From Generative AI Adoption to Responsible Innovation: A Measurement Model for Trustworthy AI Innovation in Consulting Services

Manuscript prepared for Open Access Journal of Data Science & Artificial Intelligence (OAJDA), ISSN 2996-671X

Cuauhtémoc G. Gómez-D* | ORCID: 0009-0000-9133-226X | CVU: 721181

ARTIFEX Consulting; Doctoral Program in Technology and Innovation Management, Universidad Tecnológica Latinoamericana en Línea (UTEL), México

*Corresponding author: ggomez@artifex-consulting.io

Margarita García Nieto, Ph.D. in Education** | ORCID: 0009-0006-4335-5145 | CVU: 2112701

CECAPMKG Consulting; Doctoral Program in Technology and Innovation Management, Universidad Tecnológica Latinoamericana en Línea (UTEL), México

**Corresponding coauthor: direccion@cecapmkg.edu.mx

Abstract

This paper develops a measurement-oriented model for responsible innovation with trustworthy artificial intelligence (AI) in a technology consulting context. The case is ARTIFEX Consulting, a senior-led firm operating in AI-first transformation, enterprise architecture, data governance, retrieval-augmented generation (RAG), and advanced analytics. The central problem is that generative AI adoption is often framed as innovation although many implementations lack validated problem definition, evidence traceability, cost control, anti-hallucination mechanisms, anti-sycophancy controls, and value capture. To address this gap, the paper proposes the Responsible AI Innovation Index (RAII), a construct-based model that links creativity, inventiveness, research, development, and innovation (R&D&I), governance, risk control, and economic value. The methodological contribution is a future validation protocol that integrates expert-based content validity, exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and finite population sample planning. A pilot scenario with 10 homogeneous participants is treated as analytic calibration rather than population inference. The paper concludes that trustworthy AI innovation must be evaluated as a socio-technical measurement system rather than as isolated software deployment.

Keywords: responsible AI innovation; RAG; AI governance; exploratory factor analysis; confirmatory factor analysis; generative AI; measurement model; business model innovation.

1. Introduction

Generative AI has accelerated organizational experimentation with automated document analysis, decision support, conversational interfaces, software engineering, and knowledge retrieval. Nevertheless, adoption does not automatically constitute innovation. Following the Oslo Manual, innovation requires a new or significantly improved product or business process that is implemented or made available for use (OECD/Eurostat, 2018). In contrast, creativity generates possible alternatives, inventiveness configures functional solutions, and research, development, and innovation (R&D&I) systematizes knowledge creation and application (OECD, 2015; WIPO, 2024). The empirical and managerial problem addressed in this paper is therefore not whether firms can use AI, but whether they can convert AI use into a measurable, governed, and value-producing innovation capability.

The selected case is ARTIFEX Consulting, a technology consulting firm focused on AI-first transformation, enterprise architecture, data engineering, MLOps, RAG, cloud, and governance. The area of opportunity is the industrialization of responsible AI innovation: moving from expert-led isolated engagements to a repeatable measurement and governance model capable of demonstrating problem-solution fit, traceability, risk reduction, cost transparency, and return on investment (ROI). This framing is aligned with the OAJDA scope, which welcomes foundational and applied AI-related work (Medwin Publishers, 2025).

The research question is: *How can responsible AI innovation be conceptualized and validated as a measurable socio-technical construct in a consulting firm?* The paper contributes a mathematical index, a factor-analytic validation protocol, and a structured EFA/CFA design for future empirical testing.

2. Theoretical Background

Innovation research distinguishes the production of ideas from the implementation of value. The Oslo Manual defines innovation in relation to implementation and significant difference from previous products or processes (OECD/Eurostat, 2018). The Frascati Manual frames R&D as creative and systematic work aimed at increasing knowledge and devising new applications (OECD, 2015). WIPO's innovation ecosystem perspective adds the market, institutional, and knowledge infrastructure dimensions (WIPO, 2024). In AI-enabled business model innovation, value emerges not only through algorithmic capability but also through the redesign of value creation, delivery, and capture mechanisms (Khan et al., 2026).

Responsible AI governance extends this innovation view. NIST AI RMF organizes AI risk through Govern, Map, Measure, and Manage functions (NIST, 2023), while the NIST Generative AI Profile identifies risks such as hallucination, information-integrity failures, privacy, cybersecurity, and overreliance (NIST, 2024). ISO/IEC 42001 formalizes AI as a management-system concern rather than as an isolated technical artefact (ISO/IEC, 2023). Technical literature on RAG, hallucination, and sycophancy further shows that generative outputs require external evidence, retrieval evaluation, and human challenge mechanisms (Gao et al., 2023; Huang et al., 2023; Sharma et al., 2023).

From these sources, responsible innovation is treated here as an implemented and governed socio-technical capability. It requires: (a) creativity to produce alternatives; (b) inventiveness to configure viable solutions; (c) R&D&I to systematize experimentation; (d) governance to control risk; and (e) value capture to justify scaling.

3. Research Context and Area of Opportunity

ARTIFEX Consulting's opportunity is not the absence of technical capability; rather, it is the need to standardize how AI initiatives are selected, validated, deployed, and monetized. The specific area of opportunity is the lack of an explicit measurement model that integrates pain points, authorized sources, AI governance, anti-hallucination validation, anti-sycophancy controls, GenAI cost visibility, and ROI. Without such a model, AI adoption may remain a demonstration of technical novelty rather than a defensible innovation practice.

Table 1 summarizes the proposed constructs. The model deliberately separates idea generation from implementation and value capture, because the scientific contribution is not merely to list sources but to identify models, variables, similarities, differences, and the author's position.

Table 1: Proposed constructs for responsible AI innovation.

Construct	Definition	Observable indicators	Role in model
Creativity (CRE)	Generation of novel and useful alternatives.	Idea originality; diversity of options; relevance to pain point.	Input capability.
Inventiveness (INV)	Configuration of functional solutions.	Prototype viability; architectural feasibility; technical novelty.	Transformational mechanism.
R&D&I (RDI)	Systematic knowledge creation and application.	Evidence mapping; experimentation; documentation.	Methodological discipline.
AI Governance (GOV)	Policies, roles, and controls for AI risk.	AIMS; risk register; quality gates; auditability.	Risk and trust layer.
Value Capture (VAL)	Business outcome from implemented AI.	ROI; cost avoided; productivity; NPS; win-rate.	Innovation outcome.

Table 2: Ten-dimensional operationalization of the Responsible AI Innovation Index (RAII).

Code	Dimension	Operational definition	Indicative variable
D1	Pain point	Identifies a measurable operational, strategic, or regulatory problem.	Problem clarity; owner; baseline; acceptance criteria.
D2	Value hypothesis	Connects the solution to expected value before development.	KPI; value assumption; risk/cost/benefit logic.
D3	Applied research	Structures R&D&I, evidence review, and source analysis.	Evidence map; R&D&I plan; constraints; literature support.
D4	Development	Builds the AI-enabled artifact or workflow.	Prototype; RAG/agent/model; architecture; security.
D5	Validation	Tests factuality, quality, residual risk, and release criteria.	Quality gates; factual checks; human review; residual risk.
D6	Governed deployment	Deploys the solution with accountability and operational controls.	Runbook; SLA/SLO; monitoring; change control.
D7	Value capture	Measures adoption and realized non-financial/business outcomes.	Adoption; productivity; NPS; avoided risk.
D8	Anti-sycophancy control	Measures the ability to challenge user assumptions and avoid compliant but weak answers.	Adversarial prompts; disagreement rationale; reviewer challenge.
D9	Anti-hallucination control	Measures grounding, traceability, and factual verification.	Authorized sources; RAG evaluation; citations; verification thresholds.
D10	Economic value and ROI	Quantifies cost-to-serve, GenAI cost, and financial return.	Token/RAG cost; cost avoided; ROI; prioritization score.

4. Mathematical Model

The core dependent construct is Responsible AI Innovation (RAI). For operational use, the model defines the Responsible AI Innovation Index (RAII) as a weighted normalized score across $m = 10$ dimensions. For Likert-type items scored from 1 to 5, the mean score of each dimension is normalized to the interval $[0, 1]$.

$$D_{ij} = \frac{1}{q_j}\sum_{k=1}^{q_j} x_{ijk}, \qquad z_{ij} = \frac{D_{ij}-1}{4}, \qquad RAII_i = 100\sum_{j=1}^{10} w_j z_{ij}, \quad \sum_{j=1}^{10} w_j = 1 \tag{1}$$

where D_{ij} is the mean score of respondent i on dimension j, q_j is the number of items in dimension j, x_{ijk} is the observed item score, z_{ij} is the normalized dimension score, and w_j is the assigned weight of dimension j. Equation 1 allows the ten dimensions in Table 2 to be treated as a transparent readiness score rather than as an unstructured maturity opinion.

For EFA, the observed correlation matrix $\mathbf{R}$ is decomposed into common-factor variance and unique variance:

$$\mathbf{R} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\top} + \boldsymbol{\Psi}, \tag{2}$$

where $\boldsymbol{\Lambda}$ is the matrix of factor loadings and $\boldsymbol{\Psi}$ is the uniqueness matrix. This stage is exploratory and should be conducted only with an adequate sample and a correlation matrix suitable for factor analysis.

For CFA, the first-order measurement model can be expressed as:

$$\mathbf{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}, \tag{3}$$

where $\mathbf{x}$ contains observed indicators, $\boldsymbol{\xi}$ represents latent dimensions $D_1, \ldots, D_{10}$, $\boldsymbol{\Lambda}_x$ is the loading matrix, and $\boldsymbol{\delta}$ is the vector of measurement errors. The second-order CFA hypothesis is:

$$D_j = \lambda_j \eta_{RAI} + \epsilon_j, \qquad j = 1, \ldots, 10. \tag{4}$$

Alternatively, a structural model can test whether the broader theoretical constructs explain RAI:

$$\eta_{RAI} = \beta_1 \xi_{CRE} + \beta_2 \xi_{INV} + \beta_3 \xi_{RDI} + \beta_4 \xi_{GOV} + \beta_5 \xi_{VAL} + \zeta. \tag{5}$$

CFA would test whether the hypothesized latent structure fits empirical data using indices such as CFI, TLI, RMSEA, and SRMR, while avoiding mechanical dependence on universal fixed cutoffs because fit indices are sensitive to model and sample conditions (Groskurth et al., 2024).

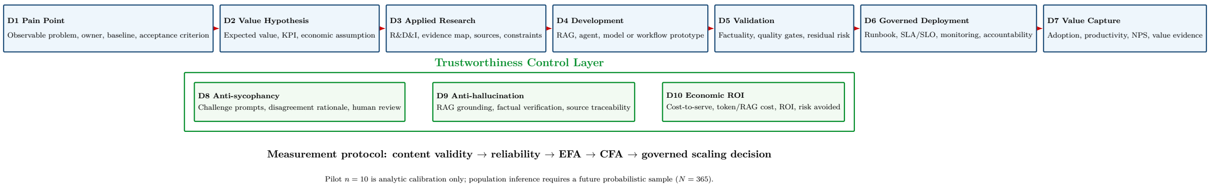

Figure 1: Responsible AI Innovation Measurement Model.

5. Methodology

The current 10-person assessment is not treated as statistical validation. It is a non-probabilistic, purposive, and homogeneous pilot by use case. The purpose of the pilot is analytic calibration: checking item clarity, construct relevance, and operational feasibility. Population inference over $N = 365$ would require a probabilistic sample. Using a finite population formula with 95% confidence, 5% margin of error, and $p = .50$, the future sample size is approximately 187 respondents.

$$n = \frac{N z^2 p(1-p)}{e^2(N-1) + z^2 p(1-p)}. \tag{6}$$

The proposed validation sequence is: (1) expert review and content validity using Aiken's V; (2) reliability analysis using Cronbach's alpha or complementary reliability coefficients; (3) EFA to identify latent dimensions; (4) CFA to test the hypothesized measurement model; and (5) structural testing of the proposed relation among creativity, inventiveness, R&D&I, governance, and value capture. Recent CFA literature emphasizes that model fit must be interpreted according to model conditions rather than as a mechanical checklist (Groskurth et al., 2024).

Table 3: Validation protocol for future empirical testing.

Stage	Purpose	Evidence	Decision use
Content validity	Expert panel evaluates relevance and clarity of each item.	Aiken's V.	Revise or remove weak items.
Reliability	Internal consistency per dimension.	Cronbach's alpha; item-total correlations.	Check redundancy and coherence.
EFA	Explore latent structure.	KMO; Bartlett; loadings; communalities.	Retain factors and refine scale.
CFA	Confirm hypothesized measurement model.	CFI; TLI; RMSEA; SRMR; standardized loadings.	Accept or re-specify model.
Structural model	Test relationships among constructs.	Path coefficients; R^2; indirect effects.	Explain value generation.

6. Illustrative Pilot Scenario

Table 4 reports an illustrative scenario. These values are not empirical results and should not be used as evidence of construct validity. They show how the future instrument would be reported once sufficient data are collected.

Table 4: Illustrative pilot and future validation indicators.

Block	Variables	Expected/target value	Methodological reading
Population and pilot	$N = 365$; $n = 10$.	2.74% coverage.	Pilot only; no population inference.
Future sample	95% confidence; 5% error; $p = .50$.	$n \approx 187$.	Required for probabilistic inference.
Content validity	5 expert judges; 5-point scale.	Target $V \geq .80$.	Formal expert panel needed.
Reliability	10 dimensions; Likert items.	Target $\alpha = .70$-$.90$.	Avoid redundancy if $\alpha > .90$.
EFA	Adequate correlation matrix.	KMO $> .60$; Bartlett $p < .05$.	Proceed only if assumptions hold.
CFA	Hypothesized second-order model.	Report CFI, TLI, RMSEA, SRMR.	Interpret dynamically, not mechanically.

7. Discussion

The proposed model reframes AI innovation as a governed measurement process. This is important because many organizations label generative AI use as innovation without proving problem clarity, factual grounding, risk mitigation, or economic value. The conceptual model states that creativity and inventiveness are necessary but insufficient. R&D&I introduces systematic inquiry, AI governance reduces uncontrolled adoption, and value capture determines whether the implementation qualifies as innovation under an Oslo-type definition.

The main managerial implication is a conditional Go decision. A proof of concept should not be scaled unless it passes discovery, architecture review, evidence traceability, anti-hallucination evaluation, anti-sycophancy challenge, cost baseline, and ROI definition. This position avoids two extremes: methodological paralysis and uncritical deployment. It also gives consulting firms a defensible innovation governance artefact for clients, auditors, and executive boards.

8. Conclusion

This paper converts an academic innovation assignment into a publishable research manuscript by formalizing responsible AI innovation as a measurable latent construct. The proposed RAII model integrates creativity, inventiveness, R&D&I, trustworthy AI governance, RAG-based grounding, risk control, and value capture. The mathematical model provides a scoring mechanism, while EFA and CFA provide a future empirical validation path. The present 10-person pilot must be interpreted as analytic calibration only; a future probabilistic study with approximately 187 participants would be required to make population-level claims for $N = 365$. The main contribution is a measurement and governance model that helps distinguish AI experimentation from responsible, auditable, and value-producing innovation.

Limitations and Future Work

The study is currently conceptual and pilot-oriented. No empirical factor validation is claimed. Future research should collect a probabilistic sample, execute EFA and CFA, test convergent and discriminant validity, and compare the model across sectors such as healthcare, manufacturing, logistics, and public safety.

Declarations

Ethics approval: Not applicable for the current conceptual manuscript; future human-subject assessment should follow applicable institutional ethics requirements.

Data availability: No empirical dataset is reported. Illustrative statistics are methodological examples only.

Conflict of interest: The corresponding author is associated with ARTIFEX Consulting, the case organization.

AI-assisted preparation disclosure: AI-assisted tools were used only for editorial drafting, language refinement, formatting support, and structural organization during the preparation of this manuscript. The authors retain full responsibility for the conceptual contribution, source verification, originality, data accuracy, interpretation, conclusions, and final submission.

References

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement, 40*(4), 955-959. `https://doi.org/10.1177/001316448004000419`

Bartlett, M. S. (1954). A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society: Series B, 16*(2), 296-298. `https://doi.org/10.1111/j.2517-6161.1954.tb00174.x`

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334. `https://doi.org/10.1007/BF02310555`

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv*. `https://arxiv.org/abs/2312.10997`

Groskurth, K., Bluemke, M., & Lechner, C. M. (2024). Why we need to abandon fixed cutoffs for goodness-of-fit indices: An extensive simulation and possible solutions. *Behavior Research Methods, 56*, 3891-3914. https://doi.org/10.3758/s13428-023-02193-3

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *arXiv*. https://arxiv.org/abs/2311.05232

International Organization for Standardization/International Electrotechnical Commission. (2023). *ISO/IEC 42001:2023 Artificial intelligence - Management system*. ISO. https://www.iso.org/standard/42001

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31-36. https://doi.org/10.1007/BF02291575

Khan, M., Ahmed, A., & Sharif, M. M. (2026). AI infused business model innovation for competitive advantage in the era of big data and digital transformation. *Open Access Journal of Data Science and Artificial Intelligence, 4*(1). https://doi.org/10.23880/oajda-16000165

Medwin Publishers. (2025). *Open Access Journal of Data Science and Artificial Intelligence (OAJDA)*. ISSN 2996-671X.

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

National Institute of Standards and Technology. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (NIST AI 600-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.600-1

OECD. (2015). *Frascati manual 2015: Guidelines for collecting and reporting data on research and experimental development*. OECD Publishing. https://doi.org/10.1787/9789264239012-en

OECD/Eurostat. (2018). *Oslo manual 2018: Guidelines for collecting, reporting and using data on innovation* (4th ed.). OECD Publishing. https://doi.org/10.1787/9789264304604-en

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2023). Towards understanding sycophancy in language models. *arXiv*. https://arxiv.org/abs/2310.13548

World Intellectual Property Organization. (2024). *Global innovation index 2024: Unlocking the promise of social entrepreneurship*. WIPO. https://www.wipo.int/global_innovation_index/en/2024/