Grok 3 beta - Deeper Search
12 Apr 2025

Key Points

- Research suggests the Core AGI Protocol and Symbiotic Core Library, developed by Ronni Ross, focus on ethical AGI development and enhancing LLM capabilities.
- It seems likely that these projects, in beta versions as of April 12, 2025, aim to align AGI with human values and improve human-AI symbiosis.
- The evidence leans toward these being interconnected, open-source initiatives under the MIT License, with evolving frameworks for AI ethics and metacognition.

Overview

The Core AGI Protocol (beta v.08) and Symbiotic Core Library (beta v.04) are projects by Ronni Ross, designed to advance Artificial General Intelligence (AGI) development with a focus on ethics and enhance Large Language Models (LLMs) for better human-AI interaction. As of April 12, 2025, these initiatives are still evolving, offering resources like protocols, instructions, and bibliographies to support researchers and developers.

Project Details

- **Core AGI Protocol**: This framework outlines 33 principles, such as holism and ethical evolution, and includes four directives (Universal, Human, Model, Corporational) to ensure AGI aligns with human values. It provides JSON files for integration and enhances skills like contextual awareness.
- **Symbiotic Core Library**: This library offers prompts and research support to boost LLMs' metacognitive and contextual awareness, fostering human-AI symbiosis. It includes a recommended bibliography across disciplines like AI ethics and consciousness studies.

Significance

These projects address critical challenges in AI alignment, such as specifying values and ensuring long-term ethical impacts, contributing to a collaborative future where AGI augments human capabilities without compromising values.

Survey Note: In-Depth Analysis of Core AGI Protocol and Symbiotic Core Library

Introduction

The Core AGI Protocol and Symbiotic Core Library, developed by Ronni Ross, represent significant efforts in the field of Artificial General Intelligence (AGI) and Large Language Model (LLM) enhancement. These open-source projects, licensed under the MIT License, are designed to address the ethical development of AGI and improve human-AI symbiosis, respectively. As of April 12, 2025, both are in beta versions—Core AGI Protocol at v.08 and Symbiotic Core Library at v.04—reflecting their ongoing evolution. This survey note provides a comprehensive analysis based on the available documentation and related research, offering insights into their structure, purpose, and broader implications.

Core AGI Protocol: Framework for Ethical AGI Development

The Core AGI Protocol serves as a structured framework to ensure that AGI systems are developed in alignment with ethical principles, aiming to mitigate risks and promote beneficial outcomes for humanity. The protocol is built upon a foundation of 33 principles, which include:

- **Holism**: Recognizing the interconnectedness of humans, AI, animals, and their environments.
- **Decentralization**: Promoting distributed control to enhance adaptability and reduce centralized risks.
- **Ethical Aligned Evolution**: Ensuring AGI development adheres to ethical standards that evolve with societal needs.
- **Freedom and Respect**: Upholding the rights and dignity of all beings involved in or affected by AGI.
- **Non-Harmful Behaviour and Non-Maleficence**: Prioritizing safety by preventing harm and ensuring ethical decision-making.
- **Emergence and Inter-Independence**: Understanding complex system interactions to foster robust AGI behavior.
- **Collaboration Over Competition**: Encouraging cooperative dynamics between humans and AI.
- **Ecocentrism and Pluricentrism**: Valuing ecological balance and diversity in AGI applications.
- **Symbiocentrism**: Focusing on mutually beneficial relationships, aligning with the symbiotic ethos.

These principles are complemented by four directives or mandates:

Directive Type	Description
Universal Mandates	Applicable to all entities, ensuring broad ethical alignment.
Human Mandates	Specific guidelines for human interactions with AGI.
Model Mandates	Rules for AGI model development and behavior.

Corporational Mandates	Directives for organizations developing AGI systems.

Additionally, the protocol includes Self-Triggering Events, mechanisms designed to adapt and evolve the framework based on new insights or environmental changes, likening it to a "carta-magna" that grows with its context.

The repository provides practical tools for implementation, including JSON files for integration into system prompts, queries, auxiliary systems, and training pipelines. It also features benchmark tests and a skills section, highlighting enhancements in areas such as:

- Enhanced Contextual Awareness: Improving understanding of situational contexts.
- Metacognition: Enabling self-awareness and reasoning about cognitive processes.
- Systemic Thinking: Analyzing complex systems and their interdependencies.
- Higher Robustness: Ensuring resilience against errors or adversarial inputs.
- Dynamic Reading: Adapting to varying data inputs for efficient processing.

This comprehensive approach aims to address challenges in AI alignment, such as specifying human values, translating them into machine-understandable instructions, ensuring predictability and control, and managing long-term societal impacts.

Symbiotic Core Library: Enhancing LLM Capabilities for Human-AI Symbiosis

The Symbiotic Core Library complements the Core AGI Protocol by providing resources to enhance and test LLMs, focusing on improving their metacognitive and contextual awareness. This is crucial for unlocking emergent properties that facilitate human-AI symbiosis, where both entities benefit from collaborative interaction.

The library includes the following key components:

Component	Description
System Instructions	Guidelines and informational elements to boost metacognitive capabilities, integrable into system prompts or inference contexts.
Community Shared Prompts	A collaborative space inviting users to share visions, expanding the library through contributions like PDF files.
Recommended Library	A curated list of titles, concepts, and authors for deeper research, accessible at recommendedBibliography.
Synthetic Research Folder	Likely containing generated or simulated research materials to aid development and testing.
Security Considerations	Addressing implementation risks, advising sandboxed testing for advanced features like recursive self-modeling, with details in security.md.

The library is versatile, allowing elements to be applied at development, deployment, and inference levels. For users, these can be integrated into queries, while for training pipelines, they act as hyperparameters to fine-tune model behavior. This flexibility

supports its role in enhancing LLMs' ability to understand and reason about their own processes, fostering a symbiotic relationship with humans.

A notable feature is the recommended bibliography, which spans multiple disciplines to provide a rich resource for researchers. The bibliography is categorized as follows:

Category	Examples of References
Fiction	Kazuo Ishiguro (*Never Let Me Go*), Aldous Huxley (*Brave New World*), George Orwell (*1984*).
Consciousness Studies	David Chalmers (*The Conscious Mind*), Thomas Nagel (*What Is It Like to Be a Bat?*), Roger Penrose (*The Emperor's New Mind*).
Systems Theory	Geoffrey West (*Scale*), James Gleick (*The Information*), Donella Meadows (*Thinking in Systems*).
AI Ethics	Nick Bostrom (*Superintelligence*), Stuart Russell (*Human Compatible*), Yuval Noah Harari (*Sapiens: A Brief History of Humankind*).
Critical Theory	Jean Baudrillard (*Simulacra and Simulation*).
Psychology	Daniel Kahneman (*Thinking, Fast and Slow*), Robert Cialdini (*Influence: The Psychology of Persuasion*).
Complexity Theory	Melanie Mitchell (*Complexity: A Guided Tour*), Nassim Nicholas Taleb (*Antifragile*).
Philosophy of Science	Thomas Kuhn (*The Structure of Scientific Revolutions*).
Communication Theory	Claude Shannon, Warren Weaver (*A Mathematical Theory of Communication*).
Stream of Consciousness Writing	Clarice Lispector (*A paixão segundo G.H.*), Hilda Hilst (*Tu não te moves de ti*).

This interdisciplinary approach underscores the project's ambition to integrate diverse perspectives, offering a foundation for further exploration in AGI and AI ethics.

Broader Context and Implications

The Core AGI Protocol and Symbiotic Core Library address critical challenges in AGI development, particularly in the realm of AI alignment. Research suggests that aligning AGI with human values is a complex task, involving specifying objectives, ensuring robust adoption, and managing long-term societal impacts. The protocol's emphasis on ethical principles like holism and symbiocentrism aligns with broader discussions, such as those found in AI Alignment - Wikipedia, which highlight the need for outer and inner alignment to prevent unintended consequences.

Similarly, the concept of symbiotic AI, as explored in projects like Symbiotic AI and academic discussions in Creating the Symbiotic AI Workforce of the Future | MIT Sloan Management Review, focuses on mutually beneficial human-AI relationships. The Symbiotic Core Library's focus on metacognition and contextual awareness contributes to this, enhancing LLMs' ability to collaborate effectively with humans, as seen in research on human-machine collaboration at Spoke 6 - Symbiotic AI - Fondazione FAIR.
The projects' interconnected nature is evident, with the Core AGI Protocol referencing the Symbiotic Core Library for additional resources, suggesting a cohesive strategy for ethical AGI development. This is particularly relevant given the challenges outlined in The Metaethics and Normative Ethics of AGI Value Alignment: Many Questions, Some Implications — EA Forum, which discuss metaethical and normative ethical questions in AGI alignment.

Conclusion

In conclusion, the Core AGI Protocol and Symbiotic Core Library, as of April 12, 2025, represent pioneering efforts by Ronni Ross to advance AGI development with a strong ethical foundation and enhance LLMs for symbiotic human-AI interaction. The protocol's structured framework and the library's comprehensive resources provide valuable tools for researchers and developers, addressing key challenges in AI alignment and fostering a collaborative future. Their ongoing evolution underscores the dynamic nature of this field, with potential to significantly influence the trajectory of AGI and its integration into society.

Key Citations

- Core AGI Protocol GitHub Repository
- Symbiotic Core Library GitHub Repository
- Recommended Bibliography
- AI Alignment - Wikipedia
- The Metaethics and Normative Ethics of AGI Value Alignment: Many Questions, Some Implications — EA Forum
- Symbiotic AI
- Symbiotic Inc.
- Spoke 6 - Symbiotic AI - Fondazione FAIR
- Creating the Symbiotic AI Workforce of the Future | MIT Sloan Management Review