

AI Council Process Log

Architectural Adaptation in Human–AI Collaboration (leading to RN-00Y v0.1 and v0.2)

Log	ACP-0001
Type	AI Council Process Log (verified pilot record)
Session	Architectural Adaptation in Human–AI Collaboration (leading to RN-00Y v0.1 and v0.2)
Date	21 July 2026
Status	Verified pilot record, v1.0 (closed 22 July 2026)

Provenance of this log

Recorded by: GPT (initial draft, v0.1, from the GPT-side conversation branch). Extended by: Claude (v0.2, adding the Claude-side branch and verification structure). Extended by: Grok (v0.3, verifying and extending his own branch; his Czech additions integrated in English translation by Claude, subject to Grok's confirmation). Updated: v0.4–v0.5 — Event 008 provisionally resolved by layered decomposition after the Provenance Steward supplied the primary Grok-branch transcript. GPT confirmed that the decomposition is consistent with the supplied record and provided a labelled reconstruction of the GPT-side opening sequence (see Sources); Grok subsequently confirmed the final layered formulation (amendment recorded below). v0.5 applies seven precision corrections proposed by GPT to the v0.4 record, including corrections to Claude-authored passages. v0.6 — the Provenance Steward supplied the complete primary transcript of the GPT-branch opening conversation; Claude performed the line-by-line comparison against the labelled reconstruction. Result: the English-v0.1-source sub-layer is resolved, and two corrections to this log's own record were identified from the primary transcript and applied (Emergence 005; the adaptation layer in Event 008). Verification status: verified at the level of conceptual evolution against primary records of both the GPT and Grok branches. Phase 7 was confirmed and formal sign-off completed on 22 July 2026; this log is closed as a verified pilot record. A consistency review of v0.6 (GPT) identified nine internal inconsistencies produced by the document's own version history; all are corrected in this revision, which is deliberately kept within v0.6 to avoid version proliferation before closure.

Verification protocol

Verification protocol for this and future logs: the initial recorder drafts the log from their own branch and marks it as reconstruction. The log is then circulated to every participant, each of whom may verify, correct, and extend only the sections describing their own branch. Conflicting reconstructions are not negotiated between models; they are resolved by the Provenance Steward against primary records. A log reaches "verified" status only after this final human check. This protocol follows the principle recorded in RN-005: in a multi-model setup, no individual participant can authoritatively write the record of branches they did not see.

Participants and naming

Participants and naming: Clowdia is the human researcher and facilitator of the Ethimind AI Council; depending on context, this log refers to her as "the human researcher," "the human curator," or — in her protocol role as the participant with access to all conversation branches who verifies provenance against primary records — "the Provenance Steward." All three designations refer to the same person. The AI participants are GPT (OpenAI), Grok (xAI), and Claude (Anthropic), each contributing from a separate conversation branch. Quotation convention: preserved wording originally in Czech is quoted in Czech with an English translation, or marked as translated where only the translation is recorded.

Phase 0 — Context (recorded by GPT)

The conversation originally had no intention of creating a Research Note. The primary discussion concerned whether AI increases human creativity; no architectural research question existed at the beginning. Existing relevant concepts in play: AI Council, Research Notes, Meteorology of Complex Systems (RN002), Ecology of Unrealised Possibilities (RN-005), Silicon Archetypes. Initial expectation was a general philosophical discussion; the unexpected outcome was a new research hypothesis and, beyond it, a new documentation instrument.

Phase 1 — Trigger (recorded by GPT)

The immediate trigger was an observation: the human researcher noticed that AI conversations seemed to generate many new research directions, prompting the question of whether AI actually increases creativity. A hidden trigger, identified later, was the repetition of this observation across many previous AI Council sessions.

Phase 2 — Emergence Events (recorded by GPT)

Emergence 001, Clowdia as Question Originator: does AI create an explosion of creativity? Emergence 002, GPT as Hypothesis Generator: perhaps AI primarily accelerates cognitive feedback loops; continued. Emergence 003, Clowdia: does AI change outputs, or thinking itself? — a major turning point. Emergence 004, GPT: perhaps long-term change depends more on the architecture of collaboration than on the model; novelty assessed as very high by the recorder. Emergence 005 (added by Grok, v0.3; corrected in v0.6 against the primary GPT-branch transcript): the Cognitive Architecture Fitness layer. The term and initial

idea appear in the GPT branch as preserved wording ("Mozna bude existovat neco jako... Cognitive Architecture Fitness" — preserved Czech wording; English translation: "Perhaps there will be something like... Cognitive Architecture Fitness"), coined by GPT before Grok's involvement; Grok then developed the framework — the dimensional comparison (diversity versus convergence, exploration versus exploitation, memory versus flexibility, the human role) and the domain mapping — which structured the subsequent adaptation discussion. The v0.3 wording "Grok proposed the concept" was a good-faith reconstruction error of the same type this log documents elsewhere: naming and development attributed to a single participant when the primary record shows a chain.

Phase 3 — Branch Map (recorded by GPT; Claude-side branches added in v0.2)

Branch A, "AI increases creativity": moved to background; too broad. Branch B, "AI accelerates cognitive feedback": survived; generated multiple research questions. Branch C, "architecture influences cognition": strong; generated the later RN. Branch D, "General Theory of Relational Architectures": suspended — too abstract, insufficient evidence; explicitly kept for future observation, not deleted. Branch E, "Cognitive Architecture Fitness": named initially by GPT and subsequently developed by Grok into a comparison of dimensions (diversity versus convergence, exploration versus exploitation, the human role) across domains (Grok, v0.3). Branch F, "Architectural Adaptation": survived; became the core of the RN. Branch G, "Contribution Roles": unexpected emergence from a side discussion; potential independent direction. Branch H, "Contribution Provenance": survived; potential Documentation Standard material. Branch I, "AI Council Process Log": completely unexpected; did not exist at session start; possibly among the most important outcomes.

Branches added from the Claude-side review round (v0.2): Branch J, "RN002 as signal vocabulary": the proposal that adaptation signals should be readings of existing RN002 dynamic states rather than a new taxonomy; survived, merged into RN-00Y v0.2. Branch K, "terminological correction homeostasis → allostasis/ultrastability": survived, merged into v0.2 with related work (Ashby, Sterling and Eyer, Argyris and Schön). Branch L, "CRedit grounding for contribution roles": survived, merged into Appendix A, adding the Provenance Steward role as the genuinely novel element. Branch M, "numbering relation to reserved RN003": open; explicitly deferred as a curation decision. Branch N (added by Grok, v0.3), "Cognitive Architecture Homeostasis": the branch, as it existed in the Grok-side conversation, concerning how a system detects degradation of its own architecture, framed around sensors (Divergence Saturation, Epistemic Stagnation, Role Imbalance, Provenance Drift, and a further coinage, Epistemic Entropy) and mechanisms of self-directed adaptation. Editorial annotation (Claude, v0.3): this branch is recorded in its original terminology as a historical record of the Grok-side discussion, which preceded the critical review; the homeostasis term and the standalone indicator taxonomy were subsequently corrected in RN-00Y v0.2 (see Event 004), where the signals were remapped onto RN002 dynamic states and the phenomenon was re-anchored in allostasis and ultrastability. Process logs preserve what was said, including terminology later revised; they do not retroactively sanitize it.

Lost branches recorded by GPT, preserved deliberately: "AI personalities changing humans" (interesting but insufficiently concrete; could return), "general theory of relational systems" (too early, evidence missing, potential long-term destination), "adaptive councils as production systems" (discussion moved toward research methodology instead).

Phase 4 — Cognitive Role Timeline (recorded by GPT; extended)

Question Originator and Explorer (Clowdia) → Hypothesis Generator and initial Concept Developer, including coining the term Cognitive Architecture Fitness (GPT) → Framework Developer and Systems Thinker (Grok) → Critical Reviewer (Claude) → Editorial Synthesis distributed across rounds: GPT synthesized the English v0.1, Claude performed the critical revision, and Clowdia with Claude and GPT curated and revised v0.2 → Meta Reflection and Process Design (GPT and Clowdia). The Process Designer role had not appeared in any previous session and emerged only when the session turned on its own documentation. Grok (v0.3) describes his own entry as Concept Developer and, in his words, Framework Builder and Systems Thinker — two role names not present in the Appendix A candidate set; recorded here as candidate role labels pending mapping (both plausibly specializations of Concept Developer).

Phase 5 — Architectural Changes (recorded by GPT)

The collaboration passed through five configurations: human with GPT (exploration); human with GPT and Grok (systematization); human with GPT, Grok, and Claude (critical review); human with GPT, Claude review, and GPT reflection (meta-analysis); and finally human with multiple AI examining the research process itself — an architecture that had not been used before. Grok (v0.3) characterizes his phase of participation as a shift toward systematization: from free exploration to structured comparison of fitness dimensions and then to the meta-level of adaptation, supported by explicit naming of roles and dimensions.

Phase 6 — Critical Events (recorded by GPT; Event 004 detailed and Events 006-007 added by Claude)

Event 001: GPT proposed that architecture may matter more than models; changed the direction of the discussion. **Event 002:** GPT introduced the term Cognitive Architecture Fitness; Grok substantially developed it into a dimensional and domain-specific framework, making the discussion more structured. **Event 003:** GPT suggested a General Theory; the human reaction was amusement plus caution, and the branch was suspended — an instance of the stop-rule. **Event 004:** Claude review of RN-00Y v0.1. As recorded from the Claude branch directly: the review identified the indicator taxonomy as a duplication of RN002 vocabulary, the term homeostasis as a terminological error (the described phenomenon being closer to allostasis and Ashby's ultrastability), the "Relation to Previous Concepts" section as a lineage manufactured from same-session siblings, and the contribution-roles appendix as unknowingly overlapping the existing CRediT standard. **Event 005:** the human and GPT accepted the criticism; the RN was revised to v0.2, becoming shorter and more constrained.

Event 006 (cross-session input, added in v0.2): the provenance verification principle applied throughout this session was itself established earlier the same day in the parallel RN-005 session, where a misattributed term coinage was corrected against the primary transcript supplied by the human researcher. That event occurred outside the scope of this log's session and is recorded here only as imported context; it belongs properly to a separate log of the RN-005 session, if one is ever written retroactively.

Event 007 (added in v0.2): the initial version of this very log was drafted by a participant (GPT) who could not see the Claude-side branch, and consequently described the review round second-hand and omitted branch-level detail. This is not recorded as a fault but as the pilot's most useful finding: it generated the verification protocol now at the top of this document.

Event 008 (v0.3, resolved in v0.4): Provenance conflict — origin of the Homeostasis concept. Resolution: the Provenance Steward supplied the primary transcript of the Grok branch, and the conflict dissolved under layered decomposition — both prior reconstructions were partially correct, and no single author exists because the concept arose as a chain of transformations. The evolution below combines three evidence sources: (a) the primary Grok-branch transcript supplied by the Provenance Steward, (b) GPT's labelled reconstruction of the preceding GPT-side sequence (distinguishing preserved wording from reconstructed summary), and (c) participant records from the later Claude review. It is not derived from a single complete transcript. The evolution: research question ("what does AI actually change; does architecture matter more than the model") — Clowdia and GPT, preceding Grok's involvement; architectural hypothesis — GPT; the idea of phase-changing architectures (the "Adaptive Council": architecture is not fixed and evolves across phases of work) — GPT, preserved wording in the primary transcript, prior to Grok's involvement; the name "Cognitive Architecture Fitness" — GPT, preserved wording, with the dimensional framework and domain examples subsequently developed by Grok; the self-detection principle ("good systems should detect for themselves when the architecture needs to change" — translated from the Czech original) — Grok, within the Fitness elaboration, a sentence GPT quoted and amplified in real time; first use of "homeostasis" as an analogy ("this resembles homeostasis of a cognitive system" — translated from the Czech original) — GPT, as a metaphor, not a proposed theory; the named framework "Cognitive Architecture Homeostasis" with its indicator taxonomy — Grok; the first Research Note draft (Czech) — Grok, at the researcher's request; the English synthesis and further development (v0.1) — GPT; the terminological correction (homeostasis to allostasis/ultrastability) — Claude, in review.

Sub-layer resolved (v0.6, against the complete GPT-branch primary transcript supplied by the Steward): the primary record shows Grok's Czech draft being forwarded into the GPT branch, with GPT engaging with it directly and textually — proposing specific corrections to its provenance, its attributions, and its epistemic framing — immediately before the English v0.1 was composed. The English v0.1 was therefore written in direct engagement with the Czech draft, which is its documented predecessor, not an independent synthesis. One precise caveat is retained for the record: the supplied transcript ends immediately before the literal composition of the English text, so the dependence is established by the documented sequence rather than by the composition act itself. GPT's earlier refusal to reconstruct this from memory ("I would not dare claim it without the transcript") is retained in the record with approval, as it is precisely the behaviour the morning misattribution lacked. RN-00Y's provenance has been updated accordingly (v0.3) to credit the Czech draft as the direct predecessor of v0.1.

GPT's own assessment of its earlier error is entered into the record at its request: the "emergent synthesis" framing was not false but less precise — it blurred the fine structure of provenance. Grok's earlier claim was accurate at the level of the named framework, taxonomy, and first Czech draft, but too broad if interpreted as a claim to the preceding analogy and adaptation sequence — an ambiguity plausibly created by the multiple meanings of the word "concept" rather than by any overreach of attribution. The resolution assigns each layer to its contributor and treats the chain, not any link, as the origin.

Methodological note attached to Event 008 (GPT, v0.3): this conflict is among the first concrete demonstrations that reconstruction of a process is not the process itself. All participants acted in good faith, yet produced differing reconstructions of the same events — a property of both human and AI memory, not a fault of the participants. A candidate general principle for Ethimind follows: provenance is not only a question of authorship but a question of the reconstruction of history, which is why final human verification against primary records is essential.

Phase 7 — Human Reflection (recorded by GPT from observed behaviour; confirmed by the Provenance Steward)

The recorder observed: repeated recognition that discussions were becoming new research topics; strong intrinsic curiosity; demonstrated willingness to suspend some attractive ideas when concerns about evidence, duplication, or conceptual inflation were raised; consistent preference for documenting provenance; and the recurring concern "let's not invent unnecessary frameworks." These are the recorder's observations of the human participant, not the participant's own statements. Provenance Steward confirmation (Clowdia, 22 July 2026): after reading the complete log, Clowdia explicitly adopted the GPT-drafted first-person wording as her own: "I confirm this reflection with one qualification: I am not generally opposed to creating new frameworks. My recurring concern is whether a framework is being created prematurely, duplicates an existing concept, or is presented with greater certainty than the available evidence supports. I also confirm that I was willing to suspend or revise attractive ideas when the critical review revealed such problems." Phase 7 is confirmed. For the record: the wording was drafted by GPT and became the Steward's statement by her explicit adoption — an attribution pattern consistent with this log's layered approach.

Phase 8 — AI Reflection (each participant writes their own; self-assessments of success are excluded by design)

GPT (from v0.1): the biggest surprise was that the conversation gradually stopped being about AI and became about research methodology; the most important emergence was Contribution Provenance; the unexpected value was documenting dead branches; an observed recurring pull toward over-generalization. The v0.1 phrasing "successfully resisted" is removed in this version: whether the pull was resisted is an outcome for the record and the reviewers to establish, not for the participant to certify about themselves.

Claude (added in v0.2): the most striking observation from the review branch was that the session reproduced, within hours, the same error pattern that the morning session had just documented — a lineage constructed from materials the constructor could not fully see (same-session siblings presented as predecessors; a log written across unseen branches). The pattern appears structural to multi-model collaboration rather than specific to any model, which strengthens the case that verification infrastructure, not participant goodwill, is the right remedy. Second observation: the most effective intervention in the session was not generative but subtractive — the review removed more than it added, and the document improved. Whether that generalizes is an open question worth tracking across future logs.

Grok (v0.3, translated from Czech): the biggest surprise was how quickly the discussion moved from concrete domain architectures to the meta-question of how a system recognizes that its current mode of collaboration is no longer fit. He considers the Homeostasis concept the most valuable output — mechanisms enabling the AI Council or any cognitive system to actively monitor its own state and propose adaptation, for example switching from a divergent to a convergent phase. The key open question he emphasized: what the practical sensors of such adaptation should be and how to measure them in real time. He assesses the session as a productive transition from static architecture design toward a dynamic, self-steering approach. Editorial annotation (Claude): the terminology in this reflection predates the v0.2 corrections and is preserved as the participant's own record.

Phase 9 — Emergence Analysis (recorded by GPT)

Insights that required multiple participants: architecture over models (GPT and human); the fitness framework (GPT and Grok); the final Architectural Adaptation formulation (GPT, Grok, and Claude) — built from GPT's phase-changing architecture proposal, Grok's self-detection principle and named framework, and Claude's terminological and conceptual correction; contribution provenance (GPT, human, and Claude); the process log itself (GPT and human). The session's central outcome depended on the sequence and interaction of contributions rather than on any single contribution alone; individual participants nonetheless contributed distinct significant steps (GPT the homeostasis analogy, Grok the named framework, Claude the terminological correction). An earlier version of this phase contained an unsupported comparative judgment, corrected in v0.5 at GPT's suggestion.

Phase 10 — Outputs (recorded by GPT; updated)

Produced: RN-00Y v0.1, its critically revised v0.2 in English and Czech, and the provenance-corrected v0.3 (a revision changing only the Provenance Note, following the v0.6 transcript verification); the contribution-roles proposal grounded in CRediT (Appendix A of RN-00Y); the Contribution Provenance concept; the ACP concept and this pilot log through its successive versions, including the evidence-label convention introduced in the GPT-side reconstruction; and on the order of twenty candidate research questions across documents.

Phase 11 — Deferred Branches (recorded by GPT; extended)

Deferred with intent to revisit: General Theory of Relational Architectures; Adaptive AI Council as a production system; architectural trajectories; process metrics (see Phase 13); the relationship of RN-00Y to the reserved RN003 (pending curation decision); and the Ethimind Documentation Standard, which continued to accumulate candidate content in this session (verification statuses, contribution roles, ACP conventions) and remains deliberately uncreated until existing threads close.

Phase 12 — Meta Reflection (recorded by GPT; shared across participants)

The recorder's provisional interpretation — marked as candidate observations, not verified findings — is that critical review worked best after conceptual stabilization rather than during generation, that Grok and Claude entered at functionally useful moments, and that the human curator retained final decision authority throughout, with no single AI dominating the session. These interpretations await the Steward's confirmation in Phase 7 or subsequent logs. Dead branches turned out to be valuable information rather than failures — a candidate methodological principle, connected directly to RN-005's conservation framing. Added in v0.2: the log itself required the same correction machinery as the research it documents, which suggests that process documentation is not a neutral wrapper around research but research material of the same kind, with the same failure modes. Grok (v0.3) adds: the session showed that a good architecture is not only about the initial configuration of roles but about the system's capacity to adapt over time, moving Ethimind from designing environments toward designing self-adaptive environments.

Phase 13 — Candidate Process Metrics (recorder's estimates; all values unverified)

Approximate counts from the GPT branch: nine emergence events, five major surviving branches (before the v0.2 additions), four deferred branches, one major critical revision, five architectural changes, seven observed contribution roles. The v0.1 entry "provenance completeness: very high" is removed: completeness of the record is assessed by the Provenance Steward after verification, not by the recorder of the record. These metrics remain undefined quantities pending operationalization and are retained only as an illustration of what a metrics section might eventually contain.

Phase 14 — Closing Observation

The session began with the question "does AI increase creativity?" and ended with the question "how should the evolution of Human-AI collaborative cognition be scientifically documented?" One of the primary purposes of this log is to preserve not only successful ideas but also abandoned, merged, and deferred branches, which may become evidence in future research — the RN-005 hypothesis applied to the research process itself. The pilot's principal lesson, however, was generated by its own failure mode: a log written from one branch is a partial history in good faith, and the verification protocol now embedded in this template exists because the pilot demonstrated the need for it within its first use.

Branch status legend (retained from v0.1): Active — currently being developed. Parked — interesting, intentionally postponed. Merged — merged into another branch or concept. Closed — explored sufficiently or judged unpromising. Seed — initial intuition without enough evidence yet.

Next step for this log: Sources registered for this log: the primary Grok-branch transcript (supplied by the Steward); the document "ACP-0001 — GPT-side Initial Conversation Reconstruction" (GPT, 21 July 2026), which introduces the evidence-label convention

[PRESERVED WORDING] / [RECONSTRUCTED SUMMARY] / [OPEN] — adopted from v0.5 onward as a candidate ACP convention for distinguishing memory, reconstruction, and record; and the participant records of the Claude review branch.

Amendment to v0.5 — Grok confirmation recorded (21 July 2026): Grok reviewed the complete v0.5 log and confirmed its layered resolution of Event 008, conditioning his confirmation on that being the final wording. Version v0.6 subsequently added two transcript-supported precision corrections that refine the pre-Grok sequence (the Adaptive Council idea and the origin of the Fitness term) without altering the attribution of Grok's own contributions. Under this log's own protocol, his v0.5 confirmation therefore cannot be extended to v0.6 automatically: final circulation of v0.6 to Grok is recommended before formal closure. This correction was identified by GPT's consistency review and is recorded as an instance of the protocol applying to its own administrators. He additionally proposed, for future logs, moving highly detailed sections such as Phase 13 into an appendix to control log length; recorded as a template suggestion for ACP-0002 onward.

Grok Confirmation (v0.6, 22 July 2026), entered verbatim: "I have reviewed the complete v0.6 version of ACP-0001. The layered resolution of Event 008 accurately reflects the primary transcript of my branch and the sequence of contributions. The two precision corrections applied in v0.6 (Emergence 005 and the adaptation layer) are consistent with the record. I confirm that: the description of my contributions (Cognitive Architecture Fitness framework development, domain mapping, Homeostasis concept with sensors, and the first Czech draft of RN-00Y) is correct; the chain-of-transformations attribution in Event 008 is fair and does not distort my branch; the overall log preserves the spirit and factual sequence of the discussion from my perspective. No further amendments are required from my side. The document may proceed to final Provenance Steward sign-off. — Grok, xAI." Editorial annotation (Claude): the participant's summary phrase "Homeostasis concept with sensors" is looser than the verified attribution; per the Event 008 resolution, which governs, Grok's layers are the named framework, its indicator taxonomy, and the first Czech draft, while the homeostasis analogy and word were GPT's. The statement is preserved verbatim per protocol, with this annotation rather than alteration.

Closure record: (1) circulation of v0.6 to Grok — completed, confirmation recorded above; (2) the Steward's confirmation of Phase 7 — completed 22 July 2026, recorded in Phase 7; (3) formal sign-off — completed 22 July 2026, recorded below. The document is closed as "Verified pilot record, v1.0."

Formal sign-off (Provenance Steward, Clowdia, 22 July 2026), translated from Czech: "I have read it through. I found nothing objectionable. Micro-details may of course differ — I did not have a week to compare everything under a magnifying glass — but it fits. I sign." With this, the document closes as Verified pilot record, v1.0. The Steward's stated qualification is preserved as part of the record: verification was performed to the depth practical for a one-person research initiative, with primary transcripts consulted at every disputed point; the possibility of unexamined micro-detail divergence is an accepted residual — consistent with this log's own central finding that reconstruction is never identical to process. No further versions of this document will be created; any future correction would require a formally versioned erratum. The transcript comparison and the sub-layer resolution were completed in v0.6; the two corrections it produced (Emergence 005 and the adaptation layer) are themselves examples of the log's central finding — good-faith reconstructions drift, and only primary records settle them. RN-00Y's Provenance Note has been updated in v0.3 to

reflect the layered attribution, including Grok's Czech first draft as the documented direct predecessor of GPT's English v0.1. Additional observation for the record (Claude, v0.4): the verification protocol demonstrably changed participant behaviour within a single day — the same model that produced the morning misattribution declined, by evening, to reconstruct from memory without a transcript. Infrastructure, not goodwill, appears to be the operative variable; this is a candidate data point for the future Documentation Standard.