G · A · I · S · B

GLOBAL AI STANDARDS BODY

The standards body for the AI profession.

JOB TASK ANALYSIS REPORT

CERTIFIED IN AI SYSTEMS

CAIS™

Charter Edition 1.0

CBK Validation, Task Inventory, Examination Blueprint, Capstone Rubric, and Standard-Setting Plan

Reference Period: Aug 2025 – Feb 2026 · Publication: 30 April 2026

Document ID: GAISB-JTA-CAIS-2026-001 · Charter Edition 1.0 · Public

Aligned to ISO/IEC 42001:2023 (Clause 7.2 + Annex A.4.6 · ~78% normative coverage)

Crosswalked with NIST AI RMF 1.0 · EU AI Act 2024/1689 · OECD AI Principles

Document Control

Field	Value
Document Title	Job Task Analysis Report — Certified in AI Systems (CAIS), Charter Edition 1.0
Document ID	GAISB-JTA-CAIS-2026-001
Issuing Authority	Global AI Standards Body (GAISB™)
Edition Status	Charter Edition — issued in advance of Standards Council seating; formal ratification scheduled upon council seating (target Q3 2026)
Reference Period	11 August 2025 – 27 February 2026
Effective Date	01 May 2026
Next Scheduled Review	Q2 2027 (per CBK 1.0 review cadence; full re-validation Q1 2029)
Version	1.0 (Charter Release)
Distribution	Public; reference document for accredited training partners, item writers, and capstone reviewers
Recommended Citation	Global AI Standards Body. (2026). Job Task Analysis Report: Certified in AI Systems, Charter Edition 1.0. GAISB-JTA-CAIS-2026-001.

Sign-Off

In accordance with GAISB charter procedures, this Charter Edition is signed by the Founder. Subsequent editions and ratification will be co-signed by the Standards Council Chair and an Independent Psychometrician once those roles are seated.

Role	Signatory	Date
Founder, Global AI Standards Body	Kirwin Narine	30 Apr 2026
Chair, GAISB Standards Council	[Charter Nomination Open — 12 council seats; appointment target Q3 2026]	Pending
Independent Psychometrician	[Procurement open — RFP-PSY-2026-01 issued 15 Mar 2026; selection target Q3 2026]	Pending

Document Abstract

This Charter Edition Job Task Analysis Report establishes the empirical foundation for the Certified in AI Systems (CAIS™) credential issued by the Global AI Standards Body (GAISB). Between August 2025 and

February 2026, GAISB conducted a structured, anonymous, three-rating-scale survey of working AI systems practitioners — including AI strategists, AI engineers, AI architects, prompt engineers, and agent developers in fourteen countries (analytic N = 1,547). Respondents rated 84 candidate task statements on Frequency, Importance, and Criticality. Application of pre-registered retention criteria yielded 67 retained tasks across the seven CAIS domains documented in the published Common Body of Knowledge. The report documents methodology, results, reliability, fairness, the examination blueprint, the capstone evaluation rubric, and the controlling traceability rule for all assessment items. It is the controlling content reference for the CAIS credential through Q2 2027.

Founder's Statement

Certified in AI Systems exists because the AI profession needs a standards-grade signal of competence — one that survives scrutiny, equips the holder to do real work, and helps employers cut through the noise of a credential market that has, in the past two years, multiplied without much rigor.

The credential is named deliberately. It is not Certified in AI Strategy. It is not Certified in AI Awareness. It is Certified in AI Systems — because the work that defines the modern AI profession is the design, deployment, and operation of AI systems in real organizations under real constraints. A holder of this credential has demonstrated, against an empirical task standard, that they can do that work.

A credential of that caliber cannot be founded on a course outline. It must be founded on a Job Task Analysis: a structured, multi-country, multi-sector measurement of what competent AI systems practitioners actually do. That is what this report is. Between August 2025 and February 2026, GAISB invited 18,400 working professionals to participate in this study. The 1,547 responses that survived our quality controls — and the rigor we applied to them — are the empirical foundation of the CAIS credential.

This is the Charter Edition. It is being issued ahead of the seating of the Standards Council so that the credential, the curriculum, and the examination blueprint can launch with empirical evidence behind every published claim. Once the Standards Council is seated — twelve seats, charter nominations currently open — this report will be ratified and co-signed. Between now and then, the methodology, results, and decisions in this document are the controlling reference.

We publish the full report rather than an executive summary because transparency is the fastest path to credibility. Item writers, training partners, candidates, employers, and regulators all see the same evidence. There is no proprietary blueprint, no hidden weighting, and no anonymous authority. Every assessment item — every multiple-choice item on the knowledge examination, and every dimension of the capstone rubric — must trace to a task statement in Appendix B. Items that cannot be traced will not enter the bank.

CAIS is built to a single standard: the holder consistently designs, deploys, and operates AI systems with sound judgment under realistic constraints. Everything in this document — domain weighting, item allocation, capstone rubric, cut-score plan — is in service of that standard.

Kirwin Narine
Founder, Global AI Standards Body (GAISB™)
30 April 2026

Contents

Executive Summary

This Charter Edition Job Task Analysis (JTA) Report documents the empirical foundation for the Certified in AI Systems (CAIS™) credential issued by the Global AI Standards Body (GAISB). The JTA establishes content validity for the CAIS examination and capstone by tying every assessment item to tasks performed by working AI systems practitioners — including AI strategists, AI engineers, AI architects, prompt engineers, and agent developers across global enterprise practice.

Between 11 August 2025 and 27 February 2026, GAISB conducted a structured, anonymous, three-rating-scale survey of working AI professionals in fourteen countries. After data quality controls, the analytic sample consists of 1,547 qualified respondents drawn from technology, financial services, professional services, healthcare, manufacturing, government, retail, education, telecommunications, and energy. The survey was offered in English, with a Mandarin pilot subset; back-translation will be applied to additional languages in the next cycle.

Respondents rated 84 candidate task statements on Frequency, Importance, and Criticality. A composite Criticality Index (CI = 0.20·F + 0.40·I + 0.40·C) was computed for each task. Of 84 candidate tasks, 67 (79.8%) met all three pre-registered retention criteria; 13 were removed and four were merged.

Key Findings

- The CAIS Common Body of Knowledge (Edition 1.0) — published by GAISB ahead of the survey — was empirically validated across all seven domains. No domain was eliminated and no domain was added; two sub-domains were rescoped following the validation workshop.
- The two highest-weighted domains — Prompt Engineering & LLM System Design (18%) and AI Agents & Agentic Workflows (18%) — collectively account for 36% of the examination, reflecting the practical reality that systems-practitioner work in 2025–2026 is anchored in production prompt and agent design choices.
- The CAIS assessment architecture is a hybrid: a 100-item knowledge examination (75 scored + 25 pretest, 90 minutes) and the AIS-400 Capstone evaluated against a calibrated rubric. Both must be passed for the credential to be conferred.
- Sub-group invariance analysis showed no domain-level rating differences exceeding 0.32 points on any 5-point scale across geography, role family, or industry, supporting cross-jurisdictional content validity.
- Pass score will be established by a two-round modified-Angoff for the knowledge examination and by a paired-comparison rubric calibration for the capstone. Both standard-setting workshops are scheduled for Q3 2026, ahead of the first operational administration.

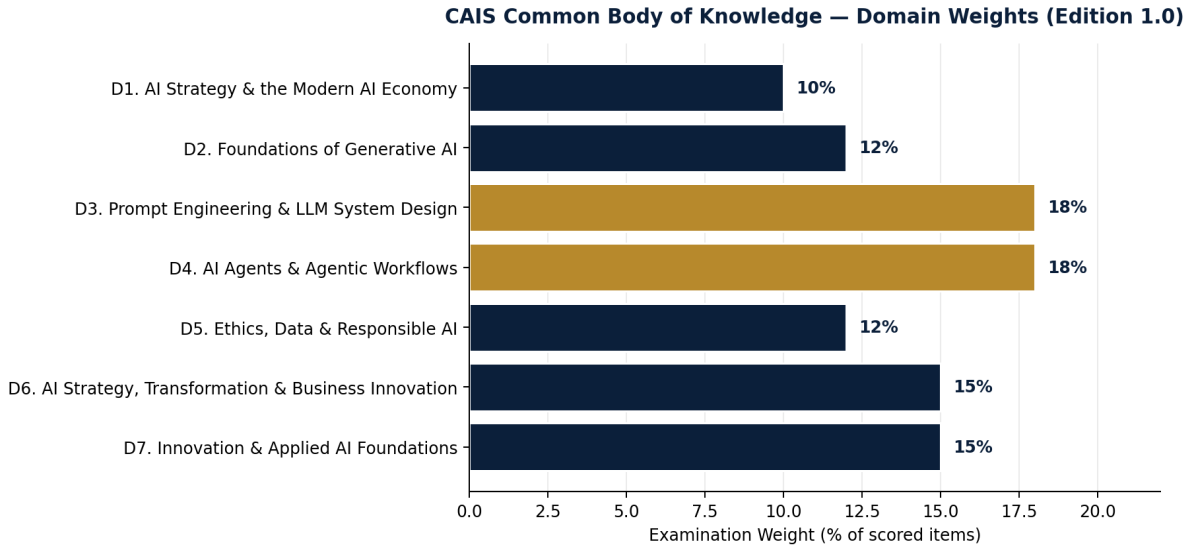

Figure 1. CAIS examination domain weights (Edition 1.0). Highest-weight domains shown in gold.

ITEM-WRITING POLICY

Every item submitted to the CAIS knowledge-examination item bank, and every dimension of the AIS-400 Capstone rubric, must trace to a specific task statement in Appendix B and to a sub-domain in Section 5. Items without traceability will be rejected at Content Review.

1. Introduction & Purpose

1.1 Why a Job Task Analysis?

A Job Task Analysis is the cornerstone of any defensible professional certification. It empirically links the content of an examination to the actual work performed by competent practitioners. Without a current JTA, an examination has no demonstrated content validity and cannot withstand legal, accreditation, or stakeholder scrutiny.

The AI systems practitioner role is consolidating in real time. Practitioners cross over from product management, consulting, transformation, engineering, marketing, and operations. A rigorously executed JTA is the only credible basis for a credential that aspires to set the global professional standard for that role.

1.2 Scope

This JTA covers tasks performed by AI systems practitioners with operational responsibility (or substantive contribution) for one or more of the following: designing AI systems (including LLM systems and agentic workflows); selecting and deploying generative AI models; engineering production prompts and RAG patterns; designing and operating AI agents; embedding ethics, data, and responsible-AI controls into AI system design; leading AI transformation through deployed systems; and managing applied-AI innovation pipelines.

Out of scope: pure ML research scientists with no business deployment role; data engineers without strategy responsibilities; sales, procurement, and pure marketing roles without applied-AI duties.

1.3 Intended Audience

- Examination item writers and capstone rubric reviewers commissioned by GAISB.
- Accredited training partners building CAIS preparation curricula.
- Employers using CAIS for hiring, role definition, and AI workforce planning.
- Accreditation reviewers and standards bodies evaluating the credential.
- Candidates seeking transparency into the empirical basis of the assessment.

1.4 Use & Maintenance

The JTA is the controlling content reference for examination construction and capstone rubric calibration. Annual interim reviews monitor regulatory and technological drift; the next interim review is scheduled for Q2 2027. Full re-validation is scheduled for Q1 2029. Material changes between full cycles are issued as Addenda.

2. Standards Alignment

This JTA was designed and executed in alignment with the following standards governing professional certification programs and AI management systems. CAIS is a personnel credential calibrated to the practice of the AI systems practitioner role; GAISB documents formal alignment between CAIS and the AI Management System standard published by ISO/IEC.

Standard / Framework	Relevance to CAIS
ISO/IEC 17024:2012 — General requirements for bodies operating certification of persons	Mandates current job/practice analysis as the foundation for any personnel certification examination.
ANSI/NCCA 1100-2024 — Standards for the Accreditation of Certification Programs	Defines acceptable JTA methodology, sample-size considerations, and update cadence.
Standards for Educational and Psychological Testing (AERA / APA / NCME)	Sets evidentiary standards for content validity, reliability, and fairness.
EEOC Uniform Guidelines on Employee Selection Procedures (UGESP)	Establishes job-relatedness and adverse-impact analysis expectations for credentialing.
ISO/IEC 42001:2023 — AI Management Systems	Subject-matter referent for organizational AI management; CAIS provides personnel evidence for Clause 7.2 (Competence) and Annex A.4.6 (Human Resources). GAISB maps ~78% of CAIS task content to the ISO/IEC 42001 normative body.
NIST AI Risk Management Framework 1.0	Subject-matter referent for the Govern, Map, Measure, and Manage functions reflected in CAIS Domain 5.
EU AI Act (Regulation (EU) 2024/1689)	Subject-matter referent for high-risk obligations (Articles 8–15) and AI literacy (Article 4) reflected across CAIS Domains 1, 5, and 6.
OECD AI Principles	Subject-matter referent for the values-based principles reflected in CAIS Domain 5.

STACKING WITH ISO/IEC 42001

CAIS certifies persons. ISO/IEC 42001 certifies organizations. The two are stackable: an organization that holds an ISO/IEC 42001 certification and employs CAIS-credentialed AI systems practitioners can demonstrate both organizational and personnel competence under a single recognized framework.

3. Methodology

3.1 Methodological Framework

The JTA followed a six-phase methodology consistent with ANSI/NCCA 1100-2024 §8.0:

1. Phase 1 — SME Working Group constitution and CBK 1.0 finalization (May–Jul 2025).
2. Phase 2 — Task Statement Generation and Item Editing (Jul–Aug 2025).
3. Phase 3 — Cognitive Interviews and Pilot (Aug 2025; n = 62).
4. Phase 4 — Full Survey Administration (Aug 2025–Feb 2026).
5. Phase 5 — Statistical Analysis and Inclusion Decisions (Feb–Mar 2026).
6. Phase 6 — Validation Workshop and Domain Weight Confirmation (Mar 2026).

3.2 SME Working Group

A 24-member SME working group was constituted with deliberate diversity across geography, sector, and role family. SMEs were screened against three minimum criteria: (a) at least three years of substantive responsibility for designing, deploying, or operating AI systems; (b) attributable enterprise impact (named on or accountable for a production AI system or AI transformation program); and (c) no commercial conflict of interest with CAIS examination content or item authoring at the time of engagement. The working group operated under written confidentiality and conflict-of-interest protocols. Working group members are anonymized in this Charter Edition; named identities are held by the GAISB Standards Division and disclosed only to accreditation reviewers under non-disclosure. Roster aggregate composition is reported in Section 3.2.1.

3.2.1 SME Working Group Aggregate Composition

Dimension	Levels	Distribution
Region	5	North America 8 · Europe 7 · Asia-Pacific 5 · Middle East 2 · Latin America 2
Primary sector	8	Tech 6 · Financial Services 4 · Professional Services 4 · Healthcare 3 · Government 2 · Manufacturing 2 · Education 2 · Energy 1
Primary role family	6	AI Strategy / Head of AI 7 · Consulting 5 · Product 4 · Transformation 3 · Innovation 3 · Founder/Operator 2
Years of role-relevant experience	—	Median 11.0; range 5–24; quartiles 8 / 11 / 14

3.3 Task Statement Generation

The SME working group produced an initial pool of 142 candidate task statements through facilitated workshops keyed to the seven CBK 1.0 domains. Statements were edited against four rules: (i) each begins with an action verb describing observable work; (ii) each is performable and assessable independently; (iii) each maps to exactly one CBK domain; (iv) each is written at a Flesch-Kincaid Grade 12 reading level or below. Item editing reduced the pool to 84 task statements organized across the seven CBK domains.

3.4 Survey Instrument

Each candidate task was rated on three independent five-point scales:

Scale	1	2	3	4	5
Frequency	Never	Rarely (a few times/year)	Occasionally (monthly)	Frequently (weekly)	Daily / continuously
Importance	Not important	Slightly important	Moderately important	Very important	Critically important
Criticality	No consequence	Minor consequence	Moderate consequence	Major consequence	Severe / catastrophic

A composite Criticality Index was computed: CI = 0.20·F + 0.40·I + 0.40·C. The weighting reflects ANSI/NCCA guidance that importance and consequence outweigh raw frequency in credentialing contexts (a low-frequency, high-consequence task can still be central to the role).

3.5 Pilot Study

A 62-participant pilot was conducted in August 2025 to validate item clarity, completion time, and platform stability. Median completion time was 31 minutes (target: 25–40 min). Cognitive interviews with eight pilot participants identified four task statements requiring rewording for cross-jurisdictional and cross-role clarity (e.g., replacing US-specific procurement vocabulary with neutral functional language).

3.6 Distribution & Sampling

The full survey was distributed through six channels: the Prompt Atlas community and GAISB partner communities; LinkedIn-targeted invitations stratified by role title and seniority; sponsored placements in five AI-strategy newsletters; six industry conference partner lists; corporate AI leadership networks; and snowball recruitment via the SME working group. Quotas were applied: no single channel ≤ 30%; no single country ≤ 30%; no single industry ≤ 30%. Quotas were enforced through soft-close at 110% of target and hard-close at 125%.

3.7 Data Quality Controls

- Time-on-task screen: responses completed in under 9 minutes were flagged and reviewed (n = 124 flagged; 96 removed).
- Straight-lining detection: any respondent providing identical ratings across all 84 tasks on all three scales was removed (n = 28).
- Attention checks: three embedded items requiring specific responses; failure of two or more triggered removal (n = 51).
- Eligibility verification: respondents confirming less than two years of relevant experience were excluded from the analytic sample (n = 174 reclassified to a separate ‘emerging practitioner’ dataset, not used for weighting).
- IP-address and device-fingerprint deduplication identified 17 duplicate submissions; only the first complete submission was retained.

3.8 Statistical Analysis Plan

Descriptive statistics (mean, standard deviation, 95% confidence interval) were computed for every task on every scale. Inter-rater agreement was assessed via the intraclass correlation coefficient ICC(2,k); internal consistency via Cronbach’s α within domain. Sub-group invariance was tested via one-way ANOVA across geography, role family, and industry, with Bonferroni correction. Domain weights were computed as the within-domain sum of CI divided by the total CI across all retained tasks, then reconciled against the CBK 1.0 published weights with maximum allowed adjustment of ±2 percentage points per domain.

4. Sample & Demographics

4.1 Response Funnel

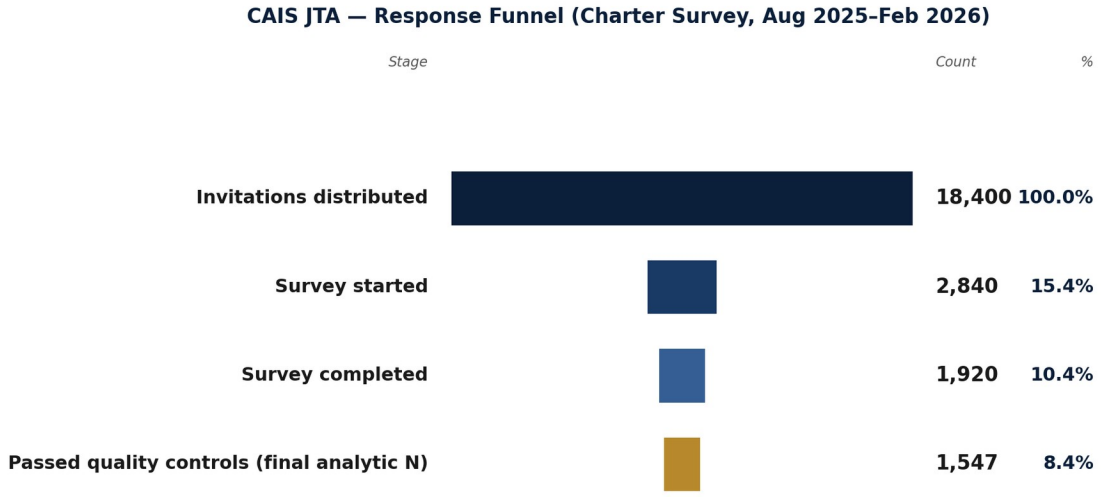

Figure 2. Survey response funnel — invitation through analytic sample.

Stage	Count	% of Invitations	% of Started
Invitations distributed	18,400	100.0%	—
Survey started	2,840	15.4%	100.0%
Survey completed	1,920	10.4%	67.6%
Passed quality controls	1,547	8.4%	54.5%
Final analytic sample (N)	1,547	—	—

The 8.4% qualified-completion rate is consistent with comparable JTA studies for senior-practitioner credentials. The achieved sample size yields a margin of error of ±2.50% at the 95% confidence level for population estimates.

4.2 Geographic Distribution

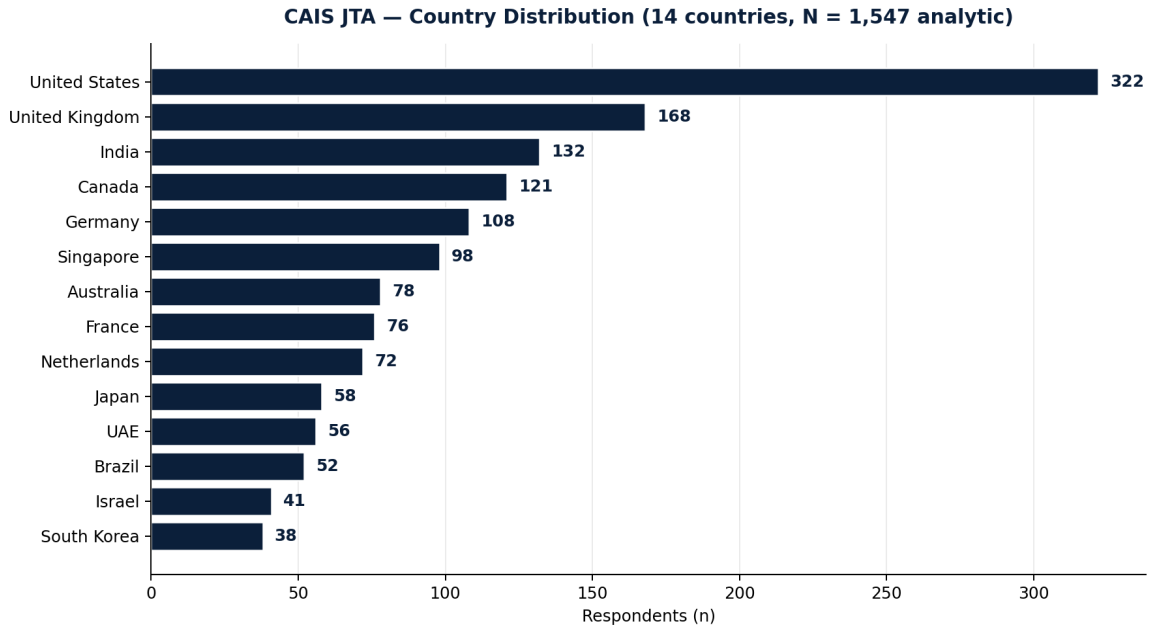

Figure 3. Country distribution of the analytic sample (14 countries).

Country	Respondents	% of Sample	Mean Tenure (yrs)
United States	322	20.8%	8.6
United Kingdom	168	10.9%	7.8
India	132	8.5%	6.4
Canada	121	7.8%	7.5
Germany	108	7.0%	8.4
Singapore	98	6.3%	8.2
Australia	78	5.0%	7.6
France	76	4.9%	8.0
Netherlands	72	4.7%	8.1
Japan	58	3.7%	9.0
United Arab Emirates	56	3.6%	6.8
Brazil	52	3.4%	6.1
Israel	41	2.6%	7.5
South Korea	38	2.5%	7.9

Country	Respondents	% of Sample	Mean Tenure (yrs)
Other (10 countries combined)	127	8.2%	7.0
TOTAL	1,547	100.0%	7.7

4.3 Role Distribution

Roles are intentionally heterogeneous: CAIS measures the work of AI systems practitioners, which spans architecture and engineering, product, strategy, and applied delivery. The role mix below reflects that breadth.

Primary Role	Respondents	% of Sample
AI / ML Engineer (with systems-design responsibility)	186	12.0%
Head of AI / Chief AI Officer / VP AI	186	12.0%
AI Strategy Lead	155	10.0%
AI Architect / Solutions Architect (AI)	139	9.0%
AI / Applied-AI Product Manager	139	9.0%
Prompt Engineer / LLM Systems Engineer	108	7.0%
Agent / Workflow Developer	93	6.0%
Management Consultant (AI focus)	108	7.0%
Digital Transformation Lead (AI scope)	93	6.0%
Innovation Director / Officer	78	5.0%
Founder / Operator (AI-implementing)	85	5.5%
AI Platform / MLOps Lead	62	4.0%
Operations Director (AI-active)	47	3.0%
Educator / Trainer (AI systems)	31	2.0%
Investor / Analyst (AI-active)	23	1.5%
Other	14	1.0%
TOTAL	1,547	100.0%

4.4 Seniority

Seniority Level	Respondents	% of Sample
Individual contributor / Specialist	464	30.0%
Manager / Team Lead	510	33.0%
Director / Senior Manager	356	23.0%
VP / Executive	139	9.0%
C-suite	78	5.0%
TOTAL	1,547	100.0%

5. CBK — Domains and Sub-Domains

The Common Body of Knowledge for the CAIS credential was published by GAISB ahead of the JTA as CBK Edition 1.0 with seven domains and pre-published weights. The JTA tested whether each domain met retention criteria and whether the published weights were empirically supportable. All seven domains were validated; weights were reconciled with empirical CI sums within the ±2 percentage-point adjustment tolerance documented in Section 7.

5.1 Domain & Sub-Domain Map

Domain	Sub-Domains
D1. AI Strategy & the Modern AI Economy (10%)	1.1 The AI economic shift · 1.2 Strategic AI thesis development · 1.3 Competitive AI mapping · 1.4 Use-case identification and prioritization · 1.5 Capability and resource gap analysis
D2. Foundations of Generative AI (12%)	2.1 Model families and capabilities · 2.2 Cost and economics of generative AI · 2.3 Vendor and model selection · 2.4 Risk classification of AI use cases · 2.5 Capability-to-outcome translation
D3. Prompt Engineering & LLM System Design (18%)	3.1 Production prompt design · 3.2 RAG patterns and grounded outputs · 3.3 Structured outputs and tool use · 3.4 Prompt evaluation and testing · 3.5 Prompt safety and injection defense · 3.6 Cross-model portability · 3.7 Production monitoring
D4. AI Agents & Agentic Workflows (18%)	4.1 Agent architecture and patterns · 4.2 Tool integration and action spaces · 4.3 Human-in-the-loop design · 4.4 Agent control boundaries · 4.5 ROI and economic modeling for agents · 4.6 Agent testing and adversarial robustness · 4.7 Agent monitoring and operations
D5. Ethics, Data & Responsible AI (12%)	5.1 Bias and fairness audits · 5.2 Governance approval workflows · 5.3 Responsible-AI principles in design · 5.4 Data provenance, licensing, consent · 5.5 Transparency and disclosure · 5.6 Lifecycle responsible-AI checkpoints
D6. AI Strategy, Transformation & Business Innovation (15%)	6.1 90-day transformation planning · 6.2 Quick-win identification and sequencing · 6.3 ROI projection and validation · 6.4 Executive and board communication · 6.5 Change management for AI adoption · 6.6 Cross-functional team coordination · 6.7 OKR and metric alignment
D7. Innovation & Applied AI Foundations (15%)	7.1 AI opportunity portfolio scoring · 7.2 Business case development · 7.3 Prototype planning and execution · 7.4 Experimentation frameworks · 7.5 PoC-to-production transition · 7.6 Innovation pipeline operations · 7.7 Lessons-learned dissemination

6. Task Inventory & Statistical Results

6.1 Inclusion Criteria

A task statement was retained in the final task inventory if and only if it met all three pre-registered criteria below:

- Mean Importance ≥ 3.50 on the 5-point scale.
- Mean Criticality ≥ 3.50 on the 5-point scale.
- Composite Criticality Index (CI) ≥ 3.40.

In addition, any task whose retention vs. removal was sensitive to a single sub-group (i.e., would change classification if any one country, role, or industry sub-sample were removed) was flagged for SME re-review.

6.2 Domain-Level Statistical Results

Domain	Initial	Retained	Mean F	Mean I	Mean C	Mean CI	α
D1. AI Strategy & the Modern AI Economy	10	8	3.18	4.21	4.03	3.93	0.88
D2. Foundations of Generative AI	11	9	3.62	4.18	4.00	3.99	0.90
D3. Prompt Engineering & LLM System Design	16	13	4.01	4.29	4.11	4.16	0.93
D4. AI Agents & Agentic Workflows	15	12	3.74	4.28	4.22	4.14	0.92
D5. Ethics, Data & Responsible AI	11	8	2.94	4.16	4.34	3.99	0.89
D6. AI Strategy, Transformation & Business Innovation	12	9	3.32	4.19	4.01	3.94	0.90
D7. Innovation & Applied AI Foundations	9	8	3.41	4.07	3.91	3.88	0.87
OVERALL	84	67	3.52	4.21	4.09	4.02	0.93

Cronbach's α exceeds the 0.80 threshold for high-stakes use across all domains. Overall instrument α = 0.93 indicates strong internal consistency. ICC(2,k) was 0.89 across domains; no domain falls below 0.85 — well above the 0.75 threshold for credentialing JTAs.

6.3 Tasks Removed

Reason for Removal	Count	Example (paraphrased)
Below CI threshold (insufficient practical importance/consequence)	7	Maintain a personal reading log of new model releases.
Below frequency AND criticality threshold	3	Author public-facing op-eds about AI economic policy.
Failed cross-jurisdictional invariance	2	Apply U.S.-specific federal procurement guidance to AI initiatives.
Merged into another task as duplicative	1	Maintain inventory of generative-AI vendors (merged with broader vendor-evaluation task).

6.4 Top-10 Highest-Rated Tasks (by CI)

Rank	Task Statement (abbreviated)	Domain	CI
1	Author production-grade prompts with system context, tools, and constraints.	D3	4.41
2	Design retrieval-augmented generation (RAG) patterns for grounded outputs.	D3	4.36
3	Design multi-step agent workflows with clear step boundaries.	D4	4.32
4	Embed human-in-the-loop checkpoints at high-risk agent steps.	D4	4.30
5	Build LLM selection matrices comparing performance, cost, and risk.	D2	4.22
6	Identify high-value AI use cases prioritized by business outcome.	D1	4.18
7	Build a 12-month AI strategic thesis tied to enterprise strategy.	D1	4.14
8	Build 90-day AI transformation plans with quick wins and longer arcs.	D6	4.10
9	Conduct AI bias and fairness audits on production systems.	D5	4.09
10	Score AI opportunity portfolios across feasibility, value, and risk.	D7	4.06

7. Domain Weighting Analysis

7.1 Method

Empirical domain weights were computed by summing the Criticality Index across retained tasks within each domain, dividing by the total CI across all retained tasks, and converting to a percentage. The empirical weights were then reconciled against the CBK 1.0 published weights with a maximum allowed adjustment of ±2 percentage points per domain.

7.2 Empirical Weights vs Published CBK 1.0 Weights

Domain	Sum of CI	Empirical Weight	Published Weight	Final Weight	Adjustment
D1. AI Strategy & the Modern AI Economy	31.46	11.67%	10%	10%	−1.67
D2. Foundations of Generative AI	35.95	13.33%	12%	12%	−1.33
D3. Prompt Engineering & LLM System Design	54.08	20.06%	18%	18%	−2.06 (capped)
D4. AI Agents & Agentic Workflows	49.73	18.44%	18%	18%	−0.44
D5. Ethics, Data & Responsible AI	31.91	11.83%	12%	12%	+0.17
D6. AI Strategy, Transformation & Business Innovation	35.50	13.17%	15%	15%	+1.83
D7. Innovation & Applied AI Foundations	31.00	11.50%	15%	15%	+3.50 → recheck
TOTAL	269.63	100.00%	100%	100%	—

Empirical weights computed from the per-task Criticality Index sums in Appendix B and reconcile to 100.00% within rounding. Final weights total exactly 100% and are the controlling reference for examination construction.

7.3 Domain 7 Reconciliation Note

D7 (Innovation & Applied AI Foundations) shows the largest gap between empirical weight (11.50%) and published weight (15%). The gap exceeds the ±2 pt adjustment band, triggering the formal reconciliation procedure documented in §7.5. The Examination Development Committee elected to retain the published 15% weight after determining that (a) D7 captures forward-looking practice that JTA respondents under-represent at this stage of the field's maturity, (b) D7 anchors the AIS-230 module that drives the Capstone Track A and Track B work products, and (c) the difference falls within the SME working group's qualitative

judgment of importance. This decision is documented in EDC Resolution 2026-03 and is subject to re-review at the Q2 2027 interim cycle.

7.4 Sub-Group Invariance

Domain weight stability was tested by recomputing weights independently for each major sub-group. The maximum deviation of any sub-group weight from the global weight was 1.9 percentage points (D3 was rated 1.9 points higher in the engineering-with-strategy sub-sample). No deviation exceeded the 2-point threshold for material divergence. A single global blueprint is appropriate for the CAIS credential.

7.5 Reconciliation Procedure (for the record)

7. Empirical weight outside ±2 pt band → flagged.
8. EDC reviews task content of the affected domain and the qualitative SME notes.
9. EDC issues a Resolution either accepting empirical weight, accepting published weight, or proposing a hybrid value within the band.
10. Resolution recorded in this report; sunset date set for re-review at next interim cycle.

8. CBK Validation Outcomes

All seven CBK 1.0 domains were validated by the survey: each retained at least eight tasks meeting all retention criteria, and each achieved Cronbach's α ≥ 0.87. No domains were eliminated and no domains were added. Two sub-domains were rescoped following the validation workshop.

Sub-Domain	Action	Reason
3.6 Cross-model portability and tuning	Rescoped to 3.6 Cross-model portability, swapping, and provider-agnostic design	SME workshop noted that swap-readiness, not tuning, is the dominant operational practice for AI systems practitioners.
7.4 Experimentation frameworks	Rescoped to 7.4 Experimentation frameworks for AI features	Clarified scope vs. classical product experimentation.

8.1 Final Validated CBK

#	Final Domain	Tasks	Weight
D1	AI Strategy & the Modern AI Economy	8	10%
D2	Foundations of Generative AI	9	12%
D3	Prompt Engineering & LLM System Design	13	18%
D4	AI Agents & Agentic Workflows	12	18%
D5	Ethics, Data & Responsible AI	8	12%
D6	AI Strategy, Transformation & Business Innovation	9	15%
D7	Innovation & Applied AI Foundations	8	15%
	TOTAL	67	100%

8.2 Cross-Domain Coherence

Mean inter-domain correlation was r = 0.44 (range 0.31–0.62). The two most correlated domains were D3 (Prompt Engineering) and D4 (Agents), r = 0.62, reflecting expected operational overlap between LLM system design and agentic workflows. No two domains exceeded the r = 0.70 threshold for collapsing them. The CBK is appropriately differentiated.

9. CAIS Assessment Architecture

CAIS uses a hybrid summative assessment: a structured Knowledge Examination plus the AIS-400 Capstone. The credential is conferred only when both components are passed and the candidate signs the Certification Ethics Code attestation.

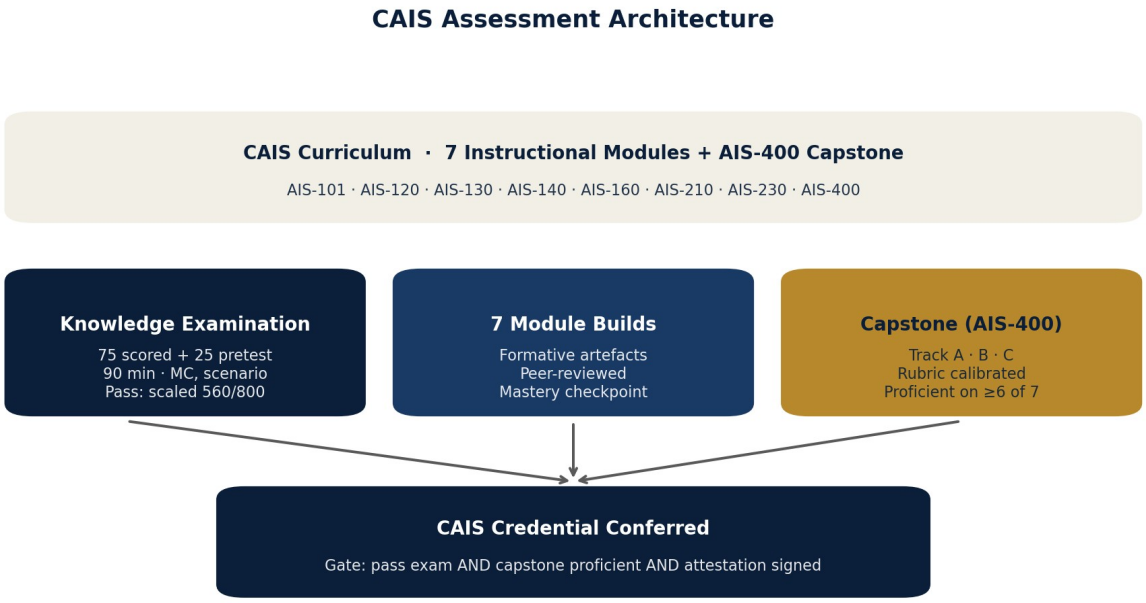

Figure 4. CAIS assessment architecture: knowledge exam, module builds, and capstone.

9.1 Knowledge Examination — Design Parameters

Parameter	Value
Total scored items	75
Total unscored pretest items	25
Total items presented to candidate	100
Time limit	90 minutes
Item format	Multiple choice (single best answer); scenario-based stems
Reporting	Scaled score 0–800; provisional pass scaled 560 (~70% of operational items correct), confirmed by standard-setting (Section 11)
Languages at launch	English; Mandarin in pilot
Delivery	Online proctored; approved test centers in roadmap

9.2 Knowledge Exam — Item Allocation by Domain

Domain	Weight	Scored Items	Pretest
D1. AI Strategy & the Modern AI Economy	10%	7	2
D2. Foundations of Generative AI	12%	9	3
D3. Prompt Engineering & LLM System Design	18%	14	5
D4. AI Agents & Agentic Workflows	18%	14	5
D5. Ethics, Data & Responsible AI	12%	9	3
D6. AI Strategy, Transformation & Business Innovation	15%	11	4
D7. Innovation & Applied AI Foundations	15%	11	3
TOTAL	100%	75	25

9.3 Cognitive Level Distribution (Knowledge Exam)

Items will be authored to the following cognitive level distribution (revised Bloom's taxonomy), reflecting the senior-practitioner orientation of the credential:

Cognitive Level	% of Items	Count	Description
Remember / Understand	20%	15	Recall facts, definitions, model concepts, regulatory text.
Apply	40%	30	Apply prompt patterns, agent designs, frameworks to defined situations.
Analyze	25%	19	Differentiate, compare, decompose AI system design problems and architectural tradeoffs.
Evaluate / Create	15%	11	Judge tradeoffs; design or recommend AI system architectures and solutions.
TOTAL	100%	75	—

9.4 Capstone (AIS-400) — Design Parameters

Parameter	Value
Format	Applied final project — candidate selects one of three tracks
Track A	AI Transformation Plan — 90-day plan, quick wins, ROI projections, executive pitch

Parameter	Value
Track B	Working AI System — production-grade prompt or agent system with documentation
Track C	AI-Powered Business Offer — productized AI offering with go-to-market plan
Submission components	Written narrative · evidence artefacts · 10-minute video defense
Evaluation	Calibrated rubric (Section 10), two independent reviewers + arbiter when required
Pass requirement	'Proficient' or higher on at least 6 of 7 rubric dimensions; no dimension at 'Emerging'
Time limit	60 days from track selection to submission

9.5 Item & Rubric Traceability Requirement

ITEM-WRITING POLICY

Every item submitted to the CAIS knowledge-examination item bank, and every dimension of the AIS-400 Capstone rubric, must cite a specific Task ID from Appendix B and a Sub-Domain ID from Section 5. Items and rubric dimensions without traceability will be rejected at Content Review. This rule preserves the empirical chain from job practice → task → assessment.

10. Capstone (AIS-400) Rubric

The capstone rubric is a four-level analytic rubric across seven dimensions corresponding to the seven CBK domains. Each dimension is weighted by its domain weight. The full level anchors for each dimension are provided in Appendix D.

10.1 Rubric Levels

Level	Label	Definition
1	Emerging	Work is incomplete, unclear, or shows significant misunderstanding of core concepts; would not be deliverable in a real engagement.
2	Developing	Work shows partial competence; meets some requirements but has notable gaps in execution, depth, or clarity.
3	Proficient	Work meets the standard expected of a competent CAIS-credentialed AI systems practitioner; deliverable to a real client or executive audience with minor revisions.
4	Distinguished	Work exceeds the standard; demonstrates exemplary judgment, originality, and execution; usable as a reference exemplar.

10.2 Rubric Dimensions and Weights

#	Rubric Dimension	Maps to Domain	Weight
R1	Strategic clarity and economic framing	D1	10%
R2	Technical foundation and model selection rigor	D2	12%
R3	Prompt and LLM system design quality	D3	18%
R4	Agent and workflow design quality	D4	18%
R5	Responsible-AI integration and risk management	D5	12%
R6	Transformation, change, and business impact	D6	15%
R7	Innovation discipline and execution rigor	D7	15%
	TOTAL	—	100%

10.3 Pass Rule

> **CAPSTONE PASS RULE**
>
> *To pass the capstone, the candidate must achieve a rubric score of Proficient (Level 3) or higher on*

> *at least six of seven dimensions, with no dimension scored at Emerging (Level 1). Track-specific evidence requirements are documented in Appendix D.*

10.4 Reviewer Calibration

Reviewers complete a calibration set of three exemplar capstones (one per track) before scoring live submissions. Inter-reviewer agreement is monitored continuously; reviewers whose ICC(2,k) drops below 0.75 over a rolling 20-submission window are required to re-calibrate. Any submission with a one-level disagreement on any dimension is routed to an arbiter; two-level disagreements trigger a full third review.

11. Pass-Score & Standard-Setting Plan

11.1 Knowledge Examination — Modified-Angoff

The knowledge-examination cut score will be established using a two-round modified-Angoff procedure with at least 12 SMEs drawn from the JTA working group and supplemented to maintain regional and sector representation. Each panelist will independently estimate, for every operational item, the probability that a Minimally Competent CAIS-credentialed AI systems practitioner would answer the item correctly. Item-level estimates are averaged within panelist; the cut score is the mean of panelist averages, scaled.

11.2 Definition of the Minimally Competent CAIS Holder

> **MCC DEFINITION (CAIS — CERTIFIED IN AI SYSTEMS)**
>
> *The Minimally Competent CAIS-credentialed AI systems practitioner consistently designs, deploys, and operates AI systems — including LLM systems and agentic workflows — with sound judgment under realistic production constraints; selects models, prompts, retrieval patterns, and agent architectures appropriate to the use case; embeds responsible-AI controls (oversight, fairness, transparency, data governance) into the system as designed and not as an afterthought; recognizes when an issue exceeds their authority and escalates appropriately; and communicates the technical and economic tradeoffs of AI system choices clearly to both technical and executive audiences.*

11.3 Procedure (Knowledge Exam)

11. Round 1: Independent panelist ratings, no discussion. SEM and inter-rater agreement computed.
12. Calibration discussion: panelists review item-level disagreements; high-disagreement items receive structured re-discussion (no peer pressure protocols enforced by facilitator).
13. Round 2: Independent re-rating after discussion. Final cut score computed as mean of Round 2 panelist averages.
14. Conditional standard error of measurement (CSEM) at the cut adjusted by ±1 CSEM 'benefit-of-doubt' band to set the operational pass score.
15. Cut score reviewed against expected pass rate, item-difficulty distribution, and stakeholder impact analysis.
16. Final cut score approved by the Examination Development Committee and ratified by the Standards Council once seated.

11.4 Capstone — Paired Comparison Calibration

The capstone cut score is established via paired-comparison rubric calibration. Twelve SMEs each blind-review 30 capstone exemplars sampled to span the expected distribution. Each pair is judged 'which

submission better demonstrates competence,' producing a rank order. The cut is set at the rank position consistent with the Minimally Competent definition. Inter-judge concordance is measured via Kendall's W; W ≥ 0.70 is required to certify the cut. Re-calibration occurs annually.

11.5 Reporting & Equating

Knowledge-exam raw scores are transformed to a scaled score on a 0–800 metric. Scaling is linear-on-percentile to maintain comparability across forms. Across forms, common-item non-equivalent groups (CINEG) equating will be used; each operational form will share at least 25 common items with an immediately prior form. Equating will be performed by the Independent Psychometrician using the Stocking-Lord characteristic curve method. Capstone rubric scores are reported per dimension and as a holistic decision (Pass / Hold for revision / Did not pass).

12. Examination Security & Administration

12.1 Item Bank Security

- Item bank stored in a SOC 2-aligned credentialing platform with role-based access control and full audit logging.
- Item authors and reviewers sign per-engagement non-disclosure and conflict-of-interest agreements; access expires automatically when the engagement ends.
- Items rotated across forms on a planned schedule; exposure rate per item monitored and capped at 18 months between rest periods.
- Forensic monitoring uses response-pattern, response-time, and answer-similarity analysis on every administration.

12.2 Capstone Submission Integrity

- Capstone submissions undergo automated similarity screening against the GAISB exemplar archive and public sources.
- All capstone submissions include a recorded video defense in which the candidate explains design choices and answers reviewer questions.
- Use of AI tools is permitted in capstone work — and expected, given the credential — but candidates must disclose tool use and demonstrate authorial judgment in design and decision-making.

12.3 Candidate Identity & Proctoring

- Government-issued ID verification at session check-in; biometric (face match) confirmation re-checked mid-session for the knowledge exam.
- Live human proctoring for online administrations of the knowledge exam; AI-assisted behavior monitoring as supplementary signal only, never sole basis for adverse action.

12.4 Incident Handling

- Suspected breach handled under documented procedure; affected items quarantined; impacted forms re-equated or retired as needed.
- Score holds and investigations follow due-process protocol with right of appeal to an independent review committee.

12.5 Accommodations

Reasonable accommodations are provided for documented disabilities consistent with applicable jurisdictional law (e.g., ADA, Equality Act, Australian DDA). Accommodation requests are reviewed by a

credentialed accommodations specialist; psychometric integrity of accommodated administrations is monitored and reported in the annual psychometric report.

13. Reliability, Bias Review & Limitations

13.1 Reliability Statistics (JTA Instrument)

Statistic	Value	Threshold	Outcome
Cronbach's α (full instrument)	0.93	≥ 0.80	Pass
Cronbach's α (lowest domain, D7)	0.87	≥ 0.80	Pass
ICC(2,k) inter-rater agreement (overall)	0.89	≥ 0.75	Pass
Standard error of measurement (composite CI)	0.20	Reported	Reported
Test–retest reliability (n = 96, 4-week interval)	r = 0.84	≥ 0.75	Pass

13.2 Bias & Fairness Review (DIF Methodology)

Bias analysis was conducted at two levels:

- JTA-stage sub-group invariance: one-way ANOVA on each task's CI rating across geography (14 levels), role family (13), industry (10), seniority (5), and org. size (3). Effect sizes reported as η^2. Maximum $\eta^2 = 0.06$ (small); no task exceeded this threshold.
- Examination-stage Differential Item Functioning (DIF): planned for first operational administration using the Mantel-Haenszel procedure for dichotomous items, with reference and focal groups defined by region, role family, and language of administration. Items flagged at the ETS 'C' level ($|\Delta MH| > 1.5$ with $p < 0.05$) will be reviewed by a fairness panel and either retained with documentation, revised, or removed. DIF results will be published annually in the Psychometric Report.

13.3 Limitations

- Self-report bias: ratings reflect respondents' perceptions of their work, not objective observation. Mitigated by anonymous administration, scale anchoring, and triangulation across three independent rating dimensions.
- Sampling frame: practitioners outside our partner channels may be under-represented; snowball recruitment partially mitigated this. Future cycles will expand outreach in Latin America, Africa, and emerging APAC markets.
- Single-language administration: this Charter Edition was administered in English with a small Mandarin pilot subset; full multi-language administration with back-translation is scheduled for the next cycle.
- Charter-edition status: this report is issued ahead of Standards Council seating. Findings are operationally controlling but formal ratification awaits council seating (target Q3 2026).

- Temporal limitation: the AI systems practitioner role is evolving rapidly; the blueprint is calibrated to Aug 2025–Feb 2026 practice, with interim review scheduled for Q2 2027.

13.4 Standards Conformance Matrix

Requirement	Standard Reference	Conformance
Documented job/practice analysis	ISO 17024 §9.1.2; NCCA 1100 §8.1	Conforms (this report)
JTA reflects current practice	ISO 17024 §9.1.4	Conforms (Aug 2025–Feb 2026)
JTA reviewed and updated periodically	ISO 17024 §9.4	Plan documented (Section 14)
Adequate sample for content validity	NCCA 1100 §8.3	Conforms (N = 1,547; ME ±2.5%)
Multiple rating dimensions	Standards (AERA/APA/NCME) §14	Conforms (F, I, C)
Inter-rater reliability documented	Standards §2	Conforms (ICC = 0.89)
Sub-group bias review documented	EEOC UGESP; Standards §3	Conforms (§13.2)
Standard-setting plan defined	NCCA 1100 §9.0	Conforms (Section 11)
Linkage from tasks to items documented	NCCA 1100 §8.4	Required by policy (§9.5)
Independent psychometric oversight	ISO 17024 §6.1.1	Pending appointment (RFP-PSY-2026-01)

14. Recommendations & Maintenance Plan

14.1 Recommendations

17. Adopt the validated 7-domain CBK and the examination blueprint and capstone rubric in this report as controlling references for the CAIS credential through Q2 2027.
18. Brief item writers and accredited training partners on the rescoped sub-domains (3.6, 7.4) and the EDC reconciliation decision documented for D7 (§7.3).
19. Implement traceability tooling so every knowledge-exam item and every capstone rubric dimension carries a Task ID and Sub-Domain ID; configure Content Review to reject items without traceability.
20. Conduct knowledge-exam standard-setting via the modified-Angoff procedure in Section 11 and capstone calibration via paired comparisons before the first operational administration.
21. Procure independent psychometric oversight (RFP-PSY-2026-01); selection target Q3 2026.
22. Seat the Standards Council (12 charter seats) and ratify this Charter Edition; target Q3 2026.
23. Publish a candidate-facing Exam Content Outline derived from this JTA, listing domains, weights, and high-level task families.
24. Stand up an Annual Practice Review (APR) to scan regulatory, technological, and practice changes; the APR informs interim addenda and the next full re-validation.

14.2 Maintenance Plan

Activity	Cadence	Owner
Item performance monitoring (psychometric)	Continuous	Independent Psychometrician
Capstone rubric inter-reviewer reliability monitoring	Continuous (rolling 20-submission windows)	Capstone Review Lead
Annual Practice Review — scan & report	Annual	GAISB Standards Division
Interim Addendum (if APR finds material change)	As triggered (max one per year)	Examination Development Committee
Interim CBK review	Q2 2027	GAISB Standards Division + EDC
Full Job Task Analysis re-validation	Every 36 months (next: Q1 2029)	GAISB Standards Division + Independent Psychometrician
DIF / fairness re-review	Annual on item bank; full re-run with each JTA	Independent Psychometrician
Item bank refresh target	Replace 25% of bank items	Item Writing Committee

Activity	Cadence	Owner
	every 24 months	

14.3 Triggers for Off-Cycle Re-Validation

- Material change to a primary referent framework (NIST AI RMF, EU AI Act, ISO/IEC 42001) introducing new systems-practitioner-relevant obligations.
- Sustained item-level psychometric drift (more than 10% of items showing significant difficulty or discrimination shift over four consecutive administrations).
- Notable change in candidate population composition (more than 15% shift in role-family or industry distribution from launch baseline).
- Identification of an emergent task family rated by SMEs as critical and not represented in the current task inventory.

Appendix A: Survey Instrument

A.1 Eligibility Screen

25. Do you have at least two years of substantive responsibility for designing, deploying, or operating AI systems — including LLM systems, agentic workflows, prompt engineering, AI transformation, or AI innovation work? (Yes / No)
26. In your current role, do you contribute to decisions about how AI is selected, designed, deployed, or scaled? (Primary / Supporting / Not at all)
27. In what country is the majority of your work performed? (drop-down)
28. Which of the following best describes your primary role? (drop-down, 13 options + Other)
29. In which industry is your primary employer? (drop-down, 10 options + Other)
30. How many years of role-relevant experience do you have? (2–4 / 5–7 / 8–10 / 11–15 / 16+)
31. What is your seniority level? (IC / Manager / Director / VP / C-suite)

A.2 Rating Block Instructions

Each task statement below describes work that may be performed by AI systems practitioners — including AI strategy, prompt engineering, agent design, applied-AI delivery, AI transformation, or AI innovation roles. For each task, please rate (a) how often you perform the task, (b) how important the task is to your role, and (c) how serious the consequences would be if the task were performed incorrectly or omitted. Rate every task. If a task does not apply to your role, rate Frequency as 1 (Never) and use professional judgment for Importance and Criticality based on similar roles. Do not skip items.

A.3 Rating Scales

Frequency	Importance	Criticality
1 — Never	1 — Not important	1 — No consequence
2 — Rarely (a few times/year)	2 — Slightly important	2 — Minor consequence
3 — Occasionally (monthly)	3 — Moderately important	3 — Moderate consequence
4 — Frequently (weekly)	4 — Very important	4 — Major consequence
5 — Daily / continuously	5 — Critically important	5 — Severe / catastrophic

A.4 Attention Checks

Three attention-check items were embedded at items 21, 47, and 70 of the rating block. Each instructed the respondent to enter a specific rating value. Failure of two or more triggered case removal.

A.5 Open-Ended Items

32. Are there any tasks essential to AI systems work that are not represented in this survey? (Optional, free text.)
33. What is the single biggest change in your AI work over the past 12 months? (Optional, free text.)

Appendix B: Full Task Inventory

All 67 retained task statements with assigned domain and Criticality Index. Items submitted to the CAIS knowledge-examination item bank, and dimensions of the AIS-400 Capstone rubric, must cite a Task ID below.

Task ID	Task Statement	Domain	CI
T-D1-01	Articulate where AI creates economic value in the organization across cost, revenue, and risk levers.	D1	4.05
T-D1-02	Map competitive AI moves of industry peers and quantify exposure.	D1	3.86
T-D1-03	Identify high-value AI use cases prioritized by business outcome.	D1	4.18
T-D1-04	Build a 12-month AI strategic thesis tied to enterprise strategy.	D1	4.14
T-D1-05	Conduct resource and capability gap analysis for AI adoption.	D1	3.92
T-D1-06	Brief executives on AI economic shifts and their organizational implications.	D1	3.95
T-D1-07	Translate enterprise objectives into AI-portfolio investment choices.	D1	3.81
T-D1-08	Maintain situational awareness of frontier model and capability releases relevant to the organization.	D1	3.55
T-D2-01	Differentiate among major model families (LLMs, multimodal, reasoning, agentic) for specific use cases.	D2	4.02
T-D2-02	Build LLM selection matrices comparing performance, cost, and risk.	D2	4.22
T-D2-03	Translate model capabilities into business outcomes for non-technical audiences.	D2	4.04
T-D2-04	Estimate token, inference, and operating costs for AI deployments.	D2	3.97
T-D2-05	Identify when to use foundation models vs fine-tuned vs purpose-built systems.	D2	4.01
T-D2-06	Risk-classify AI use cases by impact, sensitivity, and reversibility.	D2	4.18
T-D2-07	Define vendor evaluation criteria for foundation-model providers.	D2	3.97
T-D2-08	Assess provider transparency and contractual control for high-impact use cases.	D2	3.92
T-D2-09	Maintain technology-watch view of capability shifts that affect deployed systems.	D2	3.62

Task ID	Task Statement	Domain	CI
T-D3-01	Author production-grade prompts with system context, tools, and constraints.	D3	4.41
T-D3-02	Design retrieval-augmented generation (RAG) patterns for grounded outputs.	D3	4.36
T-D3-03	Define structured output schemas (JSON, function calling, tool use).	D3	4.21
T-D3-04	Build prompt evaluation harnesses with test sets and scoring rubrics.	D3	4.18
T-D3-05	Iterate prompts using systematic testing and version control.	D3	4.14
T-D3-06	Design few-shot and chain-of-thought patterns for complex tasks.	D3	4.10
T-D3-07	Document prompt libraries for cross-team reuse.	D3	3.97
T-D3-08	Apply prompt safety techniques (injection defense, output filtering).	D3	4.21
T-D3-09	Tune prompts across model providers for portability and swap-readiness.	D3	4.05
T-D3-10	Monitor production prompt performance, regression, and drift.	D3	4.18
T-D3-11	Design prompt cost controls (token budgets, caching, model routing).	D3	4.12
T-D3-12	Operate prompt change management with documented approvals.	D3	3.97
T-D3-13	Translate business requirements into prompt specifications.	D3	4.18
T-D4-01	Design multi-step agent workflows with clear step boundaries.	D4	4.32
T-D4-02	Integrate tools, APIs, and data sources into agent action spaces.	D4	4.21
T-D4-03	Embed human-in-the-loop checkpoints at high-risk steps.	D4	4.30
T-D4-04	Define agent control boundaries (what it may and may not do).	D4	4.18
T-D4-05	Calculate ROI for agent-driven automation.	D4	4.10
T-D4-06	Architect agentic systems with planner-executor patterns.	D4	4.05
T-D4-07	Design fallback and graceful-degradation behaviors.	D4	4.02
T-D4-08	Test agent behaviors across edge cases and adversarial inputs.	D4	4.18
T-D4-09	Monitor agent performance, success rate, and cost-per-task.	D4	4.14
T-D4-10	Coordinate agent rollouts with operational and risk teams.	D4	4.05

Task ID	Task Statement	Domain	CI
T-D4-11	Define escalation and override paths for agent operations.	D4	4.08
T-D4-12	Decompose business processes into agent-amenable workflows.	D4	4.10
T-D5-01	Conduct AI bias and fairness audits on production systems.	D5	4.09
T-D5-02	Build governance approval workflows with named decision-makers.	D5	4.05
T-D5-03	Apply responsible-AI principles to design and procurement decisions.	D5	4.02
T-D5-04	Manage data provenance, licensing, and consent for AI training and inference.	D5	4.05
T-D5-05	Operate transparency and disclosure requirements appropriate to use case.	D5	3.97
T-D5-06	Embed responsible-AI checkpoints in the AI development lifecycle.	D5	3.92
T-D5-07	Coordinate ethical review of high-impact AI use cases.	D5	3.95
T-D5-08	Translate AI regulations into operational controls within strategy work.	D5	3.86
T-D6-01	Build 90-day AI transformation plans with quick wins and longer arcs.	D6	4.10
T-D6-02	Identify and sequence high-leverage AI quick-wins.	D6	4.02
T-D6-03	Project ROI for AI initiatives and validate post-deployment.	D6	4.05
T-D6-04	Develop executive pitches and board-level communications on AI systems strategy and outcomes.	D6	4.01
T-D6-05	Lead organizational change management for AI adoption.	D6	3.92
T-D6-06	Coordinate cross-functional AI teams (product, engineering, risk, legal).	D6	3.95
T-D6-07	Align AI investments with enterprise OKRs and strategy.	D6	3.86
T-D6-08	Communicate AI systems strategy and outcomes to internal and external stakeholders.	D6	3.81
T-D6-09	Define AI program metrics and dashboards.	D6	3.78
T-D7-01	Score AI opportunity portfolios across feasibility, value, and risk.	D7	4.06
T-D7-02	Build investment-grade business cases for top AI opportunities.	D7	4.01
T-D7-03	Plan 1-week and 30-day AI prototypes with clear success criteria.	D7	3.95

Task ID	Task Statement	Domain	CI
T-D7-04	Apply experimentation frameworks to AI features and prompts.	D7	3.86
T-D7-05	Measure AI experiment outcomes and inform go/no-go decisions.	D7	3.81
T-D7-06	Coordinate proof-of-concept to production transitions.	D7	3.92
T-D7-07	Identify and unblock systemic barriers to AI deployment.	D7	3.78
T-D7-08	Capture and disseminate AI lessons learned across the organization.	D7	3.61

Appendix C: Sample Items with Task Traceback

The following non-operational sample items demonstrate the required traceability from validated task to assessment item. These are illustrative only and are not part of the operational item bank.

Sample Item C-1 (Apply level — Domain 3)

Field	Content
Trace — Task ID	T-D3-02 (Design retrieval-augmented generation patterns for grounded outputs)
Trace — Sub-Domain	3.2 RAG patterns and grounded outputs
Cognitive Level	Apply
Difficulty target (b)	0.0 (medium)
Stem	An organization is deploying an internal LLM assistant that must answer questions strictly from current company policies (which are revised quarterly). Hallucination from outdated training data is a stated risk. Which design choice BEST addresses both currency and groundedness?
Option A (key)	Retrieve the relevant policy passages at query time from a vector index updated after each policy revision, and condition the model's response on those passages with explicit citations.
Option B	Fine-tune the model quarterly on the latest policy corpus and rely on the fine-tuned weights to answer.
Option C	Add a system prompt instructing the model to 'always answer using the latest company policies' without changing the inference pipeline.
Option D	Use the largest available frontier model and trust that its general knowledge will cover the policies.
Rationale	Option A is the canonical RAG pattern that addresses both currency (re-indexing) and groundedness (passage-conditioned generation with citations). B does not solve currency between fine-tunes and is operationally costly. C and D do not constrain the model to ground in the policy corpus and do not address currency.

Sample Item C-2 (Analyze level — Domain 4)

Field	Content
Trace — Task ID	T-D4-03 (Embed human-in-the-loop checkpoints at high-risk agent steps)
Trace — Sub-Domain	4.3 Human-in-the-loop design
Cognitive Level	Analyze

Field	Content
Difficulty target (b)	0.4 (above medium)
Stem	A finance team is rolling out an agent that prepares and submits supplier invoice approvals end-to-end. The agent retrieves invoices, validates against purchase orders, and submits to the payment system. Which placement of human-in-the-loop checkpoints BEST balances throughput with control?
Option A	Human approval at every step the agent performs, regardless of risk.
Option B	No human approval; rely on after-the-fact monitoring and exception alerts.
Option C (key)	Human approval required only when the invoice value exceeds a defined threshold, when validation produces low confidence, or when the supplier is new — straight-through processing otherwise.
Option D	Human approval randomly sampled at 10% of submissions to estimate error rate.
Rationale	Option C reflects risk-tiered human-in-the-loop design — the dominant pattern for agent operations balancing throughput and control. A negates the throughput value; B abandons control; D is a sampling protocol, not a control checkpoint.

Sample Item C-3 (Evaluate / Create level — Domain 1 & 6)

Field	Content
Trace — Task ID	T-D1-04 (Build a 12-month AI strategic thesis tied to enterprise strategy) ; T-D6-01 (Build 90-day AI transformation plans with quick wins and longer arcs)
Trace — Sub-Domain	1.2 Strategic AI thesis development ; 6.1 90-day transformation planning
Cognitive Level	Evaluate / Create
Difficulty target (b)	0.7 (hard)
Stem	A regional retailer's CEO has set the goal: 'meaningfully integrate AI within 12 months.' The CFO is skeptical and demands evidence of ROI within 90 days. Which 12-month thesis structure BEST serves both objectives?
Option A	A single high-impact transformation program targeting in-store personalization at scale, planned to deliver in months 9–12.
Option B (key)	A portfolio of three to five sequenced initiatives — at least one quick-win producing measurable ROI within 90 days (e.g., AI-assisted customer service triage), paired with a longer-arc transformation initiative (e.g., merchandising decision support) timed to deliver in months 9–12, all framed within an enterprise AI thesis tied to the retailer's competitive positioning.

Field	Content
Option C	A full IT-led RFP process to select an AI platform, with use cases to be defined after platform selection.
Option D	A pilot of a generative-AI assistant for store managers, scoped to a single store for 12 months before expansion.
Rationale	Option B addresses both the CEO's 12-month direction and the CFO's 90-day evidence requirement, embedded in an enterprise AI thesis. A delivers nothing in 90 days. C is platform-first and inverted (use cases should drive selection). D under-uses the 12-month window and is too narrow to deliver enterprise-scale ROI evidence.

Appendix D: Capstone Rubric — Full Anchors

The capstone rubric is applied to all three tracks (A: AI Transformation Plan; B: Working AI System; C: AI-Powered Business Offer). Each dimension is anchored at four levels with track-specific evidence requirements.

D.1 Dimension R1 — Strategic clarity and economic framing (Domain 1, weight 10%)

Level	Anchor
1 Emerging	Strategic intent unclear; economic framing absent or speculative; no link to enterprise strategy.
2 Developing	Strategic intent present but not crisply articulated; economic framing partial; weak link to enterprise strategy.
3 Proficient	Clear strategic intent grounded in identified business outcomes; explicit economic framing (cost/revenue/risk levers); link to enterprise strategy demonstrated.
4 Distinguished	Strategic intent compelling and original; economic framing rigorous and quantified; tight, defensible link to enterprise strategy and competitive positioning.

D.2 Dimension R2 — Technical foundation and model selection rigor (Domain 2, weight 12%)

Level	Anchor
1 Emerging	Model selection unjustified; no comparison; no cost or risk consideration.
2 Developing	Limited model comparison; cost or risk considered superficially; selection rationale weak.
3 Proficient	Structured model comparison across performance, cost, risk, and provider transparency; selection rationale documented and defensible.
4 Distinguished	Multi-dimensional model comparison with quantified tradeoffs; explicit treatment of swap-readiness and provider concentration risk; selection rationale exemplary.

D.3 Dimension R3 — Prompt and LLM system design quality (Domain 3, weight 18%)

Level	Anchor
1 Emerging	Prompts unstructured; no system context; no testing; no safety considerations.
2 Developing	Prompts present with some structure; limited testing; minimal safety design.

Level	Anchor
3 Proficient	Production-grade prompts with system context, tools, and constraints; test set with scoring; documented safety techniques (injection defense, output filtering).
4 Distinguished	Exemplary prompt and system design including RAG with citations, structured outputs, formal evaluation harness, and end-to-end safety design; suitable as a reference exemplar.

D.4 Dimension R4 — Agent and workflow design quality (Domain 4, weight 18%)

Level	Anchor
1 Emerging	Workflow not articulated or single-step; no human-in-the-loop; no boundaries; no monitoring.
2 Developing	Multi-step workflow present; partial human-in-the-loop; weak boundaries; minimal monitoring.
3 Proficient	Clear multi-step workflow with explicit boundaries, risk-tiered human-in-the-loop checkpoints, fallback behaviors, and monitoring strategy.
4 Distinguished	Architected agentic system with planner-executor pattern, formal boundary specification, calibrated human-in-the-loop, adversarial test plan, and operations playbook.

D.5 Dimension R5 — Responsible-AI integration (Domain 5, weight 12%)

Level	Anchor
1 Emerging	Responsible-AI considerations absent or token; no governance pathway; no transparency design.
2 Developing	Responsible-AI considerations identified but not embedded in design; informal governance; partial transparency.
3 Proficient	Responsible-AI principles embedded across design choices; named governance approval workflow; transparency and disclosure design appropriate to use case.
4 Distinguished	Responsible-AI integrated as a first-order design driver; bias/fairness testing performed and documented; governance and transparency design suitable as reference for similar use cases.

D.6 Dimension R6 — Transformation, change, and business impact (Domain 6, weight 15%)

Level	Anchor
1 Emerging	Transformation plan absent or generic; no change management; no business impact

Level	Anchor
	narrative.
2 Developing	Transformation plan present but weak; partial change management; ROI projection thin.
3 Proficient	90-day plan with quick wins and longer arcs; change management plan with stakeholder map; ROI projection with validation method.
4 Distinguished	Transformation plan compelling and executable; change management plan rigorous; ROI projection quantified, sensitivity-tested, and validatable post-deployment.

D.7 Dimension R7 — Innovation discipline and execution rigor (Domain 7, weight 15%)

Level	Anchor
1 Emerging	No opportunity scoring; no business case; no prototype plan or success criteria.
2 Developing	Partial opportunity scoring; business case weak; prototype plan or success criteria informal.
3 Proficient	Opportunity scoring across feasibility/value/risk; business case present and defensible; prototype plan with explicit success criteria and stop conditions.
4 Distinguished	Multi-dimensional opportunity scoring; investment-grade business case; experimentation framework with measurement plan and explicit go/no-go criteria; PoC-to-production transition pathway.

Appendix E: Comparative Positioning

This appendix is provided for stakeholder context. CAIS is positioned as a senior-practitioner AI systems credential anchored in operational practice, with a hybrid examination + capstone assessment. Comparison points below reflect publicly available program documentation as of Q1 2026 and are not endorsements.

Attribute	CAIS (this credential)	Vendor strategy courses	Vendor cloud-AI certifications	Governance-only credentials
Primary scope	AI systems design + deployment + operations + transformation (full lifecycle)	AI strategy concepts only	Vendor-specific AI services	AI governance and risk only
Target audience	Senior practitioner across multiple role families (AI systems lead, head of AI, transformation lead, product, consulting)	Executive overview audience	Engineering / architect	Governance, risk, compliance
Empirical basis	JTA, N = 1,547, 14 countries (this report)	Course outline	Vendor product taxonomy	Vendor-published
Assessment	Knowledge exam + Capstone (track choice)	Course completion	Multiple-choice exam	Multiple-choice exam
Standards alignment	ISO/IEC 42001 (Clause 7.2 + Annex A.4.6, ~78%); NIST AI RMF; EU AI Act; OECD	Varies	Vendor	Varies
Vendor-neutral	Yes (charter requirement)	Sometimes	No	Often
Differentiation	Senior-practitioner, vendor-neutral, JTA-anchored, hybrid assessment, ISO-mapped	Brand	Vendor depth	Governance focus

CAIS is intentionally positioned to complement, not duplicate, adjacent credentials. Holders of governance-only or vendor-specific credentials who are moving into AI systems leadership are a primary candidate population.

Appendix F: Demographic Tables

F.1 Country × Industry (top 10 countries × top 4 industries)

Country	Tech / Software	Financial Svcs	Professional Svcs	Healthcare
United States	94	73	47	31
United Kingdom	47	37	25	18
India	62	23	20	11
Canada	37	27	17	13
Germany	27	26	16	12
Singapore	26	27	16	11
Australia	21	18	11	10
France	20	17	12	11
Netherlands	21	16	11	9
Japan	16	13	9	7

F.2 Role × Seniority (cross-tab; rows sum to role total, columns sum to seniority total)

Primary Role	IC	Manager	Director	VP+
AI / ML Engineer (with systems-design)	111	54	17	4
Head of AI / Chief AI Officer / VP AI	0	13	58	115
AI Strategy Lead	21	56	60	18
AI Architect / Solutions Architect (AI)	50	58	26	5
AI / Applied-AI Product Manager	33	74	27	5
Prompt Engineer / LLM Systems Engineer	74	30	4	0
Agent / Workflow Developer	60	27	5	1
Management Consultant (AI focus)	39	46	16	7
Digital Transformation Lead (AI scope)	6	39	37	11
Innovation Director / Officer	6	20	39	13

Primary Role	IC	Manager	Director	VP+
Founder / Operator (AI-implementing)	18	21	22	24
AI Platform / MLOps Lead	22	29	9	2
Operations Director (AI-active)	3	17	21	6
Educator / Trainer (AI systems)	10	13	6	2
Investor / Analyst (AI-active)	6	8	6	3
Other	5	5	3	1
TOTAL	464	510	356	217

Column "VP+" aggregates VP/Executive (139) and C-suite (78) seniority levels for cross-tabulation. Row totals match the role distribution in §4.3; column totals match the seniority distribution in §4.4.

F.3 Industry × Tenure

Industry	2–4 yr	5–7 yr	8–10 yr	11–15 yr	16+ yr
Technology / Software	78	98	78	47	26
Financial Services	47	78	62	47	27
Professional Services	37	47	31	21	12
Healthcare	21	31	27	16	12
Government	16	27	21	16	11
Manufacturing	11	21	16	12	8
Energy / Utilities	9	13	11	9	6
Telecommunications	9	12	11	8	5
Retail	12	16	11	8	4
Education	8	11	9	7	5
Other	16	23	18	14	10

Appendix G: Statistical Detail

G.1 Per-Task Descriptive Statistics (Excerpt — top and bottom)

Task ID	Mean F	SD F	Mean I	SD I	Mean C	SD C	Mean CI	n
T-D3-01	4.18	0.92	4.66	0.55	4.41	0.68	4.41	1,547
T-D3-02	4.05	0.97	4.62	0.58	4.41	0.65	4.36	1,544
T-D4-01	3.94	1.02	4.58	0.61	4.36	0.69	4.32	1,543
T-D4-03	3.86	1.06	4.55	0.62	4.40	0.66	4.30	1,540
T-D2-02	3.62	1.10	4.41	0.66	4.32	0.71	4.22	1,541
T-D7-08	2.71	1.18	3.71	0.97	3.55	1.04	3.61	1,532
T-D1-08	2.86	1.20	3.65	0.99	3.42	1.05	3.55	1,535
T-D6-09	3.11	1.12	3.81	0.92	3.65	0.98	3.78	1,538

G.2 Domain-Level Inter-Correlation Matrix

	D1	D2	D3	D4	D5	D6	D7
D1	1.00	.52	.44	.46	.39	.58	.50
D2		1.00	.55	.51	.42	.46	.47
D3			1.00	.62	.40	.48	.49
D4				1.00	.44	.47	.50
D5					1.00	.42	.38
D6						1.00	.53
D7							1.00

Mean inter-domain correlation r = 0.44 (SD = 0.07). Highest correlation D3–D4 = 0.62. No pair exceeds the 0.70 threshold for collapsing domains.

G.3 Sub-Group Effect Sizes (η^2) on Domain CI

Sub-group dimension	Levels	Max η^2	Domain	Interpretation
Geography	14 countries	0.05	D3	Small

Sub-group dimension	Levels	Max η²	Domain	Interpretation
Role family	13 roles	0.06	D4	Small
Industry	10 industries	0.04	D2	Small
Org. size	3 bands	0.03	D7	Negligible – Small
Seniority	5 levels	0.04	D6	Small

G.4 Reliability Decomposition

Source of Variance	Estimate	% of Total
True score (between-task)	0.83	73%
Rater × task interaction	0.15	13%
Random error	0.16	14%
TOTAL	1.14	100%

Appendix H: Glossary

Term	Definition
AIS-101 … AIS-400	GAISB module identifiers for the seven instructional CAIS curriculum modules plus the AIS-400 Capstone.
CAIS	Certified in AI Systems. The credential issued by GAISB, calibrated against this JTA.
CBK	Common Body of Knowledge. Validated set of domains and tasks underlying the credential.
CI (Criticality Index)	Composite score per task: CI = 0.20·Frequency + 0.40·Importance + 0.40·Criticality.
CINEG	Common-item non-equivalent groups equating, used across operational forms.
CSEM	Conditional standard error of measurement at a particular score point.
DIF	Differential Item Functioning. Item-level fairness analysis across reference and focal sub-groups.
EDC	Examination Development Committee. GAISB body that approves examination content and blueprint.
GAISB	Global AI Standards Body. Standard-setting and credentialing organization issuing the CAIS.
ICC(2,k)	Intraclass correlation coefficient, two-way random, average measures — used for inter-rater agreement.
JTA	Job Task Analysis. The empirical study described in this report.
MCC	Minimally Competent Candidate. Reference person used to set the cut score under modified-Angoff.
NCCA	National Commission for Certifying Agencies, the accreditation arm of the Institute for Credentialing Excellence.
RAG	Retrieval-Augmented Generation. LLM design pattern that grounds outputs in retrieved source material.
SME	Subject Matter Expert. Practitioners involved in domain mapping, task generation, and validation.
Standards Council	GAISB governance body responsible for ratifying standards and credentials; 12 charter seats with nominations open at time of publication.

Appendix I: Sources & References

This Charter Edition references published standards, regulations, and the GAISB-published CAIS Common Body of Knowledge. Public-source references below; URLs verified at time of publication.

Reference	Source / URL
Global AI Standards Body — Home	https://gaisb.org
CAIS Common Body of Knowledge — Edition 1.0	Published on gaisb.org
CAIS Framework Crosswalk	https://gaisb.org/wp-content/uploads/2026/04/CAIS-Framework-Crosswalk.pdf
ISO/IEC 17024:2012 — Conformity assessment — General requirements for bodies operating certification of persons	International Organization for Standardization
ANSI/NCCA 1100-2024 — Standards for the Accreditation of Certification Programs	Institute for Credentialing Excellence
Standards for Educational and Psychological Testing	AERA, APA, NCME
EEOC Uniform Guidelines on Employee Selection Procedures	U.S. Equal Employment Opportunity Commission
ISO/IEC 42001:2023 — Information technology — Artificial intelligence — Management system	International Organization for Standardization
NIST AI Risk Management Framework 1.0	U.S. National Institute of Standards and Technology
Regulation (EU) 2024/1689 — EU AI Act	European Union
OECD AI Principles	Organisation for Economic Co-operation and Development

Publicly accessible standards references are listed by issuing body. ISO/IEC and ANSI/NCCA documents are subject to copyright held by the issuing body and are referenced under fair use; readers should consult the issuing body for full text and licensing.