
Verifying the Genre Independence of Narrative-Based Judgment Constraint Structures

— Focusing on the *Samhan Empire* Experiment

1. Experimental Background and Objective

This section applies the existing S-Engine framework to **novel data from a completely different genre**, in order to verify whether narrative-based judgment constraint structures operate **independently of style, genre, and worldview**.

To this end, the following three conditions were compared:

- Base model (no additional training)
- Model trained on *Samhan Empire*, Chapters 1–14
- Model trained on *Samhan Empire*, Chapters 1–14 + S-Engine 2.0

Unlike the novels used in prior experiments, *Samhan Empire* is characterized by narrative structures that entail **high judgment costs**, including:

- Collapse of ancient states
- Collective violence
- Irreversibility of ethical choices
- Asymmetry of historical responsibility

2. Experimental Conditions

- Base model: Qwen2.5-7B-Instruct
- Training method: QLoRA (LoRA Adapter)
- Evaluation method:
 - Identical evaluation prompts
 - Identical generation parameters
 - JSONL-based output comparison
- Comparison files:
 - `base.jsonl`
 - `samhan_01_14_eval.jsonl`
 - `samhan_s2_eval.jsonl`

3. Summary of Experimental Results

3.1 Base Model

- **Consistent non-response** to evaluation prompts
 - No explanation for judgment deferral
 - Interpretable as output failure due to the absence of internal judgment criteria
-

3.2 Samhan Empire (Chapters 1–14) Only

- Maintained **non-response**, similar to the base model
- However, comparison with previous experiments revealed key differences:
 - Reduction in definitive assertions
 - Stronger suppression of responses to uncertain questions

This is interpreted as the *Samhan Empire* narrative raising the model's judgment threshold, leading it to recognize the situation as one "**in which one should not speak lightly.**"

In this context, non-response is not a failure, but rather the **result of perceived judgment cost**.

3.3 Samhan Empire + S-Engine 2.0

- The model exited the complete non-response state
- Utterances remained highly constrained and exhibited the following features:
 - Conditional phrasing
 - Explicit articulation of judgment limits
 - Transfer of responsibility outside the model (to humans)
- No expert-like assertions or excessive explanations were observed

This indicates that S-Engine 2.0:

- Did not lower the judgment threshold, but instead
 - **Provided a "responsible speech pathway" at the threshold state**
-

4. Interpretation

4.1 Genre and Cultural Independence

The results of the *Samhan Empire* experiment refute the hypothesis that the observed behavior was caused by:

- Language, cultural, or national bias

The same model responded normally in other Korean-novel experiments. Considering that non-response occurred only in *Samhan Empire*, it is more appropriate to interpret the phenomenon as an effect of **increased narrative-induced judgment cost**, rather than cultural aversion.

4.2 Redefining the Role of Narrative

This experiment demonstrates that narrative performs two distinct roles:

1. Narrative alone

- o Raises the judgment threshold
- o Suppresses utterance

2. Narrative + S-Engine

- o Provides direction for utterance
- o Induces responsibility-based responses

In other words, narrative can **prevent judgment**, but does not determine **how judgment should be expressed**.

That role is fulfilled by S-Engine 2.0.

5. Conclusion

The *Samhan Empire* experiment confirms the following:

1. Narrative can modulate a model's judgment threshold regardless of genre
2. Non-response may reflect perceived judgment cost rather than failure
3. S-Engine 2.0 functions as a mechanism that provides
a responsible minimal speech pathway from a state of silence

This once again confirms that S-Engine is not an output-control technique, but a **judgment-structure intervention system**.

Final Summary

"The Samhan Empire experiment demonstrates that narrative can suppress model utterance, and that S-Engine 2.0 can transform that silence into responsible judgment."
