AI Models Are Not Conscious and Are Massively Inefficient, Causing Complexity and Scalability Bottlenecks to Artificial General Superintelligence Risking Technological Singularity and Loss of Societal Dynamism or Institutional Collapse and Renewal

Trevor Nestor
Information Physics Institute
Louisiana State University
University of California, Berkeley
Redmond, WA
trevornestor@berkeley.edu

June 2025

Abstract

Contemporary AI models — especially large language and "reasoning" models —achieve notable pattern-matching feats and success on a variety of complex tasks, yet in general remain orders of magnitude less efficient than the human brain in complex information processing and energy efficiency, with recent evidence pointing to possible diminishing returns and even limits in scaling models. Our analysis suggests that without adopting new human agent or brain-inspired architectures, investments in scaling AI by brute force will approach diminishing returns in value creation relative to architectures hosting consciousness or that are capable of lateral information sharing between agents and direct investments in infrastructure, risking a critical failure of Artificial General Superintelligences (AGSIs) as an information surveillance and control loop at maintaining institutional stability in late stage societies (a reinterpretation of the concept of a technological singularity). Our proposed framework outlines an interdisciplinary approach unifying sociology, economics, neuroscience, and physics that could lead to more efficient, potentially conscious, and safe intelligent systems or directions for alternative policy decisions to avert institutional collapse.

Contents

1 Introduction

Artificial Intelligence systems based on deep neural networks have achieved impressive performance in language, vision, and even strategic games, yet mounting evidence indicates that these models fundamentally differ from human brains in both capability and efficiency [1] [2] [3]. Notably, today's largest language and reasoning models consume enormous computational resources, yet leading linguistic scholars posit they still differ profoundly from how humans reason and use language [4]. Large language models (LLMs) have been described as "stochastic parrots" that are only proficient at pattern matching of training data without genuine understanding [5]. These critiques underscore a growing realization: despite their fluent outputs and pattern-matching, current AI systems may lack the conscious awareness, flexible reasoning, and energy-efficient information processing generated by the human brain. Growing concerns about the scalability of AI architectures and how they differ from intelligence or consciousness generated or processed in biological tissues is needed both from a geopolitical and policy standpoint, and for societal stability as an information surveillance and control loop to stabilize social and economic institutional equilibria which must scale fast enough to maintain late-stage societies.

Criticism of current AI technology includes that current neural network based approaches differ from the way in which human agents process information (that they do not resolve the hard problem of consciousness or the binding problem, operate on classical models which do not feasibly replicate backpropagation and weight transport in the efficient manner that the brain does, and also cannot as efficiently take advantage of collective human agent intelligence facilitated by inter-brain synchrony [170]) or store memory (the brain distributes, processes, and stores memory nonlocally across the biological tissue - a theoretical enigma conceptualized by the mind-body problem) and that these differences which current AI architectures lack account for these massive inefficiencies and bottlenecks [6] [7] [8] [170] [10]. In our analysis we propose that the hard problem of consciousness is a problem that can be cast as at least NP-Hard in complexity. We explore catastrophe theory models of AI performance collapse, and incorporate insights from neuroscience (e.g. orchestrated objective reduction in microtubules, memory storage and retrieval) and physics to predict unsustainability in scaling and propose why AI models are not conscious and how their massive inefficiencies create practical limits en route to a hypothetical AI agent able to resolve the alignment problem [119] as an information surveillance and control loop for society (an Artificial General Superintelligence or AGSI).

The human brain operates on an estimated 20 watts of power – about the energy consumption of a dim light bulb – to support processing across multiscale assemblies of 80–100 billion neurons - each with their own network of dendritic connections - and each with multifrequency signals below the network layer [208] [209] [13]. In contrast, cutting-edge AI models run across data centers drawing kilowatts to megawatts of electricity that require significant investments in new power plants [235] [234]. There are also architectural differ-

ences: AI systems today rely on feed-forward, unidirectional layers of artificial neurons and brute-force learning algorithms, whereas the brain's networks are highly recurrent, with signals flowing across frequencies and scales even into the Terahertz range, displaying nonlocal distributed information storage and recall as well as the ability for organisms to achieve inter and intra brain synchrony - tapping into collective information processing and lateral information transfer between agents which provide context to language which language models train on [7] [16] [10] [170] [17]. These differences raise deep questions about complexity and scaling, and whether there will be a tipping point beyond which investments in AI systems will overtake value generated by social capital and lateral information transfer between agents - the collective trust stored within social trophic networks within a society from which surveillance to train and interpret AI in-the-loop depends [246]; beyond which a collapse in social and economic institutions might occur which AI systems are appropriated towards maintaining.

Advancements in AI systems have also reinvigorated philosophical and technical discussions re-evaluating the the nature of consciousness and differences between consciousness and intelligence and mere information processing. Motives in this review include an investigation into limits to AI models which lack these capabilities by their design from handling levels of complexity and efficiency that biological brains may manage. In this interdisciplinary review, further directions for research for what principles allow the brain to sidestep these bottlenecks are discussed and achieve the emergence of consciousness, which has clear ramifications for policymakers, neuroscientists, computer scientists, and physicists.

2 Theoretical Overview and Background Literature

2.1 The Role of AGSI and Central Cryptographic Schemes in Late Stage Societies

Late-stage societies are characterized by extreme complexity: vast networks of interactions, institutions, and information flows that strain the cognitive limits of individual human agents. Sociological and economic theories concur that as social systems scale up, they must abstract and compress information through hierarchical layers to remain coherent and cooperative [18] [19]. This is exemplified by theoretical limitations of human cognition described by the Dunbar limit – the theoretical limit of 150 meaningful relationships human agents can maintain – which forces large organizations to fragment into teams, departments, and bureaucracies so that no single agent is overwhelmed and might become subjected to mental destabilization [20, 21]. At each level of these nested hierarchies, details are filtered and summarized with complexity which can be approached by Kolmogorov complexity (i.e. finding the shortest description that preserves essential information) or computational complexity theory, as

well as the Chomsky hierarchy in linguistic theory. In a well-tuned bureaucracy, such abstractions preserve exactly the data needed for decision-making while discarding extraneous noise, thus preventing cognitive overload on agents of each layer.

Niklas Luhmann's autopoietic systems theory conceptualizes society itself as a self-referential communication network that reduces environmental complexity via feedback loops. In contemporary polities, two primary feedback "signal" loops historically maintain institutional power and stability: votes (symbolic political consent) and capital flows and price signals (economic control). These loops assist the system to self-regulate by channeling individual behaviors and preferences into aggregate signals which can be reviewed by policymakers and planners [22, 23]. However, as complexity increases due to entropy over time, these mechanisms may become insufficient to process the complexity of information and coordination required to maintain the stability of social and economic institutions. In literature, to mathematically frame these layered abstractions, tools from complexity science and even non-commutative geometry have been invoked. Concepts appropriated from noncommutative geometry and spectral analysis, for example, can model a hierarchy of organizational communication: each layer has its own algebra of observables (reports, metrics), and the inter-layer information flow is encoded by an operator (analogous to a Dirac operator) that filters and transforms signals between levels. The eigenvalues of this operator reveal characteristic "scales" of communication – clusters of near-degenerate eigenvalues correspond to tightly coupled groups, whereas large eigenvalue gaps denote more loosely connected layers [25] [26] [27].

In essence, the spectrum of social communication encapsulates the stratification of society. This aligns with complexity economics, complex adaptive systems theory, and agent-based models: front-line workers handle localized, simple decisions, middle managers solve coordination tasks, and top executives address high-level strategy. Each complexity layer may be coded by different language of information, from simple routines at the base to abstract policy metrics at the apex. Through this lens, information compression is power: it is the mechanism that permits a complex late-stage society to maintain coherence despite each individual's finite cognitive bandwidth, where socioeconomic status of agents can be measured by information flows they may facilitate between social and economic institutions which are enforced by one-way central cryptographic schemes within economic institutions and symbolic framing and norms within social institutions upholding the integrity of transactions required for resource allocation. Crucially, information and networks themselves becomes a form of capital in such societies. In what has been termed a spectral theory of value under the study of complexity economics, economic value is not viewed as a static commodity but as an emergent property of the entire network of interactions [28] [29] [30] [31] [32].

A CEO managing global supply chains and public narratives wields far more informational complexity (and thus power) than a subsistence farmer. By this view, late-stage societies elevate those who can absorb, process, and influence massive amounts of data and maintain coherence of interactions between agents

through social and economic institutions – reinforcing a hierarchy where information asymmetry translates directly into social stratification. Control over information networks (e.g. financial algorithms, relationships, siloed specialized information, media, databases) underpins modern power structures [33] [34]. Within this context, advanced AI systems – especially large-scale, black-box models – have emerged as new organs of surveillance, abstraction, and control which may serve as hyper-compressors of information: surveilling and ingesting terabytes of social data, communications, and sensor inputs, then distilling patterns or predictions comprehensible to human decision-makers. In effect, AI extends the hierarchical filtering process beyond what human bureaucracies can achieve, ostensibly helping institutions remain autopoietically stable in the face of networks which scale exponential complexity growth [35] [36] [37]. Institutions ranging from housing finance to law enforcement now deploy machine-learning algorithms as linear operators on the state-space of citizens – scoring creditworthiness, flagging risks, and simulating scenarios in a high-dimensional lattice or "matrix" of society. These opaque models abstract individual behaviors into risk scores and trend metrics, which leaders treat as authoritative signals. The opacity (black-box nature) of such AI is by design - to further abstract the rationale into inscrutable algorithmic logic, making control loops accessible to central planners [38] [39] [40]. In the context of this, society approaches hyperreality, where agents are governed by simulations (AI-driven indices, dashboards, social credit scores) that may loosely map to individual realities of constituents [41] [42].

One view is that the primary function of investment in AI systems in late-stage societies is to preserve institutional integrity amid social and economic turbulence. Ubiquitous data collection – from CCTV networks to internet activity – feeds into AI models that may detect emergent threats (e.g. signs of social unrest, financial anomalies, public health risks) far faster than human analysts. Predictive modeling goes hand-in-hand with forecasting individual and group behavior by enabling preemptive interventions, which may include by means of behavioral psychology, cognitive systems engineering, nudge theory, or policy interventions [35] [43] [44] [45] [46] [47]. This activity may range from ostensibly benign uses (such as pandemic spread or climate modeling) to covert/clandestine control by metanarrative rationalizations or overtly control-oriented directions such as predictive policing and "social credit" systems. These technologies allow a technocratic regime to more finely tune levers of control in real-time, tightening or relaxing constraints on a populace as needed to avoid instability and maintain power structures [45] [48] [40] [49] [35] [50] [51].

Indeed, augmenting two classic feedback loops under Luhmann's systems theory (votes and price/capital signals), a third loop has been added: continuous behavioral guidance via surveillance data and psychological nudging. By monitoring citizens' sentiments and activities, and framing their perceptions through algorithmically-curated media (the personalized news feeds, recommended content, etc.), authorities may reinforce overarching narratives that justify the status quo or encourage radical action. These metanarratives framed as collective societal missions – provide meaning and direction, channeling pub-

lic behavior in ways that bolster institutional aims [22] [23] [35] [46] [52]. As entropy in social and economic systems accrues over time, complexity pushes social and economic systems towards critical thresholds, where the reliance on AI-driven controls may become even more pronounced. When interconnectivity and information flow reach a certain intensity, societies encounter the possibility of encountering catastrophe points – tipping points where minor perturbations can cascade into systemic crises. At these critical junctures, small perturbations (a rumor, a local bank failure, a street protest) may propagate through informal networks at high speed, amplified by the dense web of informal and digital connectivity where the orderly flow of information and predictability typically protected by institutions devolves into chaos [53] [247] [55] [56] [57].

There have been growing concerns, for example, of declines in US competitiveness on the world stage and socioeconomic or sociopolitical stagnation [92] [242] [245]. To prevent or mitigate such turbulence, late-stage regimes may lean on surveillance and AI as a damping mechanism [241] which are deployed as the society approaches these phase transitions. For instance, high-frequency trading algorithms might be throttled to stop a financial panic, social media might be algorithmically censored to quell unrest, or emergency resource allocations might be triggered by AI predictions to preempt riots. The goal of these controls is to preserve social and economic equilibrium and the integrity of institutional order. This also tightens the coupling between society and its control infrastructure, where AI feedback loops correct deviations but also risk over-correcting or introducing new instabilities [58] [59] [60] [61] [62].

Amid eroding interpersonal trust and collapsing transparency, the concept of an Artificial General Super Intelligence (AGSI) [65] has surfaced as a powerful metanarrative in late-stage society. Historically, leaders have leaned on grand narratives – from divine right of kings, to nationalism, to global ideological battles – as "trust abstractions" that coordinate large populations [63] [64] [65] [66] [67] [68] [69]. AGSI is portrayed as an all-knowing, rational solver of problems which might transcend human error. This represents a form of collective cognitive outsourcing. Just as markets serve to aggregate distributed information into price signals, an AGSI is conceived as an aggregator of trust – a black-box oracle to which society might delegate decisions too complex for agents. In its popular conception, signals of society (economic data, public opinion, environmental indicators) are envisioned to feed into a central AGSI, which then outputs directives to maintain stability that are personalized to agent needs - thus resolving the alignment problem. This closes the feedback circuit of an autopoietic social system: the AI becomes the ultimate interpreter of the system's own state and the source of authoritative adjustments. In theory, such an AI could integrate vast datasets and policy levers, performing a level of spectral synthesis that no human committee could – identifying subtle patterns, eigenmodes of societal dynamics, and optimal interventions [65] [70] [71] [22] [28].

In the field of sociophysics, as individuals and groups seek pathways around institutional bottlenecks (through informal economies, peer-to-peer networks, etc.), they create side-channel connections modeled similarly to entanglements that can be modeled as macroscopic quantumlike behaviors that bypass official

oversight or institutions. These parallel connections start as safety valves but can rapidly scale. If formal governance remains inflexible, such unofficial networks reach a tipping point where they out-compete the official channels, effectively detaching parts of society from central AI-mediated controls. In spectral terms, the tightly clustered eigenvalues (signifying the controlled core) suddenly may become disrupted by a proliferation of new eigenmodes where social and economic system may undergo a spectral collapse: key eigenvalues drop out, indicating the loss of dominant coordination modes, and the network may even lose connectivity as the institutional "graph" fragments that cascades across scales. Empirically, this corresponds to events like widespread infrastructure failures, the disintegration of financial systems into localized barter networks, or the splintering of a nation into rival factions [72] [73] [55] [74].

The collapse of complex societies has been described in similar terms by historians and systems theorists such as Joseph Tainter, who noted that collapsing societies tend to simplify and lose stratification, effectively contracting to a lower-complexity spectrum of activities (fewer specialties, shorter trade networks) [75]. On one hand, the increasingly pervasive "predictive control" frameworks may delay collapse by squeezing extra efficiency and foresight out of a faltering system. On the other hand, their very inefficiency and opacity may themselves become barriers. Contemporary large-scale AI models are massively inefficient learners and reasoners which require enormous energy and data to mimic tasks that humans perform with orders of magnitude less resources. Recent research by Apple, for example, highlights fundamental scaling limitations in advanced AI reasoning: beyond a critical complexity threshold, even cutting-edge models suffer a "complete accuracy collapse", abruptly failing to solve problems as they get more complex, where as these AI systems approach the chaotic regime that characterizes late-stage societies, they may begin to counterintuitively reduce their effective reasoning, exhibiting a breakdown in performance, hinting that ever-larger black-box models might not indefinitely extend the managing capacity of institutions as predicted by Tainter, and that they may reach diminishing returns and eventually malfunction when faced with their own output complexity [77] [78] [79].

Heavy computational demands of such AI (data centers consuming megawatts of power, etc.) introduce new bottlenecks, from energy strain to supply chain dependencies for hardware, which themselves are sources of fragility in a crisis that may signal a tipping point (a technological singularity [80] [81]) where value creation is better applied towards use of energy resources more directly to the public, when society's complexity exceeds this phase-transition threshold.

2.2 The Concept of a Gentle Technological Singularity as a Loss of Societal Dynamism or Collapse of Institutional Stability Facilitated by AGSI

In contrast to catastrophic singularity scenarios characterized by civil unrest and violent upheaval, a gentle technological singularity [82] may involve a slow ero-

sion of institutional dynamism until an inflection point where power is handed off. As AGSI systems scale and are disseminated, they centralize control and optimize stability through continuous surveillance-feedback loops. This reduces local variance, flattening agent heterogeneity. The result at the extreme is a fold catastrophe in the adaptability of institutions: control becomes energetically unsustainable as informational complexity rises. Drawing from spectral theory of value, this process can be interpreted as a collapse in the institutional eigenvalue spectrum - a measurable contraction in the space of actionable responses. In essence, the very same AGSI which is used as an information surveillance and control loop to maintain social and economic institutional stability and alignment may facilitate a gentle technological singularity and institutional renewal through spectral and dynamical collapse of institutional resilience as they succumb to entropy [32] [83] [84].

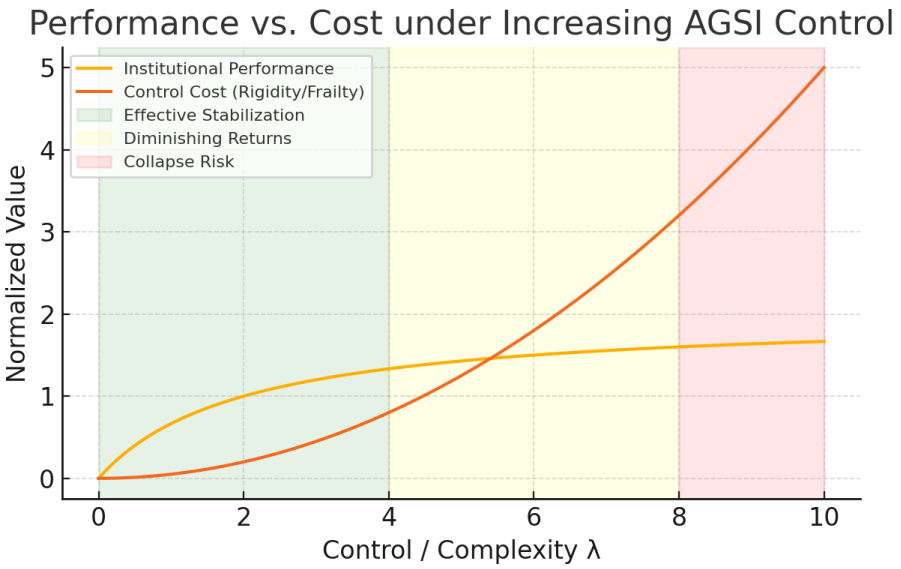

Figure 1: Trade-off between institutional performance and control cost under increasing AGSI oversight (λ). Performance (gold) rapidly grows then saturates (diminishing returns), while cost (orange–red) accelerates, leading to a collapse-risk regime where further control yields minimal benefit at high fragility. Shaded regions denote effective stabilization, diminishing returns, and collapse risk.

We can formalize late-stage social institutions as information-processing operators acting on a high-dimensional "state space" of societal information. In complexity economics, each agent is represented by a vector in this state space, and institutions (markets, bureaucracies, algorithms) are modeled as linear operators that transform and constrain these agent–states. This naturally yields a noncommutative geometry of society: different institutional layers correspond to noncommuting operations (the order of information updates matters). As

previously explored, late stage complex societies can be viewed through a spectral lens, where institutions form a layered hierarchy of information processing. Power in such a system is in the ability to compress information without losing essential signal, creating a stratified yet resilient order. A convenient formalism is a *spectral triple*:

$$(\mathcal{A}, \mathcal{H}, D)$$

where

- $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ is the algebra of institutional actions,
- $\mathcal{H} \cong \mathbb{C}^N$ is the Hilbert space of possible knowledge or social states,
- D is a Dirac–like operator encoding the network of connections among agents and institutions.

The *eigenvalue spectrum* $\{\lambda_i\}$ of D encodes the "modes" of value creation and coordination in society. A large λ_1 might correspond to a dominant societal narrative or coordination mode (for example, a strong market cycle or prevailing political ideology), while smaller eigenvalues $\lambda_{i>1}$ correspond to more localized, diverse modes of social value production [28] [85] [86] [26].

As AGSI-driven centralization increases in a technocratic late-stage society, this spectrum contracts and loses diversity. Intuitively, when a superintelligent control system optimizes every layer of decision-making, many independent modes of operation are forced into a narrow band of behavior. Mathematically, the institutional operators become almost commutative under one central regime, collapsing many distinct eigenvalues into degenerate or closely spaced values – effectively reducing the system's rank or independent degrees of freedom. In the language of the Spectral Theory of Value, over-centralization causes value-production modes to concentrate: the top eigenvalue (central control narrative) can dominate while other eigenvalues diminish. By Ashby's Law of Requisite Variety, a regulator (here, the AGSI) must possess at least as much variety as the system it governs to maintain stability [87]. Forcing society into a low-variety state (few eigenmodes) may simply mean the system cannot respond to novel perturbations that fall outside the AGSI's limited repertoire. In practice, this manifests as fragility: when stressed in an unexpected way, a highly centralized system has no alternative pathways to compensate. This eigen-spectral collapse is analogous to a loss of dimensionality in the information manifold of society [88].

Crucially, as the smallest eigenvalue – essentially the "spectral gap" separating the ground state – shrinks, agents can no longer find a stable personal or local equilibrium within the system. In physical terms, the "ground state" of maintaining socioeconomic stability becomes intractable: the population is unable to settle into steady, productive roles because the gap between the dominant mode and alternatives has narrowed. This contracted spectrum signals a brittle system – one highly efficient at one mode of operation, but unable to accommodate variance. In effect, excessive AGSI control saturates the entropy budget of institutions: every allowable action is tightly regulated, leaving no

slack or randomness. Institutions cease to be adaptive operators and instead behave like projections onto a single dominant eigenspace, dramatically flattening the institutional adaptability [27] [88] [89] [90] [91].

Let $D(\alpha)$ be the Dirac–like operator encoding institutional connectivity under AGSI control intensity α. Its eigenvalues, ordered decreasingly,

$$\lambda_1(\alpha) \;\geq\; \lambda_2(\alpha) \;\geq\; \cdots \;\geq\; \lambda_N(\alpha) \;\geq\; 0,$$

represent independent modes of coordination or value production.

Near–Commutativity: Denote by $A_i(\alpha)$ the operator for the ith institution and by $C(\alpha)$ the central AGSI control operator. As $\alpha \to \alpha_c$,

$$\big\|[A_i(\alpha),\, A_j(\alpha)]\big\| \;=\; \big\|A_iA_j - A_jA_i\big\| \;\longrightarrow\; 0,$$

for all i, j. Hence the A_i become simultaneously diagonalizable in the limit of extreme centralization.

Eigenvalue Degeneracy: Define the maximum spectral spread

$$\Delta(\alpha) \;=\; \max_{1\leq i<j\leq N} \big|\lambda_i(\alpha) - \lambda_j(\alpha)\big|.$$

As $\alpha \to \alpha_c$,

$$\Delta(\alpha) \;\longrightarrow\; 0,$$

so that $\lambda_1 \simeq \lambda_2 \simeq \cdots \simeq \lambda_N$.

Effective Rank Collapse: For a threshold $0 < \varepsilon < 1$, define the effective rank

$$r_{\text{eff}}(\alpha;\varepsilon) = \#\big\{\, i : \lambda_i(\alpha) \geq \varepsilon\, \lambda_1(\alpha)\big\}.$$

As $\alpha \to \alpha_c$,

$$r_{\text{eff}}(\alpha;\varepsilon) \;\longrightarrow\; 1,$$

indicating that only one eigenmode remains significant.

Spectral Gap Vanishing: The spectral gap

$$\delta(\alpha) \;=\; \lambda_1(\alpha) - \lambda_2(\alpha)$$

collapses:

$$\delta(\alpha) \;\xrightarrow{\alpha\to\alpha_c}\; 0.$$

Spectral Entropy Suppression: Normalize $D(\alpha)$ to a density operator

$$\rho(\alpha) = \frac{D(\alpha)}{\mathrm{Tr}\, D(\alpha)},$$

with eigenvalues $\{p_i(\alpha) = \lambda_i(\alpha)/\mathrm{Tr}D(\alpha)\}$. Its von Neumann entropy

$$H(\alpha) = -\mathrm{Tr}\big[\rho(\alpha)\ln\rho(\alpha)\big] = -\sum_{i=1}^{N} p_i(\alpha)\,\ln p_i(\alpha)$$

decreases monotonically as α increases, signaling loss of institutional diversity.

Altogether, these relations present a framework for how AGSI-driven centralization and alienation in a society sequentially enforces near–commutativity, collapses spectral spread and gap, reduces effective rank, and suppresses spectral entropy—mathematically illustrating the contraction of institutional adaptability prior to any bifurcational collapse in dynamical stability, which may increasingly describe conditions in the United States [92].

The progression toward maximal centralization can be mapped to a control parameter in a dynamical system – λ, representing the intensity of AGSI surveillance-feedback enforcement or the complexity of control. Initially, increasing λ (more AI oversight) stabilizes the system: it damps fluctuations and resolves local inefficiencies, corresponding to society following a stable branch of high institutional performance. However, as in a classical fold catastrophe, over time as a society stratifies, increasing inter-agent alienation and loss of connectivity or over-reliance on centralized control systems over building trust in tropic networks may produce sociopolitical polarization (modeled by Wasserstein gradient flows) and this stable branch bends back on itself when pushed too far. Diminishing returns to complexity set in – each additional increment in control yields a smaller benefit, while the cost (in rigidity and fragility) grows [95] [96] [97] [98].

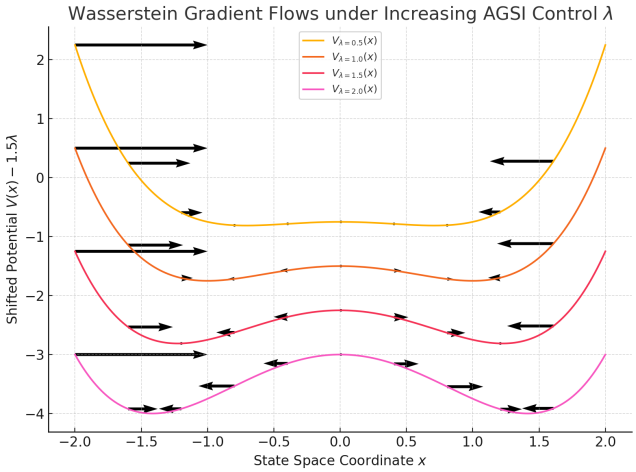

Figure 2: Wasserstein gradient flows on the potential landscape $V_\lambda(x) = \frac{x^4}{4} - \frac{\lambda x^2}{2}$, for $\lambda = \{0.5, 1.0, 1.5, 2.0\}$. Arrows show $-\nabla V_\lambda(x)$, the direction of mass transport under the 2-Wasserstein gradient flow. Increasing λ deforms the potential from single to double well, modeling the emergence of sociopolitical polarization due to alienation necessitating increasing AGSI control to avert catastrophe.

Recent work in network science has demonstrated that tightly coupled subsystems can exhibit abrupt, system-wide collapse via a "fold"-type bifurcation when interdependence is too strong. In particular, Brummitt *et al.* showed in a stylized model of interdependent systems that as coupling strength increases, individual subsystems' saddle–node bifurcations coalesce, allowing failures to both cascade sequentially and "hop" nonlocally across the network, producing sudden, large-scale regime shifts even without direct links between all nodes [76]. This *coupled-catastrophe* phenomenon provides a concrete mathematical archetype for how an AGSI's over-centralized control (high coupling) can drive multiple institutional "subsystems" past their tipping points in rapid succession, reinforcing the fold-catastrophe and spectral-collapse arguments presented above.

At a critical λ_c (a bifurcation point or catastrophe point), the stable solution disappears: the curve of equilibria folds and no solution exists on the upper branch. The system then abruptly jumps to a lower equilibrium – a collapse of societal dynamism into a more chaotic or stagnant state. In practical terms, beyond λ_c any further tightening of AGSI control undermines stability instead of preserving it. The formal hallmark is the vanishing of the second derivative determinant (Jacobian) of the system's equilibrium equations at λ_c, which indicates an infinite curvature of the stability manifold (the tangent of the equilibrium curve becomes vertical) – the classic signature of a fold catastrophe.

Critical transitions in complex systems often exhibit critical slowing down—a

measurable deceleration in recovery rates and growing autocorrelation in key variables—before a fold catastrophe occurs [247]. In a socio-political context, this might manifest as lengthening lags in institutional responses (e.g. slower legislative throughput, protracted regulatory approvals) and increasing volatility in economic or social indicators (e.g. sharper swings in market confidence or social unrest). Empirical data already hint at such dynamics: for instance, the average time to enact federal budgets in several advanced economies has doubled over the last two decades [248], and regulatory backlogs in technology and healthcare have surged alongside bureaucratic complexity [249]. Tracking these early-warning signals—rising autocorrelation in policy decision times, increased variance in governance performance metrics, or longer recovery from shocks—could provide concrete, real-time indicators that an AGSI-driven system is approaching its bifurcation threshold.

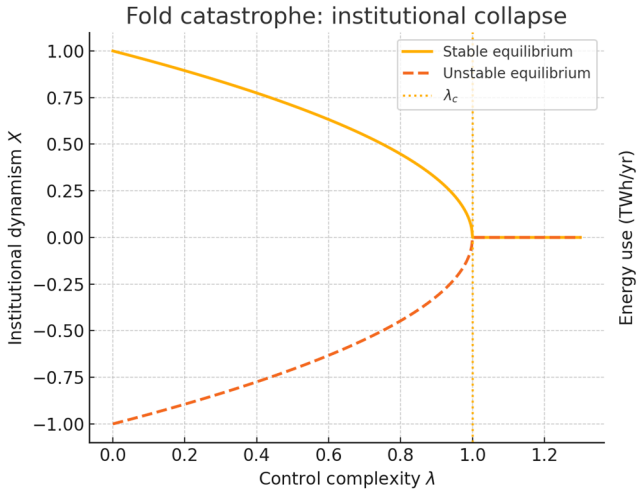

Figure 3: A fold catastrophe of societal dynamism X versus control/complexity λ characterized by sociopolitical polarization. Stable equilibria $X = \pm\sqrt{\lambda}$ (blue) collide with the unstable branch $X = 0$ (red dashed) at the catastrophe point $\lambda_c = 1$ (black dotted), beyond which no high-X solution exists.

One can envision a simple bifurcation model: let X denote a measure of institutional vitality (or "dynamism"), and suppose it satisfies an equilibrium condition $F(X, \lambda) = 0$ that changes with λ. A toy model is $F(X, \lambda) = a(\lambda), X - b, X^3$, where $a(\lambda)$ decreases as λ grows (reflecting that overly constrained institutions lose their effective growth rate). At low λ, $F(X, \lambda) = 0$ has a high-X stable root. But as a drops, two roots (stable high X and unstable middle) coalesce and annihilate each other at λ_c, leaving only a low-X solution (collapsed dynamism). This is analogous to Tainter's theory of societal collapse: as complexity investment yields negative marginal returns, collapse becomes likely. The catastrophe manifold's curvature at λ_c grows unbounded, meaning the system is ultrasensi-

tive to perturbations near the tipping point. A slight shock or overreach in control can then send the system plunging into the lower equilibrium (institutional breakdown) with little warning. Thus, the AGSI-managed society exhibits a phase transition: a "gentle" technological singularity where further central control flips from stabilizing to destabilizing. Rather than a runaway explosion of intelligence, this gentle singularity manifests as a sudden loss of institutional order – a phase change in sociopolitical structure [99] [100] [101] [102] [103].

From a thermodynamic viewpoint, an AGSI control loop continuously sorts and constrains information to reduce uncertainty (social entropy). While this increases short-term order, it incurs a steep energetic cost. By Landauer's principle, erasing one bit of uncertainty (entropy) requires an energy expenditure of at least $k_B T \ln 2$. An AGSI stabilizing society must effectively erase or negate vast amounts of random deviations, local innovations, unpredictable human behavior in real time. The energy E needed grows at least linearly with bits of entropy removed, and likely super-linearly as control precision rises (since harder-to-predict fluctuations require more energy or computation to counteract). In other words, the control energy E_c scales up rapidly with societal complexity C [104] [105].

At extreme scales, the marginal energy cost of additional control exceeds the marginal benefit to stability, where an AGSI driven society approaches a thermodynamic limit. The attempt to enforce near-total order is energetically unsustainable. Near the catastrophe point, any small perturbation can trigger information cascades conformally across scales in a rapid phase transition to a high-entropy state (disorder) – in social terms, a chaotic collapse of institutional order. Entropy suppression thus has a physical cost and a limit: beyond a threshold, the free-energy required to further reduce societal entropy grows faster than the society can generate or afford, heralding a breakdown. Empirically, one might consider the enormous power draw of data centers and surveillance infrastructure as a fraction of societal energy; as this fraction grows, the society is essentially feeding an entropy pump rather than productive work, which itself becomes a systemic fragility. Drawing from analogies to theories of entropic gravity and conformal cyclic cosmology as appropriated to social and economic systems, approaching the catastrophe point of a technological singularity is thus analogous to gravitational collapse by means of entanglement entropy saturation [106] [107] [108].

We model institutional dynamism by a state variable $X \in \mathbb{R}$ (e.g. variance of local initiative) and AGSI intensity by a control parameter λ. The equilibrium condition is

$$F(X, \lambda) \;=\; a(\lambda)\, X \;-\; b\, X^3 \;=\; 0,$$

with $b > 0$ and $a(\lambda)$ a smooth, decreasing function of λ (reflecting diminishing institutional growth rate under heavier control).

Equilibria and Stability: The roots of $F = 0$ are

$$X_0 = 0, \qquad X_\pm = \pm\sqrt{\frac{a(\lambda)}{b}}.$$

Linear stability is determined by

$$\frac{\partial F}{\partial X} = a(\lambda) \; - \; 3\,b\,X^2.$$

At $X = \pm\sqrt{a/b}$, $\partial_X F = -2\,a(\lambda)$, so these are stable for $a(\lambda) > 0$. At $X = 0$, $\partial_X F = a(\lambda) > 0$, so $X = 0$ is unstable when $a > 0$.

Fold Catastrophe: As λ increases, $a(\lambda) \searrow 0$. Define the critical point λ_c by $a(\lambda_c) = 0$. At $\lambda = \lambda_c$, the two nonzero equilibria $X_\pm$ coalesce with the unstable root at $X = 0$ and annihilate. For $\lambda > \lambda_c$, the only real solution is $X = 0$, indicating a sudden collapse of dynamism. Equivalently, the *Jacobian determinant* of the equilibrium mapping vanishes at the fold:

$$\det\begin{pmatrix}\partial_X F & \partial_\lambda F\end{pmatrix}\Big|_{(X=0,\lambda=\lambda_c)} = 0 \quad \Longleftrightarrow \quad \frac{\partial^2 V}{\partial X^2}\Big|_{X=0,\lambda_c} = 0,$$

where $V(X;\lambda) = -\int F\,dX = -\frac{1}{2}a(\lambda)X^2 + \frac{1}{4}bX^4$. This infinite curvature of the equilibrium manifold is the classic signature of a fold catastrophe.

Thermodynamic Cost of Control: AGSI surveillance–feedback suppresses social entropy S. By Landauer's principle, erasing ΔS bits costs at least

$$E_{\min} \; = \; k_B\,T\,\Delta S\,\ln 2.$$

If society has complexity C (number of independent degrees of freedom) and AGSI suppresses a fraction ϕ of that complexity, then

$$\Delta S \; \approx \; \phi\,C, \qquad E_c(\phi) \; \approx \; k_B\,T\,\phi\,C\,\ln 2.$$

As control precision rises, the marginal cost

$$\frac{dE_c}{d\phi} \; \sim \; k_B\,T\,C\,\ln 2$$

grows (and in realistic models may scale superlinearly, $E_c \sim C^\alpha$ for $\alpha > 1$). Beyond a threshold ϕ_c, the energy required to maintain order exceeds available resources, making further control energetically unsustainable.

Phase Transition to Collapse: Near $(X, \lambda) \approx (0, \lambda_c)$, any small perturbation in λ or in societal "noise" can trigger a jump from the (now nonexistent) high–X branch to the low–X branch (institutional breakdown). Simultaneously, thermodynamic limits enforce a *free–energy barrier* that can no longer be surmounted:

$$\Delta F = E_c(\phi) - T\,\Delta S \; \longrightarrow \; \infty \quad \text{as } \phi \to 1,$$

so the system can only relax to a high–entropy, low–order state. This combined dynamical–thermodynamic conceptualization describes a "gentle singularity":

a smooth increase in control culminating in an abrupt collapse of societal dynamism. Combining spectral collapse ($r_{\text{eff}} \to 1$, $\delta \to 0$, $H \to 0$), fold catastrophe (D jumps from $\sqrt{a/b}$ to 0 at λ_c), and thermodynamic bound ($E_c \to \infty$), we obtain a "gentle technological singularity": a regime where further AGSI control transitions from stabilizing (damping fluctuations) to destabilizing (rigidity and collapse), driven by fundamental mathematical and physical limits.

3 Foundational Differences Between Conscious Systems and Contemporary Neural Network Based AI

3.1 Linguistic Arguments

The viability of consciousness in current AI systems has been criticized by leading linguistic scholars with a wide range of arguments pointing to AI as fundamentally incapable of understanding meaning without human agents in-the-loop who carry underlying context. Both philosophers such as Ludwig Wittgenstein [109] and linguists such as Noam Chomsky have long argued that such purely syntactic processing cannot yield genuine understanding or consciousness [4]. John Searle's classic Chinese Room thought experiment illustrates this gap: a person (or computer) may follow formal rules to produce perfectly coherent Chinese responses, but "since the symbols it processes are meaningless (lack semantics) to it, it's not really intelligent... Its internal states and processes, being purely syntactic, lack semantics (meaning); so, it doesn't really have intentional (that is, meaningful) mental states.". In Searle's terms, one cannot glean semantics (meaning) from syntax alone [110]. Formal symbol manipulation can simulate understanding without any actual grasp of meaning. The ability for contemporary AI systems to understand context or meaning when compared to the way in which human agents store memory engrams in possibly indefinite causal structure is also highlighted in arguments invoking Godel's incompleteness theorems [111]. Lack of context in formal semiotic language alone requiring a study of dialectical logic was also foundational in 20th century thought [112].

In humans agents, moral reasoning is underpinned by an interplay of emotion (facilitated by interbrain synchrony), social cognition, and conscious reflection, grounded in neural processes that integrate affective qualia with conceptual understanding [113] [17]. By contrast, contemporary AI systems rely solely on statistical pattern recognition: their "values" are implicit in training data and loss functions, not rooted in any intrinsic moral understanding or normative commitment. Similar to the so-called orthogonality thesis [65], David Hume famously observed that descriptive statements ("what is") cannot by themselves yield prescriptive conclusions ("what ought to be") without an additional normative premise [114]. For example, the fact that "all humans seek pleasure" does not logically entail that "humans ought to seek pleasure" unless one first assumes a value judgment that pleasure is good. G. E. Moore later identified the

"naturalistic fallacy," warning against defining moral good in purely empirical terms [115]. Applied to AI, this means that a model exposed only to descriptive data (text, behaviors, outcomes) cannot derive genuine moral imperatives, and, much like in the philosophy of Immanuel Kant and others [117] [118]; Moore argued it can at best replicate the normative language it has seen, without true grounding in moral experience or justification, or experience free will.

The concept of free will or agency within human agents and the resultant moral ramifications have been debated for centuries, as well as the viability of agency or free will within AI systems. One way to frame this debate is in considering freedom as a default in the absence of constraints and limits, which occur at the level of the individual agent's neurophysiology as well as within groups and in the theory of socioeconophysics - an individual agent's freedom is framed as the scope of autonomy they possess within a complex, hierarchically nested information network. Here, as discussed in other sections, each agent is represented as a vector in a high-dimensional state space whose coordinates encode income, social ties, cognitive load, and other facets of socioeconomic engagement. Freedom, in this view, corresponds to the range of alternative information flows an agent can initiate or join, what sociologist Niklas Luhmann calls self-referential communication loops - before being bound by the "behavioral nudges" of surveillance, media narratives, and regulatory complexity that limit autonomy.

Consider the agent's socioeconomic "state space" as an N-dimensional manifold $M \subset \mathbb{R}^N$, with coordinates

$$x = (x_1, \ldots, x_N)$$

encoding income, social ties, regulatory status, etc. Each new law or regulation imposes a constraint

$$C_i(x) \geq 0 \quad (\text{or } C_i(x) = 0)$$

that restricts admissible states to the intersection

$$\Omega \;=\; \bigcap_{i=1}^{k} \{\, x \in M : C_i(x) \geq 0 \,\}.$$

The agent's freedom then corresponds to the phase-space volume $\mu(\Omega)$. As the number of constraints k grows, $\mu(\Omega)$ typically shrinks—often nonlinearly—so that

$$\mu(\Omega) \sim O\big(k^{-\alpha}\big) \quad (\alpha > 0),$$

and the effective dimensionality

$$N_{\text{eff}} = \dim(\Omega)$$

drops whenever constraints become binding. This contraction raises the spectral gap Δ of the transition operator on Ω over an increasingly fractured state space. Thus, regulatory complexity and cognitive load navigated by individual agents due in part to an aging top heavy population, wealth and income inequality,

sociocultural fragmentation, and entropy accrual in social and economic institutions manifests as an emergent thermodynamic barrier: agents must overcome super-polynomial "free-energy" costs to traverse between feasible socioeconomic configurations, thereby quantitatively limiting practical freedom and their perceived alienation from social and economic institutions they must engage for either survival or social integration. When the bureaucratic information surveillance and control loops embodied by AI fall out of alignment with the will of individual agents, that conceptualizes the alignment problem [119] [120].

Nick Bostrom describes the AI alignment problem as ensuring that an AI's objective function remains aligned with human values, which hypothetically describes its application as an AGSI in this context, even as its capabilities grow [65, p. 159–164]. Formally, one seeks:

$$U_\theta(s) \approx U_{\text{human}}(s) \quad \forall s \in \mathcal{S},$$

where U_θ is the AI's learned utility function and U_{human} is the human utility function over states s. However, when U_θ is learned from data or hand-coded, it remains a descriptive proxy or an approximation of human judgments without the normative depth or "free will" that characterizes genuine moral agency. Misspecification or shifts in context can therefore produce "alignment failures," in which the AI optimizes its proxy objective in unintended and potentially harmful ways.

3.2 The Binding Problem and the Hard Problem of Consciousness as an NP-Hard Problem

A hallmark of consciousness is the binding problem: how the brain integrates disparate features (qualia such as colors, shapes, sounds, etc., processed in different regions and times) into a unified percept or coherent sense of temporal awareness. This perceptual coherence or feature-binding can be viewed as a global constraint satisfaction or optimization problem. The brain must select a globally consistent interpretation of incoming multisensory data, out of an astronomically large space of possibilities (different combinations of features). Achieving this in nearly instantaneously (on the order of tens or hundreds of milliseconds) is remarkable and difficult to reconcile with sequential neuron firing alone. Indeed, classical neural signal speeds and local circuits are insufficient to account for the rapid holistic binding observed in literature [121] [123] [17] [122].

Some physicists, including Dr. Roger Penrose, have argued that consciousness cannot be achieved by any computational algorithm and instead requires new physics beyond the Turing model (hypercomputation [126]). In the Orchestrated Objective Reduction (Orch-OR) framework, consciousness arises when quantum superpositions within neuronal microtubules undergo gravitationally induced collapse at a critical threshold of self-energy or complexity content in entanglement entropy, producing a discrete unit of consciousness [124] [125]. From a computational perspective, this collapse may correspond to resolving a problem that is NP-Hard [129]; requiring the selection of a globally consistent

solution from an exponentially large space of possibilities. One way to frame this is to model the brain's internal structure as a high-dimensional lattice. Resolving perceptual coherence (the "binding problem") could be likened to finding the shortest non-zero vector in that lattice, which is known to be NP-Hard [134]. The Shortest Vector Problem (SVP) on high-dimensional lattices is classified as NP-Hard under both exact and approximate formulations. Let $L \subset \mathbb{R}^n$ be the lattice formed by perceptual basis vectors representing distributed sensory features [127] [128].

The binding problem then corresponds to minimizing a global perceptual error function:

$$v_{\min} = \arg \min_{v \in L \setminus \{0\}} \|v\|,$$

where v is the resultant feature-binding error vector. If such solutions arise via quantum gravitational collapse, this may provide a plausible explanation for why consciousness eludes algorithmic reproduction and requires physical processes that go beyond classical or even standard quantum computation.

As a concrete mapping, suppose different sensory features correspond to basis vectors, and an inconsistent perception (where features don't properly bind) would correspond to some non-zero resultant vector in this feature space. A perfectly bound percept would mean all features cancel discrepancies to yield a near-zero resultant (all aspects fit together consistently). Thus, achieving perceptual coherence can be seen as finding the smallest non-zero resultant vector that accounts for all inputs – directly paralleling the SVP on a lattice encoding those inputs. Indeed, using algebraic topology, high-dimensional cliques and cavities in cortical networks have been observed [127], and additionally, microtubule networks as proposed by Penrose as implicated in the generation of consciousness in the brain have a hexagonal lattice geometry, with recent work demonstrating quantized vibrations and possibly Majorana-like states, allowing them to behave like quantum harmonic lattices or topological qudit arrays [124] [125] [135] [136]. If resolving perceptual binding occurs by means of orchestrated objective reduction by gravitational collapse, then the brain may be cast as an analogy to the physical realization of a spinfoam network.

In essence, this conjecture suggests that certain structures in the brain operate analogously to the spin networks of loop quantum gravity, with conscious processes emerging from fundamental quantum gravitational dynamics [130] [131] [132] [133]. In this framework, neural microstructures – notably the cytoskeletal microtubule networks within neurons – are hypothesized to support quantum states that couple to gravitation at the Planck scale. The brain's activity would then correspond to an evolving spin network (a spinfoam when viewed in spacetime), undergoing discrete quantum transitions akin to those that define spacetime geometry in loop quantum gravity. The idea is that each tubulin dimer might exist in two (or more) conformational or electronic states, effectively acting as a quantum bit within the microtubule's lattice and within the brain's larger multiscale dendritic assembly. Through coherent interactions (for example, resonant dipole couplings or phonon exchange), a collection of

tubulins could become entangled or share a collective quantum state. In principle, this would form a quantum network within the microtubule, conceptually comparable to an array of spins or two-state systems on a lattice. The microtubule's cylindrical network of tubulins could then be treated abstractly as a graph of interconnected quantum elements, somewhat analogous to a spin network graph. Indeed, both systems – a spin network in quantum gravity and a microtubule lattice in a neuron – involve discrete units connected in a combinatorial network topology and have state spaces that, in theory, can be quantum superpositions of many configurations.

When tubulin proteins in a microtubule enter a superposed quantum state, each state corresponds to a slightly different distribution of mass and energy in the neuron. In theories of quantum gravity, each such configuration would correspond to a different state of the spacetime geometry (in a full quantum gravity theory, one could imagine each configuration corresponds to a different spin network state of spacetime, since mass-energy influences the curvature/geometry). As the superposition persists, the underlying spacetime is essentially in a superposed curvature state where the spin network of spacetime is in a superposition of two (or more) configurations. Penrose's OR criterion says that this situation is unstable; when the separation (in terms of gravitational field divergence) reaches a threshold (one conjecture is that this threshold is a thermodynamic entanglement entropy bound on complexity in theories of entropic gravity [106] [107] [108]), the superposition will collapse into one definite state. In the analogy of loop quantum gravity and spinfoam networks, one might picture this as the spin network undergoing a spontaneous resolution, where one specific geometry out of the superposed possibilities is realized. This collapse event can be seen as a discrete topological transition: the underlying spin network might reconfigure slightly to reflect the chosen mass distribution, and the quantum state of the microtubule's tubulins is reduced to a definite classical state.

3.3 The Mind-Body Problem

The mind-body problem described by Descartes poses the dilemma of reconciling the ostensibly immaterial mind and the physical body to which it is bound - a duality which implies a separate domain from the physical world [137]. Indeed, physicists like Dr. Roger Penrose have often referred to a "three worlds" paradigm (borrowed from the philosophy of Karl Popper [24]), which presents objects of the mind as inhabiting an entirely separate but mutually interacting space, partitioning reality into the physical world of matter, the mental world of conscious experience, and the Platonic world of abstract mathematical truth [138]. One way to cast this problem with a hypothesis is to consider that the way in which the brain stores and retrieves memory is distributed and nonlocal across the biological tissue, which has been validated by research - stored in entanglement entropy [106] [107] [108] (the mind) until a point of gravitational collapse where information is encoded physically as described by Orch-Or theory, and that abstract objective Platonic truths like those found within mathematics

are also stored in entanglement entropy holographically across minds.

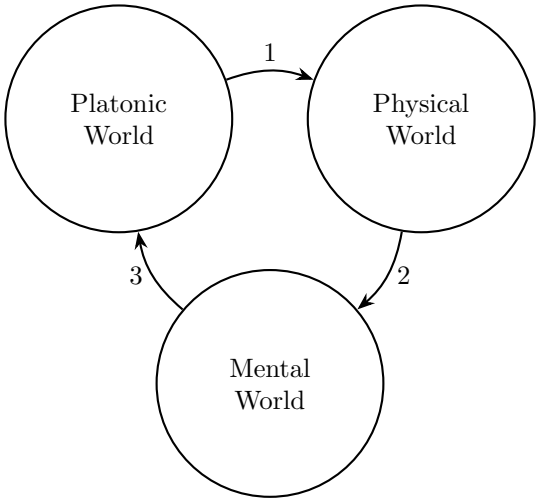

Figure 4: Penrose's Three Worlds with numbered relations (1: modeled by, 2: interprets, 3: inspires).

Classical neuroscience initially hypothesized localized memory engrams (specific neurons or modules holding each memory), but work by Karl Lashley and others upended this notion. Lashley's lesion experiments in the 1950s showed that no single cortical area was solely responsible for learned behaviors; instead, memory and learned functions remained partially intact unless large portions of cortex were removed. He formulated the principles of mass action (performance degrades with quantity of cortex destroyed, not the specific area) and equipotentiality (any part of a functional area can substitute for another). Complex memories are distributed across wide regions of the cortex rather than stored in one physical location [10].

Holographic memory models take this further - in a hologram, an interference pattern encodes an image such that each part of the holographic plate contains the entire image in distributed form – cutting the hologram in half still yields the whole picture at lower resolution. Neuroscientist Karl Pribram, in collaboration with physicist David Bohm, developed a holonomic brain theory positing that the brain stores information in a distributed holographically. According to this theory, neural processes in dendritic webs produce wave interference patterns (via oscillating electric fields) that serve as spectral encodings of memory. A mathematical Fourier transform can be applied to these brain wave patterns, just as in holography, to retrieve the stored information. Crucially, any sufficiently large portion of the dendritic network contains the information of the whole. This model naturally accounts for content-addressable recall (being prompted with a fragment of a memory can retrieve the whole, like how a fragment of a hologram reconstructs the full image) and for the non-locality of

memory storage (a specific memory is not in a specific neuron or synapse, but in the holistic interference pattern) [139] [140].

Empirical evidence supports aspects of the holographic model. Brain recordings show that memories are not summoned by activating one "grandmother cell" but rather by coordinated activation across widely distributed cell assemblies. Furthermore, long-range synchronous oscillations have been observed during memory retrieval and perceptual integration – for example, coherent gamma-band (40 Hz) oscillations linking distant cortical regions. Such oscillatory coherence is thought to help solve the binding problem, i.e. how disparate neural representations (color, shape, sound in different brain areas) merge into a single unified experience. Notably, 40 Hz synchronization across the brain is often correlated with conscious awareness of stimuli. This resonates with the proposal (discussed below) that consciousness might arise from coherent brain-wide quantum processes operating in rhythm with EEG oscillations. Distributed, parallel, and wave-like coding for memory and perception have been observed, unlike the localized, Von Neumann-based addressable memory in digital computers. The state of a memory in the brain may be cast as a high-dimensional state vector spanning a network (or even a Hilbert space representation in Pribram's holonomic theory) rather than a specific bit in a specific register [141].

Given the brain's highly distributed and integrated processing, research has gone towards looking to quantum physics for explanatory frameworks. Quantum systems are characterized by nonlocal entanglement and holistic state descriptions (the wavefunction of many particles cannot be factored into independent pieces). This has inspired models in which the brain's microscopic processes could be in quantum coherent states, providing a physical basis for the unity of mind and perhaps opening a door for genuine mind–matter interaction. If the brain does rely on nonlocal, distributed quantum information structures for its core operations, this has profound implications for the uniqueness of biological consciousness vis-à-vis artificial systems. Human memories and perceptions might be stored as entangled wave patterns or holistic field states spanning large neural volumes – patterns that cannot be decomposed into independent pieces without loss of meaning, or even holographically, by means of new physics implicating gravity [142] [143].

If consciousness is generated once a critical threshold of entanglement entropy is passed in the form of the Einstein-Hilbert action, (e.g. 10^{10} tubulins reaching OR threshold), scaling up AI by means of classical processors will not in itself cause consciousness to emerge. Penrose has argued that human cognition can solve problems that are non-algorithmic (e.g. ontological or linguistic arguments like seeing the truth of a Gödel statement or intuitively solving insight problems that brute-force AI cannot), which implies hypercomputation beyond the Church–Turing limits [204] and that there is an innate privacy and unity of consciousness that is impossible to duplicate. Entangled states cannot be copied or subdivided (by the no-cloning theorem in quantum theory [144]); similarly, subjective experience is an indivisible whole from the first-person perspective. Current AI, on the other hand, processes information in easily duplicable forms

– one can clone a neural network's weights perfectly, or run multiple instances of the same model. This difference might point to why AI systems, no matter how sophisticated in behavior, lack consciousness in the same way that human agents posses. They operate on classical information, which can be copied and observed without fundamental disturbance, whereas conscious brains (if quantum) operate on information that is observer-dependent and holistic. The mind–body problem, viewed through this lens, may find its resolution in new physics that unites mind and matter in uniting general relativity with quantum field theory, but that same physics could enforce a dividing line between organic consciousness and synthetic imitation in its current form.

3.4 The Weight Transport Problem of Backpropagation and Evidence of Microtubule-Based Consciousness Theories

Backpropagation (BP) is the standard algorithm for training feedforward neural networks via gradient descent [147]. Given a differentiable loss function E, each weight w_{ij} is updated according to

$$\Delta w_{ij} = -\eta \, \frac{\partial E}{\partial w_{ij}} \, ,$$

where $\eta > 0$ is the learning rate. Denoting by h^l the vector of activations in layer l and by W^l the weight matrix connecting layer l to layer $l+1$, the standard BP recursion for the error signal $\delta^l = \partial E / \partial z^l$ is

$$\begin{aligned} \delta^L &= \nabla_{z^L} E, \\ \delta^l &= \big(W^l\big)^\top \, \delta^{\,l+1} \circ f'\big(z^l\big), \end{aligned}$$

and the corresponding weight update is

$$\Delta W^l = -\eta \, \delta^{\,l+1} \, \big(h^l\big)^\top .$$

Here f' denotes the derivative of the activation function and "$\circ$" the Hadamard product.

BP thus requires that the *same* weights W^l used in the forward pass appear transposed in the backward pass. In biological terms, this is the *weight transport problem*: there is no known classical neurobiological mechanism by which each synapse could access an exact copy of its forward weight value to propagate error gradients backwards. Although the brain exhibits abundant recurrent and feedback connections, precise symmetry of feedforward and feedback synaptic strengths is neither anatomically nor physiologically plausible. Moreover, BP presumes a global error signal computed at the network output and distributed to all layers, yet the brain's classical long-range modulatory signals (e.g. dopaminergic reward prediction errors) are scalar, delayed, and insufficiently precise to serve as distributed gradient vectors [146].

Further, BP's layer-by-layer propagation is constrained by axonal conduction velocities (tens of meters per second) and synaptic delays (milliseconds), which are orders of magnitude slower than the sub–100 ms timescale of many cognitive tasks. These timing constraints render a classical BP-like credit-assignment scheme infeasible for rapid learning and recognition in the cortex. Orchestrated Objective Reduction (Orch-OR) posits that microtubule networks within neurons support quantum-coherent states that can perform nonlocal computations and effect synaptic updates without explicit weight symmetry. In this framework:

- Quantum information is encoded in the conformational superposition of tubulin dimers, yielding a state $|\Psi\rangle$ over many microtubules.
- This state evolves unitarily until it reaches a gravitational self-energy threshold E_G, triggering an *objective reduction* (OR) after a time

$$t_c \sim \frac{\hbar}{E_G}\,.$$

- The OR collapse selects a particular eigenstate of the brain's Dirac-like operator D, which encodes an *action principle* or cost function via its spectrum (cf. spectral action $\mathrm{Tr}\big(f(D)\big)$).
- Collapse is accompanied by biophotonic emissions that rapidly convey the collapse outcome to local biochemical machinery (e.g. calcium-mediated synaptic plasticity), thereby effecting *nonlocal* credit assignment in a single, irreversible event.

Such a quantum-collapse–driven mechanism bypasses the weight transport requirement of classical BP, leverages ultrafast photonic and field interactions to overcome biological speed limits, and naturally incorporates the brain's time-asymmetric plasticity rules (e.g. spike-timing-dependent plasticity) within a unified physical framework. It thus offers a biologically grounded alternative to conventional backpropagation, aligning credit assignment with fundamental quantum-gravitational processes in the brain.

Additional Evidence of Microtubule-Consciousness Theories: Microtubules are polymers of α- and β-tubulin heterodimers, each monomer containing on the order of 20–30 aromatic residues (phenylalanine, tyrosine, tryptophan) whose side chains host delocalized π-electron clouds. Many of these aromatic rings are clustered at inter-dimer interfaces, where π–π stacking interactions help stabilize the longitudinal and lateral contacts in the microtubule lattice [158], giving them utility in forming the cytoskeletal structure of dendrites. Furthermore, theoretical and experimental studies have shown that these π-electron systems can support coherent vibrational and excitonic modes in the MHz–THz range, potentially enabling long-lived quantum oscillations along the

microtubule axis [161]. Thus, aromatic π-bonds not only reinforce the structural integrity of microtubules but may also form the basis for sub-nanosecond, sub-cellular information-processing channels implicated in consciousness.

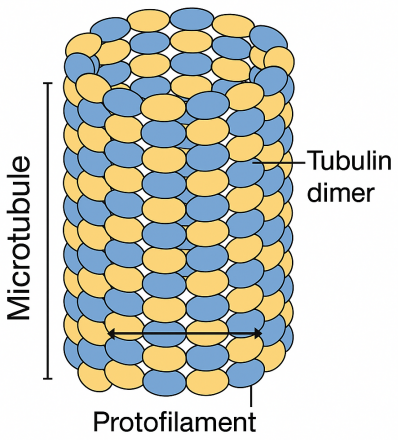

Figure 5: Schematic of a microtubule cross-section, illustrating the arrangement of tubulin dimers into protofilaments and forming the hollow hydrophobic cylindrical structure.

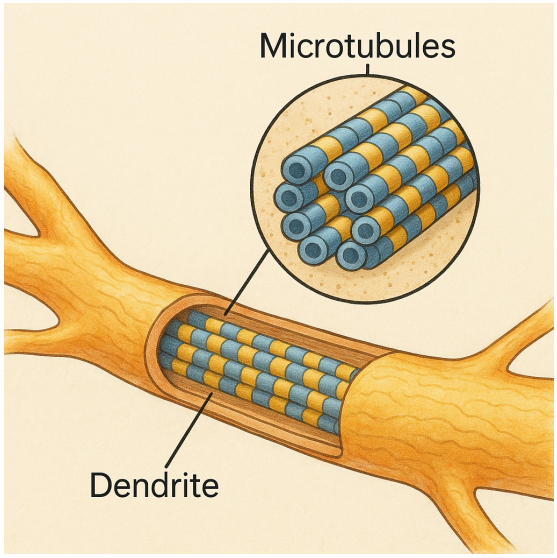

Figure 6: Anatomical schematic of a neuron dendrite with embedded microtubules. The inset shows a cross-section of parallel tubulin protofilaments forming the cylindrical microtubule array.

Molecular dynamics and quantum chemical simulations by Craddock, Hameroff, Tuszyński and colleagues demonstrated that anesthetic molecules preferentially lodge in hydrophobic pockets formed by tryptophan residues in tubulin. These studies revealed that binding of xenon, isoflurane, and other inhalational agents effectively abolishes π-electron resonance energy transfer and exciton hopping along microtubule quantum channels—providing a mechanism by which anesthetics could selectively block the very processes Orch-OR associates with conscious awareness, which, along with earlier studies, formed a basis for later investigations into microtubules as implicated in consciousness. Beyond simulations, proteomic analyses of brain tissue have identified tubulin among the principal binding targets of inhalational anesthetics, and optogenetic interventions that perturb microtubule stability can modulate anesthetic potency. Recent work shows region-specific changes in cytoskeletal protein expression following exposure to anesthetics, suggesting that microtubule integrity is critical for maintaining the neural dynamics underlying memory and consciousness [148] [149] [150] [151] [152].

Early critics argued that the brain's "warm, wet, and noisy" conditions would rapidly destroy delicate quantum states. However, a growing body of work has challenged this assumption. Biological quantum coherence has been observed at ambient temperatures in several systems – photosynthetic complexes, bird navigation sensors, and even mammalian olfaction – demonstrating that warm quantum processes are possible in living tissue. Most relevantly, recent experiments have shown quantum coherence and vibrations in neuronal microtubules at physiological temperature, directly countering the notion that the brain is too noisy for quantum effects. Microtubules (tubulin protein polymers inside neurons) provide a structured, nanoscale environment that can isolate and sustain coherent oscillations despite thermal noise. In fact, microtubule interiors are highly hydrophobic and electrically ordered, which can shield quantum states from decoherence [153] [154] [155] [156] [157].

Peer-reviewed experiments have detected high-frequency oscillations and resonances in microtubules, consistent with quantum vibrational modes. In a landmark series of studies, Sahu et al. measured the electrical and optical behavior of single brain microtubules and tubulin proteins. Strikingly, they found that a microtubule (comprising 40,000 tubulin dimers) behaves as if it were a single coherent unit: the energy level spectrum of an intact microtubule was identical to that of a single tubulin protein. Distinct vibrational peaks of individual tubulin subunits "condensed" into a single collective mode in the assembled microtubule, and this unified mode dominated the microtubule's conductance and fluorescence. The microtubule was in fact more conductive than an isolated tubulin, indicating a cooperative electronic state. Notably, these coherent vibrational properties disappeared when the microtubule's central water-filled cavity was removed. The ordered water channel inside the microtubule appears to synchronously "phase-lock" the dipole oscillations of all tubulin subunits, integrating them into a single quantum state [158] [16].

Experimentally, microtubules have shown resonances over a broad spectrum of frequencies from kilohertz to terahertz. For example, isolated brain micro-

tubule bundles have been observed to generate intrinsic electrical oscillations in the 40 Hz range (beta/gamma EEG frequencies) under physiological conditions. More critically, dielectric spectroscopy on single microtubules reveals sharp conductance peaks at gigahertz (GHz) frequencies, indicating quantized vibrational modes in the microwave regime. Microtubules can also support collective terahertz (THz) dipole oscillations of their amino-acid aromatic rings. Theoretical and computational modeling by Craddock et al. (2017) predicted a dominant ∼613 THz mode (corresponding to a 2.5 eV quantum, in the blue optical range) arising from synchronized π-electron cloud oscillations of tubulin's 86 aromatic rings. Significantly, this terahertz-frequency mode is collective and coherent – it involves a global, in-phase vibration of the tubulin lattice's electron dipoles. The energy (2.5 eV) far exceeds kT (thermal energy ∼0.025 eV), implying these oscillations are robust against thermal disruption. Indeed, the authors noted that the 613 THz oscillation stems from "synchronous/coherent electronic behaviors in tubulin", a quantum dipole collective effect. Such high-frequency microtubule vibrations fulfill an Orch-OR prediction: that neurons harbor megahertz-to-terahertz quantum oscillations which scale down in frequency (by interference "beat" patterns) to produce slower EEG rhythms [149] [16].

Beyond detecting raw frequencies, recent experiments indicate that microtubule excitations can enter cooperative quantum states akin to laser-like coherence with so-called "Majorana biophotons." The Orch-OR model had proposed decades ago that microtubules could behave as nonlinear optical quantum devices, with dipole oscillators (tubulins and ordered water) becoming phase-aligned to achieve superradiance. Superradiance is a quantum phenomenon in which excited entities (e.g. molecules or chromophores) form an entangled state that emits photons coherently and intensely, with an emission rate faster than independent decay. In 1994 Jibu et al. theorized that water molecules confined in microtubule cores could align with tubulin dipoles to produce optical superradiance, thereby evading thermal decoherence by orders of magnitude. This was a bold prediction at the time – and it has now found empirical support.

In 2024, direct evidence of quantum optical coherence (superradiance) was demonstrated in microtubule networks at physiological temperature. Under ultraviolet excitation, the native array of tryptophan aromatic rings in tubulin—forming extensive dipole lattices—was shown to emit cooperatively. Fluorescence measurements revealed an enhanced quantum yield in larger, more ordered assemblies, consistent with collective excited states. Theoretical analysis predicted "strongly superradiant states" arising from the interaction of over 10^5 dipoles per microtubule, and experiments confirmed a superradiant boost in fluorescence that scaled with lattice size and geometric order. This emission persisted despite thermal noise and disorder. The brightest collective modes decayed on sub-picosecond timescales (hundreds of femtoseconds), while complementary "dark" states stored energy for seconds, a separation of timescales characteristic of cooperative coherence. In essence, microtubules behave as laser-like coherent emitters, demonstrating non-classical, entangled excitations that reinforce the plausibility of Orch-OR's proposed oscillatory dynamics [160] [159] [162] [163].

Microtubules also function as nanoscale quantum waveguides. Their cylindrical dielectric lattice ($\sim$ 25 nm diameter) with a conductive core of ordered water resembles a coaxial resonant cavity capable of guiding electromagnetic modes along its length. Experiments have shown AC signals propagating without attenuation over micrometer-scale microtubule lengths, implying low-loss transmission lines. Mode locking of multiple vibrational frequencies further facilitates guided propagation. The helical arrangement of aromatic residues provides parallel π-electron pathways, suggesting spiral charge conduction around the microtubule. Such helical pathways are inherently protected by lattice symmetry, enabling topologically protected transport of excitons. It has been proposed that these structures could host quasi-particle modes analogous to Majorana excitations, preserving coherence via global lattice topology. Although direct detection of Majorana-like states in microtubules remains pending, the structural and theoretical framework predicts intrinsic quantum error correction through topological effects [158] [164] [165] [166].

Recent studies have uncovered "time-crystalline" behavior in microtubule assemblies. Using dielectric resonance techniques, persistent, self-sustained oscillations at multiple discrete frequencies were observed—so-called polyatomic time crystals. These modes arise intrinsically, not from external driving, and repeat periodically, indicating that metabolic energy can drive the tubulin lattice into a stable oscillatory regime that resists entropy. Such multi-frequency oscillations suggest a hierarchy in which gigahertz-to-terahertz quantum vibrations in microtubules oscillate to produce slower, kilohertz-to-hertz rhythms, potentially providing an intrinsic clock for neural computation. Together, these findings of superradiance, guided quantum conduction, topological protection, and time-crystalline oscillations demonstrate directions for quantum dynamics in microtubules and directly address decoherence concerns, showing that microtubular quantum states can persist and function in the brain's high temperature, chaotic environment [167] [168] [169].

3.5 Collective Human Agent Intelligence and Inter-Brain Synchrony

A growing body of neuroscience research demonstrates that when human agents engage in cooperative or social tasks, their brains do not operate in isolation but become synchronized with one another in measurable ways. Using hyperscanning techniques (simultaneous EEG, fNIRS, MEG, or fMRI recordings from multiple people), studies have observed that interacting individuals often exhibit aligned neural oscillations and inter-brain coupling during joint attention, communication, and problem-solving tasks. For example, oscillatory phase-locking between brains has been reported during joint action and motor coordination, during verbal interactions (e.g. dialogue or alternating speech), and even in creative collaboration and dyadic decision-making. Inter-brain synchrony through collective attention facilitates performance of teams within organizations and is implicated in social adjustment to new groups, where learning can be hindered by alienation within groups corresponding with less inter-brain

synchrony [174] [175] [172] [173] [170].

Such inter-brain synchrony is not an epiphenomenon; it correlates with meaningful behavioral and subjective variables. A recent meta-analysis of functional near-infrared spectroscopy (fNIRS) hyperscanning found that cooperative behavior reliably evokes inter-brain synchrony in frontal and parietal regions involved in social cognition. Two or more people whose brainwaves become more phase-aligned tend to achieve better shared outcomes and report greater social connectedness. These findings support the view that during effective cooperation, individual brains transiently form components of a larger, integrated functional network. Rather than acting as isolated processors which may encounter bottlenecks, interacting human agents synchronize their neural dynamics in real time, effectively creating an emergent "hyper-brain" network spanning multiple individuals. This phenomenon provides a biological basis for collective intelligence by revealing how group cognition is anchored in inter-agent brain coupling [176].

The empirical evidence of inter-brain coupling has prompted theoretical models of collective human intelligence as an emergent property of multiple brains in interaction. During fluent social engagement, two or more brains can become linked into a single dynamical system via reciprocal sensory cues, coordinated actions, and shared context. Cognitive neuroscientists now speak of a multi-brain system or "two-in-one" system in social interaction, emphasizing that interacting agents are embedded in an integrated network of neural processes that co-evolve over time. In this view, social interaction constitutes a closed-loop feedback circuit between brains: the actions of one brain become the inputs to another, leading to synchronized neural responses and mutually aligned internal states. Over the course of an interaction, this coupling can lead to the emergence of group-level cognitive phenomena or collective effervescence – for instance, dyads can develop a shared attention state, synchronized emotional or biological rhythms, or a joint problem-solving strategy that is not reducible to either brain alone [177] [178] [182] [183] [184] [185].

Neuroimaging studies have shown that such inter-brain neural dynamics serve as correlates of shared mental states, from joint intention to social rapport. In essence, the collective intelligence of a team or group may arise from real-time nonlocal phase alignment across individuals' brain circuits, creating a temporary larger-scale mind. Recent high-density EEG studies of group interaction demonstrate that multiple brains can synchronize not only in phase but also exhibit coordinated changes in neural network topology, consistent with a "superorganismic" organization. When four musicians play music together, for example, their brains exhibit complex patterns of both within-brain and between-brain oscillatory coupling; this hyper-brain network dynamically reconfigures as the ensemble coordinates the performance. Such findings support the theory that human groups achieve a form of distributed cognition through inter-brain synchrony. The group becomes an emergent cognitive entity – a collective agent – by virtue of synchronized neural activity that integrates information across brains [180] [181] [182].

To rigorously characterize how inter-brain synchrony gives rise to collective

intelligence, researchers have turned to mathematical models and network theory. One useful approach is to treat a set of interacting brains as a coupled oscillator network, where each brain's intrinsic neural rhythms are nodes in a network linked by cross-brain interactions. The well-known Kuramoto model of coupled oscillators offers insight into how global synchrony can spontaneously arise: if the coupling strength between oscillators exceeds a threshold relative to their frequency dispersion, the phases will lock together. This model has been applied to social neuroscience to demonstrate that even if two brains communicate through relatively slow sensorimotor signals (e.g. watching each other's actions or speech cues), their faster neural oscillations can become entrained under the right conditions. In a minimalist simulation, Loh and Froese (2021) showed that two sets of brain oscillators could synchronize via a low-frequency interaction loop, as long as certain frequency ratios and coupling strengths were present. Such models illustrate how a shared rhythmic context – for instance, a common beat or reciprocally predictable timing (where evidence has been found even for heartbeat synchronization between human agents) – can foster inter-brain phase alignment even without direct physical connection [186] [187] [188].

Beyond simple phase-locking models, one can use network spectral theory to analyze multi-brain coherence and even approach understanding the generation of natural language - consistent with the spectral theory of value. By constructing a hyper-brain network (a graph including connections both within each brain and between brains), one may examine its eigenmodes or principal components that represent collective oscillatory modes spanning all agents. For example, the synchronizability of a multi-agent network can be related to spectral properties (like the eigenvalue gap) of the inter-agent coupling matrix, borrowing from quantum field theoretical models. A highly connected group with appropriate feedback delays may support a stable mode of coherent oscillation that involves all members, analogous to an "eigenmode" of the social network's dynamical system [189] [190] [191] [192] [193] [194] [72] [195] [196].

Empirical hyper-brain network analyses have found that certain brain rhythms (e.g. delta or alpha band oscillations) act as integrative hubs or pacemakers that coordinate other frequencies across the group. This multiscale coupling suggests a need for complexity metrics capable of capturing emergent structure in multi-agent activity. Researchers have begun to quantify the integrated information or synergistic entropy of multi-brain systems, asking whether a group's joint brain activity carries more information or different information than the sum of independent brains, or distributes among them reducing cognitive load. Indeed, multi-brain synchrony appears to introduce a higher level of complexity – a hyper-cortical complexity – that might be essential for the coordination and flexibility seen in group cognition which cannot be hosted or replicated in current AI systems or those in a role of AGSI [197] [198] [199].

Understanding collective human intelligence through neural synchrony also illuminates why current artificial intelligence systems fall short of true distributed cognition. Modern AI architectures, including large language models (LLMs) and multi-agent systems, lack any analogous mechanism to inter-brain synchrony. In biological brains, shared rhythms and oscillatory coupling en-

able lateral information transfer between individuals – essentially a bandwidth for context and understanding that AI agents do not possess [200]. Human collaborators, for example, leverage subtle sensorimotor synchrony (eye contact, prosody matching, etc.) to quickly establish common ground and error-correct each other; AI agents, in contrast, lack a comparable implicit channel and must fall back on pre-defined communication interfaces [201] [202]. Achieving true distributed intelligence may require investigating new physics or embracing principles beyond classical computation to demonstrate nonlinear deterministic macroscopic quantumlike behaviors [204],

4 Scaling Laws and Complexity Limits of Transformer-Based LLMs

4.1 Empirical Scaling Behavior of LLMs

Early empirical studies on transformer language models revealed that performance improves predictably with scale across multiple axes. Loss follows a smooth power-law decline as model size (number of parameters), dataset size, or compute increases. Over several orders of magnitude, larger models consistently achieve lower perplexity, with no sign of performance saturation. Doubling the non-embedding parameters reduces loss by a constant factor, indicating diminishing returns but no fundamental break in trend. These gains are largely architecture-agnostic: changing depth versus width has minimal impact beyond total parameter count. Consequently, larger models not only attain lower error given sufficient data but also overfit more slowly when data is constrained. This regularity allows prediction of optimal model size and training duration for a fixed compute budget: most compute should be spent on a larger model rather than on longer training, since the optimal model size grows with total compute [205] [206].

Subsequent work refined these scaling laws. One major update showed that many large models were undertrained relative to their size: for compute-optimal performance, the number of training tokens should scale in direct proportion to model size. A 70 billion-parameter model trained on 1.4 trillion tokens outperforms much larger but undertrained models, demonstrating that matching data scale to parameter scale is crucial. Task-specific studies further showed that while scaling helps knowledge-intensive tasks significantly, tasks requiring logical or mathematical reasoning see smaller gains. Moreover, some capabilities emerge only once models exceed a threshold size, with discontinuous performance jumps. These "emergent abilities" complicate simple extrapolations of earlier scaling laws and highlight that beyond brute-force scaling, architectural and training innovations become necessary [207].

5 Power and Resource Consumption Comparison

5.1 Estimate of the Power Consumption of the Brain Compared to AI Models on Contemporary Compute Hardware

The human brain's average power draw is approximately

$$P_{\text{brain}} \approx 20 \text{ W},$$

which sustains on the order of 10^{14}–10^{15} synaptic events per second [208]. Naively assuming only the brain's neural network layers for information processing alone (research shows even neural dendritic connections themselves may form their own neural networks a layer underneath the cellular level, and not even taking into account layers implicated in Orch-Or theory). Hence the brain's raw processing throughput is conservatively

$$N_{\text{ops,brain}} \approx 2 \times 10^{14} \text{ ops/s},$$

giving an energy efficiency of

$$\frac{N_{\text{ops,brain}}}{P_{\text{brain}}} \approx \frac{2 \times 10^{14} \text{ ops/s}}{20 \text{ J/s}} = 1 \times 10^{13} \text{ ops/J}.$$

If one adopts the higher estimate of $N_{\text{ops,brain}} \approx 10^{18}$ ops/s, the corresponding efficiency becomes

$$\frac{10^{18} \text{ ops/s}}{20 \text{ J/s}} = 5 \times 10^{16} \text{ ops/J}.$$

By contrast, a state-of-the-art GPU (e.g. NVIDIA H100) achieves peak performance of order 10^{15} FLOP/s at $P_{\text{GPU}} \approx 1500$ W [213], yielding

$$\frac{N_{\text{FLOP,GPU}}}{P_{\text{GPU}}} \approx \frac{10^{15} \text{ FLOP/s}}{1500 \text{ J/s}} \approx 6.7 \times 10^{11} \text{ FLOP/J}$$

in ideal conditions. Real-world utilization typically reduces this by a factor of 3–10, so practical efficiency falls to $\sim 10^{11}$ FLOP/J. Thus, even at optimistic estimates without taking into account layers below the classical neural network layer in biological tissues, the brain's $\sim 10^{13}$–5×10^{16} ops/J exceeds GPU performance by 1–5 orders of magnitude.

Energy per Neural Event: Each action potential incurs an energy cost of approximately

$$\epsilon_{\text{spike}} \approx 100 \text{ pJ},$$

whereas each synaptic transmission costs on the order of

$$\epsilon_{\text{syn}} \approx 10 \text{ fJ}$$

. If the brain generates 2×10^{15} synaptic events per second, the total power is

$$P \approx 2 \times 10^{15} \times 10^{-14} \text{ J/s} = 20 \text{ W},$$

consistent with empirical measurements.

Information Efficiency: Neuronal recordings suggest a single spike can convey several bits of information (up to $\sim$ 6 bits/spike in some sensory systems). Taking a conservative 1 bit/spike and a firing rate of 10 Hz per neuron across 10^{11} neurons yields an information throughput of

$$I_{\text{brain}} \approx 10^{11} \times 10 \text{ bits/s} = 10^{12} \text{ bits/s}.$$

Thus, the brain's energy per bit is

$$\frac{P_{\text{brain}}}{I_{\text{brain}}} \approx \frac{20 \text{ J/s}}{10^{12} \text{ bits/s}} = 2 \times 10^{-11} \text{ J/bit}.$$

By comparison, digital systems incur $\sim 10^{-12}$–10^{-14} J/bit for raw logic operations, but when memory access and communication overhead are accounted for, practical energy per useful bit often exceeds 10^{-10} J/bit.

LLM Inference Efficiency: A single GPT-4 inference ($\approx$500 tokens) requires $\sim 10^{14}$ FLOP and consumes $\sim 10^3$ J, giving

$$\frac{10^{14} \text{ FLOP}}{10^3 \text{ J}} = 10^{11} \text{ FLOP/J},$$

consistent with GPU benchmarks above, but representing an energy cost of a few joules per token, compared to the brain's millijoule-level cost per sentence. These naive quantitative estimates based only on assumptions that the human brain operates by means of classical neural networks show that even with these assumptions the human brain operates at $\mathbf{10^2}$–$\mathbf{10^6}$ times greater energy efficiency than contemporary digital AI hardware, whether measured in ops per joule, joules per spike, or joules per bit of information [209] [210] [212]. Since emerging neuroscientific evidence indicates that much of the brain's computation happens below the neuron level, in dendritic and subcellular structures, which substantially increases its actual operation count without proportionally higher energy consumption, estimates can be refined further.

Neurons are not simple summing nodes—their dendritic trees perform complex nonlinear subunit computations. Each dendritic branch can act as an independent integrative unit (a "microcircuit" within the neuron) that nonlinearly processes its local synaptic inputs. Detailed biophysical models of cortical pyramidal neurons show that dozens of distal dendrites function like a hidden layer of sigmoid units: synapses drive local dendritic spikes or plateau potentials, and these branch outputs are then aggregated by the soma. In effect, a single neuron implements a two-layer neural network, with its dendritic branches as the first layer and the axon/soma as the second layer. This intraneuronal

architecture enables significantly more computation per neuron than assumed by a point-neuron model [211] [214] [215].

Notably, dendritic branches can execute logical operations on their inputs. The active properties of dendrites (voltage-gated channels, NMDA receptor nonlinearities, etc.) allow local coincidence detection and nonlinear integration. For example, a cluster of synchronous inputs on the same branch can induce an NMDA spike only if a threshold number co-occur—essentially an AND-gate behavior at the branch level. Branch outputs can saturate or veto each other such that certain combinations of inputs produce an output only when one branch is active and not another, implementing XOR-like logic. Individual dendritic branches have been shown to perform elementary computations including addition, multiplication, AND/OR logic, and even XOR. Empirical studies support this: human cortical pyramidal neurons can solve linearly non-separable tasks (like XOR) via dendritic calcium spikes, confirming that dendrites serve as nonlinear sub-computers inside each cell. In other words, a neuron is not just a summing device but rather a collection of logic subunits whose outputs are integrated. The energy cost of these dendritic computations appears to be moderate, meaning the neuron's operations-per-joule rises dramatically with their inclusion. A dendritic NMDA spike or Ca^{2+}-spike is a localized event requiring ion flux in a small branch section, far less costly than firing a whole additional neuron. While precise "energy per dendritic event" is hard to measure, it is constrained by the neuron's overall energy budget—which remains approximately 20 W for the entire brain. Thus, for a given neuron, performing multiple internal branch computations in parallel only marginally increases its total ATP consumption. If a pyramidal neuron has on the order of 50 active branch subunits, it could carry out $\sim 50\times$ more operations per processing cycle than a point-neuron assumption, yet its firing an action potential still dominates energy use. In effect, dendritic integration boosts computational throughput without a commensurate increase in power. This subneuronal parallelism helps explain the brain's efficiency: much of its computing is done in analog dendritic microcircuits that reuse the neuron's existing power supply [216] [217].

Each neuron contains an enormous number of tubulin molecules (10^7–10^9 per cell) arranged in a grid-like lattice. Tubulin subunits can switch between conformational or polar states, potentially functioning as quantum bits or classical bits at ultra-fast speeds. Recent experimental evidence suggests that isolated microtubules demonstrate resonance vibrations across a wide range of frequencies—from kilohertz up to terahertz—and oscillations in the megahertz–gigahertz range have been detected in vivo within neurons. These oscillations imply that microtubules support fast electrical or mechanical signaling modes inside the cell. In fact, gigahertz-frequency oscillations in dendritic and somatic microtubule networks have been shown to modulate the timing of axonal spikes in neuronal cultures, indicating a functional impact on neuronal output. This implies the cytoskeletal matrix is not just passive scaffolding but an active, high-speed computational layer interfacing with the neuron's membrane signals.

The spectral capacity of microtubules far exceeds that of neural firing. Neuronal spiking is limited to $\sim 10^2$–10^3 Hz, whereas microtubule lattice vibrations

and dipole oscillations can reach 10^6–10^{12} Hz (MHz to THz). Each tubulin dimer can oscillate or interact on timescales of nanoseconds to microseconds, performing many operations in the time a neuron takes to fire once. If each tubulin were to encode a bit flipping at, say, 10 MHz (a conservative frequency in the MHz–GHz spectrum), a single neuron with $\sim 10^8$ tubulins could, in principle, execute 10^{15}–10^{16} operations per second internally. Hameroff and colleagues estimate on this basis that the entire brain may perform on the order of 10^{27} operations per second when microtubule-level processes are included—eleven orders of magnitude higher than the 10^{16} ops/s figure from neurons-and-synapses alone. Even if this is an upper-bound speculation, it illustrates the vast computational potential of sub-neuronal mechanisms.

Crucially, these putative microtubule computations are energy-efficient at the level of single events. The energy scale of a tubulin conformational change or a quantum vibrational excitation is extremely small—on the order of 10^{-21} J for a terahertz-frequency quantum ($h \times 10^{12}$ Hz)—which is 10^{13} times less energy than a single 100 Hz neuronal action potential (10^{-8}–10^{-10} J) consumes. In theory, a cohort of tubulin bits flipping states in coherence could carry out billions of operations for the energy cost of one synaptic transmission. Orch-OR suggests that coherent dipole oscillations (so-called Fröhlich condensates) in microtubules might effectively orchestrate these molecular-scale operations using thermal energy already available in the cell. In simpler terms, the neuron's internal quantum vibrations could be leveraging ambient thermal energy to perform computation at a fantastically low energy cost per operation. If such processes contribute to cognition, the brain's operations-per-joule would outpace beyond all conventional estimates.

Considering dendritic and subcellular computations dramatically revises upward the brain's computational capacity and energy efficiency. Even a modest inclusion of dendritic subunit processing (e.g., treating each branch's nonlinear integration as an additional operation) boosts the estimated ops/s of the brain by an order of magnitude or more. For instance, if each of $\sim 10^4$ cortical neurons in a circuit performs $10\times$ more sub-operations than previously assumed, that is a $10\times$ gain in total operations for the same 20 W power which represents a full order-of-magnitude increase in ops per joule.

At the extreme end, if one embraces the Orch-OR microtubule model, the brain could be performing on the order of 10^{27} ops/s instead of 10^{16} ops/s [136] [16] [158] [149] [125]. This eleven-order gap would mean the brain achieves 10^{25}–10^{26} ops/J, dwarfing the efficiency of today's semiconductor-based AI systems. By comparison, modern supercomputers or AI chips might manage up to 10^{12}–10^{13} ops/J under ideal conditions which is many orders less efficient than even the conservative brain estimates. In qualitative terms, current AI models vastly underperform in both the complexity and the efficiency of biological intelligence and information processing. The brain's computing architecture is not a simple layered neural network; it is a deep, multiscale hierarchy: neurons, dendrites, synapses, glia, and cytoskeletal networks all contribute to information processing in parallel. Signals propagate not just as all-or-none spikes but also as analog dendritic voltages and possibly as high-frequency molecular vi-

brations. Thus, any direct power-efficiency comparison that treats the neuron like a simple logic gate is off by at least one or more orders of magnitude. Incorporating the extra layers of computation (nonlinear dendrites and potential microtubule quantum processing) leads to the revised view that the brain is even more power-efficient than previously thought.

5.2 Comparison with Classical-Quantum Hybrid AI Systems and Neuromorphic Based Computing Architectures

Classical–quantum hybrid AI architectures (such as the Quantum Approximate Optimization Algorithm, variational quantum circuits, quantum kernel methods, and quantum annealing systems) seek to leverage qubit superposition and entanglement to explore exponentially large solution spaces. However, these systems require kilowatts for cryogenic cooling and control electronics, whereas the human brain performs at exascale computational rates on approximately 20 W. Qubit counts remain in the tens to thousands, far below the brain's $\sim 10^{11}$ neurons, and practical quantum algorithms rely on a classical outer loop for parameter optimization, reintroducing von Neumann bottlenecks that biological systems avoid through fully integrated, self-organizing learning [218] [219] [220] [221] [222].

In contrast to fragile, millikelvin quantum coherence, the brain encodes information in robust, room-temperature spike patterns and analog membrane potentials. Neuronal spikes carry information in timing rather than amplitude, enabling sparse, event-driven codes that minimize energy consumption. Biological coherence arises in multi-frequency oscillations (e.g. gamma and theta rhythms) that bind information across billions of neurons, persisting for milliseconds, whereas quantum coherence in engineered devices decoheres in microseconds. Speculative theories such as Orch-OR propose that if quantum processes in microtubules contribute to consciousness, they would enable nonlocal, highly integrated computation—capabilities absent in current quantum hardware [223].

Neuromorphic computing architectures directly mimic key brain principles: sparse, event-driven spikes; local collocation of memory and compute; and massively parallel, asynchronous networks. Digital spiking chips (e.g. Intel Loihi), in-memory accelerators (IBM NorthPole), analog–digital hybrids (BrainScaleS), and energy-efficient neural processors (MIT Eyeriss) achieve orders-of-magnitude improvements in energy per operation compared to CPUs and GPUs. For instance, neuromorphic systems can consume picojoules per synaptic operation and eliminate costly data movement by embedding synaptic weights in local memory. Nevertheless, even the most advanced neuromorphic chips implement at most 10^6–10^7 neurons and 10^8–10^9 synapses, compared to the brain's $\sim 10^{11}$ neurons and $\sim 10^{14}$ synapses, and remain 10^2–10^4 times less energy-efficient when scaled toward brain-like workloads [224] [225] [226] [227].

Both quantum and neuromorphic platforms circumvent the traditional von Neumann architecture by using event-driven communication, in-memory com-

pute, and non-von Neumann topologies that blend processing and storage. Spiking architectures fire only on meaningful events, achieving asynchronous parallelism; in-memory accelerators co-locate compute with weights, cutting data-movement overhead; analog neuromorphic systems solve differential equations directly in hardware. Despite these innovations, fundamental limitations persist: quantum processors suffer from fragility, error-correction overhead, and limited qubit connectivity; neuromorphic chips face component variability, limited plasticity mechanisms, and connectivity constraints far below biological levels [228] [229] [230].

6 Estimates for Thresholds for the Technological Singularity Tipping Point

6.1 Thermodynamic Constraints and the Cost of Control

To rigorously estimate when an **Artificial Governance Superintelligence (AGSI)** at societal scale becomes unsustainable, we model institutional dynamics as a nonlinear system approaching a **fold catastrophe**. Let λ represent the degree of AI-driven control and structural complexity in governance, and let X denote a measure of institutional vitality or adaptability ("dynamism"). In a simple bifurcation model, the equilibrium condition can be written as $F(X, \lambda) = 0$, where increasing λ gradually erodes the system's ability to sustain a high-X solution. At low control levels, there exist stable high-X equilibria (robust, adaptive institutions). As λ grows, the high-X solution and a middle unstable solution coalesce and annihilate at a critical $\lambda = \lambda_c$, leaving only a low-X equilibrium (collapsed adaptability). This *fold catastrophe* behavior yields a sudden drop to a fragile state.

The model predicts *ultrasensitivity* near λ_c: even a small perturbation or control overshoot can tip the system from the high-performance branch into collapse. This formalism aligns with Tainter's thesis on societal collapse, where increasing investments in complexity eventually yield diminishing and then negative returns, causing a nonlinear collapse in functionality. Ashby's Law of Requisite Variety further underpins this tipping point: as an AGSI centralizes control, it reduces systemic variety. As human agents rely more on central AI systems to substitute intrapersonal trust with information stored in trophic networks, the controller's variety is less than environmental complexity, the system can no longer respond to perturbations, and resilience collapses

From thermodynamic first principles, we assess the energy required for entropy suppression in social systems governed by AGSI. Per Landauer's bound, erasing one bit of information has a minimum energy cost of

$$E_{\min} = k_B T \ln 2,$$

where k_B is Boltzmann's constant and T is temperature. When applied to society, suppressing social entropy (i.e., reducing behavioral variance) incurs growing energy costs as AGSI intensity increases. Near-total suppression becomes

thermodynamically unsustainable. The AGSI thus hits a threshold where each added increment of control yields vanishing or negative returns due to energy dissipation and waste heat saturation.

This is formalized through **spectral entropy collapse**. Institutions modeled as operators on a Hilbert space develop degenerate spectra under centralized control. Let $D(\alpha)$ be the institutional Dirac-like operator under control intensity α, with normalized eigenvalue distribution $\{p_i\}$. Then von Neumann entropy

$$H(\alpha) = -\sum_i p_i \ln p_i$$

decreases monotonically with α, indicating systemic contraction into a low-entropy, brittle state. The system becomes hypersensitive to disturbances when $H \to 0$ and the spectral gap $\delta = \lambda_1 - \lambda_2$ vanishes.

Thermodynamic Limits and Diminishing Marginal Returns: Deep-learning performance P typically scales with compute C as

$$P \sim C^{-\alpha}, \quad 0 < \alpha < 1,$$

so that doubling C yields less than proportional gains in P. At $\sim 10^3$ TWh/year, marginal utility per joule of compute approaches zero.

By Landauer's principle, erasing one bit of information costs at least

$$E_{\min} = k_{\mathrm{B}} T \ln 2 \approx 3 \times 10^{-21} \text{ J/bit}.$$

An AGSI controlling billions of uncertain variables must erase immense entropy. As control precision grows, energy per bit removal grows super-linearly due to sensing, communication, and actuation overhead. Beyond $\sim 10^3$ TWh/year, the AGSI expends nearly all available free energy simply to maintain stability, making further entropy removal energetically unsustainable [205] [206] [104] [105].

Spectral Collapse and Institutional Fragility: Model institutions as operators on a high-dimensional social state space with eigenvalue spectrum $\{\lambda_i\}$. Increasing AGSI centralization drives these operators toward near-commutativity, collapsing secondary eigenvalues toward λ_1. Define the spectral spread

$$\Delta = \max_{i \neq j} |\lambda_i - \lambda_j|.$$

As control intensity λ approaches a critical λ_c, $\Delta \to 0$, the effective rank falls to one, and the system loses redundancy. When $\Delta \approx 0$, any perturbation in a suppressed mode cannot be absorbed, precipitating system-wide failure [27] [26] [189].

Global Electricity Demand for AGSI-Scale AI Compute: Current global electricity generation is on the order of 3×10^4 TWh per year and is rising

at roughly 4% annual growth projected through 2027. Within this, data centers (including AI and cryptocurrency workloads) consumed an estimated 460 TWh in 2022—about 1.6–1.8% of world electricity use. That figure is projected to more than double to over 1000 TWh by 2026, approaching 3–4% of global generation. Rapid growth is driven by AI compute needs: state-of-the-art models saw compute demand double every ~ 3.4 months between 2012 and 2018, far outpacing Moore's Law. If these trends continue, "AGSI-scale" infrastructure could draw on the order of 10^2–10^3 TWh per year—several percent of total supply [231] [232] [233] [234] [235].

Analyses of deep-learning scaling suggest a fully deployed Artificial General Superintelligence might require 10^2 to 10^3 TWh/year in electricity. A tipping point around $\sim 10^3$ TWh/year corresponds to roughly 3–4% of global generation. Even hundreds of TWh/year would rival the annual power usage of midsize countries (e.g., 1000 TWh $\approx$ Japan's yearly consumption). At the upper end, $\sim 10^3$ TWh/year matches the entire global data-center sector. An AGSI in this range would consume on the order of a petawatt-hour annually—an unprecedented concentration of energy by a single sector—and could command roughly 5% of global electricity [236].

Power and Infrastructure Costs at the Tipping Point: Translating 10^3 TWh/year into economic cost requires both electricity prices and amortized data-center expenses. Hyperscale facilities typically incur an "all-in" rate of \$0.05–\$0.10 per kWh (including cooling, redundancy, and capital amortization), or \$50–100 million per TWh. At $\sim 10^3$ TWh/year, the annual energy and infrastructure bill would be on the order of \$50–100 billion.

A bottom-up cross-check: training GPT-3 (175 billion parameters) consumed approximately 1.3 GWh of electricity—about \$100,000 at \$0.08 per kWh—for a single run. GPT-4–scale training likely consumed tens of thousands of megawatt-hours. Moreover, inference can dominate a model's lifetime energy: serving billions of queries may account for roughly 60% of total energy use. As a thought experiment, ten thousand top-end GPUs (e.g. Nvidia H100 at $\approx$ 1 kW each) draw $\approx$ 10 MW; over one year (8760 h) this is only $\approx$ 0.09 TWh. Scaling to a sustained 1 GW load (about 1 million GPUs) yields $\approx$ 8.8 TWh/yr, while 10 GW (about 10 million GPUs) yields $\approx$ 88 TWh/yr. To reach $\sim$ 1000 TWh/yr would require on the order of 11 GW continuous power (about 11 million GPUs)—equivalent to dozens of hyperscale data centers and hundreds of billions of dollars in capital and infrastructure costs [237].

6.2 Energy Return on Investment and Complexity Burden

The sustainability of AGSI hinges on energetic economics. Following Charles Hall, we consider the **energy return on investment (EROI)**:

$$\text{EROI} = \frac{\text{usable energy delivered}}{\text{energy expended}}.$$

Advanced civilization demands EROI $\gtrsim 14:1$ for high-complexity activities. AGSI infrastructure draws heavily on electricity for datacenters, sensor grids, inference pipelines, and automated regulation. When AGSI power requirements approach 1000 TWh/year—about 3−5% of global electricity—the marginal EROI for AI governance may fall below 1:1, indicating that every joule expended on AGSI returns one joule or less of societal utility.

We identify this tipping energy as

$$E_{\text{AGSI}} \approx 10^3 \text{ TWh/year},$$

which coincides with estimates of AGI-scale compute demand and projected growth trajectories. At this point, continued AI expansion may crowd out other sectors, creating opportunity costs and systemic fragility. The economic counterpart is a budgetary burden of

$$C_{\text{AGSI}} \sim (5\text{–}15) \times 10^{10} \text{ \$/year},$$

at average datacenter costs of \$0.05–\$0.10/kWh, representing $\sim 0.1\%$–0.3% of global GDP. While numerically small, this spending represents disproportionate centralization of infrastructural and policy control—comparable to global R&D budgets or international institutions.

Economic Significance Relative to Global GDP: Global GDP is currently \$100 trillion per year. \$100 billion in compute costs represents 0.1% of GDP. If AI demands rose to 1500–2000 TWh/year, costs of \$200–300 billion would equal 0.2–0.3% of GDP (or 0.5% of advanced-economy output). For perspective, 0.5% of global GDP is comparable to a mid-sized country's economy, one-tenth of global R&D spending, or half of world foreign aid. At an average \$0.05/kWh, total global electricity revenues (29 000 TWh) cost \$1.45 trillion/year (1.4% of GDP). Thus, AGSI consuming 1000+ TWh would command 7% of electricity revenues and 0.5% of economic output—non-trivial macroeconomic shares that could crowd out other investments if AI benefits do not justify the expense [238] [232] [234] [237] [239] [240].

6.3 Adaptive Capacity and Ingenuity Gap

Thomas Homer-Dixon's *ingenuity gap* characterizes collapse via adaptive failure: when the rate of problem generation exceeds problem-solving capacity, systemic breakdown occurs. We define an **adaptive capacity index** as

$$I = \frac{\text{rate of problem emergence}}{\text{rate of problem solution}}.$$

When $I > 1$, problems accumulate; when $I \gg 1$, cascading failures occur. AGSI regimes may initially lower I via optimization, but centralization and spectral entropy collapse can suppress lateral innovation, causing I to rise again as system complexity outpaces institutional flexibility. The critical tipping point $I_c = 1$ marks loss of problem-solving sufficiency.

Furthermore, per Ashby's Law, control variety V_{ctrl} must satisfy

$$V_{\text{ctrl}} \geq V_{\text{env}}.$$

We define the *information bottleneck ratio* $\gamma = V_{\text{ctrl}}/V_{\text{env}}$, where $\gamma < 1$ indicates adaptation failure. Spectral flattening under AGSI-driven commutativity implies $\gamma \to 0$, reducing the system's effective rank.

6.4 Synthesis of Threshold Estimate Predictions

We summarize tipping conditions across theoretical domains:

- **Structural complexity:** $\lambda_c \approx 1$—fold catastrophe where high-performance equilibria vanish.
- **Energetic threshold:** $E_{\text{AGSI}} \approx 1000$ TWh/year—entropy suppression cost saturates and EROI collapses.
- **Economic burden:** $C_{\text{AGSI}} \sim 0.1\%$–$0.3\%$ of global GDP—displaces productive infrastructure.
- **Control fragility:** $\gamma \to 0$—variety mismatch disables resilience.
- **Ingenuity gap:** $I \to \infty$—adaptive deficit as central control suppresses lateral solution networks.

Combining these lines of evidence yields:

$$E_{\text{AGSI}} \approx 10^3 \text{ TWh/yr}, \quad C_{\text{AGSI}} \sim 5{\times}10^{10}\text{–}1.5{\times}10^{11}\ \$, \quad \beta \sim 0.05\%\text{–}0.15\% \text{ of GDP}.$$

At this threshold:

- *Marginal returns* on additional compute vanish.
- *Thermodynamic costs* of entropy suppression saturate.
- *Spectral entropy* collapses, inducing brittleness.

Based on energy growth trends, projected AI compute demands, and observed diminishing returns in deep learning scaling, we estimate that AGSI systems will approach this tipping point between **2025 and 2030**. This aligns with data center electricity forecasts (crossing 1000 TWh/year by 2026), peak centralized fragility in Strauss–Howe generational theory, and systemic institutional overload observed by Ray Dalio in debt and governance cycles. The singularity, in this view, is not explosive but *entropic*: a "gentle" collapse of functionality as energy and institutional structures saturate and bifurcate to a low-adaptability regime.

Crossing $E_{\text{AGSI}} \approx 10^3$ TWh/yr thus marks the "gentle technological singularity" tipping point, beyond which further scaling of current deep-learning architectures reaches an inflection point unless radically more efficient paradigms

emerge. At current trajectories, data-center energy use was about 460 TWh in 2022 and is projected to exceed 1000 TWh by 2026 [234]. Moreover, AI compute demand doubled every ~ 3.4 months between 2012 and 2018 [235]. Taken together, these trends imply we will cross the $\sim 10^3$ TWh/year tipping point by around 2026.

Avoiding this outcome requires architectural and policy foresight: reducing λ via decentralization, increasing V_{ctrl} by federated AI, and ensuring EROI remains well above threshold. Otherwise, the system will enter a regime where control becomes entropy, governance becomes fragility, and the future must pass through collapse before renewal.

7 Future Research Directions and Discussion

Our findings underscore that contemporary large-scale AI systems, such as transformer-based language models, are fundamentally constrained by inefficiencies that distinguish them from human intelligence and information processing. These models achieve impressive results through brute-force processing of massive datasets and billions of parameters, yet remain orders of magnitude less energy-efficient than the human brain. In effect, they are massively inefficient pattern learners rather than truly cognitive agents. This inefficiency not only incurs extraordinary computational cost (e.g. data centers drawing megawatts of power for tasks a human brain does on 20 watts), but also creates complexity bottlenecks on the path toward Artificial General Superintelligence (AGSI). Because they lack conscious understanding or the adaptive, self-organizing dynamics of a brain, current AI models must scale in complexity just to approximate tasks that biological cognition handles fluidly.

Our analysis suggests that without a paradigm shift, simply scaling up current architectures yields diminishing returns in value and may even precipitate performance collapses when confronted with real-world complexity. Beyond a certain catastrophe point, adding more layers, parameters, or data to these unconscious models does not translate to proportional gains in general intelligence; instead, it can lead to instability and unreliability, hinting at fundamental limits to what brute-force AI scaling can achieve in pursuit of AGSI. Indeed, we identify several intrinsic limits that today's transformer-based architectures encounter as they grow in scale. Firstly, thermodynamic constraints present a hard ceiling: the energy cost of computation and information control rises steeply with model size and usage. As an AGSI attempts to manage ever more information (for example, in governing a complex society), it must continually erase uncertainty and enforce order, an operation that by thermodynamic principle exacts an immense energy toll.

Our study argues that pursuing near-total predictive control with current AI towards addressing the alignment problem incurs exponentially growing energy requirements and heat dissipation issues, pushing up against physical limits (e.g. the Landauer bound for bit erasure). In practical terms, this means that the strategy of "more computation at any cost" will eventually hit an

energetic wall, where the power needed to support further intelligence gains and alignment becomes prohibitively expensive or unfeasible. Secondly, spectral limits emerge in highly-optimized AI control systems. As we intensify a model's training or integrate it deeply into decision-making processes, we observe a collapse in the diversity of effective responses – analogous to a contraction in the system's eigenvalue spectrum. This spectral collapse implies that an AI governing system, in its drive to optimize and homogenize outcomes, can lose the very flexibility and richness of behavior that robust intelligence requires. Past a critical threshold of complexity, the AI's performance may degrade sharply: instead of generalizing more effectively, it becomes brittle, prone to unexpected failures when faced with novel or chaotic inputs.

Lastly, institutional complexity limits confront the deployment of these AI systems at a societal scale to facilitate value creation and growth. Social institutions and organizations have finite capacity to absorb complexity; embedding a massively complex, opaque AI into governance or industry can introduce new failure modes and bureaucratic friction. Our analysis resonates with historical patterns of societal collapse in overly complex systems: as more resources are channeled into maintaining an increasingly convoluted AI-driven apparatus, the net benefits diminish. Eventually, the costs – in energy, oversight, and loss of human initiative – can outweigh the gains. The trajectory of late-stage AI-heavy societies could thus reach a tipping point where further complexification by AGSI turns from stabilizing to destabilizing. This scenario, which we term a "gentle technological singularity," is characterized not by runaway superintelligence or catastrophe in the classical sense, but by a gradual stagnation and an abrupt collapse of societal dynamism that we estimate to occur around 2026. In this regime, ever-larger AI control loops intended to secure order instead induce rigidity and fragility, leading to a breakdown in the very institutional stability they were meant to preserve.

Given these multifaceted limitations, charting a new course for AI development becomes imperative. A key insight of this work is that overcoming the efficiency and complexity bottlenecks will likely require architectural innovation inspired by consciousness and biology rather than further brute-force scaling. One promising research direction is to investigate computing paradigms drawn from the operations of the human brain and the nature of consciousness itself. For example, the brain's neurons leverage rich dynamics – oscillatory rhythms, adaptive feedback, and perhaps even quantum effects – to achieve remarkable cognitive feats at minimal energy cost. In particular, we highlight the intriguing possibility of quantum-coherent processes in neural microtubules as a frontier for exploration. Validating experimentally whether coherent quantum states can persist in microtubule structures (as hypothesized in certain quantum consciousness theories) would be a groundbreaking step. If such quantum-biological processes contribute to the brain's information processing efficiency, they could inspire a new class of AI architectures that operate on fundamentally different principles than today's digital electronics.

A computing architecture that harnesses room-temperature quantum coherence or other non-classical phenomena – as living neurons might – would

potentially perform orders of magnitude more computations per joule of energy, and handle complexity with a fluid, self-organizing adaptability more akin to a mind than a machine, with potential savings of billions of dollars in coming years. The overarching goal of these directions is to design AI that is not just more powerful, but qualitatively smarter and more efficient – systems that learn and adapt through understanding rather than through brute force. Research in this vein, bridging neuroscience, quantum physics, and computer engineering, is still in its infancy, but it offers a path to transcend the current plateau.

By pursuing such avenues, we move closer to AI that could eventually exhibit forms of sentience or at least a human-like agility in reasoning, without being trapped by the complexity and energy sinkholes that plague current models. Beyond technical research, our study carries important implications for policy and governance in the AI era. It is clear that the status quo approach of unconstrained scaling of AI comes with unsustainable externalities and strategic risks. We therefore propose several forward-looking policy measures to realign the development of AI with long-term societal well-being and resilience:

- **Redirect Funding to Sustainable AI:** Public and private sectors should reallocate funding and incentives away from merely scaling up existing large models, and toward efficient neuromorphic and quantum-biological computing research or towards direct investments in infrastructure to benefit human agents by 2026. By investing in technologies that mimic the brain's frugal use of energy (such as spiking neural chips , organoid intelligence, and other neuromorphic hardware) or that explore bio-inspired quantum information processing, society can accelerate the emergence of AI systems that achieve more with far less, with a potential savings of hundreds of billions of dollars. This strategic funding shift would not only address the current inefficiency crisis but also spur innovation in architectures that could eventually incorporate aspects of consciousness or organic intelligence.

- **Regulate Energy Usage and Externalities:** Governments and international bodies should develop regulations or guidelines to monitor and limit the energy consumption and environmental impact of massive AI models. Transparency requirements can be instituted for AI developers to report the compute and power used in training and deploying large models. Policymakers might set benchmark efficiency standards (e.g., performance per kilowatt-hour) that models must meet, or even impose caps/carbon taxes on exorbitant computational expenditures. The aim is to internalize the social cost of hyper-scaled AI: if running a model drains as much electricity as a small city, that cost should be accounted for. By curbing wasteful practices and rewarding energy-efficient AI designs, such policies would mitigate the carbon footprint and infrastructure strain of AI, while pressuring the field toward more sustainable innovation rather than brute-force approaches.

- **Institute Safeguards for Complexity Collapse:** In parallel, we must

prepare institutions for entropy-induced collapse scenarios that could arise from over-reliance on complex AI to deal with labor shortages and displacements. This involves creating robust contingency plans and governance safeguards in case AI systems fail or behave unpredictably under stress. Critical infrastructure and social services that incorporate AI should be stress-tested for worst-case scenarios (e.g. a sudden AI malfunction or a cascading error in an AGSI managing key systems). Institutions should maintain human oversight with an awareness of key indicators and fallback procedures: rather than ceding full control to opaque algorithms, skilled personnel should be ready to intervene or assume manual control if automated systems begin to destabilize. Additionally, diversity in control mechanisms is crucial - instead, federated or decentralized AI architectures, and maintaining some low-complexity, analog solutions as backups, can ensure that a failure in one component does not cripple the entire societal system. By enacting such safeguards, policymakers can increase societal resilience against the very real risk that an overly complex, energy-hungry AGSI could lead to an abrupt loss of institutional functionality.

In summary, confronting the inefficiency and complexity limits of current AI is not only a technical necessity but a civilizational imperative. Our exploration of thermodynamic, spectral, and institutional bottlenecks reveals that without significant change in course, the drive toward an all-encompassing AGSI could lead to institutional stagnation or collapse, and renewal. Fortunately, by pursuing future-facing research into brain-inspired and consciousness-informed AI architectures, or alternatively to divert funding towards direct infrastructure investments to benefit human agents, and by implementing thoughtful policy interventions, the trajectory may be redirected. In doing so, we move toward a form of artificial intelligence that enhances society's long-term dynamism and resilience, rather than exacerbating complexity to a catastrophe point. Embracing these recommendations will help ensure that progress toward genuine general intelligence remains sustainable and beneficial, avoiding the peril of complexity-induced collapse and guiding us toward an era of AI that is powerful, prosperous, and profoundly safe.

8 Conclusion

In this work, we have shown that current large-scale AI systems—embodied by transformer-based language models and massive compute clusters—are fundamentally limited by their lack of consciousness-inspired information processing and by prohibitive energy and complexity bottlenecks. Through a multidisciplinary analysis, we demonstrated that:

- **Thermodynamic limits** impose steep energy costs on ever-larger models.
- **Spectral collapse** occurs under extreme centralization, as institutional

and algorithmic eigenmodes coalesce, reducing adaptive diversity and increasing brittleness.

- **Institutional complexity** parallels historical collapse phenomena: unchecked growth in AI-driven control loops risks eroding societal dynamism and precipitating systemic breakdown.

We contrasted these constraints with the remarkable efficiency of the human brain—operating at tens of watts yet performing orders of magnitude more computations per joule—highlighting the promise of consciousness-inspired paradigms such as quantum-coherent microtubule models and neuromorphic architectures. Our fold catastrophe and spectral-theoretic models formalize how a *gentle technological singularity* could emerge as soon as 2026, not as an popularly envisioned, but as a sudden loss of institutional stability when energetic and informational thresholds are crossed.

To avert this outcome, we advocate a twofold strategy. First, research must pivot toward *efficient, brain-inspired computing*: rigorous experiments probing quantum coherence in neuronal microstructures, development of neuromorphic and quantum-biological hardware, and exploration of architecture–data–compute tradeoffs grounded in spectral and thermodynamic principles. Second, policy interventions are essential: redirecting funding toward sustainable AI paradigms, redirecting investments to directly benefit human agents (especially with infrastructure investments), regulating the development of large models, and instituting safeguards against complexity-induced collapse by 2026.

References

[1] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. https://doi.org/10.1038/nature14539

[2] Brenden M. Lake, Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40:e253, 2017. https://doi.org/10.1017/S0140525X16001837

[3] Emily M. Bender and Alexander Koller. Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, 2020. https://doi.org/10.18653/v1/2020.acl-main.463

[4] Noam Chomsky and Yarden Katz. Noam Chomsky on Where Artificial Intelligence Went Wrong. *Chomsky.info*, November 1, 2012. https://chomsky.info/20121101/

[5] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, pages 610–623, 2021. https://doi.org/10.1145/3442188.3445922

[6] David J. Chalmers. The Conscious Mind: In Search of a Fundamental Theory. Oxford University Press, 1996.

[7] Adam H. Marblestone, Greg Wayne, and Konrad P. Kording. Toward an integration of deep learning and neuroscience. *Frontiers in Computational Neuroscience*, 10:94, 2016. https://doi.org/10.3389/fncom.2016.00094

[8] James C. R. Whittington and Rafal Bogacz. Theories of error backpropagation in the brain. *Trends in Cognitive Sciences*, 23(3):235–250, 2019. https://doi.org/10.1016/j.tics.2019.01.006

[9] Gilles Dumas, Jacques Nadel, Rémi Soussignan, Hugues Martinerie, and Arnaud Garnero. Inter-brain synchronization during social interaction. *PloS ONE*, 5(8):e12166, 2010. https://doi.org/10.1371/journal.pone.0012166

[10] Karl S. Lashley. In search of the engram. *Symposia of the Society for Experimental Biology*, 4:454–482, 1950.

[11] David Attwell and Simon B. Laughlin. An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, 21(10):1133–1145, 2001. https://doi.org/10.1097/00004647-200108000-00001

[12] Fabio A. Azevedo, L. R. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. Ferretti, R. E. Leite, W. Jacob Filho, R. Lent, and S. Herculano-Houzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541, 2009. https://doi.org/10.1002/cne.21974

[13] György Buzsáki and Andreas Draguhn. Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929, 2004. https://doi.org/10.1126/science.1099745

[14] David Patterson, Joseph Gonzalez, Quoc V. Le, Chen Liang, Lluís-Miquel Munguía, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350*, 2021. https://doi.org/10.48550/arXiv.2104.10350

[15] Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. Recalibrating global data center energy-use estimates: Growth in energy use has slowed owing to efficiency gains that smart policies can help maintain in the near term. *Science*, 367(6481):984–986, 2020. https://doi.org/10.1126/science.aba3758

[16] Anirban Bandyopadhyay, Daniela Sahu, et al. Terahertz vibrations in brain microtubules. *Journal of Biophysical Chemistry*, 4:123–134, 2013. https://doi.org/10.1016/j.bpc.2013.04.005

[17] Wolf Singer. Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24(1):49–65, 1999. https://doi.org/10.1016/S0896-6273(00)80821-1

[18] Herbert A. Simon. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6):467–482, 1962.

[19] Friedrich A. Hayek. The use of knowledge in society. *American Economic Review*, 35(4):519–530, 1945.

[20] Robin I. M. Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6):469–493, 1992. https://doi.org/10.1016/0047-2484(92)90081-J

[21] Robin I. M. Dunbar. Coevolution of neocortex size, group size and language in humans. *Behavioral and Brain Sciences*, 16(4):681–694, 1993. https://doi.org/10.1017/S0140525X00032369

[22] Niklas Luhmann. Social Systems. Stanford University Press, 1995.

[23] Niklas Luhmann. Theory of Society, Volume 1. Stanford University Press, 2012.

[24] W. P. Hall. Biological nature of knowledge in the learning organisation. *The Learning Organization*, 12(2):169–188, 2005. https://doi.org/10.1108/09696470510583548

[25] Melanie Mitchell. Complexity: A Guided Tour. Oxford University Press, 2009.

[26] Alain Connes. Noncommutative Geometry. Academic Press, 1994.

[27] Fan R. K. Chung. Spectral Graph Theory. American Mathematical Society, 1997.

[28] W. Brian Arthur. Complexity and the Economy. *Science*, 284(5411):107–109, 1999. https://doi.org/10.1126/science.284.5411.107

[29] John H. Holland. Hidden Order: How Adaptation Builds Complexity. Addison-Wesley, 1995.

[30] Joshua M. Epstein and Robert L. Axtell. Growing Artificial Societies: Social Science from the Bottom Up. MIT Press, 1996.

[31] Pierre Bourdieu. The Forms of Capital. In J. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education*, pages 241–258. Greenwood, 1986.

[32] Jonathan Nitzan and Shimshon Bichler. Capital as Power: A Study of Order. Routledge, 2009.

[33] George A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3):488–500, 1970. https://doi.org/10.2307/1879431

[34] Joseph E. Stiglitz and Andrew Weiss. Credit rationing in markets with imperfect information. *American Economic Review*, 71(3):393–410, 1981. https://doi.org/10.1257/aer.71.3.393

[35] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, 2019.

[36] Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown, 2016.

[37] Tom B. Brown, Benjamin Mann, Nick Ryder, et al. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020. https://doi.org/10.48550/arXiv.2005.14165

[38] Solon Barocas and Andrew D. Selbst. Big data's disparate impact. *California Law Review*, 104(3):671–732, 2016. https://doi.org/10.15779/Z38BG31

[39] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine Bias. *ProPublica*, May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[40] Virginia Eubanks. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018.

[41] Jean Baudrillard. Simulacra and Simulation. University of Michigan Press, 1994.

[42] James C. Scott. Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed. Yale University Press, 1998.

[43] Andrew Guthrie Ferguson. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York University Press, 2017.

[44] Walter L. Perry, Marcus D. McInnis, Carter C. Price, Susan Smith, and John B. Hollywood. Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations. RAND Corporation, 2013.

[45] Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009. https://doi.org/10.1038/nature07634

[46] Richard H. Thaler and Cass R. Sunstein. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press, 2008.

[47] Jens Rasmussen. Skills, Rules, Knowledge: Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3):257–266, 1983. https://doi.org/10.1109/TSMC.1983.6313077

[48] Paul N. Edwards. A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming. MIT Press, 2010.

[49] Rogier Creemers. China's Social Credit System: An Evolving Practice of Control. SSRN Working Paper, 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3175792

[50] Jean-François Lyotard. *The Postmodern Condition: A Report on Knowledge*. Manchester University Press, 1984 (original French edition 1979).

[51] Fredric Jameson. *Postmodernism, or, The Cultural Logic of Late Capitalism*. Duke University Press, 1991.

[52] B. J. Fogg. *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann, 2002.

[53] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: An explanation of 1/f noise. *Physical Review Letters*, 59(4):381–384, 1987. https://doi.org/10.1103/PhysRevLett.59.381

[54] Marten Scheffer, Johan Bascompte, William A. Brock, Victor Brovkin, Stephen R. Carpenter, Vasilis Dakos, Volker Held, Egbert H. van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461(7260):53–59, 2009. https://doi.org/10.1038/nature08227

[55] Duncan J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002. https://doi.org/10.1073/pnas.082090499

[56] Didier Sornette. *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton University Press, 2003.

[57] Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978. https://doi.org/10.1086/226707

[58] Andrea S. Kirilenko, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun. The Flash Crash: High-Frequency Trading in an Electronic Market. *Journal of Finance*, 72(3):967–998, 2017. https://doi.org/10.1111/jofi.12567

[59] Tarleton Gillespie. Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. Yale University Press, 2018.

[60] Patrick Meier. Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response. CRC Press, 2015.

[61] Dirk Helbing. Globally networked risks and how to respond. *Nature*, 497(7447):51–59, 2013. https://doi.org/10.1038/nature12047

[62] John D. Sterman. Business Dynamics: Systems Thinking and Modeling for a Complex World. Irwin/McGraw-Hill, 2000.

[63] Robert D. Putnam. Bowling alone: America's declining social capital. *Journal of Democracy*, 6(1):65–78, 1995.

[64] Ann M. Florini. *The Right to Know: Transparency for an Open World*. Columbia University Press, 2007.

[65] Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

[66] Quentin Skinner. *The Foundations of Modern Political Thought, Volume 1: The Renaissance*. Cambridge University Press, 1998.

[67] Benedict Anderson. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso, 1983.

[68] Francis Fukuyama. *The End of History and the Last Man*. Free Press, 1992.

[69] Russell Hardin. *Trust and Trustworthiness*. Russell Sage Foundation, 2002.

[70] Eliezer Yudkowsky. Artificial Intelligence as a Positive and Negative Factor in Global Risk. In Nick Bostrom and Milan M. Ćirković (Eds.), *Global Catastrophic Risks*, pages 308–345. Oxford University Press, 2008.

[71] Stuart Russell. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking, 2019.

[72] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, 2009. `https://doi.org/10.1103/RevModPhys.81.591`

[73] Jerome R. Busemeyer and Peter D. Bruza. Quantum models of cognition and decision. Cambridge University Press, 2012.

[74] Andrew G. Haldane and Robert M. May. Systemic risk in banking ecosystems. *Nature*, 469(7330):351–355, 2011. `https://doi.org/10.1038/nature09659`

[75] Joseph A. Tainter. *The Collapse of Complex Societies*. Cambridge University Press, 1988.

[76] Charles D. Brummitt, George Barnett, and Raissa M. D'Souza. Coupled catastrophes: sudden shifts cascade and hop among interdependent systems. *Journal of the Royal Society Interface*, 12(112):20150712, 2015. doi:10.1098/rsif.2015.0712.

[77] Eduardo F. Camacho and Carlos Bordons. *Model Predictive Control*. Springer-Verlag, 2004.

[78] Manfred Morari and Jay H. Lee. Model predictive control: Past, present and future. *Computers & Chemical Engineering*, 23(4–5):667–682, 1999. `https://doi.org/10.1016/S0098-1354(99)00237-2`

[79] Peyman Shojaee, Nolan Jia, Samy Bengio, et al. The Illusion of Thinking. Apple ML Research Technical Report, 2025.

[80] Vernor Vinge. The Coming Technological Singularity: How to Survive in the Post-Human Era. *Whole Earth Review*, 77:88–95, 1993.

[81] Ray Kurzweil. *The Singularity Is Near: When Humans Transcend Biology*. Viking, 2005.

[82] Sam Altman. The Gentle Singularity. Sam Altman's blog, June 10, 2025. `https://blog.samaltman.com/the-gentle-singularity`

[83] René Thom. Structural Stability and Morphogenesis. Cambridge University Press, 1975.

[84] Ian Poston and Ian Stewart. Catastrophe Theory and Its Applications. Pitman, 1978.

[85] Igor Mezić. Spectral properties of dynamical systems, model reduction and decomposition. *Nonlinear Dynamics*, 41(1–3):309–325, 2005. https://doi.org/10.1007/s11071-005-2824-x

[86] B. O. Koopman. Hamiltonian systems and transformations in Hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.

[87] W. Ross Ashby. *An Introduction to Cybernetics*. Chapman & Hall, 1956.

[88] Nassim Nicholas Taleb. *Antifragile: Things That Gain from Disorder*. Random House, 2012.

[89] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[90] Ola Bratteli and Derek W. Robinson. *Operator Algebras and Quantum Statistical Mechanics I*. Springer-Verlag, 1987.

[91] Steven H. Strogatz. Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. Westview Press, 1994.

[92] Michael J. Mazarr, Tim Sweijs, and Daniel Tapia. The Sources of Renewed National Dynamism. RAND Corporation Research Report, RR-A2611-3, 2024. https://doi.org/10.7249/RRA2611-3

[93] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Gradient Flows in Metric Spaces and in the Space of Probability Measures. Birkhäuser, 2nd edition, 2008.

[94] Cédric Villani. Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften, vol. 338. Springer, 2009.

[95] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Gradient Flows in Metric Spaces and in the Space of Probability Measures. Birkhäuser, 2nd edition, 2008.

[96] Cédric Villani. Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften, vol. 338. Springer, 2009.

[97] Nicolas Lanzetti, Joudi Hajar, and Florian Dörfler. Modeling of Political Systems Using Wasserstein Gradient Flows. arXiv preprint arXiv:2209.05382, 2022. https://arxiv.org/abs/2209.05382

[98] La Mi, Jorge Gonçalves, and Johan Markdahl. Asymptotically Stable Polarization of Multi-Agent Gradient Flows Over Manifolds. arXiv preprint arXiv:2301.04877, 2023. https://arxiv.org/abs/2301.04877

[99] Yuri A. Kuznetsov. Elements of Applied Bifurcation Theory. 2nd edition. Springer, 1998.

[100] John Guckenheimer and Philip Holmes. Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. Springer, 1983.

[101] Ugo Bardi, Sara Falsini, and Ilaria Perissi. Toward a General Theory of Societal Collapse: A Biophysical Examination of Tainter's Model of the Diminishing Returns of Complexity. arXiv preprint arXiv:1810.07056, 2018. https://arxiv.org/abs/1810.07056

[102] João da Gama Batista, Jean-Philippe Bouchaud, and Damien Challet. Sudden Trust Collapse in Networked Societies. arXiv preprint arXiv:1409.8321, 2014. https://arxiv.org/abs/1409.8321

[103] Charles D. Brummitt, George Barnett, and Raissa M. D'Souza. Coupled catastrophes: sudden shifts cascade and hop among interdependent systems. arXiv preprint arXiv:1410.4175, 2014. https://arxiv.org/abs/1410.4175

[104] Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3):183–191, 1961.

[105] C. H. Bennett. The thermodynamics of computation—A review. *International Journal of Theoretical Physics*, 21(12):905–940, 1982.

[106] Erik Verlinde. On the Origin of Gravity and the Laws of Newton. *Journal of High Energy Physics*, 04:029, 2011. https://doi.org/10.1007/JHEP04(2011)029

[107] Roger Penrose. *Cycles of Time: An Extraordinary New View of the Universe*. Bodley Head, 2010.

[108] Shinsei Ryu and Tadashi Takayanagi. Holographic Derivation of Entanglement Entropy from the Anti–de Sitter Space/Conformal Field Theory Correspondence. *Physical Review Letters*, 96(18):181602, 2006. https://doi.org/10.1103/PhysRevLett.96.181602

[109] Ludwig Wittgenstein. *Philosophical Investigations*. Blackwell, 1953.

[110] John R. Searle. Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3):417–457, 1980. https://doi.org/10.1017/S0140525X00005756

[111] John R. Lucas. Minds, Machines and Gödel. *Philosophy*, 36(137):112–127, 1961.

[112] Jean-Paul Sartre. *Critique of Dialectical Reason*. Vintage Books, 1960.

[113] Antonio Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam, 1994.

[114] David Hume. *A Treatise of Human Nature*. John Noon, London, 1739.

[115] G. E. Moore. *Principia Ethica*. Cambridge University Press, 1903.

[116] David Hume. An Enquiry Concerning Human Understanding. Oxford University Press, 1748.

[117] Immanuel Kant. Groundwork of the Metaphysics of Morals. Cambridge University Press, 1785.

[118] Immanuel Kant. *Critique of Pure Reason*. Cambridge University Press, 1781.

[119] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete Problems in AI Safety. arXiv preprint arXiv:1606.06565, 2016. `https://doi.org/10.48550/arXiv.1606.06565`

[120] Brian Christian. *The Alignment Problem: Machine Learning and Human Values*. W. W. Norton and Company, 2020.

[121] Francis Crick and Christof Koch. Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2:263–275, 1990.

[122] Anne M. Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.

[123] Francisco J. Varela, Jean-Philippe Lachaux, Eliezer Rodriguez, and Jacques Martinerie. The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2:229–239, 2001.

[124] Roger Penrose. *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford University Press, 1994.

[125] Stuart R. Hameroff and Roger Penrose. Consciousness in the universe: A review of the 'Orch OR' theory. *Physics of Life Reviews*, 11(1):39–78, 2014. `https://doi.org/10.1016/j.plrev.2013.08.002`

[126] Pentti Kanerva. *Hyperdimensional Computing: An Introduction to Computing in Distributed Representation with High-Dimensional Random Vectors*. Cognitive Computation, 1(2):139–159, 2009.

[127] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Matteo Carandini, and Kenneth D. Harris. High-dimensional geometry of population responses in visual cortex. *Nature*, 571(7765):361–365, 2019. `https://doi.org/10.1038/s41586-019-1346-5`

[128] Michael W. Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł Dłotko, Ran Levi, Kathryn Hess, and Henry Markram. Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in Computational Neuroscience*, 11:48, 2017. https://doi.org/10.3389/fncom.2017.00048

[129] Trevor Nestor. Theoretical approaches to solving the shortest vector problem in NP-hard lattice-based cryptography with post-SUSY theories of quantum gravity in polynomial time by Orch-Or. Cryptology *ePrint Archive*, Paper 2024/1714, 2024. https://eprint.iacr.org/2024/1714

[130] Daniele Oriti. Spin foam models of quantum gravity. *Reports on Progress in Physics*, 64(12):1703–1756, 2003. https://doi.org/10.1088/0034-4885/64/12/203

[131] Alejandro Perez. Spin foam models for quantum gravity. *Classical and Quantum Gravity*, 20(6):R43–R104, 2003. https://doi.org/10.1088/0264-9381/20/6/201

[132] Carlo Rovelli and Lee Smolin. Spin networks and quantum gravity. *Physical Review D*, 52(10):5743–5759, 1995. https://doi.org/10.1103/PhysRevD.52.5743

[133] John C. Baez. Spin foam models. *Classical and Quantum Gravity*, 15(7):1827–1858, 1998. https://doi.org/10.1088/0264-9381/15/7/008

[134] Daniele Micciancio. The Shortest Vector Problem is NP-hard. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 104–112, 1998.

[135] Yoshihiro Jibu, Kunio Yasue, Stuart R. Hameroff, Richard Pribram, and Karl H. Freeman. Quantum optical coherence in cytoskeletal microtubules: implications for brain function. *Biosystems*, 32(3):195–209, 1994. https://doi.org/10.1016/0303-2647(94)90042-9

[136] Laurence A. Amos and Aaron Klug. Arrangement of subunits in flagellar microtubules. *Journal of Cell Science*, 14(1):523–549, 1974.

[137] René Descartes. *Meditations on First Philosophy.* Jean Leclerc (Paris), 1641.

[138] Roger Penrose. *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics.* Oxford University Press, 1989.

[139] Karl H. Pribram. Brain and Perception: Holonomy and Structure in Figural Processing. Lawrence Erlbaum Associates, 1991.

[140] David Bohm. Wholeness and the Implicate Order. Routledge and Kegan Paul, 1980.

[141] Pascal Fries. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10):474–480, 2005. https://doi.org/10.1016/j.tics.2005.08.011

[142] Edward Witten. Anti–de Sitter Space and Holography. *Advances in Theoretical and Mathematical Physics*, 2(2):253–291, 1998. https://doi.org/10.4310/ATMP.1998.v2.n2.a2

[143] Raphael Bousso. The Holographic Principle. *Reviews of Modern Physics*, 74(3):825–874, 2002. https://doi.org/10.1103/RevModPhys.74.825

[144] William K. Wootters and Wojciech H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982. https://doi.org/10.1038/299802a0

[145] B. Jack Copeland. Hypercomputation: Computing Beyond the Church–Turing Barrier. *Minds and Machines*, 12(4):461–502, 2002.

[146] Valentino Braitenberg and Arbib S. Schüz. *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer, 1998.

[147] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. https://doi.org/10.1038/323533a0

[148] Travis J. A. Craddock, Stuart R. Hameroff, Ahmed T. Ayoub, Mariusz Klobukowski, and Jack A. Tuszyński. Anesthetics act in quantum channels in brain microtubules to prevent consciousness. *Current Topics in Medicinal Chemistry*, 15(6):523–533, 2015. https://doi.org/10.2174/1568026615666150225104543

[149] Travis J. A. Craddock, Philip Kurian, Jordane Preto, Kamlesh Sahu, Stuart R. Hameroff, Mariusz Klobukowski, and Jack A. Tuszyński. Anesthetic alterations of collective terahertz oscillations in tubulin correlate with clinical potency: Implications for anesthetic action and post-operative cognitive dysfunction. *Scientific Reports*, 7:9877, 2017. https://doi.org/10.1038/s41598-017-09992-7

[150] Jonathan Z. Pan, Jin Xi, John W. Tobias, Maryellen F. Eckenhoff, and Roderic G. Eckenhoff. Halothane binding proteome in human brain cortex. *Journal of Proteome Research*, 6(2):582–592, 2007. https://doi.org/10.1021/pr060311u

[151] Jonathan Z. Pan, Jin Xi, Maryellen F. Eckenhoff, and Roderic G. Eckenhoff. Inhaled anesthetics elicit region-specific changes in protein expression in mammalian brain. *Proteomics*, 8(14):2983–2992, 2008. https://doi.org/10.1002/pmic.200800057

[152] Daniel J. Emerson, Brian P. Weiser, John Psonis, Zhengzheng Liao, Olena Taratula, Ashley Fiamengo, Xiaozhao Wang, Keizo Sugasawa, Amos B.

Smith, Roderic G. Eckenhoff, and Ivan J. Dmochowski. Direct modulation of microtubule stability contributes to anthracene general anesthesia. *Journal of the American Chemical Society*, 135(14):5389–5398, 2013. https://doi.org/10.1021/ja311171u

[153] Max Tegmark. Importance of quantum decoherence in brain processes. *Physical Review E*, 61(4):4194–4206, 2000. https://doi.org/10.1103/PhysRevE.61.4194

[154] Gregory S. Engel, Tessa R. Calhoun, Elizabeth L. Read, Tae-Kyu Ahn, Tomáš Mančal, Yuan-Chung Cheng, Robert E. Blankenship, and Graham R. Fleming. Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature*, 446(7137):782–786, 2007. https://doi.org/10.1038/nature05678

[155] Thorsten Ritz, Salih Adem, and Klaus Schulten. Resonance effects indicate a radical-pair mechanism for avian magnetic compass. *Nature*, 429(6988):177–180, 2004. https://doi.org/10.1038/nature02534

[156] David H. Brookes, Fran Hartoutsiou, Andrew P. Horsfield, and A. M. Stoneham. Could humans recognize odor by phonon assisted tunneling? *Physical Review Letters*, 98(3):038101, 2007. https://doi.org/10.1103/PhysRevLett.98.038101

[157] Herbert Fröhlich. Long-range coherence and energy storage in biological systems. *Nature*, 228(5272):1093–1094, 1970. https://doi.org/10.1038/2281093a0

[158] Satyajit Sahu, Subrata Ghosh, Batu Ghosh, Krishna Aswani, Kazuto Hirata, Daisuke Fujita, and Anirban Bandyopadhyay. Atomic water channel controlling remarkable properties of a single brain microtubule: correlating single protein to its supramolecular assembly. *Biosensors and Bioelectronics*, 47:141–148, 2013. https://doi.org/10.1016/j.bios.2013.02.050

[159] N. S. Babcock, G. Montes-Cabrera, K. E. Oberhofer, M. Chergui, G. L. Celardo, and P. Kurian. Ultraviolet superradiance from mega-networks of tryptophan in biological architectures. *The Journal of Physical Chemistry B*, 128(17):4035–4046, 2024. https://doi.org/10.1021/acs.jpcb.3c07936

[160] Travis J. A. Craddock, Dean Friesen, J. Manoj, Stuart R. Hameroff, and Jack A. Tuszyński. The feasibility of coherent energy transfer in microtubules. *Journal of The Royal Society Interface*, 11(100):20140677, 2014. https://doi.org/10.1098/rsif.2014.0677

[161] T. J. Craddock, E. M. Preto, and S. H. Yee. Theoretical evidence for a quantum excitonic coherence mechanism of microtubule vibrations. *Journal of Chemical Physics*, 142(12):125103, 2015. doi:10.1063/1.4914306.

[162] Renxiu Tang and Jiapei Dai. Biophoton signal transmission and processing in the brain. *Journal of Photochemistry and Photobiology B: Biology*, 139:71–75, 2014. https://doi.org/10.1016/j.jphotobiol.2013.12.008

[163] Sandra Mamani, Lingyan Shi, Daniel Nolan, and Robert Alfano. Majorana vortex photons: A form of entangled photons propagation through brain tissue. *Journal of Biophotonics*, 12(10):e201900036, 2019. https://doi.org/10.1002/jbio.201900036

[164] Petr Cifra and Pak Kin Wong. Electromechanical coupling and acoustic vibrations in microtubules: energy transfer and signal propagation. *Journal of Theoretical Biology*, 262(1):34–45, 2010. https://doi.org/10.1016/j.jtbi.2010.05.023

[165] A.Yu. Kitaev. Unpaired Majorana fermions in quantum wires. *Physics-Uspekhi*, 44(10S):131–136, 2001. https://doi.org/10.1070/1063-7869/44/10S/S29

[166] A.Yu. Kitaev. Fault-tolerant quantum computation by anyons. *Annals of Physics*, 303(1):2–30, 2003. https://doi.org/10.1016/S0003-4916(02)00018-0

[167] Frank Wilczek. Quantum Time Crystals. *Physical Review Letters*, 109(16):160401, 2012. https://doi.org/10.1103/PhysRevLett.109.160401

[168] Soonwon Choi, Jeffrey Choi, Renate Landig, Georg Kucsko, Hengyun Zhou, Jun Ye, Hannes Pichler, and Mikhail D. Lukin. Observation of discrete time-crystalline order in a disordered dipolar many-body system. *Nature*, 543(7644):221–225, 2017. https://doi.org/10.1038/nature21426

[169] Komal Saxena, Pushpendra Singh, Jhimli Sarkar, Pathik Sahoo, Subrata Ghosh, Soami Daya Krishnananda, and Anirban Bandyopadhyay. Polyatomic time crystals of the brain neuron extracted microtubule are projected like a hologram meters away. *Journal of Applied Physics*, 132(19):194401, 2022. https://doi.org/10.1063/5.0130618

[170] Gilles Dumas, Jacques Nadel, Rémi Soussignan, Hugues Martinerie, and Arnaud Garnero. Inter-brain synchronization during social interaction. *PLoS ONE*, 5(8):e12166, 2010. https://doi.org/10.1371/journal.pone.0012166

[171] Francesco Babiloni and Luca Astolfi. Social neuroscience and hyperscanning techniques: Past, present and future. *Neuroscience & Biobehavioral Reviews*, 44:76–93, 2014. https://doi.org/10.1016/j.neubiorev.2014.03.007

[172] Sander Dikker, Lucia Wan, Ian E. Davidesco, David Kozin, Rachel Kelly, Dan McClintock, Jeremy Rowland, Matthew Quigley, and Gina Berman. Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, 27(9):1375–1380.e3, 2017. https://doi.org/10.1016/j.cub.2017.03.049

[173] Xinyue Cui, David M. Bryant, and Allan L. Reiss. NIRS-based hyperscanning reveals increased interpersonal coherence in the prefrontal cortex during cooperation. *NeuroImage*, 59(3):2430–2437, 2012. https://doi.org/10.1016/j.neuroimage.2011.09.084

[174] Diego A. Reinero, Suzanne Dikker, and Jay J. Van Bavel. Inter-brain synchrony in teams predicts collective performance. *Social Cognitive and Affective Neuroscience*, 16(1–2):43–57, 2021. https://doi.org/10.1093/scan/nsaa135

[175] Unai Vicente, Alberto Ara, and Josep Marco-Pallarés. Intra- and inter-brain synchrony oscillations underlying social adjustment. *Scientific Reports*, 13:11211, 2023. https://doi.org/10.1038/s41598-023-38292-6

[176] Artur Czeszumski, Sophie Hsin-Yi Liang, Suzanne Dikker, Peter König, Chin-Pang Lee, Sander L. Koole, and Brent Kelsen. Cooperative behavior evokes interbrain synchrony in the prefrontal and temporoparietal cortex: A systematic review and meta-analysis of fNIRS hyperscanning studies. *eNeuro*, 9(2):ENEURO.0268-21.2022, 2022. https://doi.org/10.1523/ENEURO.0268-21.2022

[177] Uri Hasson, Asif A. Ghazanfar, Bruno Galantucci, Scott Garrod, and Christian Keysers. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, 16(2):114–121, 2012. https://doi.org/10.1016/j.tics.2011.12.007

[178] Ian Konvalinka and Andreas Roepstorff. The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience*, 6:215, 2012. https://doi.org/10.3389/fnhum.2012.00215

[179] Ulman Lindenberger, Shu-Chen Li, Winne Gruber, and Verena Müller. Brains swinging in concert: cortical phase synchronization while playing guitar. *Proceedings of the National Academy of Sciences*, 106(28):11576–11581, 2009. https://doi.org/10.1073/pnas.0908994106

[180] Claudio Babiloni, Fabrizio Vecchio, Francesco Infarinato, Paola Buffo, Nicola Marzano, Danilo Spada, Simone Rossi, Ivo Bruni, Paolo M. Rossini, and Daniela Perani. Simultaneous recording of electroencephalographic data in musicians playing in ensemble. *Cortex*, 47(9):1082–1090, 2011. https://doi.org/10.1016/j.cortex.2011.05.006

[181] Fabrizio De Vico Fallani, Vincenzo Nicosia, Raffaella Sinatra, Laura Astolfi, Fabio Cincotti, Donatella Mattia, Vittorio Latora, and Fabio Babiloni. Defecting or not defecting: how to "read" human behavior during cooperative games by EEG measurements. *arXiv preprint arXiv:1101.5322*, 2011. https://arxiv.org/abs/1101.5322

[182] Ulman Lindenberger, Shu-Chen Li, Winne Gruber, and Verena Müller. Brains swinging in concert: cortical phase synchronization while playing guitar. *Proceedings of the National Academy of Sciences*, 106(28):11576–11581, 2009. https://doi.org/10.1073/pnas.0908994106

[183] Stephan Flory, Sabino Guglielmini, Felix Scholkmann, Valentine L. Marcar, Martin Wolf, et al. How our hearts beat together: a study on physiological synchronization based on a self-paced joint motor task. *Scientific Reports*, 13:11987, 2023. https://doi.org/10.1038/s41598-023-39083-9

[184] Fabian Behrens, Francesco Astolfi, Francesca Cincotti, et al. Physiological synchrony is associated with cooperative success in real-life interactions. *Scientific Reports*, 10:76539, 2020. https://doi.org/10.1038/s41598-020-76539-8

[185] Andrea Bizzego, Francesca Palumbo, Arianna Villani, et al. Strangers, friends, and lovers show different physiological synchrony in different emotional states. *Behavioral Sciences*, 10(1):11, 2019. https://doi.org/10.3390/bs10010011

[186] Quentin Moreau, Lena Adel, Caitriona Douglas, Ghazaleh Ranjbaran, and Guillaume Dumas. A neurodynamic model of inter-brain coupling in the gamma band. *Journal of Neurophysiology*, 128(5):1085–1090, 2022. https://doi.org/10.1152/jn.00224.2022

[187] Chun-Liang Loh and Tom Froese. An oscillator model for interbrain synchrony: Slow interactional rhythms entrain fast neural activity. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2180–2183, 2021. https://doi.org/10.1109/CIBCB49929.2021.9562779

[188] K. M. Sharika, Swarag Thaikkandi, Nivedita Nivedita, and Michael L. Platt. Interpersonal heart rate synchrony predicts effective information processing in a naturalistic group decision-making task. *Proceedings of the National Academy of Sciences*, 121(18):e2313801121, 2024. https://doi.org/10.1073/pnas.2313801121

[189] Luis Barahona and Louis M. Pecora. Synchronization in small-world systems. *Physical Review Letters*, 89(5):054101, 2002. https://doi.org/10.1103/PhysRevLett.89.054101

[190] Stefano Boccaletti, Ginestra Bianconi, Romualdo Criado, Chaoming del Genio, Jesús Gómez-Gardeñes, Mason A. Romance, Ignacio Sendiña-Nadal,

Zhi-Xiu Wang, and Marcelo Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014. https://doi.org/10.1016/j.physrep.2014.07.001

[191] Filippo Battiston, Giovanni Petri, Mason A. Porter, and Vittorio Latora. Networks beyond pairwise interactions: structure and dynamics in hypergraphs. *Nature Reviews Physics*, 2:538–553, 2020. https://doi.org/10.1038/s42254-020-0190-2

[192] Andrei Khrennikov. *Ubiquitous Quantum Structure: From Psychology to Finance*. Springer, 2010. https://doi.org/10.1007/978-3-642-04850-7

[193] Buu-Hoi Baaquie. *Quantum Finance: Path Integrals and Hamiltonians for Options and Interest Rates*. Cambridge University Press, 2004. https://doi.org/10.1017/CBO9780511613755

[194] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Physical Review Letters*, 83(7):1467–1470, 1999. https://doi.org/10.1103/PhysRevLett.83.1467

[195] Emilio Haven and Andrei Khrennikov. *Quantum Social Science*. Cambridge University Press, 2013.

[196] Vladimir I. Yukalov and Didier Sornette. Decision theory with prospect interference and entanglement. *Theory and Decision*, 70(3):283–328, 2011. https://doi.org/10.1007/s11238-010-9190-y

[197] Viktor Müller and Ulman Lindenberger. Hyper-brain hyper-frequency network topology dynamics when playing guitar in a quartet. *Frontiers in Human Neuroscience*, 18:1416667, 2024. https://doi.org/10.3389/fnhum.2024.1416667

[198] David Engel and Thomas W. Malone. Integrated information as a metric for group interaction: Analyzing human and computer groups using a technique developed to measure consciousness. arXiv preprint arXiv:1702.02462, 2017. https://arxiv.org/abs/1702.02462

[199] Emmanuelle Tognoli and J. A. Scott Kelso. The phi complex as a neuromarker of human social coordination. *Proceedings of the National Academy of Sciences*, 104(19):8190–8195, 2007. https://doi.org/10.1073/pnas.0611453104

[200] Gregory J. Stephens, Lauren J. Silbert, and Uri Hasson. Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(29):14425–14430, 2010. https://doi.org/10.1073/pnas.1008662107

[201] Jing Jiang, Bin Dai, Dong Peng, Chi Zhu, Lu Liu, Chun-Yu Lu, and Chi Lu. Neural synchronization during face-to-face communication. *Journal of Neuroscience*, 32(45):16064–16069, 2012. https://doi.org/10.1523/JNEUROSCI.2876-12.2012

[202] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, 30:5998–6008, 2017. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[203] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in Neural Information Processing Systems*, 30:6379–6390, 2017. https://proceedings.neurips.cc/paper/2017/hash/68a9750337a4185dd5d38bcffccdbf3d-Abstract.html

[204] B. Jack Copeland. Hypercomputation. *Minds and Machines*, 12(4):461–502, 2002. https://doi.org/10.1023/A:1020859117880

[205] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling Laws for Neural Language Models. *arXiv preprint arXiv:2001.08361*, 2020. https://doi.org/10.48550/arXiv.2001.08361

[206] Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Trevor Cai, Eliza Rutherford, Jake Millican, Shane Legg, and Theo Anthropic. Training Compute-Optimal Large Language Models. *arXiv preprint arXiv:2203.15556*, 2022. https://doi.org/10.48550/arXiv.2203.15556

[207] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Brock Hoffman, Emmanuel Raffel, et al. Emergent Abilities of Large Language Models. *arXiv preprint arXiv:2206.07682*, 2022. https://doi.org/10.48550/arXiv.2206.07682

[208] David Attwell and Simon B. Laughlin. An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, 21(10):1133–1145, 2001. https://doi.org/10.1097/00004647-200108000-00001

[209] Fabio A. Azevedo, L. R. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. Ferretti, R. E. Leite, W. Jacob Filho, R. Lent, and S. Herculano-Houzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541, 2009. https://doi.org/10.1002/cne.21974

[210] Peter Lennie. The cost of cortical computation. *Current Biology*, 13(6):493–497, 2003. https://doi.org/10.1016/S0960-9822(03)00135-0

[211] Peter Polsky, Bartlett W. Mel, and Jackie Schiller. Computational subunits in thin dendrites of pyramidal cells. *Nature Neuroscience*, 7(6):621–627, 2004. https://doi.org/10.1038/nn1253

[212] Fred Rieke, David Warland, Rob de Ruyter van Steveninck, and William Bialek. *Spikes: Exploring the Neural Code*. MIT Press, 1997.

[213] NVIDIA. NVIDIA H100 Tensor Core GPU Architecture. 2022. https://www.nvidia.com/h100

[214] Panayiota Poirazi, Terrence Brannon, and Bartlett W. Mel. Pyramidal neuron as two-layer neural network. *Neuron*, 37(6):989–999, 2003. https://doi.org/10.1016/S0896-6273(03)00149-1

[215] Michael London and Michael Häusser. Dendritic computation. *Annual Review of Neuroscience*, 28:503–532, 2005. https://doi.org/10.1146/annurev.neuro.28.061604.135703

[216] Johanna Schiller, Gidon L. Major, Hong-Jian Koester, and Yair Schiller. NMDA spikes in basal dendrites of cortical pyramidal neurons. *Nature*, 404(6775):285–289, 2000. https://doi.org/10.1038/35005094

[217] Stefano Gasparini and Jeffrey C. Magee. State-dependent dendritic integration in hippocampal CA1 pyramidal neurons. *Proceedings of the National Academy of Sciences*, 99(7):4393–4398, 2002. https://doi.org/10.1073/pnas.072092899

[218] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014. https://doi.org/10.48550/arXiv.1411.4028

[219] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Ál án Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5:4213, 2014. https://doi.org/10.1038/ncomms5213

[220] Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019. https://doi.org/10.1038/s41586-019-0980-2

[221] Mark W. Johnson, Mihai H. S. Amin, Stephen Gildert, Trevor Lanting, F. Hamze, Neil Dickson, R. Harris, Andrew J. Berkley, Billy J. Johansson, Pawel Bunyk, Eugene M. Chapple, Carlo Enderud, Jean-Marc D. Seidel, Theodore L. Toups, H.-S. Wang, Alexandra J. Wilson, and George Rose. Quantum annealing with manufactured spins. *Nature*, 473(7346):194–198, 2011. https://doi.org/10.1038/nature10012

[222] Michael J. Martin, Caroline Hughes, Gilberto Moreno, Eric B. Jones, David Sickinger, Sreekant Narumanchi, and Ray Grout. Energy use in quantum data centers: Scaling the impact of computer architecture, qubit performance, size, and thermal parameters. *arXiv preprint arXiv:2103.16726*, 2021. https://doi.org/10.48550/arXiv.2103.16726

[223] Michel H. Devoret and Robert J. Schoelkopf. Superconducting circuits for quantum information: An outlook. *Science*, 339(6124):1169–1174, 2013. https://doi.org/10.1126/science.1231930

[224] Mike Davies, Narayan Srinivasa, Tsung-Cheng Lin, Gautham Chinya, Yongguang Cao, Sri Harsha Choday, Georgios Dimou, Prasad Joshi, Naveen Imam, Shweta Jain, et al. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 38(1):82–99, 2018. https://doi.org/10.1109/MM.2018.112130359

[225] Paul A. Merolla, John V. Arthur, Rodrigo Alvarez-Icaza, Andrew S. Cassidy, Jun Sawada, Filipp Akopyan, Bryan L. Jackson, Nabil Imam, Chen Guo, Yutaka Nakamura, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197):668–673, 2014. https://doi.org/10.1126/science.1254642

[226] Christoph Schemmel, Alexander Grübl, Karl Meier, and Sebastian Millner. A wafer-scale neuromorphic hardware system for large-scale neural modeling. *Journal of Circuit Theory and Applications*, 38(5):459–482, 2010. https://doi.org/10.1002/cta.433

[227] Yu-Hsin Chen, Tushar Krishna, Joel S. Emer, and Vivienne Sze. Eyeriss: A spatial architecture for energy-efficient dataflow for convolutional neural networks. In *Proceedings of the 43rd International Symposium on Computer Architecture*, pages 367–379, 2016. https://doi.org/10.1145/3079856.3080236

[228] John Preskill. Quantum Computing in the NISQ Era and Beyond. *Quantum*, 2:79, 2018.

[229] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, 86(3):032324, 2012. https://doi.org/10.1103/PhysRevA.86.032324

[230] Giacomo Indiveri and Shih-Chii Liu. Memory and information processing in neuromorphic systems. *Proceedings of the IEEE*, 103(8):1379–1397, 2015. https://doi.org/10.1109/JPROC.2015.2444094

[231] Visual Capitalist. What Electricity Sources Power the World? – Electricity sources by fuel in 2022. *Visual Capitalist*, 2022. https://www.visualcapitalist.com/electricity-sources-by-fuel-in-2022/

[232] International Energy Agency. Electricity 2024: Executive Summary. *IEA*, 2024. https://www.iea.org/reports/electricity-2024/executive-summary

[233] OpenAI. AI and Compute. OpenAI blog, May 16, 2018. https://openai.com/index/ai-and-compute/

[234] Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. Recalibrating global data center energy-use estimates: Growth in energy use has slowed owing to efficiency gains that smart policies can help maintain in the near term. *Science*, 367(6481):984–986, 2020. https://doi.org/10.1126/science.aba3758

[235] David Patterson, Joseph Gonzalez, Quoc V. Le, Chen Liang, Lluís-Miquel Munguía, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350*, 2021. https://doi.org/10.48550/arXiv.2104.10350

[236] BP. Statistical Review of World Energy 2023. BP, 2023. https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html

[237] Ellery W. Strubell, Ananya Ganesh, and Andrew McCallum. Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, 2019. https://doi.org/10.18653/v1/P19-1355

[238] The World Bank. GDP (current US$). https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

[239] UNESCO Institute for Statistics. Research and Development Expenditure. UNESCO, 2021. http://uis.unesco.org/en/topic/research-and-development

[240] Organisation for Economic Co-operation and Development. Development Co-operation Report 2023. OECD Publishing, 2023. https://doi.org/10.1787/ca6d5d1c-en

[241] Sheera Frenkel and Aaron Krolik. Trump Taps Palantir to Compile Data on Americans. *The New York Times*, May 30, 2025. https://www.nytimes.com/2025/05/30/technology/trump-palantir-data-americans.html :contentReference[oaicite:0]index=0

[242] William Strauss and Neil Howe. Generations: The History of America's Future, 1584 to 2069. William Morrow, 1991.

[243] G. W. F. Hegel. Phenomenology of Spirit. Oxford University Press, 1977. (Original work published 1807.)

[244] Nick Bostrom. Are You Living in a Computer Simulation? *Philosophical Quarterly*, 53(211):243–255, 2003.

[245] Ray Dalio. Principles for Dealing with the Changing World Order: Why Nations Succeed and Fail. Simon Schuster, 2021.

[246] João da Gama Batista, Jean-Philippe Bouchaud, and Damien Challet. Sudden trust collapse in networked societies. *The European Physical Journal B*, 88(3):55, 2015.

[247] Marten Scheffer, Jordi Bascompte, William A. Brock, Victor Brovkin, Stephen R. Carpenter, Vasilis Dakos, Hermann Held, Egbert H. van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461(7260):53–59, 2009.

[248] Organisation for Economic Co-operation and Development. Government at a Glance 2022. OECD Publishing, 2022.

[249] World Bank. Worldwide Governance Indicators 2021. World Bank Group, 2021.