Co-Creating Operational AI Governance

Giorgio Zampirolo,[1*] Daria Rudnik[2], Felicia Burtscher[3]

[1]Independent Researcher, London, United Kingdom; [2]Independent Consultant, Tel Aviv, Israel; [3]Independent Researcher and Consultant, Berlin, Germany
*Corresponding: giorgio.zampirolo@gmail.com

CERN IdeaSquare Journal of Experimental Innovation, 2026; 10(2): 151-156
DOI: https://doi.org/10.23726/cij.2026.1856

Supplementary Material A: Replication Pack

This pack is organised in five parts:

(A) the protocol and facilitator manual; (B) printable, fillable participant materials; (C) research documentation and fidelity forms; (D) a synthetic worked example; and (E) licence and references.

Each part can be used independently.

PART A: PROTOCOL AND FACILITATOR MANUAL

Purpose of the Replication Pack

This pack provides an operationally specified, facilitator-ready protocol for implementing the co-creation method described in 'Co-Creating Operational AI Governance.' It documents the protocol in its current, most developed form (Version 3, used with two documented adaptations in the third and most recent organisational application), alongside the version history showing how it developed across the first two applications.

The pack is designed to enable replication of the process structure, not replication of outcomes or organisational effectiveness. Governance artefacts produced through this method are expected to vary across organisational contexts, reflecting differences in risk profiles, authority structures, and regulatory environments. Several operational decisions are not yet specified by the protocol itself; these are named explicitly, not left for a facilitator to infer, in 'Facilitator Decisions Not Otherwise Specified' below.

Archival, Hosting, and Licensing

This pack, including all blank templates and the synthetic worked example, will be deposited in CIJ's own Zenodo community (zenodo.org/communities/cij) and assigned a persistent DOI prior to publication, distinct from the DOI assigned to the paper. It will be released under a CC BY 4.0 licence, matching CIJ's own copyright terms, permitting reuse and adaptation with attribution. Client-owned artefacts from the three real applications are excluded from this archive, consistent with the confidentiality terms in Part C; only the protocol, templates, version history, and synthetic example are archived. Supplementary material of this kind is not counted toward the journal's word limit for the main article.

Scope and Validity Statement

This protocol's validity claim is limited to protocol specification and preliminary process feasibility. The protocol is operationally specified to support future replication attempts; independent procedural replication has not yet been demonstrated.

In the three applications reported in the accompanying paper, preliminary process feasibility was assessed against the following criteria, applied retrospectively to records generated during and after each engagement:

- whether the planned stages could be enacted;
- whether the required participant roles were represented;
- whether a decision-rights or escalation artefact was produced;
- whether documented concerns resulted in traceable artefact changes;
- whether outputs corresponded to the stated governance dilemma.

The Feasibility Matrix in Part C applies these criteria to each application individually, rather than asserting feasibility in general terms. This pack does not claim to establish organisational effectiveness, long-term adoption, or comparative artefact quality; these remain targets for future, prospectively designed evaluation with criteria fixed before, not after, data collection.

Artefact-Assessment Dimensions

Governance artefacts produced through this protocol can be assessed for:

- Completeness: whether required components (decision owners, escalation triggers, review dates) are present.
- Structural correspondence: whether the artefact follows the expected decision-rights/escalation format.
- Substantive correspondence: whether content is consistent with session records and facilitator documentation.
- Traceability: whether elements can be linked to documented participant inputs and decisions.
- Internal coherence: whether the artefact's elements are mutually consistent.
- Contextual relevance: whether the artefact addresses the stated governance dilemma and organisational constraints.

Part D demonstrates all six dimensions applied to a worked example, including an honest account of which dimensions a synthetic example can and cannot actually support.

Protocol Version History

This pack documents Version 3, the current and most developed version of the protocol, applied in Application 3 with two documented adaptations detailed in the Protocol Fidelity Log (Part C). Earlier versions, used in the first two applications, are shown below for transparency about how the protocol developed.

Version	Applied in	Key features	Motivating observation
1	Application 1	Single sitting; three phases run consecutively; no designed opening step	NA
2	Application 2	Adds an opening step (Session 0), though not yet formally designed	Application 1 showed that running all three phases consecutively, without a planned opening or assigned roles, was associated with uneven participation and incomplete convergence
3 (current)	Application 3	Adds a fully designed Session 0 commitment step, an explicit Input/Agree distinction (RAPID-inspired) in Session 2, separated executive-sponsor and final-decider roles activated in Session 3, and a post-session assurance review window	Motivated by the need for structural, rather than purely facilitator-dependent, safeguards around participation balance, decision authority, and institutionalisation

Summary of the Three Applications

A process-level summary of the protocol's three organisational applications, consistent with Table 1 in the accompanying paper; full case-level detail, including the feasibility matrix and fidelity log, appears in Part C.

	Application 1	Application 2	Application 3
Protocol version used	Version 1	Version 2	Version 3, with two documented adaptations (see Protocol Fidelity Log)
Stages completed	Divergence (overran); Interpretive Coordination (late, rushed); Convergence not reached	Session 1 (Divergence) completed by corpus close (15 Jul 2026); Session 2 neither scheduled nor conducted	All stages completed for continuing group (n=8); second group's continuation undecided
Key process finding	Absence of assigned roles or a planned opening associated with a few participants dominating discussion	Operational concerns raised within the participating group; leadership clarified non-negotiables in return; the scheduled absence was assessed as an implementation condition, not a protocol deviation	Escalation-trigger item flagged under an explicit Agree right; required a structural amendment; independent assurance review produced further changes
Change following	Motivated introduction of a designed Session 0	Further sessions anticipated but unscheduled at corpus close; no later evidence existed	Not applicable: the most recent application

Boundary Conditions and Assumptions

The protocol is designed for, and is expected to be most readily enacted under, the following conditions:

- Participants hold decision-relevant authority or influence over AI-related practices.
- The organisation is willing to treat governance artefacts as provisional and revisable, rather than final policy.
- Facilitation gives every participant a structured opportunity to speak before senior voices dominate, rather than relying on facilitator discretion alone.
- The Agree-holder (the function assigned bounded veto authority in Session 2, Legal/Compliance in the reported applications but not necessarily so) holds a genuine, organisationally recognised mandate for that authority. This right reallocates power to a specific function rather than neutrally resolving it; its legitimacy depends on organisational mandate, the nature of the contested issue, and a clear separation of legal advice, compliance approval, and executive risk acceptance.

Outcomes will vary depending on organisational maturity, power dynamics, facilitation quality, and regulatory constraints.

<u>An open design question</u>: is a binary Agree right enough?

The current protocol distinguishes only advisory Input from a bounded Agree veto, held by whichever function has recognised authority over the concern at hand. A more granular design would separate legal interpretation, compliance approval, ethical or policy recommendation, executive risk acceptance, and the final decider's approval as distinct roles. None of the three applications to date used this more granular structure, and this pack does not present it as something already tested. It is recorded here as an open question for future protocol development: the present binary may understate

the complexity of real governance authority, particularly where the Agree-holder can flag a concern but a different, more senior actor (the final decider) lawfully accepts the residual risk.

Operational Contingencies

The guidance above describes the protocol as specified. Facilitators should expect, and plan for, the following departures:

- Minimum viable group. If invitees decline at Session 0, the session may proceed with as few as one representative each from technical, operational, legal/compliance, and leadership perspectives (a minimum of 4), though fewer than 6 total departs from the specified range and should be recorded as such: following the precedent of Application 2, where an unplanned absence was documented as an implementation condition rather than treated as invalidating the session.
- Agree-holder unavailable when an item is flagged. Log the item and defer it to the post-session assurance review rather than resolve it without appropriate authority. The session may proceed to Session 3 on that item only provisionally, with the artefact explicitly marked pending Agree sign-off.
- Quorum for the compliance-review segment. Session 2's compliance segment should not proceed to assigning binding Agree rights without the designated Agree-holder present. If absent for the full session, treat all resulting assignments as provisional pending their review.
- Absence of the executive sponsor in Session 3. Defer institutionalisation to a documented follow-up commitment rather than assign it to a substitute without equivalent organisational standing, since the mechanism depends on the sponsor's own authority over the target forum.

Facilitator Decisions Not Otherwise Specified

The protocol does not yet specify every operational decision a facilitator will need to make. These are named explicitly below so a replicator does not have to infer them from general facilitation experience.

- Bounding the governance dilemma: choose a concrete, current dilemma with genuine cross-functional stakes, not one already resolved by existing policy or dependent on information held by only one function.
- Participant eligibility: a participant is eligible if their role could plausibly be bound by the resulting artefact (i.e., they or their function would need to act differently once it exists) not merely be informed of it.
- Facilitator competence and independence: experienced in structured facilitation, without a stake in the substantive governance outcome. Where the facilitator's organisation has a consulting/commercial relationship with the client, as in this protocol's own applications, this should be disclosed, not treated as disqualifying in itself.
- Sponsor-selection criteria: the executive sponsor should have actual, current authority over the specific governance, risk, or audit forum the artefact will be embedded into, not merely general seniority.
- Boundary of the Agree-holder's mandate: agree the scope of matters within the Agree-holder's authority with the organisation before Session 0, not while facilitating Session 2; document this scope on the Agree Decision Log (Part B), and confirm separately who holds the final decider role.
- Pre-work anonymisation: the facilitator removes names and identifying role detail before circulating pre-work, retaining a private key for facilitation purposes only.
- Recording disagreement: log residual disagreement in the designated worksheet field; it is retained as data, not resolved by majority vote or facilitator override.
- Session-record validation: the read-back check built into Sessions 1 and 2 constitutes real-time validation; no separate step is required.
- Logging adaptations: any departure from the specified protocol is recorded on the Protocol Fidelity Log (Part C), not left informal.
- Defining a failed session: a session that does not complete its planned stages is a partial application, not a failure; report what was and was not completed rather than characterising the session as having failed or succeeded.
- Handling an unresolved binding item: defer to the assurance review; if still unresolved there, mark the artefact provisional pending a scheduled follow-up, rather than treating silence as resolution.
- Artefact version control: each iteration of a governance artefact carries a version number, date, and facilitator sign-off.
- Recording review-date outcomes: note whether the scheduled review actually took place and what, if anything, changed; an unmonitored review date is not evidence of institutionalisation.
- Online versus in-person delivery: the structured elicitation formats used in Sessions 1–2 were designed with in-person facilitation in mind. Their performance online has not yet been tested and should be treated as an open question, not assumed equivalent.

Intended Use Context

- Cross-functional AI-governance design sessions involving technical, operational, legal/compliance, and leadership perspectives
- Executive education or organisational learning interventions
- Early-stage or transitional AI-governance initiatives

Session Overview

Total duration: approximately 3 hours 45 minutes of synchronous facilitation across three sessions, plus a short asynchronous commitment step (Session 0) and a 3–5 business day assurance-review window following the final session.

Timings are indicative and may vary with facilitation style and organisational context.

Participants: 6–12 individuals. Roles represented: Strategy/Leadership, Operations, Technology/Data, Legal/Compliance.

Facilitator: 1 (neutral, process-focused).

Required Inputs

- Governance Challenge: a concrete and current dilemma (e.g. accountability for AI-supported decisions, escalation paths for model failure, override authority).
- Participant Group: cross-functional, with decision relevance rather than advisory status.
- Minimal Contextual Framing: enough information to establish shared reference, without pre-empting interpretation.

Method Structure

The protocol comprises a preparatory step, three facilitated sessions, and a separate individual assurance review.

Session 0: Commitment and Framing (15–20 minutes, asynchronous or short call)

Purpose: secure each participant's explicit commitment to attend all subsequent sessions before Session 1 begins; distribute the governance dilemma and pre-work in advance; designate an executive sponsor, activated in Session 3, and a final decider with documented authority to approve the resulting arrangement. These may be the same individual where organisational authority permits, but this must be confirmed, not presumed.

Facilitator guidance: confirm attendance in writing from each participant; do not proceed to Session 1 scheduling until commitments are received. Where an organisation prefers to pilot with a larger group before committing a smaller working group to the full cycle, as in Application 3, this should be a deliberate, documented decision (see Protocol Fidelity Log, Part C), not an assumed substitute for individual commitment.

Output: confirmed attendance commitments; designated executive sponsor and final decider; pre-work distributed.

Session 1: Divergence (60 minutes)

Pre-work: individual written responses to the core prompts below, submitted in advance and anonymised by the facilitator before circulation.

- What does this situation reveal as the core governance challenge for your role?
- What is at stake for your domain? What limits, obligations, or constraints shape your view?
- Where do you believe accountability should sit? At which stage of the AI-supported decision process?
- What interpretations or expectations from other roles may not align with your domain's reality?

In-session structure: framing and silent review of anonymised pre-work themes; a structured small-group rotation (1-2-4-All: individual reflection, then pairs, then groups of four, then whole-group harvest; Lipmanowicz & McCandless, 2014) that gives every participant a structured opportunity to speak in a small setting before whole-group discussion opens;

facilitator-led, visible capture of the divergence map; a read-back check in which the facilitator reads the captured map aloud and participants may immediately correct mischaracterisations before the note is finalised.

Output: a divergence map, validated by participants in the room before circulation.

Session 2: Interpretive Coordination and Compliance Check (75 minutes)

Pre-work: review of the validated divergence map.

In-session structure: recap and framing; a structured rotation (e.g. Troika Consulting) in which participants explain their position and restate another's before the whole group reconvenes; a dedicated compliance-review segment adapting the Input/Agree distinction from Rogers & 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 the reported applications, though data-protection, cybersecurity, or safety-governance functions could equally hold it. Most items are marked Input, but items within the Agree-holder's formal remit are marked Agree (a bounded veto, not the protocol's ultimate authority) meaning the artefact cannot proceed to final approval with that item unresolved; facilitator-led capture of the shared-vocabulary record and the Agree flag log; a read-back check.

Output: a shared vocabulary record and an Agree flag log with assigned rights, both participant-validated.

Session 3: Convergence and Provisional Approval (90 minutes)

Pre-work: review of the shared vocabulary record and Agree flag log.

In-session structure: recap and framing; joint construction of the governance artefact (decision-rights map/escalation pathway), items with an Agree flag cannot be marked resolved without the Agree-holder's explicit sign-off; a read-back check on the draft artefact; a mandatory reflective block in which a named accountable owner is assigned to each artefact element, a resourcing commitment is discussed and recorded, and the executive sponsor commits (in the room) to embedding the artefact into an existing governance, risk, or audit forum, with the review date entered as a condition of the session closing. Where advance scheduling is used instead of a live, in-room entry, as in Application 3, this is a documented adaptation, not the mechanism as originally specified (see Protocol Fidelity Log, Part C).

Output: a session-approved provisional artefact with named owners, a resourcing commitment, and a review date, pending the final decider's approval and the assurance review below.

Post-Session Assurance Review (3–5 business days)

The Agree-holder reviews the session-approved provisional artefact independently, outside the group setting, confirming or formally amending it in writing; material amendments are recirculated to participants with a rationale. The final decider then gives final organisational approval before the artefact is entered into the organisational record. This is not a fourth workshop session and involves no additional participants or scheduling burden.

PART B: PRINTABLE, FILLABLE PARTICIPANT MATERIALS

Blank Decision-Rights Map

To be used during Session 3 (Convergence and Provisional Approval) for joint construction of a provisional governance artefact.

AI System / Use Case: _____ Organisation / Unit: _____ Date: _____

Stage / Decision Type	AI Role	Human Role	Accountable Party / Escalation Trigger
	☐ Recommend ☐ Decide ☐ Suggest ☐ Advisory ☐ None	☐ Review ☐ Decide ☐ Override ☐ Approve ☐ Execute ☐ Monitor	
	☐ Recommend ☐ Decide ☐ Suggest ☐ Advisory ☐ None	☐ Review ☐ Decide ☐ Override ☐ Approve ☐ Execute ☐ Monitor	
	☐ Recommend ☐ Decide ☐ Suggest ☐ Advisory ☐ None	☐ Review ☐ Decide ☐ Override ☐ Approve ☐ Execute ☐ Monitor	

Session 0 Commitment Card

Participant name: _____ Role/domain: _____

☐ I confirm I can attend all three subsequent sessions.

☐ I have reviewed the governance dilemma and pre-work instructions.

Executive sponsor designated (name/role): _____

Final decider designated (name/role, may be the same as sponsor if confirmed): _____

Agree Decision Log (Session 2)

Item flagged: _____

☐ Input (advisory)

☐ Agree (bounded veto — artefact cannot proceed to final approval until resolved)

Agree-holder for this item (function/name): _____

Agreed scope of the Agree-holder's authority (confirmed before Session 0):

_____

Resolution / amendment required: _____

Sponsor Institutionalisation & Final Approval (Session 3)

Governance/risk/audit forum into which this artefact will be embedded: _____

Review date: _____ Entered live in session:

☐ Yes

☐ No — arranged in advance (record date arranged and by whom, and log as an adaptation)

Final decider approval (name/role, following any assurance-review amendments): _____

Think-Tank Design Canvas

A boundary object for structured collective sense-making across professional domains.

Session 0: Commitment and Framing

Governance challenge (facilitator completes before the session):

_____

Your role/domain (participants select one):

☐ Strategy / Leadership

- ☐ Operations / Service Delivery
- ☐ Technology / Data / AI
- ☐ Legal / Compliance / Risk
- ☐ Change Management
- ☐ Other: _____
- ☐ I confirm my commitment to attend all three sessions.

Session 1: Divergence

1. Governance challenge from your perspective: _____

2. Salient risks, responsibilities, or constraints: _____

3. Accountability and decision authority — where, and at which stage:
_____

4. Assumptions you believe others may be making: _____

Session 2: Interpretive Coordination and Compliance Check

Cross-role translation — how does each role define accountability, acceptable risk, decision authority, and escalation triggers?

Facilitator notes — technical/AI: _____ Legal/compliance: _____ Operations: _____

Strategy/leadership: _____

Key points of interpretive misalignment: _____

Emerging shared vocabulary: _____

Agree decision log — item flagged: _____ ☐ Input ☐ Agree Resolution required: _____

Session 3: Convergence and Provisional Approval

Select artefact type:
- ☐ Decision-Rights Map
- ☐ Escalation Pathway
- ☐ Accountability Matrix
- ☐ Threshold / Override Framework

Provisional agreements reached: _____

Residual disagreements: _____

Named accountable owner(s): _____

Resourcing commitment: _____

Sponsor institutionalisation — forum: _____ Review date: _____ Entered live in session: ☐ Yes ☐ No

Final decider approval (pending assurance review): _____

Canvas Usage Notes

For facilitators: give every participant a structured opportunity to speak before senior voices dominate, rather than relying on discretion alone; prioritise translation over resolution in Session 2; accept residual disagreement as legitimate in Session 3.

For participants: speak from your authentic professional perspective; domain-specific language is welcome and will be translated; uncertainty and disagreement are data, not problems to suppress.

For researchers and replicators: the canvas structures the protocol described in the accompanying paper. Completed canvases, alongside facilitator reports and the other materials listed in Part C, form part of the evidence base for process reconstruction; canvases alone are not treated as sufficient evidence of what occurred. The canvas functions as a boundary object enabling coordination across professional domains.

PART C: RESEARCH DOCUMENTATION AND FIDELITY FORMS

Data and Documentation Generated

Where this protocol is used as part of a research study, the following materials constitute the evidence corpus, consistent with the accompanying paper's methodology:

- Facilitator reports
- Formal internal and engagement communications
- Formal follow-up communications
- Documented reflections on process stages
- Formal participant-generated outputs
- Final artefacts
- Documentation comparing expected and achieved outcomes
- Assessments of artefact content and structure

Informal messages and draft artefacts are excluded from the evidence corpus. No session recordings were used in the analysis; recordings existed for the third application but were not authorised for research access. These materials support process reconstruction and reflective, within- and cross-case analysis rather than statistical evaluation.

Feasibility Matrix

The five feasibility criteria in Part A, applied individually to each application rather than asserted in general terms, together with an additional institutional-review indicator:

Criterion	Application 1	Application 2	Application 3
Required roles represented	Partial: participants spanned multiple functions; full coverage against the protocol's specified role set not separately confirmed	Partial: leadership group; full cross-functional coverage not separately confirmed	Met: cross-functional group including a Legal/Compliance representative with assigned decision authority
All phases enacted	Not met: Convergence not reached	Partial: Session 1 (Divergence) completed by corpus close; Session 2 neither scheduled nor conducted	Met, for the continuing group: all stages completed
Artefact produced	Not met	Not yet (pending)	Met
Traceable changes observed	Not met: no artefact to trace	Not yet applicable	Met: the escalation-trigger amendment and post-session change are both documented
Outputs corresponded to the stated governance dilemma.	Not met: no final artefact was produced.	Not yet assessable: no artefact had been produced by corpus closure.	Met: the resulting decision-rights/escalation artefact addressed the stated governance dilemma.
Institutional review scheduled	Not met	Not yet reached	Met: a review date was recorded, though the scheduling method (advance arrangement rather than live entry) is itself a documented adaptation; see the Protocol Fidelity Log

Protocol Fidelity Log

Application 3 used Version 3 with two documented adaptations. Neither adaptation is hidden inside the limitations section alone; both are named here against the specific protocol element they affect.

Protocol element	As specified	Adaptation actually used, and why	Consequence for interpretation
Session 0 individual commitment	Each participant individually commits to	A staged, one-group-first rollout was used instead: 17 participants split into two groups (8 and 9); only one group proceeded through the full cycle. This	The individual commitment mechanism was not tested as specified in this application; its performance under the

Protocol element	As specified	Adaptation actually used, and why	Consequence for interpretation
	attending all three sessions	was the organisation's own prudent piloting choice, decided in advance.	originally specified design remains untested
Session 3 institutionalisation	The review-date calendar entry is made live, in the room, before participants disperse	The entry was arranged in advance instead. The facilitation team credits careful advance planning (including clarifying roles ahead of each session) as a key factor in keeping the engagement on track.	Institutionalisation occurred and was documented with care, but the live, in-room mechanism specifically was not what produced it here; whether the two approaches differ in durability cannot be determined from this single case
Agree/Decide separation	Full RAPID framework (five roles: Recommend, Agree, Perform, Input, Decide)	The protocol adapts only the Input/Agree distinction from RAPID for the compliance-review segment, and separately designates a final decider (not part of RAPID) who holds ultimate approval authority; it does not implement RAPID in full. The Agree role was held by Legal/Compliance in this application but is not hard-coded to that function.	This is an adaptation inspired by RAPID, not an application of it; readers should not assume fidelity to the full framework

Confidentiality and Disclosure

Consulting client-owned artefacts produced during the three applications are not reproduced in this pack. A detailed log of one post-session review from the third application is withheld under that client's confidentiality policy, triggered by a separate, unrelated engagement with the same client; this is a data-access boundary, not an absence of documentation. The authors' organisation maintained this separate, ongoing consulting engagement with the third application's organisation, scoped independently of and prior to the AI-governance sessions, concerning an unrelated digital-transformation initiative; this relationship did not form part of the analysed intervention. Client review was limited to confidentiality risks and the factual accuracy of descriptions derived from client-owned material; participating organisations did not review, approve, or influence the authors' analytical interpretations, propositions, or conclusions, which were developed independently.

Declaration of Competing Interests

The authors' organisations delivered all three applications reported in this pack and the accompanying paper on a consulting basis, and maintained a separate, ongoing consulting engagement with the organisation involved in the third application at the time of writing (see Confidentiality and Disclosure, above). No external grant funding was received. No stock ownership, patent applications, paid expert testimony, or affiliation with the journal's editorial board apply.

Generative AI Disclosure

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, analytical claims, and citations.

Ethics and Research Governance Statement

This study analyses records generated during consulting engagements, retrospectively examined as research data; the analytical corpus closed on 15 July 2026 (Section 3 of the main paper). 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, which were developed independently. Client-owned artefacts were included in the analysis where authorised, but are not reproduced, archived, or shared publicly (Section 7 of the main paper).

PART D: SYNTHETIC WORKED EXAMPLE

Illustrative Synthetic Example (Hiring-AI Scenario)

The following trace is entirely synthetic. It was constructed to illustrate the protocol's distinctive mechanisms (divergence, translation, a bounded Agree veto, a structural amendment, and an assurance-review change) across all stages of a single fictional case. It is not a report of any of the three organisational applications, and it does not reproduce, adapt, or disclose any client-owned artefact, organisational decision, or internal process.

1. Anonymised pre-work statements (Session 1)

- Technology/AI: "The model flags likely-strong candidates; final judgement should stay human, but the flag should count for something."
- Operations/HR: "We need speed. If every flagged candidate needs a full committee review, the tool doesn't save us anything."
- Legal/Compliance: "Any automated narrowing of a candidate pool needs a documented, auditable rationale we can defend if challenged."
- Leadership: "I want to know who's accountable if we get this wrong, not just that a process exists."

2. Divergence map (Session 1 output)

The four positions above, captured visibly, show accountability located at four different points: the model's output, HR's operational discretion, a documented compliance rationale, and an unspecified "someone" from leadership's perspective.

3. Vocabulary conflict (Session 2)

Translation surfaces that “approve” means different things to different roles: to Operations, it means the model’s suggestion may proceed unless flagged; to Legal/Compliance, it means a human has affirmatively reviewed and signed off. This distinction becomes the session’s central translation task.

4. Agree decision log (Session 2)

Item flagged: whether model-driven shortlisting decisions require a logged rationale before proceeding. Marked: Agree (bias-audit implications, within the Legal/Compliance representative's agreed remit). Ruling: shortlisting cannot proceed to final approval without a bias-audit flag field in the artefact.

5. Initial artefact (Session 3, first draft)

Candidate screening: Recommend/Review, hiring manager, model-confidence trigger. Shortlisting: Suggest/Decide, HR lead, no bias-audit field yet. Final selection: Advisory/Decide, department head, override logged.

6. Binding amendment

Per the Agree ruling above, the Shortlisting row is amended before final approval to add an escalation trigger: bias audit flag.

7. Post-session change (assurance review)

Reviewing independently, the Legal/Compliance representative requests one further change: the Final Selection row's "override logged" trigger is tightened to require a written justification, not just a log entry. A residual, unresolved question is also flagged for further internal reflection: whether the same audit standard should apply when the same tool is used for internal promotions.

8. Final version and review date

Artefact version 2, approved [date]; review scheduled for [date + 6 months]; sponsor: Head of HR; review date entered live in the session.

Stage / Decision Type	AI Role	Human Role	Accountable Party / Escalation Trigger
Candidate screening	☑ Recommend	☑ Review	Hiring manager / model confidence < threshold
Shortlisting (v.1, pre-amendment)	☑ Suggest	☑ Decide	HR lead / (no bias-audit field yet)
Shortlisting (v.2, post-amendment)	☑ Suggest	☑ Decide	HR lead / bias audit flag
Final selection (v.2, post-async change)	☑ Advisory	☑ Decide	Department head / override with written justification

Worked Assessment: Applying the Artefact-Assessment Dimensions

Completeness: Yes. Each row specifies an AI role, a human role, an accountable party, and an escalation trigger; no required field is blank.

Structural correspondence: Yes. The artefact follows the stage-by-stage decision-rights/escalation format specified as Session 3's output.

Substantive correspondence: Not demonstrable from a synthetic example, since there is no underlying session record to check the artefact against. For a real application, this dimension compares the artefact's content against facilitator notes and participant outputs from that session; the escalation-trigger episode in Application 3's Session 2 is a real instance of this check being performed.

Traceability: Partially demonstrable. Each row can be traced to a stage in the stated scenario, but not to documented participant inputs, since none exist for a synthetic case.

Internal coherence: Yes. Escalation triggers are consistent with the accountability assigned at each stage: the hiring manager, accountable at the point closest to the model's output, is assigned the trigger closest to that output, rather than one belonging to a later stage.

Contextual relevance: Yes, at the level of a plausible scenario. The three stages address the stated governance dilemma and cover the decision pipeline end to end.

This split result is itself informative: two of the six dimensions (substantive correspondence and full traceability) cannot be meaningfully assessed without real session records, however well-constructed the synthetic example is. This is a genuine limit of illustrative material, not a gap in this pack, and is the reason the three real applications are reported separately (Part C) even though their artefacts cannot be shown.

PART E: LICENCE AND REFERENCES

This pack is released under a CC BY 4.0 licence upon acceptance (see Archival, Hosting, and Licensing, Part A). The following sources are cited in this pack; sources cited only in the accompanying paper are listed there.

Lipmanowicz, H., & McCandless, K. (2014). The surprising power of liberating structures: Simple rules to unleash a culture of innovation. Seattle, WA: Liberating Structures Press.

Rogers, P., & Blenko, M. (2006). Who has the D? How clear decision roles enhance organizational performance. Harvard Business Review, 84(1), 52–61.

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

Zampirolo, G., Burtscher, 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