ISSN 2007-9737

Towards a Proto Artificial General Intelligence: The Role of Large Language Model Ontologies in its Development

Christian E. Maldonado-Sifuentes[1], Mariano Vargas-Santiago[2], Samuel Solis-Gamboa[2], Grigori Sidorov[3,*], Luis Lechuga-Gutierrez[2], Francisco González-Andrade[2], María del Carmen Heras-Sánchez[4]

[1] Transdisciplinary Research for Augmented Innovation Laboratory,
Mexico

[2] Consejo Nacional de Humanidades Ciencias y Tecnologías,
Mexico

[3] Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Mexico

[4] Universidad de Sonora,
Mexico

christian.maldonado@trai-l.com, {mariano.vargas, samuel.solis,
luis.lechuga, francisco.gonzalez}@conahcyt.mx, sidorov@cic.ipn.mx, carmen@investigacion.uson.mx

Abstract. Proto Artificial General Intelligence (ProtoAGI) aims to create a versatile artificial intelligence system capable of autonomously performing diverse tasks. A foundational element of ProtoAGI is the Large Language Model (LLM) ontology, which plays a crucial role in organizing and retrieving information about different LLMs, enabling the selection of the most appropriate model for specific tasks. This ontology, the first of several designed to support ProtoAGI, addresses key challenges in managing and accessing information regarding LLM capabilities, performance, and task suitability. We present the methodology for constructing this ontology, covering data extraction, enrichment, and model recommendation using a generalized LLM API. The initial version of this ontology involved processing over a million tokens, underscoring the system's complexity and the scale of information integrated. This ontology is designed for continuous updates, ensuring that ProtoAGI remains current with the latest advancements in LLMs. The ongoing development of this ontology marks a significant step in ProtoAGI's evolution, following an initial proof-of-concept demonstrated during the 2024 eclipse, where the feasibility of integrating such a comprehensive LLM ontology into a general-purpose AI system was shown. By making this ontology accessible to the broader AI community, we aim to accelerate further advancements in AGI research and applications.

Keywords. Artificial general intelligence, large language models, ontology, hybrid intelligent systems, multi-agent systems.

1 Introduction

The aspiration to develop Artificial General Intelligence (AGI) systems with human-like cognitive capabilities across a diverse array of tasks stands at the forefront of artificial intelligence research. AGI seeks to transcend the specialized functionalities of current AI technologies, aiming instead for a breadth and adaptability of intelligence that mirrors human cognition. From the perspective of research the development of AGI also represents a pivotal goal, aimed at creating systems that emulate the

ISSN 2007-9737

comprehensive cognitive capabilities of humans across varied tasks.

AGI aspires to surpass the specialized functions typical of contemporary AI technologies, striving for a level of versatility and adaptability akin to human intellect. This ambition extends beyond the replication of human-like problem-solving and learning capacities; it also seeks to incorporate creativity, emotional intelligence, and an understanding of complex contexts and nuances. In this manuscript, we present and advancement over our original proof-of-concept:

Proto Artificial General Intelligence (ProtoAGI), a cutting-edge AGI simulation system marked by its dynamic learning capabilities and a modular, scalable architecture. ProtoAGI signifies a departure from conventional AI systems, embodying a framework designed for continual evolution and adaptation.

Central to this system is a self-enhancing knowledge base that facilitates iterative learning and reflection, thereby redefining the standards of machine cognition's flexibility and depth. We introduce a general ontology within ProtoAGI, enabling the system to effectively address and resolve diverse problems.

This ontology functions as a sophisticated mechanism that discerns and categorizes inputs, ensuring the delivery of precise and contextually appropriate outcomes. The incorporation of this ontology not only enriches the system's operational intelligence but also enhances its capability to perform across both general and specialized tasks with heightened accuracy. Our approach underscores the necessity of developing AGI systems that are not only technically proficient but also ethically informed and capable of seamless integration within human-centric environments.

The architecture of ProtoAGI is strategically crafted to propel forward the domain of AGI by tackling critical challenges such as learning efficiency, adaptability, and comprehension of human emotions and ethical considerations. Recent advancements in AGI research underscore its growing relevance and potential applications across various sectors. Studies by Naudé and Dimitri in 2020 highlight the implications of AGI for public policy, discussing the control and political challenges arising from its development [5].

Fei et al. (2022) explore a multimodal foundation model for AGI, trained on extensive data to handle various cognitive tasks, marking a significant step toward realizing AGI [3]. Yamakawa (2021) adopts a whole brain architecture approach, aiming to accelerate AGI development by mimicking brain structures and functions [12].

Additionally, Shevlin et al. (2019) discuss the inherent limitations and challenges in achieving AGI despite advancements in machine intelligence [8]. These contributions not only position ProtoAGI within the contemporary research landscape but also illuminate the system's potential to advance AGI technologies that are adaptable, ethically responsible, and capable of complex, human-like reasoning and learning.

In a previous manuscript, we introduced ProtoAGI, a novel AGI simulation system characterized by its dynamic learning capabilities and a modular, scalable architecture. The Large Language Model (LLM) ontology is a critical component of ProtoAGI, organizing and retrieving relevant information about various LLMs, enabling the selection of the most suitable model for specific tasks. This article focuses on the importance of the LLM ontology, the methodology used to develop it, and the iterative nature of its creation, marking significant progress towards completing ProtoAGI's first prototype.

2 Related Works

Research in Artificial General Intelligence (AGI) aims to develop systems with cognitive capabilities that mirror human intelligence, transcending the specialized functionalities of current AI technologies. Significant contributions to this field include the development of computational cognitive architectures and the integration of symbolic and sub-symbolic computations. The ACT-R computational cognitive architecture is a notable example, offering a framework for modeling human-like cognitive abilities.

This architecture has been validated against extensive experimental data, showcasing its potential in creating systems with general intelligence capabilities akin to those of humans [4]. Furthermore, the integration of symbolic and sub-symbolic computation is pivotal for advancing human-level artificial intelligence. This approach addresses the complex challenge of developing systems that can operate across varied tasks with adaptability and depth akin to human cognition [2].

These studies underline the importance of a holistic approach in AGI research, combining technical proficiency with sophisticated cognitive models to foster systems capable of complex reasoning and learning.

Research on Artificial General Intelligence (AGI) with human-like cognitive capabilities has been progressing with significant contributions from various scholars. Rayhan et al. [7] provide a roadmap to achieving human-level capabilities in AGI, highlighting the developmental strides necessary for such advancements. This work lays a foundational perspective on the scope and ambition of AGI research.

Tong et al. [10] explore the incorporation of affective dimensions into AGI through neural-symbolic computing. Their work aims to endow artificial systems with a more nuanced understanding of human emotions, which is crucial for real-life interactions and cognitive processes.

Jovanović [11] discusses the current limitations and barriers that prevent the realization of AGI. His analysis identifies key areas where further research and technological development are needed, providing a critical overview of the field's challenges. Sonko et al. [9] offer a critical review of the ethical considerations and challenges associated with the development of AGI. Their insights are vital for guiding future research towards responsible and ethically informed technological advancements.

Azam et al. [1] introduce a novel model of narrative memory for conscious agents, aiming to enhance the memory capabilities of AGI systems to better simulate human-like cognitive functions. This work contributes to the broader goal of creating more sophisticated and versatile AGI systems. Together, these studies underscore the diverse approaches and ongoing research efforts aimed at achieving AGI with human-like cognitive abilities, reflecting both the potential and the hurdles in the path towards such advanced artificial intelligence systems.

Despite these advancements, a significant gap in current AGI research is the absence of a comprehensive, general ontology that could unify various approaches and facilitate the development of truly generalizable AGI systems.

Existing works often focus on specific aspects or functionalities, lacking an overarching framework that integrates different cognitive models and computational architectures. Such a general ontology would be instrumental in standardizing methodologies, promoting interoperability between systems, and accelerating progress towards achieving genuine AGI. Without it, efforts remain fragmented, and the vision of fully realizing AGI's potential remains elusive.

3 Importance of LLM Ontology

The LLM ontology is essential for several reasons:

– **Organization of Information**: It systematically organizes information about various LLMs, including their capabilities, specializations, and computational requirements.

– **Facilitating Model Selection**: It aids in selecting the most suitable model for a specific task by providing detailed information and recommendations based on the model's characteristics.

– **Enhancing Efficiency**: By automating the model selection process, it reduces the time and effort required to identify the optimal LLM, thereby improving overall efficiency.

– **Continuous Improvement**: The ontology can be continuously fetched and updated, ensuring that the latest models and their attributes are always available for selection.

3.1 Optimizing ProtoAGI with LLM Ontologies

ProtoAGI is designed to be a versatile AI system capable of autonomously performing a wide range of tasks. A crucial component of this system is the Large Language Model (LLM) ontology, which plays a pivotal role in organizing and retrieving information about various LLMs.

This allows the system to select the most suitable model for a given task. The LLM ontology serves as the first of several planned ontologies that will systematically provide accessible knowledge to ProtoAGI.

The LLM ontology addresses a core challenge: Managing and accessing detailed information about different models, their capabilities, and their performance across diverse contexts. This section outlines the methodology used to create the ontology, which includes data extraction, enrichment, and model recommendations using a generalized LLM API.

3.1.1 Problem Representation

- **Ontologies**:
  - $\mathcal{L}$: Set of LLMs, $\mathcal{L} = \{L_1,\ L_2,\ \ldots,\ L_m\}$.
  - $\mathcal{A}$: Set of algorithms, $\mathcal{A} = \{A_1,\ A_2,\ \ldots,\ A_n\}$.
  - $\mathcal{D}$: Set of datasets, $\mathcal{D} = \{D_1,\ D_2,\ \ldots,\ D_p\}$.
- **Selection Variables**:
  - $x_{ijk} \in \{0,1\}$: Binary variable indicating whether LLM L_i, algorithm A_j, and dataset D_k are selected together.
- **Objective Function**:
  - Maximize $f(x_{ijk})$.
  - Here, $f(x_{ijk})$ is a function that evaluates the performance of the combination of L_i, A_j, and D_k for a specific task.
- **Constraints**:
  - Each task requires exactly one combination of L_i, A_j, and D_k:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{p} x_{ijk} = 1. \tag{1}$$

  - Additional constraints based on resource limits, compatibility, etc:

$$g(x_{ijk}) \leq R, \tag{2}$$

    where $g(x_{ijk})$ represents resource usage (e.g., computation, memory), and R is the resource limit.

3.1.2 Combinatorial Space

The combinatorial space involves all possible combinations of LLMs, algorithms, and datasets. The total number of combinations is $m \times n \times p$. A brute force approach would involve evaluating every combination:

$$\text{Total combinations} = m \cdot n \cdot p. \tag{3}$$

While this guarantees finding the global optimum, the computational cost becomes prohibitive as m, n, and p grow large, making it impractical for large-scale systems like ProtoAGI.

3.2 Advantages of LLM-guided Selection over Brute Force Search

Instead of a brute force search, ProtoAGI leverages multiple specialized LLMs for their reasoning capabilities, enabling them to intelligently guide the search for optimal combinations within the ontology.

This method offers several advantages over brute force search by balancing the trade-off between precision and computational efficiency, effectively replacing human-in-the-loop processes with AI-driven decision-making.

ISSN 2007-9737

3.2.1 Brute Force Search

- **Total combinations**: $m \times n \times p$
- **Objective function evaluation**:

$$f(L_i, A_j, D_k) \quad \forall \quad \begin{aligned} &i \in \{1, \dots, m\}, \\ &j \in \{1, \dots, n\}, \\ &k \in \{1, \dots, p\}. \end{aligned} \tag{4}$$

- **Deterministic precision**: Evaluates every possible combination, guaranteeing the optimal solution.
- **Computational cost**: Extremely high for large m, n, and p, making this approach inefficient and infeasible for large-scale tasks.

3.2.2 LLM-guided Selection

- **Total combinations**: Reduced to a manageable subset $\mathcal{S}$.
- **Stochastic approximation**:

$$\tilde{f}(L_i,\ A_j,\ D_k) \ \text{ for selected } \ (i,\ j,\ k) \in \mathcal{S}, \tag{5}$$

  where $\mathcal{S} \subseteq \{1, \dots, m\} \times \{1, \dots, n\} \times \{1, \dots, p\}$.
- **Simulated reasoning capabilities**: LLMs use their knowledge and contextual understanding to predict the most promising combinations.
- **Human-in-the-loop replacement**: LLMs streamline the decision-making process by emulating expert reasoning.

3.2.3 Mathematical Representation

- **Brute Force Search**:

  Optimal combination: $= \max_{i,j,k} f(L_i,\ A_j,\ D_k)$ with:

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} 1 = m \cdot n \cdot p. \tag{6}$$

- **LLM-guided Selection**:
  - **Subset selection**: LLMs reduce the search space to a subset $\mathcal{S}$.
  - **Stochastic approximation**:

    Approximate optimal combination $=$ $\max_{(i,j,k)\in\mathcal{S}} \tilde{f}(L_i, A_j, D_k)$.
  - **Reduced search space**:

$$|\mathcal{S}| \ll m \cdot n \cdot p. \tag{7}$$

3.2.4 Advantages of LLM-guided Selection

- **Efficiency**:
  - LLMs drastically reduce the number of combinations to evaluate, improving computational efficiency.
  - The search space is reduced from $O(m \cdot n \cdot p)$ to $O(|\mathcal{S}|)$, where $|\mathcal{S}|$ is much smaller.
- **Simulated Reasoning**:
  - LLMs utilize prior knowledge and contextual understanding to predict promising combinations, thus avoiding unnecessary evaluations.
  - Provides a probabilistic assessment of potential solutions, leading to quicker convergence to high-quality solutions.
- **Stochastic Approximation**:
  - Balances precision with computational efficiency by approximating the optimal solution.
  - Enables adaptive and dynamic decision-making, similar to expert human evaluation and refinement.
- **Human-in-the-loop Replacement**:
  - LLMs reduce reliance on manual decision-making processes, enhancing scalability and reducing human error.
  - Improves repeatability in selecting optimal combinations, especially for complex or large-scale tasks.

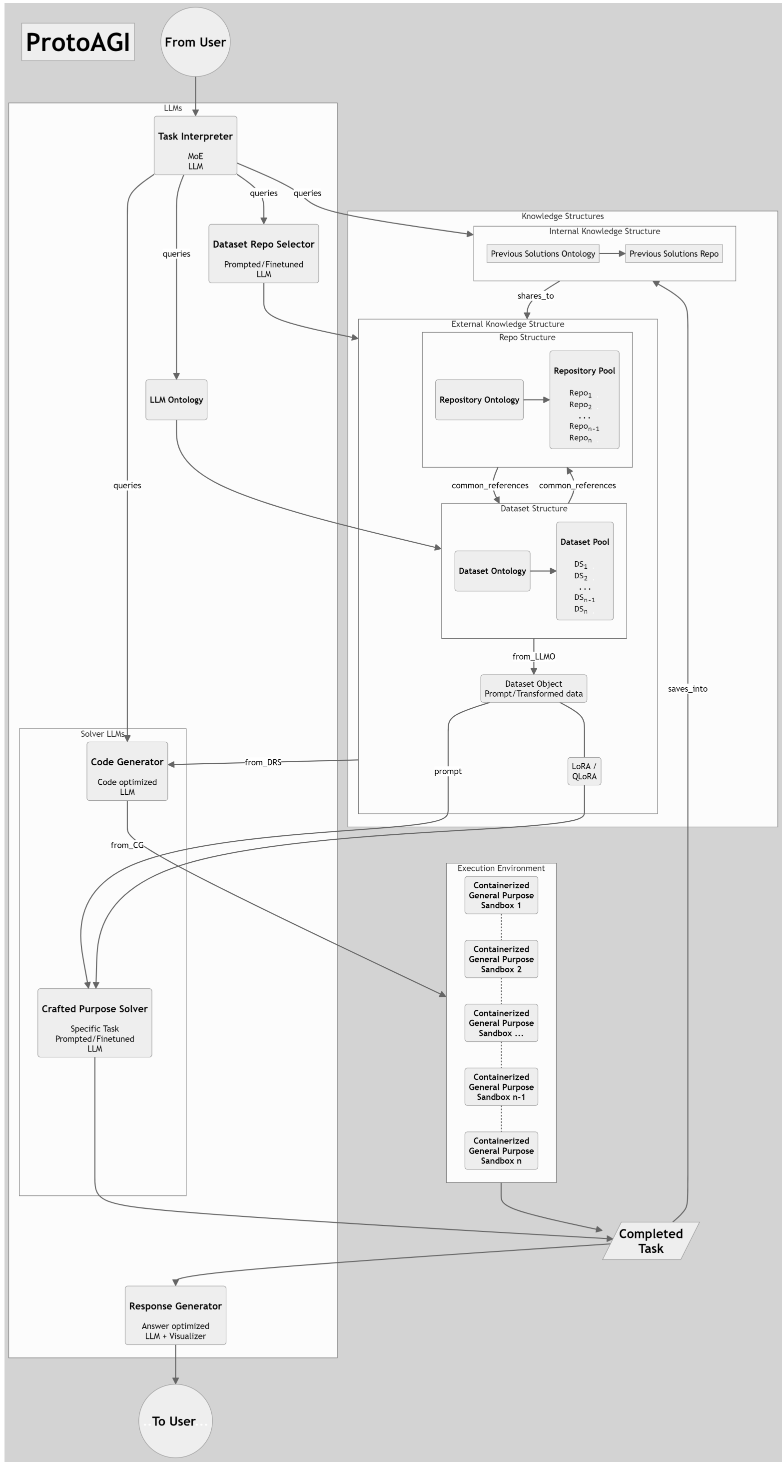

Fig. 1. ProtoAGI system overview

3.2.5 Stochastic Approximation Model

Given the LLM's ability to simulate reasoning, its selection process can be modeled as follows:

– **Initial Predictions**:

$$P(x_{ijk} = 1) = \mathsf{LLM}_{\mathsf{score}}(L_i, A_j, D_k). \quad (8)$$

– **Thresholding**:

$$\mathcal{S} = \{(i, j, k) \mid \mathsf{LLM}_{\mathsf{score}}(L_i, A_j, D_k) \geq \tau\}, \quad (9)$$

where τ is a threshold score to filter top candidates.

– **Evaluation**:

$$\tilde{f}(L_i, A_j, D_k) \quad \forall (i, j, k) \in \mathcal{S}. \quad (10)$$

– **Selection**:

$$\text{Optimal combination} \approx \max_{(i,j,k)\in\mathcal{S}} \tilde{f}(L_i, A_j, D_k). \quad (11)$$

4 Our Method

ProtoAGI is a versatile AI system designed to autonomously perform a wide range of tasks by leveraging multiple specialized Large Language Models (LLMs). A critical component of ProtoAGI is the LLM ontology, which organizes and retrieves relevant information about various LLMs to facilitate the selection of the most suitable model for specific tasks. This section details the methodology behind the creation of the LLM ontology, highlighting its integration within the ProtoAGI system.

4.1 ProtoAGI System Overview

4.1.1 Dynamic Learning Capabilities

ProtoAGI signifies a departure from conventional AI systems, embodying a framework designed for continual evolution and adaptation. Central to this system is a self-enhancing knowledge base that facilitates iterative learning and reflection, thereby redefining the standards of machine cognition's flexibility and depth.

4.1.2 Modular and Scalable Architecture

ProtoAGI is characterized by its modular, scalable architecture that allows for seamless integration of various AI components. This design ensures that the system can adapt to new tasks and incorporate advancements in AI technologies.

4.1.3 General Ontology Integration

A general ontology within ProtoAGI enables the system to effectively address and resolve diverse problems. This ontology functions as a sophisticated mechanism that discerns and categorizes inputs, ensuring the delivery of precise and contextually appropriate outcomes.

4.1.4 Ethical and Human-centric Design

Our approach underscores the necessity of developing AGI systems that are not only technically proficient but also ethically informed and capable of seamless integration within human-centric environments. ProtoAGI is strategically crafted to tackle critical challenges such as learning efficiency, adaptability, and comprehension of human emotions and ethical considerations.

4.2 Ontology Creation Method

The process of creating the LLM ontology involves several steps, including data extraction, ontology enrichment, and model recommendation. This first iteration involved processing over a million tokens to achieve the final version.

4.2.1 Data Extraction and Parsing

The first step involves extracting relevant information about different LLMs from a specified HTML source, namely from the Ollama repository [6]. Ollama serves not only as a comprehensive database but also as a framework for running LLMs in a containerized environment similar to Docker. This ensures consistent performance and ease of deployment across different systems.

The extraction process is achieved using a PHP script that parses the HTML and retrieves key details such as the model's name, URL, description, and additional metadata like the number of pulls, tags, and the last update time. This data is then encoded into a JSON structure for further processing.

4.2.2 Ollama Repository

The Ollama repository is a comprehensive database of large language models, providing detailed information about each model's capabilities, usage statistics, and other relevant metadata. The repository serves as a central hub for researchers and developers to access a wide variety of LLMs. Additionally, Ollama functions as a framework to run these models in isolated, containerized environments, similar to Docker, which facilitates easy deployment and scalability.

By utilizing the Ollama repository, we ensure that our ontology is built upon a robust and diverse dataset, encompassing the latest advancements in LLM technology. This rich source of data allows us to create a well-rounded and informative ontology that can support a wide range of applications within the ProtoAGI system. The extraction process from the Ollama repository involves several steps:

1. **HTML Parsing**: Using PHP, we parse the HTML content of the Ollama repository to identify and retrieve relevant sections containing LLM information.

2. **Data Retrieval**: Extract key details such as the model's name, URL, description, number of pulls, tags, and the last update time.

3. **JSON Encoding**: Encode the retrieved data into a structured JSON format for further processing and enrichment.

By leveraging the extensive data available in the Ollama repository, we ensure that our LLM ontology is comprehensive and up-to-date, reflecting the latest trends and capabilities in the field of large language models.

4.2.3 Tag Extraction and Ontology Enrichment

After extracting the basic information, we enrich our ontology by retrieving additional tags associated with each model from their respective URLs. This process involves fetching the HTML content of the tag pages and extracting the relevant tags using DOM and XPath queries. The enriched data is then encoded into a JSON structure. This step ensures that the ontology captures a comprehensive set of attributes for each model.

4.2.4 LLM Integration for Model Recommendations

To provide detailed descriptions and implications of each model and its associated tags, we integrate an LLM API. This API is used to generate detailed explanations and recommendations for each model based on its description and tags.

The API queries are designed to slow down request handling to avoid overloading the system. This process involved processing over a million tokens to ensure the accuracy and comprehensiveness of the final recommendations. The overall implementation involves three main steps:

1. **HTML Parsing and Data Extraction**: Using PHP, we parse the HTML content to extract models' information, including their names, descriptions, URLs, and metadata.

2. **Tag Retrieval and Ontology Enrichment**: Additional tags are retrieved from specific URLs, and the ontology is enriched with this information.

3. **Model Recommendations Using LLM API**: Detailed descriptions and recommendations are generated for each model and its tags using the LLM API.

The combination of these steps allows our system to effectively select the most suitable LLM for a given task, considering the model's specialization, computational requirements, and other relevant factors. The continuous fetching and updating mechanism ensures that the ontology remains current and relevant.

4.3 Leveraging Previous Solutions Ontology

As ProtoAGI evolves, we can leverage a previous solutions ontology to enhance efficiency. This ontology stores past successful configurations of LLMs, algorithms, and datasets, allowing the system to refer to these solutions and reduce computational requirements.

4.3.1 Improved Performance Over Time

The performance of ProtoAGI improves over time as more solutions are added to the previous solutions ontology. Let t represent time and $C(t)$ represent the cumulative computational cost up to time t. The gradient of improved performance, $G(t)$, can be modeled as:

$$G(t) = \frac{\partial P(t)}{\partial t}, \tag{12}$$

where $P(t)$ is the performance function, which improves as more solutions are added. The computational cost $C(t)$ for each new solution is: $C(t) = O(1)$ for retrieval $+O$ (evaluation cost) for adding a new solution.

4.3.2 Diminishing Computational Power Requirements

The presence of a previous solutions ontology reduces the need for extensive computation. The system can quickly refer to past successful configurations, minimizing the search space and computational cost. The cost of using the previous solutions ontology is:

$$\text{Previous solutions ontology cost: } O(1). \tag{13}$$

Combining this with our LLM-guided selection approach, the overall computational cost becomes:

$$\text{Optimized cost: } O(|\mathcal{S}|) + \sum_{i=1}^{N} O(\text{evaluation cost}_i), \tag{14}$$

where $|\mathcal{S}| \ll m \cdot n \cdot p$ and N is the number of solutions added to the ontology. As N increases over time, the retrieval cost remains $O(1)$, but the evaluation cost for adding new solutions gradually decreases due to the growing efficiency of the ontology.

In this way, ProtoAGI benefits from both efficient LLM-guided selection and the reduced computational requirements provided by the previous solutions ontology, with performance improving as more solutions are added.

4.4 Continuous Fetching and Updating

The LLM ontology is designed to be continuously fetched and updated, reflecting the dynamic nature of the AI landscape. As new models are developed and existing models are updated, the ontology can incorporate these changes, ensuring that ProtoAGI always has access to the most up-to-date information. This capability is crucial for maintaining the system's relevance and effectiveness in selecting the best LLM for a given task.

4.4.1 First Iteration and Token Utilization

The first iteration of the LLM ontology involved an extensive process of data processing and enrichment, requiring the use of over a million tokens. This significant utilization of tokens was essential for capturing the comprehensive details and nuances of each model, ensuring both accuracy and depth in the ontology.

ISSN 2007-9737

4.4.2 Data Processing

To begin with, the initial data extraction phase involved parsing large amounts of HTML content from the Ollama repository. This repository contains a vast array of information on various LLMs, including their descriptions, usage statistics, and metadata such as tags and last update times. The PHP script used in this phase had to efficiently handle and process this extensive dataset to extract relevant information accurately.

4.4.3 Token Utilization for Data Enrichment

After the initial extraction, the next phase focused on data enrichment. This step was crucial for adding more context and detailed descriptions to each LLM entry. By utilizing a significant number of tokens, we were able to:

– Generate comprehensive descriptions that cover various aspects of each LLM, including their capabilities, specializations, and ideal use cases.

– Extract and include additional metadata such as tags, which help in categorizing and understanding the specific features and functionalities of each model.

– Ensure that the data is structured in a way that facilitates easy retrieval and application within the ProtoAGI system.

4.4.4 Ensuring Robustness and Scalability

The intensive token usage was not merely for immediate accuracy but also to lay a solid foundation for future iterations of the ontology. By creating a detailed and well-organized initial dataset, we ensured that the ontology is robust and scalable. This means that as new LLMs are developed and added to the Ollama repository, the ontology can be updated and expanded without compromising its integrity or performance.

4.4.5 Continuous Improvement and Updates

Given the rapidly evolving nature of AI and LLM technologies, it is imperative that the ontology remains current. The foundational work done in the first iteration supports continuous updates and improvements. As new models are introduced and existing models are updated, the ontology can seamlessly incorporate these changes, maintaining its relevance and utility for the ProtoAGI system.

5 Results

The resulting Large Language Model (LLM) ontology is structured as a comprehensive JSON file, consisting of a total of **20,283 elements**. This file provides a well-organized, machine-readable repository of information regarding various LLMs, their capabilities, specializations, and metadata, designed to facilitate the selection of the most suitable models for specific tasks. The full ontology can be accessed at the following address[1].

5.1 Structure of the JSON File

The JSON file encodes detailed information about each LLM, which is organized in a modular structure for efficient access and retrieval. The key fields included in the ontology are as follows:

– **Model Name**: A string representing the unique name of the LLM.

– **Description**: A comprehensive description of the LLM, highlighting its key capabilities, intended use cases, and performance characteristics.

– **URL**: A link to additional information about the LLM, including documentation and source repositories.

– **Metadata**:

  – **Pulls**: An integer value representing the number of times the model has been pulled or used from its repository, providing an indication of its popularity and utilization.

[1] trai-l.com/AIG/parser/JSON/processed_llm_ontology.json

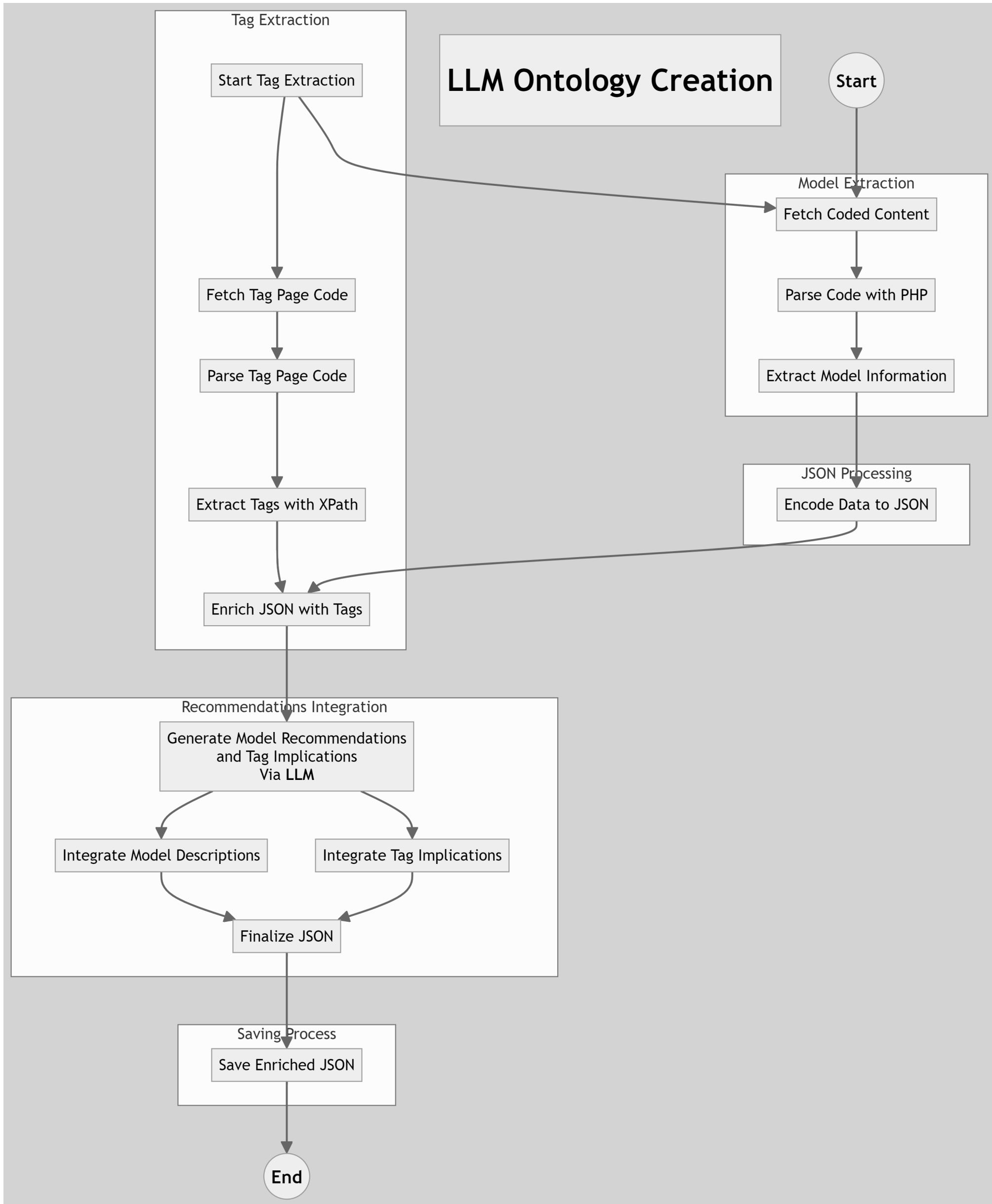

Fig. 2. Detailed flowchart for LLM ontology creation

– **Tags**: A list of categorical tags associated with the model, each providing insights into its functionalities, such as “NLP”, “Text Generation”, “Multimodal”, etc. This allows for efficient filtering and categorization.

– **Last Update**: A timestamp that specifies the last update time of the model, ensuring users are informed of the model’s most recent state.

– **Tag Descriptions**: A set of detailed explanations for each tag associated with the model, outlining the implications and relevance of each tag to the model’s performance or intended applications.

5.2 Processing and Token Utilization

The first iteration of this ontology involved extensive data processing, using over **1 million tokens** to capture the detailed information and metadata for each LLM. The process included:

– **Data extraction** from multiple sources, including model repositories and documentation.

– **Ontology enrichment** through the incorporation of additional metadata, such as tags and usage statistics.

– **Model recommendation generation** using an LLM API, which provided descriptions and tag implications based on extracted data.

This large-scale token processing ensured comprehensive coverage of each LLM's attributes, enabling the ontology to serve as a robust foundation for automating LLM selection within the ProtoAGI system.

5.3 Technical Details of the Ontology

With 20,283 total elements, the ontology covers a broad range of LLMs across various categories, providing detailed descriptions and metadata for each. The large number of elements allows for fine-grained model selection and efficient retrieval of information tailored to diverse AI tasks. The following summarizes the content:

– **LLMs covered**: The ontology includes models for natural language processing (NLP), text generation, multimodal systems, and other AI applications.

– **Metadata depth**: Each model entry contains multiple layers of metadata, providing insights into model performance, popularity, and applicability in different contexts.

– **Extensibility**: The ontology is designed for continuous updates, allowing new models and tags to be added with minimal changes to the underlying structure.

This extensive effort ensures that the ontology provides a scalable, dynamic, and comprehensive repository, laying the groundwork for efficient model selection and recommendation within the ProtoAGI system.

6 Discussion

The creation of an efficient LLM ontology introduces a transformative approach to addressing some of the most critical challenges in Artificial General Intelligence (AGI) research, especially concerning the exponential growth of computational requirements. As AI models continue to increase in complexity, the need for scalable and efficient resource management becomes paramount.

The LLM ontology developed for ProtoAGI significantly contributes to this goal by optimizing the model selection process, thereby reducing the computational overhead traditionally associated with brute force search methods. One of the primary contributions of the LLM ontology is its ability to intelligently manage computational resources. By curating a structured knowledge base of LLMs, including metadata such as model performance, capabilities, and resource consumption, the ontology enables ProtoAGI to make informed decisions when selecting models for specific tasks. This ensures that only the most appropriate models are chosen, avoiding redundant or resource-intensive computations.

This model selection optimization drastically reduces the computational load, making it a critical tool for scaling AGI systems. In the context of supercomputing, the LLM ontology becomes even more important. As we move toward the next generation of AI systems, the role of supercomputers in training and deploying massive AI models will become more prevalent.

Supercomputers equipped with hundreds of thousands of cores and GPUs are uniquely positioned to handle the computational demands of AGI systems. However, even with access to such immense computational power, resource efficiency remains vital.

The ontology enables ProtoAGI to take full advantage of supercomputing infrastructure by selecting models that best balance performance and resource use, ensuring that computational power is not wasted on suboptimal models. This efficiency allows AGI systems to scale effectively, leveraging supercomputers to perform vast numbers of parallel computations while maintaining high performance.

Additionally, the previous solutions ontology further amplifies the system's efficiency by allowing ProtoAGI to learn from its prior configurations. This repository of successful model-algorithm-dataset combinations allows the system to bypass unnecessary evaluations of previously tested solutions, significantly reducing both the search space and computational cost.

In a supercomputing environment, where every computation may consume significant time and energy, such optimizations are crucial for maintaining system performance and throughput. The LLM ontology's efficiency improvements are not just a technical advancement—they carry broader implications for the scalability and deployment of AGI systems in real-world applications.

As AGI moves closer to widespread deployment across sectors such as healthcare, finance, autonomous systems, and scientific research, it is essential that these systems are capable of operating efficiently in constrained environments, where computational resources are either limited or highly expensive. The ability to dynamically adjust model selection based on the available resources ensures that AGI systems remain adaptable, making them more practical for both research and industry.

Supercomputing also plays a critical role in accelerating AGI development. Training large-scale LLMs, simulating neural networks with human-like cognitive abilities, and processing immense datasets all require computational resources that far exceed the capacity of traditional computing environments.

The LLM ontology provides a pathway for maximizing the utility of supercomputing infrastructure by reducing the resource requirements for both training and inference tasks. As AI models and AGI systems grow larger, more intricate, and more reliant on massive datasets, the role of efficient resource allocation, powered by supercomputers, will become increasingly central to the success of AGI.

Moreover, the continuous updating capability of the LLM ontology ensures that ProtoAGI stays current with the latest advancements in LLMs, datasets, and algorithms. As new models are developed and existing models are improved, the ontology can be updated dynamically, ensuring that ProtoAGI remains relevant and at the cutting edge of AGI research. This continuous integration of new data not only improves ProtoAGI's model selection but also helps the system evolve as the field of AI progresses.

From a broader perspective, the LLM ontology has profound implications for the general AI and machine learning communities. By providing a structured, continuously updated repository of LLM capabilities, it enables researchers and developers to leverage this resource for a wide range of applications. The ontology can support everything from educational tools and research projects to industrial-scale AI deployments.

Its ability to offer insights into the strengths and weaknesses of existing models can foster innovation and collaboration across various AI domains, leading to the development of more specialized and efficient AI models. Finally, as AI research moves forward, the integration of supercomputing with intelligent, resource-efficient systems like ProtoAGI will be critical.

The LLM ontology provides a template for how supercomputing resources can be managed more effectively, ensuring that as AGI systems scale, they can maintain efficiency, reduce energy consumption, and maximize performance. This is essential for enabling AGI to fulfill its potential across a wide array of sectors, while ensuring that the ever-increasing computational demands do not become a bottleneck for future advancements.

7 Conclusion

In conclusion, the development of the LLM ontology marks a significant advancement in the ProtoAGI system. By systematically organizing and enriching information about various LLMs, this ontology facilitates the efficient selection of the

ISSN 2007-9737

most suitable models for specific tasks. This process not only improves the overall efficiency of the ProtoAGI system but also ensures that it remains adaptable and up-to-date with the latest advancements in AI models.

The LLM ontology also serves as a valuable resource for the broader AI community. Its structured and detailed repository of LLM information can be utilized in various contexts, including educational purposes, research, and practical applications in industry. By providing a reliable and continuously updated source of model capabilities, the ontology supports the development and deployment of advanced AI systems across multiple domains.

Overall, the development of the LLM ontology represents a significant step towards realizing the vision of AGI. By addressing the critical challenges of model selection and resource management, this ontology provides a robust framework for advancing AGI technologies that are both efficient and adaptable, ultimately contributing to the broader goal of achieving human-like cognitive capabilities in artificial systems. The benefits of this ontology extend beyond ProtoAGI, offering valuable insights and resources that can be leveraged by the entire AI community to foster further advancements in the field.

References

1. **Azam, M., Rafiq, T., Gul-Naz, F., Ghafoor, M., Un-Nisa, M., Malik, H. (2024).** A novel model of narrative memory for conscious agents. International Journal of Information Systems and Computer Technologies, Vol. 3, No. 1, pp. 12–22. DOI: 10.58325/ijisct.003.01.0080.

2. **Besold, T. R., Lamb, L., Icard, T. F., Miikkulainen, R. (2015).** Proceedings of the 10th international workshop on neural-symbolic learning and reasoning.

3. **Fei, N., Lu, Z., Gao, Y., Yang, G., Huo, Y., Wen, J., Lu, H., Song, R., Gao, X., Xiang, T., Sun, H., Wen, J. R. (2022).** Towards artificial general intelligence via a multimodal foundation model. Nature Communications, Vol. 13, No. 1. DOI: 10.1038/s41467-022-30761-2.

4. **Lebiere, C. (2011).** Profile: The FMS cognitive modeling group at carnegie mellon university. The IEEE Intelligent Informatics Bulletin, Vol. 12, No. 1, pp. 1–3.

5. **Naudé, W., Dimitri, N. (2019).** The race for an artificial general intelligence: implications for public policy. AI and Society, Vol. 35, No. 2, pp. 367–379. DOI: 10.1007/s00146-019-00887-x.

6. **Ollama (2024).** Large language models repository. https://ollama.com/library.

7. **Rayhan, A., Rayhan, R., Rayhan, S. (2023).** Artificial general intelligence: Roadmap to achieving human-level capabilities. DOI: 10.13140/RG.2.2.33680.79361/1.

8. **Shevlin, H. (2019).** The limits of AGI: Challenges and perspectives.

9. **Sonko, S., Adewusi, A. O., Obi, O. C., Onwusinkwue, S., Atadoga, A. (2024).** A critical review towards artificial general intelligence: Challenges, ethical considerations, and the path forward. World Journal of Advanced Research and Reviews, Vol. 21, No. 3, pp. 1262–1268. DOI: 10.30574/wjarr.2024.21.3.0817.

10. **Tong, J. C. H., Hsu, Y. F., Liau, C. J. (2024).** An exploring study on building affective artificial intelligence by neural-symbolic computing. Proceedings of the AAAI Symposium Series, Vol. 3, No. 1, pp. 592–593. DOI: 10.1609/aaaiss.v3i1.31288.

11. **Voss, P., Jovanovic, M. (2023).** Why we don't have AGI yet. DOI: 10.48550/ARXIV.2308.03598.

12. **Yamakawa, H. (2021).** The whole brain architecture approach: Accelerating the development of artificial general intelligence by referring to the brain. Neural Networks, Vol. 144, pp. 478–495. DOI: 10.1016/j.neunet.2021.09.004.

Article received on 11/05/2024; accepted on 09/07/2024.
**Corresponding author is Grigori Sidorov.*