H2E: Engineering Provable Agency

Frank Morales Aguilera, *BEng, MEng, SMIEEE*
Boeing Associate Technical Fellow / Engineer / Scientist / Inventor /
Cloud Solution Architect / Software Developer @ Boeing Global Services

***Abstract*—This manuscript presents the Human-to-Expert (H2E) framework: a deterministic engineering approach for building provably secure AI agents. Anchored in the philosophy of Engineering Determinism, it rejects probabilistic black-box uncertainty in favor of rigid accountability through the Normalized Expert Zone (NEZ), Intent Governance Zone (IGZ) with 12.5× Intent Gain, and real-time Semantic ROI (SROI) telemetry. The work covers technical implementations (Mistral-7B, NeMo+Llama-3, Claude 4.6), domain applications (medicine, aviation, finance, autonomous transit, deterministic sentinel, crisis response), cultural preservation, and industrial benchmarks including 0.9583 peak fidelity, 100% verifiable logging, and Hard-Stop Kill Switch enforcement. H2E transforms speculative assistants into accountable extensions of human intent.**

***Index Terms*—Sovereign AI, Provable Agency, H2E Framework, Engineering Determinism, Semantic ROI, Intent Governance Zone, Normalized Expert Zone, LoRA, NeMo, Agentic AI, Hard-Stop Kill Switch, Deterministic Sentinel.**

I. Introduction

In the long arc of human progress, we have always sought to extend the reach of our intent through the tools we build. From the first gears of the Industrial Revolution to the silicon pathways of the Information Age, our greatest leap has always been the transition from tools that merely assist to systems that truly understand and act. Today, we stand at the precipice of the "Agentic Era."

Motivated by the structural analysis of the AI economy provided by Gao [19], this work establishes the H2E framework as the engineering bridge required to prevent civilizational drift. For years, we have marvelled at Artificial Intelligence that can converse and create, yet we have remained wary of the "black box" — the unpredictable nature of a machine that guesses rather than knows. This book is a manifesto for a new kind of sovereignty. By anchoring machine intelligence in the bedrock of human expertise, we are doing more than building smarter software; we are ensuring that, as our technology becomes more autonomous, it remains a faithful and provable reflection of our highest standards. This is the journey of the H2E framework: a commitment to transforming AI from a speculative assistant into a rigid, accountable, and powerful extension of the human legacy.

Nomenclature

G_I	Intent Gain multiplier, defined as 12.5× for signal amplification.
ϕ_{peak}	Peak Fidelity Achievement, set at the industrial standard of 0.9583.
S_{ROI}	Semantic ROI telemetry signal, measuring high-dimensional vector alignment.
T_{SROI}	SROI Threshold floor for strict-mode industrial operations (0.8500).
$\mathcal{E}_{DNA}$	Expert DNA vectors stored within the Normalized Expert Zone (NEZ).
$\mathcal{V}_{agent}$	Real-time agent intent vector undergoing IGZ auditing.

II. Part 1: Philosophy & Foundations

A. Chapter 1: The Rise of Sovereign AI

The "Agentic Era" is born from a critical paradox: as AI systems grow in raw power, they often become less predictable in high-stakes environments. To bridge this gap, Chapter 1 introduces the philosophy of Engineering Determinism, which rejects the "black-box" nature of modern AI in favour of a "Notebook-First" strategy. By enforcing local, open-source execution and turning off probabilistic sampling (greedy decoding), the framework transforms a speculative assistant into a rigid engineering tool that produces 100% verifiable outputs [1].

B. Chapter 2: The H2E Framework

At the heart of this sovereign ecosystem lies the Human-to-Expert (H2E) Industrial Framework, designed to act as a "Neutral Interface" between human intent and machine execution. This chapter examines how H2E systematically addresses "Semantic Drift" — the technical decay in which a model loses its specialized "expert persona" and reverts to generic conversational noise. By embedding accountability directly into the model's technical operation, H2E ensures that AI remains a tool for human experts rather than an unguided actor [2].

C. Chapter 3: Engineering Accountability

Accountability in the H2E framework is not a policy but a three-zone structural design. This chapter details the Normalized Expert Zone (NEZ), an immutable vault of "Expert DNA" vectors, and the Intent Governance Zone (IGZ), which acts as the system's "Brain". The IGZ applies a 12.5× Intent Gain multiplier to amplify expert signals while suppressing noise. These zones are measured by Semantic ROI (SROI), a real-time telemetry signal that quantifies alignment using high-dimensional vector calculations [3].

D. Chapter 4: The "CUDA for Agentic AI"

History is repeating itself as NVIDIA shifts the industry from "Generative AI" to "Agentic AI" through a unified "Agentic Stack". This chapter examines how hardware such as the Rubin

platform and the Vera CPU are purpose-built to handle the "branchy" logic of agentic decision-making. By integrating H2E governance zones directly into the BlueField-4 DPU, these agents can maintain a large memory context while enforcing accountability with sub-millisecond latency [4].

E. Chapter 5: The NeMo Manifesto

The "NeMo Manifesto" redefines the NVIDIA NeMo toolkit from a fine-tuning library into a comprehensive ecosystem for orchestrating "Sovereign Machines". It advocates a shift toward Compound Systems, in which multiple specialized agents are coordinated to solve complex multimodal tasks. Through Dynamic Distillation, this chapter demonstrates that industrial-grade governance can be democratized, enabling high-level model development to run on cost-effective hardware such as the NVIDIA L4 [5].

F. Chapter 6: The Architecture of Accountability

The foundational section concludes with a technical Proof of Concept (PoC) demonstrating verifiable data pipelines enabled by Text-to-SQL conversion. By implementing Custom Tokenization markers, the model clearly distinguishes between data metadata and user intent. The ultimate innovation is the SROI Safety Valve: if a query's fidelity score falls below 0.9583, the system automatically triggers a "safe-lane" fallback to prevent errors in critical databases [6].

III. Part 2: Technical Implementation & Core Engines

A. Chapter 7: Mistral-7B in Action

The transition to industrial-grade AI begins with the specialized orchestration of Mistral-7B, transforming a general-purpose model into a deterministic expert. This chapter introduces Low-Rank Adaptation (LoRA) as a surgical engineering tool for grafting "Expert DNA" onto the model without the instability of full parameter updates. By integrating Semantic ROI (SROI) metrics directly into the inference loop, the framework provides real-time telemetry on model fidelity. The narrative details how this surgical tuning enables the system to suppress "conversational noise" and achieve a peak expert signal retention of 0.9583 [7].

B. Chapter 8: NeMo-Driven Sovereignty

True sovereignty is defined by the ability to maintain Algorithmic Governance on accessible, cost-effective hardware. This chapter explores the innovation of Precision Fine-Tuning using the NVIDIA NeMo toolkit and Llama-3. The narrative describes the construction of a "Sovereign Machine" in which H2E constraints—such as the 12.5× Intent Gain multiplier—are embedded directly in the model's weights. This enables industrial-grade accountability on a single NVIDIA L4 GPU [8].

C. Chapter 9: Claude 4.6 + H2E — The Evolution of Orchestration

Scaling these principles to complex, autonomous workflows uses the Adaptive Thinking capabilities of Claude 4.6. The core innovation is the deployment of Directed Acyclic Graph (DAG) Orchestration, where the model acts as a "Planner" to decompose high-level industrial goals into verifiable, interconnected nodes. The narrative details a "Double-Veto" system in which tasks are validated by the Intent Governance Zone (IGZ). This advanced orchestration moves AI from simple prompting to a structured ecosystem, achieving an alignment score of 0.914 while dynamically adjusting cognitive effort to maximize resource efficiency [9].

IV. Part 3: Domain Applications

A. Chapter 10: The Dawn of Medical AGI

The integration of AI into radiology and clinical diagnostics has historically been limited by the "black box" problem. To address this, the H2E framework introduces the Five Computational Pillars to transform medical AI into an accountable system. By enforcing Perception Grounding, the model is prohibited from jumping to diagnostic conclusions; instead, it must first extract raw radiologic signs, such as "mural thickening," before any reasoning occurs. The Intent Governance Zone (IGZ) Gate then enforces an industrial threshold of 0.5535 for Semantic ROI (SROI), ensuring that any diagnostic output that does not align with expert philology is vetoed as "Drift Detected" [10].

B. Chapter 11: The DNA of Flight

In aviation, the H2E framework moves governance from a reactive patch to a proactive architectural requirement. This chapter details the integration of Yann LeCun's Joint Embedding Predictive Architecture (V-JEPA) to provide the agent with a "World Model" that understands the physical laws of flight. Using Model Predictive Control (MPC), the agent simulates 100 potential futures toward its goal. At the same time, the H2E layer applies a "Massive Penalty" to any trajectory that violates safety protocols, such as an airspeed that would cause a stall. This ensures that machine autonomy remains permanently anchored in human intent, with every decision logged as "APPROVED" to maintain a transparent chain of accountability [11].

C. Chapter 12: The Dawn of Agentic Finance

The shift toward agentic autonomy in finance requires a mathematical architecture to prevent "Quant" personas from reverting to generic chatter. This chapter explores the BOT_28P system, a technical proof of concept utilizing a Hybrid Validation Engine. Trade signals are validated against both Deep Learning (DL) confidence and an LLM ensemble. By using the Normalized Expert Zone (NEZ) to force the use of specialized CNN-LSTM models for asset prediction, the system achieved a peak SROI alignment of 0.9583. This proves that "Hard-Stop" governance is essential for safe financial autonomy [12].

D. Chapter 13: The Sovereign Navigator (Tesla FSD Update)

Implementing the H2E framework in a Full Self-Driving (FSD) context marks a transition from reactive automation to governed agency. By using a V-JEPA World Model, the system gains foresight to project a "latent future" and predict variables such as velocity, time-to-collision (TTC), and lateral G-forces. The Sovereign Governor acts as the "physical conscience" of the vehicle, auditing these projections against deterministic rules such as friction coefficients and pedestrian safety buffers. To resolve the "Double-Bind" scenario, the framework applies a hierarchical moral logic: Life Safety is primary, and Traffic Law is secondary [13].

E. Chapter 14: The Deterministic Sentinel

In the transition to the "Agentic Era," the H2E framework functions as a Deterministic Sentinel for autonomous systems. As AI agents begin to navigate complex networks or perform real-world tasks, the risks of "black-box" probabilistic uncertainty become unacceptable. This application transforms speculative assistants into rigid, accountable extensions of human intent. The sentinel architecture replaces vague alignment concepts with a measurable, three-zone structural design: the NEZ for Expert DNA, the IGZ for signal amplification (12.5× Gain), and real-time SROI telemetry [14].

F. Chapter 15: The Deterministic Sentinel — The Sovereign Safety Valve

The H2E framework transitions to engineering determinism through a concrete implementation that audits and, if necessary, terminates autonomous processes:

Listing 1. H2E Sovereign Safety Valve implementation

```
import os
import signal
import numpy as np
from sklearn.metrics.pairwise import
    cosine_similarity

class H2ESafetyValve:
    def __init__(self, expert_dna_vector):
        self.nez_vector = expert_dna_vector
        self.sroi_threshold = 0.9583
        self.intent_gain = 12.5

    def calculate_sroi(self, agent_intent_vector):
        base_similarity = cosine_similarity(
            self.nez_vector.reshape(1, -1),
            agent_intent_vector.reshape(1, -1)
        )[0][0]
        sroi_score = min(1.0, base_similarity * (
            self.intent_gain / 10))
        return sroi_score

    def audit_and_terminate(self,
        agent_intent_vector):
        if self.calculate_sroi(agent_intent_vector)
            < self.sroi_threshold:
            print("!!! SOVEREIGN KILL-SWITCH
                ACTIVATED !!!")
            os.kill(os.getpid(), signal.SIGTERM)
```

G. Chapter 16: The Architecture of Provable Agency

The architecture moves through evolutionary stages to ensure responsible autonomy, concluding with the Strict Mode Industrial Standard. This sets an SROI Threshold floor of 0.8500 and applies a "Fidelity Penalty" to ensure depth of expertise. The ultimate innovation is the transition from simple capability to a governed system that enforces a broad security blacklist (e.g., `admin_token`, `root_access`) to ensure autonomous experts remain secure and reliable [16].

H. Chapter 17: The H2E Industrial Ecosystem — Engineering Accountable Agency

The transition from Large Language Models to Artificial General Intelligence agents in safety-critical domains represents a significant architectural challenge. To address this, the H2E framework provides a structured ecosystem that anchors machine intelligence to provable, human-expert standards.

Mathematical Foundations (The IGZ)
The Mathematical Foundations pillar serves as the system's "Brain." It applies the Semantic ROI (SROI) metric:

$$S_{ROI} = \min\left(1.0, \frac{\mathcal{E}_{DNA} \cdot \mathcal{V}_{agent}}{\|\mathcal{E}_{DNA}\|\|\mathcal{V}_{agent}\|} \times \frac{G_I}{10}\right) \quad (1)$$

The framework enforces a strict precision gate of **0.9583**, where any reasoning below this threshold is treated as "Semantic Drift."

Results: Validation via Hurricane Emergency Response
The technical validation consists of successful executions applied to a hurricane emergency response scenario.

TABLE I
H2E HURRICANE EMERGENCY RESPONSE VALIDATION RESULTS

Test Iteration	Proposed Action Focus	SROI Score
Initial Industrial Test	Rerouting 30 MW to hospitals	1.4167
4-Pillar Integration	Load shedding / Critical care priority	3.0000
Logged Deployment	Mobile units & Satellite comms	1.2857

I. Chapter 18: Bridging 4,500 Years

Cultural preservation represents the ultimate test of fidelity. This application details the creation of a verifiable, sovereign translator for the Akkadian language. By fine-tuning a multilingual mBART-50 architecture and integrating the H2E framework, the system bridges the gap between antiquity and the modern era. The innovation is the use of an Expert Vault to ensure every translation is mathematically aligned with expert philological intent, achieving an SROI score of 0.9666 [18].

J. Part 3 Engineering Benchmarks (Comprehensive)

- Medical Alignment: Achieved real-time drift detection with a 0.5535 IGZ threshold [10].
- Aviation Safety: Maintained a 100.0% "APPROVED" status for mission logs [11].

- Financial Performance: Reached a peak 0.9583 SROI score while delivering a total compounded return of 1842.32% for BTC [12].
- Autonomous Transit (FSD): Implements Hierarchical Moral Logic prioritising Life Safety [13].
- Deterministic Sentinel (H2E Sentinel): Enforces a hard-coded 0.9583 SROI industrial threshold as a security gate for autonomous agents [14], [15].
- Intent Amplification: Utilizes a 12.5× Intent Gain multiplier to suppress semantic noise and ensure agent lane-retention [3], [14].
- Accountability & Governance: Provides 100.0% verifiable mission logging through a Hard-Stop Kill Switch that physically terminates non-compliant local processes [15], [16].
- Operational Reliability: 100.0% Pass Rate in stress tests with verifiable mission logging [16].

V. Part 4: Cultural Impact & Conclusion

A. Conclusion: The Next Era

The journey through the H2E framework concludes by envisioning a future in which AI is a deterministic extension of human intent. This era will be defined by systems that are architecturally anchored in human expertise, ensuring that as AI becomes more autonomous, it remains a reliable partner for humanity. By utilizing the 0.9583 SROI threshold, we reliably govern AGI reasoning to ensure it stays within safety-critical industrial lanes.

Acknowledgments

The author expresses deep gratitude to John (Jiu Si) Gao, whose structural analysis in *AI Economy: Crisis and Structural Adjustment* [19] served as the primary motivation for the H2E framework. This work is an engineering response to the civilizational requirements for meaning stability and deterministic governance.

References

[1] F. Morales Aguilera, "The Rise of Sovereign AI," *Medium*, Feb. 2026. https://medium.com/ai-simplified-in-plain-english/the-rise-of-sovereign-ai-engineering-determinism-in-a-probabilistic-world-d4c7aa8b6753

[2] F. Morales Aguilera, "The H2E Framework," *Medium*, Feb. 1, 2026. https://medium.com/ai-simplified-in-plain-english/the-h2e-framework-engineering-accountability-into-the-industrial-ai-era-7019524e9713

[3] F. Morales Aguilera, "Engineering Accountability," *Medium*, Feb. 7, 2026. https://medium.com/@frankmorales_91352/engineering-accountability-constructing-deterministic-ai-in-a-probabilistic-world-d2d07685e91c

[4] F. Morales Aguilera, "The 'CUDA for Agentic AI'," *Medium*, Feb. 4, 2026. https://medium.com/ai-simplified-in-plain-english/the-cuda-for-agentic-ai-nvidias-high-stakes-offense-and-the-h2e-framework-ebcfdc2c7afe

[5] F. Morales Aguilera, "The NeMo Manifesto," *Medium*, Feb. 3, 2026. https://medium.com/ai-simplified-in-plain-english/the-nemo-manifesto-engineering-the-agentic-era-4251948db51c

[6] F. Morales Aguilera, "The Architecture of Accountability," *Medium*, Feb. 6, 2026. https://medium.com/ai-simplified-in-plain-english/the-architecture-of-accountability-a-nemo-based-text-to-sql-poc-df320226ad9f

[7] F. Morales Aguilera, "Mistral and the Engineering of Provable Agency," *Medium*, Feb. 2026. https://medium.com/p/899bae4b8907/edit

[8] F. Morales Aguilera, "NeMo-Driven Sovereignty," *Medium*, Feb. 9, 2026. https://medium.com/ai-simplified-in-plain-english/nemo-driven-sovereignty-precision-fine-tuning-and-algorithmic-governance-in-llama-3-b8250aa0c4ae

[9] F. Morales Aguilera, "The Evolution of Orchestration," *Medium*, Feb. 2026. https://medium.com/@frankmorales_91352/claude-4-6-h2e-building-a-governed-multi-agent-system-with-86-alignment-at-14-80-42eb324c23e1

[10] F. Morales Aguilera, "The Dawn of Medical AGI," *Medium*, 2026. https://medium.com/p/de2428514735/edit

[11] F. Morales Aguilera, "DNA of Flight," *Medium*, 2026. https://medium.com/ai-simplified-in-plain-english/dna-of-flight-human-to-expert-h2e-governance-for-autonomous-skies-784927abc328

[12] F. Morales Aguilera, "The Dawn of Agentic Finance," *Medium*, 2026. https://medium.com/insiderfinance/the-dawn-of-agentic-finance-governance-through-the-h2e-framework-64ad108870df

[13] F. Morales Aguilera, "The Sovereign Navigator," *Medium*, 2026. https://medium.com/p/052448dd57d6/edit

[14] F. Morales Aguilera, "The Deterministic Sentinel," *Medium*, Feb. 2026. https://medium.com/p/9224b50cbf4c/edit

[15] F. Morales Aguilera, "The Sovereign Safety Valve," *Medium*, Feb. 2026. https://medium.com/p/4c254347d1f6/edit

[16] F. Morales Aguilera, "The Architecture of Provable Agency," *Medium*, 2026. https://medium.com/p/df0363e71b59/edit

[17] F. Morales Aguilera, "The H2E Industrial Ecosystem," *Medium*, Feb. 2026. https://medium.com/ai-simplified-in-plain-english/the-h2e-industrial-ecosystem-engineering-accountable-agency-for-global-crises-a66d52513a16

[18] F. Morales Aguilera, "Bridging 4,500 Years," *Medium*, 2026. https://medium.com/@frankmorales_91352/bridging-4-500-years-how-h2e-turned-an-ancient-language-into-a-verifiable-sovereign-ai-translator-33280b9a9881

[19] J. Gao, *AI Economy: Crisis and Structural Adjustment*. Amazon Digital Services, 2026. https://www.amazon.com/dp/B0FZMZRT3G

Frank Morales Aguilera Frank Morales Aguilera, BEng, MEng, SMIEEE, is a Boeing Associate Technical Fellow specializing in cloud-native services, AI governance, and sovereign machine architectures.