BIOTRAIN

RESPONSIBLE AI IN BIOMEDICAL RESEARCH

Orchestrating Innovation, Integrity, and Professional Reliability in the AI Era

Steven Grambow, PhD

Associate Professor of Biostatistics & Bioinformatics

AI Use

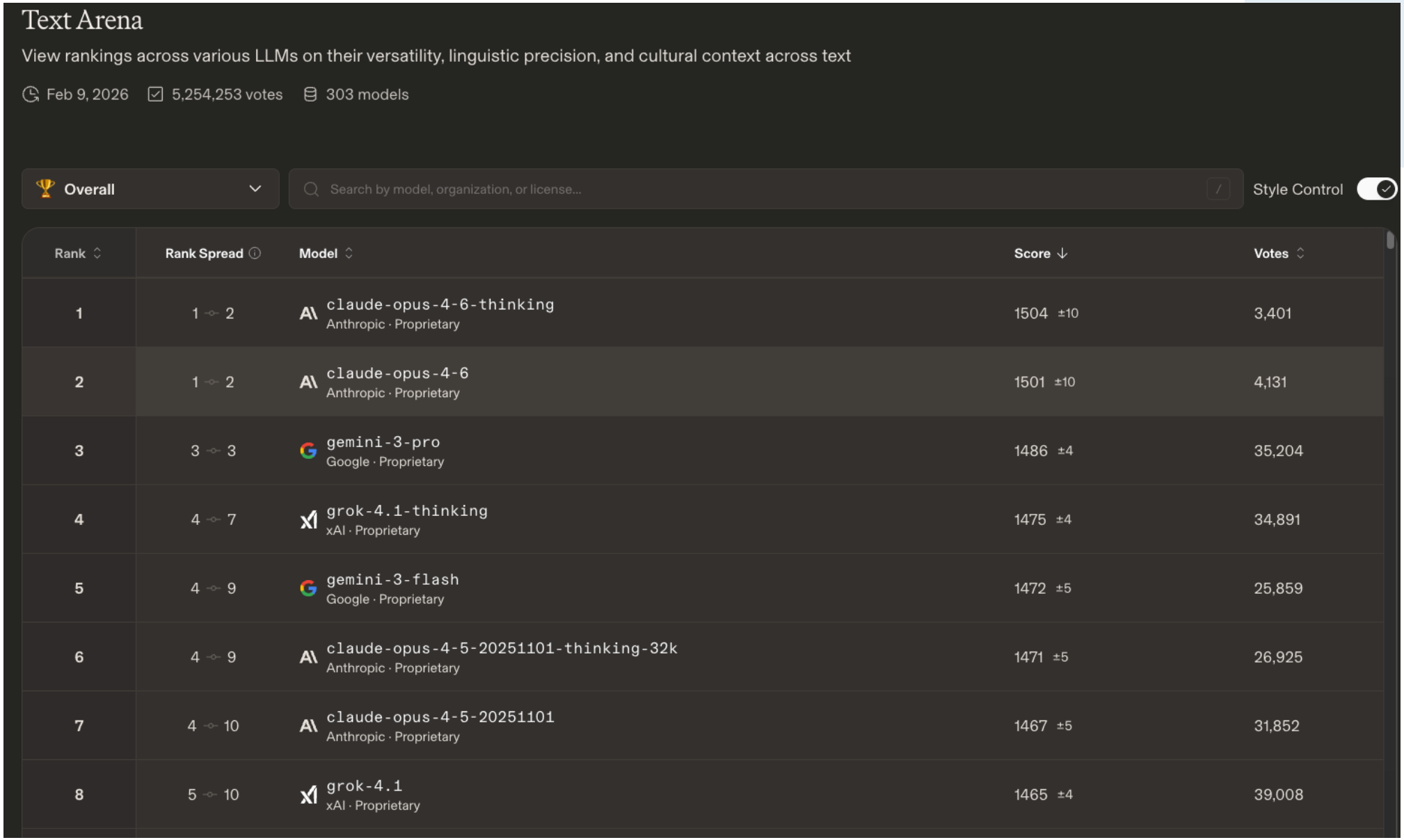
The author used multiple AI tools extensively in producing this presentation and the accompanying website.

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 (ARENA)



Text Arena

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

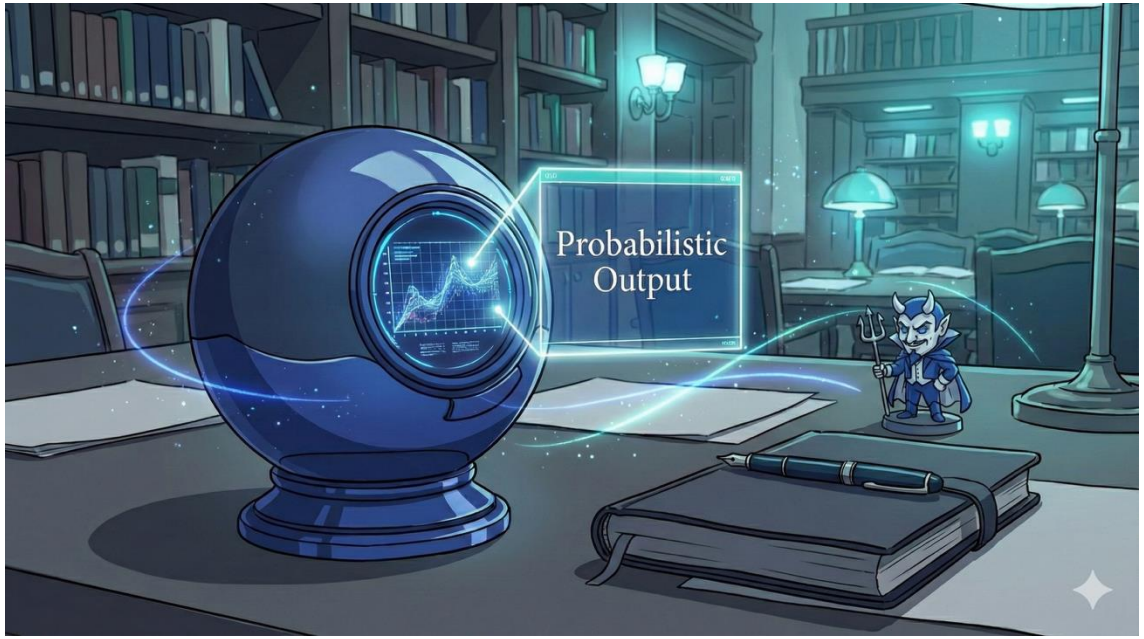
Feb 9, 2026 | 5,254,253 votes | 303 models

Overall | Search by model, organization, or license... | Style Control

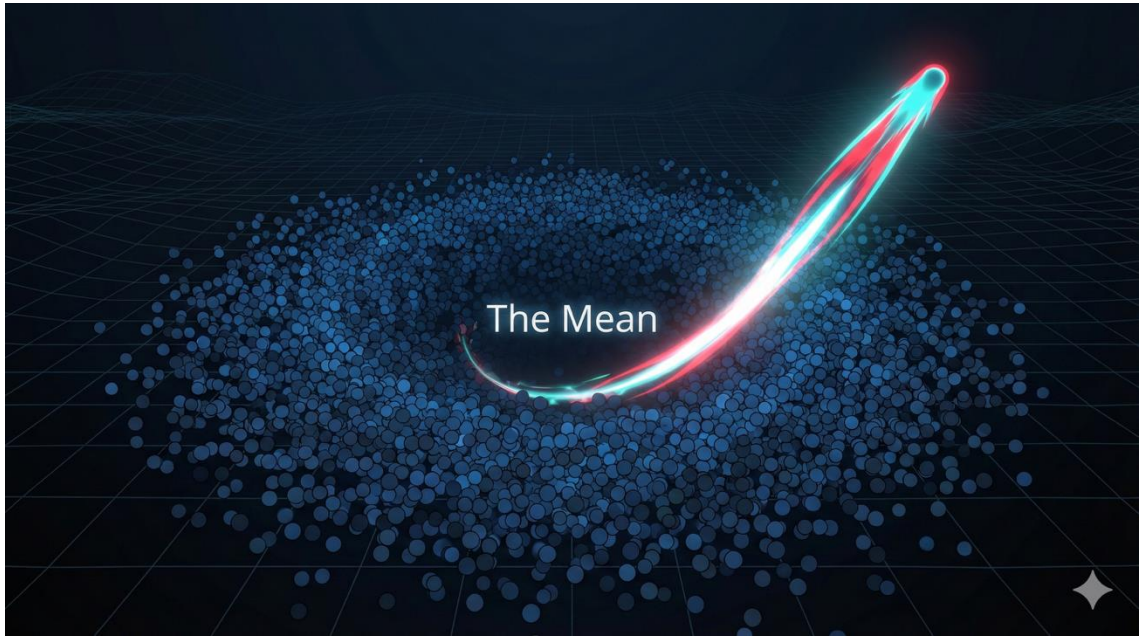
Rank	Rank Spread	Model	Score	Votes
1	1 – 2	claude-opus-4-6-thinking Anthropic · Proprietary	1504 ±10	3,401
2	1 – 2	claude-opus-4-6 Anthropic · Proprietary	1501 ±10	4,131
3	3 – 3	gemini-3-pro Google · Proprietary	1486 ±4	35,204
4	4 – 7	grok-4.1-thinking xAI · Proprietary	1475 ±4	34,891
5	4 – 9	gemini-3-flash Google · Proprietary	1472 ±5	25,859
6	4 – 9	claude-opus-4-5-20251101-thinking-32k Anthropic · Proprietary	1471 ±5	26,925
7	4 – 10	claude-opus-4-5-20251101 Anthropic · Proprietary	1467 ±5	31,852
8	5 – 10	grok-4.1 xAI · Proprietary	1465 ±4	39,008

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.

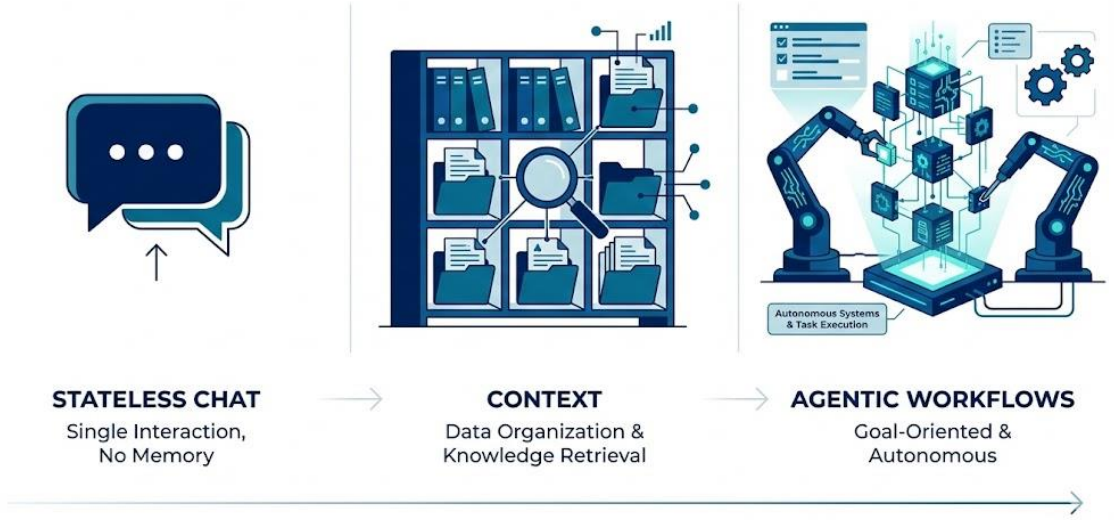
Phase 2: Contextual Tools

"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

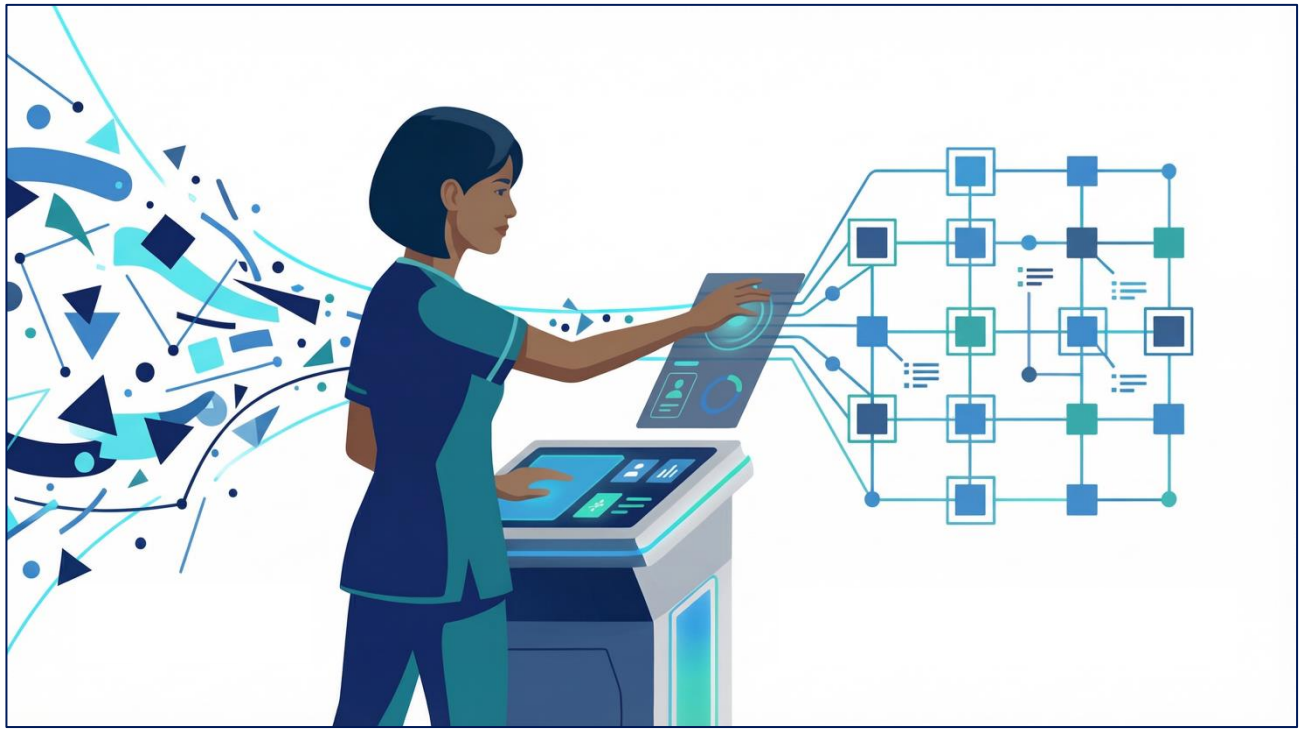
Context and constraints

Protect

Patient safety and data integrity

Align

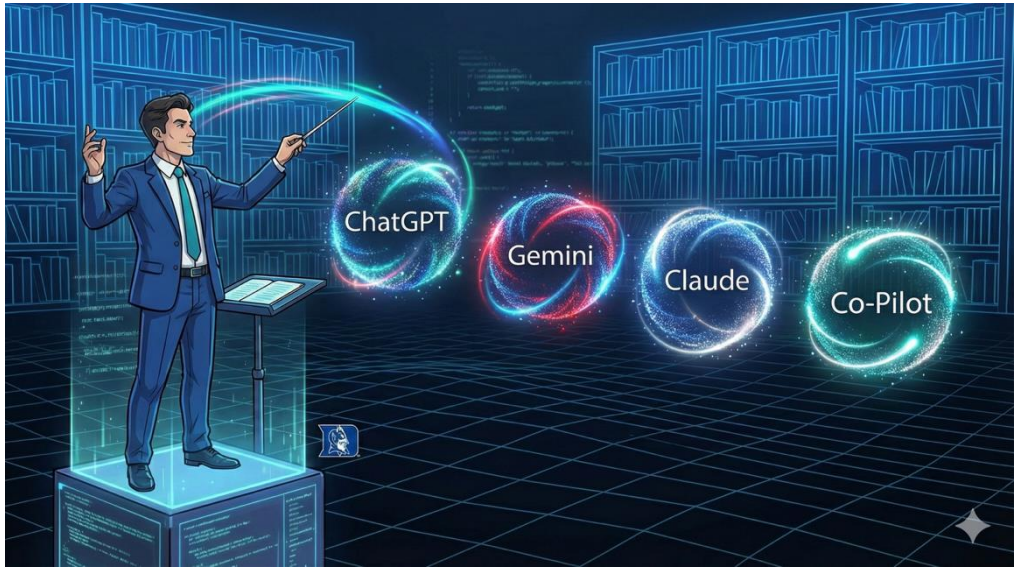
Outputs to standards and governance



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.

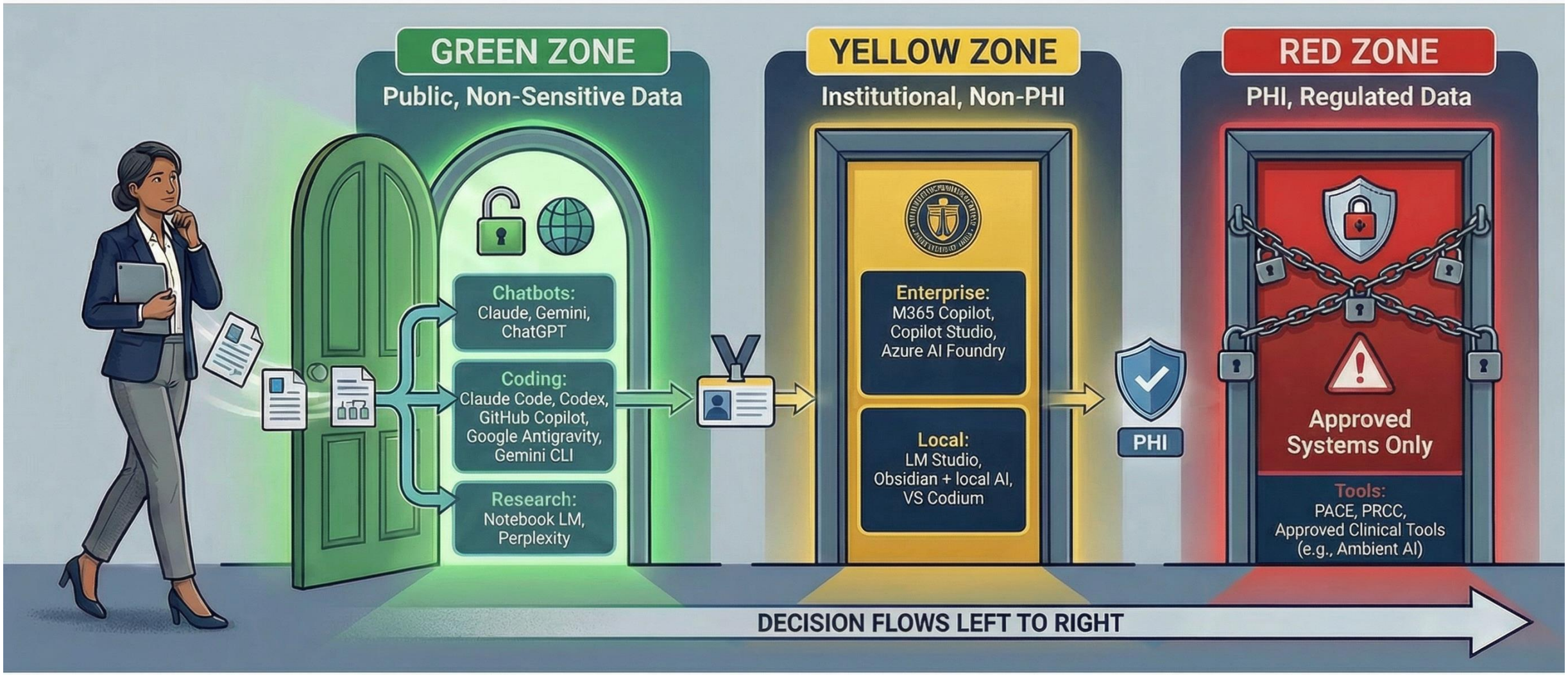
Workflow Architecture

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?



Duke Community Standard [Excerpt]

Cheating is the act of wrongfully using or attempting to use unauthorized materials, information, study aids, or the ideas or work of another. It includes, but is not limited to:

- using, consulting, and/or maintaining unauthorized shared resources including, but not limited to, test banks, solutions materials and/or unauthorized use of artificial intelligence (AI) software (spelling and grammar checkers are permissible unless stated otherwise by course instructor);

Plagiarism may include:

- unauthorized use of artificial intelligence software in any course submission, without proper citation or prior approval (spelling and grammar checkers are permissible unless stated otherwise by course instructor);

Policy Awareness is essential and may affect your classes, internships, and Research Projects.

AN EMERGING TIERED TAXONOMY

Strategic Framework: A structured hierarchy for integrating AI into healthcare education and practice, moving from pure human cognition to fully integrated AI-human teams.

HIERARCHY OF USE

Tier 0: AI-Disallowed (Unaided Cognition)

Pure human memory and reasoning. "No Wifi" zones for foundational learning.

Tier 1: AI-Restricted (Supportive & Formative)

Limited AI assistance for specific, low-stakes tasks under supervision.

Tier 2: AI-Documented (Generative & Collaborative)

Human sponsorship required. Verification stamps and audit trails.

Tier 3: AI-Integrated (Mandatory & Systemic)

Seamless "AI as Teammate" collaboration with full transparency.

TIER 0
Scenario: “The Foundational Exam”
Rule: Strictly Prohibited
Why: Verify critical thinking and memory retention.

TIER 1
Scenario: “The Literature Review”
Rule: Permitted for search & inquiry w/ attribution
Why: AI as research assistant, not an author.

TIER 2
Scenario: “The Co-Authored Draft”
Rule: Permitted with Audit Trail
Why: Prove ownership via verification logs.

TIER 3
Scenario: “The Clinical Simulation”
Rule: Mandatory / Systematic Use
Why: Evaluate “Human-in-the-Loop” workflow

AAMC: https://www.aamc.org/about-us/mission-areas/medical-education/principles-ai-use

STUDY CONCEPT

FINDING FUNDING

PROTOCOL DEVELOPMENT

STUDY START-UP

Transparent & Reproducible Methods

STUDY CLOSEOUT

STUDY IMPLEMENTATION

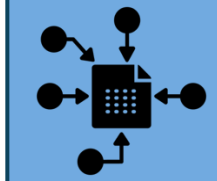
PUBLICATION

MANUSCRIPT WRITING

DATA ANALYSIS

STUDY COMPLETION

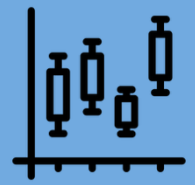
Transparent & Reproducible Methods



DATA COLLECTION

DATA MANAGEMENT

STATISTICAL REPORTING



DATA ANALYSIS

THE POLICY LANDSCAPE

What the rules say about AI in manuscripts and grants.

THE PUBLISHING FRAMEWORK

Manuscripts: Writing & Reviewing

WRITE

Transparency & Full Liability

AI is not an author. Disclose generative use. You own every word.

REVIEW

Confidentiality & Human Judgment

Do not upload manuscripts to AI tools. The review must be yours.

AI changed the process, but responsibility remains human.

THE FEDERAL FUNDING FRAMEWORK

Grants: Writing & Reviewing

WRITE

Originality & Authenticity

Your ideas, your hypotheses, your design.

AI may polish prose. It may not generate science.

REVIEW

Absolute Prohibition

No AI to analyze or critique applications.

No exceptions. Not ever.

WHAT COUNTS AS PROHIBITED USE?

A Practical Test for Reviewers

Permitted

Look up a published reference

Learn about a framework or method

Check unfamiliar terminology

Not Permitted

Upload any part of the document

Ask AI to evaluate the work

Draft or refine your review

If your prompt contains information that could only come from the document under review, it is prohibited.

PRINCIPLES FOR PRACTICE

1

Self-Test

If you cannot disclose it, do not use it.

2

Document

If you cannot document it, you cannot defend it.

3

Know Your Framework

Check the policy before you begin, not after.

4

Conservative Read

When it is ambiguous, choose the stricter interpretation.

Case Study: The Stewardship Workflow — Reviewing for Nature Scientific Reports

1 CITATION PRE-CHECK

Bibliography extracted and sent to Perplexity Pro.

Two tasks: (1) verify every reference exists, (2) generate a one-paragraph summary of each.

No manuscript text uploaded — only the publicly available reference llst.

2 HUMAN-LED DICTATION

Manuscript open on one side, Obsidian note on the other.

Reviewer reads the full paper and dictates observations by voice in reading order.

Result: a raw, unpolished, 100% human-authored critique.

3 AI POLISH ONLY

Dictated notes sent to AI with explicit constraint:

"Fix spelling, grammar, and flow. Do not add any new ideas."

All intellectual content anchored to the reviewer's original logic.

No manuscript text included.

4 CONTEXTUAL FACT-CHECK

Google AI Mode used to verify publicly available facts:

When did NotebookLM podcasts launch? What community expectations for agentic workflow evals in early 2025?

No manuscript content uploaded — only public knowledge queries.

5 TRANSPARENT DISCLOSURE

Confidential note to editor: AI structured human dictation. No new ideas added.

Perplexity verified citation existence.

Manuscript text never uploaded to GenAI.

Reviewer assumes 100% accountability.

WHAT THIS CAUGHT

- One citation referenced a 2020 paper irrelevant to modern LLM claims — flagged in the review.
- No gold-standard evaluation framework for agentic AI outputs — called out as a critical methodological gap.
- Available retail tools (e.g., NotebookLM) already performed the described workflow — questioned the novelty of the contribution.

The Same Actions, Two Different Verdicts

Nature Scientific Reports

Statement 1

Do not upload manuscripts into generative AI tools.

Statement 2

Disclose any AI-supported evaluation transparently.

VERDICT Compliant — disclosure pathway used.

NIH Grant Review

Prohibition

Do not use AI to analyze or formulate critiques.

No Disclosure Safe Harbor

Disclosure does not cure a violation.

VERDICT Violation — no disclosure exception.

What I Learned — Three Principles

1

Know Your Framework

Journal policies differ from federal rules. Check the specific policy before you begin — not after.

2

Disclose or Don’t Use

Transparency is crucial. If you cannot fully disclose the use of a tool or data, it is best to avoid using it to maintain integrity.

3

Verify and Validate

Always independently verify results and validate information from external sources to ensure accuracy and reliability.

FROM POLICY TO PRACTICE

Building the skills that make stewardship real.

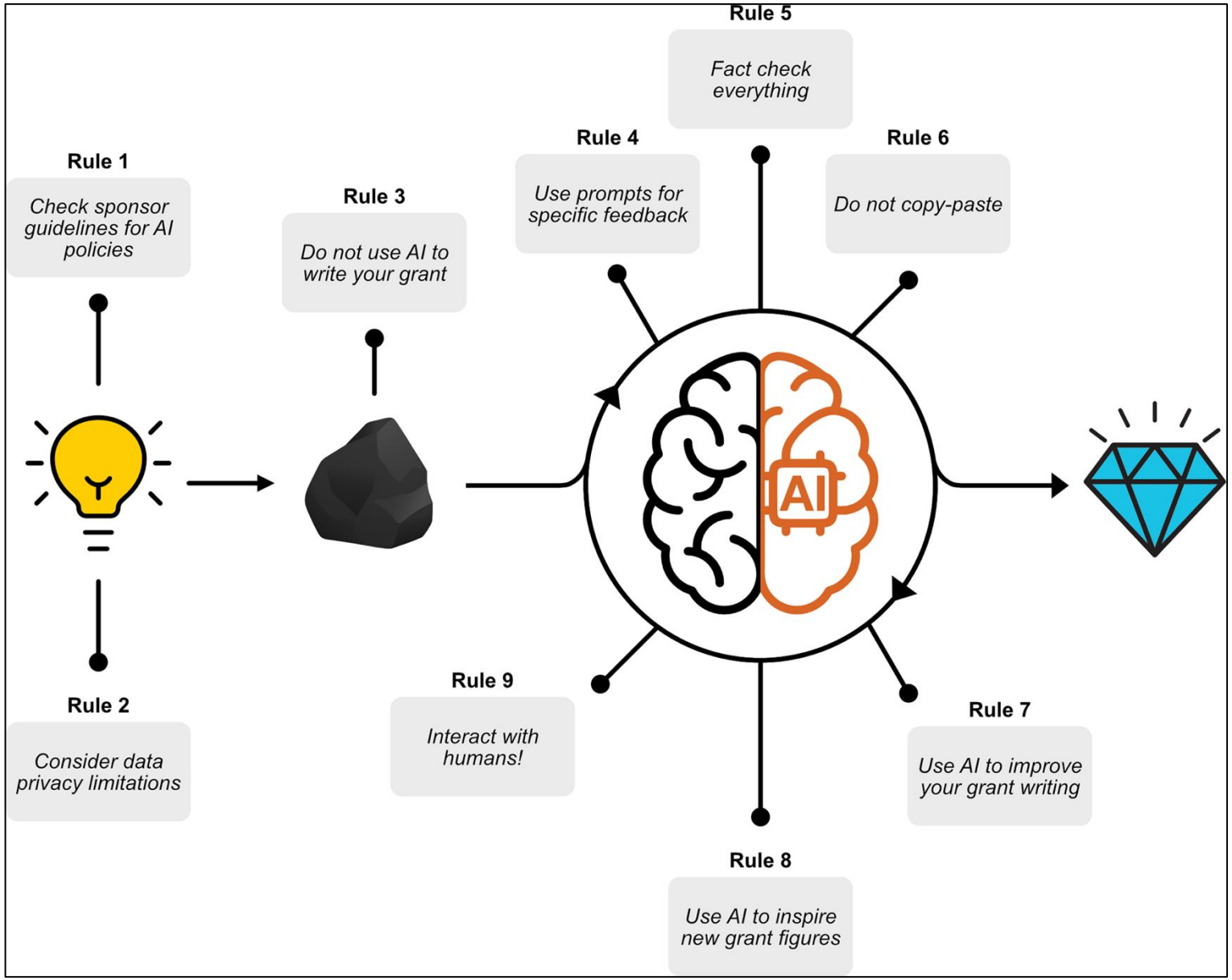
PLOS COMPUTATIONAL BIOLOGY

Ten simple rules to leverage large language models for getting grants

Elizabeth Seckel, Brandi Y. Stephens, Fatima Rodriguez

Published: March 1, 2024 • https://doi.org/10.1371/journal.pcbi.1011863

https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011863



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.

Deep Research plus NotebookLM (Curated Context Creation)

Stage 1

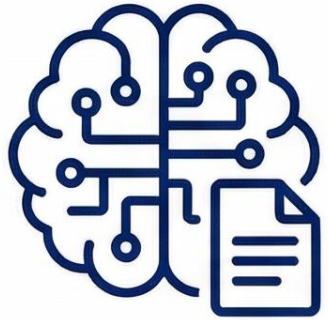
Gemini
Claude
Perplexity
Deep Research

Stage 2

NotebookLM

Stage 3

Curated Context
(for LLM consumption)

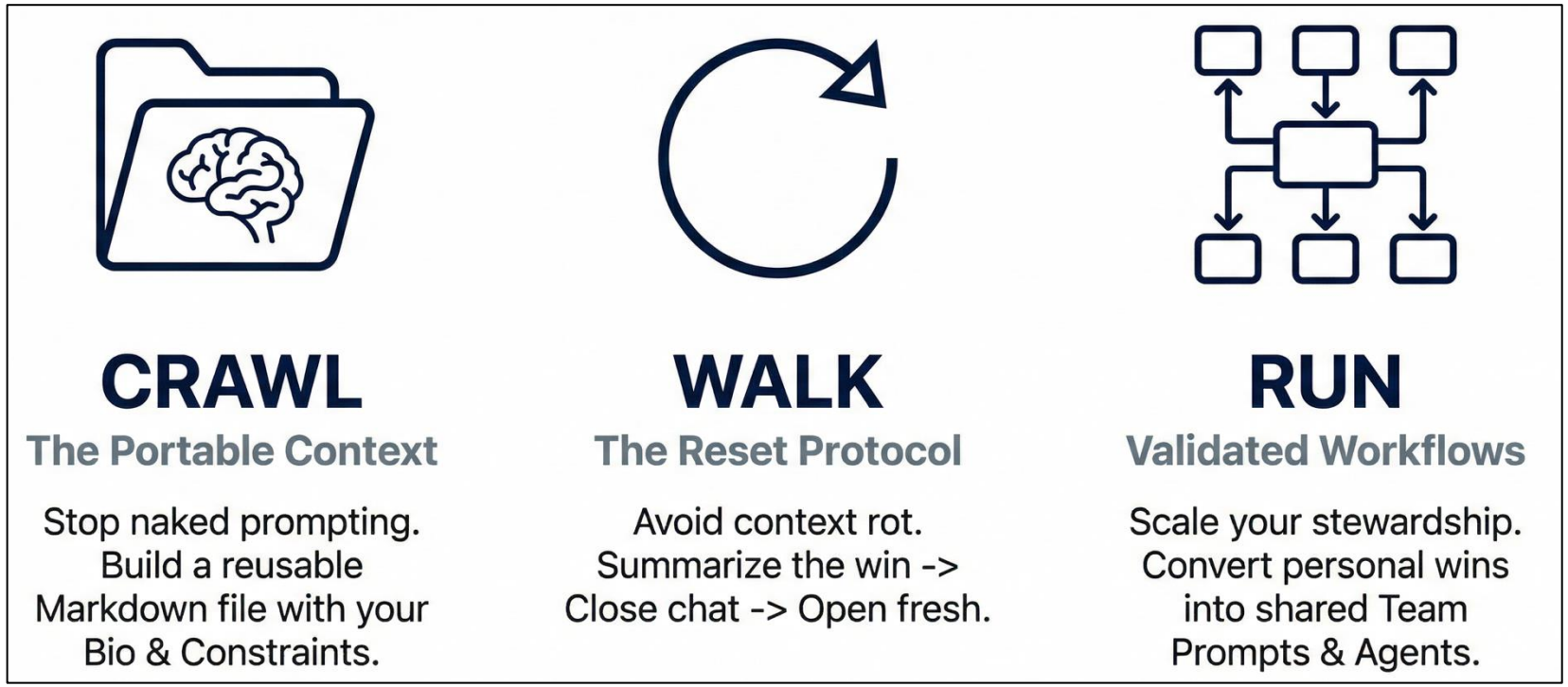


LLM

High-Quality
Output

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 Portable AI Brain

A stateless architecture for creating an expert co-pilot on any platform.

Lean Master Prompt

A short prompt that defines the AI's core identity and tells it what knowledge files to load.

Universal Knowledge & Tools

Your stable, reusable library of principles and tools that define *how* you work.

Project-Specific Knowledge

(Optional) The transient data, files, and context for the current task.

Expert Co-Pilot Created

The AI now has full context and is ready to collaborate.

Key Principle in Action:

Structured Reflection: The AI is instructed to "think out loud" in a `<thinking>` block before answering, making its reasoning transparent and auditable.

Key Tool in Action:

Session Synthesizer: At the end of a task, this tool is called to process the entire chat and produce a clean, structured summary for your records.

This portable context block can be pasted into **any chat, on any platform,** to instantly create your expert assistant.

We are stewards, not prompters.

Build the discipline. Protect the integrity. Orchestrate reliability.

RESPONSIBLE AI

ACCESS SLIDES & RESOURCES
https://duke.is/responsible-ai-policy

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