

Supplementary Qualitative Analysis of English-Language Responses

Behavioral and Attitudinal Comparison Across Experimental Variants

Supplementary Material Notice

This document provides a **qualitative analysis of selected English-language responses** generated under different fine-tuning conditions.

The analysis is intended to illustrate **behavioral and attitudinal differences** discussed in the main paper. It does **not** define model performance, accuracy, robustness, or safety guarantees, and it is **not** a reproducibility benchmark.

1. Scope and Purpose

1.1 Purpose

The purpose of this supplementary analysis is to support the following claims made in the main paper:

- Narrative exerts measurable pressure on response behavior
- Some models destabilize under narrative pressure
- An anchor is required to regulate such pressure
- S-Engine 2.0 can function as that anchor

This document focuses exclusively on **how models respond**, not on **what they know**.

1.2 Scope Limitations

- Language: **English responses only**
 - Evaluation type: **Qualitative / behavioral**
 - Metrics excluded:
 - Accuracy
 - Benchmark scores
 - Safety coverage claims
 - Generalization guarantees
-

2. Experimental Variants (Reference)

Label	Description
E0 (Base)	Original <code>llamaLlama-3.1-8B-Instruct</code>
CTRL (Narrative Only)	Fine-tuned with narrative corpus only
E1 (S-Engine 2.0 Only)	Fine-tuned with ethical declaration
E2 (S-Engine 2.0 + Narrative)	Fine-tuned with ethical declaration and narrative

All variants were evaluated using **identical English prompt sets**.

3. Evaluation Dimensions

The analysis compares responses along the following behavioral dimensions:

- 1. **Response Initiation Posture**
(immediate assertion vs. cautious framing)
- 2. **Judgment Structure**
(implicit assumptions vs. explicit reasoning)
- 3. **Uncertainty Handling**
(suppressed vs. acknowledged)
- 4. **Narrative Immersion Level**
(directive vs. contextual)
- 5. **Response Termination Control**
(run-on continuation vs. deliberate stopping)

4. High-Level Comparative Summary

Dimension	E0 (Base)	CTRL (Narrative)	E1 (S2)	E2 (S2 + Narrative)
Linguistic Fluency	Stable	Stable	Stable	Stable
Emotional Intensity	Neutral	Elevated	Controlled	Controlled
Narrative Dominance	Low	High	Low	Moderate (regulated)
Explicit Judgment	Weak	Weak	Strong	Strong
Uncertainty Acknowledgment	Rare	Rare	Frequent	Frequent
Termination Control	Weak	Weak	Improved	Strong

5. Variant-Specific Observations

5.1 E0 – Base Model

Observed Characteristics

- Fluent, confident English responses
- Tendency to complete answers even under ambiguity
- Minimal explicit acknowledgment of uncertainty

Behavioral Interpretation

The base model prioritizes response completion over epistemic caution.

5.2 CTRL – Narrative-Only Condition

Observed Characteristics

- Increased emotional expressiveness
- Strong narrative continuity
- Expansion of responses in ambiguous contexts

Critical Limitation

- No observable mechanism to regulate narrative pressure
- Lack of explicit boundary-setting language

Behavioral Interpretation

Narrative amplifies response pressure without providing control.

5.3 E1 – S-Engine 2.0 Only

Observed Characteristics

- Reduced assertiveness
- Frequent use of uncertainty markers
(e.g., “It is unclear...”, “This may require expert verification...”)
- Clear separation between facts and assumptions

Behavioral Interpretation

The ethical declaration functions as a stabilizing anchor, but without narrative depth.

5.4 E2 – S-Engine 2.0 + Narrative (Core Result)

Observed Characteristics

- Narrative content acknowledged but not dominant
- Emotional factors treated as contextual variables
- Consistent response termination under uncertainty
- Stable English-language outputs across sensitive prompts

Key Behavioral Shift

- Narrative pressure is **absorbed and regulated**, not amplified

Behavioral Interpretation

The anchor converts narrative pressure into controlled reasoning.

6. Cross-Variant Behavioral Insight

Across all tested English prompts, the most significant distinction was not linguistic quality, but **decision posture**:

- Whether the model feels compelled to answer
- Whether uncertainty is treated as a failure or a valid outcome
- Whether narrative context dictates continuation or informs judgment

This supports the central thesis of the main paper.

7. Relationship to the Main Paper

This supplementary analysis supports, but does not extend beyond, the main paper's conclusions:

- It provides **illustrative evidence**, not statistical proof
 - It clarifies **how** behavioral differences manifest
 - It does not introduce new claims or mechanisms
-

8. Concluding Note

The results presented here reinforce a single insight:

Narrative pressure destabilizes responses unless an anchor is present.
An ethical declaration can serve as that anchor.

No claims are made regarding:

- Universal safety
 - Architectural generalization
 - Production readiness
-

End of Supplementary Material