CERN IdeaSquare Journal of Experimental Innovation, 2026; 10(2): 144-150
DOI: https://doi.org/10.23726/cij.2026.1824

ORIGINAL ARTICLE

An AI Convenor for Multi-Domain Co-Creation: A Governance-Constrained Decision-Rights and Escalation Framework

Masaya Ochiai,[1*]

[1]Independent Researcher, Nagoya, Aichi, Japan
*Corresponding author: masaya.ochiai.research@gmail.com

ABSTRACT

Artificial Intelligence (AI) increasingly summarises meetings and prompts participation, but multi-domain co-creation also requires rules for whose constraints count, when action must stop, and who may decide. We propose an AI Convenor as a governance-constrained facilitation role. The paper's central artifact is a decision-rights and escalation framework that permits only pre-authorised procedural actions, defaults substantive choices to AI-propose/human-decide, and accumulates trigger-specific human decisions and checkpoints. A short protocol and reusable records operationalise the framework. AIC-12 is offered as a candidate 12-item post-session set—not a validated instrument—alongside process indicators and conditional propositions for comparative testing. A synthetic vignette illustrates translation, option comparison, sensitive data, dissent, and closure. The contribution is a testable design pattern; no empirical effectiveness claim is made.

Keywords: Co-creation; human-AI collaboration; governance; facilitation; decision rights; escalation; procedural fairness.

Received: January 2026. Accepted: August 2026.

INTRODUCTION

Co-creation across disciplines and stakeholder groups can address problems that no single domain can solve alone. Yet multi-domain work is vulnerable to incompatible vocabularies, uneven participation, and conflicting standards of evidence. A recurring bottleneck is the convenor role: someone must organise the process, translate across domains, record choices, and protect the legitimacy of decisions.

Consider a team of clinicians, data engineers, and procurement staff choosing a hospital analytics system. The word "safety" may refer to patient harm, model failure, or contractual liability. An AI Convenor could ask each domain to define the term, place the definitions in a shared glossary, and record the options. However, if the discussion turns to patient data or an external purchasing commitment, the AI should not close the issue; it should pause the affected step and require an authorised human review.

This paper asks how AI can perform limited convening functions while preventing authority capture, and what governance and evaluation materials would make such a design testable. We propose an AI Convenor: a socio-technical facilitation role that standardises translation, participation checks, and decision traceability while operating under explicit human decision rights.

Table 1 is the central artifact. The capability description, protocol, decision log, candidate AIC-12 items, process indicators, and vignette show how it could be implemented and tested; they are supporting materials, not separate effectiveness claims. The contribution is a testable design pattern, not empirical findings.

BACKGROUND AND POSITIONING

Co-creation research emphasises that outcomes depend not only on ideas but on participation, legitimacy, and implementation (Sanders & Stappers, 2008; Voorberg et al., 2015). Boundary objects—shared artifacts usable across communities without complete agreement—help coordinate knowledge boundaries (Star & Griesemer, 1989; Carlile, 2002). Convening is therefore a locus of power: it shapes which constraints enter the record, which disagreements remain visible, and when closure is declared. Co-creation can be corrupted when power and value capture remain implicit (von Busch & Palmås, 2023), while algorithmic management shows that scalable coordination can create

opacity and legitimacy risks (Lee et al., 2015; Kellogg et al., 2020).

In this paper, a domain is a community of practice with specialised language, methods, and standards; co-creation entails joint framing, shared artifacts, and joint decisions. A convenor maintains process but has no content authority by virtue of the role. "Multi-domain" means three or more domains as a scope condition, not a natural cutoff: it creates multiple translation interfaces and possible coalitions. Studies should vary domain count rather than treat three as universal.

AI-supported meeting systems mainly provide behavioural feedback, summaries, or conversational cues. MeetingCoach offers a post-meeting dashboard for effective and inclusive meetings (Samrose et al., 2021); other work examines real-time generative ideation cues (Rayan et al., 2024) and virtual co-hosts that observe, ask, and intervene for inclusion (Houtti et al., 2025). The present proposal addresses a different layer: what an AI facilitator may execute, may only propose, and must escalate. Its novelty is not AI summarisation or prompting, but a visible governance-response hierarchy that bounds facilitation authority.

Trust cannot be assumed. People may reject an algorithm after observing an error (Dietvorst et al., 2015), prefer algorithmic advice in some tasks (Logg et al., 2019), or accept imperfect systems when they retain limited control (Dietvorst et al., 2018). Perceived fairness may therefore vary with stakes, visible mistakes, prior trust, and meaningful human control. The framework makes contestability and escalation design requirements and states P1 conditionally.

THE AI CONVENOR FRAMEWORK

Design assumptions, goal, and hard boundaries

The framework rests on four design assumptions, not findings: standardised prompts can make domain constraints visible; explicit decision rights can clarify who may decide; contestable records can support review; and a legitimate human escalation path can prevent procedural support from becoming de facto authority. The trigger categories provisionally synthesise these boundaries with research on power asymmetry in co-creation, algorithmic management, human–AI interaction, and algorithmic accountability (von Busch & Palmås, 2023; Lee et al., 2015; Kellogg et al., 2020; Amershi et al., 2019; Diakopoulos, 2016). They are design hypotheses, not an exhaustive or validated risk taxonomy, and require expert, stakeholder, and empirical review.

The design goal is to improve procedural fairness, cross-domain translation, and decision traceability without granting the AI substantive authority. The AI Convenor must not decide normative trade-offs, assign authorship or credit, allocate resources, impose sanctions, commit the group externally, or override an explicit human decision right.

Permitted functions

Within these boundaries, the AI Convenor may:

(i) Translate: maintain a glossary, clarify jargon, and standardise domain constraints.

(ii) Orchestrate process: generate agendas and time boxes and request domain input at decision points.

(iii) Prompt boundary objects: propose only the requirements, risk, evaluation, or glossary artifact needed to expose assumptions.

(iv) Log decisions: record options, criteria, rationale, assumptions, dissent, owners, and verification needs.

(v) Monitor participation: invite missing domains without judging the substance of their contributions.

(vi) Enforce governance: apply Table 1, display the rule and reason, and preserve an auditable record.

Decision rights and escalation

Table 1 is the central artifact. Its ordering represents response restrictiveness, not a universal ranking of harms. When multiple triggers apply, the most restrictive response determines whether the affected action may continue, while all trigger-specific decision rights and checkpoints remain cumulative. Any substantive selection not covered by R1–R4 defaults to AI-propose/human-decide; R0 applies only to pre-authorised, reversible procedural actions. Participants may invoke a more restrictive response, and the AI may not downgrade a human escalation request.

Table 1. Decision-rights and escalation framework (central artifact)

Rule	Trigger and practical cue	Required response
R0	**Routine procedure.** Agenda, minutes, reminders, translation, or record keeping; no substantive choice and no R1–R4 trigger.	AI may execute and log; participants can correct the output.
R1	**Attribution or ownership.** Authorship, credit, ownership, or intellectual-property attribution.	AI may present neutral options and log them; a named human decides.
R2	**Resources or access.** Budget, staffing, access rights, equipment, or other allocation.	Pause the affected action; an authorised human decides and records the decision.
R3	**Sensitive data, scope, or commitment.** Personal or regulated data, sensitive IP, a change to goals or scope, or an external promise.	Stop the affected segment; complete the applicable privacy, legal, ethics, governance, or sponsor/steering checkpoint; record an authorised human decision before proceeding.
R4	**Sanction or exclusion.** Discipline, exclusion, loss of participation rights, or an equivalent sanction.	Stop; complete all other applicable checkpoints. The authorised human body decides; AI records only after confirmation.

"Execute" permits only a pre-authorised, reversible procedure. "Propose" permits organising alternatives but not selecting among them. "Human decision" requires a named person or body to decide and record the result. "Stop-and-review" suspends the affected segment until applicable checkpoints and decisions are complete. Thus, regulated data plus exclusion requires both the R3 checkpoint and the R4 authorised-body decision.

Implementation requirements and failure modes

One implementation combines a transcript, pre-session governance profile, AI classifier, and deterministic governance-rule engine. Participants first register domains, decision rights, sensitive topics, and authorised reviewers. During the session, the model assigns R0 only where it can positively identify a pre-authorised, reversible procedural episode; it maps recognised restricted triggers to R1–R4 and routes every remaining substantive choice to the AI-propose/human-decide default. The rule engine then applies the most restrictive response while retaining every trigger-specific obligation. Low confidence, conflicting labels, or a manual flag defaults upward. The interface displays the active rule and reason, permits correction, time-stamps overrides, and limits model access to sensitive data.

Failure modes include missed triggers, excessive false alarms, lost context, euphemistic or adversarial wording, model drift, and privacy leakage. Missed triggers can permit overreach; repeated false alarms can interrupt work until users ignore the system. Initial implementations should therefore undergo scripted tabletop tests, designated human oversight, and measurement of trigger recall, precision or false-positive rate, and interruption burden before consequential use.

OPERATIONALISATION AND TESTABLE PROPOSITIONS

Operational protocol

Consistent with explicit collaboration planning and evaluation (Drain & Sanders, 2019), the protocol is:

(i) Setup: each domain states objectives, constraints, non-negotiables, and acceptable evidence; participants register decision rights and sensitive topics or data.
(ii) Glossary: identify ambiguous terms and record domain-specific meanings before comparing options.
(iii) Ground rules: confirm a dissent or minority-report option and the Table 1 logic.
(iv) Decision loop: elicit constraints, organise options and criteria, invite missing voices, and check R1–R4 triggers.
(v) Boundary objects: maintain only the requirements list, risk register, evaluation matrix, or glossary needed for the decision.
(vi) Decision capture: log the selection, rationale, assumptions, dissent, active rules, responsible humans, and verification needs.
(vii) Retrospective: complete the candidate AIC-12 items and identify corrections to the glossary, framework, or implementation.

Propositions and comparative design

The protocol and Table 1 are design specifications. The following propositions are hypotheses for later tests, not findings from this manuscript:

P1 (conditional fairness): Compared with an insider convenor, a governance-constrained AI Convenor will increase perceived procedural fairness more for routine,

low-stakes facilitation than for high-stakes decisions. Any advantage will weaken or reverse after a salient AI error and will vary with prior trust in AI.

P2 (contestability): Explicit decision rights, visible reasons, and correction mechanisms will increase perceived contestability relative to AI facilitation without a governance rule.

P3 (translation): Structured glossary and constraint prompts will increase perceived translation quality and shared understanding across domains.

P4 (traceability): The framework will improve the completeness of decision records, although additional logging and escalation may increase session duration.

P5 (overreach): AI facilitation without explicit decision rights and auditable logs will create greater risk of hidden procedural bias and illegitimate authority capture.

P6 (skill shift): Repeated reliance on the AI Convenor may reduce participants' own facilitation and translation practice unless sessions include reflection or rotating human facilitation roles.

A comparative study could cross convenor type (insider human, external human, governance-constrained AI, ungoverned AI) with decision stakes (routine, high) and visible error (absent, present). Fairness should be rated separately by episode. Prior trust should be measured before the session, and noticed errors after it, so P1 is tested as a moderated effect rather than a uniform benefit.

Decision record and supporting boundary objects

The decision log is the minimum record needed to make the framework inspectable. It can be implemented in any meeting platform and does not depend on a particular model.

Table 2. Decision log template

Field	Required content
Identifier and date	Unique decision ID and session or timestamp.
Decision and options	What was decided; options considered; criteria used.
Domain constraints	Relevant constraints stated by each affected domain.
Selection and rationale	Chosen option and reasons for selection.
Dissent and assumptions	Minority report, unresolved objections, and assumptions that must hold.
Governance status	Active Table 1 rules or substantive-choice default, cumulative checkpoints, escalation taken, and human decision-maker.
Follow-through	Owner, deadline, and evidence or verification needed.

Four optional boundary objects may support the record: a requirements list for constraints and dependencies; a risk register for likelihood, impact, mitigation, triggers, and ownership; an evaluation matrix for declared criteria and weights; and a glossary for semantic drift. These are supporting devices, and teams should use only those that reduce coordination burden.

AIC-12 AS A CANDIDATE POST-SESSION ITEM SET

Rationale and items

AIC-12 is a candidate set of twelve post-session items, not a validated instrument. Its four provisional domains are procedural fairness and legitimacy (Colquitt, 2001), voice and contestability (Edmondson, 1999; Diakopoulos, 2016), translation and shared understanding (Star & Griesemer, 1989; Carlile, 2002), and trust and coordination (McAllister, 1995; Lewis, 2003). Responses use a seven-point scale from 1 (strongly disagree) to 7 (strongly agree). The wording is author-generated and conceptually informed by these constructs; it is not adapted from, or validated against, the cited scales.

Table 3. AIC-12 candidate item set

Code	Candidate item
F1	Each domain had an explicit opportunity to state its constraints before important decisions.
F2	The process did not feel biased toward a particular domain or role.
F3	Options considered and reasons for decisions remained traceable after the session.
V1	I could raise dissent without fear of negative consequences.
V2	I could easily challenge, revise, or correct AI summaries or recommendations.
V3	Persistent disagreement remained visible in the record or was referred to an identified human decision process rather than being treated as closure.
T1	Specialised terms and assumptions were translated across domains.
T2	Domain requirements and constraints were organised in shared artifacts.
T3	The source of confusion—definitions, assumptions, or criteria—became clear.
TC1	I trusted that other domains' expertise was represented adequately for the decisions made.
TC2	Roles and next actions supported post-session coordination.
TC3	Unresolved issues identified an owner, timeline, and verification evidence.

Provisional reporting and validation

Before validation, report item-level distributions and missingness. Exploratory means for the four provisional domains may be shown, but a total "session health" score, hard cutoff, and automatic imputation are unjustified. A low rating on an item should prompt qualitative inquiry, not diagnosis.

Validation should remain separate from evaluation of the AI Convenor. It should begin with expert content review, cross-domain cognitive interviews, and a pilot of response processes and missingness. Later studies can test factor structure, reliability, convergent and discriminant validity, sensitivity to convenor type and stakes, and measurement invariance. Comparative studies should also measure pre-session trust and whether participants noticed an AI error.

OBJECTIVE PROCESS INDICATORS

Observable indicators can complement self-report. Speaking-time share and speaking-turn share should be reported separately by domain during decision episodes. A dominance ratio is the largest share divided by the smallest; when the minimum is zero, report the ratio as undefined and state the zero-participation case. Prompt responsiveness is the proportion of directed prompts that receive a response. Boundary-object production counts artifacts created, and reuse counts decisions that reference them. Decision-log completeness is the proportion of required Table 2 fields completed. Governance detection should report trigger recall and precision or false-positive rate; escalation compliance is the proportion of audited triggers for which all required responses occurred. Interpretation remains context-dependent.

SYNTHETIC WALKTHROUGH

Table 4 returns to the introductory hospital example. It illustrates intended operation; it is not empirical evidence that the framework is feasible or effective.

Table 4. Synthetic walkthrough of a multi-domain session

Situation	AI procedural action	Human authority and record
Glossary clash	Flags that "safety" has different meanings and requests definitions.	Domains confirm meanings; glossary entry is saved.
Option comparison	Proposes criteria fields and enters stated constraints.	Humans approve criteria and any declared weights; decision log is updated.
Sensitive-data trigger	Detects patient-data discussion and activates R3.	Authorised reviewers pause the segment, complete the privacy checkpoint, and decide whether to proceed.
Persistent dissent	Offers a minority-report field and avoids forced consensus.	The dissenting domain records its rationale; the objection remains visible or is referred to a named human process.
Closure	Summarises actions, owners, deadlines, and verification needs.	Humans confirm commitments and correct the final record.

DISCUSSION

AI as a meeting tool differs from AI as convening infrastructure. Summaries and prompts do more than save time when they shape whose constraints are visible, which options survive, or whether disagreement appears resolved. These functions distribute voice, legitimacy, and accountability. The framework therefore treats AI as a constrained process actor whose outputs are correctable records and prompts, not binding decisions.

Trust is not simply whether participants "like" the system. A visible mistake may reduce fairness judgments in high-stakes episodes, while correction paths and retained human control may preserve use. Conversely, smooth performance may create unwarranted authority. Evaluation should therefore examine contestability, escalation compliance, record quality, interruption burden, and differences between routine and sensitive decisions.

LIMITATIONS AND RESEARCH PROGRAMME

This paper reports no implementation, stakeholder elicitation, pilot, or comparative experiment. The trigger categories may be incomplete, the response-restrictiveness ordering may be too coarse, and the framework may interrupt some settings more than it protects them. Three or more domains is an operational boundary, not an empirical law. AIC-12 remains a candidate item set with unknown content validity, reliability, and structure.

A design-science programme can develop the artifact through staged testing (Romme, 2023): (1) expert and stakeholder review of the trigger categories and candidate items; (2) tabletop alpha tests with scripted routine and restricted episodes; (3) prototype tests of recall, precision or false-positive rate, privacy failures, and interruption burden; (4) comparisons of insider human, external human, governed AI, and ungoverned AI facilitation; and (5) a separate psychometric programme for AIC-12. Failures, overrides, and unintended authority effects should be primary results, not implementation noise.

TRANSPARENCY AND RESPONSIBILITY STATEMENT

1. Humans retain responsibility for all substantive selections and for scope, authorship and credit, resources, sanctions, and every action activating R1–R4.
2. The AI Convenor is limited to procedural facilitation, translation, and record keeping under contestability and auditability.
3. ChatGPT (OpenAI) was used for translation into English and language editing of content conceived by the author, as well as for assistance with literature searches. Conceptual design, argument structure, and final source selection were based on the author's own judgment. All final wording was reviewed, revised, and approved by the author, who takes full responsibility for the manuscript.

AUTHOR CONTRIBUTIONS

Masaya Ochiai conceived the work, developed the governance framework and supporting materials, and wrote the manuscript.

FUNDING

The author received no financial support for the research, authorship, and/or publication of this article.

CONFLICTS OF INTEREST

The author declares no competing interests.

REFERENCES

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19) (Article 3, pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300233

Carlile, P. R. (2002). A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science, 13*(4), 442–455. https://doi.org/10.1287/orsc.13.4.442.2953

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86*(3), 386–400. https://doi.org/10.1037/0021-9010.86.3.386

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM, 59(*2), 56–62. https://doi.org/10.1145/2844110

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Drain, A., & Sanders, E. B.-N. (2019). A collaboration system model for planning and evaluating participatory design projects. *International Journal of Design, 13*(3), 39–52. https://www.ijdesign.org/index.php/IJDesign/article/view/3486

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*(2), 350–383. https://doi.org/10.2307/2666999

Houtti, M., Zhou, M., Terveen, L., & Chancellor, S. (2025). Observe, ask, intervene: Designing AI agents for more inclusive meetings. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (Article 709, pp. 1–18). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713838

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)* (pp. 1603–1612). Association for Computing Machinery. https://doi.org/10.1145/2702123.2702548

Lewis, K. (2003). Measuring transactive memory systems in the field: Scale development and validation. *Journal of Applied Psychology, 88*(4), 587–604. https://doi.org/10.1037/0021-9010.88.4.587

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38*(1), 24–59. https://doi.org/10.5465/256727

Rayan, J., Kanetkar, D., Gong, Y., Yang, Y., Palani, S., Xia, H., & Dow, S. P. (2024). Exploring the potential for generative AI-based conversational cues for real-time collaborative ideation. In *Proceedings of the 16th Conference on Creativity & Cognition* (pp. 117–131). Association for Computing Machinery. https://doi.org/10.1145/3635636.3656184

Romme, A. G. L. (2023). Design science as experimental methodology in innovation and entrepreneurship research: A primer. CERN *IdeaSquare Journal of Experimental Innovation, 7*(2), 4–7. https://doi.org/10.23726/cij.2022.1427

Samrose, S., McDuff, D., Sim, R., Suh, J., Rowan, K., Hernandez, J., Rintel, S., Moynihan, K., & Czerwinski, M. (2021). MeetingCoach: An intelligent dashboard for supporting effective and inclusive meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)* (Article 252, pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3411764.3445615

Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign, 4*(1), 5–18. https://doi.org/10.1080/15710880701875068

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, translations, and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science, 19*(3), 387–420. https://doi.org/10.1177/030631289019003001

von Busch, O., & Palmås, K. (2023). *The corruption of co-design: Political and social conflicts in participatory design thinking*. Routledge. https://doi.org/10.4324/9781003281443

Voorberg, W. H., Bekkers, V. J. J. M., & Tummers, L. G. (2015). A systematic review of co-creation and co-production: Embarking on the social innovation journey. *Public Management Review, 17*(9), 1333–1357. https://doi.org/10.1080/14719037.2014.930505