

Co-Creating Operational AI Governance

Giorgio Zampirolo,^{1*} Daria Rudnik², Felicia Burtscher³

¹Independent Researcher, London, United Kingdom; ²Independent Consultant, Tel Aviv, Israel; ³Independent Researcher and Consultant, Berlin, Germany

*Corresponding author: giorgio.zampirolo@gmail.com

ABSTRACT

Formal principles, assigned roles, and controls are necessary components of Artificial Intelligence (AI) governance, but organisations may still struggle to translate them into decision rights and escalation arrangements. This article presents a co-creation protocol designed to support cross-functional actors in converting regulatory and organisational requirements into explicit decision rights, escalation pathways, and revisable artefacts. Design requirements were informed by a prior survey of European AI practitioners; the protocol was developed from professional consulting experience across successive versions used in three engagements. Its current version was completed once, with two documented adaptations. Using process evidence, the study formulates two provisional propositions and identifies further untested design questions. The paper reports observed enactability in one completed, adapted application rather than validated effectiveness, and situates the protocol relative to the EU AI Act and ISO/IEC 42001 as a translation mechanism, not a compliance substitute.

Keywords: Artificial Intelligence; Governance; Co-Creation; Design Science; Experimental Innovation.

Received: February 2026. Accepted: August 2026.

1. INTRODUCTION

Despite the proliferation of AI-governance frameworks, principles, and regulatory instruments, organisations continue to report difficulty translating high-level requirements into operational decision rights, escalation pathways, and accountability arrangements. A cross-national survey found that professionals across European organisations frequently identified unclear ownership, fragmented coordination, and uneven governance literacy as barriers to implementation (Zampirolo et al., 2026). Where relevant knowledge and authority are distributed across functions, operational governance has characteristics of a wicked problem: no single actor holds all relevant knowledge, even though particular actors may hold formal authority over defined issues (Rittel & Webber, 1973; Head & Alford, 2015). In such settings, formal specification alone may be insufficient to produce workable operational arrangements.

This motivates one research question: *how can a structured co-creation protocol be specified and enacted to support cross-functional translation of AI-governance requirements into explicit, revisable decision rights and escalation arrangements, and what implementation contingencies emerge during its formative use?* This paper presents such a protocol, its design requirements derived from the survey above and translated into a facilitation procedure using the authors' professional

experience. It reports successive versions used across the complete set of three organisational engagements undertaken to date, with the current version, Version 3, completed once and with two documented adaptations; earlier, partial engagements informed its development (Section 5). The paper reports two propositions grounded in process evidence and further untested design questions. Because client-specific outputs remain the property of participating organisations, the paper claims preliminary protocol feasibility and formative design knowledge, not validated organisational effectiveness.

2. AI-GOVERNANCE AND REGULATORY CONTEXT

Contemporary scholarship conceptualises responsible AI governance through structural, relational, and procedural organisational practices (Papagiannidis et al., 2025) and identifies gaps between principles and operational application (Birkstedt et al., 2023; Batool et al., 2025). Work on participatory AI highlights the difficulty of involving diverse actors where power and expertise are unevenly distributed (Birhane et al., 2022); recent studies co-design governance artefacts with practitioners (Bogucka et al., 2024) or examine how organisations operationalise responsible-AI commitments (Ruster & Daniell, 2025; Mökander & Floridi, 2023). The studies reviewed here provide governance frameworks, auditing procedures, impact-



assessment artefacts, and systems-oriented implementation processes, but do not specify a replicable participant-level facilitation procedure for negotiating contested decision rights.

This gap sits alongside, not in competition with, formal requirements. Article 17 of the EU AI Act requires high-risk AI system providers to maintain a documented quality-management system with an accountability framework assigning responsibilities across the organisation (European Parliament and Council, 2024); ISO/IEC 42001 complements this with a broadly applicable AI management system standard (ISO/IEC, 2023).

These instruments establish formal governance requirements and management-system expectations, while the protocol proposed here addresses a different level of specification: a participant-level facilitation sequence for translating such requirements into organisational decision arrangements. The protocol here is a meso-level mechanism for that translation: its outputs (decision-rights maps, escalation pathways, versioned artefacts) may support the accountability, documentation, and review functions associated with these instruments, without themselves constituting compliance.

3. FRAMEWORK PROVENANCE AND RESEARCH DESIGN

This study draws on design science and action design research, treating artefacts and design principles as contributions arising from iterative cycles of design, use, and reflection (Hevner *et al.*, 2004; Sein *et al.*, 2011; Peffers *et al.*, 2007), and on Romme's view that organisations can themselves be designed (Romme, 2003; van Aken & Romme, 2012). Co-creation is used here in a deliberately restricted sense: intra-organisational, cross-functional co-design among actors holding decision-relevant knowledge or authority. This is narrower than stakeholder-oriented co-creation models that include both decision-makers and people affected by the problem (Bammer, 2019); employees affected by AI-supported decisions, service users, and external stakeholders fall outside the present protocol.

Critical social-design scholarship warns that participatory processes do not automatically produce beneficial outcomes and may obscure questions of power, inclusion, and whose interests are served (von Busch & Palmås, 2016). The protocol responds partly by making decision authority and residual disagreement explicit, but this does not remove the effects of participant selection, hierarchy, or organisational mandate.

The protocol's initial design requirements were informed by an online survey of 70 professionals across seven European countries, combining descriptively analysed closed-response items with thematically

analysed open-ended responses (Zampirolo *et al.*, 2026). Recurring findings translated into design choices: unclear ownership informed the decision-rights component, fragmented coordination informed the translation phases, and uneven governance literacy informed the structured elicitation and shared-vocabulary activities, using the authors' consulting experience, prior to the protocol's first application.

The protocol's three applications (the complete set to date) were examined using a retrospective qualitative multiple-case design; the analytical corpus closed on 15 July 2026. Each engagement was reconstructed as a bounded case and compared against the protocol version specified before it began [Version 1, 2, or 3 respectively (Section 5)] rather than against its current form. Artefacts were assessed for completeness, correspondence with the record, traceability to participant inputs and decisions, internal coherence, and relevance to the stated dilemma. Cross-case interpretations required comparable, evidence-traceable observations; recurrence alone was insufficient. The corpus comprised facilitator reports, formal communications, documented reflections, and formally versioned or circulated artefacts and outcome comparisons; informal, uncirculated drafts were excluded as insufficiently attributable to a decision point, though artefact versions revised during a session (including the pre- and post-amendment states in Section 5) are corpus material. No session recordings were used in the analysis; recordings existed for Application 3 but were not authorised for research access. Propositions were developed iteratively and tested against negative or qualifying evidence, each reported with an explicit evidential status (Section 6); full corpus and procedure detail appear in Supplementary Material A.

All authors reviewed each case and made significant methodological decisions collaboratively, distinguishing documented observations from researcher interpretation; this is collaborative interpretation through negotiated agreement, not inter-rater reliability, since no independent coding agreement was statistically assessed. The authors' roles spanned protocol development, facilitation, documentation, and analysis, creating confirmation-bias risks that this procedure limits rather than eliminates. Generative AI (Claude, Anthropic) assisted with drafting, literature discovery, and preliminary citation checking; all citations and source-dependent claims were independently checked by the authors against original sources, who take full responsibility for all content, claims, and citations. Ethics and data-governance considerations appear under Ethics and Research Governance, below.

4. CO-CREATION PROTOCOL

The protocol is designed for a cross-functional group of six to twelve participants spanning technical, operational, legal/compliance, and leadership

perspectives, convened around a bounded AI-governance dilemma. It comprises a preparatory step, three facilitated sessions, and a separate individual review; facilitator-level detail is in Supplementary Material A.

Session 0 (Commitment and Framing) distributes the governance dilemma and preparatory material in advance and secures each participant's explicit commitment to attend all subsequent sessions. The organisation designates an executive sponsor, activated in Session 3, and a final decider with authority to approve the arrangement; the roles may coincide but this must be confirmed, not presumed.

Session 1 (Divergence) surfaces differing interpretations through de-identified individual pre-work and a structured small-group rotation (adapting Lipmanowicz & McCandless, 2014) giving every participant a structured opportunity to contribute before whole-group discussion opens, closing with a facilitator-led, participant-validated capture of the divergence map. De-identification is a safeguard, not a guarantee, since role-specific content in a small group may still identify a participant.

Session 2 (Interpretive Coordination) uses a further structured rotation to make role-specific interpretations mutually legible and establish shared vocabulary before assigning responsibilities, and includes a compliance-review segment adapting the Input/Agree distinction from Rogers and Blenko's (2006) RAPID framework. The Agree role is assigned before the session to whichever function holds recognised authority over the relevant concern, Legal/Compliance in Application 3, though other functions could equally hold it. Most items are marked advisory input, but items within the Agree-holder's remit are marked Agree (a bounded veto, not the protocol's ultimate authority) such that the artefact cannot proceed to final approval while unresolved; where the Agree-holder is unavailable, the item may proceed provisionally pending sign-off (Supplementary Material A).

Session 3 (Convergence and Provisional Approval) produces a session-approved provisional artefact, typically a decision-rights map and escalation pathway, with named element owners for individual decision items and a resourcing commitment. The Agree-holder then conducts a separate individual assurance review within three to five business days, confirming or amending the artefact; material amendments are recirculated with a rationale before the final decider approves and the executive sponsor embeds the artefact, with a review date, into an existing forum (Supplementary Material A).

5. THREE FORMATIVE APPLICATIONS

Table 1. The protocol's three formative applications, the complete set to date, February–July 2026; fidelity and feasibility detail appear in Supplementary Material A.

	Application 1	Application 2	Application 3
Period	Feb 2026	Jul 2026	Jul 2026
Context	Southern European professional-services org, ~50 employees; 7 participants, in person	North American nonprofit (~22 core staff, ~100 volunteers); 5 of 6 scheduled attended; online	Southern European project-based industrial org; ~450 employees; two groups (8, 9 participants), 17 at Session 0; in person
Functions represented	Not confirmed	Leadership; operational coverage not confirmed	Technical, operational, legal/compliance, leadership (continuing group)
Protocol version	V1: single-sitting; no Session 0	V2: informal Session 0; no decision right/sponsor/async	V3, two documented adaptations (Suppl. Material A)
Stages completed	Divergence overran; Coordination rushed; Convergence not reached	Session 1 (Divergence) completed by corpus close (15 Jul 2026); Session 2 neither scheduled nor conducted	All stages completed, continuing group (n=8); second group undecided
Key finding	Dominance by a few participants absent a planned opening	Operational concerns raised; leadership clarified non-negotiables	Escalation item under Agree right; structural amendment; further change at assurance review
Change following	Motivated a designed Session 0	Further sessions anticipated but unscheduled at corpus close; no later evidence existed	Not applicable: the most recent application

Table 1 summarises the three engagements. Across them, the design of the protocol's opening step (Session 0) shows a graded pattern: absent, then present but undesigned, then fully specified. The facilitation team's records associate the third application's designed opening with more complete preparation than the second, though this is confounded with the concurrent shift from online to in-person delivery (Section 7). The most fully documented process episode is Application 3's Session 2: an escalation-trigger issue tied to a previous operational incident was flagged under the Agree right. Rather than being logged as one input among several, it required a structural amendment before the artefact could proceed toward final entry; the subsequent assurance review then produced a further, separate change rather than confirming the artefact unchanged.

6. DESIGN PROPOSITIONS AND PROSPECTIVE QUESTIONS

Consistent with design-science guidance on formulating context-bounded design propositions (Romme, 2003), the paper reports two propositions grounded in process evidence and separately identifies related design questions not yet tested. The associated boundary conditions are reported explicitly, consistent with the view that reusable design knowledge is context-dependent rather than universally applicable (Chandra Kruse, 2025). The first proposition draws on a primary episode in Application 2 plus a contrast in Application 1; the second rests on a single episode in Application 3. A summary of the documentary basis appears in Table 2; boundary conditions below are hypothesised, not observed, since each rests on a single case.

Translation between operational concerns and strategic constraints. Within a leadership group holding uneven access to operational information, structured exchange may help make operational concerns and strategic constraints mutually visible, particularly where early participation is protected from hierarchical dominance through structured turn-taking and participants hold sufficient authority to engage substantively. In Application 2, operational concerns were raised within the group and met with clarified non-negotiables in return; Application 1, lacking a designed opening, instead saw a few participants dominate, though the applications differ in organisation, delivery mode, and protocol maturity, so this contrast does not isolate the effect of the opening structure. Status: single-case process observation with a non-equivalent cross-case contrast; distinct operational representation within the group was not separately confirmed (Table 1).

Formally assigned decision authority and traceable artefact amendment. Where a contested governance concern falls within a formally assigned Agree right, requiring its resolution before final approval may produce a traceable structural amendment. Application 3's escalation-trigger episode (Section 5) is the evidential basis; Applications 1 and 2 did not reach this stage. Status: single-case process observation, not yet demonstrated across multiple applications.

Related design questions, not yet tested in any application:

Spacing and reconsideration: whether spacing creates a genuine opportunity for reconsideration or merely a scheduling gap; revisions between sessions are suggestive but do not establish incubation as causal (Sio & Ormerod, 2009).

Durable institutionalisation: whether institutionalisation commitments produce durable continuity, not yet observed.

Granularity of authority: whether a bounded Agree right is granular enough, or whether legal interpretation, compliance approval, and executive risk acceptance warrant separate roles (Supplementary Material A).

Co-creation and formal authority: the protocol distinguishes a co-created provisional artefact from the formally authorised artefact that results after assurance review. Co-creation therefore does not, by itself, determine the final organisational record; formally designated roles retain that authority, with amendments traceable and recirculated to participants.

Table 2. Evidential basis and status of the two directly supported propositions.

Proposition	Documentary basis	Observed mechanism	Contrary / limiting evidence	Evidential status
Translation between operational concerns and strategic constraints (Application 2)	Facilitator report, engagement communications, follow-up reflections	Operational concerns raised within the participating group; leadership clarified non-negotiables in return	Application 1, lacking a designed opening, saw a few participants dominate instead (non-equivalent contrast)	Single-case observation; operational representation not separately confirmed
Formally assigned decision authority and traceable amendment (Application 3)	Facilitator report, artefact versions, follow-up review, engagement communications	Escalation-trigger item flagged under an explicit Agree right; required a structural amendment before the session-approved artefact could proceed; a separate assurance review produced a further change	Applications 1 and 2 did not reach this stage	Single-case process observation

7. VALIDITY, LIMITATIONS, AND CONCLUSION

The validity claim is limited to protocol specification and preliminary process feasibility, assessed against whether planned stages could be enacted, required roles were represented, a decision-rights artefact was produced, documented concerns produced traceable changes, and outputs corresponded to the stated dilemma. Only Application 3 met all five criteria; Applications 1 and 2 primarily informed development. The protocol is operationally specified to support future replication attempts; independent procedural replication has not yet been demonstrated. Supplementary Material A states this claim in the same terms, operationalised through a fidelity log and feasibility matrix.

This analysis covers only three engagements and cannot establish saturation, effectiveness, prevalence, or universal applicability; two of three involved Southern European organisations, limiting contextual variation. The second application remained incomplete at corpus closure (further sessions were envisaged but none scheduled or conducted by 15 July 2026, so its contribution is limited to the first session, not a selectively excluded later stage) and its online delivery and lesser protocol maturity are confounded, so comparisons involving it are suggestive rather than isolating a single cause. Version 3's two documented departures from specification are logged in Supplementary Material A's fidelity log. Application 3 is treated as one organisational case with two embedded groups; complete evidence derives only from the group that continued, and a selection effect cannot be ruled out. Reliance on formal, attributable records may also underrepresent informal dissent, so the corpus better supports reconstructing formalised decisions than capturing interpersonal dynamics.

The authors had access to one post-session review log, but the client did not authorise its reproduction or public sharing; independent readers cannot directly inspect that part of the evidence chain. The research team's dual role in developing, applying, and analysing the protocol creates confirmation-bias risks that Section 3's procedure limits rather than eliminates, further disclosed in the Declaration of Competing Interests, below.

This study offers a facilitator-ready co-creation protocol, an account of its development and one completed enactment of its current version, and two propositions grounded in process evidence alongside further design questions. The evidence supports preliminary claims about enactability and traceable artefact revision, but not organisational effectiveness, durability, or comparative superiority. Future work should pursue independent replication using prospectively defined feasibility indicators, participant or organisational validation, assessment by researchers not involved in facilitation, and testing in online delivery.

DECLARATION OF COMPETING INTERESTS

The authors' organisations delivered all three applications on a consulting basis, and maintained a separate, ongoing consulting engagement with the organisation involved in the third application (Section 7). No external grant funding was received. No stock ownership, patent applications, paid expert testimony, or affiliation with this journal's editorial board apply.

ETHICS AND RESEARCH GOVERNANCE

This study analyses records generated during consulting engagements, retrospectively examined as research data; the analytical corpus closed on 15 July 2026 (Section 3). Participating organisations authorised the authors to describe the protocol and process-level findings for academic publication. Client review was limited to confidentiality risks and factual accuracy of client-owned material; organisations did not review, approve, or influence the authors' analytical interpretations or conclusions, developed independently. Client-owned artefacts were included in the analysis where authorised, but are not reproduced, archived, or shared publicly (Section 7).

REFERENCES

- Bammer, G. (2019). Key issues in co-creation with stakeholders when research problems are complex. *Evidence & Policy*, 15(3), 423–435. <https://doi.org/10.1332/174426419X15532579188099>
- Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics*, 5, 3265–3279. <https://doi.org/10.1007/s43681-024-00653-w>
- Birhane, A., Isaac, W., Prabhakaran, V., Díaz, M., Elish, M. C., Gabriel, I., & Mohamed, S. (2022). Power to the people? Opportunities and challenges for participatory AI. *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*, Article 6, 1–8. <https://doi.org/10.1145/3551624.3555290>
- Birkstedt, T., Minkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: Themes, knowledge gaps and future agendas. *Internet Research*, 33(7), 133–167. <https://doi.org/10.1108/INTR-01-2022-0042>
- Bogucka, E., Constantinides, M., Šćepanović, S., & Quercia, D. (2024). Co-designing an AI impact assessment report template with AI practitioners and AI compliance experts. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1), 168–180. <https://doi.org/10.1609/aies.v7i1.31627>
- European Parliament and Council. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union.
- Head, B. W., & Alford, J. (2015). Wicked problems: Implications for public policy and management. *Administration & Society*, 47(6), 711–739. <https://doi.org/10.1177/0095399713481601>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- ISO/IEC. (2023). *ISO/IEC 42001:2023 — Information technology — Artificial intelligence — Management system*. International Organization for Standardization.
- Chandra Kruse, L. (2025). Design principles unveiled: Capturing innovation insights for future design breakthroughs. *CERN IdeaSquare Journal of Experimental Innovation*, 9(1), 5–8. <https://doi.org/10.23726/cij.2025.1582>

- Lipmanowicz, H., & McCandless, K. (2014). *The surprising power of liberating structures: Simple rules to unleash a culture of innovation*. Liberating Structures Press.
- Mökander, J., & Floridi, L. (2023). Operationalising AI governance through ethics-based auditing: an industry case study. *AI and Ethics*, 3(2), 451–468. <https://doi.org/10.1007/s43681-022-00171-7>
- Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *Journal of Strategic Information Systems*, 34(2), 101885. <https://doi.org/10.1016/j.jsis.2024.101885>
- Peffer, 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>
- Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), 155–169. <https://doi.org/10.1007/BF01405730>
- Rogers, P., & Blenko, M. (2006). Who has the D? How clear decision roles enhance organizational performance. *Harvard Business Review*, 84(1), 52–61.
- Romme, A. G. L. (2003). Making a difference: Organization as design. *Organization Science*, 14(5), 558–573. <https://doi.org/10.1287/orsc.14.5.558.16769>
- Ruster, L. P., & Daniell, K. A. (2025). How to operationalize responsible use of artificial intelligence. *MIS Quarterly Executive*, 24(2), Article 6, 1–19. <https://doi.org/10.17705/2msqe.00116>
- Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly*, 35(1), 37–56. <https://doi.org/10.2307/23043488>
- Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin*, 135(1), 94–120. <https://doi.org/10.1037/a0014212>
- van Aken, J. E., & Romme, A. G. L. (2012). A design science approach to evidence-based management. In D. M. Rousseau (Ed.), *The Oxford handbook of evidence-based management* (pp. 43–57). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199763986.013.0003>
- von Busch, O., & Palmås, K. (2016). Social means do not justify corruptible ends: a realist perspective of social innovation and design. *She Ji: The Journal of Design, Economics, and Innovation*, 2(4), 275–287. <https://doi.org/10.1016/j.sheji.2017.07.002>
- Zampirolo, G., Burtcher, F., Mariotti, Y., Rudnik, D., & Aderemi-Makinde, M. (2026). The AI governance paradox: Ethical awareness without operational clarity. In R. Kimmons (Ed.), *Artificial Intelligence ethics, law, and policy* [Working title]. IntechOpen. <https://doi.org/10.5772/intechopen.1015694>