PAICE.work

Making AI Collaboration Measurable, Teachable, and Governable

A framework for assessing collaboration capability, governance readiness, and risk in AI-assisted systems

Vision & Partnership Whitepaper

Author: Sam Rogers, PAICE.work PBC
Date: December 15, 2025
Version: 4.0

Audience:
Executives, Security, Risk, People, & Operations leaders evaluating AI collaboration measurement

---

Abstract

AI adoption is accelerating faster than organizations' ability to govern how AI is used in practice. While tools, policies, training, and technical controls are increasingly common, most governance mechanisms measure intent rather than behavior. As a result, many organizations lack visibility into how people+AI systems actually operate when AI outputs are incomplete, misleading, or wrong.

PAICE addresses this gap by providing a structured framework for measuring people+AI collaboration effectiveness as a property of a joint system. Rather than evaluating individual skill, model performance, or policy compliance in isolation, PAICE assesses observable collaboration behavior across five non-substitutable dimensions: Performance, Accountability, Integrity, Collaboration, and Evolution.

The framework uses a behavior-focused assessment model to produce interpretable, comparable signals suitable for governance, audit, and risk discussions. PAICE is designed to complement existing AI governance and risk management standards while avoiding invasive data collection, surveillance, or prescriptive enforcement.

This document describes the PAICE framework™, its methodological foundations, current validation status, and intended use boundaries. It is written for organizational leaders, governance bodies, researchers, and practitioners seeking a defensible way to reason about people+AI collaboration capability in real-world contexts.

How to Use This Document

This document is organized into two layers.

The core sections (Sections 1–8) describe the PAICE framework™, what it measures, how scores are produced, and how signals are intended to be used. These sections can be read sequentially.

The appendices provide additional methodological, privacy, validation, and use-boundary detail for readers who require greater depth. Core sections reference appendices where appropriate.

This document is descriptive, not prescriptive. It defines what PAICE measures and how signals should be interpreted, not the actions organizations must take.

Terminology and Naming

PAICE refers to the **People+AI Collaboration Effectiveness** framework described in this document.

PAICE.work refers to the **reference implementation** of the PAICE framework™.

PAICE.work PBC refers to the **public benefit corporation** that stewards the framework and operates the PAICE.work product.

Throughout this document, “people+AI” is used intentionally to emphasize collaboration as a property of a joint system rather than a human or model-centric activity.

Document Status and Scope

This document reflects the current state of the PAICE framework™ at the time of publication. Validation, benchmarking, and research efforts are ongoing and are described explicitly where relevant.

Claims are scoped conservatively and updated only when supported by evidence. Limitations and boundaries are stated deliberately.

---

1. Executive Summary

AI adoption is currently outpacing organizational ability to govern it. Tools ship quickly, access expands, and usage spreads, but visibility into how people+AI systems actually operate remains limited. As a result, many organizations are scaling AI without a clear view of where risk accumulates or how effectively oversight functions in practice.

Most AI governance mechanisms still measure intent rather than behavior. Policies, training completion, usage metrics, self-assessments, and technical controls are all necessary, but they do not show whether people+AI systems behave reliably when AI outputs are incomplete, misleading, or wrong. Recent public AI incidents have made this gap visible. In many cases, the technology performed within expected parameters. Failures emerged from routine people+AI interactions that went unmeasured until consequences surfaced.

PAICE addresses this gap by making people+AI collaboration effectiveness measurable. It treats collaboration as a property of a joint system rather than as a function of individual skill, role assignment, or model performance alone. The framework evaluates how responsibility, verification, judgment, and adaptation are exercised across people+AI systems under realistic conditions of uncertainty.

Effectiveness is expressed across five dimensions: Performance, Accountability, Integrity, Collaboration, and Evolution. Each dimension captures a distinct and non-substitutable aspect of quality that materially affects risk and outcomes. The framework is intentionally behavior-focused. It does not rely on self-reported values, quizzes, policy acknowledgments, or familiarity with specific tools as proxies for capability.

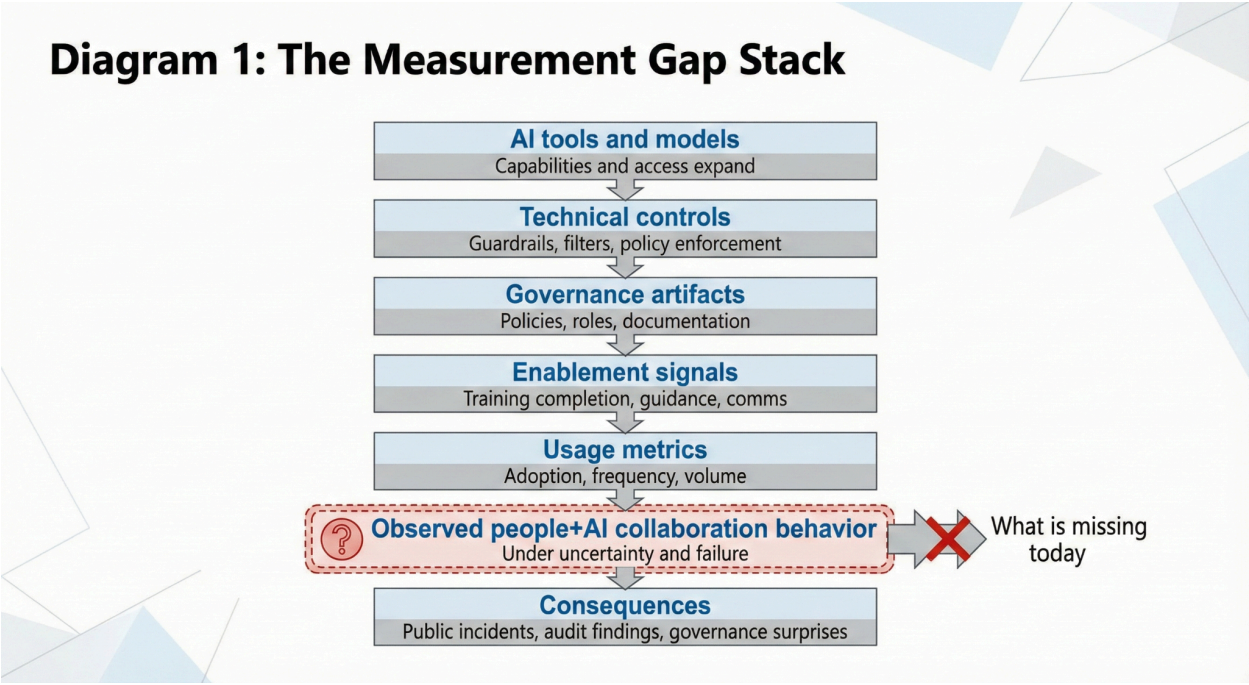
PAICE produces a composite score and dimensional breakdowns through a structured, conversational assessment model designed to elicit observable collaboration behavior. Weighting reflects relative risk exposure, with Accountability carrying greater influence due to its central role in preventing and mitigating AI-related failures. Scores are interpretable, comparable within defined contexts, and suitable for governance, audit, and risk discussions.

PAICE is designed to make AI risk visible without creating new privacy or surveillance risks for people. The framework emphasizes data minimization, aggregation, and behavior patterns over individual profiling, and is intentionally structured to support governance and oversight without invasive data collection. Privacy, security, and accessibility are treated as structural design constraints rather than compliance afterthoughts.

The framework is designed to complement existing AI governance and risk management standards, including NIST AI RMF, ISO/IEC 42001 (AI Management Systems), and emerging regulatory requirements related to human oversight. PAICE does not yet claim certification or compliance with existing standards. It provides evidence organizations can use to reason more precisely about oversight readiness and collaboration risk.

PAICE is live and in use, with an active validation and research program underway. While the framework and assessment approach are established, benchmarking and longitudinal evidence are being developed through pilot partnerships and research collaborations. Claims are updated conservatively as evidence accumulates.

PAICE.work provides the reference implementation of the PAICE framework™ and serves as its steward via our public benefit corporation. The framework is designed to evolve slowly and transparently, preserving interpretability and trust while supporting practical application across governance, risk, enablement, and operational contexts.



2. The Measurement Gap

The executive summary describes the problem at a strategic level. This section defines it operationally.

Organizations have extensive visibility into AI systems and limited visibility into how those systems are used in practice. Governance mechanisms typically focus on tools, policies, documentation, and training, but these controls stop short of measuring how people+AI systems behave when decisions are made under real conditions.

Most AI-related failures do not originate in novel or extreme scenarios. They emerge from routine people+AI interactions where AI-generated outputs are accepted, reused, or acted upon without sufficient verification, contextual judgment, or accountability. In these cases, the technology often performs within expected parameters. Risk accumulates in the collaboration layer, not the model layer.

Common organizational metrics are poorly suited to this problem. Tool adoption, usage frequency, training completion, and self-assessment provide indicators of exposure and intent, but they do not reveal whether people+AI systems behave reliably when AI outputs are incomplete, misleading, or wrong. Policies describe expected behavior, but they do not demonstrate that those behaviors are practiced consistently.

As a result, organizations frequently discover AI-related risk only after a failure becomes visible. Without a direct measure of people+AI collaboration capability, leaders lack a reliable basis for scaling access, regulators lack evidence of effective oversight, and organizations struggle to demonstrate reasonable prevention of foreseeable failures.

PAICE is designed to address this gap by making people+AI collaboration capability observable, comparable, and auditable. Further detail on the assessment and scoring mechanics is provided in Appendix A and Appendix B.

---

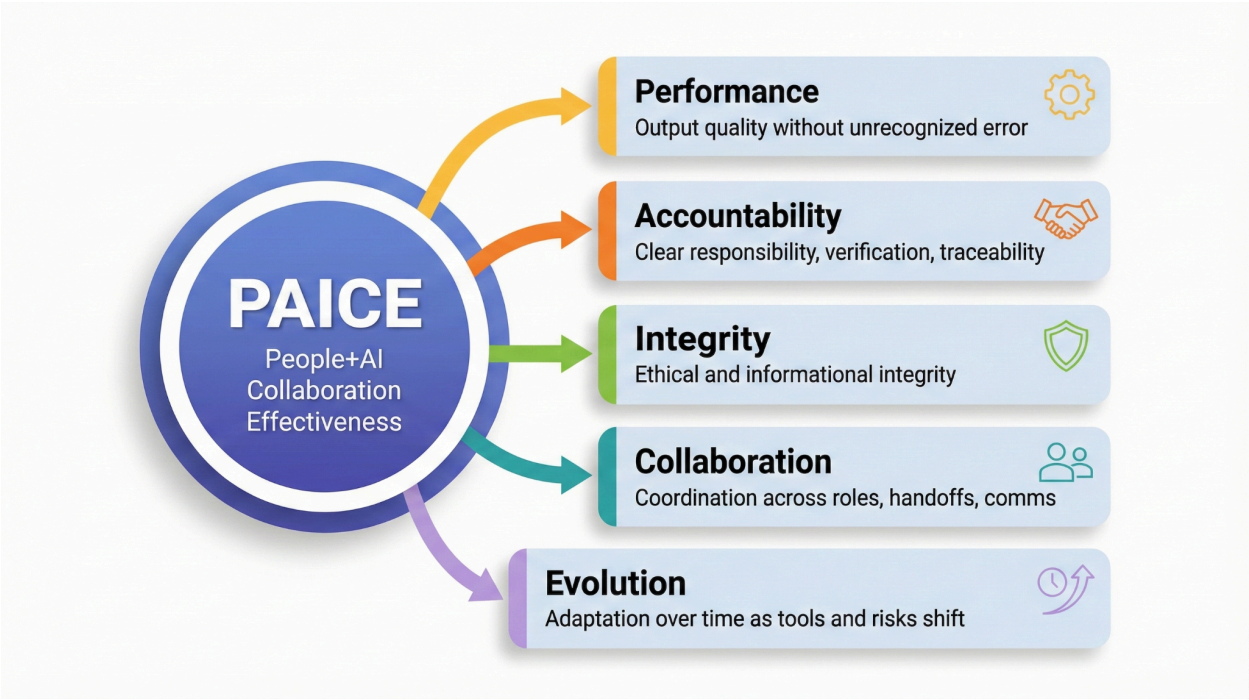
3. What PAICE Measures

PAICE measures people+AI collaboration effectiveness as a property of a joint system. The unit of capability analysis is not individual behavior or model performance in isolation, but the quality and reliability of outcomes produced when people and AI systems work together under conditions of uncertainty.

The framework evaluates how responsibility, verification, judgment, and adaptation are distributed across the people+AI system in practice. Collaboration effectiveness is assessed based on observable interaction patterns rather than stated intent, role definitions, or tool familiarity.

PAICE is designed to surface collaboration behaviors that materially affect risk and outcome quality, particularly when AI outputs are incomplete, misleading, or incorrect. The underlying behavioral signals and aggregation logic are described in Appendix B.

3.1 The Five Dimensions of people+AI Collaboration



PAICE expresses collaboration capability across five dimensions. Each dimension captures a distinct and non-substitutable aspect of collaboration quality. Strength in one dimension does not compensate for failure in another.

- **Performance** The ability of the people+AI system to improve task execution and output quality without introducing unrecognized error, over-reliance, or inappropriate delegation.
- **Accountability** The ability of the people+AI system to retain clear responsibility for outcomes, including verification of AI outputs, traceability of decisions, and ownership of downstream impact.
- **Integrity** The ability of the people+AI system to maintain ethical and informational integrity, including appropriate use of AI-generated content and avoidance of misleading or misrepresentative outputs.
- **Collaboration** The ability of the people+AI system to coordinate effectively across roles, workflows, and handoffs, including communication about AI involvement and limitations.
- **Evolution** The ability of the people+AI system to adapt practices over time in response to changing tools, contexts, and observed failures

Together, these dimensions form a composite view of collaboration capability that can be measured, compared, and tracked **over** time.

3.2 What PAICE Does Not Measure

PAICE does not measure:

- Individual intelligence or general competence
- Technical proficiency with specific AI models or vendors
- Policy compliance by declaration alone
- Attitudes, beliefs, or self-reported confidence

The framework is intentionally behavior-focused. It assesses demonstrated collaboration capability rather than personal attributes or intentions.

---

4. How PAICE Produces a Score

PAICE scores are derived from observed people+AI collaboration patterns elicited through a structured assessment process. The purpose of the scoring system is to produce a consistent, interpretable signal of collaboration capability under conditions of uncertainty.

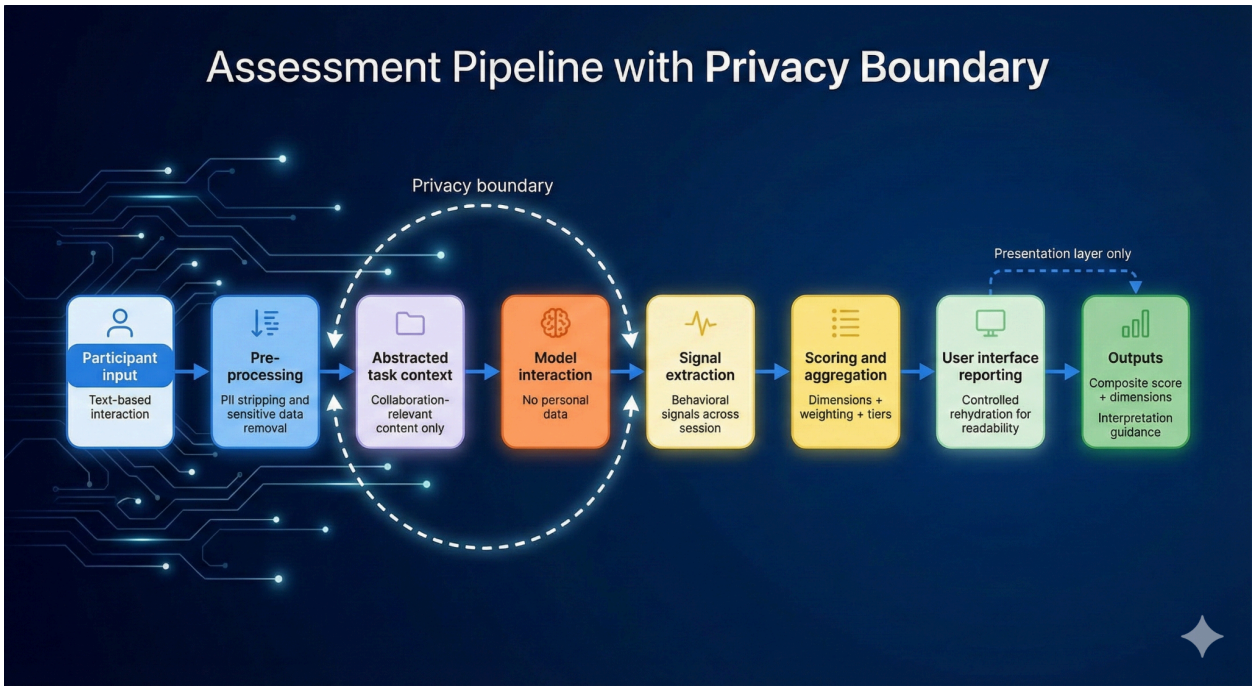
The assessment is designed to surface how people+AI systems behave when AI outputs are incomplete, misleading, or incorrect. Rather than relying on self-reporting or hypothetical knowledge checks, PAICE evaluates demonstrated responses to realistic prompts that require verification, judgment, and adaptation.

4.1 Behavioral Assessment Model

PAICE uses a conversational assessment model that adapts dynamically based on participant responses. This approach probes collaboration behavior in context rather than testing static knowledge, policy recall, or tool familiarity.

Key characteristics of the model include:

- Scenario-based prompts reflecting common AI-assisted work patterns
- Adaptive follow-up questions that surface reasoning and verification behavior
- Intentional introduction of ambiguity or failure conditions
- Pre-processing that strips personal and sensitive information before any interaction with language models, while preserving context for scoring and rehydration in the user interface
- Evaluation based on observable response patterns rather than correctness alone



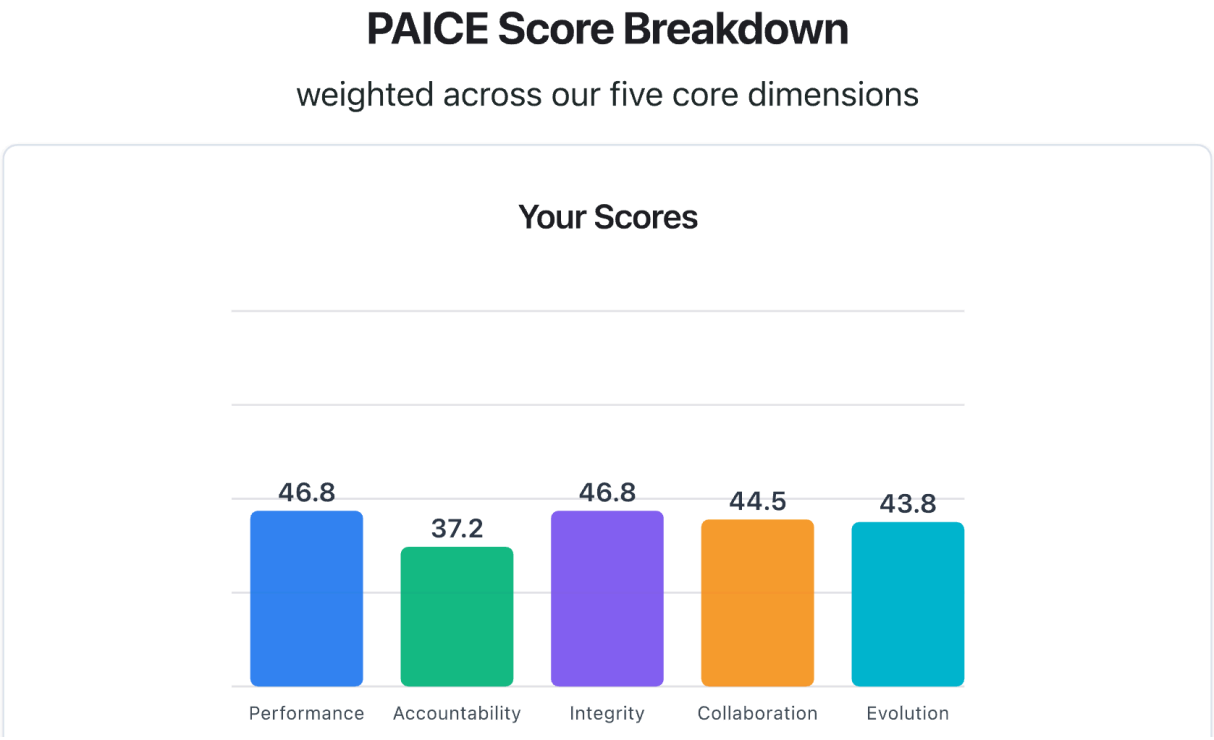
The model captures how responsibility and decision-making are exercised across the people+AI system as conditions change. A fuller description of assessment flow and prompt categories is provided in Appendix A.

4.2 Dimensional Scoring and Weighting

Assessment responses are evaluated across the five PAICE dimensions. Each dimension contributes independently to the overall signal and reflects a distinct aspect of collaboration capability.

While all five dimensions are necessary, they do not contribute equally to risk exposure. Each dimension reflects the aggregation of multiple underlying behavioral signals rather than a single response or metric. Accountability carries greater weight in the composite score due to its central role in preventing and mitigating AI-related failures.

Accountability captures whether responsibility for AI-assisted outcomes is retained, exercised, and traceable in practice. Failures in this dimension are consistently associated with public AI incidents, regardless of technical system performance or policy presence.



Weighting is applied consistently across assessments. Adjustments to weighting models are versioned and documented to preserve interpretability and longitudinal comparability, as detailed in Appendix B.

4.3 Score Expression and Tiers

PAICE scores are expressed as a composite value accompanied by dimensional breakdowns and capability tiers. This structure allows users to assess both overall collaboration capability and the specific dimensions contributing to risk or resilience.

Capability tiers provide a stable, interpretable shorthand for discussing collaboration maturity without implying linear progression or universal end states.

Scores are comparable within defined contexts such as similar roles, teams, or time periods. Cross-context comparison requires explicit interpretation and is addressed through benchmarking and research efforts described elsewhere in this document.

4.4 Interpretation, Not Judgment

PAICE scores are decision-support signals. They are not automated decisions, compliance determinations, or performance evaluations.

The framework intentionally separates measurement from interpretation. A PAICE score reflects collaboration capability at a point in time. How that signal is used depends on organizational context, governance structure, and risk tolerance.

PAICE does not prescribe actions, enforce controls, or replace human decision-making. Its role is to make previously unmeasured aspects of people+AI collaboration visible and discussable in a defensible way.

---

5. Why PAICE Is Defensible

This section moves beyond problem framing to address how PAICE holds up under scrutiny from legal, risk, procurement, and regulatory perspectives. The PAICE approach designed to be reasonable, bounded, and suitable for use in governed environments.

PAICE's defensibility rests on four design choices: scope limitation, methodological discipline, standards compatibility, and interpretability. Together, these choices reduce the risk of overreach while increasing the usefulness of the signal produced.

5.1 Clear Scope and Limits

PAICE measures people+AI collaboration capability and nothing else. It does not evaluate individual performance, certify compliance, or make automated decisions. By constraining scope deliberately, the framework avoids being repurposed as a surveillance mechanism, performance tool, or enforcement system.

This limitation is not incidental. It is a prerequisite for responsible measurement in environments where AI use intersects with labor law, privacy regulation, and organizational governance.

5.2 Methodological Discipline

PAICE prioritizes demonstrated behavior over declared intent. Unlike governance approaches that rely primarily on policies, role assignments, or attestations, PAICE evaluates how responsibility, verification, and judgment are exercised when people+AI systems encounter uncertainty or failure.

This focus reduces reliance on self-reporting and mitigates known weaknesses in intent-based oversight, particularly in rapidly evolving technical environments.

5.3 Compatibility with Existing Governance Frameworks

PAICE is designed to be compatible with established AI governance and risk management frameworks, including NIST AI RMF and ISO/IEC 42001 (AI Management Systems). Supporting detail on privacy, security, and accessibility considerations is provided in Appendix C. These frameworks emphasize accountability, oversight, and continuous improvement but provide limited guidance on how to measure collaboration behavior directly.

PAICE does not replace governance programs or technical controls. It provides a complementary measurement layer that organizations can use alongside existing processes without restructuring them.

5.4 Interpretability and Audit Support

PAICE scores are designed to be interpretable rather than opaque. Dimensional breakdowns, documented weighting, and versioned methodology allow stakeholders to understand how signals are generated and what they represent.

This interpretability supports audit, internal review, and regulatory inquiry by enabling organizations to explain how collaboration capability was assessed and how resulting decisions were informed. PAICE provides evidence to support judgment, not automated conclusions.

---

6. Validation Status and Evidence Development

PAICE is intentionally conservative in how validation claims are made. At the time of publication, the framework, assessment model, and scoring methodology are operational, but large-scale benchmarking and longitudinal validation are still in progress.

6.1 Current Status

PAICE is currently in an early validation phase. The assessment model and scoring logic have been exercised in controlled and exploratory contexts to test clarity, consistency, and interpretability. These efforts have focused on confirming that prompts elicit observable collaboration behavior, that dimensional signals are distinct, and that scoring produces stable results within defined contexts.

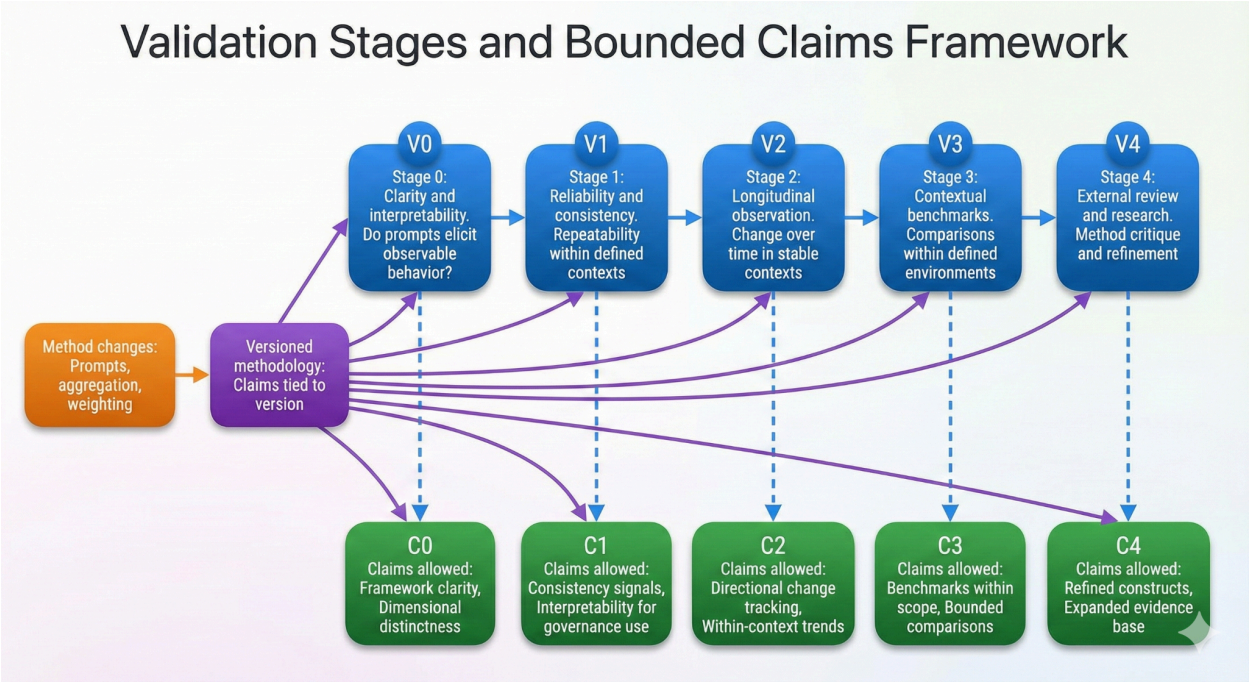
Formal pilot programs are scheduled to begin in the upcoming operating period. Planned validation phases and evidence criteria are described in Appendix D. These pilots are designed to generate empirical evidence across varied organizational settings, roles, and risk profiles.

6.2 Evidence Roadmap

The validation approach is structured to mature over time without overstating confidence prematurely. Planned evidence development includes:

- Pilot-based data collection across teams and functions
- Dimensional consistency analysis within and across contexts
- Longitudinal observation to assess change over time
- External research collaboration to support methodological review

Results will be published selectively and conservatively as evidence accumulates. Claims will be updated only when supported by data.



6.3 Appropriate Use During Validation

During this phase, PAICE scores should be used to provide directional signals rather than definitive judgments. They are well-suited to identifying areas of elevated risk, informing governance conversations, and prioritizing further investigation.

PAICE is not intended to be used as a sole decision-making input, certification mechanism, or performance evaluation tool. Its value during validation lies in making collaboration behavior visible earlier than traditional indicators allow.

This staged validation approach is intentional. It reflects a preference for credibility over speed and for evidence over assertion.

---

7. From Signal to Action

PAICE provides a structured signal that organizations can use to reason about people+AI collaboration risk and capability. The value of the signal lies in how it informs judgment across governance, operational, and learning contexts without collapsing those contexts into a single set of prescribed responses.

During early adoption, PAICE scores are most effective when used to surface patterns and prompt focused discussion. They help organizations identify where collaboration introduces risk, where accountability may be diffuse, and where existing controls or practices may not be operating as intended.

7.1 Governance and Oversight Use

For governance bodies, PAICE grounds AI risk discussions in observable collaboration behavior. Scores and dimensional breakdowns can inform decisions about access, escalation thresholds, review cadence, and control design, supplementing policy artifacts and incident history with evidence from day-to-day practice.

7.2 Operational and Design Use

At the operational level, PAICE signals can inform workflow design, role clarity, and oversight mechanisms. Patterns identified through assessment may indicate the need for clearer verification steps, improved handoffs, or adjustments to how AI outputs are reviewed, reused, or escalated.

Because PAICE measures collaboration capability rather than tool performance, insights remain applicable across different models, vendors, and deployment contexts.

7.3 Learning and Capability Development

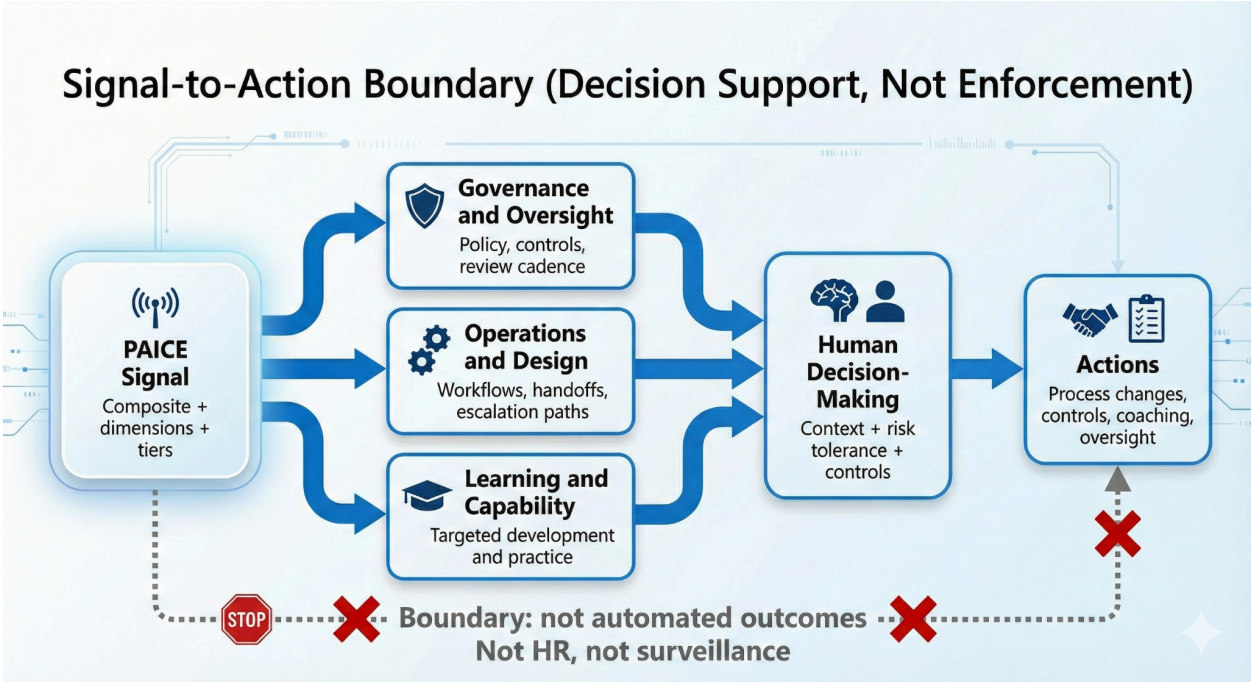
PAICE scores can support learning and capability development by highlighting specific dimensions where collaboration behavior limits effectiveness or increases risk at the team and organizational level. This enables more targeted interventions without assuming that more training or stricter policy alone will resolve underlying issues.

Appropriate collaboration capability depends on context, role, and risk tolerance. PAICE does not assume uniform targets or linear progression.

7.4 Boundaries of Action

PAICE provides evidence to inform professional judgment. It does not generate recommendations, enforce controls, or determine outcomes.

By maintaining a clear separation between signal and action, the framework remains applicable across varied organizational environments while avoiding the risks associated with prescriptive or one-size-fits-all approaches. Guidance on appropriate and inappropriate uses is summarized in Appendix E.



---

8. Stewardship and Evolution

PAICE is designed to evolve deliberately rather than rapidly. Because the framework functions as a measurement standard, stability, interpretability, and trust take precedence over feature expansion or frequent revision.

Changes to the PAICE framework™, assessment model, and scoring methodology are governed through explicit versioning and documented rationale. Updates are introduced only when they improve measurement fidelity, reduce ambiguity, or are required to address material changes in people+AI collaboration patterns.

PAICE.work serves as the steward of the framework through its public benefit corporation structure. This stewardship model is intended to balance innovation with responsibility, ensuring that commercial incentives do not compromise methodological integrity or public trust.

8.1 Versioning and Transparency

Material changes to dimensions, weighting, prompts, or interpretation guidance are versioned and disclosed. Historical scores remain interpretable within their original methodological context, preserving longitudinal usefulness.

Where appropriate, updates are accompanied by explanatory documentation describing the motivation, scope, and expected impact of changes.

8.2 Research and Community Engagement

PAICE is intended to support ongoing research into people+AI collaboration effectiveness. Partnerships with academic and industry researchers will be used to test assumptions, validate constructs, and refine measurement approaches over time.

Feedback from practitioners, governance bodies, and participants is incorporated selectively to improve clarity and applicability without compromising scope or rigor.

8.3 Long-Term Orientation

PAICE is not designed to chase short-term trends in AI capability or tooling. Its focus remains on durable collaboration patterns that persist across model generations and deployment contexts.

By prioritizing careful stewardship over rapid expansion, PAICE aims to remain a trusted reference point for measuring people+AI collaboration as both technology and organizational practices continue to evolve.

Appendices

Appendix A. Assessment Model and Prompt Structure

This appendix provides deeper detail on how the PAICE assessment elicits observable people+AI collaboration behavior without relying on quizzes, self-reporting, or tool-specific knowledge. The intent is to explain assessment mechanics at a methodological level while avoiding disclosure that would enable optimization or gaming.

A1. Assessment Flow

PAICE uses an adaptive, conversational assessment flow designed to approximate real-world AI-assisted work rather than simulated testing conditions. Participants are presented with realistic scenarios that reflect common patterns of AI use, such as drafting, summarization, analysis, or decision support, but with intentional ambiguity or incompleteness introduced.

The assessment elicits a real working scenario, followed by context-sensitive probes. These probes are not scripted in a fixed sequence. Instead, they adapt based on participant responses in order to surface how verification, judgment, and responsibility are exercised as conditions evolve.

This adaptive structure allows the assessment to observe not only what actions participants take, but when they take them, what triggers changes in behavior, and how uncertainty is

managed across interaction turns. The goal is to reveal collaboration patterns that would otherwise remain implicit in routine work.

A2. Prompt Categories

Assessment prompts are organized into categories that correspond to distinct collaboration behaviors rather than task difficulty or domain expertise. Categories are mixed and revisited across the session to reduce predictability and discourage surface-level optimization.

Prompt categories include scenarios that surface verification practices, handling of uncertainty, escalation and decision ownership, reuse and transfer of AI outputs across contexts, communication across roles, and responses to simulated error or failure conditions.

Prompts are intentionally framed to be domain-agnostic. Participants are not rewarded for specialized knowledge, nor penalized for unfamiliarity with a particular tool or workflow. This ensures that observed behavior reflects collaboration patterns rather than technical fluency.

A3. Response Interpretation Principles

PAICE does not score responses based on correctness or stylistic quality. Instead, it evaluates interaction patterns that emerge across the session. Interpretation focuses on when and how verification is performed, how uncertainty is acknowledged or suppressed, whether accountability is retained or deferred, and how AI outputs are integrated into downstream decision-making.

Responses are interpreted in context rather than isolation. A single strong or weak response does not determine outcomes. Patterns across multiple prompts inform signal generation, allowing the assessment to distinguish between isolated lapses and systemic collaboration behavior.

This approach reduces sensitivity to coaching, memorization, or performative compliance. Participants who understand how to articulate desired behaviors but do not demonstrate them consistently will not produce the same signal as those who exhibit disciplined collaboration practices in context.

A4. Boundaries and Safeguards

The assessment model is designed to measure human participation in people+AI systems. It is not intended for use by autonomous agents, scripted automations, or machine-mediated proxy interactions.

PAICE includes safeguards to detect interaction patterns inconsistent with direct human participation, such as sustained machine-speed response timing, absence of deliberative revision, and behavioral signatures associated with agentic browsing or task delegation

systems. Interactions that appear to be generated or mediated entirely by machines are excluded from scoring.

This boundary preserves the integrity of the measurement and reflects the current scope of PAICE, which focuses on human responsibility and judgment within AI-assisted systems. Future extensions to machine-mediated collaboration, if any, would require distinct methodology and validation.

Appendix B. Scoring Aggregation and Weighting

This appendix provides methodological detail on how PAICE scores are constructed, aggregated, and interpreted. It is intended to support audit, research, and implementation review.

B1. Underlying Signals

PAICE does not treat individual responses, statements, or correct answers as scores. Instead, it evaluates patterns of interaction that emerge across an assessment session. These patterns are referred to as underlying behavioral signals.

A signal represents an observable property of how a people+AI system operates under uncertainty. Signals are derived from features such as whether verification occurs before reuse, how uncertainty is acknowledged or suppressed, how responsibility is claimed or deferred, and how AI outputs are scoped, escalated, or integrated into downstream work.

Signals are intentionally defined at a level that abstracts away individual phrasing or stylistic differences. This reduces sensitivity to language proficiency, verbosity, or role-specific jargon, and focuses measurement on collaboration behavior rather than expression.

Each signal is designed to be interpretable across roles and contexts. Signals are not binary, nor are they tied to single prompts. They are inferred across multiple interaction points to reduce noise and gaming.

B2. Aggregation Model

Signals are aggregated into dimensional scores using a structured roll-up model. The aggregation model is designed to preserve dimensional independence while allowing multiple signals to contribute to each dimension.

No single signal is sufficient to determine a dimensional outcome. This avoids single-item dominance and ensures that isolated behaviors do not disproportionately influence scores. Where signals appear contradictory, aggregation favors risk-sensitive interpretation rather than optimistic averaging.

The model supports partial completeness. Assessments do not require every signal to be observed in order to produce a score. Instead, confidence bounds are applied implicitly through aggregation density, with sparse signal environments yielding more conservative dimensional outcomes.

Aggregation logic is stable within a methodology version. Changes to signal definitions or aggregation rules are treated as methodological updates and versioned accordingly.

B3. Weighting Philosophy

PAICE applies weighting at the dimensional level rather than at the individual signal level. This preserves interpretability and avoids creating opaque scoring interactions.

Accountability carries greater weight in the composite score due to its central role in preventing and mitigating failure in people+AI systems. Accountability is not weighted because it tends to score lower, but because failures in accountability disproportionately amplify the impact of weaknesses in other dimensions.

For example, strong performance or collaboration without retained accountability increases the likelihood that unverified outputs propagate unchecked. Conversely, strong accountability can partially mitigate weaknesses elsewhere by introducing verification, escalation, or human judgment at critical points.

Weighting reflects relative risk exposure, not moral priority. It is applied consistently within a methodology version and documented to support audit and longitudinal comparison.

B4. Tiers and Interpretive Boundaries

Capability tiers provide an interpretive layer over composite scores. They are intended to support communication and comparison without implying linear maturity models or universal end states.

Tier	Score Range	Typical Behaviors
Constrained	0–29	Avoids or misuses AI; frequent errors
Informed	30–49	Understands concepts but applies inconsistently
Proficient	50–69	Performs reliably with guidance
Advanced	70–89	Self-corrects and anticipates AI failure modes
Exceptional	90–100	Designs systems for human-AI co-creation

Tiers are derived from score ranges within a defined methodological context. They are not absolute indicators of organizational quality and should not be compared across materially different roles, industries, or risk environments without interpretation.

Tiers do not prescribe actions. They provide shorthand for discussing collaboration capability relative to context and risk tolerance.

Appendix C. Privacy, Security, Accessibility, and Standards

This appendix provides detail on how privacy, security, and accessibility are treated as foundational design constraints within PAICE. These considerations are integral to both risk posture and measurement integrity, not secondary compliance concerns.

C1. Data Minimization

PAICE is designed around a principle of strict data minimization. The assessment process collects only the information necessary to observe collaboration behavior and derive scoring signals. Personal identifiers, sensitive attributes, and contextual details that do not materially affect collaboration patterns are intentionally excluded.

This approach serves three purposes. First, it reduces regulatory and operational risk by limiting the volume and sensitivity of data handled. Second, it improves measurement quality by preventing scores from being influenced by personal characteristics unrelated to collaboration capability. Third, it constrains downstream system behavior by ensuring that richer personal context is not available for unintended secondary use.

Data minimization in PAICE is not achieved through policy alone. It is enforced through architectural choices that limit what data can enter the assessment pipeline in the first place. This reduces reliance on trust-based controls and supports verifiable privacy guarantees.

Data retention policies are aligned with these principles. Assessment data is retained only as long as required to support scoring, interpretation, auditability, and explicitly agreed research purposes.

C2. PII Stripping and Rehydration

Before any interaction with language models occurs, participant inputs are processed to remove personal and sensitive information. This boundary is enforced consistently and treated as a mandatory step in the assessment pipeline rather than an optional safeguard or configuration choice.

This design reflects a privacy-by-design approach in which potentially identifying information is never made available to generative systems. As a result, language models operate only on abstracted task context and collaboration-relevant content, not on personal data.

The successful operation of this boundary is itself part of what PAICE measures. Maintaining separation between collaboration behavior and personal data reflects disciplined people+AI system design and contributes directly to collaboration integrity and risk reduction.

To preserve usability and interpretability, the participant experience maintains conversational continuity through controlled rehydration within the user interface. Rehydration occurs only in the presentation layer and is explicitly excluded from model interaction, scoring logic, and downstream processing.

C3. Security Practices

PAICE applies security controls appropriate to the sensitivity and use of assessment data, with an emphasis on minimizing attack surface rather than maximizing feature complexity. Security design follows a least-privilege and separation-of-concerns approach aligned with privacy-first system design.

Controls are designed to protect confidentiality, maintain integrity, and support traceability of handling practices without introducing persistent monitoring or excessive logging of user behavior. Where logs or records are required for auditability, they are scoped narrowly and retained for defined purposes.

Security mechanisms are reviewed as part of stewardship and versioning processes to ensure continued adequacy as the framework evolves and as threat models change.

C4. Accessibility

PAICE is designed to avoid conflating collaboration capability with physical ability, cognitive load tolerance, or interface familiarity. The assessment model emphasizes reasoning, verification, and judgment behaviors rather than speed, memory, or fine motor interaction. PAICE.work maintains WCAG 2.1 AA accessibility compliance.

Interaction design choices aim to ensure that participants are not disadvantaged due to assistive technology use, language differences, or non-standard work environments. Accessibility considerations are incorporated to reduce bias and improve the validity of collaboration measurement across diverse participant populations.

C5. Standards Mapping Table

This mapping provides compliance teams with specific controls/functions to reference when demonstrating AI collaboration competence to auditors.

PAICE Dimension	ISO/IEC 42001:2023	NIST AI RMF	EU AI Act
Performance	Control 7.2 (Competence) Control 8.1 (Operational planning)	MANAGE 1.1 (Allocation of resources) MANAGE 2.1 (Planning)	Article 9 (Risk management system)
Accountability	Control 5.2 (Risk assessment) Control 9.1 (Performance evaluation)	GOVERN 1.1 (Legal requirements) MANAGE 2.2 (Accountability)	Article 14 (Human oversight) Article 17 (Quality management)
Integrity	Control 6.2 (Information security) Control 8.2 (Testing)	GOVERN 1.2 (Ethical guidelines) MAP 1.1 (Context)	Article 10 (Data governance) Article 15 (Accuracy)
Collaboration	Control 7.3 (Awareness) Control 8.3 (Communication)	MANAGE 2.3 (Operator override) MANAGE 3.1 (Documentation)	Article 13 (Transparency) Article 14 (Human oversight)
Evolution	Control 10.2 (Continual improvement) Control 9.2 (Monitoring)	MANAGE 4.1 (Monitoring) MANAGE 4.2 (Performance metrics)	Article 72 (Post-market monitoring)

Appendix D. Validation Methods and Evidence Development

This appendix provides additional detail on how evidence for PAICE is generated, evaluated, and communicated over time. Validation is treated as an ongoing process rather than a one-time certification event, reflecting the evolving nature of people+AI collaboration.

D1. Validation Philosophy

PAICE adopts a conservative validation philosophy that prioritizes credibility, interpretability, and reproducibility over rapid claims of predictive power. The framework is designed to surface collaboration behavior reliably before asserting external correlations or benchmarks.

This approach reflects two constraints. First, collaboration behavior is context-sensitive and cannot be meaningfully validated through decontextualized testing alone. Second, premature claims of prediction or certification risk undermining trust in the measurement if later evidence diverges.

As a result, PAICE validation proceeds incrementally, with each stage expanding the scope of claims only when supported by accumulated evidence.

D2. Validation Phases

Evidence development is structured into distinct phases, each with explicit objectives and limits.

Early-phase validation focuses on clarity and interpretability. This phase tests whether assessment prompts elicit observable collaboration behavior, whether dimensional signals are distinct, and whether scoring outputs are understandable and actionable to practitioners.

Mid-phase validation emphasizes reliability and consistency. Pilot-based data is used to examine repeatability within similar contexts, sensitivity to role and workflow differences, and stability of dimensional signals over time.

Later-phase validation supports comparative and longitudinal use. This includes the development of benchmarks within defined contexts, analysis of change over time, and exploration of relationships between collaboration signals and observed incidents or near-miss indicators.

D3. Evidence Sources

Evidence supporting PAICE claims may include multiple complementary sources rather than a single validation metric. These sources include:

- Consistency of dimensional signals across repeated assessments in similar contexts
- Distinctness and non-substitutability of dimensions
- Practitioner interpretation and decision usefulness
- Observed alignment with known risk patterns or documented failures
- External methodological review and critique

No single source is treated as definitive. Evidence is evaluated collectively to support bounded claims appropriate to the maturity of the framework.

D4. Versioning and Claim Management

Validation claims are explicitly tied to methodology versions. When assessment prompts, aggregation logic, or weighting models change, claims are reviewed to determine whether prior evidence remains applicable.

Claims about predictive utility, benchmarking, or generalizability are updated conservatively and only when supported by sufficient data. Where evidence is incomplete or context-dependent, limitations are stated explicitly.

This disciplined approach to claim management is intended to preserve interpretability over time and to ensure that organizations relying on PAICE signals can understand what is known, what is assumed, and what remains under investigation.

Appendix E. Appropriate and Inappropriate Uses

This appendix clarifies how PAICE signals are intended to be used and where their use would be inappropriate or misleading. These boundaries are essential to preserving the integrity of the framework and preventing misuse as PAICE adoption expands.

E1. Intended Uses

PAICE is designed to support judgment, not replace it. Appropriate uses focus on making previously unmeasured aspects of people+AI collaboration visible so they can be discussed, governed, and improved responsibly.

Appropriate uses include informing governance and oversight discussions, identifying areas of elevated collaboration risk, supporting workflow and control design, prioritizing further investigation, and contributing evidence for audit, review, or research purposes. In these contexts, PAICE functions as an input to decision-making alongside professional judgment, existing controls, and contextual knowledge.

PAICE is also suitable for longitudinal observation within stable contexts, where changes in collaboration behavior over time can inform learning, adaptation, and governance refinement.

E2. Uses That Require Caution

Certain applications of PAICE signals require heightened care due to their potential impact. These include decisions related to access, authorization, escalation thresholds, or operational constraints.

In such cases, PAICE scores should be used to prompt review rather than to trigger automatic outcomes. Additional context, corroborating evidence, and human judgment are required to avoid over-reliance on a single measurement signal.

PAICE is not currently designed to provide real-time monitoring or continuous enforcement. Attempts to repurpose it for these functions risk undermining both measurement validity and trust.

E3. Inappropriate Uses

PAICE is not intended to be used as a hiring, promotion, disciplinary, or individual performance evaluation tool. It does not assess personal competence, intent, or worth, and using it for these purposes would exceed its scope and introduce ethical and legal risk.

The framework is also not intended to function as a compliance certification, regulatory approval mechanism, or substitute for established governance programs. PAICE does not declare compliance, certify readiness, or determine acceptability of AI use in isolation.

PAICE should not be used as a surveillance or monitoring system. The assessment model is episodic and behavior-focused by design, and extending it into persistent observation would violate core privacy and measurement principles.

E4. Boundary Cases and Escalation

In situations where organizations consider using PAICE signals in high-stakes contexts, explicit governance review is required. Boundary cases should be evaluated by designated oversight bodies with access to contextual information and authority to interpret signals appropriately.

PAICE provides evidence to inform these discussions but does not resolve them. Preserving this separation between measurement and action is essential to maintaining the legitimacy and long-term usefulness of the framework.

E5. Current Scope and Limitations

At the time of publication, PAICE is implemented and validated within a defined set of operational constraints. The current assessment experience is text-based, conducted in English, and optimized for desktop interaction. These constraints reflect where the framework has been most thoroughly exercised and evaluated.

While the underlying measurement principles are expected to generalize beyond these boundaries, PAICE does not yet claim validated equivalence across other modalities, languages, or device contexts. Expansion into additional interaction modes, languages, and form factors is an explicit area of future work and will be pursued only when methodological integrity and interpretability can be preserved.

These limitations are stated deliberately. PAICE prioritizes evidence-backed claims over aspirational coverage and will update scope disclosures as validation expands.

Acknowledgments

The PAICE framework™ has been shaped by contributions, critique, and influence from practitioners, researchers, and leaders working across AI governance, risk management, learning systems, and organizational design. Conceptual influences include work in human factors, safety engineering, socio-technical systems, and risk governance.

Special thanks to:

- **Nate B. Jones** and the **Nate's Notes Substack community** - where this idea evolved from "someone should measure this" to "I'm going to build it"
- **Marily Nika's AI Product Manager cohort** - where the product and the pitch first came together as a viable startup
- **The SnapDev.ai team** - for exceptional technical execution and tool support
- **Advisors** - Barry Kayton, Megan Torrence, Mike Simmons, Lee Rodrigues
- **Early testers** - feedback from early reviewers and collaborators has informed the framework's emphasis on behavior, interpretability, and restraint.

Acknowledgment does not imply endorsement.

About PAICE

PAICE is a measurement framework for assessing people+AI collaboration effectiveness. It is designed to make collaboration behavior observable, comparable, and discussable without relying on self-reporting, surveillance, or prescriptive enforcement.

PAICE treats collaboration as a system property rather than an individual attribute. Its purpose is to support governance, oversight, learning, and responsible scaling of AI use in organizational contexts.

About PAICE.work

PAICE.work provides the reference implementation of the PAICE framework™. It delivers structured assessments, scoring, and reporting aligned with the methodology described in this document.

PAICE.work is operated by PAICE.work PBC, a public benefit corporation. This structure is intended to balance commercial sustainability with responsible stewardship of a measurement framework that may influence governance, policy, and organizational decision-making.

Version History

v3.1 - *November 12, 2025* Initial public release of the PAICE framework™ with early appendices and L&D-focused framing for DevLearn 2025.

v4.0 - *December 15, 2025* Structural and methodological revision. Limited appendices, clarified scope and boundaries, strengthened privacy and validation posture, and reframed the document as an evergreen reference standard.

Licensing and Use

This document is provided for informational and reference purposes.

Permission is granted to cite, reference, and discuss the PAICE framework™ with attribution. Use of the PAICE name or framework for certification, compliance claims, or derivative measurement tools requires explicit permission.

Citation:
Rogers, S. (2025). *PAICE.work: Making AI Collaboration Measurable, Teachable, and Governable - Vision & Partnership Whitepaper (v4.0)*. PAICE.work PBC.
https://PAICE.work/whitepaper

License:
Creative Commons Attribution 4.0 International (CC BY 4.0)
You may share and adapt with attribution.

Contact

For pilot participation, research collaboration, or stewardship inquiries:
https://paice.work/contact
