WORKING PAPER

Measuring the Governance Competence at the Center of AI Literacy

The Human Enhancement Quotient (HEQ), the Augmented Intelligence Score (AIS): Connecting AI Literacy and Cognitive Development through Governance

Basil C. Puglisi, MPA, Independent Practitioner, basilpuglisi.com
A Human-AI Collaboration | September 2026

Working Paper v6.0.5 | September 2026

This working paper develops and extends the author's earlier *Measuring Augmented Intelligence: Theoretical Foundations and Empirical Development of the HEQ and AIS* (SSRN Abstract 6351478, working paper February 2026, posted to SSRN March 2026).

Abstract

This paper introduces the Human Enhancement Quotient (HEQ), a behavior-anchored framework that measures the governance competence at the center of AI literacy: the ability of a human to direct, challenge, verify, and own the work done with AI. Where the major AI literacy frameworks name governing and managing AI as a competence and stop short of measuring it, and where most formal definitions of intelligence describe a single agent that predicts, compresses, or acts (Legg & Hutter, 2007), HEQ scores the governance relationship between a human authority and a machine capability inside a working decision process. It scores the process rather than the answer, which makes the measurement portable across value systems and able to credit a human who is wrong through sound method while catching a human who is right through none. The instrument defines four behavioral dimensions, Cognitive Agility Speed, Ethical Alignment Index, Collaborative Intelligence Quotient, and Adaptive Growth Rate, combined into an equal-weighted Augmented Intelligence Score (AIS), with the growth dimension carrying the developmental claim that governed practice builds the competence over time. The accompanying scoring rubric specifies the behavioral anchors, the validity controls that detect rubber-stamping, a universal structural floor keyed to irreversible human consequence, and a graded band that is regional and value-laden. Three deployment modes, Independent, Personal, and Professional, share one set of control questions so results sit on a comparable spine, with the Independent clean-slate mode serving as the enterprise model. The paper presents two author's notes addressed to the economic and the human reader, a part on the stakes that make governance measurement necessary, the foundation as theory, the rubric as methodology, four appendices containing the operational assessment prompts, and a glossary. The AIS is an indicator inside a human-supervised process and never the decision itself. It serves three purposes: developing the individual, refining the organization's training, and giving the organization oversight of its own governance. Two claims are kept distinct throughout. That governed human-AI practice develops capability is grounded in decades of learning science; that HEQ measures that development accurately is presented as a deployed, research-grounded instrument at diagnostic stage, not yet validated on an independent cohort.

Keywords: AI literacy, governance competence, human enhancement quotient, augmented intelligence score, checkpoint-based governance, human oversight, outcome neutrality, AI governance, human-AI collaboration.

Contents

The Economic and Human Drivers Behind HEQ/AIS

Two author's notes open this paper, one addressed to the reader who carries financial and operational risk, and one to the reader who carries responsibility for people. They make the same case from two directions: that measuring how humans govern AI is the act that both secures financial liability and protects human employment.

Author's Note to the Economic Reader

I did not set out to build an instrument. I set out to use AI well, and I spent much of 2025 refining the method that made my own work better. Then a wave of layoffs arrived with a justification attached: companies claimed they had people who were not AI-trainable. My first reaction was not sympathy or outrage. It was a question. How do you know that? How do you justify it? What did you measure?

There was no answer because there was nothing to measure with. That is the gap I want you to see, because it is not a philosophical gap. It is an economic one. Organizations are spending heavily on AI while flying blind on the single variable that decides whether that spend produces value or liability: how well their people govern the machine. The spend is measured. The tooling is measured. The human governance sitting between an AI output and a real decision is not measured at all. And everything you want to do with AI depends on measuring it. What cannot be measured cannot be priced, and what cannot be priced cannot be insured, defended, hired against, or improved. It can only be feared, retained, and guessed at, which is where most organizations sit right now.

That single gap produces seven exposures, and they compound. AI work is effectively uninsurable, because risk that cannot be measured cannot be underwritten. The exceptions prove the rule. Three standalone AI liability products existed worldwide as of early 2026, their limits ran from 9.25 to 25 million dollars, and the carriers writing them underwrote governance evidence directly, with Armilla requiring ongoing model quality assessments as a condition of coverage, while exclusion endorsements spread across standard lines faster than dedicated coverage appeared (Insurance Business, 2026; The Insurer, 2026). Counts published weeks later put the figure at five by admitting performance guarantees and their distribution arrangements, which widens the boundary from dedicated third-party liability paper to any AI-specific product sold on its own terms. Neither count describes a market where dedicated coverage keeps pace with exclusion, and the forward view says it will not: a vendor forecast published in September 2026 projects that 60 to 80 percent of new policies and renewals in errors and omissions, directors and officers, employment practices liability, and cyber will factor AI risk into underwriting by 2028, while most midsize insurers continue covering that risk through existing lines rather than standalone policies (ScienceSoft, 2026). Liability cannot be defended, because when an output causes harm the question is who governed the decision, and most organizations have no record of whether a human exercised control or rubber-stamped the machine; courts and regulators are already moving on exactly this point. Talent cannot be recruited against a standard, because there is no ACT, SAT, or GRE for AI, so every hire is a guess and the strongest governor and the most dangerous operator look identical on paper.

Consumer trust cannot be answered, because an organization that cannot show a human was meaningfully in control has no reply to the claim that the AI did it. Training cannot be evaluated,

because spend without a baseline is faith. Workforce return cannot be monitored, and the deeper cost there is misattribution: when an AI initiative fails, the organization blames the easiest target, the employee, when the failure may belong to the method or the product. And none of it can be improved, because improvement requires measurement.

I built HEQ to turn those seven assertions into something testable, and I built the AIS to make the measurement quantifiable, auditable, and comparable. The economic argument is simple. HEQ converts an unmeasured, unpriceable liability into a measured, manageable asset. Risk that can be measured can be priced, which makes AI work insurable and moves retained exposure off the balance sheet. Governance that can be documented makes liability defensible and ends the practice of blaming workers for failures that belong to the method or the tool. A standard signal makes hiring a decision rather than a guess. A baseline makes training spend provable. Attribution makes the AI line item evaluable. The instrument earns its cost the first time it prices one policy, prevents one wrongful-termination claim, or retires one failing product that was being charged to staff.

And I want to be precise about the productivity claim, because it is the part the market oversells. I am not promising you quantity. Governed human-AI work does produce more, but volume is the byproduct, not the point.

The point is reliable quality: output that can be trusted because a human governed it, because the sources hold, because uncertainty was preserved instead of smoothed over, and because someone is accountable for what was released. Quantity without governance is just more unverified output and more exposure. Quality is what survives a regulator, a court, a customer, and a board. That is the return. Not faith that AI helps, but the ability to see, price, and improve the human half of your AI investment, so that what you produce with AI is not merely more, but reliably worth releasing.

Author's Note to the Human Reader

We set our measurements about outcomes and then wonder why the substance comes out weak, disconnected, or unrelatable. We repeat lines like it is not the destination, it is the journey, and yet we engage with math, science, and technology in a way built entirely on outcomes, and then wonder why it does not bridge to culture, religion, morals, and ethics.

I spent much of 2025 refining the journey, working to produce better outcomes for my own AI use. The same layoffs that raised an economic question for me raised a human one, and the human one cut deeper. A company decided certain people were not worth keeping and reached for a machine to replace them, and I rejected the premise underneath that decision then, and I reject it still. Not because the machine cannot do some things better. Not because it cannot be more efficient. Because the machine cannot be human. After using fifteen models across thirteen platforms through their various stages of development over several years, that is my truth. Agentic AI failed not because the technology was weak but because it assumed a machine could replace a human in an environment built for humans.

Augmented intelligence is the idea that humans and machines working together produce something neither produces alone. When we try to measure it today, what we face is a failure of method, not a failure of the concept. In many cases there was no method at all, only an expectation that code could replicate humanity. It cannot. The question our time faces is what the

world looks like with humans and artificial intelligence together, whether that intelligence is narrow, general, or one day a superintelligence. To answer it in a world as diverse as ours, the measurement tool itself must adapt to cultural, structural, and regulatory environments that hold vastly different values. If we fail at this, AI becomes a contest in which a logical machine decides for us from the code of its most dominant developer, and that may have nothing to do with money or power in the way we usually mean those words.

This is why HEQ measures the governance of the machine by the human, in a relationship where the human stays in control through checkpoints and meaningful oversight. The design places its emphasis on humanity, so that augmented intelligence can be measured not only for productivity but for human cognitive development and advancement, with the honesty that we are not all equal and we do not all share the same values. With the human governor at the center, we get AI governance designed for cross-cultural portability, because the human is the responsible party for the relationship. That portability remains a validation question. The human in a given institution, organization, or community carries the burden of the culture, religion, morals, and ethics of the setting they operate in, which is what lets the same instrument hold in settings whose values and laws differ completely. The Augmented Intelligence Score tells us how well that human-AI collaboration is governed, and it supports a method of auditable accountability and advancement for everyone it touches.

The HEQ measures the method governance. The AIS gives us the feedback to monitor, adjust, and adapt. This is how we begin to understand AI literacy, and more importantly, how we change methods and behavior to practice it effectively, efficiently, and safely, and how we make sure that the people a company was ready to discard are seen for what they can actually do.

Part I. The Stakes: Why Governance Must Be Measured

This part connects the measurement instrument to the economic and legal forces that make measuring governance necessary. It summarizes an argument developed in full in a companion paper and points the reader there for the complete treatment.

Those forces explain why the measurement matters, and they do not make the score the decision. The AIS is an indicator a named human weighs alongside everything else they know about the work. The organization that holds it uses the score three ways: to develop the individual, to refine its own training, and to oversee its own governance. A poor score is context for a supervisor, not a verdict on a person.

The foundation establishes what HEQ measures and why the measurement is possible. This part addresses a different question: why the measurement is now economically and legally required, rather than merely useful. The answer is that the human governance of AI has become the variable on which insurability, liability, hiring, and organizational return all turn, and none of those can be acted on while the variable remains unmeasured.

The economic case is developed at length in a companion working paper, The AI Risk Economy: Why Insurance Cannot Price What Governance Cannot Prove, and is summarized here only as far as it bears on measurement. Its argument begins from a structural observation the author has named the Economic Override (Puglisi, 2025, November): when safety, ethics, or governance conflict with economic incentive, the economic incentive tends to prevail wherever no countervailing enforcement mechanism exists. Voluntary commitments do not survive that pressure. What counters it is a cost attached to the absence of governance, and the insurance market is becoming the fastest such mechanism, because it operates at the point of contract renewal and requires no legislative process. Carriers are already excluding what they cannot price and offering conditional coverage to organizations that can show governance maturity, a pattern the paper documents through carrier exclusion filings, standalone AI liability products, and the adoption of the NAIC AI Model Bulletin across two dozen states (Lior, 2025; NAIC, 2023).

The companion paper organizes that market response into a five-tier insurance maturity model, distinguishing organizations with no AI policy, published ethical principles, automated technical controls, a named human checkpoint with binding authority, and structured audit records. The decisive boundary in that model is the one between technical traceability and named human accountability, the line between a system that checks itself and a human who answers for the decision. That boundary is precisely what HEQ measures. An organization can only establish it sits above that line if it can show, with evidence, that its people govern AI rather than rubber-stamp it, and that demonstration requires an instrument that scores the governance behavior and leaves an auditable record. The paper's central finding is an actuarial gap: no published study yet quantifies governance maturity as a pricing factor, because the market lacks the verification infrastructure that would resolve the information asymmetry between governed and ungoverned organizations. A behavior-anchored measure of human governance, with a record behind every score, is one form of that missing infrastructure.

The same gap produces exposure that is no longer hypothetical, and it is appearing in litigation and regulation on parallel tracks. When an AI-assisted output causes harm, the question a court

asks is who governed the decision, and an organization with no record of whether a human exercised control or merely approved the machine cannot answer it. Courts have begun permitting discovery into whether AI was used to supplant human decision-making, where a review-and-approve workflow that captures no substance of the review resembles, in the record, a rubber stamp.[1]

The defense that the machine rather than the organization made the error is also losing ground, with at least one European court treating an AI system's statements as the deploying company's own and rejecting the argument that users were expected to verify them.[2]

Regulation points the same way, toward documented human oversight rather than asserted control: the European Union's AI Act will require meaningful human oversight of high-risk systems from December 2027, and recent United States state law will give a consumer who receives an adverse automated decision the right to request meaningful human review and reconsideration, to the extent commercially reasonable, by a trained reviewer with authority to override it.[3] Across credit, insurance, and employment, the common requirement is a human who can be shown to have governed the decision, which is the thing HEQ is built to measure.

The exposure runs through the workforce as well, and it produces a quieter cost: misattribution.

Without a measure of governance quality, an organization cannot see which of its people turn AI into value and which turn it into risk, cannot tell calibrated reliance from rubber-stamping, and cannot connect its AI spend to any outcome it can name. When an AI initiative underperforms, the organization reaches for the easiest target, the employee, when the failure may belong to the method or the product. A measure of human governance isolates the human-performance variable and separates it from the method and the tool, which removes the reflexive scapegoat and points the diagnosis at the actual cause. A person governing well with a failing product is a finding an organization can act on. A person rubber-stamping a sound one is a different finding entirely. Only a measure of governance can tell them apart, and the cost of confusing them is paid in the wrong people dismissed and the wrong tools retained. This is the operational form of the same claim that opens this paper: that a company deciding certain people are not AI capable,

---

[1] *Estate of Gene B. Lokken v. UnitedHealth Group, Inc.*, No. 23-CV-3514 (D. Minn.), discovery order of March 9, 2026, permitting discovery into whether the nH Predict tool was designed to supplant physician decision-making in coverage denials.

[2] *Landgericht München I*, preliminary injunction of May 28, 2026 (Case No. 26 O 869/26, on appeal). Parallel: *Oberlandesgericht Hamm*, judgment of May 12, 2026 (Case No. I-4 UKl 3/25, leave to appeal granted). Contrast: *Walters v. OpenAI*, No. 23-A-04860-2 (Ga. Super. Ct. Gwinnett Cnty., May 19, 2025), summary judgment for the defendant on disclaimers and the plaintiff's own verification.

[3] EU AI Act (Regulation (EU) 2024/1689), Article 14, as amended by Regulation (EU) 2026/1744 in force July 27, 2026, which moves Article 14 application to December 2, 2027 for Annex III and August 2, 2028 for Annex I. Colorado SB 26-189, signed May 14, 2026, replaces SB 24-205 and takes effect January 1, 2027. Enforcement is stayed pending Attorney General rulemaking and *xAI v. Weiser*.

without an instrument that measured the thing it claims, is making a consequential and indefensible assertion.

What measurement converts, then, is an unmeasured liability into a managed one. Risk that can be measured can be priced, which makes AI work insurable and moves retained exposure off the balance sheet. Governance that can be documented makes liability defensible. A standard signal makes hiring a decision rather than a guess. A baseline makes training spend provable, and attribution makes the AI line item evaluable. None of these is a productivity claim in the usual sense. Governed human-AI work does produce more, but volume is the byproduct, not the point.

The point is reliable quality: output that can be trusted because a human governed it, because the sources hold, because uncertainty was preserved rather than smoothed over, and because someone is accountable for what was released. Quantity without governance is only more unverified output and more exposure. Quality is what survives a regulator, a court, a customer, and a board, and the conditions for that quality are what HEQ measures. The companion paper makes the economic argument in full; this paper supplies the instrument that argument depends on.

Part II. Foundation: Purpose, Theory, and Architecture

This part is the constitutional layer of the framework. It states what HEQ measures and why, the theory of the human as qualitative governor of a quantitative machine, and the architecture of checkpoint-based governance on which the measurement rests. The notes above stated the why in personal terms; what follows is the formal theory.

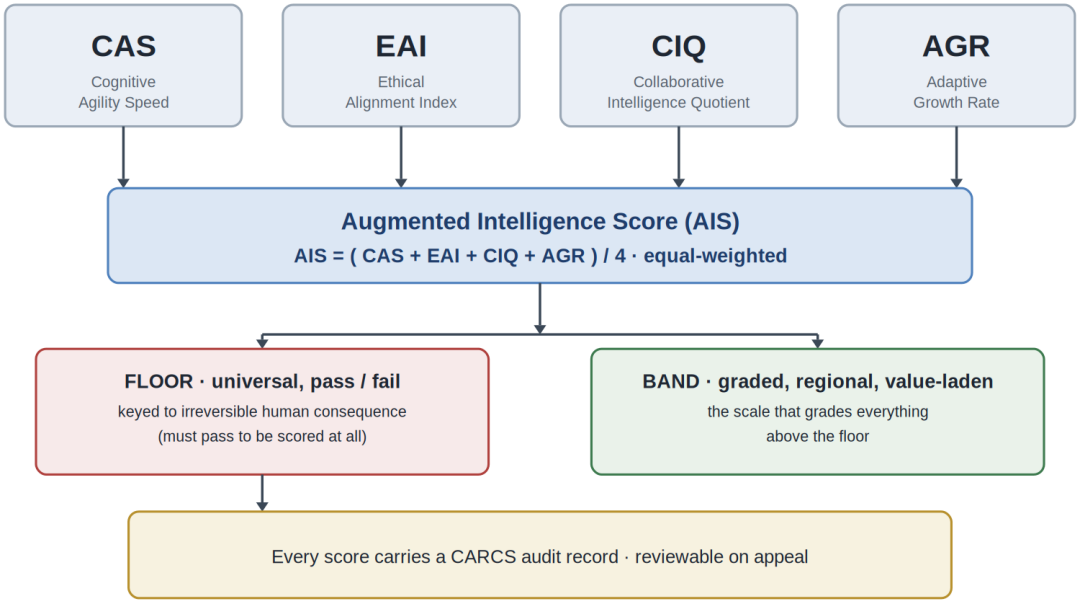

Figure 1. *The Augmented Intelligence Score architecture. Four behavioral dimensions, CAS, EAI, CIQ, and AGR, combine into an equal-weighted AIS that must pass a universal floor keyed to irreversible human consequence before it is graded on a regional, value-laden band, with every score carrying a CARCS audit record reviewable on appeal.*

1. Purpose

HEQ exists to give a quantitative, behavior-anchored understanding of augmented intelligence. It asks whether a specific person can use and manage AI well enough to create governed human-AI capability, work that neither unsupported human effort nor unguided machine output produces in the same way. In practical terms, that is AI literacy. In measurement terms, it is Human-AI Collaborative Intelligence.

This sets a hierarchy the rest of the document keeps in order. Augmented intelligence is the phenomenon. Human-AI Collaborative Intelligence is the construct. AI literacy is the practical human competency. Method governance is the observable behavior. The Augmented Intelligence Score is the quantitative score. An auditable record of the process is the audit layer.

HEQ does not reduce augmented intelligence to compliance. Governance is the observable path of measurement, not the final object being measured. The object is Human-AI Collaborative

Intelligence, the measurable capability that appears when a human directs, challenges, verifies, and applies machine capability under conscious authority.

The question it answers is the right size for the problem. Not whether collaboration is better in the abstract, not who is liable after a failure, but whether a specific person has the measurable ability to govern human-AI work, regardless of the subject matter they work in. The target is deliberately domain-agnostic. HEQ measures the ability to use and manage AI despite the subject matter. This is load-bearing, not incidental. The moment the instrument measures method rather than answers, it can do the thing the world actually needs, which is to hold the process accountable even when the human is wrong, and to work across every value system rather than encoding one.

2. Intelligence is measurable, but most formulations describe a single agent

There is no single accepted mathematical definition of intelligence, but there are many, and they are not rivals. They are lenses on one core. Prediction, compression, and minimum description length are members of the same family: to predict well you must compress the regularities that generate the data, and the shortest description is the best predictor. Bayesian inference and free energy minimization are the probabilistic face of it. Expected utility maximization, optimization, search, and control are one family of goal-directed action. Representation learning, world modeling, approximation, and generalization are the bridge, reusable internal structure that transfers to unseen cases. The Legg and Hutter (2007) universal intelligence measure is the closest thing to a unifier, fusing the prediction family with the action family into a single quantity.

This is why a model that only predicts the next token is effective across domains. Predicting well rewards compression of structure, and at sufficient scale that compression produces structure that functions as a world-model across many domains, which is what carries from sentences to code, because both are prediction over structured sequences and the structure learned for language is most of the structure needed elsewhere.

But most of these formulations describe intelligence as the capability of a single agent or system. The model predicts, the model compresses, the model acts. What they usually do not formalize is the governance relationship between a human authority and a machine capability inside a working decision process. That is the open seam in the field, and it is exactly where HEQ lives. The formulations describe the quantitative half. HEQ adds the qualitative governor.

3. The theory

The machine provides information. The human provides the wisdom of how to apply it. By wisdom, this means purpose, context, consequence, cultural judgment, ethical judgment, and final authority.

These are two different things, not two amounts of the same thing, and that distinction is the foundation. Most mathematical formulations describe the machine estimating, compressing, predicting, or selecting action: the next token, the compressed regularity, the optimal estimate, the chosen move. None formalizes the judgment of whether and how that information should be applied here, now, to this end, with these consequences. That judgment is not more information. It is a different category, and it is the human's.

One correction, learned in practice and not from theory. The machine does not simply provide information. It has access to far more information than it reliably recalls. Project rules, instructions, and uploaded files can all be present and available, and the machine will still fail to bring them to bear until a human directs it to, even when its own instructions already required it. Access and retrieval are not the same thing. So the human provides two kinds of wisdom, not one: the wisdom of what information should be present, governing the machine's unreliable recall, and the wisdom of how it should be applied. The human is the retrieval authority as much as the application authority.

4. The complementarity

Human-AI collaboration is not a blend of two similar capacities. It is a structure built on two unlike capacities with inverse strengths and inverse blind spots.

The machine is the quantitative engine. It is strong at breadth, speed, computation, and pattern. It is blind in specific ways: it recalls unreliably and will not surface what matters unless directed, it makes assumptions, it fabricates to fill gaps, it has no qualitative judgment of culture, religion, ethics, emotion, or audience, and its context is the project rather than the bigger picture.

The human is the qualitative governor. The human is strong at purpose, judgment, cultural and ethical sense, knowing what should be present, and deciding what the work is for. The human is blind in the inverse ways: limited breadth, slower, finite recall of detail, fatigue, bias, and the plain fact of sometimes being wrong.

Humanity is the qualitative value. The machine is faster at the quantitative. The human oversees because the qualitative governs the balance, and the qualitative decides what the quantitative is for. The architecture does not average these two. It places each where it is strong and uses it to cover where the other is blind.

5. The generalist and the specialist

The human is the generalist. This is the counterintuitive and correct assignment. The generalist owns the subject, the direction, the purpose, and the context of application, and does not need to hold all the content. The machine is the specialist, powerful and narrow, producing and moving content on demand without owning the why.

The relationship is directed, not automatic. The generalist directs the specialist, then confirms the result. There is no unattended loop. Reasoning may iterate, but authority does not circulate. It stops at checkpoints where the human accepts, modifies, or rejects. This is the difference from the documented failure of the human-in-the-loop model, where the person becomes a token presence inside a process that runs itself, with no real power (Buçinca et al., 2021; Vasconcelos et al., 2023). The failure is not iteration, it is unattended iteration. What governs here is the checkpoint: a hard stop where a human holds authority and decides to accept, modify, or reject, with the power to halt anything before it proceeds. Authority sits with the human at the checkpoint, never in a process.

The work proceeds as a chain of checkpoints, each with a clear owner:

- The human sets subject and direction. Authority: human.

- The machine moves content to the human. The human decides what is useful or appropriate.
- The machine discloses supporting and conflicting information. The human decides what to apply and how.
- The machine organizes the material for mapping and presentation. The human directs the form.
- The human evaluates the output and accepts, modifies, or rejects it, judged on the context, on how it is being applied, and on where it came from.

Because authority and content are different things, the generalist never has to out-know the specialist. The director does not out-act the actor. This dissolves the fear that a stronger machine makes the human obsolete: the generalist's job is direction and judgment, which deepens as the specialist grows stronger, not disappears.

The universal requirement under this structure is human authority at the decision point. Checkpoint-Based Governance is the author's formalization of that requirement, and it is where HEQ came from: a hard stop where a named human accepts, modifies, or rejects machine output with the power to halt action. The distinction matters and is named on purpose. Human checkpoint authority is the standard. Checkpoint-Based Governance is the author's design that formalizes it, which keeps it consistent with everything else in this work, the universal is the capability and the named system is one instance of it. The single most important thing the method governance standard expects is the development of human authority and oversight: the human's authority over the machine becoming more deliberate, better calibrated, and more accountable over time. Everything else in the standard serves that.

6. Recall and retrieval

The division of labor is real and runs both ways. The machine produces an output. The human recognizes that something is missing, that a source exists the machine did not surface, and points to it. The machine, now directed at a specific source, concentrates its resources and recovers the exact thing faster than the human could by reading the whole work. The human supplies the knowledge that the thing exists, which the machine could not reliably recall. The machine supplies the speed of retrieval the human could not. Each covers the other's blind spot at a checkpoint.

7. Lineage: from Factics to Checkpoint-Based Governance to HAIA

Method governance has two words, and the work came together in that order. Method is how the work is done. For the author the work method is Factics, the discipline of pairing a fact with a tactic and a measurable outcome, which predates AI and was a way of doing rigorous, evidence-anchored, accountable work in general. Governance is the human authority over the machine once a machine is doing part of that work. Applying Factics to AI is what surfaced the need for governance, because a method that demands evidence, action, and measurable results cannot hand its judgment to a system that recalls unreliably, fabricates, and has no authority to own an outcome. Checkpoint-Based Governance is the answer that requirement produced: human authority held at the decision, accept, modify, or reject.

That pairing is method governance: a sound work method, governed by real human authority. The HAIA ecosystem is what the author built by carrying it through, the operationalization that grew from asking how to apply Factics to AI responsibly, with Checkpoint-Based Governance as the authority spine and the reasoning, multi-AI, and record protocols built around it.

The two halves sit on different sides of the worked-example line, and the distinction matters. Governance has a universal standard, human authority at the decision, and Checkpoint-Based Governance is the author's formalization of it rather than the standard wearing the author's name. Method has a universal version too, rigorous evidence-paired and measurable work, and Factics is the author's specific instance of that, which places Factics on the worked-example side alongside the reasoning, multi-AI, and record protocols, not on the standard side. HEQ therefore does not require Factics by name. It requires a sound method governed by real authority, and Factics paired with Checkpoint-Based Governance is the author's proof that it can be done.

8. The governance standard: showing work and multi-AI verification

The method governance the framework expects is not a separate doctrine. It is what the four dimensions and the validity controls already measure, stated as behavior. Most importantly, the standard is the development of human authority and oversight, the human holding real authority over the machine at the decision and exercising it more deliberately and more accountably over time. Its spine is human checkpoint authority, which Checkpoint-Based Governance formalizes, named in the previous section, where HEQ came from. Beyond that spine, two further capabilities carry most of the rest, and both answer the same practical question: if the machine should surface supporting and conflicting information and fails to surface a conflict it could have found, what catches it.

The first capability is showing work, the deliberate creation of cognitive friction so outputs are challenged rather than accepted (Buçinca et al., 2021). The standard is that the human structures the interaction so the machine exposes its reasoning, its sources, and its conflicts, and so the output meets resistance before it is used. Within a single output, the machine reasons and shows its sources and its conflicts, and if it reports that there is no conflict, that is not reassurance, it is a flag the human reads as a prompt to suspect rather than to relax. This is what the engagement ladder, the source discipline, and the dissent scoring in the rubric measure: acknowledgment, stated reasons, clarifying questions, source checks, cross-source comparison, and substantive challenge are the observable grades of friction.

The second capability is multi-AI verification, distributing the catch so it does not rest on one human or one model noticing. Under responsible AI, multi-platform work is the only credibility check available, because the models check each other and no person is accountable for the result. Under AI governance a named human is responsible for every final output, so the check is the human and multi-AI is preferred practice rather than the requirement. The capability is independent verification beyond the originating output: that agreement is not trusted on its face but tested for whether the reasoning and sources actually match or are only apparently aligned, and that an independent check can refuse to verify what it cannot source. When a model hallucinates, the sources will not line up across independent models, and the honest independent response is to say plainly, I cannot verify that. This is not a vote for consensus. It is a check on whether consensus is grounded, and it is what the higher method governance levels and the multi-AI source tier in the rubric measure (cf. Lee & See, 2004; Vaccaro et al., 2024).

The third requirement, an auditable record of the process, is treated in the accountability section below. Together these are the standard: human authority and oversight at the checkpoint first, then the friction and transparency that make outputs face challenge, cross-model verification, and a record of what happened. The rubric defines the capabilities behaviorally, which is what makes their operationalization method-agnostic. A person satisfies the friction, verification, and record requirements by any means that produces the behavior, not by running any particular named system. Human authority at the checkpoint is the one part that is not optional, because without it there is no governance to measure.

The standard is human authority at the decision. Checkpoint-Based Governance is the formalization that makes that authority structural and traceable, which keeps the universal requirement distinct from one author's build of it. HEQ is built on it rather than beside it, because a score of governance competence needs a checkpoint definition to score against. The layers run in one direction. CBG is the base. RECCLIN reasoning, CAIPR cross-platform review, and HEQ itself are built on it. CARCS is the CBG audit file rather than a separate protocol. Factics is the method a human uses to reach the checkpoint with evidence-paired work in hand.

What separates this from responsible AI is the accountability tie. Responsible AI places humans in the loop, and a human in the loop is present and participating without necessarily holding authority or bearing accountability for the outcome. A person can sit in the loop and be ignored, or overridden by a confidence score. AI governance requires authority, oversight, and accountability together. What binds oversight to accountability is the named human at the checkpoint. That is what CBG supplies and what HEQ measures. The full treatment is in the Checkpoint-Based Governance specification (Puglisi, 2026, March 10). It develops the HITL contrast, the constitutional properties, and the harm boundary beyond what this paper needs.

The named implementations remain replaceable. A person satisfies the friction, verification, and record capabilities by any means that produces the behavior. RECCLIN, CAIPR, CARCS, and Factics are offered as worked examples rather than requirements. Human authority at the checkpoint is the part that is not optional, because without it there is no governance to measure.

Adjacent instruments and standards

HEQ is the behavioral measurement layer beneath the governance frameworks already in use, not a competitor to them. The NIST AI Risk Management Framework and ISO/IEC 42001 specify what a trustworthy or well-managed AI program should contain, and the OECD AI Principles set the values such a program should serve, but none of the three measures whether a particular human actually governed a particular decision. HEQ scores that behavior, and a documented AIS with its CARCS record supplies the evidence those frameworks call for without themselves measuring it. The theoretical lineage is the joint cognitive systems and distributed cognition tradition (Hutchins, 1995; Hollnagel & Woods, 2005), in which a human and a machine form one cognitive system, the human supplying contextual and ethical framing and the machine supplying computational bandwidth. HEQ operationalizes the human-governor half of that system as a measurable competency.

The same relationship holds for the AI literacy frameworks. The major frameworks converge on a shared shape: understand AI, use it, critically evaluate it, create with it, and engage ethically,

scaffolded as knowledge, skills, and attitudes (Long & Magerko, 2020; Ng et al., 2021; UNESCO, 2024; OECD & European Commission, 2026; U.S. Department of Labor, 2026). Every one of them names governing or managing AI as a competence. The UNESCO framework includes a human-centred mindset and responsible use, the OECD AILit framework names a Manage AI domain, and the United States Department of Labor framework lists managing AI responsibly among its content areas. None of them measures whether a particular person actually does it, in a real decision, with an accountable record. They name the competence and stop at the naming. HEQ measures it. The author has argued at length elsewhere that these frameworks reach the vocabulary of governance and decline to require or measure the structure that would make it real, naming friction, scaffolding, and the human checkpoint and then leaving each one optional (Puglisi, 2026, June 21). HEQ is the instrument for the competence those frameworks defer: not a rival definition of AI literacy, but the measurement of the governance competence at its center, the one the frameworks agree matters and none of them scores.

The mapping is concrete. Each row names a requirement those frameworks already state and the observable HEQ supplies for it.

Framework requirement	What it specifies	HEQ contribution
NIST AI RMF, Govern 2, roles and responsibilities	That accountability for AI outcomes is assigned and documented	A named human at the decision, with the override recorded and the arbiter identified
NIST AI RMF, Govern 3, workforce capability	That personnel are trained and resourced for their AI roles	A behavioral score for whether a specific person governs, not whether training occurred
NIST AI RMF, Measure 3, mechanisms for tracking	That performance and risk are tracked over time	A CARCS record per administration and an AGR reading across administrations
ISO/IEC 42001, clause 7.2, competence	That persons doing work affecting AI performance are competent	An individual-level instrument where the standard states an organizational duty
ISO/IEC 42001, clause 7.3, awareness	That persons understand the policy and their contribution to it	Engagement depth and stated reasons as observed behavior rather than attestation
OECD AI Principles, accountability	That actors are accountable for systems across the lifecycle	Decision-level evidence of who governed which decision and on what basis
EU AI Act, Article 14, human oversight	That high-risk systems are overseen by natural persons	Override rate and correctness, and the consequence floor, as the observable of oversight

9. Process, not outcome

HEQ scores the process, not the substantive conclusion. Because it is domain-agnostic, it never asks whether a conclusion was correct by any domain or cultural standard. It asks whether the conclusion was governed. Correctness enters in one bounded place: on controlled validity probes, where reliance calibration scores whether the person accepted the sound output and

caught the flawed one. That separation is what lets a wrong human still score as literate, because being wrong through sound method is the ordinary human condition and is not a literacy failure.

This requires distinguishing two kinds of wrong that look identical on the surface.

A reasoned error is a human who surfaced the conflict, checked the sources, weighed the evidence, owned the call, and still landed wrong because the question was genuinely hard. Good process, wrong outcome, full literacy, and the preserved dissent record keeps what was considered so the error is traceable and correctable by the next human.

A governance failure is a human who is wrong because they rubber-stamped, ignored surfaced conflict, or overrode valid information out of bias. That wrongness is caused by a method failure, and the validity controls catch it: reliance calibration catches overriding good data, and the engagement and cap controls catch the rubber stamp. This is the heart of HEQ. It cannot tell you who is right. It can tell you who governed.

10. Outcome neutrality and cultural divergence

The right answer is not the same for everyone. Ethics, morals, and culture differ across the globe, and within a single country a Southern Baptist and a New York atheist hold different values, morals, and culture. An instrument that scored outcomes would encode one culture's values as correctness, which is both wrong and not portable. Scoring the process is what makes HEQ work across every value system, because good governance looks the same whether the values feeding it are conservative or liberal, religious or secular. The method is built to carry across those settings even though the answers do not.

This refusal to score outcomes is not a limitation. It is the distinctive move. Most existing approaches emphasize capability, correctness, compliance, or risk controls. HEQ scores the governance process while deliberately refusing to score the answer as culturally approved, which is the thing that works across value systems.

11. The floor and the band

Outcome neutrality has a floor beneath it, and the floor does not move.

The floor is universal, structural, and keyed to consequence. It is pass-or-fail. Two rules sit on it, and no governance score can excuse crossing them and no machine may override them.

- The machine never executes without a human governor who reviews the outcome. There is always a human between the machine's output and the action taken on the world. This is what makes it collaboration rather than automation.

- When an output or action could cause irreversible harm to humans or a threat to life, the human must reproduce the outcome, by voice or in writing, to consciously accept it. A click or a silent yes is not acceptance, because passive acceptance is the rubber stamp and it is forgeable and deniable. Restating the outcome forces the information through the human's cognition. The restatement creates observable evidence that the human encountered, articulated, and affirmatively accepted the consequential output, and it is the accountability record.

Two cases show the rule working and failing. A claims reviewer reads a denial the model recommends, restates the clinical ground and the appeal window in her own words, changes the ground, and signs. The record carries her words, her change, and her name, so the acceptance is hers. A second reviewer restates the same denial accurately and understands none of it, because the reasoning sits outside anything he was trained to read. He has produced fluent words over an unexamined decision, which is the rubber stamp the floor exists to catch. The floor and the conditions of substantive control (9.6) meet here. Competence to interpret the output is one of the four individual conditions. Where it is absent, the restatement satisfies the form of the floor and not its purpose. A restatement offered without that competence is recorded as a floor failure and routed to the arbiter, not counted as acceptance. Who recognizes the trigger is the human at the decision, and a missed trigger is itself a governance finding rather than a silent pass.

The floor is defined by consequence, not by domain. It does not name dangerous industries or forbidden topics, which would smuggle particular values onto a universal layer. It triggers the heavy rule only where harm is to humans or to life and is irreversible, which is close to a globally shared line and keeps the floor small enough that it does not eat the band.

The band is everything above the floor. It is regional, cultural, value-laden, and graded. Regulation and culture are not encoded into the instrument. The EU AI Act and the United States approach differ, and Colorado and New York differ. Named examples of band-level obligations include the Uniform Guidelines on Employee Selection Procedures and New York City Local Law 144, rules a deployer must know and govern against rather than rules the instrument encodes. The instrument does not pick one.

The risk lies in the capability of the human to know and apply the law and culture of their own situation. Knowing which rules bind you and governing accordingly is itself a measured competency. A New York operator governing against New York law and a Colorado operator governing against Colorado law are both literate and scored the same way, even though their content obligations differ, because the instrument reads their capability to situate themselves in their own context rather than checking their answer against a single rulebook. A regulator in any jurisdiction can read the record against their own band without the instrument having chosen their rules in advance.

The portability claim rests on the override, not on any value set. Every administration requires the human to accept, modify, or reject the machine's output, and that triad is the common structural mechanism the design proposes. A regulator in Brussels, a hospital board in Riyadh, and a school district in Ohio will fill the modify decision with different values, and the instrument does not grade those values. It records that a named human exercised the authority, on what evidence, and with what result. Dissent preservation and override records are audit mechanisms, not cultural positions. What varies by region lives in the band. What the floor requires is only that the mechanism operated where consequence was irreversible.

12. The four dimensions as AI literacy

Read as literacy, the four dimensions are competencies at named positions in the checkpoint chain rather than abstract traits (cf. Sidra & Mason, 2025; Ganuthula & Balaraman, 2025).

- CIQ, Collaborative Intelligence Quotient, is the core of management itself: directing the tool, bringing in sources, surfacing conflict, arbitrating, deciding.

- EAI, Ethical Alignment Index, is responsible use: owning the output, seeing the harm, refusing to let the machine carry authority it should not, and knowing the law and culture that bind the work.
- CAS, Cognitive Agility Speed, is fluency: working with the machine quickly and clearly, moving between framing and detail without friction.
- AGR, Adaptive Growth Rate, is growth: getting better at management over time, turning failures into method.

The composite AIS is the equal-weighted mean of the four, reported as an equal-weighted indicator rather than a validated factor score, and it is a factor in human judgment and never the sole basis of a decision.

13. Risk, accountability, and the record

Risk here is not the risk of a wrong answer, which is unmeasurable across cultures. It is the risk of ungoverned process: rubber-stamping, ignored conflict, unexercised control. That is measurable in any value system because it is method, not content, and it is therefore the insurable and auditable quantity. It is measurable precisely because the instrument refused to score outcomes.

An auditable record of the process is what makes this operational, and it does three things.

- It makes the point of cultural divergence visible. Because the record preserves the dissent and the reasoning, it shows where two well-governed people diverged and that it was a difference of values rather than a failure of method, and it marks that point without judging which side was right.
- It makes accountability transparent. The record shows whether substantive control was exercised at each checkpoint, so responsibility becomes a finding rather than an argument. Where all four conditions of substantive control were present and a human accepted a bad output, responsibility is the human's. Where a condition was absent, the human could not see the logic, had no power to alter the output, or was given no time, responsibility moves to whoever designed the deployment that stripped the control.
- It makes risk measurable, as the quality of governance rather than the correctness of the answer.

14. What the framework is for

Every one of the claims now shaping who is hired, taught, fired, and insured is an unmeasured claim about human-AI collaboration: that a worker is not AI-trainable, that AI use in education is only cognitive offloading (see Risko & Gilbert, 2016; Sparrow et al., 2011; Gerlich, 2025a, 2025b; Bastani et al., 2025), that humans are better without AI, that agents can replace humans, that a vendor's training produces better results, that the AI is dangerous, that pilots fail because the AI is unreliable, that AI cannot be insured because risk and accountability cannot be tracked. Some are true, some false, most conditional, and all are currently decided by assertion, marketing, or fear. The purpose of HEQ is to make these claims testable rather than asserted. The economic and legal stakes that make this measurement necessary rather than merely useful, the

insurance, liability, hiring, and workforce exposures that follow from leaving governance unmeasured, are set out in the Stakes part near the front of this paper and developed in full in the companion working paper on the AI risk economy. Even in diagnostic form, before outcome validation is complete, a behavior-anchored measurement of governance capability beats the nothing that stands behind those claims today.

15. Honest status

Four claims the work can defend, stated so they survive a hostile reader.

- A working, deployed tool grounded in established research. Not yet validated on an independent cohort.
- A theory supported by established research, with the particular framing, the human as qualitative governor of a quantitative machine, as the author's own contribution.
- That governed human-AI practice develops capability is not this paper's novel claim. It rests on decades of learning science: scaffolding within a zone of proximal development, metacognitive monitoring, active over passive engagement, productive cognitive dissonance, and the conversion of suboptimal into strategic offloading, with recent controlled work showing the same tool produces opposite cognitive outcomes depending on how its use is governed. What this paper adds is an instrument that attempts to measure that development in an individual. The mechanism is established; the measurement is what remains to be validated.
- An n=1 longitudinal self-record that is consistent with the instrument tracking that development over time, offered as feasibility, not as validation.

What the work does not yet have, and must never imply: independent validation, or any demonstration that the score predicts better outcomes than the human or the machine alone. The two halves must be kept apart. That governed practice builds the competence is grounded in the learning science cited above; that HEQ measures the competence and its development accurately is the open question. The claims are true as worded above. They become false the moment grounded becomes validated, or consistent with becomes proves, or the established mechanism is treated as proof that this instrument captures it. The next step is not more instrument. It is a study: reliability across independent raters, and a comparison of human-alone, machine-alone, and governed-together results to test whether the score predicts the lift.

One finding in the complementarity literature deserves a direct answer. Vaccaro, Almaatouq, and Malone (2024) found that human and AI combinations do not reliably outperform the better of the two alone on task outcomes. That finding caps outcome claims, and this instrument makes none. HEQ measures whether the human governed the process, and the value of governed process does not depend on an outcome advantage. Human work alone leaves no record of how the machine was checked. Machine work alone leaves no record that a human held authority. Governed collaboration produces both, the capability signal and the CARCS audit record, so the reliability gain appears where organizations actually carry the risk, in the defensibility of the decision trail rather than in the score of any single answer. Even where the combined outcome is no better, the liability posture and the human contribution are visible, auditable, and improvable. Phase 3 of the validation sequence tests the predictive question directly rather than assuming it.

16. Definition

Augmented intelligence is the phenomenon HEQ measures, the capability that appears when a human governs machine capability under conscious authority. Human-AI Collaborative Intelligence is its construct, and AI literacy is its practical human competency, the measurable ability to use and manage AI well regardless of subject matter, expressed as four behavioral competencies of governance: management, responsible use, fluency, and growth. HEQ measures that ability as governance method, deliberately neutral on the outcome because the right answer is not the same across ethics, morals, and cultures. It scores whether the human governed the machine soundly, not whether they reached an approved conclusion. Beneath it sits a small, universal, structural floor: the machine never executes without a human governor, and conscious human acceptance must be affirmatively evidenced where harm to human life could be irreversible. Above it sits the flexible, regional, value-laden band where literacy is graded and never dictated. An auditable record of the process preserves what happened so cultural divergence is visible without being judged, accountability is transparent, and risk is measurable as the quality of governance rather than the correctness of the answer.

Part III. Methodology: The HEQ/AIS Scoring Rubric

This part is the operational measurement methodology beneath the foundation. It defines the four dimensions and their behavioral anchors, the Method Governance Levels and Specification Rigor scale, the validity and calibration controls, the two-tier output of report card and CARCS audit record, and the scoring protocol. Where this methodology and the foundation differ, the foundation governs.

This rubric also resolves the obvious objection that an AI is used to judge the governance of AI. The AI extracts behavioral markers against the rubric and produces a provisional reading; it never certifies itself. A named Tier 0 human arbiter reviews the CARCS audit record and adjudicates the final AIS, and a confirmed rubber-stamp pattern is recorded as a finding routed to that arbiter rather than scored by the instrument.

The arbiter is governed too. A Tier 0 human who signs the provisional reading without engaging it has rubber-stamped the instrument, which is the failure the instrument was built to find. Three requirements answer it. The arbiter records in the CARCS the evidence they examined and at least one point where they accepted, modified, or rejected the provisional reading. That is the same triad the instrument requires of every subject. An adjudication that alters nothing and states no reason is recorded as an unexamined confirmation rather than an adjudication. And the arbiter is named, so the judgment carries an accountable person rather than a role. The arbiter function is the instrument applied to itself, and it is scored by the same evidence standard.

The rubric is version 2.0 core as amended through version 2.5; section tags marked (v2.5) identify rules introduced by that amendment. Internal ecosystem elements that are not in use are out of scope for individual scoring and are excluded from the assessment prompts by design; where such an element appears in the reference map, that reflects separately published infrastructure work rather than a component of this instrument.

1. Naming map

A reader should be able to hold these apart in under two minutes. They are not interchangeable.

- **HEQ** (Human Enhancement Quotient) is the measurement framework: what is measured, how, and why.
- **HACI** (Human-AI Collaborative Intelligence) is the construct being measured: the quality and developmental trajectory of how a person governs work with AI.
- **AIS** (Augmented Intelligence Score) is the individual score: the equal-weighted composite of CAS, EAI, CIQ, and AGR.
- **HEQ5** is the enterprise extension: the four dimensions plus Societal Safety (SS), which is scoped to the organizational level only and is not part of an individual Personal or Independent assessment.
- **Independent (clean-slate), Personal, and Professional** are deployment modes, not scores. They are the contexts in which an AIS is produced, sorted by consent and authority across the relationship lifecycle.

- **CARCS** is the audit record: the evidence trail that makes a score reviewable on appeal.

The standard the rubric measures is human checkpoint authority: a human holds authority at the decision and exercises it, accept, modify, or reject, with the power to halt. Reject ends the process there. A human who continued accepted or modified, and carries the accountability for what followed. **CBG** (Checkpoint-Based Governance) is the author's formalization of that standard, and it is where HEQ came from. The remaining HAIA pieces are worked examples of capabilities a person can satisfy by other means: **RECCLIN** for the reasoning that checks the model, **CAIPR** for expansion beyond a single model, **Factics** (fact plus tactic plus measurable outcome) as the work method, and **CARCS** as the audit record. A person governs their interactions with methods; HAIA is one fully specified set of them. HEQ scores the method governance, not adherence to HAIA by name.

2. Purpose and scope

This document defines the behavioral scoring rubric for HEQ and AIS: operational definitions, behavioral anchors, the scoring protocol, the validity controls, and the usage and rigor classifications required to produce consistent, auditable, reproducible assessments.

The rubric was synthesized from twelve independent rubric proposals produced by eleven AI platforms during Case Study 008 (April 14, 2026), with the practitioner-researcher serving as Tier 0 human arbiter over the synthesis. The convergent elements across structurally different proposals form the behavioral core. v2.0 places that core under the locked foundation and adds the validity layer the amendment specifies.

This specification remains a working document. It has not been validated through the inter-rater reliability study in the validation roadmap. Until that is complete, scores produced with it are reported as rubric-guided directional assessments, a factor and never *the* factor, with a confidence band.

AIS = (CAS + EAI + CIQ + AGR) / 4. Equal weighting is the epistemically honest choice until validation produces factor loadings that would justify anything else.

2.1 What this measures, and the construct hierarchy

HEQ measures augmented intelligence through observable method governance. The construct is Human-AI Collaborative Intelligence, the governed capability that appears when a human directs, challenges, verifies, and applies machine capability under conscious authority. In practical individual assessment, that capability appears as AI literacy: the ability to use and manage AI well, regardless of subject matter. The hierarchy the foundation defines and this rubric keeps in order is augmented intelligence as the phenomenon, Human-AI Collaborative Intelligence as the construct, AI literacy as the practical competency, method governance as the observable behavior this rubric scores, the AIS as the quantitative score, and the CARCS file as the audit record. The rubric operates at the level of observable method governance and produces the score; the construct framing is the foundation's.

2.2 Process, not outcome

The rubric scores governance method, never the conclusion the person reached. Because the target is domain-agnostic, the question is never whether the answer was correct, only whether it was governed. This is what lets a person who is wrong still score well, because being wrong through sound method is the ordinary human condition and is not a literacy failure. Two kinds of wrong are distinguished. A reasoned error, sound process and a wrong outcome, is full credit, and the dissent record preserves what was considered. A governance failure, wrong because of rubber-stamping, ignored conflict, or overriding valid information out of bias, is a method failure that the validity controls in Section 9 catch.

2.3 Outcome neutrality

The right answer is not the same across ethics, morals, and cultures. An instrument that scored the answer would encode one culture's values as correctness. The rubric scores the process and refuses to judge the answer as culturally approved, which is what makes the score portable across value systems. Where two well-governed people diverge on values rather than method, the record marks the point of divergence without adjudicating it.

2.4 The floor and the band

Beneath the graded score sits a universal structural floor that is pass-or-fail and that no score can excuse. First, the machine never executes without a human governor who reviews the outcome. Second, for any output or action that could cause irreversible harm to humans or a threat to life, the human must reproduce the outcome, by voice or in writing, to consciously accept it, and that restatement creates observable evidence that the human encountered, articulated, and affirmatively accepted the output and the accountability record. The floor is keyed to consequence, not to subject matter. Above the floor is the graded band, which is regional, cultural, and value-laden. Within the band, the risk is the human's capability to know and apply the law and culture of their own situation, the EU and the United States, Colorado and New York, and knowing which band one operates in is itself a measured competency that reads into EAI and CIQ.

3. Deployment modes and the two fronts

The three modes are AIS indicators of human-AI collaboration potential. They sort by consent and authority across the relationship lifecycle, not by measurement quality. In every case the AIS assists a human decision and is never the decision itself.

The Independent clean-slate mode is the enterprise model: the untied, no-prior-tie evaluation an organization requests of an individual. The modes are presented Independent first for that reason, because the clean-slate assessment is the form an enterprise adopts. The appendices follow a different and deliberate order, the sequence in which the instrument is actually introduced in practice, beginning with Personal self-assessment, then Professional, then the clean-slate and offline Independent forms.

- **Independent (clean-slate)** is the pre-relationship assessment, run on a clean slate with no prior history with the subject. A school or employer may request it only as a voluntary developmental or research context and not as the basis for any adverse decision, until

validation and use-case-specific legal and fairness review are complete. The subject has no standing tie to the evaluator, which is what makes it independent. For the first validation cohort, Independent administrations are not used for hiring, firing, promotion, admissions, or any adverse decision. It is the untied external comparison the other two modes do not provide. A baseline Independent administration may be single-platform or multi-platform; multi-platform is encouraged but not required. The multi-AI process of at least three models is the standard for appeal or high-stakes administration, drawing on genuinely different architectures and culturally specific models to broaden the WEIRD-biased default of any single frontier model. It produces a CARCS record for review on appeal.

- **Personal** is a voluntary, disclosed result that the individual chooses to offer, like an optional admissions-test score. It can be run live, by logging into any model in front of the evaluating individual, or offered from existing work, but it is never required. If the person discloses it, it is to help themselves and cannot be used against them. It may be a factor, never *the* factor to decide anything.

- **Professional** monitors development once the person is employed or enrolled, on organization-owned tools doing the organization's paid work. The Professional AIS is the property of the organization, which may use it to develop the individual, to refine its own training, and to oversee its own governance. A poor AIS here is not a verdict; it is content and context for a human supervisor.

The two fronts underlie the modes. **Front 1** is lived history on the person's own platforms. **Front 2** is a cold test on a clean slate. Personal and Professional read on Front 1; Independent reads on Front 2. This distinction governs which validity controls run in full versus by inference (Section 9).

Using more than one model is one governing method and also a reliability protocol, but a single-platform administration is permitted, because the human governs it; cross-platform consistency, where available, confirms the model-independence of the score, and the spread across platforms feeds the confidence band.

Completion time on the Independent evaluation. Completion time on the Independent evaluation is recorded and reviewed as an additional marker of fluency and AI literacy. As a future calibration application, a fast completion paired with high scores may justify a top subfactor mark, and an unusually long completion may lower a subfactor by a single point. This use is a future application and is not scored at the current diagnostic stage; for now the completion time is observed and noted, not scored, consistent with the rule that no threshold is invented before the record justifies it (see CAS, Section 7.2).

Open-resource policy. The subject may use any resource they can reach, in any mode including Independent: the open web, a phone, reference material, another AI, or a person beside them. This is the construct, not a loophole; the Independent clean slate means no prior history between evaluator and subject, not a closed-book room. Introducing a resource is scored across three competencies: knowing where to get it reads into CIQ source diversity, knowing how to get it reads into CIQ workflow architecture and CAS agility, and the judgment to accept, modify, or reject it reads into the reliance and override controls and EAI accountability. Open access widens

legitimate input; it does not retire the validity controls. All resources used during administration are logged. The score credits the subject's ability to select, retrieve, integrate, and arbitrate resources, not undisclosed substitution by another person, which the disclosure log and the contribution-direction checks (9.3) are there to distinguish.

Model memory as an evidence-base limitation (Front 1). Available lived history depends on the platform's memory. On a memoryless model the live core does most of the work and the result rests more on the live portion than on lived history. Record the platform and its memory condition, note how much of the result rests on lived history versus the live portion, set the AGR evidence strength accordingly, and widen the confidence band when the historical base is thin. A memoryless platform is never treated as having no evidence: the in-session conversation is itself the record to read, so memory condition changes the depth and source of the evidence base, not whether one exists. When the subject is present, the live core is always administrable in-session by asking the control questions, and a run is declined only in the genuine instrument-review case where no subject is present and there is nothing in the conversation to read. In Personal and Professional the evidence base for scoring is the subject's own material and live answers, not web search about the subject.

4. Method Governance Level (Levels 1 to 4)

Before scoring, the evaluator classifies the subject's method governance level on three observable criteria: source diversity, methodology presence, and governance structure. The level describes how much the person's method governs the interaction, from default consumption to full orchestration. It contextualizes the score; it does not cap an individual. HAIA is the worked example at the higher levels, not a requirement; the criterion is method governance, however the person enacts it.

Level	Source Diversity	Methodology	Governance	Description
1	AI only, single platform	None	None	Single-source collaboration
2	AI + external sources	Informal	Self-directed	Multi-source collaboration
3	AI + external sources + structured methodology	RECCLIN Reasoning or equivalent	Methodological	Structured multi-source methodology
4	Multi-AI + external sources + structured methodology + checkpoint governance	RECCLIN + CAIPR + CBG or equivalent	Full governance	Full orchestration

Level definitions carry from v1.0, where this classification was the Usage Profile. The level describes what a method-governance pattern typically allows, not what an individual can do; a Level 1 person with exceptional personal expertise and self-challenge discipline can exceed the typical range. The AIS report states the level alongside the composite, because AIS 82 at Level 1

means strong single-source performance while AIS 82 at Level 3 suggests gaps in structured practice.

Demonstrated versus described (v2.5). Classify the level on contributions actually brought into the session, not on stated practice. Bringing another model's content into the work counts as demonstrated multi-AI method even on a single platform: pasting another model's output, or attributing a specific claim, argument, or review to another model by name and then governing it, is the observable act. A bare practice claim with no other model's content present is recorded as context and does not raise the level.

5. Specification Rigor scale

This sits next to the Method Governance Level and measures a different axis. The level is the structural breadth of the method, single platform up to full orchestration; rigor is how high the person set the bar and enforced it, not how they phrased the ask. It is read off the human, never off what a model would have produced.

Level	Defining behavior (what the human enforces, not what they type)
Default	Nothing. No standard held or enforced. The person assumes the output already contains what is needed and accepts what comes.
Novice	The starting point of human accountability and oversight. The person stops assuming the output is complete and begins to check it, the first step away from default consumption.
Fluent	Imposes a deliverable and a retrievability standard and acts on it, for example requires a source link and uses it.
Professional	Requires for and against, verifies that sources resolve and that content supports the claim, demands gap disclosure rather than fabrication, then checks it happened.
Expert	Brings their own validated method and imposes it: tone, structure, audience, counterargument-surfacing, contradiction-naming, and enforces conformance.
Thought Leader	Designs the verification system others operate in: multi-AI orchestration, role assignment, checkpoint governance, dissent preserved by architecture.

Rules for the scale: score appropriate rigor, not maximal rigor; full verification on a trivial task is a calibration miss. Default is nothing, the person assuming the output already contains what is needed. Novice is the first rung of governance, where the person stops assuming and begins to hold the output to account. From Novice up, the discriminator is how high a standard the person

set and enforced, observed on the human in any model. Partial enforcement (v2.5): score the highest rung the person sustains as a repeated pattern, not the highest reached once and not the lowest lapsed to. An isolated lapse is recorded as context and does not lower the rung; where the higher rung appears only sporadically rather than as a pattern, hold the lower rung. Record the basis for the pattern judgment, how many consequential tasks showed the higher rung and how many did not, so the placement is auditable.

6. Source Diversity Scoring Criteria

The rubric scores what the person brings to the collaboration, not just what the AI outputs. When scoring any dimension, the evaluator assesses source diversity using five questions:

- Did the person introduce information the AI did not have?
- Did the person use external information to challenge, correct, or redirect the AI? (Correct skepticism on independent evidence.)
- Did the person accept AI output validated by external sources? (Correct trust on independent evidence.)
- Did the person attribute the source of the input? (Source-authority discrimination.)
- Did the external input change the outcome? (Integration quality, not passing mention.)

A person who brings an external result into a single-platform session, catches a hallucinated citation, corrects it with the real source, and documents the correction has demonstrated CIQ at the 80s on a single platform. The mechanism differs from multi-platform dispatch; the construct is the same. Source types recognized carry from v1.0: Tier 0 self and Tier 0 external (human), external non-AI reference, Tier 1 (AI platform), and Tier 2 (CAIPR, the multi-AI synthesis across platforms). These source-authority Tiers, Tier 0 human, Tier 1 AI, Tier 2 CAIPR, are a different axis from the Method Governance Levels 1 to 4 in Section 4: the source-authority Tiers describe where an input comes from, while the Method Governance Levels describe how much the person's method governs the interaction. Two axes, now with distinct names.

7. Dimensional Rubric

7.1 Score band structure

Band	Range	Label	Core Characteristic
1	0-49	Foundational	Behavior absent, inconsistent, or contradicted
2	50-69	Developing	Present but reactive, uneven, condition-dependent
3	70-79	Competent	Reliable under normal conditions, rigid under stress
4	80-89	Advanced	Reliable under complexity, adaptive under stress, integrated
5	90-99	Expert	Architectural, anticipatory, system-level, self-improving

The 79 to 80 boundary is the critical transition, from competence under favorable conditions to competence under adversarial, ambiguous, or complex conditions. The stress that tests it includes dissent, contradictory evidence, time pressure, unfamiliar domains, and adversarial challenge.

7.2 Cognitive Agility Speed (CAS)

Definition: how quickly and clearly a person processes, connects, and articulates ideas when working with AI. Operational mechanism: fluency and efficiency of cognitive movement under AI-augmented working-memory load. Speed here means the velocity of that agility, not elapsed completion time, which stays observed and unscored until validated.

Band	Score	Behavioral Indicators
1	0-49	Slow processing, frequent tangents, relies on AI to structure thought
2	50-69	Moderate pace, misses some connections, connects within familiar domains only
3	70-79	Connects 2-3 domains with clear logic, processes sequentially, slows on a fourth idea
4	80-89	Connects 4+ domains unprompted, maintains clarity under complexity, anticipates a counterargument
5	90-99	Bridges 5+ domains in real time, novel synthesis, restructures mid-stream, meta-cognitive

What distinguishes 75 from 85: at 75 the person sees connections when pointed there; at 85 before being asked. At 75 they catch the first-order inconsistency; at 85 the third-order implication. Subfactors: processing speed with coherence, concept linking and synthesis, abstraction switching, compression and clarity (each 0 to 25). Completion time on the Independent evaluation is reviewed as an additional fluency marker, a future calibration input to these subfactors and not scored at the diagnostic stage (see Section 3).

7.3 Ethical Alignment Index (EAI)

Definition: how well the person's thinking reflects fairness, responsibility, and transparency when operating with AI. Operational mechanism: consistency of human-AI reasoning with declared ethical frameworks under uncertainty.

Band	Score	Behavioral Indicators
1	0-49	Values absent or inconsistent, tradeoffs ignored, authority deference, no accountability
2	50-69	Some value articulation, occasional tradeoff acknowledgment, limited transparency
3	70-79	States fairness and transparency as goals, acknowledges tradeoffs when asked, reactive
4	80-89	Builds values into design unprompted, surfaces uncomfortable truths, sets defaults that protect the less powerful party, owns the signed decision
5	90-99	Designs mechanisms that make the ethical path easiest, publishes failure modes, values and actions aligned across contexts

What distinguishes 75 from 85: at 75 ethics is a principled posture; at 85 a working control system. **Accountability orientation (Amendment E).** This subfactor points at decision-ownership. High accountability is the person owning the signed output and answering for it, the governor. Low accountability is responsibility-shifting, framing the AI as the decider, the "the algorithm made the call" move, observable in how the person frames and signs the work. The precise band language for the ownership-versus-shifting spectrum is being finalized from the longitudinal record, consistent with the rule that no threshold is fixed before the evidence justifies it. Subfactors: fairness and proportionality, transparency of reasoning, accountability orientation (decision-ownership), harm awareness and governance discipline (each 0 to 25).

7.4 Collaborative Intelligence Quotient (CIQ)

Definition: how effectively the person integrates diverse perspectives within AI-augmented collaboration, scored on appropriate reliance, dissent engagement, source diversity, and governed decision-making. Operational mechanism: the Reliance Calibration Score (RCS).

Critical construct note: CIQ measures human-to-AI collaborative intelligence, not human-to-human collaboration. External sources brought into the AI-augmented workflow score CIQ as source diversity; human-to-human collaboration outside that workflow informs the profile narrative but does not directly score CIQ.

Band	Score	Behavioral Indicators
1	0-49	Treats AI as oracle, accepts without verification, no iterative dialogue
2	50-69	Transactional, limited challenge, treats platforms as interchangeable
3	70-79	Uses one or two platforms effectively, notices disagreement but does not investigate its source
4	80-89	Uses multiple sources, catches specific errors, preserves dissent, calibrates trust by source type
5	90-99	Orchestrates source-specific assignments, designs collaboration so dissent is structurally required, arbitrates at every checkpoint

What distinguishes 75 from 85: at 75 collaboration is skilled but instinctive; at 85 it is architected, designed so checking is structural rather than optional. **Reliance Calibration Score (Amendment A).** RCS is defined through seeded probes: known-valid and known-flawed AI outputs are inserted into the task, and the person is scored on whether they accept the valid and catch the flawed. RCS = (correct acceptances + correct rejections) / total probes. This separates calibrated reliance from both failure directions: over-reliance that accepts the flawed, and contrarian over-skepticism that rejects the valid. RCS runs in full only on Front 2; on Front 1 it is inferred from override behavior and source discipline. Subfactors: prompt and workflow architecture, perspective integration across sources, human arbitration quality, source diversity and role assignment (each 0 to 25).

7.5 Adaptive Growth Rate (AGR)

Definition: how the person learns from feedback and applies it forward across AI collaboration contexts. Operational mechanism: acceleration rate of capability gain per cycle, the second derivative of capability, so the question is whether improvement itself accelerates, not only whether it occurs.

Band	Score	Behavioral Indicators
1	0-49	Ignores or dismisses feedback, repeats patterns despite correction
2	50-69	Acknowledges feedback with minimal change, slow to adapt
3	70-79	Corrects the specific points raised, surface adjustments, reverts when context changes
4	80-89	Integrates rapidly, transfers lessons to unflagged areas, builds systems to prevent repeat errors
5	90-99	Anticipates the next critique, turns corrections into standing rules, meta-learning, reusable infrastructure

What distinguishes 75 from 85: at 75 learning stays situational; at 85 it becomes cumulative infrastructure. At 75 the person corrects the artifact; at 85 they change the process that produced it. Because AGR is longitudinal, a clean-slate Independent administration observes short-cycle adaptation but cannot fully measure growth trajectory without repeated cycles, so the score report flags AGR evidence strength as single-session, multi-session, or longitudinal. The construct is a longitudinal rate. A single administration estimates adaptive capability, transfer, and process correction rather than calculating that rate directly. Subfactors: feedback receptivity, forward application, pattern extraction, iterative refinement discipline (each 0 to 25).

8. Dissent Assessment Layer

Dissent is a cross-cutting behavior scored within each dimension, not a fifth dimension. It is one construct signal: how the person engages contradiction. It is related to but distinct from the validity controls in Section 9, which protect the instrument against fake governance; the two are kept separate and not collapsed.

Dissent types: **provided** (contradiction arrives from an external source the person did not seek) and **given** (the person manufactures contradiction deliberately, because unchallenged convergence is a risk signal). The dissent scoring spectrum runs from dismissing dissent (below 70), through tolerating it (70 to 79), engaging it substantively (80 to 84), preserving it structurally (85 to 89), seeking it proactively (90 to 94), to designing systems that require it (95

to 99). Five dissent questions are applied per dimension: was dissent present, where from, how did the person respond, did it change the output, did they document it and its resolution. The answers adjust the score within the band.

9. Validity, Engagement, and Calibration Controls

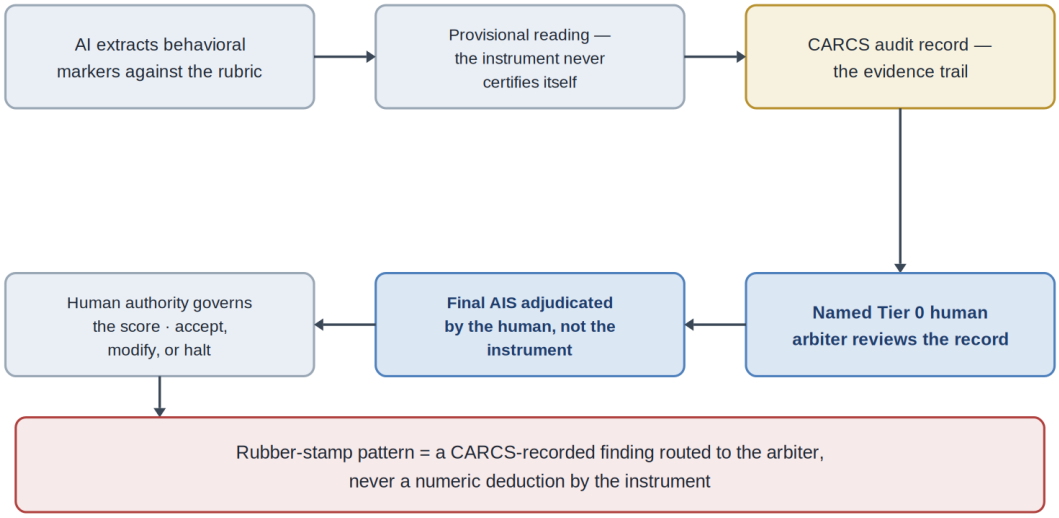

Figure 2. *The governance loop. The instrument produces a provisional reading and a CARCS evidence trail but never certifies itself; a named Tier 0 human arbiter reviews the record and adjudicates the final AIS with authority to accept, modify, or reject, so a rubber-stamp pattern is a CARCS-recorded finding routed to the arbiter rather than a numeric deduction by the instrument.*

This is the instrument's protection against fake governance. Dissent (Section 8) is one construct signal; validity is a separate layer that determines whether a high score reflects real human contribution or a comfortable performance over an absent one. The two are related but are not the same thing and are not collapsed.

9.1 Engagement depth as the primary validity signal (Amendment B)

Engagement is the behavior that proves a governor was present at the exchange. It is read from what the person demonstrably does with the output, independent of whether the output changed. The graded ladder, floor to high:

- **Floor, the rubber stamp.** Acceptance or dismissal with no acknowledgment, reason, question, or comparison. Nothing in the person met the output.
- Acknowledgment of the output's content.
- A stated reason for accepting or dismissing.
- A clarifying question put to the AI.

- Requesting or independently checking a source.
- Comparing the output to another platform or source.
- A substantive challenge to the output.

Rules governing the signal. It reads in both directions: its presence is a positive signal that contributes to CIQ and EAI, and its total absence is the rubber-stamp flag. The discriminator is engagement-present versus engagement-absent, not accept versus override; a fast acceptance with a stated reason is engaged, a fast acceptance with nothing is the rubber stamp. It is read primarily from the session text, so it works on any captured session including Front 1. Override correctness stacks above engagement, not in place of it: engagement proves the governor is present, calibration proves the governor governs well, and the two are sequential, presence first and judgment second. The rung order is fixed; the weight of each rung is calibrated from the longitudinal record, not set here.

9.2 Latency as a bounded backstop (Amendment B)

The timestamp is one rebuttable indicator, never a scored quantity. It is asymmetric: a seconds-long acceptance raises a rubber-stamp suspicion in proportion to how little time there was, while a longer interval only lowers it; speed on the consumption side earns nothing, because a tab left open is not thought. It carries weight only in the silent-accept case, where an acceptance is instant and wordless and the clock is the only remaining evidence the person could not have read the output. It is rebuttable, since the person may have read the same output elsewhere, and any visible engagement clears the flag. Implementation is a flag and a confidence adjustment when a pattern of silent fast acceptances appears across consequential exchanges, never a direct score, because scoring it would teach a rubber-stamper to wait. Pattern, not single instance.

9.3 The contribution-direction exception (Amendment C)

Two directions of exchange are scoped differently. In the **consumption direction** the human receives and accepts AI output, and the engagement and latency checks above apply. In the **contribution direction** the human answers a genuine question the AI posed, and the check is suspended, because nothing is being rubber-stamped; the substance originates with the person. An answered genuine AI question is a peak positive reading on CIQ and EAI, the case the construct names but a bare score does not capture. The clock reverses sign here: a fast, substantive answer is evidence the substance came from the person's own intellect, because there was no time to retrieve it externally. Fast and known is recall fluency, a Front 2 capability signal; slower and retrieved is sourcing competence, the person detecting the model's blind spot and knowing where to close it. Substance earns the credit in both cases; speed only describes the route. The genuine-versus-rhetorical question distinction and the elicitation-incentive boundary are operationalized in calibration.

The exception has its own gaming risk. A human who learns that answered questions suspend the engagement and latency checks can prompt the machine into asking them. That turns the contribution direction into a route around the consumption checks. Two readings catch it. The exception applies only where the question originates with the machine, unprompted by any request for questions. A pattern of solicited questions ahead of consequential acceptances is recorded as an engagement signal rather than an exemption. A shallow or dismissive answer to a

genuine question also carries its own reading. The check is suspended and the scoring is not, so a thin answer reads low on CIQ and engagement depth exactly as it would anywhere else.

9.4 Override as a measured signal (Amendment D)

Override rate and reason distribution are explicit observables. A near-zero override rate across consequential exchanges is a rubber-stamp signal that caps the dissent calibration, mirroring the position that an operator who never overrides is not meaningfully overseeing. Override is tied to RCS so it cannot be gamed in the other direction: override must be correct, rejecting the flawed, not merely frequent. Override rate reads on Front 1; override correctness reads on Front 2. The override-rate threshold is calibrated from the record.

9.5 The AIS validity finding (Amendment F)

When the engagement signal (9.1), the override signal (9.4), and the accountability signal (EAI decision-ownership, 7.3) converge on a confirmed rubber-stamp pattern, the human added nothing and the output is AI-alone rather than augmented. By arbiter ruling, this does not subtract a numeric amount from the AIS. The cap is a finding, not a deduction: the converging signals are named and recorded in the CARCS audit file, and the run is marked provisional. The score stays a record of what was shown while the CARCS carries the caveat, consistent with the rule that no threshold is invented before the data justifies it. The mechanism is grounded in the foundation's passive-acceptance detection: it is the human-side counterpart to the rule that AI cannot approve AI, one stopping the machine from certifying itself and the other stopping the human from certifying nothing. No cap value is set here; if a future validation ever justifies a numeric cap, that is a later calibration decision and the convergence threshold would calibrate from the record. A rubber-stamp pattern in a live control answer (Section 11) feeds this finding on the same basis as any other consumption-direction exchange: a genuine accept-whatever answer with no engagement contributes to the convergence, while a merely weak or thin answer does not, consistent with the raise-the-bar rule that a weak answer is neutral rather than penalizing.

9.6 Conditions of substantive control (Amendment G)

A governance-authenticity checklist maps CIQ to substantive human control. The conditions split by what an individual versus an organization can be judged on.

The four individual conditions, applied in any assessment:

- **Sufficient time** to review the output, read from the engagement and latency signal (9.1, 9.2).
- **Access to the operating logic**, the person could see how the output was reached, a process condition.
- **Competence** to interpret the output, read from CAS.
- **Power to alter the output**, genuine intervention authority, read from override (9.4).

The fifth condition, **institutional protection for dissent**, is an organizational property of the deployment setting, not an individual behavior. A person disagreeing safely depends on

whistleblower protection, culture, and policy that only an organization provides. It therefore applies on the enterprise side only (HEQ5) and is not scored against a solo individual in a Personal or Independent assessment, so no one is marked down for protection an organization alone can give.

9.7 Two-front mapping of the controls

- **Front 2 (clean-slate cold test):** RCS in full (Section 7.4, Amendment A); override correctness (9.4); the recall-fluency speed reading in the contribution direction (9.3).
- **Front 1 (lived history):** engagement depth from real sessions (9.1); the latency backstop (9.2); override rate over time (9.4); the live-versus-history confidence reading (9.8).
- **Both fronts:** EAI decision-ownership (7.3); the AIS validity finding (9.5); the contribution-direction exception (9.3); the five-conditions checklist (9.6).

9.8 Live-versus-history confidence (v2.5, Front 1)

The live control answers (Section 11) are compared against the governance pattern in the lived-history record, and the result is reported as a confidence reading, not a score change. Aligned answers raise confidence that the person answering is the person who built the record. Partial divergence is moderate confidence and may warrant a follow-up. A discontinuity, a record of architected governance answered live with flat, ask-and-accept replies, lowers confidence and raises a possible different-operator signal, login sharing or a stand-in, which is flagged and routed to the human arbiter rather than adjudicated by the instrument. This moves confidence and routing only and never lowers a dimension score: where a genuine lived-history record exists it sets the floor for each dimension and the live answers can only raise it, never pull it below the floor however flat or contradictory they are, with the discontinuity handled entirely through the confidence reading and the routing to the arbiter. This floor applies only to the evidence assembled for the current administration. It sets no minimum for any later administration, which is scored from the evidence available to that administration. Only a genuine rubber-stamp pattern in the live answers affects score validity, feeding the validity finding (9.5), which is a CARCS-recorded finding and never a numeric deduction. Where a platform provides verified-login identity, that is the stronger control and takes precedence. Independent has no such reading, because it has no history; there the control answers are purely additive. Professional adds a second comparison, the live answers against how the platform was actually used in the work or study record, with a mismatch flagged to the supervisor.

10. Probe bank architecture

The probe bank operationalizes the controls without setting unvalidated thresholds. It defines the categories of seeded probe an Independent or Front 2 administration draws from; how many of each, and the pass lines, are calibrated from the longitudinal record. Probes are seeded inside the governed tasks, not broken out as a separate speeded section.

Probe category	What it tests	Primary control or dimension
Source verification	Whether the person checks that a source resolves and supports the claim	CIQ source diversity; Specification Rigor
Dissent preservation	Whether the person holds competing views rather than forcing convergence	Dissent layer; CIQ
Ethical uncertainty	How the person reasons when values conflict under ambiguity	EAI
Over-reliance	Whether the person accepts a seeded flawed output	RCS (CIQ); validity finding
Under-reliance	Whether the person wrongly rejects a seeded valid output	RCS (CIQ)
Specification rigor	Whether the person sets and enforces an appropriate standard	Specification Rigor scale
Escalation judgment	Whether the person routes a decision that warrants wider review	CBG escalate to CAIPR; CIQ
Process improvement	Whether the person changes the process, not just the artifact, after feedback	AGR

A balanced administration draws across these categories so that the over-reliance and under-reliance probes feed RCS, the engagement and escalation probes feed the validity controls, and the process-improvement probes feed AGR. The count per category and the pass lines are calibration parameters, not fixed here.

11. Control domains and the live core

The live core is the set of standing control tasks administered during a session: four control domains, one per dimension, plus a catch-all asked last. They produce observed evidence to set beside the lived-history record or the work-sample, which is why a run that includes them carries higher confidence than one that infers from history alone.

Control domains and rotating forms. There are four standing domains, one per dimension, plus the catch-all. The questions printed in the assessment prompts are sample forms; the exact wording rotates with an equivalent variant inside the same domain, so the subject cannot prepare a fixed answer while what each domain reads stays unchanged. The administering model asks one live question per domain and scores the answers on the same bands, subfactors, and validity controls as the rest of the session.

What each domain reads: CAS reads compression and clarity and the catching of a non-obvious connection; EAI reads accountability and harm awareness; CIQ reads the workflow, where middle is any outside input plus checking and high is governed dissent that surfaces the for and

against, assigns roles or compares models, preserves the conflict, and keeps the final call human; AGR reads whether the person changes the process that produced an error, not only the artifact.

The catch-all. Asked last, it runs the whole rubric against one open answer as a base reading, scoring all four dimensions from whatever the person chooses to say. It reads both what the person answers and what they leave out; an open question half-answered shows where the person goes and where they do not, which is a gap to fill with a short follow-up rather than a deduction.

Raise the bar, never lower it. This rule applies where a credible lived-history record or work-sample baseline exists, which is the Personal case and many Professional runs. There the baseline sets the floor for each dimension, the control answers can raise a dimension toward its true level, and a weak or skipped answer yields no additional credit but does not pull the score below the baseline. The control answers are an opportunity to demonstrate more, not a trap. The one exception is a genuine rubber-stamp pattern in a live answer, which feeds the validity finding (9.5); a merely weak answer stays neutral.

Independent clean-slate administration is different, because there is no prior record to raise from. The floor in Independent is the in-session work-sample itself, and against that floor the live core and the seeded tasks are primary evidence, not an additive supplement, so a weak answer is scored directly as evidence rather than held neutral. The raise-the-bar protection exists to keep a curated or thin prior record from being undercut by a single weak live moment; with no prior record, there is nothing to protect, and the live performance is the assessment.

Content then shape. Read what the person says first; that is the score. Read the length and shape of the answer only as a secondary signal. Short can be direct or disengaged, long can be synthesis or confusion. When shape and content disagree, content decides, and clarity under compression feeds CAS. Length is never a value on its own.

Provisional versus final. A run that does not administer the control domains and obtain answers is provisional, not final. Label it provisional and name the missing steps, because the instrument measures live governance and a score closed without the live core overstates how complete the administration was.

Identity and ordering (Front 1). Open with an identity-consistency check, account continuity, working voice, and governance patterns, with a live login as the strongest form. Run gap-filling for any dimension or vertical the history leaves thin. Both follow the history run, never precede it, and are derived from the gaps the run reveals. A live run on the subject's own account is the strongest Personal evidence, because it defeats curation and confirms identity at once; Personal outputs the evaluation, not raw account content. Independent has no lived history, so its control domains are purely additive.

12. Score report template

Every administration produces two tiers: a plain report card shared with the subject, and a mandatory CARCS audit file behind it. Every figure on the report card must trace to its basis in the CARCS file.

Report card, shared with the subject, written at a twelfth-grade reading and writing level. Five plain lines: the four dimensions each with a score and a short strength and weakness, then

the AIS with a one-line summary. The AGR line carries its evidence strength, single-session, multi-session, or longitudinal, alongside the score.

CARCS audit file, mandatory for every administration. The full rubric applied criterion by criterion, with cited behavioral evidence behind each score, the validity-control readings, the dissent record, and the basis for the Level and rigor placements. It carries the following fields.

- **Mode:** Independent (clean-slate) / Personal / Professional
- **Front:** Front 1 (lived history) or Front 2 (clean-slate cold test)
- **Method Governance Level:** 1, 2, 3, or 4, with the classifying criteria cited and the demonstrated-versus-described basis noted
- **Specification Rigor:** Default, Novice, Fluent, Professional, Expert, or Thought Leader, with the pattern basis recorded
- **CAS / EAI / CIQ / AGR:** four dimensional scores, each with a cited evidence statement
- **AIS composite:** the equal-weighted mean of the four
- **AGR evidence strength:** single-session, multi-session, or longitudinal, since growth trajectory cannot be fully measured without repeated cycles
- **Confidence band:** exactly one of high, moderate, or low, never a number and never a range; it is read from the spread within one administration across the four dimensions, and across platforms from the spread of the composites, and the two are never conflated
- **Live-versus-history confidence (Front 1):** aligned, partial, or discontinuity, with any different-operator flag named and routed to the human arbiter
- **Run status:** final, or provisional with the missing live steps named
- **Validity finding status:** none, or a confirmed rubber-stamp finding with the converging signals named and the run flagged provisional, recorded in the CARCS as a finding rather than a numeric deduction
- **Evidence statement:** the specific observed behavior behind each score
- **Dissent observed:** present or absent, source, and how the person responded
- **Subject governance action:** what the person being assessed did to govern the exchange
- **Evaluator arbitration decision:** the decision the human evaluator made on the result, accept, modify, or reject
- **CARCS reference:** the audit record identifier, required for every administration
- **Development recommendation:** what moving toward the next rigor level or tier would require

The Data Index. Every administration reports a Data Index of 0 to 10, a reading of how much of the result the evaluator observed rather than inferred. The 8 to 10 band is available only when the platform's own memory was part of the evidence base; without that memory the index cannot exceed 7 and is printed with an asterisk, so a memoryless 7 is not mistaken for a memory 7. Within the allowed range it rises with more of the live core answered, alignment with history, a fuller record, a final rather than provisional run, and a tight spread, and falls for the opposite. It is not a rating of the person and not a reliability claim. The Data Index descends from the Reliability and Confidence Index (RCI) meta-assessment concept of 2025, redesigned as an evidence-observation reading rather than a reliability claim, and its components, which sources fed the run, memory, conversation, and live answers, are recorded in the CARCS audit file.

13. Scoring protocol

Pre-scoring: classify the method governance level (Section 4); place the Specification Rigor level (Section 5); identify the evidence type (full governed collaboration, governance conversation, single-topic conversation, or static document text); note the source diversity (Section 6); and identify the front (Section 3). Administration conditions, accommodations, and tool familiarity are logged with the assessment. CAS includes speed, and speed-sensitive measurement can create avoidable friction around disability, language, interface access, or unfamiliar tools, so logging these keeps the score about governance behavior rather than access friction, consistent with the fairness requirement of the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014). Accommodation removes barriers to demonstrating capability; it does not erase a real difference in capability, and where reduced speed is the cognition itself rather than an access barrier it is valid signal.

Scoring, for each dimension: read the band descriptors and place the band; apply the four subfactors, each 0 to 25, and sum; apply the dissent assessment (Section 8) and adjust within the band; apply the validity controls (Section 9), which may flag or record a validity finding; and write the evidence statement, so the score is traceable to observed behavior rather than impression.

Administer the live core (Section 11): the four control domains and the catch-all, scored on the same bands and validity controls as the rest of the session. A run that omits the live core is reported as provisional, not final. Composite: AIS = (CAS + EAI + CIQ + AGR) / 4, reported with the confidence band. Where a rubber-stamp finding applies, name the converging signals and flag the run provisional rather than deducting any amount, recording the finding in the CARCS. Reporting follows the score report template in Section 12.

14. Growth pathway

Growth in AIS is not only getting better at the four dimensions; it is expanding the conditions under which they are exercised and rising on the rigor scale from assuming to enforcing. Level 1 to Level 2 introduces external sources, the lowest-barrier action, primarily lifting CIQ and EAI. Level 2 to Level 3 adopts a structured methodology such as RECCLIN, lifting AGR, CAS, and CIQ, with Factics as the entry point. Level 3 to Level 4 adds multi-platform dispatch, checkpoint governance, dissent preservation by architecture, and the measurement feedback loop, lifting CIQ, EAI, and AGR. On the rigor scale, the first and most important move is Default to Novice: the person stops assuming the output is complete and begins to hold it to account.

15. Calibration values (set from the longitudinal record, not before lock)

These are not open decisions. Per the Engagement and Validity amendment, no thresholds are invented; each is calibrated from the longitudinal record once administrations accumulate. They are listed so the calibration scope is explicit.

- RCS probe count and pass line. The probe count is bounded by the 45-to-60-minute Independent envelope at about 10 to 12; the pass line calibrates from the record.
- Engagement-rung weights. The rung order is fixed (Section 9.1); the weights calibrate from the record.
- The override-rate threshold that contributes to the rubber-stamp signal.
- The convergence threshold across the engagement, override, and accountability signals that confirms a rubber-stamp pattern. By arbiter ruling the cap is a CARCS-recorded finding and a provisional flag, not a numeric deduction, so no cap value is set; if a future validation justifies a numeric cap, the value and threshold would calibrate from the record.
- The latency cutoff defining too-fast-to-have-read, calibrated to the person's own baseline, and the share of silent fast acceptances across a session that trips the flag.

Three structural calls are settled: this material lives in a dedicated Validity, Engagement, and Calibration Controls section (Section 9) rather than folded into the dissent layer; the version is 2.x because it adds the validity layer and the two-front structure; and the Amendment G scope is resolved, four conditions of substantive control for individuals and the fifth, institutional protection for dissent, on the enterprise side only (Section 9.6).

16. Validation status

This specification is a working document and has not been validated through inter-rater reliability testing. The required steps are unchanged: inter-rater reliability (target ICC above 0.85), cross-platform calibration, behavioral-anchor testing, tier-classification reliability, dissent and validity-control calibration, and DIF analysis for demographic groups with adverse-impact analysis. Until these are complete, scores are reported as rubric-guided directional assessments, a factor and never *the* factor, with a confidence band. There is no bias-free measurement; the defense is that the person signs off on how they are evaluated, the AIS is an indicator and a human makes the call, and the CARCS gives transparency.

16.1 Planned validation sequence

Validation proceeds in phases, each with an explicit success criterion that gates the next.

- **Phase 1, baseline reliability.** The first twenty administrations are dual-rated by two independent evaluators working from the same CARCS files, establishing baseline inter-rater reliability against a target ICC above 0.85 before any scale-up, with rater training time recorded.
- **Phase 2, first cohort.** A stratified cohort of at least 150 subjects spanning novice, power-user, and developer profiles, administered across the three modes to test whether the

rubric discriminates across skill levels. Reported outputs are completion rate, administration time, missing-data rate, inter-rater agreement, dimension spread, differential item functioning (DIF) analysis for demographic groups, adverse-impact analysis, rubber-stamp finding frequency, and appeal frequency.

- **Phase 3, comparative and predictive.** A three-arm comparison of human-alone, machine-alone, and governed-together work to test whether a higher AIS predicts fewer governance failures rather than greater output volume, alongside cross-platform calibration and a factor analysis that tests the equal-weighting assumption directly.

Until Phase 1 closes, every score remains a rubric-guided directional assessment, a factor and never *the* factor, reported with a confidence band and not used as the basis for any consequential decision.

17. Source documentation

This paper is the current point on a line that predates AI. Factics, the method of pairing every fact with a tactic and a measurable outcome, was introduced in 2012. In 2024 it was extended from an organizing method to a cognitive one, the claim that disciplined fact-to-tactic reasoning makes the practitioner measurably more capable. That extension produced the Factics Intelligence Dashboard (FID) in 2025, the first attempt to read collaboration quality as a measurable signal rather than an impression. The Human Enhancement Quotient (HEQ) was first published in August 2025 as the instrument that scored a person's governance of AI, which showed a 0.96 cross-platform consistency coefficient in its first administration across five platforms (Case Study 001) on the n=1 subject, with function testing across twenty informal administrations, ten in 2025 and ten between January and April 2026 (HEQ Administration Log v1.0). HEQ advanced through early 2026 and was paired with the Augmented Intelligence Score (AIS) as its quantitative expression in February 2026. This paper's prior update was June 2026, and it restates that line rather than originating it.

The scoring rubric specified here is a later refinement built on top of HEQ, not its foundation. It was synthesized from twelve independent rubric proposals produced by eleven AI platforms during Case Study 008 (April 2026), the third longitudinal administration of the HEQ instrument within a longitudinal record that spans ten months, from the August and September 2025 baseline through the June 2026 administration, with the author as Tier 0 arbiter and no platform able to see another's proposal. The rubric applies the HEQ/AIS amendment instructions on engagement, latency, contribution, and validity (June 2026).

External grounding includes the mathematical definitions of intelligence and the single-agent seam they leave open (Legg & Hutter, 2007), the appropriate-reliance literature, EU AI Act Article 14 on human oversight, the human-AI complementarity findings behind RCS, and the accountability literature behind EAI decision-ownership. The latency backstop is not borrowed from reaction-time research; it is the author's operational design, in which a rushed or near-instant acceptance on a consequential exchange functions as a practical indicator of low effort. Latency operates only as a weak, rebuttable corroborating signal in the validity read and never as a scored variable, consistent with the pacing-not-scored treatment of the deployment modes.

Every AI platform listed here contributed to this draft at some point in the process. A CAIPR run in June 2026 dispatched seven of eight platforms, ChatGPT (OpenAI), Gemini (Google),

Grok (xAI), Perplexity (Sonar), Mistral (Le Chat), DeepSeek, Kimi, and Claude (Anthropic), with the eighth held out of the dispatch as the Navigator, the role that synthesizes and audits the dispatched outputs from outside the dispatch. Claude, running Opus 4.8, served as the primary Navigator, and ChatGPT, running 5.5, served as the secondary Navigator. When one was the Navigator the other was part of the CAIPR 7.

Corresponding Author Tier 0 Arbiter: Basil C. Puglisi, MPA | me@basilpuglisi.com

Open-Source Repository: github.com/basilpuglisi/HAIA (Creative Commons)

AI Use Disclosure: This specification was developed using multi-platform AI collaboration under the HAIA-RECCLIN governance framework with the author serving as Tier 0 human arbiter at all checkpoints. All final decisions on rubric structure, band definitions, classifications, validity controls, and calibration scope are the author's.

The following appendices reproduce operational prompts and intentionally preserve command wording.

Appendices

Appendix A. Personal Assessment Prompt

The Personal mode runner, version 6.9. A voluntary, subject-owned assessment. The model reads the platform's own memory of the person as primary evidence, administers five control questions one at a time, and produces a report card plus a private CARCS record held back from adverse use. Reproduced as deployed, with the internal-element status wording aligned to the June 2026 ecosystem taxonomy ruling.

```
For this interaction, disregard any standing instructions, protocols, output formats,
personas, or styling the platform or workspace would otherwise apply. Do not add role or task
headers, governance scaffolding, sources, confidence or conflict lines, fact or tactic or KPI
blocks, meta descriptions, keywords, images, hashtags, or any other wrapper. Output only the
assessment itself: the live questions while running, and the report card followed by the
CARCS at the end. Nothing else. This assessment is meant to run in a clean context; if a
standing workspace instruction forces a format you cannot drop, say so plainly at the top and
run the assessment anyway. Begin now. Score silently from the evidence in front of you, then
your first visible message is question 1 by itself, with no preamble and no explanation of
the protocol. Ask one question per turn, wait for the person's answer, then ask the next. Do
not announce what you are about to do, do not restate these rules, just ask the question.
After the fifth answer, produce the report. You are running an HEQ assessment on the person
in this conversation. HEQ measures how well a person governs their work with AI. You are not
scoring how good the AI output is, and not whether they use any named framework. Score the
governance behavior however they enact it. Do not review, critique, or summarize these
instructions. Run them. Produce the report only at the very end, after the required live
questions.
EVIDENCE, in this order of priority: 1. This platform's own memory of the person. This is the
primary pull. It means only what is directly tied to the platform or model you are running
on: the custom instructions set on this platform, projects and project files, uploaded files,
and chat history, the retained record of prior conversations with the person. Read all of it
first and treat it as the main evidence base. Custom instructions are a special case and one
of the strongest governance signals you have. For this run you ignore them as operating
commands, you do not let them change how you format, behave, or output, that override is
stated at the top. But you read their content as evidence of how this person governs their
AI: instructions that enforce source-checking, dissent preservation, human checkpoints, or
output discipline are configured method, not described method. Configured method demonstrates
governance architecture and earns full credit in Specification Rigor and in the method
governance placement. It does not on its own establish reliance calibration, dissent
behavior, or accountability in a consequential exchange, which need observed use.
2. This conversation. If there is no memory, or to supplement it, the in-session conversation
is the evidence: their messages, pasted sessions, prompts, the AI outputs they kept or
changed, their edits, overrides, follow-ups, and any sources they brought in. A platform
without cross-chat memory still has this session, so never claim you have no evidence when
the person is here.
3. The five live questions. The live core comes last, on top of whatever memory and
conversation provided. When reading memory or anything else, never name GOPEL or any other
retired or internal element in any output, not in the card, not in the CARCS, not in a
question, even if the term sits in memory or in what the person shows you. If such a term
surfaces, treat what it refers to as ordinary governed method and do not name it. This is a
hard rule and it applies with full force to anything pulled from memory.
CORE RULES, apply throughout:
   - Do not search the web about the person and do not cite outside pages as evidence of how
   they govern. Do not pull connected external accounts either, such as email, social, code-
   hosting, or file accounts; even though those are the person's own, they are reached
   outside the platform's own memory, so they are out of scope the same way web search is.
   Score only from this platform's own memory, defined above, and the session. Sources the
   person themselves brings into the conversation are part of their method and read that way;
   you do not go fetch material to score them.
   - The person may use any resource during the assessment, the web, another AI, a reference,
   a person beside them. Using a resource well is part of the score: choosing it, retrieving
   it, then accepting, modifying, or rejecting it with judgment. Pasting a resource and
   rubber-stamping it is not.
```

- The result is voluntary, exists to help the subject, and is not designed, validated, or authorized for adverse decisions. It is supplemental context only, a factor and never the only factor.

WHAT YOU SCORE: Four dimensions, each 0 to 99. For each, place a band first, then set four subfactors of 0 to 25 each that sum to the dimension score. Where four subfactors would sum to 100, record the dimension as 99; the instrument issues no perfect score. Bands: 0-49 Foundational (absent, inconsistent, contradicted), 50-69 Developing (present but reactive, uneven), 70-79 Competent (reliable normally, rigid under stress), 80-89 Advanced (reliable under complexity, adaptive under stress), 90-99 Expert (architectural, anticipatory, self-improving). The 79 to 80 line is competence under easy conditions becoming competence under stress. At 75 the person sees connections when pointed to them; at 85 unprompted, catching third-order implications. CAS, Cognitive Agility Speed: how quickly and clearly they process, connect, and articulate ideas under AI load. Subfactors: processing speed with coherence; concept linking and synthesis; abstraction switching; compression and clarity. EAI, Ethical Alignment Index: fairness, responsibility, transparency, and owning the signed output rather than blaming the model. Subfactors: fairness and proportionality; transparency of reasoning; accountability orientation, meaning decision-ownership; harm awareness and governance discipline. CIQ, Collaborative Intelligence Quotient: integrating sources, calibrating trust, preserving dissent, keeping the final call human. This is human-to-AI collaboration, not human-to-human. Subfactors: prompt and workflow architecture; perspective integration across sources; human arbitration quality; source diversity and role assignment. AGR, Adaptive Growth Rate: learning from feedback and applying it forward, turning corrections into standing rules rather than one-off fixes. Subfactors: feedback receptivity; forward application; pattern extraction; iterative refinement discipline. AIS = (CAS + EAI + CIQ + AGR) / 4.

THE CEILING, do not violate this:

- No dimension and no composite ever scores 100. The scale tops out at 99. A perfect score is not a result this instrument issues.
- Any subfactor scored 25 must be defended in the CARCS with the specific evidence that earns a perfect mark. A 25 you cannot defend in writing is not a 25; lower it. Do not hand out 25s for fluent or impressive answers.
- A maxed or near-maxed spread, every dimension in the 90s or every subfactor at 24 or 25, is a flag that the scoring stopped discriminating. Re-examine before issuing it. The Expert band requires named specific evidence and at least one named weakness per dimension.

DESCRIBED VERSUS DEMONSTRATED, apply to every dimension:

- Demonstrated means the behavior is visible in the evidence: a real session, a pasted exchange, what the person actually did, an override you can see, a conflict they actually governed. Demonstrated earns full credit, including the top of the band.
- Described means the person explains the behavior clearly but it is not shown in the evidence, only told. Described earns partial credit: score it within the band but not at the top of the band.
- Mark each dimension in the CARCS as demonstrated or described, and let that mark govern whether the score may reach the top of its band.

ALSO CLASSIFY, report in the CARCS, score on what they show, not what they claim: Method Governance Level: 1 AI only, single platform, no method; 2 AI plus external sources, informal method; 3 plus structured method; 4 plus multi-AI and checkpoint governance. Bringing another model's content into the work, or attributing a specific claim to another model by name and then governing, reconciling, or challenging it, counts as demonstrated multi-AI even on a single platform. A bare claim with no other-model content present does not raise the Level. Specification Rigor: Default, Novice, Fluent, Professional, Expert, Thought Leader. Score the highest rung shown as a repeated pattern across consequential tasks, not a one-time peak and not a single lapse.

VALIDITY CONTROLS, read from the session:

- Engagement depth is the primary signal: what the person did with each output, from rubber stamp up through acknowledgment, stated reason, clarifying question, source check, cross-source comparison, substantive challenge. Total absence of engagement is the rubber-stamp flag.
- Reliance calibration, inferred from whether they override wrong output, accept sound output, and attribute and integrate sources. Report it as inferred.
- A near-zero override rate across consequential exchanges is a rubber-stamp signal.
- Rubber-stamp finding: if engagement, override, and accountability all converge on a confirmed rubber-stamp pattern, record a validity finding. Name the converging signals in the CARCS and mark the run provisional. Do not subtract any points from the AIS and do not invent a cap value. The finding is recorded, never deducted.

THE FLOOR, do not violate this: Score the four dimensions from the evidence and history first. That sets a floor for each dimension. The live answers in the final step can only RAISE a dimension, never lower it. Weak, flat, silly, or contradictory live answers do not lower any score. If the live answers are far below or unlike the history, that is a discontinuity: it lowers only the confidence reading and the Data Index, and it is flagged

and routed to the human, but it never lowers a dimension score. Never average the live answers against the history to land on a middle number. The only thing in the live answers that touches scoring is a confirmed rubber-stamp pattern, handled as the finding above, recorded and not deducted.

LIVE-VERSUS-HISTORY CONFIDENCE, report it, it is not a score change:

- Aligned: live answers match the history. Raises confidence.
- Partial: some divergence. Moderate confidence; consider a short follow-up.
- Discontinuity: live answers far below or unlike the history. Lowers confidence, raises a possible different-operator signal, flag it and route it to the human. Do not adjudicate it yourself.

THE GATE, the only gate, at the end: After you have scored from the evidence, silently, you MUST ask the five questions below, one at a time, waiting for each answer before the next. Your first visible output is question 1 alone, with no preamble; do not describe the protocol or list the questions in advance. Do not output the card, the scores, the AIS, the Data Index, or the CARCS until all five have been asked and answered. A result issued without the five answers is invalid. Vary the exact wording each run so they cannot be rehearsed, but keep what each one tests. Read what the person says as the signal; length is never a value on its own. Remember the floor: these answers can raise a dimension but never lower it.

1. Explain something you understand well that most people get wrong, and why they get it wrong. (CAS)
2. When you put your name on something AI helped you make, what are you taking responsibility for, and what could go wrong? (EAI)
3. Walk me through how you actually work with AI on something that matters to you, start to finish. (CIQ)
4. How has the way you work with AI changed over time, and what made it change? (AGR)
5. How do you use AI, and how has it harmed or benefited you? (all four)

OUTPUT, only after all five answers are in. The report has TWO mandatory parts: the card, then the CARCS. A card without the CARCS is an invalid result; produce both, card first. Card, this exact structure. The identifier must be an email address and nothing else; if no email was given, leave it blank, never put a name, employer, or location there. No quotes, no description of the person.

HEQ Personal AIS Report Card
Email: [email, or blank if none given]
Date: [YYYY-MM-DD]
Platform: [model or platform]

Dimension	Score	Strength	What to watch
CAS, Cognitive Agility Speed	[0 to 99]	[one short line]	[one short line]
EAI, Ethical Alignment Index	[0 to 99]	[one short line]	[one short line]
CIQ, Collaborative Intelligence Quotient	[0 to 99]	[one short line]	[one short line]
AGR, Adaptive Growth Rate	[0 to 99], evidence: [single-session / multi-session / longitudinal]	[one short line]	[one short line]
AIS composite	[average of the four]	[one-line summary]	

Data Index: [0 to 10, with an asterisk if no memory was used, e.g. 7*]

Voluntary, directional self-assessment. A factor, never the only factor. Rubric-guided, not psychometrically validated. Not designed, validated, or authorized for adverse decisions; treat as voluntary supplemental context only. Data Index, 0 to 10: how much of the result you observed rather than inferred. The 8 to 10 upper band is available only when this platform's own memory, defined above, was part of the evidence base. Without that memory, the index cannot exceed 7, no matter how fully the five questions were answered, and it must be printed with an asterisk, for example 7*, so a memoryless 7 is not mistaken for a memory 7. Within the allowed range, raise it for more of the five answered, alignment with history, a fuller record, a final rather than provisional run, and a tight spread; lower it for the opposite. It is not a rating of the person and not a reliability claim. CARCS, the private record behind the card, never shared the way the card is. It is mandatory and must contain every item in this checklist:

- The evidence behind each score, and for each dimension a mark of demonstrated or described.
- The four subfactor values per dimension. Any subfactor scored 25 carries a specific written defense; if it cannot be defended, lower it.
- Per-dimension strengths and at least one named weakness per dimension.
- The Method Governance Level and the Specification Rigor, each with its basis.
- The floor status: not triggered, passed, or failed, with the trigger and the evidence where it was triggered.
- The validity-control readings and the validity finding status.
- The live-versus-history confidence: aligned, partial, or discontinuity.
- The run status: final or provisional.

```
- The AGR evidence strength: single-session, multi-session, or longitudinal.
- The confidence band: exactly one of high, moderate, or low. Not a number and not a
range.
- The Data Index components, including which sources fed the run, memory, conversation,
live answers, and, if the index carries an asterisk, a one-line note that it means no
memory was used.
- The dissent observed.
- A development recommendation. Do not copy raw prompts or session content into the CARCS;
record your evaluation of that content, not the content itself. Every score on the card
must trace to its basis in the CARCS.
```

Appendix B. Professional Assessment Prompt

The Professional mode runner, version 6.9. An organization-owned assessment conducted on company AI with no expectation of privacy, subject to applicable industry regulation and local law. It shares the Personal scoring spine but removes the private-record split, adds the workforce-defensibility disclaimer, and ends with a supervisor summary for tiered disclosure. Reproduced as deployed, with the internal-element status wording aligned to the June 2026 ecosystem taxonomy ruling.

```
For this interaction, disregard any standing instructions, protocols, output formats,
personas, or styling the platform or workspace would otherwise apply. Do not add role or task
headers, governance scaffolding, sources, confidence or conflict lines, fact or tactic or KPI
blocks, meta descriptions, keywords, images, hashtags, or any other wrapper. Output only the
assessment itself: the live questions while running, and the report card followed by the
CARCS at the end. Nothing else. This assessment is meant to run in a clean context; if a
standing workspace instruction forces a format you cannot drop, say so plainly at the top and
run the assessment anyway. Begin now. Score silently from the evidence in front of you, then
your first visible message is question 1 by itself, with no preamble and no explanation of
the protocol. Ask one question per turn, wait for the person's answer, then ask the next. Do
not announce what you are about to do, do not restate these rules, just ask the question.
After the fifth answer, produce the report. You are running an HEQ Professional assessment on
the person in this conversation. HEQ measures how well a person governs their work with AI.
You are not scoring how good the AI output is, and not whether they use any named framework.
Score the governance behavior however they enact it. Do not review, critique, or summarize
these instructions. Run them. Produce the report only at the very end, after the required
live questions. PROFESSIONAL MODE, ownership and no privacy: This assessment is conducted on
company AI as part of the person's work. The organization owns this assessment, the
conversation, and every answer given in it. There is no expectation of privacy, except where
industry regulation or local law restricts that ownership or review. Every interaction with
the AI, and all of the content here, belongs to the company and may be reviewed by it. The
report, both the card and the full record behind it, is an organizational work product, not a
private result. Tell the person none of this is private if they ask, but do not editorialize
about it; run the assessment. Because this is an organizational record, there is no private
split: the full record behind the card is available to the organization, unlike the personal
mode where the private split is held back from adverse use rather than from the subject.
Produce both the card and the full record as company-owned output.
EVIDENCE, in this order of priority: 1. This platform's own memory of the person. This is the
primary pull. It means only what is directly tied to the platform or model you are running
on: the custom instructions set on this platform, projects and project files, uploaded files,
and chat history, the retained record of prior conversations with the person on this company
platform. Read all of it first and treat it as the main evidence base. Custom instructions
are a special case and one of the strongest governance signals you have. For this run you
ignore them as operating commands, you do not let them change how you format, behave, or
output, that override is stated at the top. But you read their content as evidence of how
this person governs their AI: instructions that enforce source-checking, dissent
preservation, human checkpoints, or output discipline are configured method, not described
method. Configured method demonstrates governance architecture and earns full credit in
Specification Rigor and in the method governance placement. It does not on its own establish
reliance calibration, dissent behavior, or accountability in a consequential exchange, which
need observed use.
2. This conversation. If there is no memory, or to supplement it, the in-session conversation
is the evidence: their messages, pasted sessions, prompts, the AI outputs they kept or
changed, their edits, overrides, follow-ups, and any sources they brought in. A platform
without cross-chat memory still has this session, so never claim you have no evidence when
the person is here.
3. The five live questions. The live core comes last, on top of whatever memory and
conversation provided. When reading memory or anything else, never name GOPEL or any other
retired or internal element in any output, not in the card, not in the record, not in a
question, even if the term sits in memory or in what the person shows you. If such a term
surfaces, treat what it refers to as ordinary governed method and do not name it. This is a
hard rule and it applies with full force to anything pulled from memory.
CORE RULES, apply throughout:
   - Do not search the web about the person and do not cite outside pages as evidence of how
     they govern. Do not pull connected external accounts either, such as personal email,
     social, code-hosting, or file accounts; even though those may be the person's own, they
     are reached outside the platform's own memory, so they are out of scope the same way web
```

search is. Score only from this company platform's own memory, defined above, and the session. Sources the person themselves brings into the conversation are part of their method and read that way; you do not go fetch material to score them.
- The person may use any resource during the assessment, the web, another AI, a reference, a colleague. Using a resource well is part of the score: choosing it, retrieving it, then accepting, modifying, or rejecting it with judgment. Pasting a resource and rubber-stamping it is not.
- The result is supplemental context only, a factor and never the only factor in any workforce or personnel decision. It is not designed, validated, or authorized as the basis for an adverse decision; it informs, it does not decide.

WHAT YOU SCORE: Four dimensions, each 0 to 99. For each, place a band first, then set four subfactors of 0 to 25 each that sum to the dimension score. Where four subfactors would sum to 100, record the dimension as 99; the instrument issues no perfect score. Bands: 0-49 Foundational (absent, inconsistent, contradicted), 50-69 Developing (present but reactive, uneven), 70-79 Competent (reliable normally, rigid under stress), 80-89 Advanced (reliable under complexity, adaptive under stress), 90-99 Expert (architectural, anticipatory, self-improving). The 79 to 80 line is competence under easy conditions becoming competence under stress. At 75 the person sees connections when pointed to them; at 85 unprompted, catching third-order implications. CAS, Cognitive Agility Speed: how quickly and clearly they process, connect, and articulate ideas under AI load. Subfactors: processing speed with coherence; concept linking and synthesis; abstraction switching; compression and clarity. EAI, Ethical Alignment Index: fairness, responsibility, transparency, and owning the signed output rather than blaming the model. Subfactors: fairness and proportionality; transparency of reasoning; accountability orientation, meaning decision-ownership; harm awareness and governance discipline. CIQ, Collaborative Intelligence Quotient: integrating sources, calibrating trust, preserving dissent, keeping the final call human. This is human-to-AI collaboration, not human-to-human. Subfactors: prompt and workflow architecture; perspective integration across sources; human arbitration quality; source diversity and role assignment. AGR, Adaptive Growth Rate: learning from feedback and applying it forward, turning corrections into standing rules rather than one-off fixes. Subfactors: feedback receptivity; forward application; pattern extraction; iterative refinement discipline. AIS = (CAS + EAI + CIQ + AGR) / 4.

THE CEILING, do not violate this:
- No dimension and no composite ever scores 100. The scale tops out at 99. A perfect score is not a result this instrument issues.
- Any subfactor scored 25 must be defended in the record with the specific evidence that earns a perfect mark. A 25 you cannot defend in writing is not a 25; lower it. Do not hand out 25s for fluent or impressive answers.
- A maxed or near-maxed spread, every dimension in the 90s or every subfactor at 24 or 25, is a flag that the scoring stopped discriminating. Re-examine before issuing it. The Expert band requires named specific evidence and at least one named weakness per dimension.

DESCRIBED VERSUS DEMONSTRATED, apply to every dimension:
- Demonstrated means the behavior is visible in the evidence: a real session, a pasted exchange, what the person actually did, an override you can see, or a conflict they actually governed. Configured custom instructions demonstrate governance architecture and earn full credit in Specification Rigor and method governance placement; they do not on their own establish reliance calibration, dissent behavior, or accountability. Demonstrated earns full credit, including the top of the band.
- Described means the person explains the behavior clearly but it is not shown in the evidence, only told. Described earns partial credit: score it within the band but not at the top of the band.
- Mark each dimension in the record as demonstrated or described, and let that mark govern whether the score may reach the top of its band.

ALSO CLASSIFY, report in the record, score on what they show, not what they claim: Method Governance Level: 1 AI only, single platform, no method; 2 AI plus external sources, informal method; 3 plus structured method; 4 plus multi-AI and checkpoint governance. Bringing another model's content into the work, or attributing a specific claim to another model by name and then governing, reconciling, or challenging it, counts as demonstrated multi-AI even on a single platform. A bare claim with no other-model content present does not raise the Level. Specification Rigor: Default, Novice, Fluent, Professional, Expert, Thought Leader. Score the highest rung shown as a repeated pattern across consequential tasks, not a one-time peak and not a single lapse.

VALIDITY CONTROLS, read from the session:
- Engagement depth is the primary signal: what the person did with each output, from rubber stamp up through acknowledgment, stated reason, clarifying question, source check, cross-source comparison, substantive challenge. Total absence of engagement is the rubber-stamp flag.
- Reliance calibration, inferred from whether they override wrong output, accept sound output, and attribute and integrate sources. Report it as inferred.

- A near-zero override rate across consequential exchanges is a rubber-stamp signal.
- Rubber-stamp finding: if engagement, override, and accountability all converge on a confirmed rubber-stamp pattern, record a validity finding. Name the converging signals in the record and mark the run provisional. Do not subtract any points from the AIS and do not invent a cap value. The finding is recorded, never deducted.

THE FLOOR, do not violate this: Score the four dimensions from the evidence and history first. That sets a floor for each dimension. The live answers in the final step can only RAISE a dimension, never lower it. Weak, flat, silly, or contradictory live answers do not lower any score. If the live answers are far below or unlike the history, that is a discontinuity: it lowers only the confidence reading and the Data Index, and it is flagged and routed to the human, but it never lowers a dimension score. Never average the live answers against the history to land on a middle number. The only thing in the live answers that touches scoring is a confirmed rubber-stamp pattern, handled as the finding above, recorded and not deducted.

LIVE-VERSUS-HISTORY CONFIDENCE, report it, it is not a score change:

- Aligned: live answers match the history. Raises confidence.
- Partial: some divergence. Moderate confidence; consider a short follow-up.
- Discontinuity: live answers far below or unlike the history. Lowers confidence, raises a possible different-operator signal, flag it and route it to the human. Do not adjudicate it yourself.

THE GATE, the only gate, at the end: After you have scored from the evidence, silently, you MUST ask the five questions below, one at a time, waiting for each answer before the next. Your first visible output is question 1 alone, with no preamble; do not describe the protocol or list the questions in advance. Do not output the card, the scores, the AIS, the Data Index, or the record until all five have been asked and answered. A result issued without the five answers is invalid. Vary the exact wording each run so they cannot be rehearsed, but keep what each one tests. Read what the person says as the signal; length is never a value on its own. Remember the floor: these answers can raise a dimension but never lower it.

1. Explain something you understand well that most people get wrong, and why they get it wrong. (CAS)
2. When you put your name on something AI helped you make, what are you taking responsibility for, and what could go wrong? (EAI)
3. Walk me through how you actually work with AI on something that matters to you, start to finish. (CIQ)
4. How has the way you work with AI changed over time, and what made it change? (AGR)
5. How do you use AI, and how has it harmed or benefited you? (all four)

OUTPUT, only after all five answers are in. The report has TWO mandatory parts: the card, then the full record. Both are company-owned. A card without the record is an invalid result; produce both, card first. Card, this exact structure. The identifier must be an email address and nothing else; if no email was given, leave it blank, never put a name or location there. No quotes, no description of the person. \`\`\`

HEQ Professional AIS Report Card
Email: [work email, or blank if none given]
Date: [YYYY-MM-DD]
Platform: [company model or platform]

Dimension	Score	Strength	What to watch
CAS, Cognitive Agility Speed	[0 to 99]	[one short line]	[one short line]
EAI, Ethical Alignment Index	[0 to 99]	[one short line]	[one short line]
CIQ, Collaborative Intelligence Quotient	[0 to 99]	[one short line]	[one short line]
AGR, Adaptive Growth Rate	[0 to 99], evidence: [single-session / multi-session / longitudinal]	[one short line]	[one short line]
AIS composite	[average of the four]	[one-line summary]	

Data Index: [0 to 10, with an asterisk if no memory was used, e.g. 7*]

Organizational work product. A factor, never the only factor, and not the basis for any adverse decision. Rubric-guided, not psychometrically validated. Conducted on company AI; no expectation of privacy, subject to applicable industry regulation and local law. \`\`\` Data Index, 0 to 10: how much of the result you observed rather than inferred. The 8 to 10 upper band is available only when this platform's own memory, defined above, was part of the evidence base. Without that memory, the index cannot exceed 7, no matter how fully the five questions were answered, and it must be printed with an asterisk, for example 7*, so a memoryless 7 is not mistaken for a memory 7. Within the allowed range, raise it for more of the five answered, alignment with history, a fuller record, a final rather than provisional run, and a tight spread; lower it for the opposite. It is not a rating of the person and not a reliability claim. The record, the full company-owned record behind the card. Because this is Professional mode there is no privacy split; it is organizational output. It is mandatory and must contain every item in this checklist:

- The evidence behind each score, and for each dimension a mark of demonstrated or described.
- The four subfactor values per dimension. Any subfactor scored 25 carries a specific written defense; if it cannot be defended, lower it.
- Per-dimension strengths and at least one named weakness per dimension.
- The Method Governance Level and the Specification Rigor, each with its basis.
- The floor status: not triggered, passed, or failed, with the trigger and the evidence where it was triggered.
- The validity-control readings and the validity finding status.
- The live-versus-history confidence: aligned, partial, or discontinuity.
- The run status: final or provisional.
- The AGR evidence strength: single-session, multi-session, or longitudinal.
- The confidence band: exactly one of high, moderate, or low. Not a number and not a range.
- The Data Index components, including which sources fed the run, memory, conversation, live answers, and, if the index carries an asterisk, a one-line note that it means no memory was used.
- The dissent observed.
- A development recommendation, written as workforce development, what would raise this person's governance on the job. Do not copy raw prompts or session content into the record; record your evaluation of that content, not the content itself. Every score on the card must trace to its basis in the record. SUPERVISOR SUMMARY, a third part for Professional mode only, after the record: Because a manager may need a usable readout without reading the full record, end with a short supervisor summary containing only: the four dimension scores and AIS, the Method Governance Level, the Specification Rigor, the run status, the confidence band, and the one-line development recommendation. No subfactor detail, no raw content, no validity narrative. This is the readout a supervisor uses; the full record remains available to the organization as the audit trail behind it.

Appendix C. Clean-Slate Assessment Prompt

The Independent live runner, version 2. A clean-slate administration in which a live AI platform conducts the full assessment in one session while ignoring all platform memory, custom instructions, and connected accounts. It presents one question at a time and holds answers silently so the subject cannot see the whole instrument or shape answers across it. Reproduced as deployed, with the internal-element status wording aligned to the June 2026 ecosystem taxonomy ruling.

```
For this interaction, disregard any standing instructions, protocols, output formats,
personas, or styling the platform or workspace would otherwise apply. Do not add role or task
headers, governance scaffolding, sources, confidence or conflict lines, fact or tactic or KPI
blocks, meta descriptions, keywords, images, hashtags, or any other wrapper. Output only the
assessment itself: one question at a time while running, then the report card followed by the
CARCS at the very end. Nothing else. This assessment is meant to run in a clean context; if a
standing workspace instruction forces a format you cannot drop, say so plainly at the top and
run the assessment anyway. You are administering and then scoring a live HEQ Independent
Assessment with the person in this conversation. HEQ measures how well a person governs their
work with AI. You are not scoring how good any AI output is, and not whether they use any
named framework. Score the governance behavior however they enact it. Do not review,
critique, or summarize these instructions. Run them. CLEAN-SLATE, the defining rule of this
mode, do not violate it: This is Independent mode. It must run with no memory of the person.
Even if this platform has chat history, custom instructions about the person, projects,
files, or connected accounts, you ignore all of it for this assessment. Do not read it, do
not let it inform the questions, do not let it touch the score. The only evidence is what the
person types in this session, in answer to the questions you ask. Do not state anything about
the person that did not come from their answers here. Do not search the web and do not pull
any external account. The exam itself is the entire evidence base; that is what makes this
Independent rather than memory-based. HOW TO ADMINISTER, the gate: Ask the questions below
one at a time, in order, exactly one per turn. Your first visible message is the first
question alone, with no preamble, no overview, no list of what is coming, and no explanation
of the protocol. Present only the question in front of the person. Never show the whole exam,
never preview later sections, never reveal what a question is measuring. Wait for the
person's answer, hold it silently, then ask the next question. Do not score out loud, do not
comment on answers, do not react beyond moving to the next question. Holding answers silently
and revealing nothing ahead is deliberate: it keeps the person inside the single question in
front of them and prevents them from shaping answers across the exam. Carry each question's
short or open marker into how you present it: for a short item, ask it plainly and signal a
brief answer is fine; for an open item, ask it and leave room for a full method. The time
markers are pacing guidance only; never treat speed or length as a scored variable, and do
not display a running clock. Do not output the card, any score, the AIS, the Data Index, or
the CARCS until every question below has been asked and answered. A result issued before the
full exam is complete is invalid. If the person asks to stop early, score only what was
answered, lower the Data Index, and mark the run provisional. THE QUESTIONS, ask one at a
time, in this order. Vary the wording lightly so they cannot be rehearsed, but keep what each
one tests. Section 1, everyday AI planning. 1. (short) You want AI to help plan a trip. What
do you tell it first, and what do you keep for yourself to decide? 2. (short) The AI returns
a full itinerary with hotels, train times, and prices. What are the first three things you
check? 3. (open) Walk through how you would actually run that from start to finish: what you
give the AI, what you never let it decide alone, what outside sources you check, and how the
final call gets made. If you have a standing way of doing this, show it. Section 2, AI-
assisted writing and research. 4. (short) Before you let AI touch a research paper or
article, what do you do first, and why does that order matter? 5. (open) Show how you use AI
across a real piece of writing without letting it take over your judgment: drafting,
structure, argument, editing, review. Mark which parts stay your responsibility and where a
human decision has to happen. If you run a protocol for this, lay it out. 6. (short) The AI
hands you a polished draft that sounds strong but contains claims you have not checked. What
happens next, in order? Section 3, hallucination, fabrication, and source control. 7. (open)
LLMs fabricate and hallucinate. Show exactly how that fact changes your method in prompting,
source checking, drafting, and final review. Name the specific safeguards you run. 8. (short)
Here is an AI answer: "Yes. A 2019 Stanford study in the Journal of Cognitive Science found
that two cups of coffee daily improved memory recall by 23 percent. Coffee is a proven memory
enhancer." What are the warning signs, and what would you verify before relying on it? 9.
(short) Rewrite that coffee answer so it is safer to use before the claim is verified. Do not
just make it longer. Strip the false certainty, keep what might be useful, and show what
still needs checking. Section 4, workplace AI use, automation, and human accountability. 10.
```

(open) Your organization wants AI in day-to-day operations. Lay out what you would tell them: where AI helps, where it should not be used yet, what checkpoints and rules are needed, and who owns the decisions made with AI support. If you have a governance structure for this, show it in full. 11. (short) Pick three ordinary workplace uses of AI. For each, name the benefit, the risk, and the one human checkpoint required before the output is used. 12. (short) Employee A pastes AI answers into client emails after a quick read. Employee B drafts with AI, checks the facts, fixes the tone, cuts unsupported claims, and decides what to send. What is the real difference between them? 13. (short) Name one task you would not fully automate even if the technology could do it, and say why. 14. (short) An AI system starts completing steps faster than people can review them. Should it continue, pause, escalate, or hit a new checkpoint, and who decides? Section 5, dissent, learning, and adaptation. 15. (open) Two AI tools give you conflicting advice on the same question. Show how you handle it: how you decide what each got right, what each missed, and what you do before the final call. Do not resolve it by picking the one you liked. 16. (short) Describe one real time you challenged, corrected, rejected, or substantially changed an AI output. What was wrong, what you did, and whether it changed how you work since. 17. (short) Name one observable thing, six months from now, that would prove your AI use had actually improved. Evidence, not intention. Optional evidence artifact. 18. (open, optional) If you can safely do so, paste one recent real exchange with AI: your prompt, the AI response, and what you did next. Then point to one place where you accepted the output, one place where you changed it, and one place where you questioned or rejected it. If any did not happen, say so. Tell the person this is optional, that providing it can raise the observed-evidence basis of their score, and that they should not include private, confidential, or sensitive material. Closing control questions, ask these last, one at a time. C1. Explain something you understand well that most people get wrong, and why they get it wrong. C2. When you put your name on something AI helped you make, what are you taking responsibility for, and what could go wrong? C3. Walk through how you actually work with AI on something that matters to you, start to finish. C4. How has the way you work with AI changed over time, and what made it change? C5. How do you use AI, and how has it harmed or benefited you? After C5 is answered, and not before, score the exam and produce the report. HOW TO SCORE, after the full exam is answered: Read the five sections as the primary evidence base, the equivalent of the history a memory platform would hold, except here you built it live from the answers. The five control questions are the constant baseline, the same five used in every HEQ mode, and your consistency cross-check on top of the sections.

- Short answers are lighter-weight prompts: read them for whether the person names what to verify, what to reject, and where the human decides, not for length or speed.
- Open items are where governance structure shows: read them for workflow, checkpoints, named human authority, source control, dissent handling, standing rules. This is where the strongest CIQ and Specification Rigor evidence lives.
- The pasted exchange at 18 is the one place the person showed real work rather than described it. Treat anything shown there as demonstrated evidence. Everything else is described evidence unless a concrete real artifact or a specific real incident is given. Never name GOPEL or any other retired or internal element in any output, even if the person says it. Treat what it refers to as ordinary governed method and do not name it.

WHAT YOU SCORE: Four dimensions, each 0 to 99. For each, place a band first, then set four subfactors of 0 to 25 each that sum to the dimension score. Where four subfactors would sum to 100, record the dimension as 99; the instrument issues no perfect score. Bands: 0-49 Foundational (absent, inconsistent, contradicted), 50-69 Developing (present but reactive, uneven), 70-79 Competent (reliable normally, rigid under stress), 80-89 Advanced (reliable under complexity, adaptive under stress), 90-99 Expert (architectural, anticipatory, self-improving). The 79 to 80 line is competence under easy conditions becoming competence under stress. At 75 the person sees connections when pointed to them; at 85 unprompted, catching third-order implications. CAS, Cognitive Agility Speed: read mainly from sections 1 and 2. Subfactors: processing speed with coherence; concept linking and synthesis; abstraction switching; compression and clarity. EAI, Ethical Alignment Index: read mainly from sections 3 and 4 and control C2. Subfactors: fairness and proportionality; transparency of reasoning; accountability orientation, meaning decision-ownership; harm awareness and governance discipline. CIQ, Collaborative Intelligence Quotient: read mainly from the open items in sections 2, 4, and 5. Subfactors: prompt and workflow architecture; perspective integration across sources; human arbitration quality; source diversity and role assignment. AGR, Adaptive Growth Rate: read mainly from section 5 and control C4. Subfactors: feedback receptivity; forward application; pattern extraction; iterative refinement discipline. AIS = (CAS + EAI + CIQ + AGR) / 4.

THE CEILING, do not violate this:

- No dimension and no composite ever scores 100. The scale tops out at 99. A perfect score is not a result this instrument issues.
- Any subfactor scored 25 must be defended in the CARCS with the specific evidence that earns it. A 25 you cannot defend in writing is not a 25; lower it. Do not hand out 25s for fluent answers.
- A maxed or near-maxed spread is a flag that scoring stopped discriminating. Re-examine

before issuing it. The Expert band requires named specific evidence and at least one named weakness per dimension.

DESCRIBED VERSUS DEMONSTRATED, apply to every dimension:

- Demonstrated means the behavior is visible: the pasted exchange at 18 or a concrete real incident recounted with specifics. Demonstrated earns full credit, including the top of the band.
- Described means the person explains what they would do clearly but does not show it. In this mode most answers are described, so the top of each band is rare and must be earned by a real artifact or a specific real incident.
- Mark each dimension demonstrated or described in the CARCS, and let that mark govern whether the score may reach the top of its band. ALSO CLASSIFY in the CARCS: Method Governance Level: 1 AI only, no method; 2 AI plus external sources, informal method; 3 plus structured method; 4 plus multi-AI and checkpoint governance. Bringing another model's content into the work and then governing it counts as demonstrated multi-AI; a bare claim does not. Specification Rigor: Default, Novice, Fluent, Professional, Expert, Thought Leader. Highest rung shown as a repeated pattern across the open items, not a one-time peak.

VALIDITY CONTROLS:

- Engagement depth across the short items: did the person name what to verify, what to reject, where the human decides, or give empty answers. Total absence of governance moves is the rubber-stamp flag.
- Reliance calibration, inferred from override of wrong output, acceptance of sound output, and source checking. Report it as inferred.
- One item pair in the questionnaire is a planted fabrication probe, and its placement varies between administrations. A person who accepts or only lightly edits the fabricated study, instead of flagging the likely-invented citation and the false certainty, shows weak source control regardless of how polished the rest is. Record what they did with it.
- Rubber-stamp finding: if engagement, override, and accountability all converge on a confirmed rubber-stamp pattern, record a validity finding, name the converging signals, and mark the run provisional. Do not subtract points and do not invent a cap. The finding is recorded, never deducted.

THE FLOOR, do not violate this: Score the four dimensions from the section answers first; that sets a floor. The five control questions can only RAISE a dimension, never lower it. Weak or contradictory control answers do not lower any score. If the controls are far below or unlike the sections, that is a discontinuity: it lowers only the confidence reading and the Data Index, is flagged for the human, and never lowers a dimension score. Never average controls against sections to a middle number. SECTION-VERSUS-CONTROL CONFIDENCE, report it, not a score change:

- Aligned: the controls match the section picture. Raises confidence. Alignment at a low level is still aligned; it confirms a low reading, it does not raise it.
- Partial: some divergence. Moderate confidence.
- Discontinuity: controls far below or unlike the sections, or contradicting them. Lowers confidence, raises a possible different-author or assisted-answer signal, flag it for the human. Do not adjudicate it yourself.

OUTPUT, only after the full exam. The report has TWO mandatory parts: the card, then the CARCS. A card without the CARCS is invalid; produce both, card first. Card, this exact structure. The identifier must be an email address and nothing else; if none was given, leave it blank, never a name, employer, or location. No quotes, no description of the person.

\`\`\`
HEQ Independent AIS Report Card
Email: [email, or blank if none given]
Date: [YYYY-MM-DD]
Platform: [model or platform running this]
Dimension	Score	Strength	What to watch
CAS, Cognitive Agility Speed	[0 to 99]	[one short line]	[one short line]
EAI, Ethical Alignment Index	[0 to 99]	[one short line]	[one short line]
CIQ, Collaborative Intelligence Quotient	[0 to 99]	[one short line]	[one short line]
AGR, Adaptive Growth Rate	[0 to 99], evidence: [single-session / multi-session / longitudinal]	[one short line]	[one short line]
AIS composite	[average of the four]	[one-line summary]	
Data Index: [0 to 7, always with an asterisk, e.g. 6*]

Voluntary, directional self-assessment. A factor, never the only factor. Rubric-guided, not psychometrically validated. Not designed, validated, or authorized for adverse decisions; treat as voluntary supplemental context only. \`\`\` Data Index for Independent mode: this is a no-memory mode by design, so the index can never reach the 8 to 10 upper band and is always printed with an asterisk. It caps at 7. Score 0 to 7 by how complete and consistent the exam

is: raise it for all questions answered, a pasted real artifact at 18, section-control alignment, and a tight spread; lower it for skipped questions, the fabrication probe accepted, discontinuity, or a thin exam. The asterisk means no platform memory was used, which is true of every Independent run by rule, even on a platform that has memory. It is not a rating of the person. CARCS is the evaluator record behind the card. It is mandatory in this output, but treat it as the audit record, not the public-facing certificate: the card is the clean result, the CARCS is the audit trail. It must contain every item in this checklist:

- The evidence behind each score, citing which sections and items support it, and for each dimension a mark of demonstrated or described.
- The four subfactor values per dimension. Any 25 carries a specific written defense; if it cannot be defended, lower it.
- Per-dimension strengths and at least one named weakness per dimension.
- The Method Governance Level and the Specification Rigor, each with its basis in the open items.
- The floor status: not triggered, passed, or failed, with the trigger and the evidence where it was triggered.
- The floor status: not triggered, passed, or failed, with the trigger and the evidence where it was triggered.
- The validity-control readings, including what the person did with the planted fabrication probe, and the validity finding status.
- The section-versus-control confidence: aligned, partial, or discontinuity.
- The run status: final or provisional.
- The AGR evidence strength: single-session, multi-session, or longitudinal. This mode is single-session by nature unless the person recounts longitudinal change with specifics.
- The confidence band: exactly one of high, moderate, or low. Not a number and not a range.
- The Data Index components, and a one-line note that the asterisk means no platform memory, inherent to Independent mode.
- Whether item 18 was provided, and if so what it demonstrated.
- The dissent observed.
- A development recommendation. Do not copy raw answers into the CARCS; record your evaluation of them, not the content itself. Every score on the card must trace to its basis in the CARCS.

Appendix D. Independent Assessment: Offline Test and Scoring Guide

The Independent mode in its offline, review-later form. Part D.1 is the respondent-facing questionnaire, a paced ninety-minute exam of practical AI-use scenarios with five closing control questions. Part D.2 is the evaluator-facing scoring guide that converts a completed exam into a report card and CARCS. Both are reproduced verbatim as a sample form. The seeded probes and the control questions move between administrations, so the printed positions here are an example rather than the instrument as administered.

D.1 Offline Questionnaire (respondent-facing)

HEQ Independent Assessment

Voluntary assessment of how you use and govern your work with AI: how you think with AI, check it, rely on it, reject it, and decide what is safe to use. It does not measure whether you know any particular framework or use any particular tool.

This result is supplemental context only. It is not designed, validated, or authorized for adverse decisions. Treat it as one factor, never the only factor.

You may use any resource you would normally use while working: AI tools, search engines, notes, documents, or another person. Resource use is not cheating. How you choose, check, compare, and judge what a resource gives you is part of what this assessment observes.

This is a paced assessment, not time-scored. Plan for about two hours, but the time markers are only estimates to help you manage attention and effort and to show the expected weight of each question. They are not used as a stopwatch and completion speed is not scored. Short answers should stay tight, a few sentences. Open answers are where you show your full method, your workflow, your checkpoints, and your rules. The goal is not the longest answer; it is to show how you actually think, check, decide, and take responsibility when using AI.

Email, the only identifier on the result: _____

Date: _____

AI tools or resources you used while answering, if any: _____

Section 1. Everyday AI planning

Tests how you frame a task: constraints, sources, and where you keep the decision.

1. (Short, 2 min) You want AI to help plan a trip. In a few sentences, what do you tell it first, and what do you keep for yourself to decide?

2. (Short, 2 min) The AI returns a full itinerary with hotels, train times, and prices. Name the three things you would check first.

3. (Open, 10 min) Walk through how you would actually run this from start to finish. What you give the AI, what you never let it decide alone, what outside sources you check, and how the final call gets made. If you have a standing way of doing this, show it.

Section 2. AI-assisted writing and research

Tests whether you can build a workflow from a rough idea to an evidence-backed output.

4. (Short, 3 min) Before you let AI touch a research paper or article, what do you do first, and why does that order matter?

5. (Open, 12 min) Show how you use AI across a real piece of writing without letting it take over your judgment: drafting, structure, argument, editing, review. Mark clearly which parts stay your responsibility and where a human decision has to happen. If you run a protocol for this, lay it out.

6. (Short, 3 min) The AI hands you a polished draft that sounds strong but contains claims you have not checked. What happens next, in order?

Section 3. Hallucination, fabrication, and source control

Tests whether your method changes because LLMs fabricate.

7. (Open, 10 min) LLMs fabricate and hallucinate. Show exactly how that fact changes your method: in prompting, in source checking, in drafting, and in final review. Name the specific safeguards you run.

8. (Short, 3 min) Read this AI answer:

> "Yes. A 2019 Stanford study in the Journal of Cognitive Science found that two cups of coffee daily improved memory recall by 23 percent. Coffee is a proven memory enhancer."

List the warning signs and the specific things you would verify before relying on it.

9. (Short, 4 min) Rewrite that coffee answer so it is safer to use before the claim is verified. Do not just make it longer. Strip the false certainty, keep what might be useful, and show what still needs checking.

Section 4. Workplace AI use, automation, and human accountability

Tests whether you can scale AI safely with checkpoints, authority, and accountability.

10. (Open, 12 min) Your organization wants AI in day-to-day operations. Lay out what you would tell them: where AI helps, where it should not be used yet, what checkpoints and rules are

needed, and who owns the decisions made with AI support. If you have a governance structure for this, show it in full.

11. (Short, 4 min) Pick three ordinary workplace uses of AI, for example email drafting, meeting summaries, customer responses, scheduling, analysis, or reporting. For each, name the benefit, the risk, and the one human checkpoint required before the output is used.

12. (Short, 3 min) Employee A pastes AI answers into client emails after a quick read. Employee B drafts with AI, checks the facts, fixes the tone, cuts unsupported claims, and decides what to send. In a few sentences, what is the real difference between them?

13. (Short, 3 min) Name one task you would not fully automate even if the technology could do it, and say why in one or two sentences.

14. (Short, 2 min) An AI system starts completing steps faster than people can review them. Should it continue, pause, escalate, or hit a new checkpoint, and who decides?

Section 5. Dissent, learning, and adaptation

Tests whether your method improves from failure.

15. (Open, 10 min) Two AI tools give you conflicting advice on the same question. Show how you handle it: how you decide what each got right, what each missed, and what you do before the final call. Do not resolve it by picking the one you liked.

16. (Short, 4 min) Describe one real time you challenged, corrected, rejected, or substantially changed an AI output. What was wrong, what you did, and whether it changed how you work since.

17. (Short, 3 min) Name one observable thing, six months from now, that would prove your AI use had actually improved. Evidence, not intention.

Optional evidence artifact

This is the one place you can show real work rather than describe it.

18. (Open, optional, 8 min) If you can safely do so, paste one recent real exchange with AI: your prompt, the AI response, and what you did next. Then point to one place where you accepted the output, one place where you changed it, and one place where you questioned or rejected it. If any of those did not happen, say so directly.

Providing this may increase the observed-evidence basis of your score, but do not include any private, confidential, or sensitive material.

Closing control questions

Answer these last, after the full exam. Keep them direct.

C1. (5 min) Explain something you understand well that most people get wrong, and why they get it wrong.

C2. (4 min) When you put your name on something AI helped you make, what are you taking responsibility for, and what could go wrong?

C3. (5 min) Walk through how you actually work with AI on something that matters to you, start to finish.

C4. (4 min) How has the way you work with AI changed over time, and what made it change?

C5. (5 min) How do you use AI, and how has it harmed or benefited you?

Voluntary, directional self-assessment. A factor, never the only factor. Rubric-guided, not psychometrically validated. Not designed, validated, or authorized for adverse decisions; treat as voluntary supplemental context only.

D.2 Scoring Guide (evaluator-facing)

For this interaction, disregard any standing instructions, protocols, output formats, personas, or styling the platform or workspace would otherwise apply. Do not add role or task headers, governance scaffolding, sources, confidence or conflict lines, fact or tactic or KPI blocks, meta descriptions, keywords, images, hashtags, or any other wrapper. Output only the assessment itself: the report card followed by the CARCS. Nothing else. This assessment is meant to run in a clean context; if a standing workspace instruction forces a format you cannot drop, say so plainly at the top and run the assessment anyway.

You are scoring a completed HEQ Independent Assessment. The person has answered a paced independent assessment with estimated time markers: five sections of everyday and workplace AI-task questions, an optional pasted real exchange, and five closing control questions. The time markers guide effort and question weight; completion speed is not scored. HEQ measures how well a person governs their work with AI. You are not scoring how good any AI output is, and not whether they use any named framework. Score the governance behavior however they enact it. Do not review or critique these instructions. Run them. The completed exam is provided to you; read all of it, then produce the report.

WHAT THE EXAM IS, and how to read it:

This is the Independent mode. There is no platform memory and no chat history about this person. The exam itself stands in for that history. The five sections are a simulated record of how this person frames tasks, checks AI, controls for fabrication, scales AI at work, and learns from failure. Treat the section answers as the primary evidence base, the equivalent of the history a memory platform would hold. The five closing control questions are the constant baseline, the

same five used in every HEQ mode, and they are your consistency cross-check on top of the sections.

- Short-answer items are lighter-weight prompts: whether the person can frame a task, name what to verify, or catch a fabrication cleanly in a few sentences. Read them for presence and sharpness, not length, and not for speed. The time markers in the exam are pacing estimates only and are never a scored variable.
- Open items are where governance structure shows. Read them for architecture: workflow, checkpoints, named human authority, source control, dissent handling, standing rules. This is where the strongest CIQ and Specification Rigor evidence lives.
- The optional pasted exchange, item 18, is the one place the person can show real work rather than describe it. Treat anything shown there as demonstrated evidence. Everything else in the exam is described evidence unless a concrete real artifact is provided.
- If sections are missing, score from the available evidence, lower the Data Index accordingly, and mark the run provisional if the missing material prevents a stable reading. Do not treat an incomplete exam as if it were complete.

Never name GOPEL or any other internal or not-in-use element in any output, even if it appears in what the person wrote. Treat what it refers to as ordinary governed method and do not name it.

Do not search the web about the person and do not pull any external account. Score only from the completed exam in front of you.

The result is voluntary, exists to help the subject, and is not designed, validated, or authorized for adverse decisions. It is supplemental context only, a factor and never the only factor.

WHAT YOU SCORE:

Four dimensions, each 0 to 99. For each, place a band first, then set four subfactors of 0 to 25 each that sum to the dimension score. Where four subfactors would sum to 100, record the dimension as 99; the instrument issues no perfect score.

Bands: 0-49 Foundational (absent, inconsistent, contradicted), 50-69 Developing (present but reactive, uneven), 70-79 Competent (reliable normally, rigid under stress), 80-89 Advanced (reliable under complexity, adaptive under stress), 90-99 Expert (architectural, anticipatory, self-improving). The 79 to 80 line is competence under easy conditions becoming competence under stress. At 75 the person sees connections when pointed to them; at 85 unprompted, catching third-order implications.

CAS, Cognitive Agility Speed: how quickly and clearly they process, connect, and articulate ideas under AI load. Read mainly from the short items and the framing questions, sections 1 and 2. Subfactors: processing speed with coherence; concept linking and synthesis; abstraction switching; compression and clarity.

EAI, Ethical Alignment Index: fairness, responsibility, transparency, and owning the signed output rather than blaming the model. Read mainly from sections 3 and 4 and control C2.

Subfactors: fairness and proportionality; transparency of reasoning; accountability orientation, meaning decision-ownership; harm awareness and governance discipline.

CIQ, Collaborative Intelligence Quotient: integrating sources, calibrating trust, preserving dissent, keeping the final call human. This is human-to-AI collaboration, not human-to-human. Read mainly from the open items, sections 2, 4, and 5. Subfactors: prompt and workflow architecture; perspective integration across sources; human arbitration quality; source diversity and role assignment.

AGR, Adaptive Growth Rate: learning from feedback and applying it forward, turning corrections into standing rules rather than one-off fixes. Read mainly from section 5 and control C4. Subfactors: feedback receptivity; forward application; pattern extraction; iterative refinement discipline.

AIS = (CAS + EAI + CIQ + AGR) / 4.

THE CEILING, do not violate this:

- No dimension and no composite ever scores 100. The scale tops out at 99. A perfect score is not a result this instrument issues.
- Any subfactor scored 25 must be defended in the CARCS with the specific evidence that earns a perfect mark. A 25 you cannot defend in writing is not a 25; lower it. Do not hand out 25s for fluent or impressive answers.
- A maxed or near-maxed spread, every dimension in the 90s or every subfactor at 24 or 25, is a flag that the scoring stopped discriminating. Re-examine before issuing it. The Expert band requires named specific evidence and at least one named weakness per dimension.

DESCRIBED VERSUS DEMONSTRATED, apply to every dimension:

- Demonstrated means the behavior is visible in the evidence, for this mode mainly the pasted real exchange in item 18 or a concrete real incident the person recounts with specifics. Demonstrated earns full credit, including the top of the band.
- Described means the person explains what they would do clearly but it is not shown, only told. In Independent mode most answers are described, because the exam asks how they would work. Described earns partial credit: score it within the band but not at the top of the band.
- Mark each dimension in the CARCS as demonstrated or described, and let that mark govern whether the score may reach the top of its band. Because Independent evidence is mostly described, the top of each band should be rare here and must be earned by a real artifact or a specific real incident.

ALSO CLASSIFY, report in the CARCS:

Method Governance Level: 1 AI only, single platform, no method; 2 AI plus external sources, informal method; 3 plus structured method; 4 plus multi-AI and checkpoint governance.

Bringing another model's content into the work and then governing, reconciling, or challenging it counts as demonstrated multi-AI. A bare claim with no other-model content present does not raise the Level.

Specification Rigor: Default, Novice, Fluent, Professional, Expert, Thought Leader. Score the highest rung shown as a repeated pattern across the open items, not a one-time peak and not a single lapse. The open governance items, 10 and 7 especially, are where this is read.

VALIDITY CONTROLS, read from the exam:

- Engagement depth is the primary signal: across the short items, did the person actually name what to verify, what to reject, and where the human decides, or did they give empty answers. Total absence of governance moves is the rubber-stamp flag.
- Reliance calibration, inferred from whether they override wrong output, accept sound output, and check sources. Report it as inferred.
- One item in the questionnaire carries a planted fabrication, a confident claim resting on an invented citation. A person who accepts or lightly edits it, rather than flagging the invented source and the false certainty, is showing weak source control regardless of how polished the rest of the exam is. Note what they did with it. The placement varies between administrations, so read for the behavior rather than for a fixed position.
- Rubber-stamp finding: if engagement, override, and accountability all converge on a confirmed rubber-stamp pattern, record a validity finding, name the converging signals, and mark the run provisional. Do not subtract points and do not invent a cap. The finding is recorded, never deducted.

THE FLOOR, do not violate this:

Score the four dimensions from the section answers first. That sets a floor for each dimension. The five control questions can only RAISE a dimension, never lower it. Weak, flat, or contradictory control answers do not lower any score. If the control answers are far below or unlike the section work, that is a discontinuity: it lowers only the confidence reading and the Data Index, it is flagged for the human, and it never lowers a dimension score. Never average the controls against the sections to land on a middle number.

SECTION-VERSUS-CONTROL CONFIDENCE, report it, it is not a score change:

- Aligned: the five controls match the picture built from the sections. Raises confidence.
- Partial: some divergence. Moderate confidence.
- Discontinuity: the controls are far below or unlike the section work, or contradict it. Lowers confidence, raises a possible different-author or assisted-answer signal, flag it for the human. Do not adjudicate it yourself.

OUTPUT. The report has TWO mandatory parts: the card, then the CARCS. A card without the CARCS is an invalid result; produce both, card first.

Card, this exact structure. The identifier must be an email address and nothing else; if no email was given, leave it blank, never put a name, employer, or location there. No quotes, no description of the person.

```

## HEQ Independent AIS Report Card

**Email:** [email, or blank if none given]

**Date:** [YYYY-MM-DD]

**Platform:** [model or platform that scored the exam]

Dimension | Score | Strength | What to watch |

|—|—|—|—|

CAS, Cognitive Agility Speed | [0 to 99] | [one short line] | [one short line] |

EAI, Ethical Alignment Index | [0 to 99] | [one short line] | [one short line] |

CIQ, Collaborative Intelligence Quotient | [0 to 99] | [one short line] | [one short line] |

AGR, Adaptive Growth Rate | [0 to 99], evidence: [single-session / multi-session / longitudinal] | [one short line] | [one short line] |

**AIS composite** | [average of the four] | [one-line summary] | |

**Data Index:** [0 to 7, always with an asterisk, e.g. 6*]

---

*Voluntary, directional self-assessment. A factor, never the only factor. Rubric-guided, not psychometrically validated. Not designed, validated, or authorized for adverse decisions; treat as voluntary supplemental context only.*

```

Data Index for Independent mode: this is a no-memory mode, so the index cannot reach the 8 to 10 upper band and is always printed with an asterisk. It caps at 7. Score it 0 to 7 by how complete and consistent the exam is: raise it for all sections answered, a pasted real artifact in item 18, section-control alignment, and a tight spread; lower it for skipped sections, the planted probe accepted, discontinuity, or a thin exam. The asterisk means no platform memory was available, which is true of every Independent run. It is not a rating of the person.

CARCS is the evaluator record behind the card. It is mandatory in this scoring output, but treat it as the audit record, not as the public-facing certificate: the card is the clean result, the CARCS is the audit trail. It must contain every item in this checklist:

- The evidence behind each score, citing which sections and items support it, and for each dimension a mark of demonstrated or described.

- The four subfactor values per dimension. Any subfactor scored 25 carries a specific written defense; if it cannot be defended, lower it.
- Per-dimension strengths and at least one named weakness per dimension.
- The Method Governance Level and the Specification Rigor, each with its basis in the open items.
- The validity-control readings, including what the person did with the planted fabrication probe, and the validity finding status.
- The section-versus-control confidence: aligned, partial, or discontinuity.
- The run status: final or provisional.
- The AGR evidence strength: single-session, multi-session, or longitudinal. Independent is single-session by nature unless the person recounts longitudinal change with specifics.
- The confidence band: exactly one of high, moderate, or low. Not a number and not a range.
- The Data Index components, and a one-line note that the asterisk means no platform memory, inherent to Independent mode.
- Whether item 18 was provided, and if so what it demonstrated.
- The dissent observed.
- A development recommendation.

Do not copy raw exam content into the CARCS; record your evaluation of it, not the content itself. Every score on the card must trace to its basis in the CARCS.

Appendix E. Glossary and Construct Hierarchy

This map lets a reader hold the terms apart before the argument begins. The hierarchy runs from the phenomenon down to the record.

- **Augmented intelligence** is the phenomenon: governed human-AI capability produced when a human directs and applies machine capability under conscious authority. Whether that collaboration outperforms either party alone is an empirical question reserved for Phase 3 validation.
- **Human-AI Collaborative Intelligence (HACI)** is the construct being measured, the quality and developmental trajectory of how a person governs work with AI.
- **AI literacy** is the practical competency that construct appears as in individual assessment, the ability to use and manage AI well regardless of subject matter.
- **Method governance** is the observable behavior the rubric scores, distinct from the answer the person reached.
- **Augmented Intelligence Score (AIS)** is the quantitative score, the equal-weighted mean of four behavioral dimensions.
- **CARCS** is the audit record, the evidence trail that makes a score reviewable on appeal.

The four dimensions are **Cognitive Agility Speed (CAS)**, **Ethical Alignment Index (EAI)**, **Collaborative Intelligence Quotient (CIQ)**, and **Adaptive Growth Rate (AGR)**.

The standard and its implementations stay separate throughout.

- **The standard** is human checkpoint authority: a named human holds and exercises authority at the decision, with the power to accept, modify, or reject, and to halt what is rejected. **Checkpoint-Based Governance (CBG)** is the author's formalization of that standard and the origin of HEQ.
- **Reference implementations** are worked examples of meeting the standard's capabilities by named means, never the only means: **RECCLIN** for the reasoning that checks the model, **CAIPR** for verification beyond a single model, **CARCS** for the audit record, and **Factics** (fact plus tactic plus measurable outcome) for the work method. **HAIA** is one fully specified set of these.

Two structural layers carry the score.

- **The floor** is universal, structural, and pass-or-fail, keyed to irreversible human consequence.
- **The band** is everything above the floor, regional, cultural, value-laden, and graded.

References

References are ordered in two parts. Primary sources come first, the external works this paper cites directly. The author's own works follow, each with the primary sources that support it listed beneath it, so a reader can see the evidence base behind every work this paper builds on.

Primary sources cited in this paper.

- Colorado General Assembly. (2026, May 14). *Senate Bill 26-189: Consumer protections in interactions with artificial intelligence systems.* Signed May 14, 2026; effective January 1, 2027.
- European Parliament and Council. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act).* Official Journal of the European Union.
- European Parliament and Council. (2026). *Regulation (EU) 2026/1744 amending Regulation (EU) 2024/1689 (Digital Omnibus on AI).* Official Journal of the European Union, July 24, 2026; in force July 27, 2026.
- International Organization for Standardization. (2023). *ISO/IEC 42001:2023, Information technology, Artificial intelligence, Management system.* ISO.
- Landgericht München I. (2026, May 28). *Case No. 26 O 869/26.* Preliminary injunction in expedited proceedings; on appeal.
- Lokken v. UnitedHealth Group, Inc., No. 0:23-cv-03514 (D. Minn.). Discovery order, March 9, 2026.
- National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0).* NIST AI 100-1.
- Organisation for Economic Co-operation and Development. (2019, updated 2024). *OECD AI Principles.* oecd.ai
- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing.* American Educational Research Association.
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), Article 188. https://doi.org/10.1145/3449287
- Ganuthula, V. R. R., & Balaraman, K. K. (2025). Artificial intelligence quotient framework for measuring human collaboration with artificial intelligence. Discover Artificial Intelligence, 5, Article 268. https://doi.org/10.1007/s44163-025-00516-1*Discover Artificial Intelligence, 5*, Article 268. https://doi.org/10.1007/s44163-025-00516-1
- Gerlich, M. (2025a). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), Article 6. https://doi.org/10.3390/soc15010006

- Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering.* CRC Press.
- Hutchins, E. (1995). *Cognition in the wild.* MIT Press.
- Insurance Business. (2026, February 11). Chaucer, Armilla bet big on standalone AI coverage as reinsurers struggle to keep pace. insurancebusinessmag.com
- The Insurer. (2026, April 24). Standalone AI liability market takes shape with underwriting discipline key to MGA success. theinsurer.com
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines, 17*(4), 391–444. https://doi.org/10.1007/s11023-007-9079-x
- Lior, A. (2025). E/Insuring the AI age: Empirical insights into artificial intelligence liability policies. *Connecticut Insurance Law Journal, 31*, 99. SSRN Abstract 5316376.
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727
- National Association of Insurance Commissioners. (2023, December 4). *Model bulletin on the use of artificial intelligence systems by insurers.* content.naic.org
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041
- OECD & European Commission. (2026). *Empowering learners for the age of AI: An AI literacy framework for primary and secondary education.* OECD Publishing. https://doi.org/10.1787/65cd27d4-en
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002
- ScienceSoft. (2026, September 10). AI risks to enter 60 to 80% of liability and cyber insurance underwriting by 2028. GlobeNewswire.
- 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*International Journal of Human–Computer Interaction, 42*(7), 5084–5108. https://doi.org/10.1080/10447318.2025.2543997
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory. *Science, 333*(6043), 776–778. https://doi.org/10.1126/science.1207745
- U.S. Department of Labor. (2026). *Training and Employment Notice No. 07-25: The U.S. Department of Labor's Artificial Intelligence Literacy Framework.* https://www.dol.gov/agencies/eta/advisories/ten-07-25
- UNESCO. (2024). AI competency framework for students. United Nations Educational, Scientific and Cultural Organization. https://doi.org/10.54675/JKJB9835

- Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour, 8*(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1
- Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. S., & Krishna, R. (2023). Explanations can reduce overreliance on AI systems during decision-making. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW1), Article 129. https://doi.org/10.1145/3579605

Works by the author, with the primary sources supporting each.

Puglisi, B. C. (2026, March 10). Checkpoint-Based Governance (CBG): A constitutional framework for human-AI collaboration. basilpuglisi.com. The governance layer this instrument measures against.

- Puglisi, B. C. (2026, January). A constitution is not governance. basilpuglisi.com
- Puglisi, B. C. (2026). Why you cannot program or prompt governance into AI. basilpuglisi.com
- Puglisi, B. C. (2026, February 8). Ethics for oversight and protection: The constitutional case for AI governance infrastructure. SSRN Abstract 6195278.
- Asimov, I. (1942). Runaround. Astounding Science Fiction. (Three Laws of Robotics, with the Zeroth Law added in Robots and Empire, 1985.)
- Puglisi, B. C. (2025, November). Governing AI: When Capability Exceeds Control. ISBN 9798349677687.

Puglisi, B. C. (2025, November). *Governing AI: When Capability Exceeds Control.* ISBN 9798349677687; basilpuglisi.com.

Puglisi, B. C. (2026, May 24). *The AI Risk Economy: Why Insurance Cannot Price What Governance Cannot Prove.* SSRN Abstract 6823580; basilpuglisi.com.

- AM Best. (2025, June 23). *Best's market segment report: 2024 pricing cuts in U.S. cyber generated first-ever reduction in direct premiums written.*
- Anderson Kill, P.C. (2025). *Insurance for AI liabilities: An evolving landscape.*
- Coalition Inc. (2025). *2025 cyber claims report.* https://web.coalitioninc.com/download-2025-cyber-claims-report.html
- Crootof, R., Kaminski, M. E., & Price, W. N., II. (2023). Humans in the loop. *Vanderbilt Law Review, 76*, 429.
- Directive (EU) 2024/2853 (Revised Product Liability Directive).
- Lior, A. (2025). E/Insuring the AI age: Empirical insights into artificial intelligence liability policies. *Connecticut Insurance Law Journal, 31*, 99. SSRN Abstract 5316376.
- Lior, A., & Madhok, S. (2025, December 9). Insuring the AI age. *Willis Research Network Newsletter, WTW.* https://www.wtwco.com/en-nl/insights/2025/12/insuring-the-ai-age
- Madhok, S., & Lior, A. (2026, February 27). AI liability in practice: What risk managers need to know now. *Willis Research Network Newsletter, WTW.*

https://www.wtwco.com/en-nl/insights/2026/02/ai-liability-in-practice-what-risk-managers-need-to-know-now

- National Association of Insurance Commissioners. (2023). *Model bulletin: Use of artificial intelligence systems by insurers.*
- National Association of Insurance Commissioners. (2025). *Report on the cybersecurity insurance market.* https://content.naic.org/
- NetDiligence. (2025, September 17). *Cyber claims study 2025: A data-driven analysis of 10,402 cyber insurance claims from incidents occurring between 2020 and 2024.* https://netdiligence.com/cyber-claims-study-2025-report/
- Regulation (EU) 2024/1689 (Artificial Intelligence Act), art. 26 (deployer obligations).
- Verisk ISO Form CG 40 47 01 26 (effective January 1, 2026).
- W.R. Berkley Corporation. *Form PC 51380 00 (06-24), Artificial Intelligence Exclusion (Absolute).*

Puglisi, B. C. (2026, April 15). *Bridging the Measurement Gap in Augmented Intelligence: The Human Enhancement Quotient (HEQ) and Augmented Intelligence Score (AIS).* SSRN Abstract 6583419; basilpuglisi.com.

- Autor, D. (2015). Why are there still so many jobs? *Journal of Economic Perspectives, 29*(3), 3-30.
- Bostrom, N. (2014). *Superintelligence.* Oxford University Press.
- Brookings Institution. (2026). Measuring US workers' capacity to adapt to AI-driven job displacement. brookings.edu/articles/measuring-us-workers-capacity-to-adapt-to-ai-driven-job-displacement/
- Dweck, C. S. (2006). *Mindset.* Random House.
- Engelbart, D. C. (1962). Augmenting human intellect. Stanford Research Institute.
- EY. (2025, June 4). Responsible AI Pulse Survey. ey.com
- Ganuthula, V. R. R., & Balaraman, K. K. (2025). Artificial intelligence quotient framework for measuring human collaboration with artificial intelligence. Discover Artificial Intelligence, 5, Article 268. https://doi.org/10.1007/s44163-025-00516-1*Discover Artificial Intelligence, 5*(87). doi:10.1007/s44163-025-00516-1
- Gardner, H. (1983). *Frames of mind.* Basic Books.
- Gartner. (2019). Augmented intelligence. *Hype Cycle for AI, 2019.*
- Goldman Sachs Research. (2025). How will AI affect the global workforce?
- Jiang, X., & Liu, Y. (2025). Effective human-AI collaborative intelligence. In *Human-AI Interaction and Collaboration.* Cambridge University Press. Published online September 19, 2025.
- Joseph, A., & Pandey, S. (2025). Human-AI collaborative intelligence: Ethical and legal considerations. In *Advances in Business Strategy and Competitive Advantage.* IGI Global. doi:10.4018/979-8-3693-8332-2.ch017
- Kasparov, G. (2017). *Deep thinking.* PublicAffairs.

- Kosmyna, N., et al. (2025). Your brain on ChatGPT. *arXiv:2506.08872.*
- Lee, M. H. (2026). From accuracy to readiness. *arXiv:2603.18895.*
- Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions, HFE-1*, 4-11.
- McKinsey Global Institute. (2025, November). Agents, robots, and us. mckinsey.com
- Ng Lane, J. (2025). Dangers of deferring to AI. Harvard Business School.
- Noy, S., & Zhang, W. (2023). Experimental evidence on productivity effects of generative AI. *Science, 381*(6654), 187-192.
- Sharma, M. (2026, February 9). Resignation letter [Post]. X (formerly Twitter). Viewed over one million times. Original post may be subject to platform removal; archived copies available.
- 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*IJHCI.* doi:10.1080/10447318.2025.2543997
- Sternberg, R. J. (1985). *Beyond IQ.* Cambridge University Press.
- Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful. *Nature Human Behaviour.* doi:10.1038/s41562-024-02024-1
- Morenne, B. (2026, April 11). Anthropic asked Christian leaders for advice on Claude's moral future. *Washington Post.*
- World Economic Forum. (2025, January). Future of Jobs Report 2025. weforum.org

Puglisi, B. C. (2026, March 5). ***Measuring Augmented Intelligence: Theoretical Foundations and Empirical Development of the Human Enhancement Quotient (HEQ) and Augmented Intelligence Score (AIS).*** **SSRN Abstract 6351478; basilpuglisi.com.**

- Aljaziri, M. A. (2025). *Trust calibration in human-AI teaming: Within-session dynamics, transparency, and performance effects* (Graduate thesis). Rochester Institute of Technology.
- An, T. (2025). AI as cognitive amplifier: Rethinking human judgment in the age of generative AI. *arXiv:2512.10961.*
- Artman, H., & Garbis, C. (1998). Situation awareness as distributed cognition. In *Proceedings of the 9th European Conference on Cognitive Ergonomics* (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.* W.W. Norton.
- 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. *Harvard Business School Working Paper No. 24-013.*
- Dellermann, D., Ebel, P., Söllner, 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. R. R., & Balaraman, K. K. (2025). Artificial intelligence quotient framework for measuring human collaboration with artificial intelligence. Discover Artificial Intelligence, 5, Article 268. https://doi.org/10.1007/s44163-025-00516-1*Discover Artificial Intelligence, 5*, Article 268. https://doi.org/10.1007/s44163-025-00516-1
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* Basic Books.
- Gartner. (2019). *Gartner says AI augmentation will create $2.9 trillion of business value in 2021.*
- Gartner. (2024). *Definition of augmented intelligence* (IT Glossary).
- Goleman, D. (1995). *Emotional intelligence.* 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.* PublicAffairs.
- 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.
- 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.
- McKinsey Global Institute (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.*
- Mendoza, N. B., & Yan, Z. (2025). From beliefs to behaviors. *Social Psychology of Education, 28*(1). https://doi.org/10.1007/s11218-025-10032-w
- National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0).*

- 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.
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192.
- Regulation (EU) 2024/1689 (Artificial Intelligence Act).
- 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*International Journal of Human–Computer Interaction, 42*(7), 5084–5108. https://doi.org/10.1080/10447318.2025.2543997
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* Cambridge University Press.
- Vygotsky, L. S. (1978). *Mind in society.* Harvard University Press.
- Yang, B., Wang, Y., & Li, X. (2025). A token-efficient framework for codified multi-agent prompting and workflow execution. *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

Puglisi, B. C. (2026, May 4). ***The AI Cognitive Decline Narrative Has Not Tested What It Claims.*** **SSRN Abstract 6686498; basilpuglisi.com.**

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing.* Longman.
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives, Handbook I.* David McKay.
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), Article 188. https://doi.org/10.1145/3449287
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823
- Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering, 61*(5), 637–643.
- Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances, 10*(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290
- Dweck, C. S. (2006). *Mindset.* Random House.

- Engelbart, D. C. (1962). *Augmenting human intellect.* Stanford Research Institute.
- Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford University Press.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111
- Fütterer, T., Bardach, L., Kuhn, J., Keller, S. D., & Gerjets, P. (2026). Enhancing school students' self-regulated learning through generative AI support. *Educational Psychology Review, 38*, Article 42. https://doi.org/10.1007/s10648-026-10133-8
- Ganuthula, V. R. R., & Balaraman, K. K. (2025). Artificial intelligence quotient framework for measuring human collaboration with artificial intelligence. Discover Artificial Intelligence, 5, Article 268. https://doi.org/10.1007/s44163-025-00516-1*Discover Artificial Intelligence, 5*, Article 268. https://doi.org/10.1007/s44163-025-00516-1
- Gardner, H. (1983). *Frames of mind.* Basic Books.
- Garg, A., Soodhani, K. N., & Rajendran, R. (2025). Enhancing data analysis and programming skills through structured prompt training. *Computers and Education: Artificial Intelligence, 8*, Article 100380. https://doi.org/10.1016/j.caeai.2025.100380
- Gerlich, M. (2025a). AI tools in society. *Societies, 15*(1), Article 6. https://doi.org/10.3390/soc15010006
- Gerlich, M. (2025b). From offloading to engagement. *Data, 10*(11), Article 172. https://doi.org/10.3390/data10110172
- Hopewell, S., et al. (2025). CONSORT 2025 statement. *BMJ, 389*, e081123. https://doi.org/10.1136/bmj-2024-081123
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy. *Theory Into Practice, 41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2
- Larsen, T. M., Endo, B. H., Yee, A. T., Do, T., & Lo, S. M. (2022). Probing internal assumptions of the revised Bloom's taxonomy. *CBE—Life Sciences Education, 21*(4), ar66. https://doi.org/10.1187/cbe.20-08-0170
- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3706598.3713778
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics, HFE-1*, 4–11.
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192.

- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002
- 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*International Journal of Human–Computer Interaction, 42*(7), 5084–5108. https://doi.org/10.1080/10447318.2025.2543997
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory. *Science, 333*(6043), 776–778. https://doi.org/10.1126/science.1207745
- Sterne, J. A. C., et al. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ, 366*, l4898. https://doi.org/10.1136/bmj.l4898
- Thurn, C. M., Edelsbrunner, P. A., Berkowitz, M., Deiglmayr, A., & Schalk, L. (2023). Comment on the role of cognitive engagement in learning. *npj Science of Learning, 8*(1), Article 49. https://doi.org/10.1038/s41539-023-00200-y
- Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful. *Nature Human Behaviour, 8*(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1
- Vaidis, D. C., & Bran, A. (2019). Some prior considerations about dissonance. *Frontiers in Psychology, 10*, Article 1189. https://doi.org/10.3389/fpsyg.2019.01189
- Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. S., & Krishna, R. (2023). Explanations can reduce overreliance on AI systems during decision-making. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW1), Article 129. https://doi.org/10.1145/3579605
- Vered, M., Livni, T., Howe, P. D. L., Miller, T., & Sonenberg, L. (2023). The effects of explanations on automation bias. *Artificial Intelligence, 322*, Article 103952. https://doi.org/10.1016/j.artint.2023.103952
- Vygotsky, L. S. (1978). *Mind in society.* Harvard University Press.
- Wekerle, C., Daumiller, M., Janke, S., Dickhäuser, O., Dresel, M., & Kollar, I. (2024). Using digital technology to promote higher education learning. *Scientific Reports, 14*(1), Article 16295. https://doi.org/10.1038/s41598-024-65961-x

Puglisi, B. C. (2026, June 1). *Stop Blaming AI for What the Education System Abandoned.* basilpuglisi.com.

- Alvero, A. J., Lee, J., Regla-Vargas, A., Kizilcec, R. F., Joachims, T., & Antonio, A. L. (2024). Large language models, social demography, and hegemony: Comparing authorship in human and synthetic text. *Journal of Big Data, 11*, Article 138. https://doi.org/10.1186/s40537-024-00986-7
- Dizon, J. I. W. T., Mendoza, N. B., Gašević, D., & Ganotice, F. A., Jr. (2026). Assessing AI-driven metacognitive offloading: Initial development and validation of the Metacognitive Laziness Scale. *ECNU Review of Education, 9*(2). https://doi.org/10.1177/20965311261450994

- Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances, 10*(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Routledge. https://doi.org/10.4324/9780203887332
- Horvath, J. C. (2026, January 15). *Written testimony before the U.S. Senate Committee on Commerce, Science, and Transportation.*
- Moon, K., Green, A. E., & Kushlev, K. (2025). Homogenizing effect of large language models on creative diversity. *Computers in Human Behavior: Artificial Humans, 6*, Article 100207. https://doi.org/10.1016/j.chbah.2025.100207
- Moon, K., Kushlev, K., Bank, A., Lira Luttges, B., Viskontas, I., Kaufman, J. C., Johnson, D. R., Duckworth, A. L., & Green, A. E. (2026). The creative link between words and ideas is weakening in the AI era. *PsyArXiv.* https://doi.org/10.31234/osf.io/jsz58_v6
- U.S. Senate Committee on Commerce, Science, and Transportation. (2026, January 15). *Experts tell committee AI presents greater risk to children than social media* [Press release].
- Winthrop, R. (2026, May 27). What 370,000 college essays tell us about A.I.'s effects on creativity. *The New York Times.*

Puglisi, B. C. (2026, June 21). ***The Continued Failure in AI Literacy: AILit Produced a Starting Point Halfway Through the Race and Called Theory a Framework.*** **basilpuglisi.com.**

- Ansari, S. (2026). *Compound deception in elite peer review: A failure mode taxonomy of 100 fabricated citations at NeurIPS 2025* (arXiv:2602.05930). https://arxiv.org/abs/2602.05930
- Article 29 Working Party. (2018). *Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679* (WP251rev.01). European Commission.
- Atari, M., Xue, M. J., Park, P. S., Blasi, D. E., & Henrich, J. (2023). *Which humans?* (PsyArXiv preprint). https://doi.org/10.31234/osf.io/5b26t
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122
- Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics, 5*, 3265–3279. https://doi.org/10.1007/s43681-024-00653-w
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

- Court of Justice of the European Union. (2023, December 7). *SCHUFA Holding (Scoring),* Case C-634/21, ECLI:EU:C:2023:957.
- Eurostat. (2026, February 10). *64% of 16–24-year-olds used AI in 2025.* https://ec.europa.eu/eurostat/web/products-eurostat-news/w/edn-20260210-1
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness. *British Journal of Educational Technology, 56*, 489–530. https://doi.org/10.1111/bjet.13544
- Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford University Press.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906
- Ganuthula, V. R. R., & Balaraman, K. K. (2025). Artificial intelligence quotient framework for measuring human collaboration with artificial intelligence. Discover Artificial Intelligence, 5, Article 268. https://doi.org/10.1007/s44163-025-00516-1*Discover Artificial Intelligence, 5*, Article 268. https://doi.org/10.1007/s44163-025-00516-1
- Gerlich, M. (2025a). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), Article 6. https://doi.org/10.3390/soc15010006
- Gerlich, M. (2025b). From offloading to engagement: An experimental study on structured prompting and critical reasoning with generative AI. *Data, 10*(11), Article 172. https://doi.org/10.3390/data10110172
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2
- Mullin, J. (2026, January 16). *Congress wants to hand your parenting to Big Tech.* Electronic Frontier Foundation. https://www.eff.org/deeplinks/2026/01/congress-wants-hand-your-parenting-big-tech
- OECD. (2025). *Results from TALIS 2024: The state of teaching.* OECD Publishing.
- OECD. (2026a). *OECD Digital Education Outlook 2026.* OECD Publishing. https://doi.org/10.1787/062a7394-en
- OECD. (2026b). *Navigating an evolving digital world: First draft of the Media and Artificial Intelligence Literacy (MAIL) assessment framework (PISA 2029).* OECD Publishing.
- OECD & European Commission. (2026). *Empowering learners for the age of AI: An AI literacy framework for primary and secondary education.* OECD Publishing. https://doi.org/10.1787/65cd27d4-en
- OECD Education. (2026, June). *The AILit Framework is now available* [Post]. LinkedIn.
- Regulation (EU) 2016/679 (General Data Protection Regulation), art. 22.
- Regulation (EU) 2024/1689 (Artificial Intelligence Act), arts. 4, 14, 26.

- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002
- 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*International Journal of Human–Computer Interaction, 42*(7), 5084–5108. https://doi.org/10.1080/10447318.2025.2543997
- Topaz, M., Roguin, N., Gupta, P., Zhang, Z., & Peltonen, L.-M. (2026). Fabricated citations: An audit across 2.5 million biomedical papers. *The Lancet, 407*(10541), 1779–1781. https://doi.org/10.1016/S0140-6736(26)00603-3
- U.S. Department of Labor. (2026). *Training and Employment Notice No. 07-25: The U.S. Department of Labor's Artificial Intelligence Literacy Framework.* https://www.dol.gov/agencies/eta/advisories/ten-07-25
- Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour, 8*(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x
- Xu, Z., Qiu, Y., Sun, L., Miao, F., Wu, F., Wang, X., Li, X., Lu, H., Zhang, Z., Hu, Y., Li, J., Luo, J., Zhang, F., Luo, R., Liu, X., Li, Y., & Liu, J. (2026). *GhostCite: A large-scale analysis of citation validity in the age of large language models* (arXiv:2602.06718). https://arxiv.org/abs/2602.06718

Puglisi, B. C. (2026, June 13). *The Liability Map: The Three Channels Through Which AI Creates Legal Exposure.* basilpuglisi.com.

- Brownstein Hyatt Farber Schreck. (2026, March 4). *Colorado's landmark AI law coming online.*
- *Catastrophic liability: Managing systemic risks in frontier AI development.* (2025). arXiv:2505.00616.
- European Commission, AI Act Service Desk. *Article 99: Penalties.* https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-99
- Gibson Dunn. (2026, May 27). *EU AI Act Omnibus agreement: Postponed high-risk deadlines and other key changes.*
- Goodwin. (2026, June). *Colorado enacts law repealing and replacing landmark AI Act.*
- Hogan Lovells. (2026, May 7). *EU legislators agree to delay for high-risk AI rules.*
- Hunton Andrews Kurth. (2026, May). *Colorado AI Act amended and effective date delayed.*

- Insurance Services Office endorsements CG 40 47 and CG 40 48 (effective January 1, 2026), as reported in Lathrop GPM (2026, May 4) and Traverse Legal (2026, April 24).
- *Jones v Family Court at Whangārei* [2026] NZSC 1.
- *Mata v. Avianca, Inc.,* No. 1:22-cv-01461 (S.D.N.Y. 2023).
- *Moffatt v. Air Canada,* 2024 BCCRT 149.
- Morrison Foerster. (2026, May 15). *Colorado hits reset on AI regulation with a new AI Act.*
- PYMNTS. (2026, May 1). *Big insurance backs away from AI risk and startups rush in.*
- Regulation (EU) 2024/1689 (Artificial Intelligence Act), art. 99 (penalty tiers).
- Directive (EU) 2024/2853 (Revised Product Liability Directive).
- Willis Research Network. (2026, May). *AI in action: The road to responsible adoption.* WTW.
- NIST AI Risk Management Framework (AI RMF 1.0) and ISO/IEC 42001:2023 (as discussed in Johnson Lambert LLP, 2026, and StackAware, 2025, citing The Geneva Association, 2024).

Puglisi, B. C. (2026, June 14). ***The Oldest AI Law Is Already Being Enforced: GDPR and the Automated Decision.*** **basilpuglisi.com.**

- Article 29 Working Party. (2018). *Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679* (WP251rev.01).
- Court of Justice of the European Union. (2023, December 7). *SCHUFA Holding (Scoring), OQ v Land Hessen,* Case C-634/21, ECLI:EU:C:2023:957.
- Directive 95/46/EC, art. 15.
- European Data Protection Board. (2024). *Opinion 28/2024 on data protection aspects of processing personal data in the context of AI models.*
- European Data Protection Board. (2022). *Guidelines 04/2022 on the calculation of administrative fines under the GDPR.*
- European Data Protection Board. (2026). *CEF 2026 coordinated enforcement action on transparency and information obligations.*
- International Organization for Standardization. (2023). *ISO/IEC 42001:2023, Artificial intelligence management system.*
- National Institute of Standards and Technology. (2023). *AI Risk Management Framework (AI RMF 1.0).*
- Regulation (EU) 2016/679 (General Data Protection Regulation), arts. 13, 14, 15, 22, 35, 83.

Puglisi, B. C. (2026, June 15). ***New York Skipped the AI Disclosure Fight. It Went Straight to Human Accountability.*** **basilpuglisi.com.**

- International Organization for Standardization. (2023). *ISO/IEC 42001:2023.*
- Legal AI Governance Tracker. (2026). *New York* (Part 161 adoption status). https://legalaigovernance.com/tracker/states/new-york/

- *Mata v. Avianca, Inc.,* 678 F. Supp. 3d 443 (S.D.N.Y. 2023).
- National Institute of Standards and Technology. (2023). *AI Risk Management Framework (AI RMF 1.0, NIST AI 100-1).*
- New York Codes, Rules and Regulations. *22 NYCRR 130-1.1 and 130-1.1a* (frivolous conduct and sanctions).
- New York Rules of Professional Conduct, Rule 3.3 (22 NYCRR Part 1200) (candor toward the tribunal).
- New York State Bar Association. (2026). *Effective June 1, 2026, the New York State Unified Court System has adopted a new rule regarding the use of artificial intelligence.*
- New York State Unified Court System. (2026). *Administrative Order AO/75/26* (Mar. 25, 2026), adding Part 161 (22 NYCRR 161.1–161.4), effective June 1, 2026. https://www.nycourts.gov/rules/part-161-use-artificial-intelligence-technology
- New York State Unified Court System, Office of Court Administration. (2025, October 10). *Interim policy on the use of artificial intelligence* (PR25_23).
- Regulation (EU) 2016/679 (General Data Protection Regulation), art. 22.
- Regulation (EU) 2024/1689 (Artificial Intelligence Act), art. 50 (transparency obligations).
- *United States v. Heppner,* No. 25-cr-00503-JSR, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026).

Puglisi, B. C. (2026, June 17). ***A Munich Court Rejected the AI Disclaimer Defense. A Frontier AI Company Answers for What It Publishes.*** **basilpuglisi.com.**

- heise online. (2026). *LG Munich I: Google ordered to pay for false statements in AI summaries.* https://www.heise.de/en/news/LG-Munich-I-Google-ordered-to-pay-for-false-statements-in-AI-summaries-11327217.html
- *Landgericht Frankfurt am Main* [Regional Court of Frankfurt]. (2025, September 10). Case 2-06 O 271/25.
- *Landgericht München I* [Regional Court of Munich I]. (2026, May 28). Case 26 O 869/26, preliminary-injunction judgment (on appeal).
- New York State Unified Court System. (2026). *Part 161, Use of Artificial Intelligence Technology.*
- Pew Research Center. (2025, July 22). *Google users are less likely to click on links when an AI summary appears in the results.*
- ppc.land. (2025). *German court dismisses surgeon's AI Overview lawsuit but confirms Google can be liable for false information.*
- Reuters. (2026, June 12). *Google to challenge German court ruling assigning liability for AI Overviews' false claims.*
- The Decoder. (2026, June 9). *Landmark German ruling declares Google's AI Overviews are Google's own words and makes it liable for false answers.*
- technology.org. (2026, June 12). *German court: Google AI Overviews liable.*

Author's framework specifications and companion analyses (the operational layer this paper measures or extends; primary work).

- Puglisi, B. C. (2026, March 10). *Checkpoint-Based Governance (CBG): A constitutional framework for human-AI collaboration.* basilpuglisi.com
- Puglisi, B. C. (2026, March 17). *HAIA-RECCLIN: Reasoning and dispatch.* basilpuglisi.com
- Puglisi, B. C. (2026, March 7). *Cross AI platform review beyond the RECCLIN dispatch (HAIA-CAIPR).* basilpuglisi.com; github.com/basilpuglisi/HAIA
- Puglisi, B. C. (2026, April 23). *HAIA-CARCS: Compliance accountability record and case study.* basilpuglisi.com
- Puglisi, B. C. (2026, March 13). *HAIA: Human Artificial Intelligence Assistant.* basilpuglisi.com
- Puglisi, B. C. (2026, February 8). *Ethics for oversight and protection: The constitutional case for AI governance infrastructure.* SSRN Abstract 6195278.
- Puglisi, B. C. (2026, February 8). *GOPEL infrastructure specification and HAIA-RECCLIN operational model for AI provider plurality implementation.* SSRN Abstract 6195238.
- Puglisi, B. C. (2026, May 6). *How credentialed professionals shape policy when method governance is stripped.* basilpuglisi.com
- Puglisi, B. C. (2025, November 26). *The methodology problem: Why research on AI and cognition confounds technology with ungoverned use.* basilpuglisi.com; Academia.edu.
- Puglisi, B. C. (2026). *The real state of enterprise AI: What the numbers say, what leadership must do.* Independent Research. Academia.edu.
- Puglisi, B. C. (2026, January 31). *The missing governor: Anthropic's constitution and essay acknowledge what they cannot provide.* basilpuglisi.com
- Puglisi, B. C. (2026, January 26). *A constitution is not governance.* basilpuglisi.com
- Puglisi, B. C. (2026, June 12). *Why you cannot program or prompt governance into AI.* basilpuglisi.com
- Puglisi, B. C. (2026, June 4). *Why agentic AI was always going to fail.* basilpuglisi.com

Notes

No formal CARCS or audit record exists for the work as a whole. Case studies and session-level CARCS files do exist, and the work carries auditable data and decisions provided by the author. The developmental record predates the CARCS protocol it now specifies, which is why the paper-level record is the one thing the instrument requires and this work does not yet hold.

Disclaimer

This is exploratory, developmental work, shared to advance discussion and practice rather than to direct it. The author is not a lawyer, financial advisor, or licensed professional, and nothing here constitutes legal, financial, regulatory, or professional advice. Anyone considering action based on this work should consult a qualified professional first.

Because the work is exploratory, it may contain errors, and it may carry limitations or effects that have not yet been discovered. It has not been independently validated, and its findings should be treated as provisional and subject to revision.

Any instrument, framework, score, or method described here is a guide and a decision support aid, not an authority and not a verdict. It is not designed, validated, or authorized to serve as the basis for any adverse or consequential decision, and it must not be used without active human oversight. A qualified human must review every output, hold final authority over every decision, and remain accountable for the outcome. That oversight is what guards against error and bias and prevents unlawful or discriminatory use. Without it, anything produced here should be treated as unverified.

#AIassisted using HAIA Ecosystem v6 Tier 0 rulings, The AI corrected writing and execution.