2nd Annual Artificial Intelligence & Medicine Symposium

THE HORIZON MANDATE

Orchestrating Professional Reliability in the AI Era

Part 1: The Disorientation & The Shift

Part 2: The Framework & Workflow

Part 3: The Toolkit & Closing

Steven Grambow, PhD

Associate Professor of Biostatistics & Bioinformatics

Disclosures

Serve on multiple Data Monitoring Committees for WCG Consulting and Gilead Sciences, Inc.

[DRAFT] Written by a human | Boosted by AI | Transparency is part of the process

THE VIEW FROM THE WINDOW (LM ARENA)

Text Arena

View rankings across various LLMs on their versatility, linguistic precision, and cultural context across text

Last Updated: Jan 16, 2026 | Total Votes: 5,040,489 | Total Models: 298

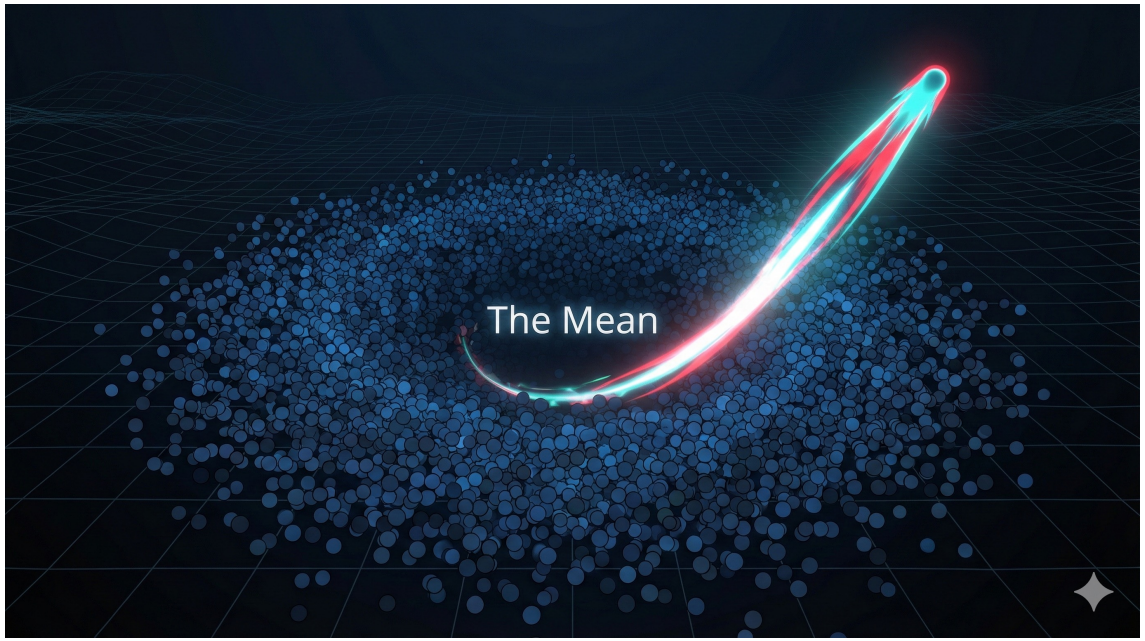
Overall | Search by model name... / | Style Control

Rank	Rank Spread	Model	Score	95% CI (±)	Votes	Organization	License
1	1 ↔ 1	gemini-3-pro	1490	±5	27,827	Google	Proprietary
2	2 ↔ 4	grok-4.1-thinking	1477	±5	27,985	xAI	Proprietary
3	2 ↔ 8	gemini-3-flash	1472	±6	13,245	Google	Proprietary
4	2 ↔ 8	claude-opus-4-5-20251101-thinking-32k	1470	±5	19,898	Anthropic	Proprietary
5	3 ↔ 9	claude-opus-4-5-20251101	1467	±5	21,241	Anthropic	Proprietary
6	3 ↔ 9	grok-4.1	1465	±5	32,015	xAI	Proprietary
7	3 ↔ 10	gemini-3-flash (thinking-minimal)	1462	±7	9,644	Google	Proprietary
8	3 ↔ 15	ernie-5.0-0110	1459 Preliminary	±9	4,829	Baidu	Proprietary
9	5 ↔ 14	gpt-5.1-high	1458	±5	24,439	OpenAI	Proprietary

THE STOCHASTIC TRAP

Stochastic token generators predict the most likely next token. Without direction, outputs trend toward average patterns in training data.

“Phase 1: The Oracle” *Probabilistic output without direction*



Most likely ≠ most useful

Key Insight: The model is not broken. It’s doing exactly what it’s designed to do, predict the most likely next token.

THE EVOLUTION: CHAT TO AGENTS

We are witnessing a fundamental shift in AI capability, moving from simple conversational interfaces to autonomous intelligent systems capable of executing complex tasks.

Phase 1: Stateless Chat

"The Cocktail Party" — Fun, conversational, but forgetful and disconnected.

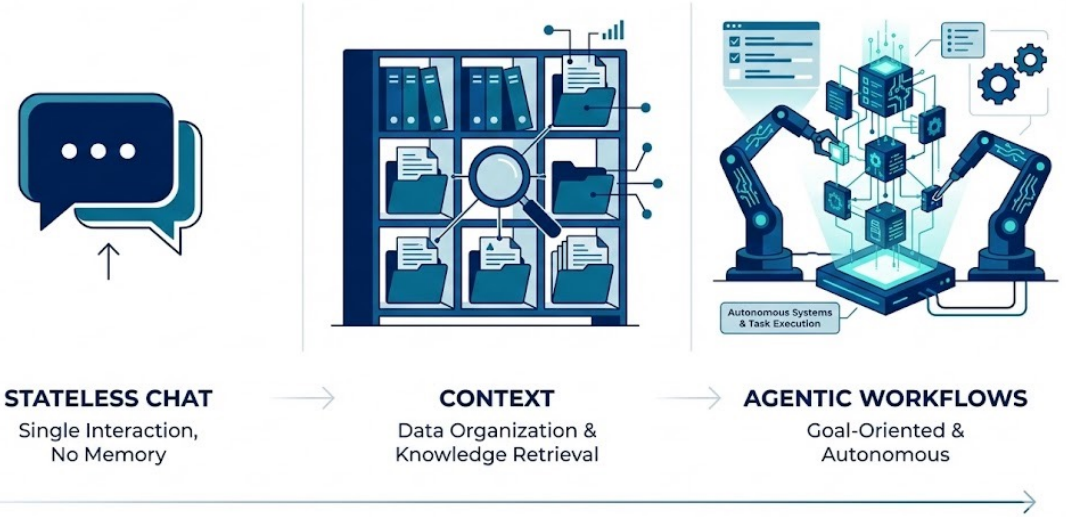
Phase 2: Contextual Tools

"The Library" — Grounded in data, RAG-enabled, capable of reference.

Phase 3: Intelligent Agents

"The Workbench" — Goal-oriented, tool-using, multi-step reasoning.

From Conversational Interfaces to Autonomous Intelligence Systems



Key Insight: We're not replacing chat—we're adding layers of capability.

FROM CHAOS TO STEWARDSHIP

Curate

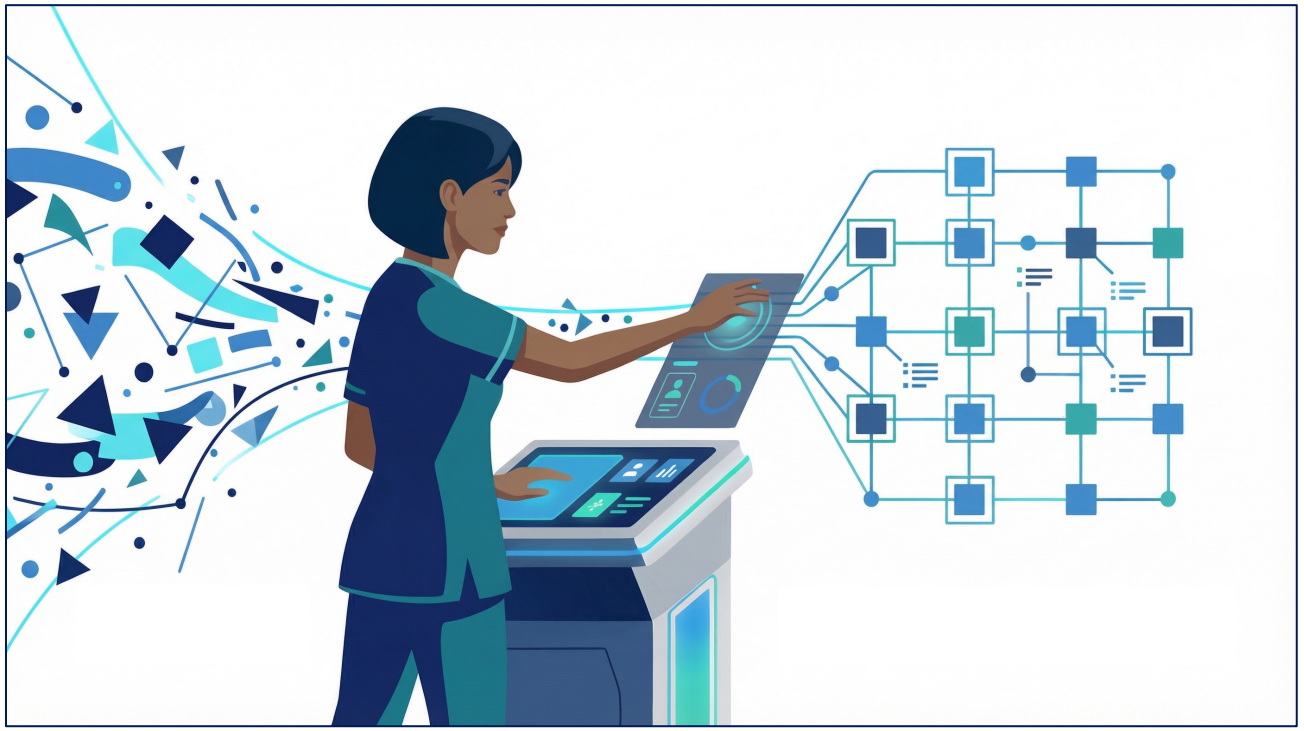
Context and constraints

Protect

Patient safety and data integrity

Align

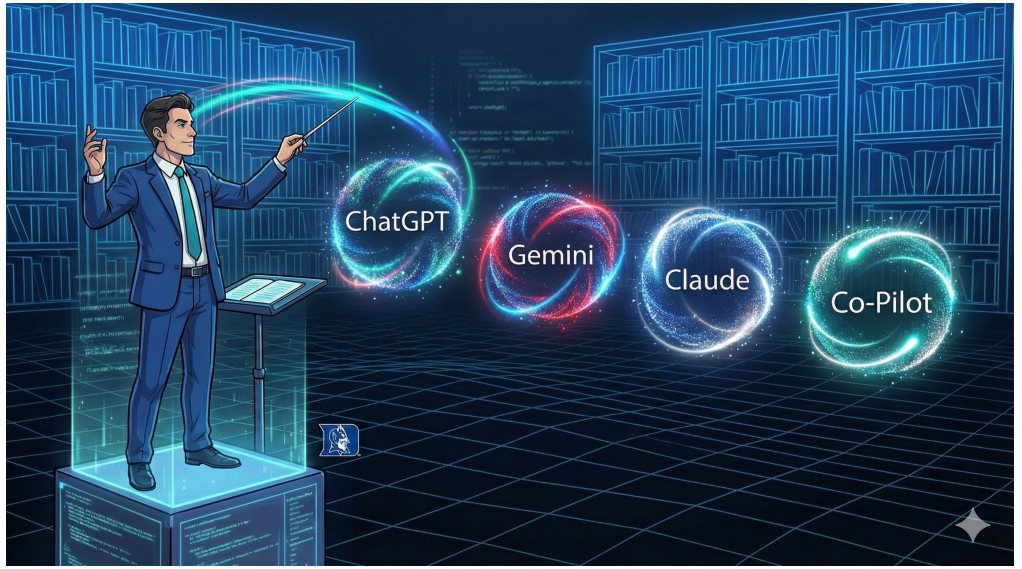
Outputs to standards and governance



Key Insight: Our job is to build the human layer that turns chaos into capability.

THE NEW CORE COMPETENCY

The Engine: Enthusiastic, Fast, Stochastic.



The Human: Critical, Directed, Accountable.

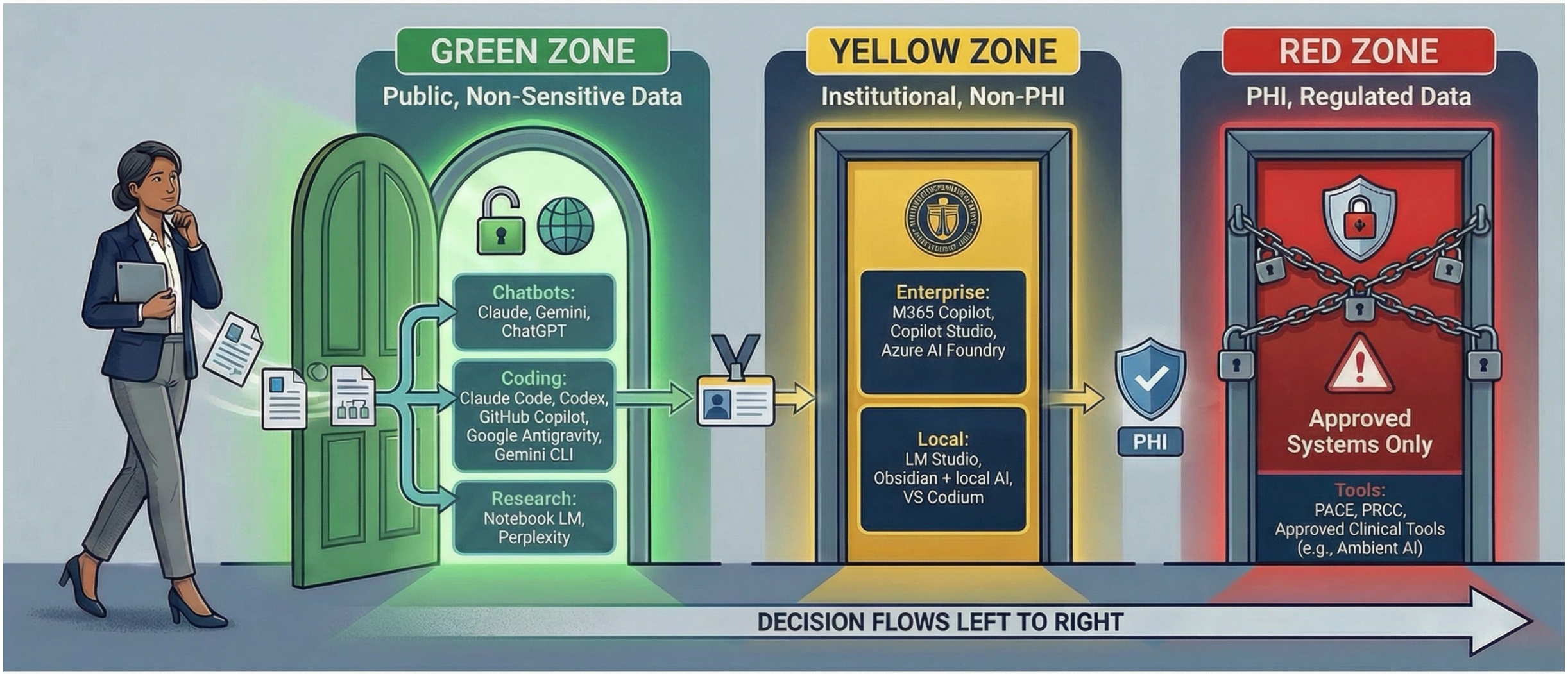
Workflow Architecture

Raw Intent → Stewardship Layer → Professional Reliability

Key Insight: Orchestrate the model; don't outsource judgment.

The Sovereignty Gate: Zone → Tools

Before you touch any tool, ask: where do these data belong?



WHY CONTEXT BREAKS

These aren’t user errors. They’re model properties.

Cognitive Load

Models have attention limits. Overload degrades output.

Context Rot

Performance decays as context grows. More is not better.

Lost in the Middle

Mid-context information gets ignored. Position matters.

Key Principle: Context stuffing **is not** context engineering

Curate, don’t dump!

Context Engineering: The Discipline

The steward's workflow for reliable AI-augmented output

Define Intent

Task, audience, success criteria

Curate Context

Select sources, Distill, don't dump.

Set Constraints

Rules, format, guardrails.

Validate Output

Human review. Verify claims.

Disclose Process

Trace the work. Document AI role.

Verified Output

Key Principle: Manage cognitive load. Durable context in project files; transient context in chat. Summarize and reset.

WORKFLOWS IN ACTION

PEER REVIEW

Define Intent
I review first; AI assists

Curate Context
Verify references; summarize sources

Set Constraints
Clean my dictation; preserve my voice

Validate Output
Human judgment; check accuracy

Disclose Process
Document AI role to editor

LITERATURE SYNTHESIS

Define Intent
Scope research question; cast wide net

Curate Context
Verify sources; extract key references

Set Constraints
Source-grounded only; no hallucination

Validate Output
Check synthesis against sources

Disclose Process
Cite sources; document AI role

PROGRESS REPORT

Define Intent
Clear deliverable; know your audience

Curate Context
Aligned tech stack = built-in context

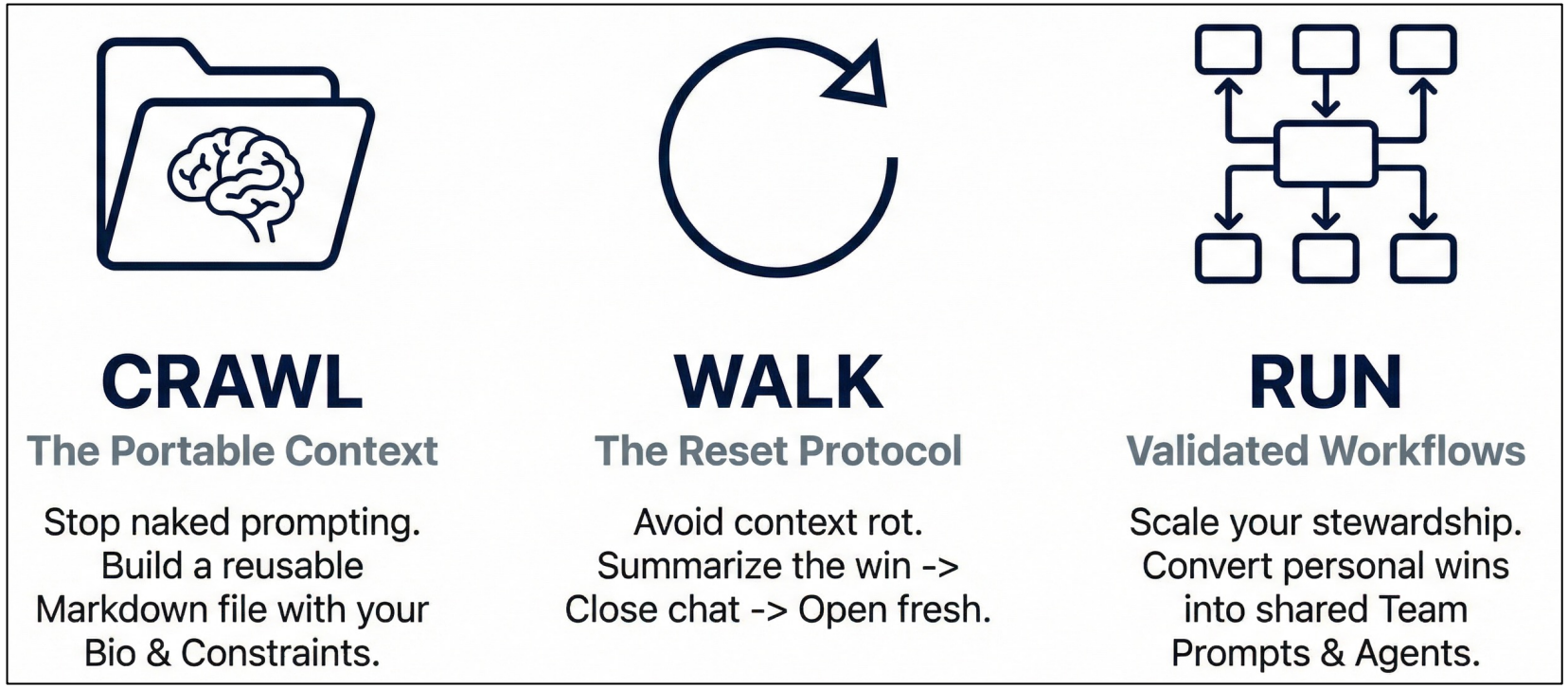
Set Constraints
Query within your own data

Validate Output
Human shapes the final deliverable

Disclose Process
Acknowledge AI assistance

PERSONAL INFRASTRUCTURE

From Stateless Chat to Durable Assets



Key Insight: **Amateurs prompt. Professionals build.** Your prompt is transient; your *context file* is an asset that compounds over time.

THE HORIZON MANDATE

We are stewards, not prompters.

Build the discipline. Protect the integrity. Orchestrate reliability.

ACCESS SLIDES & RESOURCES
https://duke.is/horizon-mandate

[DRAFT] Written by a human | Boosted by AI | Transparency is part of the process