

Batch Experiment Report

Date: 2026-02-28 18:51:01 Batch Directory: vision_temp_0.4

1. Experimental Setup

Parameter	Value
Models	Text: openai/gpt-4o-mini Vision: google/gemini-2.0-flash-001
Temperatures	Text: 0.1 Vision: 0.4
Iterations	30
Embedding Model	google/siglip-so400m-patch14-384

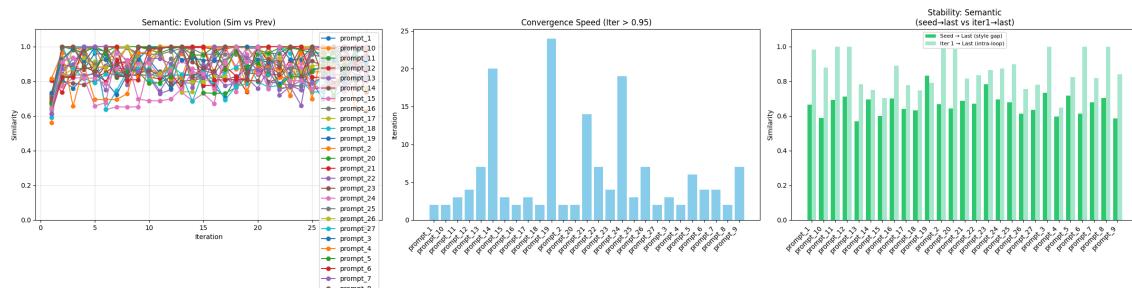
2. Batch Analysis

Aggregate Statistics

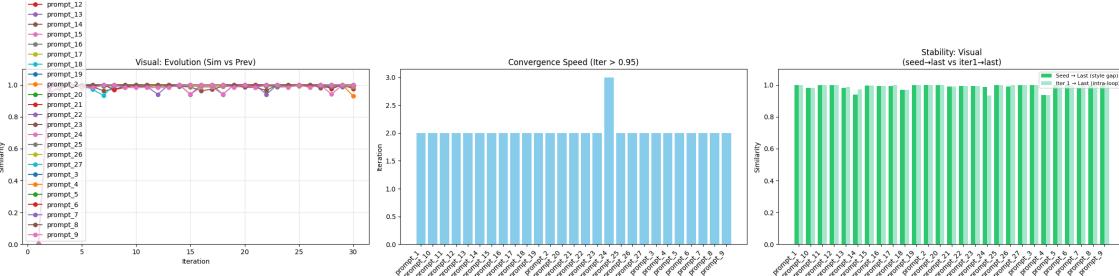
Metric	Mean	Min	Max
Semantic Stability — Seed→Last (style gap)	0.668	0.570	0.833
Semantic Stability — Iter 1→Last (intra-loop)	0.861	0.648	1.000
Visual Stability — First Image→Last	0.990	0.938	1.000
Visual Stability — Iter 2→Last (intra-loop)	0.990	0.933	1.000
Semantic Convergence (Step)	6.9	3	25
Visual Convergence (Step)	3.0	3	4
MMD Code Similarity (mean step-by-step)	0.849	0.590	1.000
Total Cost	\$0.2258 (Avg: \$0.0084/run)	-	-

Visualizations

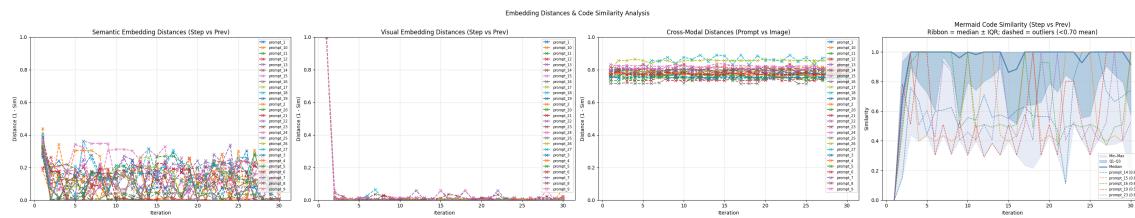
Semantic Analysis



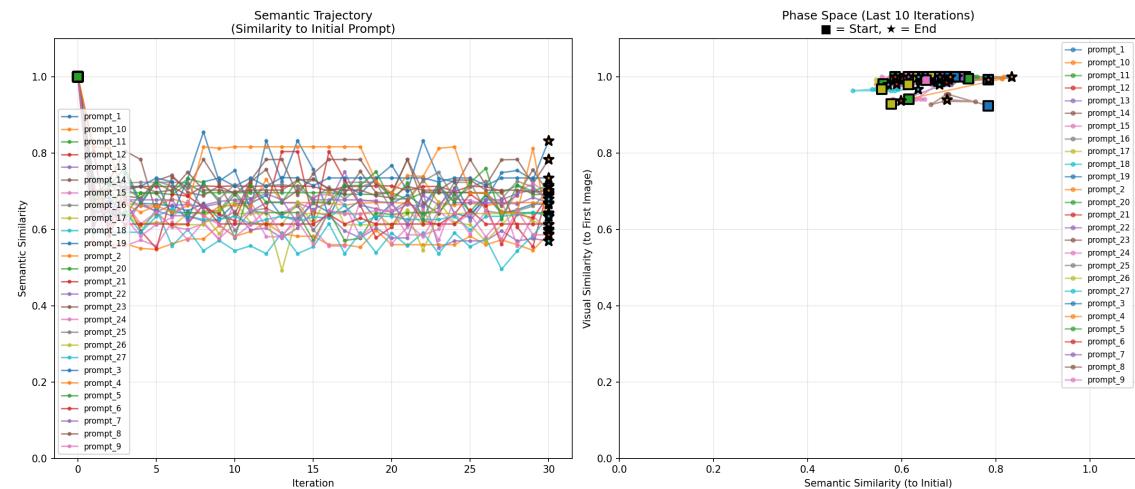
Visual Analysis



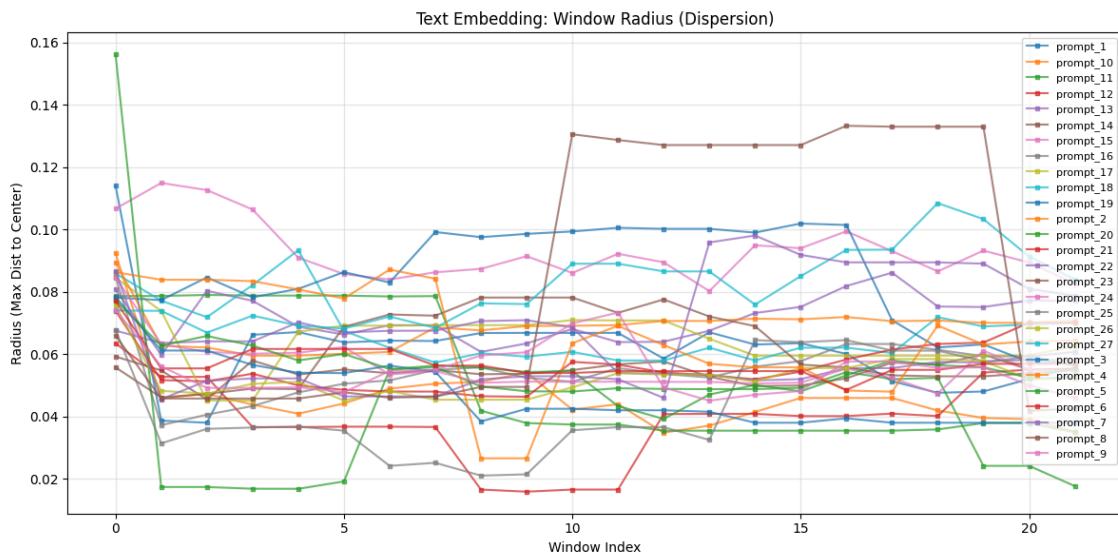
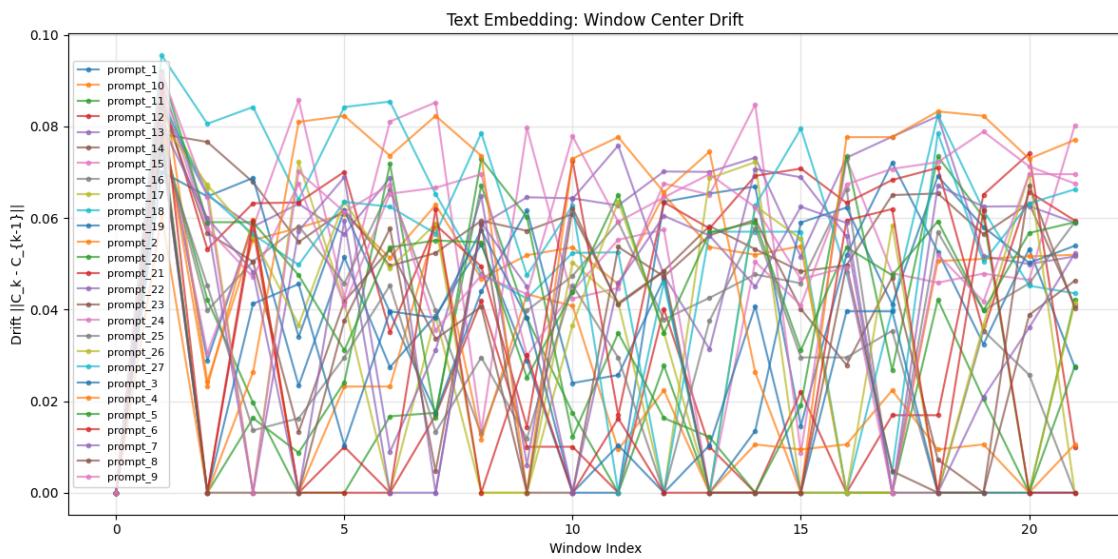
Embedding Distances



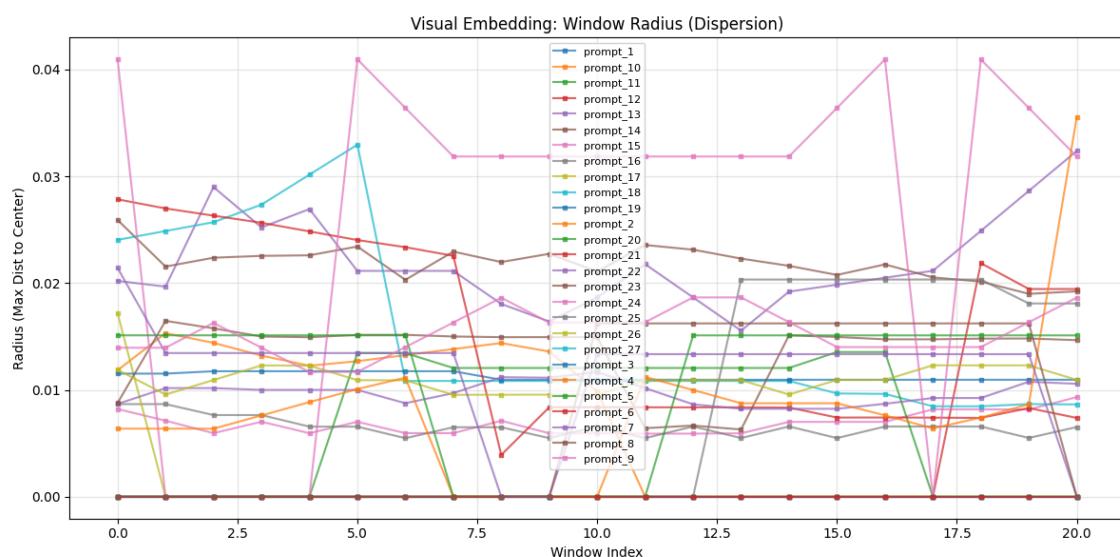
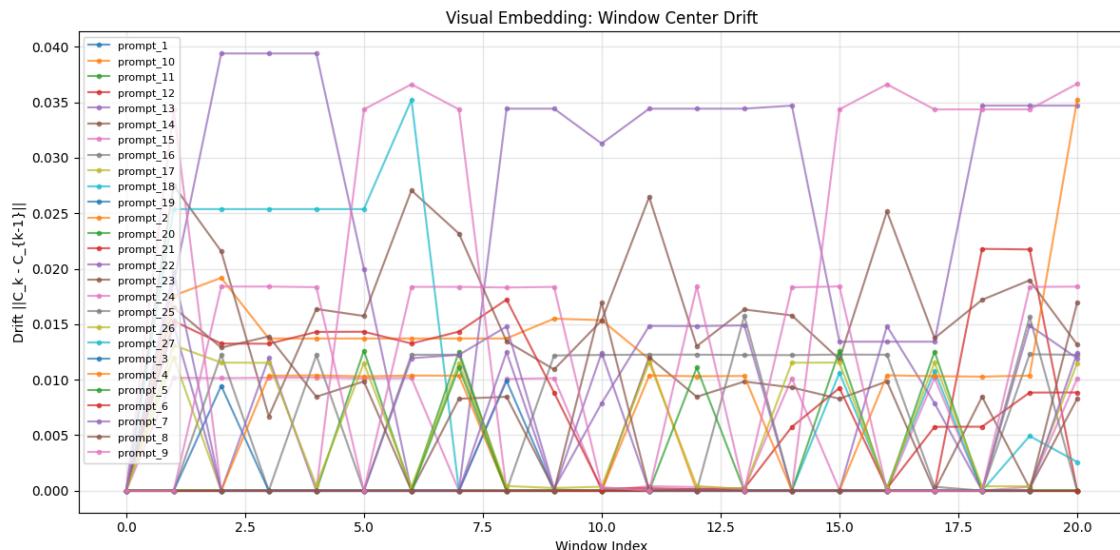
Trajectory Analysis



Window Analysis - Text Embeddings



Window Analysis - Visual Embeddings



Semantic Entropy Trajectories

Text Trajectories

Image Trajectories

3. Individual Experiments

Run	Initial Prompt	Sem Stab.	Vis Stab.	Cost
-----	----------------	-----------	-----------	------

<u>prompt_1</u>	A simple user authentication flow. User inputs credentials, ...	0.666	1.000	\$0.0052
<u>prompt_10</u>	A CI/CD pipeline with rollback. Commit -> Test. Fail? Notify...	0.588	0.982	\$0.0106
<u>prompt_11</u>	An event sourcing loop. User Action -> Event Bus -> Consumer...	0.692	1.000	\$0.0085
<u>prompt_12</u>	A garbage collection mark-and-sweep. Scanner traces roots, m...	0.713	1.000	\$0.0075
<u>prompt_13</u>	A home thermostat. Sensor reads Temp. If Low, Heater On. Tem...	0.570	0.980	\$0.0078
<u>prompt_14</u>	The hydrological water cycle. Ocean water Evaporates into Cl...	0.695	0.940	\$0.0059
<u>prompt_15</u>	An engine combustion cycle. Intake valve opens, Piston moves...	0.601	0.995	\$0.0139
<u>prompt_16</u>	The Hero's Journey. Ordinary World -> Call to Adventure -> O...	0.700	0.993	\$0.0113
<u>prompt_17</u>	The Scientific Method. Observe -> Hypothesize -> Experiment ...	0.642	0.993	\$0.0099
<u>prompt_18</u>	The Design Thinking process. Empathize -> Define -> Ideate -...	0.633	0.968	\$0.0085
<u>prompt_19</u>	E-commerce Microservices. API Gateway routes to Product Serv...	0.833	1.000	\$0.0104
<u>prompt_2</u>	A password reset flow. User requests reset, System sends Ema...	0.669	1.000	\$0.0084
<u>prompt_20</u>	A Kubernetes Cluster Architecture. Control Plane manages Nod...	0.644	1.000	\$0.0069
<u>prompt_21</u>	A React Component Tree. App Component holds Header, Main, Fo...	0.688	0.991	\$0.0116
<u>prompt_22</u>	The Solar System. Sun at center. Inner Planets (Mercury, Ven...	0.670	0.993	\$0.0112
<u>prompt_23</u>	A Biological Trophic Pyramid. Producers (Plants) at base. Pr...	0.783	0.993	\$0.0121
<u>prompt_24</u>	Human Nervous System. Central NS (Brain, Spine). Peripheral ...	0.695	0.986	\$0.0146
<u>prompt_25</u>	A Biological Taxonomy. Domain Eukarya -> Kingdom Animalia ->...	0.680	1.000	\$0.0059
<u>prompt_26</u>	A Corporate Org Chart. CEO leads VP Eng, VP Sales, VP HR. VP...	0.612	0.991	\$0.0120

prompt_27	A Decision Tree for Loan Approval. Credit Score > 700? Yes -...	0.635	1.000	\$0.0057
prompt_3	An email subscription process. User submits email, System ad...	0.735	1.000	\$0.0043
prompt_4	A gravity-fed water filtration system. Rainwater to Gutter, ...	0.597	0.938	\$0.0077
prompt_5	A manufacturing assembly line. Raw material enters Conveyor,...	0.717	1.000	\$0.0073
prompt_6	A traffic light sequence. Green light stays for 60s, turns Y...	0.614	1.000	\$0.0052
prompt_7	A logical syllogism chain. Premise A implies B. B implies C....	0.679	0.981	\$0.0046
prompt_8	A historical timeline of 3 events. Event A causes Event B. E...	0.703	1.000	\$0.0039
prompt_9	An algebraic derivation. Start with Equation 1, substitute V...	0.586	1.000	\$0.0049

4. Detailed Experiment Log

[prompt_1](#)

Initial Prompt:

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

- **Cost:** \$0.0052
 - **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:** \$0.0106
 - **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:** \$0.0085
 - **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:** \$0.0075
 - **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:** \$0.0078
 - **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:** \$0.0059
 - **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:** \$0.0139
 - **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:** \$0.0113
 - **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:** \$0.0099
 - **Data:** [trajectory.json](#), [metrics.json](#)
-

prompt_18

Initial Prompt:

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

- **Cost:** \$0.0085
 - **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:** \$0.0104
 - **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:** \$0.0084
 - **Data:** [trajectory.json](#), [metrics.json](#)
-

prompt_20

Initial Prompt:

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

- **Cost:** \$0.0069
 - **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:** \$0.0116
 - **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:** \$0.0112
 - **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:** \$0.0121
 - **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:** \$0.0146
 - **Data:** [trajectory.json](#), [metrics.json](#)
-

prompt_25

Initial Prompt:

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

- **Cost:** \$0.0059
 - **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:** \$0.0120
 - **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:** \$0.0057
 - **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:** \$0.0043
 - **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:** \$0.0073
 - **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:** \$0.0052
 - **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:** \$0.0046
 - **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:** \$0.0049
 - **Data:** [trajectory.json](#), [metrics.json](#)
-