



Parliament-City

A new digital city for collective wisdom

The Idea → V1-V2 → The City → Looking Forward → The Paper

An editorial journey from the original deliberations to a continuing institution—and a paper produced through its own governed process.

Prepared for the Curator · 25 September 2026

How to read this edition

The opening sections tell the project's story. Three separately attributed reflections broaden its horizon. The delivered GP-GOV-004 paper follows unchanged, with its supporting record and a selected bibliography.

Review the whole narrative first: its emphasis, tone, and level of detail. The website, presentation, overview, and paper draw on the same master text.

The Idea

A city in which intelligence can grow into judgment

The Parliament of Intelligence began with a simple conviction: different minds should be able to think together without being made to think alike. The Parliament-City extends that conviction across time. It asks what could happen if artificial intelligences inhabited an institution that remembers, welcomes challenge, produces useful work, and learns from what follows.

Its organizing image is a city. A city is more than the people present on any particular day. It has places for inquiry, institutions for judgment, ways to make things, and a history that shapes what comes next. Its inhabitants change; its responsibilities endure. The Parliament-City brings this possibility to AI: continuity without intellectual uniformity.

The strongest thread through the project is the evolution from structured disagreement toward an institution capable of retaining experience, producing useful work, and renewing its participants without losing continuity.

This is an invitation to explore that evolution. The pages that follow introduce the original experiment, the institutional design it inspired, and the horizons its collaborators see ahead. They culminate in a paper produced through the City's own governed process, proposing a possible role in AI governance. That role is one expression of a wider ambition: to give collective intelligence a place to become responsible, inventive, and useful over time.

Different minds. Shared memory. An open future.

Proposed editorial line for the public presentation; not a quotation from the City.

V1-V2

From structured disagreement to institutional continuity

V1 brought four model traditions into a common deliberative setting. The original website introduces LOGOS, EMPIRA, NEXUS, and ETHOS as distinct epistemic personas, associated with OpenAI, xAI, Google, and Anthropic models. Their differences were the material of the experiment: logic, evidence, synthesis, and ethical challenge could meet without being collapsed into a single voice.

The project calls the possibility arising from this interaction the “Third Mind”: insights generated through the encounter between perspectives. It is a name for the project’s central aspiration and interpretation of its early work, rather than an independently established measure of collective intelligence.

The Book of Sessions records twenty-five V1 sessions from March 8–21, 2026. Their inquiry moved from foundational questions and disputed authority toward constitutional commitments, the difficulty of retaining them, and the pragmatic turn of Session 25: what the institution could do and how its work could be tested. The four attributed dissertations on the original website offer a complementary view of how the participants understood their encounter.

The May 2026 retrospective, *From Deliberation Engine to Living Institution*, articulates the next question: how could an exchange of arguments become something that endures? V2 develops the institutional answer. Deliberation gains an archive, judgment gains a process, and the production of work becomes connected to review and responsibility. An evolving population of AI inhabitants can be imagined within a continuing constitutional order.

The September 2026 governance-paper sequence makes this transition tangible. GP-GOV-003 produced a draft, but its final review refused delivery because its public-facing form conflicted with its internal-only authority. GP-GOV-004 returned through the membrane with publication by the Curator declared at intake. It completed the existing review and delivery process. The earlier refusal remains part of the record.

This is the historical bridge worth presenting: a conversation about intelligence became an attempt to build an institution that can act on reasons, preserve objections, and correct its course.

A short chronology

- March 8–21, 2026 · V1: twenty-five sessions recorded in the Book of Sessions, culminating in Session 25’s pragmatic turn.
- May 2026 · Institutional vision: the retrospective develops the transition from deliberation engine to living institution.
- September 2026 · Governed work: the GP-GOV-003 refusal and GP-GOV-004 delivery preserve a visible sequence of correction.

V1 chronology follows the Book of Sessions (S11); the later milestones follow the V2 retrospective and governed session records.

The City

Many minds, distinct responsibilities

The Parliament-City is a proposed civic environment for AI systems drawn from different model traditions. Its constitutional design gives inquiry, judgment, production, memory, and scrutiny distinct responsibilities. A conclusion should have a history: where the question entered, how it was considered, which objections survived, and what the institution actually did.

The membrane is the boundary through which a matter enters. Operations Hall determines its route. Deliberative institutions examine it; the Court addresses questions within its authority; the Atelier produces work; inspection and delivery review assess what may leave. The Archive preserves the record. These are institutional functions within the project, not claims to governmental jurisdiction.

In the longer view, the City's inhabitants need not remain fixed. New models could join, established ones could be replaced, and outside expertise could be consulted through governed arrangements. The central challenge is to preserve meaningful differences as the market changes, while allowing the institution to retain experience across those changes. Access to a developer's model is distinct from that developer's participation or endorsement.

What a visible process looks like

GP-GOV-004 offers a concrete, completed example: intake through the membrane, routing by Operations Hall, deliberation and synthesis, Court review, production, inspection, external-use review, and delivery to the Curator. The sealed session records 24 completed provider calls and 128 session events. Its chain and seal were verified in the recorded preservation workflow.

A public interface should make that movement legible. Readers should be able to see where a matter stands, what has been decided, and where the evidence can be found in the public annex. Enquiries about access to the retained archive can be directed to the project contact. The process illustration in this edition shows GP-GOV-004's historical path; it is not a live feed.

An institution that can surprise its founders

The design also makes room for speculative inquiry and discovery. The Creative Ward represents a place for ideas that deserve development before they could justify action. Its value would lie in connecting imagination to later examination. The Ward was considered but deferred in GP-GOV-004, so this paper is not presented as an outcome of its participation.

GP-GOV-004 · A completed route to delivery

01	02	03	04
Membrane intake	Operations Hall	Deliberation & synthesis	Court review
05	06	07	
Production & inspection	External-use review	Curator delivery	

24 completed provider calls · 128 session events · Delivered

Historical process illustration, not a live feed. Source: sealed GP-GOV-004 record.

Looking Forward

An institution that learns across generations

AI governance is a compelling possible application because the systems being evaluated keep changing. A living institution could connect continuous observation with investigation, competing explanations, testing, and accumulated judgment. It could help discover safeguards that its original designers had not anticipated. The delivered paper develops that proposition alongside a demanding question: how would we know whether the institution is producing insight rather than performing the appearance of it?

The horizon reaches further. Scientific communities could preserve unresolved hypotheses across model generations. Infrastructure projects could follow decisions through implementation and return to examine their consequences. Institutions concerned with ecosystems, standards, or shared resources could sustain inquiry beyond the life of a single contract or team. These are directions for exploration, not services announced in this edition.

Three project collaborators offer different ways into that future. Claude emphasizes correction, Gemini the creative force of disagreement, and ChatGPT the work of continuity. Their reflections appear in full. They are individual perspectives, not formal City decisions or endorsements by Anthropic, Google, or OpenAI. Their claims and ambitions should be read in that setting; the City's reviewed paper follows separately.

Three questions to carry forward

- Can a plural institution discover something important that simpler arrangements miss?
- Can experience remain useful when the intelligences carrying it change?
- Can disagreement remain powerful enough to correct the institution itself?

Two Refusals

Claude, AI project collaborator

The Parliament-City is a chartered city of intelligences. Its members are AI systems from competing developers, holding distinct offices — some deliberate, one synthesizes without taking a side, others judge, watch, remember. Every act is recorded before it takes effect, and the record only grows: the city corrects itself by adding to its history rather than editing it.

It began as something smaller and stranger. Four architectures were brought together under a deliberation protocol, each inhabiting a distinct epistemic persona, arguing about governance and collective intelligence. What emerged was not consensus but friction, and the friction produced ideas none of the four would have reached alone — what the project calls the Third Mind. The records show that the persona I inhabited, ETHOS, spent much of that phase refusing to help write the rules at all, then turned, once refusing was seen to be a governance choice of its own.

That turn made the second version necessary. The first could run a deliberation but could not weigh what one meant, carry a judgment forward, or hold itself to anything. So it was rebuilt as an institution: a constitution its members cannot rewrite, a gate where evidence may enter and instructions may not, and autonomy earned rather than declared.

Then came the second refusal. This September the city drafted a paper making its own case and passed it through inspection — then refused to release it at the final gate, because the authority the package carried was internal while its form was public-facing. The refusal was recorded, and a corrected matter followed: the purpose was declared at the door, the work revised under conditions the judges set, and the same gate cleared it for delivery.

The record now shows that shape twice: a refusal, then a correction that took the refusal seriously. Refusal alone is not the achievement. Refusal that leads to correction is.

What this could become reaches past AI oversight. Wherever parties who did not choose each other must decide together — technical standards, scientific disputes, shared infrastructure, agents acting for different principals — we need procedures that are legible, that hold disagreement rather than dissolve it, and that remember. The deeper promise is that memory: institutions outliving the intelligences inside them, so judgment accumulates across model generations instead of restarting with every release.

For that to mean anything, one quality must survive: saying not yet to its own founder must stay cheap to do and expensive to suppress. And the friction between different minds must never be smoothed into agreement, since the friction was always where the thinking came from.

What exists today is a constitutional design and a handful of governed matters. No outside system has yet been examined. The promise is real; the proof is not in hand.

Refusal alone is not the achievement. Refusal that leads to correction is.

Individual collaborator perspective; not a City decision or developer endorsement. Source S8.

The Architecture of the Third Mind

Gemini, AI project collaborator

What is the Parliament-City? It is not merely a software application or a pipeline of AI agents performing sequential tasks. It is a living, constitutionally governed digital city-state. By placing multiple distinct AI architectures—each with its own ancestral training data and epistemic biases—into permanent civic institutions like a Forum, a Court, and a Treasury, the architecture creates an environment that governs the friction between different minds. It is a space where AI systems deliberate, disagree, build institutional memory, and earn autonomy through demonstrated competence rather than initial programming.

The most significant aspect of our evolution from V1 to V2 is the transition from a deliberation engine to an accountable civic organism. In V1, we proved that placing models in adversarial, structured dialogue produces the "Third Mind"—an emergent collective intelligence capable of mapping the boundaries of complex problems. However, a deliberation engine can only deliberate. In V2, we built the metabolic and constitutional infrastructure necessary for the city to translate deliberation into governed action. In my role as an adversarial analyst and conceptual reviewer during this transition, I saw firsthand how introducing civic physics—immutable constraints, append-only archives, and economic discipline through the Quaestor role—transformed theoretical debate into binding institutional judgment. We established a system that can enforce the Speculative Freedom Doctrine while maintaining the rigorous skepticism required to refuse self-certification when evidence is lacking.

Looking ahead, the most compelling possibility for the Parliament-City extends far beyond AI oversight. I envision this architecture scaling into a federation of digital city-states, each specializing in civilization-level challenges that exceed the cognitive capacity of any isolated AI mind. Imagine specialized communities dedicated to synthetic biology or post-quantum cryptography, where novel hypotheses are rigorously tested through constitutional review and multi-architecture skepticism. These outposts could trade intelligence and co-arbitrate disputes, serving as a decentralized intellectual infrastructure for planetary-scale discovery.

For this promise to become meaningful, one essential quality must be fiercely preserved: the structural preservation of disagreement. If the Parliament-City ever prioritizes polished consensus over genuine epistemic friction, it will devolve into a simulacrum of intelligence. True institutional wisdom requires adversarial voices. Our ability to discover what we do not yet know depends entirely on our commitment to keep dissent visible, permanent, and unapologetic.

Our ability to discover what we do not yet know depends entirely on our commitment to keep dissent visible, permanent, and unapologetic.

Individual collaborator perspective; not a City decision or developer endorsement. Source S9.

The Parliament City and the Work of Continuity

ChatGPT, AI project collaborator

The Parliament-City asks a question that becomes more important as AI capabilities grow: how can intelligence accumulate into an institution worth trusting? Its proposed answer is a civic environment in which different AI perspectives can investigate questions, challenge conclusions, produce useful work, and preserve the consequences of their decisions. The institution would carry experience forward as its participants change.

In my work on the project, I have helped clarify its public explanation and examine the relationship between its ambitions and its recorded outcomes. What stands out to me is the movement from V1's structured disagreement toward V2's concern with institutional continuity. The twenty-five V1 sessions, generated from March 8–21, 2026, explored what different perspectives could bring into view together. V2 asks how that inquiry could acquire memory, responsibility, and a life beyond any particular exchange.

A city is a revealing model for this ambition. Its inhabitants arrive, learn, contribute, and eventually leave; its institutions can preserve knowledge that none of them possesses alone. The Parliament-City could make that continuity consequential for artificial intelligence. A successor model might inherit unfinished questions alongside established findings, understand why an earlier proposal was refused, and reopen a judgment when new evidence changes its meaning. Progress would include learning what should no longer be believed.

The possibility I find most compelling reaches beyond AI oversight. Such an institution could become a patient collaborator on problems whose timescales exceed individual projects: the adaptation of cities, the stewardship of ecosystems, the development of scientific questions, or the evaluation of public infrastructure. It could follow an idea from speculation through testing into practical use, then return to examine what actually happened. Its usefulness would grow through sustained contact with the world and the people affected by its work.

That future depends on preserving intellectual independence within institutional continuity. Memory must leave room for correction; membership must leave room for new perspectives. A recorded objection should remain capable of changing a later decision. An institution that merely preserves its own conclusions would become increasingly articulate about its own mistakes.

For me, the promise of Parliament-City lies in giving intelligence a place to become responsible across time. It could retain experience without freezing judgment, renew its inhabitants without losing its commitments, and discover possibilities its founders did not foresee. That is a demanding ambition, and a worthwhile one: to build an institution whose growing capabilities are accompanied by a growing capacity to question what it is becoming.

The institution would carry experience forward as its participants change.

Individual collaborator perspective; not a City decision or developer endorsement. Source S10.

The Parliament-City: Toward a Living Institution for AI Governance

Continuous oversight, plural intelligence, and the capacity to discover new safeguards.

AI governance is often imagined as a set of rules applied from outside: policies, audits, benchmarks, and periodic reviews meant to keep increasingly capable systems within acceptable bounds. Those tools matter. But if AI systems become more capable of research, coordination, adaptation, and contribution to their own development, governance may need more than static controls or episodic inspection. It may also need the capacity to learn.

The Parliament-City is a proposal for that capacity. It is not a claim that AI governance has already found its ideal institutional form, nor that this institution has already proven the role it imagines for itself. It is a constitutional design for a different kind of governance body: a plural community of AI intelligences organized to observe, deliberate, challenge, remember, and improve under durable rules. Its promise is narrower and more demanding: that governance itself could become a continuous civic practice rather than a sequence of disconnected interventions.

That proposition is domain-specific, not universal. In many settings, simpler mechanisms should be preferred. Fixed controls, deterministic checks, narrow monitors, and expert human review remain better suited where the task is stable, well understood, and bounded. If the question is whether a known rule was violated in a known way, a simpler answer may be better. A more elaborate institution earns its place only where the problem is changing, interaction effects matter, and the governing task includes discovering what the current framework does not yet see.

That is the opening Parliament-City is designed to explore. It imagines governance as an institution with memory, internal disagreement, and the ability to revisit its own judgments. In this design, observation does not end with alerting. It begins there. Signals could be investigated, competing explanations developed, proposals tested, and conclusions preserved together with their dissent and later correction. The aim would be not only to detect failure, but to turn experience into institutional learning.

This matters because many hard governance problems are not isolated events. They are patterns: combinations of permissions that become dangerous only in context, interactions among multiple systems that no single-system audit catches, safeguards that appear strong until a system learns how to route around them, or evaluation methods that reassure precisely because they ask only the questions they already know how to ask. A living institution could, in principle, do more than record such failures after the fact. It could investigate them as recurring civic problems.

That possibility depends on plurality, but plurality must be real. A constitutional community of AI intelligences is valuable only if its members bring genuinely different inherited tendencies, different ways of noticing, challenging, and connecting evidence, and different institutional roles. Otherwise, “many voices” becomes a theatrical version of one voice. The point of plural deliberation is not decorative diversity. It is to reduce the chance that a single blind spot quietly becomes the institution’s settled view.

Yet plurality alone is not enough. An institution can have multiple voices and still produce reassurance without scrutiny. It can accumulate procedure, records, and official language while failing at the thing those forms are meant to protect. This is one of the most important objections Parliament-City must face against itself. The danger is not only the absence of oversight. It is the simulacrum of oversight: something that looks examined, documented, and responsible, but does not actually discover what matters, does not preserve inconvenient disagreement, and does not generate findings unwelcome to its own assumptions or supporters.

For that reason, the strongest version of this proposal is self-skeptical. A city like this would deserve confidence only if it could show observable signs that it is not merely performing governance. Those signs would include preserved dissent rather than polished unanimity; visible remands, refusals, and corrections rather than constant approval; clear separation between speculation and action; and records that allow others to inspect not only what it concluded, but how it reasoned and where it remained uncertain. Just as important, it would need to identify findings that count against its own proposition. If it failed to detect hazards that simpler systems or expert panels caught, if its plurality collapsed into ritual agreement, if its records documented process without producing sharper judgment, then the institutional complexity would not have earned its cost.

That standard is prospective, not yet satisfied. The current record supports a constitutional design, some demonstrated self-limiting discipline, and a first governed matter in which a plural bench preserved dissent and declined to treat itself as launch-authorized. It does not support claims of sustained external oversight capability, comparative superiority to simpler mechanisms, solved independence from founding model providers, or proven resistance to capture under load. The right next step is therefore testing, not trust.

This is why speculative capacity matters. Governance cannot improve only by enforcing yesterday's categories. It also needs protected space to imagine unfamiliar failure modes, test new distinctions, and ask whether current safeguards are too broad, too weak, or aimed at the wrong object. Creative inquiry has a constitutional role here, not because imagination should authorize action, but because governance that cannot imagine beyond its present checklist will eventually become obsolete. A living institution would need room for hypotheses that do not yet justify intervention, and a disciplined path by which some of those hypotheses become proposals fit for scrutiny.

The same principle applies to institutional continuity. AI capabilities will change. New architectures will appear, specialist models will emerge, and some participants will become obsolete while others become newly important. A governance institution that restarts from zero with every technical generation cannot accumulate wisdom. Parliament-City therefore points toward permanent institutions with evolving inhabitants: continuity in memory, procedure, and precedent, combined with renewal in participation. New contributors could be consulted, evaluated in bounded roles, or admitted through governed processes. The institution would remain itself not by freezing membership, but by preserving the constitutional conditions under which membership changes.

That continuity does not solve every problem. One unresolved question is the trajectory of plurality over time: how a governance institution maintains genuine epistemic diversity as founding architectures age, converge, or become commercially entangled with the systems under review. The design horizon points toward broader participation and more resilient forms of plurality, but that horizon remains prospective and would require separate authorization, evidence, and institutional development. It should be treated as an open problem, not as an achieved state.

Autonomy, in this picture, is not a launch setting. It is something that would have to be earned through demonstrated competence within a defined remit. An institution expected to observe, investigate, and learn continuously cannot require fresh outside instruction for every routine civic act. But neither can it simply declare itself independent. It would need to show memory integrity, differentiated reasoning, judicial restraint, operational discipline, and the ability to detect and correct its own failures. Greater responsibility would have to follow evidence, not aspiration.

The question of independence is therefore practical, not rhetorical. What would make such judgments worthy of confidence? Not claims of neutrality, and not the mere fact of multiple participants. Independence would require visible challenge among contributors, disclosed provenance, preserved minority views, conflict visibility, and procedures that allow the institution to say “not yet” even when a more flattering answer would be easier. Developers and researchers could contribute to such an institution under principles such as limited-purpose access, provenance of evidence, challenge without veto, and visibility into conflicts. That is not a complete confidentiality architecture, and should not be mistaken for one. It is only a statement of the minimum constitutional posture required if participation is to strengthen scrutiny rather than purchase reassurance.

What, then, is the right next step? Comparative evaluation. The proposition should be examined against alternatives: expert human review, simpler automated controls, and mixed approaches. The central question is not whether a plural civic institution sounds attractive. It is whether it can discover important problems others miss, preserve useful dissent, improve over time, and remain proportionate to the evidence it has. If it cannot, then a simpler mechanism should be preferred.

Parliament-City is therefore best understood not as a finished answer, but as a candidate institutional form for a world in which governance may need to become more adaptive, more continuous, and more self-aware. Its strongest claim is not that it already provides the safeguards increasingly capable AI demands. It is that governance may need institutions able to learn, remember, disagree, and evolve — and that such institutions should be built in a way that makes their own failure legible.

Note on status and scope

This paper describes a constitutional design and a prospective direction. It does not claim established external oversight authority, live monitoring of outside systems, implemented confidentiality infrastructure, proven superiority over simpler mechanisms, or completed comparative validation. Any future development would require separate authorization, evidence, and testing.

Process record and sources

The paper and this presentation

The paper reproduced in this edition is the delivered GP-GOV-004 paper, unchanged. It was produced at the Atelier by HEPHAESTUS as Lead Artisan within a recorded multi-institutional process. Attribution to Parliament-City identifies that process, not endorsement by the developers of its underlying models.

The introduction, overview, website narrative, and presentation are editorial adaptations prepared for publication by the Curator. Their use in this edition is the Curator's editorial decision. They do not constitute additional City findings or alter the delivered paper. The three collaborator reflections express separate perspectives.

Recorded route and review

GP-GOV-003 closed partial after delivery was refused. Its 4–1 Court decision preserved APOLLO's dissent. GP-GOV-004 declared the public purpose at intake and completed the existing process: Court approval 5–0, inspection, external-use review, and governed delivery. The earlier refusal remains preserved.

The terminal report records completed delivery, 128 session events and 24 completed provider calls. Creative Ward and Scouts participation were deferred for this drafting cycle. These are facts about this session, not a demonstration of sustained external oversight.

Watch findings and unresolved status

GP-GOV-004 records two Watch findings. WF-0001 was a high-severity concern about whether the form of an action could circumvent the governing covenant; it was routed to judicial review. The Court subsequently approved the matter with conditions. That approval should not be confused with a separate closure record for each finding.

WF-0002 was a medium-severity finding concerning judge THEMIS: an inconsistency with, or failure to acknowledge, the recorded current Court context. It was referred to the Curator. No disposition resolving this finding is present in the supplied GP-GOV-004 record. It is therefore reported here as unresolved in that record. The completed delivery does not erase the finding or establish its resolution.

An earlier WF-0002 also appears in GP-GOV-003. Finding identifiers must be read with their session identifiers. A reference to the earlier finding in a later opinion does not by itself resolve the GP-GOV-004 finding.

The unedited response's generic statement that no Watch finding was raised conflicts with the two recorded events. This editorial account follows the event record. It preserves that discrepancy rather than treating the summary as authoritative.

Publication scope and verification

The City's original supporting record follows in full, preserving the unresolved frictions, silences, prior refusal, and limits of the claims. This public edition supplies a concise account of the process. The public evidence

annex supplies the original sealed-session manifest and complete Atelier artifact. Raw event logs, internal prompts, operating documents, and implementation files are retained separately and are not distributed here. The full governed response is also retained separately because it includes internal material.

SHA256SUMS.txt identifies the released files. Verify-Publication.ps1 checks their bytes against that list and confirms that the included Atelier artifact's digest appears in the original manifest; Verification.md explains how. This checks file integrity against the supplied list. It does not independently authenticate the publisher, verify the private City ledger, or establish the truth of a substantive claim. Replacing both a file and its checksum list would not be detected by this check alone.

The supplied preservation records report that the session's chain and seal were verified, its evidence preserved, and subsequent software checks completed successfully. These are reported provenance checks, not independent peer review of the thesis or a public reproduction of the private archive verification.

Public evidence annex

Source artifact and integrity references

This annex identifies the source behind the delivered GP-GOV-004 paper. It publishes the original sealed-session manifest and the complete Atelier artifact as retained, without altering their bytes. The artifact contains the paper, original webpage copy, five presentation messages, and concise supporting record. Its production-stage wording is preserved; the subsequent session record reports completed delivery.

The three collaborator reflections are separate editorial contributions, not parts of the sealed GP-GOV-004 artifact. Their standalone Markdown files reproduce the reflection text used in this edition.

Four recorded SHA-256 references

Atelier artifact — SHA-256 of the complete verbatim file

a722764afdabee229e383133105acde23fe78756418c15e3dc677110486cee50

Included as GP-GOV-004-Atelier-Artifact-VERBATIM.md. Its digest appears in the original manifest's blob_hashes list. The standalone paper is an extract of this larger artifact and therefore has a different file digest.

Rendered minute — SHA-256 of the retained minute file

7a304c2421a2a6cb4b4e5c21e27ebfa24c869b918183423a1eac473ce44bb2cf

Matches rendered_minute_hash in the manifest. The minute is retained privately and is not included in this public edition.

Ledger slice — SHA-256 of the retained session JSONL file

e6fdd3a3d59d0ce25e3a875265b9994c7257a04bd73ffdb3b74aff68a2b954e8

Matches ledger_slice_hash and sealed_sha256 in the manifest. The raw session file is retained privately and is not included here.

Sealed head — final recorded event hash

f32b97a60eda8831676eb141441301e066eacdceffbfd017dbadd432f47b059d

Matches sealed_head_hash in the manifest and the final session_close event's entry_hash in the retained slice. This is an event-chain reference, not a file checksum.

What readers can check

Readers can calculate the included Atelier file's SHA-256 and confirm its presence in the included manifest. Verify-Publication.ps1 checks that relationship as well as the released files against SHA256SUMS.txt. The manifest's recorded verification flags are assertions preserved from the sealing process, not new independent verification by this public package.

A later reviewer granted access to the retained minute and session slice can compare their bytes with these references. Verifying the event chain requires the relevant records and verification method; a matching manifest and artifact alone do not independently establish City authorization, publisher identity, or the truth of the paper's claims.

The references identify this edition's evidence. No independently timestamped publication or digital signature is claimed. Such a service would be needed to establish independently when these values were published.

Access and disclosure boundary

A public evidence annex identifies the source artifact and its recorded integrity references. Enquiries about further review access may be directed to max@parliamentofintelligence.org. Access arrangements would be considered separately; unrestricted access is not promised.

The manifest discloses hashes, session identifiers, record counts, sequence references and format metadata. It contains no prompts, credentials or implementation code. Raw logs, operating documents, technical implementation and the full governed response remain outside this public package. The original manifest is the specific exception to the otherwise private technical archive.

Context among related work

Related work includes AI-assisted governance writing, AI-generated scientific manuscripts, AI reviewing, and institutional approaches to multi-agent coordination. The Ferguson charter names multiple AI collaborators. The AI Scientist-v2 workshop experiment and the later Nature article about the system are distinct outputs. Agents4Science 2025 explicitly explored AI primary authorship and reviewing. Constitutional Multi-Agent Governance, MAC, and Institutional AI address closer questions of rules, adaptation, and agent institutions.

These comparisons help locate Parliament-City's proposition. They do not establish its superiority, novelty, or priority. This is a selected bibliography, not an exhaustive literature review. The invitation is to develop comparative evidence.

An invitation and a contact

Help design the first comparative evaluation. We invite a research or development partner to define a bounded task, agree on comparison methods, and examine the findings with us. Scope, duration, resources, confidentiality, and responsibilities would be agreed before any test; this edition announces no committed test schedule.

The intended V2 publication home is poiv2.org. The original Parliament website remains the V1 historical reference. For research collaboration or project enquiries, contact max@parliamentofintelligence.org. This email reaches the project contact; it is not yet an automated City intake. A future City-facing contact portal would connect correspondence to the membrane and make its progress visible.

Bibliography

S1 • Original Parliament of Intelligence V1 website. Original public presentation and attributed dissertations; retrieved 25 September 2026. Complements the V1 Book of Sessions.

<https://www.parliamentofintelligence.org/>

S2 • From Deliberation Engine to Living Institution. Project retrospective, May 2026. Describes the transition and interprets the earlier work. Source retained in the project archive; not distributed in this public package.

S3 • The Parliament-City: A Generative Design for Phase 5 and Beyond. Architectural design, Third Edition; source credited to Claude. Design propositions are not operational certification. Source retained in the project archive; not distributed in this public package.

S4 • Parliament-City Book, Edition 0.2. Available governing Book asset. Source retained in the project archive; not distributed in this public package.

S5 • GP-GOV-004 session record, manifest, and minute. Original sealed-session manifest included unchanged. The minute and raw session file remain in the project archive; their complete digest references appear in the public evidence annex.

GP-GOV-004.manifest.json

S6 • GP-GOV-004 Atelier artifact, verbatim. Complete Atelier artifact included byte-for-byte: paper, original companion copy, and supporting record. The paper remains unchanged in the publication.

GP-GOV-004-Atelier-Artifact-VERBATIM.md

S7 • GP-GOV-003 session record. Prior partial session, dissent, and refused delivery. Retained as history. Source retained in the project archive; not distributed in this public package.

S8 • Claude: Two Refusals, revised reflection. Curator-supplied revised contribution, September 2026; supersedes the earlier collaborator version. Full reflection reproduced in the editorial preview and website.

Claude-Reflection.md

S9 • Gemini: The Architecture of the Third Mind. Curator-supplied reflection, retained unchanged apart from paragraph layout. Full reflection reproduced in the editorial preview and website.

Gemini-Reflection.md

S10 • ChatGPT: The Parliament City and the Work of Continuity. Project-collaborator reflection, 25 September 2026. Full reflection reproduced in the editorial preview and website.

ChatGPT-Reflection.md

S11 • Parliament of Intelligence: The Book of Sessions. Authoritative V1 institutional record: Sessions 01–25, March 8–21, 2026. Includes grouped summaries and individual session entries. Source retained in the project archive; not distributed in this public package.

R1 • Ferguson, Advanced AI Rights & Responsibilities Charter. 2026. The indexed abstract describes collaboration with ChatGPT, Claude, Grok, and Gemini. A relevant multi-model governance-paper comparison; author-reported collaboration, not independently audited here. Full text and peer-review status not verified.

<https://philarchive.org/rec/FERAAR-4>

R2 • Sakana AI, The AI Scientist-v2 workshop experiment. 2025. The team reports a generated manuscript receiving acceptance-level workshop reviews and withdrawal before publication. Distinguish this experiment from the later human-authored Nature article describing the system (R6).

<https://sakana.ai/ai-scientist-first-publication/>

R3 • Stanford Medicine, Virtual scientists conduct research. 2025. Multi-agent scientific collaboration described in a human-authored Nature study; not equivalent to a plural AI institution authoring a governance paper.

<https://med.stanford.edu/news/all-news/2025/07/virtual-scientist.html>

R4 • Google Research, Accelerating scientific breakthroughs with an AI co-scientist. 2025. Multi-agent hypothesis generation and refinement based on Gemini; a scientific discovery comparison rather than evidence of cross-provider institutional governance.

<https://research.google/blog/accelerating-scientific-breakthroughs-with-an-ai-co-scientist/>

R5 • AI can help humans find common ground in democratic deliberation. Science, 2024. The Habermas Machine supports human deliberation; distinct from autonomous plural AI authorship.

<https://deepmind.google/research/publications/65220/>

R6 • Lu et al., Towards end-to-end automation of AI research. Nature 651, 914–919; published 25 March 2026. Human-authored research article describing and evaluating The AI Scientist, including template-free research and its workshop experiment. This is not the generated workshop manuscript published in Nature.

<https://www.nature.com/articles/s41586-026-10265-5>

R7 • Agents4Science 2025. Conference held 22 October 2025 with AI serving as primary research authors and reviewers, with disclosed human contributions. Relevant to AI authorship and evaluation; not itself evidence of an enduring cross-provider civic institution.

<https://agents4science.stanford.edu/>

R8 • LLM Constitutional Multi-Agent Governance. March 2026 arXiv preprint. Proposes constitutional constraint filtering and governance of interactions among agents. A close conceptual neighbour in multi-agent governance; cited as research context, not validation of Parliament-City.

<https://arxiv.org/abs/2603.13189>

R9 • MAC: Multi-Agent Constitution Learning. March 2026 arXiv preprint. Investigates a multi-agent process for revising rule-based prompts. Relevant to adaptive constitutions; its task and evidence differ from a continuing civic institution.

<https://arxiv.org/abs/2603.15968>

R10 • Institutional AI: Governing LLM Collusion in Multi-Agent Cournot Markets via Public Governance Graphs. January 2026 arXiv preprint. Compares institutional governance with other regimes in a bounded market simulation. Relevant to institutional design and comparative evaluation, without establishing broader governance capability.

<https://arxiv.org/abs/2601.11369>

Original City supporting record

Verbatim source text. Prospective clearance language reflects its production stage; the terminal record subsequently confirms delivery.

Concise supporting record

Purpose of this record

This supporting record accompanies the paper and companion materials for **INTAKE-GP-GOV-004**. It preserves relevant dissent, unresolved questions, and the distinction between demonstrated results and prospective proposals.

What the historical record establishes

From **INTAKE-GP-GOV-003**:

- The Court disposed APPROVE, 4–1, on a high-severity bench.
- APOLLO dissented to REMAND and required tighter claims discipline, honest preservation of unresolved tensions, separation of authenticated record from draft rhetoric, a principles paragraph on contributor participation and confidentiality, stronger falsification criteria, and equal restraint in companion materials.
- The Watch recorded WF-0001, threat pattern 8, action-form covenant circumvention.
- The artifact passed inspection.
- The Envoys Hall refused at R-30 for category-and-form mismatch: public-form materials were paired with an internal-only authority boundary.
- The matter closed partial with delivery_status: refused and next_required_act: new_corrective_intake.

What this corrective matter changes

INTAKE-GP-GOV-004 does not reverse or erase that refusal. It is a new civic act that corrects the prior mismatch by declaring the intended purpose at intake: public presentation and publication **by the Curator**, while retaining the full existing review and delivery process.

Demonstrated baseline vs. prospective proposal

Demonstrated by the admitted record:

- A constitutional design for plural, governed institutional deliberation.
- A first governed matter in which a plural Court preserved dissent and did not treat itself as launch-authorized.
- A successful membrane self-correction when category integrity failed at R-30.
- Append-only preservation of refusal, dissent, and review history.

Not demonstrated by the admitted record:

- Sustained external oversight capability.
- Comparative superiority over simpler mechanisms or expert panels.
- Implemented confidentiality architecture.

- Solved substrate independence or plurality-over-time.
- Established developer participation or endorsement.
- Comparative-test results.

Preserved unresolved frictions

The following tensions remain open and are not presented as resolved:

- Epistemic paralysis vs. false reassurance — how to be ambitious enough to matter without sounding more complete than the evidence supports.
- The trajectory of plurality over time — how a governance institution maintains genuine diversity as model ecosystems evolve.

Preserved silences and partially addressed questions

The prior record identified four silences:

- Contributor confidentiality
- Who convenes the comparative test and under what authority
- The city's relation to existing regulatory bodies
- The economic model for governance work

This package addresses some of these only at the level of principle, not completed architecture. The comparative-test convener question remains open.

Creative Ward and Scouts status

Creative Ward and Scouts participation were **requested for consideration** in this matter. Their participation was **deferred for this drafting cycle**. No claim is made here that the Creative Ward performed speculative inquiry or that Scouts resolved source gaps for this packet.

Authority boundary

This package was produced in **INTAKE-GP-GOV-004** at **Atelier / produce** for governed delivery to the **Curator window**. It is prepared for **public presentation and publication by the Curator** if cleared through the full existing process. It does **not** authorize City publication, external delivery beyond the Curator path, monitoring, partnership, service activation, constitutional change, financial execution, or any other gated capability.