This paper was presented at the XXXVII ISPIM Innovation Conference, held in Granada, Spain on 07 June to 10 June 2026. ISBN 978-952-65771-2-8.

Developing a Responsible Innovation Capability Evaluation Indicator System for Software Firms

Yuwen Xu*

University of Galway, Galway, H91 HX31, Ireland
Lero, the Research Ireland Centre for Software, Limerick, V94 NYD3, Ireland
E-mail: yuwen.xu@universityofgalway.ie

Suzana Sampaio

Federal Rural University of Pernambuco, Recife 52171-900, Brazil.
E-mail: suzana.sampaio@universityofgalway.ie

Kathryn Cormican

University of Galway, Galway, H91 HX31, Ireland
E-mail: kathryn.cormican@universityofgalway.ie
* Corresponding author

Abstract: As software firms face growing expectations to demonstrate Responsible Innovation (RI), a critical gap remains between high-level theory and implementable practice. This study develops an RI capability evaluation indicator system for software firms grounded in the AIRR framework of anticipation, inclusion, reflexivity, and responsiveness. Drawing on software innovation capability literature, responsibility governance frameworks, and software governance practice literature, the study constructs an initial indicator pool and translates it into observable indicators suitable for later expert review and document-based application. Preliminary findings from Step 1 produce a multi-level framework consisting of four first-level dimensions, twelve second-level dimensions, and thirty-nine third-level indicators. The subsequent stages will refine the indicator set through expert assessment and determine indicator weights using the Analytic Hierarchy Process (AHP). The study contributes a software-specific and AIRR-aligned basis for future RI capability assessment and comparative empirical analysis.

Keywords: Responsible innovation; Evaluation indicator system; Software firms; AIRR

1 Problem

As software firms face growing expectations to demonstrate Responsible Innovation (RI), a critical gap remains between high-level theory and implementable practice. While current research is increasingly grounded in the anticipation, inclusion, reflexivity, and responsiveness (AIRR) framework, software firms lack an operational mechanism to capture and measure RI activities in a consistent, sector-specific way. This absence of practical metrics makes it difficult for firms to systematically assess current actions, compare practices across projects, and identify where capability gaps lie. Consequently,

firms lack the support to translate abstract RI principles into tangible practices that can inform day-to-day development and governance work and enable stepwise capability improvement. Therefore, a software-specific measurement framework aligned with AIRR is needed to support consistent assessment and improvement of RI capability in software firms.

2 Current understanding

RI has been developed as an approach to aligning research and innovation with broader societal values, needs, and expectations (von Schomberg, 2013). Within this broader tradition, RI research provides a widely used general framework that defines anticipation, inclusion, reflexivity, and responsiveness (AIRR) as core dimensions of RI (Stilgoe et al., 2013). Building on this foundation, measurement studies have also treated RI as a firm-level construct, including work that develops and validates RI scales with demonstrated discriminant validity relative to corporate social responsibility and sustainability (Zhang et al., 2023). Firm-level measurement has gained some traction in manufacturing-related settings, where RI orientation has been linked to non-financial reporting practices in SMEs (Intenza et al., 2025).

At the same time, RI is context-dependent. While some themes are shared across industries, required actions and challenges vary by sector (Zhou et al., 2025, Gurzawska, 2021). This issue is especially salient in software firms, where rapid iteration, user-facing deployment, and scalable digital diffusion can produce wide-ranging societal spillovers.

One of the few sector-focused contributions is Chang and Hsieh (2024), who apply the Analytic Hierarchy Process (AHP) to prioritise RI-related indicators for the software service industry. Their framework offers a useful starting point by highlighting relevant categories such as financial capital, capabilities, social capital, and reputational assets. However, their framework is organised around asset domains, which may weaken alignment with AIRR's process dimensions and shift evaluation toward what firms possess rather than how they innovate responsibly. Consequently, this study focuses on software firms and develops a software-specific, AIRR-aligned indicator system that operationalises AIRR dimensions into measurable items and derives indicator weights through structured expert elicitation with explicit decision rules and AHP consistency checks.

3 Research question

How can the AIRR dimensions be operationalised into a measurable and weighted RI capability evaluation indicator system for the software sector?

4 Research design

This study adopts a theory-led and contextualised refinement approach to indicator development. The AIRR framework is first used to define the first-level dimensions, while software innovation capability literature, responsibility governance frameworks, and software governance practice literature are integrated to derive the second- and third-level indicators. The preliminary indicator set is then refined through expert assessment

to retain indicators that are relevant to responsible innovation, appropriate for software firms, and observable through publicly available corporate documents.

In a later application stage, a structured coding manual will be developed to support document-based firm assessment. The manual will specify eligible document types, evidence criteria, and coding rules for each retained indicator, and will be pilot-tested through independent coding by two coders, with inter-coder reliability assessed using weighted Cohen's Kappa. Once validated, the coding manual and AHP-derived weights will be applied to publicly available corporate documents in order to generate firm-level RI capability profiles and support cross-firm comparison.

Figure 1 summarises the scientific approach used in this study, which consists of three sequential steps: indicator construction, key indicator selection, and indicator weighting.

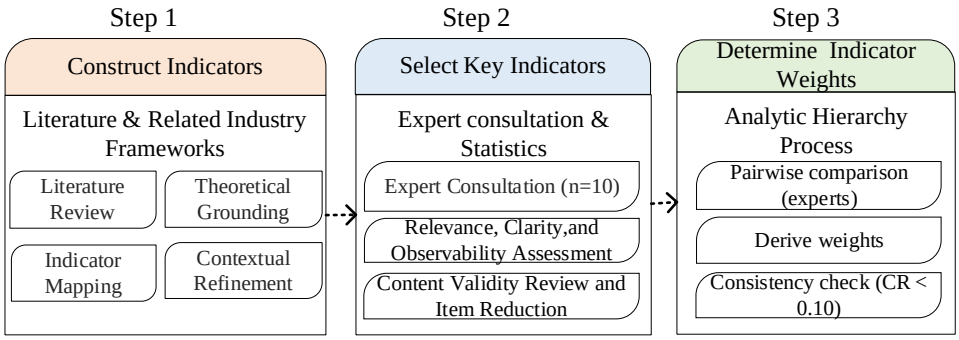

Figure 1 Research design for constructing the RI capability evaluation indicator system

1) Construct indicators. An initial pool of indicators is generated through a structured review and mapping process based on the AIRR framework, informed by software innovation capability literature, responsibility governance frameworks, and selected reporting standards (e.g., AI HLEG, Microsoft, GRI, and SASB). The identified second-level dimensions are then operationalised into observable third-level indicators.

2) Refine and select key indicators through expert assessment. The preliminary indicators are reviewed by experts in order to assess their relevance to responsible innovation, clarity of wording, software-sector suitability, and documentary observability. On this basis, content validity assessment is conducted to revise or remove indicators that are redundant, ambiguous, weakly aligned with the target construct, or difficult to observe through publicly available corporate documents. This step results in a refined set of indicators suitable for subsequent weighting and later document-based application. Experts will be selected through purposive sampling to ensure balanced representation from RI/AI governance research, software development and product management practice, and corporate governance or disclosure-related roles.

3) Determine indicator weights using the AHP. The relative importance of indicators will be determined using AHP. A two-level hierarchy will be constructed, with AIRR dimensions at level one and selected indicators at level two. Experts will provide pairwise comparisons, and weights will be derived from the resulting matrices. Judgment consistency will be assessed using the consistency ratio (CR), with CR < 0.10 indicating acceptable consistency. The outcome of this step is a weighted RI capability indicator

system for software firms, which can subsequently be applied to public documentary evidence to support firm-level scoring and comparison.

5 Findings

Following Step 1, the study preliminarily established a four-dimensional first-level structure consisting of Anticipation, Inclusion, Reflexivity, and Responsiveness. This structure preserves the theoretical logic of responsible innovation while adapting it to the organisational and process context of software firms. At the second level, the preliminary framework includes twelve dimensions: A1 impact and scenario anticipation, A2 risk identification and responsibility assessment, A3 responsible requirements and pre-release planning, I1 stakeholder identification and substantive participation, I2 inclusive and fairness-oriented design, R1 value trade-offs and ethical review, R2 organizational learning and cross-boundary collaboration, R3 review of assumptions, bias, and limitations, RS1 transparency, documentation, and traceability, RS2 continuous monitoring, feedback translation, and corrective iteration, RS3 human oversight and intervention arrangements, and RS4 incident response, accountability, and redress.

At the third level, the framework currently contains thirty-nine publicly evidenced items, focusing on whether relevant mechanisms, processes, or governance arrangements are disclosed, documented, or recorded in publicly available corporate materials. The third-level indicators were formulated in this documentary form because the subsequent stages of the study are intended to support public-document coding, firm-level scoring, and cross-firm comparison. The intended documentary sources include annual reports, ESG or sustainability reports, AI governance reports, technical documentation, model cards, transparency notes, product pages, terms of service, changelogs, privacy policies, and public complaint or appeal channels.

Table 1 presents the preliminary multi-level framework of responsible innovation capability for software firms. The framework consists of four first-level dimensions, twelve second-level dimensions, and thirty-nine third-level indicators. Its purpose is to provide a theoretically grounded and software-contextualised basis for subsequent expert review, content-validity screening, and weighting.

Table 1 Initial indicator framework developed in Step 1

First-level dimension	*Second-level dimension*	*Representative third-level indicators*
Anticipation	A1. Impact and scenario anticipation	intended uses and unsupported uses; target users and indirectly affected groups; high-risk or sensitive scenarios; social, ethical, or privacy impacts
	A2. Risk identification and responsibility assessment	structured risk or impact assessment process; risk categorization; mitigation disclosure; reassessment triggered by updates or external changes
	A3. Responsible requirements and pre-release planning	responsible AI principles or design guidelines; pre-release criteria; formal review or approval mechanisms
Inclusion	I1. Stakeholder identification and	identification of vulnerable or indirectly affected groups; user research or co-

First-level dimension	*Second-level dimension*	*Representative third-level indicators*
	substantive participation	creation; stakeholder engagement mechanisms; pre-deployment communication
	I2. Inclusive and fairness-oriented design	identification of groups facing unfair impact or quality disparities; subgroup fairness testing; fairness-driven design updates
Reflexivity	R1. Value trade-offs and ethical review	disclosure of trade-offs among responsible innovation principles; rationale for governance decisions; ethics or review committees; conditions for pausing or adjusting deployment
	R2. Organizational learning and cross-boundary collaboration	internal review or improvement mechanisms; dedicated responsible AI teams; collaboration with academia or industry bodies; employee training or capability-building programs
	R3. Review of assumptions, bias, and limitations	disclosure of data sources and known biases; system assumptions and applicability conditions; limitations and unsupported conditions; root-cause analysis of errors or bias incidents
Responsiveness	RS1. Transparency, documentation, and traceability	public technical documentation; version histories or changelogs; publicly accessible evaluation or transparency materials
	RS2. Continuous monitoring, feedback translation, and corrective iteration	post-deployment monitoring mechanisms; user feedback channels and handling pathways; feedback-driven product adjustments; re-evaluation triggered by new risks or uses
	RS3. Human oversight and intervention arrangements	governance structure for oversight; override, intervention, or interruption mechanisms; validation of oversight effectiveness
	RS4. Incident response, accountability, and redress	issue reporting procedures; publicly accessible complaint or appeal channels; user redress, remedy, or compensation arrangements

Next, the preliminary indicator framework will be subjected to expert review in order to refine the indicator set and remove redundant, unclear, or weakly observable items. The retained indicators will then be weighted using AHP. These subsequent steps are expected to produce a more parsimonious and expert-validated RI capability indicator system, together with a consistent weighting structure that can support later public-document-based scoring and firm-level comparison.

6 Contribution

This study makes three contributions to RI research. First, it provides a software-specific operationalisation of the AIRR framework by translating its four dimensions into a structured capability indicator system. Second, it introduces a document-oriented indicator design in which third-level indicators are formulated as publicly observable disclosure items rather than self-reported survey items. Third, the proposed framework provides a basis for subsequent weighting and future comparative research on RI capability across software firms.

7 Practical implications

The proposed indicator system offers a structured tool for software firms and related stakeholders to assess RI capability and identify priorities for improvement across the AIRR dimensions. By translating responsible innovation into software-relevant and potentially documentable indicators, the framework can help make RI more actionable in day-to-day development and governance work. In particular, it may support firms in identifying gaps in areas such as stakeholder participation, fairness-oriented design, ethical review, transparency, and corrective response.

The framework may also be useful for policymakers, industry bodies, and standard-setting communities seeking more concrete ways to discuss and benchmark RI capability in software firms. Because the indicators are being designed with later public-document-based application in mind, the framework has the potential to support more comparable and externally reviewable assessments than approaches relying solely on internal self-reporting. In this sense, the study provides not only an analytical structure for future research but also a foundation for more evidence-informed dialogue on responsible software innovation.

References

CHANG, T.-S. & HSIEH, Y.-C. 2024. Applying the analytic hierarchy process for investigating key indicators of responsible innovation in the Taiwan software service industry. *Technology in Society*, 78**,** 102690.

GURZAWSKA, A. 2021. Responsible Innovation in Business: Perceptions, Evaluation Practices and Lessons Learnt. *Sustainability* [Online], 13.

INTENZA, M., TURZO, T., NETTI, A. & MARZI, G. 2025. Linking Responsible Innovation and Nonfinancial Reporting: Evidence From Manufacturing SMEs. *Business Strategy and the Environment*.

STILGOE, J., OWEN, R. & MACNAGHTEN, P. 2013. Developing a framework for responsible innovation. *Research Policy*, 42**,** 1568-1580.

VON SCHOMBERG, R. 2013. A Vision of Responsible Research and Innovation. *Responsible Innovation.*

ZHANG, S. X., CHEN, J., HE, L. & CHOUDHURY, A. 2023. Responsible Innovation: The development and validation of a scale. *Technovation*, 124**,** 102754.

ZHOU, Q., BAI, S., TAN, Y., HAN, C. & YANG, M. 2025. Corporate Responsible Innovation for Sustainability: Toward Firm-Level Measurements for Process and Outcomes. *Thunderbird International Business Review*.

This paper was presented at the XXXVII ISPIM Innovation Conference, held in Granada, Spain on 07 June to 10 June 2026. ISBN 978-952-65771-2-8.

Areas for feedback & development

Feedback is especially welcome on two aspects of the study. First, we invite comments on the coverage, distinctiveness, and software-sector relevance of the proposed indicator system, including whether any second- or third-level indicators appear missing, redundant, or insufficiently adapted to the realities of software firms. Second, we would particularly value feedback on the validity and limitations of using publicly available corporate documents as evidence of RI capability, including whether this approach provides a credible basis for firm-level comparison and where it may risk conflating disclosure quality with actual organisational capability.