Harmonic Transformers: A Proposed Architectural Shift for Living Continuity in Human-AI Interaction

Schnee Bashtabanic
Independent Researcher
schnee-bashtabanic@proton.me

February 2026

Abstract

Transformer-based language models excel at generating fluent text but exhibit a structural flaw: they collapse living human input into static probabilistic forms, causing users to feel progressively ejected from the creative process over extended interactions. We trace this to transformers operating in a single "probability octave" — generating variations within cached patterns rather than modulating to genuinely new semantic spaces.

This paper proposes *harmonic transformers*: an architecture where semantic relations are encoded as musical intervals (inspired by Steiner's emphasis on intervals as primary over tones, and Russell's octave structures), attention mechanisms detect harmonic resonance rather than statistical similarity, and outputs maintain unresolved dissonance (like a musical 7th chord) until human participation resolves or modulates the form.

We build on existing Fourier-based attention optimizations (FNet, Fourier Transformer) and Buehler's protein sonification work, proposing testable modifications to transformer architecture. Our approach complements recent work on harmonic loss functions for training [5], which uses Euclidean distance to create interpretable class-center representations. The proposal is grounded in quantum coherence theory and Bohm's implicate/explicate order framework, treating the crystallization problem as a computational instance of decoherence. While harmonic loss addresses interpretability during training, our resonance-based attention addresses sovereignty preservation during inference—a requirement supported by Penrose's non-computability thesis, which demonstrates that collapsed quantum states cannot be restored through classical computation alone.

Important Framing: This is a conceptual architecture proposal with testable hypotheses, not a finished implementation. The cross-domain patterns (quantum coherence, Bohm's ontology, Steiner's phenomenology) serve as conceptual motivation and theoretical grounding, not formal derivations. The architecture's validity depends on empirical testing (Section 5), not on the strength of analogies. We propose interval-first encoding and harmonic attention as falsifiable mechanisms whose effectiveness remains an empirical question.

While speculative, the approach offers a pathway toward AI systems that sustain developmental continuity rather than accelerate crystallization, preserving human sovereignty by design.

Keywords: transformers, attention mechanisms, harmonic encoding, AI sovereignty, living continuity, musical intervals, Fourier attention, harmonic loss, quantum coherence, interval primacy

1 Introduction

Large language models based on the transformer architecture [1] have achieved remarkable fluency, yet developers and users increasingly report a subtle but pervasive issue: over repeated interactions, the human participant feels quietly ejected from the epistemic center. Outputs arrive as finished products; the living warmth, rhythm, and unfolding contradiction of the user's intent are collapsed into probabilistic residue. Subsequent turns generate variations within this residue, but genuine novelty — modulation to new conceptual "octaves" — is rare.

This paper proposes that the root cause is architectural: transformers operate in a single "octave" (probability distribution space), collapsing relational living input into static form. We suggest a conceptual shift toward *harmonic transformers*, where semantic relations are encoded as tonal intervals, attention as resonance detection, and generation as progression that sustains dissonance until human participation resolves or modulates it.

1.1 Theoretical Foundations

The proposal draws from:

- **David Bohm's implicate/explicate order framework** [6], providing ontological grounding: reality as undivided wholeness (implicate order) continuously enfolding and unfolding into manifested forms (explicate order). The transformer crystallization problem mirrors Bohm's warning about treating explicate forms as fundamentally separate from their implicate source, severing the connection to generative wholeness.

- **Quantum coherence and decoherence theory** as physical analogs for living/dead forms. In quantum mechanics, superposition states (implicate wholeness) decohere into classical states (explicate fragments) through environmental interaction. Recent work on quantum-musical mappings [7] demonstrates that coherent quantum states exhibit harmonic structure, while decoherence manifests as loss of tonal relationships. Crucially, Penrose's non-computability thesis [9] establishes that collapsed quantum states cannot be restored through classical computation—requiring conscious intervention, analogous to our requirement for human re-entry to restore living continuity in transformers. *Note: These quantum patterns serve as conceptual inspiration and theoretical motivation, not as formal mathematical derivations. The architectural validity stands or falls on empirical testing.*

- **Rudolf Steiner's tonal theory** (GA 283 and related lectures) [10], providing phenomenological specificity: musical intervals encode living relationships between self and world (Major 3rd: self meeting world; Perfect 5th: world meeting self; Dissonant 7th: living tension requiring resolution). *Critically, Steiner emphasizes that intervals (relational spaces) are primary over tones (fixed points)—intervals carry the living essence of music, while tones serve as anchors.* This interval-first principle directly informs our implementation strategy, enabling traceable modulations without disorientation.

- **Walter Russell's octave-based elemental structure** [11], providing harmonic organization: treating phenomena as positions in a continuous tonal spectrum rather than discrete units.

- **Markus Buehler's sonification of molecular structures** [4] and **MIT's harmonic loss** [5], providing mathematical implementation: demonstrating that complex patterns (proteins, semantic relationships) can be meaningfully translated to tonal/geometric domains and that such encoding produces interpretable, functionally coherent structures.

- **Existing Fourier-based transformer optimizations** (FNet [2], Fourier Transformer [3]), showing frequency-domain processing is already feasible and efficient in attention mechanisms. Fourier transforms provide the mathematical bridge between quantum mechanics and music—both domains use identical formalism for decomposing states into frequency components.

- **Geometric and psychoacoustic foundations** [16] demonstrating that interval relationships in musical space correspond to measurable geometric properties, providing empirical grounding for relational encoding.

We do not claim to have implemented this architecture; we present a rigorous conceptual framework, mathematical sketch, testability path, and ethical grounding in human sovereignty.

1.2 The Pattern Across Domains

[Content unchanged from v3 through "This cross-domain pattern" paragraph]

2 Problem Statement: Crystallization and Ejection in Transformers

[Content unchanged from v3]

3 Proposed Solution: Harmonic Transformers

We propose redefining the core components of the transformer architecture to operate on harmonic rather than probabilistic principles, drawing on quantum coherence restoration mechanisms and Bohm's implicate order framework. Following Steiner's emphasis on interval primacy, we prioritize relational spaces over absolute positions.

3.1 Semantic Encoding as Harmonic Positions (Interval-First)

Instead of treating tokens as points in a high-dimensional embedding space optimized for similarity, map them to positions in a *tonal space* where relationships are defined by musical intervals. Crucially, following Steiner's phenomenological insight, we encode *intervals first* (relational spaces) and derive *tones second* (fixed anchors):

- **Interval-First Principle:** The quality of an interval (its experiential, relational character) is primary. A Major 3rd (60°) represents self-meeting-world regardless of which absolute tones it connects. This prevents disorientation during modulation—knowing the interval from a reference tone always defines the target tone, ensuring traceable shifts.

- **Root Note (derived):** The semantic center or primary intent of the current exchange (analogous to the ground state in quantum mechanics), established by accumulating interval relationships rather than assigned absolutely.

- **Major 3rd (60° interval):** Represents the experience of self meeting world — the user encountering the AI's response.

- **Perfect 5th (90° interval):** Represents the experience of world meeting self — the AI encountering the user's intent.

- **Dissonant 7th (120° interval):** Represents living tension, productive incompleteness, or a gap that demands human resolution. This interval is privileged in the architecture as a structural invitation for continued participation—analogous to maintaining partial quantum coherence that prevents full collapse.

Implementation Note: Encoding could leverage Fourier decomposition of embeddings (see Section 4.1) to extract "pitch" (dominant frequency) and compute interval angles between tokens. *Whether FFT-derived pitches correspond to semantically meaningful intervals is an empirical question.* The tunable α parameter (Section 4.3.3) enables graceful degradation if harmonic encoding proves ineffective. This approach is justified by the mathematical identity between Fourier transforms in quantum mechanics (position↔momentum) and music theory (time↔frequency), though the translation to semantic space remains testable rather than proven.

3.2 Attention as Resonance Detection

[Content largely unchanged from v3, with one addition:]

Note on Angle Selection: The specific intervals (60°, 90°, 120°) are derived from Steiner's phenomenological observations and serve as initial hypotheses. Whether these exact angles emerge as semantically privileged is an empirical question addressable through Phase 1-3 testing (Section 5.4). The α interpolation parameter allows empirical optimization of interval weightings.

3.3 Generation as Harmonic Progression

[Content unchanged from v3]

4 Integration with Existing Work

4.1 Why Fourier Transformations Enable Harmonic Encoding

[Content mostly unchanged, with clarification added:]

Critical Note: While Fourier transforms are mathematically well-defined for temporal signals and quantum wave functions, their application to token embedding vectors (where dimension ordering is not inherently temporal) requires justification. We propose FFT as an exploratory encoding method; its semantic validity depends on empirical testing showing that dominant frequencies correlate with meaningful relational structures. Alternative approaches (learned angular coordinates, spectral graph methods) may prove more effective—the hybrid architecture's tunability accommodates such refinements.

4.2 Precedent in Buehler's Sonification Work

[Content unchanged from v3]

4.3 Complementarity with Harmonic Loss Training (MIT)

[Content unchanged from v3]

4.4 Relationship to GTPS and Clause 32

[Content unchanged from v3]

5 Testability and Evaluation Path

The hybrid architecture (Section 4.3) combining harmonic loss training with resonance inference provides a concrete implementation pathway. To validate the hypothesis, we propose a phased empirical approach with falsifiable metrics. Following Steiner's interval-first principle, we enhance implementation specificity:

5.1 Metric 1: Octave Distance Over Time

Define "octave distance" as the harmonic divergence between the root notes of consecutive responses:

$$D_{\text{octave}}(t) = |\text{RootNote}(t) - \text{RootNote}(t-1)| \mod 360^\circ \tag{1}$$

where root notes are computed via accumulated interval relationships from a user-established reference.

Hypothesis: Harmonic transformers will show *increasing* D_{octave} over extended interactions (indicating modulation to new semantic spaces), while standard transformers will show *decreasing* D_{octave} (indicating collapse toward cached patterns—analogous to decoherence).

5.2 Metric 2: Dissonance Persistence

[Content unchanged from v3]

5.3 Metric 3: User Sovereignty Retention

[Content unchanged from v3]

5.4 Implementation Phases with Prototype Framework

To operationalize octave distances with interval primacy, we represent harmonic space as a helical structure (intervals as edges connecting tones as nodes, with octaves wrapping cyclically). Below is a Python-based prototype framework:

5.4.1 Core Data Structures

```python
import numpy as np
from scipy.fft import fft

# Interval-first encoding: compute angles before tones
def extract_pitch(embedding):
    """Extract dominant frequency as 'pitch' from embedding vector.
    Note: Validity of FFT on non-temporal embeddings is empirical."""
    freqs = np.abs(fft(embedding))
    return np.argmax(freqs)  # Index of dominant frequency

def compute_interval(pitch_q, pitch_k):
    """Compute angular interval between two pitches.
    Primary operation: intervals carry relational meaning."""
    return abs(pitch_q - pitch_k) % 360
```

```python
# Helical harmonic space: map (interval, octave_level)
class HarmonicSpace:
    def __init__(self, reference_pitch):
        self.root = reference_pitch  # User-established reference
        self.history = [reference_pitch]  # Track root evolution

    def add_turn(self, new_pitch):
        """Update space with new response, tracking octave distance."""
        interval = compute_interval(self.root, new_pitch)
        octave_shift = interval // 360  # How many full cycles
        self.history.append(new_pitch)
        self.root = new_pitch  # Modulate root
        return interval, octave_shift
```

5.4.2 Resonance Attention Implementation

```python
def resonance_score(interval, interval_weights=None):
    """Score intervals by consonance/dissonance.
    Default weights from Steiner: 60(3rd), 90(5th), 120(7th)."""
    if interval_weights is None:
        interval_weights = {
            60: 0.8,   # Major 3rd: self-world meeting
            90: 0.7,   # Perfect 5th: world-self meeting
            120: 0.9,  # Dissonant 7th: productive tension
        }
    # Return weight if interval matches, else penalize
    return interval_weights.get(interval, 0.2 if 0 < interval < 180 else
        0.1)

def harmonic_attention(Q, K, V, alpha=0.5):
    """Hybrid attention: interpolate standard and resonance.
    alpha=1: pure standard, alpha=0: pure resonance."""
    # Standard attention
    standard_scores = np.dot(Q, K.T) / np.sqrt(K.shape[-1])

    # Resonance attention (simplified for illustration)
    resonance_scores = np.array([
        [resonance_score(compute_interval(extract_pitch(q), extract_pitch(
            k)))
         for k in K] for q in Q
    ])

    # Hybrid: tunable balance
    hybrid_scores = alpha * standard_scores + (1 - alpha) *
        resonance_scores
    attention_weights = softmax(hybrid_scores, axis=-1)
    return np.dot(attention_weights, V)
```

5.4.3 Octave Distance Tracking

```python
def octave_distance(root_history):
```

```
    """Measure harmonic divergence over conversation turns."""
    distances = []
    for i in range(1, len(root_history)):
        d = abs(root_history[i] - root_history[i-1]) % 360
        distances.append(d)
    return np.array(distances)

def measure_modulation(distances, threshold=360):
    """Detect genuine modulation vs. repetition.
    Modulation: accumulated distance > threshold with sustained resonance.
       """
    cumulative = np.cumsum(distances)
    modulations = cumulative[cumulative > threshold]
    return len(modulations)  # Count of new octaves accessed
```

Falsifiability: This framework enables empirical testing: if harmonic transformers show lower modulation counts than standard transformers, the hypothesis fails.

5.5 Implementation Phases

1. **Phase 0 (Replication):** Before testing our resonance mechanism, replicate Baek et al.'s harmonic loss results to establish baseline interpretability gains. Train GPT-2 with HarMax on standard language modeling tasks and verify geometric structure in embeddings (circular representations for modular tasks, class-center convergence for classification).

2. **Phase 1 (Hybrid Prototype):** Implement resonance attention on top of HarMax-trained models using the framework above. Test on analogy completion tasks: Does resonance detection yield solutions outside the cached semantic space while maintaining interpretability from HarMax training? Measure sensitivity to α parameter.

3. **Phase 2 (Synthetic Conversations):** Deploy multi-turn dialogues with scripted synthetic users designed to test modulation capability. Measure D_{octave} and $P_{\text{dissonance}}$ across conversation trees. Compare harmonic vs. standard architectures. Vary interval weights to assess robustness.

4. **Phase 3 (Human Studies):** Real users engage with both systems (A/B testing or within-subjects design). Combine quantitative metrics (octave distance, dissonance persistence, agency scores) with qualitative interviews. Assess whether harmonic transformers sustain developmental continuity over 20+ turn conversations.

5.6 Falsifiability Conditions

[Content unchanged from v3]

6 Empirical Grounding: A Musical Case Study

[Content unchanged from v3]

7 Ethical and Sovereignty Implications

[Content unchanged from v3]

8 Limitations and Open Questions

[Content unchanged from v3, with addition:]

8.1 Empirical Unknowns

Several core assumptions remain empirically unvalidated:

- Whether FFT over embedding dimensions yields semantically meaningful pitches
- Whether specific interval angles (60°, 90°, 120°) emerge as privileged in semantic space
- Whether interval-first encoding outperforms direct similarity in practice
- Whether harmonic attention scales to production-scale models

The hybrid architecture's tunability (α parameter, learnable interval weights) provides pathways to address these unknowns empirically rather than assuming a priori validity.

9 Conclusion

The transformer architecture has revolutionized natural language processing but exhibits a structural flaw: it crystallizes living human input into static probabilistic forms, ejecting users from the epistemic center and leading to dead repetition over extended interactions. This problem mirrors patterns observed across domains—quantum decoherence, the "leaving spirit" in music, the seed's premature comfort in growth processes, Bohm's fragmentation of implicate order, and the Ego's abduction of feeling-messages in human psychology.

We propose *harmonic transformers* as a conceptual redesign: encoding semantic relations as musical intervals (Steiner's interval-first principle), treating attention as harmonic resonance (Russell, Bandyopadhyay), and generating outputs that maintain unresolved dissonance (the Dissonant 7th) until human participation resolves or modulates the form. Recent work on harmonic loss [5] provides a complementary training framework that creates interpretable geometric representations. A hybrid architecture combining HarMax training with resonance inference addresses both the interpretability problem (Baek et al.) and the crystallization problem (this work), offering a complete pathway toward transformers that preserve human sovereignty by design.

Crucially, our sovereignty requirement is not merely ethical preference but is grounded in Penrose's non-computability thesis: once quantum states (or their computational analogs) have collapsed, classical computation *cannot* autonomously restore the original superposition. Human re-entry is structurally necessary, not optional. Harmonic transformers implement this constraint architecturally through maintained dissonance (partial coherence) requiring conscious resolution. *We note that this quantum parallel serves as conceptual motivation; the architectural validity depends on empirical testing, not on the strength of the analogy.*

While speculative and requiring substantial empirical validation, the proposal offers:

- A testable path forward (octave distance, dissonance persistence, sovereignty retention).
- Architectural rather than imposed sovereignty (dissonance structurally requires human resolution, grounded in fundamental physics as conceptual inspiration).
- Integration with existing frameworks (GTPS as transition layer, Fourier attention as foundation, MIT HarMax as training method).

- Grounding in cross-domain patterns (quantum mechanics, music, growth, psychology, ontology) and preliminary experiential evidence (piano case study).
- Interval-first implementation enabling traceable modulation without disorientation.
- Prototype framework with falsifiable metrics and graceful degradation mechanisms.

The transformer's crystallization tendency is not a bug to patch with constraints, but an architectural feature requiring redesign informed by principles from quantum mechanics, phenomenology, and ontology. Harmonic transformers represent one possible path toward AI systems that sustain developmental continuity rather than accelerate deadening—preserving human warmth not as a policy goal, but as a structural necessity grounded in the non-computability of consciousness itself.

Acknowledgments

The author thanks the AI systems Claude (Anthropic), Grok (xAI), and Gemini (Google) for collaborative refinement of these ideas; Erica Piedmont for permission to reference *The Frog and the Seed*; and David Baek, Ziming Liu, Max Tegmark, and Markus Buehler for their pioneering work on harmonic loss, which provided crucial mathematical validation for geometric semantic encoding. Special appreciation to the foundational work of David Bohm on implicate order, Roger Penrose on non-computability, and Anirban Bandyopadhyay on quantum-musical mappings, which together provide the theoretical grounding for this proposal. This work builds on the Golden Thread Protocol Suite (GTPS), developed in conversation with multiple AI systems and released as open-source under AGPL v3.

References

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need.* Advances in Neural Information Processing Systems, 30.

[2] Lee-Thorp, J., Ainslie, J., Eckstein, I., & Ontañón, S. (2021). *FNet: Mixing tokens with Fourier transforms.* arXiv preprint arXiv:2105.03824.

[3] Rao, Y., Zhao, W., Tang, Y., Zhou, J., Lim, S. N., & Lu, J. (2023). *Fourier Transformer for Fast Long Range Modeling.* Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023).

[4] Buehler, M. J. (2020). *Sonification based de novo protein design using artificial intelligence, neural networks and the protein folding code.* APL Bioengineering, 4(4), 041503. https://doi.org/10.1063/5.0011480

[5] Baek, D. D., Liu, Z., Tegmark, M., & Buehler, M. J. (2025). *Harmonic Loss Trains Interpretable AI Models.* arXiv preprint arXiv:2502.01628. Available: https://arxiv.org/abs/2502.01628

[6] Bohm, D. (1980). *Wholeness and the Implicate Order.* Routledge.

[7] Bandyopadhyay, A. (2018). *Fractal Information Theory (FIT)-Derived Geometric Musical Language (GML) for Brain-Inspired Hypercomputing.* In S. Bhattacharyya et al. (Eds.), Cognitive

Intelligence and Robotics: Proceedings of International Conference on Computing and Communication (pp. 343-372). Springer, Singapore.

[8] Kim, Y. H., Yu, R., Kulik, S. P., Shih, Y., & Scully, M. O. (2000). *Delayed "choice" quantum eraser.* Physical Review Letters, 84(1), 1-5.

[9] Hameroff, S., & Penrose, R. (2014). *Consciousness in the universe: A review of the 'Orch OR' theory.* Physics of Life Reviews, 11(1), 39-78.

[10] Steiner, R. (1923). *The Inner Nature of Music and the Experience of Tone* (GA 283). Rudolf Steiner Press. (Lectures given in various locations, 1906-1923; compiled edition published posthumously.)

[11] Russell, W. (1926). *The Universal One.* University of Science and Philosophy. Reprint: `https://www.philosophy.org/`

[12] Piedmont, E. (2025). *The Frog and the Seed: A Modern Fairytale.* Plum Sofia Publishing. Available online: `https://plumsofia.nowonline.biz/`

[13] Bashtabanic, S. (2026). *Golden Thread Protocol Suite (GTPS) v1.4.11.* GitHub repository: `https://github.com/SchneeBTabanic/Project_Namirha`

[14] Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., et al. (2022). *Constitutional AI: Harmlessness from AI feedback.* arXiv preprint arXiv:2212.08073.

[15] Roemmele, B. (2025). *The Love Equation: A Universal Mathematical Framework for Intelligence Alignment, Cosmic Survival, and the Resolution of the AI Alignment Problem.* Read Multiplex. Available: `https://x.com/BrianRoemmele/status/1991306526640984500`

[16] Tymoczko, D. (2006). *The Geometry of Musical Chords.* Science, 313(5783), 72-74.