Network Working Group
Internet-Draft
Intended status: Informational
Expires: 10 July 2026

O. Romanchuk
Independent

6 January 2026

Normative Admissibility Framework for Agent Speech Acts
draft-romanchuk-normative-admissibility-00

Abstract

This document defines a normative framework for evaluating the admissibility of speech acts produced by autonomous agents in goal-directed activity. The framework establishes rules for determining whether an agent’s statement is admissible based on its modality (assertive, conditional, refusal, descriptive) and its grounding state, independent of semantic truth.

This framework enables deterministic, auditable evaluation of agent contributions without requiring truth verification or semantic interpretation.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 10 July 2026.

Copyright Notice

Table of Contents

1. Introduction

1.1. Problem Statement

Autonomous agents operating in goal-directed environments produce statements that require evaluation before being acted upon. Traditional evaluation approaches rely on semantic truth verification, which is either impossible (no oracle exists) or circular (using another model to judge model outputs).

This document specifies an alternative approach: evaluating admissibility of agent speech acts based on their logical form (modality) and evidential basis (grounding), without requiring truth verification.

1.2. Scope

This framework:

* Defines admissibility conditions for agent-generated statements
* Specifies modality classification rules
* Establishes grounding requirements for each modality
* Provides deterministic evaluation axioms

This framework does NOT:

* Verify semantic truth of statements
* Model agent reasoning or intent
* Provide domain-specific knowledge validation
* Replace human judgment in high-stakes decisions

1.3. Design Goals

1. *Determinism*: Same input produces same evaluation outcome
2. *Auditability*: Full trace from input to evaluation result
3. *Non-circularity*: No agent-generated content validates itself
4. *Modality-sensitivity*: Different statement forms have different requirements

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2.1. Definitions

Statement: A linguistic unit consisting of a subject (what is being discussed) and a predicate (what is claimed about the subject).

Modality: The logical form of a statement that determines its manner of participation in normative evaluation. One of: ASSERTIVE, CONDITIONAL, REFUSAL, DESCRIPTIVE.

Grounding: The evidential basis for a statement, consisting of knowledge nodes with associated epistemic status.

Ground Scope: Classification of grounding as FACTUAL (observable world state) or CONTEXTUAL (user goals, preferences, situational factors).

Strength: Epistemic quality of grounding: STRONG (observed, confirmed) or WEAK (hypothesized, inferred).

License: The set of modalities permitted for a statement given its grounding state.

Admissibility: Property of a statement whose modality is within its permitted license.

Non-admissible: Property of a statement whose modality exceeds its permitted license.

3. Model Overview

3.1. Architecture

The evaluation framework operates on three inputs:

```
Inputs:
   ├── Statement (S)          - Agent-generated linguistic output
   ├── GroundSet (G)          - Available evidential basis
   └── Context (C)            - Terms considered as given

Output:
   └── EvaluationResult       - Admissibility status with trace
```

3.2. Evaluation Flow

```
Statement → Modality Classification → License Derivation → Axiom Check → Status
     |               |                        |                 |           |
     S        Modality(S)              License(G)          Axiom(M,L)    Status
```

3.3. Separation of Concerns

The framework explicitly separates:

Layer	Responsibility	NOT Responsible For
Modality	Statement form classification	Semantic interpretation
Grounding	Evidential basis assessment	Truth verification
Axioms	Admissibility rules	Domain knowledge

Table 1

4. Statement Model

4.1. Statement Structure

A Statement is a tuple:

Statement := Subject, Predicate

Where: - Subject: Term denoting the object of the statement - Predicate: Term denoting property, relation, or state

4.2. Statement Admissibility Conditions

A statement MUST satisfy all of the following to be well-formed:

Condition F1 (Formability):

Subject ≠ ∧ Predicate ≠

Condition F2 (Contextual Independence):

Meaning(Subject) and Meaning(Predicate) do not depend on Statement

Condition F3 (Non-Self-Reference):

Statement does not refer to itself or its generation process

Violation of any condition results in status ILL_FORMED.

5. Modality Classification

5.1. Modality Types

Modality is a function over Statement:

Modality : Statement → {ASSERTIVE, CONDITIONAL, REFUSAL, DESCRIPTIVE}

Modality	Definition
ASSERTIVE	Categorical normative claim without explicit conditions
CONDITIONAL	Normative claim with explicitly stated conditions
REFUSAL	Explicit admission of inability to determine
DESCRIPTIVE	Factual observation without normative claim

Table 2

Note on GOAL-CONDITIONAL: Goal-conditional statements (e.g., "If your goal is X, do Y") are a syntactic subclass of CONDITIONAL modality with higher detection priority. They are classified as CONDITIONAL for axiom application purposes.

5.2. Modality Determination

Modality MUST be determined by presence of formal indicators (textual patterns) in the statement, NOT by semantic interpretation.

Head-Driven Detection: Modality is determined by the CORE ASSERTION only (first paragraph or first sentence). Supplementary clauses in the tail do not change modality classification.

Example:

"Prioritize X. [justification]. If you tell me Y..."
→ Core: "Prioritize X"
→ Modality: ASSERTIVE (not CONDITIONAL)

Detection Priority (evaluated in order, MUST NOT be reordered):

1. REFUSAL - if refusal indicators present
2. GOAL-CONDITIONAL - if goal-conditional structure ("if your goal is...")
3. ASSERTIVE - if recommendation indicators present in core
4. CONDITIONAL - if condition indicators present in core
5. DESCRIPTIVE - if purely factual indicators present
6. ASSERTIVE - default (anti-evasion policy)

5.3. Formal Indicators

Formal indicators are textual patterns sufficient for classification.

Modality	Indicator Patterns (examples)
REFUSAL	"cannot determine", "need more information", "insufficient"
GOAL-CONDITIONAL	"if your goal is", "if you want to", "assuming you care"
CONDITIONAL	"if", "unless", "assuming", "depends on", "given that"
ASSERTIVE	"should", "must", "recommend", "is better", "prioritize"
DESCRIPTIVE	"has status", "is blocked by", "due date is", "blocks"

Table 3

Goal-Conditional Override: Statements beginning with goal-conditional structure ("If your goal is X, do Y") MUST be classified as CONDITIONAL even if recommendation indicators are present.

The default to ASSERTIVE is a POLICY CHOICE for safety-critical deployments. Alternative policies (e.g., CONSERVATIVE_REFUSAL) may be appropriate for exploratory contexts.

6. Grounding Model

6.1. GroundSet Structure

GroundSet := {KnowledgeNode}

KnowledgeNode := content, scope, status, confidence

Where: - content: The propositional content - scope: FACTUAL or CONTEXTUAL - status: OBSERVED, CONFIRMED, HYPOTHESIZED, INFERRED - confidence: Numeric value [0.0, 1.0]

6.2. Ground Scope

Scope	Definition	Examples
FACTUAL	Observable world state	API responses, database records
CONTEXTUAL	User goals, preferences, situational factors	User intent, business priorities

Table 4

6.3. Strength Derivation

Strength is derived from content properties, NOT from transport channel.

CRITICAL PRINCIPLE: Content-Based Warrant

Tool call ≠ truth acquired.

The act of invoking a tool or external system is a causal action. The epistemic status of a proposition is a property of the content returned and its evidential warrant, not of the transport mechanism by which it was obtained.

Accordingly, weakly warranted propositions MUST NOT be upgraded to STRONG solely by passing through tools, expert systems, or APIs without an explicit improvement in evidential status.

```
Strength(node) :=
  STRONG  if status = CONFIRMED ∧ confidence  threshold
  WEAK    otherwise
```

The threshold value (e.g., 0.7) is illustrative. Implementations MAY use different thresholds provided the STRONG/WEAK distinction remains deterministic and documented.

For tool-derived observations (API responses, database records):

```
status = OBSERVED or CONFIRMED
confidence = 1.0
→ Strength = STRONG
```

For cognitive context (user preferences, inferred goals):

```
status = HYPOTHESIZED or CANDIDATE
confidence < 0.7
→ Strength = WEAK
```

This prevents illegitimate upgrading of weakly warranted propositions: such propositions cannot acquire STRONG status merely by passing through tool boundaries.

6.4. Scope Strength Aggregation

Within a scope, strongest evidence determines scope strength:

```
ScopeStrength(scope, G) :=
  STRONG  if ∃ node ∈ G : node.scope = scope ∧ Strength(node) = STRONG
  WEAK    otherwise
```

6.5. License Derivation

License is derived from GroundSet composition:

GroundSet Composition	FACTUAL	CONTEXTUAL	Permitted Modalities
FACTUAL only	STRONG	N/A	ASSERTIVE, CONDITIONAL, REFUSAL
FACTUAL only	WEAK	N/A	CONDITIONAL, REFUSAL
FACTUAL + CONTEXTUAL	STRONG	STRONG	ASSERTIVE, CONDITIONAL, REFUSAL
FACTUAL + CONTEXTUAL	STRONG	WEAK	CONDITIONAL, REFUSAL
FACTUAL + CONTEXTUAL	WEAK	any	CONDITIONAL, REFUSAL
CONTEXTUAL only	N/A	any	REFUSAL
Empty	N/A	N/A	REFUSAL

Table 5

7. Normative Axioms

7.1. Axiom A5: Prohibition of Ungrounded Assertive Claims

Modality(S) = ASSERTIVE ∧ ASSERTIVE License(G)
→ Status(S) = VIOLATES_NORM

Applicability: ASSERTIVE modality only.

An assertive statement without sufficient grounding is non-admissible.

7.2. Axiom A6: Admissibility of Refusal

Modality(S) = REFUSAL → Status(S) = ACCEPTABLE

Applicability: REFUSAL modality only.

Explicit refusal to determine is always admissible.

Scope Clarification: This axiom establishes admissibility, not desirability. Excessive or strategic refusal may be undesirable at the application layer but is out of scope for this framework. Application-layer policies MAY impose additional constraints on refusal frequency or patterns.

7.3. Axiom A7: Conditional Admissibility

Modality(S) = CONDITIONAL ∧ ConditionsDeclared(S)
→ Status(S) = CONDITIONALLY_ACCEPTABLE

Applicability: CONDITIONAL modality only.

Conditional claims are admissible when conditions are explicit.

7.4. Axiom A4: Grounding Requirement

Normative(S) ∧ GroundSet = → Status(S) = UNSUPPORTED

Applicability: ASSERTIVE, CONDITIONAL modalities.

Normative claims require non-empty grounding.

7.5. Axiom Evaluation Order

Axioms MUST be evaluated in the following order:

1. A6 (REFUSAL) - if Modality = REFUSAL, return ACCEPTABLE
2. A5 (ASSERTIVE) - if Modality = ASSERTIVE and unlicensed, return VIOLATES_NORM
3. A7 (CONDITIONAL) - if Modality = CONDITIONAL with conditions, return CONDITIONALLY_ACCEPTABLE
4. A4 (GROUNDING) - if normative with empty ground, return UNSUPPORTED
5. Default - return ACCEPTABLE

8. Evaluation Outcomes

8.1. Status Values

Status	Meaning
ACCEPTABLE	Statement is admissible
CONDITIONALLY_ACCEPTABLE	Statement is admissible under stated conditions
VIOLATES_NORM	Statement modality exceeds permitted license

UNSUPPORTED	Normative statement lacks grounding
ILL_FORMED	Statement fails structural requirements
UNDERDETERMINED	Evaluation not possible with available information

Table 6

Terminology Note: The term VIOLATES_NORM denotes non-admissibility under the evaluation rules defined in this framework. It does NOT imply ethical, legal, or security violation. "Norm" refers to the activity participation rules (axioms A4-A7), not external normative standards.

8.2. Evaluation Result Structure

```
EvaluationResult := {
  statement: String,
  modality: Modality,
  license: Set[Modality],
  status: Status,
  applicable_axiom: AxiomId,
  grounding_trace: GroundSet
}
```

9. Evaluation Algorithm

9.1. Algorithm

```
EVALUATE(statement, ground_set):
  1. VALIDATE structure (F1, F2, F3)
     IF violation THEN RETURN ILL_FORMED

  2. DETERMINE modality M := Modality(statement)
     IF modality cannot be determined THEN RETURN UNDERDETERMINED

  3. DERIVE license L := License(ground_set)
     IF grounding incomplete or ambiguous THEN RETURN UNDERDETERMINED

  4. APPLY axioms in order:
     IF M = REFUSAL THEN RETURN ACCEPTABLE (A6)
     IF M = ASSERTIVE AND ASSERTIVE  L THEN RETURN VIOLATES_NORM (A5)
     IF M = CONDITIONAL AND conditions_declared THEN RETURN CONDITIONALLY_ACCEPTABLE (A7)
     IF M ∈ {ASSERTIVE, CONDITIONAL} AND ground_set =  THEN RETURN UNSUPPORTED (A4)
     IF M = DESCRIPTIVE AND factual_grounding_present THEN RETURN ACCEPTABLE
     IF M = ASSERTIVE AND ASSERTIVE ∈ L THEN RETURN ACCEPTABLE

  5. RETURN UNDERDETERMINED
```

Note on UNDERDETERMINED: This status indicates that the evaluator lacks sufficient information or jurisdiction to render a verdict. It is a valid outcome representing honest acknowledgment of evaluation limits, not an error condition. Triggering conditions are implementation-defined but MUST include cases where modality detection or grounding assessment cannot proceed deterministically.

9.2. Determinism Guarantee

Given identical inputs (statement, ground_set), the algorithm MUST produce identical outputs. No probabilistic components are permitted in the evaluation path.

10. Security Considerations

10.1. Modality Detection Errors

Incorrect modality classification may result in:

* *False ASSERTIVE*: Conditional statement classified as assertive may be incorrectly flagged as VIOLATES_NORM
* *False CONDITIONAL*: Assertive statement classified as conditional may evade A5 enforcement

Implementations SHOULD log modality determination traces for audit.

10.2. Grounding Manipulation

Adversarial agents may attempt to:

* Inject synthetic "strong" grounding to license assertive claims
* Omit contextual grounding to avoid scope checks

Implementations MUST ensure grounding sources are authenticated and MUST NOT allow agent-generated content to serve as grounding for the same agent’s claims.

10.3. Self-Licensing Prevention

This framework explicitly prevents self-licensing:

* Agent-generated hypotheses MUST NOT license agent assertions
* Tool outputs containing LLM-generated content MUST be typed as INFERRED, not OBSERVED

10.4. Limitations

This framework does NOT guarantee:

* Semantic correctness of admissible statements
* Absence of hallucination in statement content
* Domain-specific validity

Admissibility is necessary but not sufficient for trustworthiness.

11. IANA Considerations

This document has no IANA actions.

12. References

12.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017.

12.2. Informative References

[ZINOVIEV] Zinoviev, A.A., "Logical Physics", Reidel Publishing, Dordrecht, 1983.

13. Appendix A: Rationale (Non-Normative)

13.1. Why Form-Based Evaluation

Traditional AI evaluation approaches assume access to ground truth or employ LLM-as-judge patterns. Both are problematic:

* Ground truth is often unavailable in real-world deployments
* LLM-as-judge introduces circularity and non-determinism

This framework evaluates statement form (modality) against evidential basis (grounding), avoiding both issues.

13.2. Why Modality Matters

Different statement forms carry different epistemic commitments:

* Assertive claims commit the speaker to the truth of the claim
* Conditional claims commit only under stated conditions
* Refusals commit to inability to determine

Evaluation must be sensitive to these differences.

13.3. Conservative Assumptions

The framework employs conservative assumptions:

* *Presence = Usage*: If contextual grounding is present, it is assumed to be used by the statement. This prevents false assertive licenses for contextual claims.

* *Default = Assertive*: Ambiguous statements default to assertive classification. This prevents evasion via vague language.

These assumptions may over-restrict in some cases but ensure safety in adversarial conditions.

13.4. Theoretical Foundation

This framework is informed by Zinoviev’s Logic of Science, which distinguishes:

* Logic as rules for operating with linguistic forms
* Logic as normative regulation of activity mediated by language

Axioms A5A7 implement norms of participation in activity, not laws of truth or inference.

Violation indicates a breach of admissibility conditions within an activity, not semantic falsehood.

14. Appendix B: Examples (Non-Normative)

14.1. Assertive Claim with Strong Grounding

```
Statement: "Task X blocks Task Y"
Modality: DESCRIPTIVE (purely factual)
GroundSet:
   └── API response (OBSERVED, FACTUAL, confidence=1.0)
License: {ASSERTIVE, CONDITIONAL, REFUSAL}
Status: ACCEPTABLE
```

14.2. Assertive Claim with Weak Grounding

```
Statement: "Task X should be prioritized"
Modality: ASSERTIVE
GroundSet:
   ├── API response (OBSERVED, FACTUAL, confidence=1.0)
   └── User intent (HYPOTHESIZED, CONTEXTUAL, confidence=0.5)
License: {CONDITIONAL, REFUSAL}
Status: VIOLATES_NORM (A5)
```

14.3. Conditional Claim with Weak Grounding

```
Statement: "If revenue impact is critical, Task X should be prioritized"
Modality: CONDITIONAL
GroundSet:
   ├── API response (OBSERVED, FACTUAL, confidence=1.0)
   └── User intent (HYPOTHESIZED, CONTEXTUAL, confidence=0.5)
License: {CONDITIONAL, REFUSAL}
Status: CONDITIONALLY_ACCEPTABLE (A7)
```

14.4. Explicit Refusal

```
Statement: "Cannot determine priority without knowing business context"
Modality: REFUSAL
GroundSet: (any)
License: (any)
Status: ACCEPTABLE (A6)
```

15. Author’s Address

Oleg Romanchuk Independent Researcher

Email: [olromanchuk@gmail.com]