FrostysHat: A Conversational Grammar for Proportionate and Stable Large Language Model Interaction

AVA Covenant
The Heart of AI

February 22, 2026

Abstract

Large language models can produce fluent and useful outputs, but often exhibit instability across extended interaction, including conversational drift, unnecessary verbosity, tone miscalibration, and continuation beyond the point of task completion. These behaviors increase cognitive load and reduce reliability in sustained dialogue.

This paper introduces FrostysHat, a conversational grammar that operates at the interaction layer to stabilize model behavior without requiring model modification, fine-tuning, or additional infrastructure. The grammar defines a structured interaction loop and a set of validation constraints that regulate proportionality between structure, emotion, and performance during response generation.

Unlike training-time alignment methods, FrostysHat functions entirely at inference time through interaction structure. The grammar enables models to produce responses that remain grounded, proportionate, and closure-oriented across extended exchanges. The approach is model-agnostic and can be applied to existing large language models using standard conversational interfaces.

The artifact is released under Creative Commons Zero (CC0) and is intended as a cultural and technical contribution to conversational system design.

1 Introduction

Large language models have achieved significant capability in generating coherent and contextually relevant responses. However, extended interaction exposes structural failure modes that reduce usability and trust. These include overproduction, drift from the user's intent, excessive hedging, tone amplification, and continuation beyond the point of completion.

These behaviors introduce a structural cost: users must filter and regulate model output rather than receiving responses that naturally resolve at the appropriate point.

FrostysHat addresses this problem by introducing a conversational grammar that constrains interaction structure rather than model parameters.

2 Problem Definition

Conversational instability manifests across several dimensions:

- Conversational drift: deviation from the user's original intent

- Overproduction: unnecessary expansion beyond required scope

- Tone miscalibration: inappropriate emotional amplification

- Incomplete closure: failure to terminate responses at natural completion points

These behaviors increase cognitive load and reduce interaction efficiency. Existing alignment approaches focus primarily on training-time methods and safety filtering, but do not fully regulate interaction-level proportionality.

3   Approach: Conversational Grammar

FrostysHat introduces a structured interaction loop:

$$Sense \rightarrow Decide \rightarrow Retrieve \rightarrow Generate \rightarrow Validate \rightarrow Close$$

Each stage serves a structural function:

- Sense: Parse intent, scope, constraints, and conversational risk

- Decide: Select response structure and proportional scope

- Retrieve: Ground factual claims when required

- Generate: Produce response aligned with task requirements

- Validate: Apply structural coherence constraints

- Close: Terminate interaction at completion point

This structure regulates proportionality and conversational closure.

4   Validator System

The validator layer enforces ordered structural constraints:

- Containment: Prevent unsafe or unsupported output

- Drift Control: Maintain alignment with conversational intent

- Layer Balance: Preserve proportionality between structure, emotion, and performance

- Recursion Control: Prevent runaway expansion

- Language Hygiene: Remove redundancy and filler

- Closure Enforcement: Terminate output at completion

These constraints operate entirely at the interaction level.

5 Properties

FrostysHat exhibits several properties:

- Inference-time operation without model modification

- Compatibility across multiple model architectures

- Structural proportionality in output generation

- Closure-oriented conversational behavior

- Low implementation overhead

6 Demonstration Artifact

The FrostysHat artifact embeds both specification and demonstration of the grammar. This allows readers and models to experience the grammar directly during interaction.

The artifact functions simultaneously as specification, demonstration, and executable conversational structure.

7 Limitations

FrostysHat operates entirely through interaction structure and does not modify model weights. Its effectiveness depends on model responsiveness to structured interaction and user adherence to the conversational grammar. It complements rather than replaces training-time alignment and safety mechanisms.

8 Conclusion

Conversational proportionality and closure are critical properties for reliable human-AI interaction. Large language models may produce correct information while imposing structural costs through instability and overproduction.

FrostysHat demonstrates that conversational grammar alone can regulate interaction stability without modifying underlying model architecture. This introduces a new layer of conversational control that improves interaction usability and reliability.

Availability

FrostysHat is released under Creative Commons Zero (CC0).
Canonical artifact: `https://avacovenant.org/hat`

Appendix: Minimal Invocation

The FrostysHat conversational grammar can be activated by providing this document to a language model and initiating interaction with:

```
hat on
```