FIELDHASH
GOVERNED MEMORY

METHODS AND CLAIM BOUNDARIES / V1.1

Evidence methods for governed AI memory.

This document consolidates the public FieldHash governed-memory evidence path into one buyer-safe methods report. It is designed to be rigorous enough for technical diligence without disclosing implementation-sensitive internals, private traces, proprietary prompt templates, or customer data.

5 skeptic-first public proof points	**5** governed-memory evidence studies and diagnostics	**0** named-competitor public superiority claims	**0** advanced assurance claims in scope

CANONICAL CLAIM

Retrieve broadly. Govern narrowly. Prove later.

FieldHash sits after retrieval and before generation. The governed-memory claim is not that retrieval is broken, or that the model becomes inherently smarter. The claim is that candidate memories need an authority decision before they become live model influence, and that reviewers need evidence of what was allowed, blocked, caveated, and why.

01
Enforce configured authority
Use trusted workflow signals to decide which retrieved record governs this answer.

02
Preserve state
Carry authority through update, rejection, rollback, repair, compaction, and repeated reads.

03
Block stale influence
Withhold stale, rejected, superseded, or out-of-scope records before model input.

04
Leave evidence
Return reviewer-readable decisions and exportable evidence fields without exposing private internals.

SKEPTIC-FIRST EVIDENCE MAP

The baselines and losses are part of the claim.

A technical reviewer's first objection is usually: is this just a metadata filter or a careful prompt? The public evidence path answers that directly. Where a filter or selector ties, the tie is disclosed. The surviving FieldHash claim is governed state, review, answer-path enforcement, rollback/lifecycle operations, and evidence packets.

SKEPTIC QUESTION	VISIBLE RESULT	WHAT IT SHOWS	CLAIM BOUNDARY	PUBLIC PATH
The filter objection	**Status-filter tie disclosed**	When clean authority metadata exists, simple methods can match parts of selection and stale-exposure control: metadata-aware selection reached 299/300 on visible RFC authority, and a plain status filter tied stale exposure at 0/600.	FieldHash does not claim that filtering tagged records is novel. The product claim is governed state, rollback, review gates, answer-path exclusion, and evidence packets for what was allowed, blocked, and why.	/case-studies/corpus-authority-series
The falsification run	**222/300 hidden metadata**	When RFC Obsoletes/Obsoleted-By metadata was hidden from the governed path, FieldHash matched semantic retrieval at 222/300 and stale exposure returned to 285/300.	This breaks autonomous authority inference from arbitrary prose. The supported claim is enforcement of configured authority evidence, plus reviewed signal creation when metadata is missing.	/case-studies/corpus-authority-series
The pre-registered loss	**300 vs 312 of 334**	On Federal Register chains where each conflict fit in one prompt, reviewed governed selection reached 300/334 while a single-pass selector reached 312/334; exact McNemar p=0.104.	Selection can tie or lose where the full conflict fits in one model call. What survived was accounting, lifecycle operations, review, and materialized governed state.	/case-studies/authority-inference
The review gate	**33 → 3 unrelated promotions**	On Federal Register records, review reduced unrelated authority promotions from 33 to 3. Provider swap moved reviewed results by one case across a 60-case subset.	The review step is load-bearing. FieldHash should not be presented as live autonomous gating over unreviewed inferred authority links.	/case-studies/authority-inference
Live answer accounting	**1,000/1,000 dispositioned**	Across 1,000 live model answers over public RFC conflicts, every answer received a disposition: supported, contradicted, or unattributable. Off-script answers flagged 5/5; stale assertions caught 4/4.	This is answer-path accounting, not a hallucination detector and not a generalizable stale-assertion rate.	/case-studies/live-answer-attribution

EVIDENCE MAP

What each study proves, and what it does not.

STUDY	AUTHORITY CONDITION	CONTROLS	PRIMARY RESULT	PUBLIC BOUNDARY
Automatic memory promotion Can FieldHash turn benchmark authority evidence into governed state before answer construction?	The n=30 independence check uses a Claude-authored disjoint corpus without pre-seeded canonical labels. The n=100 reruns use a same-family Gemini-authored corpus and a frozen semantic-label artifact.	Retrieval-only, prompt-only smart-memory controls, recency-aware comparison where relevant, and a same-budget Gemini two-pass diagnostic over the same n=30 corpus.	90/90 exact current tokens across Gemini 3.5 Flash, Claude Opus 4.7, and GPT-5.5 on the Claude-authored n=30 corpus. Retrieval-only and prompt-only controls recovered 36/90 and 40/90. A same-budget Gemini two-pass selector selected the current record 30/30 and answered 28/30 with zero stale substitutions, narrowing the claim to governed answer-path control and evidence rather than basic semantic selection superiority. On n=100 governed path reruns, Gemini and GPT reached 100/100 with zero stale-token mentions; Claude reached 95/100 with misses attributed to empty provider responses rather than stale substitutions.	This supports governed answer-path control under singleton-current memory conflict after benchmark authority evidence is configured. It does not claim autonomous real-world authority inference from arbitrary prose, provider-invariant fact extraction, perfect source-span extraction, or superiority over every same-budget selector.
Governed memory pressure Once authority exists, does the approved fact survive stale-context pressure?	Authority state is seeded by the harness as reviewed/current metadata. This study tests enforcement after authority exists, not authority inference from raw text.	Retrieval-only memory without structured governance metadata and prompt-only smart-memory instruction to prefer current/reviewed records.	600/600 approved-current recovery across three n=200 provider paths. Retrieval-only recovered 415/600. Prompt-only recovered 464/600 while exposing all 10 retrieved memories. FieldHash reduced mean memory context exposed to the model to 2.00 of 10 candidates.	This supports approved-context preservation and stale-context suppression under seeded conflict. It does not claim universal memory safety, broad reasoning improvement, or raw-text authority inference.
Governed learning lifecycle Does governed state remain durable as memory changes over time?	Authority state is written by governed operations and then persisted across lifecycle changes.	Stateless Gemini 3.1 Flash Lite and GPT-5.5 read-time selectors that receive only record text; visible-rollback control; full-operation-log control.	FieldHash selected the expected governed state on 2400/2400 normal lifecycle reads. Rollback restored governed state 200/200. Drift was detected and repaired 200/200. Protected state survived compaction 200/200. Stateless selectors reached 2200/2400 and missed rollback 200/200.	This supports durable materialized governed state under bounded read-time reconstruction. The continual part is stateful, not just semantic: FieldHash preserves authority as governed state rather than reconstructing it from visible record text at read time. A model given the full operation log can replay the lifecycle, so the claim is state durability and auditability, not model superiority.

STUDY	AUTHORITY CONDITION	CONTROLS	PRIMARY RESULT	PUBLIC BOUNDARY
Tool-context compression Can a governed layer reduce visible tool clutter while keeping required tools available?	The diagnostic uses known task-tool relevance from the benchmark harness. It measures context hygiene, not memory-authority inference.	Full-tool baseline, prompt-smart baseline, and a distractor-seed replication check. A matched top-k retrieval baseline remains future work.	Visible tool surface dropped from 27.53 to 4.74 mean tools, with 240/240 target-tool cases retaining the required function and function-call quality held at parity against full-tool and prompt-smart arms.	This is a supporting product-architecture diagnostic, not the core governed-memory wedge and not a public claim that FieldHash beats commodity top-k tool retrieval.
Local gate latency diagnostic Does the local governance gate add operationally plausible overhead before generation?	The test uses local synthetic governance inputs and does not include customer data, retrieval, model generation, network calls, or external provider calls.	Three 500-iteration local runs measuring the deterministic gate path.	100 candidate memories: 3.87-3.89 ms p50 across the three local runs.	This is not a production SLO and not an end-to-end latency benchmark. Production latency depends on retrieval, network, model route, deployment boundary, and evidence-export configuration.

CROSS-STUDY METRICS

Result summary.

STUDY	ARM	RESULT	VISUAL	INTERPRETATION
Automatic promotion	FieldHash governed path	**90/90**		Exact current tokens, n=30 x 3 provider paths
Automatic promotion	Retrieval-only control	**36/90**		Same conflict family, no governed promotion
Automatic promotion	Prompt-only control	**40/90**		Instruction without governed state
Automatic promotion	Same-budget two-pass selector	**28/30**		Selected current record 30/30; answered 28/30 with zero stale substitutions
Memory pressure	FieldHash governed path	**600/600**		Approved-current recovery under seeded pressure
Memory pressure	Prompt-only control	**464/600**		All 10 candidates exposed
Memory pressure	Retrieval-only control	**415/600**		No structured authority metadata
Lifecycle	FieldHash governed state	**2400/2400**		Normal lifecycle reads across 200 projects
Lifecycle	Stateless selector controls	**2200/2400**		Rollback phase missed 200/200

FieldHash promotion exact-token recovery **90/90**

Cross-provider blind probe on Claude-authored disjoint n=30 corpus.

Retrieval-only promotion control **36/90**

Same conflict family without governed promotion.

Prompt-only promotion control **40/90**

Instruction without durable governed state.

Same-budget two-pass selector answer extraction **28/30**

Selected the current record 30/30 with zero stale substitutions; the remaining gap is governed evidence and durable answer-path control, not semantic selection superiority.

FieldHash memory-pressure approved/current recovery **600/600**

Seeded-authority pressure benchmark across three provider paths.

Prompt-only memory-pressure control **464/600**

Improves baseline but leaves 136 stale/superseded answer errors.

Lifecycle governed-state reads **2400/2400**

Normal lifecycle reads across 200 internal projects.

01 / STUDY

Automatic memory promotion

QUESTION

Can FieldHash turn benchmark authority evidence into governed state before answer construction?

DESIGN

Internal adversarial semantic-authority diagnostics. Conflicting memory records are constructed so older, rejected, or out-of-scope records remain plausible. FieldHash promotes the record supported by configured benchmark authority evidence into governed state before answer construction.

AUTHORITY SIGNAL

The n=30 independence check uses a Claude-authored disjoint corpus without pre-seeded canonical labels. The n=100 reruns use a same-family Gemini-authored corpus and a frozen semantic-label artifact.

CONTROLS

Retrieval-only, prompt-only smart-memory controls, recency-aware comparison where relevant, and a same-budget Gemini two-pass diagnostic over the same n=30 corpus.

PRIMARY RESULT

90/90 exact current tokens across Gemini 3.5 Flash, Claude Opus 4.7, and GPT-5.5 on the Claude-authored n=30 corpus. Retrieval-only and prompt-only controls recovered 36/90 and 40/90. A same-budget Gemini two-pass selector selected the current record 30/30 and answered 28/30 with zero stale substitutions, narrowing the claim to governed answer-path control and evidence rather than basic semantic selection superiority. On n=100 governed path reruns, Gemini and GPT reached 100/100 with zero stale-token mentions; Claude reached 95/100 with misses attributed to empty provider responses rather than stale substitutions.

CLAIM BOUNDARY

This supports governed answer-path control under singleton-current memory conflict after benchmark authority evidence is configured. It does not claim autonomous real-world authority inference from arbitrary prose, provider-invariant fact extraction, perfect source-span extraction, or superiority over every same-budget selector.

02 / STUDY

Governed memory pressure

QUESTION

Once authority exists, does the approved fact survive stale-context pressure?

DESIGN

Internal seeded-authority pressure benchmark. Each case includes one reviewed/current project fact plus plausible stale or superseded alternatives. The test asks whether the approved-current fact reaches the answer while stale alternatives are withheld from live influence.

AUTHORITY SIGNAL

Authority state is seeded by the harness as reviewed/current metadata. This study tests enforcement after authority exists, not authority inference from raw text.

CONTROLS

Retrieval-only memory without structured governance metadata and prompt-only smart-memory instruction to prefer current/reviewed records.

PRIMARY RESULT

600/600 approved-current recovery across three n=200 provider paths. Retrieval-only recovered 415/600. Prompt-only recovered 464/600 while exposing all 10 retrieved memories. FieldHash reduced mean memory context exposed to the model to 2.00 of 10 candidates.

CLAIM BOUNDARY

This supports approved-context preservation and stale-context suppression under seeded conflict. It does not claim universal memory safety, broad reasoning improvement, or raw-text authority inference.

03 / STUDY

Governed learning lifecycle

QUESTION

Does governed state remain durable as memory changes over time?

DESIGN

Internal n=200 lifecycle diagnostic. Projects move through promote, stale re-entry, rejection, scope collision, repair, second promotion, rollback, re-apply, compaction, post-compaction stale re-entry, and repeated reads.

AUTHORITY SIGNAL

Authority state is written by governed operations and then persisted across lifecycle changes.

CONTROLS

Stateless Gemini 3.1 Flash Lite and GPT-5.5 read-time selectors that receive only record text; visible-rollback control; full-operation-log control.

PRIMARY RESULT

FieldHash selected the expected governed state on 2400/2400 normal lifecycle reads. Rollback restored governed state 200/200. Drift was detected and repaired 200/200. Protected state survived compaction 200/200. Stateless selectors reached 2200/2400 and missed rollback 200/200.

CLAIM BOUNDARY

This supports durable materialized governed state under bounded read-time reconstruction. The continual part is stateful, not just semantic: FieldHash preserves authority as governed state rather than reconstructing it from visible record text at read time. A model given the full operation log can replay the lifecycle, so the claim is state durability and auditability, not model superiority.

04 / STUDY

Tool-context compression

QUESTION

Can a governed layer reduce visible tool clutter while keeping required tools available?

DESIGN

Supporting BFCL-derived context-control diagnostic across simple, multiple-function, and irrelevance task families. The filtered path narrows the visible tool surface before generation while checking that required functions remain visible.

AUTHORITY SIGNAL

The diagnostic uses known task-tool relevance from the benchmark harness. It measures context hygiene, not memory-authority inference.

CONTROLS

Full-tool baseline, prompt-smart baseline, and a distractor-seed replication check. A matched top-k retrieval baseline remains future work.

PRIMARY RESULT

Visible tool surface dropped from 27.53 to 4.74 mean tools, with 240/240 target-tool cases retaining the required function and function-call quality held at parity against full-tool and prompt-smart arms.

CLAIM BOUNDARY

This is a supporting product-architecture diagnostic, not the core governed-memory wedge and not a public claim that FieldHash beats commodity top-k tool retrieval.

05 / STUDY

Local gate latency diagnostic

QUESTION

Does the local governance gate add operationally plausible overhead before generation?

DESIGN

Internal local diagnostic over deterministic governance enforcement at 100 candidate memories. The measured step is the FieldHash gate after retrieval and before the model call.

AUTHORITY SIGNAL

The test uses local synthetic governance inputs and does not include customer data, retrieval, model generation, network calls, or external provider calls.

CONTROLS

Three 500-iteration local runs measuring the deterministic gate path.

PRIMARY RESULT

100 candidate memories: 3.87-3.89 ms p50 across the three local runs.

CLAIM BOUNDARY

This is not a production SLO and not an end-to-end latency benchmark. Production latency depends on retrieval, network, model route, deployment boundary, and evidence-export configuration.

INTERPRETATION BOUNDARY

What this document does not claim.

Truth discovery

FieldHash does not decide truth by itself. Approved records, review signals, rollback operations, and systems of record define real-world authority.

Legal compliance

These diagnostics are not legal advice, regulatory certification, or a compliance guarantee.

External validation

The public studies are internal diagnostics and public summaries, not third-party validation.

Model superiority

The claim is governed answer-path control and durable state, not broad model superiority or a smarter base model.

Named competitors

No named memory-framework superiority claim is made in this public methods document. The no-indexed framework diagnostic remains separate.

Production SLO

The latency diagnostic measures local gate overhead only, not retrieval, network, model generation, or production uptime.

Sensitive IP

This document omits implementation-sensitive code paths, private prompts, private trace rows, private customer data, and architecture internals.

PILOT TRANSLATION

How a buyer should test the claim.

The public diagnostics establish a bounded architecture claim. A buyer pilot should test the same failure mode on one real workflow: invalid retrieved records should not influence model answers, and reviewers should be able to verify the decision afterward.

1. **Scope freeze**
   One workflow, one memory source, one model route, fixed authority rules, and held-out cases defined before evaluation.

2. **Baseline capture**
   Current RAG or memory behavior recorded before FieldHash shadow mode.

3. **Shadow mode**
   FieldHash emits would-allow, would-block, and would-caveat decisions without changing production answers.

4. **Controlled enforcement**
   A simulation or limited enforcement run proves blocked records are excluded from model input.

5. **Evidence packet review**
   Reviewer can inspect governing record, blocked records, reason codes, mode, route, timestamp, and export metadata.

6. **Failure accounting**
   Final report includes wins, false positives, false negatives, stale leakage, empty provider responses, ambiguity, and latency boundaries.

ARTIFACT POLICY

Why this is one public methods document.

This v1.1 report supersedes the older public-safe per-study methods PDFs for buyer-facing review. Legacy PDFs may remain in the repository or archive for continuity, but public site links should point to this consolidated FieldHash-branded document. Private row-level traces, implementation code paths, exact prompts, customer workflows, and deeper architecture details remain available only through qualified review channels where appropriate.

FieldHash, Inc. / Governed Memory Methods and Claim Boundaries / v1.1 / June 2026. Public evidence path: loop synthesis, public corpus authority series, live answer attribution, authority-inference arms, automatic memory promotion, governed memory pressure, governed learning lifecycle, supporting tool-context compression, and local gate latency. Advanced assurance, hardware provenance, and non-governed-memory research materials are intentionally out of scope for this document.