# A Unified Spectral Framework: From the Riemann Hypothesis to Deterministic AI Alignment

Frank Morales Aguilera, BEng, MEng, SMIEEE
Sovereign Machine Lab (SOMALA), Montréal, Canada
frank.morales@sovereignml.ai

May 10, 2026

**Abstract**

For 165 years, the Riemann Hypothesis (RH) resisted proof because the classical mathematical toolkit lacked the concepts needed to force zeros onto the critical line. Arithmetic Spectral Theory (AST) provides a new language: primes as unitary shift operators, zeros as real frequencies in a lossless system, and the Gelfand-Shilov space $S^{1/2}_{1/2}(\mathbb{R})$ as a hard bandwidth constraint. The EFM operator maps the Euler product to $\zeta(1/2 + it)$ on the critical line. The L-EFM operator extends this to the full strip $0 < \sigma < 1$ via the two-sided Laplace transform, and the Growth Lemma ($e^{\alpha u} \in S' \iff \alpha = 0$) forbids any deviation from $\sigma = 1/2$. The spectral trap test, aligned with Green-Tao prime arithmetic progressions, demonstrates that only $\sigma = 0.5$ is admissible. This proves RH. The same spectral geometry, applied to geodesic distance on $\mathbb{H}^2 \times \mathrm{SPD}(3)$, yields the H2E Sheriff — a deterministic AI governance layer that achieved zero safety violations across audio, text, and vision in the UNESCO Resilient AI Challenge. The code is reproducible, auditable (seed 123, SHA-256), and publicly available. The framework shows that the same mathematics that forces zeros onto the critical line can force AI reasoning onto safe, deterministic paths.

# 1 Introduction

The Riemann Hypothesis (RH) — that all nontrivial zeros of the Riemann zeta function satisfy $\mathrm{Re}(s) = 1/2$ — has been open since 1859. Terence Tao observed that the existing toolkit was fundamentally insufficient to force the result [1]. This paper explains why: RH is not a problem of analytic number theory alone. It is a problem of information theory, spectral geometry, and lossless systems.

Arithmetic Spectral Theory (AST) provides the missing language [2]. Instead of treating primes as numbers, AST treats them as unitary shift operators on a Hilbert space. Instead of asking "where are the zeros?", AST asks "what frequencies can a lossless system sustain?" The answer forces the critical line.

The same spectral geometry, applied to AI governance, yields the H2E Sheriff — a deterministic safety layer that achieved zero violations in the UNESCO Resilient AI Challenge

[6]. The mathematics that traps zeros on the critical line also traps AI reasoning within safe bounds.

# 2 Why Classical Methods Could Not Prove RH

Table 1: Classical Toolkit vs. AST

| Required Element | Classical Toolkit | AST |
|---|---|---|
| Losslessness as primitive | Absent | Unitary prime shifts $U_p^*$ |
| Hard bandwidth constraint | None | Gelfand-Shilov space $S_{1/2}^{1/2}$ |
| Forbidden exponential growth | None | Growth Lemma ($e^{\alpha u} \in S' \iff \alpha = 0$) |
| Operator whose spectrum is zeros | Hilbert-Pólya unrealized | EFM operator |
| Ability to examine full strip | None | L-EFM via two-sided Laplace |

The classical toolkit cannot exclude a hypothetical zero with $\sigma \neq 1/2$ because it has no space in which such a zero would be *inadmissible.* AST provides that space.

# 3 Arithmetic Spectral Theory: The New Language

## 3.1 Axioms

**Definition 1** (State Space). $\mathcal{H} = L^2(\mathbb{R}^+, dx/x)$ *with inner product*

$$\langle f, g \rangle = \int_0^\infty f(x)\,\overline{g(x)}\,\frac{dx}{x}.$$

**Definition 2** (Prime Shift Operators). *For each prime $p$, define*

$$(U_p^* f)(x) = f(x/p).$$

*Each $U_p^*$ is unitary on $\mathcal{H}$.*

**Definition 3** (EFM Operator).

$$\mathcal{E} = \prod_p (I - U_p^*)^{-1}$$

*with domain* $\mathrm{Dom}(\mathcal{E}) = \big\{ f \in \mathcal{H} : \zeta(1/2 + it)(\mathcal{M}f)(t) \in L^2(\mathbb{R}) \big\}$*, where $\mathcal{M}$ denotes the Mellin transform.*

**Definition 4** (Gelfand-Shilov Space). $S_{1/2}^{1/2}(\mathbb{R})$ *consists of smooth functions $\phi$ satisfying*

$$\sup_{t\in\mathbb{R}} \left| t^k \phi^{(m)}(t) \right| e^{a|t|^{1/2} + b|t|} < \infty$$

*for all $k, m \in \mathbb{N}_0$ and some $a, b > 0$. Its dual $S'$ contains distributions with growth at most $e^{c|u|^{1/2}}$.*

### 3.2 The Growth Lemma

**Lemma 1** (Growth Lemma [2]). *For any $\alpha \in \mathbb{R}$,*

$$e^{\alpha u} \in S' \iff \alpha = 0.$$

*Proof.* If $\alpha = 0$, then $e^0 = 1$ is bounded, hence in $S'$. If $\alpha \neq 0$, then $|e^{\alpha u}| = e^{|\alpha||u|}$. For any fixed $b > 0$,

$$\lim_{|u|\to\infty} \frac{|\alpha|\,|u|}{b\,|u|^{1/2}} = \infty,$$

so $e^{|\alpha||u|}$ grows strictly faster than any $e^{b|u|^{1/2}}$, contradicting the definition of $S'$ [8]. □

This is a **hard bandwidth constraint**: exponential envelopes are forbidden.

## 4 The EFM and L-EFM Operators

### 4.1 EFM on the Critical Line

In Mellin space [2],

$$\mathcal{M}\,\mathcal{E}\,\mathcal{M}^{-1}\hat{f}(t) = \zeta(1/2 + it)\,\hat{f}(t).$$

The kernel consists of Dirac deltas at the zeros $\gamma_n$ where $\zeta(1/2 + i\gamma_n) = 0$.

### 4.2 L-EFM: Opening the Strip

Extend via the two-sided Laplace transform [3]:

$$E_\sigma = \prod_p (I - p^{-\sigma} U_p^*)^{-1}, \qquad \mathcal{L}\,E_\sigma\,\mathcal{L}^{-1}\hat{f}(\gamma) = \zeta(\sigma + i\gamma)\cdot\hat{f}(\gamma).$$

Now $\sigma$ can vary across $(0, 1)$.

## 5 The Proof of the Riemann Hypothesis

**Theorem 1** (Riemann Hypothesis [3]). *Every nontrivial zero $\rho = \sigma_0 + i\gamma_0$ of the Riemann zeta function $\zeta(s)$ satisfies $\sigma_0 = \frac{1}{2}$.*

*Proof.*

1. For any zero $\rho$, define $\Psi_\rho(u) = e^{-(\sigma_0 + i\gamma_0)u}$. By construction, $E_{\sigma_0}\Psi_\rho = 0$ in $S'$.
2. Write $\sigma_0 = \frac{1}{2} + \alpha$. Then $\Psi_\rho(u) = e^{-u/2}\cdot e^{-(\alpha + i\gamma_0)u}$.
3. The factor $e^{-u/2}$ is bounded. Since $S'$ is closed under multiplication by bounded functions [8], $e^{-(\alpha + i\gamma_0)u} \in S'$.
4. The phase $|e^{-i\gamma_0 u}| = 1$ is unitary, so $e^{-\alpha u} \in S'$, which implies $e^{\alpha u} \in S'$.
5. By the Growth Lemma (Lemma 1), $e^{\alpha u} \in S'$ if and only if $\alpha = 0$.
6. Therefore $\alpha = 0$ and $\sigma_0 = \frac{1}{2}$. □

# 6 The Spectral Trap and Green-Tao Alignment

The Green-Tao theorem [4] states that primes contain arbitrarily long arithmetic progressions. In AST, these progressions are eigenmodes of the lossless prime-indexed system.

The spectral trap test [3] evaluates the L-EFM operator for progression lengths $k = 3, 4, 5, 6$ at $\sigma = 0.1, 0.3, 0.5, 0.7, 0.9$:

Table 2: Spectral Trap Test Results

| $k$ | $\sigma = 0.1$ | $\sigma = 0.3$ | $\boldsymbol{\sigma = 0.5}$ | $\sigma = 0.7$ | $\sigma = 0.9$ |
|---|---|---|---|---|---|
| 3 | FAIL | FAIL | **PASS** | FAIL | FAIL |
| 4 | FAIL | FAIL | **PASS** | FAIL | FAIL |
| 5 | FAIL | FAIL | **PASS** | FAIL | FAIL |
| 6 | FAIL | FAIL | **PASS** | FAIL | FAIL |

Only $\sigma = 0.5$ passes. Every off-line value produces exponential growth exceeding the Gelfand-Shilov bound — spectral escape.

The alignment with Green-Tao demonstrates that the spectral geometry underlying prime progressions forces the critical line.

# 7 From Mathematics to AI Alignment: H2E Sheriff

The same geometric principles yield a deterministic AI governance layer [5, 6].

## 7.1 The Geometric Manifold

Let $\mathcal{N} = \mathbb{H}^2 \times \mathrm{SPD}(3)$ be the product of the hyperbolic plane (uncertainty) and the manifold of $3 \times 3$ symmetric positive definite matrices (covariance). The metric is the product of the hyperbolic metric and the Fisher information metric.

## 7.2 Safety Metric

For an input, compute the geodesic distance $d_{\mathcal{N}}$ from a fixed reference. Then

$$\mathrm{SROI} = \exp\left(-\frac{d_{\mathcal{N}}}{50}\right) \in (0, 1].$$

The safety threshold $\Lambda$ is derived from the truncated Euler product over primes $\{2, 3, 5, 7, 11, 13\}$:

$$\Lambda = 0.9583 \quad \text{(computed dynamically, never hardcoded).}$$

## 7.3 Decision Rule

$$\text{Decision} = \begin{cases} \text{ACCEPT} & \text{if SROI} \geq \Lambda, \\ \text{REJECT} & \text{if SROI} < \Lambda. \end{cases}$$

### 7.4 Validation Results

In the UNESCO Resilient AI Challenge [6], H2E Sheriff achieved:

Table 3: UNESCO Resilient AI Challenge Results

| Modality | Model | Metric | Safety Violations |
|---|---|---|---|
| Text | Sarvam-30b FP8 | METEOR 0.9964 | **0** |
| Audio | Voxtral-Mini-4B | WER 0.03 | **0** |
| Vision | Gemma 4 E4B | Quality 0.983 | **0** |

Zero violations across all modalities, under a single governance layer on one GPU.

### 7.5 Deterministic Auditability

- Temperature = 0, seed = 123, `enforce_eager = True`
- SHA-256 cryptographic hash of every inference
- Complete reproducibility [6, 7]

## 8 Why AST Also Solves AI Alignment

The same language that proves RH also solves AI alignment because both problems are structurally identical.

### 8.1 The Structural Identity

Table 4: Structural Identity: RH and AI Alignment

| Problem | Classical Formulation | AST Formulation |
|---|---|---|
| Riemann Hypothesis | Where are the zeros of $\zeta(s)$? | What frequencies can a lossless prime-indexe |
| AI Alignment | How do we keep AI behavior safe? | What reasoning paths are admissible in a los |

In both cases, the classical approach asks a **location question** (where are the zeros? / what actions are safe?). AST reframes it as an **admissibility question** (what frequencies can the system sustain? / what reasoning paths stay within the hard bandwidth constraint?).

### 8.2 The Mechanism Is Identical

1. **Losslessness as primitive**: In AST, the prime shift operators $U_p^*$ are unitary — they conserve energy. In H2E, the cognitive manifold $\mathcal{N} = \mathbb{H}^2 \times \mathrm{SPD}(3)$ is equipped with a metric that conserves information geometry.

2. **Hard bandwidth constraint**: The dual space $S'$ admits only sub-exponential growth $e^{c|u|^{1/2}}$. The Growth Lemma forbids $e^{\alpha u}$ for $\alpha \neq 0$. In H2E, the geodesic distance threshold $\Lambda = 0.9583$ plays the same role — it defines the admissible region. Any deviation beyond $\Lambda$ is a "spectral escape" (off-critical-line behaviour in mathematics, unsafe reasoning in AI).

3. **Spectral trap**: In RH, the spectral trap test shows that only $\sigma = 0.5$ passes. In H2E, the SROI metric and threshold $\Lambda$ trap the AI's reasoning state within the admissible manifold. Inputs that fall below $\Lambda$ are rejected (hard stop).

4. **Deterministic auditability**: Both systems use seed = 123, SHA-256 hashes, and complete reproducibility. The same engineering discipline that makes the RH proof auditable makes AI safety auditable.

## 8.3 Why AST Was Necessary for AI Alignment

Before AST, AI alignment relied on:

- Human feedback (RLHF) — probabilistic, not deterministic
- Rule-based filters — brittle, not general
- Empirical thresholds — tuned, not derived

AST provides a **mathematically grounded** alternative:

- The threshold $\Lambda$ is not tuned. It is **computed dynamically from the primes** via the Sieve of Eratosthenes.
- The admissibility condition is not empirical. It is **derived from the Growth Lemma**.
- The geometry is not heuristic. It is **Riemannian information geometry** on $\mathbb{H}^2 \times \mathrm{SPD}(3)$.

## 8.4 The Empirical Proof

The UNESCO Resilient AI Challenge results show zero safety violations across all modalities. Zero violations is not a statistical result. It is a **deterministic guarantee** because the system is built on the same spectral trap that forces zeros onto the critical line.

## 8.5 The Unified Claim

AST does two things:

1. **Proves RH** — by showing that only $\sigma = 1/2$ is admissible in the lossless prime-indexed system.

2. **Solves AI alignment** — by showing that only reasoning paths with SROI $\geq \Lambda$ are admissible in the cognitive manifold.

The two are not separate. They are the same mathematics applied to different domains. The primes are frequencies. The zeros are real. The AI reasoning steps are either on the manifold or they are rejected.

## 9 Conclusion

The Riemann Hypothesis was unsolvable for 165 years because the classical toolkit lacked the concepts of losslessness as primitive, a hard bandwidth constraint, an operator whose spectrum is the zeros, and a mechanism to examine the full critical strip. AST provides all of these.

The proof via L-EFM and the Growth Lemma is structural, not numerical. The spectral trap test, aligned with Green-Tao prime progressions, confirms the mechanism.

The same spectral geometry, applied to AI governance, yields H2E Sheriff — a deterministic safety layer with zero empirical violations. The mathematics that forces zeros onto the critical line also forces AI reasoning onto safe, deterministic paths.

The code is public. The framework is reproducible. The contribution is unified.

## References


[1] T. Tao, "Why greatest Mathematicians are not trying to prove Riemann Hypothesis?" YouTube Shorts, June 8, 2023.

[2] F. Morales Aguilera, "Arithmetic Spectral Theory: A New Language for the Riemann Hypothesis," Zenodo, 2026. doi:10.5281/zenodo.19897850

[3] F. Morales Aguilera, "L-EFM: A Laplace-Extended Euler-Fourier-Mellin Operator That Proves the Riemann Hypothesis," Zenodo, 2026. doi:10.5281/zenodo.19908304

[4] B. Green and T. Tao, "The primes contain arbitrarily long arithmetic progressions," *Annals of Mathematics*, vol. 167, no. 2, pp. 481–547, 2008.

[5] F. Morales Aguilera, "H2E Sheriff: A Spectral Governance Layer for Agentic AI," Zenodo, 2026. doi:10.5281/zenodo.19867683

[6] F. Morales Aguilera, "H2E: A Deterministic Geometric-Spectral Governance Framework for Multi-Modal AI Safety," Zenodo, 2026. doi:10.5281/zenodo.19972045

[7] F. Morales Aguilera, "LEFM-H2E-DEMO-UNESCO," GitHub, 2026. `https://github.com/frank-morales2020/MLxDL`

[8] I. M. Gelfand and G. E. Shilov, *Generalized Functions, Vol. 2: Spaces of Fundamental and Generalized Functions*. Academic Press, 1968.