Artificial Intelligence Quotient: A Systematic Survey and Analytical Framework for a Nascent Multidisciplinary Construct

Wim Sweldens* Bohang Sun† Sushma Nagaraja Grellscheid‡

31 May 2026

Abstract

The term Artificial Intelligence Quotient (AIQ) has emerged across multiple disciplines including cognitive psychology, computer science, education, enterprise consulting, and marketing, each producing distinct frameworks for what AIQ is and how it should be measured. As there is no systematic comparative analysis to date, this paper surveys ten independently developed AIQ approaches and introduces a two-dimensional analytical model to differentiate them: whether AIQ is conceived as a stable trait or a perishable skill, and what each framework treats as its primary object of measurement. The analysis reveals that most frameworks cluster around a shared assumption that AIQ is best measured through performance, while calibration, the ability to accurately read where AI succeeds and fails, remains largely implicit. We argue that calibration is the central mechanism through which knowledge becomes performance, and that its absence from most frameworks represents a significant conceptual gap. The parallel emergence of AIQ across so many fields reflects a genuine shared intuition that deserves coordinated academic attention.

1 Why AIQ, Why Now

Artificial intelligence (AI) has rapidly transitioned from a specialized technical domain to a pervasive component of everyday cognitive and organizational processes. It now plays a central role in activities such as writing, decision-making, diagnosis, design, and communication. This transformation has outpaced the development of corresponding frameworks for evaluating human capability in AI-mediated environments. While established constructs exist for assessing knowledge, reasoning ability, and social intelligence, there remains no widely accepted framework for measuring how effectively individuals interact with and leverage artificial intelligence systems. The recent publication of *Magnifica Humanitas* (Leo XIV, 2026), the first papal encyclical devoted to artificial intelligence, calls explicitly for "informed users" as a structural prerequisite for legitimate AI governance, alongside robust legal frameworks and independent oversight.

In response to this gap, the concept of Artificial Intelligence Quotient (AIQ) has emerged across several domains within a remarkably short time frame. Researchers in cognitive psychology, computer science, and education, alongside practitioners in enterprise consulting and marketing, have independently introduced the term to describe various dimensions of human–AI capability. Notably, these efforts have largely developed in parallel, with minimal cross-referencing or coordination.

This convergence is unlikely to be coincidental. Rather, it suggests the articulation of a shared underlying problem: existing measures of human capability are insufficient for a context in which cognitive tasks are increasingly distributed between humans and AI systems. The central question, therefore, is whether these independently developed AIQ frameworks refer to a common construct or whether the shared terminology obscures substantive conceptual divergence.

*Department of Computer Science, University of Leuven, Belgium; Co-founder, Kiswe Inc. wim@sweldens.com
†Department of Computer Science and Technology, University of Cambridge, UK. bs835@cam.ac.uk
‡Department of Informatics, University of Bergen, Norway. Department of Biosciences, Durham University, UK sushma.grellscheid@uib.no

This paper addresses that question through a structured comparative analysis. We survey ten distinct AIQ frameworks, introduce two analytical dimensions that capture their fundamental differences, and map the resulting conceptual landscape. The objective is not to propose a unified definition of AIQ, but to render visible an emerging interdisciplinary conversation that remains fragmented across domains.

2 The Quotient Tradition: From IQ to EQ to AIQ

A further consideration concerns the historical and conceptual controversies associated with quotient-based measures themselves. The earliest and most influential example is the intelligence quotient (IQ), introduced by Alfred Binet and Théodore Simon in the early twentieth century Binet and Simon (1905). Constructs such as IQ have never functioned as purely neutral scientific measurements. Rather, they have historically embodied normative assumptions regarding which forms of cognition are socially valued, how intelligence should be operationalized, and who benefits from particular systems of classification. IQ testing has long been criticized for reducing complex and culturally situated forms of human capability into simplified scalar metrics that may obscure important contextual, social, and ethical dimensions of intelligence.

Although initially intended as a practical tool for identifying educational needs, IQ evolved into a generalized measure of cognitive ability with widespread institutional applications. The appeal of such quotients lies in their standardization and comparability. Numerical representations enable efficient communication, benchmarking, and decision-making across contexts. However, this reduction of complex constructs to single metrics has consistently attracted criticism, including that IQ mistakes one dimension of human capability for the whole, that it encodes cultural assumptions as universal truths, that it creates false precision around something irreducibly complex. Despite these largely valid criticisms IQ persists because of the tangible need it addresses — a reliable, standardized way to compare human cognitive capability.

A similar trajectory can be observed in the development of emotional intelligence (EQ). When Daniel Goleman popularized EQ in the 1990s, Goleman (1995), he was responding to a genuine perceived gap: IQ predicted performance in controlled settings but left unexplained why some highly intelligent people failed in organizations and social environments while others with more modest cognitive scores thrived. EQ named something that practitioners already sensed was important — to capture dimensions of human capability not accounted for by IQ, particularly those related to emotional awareness and interpersonal effectiveness. It faced the same criticisms as IQ: reductive, imprecise, culturally contingent. It achieved the same result: widespread adoption because the underlying relevance in organizational and social contexts was real.

The pattern is consistent across IQ and EQ. A new quotient emerges when existing measures fail to capture something that practitioners and researchers sense is consequential. It is contested on the grounds that it oversimplifies, yet it persists because the gap it addresses is genuine. Domain-specific quotients have continued to proliferate — Forrester's Robotics Quotient (RQ) from 2015 is one example — each responding to a new gap in existing measurement frameworks.

AIQ follows this established pattern. It emerges in response to a recognized limitation in existing frameworks: neither IQ nor EQ adequately captures the human competencies required to effectively interact with AI systems, including how appropriately they will trust its outputs, or how effectively they will navigate the places where it fails. The increasing integration of AI into cognitive workflows has introduced new forms of human capability, including the variable ability to interpret AI outputs, calibrate trust, and integrate machine-generated insights into decision-making processes. The gap in the existing measurement landscape for this new human capability has resulted in multiple communities independently reaching for the same solution: a quotient for artificial intelligence.

The development of AIQ can be visualized within a broader historical tradition of "quotient" constructs designed to quantify the complex dimensions of human capability. Any attempt to construct a standardized measure of human capability in relation to AI systems risks reproducing similar problems of reductionism, overconfidence in quantification, and hidden value assumptions embedded within assessment design. The question is therefore not only whether AIQ can be measured reliably, but also which forms of human–AI interaction are being privileged, rewarded, or excluded by particular frameworks. Recognizing these historical critiques does not invalidate the AIQ concept; rather, it highlights the importance of transparency regarding underlying assumptions, careful attention to context, and caution against treating AIQ scores as exhaustive representations of human capability.

The history of IQ testing is, however, considerably more fraught than its practical origins suggest. Within two decades of Binet and Simon's work, intelligence testing was appropriated by psychologists active in the eugenics movement in the United States and deployed to justify racial hierarchy, immigration restriction, and forced sterilization. Cave (2020) documents how the concept of intelligence has historically functioned as a hegemonic ideology that legitimates dominance hierarchies along racial, gender and class lines, and argues that this legacy continues to shape contemporary AI debates. The mechanism underlying these abuses is consistent across their varied historical manifestations: IQ was framed as a fixed, intrinsic property of individuals — immutable in the manner of a biological characteristic — and thereby positioned as a legitimate basis for differential treatment, exclusion, and ranking. Constructs that purport to measure what individuals fundamentally are, rather than what they have learned, carry an inherent potential for hierarchical misappropriation.

This historical legacy is directly relevant to the design of AIQ frameworks, and not only as a cautionary parallel. It also indicates a structural condition that any credible AIQ framework must satisfy. A framework that conceptualizes AIQ as a stable trait — fixed within individuals and substantially resistant to development — reproduces the same structural risk as IQ: it functions as a sorting mechanism that may naturalize inequality rather than support equitable development. By contrast, a framework that conceptualizes AIQ as a perishable skill — assessed not against a fixed standard but against a continuously shifting technological frontier — is ontologically distinct. Such a framework makes no claim regarding what individuals fundamentally are. It evaluates only the degree to which an individual has developed the capacity to navigate the current state of AI capability, a state that is itself subject to continuous change. Under this design, a low AIQ score carries no implication regarding innate capacity; it reflects only the current state of a learnable and improvable competence. Furthermore, because the frontier itself evolves, even high-scoring individuals must engage in ongoing recalibration. No individual begins with an inherent advantage relative to AI systems that did not previously exist. The egalitarian implications of this design are not incidental but follow structurally from the nature of what is being assessed.

The viability of AIQ, as with its predecessors, depends on two conditions: first, that it captures a meaningful and distinct dimension of human capability; and second, that its measurement frameworks accurately reflect this dimension without introducing misleading precision. These considerations motivate the present survey.

3 The Landscape: Ten Approaches

The term AIQ has been independently coined across five distinct domains: cognitive psychology, computer science, education and workplace assessment, enterprise consulting, and marketing. The ten approaches surveyed below are presented roughly chronologically and appear to have evolved independently.

(1) Klein, Hoffman, and Mueller / ShadowBox Klein (2020); Klein et al. (2023)

The earliest approach in this survey predates the generative AI moment entirely. Gary Klein, Robert Hoffman, and Shane Mueller, working through ShadowBox LLC and funded in part by the Air Force Research Laboratory, developed AIQ as a practitioner toolkit designed to help people build appropriate mental models of the specific AI systems they use. Their starting observation was that many promising AI systems were being rejected or underused because operators did not understand them well enough to trust them appropriately. AIQ in their framework is not a score but a set of five instruments — including a cognitive tutorial, a ShadowBox scenario tool, and a self-explaining scorecard — designed to improve what they call appropriate trust: not blind reliance on AI, nor reflexive skepticism, but calibrated judgment about when a specific system can and cannot be trusted. Their work is grounded in naturalistic decision-making research and is explicitly practical in orientation.

(2) Qin, Lu et al. Qin et al. (2025)

Qin et al. (2025) introduced AIQ as a measurable individual difference — a person's stable ability to use AI to perform a wide variety of tasks. Their approach is the most empirically rigorous in this survey. Across five studies using archival data, longitudinal experiments, and controlled tasks, they demonstrated that a general AIQ factor can be statistically extracted from individual performance on AI-assisted tasks, that this factor is stable over time, and that it predicts future performance across different AI systems and task types. Crucially, they show that AIQ is distinct from IQ, emotional intelligence, AI literacy, and computer literacy.

Their methodology is explicitly modeled on how IQ was established as a construct — find a general factor, demonstrate stability, demonstrate predictive validity.

(3) Forrester Research Forrester Research (2024, 2026)

In March 2024, Forrester Research introduced AIQ as an enterprise workforce readiness metric, building on their earlier Robotics Quotient (RQ) framework from 2015. Defined as the readiness of individuals, teams, and organizations to adapt to, collaborate with, trust, and generate business results from AI, Forrester's AIQ is measured through a 12-statement survey assessing four competencies: understanding of AI capabilities and limitations, hard skills including prompt engineering, confidence and motivation, and ethics, risk and privacy awareness. Forrester has deployed this assessment to enterprise clients across the United States, United Kingdom, Germany, France, and Australia, and published longitudinal data showing that employee readiness has failed to keep pace with organizational AI deployment. Their AIQ is primarily an organizational diagnostic tool rather than an individual credential.

(4) Value Creation Innovation Institute / VCII Value Creation Innovation Institute (2024)

The Value Creation Innovation Institute, based in Singapore, developed AIQ as a structured commercial assessment framework built on six pillars: prompting skill, training competence, testing and evaluation, optimization, understanding AI limitations, and humanizing AI interaction. VCII offers two assessments — an AI Core Knowledge test and a scenario-driven AIQ Aptitude Test — developed over what they describe as 5,000 human hours of research and testing. Their explicit ambition is for AIQ to become a global standard alongside IQ and EQ. The framework acknowledges that AI capabilities evolve and that assessments must be continuously updated, though it stops short of building temporal decay into the credential itself.

(5) Ganuthula and Balaraman Ganuthula and Balaraman (2025a,b)

Published first as an arXiv preprint in February 2025 and subsequently in the Springer journal Discover Artificial Intelligence, this framework introduces AIQ as a multidimensional measure of human-AI collaborative intelligence. The authors propose eight dimensions including prompt engineering intelligence, critical evaluation capability, integration intelligence, and adaptive learning capability. The framework is explicitly positioned for educational and professional contexts, drawing on established cognitive assessment methodologies and item response theory. It represents the most academically complete framework in the survey in terms of dimensional specification, though it remains primarily theoretical with implementation envisioned through adaptive digital assessment platforms.

(6) aiq.works aiq.works (2025)

A commercial assessment platform launched in November 2025, aiq.works offers what it describes as a research-validated AIQ assessment across eight dimensions, using over 400 calibrated items and item response theory methodology. The platform appears to be closely related to the Ganuthula and Balaraman academic framework, though the precise relationship is not publicly documented. It represents the most methodologically rigorous commercial implementation of AIQ assessment currently available.

(7) WSU / MDPI Pereyda and Holder (2025)

A research team at Washington State University published an AIQ framework in December 2025 that is fundamentally different from all others in this survey: it measures the intelligence of AI systems rather than human capability. Using two domain measures — dissimilarity and complexity — the framework generates an AIQ score for an AI agent reflecting how broadly intelligent it is across different task domains. This approach shares the name and the IQ analogy with the others but belongs to an entirely different research tradition. It is included here for completeness and because it illustrates the breadth of the conceptual territory the acronym currently occupies.

(8) Sweldens Sweldens (2025)

One of the present authors introduced AIQ in a series of Substack essays beginning in September 2025, defining it as a measure of a person's ability to navigate the jagged frontier of AI capability Dell'Acqua et al. (2023) — the uneven and shifting boundary where AI performs reliably in some domains and fails unpredictably in others. The framework proposes five dimensions: context awareness, prompting, anticipating failures, verification and ethics, and workflow integration. Two features distinguish it from other approaches in this survey. First, the explicit centering of the jagged frontier as the object of measurement: AIQ is not about general performance with AI but specifically about the ability to read where AI can and cannot be trusted. Second, a temporal decay mechanic built into the credential: because the frontier shifts as AI capabilities evolve, an AIQ score expires and requires periodic recalibration.

(9) Drake / Soulcraft Drake (2026)

Evan Drake of Soulcraft, writing in April 2026, coined AIQ to describe an organization's presence, sentiment, and position in AI-generated outputs — essentially an evolution of search engine optimization into the generative AI era. This is a marketing concept rather than a measure of human capability, and it points the lens at organizations rather than individuals.

(10) Nebuli / aiq.org Nebuli (2023)

The research lab Nebuli, operating under the domain aiq.org, uses AIQ to stand for Augmented Intelligence Quotient, referring to their methodology for building human-centric, vertically specialized AI models. Neither approach is primarily concerned with measuring human capability in relation to AI, and both are noted here to illustrate the breadth of territory the acronym currently covers.

Despite their diversity, these approaches share a common intuition: that existing measures of human capability do not adequately capture the competencies required for effective interaction with AI systems.

4 Two Analytical Dimensions

The surveyed approaches share a name and a broad intuition but diverge sharply along two fundamental dimensions when examined closely.

Despite methodological diversity from psychometric instruments, practitioner toolkits, scenario-based assessments and some differences in domain or purpose, the deepest differences concern two analytical dimensions for mapping the most important differences across the landscape.

4.1 Trait versus Perishable Skill

The first dimension concerns whether AIQ is conceptualized as a stable individual trait or a perishable skill. The distinction is important and is not addressed explicitly in most frameworks. Trait-based approaches treat AIQ as relatively fixed and suitable for selection and ranking, in contrast to skill-based approaches that treat it as learnable, improvable, and subject to decay as technologies evolve.

A trait can be measured, ranked, and expected to be stable over time and across contexts. The classic example is IQ: it can be improved somewhat through practice and education, but the underlying capacity is largely dispositional. The hypothesis would be that while training may be of marginal value, it cannot fundamentally transform a subject's AIQ. A skill however, is learned, practiced, and time-sensitive. It can be developed deliberately, and can also decay when not maintained or become obsolete when the environment changes. Some examples of such skills are typing and operating a slide rule, both of which are now obsolete. A key differentiator between traits and skills are that the latter can be refined through education, practice, and continuous development. Crucially, a skill can also be rendered irrelevant by shifts in the surrounding technology.

Among the approaches surveyed, Qin et al. (2025) occupy the clearest trait position. Their entire empirical apparatus is designed to demonstrate that AIQ is a stable individual difference that is consistent across time, tasks, and different AI systems. Their methodology is explicitly modeled on how IQ was established, and their finding that chess players maintain stable human-AI performance over eighteen years is a direct argument for dispositional stability. The implication of this hypothesis is that training would not substantially affect AIQ, leading to the consequence that the right organizational response is selection rather than development.

Most other approaches implicitly treat AIQ as a skill without theorizing it explicitly. Forrester Research (2024) tracks scores over time and frames AIQ as something organizations must invest in and raise. Value Creation Innovation Institute (2024) describes continuous adaptation as a core pillar. Ganuthula and Balaraman (2025a) include adaptive learning capability as one of their eight dimensions. These framings lean toward skill but stop short of the logical consequence: if AIQ is a skill that can be built, it can also decay. Furthermore, in a changing environment, deploying outdated AIQ scores would risk miscalibration and underscores the need for continuous measurement, education and adaptation.

This skill-based interpretation is also consistent with evidence from the AI systems literature. Chain-of-thought prompting shows that structured instructions can substantially improve model performance on complex reasoning tasks (Wei et al., 2022), while Reflexion shows that critique, reflection, and revision can improve language-agent performance over single-pass generation (Shinn et al., 2023). More fine-grained work on prompt tone likewise suggests that linguistic framing can affect model outputs (Dobariya and Kumar, 2025). These findings do not establish AIQ as a psychometric construct, but they indicate that effective AI use depends partly on learnable interaction practices rather than fixed individual traits alone.

The distinction between trait-based and skill-based conceptualizations of AIQ is not merely analytical in character. It carries significant ethical implications that the field has not yet addressed explicitly.

A trait-based AIQ framework, in a manner analogous to IQ, risks functioning as a mechanism of selection rather than development. If AIQ is primarily understood as a stable individual difference — a capacity distributed unequally across individuals in ways that are substantially resistant to change — then the organizational logic it encourages is one of sorting rather than upskilling. Access to AI-augmented roles and responsibilities would accordingly be allocated on the basis of assessed AIQ, with low-scoring individuals potentially excluded from participation rather than supported toward improvement. Because trait-based frameworks carry implicit assumptions regarding the dispositional, and potentially innate, nature of the capacities they assess, they are susceptible to the form of ideological capture that Cave (2020) documents in the history of IQ research: the normalization of socially produced inequalities as expressions of natural difference.

A skill-based conceptualization of AIQ avoids this structural risk. If AIQ is understood as a measure of how effectively an individual navigates a technological frontier that is external, continuously shifting, and indifferent to the prior characteristics of any individual, then the credential assesses a relationship between person and frontier rather than an intrinsic property of the person. No individual possesses a priori calibration with respect to AI systems that have not previously existed. The relevant competencies are acquired through practice and experience, subject to decay in the absence of continued engagement, and require renewal as the frontier evolves. Under this design, AIQ functions as a developmental target accessible in principle to all individuals, rather than as a fixed attribute differentially distributed at the outset. The broader implications for equity are therefore not peripheral to the design of AIQ frameworks but constitute a central consideration that bears directly on the choice between trait-based and skill-based conceptualizations. Frameworks that adopt a trait-based position without engaging with these implications risk reproducing, within the domain of AI capability assessment, the hierarchical structures that have historically attended the misuse of intelligence constructs.

4.2 Object of Measurement: Knowledge, Calibration, and Performance

The second dimension concerns what is being measured. Three distinct objects can be identified:

- **Knowledge**: understanding of AI systems and their properties
- **Performance**: effectiveness in AI-assisted tasks
- **Calibration**: the ability to accurately assess when AI can and cannot be trusted

Most frameworks measure either knowledge or performance, while calibration remains implicit.

The second analytical dimension concerns the object of measurement itself. When a framework claims to assess AIQ, what precisely has a high-scoring individual demonstrated? The answer is more complex than it initially appears, because at least three distinct constructs may in principle be measured, and these constructs stand in a structured relationship to one another.

The first construct is knowledge: an individual's understanding of artificial intelligence systems, including how such systems function, where their limitations lie, and under what conditions they are prone to failure. This corresponds broadly to conventional notions of AI literacy and can be evaluated through factual assessment and conceptual reasoning. Although such knowledge constitutes a necessary foundation for effective AI use, it is not sufficient in itself. Individuals may possess substantial theoretical understanding of AI while nonetheless demonstrating poor judgment in practical interaction with these systems.

The second construct is performance: the observable quality of human–AI task execution. This includes the quality of outputs produced, the efficiency of task completion, and the avoidance of consequential errors. Performance-based approaches, such as those developed by Qin et al. (2025), possess the advantage of direct observability and quantifiability. However, successful performance on familiar or routine tasks does not necessarily indicate the presence of robust judgment. Individuals may achieve high levels of fluency through repetition and procedural familiarity while remaining vulnerable to serious failures when operating near the boundaries of AI capability.

Between knowledge and performance lies the mechanism that connects them: calibration. Calibration refers to the capacity to accurately assess where AI systems are reliable and where they are not; that is, to determine when trust in an AI system is warranted, when skepticism is necessary, and when human intervention becomes essential. Calibration is therefore fundamentally meta-cognitive in nature. It concerns not merely the effective use of AI, but the ability to evaluate how, when, and to what extent AI systems should be relied upon.

Knowledge enables Calibration. Calibration produces Performance.

This relationship has important implications for the design of AIQ frameworks. Most existing approaches measure either the inputs to AI-assisted capability or its observable outputs, while leaving the intervening mechanism implicit. Frameworks organized around competency bundles — including those proposed by Ganuthula and Balaraman (2025a), Forrester Research (2024), Value Creation Innovation Institute (2024), and aiq.works (2025) — incorporate components that implicitly require calibration, such as critical evaluation and understanding AI limitations. However, calibration remains one competency among many rather than the central organizing construct. Conversely, performance-oriented approaches — most notably Qin et al. (2025) — infer capability directly from task outcomes without establishing whether participants understand why they succeeded or whether such performance would generalize to unfamiliar contexts near the frontier of AI capability.

A notable exception is the work of Klein et al. (2020, 2023), whose entire toolkit is organized around understanding how a specific system fails and building appropriate trust accordingly. Their framework explicitly centers calibration and the development of appropriate trust in AI systems and is specifically designed to help users understand how and where a deployed system fails, thereby improving judgment in operational settings. The limitation is scope, in that their approach is designed for a single deployed system in a specific operational context. What they measure is calibration relative to one tool, rather than the broader problem of calibration relative to the continually shifting and heterogeneous frontier of AI capability more generally.

4.3 The Interaction Between the Two Dimensions

The two analytical dimensions outlined above interact in ways that have significant implications for the conceptualization and design of AIQ frameworks.

If AIQ is understood primarily as a stable trait, the precise object of measurement becomes comparatively less consequential. Under a trait-based interpretation, the primary objective is to identify a durable underlying signal of capability, and observable performance may serve as an adequate proxy for that signal. Stability across contexts and over time is assumed to be the defining characteristic of the construct.

However, different implications follow if AIQ is conceptualized primarily as a perishable skill and calibration is treated as the central mechanism underlying effective human–AI interaction. Under these assumptions, the continual evolution of AI capability introduces a structural problem: the boundary between reliable and unreliable AI performance is itself dynamic. Consequently, calibration that was accurate at one point in time may subsequently become systematically inaccurate as AI systems evolve.

An individual who possessed an accurate understanding of the limitations of AI systems in 2022 may therefore become substantially miscalibrated in relation to systems available in 2025, particularly if newer systems fail in qualitatively different ways. In such cases, domain knowledge may remain substantial and performance on familiar tasks may continue to appear competent, while the individual's underlying model of where AI systems can and cannot be trusted has become outdated. The resulting risk is not merely reduced effectiveness but misplaced confidence grounded in obsolete assumptions about system capability.

This interaction between the two dimensions has direct consequences for credential design. If AIQ is treated as a perishable skill and calibration is regarded as its central component, then temporal decay is not simply an optional feature of assessment frameworks but a logical requirement. Any credential purporting to measure calibrated human–AI capability must account for the changing nature of the technological frontier it evaluates. Under such a framework, static or non-expiring certification risks misrepresenting current capability by validating forms of calibration that may no longer correspond to contemporary AI systems.

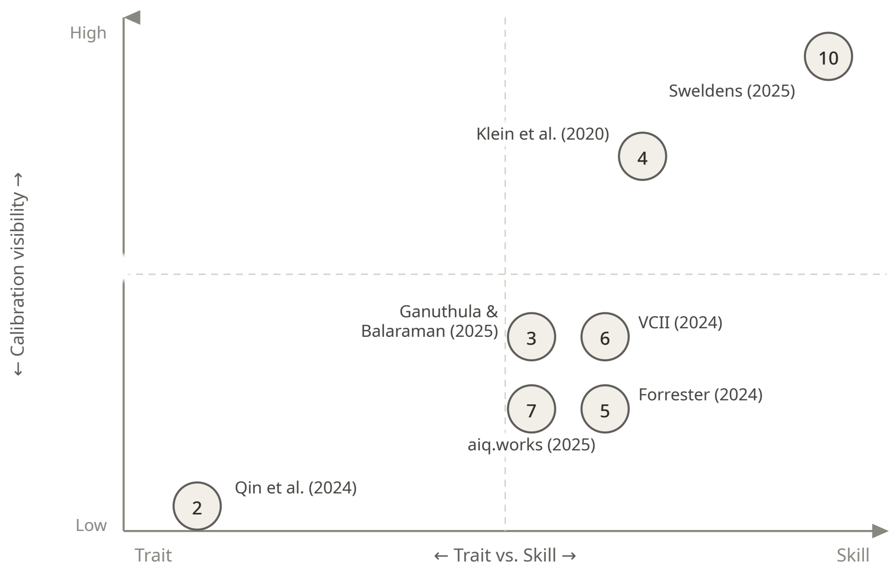

Figure 1: Two-dimensional representation of AIQ frameworks. The horizontal axis represents the extent to which AIQ is conceptualized as a stable trait (left) versus a learnable and potentially perishable skill (right). The vertical axis represents the degree to which calibration is explicitly foregrounded as the primary object of measurement, from low (bottom) to high (top). Each framework is identified by its number from Table 1. Three approaches — WSU/MDPI (1), Nebuli/aiq.org (8), and Drake/Soulcraft (9) — are excluded from the figure as they do not primarily assess human capability in relation to AI systems.

5 Mapping the Landscape

Applying the two analytical dimensions introduced in Section 4 to the frameworks surveyed in Section 3 produces a structured representation of the contemporary AIQ landscape. Table 1 positions each framework along two axes: (1) the extent to which AIQ is conceptualized as a stable trait versus a learnable and potentially perishable skill, and (2) the degree to which calibration is explicitly foregrounded as the primary object of measurement.

Notably, no framework occupies the space defined by both trait-based conceptualization and calibration-centered measurement, suggesting an unexplored area of inquiry.

The clustering also reveals rapid parallel development, indicating that the field is responding to shared external pressures rather than cumulative theoretical progress.

5.1 The concentration in the lower-right quadrant

Four approaches — Ganuthula and Balaraman (3), Forrester (5), VCII (6), and aiq.works (7) — cluster within the lower-right quadrant of the matrix. Despite substantial differences in methodology, application domain, and institutional setting, these frameworks converge on a shared conceptual position: AIQ is understood primarily as a learnable skill rather than a stable trait, and effective AI capability is assessed principally through performance and competency measures. Within these frameworks, calibration is present, but generally as one competency among several rather than as the central organizing construct. Collectively, this cluster represents the emerging mainstream consensus within the AIQ landscape. The fact that all four approaches emerged within the narrow time window of 2024–2025 further suggests that this consensus developed rapidly and largely without explicit coordination among the groups involved.

Table 1: AIQ approaches across two dimensions

#	Study	Year	Trait–Skill (0–100)	Calibration Visibility (0–100)	Quadrant
1	WSU/MDPI	2025	N/A	N/A	Outside matrix
2	Qin, Lu et al.	2024	10	5	Bottom left
3	Ganuthula & Balaraman	2025	55	35	Bottom right
4	Klein, Hoffman, Mueller	2020	70	75	Top right
5	Forrester	2024	65	30	Bottom right
6	VCII	2024	60	35	Bottom right
7	aiq.works	2025	55	30	Bottom right
8	Nebuli/aiq.org	2023	N/A	N/A	Outside matrix
9	Drake/Soulcraft	2026	N/A	N/A	Outside matrix
10	Sweldens	2025	95	95	Top right

5.2 The isolated upper-right quadrant

Only two approaches occupy the upper-right quadrant, where AIQ is conceptualized both as a skill and as a construct fundamentally centered on calibration. Klein, Hoffman, and Mueller (4), beginning in 2020, developed a framework explicitly focused on understanding how specific AI systems fail and on cultivating appropriate trust in those systems. One of the co-authors of this article, Sweldens (10), writing in 2025, similarly foregrounds calibration, but defines the object of measurement more broadly as the “jagged frontier” of AI capability — the uneven and continuously shifting boundary between reliable and unreliable AI performance. In both approaches, calibration is not treated as a secondary component of capability but as its central mechanism. The conceptual distance between this quadrant and the dominant lower-right cluster is notable. No frameworks occupy an intermediate position between these two orientations. Existing approaches either embed calibration within broader competency structures or elevate it to the primary object of assessment, with little evidence of gradual theoretical transition between these positions.

At the same time, important differences remain between the two upper-right approaches. Klein’s framework conceptualizes calibration locally: it concerns the development of appropriate trust in a specific deployed system operating within a defined context. Sweldens’ framework conceptualizes calibration more generally, as the ability to navigate the shifting frontier of AI capability across systems, domains, and time. This distinction has direct implications for credential design. Because Klein’s framework concerns calibration relative to a stable system, it does not inherently require credential expiration. By contrast, Sweldens’ framework logically entails temporal decay, since the frontier to which calibration refers is itself continuously evolving.

5.3 The absence of an upper-left quadrant

No surveyed framework conceptualizes AIQ simultaneously as a stable trait and as a calibration-centered construct. At first glance, this absence appears theoretically coherent: traits imply stability, whereas calibration appears to require continuous adaptation to a changing technological environment. However, the absence of frameworks in this quadrant should not be interpreted as evidence that such a position is conceptually impossible.

Calibration itself may possess trait-like dimensions. The capacity to accurately assess the limits of a system, maintain epistemic restraint, and resist unwarranted confidence in persuasive outputs may vary systematically across individuals in ways that are relatively stable across domains. Some individuals appear dispositionally better calibrated than others, independent of specific technical expertise, suggesting that calibration may partly reflect enduring cognitive tendencies rather than solely learned competencies. No existing AIQ framework has directly investigated this possibility. The empty upper-left quadrant therefore represents not a conceptual contradiction but an unexplored area of inquiry. More specifically, it raises a question that current frameworks have largely overlooked: whether there exists a stable human disposition toward appropriate trust in AI systems, and how such a disposition might interact with the learned skills developed through education and experience.

5.4 The significance of chronology

The chronology embedded within the matrix is itself analytically revealing. Klein's framework, originating in 2020, predates the widespread adoption of generative AI and emerged in the context of earlier, more specialized AI systems. By contrast, the dominant lower-right cluster and the Sweldens framework all emerged during 2024–2025, a period characterized by rapid diffusion of generative AI technologies. The near-simultaneous appearance of multiple AIQ frameworks within such a compressed timeframe, and with little apparent cross-awareness, suggests that the field is responding to a shared external pressure rather than developing cumulatively through established theoretical traditions.

This parallel emergence presents both an opportunity and a risk. It represents an opportunity because it demonstrates substantial intellectual and institutional attention directed toward the problem of human capability in AI-mediated environments. At the same time, it presents a risk because independently developed frameworks may address closely related problems using incompatible conceptual assumptions, measurement strategies, and validation criteria. Without greater coordination, the field may evolve into a fragmented landscape of parallel but non-interoperable approaches.

6 Integrated Matrix Analysis

The matrix presented in the previous section should not be interpreted as a ranking of frameworks. No quadrant is inherently superior, and the placement of individual approaches necessarily involves interpretive judgment that may reasonably be contested. The primary function of the matrix is instead analytical: it makes explicit a series of assumptions that many frameworks leave implicit. Three such assumptions warrant particular examination.

6.1 The limitations of the emerging mainstream consensus

The four approaches clustered in the lower-right quadrant collectively represent the emerging mainstream understanding of AIQ. Across these frameworks, AIQ is conceptualized primarily as a learnable capability that can be developed through education and practice, measured through observable performance and competency, and applied within educational and workplace settings. These assumptions are neither arbitrary nor implausible. AIQ is almost certainly at least partly learnable, performance provides a practical proxy for capability, and the demand for workforce-oriented assessment is both immediate and substantial.

Nevertheless, the convergence of these frameworks also reveals a shared limitation. They evaluate what individuals can accomplish using AI systems without systematically examining whether individuals possess accurate judgment regarding when those systems should or should not be trusted. High performance on familiar tasks does not necessarily imply robust calibration at the frontier of AI capability. Individuals may become highly fluent users of AI tools — efficient, productive, and apparently competent — while simultaneously over-trusting outputs precisely in the contexts where AI systems are most likely to fail. Calibration is tested most clearly in edge cases, unfamiliar conditions, and situations in which AI systems generate confident but incorrect outputs. Frameworks that assess aggregate performance across routine tasks are unlikely to detect this failure mode. As a result, they may assign high AIQ scores to individuals who remain substantially miscalibrated in the contexts where judgment matters most.

This limitation is not merely methodological. As AI systems become increasingly integrated into high-stakes domains — including medical diagnosis, legal reasoning, financial decision-making, and safety-critical operations — the consequences of miscalibration become correspondingly more severe. Assessment systems that cannot distinguish between genuine calibration and fluent over-trust may therefore prove inadequate for the environments in which they are intended to operate.

6.2 Temporal decay as a defining differentiator

All frameworks surveyed in this paper acknowledge, either explicitly or implicitly, that AI systems are evolving rapidly. Most approaches treat this primarily as a problem of curricular maintenance: assessment content must be periodically updated to reflect new capabilities, models, and tools. Such an approach is sufficient if

AIQ is understood primarily in terms of knowledge acquisition or observable performance, since knowledge can be refreshed and performance can be re-evaluated.

However, stronger implications follow if calibration is treated as the central mechanism underlying AIQ and if AIQ itself is conceptualized as a perishable skill rather than a stable trait. Under these assumptions, the movement of the technological frontier becomes central. Calibration that was accurate in 2022 may become systematically inaccurate by 2025, not because individuals have forgotten prior knowledge, but because the capabilities and failure modes of AI systems have changed. An individual's internal model of where AI systems succeed and fail may therefore become outdated while continuing to generate high subjective confidence. Importantly, such outdated calibration may be difficult to detect externally and often difficult for individuals themselves to recognize.

This has direct implications for credentialing. A non-expiring AIQ credential does not merely risk incompleteness; it may become actively misleading by certifying competence relative to a technological frontier that no longer exists. Temporal decay is therefore not simply a design preference but a logical consequence of two prior assumptions: first, that AIQ is a skill rather than a stable trait, and second, that calibration rather than performance constitutes the central object of measurement. Frameworks that recognize the rapid evolution of AI systems while issuing effectively permanent credentials have not fully followed these assumptions to their logical conclusion.

6.3 The under-recognition of Klein's framework

Among the frameworks surveyed, the work of Klein, Hoffman, and Mueller occupies a distinctive and relatively isolated position. Beginning in 2020, and developed within high-stakes operational contexts such as military systems, industrial AI, and medical decision support, their framework represents the only pre-generative-AI approach that explicitly centers calibration as the primary object of intervention. The framework is organized entirely around the problem of appropriate trust: not simply whether users can operate a system effectively, but whether they understand when that system should and should not be trusted.

Despite the relevance of this work, it appears to have had limited influence on more recent AIQ literature. Academic publications emerging during 2024 and 2025 rarely reference it, and commercial assessment platforms similarly do not appear to engage with it substantively. This absence represents a significant missed opportunity. Klein and colleagues devoted considerable effort to developing practical tools for addressing calibration in precisely the kinds of environments where the consequences of miscalibration are immediate and measurable. Contemporary AIQ research would likely benefit from engaging more directly with this earlier body of work rather than independently rediscovering many of the same conceptual problems.

6.4 Parallel emergence as a coordination problem

As noted in Section 1, the independent emergence of multiple AIQ frameworks within a short period strongly suggests that a genuine conceptual gap is being recognized across domains. At the same time, this parallel emergence also constitutes a coordination problem. Multiple research groups and practitioners are addressing closely related questions with limited awareness of one another, producing frameworks that are difficult to compare systematically, validate against one another, or integrate cumulatively. The resulting landscape consists of parallel efforts that share a broad intuition while lacking shared terminology, methodological standards, and common empirical foundations.

Such fragmentation is not unusual during the early stages of a developing field. The historical trajectories of both IQ and EQ research exhibit similar periods of parallel development prior to eventual consolidation around more standardized constructs and measurement frameworks. However, such consolidation requires coordination, and coordination requires at minimum that researchers and practitioners working on related problems are aware of one another's existence. The matrix developed in this paper represents a preliminary attempt to increase that visibility. More substantial progress will likely require sustained collective efforts, including shared terminology, cross-framework validation studies, and potentially institutional structures capable of facilitating systematic interdisciplinary dialogue.

7 Open Questions

The matrix developed in this paper provides a structured account of the current AIQ landscape. Equally importantly, however, it reveals several areas in which the field remains conceptually and empirically underdeveloped, of which four open questions stand out as particularly consequential.

The trait vs. skill question remains empirically unresolved

Qin and colleagues have produced some of the strongest evidence to date suggesting that AIQ functions as a stable individual difference. However, their empirical studies were conducted within relatively narrow task domains and using AI systems that largely predate the widespread adoption of generative AI models. It therefore remains unclear whether the same findings generalize to large language models and other contemporary generative systems. The implications of this uncertainty are substantial. If AIQ proves to function primarily as a stable trait, then many current assumptions underlying AIQ education, training, and credentialing initiatives may require reconsideration. Conversely, if AIQ is primarily a skill that develops and decays through interaction with evolving systems, then educational and certification approaches become significantly more defensible. Resolving this question will require longitudinal research specifically designed to examine the relative stability and temporal decay of AIQ within rapidly changing generative AI environments.

The unexplored quadrant deserves investigation

The absence of frameworks occupying the upper-left quadrant of the matrix — that is, approaches treating AIQ as both trait-like and calibration-centered — identifies an important and largely unexplored line of inquiry. Existing frameworks implicitly assume that calibration is primarily adaptive and therefore skill-based. However, this assumption has not been systematically tested. It is plausible that individuals differ dispositionally in their tendency toward appropriate epistemic caution in relation to AI systems. Some individuals appear naturally more resistant to persuasive but potentially unreliable outputs, while others appear more inclined toward uncritical acceptance. Such tendencies may reflect relatively stable individual differences in calibration-related cognition, independent of specific technical expertise or training. Whether these tendencies are genuinely trait-like, and whether they predict AIQ outcomes independently of other variables, remain open empirical questions. Investigation of this possibility may provide a productive bridge between trait-oriented and skill-oriented conceptions of AIQ.

The problem of general versus domain-specific AIQ

A further unresolved issue concerns whether AIQ should be understood as a single general capability or as a collection of domain-specific competencies. Most current frameworks implicitly assume that AIQ generalizes across tools, tasks, and domains. However, if the frontier of AI capability is fundamentally jagged — highly reliable in some contexts while unpredictably unreliable in others — then calibration itself may be domain-dependent. An individual who is well calibrated regarding AI-assisted legal reasoning may nonetheless be poorly calibrated in medical diagnosis, scientific interpretation, or financial forecasting. If this is the case, then AIQ may not constitute a unitary construct but rather a family of partially independent competencies. This distinction has important implications for assessment design. If AIQ is domain-specific, then generalized single-score credentials may obscure meaningful variation in capability across contexts and produce misleading claims of competence.

The validation gap

Across the AIQ landscape, perhaps the most fundamental empirical question remains insufficiently addressed: whether higher AIQ scores actually predict improved outcomes in real-world AI-assisted work. At present, many frameworks assume rather than demonstrate that their measures correspond to meaningful differences in practical capability. Without evidence establishing predictive validity, AIQ assessments remain grounded primarily in theoretical assumptions rather than demonstrated utility. The development of robust validation studies linking AIQ measures to consequential real-world outcomes is therefore arguably the single most important next step for the field.

Recent work on evaluation at the edge of human comprehension points toward one possible empirical route. Marro et al. (2026) propose a benchmarking framework in which humans act as bounded verifiers, evaluating localized claims when full comprehension may be infeasible. Although this does not validate AIQ scores directly, it suggests that calibration-like behavior – knowing what to check and how to preserve judgment under partial comprehension – can be operationalized in measurable human–AI evaluation workflows.

Equity, access, and the potential to reduce the AI divide

A further question concerns the relationship between AIQ frameworks and the unequal distribution of AI capability across populations. If AIQ is partly a function of the quality and sophistication of AI systems to which individuals have had access, then assessment scores may inadvertently reflect structural inequalities in access to frontier technology rather than differences in underlying competence or calibration ability. AI tools of substantially different capability are currently distributed unequally across income levels, geographies, organizational sizes, and educational contexts. An AIQ framework designed without explicit attention to this structural condition risks replicating the legitimating function that IQ has historically performed: certifying as individual capacity what is in part a product of differential opportunity.

This concern, however, also points toward a constructive possibility that the field has not yet fully articulated. Because AIQ, properly conceptualized, is a perishable skill measured against a moving frontier rather than a fixed property of individuals, it is in principle accessible to anyone with access to current AI systems. A well-designed AIQ framework could therefore function not merely as a credentialing mechanism but as an instrument for broadening participation in AI-mediated work — identifying where access gaps exist, directing educational and developmental investment toward underserved populations, and contributing to a reduction in the digital and AI divide rather than its entrenchment. Whether AIQ frameworks are designed in ways that realize this potential or inadvertently foreclose it represents one of the most consequential design choices the field faces, and one that deserves sustained attention from researchers, practitioners, and policymakers alike.

8 Conclusion: A Timely Concept

The most significant finding of this survey does not concern any individual framework in isolation, but rather the collective pattern that emerges across them. Researchers and practitioners working in cognitive psychology, computer science, education, enterprise consulting, and marketing have independently converged upon the same conceptual term within a remarkably compressed period of time and with little apparent awareness of one another's work. This parallel emergence should not be dismissed as coincidence or conceptual noise. Rather, it constitutes evidence that a genuine conceptual gap is being recognized across multiple domains simultaneously. Existing measures of human capability appear increasingly insufficient for environments in which individuals interact continuously with highly capable AI systems, and the inadequacy of those measures has become salient across disciplines as methodologically distinct as psychometrics and commercial strategy.

The two analytical dimensions introduced in this paper — whether AIQ is conceptualized as a stable trait or as a perishable skill, and the extent to which calibration is treated as the central object of measurement — provide an initial framework for comparing approaches that have thus far developed largely in isolation. Applying these dimensions reveals both a developing consensus and a significant conceptual limitation. The emerging consensus is that AIQ is best understood as a learnable capability that can be cultivated through education, training, and assessment. The principal limitation concerns calibration: the mechanism through which knowledge is translated into reliable performance and the component most vulnerable to failure in precisely those situations where judgment is most consequential.

The analysis presented in this paper also points toward a design imperative that extends beyond methodological preference. Cave (2020) has documented the mechanisms through which quotient-based constructs have historically been appropriated to legitimate dominance hierarchies, identifying the framing of such constructs as fixed, intrinsic human properties as the structural condition that enables this misappropriation. AIQ frameworks that adopt a trait-based conceptualization inherit this structural vulnerability. The field would therefore benefit from an explicit and collective commitment to designing AIQ as a perishable skill assessed against a moving technological frontier, rather than as a stable property of individuals. This commitment is not merely an ethical preference. It follows from the nature of the construct itself: the frontier evolves, calibration decays in the absence of continued engagement, and no individual possesses prior familiarity with AI systems that have not yet existed. Frameworks that proceed on trait-based assumptions risk not only methodological inadequacy but the reproduction of the hierarchical conditions that have historically attended the quantification of human cognitive capacity.

The field currently occupies an early but strategically important stage of development. The parallel emergence that characterizes the present moment represents a substantial opportunity. Intellectual attention, institutional interest, and methodological diversity are all clearly present. However, opportunities for coordination are unlikely to remain indefinitely available. As frameworks become institutionalized, organizations invest in particular assessment models, and commercial incentives consolidate around specific methodologies, the difficulty of establishing interoperability and conceptual convergence will increase. The historical development of both IQ and EQ research demonstrates that early fragmentation can persist for extended periods once competing paradigms become established.

The urgency of this coordination has been underscored by developments beyond the academy. The papal encyclical *Magnifica Humanitas* (Leo XIV, 2026), released on the 135th anniversary of *Rerum Novarum*, draws an explicit parallel between the industrial and AI revolutions and confirms that human-AI capability is a matter of moral and political concern at the highest levels. The field of AIQ research has converged from many directions on the same problem independently, confirming that the conversation this paper calls for is timely.

At present, the field arguably requires not additional frameworks but greater coordination among existing efforts. Researchers, practitioners, and institutions working on AIQ would benefit from systematic comparison of assumptions, cross-validation of instruments, and the development of shared terminology capable of supporting cumulative progress. Whether such coordination emerges through collaborative working groups, dedicated research consortia, or more formal institutional structures is ultimately secondary to the broader requirement that sustained interdisciplinary dialogue occur at all. The present paper is intended as a contribution toward initiating that process.

Author Disclosure

AI tools (Claude and ChatGPT) were used to conduct web searches, synthesis of search results, collaborative drafting of all sections, identification of structural issues during read-through, and critical questioning of analytical claims during development. WS has independently developed an AIQ framework included as one of the ten approaches surveyed in this paper. WS, BS and SNG contributed the core analytical insight (the two dimensions and the calibration argument), the selection of sources beyond what search returned, and refinement of arguments from draft to final manuscript. We acknowledge Jianlin Yu, University of Oxford for critical comments and helpful discussions.

References

aiq.works (2025). AIQ assessment. https://aiq.works. Accessed 29 May 2026.

Binet, A. and Simon, T. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, 11:191–244. https://doi.org/10.3406/psy.1904.3675.

Cave, S. (2020). The problem with intelligence: Its value-laden history and the future of AI. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 29–35. https://doi.org/10.1145/3375627.3375813.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. Working Paper 24-013, Harvard Business School. https://www.hbs.edu/faculty/Pages/item.aspx?num=64700.

Dobariya, O. and Kumar, A. (2025). Mind your tone: Investigating how prompt politeness affects LLM accuracy. arXiv preprint. https://arxiv.org/abs/2510.04950.

Drake, E. (2026). What's your AIQ? LinkedIn Pulse. https://www.linkedin.com/pulse/whats-your-aiq-evan-drake-jeztc/.

Forrester Research (2024). The forrester artificial intelligence quotient (AIQ) assessment. Forrester report RES180710. https://www.forrester.com/report/the-forrester-artificial-intelligence-quotient-aiq-assessment/RES180710.

Forrester Research (2026). AIQ 2.0: Employees (still) aren't ready to succeed with workforce AI. Forrester report.

Ganuthula, V. R. R. and Balaraman, K. K. (2025a). Artificial intelligence quotient (AIQ): A novel framework for measuring human-AI collaborative intelligence. *Discover Artificial Intelligence*, 5:516. https://doi.org/10.1007/s44163-025-00516-1.

Ganuthula, V. R. R. and Balaraman, K. K. (2025b). Artificial intelligence quotient (AIQ): A novel framework for measuring human-AI collaborative intelligence. arXiv preprint. https://arxiv.org/abs/2503.16438.

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

Klein, G. (2020). AIQ: Artificial intelligence quotient. Psychology Today. https://www.psychologytoday.com/us/blog/seeing-what-others-dont/202007/aiq-artificial-intelligence-quotient.

Klein, G., Borders, J., Hoffman, R., and Mueller, S. (2023). Helping people get smart about the smart machines they are using. About Work. https://medium.com/about-work/helping-people-get-smart-about-smart-machines-they-are-using-f9e0095846fe.

Leo XIV (2026). Magnifica humanitas: Encyclical letter on safeguarding the human person in the time of artificial intelligence. Holy See. https://www.vaticannews.va/en/pope/news/2026-05/pope-leo-xiv-first-encyclical-magnifica-humanitas.html.

Marro, S., Yu, J., La Malfa, E., Deb, O., Li, J., Yang, Y., Abraham, E., Sengupta, S., Sommerlade, E., Wooldridge, M., and Torr, P. (2026). Benchmarking at the edge of comprehension. arXiv preprint. https://arxiv.org/abs/2602.14307.

Nebuli (2023). Augmented intelligence quotient. https://aiq.org. Accessed 29 May 2026.

Pereyda, C. and Holder, L. (2025). The artificial intelligence quotient (AIQ): Measuring machine intelligence based on multi-domain complexity and similarity. *AI*, 6(12):313. https://doi.org/10.3390/ai6120313.

Qin, X., Lu, J. G., Chen, C., Zhou, X., Gan, Y., Li, W., and Song, L. (2025). Artificial intelligence quotient (AIQ). OSF Preprints. https://doi.org/10.31234/osf.io/qjm3r.

Shinn, N., Cassano, F., Berman, E., Gopinath, A., Narasimhan, K., and Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 36. https://arxiv.org/abs/2303.11366.

Sweldens, W. (2025). AIQ: Artificial intelligence quotient. Substack newsletter. https://wimsweldens.substack.com.

Value Creation Innovation Institute (2024). What is your AI quotient score? https://www.vciinstitute.com/blog/what-is-your-aiq-score-and-why-does-it-matter.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pages 24824–24837. https://arxiv.org/abs/2201.11903.