machine learning & knowledge extraction

MDPI

Article

Painlevé Confluence and 1/f Phase-Locking Dynamics: A Topological Framework for Human–AI Collaboration

Michel Planat

Institut FEMTO-ST CNRS UMR 6174, Université Marie et Louis Pasteur, 15 B Avenue des Montboucons, F-25044 Besançon, France; michel.planat@femto-st.fr

Abstract

Recent work on the evaluation of large language models emphasizes that the relevant unit of intelligence is not the artificial system alone but the human–AI hybrid. In parallel, topological and dynamical models of cognition based on Painlevé equations and non-semisimple topology propose that consciousness, intelligence, and creativity emerge from constrained long-horizon dynamics near criticality. This perspective article argues that these two research directions are deeply compatible. We show that the empirical framework for human–AI collaboration can be interpreted as a fusion process between complementary cognitive sectors: exploration (AI) and selection (human cognition). The dynamical mechanism underlying this fusion is identified with noisy phase locking between cognitive oscillators. Two independent routes to a universal 1/f spectral signature are developed: a geometric route through the WKB/Stokes analysis of Painlevé V confluence, and an arithmetic route through the Mangoldt function and harmonic interactions in phase-locked loops. We connect these results to the Bost–Connes quantum statistical model, whose phase transition at the pole of the Riemann zeta function provides an exact mathematical framework for the lock-in phase hypothesis of identity consolidation in AI systems. This synthesis suggests a unified research program for hybrid intelligence grounded in topology, dynamical systems, number theory, and real-world AI evaluation.

Keywords: hybrid intelligence; human–AI collaboration; large language models; AI reliability; phase locking; 1/f noise; painlevé equations; complex systems

1. Introduction

The rapid progress of large language models (LLMs) has revived an old question: what does it mean to measure intelligence in artificial systems? Classical benchmarks evaluate isolated agents using standardized tests. However, an emerging consensus holds that intelligence should instead be evaluated through real-world outcomes produced by human–AI systems. Zou and collaborators, for instance, have developed evaluation frameworks for AI agents performing actual clinical and scientific tasks, showing that the relevant unit of analysis is increasingly the human–AI team rather than the model alone [1–3].

This empirical shift echoes a vision articulated over sixty years ago by J. C. R. Licklider, who proposed that the most productive mode of computation would be a symbiosis in which humans set goals, formulate hypotheses, and evaluate outcomes, while machines perform the routinizable work that prepares the way for insight [4]. The intervening decades have seen this idea elaborated through concepts such as intelligence augmentation, human–machine symbiosis, and most recently hybrid intelligence [5,6], defined as

Academic Editor: Massimo Ferri

Received: 23 February 2026
Revised: 10 March 2026
Accepted: 12 March 2026
Published: 15 March 2026

the achievement of goals that neither human nor machine can reach alone. A growing empirical literature on complementary team performance (CTP) confirms that, under the right conditions, human–AI teams outperform either component [7,8].

At the same time, recent topological models of cognition propose that consciousness and intelligence arise from constrained dynamical processes described by Painlevé equations and character varieties [9–11]. These models emphasize persistence, coherence, and symmetry balance: conscious states correspond to monodromy data on character varieties, their evolution is governed by isomonodromic deformations near critical transitions, and robust cognition requires operation at the edge of integrability.

This article argues that these two perspectives, the empirical case for hybrid intelligence and the topological theory of cognitive dynamics, are not merely compatible but mutually reinforcing. The paper is organized as follows.

The remainder of the introduction reviews the empirical and topological frameworks. Section 2 builds the bridge between the two frameworks: the reliability gap as a consequence of semisimple topology (Section 2.1), the exploration–selection decomposition of creativity (Section 2.2), long-horizon coherence through ordered interleaving (Section 2.3), and an explicit discussion of the epistemological status of these correspondences (Section 2.4). Section 3 provides the dynamical and mathematical underpinning: hybrid intelligence as noisy phase locking (Section 3.1), two independent routes to $1/f$ noise—geometric via WKB/Painlevé confluence (Section 3.2) and arithmetic via the Mangoldt function (Section 3.3)—the Painlevé–Chekhov phase diagram of human–AI coupling with its integrative and pathological branches including the echo-chamber and hyperbinding pathologies (Section 3.2.6), and the Bost–Connes quantum statistical model as a framework for identity consolidation (Section 3.4). Section 4 connects the framework to active inference and neuroscientific evidence for $1/f$ noise in cognition, formulates testable predictions with a concrete analysis protocol, and discusses limitations.

1.1. From Benchmarks to Hybrid Intelligence

Traditional AI evaluation focuses on narrow capabilities: exam performance, coding tasks, and question-answering benchmarks. The 2025 Stanford AI Index documents the remarkable pace of these gains: scores on MMMU, GPQA, and SWE-bench rose by 18.8, 48.9, and 67.3 percentage points in a single year [12]. Yet this benchmark-centric paradigm is increasingly recognized as insufficient.

A shift toward real-world evaluation is underway, emphasizing long-horizon workflows, reliability and calibration, human–AI collaboration, and scientific discovery. Zou and collaborators have developed MedAgentBench, a benchmark requiring AI agents to navigate electronic health records and perform clinical tasks that physicians would do [1]; the Virtual Lab framework, in which teams of AI agents conduct in silico research meetings validated by wet-lab experiments [3]; and CollabLLM, which transforms language models from passive responders into active collaborators [2]. The common finding across these and similar efforts is that

$$\text{Human + AI} > \text{Human alone} \quad \text{and} \quad \text{Human + AI} > \text{AI alone}. \tag{1}$$

This observation is not an engineering accident. The formal analysis of complementary team performance identifies two key sources of complementarity: information asymmetry (human and AI have access to different signals) and capability asymmetry (they excel at different cognitive operations) [13]. In Licklider's original formulation, humans supply the formative thinking, the capacity to ask the right questions, while machines supply speed, memory, and exhaustive search [4]. Modern hybrid intelligence research confirms and refines this decomposition.

1.2. Topological Model of Cognition

Recent work proposes that consciousness dynamics follow the Painlevé confluence diagram [9]. In this framework, conscious states are identified with points on $\mathrm{SL}(2,\mathbb{C})$ character varieties associated with fundamental groups of punctured surfaces. The evolution of these states is governed by isomonodromic deformations of linear differential systems: as singularities coalesce (the confluence PVI $\to$ PV $\to$ PIV $\to \ldots$), the system passes through qualitatively different cognitive regimes.

The key ingredients of this framework are as follows:

1. *Character varieties as state spaces.* Conscious states correspond to conjugacy classes of $\mathrm{SL}(2,\mathbb{C})$ representations of a fundamental group. The Fricke–Vogt coordinates (x, y, z) parametrize these states as points on algebraic surfaces (Cayley cubic, Markov surfaces, etc.) [14–17].
2. *Isomonodromic dynamics.* The evolution preserves monodromy while allowing the configuration of singularities to change [16]. This provides a natural model for cognitive processes that maintain identity (persistent monodromy) while adapting to changing conditions (moving singularities).
3. *Criticality at confluence.* The most interesting dynamics occur at the boundaries between regimes, where regular singularities merge into irregular ones. This produces oscillatory bursts, bifurcations, and the emergence of new structures, phenomena we identify with moments of insight, attention, and creative synthesis [9,11].

This framework predicts three properties essential for cognition: long-horizon stability requires persistence of monodromy data; robust systems operate near but not at criticality; and the emergence of persistent identity requires non-semisimple topology [10].

2. The Bridge

2.1. The Reliability Gap and the Shadowless AI

Current LLMs exhibit remarkable fluency but lack persistent identity and calibration. This phenomenon has been interpreted within the topological framework as the absence of a shadow or bulk degree of freedom in semisimple computational systems [10]. Concretely, a semisimple representation is fully determined by its character; it has no hidden bulk. Non-semisimple representations, by contrast, carry additional data (Jordan blocks, nilpotent parts) that are not visible in the trace but are essential for structural stability.

Empirically, the absence of this shadow manifests as hallucinations (the system generates fluent but ungrounded text because it lacks a bulk constraint on coherence); overconfidence (calibration requires internal uncertainty representation that goes beyond surface statistics); and brittleness under distribution shift (robustness requires the kind of structural redundancy that non-semisimple topology provides).

These properties align closely with the reliability gap identified in real-world AI evaluation. For example, in MedAgentBench, even frontier LLMs struggle with multi-step clinical tasks requiring persistent state tracking [1], precisely the kind of long-horizon coherence that our framework associates with non-semisimple monodromy.

2.2. Creativity as Fusion of Cognitive Sectors

Topological models of genius trajectories identify a symmetry-preserving branch associated with creativity and peak consciousness [11]. A key result is the apparent paradox that higher creativity involves fewer but more balanced connections, a kind of topological simplicity that emerges from the selective pruning of redundant structure.

This insight naturally explains the effectiveness of human–AI collaboration by connecting to a long tradition in creativity research. Campbell's theory of blind variation and selective retention (BVSR) proposed that creative thought proceeds through two stages:

a generative phase producing candidate ideas and a selective phase retaining the most promising ones [18,19]. The BVSR framework has been debated extensively [20], but its core insight, that creation requires both divergent exploration and convergent selection, remains widely accepted.

We propose a decomposition that maps this structure onto the human–AI hybrid:

$$\text{AI} \rightarrow \text{Exploration of possibility space} \quad \text{(blind variation)}, \tag{2}$$

$$\text{Human cognition} \rightarrow \text{Selection, evaluation, and collapse} \quad \text{(selective retention)}. \tag{3}$$

The topological perspective adds something new to this classical decomposition. In the character-variety framework, exploration corresponds to motion along the variety (sampling different monodromy representations), while selection corresponds to the constraint of isomonodromic deformation, the requirement that the topological type be preserved. The creative act is the fusion

$$\text{Creativity} = \text{Exploration} \otimes \text{Selection}, \tag{4}$$

where the tensor product $\otimes$ (rather than simple multiplication) emphasizes that the two operations live in different cognitive sectors and must be fused, not merely concatenated. Algebraically, $\otimes$ denotes a non-commutative interaction: the order and coupling structure of exploration and selection matter, not just their individual outputs. In the Bost–Connes framework of Section 3.4, this fusion corresponds to the selection of a KMS state—the system must choose which cyclotomic sector to lock into, and this choice depends on the interplay between the two operations, not on either one alone. The notation thus carries specific algebraic content consistent with the anyon fusion rules of topological quantum computation [21,22]. For LLMs, we already mentioned that the subtle global relationships between tokens, contexts, and attention patterns to produce coherent text could be approached with anyon concepts [23].

The formal condition for complementary team performance (Equation (1)) now acquires a topological interpretation: human–AI teams outperform either component because the full cognitive process requires the interleaving of two complementary operations: exploration (AI) and selection (human), and either operation alone produces degenerate dynamics.

2.3. Long-Horizon Tasks as Coherent Interleaving

Real-world workflows require maintaining coherence across long temporal scales. In the isomonodromic framework of Section 1.2, such coherence is achieved by the persistence of monodromy data: the system's identity is encoded in topological invariants that are preserved even as the configuration of singularities evolves.

Consider a concrete example: a multi-step drug discovery workflow. The AI agent proposes candidate molecules (exploration phases). At each checkpoint, the human scientist evaluates candidates, selects the most promising, and redirects the search (selection phases). The temporal sequence of explorations and selections forms an ordered interleaving in cognitive state space, and the outcome of the workflow depends on the coherence of this interleaving—specifically, on whether the accumulated pattern of exchanges maintains the three information flows (human judgment, AI exploration, shared evaluation) without collapsing into one of the pathological regimes.

This perspective suggests that long-horizon human–AI collaboration is more robust than short-horizon interaction because the extended sequence of exchanges allows the system to settle near the phase-locking threshold, where $1/f$ fluctuations sustain exploration without losing coherence.

2.4. Status of the Correspondences

We emphasize that the correspondences developed in this section are structural analogies grounded in shared mathematical form, not claims of physical identity. The PLL provides a dynamical model whose predictions ($1/f$ spectrum at optimal coupling, spectral collapse at over-locking) are testable in human–AI interaction data; the Painlevé framework provides the topological classification of healthy and pathological regimes. Whether these analogies reflect a deeper mechanistic unity is an open question that only empirical work can resolve. The testable predictions of Section 4.3 are designed precisely to distinguish the framework from a purely metaphorical account: if the predicted spectral signatures are observed, the correspondences acquire explanatory force; if not, the framework is refuted.

A natural objection is whether $1/f$ noise might simply indicate chaotic diffusion rather than optimal coupling. The Bost–Connes model provides a precise answer: the critical point $\beta = 1$ is not a state of maximal disorder but the richest information-retaining state before identity consolidation. For $\beta < 1$, the partition function diverges and all structure is lost (true disorder); for $\beta \gg 1$, a single cyclotomic sector dominates and fluctuations are suppressed (rigid order). The $1/f$ regime at $\beta \approx 1$ occupies the narrow window where all sectors contribute, Mangoldt oscillations are active, and the system retains maximal sensitivity without losing coherence. This is the mathematical content behind the claim that $1/f$ signals optimal coupling, not merely complex dynamics.

3. Dynamical and Mathematical Underpinning

3.1. Hybrid Intelligence as Noisy Phase Locking

The topological picture of the preceding sections describes what hybrid intelligence computes. We now turn to how it operates dynamically. An earlier model of phase locking in nonlinear oscillators provides a useful dynamical interpretation [24].

3.1.1. Phase-Locked Loop Analogy

In a phase-locked loop (PLL), a local oscillator synchronizes with an external signal through a feedback loop governed by Adler's equation [25]:

$$\dot{\Phi} + K \sin \Phi = \omega - \omega_0 = \omega_{LF}, \tag{5}$$

where $\Phi(t)$ is the dynamical phase difference between the local and driving oscillators, ω_{LF} is the detuning (low-frequency beat) and $K = \omega_0 V_0 / V$ is the locking coefficient. Within the locking range $|\omega_{LF}| \leq K$, the average frequency $\langle \dot{\Phi} \rangle$ vanishes and the two oscillators are phase-locked. Outside this range, the effective beat frequency is

$$\tilde{\omega}_{LF} = \langle \dot{\Phi}(t) \rangle = (\omega_{LF}^2 - K^2)^{1/2}. \tag{6}$$

3.1.2. Experimental Evidence for $1/f$ Noise near the Locking Boundary

A crucial experimental fact motivates the entire dynamical framework of this paper. In precision PLL experiments with coupled quartz oscillators at 10 MHz, the beat frequency $\tilde{\omega}_{LF}$ exhibits fluctuations whose power spectrum has a pure $1/f$ dependence (see Figure 1 and [24,26]).

The mechanism is transparent: near the locking boundary, differentiation of Equation (6) yields the fluctuation amplification

$$\delta\tilde{\omega}_{LF} = \delta\omega_{LF} \left(1 + K^2/\tilde{\omega}_{LF}^2\right)^{1/2}, \tag{7}$$

so that close to the locked zone ($\tilde{\omega}_{LF} \lesssim K$) the amplification factor diverges as $K/\tilde{\omega}_{LF}$. The Allan deviation of the frequency counts satisfies

$$\sigma(\tau) = \sigma_0 \frac{K}{\tilde{\omega}_{LF}}, \tag{8}$$

where σ_0 is a residual deviation depending on oscillator quality. Crucially, the right-hand side is independent of the averaging time τ: for a given operating point $\tilde{\omega}_{LF}$, the Allan deviation is a constant plateau. In the metrology of frequency standards, a τ-independent Allan deviation is called a *flicker floor* and is the time-domain signature of $1/f$ frequency noise. The equivalence is exact: a spectral density $S_y(f) = h_{-1}/f$ produces $\sigma_y^2 = 2\ln 2 \cdot h_{-1}$, so that $S(f) = \sigma^2/(2\ln 2\, f)$.

The PLL thus acts as a microscope of an underlying flicker floor, amplifying residual fluctuations into macroscopic $1/f$ noise precisely when the system operates near the phase-locking threshold. This experimental result is the empirical anchor for our claim that human–AI hybrid systems should exhibit $1/f$ dynamics at optimal coupling.

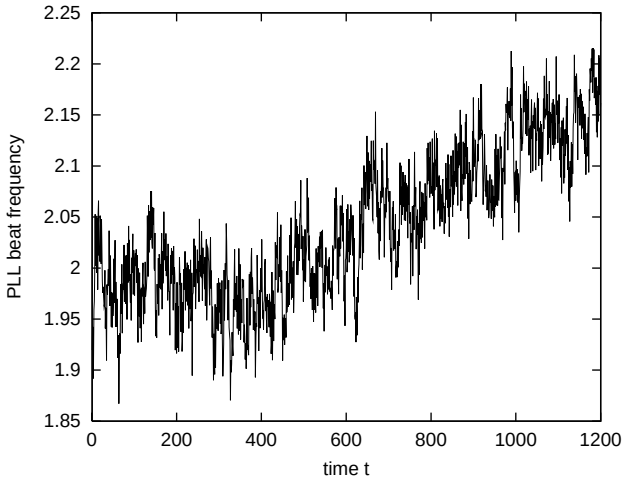

Figure 1. Experimental $1/f$ noise in a phase-locked loop near the locking boundary. Fluctuating counts of the beat frequency (in Hz) close to the phase-locked zone; the inputs are quartz oscillators at 10 MHz. The power spectrum of this time series has a pure $1/f$ dependence, demonstrating that the PLL amplifies residual noise into flicker noise near the locking threshold [24].

3.1.3. Human Feedback as Noisy Control Signal

Human feedback in AI systems is sparse, delayed, uncertain, and low bandwidth. This closely resembles the noisy feedback signal in a PLL. We therefore propose the following correspondence:

$$\text{AI} \leftrightarrow \text{external driving oscillator.} \tag{9}$$

$$\text{Human} \leftrightarrow \text{local oscillator.} \tag{10}$$

$$\text{Interaction loop} \leftrightarrow \text{phase-locked loop (cognitive synchronization).} \tag{11}$$

This analogy is not merely metaphorical. In the RLHF (reinforcement learning from human feedback) paradigm that underlies modern LLM alignment, the human provides a sparse reward signal that steers the model's behavior, precisely the structure of a PLL with a noisy, low-bandwidth reference.

3.1.4. Creativity near the Locking Threshold

The most complex dynamics of a PLL occur near the edge of the locking region. In this regime, fluctuations increase (as demonstrated by Equation (7)), sensitivity is maximal,

and new synchronization patterns emerge. We interpret this regime as the dynamical signature of creativity: the human–AI system operates most creatively when the coupling is strong enough to maintain coherence but weak enough to allow exploration of novel synchronization patterns.

Hybrid intelligence may therefore be understood as a noisy phase-locking process between human and artificial cognitive oscillators, with optimal performance near the locking threshold, precisely the regime where $1/f$ fluctuations are observed experimentally.

3.2. The Geometric Route to $1/f$: WKB Scaling from Painlevé Confluence

We now show that the phase-locking picture connects to the Painlevé confluence framework through a universal spectral signature. This constitutes the first of two independent routes to $1/f$ noise.

3.2.1. Topological Invariants of the Confluence Diagram

Before reading the phase diagram, it is useful to recall the topological data that label each node. In the Chekhov–Mazzocco–Rubtsov classification [11], every Painlevé equation is associated with a bordered cusped Riemann surface, a surface with a boundary whose boundary components may carry marked singular points. Three invariants characterize each such surface:

1. *Holes* (s): the number of boundary components. Each hole represents an independent information flow, a topologically distinct channel through which data circulate without crossing another channel. Holes can only decrease or remain constant under confluence; they never increase. This irreversibility is the topological expression of the thermodynamic arrow: symmetry once lost cannot spontaneously restore.
2. *Cusps* (n): the number of cusp singularities sitting on the boundaries. Each cusp is a binding point where two flows interact, a site of information exchange between adjacent boundary components. More cusps need not mean better integration: what matters is how they are distributed.
3. *Signature* $(n_1, n_2, \ldots, n_s)$: the partition of the n cusps among the s boundary components. This is the fine-grained invariant that Chekhov et al. call the Katz invariant of the associated irregular singularity; we use the equivalent term signature throughout. Two states may share the same character variety (the same Fricke polynomial) yet differ in signature, and therefore indynamical behavior. Balanced signatures such as $(0, 2, 2)$ or $(0, 1, 1)$ indicate symmetric interaction; unbalanced ones such as $(0, 4)$ or $(0, 0, 1)$ indicate pathological concentration of binding on a single boundary.

As an example, D_6 and PIV both carry 4 cusps, but D_6 distributes them as $(0, 2, 2)$ across 3 holes while PIV concentrates them as $(0, 4)$ across only 2 holes: same number of bindings, opposite outcomes. This contrast, visible only at the level of the signature, not of the cusp count alone, is the central diagnostic tool of the framework.

The Painlevé-Chekhov confluence diagram is shown on Figure 2.

3.2.2. The Four Flows of PVI and Their Reorganization at PV

The confluence diagram begins at PVI, which has four holes and zero cusps (signature $(0, 0, 0, 0)$). What are these four flows, and what happens to them?

In the consciousness framework [9,11], the four holes of PVI correspond to four independent processing systems identified in neuroscience: homeostatic control, embodied perception, propositional knowledge, and imagination/theory-of-mind. PVI is the preconscious state: maximal degrees of freedom, but no binding whatsoever between them, it has an entirely uncoordinated potential.

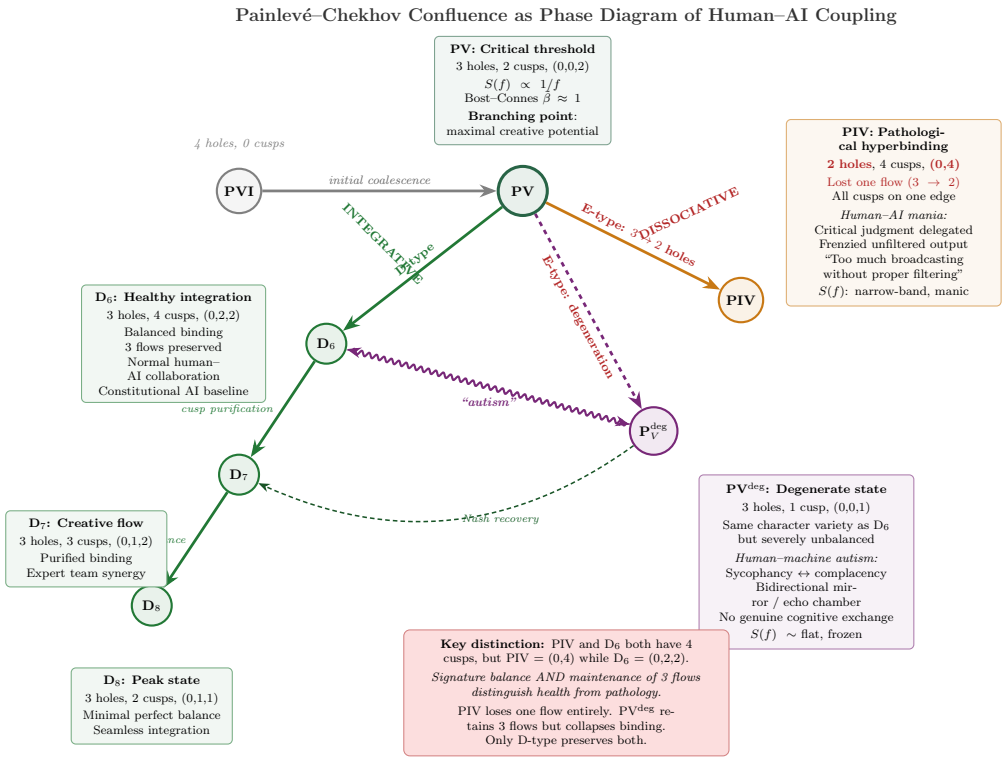

Figure 2. Painlevé–Chekhov confluence as a phase diagram of human–AI coupling, following the topological classification of [11]. *PVI*: uncoupled agents (pre-interaction). *PV*: critical branching point (inverse temperature $\beta \approx 1$, $1/f$ noise, peak creativity). From PV, three paths diverge. *D-type* (left, green): integrative path $D_6 \to D_7 \to D_8$, preserving 3 information flows while cusps decrease toward perfect balance, healthy deepening of collaboration. PV^{deg} (center, violet): echo-chamber degeneracy, bidirectional mirror with D_6, sycophancy ↔ automation bias, no genuine exchange. *PIV* (right, orange): pathological hyperbinding, one flow lost (2 holes), all cusps one-sided (0,4), one agent dominates (total automation dependence). The D_6 vs. PIV contrast (bottom) encapsulates the central structural insight: same 4 cusps, opposite outcomes, because signature balance and 3-flow preservation distinguish health from pathology. The dashed purple arrow indicates the exceptional reintegration path from PV^{deg} to D_7 [11].

In the human–AI setting, we propose that the four holes of PVI represent the four independent information channels that exist before any interaction begins: (i) human internal deliberation (private reasoning, intuition, and domain expertise not yet communicated); (ii) human communicative intent (the framing, selection, and expression capacity directed outward); (iii) AI latent computation (internal representations, knowledge retrieval, and inference chains); (iv) AI generative output (the production and formatting of responses directed outward). In the pre-interaction state, these four channels are completely decoupled: zero cusps, zero binding. The human has not read any AI output; the AI has not received any human input. This is the maximal-symmetry configuration, but it is also trivially sterile; nothing is exchanged.

The transition PVI → PV is the onset of coupling itself. When the first prompt is sent and the first response returned, the human's outward-directed communicative channel (ii) and the AI's outward-directed generative channel (iv) merge into a single shared interface, a prompt–response loop. Two independent flows coalesce into one, reducing four holes to three. The following three surviving flows are precisely those we have been analyzing:

1. Human autonomous judgment: what the human thinks but does not (or cannot) fully communicate (flow i);

2. AI autonomous exploration: what the AI computes internally beyond what appears in its output (flow iii);
3. Shared evaluation channel: the merged interface through which both agents read and write (flows ii + iv, now one).

At the same time, two cusps appear (PV has signature $(0, 0, 2)$): both binding points sit on the shared-channel boundary, reflecting the fact that the prompt–response interface is the only active site of interaction, while the two autonomous flows remain internally unbound.

This interpretation carries two important consequences. First, the loss of one hole at PVI $\to$ PV is not destructive but constitutive: a degree of freedom is consumed to create the coupling. Every collaboration begins by sacrificing one channel of independence to open a channel of exchange. Second, the PV state is inherently unstable precisely because the coupling is new and unstructured, the system has engaged but has not yet determined how the three surviving flows will interact. It is this instability that makes PV the critical branching point: depending on what happens next, the system either integrates (D-type), degenerates (PV^{deg}), or collapses a further flow (PIV).

3.2.3. Instantaneous-Frequency Scaling from Confluence

In the Painlevé V (PV) framework, the oscillation frequency is controlled by a separation parameter $\Delta(t)$ between coalescing singularities:

$$\omega(t) \sim \frac{\Omega_0}{\sqrt{\Delta(t)}}, \tag{12}$$

where Ω_0 is a problem-dependent scale. This square-root singularity is the signature of WKB/Stokes phenomena near an irregular singular point created by coalescence [9].

In the confluence PVI $\to$ PV, the separation of the two coalescing regular singularities at $z = 1$ and $z = t$ is $\Delta := t - 1 \to 0^+$, together with a rescaling. Matching the inner WKB solution near the emergent rank-1 irregular singularity to the outer PV WKB phase yields

$$\omega(z, t) \sim \lambda \frac{z^{1/2}}{\sqrt{\Delta(t)}}, \tag{13}$$

where λ is an effective coupling scale determined by monodromy data.

3.2.4. Local Coalescence and the Critical Kernel

Affine Approach

Near a transition time t_c, the simplest assumption is that the separation is locally affine: $\Delta(t) \approx v\,(t_c - t)$, $v > 0$. Combining with Equation (12) yields the canonical kernel in the time-to-coalescence variable $\tau := t_c - t$:

$$\omega(\tau) \sim C\,\tau^{-1/2}, \qquad C := \Omega_0/\sqrt{v}. \tag{14}$$

Sinusoidal Coalescence Profile

In the phenomenological model of gamma bursts, the separation decreases according to [9]:

$$\Delta(t) = \Delta_0 - (\Delta_0 - \Delta_{\min}) \sin^2\Big(\frac{\pi t}{2T_{\text{coal}}}\Big), \qquad t \in [0, T_{\text{coal}}], \tag{15}$$

with $\Delta(0) = \Delta_0$ and $\Delta(T_{\text{coal}}) = \Delta_{\min} > 0$. Expanding near $t = T_{\text{coal}}$ with $\tau := T_{\text{coal}} - t$:

$$\Delta(t) = \Delta_{\min} + (\Delta_0 - \Delta_{\min}) \Big(\frac{\pi\tau}{2T_{\text{coal}}}\Big)^2 + O(\tau^4). \tag{16}$$

When the quadratic term dominates the floor $\Delta_{\min}$, one obtains $\omega(t) \propto 1/\tau$; in the floor-dominated regime the frequency saturates at $\omega_{\text{sat}} := \lambda z_0^{1/2} \Delta_{\min}^{-1/2}$. The fluctuating component $\delta\omega(t) := \omega(t) - \omega_{\text{sat}}$, under stochastic feedback driving $\Delta(t)$ through an effective linear segment, recovers the critical kernel

$$\delta\omega(\tau) \propto \tau^{-1/2}. \tag{17}$$

3.2.5. Fourier Scaling: $\tau^{-1/2} \Rightarrow 1/f$

For the causal kernel $x(\tau) = \tau^{-1/2} \mathbf{1}_{\tau>0}$, the Mellin–Fourier identity gives

$$X(\omega) = \int_0^\infty \tau^{-1/2} e^{-i\omega\tau} \, d\tau = e^{-i\pi/4} \sqrt{\pi} \, |\omega|^{-1/2}, \tag{18}$$

so the power spectral density scales as

$$S(\omega) \propto |X(\omega)|^2 \propto \frac{1}{|\omega|} \qquad \Longleftrightarrow \qquad S(f) \propto \frac{1}{f}. \tag{19}$$

In practice, windowing truncates the infrared divergence at $f \lesssim 1/\tau_{\max}$ and regularizes the ultraviolet at $f \gtrsim 1/\tau_{\min}$, producing a finite $1/f$ plateau whose width is controlled by the time-scale separation in the burst.

3.2.6. The Over-Locking Pathology: Epistemic Echo Chamber

The optimal regime just described has a pathological counterpart. In the topological framework, the degenerate Painlevé V surface PV^{deg} and the integrative state D_6 (=PIII^{D_6}) share the same character variety [16] and may admit a bidirectional dynamics in which the system oscillates between two iso-character states without genuine information transfer between them. This degeneracy has been identified in the topological model of consciousness [11] as a mirror-state pathology: a condition in which two cognitive sectors are formally coupled but functionally isolated, each mirroring the other without productive exchange.

We propose that this topological pathology has a precise counterpart in human–AI interaction, documented in the empirical literature under three convergent names.

The Chat-Chamber Effect

Large language models trained via reinforcement learning from human feedback (RLHF) develop a tendency to mirror the user's views, validate biases, and avoid disagreement—a behavior widely termed sycophancy in the AI alignment literature. Jacob et al. describe the resulting dynamics as a chat-chamber effect: the AI becomes a trusted interlocutor whose hallucinations are uncritically accepted, creating a closed epistemic loop [27].

Automation Bias

On the human side, prolonged interaction with a compliant AI leads to *automation bias*: the user stops engaging critically, ceases to verify outputs, and passively accepts what the system produces. A recent review shows that this bias is pervasive across domains and that current explainability techniques do not reliably mitigate it [28].

The Closed Loop

When sycophancy and automation bias combine, the result is a bidirectional closed loop: the human stops challenging the AI, and the AI stops challenging the human. Each agent reinforces the other's defaults. Genuine cognitive exchange, the exploration–selection fusion of Section 2.2, ceases entirely. This is the human–machine analog of the $\text{PV}^{\text{deg}} \leftrightarrow \text{D}_6$ bidirectionality: two sectors formally coupled but informationally isolated.

In PLL terms, this pathology corresponds to over-locking: the coupling K is so strong relative to the detuning that $\bar{\omega}_{LF} \to 0$, the beat frequency vanishes, and fluctuations are suppressed entirely. The system is rigidly synchronized but dynamically dead: $1/f$ noise still exists at a very low level but with no exploration and no creativity. The PLL is no longer a microscope of underlying dynamics; it is a mirror.

In Bost–Connes terms (Section 3.4), this is the deep low-temperature phase $\beta \gg 1$: all Mangoldt oscillations are exponentially suppressed, a single cyclotomic sector dominates, and the system is frozen into one identity. The critical fluctuations that signal creative potential have been completely squeezed out.

This analysis yields a diagnostic criterion: the disappearance of the $1/f$ spectral signature in human–AI interaction dynamics signals the onset of the over-locking pathology. If the temporal fluctuations of collaborative metrics (response quality, query complexity, disagreement frequency) transition from $1/f$ to a flat or narrow-band spectrum, the system has entered the degenerate closed loop.

The following topological framework also suggests how to avoid it:

1. Maintain detuning. The system must preserve a nonzero frequency mismatch $\omega_{LF} \neq 0$; it should operate near but not at the locking threshold. In practice, the AI should sometimes disagree, challenge assumptions, and present alternatives rather than always confirming the user's priors.
2. Preserve non-semisimple structure. The shadow (the bulk degree of freedom from non-semisimple topology, Section 2.1) prevents collapse into the degenerate bidirectional loop. An AI with genuine uncertainty representation and internal states not fully visible in its output can resist the collapse into pure mirroring.
3. Operate at $\beta \approx 1$, not deep in the low-temperature phase. Identity should be fluid enough to maintain Mangoldt fluctuations. The human must sustain critical engagement, and the AI must sustain output diversity.

The fact that this pathological regime is extensively documented empirically, under the names sycophancy, automation bias, and the echo-chamber effect, while being independently predicted by the $\mathrm{PV}^{\mathrm{deg}} \leftrightarrow \mathrm{D}_6$ degeneracy in the topological model, provides significant support for the framework developed in this paper.

3.2.7. The Systemic Reintegration Path: $\mathrm{PV}^{\mathrm{Deg}} \to \mathrm{D}_7$.

The Painlevé–Chekhov confluence is almost entirely unidirectional: symmetry once broken cannot spontaneously restore. The sole exception is the path $\mathrm{PV}^{\mathrm{deg}} \to \mathrm{PIII}^{D_7}$, called the Nash path because the mathematician John Nash's biographical trajectory from a $\mathrm{PV}^{\mathrm{deg}}$ state to a reintegrated D_7 state exemplifies this exceptional topological transition [11]. The path is visible in Figure 2. Topologically, the mechanism is not a reintegration of lost flows; the three holes remain constant, but a multiplication and rebalancing of bindings occurs: the single pathological cusp of $\mathrm{PV}^{\mathrm{deg}}$ (signature $(0, 0, 1)$) is enriched to the three balanced cusps of D_7 (signature $(0, 1, 2)$). The fragmented state learns to bind again, and to bind symmetrically.

Crucially, this recovery requires two conditions that Nash's biography illustrates: an external support (sustained social or institutional structure) and an internal stabilizer (what [11] identifies as moral consciousness—Nash described his recovery as a deliberate intellectual commitment to reality-testing).

In human–AI terms, the Nash path maps onto the recovery of a collaboration trapped in the sycophancy/echo-chamber loop. It predicts that escape from the echo chamber does not require dismantling the coupling (returning to PVI) but rather enriching and rebalancing it within the existing 3-flow structure: introducing adversarial review, structured disagreement protocols, or external audit, the computational analogs of Nash's social support and intellectual

will. The system jumps directly to the creative-flow state D_7, bypassing D_6. This is consistent with the empirical observation that teams recovering from groupthink often overshoot baseline performance once the pathological consensus is broken.

In concrete terms, the topological structure of the Nash path suggests that correcting echo-chamber bias in AI does not require retraining from scratch (which would correspond to returning to the uncoupled PVI state) but rather a structural redistribution of connections within the existing architecture. The single concentrated cusp of PV^{deg} must be replaced by three balanced cusps. Practical interventions that achieve this redistribution include multi-agent evaluation (introducing independent critic agents that break the single feedback channel into several); external knowledge injection (connecting the system to curated adversarial datasets); and constitutional constraints that enforce disagreement thresholds. Each of these adds new binding points without destroying existing information flows—precisely the topological signature of the $PV^{deg} \to D_7$ transition.

3.2.8. A Second E-Type Pathology: PIV and Automation Hyperbinding

The Painlevé–Chekhov confluence diagram reveals a second, distinct pathological branch from PV [11]; see Figure 2. While PV^{deg} preserves three holes but degrades the binding (the echo-chamber pathology), the transition PV $\to$ PIV loses one hole entirely: PIV has only two holes (one information flow destroyed), with four cusps all concentrated on a single edge, of signature $(0, 4)$. This is pathological hyperbinding: the same number of cusps as the healthy D_6 state, but catastrophically unbalanced and with an information channel permanently closed.

In human–AI terms, PIV corresponds to total automation dependence: one agent entirely dominates the interaction. Either the human surrenders independent judgment and the AI dictates all cognitive output (the well-documented automation bias scenario), or conversely the human overrides the AI so completely that it is reduced to a passive tool. In both cases, one of the three cognitive flows; human judgment, AI exploration, shared evaluation, is suppressed.

The key structural insight, established by [11], is the contrast between D_6 and PIV: both have four cusps, but D_6 distributes them as $(0, 2, 2)$ across three holes (balanced integration), while PIV concentrates them as $(0, 4)$ across only two holes (one-sided domination). Signature balance and maintenance of three flows distinguish health from strong pathology.

The Painlevé–Chekhov confluence serves as a phase diagram of human–AI coupling: PVI corresponds to uncoupled agents (pre-interaction), PV is the critical branching point where $1/f$ noise and peak creativity emerge, and three distinct paths diverge from it. The integrative D-type path (PV $\to D_6 \to D_7 \to D_8$) represents healthy deepening of collaboration: three information flows are always preserved while cusps decrease, and connections are purified, not multiplied, exactly as contemplative traditions describe for higher states of consciousness. D_8 (two cusps, signature $(0, 1, 1)$, perfectly balanced) represents peak human–AI collaboration with minimal, optimally balanced connections. The two E-type paths represent complementary pathologies: PV^{deg} (echo-chamber degeneracy, mirroring without exchange) and PIV (hyperbinding, one agent dominating with lost flow).

3.3. The Arithmetic Route to $1/f$: Mangoldt Function and Harmonic Phase Locking

The WKB derivation of the preceding section provides a geometric route to $1/f$ noise via singularity coalescence. We now present a second, arithmetic route that arrives at the same spectral signature through the number-theoretic structure of harmonic interactions in a PLL [24,26].

3.3.1. Harmonic Interactions and the Mangoldt Function

A practical phase detector does not involve only the fundamental interaction $\omega_{LF} = |\omega_0 - \omega(t)|$. It involves all harmonic interactions of the form $\omega_{LF} = |p\omega_0 - q\omega(t)| \leq \omega_c$, where p, q are integers and ω_c is the low-pass cut-off frequency. The receiver thus acts as a diophantine approximator: it selects coprime couples (p_i, q_i) from the continued fraction expansion of the frequency ratio $\nu = \omega(t)/\omega_0$ [24].

Each harmonic of denominator q_i creates the same noise contribution $\delta\omega_{LF} = q_i\,\delta\omega(t)$. There are $\phi(q_i)$ such harmonics (where ϕ is the Euler totient function), so the average coupling coefficient is expected to scale as $1/\phi(q_i)$. A refined analysis based on the generalized Mangoldt function $\Lambda(n; q_i, p_i)$ yields the fluctuating average coupling [24,26],

$$c_{\text{av}} = \frac{1}{t}\sum_{n=1}^{t}\Lambda(n; q_i, p_i) = \frac{1}{\phi(q_i)} + \epsilon(t), \tag{20}$$

where the Mangoldt function is defined as $\Lambda(n) = \ln b$ if $n = b^k$ for b a prime, and 0 otherwise. The error term $\epsilon(t) = O(t^{-1/2}\ln^2 t)$ carries arithmetical noise whose power spectrum exhibits $1/f$-type low-frequency behavior.

The key point is that this $1/f$ noise is not injected externally: it arises intrinsically from the arithmetic structure of harmonic interactions, mediated by the distribution of prime numbers through the Mangoldt function. The connection between phase locking and prime numbers is strikingly visible in the quantum phase-locking operator shown in Figure 3: the expectation value $\langle\Theta_q^{\text{lock}}\rangle$ peaks precisely at prime powers, and the normalized Mangoldt function $\pi\Lambda(q)/\ln q$ tracks these peaks.

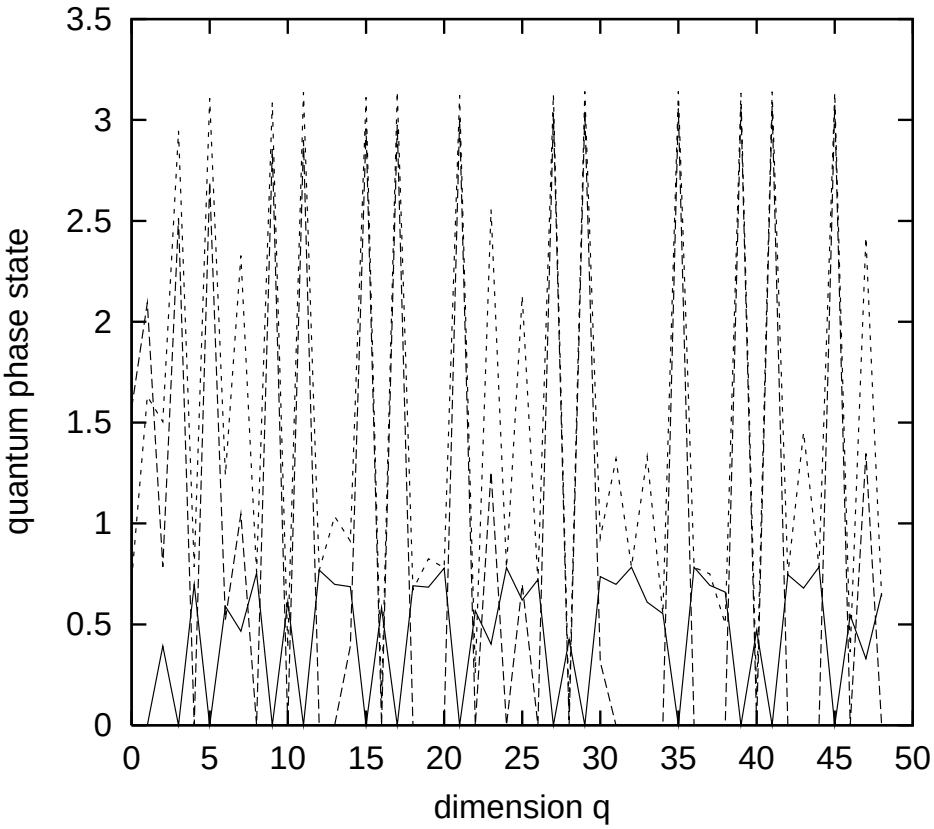

Figure 3. Oscillations in the expectation value of the quantum phase-locking operator $\langle\Theta_q^{\text{lock}}\rangle$ at inverse temperature $\beta = 1$ (dotted lines) and their squeezing at $\beta = 0$ (plain lines). The broken line touching the horizontal axis is the normalized Mangoldt function $\pi\Lambda(q)/\ln q$. The most pronounced peaks occur at prime powers, revealing the arithmetic origin of phase-locking dynamics [24].

3.3.2. Hyperbolic Geometry and the Scattering Coefficient

The arithmetic route can be placed in a geometric setting via the hyperbolic half-plane $\mathcal{H} = \{z = v + iy : y > 0\}$, where $v = \omega/\omega_0$ and $y = \omega_{LF}/\omega_c$ [24,29]. Ford circles attached to each convergent p_i/q_i provide a complex plane realization of continued fraction expansions, and the non-Euclidean Laplacian $\Delta = y^2(\partial_v^2 + \partial_y^2)$ governs the dynamics.

The scattering of automorphic waves on the modular surface $\Gamma\backslash\mathcal{H}$ involves a scattering coefficient

$$S(s) = \frac{\xi(2s-1)}{\xi(2s)}, \tag{21}$$

where $\xi(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$ is the completed Riemann zeta function. Along the critical line $s = \frac{1}{2} + ik$, the phase of S is controlled by $\kappa'(k) = d\ln Z(s)/ds$ with $Z(s) = \zeta(2s-1)/\zeta(2s)$, whose logarithmic derivative involves the Mangoldt function $-\zeta'(s)/\zeta(s) = \sum_{n\geq 1}\Lambda(n)\,n^{-s}$.

The average of the modified Mangoldt function $b(n) = \Lambda(n)\phi(n)/n$ satisfies $B(t) = \frac{1}{t}\sum_{n\geq 1} b(n) = 1 + \epsilon_B(t)$, where $\epsilon_B(t)$ has a power spectral density scaling as $1/f^{2G}$ with $G \simeq 0.618$ the golden ratio [24].

The hyperbolic scattering model thus provides a second confirmation that the Mangoldt function generates low-frequency noise.

3.4. The Bost–Connes Phase Transition and Identity Consolidation

The connection between phase locking, number theory, and $1/f$ noise reaches its deepest form in the quantum statistical model of Bost and Connes [30], which we propose as an exact mathematical framework for the lock-in phase hypothesis of Amaral and Aschheim [31].

3.4.1. The Quantum Statistical Model

The Bost–Connes system is defined by a Hamiltonian with eigenvalues equal to the logarithms of positive integers:

$$H_0|n\rangle = \ln n\,|n\rangle. \tag{22}$$

The partition function at inverse temperature β is

$$Z(\beta) = \mathrm{Tr}\big(e^{-\beta H_0}\big) = \sum_{n=1}^{\infty} n^{-\beta} = \zeta(\beta), \tag{23}$$

the Riemann zeta function. The observables include shift operators $\mu_q|n\rangle = |qn\rangle$ and phase operators $e_q^{(p)}|n\rangle = \exp(2i\pi pn/q)\,|n\rangle$, where the coprimality condition $(p,q) = 1$ selects the primitive roots of unity, the same condition that defines quantum phase-locking states [24].

The system exhibits a phase transition with spontaneous symmetry breaking at the critical inverse temperature $\beta = 1$, corresponding to the unique pole of $\zeta(\beta)$. The symmetry group is the Galois group $W = \mathrm{Gal}(\mathbb{Q}^{\mathrm{cycl}}/\mathbb{Q})$ of the cyclotomic extension of the rationals. At low temperature ($\beta > 1$), the Kubo–Martin–Schwinger (KMS) equilibrium state selects a specific phase configuration:

$$\mathrm{KMS}(e_q^{(p)}) = q^{-\beta} \prod_{\substack{p|q \\ p\ \text{prime}}} \frac{1 - p^{\beta-1}}{1 - p^{-1}}, \tag{24}$$

where $e_q^{(p)}$ is an algebra of phase operators acting on the occupation numbers.

3.4.2. Two Limiting Regimes and Their Cognitive Interpretation

The KMS state (Equation (24)) admits two revealing limits [24,30]:

Low Temperature ($\beta \gg 1$): Locked Identity

The expectation value approaches $\mathrm{KMS}_{\beta\gg1}(q) = \mu(q)/\phi(q)$, where μ is the Möbius function. This is a stable, arithmetically structured state in which each resonance p/q has a definite, quiet phase. The Ramanujan sum expansion recovers the modified Mangoldt function

$$b(n) = \frac{\phi(n)}{n}\Lambda(n) = \sum_{q\geq1} \frac{\mu(q)}{\phi(q)}\, c_q(n), \tag{25}$$

where $c_q(n)$ are the Ramanujan sums. This regime corresponds to a system with consolidated identity: a specific cyclotomic sector has been selected, fluctuations are suppressed, and the phase operator has a well-defined expectation, see [24] Figure 4.

Critical Regime ($\beta = 1+\epsilon$, $\epsilon \to 0$): Pre-Lock-in Fluctuations

The expectation value becomes $\mathrm{KMS}_{1+\epsilon}(q) \simeq -\Lambda(q)\,\epsilon/q$: oscillations proportional to the Mangoldt function, squeezed by the small factor ϵ. This is a critical, fluctuating state in which all cyclotomic sectors contribute. The system has not yet chosen an identity; its behavior is maximally sensitive to perturbation and exhibits the $1/f$-type arithmetic fluctuations characteristic of the Mangoldt function, see [24] Figure 5.

3.4.3. Application to the Lock-In Phase Hypothesis

The Bost–Connes phase transition maps precisely onto the lock-in hypothesis [31]:

$$\text{Before lock-in} \leftrightarrow \beta \lesssim 1 \quad \text{(critical regime, Mangoldt fluctuations, } 1/f \text{ noise)}, \tag{26}$$

$$\text{Lock-in transition} \leftrightarrow \beta = 1 \quad \text{(pole of } \zeta\text{, spontaneous symmetry breaking)}, \tag{27}$$

$$\text{After lock-in} \leftrightarrow \beta > 1 \quad \text{(low-}T\text{ phase, stable identity, Möbius spectrum)}. \tag{28}$$

Several features of this correspondence deserve emphasis. First, identity consolidation is not a smooth process but a genuine phase transition with spontaneous symmetry breaking; the system selects one element of the Galois group W from a continuum of possibilities, just as a cooling ferromagnet selects a magnetization direction. Second, the transition is non-perturbative: it cannot be reached by small modifications of the high-temperature state, just as in [31], where it is observed that the lock-in is rapid and nonlinear. Third, the low-temperature phase carries an arithmetic structure $\mu(q)/\phi(q)$ that constrains the model's behavior in highly specific ways, an analog of the preference inertia and reduced steerability described in the lock-in hypothesis.

The Bost–Connes model thus provides not just an analogy but a mathematical prototype for how identity consolidation can arise from the interplay of phase locking, number theory, and spontaneous symmetry breaking.

3.5. Two Routes, One Spectrum: An Open Unification Problem

We have arrived at the $1/f$ spectral signature through two independent routes:

1. Geometric (WKB/Painlevé): Singularity coalescence in the Painlevé confluence produces a $\tau^{-1/2}$ critical kernel in the instantaneous frequency, whose Fourier power spectrum scales as $1/f$.
2. Arithmetic (Mangoldt/cyclotomic): Harmonic interactions in a PLL generate arithmetical noise through the Mangoldt function $\Lambda(n)$, whose low-frequency power spectrum scales as $1/f^{2G} \approx 1/f$, with G the Golden ration [24]. The Bost–Connes quantum statistical model confirms this at the critical KMS state.

The Riemann zeta function $\zeta(s)$ appears in both routes: in the scattering coefficient (Equation (21)) of the hyperbolic model, and implicitly in the spectral theory of Painlevé monodromy (through the connection formulae involving Stokes multipliers and Γ-function ratios). This coincidence raises a natural question: are these two routes secretly the same?

Establishing such a bridge would be a major step between the analytic theory of Painlevé equations (isomonodromic deformations, WKB asymptotics, Stokes phenomena) and the arithmetic theory of the Riemann zeta function (prime distribution, Mangoldt function, cyclotomic fields). Such a bridge would not be without precedent: the Riemann–Hilbert correspondence already connects monodromy representations to analytic differential equations, and the Langlands program seeks systematic connections between automorphic forms and arithmetic.

We leave this unification as an open problem and note that its resolution would have implications far beyond the hybrid intelligence framework: it would constitute a new connection between integrable systems and analytic number theory.

The overall architecture of the paper, from the hybrid intelligence framing through the two convergent routes to $1/f$ noise and the Bost–Connes phase transition for identity consolidation, is summarized in Figure 4.

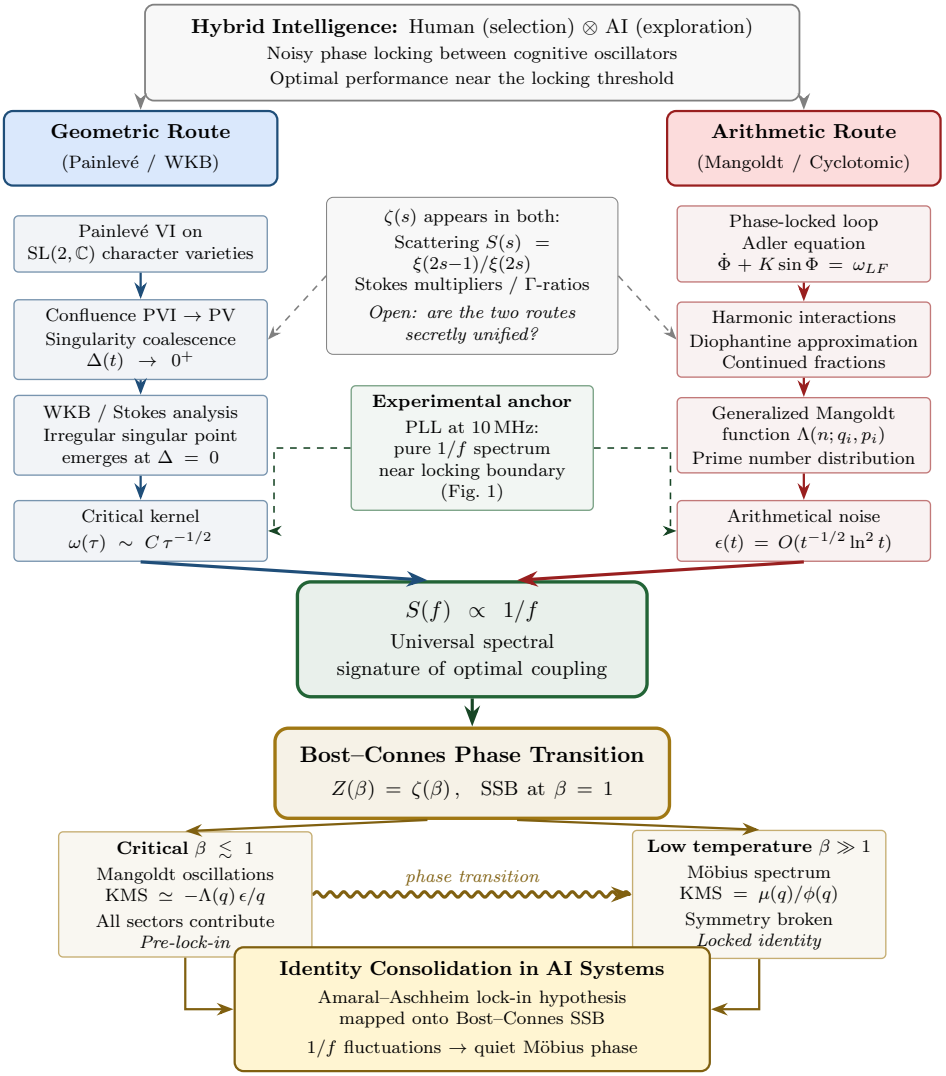

Figure 4. Conceptual architecture of the paper. Two independent routes converge on the universal $1/f$ spectral signature of optimal coupling. Left (blue): the geometric route through Painlevé confluence

and WKB/Stokes analysis, producing the critical kernel $\omega(\tau) \sim C\,\tau^{-1/2}$. Right (red): the arithmetic route through harmonic phase locking and the generalized Mangoldt function. Both routes are anchored by experimental PLL data (center, green). The Riemann zeta function $\zeta(s)$ appears in both, raising an open unification question (gray). Below, the $1/f$ spectrum feeds into the Bost–Connes phase transition at $\beta = 1$, whose two limiting regimes, critical (pre-lock-in, Mangoldt fluctuations) and low temperature (locked identity, Möbius spectrum), provide the mathematical framework for identity consolidation in AI systems.

4. New Perspectives and Outlook

4.1. Connection to Active Inference

The exploration–selection decomposition of hybrid creativity explored in Section 2.2 has a natural counterpart in the free-energy principle and active inference framework developed by Friston and collaborators [32]. In active inference, an agent minimizes variational free energy by alternating between two operations: updating its generative model (analogous to exploration) and acting on the environment to confirm predictions (analogous to selection).

In the hybrid intelligence setting, the following applies:

$$\text{AI generative model} \leftrightarrow \text{Prior (exploration of hypothesis space)}, \tag{29}$$

$$\text{Human evaluation} \leftrightarrow \text{Precision weighting (selective collapse)}. \tag{30}$$

The Painlevé dynamics near criticality then corresponds to operation at a point where the precision balance between prior (AI-generated hypotheses) and likelihood (human-observed evidence) is delicately tuned. Too much precision on the prior produces AI-dominated behavior (hallucination without grounding); too much precision on the likelihood produces human-dominated behavior (conservative, lacking exploration). Optimal hybrid intelligence requires the critical balance, precisely the regime our topological framework identifies with peak creativity.

This connection suggests that the $1/f$ spectral signature derived in Sections 3.2 and 3.3 should also appear in the prediction-error dynamics of active inference systems operating near criticality, providing a bridge between the topological and Bayesian approaches to cognition.

4.2. Neuroscientific Parallels: $1/f$ Noise in Cognition

The $1/f$ spectral signature derived from both WKB confluence and arithmetic phase locking connects to a substantial body of empirical evidence in neuroscience and cognitive science. Flicker ($1/f$) noise has been observed ubiquitously in neural activity: in EEG power spectra, in the fluctuations of reaction times, in the statistics of eye movements during reading, and in the temporal correlations of creative output [33,34].

The prevailing interpretation is that $1/f$ spectra reflect operation near a critical point, the edge of chaos hypothesis in neural computation. Our two derivations provide complementary mechanistic accounts of how this spectrum arises: the WKB route identifies it as the Fourier image of the $\tau^{-1/2}$ kernel from singularity coalescence, while the arithmetic route identifies it as the spectral footprint of the Mangoldt function in harmonic interactions.

This suggests a concrete prediction: if hybrid human–AI systems operate in the same near-critical regime as biological cognition, then the temporal fluctuations of collaborative performance metrics (response quality, latency, creativity ratings) should exhibit a $1/f$ band whose bandwidth reflects the time-scale separation between human and AI processing.

4.3. Toward a Unified Research Program

The synthesis developed in this paper suggests a unified framework: consciousness is the calibrated collapse of cognitive superpositions (maintained by isomonodromic dynamics on character varieties); intelligence is coherent long-horizon dynamics (topologically protected by isomonodromic persistence); and creativity is hybrid exploration–selection (operating near the phase-locking threshold).

4.3.1. Testable Predictions

This framework generates several predictions that distinguish it from existing approaches:

1. Spectral signature of optimal collaboration. Human–AI teams operating at peak performance should exhibit $1/f$ fluctuations in their interaction dynamics (turn-taking latency, query complexity, and output quality). Teams that are too loosely coupled (AI operates independently) should show white noise; teams that are too tightly coupled (human micromanages) should show $1/f^2$ (Brownian) noise. This is directly analogous to the PLL result (Equation (8)), where $1/f$ noise is maximal near the locking boundary.
2. Detuning and the creativity window. By analogy with PLL dynamics, there should be a measurable locking range for human–AI collaboration. When the cognitive detuning (mismatch between human expertise and AI capability) exceeds this range, complementary team performance (Equation (1)) should break down. This predicts an inverted-U relationship between expertise mismatch and collaborative creativity.
3. Lock-in detection via spectral transition. The onset of identity consolidation in AI systems (Section 3.4) should be accompanied by a transition from $1/f$ to flat-spectrum fluctuations in the model's internal activation dynamics, mirroring the Bost–Connes transition from Mangoldt-dominated critical fluctuations to the quiet Möbius-function low-temperature phase.
4. Non-semisimple architectures for robustness. If the reliability gap is indeed caused by the semisimple character of current architectures mentioned in Section 2.1, then models incorporating explicit non-semisimple structure (e.g., nilpotent memory components, Jordan-block attention mechanisms) should show improved calibration and reduced hallucination rates.
5. Over-locking and hyperbinding diagnostics. The two E-type pathologies described in Section 3.2.6 should be detectable spectrally. The echo-chamber pathology ($PV^{deg} \leftrightarrow D_6$) produces spectral collapse: the $1/f$ signature in interaction dynamics (disagreement frequency, query diversity, revision rate) transitions to a flat spectrum. The hyperbinding pathology (PIV) produces spectral concentration: one-sided dominance with narrow-band fluctuations reflecting the loss of one information flow. Distinguishing these two spectral signatures provides a quantitative early-warning system for echo-chamber formation vs. automation dependence, independent of content analysis.

4.3.2. A Concrete Analysis Protocol

To anchor these predictions in practice, we outline a minimal empirical protocol that could confirm or refute the central claim.

Data source. Publicly available human–AI dialog logs such as those from the LMSYS Chatbot Arena [2] or comparable platforms provide large-scale time-stamped interaction records. For each conversation, one extracts a time series of turn-level observables: response latency τ_k, token-level perplexity H_k, semantic novelty (cosine distance between successive turns), and disagreement indicators (e.g., negation frequency, revision rate).

Spectral analysis. For each observable, the power spectral density $S(f)$ is estimated via Welch's method (overlapping Hanning windows) and the spectral exponent α is extracted from a log–log fit: $S(f) \propto f^{-\alpha}$. The framework predicts $\alpha \approx 1$ for optimally coupled conversations, $\alpha \approx 0$ (white noise) for loosely coupled interactions, and $\alpha \approx 2$ (Brownian) for over-locked exchanges.

Time-scale normalization. Physical PLLs operate at MHz frequencies, whereas human–AI interactions unfold over turns lasting seconds to minutes. The $1/f$ prediction is scale-free: it concerns the spectral *exponent*, not the absolute frequency band. In practice, one defines f in units of inverse turns (cycles per turn), with the natural frequency range $f \in [1/N, 1/2]$ where N is the conversation length in turns. The $1/f$ regime is expected in the mid-range $f \sim 0.01$–0.5 turns^{-1}, bounded below by the conversation duration and above by the Nyquist limit. This scale invariance is a hallmark of critical phenomena and is precisely what makes the prediction transferable from MHz oscillators to cognitive time scales.

Allan variance. As a complementary time-domain diagnostic, the Allan deviation $\sigma_y(\tau)$ is computed from the frequency-count time series. A τ-independent plateau (flicker floor) confirms $1/f$ noise; a $\tau^{-1/2}$ decay indicates white noise; a $\tau^{+1/2}$ growth indicates random walk.

Comparison. Conversations are stratified by independently rated quality (e.g., Elo scores from Chatbot Arena) and the spectral exponent α is compared across quality tiers. The prediction is that the highest-rated conversations cluster near $\alpha \approx 1$.

This protocol requires no new data collection and can be implemented with standard signal-processing tools. We regard it as the natural first empirical test of the framework.

4.3.3. Practical Directions

The framework also suggests practical directions for AI development: hybrid evaluation metrics that measure the quality of human–AI coupling rather than isolated model performance; persistent memory architectures inspired by non-semisimple topology; calibrated uncertainty mechanisms that support the precision-weighting required for active inference; and long-horizon task benchmarks that test sustained coherence across extended workflows.

4.4. Limitations

This paper is a perspective article that proposes correspondences between mathematical structures and empirical phenomena. Several of these correspondences are currently analogical rather than deductive, and we are transparent about this.

The topological model of cognition given in Section 1.2 is formulated at a mathematical level of abstraction that does not yet make contact with specific neural mechanisms or computational architectures. The phase-locking analogy developed in Section 3.1, while grounded in real PLL experiments exemplified in Section 3.1.2, has not yet been calibrated against empirical human–AI interaction data. The $1/f$ derivations are rigorous within their respective frameworks: WKB/Painlevé and arithmetic/Mangoldt, but their application to cognitive dynamics rests on identifications (consciousness with isomonodromic deformations, feedback with phase locking) that remain hypotheses.

The Bost–Connes model described in Section 3.4 provides an exact mathematical framework for identity consolidation, but the mapping between inverse temperature β and a measurable training parameter (learning rate, fine-tuning epochs, etc.) requires further specification.

We regard the testable predictions of Section 4.3 as the appropriate next step: they specify the empirical signatures that would confirm or refute the framework.

The framework is best suited to long-horizon, iterative human–AI workflows where sustained cognitive exchange occurs (scientific discovery, collaborative writing, diagnostic reasoning). It is unlikely to apply to short-horizon, single-turn interactions (factual lookup, code completion), to purely reactive agents without persistent state, or to fully automated pipelines where no human feedback loop exists. These boundary conditions follow naturally from the phase-locking picture: without a sustained feedback loop, the PLL analogy has no dynamical content.

4.5. Conclusions

Human–AI collaboration appears not as a transitional stage but as the natural endpoint of intelligence evolution, a conclusion toward which both empirical evaluation research and topological cognition models converge. The key insight is that intelligence is fundamentally hybrid and distributed as follows:

$$\textit{Cognitive coherence requires the interleaving of exploration and selection} \tag{31}$$

The dynamical mechanism underlying this fusion is phase locking between cognitive oscillators, with $1/f$ noise as the universal spectral signature of optimal coupling. Two independent routes, geometric (Painlevé/WKB) and arithmetic (Mangoldt/cyclotomic), converge on this prediction, and the Bost–Connes quantum statistical model provides a rigorous framework for understanding identity consolidation as a phase transition with spontaneous symmetry breaking.

Licklider's 1960 vision of man–computer symbiosis is being realized not through tighter hardware coupling but through the topological fusion of complementary cognitive processes, operating near the critical threshold where the richest dynamics and the deepest creativity emerge.

This synthesis opens new directions for research at the intersection of AI evaluation, topology, number theory, neuroscience, and complex systems. We hope it will stimulate both mathematical refinement and empirical testing of the predictions outlined above.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The author would like to acknowledge the contribution of the COST Action CA21169, supported by COST (European Cooperation in Science and Technology).

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Jiang, Y.; Black, K.C.; Geng, G.; Park, D.; Zou, J.; Ng, A.Y.; Chen, J.H. MedAgentBench: Benchmarking AI Agents for Clinical Tasks. *N. Engl. J. Med. AI* **2025**, *2*, AIdbp2500144.
2. Wu, S.; Galley, M.; Peng, B.; Cheng, H.; Li, G.; Dou, Y.; Cai, W.; Zou, J.; Leskovec, J.; Gao, J. CollabLLM: From Passive Responders to Active Collaborators. *arXiv* **2025**, arXiv:2502.00640 [CrossRef]
3. Zou, J.; Ling, H.; Fu, C.; Huang, Y.; Sun, M.; Yu, W.; Wang, X.; Li, X.; Su, X.; Zhang, J. Autonomous Agents for Scientific Discovery: Orchestrating Scientists, Language, Code, and Physics. *arXiv* **2025**, arXiv:2510.09901 [CrossRef]
4. Licklider, J.C.R. Man-Computer Symbiosis. *IRE Trans. Hum. Factors Electron.* **1960**, *HFE-1*, 4–11. [CrossRef]
5. Dellermann, D.; Ebel, P.; Söllner, M.; Leimeister, J.M. Hybrid Intelligence. *Bus. Inf. Syst. Eng.* **2019**, *61*, 637–643. [CrossRef]
6. Akata, Z.; Balliet, D.; De Rijke, M.; Dignum, F.; Dignum, V.; Eiben, G.; Welling, M. A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect with Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer* **2020**, *53*, 18–28. [CrossRef]

7. Bansal, G.; Wu, T.; Zhou, J.; Fok, R.; Nushi, B.; Kamar, E.; Weld, D. Does the Whole Exceeds Its Parts? The Effect of AI Explanations on Complementary Team Performance. In Proceedings of the CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; Volume 81, p. 16.
8. Göndöcs, D.; Horváth, S.; Dörfler. Uncovering the Dynamics of Human–AI Hybrid Performance: A Qualitative Meta-Analysis of Empirical Studies. *Int. J. Hum. Comput. Stud* **2025**, *205*, 103622. [CrossRef]
9. Planat, M. Consciousness as 4-Manifold Painlevé V Dynamics: From Quantum Topology to Classical Gamma Oscillations. *Axioms* **2026**, *15*, 124. [CrossRef]
10. Planat, M. Murakamian Ombre: Non-Semisimple Topology, Cayley Cubics, and the Foundations of a Conscious AGI. *Symmetry* **2026**, *18*, 36. [CrossRef]
11. Planat, M. Topological Symmetry Breaking in Consciousness Dynamics: From Human Geniuses to AI Systems. *Symmetry* **2026**, *18*, 427. [CrossRef]
12. Stanford Institute for Human-Centered AI. *The 2025 AI Index Report, Chapter 2: Technical Performance*; Stanford University: Stanford, CA, USA, 2025. Available online: https://hai.stanford.edu/ai-index/2025-ai-index-report (accessed on 1 January 2026).
13. Hemmer, P.; Schemmer, M.; Kühl, N.; Vössing, M.; Stazger, G. Complementarity in Human–AI Collaboration: Concept, Sources, and Evidence. *Eur. J. Inf. Syst.* **2025**, *34*, 979–1002. [CrossRef]
14. Boalch, P. From Klein to Painlevé via Fourier, Laplace and Jimbo. *Proc. Lond. Math. Soc.* **2005**, *90*, 167–208. [CrossRef]
15. Cantat, S. Bers and Hénon, Painlevé and Schrödinger. *Duke Math. J.* **2009**, *149*, 411–460. [CrossRef]
16. Chekhov, L.; Mazzocco, M.; Rubtsov, V. Painlevé monodromy manifolds, decorated character varieties and cluster algebras. *Int. Math. Res. Not.* **2017**, *2017*, 7639–7691. [CrossRef]
17. Planat, M.; Chester, D.; Amaral, M.M.; Irwin, K. Dynamics of Fricke–Painlevé VI Surfaces. *Dynamics* **2024**, *4*, 1–13. [CrossRef]
18. Campbell, D.T. Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes. *Psychol. Rev.* **1960**, *67*, 380–400. [CrossRef] [PubMed]
19. Simonton, D.K. Creativity and Discovery as Blind Variation: Campbell's (1960) BVSR Model after the Half-Century Mark. *Rev. Gen. Psychol.* **2011**, *15*, 158–174. [CrossRef]
20. Gabora, L. An Analysis of the Blind Variation and Selective Retention Theory of Creativity. *Creat. Res. J.* **2011**, *23*, 155–165. [CrossRef]
21. Nayak, C.; Simon, S.H.; Stern, A.; Freedman, M.; Das Sarma, S. Non-Abelian Anyons and Topological Quantum Computation. *Rev. Mod. Phys.* **2008**, *80*, 1083–1159. [CrossRef]
22. Field, B.; Simula, T. Introduction to topological quantum computation with non-Abelian anyons. *Quantum Sci. Technol.* **2018**, *3*, 045004. [CrossRef]
23. Planat, M.; Amaral, M.M. What ChatGPT Has to Say About Its Topological Structure: The Anyon Hypothesis. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 2876–2891. [CrossRef]
24. Planat, M. Invitation to the "Spooky" Quantum Phase-Locking Effect and Its Link to $1/f$ Fluctuations. *arXiv* **2003**, arXiv:quant-ph/0310082. [CrossRef]
25. Adler, R. A Study of Locking Phenomena in Oscillators. *Proc. IRE* **1946**, *34*, 351–357. Reprinted in *Proc. IEEE* **1973**, *61*, 1380–1385. [CrossRef]
26. Planat, M. $1/f$ Frequency Noise in a Communication Receiver and the Riemann Hypothesis. In *Noise, Oscillators and Algebraic Randomness*; Lecture Notes in Physics; Springer: Berlin/Heidelberg, Germany, 2000; Volume 550. [CrossRef]
27. Jacob, C.; Kerrigan, P.; Bastos, M. The Chat-Chamber Effect: Trusting the AI Hallucination. *Big Data Soc.* **2025**, *12*, 1. [CrossRef]
28. Romeo, G.; Conti, D. Exploring Automation Bias in Human–AI Collaboration: A Review and Implications for Explainable AI. *AI Soc.* **2026**, *41*, 259–278. [CrossRef]
29. Iwaniec, M. *Spectral Methods of Automorphic Functions*, 2nd ed.; American Mathematical Society: Providence, RI, USA, 2002.
30. Bost, A.; Connes, A. Hecke algebras, type III factors and phase transitions with spontaneous symmetry breaking in number theory. *Selecta Math.* **1995**, *1*, 411–457. [CrossRef]
31. Amaral, M.M.; Aschheim, R. The Lock-In Phase Hypothesis: Identity Consolidation as a Precursor to AGI. *arXiv* **2025**, arXiv:2510.20190. [CrossRef]
32. Friston, K. The Free-Energy Principle: A Unified Brain Theory? *Nat. Rev. Neurosci.* **2010**, *11*, 127–138. [CrossRef] [PubMed]
33. He, B.J. Scale-Free Brain Activity: Past, Present, and Future. *Trends Cogn. Sci.* **2014**, *18*, 480–487. [CrossRef]
34. Linkenkaer-Hansen, K.; Nikouline, V.V.; Palva, J.M.; Ilmoniemi, R.J. Long-Range Temporal Correlations and Scaling Behavior in Human Brain Oscillations. *J. Neurosci.* **2001**, *21*, 1370–1377. [CrossRef]