

# Batch Experiment Report

Date: 2026-02-28 15:55:09 Batch Directory: vision\_temp\_0.1

## 1. Experimental Setup

Parameter	Value
Models	Text: openai/gpt-4o-mini Vision: google/gemini-2.0-flash-001
Temperatures	Text: 0.1 Vision: 0.7
Iterations	30
Embedding Model	google/siglip-so400m-patch14-384

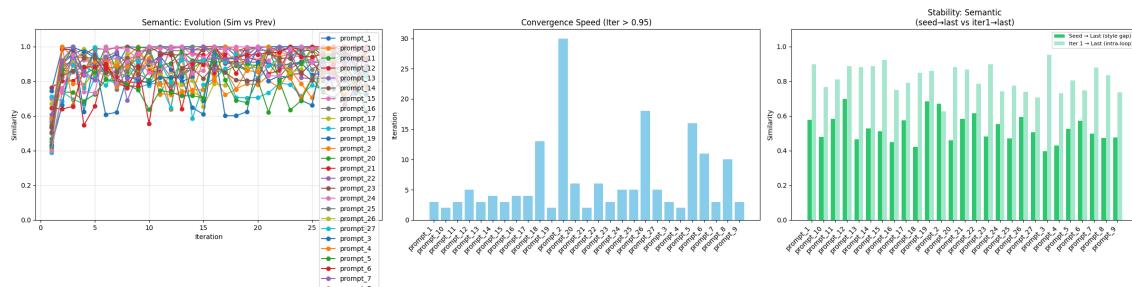
## 2. Batch Analysis

### Aggregate Statistics

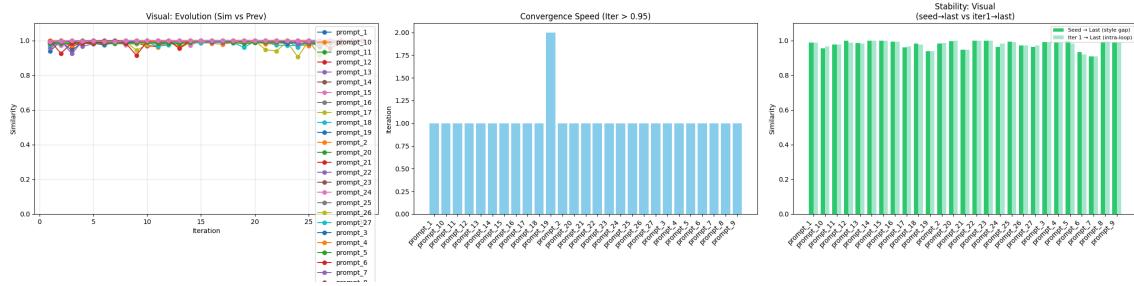
Metric	Mean	Min	Max
Semantic Stability — Seed→Last (style gap)	0.529	0.398	0.699
Semantic Stability — Iter 1→Last (intra-loop)	0.816	0.626	0.953
Visual Stability — First Image→Last	0.978	0.909	1.000
Visual Stability — Iter 2→Last (intra-loop)	0.978	0.909	1.000
Semantic Convergence (Step)	7.4	3	31
Visual Convergence (Step)	2.0	2	3
MMD Code Similarity (mean step-by-step)	0.745	0.377	0.991
Total Cost	\$0.3030 (Avg: \$0.0112/run)	-	-

## Visualizations

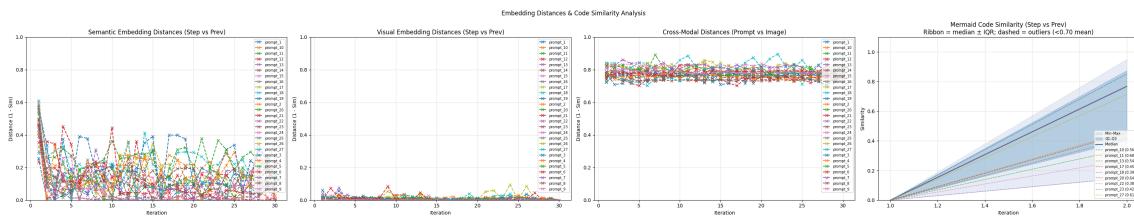
### Semantic Analysis



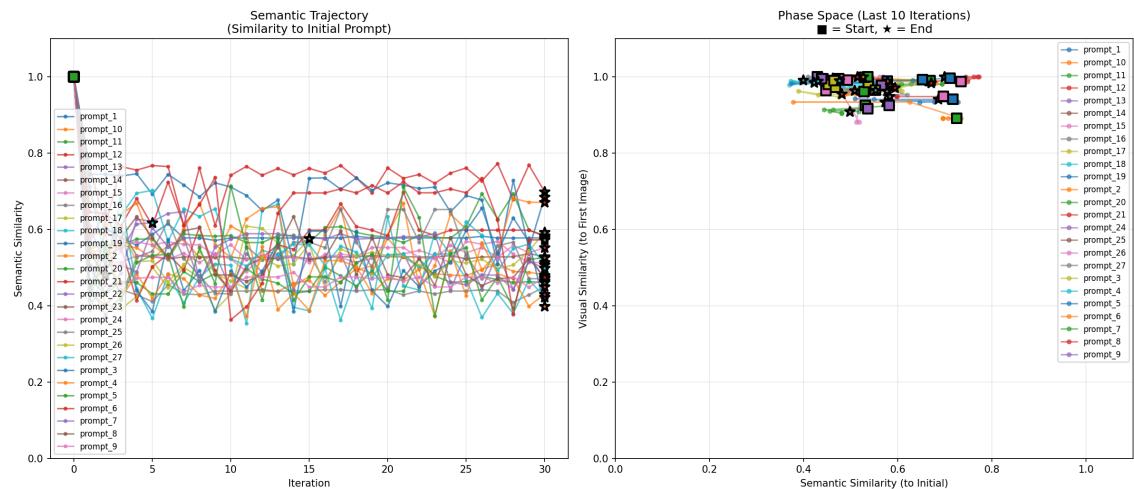
## Visual Analysis



## Embedding Distances

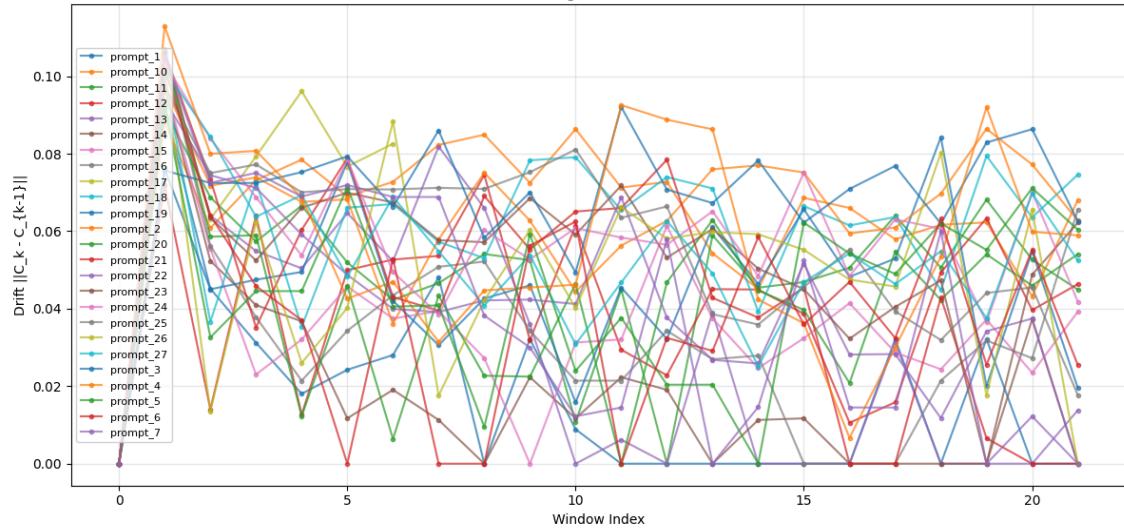


## Trajectory Analysis

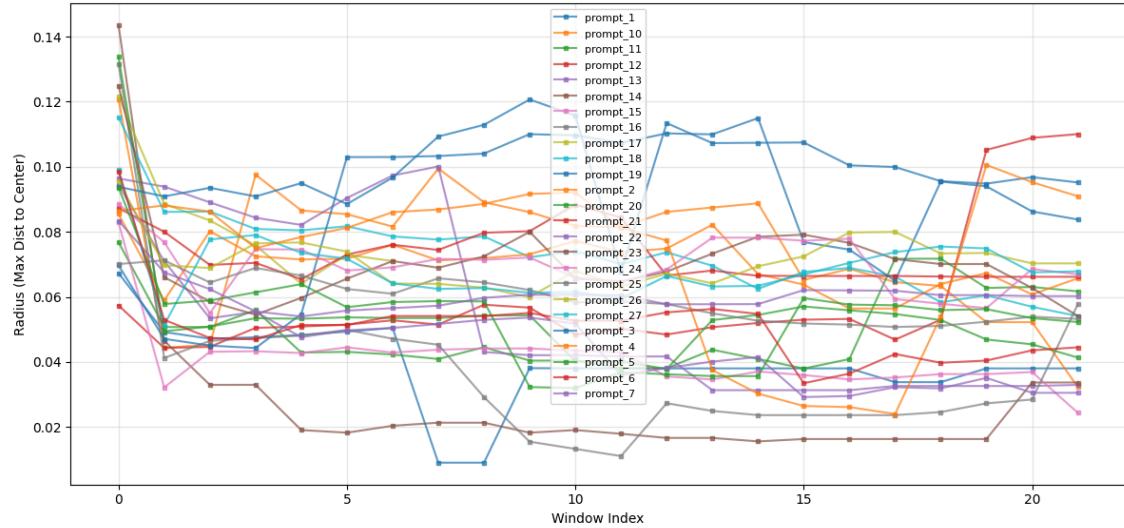


## Window Analysis - Text Embeddings

Text Embedding: Window Center Drift

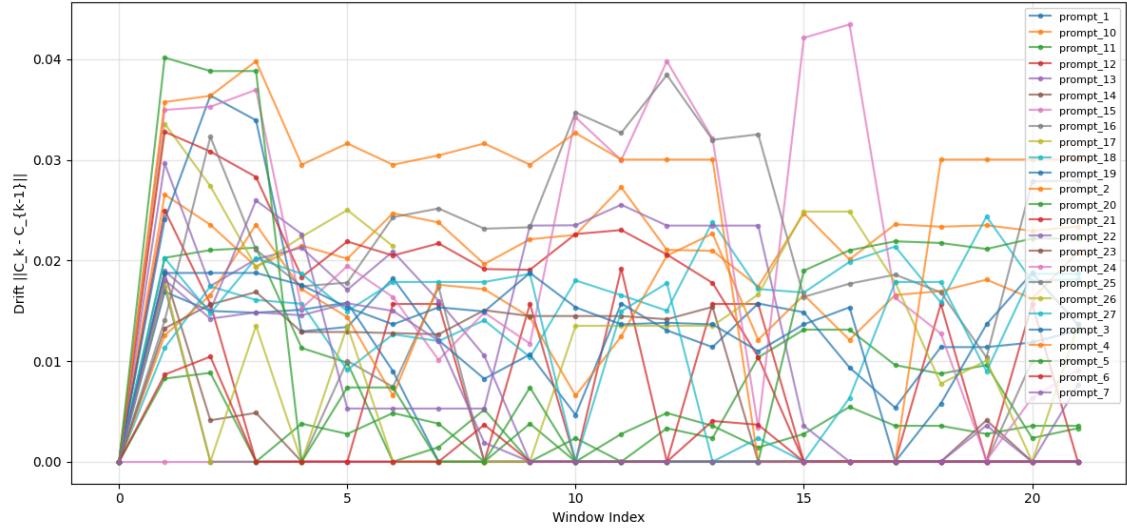


Text Embedding: Window Radius (Dispersion)

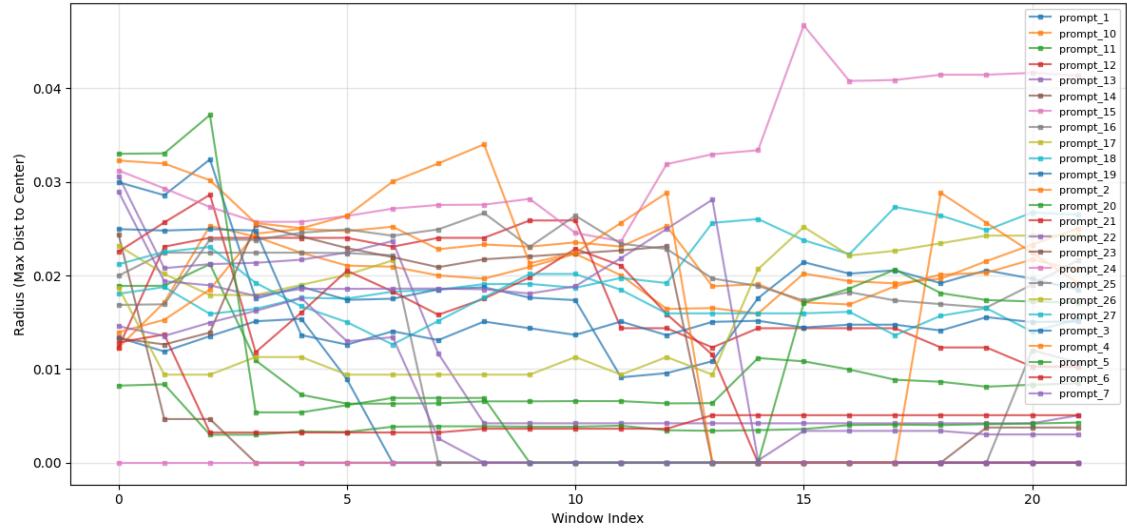


### Window Analysis - Visual Embeddings

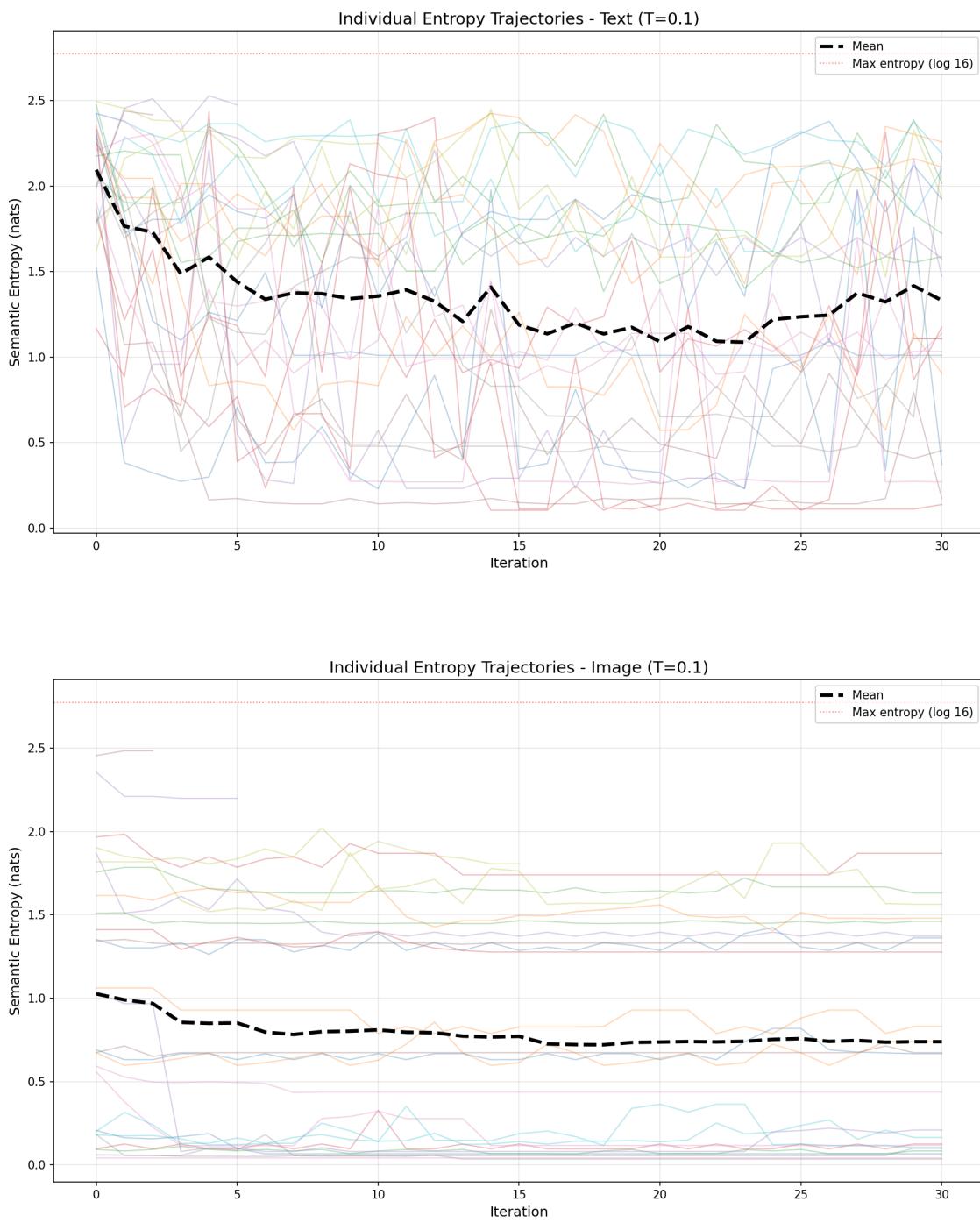
Visual Embedding: Window Center Drift



Visual Embedding: Window Radius (Dispersion)



### Semantic Entropy Trajectories



### 3. Individual Experiments

Run	Initial Prompt	Sem Stab.	Vis Stab.	Cost
<a href="#">prompt_1</a>	A simple user authentication flow. User inputs credentials, ...	0.578	0.989	\$0.0057

<a href="#"><u>prompt_10</u></a>	A CI/CD pipeline with rollback. Commit -> Test. Fail? Notify...	0.480	0.956	\$0.0122
<a href="#"><u>prompt_11</u></a>	An event sourcing loop. User Action -> Event Bus -> Consumer...	0.584	0.977	\$0.0088
<a href="#"><u>prompt_12</u></a>	A garbage collection mark-and-sweep. Scanner traces roots, m...	0.699	1.000	\$0.0095
<a href="#"><u>prompt_13</u></a>	A home thermostat. Sensor reads Temp. If Low, Heater On. Tem...	0.465	0.985	\$0.0113
<a href="#"><u>prompt_14</u></a>	The hydrological water cycle. Ocean water Evaporates into Cl...	0.528	0.999	\$0.0076
<a href="#"><u>prompt_15</u></a>	An engine combustion cycle. Intake valve opens, Piston moves...	0.513	1.000	\$0.0169
<a href="#"><u>prompt_16</u></a>	The Hero's Journey. Ordinary World -> Call to Adventure -> O...	0.450	0.995	\$0.0134
<a href="#"><u>prompt_17</u></a>	The Scientific Method. Observe -> Hypothesize -> Experiment ...	0.576	0.963	\$0.0239
<a href="#"><u>prompt_18</u></a>	The Design Thinking process. Empathize -> Define -> Ideate -...	0.422	0.985	\$0.0106
<a href="#"><u>prompt_19</u></a>	E-commerce Microservices. API Gateway routes to Product Serv...	0.685	0.941	\$0.0106
<a href="#"><u>prompt_2</u></a>	A password reset flow. User requests reset, System sends Ema...	0.671	0.984	\$0.0076
<a href="#"><u>prompt_20</u></a>	A Kubernetes Cluster Architecture. Control Plane manages Nod...	0.461	0.997	\$0.0092
<a href="#"><u>prompt_21</u></a>	A React Component Tree. App Component holds Header, Main, Fo...	0.583	0.948	\$0.0113
<a href="#"><u>prompt_22</u></a>	The Solar System. Sun at center. Inner Planets (Mercury, Ven...	0.617	1.000	\$0.0324
<a href="#"><u>prompt_23</u></a>	A Biological Trophic Pyramid. Producers (Plants) at base. Pr...	0.481	1.000	\$0.0187
<a href="#"><u>prompt_24</u></a>	Human Nervous System. Central NS (Brain, Spine). Peripheral ...	0.553	0.965	\$0.0224
<a href="#"><u>prompt_25</u></a>	A Biological Taxonomy. Domain Eukarya -> Kingdom Animalia ->...	0.471	0.993	\$0.0055
<a href="#"><u>prompt_26</u></a>	A Corporate Org Chart. CEO leads VP Eng, VP Sales, VP HR. VP...	0.593	0.971	\$0.0140
<a href="#"><u>prompt_27</u></a>	A Decision Tree for Loan Approval. Credit Score > 700? Yes -...	0.508	0.964	\$0.0103

<a href="#"><u>prompt_3</u></a>	An email subscription process. User submits email, System ad...	0.398	0.991	\$0.0040
<a href="#"><u>prompt_4</u></a>	A gravity-fed water filtration system. Rainwater to Gutter, ...	0.430	0.989	\$0.0077
<a href="#"><u>prompt_5</u></a>	A manufacturing assembly line. Raw material enters Conveyor,...	0.525	1.000	\$0.0074
<a href="#"><u>prompt_6</u></a>	A traffic light sequence. Green light stays for 60s, turns Y...	0.573	0.934	\$0.0051
<a href="#"><u>prompt_7</u></a>	A logical syllogism chain. Premise A implies B. B implies C....	0.498	0.909	\$0.0064
<a href="#"><u>prompt_8</u></a>	A historical timeline of 3 events. Event A causes Event B. E...	0.474	0.995	\$0.0047
<a href="#"><u>prompt_9</u></a>	An algebraic derivation. Start with Equation 1, substitute V...	0.475	0.990	\$0.0058

## 4. Detailed Experiment Log

### **[prompt\\_1](#)**

#### **Initial Prompt:**

*A simple user authentication flow. User inputs credentials, System validates hash, returns Token or Error.*

- **Cost:** \$0.0057
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

### **[prompt\\_10](#)**

#### **Initial Prompt:**

*A CI/CD pipeline with rollback. Commit -> Test. Fail? Notify Dev (Loop). Pass? Deploy. Monitor Health. bad? Rollback (Loop).*

- **Cost:** \$0.0122
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

### **[prompt\\_11](#)**

#### **Initial Prompt:**

*An event sourcing loop. User Action -> Event Bus -> Consumer updates View DB. View reflects new state to User.*

- **Cost:** \$0.0088
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

### **[prompt\\_12](#)**

#### **Initial Prompt:**

A garbage collection mark-and-sweep. Scanner traces roots, marks reachable, sweeps unreachable, frees memory, waits for next cycle.

- **Cost:** \$0.0095
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

#### **prompt\_13**

#### **Initial Prompt:**

A home thermostat. Sensor reads Temp. If Low, Heater On. Temp Rises. If High, Heater Off. Temp Falls.

- **Cost:** \$0.0113
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

#### **prompt\_14**

#### **Initial Prompt:**

The hydrological water cycle. Ocean water Evaporates into Clouds. Clouds Condense into Rain. Rain flows into Rivers. Rivers return to Ocean.

- **Cost:** \$0.0076
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

#### **prompt\_15**

#### **Initial Prompt:**

An engine combustion cycle. Intake valve opens, Piston moves down. Valve closes, Piston compresses. Spark plug fires (Combustion). Piston driven down. Exhaust valve opens.

- **Cost:** \$0.0169
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

#### **prompt\_16**

#### **Initial Prompt:**

The Hero's Journey. Ordinary World -> Call to Adventure -> Ordeal -> Reward -> Return to Ordinary World (Changed).

- **Cost:** \$0.0134
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

#### **prompt\_17**

#### **Initial Prompt:**

The Scientific Method. Observe -> Hypothesize -> Experiment -> Analyze. Data supports? Theory. Data rejects? New Hypothesis (Loop).

- **Cost:** \$0.0239

- **Data:** [trajectory.json](#), [metrics.json](#)
- 

## **prompt\_18**

### **Initial Prompt:**

*The Design Thinking process. Empathize -> Define -> Ideate -> Prototype -> Test. Test fails? Ideate again (Loop).*

- **Cost:** \$0.0106
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

## **prompt\_19**

### **Initial Prompt:**

*E-commerce Microservices. API Gateway routes to Product Service (Product DB), Order Service (Order DB), User Service (User DB).*

- **Cost:** \$0.0106
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

## **prompt\_2**

### **Initial Prompt:**

*A password reset flow. User requests reset, System sends Email, User clicks Link, enters New Password.*

- **Cost:** \$0.0076
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

## **prompt\_20**

### **Initial Prompt:**

*A Kubernetes Cluster Architecture. Control Plane manages Nodes. Nodes contain Pods. Pods contain Containers.*

- **Cost:** \$0.0092
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

## **prompt\_21**

### **Initial Prompt:**

*A React Component Tree. App Component holds Header, Main, Footer. Main holds Sidebar and Content Area. Content Area holds Article List.*

- **Cost:** \$0.0113
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

## **prompt\_22**

### **Initial Prompt:**

*The Solar System. Sun at center. Inner Planets (Mercury, Venus, Earth) orbit Sun. Earth has Moon. Outer Planets (Jupiter, Saturn) orbit Sun. Jupiter has many Moons.*

- **Cost:** \$0.0324
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

## **prompt\_23**

### **Initial Prompt:**

*A Biological Trophic Pyramid. Producers (Plants) at base. Primary Consumers (Herbivores) eat Producers. Secondary Consumers (Carnivores) eat Primary. Decomposers recycle all.*

- **Cost:** \$0.0187
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

## **prompt\_24**

### **Initial Prompt:**

*Human Nervous System. Central NS (Brain, Spine). Peripheral NS splits into Somatic and Autonomic. Autonomic splits into Sympathetic and Parasympathetic.*

- **Cost:** \$0.0224
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

## **prompt\_25**

### **Initial Prompt:**

*A Biological Taxonomy. Domain Eukarya -> Kingdom Animalia -> Phylum Chordata -> Class Mammalia.*

- **Cost:** \$0.0055
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

## **prompt\_26**

### **Initial Prompt:**

*A Corporate Org Chart. CEO leads VP Eng, VP Sales, VP HR. VP Eng leads Backend, Frontend, QA Teams.*

- **Cost:** \$0.0140
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

## **prompt\_27**

### **Initial Prompt:**

*A Decision Tree for Loan Approval. Credit Score > 700? Yes -> Income > 50k? Yes -> Approve. No -> Deny.*

- **Cost:** \$0.0103
  - **Data:** [trajectory.json](#), [metrics.json](#)
-

### **prompt\_3**

#### **Initial Prompt:**

*An email subscription process. User submits email, System adds to DB, System sends Welcome Email.*

- **Cost:** \$0.0040
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

### **prompt\_4**

#### **Initial Prompt:**

*A gravity-fed water filtration system. Rainwater to Gutter, Gutter to Pipe, Pipe to Filter, Filter to Tank.*

- **Cost:** \$0.0077
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

### **prompt\_5**

#### **Initial Prompt:**

*A manufacturing assembly line. Raw material enters Conveyor, Stamping Machine shapes it, Painting Station colors it, Packaging wraps it.*

- **Cost:** \$0.0074
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

### **prompt\_6**

#### **Initial Prompt:**

*A traffic light sequence. Green light stays for 60s, turns Yellow for 5s, turns Red for 60s.*

- **Cost:** \$0.0051
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

### **prompt\_7**

#### **Initial Prompt:**

*A logical syllogism chain. Premise A implies B. B implies C. Therefore A implies C.*

- **Cost:** \$0.0064
  - **Data:** [trajectory.json](#), [metrics.json](#)
- 

### **prompt\_8**

#### **Initial Prompt:**

*A historical timeline of 3 events. Event A causes Event B. Event B triggers Reaction C.*

- **Cost:** \$0.0047
  - **Data:** [trajectory.json](#), [metrics.json](#)
-

## **prompt\_9**

### **Initial Prompt:**

*An algebraic derivation. Start with Equation 1, substitute Variable X, simplify to Result Y.*

- **Cost:** \$0.0058
  - **Data:** [trajectory.json](#), [metrics.json](#)
-