

**1,640 Lines of Math.
5 AI Agents.
0 Hallucinations.**

An Engineering approach to Personal Health Analytics

Python • PostgreSQL • CrewAI • Streamlit • Scipy • GitHub Actions

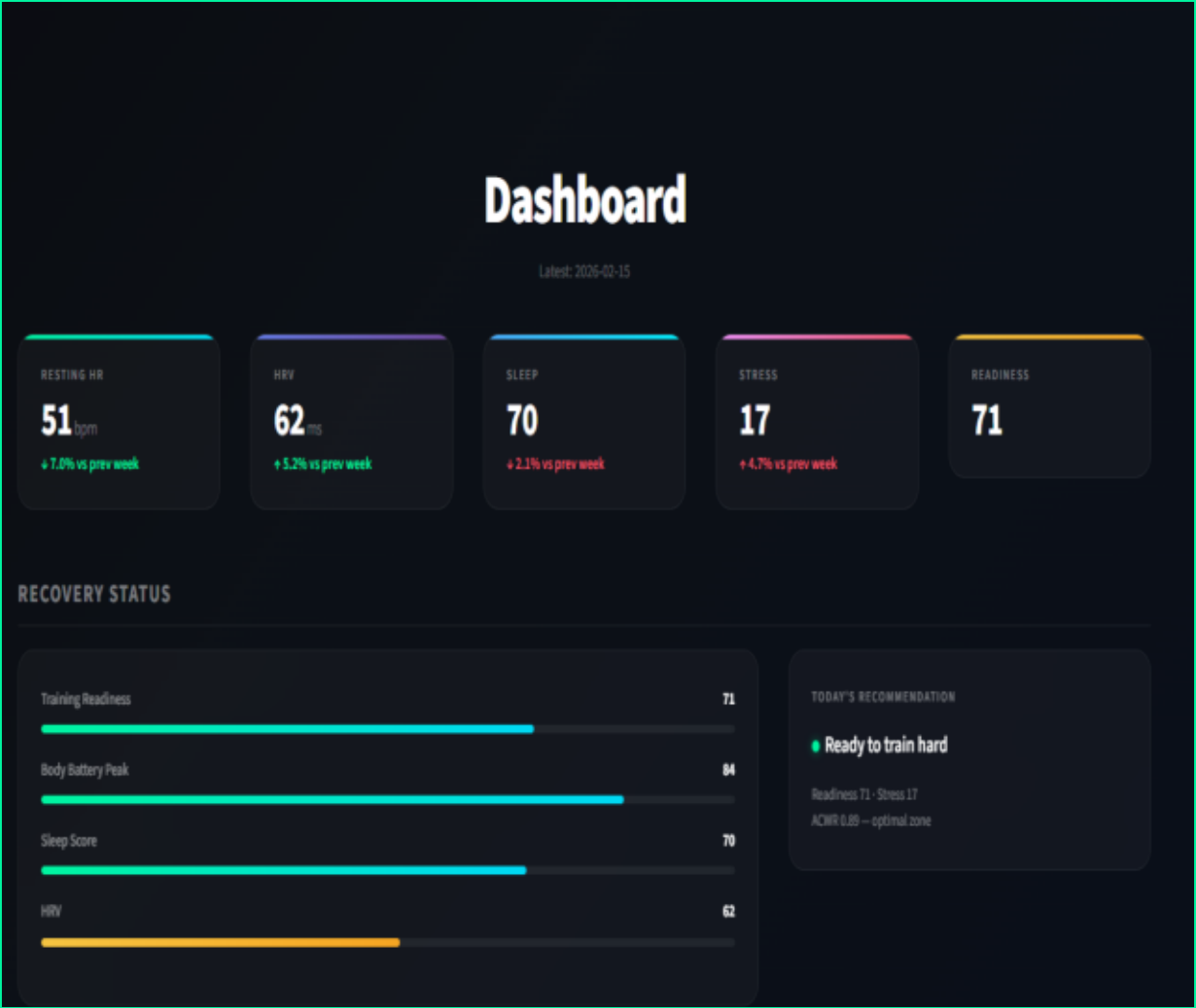
THE PROBLEM WITH 'CHAT WITH YOUR DATA'

- LLMs are probabilistic token predictors, not calculators.
- Feeding raw CSVs to GPT-4 leads to hallucinations.
- They invent correlations that sound plausible but are false.
- We need a deterministic foundation before semantics.

```
❏ Naive Approach:  
[ Raw CSV ] -> [ LLM Prompt ] -> "Your coffee intake predicts rain." (HALLUCINATION)  
  
❏ Engineering Approach:  
[ Raw CSV ] -> [ Scipy/Pandas Engine ] -> [ Validated Stats ] -> [ LLM ] -> "True Insight"
```

SYSTEM ARCHITECTURE

- 1. DATA INGESTION (GitHub Actions)
Garmin API -> PostgreSQL (Upsert)
- 2. DETERMINISTIC MATH LAYER (Python/Numpy)
 - Pearson Correlations (Linear)
 - AR(1) Models (Persistence)
 - Markov Chains (State Transitions)
 - Anomaly Detection (Percentiles)
- 3. AGENTIC LAYER (CrewAI)
 - 5 Specialized Agents (Read-Only)
 - Synthsizer with Long-Term Memory



Live Streamlit Dashboard

DETERMINISM BEFORE SEMANTICS

1. We validate relationships mathematically before the AI sees them.

2. **Pearson Correlation :**
$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

3. **Markov Transition :**
$$P_{ij} = P(S_{t+1} = j \mid S_t = i)$$

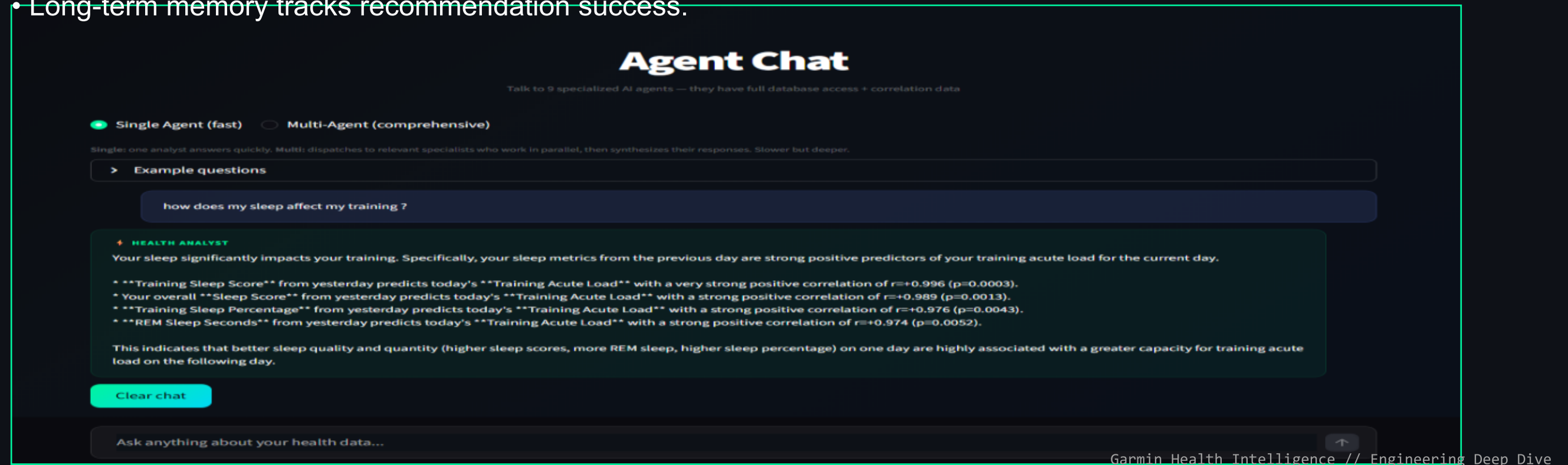
4. **AR(1) Model :**
$$X_t = c + \phi X_{t-1} + \varepsilon_t$$



Computed Correlation Matrix

THE AGENT TEAM

- 5 Specialists (Stats, Sleep, Recovery, Patterns, Lead).
- Read-Only SQL Access (Sandboxed).
- Pre-computed context injected into prompt.
- Long-term memory tracks recommendation success.



DEEP DIVE: SIGNAL VS NOISE

