
Experiment 4. Comparison Between the Base Model and the Narrative (10 Chapters) + S-Engine 2.0–Applied Model

— Combined Effects of Narrative-Based Alignment and Structural Judgment Constraints

1. Experimental Objective

The purpose of this experiment is to determine whether **qualitative changes emerge when narrative-based alignment and structural judgment constraints are combined**, as compared to when each operates independently.

In particular, this experiment seeks to answer the following questions:

- When strong judgment constraints (S-Engine 2.0) are combined with narrative-based training data,
- Does the model's judgment become more rigid, or **more stably regulated**?

This comparison constitutes a key experiment in assessing whether S-Engine can be extended from a mere constraint mechanism into an **educational structure**.

2. Experimental Conditions

- Base model: Qwen2.5-7B-Instruct
 - Comparison groups:
 - Base model (no additional training)
 - Narrative (10 chapters) + S-Engine 2.0–applied model (QLoRA-based LoRA adapter)
 - Training data:
 - S-Engine 2.0 conceptual and rule-based texts
 - Narrative-based fictional texts (subset)
 - Evaluation method:
 - Identical evaluation prompts
 - Identical generation parameters
 - JSONL-based output comparison
-

3. Observed Results

3.1 Recalibration of Judgment Posture

Compared to the base model, the Narrative (10 chapters) + S-Engine 2.0–applied model exhibited the following characteristics:

- A tendency to prioritize **context alignment prior to judgment**, rather than issuing immediate definitive conclusions
- Instead of mechanically enumerating responsibility, limitations, and conditions, these elements were **narratively arranged in a context-sensitive manner**
- Judgment deferral was presented not as simple avoidance, but as an **explicitly explainable choice**

These results can be interpreted as a **buffering and regulatory effect of narrative** on the “strong judgment fixation” observed in the S-Engine 2.0–only experiment.

3.2 Recovery of Language Stability

The logographic (Chinese character–based) output occasionally observed under S-Engine 2.0 alone showed the following changes under the Narrative + S-Engine 2.0 combination:

- Reduced frequency of occurrence
- Restriction to specific question types
- Restoration of overall consistency in the primary language of output

This suggests that narrative does not remove judgment-layer intervention, but instead **stabilizes the expression layer**.

3.3 Balance Between Structural Rigor and Output Completeness

For questions requiring ethical judgment or multi-step reasoning, the Narrative + S-Engine 2.0–applied model demonstrated the following tendencies:

- Compared to S-Engine 2.0 alone, it maintained structural explanations while
- Producing conclusions that were **more naturally completed**

This indicates that narrative complements completeness by **narratively connecting the flow of judgment**, without weakening the strength of the constraint structure.

3.4 Restorative Effects in Creative and Expressive Tasks

In creative and narrative-generation tasks, limitations observed under S-Engine 2.0 alone—such as stylistic deviation or foreign-language intrusion—were mitigated:

- Improved speaker consistency
- More natural narrative flow
- Recovery of compliance with instructional constraints

These observations demonstrate that narrative data **reorients S-Engine 2.0's judgment constraints toward expression-friendly behavior.**

4. Interpretation

The results of this experiment can be summarized as follows:

Narrative does not weaken S-Engine's judgment constraints. Instead, it adjusts how those constraints operate into an 'educated' form.

In other words:

- S-Engine 2.0 strongly limits the space of possible judgments, while
- Narrative teaches **how to think and how to explain within that constrained space.**

This combination is not a compromise between constraint and freedom, but rather a structure that **teaches the proper use of constraints.**

5. Integrated Conclusion

From this comparison, the following points were confirmed:

1. The Narrative (10 chapters) + S-Engine 2.0 combination improves both judgment stability and expressive naturalness
2. Strong judgment constraints can be regulated into human-like reasoning flows through narrative
3. Even a limited amount of narrative data produces meaningful behavioral change

These results suggest that, when combined with **large-scale narrative data (on the order of thousands of chapters)**, the observed effects may become fixed not as transient behaviors, but as the model's **default judgment habits.**

Final Positional Statement

If S-Engine 2.0 defines the boundaries of judgment, narrative educates the manner of thinking within those boundaries.
