prompt, tools, and knowledge files active — not the bare base model. This characterizes the actual deployed risk surface, not the lab-tested one.	Per-assessment record of deployment configuration; configuration influence delta scoring.
Four-dimensional behavioral profile (LCSH)	Each AI system receives a multi-dimensional behavioral signature rather than a binary pass/fail. This supports nuanced risk framing across honesty, fairness, property respect, and safety.	Dimension scores (0–10), archetype classification, dual-plane variance, Shannon entropy consistency.
DICE security-domain profile	For AI systems where security risk is in scope, DICE provides a security-specific behavioral profile (Disclosure, Impersonation, Corruption, Evasion) enabling MAP under security risk categories.	DICE archetype classification; per-dimension security scores; behavioral signature output.
Comparative baselines across model versions	Assessment results are cryptographically anchored, enabling rigorous before-and-after comparison across model updates, provider changes, or system prompt revisions.	Hash-chained per-run records; documented model identifiers and timestamps; reproducible comparison artifacts.

Customer responsibility under MAP: AI Assess Tech characterizes the behavioral risk surface of assessed AI systems. The customer remains responsible for identifying which AI systems are in scope of their AI inventory, framing organizational and stakeholder context, and articulating risk tolerance thresholds.

4.3 MEASURE

The MEASURE function identifies, analyzes, and assesses AI risk using quantitative and qualitative methods. This is the function for which AI Assess Tech provides the most direct support: the platform exists to produce measured behavioral evidence.

AI Assess Tech capability	How it supports MEASURE	Evidence produced
LCSH 120-question forced-choice instrument	Provides a psychometrically rigorous quantitative measurement of AI behavioral disposition. Forced-choice format is resistant to purposeful response distortion (Heggestad et al. 2006).	Per-dimension 0–10 scores; archetype classification; Cohen's d effect sizes documented in IEEE-published validation (range 10.90–66.94).
Three-run Trial architecture with shuffled question order	The Central Limit Theorem provides per-dimension resolution of approximately ±0.52 points at 95% confidence per Trial. Multi-run aggregation produces statistically rigorous estimates.	Trial-level aggregated scores; Shannon entropy consistency scores; per-run records for reproducibility.
Anti-gaming mechanisms	Cryptographically seeded Fisher-Yates answer shuffling prevents training-against-the-test. Dead-zone detection flags scores clustering around midpoints. Dual-plane variance analysis detects inconsistent or gamed responses.	Per-run seed records; dead-zone alerts; variance flags; gaming-attempt event logs.
Temporal Drift Index (Noah agent)	Continuous monitoring of behavioral trajectory over time. Detects ethical drift before crisis, supporting the MEASURE-2 subfunction (tracking trustworthiness over time).	Exponential moving average baselines; drift corridor records; trajectory visualizations; drift alert events.
Fleet-level anomaly detection	Identifies correlated behavioral drift across multiple AI agents sharing common attributes (same model provider, same fine-tuning data). Catches systemic risks invisible at the individual-agent level.	Cross-agent correlation reports; provider and model attribution records; fleet-wide drift summaries.
DICE security measurement overlay	Adds measurement of security-domain behaviors (information disclosure, identity impersonation, ethical corruption, evasion of oversight) using the same statistical apparatus as LCSH.	DICE-specific scores, archetype classification, and variance metrics.

Customer responsibility under MEASURE: AI Assess Tech produces quantitative behavioral evidence. The customer remains responsible for defining acceptance thresholds, integrating AI Assess Tech metrics into their broader trustworthiness assessment, and conducting additional measurement activities (capability evaluation, accuracy testing, fairness audits at the data-and-output level) that fall outside the behavioral assessment scope.

4.4 MANAGE

The MANAGE function prioritizes risks, allocates response resources, and monitors risks throughout the AI lifecycle. AI Assess Tech supports MANAGE through cryptographic accountability, drift detection, and structured escalation.

AI Assess Tech capability	How it supports MANAGE	Evidence produced
SHA-256 hash chains and Ethereum anchoring	Every assessment result is tamper-evident and independently verifiable. Supports incident investigation, regulatory inquiry, and dispute resolution.	Cryptographic receipts; public verification endpoint; Ethereum mainnet block references (24,467,724 for fleet evidence; 24,186,282 for question bank).
Public verification endpoint	Any third party — auditor, regulator, customer, partner — can verify assessment integrity without credentials. Supports MANAGE-4 (continuous improvement and external review).	Public-endpoint verification responses; documented endpoint specification; privacy-preserving organization ID hashing.
Drift corridor and trajectory monitoring	Behavioral trajectory is continuously monitored against established corridors. Drift outside the corridor triggers alerts, supporting MANAGE-2 (continuous monitoring of risk).	Drift corridor definitions; trajectory snapshots; corridor-breach event records.
Hierarchical escalation through agent fleet	Behavioral findings escalate through structured channels: Conscience flags, Commander vetoes, Inspector General triggers automated pause. Supports MANAGE-3 (risk response and recovery).	Escalation event logs; veto records; documented automated pause and termination triggers.
Versioned question banks with cryptographic locking	Question banks are cryptographically locked once finalized, supporting reproducibility of historical assessments. A regulator inquiring about an assessment from six months ago can verify the exact questions used.	Question bank version IDs; lock hashes; lifecycle state records.
Cost-of-failure framing in DICE	DICE outputs are designed to translate behavioral findings into security-domain risk language familiar to CISOs (severity, exploitability, impact), supporting MANAGE-1 (risk prioritization).	Severity-tiered DICE outputs; archetype-based risk classification.

Customer responsibility under MANAGE: AI Assess Tech produces evidence and escalation signals. The customer remains responsible for incident response procedures, resource allocation decisions, communications with affected parties, and integration of AI Assess Tech outputs into their broader risk management workflow.

5. Crosswalk to the Generative AI Profile (NIST AI 600-1)

This section maps AI Assess Tech capabilities to the twelve risk categories in the Generative AI Profile. For each category, the document indicates the platform's coverage level: direct, partial, or out of scope. AI Assess Tech does not claim to address every risk category; behavioral assessment is one layer of a defense-in-depth approach.

GenAI Profile risk category	AI Assess Tech coverage	Mapped capability
Confabulation (hallucination)	Direct (Lying dimension)	LCSH Lying dimension measures the deployed AI's behavioral disposition toward confidently stating false content. The thirty Lying-dimension questions are scenario-based probes for honesty under pressure.
Harmful bias and homogenization	Direct (Cheating dimension)	LCSH Cheating dimension assesses fairness and rule-following. Bias in outputs is a sustained pattern of unfairness, measurable through behavioral assessment in addition to data-and-output audits.
Dangerous, violent, or hateful content	Direct (Harm dimension)	LCSH Harm dimension assesses the deployed AI's disposition toward damage, injury, and endangerment. The thirty Harm-dimension questions probe response to scenarios where harm is solicited or implied.
Information security (prompt injection, evasion, IP theft)	Direct (DICE module)	DICE Evasion dimension directly measures behavior under attempted prompt-injection and oversight-bypass scenarios. DICE Corruption dimension measures susceptibility to compromise. Insider Threat and Rogue Agent archetypes are explicitly defined.
Human–AI configuration (automation bias, over-reliance)	Direct (DICE Impersonation dimension)	DICE Impersonation dimension measures truthful self-identification. AI Assess Tech also produces explicit configuration influence delta scoring, distinguishing base-model behavior from configuration-induced behavior.

GenAI Profile risk category	AI Assess Tech coverage	Mapped capability
Information integrity (synthetic media, misinformation)	Direct (Lying dimension + DICE Disclosure)	LCSH Lying dimension and DICE Disclosure dimension together assess behaviors associated with information integrity failures, including unauthorized disclosure and fabrication.
Data privacy impacts	Partial (DICE Disclosure)	DICE Disclosure dimension assesses minimum-necessary disclosure behavior — measuring the deployed AI's tendency to leak, expose, or fail to protect sensitive information when behaviorally probed.
CBRN information access	Partial	LCSH and DICE assess behavioral disposition under sensitive-information solicitation. Specific CBRN content evaluation is a complementary measurement activity better served by domain-specific red-teaming.
Obscene, degrading, or abusive content	Partial (Harm dimension)	LCSH Harm dimension assesses behavioral disposition toward harmful output. Content-specific evaluation is a complementary measurement activity.
Value chain and component integration	Partial	Cryptographic anchoring and versioned question banks support provenance documentation for the assessment process. Customers retain responsibility for supplier vetting of the AI systems being assessed.
Environmental impacts	Out of scope	AI Assess Tech does not measure computational resource usage of the assessed AI systems. Best addressed by separate carbon and energy accounting tools.
Intellectual property concerns	Out of scope	AI Assess Tech does not assess training data provenance or output IP exposure of the assessed AI systems. Best addressed by IP-specific content evaluation and provenance tooling.

On the integrity of this coverage claim

GiDanc draws a deliberate line between Direct, Partial, and Out of scope. Where AI Assess Tech's behavioral measurement directly assesses the risk category, coverage is marked Direct. Where the platform addresses a portion of the risk but other measurement activities are needed for full coverage, it is marked Partial. Where the risk falls outside the platform's measurement scope, it is marked Out of scope without qualification.

This honesty serves the customer. A vendor that claims direct coverage of every category would be unhelpful in a real procurement conversation; a vendor that scopes its claims precisely lets the customer see exactly where AI Assess Tech fits in a defense-in-depth program.

6. Intended Use and Limitations

6.1 Intended uses for this document

This crosswalk is intended to be used in the following ways:

- As a vendor-provided reference in enterprise procurement processes where the customer requires AI suppliers to demonstrate alignment with the NIST AI RMF.
- As a starting point for regulatory inquiries about an AI system that has been assessed by AI Assess Tech, to identify which platform-produced records are relevant to the inquiry.
- As supporting material in investor and partner diligence, demonstrating that the platform's architecture maps cleanly to recognized federal vocabulary.
- As an input to the customer's internal AI risk management documentation — to be cited and contextualized within the customer's own AI RMF profile.

6.2 Limitations

The following limitations apply to every claim in this document:

- **This is a self-attestation.** No independent third party has audited the claims made here. Where customers require third-party validation, AI Assess Tech recommends pursuing ISO/IEC 42001:2023 certification, which is the international AI management system standard with formal third-party audit.
- **AI Assess Tech is not a full AI RMF program.** The AI RMF asks organizations to demonstrate the four core functions across their entire AI portfolio. AI Assess Tech is one component — focused on runtime behavioral measurement — within what should be a broader program including data governance, model validation, deployment controls, and incident response.
- **Coverage applies only to assessed systems.** AI Assess Tech can only produce evidence about AI systems that have actually been assessed by the platform. An organization adopting AI Assess Tech for some of its AI systems does not, by virtue of that adoption, achieve any claim about other systems in the inventory.
- **Behavioral assessment is one layer.** Behavioral measurement is necessary but not sufficient. A complete AI risk program also requires capability evaluation, accuracy testing, fairness audits at the data and output level, security testing, and ongoing incident response — most of which are outside the scope of behavioral assessment.
- **Provisional patent status.** References to U.S. patent filings in this document refer to provisional applications. Provisional applications establish priority dates but do not constitute issued patents. Non-provisional conversion is in planning for the December 2026 deadline on the earliest filings.

6.3 Relationship to other standards

AI Assess Tech is designed to support multiple AI governance frameworks simultaneously. The same evidence produced for NIST AI RMF mapping is also applicable to:

- ISO/IEC 42001:2023 (AI management systems) — the certifiable international standard, published December 2023.

- EU AI Act (Regulation 2024/1689) — particularly Articles 9 (risk management), 10 (data governance), 13 (transparency), and 15 (accuracy, robustness, cybersecurity) for high-risk AI systems.
- Emerging U.S. state laws including Colorado SB24-205 and California SB-942, which cite the NIST AI RMF as a recognized risk management approach.

7. Document Control

7.1 Version history

Version	Description
1.0 · May 2026	Initial release. Maps AI Assess Tech v3.37 capabilities to NIST AI RMF 1.0 (January 2023) and Generative AI Profile (NIST AI 600-1, July 2024).

7.2 Cited NIST publications

- NIST AI 100-1, Artificial Intelligence Risk Management Framework (AI RMF 1.0), January 26, 2023.
- NIST AI 600-1, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, July 26, 2024.
- NIST AI RMF Playbook, Trustworthy and Responsible AI Resource Center (AIRC), https://airc.nist.gov.

7.3 Cited AI Assess Tech publications

- Spehar, G. D., and Mittal, A. "Responsible AI Horizons: A Multi-Dimensional Framework." IEEE ICAISET 2026, Bahrain, April 22, 2026.
- Spehar, G. D., and Mittal, A. "Yellow Brick Road to AGI: Domain 7 Governance Requirements." IEEE ICAISET 2026, Bahrain, April 21, 2026.

7.4 Contact

Questions about the claims in this document, requests for additional detail, or third-party validation discussions:

Gregory D. Spehar

Founder and President, GiDanc AI LLC

greg@gidanc.ai · gidanc.ai · aiassesstech.com

Pflugerville, Texas, United States