ARC Framework - Scholarly Derivation - Applying the ARC Framework (Accurate, Reliable, Contextual) for Human-AI Contextualization: A Model for Assuring AI Integrity – Edition V1.0

Aligned to APA 7th Edition Standards and ARC Tier-1 & Tier-2 Validation

> *"Context is not decoration—it is structure."*
> *Meaning emerges only where relevance meets responsibility.*
> **— ARC Framework™ Principle**

Abstract

As artificial intelligence (AI) systems increasingly mediate knowledge work, decision-making, and strategic operations, the integrity of human-AI contextualization—how humans frame, interpret, and operationalize AI outputs—has become a pressing ethical and epistemic concern. This paper applies the ARC Framework—Accurate, Reliable, Contextual—as a formal mechanism to structure, audit, and enhance the integrity of AI outputs. Grounded in peer-reviewed research and institutional standards, this derivation validates the ARC Framework's ability to mitigate AI hallucinations, bias propagation, and context drift. By embedding Tier-1 rigor through role designation, source citation, and verifiability, and advancing to Tier-2 for judgment and commitment, the ARC Framework elevates AI from a tool of convenience to a partner in trustworthy intelligence. Recent industry discourse also affirms ARC's relevance, as seen in the Forbes Technology Council's position that "contextualization is the key to unlocking generative AI's potential."

Introduction

The rapid integration of AI systems across domains such as education, governance, climate modeling, and healthcare has introduced urgent questions about the trustworthiness of machine-generated insights (Floridi & Cowls, 2022). Many generative AI systems, including large language models, suffer from issues such as hallucinated facts, unreliable logic, and context misalignment (Raji et al., 2020). These risks are not technical glitches—they are epistemological challenges. To address them, the ARC Framework offers a structured, human-centered protocol designed to assure AI output quality across three core dimensions: Accuracy, Reliability, and Contextuality (Woods, 2025).

Recent industry discourse supports ARC's focus. As the Forbes Technology Council (2023) asserts, "contextualization is the key to unlocking generative AI's potential," reinforcing ARC's foundational claim that relevance and interpretation must be driven by human framing.

Clarifying ARC's Core Structure: Pillars vs. Tiers

The ARC Framework is built upon three epistemic pillars: **Accuracy**, **Reliability**, and **Contextuality**. These are not tied to a specific tier or stage; rather, they form the evaluative foundation used to assess AI outputs at all levels. Think of them as the "why" behind ARC—why we audit AI outputs, why we structure human-AI interactions, and why integrity matters.

To apply these pillars in practice, ARC introduces two tiers:

- **Tier-1: Validation & Rigor** – Ensuring AI responses are verifiable, sourced, and role-appropriate using structured protocols.
- **Tier-2: Decision & Refinement** – Guiding users through analysis, re-framing, and context-aligned commitment.

Contextualization in the Scholarly Landscape

ARC's emphasis on contextualization aligns with a growing body of scholarly work across human-computer interaction (HCI), AI education, and systems theory. Eguchi et al. (2021) emphasize the importance of culturally responsive, context-aware AI education for students. Hirsch et al. (2023) explore HCI-based contextualization in cultural heritage systems. Murphy and Largacha-Martinez (2023) define contextualization as essential for aligning AI with real-world organizational complexity. Addas (2020) underscores that neglecting context leads to system failure. And Nardi (1996) offers a foundational perspective on context in activity theory and interface design.

These works collectively affirm ARC's claim: AI cannot produce relevant or trustworthy output unless the user supplies a frame. Context isn't an afterthought—it is the epistemic foundation.

Pillars in Practice

1. Accuracy: Verifiable, Fact-Based Outputs

Accuracy in AI is not simply about avoiding errors—it is about aligning machine output with domain-valid knowledge (Barocas, Hardt, & Narayanan, 2019). The ARC Framework mandates source-backed responses, requiring that all factual claims be explicitly tied to verifiable, timestamped references.

For instance, in climate science or medical diagnostics, even minor inaccuracies can lead to cascading failures (NIST, 2023). The ARC Framework mandates source-backed responses, requiring that all factual claims be explicitly tied to verifiable, timestamped references.

Example: "According to the IPCC Sixth Assessment Report (2022), carbon emissions must peak by 2025 to remain within 1.5°C warming targets."

Such specificity transcends vague generalizations, grounding AI claims in reproducible knowledge

2. Reliability: Traceable Reasoning and Institutional Standards

Ribeiro, Singh, and Guestrin (2016) argue that users must understand why an AI system produced a given output, not just what it produced. This aligns with ARC's call for "reasoned transparency," requiring that logic chains and institutional sources be traceable and explainable (Woods, 2025).

The U.S. National Institute of Standards and Technology's (NIST) AI Risk Management Framework explicitly states that "traceability, explainability, and transparency are foundational to responsible AI" (NIST, 2023, p. 14).

3. Contextuality: Contextual: Human-in-the-Loop Framing and Relevance

Context is the operating environment in which AI outputs are interpreted and made meaningful. Building on the work of Kahneman (2011), who demonstrates that human reasoning is inherently context-bound, as well as Dervin's (1998) theory of sense-making and Schön's (1983) reflective practitioner model, the ARC Framework asserts that context is the invisible structure shaping meaning. Without intentional framing, AI can produce technically accurate but practically irrelevant responses—a form of epistemic drift.

To mitigate this, ARC mandates user-defined roles, goals, and domains (e.g., "Act as a medical ethicist advising on gene editing policy") to ensure that outputs remain relevant, tone-appropriate, and decision-aligned. This emphasis on Contextual Intelligence shifts curation power back to the human, restoring interpretive authority where it belongs and making the user an active participant in meaning-making, not a passive recipient of machine-generated logic.

Tier-1: ARC as Scholarly Protocol for Validation and Rigor

Tier-1 is where the ARC Framework transitions from principle to practice. While the epistemic pillars of Accuracy, Reliability, and Contextuality define what trustworthy AI should be, Tier-1 outlines how to get there through a structured validation protocol. This protocol ensures that AI-generated responses meet minimum scholarly thresholds and are suitable for critical decision-making across domains.

The Tier-1 protocol consists of three interdependent components:

- **Act As**: The AI must assume a clearly defined expert role to contextualize its output (e.g., "Act as a climate policy advisor").
- **Reference**: All claims must be substantiated by credible sources—ideally peer-reviewed, institutional, or timestamped for domain relevance.
- **Cite Sources**: Every factual assertion must include traceable, properly formatted citations (APA 7th or discipline-appropriate style) to enable transparency and independent verification.

While the ARC Framework emphasizes explicit, verifiable citations wherever possible, it also acknowledges the current architectural limitations of large language models. In some cases, references may be implicitly indicated through recognizable synthesis patterns rather than formally cited sources.

As LLM capabilities continue to advance toward higher standards of traceability and epistemic accountability, ARC's structure is designed to accommodate and guide that evolution. Until then, active human oversight remains indispensable to validate the credibility, relevance, and contextual framing of all referenced material.

This protocol is not merely a formatting guideline; it is a trust mechanism that enables human users to evaluate credibility, challenge assumptions, and ensure alignment with context-specific standards. Each of the ARC pillars—Accuracy, Reliability, and Contextuality—is operationalized through this three-part structure, providing a robust foundation for responsible human-AI contextualization and decision framing.

The ARC Framework's emphasis on traceability, verifiability, and contextual oversight aligns directly with the U.S. National Institute of Standards and Technology's AI Risk Management Framework (NIST AI RMF), which identifies these elements as foundational to responsible AI. This compatibility positions ARC not only as a scholarly framework, but as an actionable model for NIST-aligned AI governance.

To demonstrate interoperability, *Figure 1* illustrates how ARC's Tier-1 validation protocol aligns with the institutional principles outlined in the U.S. National Institute of Standards and Technology's AI Risk Management Framework (NIST AI RMF). This crosswalk reinforces ARC's operational legitimacy as a model for compliant, responsible AI deployment across sectors.

Tier-2: Contextual Intelligence Tier — From Verification to Judgment

Tier-1 ensures credibility. Tier-2 ensures human responsibility. While Tier-1 validates AI outputs based on structured rigor, Tier-2 shifts the focus to critical human engagement—guiding users to analyze blind spots, reframe insights, and commit to context-aligned decisions. This phase is where human discernment reclaims center stage, turning passive interaction into *intentional contextualization and epistemic responsibility*.

- 🔍 **Analyze: Uncover Blind Spots and Assumptions**
  The "Analyze" step encourages users to interrogate what might be missing or distorted in an AI-generated response. This echoes Kahneman and Tversky's (1984) framing effect theory, which explains how people unconsciously accept certain premises unless prompted to question them. ARC embeds this meta-awareness by asking: *What are the risks, limitations, or implicit biases embedded in this output?*

- ↺ **Reframe / Refine: Shift Perspective**
  Reframing means taking the same data and looking at it through an alternate lens—be it behavioral science, equity, stakeholder impact, or mission alignment. This reflects Donald Schön's (1983) notion of the reflective practitioner and Karl Weick's (1995) dynamic model of sensemaking. ARC asks: *How would this insight shift if viewed through a different theoretical or strategic framework?*

- ⚖️ **Commit: Make Context-Aligned Decisions**
  Commitment is the moment of intentional action. Drawing on Dervin's (1998) theory of sense-making, ARC frames this as the user's responsibility to bridge the knowledge gap with informed, situated choices. It challenges users to ask: *Given the trade-offs and risks, what am I prepared to stand behind and execute?*

Tier-2 does not replace AI with human judgment; it restores **human agency** as the ultimate orchestrator of meaning and direction.

Based on trade-offs, risks, and context, what decision am I prepared to own?

The Curation Continuum: Intelligence as a Living Process

AI use is not a one-time transaction—it is a recursive cycle of of interpretation, refinement, and epistemic contextualization. The ARC Framework formalizes this reality through the **Curation Continuum**, a model inspired by systems thinking, organizational learning, and iterative design. ARC positions curation not as an afterthought, but as an **ongoing protocol** that safeguards human-AI collaboration from epistemic drift.

Scholarly Foundations:

- **Nonaka & Takeuchi (1995):** Knowledge creation is a continuous spiral that moves between tacit and explicit understanding.
- **Bawden & Robinson (2020):** Effective digital knowledge work requires sustained, intentional curation.
- **Snowden (2007):** The Cynefin Framework emphasizes the need for iterative, context-sensitive sensemaking in complex systems.

ARC is not a filter applied after the fact. It is a recursive engagement system—a way to ensure that decisions evolve as **context shifts**, **goals adapt**, and **new information emerges**.

The **Curation Continuum** reinforces three essential truths:

- ⟳ **Context must be refreshed**: Prior assumptions may no longer apply as conditions change.
- ↺ **Reliability must be recalibrated**: New data, models, or sources may shift what is considered valid.
- ✅ **Accuracy must be re-verified**: Outputs must be checked against updated references and constraints.

This approach protects against static reliance on AI and fosters **epistemic agility**. With ARC, the user doesn't just manage AI outputs—they **curate an evolving understanding**, staying aligned with reality through intentional, iterative stewardship.

Conclusion: Toward Trustworthy Human-AI Contextualization

The ARC Framework for Human-AI Contextualization offers a rigorously structured and philosophically grounded model for trusted AI use. It is more than a checklist—it is a dynamic practice of epistemic stewardship. Unlike generic human-AI collaboration models, ARC positions humans as the custodians of context—responsible for ensuring that AI outputs remain aligned with evolving ethical, cultural, and domain-specific meaning. While industry is now awakening to the power of context (Forbes Technology Council, 2023), ARC provides the tools to implement it. Grounded in scholarship, validated through applied use, and scalable across domains, ARC equips humans to frame, question, and commit to AI outcomes with clarity and care.

Figure 1

ARC Operational Alignment with NIST AI Risk Management Framework

NIST AI RMF Principle	ARC Tier-1 Implementation Example
Traceability	*Cite Sources:* All factual claims must be traceable to timestamped, domain-valid sources.
Explainability	*Reference + Role Clarity:* Outputs reflect declared expertise and rational justification.
Transparency	*Act As:* Outputs explicitly state AI's role, assumptions, and intended function.
Contextual Awareness	*Contextuality Pillar:* Human defines role, goal, and domain, ensuring relevance and alignment.

Note. ARC's Tier-1 validation protocol operationalizes key NIST AI RMF principles, positioning ARC as both academically rigorous and practically aligned with institutional AI governance frameworks. This alignment highlights how ARC protocols operationalize the AI governance principles outlined in NIST AI RMF 1.0 (2023).

Selected References (APA 7)

Addas, S. (2020). Contextualization in human-computer interaction. Transactions on Human-Computer Interaction, 2(4).

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning. https://fairmlbook.org

Bawden, D., & Robinson, L. (2020). Introduction to information science (2nd ed.). Facet Publishing.

Dervin, B. (1998). Sense-making theory and practice. Journal of Knowledge Management, 2(2), 36–43.

Eguchi, A., et al. (2021). Contextualizing AI education for K-12. PMC.

Floridi, L., & Cowls, J. (2022). A unified framework for AI in society. Harvard Data Science Review, 4(1).

Floridi, L. (2022). Artificial intelligence and its limits: An epistemological framework. Philosophy & Technology, 35(3), 1–23. https://doi.org/10.1007/s13347-022-00532-1

Forbes Technology Council. (2023, August 30). Contextualization: The key to unlocking generative AI's potential. https://www.forbes.com/sites/forbestechcouncil/2023/08/30/contextualization-the-key-to-unlocking-generative-ais-potential/

Hirsch, L., et al. (2023). Human-computer interaction for contextualizing cultural heritage. Applied Sciences, 14(17).

Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

Mittelstadt, B. D. (2023). AI as an epistemic technology. Science and Engineering Ethics, 29(1), 1–18. https://doi.org/10.1007/s11948-023-00451-3

Murphy, J. W., & Largacha-Martinez, C. (2023). Contextualizing AI. In Contextualizing AI in Practice (pp. 19–35). Springer.

Nardi, B. A. (1996). Context and consciousness: Activity theory and HCI. MIT Press.

National Institute of Standards and Technology (NIST). (2023). *AI Risk Management Framework (AI RMF 1.0).* U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company. Oxford University Press.

Raji, I. D., & Buolamwini, J. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33–44. https://doi.org/10.1145/3351095.3372873

Raji, I. D., et al. (2020). Closing the AI accountability gap. FAccT 2020 Proceedings.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Explaining the predictions of any classifier. KDD '16 Proceedings.

Schön, D. A. (1983). The reflective practitioner. Basic Books.

Shneiderman, B. (2020). Human-centered artificial intelligence: Three fresh ideas. AIS Transactions on Human-Computer Interaction, 12(3), 109–124. https://doi.org/10.17705/1thci.00131

Ehsan, U., & Riedl, M. O. (2020). Human-centered explainable AI: Towards a reflective sociotechnical approach. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–12. https://doi.org/10.1145/3313831.3376592

Weick, K. E. (1995). Sensemaking in organizations. Sage Publications.

Woods, E. (2025). *Applying the ARC Framework (Accurate, Reliable, Contextual) for Human-AI Contextualization: A Model for Assuring AI Integrity* [Scholarly paper]. ARC Initiative.
