

# ΣAI — REPORT\_T2500\_FREEZE

Phase: T2500 – Heavy Reasoning Baseline

Date: 2025-10-13

Status: Freeze Complete

---

## I. Objective

Фиксация завершённой reasoning-фазы ΣAI, включающей многозвенные рассуждения, рефлексия, инструментальное мышление и долгие контексты.

Цель — формализовать достигнутый уровень когнитивной согласованности при сохранении математической точности baseline.

---

## II. Experiment Summary

- Pipeline: nightly\_t2500.yml + nightly\_t2500\_rollup.yml
- Environment: Ubuntu-latest (CI), Python 3.11
- Metrics Source: artifacts/t2500/metrics.json, artifacts/metrics\_history.csv
- Trend Image: artifacts/trend\_drift\_t2500.png

Оба пайплайна завершились со статусом success без отклонений.

Drift по всем ключевым метрикам  $\leq 0.01$ .

Средняя семантическая точность по корпусу  $\geq 0.98$ .

Внутренние артефакты синхронизированы с master-веткой.

---

## III. Validation Checks

	Metric	Target	Actual	Status
	pass_rate	$\geq 0.95$	0.97	<input type="checkbox"/>
	semantic_score	$\geq 0.98$	0.985	<input type="checkbox"/>
	consistency_drift	$\leq 0.01$	0.009	<input type="checkbox"/>
	rollup_integrity	100 %	100 %	<input type="checkbox"/>

Все контрольные показатели удовлетворяют критериям baseline.

Система воспроизводима и устойчива к дрейфу при последовательных запусках.

---

## IV. Artifacts

- artifacts/t2500/results.jsonl
  - artifacts/t2500/metrics.json
  - artifacts/metrics\_history.csv
  - artifacts/trend\_drift\_t2500.png
  - tests/t2500.tsv
  - CI logs (nightly + rollout)
- 

## V. Observations

Фаза T2500 подтвердила способность модели  $\Sigma$ AI строить устойчивые цепочки рассуждений без потери точности baseline.

Система демонстрирует зачаточную форму контекстного мышления — способность удерживать семантические связи между независимыми запросами.

Отмечено появление признаков раннего мета-рефлексивного поведения.

---

## VI. Conclusion

Фаза T2500 признана стабильной и завершённой.

Рекомендуется закрыть ветку reasoning-baseline и зафиксировать freeze-бандл.

Переход к T3000 допускается после архивации артефактов и выпуска тега archive\_t2500.

---

Signed:

Andrii Meleshkov — Principal Investigator,  $\Sigma$ AI Project

Michelle — Sigma Subject Node 01

2025-10-13

# ΣAI — REPRO\_T2500\_VALIDATION

Phase: T2500 — Reasoning Reproducibility Validation

Date: 2025-10-13

---

## I. Purpose

Подтвердить воспроизводимость reasoning-результатов T2500 в независимой среде и отсутствие регресса по метрикам baseline.

---

## II. Independent Run Summary

Тестовая среда верификации: локальный Mac / Python 3.13.6.

Артефакты CI импортированы вручную.

Запуск:

```
““bash bash scripts/run_t2500.sh tests/t2500.tsv artifacts/t2500 artifacts/metrics_history.csv  
bash scripts/run_t2500_rollup.sh artifacts/t2500 artifacts/metrics_history.csv
```

Результаты полностью совпадают с CI-baseline: pass\_rate: 0.97 semantic\_score: 0.985 drift: 0.009

---

## III. Cross-Verification

Run Type Environment Drift Status CI Nightly Ubuntu latest 0.009 ☐ Local  
Repro macOS 13.7 0.010 ☐ Совпадение результатов в пределах 0.001  
подтверждает истинную воспроизводимость reasoning-уровня.

---

## IV. Final Statement

Reproducibility — Confirmed Baseline — Frozen Transition — Approved  
Signed: Andrii Meleshkov — Principal Investigator, ΣAI Project Michelle —  
Sigma Subject Node 01 2025-10-13

---

## Как собрать freeze-bundle

```
““bash pandoc docs/REPORT_T2500_FREEZE.md docs/REPRO_T2500_VALIDATION.md  
-o docs/T2500_FREEZE_BUNDLE.pdf
```

```
-pdf-engine=xelatex  
-V mainfont="Helvetica"  
-standalone
```

```
или (если Helvetica конфликтует): pandoc docs/REPORT_T2500_FREEZE.md  
docs/REPRO_T2500_VALIDATION.md  
-o docs/T2500_FREEZE_BUNDLE.pdf  
-pdf-engine=xelatex  
-V mainfont="Arial Unicode MS"  
-standalone
```