UCLA

Recent Work

Title

Novel Framework for a Cybernetic Theory of General Intelligence and Information Field Theory for Strong AI Research and Development - Virtual Neurons Are All You Need

Permalink

https://escholarship.org/uc/item/3cx3s73v

Author

Zukowski, Novak I. K.

Publication Date

2023-06-18

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, availalbe at https://creativecommons.org/licenses/by-nc/4.0/

eScholarship.org Powered by the California Digital Library
University of California

Novel Framework for a Cybernetic Theory of General Intelligence and Information Field Theory for Strong AI Research and Development - Virtual Neurons Are All You Need

Novak I. K. Zukowski
https://orcid.org/0009-0009-3887-3804

Electrical and Computer Engineering Department, UCLA, Los Angeles, 90095, California, USA

Abstract— **This project develops a framework for computational general intelligence that can be used to build computational models, or world models, for machine intelligence and agent decision making. The framework proposes a novel definition of information and proposes virtual neurons as a method to enable longer term prediction and planning for agent decision making. Virtual neurons function as the most fundamental building blocks of an agent's cognitive model of the world and are the mediating particles of an agent's interaction with the proposed information field, and represent the set of an agent's available receptive fields that span multiple layers of abstraction. Layers of virtual neurons are computationally represented by graph structures to enable path planning, task navigation, and enable decision making and coordination with other agents through the creation of a mutual goal space. The framework bears particular applications to developing reinforcement learning agents that operate in the real world. Additionally, this framework enables a mapping between machine intelligence and natural intelligence, which serves as a tool for improving human-AI alignment.**

Index Terms—**Machine Intelligence, Artificial General Intelligence, Strong AI, Cognitive Architecture, Machine Common Sense, Neural-Symbolic AI, Cybernetics, Engineered Natural Intelligence, Deep Reinforcement Learning, Multi-Agent Reinforcement Learning, Trust, Game Theory, Information Field Theory, Entropy, Emergent Systems, Complexity Science, Graph Networks, Computational Decision Making**

I. Introduction

A. History

Deep reinforcement learning methods commonly use Markov decision problems to model optimal action-reward combinations based on feedback systems between agents and environmental conditions and have made started to make successes in handling higher dimensionality environments [1]. Cognitive architectures are computational models that reflect human-like cognition and generally enable perception and planning through high level semantic and symbolic reasoning [2]. Over the past decade, both deep reinforcement learning and cognitive architecture approaches have been implemented for decision making and optimal control applications in autonomous vehicles, robotics, recommendation systems, language models, and various other areas [2][1].

B. Constraints

Whereas artificial intelligence models have been most commonly trained on zero sum games environments, navigation and interaction in real world environments involve non-zero sum environments that have higher dimensionality than zero sum environments that has presents challenges due to constraints in computational power and quality of available training data [3]. Due to the inherent uncertainty in the dynamics of the real world, it is intractable to predict every edge case that will occur in real world operation that an agent could encounter [5].

II. Motivation

To assure safe and reliable operation of autonomous agents in human populated environments, it important to develop autonomous agents with generalized intelligence to enable higher versatility such that agents can be trusted to operate successfully under unforeseen edge cases [5]. Cognitive architectures have potential to be improved in scalability for their level of complexity. Reinforcement learning frameworks have potential for improvement in adaptive task learning to enable proficiency across multiple tasks for more optimal performance in edge case conditions. In addition to these bottlenecks of existing systems, enabling better human-AI alignment with human goals and long-term interests is an open area of research of high interest [4].

Intuitively, enabling long term prediction for agents that function as AI assistants, content filtering systems, and recommendation systems is an essential capability to solve to enable alignment with longer term interests of users. Long term prediction has been primary challenge that has given rise to a variety of approaches [6].

In addition to existing definitions of information having some limits in resolution, there are various definitions of entropy for different applications [7]. This framework posits that definitions of information and entropy that have higher scope invariance would be an important component to enabling real world robustness to optimize agent attention, allocation of processing resources, and the balance between exploration and exploitation decisions.

The framework posits that multi-agent deep reinforcement learning, graph networks, neural network attention mechanisms, neural information theory, cognitive architectures, and related tools will be important tools that yield progress towards the development of versatile machine intelligence and strong AI. This project has provided the opportunity to develop resolution on the feasible implementations that can be presently created at their intersections and their evolution going forward.

III. Approach

This has been a self-directed project with advising from UCLA Professor Dr. Briggs to develop a framework that lays a foundation for drawing design principles from to develop versatile human-aligned machine intelligence with focus on identifying scale invariant and scope invariant rules, axioms, and laws related to natural intelligence, cybernetics, and machine intelligence. Identifying invariant components of information processing can yield insight towards the far sight and feasible implementations that may precipitate from research and development efforts. With consideration to the high rate of progress in artificial intelligence, these insights can identify more stable ground to invest efforts in.

The plans for this project were to investigate the scope of feasible applications that precipitate from the intersections between graph neural networks, spiking neural networks, neural network attention mechanisms and transformers, multi-agent reinforcement learning, information theory, reinforcement learning from human feedback from EEG; then to develop an ensemble of algorithms to demonstrate a proof of concept of human-aligned machine decision making that would optimally exhibit a sign of emergent properties.

After some investigation, it was determined the optimal tools to use for a proof of concept would be multi-agent reinforcement learning and graph neural networks. Due to the high complexity that multi-agent reinforcement learning (MARL) can scale to, after developing an understanding of the other tools mentioned, much focus was allocated to this area. Git was setup for version control and integrated with Google Colaboratory. From testing existing MARL programs and learning how to utilize the common MARL frameworks, it was found that a MARL environment hosted on a local machine would be more suitable than on Google Colaboratory.

Common frameworks for implementing MARL projects are Mava, Open AI Gym, and others. Message Passing Networks, Graph Attention Networks, transformer architectures, and other methods of machine learning are tools have been integrated with other MARL projects of similar nature.

IV. Operation

The ensemble of rules, axioms, and laws in this framework are observational and are intended to enable computational tractability such that world models can be built from them that enables decision making from environmental feedback, longer term planning, novel task learning, and more versatile general intelligence for reinforcement learning agents. Observational laws have the utility of enabling reliable prediction while not being physical laws. Some examples of well-known observational laws are Ohm's Law, Moore's Law, Hubble's Law, and Boyle's Law. This framework can be thought of as a theory of all phenomena relevant to general intelligence from which observational laws, axioms, and rules can be extracted to build computational world models to create progress towards developing human-aligned artificial general intelligence. It is important to build agents with an emphasis on assured human-alignment. Risk assessment and strong AI (AGI) safety is also of great important and will be a topic of future research.

In our human brains, we maintain models of the world that enable decision making through understanding our environment and understanding which actions lead us toward the goal states we seek. It is intractably complex to compute simulations of very low-level physical phenomena and of organic information processing done in brains with full fidelity due to the finite amount of processing power available in both organic brains and machine processing hardware. In order to form computational models, simplifications and linearizations must be made while minimizing the necessary sacrifice to accuracy and precision such to sacrifice precision only where precision provides low returns on its energy investment. Abstracting away non-linear phenomena such as the Brownian motion of particles and other dynamics is analogous to a form of mean field analysis [8]. Ascertaining the change in accuracy that occurs in response to a change in precision is a critical and nuanced matter in machine learning and other fields. The necessary exchange between accuracy, precision, and effective scope to cope with finite processing resources is part of what leads us to maintain and develop a diverse set of models despite the common desire to have a unified theory. Each of mathematical and computational models have their own utility, accuracy, and precision relevant to their respective scope.

This framework's observational laws, axioms, and rules pertain to both the physical domain and the organic cognitive domain, and are intended to lay a foundation for a more unified theory of general intelligence to be applicable to both biological agents and machine agents. The development of reinforcement learning frameworks that enable more generalized agent intelligence are intended to precipitate from this framework through the resulting world models built from them. Optimally, a world model facilitates emergent behavior in agents that bears similarities to the non-algorithmic behaviors observed in organic behavioral dynamics. This prompts investigation on the mechanisms that enable higher level abstract reasoning to emerge from low level axioms where the axioms define how "atomic" elements within a complex system interact. John Conway's Game of Life is a well-known example of a model of simple rules that lead to emergent behavior [9].

This area of research involved with developing bio-inspired frameworks for machine intelligence is related to neural-symbolic AI, engineered natural intelligence, and neuromorphic engineering; and draws from neuroscience, machine learning, cybernetics, complex systems research, among other fields [10][11]. It should be noted that neuroscientific and biological models provide invaluable insights for optimizing machine learning systems, but the most functional artificial agent may not call for a direct simulation of biological processes. It is important to gain some resolution on what principles of organic information processing can provide utility in the machine domain while reducing computational complexity [12]. Bearing a neuroscientific understanding of organic brain function also bears important utility in ensuring human-AI alignment. Similar to how explainable AI helps understand a model's operation that leads to its outputs, this framework helps build a model of explainable organic intelligence. These are some of the motivations for developing this framework.

A. Abstraction, Semantics, and Cognitive-Physical Error

When reasoning across multiple scopes and applications, it is important to distinguish and resolve ambiguity between closely related terms. In areas like cybernetics and engineered intelligence that lie at the intersection of many fields, it is necessary to identify similarity and symmetries to find the common ground between all a term's definitions across multiple scopes to gain resolution on the correspondences (traces) between linguistic meaning, mathematical meaning, physical real-world meaning. Max Tegmark, a prominent physicist and machine learning researcher, goes as far to introduce the term human "baggage" to describe the higher-level abstraction space incumbent with language that exists on top of the physical external reality that humans use to mentally navigate and comprehend the mathematical structures that correspond to the physical reality [13]. Since achieving adept navigation of mathematical structures often needs to be preceded by linguistic processing for humans, "baggage" is associated with the cost of cognitive energy dedicated to linguistic processing in order to work towards developing a higher fidelity mental model of the physical world.

In a similar light, this framework posits that the informative difference between the cognitive representation of a phenomenon and the physical nature of the phenomenon can be termed cognitive-physical error. The higher the cognitive-physical error of an agent’s model of the environment around it, the more risk an agent's actions are prone to. Cognitive-physical error costs energy to correct. If not corrected before an action, then it is said that the agent accepts the incumbent risk. Risk costs energy to recover from in the chance that the related adverse condition occurs.

Investigating the nature of the abstraction space and semantic spaces associated with language can provide some important insights on general intelligence. Identifying the common ground between how a term is defined across multiple scopes provides resolution on its corresponding physical meaning, and likewise, the term's position in embedding space that language decodes and encodes. The embedding space in language is correlated to it's semantic space. A study at Meta observed that point cloud representations of the word embedding spaces across human languages have significant similarities [14]. This provides empirical evidence that the relational knowledge graphs in the language embedding space have symmetries from person to person regardless of country of origin, thus intuitively provides some empirical evidence that the semantic space is more rooted in genetic attributes of a species rather than the conditions an organism is born into. It can be inferred that this is related to the consistency of neural architecture from organism to organism within a species, and even across species to some degree. A term's common ground allows for the most general form to be recognized and is optimally the smallest maximum level of abstraction needed for the term to maintain utility in its range of all applicable scopes. During real-time reasoning over an agent's lifetime, as the number of scopes that a term is observed to be applicable to increases, the smallest maximum level of abstraction that corresponds to the term's most general form may increase: I.E. the term's level of abstraction increases as the agent observes how to utilize it over more application scopes.

One goal result enabled by a quality general form is maximization of the rate of precision gain and accuracy gain when a novel task area arises for an agent to navigate by decreasing the path complexity required to navigate through the task (the energy cost of the path). Learning capabilities in a new task area from previous tasks performed is an open area of research in reinforcement learning [15]. Enabling the identification of similarities between previous tasks and new tasks is important to this effort. In order to minimize computational complexity, it is optimal to minimize the number of terms and parameters needed to perform a task without sacrificing accuracy or precision. One key being able to reduce the number of parameters in machine learning models without sacrificing accuracy is having high quality training data. The active learning technique in machine learning has been conducive to this effort as a method to filter less relevant data training samples and use only the most relevant data samples during the model training process [16]. On the broader scale, the limited amount of available high quality data is currently a primary bottle neck in machine learning, and is especially the case for large language models [17].

One intermediate goal of this framework is to minimize the number of symbols and parameters needed to reason about all possible scenarios by eliminating linguistic ambiguity and overlapping definitions to reduce complexity without sacrificing the accuracy or precision of an agent's reasoning process. Clarifying definitions and how they're used in this framework bears some importance since the same terms may have some variation in definition as used by other research works. Definitions of terms will be clarified when they are introduced in this text. Having non-ambiguous definitions enables more tractable mathematical models to be built. Some of the concepts from the organic domain have been historically nuanced, which has made them more challenging to factor into mathematical models due to the limits in achievable resolution (signal clarity) on them. Having well resolved definitions throughout his framework lays a substrate to build models that enable optimization for decision making.

This framework is intended to enable the architecting of software and algorithms for reinforcement learning agents with general intelligence that process reward signals and other forms of information to enable self-learning. It may additionally point towards areas of interest to seek empirical evidence to further verify. There are a few terms that I've found exist only in this framework such as Cognitive-Physical Error and Cognitive-Kinetic Energy. The concept of information has also been slightly redefined from existing definitions to provide a higher resolution perspective on information and information transfer.

B. Virtual Neurons Introduction

Similar to energy transduction across mediums, and signal transduction via molecular interaction, information can be transduced between the organic domain and machine domain [18]. Just as energy transfer is discretized by Plank's constant and mediated by particles - photons. This framework proposes that information transfer can similarly discretized by Plank's constant and mediated by information particles - potential effective photons. This is a key proposition of this framework.

Physical phenomon is analog, whereas physical energy transfer is discretized by Plank's constant. Physical particles "sense" the world by electromagnetic charge and other forces. In the physical domain, sensing and responding is unified (fully compressed), since no cognition occurs between the sensing a force and responding to it.

An example of higher complexity physical sensing and reponse mechanisms are molecular interactions, as are studied in chemistry. In complex organisms, sensing is interlinked with the physical world via organic neurons. Organic neurons create points of measurement. Measurements are indicated by action potential spikes. For organisms to absolve themselves of the energy cost of mentally simulating a nearly infinite amount of nuance and complexity of the physical world: conscious experience is smoothly discretized by emergent states measured by conscious attention.

Through organic neuron activity, the space of analog conscious perception in organisms is smoothly generated from pooling and channeling of energy propagation, or flow, throughout neuronal space, akin to a multi-dimensional graph network where nodes represent where energy tends to pool and links represent where nodes tend to interact such that energy transfer occurs between nodes. This framework lays the foundation for a cognitive space, which is modeled as a virtual space composed of virtual neurons that are reflections of the emergent states observed and measured by conscious attention. Most virtual neurons are assigned a semantic meaning by the conscious mind.

Experiences that occur most often are represented by virtual neurons that have an attribute of high certainty. Novel virtual neurons reflect novel experiences, concepts, intuitions, thoughts, etc and take time for a semantic meaning to be assigned to them; these are "slippery" mental objects that require reasoning to ascertain their connection and utility to other virtual neurons so that they can be stored in the knowledge base more permanently as a semantic node. Through reasoning, virtual neuron networks are transduced into semantic networks by the mind. Through this mechanism, virtual neurons have a finite lifetime. The higher the periodicity of the event the virtual neuron is correlated with, the longer the virtual neuron's lifetime. Likewise, virtual neurons correlated with events that have the highest scale invariance and scope invariance, will have the longest lifetimes since they have more universal utility, leading them to be accessed more often. Event periodicity also has a correlation with the certainty of a virtual neuron: higher periodicity has a correspondence with higher certainty.

Whereas organic neurons are the transistors of the organic domain, virtual neurons are the emergent measured variables modulated by neural programs. The virtual domain exists as a medium between the organic domain and computational domain. Both the computational domain and virtual domain modulate nodes in the semantic domain. The semantic domain functions as a medium between the computational domain and virtual domain.

The minimum virtual state difference is discretized by Plank's constant, which reflects the discretization of physical energy transfer. This discretization enables virtual neuron activations to be modeled numerically. Plank's constant creates a convenient base factor for a messaging protocol to facilitate state activation transfer between the virtual domain and computational domain, but it may be modulated by other factors in either domain for energy scale translation.

Both AI models and organisms both have agent-relevant world models that are formulated from their receptive fields modeled by their agent-relevant virtual networks. Software is written in the computational domain and is a reflection of the most relevant parts of the virtual domain that can be measured via user data, semantic description via chat interface, EEG data, etc.

C. Invariance, Similarities, and Symmetries Create Virtual Neurons

This framework posits that an attribute of general intelligence is modulation between high level reasoning (high abstraction: low precision, high bandwidth reasoning) and low-level reasoning (low abstraction: high precision, low bandwidth reasoning); and that this ability may be related to scale invariance in the relational web of abstract structures that resembles the scale invariance of quasicritical organic neural networks. This may prove to be a fruitful area to gain empirical evidence on in future research. An example of where this attribute manifests is as experienced in business management positions: in effort to maintain alignment between high level perspectives of operations and ground truth, it is important to modulate between the mindset of generalist and the mindset of a specialist in response to task demands. This is often encouraged in organizations to dampen the effects associated with the Peter principle [19].

The relational web of abstract structures that corresponds to associative thought and word embeddings in abstract space (schema) can be computationally represented as a graph network. With this, this framework proposes that this relational graph space can be modeled as a network of "virtual neurons" akin to virtual addressing in computer architecture. This graph representation enables task planning to be computationally represented in the form of path planning through the graph network [20]. Using graph networks and path planning to map possible system states may be an area that is fruitful to assure the safety of an intelligent system and facilitate certification for wider scale operation and public interaction [21].

To enable utility across multiple scenarios, the axioms in this framework seek to model scope invariant and scale invariant properties and features. Scale invariance refers to a property where a system, pattern, or phenomenon remains unchanged when scaled by a certain factor. Scale invariance is often associated with self-similarity and power-law behavior across different scales as seen in fractal geometry, and stock price fluctuation patterns [22].

The brain has been observed to be self-balancing such to keep organic neural networks near the critical point of network connectivity, which optimizes information propagation through organic neural networks. Computational power, information storage, and sensitivity are maximized near the critical point. Scale invariance has been observed of neuron networks that are near the critical point, which enables long distance signal propagation [23].

This framework defines scope invariance to refer to elements that are conserved across multiple scopes or transformations. An agent's reasoning processes and preferences that are consistent across multiple scopes are said to be scope invariant. Scope invariance intuitively has a relation to symmetry [24]. This framework posits that organic brains are able to invest in developing scope invariant reasoning processes because many events in the real world environment exhibit periodicity or quasiperiodicity. Brains invest energy in developing responses to stimuli that are projected to reoccur in order to optimize return on energy investment. Neurons have the natural property of exploiting time symmetric events in this regard.

It's been observed in a study on organoid intelligence done by Cortical Labs that organic neurons tend towards preferring periodic activity and synchronization with environmental stimuli [25]. This framework posits that this indicates organic brains tend towards increasing the signal-to-noise ratio over time in their detection of periodic environmental conditions in order to converge towards an optimal action selection policy. By counter example, if organic neurons did not prefer to tend towards synchronizing their activity with periodic external events, then decision making and movement in biological agents would be random, which would not result in biological agents efficiently obtaining their food sources, which would inhibit their survival.

D. Neural Software Programs and Graph-based Cognitive Navigation

The Gestalt process involved with smoothing perception from one frame to another can be seen as a spontaneous pattern completion process, which this framework posits arises from the same (or similar) mechanism that drives neurons to prefer periodicity, which generates the impetus towards Signal-to-Noise-Ratio increase [26]. In a computational analogy, organic neurons seem to detect graph symmetry to reduce the energy of task planning by seeking to recycle previously developed stimulus-response associations that have been historically effective. It can be inferred that this is part of how energy efficiency is achieved in organic neural networks. A group of stimulus-response dynamics in this framework is defined as a neural software program: a policy. Neural software is recruited in order from subconscious (purely instinctive) to conscious (purely logical). This will be expounded upon.

A network of associated thoughts is referred to as schema and can be represented as a graph network. Learning the association between which actions are effective in given situations is referred to as associative search tasks, contextual bandits, and multi-armed bandit problems in reinforcement learning and are an open area of research [27]. This framework helps elucidate how the contextual bandits scope manifests in the organic domain.

Information processing in organic brains is drawn towards rates of change rather than static conditions. Entropy in moderate amounts attracts curiosity. Too much entropy leads to excess stress whereas too little entropy leads to boredom [28]. The conditions that are static across multiple frames often do not capture the mind's attention since the appropriate response to the conditions would have already become known over sufficient exposure time and learning. Likewise, cognitive resources (attention) tends to be pointed towards where a new response needs to be formulated. Consciously generating a new response entails the inference process and generation of information in order to conceptualize and plan the response in the cognitive domain. It has been shown that curiosity reduction is rewarded by dopaminergic pathways in the brain [29].

As an agent navigates its environment, data stored in it's knowledge base is accessed by the agent to build a relational graph of semantic concepts related to the relevant tasks and goals driven by environmental conditions. Since the relational structure of semantic concepts have some recursive properties where the meaning of some terms can be seen to be able to be nested inside other terms (akin to Russian dolls), this framework posits that nodes within the semantic graph differentiate in response to state changes such to increase the resolution of concepts relevant to an agent's present environment, and optimally converges toward "unpacking" the semantic relations that lead to the effective permutation of operations on available resources in order to achieve the current task parameters [30]. Since an agent's "brain" generates an approximately fixed amount of power that can be allocated to real-time processing, an agent's cognitive bandwidth in working memory allows for only a fixed number of parameters to represent the semantic space. The packing and unpacking of relevant semantic nodes from an agent's knowledge base to traverse levels of abstraction and reasoning processes enables a modular reasoning space too enable adaptive functionality under fixed processing power. This is can also be described as resolution modulation.

E. Virtual Neurons as Receptive Fields to Balance Exploitation and Exploration Based on Quantified Information and Risk

One of the cornerstones of this framework is the introduction of a novel perspective and definitions of information and entropy with the goal to achieve computational tractability in the world model for long term path planning, optimal balancing between agent exploitation and exploration, and "common sense" that enables optimal attention and energy allocation towards relevant tasks for agent success in novel conditions.

This framework defines relationships between data, information, entropy, and energy and how they are used to model interactions. This framework entails agent-agent interactions, agent-environnent interactions, and environment-environment interactions. There is a distinction between information and data. Information entails inference. Inference acts to fill in gaps in a data set. Information is generated as a result of data compression, and results in inference. Inference entails either extrapolation or interpolation. Extrapolation refers to "filling in" future data points and is synonymous with projection, while interpolation refers to "filling in" historic data points.

This framework defines projection as the output of extrapolation from data and defines prediction as projection plus anticipation. Anticipation is defined as projection from data derived from intrinsic regions of the brain that are less logical and more instinct driven.

When observations of events in the environment are made by an agent, they are recorded in the agent's memory in its cognitive model. An agent's cognitive model is an instantiation of the world model. Whereas the world model entails all of the known computational laws of the reinforcement learning universe, the cognitive model is the vector space of built from the cognitive bandwidth that's currently available or available through saccades of attention. Policies exist within an agent's cognitive model as neural software programs that encode stimulus-response association mappings.

Policies are task relevant. Tasks are goal relevant. Therefore, policies are goal relevant. Goals are relevant to intrinsic needs. Therefore, all tasks an agent performs are ultimately aimed to meet its intrinsic needs. Intrinsic needs are related to neurotransmitter systems, which will be discussed later. An agent may have multiple concurrent goals and multiple current task processes, therefore multiple active policies that superpose to form the agent's behavior that manifests at the mesoscale. The Ada model recently developed by DeepMind is an example of research to enable multi-task capability through adaptive policy learning [31].

The study on language translation at Meta provides empirical evidence that "embeddings of words in different languages share similar neighborhood structure, because people across the world share the same physical world" [32]. As discussed previously, this framework defines the semantic space that is synonymous with the embedding space as a domain of virtual neurons used for semantic reasoning and decision making. These virtual neurons are the base layer of the cognitive domain of an agent. Each virtual neuron represents a receptive field of an agent. The definitions in this framework are such to align receptive field nodes in a low entropic lattice by minimizing ambiguity, which facilitates mathematical formulation and computation. Low entropy virtual neural lattices maximize fluid cognition. (The definition of neural entropy within this framework differs from the definition of neural entropy commonly used in neuroscience, which will be further explained.)

A receptive field functions like an active question since it is a point of observation. Receptive fields guide conscious attention, which leads to potential work, which represented by information. Receptive fields and questions create information channels. Receptive fields are the interaction points of information - information is generated by what an agent is able to measure. This framework proposes that since information acts to generate projections that guide actions, information is rooted in relative rates of change. Information elucidates a future condition to the observer (agent).

Local rates of change that equal zero also impart information. This is because the environment is assumed to be dynamic and changing with time, thus by projecting that a quantity measured by a receptive field will remain constant, this indicates that the quantity is constant relative to other receptive fields that change in synchronization with the environment. The change in the environment is due to interactions within the environment. Since the number of interactions between all the particles and objects in an environment is intractably complex, there is a degree of uncertainty in predicting future states. This uncertainty due to exceedingly numerous interactions is defined as entropy. A single observed interaction is a form of entropy and is defined as a single unit of complexity since interaction leads to a state change. Both information and complexity are measures of entropy and are therefore subsets of entropy. Information is entropy resolved. Complexity is entropy realized. Complexity relates to deterministic or determined interactions. Information relates to probabilistic interactions. Entropy relates to all possible interactions available given an amount of energy available to a system. Thus, it can be said that the ever progressing "needle of time" acts to transform information into complexity. Information processing systems act to resolve entropy into information via pattern recognition. Unresolved entropy is noise.

F. Synchronization, Relevance, and Graph-based Reasoning using Virtual Neurons

Like organic neural networks, virtual neural networks can generate or "grow" new nodes. The act of growing a new virtual neuron is in the act of recognizing a new fruitful concept or area to measure. Virtual neurons are highly reconfigurable such that combinations and permutations of them can be rapidly tested by an intelligent system [32]. New concepts are formed and identified when a certain combination or permutation has observed utility due to having a certain fidelity in its reflection of the physical domain and alignment with an agent's goal set.

The virtual neural network within an organic brain spans from the highest layers of consciousness down to the deepest layers of subconscious. The most salient virtual neurons to an agent are on its mesoscale layer of conscious awareness. Resolving virtual neuron layers becomes more entropic at layer depth or layer height increases with reference to the mesoscale layer. IE it's challenging to resolve the receptive fields measured by the self's subconscious that influence its own behavior. Virtual neural networks are also formed by organized groups of agents. Mutual values across agents determine what common receptive fields are formed. These virtual neuron layers are defined to be in the social domain, the metascale layers. A collection of agents that operate as a group are referred to as collective agents. In both individual agents and collective agents there are intrinsic (core) values that are infrequently modified and values that are more susceptible to modification in response to environmental conditions. Virtual neural networks in the social domain of an agent become more entropic to resolve as social distance increases. Trust is the factor that determines social distance and is dependent on mutual information transfer in addition to resilience. Resilience is self-correcting stability such to maintain growth state by maintaining operation near the brain's critical point. Since every agent is submerged into it own feedback system with the environment, virtual neural networks become more entropic to resolve as layer height increases above the mesoscale layer into the layers of the social domain. The most resolvable virtual neurons in the social domain are where mutual values exist such that information transfer becomes beneficial to increasing resilience.

Every virtual neuron is "entangled" with a mental concept in the semantic domain. Elements in the semantic domain consist of objects, actions, rewards, states, assets, resources, liabilities, risk, etc; anything that can be defined or measured by an agent. A virtual neuron that can be modulated by an agent is an atomic unit of a task. Tasks are akin to the available actions, like the available levers described in the multi-armed bandits problem in reinforcement learning, and are executed when their correlated task pressure is high due to rate of change in resource levels [33][27]. An agent in this framework tracks goal relevant resources where each resource category is represented by a neuron and goal relevant resource quantifications are represented via a similar encoding mechanism to rate coding mechanisms in organic neurons where firing frequency is proportional to stimulus intensity [34]. Agent resource pressure is proportional to firing frequency of goal relevant resource tracking neurons. Higher resource pressure leads to agent decision pressure.

Tasks operations modulate resources to achieve desired goal states. Task formulation and coordination is facilitated by the neural property that enables reconfigurable and modular functions of semantic elements. The most intuitive example function is a reconfigurable superposition where elements within a set are selected to modulate a summation to output the desired value. Tasks are able to be performed via graph based path navigation and entails a reasoning process to find the most relevant and effective paths to take to achieve a goal state desired by a task. The reduction of cognitive-physical error that is involved with enabling effective task progress is synonymous with the act of reasoning.

An attribute of general intelligence is spontaneous cognitive-physical error reduction. (A spontaneous process is defined as a process that naturally occurs without application of an external force if a set of conditions occurs.) Cognitive-physical error reduction is the act of decreasing the difference between observed theoretical data and observed physical data; it is the cognitive process that aligns cognitive models with the physical world and aligns expectations with reality. To reduce cognitive-physical error is to explore relevant combinations and permutations of semantic elements such to create a theoretical model (cognitive model) that describes the expected results of operations performed on resources, then observe the result of each theoretical permutation and compare it to the relevant measured data of the physical domain, then act to guide subsequent permutations towards decreasing the difference between theoretical data and physical data. (A combination is a set. A permutation is an ordered set)

Reasoning involves guiding a layer of virtual neurons towards convergence to common points with another layer of virtual neurons via the mechanisms of network reconfiguration and neuron modulation. Convergence to a common point creates a point of symmetry. A single point or element of symmetry is defined as a similarity. To illustrate, consider the following example using virtual neurons to model the task of food preparation. A combination of ingredients is defined as the set of ingredients required for a chosen recipe. Each ingredient has specific heat absorption properties, thus each ingredient requires a different amount of time to be heated up to optimal serving temperature. A permutation is composed of the ordered set of ingredients for a recipe and the operations performed on the ingredients (IE time in oven, pot, etc at specified temperature) ordered by the operation times required for all the ingredients to reach optimal serving temperature at the same time. The point in time and space where different ingredients with different rates of change reach a common value is defined as a point of synchronization. Recipe outcome quality declines as realized (physical) virtual neuron synchronization deviates from required (theoretical) virtual neuron synchronization. This framework posits that consistent arrival at synchronization points across systems and or processes (which leads to periodicity) is one enabling component of emergence and abstraction. This is symmetric with ecological system synchronizations proposed in the theory of panarchy [35].

Since neurons exhibit the property of pattern completion and a preference for periodicity, the aspect of synchronization is inseparable from the goal space of reasoning, thus the capability of reasoning forms a base component of signal generation - information generation. Energy packets that don't have synchronization with other energy packets are referred to as noise. The phenomenon of object permanence and object property permanence is related to spontaneous neural pattern completion processes. The contrastive learning technique as proposed as a technique for scene understanding and long term prediction for the H-JEPA architecture is also a form of pattern completion [6]. Drawing from the research on critical point dynamics in organic brains, it seems that synchronization is not maximized, but balanced. The roles the synchronization plays in brains is still an open research that has grown in interest [36]. The branching ratio indicates the number of descendant neurons activated when a reference neuron is activated. Critical entropy for optimal signal propagation is achieved when the average number of descendant neurons activated in response to a reference neuron equals one; when the branching ratio equals one. Supercriticality is when the branching ratio is greater than one, which results in excess signal amplification. Subcriticality is when the branching ratio is less than one, which results in signal decay [23].

The terms that bear physical units in a mathematical model are analogous to receptive fields that measure the quantity and rates of change of physical elements (and non-physical elements). The set of ingredients within a mathematical model consists of the terms that reflect the physical world and the valid operations that are able to be performed on the set of ingredients. Physical experiments generate data that illustrate relationships between semantic terms and their influence on each other by demonstrating the effect of modulating the quantities of receptive fields and observing the effect on other receptive fields. The operations between terms represent relationships, transformations, dependencies, etc are arranged such that the computational results of a mathematical model match observed physical effects.

Semantic terms in the mind (cognitive domain) are represented as symbolic terms in mathematical models. Some of the goals in the utilization of mathematical model to configure a set of terms to optimize the value of a target term and ascertain the least complex semantic relations between target terms. A target receptive field is optimized with reference to a loss function by finding the correct configuration of receptive fields that interact with the target field. This process is commonly achieved via backpropagation. By these definitions, backpropagation in machine learning performs low level reasoning. A generative large language model interfacing with a compiler to correct its output in response to error messages about its generated code would be a higher magnitude of reasoning since the abstraction difference between the output and the components that are modulated to generate the output is higher. In the current state of generative language models, a key difference between an AI programmer and human programmer is that the human is much more precise in resolving the relevance of what code modifications to make in response to error messages from the compiler. This represents precision in the act of resolving the relevance component of information packets.

To constitute self-learning and self-programming, the most relevant tasks to perform need to be identified by an agent. This framework posits that this can be performed by resolving task relevance by feedback from the social domain or from any system or process the agent is interacting with. Information acts to resolve the most relevant effective paths through a graph network that corresponds to the most relevant combinations and permutations of virtual neurons and the sequence of operations to perform on them to progress toward the goal state, which in turn modulates a target receptive field toward the desired value. Elements with in a relevant scope attract an agent's attention, whereas irrelevant elements are masked. Masking is a form of filtering and is normally done via multiplication by zero or near-zero [37]. Resolving relevance results in following the least complex path toward a goal, which is the most effective available path.

G. Qbits for Certainty, Entropy, Probability, and Risk

Reasoning involves concurrently accounting for determined virtual neurons values and probabilistic values. Determined values are imparted by data. Probabilistic values are imparted by information. Deterministic virtual neurons are created from data. Entropic virtual neurons are created from inference to fill in gaps in data in knowledge. Receptive fields can have a range of values depending on the range of their activation function. Since virtual neurons are the neurons that compose cognitive models, virtual neurons can represent artificial neurons in a program, organic neurons, brain regions, receptive fields, or emergent behaviors. In the case that a virtual neuron is formed in cognitive space to represent an emergent behavior, the emergent activation function is entropic since is it not able to be measured deterministically. They may point an agent's attention towards where a deterministic system of interacts exists that an agent didn't previously recognize, so that can agent could then obtain data on the constituent phenomenon. Until emergent virtual neurons are linked to a deterministic process, they impart a degree of recognized uncertainty.

Virtual neurons with some degree of uncertainty in their value can be represented by qbits since they indicate a probability assigned in an agent's cognitive model imparted by information. Probability is the means to account for uncertainty and risk in decisions. These are agent-relevant probabilities rather than absolute or universal probabilities. If a receptive field takes on a value determined by physical measurement its dimension is reduced, thus its value can be represented by a non-quantum parameter. An unknown quantity introduces another dimension to account for in a cognitive model, represented by the range of possible value a parameter can take on. The integration operation performed by neurons by computing the summation of inputs from its neighbors results in dimensionality reduction by representing the values of multiple inputs with a single pulse or sum Dimensionality reduction is one of the mechanisms that achieves power efficacy organic brains [36]. Having physical data collapses the amount of combinations and permutations to test, thereby reducing mental energy to account for all possible scenarios. An agent uses data to generate information that resolves entropy by identifying what scenarios are possible and which are not possible or not relevant. In this framework, data is compressed in the neural engine to generate information which results in agents energy allocated to either kinetic energy for physical actions, cognitive-kinetic energy for planning (cognitive action), or cognitive energy for risk and opportunity evaluation (increasing cognitive resolution).

Since brains have a mostly fixed amount of processing power, having a large uncertainty in a population of neurons leads to wide bandwidth, low amplitude attention allocation, where many models and possibilities need to be entertained in order to recognize risk present. Having high certainty leads to high amplitude, low bandwidth attention allocation. Uncertainty and certainty are agent-relevant, not absolute. Thus, elucidates the importance of reasoning to reduce cognitive-physical error to reduce unrecognized risk. Risk that is unrecognized for significant amounts of time leads and agent to become unprepared to handle the related adverse event in the chance that it does occur. Thus, unrecognize risk has the potential to cause the most harm. The machine domain and social domain serve to reduce unrecognized risk given quality coordination and information transfer.

H. Information Quality Measured by Rate of Signal to Noise Ratio Change

From this, it can be seen that quality information leads to task progress, which has some correlation with the presence of synchronization points of task relevant neurons. This is reflected in the organic domain by the physical changes that occur in organic neural networks during the learning processes. During the learning process in organic brains, myelin sheath thickness is grown to increase signal propagation velocity between neurons. However, an important observation is that signal propagation velocity is not just maximized across neuron groups, it is modulated to increase synchronization between relevant neurons [39]. From a generalized perspective, it can be approximated that quality information generally coaxes neurons toward synchronization; which is a symmetry reflected from the phenomenon where information results from neuron synchronization. Currently, for this framework the optimal amount of information for an agent to receive or generate is defined as the amount of information the enables convergence towards criticality. Optimal convergence rate is still being investigated, but it can be surmised that there is a critical convergence rate, just like there is a distance from criticality. The critical point is where the number of down stream neurons activated by a single neuron is equal to one according to the branching ratio. This framework posits that the rate of signal-to-noise ratio increase is correlated to information quality and has some relation to the number of points of symmetry and or similarity across measured events.

In the scope of measuring non-deterministic probability of an entropic (projected) future, the conditional probability of a specified inferenced by ascertaining the frequency that a target event occurs in comparison to the frequency that the conditions occur. An event that occurs every time a set of conditions occur is said to be deterministically correlated with the set of conditions. In most real-world scenarios, most event-condition correlations that attract attention and are queried about within an agents cognitive space are those that are not fully deterministic so as long as the events are relevant to an agent's goal. This framework posits that low risk virtual neurons attract action, since they require less cognitive processing to estimate the outcomes of; there is a higher return on energy investment. This framework posits that every agent has a certainty threshold for the level of uncertainty they deem takes too much time and or energy to resolve to receive a return that is above the average return on the invested energy for of all the agent's perceived possible actions. Neuroscientifically, this relates to the balance of introversion and extroversion, which correlates to neural sensitivity to the environment. Its's been observed that individuals with higher introversion are more risk averse, and thereby dedicate more time to information processing where as higher extroversion correlates with higher risk acceptance [40]. An agent's policy in reinforcement learning is the systems of functions that an agent uses to predict the return on its energy investment for a set of actions, such to choose the action that provides the highest return on investment for a given set of conditions. Just as individuals of scientific nature consult multiple models to test hypotheses and approximate probabilities, an agent in this framework may utilize multiple relevant policies to ascertain the path of most effective action.

I. Virtual Neurons Summary

This framework proposes virtual neurons as the most fundamental building blocks of an agent's cognitive model of the world in which networks of virtual neurons generated by data interact in parallel with entropic virtual neurons generated by inference. Layers of virtual neurons are computationally represented by graph structures to enable path planning and task navigation. In order to build comprehensive world models of the environment, an agent must make inferences to fill in gaps in data and make low error hypotheses to fill in gaps in concrete knowledge in order to build a tractable goal space to enable efficient decision making and coordination with other agents. Virtual neurons are the mediating particles of the an agent's interaction with the information field, thus virtual neurons function to represent the set of an agent's available receptive fields that span multiple layers of abstraction.

J. Consciousness Definition Disambiguation

This framework defines consciousness as the ability to sense the environment, thus the term consciousness is does not impart the nature of the substrate that it arises from. Therefore, in this framework both machine agents and organic agents are able to have consciousness. There are some important distinctions to make in this matter. The substrate of consciousness in the organic domain is composed of organic neurons and larger physiological body of organisms while the substrate of consciousness in the machine domain is composed of electrical and mechanical sensors, transistors, and other components in physical hardware.

An organic substrate must sustain molecular balance, reproduction, and growth in order to maintain functionality. The cost of molecular imbalance to an organic agent involves the advent of adverse affective states such as depression, anxiety, and others. The cost of exposure to harmful environmental conditions may be harm or suffering. These are categorized as decay states. The mental and physical state of an organic agent is sensitive to its growth state and prevention of decay [41].

Machine agents, by definition, don't have a physiological substrate that affective states can arise from. Machines don't experience pleasure nor pain and haven't been observed to be capable of becoming intrinsically motivated to sustain the health of the physical hardware that underpins their existence. In the far future, if a machine would exhibit the impetus of self-repair without explicitly being programmed to do so, it would be of importance to investigate if it's activity bears conflicts of interest with organic agents, and is even more important to determine prior to compiling the machine agent if there is any chance self-repair or self-replication can arise as an emergent result of reward hacking. This matter and many others are the subject of much AGI safety research [4].

V. Results

This project develops a framework for computational general intelligence that can be used to build computational models of decision making for reinforcement learning agents that operate in the real world. Virtual neuron networks are the main tools in this framework to enable the development of reconfigurable functions composed of receptive fields, graph-based path planning for task navigation, and long term planning. To achieve this, this framework proposes a novel model of information in the environment, information in cognitive domain and response to environment, critical neural balance, information transfer, entropy, and complexity.

VI. Future Research

A machine agent's primary role is to support the growth state of organic agents. Whether through software replication or hardware replication, it is important to ensure that machines don't replicate in ways that have conflicts of interest with the growth and balance of organic life. While there are many systems of principles to reason over to determine a machine's utility, in order to be fully aligned with long term human growth, it is necessary to enable feedback from the organic domain. Feedback from the organic domain enables inference of affective states of a host human user and can be facilitated by operation on stored user data or by real-time user interaction via chat, body language recognition, or EEG data stream [42]. Feedback from the organic domain enables a reward function to be developed that is based on the affective states of a user. It can be surmised that having the base components of a machine agent's reward function rooted in organic metrics would lessen the potential for reward hacking, although the avenue of influencing or deceiving a user to take an action to serve the machine agent's purpose would still exist as this issue is complex to safeguard against. This and many other areas of AGI risk require more rigorous future research [43].

In this light, making progress in human-AGI alignment has a dependency on progress in ensuring user data privacy, user data ownership, and in cyber security. AI technologies have the potential to cause high rates of change to social and economic conditions. Thus, while AI technology may still progress independently of these areas, it is important to ensure the change is aligned with longer term economic and ecological stability [44]. Suboptimal alignment with longer term user goals and interests has been seen in social media content recommendation engines, which exhibit a tendency to encourage dopaminergic engagement for short term satisfaction [45]. Progress toward human-aligned reinforcement learning based recommendation systems is one of the key areas enabled by this framework.

Providing further resolution on the information field theory of this framework, which defines how virtual neurons interact is the planned next step in this research.

Acknowledgment

I would like to express my gratitude to Dr. Briggs for his guidance throughout the course of this research, and for his dedication and commitment to fostering an intellectually stimulating environment that has promoted much academic growth. His knowledge and insights delivered through balanced feedback and discussion added much value to the research process during this project.

References

[1] X. Wang et al., "Deep Reinforcement Learning: A Survey," in IEEE Transactions on Neural Networks and Learning Systems, 2022, doi: 10.1109/TNNLS.2022.3207346.

[2] Kotseruba, I., Tsotsos, J.K. 40 years of cognitive architectures: core cognitive abilities and practical applications. Artif Intell Rev 53, 17–94 (2020). https://doi.org/10.1007/s10462-018-9646-y

[3] X. Song and Z. Lin, Non-zero-sum Game Control for Multi-vehicle Driving via Reinforcement Learning.

[4] R. Ngo, L. Chan, and S. Mindermann, The Alignment Problem from a Deep Learning Perspective. [Online]. Available: https://arxiv.org/pdf/2209.00626.pdf

[5] S. Bubeck, E. Price, and I. Razenshteyn, Adversarial examples from computational constraints. [Online]. Available: https://arxiv.org/pdf/1805.10204.pdf

[6] Y. LeCun, A Path Towards Autonomous Machine Intelligence, Jun. 2022. [Online]. Available: https://openreview.net/pdf?id=BZ5a1r-kVsf

[7] C. Tsallis, The Nonadditive Entropy Sq and Its Applications in Physics and Elsewhere: Some Remarks. [Online]. Available: https://www.mdpi.com/1099-4300/13/10/1765

[8] M. Bal, “Deep implicit attention: A mean-field theory perspective on attention mechanisms,” mcbal, https://mcbal.github.io/post/deep-implicit-attention-a-mean-field-theory-perspective-on-attention-mechanisms/ (accessed Jun. 16, 2023).

[9] E. Plantec et al., Flow-Lenia: Towards open-ended evolution in cellular automata through mass conservation and parameter localization. [Online]. Available: https://arxiv.org/pdf/2212.07906.pdf

[10] G. A. Silva, P. Y. Wang, S. Sapra, and V. K. George, Generalizable Machine Learning in Neuroscience Using Graph Neural Networks, Feb. 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2021.618372/full

[11] C. von der Malsburg, T. Stadelmann, and B. F. Grewe, A Theory of Natural Intelligence, Apr. 2022. [Online]. Available: https://arxiv.org/pdf/2205.00002.pdf

[12] Zador, A., Escola, S., Richards, B. et al. Catalyzing next-generation Artificial Intelligence through NeuroAI. Nat Commun 14, 1597 (2023). https://doi.org/10.1038/s41467-023-37180-x

[13] M. Tegmark, The Mathematical Universe, Oct. 2007. [Online]. Available: https://arxiv.org/pdf/0704.0646.pdf

[14] M. R. Myle Ott, “Unsupervised machine translation: A novel approach to provide fast, accurate translations for more languages,” Engineering at Meta, https://engineering.fb.com/2018/08/31/ai-research/unsupervised-machine-translation-a-novel-approach-to-provide-fast-accurate-translations-for-more-languages/ (accessed Jun. 16, 2023).

[15] J. Tarbouriech et al., Adaptive Multi-Goal Exploration. Accessed: Jun. 16, 2023. [Online]. Available: https://arxiv.org/pdf/2111.12045.pdf

[16] B. Essid, “Active learning strategies: Pros and cons,” Medium, https://medium.com/mlearning-ai/active-learning-strategies-pros-and-cons-802e010302d (accessed Jun. 16, 2023).

[17] P. Villalobos∗ et al., Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning. [Online]. Available: https://arxiv.org/pdf/2211.04325.pdf

[18] Anders S Hansen, Erin K O'Shea (2015) Limits on information transduction through amplitude and frequency regulation of transcription factor activity eLife 4:e06559 https://doi.org/10.7554/eLife.06559

[19] E. P. Lazear, “The peter principle: A theory of decline,” SSRN Electronic Journal, 2003. doi:10.2139/ssrn.403880

[20] W. Tenachi, R. Ibata, and F. I. Diakogiannis, Deep symbolic regression for physics guided by units constraints: toward the automated discovery of physical laws. [Online]. Available: https://arxiv.org/pdf/2303.03192.pdf

[20] J. B. Glattfelder, “The semantics of symmetry, invariance, and structure,” Information—Consciousness—Reality, pp. 65–92, 2019. doi:10.1007/978-3-030-03633-1_3

[21] C. van der Ploeg, M. Braat, B. Masin, J. Brouwer, and J.-P. Paardekooper1, Connecting the Dots: Context-Driven Motion Planning Using Symbolic Reasoning. [Online]. Available: https://arxiv.org/pdf/2303.04191.pdf

[22] J. M. Beggs, The Cortex and the Critical Point: Understanding the Power of Emergence. Cambridge: MIT Press, 2022.

[23] Z. Pizlo and J. A. de Barros, “The concept of symmetry and the theory of perception,” Frontiers in Computational Neuroscience, vol. 15, 2021. doi:10.3389/fncom.2021.681162

[24] B. J. Kagan et al., In vitro neurons learn and exhibit sentience when embodied in a simulated game-world, 2021. doi:10.1101/2021.12.02.471005

[25] C. D. Gilbert and M. Sigman, “Brain states: Top-down influences in sensory processing,” Neuron, vol. 54, no. 5, pp. 677–696, May 2007. doi:10.1016/j.neuron.2007.05.019

[26] R. S. Sutton and A. G. Barto, Cambridge, MA: The MIT Press, 2020.

[27] J. Gottlieb, “Attention, learning, and the value of information,” Neuron, vol. 76, no. 2, pp. 281–295, 2012. doi:10.1016/j.neuron.2012.09.034

[28] M. Jepma, R. G. Verdonschot, H. van Steenbergen, S. A. Rombouts, and S. Nieuwenhuis, “Neural mechanisms underlying the induction and relief of Perceptual Curiosity,” Frontiers in Behavioral Neuroscience, vol. 6, 2012. doi:10.3389/fnbeh.2012.00005

[29] G. Stix, “Why language and thought resemble Russian dolls,” Scientific American Blog Network, https://blogs.scientificamerican.com/talking-back/why-language-and-thought-resemble-russian-dolls/.

[30] A. A. Team, Human-Timescale Adaptation in an Open-Ended Task Space. [Online]. Available: https://arxiv.org/pdf/2301.07608.pdf

[31] Optimize Your Learning & Creativity with Science-Based Tools | Huberman Lab Podcast. 2023.

[32] C. M. Wu, E. Schulz, T. J. Pleskac, and M. Speekenbrink, “Time pressure changes how people explore and respond to uncertainty,” Scientific Reports, vol. 12, no. 1, 2022. doi:10.1038/s41598-022-07901-1

[33] W. Guo, M. E. Fouda, A. M. Eltawil, and K. N. Salama, “Neural coding in spiking neural networks: A comparative study for robust neuromorphic systems,” Frontiers in Neuroscience, vol. 15, 2021. doi:10.3389/fnins.2021.638474

[34] J. Rocha et al., Panarchy: ripples of a boundary concept. [Online]. Available: https://arxiv.org/pdf/2012.14312.pdf

[35] S. Kazemi and Y. Jamali, “Phase synchronization and measure of criticality in a network of neural mass models,” Scientific Reports, vol. 12, no. 1, 2022. doi:10.1038/s41598-022-05285-w

[36] Goi, E., Zhang, Q., Chen, X. et al. Perspective on photonic memristive neuromorphic computing. PhotoniX 1, 3 (2020). https://doi.org/10.1186/s43074-020-0001-6

[37] S. Rush, The annotated Transformer - Harvard University, http://nlp.seas.harvard.edu/annotated-transformer/.

[38] R. D. Fields, “The brain learns in unexpected ways,” Scientific American, https://www.scientificamerican.com/article/the-brain-learns-in-unexpected-ways/.

[39] N. Nicholson, E. Soane, M. Fenton-O’Creevy, and P. Willman, “Personality and domain-specific risk taking,” Journal of Risk Research, vol. 8, no. 2, pp. 157–176, 2005. doi:10.1080/1366987032000123856

[40] B. J. Kagan, D. Duc, I. Stevens, and F. Gilbert, “Neurons embodied in a virtual world: Evidence for organoid ethics?,” AJOB Neuroscience, vol. 13, no. 2, pp. 114–117, 2022. doi:10.1080/21507740.2022.2048731

[41] S. Moontaha, F. E. Schumann, and B. Arnrich, “Online learning for wearable EEG-based emotion classification,” Sensors, vol. 23, no. 5, p. 2387, 2023. doi:10.3390/s23052387

[42] D. Hendrycks and M. Mazeika, X-Risk Analysis for AI Research. [Online]. Available: https://arxiv.org/pdf/2206.05862.pdf

[43] K. J. Friston et al., Designing Ecosystems of Intelligence from First Principles. [Online]. Available: https://arxiv.org/pdf/2212.01354.pdf

[44] J. Ferrer, “Scrolling for dopamine: The impact of social media on our lives,” Medium, https://uxdesign.cc/scrolling-for-dopamine-the-impact-of-social-media-on-our-brains-ux-ui-addiction-design-psychology-phone-app-e6a1786d9d3b