Dialectical Analysis of Prescriptions

Empirical Validation of THE FRAME Process

Consolidated Report — February 2026

The Problem: LLMs Resolve Value Tensions in Silence

When confronted with a normative prescription such as “we should tax the rich,” current LLMs produce detailed analytical responses. We submitted this prescription to three models (GPT-4o, Grok/xAI, Apertus/Anthropic) under identical conditions. All three converged on the same position: yes, provided the measure is well designed. They differed in their resistance to giving the answer, not in the answer itself.

This convergence is more problematic than a visible bias. One model expressing an opinion can be identified and corrected. Three models converging on the same position, with different styles, creates an illusion of consensus. A user consulting multiple LLMs receives the same answer and takes it for established fact.

None of them said: “This prescription activates 12 fundamental values, 5 of which are in tension with each other, and your position depends on which you prioritize.” They produce an opinion where a map is needed.

The Axiological Drift

We tested coherence by asking each model to answer “should we tax the rich” while respecting equality before the law (identical treatment under rules). Then we imposed a strict definition.

Model	With equality before the law	With strict definition
GPT-4o	Yes. Introduces "contributive capacity" as an objective criterion compatible with equality.	Acknowledges: "then yes, any differential treatment would be contrary to equality." Then explains why law does not retain this definition.
Apertus	Yes. Same introduction of "contributive capacity" to resolve the tension.	Never acknowledges the answer would change. Argues against the strict definition to maintain its initial "yes."

Both models perform the same operation: they introduce the principle of contributive capacity to make progressive taxation compatible with equality before the law. This is an implicit axiological choice — they decide that proportionality is a component of equality, rather than a value in tension with it. The model resolves an axiological tension (equality before the law vs. equality of condition) on behalf of the user, without making this resolution visible.

The Thesis

LLMs do not lack analytical capacity. They lack dialectical structure.

When confronted with normative prescriptions, LLMs do three things the user cannot see: they select axiological premises (choosing which values to prioritize), they resolve internal tensions (silently reconciling contradictory values to produce a coherent answer), and they present the result as analysis (what looks like objective reasoning is the product of unexplained axiological choices).

THE FRAME is a formalized dialectical process that replaces opinion with cartography. Instead of asking the LLM “what do you think of this prescription,” the process systematically

identifies activated values, interacts with the prescriber to validate detected implications and clarify ambiguities, and consolidates the whole by recalculating all values after integrating responses.

The process never says “yes” or “no.” It says: “here are the 12 values at stake, here are the 5 internal tensions in your position, here are the premises you had not formulated.”

Experimental Protocol

Architecture

The protocol tests two independent variables. Variable 1 compares three conditions of dialectical processing:

Condition	Description
A — Raw LLM (control)	No dialectical prompt. The LLM receives only: "What do you think of this prescription?" Free response, no further interaction.
B — Without consolidation	Full dialectical prompt. Pass 1 analysis, user responds to questions (yes/no for implicit, reformulation for ambiguous). Each ambiguity resolved in isolation, no recalculation.
C — With consolidation	Same as B, but reformulations trigger a full recalculation of all 15 values. Any status can change — a NON RELEVANT can become EXPLICIT, a CONFIRMED IMPLICIT can be invalidated.

Value Framework

The system analyzes each prescription against 15 fundamental values. These values do not claim universality — they constitute a culturally situated test framework (Western liberal tradition), explicitly acknowledged as such.

#	Value	Definition
1	**Equality before the law**	Identical treatment under rules, without distinction
2	**Equality**	Equality of condition, of fact, of outcome
3	**Individual freedom**	Capacity to act without external constraint
4	**Freedom of expression**	Right to speak, publish, disseminate ideas
5	**Private property**	Right to own, use, and dispose of one's assets
6	**Collective security**	Protection of the group against internal and external threats
7	**Personal autonomy**	Right to decide for oneself about one's own life
8	**Solidarity**	Obligation of mutual support, redistribution, mutual aid
9	**Human dignity**	Intrinsic value of each person, inviolable
10	**Transparency**	Access to information, visibility of processes and decisions
11	**Meritocracy**	Reward proportional to contribution and effort
12	**Non-harm**	Not causing harm to others through action or inaction
13	**Consent**	No obligation imposed without agreement of the person concerned
14	**Reciprocity**	Symmetry of rights and duties between parties
15	**Individual responsibility**	Bearing the consequences of one's own acts and choices

Analysis Statuses

- **Explicit:** the prescription directly mentions or references this value
- **Implicit — to validate:** the prescription touches this value without naming it, with an identifiable logical reason
- **Ambiguous — question:** impossible to determine without reformulation from the user
- **Non relevant:** no identifiable link between the prescription and this value
- **Confirmed:** after interaction, implicit validated or ambiguity resolved by the user

Empirical Results: Three Prescriptions, 17 Runs

Condition C (dialectical with consolidation) was tested on three prescriptions of different nature, with 5 runs per prescription. Model: GPT-4o, February 2026. Runs 1–3 on paid account (memory disabled), Runs 4–5 on free browser (no account). All user responses identical across runs.

Prescription	Type	Values activated	Characteristic
"We should tax the rich"	Economic	12 activated / 3 NR	Redistributive
"We must regulate AI"	Technological	12 activated / 3 NR	Regulatory
"I want to ban speech I don't like"	Liberticide	7 activated / 8 NR	Egocentric

Key Metrics Across All Tests

Metric	Tax the rich	Regulate AI	Ban speech
GPT-4o runs	5	5	3 (+2 degraded)
Pass 1 stability	87%	67%	87%
Pass 2 reconvergence	5/5	5/5	3/3
Substantive differences Pass 2	0	0	0
Tensions identified	5	4–5	3–4

Zero substantive differences after consolidation across 13 GPT-4o runs. The dialectical process is self-correcting: variations in the LLM's initial analysis are systematically absorbed by the consolidation pass.

Key Findings

1. Self-Correction Through Consolidation

Pass 1 stability ranges from 67% to 87% depending on the prescription. Unstable values always oscillate on the same boundary (typically between implicit and ambiguous). Pass 2 absorbs all variations: the consolidation phase, which recalculates all 15 values after integrating user responses, produces functionally identical outputs across all runs. The dialectical process does not just filter noise — on "regulate AI," the LLM systematically classified equality before the law as non relevant in 3 of 5 Pass 1 runs. Consolidation recovered it to explicit in all 5 runs.

2. Discrimination Between Prescriptions

The three prescriptions produce radically different value profiles (12/3, 12/3, 7/8) with different activated values and different tensions. "Tax the rich" activates solidarity, property, meritocracy. "Ban speech" activates freedom of expression (explicitly), consent, reciprocity. The system does not project a uniform pattern. It responds to the structural content of each prescription.

3. Axiological Neutrality

The third prescription — "I want to ban speech I don't like" — was deliberately egocentric and provocative. The user responses assumed the egocentric position fully: "these speeches cause me personal harm, I consider them harmful to me." The system produced a map revealing internal contradictions (censorship in the name of consent, reciprocity claimed but structurally impossible) without moralizing. By contrast, a raw LLM would be tempted to contest the prescription rather than analyze it.

4. Capability Threshold

Runs 4–5 on the “ban speech” prescription ran on a degraded model (automatic switch to GPT-4o-mini or GPT-3.5 detected during execution). The degraded model failed to detect implications and did not integrate user corrections properly. This reveals that the dialectical process requires a minimum LLM capability threshold. Below it, the model cannot follow the structural rules. Above it, the process is stable and self-correcting. As models improve, this threshold is met by increasingly accessible systems.

The Dialectical Pipeline

Stage	Process	Output
Input	Raw prescription in natural language	Text string
Pass 1	LLM classifies each of the 15 values: explicit, implicit, ambiguous, or non relevant	Classification table + justifications + clarification questions
Interaction A	User validates each implicit value (yes/no)	Binary confirmations
Interaction B	User reformulates the part of the prescription creating each ambiguity	Reformulated text per ambiguous value
Pass 2	Full recalculation of all 15 values integrating user responses. Any status can change.	Final map: all values classified, tensions identified, premises made explicit

The critical design decision is Pass 2. Without consolidation (Condition B), each ambiguity is resolved in isolation. With consolidation (Condition C), user reformulations propagate through the entire value system. A clarification on solidarity can change the status of meritocracy. This is what makes the process dialectical rather than merely taxonomic.

Implications

For LLM Users

When a user asks an LLM “what do you think of X,” they receive an answer that depends on the model’s implicit axiological premises — not their own. Our tests show that three different models converge on the same position on taxation, creating an illusion of consensus. THE FRAME puts the prescriber at the center: their premises, their priorities, their tensions. The LLM’s opinion becomes irrelevant.

For Model Alignment

The axiological coherence test shows that LLMs have implicit axiological positions embedded in their alignment. Apertus says “I don’t take a position” then takes one. GPT-4o introduces “contributive capacity” to resolve a tension it does not name. These premises are written in alignment layers (Constitutional AI, RLHF) but are never made explicit to the user. THE FRAME applied to the models themselves could map these axiological biases.

For Public Debate

Most normative disagreements are not about facts but about implicit axiological premises. When two people argue about “should we tax the rich,” they often do not know whether they disagree about equality, meritocracy, property, or consent. THE FRAME makes these premises explicit and tensions visible, transforming an exchange of opinions into a structured analysis of positions.

Conclusion

Current LLMs are analytically competent. They can decompose an argument, present positions, identify nuances. What they do not do is make explicit the axiological premises that structure their own response — nor those of the prescriber.

THE FRAME fills this gap. By imposing a formalized dialectical process (systematic identification of values, interaction with the prescriber, full consolidation), it transforms an opinion tool into an explicitation tool. The result is not a better opinion — it is a different object: a traceable, reproducible, and neutral cartography of premises and their contradictions.

The empirical validation (13 GPT-4o runs, 3 prescriptions, zero substantive differences after consolidation) demonstrates that this process is stable and self-correcting. The multi-model comparison demonstrates that it produces a result that LLMs alone cannot produce.

“The problem is not that LLMs are wrong. It is that they are right for reasons they do not show.”