The Human Enhancement Quotient (HEQ)

Measuring Collaborative Intelligence for Enterprise AI Adoption

A Quantitative Framework Built on the Factics Methodology

Basil C. Puglisi, MPA

Human-AI Collaboration Strategist | AI Governance Consultant

Creator of Factics, Growth OS, HAIA-RECCLIN, and HEQ

IMPORTANT: SCOPE AND INTENDED USE

HEQ: The First Integrated Framework Combining Governance Architecture, Measurement, and Organizational Deployment

This framework addresses a critical enterprise gap: organizations need to measure AI collaboration capability, but no structured methodology exists. HEQ provides **auditable structure where currently only managerial intuition exists**.

What HEQ Delivers: A developmental framework with high cross-platform consistency (0.96 consistency score across five AI platforms) that organizations can pilot to bring structured assessment to AI readiness evaluation. HEQ is designed as **a factor, not THE factor** in workforce decisions. Scores inform development conversations, identify training needs, and provide one data point among many in talent management.

Current Status: Enterprise Pilot Edition. Cross-platform consistency testing demonstrates technical stability. Multi-user validation studies (n=100+) are underway per the Research Agenda. Organizations adopting now become validation partners shaping the enterprise standard.

Prompt Evolution: HEQ prompt instruments were published Q2/Q3 2025. Advanced interactive prompt work, including questionnaire integration into evaluations, is scheduled for Q1 2026 release.

Prompt Evolution: HEQ prompt instruments were published Q2/Q3 2025. Advanced interactive prompt work, including questionnaire integration into evaluations, is scheduled for Q1 2026 release.

Technical Validation: Cross-platform consistency score of 0.96 across five AI platforms (ChatGPT, Claude, Gemini, Perplexity, Grok) under identical prompts. Informal cross-user testing (n=10) showed dimensional score variance within ±4 points. This establishes technical stability; formal multi-user validation is in progress.

Origin Context: HEQ emerged from observing that organizations were making workforce decisions about "AI trainability" with zero objective methodology. When companies publicly referenced AI skill gaps in workforce restructuring, no standardized assessment existed to define what "AI trainable" meant or to give employees fair opportunity to demonstrate or develop capability. HEQ provides structured methodology to **prevent arbitrary decisions**, not to rationalize them.

Enterprise Adoption: Organizations adopting HEQ become validation partners in generating empirical performance data for enterprise-grade deployment. Early adopters contribute anonymized pilot data to accelerate validation while receiving structured methodology for workforce development planning.

LEGAL DEFENSIBILITY: Why Structure Is Safer Than Gut Feeling

Courts have consistently ruled that structured processes with documentation are less discriminatory than pure subjective judgment, even when the structure is imperfect. Using HEQ as **one factor among many** in a documented decision process is **more**

legally defensible than unstructured manager intuition, not less. The legal risk matrix:

- **Manager gut feeling alone:** HIGH RISK. No documentation, arbitrary, impossible to defend against disparate impact claims.

- **HEQ as one factor + interviews + work samples:** MEDIUM RISK. Documented process demonstrates good faith effort at fairness.

- **HEQ as sole determinant:** VERY HIGH RISK. Misrepresents developmental framework as validated test.

RESPONSIBLE USE IN EMPLOYMENT CONTEXTS

Organizations using HEQ for hiring, promotion, or workforce planning must:

- **Use HEQ as one factor among many.** Combine with interviews, work samples, performance history, and manager judgment.

- **Document decision rationale.** Show how HEQ scores informed (but did not determine) the outcome.

- **Monitor for adverse impact.** Track whether HEQ scores disproportionately affect protected classes and adjust if needed.

- **Provide development pathways.** Low HEQ score triggers training plan, not automatic negative outcome.

- **Respect validation status.** Treat scores as directional indicators with ±4 point variance until formal validation completes.

This approach should be legally defensible because it demonstrates good faith effort to add structure and transparency where previously only subjective judgment existed.

Table of Contents

1. Executive Summary

The Enterprise Problem. Organizations lack a quantitative method to assess how effectively employees collaborate with AI. Hiring managers cannot measure a candidate's capacity for AI partnership. Training programs cannot prove ROI. Performance reviews cannot track whether AI collaboration skills are developing or stagnating. The market has AI fluency (can you use the tool?) but not AI intelligence (does using the tool make you measurably better?).

The Stakes. McKinsey Global Institute's November 2025 research finds that 72 percent of skills demanded by employers now operate in "shared" mode between humans and AI. By 2030, human-AI collaboration could unlock $2.9 trillion in US economic value, but only if organizations can identify, develop, and deploy people who collaborate effectively with intelligent systems. The bottleneck is measurement.

The Solution. The Human Enhancement Quotient (HEQ) provides a quantitative framework for measuring human-AI collaborative intelligence across four dimensions:

- **Cognitive Adaptive Speed (CAS):** How quickly the employee generates accurate insight when augmented by AI
- **Ethical Alignment Index (EAI):** How consistently reasoning aligns with declared ethical frameworks under uncertainty
- **Collaborative Intelligence Quotient (CIQ):** The ratio of correct trust to correct skepticism when evaluating AI outputs, measured via Reliance Calibration Score (RCS)
- **Adaptive Growth Rate (AGR):** How rapidly capability improves through repeated AI interaction

Enterprise Applications. HEQ enables four talent management use cases:

- **Pre-Employment Assessment:** Screen candidates for AI collaboration capability before hiring
- **Performance Reviews:** Track employee growth in human-AI partnership over time
- **Training Validation:** Measure whether AI education programs produce measurable enhancement
- **Adoption Readiness:** Identify which employees can lead AI transformation initiatives

The Training Target: "Trust but Verify." CIQ consistently scores lowest across all validation testing. The Reliance Calibration Score (RCS), which measures how often users correctly accept valid AI output and correctly reject AI errors, is the primary operational metric for reducing automation bias risk. Think of RCS as the "Trust but Verify" metric: it captures whether employees know *when* to trust AI and *when* to verify independently. Enterprise training programs should prioritize RCS improvement as the highest-leverage intervention.

The Foundation. HEQ emerges from fifteen years of practitioner research. The Factics methodology (2011–2024) established that structured information-to-action

frameworks increase applied human intelligence. The Factics Intelligence Dashboard (FID) operationalized measurement. HEQ extended FID into a collaborative intelligence metric with a progressive validation pathway: technical stability validation achieved 0.96 inter-model consistency across five AI platforms (ChatGPT, Claude, Gemini, Perplexity, Grok) under identical prompts; preliminary cross-user testing (n=10) demonstrated dimensional score variance within ±4 points with consistent identification of CIQ as lowest-scoring dimension; formal multi-user psychometric validation (n=100+) is underway per the 2026 Research Agenda.

The Governance Layer. HAIA-RECCLIN provides the checkpoint-based governance mechanism ensuring human authority over AI-assisted decisions. Growth OS establishes the organizational culture enabling sustainable AI adoption. Together, they create the operational infrastructure for deploying HEQ at enterprise scale.

Current Status. HEQ has completed technical stability validation and preliminary cross-user consistency testing. The framework is offered for enterprise pilots and scholarly collaboration while formal multi-user psychometric validation proceeds per the Research Agenda. Organizations adopting now become validation partners shaping the enterprise standard.

The Opportunity. Organizations that measure collaborative intelligence will identify high-potential AI collaborators, validate training investments, and build workforce capacity for the partnership economy. Those that do not will hire blind, train without proof, and lose the $2.9 trillion race.

Governance without measurement is control. Measurement without growth is stagnation. HEQ provides both.

2. The Enterprise Measurement Gap

2.1 The Problem: No Metric for Partnership Quality

Traditional intelligence measurement was designed for isolated human cognition. The theoretical expansions by Howard Gardner (multiple intelligences, 1983), Robert Sternberg (triarchic theory, 1985), Daniel Goleman (emotional intelligence, 1995), and Carol Dweck (growth mindset, 2006) broadened the concept but did not address humans thinking alongside AI systems.

Human-Systems Integration (HSI) metrics have existed for decades, but they remained engineering-focused on workload, error rates, and interface usability. What was missing: a collaborative intelligence quotient that measures the cognitive amplification of the human-AI pair, not just task performance or individual capability.

Garry Kasparov articulated the paradigm shift in *Deep Thinking* (2017). When humans played against computers alone, the machine often won. When humans worked with computers, they dominated both human-only and machine-only teams. The computer calculated possibilities. The human decided which ones mattered. This "Centaur" model established that human-AI teaming could exceed either component operating alone. The question became: how do we measure the quality of that partnership?

2.2 The Market Reality: 72% Shared Skills

McKinsey Global Institute's November 2025 analysis quantifies the transformation. Their Skill Change Index (SCI), built on 3.4 million occupation-skill mappings, classifies skills into three categories:

- **People-led:** Greater than 55 percent of time in non-automatable activities (empathy, conflict resolution, design thinking)
- **AI-led:** Greater than 55 percent in automatable activities (data entry, financial processing, pattern recognition)
- **Shared:** The middle ground where humans and AI work together

The central finding: 72 percent of skills operate in shared mode. Most knowledge work exists in a partnership zone where neither pure human effort nor pure AI automation is optimal. By 2030, optimizing that partnership could unlock $2.9 trillion in US economic value.

Labor market data confirms the urgency. Demand for AI fluency in US job postings grew sevenfold between 2023 and 2025, faster than any other skill category. Approximately eight million Americans now work in occupations where job postings require at least one AI-related skill. Simultaneously, job posting mentions are declining for routine writing and research. The message from employers: the ability to work effectively with AI is no longer optional.

By mid-2025, practitioners increasingly reported that AI augmentation materially affected their cognitive output. Empirical research began supporting these observations: Brynjolfsson, Li, and Raymond (2023) documented measurable

productivity lift with heterogeneous effects across worker skill levels in call center environments. The field moved from theoretical speculation toward quantifiable measurement.

2.3 Fluency Is Not Intelligence

MGI's terminology frames this as "AI fluency," defined as "the ability to use and manage AI tools." This framing captures adoption but not depth.

The Factics methodology distinguishes between fluency and intelligence:

- **Fluency:** Can you use the tool? Binary. You can or you cannot.
- **Intelligence:** Does using the tool make you measurably better? Developmental. It grows or it stagnates.

Organizations investing in AI training programs face a specific gap: they cannot answer whether individuals are "AI trainable" or whether their training investments worked. Completion certificates prove attendance, not capability. HEQ answers the quantitative question that fluency metrics cannot: how do we measure one's collaborative intelligence with AI, and how does it grow after use, training, or education?

FID emerged from this specific enterprise need. Factics provided a method to present materials more intelligently and to become more intelligent through structured AI engagement. FID turned AI's analytical capacity back on the user. HEQ extends FID into a metric that organizations can deploy across hiring, performance management, and training validation.

2.4 The Cultural Signals

Cultural signals amplify the empirical findings. Former Google Chief Business Officer Mo Gawdat's "borrowing intelligence" framework positions AI collaboration not as incremental efficiency but as access to cognitive capacity previously unavailable. Venture capitalist Nic Carter's April 2025 observation that refusing AI assistance now functions like "deducting 30 IQ points" crystallized the stakes for knowledge workers.

These signals measure different phenomena than productivity studies. Where Brynjolfsson et al. (2023) measure task efficiency (fact), Carter and Gawdat measure cognitive strategy (tactic). The synthesis: productivity floors are rising while the cognitive ceiling is being redefined. Organizations need metrics for both.

2.5 The Autonomous AI Counter-Thesis

HEQ assumes human-AI partnership creates value. A competing thesis holds that AI will become fully autonomous for most knowledge work, rendering partnership metrics irrelevant.

The Counter-Argument. If AI capability continues accelerating, the 72 percent "shared" skills identified by MGI may shift to 90 percent or more "AI-led" within a decade. In this scenario, HEQ measures a shrinking domain.

The Response. Even under aggressive AI autonomy projections, three factors preserve HEQ relevance:

- **Regulatory Mandate.** EU AI Act and emerging frameworks require human oversight for high-risk AI applications. Human-in-the-loop is legally required regardless of AI capability.

- **Accountability Gap.** Autonomous AI cannot be held accountable for decisions. Human checkpoint authority persists for decisions with legal, ethical, or reputational consequences.

- **Transition Period.** Even if full autonomy arrives, the transition period (estimated 5-15 years) requires partnership metrics. HEQ addresses the present and near-term, not the speculative long-term.

Monitoring Commitment. This framework will be reassessed if AI autonomy benchmarks indicate human collaboration adds negative or negligible value across task categories.

2.6 Scientific Foundation: Why Structured Oversight Matters

HEQ operationalizes findings from Harvard, MIT, and Microsoft research confirming that structured human oversight is essential for AI deployment. This is not theoretical; it is empirically validated.

The Harvard Oversight Paradox (Lin, Greenstein, & MacCormack, 2024). A field experiment with over 1,000 participants revealed that human-in-the-loop processes often degrade performance when AI provides convincing explanations. Participants were **19.4% more likely to defer to incorrect AI recommendations** when the AI used authoritative language. This creates an "Oversight Paradox" where better-sounding AI leads to worse human oversight. HEQ's CIQ dimension and RCS metric directly address this by measuring whether users can resist automation bias.

The MIT Productivity Study (Noy & Zhang, 2023). Foundational research found that generative AI decreased task time by 40% and increased output quality by 18%, specifically helping lower-skilled workers catch up. However, without governance, the quality boost plateaued. AI democratizes speed and raises the floor, but does not automatically raise the ceiling for experts without advanced workflows. HEQ measures who achieves "Cognitive Amplification" versus who merely banks time savings.

The Deloitte Cautionary Tale (CJPI, 2025). A $440,000 government report produced by Deloitte was found to contain AI-hallucinated data and fictitious case studies. Senior reviewers assumed the polished AI output implied factual accuracy, skipping source verification. This is the cost of passive oversight. HEQ's Synergy metric (S) and HAIA-RECCLIN's "Conflict Documentation" protocol exist precisely to prevent this failure mode.

The Codified Prompting Validation (Yang, Wang, & Li, 2025). Microsoft-affiliated research confirmed that structured, role-based AI interaction improves reasoning accuracy by **10-20%** and reduces token usage by over 40% compared to

conversational prompting. This validates HAIA-RECCLIN's role assignment methodology: structured interaction is empirically superior to unstructured chat.

The Cognitive Amplifier Theory (An, 2025). Empirical study showing that AI acts as a "Cognitive Amplifier" for experts, allowing them to achieve quality levels impossible for novices even with AI. Domain expertise is not obsolete; it is the fuel for the amplifier. HEQ identifies who has "Amplifier Capability" (high domain skill + high AI fluency) versus who merely uses AI as a crutch.

Implication for Enterprise. These findings converge on a single conclusion: organizations need structured measurement of human-AI collaboration quality. Passive oversight fails. Unstructured prompting underperforms. Expert amplification requires deliberate cultivation. HEQ provides the measurement framework; HAIA-RECCLIN provides the governance architecture.

3. The HEQ Framework

3.1 Four Dimensions of Collaborative Intelligence

HEQ measures human-AI collaborative intelligence across four dimensions, each capturing a distinct aspect of effective partnership. These dimensions evolved from the six-domain Factics Intelligence Dashboard (FID), condensed to reflect dynamic enhancement over time rather than static skill at a moment:

FID Domains	HEQ Dimension	What It Measures
Verbal / Linguistic	**Cognitive Adaptive Speed (CAS)**	Rate of accurate insight generation given AI-augmented working memory load
Analytical / Logical	**Ethical Alignment Index (EAI)**	Consistency of human-AI reasoning with declared ethical frameworks under uncertainty
Creative + Strategic	**Collaborative Intelligence (CIQ)**	Appropriate Reliance: ratio of correct trust (accepting valid AI) to correct skepticism (rejecting AI errors)
Emotional + Adaptive	**Adaptive Growth Rate (AGR)**	Acceleration rate of capability gain per AI interaction cycle

Table 1: FID to HEQ Domain Mapping

Dimensional Overlap. The four HEQ dimensions exhibit some interdependence, inherited from the six FID categories from which they derive. CAS improvement often correlates with AGR acceleration. EAI and CIQ interact when ethical judgment requires calibrated trust in AI outputs. This overlap reflects authentic cognitive integration rather than measurement error: real-world intelligence enhancement operates through interconnected rather than isolated capacities.

3.2 The HEQ Formula

$$\text{HEQ} = (\text{CAS} + \text{EAI} + \text{CIQ} + \text{AGR}) \div 4$$

Each dimension scores on a 0-100 scale. The HEQ score is the arithmetic mean. This simple formula enables transparent calculation while the dimensional breakdown provides diagnostic specificity. When adequate collaboration history exists (≥1,000 interactions across ≥5 domains), longitudinal evidence receives up to 70% weight, with live assessment scenarios weighted ≥30%. Precision bands reflect evidence quality and target ±2 points for decision-making applications.

3.3 Scoring Rubric

Score Range	Interpretation
0-20	Minimal capability; significant development needed before AI-integrated roles
21-40	Emerging capability; structured training recommended before deployment
41-60	Developing capability; ready for supervised AI collaboration with mentorship
61-80	Strong capability; effective independent AI collaboration; can mentor others
81-100	Expert capability; can lead AI transformation initiatives and train teams

Table 2: HEQ Scoring Interpretation for Enterprise

3.4 Synergy Validation

While HEQ measures the quality of collaboration, synergy validation ensures its utility. A high HEQ score is only meaningful if the Human+AI team outperforms the AI operating alone. Without this validation, a high score could theoretically exist even if the human introduced "cognitive clutter" that slowed a capable AI.

The Synergy Metric (S). S = HEQ_Score – AI_Baseline. If S is negative, the human did not add value; they subtracted it. In preliminary trials, HAIA-RECCLIN protocols generated positive S in the majority of cases, whereas unstructured prompting more frequently resulted in negative S.

This synergy check defends against the "human in the loop for bureaucracy's sake" critique. HAIA-RECCLIN checkpoints exist not to slow AI down but to ensure the human genuinely contributes intelligence to the collaboration.

3.4.1 Operationalizing AI_Baseline

To compute the Synergy Metric, organizations should establish **Benchmark Samples** for key roles rather than running continuous dual-execution (which would double compute costs). Quarterly audit of 5-10 representative tasks per role provides sufficient data for trend analysis:

- **Sample Audit Protocol.** Select representative tasks quarterly. Run each task through AI alone (no human collaboration) and score the output using the same rubric applied to Human+AI outputs. Compare to Human+AI performance on identical tasks.
- **Task Categories.** Establish baselines for distinct task types (research synthesis, code generation, strategic analysis, creative content). AI competence varies by domain.
- **Threshold Setting.** Define the minimum S value that justifies human involvement. For high-stakes decisions, S > 5 may be required. For routine tasks, S > 0 may suffice.
- **Negative S Response.** If S is consistently negative for a user-task combination, options include: (a) additional CIQ training, (b) restructuring the collaboration protocol, (c) removing the human checkpoint for that task category.

Enterprise Application. Synergy validation answers the "human in the loop for bureaucracy's sake" critique. Checkpoints are justified when S > 0; they are questioned when S ≤ 0. Sample audits make this measurement economically feasible.

3.5 The CIQ Challenge: Appropriate Reliance

Across validation trials, CIQ consistently scored lowest (85-91 range compared to 88-96 for other dimensions). Users struggled with Appropriate Reliance, either over-trusting AI creative outputs or under-trusting AI data analysis. This aligns with "Jagged Frontier" research (Dell'Acqua et al., 2023), which shows that AI competence varies unpredictably across tasks.

The low CIQ scores indicate that users did not consistently know which tasks were inside or outside the AI's competence zone, leading to hesitation or redundancy. This validates the need for the **Reliance Calibration Score (RCS)**: a sub-metric tracking how often the human correctly intervenes when the AI is wrong.

HAIA-RECCLIN's Checkpoint-Based Governance is positioned as the solution. By forcing explicit approval of Confidence and Sources at each checkpoint, the system trains users to calibrate trust appropriately. For enterprise deployment, CIQ represents both the highest-value training target and the clearest differentiator between employees who collaborate effectively with AI and those who merely use it.

3.6 The Critical Discovery: Measurement Produces Growth

The most significant finding was unexpected: measurement itself produced growth. Users who engaged with HEQ assessment showed improved performance in subsequent interactions. The act of structured evaluation catalyzed enhancement rather than merely recording it.

This discovery demanded a conceptual shift. The Factics Intelligence Dashboard asked: *How do you think?* The question needed to become: *How far can you grow?*

3.6.1 Mechanism Hypothesis

Three factors may explain why HEQ assessment produces growth:

- **Metacognitive Activation.** The structured output format (Role, Task, Sources, Conflicts, Confidence, Decision) forces users to articulate their reasoning process. Articulation builds awareness; awareness enables improvement.
- **Dimensional Feedback.** Receiving scores across four dimensions identifies specific weaknesses. Users naturally focus subsequent effort on lowest-scoring areas (typically CIQ).
- **Checkpoint Habituation.** Repeated exposure to governance checkpoints trains the cognitive pattern of pausing, evaluating, and deciding rather than accepting AI output passively.

Testable Prediction. If this mechanism hypothesis is correct, users who receive dimensional feedback will show greater improvement than users who receive only composite scores. This will be tested in the 2026 validation studies.

Testable Hypothesis. Formal design details for mechanism testing are documented in the Research Agenda (Section 9.2). The proposed validation employs a minimal longitudinal design with baseline, intervention (HAIA-RECCLIN structured prompting), and follow-up windows. Control comparison: unguided prompting versus structured prompting. Success criterion: statistically and practically meaningful lift in at least two dimensions within a defined window, with effect sizes reported.

4. Enterprise Applications: HEQ as Supporting Evidence

Origin Context. HEQ's research origins examined how organizations were making workforce decisions about "AI trainability" without any objective methodology. When companies like Microsoft publicly referenced AI skill gaps in workforce restructuring, no standardized assessment existed to define what "AI trainable" meant or to give employees a fair opportunity to demonstrate or develop capability. HEQ emerged to provide **structured methodology where none existed**, not to rationalize arbitrary decisions but to **prevent them**.

Positioning. HEQ is designed as **a factor, not THE factor** in workforce decisions. Scores inform development conversations, identify training needs, and provide one data point among many in talent management. HEQ scores must never serve as sole or primary determinant in hiring, promotion, or termination decisions. The four applications below describe how HEQ can **support** organizational decision-making, not replace human judgment.

4.1 Pre-Employment Capability Assessment

The Problem. Hiring managers currently assess AI collaboration capability through unstructured interviews, resume keywords, or gut feeling. This creates both false positives (hiring candidates who cannot collaborate effectively) and false negatives (rejecting candidates who could excel with development). No methodology exists to provide candidates fair opportunity to demonstrate capability.

HEQ as Supporting Evidence. Administer the HAIA Intelligence Snapshot as one component of a multi-factor hiring process. HEQ scores provide:

- Dimensional data on specific strengths and development areas
- CIQ score indicating trust calibration capability
- Synergy metric suggesting whether candidate adds value to AI collaboration
- Baseline for onboarding development planning if hired

Critical Boundary. HEQ scores must be considered alongside interviews, work samples, references, and other established hiring criteria. A low HEQ score alone does not justify rejection; it identifies a development area. A high HEQ score alone does not guarantee success; it indicates one dimension of capability. Organizations using HEQ in hiring must document that it serves as supporting evidence, not primary determinant.

Implementation. The assessment integrates into existing hiring workflows. Candidates receive assessment instructions, complete the HAIA Intelligence Snapshot (approximately 45-60 minutes), and results are delivered to hiring managers within 24 hours alongside other candidate evaluation data.

4.2 Performance Reviews During Employment

The Problem. Traditional performance reviews measure task completion and behavioral competencies but cannot track whether an employee's AI collaboration

capability is improving, stagnating, or declining. When organizations make decisions about "AI readiness," they lack objective longitudinal data.

HEQ as Supporting Evidence. Conduct HEQ assessments at regular intervals (quarterly or semi-annually) to track growth trajectory. The Adaptive Growth Rate (AGR) dimension specifically measures capability improvement over time. Managers receive:

- Longitudinal HEQ trend showing growth or stagnation
- Dimensional comparison identifying which capabilities are developing
- Benchmark comparison against team and organizational averages
- Specific development recommendations based on lowest-scoring dimensions

Critical Boundary. HEQ trends inform development planning conversations, not performance ratings or compensation decisions. A declining HEQ score triggers a development intervention, not a performance penalty. An improving HEQ score indicates readiness for expanded AI-integrated responsibilities, not automatic promotion. HEQ data supplements traditional performance metrics; it does not replace them.

Implementation. HEQ assessment becomes a component of the performance review cycle focused on **development planning**, not performance evaluation. Results feed into coaching conversations and training recommendations.

4.3 Training Program Validation

The Problem. Organizations invest in AI training programs but cannot prove they work. Completion rates measure attendance. Satisfaction surveys measure enjoyment. Neither measures capability change. Without objective measurement, training budgets are justified by faith, not evidence.

HEQ as Supporting Evidence. Conduct HEQ assessments before and after training programs to measure actual capability enhancement. Training ROI becomes quantifiable:

- Pre-training baseline HEQ establishes starting capability
- Post-training HEQ measures actual enhancement
- Delta HEQ (post minus pre) quantifies training effectiveness at group level
- Dimensional analysis identifies which aspects of training produced results

Critical Boundary. Training validation uses **group-level** Delta HEQ to assess program effectiveness, not individual scores to evaluate trainees. A participant with low post-training HEQ is not "failing"; the training program may need refinement for their learning style. Individual scores inform personalized follow-up coaching, not training completion status.

Implementation. Training programs incorporate HEQ assessment as a program evaluation component. Program designers use aggregate dimensional results to refine curriculum. Finance teams use group Delta HEQ to calculate training ROI.

Programs that produce statistically significant group-level HEQ improvement justify continued investment.

4.4 AI Adoption Readiness Assessment

The Problem. Organizations planning AI transformation cannot identify which employees are ready to lead adoption initiatives. Technical skills are visible; collaborative intelligence is not.

HEQ as Supporting Evidence. Assess the current workforce to create an AI Adoption Readiness Map for **workforce planning purposes**. Aggregate HEQ scores provide organizational visibility into readiness distribution:

HEQ Range	Readiness Category	Development Recommendation
81-100	**AI Champions**	Invite to lead transformation initiatives; train others; pilot new AI tools
61-80	Early Adopters	Offer AI-integrated project opportunities; mentor developing employees
41-60	Developing	Prioritize for AI collaboration training; provide supervised AI project experience
21-40	Foundational	Enroll in basic AI literacy training; identify learning support needs
0-20	Pre-Foundational	Assess barriers to AI collaboration; provide individualized development support

Table 3: AI Adoption Readiness Categories

Critical Boundary. Readiness categories inform **development planning and training resource allocation**, not employment status. A "Pre-Foundational" score identifies a development need, not a termination candidate. The table above shows **development recommendations**, not employment actions. Organizations must not use readiness categories to justify layoffs, demotions, or punitive actions.

False Negative Warning. Low scores may indicate **prompting literacy gaps** rather than cognitive capability gaps. An experienced professional who struggles with AI interface conventions (typing speed, prompt structure, chat-based interaction) may score poorly despite strong domain judgment. Remediation must assess interface training needs before capability conclusions. This is particularly relevant for preventing age discrimination claims where older workers may score low due to unfamiliarity with chat interfaces, not lack of analytical capability.

Implementation. Conduct organization-wide HEQ assessment before major AI initiatives. Results inform deployment sequencing (invite AI Champions to lead), training resource allocation (prioritize Developing category), and change management strategy (leverage Early Adopters as peer coaches).

4.5 Ethical Safeguards

HEQ is designed as a development tool, not a punishment mechanism. Ethical deployment requires:

- **Prohibited Uses:** HEQ scores shall not be used as the sole or primary factor in hiring, firing, or promotion decisions. Scores inform development conversations; they do not determine employment outcomes.

- **Required Disclosure:** Any organizational deployment must inform participants that HEQ assessment is occurring and explain how results will be used.

- **Appeal Process:** Employees who believe their HEQ score does not reflect their capability must have access to reassessment.

- **Context Consideration:** Scores must be interpreted alongside other performance data, not in isolation.

4.6 CIQ Development: Training Appropriate Reliance

CIQ (Collaborative Intelligence Quotient) measures the ratio of correct trust to correct skepticism. Across validation trials, CIQ scored lowest (85-91 vs. 88-96 for other dimensions). This section provides a training methodology.

4.6.1 The CIQ Training Protocol

- **Bait Injection.** Training sessions include AI outputs with intentional errors (plausible but incorrect facts, flawed calculations, biased recommendations). Trainees must identify errors before proceeding.

- **Trust Calibration Journaling.** After each AI interaction, trainees record: (a) their confidence in the AI output before verification, (b) actual accuracy after verification, (c) the delta between expected and actual. Over time, this calibrates intuition.

- **Domain Mapping.** Trainees map AI competence zones for their specific work domain. Where does the AI reliably excel? Where does it reliably fail? Where is performance unpredictable (the "Jagged Frontier")?

- **Reliance Calibration Score (RCS).** A sub-metric tracking: (correct acceptances + correct rejections) ÷ total decisions. Target RCS > 0.85 before deployment in high-stakes AI collaboration roles.

Implementation. A 2-hour CIQ calibration workshop using bait protocols can be integrated into existing AI training programs. Pre/post RCS measurement validates workshop effectiveness. Organizations should prioritize CIQ training for employees in the 41-80 HEQ range, where the marginal return on training investment is highest.

4.7 Enterprise Pilot Scorecard

The following metrics constitute the minimum measurement set for enterprise HEQ pilots. This scorecard packages the operational metrics referenced throughout this paper into a single implementation artifact.

Metric	Definition	Target
Baseline HEQ	Pre-intervention composite score (0-100)	Establish within first week
Post HEQ	Post-intervention composite score (0-100)	Measure at 30, 60, 90 days
Delta HEQ	Post minus Baseline score	> +5 points for training ROI
Synergy (S)	HEQ_Score minus AI_Baseline	S > 0 (human adds value)
RCS	Reliance Calibration Score	> 0.85 for high-stakes roles
Audit Trail %	Decisions with full governance documentation	100% for regulated workflows
Override Rate	Human corrections of AI recommendations	Track trend (not target)
Checkpoint Time	Average minutes per governance checkpoint	Track efficiency over time

Table 4: Enterprise Pilot Scorecard

Pilot Verification of Business Impact. The $2.9 trillion opportunity cited from McKinsey Global Institute provides market context but does not guarantee organizational results. Enterprise pilots must independently verify business impact using internal metrics: Delta HEQ distribution across participants, S positive rate (percentage of users who add value versus subtract it), RCS improvement trajectory, and correlation between HEQ gains and measurable business outcomes (error reduction, cycle time, decision quality). Pilot reports should document both HEQ metrics and business KPIs to establish local validity before scaling.

5. HEQ5: Organizational Deployment

5.1 Why the Fifth Dimension

The four-dimension HEQ model measures individual human-AI collaborative intelligence. Organizations operate at a different scale with different obligations. Individual enhancement is necessary but insufficient for enterprise deployment.

HEQ5 extends the framework by adding a fifth dimension: **Societal Safety.** This dimension measures how human-AI collaboration impacts stakeholders beyond the immediate user, including affected communities, downstream systems, and societal structures.

The original four-dimension HEQ remains valid for individual assessment. HEQ5 extends this framework for organizational contexts where societal impact measurement is required. Practitioners may use HEQ for personal cognitive development tracking and HEQ5 for enterprise deployment, regulatory compliance, and stakeholder accountability.

5.2 The Five Dimensions

Dimension	Focus	Indicator
Cognitive Adaptive Speed (CAS)	Idea connection and analysis velocity	% improvement in solution speed
Ethical Alignment Index (EAI)	Moral reasoning and policy consistency	% ethical agreement across AIs
Collaborative Intelligence Quotient (CIQ)	Human-AI team synergy	Reliance Calibration Score from multi-AI sessions
Adaptive Growth Rate (AGR)	Learning from feedback and iteration	% faster adoption per cycle
Societal Safety (SS)	Impact on stakeholders, systems, and society	Risk-adjusted deployment score; audit compliance rate

Table 5: HEQ5 Five Dimensions for Enterprise

$$\text{HEQ5} = (\text{CAS} + \text{EAI} + \text{CIQ} + \text{AGR} + \text{SS}) \div 5$$

5.3 Societal Safety: Operationalization

Societal Safety measures outcome impacts at organizational scale, complementing Ethical Alignment's focus on process fairness. Components include:

- **Stakeholder Impact Assessment:** Documented analysis of effects on affected communities

- **Downstream System Risk:** Evaluation of cascade effects in interconnected systems

- **Regulatory Compliance:** Alignment with EU AI Act, ISO 42001, NIST AI RMF requirements

- **Audit Trail Completeness:** Percentage of decisions with full governance documentation

5.3.1 Automation, Displacement, and the Growth Imperative

McKinsey Global Institute (2025) projects that current AI technologies could "theoretically automate more than half of current US work hours," while emphasizing this is "not a forecast of job losses." It is, however, **a forecast of task migration**. The distinction matters. Technical automation potential describes what AI can do. Actual displacement describes organizational choices about what AI should do. While jobs may remain stable, the *work within those jobs* shifts fundamentally.

HEQ's governance framework ensures human value-add in contexts where automation is technically possible but partnership is preferable. The Societal Safety dimension explicitly measures whether AI deployment creates stakeholder harm, including workforce displacement without transition support. Organizations using HEQ commit to measuring not just efficiency gains but distributional effects.

More fundamentally, the Growth Operating System philosophy that underlies HEQ reframes the automation question entirely. Organizations driven by Growth OS respond to AI capability not by downsizing but by adding and growing: expanding what humans can accomplish, creating new roles that use augmented capability, and developing workforce capacity that did not previously exist. Displacements met with adding and growing will supersede those that default to reduction.

This reflects a governance commitment, not an empirical prediction. Organizations using HEQ5 commit to monitoring Societal Safety by tracking whether AI deployment is accompanied by workforce development investment. This is a measurement requirement ensuring transparency about distributional effects, not a guarantee of outcomes.

The Growth Philosophy. HEQ is designed for organizations that treat AI as an amplifier of human capability (growth trajectory) rather than a replacement mechanism (extraction trajectory). The Societal Safety metric distinguishes between these approaches: automation without growth is extraction; automation with growth is transformation. Organizations pursuing extraction models should use different governance frameworks; HEQ measures whether the growth choice is made deliberately, with documented assessment of effects on affected communities.

5.4 Regulatory Alignment

HEQ5 artifacts map to emerging regulatory requirements:

Obligation	HAIA-RECCLIN Artifact	CBG Checkpoint	HEQ5 Metric
Traceability (EU AI Act Art. 12)	Role assignment log; Source citations	Record phase documentation	Audit trail completeness %
Risk Management (NIST AI RMF)	Conflict documentation; Confidence scoring	Arbitrate phase review	SS risk-adjusted score
Human Oversight (ISO 42001)	Navigator role; Decision point	Mandatory approval gates	Human override rate
Post-Market Monitoring	Expiry notation; AGR tracking	Continuous monitoring alerts	GR trajectory over time

Table 6: Regulatory Compliance Mapping

5.5 The Complete Architecture

Layer	Function	Outcome
Growth OS	Defines culture and governance rhythm	Sustainable, transparent AI adoption
HAIA-RECCLIN	Distributes cognitive tasks under human oversight	Auditable, bias-balanced output
HEQ / HEQ5	Measures human-AI enhancement performance	Quantified intelligence growth (0-100 scale)

Table 7: The Complete Architecture

Governance absent measurement becomes opacity. Measurement absent culture becomes surveillance. HEQ5 provides both, anchored in human cognitive sovereignty.

6. Governance Architecture

HEQ measurement operates within a governance infrastructure ensuring human authority, auditability, and continuous improvement. This section documents the operational framework enabling enterprise deployment.

6.1 Growth OS: The Cultural Foundation

Growth OS is the organizational and cultural foundation for scalable AI governance. It integrates people, data, and process under one rhythm so that growth, not efficiency, becomes the measure of success.

The three pillars of Growth OS:

- **Trust and Transparency:** Visible governance, ethical escalation, human decision checkpoints
- **Rhythm and Culture:** Iterative feedback loops that compound learning and adaptability
- **Outcome Anchoring:** ROI reframed around growth metrics such as revenue per employee and customer lifetime value

Growth OS is the environment in which governance and measurement operate. Without it, frameworks remain theoretical. With it, they become operational.

6.2 HAIA-RECCLIN: The Governance Mechanism

HAIA-RECCLIN stands for Human Artificial Intelligence Assistant with Roles: Researcher, Editor, Coder, Calculator, Liaison, Ideator, Navigator. This framework represents a structured multi-AI collaboration approach where AI systems serve as assistants to human authority across distinct functional domains with distributed checkpoint authority. To learn more about HAIA-RECCLIN implementation and resources, visit basilpuglisi.com/haia-recclin.

Role	Function and Checkpoint Authority
Researcher	Gathers evidence, validates sources, identifies conflicts between claims, flags unverified assertions
Editor	Synthesizes conflicting information, preserves dissenting viewpoints, maintains narrative coherence
Coder	Implements technical solutions, validates functionality, documents implementation decisions
Calculator	Quantifies risks and returns, validates statistical claims, ensures mathematical rigor
Liaison	Coordinates across organizational boundaries, manages stakeholder communication
Ideator	Proposes alternative approaches, challenges assumptions, expands solution space
Navigator	Guides decision makers through tradeoffs, maintains process integrity, coordinates role interactions

Table 8: HAIA-RECCLIN Seven Roles

External Validation: The People-Led, AI-Led, Shared Taxonomy.
McKinsey Global Institute's skill classification provides independent validation for RECCLIN's role-based governance architecture. Their taxonomy distinguishes people-led skills (including interpersonal conflict resolution, design thinking, and empathy-dependent judgment) which remain challenging for machines to replicate. In RECCLIN terms, these map to the Navigator and Liaison roles, where human contextual understanding and stakeholder relationships cannot be delegated.

AI-led skills include data entry, financial processing, and pattern recognition at scale, where people step back from hands-on work to focus on design, validation, and exception handling. In RECCLIN terms, these map to Researcher and Calculator roles when operating in high-automation mode.

Shared skills constitute the 72 percent middle ground where "machines handle routine tasks while people frame problems, provide guidance to AI agents and robots, interpret results, and make decisions." This describes precisely the collaboration that RECCLIN's checkpoint architecture is designed to govern.

The MGI research confirms that role assignment is not arbitrary preference but reflects genuine differences in where human judgment adds value. RECCLIN operationalizes what MGI describes theoretically: a structured method for determining which aspects of a task belong in people-led, AI-led, or shared execution modes.

6.3 The Governance Cycle

HAIA-RECCLIN operates through a four-phase governance cycle:

- **Initiate:** Assign roles and intent. Human defines the scope and objective.
- **Collaborate:** Cross-AI dialogue generates balanced insight. Multiple platforms analyze from different perspectives.
- **Arbitrate:** Human oversight approves, redirects, or rejects outputs. No decision executes without human authorization.
- **Record:** Every reasoning step becomes part of an ethical audit log. Documentation enables forensic analysis.

6.4 Learning Through Prompting

The critical insight of HAIA-RECCLIN is that **custom prompts make users learn as they prompt.** The structured output format forces cognitive engagement:

- **Role:** Stating the assigned role first creates immediate task clarity
- **Task:** Explicit understanding of the request prevents misalignment
- **Sources:** APA-style citations with verification requirements build evidence awareness
- **Conflicts:** Documenting dissent rather than forcing consensus preserves intellectual honesty
- **Confidence:** 0-100% scoring with justification calibrates certainty against evidence

- **Expiry:** Time-sensitivity notation prevents stale information from persisting
- **Fact to Tactic to KPI:** The Factics loop embedded in every response
- **Decision:** Human arbitration point with explicit recommendation plus alternatives

This structure mirrors findings in educational Hybrid Intelligence research showing that active co-regulation of AI (rather than passive acceptance) is the key driver of learning gains.

6.5 Checkpoint-Based Governance (CBG)

Checkpoint-Based Governance provides the protocol-driven framework for structuring human-AI collaboration through mandatory decision points. The core architectural principles:

Human Authority Preservation. Humans retain final decision rights at defined checkpoints. AI systems contribute intelligence but never execute decisions autonomously.

Systematic Evaluation. Decision points apply predefined criteria consistently, preventing ad hoc judgment and supporting reproducible oversight.

Documented Arbitration. Every checkpoint decision generates a record including the input evaluated, criteria applied, decision rendered, and responsible party identified.

Trust Calibration Training. Each checkpoint functions as a "forced calibration" event. By requiring the user to explicitly approve the Confidence score and Sources before proceeding, the system combats Automation Bias (mindless acceptance) and trains the user to distinguish between high-competence and low-competence AI outputs. This addresses the "Jagged Frontier" challenge by institutionalizing productive friction.

Continuous Monitoring. The framework includes mechanisms for detecting automation bias drift (humans defaulting to AI recommendations without genuine review) and model performance degradation.

CBG formally links to Shneiderman's Human Control frameworks and Guszcza's Friction metric. Checkpoints represent institutionalized, productive friction that prevents automation bias and ensures accountability. The system accepts efficiency costs in exchange for traceable responsibility.

Workflow Redesign, Not Task Automation. McKinsey's central finding challenges conventional AI deployment: "Integrating AI will not be a simple technology rollout but a reimagining of work itself." Organizations treating AI as task automation (which steps can AI do instead of humans?) capture a fraction of the potential value. Those redesigning workflows around human-AI partnership (how should humans and AI work together so combined output exceeds what either produces alone?) capture the full $2.9 trillion opportunity.

CBG is workflow redesign methodology. Each checkpoint restructures the interaction between human and AI from passive consumption to active collaboration. Task automation asks: "Which steps can AI do instead?" Workflow redesign asks: "How should humans and AI work together for combined output exceeding either alone?" CBG answers the second question.

7. Validation Evidence

7.1 Multi-AI Validation Results

HEQ was tested across five independent AI architectures: ChatGPT, Claude, Gemini, Perplexity, and Grok. The five-AI comparison showed notable consistency with a mean of 91.4 ± 2.9 and inter-model output agreement of 0.96 under identical prompts. See Methods Appendix for computation details and limitations.

Formal Validation (n=1). The formal validation involved single-user testing with the framework author, establishing system consistency across AI platforms.

Informal Cross-User Testing (n=10). Beyond the formal study, the HAIA Intelligence Snapshot prompt was tested by 10 additional individuals using their own ChatGPT-4 accounts. All participants returned data consistent with expectations, indicating prompt stability across different users, contexts, and interaction styles. While this does not constitute formal psychometric validation, it provides preliminary evidence of cross-user consistency.

7.2 Case Study 001: Thought Leader Validation

The HAIA Intelligence Snapshot prompt (v2.0 and v3.0 iterations) was administered to five AI platforms simultaneously. Each platform evaluated the same user work sample and produced independent HEQ scores.

Convergent Findings:

- Despite differences in model architecture and training, composite scores converged within a narrow band (89-94)
- All models identified strong systems thinking, ethical grounding, and adaptive learning
- CAS, EAI, and AGR consistently rated high across platforms
- CIQ (Collaborative Intelligence) rated lowest across all platforms
- Claude's conservative scoring offset Gemini's optimism, validating triangulation as a bias buffer

This shows that the snapshot measures consistent cognitive and behavioral signals rather than isolated model-specific artifacts.

Model-Specific Contributions:

- **ChatGPT:** Emphasized meta-cognition and scalability; executive-ready output format
- **Gemini:** Formalized the three-step process (Co-creation, Context Injection, Scoring)
- **Perplexity:** Introduced trust-building transparency language and disclosure requirements
- **Grok:** Added confidence bands (± ranges) and directional disclaimers

- **Claude:** Delivered enterprise-grade scoring rubric with weighted RCI and structured reporting

7.3 Meta-Insight: HAIA as a Living System

The case study confirmed that HAIA itself is a living system. It evolved in real time using collective AI intelligence, showing its own premise that structured collaboration yields superior outcomes.

The significant finding: HAIA excels at AI-to-AI collaboration but must now focus on measuring and enabling human-to-human collaboration to complete the loop. This informs the CIQ enhancement priority in the research agenda.

7.4 Cultural and Demographic Limitations

The HEQ framework was developed and tested on Western AI models (ChatGPT, Claude, Gemini, Perplexity, Grok) using English-language prompts with a Western user population. This reflects potential WEIRD (Western, Educated, Industrialized, Rich, Democratic) bias in both AI training data and assessment design.

Explicit Restrictions. Until cross-cultural validation is complete, the following restrictions apply:

- HEQ scores should be interpreted with explicit caution outside Western organizational contexts
- Organizations deploying HEQ in non-Western markets must disclose validation limitations to all participants
- Cross-cultural deployment requires local validation pilot before scaling
- The Societal Safety (SS) dimension is particularly sensitive to cultural variation and should not be scored in non-Western contexts without local adaptation

Validation Requirement. Cross-cultural validity testing on non-Western AI platforms (e.g., Ernie Bot, YandexGPT, Jais) and diverse user populations is a **gating requirement for global deployment**, not an optional enhancement. See Research Agenda Section 9.2 for specific validation studies.

7.5 Validation Limitations and Mitigation

The current validation has three boundaries that enterprise adopters must understand:

Boundary 1: Sample Size. HEQ validation follows a progressive pathway: technical stability validation (0.96 inter-model consistency) is complete; preliminary cross-user consistency testing (n=10, ±4 point variance) establishes prompt reliability; formal multi-user psychometric validation (n=100+) is underway. Enterprise pilots must generate multi-user data before scaling deployment.

Boundary 2: Inter-Model Agreement ≠ Construct Validity. High agreement across AI platforms (0.96 coefficient) may reflect consistent prompt interpretation rather than valid measurement of underlying cognitive constructs. The prompt

produces stable outputs; whether those outputs measure what they claim to measure requires independent psychometric validation.

Boundary 3: Western Platform Bias. Validation occurred on ChatGPT, Claude, Gemini, Perplexity, and Grok. Cross-cultural validity on non-Western platforms (Ernie Bot, YandexGPT) and diverse user populations is untested.

Mitigation Pathway. Enterprise pilots should treat HEQ scores as directional indicators during the 2026 validation period, not as deterministic metrics. Scores inform development conversations; they do not determine employment outcomes (see Section 4.5 Ethical Safeguards). Formal multi-user validation studies are proposed in the Research Agenda.

8. Intellectual Foundation

HEQ emerges from fifteen years of practitioner research. This section documents the intellectual lineage establishing credibility and theoretical grounding.

8.1 Teachers NOT Speakers (2011-2012)

The foundational insight came from observing professional conferences. The standard model positioned speakers as authorities transmitting knowledge to passive audiences. The result was entertainment without implementation.

In February 2011, during a Social Media Club presentation at Social Media Week, a different principle was articulated: *"Everyone who attends is a participant and hopefully a teacher too."* This was the seed of a methodology that would evolve over the next decade into a comprehensive framework for intelligence enhancement.

By February 2012, this principle had been formalized into the Teachers NOT Speakers philosophy at Social Media Action Camp (#SMWsmac) during Social Media Week NYC. The approach required hands-on sessions focused on active implementation rather than passive listening, presenters who functioned as educators rather than self-promoters, and participants who left with executable knowledge.

8.2 Auditable Architecture and Digital Factics (2012)

The October 2012 NYXPO event at the Javits Center introduced a critical concept: **auditable architecture.** Every professional on the Social Media Track was required to provide a syllabus-style takeaway. Sessions had to deliver a roadmap that attendees could use to implement digital strategies immediately.

This concept, that learning should be traceable, verifiable, and executable, became the structural foundation for Factics. It foreshadows the HAIA-RECCLIN governance framework's emphasis on auditable processes and human oversight.

On November 27, 2012, *Digital Factics: Twitter* was published through Digital Media Press (magcloud.com/browse/issue/471388). This 58-page publication documented the methodology for the first time, establishing Factics as a formalized framework with published provenance.

8.3 The Factics Methodology

Over the following years, Factics evolved from event methodology into a comprehensive framework for converting information into action. The core formula:

$$\text{Facts + Tactics + KPIs = Factics}$$

The complete Factics loop requires:

- **Facts and Data:** Verified evidence, not opinion or assumption
- **Tactics and Strategy:** Executable actions, not recommendations or suggestions

- **Outcomes and Goals:** Defined success states, not vague aspirations
- **KPIs:** Measurable indicators that convert belief into testable commitment

This loop forces clarity, testability, and accountability. Content without this structure becomes entertainment instead of value delivery. Policy analysis lacking actionable tactics generates awareness without capability.

8.4 The Intelligence Enhancement Thesis (February 2024)

By February 2024, over a decade of practice had produced an observation that demanded explicit articulation. A blog post on basilpuglisi.com made the thesis clear:

> *"The argument I am putting forward is that Factics increases applied human intelligence. This is not a proven fact. It is a defined position offered for validation."*

The post defined intelligence not as an innate trait or a score, but as *"the applied ability to reason clearly, decide effectively, learn faster, and predict consequences under uncertainty."*

The Internal Pathway. When Factics becomes a personal operating standard, claims attach to facts, recommendations declare tactics, and tactics declare KPIs. The mind stops optimizing for persuasion and starts optimizing for accuracy.

The External Pathway. When people consume content built inside the same loop, the content functions as cognitive scaffolding. It trains how to move from evidence to action to measurement.

8.5 Complete Timeline

Date	Milestone	Contribution
Feb 2011	SMW Presentation	"Everyone is a participant and teacher" peer learning model
Feb 2012	#SMWsmac NYC	Teachers NOT Speakers philosophy formalized
Oct 2012	NYXPO Javits	Auditable Architecture concept introduced
Nov 2012	Digital Factics: Twitter	First publication documenting Factics methodology
2012-2023	Methodology Refinement	Facts + Tactics + KPIs formalized through consulting practice
Feb 2024	**Intelligence Thesis**	**"Factics increases applied human intelligence" declared**
2024	FID Development	Six-domain measurement operationalizes the thesis
Sept 2025	HEQ v1.0 Published	Evolution from measurement to enhancement trajectory
Q2/Q3 2025	HEQ Prompts Published	Prompt instruments released for enterprise evaluation workflows
Q2/Q3 2025	HEQ Prompts Published	Prompt instruments released for enterprise evaluation workflows
Feb 2025	AIQ Framework	Independent academic validation of convergent architecture
Dec 2025	**Cross-User Testing**	**Informal n=10 prompt consistency validation via ChatGPT-4**

Table 9: Complete Factics to HEQ Timeline (2011-2025)

8.6 The Factics Intelligence Dashboard (FID)

The Factics Intelligence Dashboard operationalized the February 2024 thesis. If Factics methodology increases applied human intelligence, that increase should be measurable. FID provided the measurement instrument with six domains:

Domain	Definition
Verbal / Linguistic	Clarity, adaptability, and persuasion in communication
Analytical / Logical	Reasoning, structure, and problem-solving accuracy
Creative	Originality, ideation, and practical innovation
Strategic	Foresight, goal alignment, and systems thinking
Emotional / Social	Empathy, leadership, and audience awareness
Adaptive Learning	Ability to integrate new tools, data, and systems efficiently

Table 10: FID Six-Domain Model

8.7 Scientific Context: Related Work

HEQ enters a field with significant prior work. Understanding this lineage positions the framework accurately:

Foundational Frameworks. Shneiderman's Human-Centered AI established levels of automation and control transitions. Guszcza's "team member" concept introduced metrics including Friction (productive delay that enables oversight), Variance (consistency across similar decisions), and Bias Differential (systematic error introduced by human-AI vs. human-alone). Lee's taxonomy categorized collaboration styles and process metrics.

Contemporary Research. Stanford HAI, MIT, and Microsoft Research have developed frameworks including the Team Effectiveness Assessment (TEA) measuring team process effectiveness. Bansal et al. (2019) formalized "Appropriate Reliance" as the balance between accepting valid AI output and rejecting AI errors. The "Jagged Frontier" research (Dell'Acqua et al., 2023) showed that AI competence varies unpredictably across tasks, making human judgment about when to trust AI critical.

Convergent Development: Independent Confirmation. The theoretical necessity of collaborative intelligence measurement was confirmed when the AIQ framework (Ganuthula & Balaraman, arXiv 2025) emerged independently, with formal publication occurring shortly after HEQ's initial articulation. AIQ proposed eight dimensions for measuring human-AI collaborative intelligence: Strategic AI Understanding, Prompt Engineering Intelligence, Critical Evaluation Capability, Integration Intelligence, Adaptive Learning Capability, Ethical Judgment, Context Sensitivity, and Creative Synthesis. This convergent evolution across independent researchers validates the underlying construct: human-AI collaboration quality is measurable, and the field recognizes this need simultaneously from multiple directions. HEQ and AIQ represent complementary approaches with different emphases: HEQ uniquely integrates governance architecture and organizational deployment; AIQ emphasizes psychometric breadth and dimensional comprehensiveness.

Comparative Framework Positioning

Dimension	HEQ	AIQ	Other Frameworks
Construct breadth	4 core dimensions (+1 organizational)	8 dimensions (broader)	2–3 typically
Governance integration	✓ HAIA-RECCLIN + CBG	Not specified	Rarely
Organizational deployment	✓ Growth OS	Not specified	Rarely
Validation status	Progressive pathway (0.96 + n=10)	Theoretical (arXiv preprint)	Varies
Unique strength	**Integrated governance**	**Dimensional depth**	**Varies by tool**

Table: Comparative Framework Positioning (HEQ vs AIQ vs Other Frameworks)

Workforce-Level Measurement: The McKinsey Skill Change Index.
McKinsey Global Institute (2025) developed the Skill Change Index (SCI), a time-weighted measure of automation's potential impact on skills used in today's workforce. Their methodology integrates four inputs: employment data across approximately 800 occupations from the Bureau of Labor Statistics, detailed work activities (DWAs) from O*NET, roughly 34,000 skills linked to occupations from Lightcast, and McKinsey's proprietary automation adoption model. Using GPT-4o to map 3.4 million occupation-DWA-skill combinations, they classified skills into three categories: people-led (greater than 55 percent of time in non-automatable activities), AI-led (greater than 55 percent in automatable activities), and shared (the middle ground).

Their central finding, that 72 percent of skills operate in shared mode, validates the theoretical premise underlying HEQ: most knowledge work exists in a partnership zone where neither pure human effort nor pure AI automation is optimal. The question becomes how to optimize that partnership.

The SCI and HEQ operate at different scales and answer different questions. SCI measures skill exposure at the occupation-cluster level using labor market data. HEQ measures partnership quality at the individual-session level using structured assessment protocols. These scales are complementary: SCI identifies which skills will require partnership; HEQ measures whether the partnership is working for a specific person in a specific context.

HEQ's Distinct Contribution. Where other frameworks measure collaboration quality, HEQ/HEQ5 uniquely embeds governance (Checkpoint-Based Governance) and organizational culture (Growth OS) as co-equal dimensions. HEQ measures the enhancement trajectory of the individual within a governed system, not just team process effectiveness or task performance.

9. The 2026 Research Agenda

9.1 Current Status: What Is In Action

The framework has completed technical stability validation and preliminary cross-user consistency testing. The following activity is currently underway:

Third-Party Prompt Evaluations. Volunteer participants are testing HAIA-RECCLIN prompts in their own workflows. This preliminary testing gathers qualitative feedback on prompt usability, structured output clarity, and perceived cognitive engagement. Results will inform protocol refinement before formal multi-user validation.

Informal Cross-User Consistency. Ten individuals have tested the HAIA Intelligence Snapshot prompt using their personal ChatGPT-4 accounts. All returned data consistent with expectations, providing preliminary evidence of prompt stability across different users, contexts, and interaction styles.

9.2 Proposed Priorities (Not Yet In Action)

The following research priorities are proposed but not yet implemented:

Multi-User Validation. Expansion from current testing to formal multi-user validation (n=30+) across demographic contexts. This requires controlled experiments comparing task outcomes for groups using HAIA-RECCLIN versus standard prompting versus AI-only. Process metrics will include communication overhead (time spent prompting/refining), cognitive load (via NASA-TLX surveys), and trust calibration (pre/post-task trust scales).

Memory-Enabled Longitudinal Protocols. Platforms with persistent memory (OpenAI memory controls, Claude memory) enable true longitudinal assessment. Proposed protocols will track HEQ trajectories over months rather than sessions, measuring whether structured governance produces sustainable enhancement.

CIQ Enhancement and Adversarial Trust Testing. Future studies will use "bait" protocols, intentionally injecting plausible but incorrect information into AI outputs, to measure the user's Reliance Calibration Score (RCS). A high HEQ requires the user to catch these errors, proving they are a critical co-pilot, not a passive passenger.

Cross-Cultural Validation. Testing on non-Western AI platforms (Ernie Bot, YandexGPT) and diverse user populations to address WEIRD bias limitations.

Interactive Prompt Development. Building on the Q2/Q3 2025 prompt instrument release, advanced interactive prompt work is scheduled for Q1 2026. This includes questionnaire integration into evaluations, enabling structured self-assessment that feeds directly into HEQ scoring workflows. The goal is seamless integration with enterprise talent management systems.

Interactive Prompt Development. Building on the Q2/Q3 2025 prompt instrument release, advanced interactive prompt work is scheduled for Q1 2026. This includes questionnaire integration into evaluations, enabling structured self-

assessment that feeds directly into HEQ scoring workflows. The goal is seamless integration with enterprise talent management systems.

9.3 Enterprise Pilot Pathway

Organizations interested in piloting HEQ for talent management applications should contact basilpuglisi.com. Pilot programs include:

- Pre-employment assessment integration with existing hiring workflows
- Performance review HEQ tracking implementation
- Training program ROI measurement through pre/post HEQ assessment
- Organization-wide AI Adoption Readiness mapping

10. Methods Appendix

10.1 Cross-Platform Consistency Score

What This Metric Is: The 0.96 Cross-Platform Consistency Score measures how consistently different AI platforms score the same user under identical prompts. It is computed using Intraclass Correlation Coefficient (ICC), two-way random effects model, absolute agreement, single measures.

Technical Note: ICC calculated on identical prompts across varying models measures model consensus on scoring, not human trait stability. This is an engineering observation of prompt reliability, not a psychometric claim about the underlying construct.

What This Metric Is Not: This is not psychometric reliability (test-retest stability), inter-rater reliability (human scorer agreement), or construct validity (evidence that HEQ measures what it claims). The term "reliability" is reserved for formal psychometric validation.

Procedure: Five AI platforms received identical prompts and produced scores across four HEQ dimensions. ICC was computed across the 5×4 score matrix.

Formal Sample Size: Technical validation sample: 1 user (author), 5 platforms, 4 dimensions, 1 assessment session per platform. Cross-user consistency sample: 10 additional users via ChatGPT-4.

Informal Cross-User Testing (n=10): Ten additional users tested the prompt using personal ChatGPT-4 accounts. Observable findings:

- Dimensional score variance: ±4 points across users
- CIQ identified as lowest-scoring dimension: 10/10 users
- Rubric compliance (all required output fields present): 100%
- Composite score range: 78-94

Model Versions: ChatGPT-4 (September 2025), Claude 3.5 Sonnet, Gemini 1.5 Pro, Perplexity (default), Grok 2. Temperature settings: default for each platform.

Limitations: The 0.96 coefficient measures inter-model output agreement, not construct validity, test-retest reliability, or generalizability across users. High agreement may reflect consistent prompt interpretation rather than valid measurement of underlying constructs. The informal n=10 testing provides preliminary cross-user consistency evidence but does not constitute formal psychometric validation.

10.2 Psychometric Validation Roadmap

The following validation studies are required before HEQ can claim psychometric validity:

Construct Validity Plan. Correlate HEQ scores with established measures of related constructs (cognitive flexibility assessments, trust calibration tasks, learning

agility instruments). Discriminant validity: HEQ should not correlate highly with unrelated constructs (e.g., personality traits).

Test-Retest Reliability Plan. Administer HEQ to the same users at two time points (2-4 weeks apart) with no intervention. Target: ICC > 0.80 for composite score, > 0.70 for individual dimensions.

Inter-Rater Reliability Plan. Have multiple trained human raters score the same user outputs independently. Target: ICC > 0.85. This validates that the scoring rubric produces consistent results across evaluators.

Measurement Invariance Plan. Test whether HEQ measures the same construct across demographic groups (age, gender, cultural background, AI experience level). Differential item functioning (DIF) analysis required before claiming cross-population validity.

Status. None of these validation studies are complete. Current evidence supports prompt consistency and inter-model agreement only. Enterprise adopters should treat HEQ scores as directional indicators until formal validation is published.

10.3 Scoring Guidelines

Score interpretation guidelines for enterprise deployment:

- **0-20:** Minimal demonstrated capability; significant development needed
- **21-40:** Emerging capability; foundational application present
- **41-60:** Developing capability; consistent application across contexts
- **61-80:** Strong capability; adaptable integration across domains
- **81-100:** Expert capability; sophisticated application with strategic depth

Full prompt templates and scoring instructions are available in the GitHub repository (github.com/basilpuglisi/HAIA).

10.4 Replication Package

The following materials are available for independent replication:

- HAIA Intelligence Snapshot prompts (v1.0, v2.0, v3.0)
- Scoring rubric and dimension definitions
- Case Study 001 raw outputs (anonymized)
- ICC computation script

Repository: github.com/basilpuglisi/HAIA

11. References

11.1 Primary Sources

Puglisi, B. C. (2012). Digital Factics: Twitter. Digital Media Press. magcloud.com/browse/issue/471388

Puglisi, B. C. (2024, February). The intelligence enhancement thesis. basilpuglisi.com/factics-make-us-more-intelligent

Puglisi, B. C. (2025a). From metrics to meaning: Building the Factics Intelligence Dashboard. basilpuglisi.com/fid

Puglisi, B. C. (2025b). The Human Enhancement Quotient: Measuring Cognitive Amplification Through AI Collaboration. basilpuglisi.com/HEQ

Puglisi, B. C. (2025c). Governing AI: When Capability Exceeds Control. Amazon.

Puglisi, B. C. (2025d). Digital Factics X: The Success Guide to Mastering Business Growth on X. Amazon.

Puglisi, B. C. (2025e). HAIA-RECCLIN: The Multi-AI Governance Framework for Responsible AI. basilpuglisi.com/haia-recclin

11.2 Theoretical Foundations

Bansal, G., et al. (2019). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. *CHI Conference on Human Factors in Computing Systems.*

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. *NBER Working Paper 31161*. nber.org/papers/w31161

Carter, N. [@nic__carter]. (2025, April 15). I've noticed a weird aversion to using AI... [Rejecting it is like] deducting 30 IQ points [Post]. X. x.com/nic__carter

Dell'Acqua, F., et al. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Working Paper.*

Dweck, C. S. (2006). *Mindset: The New Psychology of Success*. Random House.

Ganuthula, V., & Balaraman, A. (2025). Artificial Intelligence Quotient (AIQ): A novel framework for measuring human-AI collaborative intelligence. *arXiv preprint arXiv:2503.16438.*

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books.

Gawdat, M. (2024-2025). On AI collaboration as "borrowing" cognitive capacity. Various interviews including *Diary of a CEO* and *Nordic Business Forum*. See also:

Gawdat, M. (2022). *Scary Smart: The Future of Artificial Intelligence and How You Can Save Our World*. Macmillan.

Goleman, D. (1995). *Emotional Intelligence: Why It Can Matter More Than IQ*. Bantam Books.

Kasparov, G. (2017). *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*. PublicAffairs.

Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge University Press.

Zhang, Y., et al. (2020). Why is AI not a panacea for data workers? An interview study on human-AI collaboration. *Proceedings of the ACM on Human-Computer Interaction*.

11.3 Scientific Validation Research (Cited in Section 2.6)

An, T. (2025). AI as Cognitive Amplifier: Rethinking Human Judgment in the Age of Generative AI. *arXiv preprint arXiv:2512.10961*. arxiv.org/pdf/2512.10961.pdf

CJPI Insights. (2025, November 5). Why Deloitte's $440,000 AI Report Is a Warning to Every Organisation Using Artificial Intelligence. *CJPI*. cjpi.com/insights/why-deloittes-440000-ai-report-is-a-warning-to-every-organisation-using-artificial-intelligence/

Lin, C., Greenstein, S., & MacCormack, A. (2024). Narrative AI and the Human-AI Oversight Paradox in Decision-Support Systems (Working Paper 25-001). *Harvard Business School*. hbs.edu/ris/Publication%20Files/25-001_8ebbe0cb-2a19-453c-9014-1e301e8dd2fb.pdf

Noy, S., & Zhang, W. (2023). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *MIT Economics Working Paper*. economics.mit.edu/sites/default/files/inline-files/Noy_Zhang_1.pdf

Yang, B., Wang, Y., & Li, X. (2025). A Token-Efficient Framework for Codified Multi-Agent Prompting and Workflow Execution. *arXiv preprint arXiv:2507.03254*. arxiv.org/pdf/2507.03254.pdf

11.4 Workforce Research

Yee, L., Madgavkar, A., Smit, S., Krivkovich, A., Chui, M., Ramirez, M. J., & Castresana, D. (2025, November). *Agents, robots, and us: Skill partnerships in the age of AI*. McKinsey Global Institute. mckinsey.com/mgi

11.5 Regulatory References

European Union. (2024). Regulation (EU) 2024/1689 (AI Act). *Official Journal of the European Union*. eur-lex.europa.eu/eli/reg/2024/1689/oj

International Organization for Standardization. (2023). ISO/IEC 42001:2023 Artificial intelligence management system. iso.org

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10

11.6 Repository and Contact

GitHub Repository: github.com/basilpuglisi/HAIA

Contact: me@basilpuglisi.com

HAIA-RECCLIN Resources: basilpuglisi.com/haia-recclin

Social: @BasilPuglisi

Appendix A: Role-Specific Value Propositions

A.1 For the CHRO / Head of People

HEQ segments your workforce into five AI readiness tiers. You can identify who leads transformation (AI Champions, 81-100), who needs development (Developing, 41-60), and who may need role reassignment (Pre-Foundational, 0-20).

Key Metrics: Workforce HEQ distribution curve. Delta HEQ per training program. AGR trajectory by employee cohort.

Decision Enabled: Which employees lead AI initiatives? Where should training investment concentrate? Who mentors whom?

A.2 For the CFO / Finance

Training programs currently prove attendance, not capability. HEQ enables cost-per-point improvement calculations. Pre-employment HEQ reduces mis-hire costs (1.5-2x annual salary per bad hire). Synergy validation justifies headcount by proving humans add value versus automation.

Key Metrics: Training ROI via Delta HEQ. Cost per HEQ point improvement. Synergy metric (S) by department.

Decision Enabled: Which training programs produce measurable capability? Does human oversight add or subtract value in specific workflows?

A.3 For the CTO / Technology

HAIA-RECCLIN creates audit trails for EU AI Act compliance. HEQ5's Societal Safety dimension maps to ISO 42001 requirements. The 0.96 inter-model output agreement suggests platform-agnostic deployment across your AI stack.

Key Metrics: Audit trail completeness percentage. Human override rate. Regulatory compliance score.

Decision Enabled: Which AI governance architecture meets regulatory requirements? How do we prove human oversight in high-risk applications?

A.4 For the CEO / Founder

The $2.9 trillion opportunity requires a workforce that collaborates with AI, not just uses it. Competitors measuring collaborative intelligence will identify and develop talent faster. HEQ is a competitive moat for talent strategy.

Key Metrics: Organization-wide HEQ mean versus industry benchmark. Percentage of workforce at AI Champion level (81+).

Decision Enabled: Are we building partnership capability or just deploying tools? How do we outcompete for AI-native talent?

Attribution and Ethical Use Notice

Attribution Requirements. Any use of HEQ, HEQ5, HAIA-RECCLIN, or related frameworks in research, publications, training materials, or commercial applications must include proper attribution to Basil C. Puglisi and this white paper. Suggested citation:

> *Puglisi, B. C. (2025). The Human Enhancement Quotient (HEQ): Measuring Collaborative Intelligence for Enterprise AI Adoption (Version 4.3.3). White Paper: Enterprise Pilot Edition. basilpuglisi.com/HEQ*

Permitted Uses. This white paper may be used for: scholarly review and academic citation, enterprise pilot programs with proper attribution, internal organizational assessment with disclosure to participants, derivative research that extends or tests the framework with citation.

Prohibited Uses. The following uses are not permitted without written authorization: commercial products or services based on HEQ without licensing agreement, claims of psychometric validity beyond what this paper explicitly states, use of HEQ scores as sole determinant in employment decisions, removal of attribution or misrepresentation of authorship.

Ethical Use Commitment. Organizations deploying HEQ commit to: disclosing assessment to all participants, using scores for development rather than punishment, providing appeal processes for disputed scores, interpreting results alongside other performance data, and monitoring for unintended bias in deployment.

Contact for Licensing. For commercial licensing, enterprise partnerships, or research collaboration inquiries: basilpuglisi.com

— End of White Paper: Enterprise Pilot Edition —