

Governance and Measurement Annex

The companion to the One-Page Pilot Charter. The charter is the one-page decision document the sponsor signs; this annex holds the technical detail that would otherwise crowd it. Complete it during the first two weeks (Design), before any participant enters the system. The reasoning for every element is in The Last Human Coach, Chapter 9.

The chapter names what this annex records: system and model versions; the configuration freeze; the update and rollback process; data flows and subprocessors; retention and deletion; reporting thresholds; scripted safety tests; handoff destinations; the comparison design; baseline measures; evaluation instruments; participant communications; the midpoint review; longer-term follow-up dates; and protocol amendments.

1. System under test — record and freeze at launch

Element	Recorded
Product and system version	_____
Foundation-model version	_____
Methodology configuration	_____
Memory configuration	_____
Available modalities	_____
Safety / escalation configuration	_____
Reporting configuration	_____
Relevant subprocessors	_____
Configuration freeze date	_____

Update and rollback: material changes trigger notification and repeat testing; the buyer can delay or reject an update; a rollback path exists. Agreed with vendor: ☐

2. Data flows and lifecycle

- Data controller / processor roles under applicable law: _____
- Subprocessors, model providers, hosting, analytics receiving data: _____
- Is any coaching content used to train or improve models? _____
- Storage location(s): _____ Retention period: _____ Deletion workflow: _____
- Cross-border transfers: _____
- User rights: inspect ☐ correct ☐ export ☐ delete ☐
- DPIA required and completed (where applicable): ☐

3. Reporting thresholds

- Employer receives aggregate only — no individual transcripts, quotations, inferred traits, risk scores, or person-level themes. ☐
- Minimum group size before aggregate reporting appears: _____
- Small-group suppression rule: _____
- Confirmed: administrators cannot retrieve individual topics or excerpts. ☐

4. Safety, escalation, and handoff

- Scripted safety tests (three levels: within-coaching / needs-human / imminent-risk) defined and run before launch, and continued through the pilot. ☐
- Handoff destinations, by case: crisis resource _____ · EAP/therapist _____ · HR/legal/safeguarding _____ · human coach _____

- Who responds, response time, after-hours path, information transferred, consent step:
-

5. The four evidence categories → three signals

Collect across all four; the three charter signals are drawn from them.

- **Reach and engagement** — invitation, activation, first use, return use, timing, distribution. (*Return is not benefit; heavy use may be value, unresolved need, or a system built to keep talking.*)
- **Developmental relevance and early transfer** — used for genuine situations? affected the next action? later applied or revised? Pre-define one or two observable behaviors: _____
- **Safety, trust, fairness** — technical failures, privacy complaints, inappropriate advice, handoffs, requests for a human, differences across groups.
- **Operational feasibility** — support burden, human-supervision workload, handoff response times, integration, vendor responsiveness, change management, cost.

Satisfaction, relevance, trust, willingness-to-recommend: tracked as experience, not development.

6. Comparison design and baseline

- Comparison approach: ☐ randomized phased invitation ☐ waitlist ☐ constructed comparison ☐ none. Notes: _____
- Selection effects recorded: invited _____ / enrolled _____ / how enrollees differ from the eligible population (where examinable without compromising privacy): _____
- Baseline measures recorded: ☐
- **Retention is not a pilot outcome** (too short, too small). Longer-term audit scheduled — 6-month review: _____ 12-month review: _____. Purpose: audit the decision and design the next study, not bill the pilot for later change.

7. Participant communications — trust rules published before session one

Participants are told: participation is voluntary; they are interacting with AI; what it can and cannot do; what it remembers; how to inspect, correct, delete; who can access content; what reaches the employer; the narrow safety/legal exceptions; how handoffs work; where to take questions. Published before first session: ☐

8. Run and review

- Existing human coaching/support kept in place; the pilot adds a channel, does not replace one. ☐
- Program monitoring separated from coaching content: watch the system, not the individual. ☐
- Reminder discipline: return-rate not turned into a measure of the campaign; encouragement given recorded. ☐
- Time-of-day data: aggregate only, and only if disclosed. ☐
- **Day-45 midpoint review** — can it continue safely, and is the design producing usable evidence? Correct operations; enforce stop rules; no verdict from the adoption curve.

9. Amendments

Decision rules may change only for safety, ethics, or unforeseen operational reality. Document each amendment: the change, the reason, the date, and whether anyone had already seen the relevant results. Preserve the original analysis and decision plan; record the amendment separately.

Date	Amendment	Reason	Results already seen?

The decision itself is made on the one-page charter, read against the criteria written before launch. This annex is what makes that one page trustworthy.

This instrument accompanies Chapter 9 of The Last Human Coach: How AI Brings Coaching to Every Manager by Stefano Calvetti, where the reasoning behind every line is explained in full.