

Batch Experiment Report

Date: 2026-02-28 15:52:07 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

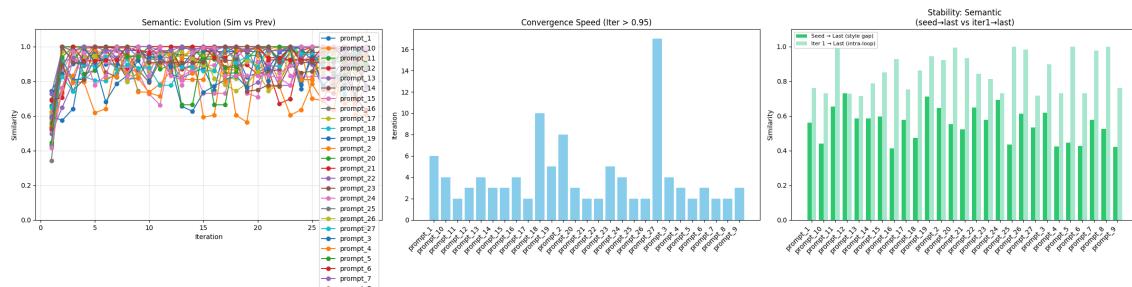
2. Batch Analysis

Aggregate Statistics

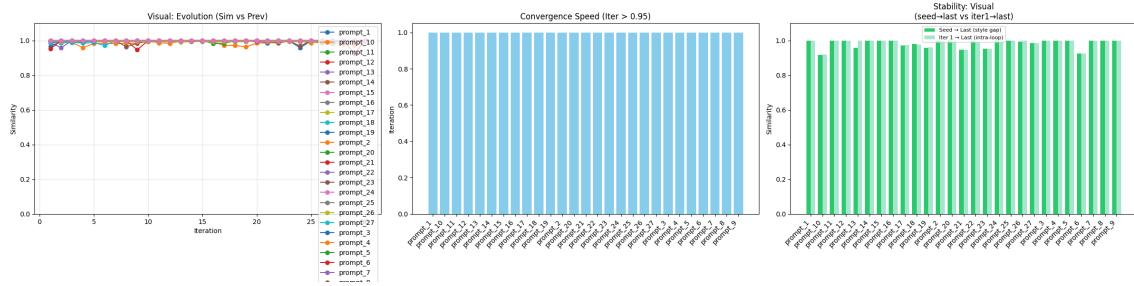
Metric	Mean	Min	Max
Semantic Stability — Seed→Last (style gap)	0.556	0.414	0.732
Semantic Stability — Iter 1→Last (intra-loop)	0.855	0.714	1.000
Visual Stability — First Image→Last	0.985	0.918	1.000
Visual Stability — Iter 2→Last (intra-loop)	0.987	0.919	1.000
Semantic Convergence (Step)	5.1	3	18
Visual Convergence (Step)	2.0	2	2
MMD Code Similarity (mean step-by-step)	0.882	0.424	1.000
Total Cost	\$0.2226 (Avg: \$0.0082/run)	-	-

Visualizations

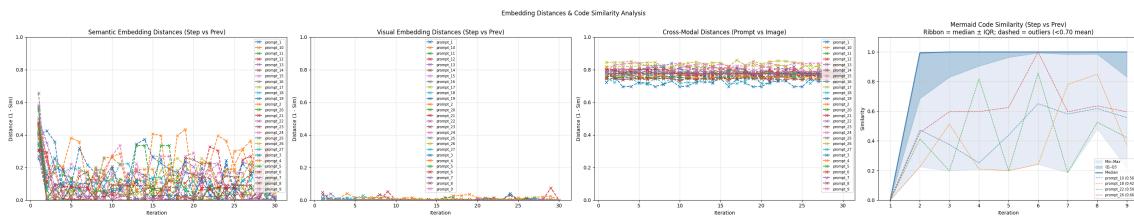
Semantic Analysis



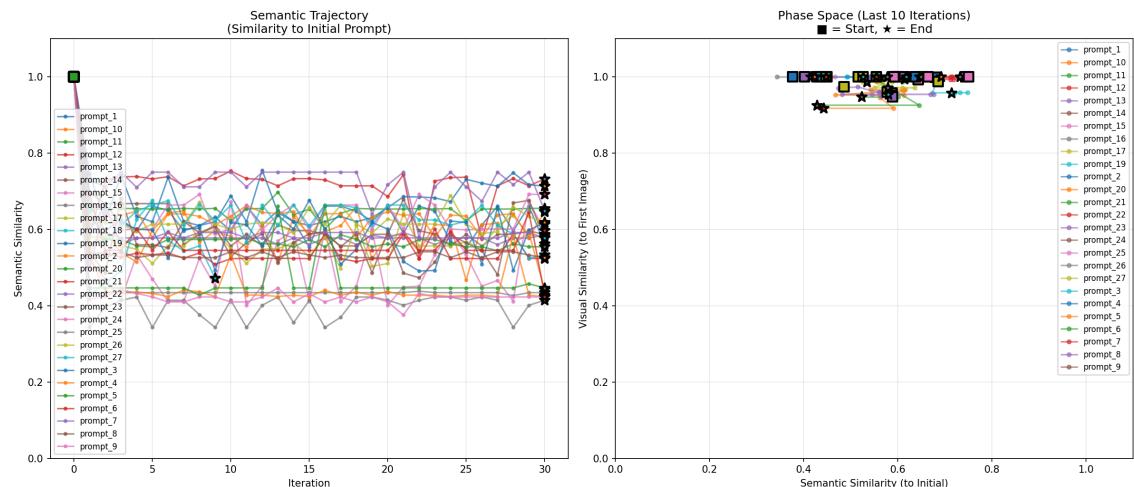
Visual Analysis



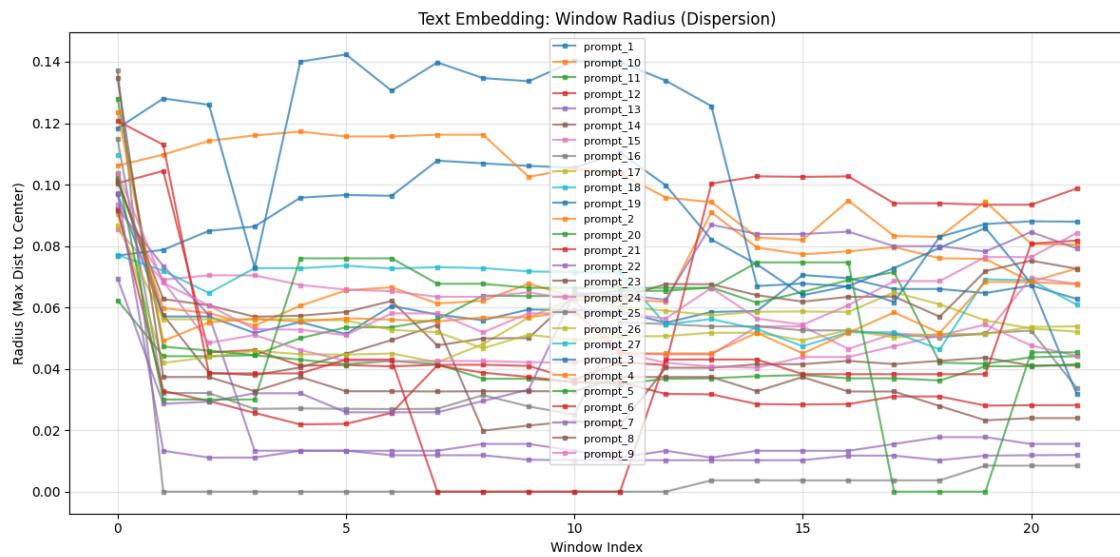
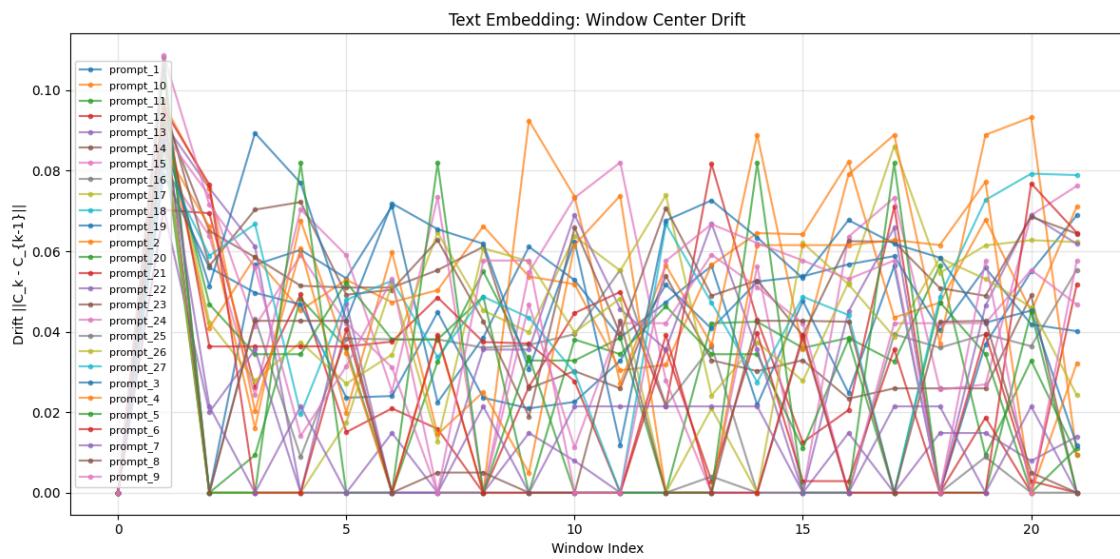
Embedding Distances



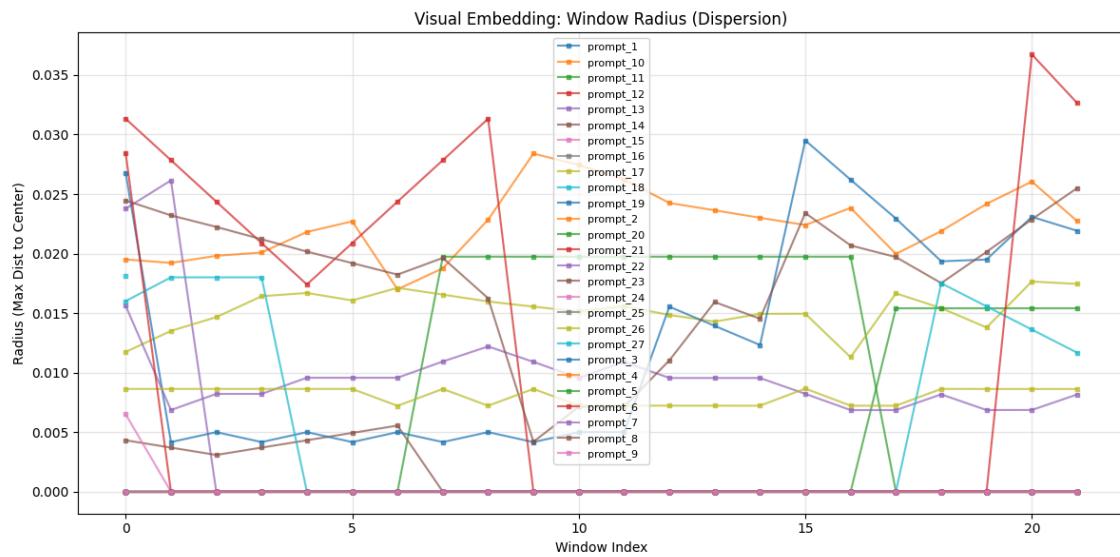
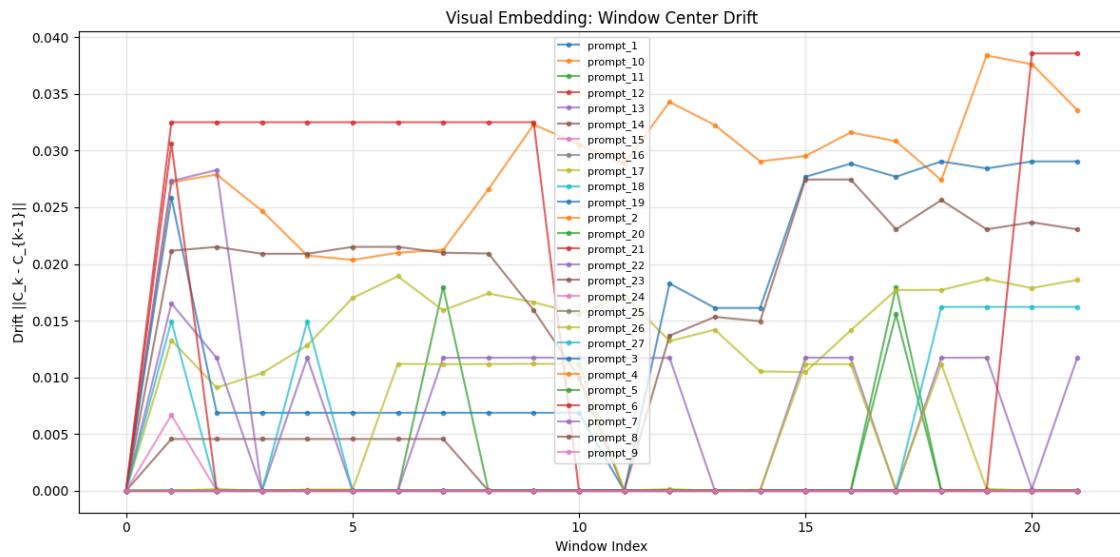
Trajectory Analysis



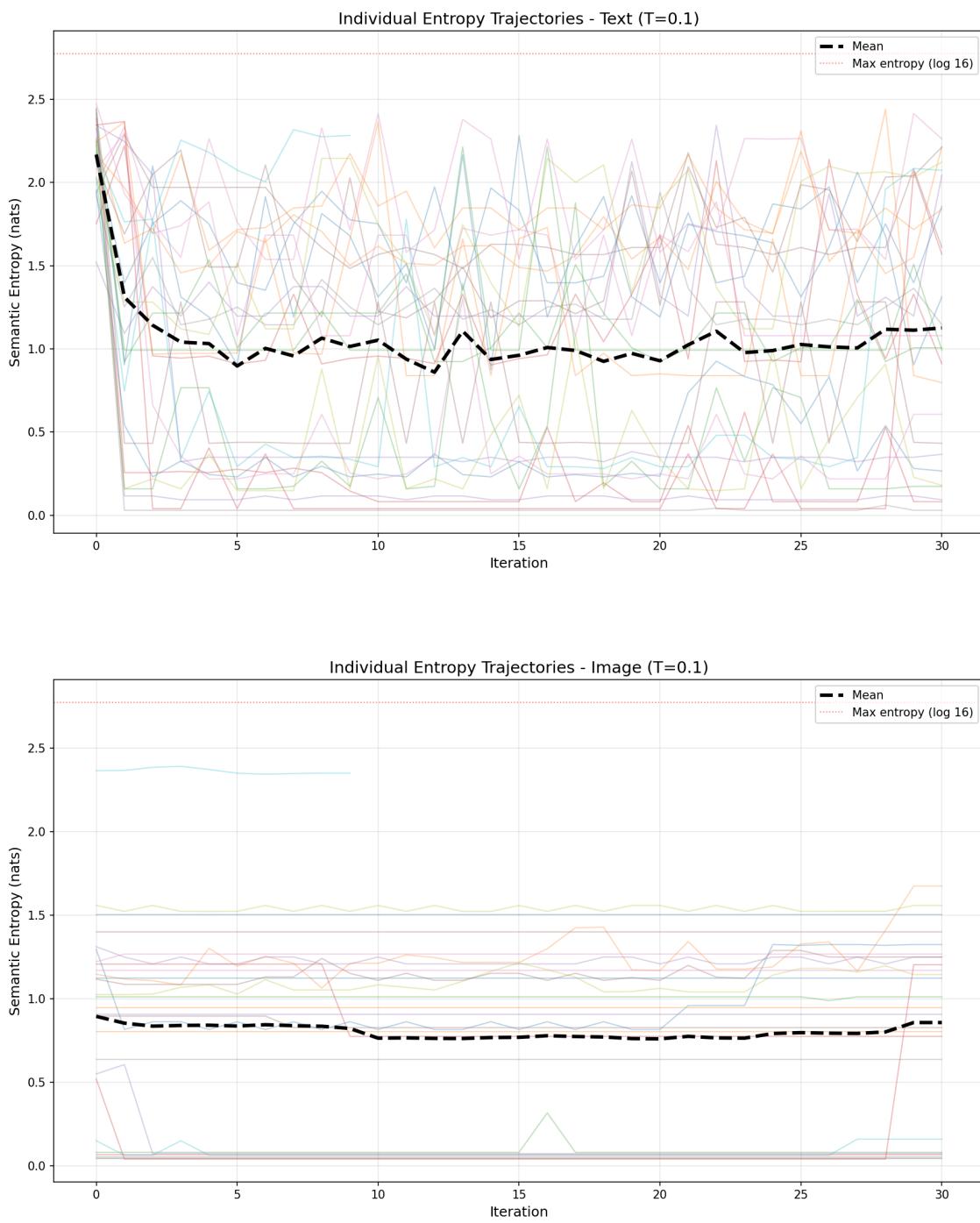
Window Analysis - Text Embeddings



Window Analysis - Visual Embeddings



Semantic Entropy Trajectories



3. Individual Experiments

Run	Initial Prompt	Sem Stab.	Vis Stab.	Cost
prompt_1	A simple user authentication flow. User inputs credentials, ...	0.562	1.000	\$0.0047

<u>prompt_10</u>	A CI/CD pipeline with rollback. Commit -> Test. Fail? Notify...	0.442	0.918	\$0.0124
<u>prompt_11</u>	An event sourcing loop. User Action -> Event Bus -> Consumer...	0.655	1.000	\$0.0077
<u>prompt_12</u>	A garbage collection mark-and-sweep. Scanner traces roots, m...	0.732	1.000	\$0.0096
<u>prompt_13</u>	A home thermostat. Sensor reads Temp. If Low, Heater On. Tem...	0.585	0.960	\$0.0126
<u>prompt_14</u>	The hydrological water cycle. Ocean water Evaporates into Cl...	0.587	0.999	\$0.0078
<u>prompt_15</u>	An engine combustion cycle. Intake valve opens, Piston moves...	0.597	1.000	\$0.0101
<u>prompt_16</u>	The Hero's Journey. Ordinary World -> Call to Adventure -> O...	0.414	1.000	\$0.0080
<u>prompt_17</u>	The Scientific Method. Observe -> Hypothesize -> Experiment ...	0.578	0.972	\$0.0115
<u>prompt_18</u>	The Design Thinking process. Empathize -> Define -> Ideate -...	0.472	0.980	\$0.0131
<u>prompt_19</u>	E-commerce Microservices. API Gateway routes to Product Serv...	0.713	0.958	\$0.0109
<u>prompt_2</u>	A password reset flow. User requests reset, System sends Ema...	0.646	1.000	\$0.0076
<u>prompt_20</u>	A Kubernetes Cluster Architecture. Control Plane manages Nod...	0.554	1.000	\$0.0071
<u>prompt_21</u>	A React Component Tree. App Component holds Header, Main, Fo...	0.523	0.947	\$0.0090
<u>prompt_22</u>	The Solar System. Sun at center. Inner Planets (Mercury, Ven...	0.649	1.000	\$0.0107
<u>prompt_23</u>	A Biological Trophic Pyramid. Producers (Plants) at base. Pr...	0.579	0.954	\$0.0089
<u>prompt_24</u>	Human Nervous System. Central NS (Brain, Spine). Peripheral ...	0.692	1.000	\$0.0129
<u>prompt_25</u>	A Biological Taxonomy. Domain Eukarya -> Kingdom Animalia ->...	0.434	1.000	\$0.0050
<u>prompt_26</u>	A Corporate Org Chart. CEO leads VP Eng, VP Sales, VP HR. VP...	0.614	0.994	\$0.0108
<u>prompt_27</u>	A Decision Tree for Loan Approval. Credit Score > 700? Yes -...	0.534	0.987	\$0.0054

<u>prompt_3</u>	An email subscription process. User submits email, System ad...	0.618	1.000	\$0.0040
<u>prompt_4</u>	A gravity-fed water filtration system. Rainwater to Gutter, ...	0.424	1.000	\$0.0076
<u>prompt_5</u>	A manufacturing assembly line. Raw material enters Conveyor,...	0.447	1.000	\$0.0069
<u>prompt_6</u>	A traffic light sequence. Green light stays for 60s, turns Y...	0.428	0.926	\$0.0051
<u>prompt_7</u>	A logical syllogism chain. Premise A implies B. B implies C....	0.578	1.000	\$0.0047
<u>prompt_8</u>	A historical timeline of 3 events. Event A causes Event B. E...	0.526	1.000	\$0.0039
<u>prompt_9</u>	An algebraic derivation. Start with Equation 1, substitute V...	0.423	1.000	\$0.0044

4. Detailed Experiment Log

[prompt_1](#)

Initial Prompt:

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

- **Cost:** \$0.0047
 - **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:** \$0.0124
 - **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:** \$0.0077
 - **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:** \$0.0096
 - **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:** \$0.0126
 - **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:** \$0.0078
 - **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:** \$0.0101
 - **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:** \$0.0080
 - **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:** \$0.0115

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

prompt_18

Initial Prompt:

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

- **Cost:** \$0.0131
 - **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:** \$0.0109
 - **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:** \$0.0071
 - **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:** \$0.0090
 - **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:** \$0.0107
 - **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:** \$0.0089
 - **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:** \$0.0129
 - **Data:** [trajectory.json](#), [metrics.json](#)
-

prompt_25

Initial Prompt:

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

- **Cost:** \$0.0050
 - **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:** \$0.0108
 - **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:** \$0.0054
 - **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:** \$0.0076
 - **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:** \$0.0069
 - **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:** \$0.0047
 - **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:** \$0.0039
 - **Data:** [trajectory.json](#), [metrics.json](#)
-

prompt_9

Initial Prompt:

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

- **Cost:** \$0.0044
 - **Data:** [trajectory.json](#), [metrics.json](#)
-