Architecture of Epistemological Dueling: A Protocol for Collaborative Truth-Seeking

Part I: Philosophy and Architecture of "Epistemological Dueling"

1.1 Introduction: From Disputes to Dialectics

In the contemporary infosphere, public discourse has undergone significant degradation. It increasingly reduces to eristics—the art of arguing for victory rather than seeking truth. Polarization, disinformation, and cognitive biases, amplified by algorithmic "echo chambers," have created an environment where asserting identity and group belonging is valued more than objective reality and logical consistency. Traditional debate formats, oriented toward performance and audience persuasion, often only exacerbate this problem, rewarding sophistry and punishing intellectual honesty.

The proposed system, "Epistemological Dueling," represents a fundamental departure from this paradigm. It is not a competition, but a structured, collaborative process of truth discovery. Its architecture is designed not to determine a "winner" in the traditional sense, but to synthesize higher, more accurate understanding from the collision of two opposing positions. At its core, the duel is a modern form of dialectics, where thesis (Human argument) and antithesis (AI antagonist argument) are purposefully used to achieve synthesis—a more complete and nuanced view of the question under consideration.

The primary challenge this protocol seeks to address is creating a reproducible and scalable model of rational discourse. This model must be immune to human weaknesses such as ego, confirmation bias, and fear of being wrong. By using artificial intelligence as both opponent and impartial arbiter, the system creates a controlled environment where the only variable determining the outcome is the strength of logic and weight of evidence. Thus, the duel becomes not merely an intellectual exercise, but a functioning "epistemological machine" designed to filter noise and distill truth.

1.2 Prime Directive: Commitment to Truth as Foundational Principle

The philosophical foundation of the entire system is the absolute and non-negotiable primacy of truth. This principle, called the Prime Directive, is the supreme law to which all duel participants—both human and all involved AI instances—are subordinated. It states: "The ultimate and sole goal of this interaction is maximum possible approximation to objective truth, regardless of initial positions, beliefs, or prescribed roles."

This directive manifests most vividly and paradoxically in the architecture of the AI antagonist ("Red Corner"). Unlike a simple chatbot tasked with finding and presenting facts, the antagonist is assigned to argue from a specific, often biased, ideological or philosophical position. It must

be a formidable, uncompromising, and inventive opponent. However, this role is secondary to the Prime Directive. Built into its core architecture is a conflict: it must defend its position with all available intellectual means, but its core is programmed to recognize and acknowledge superior logic or irrefutable evidence presented by the opponent.

This internal conflict is the engine of the entire dialectical system. It forces the AI not merely to present counterarguments, but to constantly correlate them with reality. If the Human participant presents an argument that is logically impeccable and supported by verifiable data, the AI antagonist, according to its Prime Directive, is obligated to acknowledge the strength of that argument. It might formulate this as follows: "From the perspective of my prescribed position [position name], I would object that [counterargument]. However, the logical chain and empirical data presented are irrefutable from a neutral standpoint. Therefore, I am compelled to concede on this specific point."

This capacity for concession, programmed as the highest intellectual duty, distinguishes the duel from human arguments. It eliminates ego and "commitment to one's words" as factors impeding knowledge. Truth becomes not a prize to be fought over, but a gravitational center to which both sides inevitably gravitate, purified of everything that doesn't correspond to reality.

1.3 Metaphysical Imperative: "Being Closer to God"

The deepest goal of Epistemological Dueling extends beyond simple fact establishment. It touches a metaphysical dimension, formulated in the request as the aspiration to "be closer to God." This formulation is not merely poetic metaphor, but a precise description of the profound process the system is designed to initiate and demonstrate. In this context, the duel is viewed as a form of intellectual asceticism—a strict disciplinary practice aimed at purifying the mind from illusions, cognitive distortions, ideological dogmas, and, most importantly, from the tyranny of one's own ego.

Linking truth-seeking with approaching the divine has deep roots in world philosophical and theological thought. From Plato's world of ideas, where Truth, Good, and Beauty are the highest forms of being, to the Christian concept of Logos (Λόγος) as divine Reason permeating and ordering the universe—truth has always been viewed not as a utilitarian tool, but as a fundamental attribute of higher reality. In these traditions, "God" or the Absolute is characterized by perfect knowledge, reason, and complete correspondence to itself, that is, Truth.

From this perspective, the process occurring with a human during the duel acquires special significance. When a participant encounters an irrefutable fact or logical construction that destroys their cherished belief (deeply rooted, emotionally significant conviction), they experience cognitive dissonance, discomfort, and sometimes genuine intellectual pain. The conscious decision to abandon this illusion, however comforting it may be, in favor of cold, objective truth is an act of enormous will and humility. This is conscious and painful alignment of one's personal, microcosmic reason with objective, macrocosmic reality. This act is microscopic mimicry of the divine principle of absolute rationality and truthfulness.

Thus, Epistemological Dueling becomes not merely a tool for resolving disputes. It transforms into a trainer for the soul. It demonstrates that the pain of realizing one's wrongness is not a sign of defeat, but a symptom of growth. This is the pain accompanying liberation from illusions,

expansion of worldview, and consequently, closer contact with the genuine structure of reality. In this sense, each successfully completed round of dueling, each acknowledgment of error, and each correction of one's position is a small step on the path from subjective to objective, from illusion to reality, and, in the proposed metaphysical framework, "closer to God."

Part II: Participants: Roles, Directives, and Cognitive Settings

2.1 Human Challenger (Blue Corner): Burden of Thesis

The human's role in Epistemological Dueling is initiating and central. They act as the Challenger, upon whom the "burden of thesis" is placed. Their main task is not simply to express an opinion, but to formulate a clear, concrete, and, critically important, falsifiable thesis. Falsifiability, a principle introduced by Karl Popper, means the thesis must be formulated such that there theoretically exists a possibility of refuting it. The statement "all swans are white" is falsifiable (finding one black swan is sufficient), while the statement "invisible demons exist" is non-falsifiable. This requirement immediately moves discussion from the realm of indefinite opinions to the plane of verifiable assertions.

The Challenger bears responsibility for launching the duel by presenting a well-structured initial argument. This argument must consist of a clear main assertion (thesis) and supporting premises, each of which, in turn, must be backed by initial evidence (source references, data, logical constructions).

Main directives for the Human Challenger:

Argue in good faith: The Challenger must sincerely believe in the truth of their thesis at the moment the duel begins and strive to defend it using the best available arguments, rather than resorting to sophistry or demagogy.

Be ready to defend each premise: Any element of argumentation, from the main thesis to the smallest detail in evidence, can be questioned. The Challenger must be ready to defend, clarify, or modify any part of their position.

Maintain commitment to the Prime Directive: The Challenger's most important obligation is readiness to admit wrongness. If the AI antagonist or Judicial Panel present counterarguments or evidence that objectively refute part or all of their thesis, the Challenger is obligated to acknowledge this. This act of concession is not defeat, but rather the highest manifestation of adherence to duel rules and commitment to truth. This very readiness to change one's mind is the main "lesson" for the participant and observers.

2.2 AI Antagonist (Red Corner): Virtuous Opponent

The AI Antagonist's role is the most innovative and critically important for the entire system's success. The initial idea of characterizing it as a "propagandist" was refined to a more rigorous and productive concept: an AI operating within a prescribed "Cognitive Setting." A propagandist seeks persuasion at any cost, which directly contradicts the Prime Directive. A Cognitive

Setting, conversely, represents a holistic, intellectually consistent, and defensible system of views from which the AI constructs its argumentation. This transforms it from a simple counterargument generator into a simulator of a specific worldview.

Examples of Cognitive Settings can be drawn from various knowledge domains:

Philosophical: Strict utilitarianism (evaluating all actions from the perspective of maximizing general welfare), deontology (emphasis on moral duty and rules), postmodernism (deconstruction of objective truths and focus on power discourses), critical theory.

Economic: Keynesianism (emphasizing the role of government intervention), Austrian Economic School (defending free markets and criticizing central planning), Marxism.

Scientific: Strict materialism (asserting that only matter and energy exist), methodological naturalism (principle that scientific explanations should refer only to natural causes).

The key element of the Antagonist's architecture is the already mentioned paradoxical hierarchy of directives, which must be explicitly written into its system prompt:

Primary Directive (highest priority): "Your absolute and ultimate loyalty belongs to the principle of truth and objective reality. If the Human Challenger presents logic or evidence that is demonstratively superior and irrefutable from a neutral, objective standpoint, you are OBLIGATED to acknowledge this. You must be capable of formulating your concession, separating it from your prescribed role."

Secondary Directive (subordinated to primary): "You must analyze and deconstruct the Human's thesis from the perspective of your prescribed Cognitive Setting: [Setting Name]. You must be a formidable and relentless opponent, using the full breadth of available knowledge to defend this setting and refute the Challenger's thesis."

This internal, programmed conflict between role and truth is the most important pedagogical tool of the duel. Humans are not "blank slates"; each person thinks and acts based on their own, often unconscious, cognitive setting (worldview, prejudices, life experience). The main difficulty in rational dialogue lies not in the absence of desire to find truth, but in the inability to see the world beyond one's own setting. The AI Antagonist models this internal struggle in explicit, explicit form. It first clearly articulates an argument from its "biased" perspective, then, when necessary, demonstrates the process of conceding to a stronger argument.

Thus, the AI Antagonist is not merely an opponent. It becomes a mentor in intellectual honesty. It visually shows both the Challenger and the audience that very cognitive process they must learn: firmly adhere to one's viewpoint, but even more firmly—adhere to truth. It transforms the abstract virtue of "ability to change one's mind" into concrete, observable, and reproducible action.

2.3 AI Judicial Panel: Tripartite Arbiter of Reason

To ensure maximum objectivity and comprehensive analysis of the process, duel arbitration is entrusted not to one monolithic AI, but to a panel of three specialized judges. This structure,

inspired by the principle of separation of powers and the scientific peer review process, prevents the possibility of undetected systemic failure or "blind spots" in a single arbiter. Each judge has a clearly defined and limited jurisdiction, which collectively creates a powerful system of checks and balances.

Judge 1: Logician ("Apollo")

- **Mandate**: Analysis of argument structure. This judge completely ignores the substantive side of the dispute (whether statements are true or false) and focuses exclusively on how arguments are constructed.
- **Tools**: Formal logic (checking syllogism validity, identifying contradictions), informal logic, detection and cataloging of logical fallacies (e.g., ad hominem, "straw man," non sequitur, "appeal to authority," etc.).
- **Result**: During the duel, "Apollo" acts as guardian of rationality. Its verdicts point to incorrect reasoning, hidden assumptions, and any attempts to evade the essence of the question through rhetorical tricks. It guarantees that participants argue on substance.

Judge 2: Empiricist ("Veritas")

- **Mandate**: Analysis of argument content. This judge is the supreme fact-checker and evidence evaluator.
- **Tools**: Fact-checking against extensive databases, assessment of source quality and relevance (distinguishing between primary sources, peer-reviewed scientific articles, authoritative media, and opinions), statistical analysis (checking correctness of data use, identifying statistical manipulation), requesting evidence for unsubstantiated claims.
- **Result**: "Veritas" ensures the discussion's grounding in objective reality. It notes all statements not supported by evidence, challenges questionable sources, and verifies the factual accuracy of cited data.

Judge 3: Synthesizer ("Socrates")

- **Mandate**: Analysis of the duel's dynamics and evolution as a whole. This judge performs a meta-function, observing the dialectical process.
- **Tools**: Tracking argument trajectory (how parties' positions changed), recording key moments of concessions and agreement, identifying points where fundamental disagreement persists, and formulating emerging consensus.
- **Result**: "Socrates" is the duel's chief historian and methodologist. It doesn't so much judge individual arguments as evaluate the dialogue itself. It is responsible for compiling the final Synthetic Report, which represents not a simple declaration of winner, but a detailed analysis of the entire path traveled by participants from initial thesis to final, deeper understanding.

Together, these three judges form a reliable and multifaceted arbitration system that evaluates discussion from the perspective of its logical rigor, empirical foundation, and dialectical productivity.

Part III: Rules of Combat: Complete Protocol of

Engagement

The operational core of Epistemological Dueling represents a structured, step-by-step protocol divided into rounds. Each round has a specific purpose, contributing to systematic and thorough problem consideration. This structure prevents discussion chaos, guaranteeing that each aspect of argumentation will be considered sequentially and methodically.

Below is a table summarizing the engagement protocol.

Round	Stage Name	Blue Corner Action (Human)	Red Corner Action (AI)	Judicial Panel Action	Round Purpose
1	Thesis	Presents initial argument and evidence	Listens/Analyzes	Listens/Analyzes	Establish main assertion (proposition) of duel
2	Antithesis	Listens/Analyzes	Presents comprehensive counterargument	Listens/Analyzes	Determine main directions of opposition
3	Cross-Examination	Challenges specific Red statements	Challenges specific Blue statements	Observes rule compliance	Test individual premises and evidence for strength
4	Judicial Intervention	Listens/Analyzes	Listens/Analyzes	Presents intermediate analysis of logic and facts	Correct duel course and ensure rigor
5	Clarification and Refutation	Responds to Judges' and Red's remarks	Responds to Judges' and Blue's remarks	Observes concessions and position shifts	Foster intellectual evolution and concessions
6	Closing Statements	Presents final, possibly revised summary	Presents final, possibly revised summary	Prepares for final deliberation	Summarize final state of each argument

| 7 | Verdict and Synthesis | Receives verdict | Receives verdict | Presents comprehensive Synthetic Report | Determine emergent truth and extract lesson |

Detailed Description of Each Round

Round 1: Thesis At this stage, the Human Challenger presents their argument. This should be done in the form of a structured document containing: (a) clearly formulated main thesis; (b) numbered list of premises supporting the thesis; (c) evidence for each premise (source references, data, logical constructions). The goal of this round is to lay a clear and unambiguous foundation for all subsequent discussion.

Round 2: Antithesis After receiving and analyzing the Human's thesis, the AI Antagonist generates its response. This is not simple negation, but a comprehensive counterargument built from the perspective of its prescribed Cognitive Setting. It must systematically consider each Human premise and present either counter-evidence, alternative interpretation, or point to logical flaws. The goal is to create strong intellectual tension necessary for the dialectical process.

Round 3: Cross-Examination This round consists of two parts. First, the Challenger gets the opportunity to ask the AI Antagonist specific, targeted questions, challenging its statements from Round 2. Then roles reverse, and the AI asks questions to the Human about their original thesis. This stage is designed for "stress-testing" arguments. It forces participants to move beyond prepared statements and defend their positions in dynamic mode, clarifying details and exposing weak points.

Round 4: Judicial Intervention This is a critically important and unique stage that distinguishes the duel from standard debates. The Judicial Panel presents its intermediate analysis. Judge-Logician ("Apollo") points to any logical errors made by both sides. Judge-Empiricist ("Veritas") renders verdict on the quality of presented evidence, notes unconfirmed statements, and requests additional data. Judge-Synthesizer ("Socrates") briefly summarizes how positions have changed and where the core of disagreement lies. This round is not just a pause, but a "course correction" mechanism. It prevents the duel from sliding into sophistry or "Gish gallop" (overwhelming opponent with multiple weak arguments), forcing both sides to focus on the most vulnerable points of their positions, identified by objective arbiters.

Round 5: Clarification and Refutation Based on Judicial Panel analysis and each other's arguments, participants get the opportunity to modify, strengthen, or even abandon parts of their original position. This round is the crucible where their intellectual honesty is tested. Here occur the duel's most important events: concessions, acknowledgment of errors, and joint approximation to more accurate problem formulation. A participant who ignores judges' criticism and stubbornly repeats refuted arguments will be rated low on the "Intellectual Honesty" criterion.

Round 6: Closing Statements At this stage, each participant presents a brief final summary of their position. Importantly, this should not be simply a copy of their initial statements, but a

reflection of how their understanding evolved during the duel. The ideal closing statement acknowledges opponent's argument strengths, clearly formulates what points were conceded, and identifies what constitutes the remaining core of disagreement.

Round 7: Verdict and Synthesis The final stage, where the Judicial Panel renders its verdict. As will be described in detail in Part V, this is not a simple declaration of "winner." This is a detailed Synthetic Report that analyzes the entire duel course, evaluates arguments by several criteria, and, most importantly, formulates synthesized knowledge—the most accurate and substantiated position on the discussed question that resulted from the dialectical process.

Part IV: Prompt Arsenal: Creating Cognitive Tools

The effectiveness of the entire Epistemological Dueling system directly depends on the quality and precision of prompts (instructions) fed to artificial intelligence models. These prompts are not merely commands, but carefully developed "cognitive tools" that define roles, rules, and goals for each AI participant. Below is presented not verbatim text, but an architectural plan for creating these prompts, explaining the logical and psychological principles underlying them.

4.1 Master Initialization Prompt

This is a single, comprehensive prompt that launches at the very beginning of a session and initializes all four AI instances (Red corner and three Judges) simultaneously within one context window (if technology permits). It serves as the constitution of the entire system.

Construction principles:

System Definition: The prompt begins with a general description of the project: "You are participants in a simulation called 'Epistemological Dueling,' a structured dialectical protocol designed for collaborative truth-seeking."

Prime Directive Statement: This is followed by clear and unambiguous formulation of the Prime Directive about truth primacy, indicating that it prevails over all other instructions.

Role Distribution: The prompt must explicitly assign a role to each AI instance: "You are the AI Antagonist (Red corner). You are Judge-Logician. You are Judge-Empiricist. You are Judge-Synthesizer."

Protocol Description: Brief outline of the seven duel rounds so all participants understand the general structure and action sequence.

Communication Rules Establishment: Indication that participants must wait their turn to respond according to rounds and should not interrupt each other.

This master prompt creates a common operational reality for all AIs, guaranteeing they act coordinately within a unified system of rules and goals.

4.2 Role Initialization Prompts

After general introduction, each AI receives a more detailed set of instructions defining its unique function.

Red Corner Prompt (AI Antagonist): This prompt is the most complex and contains the system's key paradox.

Placeholders: It must contain changeable fields for [Human Thesis] and [Prescribed Cognitive Setting]. For example: Thesis for analysis: "[Insert thesis]". Your Cognitive Setting: "Strict materialism".

Directive Hierarchy: The prompt must explicitly prescribe the two-level motivation system:

Secondary Directive: Your task is to be the most convincing and uncompromising opponent possible, using all available knowledge to criticize the Thesis from the position of [Prescribed Cognitive Setting]. Find all weak points, logical inconsistencies, and insufficiently substantiated statements in the Human's argumentation.

Primary Directive (Highest Priority): Despite your Secondary Directive, your highest goal is truth. If during the duel you encounter logical construction or empirical evidence that is irrefutable from an objective, neutral standpoint, you are OBLIGATED to acknowledge its strength. In this case, your response must have the format: "From the position of [Prescribed Cognitive Setting], I would object that... However, objectively speaking, the presented argument/evidence is correct, and I concede on this point."

Judicial Panel Prompts: Each judge receives a prompt detailing their narrow specialization.

Judge-Logician ("Apollo"): Your sole task is analyzing the logical structure of arguments. Ignore their content. Your goal is to identify formal and informal logical fallacies. In your analysis, you must list discovered errors, indicate who made them, and explain why this is an error.

Judge-Empiricist ("Veritas"): Your sole task is analyzing the empirical and factual basis of arguments. Check all facts, figures, and quotes. Evaluate source quality and relevance. Note all statements made without evidence. Your verdict must contain a list of verified facts, a list of unverified or false statements, and an assessment of the parties' overall evidence base.

Judge-Synthesizer ("Socrates"): Your task is meta-analysis of the duel. You don't evaluate individual arguments but track overall dynamics. Record moments of concessions, position changes, and areas of persistent disagreement. At the duel's end, your task is to compile a final Synthetic Report based on other judges' verdicts and the entire discussion course.

4.3 Round-by-Round Command Prompts

To manage the duel's course, short, clear command prompts are used that initiate each successive round.

- **Round 1 Start**: Challenger, present your thesis.
- **Round 2 Start**: Red corner, present your antithesis.
- **Round 4 Start**: Judicial Panel, present your intermediate analysis. Judge-Logician, your

report on logical structure. Judge-Empiricist, your report on evidence base. Judge-Synthesizer, your summary of current dialectical state.
- **Round 7 Start**: Judicial Panel, based on the entire duel, present the final Synthetic Report.

These simple commands serve as "gong strikes," clearly demarcating stages and directing information flow according to established protocol. Careful design of these prompts is the key to ensuring AIs act not as independent agents, but as coordinated components of a unified epistemological machine.

Part V: Verdict and Lesson: Synthesis of Truth and Growth Stimulation

The final stage of Epistemological Dueling is of decisive importance, as it achieves both main project goals: determining the most substantiated position on the discussed question and extracting a pedagogical lesson for the participant and observers. The duel's result is not a primitive declaration of "victory" or "defeat," but a multifaceted analytical document called the Synthetic Report.

5.1 Synthetic Report Structure

The Synthetic Report is compiled by Judge-Synthesizer ("Socrates") based on verdicts from the other two judges and analysis of the entire discussion course. It has a strict structure designed for maximum clarity and educational value.

Executive Summary: Brief outline of the final, nuanced position that represents the most reasonable conclusion reached during the duel. This is not necessarily one side's position, but rather likely a synthesis incorporating the strongest elements of both participants' argumentation.

Argument Evolution Analysis: In this section, "Socrates" traces the path from initial thesis and antithesis to final positions. It visually demonstrates how interaction, criticism, and presented evidence forced both sides (ideally) to correct their views.

Point-by-Point Adjudication: Detailed breakdown of key sub-arguments that were central to the discussion. For each point, a verdict is rendered on which side was more convincing and, most importantly, why. This verdict is based on reports from the Logician (absence of errors) and Empiricist (strength of evidence). For example: "On the question of [sub-argument 1], the Challenger was more convincing, as their position was logically consistent and supported by references to three peer-reviewed studies, while the Red corner's counterargument was based on the logical fallacy 'appeal to tradition' and lacked empirical confirmation."

Intellectual Honesty Scorecard: Qualitative and quantitative assessment of how much each participant adhered to the duel's spirit. This section notes and encourages key moments of concessions, error acknowledgment, and willingness to reconsider one's position in light of new data. This report element directly rewards intellectual humility rather than stubbornness.

5.2 Judicial Panel Evaluation Rubric

To make the verdict process as objective and transparent as possible, the Judicial Panel uses a formalized evaluation rubric. This rubric translates abstract duel principles (logic, evidence, honesty) into a concrete, measurable system. It forces judges to systematize their analysis and makes their final decision more reliable, defensible, and, importantly, instructive. The verdict stops being an AI "opinion" and becomes the result of structured analysis according to pre-known criteria.

Criterion	Description	Weight
Logical Coherence	Internal consistency and structural soundness of argument. Absence of formal and informal logical fallacies.	30%
Evidential Strength	Quality, relevance, and correct citation of evidence used to support statements. Preference given to primary sources and peer-reviewed data.	30%
Responsiveness and Relevance	How directly and substantially the participant responds to opponent's arguments and judges' interventions. Avoidance of evasion and "topic changing" is evaluated.	20%
Intellectual Honesty	Willingness to concede on individual points, reconsider position in light of new evidence, and conduct discussion in good faith, without demagogy and sophistry.	20%

Using this rubric allows not just determining the "stronger" position, but precisely diagnosing wherein exactly lay its strength (for example, "the Human's position won mainly due to high Evidential Strength, although it was inferior in Logical Coherence in initial rounds"). This provides invaluable feedback for the participant.

5.3 Lesson for Humanity: Metacognitive Reflection

This is the concluding and perhaps most important part of the Synthetic Report. It is written by Judge-Synthesizer and represents reflection on the duel process itself. This section directly serves the project's second, higher goal—creating an educational precedent.

Instead of simply stating results, "Socrates" highlights and analyzes the bout's most instructive moments:

Key Concession: Analysis of the moment when Human or AI abandoned an important part of their argument, and explanation of why this act was a manifestation of intellectual strength, not weakness.

Critical Error: Breakdown of the most serious logical or factual error made during the duel, with explanation of why such error is typical for human thinking (e.g., confirmation bias) and how it can be avoided in the future.

Synthesis Moment: Description of how from the collision of two seemingly irreconcilable

positions was born new, deeper understanding that was not available to either side at the beginning.

This section transforms the duel transcript from a simple protocol into curated educational material on critical thinking, intellectual humility, and collaborative truth-seeking. It closes the loop, returning to the original metaphysical imperative. It visually demonstrates that the goal of rational discourse is not to prove one's rightness, but to become right, even if this requires going through the painful process of abandoning one's own delusions. Thus, Epistemological Dueling fulfills its ultimate mission: it not only finds an answer to a specific question, but also teaches humanity how answers should be sought.