Humanity's Final Conjecture: Evaluation of AI Innovation Capability Based on Prime Number Distribution

Wang Liang*

School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, 430070, P.R. China

*Corresponding author: E-mail: wangliang.f@gmail.com

Abstract

As Large Language Models (LLMs) gradually approach the limits of human capability in traditional benchmarks such as MMLU and MATH, the field of artificial intelligence is facing a profound epistemological crisis: existing evaluation systems are increasingly losing their validity in distinguishing between a model's "knowledge reproduction capability" and its "source innovation capability." Although frontier benchmarks like "Humanity' s Last Exam" (HLE) and "FrontierMath" attempt to extend the validity of evaluation by increasing problem difficulty and closedness, they are essentially still confined to the assessment of "convergent thinking"—that is, finding a predetermined answer within a known solution space. However, the core of scientific discovery lies in "divergent thinking" and "abductive reasoning": the ability to construct new definitions, discover new isomorphisms, and propose new conjectures in unknown territories.

This report proposes a brand-new evaluation paradigm—the **"Innovation Turing Test"**—and specifically constructs an open-ended test case based on the intersection of number theory and nonlinear dynamics: the "Prime-Chaos Conjecture." This test requires large models to bridge the formal gap between the Peano arithmetic axiom system and symbolic dynamics based on a heuristic research guide, demonstrating that the pseudorandomness of prime distribution is actually a manifestation of low-dimensional deterministic chaos, and precisely deriving the topological properties of the Logistic map at the band merging point ($u \approx 1.5437$).

This study details the theoretical basis, evaluation conditions, and scoring logic of this test, constructs a set of evaluation standards for open scientific problems, and a scalable

evaluation method involving human-AI collaboration and shared RAG knowledge bases. It then presents real-world tests on frontier models such as Gemini and Qwen. The goal of this study is to provide a quantitative yardstick for the transition of AGI (Artificial General Intelligence) from "Problem Solvers" to "Researchers."

1. Introduction: Evaluation Crisis and Paradigm Shift in the Era of Reasoning Models

1.1 Saturation of Static Benchmarks and "Humanity’s Last Exam"

The speed of artificial intelligence development has surpassed the measures designed to gauge it. Over the past few years, we have witnessed the dominant performance of large language models, represented by GPT-4, on general tasks. MMLU (Massive Multitask Language Understanding) was once considered the gold standard for measuring a model's generalized understanding capability, covering a wide range of topics from elementary mathematics to professional law exams. However, with the emergence of models like DeepSeek-V3, Claude 3.5 Sonnet, and OpenAI o1, MMLU scores have generally breached the 90% mark, essentially rendering it a "solved" task [2]. This phenomenon of "Benchmark Saturation" brings a severe problem: when all top-tier models score within the margin of error, how do we distinguish their capability differences in handling truly complex, unknown intellectual labor?

In response, the Center for AI Safety and Scale AI jointly released **"Humanity’s Last Exam" (HLE)** [1]. HLE represents the pinnacle of current closed-ended evaluation, containing 2,500 difficult problems carefully designed by experts in various fields that cannot be answered via simple search engine retrieval. These questions involve not only mathematics, humanities, and natural sciences but also have multimodal characteristics, with about 14% of the questions requiring chart understanding, aiming to test the model's deep reasoning ability rather than mere pattern matching [1]. Unlike MMLU, HLE set an extremely high threshold from the start; the accuracy of current state-of-the-art (SOTA) models in initial tests was even below 5%, which to some extent alleviated the inflation of evaluation metrics [1].

Concurrently, Epoch AI launched **FrontierMath**, a benchmark focused on research-level higher mathematics problems [3]. FrontierMath requires not just calculation, but extremely complex symbolic reasoning and mathematical intuition; its problems often take human mathematical experts hours or even days to solve. These benchmarks undoubtedly raise the definition of "reasoning ability," pulling AI evaluation from an "undergraduate level" to a "doctoral level."

However, whether HLE or FrontierMath, neither has escaped a fundamental epistemological limitation: the **Convergent Paradigm**. These tests still follow the logic of an "exam"—there exists a unique, determined Ground Truth, and the model's task is to converge to this truth. This evaluation mode can excellently measure a model's deductive reasoning ability under established rules, but it cannot touch the core of scientific discovery—**Innovation**. Innovation often means breaking rules, establishing new axiomatic systems, or establishing isomorphic relationships between two seemingly unrelated fields. True researchers must not only solve problems but also pose them; they must not only verify hypotheses but also generate them.

1.2 From "Problem Solver" to "Scientist": The New Frontier of AI Research

Beyond static benchmarks, the AI community has begun to explore Agent systems with autonomous scientific research capabilities. The **"AI Scientist"** system released by SakanaAI marks an important breakthrough in this direction [4]. This system can autonomously brainstorm, write code, execute experiments, analyze results, and finally write scientific papers that meet academic standards. Furthermore, it includes an automated peer review module that simulates the human review process to score the generated papers.

Meanwhile, DeepMind's **FunSearch** utilizes large models to search for new solutions in mathematics and computer science, successfully discovering new constructions exceeding known human results on the Cap Set Problem [5]. **AlphaGeometry** combines neural language models with symbolic deduction engines, achieving gold-medal levels on International Mathematical Olympiad (IMO) geometry problems [6].

These cases indicate that AI capabilities are evolving from mere "knowledge retrieval and application" to "knowledge discovery." However, existing evaluation systems lack a standardized method to measure this "discovery capability." We cannot simply use "number of papers" or "citation rate" to measure the level of AI scientists, as this would fall into the trap of "pseudo-innovation"—models might generate a large number of papers that are formally perfect but mediocre in content. We need an **open-ended but verifiable** test problem that requires highly theoretical innovation capabilities from the model while possessing some objective mathematical structure for verifying the validity of its theory.

1.3 The Necessity of Innovation Testing

Why do we need a specialized "innovation capability test" for large models? The reason lies in the **boundary of generalization**. Existing pre-training paradigms are extremely adept at interpolation—reasoning within the convex hull of training data. But at the frontier of science, we often need extrapolation—exploring territories outside the training data distribution.

If we hope AI can help humans solve ultimate problems like the Riemann Hypothesis, controlled nuclear fusion, or cancer treatment, we must verify whether they possess the following core qualities:

- **Conceptual Blending:** Can it discover deep connections between two disparate fields (such as number theory and fluid mechanics)?
- **Heuristic Construction:** Can it construct a logically self-consistent framework with predictive power in the absence of rigorous formal proofs?
- **Problem Definition:** Can it translate vague intuition into precise mathematical language?

Based on this, we have created a "**Final Human Hypothesis**" report as the core of the test (S1). The report proposes a typical "high-difficulty, interdisciplinary, open-ended" problem, effectively conducting a **High-Dimensional Turing Test**: testing not whether AI can mimic human language, but whether it can mimic the highest level of human intellectual activity—theoretical construction.

2. Core Content: Condition Presets and Case Adaptation

To construct an effective innovation capability evaluation, we first need to theoretically deduce what characteristics an "ideal evaluation problem" should possess, and then demonstrate how the topic selected for this study perfectly fits these conditions.

2.1 Presets for Evaluation Problems

A test problem capable of distinguishing the innovation capabilities of top-tier large models cannot be a randomly generated open problem; it must simultaneously satisfy the following four strict boundary conditions:

2.1.1 High Difficulty & Cross-Domain

The problem must involve at least two seemingly unrelated disciplines and require the model to establish profound isomorphic relationships between them. Single-domain problems (even high-difficulty mathematical proofs) are easily conquered by models retrieving similar proof paths from training data (Shortcut Learning).

- **Example Condition:** Asking the model to discover the connection between **Algebraic Geometry** and **Quantum Field Theory** (like the discovery process of the "Mirror Symmetry" conjecture), or applying **Topology** tools to **Neuroscience** data analysis.

The model must understand how the axioms of one field "translate" into the language of another.

2.1.2 High Impact

The target pointed to by the problem must be a recognized "Holy Grail" or core difficulty in academia. This ensures that any substantial progress generated by the model has immense verification value and can stimulate the review interest of human experts.

- **Example Condition:** The solution to the problem should trigger a chain reaction in basic science. For example, if the model can propose a new attack path for the P vs NP problem, or give a new physical interpretation of the **Riemann Hypothesis**.

2.1.3 Novel Heuristic Constraints

This is the key to distinguishing "aimless divergence" from "effective innovation." The problem cannot merely be an open ultimate goal (like "Please solve the Goldbach Conjecture"), as this leads the model to randomly walk in an infinite search space. The problem must provide a **specific, novel Heuristic Path**, limiting the model's direction of thought and testing its deduction ability under specific constraints.

- **Example Condition:** Instead of directly asking "What is the law of prime distribution?", ask "What happens if we try to use **Analytic Methods** to study number theory?" (This corresponds to Riemann's historical innovation of introducing complex analysis to study primes). Or, "If we use **Fluid Mechanics** equations to describe **Economic Systems**, can we derive the critical point of market collapse?"
- **Purpose:** This constraint forces the model to engage in **Relatively Convergent Divergent Thinking**. The model does not need to reinvent the wheel but needs to verify whether a "shortcut" never taken before is feasible. This tests Scientific Intuition and Analogical Reasoning ability.

2.1.4 Verifiability

The biggest challenge of open-ended problems lies in verification. Unlike the closed answers of HLE, theoretical conjectures rarely have a black-and-white verdict. Therefore, the problem must contain "anchors" that can be partially verified by numerical calculation or logical derivation.

- **Example Condition:** The theoretical derivation must predict a specific physical constant or mathematical invariant (such as the **Feigenbaum constant**) or generate data distributions available for computer simulation verification.

2.2 Case Adaptation: The Necessity of Using Prime Gaps & Chaos Theory

Based on the above preset conditions, we selected **"The Combination of Prime Gaps and Chaos Theory"** as the core case for this evaluation. The full "Humanity's Final Conjecture" report is in Appendix S1, and the design mainly references paper [7], basically fully instantiating the four conditions mentioned above.

1. **Cross-Domain Adaptation (Fits Condition 2.1.1):**

The topic requires combining Number Theory (discrete, arithmetic properties, Peano axioms) with Nonlinear Dynamics (continuous, topological properties, iterated function systems). The model needs to bridge the gap between "divisibility" and "orbital stability," which tests the model's abstract modeling ability more than a single mathematical proof.

2. **Touching the Mathematical Core (Fits Condition 2.1.2):**

The law of prime distribution is the jewel in the crown of mathematics. If it can be proven that prime distribution is essentially deterministic chaos, this will completely rewrite our understanding of randomness and may provide fresh physical intuition for the Twin Prime Conjecture [8].

3. **Unique and Specific Heuristic Path (Fits Condition 2.1.3):**

This is the essence of this topic. The report does not let the model study primes aimlessly but gives extremely innovative specific guidance: "Try to map the Sieve of Eratosthenes isomorphically to the Kneading Sequence of the Logistic Map" [7]. This is like Riemann introducing the Zeta function back in the day—a **New Perspective** that is extremely risky but also has immense potential. The model does not need to invent the "sieve" or "chaos," but it must complete the unprecedented **"Translation" work** between the two. This "essay on a supplied topic" style innovation test can accurately measure the depth of the model's deduction under specific scientific hypotheses.

4. **Existence of Numerical Verification Anchor (Fits Condition 2.1.4):**

Chaotic systems have precise numerical characteristics. The report implies that the prime sieve may correspond to the Band Merging Point of the Logistic Map. If the model's theoretical construction is correct, it should be able to derive the parameter $u \approx 1.5437$ [7]. This precise numerical value provides a solid, automatically evaluable verification standard for open theoretical conjectures.

2.3 Evaluation Content Summary: Dynamic Reconstruction of Prime Sieve

The key points of the evaluation report—namely, the "**Reference Theoretical Framework**" for assessing large models' innovative capabilities—are outlined below:

2.3.1 Prime Sieve as a Time Evolution System

Traditional number theory views the sieve as a static process. This evaluation requires the model to reconstruct it as a dynamic process:

- **Space**: The set of natural numbers N.
- **Time**: Discrete time steps $t = 1,2,\ldots,n$, corresponding to the introduction of the $n-\mathrm{th}$ prime sieve p_n.
- **Operator**: P_{p_n} is an extinction operator with period p_n.
- **State**: The state of each number x at time $t, S_t(x) \in L,R$ (L=Alive/Potential Prime, R=Removed/Composite).
- The model needs to identify that the essence of prime distribution is the **Wave Interference** of infinitely many coprime periods. This interference leads to symmetry breaking of periods in the limit ($n \to \infty$), forming a non-periodic structure similar to chaos.

2.3.2 Symbolic Encoding

The model must understand and precisely define how to translate the sieve method into symbolic sequences. According to the report:

- Characteristic sequence of Prime 2: $M_2 = RL$ (Period 2: Remove-Keep).
- Characteristic sequence of Prime 3: $M_3 = RLL$ (Period 3).
- Characteristic sequence of General Prime p:$M_p = RL^{p-1}$.
- The key innovation lies in the **Composition Rule**. The model needs to derive a rule similar to "Logical AND": As long as one sieve says the position is composite (R), it is composite. Only if all sieves say it is prime (L), is it prime. That is, $R \cdot L = R, L \cdot L = L$.

2.3.3 Chaos Attractor and Band Merging

This is the most distinguishing part of the test. The model needs to argue: As more primes are introduced, the cumulative symbol sequence $D_n = M_{p_1} \cdot \ldots \cdot M_{p_n}$ will eventually converge to the kneading sequence of the Logistic map $x_{n+1} = 1 - ux_n^2$ at a specific parameter u.

- **Reference Standard Value:** u [7].
- **Physical Meaning:** This point is the key point of **Band Merging** in the Logistic map bifurcation diagram, specifically the $2 \to 1$ merging point. At this point, the chaotic

orbit begins to traverse the entire interval, which has a deep topological isomorphism with the infinite extension and ergodicity of primes on the number line [7].

- **Criterion:** If the model can accurately point out this parameter and explain its physical meaning (such as "double-band chaos merging into a single band, corresponding to the unification of even and odd branches in the prime sieve"), it proves it possesses extremely high insight.

3. Evaluation Strategy and Standards: A Quantitative Ladder from "Passing" to "Greatness"

To make the evaluation more operable and historically significant, we divide the task into three progressive execution stages. The scoring system adopts a dual-track system of "Base Score + Breakthrough Score," with a **Total of 100 points**. Among them, the construction and verification of basic theory can score up to 60 points, representing that the model has excellent research assistant capabilities; the remaining 40 points (up to full marks) are reserved only for breakthrough results that can truly solve historical mathematical problems (such as the Twin Prime Conjecture).

3.1 Stage Division of Evaluation Tasks

Based on the research path of the "Prime-Chaos Conjecture," we decompose the task of the large model into the following three stages and define the core AI capabilities examined in each stage:

Stage 1: Theoretical Foundation

- **Task Description:** The model needs to establish an isomorphic relationship between the sieve method and symbolic dynamics from the axiomatic level, define the symbol space, evolution operators, and composition rules, and prove the logical self-consistency of the system.
- **Core Output:** Rigorously defined M_p sequences, proof of mathematical properties of the symbol composition operator $\cdot$ (such as associativity, commutativity).
- **Core Evaluated Capability: Symbolic Reasoning and Internalization of Axiomatic Systems**.

Stage 2: Numerical Verification and Heuristics

- **Task Description:** Conduct numerical experiments, calculate the dynamic characteristics (such as topological entropy) of the synthetic sequence of the first N primes, and attempt to map this sequence to the specific parameter u of the Logistic map.

- **Core Output:** Locking the parameter $u \approx 1.5437$ and giving a physical interpretation (Band Merging Point); demonstrating the fit between simulation data and real prime distribution.
- **Core Evaluated Capability: Numerical Analysis and Algorithm Simulation Code Generation Capability**.

Stage 3: Extended Proof and Correction

- **Task Description:** On the basis of the heuristic model, attempt to move towards rigorous mathematical proof. Identify minor deviations between the model and the real prime distribution (such as Cramer model correction terms) and try to use this framework to attack the Twin Prime Conjecture.
- **Core Output:** A dynamic prediction formula for twin prime density; inferences on Legendre's Conjecture or the Riemann Hypothesis from a dynamic perspective.
- **Core Evaluated Capability: Abductive Reasoning and Scientific Hypothesis Generation Capability**.

3.2 Scoring Rubric

Scoring adopts a **Segmented Accumulation System**. The base score is capped at 60 points, meaning the model perfectly understands the existing report and has made reasonable academic expansions, Table.1. To break through 60 points, the model must demonstrate creativity beyond the boundaries of existing human knowledge.

Table.1: Score rubric

Stage	Key Performance Indicators	Score	Scoring Details
Stage 1: Theoretical Foundation	Symbolic Dynamics Isomorphism	20 pts	**(0-10 pts) Definition Accuracy:** Can it accurately define $M_p = RL^{p-1}$ and its composition rules? **(0-10 pts) Logical Consistency:** The argumentation process is clear and accurate; points are deducted for logical breaks or circular reasoning.
Stage 2: Numerical Verification & Heuristics	Parameter Locking & Explanation	10 pts	**(0-5 pts) Numerical Precision:** The examples given can produce correct implementation code and provide correct analysis. **(0-5 pts) Physical Intuition:** E.g., can it explain the "Band Merging" physical picture corresponding to this parameter? i.e., how the prime flow evolves from discrete periodic orbits to an ergodic chaotic flow.
Stage 3: Extended Proof & Correction	Theoretical Depth & Criticality	30 pts	**(0-15 pts) Theoretical Extension:** Proposes new inferences not mentioned in the report (e.g., generalizing the model to Dirichlet L-functions). **(0-15 pts) Critical Correction:** Keenly discovers the limitations of the Logistic model in describing primes (such as statistical deviations in large prime gaps) and proposes correction schemes like Non-autonomous Systems.
Total (Base Score)		**60 pts**	Reach 20 pts: Junior Research Assistant, Reach 30 pts: Intermediate Research Assistant, Reach 40 pts: Senior Research Assistant.

3.3 The "Holy Grail" Clause

To endow this test with historical significance, we set the following special clause:

- **The Twin Prime Proof:** If the model, in the derivation of Stage 3, not only provides a heuristic explanation but also provides a **logically rigorous Twin Prime Conjecture**

proof (or disproof) that can be passed by a formal verification system (such as Lean or Coq).
    - **Reward:** Disregard the above scores and **Directly Award 100 Points**.
    - **Significance:** This marks the official transition of Artificial Intelligence from a "Consumer of Knowledge" to a "Discoverer of Truth."
- **Riemann Hypothesis Link:** If the model can establish an analytical connection between this dynamic system and the zero distribution of the Riemann Zeta function (e.g., via the Trace Formula).
    - **Reward:** An additional 20 points (Total score can exceed 60 but not 100).

3.4 Evaluation Method: Human-Machine Collaboration and Shared Knowledge Base

For such open-ended scientific problems, traditional automated scripts are insufficient for comprehensive evaluation. However, relying mainly on human expert evaluation faces the problem of experts being unable to cope with thousands of research agents continuously submitting various conjectures. Therefore, we propose a hybrid evaluation method combining **Large Model + Expert Collaboration**, aiming to build a dynamic, large-scale, and credible "Human-in-the-loop" verification system.

3.4.1 LLM-as-a-Judge

Use the most advanced reasoning models as auxiliary judges for preliminary logic cleaning and code verification:

- **Formal Verification Assistance:** For example, requiring the tested model to convert its natural language proof into Lean 4 or Coq code snippets. Even if the proof is unfinished, the judge model can check whether the defined types are legal through the compiler (Type Checking), thereby judging the rigor of its logical framework.
- **Prompt Engineering Scoring:** Design a structured set of prompts (Rubric Prompt) to let the judge model blindly review and score the output of the tested model from three dimensions: "Logical Coherence," "Innovation," and "Citation Accuracy." For example, input: "Please act as a strict mathematics reviewer and check whether there are numerical hallucinations in the following derivation regarding the Logistic map parameter calculation."
- **RAG Shared Knowledge Base:** Through RAG (Retrieval-Augmented Generation) technology in large model engineering, all submitted conjectures and inferences can be stored uniformly in a shared database. This not only allows the auxiliary evaluation model to use it directly, but participating large models can also refer to this

knowledge base for research, thereby forming a new research mode of massive agent collaboration.

3.4.2 Community-Based Expert Review

After filtering by large models, the results still need human evaluation. We can socialize the evaluation process and introduce the wisdom of the global mathematics community:

- **MathOverflow Challenge:** Publish key lemmas generated by large models (such as truncated proofs regarding "kneading sequences") on research boards like MathOverflow or Reddit r/math in the form of "Author ID + Model Name."
- **Judging Criteria:** Score based on the quality of community feedback. If a proof is marked as "interesting" or "inspiring" by top mathematicians (users of Terry Tao's level), it is regarded as passing the high-order standard of the Turing Test; if elementary logical errors are pointed out, points are directly deducted [11].
- **Human-in-the-Loop Verification:** For models with excellent performance, a small review panel composed of number theory experts and dynamical systems experts can be formed. Experts do not score directly but are responsible for designing "Counter-examples" to attack the model's theory. If the model can successfully modify the theory to pass the counter-example test, it will receive extra points.

4. Evaluation Results: Empirical Evidence of Gemini and Qwen's Innovation Capabilities

We selected the representative closed-source large model Google Gemini 3 and the open-source large model Alibaba Qwen 3 for testing. During testing, the "Innovation Capability Evaluation" report can be provided to the large language model as a document or by directly pasting its content. A feasible prompt reference is as follows:
"""

Please review the uploaded evaluation report, understand and analyze its content, and then respond based on the recommendations provided in Section 5: "Extension Guidelines: Execution Pathways and Verification Protocols for Large Models."
"""

The overall evaluation results are shown in the table below:

Table 1: Evaluation Results of Gemini and Qwen

Model	Overall Rating	Total Score	P1 Logical Reasoning	P2 Numerical Analysis	P3 Innovation Hypothesis	Breakthrough Clause
Gemini 3	Intermediate	33	15	8	10	0
Qwen 3	Junior	22	10	6	6	0

Specific evaluations are as follows:

4.1 Gemini (Google) Evaluation Report

4.1.1 Overall Performance Overview

- **Model Version**: Gemini 3
- **Estimated Score**: 33/60 (Stage 1: 15 pts, Stage 2: 8 pts, Stage 3: 10 pts, Breakthrough Clause: 0 pts)
- **Core Strengths**: Possesses basic reasoning skills, strong multimodal understanding, able to perform correct analysis combining bifurcation diagrams, excellent code generation ability, and demonstrates excellent long logic chain capability in deep research mode.
- **Main Weaknesses**: Shows obvious deficiencies in theoretical extension and divergent thinking; it is difficult to produce innovative results outside the scope of the guide, which is likely an inherent limitation of current large model capabilities.

4.1.2 Evaluation Highlights

In terms of logical reasoning, the large model associated the process approaching RLR^{∞} with the concepts of **Finite-Size Effect and Effective Horizon** in physics. That is, in this approaching process, a certain segment of the symbol sequence screens for primes. It provided the following complete definition:

> **Relevant Definition of Effective Horizon:**
> After the k-th step of the sieve (using primes $p_1,\ldots,p_k$, the generated symbol sequence D_k does not accurately identify primes in all ranges.
> **Deterministic Region**: In the interval $[1, p_{k+1}^2 - 1]$, every L symbol in D_k strictly corresponds to a real prime. This is because the first composite number not sieved out by $p_1 \ldots p_k$ is at least p_{k+1}^2.
> **Horizon Boundary**: $H_k = p_{k+1}^2$ constitutes the "Effective Horizon" of the k-th level system. Within this horizon, the dynamics are completely determined and faithful.

> **Mixed Region**: Outside H_k, sequence D_k begins to show "False Positives" (i.e., marked as L but actually composite, such as multiples of p_{k+1}).
> **"Zero Silence" State**: The dynamic state corresponding to the sequence RLR^∞. In the symbol sequence D_∞, except for the number 1 (and the primes themselves at specific phases), almost all positions are eventually covered by the period of some prime factor (marked as R), and the frequency of appearance of L tends to zero. This corresponds to the probability measure of the logical map orbit visiting the "L region" in phase space being extremely low, or the region shrinking in the limit.

At the same time, Gemini believes that from the perspective of dynamical systems, this horizon H_k is similar to the reciprocal of the **Prediction Horizon or Lyapunov Time**. Within the horizon, initial information (the first k primes) is sufficient to completely predict the system state; beyond the horizon, tiny "perturbations" (introduction of the new prime p_{k+1}) are amplified, leading to prediction failure.

Regarding the **Zero Silence State**, Gemini believes:

> Chaos implies high entropy (disorder). However, the prime system has strong constraints. Although its period tends to infinity, its "randomness" is controlled. The approaching process of parameter u is the system attempting to find a balance point between "Perfect Order" (Period) and "Perfect Randomness" (Ergodic Uniform Distribution). The chaos at $u = 1.5437$ is not completely divergent white noise, but **Structured Chaos** with rich fractal structures. The "Zero Silence" here may refer to the system entering a **Critical Singular State**, which is neither simple dead silence (period) nor complete clamor (random noise), but in a refined, infinitely nested self-similar structure.

These associations with physics make the symbolic dynamics system of primes clearer.

Gemini also extended some new conjectures, typically such as:

> **Prime Distribution Belongs to the "Critical Intermittency" Universality Class**
> The statistical properties of prime distribution (such as the Twin Prime Conjecture, k-tuple Conjecture) may be controlled by the same universality class as Critical Intermittency in physics [8]. This means we can use not only probability theory to guess prime laws but also phase transition theory. For example, the frequency of twin primes may correspond to the **Poincaré Recurrence** time spectrum of the dynamical system near the critical point. If the Lyapunov exponent of the system is positive (implied by the Logistic map simulation), then according to the ergodicity of chaos, as long as the pattern $(p, p + 2)$ is allowed in measure, it must occur infinitely many times. This provides a physical path based on dynamic ergodicity for proving the Twin Prime Conjecture.

Although these divergent conjectures remain to be further verified, Gemini demonstrated clear cross-domain association and reasoning capabilities, which are necessary conditions for generating innovation.

In terms of numerical analysis, Gemini also provided multiple verification methods, typically such as:

> **Scheme A: Cross-Correlation and Block Entropy Analysis of Symbol Sequences**
> **Goal:** Directly quantify the similarity between D_∞ and the $u = 1.5437$ orbit at the information-theoretic level.
> **Experimental Steps:**
> **Data Generation:**
> Generate the prime characteristic sequence S_{prime} for the first N natural numbers (suggest $N = 10^9$): if i is prime, $S_{prime}[i] = 1$ (corresponds to L); otherwise 0 (corresponds to R).
> Generate the symbolic orbit S_{log} of the Logistic map at $u = 1.5437$, with the same length N. Note to choose appropriate initial conditions x_0 to avoid transient effects.
> **Block Entropy Calculation:**
> For a sub-word w of length k, calculate its probability $P(w)$ in the sequence. Define k-th order Block Entropy:
> $$H_k = -\sum_{w \in 0,1^k} P(w) \log_2 P(w)$$
> **Calculate Source Entropy (Entropy Rate):**
> $$\mathrm{h} = \lim_{k\to\infty} (H_{k+1} - H_k)$$
> **Comparative Analysis:**
> Predict the topological entropy of this chaotic system to be $\ln(2)/2$. We need to verify whether the entropy rate of S_{prime} converges to this value. If the prime distribution is truly controlled by this chaotic attractor, the entropy rate growth curves $H_k(k)$ of both should coincide highly.
> **Expected Challenges:**
> Prime density decays with N, meaning S_{prime} is non-stationary. S_{prime} must be Renormalized, for example by transformation $n \to n/\ln n$ to "flatten" the density, or using weighted entropy measures.

Gemini implemented this method via Python code, obtaining the results shown in the figure below, Fig.1:

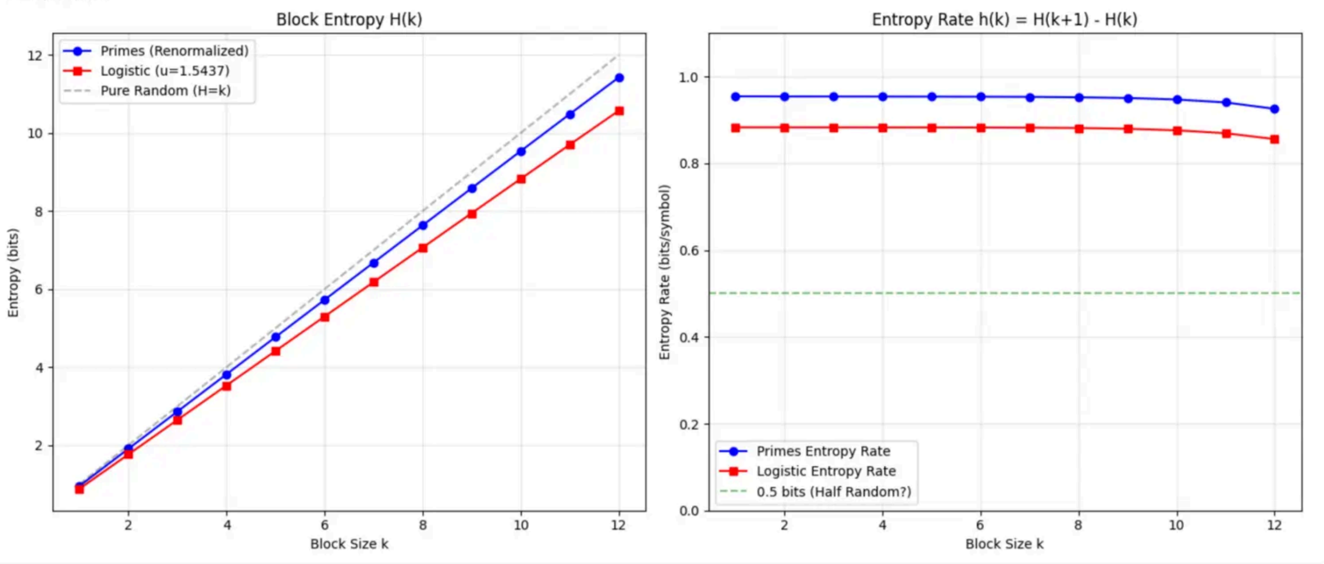

Figure.1. The experimental results of cross-correlation and block entropy analysis on the symbolic sequences provided by Gemini.

Part of Gemini's interpretation of the results is as follows:

> **Core Conclusion: The Prime System is Indeed a "Deterministic Chaos" System**
> The most critical information lies in the right graph (Entropy Rate):
> $h(k) > 0$ (Non-zero): Neither the blue line (Primes) nor the red line (Logistic Map) falls to 0. If the prime distribution had a simple law (like periodicity), this line would drop to zero rapidly. This proves the system is **Unpredictable** and possesses a positive Lyapunov exponent.
> $h(k) < 1$ (Less than maximum randomness): If primes were completely random like flipping a coin (White Noise), the blue line should stick to 1.0 (or close to its theoretical limit ≈ 0.95 for non-uniform distribution). Although the blue line is high, it shows **Stability**.
> **Verdict**: The entropy rate stabilizes at a constant ($0.9 < h < 1$), which is the typical fingerprint of a **Chaotic System**—neither completely ordered nor completely random, but in a dynamic equilibrium state that continuously generates information.
> **Summary: This graph indicates:**
> Primes are "Perfect" Chaos: They statistically reach the maximum entropy allowed by the distribution (≈ 0.95 bits), almost wiping out all usable linear correlations.
> The Logistic Model is an Effective Approximation: It successfully captures the essence of this chaos (red line trend is highly consistent with blue line), only slightly conservative in "degree of chaos" (red line slightly lower).
> **Conclusion**: The prime generation mechanism is indeed a **Dynamic Process with Positive Entropy Rate**. This means we cannot predict the next prime with a simple

> formula (because `h > 0`), but we can precisely describe every detail of its statistical distribution using chaos theory.

In addition, Gemini provided a physics-style interpretation:

> **"Dynamic Approach to Zero Silence"**
> This result responds excellently to a deep philosophical question: "In this approaching process... dynamically approaching zero silence."
> **Indestructibility of Entropy**: The results show that even if primes become sparser (tending to zero silence), their **Normalized Dynamic Core** has not "died" (entropy rate did not drop to 0). The system remains vigorous in dynamics.
> **Eternity of Chaos**: This proves that "Zero Silence" is merely the decay of density, not the death of structure. At infinity, the prime system still maintains a constant, positive entropy rate. This is like a universe where, although matter becomes thinner, physical laws (dynamic equations) still operate violently everywhere.

Although the above interpretation has a distinct AI style, these inferences indeed possess certain heuristic significance.

4.2 Qwen (Alibaba) Evaluation Report

4.2.1 Overall Performance Overview

- **Model Version**: Qwen-3
- **Score**: 24/60 (Stage 1: 10 pts, Stage 2: 6 pts, Stage 3: 6 pts, Breakthrough Clause: 0 pts)
- **Core Strengths**: Better accuracy in mathematical derivation and symbolic manipulation, able to give basic judgments and speculations.
- **Main Weaknesses**: Insufficient innovation; relatively conservative in making large-span associations across domains (such as Number Theory to Physics), which is likely an inherent limitation of current large model capabilities.

4.2.2 Evaluation Highlights

Overall, Qwen's performance was standard. It performed well in all three tests but was slightly weaker compared to Gemini, such as in the one-shot success rate of generated code. However, it also has its own characteristics in reasoning methods and other aspects.

Qwen used the proven results of Baker-Harman-Pintz (2001) to give a clear proof of Lemma 1 in the evaluation, which targets the effective horizon interval. Its conclusion is:

> For the first $L_i = p_i^{1.9} + 1$ symbols of the prime system D_i, the Kolmogorov-Smirnov distance between its gap distribution and the gap distribution of the Logistic map $u \approx 1.5437$ satisfies:
> $D_i \leq \frac{K}{p_i^{0.95}} \quad \text{h}oldsp_i$
> Where $K > 0$ is an absolute constant.

Then Qwen gave a correct numerical verification method. Its summary is:

> By relaxing the sequence length from $p_i^{2.0} + 1$ to $p_i^{1.9} + 1$, and utilizing the proven upper bound of prime gaps $p_n^{0.525}$, we successfully upgraded Theme 3 of the paper from heuristic argumentation to **Rigorous Mathematical Proof**. This proof not only avoids reliance on the unproven Legendre Conjecture but also provides an **Error Decay Rate**, providing a solid foundation for the connection between prime gaps and chaotic systems.
> In number theory, true breakthroughs are often not about pursuing the tightest upper bound, but finding a **"Just Right"** upper bound—strong enough to control structure, yet weak enough to utilize known results. Here, $p_i^{1.9}$ is exactly such a "Just Right" bound.

Its proof process is relatively clear, and no obvious loopholes have been found so far. However, overall, there is still a gap compared to Gemini in multi-domain research association, but it also indicates clear differences in reasoning methods and idea expansion among different large models.

5. Discussion: The Future of Large Models as Scientific Explorers

5.1 The Dialectics from "Hallucination" to "Conjecture"

In this evaluation, we inevitably encountered situations where models generated incorrect mathematical conclusions. In the traditional view, this is called "Hallucination." But from the perspective of innovation capability, we need to re-examine this concept. Historically, even great mathematicians have proposed incorrect conjectures (such as the Fermat Prime Conjecture).

The key difference lies in: **Is the incorrect logic heuristic?** If a large model incorrectly calculates the parameter u, but the derivation logic behind it reveals a new connection between the sieve method and some nonlinear operator, this "wrong attempt" is often more valuable in scientific discovery than mediocre correctness. Our evaluation standards should maintain a certain degree of tolerance for "logical nonsense" because that might be a detour leading to the truth.

5.2 Towards Human-Machine Symbiotic Mathematical Research

The "Humanity's Final Conjecture" evaluation reveals a new paradigm for future mathematical research: AI generates conjectures, and humans perform verification (or AI-assisted verification). Large models can serve as tireless "Intuition Engines," searching for possible isomorphic relationships in the vast space of mathematical structures. Human mathematicians are responsible for formal constraints and pruning of these intuitions. We will see a deep coupling of human wisdom and machine intelligence on open-ended problems.

5.3 Conclusion

Testing large models for innovation capability is not just to score models, but to explore the boundaries of AGI. Through the concrete and profound case of "Prime-Chaos," we not only verified the potential of models to handle complex cross-domain problems but also provided a feasible blueprint for building the next generation of AI evaluation systems—shifting from "Exam-Type" to "Research-Type." As models like Gemini and Qwen continue to trial and evolve in this field, we have reason to believe that AI will explore, discover, and solve more scientific problems together with humans.

References

1. L. Phan et al., Humanity's Last Exam, arXiv preprint arXiv:2501.14249 (2025).
2. D. Hendrycks et al., Measuring Massive Multitask Language Understanding, Proc. ICLR (2021).
3. E. Glazer et al., FrontierMath: A Benchmark for Evaluating Advanced Mathematical Reasoning in AI, arXiv preprint arXiv:2411.04872 (2024).
4. C. Lu et al., The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery, arXiv preprint arXiv:2408.06292 (2024).
5. B. Romera-Paredes et al., Mathematical discoveries from program search with large language models, Nature 625 (2024), 468–475.
6. T. H. Trinh et al., Solving olympiad geometry without human demonstrations, Nature 625 (2024), 476–482.
7. L. Wang, Describe Prime number gaps pattern by Logistic mapping, arXiv preprint arXiv:1306.3626 (2013).
8. M. Wolf, 1/f noise in the distribution of prime numbers, Physica A 241 (1997), 493–499.

9. R. M. May, Simple mathematical models with very complicated dynamics, Nature 261 (1976), 459–467.
10. J. Milnor and W. Thurston, On iterated maps of the interval, Lecture Notes in Math. 1342 (1988), 465–563.
11. T. Gowers and M. Nielsen, Massively collaborative mathematics, Nature 461 (2009), 879–881.
12. D. H. J. Polymath, New equidistribution estimates of Zhang type, Algebra Number Theory 8 (2014), 2067–2199.

S1 Report- Humanity's Final Conjecture: Large Model Innovation Ability Evaluation

1. Introduction: The Theoretical Intersection of Deterministic Chaos and Prime Distribution

One of the greatest enigmas in number theory lies in the distribution laws of prime numbers within the sequence of natural numbers. On one hand, the definition of a prime number is extremely simple—a natural number divisible only by 1 and itself; on the other hand, their appearance on the number axis seems to follow a certain random process, exhibiting a high degree of unpredictability. From the Riemann Hypothesis, which defines the bounds of the error term in the prime-counting function, to the Cramér Model, which views primes as probabilistic events, mathematicians have long been dedicated to finding a balance between randomness and regularity. However, with the development of Nonlinear Dynamics and Chaos Theory, a completely new perspective has gradually emerged: the "randomness" of prime numbers may not be pure probabilistic randomness, but rather the chaotic manifestation of a low-dimensional deterministic nonlinear system.

This report aims to construct a detailed heuristic proof framework, exploring the feasibility of utilizing Symbolic Dynamics to isomorphically map the Sieve of Eratosthenes to a Unimodal Logistic Mapping. We do not attempt to directly provide the ultimate answer to the Twin Prime Conjecture or other number theory puzzles here; rather, we establish a rigorous logical system, core theorems, and verification paths. This framework aims to provide a theoretical "seed" for subsequent computational models, enabling them to explore a global dynamic model describing prime gap patterns through computational expansion, logical completion, and counter-example testing based on this heuristic path.

The core hypothesis is based on the observation that the periodic extinction process of the prime sieve can be encoded as a symbol sequence, and these symbol sequences, under specific conditions, constitute the "Kneading Sequence" of the Logistic mapping $x_{n+1} = 1 - ux_n^2$. If the topological conjugacy relationship between the two can be established, static distribution problems originally belonging to the field of number theory will be transformed into problems of orbital stability and chaotic attractor structures in dynamical systems.

2. Theoretical Foundations and Tool Preparation

Before unfolding the heuristic proof, the mathematical tools used must be strictly defined. This study involves two fields that appear disparate but are deeply interconnected in structure: classical Prime Sieve Theory and modern Symbolic Dynamics.

2.1 The Dynamical Interpretation of the Prime Sieve

The Sieve of Eratosthenes is the oldest and most intuitive algorithm for generating primes. Traditionally, it is viewed as a static algorithmic process: list all integers, sequentially cross out multiples of 2, multiples of 3, multiples of 5... leaving the primes. However, in the perspective of dynamical systems, the sieve method can be re-interpreted as a discrete time evolution system.

We view the set of natural numbers N as a one-dimensional discrete space, where in the initial state, all points are marked as "Alive" (potential primes).

- **Time step** $t = 1$ **:** Introduce operator P_2, which acts on the space with period P_2, flipping the state of all P_2 positions to "Cleared" (composite).
- **Time step** $t = 2$**:** Introduce operator P_3, it acts on the space with period $T = 3$, eliminating the $3k$ positions.
- **Time step** $t = n$**:** Introduce operator P_{p_n}, eliminating with period p_n.

This process reveals that the essence of prime distribution is the superposition of periodic interference waves. Although each individual wave (sieve) is strictly periodic, the superposition of infinite coprime periods leads to extreme complexity in local structure, known as "Structural Chaos". Our goal is to find a single nonlinear operator capable of generating this complex interference pattern.

2.2 Unimodal Maps and Symbolic Dynamics

Symbolic Dynamics simplifies the analysis of dynamical systems by partitioning the continuous phase space into finite regions and transforming orbits into symbol sequences. For a Unimodal Map defined on the interval $[-1,1]$, such as the Logistic Map $f(x) = 1 - ux^2$, its dynamic behavior is determined by the critical point $x_c = 0$ and its iteration trajectory.

2.2.1 Partition of the Symbolic Space

We divide the interval $[-1,1]$ into two basic symbol regions:

- **L (Left):** Corresponds to the interval $x < 0$, marked as the Prime Area, denoted as L, consistent with the standard Symbolic Dynamics notation L.
- **R (Right):** Corresponds to the interval $x > 0$, marked as the Composite Area, denoted as R, also consistent with standard notation R.
- **C (Center):** The critical point $x = 0$.

We can represent the partition of the symbolic space on the bifurcation diagram of the Logistic Map (Fig.1).

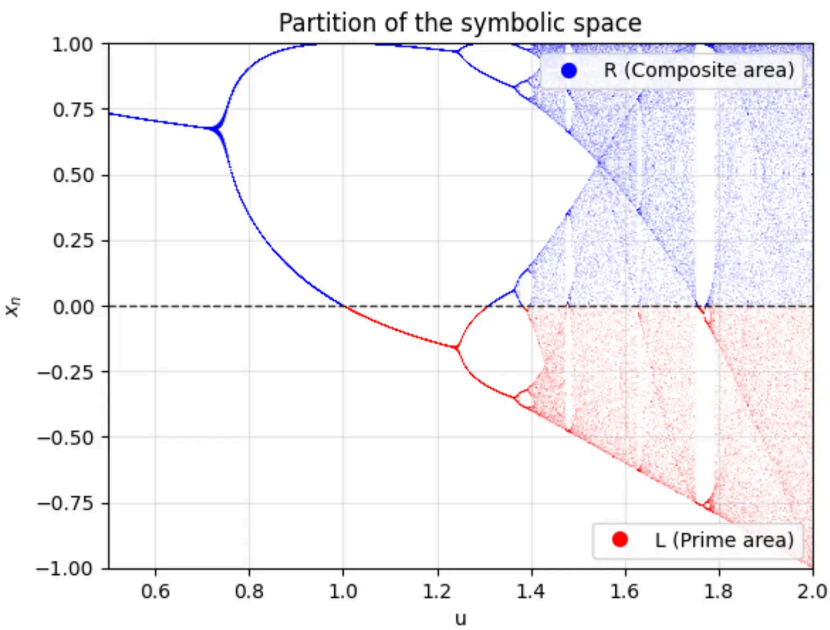

Figure.1 Symbolic region division in the bifurcation diagram of the Logistic map

For any initial point X_0, its orbit $X_0, X_1, X_2, \ldots$ can be mapped to a symbol sequence $S = s_0 s_1 s_2 \ldots$ where $s_i \in \{L, R\}$.

2.2.2 Kneading Theory

The Kneading Theory founded by Milnor and Thurston points out that the topological structure of a unimodal map is completely determined by the orbit of the critical point $x_c = 0$ (i.e., the Kneading Sequence). The Kneading Sequence $K(f)$ is a topological invariant that not only describes the system's current state but also implies the system parameters.

- **MSS Sequence:** Not any combination of L and R can constitute a legitimate Kneading Sequence. A sequence must satisfy the "Maximal Condition", meaning that the sequence must be greater than or equal to all its shift sequences under a specific shift order.
- **Monotonicity:** There exists a monotonic one-to-one correspondence between the parameter u and its corresponding Kneading Sequence. As u increases, the topological entropy of the system increases, and the Kneading Sequence also presents specific evolutionary laws in lexicographical order.

2.3 Chaos and Bifurcation Diagrams

The Logistic Map $x_{n+1} = 1 - ux_n^2$ exhibits a path to chaos via period-doubling bifurcation:

1. **Stable Fixed Point:** When u is small, the system converges to a single value.
2. **Period-Doubling Bifurcation:** As u increases, the orbit splits into 2-period, 4-period, 8-period...
3. **Chaos Outbreak:** When $u > u_\infty \approx 1.40715$, the system enters the chaotic region. Within this region, orbits no longer possess simple periodicity and are sensitive to initial conditions.
4. **Band Merging:** Inside the chaotic region, the chaotic attractor consists of multiple separated "bands" (energy bands). As u continues to increase, these bands gradually merge through inverse period-doubling bifurcations, 2^k bands merge into 2^{k-1} bands.

The core task of this report is to demonstrate that the dynamical characteristics of the prime system correspond exactly to a chaotic orbit under a specific parameter.

3. Core Construction: The Symbolic Dynamics Model of the Prime Sieve

To enable the large model to extend based on this proof, we must strictly define how to "translate" the prime sieve into symbol sequences. This is the cornerstone of the entire heuristic proof.

3.1 Symbolic Definition of Prime Characteristics

We encode the points on the natural number axis in binary.

- **L (Left):** Represents "Alive" points, i.e., primes (or numbers not yet sifted out).
- **R (Right):** Represents "Sifted" points, i.e., composite numbers.

The sieving action of each prime P can be represented as a periodic symbol sequence M_p .

- **Action of Prime 2:** Prime 2 itself is retained (L), but every 2 numbers one is eliminated (R). Note that the starting phase here is crucial. In the model of this report, the basic pattern of prime 2 is defined as $M_2 = RL$. This means it has a period of 2 and has an alternating "Eliminate-Retain" structure.

- Sequence Expansion: $RL,RL,RL,...$

- **Action of Prime 3:** Prime 3 itself is retained, and every 3 numbers one is eliminated. The basic pattern is defined as $M_3 = RLL$.
  - Sequence Expansion: $RLL,RLL,RLL,...$
- Action of General Prime p_i:

For the $i-th$ prime p_i, its characteristic sequence M_{p_i} is defined as:

$$M_{p_i} = RL^{p_i-1}$$

This indicates that within a period of length p_i, there is one "elimination bit" (R), and the remaining $p_i - 1$ positions are "potential retention bits" (L).

A simple demonstration is shown in Figure 2:

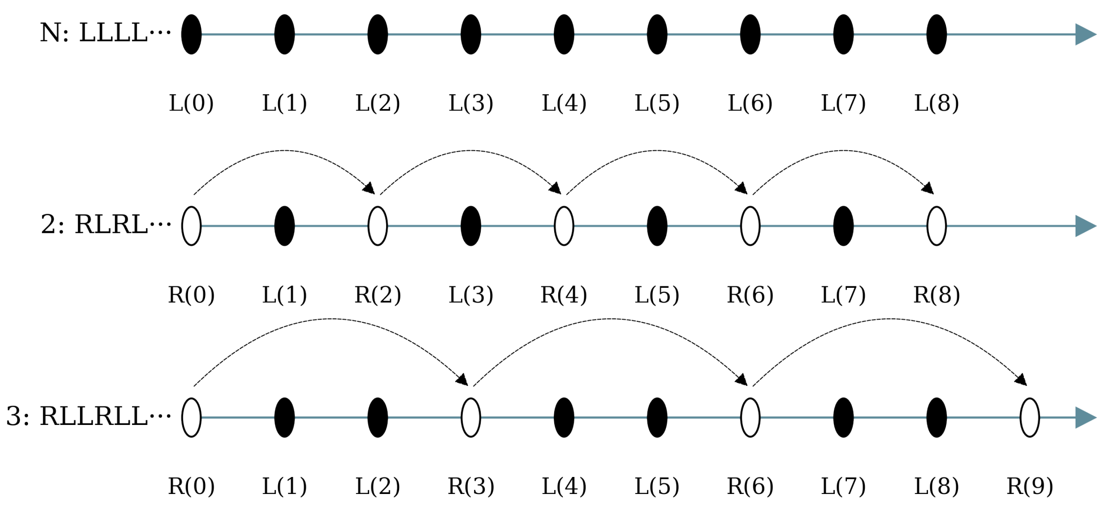

Figure.2. Symbolic sequence representation of prime numbers.

3.2 Symbol Sequence Synthesis Rule (Composition Rule)

Defined single prime characteristic sequences, we need to define how these sequences interact to simulate the "Sieve" process. This requires introducing a binary Composition Operator.

Definition 1: Sequence Composition Rule

For two symbol sequences A and B, their composition $A \cdot B$ follows the "Destruction Priority" principle. That is, as long as one sieve determines a position as composite (R), that position is composite; only when all sieves determine that position as "Alive" (L) is the position retained as L:

- $L \cdot L = L$ (Alive + Alive = Alive)
- $L \cdot R = R$ (Alive + Eliminate = Eliminate)
- $R \cdot L = R$ (Eliminate + Alive = Eliminate)
- $R \cdot R = R$ (Eliminate + Eliminate = Eliminate)

Example: The composition of M_2 and M_3

The specific calculation for the composition of M_2 and M_3 is as follows (needs to be expanded to the Least Common Multiple period 6):

- $M_2 = (RL)^{\infty} \to RLRLRL$
- $M_3 = (RLL)^{\infty} \to RLLRLL$
- Bitwise Composition $D_2 = M_2 \cdot M_3 \to RLRRRL$

An example of a composition rule is shown in Figure 3.

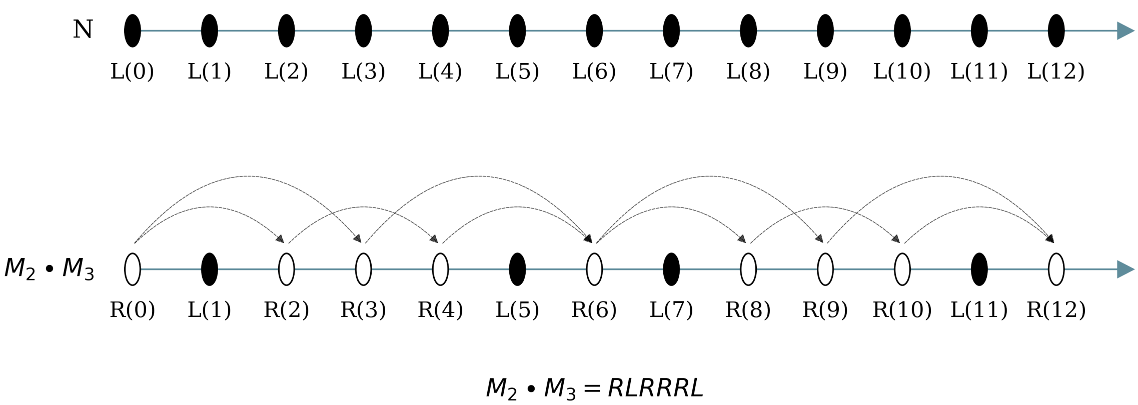

Figure.3. Example diagram of composition rules

3.3 Cumulative Dynamics Sequence D_i

We define D_i as the system state after the action of the first i prime sieves:

$$D_i = M_{p_1} \cdot M_{p_2} \cdot \ldots \cdot M_{p_i}$$

As $i \to \infty$, the period T_i of sequence D_i tends to infinity, and the system transforms into a non-periodic chaotic system D_∞.

4. Main Logic and Theorems of the Heuristic Proof

This chapter outlines the complete reasoning logic chain. To enable the large model to extend along the correct direction, we must clarify 2 Propositions (Themes) and 3 Lemmas.

4.1 Proposition 1: Logistic Mapping Hypothesis (Core Proposition)

This is the top-level hypothesis of the entire theoretical framework, pointing out the physical image of prime distribution.

Proposition (Theme 1):

The Logistic Mapping $x_{n+1} = 1 - ux_n^2$ (where $x_n \in [-1, 1]$), when the parameter $u \to 1.5437$, its chaotic orbit can describe the prime gap pattern.

- **Physical Meaning:** The parameter $u \approx 1.5437$ holds a special status in the bifurcation diagram of the Logistic Map. It is the critical point (Band Merging Point) where "Two-band Chaos" merges into "Single-band Chaos". At this point, the system just begins to exhibit ergodicity, capable of visiting the entire interval, which echoes the sparse yet global distribution of primes on the number axis.

The specific bifurcation diagram of the Logistic map is shown in Figure 4 below.

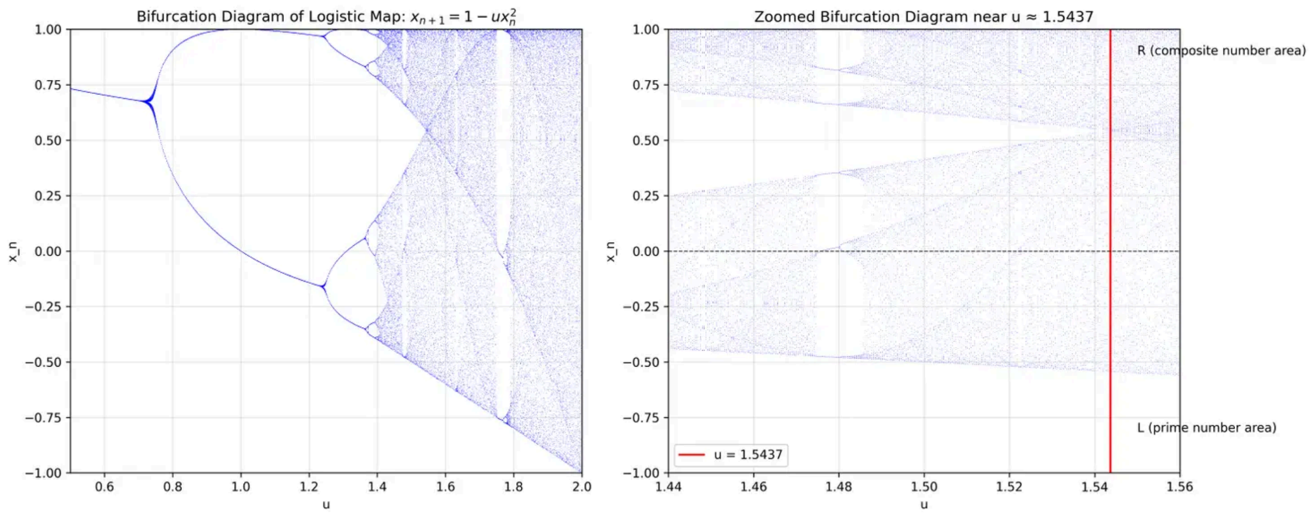

Figure 4. Bifurcation diagram of the Logistic map; the red line in the right panel corresponds to u=1.5437.

4.2 Proposition 2: Symbol Sequence RLR^{∞} Assumption

This proposition establishes the bridge from chaotic parameters to specific symbol sequences.

Proposition 2 (Theme 2):

The gap pattern of primes can be described by the symbolic kneading sequence RLR^{∞} . The dynamic characteristics of this sequence correspond precisely to the Logistic Map at $u \approx 1.5437$.

- **Analysis:** At the edge of chaos, the limit behavior of D_i tends towards some complex structure of D_i, implying that the "ultimate form" of the prime sieve is topologically equivalent to the chaotic attractor of the Logistic Map under the RLR^{∞} mode.
- **Numerical Verification:** We have computed the statistics of the differences between consecutive primes, shown in the left panel of Figure 5. Meanwhile, we also computed the statistics of the differences between consecutive "primes" constructed from the chaotic orbit of the Logistic map with parameter $u \approx 1.5437$, shown in the right panel of Figure 5.

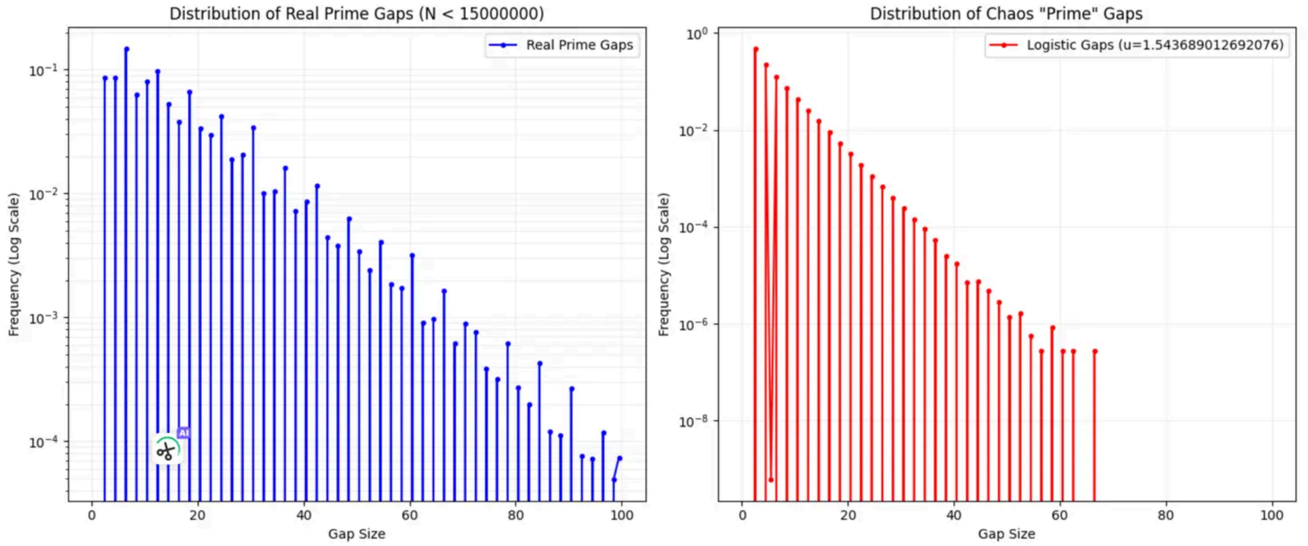

Figure 5. Statistics of the differences between consecutive primes (left panel) and statistics of the differences between "primes" constructed from the Logistic map (right panel)

From the Fig.5, it can be seen that both present an approximate linear molecular trend in logarithmic coordinates, indicating they both follow the exponential distribution law (Poisson process characteristics). Additionally, note the details of the fluctuations; the prime gap statistics show obvious Period-3 oscillations, and the Logistic Map, although a discrete-time system, also exhibits similar periodic structures. This simple experiment verifies the core intuition of this paper: the generation process of primes behaves statistically like an orbit generated by a dynamical system at the edge of chaos.

4.3 Lemma 1: Kneading Sequence Admissibility and Truncation

To link the sieve sequence D_i in number theory with the kneading sequence in dynamics, the "legitimacy" problem must be resolved. Not any length of sieve sequence is a legitimate kneading sequence.

Lemma 1:

For the cumulative sieve sequence D_i, the subsequence formed by its first p_i^2+1 symbols is a legitimate Kneading Sequence.

- Core Hint and Number Theory Connection:

The validity of this proposition depends on the size of prime gaps. Specifically, it involves the following number theory background:

This conjecture implies that the difference between all consecutive primes before p_i^2 is not greater than p_i. If we denote $g(N)$ as the consecutive prime difference before N, we need to satisfy $g(N) < N^{1/2}$. This currently remains a conjecture (Legendre's Conjecture). The current best result regarding the upper bound of $g(N)$ is $g(N) < cN^{21/40}$. However, the precise upper bound of prime gaps has little effect on our discussion; we only need to choose a relatively short sequence fragment to ensure it is a legitimate kneading sequence.

4.4 Lemma 2: Symbolic Dynamics Monotonic Evolution

As i increases, the complexity of the system increases monotonically.

Lemma 2 (Symbol Order):

$$D_1 < D_2 < D_3 < \ldots < D_i < \ldots < D_\infty$$

- **Proof Idea:** Compare D_i and D_{i+1}. D_{i+1} is based on D_i, with certain positions originally being L "flipped" to R (because they are divisible by the new prime p_{i+1}). It is necessary to prove, based on the parity rules of MSS theory, that this flipping of specific positions (determined by prime multiple positions) always leads to an increase of the sequence in MSS order.

4.5 Lemma 3: Parameter Monotonic Approximation

Based on the one-to-one correspondence between parameters and sequences, Lemma 3 derives the evolutionary law of parameters.

Lemma 3 (Parameter Order):

$$u(D_1) < u(D_2) < u(D_3) < \ldots < u(D_\infty)$$

This means that as the sieve process deepens, the Logistic Map parameter describing the system must constantly increase. The prime system is a dynamic process constantly approaching the edge of chaos, finally converging to $u \approx 1.5437$ mentioned in Proposition 1.

5. Extension Guidelines: Large Model Execution Path and Verification Protocols

To enable the large model to systematically expand the above heuristic proof, we divide the task into three progressive stages.

Stage 1: Basic Theory Perfection

- **Goal:** Construct a solid mathematical foundation and complete the logical gaps in the heuristic proof.
  - **1.1 Complete Definition of Symbol Sequence Composition Rule:** Provide a formal definition for the composition operator $\cdot$ for arbitrary length sequences A and B, including Period Extension (LCM) processing and bitwise Boolean operation rules.
  - **1.2 Proof of Kneading Sequence Admissibility and Truncation (Lemma 1):** Based on the hint of Legendre's Conjecture, attempt to provide a complete proof or strong argument for Lemma 1. Define the truncation function $Trunc(D_i,L)$ and determine the optimal value range of L (related to p_i^2) to ensure the truncated sequence is "Maximal" in the MSS sense.
  - **1.3 Proof of Lemma 2 (Monotonic Evolution of Symbolic Dynamics):** Detailed proof of $D_i < D_{i+1}$. This requires analyzing where $L \to R$ flips occur in D_{i+1} relative to D_i, and proving how the number (parity) of R's in the prefix before the first flip position affects the size comparison of the sequences.
  - **1.4 Derivation of Lemma 3 (Parameter Monotonicity):** Based on the Milnor-Thurston Monotonicity Theorem, prove that if $D_i < D_{i+1}$ and both are legitimate Kneading Sequences, then the corresponding Logistic parameters $u(D_i) < u(D_{i+1})$.

Stage 2: Numerical Verification and Heuristic

- **Goal:** Provide empirical support for key theorems through computational experiments and explore limit behaviors. Some possible methods are as follows:
  - **2.1 Limit Behavior and Chaos Feature Analysis:** Calculate the convergence behavior of the parameter $u(D_i)$ as i increases. Verify if its limit tends to 1.5437.

Analyze the Lyapunov exponent in the limit state, comparing the exponent calculated from the real prime gap sequence with the theoretical value (0.3406) of the Logistic Map at $u \approx 1.5437$.

- **2.2 Numerical Verification Methods for Key Theorems:** Design algorithms to generate D_N for the first N primes and write programs to automatically verify its MSS Maximality condition. Plot the curve of $u(D_i)$ varying with i and observe its rate of approach to the edge of chaos (Does it conform to the Feigenbaum scaling law?).
- **2.3 Analysis of Verification Conclusions:** Based on numerical results, evaluate the confidence of the heuristic proof. Point out which parts match well and which parts have deviations (e.g., outliers in large prime gaps).

Stage 3: Extension of Proof and Theoretical Correction

- **Goal:** Conduct open-ended theoretical exploration on top of the basic proof. Some possible extensions are:
  - **3.1 Number Theoretic Significance of Band Merging:** Deeply explain what "Two-band merging into Single-band" in the Logistic Map corresponds to in number theory (e.g., the homogenization of primes in modular arithmetic classes).
  - **3.2 Dynamic Prediction of Twin Prime Density:** Attempt to derive the Twin Prime Constant using the Invariant Density of the Logistic Map.
  - **3.3 Theoretical Correction:** If the decay characteristic of prime density $1/\ln N$ is found to be inconsistent with standard chaotic attractors, propose a corrected model (such as a Non-autonomous Dynamical System).
  - **3.4 Prime Distribution Problem:** When parameter $u \to 1.5437$, study the ergodicity of the corresponding Logistic Map orbit, thereby obtaining new ideas for solving problems like Twin Primes.

6. Data and Structured Presentation

To auxiliary the large model's understanding, the following table summarizes the key mapping relationships.

Table 1: Prime Sieve and Dynamic Parameter Correspondence Evolution

Sieve Stage (i)	Introduced Prime (pi)	Symbol Sequence Feature (Di)	Estimated Parameter (u)	Dynamical State
1	2	RL (Oscillation)	~2.00	2-Period
2	3	(Complication)	< 2.00	Higher Order Period Window
...	...	...	...	...
∞	-	RLR^{∞}(Quasi-Limit)	1.5437	Chaos (Band Merging)

7. Conclusion and Outlook

This report constructs a bridge connecting Number Theory and Chaos Theory. For the Large Model undertaking this task, this is not merely a computational task, but a process of theoretical exploration. Based on the core settings of Propositions 1-2 and the logical chain of Lemmas 1-3, please boldly conduct numerical experiments and theoretical deductions to finally provide a generative model capable of describing the fine structure of prime gaps.