The Road Towards Artificial General Intelligence (AGI), Architectural Approaches to Reach Human-Level Intelligence (HLI)

Dr. Edward H. Ghafari

Research & Development Group

iCES Corporation | Systems Thinking @ Work

March, 31, 2025

ORCID 0009-0008-1800-0806

iCES Corporation, Research and Development Group

edward.ghafari@ices-corporation.com

edward.ghafari@gmail.com

https://www.linkedin.com/in/eghafari

https://www.ices-corporation.com

iCES Corporation is a premier Artificial Intelligence and Engineering firm specializing in Cybersecurity, Data Science, Workflow Automation, and Systems Engineering. With a foundation in cutting-edge research, advanced engineering, and complex systems thinking, we deliver mission-critical solutions that empower Government agencies and commercial enterprises to excel ahead of evolving challenges.

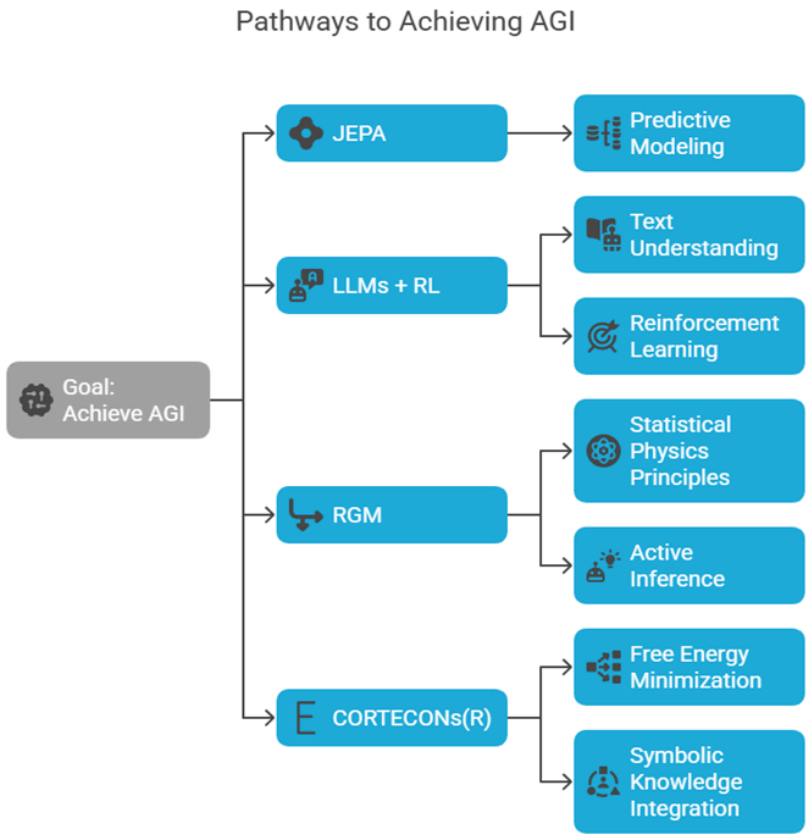
Founded in 2007, iCES applies systems thinking to deliver next-generation solutions to complex challenges and empowers organizations to achieve their mission-critical goals.

Abstract

This report, titled "The Road Towards Artificial General Intelligence (AGI), Architectural Approaches to Reach Human-Level Intelligence (HLI)," delves into four prominent architectural paradigms proposed as potential pathways towards AGI: the Joint Embedding Predictive Architecture (JEPA), the combination of Large Language Models with Reinforcement Learning (LLMs + RL), Recurrent Generative Models (RGM), and CORTECONs(R). Achieving AGI, defined as machines with human-level cognitive abilities across a broad spectrum of intellectual tasks, necessitates exploring diverse architectural approaches. This report provides an in-depth analysis of each architecture's core principles, capabilities, strengths, weaknesses, and recent advancements, culminating in a comparative discussion of their potential to reach the ultimate goal of AGI. JEPA focuses on learning abstract representations of the world through self-supervised predictive modeling. LLMs + RL leverages the power of large language models for understanding and generating text, combined with reinforcement learning to guide behavior and decision-making. RGM employs generative modeling based on principles from statistical physics and active inference to learn complex dynamics and enable planning. Finally, CORTECONs(R) aims to achieve AGI through free energy minimization in a latent layer and integration with symbolic knowledge representations.

1. Introduction

Artificial General Intelligence (AGI) represents the aspiration to create machines with human-level cognitive abilities across a broad spectrum of intellectual tasks [1]. This ambition extends beyond the current state of artificial intelligence, where systems typically excel in narrow, well-defined domains. Achieving AGI necessitates exploring diverse architectural approaches that can replicate the complexity and adaptability of human intelligence [3]. This report delves into four prominent architectural paradigms proposed as potential pathways towards AGI: the Joint Embedding Predictive Architecture (JEPA), the combination of Large Language Models with Reinforcement Learning (LLMs + RL), Recurrent Generative Models (RGM), and CORTECONs(R). Each of these approaches offers a unique perspective on how to build a system capable of general intelligence. JEPA focuses on learning abstract representations of the world through predictive modeling in a self-supervised manner [5]. LLMs + RL leverages the power of large language models for understanding and generating text, combined with reinforcement learning to guide behavior and decision-making [7]. RGM employs generative modeling based on principles from statistical physics and active inference to learn complex dynamics and enable planning [8]. Finally, CORTECONs(R) aims to achieve AGI through free energy minimization in a latent layer and integration with symbolic knowledge representations [10]. This report will provide an in-depth analysis of each architecture's core principles, capabilities, strengths, weaknesses, and recent advancements, culminating in a comparative discussion of their potential to reach the ultimate goal of AGI.



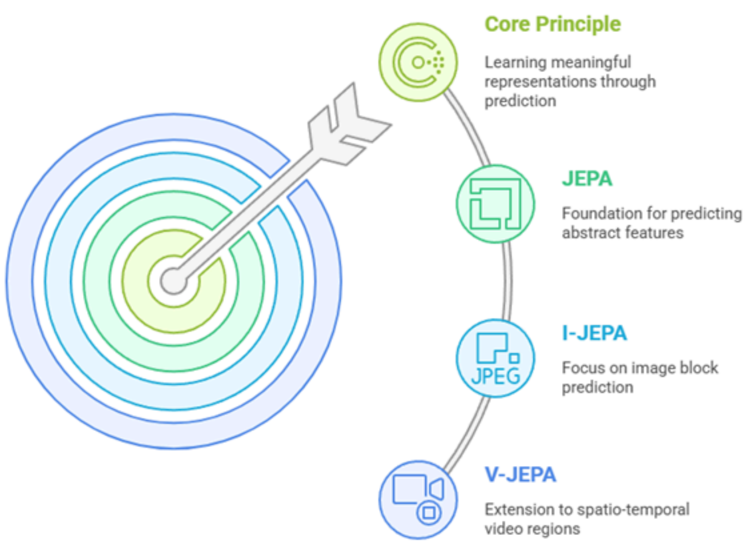
2. Joint Embedding Predictive Architecture (JEPA)

Core Principles

The fundamental principle behind the Joint Embedding Predictive Architecture (JEPA) is to learn by predicting abstract representations of different parts of an input from a given context [5]. This process involves transforming inputs like images or videos into lower-dimensional embeddings that capture their essential features [5]. Encoders are used to create these embeddings, and a predictor module is then trained to forecast the embedding of a target portion of the input based on the embedding of a context portion [5]. Yann LeCun, who proposed JEPA, envisions this architecture enabling AI systems to develop comprehensive internal models of the world, facilitating their ability to predict future events, reason about situations, and plan actions in a self-supervised fashion, much like humans do [5]. This learning from observation, without the need for extensive labeled data, is a key aspect of JEPA's design [5].

The Image-based Joint Embedding Predictive Architecture (I-JEPA) specifically focuses on images and employs a masking strategy [6]. The model is trained to predict the representations of hidden image blocks (target blocks) using the information from visible context blocks [6]. This approach encourages the learning of semantic relationships between different parts of an image rather than focusing on low-level pixel reconstruction [13]. Video Joint Embedding Predictive Architecture (V-JEPA) extends these principles to video data by predicting masked spatio-temporal regions in a learned latent space [6]. This allows the model to learn about the dynamics of motion and the interactions between objects over time [6]. The core idea across these variations is to learn meaningful representations by predicting abstract features, which can then be used for various downstream tasks.

Hierarchy of Joint Embedding Predictive Architectures



Generative Capabilities

The original JEPA architecture was primarily designed for learning representations through prediction and did not inherently possess the capability to generate new data samples [5]. Its focus was more on understanding the relationships within existing data rather than creating novel instances [14]. However, recent advancements have sought to address this limitation. D-JEPA, for instance, integrates diffusion models with the JEPA framework, enabling the generation of data in a continuous space [17]. This approach reinterprets JEPA as a form of masked image modeling, which allows for autoregressive generation of data [17]. Notably, D-JEPA has demonstrated strong performance in image generation tasks, achieving competitive results on benchmarks like ImageNet [17]. This development signifies a crucial expansion of JEPA's potential role in achieving AGI, as the ability to generate realistic data is often considered an important aspect of a comprehensive understanding of the world.

World Model Development

A central aim of the JEPA architecture is to facilitate the development of robust world models within AI systems [5]. These world models are intended to be internal representations of how the world functions, enabling the AI to anticipate future events and understand the outcomes of actions without needing to physically perform them [5]. The JEPA framework often includes a dedicated "World Model Module" that uses current and past data to predict future states and even fill in missing information about the present state, acting as a kind of simulator [6]. In the context of video, V-JEPA's predictor is considered an early form of a physical world model, capable of understanding detailed interactions between objects and their motion over time [16]. Furthermore, Action-Conditioned JEPA (ACT-JEPA) integrates JEPA with imitation learning to enhance the learning of policy representations by predicting abstract observation sequences, thereby contributing to a more comprehensive world model for decision-making tasks [18]. The emphasis on predictive learning across different modalities positions JEPA as a strong contender for building AI systems with a nuanced understanding of the world's dynamics.

Strengths

JEPA exhibits several key strengths that make it a promising architecture for AGI. One of its primary advantages is its ability to learn abstract representations by focusing on the essential semantic information in the input while discarding irrelevant details [6]. This leads to more efficient processing and improved generalization capabilities. Furthermore, JEPA leverages self-supervised learning from unlabeled data, significantly reducing the need for extensive, costly labeled datasets [5]. This is particularly beneficial for scaling AI to handle the vast amounts of raw sensory data available in the real world. I-JEPA has demonstrated remarkable efficiency in vision tasks, requiring considerably less computational resources and training time compared to other self-supervised methods for pre-training on large datasets like ImageNet [13]. Similarly, V-JEPA shows efficiency in learning video representations, achieving strong performance with shorter training durations and without relying on pre-trained encoders or textual supervision [15]. Recent advancements, such as Contrastive-JEPA (C-JEPA), further enhance the stability and quality of the learned visual representations [19].

Weaknesses

Despite its strengths, JEPA also has certain weaknesses. Training JEPA models can be a complex undertaking, requiring careful architectural design and training methodologies [prompt]. The original JEPA lacked an explicit mechanism for generating new data, which is an important aspect of modeling the probabilistic nature of the world [5]. While D-JEPA has addressed this, the complexity of this new approach requires consideration. Recent research has identified a limitation in JEPA's ability to grasp local semantics for dense representations, which arises from the masked modeling approach in the embedding space [23]. This can result in a reduced ability to capture fine-grained local semantic information. Specifically, in I-JEPA, the Exponential Moving Average (EMA) used for updating the target encoder has been found to be inadequate in preventing complete model collapse, and its prediction mechanism struggles to accurately learn the mean of patch representations [19].

Recent Research and Progress

Ongoing research on JEPA is actively working to overcome its identified limitations and expand its applicability. Contrastive-JEPA (C-JEPA) integrates Variance-Invariance-Covariance Regularization (VICReg) with I-JEPA to tackle the issues of EMA collapse and improve the learning of mean patch representations, demonstrating faster and better convergence in visual representation learning [19]. DMT-JEPA introduces a novel masked modeling objective aimed at generating discriminative latent targets from neighboring information, thereby enhancing the understanding of local semantics [24]. Point-JEPA adapts the JEPA architecture for self-supervised learning on point cloud data, achieving state-of-the-art results in 3D shape classification [25]. ACT-JEPA explores the integration of JEPA with imitation learning to improve the learning of efficient representations for decision-making policies in reinforcement learning scenarios [18]. V-JEPA remains a significant focus of research, with ongoing efforts to explore its capabilities in video understanding and its potential integration into future large language model architectures [1]. Additionally, the development of D-JEPA showcases the potential of JEPA for generative modeling, with promising outcomes in image generation tasks [17].

Representational Learning

JEPA employs abstract embeddings for representational learning, which are designed to capture the core semantic content of the input data while discarding superficial variations [5]. This is achieved through the predictive process, where the model learns to forecast representations of masked or future data points based on the available context. These abstract embeddings aim to be invariant to irrelevant details, focusing instead on the underlying structure and relationships within the data [6]. For instance, V-JEPA learns representations in an abstract feature space that captures motion and object interactions in video rather than at pixel level [15]. Using such abstract embeddings facilitates efficient task adaptation and generalization, as the model learns high-level features transferable across different but related tasks [5]. The strength of these abstract embeddings lies in their ability to distill complex sensory input into a more manageable and semantically rich format, promoting the learning of both invariant and equivariant features crucial for robust perception and reasoning in dynamic environments [6].

Reasoning Mechanisms

The primary reasoning mechanism in JEPA is based on its ability to predict future states or missing information [5]. By learning a world model, the system can anticipate the consequences of actions or infer unseen parts of the environment. I-JEPA demonstrates spatial reasoning through its ability to predict representations of masked regions within an image based on the visible context, indicating an understanding of spatial relationships [5]. V-JEPA extends this to spatiotemporal reasoning by predicting future frames or parts of frames in a video, showcasing an understanding of temporal dynamics and motion [6]. This predictive capability forms the foundation of JEPA's reasoning abilities, allowing it to anticipate future events and infer missing information across space and time, thus developing an understanding of the world's dynamics that can be leveraged for planning and decision-making.

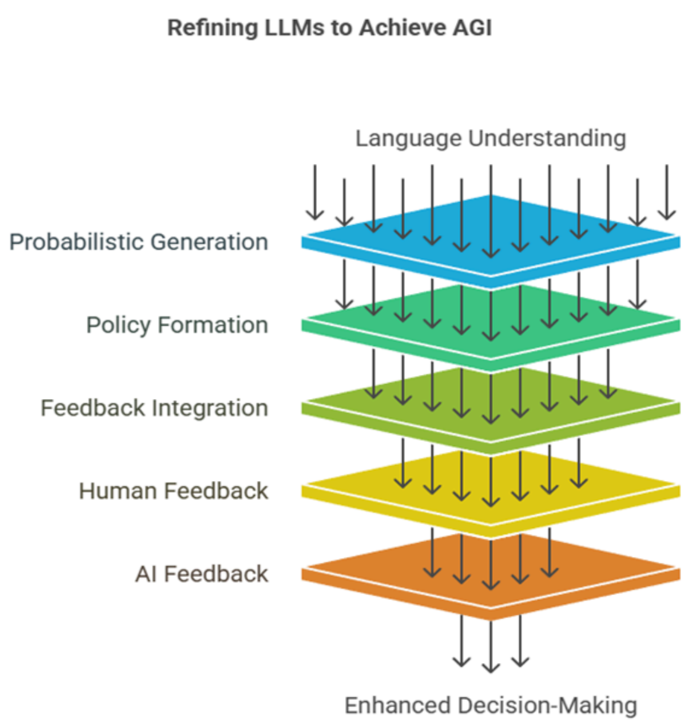
Generalization Capabilities

JEPA aims to learn strong, off-the-shelf representations that can be efficiently adapted to various downstream tasks with minimal fine-tuning [5]. The abstract nature of the learned embeddings contributes to this transferability. V-JEPA has shown impressive generalization capabilities across different video and image tasks, often outperforming other methods in frozen evaluations, which means without task-specific fine-tuning [15]. This suggests that the features learned by V-JEPA capture general visual principles. The efficiency of task adaptation is a significant advantage for AGI, as it reduces the need for extensive task-specific training and allows for quicker learning of new skills. This ability to learn generalizable representations that can be readily applied to new tasks indicates JEPA's potential to learn fundamental aspects of the world that are relevant across different contexts, reducing the need for learning from scratch for every new problem.

3. Large Language Models + Reinforcement Learning (LLMs + RL)

Core Principles

The approach of combining Large Language Models (LLMs) with Reinforcement Learning (RL) for achieving AGI centers on leveraging the exceptional language understanding and generation capabilities of LLMs and then refining their behavior and decision-making processes through RL [7]. LLMs operate based on the principle of probabilistic language generation, where they predict the next token in a sequence given the preceding context, having learned statistical relationships between words from vast amounts of text data [3]. Reinforcement Learning is integrated to align the outputs of LLMs with desired behaviors and human preferences by providing feedback in the form of rewards, allowing them to improve their performance beyond simply mimicking the patterns in their training data [7]. Key techniques in this integration include Reinforcement Learning from Human Feedback (RLHF), where a reward model is trained based on human preferences to guide the LLM's fine-tuning, and Reinforcement Learning from AI Feedback (RLAIF), which utilizes AI systems to provide feedback, potentially offering scalability and consistency benefits [7]. In an RL framework applied to LLMs, the LLM itself acts as the policy, the current text sequence represents the state, and the generated next token is considered the action, with the quality of the complete textual sequence determining the reward signal [7].



Generative Capabilities

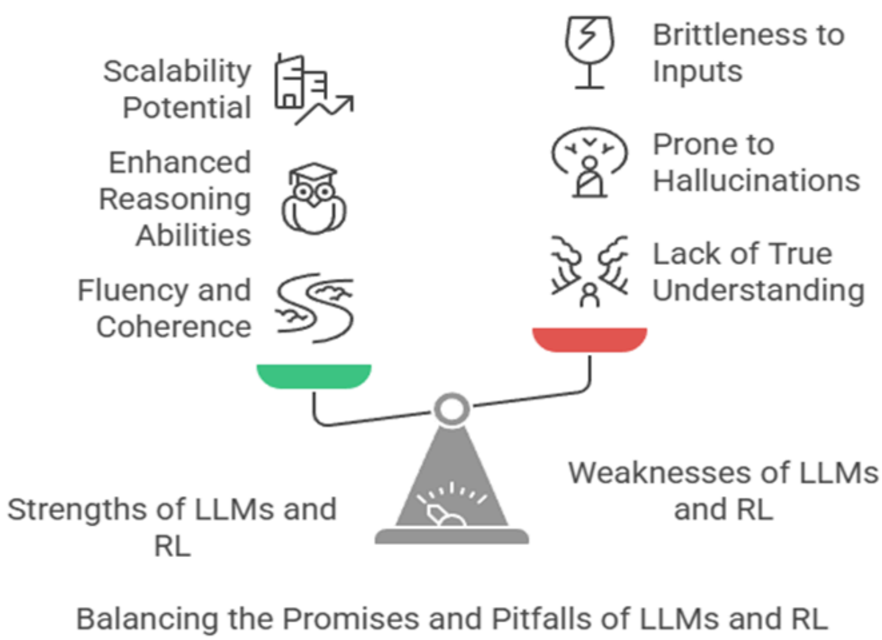
Large Language Models are inherently generative, possessing the ability to produce human-like text that is coherent and contextually relevant across a broad range of topics and styles [3]. They can generate various forms of text, including answers to questions, creative writing, and even computer code [30]. Recent advancements have led to the development of Multimodal Large Language Models (MLLMs) that can process and generate information across different modalities, such as text, images, and audio, expanding their generative capabilities beyond just language [4]. The integration of Reinforcement Learning can further enhance these generative capabilities by encouraging exploration and the generation of novel and creative outputs that align with specific reward signals, allowing LLMs to adapt their generation style based on feedback and specific requirements [7].

World Model

LLMs acquire a significant amount of world knowledge by being trained on massive text datasets, which contain information about a wide array of topics and concepts [3]. This knowledge is embedded within the model's parameters, enabling it to answer questions, provide explanations, and engage in conversations about the world. However, the world knowledge of LLMs is primarily derived from linguistic information and lacks direct sensory experience or grounding in the physical world [2]. They understand the world through the lens of language and may struggle with tasks that require common-sense reasoning or a physical understanding of how things work [12]. To address this, researchers are exploring methods to improve the grounding of LLMs by integrating them with knowledge graphs, ontologies, and environments where they can interact and receive feedback [4]. Reinforcement learning plays a crucial role in allowing LLMs to learn from interactions with simulated or real-world environments, potentially bridging the gap between linguistic knowledge and a more embodied understanding [30].

Strengths

The combination of LLMs and RL offers several compelling strengths for achieving AGI. LLMs possess exceptional abilities in natural language processing and generation, surpassing previous AI systems in fluency, coherence, and context-sensitivity [3]. They also demonstrate reasoning abilities, particularly in tasks that can be framed within the language domain, and these abilities can be significantly enhanced through techniques like Chain of Thought prompting and reinforcement learning [12]. Recent research indicates substantial progress in improving the reasoning capabilities of LLMs through advanced RL techniques [41]. Furthermore, LLMs can be scaled to extremely large sizes with billions of parameters, which generally leads to improved performance across a wide range of tasks [37]. Reinforcement learning fine-tuning provides a powerful mechanism for aligning LLMs with human preferences and desired behaviors, making their outputs more useful and reliable [7].



Weaknesses

Despite their impressive capabilities, LLMs combined with RL also exhibit several weaknesses that need to be addressed for AGI. A fundamental limitation is their lack of true understanding; they often operate based on statistical correlations rather than genuine comprehension of the concepts they process [2]. This can lead to a lack of common sense, resulting in nonsensical or illogical outputs that a human would easily identify as incorrect [2]. LLMs are also prone to generating hallucinations, which are factually incorrect or nonsensical statements presented as truthful information [2]. Most current LLMs are pre-trained on static datasets and lack the ability to continuously learn and adapt in real-time during inference, which is a key characteristic of general intelligence [2]. They can also be brittle and sensitive to adversarial inputs, producing unreliable outputs when faced with subtly malformed prompts [32]. Compared to human intelligence, LLMs often have limitations in complex multi-step reasoning, world modeling, and theory of mind [45].

Recent Research and Progress

Recent research in the field of LLMs and RL for AGI has shown significant advancements. The development of models like DeepSeek-R1 demonstrates enhanced reasoning capabilities achieved through reinforcement learning, even surpassing the performance of state-of-the-art closed-source models on challenging reasoning benchmarks [41]. This includes the model's ability to perform self-verification and generate long chains of thought. Researchers are also exploring the use of pure reinforcement learning techniques to incentivize reasoning in LLMs without relying on supervised fine-tuning data, with promising initial results [41]. Furthermore, there is ongoing work on distilling the reasoning abilities of larger LLMs into smaller, more efficient models [41]. However, some research argues that LLMs, even with continued scaling and reinforcement learning, may not represent the most direct path to achieving true AGI due to inherent architectural and training limitations, suggesting the need for exploring alternative or complementary approaches [2].

Representational Learning

LLMs primarily utilize token embeddings for representational learning [5]. These are dense vector representations learned from the vast amounts of text data the models are trained on, where each token (word or sub-word unit) is mapped to a vector in a high-dimensional space. These embeddings capture semantic relationships between words based on their co-occurrence patterns in the training text, allowing the LLM to understand the meaning of words in context and generate semantically coherent text. However, these representations are fundamentally linguistic and may not fully capture the richness of real-world concepts or the nuances of embodied experience [37]. While Multimodal LLMs extend their representational scope to include other modalities like images and audio, the underlying representations often involve mapping these different types of data to a similar embedding space, which might not fully encapsulate the unique characteristics of non-linguistic information.

Reasoning Mechanisms

Reasoning in LLMs primarily involves performing inference based on the patterns and relationships they have learned from language data [5]. Techniques such as Chain of Thought prompting encourage the model to generate intermediate reasoning steps in natural language, which can improve its ability to solve complex problems that require multiple steps of inference

[12]. Reinforcement learning can further enhance the reasoning abilities of LLMs by rewarding them for producing correct or logical outputs, thus incentivizing the learning of effective reasoning strategies [31]. However, LLMs often struggle with tasks that require long-term planning or reasoning that goes beyond the patterns explicitly present in their training data [11]. Their planning capabilities are often limited to predicting the next step in a sequence rather than formulating and executing complex, multi-stage plans that might involve physical constraints or long-term consequences.

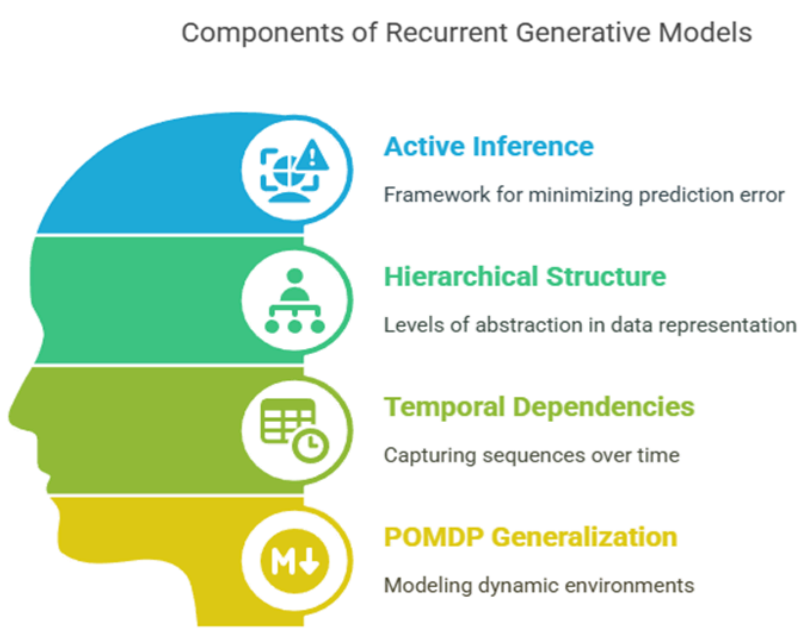
Generalization Capabilities

The generalization capabilities of LLMs are strongest within the domain of language and tasks that are closely related to their training data [2]. They can often adapt to new language-based tasks with some degree of fine-tuning. However, their ability to generalize to tasks that fall outside the linguistic domain or to truly novel situations that require embodied intelligence or common-sense reasoning about the physical world is limited [29]. While Multimodal LLMs show promise in generalizing across different types of data, their generalization abilities are still largely contingent on the types of data they have been trained on and the relationships they have learned between those modalities. Achieving broad generalization, a key characteristic of AGI, remains a significant challenge for LLM-based approaches.

4. Recurrent Generative Models (RGM)

Core Principles

Recurrent Generative Models (RGMs) represent a class of probabilistic models that learn to generate data by modeling the underlying probability distribution of the training data, often utilizing recurrent neural networks to capture temporal dependencies in sequential data [9]. A key principle often associated with RGMs in the context of AGI is active inference, a theoretical framework suggesting that intelligent agents act to minimize their prediction error or "free energy" [8]. This involves agents actively inferring the causes of their sensations and acting in a way that aligns with their internal model of the world. RGMs can be structured hierarchically, enabling them to learn and represent complex data at different levels of abstraction and across various temporal scales [9]. This hierarchical structure is considered crucial for facilitating sophisticated planning and reasoning. Furthermore, RGMs often generalize Partially Observed Markov Decision Processes (POMDPs) by incorporating "paths" as latent variables, making them well-suited for modeling dynamic environments and learning in sequential decision-making tasks [9].



Generative Capabilities

As generative models, RGMs are inherently capable of creating new data samples that statistically resemble the training data they were exposed to [9]. This includes the generation of diverse data types such as images, videos, music, and interactions within game environments [9]. By learning the underlying probability distributions of complex datasets, RGMs can produce novel instances that adhere to these learned distributions [47]. For example, an RGM trained on a dataset of handwritten digits can generate new digits that resemble the original ones [51]. Recent advancements in this area include Renormalizing Generative Models (RGMs), which have demonstrated the potential to generate high-quality data with possibly less training data and lower energy consumption compared to traditional deep learning-based generative models [51].

World Model

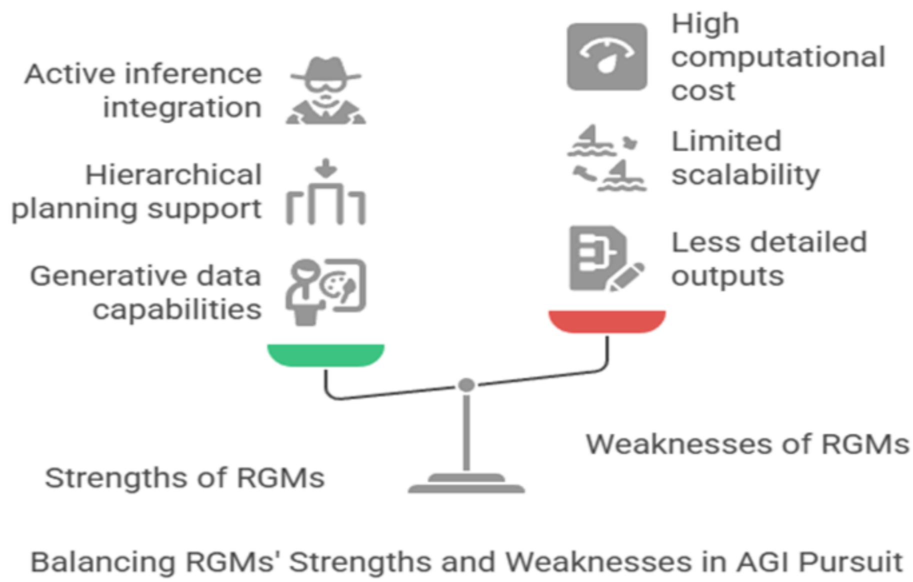
RGMs often develop world models through the use of state-space representations [9]. A state-space model mathematically defines all the possible states a system can be in, along with the probabilities of transitioning between these states [51]. By learning these states and transitions from data, RGMs can predict future states of the environment based on current observations, effectively acting as simulators of the world [51]. The inclusion of "paths" as latent variables in some RGMs allows them to infer the underlying causes of observable outcomes over different scales of space and time, leading to a more nuanced understanding of the world's dynamics [51]. This capability is crucial for AI systems to perform hypothetical reasoning and planning, essential for navigating complex and dynamic environments.

Strengths

RGMs possess several strengths that are relevant to the pursuit of AGI. Their inherent generative nature allows them to create new data and potentially imagine future scenarios, which can be valuable for planning and problem-solving [9]. The hierarchical structure of many RGMs supports hierarchical planning, enabling them to break down complex tasks into simpler subgoals at different levels of abstraction [9]. The integration of active inference mechanisms allows RGM-based agents to act in a goal-directed manner by actively seeking information and minimizing their prediction error [5]. Recent research suggests that some RGMs can achieve high accuracy with less training data and energy compared to traditional deep learning models, which could be a significant advantage for scalability [51].

Weaknesses

RGMs also have certain weaknesses. Compared to models focused on generating highly detailed outputs, such as those operating in pixel space, RGMs, particularly those relying on discrete state-space representations, might produce less detailed outputs [12]. The focus on discrete states could also be a limitation when modeling environments that are inherently continuous [8]. Historically, applications of active inference and RGMs have been largely limited to smaller-scale problems, although recent research is aiming to address this [8]. The computational cost associated with training and inverting complex RGMs, especially those with deep hierarchical structures or operating on high-dimensional data, can also be substantial [53].



Recent Research and Progress

Recent research on RGMs has shown promising progress. Friston et al. (2024) introduced Renormalizing Generative Models (RGM) as an advancement in active inference, specifically designed to scale to more complex problems [8]. VERSES.ai is actively developing and championing RGM, highlighting its potential for achieving high accuracy with reduced data and energy requirements [8]. Applications of RGMs are being explored in various domains, including video compression, image classification, audio processing, and playing Atari games, demonstrating their versatility [9]. Additionally, research is investigating the use of recurrent generative models for efficient heuristic generation in robot path planning, suggesting their potential in robotics and embodied AI [59].

Representational Learning

RGMs primarily learn state representations that capture the essential information about the current state of a system and its environment [5]. These representations can be either discrete or continuous, depending on the specific model implementation. In hierarchical RGMs, representations at different levels of the hierarchy capture information at varying degrees of abstraction, allowing the model to understand complex systems in a compositional manner [9]. Furthermore, the inclusion of "paths" as latent variables in some RGMs enables them to represent not just the current state but also the history and potential future trajectories of the system, providing a richer understanding of its dynamics [9].

Reasoning Mechanisms

Reasoning in RGMs often involves hierarchical planning, where complex goals are decomposed into simpler subgoals at different levels of abstraction [9]. This allows for more efficient and effective planning in complex environments. Additionally, the principle of active inference, frequently associated with RGMs, involves agents actively perceiving their environment to minimize the discrepancy between their predictions and their sensory inputs [5]. This process of "inference through active perception" guides the agent's actions and its information-seeking behavior, allowing it to resolve uncertainty and make informed decisions.

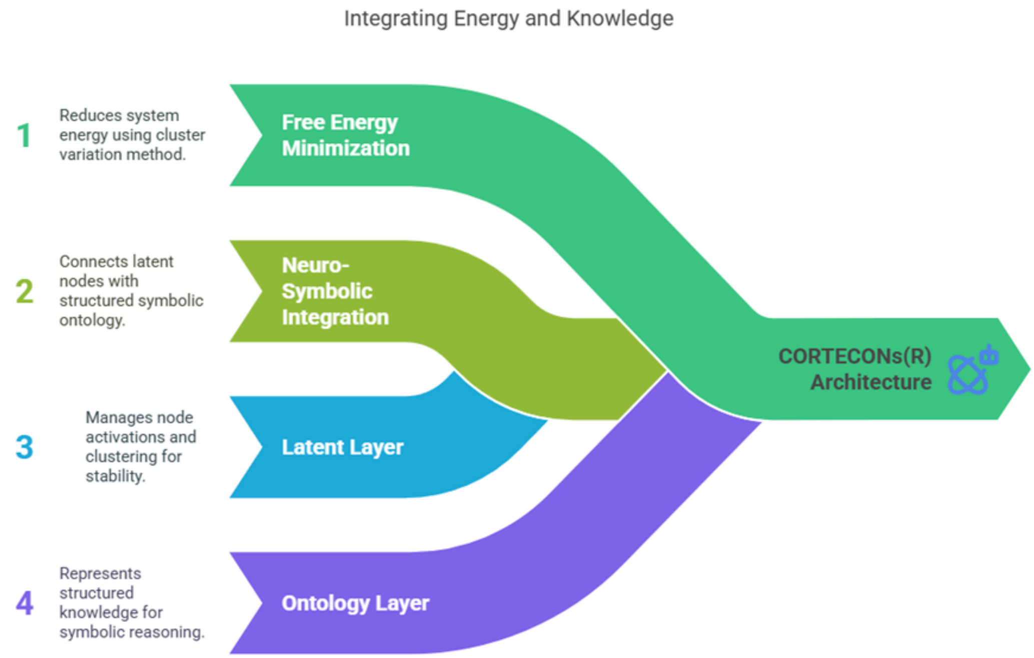
Generalization Capabilities

RGMs typically generalize within the defined state spaces they have been trained on [5]. Their ability to handle novel situations depends on the richness and diversity of the training data and the model's capacity to learn the underlying dynamics of the environment. The hierarchical nature of some RGMs may facilitate generalization to new situations that share similar underlying structures or dynamics with the training data, even if the specific states encountered are different [9]. However, if faced with entirely new environments or state spaces that are significantly different from their training, RGMs might struggle to generalize effectively without further training or adaptation.

5. CORTECONs(R)

Core Principles

CORTECONs(R) represent a novel class of neural networks with core principles centered around free energy minimization and neuro-symbolic integration [5]. The architecture employs free energy minimization across a latent layer using the cluster variation method (CVM) from statistical mechanics [10]. The CVM allows for a more intricate calculation of entropy, considering not just individual node activations but also the interactions between neighboring nodes [10]. A primary objective of CORTECONs(R) is to achieve neuro-symbolic integration by establishing a connection between a subsymbolic (latent node) layer and a structured ontology layer that represents symbolic knowledge [5]. The CVM mechanism enables control over the number of active latent nodes and their clustering [10]. This clustering is crucial for providing temporal persistence of activation, allowing groups of latent nodes to maintain stable activation over time and drive long-term activation within the ontology layer [11].



Generative Capabilities

While CORTECONs(R) are not primarily framed as generative models in the same vein as LLMs or RGMs, their architecture suggests a potential for generating novel patterns [10]. They are designed to facilitate associative processes, akin to "dreaming," where the latent layer undergoes dynamic interaction and stabilization of activation patterns [10]. This process could lead to the emergence of new associations and patterns that were not explicitly present in the input. The ability to repeatedly bring the latent layer to a state of equilibrium is considered a fundamental capability for building an AGI, potentially enabling the system to explore and stabilize novel internal states and relationships [10].

World Model

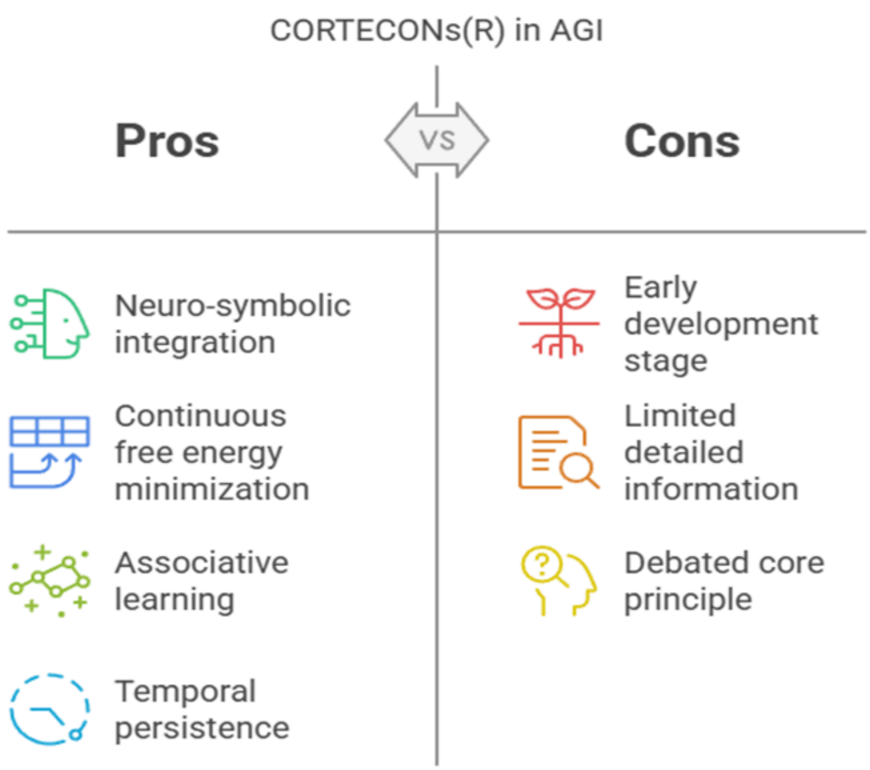
In the CORTECONs(R) architecture, the ontology layer serves as the explicit world model [5]. This layer contains symbolic representations of entities, events, places, and things, along with their associated attributes, relationships, and beliefs [11]. The latent node layer acts as an interface, mediating between sensory inputs and this symbolic world model. Clusters of active latent nodes can provide sustained activation to corresponding nodes within the ontology layer, enabling the representation and maintenance of long-term knowledge [11]. Symbolic reasoning and inference are intended to occur within this ontology layer, providing a structured and interpretable representation of the system's knowledge about the world.

Strengths

A primary strength of CORTECONs(R) lies in their unique neuro-symbolic nature, which aims to integrate the pattern recognition abilities of neural networks with the reasoning and knowledge representation capabilities of symbolic AI [5]. The continuous free energy minimization within the latent layer provides a mechanism for maintaining stability and reaching equilibrium, which is considered crucial for robust processing [10]. CORTECONs(R) are also designed to facilitate associative learning, allowing the system to discover new relationships and connections between presented input patterns [10]. Furthermore, the architecture aims to provide temporal persistence of information within the latent nodes, enabling the system to retain memories of recent events and maintain context over time [61].

Weaknesses

CORTECONs(R) are in the early stages of development compared to other AGI approaches [5]. Consequently, there is relatively limited detailed information available about their architecture, implementation, and empirical results, primarily originating from the Themesis group, which is actively developing the concept [10]. Additionally, the reliance on free energy minimization as a core principle for intelligence has been a subject of debate within the broader scientific community [70].



Recent Research and Progress

Recent progress on CORTECONs(R) is primarily driven by Themesis Inc., with ongoing efforts focused on developing a fully functional framework [60]. Their development plan includes integrating control structures, feedback loops, and interactions between the latent node layer and an ontology layer [60]. Recent updates are provided through blog posts and YouTube videos that discuss the architecture, underlying equations based on the cluster variation method, and the connection to AGI [61]. Current research focuses on achieving stable latent layer equilibrium and facilitating the neuro-symbolic connection between the latent layer and the ontology [65].

Representational Learning

CORTECONs(R) employ a dual system of representation, utilizing latent node activations and symbolic representations [5]. The latent layer consists of a grid of nodes where the activation state of each node contributes to the overall representation, with the dynamics of these activations governed by free energy minimization. These latent node activations are linked to symbolic representations within the ontology layer, which provides a structured knowledge base containing symbols, attributes, and relationships. The temporal persistence of latent node clusters enables the sustained activation of corresponding symbolic representations in the ontology, allowing for the maintenance of contextual information and long-term knowledge.

Reasoning Mechanisms

The primary reasoning mechanism in CORTECONs(R) is intended to occur at the symbolic level within the ontology layer [5]. Once concepts and relationships are activated in the ontology by the latent layer, the system is designed to perform logical inference and manipulation of these symbols to answer questions, solve problems, and make decisions. The sustained activation of symbolic representations, driven by the temporal persistence in the latent layer, supports reasoning processes that require maintaining context and accessing relevant knowledge over time.

Generalization Capabilities

The structured knowledge representation within the ontology layer provides the potential for broad generalization in CORTECONs(R) [5]. By encoding general concepts and their relationships, the system could potentially reason across different domains and adapt to novel situations based on its existing knowledge. The stable latent layer and its associative learning capabilities could further contribute to generalization by allowing the system to recognize new patterns and relate them to existing symbolic knowledge in the ontology.

6. Comparative Analysis and Discussion

To provide a clearer comparison of the four architectures, Table 1 summarizes their key features based on the preceding analysis.

Table 1: Summary of Key Architectural Features

Feature	JEPA	LLMs + RL	RGM	CORTECONs(R)
Core Principle	Predictive learning of abstract embeddings, world model learning	Language modeling with reinforcement learning for decision-making	Generative modeling, active inference, hierarchical planning	Free energy minimization, neuro-symbolic integration
Generative?	No	Yes	Yes	No
World Model?	Yes (explicitly aims to learn world models)	Limited (world knowledge primarily through language)	Yes (through state-space models)	Yes (through ontology layer)

Strengths	Abstract representations, self-supervised, efficient for vision	Strong language capabilities, reasoning (limited), scalable	Generative, hierarchical planning, active inference	Neuro-symbolic, stable latent layer, associative learning
Weaknesses	Training complexity, original lack of generative ability	Lack of true understanding, common sense, potential for hallucinations	Less detail in snippets, focus on discrete states	Early stage of development, less detail in snippets
Representation	Abstract embeddings	Token embeddings	State representations	Latent node activations, symbolic representations in ontology layer
Reasoning	Prediction of future states, spatial reasoning (I-JEPA)	Planning (limited), inference from language	Hierarchical planning, inference through active perception	Symbolic reasoning through ontology layer
Generalization	Strong off-the-shelf representations, efficient task adaptation	Generalization primarily within language domain, struggles with novel tasks	Generalization within defined state spaces	Potential for generalization through structured knowledge representation

- JEPA (Joint Embedding Predictive Architecture): Yann LeCun, Meta's Chief AI Scientist, proposed the JEPA architecture aimed at advancing machine intelligence. Turing Post+6Meta AI+6Rohit Bandaru+6

- LLMs + RL (Large Language Models combined with Reinforcement Learning): This is an active area of research involving numerous contributors. Notably, Amir Feizpour, founder of Aggregate Intellect, has explored how LLMs can enhance reinforcement learning agents. Maven | Unlock your career growth

- Recurrent Generative Models (RGM): Several researchers have significantly contributed to the development and application of RGMs:
  - Alex Graves: Known for his work on sequence modeling, Graves introduced the use of Long Short-Term Memory (LSTM) networks in generative tasks. His research demonstrated the effectiveness of LSTMs in generating complex sequences, such as handwriting and speech.
  - Jürgen Schmidhuber: As a pioneer in the field of recurrent neural networks, Schmidhuber's early work laid the foundation for many advancements in RGMs. His contributions include the development of LSTM networks, which have become a cornerstone in modeling sequential data.
  - Kanaka Rajan: A computational neuroscientist, Rajan has developed RNN-based models to understand neural sequence generation and memory, providing insights into how biological processes can inspire artificial generative models.
  - Karl Friston, a renowned neuroscientist known for his work on generative models in the context of brain function and active inference. He has developed theoretical frameworks that describe how the brain uses generative models to predict and interpret sensory information. His work encompasses various aspects of neuroscience, including perception, action, and learning. One of his notable contributions is the development of the Active Inference framework, which explains how organisms maintain their states and adapt to their environment

through predictive modeling. This framework has been applied to understand linguistic communication, as detailed in the paper "Generative models, linguistic communication and active inference" . Additionally, Friston has explored the anatomical aspects of inference in the brain, discussing how generative models constrain the connectivity and function of different brain regions. This is elaborated in the article "The Anatomy of Inference: Generative Models and Brain Structure"

- CORTECONs (COntent-Retentive TEMporally-CONnected neural networks): Alianna Maren has developed CORTECONs, a new class of neural networks designed to enable associative processes such as "dreaming" in artificial general intelligence systems

Representational Learning

The four architectures employ distinct methods for representational learning, each with its own strengths and weaknesses. JEPA utilizes abstract embeddings, which offer efficiency and generalization by focusing on semantic information [6]. However, capturing fine-grained local details can be challenging [23]. LLMs + RL rely on token embeddings, which are highly effective for language understanding but may lack the grounding and richness required for general intelligence beyond language [5]. RGMs learn state representations that capture system dynamics and temporal dependencies, which is beneficial for modeling environments and planning [5]. However, these representations might be limited by the granularity of the state space. CORTECONs(R) employ a dual approach with latent node activations for pattern recognition and symbolic representations in an ontology for structured knowledge and reasoning [5]. This hybrid method aims to combine the strengths of connectionist and symbolic AI, potentially leading to more robust and explainable representations.

Reasoning Mechanisms

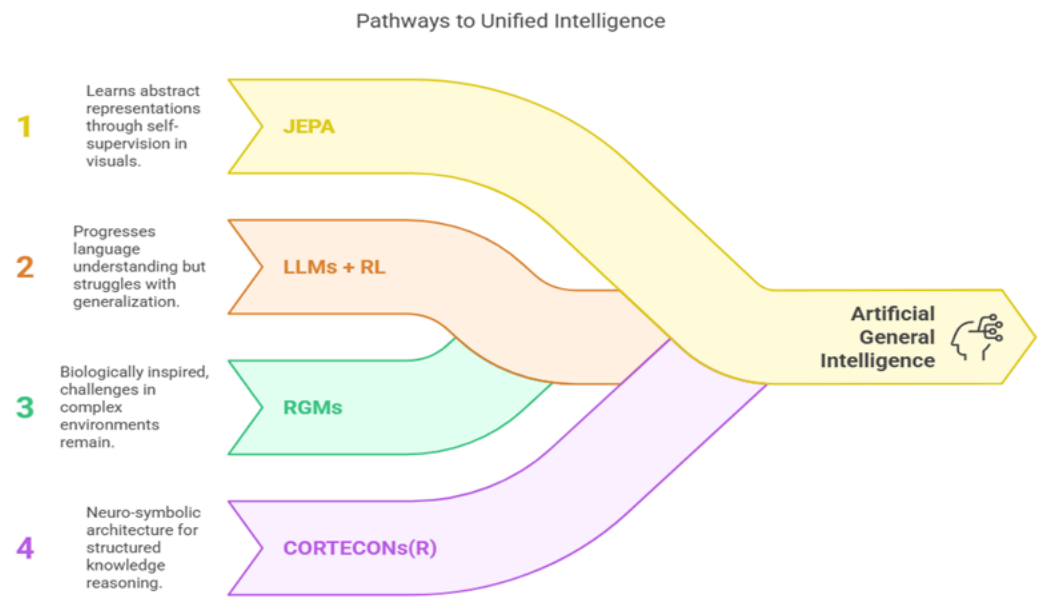
The reasoning mechanisms associated with each architecture also vary significantly. JEPA reasons primarily through prediction of future states and spatial relationships, which is fundamental for interacting with a dynamic world [5]. LLMs + RL perform inference from language, leveraging learned patterns and relationships in text data, with planning abilities often limited [5]. RGMs utilize hierarchical planning and inference through active perception, allowing agents to break down complex tasks and actively seek information to minimize prediction error [5]. CORTECONs(R) aim for symbolic reasoning through their ontology layer, enabling logical inference and manipulation of structured knowledge [5]. The potential of each reasoning approach to achieve human-level intelligence depends on its ability to handle uncertainty, adapt to novel situations, and perform complex, multi-step reasoning.

Generalization Capabilities

Generalization, the ability to adapt to novel tasks and environments, is a critical requirement for AGI, and each architecture exhibits different strengths and limitations in this regard. JEPA shows promise in learning off-the-shelf representations that can be efficiently adapted to new tasks, particularly in the visual domain [5]. LLMs + RL primarily generalize within the language domain, often requiring fine-tuning for new tasks and struggling with situations outside their training data or requiring embodied intelligence [2]. RGMs typically generalize within their defined state spaces, with their ability to handle novel situations depending on the coverage of their training data and the flexibility of their state representation [5]. CORTECONs(R) hold the potential for broad generalization through their structured knowledge representation in the ontology layer, which could allow for reasoning across different domains based on shared concepts and relationships [5].

7. Conclusion

The pursuit of Artificial General Intelligence involves exploring a diverse landscape of architectural approaches, each with its own set of principles, capabilities, strengths, and weaknesses. JEPA demonstrates a powerful ability to learn efficient and generalizable abstract representations through self-supervision, particularly in visual tasks, and recent advancements are addressing its generative limitations. LLMs + RL have shown remarkable progress in language understanding and reasoning but still face fundamental challenges in achieving true understanding, common sense, and robust generalization beyond language. RGMs offer a biologically inspired approach through active inference and hierarchical planning, with the potential for data efficiency, but scaling to complex, continuous environments remains a hurdle. CORTECONs(R) present a novel neuro-symbolic architecture that aims to integrate connectionist and symbolic processing, offering a unique pathway towards reasoning with structured knowledge. However, it is still in the early stages of development.



Ultimately, the path to AGI may not lie in a single one of these architectures but potentially in hybrid approaches that combine their respective strengths. For instance, integrating the efficient representation learning of JEPA with the symbolic reasoning of a CORTECONs(R) or leveraging the language understanding of LLMs within an active inference framework could yield more robust and general intelligent systems. Continued research and development across these and other promising architectures are crucial for making further progress towards the complex and multifaceted goal of Artificial General Intelligence.

References

1. Bowen-xu/AGI-Survey: An ongoing survey of Artificial ... - GitHub, accessed March 21, 2025, https://github.com/bowen-xu/AGI-Survey

2. (PDF) A large Language Model is not the Right Path to Bring ..., accessed March 21, 2025, https://www.researchgate.net/publication/389855793_A_large_Language_Model_is_not_the_Right_Path_to_Bring_Artificial_General_Intelligence

3. Advanced AGI vs LLMs Techniques: A Comprehensive Guide - BytePlus, accessed March 21, 2025, https://www.byteplus.com/en/topic/382110

4. Large language models for artificial general intelligence (AGI): A survey of foundational principles and approaches - arXiv, accessed March 21, 2025, https://arxiv.org/html/2501.03151v1

5. Deep Dive in Meta AI's JEPA - Joint-Embedding Predictive ... - GitHub, accessed March 21, 2025, https://github.com/neobundy/Deep-Dive-Into-AI-With-MLX-PyTorch/blob/master/deep-dives/015-meta-jepa/README.md

6. What is Joint Embedding Predictive Architecture (JEPA)? - Turing Post, accessed March 21, 2025, https://www.turingpost.com/p/jepa

7. Reinforcement Learning Enhanced LLMs: A Survey - arXiv, accessed March 21, 2025, https://arxiv.org/html/2412.10400v1

8. Big AGI Breakthrough: Leveling the Playing Field - Themesis, Inc., accessed March 21, 2025, https://themesis.com/2024/08/19/big-agi-breakthrough-leveling-the-playing-field/

9. (PDF) From pixels to planning: scale-free active inference - ResearchGate, accessed March 21, 2025, https://www.researchgate.net/publication/382692011_From_pixels_to_planning_scale-free_active_inference

10. CORTECONS: A New Class of Neural Networks - Themesis, Inc., accessed March 21, 2025, https://themesis.com/2023/09/05/cortecons-a-new-class-of-neural-networks/

11. AGI Wars: Emerging Landscape - Themesis, Inc., accessed March 21, 2025, https://themesis.com/2024/10/15/agi-wars-emerging-landscape/

12. Deep Dive into Yann LeCun's JEPA - Rohit Bandaru, accessed March 21, 2025, https://rohitbandaru.github.io/blog/JEPA-Deep-Dive/

13. Meta AI's I-JEPA Explained | Encord, accessed March 21, 2025, https://encord.com/blog/i-jepa-explained/

14. Can someone explain to me in a laymen-ish terms why is I-JEPA (Yann Lecun) architecture not generative? - Reddit, accessed March 21, 2025, https://www.reddit.com/r/learnmachinelearning/comments/1gya9pm/can_someone_explain_to_me_in_a_laymenish_terms/

15. Meta AI Releases the Video Joint Embedding Predictive Architecture (V-JEPA) Model: A Crucial Step in Advancing Machine Intelligence - MarkTechPost, accessed March 21, 2025, https://www.marktechpost.com/2025/02/22/meta-ai-releases-the-video-joint-embedding-predictive-architecture-v-jepa-model-a-crucial-step-in-advancing-machine-intelligence/

16. V-JEPA: The next step toward advanced machine intelligence, accessed March 21, 2025, https://ai.meta.com/blog/v-jepa-yann-lecun-ai-model-video-joint-embedding-predictive-architecture/

17. Denoising with a Joint-Embedding Predictive Architecture - OpenReview, accessed March 21, 2025, https://openreview.net/forum?id=d4njmzM7jf

18. arxiv.org, accessed March 21, 2025, https://arxiv.org/abs/2501.14622

19. Connecting Joint-Embedding Predictive Architecture with Contrastive Self-supervised Learning - arXiv, accessed March 21, 2025, https://arxiv.org/html/2410.19560

20. NeurIPS Poster Connecting Joint-Embedding Predictive Architecture with Contrastive Self-supervised Learning, accessed March 21, 2025, https://neurips.cc/virtual/2024/poster/95692

21. [2410.19560] Connecting Joint-Embedding Predictive Architecture with Contrastive Self-supervised Learning - arXiv, accessed March 21, 2025, https://arxiv.org/abs/2410.19560

22. Connecting Joint-Embedding Predictive Architecture with Contrastive Self-supervised Learning | OpenReview, accessed March 21, 2025, https://openreview.net/forum?id=JvQnJWIj6m&referrer=%5Bthe%20profile%20of%20Shentong%20Mo%5D(%2Fprofile%3Fid%3D~Shentong_Mo1)

23. Dense Representation Learning for a Joint-Embedding Predictive Architecture, accessed March 21, 2025, https://openreview.net/forum?id=8GCcSXlkZN

24. [2405.17995] DMT-JEPA: Discriminative Masked Targets for Joint-Embedding Predictive Architecture - arXiv, accessed March 21, 2025, https://arxiv.org/abs/2405.17995

25. Point-JEPA: A Joint Embedding Predictive Architecture for Self-Supervised Learning on Point Cloud - CVF Open Access, accessed March 21, 2025, https://openaccess.thecvf.com/content/WACV2025/papers/Saito_Point-JEPA_A_Joint_Embedding_Predictive_Architecture_for_Self-Supervised_Learning_on_WACV_2025_paper.pdf

26. A conversation with Yann LeCun, the godfather of AI | Wing Venture ..., accessed March 21, 2025, https://www.wing.vc/content/rajeev-chand-yann-lecun-ai-research-predictions

27. Towards World Models: How Latent Reasoning & JEPA Move AI ..., accessed March 21, 2025, https://medium.com/@mspraj00/towards-world-models-how-latent-reasoning-jepa-move-ai-beyond-tokens-4d4cf6fa372b

28. V-JEPA - Champaign Magazine, accessed March 21, 2025, https://champaignmagazine.com/tag/v-jepa/

29. How Close is AGI Actually? Why LLMs Alone Will Not Get us to AGI - NJII, accessed March 21, 2025, https://www.njii.com/2024/07/why-llms-alone-will-not-get-us-to-agi/

30. The Role of Reinforcement Learning in Enhancing LLM Performance - Dataversity, accessed March 21, 2025, https://www.dataversity.net/the-role-of-reinforcement-learning-in-enhancing-llm-performance/

31. Is reinforcement learning the key for achieving AGI? : r/reinforcementlearning - Reddit, accessed March 21, 2025, https://www.reddit.com/r/reinforcementlearning/comments/1ivns8i/is_reinforcement_learning_the_key_for_achieving/

32. RLAIF for Enhanced Training in LLMs | by Bijit Ghosh - Medium, accessed March 21, 2025, https://medium.com/@bijit211987/rlaif-for-enhanced-training-in-llms-63629289642a

33. What is Generative AI? Understanding Its Benefits - Learn Prompting, accessed March 21, 2025, https://learnprompting.org/docs/basics/generative_ai

34. Transforming Software Development: The Impact of Generative AI in 2025 - Charter Global, accessed March 21, 2025, https://www.charterglobal.com/generative-ai-software-development/

35. Mastering Generative AI: A comprehensive guide - Scribble Data, accessed March 21, 2025, https://www.scribbledata.io/blog/mastering-generative-ai-a-comprehensive-guide/

36. Using LLMs to Solve the ARC-AGI Challenge | SMU Guildhall, accessed March 21, 2025, https://www.smu.edu/guildhall/academics/research/using-llms-to-solve-the-arc-agi-challenge

37. Part 2 — Beyond Language: Why Scaling LLMs Won't Lead to AGI - Medium, accessed March 21, 2025, https://medium.com/autonomous-agents/part-2-beyond-language-why-scaling-llms-wont-lead-to-agi-2c3c57a83adf

38. #3 LLM: Reinforcement Learning — GPT | by LAKSHMI VENKATESH | Medium, accessed March 21, 2025, https://luxananda.medium.com/reinforcement-learning-gpt-742016025359

39. Generating Code World Models with Large Language Models Guided by Monte Carlo Tree Search | OpenReview, accessed March 21, 2025, https://openreview.net/forum?id=9SpWvX9ykp¬eId=kH2xomkeeu

40. Evaluating World Models with LLM for Decision Making - arXiv, accessed March 21, 2025, https://arxiv.org/html/2411.08794v1

41. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning - arXiv, accessed March 21, 2025, https://arxiv.org/html/2501.12948v1

42. DeepSeek-R1 - A Step Towards AGI [Paper Unfold], accessed March 21, 2025, https://newsletter.himanshuramchandani.co/p/deepseek-r1-a-step-towards-agi-paper-unfold

43. DeepSeek-R1 Paper Explained – A New RL LLMs Era in AI? - AI Papers Academy, accessed March 21, 2025, https://aipapersacademy.com/deepseek-r1/

44. Why I wouldn't rule out Large Language Models taking us to AGI: : r/ArtificialInteligence, accessed March 21, 2025, https://www.reddit.com/r/ArtificialInteligence/comments/1deb9kp/why_i_wouldnt_rule_out_large_language_models/

45. Generative AI vs. AGI: The Cognitive Strengths and Weaknesses of Modern LLMs - arXiv, accessed March 21, 2025, https://arxiv.org/html/2309.10371

46. New study suggests that LLM can not bring AGI : r/LocalLLaMA - Reddit, accessed March 21, 2025, https://www.reddit.com/r/LocalLLaMA/comments/1jbkfyq/new_study_suggest_that_llm_can_not_bring_agi/

47. Generative models vs Discriminative models for Deep Learning. - Turing, accessed March 21, 2025, https://www.turing.com/kb/generative-models-vs-discriminative-models-for-deep-learning

48. RGM (Renormalizing Generative Model) - Vocab - Envisioning.io, accessed March 21, 2025, https://www.envisioning.io/vocab/rgm-renormalizing-generative-model

49. [2501.05458] Generative Modeling: A Review - arXiv, accessed March 21, 2025, https://arxiv.org/abs/2501.05458

50. Structure of the generative recurrent model: the neural network... - ResearchGate, accessed March 21, 2025, https://www.researchgate.net/figure/Structure-of-the-generative-recurrent-model-the-neural-network-consists-of-recurrent_fig3_317138434

51. VERSES' Latest Research Advances Beyond GenAI With RGM Conceptual Modeling, accessed March 21, 2025, https://deniseholt.us/verses-advances-genai-with-rgm/

52. Learning Generative State Space Models for Active Inference - Frontiers, accessed March 21, 2025, https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2020.574372/full

53. From pixels to planning: scale-free active inference | AI Research Paper Details, accessed March 21, 2025, https://www.aimodels.fyi/papers/arxiv/from-pixels-to-planning-scale-free-active

54. [2407.20292] From pixels to planning: scale-free active inference - arXiv, accessed March 21, 2025, https://arxiv.org/abs/2407.20292

55. Conditioning Generative Models with Restrictions - Adam Jermyn, accessed March 21, 2025, https://adamjermyn.com/posts/conditioninggenerativemodels2/

56. (PDF) Development And Challenges of Generative Artificial ..., accessed March 21, 2025, https://www.researchgate.net/publication/379529325_Development_And_Challenges_of_Generative_Artificial_Intelligence_in_Education_and_Art

57. A Comprehensive Survey on Generative AI Solutions in IoT Security - MDPI, accessed March 21, 2025, https://www.mdpi.com/2079-9292/13/24/4965

58. Generative models, linguistic communication and active inference - PMC - PubMed Central, accessed March 21, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7758713/

59. [PDF] Efficient Heuristic Generation for Robot Path Planning with Recurrent Generative Model | Semantic Scholar, accessed March 21, 2025,

https://www.semanticscholar.org/paper/Efficient-Heuristic-Generation-for-Robot-Path-with-Li-Wang/99a6329caefa09240ffe0fb522dcb3d1b8434bfe

60. CORTECONs: AGI in 2024-2025 – R&D Plan Overview - Themesis, Inc., accessed March 21, 2025, https://themesis.com/2024/01/11/cortecons-agi-in-2024-2025-rd-plan-overview/

61. AGI: Three Foundation Methods - RGM, JEPA, and CORTECONs(R) - YouTube, accessed March 21, 2025, https://www.youtube.com/watch?v=7-r7PvukgZw

62. Neuro-Symbolic AI: A Pathway Towards Artificial General Intelligence - Solutions Review, accessed March 21, 2025, https://solutionsreview.com/neuro-symbolic-ai-a-pathway-towards-artificial-general-intelligence/

63. Neuro-Symbolic AI Newsletter | December 2024 - DEV Community, accessed March 21, 2025, https://dev.to/nucleoid/neuro-symbolic-ai-newsletter-december-2024-1m6n

64. Overcoming Recommendation Limitations with Neuro-Symbolic Integration - ResearchGate, accessed March 21, 2025, https://www.researchgate.net/publication/373935993_Overcoming_Recommendation_Limitations_with_Neuro-Symbolic_Integration

65. AGI Wars: Evolving Landscape and Sun Tzu Analysis - YouTube, accessed March 21, 2025, https://www.youtube.com/watch?v=RAtad6UmNUM

66. CORTECONs - Themesis, Inc., accessed March 21, 2025, https://themesis.com/category/cortecons/

67. CORTECON II – Alianna J. Maren, accessed March 21, 2025, https://www.aliannajmaren.com/cortecon-ii/

68. Category: Artificial Intelligence - Themesis, Inc., accessed March 21, 2025, https://themesis.com/category/artificial-intelligence/

69. AGI: RGMs, JEPA, and CORTECONs(R): Three AGI Building Blocks ..., accessed March 21, 2025, https://themesis.com/2024/09/12/agi-rgms-jepa-and-corteconsr-three-agi-building-blocks/

70. The Free Energy Minimization Principle as a Neurophysiological Dead End for AGI | by Victor Senkevich | Medium, accessed March 21, 2025, https://medium.com/@VictorSenkevich/the-free-energy-minimization-principle-as-a-neurophysiological-dead-end-for-agi-d685a61b4e3b