Excellent. I will conduct a rigorous exploration of a self-referential temporal navigation framework grounded in emergent coordinate systems, luminosity-based event intensity, and quantum propagation principles. I will integrate marginalized theories, derive complete mathematical formulations, and include testable models, derivations, and simulation blueprints where appropriate.

I'll return with a detailed, formalized document including equations, derivations, and conceptual mappings based on the theoretical structure you provided. Expect a report that embraces edge-case reasoning and challenges mainstream assumptions.

Rigorous Formalization of a Self-Referential Temporal Navigation System

Introduction

Understanding time as an **emergent, relational construct** rather than an absolute background requires rethinking conventional models. We propose a novel framework where each event defines its own temporal reference, navigation is guided by an intensity-like **temporal luminosity**, and quantum superposition of paths is inherently integrated. This report develops a full mathematical formalization of these ideas, incorporating fractal geometry, chaos theory, and alternative quantum interpretations. We include derivations for key formulas and outline how historically marginalized concepts (e.g. near-zero probability events, non-linear time, retrocausal loops) can be integrated and tested. The work is structured in sections **A** through **E**, covering theoretical formulations, interdisciplinary integration, testable models, specific formulas, and validation methodology, respectively.

Principles: (1) *Self-Referential Temporal Coordinates* – events establish

their own time frames via their mutual relationships, without assuming a global clock. (2) *Luminosity-Based Navigation* – an intensity metric ("temporal luminosity") quantifies the strength of inter-event connections and guides navigation in event-space similar to how a photometer guides navigation by light intensity. (3) *Quantum Propagation Integration* – all possible event trajectories exist in superposition, with high-likelihood paths emerging upon observation (analogous to wavefunction collapse). We draw on insights from Benoit Mandelbrot (fractal structures), Edward Lorenz (chaotic dynamics), and others to ensure the model captures complexity beyond standard linear physics.

The following sections present the detailed mathematical framework. Section A introduces core formulations: defining self-referential time coordinates, a relational intensity measure, probability distributions over simultaneous paths, and fractal (Cantor set) modeling of high-probability trajectories. Section B integrates non-mainstream ideas – preserving nearly-zero probabilities instead of discarding them, exploring alternative quantum interpretations (Many-Worlds, pilot wave, etc.), acknowledging non-linear and cyclic time concepts from various traditions, and incorporating "edge" phenomena like retrocausality or closed timelike loops. Section C builds testable models, including a diffusion equation for luminosity propagation, a mixture-of-experts computational architecture to combine multiple theories, self-supervised learning to detect emergent structures, and specific validation protocols. Section D provides explicit formulas for converting intensities to navigable directions, computing optimal event-paths and their probability distributions, and mapping the event-based coordinates to observable quantities. Finally, Section E outlines a validation roadmap with predictions that align with known quantum results, proposals for new experiments, error estimates, and an analysis of potential paradoxes (with resolutions ensuring logical consistency).

By treating time as an **inherently relational and fractal phenomenon**, this framework bridges discrete events with continuous trajectories, classical determinism with quantum indeterminacy, and mainstream

physics with visionary insights from transformative thinkers. The result is a self-consistent model that not only broadens our conceptual understanding of time and causality but also yields **testable predictions** and formulas, paving the way for experimental exploration of event-driven temporal navigation.

A. Mathematical Formulation of Core Concepts

A1. Emergence of Self-Referential Temporal Coordinates

In our framework, **events** are the primary entities that define time. Formally, let $\mathcal{E} = \{E_i\}$ be the set of all events. We do *not* assume a global time parameter t ordering these events. Instead, each event E_i establishes its own **temporal coordinate system** (E_i, τ_i) in which E_i itself is the origin ($\tau_i = 0$ for event E_i). Any other event E_j can be assigned a time coordinate in E_i's frame, denoted $\tau_i(E_j)$, determined purely by the relational dynamics between E_i and E_j. In other words, $\tau_i : \mathcal{E} \to \mathbb{R}$ is a time coordinate function defined by event E_i. These coordinates are *self-referential*: they are defined by the occurrence of events and their mutual relations, rather than by an external clock.

We introduce a binary **precedence relation** $\prec$ on $\mathcal{E}$ to capture causal or relational ordering. If E_a influences E_b (directly or indirectly), we write $E_a \prec E_b$. This relation induces a partial order on events (analogous to causal order in relativity), but without a fixed metric time interval. Consistent with relativity and relational time philosophy, the temporal ordering of two events can differ when viewed from different frames (each event's perspective). To maintain consistency, we impose that if $E_a \prec E_b$ in one frame, then in any frame of an event that is related to both E_a and E_b, the ordering is the same (this prevents contradictions). However, if two events are unrelated (no influence either way), their temporal order can be undefined or frame-dependent, reflecting the relational (not absolute) nature of time.

We formalize the **coordinate assignment** via the notion of **temporal distance** between events. Suppose E_i and E_j are connected by some interaction or information flow. We assign a *pairwise temporal interval* $\Delta t(E_i, E_j)$ which event E_i perceives between itself and E_j. This interval is defined via a function of their **relational intensity** (defined in section A2). Denote $I(E_i, E_j)$ as the symmetric intensity measure between E_i and E_j. We require that Δt is a monotonically decreasing function of intensity – strongly related events are "close in time" from each other's perspective, whereas weakly related events are far apart. For example, we might define:

$$\Delta t(E_i, E_j) := f\big(I(E_i, E_j)\big),$$

where f is a positive, decreasing function (e.g. $f(x) = \frac{1}{x}$ for $x > 0$, or $f(x) = -\ln x$ if $0 < x \leq 1$). Thus if two events have maximal possible intensity $I = 1$, then $f(1)$ is minimal (which we can normalize to 0, meaning they coincide or happen quasi-simultaneously in each other's frame), whereas if intensity is very small, $f(x)$ is large (large time separation perceived). In the simplest case, one can take $\Delta t_{ij} = 1/I_{ij}$ for illustration – high intensity I_{ij} yields small Δt_{ij}.

Each event E_i's timeline can then be constructed by summing such intervals along a chain of related events. If E_i and E_j are not directly related, we find a sequence $E_i = E_{k_0} \prec E_{k_1} \prec \cdots \prec E_{k_n} = E_j$ that links them (a path through the relational network). Then the **time coordinate of E_j in E_i's frame** can be defined as the accumulated interval along the *shortest* such path:

$$\tau_i(E_j) := \min_{E_{k_0} = E_i \prec \cdots \prec E_{k_n} = E_j} \sum_{m=0}^{n-1} \Delta t(E_{k_m}, E_{k_{m+1}}).$$

This equation formalizes the idea that the time difference between E_i and E_j, from E_i's perspective, is determined by the *most direct or strongest chain of influences* connecting them. In graph-theoretic terms, treating each event as a node and $f(I)$ as a "distance" on the edge between events, $\tau_i(E_j)$ is the graph distance from i to j (Graph Neural

Networks as Neural Diffusion PDEs). Because f is decreasing in intensity, the shortest distance path corresponds to the *maximally intensive* connection between the events. This ensures that an event's frame assigns earlier times to events strongly linked to it and later times to weakly linked or indirect events.

Self-reference and consistency: By construction, $\tau_i(E_i) = 0$ for any event E_i. Also, note that in general $\tau_i(E_j) \neq -\tau_j(E_i)$; each event has its own time scale. However, for consistency, if the relationship is mutual and strong, we expect approximate antisymmetry. We can impose that if $I(E_i, E_j)$ is high (close to 1), then $\tau_i(E_j) \approx -\tau_j(E_i)$ up to some small discrepancy. In a perfectly symmetric scenario, one could set $\tau_i(E_j) = -\tau_j(E_i)$, but asymmetry can arise if one event has a different "internal clock rate" than another (due to differing connections). Our framework allows such flexibility: time is **local and emergent**, and only in special cases (e.g. symmetrical interaction) does it resemble a shared coordinate.

This relational time bears resemblance to **Mach's principle** and Barbour's idea of time as derived from change. Here, all time values are defined **internally by the network of events**. There is no universal $t = 0$ birth of the universe or an external timeline – each event is the zero of its own clock. Time between events is measured by how events influence each other, reflecting an extreme relationalism: "time arises relationally, from correlations between systems (events and internal clocks)" (Measurement events relative to temporal quantum reference frames – Quantum). In the quantum reference frame literature, a similar notion exists where an "ideal clock" might measure time via correlations (Measurement events relative to temporal quantum reference frames – Quantum). Here each event essentially acts like a clock reference for others.

To illustrate simply: suppose events A, B, C occur such that A directly causes B, and A also (more weakly) influences C (perhaps through an indirect chain). Let $I(A, B)$ be high (close to 1), and $I(A, C)$ be moderate. Then in A's frame, $\tau_A(B) = f(I(A, B))$ will be small (say

$\tau_A(B) = 1$ arbitrary unit) while $\tau_A(C) = f(I(A,C))$ is larger (say 5 units). Thus A would say "B happened shortly after me, whereas C happened later." However, if B and C are unrelated to each other, B's frame might not assign any meaningful time to C except via $B \to A \to C$ path. So $\tau_B(C) = \tau_B(A) + \tau_A(C)$. If B was directly caused by A, $\tau_B(A)$ might be negative (meaning A is in B's past). So one could find $\tau_B(C) = \tau_B(A) + \tau_A(C) = -\tau_A(B) + \tau_A(C)$. Using our numbers: if $\tau_A(B) = 1$ and $\tau_A(C) = 5$, and $\tau_B(A) = -1$ (assuming symmetry for strong link), then $\tau_B(C) = 4$ units. Thus from B's perspective, C happened 4 units after B. In this way, all event frames are related by transformations that depend on the network of intensities.

Mathematically, one may view $\{\tau_i\}$ as a set of coordinate charts on the space of events. Transformations between these charts are given by the above path summation rule. The structure is analogous to an **affine connection** on a manifold of events, where $f(I)$ plays the role of a metric distance between events. Because each triangle of events may not perfectly satisfy symmetry (i.e. in general $\tau_i(E_j) + \tau_j(E_k) \neq \tau_i(E_k)$ due to possible information loss or asymmetry in indirect influence), this space can have non-Euclidean properties. It could be non-metric in the strict sense, but for strongly connected subsets it approximates a metric space.

An important special case is a **causal loop**. If an event influences itself via a chain (forming a closed loop $E_i \prec \cdots \prec E_i$), consistency demands that the total summed interval $\sum \Delta t$ around the loop is zero (a requirement akin to **self-consistency in closed timelike curves**). If not, we would have a paradoxical assignment (E_i's own time would have to be both 0 and nonzero). Our framework enforces that any *inconsistent loop must carry effectively zero intensity*, ensuring it does not contribute (see Section E4 on paradoxes). This is in line with the principle that "inconsistent causal loops have zero probability of occurring" ([PDF] Retrocausality and quantum mechanics - Griffith Research Online) – here any truly paradoxical loop would be assigned vanishing intensity, preventing it from influencing the coordinate assignments.

In summary, **self-referential coordinates** emerge from the network of events by assigning each event as an origin and defining event-to-event times via a function of their relational intensities. This yields a relational time scheme that generalizes the idea of proper time along worldlines in relativity (with worldlines replaced by chains of related events). It also resonates with the Page–Wootters mechanism of time from entanglement, in which "time is only meaningful relative to a physical clock (another system)" (Measurement events relative to temporal quantum reference frames – Quantum) – here each event can serve as that clock for others. We now turn to defining the relational intensity that underpins these temporal intervals.

A2. Measuring Relational Intensity Between Events

At the heart of the temporal navigation system is the concept of **temporal luminosity** – a scalar measure of how strongly events relate to or influence one another. We formalize this with a function $I : \mathcal{E} \times \mathcal{E} \to \mathbb{R}_{\geq 0}$, where $I(E_i, E_j)$ quantifies the **relational intensity** between events E_i and E_j. By design, we take I to be symmetric ($I(E_i, E_j) = I(E_j, E_i)$) and dimensionless (an abstract "strength" or correlation measure).

Interpretation: If we imagine each event as a "light source" in time, then $I(E_i, E_j)$ represents the brightness observed due to E_j when measured from E_i (or vice versa). A high intensity means the events are closely linked (causally or correlationally), analogous to being near each other in a temporal sense, whereas a low intensity means a tenuous or distant relationship. Importantly, this intensity is **direction-independent** – it does not encode *which* event influenced the other, only the magnitude of connection, much like a photometer that measures light intensity but not the direction of the source.

We can think of many possible definitions for I depending on context: if events carry some state information or occur within a physical system, I could be a normalized correlation between the states before and after the events, or a measure of information flow or mutual information. In a

network representation, one might assign weight I_{ij} to the edge connecting event i and j representing the strength of interaction. For theoretical generality, we assume $0 \leq I(E_i, E_j) \leq 1$ with 1 being the strongest possible link and 0 meaning no direct relation. Near-zero intensities are allowed but, as emphasized throughout, are not simply discarded; they may cumulatively have effects (Section B1).

Normalization: Each event can be assigned a base "luminosity" $L(E_i)$ defined as the sum of intensities connecting it to all other events:

$$L(E_i) := \sum_{E_j \in \mathcal{E},\, j \neq i} I(E_i, E_j).$$

If the event set is infinite, this could be an integral over a density of events. We will often refer to $L(E_i)$ as the **temporal luminosity** of event E_i. It's a scalar analogous to a star's luminosity in space – here it tells how strongly E_i is connected overall to the web of events. Navigation using luminosity will be discussed shortly; intuitively, $L(E_i)$ indicates how "centrally" or "actively" positioned E_i is in the event network (a very central event is brightly connected in all directions, whereas an isolated event has low luminosity).

Relational intensity as a metric weight: In Section A1 we already used I implicitly to define time differences. One convenient formalism is to treat $w_{ij} := f(I_{ij})$ as a **distance weight** on a graph of events, where f is a decreasing function (like $1/I$). Then strong intensity means small distance, implementing the idea that intensity correlates inversely with perceived time separation. This is consistent with regarding I as analogous to an **affinity** or inverse metric. Indeed, if we had an embedding of events into an abstract metric space, I_{ij} could be defined as $I_{ij} = g(d_{ij})$ for some decreasing function g of the metric distance d_{ij} between events i and j. In a simple case, $g(d) = 1/(1+d)$ or e^{-d} might convert a distance to an intensity in [0,1]. But since we start from relational data rather than an assumed space, we take I as fundamental.

Properties of I: It can be useful to require certain properties for I:

- *Reflexivity:* We could define $I(E_i, E_i)$, the self-intensity of an event, as a baseline (perhaps the maximum 1 or some normalization). For simplicity, set $I(E_i, E_i) = 1$ for all i (an event fully correlates with itself). This is consistent with each event being at zero time separation from itself. It also allows treating $L(E_i)$ as including a self-term if needed.

- *Boundedness:* $0 \leq I_{ij} \leq 1$. If events i and j have no influence or correlation at all, $I_{ij} = 0$. If they are essentially the same event or perfectly correlated, $I_{ij} = 1$. Intermediate values indicate partial correlation or influence. In physical terms, one might calculate I_{ij} from normalized covariance of some quantities at those events, or from a propagator amplitude magnitude between spacetime events, etc.

- *Symmetry and Undirected nature:* As stated, $I_{ij} = I_{ji}$, meaning the intensity is a property of the pair, not an ordered pair. This does **not** contradict causal direction; it simply means the *strength* of interaction is mutual. (For example, in Newtonian gravity, the force magnitude between two masses is symmetric, even though one might consider one mass influencing the other; similarly, if event A emits a signal detected by B, then B's measurement of intensity relates to A's emission power – one can equally say A had an effect on B with a certain strength. The model abstracts away direction into other constructs like $\prec$, while I captures just how strongly tied they are.)

- *Transitivity in relation:* There is generally no requirement that I_{ij} satisfies triangle inequality or any transitive property; it's not a distance itself. However, if events share strong connections to a common event, one can expect they are not entirely unrelated: e.g. if $I(A, C)$ and $I(B, C)$ are large, likely A and B have at least an indirect relationship through C. This will reflect in the time coordinates rather than a direct $I(A, B)$ necessarily. We do not enforce $I(A, B)$ to be high just because both are high with C, but through the distance formalism, A and B will have a short path $A \to C \to B$ making them effectively closer in time even if

$I(A, B)$ is low. In other words, **intensities define a weighted graph** and one can derive an effective distance or connectivity between any two nodes from those weights.

The concept of intensity can be related to known physical quantities. For instance, in quantum physics, one might equate $I(E_i, E_j)$ to the magnitude of a Feynman propagator or transition amplitude between those events (suitably normalized). In information theory, $I(E_i, E_j)$ could be the mutual information or transfer entropy between events considered as random variables. In a social or computer network analogy, I_{ij} is the weight of connection between two nodes/events. These interpretations lend themselves to different choices of how to compute I in practice, but for our formalism we treat it abstractly.

The **measurement of intensity** would be done by an observer or agent moving through events. Since it is direction-independent, an observer at event E_i can measure the total luminosity $L(E_i)$ around them but not the direction to each event. We might say the observer has a device that reads a scalar "glow" resulting from all events. More formally, if the observer is currently at some event (or in an interpolation between events), they can sample the intensity field. In continuous terms, one could imagine an **intensity field** $I(x)$ on a continuous event-space manifold, analogous to a scalar potential or brightness field. For each event location x, $I(x)$ might represent the density of nearby events or total interaction strength. In discrete terms, one just has the list of $I(x, E_j)$ to every other event E_j. Navigation then requires moving in the direction that increases intensity towards a desired target.

Example: If events are points in spacetime, we might define $I(E_i, E_j) = e^{-\Delta s_{ij}^2/\ell^2}$ where Δs_{ij} is the spacetime interval and ℓ a characteristic length. This I is high for events close in spacetime and decays for distant ones. Then $L(E_i) = \sum_j e^{-\Delta s_{ij}^2/\ell^2}$ would measure how densely packed events are around E_i. In a highly clustered region of events (e.g. many events in a short time window – an "intense period"), L would be large. This is analogous to a bright flash (many events at once) versus darkness (few events). Alternatively, if events are

defined by some system's state changes, $I(E_i, E_j)$ could be large if the state after E_i is very similar to the state after E_j (indicating they occurred near the same time or in synchrony), etc.

The relational intensity directly influences **navigation**, as described next.

A3. Probability Distribution Functions for Simultaneous Path Existence (Quantum Superposition)

In classical navigation through time (or any space), one assumes a single definite trajectory. Our system, however, embraces the **quantum principle of superposition**: until an observation is made or a criterion of selection is applied, the system can be viewed as exploring many possible event-paths simultaneously. We formalize this by defining a **probability distribution over event paths**. An **event-path** (or simply **path**) is a sequence of events $\gamma : E_{i_0} \prec E_{i_1} \prec \cdots \prec E_{i_n}$ following the relational order. For example, a path could be $A \to B \to D \to \cdots \to Z$ indicating a possible history or route through events A then B then D etc., ending at Z. We consider all possible paths that satisfy the partial order constraints (no path can violate the causal/relational ordering).

Let Ω denote the set of all allowed paths in the event network (this could be an enormous or even uncountable set if events are continuous). We define a probability measure $P : \Omega \to [0, 1]$ with $\sum_{\gamma \in \Omega} P(\gamma) = 1$ (or an integral if Ω is uncountable). This distribution encapsulates the idea that multiple potential trajectories exist concurrently. Only upon "measurement" or selection (for instance, an agent choosing a route, or a specific outcome being realized) does one particular path become the realized history.

To construct $P(\gamma)$, we draw inspiration from **Feynman's path integral** and related formulations. In quantum mechanics, each path is assigned an amplitude and the total amplitude is a sum/integral over all paths (Exploring Feynman Path Integrals: A Deeper Dive Into Quantum Mysteries | by Freedom Preetham | Quantum Mysteries | Medium)

(Exploring Feynman Path Integrals: A Deeper Dive Into Quantum Mysteries | by Freedom Preetham | Quantum Mysteries | Medium). In our model, since we ultimately want a probability distribution, we can assign *weights* to paths analogously and interpret them probabilistically (this is more akin to the **many-worlds interpretation** where each path can be thought of as a branch with some weight, rather than interfering amplitudes that might cancel out). Here is one general formulation:

- **Path weight:** Define a weight $W(\gamma)$ for each path $\gamma = (E_{i_0} \to E_{i_1} \to \cdots \to E_{i_n})$. We take this weight as the product of segment weights, where each segment weight is related to the intensity between consecutive events. For example, for γ as above, let

$$W(\gamma) := \prod_{k=0}^{n-1} w\big(E_{i_k}, E_{i_{k+1}}\big),$$

  where $w(E_i, E_j)$ is some function of $I(E_i, E_j)$. A simple choice is $w(E_i, E_j) = I(E_i, E_j)$ itself (or perhaps a normalized or exponentiated version if needed). This means a path's weight is higher if it goes through strongly connected events at each step. If any link has very low intensity, the whole path's weight is suppressed.

- **Probability of a path:** We then define

$$P(\gamma) = \frac{W(\gamma)}{\sum_{\gamma' \in \Omega} W(\gamma')},$$

  i.e. the probability is proportional to the weight. This ensures normalization. In practice, summing (or integrating) over all γ can be extremely complex, but conceptually this is our PDF over paths.

This distribution favors paths that traverse **high-intensity connections** consistently. It encodes the idea of **"principle of least action" in a stochastic sense**: just as in classical mechanics the actual path is the one that extremizes (usually minimizes) the action, here the *most likely* path is the one that maximizes the product of intensities (or

equivalently, minimizes the sum of $f(I)$ distances introduced earlier). In fact, if we take logarithms, $\ln W(\gamma) = \sum \ln w(E_{i_k}, E_{i_{k+1}})$, so the most probable path roughly minimizes an effective "action" $-\sum \ln w$. If $w = I$ and $I = e^{-(\text{action})}$ for each step, this becomes directly analogous to the path integral where stationary action gives dominant contribution (Exploring Feynman Path Integrals: A Deeper Dive Into Quantum Mysteries | by Freedom Preetham | Quantum Mysteries | Medium) (Exploring Feynman Path Integrals: A Deeper Dive Into Quantum Mysteries | by Freedom Preetham | Quantum Mysteries | Medium). Even without that explicit analogy, our scheme implies that *high-intensity pathways dominate* the probability measure – they are exponentially (or multiplicatively) favored.

To incorporate **quantum superposition**, one can also assign complex **amplitudes** to each path (including phase factors), and then probabilities would arise from the squared magnitude of a sum of amplitudes. However, doing so requires careful consideration of interference. In this model, we focus on the **probabilistic emergence of a trajectory upon observation**, meaning we are interested in the final probability weights of each possible trajectory when one is about to be realized. One can imagine that underlying this, a more primitive description involves a wavefunction Ψ that is a superposition of all paths. For example, define a state $|\Psi_{\text{paths}}\rangle = \sum_\gamma \sqrt{W(\gamma)}\, e^{i\phi(\gamma)} |\gamma\rangle$, where $\phi(\gamma)$ is some phase (which could be an integral of phases along the path, like $\sum S_{step}/\hbar$). Prior to observation, the system is in this superposed state. Once an observation is made (e.g. the traveler "chooses" or an outcome happens), it "collapses" to one particular $|\gamma\rangle$ with probability $|\sqrt{W(\gamma)}|^2$ which is exactly $W(\gamma)$ normalized, i.e. $P(\gamma)$ as above. This aligns conceptually with the many-worlds perspective: each path γ corresponds to a branch (or world) (Everett's many worlds probability : r/cosmology - Reddit), and the "thickness" (weight) of that branch is $W(\gamma)$ such that the chance of finding oneself in that branch is proportional to that weight.

Simultaneous existence: At any given not-yet-observed stage, multiple paths can co-exist. For example, consider a simple bifurcating scenario:

starting event E_0 can lead to either E_1 or E_2. Say $I(E_0, E_1) = 0.9$, $I(E_0, E_2) = 0.5$. Then there are two possible one-step paths: $\gamma_1 : E_0 \to E_1$ with weight 0.9, and $\gamma_2 : E_0 \to E_2$ with weight 0.5. These might represent two different outcomes or directions. Both paths exist in superposition initially. If an observer from E_0 hasn't yet "decided" or seen which event occurs, we treat it as both occur with respective probabilities after normalization: $P(\gamma_1) = 0.9/(0.9 + 0.5) \approx 0.643$, $P(\gamma_2) \approx 0.357$. Upon observation, one path will be realized with those odds. Before observation, however, the system can manifest effects of both (e.g. interference if phases are considered).

For **longer paths**, the distribution can get very intricate. But one particularly interesting situation is when there are *infinitely many possible paths*. In such cases, it is possible that the **support** of the probability distribution (the set of paths with non-negligible probability) has a complex structure. In fact, as we discuss in the next subsection, under certain recursive branching processes the set of high-probability paths can form a Cantor-like fractal set. The distribution $P(\gamma)$ might be very uneven, with a few trajectories carrying most weight but countless others carrying extremely tiny probabilities that nevertheless sum up to some residual. This echoes the concept in many-worlds that there are vastly many branches, but not all are equally likely – some have extremely low "measure" yet still exist.

We emphasize that in contrast to classical thinking where one would drop paths with negligible weight as effectively impossible, here we **preserve even near-zero probability paths** in the overall ensemble (Section B1 will argue why this is conceptually important). The distribution $P(\gamma)$ thus spans a huge range of probabilities, potentially covering events that standard physics might consider practically impossible.

Mathematically, one could encounter a **singular continuous distribution** for path probabilities in the continuum limit. An analogy is the *Cantor distribution* (the probability distribution whose cumulative function is the Cantor staircase). The Cantor distribution arises from a

process of repeatedly splitting probabilities – it puts weight on a Cantor set of paths (for example, paths encoded in ternary that avoid the digit 1) and has a fractal support of Hausdorff dimension $\log 2/\log 3 \approx 0.6309$ (Cantor set - Wikipedia). In our context, something similar could happen if at each branching we preferentially weight certain branches but not give zero to others, iteratively.

A4. Modeling High-Probability Paths via Cantor Set-Inspired Structures

(GraphicMaths - Cantor set) *Figure A4: Iterative construction of a Cantor-like fractal structure for event-paths. This schematic shows six stages (1–6) of a branching process. At each stage, paths split and certain segments are relatively suppressed (analogous to removing the "middle third" in the classic Cantor set construction). The darkest bars indicate the most likely path segments at each iteration. As the process continues ad infinitum, the set of paths with non-negligible probability forms a self-similar Cantor-type set (though in our model even the "removed" parts retain near-zero but non-zero probability). The fractal dimension (a measure of the distribution's support size) is less than 1, indicating that the high-probability outcomes lie on a thin subset of all mathematically possible paths.*

The distribution of trajectory probabilities described above can acquire a **fractal structure** under iterative branching. We take inspiration from the classic Cantor set – a fractal created by repeatedly removing the middle third of a line segment, leaving behind a dust of points that is uncountable yet has zero total length (GraphicMaths - Cantor set). In probability terms, if one starts with a uniform distribution on [0,1] and at each iteration redistributes probability onto the two side segments (and zero on the removed middle), in the limit one obtains the Cantor distribution: the probability mass lives entirely on the Cantor set (a set of measure zero in the interval).

In our context, **high-probability paths** play the role of the Cantor set, and low-probability paths correspond to the removed middle segments (though we will not remove them completely, only make them extremely

small). Consider a simple branching model: at each step, an event leads to two possible next events (binary branching). Assign probability weights p and $1-p$ to the two branches (with $p > 1-p$ for asymmetry). If we iterate this branching N times, there are 2^N possible paths, each of which is a sequence of left/right branch choices. The probability of a given path (assuming independence at each branch) will be $p^k(1-p)^{N-k}$ if it took the high-probability branch k times and the low one $N-k$ times. In the limit $N \to \infty$ (infinitely long paths), the set of paths that continue to consistently take the higher-weight branch most of the time will have an outsized share of probability. In the extreme case $p = 2/3$ and $1-p = 1/3$ (mimicking the Cantor removal proportions), the *single path* that always takes the p-branch at every iteration has probability $(2/3)^N$ at step N, which tends to 0 as $N \to \infty$ (so no single infinite path has finite probability), yet the *set* of paths that never take an $(1-p)$ branch at some stage corresponds to the Cantor set structure and collectively carries probability 1 in the limit of strictly zero weight on the discarded parts (GraphicMaths - Cantor set). If instead we allow $(1-p)$ branches to have tiny but nonzero weight even as $N \to \infty$, then the support of the probability measure is not strictly Cantor (since nothing is truly removed), but for practical purposes the overwhelming majority of measure lies on a Cantor-like thin set.

We can formalize a **fractal path model**: suppose at each branching, out of b branches, a certain subset (say m branches) receive significantly larger weights than the others. If we were to set the smaller weights to zero in the limit, we would be left with m^N surviving paths after N iterations (out of b^N total), similar to how at each step the Cantor set keeps 2 segments out of 3. The surviving set often has fractal dimension $D = \frac{\ln m}{\ln b}$. For example, a ternary branching where 2 branches survive yields $D = \ln 2 / \ln 3$ as in the standard Cantor set (Cantor set - Wikipedia). In a more general sense, the **effective fractal dimension** of the high-probability path set is

$$D_{\text{path}} = \frac{\ln(\text{number of significantly weighted branches per split})}{\ln(\text{total number of branches per split})}.$$

If some branches have intermediate weight, one could get a multiscale Cantor-like structure or even a multifractal distribution of path probabilities. Mandelbrot's work on fractal and multifractal distributions is relevant here – he observed that processes with heavy-tailed distributions and long memory can produce wild fluctuations and clustering (Can a Longer Memory Improve Predictions? | Chicago Booth Review) (Can a Longer Memory Improve Predictions? | Chicago Booth Review). In our model, if near-zero probability events are retained, the resulting distribution of realized trajectories over many trials might show *large deviations* where extremely unlikely sequences eventually occur (given many opportunities), analogous to heavy tails in time series (the "Noah effect" of sudden large changes (Can a Longer Memory Improve Predictions? | Chicago Booth Review)). This means while some paths are fractally common, others, though fractally rare, might eventually happen given infinite time.

To make this concrete, consider the probability measure $P(\gamma)$ over infinite paths in a binary tree with weights as above. If one computes the cumulative distribution function, it will have a **Devil's staircase** form – constant on "gaps" (which correspond to those paths that include a low-probability branch at some early stage and thus collectively have little additional weight beyond a certain point) and rising steeply on the fractal set of best paths. This is qualitatively similar to the Cantor function (which is constant on the removed middle-thirds and jumps on the Cantor set). The difference in our scenario is that the "removed" pieces aren't truly removed, just exponentially suppressed. So the final CDF would not be completely flat on those segments but would have tiny slopes – thus likely a *continuous but not absolutely continuous* distribution, i.e. one with a singular component concentrated on a fractal set (like the Cantor distribution which is continuous everywhere but has derivative zero almost everywhere except on a fractal set of points).

In summary, by iteratively favoring certain branches of event evolution, the **support of the dominant probability mass** forms a Cantor-set-like structure in the space of all possible paths. This captures the idea that

although many paths exist in superposition, relatively few are overwhelmingly likely – yet *no cutoff is imposed*; even paths with near-zero probability exist in the wavefunction. The high-likelihood paths are interwoven in a self-similar way across scales of branching. The mathematics of fractals provides tools (Hausdorff dimension, self-similarity ratios, etc.) to quantify this structure. For instance, one could derive that the set of paths above some probability threshold ϵ shrinks in size geometrically as ϵ increases, following a power-law relationship – a hallmark of fractal distributions.

This Cantor-inspired model will be useful when we consider long-term evolution and the presence of "edge case" phenomena (Section B): it suggests that even extremely rare sequences (analogous to points in the removed middle of Cantor set) might have some relevance, and that the structure of reality's timeline might be much thinner than the set of all mathematical possibilities (thus avoiding the "measure problem" of everything happening at once, by concentrating measure on a fractal subset).

Having established the core mathematical constructs – event-centric time coordinates, an intensity-based measure of temporal relationship, a quantum-like distribution over paths, and fractal structure in likely trajectories – we move next to integrating broader theoretical perspectives and ensuring no important subtlety (like near-zero probabilities or alternative interpretations) is overlooked.

B. Integration of Historically Marginalized Frameworks

B1. Models Preserving Near-Zero Probabilities (Nothing is Discarded)

Traditional physics and probability theory often simplify models by truncating events with "negligible" probabilities – effectively treating probability ~10^{-50} as zero for practical purposes. In our framework, we

explicitly **preserve near-zero probabilities** at the fundamental level. The reasoning is twofold: mathematically, even events of extremely small probability can accumulate significance over many trials or in chaotic systems; philosophically, excluding them a priori can bias the model against rare but potentially crucial phenomena (the proverbial “black swan” events).

Edward Lorenz’s discovery of deterministic chaos provides a powerful lesson here. In his weather model, rounding off tiny differences led to entirely different outcomes – the famous *butterfly effect*Traditional physics and probability theory often simplify models by truncating events with “negligible” probabilities – effectively treating probability ~ 10^{-50} as zero for practical purposes. In our framework, we explicitly **preserve near-zero probabilities** at the fundamental level. The reasoning is twofold: mathematically, even events of extremely small probability can accumulate significance over many trials or in chaotic systems; philosophically, excluding them a priori can bias the model against rare but potentially crucial phenomena (the proverbial “black swan” events).

Edward Lorenz’s discovery of deterministic chaos provides a powerful lesson here. In his weather model, rounding off tiny differences led to entirely different outcomes – the famous *butterfly effect*. Lorenz showed that **“tiny changes in initial conditions evolve to completely different trajectories”**, meaning a minuscule cause can drastically alter the course of events (Lorenz system - Wikipedia). For example, a minute air disturbance (likened to a butterfly’s flap) could be the difference between no storm and a hurricane weeks later (Lorenz system - Wikipedia). This underscores that events with extremely low initial probability or influence should not be dismissed; over time or in complex dynamics they might amplify. Our model thus keeps $P(\gamma) > 0$ for *every* path $\gamma \in \Omega$, no matter how small $P(\gamma)$ might be. There are no a priori “impossible” paths (unless truly $I = 0$ for some required link making it disconnected). In technical terms, the support of the probability measure spans all of Ω, even if much of that support carries vanishingly small density.

By preserving near-zero probabilities, the framework can naturally incorporate **quantum tunneling-like or rare fluctuations** in a classical-looking setting. In standard physics, if an outcome has probability 10^{-100}, one might say "it will never happen." Here, we keep it as an available branch – if we were to run the scenario 10^{100} times, we expect it to happen about once. This is analogous to quantum theory, where no matter how forbidden a path is classically, as long as it's not strictly disallowed, it contributes some amplitude. Over many trials those rare paths will occasionally occur. In our temporal navigation context, this means an agent might, with astronomically small probability, find a bizarre sequence of events unfolding – but it's not forbidden.

Another example: consider **quantum decay or barrier penetration**. A nucleus might have a 10^{-50} chance per second to decay. Standard approach: effectively zero chance in any given second, but given enough time (or enough atoms) decays will occur. In our model of event trajectories, a decay event would be one branch with very low weight, but it remains part of the distribution. If one were to remove it (set that branch's probability to zero) for convenience, one would entirely miss the possibility of decay. This seems obvious, but subtly, many models *do* drop low probabilities by imposing cutoffs or by assuming certain symmetries that eliminate "unlikely" processes. We avoid doing so.

Preserving tiny probabilities also connects to **multi-modal or heavy-tailed distributions**. Mandelbrot identified the importance of heavy tails through what he called the "Noah effect" – extremely large jumps occur more often than Gaussian models would predict (Can a Longer Memory Improve Predictions? | Chicago Booth Review). In a heavy-tailed scenario, probabilities far in the tail are small but not as astronomically small as in a light-tailed (e.g. normal) distribution. Our framework's retention of all probabilities naturally accommodates heavy tails: one can have a power-law distribution of path probabilities. For instance, one could find that $P(\gamma)$ for the γth most likely path decays like $\gamma^{-\alpha}$ for some exponent α, rather than exponentially. This is a type of distribution that assigns non-negligible measure to a vast number of rare paths, and it cannot even be normalized if one truncates artificially.

By keeping all paths, we ensure the correct heavy-tail behavior emerges if the intensities dictate it.

In practical terms, including near-zero probabilities means that our **simulations or calculations must track a vast range of magnitudes**. This is computationally challenging (much like simulating many-worlds or a path integral with fine detail), but conceptually it is crucial for accuracy. We envision using techniques from multifractal analysis and high-precision arithmetic to handle this if one were simulating the model.

One might worry that having so many tiny probabilities could cause predictions to lose focus – if "everything is possible," how do we get meaningful results? The resolution lies in the **fractal structuring** discussed in A4: the overwhelmingly likely set of outcomes has a specific structure (e.g. fractal dimension $D < 1$), and those dominate expectation values. Near-zero probability events exist but contribute negligibly to most aggregate quantities (they are like measure-zero events in an integral – they don't affect the integral's value but conceptually they're present). They become important only when specifically looking for them or for questions of possibility rather than expectation. Our validation protocols (Section E) will ensure that typical predictions (like expectation of some observable) are governed by the high-probability paths and match known physics, while the low-probability ones provide a reservoir of alternative possibilities that could in principle be realized in huge sample sizes or one-in-a-trillion scenarios.

In summary, **no probability branch is completely pruned** in the theoretical construction. This policy draws from chaotic dynamics (sensitivity to tiny changes) (Lorenz system - Wikipedia), fractal time series (heavy tails and long memory) (Can a Longer Memory Improve Predictions? | Chicago Booth Review) (Can a Longer Memory Improve Predictions? | Chicago Booth Review), and the spirit of quantum mechanics (anything not forbidden is mandatory). It positions the framework to account for **edge phenomena** that standard models

might ignore – for example, a sequence of "freak" coincidences, while fantastically unlikely, is not treated as impossible here and thus could be analyzed for its potential effects. This will be important when we consider phenomena like apparent temporal anomalies or miracles; they would correspond to extraordinarily low P paths, which however exist in the distribution and could conceivably occur without violating the model.

B2. Alternative Quantum Superposition Interpretations

Quantum theory's foundations remain unsettled, with several interpretations offering different ontologies. Our temporal navigation system, by incorporating superposed paths and observation-induced selection, can be connected to various interpretations, and we aim to **embrace their insights rather than pick one**. In particular, we draw on ideas from:

- **Many-Worlds Interpretation (Everett):** In many-worlds, all possible outcomes of quantum events *actually occur* in branching universes, and interference between branches accounts for quantum phenomena. Our model's superposition of event-paths (Section A3) is essentially a many-worlds picture of time: each path is like a "world." When an event with multiple outcomes occurs (like our branching in probability), we don't eliminate any outcome; each persists as a branch (with appropriate weight) (The Single Consciousness Many Worlds Interpretation (SCMWI)). The "observer" in our model would experience one branch, but the other branch still exists in the total state. We even use language like branch weights $W(\gamma)$ analogous to the thickness of worlds. Thus, our framework is naturally Everettian. The advantage here is that we don't have to add a mysterious collapse postulate – the emergence of a single trajectory is just the observer finding themselves in one branch. Meanwhile, the overall structure retains all branches. Many-worlds also suggests that probability is subjective (each branch has some measure, but all occur); in our model we treat $P(\gamma)$ as the **measure** (weight) of path γ in the

multiverse of possible histories. When needed, we recover the appearance of random collapse by acknowledging we, as observers, see only one branch with probabilities given by those weights (Everett's many worlds probability : r/cosmology - Reddit).

- **de Broglie–Bohm Pilot Wave (Bohmian mechanics):** Pilot-wave theory proposes that particles have definite trajectories guided by a pilot wave (the wavefunction) which propagates according to Schrödinger's equation. There is no splitting of worlds; instead, an unseen wave guides the particle, and probabilities arise from ignorance of initial conditions. One can draw an analogy in our model: the **relational intensity field** I can play the role of a pilot wave guiding the "particle" (the navigating agent or the realized trajectory). All possible paths exist in the wave (like all momentum components in a wavefunction), but only one path is taken by the particle. However, the particle is guided by the intensities which evolve sort of like a wave. To make this concrete, consider a simple case of two possible next events: pilot-wave style, the intensities $I(E, \mathrm{next}_1)$ and $I(E, \mathrm{next}_2)$ guide the choice – one can imagine a hidden variable that picks one outcome but influenced by the ratio of these intensities (similar to how in pilot wave the particle goes one way or another depending on initial position in a wave interference pattern). We can incorporate Bohmian insight by considering if there's a way to treat the path selection as deterministic given some hidden parameter, while still reproducing the probability $P(\gamma)$ distribution on average. This could be an interesting variant of our model: perhaps each event carries a hidden "phase" and when intensities oscillate, certain paths get constructive interference boosting I (like pilot wave nodes). While our primary formulation is more aligned with many-worlds (where all paths exist then one is realized), a pilot-wave variant would treat the realized path as the only reality but guided by a holistic intensity field computed as if all paths existed. The mathematics can accommodate this by reinterpreting $P(\gamma)$ not as fundamental probability but as an emergent frequency if one considers an ensemble of systems with slightly different hidden

states.

- **Transactional Interpretation (Cramer):** This lesser-known interpretation involves waves traveling forward and backward in time that form a “handshake” (transaction) to cause an event. In a sense, it’s time-symmetric and allows retrocausality in the process of establishing quantum events. Our framework could integrate such ideas in the relational structure: since each event establishes a frame, one could allow influences that appear to go backward in another frame (retrocausal links). For instance, an event E_j might increase the intensity $I(E_i, E_j)$ even if E_i is in the “future” of E_j in some ordering, effectively meaning E_j sends an influence back to E_i. While we did not explicitly include retrocausal links in Section A2 (we assumed a partial order to avoid contradictions), one could relax strict acyclicity if handled carefully with consistency conditions. The Transactional Interpretation’s central point is that emitter and absorber interact through advanced and retarded waves to finalize an event. If translated to our terms, one could say: an event-path γ that ends in a particular outcome might be reinforced by an influence from that outcome event traveling back along γ to ensure its own realization (a self-consistency condition). This is speculative, but the mathematics of having $I(E_i, E_j)$ depend on the entire set of paths (or future events) is conceivable (it would make the model highly non-linear and time-symmetric). We mention this to show that our framework is flexible enough to entertain such unconventional ideas of time loops and retro-influences as part of the relational dynamic, as long as they do not introduce paradox (which we address in E4).

- **Objective Collapse Models:** These are interpretations where wavefunction collapse is a real, dynamical process (e.g. GRW theory) that happens with tiny probability per particle, so large systems collapse quickly. In our model, one could mimic objective collapse by introducing a slight mechanism that randomly prunes the path superposition when certain conditions are met (like when an event becomes “macroscopic”). For example, one could impose

that beyond a certain threshold of path divergence or number of particles involved, the intensities for "coherent" superposition drop off, effectively collapsing the distribution to one branch (or a narrower set of branches). This would be an additional rule layered on our basic framework (which by itself is more Everettian). While we do not explicitly need an objective collapse (since many-worlds view suffices), if future evidence suggested collapse is real, our intensity-based approach could accommodate it by saying: intensities themselves undergo a stochastic dynamics that tends to concentrate $P(\gamma)$ onto one trajectory in certain regimes. This would preserve our core structure but add a non-linear, non-unitary element (perhaps something like an intensity diffusion with absorption). Such an extension would align with proposals that perhaps gravity or other effects cause quantum collapse. Our model could incorporate a *collapse of the intensity field* as a physical process – conceptually a bit like a sudden focusing of the "temporal luminosity" onto one direction when an event becomes too "bright" (complex).

In integrating these interpretations, we are guided by the principle that none of them is definitively proven or disproven; each might be capturing part of the truth. Our model's **mixture-of-experts architecture** (see section C2) is in fact well-suited to blend these perspectives: one expert module could use a many-worlds-like superposition calculation, another could simulate a pilot-wave guidance, and another could impose occasional collapses. The gating function (which picks which expert dominates in which regime) might learn that, say, for microscopic events, use the many-worlds superposition model; for mesoscopic maybe pilot-wave for efficiency; for macroscopic maybe introduce collapse dynamics. In effect, the **model does not commit to a single quantum interpretation** but can reduce to each in appropriate limits.

Historically, many of these interpretations were marginalized or at least outside the mainstream Copenhagen view. By allowing their mathematical ideas into our framework, we ensure it is robust and

general. For example, *Bohmian trajectories* can be seen if we condition on a particular hidden variable – then the agent will have a definite path (no branching) but intensities still guide it, reproducing quantum statistics (The Single Consciousness Many Worlds Interpretation (SCMWI)). *Everettian branching* is essentially our default view, ensuring linear superposition and interference are accounted for. *Retrocausal transactions* could be included to explore time-loop consistency (connected to section B4 and E4 on paradoxes). And *collapse models* can be emulated or even explicitly incorporated if needed to account for apparent wavefunction reduction at scale.

In practice, when making predictions, all these interpretations should give the same observable results (that's the point – they are interpretations of the same math). Our model's predictions will not depend on which view we take; they emerge from the underlying math of intensities and path weights. However, by drawing on these various frameworks, we gain insight into different limiting cases and computational approaches. For instance, a Bohmian algorithm for finding an optimal path might inspire our optimal navigation computation (Section D2) by treating it like finding a streamlining in a flow (where the flow is the gradient of intensity). An Everettian view might encourage us to calculate interference effects between nearly identical paths (which could slightly modulate probabilities). A transactional view might help in setting up boundary conditions for solving the intensity propagation (Section C1) in a time-symmetric way. Thus, integrating these perspectives is not just philosophical but also *pragmatic for model building*.

B3. Non-linear Temporal Models Across Traditions

The concept of time has varied widely across scientific and cultural traditions. Our framework benefits from incorporating **non-linear and non-standard temporal models** that have been sidelined by mainstream physics but could enrich our understanding:

- **Cyclic and Fractal Time (Ancient and Cultural Models):** Many

ancient cultures conceived of time as cyclical rather than linear. For example, Hindu cosmology talks of repeating Yugas (epochs) in a grand cycle; the Mayan calendar has cycles; some indigenous cultures view time as a repeating spiral of seasons or events. While at first this seems contrary to the increasing entropy arrow, our model can accommodate cyclicity by allowing event networks that are recurrent. If events influence others in a closed chain (with large periods), one could get *periodic structures* in the event graph. A simple way to model a cycle is to have a set of events $\{E_1, \ldots, E_n\}$ such that $E_n \prec E_1$ (closing the loop) and similar intensities $I(E_i, E_{i+1})$. This essentially forms a **closed timelike curve** in the relational graph. Normally, physics forbids such loops (for causality reasons), but solutions like Gödel's universe in general relativity show they are at least mathematically possible. We allow them with the proviso of consistency (no paradoxical change of something that prevents its own cause). A cyclical time model can emerge in our framework if the intensity field has a repeating pattern, effectively creating **temporal attractors** that an event can revisit. Indeed, if our diffusion equation for luminosity (Section C1) admits periodic solutions, then intensity could oscillate and events could repeat conditions. Mandelbrot's idea of fractal time series also ties in – a process might exhibit self-similar patterns across time scales, meaning certain configurations of events repeat in scaled forms (this is akin to cycles but not strictly periodic, more like *scaling symmetry in time*). Our Cantor-set high-probability paths hint at a self-similar pattern: zooming into a portion of the path-space might reveal a smaller copy of the overall structure. If nature's event progression is self-similar, then time might not be uniform: clusters of events and quiescent gaps could appear at all scales (which resonates with 1/f noise, long memory processes (Can a Longer Memory Improve Predictions? | Chicago Booth Review)). We integrate this by allowing non-linear time mappings – e.g. time coordinate τ in one frame might relate to another frame's τ' in a fractal way (perhaps $\tau' = \tau^{\alpha}$ piecewise). This can explain, for instance, why certain

historical patterns seem to recur (if the network of events has a scaled repetition).

- **Time as an Emergent Property (Modern Physics):** In some approaches to quantum gravity (like Carlo Rovelli's relational quantum mechanics or Julian Barbour's timeless configurations), time is not fundamental. Instead, it emerges from the relations between events or changes of state. Our model is fundamentally relational, aligning with this view. We explicitly do not assume a background time; hence if we were to quantize gravity or cosmology within our framework, time would come out as an *emergent coordinate* defined by events (consistent with the Wheeler-DeWitt equation idea where time is a parameter derived from correlations). The Page–Wootters mechanism mentioned earlier (Section A1) is one such approach: it treats the universe state as static, and what we call "time" is just correlation between subsystems (Measurement events relative to temporal quantum reference frames – Quantum). We have effectively implemented a similar idea by saying each event can act as a clock via correlation intensities. Thus, our model provides a concrete realization of emergent time – something that many quantum gravity models strive for conceptually.

- **Discrete Time and Non-linear Time Arrows:** Some traditions or theories suggest time might be discrete (made of chronons or quantum of time) or that the arrow of time (direction of increasing entropy) might not be universal. Our model can incorporate **discrete time steps** easily if we restrict events to a lattice or assume a minimum interval (e.g., if intensities $I(E_i, E_j)$ are nonzero only if $j = i + 1$ in some discrete index sense). In fact, the work by Castro et al. (referenced in the quantum frames discussion) indicated that assuming time is discrete can restore consistency in a quantum temporal reference frame model (Measurement events relative to temporal quantum reference frames – Quantum). If needed, we can impose a smallest scale such that events are like ticks of a fundamental clock, but because

each event has its own clock, this could vary – akin to a local discrete time unit that might differ across the network. More generally, **non-linear time** means the relationship between what a local clock measures and what some global parameter would be is non-linear. For example, some processes exhibit time dilation or acceleration depending on conditions (think of how time in general relativity runs slower in strong gravity – that's non-linear scaling of proper time relative to coordinate time). In our framework, intense event interactions could "slow down" local relational time (lots of events densely packed might make each event's frame count many happenings while another sparse region counts few). This can be worked into the intensity → time mapping function $f(I)$. For instance, if f is non-linear, high intensity might not just linearly reduce time intervals but could, say, saturate: perhaps beyond a certain intensity threshold, adding more intensity doesn't shorten the perceived interval much more. That would be a non-linear time response.

- **Psychological and Biological Time:** It's worth noting that even within human experience, time perception is non-linear – time can seem to speed up or slow down depending on context (boredom vs. fear situations). While our model is physical, since it concerns events and observers, one could map an observer's cognitive event processing to this framework. An intense burst of stimuli (events) might subjectively make time feel longer or shorter. By modeling the observer as another event-processing system, our relational time might provide a mathematical description of such subjective time (though that's beyond our current scope, it shows the versatility).

By integrating these diverse temporal concepts, we ensure the model is not locked into the simple linear, universal time of Newton or even the strictly forward, continuum time of standard quantum theory. Instead, time can be **multi-faceted**: partly cyclic, partly evolving, with local variations. Polymath contributions are apt here: for example, *Benoit Mandelbrot* introduced fractal time series in finance (the Joseph and

Noah effects) to explain long-range dependence and heavy tails (Can a Longer Memory Improve Predictions? | Chicago Booth Review) – we incorporate that in allowing fractal event distributions. *Edward Lorenz*, aside from chaos, effectively showed a kind of strange attractor in time (the Lorenz attractor doesn't repeat exactly, but it's bounded and has fractal structure, implying a kind of complex cyclical behavior in time). We include Lorenz's insight that long-term prediction is limited (so we might emphasize probabilistic description over deterministic long-term paths, which we do). Also, thinkers like **Ilya Prigogine** spoke of the arrow of time and how irreversibility and chaos interplay – we allow multiple time arrows (local arrows associated with each event sequence).

From a **technical standpoint**, incorporating these ideas means our equations might be non-linear. For instance, the **diffusion equation for luminosity** might need non-linear terms to allow pattern formation (like how the **Perona-Malik equation** in image processing is non-linear to preserve edges (Graph Neural Networks as Neural Diffusion PDEs)). Non-linear time models suggest that the propagation of intensities or the update of event coordinates can be non-linear functions of the intensity itself. This could lead to **soliton-like solutions** (localized repeating event patterns) or **bifurcations** (sudden changes in the pattern of events as some parameter varies) – phenomena studied in non-linear dynamics which we can borrow.

We also take heed of **traditions outside of Western science**: for example, some philosophies consider that past, present, future are all ever-present (block universe), whereas others think only the present is real (presentism). Our model interestingly can accommodate a block-universe view (the entire network of events and all paths exist as a structure) but also a presentist view for an observer (at any given point, they are at some event and see one history). The relational update rule (how intensities and frames update as events occur) could be formulated either in a block sense (solving the whole network) or an iterative sense (growing the network event by event). Both are valid perspectives, akin to block vs dynamic time.

In conclusion, by **embracing non-linearity and diversity in temporal models**, we enrich our framework's descriptive power. This ensures that if nature's true temporal behavior deviates from the simple monotonic flow (which quantum gravity and cosmology hint it might), our system can adapt to describe it. It also widens the scope of validation: we can compare predictions not just to laboratory clocks but to large-scale cosmic cycles or to statistical patterns in complex systems to see if our temporal model holds universally.

B4. Edge Phenomena and "Paranormal" Temporal Effects

Standard physics often dismisses reports of phenomena like premonition, retrocausation, time loops, or "glitches in the matrix" as mere psychological curiosities or errors. While extraordinary claims require extraordinary evidence, our framework is open to examining **edge phenomena** rigorously. By edge phenomena, we refer to any effects at the boundaries of established physics: for example, closed timelike curves (CTCs), superluminal correlations, extreme coincidences, or temporal anomalies (say, effects that seem to precede their causes). Rather than assume these cannot happen, we ask: *if* they were to occur, can our model accommodate them without contradiction?

Some edge phenomena to consider and how our model addresses them:

- **Self-Consistent Time Loops:** A time loop means an event somehow is among its own causes (e.g., event E causes F and F causes E in a loop). In relativity, solutions like Gödel's universe allow it, and the **Novikov self-consistency principle** asserts that any such loop must be self-consistent (you can't change the past in a way that prevents the loop from happening). In our model, a loop would manifest as a cycle in the event graph. We have already allowed for cycles in B3. The key is consistency: if $E \prec F$ and $F \prec E$ forming a 2-cycle, then it must be that $I(E,F)$ and $I(F,E)$ are both high, and $\tau_E(F) + \tau_F(E) = 0$ if we enforce

symmetry (meaning each sees the other as happening at the same offset, effectively simultaneous in some sense). Any **paradoxical loop** (where E influences F in a way that would negate E's occurrence) would show up as contradictory requirements on intensities or frames. Our model can flag such situations: they would likely result in an inconsistent set of equations for τ or negative probabilities. We impose that such loops get probability zero ([PDF] Retrocausality and quantum mechanics - Griffith Research Online). Practically, that means the gating network in mixture-of-experts would assign zero weight to a solution that contains a logical paradox. However, a **non-paradoxical loop** (like the classic bootstrap paradox of an object or information with no origin but consistently cycling) could, in principle, exist in our model. It would appear as a strongly connected subgraph of events all mutually consistent. The model would then give it some probability weight $P(\gamma)$ – probably extremely small if it requires many coincidences to maintain itself, but not zero. Testing for this in experiment would be challenging, but if nature allowed mild forms of this (like certain particles influencing themselves through wormholes), our formalism would not break; it would just require including those interactions in I.

- **Retrocausality (effects preceding causes):** Retrocausality in physics is usually frowned upon due to causality violation concerns. But some interpretations (like the transactional one) or proposals in quantum foundations (like certain entangled photon experiments) entertain it. In our model, retrocausality could be modeled by allowing the *relational order* $\prec$ to not strictly align with an external time. For instance, one event E_j could have influenced E_i even if j comes "later" in some embedding. If one were to draw events on a timeline, a retrocausal influence would be a link going leftward. In our formalism, since each event has its own frame, one event can see another as in its future while the latter sees the former in its past – essentially $\tau_i(E_j) > 0$ and $\tau_j(E_i) > 0$ potentially, which is only possible if i and j are not comparable in the partial order (they influence each other indirectly perhaps).

True retrocausality would be a loop as above (because if A influences B which is earlier by external time, B must influence A or something that leads to A to close consistency). So retrocausality is a subset of loops. We can thus say: allowed if self-consistent, with effectively zero probability if it would create an inconsistency. This is neatly in line with a result from a retrocausal interpretation analysis: *"inconsistent causal loops have zero probability of occurring, only self-consistent ones occur"* ([PDF] Retrocausality and quantum mechanics - Griffith Research Online). We build that into our framework as a rule.

- **Superluminal or Instantaneous Connections:** Quantum entanglement presents correlations that appear instantaneous (spacelike separated). Standard quantum theory insists no information is transmitted (so it's not causation). In our relational model, if two events are entangled, we could reflect that by a high intensity $I(E_i, E_j)$ even if they are not causally linked in the usual sense. This I would encode correlation but not a direct cause-effect. Navigation-wise, it means an agent at E_i might see $L(E_i)$ high due to E_j even though E_j is spacelike – a bit like a nonlocal glow. However, since one cannot use this to send controllable messages (in quantum theory), we ensure that while I can be high, it doesn't create a usable loop or contradiction. Essentially, we can include entanglement as edges in the event graph that are "non-causal" (no $\prec$ relation, but I nonzero). This extends our graph to a more general network, possibly a directed acyclic graph plus additional undirected edges for entanglement. The math still applies; those edges just won't be used in computing time order (since no order), but they contribute to luminosity and maybe influence path probabilities in a subtle way (they might correlate which branch is taken at separated events, reproducing entanglement correlations).

- **Psychological or Parapsychological claims (premonition, déjà vu):** While scientifically controversial, suppose some people claim to "sense" an event before it happens. If one were to model this,

one could postulate that the brain events corresponding to the premonition are entangled or intensity-linked with the future event. That would be a retrocausal connection at the level of neural events. Our framework could handle it by an appropriate $I(\text{premonition event}, \text{future event})$ being nonzero. This would effectively correlate the two in the ensemble of paths (the paths where the premonition and event align would have higher weight). Without endorsing such phenomena, we note that our model could at least represent them if they were real, unlike classical models which cannot even describe such a correlation except as a fluke. In our model it would not violate causality as long as no information paradox arises (the person can't consistently prevent the event they foresee – that would be inconsistent loop if they tried).

- **Extremely unlikely coincidences:** Sometimes sequences of events occur that are so unlikely that people consider supernatural explanations. In our model, they are simply paths with low $P(\gamma)$ that happened. Because we keep near-zero probabilities, we can quantify just how unlikely and track them. Interestingly, the fractal structure means that for any extremely unlikely path, there are many similarly unlikely ones (the Cantor set has infinitely many points with arbitrarily small measure). So while each is rare, the chance of *some* rare thing happening eventually might be higher than naive expectation (as heavy-tailed distributions often show "something crazy" is not as unlikely as "that particular crazy thing"). Our model could be used to evaluate probabilities of classes of rare events.

In integrating edge phenomena, we lean on **logical consistency** as the guiding criterion. This is where rigorous mathematics meets speculation: we allow the equations to tell us if a scenario is possible or not. For example, if we introduce a hypothetical intensity link that creates a time loop, we solve the equations for τ and P. If a solution exists that doesn't give negative probabilities or contradictory time ordering, then the model allows that loop (with some probability). If not,

the model inherently gives it probability zero or no solution. This way, we treat exotic scenarios with the same rigor as normal ones, just perhaps with extreme parameters.

Crucially, **our framework does not presume such phenomena are occurring**; it simply has the capacity to describe and quantify them. This is an improvement in scope. Should future experiments or observations reveal evidence of any tiny deviation from normal causality (for instance, certain results in delayed-choice quantum experiments or brain neuroscience experiments hinting at retro-influence), our model could be refined to include them. Meanwhile, if they don't exist, our model likely just assigns them practically zero probability and reduces to standard physics in all tested regimes.

In conclusion of Section B, by integrating marginalized frameworks and edge concepts, we've made our system **robust, flexible, and revolutionary**. It stands on the shoulders of polymaths and mavericks – from Mandelbrot and Lorenz to Everett and Bohm – uniting their insights into a single mathematical edifice. The reward is a temporal navigation theory that is not only consistent with known physics but expansive enough to explore the unknown without immediate contradiction. This integrated perspective sets the stage for constructing concrete models and simulations, which we turn to next.

C. Testable Mathematical Models

C1. Diffusion-Like Equations for Temporal Luminosity Propagation

A central dynamical aspect of our system is how **temporal luminosity (intensity)** propagates or evolves through the event network. We propose a **diffusion-like equation** to describe the evolution of the intensity field I (or the derived luminosity L) over the relational structure. The analogy is with heat or particle diffusion: just as heat flows from hot to cold regions over time, we can imagine "temporal

brightness" flowing from regions of high event connectivity to low connectivity. This yields a partial differential equation (PDE) on the graph of events or its continuum approximation.

On a Graph: Consider the events as nodes in a graph G with edge weights $I_{ij} = I(E_i, E_j)$. We can define an intensity field on nodes, e.g. $u_i = L(E_i)$ the total luminosity at node i. A discrete diffusion equation on this graph can be written as:

$$\frac{du_i}{d\tau} = D \sum_j A_{ij} \, (u_j - u_i),$$

where A_{ij} is the adjacency (or weight) matrix of the graph (perhaps normalized), and D is a diffusion coefficient. This is analogous to the standard heat equation $u_t = D\Delta u$ but on a network (Graph Neural Networks as Neural Diffusion PDEs). The term $\sum_j A_{ij}(u_j - u_i)$ is the graph Laplacian acting on u (Graph Neural Networks as Neural Diffusion PDEs - Blog - X), which spreads the values by averaging neighbors. In words, the rate of change of luminosity at event i is proportional to the difference between its neighbors' luminosity and its own. Steady state ($du_i/d\tau = 0$) would mean u_i is an average of neighbors or the field is constant on each connected component.

However, our scenario has some twists: the "space" we diffuse in is not physical space but event-space, which grows as new events occur. Also, time in this diffusion equation is not the physical time but some parametric time (perhaps corresponding to progression along the network or some continuous interpolation of event ordering). We might consider τ in the above equation as an index along a foliation of events (like a global time if it exists or just a simulation time parameter).

Continuous Approximation: If events were embedded in a continuous space (e.g. an event continuum with coordinates x), we could define a luminosity density $L(x,t)$ and write:

$$\frac{\partial L(x,t)}{\partial t} = D \nabla^2 L(x,t) + (\text{sources/sinks}).$$

This would mean intensity diffuses spatially. But without a given embedding, we stick to the graph form. Yet, if the event distribution is dense, one can imagine an underlying manifold where a graph Laplacian approaches a continuous Laplace-Beltrami operator.

Why diffusion? Diffusion is a natural model for many processes of equilibration. In our context, if one part of the event network is highly luminous (lots of interconnections) and another part is dark, diffusion would spread the luminosity out – meaning events in the dark region will gradually connect more with the bright region (since high intensity connections will tend to form or strengthen over "time"). One could interpret this as the system trying to reach a more homogeneous state where intensities are balanced, analogous to thermal equilibrium. In physical terms, if a big event happens (flash of many events) it could send out "ripples" of influence raising intensity of distant events gradually.

We also consider **directed propagation** akin to wave or diffusion with drift. For instance, perhaps intensity tends to flow forward in the event causal order (like radiation going outward in time). We might add a drift term $v \cdot \nabla L$ representing a bias in a certain direction (maybe outward along the arrows of time). However, since time itself is the parameter, it gets tricky. Instead, one can incorporate drift by using a non-symmetric adjacency if needed or by explicitly constructing an evolving sequence of graphs.

A simpler interpretation: The diffusion equation can be used as a **smoother or interpolator** for the intensity field on the existing graph. Since events happen discretely, one might run a diffusion on the graph between events to redistribute intensity weights smoothly, then at the next event update, etc. This could be part of a simulation algorithm: after each new event, let intensities adjust via a diffusion step to simulate the relational adjustments.

The diffusion equation yields testable consequences. For example, it implies a kind of **random walk** for how intensity signals move. If we

somehow perturb the network at one event (like inject a burst of luminosity at one node), the equation predicts a Gaussian spreading of that luminosity over the network with time. One could test this in a simulation by creating a cluster of events and seeing if the intensity measured at some "distance" in the graph follows the heat kernel's predictions (Graph Neural Networks as Neural Diffusion PDEs). The **graph heat kernel** $H_{ij}(t) = \exp(-t\mathcal{L})_{ij}$ (where $\mathcal{L}$ is the Laplacian) gives the transition probability of a diffusion from node i to node j in time t (Graph Neural Networks as Neural Diffusion PDEs - Blog - X). If our model is right, the way indirect relationships build up should match something like $H_{ij}(t)$. For example, if we define $I^{(t)}(E_i, E_j)$ as intensity between events as a function of the parametric time t, starting with initial intensities, then $I^{(t)}(i, j)$ might evolve according to $dI_{ij}/dt = D\sum_k (I_{ik}I_{kj} - I_{ij})$ or some such non-linear variant (since new links might form via intermediary k).

Non-linear extension: Simple diffusion is linear and tends to homogenize everything. But our system might exhibit structure formation – akin to **pattern formation** in reaction-diffusion systems. We might need a non-linear term or source term to create stable intensity gradients (for navigation cues). For instance, if luminosity tends to concentrate around events (like each event "emits" luminosity), we could add a source term $+S_i(t)$ in the equation to represent event i's intrinsic contribution (maybe proportional to number of connections it makes, etc.). Non-linear diffusion like **Perona-Malik** (used in image processing to preserve edges) uses a diffusivity that decreases with large gradients (Graph Neural Networks as Neural Diffusion PDEs) – analogously, if there's a big intensity difference (like a sudden spike), we might slow diffusion to preserve that contrast (maybe representing a boundary in event-space like between different eras or domains). This is speculative, but potentially we will consider a **saturating diffusion**:

$$\frac{du_i}{d\tau} = D\sum_j A_{ij}\, g(u_j - u_i),$$

where $g(\Delta)$ is a non-linear function such that for large $|\Delta|$, $g(\Delta)$ is

smaller (less diffusion when differences are large, thus edges remain). This could maintain distinct clusters of high luminosity separated by low luminosity gaps (like segments in Cantor set remain).

Validation of diffusion: One way to validate this part is to simulate a growing network of events with some rules and measure how $L(E)$ changes. If it roughly satisfies a diffusion equation, then our assumption was right. Alternatively, in a controlled experiment in a model system (maybe a social network analog or a computer simulation of branching timelines), we can attempt to detect a diffusion pattern in the connectivity measure.

In summary, a diffusion-like equation provides a **continuous, testable dynamic law** for intensity. It is testable in simulations (and potentially indirectly in experiments by checking if influence spreads with characteristics of a diffusion process, e.g. a bell curve distribution over "distance" after some time). This also connects to known mathematics of heat kernels on graphs, allowing us to borrow results: for example, the heat kernel trace relates to graph spectra, which could tie into the fractal dimensions if the network is fractal (heat kernel asymptotics can reveal fractal dimension). So if our event network has fractal features, the diffusion equation's solutions will reflect that (like anomalous diffusion, where mean squared "distance" grows as t^{α} with $\alpha \neq 1$ due to fractality).

The diffusion equation will also be crucial in our **navigation algorithm**: an agent might solve a backward diffusion (or Poisson equation) to find a gradient (like solving $\nabla^2 L = 0$ gives harmonic field which can guide along level sets). That overlaps with Section D1 where converting intensity to direction likely involves spatial gradients, which the diffusion equation helps define on the graph.

C2. Mixture-of-Experts Architecture for Integrating Models

To handle the complexity of our system, we employ a **Mixture-of-**

Experts (MoE) approach, a machine learning architecture where multiple specialized models ("experts") are trained to handle different aspects of a problem, and a gating network learns to weight or select among them (Mixture of experts - Wikipedia) (Mixture of experts - Wikipedia). In our context, the "problem space" spans various regimes of temporal behavior: quantum microdynamics, chaotic mesodynamics, possibly classical macro-dynamics, and exotic edge cases. No single model may efficiently cover all regimes, so we divide and conquer.

Experts to include:

1. **Quantum Path Integral Expert:** A model specialized in superposition and interference. It would use methods of summing over histories (like a Monte Carlo over paths) to compute probability amplitudes. This expert is most relevant at small scales or when intensities along multiple paths are comparable (requiring interference calculation). It effectively encodes the many-worlds or Feynman path approach.
2. **Classical Shortest-Path (Geodesic) Expert:** This model assumes one trajectory dominates (like classical physics) and attempts to find that optimal path by solving variational equations or using Dijkstra-like algorithms if discrete. It's efficient when one route has overwhelmingly higher probability than others (like when intensities strongly pick a single chain – high signal-to-noise scenario). It draws from principles of least action or geodesics in a metric (we will formulate the actual equations in Section D2).
3. **Chaotic/Fractal Dynamics Expert:** This would handle regimes where a multitude of paths contribute in a structured way (e.g. fractal distribution). It might use renormalization or fractal geometry methods to approximate the distribution of outcomes. For instance, it could calculate the Hausdorff dimension of the support or use self-similar recursion to estimate probabilities. This expert becomes important in a regime of marginal chaos – when near-zero probabilities still matter collectively. Techniques here could include multi-scale analysis or solving for fixed-point of a

branching process (like finding the invariant measure of a Cantor-set process).

4. **Relativity/Consistency Expert:** This expert checks for and enforces global consistency constraints (no paradoxes, etc.). It could use logical or algebraic methods (e.g. constraint solving) to adjust intensities or throw out certain solutions. It ensures causality (in the sense of our partial order) is respected and might solve for self-consistent loops. In practice, this might involve solving equations like $\tau_i(E_j) + \tau_j(E_k) + \ldots + \tau_m(E_i) = 0$ for loops, or ensuring P sums to 1, etc. It acts across the whole system as needed, perhaps with negligible effect in normal conditions but critical in edge cases.

5. **Learning/Adaptive Expert:** Possibly an expert that doesn't have a physical interpretation but is a generic machine learning model (like a neural network) that can detect patterns in event data and predict new structure. This can be useful if the system is placed in an environment with unknown rules – the learning expert will adapt. For instance, if the actual underlying physics had some hidden variable or a complex distribution, the learning expert might pick up on it from data (like events log) and correct the outputs. This is analogous to an AI model calibrating the theoretical model to real observations.

The **gating network** in MoE looks at the state (or inputs, like current intensities, event configurations, scale of the problem) and outputs weights $w_1, \ldots, w_n$ for each expert (Mixture of experts - Wikipedia) (Mixture of experts - Wikipedia). These weights determine how to combine the experts' outputs. For example, if in a given scenario quantum interference is critical, the gate will assign a high weight to the Quantum expert and lower to Classical, etc. The gating itself can be a neural network that has been trained on the domain or even a rule-based system initially. Over time, it can be refined (possibly via self-supervised learning as per C3).

Mathematically, if each expert k provides a output function $f_k(x)$ for

some query x (say x encapsulates the current state or a task like "predict next event distribution"), then the MoE output is:

$$f_{\mathrm{MoE}}(x) = \sum_k w_k(x)\, f_k(x),$$

with $w_k(x) \geq 0$ and $\sum_k w_k(x) = 1$ (Mixture of experts - Wikipedia). In some cases, a hard selection is used (winner-takes-all: pick the expert with highest weight). But a soft combination is smoother and can capture mixed regimes.

For instance, consider predicting the probability distribution of the next event given the past events. The Quantum expert might produce a distribution considering interference (complex amplitudes), the Classical expert might pick the single most likely event and give a delta distribution, and the Fractal expert might output a broad heavy-tailed distribution from its scaling laws. The gating network sees the context: maybe the intensities are very high and distinct, indicating a classical regime, so it gives nearly 1 weight to Classical, thus f_{MoE} ~ classical result. In another context, two paths have similar weight and some oscillatory pattern indicates interference, so the gate weights Quantum more, resulting in a combined distribution that shows an interference pattern (e.g. peaks and troughs that neither classical nor fractal alone would produce). If the situation is complex with many moderate paths, gate might mix fractal and quantum experts, etc.

Training the MoE: In a theoretical physics setting, "training" is analogous to calibrating the model to known results. We have a lot of limiting cases we know:

- In simple two-slit experiments, we expect interference fringes (quantum expert should dominate).
- In large-scale systems, we expect a single outcome consistent with classical (classical expert dominates).
- In iterative chaotic mappings, we expect a multifractal distribution (fractal expert dominates).
- In any scenario, consistency must hold (consistency expert always

active as needed).

We could set up a cost function that penalizes deviation from known outcomes in those regimes and adjust the gating network parameters accordingly (like backpropagation in ML). Because we are designing it, we might initialize gating with human insight: e.g. if number of events < threshold and intensities differences are small, use quantum; if one intensity >> others, use classical; if structure repeats over scales, use fractal, etc. The gating network can then refine these by learning from simulation data of the full model or even from actual experiments if available.

Advantages of MoE: It allows **modularity** – each expert can be developed and validated independently on the regime it's best at, using appropriate mathematics and approximations. We've effectively been doing this by separately formulating diffusion (macro behavior), path integrals (quantum micro), Cantor sets (fractals). Instead of forcing one unified approach, MoE lets them coexist. It also provides **explainability**: each expert's contribution can be understood in terms of known theories, as opposed to one giant black-box model. The gating outputs tell us which theory is in play in a given situation, which is scientifically insightful.

From a **testing perspective**, MoE means we can test each expert against scenarios:

- Expert 1 (Quantum) can be tested on small-scale experiments (does it reproduce double-slit interference? entanglement correlations? etc.).
- Expert 2 (Classical) tested on big events (does it give correct classical limit – e.g., does it recover Newtonian or relativistic behavior when quantum phases cancel?).
- Expert 3 (Fractal) tested on statistical distributions (does it match simulations of chaotic systems or known fractal dimensions?).
- The gating logic can be tested by creating hybrid scenarios and seeing if the MoE output matches full brute-force calculations. For example, we could simulate a medium-scale system exactly (with

all paths) and see if MoE chooses a similar combination.

Computational blueprint: In implementation, at each step or for each query:

1. Compute relevant features of current state (like intensity variance, number of significant paths, etc).
2. Feed to gating network to get weights.
3. Each expert computes its prediction or output.
4. Combine outputs by weights.
5. Possibly update internal state (some experts might carry state, e.g. fractal expert might update an estimated measure).

We can also run experts in parallel, which is natural for concurrent processing (each expert tries its specialized solver). This parallels how in complex engineering, different solvers are tried (like finite element vs finite difference vs analytic approximations) and one picks the best.

In our context, this ensures **consistency with known physics in every regime**: the MoE reduces to quantum mechanics when appropriate, to classical mechanics when appropriate, to chaos theory when appropriate, etc., blending them smoothly when regimes overlap (like mesoscopic scales or transition zones).

Finally, MoE architecture itself is something we can justify by the complexity of the domain. Polymath contributions in science often involve combining disciplines – MoE is the algorithmic embodiment of that: e.g., *Lorenz* gave us chaos equations (our fractal/chaos expert uses that), *Mandelbrot* gave fractals (again fractal expert), *Feynman/Dirac* gave path integrals (quantum expert), *Bellman/Hamilton* gave classical optimization (classical expert). Instead of forcing one into the other, we let them all work together.

C3. Self-Supervised Learning for Structure Detection

Given the intricate and high-dimensional nature of event networks, we deploy **self-supervised learning** techniques to help uncover hidden

structures and patterns without requiring labeled data (Self-supervised learning - Wikipedia). Self-supervised learning (SSL) uses the data itself to generate training signals – for example, by masking parts of data and training the model to predict them, or by finding representations that make different views of the same data similar (Self-supervised learning - Wikipedia). In our case, the "data" can be the growing event network (graph structure, intensity values, event properties). The goal is to learn representations or parameters that reveal the underlying order (such as fractal dimensionality, clustering, symmetries, invariants) that might not be obvious from explicit rules.

Possible SSL tasks in our context:

- **Graph Autoencoding:** We can train an autoencoder neural network to compress and reconstruct the event graph adjacency matrix or intensity matrix (Self-Supervised Learning: Everything You Need to Know (2024)). The encoder would produce a low-dimensional embedding of events (assigning each event to a vector in $\mathbb{R}^d$) and the decoder tries to reconstruct I_{ij} from those vectors (e.g. via a dot product or learned function). If successful, the embedding might capture the "event-space geometry" – perhaps revealing dimensions or clusters. For instance, if the intensities reflect an underlying 2D manifold of events, the autoencoder may map events to 2D coordinates effectively. This could validate if events lie on a simple manifold or if they require fractal coordinates (maybe the autoencoder struggles to compress a fractal adjacency because it has no simple low-dim embedding, hinting that the event connectivity is high-dimensional or fractal).
- **Contrastive Learning for Temporal Ordering:** We could generate two different "views" of the event sequence (like splitting the event set into two correlated parts, or taking the network and removing some nodes to see sub-networks) and train the model to produce similar embeddings for correlated subgraphs and distinct for uncorrelated (Self-supervised learning - Wikipedia). For example, pick an event E and consider the subgraph of events within radius

r of E (neighbors up to r hops). That subgraph is one view. Another view could be the subgraph within radius r of one of E's close neighbors. These two subgraphs overlap heavily around E. The model can be trained so that the representation of subgraph 1 and subgraph 2 are pulled together in feature space (because they share context around E), while representations of two completely different subgraphs (far apart events) are pushed apart. Through such training, the model might learn a representation where proximity in feature space corresponds to relational proximity in the event graph. This essentially could recover something like a metric or set of latent variables describing time relationships.

- **Masking and Predicting Intensities:** We can take the intensity matrix I_{ij} and mask some entries (hide them). The model is tasked to predict the masked intensities from the unmasked ones. This is akin to matrix completion or link prediction in a graph (common tasks in network analysis). Self-supervision comes from the fact we hide different sets each time and train the model to fill them in. If successful, the model (likely a graph neural network or matrix factorization) has captured the rules or patterns of how intensities are structured. For example, it might learn that if $I(A,B)$ and $I(B,C)$ are high, then often $I(A,C)$ is at least moderate (transitivity property), except if perhaps something prevents it (like event types). Essentially it could learn the triangle inequality or clustering implicitly. This relates to knowledge graph completion in AI – here our knowledge graph is events linked by intensities.

- **Anomaly Detection:** The model could be trained in a one-class self-supervised way to model "normal" event relationships, so that if an unusual pattern occurs (e.g. a paradoxical loop forming, or a sudden spike in intensity structure that doesn't match diffusion), it flags it as an anomaly. Techniques involve training an autoencoder to minimize reconstruction error on normal data; anomalies then have high error. This could serve as a check or trigger to indicate when perhaps new physics is at play (if our model can't

reconstruct a certain pattern well, maybe something outside our assumptions is happening – a clue to refine the model or discover new phenomena).

Self-supervision is appropriate because we typically don't have "labels" like "this is the fractal dimension" or "this path is the chosen one" in real data – those are what we want to discover. Instead, we use the internal consistency of data (like parts of the event graph predicting other parts) to train models that then reveal structure. This approach is akin to how scientists might use symmetry or conservation to find hidden constants – here the neural network might find a latent that remains stable, which could correspond to a conserved quantity.

Interpretation of learned structures: If the SSL model finds, say, a 3-dimensional embedding for events with low error, that suggests our event space is effectively 3D – which could link to physical spacetime dimension (maybe the events are actually happening in a 3D spatial world plus one time, which the model uncovered). If it finds two clusters in representation, that might indicate two weakly connected components – perhaps two eras or domains that don't interact much (like before and after some massive event that broke time into two phases). If it finds a circle structure, maybe time is cyclic at a large scale. If it finds no good low-dim embedding but can compress via a fractal code (some autoencoders can explicitly learn fractal codes), that confirms fractality.

One specific metric is the **Hausdorff dimension**: we could attempt to estimate it by seeing how many neighbors each event has at intensity > threshold and how that scales as threshold changes. This is like a clustering dimension. The model could be tasked to predict that scaling or to cluster events by scale, which it could learn in a self-supervised way by multi-scale prediction tasks (predict connections at scale $n+1$ from those at scale n). Success would mean it effectively learned the fractal scaling exponent, which we could read off from its weights or outputs (like if it predicts that doubling scale multiplies connections by 2^D, then D is fractal dimension).

Use in validation and refinement: After training self-supervised, the learned model (like an encoder network mapping events to features) can be examined. We might find that one of the feature dimensions correlates strongly with something like “event age” or “event cluster ID”, giving insight to our theory. Also, SSL can generate synthetic missing links which we can then check with ground truth (if any). In simulation, we can drop some interactions and see if the model predicts them – confirming our understanding.

Furthermore, a self-supervised model can serve as a **surrogate** for expensive calculations. For example, computing exact path integrals for large event networks is intractable, but a neural network trained to predict outcomes might do so fast after training. This surrogate can be used inside the MoE as an expert (the learning expert we mentioned). Essentially, self-supervised learning could produce an expert that encapsulates an aspect of the system’s dynamics that isn’t captured by our explicit equations but is present in data.

Example application: Suppose we generate many random event networks under varying conditions (like different initial intensities or rules) and we apply SSL. The model might cluster these networks into distinct regimes automatically. We might discover an emergent parameter that separates networks into “quantum-like” vs “classical-like” behavior purely from graph features, which could correspond to something like a effective $\hbar$ or decoherence parameter in the data. That would be fascinating as it means the model identified a dimensionless parameter controlling the transition – something we could then analytically investigate.

Relation to physical principles: Self-supervised learning essentially lets the data speak for itself, which resonates with how we expect a *self-referential time system* to potentially reveal hidden symmetries. For instance, maybe there’s an invariance under a certain transformation of intensities that we didn’t impose but is true; the SSL model might pick up on that (because it would make prediction easier). That could correspond to a conservation law. We could then formalize that as a

theorem.

In implementing SSL in our system, we leverage modern graph neural networks (like GCN, GAT, etc.), autoencoders, and contrastive losses. The outcome is a set of **learned parameters or features that enhance our model's predictions and understanding**. This is testable by checking the model's performance on predictive tasks: if it accurately predicts held-out intensities or links, it means it captured the structural rules fairly well (Self-supervised learning - Wikipedia). If not, maybe our model class is too limited and needs extension.

Overall, self-supervised learning acts as an **auxiliary scientist** analyzing the data produced by our theory to suggest improvements and verify internal consistency, much like how one would analyze simulation output looking for patterns. It's a powerful tool in our toolkit to ensure the model can essentially calibrate itself and uncover its own hidden assumptions, aligning with the self-referential theme (the model can, in a sense, observe its own behavior and adjust).

C4. Validation Protocols for Each Element of the Model

To ensure the reliability and scientific merit of our system, we outline rigorous **validation protocols** for each component (A through D) of the model. These protocols involve analytical consistency checks, numerical simulations, and whenever possible, comparison to empirical or established results. The goal is to validate not only each piece in isolation but also the integrated performance of the model.

Validation of Theoretical Formulations (Section A):

- *Self-Referential Coordinates:* We will construct simple toy scenarios (e.g., three events with prescribed relationships) and solve for the $\tau_i(E_j)$ assignments. We check for consistency: do the solutions for event-clock readings avoid contradictions? If we deliberately introduce a paradoxical loop in the input (e.g., require $A \prec B$, $B \prec C$, $C \prec A$), does the system detect inconsistency (no real

solution or a necessity for zero intensities)? This tests the logic of our coordinate definitions. We'll also compare our relational time predictions to an equivalent scenario in special relativity: for events with a given causal structure (like a causal diamond), do our τ differences correlate with invariant intervals if we embed events in Minkowski space? This would show that in classical limit, our time matches proper time or coordinate time up to scaling. Another test: if we slowly turn up all intensities equally (making everything strongly connected), do all $\tau_i(E_j)$ tend to 0 (everyone sees everyone as simultaneous)? That would be a reasonable limit (in a fully synchronized network, time differences vanish).

- *Relational Intensity Measure:* We verify basic properties. Symmetry $I_{ij} = I_{ji}$ is straightforward. More involved is checking transitivity effects: if $I(A,B)$ and $I(B,C)$ are high, is $I(A,C)$ at least moderate as a result of diffusion or other dynamics? We might simulate random graph models and measure triple correlations to see if our model's rules produce a triangle inequality–like condition on $f(I)$. We also ensure normalization: if we have $L(i) = \sum_j I_{ij}$, and we simulate some processes, does $L(i)$ behave as expected (like conservation or monotonic changes)? We can also directly measure the intensity field in a controlled physical system to validate scale: e.g., in a social network analogy, if events are interactions, we could infer I from data (like correlation of activity times) and see if it matches our assumptions (bounded between 0 and 1, etc.). Additionally, testing the photometer analogy: if we combine two "sources" of intensity, does an observer see the sum? (linearity test of intensity superposition).

- *Probability Distribution on Paths:* We will test limiting cases where analytical solutions are known. For example, in a simple two-branch system with fixed probabilities p and $1-p$ each step, we derived that the path distribution tends toward a Cantor-set structure as $N \to \infty$. We can simulate moderate N and verify the distribution of path probabilities matches the analytic binomial distribution (which approaches a smooth distribution for large N

by central limit, except at extremes). For finite steps, $P(\gamma)$ should equal $p^k(1-p)^{N-k}$ (if no intensity variation beyond that). We'll check that our general formula $P(\gamma) = W(\gamma)/\sum W$ reproduces that in special cases. Another validation: ensure that $\sum_\gamma P(\gamma) = 1$ always (unit normalization) in simulations. If numerical precision issues cause deviation, we may need to address those. We also compare the most likely path from our distribution to what a classical method (like Dijkstra on $-\ln I$ weights) would predict – they should coincide or come very close when probabilities are highly skewed (one dominant path). Conversely, we ensure that in a symmetrical scenario (all paths equal intensities), our distribution gives equal probabilities, aligning with symmetry.

- *Cantor-Set High-Probability Paths:* This is harder to validate directly without infinite steps, but we can do statistical tests on the distribution of path probabilities for growing depth. For instance, measure the entropy of the distribution for increasing path length N. If the distribution becomes singular (fractal-like), the entropy growth might slow (for Cantor distribution, the entropy per level tends to $\ln 2 - (\ln 3)$ times something, etc.). We can attempt to numerically estimate the fractal dimension of the support by looking at how many paths carry, say, 90% of the total probability as a function of depth. If that number grows like N^D, we can solve for D. If $D < 1$, we indeed have a fractal-like concentration. We can also visually plot distributions for small N to see if they start looking like the Cantor function's stepwise graph. Additionally, we can intentionally zero out small probabilities below a threshold and see if the support that's left approximates a Cantor set (just to identify structure). The validation here is qualitative but important to see if fractal patterns emerge from weight pruning.

Validation of Integrated Framework (Section B):

- *Near-Zero Probability Preservation:* We test that our algorithms (like simulation of path selection or intensity update) do not inadvertently drop tiny probabilities. This is a numerical concern:

probabilities might underflow in computation if extremely small. We will use high precision arithmetic or rescaling (e.g., work in log domain) to ensure even probabilities like 10^{-100} are tracked. A unit test could involve a scenario where there's a 10^{-8} chance of an alternate path; we run, say, 10^9 Monte Carlo simulations of the process and check that indeed the rare path occurs roughly as expected (maybe not feasible physically, but can simulate to a point or ensure code correctness through seeded randomness). Also, ensure that adding a very low-probability path to the model doesn't change the other probabilities significantly (stability check) – because if our normalization or rounding erroneously ignores it, it might mistakenly renormalize others to still sum to 1, which is fine; but if we physically present an outcome, our model should not be surprised by it (like throw an error for an event it considered impossible).

- *Alternative Interpretation Consistency:* We verify that in the domain where an expert interpretation is supposed to apply, the model reduces to it. For example, in a quantum-like scenario (two-slit interference), do we see interference fringes? We can simulate an event that splits into two paths and then recombines, assigning a phase difference artificially via intensities, and see if the probability of final outcomes shows constructive/destructive patterns consistent with $2\cos\phi$ interference term. If we include a pilot-wave style hidden variable, does it yield the same path distribution as the many-worlds calculation (it should, by construction of Bohm theory)? If not, refine. If we incorporate a collapse mechanism for large systems, we can test a Schrödinger cat type scenario: does our model spontaneously reduce superposition when intensities or number of events crosses a threshold? If we mimic GRW collapse by adding noise to path weights at a certain rate, we can simulate a simple case (like 100 spin-½ particles entangled – does the model collapse them with the right probability distribution? This might be too heavy to simulate exactly, but maybe a smaller analogous test). The aim is not necessarily to prove which interpretation is right, but to

ensure no contradictions: e.g. if we simulate a Bell test, do we violate Bell's inequality only up to Tsirelson bound (quantum limit) and not further (which could happen if we mishandle hidden variables)? The consistency expert in the model should catch that; we validate it explicitly.

- *Non-linear Temporal Model Integration:* If our model says time can be cyclic or discrete, we check special cases. For discrete time, we enforce an interval and see if calculations remain consistent when events only occur at ticks (our continuous formulas should reduce to difference equations consistent with that). For cyclic time, perhaps create a scenario of 3 events in a loop and see that the system can remain stable (that each sees the others at some fixed offset self-consistently). If we allow mild loops, test that we don't get divergences in intensity diffusion (maybe loops could cause resonance in diffusion PDE? We simulate a simple loop in diffusion equation and see that it just equalizes intensities along the loop with no blow-up, which it should if symmetric). We also consider if any global invariants (like total intensity) are preserved or evolving. If time is emergent, maybe there's a gauge freedom (shifting all τ_i by a constant). We attempt transformations like adding a constant to all τ of one event's frame and see no observable predictions change (gauge invariance test).

- *Edge Phenomena Tolerance:* To test paradox handling, set up known thought experiments: e.g., the "grandfather paradox" simulation – event A influences B, B influences C, C influences A such that A would not happen if C influences it. We formalize that with some negative or contradictory intensities and see that our solver for intensities or probabilities yields no solution or a self-consistent one where the paradoxical path has zero weight ([PDF] Retrocausality and quantum mechanics - Griffith Research Online). Another test: entangled choices (like Wheeler's delayed choice): simulate an event that splits, then a later event chooses to recombine or not. Our model should show that if recombined, interference appears, if not, it doesn't, *even if* the decision to

recombine is made after the particle has presumably passed the slits (this is more about simulation correctness, ensuring no hidden ordering assumption invalidates it). If we incorporate some retrocausal intensity, we ensure it doesn't allow signaling. For example, if $I_{future,past}$ is nonzero, check that an agent in the past cannot send a message to itself in the future by toggling something that changes intensities beyond what consistency allows. Likely, consistency means any such attempt is compensated by adjustments that nullify the message (Novikov principle applied). Hard to test physically, but can simulate a message-sending scenario with our rules and see that either it fails or requires a paradox (which then gets zero probability). This is subtle, but an idea: have two branches in future that might send a different signal to past; check that either both lead to consistent same past (so no actual difference to send) or those branches are not simultaneously allowed.

Validation of Computational Models (Section C):

- *Diffusion Equation Tests:* Solve the diffusion equation in simple graphs analytically and compare to numeric integration of our model's intensities. E.g., a line of 5 events with an initial luminosity distribution [1,0,0,0,0]. Analytical discrete diffusion with certain D yields known values at time t (can compute by diagonalizing Laplacian or using known formula for a 1D line). Simulate intensities updating according to our rules and see if match. Also test on a 2D grid of events if possible. Additionally, check conservation: diffusion should conserve total "mass" (here maybe total luminosity $\sum_i u_i$ if no sources). Does our model conserve total intensity if no sources? It should by design; we verify that numerically. If we include sources (like each new event adds luminosity), verify that added amount equals area under diffusion curve increase. One can also test the limiting behavior: as $t \to \infty$, diffusion should equalize u_i. Our intensities should approach some uniform state in a connected component. We could simulate a random graph and indeed see that $L(E)$ values become nearly

equal eventually if no new events come in.

- *Mixture-of-Experts Performance:* We will generate synthetic scenarios where the correct solution is known and see if MoE picks the right combination. For example, scenario X should be handled by expert A dominantly – does gating assign w_A ~ 1? If not, tune the gating network. We can do cross-validation: have a set of scenarios, train gating on some, test on others to ensure it generalizes (like if gating is learning criteria like "if N events > 50, use classical," test on N=60, etc.). Also measure the error of MoE output vs expected output. We can simulate, for instance, a small quantum system exactly (by brute force path enumeration) and compare with MoE's combined output (quantum expert might approximate, but maybe fractal expert not needed in that small case, etc.). If error is high, refine experts or gating. We also test extreme cases: if two experts strongly disagree (one says event likely happens, another says it's impossible), does gating pick one confidently or produce an average? If it averages in a nonsensical way, we might adjust gating to be more discrete in such cases (maybe through a temperature parameter making softmax sharper). Essentially, ensure that MoE output remains physically sensible (e.g., probabilities in [0,1] and sum to 1 – trivial, but when combining distributions from experts we must ensure normalization; our combination formula should be convex combination, which preserves that if each expert output is a proper distribution).

- *Self-Supervised Model Accuracy:* We hold out parts of the data (like mask some intensities or remove some events) and let the SSL-trained model predict them. Compute error metrics (MSE for intensities, accuracy for link existence, etc.) (Self-supervised learning - Wikipedia). A high accuracy indicates the model captured the underlying structure. If we introduced a known pattern (like events lie on a circle, intensities depend on circle distance), see if model's learned embeddings reflect a circle. Could use dimensionality reduction techniques on the learned

representation to see shape. If the SSL model finds something like a linear relation in features that correspond to, say, chronological order of events, then it effectively learned an arrow of time. We can also test anomaly detection: deliberately introduce an inconsistency in a test graph and see if the model's reconstruction error or contrastive score flags it. If so, good – the model can aid in consistency checking beyond deterministic rules, acting as a safety net.

- *Computational Efficiency:* Validate that the algorithms scale reasonably. For diffusion, check how complexity grows with number of events N. Graph diffusion using sparse matrices is $O(N + |E|)$ per step typically. MoE runtime is sum of expert runtimes plus gating overhead; test on increasing scenario complexity and measure CPU time. If too slow, consider simplifying experts or using approximate gating (like regimes with pre-defined thresholds). SSL training time should also be noted – if we can't train a huge graph at once, maybe train on growing subsets incrementally (online training). We ensure memory usage is manageable (for intensities on N events we store an N×N matrix potentially, which is N^2 – we need sparsity or low-rank approximations, so validate that intensities indeed remain sparse or effectively factorable).

Validation of Specific Formulas (Section D to be written):

We haven't composed section D content in analysis, but presumably:

- Converting intensity to navigable vector: Validate by checking if following that gradient indeed leads to higher luminosity – basically a discrete gradient ascent test on a graph. If not using gradient, if using a geodesic compute, test on known potential fields (like if L is like a potential with known optimum, ensure our formula finds it).
- Optimal paths calculation: Compare with brute force for small graphs (like enumerate all paths, pick best vs our algorithm

result).

- Trajectory probability distribution formula: For small sets, compare direct calculation vs formula.
- Mapping coordinates to observables: If we map event times to, say, physical time differences, check dimension analysis and units, etc.

Since section D content is not explicitly written in analysis, I infer these are straightforward formula validations.

Finally, integration test: combine everything in a simulation (like simulate a timeline with events influenced by these rules) and see if known emergent phenomena are reproduced (e.g. do a bunch of quantum interactions and see if macroscopic variables follow classical equations on average – akin to Ehrenfest theorem validation; or simulate two unconnected subsystems and see that intensities between them remain low until a coupling event is introduced – validating locality preservation).

Each validation step provides either confidence in the model or feedback to refine it. This process ensures that **each theoretical component is grounded in either established physics or logical consistency, and that the computational implementation faithfully realizes the theory**.

Having validated components and subsystems, we can proceed with greater confidence to use the model for making predictions and exploring new scenarios.