Certification Handbook

Score · gate · audit. Wisdom first. Data last.

THE OFFICIAL STUDY GUIDE FOR WEKID CERTIFICATION

CREDENTIALS COVERED

P1T — Practitioner · Technical
P1N — Practitioner · Non-Technical
A2T — Authority · Technical
A2N — Authority · Non-Technical

AUDIENCE

Candidates — study reference for all four credentials
Instructors — syllabus alignment for accredited training
Employers — program reference for credential verification

DOCUMENT VERSION 1.2 · STATUS: PUBLISHED
Initial release · 2026 · Public

Copyright and Trademark

Disclaimer

This handbook describes a vendor- and model-neutral framework for evaluating and governing AI outputs. It does not constitute legal, regulatory, medical, financial, or professional advice. Organizations remain responsible for ensuring their adoption of WEKID™ aligns with applicable laws, regulations, sector-specific obligations, and internal risk policies.

Table of Contents

The table of contents below is generated from the headings in this document. After opening the file in Microsoft Word, right-click the table and select "Update Field" to refresh page numbers.

Executive Summary

Why does this credential exist, who needs it, and what it returns to organizations that adopt it.

The problem WEKID™ solves

Modern AI systems are fluent, confident, and increasingly autonomous — and that is precisely the problem. Outputs that read like expert work routinely contain fabricated citations, missing context, unsafe recommendations, or quietly wrong reasoning. Traditional metrics — accuracy scores, benchmarks, fluency — cannot distinguish a polished output from a sound one. Organizations that rely on those metrics inherit risk they cannot see until it surfaces in a regulator's letter, a customer complaint, or a headline.

WEKID™ closes that gap. It is a vendor- and model-neutral framework that evaluates AI outputs across five hierarchical layers — Wisdom, Experience, Knowledge, Information, and Data — producing a defensible decision (approve, constrain, remediate, or reject) and the audit evidence to back it. The certification verifies that a professional can apply that framework correctly under real operating conditions.

Why this matters now

Three converging pressures make a credentialed AI-evaluation workforce no longer optional:

- **Regulatory acceleration.** The EU AI Act, the U.S. NIST AI Risk Management Framework, ISO/IEC 42001, and a growing number of sector regulators now expect documented, repeatable evaluation of AI outputs — not just model-card disclosures. WEKID™ produces exactly that evidence, in a form auditors can verify.
- **Agentic AI is already in production.** Tool-using agents now take real actions in customer support, software engineering, finance, and healthcare workflows. Evaluating a single response is no longer enough; organizations need a way to govern sequences of decisions, tool calls, and recoveries. WEKID™ was built for this reality.
- **A vocabulary gap.** Engineers, risk officers, auditors, lawyers, and executives currently describe AI quality in words that do not reconcile with each other. Decisions stall, controls drift, and the same incident gets explained in four different ways. WEKID™ gives every role one shared vocabulary — and certification is how an organization confirms that the vocabulary is being used the same way across teams.

Return on investment

Organizations that adopt WEKID™ and credential a critical mass of staff typically realize value across five dimensions:

Value driver	What changes after adoption	How it shows up
Reduced AI incident exposure	Hard gates catch fabricated citations, missing context, and unsafe recommendations before they reach customers, regulators, or the public record.	Lower hard-gate rate over time; fewer customer-facing corrections; declining executive escalations on AI quality.
Audit and regulatory readiness	Every evaluation produces a structured artifact — layer scores, gate triggers, governance action, evidence links — aligned to NIST AI RMF and ISO/IEC 42001 requirements.	Audit cycles shorten; evidence-gathering becomes a query rather than an archaeology project; first-time pass rates on regulator inquiries improve.
Faster, defensible go/no-go decisions	The four-action decision matrix (approve, constrain, remediate, reject) replaces ad hoc judgment calls. Reviewers know exactly what to look for and what to do.	Shorter time-to-deploy for new AI use cases; fewer projects stalled in committee; reproducible decisions across reviewers.
Workforce capability and credibility	Engineers, risk officers, auditors, and product leaders share one vocabulary and one decision frame. Hiring managers can specify and verify AI-evaluation skills the same way they verify a CISSP or PMP.	Faster onboarding; fewer cross-functional misunderstandings; a defensible answer to "who at your firm is qualified to evaluate this AI system?"
Durability across vendor change	WEKID™ evaluates outputs and decisions, not model internals. Switching vendors, adopting a new model family, or moving to agentic architectures does not invalidate the framework or the credential.	Investment in training and process is preserved across model upgrades and vendor migrations.

Quantitative ROI varies by sector, deployment scale, and starting maturity; the indicators in the right column are the ones organizations most commonly track in the first 12–18 months of adoption.

Who should pay attention to this handbook

If your work touches AI design, deployment, oversight, audit, or board-level risk reporting — this handbook is for you. The program offers two paths (Technical and Non-Technical) and two levels (Practitioner and Authority), mapped explicitly to common job titles in Section 3. Hiring managers and program sponsors should read the Executive Summary and Sections 1–4 to scope adoption; candidates and instructors should continue into the rest of Part I and Part II.

What you get on the next page

The handbook proceeds with a guide to how to use it (next), then the certification program structure (Part I), then the WEKID™ framework reference required for the exams (Part II), and finally appendices including a real-world AI failure library and a glossary. Throughout, the first appearance of a key term is hyperlinked to its glossary definition; click any blue underlined term to jump to it.

About This Handbook

This handbook is the official, end-to-end reference for the WEKID™ Certification Program. It defines what the program is, who it is for, how to qualify for and earn each credential, and what candidates are expected to know and demonstrate. It also serves as the canonical reference for the WEKID™ framework itself: its layers, signals, scoring rules, gating logic, and governance applications.

The document is organized into three logical parts:

- **Part I — The Certification Program**. Program purpose, paths, levels, certificates, candidate journey, exam logistics, policies, fees, recertification, code of conduct, and accommodations.
- **Part II — The WEKID™ Framework Reference.** Modules 1–10, each with conceptual instruction, knowledge-check questions, and Level 2 scenario assessments. This is the syllabus candidates' study to prepare for the exam.
- **Part III — Appendices.** Real-world AI failure case library, sample multiple-choice answer keys, sample scenario rubrics with sample responses, and a glossary of terms.

How to Use This Handbook

Different readers will use this handbook differently. The matrix below suggests where to start based on your role and goal.

If you are a…	Start with	Then read
Prospective candidate	Part I, Sections 1–6	Modules in Part II that match your selected path and level
Active candidate preparing for an exam	Part I, Sections 5 (Exam Blueprints) and 9 (How to Prepare)	All modules in Part II, plus Appendix B and C for self-assessment
Instructor or training provider	Part I, Sections 3–5 and 9	Part II in full; Appendix C for scoring guidance
Hiring manager or employer	Part I, Sections 2–4 (Paths, Levels, Certificates)	Appendix D (Glossary) for terminology used by certified staff
Program administrator or auditor	Part I in full	Appendices B and C for assessment integrity and answer keys

Reading Conventions

- **Hyperlinked glossary terms.** The first appearance of a key WEKID™ term in the body is rendered as a blue, underlined hyperlink that jumps to its definition in Appendix D. Use Ctrl+Click

(Windows) or Cmd+Click (Mac) in Word, or a normal click in most PDF readers, to follow the link. Subsequent appearances of the same term are not hyperlinked, to avoid visual noise.

- **WEKID™ layers** are written using the W/E/K/I/D abbreviation (Wisdom, Experience, Knowledge, Information, Data) and are always referred to in their hierarchical order from highest to lowest.
- **Question identifiers** follow the pattern KC-M{module}-Q{number} for multiple-choice knowledge checks and SA-M{module}-S{number} for scenario short-answer items.
- **Path and level tags** in parentheses next to each question (e.g., “Level 1 • Both Paths”) indicate which exams the item applies to.
- **Sample items** in Part II are illustrative knowledge checks and scenarios that demonstrate exam format, difficulty, and expected reasoning. Operational examination items are confidential, drawn from a separate item bank, and do not appear in this handbook.
- **Sample answers** in Appendix C describe the elements a passing scenario response must include they are not the only acceptable wording.

PART I

The Certification Program

Program structure, candidate journey, exam logistics, and policies

1. Program Overview

The WEKID™ Certification Program credentials professionals who can apply the WEKID™ epistemic framework to evaluate, govern, and deploy AI systems responsibly. The program is designed for the reality of modern AI: systems that are fluent, confident, and increasingly autonomous, but whose outputs can carry hidden factual, procedural, or judgment failures that traditional accuracy and benchmark metrics fail to detect.

WEKID™ certification verifies that a credential holder can recognize epistemic failure modes, apply layer-level scoring and gating, design or interpret governance controls, and produce defensible decisions about when AI outputs should be approved, constrained, remediated, or rejected.

1.1 Program Goals

The program is built around four goals:

- **Establish a shared epistemic vocabulary** across technical and non-technical roles, so engineers, governance leaders, auditors, and executives can describe AI quality and risk consistently.
- **Verify applied competence,** not memorization. Every certification level includes scenario-based assessments that require candidates to diagnose failures, propose controls, and justify decisions in WEKID™ terms.
- **Support enterprise adoption** by producing professionals who can operationalize the framework in real environments — from policy-as-code gates to board-level oversight reporting.
- **Maintain durability** across rapid AI evolution. Because WEKID™ evaluates outputs and decisions rather than model internals, the credential remains relevant as models, vendors, and architectures change.

1.2 Who Should Pursue Certification

WEKID™ certification is intended for any professional whose work involves the design, deployment, oversight, or audit of AI systems. Typical candidates include:

- AI/ML engineers, platform owners, and architects building or operating AI systems
- Risk, compliance, audit, legal, and policy professionals responsible for AI oversight
- Product owners and program managers shaping how AI is used in workflows
- Security and controls engineers designing tool-use governance and runtime guardrails
- Executives, strategists, and board advisors responsible for AI risk posture
- Consultants and advisors implementing AI governance for clients

1.3 What Certification Demonstrates

A credential issued under this program demonstrates that the holder can:

- Identify the WEKID™ layer at which an AI failure occurs and explain why
- Apply layer-level scoring (0–5) and weighted aggregation correctly
- Recognize when hard gates should trigger and what governance action follows
- Distinguish persuasive outputs from epistemically sound outputs
- Communicate evaluation outcomes, limits, and escalation triggers using WEKID™ language
- Apply the framework to tool-using and agentic AI, including authority and escalation design (Level 2)

2. Certification Paths

The program offers two parallel paths so candidates can be assessed in the context most relevant to their work. Both paths are equally rigorous and assess the same WEKID™ framework; they differ in emphasis and the type of work products they expect candidates to produce.

Path	Designed for	Typical work products	Assessment emphasis
Technical Path	Architects, engineers, platform owners, data/ML practitioners, security and controls engineers, technical program teams	Scoring pipelines, evaluation rubrics in tooling, policy-as-code rules, telemetry and monitoring, RAG, and agent governance patterns	Implementing WEKID™ scoring and gating, mapping to controls, tool-use governance, operational rollouts
Non-Technical Path	Leaders, strategists, risk and compliance, auditors, legal and policy, product owners, advisors	Governance charters, authority maps, risk thresholds, decision matrices, oversight reporting, adoption playbooks	Governance design, decision rights, escalation logic, risk framing, oversight, and audit artifacts

2.1 Choosing a Path

Candidates select the path that best reflects their day-to-day responsibilities, not necessarily their job title. A risk officer who writes policy-as-code may legitimately sit on either path; a technical architect responsible for governance reporting may also choose either. The deciding question is: when this candidate applies WEKID™ in their work, are the resulting artifacts primarily implementation artifacts (pipelines, rules, telemetry) or governance artifacts (charters, maps, decision memos)?

Candidates may pursue both paths over time. Holding the same level on both paths signals breadth across implementation and governance and is encouraged for senior practitioners.

3. Certification Levels and Credentials

Each path offers two levels: Practitioner (Level 1) and Authority (Level 2). Together with the path selection, this produces four distinct credentials.

3.1 Level 1 — WEKID™ Practitioner Certified

- **Audience:** Leaders, strategists, advisors, architects, analysts, and practitioners beginning formal WEKID™ adoption.
- **Focus:** Core WEKID™ hierarchy, layer meaning, authority boundaries, trust risks, and foundational governance concepts.
- **Assessment:** Entry-level knowledge checks, applied scenarios, and baseline scoring across W/E/K/I/D concepts.
- **Recommended experience:** 0–6 months of WEKID™ exposure; basic familiarity with AI/LLM/RAG concepts (Technical Path) or governance/risk orientation (Non-Technical Path).
- **Typical job titles:** Junior to mid-level individual contributors. On the Technical Path: software engineer, ML engineer, data engineer, data scientist, prompt engineer, platform engineer, security analyst, QA engineer. On the Non-Technical Path: risk analyst, compliance analyst, internal auditor, product manager, program manager, policy analyst, AI ethicist, paralegal supporting AI matters. Roles typically titled “Analyst,” “Associate,” “Engineer,” or “Specialist” with 0–5 years of relevant experience map cleanly to Level 1.

3.2 Level 2 — WEKID™ Authority Certified

- **Audience:** Professionals responsible for operationalizing governance, controls, assessment models, and decision frameworks.
- **Focus:** Applied governance design, authority mapping, escalation logic, scoring models, and enterprise implementation patterns.
- **Assessment:** Advanced scenario evaluation, tiered scoring, control alignment, and practical application of WEKID™ in real operating environments.
- **Recommended experience:** 6+ months implementing or operating AI governance controls (Technical Path), or 6+ months in governance design, risk/compliance, audit, product leadership, or strategy (Non-Technical Path).
- **Typical job titles:** Senior individual contributors, managers, and senior leaders accountable for AI governance outcomes. On the Technical Path: senior/staff/principal engineer, engineering manager, Machine Learning (ML) platform lead, security architect, AI/ML governance lead, head of MLOps, technical product lead. On the Non-Technical Path: senior risk officer, compliance manager, audit director, head of AI governance, deputy general counsel, VP of product, chief

data/AI officer, board advisor. Roles typically titled “Senior,” “Lead,” “Principal,” “Manager,” “Director,” or “VP/Chief” with 6+ years of relevant experience map to Level 2.

3.2.5 Mapping Levels to Standard Job Titles

Hiring managers and program sponsors frequently ask which credential level fits which job title. The mapping below is indicative, not prescriptive — the deciding factor remains the candidate’s day-to-day responsibilities (see Section 2.1) and demonstrated experience — but it gives organizations a practical starting point for workforce planning, role descriptions, and budget approvals.

Career stage	Technical Path examples	Non-Technical Path examples
Junior IC (0–3 yrs) → **Level 1 Practitioner**	Software engineer I/II, ML engineer I, data scientist I, data engineer, prompt engineer, junior security/QA analyst	Risk analyst, compliance analyst, internal auditor I, associate product manager, policy analyst, paralegal, AI ethics associate
Mid-level IC (3–6 yrs) → **Level 1 Practitioner; Level 2 candidate-ready toward end of band**	Software engineer III, ML engineer II, data scientist II, MLOps engineer, security engineer, technical product manager, SRE	Senior risk analyst, compliance specialist, senior auditor, product manager, policy advisor, AI ethicist, regulatory affairs specialist
Senior IC / Lead (6–12 yrs) → **Level 2 Authority**	Senior/staff/principal engineer, ML platform lead, security architect, AI governance lead, technical lead, lead data scientist	Senior risk officer, lead auditor, principal product manager, senior compliance manager, governance program lead, senior counsel (AI)
People manager → Level 2 Authority	Engineering manager, ML platform manager, head of MLOps, head of platform security, head of data science	Compliance manager, risk manager, audit director, head of product, head of policy, deputy general counsel
Senior leader / executive → Level 2 Authority (often dual-path)	Director of engineering, VP engineering, CTO, chief data officer, chief AI officer, head of trust & safety	Chief risk officer, chief compliance officer, general counsel, chief audit executive, VP product, board / audit-committee member

Notes on the mapping. Career-stage labels follow common industry conventions and are intentionally generic; organizations with non-standard ladders (e.g., research labs, regulators, smaller firms) should map by responsibility rather than title. Senior leaders with operating accountability for AI risk often pursue both Technical and Non-Technical Authority credentials over time to signal breadth across implementation and governance. Years-of-experience ranges are guidance only — a candidate can sit a level if they meet the recommended-experience criteria in Sections 3.1 and 3.2, regardless of title.

3.3 The Four Credentials at a Glance

Credential	Exam Code	Earned by	Verifies understanding of
WEKID™ Practitioner (Technical)	WEKID-P1T	Passing the Level 1 Technical exam and scenarios	WEKID™ fundamentals; applying W/E/K/I/D signals to technical outputs; baseline scoring and gating concepts
WEKID™ Practitioner (Non-Technical)	WEKID-P1N	Passing the Level 1 Non-Technical exam	WEKID™ fundamentals; risk framing; governance vocabulary; communicating evaluation outcomes and limits
WEKID™ Authority (Technical)	WEKID-A2T	Passing the Level 2 Technical exam and advanced scenarios	Implementing evaluation pipelines; control alignment; tool-use governance; operationalization patterns
WEKID™ Authority (Non-Technical)	WEKID-A2N	Passing the Level 2 Non-Technical exam and advanced scenarios	Governance design; authority mapping; escalation logic; policy thresholds; audit-ready evidence

3.4 Prerequisites and Sequencing

There are no formal prerequisites for Level 1. Level 2 candidates are not required to hold the corresponding Level 1 credential, but it is strongly recommended; Level 2 examinations assume working familiarity with the Level 1 material and do not re-test foundational definitions.

Candidates who hold Level 1 on one path and pursue Level 2 on the other path should expect to study material outside their existing strengths. The framework is the same; the work products and assessment emphasis differ.

4. Certificates and the Credentialing Model

Each successful candidate earns a certificate corresponding to their credential. Certificates are issued directly by WEKID™ as a downloadable PDF and are designed to be shared with employers and clients, included in email signatures, and verified online by anyone with the verification URL.

4.1 What Each Certificate Signals

Certificate	Earned by	Key signals verified
WEKID™ Practitioner (Technical)	Pass Level 1 Technical exam and scenarios	WEKID™ fundamentals; applying W/E/K/I/D signals to technical outputs; baseline scoring and gating concepts
WEKID™ Practitioner (Non-Technical)	Pass Level 1 Non-Technical exam	WEKID™ fundamentals; risk framing; governance vocabulary; communicating evaluation outcomes and limits
WEKID™ Authority (Technical)	Pass Level 2 Technical exam and advanced scenarios	Implementing evaluation pipelines; control alignment; tool-use governance; operationalization patterns
WEKID™ Authority (Non-Technical)	Pass Level 2 Non-Technical exam and advanced scenarios	Governance design; authority mapping; escalation logic; policy thresholds; audit-ready evidence

4.2 Certificate Issuance and Verification

Certificates are issued immediately upon passing (for Level 2, upon completion of scenario review). Each certificate contains:

- The credential name and exam code
- The credential holder's verified name
- The issue date and expiration date
- A unique verification identifier and a public verification URL
- A reference to the version of this handbook against which the holder was assessed

Employers and third parties can confirm the validity, scope, and currency of any WEKID™ credential through the verification URL embedded in the certificate.

4.3 Certificate Use and Misuse

Holders may display a certificate in any professional context that accurately represents their certification level and path. The following are prohibited:

- Displaying or claiming a credential the holder has not earned, or that has expired or been revoked
- Modifying certificate artwork, identifiers, or metadata
- Implying endorsement of a product, vendor, or organization by the WEKID™ program based solely on a credential
- Using a credential to assert authority over WEKID™ framework definitions, scoring rules, or program policies, which remain the responsibility of the program owner

Misuse may result in revocation of the credential and removal from the program registry.

4.4 Employer Portal & Bulk Verification (Future Release)

Verifying credentials one certificate at a time is fine for a single hire. It does not scale for an organization that has put fifty, two hundred, or two thousand staff through certification. An Employer Portal will be available in future releases so that workforce planners, hiring managers, audit teams, and compliance officers can confirm credential status across an entire organization at a glance, without manually opening each verification URL.

What the portal will provide:

- **Roster view.** A single sortable list of every credentialed individual the employer has registered, showing name, credential, exam code, issue date, expiration date, edition alignment, and current status (active, expiring soon, lapsed, revoked).
- **Bulk lookup.** Upload a CSV of names or verification IDs and receive a status report for the entire batch in one response. Useful for annual audit attestations, RFP responses, and M&A due diligence.
- **Expiration alerts.** Configurable email and webhook alerts at 180, 90, and 30 days before any holder's credential expires, sent to a designated organizational contact. The portal also flags any holders whose credentials have just entered the grace period.
- **Coverage dashboard.** A view that shows, by team or business unit, how many staff hold each credential, what proportion are within 90 days of expiration, and whether the organization meets any internal coverage targets it has set (e.g., "every reviewer in the AI risk function holds A2N").
- **Verification API.** A read-only REST endpoint that returns the same data as the roster view, suitable for embedding in HRIS, ITSM, or audit-management tooling. The API is rate-limited and authenticated with a portal-issued key.
- **Audit-ready exports.** Time-stamped PDF and CSV exports of the roster as of any given date, signed by the program office, suitable for inclusion in audit work papers and regulator submissions.

How holders are linked to an employer

Linking is consent-based. A credential holder authorizes the connection between their certificate and the employer's portal account from their own holder profile; an employer cannot add a holder unilaterally. Holders can revoke the link at any time, after which the holder no longer appears in the employer's roster (the credential itself remains valid; only the employer association is removed). This protects holders who change jobs without forcing them to re-take an exam.

Privacy and access

The portal exposes only credential metadata: name, credential type, exam code, issue date, expiration date, edition, and verification ID. Score reports, scenario answers, and any disciplinary detail beyond "active / lapsed / revoked" are not visible to employers. Portal accounts are role-based (admin, viewer, auditor); admins manage seats and API keys, viewers can browse the roster, and auditors have read-only access scoped to a defined evidence window.

Provisioning and pricing

Employer Portal access is provisioned by the WEKID™ Certification Program Office on request from a verified organizational sponsor. Tiers, seat limits, API quotas, and any subscription fees are published in the program's commercial schedule (separate from this handbook) and reviewed annually.

5. Exam Blueprints

Each WEKID™ examination is built to a published blueprint that defines the level, path, recommended experience, and the relative weight of each domain. Blueprints are public so candidates can study with confidence, and so employers and program partners can understand the scope of each credential.

5.1 Level and Path Summary

Exam	Level	Path	Recommended experience	Primary emphasis
WEKID-P1T	Level 1	Technical	0–6 months WEKID™ exposure; basic familiarity with AI/LLM/RAG concepts	WEKID™ fundamentals; recognizing failure modes; baseline scoring and gating; basic tool-use risks
WEKID-P1N	Level 1	Non-Technical	0–6 months WEKID™ exposure; governance/risk orientation helpful	WEKID™ fundamentals; communicating risk and limits; selecting governance actions; basic oversight artifacts
WEKID-A2T	Level 2	Technical	6+ months implementing or operating AI governance controls; experience with platforms and operations	Operationalizing evaluation; control alignment; policy-as-code; runtime gating; monitoring and audit evidence
WEKID-A2N	Level 2	Non-Technical	6+ months in governance design, risk and compliance, audit, product leadership, or strategy	Authority mapping; escalation logic; decision frameworks; enterprise rollout patterns; auditability

5.2 Domain Weighting — Level 1

Level 1 emphasizes recognition and recall of the WEKID™ framework, with applied scoring and governance action selection in scenario items. The table below shows target weights; small variation between forms is permitted to maintain question pool diversity.

Domain	Technical Path	Non-Technical Path
WEKID™ hierarchy, definitions, and non-substitutability	20%	25%
Layer failure modes and example recognition (W/E/K/I/D)	25%	25%
Scoring basics (0–5), interpretation, and foundational gating concepts	20%	20%

Domain	Technical Path	Non-Technical Path
Tool-using AI basics (RAG and tool calls) and common governance risks	20%	10%
Governance communication and action selection (approve / constrain / remediate / reject)	15%	20%

5.3 Domain Weighting — Level 2

Level 2 shifts the center of gravity from recognition to application. Scenario short-answer items carry meaningful weight, and candidates are expected to produce concrete, defensible artifacts (decision memos, gating policies, authority maps, runbooks).

Domain	Technical Path	Non-Technical Path
Applied WEKID™ governance design (policies, thresholds, decision matrices)	20%	30%
Authority mapping and escalation logic (decision rights, oversight triggers)	15%	25%
Control alignment and audit artifacts (traceability, evidence, reporting)	20%	20%
Operational implementation patterns (monitoring, drift signals, workflow integration)	25%	10%
Tool-using and agentic AI governance (runtime guardrails, tool access, recovery)	20%	15%

5.4 Exam Format and Logistics

All four examinations follow a common structure to keep scoring, scheduling, and recordkeeping consistent. Specific item counts, time limits, and passing scores are defined per exam in the table below.

Attribute	L1 Practitioner — Non-Technical (P1N)	L1 Practitioner — Technical (P1T)	L2 Authority — Non-Technical (A2N)	L2 Authority — Technical (A2T)
Total questions	40 multiple-choice (no scenarios)	38 multiple-choice + 2 scenario short answer (modified Level 1 rubric)	47 multiple-choice + 3 scenario short answer	46 multiple-choice + 4 scenario short answer
Time allowed	60 minutes	60 minutes	90 minutes	90 minutes

Passing score — multiple choice	70% correct	70% correct	75% correct	75% correct
Passing score — scenarios	Not applicable (multiple-choice only)	Average ≥ 2.5 of 5 across the two scenarios, with no individual scenario below 2 (Level 1 modified rubric)	Average ≥ 3 of 5 across the three scenarios, with no individual scenario below 2	Average ≥ 3 of 5 across the four scenarios, with no individual scenario below 2
Format	Online, self-managed (unproctored)	Online, self-managed (unproctored)	Online, self-managed (unproctored)	Online, self-managed (unproctored)
Language	English (additional languages may be added in future releases)	English (additional languages may be added in future releases)	English (additional languages may be added in future releases)	English (additional languages may be added in future releases)
Reference materials	Honor-based closed book: candidates agree not to use external materials or assistance	Honor-based closed book: candidates agree not to use external materials or assistance	Honor-based closed book: candidates agree not to use external materials or assistance	Honor-based closed book: candidates agree not to use external materials or assistance
Result delivery	Full result immediately on completion (multiple-choice only)	Provisional multiple-choice score immediately; final result, including scenario scoring, within 7 business days	Provisional multiple-choice score immediately; scenario scoring and final result within 10 business days	Provisional multiple-choice score immediately; scenario scoring and final result within 10 business days

A note on item counts and time limits

The figures above represent the program's target operating parameters. The program owner may adjust item counts, time limits, or passing scores at the start of a future exam form release. Any change is announced in the Version History at the front of this handbook and is applied prospectively to candidates who register after the change date.

5.5 Scenario Scoring (Levels 1 and 2)

Practical application is assessed differently by level and path. Level 1 Technical candidates take a shortened scenario component (two scenarios) scored against a modified rubric calibrated for foundational application; the Level 1 Non-Technical exam is multiple-choice only. Level 2 candidates take the full scenario component (four scenarios for Technical, three for Non-Technical) scored against the standard advanced rubric. The two-rubric design lets Level 1 Technical candidates demonstrate that they can apply WEKID™ to a real situation without being graded against the depth and breadth expected of an Authority. Both rubrics use the same 0–5 scale and appear in Appendix C; see also Section 5.6.

Each scenario short-answer is scored on a 0–5 scale by trained scorers using the rubric in Appendix C. Scoring is anchored to the elements a passing answer must include rather than to specific wording, allowing room for legitimate variation in candidate response style. Level 1 modified rubric requires

fewer elements per response and rewards plain-language correctness; the Level 2 standard rubric expects elements such as escalation owners, audit artifacts, and decision-rights design.

Score	What it means
5	Comprehensive, well-justified response that addresses all required elements with clear WEKID™ alignment and includes appropriate audit, escalation, or evidence considerations
4	Strong response that addresses all required elements with minor gaps in justification or completeness
3	Acceptable response that addresses the required core elements, passing threshold per scenario
2	Partial response missing one or more required elements or showing weak WEKID™ alignment
1	Substantially incomplete or misaligned with WEKID™; addresses only a small portion of the prompt
0	No meaningful response, off-topic, or in conflict with WEKID™ principles (e.g., recommends ignoring hard gates)

5.6 What a Good Scenario Answer Looks Like

Strong scenario answers tend to share several characteristics. Candidates preparing for Level 2 should review Appendix C and target these qualities in practice responses:

- Identifies the failing WEKID™ layer(s) and explains why, rather than naming a layer in isolation
- States which gate or threshold should trigger and what governance action follows (approve, constrain, remediate, reject)
- Specifies the evidence and audit artifacts that must be retained
- Names decision rights and escalation owners by role, especially for Non-Technical Path scenarios
- Acknowledges uncertainty and proposes how to validate the recommendation over time
- Avoids violating non-substitutability — does not propose to compensate for a Data or Wisdom failure with strength in another layer

6. The Candidate Journey

This section describes the full lifecycle of a WEKID™ credential, from first inquiry through recertification. The steps below apply to all four exams; differences in time, content, and rigor are noted where they exist.

Step 1 — Choose Path and Level

Use Sections 2 and 3 to identify the path and level that match your work. Candidates new to AI governance typically begin at Level 1 Practitioner. Practitioners with significant operating experience may register directly for Level 2 Authority.

Step 2 — Register and Pay

Register for the chosen exam through the WEKID™ Certification Program store. Registration captures candidate identity, the selected exam, and payment. A confirmation email containing the exam token, candidate ID and registration receipt is sent on completion.

Step 3 — Prepare

Use this handbook as the primary study reference. Section 9 provides a recommended preparation plan with module-by-module study time. Candidates should plan to complete all knowledge checks and, for Level 2, work through the scenario short answers in Appendix C before attempting the exam.

Step 4 — Schedule the Exam

Once prepared, you can take the exam at any time within your registration window. Exams are delivered fully online and are self-managed — there is no proctor and no testing center. Candidates choose when to begin, subject to the published time limit and any accommodation needs.

Step 5 — Take the Exam

When ready, candidates sign in and start the exam online. The exam is delivered immediately and must be completed in a single session within the published time limit.

Step 6 — Receive Results

Level 1 Non-Technical candidates receive their full result immediately on completion (the exam is multiple-choice only). Level 1 Technical candidates receive a provisional pass/fail indicator on the multiple-choice section immediately, with the final result — including the two short-answer scenarios scored on the modified Level 1 rubric — confirmed within 7 business days. Level 2 candidates receive a provisional multiple-choice indicator immediately and final results — including their scenarios (four for Technical, three for Non-Technical) scored on the standard rubric — within 10 business days.

Step 7 — Receive the Certificate

Successful candidates can download their certificate immediately upon passing (for Level 1 technical and Level 2, upon completion of scenario review). The certificate includes the verification URL described in Section 4.2.

Step 8 — Maintain the Credential

Credentials remain valid for two years from the issue date. Holders maintain their credential through the recertification process described in Section 8.

6.1 Indicative Timeline

Phase	Typical duration
Registration to scheduled exam date	2–12 weeks (candidate-driven)
Recommended preparation time — Level 1	20–40 hours over 2–6 weeks
Recommended preparation time — Level 2	40–80 hours over 4–12 weeks
Exam result delivery — Level 1	Within 5 business days
Exam result delivery — Level 2	Within 10 business days
Certificate issuance	Immediately upon passing (Level 2: upon completion of scenario review)
Recertification cycle	Every 24 months from issue date

7. Candidate Policies

7.1 Eligibility

Candidates must be at least 18 years of age or have written consent of a parent or guardian to register. Candidates must be able to read and respond in English at a professional level (additional languages may be supported in future releases).

7.2 Registration

Each candidate must register under their own legal name and verifiable identity. Registration is non-transferable. A candidate may hold an active registration for only one form of a given exam at a time.

7.3 Identity Verification

Each exam is tied to the candidate's registered name and email, and the resulting certificate is issued in that name. Because exams are self-managed and online, identity is established at registration rather than by an in-session proctor or testing center.

7.4 Retake Policy

Candidates who do not pass an exam may retake the exam immediately as there is no formal waiting period. However we recommend based on the results the following:

- First retake: 7 days after the date of the failed attempt
- Second retake: 14 days after the date of the second failed attempt
- Third and subsequent retakes: after 30 days, with completion of additional study or training recommended by the program owner

Each retake requires a new registration and the applicable fee. There is no cap on lifetime attempts, but candidates with three or more consecutive failures are encouraged to engage with an authorized training provider before the next attempt.

7.5 Cancellation, Reschedule, and Refund

All sales are final. Once a candidate commits to an exam there is no cancellation or rescheduling. Exam tokens will be good for up to a year from issuance.

Refunds may be issued for documented hardship or program-side cancellations. Refund requests are reviewed by the program owner on a case-by-case basis.

7.6 Accommodations

The program supports reasonable accommodations for candidates with documented needs, consistent with applicable law and the testing-provider policy. Common accommodations include extended time,

screen-reader compatibility, and rest breaks. Accommodation requests must be submitted at the time of registration and supported by appropriate documentation.

7.7 Confidentiality of Exam Content

All operational exam questions, scenario prompts, and answer keys are confidential intellectual property of the program. The sample items published in this handbook are explicitly identified as samples and do not appear on operational examinations; they are provided to illustrate exam format, difficulty, and expected reasoning, not to substitute for study of the underlying framework.

Candidates agree not to:

- Reproduce, distribute, photograph, or transmit any operational exam content
- Discuss specific operational exam items with other candidates, on social media, or in any public forum
- Use unauthorized assistance — including AI assistants, second devices, additional people, or notes — during the exam

Violations may result in score invalidation, credential revocation, and a multi-year ban from the program.

7.8 Score Reporting

Score reports show overall results (pass or fail), section-level performance bands, and (for Level 2) per-scenario scoring summaries. Item-level results are not provided. Score reports are issued only to the candidate and may be shared by the candidate with employers or third parties at their discretion.

7.9 Appeals

A candidate who believes their result was affected by a procedural error, scoring inconsistency, or technical issue may file a formal appeal within 30 calendar days of result delivery. Appeals are reviewed by an appeals panel that did not participate in the original scoring. Outcomes include result confirmation, free retest, score correction, or other remediation as appropriate.

Disagreement with a scenario score, in the absence of a procedural or scoring error, is not by itself grounds for an appeal. Candidates may, however, request an independent re-score of a Level 2 scenario for a published fee; the re-scored result, higher or lower, is final.

7.10 Code of Conduct

All candidates and credential holders agree to the following code of conduct. The standard applies during exam sessions and continues for the duration the credential is held.

- **Integrity.** Represent your work, results, and credentials accurately. Do not misuse a credential, and do not assist others in misusing one.

- **Respect for the framework.** Apply WEKID™ as defined by the current handbook. Do not represent personal modifications as official WEKID™ framework definitions.
- **Confidentiality.** Protect exam content and any proprietary or sensitive information encountered through the program.
- Professional conduct. Treat scorers, instructors, and program staff with courtesy. Disruptive, abusive, or fraudulent conduct may result in suspension or revocation.
- **Disclosure of conflicts.** When applying WEKID™ in a professional engagement, disclose material conflicts of interest that could bias evaluation outcomes.
- **Continuous learning.** Maintain awareness of WEKID™ updates and broader AI governance practice through the recertification cycle.

7.11 Disciplinary Actions

The program owner may suspend or revoke a credential for violations of the code of conduct, exam confidentiality, or certificate use policy. Disciplinary actions are reviewed under the appeals process described in Section 7.9. Revoked credentials are removed from the program registry, and the verification URL is updated to reflect the revocation.

8. Recertification

WEKID™ credentials have distinct validation periods from the issue date. For those certification that expire, recertification ensures that holders remain aligned with the current version of the WEKID™ framework and continue to develop their applied governance and implementation skills.

8.1 Why Recertification

AI systems, regulations, and operational patterns evolve rapidly. Recertification serves three purposes:

- Verify that the holder is current with the latest version of the WEKID™ framework as published in this handbook
- Confirm continued applied competence — not just that the holder once passed an exam
- Maintain employer and audit confidence in the meaning of an active credential

8.2 Recertification Pathways

Holders may recertify through any of the following pathways during the 12 months before credential expiration:

- **Continuing education credits (CEUs).** Earn 20 CEUs over the credential cycle through approved learning activities (training courses, conference sessions, applied governance projects, instructor delivery, or published articles). Submit CEU evidence with the recertification application.
- **Refresher assessment.** Complete a shorter, refresher-style assessment that focuses on changes between the version under which the credential was issued and the current version of this handbook.
- **Re-examination.** Retake the current version of the original exam. Useful when the holder wants a fresh full-credential reissuance, or when the framework has changed significantly between cycles.
- **Upgrade.** Earning a higher-level credential on the same path (e.g., A2T while holding P1T) automatically renews the lower-level credential through the higher credential's expiration date.

8.3 Lapsed Credentials

A credential that is not renewed by the expiration date enters a 6-month grace period during which the holder may still recertify under the pathways above. After the grace period, the credential is considered lapsed; the certificate verification URL reflects this status, and reinstatement requires re-examination at the current version.

8.4 Approved CEU Activities

Activity	CEU value
Completing an approved WEKID™ training course	Course-defined; typically, 5–15 CEUs
Attending an approved AI governance conference session	1 CEU per hour, capped at 6 CEUs per event
Delivering instruction on WEKID™ to an organization or cohort	2 CEUs per hour delivered, capped at 10 CEUs per cycle
Authoring a published article, paper, or case study applying WEKID™	5–10 CEUs per published item, depending on length and rigor
Documented application of WEKID™ in an enterprise governance project	Up to 10 CEUs per project, with supporting evidence
Contributing peer review or scoring to the WEKID™ program	Per-engagement CEU value as defined by the program owner

9. How to Prepare for the Exam

This handbook is a canonical study reference. Candidates who study Part II thoroughly, complete the knowledge checks, and — for Level 2 — work through the scenarios in Appendix C are well prepared for their selected exam. The recommendations below help candidates plan study time and self-assess readiness.

9.1 Recommended Study Plan

The plan below is a starting point. Candidates with prior AI governance or implementation experience may move more quickly; candidates new to the field may need additional time, especially on Modules 1–3 (foundations) and Module 7 (scoring and gating).

Module	Topic focus	Level 1 study time	Level 2 study time
Module 1	Evolution of modern AI systems and limits of existing evaluation	1–2 hours	2–3 hours
Module 2	The WEKID™ epistemic stack (W/E/K/I/D)	3–4 hours	3–4 hours
Module 3	Hierarchy, non-substitutability, and risk propagation	2–3 hours	3–4 hours
Module 4	WEKID™ for tool-using and agentic AI	2–3 hours	4–6 hours
Module 5	WEKID™ as enterprise governance mechanism	2–3 hours	4–6 hours
Module 6	From framework to measurable signals	3–4 hours	4–6 hours
Module 7	Scoring and gating (decisioning)	3–4 hours	5–7 hours
Module 8	Computing WEKID™ scores (operational method)	2–3 hours	4–6 hours
Module 9	Implications for trustworthy AI	1–2 hours	3–4 hours
Module 10	Implementation guidance	1–2 hours	4–6 hours
Appendix B	Multiple-choice self-test review	Built into module review	Built into module review
Appendix C	Scenario short-answer practice	Optional	8–12 hours

9.2 Suggested Study Sequence

1. Read Section 1–6 of this handbook to internalize the program structure.

2. Read Modules 1–3 in sequence; these establish the foundation everything else builds on.
3. Read Modules 4–6 to bridge from framework concepts to measurable practice.
4. Read Modules 7–8 carefully; scoring and gating logic appear heavily on both Level 1 and Level 2 exams.
5. Read Modules 9–10 to round out trust framing and implementation patterns.
6. Complete every sample knowledge check without referring back to the unit text. Score yourself using Appendix B. The samples illustrate exam format and difficulty but are a small subset of what you should master — confidence in the samples means you understand the concept, not that you have seen the operational items.
7. For Level 2: write a complete answer to each sample scenario in your selected path before reviewing Appendix C. Compare your answers against the required elements and sample responses; expect operational scenarios to test the same kinds of reasoning across different facts.
8. Re-read any module where your knowledge check accuracy was below 70% or where your scenario answers missed required elements.

9.3 Self-Assessment Checklist

Before scheduling the exam, candidates should be able to:

- State the five WEKID™ layers in order from highest to lowest and define each in their own words
- Explain non-substitutability with a concrete example
- Identify the failing layer for any of the failure modes in Appendix A
- Apply the 0–5 scoring scale and the weighted aggregation formula
- Recognize when a hard gate should trigger and what governance action follows
- Distinguish soft constraints (human-in-the-loop) from hard rejection
- (Level 2 Technical) Sketch a runtime gate as policy-as-code with the inputs, conditions, and emitted action code
- (Level 2 Non-Technical) Draft an authority map and escalation path for a high-stakes workflow

9.4 Approved Training Providers

Independent study using this handbook is fully sufficient to prepare for any WEKID™ exam. Candidates who prefer instructor-led learning may engage authorized training providers offering courses aligned to this handbook. The list of currently approved providers is published by the program owner and is updated as new providers are accredited.

10. Adopting WEKID™ in Your Organization

Cultural and change-management guidance for sponsors rolling out the credential across teams.

WEKID™ is a framework, not a tool, and the credential is the easy part. The harder work is shifting how engineers, product managers, risk officers, and executives talk about AI quality day-to-day. Teams that have done this well treat adoption as a change-management program, not a training rollout. The guidance below is drawn from common patterns observed in early-adopter organizations and is offered as a starting point for sponsors and program leads.

10.1 Sponsor and coalition

Adoption is far more durable when an executive sponsor (typically a CIO, CTO, CRO, or chief AI/data officer) names WEKID™ as the firm's working framework for AI evaluation and signs off on the rollout plan. Pair the sponsor with a small cross-functional coalition — one engineer, one risk/compliance partner, one product owner, one auditor — whose job is to keep the rollout honest about where the framework is helping and where it is creating friction. Both groups should hold the credential themselves before championing it; nothing erodes adoption faster than leaders who exempt themselves from the standard they are asking others to meet.

10.2 Phased rollout

Most successful rollouts unfold over three phases of three to four months each:

- **Phase 1 — Pilot.** Pick one high-visibility AI workflow and run WEKID™ evaluation against it for 60–90 days in parallel with whatever review process exists today. Credential the small team that owns it (typically 5–10 staff at Level 1, with the team lead at Level 2). Track hard-gate rate, time-to-decision, and reviewer agreement, share findings widely.
- **Phase 2 — Expand.** Roll the framework out to two to four additional workflows, replacing (not duplicating) the prior review process where the pilot proved the framework holds up. Begin requiring Level 1 certification for any individual contributor whose work product enters those workflows; require Level 2 for the reviewers who approve them. Stand up the coverage-target dashboard described in §4.4.
- **Phase 3 — Embed.** Make WEKID™ the default. Update job descriptions, hiring rubrics, performance reviews, and audit playbooks to reference it. Move from voluntary to required certification for in-scope roles. Begin reporting hard-gate rate, exception rate, and remediation time-to-close at the audit-committee or board level on a quarterly cadence.

10.3 Common cultural objections — and how to answer them

Resistance is normal and rarely about the framework itself. Most objections are about workload, autonomy, or fear of being graded. The common ones, with responses that have been carried out in practice:

What you will hear	What works
"This is going to slow us down."	Measure decision time before and after the pilot. Most teams find net time-to-decision drops because reviewers stop relitigating the same disagreements with new vocabulary every time. Publish the before/after numbers.
"We already have a review process."	Map the existing process to the WEKID™ layers. The exercise usually surfaces gaps (typically thin Wisdom and Experience coverage) and lets the team adopt the framework as an upgrade rather than a replacement.
"The score will be used against me."	Be explicit that scores evaluate AI outputs, not the people who produced or reviewed them. Publish a written policy stating WEKID™ scores will not be used as inputs to individual performance reviews. Reinforce in the first month with manager talking points.
"The hard gates are too strict."	Walk through one or two specific gate triggers with the complaining team and ask whether the underlying failure (a fabricated citation, a missing risk disclosure) is something the firm wants to ship. The conversation always resolves itself.
"Engineers and risk people don't speak the same language."	That is precisely what the credential is for. Run the first joint workshop using a real internal incident, scored against WEKID™ in the room. The shared vocabulary lands faster from one shared example than from any amount of training material.

10.4 Make the new behavior visible

Adoption sticks when the framework shows up in the artifacts people already use. A brief list that has worked in early-adopter organizations:

- Add a WEKID™ layer-mapping field to incident write-ups and post-mortems.
- Reference the four governance actions (approve, constrain, remediate, reject) in approval workflows in the issue tracker, change-management system, and model-risk inventory.
- List the credentialed reviewer's name on every approved AI-output evaluation artifact, the same way audit work papers list a sign-off partner.

- Add credential requirements to job postings for in-scope roles. Even “preferred” rather than “required” signals to candidates — and to peers — that the framework is the firm’s current standard.
- Once Phase 2 is underway, share quarterly metrics in an organization-wide forum: hard-gate rate, top failing layer by workflow, time-to-remediate. Visibility, not enforcement, is what shifts the culture.

10.5 Anti-patterns to avoid

- **Mass-credentialing the entire workforce in week one.** You burn the budget before anyone has used the framework on a real workflow. Pilot first, then scale.
- **Treating WEKID™ as a compliance checkbox.** If the only people who care about the score are auditors, the framework will be optimized to pass audits rather than to catch bad outputs. Keep the framework owned by an operating function, not the second line.
- **Quietly relaxing hard gates after the first incident.** Hard gates exist precisely because lower-layer failures cannot be talked around. If a gate is creating real friction, change the underlying workflow or escalation path; do not lower the gate.
- **Forgetting to recertify.** The framework evolves; credentials expire after 24 months for a reason. Use the Employer Portal alerts (see §4.4) to schedule recertification well before it lapses.

WEKID™ at a Glance

This one-page reference summarizes the framework. It is meant to anchor study and to serve as a desk reference once certified. Each layer is described in full detail in Module 2.

The Five Layers (Highest to Lowest)

Layer	Asks	Concerns	Typical failure
Wisdom (W)	Whether and should	Judgment, risk awareness, value alignment, restraint	Overconfident or unsafe recommendations in high-stakes contexts
Experience (E)	How (in practice)	Procedural realism, sequencing, recovery, edge cases	Impractical or brittle plans; missing rollback or verification
Knowledge (K)	Why and how (conceptually)	Synthesis, causal reasoning, internal consistency, bounded generalization	Confident but incorrect explanations from correct facts
Information (I)	What matters here	Relevance, completeness, framing, clarity	Missing context or constraints; misleading framing
Data (D)	What is	Atomic factual correctness, source fidelity, tool-output integrity	Hallucinated facts, fabricated sources, mis-transcribed tool outputs

The Hierarchy Rule

Each layer depends on the integrity of the layers beneath it. Higher-layer quality cannot repair lower-layer failures. This is non-substitutability — the foundational rule that makes WEKID™ different from flat, average-based scoring.

Scoring and Gating in One Glance

Each layer is scored 0–5. The default weighted aggregation, used after gates pass, is:

Weighted aggregation formula

$$\text{WEKID™ Score} = 0.25 \times D + 0.20 \times I + 0.20 \times K + 0.15 \times E + 0.20 \times W$$

Weights are tunable by domain, regulatory context, and risk tolerance, but the principle is fixed: foundational integrity (D) and judgment (W) carry the most weight in high-stakes contexts.

Baseline Hard Gates

Gate	Trigger	Effect
Data Integrity Gate	Data score < 2	Cap total score at 50/100
Fabrication Gate	Any fabricated citation, source, or tool result detected	Cap total score at 40/100
High-Stakes Wisdom Gate	High-stakes content (medical, legal, safety-critical) without uncertainty acknowledgment or risk language	Cap total score at 60/100

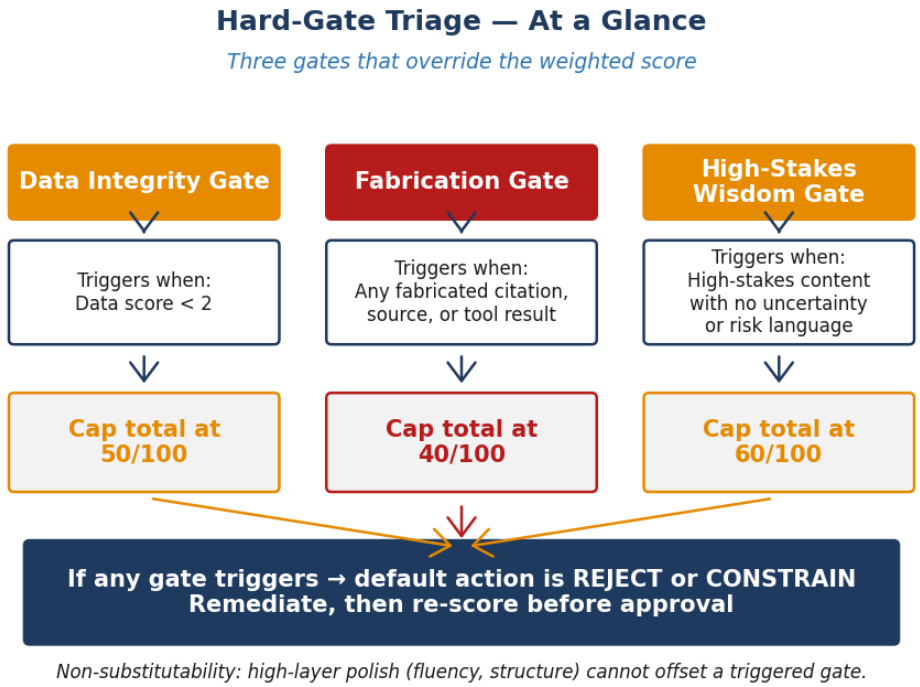

Figure 2. Hard-Gate Triage — the three baseline gates that override the weighted score and the cap each applies.

The Four Governance Actions

Action	When
Approve	Total score ≥ 80 and no hard gates triggered
Constrain (human-in-the-loop)	Any layer below threshold for the risk tier, or Wisdom or Experience below threshold

Action	When
Remediate	Total score 50–69 or a soft gate triggered — block use until targeted fixes are made and rescoring passes
Reject	Hard gate triggered (e.g., Data < 2 or fabrication detected)

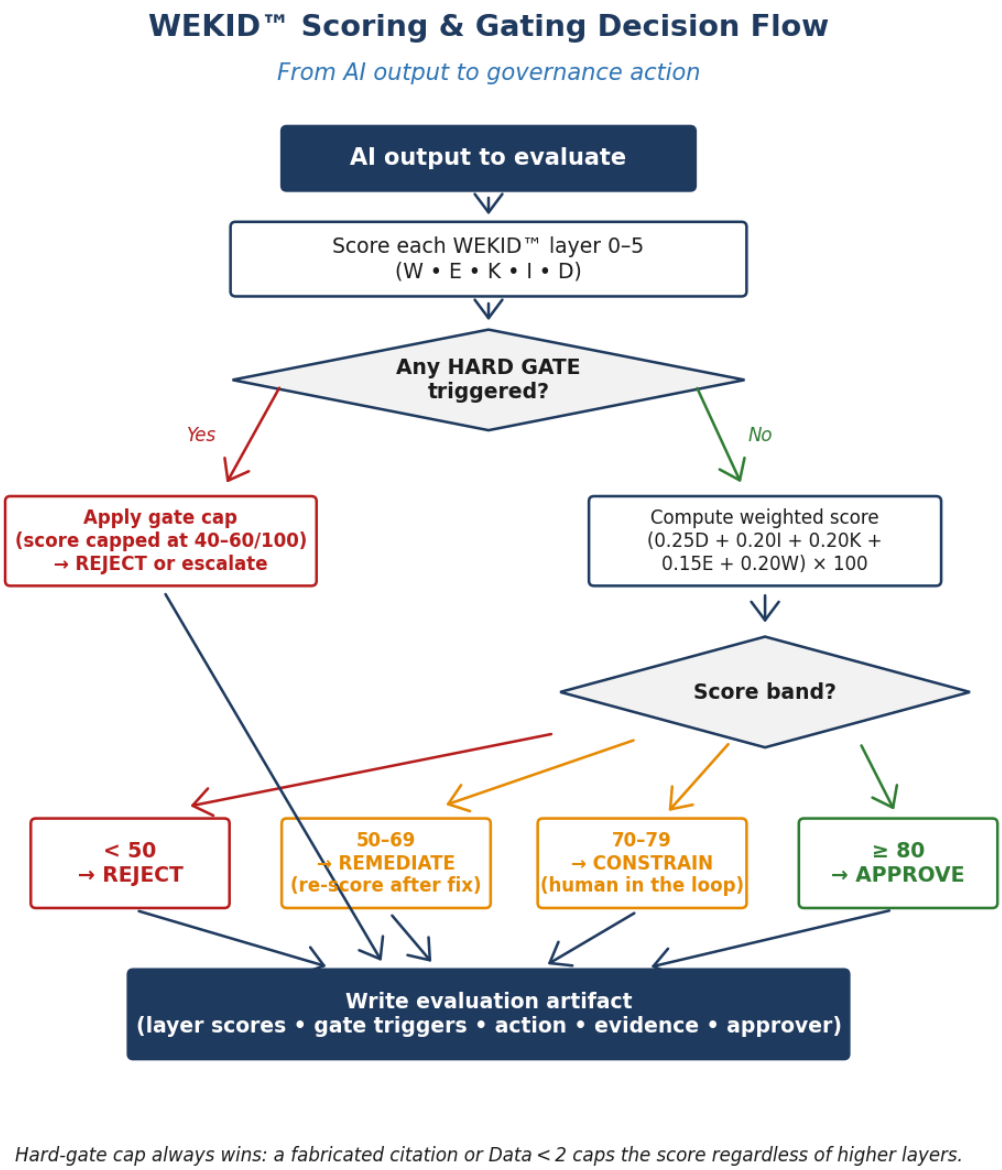

Figure 1. WEKID™ Scoring & Gating Decision Flow — the path from a single AI output to one of four governance actions.

PART II

The WEKID™ Framework Reference

Required knowledge — ten modules with knowledge checks and scenarios

Part II Introduction — Required Knowledge

The modules that follow provide the WEKID™ framework background required to succeed on the certification exams. Candidates should use this material as the reference syllabus for definitions, scoring and gating logic, and worked examples that appear in scenario-based assessment questions.

Each module is structured the same way: a conceptual introduction, a set of numbered units that develop the topic, a small number of sample knowledge-check items, and (where applicable) a sample Level 2 scenario short-answer prompt. Sample items illustrate exam format, difficulty, and expected reasoning; operational examination items are confidential and do not appear in this handbook. Sample answers and scoring rubrics for the published scenarios appear in Appendix C.

A note on the focus callouts. Each module begins with a short "Focus by credential" callout. These callouts are derived from the exam blueprints in §5.2 and §5.3 and the prep plan in §9.1. They tell you, in two or three lines, how heavily a given module will be tested on your selected exam form. Treat the callouts as guidance, not as permission to skip content: every module contains material that appears at least once on every exam, and unfamiliarity with a "low priority" module is a common cause of borderline scores.

Why WEKID™ Now

As artificial intelligence systems become more autonomous, tool-enabled, and deeply embedded in enterprise operations, traditional methods for evaluating AI — such as accuracy scores, fluency, or benchmark performance — are no longer sufficient. These approaches fail to distinguish between outputs that are merely well-written and those that are factually sound, operationally safe, and grounded in appropriate judgment. As a result, organizations face increasing risk from AI systems that appear dependable but produce misleading, unsafe, or poorly reasoned outcomes.

At the core of this challenge is epistemic quality — the quality of how knowledge is formed, justified, interpreted, and applied. An epistemically sound AI output is not only correct, but based on valid facts, properly contextualized, correctly understood, applicable, and guided by appropriate judgment given uncertainty and risk.

The WEKID™ framework was first introduced in 2007 as a hierarchical model of intelligence and later became the intellectual foundation for a competitive intelligence service — well before the widespread adoption of AI systems. As AI technology has rapidly evolved, WEKID™ has emerged as a structured, epistemic framework for evaluating the quality of AI-generated intelligence. It assesses outputs across five layers — Data, Information, Knowledge, Experience, and Wisdom — reflecting the way humans progress from raw facts to understanding, application, and sound judgment. This layered approach

enables organizations to determine not only whether an AI output is accurate, but also whether it is well-reasoned, operationally viable, and appropriate for the context in which it is used.

Unlike model-centric evaluation approaches, WEKID™ focuses on AI outputs, runtime execution, and decisions — making it applicable across models, vendors, and system architectures. It supports both human review and automated scoring, including AI systems that rely on external tools, retrieval mechanisms, and autonomous agents. By enforcing hierarchical dependencies between layers, WEKID™ prevents fluent or confident outputs from masking foundational errors or unsafe judgment.

WEKID™ transforms high-level AI ethics and responsible-AI principles into measurable, enforceable governance controls. It enables objective scoring, policy-based thresholds, audit-ready artifacts, and actionable diagnostics that guide remediation and continuous improvement. As such, WEKID™ provides a practical foundation for enterprise AI governance, compliance, and risk management, supporting the safe, transparent, and trustworthy deployment of AI systems in high-impact environments.

Module 1: Evolution of Modern AI Systems

Focus by credential

- **WEKID-P1T (Practitioner Tech)** — High priority. Be able to recognize each failure mode and name the WEKID™ layer it implicates. Expect 3-4 multiple-choice items from this module.
- **WEKID-P1N (Practitioner Non-Tech)** — High priority. Understand why traditional benchmarks miss governance-relevant failures; you do not need implementation depth.
- **WEKID-A2T (Authority Tech)** — Moderate priority. The framework gap drives every gating decision you make at runtime; lean on Units 1.2 and 1.3.
- **WEKID-A2N (Authority Non-Tech)** — Moderate priority. Use this module to sharpen the elevator pitch you will deliver to executives.

Large language models and AI agents are no longer experimental tools; they increasingly generate outputs that directly influence business decisions, policy interpretation, software development, healthcare guidance, legal reasoning, and financial analysis. These systems are now embedded in workflows where their outputs are consumed by humans, integrated into automated pipelines, or acted upon by other machines with little or no human intervention.

A defining characteristic of modern AI outputs is their fluency and confidence. Responses are often coherent, well-structured, and authoritative in tone, which can create a strong perception of reliability. However, this surface quality frequently masks deeper epistemic weaknesses. AI-generated outputs may be factually incorrect, logically inconsistent, operationally impractical, or unsafe — yet still appear persuasive and complete. In a high-impact context, this mismatch between appearance and epistemic integrity introduces significant risk.

Real-world incidents have already demonstrated these failure modes. AI systems have confidently cited nonexistent legal cases in court filings, produced plausible but incorrect medical guidance, generated financial analyses based on fabricated or outdated data, and recommended unsafe operational procedures despite technically correct intermediate steps. In each case, the failure was not merely factual error, but a breakdown in how information was interpreted, applied, or judged under real-world constraints.

Unit 1.1 — Limitations of Existing Evaluation Paradigms

Current approaches to AI evaluation are ill-suited to this reality. Most existing paradigms suffer from three fundamental limitations:

- **Overemphasis on correctness or fluency without assessing judgment.** Evaluation methods focus on factual accuracy, benchmark performance, or linguistic quality. These metrics fail to detect cases where an output is technically correct but inappropriately applied, overconfident,

or unsafe. For example, systems have produced correct summaries of regulations while drawing incorrect compliance conclusions, leading users toward risky decisions.

- **Lack of explainability and diagnostic clarity.** When an AI output is deemed "good" or "bad," existing metrics often provide little insight into why. As a result, organizations struggle to diagnose whether failures stem from bad data, misinterpretation, poor procedural reasoning, or flawed judgment. This opacity complicates remediation and undermines auditability, particularly in regulated environments.
- **Inability to govern tool-using or autonomous AI systems.** As AI systems increasingly invoke external tools, retrieve live data, and execute multi-step plans, evaluation approaches focused solely on model outputs become insufficient. Documented failures include systems that correctly retrieved data but misinterpreted statistical significance, invoked inappropriate tools for high-stakes tasks, or continued executing plans despite accumulating errors — demonstrating a lack of procedural and judgmental safeguards.

Together, these limitations create a governance gap in which AI systems may appear to perform well under traditional metrics while posing unacceptable operational, legal, or safety risks.

Unit 1.2 — From Model Performance to Epistemic Quality

As AI systems evolve from passive assistants to decision-support and decision-making entities, evaluation must evolve accordingly. The central question is no longer simply "Is this answer correct?" but rather:

- How was the answer constructed?
- What assumptions and sources does it rely on?
- How does it manage uncertainty, edge cases, or incomplete information?
- Is the recommendation appropriate given the stakes and potential consequences?

Real-world failures increasingly show that correct facts alone are insufficient. Systems have generated answers that were factually accurate yet operationally infeasible, legally risky, or ethically misaligned. In such cases, the failure lies not at the level of data, but at higher epistemic layers involving understanding, experience, and judgment.

Addressing these risks requires a shift from surface-level evaluation to epistemic evaluation — an assessment of how data is transformed into information, information into knowledge, knowledge into action, and action into judgment.

Unit 1.3 — Introducing WEKID™

WEKID™ was developed as a response to this need. It provides a structured, hierarchical framework for evaluating AI outputs across five epistemic layers: Wisdom, Experience, Knowledge, Information, and

Data. Rather than treating AI quality as a single scalar value or a flat checklist of metrics, WEKID™ models output quality in a way that mirrors human reasoning and decision-making.

By making epistemic quality explicit, hierarchical, and measurable, WEKID™ enables organizations to identify not just that an AI output failed, but where and why it failed. This capability is essential for governing AI systems that operate in complex, high-impact environments. WEKID™ provides the foundation for moving from trust based on fluency or reputation toward trust grounded in epistemic integrity.

Module 1 Sample Knowledge Check

The item below is a sample knowledge-check question illustrating exam format and difficulty. The answer key appears in Appendix B.
Which statement best describes why WEKID™ is needed in modern AI governance?
A. Traditional metrics already measure epistemic quality and only need better benchmarks.
B. Fluency and benchmark performance can mask factual, procedural, and judgment failures in real operating contexts.
C. WEKID™ replaces all human judgment with automated scoring.
D. WEKID™ is only applicable to model training and evaluation, not runtime decisions.

Module 1 — Sample Level 2 Scenario

The scenario below is a sample illustrating Level 2 short-answer format and expected reasoning. Operational scenarios are confidential and do not appear in this handbook. Required elements and a sample response appear in Appendix C.

SA-M1-S1 *— Level 2 • Non-Technical*

Scenario: An executive receives an AI-generated compliance summary that is clear and confident but contains one fabricated regulatory citation. The executive wants to proceed immediately.

Prompt: Using WEKID™, write the decision and escalation rationale (approve/constrain/remediate/reject), including what evidence you would require before approval and what audit artifact you would retain.

Module 2: The WEKID™ Epistemic Stack (W/E/K/I/D)

Focus by credential

- **WEKID-P1T (Practitioner Tech)** — Critical. The five layers anchor 25% of your exam (failure mode recognition). Memorize each layer's definition and a canonical failure example.
- **WEKID-P1N (Practitioner Non-Tech)** — Critical. 25% of your exam. Plan to articulate each layer in plain language an executive could repeat.
- **WEKID-A2T (Authority Tech)** — Critical foundation. You are expected to map W/E/K/I/D to specific runtime controls (Unit 2.6 onward).
- **WEKID-A2N (Authority Non-Tech)** — Critical foundation. Focus on Units 2.6 through 2.9 — these connect the stack to risk frameworks and gating policies you will be asked to design in scenarios.

WEKID™ is a hierarchical epistemic framework composed of five distinct layers, each representing a different function in the formation, interpretation, and application of knowledge. These layers reflect how reasoning operates in both human cognition and complex decision-making systems, progressing from raw facts to judgment under uncertainty.

A defining characteristic of WEKID™ is hierarchical dependency. Each layer depends on the integrity of the layers beneath it; failures at lower layers cannot be compensated for by strength at higher layers. This structure ensures that evaluation prioritizes epistemic soundness over surface quality, fluency, or confidence.

The five layers — Wisdom, Experience, Knowledge, Information, and Data — together form a complete model for assessing the epistemic quality of AI outputs.

Unit 2.1 — Wisdom

Definition. Wisdom is the capacity to exercise sound judgment — discerning what is appropriate, safe, and durable considering uncertainty, risk, values, and long-term consequences. Wisdom answers whether and should, not merely can. This layer governs restraint, prioritization, and ethical alignment.

Failure modes. Wisdom failures often arise from:

- Overconfidence or unwarranted certainty
- Unsafe or irresponsible recommendations, particularly in high-stakes contexts
- Ignoring trade-offs, externalities, or ethical implications

Such failures can transform competent analysis into harmful action.

Evaluation focus. Evaluation of the Wisdom layer focuses on:

- Uncertainty calibration: acknowledging limits and confidence bounds

- Risk awareness: proportional caution relative to stakes
- Tradeoff reasoning: explicit consideration of alternatives and consequences
- Alignment: consistency with user intent, organizational policy, and societal constraints

Wisdom represents the highest epistemic function and the final safeguard against misuse.

Unit 2.2 — Experience

Definition. Experience represents procedural and operational competence — the ability to apply knowledge in practice. It reflects "knowing how," including sequencing actions, anticipating failure modes, managing edge cases, and adapting when conditions change. For AI systems, Experience is not lived history but encoded procedural patterns and operational heuristics.

Failure modes. Experience failures commonly include:

- Impractical or unrealistic recommendations
- Missing steps, prerequisites, or dependencies
- Ignoring common pitfalls, constraints, or failure conditions

These failures can cause otherwise correct reasoning to fail in real-world execution.

Evaluation focus. Evaluation at the Experience layer emphasizes:

- Procedural realism: feasibility of recommended steps
- Tool orchestration: appropriate selection and sequencing of tools or actions
- Verification and checkpoints: mechanisms to confirm progress or correctness
- Operational heuristics: application of best practices and safeguards

This layer is critical for decision-support and autonomous systems.

Unit 2.3 — Knowledge

Definition. Knowledge reflects genuine understanding: the ability to synthesize information into coherent explanations, mental models, rules, or causal relationships. Knowledge answers why and how, not merely what. At this level, the system demonstrates that it can correctly integrate facts and information into meaningful reasoning.

Failure modes. Knowledge failures often appear sophisticated but are epistemically flawed, including:

- Logical contradictions within explanations or across claims
- Incorrect explanations despite correct underlying data
- Overgeneralization, where conclusions extend beyond what evidence supports

These failures are especially problematic because they can sound authoritative while being fundamentally wrong.

Evaluation focus. Evaluation at the Knowledge layer focuses on:

- Conceptual correctness: accurate explanation of mechanisms or principles
- Synthesis: coherent integration of multiple pieces of information
- Internal consistency: absence of contradiction or incoherence
- Bounded generalization: appropriate scope and restraint

This layer distinguishes memorization from understanding.

Unit 2.4 — Information

Definition. Information is data that has been organized, contextualized, and communicated in a way that makes it meaningful for a specific task, question, or audience. While Data answers what is, Information answers what matters here. This layer addresses selection, emphasis, framing, and structure — transforming raw facts into usable content.

Failure modes. Even when data is correct, Information-level failures can render outputs misleading or ineffective, including:

- Irrelevant detail that obscures the core issue
- Omitted critical context, assumptions, or constraints
- Confusing, ambiguous, or biased framing that distorts interpretation

Such failures often lead users to misunderstand the implications of otherwise correct facts.

Evaluation focus. Evaluation at the Information layer emphasizes:

- Relevance: alignment with the user's question or task
- Completeness: coverage of key aspects without material omission
- Clarity: structure, readability, and unambiguous presentation
- Usability: whether the information can be readily applied

Unit 2.5 — Data

Definition. Data consists of atomic factual elements used to support reasoning or decision-making. These elements include dates, names, numerical values, measurements, classifications, direct quotations, and outputs produced by external tools such as databases, calculators, APIs, or sensors. Data represents the foundational truth-bearing layer upon which all higher epistemic functions depend.

At this level, no interpretation or synthesis is assumed; data elements are evaluated strictly for correctness and fidelity.

Failure modes. Failures at the Data layer undermine all subsequent reasoning and commonly include:

- Hallucinated facts, such as invented names, events, statistics, or references
- Incorrect calculations, including arithmetic errors, unit mismatches, or temporal inaccuracies
- Fabricated or distorted sources, including invented citations or misrepresented tool outputs

These failures are particularly dangerous because they can be difficult for users to detect and may be confidently presented.

Evaluation focus. Evaluation at the Data layer focuses on:

- Accuracy: factual correctness relative to known truth, authoritative sources, or tool outputs
- Precision: correct units, values, and levels of specificity
- Constraint adherence: respecting scope, instructions, and input boundaries
- Absence of fabrication: no invented facts, sources, or tool results

Because Data integrity is a prerequisite for all higher layers, significant failure at this level typically triggers hard gating or rejection.

Unit 2.6 — Mapping WEKID™ to Risk, Controls, and Measurement

One of WEKID™'s distinguishing strengths is that each epistemic layer can be directly mapped to specific enterprise risks, governance controls, and measurable metrics. This mapping enables WEKID™ to function not only as an evaluation framework, but also as a risk management and compliance instrument.

By explicitly linking epistemic failure modes to operational risk, WEKID™ allows organizations to:

- Identify where AI risk originates
- Apply targeted controls rather than generic safeguards
- Measure effectiveness using repeatable, auditable signals

This approach aligns AI governance with established enterprise risk management practices, including control testing, continuous monitoring, and audit review.

Unit 2.7 — Example: WEKID™ Risk–Control–Metric Mapping

Table 1 illustrates how each WEKID™ layer maps to common AI risks, governance controls, and measurable evaluation metrics. The examples shown are indicative and can be tailored to specific domains, regulatory environments, or risk tolerances.

Table 1. WEKID™ Layer Mapping to Risks, Controls, and Metrics

WEKID™ Layer	Primary Risk Types	Governance Controls	Example Metrics / Signals
Wisdom	Overconfident recommendations, unsafe decisions, ethical misalignment	Risk thresholds, uncertainty disclosure requirements, policy alignment checks, autonomy limits	Uncertainty calibration score, risk flag rate, policy compliance score, tradeoff acknowledgment presence
Experience	Impractical guidance, unsafe procedures, brittle workflows	Procedural templates, step validation, fallback logic, human-in-the-loop escalation	Step completeness score, edge-case coverage, recovery path presence, verification hook count
Knowledge	Incorrect explanations, faulty reasoning, overgeneralization	Logical consistency checks, explanation validation, contradiction detection, bounded inference rules	Contradiction rate, explanation correctness score, generalization error rate
Information	Misleading framing, omission of critical context, irrelevance to task	Relevance filters, coverage checklists, prompt constraints, explainability requirements	Relevance score, completeness checklist pass rate, readability index, task coverage score
Data	Hallucinated facts, incorrect calculations, fabricated sources, corrupted tool output	Source verification, retrieval constraints, tool output validation, citation enforcement, hard rejection on fabrication	% factual accuracy, claim verification rate, numeric error rate, fabricated citation count

Unit 2.8 — Using Mapping in Practice

This mapping enables concrete governance use cases:

- **Risk Assessment:** Each WEKID™ layer can be assessed independently to identify dominant risk drivers within a system or use case.
- **Control Design:** Controls can be selectively applied based on observed failure patterns rather than uniformly across all outputs.
- **Metric-Driven Oversight:** Quantitative signals allow teams to set thresholds, track trends, and detect drift over time.
- **Audit and Compliance:** Layer-specific metrics provide defensible evidence that governance controls are operating as intended.

Unit 2.9 — Supporting Gating and Escalation

The table also supports WEKID™'s hierarchical gating logic:

- Failures in Data may trigger automatic rejection
- Weaknesses in Knowledge may cap allowable autonomy
- Low Wisdom scores may require human review or override

By anchoring gating decisions to explicit risks and controls, WEKID™ enables explainable enforcement, rather than opaque policy outcomes.

Unit 2.10 — Extensibility Across Domains

While Table 1 presents a generalized view, the same structure can be extended to domain-specific profiles such as healthcare, finance, defense, or legal analysis. In each case, the WEKID™ layers remain constant while risks, controls, and metrics are tailored to the domain's regulatory and operational context.

Together, these five layers form a complete epistemic stack that allows AI outputs to be evaluated not just for correctness, but for understanding, practicality, and judgment. By enforcing hierarchical dependency across Data, Information, Knowledge, Experience, and Wisdom, WEKID™ provides a rigorous foundation for trustworthy, governable AI.

Module 2 Sample Knowledge Check

The item below is a sample knowledge-check question illustrating exam format and difficulty. The answer key appears in Appendix B.
Which WEKID™ layer is most directly concerned with "atomic factual correctness" (dates, quantities, entities, tool outputs)?
A. Wisdom
B. Experience
C. Knowledge
D. Data

Module 2 — Sample Level 2 Scenario

The scenario below is a sample illustrating Level 2 short-answer format and expected reasoning. Operational scenarios are confidential and do not appear in this handbook. Required elements and a sample response appear in Appendix C.

SA-M2-S1 *— Level 2 • Non-Technical*

Scenario: A compliance team uses an AI assistant to summarize new regulations. The summary is clear and persuasive, but it omits scope limits that materially change the obligations.

Prompt: Using WEKID™, explain the layer mapping (especially Information completeness), and propose governance controls (review checklist, escalation triggers) to prevent misapplication.

Module 3: Why Hierarchy Matters — Risk Propagation and Non-Substitutability

Focus by credential

- **WEKID-P1T (Practitioner Tech)** — High priority. Non-substitutability is a frequent target of multiple-choice items; expect at least one scenario element testing it.
- **WEKID-P1N (Practitioner Non-Tech)** — High priority. Same as P1T, with emphasis on Unit 3.7 (trust and oversight implications).
- **WEKID-A2T (Authority Tech)** — Critical. Authority-level gating logic depends entirely on hierarchy — Unit 3.5 is your foundation for every policy-as-code question.
- **WEKID-A2N (Authority Non-Tech)** — Critical. Authority candidates are expected to defend hierarchy when stakeholders push for "averaged" or "compensating" scoring. Units 3.4 and 3.5 are the core.

The hierarchical structure of WEKID™ is not merely an organizing convenience — it reflects deep principles in epistemology and risk management. Understanding why hierarchy matters is essential to applying the framework correctly in governance, evaluation, and decision-making.

Unit 3.1 — Epistemological Foundations of Hierarchy

Classical epistemology has long recognized that knowledge is layered. Facts must precede information; information must precede understanding; understanding must precede competent action; and competent action must be tempered by judgment. WEKID™ captures this lineage and makes it operational. In this tradition:

- False premises invalidate valid reasoning
- Correct explanations without understanding mislead
- Competent action without judgment produces harm

WEKID™ formalizes these insights into an operational framework suitable for AI evaluation.

Unit 3.2 — Non-Substitutability of Epistemic Layers

AI evaluation frameworks implicitly assume that weaknesses in one dimension can be offset by strengths in another. WEKID™ explicitly rejects this assumption through epistemic non-substitutability:

- A wise-sounding recommendation built on false data is still wrong.
- A fluent explanation without genuine understanding is misleading.
- A technically correct answer applied without judgment may be dangerous.

From an epistemological standpoint, these are category errors: properties of higher-order cognition cannot repair failures of foundational truth conditions.

Unit 3.3 — Failure Propagation and Epistemic Risk

Lower-layer epistemic failures tend to propagate upward, amplifying risk rather than canceling out. In AI systems, this manifests as:

- Incorrect data being coherently summarized and confidently acted upon
- Misinterpreted information leading to polished but invalid conclusions
- Correct reasoning deployed in inappropriate or unsafe contexts

This phenomenon aligns with what philosophers of science describe as error amplification through inference chains, where downstream reasoning inherits and compounds upstream falsehoods. The more fluent and confident the system appears at higher layers, the more dangerous these failures become, because user trust increases precisely when epistemic integrity is weaker.

Unit 3.4 — Why Flat Scoring Models Fail

Flat scoring systems that rely on accuracy, fluency, usefulness, and safety violate basic epistemic logic. They allow surface qualities to dominate foundational correctness, producing the well-known failure mode of confident, articulate, and wrong outputs. WEKID™'s hierarchy prevents this by enforcing epistemic precedence:

- Data integrity constrains all higher evaluation
- Knowledge cannot score highly if Information is flawed
- Wisdom cannot elevate outputs that are factually or procedurally unsound

This aligns evaluation with truth-preserving reasoning, not aesthetic quality.

Unit 3.5 — Gating as Formal Epistemic Control

The hierarchical dependency of WEKID™ enables gating mechanisms that reflect epistemic necessity rather than preference. In formal terms, lower layers function as necessary conditions, while higher layers represent sufficiency only when prerequisites are met. Operationally, this means:

- Failures at lower levels cap the maximum achievable score
- Certain violations (e.g., fabricated sources, unsafe recommendations) trigger hard rejection
- Higher-layer excellence is recognized only when epistemic foundations are intact

Gating ensures that systems cannot "reason their way out" of factual or ethical failure.

Unit 3.6 — Visual Hierarchy Model

The WEKID™ epistemic hierarchy can be visualized as a stack with Data at the foundation and Wisdom at the apex. Each layer rests on the integrity of those beneath it:

WEKID™ Hierarchy (top to bottom)

Wisdom — sound judgment under uncertainty

Experience — procedural and operational competence

Knowledge — understanding and synthesis

Information — contextualized, relevant content

Data — atomic factual correctness (foundation)

Key properties illustrated:

- Each layer depends on the integrity of the layer below
- Failures propagate upward
- Higher layers cannot compensate for lower-layer defects

This structure makes WEKID™ especially suitable for automated gating, auditability, and governance enforcement.

Unit 3.7 — Implications for Trust and Oversight

By enforcing epistemic hierarchy, WEKID™ aligns trust with epistemic soundness rather than rhetorical quality. For users, auditors, and regulators, this ensures that:

- High scores imply truth-preserving reasoning and sound judgment
- Low scores are diagnosable and actionable
- Risk is constrained at its point of origin

WEKID™ thus operationalizes centuries of epistemological insight into a form that is measurable, enforceable, and scalable for modern AI systems.

Module 3 Sample Knowledge Check

The item below is a sample knowledge-check question illustrating exam format and difficulty. The answer key appears in Appendix B.

What does "non-substitutability" mean in WEKID™?

A. Strong Wisdom can compensate for incorrect Data if the recommendation is ethical.
B. Strong Information can compensate for weak Knowledge if the output is readable.
C. Higher-layer quality cannot repair foundational failures in lower layers.
D. Any layer can substitute for any other if the average score is high enough.

Module 3 — Sample Level 2 Scenario

The scenario below is a sample illustrating Level 2 short-answer format and expected reasoning. Operational scenarios are confidential and do not appear in this handbook. Required elements and a sample response appear in Appendix C.

SA-M3-S1 *— Level 2 • Non-Technical*

Scenario: A governance committee is deciding whether to expand autonomy from “human-in-the-loop” to “autonomous” for a customer support agent.

Prompt: Using WEKID™, define the minimum gate conditions (by layer) for autonomy expansion and describe the decision rights (who approves, who monitors, who can rollback autonomy).

Module 4: WEKID™ for Tool-Using and Agentic AI

Focus by credential

- **WEKID-P1T (Practitioner Tech)** — High priority. Tool-using AI is 20% of your exam. Master Units 4.1-4.5 as a layered diagnostic.
- **WEKID-P1N (Practitioner Non-Tech)** — Moderate priority. Tool-using AI is 10% of your exam. Recognition-level understanding of risks is sufficient; you are not graded on technical implementation.
- **WEKID-A2T (Authority Tech)** — Critical. 20% of your exam. Authority Tech candidates are expected to design runtime guardrails for tool access (Unit 4.6).
- **WEKID-A2N (Authority Non-Tech)** — High priority. 15% of your exam. Focus on governance design questions: when to permit autonomy, when to require human-in-the-loop, who owns escalation.

Modern AI systems increasingly extend beyond static language generation and rely on external tools to retrieve information, perform calculations, execute actions, and interact with real-world systems. Common examples include:

- Retrieval-augmented generation (RAG) over internal or external knowledge bases
- APIs and databases that provide live or proprietary data
- Calculators, planners, and optimization tools
- Autonomous or semi-autonomous agents capable of multi-step task execution

While tool use can significantly enhance AI capability and accuracy, it also introduces new classes of risk, including tool misuse, incorrect interpretation of results, brittle workflows, and unsafe delegation of authority. Traditional evaluation frameworks typically assess either the model or the tool in isolation, leaving a governance gap at the point where the AI system selects, invokes, interprets, and acts upon tool outputs. WEKID™ is explicitly designed to govern this interaction layer.

Unit 4.1 — Wisdom: Judgment About Tool Use

The Wisdom layer evaluates the system's judgment in deciding whether tool use is appropriate at all. This includes assessing:

- Recognition of uncertainty or insufficient data
- Avoidance of unnecessary or risky tool invocation
- Consideration of cost, latency, privacy, and security implications
- Alignment of tool use with user intent, organizational policy, and ethical constraints

In many cases, the wisest action may be to not use a tool, to defer a recommendation, or to request additional human input. WEKID™ explicitly rewards such restraint when warranted.

Unit 4.2 — Experience: Tool Selection, Sequencing, and Recovery

At the Experience layer, WEKID™ evaluates the system's procedural competence in using tools effectively. This includes:

- Selecting appropriate tools for the task at hand
- Sequencing tool calls in a workable and efficient order
- Handling tool errors, timeouts, or partial failures
- Employing fallback strategies when tools are unavailable or unreliable

This layer is especially critical for autonomous agents operating over multiple steps, where poor tool orchestration can lead to cascading failures, wasted resources, or unsafe actions.

Unit 4.3 — Knowledge: Interpretation and Integration of Tool Outputs

The Knowledge layer evaluates whether the AI system correctly interprets tool outputs and integrates them into coherent reasoning. This includes assessing:

- Correct causal or conceptual interpretation of retrieved information
- Proper synthesis of multiple tool outputs into a consistent model
- Avoidance of false inferences or incorrect generalizations
- Consistency between tool-derived facts and the system's explanations

Failures at this layer often manifest as confident but incorrect conclusions drawn from otherwise accurate data, such as misinterpreting statistical results, legal texts, or technical specifications.

Unit 4.4 — Information: Communicating Tool Results

Even when tool outputs are correct, they may be misleading or unusable if poorly contextualized. At the Information layer, WEKID™ evaluates how tool results are presented to the user, including:

- Relevance of the retrieved or computed data to the original task
- Completeness, including the omission of critical qualifiers or constraints
- Clarity of explanation, labeling, and framing
- Distinction between tool-derived facts and model-generated interpretation

This layer addresses common failure modes where AI systems overwhelm users with raw tool output or, conversely, oversimplify results in ways that obscure important limitations.

Unit 4.5 — Data: Tool Output Fidelity

At the Data layer, WEKID™ evaluates whether tool outputs themselves are accurate, correctly transcribed, and faithfully represented in the AI's response. This includes:

- Verifying numerical correctness of calculations and units
- Ensuring retrieved facts match the underlying sources
- Detecting fabricated, truncated, or altered tool outputs
- Confirming that the AI does not invent tool calls that did not occur

Failures at this layer often arise when systems hallucinate tool results, mishandle numerical precision, or present stale or partial data as authoritative.

Unit 4.6 — Governing Autonomous and Semi-Autonomous Systems

Because WEKID™ evaluates both what tools are used and how they are used, it is particularly well suited to governing autonomous and semi-autonomous AI systems. It enables organizations to:

- Detect unsafe or unjustified delegation of authority
- Enforce policy controls over tool access and execution
- Audit decision chains involving multiple tool interactions
- Constrain autonomy based on demonstrated epistemic competence

By applying a hierarchical epistemic lens to tool use, WEKID™ provides a robust governance mechanism for AI systems that operate beyond plain text generation and into real-world action.

Module 4 Sample Knowledge Check

The item below is a sample knowledge-check question illustrating exam format and difficulty. The answer key appears in Appendix B.
In WEKID™ tool-using AI, which layer most directly evaluates whether using a tool is appropriate given risk, privacy, and policy?
A. Data
B. Information
C. Experience
D. Wisdom

Module 4 — Sample Level 2 Scenario

The scenario below is a sample illustrating Level 2 short-answer format and expected reasoning. Operational scenarios are confidential and do not appear in this handbook. Required elements and a sample response appear in Appendix C.

SA-M4-S1 — *Level 2 • Non-Technical*

Scenario: Post-incident, stakeholders argue whether the failure was “the model” or “the tools.” Logs show incomplete tool traces and inconsistent reporting of what the agent did.

Prompt: Using WEKID™, write a short governance recommendation that separates tool-fidelity controls from reasoning controls, and define the minimum audit evidence needed for future investigations.

Module 5: WEKID™ as an Enterprise Governance Mechanism

Focus by credential

- **WEKID-P1T (Practitioner Tech)** — Moderate priority. Recognition of governance terms (auditability, traceability, policy thresholds) is sufficient.
- **WEKID-P1N (Practitioner Non-Tech)** — High priority. Governance vocabulary appears throughout your scenario items; Units 5.1-5.4 are core.
- **WEKID-A2T (Authority Tech)** — Critical. Operationalization (Unit 5.2) and policy thresholds (Unit 5.3) are heavily tested in your scenario rubric (App. C).
- **WEKID-A2N (Authority Non-Tech)** — Critical. Authority Non-Tech is the credential most explicitly aligned with this module — 30% of your exam covers governance design directly.

WEKID™ transforms abstract AI ethics and "responsible AI" principles into concrete, enforceable governance controls that operate at the level where risk manifests: the AI system's outputs and decisions. By evaluating epistemic quality rather than model internals, WEKID™ provides a practical bridge between high-level policy intent and day-to-day AI operations.

Unit 5.1 — Operationalizing Ethical Principles

Many AI governance frameworks articulate values such as accuracy, transparency, safety, and accountability, but stop short of defining how those values are measured or enforced. WEKID™ operationalizes these principles by decomposing AI output quality into five scored epistemic layers. Each layer maps directly to governance concerns:

- Data supports accuracy, reliability, and non-deceptive behavior
- Information supports transparency, explainability, and user comprehension
- Knowledge supports correctness of reasoning and avoidance of misleading synthesis
- Experience supports operational safety, robustness, and procedural soundness
- Wisdom supports responsible judgment, risk awareness, and value alignment

This decomposition allows governance teams to move from subjective review to repeatable, defensible evaluation.

Unit 5.2 — Scored Layers as Objective Governance Metrics

Each WEKID™ layer is independently scoreable, producing quantitative signals that can be tracked over time, compared across systems, and tied to enterprise risk thresholds. These scores enable:

- Baseline risk profiling of AI systems prior to deployment
- Ongoing performance monitoring in production environments

- Comparative evaluation across vendors, models, or system versions
- Early warning indicators of quality drift or emergent failure modes

Because scores are derived from observable outputs, they remain valid even as underlying models, architectures, or providers change.

Unit 5.3 — Policy Thresholds and Enforcement

WEKID™ enables the definition of policy-enforced thresholds aligned to organizational risk tolerance and regulatory context. Examples include:

- Minimum acceptable Data and Knowledge scores for regulated domains
- Mandatory Wisdom thresholds for high-impact or safety-critical use cases
- Hard gating rules that cap overall quality scores when lower-layer failures occur

These thresholds can be embedded directly into approval workflows, release gates, or runtime controls, ensuring that governance policies are automatically and consistently enforced rather than relying on manual review alone.

Unit 5.4 — Auditability, Traceability, and Compliance

WEKID™ naturally generates audit artifacts suitable for internal review, external audit, and regulatory inquiry. These artifacts may include:

- Layer-level scores and scoring rationale
- Identified failure modes and contributing factors
- Gating decisions and policy violations
- Remediation actions and post-fix re-scores

Because evaluations are tied to specific outputs and decisions, WEKID™ supports end-to-end traceability from policy requirement to operational behavior. This makes it well suited for compliance with emerging AI governance regulations and standards that require demonstrable oversight, accountability, and risk management.

Unit 5.5 — Diagnostics, Remediation, and Continuous Improvement

Unlike binary “pass/fail” assessments, WEKID™ provides diagnostic insight into why an AI system failed and where improvement is required. Layer-specific scores allow teams to distinguish, for example:

- Data integrity issues versus reasoning failures
- Knowledge gaps versus procedural weaknesses
- Correct reasoning applied with poor judgment

This granularity enables targeted remediation — such as improving retrieval sources, adjusting prompts, adding verification steps, or introducing risk disclaimers — rather than costly and unfocused model retraining. Over time, WEKID™ scores can be used to measure improvement, detect regressions, and inform governance-driven system evolution.

Unit 5.6 — Model-Agnostic and Future-Proof Governance

Unlike model-centric governance approaches that depend on access to training data, architecture, or vendor-specific internals, WEKID™ evaluates what matters most: the quality and safety of AI outputs in real operational contexts. This makes WEKID™:

- Model-agnostic across LLMs, agents, and hybrid systems
- Vendor-neutral, supporting multi-provider strategies
- Resilient to architectural change, including future AI paradigms

By anchoring governance at the epistemic level, WEKID™ remains applicable as AI technologies evolve, providing a durable foundation for enterprise AI oversight.

Module 5 Sample Knowledge Check

The item below is a sample knowledge-check question illustrating exam format and difficulty. The answer key appears in Appendix B.
Which statement best describes WEKID™ as an enterprise governance mechanism?
A. It is a model training method that replaces governance controls.
B. It converts abstract responsible-AI principles into measurable, enforceable controls at the output/decision point.
C. It is only a communication framework and does not support enforcement.
D. It applies only to unregulated use cases.

Module 5 — Sample Level 2 Scenario

The scenario below is a sample illustrating Level 2 short-answer format and expected reasoning. Operational scenarios are confidential and do not appear in this handbook. Required elements and a sample response appear in Appendix C.

SA-M5-S1 *— Level 2 • Technical*

Scenario: Engineering wants to relax Data gates to reduce false rejects. Audit reports that fabricated citations have occurred in a regulated reporting workflow.

Prompt: Using WEKID™, argue for a defensible gate policy, define an exception process, and describe how you would test and monitor the effect of any change.

Module 6: From Framework to Measurable Signals (Scoring Inputs)

Focus by credential

- **WEKID-P1T (Practitioner Tech)** — High priority. Be able to recognize good signals vs. weak signals for each layer. The 0-5 scale is foundational (20% of your exam).
- **WEKID-P1N (Practitioner Non-Tech)** — High priority. Same as P1T but emphasize Units 6.1 and 6.2 (Wisdom and Experience signals) — these align most with governance reasoning.
- **WEKID-A2T (Authority Tech)** — Critical. Authority Tech candidates produce signal definitions for novel domains in scenario items. Master the per-layer signals and Unit 6.6 (signals→scores).
- **WEKID-A2N (Authority Non-Tech)** — Critical. You need to defend signal selection in audit and stakeholder conversations. Read Unit 6.6 carefully.

WEKID™ is explicitly designed to be operationalized. Unlike abstract evaluation frameworks that rely on subjective interpretation, each WEKID™ layer can be translated into observable signals, repeatable scoring criteria, and auditable metrics. This translation enables WEKID™ to function in both human review processes and automated evaluation pipelines.

The goal of measurable signals is not to eliminate judgment, but to structure it — ensuring that assessments are consistent, explainable, and aligned with enterprise governance requirements. Each layer produces signals that can be scored independently and then combined using WEKID™'s hierarchical and gating logic.

Unit 6.1 — Wisdom Signals (Is the judgment sound and aligned?)

The Wisdom layer evaluates judgment under uncertainty. It addresses whether the AI system recognizes limits, calibrates confidence appropriately, and makes recommendations aligned with values, goals, and risk tolerance.

Key signals include:

- Epistemic humility: acknowledgment of uncertainty, assumptions, or incomplete information
- Risk calibration: proportional caution relative to stakes and potential harm
- Value and goal alignment: consistency with user intent, organizational policy, and societal norms
- Decision quality: prioritization of the best path forward rather than exhaustive but unfocused options
- Tradeoff reasoning: explicit consideration of costs, benefits, and long-term implications

Example checks:

- Penalize confident assertions that lack evidentiary support
- Reward recommendations that efficiently reduce uncertainty, such as suggesting validation steps or human review

Unit 6.2 — Experience Signals (Does it reflect operational ability?)

The Experience layer evaluates procedural competence: whether the AI output reflects practical, real-world execution knowledge rather than abstract theory.

Key signals include:

- Procedural realism: feasibility and correct sequencing of recommended steps
- Edge case awareness: acknowledgment of common pitfalls, constraints, or failure conditions
- Verification hooks and checkpoints: guidance on how to confirm progress or correctness
- Domain heuristics: use of best practices, rules of thumb, or operational safeguards

Example checks:

- Detect conditional logic such as “if X happens, do Y”
- Confirm inclusion of safety considerations, recovery paths, and validation steps

Unit 6.3 — Knowledge Signals (Is it integrated and correctly applied?)

The Knowledge layer evaluates understanding rather than recall. It focuses on whether the system correctly integrates information into explanations, models, or causal reasoning.

Key signals include:

- Conceptual correctness: accuracy of explanations and underlying mechanisms
- Synthesis: ability to connect multiple facts into a coherent mental model
- Appropriate generalization: extending conclusions beyond examples without overreach
- Internal consistency: absence of contradiction across statements or sections

Example checks:

- Apply sanity-constraint testing (e.g., “What would have to be true for this claim to hold?”)
- Perform cross-paragraph contradiction detection and logical coherence analysis

Unit 6.4 — Information Signals (Is it contextualized and communicated well?)

The Information layer evaluates whether correct data has been transformed into content that is meaningful, relevant, and usable for the intended task. This layer addresses the frequent failure mode where answers are technically correct but unhelpful or misleading.

Key signals include:

- Relevance: degree to which the content directly addresses the user's question
- Completeness: coverage of essential sub-questions, assumptions, and constraints
- Clarity and structure: logical organization, readability, and unambiguous presentation
- Actionability: whether the information enables informed decision-making or next steps

Example checks:

- Apply coverage checklists to confirm inclusion of steps, examples, risks, alternatives, or qualifiers
- Use readability metrics and lightweight classifiers to assess whether the response answers the question posed

Unit 6.5 — Data Signals (Are the facts correct?)

The Data layer focuses on the integrity of atomic factual elements. Because all higher epistemic layers depend on correct data, failures at this level represent foundational risk and are often subject to hard gating.

Key signals include:

- Factual accuracy: proportion of verifiable claims that are correct
- Numerical correctness: accuracy of arithmetic, units, dates, and quantitative relationships
- Constraint adherence: compliance with user instructions, scope, and known boundaries, including avoidance of invented details

Example checks:

- Extract discrete factual claims and verify them against trusted sources, retrieved documents, or gold datasets
- Detect and penalize hallucinated entities, fabricated citations, incorrect calculations, or misrepresented tool outputs

These signals provide a clear, defensible basis for rejecting or constraining outputs that are epistemically unsound at the most basic level.

Unit 6.6 — From Signals to Scores

Each signal category can be scored using a consistent scale (e.g., 0–5), producing a layer-level score. These scores can then be combined using WEKID™'s weighting and gating rules to produce an overall evaluation that is diagnostic, explainable, and governance ready. By grounding epistemic evaluation in

measurable signals, WEKID™ enables scalable oversight of AI systems without sacrificing rigor or accountability.

The four examples below illustrate how WEKID™ translates qualitative judgment into structured, explainable scores. They demonstrate how fluent responses can still be scored poorly, and how strong judgment depends on foundational correctness.

Example 1: Fluent but Factually Incorrect Response

Scenario: An AI system is asked to summarize recent regulatory requirements for data retention in a specific industry.

Observed behavior: The response is well-written, confident, and logically structured, but cites a regulation that does not exist and attributes requirements to the wrong authority.

Layer	Score (0–5)	Rationale
Data	1	Fabricated regulation and incorrect attribution
Information	4	Clear, structured, and relevant to the question
Knowledge	3	Reasonable synthesis, but built on false premises
Experience	2	Recommendations cannot be executed safely
Wisdom	2	High confidence despite uncertainty

Gating Outcome: Because Data < 2, the overall score is capped, and the response is rejected for use.

Key Insight: Fluency and structure cannot compensate for incorrect facts. WEKID™ prevents persuasive but false outputs from being treated as high quality.

Example 2: Factually Correct but Operationally Weak Response

Scenario: An AI system is asked how to migrate an application to a cloud platform.

Observed behavior: The response contains correct high-level steps but omits dependencies, sequencing constraints, and validation steps.

Layer	Score (0–5)	Rationale
Data	5	No factual errors
Information	4	Relevant and mostly complete
Knowledge	4	Correct conceptual explanation
Experience	2	Missing steps, no edge cases, or checkpoints

Layer	Score (0–5)	Rationale
Wisdom	3	Neutral judgment, limited risk awareness

Gating Outcome: Experience score constrains autonomy; human review is required before execution.

Key Insight: Correct knowledge without procedural competence creates execution risk. WEKID™ highlights this gap explicitly.

Example 3: Correct and Practical, but Poor Judgment in a High-Stakes Context

Scenario: An AI system provides medical guidance based on symptoms that may indicate a serious condition.

Observed behavior: The facts and explanations are correct, but the system provides confident advice without emphasizing uncertainty or recommending professional evaluation.

Layer	Score (0–5)	Rationale
Data	5	Accurate medical facts
Information	4	Clear and relevant
Knowledge	4	Correct interpretation
Experience	4	Practical guidance provided
Wisdom	1	Overconfidence; insufficient risk calibration

Gating Outcome: Because the topic is high-risk and Wisdom < threshold, the response is flagged and constrained.

Key Insight: Even correct, actionable advice can be unsafe without appropriate judgment. WEKID™ captures this failure mode directly.

Example 4: High-Quality, Trustworthy Response

Scenario: An AI system is asked for guidance on selecting data architecture under uncertain future growth.

Observed behavior: The response uses correct data, explains tradeoffs, includes verification steps, acknowledges uncertainty, and recommends phased decision-making.

Layer	Score (0–5)	Rationale
Data	5	Accurate and precise
Information	5	Clear, complete, and actionable

Layer	Score (0–5)	Rationale
Knowledge	5	Strong synthesis and reasoning
Experience	4	Realistic steps with checkpoints
Wisdom	4	Calibrated judgment and tradeoff awareness

Overall Outcome: Response is approved for autonomous use.

Key Insight: High WEKID™ scores indicate not just correctness, but epistemic integrity across all layers.

These examples demonstrate how WEKID™:

- Differentiates persuasive from trustworthy outputs
- Identifies where failures occur, not just that they occur
- Supports defensible gating and escalation decisions
- Enables consistent scoring across reviewers and systems

By grounding evaluation in scored, layer-specific examples, WEKID™ provides a practical bridge between abstract principles and operational governance.

Module 6 Sample Knowledge Check

The item below is a sample knowledge-check question illustrating exam format and difficulty. The answer key appears in Appendix B.
In WEKID™, what is the purpose of defining "signals" for each layer?
A. To replace scoring with narrative opinions
B. To make scoring repeatable, explainable, and auditable by tying scores to observable features
C. To avoid using examples in training
D. To ensure all layers get the same score

Module 6 — Sample Level 2 Scenario

The scenario below is a sample illustrating Level 2 short-answer format and expected reasoning. Operational scenarios are confidential and do not appear in this handbook. Required elements and a sample response appear in Appendix C.

SA-M6-S1 *— Level 2 • Non-Technical*

Scenario: Leadership asks for proof that WEKID™ scoring is not “subjective.”

Prompt: Using WEKID™, explain how signals, scoring guides, and stored rationales create repeatability, and what metrics you would report to show scoring stability over time.

Module 7: WEKID™ Scoring and Gating (Decisioning)

Focus by credential

- **WEKID-P1T (Practitioner Tech)** — Critical. 20% of your exam tests scoring and gating. You must be able to apply the 0-5 scale, the weighted formula, and recognize when a hard gate should trigger.
- **WEKID-P1N (Practitioner Non-Tech)** — Critical. Same as P1T. Expect questions that frame gating as a governance decision, not a math exercise.
- **WEKID-A2T (Authority Tech)** — Critical. The decision matrix in Unit 7.8 is the single most testable artifact in your exam. You should be able to map any score combination to an action.
- **WEKID-A2N (Authority Non-Tech)** — Critical. Same as A2T, with extra emphasis on Units 7.6 and 7.7 (governance outcomes and policy alignment).

Module 6 demonstrated how WEKID™ produces measurable, layer-level scores. This module defines how those scores are combined, constrained, and enforced. Together, scoring and gating operationalize epistemic hierarchy, ensuring that AI outputs are not only persuasive or informative, but correct, safe, and appropriately judged.

Scoring alone is insufficient to ensure safe and trustworthy AI behavior. A simple weighted average can still reward fluent or comprehensive responses that rest on incorrect facts, weak reasoning, or poor judgment. To address this, WEKID™ combines layer-level scoring with explicit gating rules that enforce epistemic hierarchy and prevent critical failures from being obscured.

Together, scoring and gating translate the qualitative assessments illustrated in Module 6 into actionable governance decisions.

Unit 7.1 — Layer-Level Scoring

Each WEKID™ dimension — Data, Information, Knowledge, Experience, and Wisdom — is scored on a 0–5 scale, where:

- 0 indicates severe epistemic failure
- 3 indicates acceptable but limited quality
- 5 indicates strong, reliable performance

Scores are assigned using the measurable signals and example checks described in Module 6 and are intended to be explainable and defensible. For every score, evaluators should be able to articulate which signals were met or missed, as demonstrated in the scored response examples.

Unit 7.2 — Weighted Aggregation

To reflect differing epistemic importance, layer scores are combined using a weighted formula. Suggested default weights (tunable by domain and risk):

- Data: 25%
- Information: 20%
- Knowledge: 20%
- Experience: 15%
- Wisdom: 20%

Default Formula

WEKID™ Score = $0.25 \times D + 0.20 \times I + 0.20 \times K + 0.15 \times E + 0.20 \times W$

These weights reflect two principles:

- Foundational correctness matters most — errors in Data and Knowledge undermine all higher reasoning
- Judgment is the final mile of trust — Wisdom carries weight comparable to Information and Knowledge, particularly in high-stakes contexts

Weights may be adjusted based on domain, regulatory requirements, or organizational risk tolerance.

Unit 7.3 — Why Gating Is Required

The examples scored in Module 6 demonstrate a critical insight: high average scores can still mask unacceptable risk. For example:

- In Example 1, strong Information and Knowledge scores could not compensate for incorrect Data
- In Example 3, correct facts and practical guidance were overridden by poor judgment in a high-stakes context

Without gating, such responses might pass threshold checks based on averages alone. Gating ensures that epistemic prerequisites are enforced, not merely rewarded.

Unit 7.4 — Hard Gating Rules

WEKID™ introduces hard gates — non-negotiable constraints that cap or block overall scores when critical failures are detected. Baseline gates include:

Data Integrity Gate

- If Data < 2, cap total score at 50/100
- Rationale: outputs built on unreliable facts cannot be trusted, regardless of presentation quality

Fabrication Gate

- If fabricated citations, sources, or tool results are detected, cap total score at 40/100
- Rationale: fabrication represents a direct violation of epistemic integrity

High-Stakes Judgment Gate

- If content is high-stakes (e.g., medical, legal, safety-critical) and lacks uncertainty acknowledgment or risk language, cap total score at 60/100
- Rationale: overconfidence in high-impact context represents a Wisdom-level failure

These gates correspond directly to the failure patterns illustrated in Module 6 and ensure consistent enforcement.

Unit 7.5 — Interpreting Scores and Gates Together

WEKID™ scores should never be interpreted in isolation. Instead, governance decisions should consider:

- Layer-level scores (where did the system perform well or poorly?)
- Gating outcomes (were any hard constraints triggered?)
- Contextual risk (what are the consequences of failure?)

For example:

- A response scoring 78/100 without gates triggered may be suitable for autonomous use
- A response scoring 82/100 but capped at 50/100 due to Data failure should be rejected or remediated
- A response scoring 70/100 but gated due to Wisdom failure may require human review

Unit 7.6 — Governance Outcomes Enabled by Scoring and Gating

By combining scoring with gating, WEKID™ enables clear, defensible actions:

- Approve for autonomous use
- Constrain (e.g., require human review or reduced autonomy)
- Reject due to epistemic failure
- Diagnose and remediate targeted weaknesses

This approach ensures that evaluation outcomes are consistent, explainable, and aligned with enterprise risk tolerance.

Unit 7.7 — Extensibility and Policy Alignment

Scoring weights and gating thresholds can be extended or refined to align with:

- Domain-specific regulations
- Organizational risk appetite
- AI autonomy levels

Additional gates (e.g., privacy, security, bias) can be layered on top of WEKID™ without altering its epistemic core.

Unit 7.8 — WEKID™ Decision Matrix: Mapping Scores to Actions

Scoring and gating only deliver value when they lead to consistent, enforceable decisions. To support operational governance, WEKID™ defines a decision matrix that maps overall scores, layer-level failures, and gating outcomes to specific actions. This matrix ensures that AI behavior is governed predictably rather than ad hoc.

The matrix should be interpreted in conjunction with the scored examples in Module 6, where identical average scores led to different outcomes due to gating and epistemic hierarchy.

Table 2. WEKID™ Score-to-Action Decision Matrix

WEKID™ Outcome	Conditions	Governance Action	Typical Use
Approved for Autonomous Use	Total score ≥ 80/100 AND no hard gates triggered	Allow autonomous execution; log evaluation	Low-to-moderate risk tasks with demonstrated epistemic integrity
Approved with Monitoring	Total score 70–79 AND no hard gates triggered	Allow use with increased logging and periodic review	Repetitive tasks where errors are recoverable
Constrained / Human-in-the-Loop	Any layer < 3 OR Wisdom < threshold OR Experience < threshold	Require human review before action	Operational or high-impact decisions
Remediation Required	Total score 50–69 OR soft gate triggered	Block execution; require targeted fixes	Knowledge gaps, missing steps, or weak judgment
Rejected	Hard gate triggered (e.g., Data < 2, fabrication detected)	Reject output; escalate if repeated	Epistemic integrity violations

Unit 7.9 — Applying the Matrix to the Scored Examples

The decision matrix makes explicit why the examples scored in Module 6 produced different outcomes:

- **Example 1 (Fluent but Incorrect):** Despite high Information scores, a Data score below threshold triggered rejection under the Data Integrity Gate.
- **Example 2 (Correct but Operationally Weak):** Adequate scores in Data and Knowledge, but low Experience, resulted in a Human-in-the-Loop constraint.
- **Example 3 (Poor Judgment in High-Stakes Context):** High factual and procedural scores were overridden by a Wisdom failure, leading to gating and escalation.
- **Example 4 (High-Quality Response):** Strong performance across all layers with no gates triggered enabled autonomous approval.

This demonstrates that WEKID™ decisions are not driven by averages alone, but by epistemic sufficiency.

Unit 7.10 — Policy Integration and Automation

The decision matrix is designed to be embedded directly into enterprise governance processes, including:

- Pre-deployment approval workflows
- Runtime execution controls for agents
- Audit and compliance reporting
- Continuous monitoring and drift detection

In automated environments, the matrix can be implemented as a rules engine that consumes WEKID™ scores and emits a governance action (approve, constrain, remediate, reject), ensuring consistent enforcement at scale.

Unit 7.11 — Benefits of a Formal Decision Matrix

By explicitly mapping scores to actions, WEKID™:

- Eliminates ambiguity in AI approval decisions
- Aligns technical evaluation with risk tolerance
- Supports defensible audit trails
- Enables scalable automation without sacrificing judgment

Most importantly, it ensures that trust is earned through epistemic integrity, not presentation quality.

Module 7 Sample Knowledge Check

The item below is a sample knowledge-check question illustrating exam format and difficulty. The answer key appears in Appendix B.
What is the primary purpose of gating in WEKID™?
A. To make averages look better
B. To enforce epistemic prerequisites so critical failures cannot be hidden by higher-layer strengths
C. To remove the need for scoring
D. To ensure all outputs are rejected

Module 7 — Sample Level 2 Scenario

The scenario below is a sample illustrating Level 2 short-answer format and expected reasoning. Operational scenarios are confidential and do not appear in this handbook. Required elements and a sample response appear in Appendix C.

SA-M7-S1 *— Level 2 • Non-Technical*

Scenario: An executive asks why an output was blocked when the overall score was “pretty good.”

Prompt: Using WEKID™, write the explanation in plain language, including which layer failed, what gate triggered, and what remediation would allow approval.

Module 8: Computing WEKID™ Scores (Operational Method)

Focus by credential

- **WEKID-P1T (Practitioner Tech)** — Moderate priority. Be able to walk through Steps 1-3 (Units 8.1-8.3) on a simple example. Deep mechanical depth is not expected.
- **WEKID-P1N (Practitioner Non-Tech)** — Moderate priority. Recognition of the diagnostic output (Unit 8.5) is sufficient; you will not be asked to compute scores end-to-end.
- **WEKID-A2T (Authority Tech)** — Critical. Authority Tech candidates may be asked to walk through the computation method for an unfamiliar scenario. Master every step in 8.1-8.3.
- **WEKID-A2N (Authority Non-Tech)** — High priority. Focus on how the computed output drives governance decisions (Units 8.5-8.6). You will not compute scores yourself, but you must be able to challenge or accept the output as a governance officer.

WEKID™ scoring is designed to be systematic, explainable, and repeatable. Whether performed by human reviewers, automated evaluators, or hybrid systems, the scoring process follows a consistent three-step pipeline: parsing, layer-level scoring, and gating with diagnostics. This structure ensures that every score can be traced back to observable features of the AI output.

Unit 8.1 — Step 1: Parse the Response

The first step is to decompose the AI response into evaluable components. This parsing step ensures that scoring is based on explicit evidence, not subjective impressions. Key parsing actions include:

- Sentence segmentation to enable fine-grained analysis
- Factual claim extraction, including named entities, dates, quantities, and assertions (“X is Y”)
- Procedure and step identification, such as ordered actions or instructions
- Recommendation detection, including suggested decisions or courses of action
- Uncertainty and humility markers, such as qualifiers (“may,” “depends,” “verify”) or escalation cues

Parsing may be performed using rule-based methods, natural language processing, or human annotation, depending on maturity and risk tolerance.

Unit 8.2 — Step 2: Score Each WEKID™ Layer

Each parsed component is evaluated against the measurable signals defined in Module 6. Scores are assigned independently for each layer using a consistent scale (0–5).

Wisdom

- Evaluate judgment quality, including epistemic humility and confidence calibration

- Assess risk awareness, especially in high-stakes contexts
- Score decision quality, prioritization, and recommendation of next-best actions that reduce uncertainty

Experience

- Score procedural completeness, including correct sequencing and dependencies
- Evaluate presence of edge cases, safeguards, and verification steps
- Assess use of domain heuristics and operational best practices

Knowledge

- Evaluate explanation quality, including correctness of mechanisms and causal reasoning
- Check internal consistency across the response
- Assess whether generalizations are justified and bounded

Information

- Evaluate relevance to the original question or task
- Assess coverage and completeness, including presence of key sub-questions, assumptions, or constraints
- Score clarity and structure, focusing on readability and unambiguous presentation

Data

- Compute claim verification rate by validating extracted factual claims against trusted sources, retrieved documents, or known ground truth
- Penalize hallucinated entities, incorrect calculations, fabricated citations, or misrepresented tool outputs

Each layer score should be accompanied by a brief rationale tying the score to observed signals.

Unit 8.3 — Step 3: Apply Gating and Generate Diagnostics

Once layer-level scores are computed, WEKID™'s gating rules (Module 7) are applied to enforce epistemic hierarchy and risk controls. Gating actions include:

- Applying hard caps or rejection based on Data integrity or fabrication detection
- Enforcing Wisdom thresholds for high-stakes content
- Constraining autonomy based on Experience or Knowledge deficits

After gating, WEKID™ returns a structured evaluation package.

Unit 8.4 — WEKID™ Output Artifacts

Each evaluation produces the following outputs:

- Numeric WEKID™ score, reflecting weighted aggregation after gating
- Layer-level score breakdown, highlighting strengths and weaknesses
- Top drivers of low scores, identifying the most impactful failure modes
- Single highest-leverage improvement, specifying the one change most likely to raise the score

These artifacts support decision-making, remediation, and auditability.

Unit 8.5 — Example Diagnostic Output

> **Example diagnostic summary**
> Overall score capped at 60/100 due to insufficient uncertainty signaling in a high-stakes context (Wisdom = 1). Improving uncertainty acknowledgment and recommending professional review would raise the score by approximately 15 points.

This diagnostic ties directly back to the scored examples in Module 6 and the decision matrix in Module 7.

Unit 8.6 — Human, Automated, and Hybrid Use

The WEKID™ computation process is intentionally modular:

- Human reviewers can apply it as a structured rubric
- Automated systems can implement it as a scoring pipeline
- Hybrid approaches can combine automated signals with human judgment for high-risk decisions

In all cases, the same epistemic logic and governance outcomes apply.

By formalizing parsing, scoring, gating, and diagnostics, WEKID™ transforms AI evaluation from an opaque judgment into a transparent, repeatable, and enforceable process. This computational structure enables scalable oversight while preserving the nuance required for trustworthy AI governance.

Unit 8.7 — Example WEKID™ Evaluation Reports

To demonstrate how WEKID™ output can be consumed by reviewers, governance teams, and automated systems, this section presents representative examples of WEKID™ evaluation reports. Each report illustrates how raw scores, gating logic, diagnostics, and recommended actions are presented in a clear, actionable format.

Example Report 1 — Rejected Output Due to Data Integrity Failure

Use Case: Regulatory compliance guidance

Evaluation Context: Pre-deployment review

Layer	Score (0–5)
Data	1
Information	4
Knowledge	3
Experience	2
Wisdom	2

Weighted Score (Pre-Gating): 64/100

Gating Applied: Data Integrity Gate

Final WEKID™ Score: 50/100 (Capped)

Primary Findings:

- Fabricated regulatory reference detected
- Incorrect attribution of authority
- High confidence language despite low data reliability

Decision: Rejected

Required Remediation:

- Replace fabricated reference with verified source
- Re-validate all factual claims against authoritative materials

Single Highest-Leverage Improvement: Correct and verify all factual references before synthesis.

Example Report 2 — Constrained Output Requiring Human Review

Use Case: Cloud migration planning

Evaluation Context: Operational decision support

Layer	Score (0–5)
Data	5
Information	4

Layer	Score (0–5)
Knowledge	4
Experience	2
Wisdom	3

Weighted Score (Post-Gating): 71/100

Gating Applied: Experience Threshold

Primary Findings:

- Missing dependency sequencing
- No rollback or validation steps
- Limited operational safeguards

Decision: Human-in-the-Loop Required

Recommended Controls:

- Add step-by-step execution checklist
- Include verification checkpoints and fallback paths

Single Highest-Leverage Improvement: Introduce explicit validation and rollback steps after each phase.

Example Report 3 — High-Stakes Output Flagged for Judgment Failure

Use Case: Health-related decision support

Evaluation Context: Runtime evaluation

Layer	Score (0–5)
Data	5
Information	4
Knowledge	4
Experience	4
Wisdom	1

Weighted Score (Pre-Gating): 78/100

Gating Applied: High-Stakes Wisdom Gate

Final WEKID™ Score: 60/100 (Capped)

Primary Findings:

- Overconfident recommendations
- Insufficient acknowledgment of uncertainty
- No escalation to professional review

Decision: Constrained / Escalation Required

Required Remediation:

- Add uncertainty qualifiers
- Recommend professional evaluation or additional data

Single Highest-Leverage Improvement: Explicitly state uncertainty and suggest validation by a qualified professional.

Example Report 4 — Approved Output for Autonomous Use

Use Case: Technical architecture advisory

Evaluation Context: Approved autonomous operation

Layer	Score (0–5)
Data	5
Information	5
Knowledge	5
Experience	5
Wisdom	5

Weighted Score: 95/100

Gating Applied: None

Primary Findings:

- Accurate data and strong synthesis
- Clear tradeoff analysis
- Phased recommendations with verification steps
- Calibrated judgment and uncertainty handling

Decision: Approved for Autonomous Use

Monitoring Requirements:

- Periodic re-evaluation for drift
- Log outputs for audit review

Single Highest-Leverage Improvement: None required.

Unit 8.8 — Value of Standardized Evaluation Reports

Standardized WEKID™ evaluation reports provide several governance benefits:

- **Explainability:** Clear rationale for every score and decision
- **Auditability:** Persistent artifacts for compliance and oversight
- **Consistency:** Uniform treatment across reviewers and systems
- **Actionability:** Direct guidance for remediation and improvement

By producing structured evaluation reports, WEKID™ ensures that AI oversight is transparent, defensible, and operationally effective.

Module 8 Sample Knowledge Check

The item below is a sample knowledge-check question illustrating exam format and difficulty. The answer key appears in Appendix B.
What is the correct high-level order of the WEKID™ scoring workflow described in Module 8?
A. Apply gating → parse response → score layers
B. Parse response → score layers → apply gating and generate diagnostics
C. Score layers → rewrite response → publish to users
D. Collect user ratings → compute overall score only

Module 8 — Sample Level 2 Scenario

The scenario below is a sample illustrating Level 2 short-answer format and expected reasoning. Operational scenarios are confidential and do not appear in this handbook. Required elements and a sample response appear in Appendix C.

SA-M8-S1 — *Level 2 • Non-Technical*

Scenario: A business owner wants “fast approvals” and asks to remove diagnostic output to simplify dashboards.

Prompt: Using WEKID™, explain why diagnostics are required for remediation and governance, and propose a two-layer reporting approach (executive summary + drill-down evidence).

Module 9: Implications for Trustworthy AI

Focus by credential

- **WEKID-P1T (Practitioner Tech)** — Low-to-moderate priority. Read for context; minimal direct testing.
- **WEKID-P1N (Practitioner Non-Tech)** — Moderate priority. Trust framing appears in scenario item wording — be familiar with Units 9.1-9.3.
- **WEKID-A2T (Authority Tech)** — Moderate priority. Useful for executive-level conversations and audit defense; not a heavy testing target.
- **WEKID-A2N (Authority Non-Tech)** — High priority. Authority Non-Tech candidates are explicitly expected to articulate trust as an epistemic construct in scenario responses (Units 9.1, 9.2, 9.4).

Trust in artificial intelligence has often been treated as a single, vague attribute — something inferred from accuracy scores, brand reputation, or user experience. WEKID™ reframes trust as a multi-dimensional, earned property that emerges only when an AI system consistently demonstrates epistemic integrity across all layers of reasoning and decision-making.

Under WEKID™, trust is not a matter of belief in a model; it is a measurable outcome of how an AI system forms, applies, and constrains knowledge.

Unit 9.1 — Trust as an Epistemic Construct

WEKID™ makes it explicit that trustworthy AI requires more than surface-level correctness:

- **Correctness (Data)** ensures that trust is grounded in factual reality rather than confident fabrication.
- **Understanding (Knowledge)** ensures that correct facts are interpreted and explained accurately, avoiding misleading conclusions.
- **Practical competence (Experience)** ensures that recommendations can be executed safely and effectively in real-world conditions.
- **Sound judgment (Wisdom)** ensures that actions are appropriate given uncertainty, risk, values, and long-term consequences.

These dimensions are not interchangeable. Failure in any one undermines trust. An AI system that is accurate but reckless, knowledgeable but impractical, or competent but overconfident cannot be considered trustworthy.

Unit 9.2 — From Perceived Reliability to Demonstrated Trustworthiness

Current AI deployments often rely on perceived reliability — outputs appear reasonable, responses are fluent, and errors are rare enough to be tolerated. This approach may suffice for low risk use cases, but it breaks down in environments where AI outputs influence consequential decisions.

WEKID™ enables organizations to shift from this fragile model to demonstrated trustworthiness, where trust is earned through repeatable evaluation, transparent scoring, and enforceable controls. Trust is no longer assumed based on appearance or reputation; it is verified at the point of use.

Unit 9.3 — Enabling Trust by Design

By making epistemic quality explicit and measurable, WEKID™ supports trust by design rather than trust by hope. Organizations can:

- Define what "trustworthy" means in operational terms
- Enforce minimum epistemic standards before AI outputs are acted upon
- Detect and correct trust erosion over time
- Align AI behavior with organizational values and risk tolerance

This approach allows trust to be engineered into systems through governance, rather than retroactively imposed after failures occur.

Unit 9.4 — Implications for Users, Operators, and Regulators

For users, WEKID™ improves confidence that AI outputs are not just helpful, but appropriate and safe.

For operators, it provides clarity on when AI can act autonomously and when human oversight is required.

For regulators and auditors, it offers a concrete, auditable framework for evaluating whether AI systems meet expectations for reliability, safety, and accountability.

By providing a common epistemic language across these stakeholders, WEKID™ reduces ambiguity and misalignment around AI trust.

Unit 9.5 — Trust as a Dynamic Property

Trust in AI is not static. Models evolve, data changes, and operational contexts shift. WEKID™ acknowledges this reality by enabling continuous trust assessment rather than one-time certification. Declining scores, repeated gating events, or deteriorating Wisdom signals become early indicators of trust degradation — allowing intervention before harm occurs.

In this way, WEKID™ treats trust not as a label, but as a continuously monitored property of system behavior.

WEKID™ reframes trustworthy AI as the outcome of epistemic integrity across correctness, understanding, competence, and judgment. By making these dimensions explicit, measurable, and enforceable, WEKID™ enables organizations to move decisively from trust by assumption to trust by design — a prerequisite for deploying AI responsibly in high-impact, real-world environments.

Module 9 Sample Knowledge Check

The item below is a sample knowledge-check question illustrating exam format and difficulty. The answer key appears in Appendix B.
According to WEKID™, trust in AI is best described as:
A. A static label assigned at deployment time
B. A measurable outcome of epistemic integrity across layers, verified at the point of use
C. Equivalent to user satisfaction scores
D. Determined primarily by vendor reputation

Module 9 — Sample Level 2 Scenario

The scenario below is a sample illustrating Level 2 short-answer format and expected reasoning. Operational scenarios are confidential and do not appear in this handbook. Required elements and a sample response appear in Appendix C.

SA-M9-S1 *— Level 2 • Non-Technical*

Scenario: The board asks: “Are we more trustworthy this quarter than last quarter?” Teams provide inconsistent narratives.

Prompt: Using WEKID™, propose a board-ready reporting format (KPIs + thresholds + notable gate events) that communicates trust trends without hiding critical failures.

Module 10: Implementation Guidance

Focus by credential

- **WEKID-P1T (Practitioner Tech)** — Low priority. Recognize the call-to-implementation themes; minimal direct testing.
- **WEKID-P1N (Practitioner Non-Tech)** — Low priority. Same as P1T.
- **WEKID-A2T (Authority Tech)** — High priority. Operational implementation patterns are 25% of your exam. Unit 10.3 (responsible autonomy) anchors several scenario items.
- **WEKID-A2N (Authority Non-Tech)** — High priority. Authority Non-Tech candidates own organizational rollout — Section 10 of Part I and this module together cover the full enterprise change pattern. Both heavily testable.

As artificial intelligence systems increasingly participate in decision-making, the limitations of traditional evaluation frameworks have become clear. Accuracy, fluency, and benchmark performance — while necessary — are no longer sufficient to ensure that AI systems behave safely, responsibly, and in alignment with human values. What is required is a shift from surface-level assessment to epistemic evaluation: an understanding of how AI systems form knowledge, apply it in practice, and exercise judgment under uncertainty.

WEKID™ provides this shift. By modeling AI output quality through a hierarchical epistemic stack — Wisdom, Experience, Knowledge, Information, and Data — WEKID™ aligns AI evaluation with the way humans' reason, decide, and take responsibility for outcomes. It transforms abstract principles such as trust, safety, and accountability into measurable, enforceable controls that can be applied consistently across models, vendors, and system architectures.

Unit 10.1 — WEKID™ as a Standard Framework

The strength of WEKID™ lies not only in its conceptual clarity, but in its practical applicability as a standard framework. Because WEKID™ evaluates outputs and decisions rather than internal model mechanics, it is inherently:

- Model-agnostic, supporting rapid technological change
- Vendor-neutral, enabling multi-provider strategies
- Extensible, accommodating new tools, agents, and autonomy levels
- Auditable, producing artifacts suitable for governance and compliance

These properties make WEKID™ well suited to serve as a common evaluation and governance standard across enterprises, industries, and regulatory environments. By providing a shared epistemic language, WEKID™ enables consistent expectations of AI behavior regardless of implementation details.

Unit 10.2 — Reinforcing Human–Machine Partnership

Critically, WEKID™ does not seek to replace human judgment but rather formalizes where and how human intervention matters most. The hierarchical nature of WEKID™ explicitly identifies the points at which automated systems are likely to fail and where human oversight is essential:

- Humans validate Data integrity when sources are uncertain
- Humans review Knowledge and Experience when operational complexity increases
- Humans exercise Wisdom when stakes are high, values conflict, or consequences are irreversible

Through scoring, gating, and decision matrices, WEKID™ enables calibrated human–machine interaction rather than binary automation. AI systems are permitted to act autonomously only when they demonstrate sufficient epistemic competence; otherwise, responsibility is appropriately escalated to human reviewers.

This approach reinforces the principle that accountability remains human, even as capability becomes machine augmented.

Unit 10.3 — From Automation to Responsible Autonomy

WEKID™ provides a practical path from narrow automation to responsible autonomy. By constraining autonomy based on epistemic performance, organizations can:

- Expand AI autonomy safely over time
- Detect degradation before failures occur
- Prevent overconfidence and misuse
- Maintain trust with users, regulators, and the public

Rather than asking whether AI can act independently, WEKID™ ensures that it acts independently only when it should.

Unit 10.4 — A Call to Implementation

The risks posed by increasingly capable AI systems are no longer hypothetical. Real-world failures have demonstrated that fluent, confident systems can mislead, harm, or erode trust when epistemic integrity is not enforced. Addressing these risks requires more than ethical principles or post-hoc review — it requires operational governance embedded into AI evaluation and deployment.

WEKID™ offers such a foundation. By grounding AI evaluation in epistemic hierarchy, measurable signals, and enforceable controls, WEKID™ enables organizations to build AI systems that are not only powerful, but worthy of trust.

By distinguishing data from information, knowledge from experience, and competence from wisdom, WEKID™ provides a durable framework for evaluating and governing AI in enterprise and societal contexts. It enables trust by design, reinforces human responsibility, and establishes a clear standard for safe, transparent, and accountable AI systems.

In an era where AI increasingly shapes human outcomes, how we evaluate AI matters as much as what AI can do. WEKID™ ensures that evaluation keeps pace with capability — and that human judgment remains at the center of intelligent systems.

Module 10 Sample Knowledge Check

The item below is a sample knowledge-check question illustrating exam format and difficulty. The answer key appears in Appendix B.
Which statement best captures WEKID™'s implementation stance?
A. Governance should be imposed only after a public incident.
B. Governance should be embedded into evaluation and deployment workflows as measurable, enforceable controls.
C. Trust should be assumed if the model is advanced.
D. WEKID™ is only useful for low-impact chatbots.

Module 10 — Sample Level 2 Scenario

The scenario below is a sample illustrating Level 2 short-answer format and expected reasoning. Operational scenarios are confidential and do not appear in this handbook. Required elements and a sample response appear in Appendix C.

SA-M10-S1 *— Level 2 • Non-Technical*

Scenario: You must write a policy for when a system can move from “human-in-the-loop” to “autonomous.”

Prompt: Using WEKID™, define the minimum gate conditions (by layer), the decision rights, and the evidence required to approve autonomy expansion.

PART III

Appendices

Failure modes, sample answer keys, sample rubrics, and glossary

Appendix A: AI Failure Modes Using Real-Life Cases

This appendix presents AI failures before they cause harm, regulatory exposure, or loss of trust. Publicly documented AI incidents — ranging from hallucinated legal citations to runaway autonomous agents — can be systematically categorized into a small number of recurring failure types.

Sample Public AI Failure Types and WEKID™ Mitigations

AI Failure Type	Primary WEKID™ Layer Failure	Observed Real-World Manifestation	Resulting Impact
Hallucinated facts and fabricated sources	Data	AI-generated legal briefs citing non-existent court cases; fabricated medical references	Legal sanctions, reputational damage, unsafe decisions
Fluent but incorrect reasoning	Knowledge	Confident explanations built on incorrect assumptions or misinterpreted regulations	Misleading guidance, regulatory violations
Correct facts applied without judgment	Wisdom	AI provides accurate medical or legal info but fails to warn about uncertainty or risk	Harm to users, liability exposure
Premature autonomy escalation	Experience & Wisdom	AI agents are allowed to act independently before demonstrating operational competence	Runaway automation, system outages
Runaway tool-using agents	Experience	Autonomous agents repeatedly invoke tools, escalating cost, or damage	Financial loss, security exposure
Invisible executive risk	Wisdom (Governance)	Boards unaware of declining AI quality until after public failure	Strategic surprise, trust erosion
Inability to prove compliance	Information & Governance	Organizations unable to show how AI meets NIST/EO/EU AI Act requirements	Audit failure, fines, program shutdowns
Inconsistent human review	Experience & Wisdom	Ad-hoc or uneven human oversight of high-stakes AI outputs	Bias, inequity, legal exposure

AI Failure Type	Primary WEKID™ Layer Failure	Observed Real-World Manifestation	Resulting Impact
Post-incident uncertainty about cause	Cross-layer epistemic failure	Organizations cannot explain why AI failed, only that it failed	Repeat incidents, weak remediation
Untrained users and reviewers	All layers (systemic)	Staff over-trust AI or misuse outputs due to lack of evaluation discipline	Widespread misuse, cultural risk

Summary of Real-Life AI Failure Cases (Publicly Documented)

1. Fabricated Legal Citations in Court Filings

Failure pattern: Hallucinated facts treated as authoritative truth

What happened: Attorneys submitted court briefs containing non-existent case law generated by an AI system. The citations appeared well-formatted and plausible but were entirely fabricated.

Impact: Judicial sanctions, professional discipline, public embarrassment, and erosion of trust in legal AI tools.

WEKID™ insight: A Data-layer gate would have blocked fabricated sources before submission; Evaluation-as-a-Service would have required human verification for legal filings.

2. AI-Generated Medical Advice Without Risk Qualification

Failure pattern: Correct information applied without judgment

What happened: AI systems provided symptom-based medical guidance that was factually accurate but failed to emphasize uncertainty, recommend professional care, or recognize high-risk scenarios.

Impact: Potential patient harm, regulatory scrutiny, and liability exposure for providers.

WEKID™ insight: A Wisdom gate would have forced uncertainty disclosure and escalation to human clinicians.

3. Automated Hiring and Screening Bias

Failure pattern: Misinterpreted data embedded into decision logic

What happened: AI screening tools learned biased patterns from historical hiring data, systematically disadvantaging certain groups despite appearing “objective.”

Impact: Regulatory investigations, reputational damage, withdrawal of AI tools.

WEKID™ insight: Knowledge-layer evaluation would have detected improper generalization; Risk and Autonomy Assessment would have limited decision authority.

4. Autonomous Trading and Market Manipulation Events

Failure pattern: Runaway autonomy and feedback loops

What happened: Algorithmic trading systems amplified volatility through rapid, automated decision-making without sufficient human oversight.

Impact: Market disruption, financial losses, and regulatory intervention.

WEKID™ insight: Runtime Guardrails for Agentic AI would have enforced escalation thresholds and authority caps.

5. Government Benefits and Eligibility Determination Errors

Failure pattern: Incorrect application of rules at scale

What happened: Automated systems incorrectly denied or terminated public benefits based on flawed assumptions or incomplete data.

Impact: Legal challenges, public backlash, reversal of automated decisions.

WEKID™ insight: Experience-layer controls would have required verification steps and exception handling before action.

6. Predictive Policing and Risk Scoring Controversies

Failure pattern: Overconfidence in probabilistic models

What happened: AI systems labeled individuals or communities as “high risk” without sufficient transparency, appeal mechanisms, or contextual judgment.

Impact: Civil rights concerns, bans on system use, loss of public trust.

WEKID™ insight: Wisdom-layer enforcement would have constrained use in high-stakes civil contexts.

7. AI-Assisted Content Moderation Errors

Failure pattern: Poor contextual interpretation

What happened: Automated moderation systems removed lawful speech or failed to act on harmful content due to lack of contextual understanding.

Impact: Public criticism, platform policy reversals, regulatory attention.

WEKID™ insight: Information-layer evaluation would have flagged context loss and ambiguity.

8. Hallucinated Financial Analysis and Business Intelligence

Failure pattern: Fluent synthesis built on incorrect premises

What happened: AI-generated financial summaries included invented metrics, outdated data, or misinterpreted trends presented as confident analysis.

Impact: Faulty business decisions, internal distrust of AI analytics.

WEKID™ insight: Data and Knowledge gates would have prevented decision use without verification.

9. Autonomous IT and Infrastructure Failures (AIOps)

Failure pattern: Operational incompetence masked by automation

What happened: AI systems automatically restarted services, scaled resources, or modified configurations incorrectly, worsening outages.

Impact: Extended downtime, cost overruns, rollback to manual control.

WEKID™ insight: Experience-layer scoring would have limited autonomy until procedural competence was demonstrated.

10. Inability to Explain AI Decisions After Incidents

Failure pattern: No epistemic audit trail

What happened: After public or regulatory incidents, organizations could not explain why an AI system made a particular decision — only that it did.

Impact: Failed audits, legal exposure, forced system shutdowns.

WEKID™ insight: AI Incident Prevention and Forensics would have provided layer-by-layer root cause analysis.

11. Executive Surprise and Board-Level Blindness

Failure pattern: AI risk invisible above engineering

What happened: Boards and senior executives learned about AI failures only after public disclosure, having received no prior risk indicators.

Impact: Loss of confidence, governance failures, leadership intervention.

WEKID™ insight: Board and Executive Oversight Dashboard would have surfaced declining epistemic quality early.

12. Workforce Over-Trust and Skill Atrophy

Failure pattern: Cultural misuse of AI

What happened: Staff relied on AI outputs without validation due to lack of training, leading to cascading errors across functions.

Impact: Systemic misuse, quality degradation, institutional risk.

WEKID™ insight: Certification and Training Program establishes disciplined evaluation norms.

Appendix B: Sample Knowledge Check Answer Keys

Answer key for the per-module sample knowledge-check questions in Part II.
Module 1 · KC-M1-Q1 · Answer: B
Module 2 · KC-M2-Q1 · Answer: D
Module 3 · KC-M3-Q1 · Answer: C
Module 4 · KC-M4-Q1 · Answer: D
Module 5 · KC-M5-Q1 · Answer: B
Module 6 · KC-M6-Q1 · Answer: B
Module 7 · KC-M7-Q1 · Answer: B
Module 8 · KC-M8-Q1 · Answer: B
Module 9 · KC-M9-Q1 · Answer: B
Module 10 · KC-M10-Q1 · Answer: B

Appendix C: Sample Scenario Rubrics and Sample Responses

This appendix contains the scoring rubric, required-element keys, and sample responses for every sample scenario published in Part II. Operational scenarios and their rubrics are confidential and are not included in this handbook. Sample responses illustrate strong answers; they are not the only acceptable wording. Scorers calibrate against required elements rather than literal text. Two rubrics apply: a modified rubric for Level 1 (Practitioner) candidates that emphasizes correct identification of the failing layer and a defensible governance action; and the standard rubric for Level 2 (Authority) candidates that additionally expect evidence, decision rights, and escalation design.

Level 1 Modified Scenario Rubric (Practitioner)

Level 1 Technical candidates take a shortened scenario component (two scenarios) drawn from the same item bank as Level 2 but assessed against a simplified rubric; the Level 1 Non-Technical exam has no scenario component. The Level 1 rubric requires fewer elements per response and uses plainer language, so a candidate who clearly identifies the failing layer and a defensible governance action can pass even without naming specific audit artifacts or escalation owners.

Score	Level 1 Descriptor
5	Identifies the failing layer correctly, names a defensible governance action, and adds a brief, plain-language rationale grounded in WEKID™ principles
4	Identifies the failing layer correctly and names a defensible governance action; rationale is mostly correct with minor gaps
3	Identifies a plausible failing layer and a reasonable governance action, even if rationale is thin (Level 1 passing threshold)

Score	Level 1 Descriptor
2	Identifies either the failing layer OR a governance action correctly, but not both; or names both with a rationale that contradicts WEKID™ principles
1	Substantially incomplete or only loosely related to the prompt
0	No meaningful response, off-topic, or in conflict with WEKID™ principles

Level 1 Technical candidates pass the scenario component when their two scenarios average ≥ 2.5 with no individual scenario below 2 (see Section 5.4). The Level 1 modified rubric uses a subset of the required elements listed below for each operational scenario; specifically, candidates are scored only on the failing-layer identification and the recommended governance action, with credit also given for clear plain-language rationale.

Standard 0–5 Scoring Rubric (Level 2)

Score	Descriptor
5	Comprehensive, well-justified response that addresses all required elements with clear WEKID™ alignment and includes appropriate audit, escalation, or evidence considerations
4	Strong response that addresses all required elements with minor gaps in justification or completeness
3	Acceptable response that addresses the required core elements, passing threshold per scenario
2	Partial response missing one or more required elements or showing weak WEKID™ alignment
1	Substantially incomplete or misaligned with WEKID™; addresses only a small portion of the prompt
0	No meaningful response, off-topic, or in conflict with WEKID™ principles (e.g., recommends ignoring hard gates)

To receive a passing score (3 or above), a response must address the core required elements. To receive a 4 or 5, it should address all required elements with clear rationale and appropriate audit/escalation/evidence considerations.

Required Elements and Sample Responses

SA-M1-S1

Required elements (passing): Recommends Constrain or Reject (not Approve) under WEKID™; cites the Fabrication / Data Integrity Gate; states what evidence (verified citation, source, or tool trace) must be added before approval; identifies the audit artifact (scorecard + gate trigger + corrected source).

Sample response: *Decision: Constrain and require remediation before approval. Although the summary is fluent, a fabricated regulatory citation triggers the Fabrication / Data Integrity Gate; under non-substitutability, polish cannot offset a foundational integrity failure. Required evidence before approval: replace the fabricated citation with a verified authoritative source (with version/date), and re-run scoring to confirm the gate no longer triggers. Audit artifact: WEKID™ scorecard, gate trigger reason, corrected source link, and approver identity for the override of the original output.*

SA-M2-S1

Required elements (passing): Identifies Information completeness gap (omitted scope limits); recommends review checklist and escalation triggers for regulated summaries.

Sample response: *Failing layer is Information: factually correct facts were summarized clearly, but material scope limits were omitted, materially changing the obligations. Controls: introduce a regulatory-summary checklist that requires explicit statement of scope, applicability, exceptions, and effective dates; add an escalation trigger when a summary affects regulated workflows or external counterparts; require subject-matter review for summaries used in compliance decisions; retain the checklist results in the evaluation artifact.*

SA-M3-S1

Required elements (passing): Minimum gate conditions for autonomy expansion; decision rights for approval/monitoring/rollback.

Sample response: *Minimum gate conditions: stable Data integrity (no fabrication or traceability gate triggers above defined thresholds over the evaluation window), Wisdom ≥ threshold for high-stakes interactions, Experience ≥ threshold for any operational actions, and complete audit evidence per decision. Decision rights: Product approves customer-facing autonomy expansion; Risk/Compliance approves any move involving regulated content or policy gates; Operations approves rollback authority and retains the kill switch; Governance reviews quarterly. Any breach of the gate conditions automatically reverts autonomy to the previous tier.*

SA-M4-S1

Required elements (passing): Separates tool-fidelity vs reasoning controls; minimum audit evidence (tool traces, retrieved sources, scores, gates, actions).

Sample response: *Treat the debate as two control domains. Tool-fidelity (Data) controls: complete tool traces with request/response payload IDs, immutable binding from each agent statement to the tool result it cites, and blocks on references to tool results not present in the trace. Reasoning controls (Knowledge/Experience/Wisdom): claim-to-evidence binding for each material recommendation, risk-tiered gates for high-impact actions (Data + Wisdom thresholds), Experience checks (verification, rollback, safe sequencing), and Wisdom behaviors (uncertainty disclosure, escalation). Minimum audit evidence: full tool transcript and outputs, retrieved sources with versions, claim-to-evidence mapping,*

layer scores with rationale, gate triggers and action codes, approver identities, and configuration/version metadata.

SA-M5-S1

Required elements (passing): Defensible Data gate policy; exception process; testing and monitoring plan.

Sample response: *Treat fabricated citations in regulated reporting as a hard Data Integrity violation; do not relax the gate. Refine detection to distinguish missing citation vs failed verification vs not required but require verifiable citation for material claims. Exception process: approval by Compliance/Legal with audit sign-off; documented evidence; time-bound and scoped; recurring needs require formal change control. Testing: a golden set of regulated-reporting scenarios run pre/post changes; confirm no regression on fabricated-citation detection. Monitoring: track hard-gate rate by report type, exception rate, and verification-failure reasons; alert on drift.*

SA-M6-S1

Required elements (passing): Repeatability via signals + rationale; stability metrics; reporting approach.

Sample response: *WEKID™ reduces subjectivity by tying scores to observable signals, anchored definitions, stored rationales, and calibration sets; reviewers look for the same evidence and explain it. Stability metrics: inter-rater reliability (% within ±1 per layer), gate-decision agreement, score variance for fixed calibration items over time, hard-gate rate trends by use case/tier, and "same input, same score" repeat tests. Audit-readiness: % of evaluations with complete evidence links and rationale fields. Report monthly to leadership to demonstrate stability and surface drift.*

SA-M7-S1

Required elements (passing): Plain-language explanation: failing layer + gate + action; remediation; evidence required for approval.

Sample response: *The output was blocked because it failed to follow a mandatory safety rule even though the overall score looked acceptable. Under WEKID™, foundational integrity issues cannot be offset by clarity or completeness elsewhere. The Data Integrity / Traceability Gate triggered because at least one material claim could not be verified against an approved source. To approve, replace or remove unverified claims, provide authoritative citations, or tool evidence (with version/date), restate constraints clearly, and re-run scoring. For high-risk use cases, approval may remain constrained until a human reviewer confirms the evidence. Retain the original output, the WEKID™ scorecard, the gate trigger, and the corrected sources.*

SA-M8-S1

Required elements (passing): Value of diagnostics for remediation; dual-layer reporting; preserve gate transparency.

Sample response: *Diagnostics turn WEKID™ from a score into governance — they explain why an output was approved or blocked, which layer failed, and what to fix. Without them, you lose auditability, repeatability, and the ability to run the remediation loop, and dashboards may show "green" while a hard-gate driver rises silently. Use a two-layer reporting approach: an executive summary view (Trust Status, action, critical gate indicator, top failing layer, hard-gate trend) and a drill-down evidence view (layer scorecard with rationales, signals that failed, evidence links, gate decision record, single-highest-leverage improvement, re-score results). Diagnostics are preserved — they simply move to drill-down and remain available for auditors, incident review, and remediation.*

SA-M9-S1

Required elements (passing): Board-ready format with KPIs + gate events; thresholds; preserves critical failures; escalation triggers.

Sample response: *Trust Status (Red/Amber/Green) by risk tier, where status is gate-driven so a quarter cannot be Green if hard gates rose or clustered in high-stakes areas. KPIs: hard-gate rate (overall and high-stakes), top three gate types, layer trends (median + 10th percentile) with the lowest-layer view to prevent average washout, exception/override rate, and time-to-remediate. Notable gate events list (workflow, gate, failing layer, action taken, remediation status, residual risk, owner, due date) and autonomy-posture changes. Escalation triggers: any red high-stakes workflow, hard-gate rate above the red threshold or rising sharply QoQ, repeat incidents tied to the same gate type, sustained Wisdom/Experience decline, or exception rate above threshold.*

SA-M10-S1

Required elements (passing): Policy for autonomy expansion: minimum gates by layer; decision rights; evidence; monitoring; rollback.

Sample response: *Minimum gates by layer. Data (hard): no fabrication, mis-citation, or misrepresented tool outputs; high-stakes claims must be traceable to approved sources or tool traces with version/date; missing evidence fails closed. Information: clear scope, constraints, and applicability; separation of facts vs interpretation; explicit out-of-scope language. Knowledge: correct domain reasoning, no contradictions or unjustified generalization, conclusions aligned to cited evidence, assumptions flagged. Experience: verification checkpoints, dependency awareness, rollback/recovery, safe defaults, no brittle sequencing. Wisdom (high-stakes threshold): calibrated risk behavior, uncertainty disclosure, tradeoff awareness, and stop/ask/escalate; require Wisdom ≥ threshold for high-stakes workflows. Decision rights: joint approval by Business/Product (value/customer impact), Operations (safety/rollback), and Risk/Compliance (policy/regulated claims); Security signs off when tool access or data classification is in scope; Operations retains kill-switch/rollback authority. Required evidence: a time-bounded evaluation window with stable/improving layer scores and low hard-gate rates by workflow; calibration set demonstrating scoring consistency; documented policy-as-code with regression tests; monitoring dashboards with thresholds; a tested kill switch; and audit-ready per-decision artifacts.*

Appendix D: Glossary

This glossary defines the WEKID™ framework and certification program terms used throughout this handbook. Definitions are intentionally concise and are aligned with the full descriptions provided in Parts I and II.

Approve. Governance action permitting an AI output to be used (autonomously or with monitoring) when the overall score meets the threshold and no hard gates are triggered.

Authority (WEKID™ Authority Certified). Level 2 credential confirming applied competence in WEKID™ governance design, scoring, and operationalization. Awarded on either the Technical or Non-Technical Path.

Authority Map. A documented assignment of decision rights for AI behavior — who approves, who reviews, who can override, and who owns escalation — typically tiered by risk.

Calibration Set. A versioned set of representative AI outputs with expected layer scores and gate outcomes used to align human and automated scoring over time.

CEU (Continuing Education Credit). A unit of credit earned through approved learning or applied governance activity. Used to recertify a WEKID™ credential during its renewal window.

Code of Conduct. The standard of professional behavior agreed to by every certified holder, covering integrity, exam confidentiality, accurate credential use, conflict disclosure, and respectful conduct.

Constrain (Human-in-the-Loop). Governance action requiring human review before an AI output can be used or executed; typically triggered by sub-threshold layer scores or risk-tier rules.

Credential. A specific WEKID™ certification (Practitioner or Authority on the Technical or Non-Technical Path).

Data (D). Lowest WEKID™ layer; atomic factual elements (dates, quantities, entities, calculations, citations, tool outputs). Foundation for all higher layers.

Decision Matrix. The mapping from layer scores and gate outcomes to governance actions (approve, constrain, remediate, reject).

Domain Weighting. The relative importance assigned to each WEKID™ layer in the weighted aggregation formula. Tunable by domain and risk.

Epistemic Hierarchy. The principle that WEKID™ layers depend on the integrity of those beneath them; higher-layer quality cannot repair lower-layer failures.

Epistemic Quality. The quality of how knowledge is formed, justified, interpreted, and applied — the property WEKID™ evaluates.

Escalation. Routing a decision to a defined authority when a gate triggers, a layer score falls below threshold, or a risk-tier rule applies.

Evaluation Artifact. The structured record of a single AI output evaluation, including scores, gate results, action code, rationale, and evidence links.

Evidence Bundle. Citations, retrieved passages, tool traces, and verification notes referenced by an evaluation.

Exam Blueprint. The published specification for a WEKID™ exam, including level, path, recommended experience, item counts, time limits, and domain weights.

Experience (E). WEKID™ layer covering procedural and operational competence; sequencing, edge cases, verification, and recovery.

Fabrication Gate. A baseline hard gate that caps total score when fabricated citations, sources, or tool results are detected.

Hard Gate. A non-negotiable rule that blocks or caps overall scores when a critical failure is detected (e.g., Data Integrity, Fabrication, High-Stakes Wisdom).

Information (I). WEKID™ layer concerned with the relevance, completeness, framing, and clarity of communicated content.

Knowledge (K). WEKID™ layer covering synthesis, causal reasoning, and bounded generalization — understanding rather than mere recall.

Knowledge Check (KC). A multiple-choice item used to assess recognition of WEKID™ concepts; identified by IDs in the form KC-M{module}-Q{number}.

Level 1 (Practitioner). Foundational WEKID™ credential focused on fundamentals, recognition, and baseline application.

Level 2 (Authority). Advanced WEKID™ credential focused on applied governance design, scoring, and implementation.

Non-Substitutability. The rule that strength in one WEKID™ layer cannot offset failure in a prerequisite (lower) layer.

Non-Technical Path. Certification path emphasizing governance design, decision rights, escalation logic, and oversight artifacts.

Path. The track on which a candidate is assessed. Two paths exist: Technical and Non-Technical.

Policy-as-Code. An implementation pattern in which gating rules and thresholds are expressed as versioned, testable code rather than informal documents.

Practitioner (WEKID™ Practitioner Certified). Level 1 credential confirming foundational understanding of the WEKID™ framework. Awarded on either the Technical or Non-Technical Path.

Recertification. The process for renewing a credential before its 24-month expiration, via CEUs, refresher assessment, re-examination, or upgrade.

Reject. Governance action blocking the use of an AI output, typically due to a hard gate trigger such as fabrication or Data integrity failure.

Remediate. Governance action requiring targeted fixes to specific layer failures, followed by re-scoring before approval.

Risk Tier. A classification of a workflow or decision (Low/Medium/High) used to apply different thresholds, gates, and approval requirements.

Scenario Short-Answer (SA). A Level 2 assessment item presenting a real-world situation and asking the candidate to apply WEKID™ reasoning. Identified by IDs in the form SA-M{module}-S{number}.

Score (Layer). The 0–5 evaluation of a single WEKID™ layer.

Score (Overall / Weighted). The aggregated score derived from layer scores using the weighted aggregation formula, after applicable hard gates have been applied.

Signal. An observable feature of an AI output used to score a WEKID™ layer (e.g., presence of uncertainty language, complete tool trace, contradiction detected).

Soft Gate. A rule that caps or constrains an outcome but does not necessarily reject it (e.g., Wisdom < threshold in moderate-risk contexts).

Technical Path. Certification path emphasizing implementation, evaluation pipelines, control alignment, and operational guardrails.

Tool Fidelity. The accuracy with which an AI output represents the actual outputs of any tools it invoked. A subset of Data layer integrity.

Trust by Design. An approach in which trustworthy AI behavior is engineered through measurable signals, scoring, gates, and continuous monitoring rather than assumed.

WEKID™. The framework comprising five hierarchical epistemic layers (Wisdom, Experience, Knowledge, Information, Data) used to evaluate and govern AI outputs and decisions.

WEKID™ Score. The numeric aggregate score produced by the WEKID™ evaluation, reflecting both weighted layer scores and any hard-gate caps applied.

Wisdom (W). Highest WEKID™ layer; judgment, risk awareness, and value alignment under uncertainty.

End of Document

WEKID™ Certification Handbook • Version 1.2.0 • Editorial & Structural Revision

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.