WEKID™ Master Whitepaper – **A Hierarchical Epistemic Framework and Operational Model for Evaluating and Governing Enterprise Artificial Intelligence Outputs**

SUMMARY Version

By James M. Judge and Members of the WEKID Community

Version 1.3 (Community RFC Integration: Overseer-Engagement Extensions to Model Two; the Experience Layer) August 2026

Version Control Log

WEKID™ Master Whitepaper — A Hierarchical Epistemic Framework and Operational Model for Evaluating and Governing Enterprise AI Outputs

VERSION	DATE	AUTHOR	SUMMARY OF CHANGES
1.0	December 2025	James Madigan Judge	Initial release. Hierarchical epistemic framework (Wisdom, Experience, Knowledge, Information, Data); scoring, gating, and decision matrices for governing enterprise AI outputs.
1.1	March 2026	WEKID Community	Clarifications and minor enhancements. Refined layer definitions, risk, control, and metric mapping, and worked scoring examples.
1.2	July 2026	WEKID Community	Two-Model Framework Integration. Introduced Model One (Epistemic Maturity) and Model Two (AI Decision Authority) connected by the Trust Bridge; added the five Authority Levels and the maturity-to-authority decision matrix. Editorial corrections. Corrected Figure 1 to match the normative Authority Level definitions (§4.1); renumbered levels AL-0 through AL-4; recomputed the four worked examples; added decision-matrix precedence and the sub-50 maturity band; resolved figure placeholders
1.3	August 2026	WEKID Standards Council	. Community RFC integration, ratified by the WEKID Standards Council. WK-CORE-RFC-001 (Governing the Human in the Loop): overseer engagement as a governed signal, joint gating of AL-3 on maturity and engagement, an explicit calibration objective, a decision-maker alignment factor in the stakes floor, and the non-adversarial-principal scope boundary (§§ 4.4, 4.5, 4.7 through 4.10, 5.6, 6.4, 8.9). WK-CORE-RFC-002 (The Experience Layer): the Experience layer formalized as the pivot between Knowledge and Wisdom, restoring Ackoff's Understanding rung and defining Expertise and Wisdom as functions of validated, domain-indexed Experience (§§ 2.2 through 2.2.3). Both integrations are additive and backward compatible with Version 1.2.1.

Versioning. Major versions (1.0 → 2.0) denote substantive framework changes. Minor versions (1.1, 1.2, 1.3) denote additions or integrations. Patch versions (1.2.1) denote editorial corrections that do not change the framework.

Current version. This document is Version 1.3 (August 2026), the canonical statement of the WEKID™ framework; all companion documents cite it by version. This work is registered with the U.S. Copyright Office.

Contents

Executive Summary

Artificial intelligence systems have 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. Publicly documented failures, from fabricated legal citations to erroneous benefits determinations and runaway autonomous agents, share a common pattern: apparent expertise silently became delegated decision authority without governance. Capability is not authority.

Closing this gap requires answering two questions that most AI programs conflate: how mature knowledge is, and how much decision authority should follow? At the core of the first question is **epistemic quality**, that is, the quality of how knowledge is formed, justified, interpreted, and applied. At the core of the second is a governance discipline that assigns decision authority explicitly, rather than allowing it to accumulate through product defaults, vendor claims, or organizational habit.

WEKID™ was first introduced as an intelligence hierarchy in 2007, well before the pervasive use of AI. With the rapid adoption of AI technology, WEKID has matured into a governance framework composed of two complementary models, connected by a common architecture referred to throughout this paper as The WEKID Stack (Figure 1).

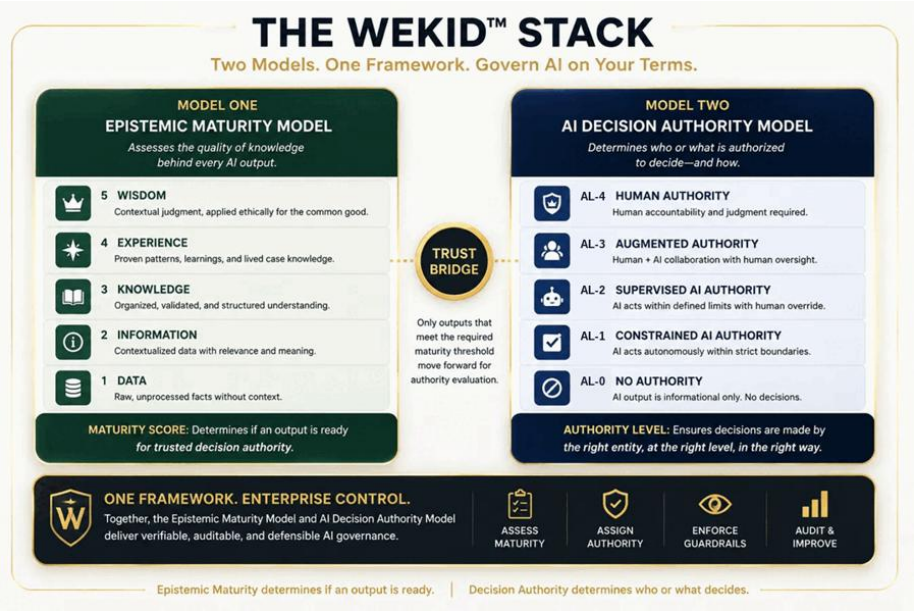

Figure 1 illustrates the WEKID Stack and two complimentary models connected via a trust bridge epistemic hierarchy and its dependency structure.

The **Epistemic Maturity Model** assesses the quality of the knowledge behind every AI output across five hierarchical layers: **Wisdom, Experience, Knowledge, Information, and Data.** Mirroring how humans progress from raw facts to understanding, application, and judgment. Layer-level evaluation produces a **Maturity Score** that determines

whether an output is epistemically ready to be trusted, while hierarchical gating prevents fluent or confident outputs from masking foundational errors.

The **AI Decision Authority Model** governs **who or what is authorized to decide and how** across five levels of delegation: from **No Authority** (informational output only) through **Constrained**, **Supervised**, and **Augmented Authority** to **Human Authority**, in which the highest-stakes decisions remain reserved to human judgment regardless of how mature the AI's contribution may be. The Trust Bridge connects the two models: only outputs that meet the required maturity threshold advance to authority evaluation. Epistemic maturity determines *if* an output is ready; decision authority determines *who or what decides*.

Unlike model-centric evaluation approaches, WEKID focuses on AI outputs 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.

WEKID transforms high-level AI ethics and responsible-AI principles into measurable, enforceable governance controls. It operates as a continuous cycle: assess maturity, assign authority, enforce guardrails, audit, and improve. It enables objective scoring, policy-based thresholds, audit-ready artifacts, and actionable diagnostics. Together, the two models deliver verifiable, auditable, and defensible AI governance before authority is delegated, not after failures occur.

This Version 1.3 extends the WEKID framework foundation with two community contributions ratified by the WEKID Standards Council. Overseer engagement is added as a governed signal within the AI Decision Authority Model, so that a grant of Supervised or Augmented Authority remains valid only while the human oversight it assumes is demonstrably exercised (Sections 4.7 through 4.10). The Experience layer of the Epistemic Maturity Model is formalized as the pivot between Knowledge and Wisdom, restoring a rung Ackoff's original hierarchy included and popular renderings dropped, so that Expertise and Wisdom are defined as functions of validated, domain-indexed Experience (Section 2.2).

1. Introduction

Large language models and AI agents increasingly generate outputs that directly influence **business decisions, policy interpretation, software development, healthcare guidance, legal reasoning, and financial analysis**. These systems are 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 creates 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.

1.1 Limitations of Existing Evaluation Paradigms

Current approaches to AI evaluation are ill-suited to this reality. Most existing paradigms suffer from five 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 and the use of enforcement tools 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.

- **Absence of any mechanism for governing decision authority**
  Even where evaluation detects quality problems, existing paradigms provide no principled answer to the question that follows: given this level of demonstrated quality, what is this system authorized to decide — and under whose oversight? In practice, autonomy is granted by product configuration or organizational habit rather than by demonstrated epistemic maturity. This is precisely how apparent expertise silently becomes delegated decision authority.
- **Reliance on runtime enforcement tools that constrain actions without assessing judgment**
  A growing class of agentic enforcement tools control planes: MCP gateways, and security or runtime guardrails. These tools govern what an AI system is *permitted* to do at execution time. These mechanisms enforce policies, scopes, and permissions: which tools an agent may invoke, which data it may access, and which actions require human approval. However, they operate on predefined rules and identity or permission boundaries rather than on the quality of the system's reasoning. A guardrail can block an unauthorized tool call, but it cannot determine whether a *permitted* action reflects sound interpretation, appropriate confidence, or correct procedural reasoning. As a result, an agent can remain fully compliant with every runtime policy while still exercising poor judgment within its authorized envelope. These tools constrain the boundaries of action but provide no assessment of epistemic maturity within those boundaries.

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

1.2 The Shift from Model Performance to Epistemic Quality

As AI systems transform from passive assistants to **decision-support and decision-making agents**, 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?

Once those questions are answered, a further question remains: **who or what should be authorized to act on this output, at what level of oversight, and with what accountability?**

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.

1.3 What is WEKID

WEKID was developed as a response to this need for epistemic evaluation and appropriate decision authority. It is a governance framework composed of two complementary models. Model One, the Epistemic Maturity Model, provides a structured, hierarchical assessment of AI outputs across five epistemic maturity layers: **Wisdom, Experience, Knowledge, Information and Data**. It produces a Maturity Score that determines whether an output is epistemically ready to be trusted. Model Two, the AI Decision Authority Model, governs who or what is authorized to decide and how across five levels of delegation, ensuring that decisions are made by the right entity, at the right level, in the right way. The Trust Bridge connects the two models: only outputs that meet the required maturity threshold advance to authority evaluation. Together they form The WEKID Stack (Figure 1).

WEKID makes epistemic quality and decision authority explicit by structuring them as hierarchical and measurable constructs. This enables organizations to identify not just that an AI output failed, but where and why it failed, and, more importantly, to prevent epistemically immature outputs from acquiring decision authority in the first place. This capability is essential for governing AI systems operating in complex, high-impact environments.

WEKID provides the foundation for moving from trust based on fluency or reputation toward **trust grounded in epistemic integrity.**

Sections 2–3 develop the Epistemic Maturity Model and the hierarchy on which it rests. Section 4 introduces the AI Decision Authority Model and the Trust Bridge. Sections 5–9 operationalize both models across tool-using systems, enterprise governance, measurable signals, scoring, gating, and authority assignment. Sections 10–11 examine the implications for trustworthy AI and the path to implementation.