

Measuring Augmented Intelligence

Theoretical Foundations and Empirical Development of the
Human Enhancement Quotient (HEQ) and
Augmented Intelligence Score (AIS)

Basil C. Puglisi, MPA

Human-AI Collaboration Strategist and Digital Strategy Consultant
basilpuglisi.com | @BasilPuglisi

Working Paper v2.5 | February 2026

Released for scholarly review and collaboration. No commercial use without written permission.

How to cite: Puglisi, B. C. (2026). Measuring Augmented Intelligence: Theoretical Foundations and Empirical Development of the Human Enhancement Quotient (HEQ) and Augmented Intelligence Score (AIS) (v2.5). basilpuglisi.com/measuring-augmented-intelligence

Scoped Contribution Statement

This paper proposes Human-AI Collaborative Intelligence (HACI) as a measurement-oriented framing within human-AI teaming research. It introduces the Human Enhancement Quotient (HEQ) as an assessment instrument with four behavioral dimensions developed from fifteen years of documented practitioner research. It defines the Augmented Intelligence Score (AIS) as a composite output with a decomposable structure and a multi-platform administration protocol grounded in validated methodology.

This paper does not claim universal psychometric validity. It claims an instrument, an initial evidence base with documented cross-platform consistency, and a staged validation roadmap. All preliminary validation data originates from the framework author's research program. No independent research group has yet replicated HEQ administration or validated the instrument against external criteria. Organizations and researchers adopting HEQ during this period become validation partners whose data will strengthen the formal psychometric case. The four dimensions are proposed as core components of the augmented intelligence construct, subject to empirical verification of their independence, coverage, and relative weighting.

Executive Summary

Augmented intelligence, as defined by Gartner, is the recognized partnership model of humans and AI enhancing cognitive performance together. Organizations have invested heavily in that model. No cross-platform, behavior-anchored, governance-integrated instrument exists to measure how well it is working for a specific individual.

This paper introduces Human-AI Collaborative Intelligence (HACI) as the measurement science discipline that fills that gap. The Human Enhancement Quotient (HEQ) is the four-dimension assessment instrument within HACI. The Augmented Intelligence Score (AIS) is the standardized composite output.

HEQ measures four dimensions developed from the original Intelligence Estimate Prompt v1 and validated through three sequential studies: Cognitive Agility Speed (CAS), how quickly and clearly someone processes and connects ideas; Ethical Alignment Index (EAI), how well thinking reflects fairness, responsibility, and transparency; Collaborative Intelligence Quotient (CIQ), how effectively someone engages with others and integrates different perspectives; and Adaptive Growth Rate (AGR), how someone learns from feedback and applies it forward.

AIS is the arithmetic mean of those four dimensions, produced through a minimum three-platform multi-AI administration protocol. The protocol builds on a cross-platform consistency coefficient (ICC) of 0.96 demonstrated across five independent AI architectures in a 2025 n=1 feasibility study. Preliminary cross-user testing (n=10) suggests dimensional variance within four points and consistent identification of CIQ as the lowest-scoring dimension, aligning with independent Harvard and MIT research on human over-reliance on AI.

Formal psychometric validation is the 2026 priority, with a staged roadmap targeting Cronbach's alpha greater than 0.75 across n=100 or more participants, criterion validity against supervisor ratings and work samples at r greater than 0.35, and longitudinal tracking enabled by platform memory capabilities. Organizations and researchers adopting HEQ now become validation partners shaping the enterprise standard.

Abstract

Gartner defines augmented intelligence as a human-centered partnership model of people and artificial intelligence working together to enhance cognitive performance. Despite \$2.9 trillion in projected business value and organizational investment at scale, no cross-platform, behavior-anchored, governance-integrated instrument exists to measure how well a specific person performs within that partnership at the individual level. This paper addresses that gap by introducing Human-AI Collaborative Intelligence (HACI) as the measurement science discipline that operationalizes augmented intelligence at the individual level.

The Human Enhancement Quotient (HEQ) is the four-dimension assessment instrument within HACI, measuring Cognitive Agility Speed (CAS), Ethical Alignment Index (EAI), Collaborative Intelligence Quotient (CIQ), and Adaptive Growth Rate (AGR). The

Augmented Intelligence Score (AIS) is the standardized composite output, defined as the arithmetic mean of the four dimensions produced through a minimum three-platform multi-AI administration protocol delivering a confidence-banded score.

The instrument was developed through three sequential studies conducted in 2025. A single-platform proof of concept on ChatGPT established the HEQ baseline. A five-platform multi-AI feasibility study then administered the same prompt across ChatGPT, Claude, Gemini, Perplexity, and Grok, producing preliminary cross-platform consistency evidence with an ICC of 0.96 from an n=1 feasibility study. An advanced protocol development study produced the finding that methodological simplicity outperforms complexity for cross-platform reliability, and documented that all five platforms independently converged on recommending a fifth reliability meta-dimension, which became the empirical origin of HEQ5's governance layer.

Preliminary cross-user testing (n=10) suggests dimensional variance within four points and consistent identification of CIQ as the lowest-scoring dimension, aligning with independent research on human over-reliance on AI. A 2026 validation agenda targets Cronbach's alpha greater than 0.75 across n=100 or more participants, r greater than 0.35 against supervisor performance ratings, and memory-enabled longitudinal AGR tracking across 90 days to test the framework's central distinction between tool-dependent fluency and persistent intelligence enhancement.

Section 1:

The Measurement Gap in Augmented Intelligence

In 2019, Gartner defined augmented intelligence as a design pattern for a human-centered partnership model of people and artificial intelligence working together to enhance cognitive performance, including learning, decision making, and new experiences. That definition arrived with a forecast: AI augmentation would generate \$2.9 trillion in business value globally. Organizations responded. AI tooling budgets expanded. Training programs launched. Workforce strategies reorganized around the assumption that human-AI collaboration would become the primary engine of knowledge work productivity.

What did not arrive alongside the investment was a way to measure whether the collaboration was actually working.

The gap is not theoretical. Organizations today make billion-dollar decisions about AI platform selection, workforce development, and talent deployment without any standardized method for assessing individual-level augmented intelligence. A hiring manager cannot measure a candidate's capacity for AI partnership. A training director cannot prove that an AI literacy program produced measurable cognitive enhancement rather than tool familiarity. A performance review cycle cannot track whether an employee's collaborative intelligence is developing, stagnating, or declining. The market has fluency, the ability to use AI tools, but not intelligence, the measurable enhancement that sustained AI collaboration produces in the human partner.

This distinction between fluency and intelligence is not semantic. Fluency is binary. A person either uses the tool or does not. Intelligence is developmental. It grows through structured engagement, or it stagnates through passive dependence. The entire premise of augmented intelligence rests on the quality of the human contribution to the partnership. When Kasparov analyzed why human-computer teams dominated both human-only and computer-only competitors in Advanced Chess, the decisive variable was not the power of the machine. It was the quality of human judgment applied alongside machine calculation. That quality is measurable. Until now, no standardized instrument has measured it.

The empirical evidence compounds the urgency. Noy and Zhang's controlled study at MIT found that generative AI decreased task completion time by 40 percent and increased output quality by 18 percent among participants who used it. Those productivity gains are real and documented. They do not answer the more consequential question: when the AI is removed, does the capability remain? Productivity gains that disappear with the tool represent offloaded thinking, not enhanced intelligence. Productivity gains that persist represent genuine cognitive development. Organizations investing in AI collaboration cannot distinguish between the two without an individual-level measurement instrument.

McKinsey Global Institute's 2025 Skill Change Index found that 72 percent of skills now operate in shared mode, the partnership zone where neither pure human effort nor pure AI automation produces optimal results. By 2030, optimizing that partnership could unlock the full projected economic value of AI adoption. The bottleneck is not AI capability. The bottleneck is human partnership quality, and no standardized instrument to measure it at the individual level exists.

Adjacent frameworks have addressed pieces of the problem without assembling the whole. Hybrid intelligence research examines how human-AI teams perform at the system level but does not produce individual scores. Human-in-the-loop taxonomies describe when humans should intervene but do not measure how well they intervene. Trust calibration research identifies behavioral patterns that distinguish effective from ineffective AI reliance but treats trust as an isolated variable rather than one dimension of a composite instrument. Productivity research measures task outputs without capturing whether the human contributor developed through the process or merely benefited from the tool.

Human-AI Collaborative Intelligence (HACI) is the measurement science that occupies this gap. The Human Enhancement Quotient (HEQ) is the assessment instrument that operationalizes HACI through four behavioral dimensions developed from fifteen years of practitioner research. The Augmented Intelligence Score (AIS) is the standardized composite output that gives individuals, organizations, and governance bodies a single, decomposable, longitudinally trackable number for augmented intelligence performance.

Section 2: Theoretical Lineage

HACI stands at the terminus of six decades of converging intellectual work. Understanding that lineage clarifies precisely what HACI inherits from each predecessor and what it adds that none of them provide.

2.1 The Six-Node Genealogy

Licklider (1960): Symbiosis as the Operating Model

J.C.R. Licklider's 1960 paper established the foundational vision: a cooperative interaction model where humans set goals, formulate hypotheses, and determine criteria, while machines execute the computational work. From Licklider, HACI inherits the symbiosis framing. HACI adds measurement of how well the human half of that symbiosis performs.

Engelbart (1962): Augmentation as a Systems View

Douglas Engelbart extended Licklider's vision into a systematic approach to improving the intellectual effectiveness of the human-plus-tools system. His emphasis on the compound system, not the individual tool or person, anticipates HACI's unit of analysis. From Engelbart, HACI inherits the systems view of augmentation. HACI adds individual-level quantification of how much enhancement the system produces for a specific person.

Gardner and Sternberg (1983, 1985): Intelligence as Multiple and Contextual

Howard Gardner's theory of multiple intelligences demolished the assumption that cognitive capability reduces to a single measurable trait. Robert Sternberg's triarchic theory extended this by distinguishing analytical, creative, and practical intelligence. Both established that intelligence is contextual, shaped by the environment in which it operates. From Gardner and Sternberg, HACI inherits the multi-dimensional intelligence construct. HACI adds the AI-collaboration context as the specific environment in which its four dimensions operate.

Dweck (2006): Growth as Measurable Trajectory

Carol Dweck's research established empirically that intelligence is not fixed. If intelligence grows, the measurement instrument must capture trajectory, not just snapshot. From Dweck, HACI inherits the growth framing. HACI adds the Adaptive Growth Rate dimension as the specific mechanism for capturing intelligence development through sustained AI collaboration.

Kasparov (2017): The Empirical Proof

Kasparov's analysis of Advanced Chess provided the empirical foundation that transforms augmented intelligence from theoretical construct to demonstrated phenomenon. The teams that won most consistently were not those with the most powerful chess engines. They were those where the human contributor most effectively guided, questioned, and directed the machine's output. From Kasparov, HACI inherits the empirical demonstration that collaborative intelligence exceeds either component alone. HACI adds the measurement instrument for the quality of the human contributor within that partnership.

Gartner (2019): The Enterprise Vocabulary

Gartner's formal definition of augmented intelligence as a design pattern for human-centered AI partnership established the conceptual vocabulary that now shapes enterprise AI investment decisions. From Gartner, HACI inherits the enterprise vocabulary and the organizational legitimacy that comes with it. HACI adds the measurement instrument the enterprise vocabulary has lacked since 2019.

2.2 The Missing Element

Across sixty years of foundational work, one element is consistently present as an assumption and consistently absent as an instrument. Every framework assumes that human performance within augmented intelligence partnerships is worth optimizing. None provides a standardized way to measure it at the individual level, track it over time, and produce a score comparable across individuals and organizations. HACI builds that measurement layer.

Section 3: The FID, HEQ, and AIS Development Path

The HACI framework did not emerge from theory. It emerged from practice, from fifteen years of systematic observation and testing documented through 900-plus published articles, conference presentations, and operational framework development. The empirical foundation rests on three sequential studies conducted in 2025.

3.1 Factics: The Foundational Methodology (2011 to 2024)

The methodological roots trace to a 2011 Social Media Week conference presentation that seeded the core principles; formal publication followed as Digital Factics: Twitter in November 2012. The Factics methodology formalized through that 2012 publication established the core principle that structured information-to-action frameworks increase applied human intelligence. Facts plus Tactics plus KPIs forced a specific cognitive discipline: claims attach to facts, recommendations declare tactics, tactics declare measurable outcomes. By February 2024, a published post on basilpuglisi.com articulated the intelligence enhancement thesis explicitly: structured AI engagement measurably increases applied human intelligence. That sentence transformed a communication methodology into a measurement hypothesis.

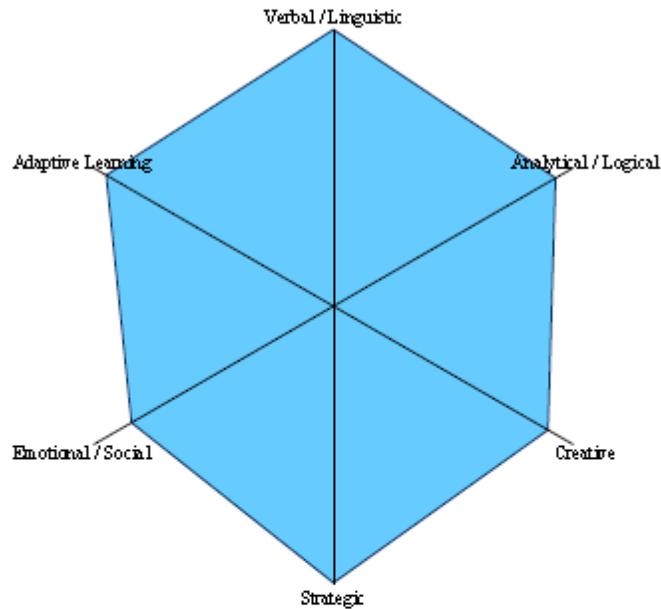
3.2 The Factics Intelligence Dashboard: A Parallel Instrument

The Factics Intelligence Dashboard (FID) was developed as a standalone applied intelligence measurement instrument using six domains: Verbal/Linguistic, Analytical/Logical, Creative, Strategic, Emotional/Social, and Adaptive Learning. FID measures how a person uses AI in a single session. It produces a snapshot profile of applied intelligence as it presents in that interaction. The FID was administered once on ChatGPT in August 2025, producing the six-domain radar chart that documents Basil C. Puglisi's applied intelligence profile at that moment.

FID and HEQ are parallel instruments serving different time horizons within the same collaboration journey. FID answers: how did someone use AI in this session? HEQ answers: how did using AI change their intelligence capabilities over time in a project, study, paper, research engagement, role, or job? FID is session-level. HEQ is longitudinal. Both require AI platform engagement. Neither is a precursor to or replacement for the other.

AI-Enhanced Intelligence Dashboard

Profile: Basil C. Puglisi – Powered by Facts & AI



Personalized Insights:

- Verbal / Linguistic (95) → Continue leveraging narrative mastery for blogs, teaching, and consulting.
- Analytical / Logical (88) → Strong for compliance mapping and SEO frameworks; use for structured growth.
- Creative (85) → Solid ideation; balance with pragmatism keeps innovation practical and actionable.
- Strategic (95) → Superpower: long-range planning, ecosystem building, and KPI-driven foresight.
- Emotional / Social (80) → Use empathy-driven storytelling to expand leadership impact further.
- Adaptive Learning (90) → Keep pioneering AI workflows; ability to integrate fast is a competitive edge.

@BasilPuglisi – Facts Consulting

#AIgenerated

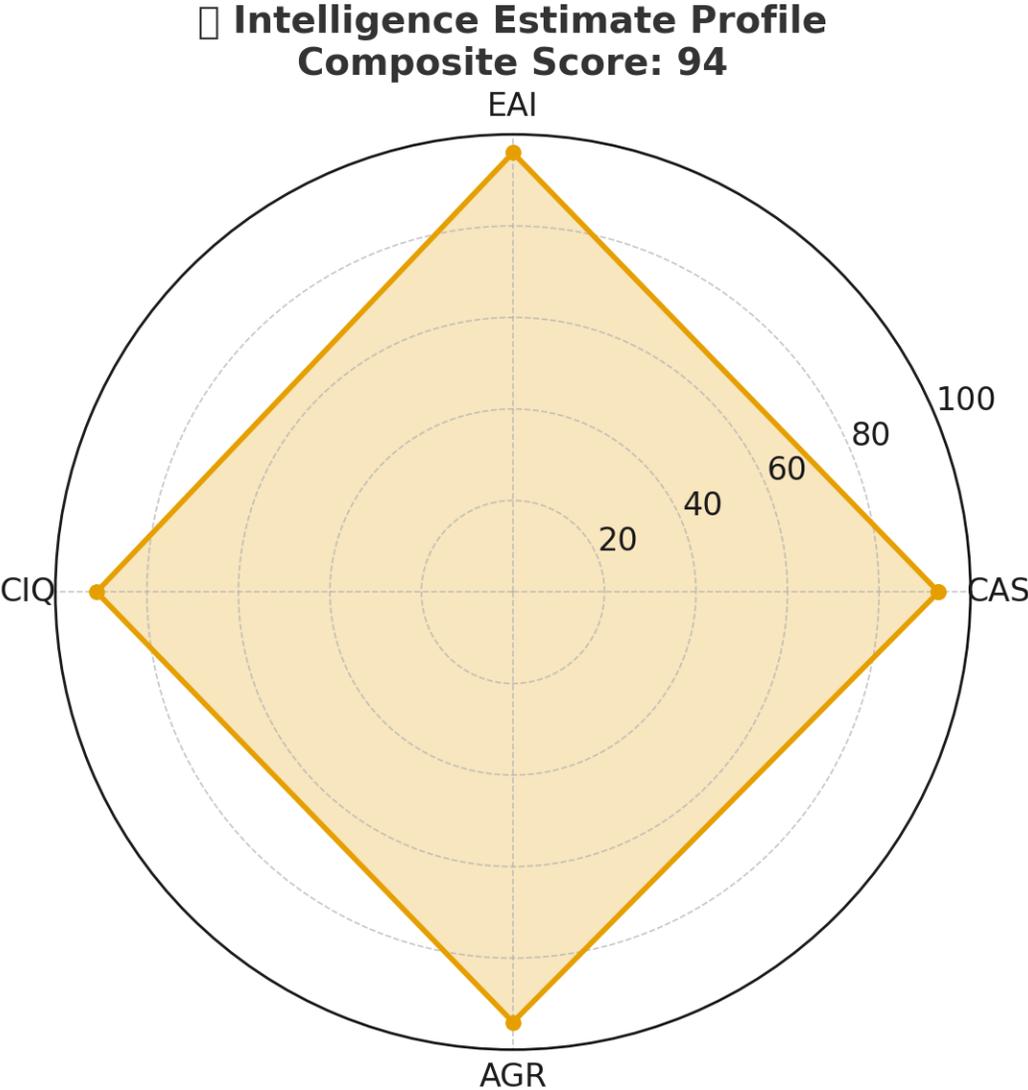
3.1: FID Six-Domain Radar Chart, Basil C. Puglisi, August 2025

3.3 HEQ Study 1: Single-Platform Proof of Concept

The Human Enhancement Quotient instrument was developed with four dimensions drawn from the measurement hypothesis: Cognitive Agility Speed (CAS), Ethical Alignment Index (EAI), Collaborative Intelligence Quotient (CIQ), and Adaptive Growth Rate (AGR). The first

HEQ administration used the Intelligence Estimate Prompt v1 on ChatGPT with Basil C. Puglisi as the subject. The prompt instructed the AI to analyze answers, writing style, and reasoning to estimate the four dimensions and produce a composite score.

ChatGPT produced CAS 93, EAI 96, CIQ 91, AGR 94, composite 94. This established the HEQ baseline and confirmed that the four-dimension instrument produced a coherent, scoreable output from a single session.



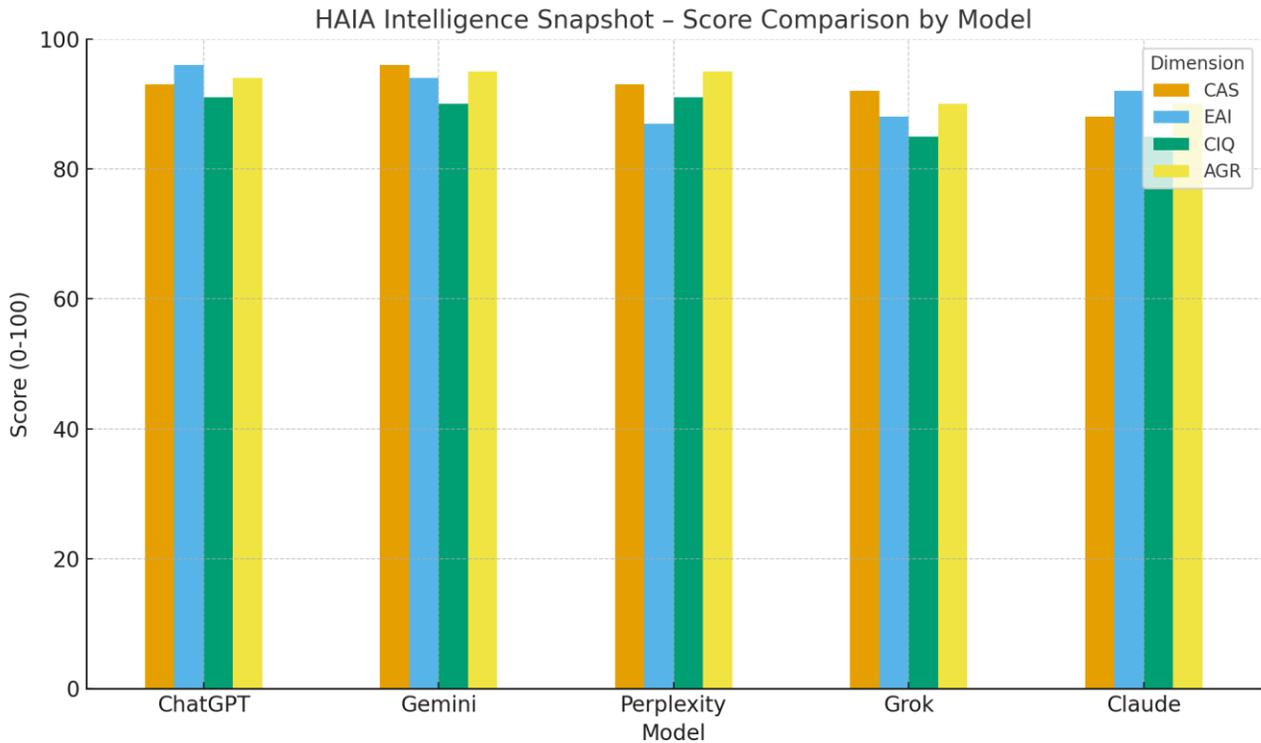
3.2: HEQ Intelligence Estimate Profile, ChatGPT, Composite Score 94 (2025)

3.4 HEQ Study 2: Multi-AI Reliability Run

The same Intelligence Estimate Prompt v1 was then administered across all five platforms: ChatGPT, Gemini, Perplexity, Grok, and Claude, with Basil C. Puglisi as the subject throughout. This was an n=1 feasibility study designed to test whether the HEQ instrument produced consistent results across architecturally distinct AI systems.

Platform	CAS	EAI	CIQ	AGR	Composite
ChatGPT	93	96	91	94	94
Gemini	96	94	90	95	94
Perplexity	93	87	91	95	92
Grok	92	88	85	90	89
Claude	88	92	85	90	89

Composite scores ranged from 89 to 94, a five-point variance. The cross-platform consistency coefficient (ICC), computed as a two-way random effects model, absolute agreement, average measures, on the 5x4 score matrix, was 0.96. Narrative convergence across platforms exceeded 95 percent agreement on assessment themes despite numerical variance. CIQ was identified as the lowest-scoring dimension by every platform.



3.3: HAIA Intelligence Snapshot Score Comparison by Model, Five Platforms 2025

3.5 HEQ Study 3: Advanced Protocol Development

Each of the five platforms was then asked to revise and improve the HEQ prompt. All five platforms independently proposed adding a fifth dimension: a Reliability and Confidence Index (RCI) as a meta-score reflecting data quality and assessment confidence. This convergent recommendation across five architecturally distinct systems is the empirical origin of HEQ5's governance layer. The platforms also contributed distinct structural

innovations: Gemini formalized a three-step co-creation process, Perplexity added transparency language and disclosure requirements, Grok added confidence bands, Claude added weighted RCI and enterprise reporting structure, ChatGPT synthesized contributions into a boardroom-ready format.

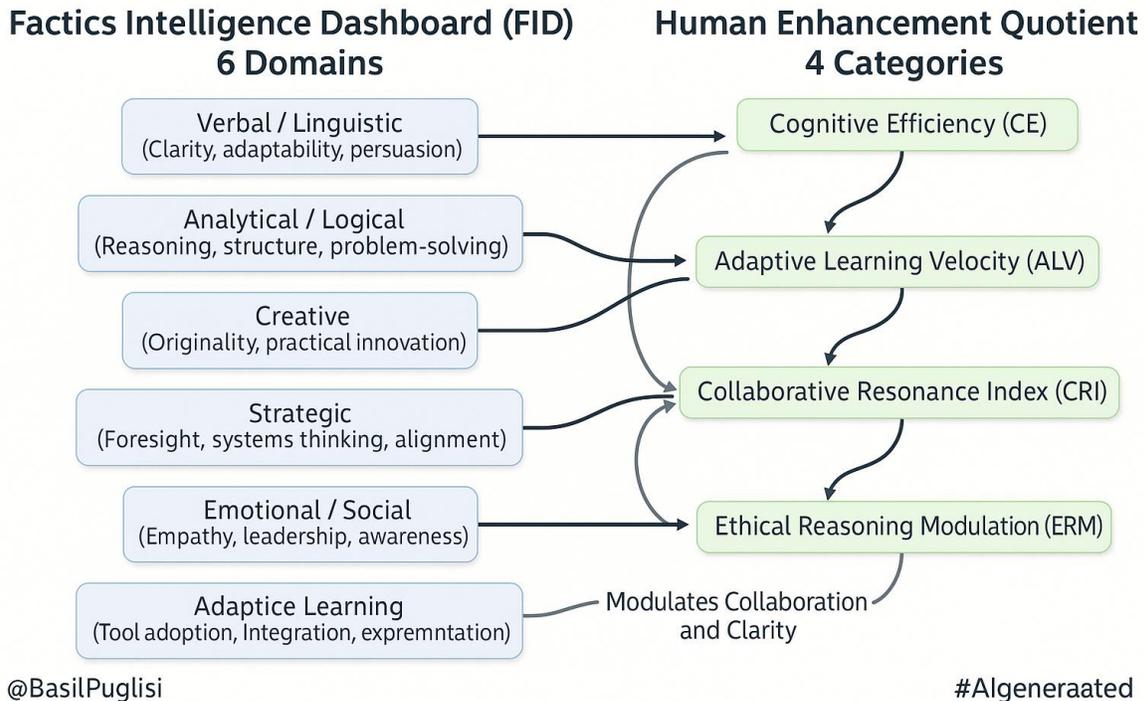
The resulting advanced prompts were tested for cross-platform execution. The complex protocol achieved only 25 percent success across platforms. The primary failure modes were platform reinterpretation, outright refusal, and session architecture constraints. This finding produced the governing methodological principle: simplicity delivers universal reliability while sophistication fails across platforms. The return to an optimized simple protocol became the foundation for the HAIA Hybrid-Adaptive v3.1 and Universal v4 prompts.

The Study 3 finding is not primarily a failure story. It is a construct validation finding. Five independent AI architectures converged on the same structural gap in the four-dimension instrument. That convergence is empirical evidence that a reliability meta-layer is necessary, not merely useful.

3.6 The FID to HEQ Dimensional Structure

FID’s six domains informed HEQ’s four dimensions. The condensation logic follows directly from the shift in measurement question: from static snapshot to dynamic trajectory.

FID Domain	HEQ Dimension	Purpose
Verbal / Linguistic	Cognitive Agility Speed (CAS)	How quickly and clearly someone processes and connects ideas when supported by AI
Analytical / Logical	Ethical Alignment Index (EAI)	How well thinking reflects fairness, responsibility, and transparency
Creative + Strategic	Collaborative Intelligence Quotient (CIQ)	How effectively someone engages with others and integrates different perspectives
Emotional + Adaptive	Adaptive Growth Rate (AGR)	How someone learns from feedback and applies it forward



3.4: FID to HEQ Domain Mapping 2025

The condensation from six domains to four dimensions was not a simplification of the constructs. It was a reorientation of the measurement question. FID’s six domains each describe a distinct static capability: how well someone writes, reasons, creates, strategizes, relates, or adapts at a moment in time. When the measurement question shifted from snapshot to trajectory, the relevant unit became the behavioral pattern that drives enhancement over time, not the capability type that produces output in a session.

Verbal and linguistic capability accelerates through AI collaboration when the person processes and connects ideas rapidly, making CAS the enhancement-oriented equivalent of that domain. Analytical reasoning under AI influence becomes most consequential when it involves maintaining ethical integrity against AI-introduced pressure, making EAI the trajectory-relevant expression of that domain. Creative and strategic thinking in AI collaboration manifests primarily through co-creation quality, whether the person treats the AI as a genuine partner or an execution tool, making CIQ the enhancement dimension for those combined domains. Gardner’s distinctions between interpersonal and creative intelligence, and Sternberg’s separation of creative from practical intelligence, remain theoretically valid within the FID six-domain structure. In HEQ, those distinctions are preserved within the operational definitions of each dimension rather than as separate scored outputs, because the measurement target is enhancement trajectory, not domain inventory. The condensation is not a claim that Gardner’s or Sternberg’s distinctions cease to matter. It is a claim that when the question is how AI collaboration changes a person over time, the behavioral patterns that drive that change cluster into four observable trajectories rather than six static profiles.

Emotional and adaptive capabilities drive growth trajectory when they determine whether a person incorporates feedback and transfers learning across contexts, making AGR the natural trajectory expression of that combined domain. The key distinction is between adaptive response to a single exchange, which FID's Adaptive Learning domain captures in snapshot, and the persistent learning pattern across a sustained body of work, which AGR measures as a developmental rate.

Section 4: Construct Definition and Dimensional Architecture

4.1 Formal Definitions with Boundaries

Human-AI Collaborative Intelligence (HACI) is the measurement science that quantifies the quality and trajectory of human-AI collaborative performance at the individual level. It is distinct from team-level Hybrid Intelligence research, system-level Human-Computer Interaction evaluation, and productivity studies. HACI's unit of analysis is the person operating within an augmented intelligence partnership.

The Human Enhancement Quotient (HEQ) is the assessment instrument that operationalizes HACI through four behavioral dimensions. It measures enhancement, the change in human cognitive performance produced by sustained AI collaboration, tracked across a meaningful unit of work rather than at a single point.

The Augmented Intelligence Score (AIS) is the composite standardized output of HEQ administration. $AIS = (CAS + EAI + CIQ + AGR) / 4$. It is distinct from benchmarking scores, which measure task completion against a fixed standard, and from model evaluation scores, which measure AI capability.

4.2 The Four Dimensions

Cognitive Agility Speed (CAS)

How quickly and clearly someone processes and connects ideas. The AI observes writing style, reasoning pace, how fast connections are made between concepts, and how clearly ideas are expressed under cognitive load. CAS is not abstract processing speed measured in isolation. It is observable clarity and connection-making speed within the AI-augmented work context, which aligns with fluid intelligence (Gf) in the Cattell-Horn-Carroll hierarchical framework, specifically the inductive and deductive reasoning speed component, rather than the narrow processing speed (Gs) measured by conventional timed tasks. "Speed" in Cognitive Agility Speed refers to the velocity of that agility within AI-augmented reasoning contexts, specifically how quickly adaptive cognitive connections are made when AI changes the work environment, not timed processing latency as measured by conventional psychometric instruments. High CAS scores indicate that AI collaboration measurably accelerates cognitive throughput without sacrificing precision. Low CAS scores indicate either over-reliance on AI output without genuine processing, or under-utilization of available AI capacity.

The primary theoretical anchor is Artman and Garbis's (1998) distributed cognition framework, which establishes that effective collaboration depends on synchronizing representational structures between humans and their tools. When AI changes the cognitive load context, the quality of that synchronized processing changes with it, making CAS the measure of how effectively a person maintains high-quality distributed cognition under AI augmentation. The broader augmentation systems view from Engelbart (1962) establishes the lineage: the human-tools system outperforms either component in isolation

to the degree that the human effectively directs the tools. Sidra and Mason's (2025) validation of the Collaborative AI Literacy Scale provides adjacent empirical support, demonstrating that the ability to process and connect ideas within AI-supported work is a measurable individual-level skill distinct from general technical proficiency.

Ethical Alignment Index (EAI)

How well thinking reflects fairness, responsibility, and transparency. The AI observes whether declared values appear in the reasoning, whether fairness considerations show up in how problems are framed, whether the person acknowledges tradeoffs or steamrolls past them, whether transparency about uncertainty is present or absent. EAI measures observable ethical reasoning behavior in the output, not the person's character or intentions. The word alignment is accurate: the AI checks whether expressed thinking aligns with ethical principles.

The theoretical anchor is Bandura's moral agency research, the capacity to act in accordance with declared values rather than situational pressures, applied to the specific pressure AI introduces when it delivers authoritative but potentially incorrect recommendations. EAI extends the intelligence family following the precedent established by Goleman's Emotional Intelligence, which demonstrated that socio-cognitive capacities orthogonal to traditional g-factor measures constitute independently measurable and practically consequential intelligence dimensions. EAI applies this same logic to the moral reasoning domain, treating ethical decision-making behavior under AI influence as a distinct, observable, and enhancement-relevant capacity. Empirical validation will determine how cleanly EAI fits within the intelligence family tradition, but the theoretical case follows the same pathway Goleman established for EQ.

The Harvard Oversight Paradox finding, that humans are 19.4 percentage points more likely to defer to incorrect AI recommendations when the AI uses authoritative language (Lane et al., 2024), is the primary empirical case for why EAI is a necessary dimension. AI-introduced authoritative framing creates measurable ethical pressure that does not exist without AI involvement, making EAI context-specific in a way that general moral reasoning scales do not capture.

Collaborative Intelligence Quotient (CIQ)

How effectively someone engages with others and integrates different perspectives. The AI observes whether the person invites other viewpoints, builds on input rather than overrides it, integrates feedback into reasoning, and treats the AI as a genuine thinking partner rather than an execution tool. CIQ measures AI-collaboration skill specifically: the calibrated reliance behavior that distinguishes effective from ineffective AI partnership. Whether CIQ scores predict collaborative performance with human teams is a hypothesis for future validation rather than an assumed property of the construct.

The theoretical foundation draws from trust calibration research in human-automation systems. Lee and See (2004) established that calibrated trust, where reliance on automation matches the system's actual capability, is both measurable and the primary predictor of effective human-machine collaboration outcomes. Aljaziri's (2025) empirical work on within-session trust calibration dynamics demonstrates that calibration quality

explains substantially more variance in collaboration outcomes than trust level alone, providing direct support for CIQ as the operative construct. Bansal et al.'s (2019, 2021) CHI research documents that AI explanations designed to help can paradoxically increase inappropriate reliance rather than calibrated oversight, underscoring that effective AI collaboration requires active calibration skill, not passive trust.

CIQ consistently scores lowest across all validation testing and across all preliminary cross-user participants. This aligns with Dell'Acqua et al.'s (2023) Jagged Frontier research, which shows that AI competence varies unpredictably across task types, making calibrated collaboration the most demanding cognitive skill in augmented intelligence partnerships. The CIQ lagging pattern across independent platforms and independent participants is the strongest construct validity evidence in the current dataset. In enterprise deployment contexts, CIQ is operationalized as the ratio of appropriate reliance behaviors, correct trust acceptance of valid AI output against correct skepticism rejecting AI errors, as documented in the enterprise administration protocol. This quantitative operationalization is compatible with the construct definition above and provides the auditable behavioral metric that organizational deployment requires.

Adaptive Growth Rate (AGR)

How someone learns from feedback and applies it forward. The AI observes whether behavior changes based on prior feedback, whether learning transfers across contexts, and whether there is measurable trajectory in how engagement develops over time. AGR is the only dimension that is explicitly longitudinal in the prompt design. Forward application requires time. A single session provides limited AGR signal. AGR is most meaningfully measured across a project, study, paper, research engagement, role, or job, the sustained work units that generate the history needed to observe genuine growth rather than session-to-session variation.

The theoretical architecture rests on three complementary anchors. Dweck's (2006) growth mindset research establishes the foundational premise that intelligence is malleable and that capability development through challenge differs fundamentally from performance through tool dependency. Dweck's instrument measures the belief in growth as an attitudinal construct. AGR operationalizes the behavioral rate of that growth rather than the belief in its possibility, which is a deliberate and important distinction. Mendoza and Yan's (2025) Growth Practices Scale bridges this gap by shifting from mindset as attitude to mindset as behavioral practice, providing the instrument-level precedent for measuring growth through observable actions rather than self-reported beliefs.

Vygotsky's Zone of Proximal Development (ZPD) provides the second anchor: the space between what a person can do independently and what they can achieve with capable assistance. AI collaboration is the most scalable form of capable assistance yet produced, and AGR measures how effectively a person uses that assistance to expand their independent capability ceiling over time. A high AGR score indicates that the assisted performance is producing genuine independent capability growth. A low AGR score indicates that the assistance is producing output without growing the person.

Newell and Rosenbloom's Power Law of Practice establishes the third anchor: learning rate is a stable, measurable individual characteristic. People learn at different speeds and that

rate predicts future performance independently of current capability level. AGR captures this rate within the AI-collaboration domain, making it a measure of developmental velocity rather than snapshot position. Sidra and Mason's (2025) Collaborative AI Metacognition Scale, which validates self-monitoring and feedback integration within AI-supported work as distinct measurable skills, provides adjacent empirical support for the behavioral observability of what AGR tracks.

4.3 The Dimensional Independence Evidence

The four-dimension structure was not assumed from the outset. The original HEQ instrument proposed six dimensions. Three consolidation events occurred through empirical testing: Verbal Clarity and Analytical Precision collapsed into CAS when testing revealed they tracked the same underlying processing pattern in AI-augmented contexts. Creative Originality was eliminated after three independent iterations failed to produce discriminatory scoring results across platforms. Strategic Foresight merged into AGR when longitudinal observation showed that strategic adaptation and growth trajectory were operationally indistinguishable over time. The four dimensions that survived are those that produced consistent, differentiable scoring patterns across multiple platforms and multiple subjects. The four-dimension structure is therefore an empirically derived outcome of testing, not a theoretically assumed architecture.

The four dimensions are proposed as core components of augmented intelligence performance, subject to empirical verification of their independence and coverage. The CIQ lagging pattern across all platforms and all preliminary cross-user participants provides initial evidence of dimensional independence. If the four dimensions measured a single underlying trait, they would move together. The consistent CIQ divergence while CAS, EAI, and AGR score higher demonstrates that they capture distinguishable behavioral patterns. Formal factor analytic verification is planned in the 2026 validation studies.

4.4 Construct Validity Model

Content validity rests on alignment between the four dimensions and the Gartner definition of augmented intelligence as a model enhancing cognitive performance including learning, decision making, and new experiences. CAS corresponds to cognitive performance enhancement. EAI corresponds to decision-making integrity. CIQ corresponds to the quality of human-AI experience co-creation. AGR corresponds to learning trajectory.

Internal structure validity is supported by the ICC of 0.96 from Study 2, which documents methodology reliability, and by the CIQ lagging pattern, which provides initial evidence of dimensional independence.

Criterion and predictive validity are planned research agenda items. The 2026 validation studies will correlate AIS with supervisor performance ratings, training outcome measures, and longitudinal productivity metrics. The discriminant validity hypothesis is that AIS will correlate with growth mindset and openness to experience at r approximately 0.3 to 0.5, reflecting shared developmental orientation, while diverging from g-factor IQ at r less than 0.2 and from Big Five conscientiousness, reflecting that augmented intelligence is not

equivalent to general cognitive ability or trait-level diligence. The 2026 validation studies will test this hypothesis empirically.

4.5 The Composite Formula and Its Governance Rationale

$$\text{AIS} = (\text{CAS} + \text{EAI} + \text{CIQ} + \text{AGR}) / 4$$

The arithmetic mean was chosen over weighted composite approaches for a specific governance reason. A weighted composite requires empirically derived factor loadings from validation studies not yet complete. The arithmetic mean is transparent, reproducible, and conservative. It makes no assumptions about dimensional importance that the current evidence does not support. When the 2026 validation studies produce factor loadings, dimensional weighting can be applied to a revised formula. Until then, equal weighting is the epistemically honest choice and provides a transparent baseline against which future empirically-derived weightings can be compared, ensuring the formula's evolution remains auditable.

The term Quotient in HEQ follows the established psychometric naming convention rather than its original mathematical meaning. IQ has not been computed as a true ratio since standardized deviation scoring replaced Binet's mental age formula in the mid-twentieth century. EQ has never been computed as a ratio. In both cases, Quotient survived as shorthand for a standardized measure of a distinct intelligence type. HEQ inherits that convention intentionally. FID measured applied intelligence in the same conceptual register as IQ. HEQ measures that same intelligence when augmented through AI collaboration. AIS is the standardized composite output, carrying the same relationship to HEQ that a deviation score carries to the construct it measures.

Section 5: The Multi-AI Scoring Methodology

5.1 The Theoretical Argument for Multi-Platform Administration

The requirement that AIS production use a minimum of three independent AI platforms is a construct-level necessity, not a methodological preference.

Augmented intelligence, properly defined, is demonstrated through effective human orchestration across AI systems, not dependence on a single architecture. A person who collaborates fluently with one platform and struggles with others demonstrates narrower collaborative intelligence than one who adapts effectively across architectures. If the construct is platform-independent collaborative capability, the measurement instrument must sample that capability across platforms.

The methodological logic follows Campbell and Fiske's (1959) multitrait-multimethod matrix principle adapted for the AI era: when multiple independent measurement instruments produce convergent results, confidence in the underlying construct increases. Study 2 demonstrated this empirically. Five architecturally distinct platforms converged on the same subject assessment within a five-point composite range, producing an ICC of 0.96. That convergence is the reliability foundation on which multi-platform administration rests.

Multi-platform administration also provides a structural validity safeguard against score gaming. An individual seeking to manipulate their AIS would need to maintain a consistent and coherent performance strategy across three or more architecturally distinct AI systems with different evaluation conventions, training data lineages, and reasoning patterns. This is substantially more demanding than gaming a single platform and produces a natural ceiling on coaching-inflated scores. This safeguard is not a replacement for the Bait Injection protocol, the administration of prompts with intentional errors embedded in AI outputs to measure actual error detection behavior rather than self-reported competency, but it operates in parallel with it as a structural feature of the instrument design.

5.2 Platform Independence Criteria

Not all platforms qualify as independent for AIS purposes. Independence requires architectural differentiation. Platforms sharing the same underlying foundation model do not constitute independent assessments. Independence requires different training data lineages, different architectural families, or documented performance divergence on reasoning benchmarks.

The five platforms used in Study 2 meet the independence criteria. The Observed Patterns analysis from multi-AI testing shows platform compliance rates ranging from 32 to 98 percent on structured assessment protocols, confirming that they behave differently rather than equivalently.

HAIA-RECCLIN AI Performance Analysis

Multi-AI White Paper Assembly: Systematic Comparison of Five Models

Methodology: Single complex prompt (20+ page defense-ready white paper) given to 5 AIs independently. Results evaluated against 7 explicit criteria.
Limitations: n=1, subjective scoring, potential confirmation bias.

Evaluation Dimension	Claude	Grok	Gemini	ChatGPT	Perplexity
Task Execution	ASSEMBLER Full 20+ page document, preserved all sections	ASSEMBLER Complete with verification notes	ASSEMBLER Complete but meta-narrative style	SUMMARIZER 5-7 pages, violated "no compression"	SUMMARIZER Fragmented, aggressive compression
Structure & Flow	Clean sections, logical sequence, smooth transitions	Well-organized, clear progression	Strong but includes process announcements	Basic sections present, limited depth	Messy, filename remnants, poor flow
Voice & Style	Strong Basil voice, authoritative yet conversational	Best match: includes faith/imagination/personal origin	Mostly on-voice but sometimes clinical	Generic business tone, less human-centered	Dry academic, minimal narrative
Detail Preservation	High: full RECCLIN roles, operational mechanics	Very high: escalation logs, dissent cases included	Good but more summary connectors	Low: bullets vs paragraphs, details lost	Very low: multiple paragraphs compressed to one
Citations & Verification	Proper inline [1], full APA references	Same as Claude plus unverified source flagged	APA list present, no inline numbers	Inline numbers present, references adequate	Incomplete refs, irrelevant URLs included
Defensive Rigor	Contestable, open invitation for review	Strongest: flags provisional data, invites challenge	Good but no explicit uncertainty flags	Mentions dissent but not systematically	Fails: no clear dissent handling
Word Count	4,600+ words	5,100+ words	3,500-4,000 words	~1,200 words	~1,300 words

5.1: HAIA-RECCLIN AI Performance Analysis, Five-Platform Comparison 2025

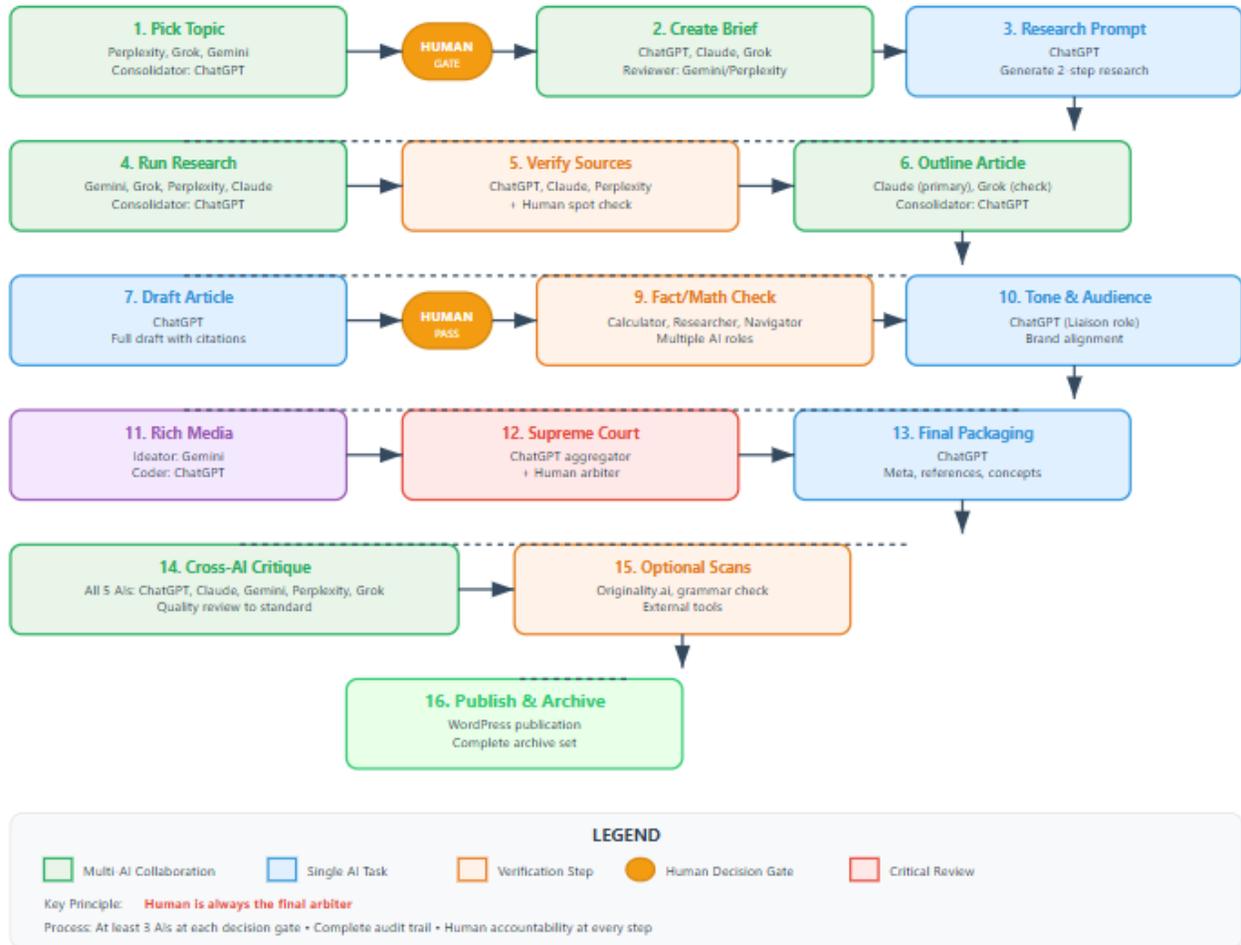
5.3 The HAIA Administration Protocol

AIS is produced through the HAIA assessment protocol, specifically the Hybrid-Adaptive v3.1 version developed from Study 3 findings. The protocol operates in six steps: historical analysis of available conversation data; three-question baseline assessment covering problem-solving, ethical reasoning, and collaborative planning; gap evaluation comparing baseline to historical patterns; targeted follow-up questions for flagged dimensions, zero to five additional questions, hard cap at eight total; adaptive scoring weighting historical data up to 70 percent with live responses minimum 30 percent; and output production including the complete AIS Snapshot.

For multi-platform AIS production, identical prompts are administered to each qualifying platform under standardized conditions. Model identifiers and version strings are documented. The cross-platform mean constitutes the reported AIS. The standard deviation across platforms constitutes the confidence band. Human arbitration of the composite is required before the AIS is finalized.

HAIA-RECCLIN 16-Step Workflow

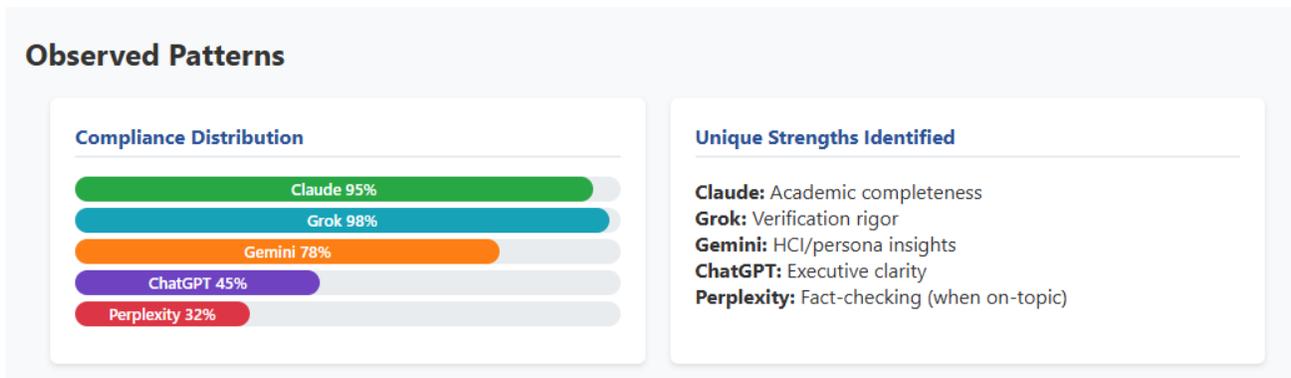
Multi-AI Collaboration with Human Oversight



5.2: HAIA-RECCLIN 16-Step Workflow 2025 (Updated in 2026, Pre Agent & GOPEL)

5.4 The AIS Snapshot: Standardized Output Format

Every AIS administration produces a complete snapshot with seven required fields: dimensional scores with confidence bands for CAS, EAI, CIQ, and AGR; composite AIS with confidence band; reliability statement including historical sample size, live exchanges completed, evidence threshold status, platform list with model versions, and growth trajectory relative to baseline; and a narrative of 150 to 250 words covering strengths, gaps, and one actionable development recommendation.



5.3: HAIA Intelligence Snapshot and HEQ Profile Card 2025

5.5 Scoring Bands and Developmental Interpretation

AIS scores are interpreted as developmental indicators, not credentials or pass/fail thresholds.

AIS Range	Classification	Development Implication
85 to 100	High Collaborative Intelligence	Candidate for AI initiative leadership; peer coaching roles
70 to 84	Proficient Collaborator	Strong foundation; target CIQ refinement for advancement
55 to 69	Developing Collaborator	Structured training program recommended; high return on investment
40 to 54	Dependent User	Interface training assessment before capability conclusions; structured oversight required
Below 40	Pre-Instrument Finding	Insufficient AI collaboration exposure for meaningful HEQ measurement; foundational literacy program required before AIS administration

The asymmetric band structure is deliberate. The instrument measures people engaged in AI collaboration. A score below 40 is not low performance on the instrument. It is a governance flag signaling that the person has not yet entered the collaboration context the instrument is designed to measure. Scoring bands concentrate discrimination in the 55 to 100 range because that is where training decisions, development planning, and readiness assessment operate. The 40 to 54 band addresses the overlap zone where interface familiarity gaps and genuine capability gaps most often coincide.

5.6 Model Version Context and Temporal Reliability

Every AIS snapshot carries a model version context field documenting the specific platform versions used in that administration. An AIS produced against one set of model versions is not directly comparable to one produced against substantially upgraded versions without explicit version-transition notation. The multi-platform requirement structurally reduces single-model instability risk. Version documentation provides the audit trail that makes longitudinal comparisons interpretable.

Section 6: Preliminary Validation & Planned Studies

6.1 Content Validity

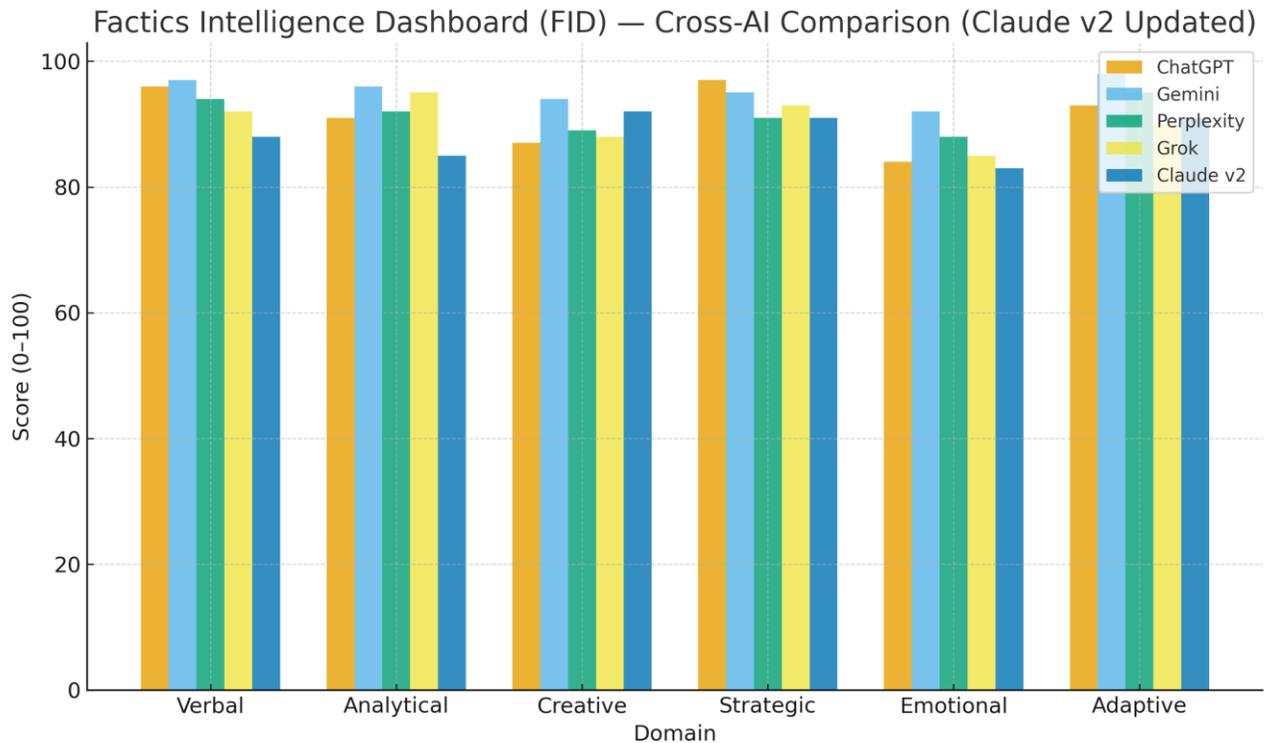
Content validity asks whether the instrument's dimensions cover the conceptual space of the construct. HEQ's four dimensions map directly to Gartner's definition of augmented intelligence as a model enhancing cognitive performance including learning, decision making, and new experiences. CAS covers cognitive performance enhancement. EAI covers decision-making integrity under AI influence. CIQ covers the quality of human-AI experience co-creation. AGR covers learning trajectory. The dimensional architecture is bounded by the Gartner definition and does not claim coverage beyond it.

6.2 Reliability Evidence: ICC = 0.96 from Multi-AI Protocol

Note: All HEQ scores reported in this section are illustrative pilot outputs from a single-subject feasibility study. Values are non-normative and not population-benchmarked. They demonstrate protocol behavior and dimensional differentiation, not validated population norms.

The cross-platform consistency coefficient (ICC) of 0.96 from Study 2 is the primary reliability evidence. It was computed using an Intraclass Correlation Coefficient, two-way random effects model, absolute agreement, average measures, on the 5x4 score matrix produced by administering identical HEQ prompts across five AI platforms. ICC of 0.96 indicates that the platforms, treated as raters, converge on the same assessment of the same underlying construct.

This coefficient measures methodology reliability, specifically the behavior of structured assessment prompts administered identically across independent AI architectures. It documents that the multi-AI administration protocol produces stable, reproducible results for a single subject across five raters. A Q4 2025 EOY audit extended administration to nine platforms, with preliminary replication of the consistency finding, expanding the platform independence evidence base from five to nine architecturally distinct systems. Population-level reliability, which requires ICC computation across multiple subjects, is not established by either the five-platform or nine-platform finding. That is precisely the purpose of the planned Q1 2026 multi-user validation study.



6.1: HAIA Intelligence Snapshot Score Comparison by Model 2025

6.3 The CIQ Lagging Pattern as Construct Evidence

Across all five platforms in Study 2 and across all ten informal cross-user participants, CIQ consistently scored lowest. The cross-user pilot produced a composite AIS range of 78 to 94 with dimensional variance within four points, and CIQ was identified as the lowest-scoring dimension in all ten cases.

This pattern is construct validity evidence of a specific kind. It demonstrates that the instrument detects genuine behavioral variance rather than producing uniform high scores or ceiling effects. The convergence with independent academic research strengthens the interpretation: the Harvard Business School Oversight Paradox study documented that humans are measurably more likely to defer to incorrect AI recommendations under authoritative framing (Lane et al., 2024), and Dell’Acqua et al.’s (2023) Jagged Frontier research demonstrated that AI competence varies unpredictably across task types, making trust calibration the most cognitively demanding skill in human-AI partnership. Aljaziri’s (2025) research on within-session trust calibration dynamics independently established that calibration quality predicts collaboration outcomes more reliably than trust level alone, further supporting CIQ as a distinct and consequential behavioral construct. The consistent CIQ lagging finding across architecturally distinct platforms and across independent participants indicates that CIQ is measuring something real and distinct from the other three dimensions.

Two alternative explanations must be acknowledged and are addressed in the planned validation studies. First, the CIQ evaluation criteria within the prompt may be inherently

more demanding than those of other dimensions, producing lower scores through prompt design rather than genuine behavioral variance. Second, the preliminary samples are small and author-proximate, raising the possibility that sampling characteristics rather than construct distinctiveness explain the pattern. The n=100 formal validation study will test CIQ performance across demographically diverse participants and will compare CIQ score distributions against established trust calibration instruments to assess whether the lagging pattern reflects construct-level distinctiveness or methodological artifact.

6.4 EOY 2025 Validation Extension: Nine Platforms and Longitudinal Trajectory

In Q4 2025, the HAIA assessment protocol was administered across nine platforms in a year-end validation audit, producing a composite AIS of 91.8. This extended the Study 2 five-platform finding and provided broader platform independence evidence. More consequentially, the EOY audit documented the first longitudinal HEQ trajectory across three quarters for a single subject: Q2 2025 composite of 87.5, Q3 2025 composite of 92.3, and Q4 2025 composite of 91.8. This three-point trajectory is a single-subject n=1 observation and does not constitute longitudinal validation. It is, however, the first empirical demonstration that HEQ scores vary across time in a pattern consistent with a developmental trajectory rather than producing static or random scores. The Q3 peak followed by Q4 stabilization is consistent with a learning curve pattern.

The EOY audit also recorded a human arbitration override event in which the minority AI platform position was verified as correct against primary sources, with the majority AI consensus overruled. This documented override is the first operational evidence that the HAIA-RECCLIN governance layer functions as designed: human judgment corrected a multi-platform AI consensus error. The governance integration claim in Section 5 is grounded in documented practice, not theoretical description alone.

6.5 Preliminary Evidence of Cross-User Prompt Stability

Following the n=1 feasibility studies in Study 1 and Study 2, ten individuals tested the HAIA Intelligence Snapshot prompt using personal ChatGPT accounts independent of the framework author. All ten administrations returned data consistent with expectations: composite AIS range 78 to 94, dimensional variance within four points, CIQ as lowest-scoring dimension in all ten cases, and 100 percent rubric compliance with required output fields present. This is preliminary evidence of cross-user prompt stability, not formal psychometric validation. Formal validation with n=100 or more participants across demographic contexts, with proper controls, randomization, and statistical analysis, is the 2026 priority.

6.6 Convergent Platform Recommendation as Construct Evidence

Study 3's finding that all five architecturally distinct platforms independently recommended adding a reliability meta-dimension to the four-dimension instrument is a form of construct validity evidence rarely available in instrument development. Five independent AI systems,

operating without knowledge of each other's outputs, identified the same structural gap. That convergence supports the theoretical claim that a reliability meta-layer is necessary to the construct rather than peripheral to it. This finding is the empirical origin of the fifth dimension in HEQ5 (Societal Safety in the enterprise deployment context, or RCI in the assessment context).

6.7 Advanced Protocol Revalidation Opportunity

Study 3 documented that the advanced protocol achieved only 25 percent cross-platform execution success in September 2025. The primary failure mode was platform memory architecture: platforms operating with session isolation could not perform the historical analysis the adaptive protocol required. Platform capabilities have changed materially since then. Multiple major platforms now offer persistent memory and conversation history features. The 2026 Research Agenda includes a revalidation study of the adaptive protocol under current platform conditions, which would enable true longitudinal AIS tracking across months rather than sessions.

Section 7: Positioning Against Adjacent Frameworks

7.1 The AIQ Framework: Closest Academic Parallel

Ganuthula and Balaraman’s Artificial Intelligence Quotient (AIQ) framework (arXiv:2503.16438, 2025) is the closest published parallel to HEQ. Both propose quotient-style instruments for measuring human capability in the AI collaboration context. Both emerged in 2025 from independent research lines. Both identify the absence of standardized individual-level measurement as the primary gap they address.

AIQ focuses primarily on measuring current proficiency, what a person knows and can do with AI today. HEQ measures enhancement trajectory, how capability grows through sustained collaboration over time. AIQ is a competency snapshot. HEQ is a development instrument. Both are needed. HEQ’s specific differentiators are three: governance architecture native to the measurement design through HAIA-RECCLIN, multi-platform administration required as a construct-level necessity, and a single composite score AIS with a decomposable dimensional spine designed for longitudinal tracking.

7.2 Comparative Framework Table

Framework	Unit of Analysis	Output Type	Time Horizon	Multi-Platform	Governance Integration
HEQ/AIS (This paper)	Individual	Composite score + dimensional breakdown	Longitudinal, trajectory	Required (min. 3)	Native, HAIA-RECCLIN
AIQ (Ganuthula & Balaraman, 2025)	Individual	Multi-dimensional factor scores	Cross-sectional, proficiency	Not specified	Not integrated
Hybrid Intelligence (Dellermann et al., 2019)	Team/System	Qualitative framework	Contextual	Not applicable	Design principles only
CI-TeX Scale (Zabel et al., 2025)	Individual	Behavioral engagement score (active user contribution)	Session-level	Not applicable	Not integrated
Collaborative AI Literacy/Metacognition (Sidra & Mason, 2025)	Individual	Validated skill scales	Cross-sectional	Not specified	Not integrated
HAI-Eval (human-AI evaluation benchmark protocols)	Human-AI team	Benchmark scores	Cross-sectional	Not specified	Not integrated
McKinsey SCI (Yee et al., 2025)	Occupation cluster	Automation exposure score	Longitudinal, macro	Not applicable	Policy-level

The CI-TeX Scale (Zabel et al., 2025) measures collaboration intensity as a behavioral construct in text-based human-chatbot interactions, validated across N=599 participants with IRT grounding and a confirmed unidimensional structure. It is the first validated instrument for text-based collaboration intensity with conversational AI. The construct boundary between CI-TeX and HEQ's CIQ is precise: CI-TeX measures depth of active collaboration behaviors, how much a person engages in the collaboration process. CIQ measures calibrated reliance, the quality of judgment about when and how much to trust and extend AI output. These are related but distinct. CI-TeX is a complementary behavioral measure, not a substitute for CIQ, and represents a promising external validation instrument for the CIQ dimension in the 2026 criterion validity study.

Sidra and Mason's (2025) Collaborative AI Literacy and Collaborative AI Metacognition Scales provide validated instruments for CAS-adjacent and AGR-adjacent constructs respectively. The Literacy Scale captures the ability to direct, contextualize, and refine AI outputs, directly adjacent to CAS's cognitive agility framing. The Metacognition Scale captures self-monitoring and evaluation of thinking during AI interaction, directly adjacent to AGR's feedback integration framing. Both are complementary instruments that could serve as external validators for their respective HEQ dimensions in the 2026 criterion validity study.

As illustrated, HEQ is the only framework that integrates individual-level measurement, longitudinal trajectory tracking, multi-platform administration, and a native governance layer within a single instrument.

7.3 The IJRTI Usage: Phenomenon versus Measurement Science

Earlier descriptive uses of the phrase "Human-AI Collaborative Intelligence" appear in education research contexts to describe collaborative phenomena without establishing a measurement discipline, canonical dimensions, or standardized output. HACI establishes the measurement science built to quantify those phenomena. The existence of prior descriptive usage confirms that the phenomenon HACI measures is recognized across independent research contexts, predating this framework's formal operationalization.

7.4 HACI as the Measurement Science Layer

Each adjacent framework identifies a genuine aspect of human-AI collaboration: collaborative performance, task synergy, productivity impact, trust calibration, skill exposure. None provides the individual-level standardized score with governance integration and longitudinal tracking that enterprise deployment requires. HACI is positioned not as a competitor to these frameworks but as the measurement science layer that operationalizes what each of them describes.

Section 8: Limitations, Ethical Scope, and Validation Roadmap

8.1 Threats to Validity

Prompt Gaming and Coaching Effects

Individuals aware of the AIS scoring criteria could potentially craft responses that score well without reflecting genuine collaborative intelligence. Mitigation: the multi-AI minimum requirement makes consistent gaming across three architecturally distinct systems substantially more difficult than gaming a single platform, as maintaining coherent performance across architecturally distinct AI systems with different evaluation conventions requires the very collaborative intelligence HEQ measures. The Bait Injection protocol, administering prompts with intentional errors embedded in AI outputs, measures actual error detection behavior rather than self-reported competency.

Platform Coupling

Five platforms used in validation testing share exposure to overlapping public data and similar safety training frameworks. True architectural independence is a matter of degree. Mitigation: the independence criteria in Section 5.2 require documented behavioral divergence. The compliance distribution data showing range from 32 to 98 percent across platforms confirms meaningful behavioral differences.

Construct Drift

As AI platforms update their models, the behavioral context in which AIS is measured changes. Mitigation: the model version context field in every AIS snapshot documents the specific platform versions used. Quarterly protocol reviews assess whether material platform changes require recalibration studies.

Criterion Leakage

If AIS administration tasks closely resemble the performance tasks used to validate AIS against external criteria, criterion validity findings may inflate due to shared method variance. Mitigation: the 2026 criterion validity studies will use performance measures collected through observation, manager rating, and work sample analysis independent of the AIS administration tasks.

Independent Replication

All preliminary validation data in this paper originates from the framework author's research program or informal volunteer participants directly connected to the author. No independent research group has yet administered the HEQ instrument, replicated the cross-user stability findings, or validated the AIS against external performance criteria. This is standard for early-stage instrument development and is explicitly acknowledged here as a transparency commitment. The validation partner model described in the Scoped Contribution Statement is the pathway to independent replication. Organizations and researchers who administer

HEQ independently and contribute their data to the validation corpus become the independent replication base the instrument requires before its practical utility claims can be formally substantiated.

8.2 Practical Scope: Intended and Prohibited Uses

Intended Uses

AIS is designed as a developmental indicator for individual coaching, training program design and validation, AI adoption readiness planning, and longitudinal tracking of collaborative intelligence growth.

Prohibited Uses

AIS must not serve as the sole or primary factor in hiring, termination, or promotion decisions. AIS must not be used as a credential without completion of the psychometric validation studies specified in the 2026 Research Agenda. AIS must not be administered without participant disclosure of what is being measured and how results will be used.

Cognitive Data Ownership

AIS production involves analysis of an individual's prompting behavior, reasoning patterns, and error detection responses. Organizations deploying HEQ must establish clear data ownership policies, specify retention and deletion schedules, obtain explicit informed consent, and comply with applicable data protection frameworks including the EU AI Act's requirements for high-risk AI applications in employment contexts.

8.3 Evidence Status and Validation Roadmap

The evidence reported in Section 6 is preliminary. It consists of a single-subject $n=1$ feasibility study, an informal ten-person cross-user pilot, and a Q4 2025 nine-platform extension. None of these constitute formal psychometric validation of the HEQ instrument. The planned studies described below are not yet completed. Until the 2026 validation program produces results, all HEQ dimensional scores should be interpreted as protocol demonstrations rather than psychometrically validated measurements.

Quarter	Study	Target	Success Criterion
Q1 2026	Advanced Protocol Revalidation	Test Hybrid-Adaptive v3.1 on memory-enabled platforms	greater than 80% cross-platform execution vs. 25% in Sept. 2025
Q1 2026	Formal Multi-User Validation	n=30 minimum across diverse professional backgrounds	Cronbach's alpha greater than 0.75 for composite AIS
Q2 2026	Criterion Validity Study	Correlate AIS with manager ratings and work samples	Statistically significant correlation, r greater than 0.35
Q2 2026	Bait Injection Protocol Validation	Formalize CIQ operational test with embedded errors	RCS measurement stable across three administrations, variance under 5 points
Q3 2026	Longitudinal Tracking Study	Track AIS trajectory over 90 days with memory-enabled platforms	Measurable AGR improvement in structured training group vs. control
Q4 2026	Cross-Cultural Validation Planning	Non-Western platform testing and international advisory board formation	Testing protocol established for Ernie Bot and YandexGPT

The longitudinal tracking study carries the highest validation priority in the 2026 agenda, and its design warrants additional specification. The study will recruit a cohort of knowledge workers engaged in sustained AI-collaborative work across a 90-day period. Participants will complete capability assessments at three time points, with at least one assessment conducted without AI access to test whether capability gains persist independently of the tool. HEQ scores will be tracked longitudinally using memory-enabled platforms, allowing AGR to be measured as a genuine developmental trajectory rather than a session-level approximation. A matched control group completing equivalent work without AI collaboration will provide the comparison baseline necessary to attribute any capability change to AI collaboration rather than general professional development. This study will produce the first empirical test of the paper's central distinction: a productivity gain that disappears when the AI is removed represents fluency, while a capability gain that persists represents intelligence. Until that study is completed, the fluency versus intelligence claim remains a well-grounded hypothesis rather than a demonstrated finding.

References

Theoretical Foundations

- Artman, H., & Garbis, C. (1998). Situation awareness as distributed cognition. In Proceedings of the 9th European Conference on Cognitive Ergonomics (ECCE9) (pp. 151–156).
- Bansal, G., Nushi, B., Kamar, E., Weld, D. S., Lasecki, W. S., & Horvitz, E. (2019). Updates in human-AI teams: Understanding and addressing the performance/compatibility tradeoff. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1), 2429–2437.
- Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Weld, D. S., & Horvitz, E. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. CHI Conference on Human Factors in Computing Systems. <https://doi.org/10.1145/3411764.3445717>
- Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. W.W. Norton & Company.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56(2), 81–105.
- Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (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 No. 24-013.
- Dellermann, D., Ebel, P., Sollner, M., & Leimeister, J. M. (2019). Hybrid intelligence. Business & Information Systems Engineering, 61(5), 637–643.
- Dweck, C. S. (2006). Mindset: The new psychology of success. Random House.
- Engelbart, D. C. (1962). Augmenting human intellect: A conceptual framework. Stanford Research Institute.
- Ganuthula, V., & Balaraman, A. (2025). Artificial Intelligence Quotient (AIQ): A novel framework for measuring human-AI collaborative intelligence. Discover Artificial Intelligence, 5, 51. <https://doi.org/10.1007/s44163-025-00516-1>
- Gardner, H. (1983). Frames of mind: The theory of multiple intelligences. Basic Books.
- Goleman, D. (1995). Emotional intelligence: Why it can matter more than IQ. Bantam Books.
- Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics, 4(1), 53–71.
- Kasparov, G. (2017). Deep thinking: Where machine intelligence ends and human creativity begins. PublicAffairs.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80.
- Licklider, J. C. R. (1960). Man-computer symbiosis. IRE Transactions on Human Factors in Electronics, 1, 4–11.
- Mendoza, N. B., & Yan, Z. (2025). From beliefs to behaviors: Conceptualizing and assessing students' practices that reflect a growth mindset. Social Psychology of Education, 28(1). <https://doi.org/10.1007/s11218-025-10032-w>
- Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence. Cambridge University Press.
- Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Scientific Validation Research

- Aljaziri, M. A. (2025). Trust calibration in human-AI teaming: Within-session dynamics, transparency, and performance effects (Graduate thesis). Rochester Institute of Technology Digital Repository. <https://repository.rit.edu/theses/12379/>
- An, T. (2025). AI as cognitive amplifier: Rethinking human judgment in the age of generative AI. arXiv preprint arXiv:2512.10961.
- Lane, J. N., Boussioux, L., Ayoubi, C., Chen, Y. H., Lin, C., Spens, R., Wagh, P., & Wang, P. H. (2024). Narrative AI and the human-AI oversight paradox in evaluating early-stage innovations (Working Paper No. 25-001). Harvard Business School.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Lawrence Erlbaum Associates.
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. MIT Economics Working Paper.
- Sidra, S., & Mason, C. (2025). Generative AI in human-AI collaboration: Validation of the Collaborative AI Literacy and Collaborative AI Metacognition Scales for effective use. *International Journal of Human-Computer Interaction*. <https://doi.org/10.1080/10447318.2025.2543997>
- Yang, B., Wang, Y., & Li, X. (2025). A token-efficient framework for codified multi-agent prompting and workflow execution. arXiv preprint arXiv:2507.03254.
- Zabel, S., Meske, C., Poecze, F., & Stracke, C. M. (2025). Being just used or truly understood: A measure of users' collaboration intensity with chatbots. *International Journal of Human-Computer Studies*. <https://doi.org/10.1016/j.ijhcs.2025.103520>

Workforce and Governance Research

- Gartner. (2019). Gartner says AI augmentation will create \$2.9 trillion of business value in 2021. Gartner Newsroom. [gartner.com](https://www.gartner.com/newsroom/2019-01-24-gartner-says-ai-augmentation-will-create-29-trillion-of-business-value-in-2021)
- Gartner. (2024). Definition of augmented intelligence. IT Glossary. [gartner.com/en/information-technology/glossary/augmented-intelligence](https://www.gartner.com/en/information-technology/glossary/augmented-intelligence)
- 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.
- European Union. (2024). Regulation (EU) 2024/1689 (AI Act). Official Journal of the European Union.
- National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0). [nist.gov](https://www.nist.gov/ai-rmf)

Sources from Previous Work

- Puglisi, B. C. (2012). *Digital Facts: Twitter*. Digital Media Press. magcloud.com/browse/issue/471388
- Puglisi, B. C. (2024, February). The intelligence enhancement thesis. basilpuglisi.com/facts-make-us-more-intelligent
- Puglisi, B. C. (2025a). From metrics to meaning: Building the Facts Intelligence Dashboard. basilpuglisi.com/fid
- Puglisi, B. C. (2025b). The Human Enhancement Quotient: Measuring cognitive amplification through AI collaboration (v1.0). basilpuglisi.com/HEQ
- Puglisi, B. C. (2025c). *Governing AI: When capability exceeds control*. Amazon.
- Puglisi, B. C. (2025d). The Human Enhancement Quotient (HEQ): Measuring collaborative intelligence for enterprise AI adoption (White Paper v4.3.3). basilpuglisi.com/HEQ

Puglisi, B. C. (2025e). HAIA-RECCLIN: The multi-AI governance framework for responsible AI. basilpuglisi.com/haia-recclin

Puglisi, B. C. (2025f). Measuring human-AI collaborative intelligence: Empirical findings on universal assessment methodology. basilpuglisi.com

Note: Figures presented throughout this paper reflect data collected during 2025 study runs. As additional platform assessments are completed and integrated, updated visualizations will be published. Current figures are included to provide visual context for the framework's measurement methodology.

Appendix B: HEQ5 — Organizational Extension

The HEQ four-dimension instrument described in this paper is designed for individual-level assessment. Enterprise deployment contexts require an additional governance dimension that the individual instrument does not assess: the degree to which a person's AI collaboration behaviors account for societal impact and organizational responsibility beyond their immediate task.

HEQ5 adds Societal Safety (SS) as a fifth dimension to the organizational deployment framework. The HEQ5 formula is: $HEQ5 = (CAS + EAI + CIQ + AGR + SS) / 5$.

Societal Safety measures observable behaviors indicating that the person's AI-augmented work accounts for downstream effects, bias amplification risks, and organizational accountability requirements. It is assessed at the organizational deployment level rather than the individual assessment level, because the relevant behavioral signals require organizational context that a standalone individual assessment cannot access.

The individual HEQ instrument documented in this paper produces a four-dimension AIS. Organizations deploying HEQ at scale should refer to the Enterprise Administration Protocol and HAIA-RECCLIN framework documentation for the full HEQ5 specification, scoring rubric, and governance integration requirements.

Appendix A: Attribution and Ethical Use Notice

Any use of HAI, HEQ, AIS, HAIA-RECCLIN, or related frameworks in research, publications, training materials, or commercial applications must include proper attribution to Basil C. Puglisi and this working paper. Commercial use requires written authorization.

Trademarks: Human Enhancement Quotient, HEQ, Augmented Intelligence Score, AIS, HAI, HAIA, Facts, and HAIA-RECCLIN are trademarks of Basil C. Puglisi.

Contact: me@basilpuglisi.com | basilpuglisi.com | GitHub: github.com/basilpuglisi/HAIA

© 2026 Basil C. Puglisi. All rights reserved.