DISCLAIMER

This document provides a technical overview of the architecture, methodology, and evaluation metrics for Project HERAKLITUS - *a hybrid AI system integrating natural language models with formal verification frameworks for dialectical reasoning.*

NOTE

this is a personal project-

‘ziu-hitsu essays’ logged at kafkaclitus.substack.com

Project HERAKLITUS

Hybrid AI for Dialectical and Formal Reasoning

Overview

Project HERAKLITUS introduces a hybrid AI framework that bridges the divide between natural language fluency and structured logical verification.
Where traditional Large Language Models (LLMs) predict plausible next tokens through probabilistic methods, HERAKLITUS advances further: it inductively generates hypotheses, rigorously verifies them deductively, and dialectically refines its conclusions.

By integrating a Mistral-based LLM with Lean/Coq proof systems through a custom semantic bridge architecture, HERAKLITUS moves beyond pattern recognition into verifiable, creative philosophical reasoning.

System Architecture

HERAKLITUS has four core subsystems:

1. Semantic Bridge Layer

- Frame-semantic parsing of natural language inputs
- Modal logic translation (including epistemic and doxastic operators)
- Bidirectional verification with uncertainty quantification

2. Mistral LLM Core

- Argument structure detection and classification
- Philosophical dialogue modeling
- Probabilistic conceptual reasoning integration

3. Lean/Coq Proof Engine

- Natural language translation to formal proof structures
- Modal logic verification and theorem validation

- Counterexample detection and proof search optimization

4. Dialectical Controller

- Multi-level knowledge hierarchy management
- Epistemic state tracking and adjustment
- Automated synthesis of multi-perspective philosophical arguments

File Structure

Heraklitus/

```
├── main.py
├── model/
│   ├── llm.py
│   └── tokenizer.py
├── processing/
│   ├── dialectic.py
│   ├── proof_engine.py
│   └── semantic_bridge.py
├── reasoning/
│   ├── commonsense.py
│   ├── strategic.py
│   └── causal.py
├── config/
│   └── style_guide.py
```

Each directory reflects a key functional domain: model instantiation, proof validation, semantic parsing, and strategic reasoning orchestration.

Methodology

Hybrid Reasoning Pipeline

HERAKLITUS operates through a five-stage reasoning pipeline:

1. **Query Analysis**
   Extracts logical structures, frame semantics, and philosophical topics.

2. **Formal Translation**
   Converts natural language arguments into Lean/Coq formal proofs with modal extensions.

3. **Multi-Level Verification**
   Validates logical consistency, epistemic coherence, and uncertainty propagation.

4. **Dialectical Refinement**
   Generates counterarguments, tracks assumptions, and refines conclusions.

5. **Zuihitsu Synthesis**
   Produces uncertainty-aware philosophical essays reflecting both certainty and epistemic humility.

Knowledge Hierarchy

HERAKLITUS organizes knowledge across four ascending layers:

- **Level 1**: Formally verified theorems (strict deductive validation)
- **Level 2**: Probabilistic beliefs (with explicit confidence scores)
- **Level 3**: Hypothetical assumptions (traceable, defeasible)
- **Level 4**: Meta-cognitive evaluations (reflection on reasoning pathways)

The Dialectical Controller manages transitions across these layers, enabling both logical rigor and creative adaptability.

Technical Stack

Component	Version	Purpose
Mistral LLM	v0.4.2	Natural language understanding and generation
Lean Proof Assistant	v4.3.0	Formal logic verification with modal extensions
Coq Proof Assistant	v8.18.0	Theorem proving, proof optimization
WordNet	-	Lexical semantics for conceptual grounding

Evaluation Metrics

Technical Metrics

- **Proof Completion Rate**: Percentage of hypotheses successfully proven.
- **Translation Accuracy**: Fidelity from natural language to formal logic representations.
- **Computation Efficiency**: Speed and memory utilization during proof search and verification.

Philosophical Quality Metrics

- **Argument Novelty**: Originality and depth of generated arguments.
- **Dialectical Progression**: Quality of multi-perspective engagement and counterargument integration.
- **Conceptual Synthesis**: Ability to unify divergent philosophical insights coherently.

Research Contributions

- Development of a novel semantic bridge architecture integrating formal logic and NLP.
- Embedding structured uncertainty quantification within philosophical reasoning outputs.
- Modal logic validation applied to LLM-generated natural language discourse.
- Scalable methodology for structured, verifiable philosophical argument generation.

Applications

- **Legal Reasoning and Structured Argumentation**
  -case law reasoning, argument validation
- **Philosophical Research Automation**
  -dialectical exploration, hypothesis generation
- **AI Ethics and Governance Systems**
  -transparency, epistemic auditability

- **Education in Logic, Rhetoric, and Philosophy**
  -automated dialectical learning environments

Conclusion

Project HERAKLITUS marks a critical step beyond surface-level AI fluency toward epistemically rigorous, dialectically creative machine reasoning.
By synthesizing Lean/Coq theorem proving, semantic parsing, and probabilistic inference, HERAKLITUS not only "talks" but "thinks" \u2014 tracing the ancient pathways of logos through digital architecture.

Keywords: Hybrid AI, Formal Verification, Dialectical Reasoning, Semantic Bridge, Modal Logic, Creative AI, Computational Philosophy