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), 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." 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. It is an invitation to move beyond the era of probabilistic uncertainty and toward an engineering era of determinism. 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. The era of the "Sovereign Machine" has arrived. It is time to engineer agency with purpose.

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).

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 (up to 1M tokens) 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 (e.g., [SCHEMA_START]), 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, demonstrating that even smaller models can outperform larger "black-box" systems under H2E constraints [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, democratizing the ability for organizations to run secure, expert-aligned agents in edge or on-premises environments without relying on third-party cloud providers [8].

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

The final chapter of Part 2 scales these principles to complex, autonomous workflows using 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—such as technical writing or security reviews—are executed in parallel or sequentially based on explicit dependencies, with each step 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 (86% in specific industrial stress tests) 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—the inability to demonstrate mathematically that a model's output aligns with expert intent. 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 conclusions; instead, it must first extract raw radiologic signs, such as "mural thickening" or "fat stranding," 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". This transition moves healthcare from "Probabilistic AI" to "Accountable AI" through strict governance gates [10].

B. Chapter 11: The DNA of Flight

In aviation, where the margin for error is zero, 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. Embedding physical reasoning with H2E gates ensures autonomous flight remains within expert safety manifolds [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 or hallucinated trends. This chapter explores the BOT_28P system, a technical proof of concept that utilizes a Hybrid Validation Engine. The innovation is a Double-Veto system that validates trade signals against both Deep Learning (DL) confidence and an LLM ensemble consisting of DeepSeek and Qwen. 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 and eliminating semantic noise [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. Modern autonomous systems often operate as “black boxes,” with the path from perception to actuation opaque. By using a V-JEPA World Model as the “Agent,” 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 “legal and 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. By authorizing “Approved Exceptions,” the Expert allows the vehicle to violate a legal boundary if it is necessary to preserve human life and the path is clear [13].

E. Chapter 14: The Deterministic Sentinel

In the rapid 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 via RentAHuman.ai, 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 from abstract policy to engineering determinism through a concrete implementation that audits and, if necessary, terminates autonomous processes. The following implementation demonstrates the H2E Sentinel Gate:

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):
        # NEZ: Immutable vault of Expert DNA
        self.nez_vector = expert_dna_vector
        self.sroi_threshold = 0.9583 # Industrial
            Peak Fidelity
        self.intent_gain = 12.5      # IGZ Signal
            Multiplier

    def calculate_sroi(self, agent_intent_vector):
        """Telemetry signal for expert alignment."""
        base_similarity = cosine_similarity(
            self.nez_vector.reshape(1, -1),
            agent_intent_vector.reshape(1, -1)
        )[0][0]
        # Apply 12.5x Intent Gain to suppress
            semantic noise
        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 !!!")
            # Physically terminate the local process
                 for safety
            os.kill(os.getpid(), signal.SIGTERM) #
                Hard-Stop
```

G. Chapter 16: The Architecture of Provable Agency

The architecture moves through distinct 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].

1) Part 3 Engineering Benchmarks (Comprehensive): The comprehensive benchmarks for Part 3 now include the high-fidelity strict-mode standards required for industrial-scale complexity:

- Medical Alignment: Achieved real-time drift detection with a 0.5535 IGZ threshold.
- Aviation Safety: Maintained a 100.0% “APPROVED” status for descent and airspeed signals in mission logs.
- Financial Performance: Reached a peak 0.9583 SROI score while delivering a total compounded return of 1842.32% for BTC.
- Autonomous Transit (FSD): Implements Hierarchical Moral Logic where Life Safety is primary, and Traffic Law is secondary, authorizing “Approved Exceptions” to preserve human life.
- Deterministic Sentinel (H2E Sentinel): Enforces a hard-coded 0.9583 SROI industrial threshold as a security gate for autonomous agents.
- Intent Amplification: Utilizes a 12.5× Intent Gain multiplier to suppress semantic noise and ensure agent lane-retention.
- Accountability & Governance: Provides 100.0% verifiable mission logging through a Hard-Stop Kill Switch that physically terminates non-compliant local processes.

- H2E Strict Mode Floor: Elevates the SROI Threshold to a high floor of 0.8500 for general autonomous experts.
- Fidelity Penalty: Rejects any response shorter than 25 characters to ensure a depth of expertise.
- Security Governance: Expanded blacklist includes `admin_token`, `root_access`, and `config_bypass` to prevent exfiltration attempts.
- Operational Reliability: 100.0% Pass Rate in stress tests with verifiable mission logging via the Hard-Stop Kill Switch.

V. Part 4: Cultural Impact & Conclusion

A. *Chapter 17: Bridging 4,500 Years*

Cultural preservation represents the ultimate test of fidelity, where the H2E framework is used to recover "lost" voices from history. This application details the creation of a verifiable, sovereign translator for the Akkadian language. By fine-tuning a multilingual mBART-50 architecture on curated ancient texts 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 [17].

B. *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 and predictable partner for humanity.

1) *Final Mini-Book Structure at a Glance:*

1) Part 1: Foundations — Philosophical and strategic cornerstones (Chapters 1–6).
2) Part 2: Core Implementations — Technical deep-dives with Mistral-7B, Llama-3, and Claude 4.6 (Chapters 7–9).
3) Part 3: Domain Applications — Scaling sovereignty across Medicine, Aviation, Finance, Autonomous Transit, Deterministic Sentinel, and The Architecture of Provable Agency (Chapters 10–16).
4) Part 4: Cultural Impact & Conclusion — Preserving the past and engineering the future legacy (Chapter 17 & Conclusion).

References

[1] F. Morales Aguilera, "The Rise of Sovereign AI," *Medium*, Feb. 2026.

[2] F. Morales Aguilera, "The H2E Framework," *Medium*, Feb. 1, 2026.

[3] F. Morales Aguilera, "Engineering Accountability," *Medium*, Feb. 7, 2026.

[4] F. Morales Aguilera, "The "CUDA for Agentic AI"," *Medium*, Feb. 4, 2026.

[5] F. Morales Aguilera, "The NeMo Manifesto," *Medium*, Feb. 3, 2026.

[6] F. Morales Aguilera, "The Architecture of Accountability," *Medium*, Feb. 6, 2026.

[7] F. Morales Aguilera, "Mistral-7B: Engineering Accountability Through Code," *Medium*, Feb. 2026.

[8] F. Morales Aguilera, "NeMo-Driven Sovereignty: Precision Fine-Tuning with Llama-3," *Medium*, Feb. 9, 2026.

[9] F. Morales Aguilera, "The Evolution of Orchestration: Mastering Claude 4.6 and the H2E Framework," *Medium*, Feb. 2026.

[10] F. Morales Aguilera, "The Dawn of Medical AGI: Engineering Accountability through the H2E Framework," *Medium*. [Online]. Available: https://medium.com/p/de2428514735/edit

[11] F. Morales Aguilera, "DNA of Flight: Human-to-Expert (H2E) Governance for Autonomous Skies," *Medium*. [Online]. Available: https://medium.com/p/784927abc328/edit

[12] F. Morales Aguilera, "The Dawn of Agentic Finance: Governance through the H2E Framework," *Medium*. [Online]. Available: https://medium.com/@frankmorales_91352/the-dawn-of-agentic-finance-governance-through-the-h2e-framework-64ad108870df

[13] F. Morales Aguilera, "The Sovereign Navigator: Implementing H2E Governance in Tesla's FSD World Model," *Medium*. [Online]. Available: https://medium.com/p/052448dd57d6/edit

[14] F. Morales Aguilera, "The Deterministic Sentinel," *Medium*, Feb. 2026.

[15] F. Morales Aguilera, "The Deterministic Sentinel — The Sovereign Safety Valve," *Medium*, Feb. 2026.

[16] F. Morales Aguilera, "The Architecture of Provable Agency: From Functional Autonomy to H2E Governance," *Medium*. [Online]. Available: https://medium.com/p/b9123bd81d74/edit

[17] F. Morales Aguilera, "Bridging 4,500 Years: How H2E Turned an Ancient Language into a Verifiable, Sovereign AI Translator," *Medium*. [Online]. Available: https://medium.com/p/33280b9a9881/edit

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.