

DOCTORAL CONSORTIUM

# CO-CREATING HUMAN-CENTRED AND RESPONSIBLE AI GOVERNANCE IN CONSULTING SMEs

LUDMILA JIŘÍČKOVÁ

Prague University of Economics and Business, Faculty of Informatics and Statistics,  
Prague, Czechia  
[ludmila.malinova@vse.cz](mailto:ludmila.malinova@vse.cz)

Small and medium-sized enterprises (SMEs) in consulting increasingly rely on AI tools in client work, yet often lack proportionate and human-centred governance approaches that translate high-level “responsible AI” principles into everyday organisational practice. This doctoral research proposes AIROF – an adaptive AI Risk & Opportunity Framework – grounded in fourth-order cybernetics to enable co-created governance through explicit roles, feedback loops, and continuous contextual adaptation. Using a Design Science Research approach, the study synthesises theoretical foundations (AI governance, risk management, Balanced Scorecard logic, fourth-order cybernetics) and triangulates them with empirical insights in consulting SMEs to design and validate AIROF. The planned instantiation (AIROF Scorecard) will support decision-making without requiring disruptive process redesign. This paper contributes a research agenda and evaluation plan for co-creating human-centred, responsible digital futures in resource-constrained knowledge organisations.

DOI

[https://doi.org/  
10.18690/um.fov.4.2026.80](https://doi.org/10.18690/um.fov.4.2026.80)

ISBN

978-961-299-152-4

**Keywords:**

responsible digital futures,  
human-centred design,  
SMEs,  
cybernetics,  
risk and opportunity  
management,  
design science research



University of Maribor Press

## 1 Introduction

The 2026 conference theme “Co-Creating Human-Centred and Responsible Digital Futures” highlights a core challenge of organisational AI adoption: digital futures become “responsible” not only through technology choices, but through governance arrangements that keep humans meaningfully in the loop and enable participation, accountability, and learning over time (*European Commission, 2021; UNESCO, 2021*).

In consulting SMEs, AI tools (including generative AI) can accelerate knowledge work and augment service delivery, but they also introduce risks that are tightly coupled with client trust, confidentiality, and professional judgement (*Davenport & Mittal, 2023; Kolbjørnsrud, 2024*). The tension is intensified by rapid AI capability evolution, which compresses governance adaptation cycles from years to months (*Russell & Norvig, 2021; Haenlein & Kaplan, 2019*).

This doctoral research addresses the problem by proposing **AIROF (Artificial Intelligence Risk & Opportunity Framework)**: an adaptive, systemic meta-framework for governing AI use through co-creation (roles, feedback loops, shared sense-making), grounded in fourth-order cybernetics as a theory of context-aware self-governance and continuous learning.

### 1.1 Research aim and positioning

The dissertation’s aim is to identify theoretical foundations for an adaptive framework to manage AI risks and opportunities in consulting SMEs (*Nonaka & Takeuchi, 1995; Davenport & Mittal, 2023*), and to design a practical instrument (AIROF Scorecard) to support adoption and evaluation in at least 15 SME organisations.

This Doctoral Consortium paper focuses on the **conceptual and methodological** contribution (AIROF framing, research questions, and evaluation plan) aligned to the conference theme.

## 2 Problem definition

### 2.1 Practical problem

SMEs face disproportionate constraints in AI adoption: limited governance capacity, limited specialist roles, and increased dependency on external AI providers (*Sánchez et al., 2025; Tominc et al., 2024*). Yet they operate in environments where responsible AI practice requires clear accountability, risk controls, and transparency. Recent OECD evidence shows persistent adoption gaps between SMEs and large firms, and highlights enabling factors such as skills, inputs, and finance implying that governance instruments must be lightweight, implementable, and adapted to SME realities (*OECD, 2024; OECD, 2019*).

In parallel, formal governance frameworks and regulations are expanding. The **EU AI Act** establishes harmonised rules for AI, reinforcing the need for structured governance and risk management across the AI lifecycle (*European Commission, 2024*).

The **NIST AI RMF** provides a voluntary framework to manage AI risks to individuals, organisations, and society (*NIST, 2023*).

**ISO/IEC 42001** introduces an AI management system standard intended to structure organisational governance of AI risks and opportunities (*ISO/IEC 42001, 2023*).

However, SMEs – especially in knowledge-intensive services—still require a **contextual, human-centred operational bridge** from these high-level expectations to everyday decision-making and learning cycles.

### 2.2 Research gap and theoretical framing

The dissertation hypothesises that **fourth-order cybernetics** provides a suitable meta-framework for adaptive AI governance because it emphasises context, systemic integration, and the human actor's reflexive role in steering socio-technical systems (*Umpleby et al., 2019; Medvedeva & Umpleby, 2024*).

This perspective naturally aligns with co-creation: governance is not a one-time policy artefact but an evolving set of relationships, roles, and feedback loops that shape responsible digital futures over time (*Božičnik & Mulej, 2011; Chiolerio, 2020*).

### 2.3 Research questions

The dissertation defines one main and four sub-questions:

**RQ0 (main):** How can a conceptual framework AIROF be designed and empirically validated for the adaptive management of risks and opportunities of the use of AI tools in consulting SMEs, using fourth-order cybernetics as a meta-framework?

**RQ1:** Which concepts and principles from fourth-order cybernetics, AI governance, risk management, and the Balanced Scorecard are relevant and sufficiently robust for the design of an adaptive framework for managing risks and opportunities of AI in consulting SMEs?

**RQ2:** What are the key needs, expectations, and problem areas associated with the use of AI tools in consulting services SMEs?

**RQ3:** How can a practical tool (AIROF Scorecard) be designed, based on theoretical foundations and empirical findings, using the logic of the Balanced Scorecard, in order to support the management of risks and opportunities of AI in consulting SMEs?

**RQ4:** What value does the implementation of the AIROF Scorecard bring to consulting organisations in the management of risks and opportunities of AI, in terms of reducing the impact of risks, increasing the utilisation of opportunities, improving transparency, and supporting decision-making?

### 2.4 Contribution statement

This doctoral research aims to contribute mainly two artefacts:

A conceptual contribution: AIROF as an adaptive, cybernetically grounded governance meta-framework that treats responsible AI as a co-created

socio-technical capability (roles, relationships, and feedback), rather than a static control checklist.

A methodological contribution: a Design Science Research process and evaluation plan tailored to consulting SMEs, connecting theory synthesis with empirical validation and a lightweight artefact instantiation (*Hevner et al., 2004; Peffers et al., 2007*).

A practical contribution: a proportionate governance approach that can complement established frameworks (EU AI Act, NIST AI RMF, ISO/IEC 42001) while remaining implementable in resource-constrained organisations as an operationalised form of AIROF through the AIROF Scorecard tool.

### **3 Methodology**

#### **3.1 Research approach**

The dissertation is positioned within a Design Science Research logic, connecting theory, empirical insights, and artefact design through iterative cycles. This is consistent with the Doctoral Consortium's focus on the "research journey" and on sharpening a proposal through feedback and planned validation.

#### **3.2 Design Science Research process and phases**

The dissertation structures the work into phases that progress from diagnosis to conceptual design, artefact creation, implementation, and reflexive evaluation.

**Table 1: Research phases, methods, and outputs (planned)**

| Phase                    | Purpose                                                              | Main methods                                                                                                    | Key outputs                                                            |
|--------------------------|----------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| Diagnostic               | Define AI governance and risk–opportunity problem in consulting SMEs | Literature synthesis; initial SME probing                                                                       | Problem definition; requirements for a lightweight governance approach |
| Conceptual design        | Construct AIROF framework                                            | Theory integration (4th-order cybernetics; AI governance; risk management; Balanced Scorecard); empirical input | AIROF constructs; co-creation logic and role design                    |
| Artefact (instantiation) | Create AIROF Scorecard concept/prototype                             | Scorecard modelling; tool prototyping (Excel/VBA)                                                               | Initial tool concept for structured risk/opportunity monitoring        |
| Implementation           | Deploy in SMEs                                                       | Onboarding + training; guided adoption                                                                          | Tool use in real organisational context                                |
| Reflexive evaluation     | Evaluate contribution and adjust                                     | Post-use surveys; follow-up interviews; triangulation                                                           | Evidence on perceived value; refinements; design knowledge             |

Source: Author's own work

### 3.3 Empirical strategy and evaluation design

The dissertation proposes triangulated validation methods, including semi-structured interviews with management and key staff, the analysis of organisational documents where available (e.g., internal policies), and post-implementation questionnaires after approximately one month of tool usage. It also proposes evaluation across 15 SME size categories (five micro, five small, five medium) within the consulting domain.

This evaluation plan aligns with a human-centred and responsible digital futures frame in two ways:

First, it treats governance effectiveness as **experienced** by human actors (usability, decision support, transparency). Second, it explicitly anticipates adaptation based on feedback, reflecting cybernetic learning rather than one-off compliance exercises (*Argyris & Schön, 1978*).

## 4 Preliminary and expected results

### 4.1 AIROF as a co-creation governance loop

The preliminary outcome is the conceptual structure of AIROF: an adaptive framework integrating AI governance principles, risk and opportunity management, and a scorecard logic inspired by Balanced Scorecard (*Kaplan & Norton, 1996; Kaplan & Norton, 2001*) to keep governance actionable and multi-perspective.

To maintain proportionality for SMEs, AIROF is defined to support decision-making without forcing immediate redesign of core business processes.

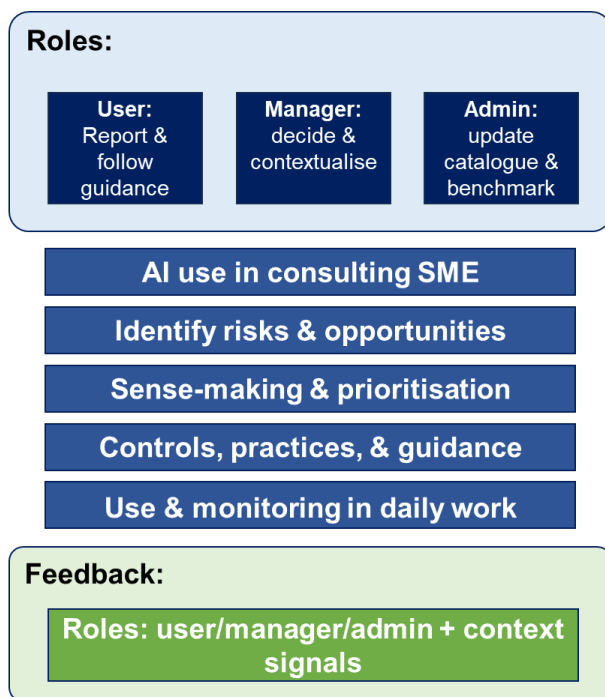


Figure 1: AIROF co-creation loop (conceptual view)

Source: Author's own work.

This representation directly reflects the dissertation's emphasis on adaptive governance cycles, explicit roles (user/manager/admin), and the possibility of cross-organisational learning via benchmarking.

## 4.2 AIROF Scorecard positioning

The dissertation introduces AIROF Scorecard as the practical instrument to operationalise AIROF through four perspectives and three access/interaction roles, implemented as a lightweight tool (prototype in spreadsheet form).

In this Bled Doctoral Consortium paper, the Scorecard is treated as the planned instantiation and evaluation vehicle; the detailed operational design (metrics, UX, automation, and governance workflows) is under preparation as the next steps of the dissertation.

## 4.3 Expected evaluation outcomes and validation criteria

The final evaluation will test whether AIROF (and its instantiation) improves working with AI tools in organisational practice. The focus is on consulting SMEs:

- **Transparency:** clearer shared understanding of AI risks/opportunities and responsibilities.
- **Decision support:** better prioritisation and management attention in AI adoption choices.
- **Risk impact reduction and opportunity realisation:** evidence via post-use feedback and benchmarking.
- **Adaptability:** the ability to update the governance catalogue and practices as AI capabilities and regulatory expectations evolve (e.g., in response to the EU AI Act and organisational learning).

# 5 Future development

## 5.1 PhD work plan and timeline

Figure 2 outlines the overall research design of the dissertation. The theoretical background – grounded in fourth-order cybernetics, AI governance, risk management, and the Balanced Scorecard – has been completed. The subsequent phases, including empirical data collection among consulting SMEs, the design of the AIROF framework and its scorecard tool, as well as validation through case studies, are currently in progress.

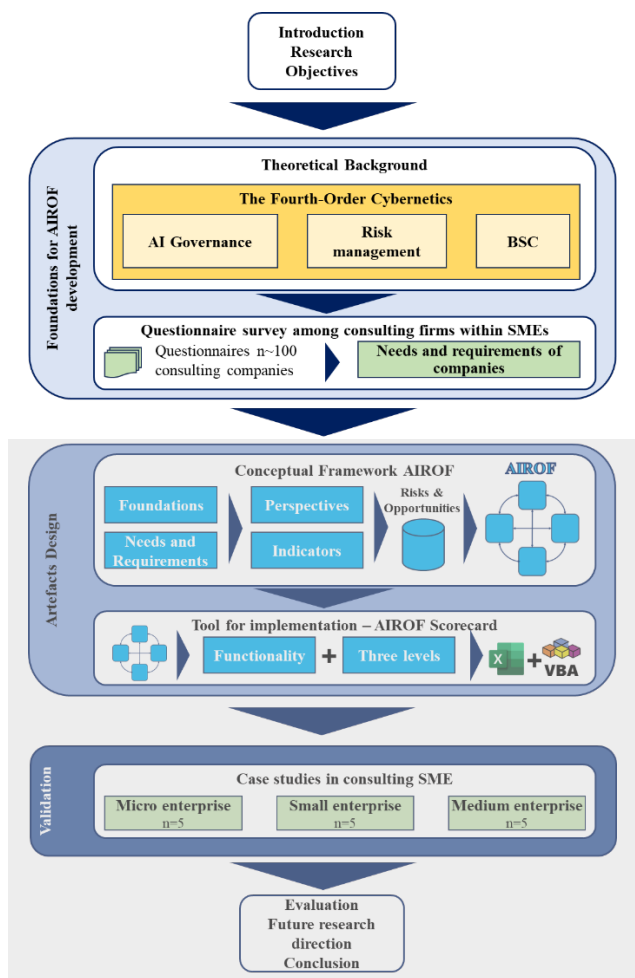


Figure 2: Remaining PhD work (high-level timeline)

Source: Author's own work

## 5.2 Ethical considerations and responsible research practice

The research involves human participants in SMEs (interviews, surveys) and therefore requires informed consent, careful handling of potentially sensitive organisational information, and clear separation of client data from research data. The research also concerns AI governance, which intersects with risk to individuals and society – consistent with the emphasis in NIST AI RMF on managing AI risks to individuals, organisations, and society.

AI governance in Europe is also a moving regulatory target; aligning AIROF's constructs with the EU AI Act's lifecycle and risk-based logic is part of ensuring that the designed governance approach supports responsible digital futures rather than encouraging superficial compliance (*Floridi & Cowls, 2021; Hagendorff, 2020*).

Similarly, the OECD's values-based principles for trustworthy AI provide a relevant normative baseline for "human-centred" co-creation.

### 5.3 Planned dissemination

This Doctoral Consortium paper will be used to refine the research design through feedback. The next phase will focus on the AIROF Scorecard operationalisation and implementation details (measurement model, tool features, and evaluation outcomes), thereby separating conceptual framing from artefact-level contribution.

#### Disclosure of AI use

In the preparation of this manuscript, the author used ChatGPT-5, a generative AI tool developed by OpenAI, for translation and language correction when translating text from Czech to English. The tool was used solely to improve linguistic clarity and ensure accurate translation (with author's corrections afterwards), without generating or altering content. All intellectual and conceptual work remains the responsibility of the author.

#### References

- Argyris, C., & Schön, D. (1978). *Organizational learning: A theory of action perspective*. Addison-Wesley.
- Božičnik, S., & Mulej, M. (2011a). A new—fourth order cybernetics and sustainable future. *Kybernetes*, 40(5/6), 670–684. <https://doi.org/10.1108/03684921111142232>
- Davenport, T. H., & Mittal, N. (2023). *All-in on AI: How smart companies win big with artificial intelligence*. *Harvard Business Review Press*.
- European Commission. (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*.
- European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. *Official Journal of the European Union*, L 1689. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Floridi, L. (2021). *Ethics, governance, and policies in artificial intelligence*. Springer.
- Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61(4), 5–14. <https://doi.org/10.1177/0008125619864925>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://www.jstor.org/stable/25148625>

- International Organization for Standardization, & International Electrotechnical Commission. (2023). ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system. ISO/IEC.
- Kaplan, R. S., & Norton, D. P. (1996). The balanced scorecard. Harvard Business School Press.
- Kaplan, R. S., & Norton, D. P. (2001). Strategy-focused organization. Harvard Business School Press.
- Kolbjørnsrud, V. (2024). Designing the intelligent organization: Six principles for human–AI collaboration. *California Management Review*, 66(2), 44–64. <https://doi.org/10.1177/00081256231211020>
- Medvedeva, T. A., & Umpleby, S. A. (2024b). Cybernetics of large social systems: The example of economics. *Systems Research and Behavioral Science*, 41(3), 524–533. <https://doi.org/10.1002/sres.2991>
- 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://www.nist.gov/itl/ai-risk-management-framework>
- Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford University Press.
- Organisation for Economic Co-operation and Development. (2019). OECD principles on artificial intelligence. OECD Publishing. <https://www.oecd.org/going-digital/ai/principles/>
- Organisation for Economic Co-operation and Development. (2024). Update on the implementation of the OECD AI principles. OECD Publishing. <https://oecd.ai/en/ai-principles>
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Russell, S. J., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
- Tominc, P., Oreški, D., Čančer, V., & Rožman, M. (2024). Statistically significant differences in AI support levels for project management between SMEs and large enterprises. *AI*, 5(1), 136–157. <https://doi.org/10.3390/ai5010008>
- Umpleby, S. A., Medvedeva, T. A., & Lepskiy, V. E. (2019). Recent developments in cybernetics, from cognition to social systems. *Cybernetics and Systems*, 50(4), 367–382. <https://doi.org/10.1080/01969722.2019.1574326>

## About the author

**Ludmila Jiříčková** is a doctoral researcher at the Department of Systems Analysis, Faculty of Informatics and Statistics, Prague University of Economics and Business, Czech Republic. Her research focuses on AI governance, systemic approaches to managing risks and opportunities of artificial intelligence, and the application of fourth-order cybernetics in organisational contexts. She combines academic research with practical experience in consulting, digital business, and organisational development. Her work emphasises human-centred and adaptive governance frameworks for SMEs, particularly in knowledge-intensive services. She is currently developing the AIROF framework and its practical implementation tool, the AIROF Scorecard, within a Design Science Research paradigm.

## Summary

This paper addresses the challenge of translating high-level AI governance principles into practical, human-centred approaches suitable for consulting SMEs. It introduces AIROF (Artificial Intelligence Risk & Opportunity Framework), an adaptive meta-framework grounded in fourth-order cybernetics, enabling co-created governance through feedback loops, explicit roles, and continuous learning. The research adopts a Design Science Research methodology, integrating theoretical foundations from AI governance, risk management, and Balanced Scorecard logic with empirical insights from SMEs. The planned AIROF Scorecard serves as a lightweight implementation tool supporting decision-making without requiring disruptive process changes. The study contributes a conceptual framework, a

methodological approach, and a practical governance solution aligned with responsible digital futures in resource-constrained organisations.