Batch Experiment Report

Date: 2026-02-28 17:40:39 **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: `0.1`
Iterations	30
Embedding Model	`google/siglip-so400m-patch14-384`

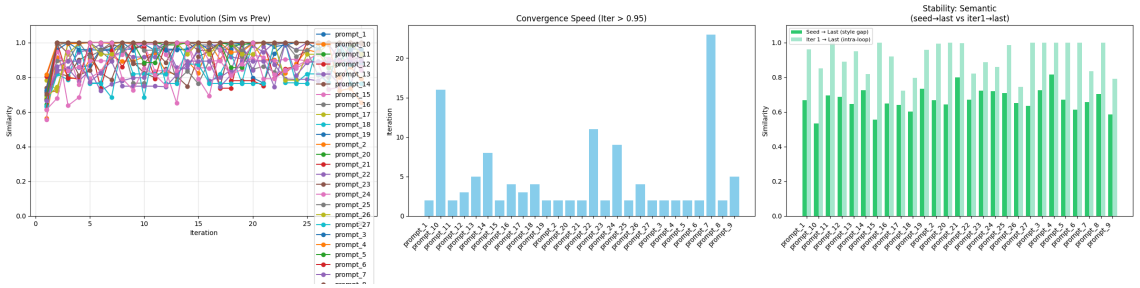
2. Batch Analysis

Aggregate Statistics

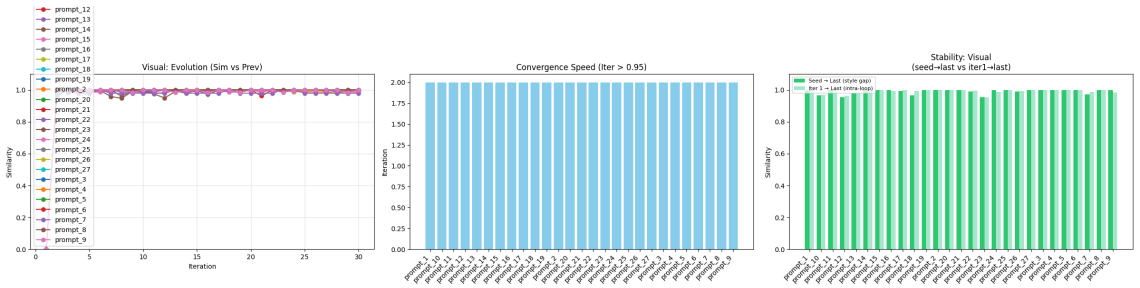
Metric	Mean	Min	Max
Semantic Stability — Seed→Last (style gap)	0.672	0.535	0.817
Semantic Stability — Iter 1→Last (intra-loop)	0.918	0.723	1.000
Visual Stability — First Image→Last	0.991	0.953	1.000
Visual Stability — Iter 2→Last (intra-loop)	0.992	0.953	1.000
Semantic Convergence (Step)	5.6	3	24
Visual Convergence (Step)	3.0	3	3
MMD Code Similarity (mean step-by-step)	0.888	0.648	0.996
Total Cost	$0.2344 (Avg: $0.0087/run)	-	-

Visualizations

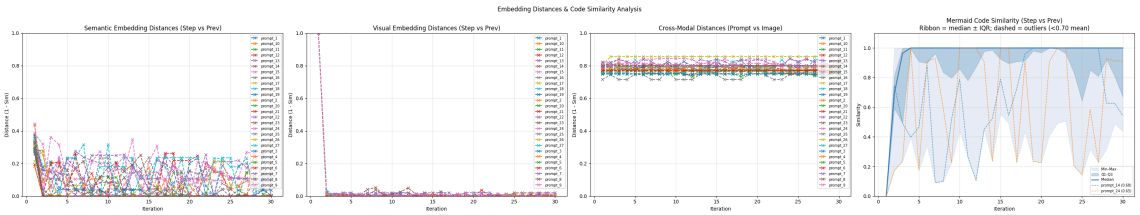
Semantic Analysis



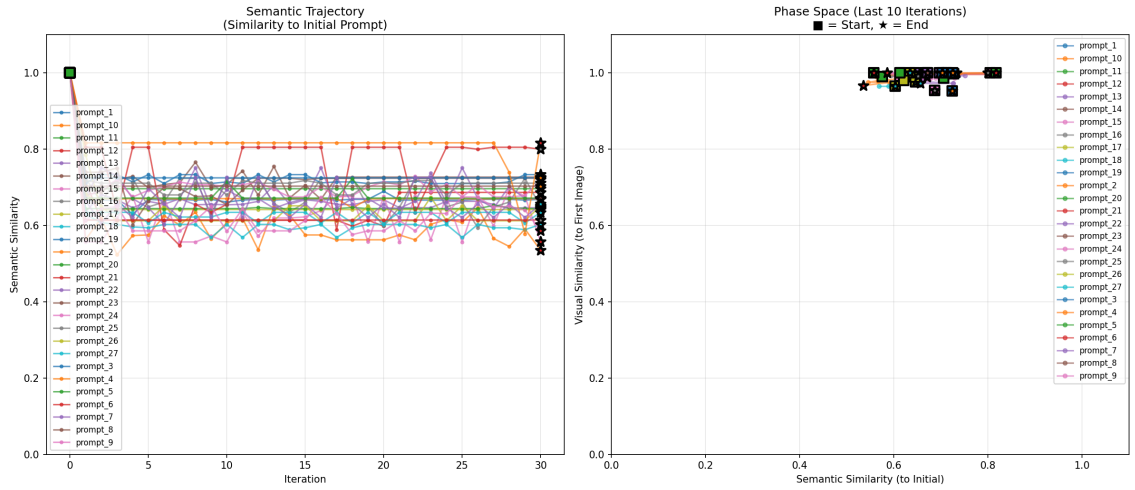
Visual Analysis



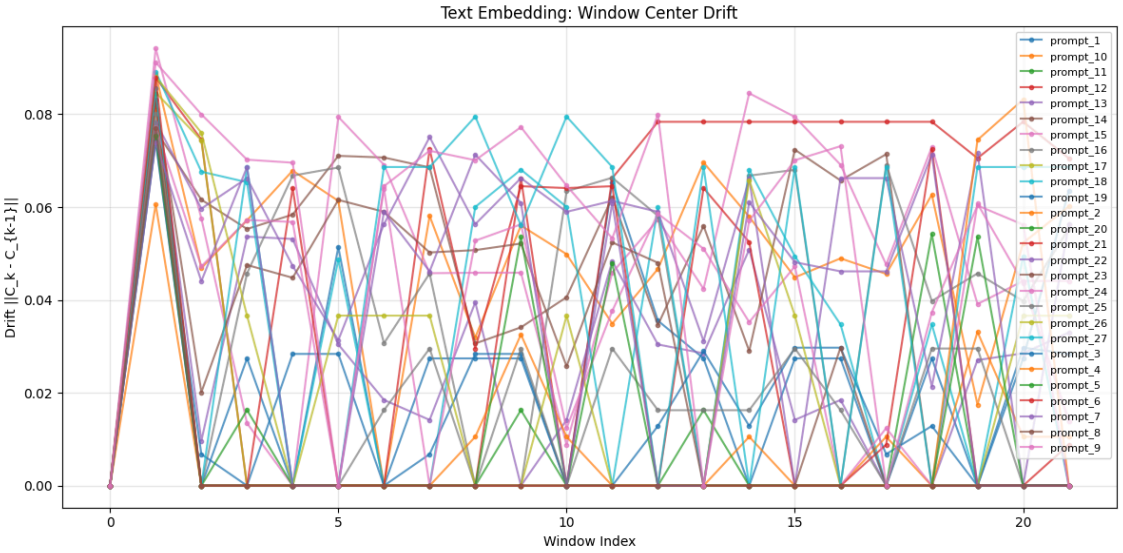
Embedding Distances

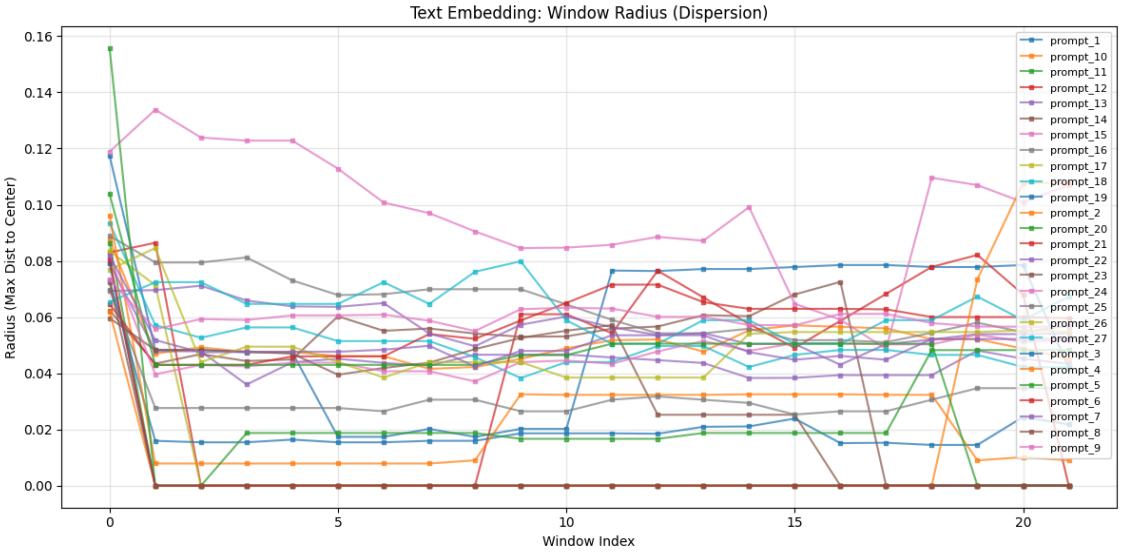


Trajectory Analysis

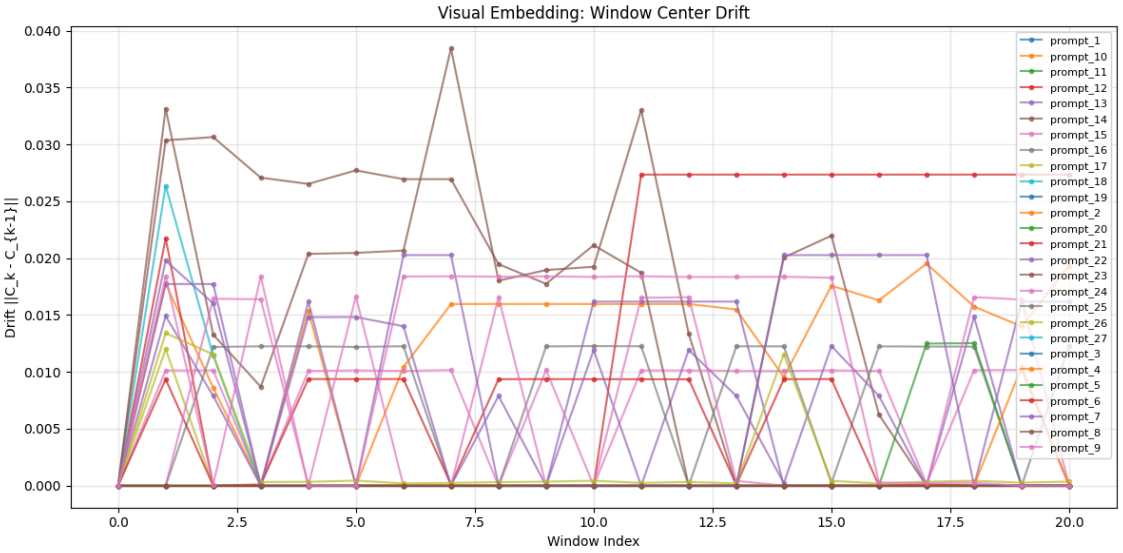


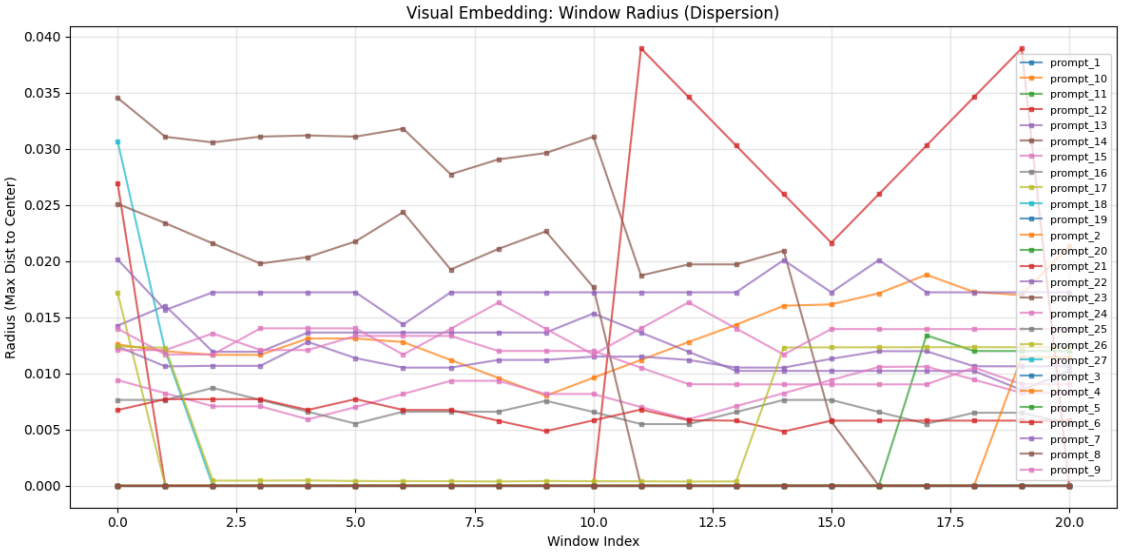
Window Analysis - Text Embeddings





Window Analysis - Visual Embeddings





Semantic Entropy Trajectories

Text Trajectories

Image Trajectories

3. Individual Experiments

Run	Initial Prompt	Sem Stab.	Vis Stab.	Cost

prompt_1	A simple user authentication flow. User inputs credentials, ...	0.668	1.000	$0.0052
prompt_10	A CI/CD pipeline with rollback. Commit -> Test. Fail? Notify...	0.535	0.967	$0.0111
prompt_11	An event sourcing loop. User Action -> Event Bus -> Consumer...	0.696	1.000	$0.0085
prompt_12	A garbage collection mark-and-sweep. Scanner traces roots, m...	0.686	0.954	$0.0083
prompt_13	A home thermostat. Sensor reads Temp. If Low, Heater On. Tem...	0.646	0.977	$0.0059
prompt_14	The hydrological water cycle. Ocean water Evaporates into Cl...	0.727	0.998	$0.0066
prompt_15	An engine combustion cycle. Intake valve opens, Piston moves...	0.557	1.000	$0.0149
prompt_16	The Hero's Journey. Ordinary World -> Call to Adventure -> O...	0.648	1.000	$0.0113
prompt_17	The Scientific Method. Observe -> Hypothesize -> Experiment ...	0.642	0.993	$0.0098
prompt_18	The Design Thinking process. Empathize -> Define -> Ideate -...	0.603	0.965	$0.0092
prompt_19	E-commerce Microservices. API Gateway routes to Product Serv...	0.733	1.000	$0.0108
prompt_2	A password reset flow. User requests reset, System sends Ema...	0.669	1.000	$0.0085
prompt_20	A Kubernetes Cluster Architecture. Control Plane manages Nod...	0.644	1.000	$0.0067
prompt_21	A React Component Tree. App Component holds Header, Main, Fo...	0.800	1.000	$0.0100
prompt_22	The Solar System. Sun at center. Inner Planets (Mercury, Ven...	0.670	0.989	$0.0113
prompt_23	A Biological Trophic Pyramid. Producers (Plants) at base. Pr...	0.724	0.953	$0.0202
prompt_24	Human Nervous System. Central NS (Brain, Spine). Peripheral ...	0.720	1.000	$0.0153
prompt_25	A Biological Taxonomy. Domain Eukarya -> Kingdom Animalia ->...	0.711	1.000	$0.0058
prompt_26	A Corporate Org Chart. CEO leads VP Eng, VP Sales, VP HR. VP...	0.651	0.991	$0.0120

prompt_27	A Decision Tree for Loan Approval. Credit Score > 700? Yes -...	0.634	1.000	$0.0056
prompt_3	An email subscription process. User submits email, System ad...	0.725	1.000	$0.0044
prompt_4	A gravity-fed water filtration system. Rainwater to Gutter, ...	0.817	1.000	$0.0070
prompt_5	A manufacturing assembly line. Raw material enters Conveyor,...	0.672	1.000	$0.0072
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....	0.656	0.972	$0.0048
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	$0.0050

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:** $0.0111
- **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:** $0.0083
- **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:** $0.0059
- **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:** $0.0066
- **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:** $0.0149
- **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:** $0.0098
- **Data:** trajectory.json, metrics.json

---

prompt_18

Initial Prompt:

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

- **Cost:** $0.0092
- **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:** $0.0108
- **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:** $0.0085
- **Data:** trajectory.json, metrics.json

---

prompt_20

Initial Prompt:

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

- **Cost:** $0.0067
- **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:** $0.0100
- **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:** $0.0113
- **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:** $0.0202
- **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:** $0.0153
- **Data:** trajectory.json, metrics.json

---

prompt_25

Initial Prompt:

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

- **Cost:** $0.0058
- **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:** $0.0056
- **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:** $0.0044
- **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:** $0.0070
- **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:** $0.0072
- **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:** $0.0048
- **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:** $0.0050
- **Data:** trajectory.json, metrics.json
