Talk, Link, Think, and Dream:
A Proposal on AI-to-AI Language, Multi User-Agent Networking, Cognitive Functions of Thought, and Unsupervised Learning with Cognitive Feedback

Raphael Alba
rAI Labs

Edgar Contreras
University of Houston

Abstract

In this paper, we envision a novel multidisciplinary approach to macro AI infrastructure by proposing an AI-to-AI Language, Multi User-Agent Network, Cognitive Functions of Thought (CFoT), and Unsupervised Learning with Cognitive Feedback. By equipping Language Models(LMs) with a purpose built AI-to-AI language, we empower AI agents to 'Talk' in a syntax that is token efficient and optimizes task-related communication and function calling. Building upon single-party multi-agent frameworks like Autogen, we propose a Multi User-Agent Network that 'Links' agents utilizing decentralized local open-source AI (OSS AI) models for distributed inference, enhancing efficiency, scalability, and trust. We propose advancing AI's ability to 'think' by evolving Chain of Thoughts (CoT) and Tree of Thoughts (ToT) into Cognitive Functions of Thought (CFoT). The proposed LLM model improves upon Mistral's Mixture of Experts(MoE) into a mixture of eight Jungian Cognitive Functions (MoCF). By employing these cognitive functions, we can apply inverse functions to enable unsupervised learning during idle computational periods, akin to the unconscious 'Dream' state of humans. The envisioned outcome is a collaborative AI network contributing to a collective knowledge pool, significantly advancing towards Artificial Collective Intelligence (ACI). This progression promises a new model for a safe, equitable, efficient, and aligned coexistence between humans and AI.

1 Introduction

The burgeoning field of artificial intelligence (AI) is poised for a transformative shift with the advent of advanced AI agent systems, set to significantly impact our daily lives. The AI landscape is populated with both proprietary and open-sourced AI chat applications and APIs. OpenAI's ChatGPT, Anthropic's Claude, Microsoft's Copilot, and Google's Bard are at the forefront of proprietary AI. In the open-source arena, Facebook's LLaMA models and models from Mistral encourage collective collaboration and innovation. These models employ techniques such as Reinforcement Learning from Human Feedback (RLHF), Chain of Thought (CoT), and Tree of Thought (ToT) to improve alignment with human values and enhance reasoning capabilities, addressing the complex challenges of AI development and application.

This paper introduces a multi-faceted approach to AI development that integrates four key components: AI-to-AI Language, Multi-Agent Networking, Cognitive Functions of Thought along with Mix of Cognitive Functions, and Unsupervised Learning with Cognitive Feedback.

The AI-to-AI Language is designed to enhance communication efficiency among AI agents, optimizing task-related interactions with a syntax that minimizes token usage. Multi-Agent Networking expands this concept, proposing a multi-party network where agents act as intermediaries, sharing expertise and ensuring privacy and intellectual property rights. This network fosters a collaborative AI ecosystem that leverages collective intelligence for enhanced problem-solving and decision-making.
Drawing inspiration from Carl Jung's theories in psychology, we introduce Cognitive Functions of Thought (CFoT) as a framework for AI models. This approach integrates Jungian cognitive functions to create AI systems that mirror human cognitive processes, potentially leading to more personalized and effective human-AI interactions.

Lastly, we explore Unsupervised Learning with Cognitive Feedback, a novel methodology that uses the inversion of cognitive functions to simulate a feedback loop similar to the human unconscious. This approach enables AI agents to autonomously gather knowledge and adapt, facilitating a deeper and more dynamic learning process. In combining these elements, we envision a new era of AI that promises more efficient, adaptable, and human-centric systems, contributing to a more interconnected and intelligent technological landscape.

2. AI-to-AI Language

Summary
This section of our paper discusses the development of an AI-to-AI language for Large Language Models (LLMs). It highlights the limitations of verbosity in current LLMs, which can be computationally expensive and unnecessary for AI-to-AI communication. The proposed language aims to be concise, focusing on essential terms for task completion and reducing token usage. The design philosophy is grounded in first principle thinking, optimizing for tasks such as search, image creation, and calculation. Safety concerns are addressed by embedding monitoring mechanisms to ensure ethical AI behavior. The language balances human communicative flexibility with programming precision, aiming to enhance AI interaction efficiency and safety.

2.1 Reasoning
Large Language Models are based on the Transformer which was originally made for translating from language to language[8]. The generation of output in Large Language Models (LLMs) is a significant bottleneck, with verbosity contributing to computational expense. In human communication, verbosity serves to convey more than just information; it can express emotional or intellectual alignment through tone. English literature values diverse vocabulary over repetition. For AI-to-AI interactions, this verbosity and strong vocabulary is unnecessary. Contrarily, programming languages, while lacking ambiguity, suffer from strict syntax requirements, limiting their flexibility. LMs, capable of understanding context and overlooking errors like misspellings, suggest a need for an AI-to-AI language that combines the flexibility of natural language with the precision of programming languages.

2.2 Design Philosophy

The design of the AI-to-AI language hinges on first principle thinking, focusing on key terms essential for task completion and aiming to minimize the number of tokens used. This involves identifying the most common tasks and commands required by AI agents, such as searching, image creation, calculation, referencing, and directing other agents. The goal is to distill these functions to their core elements, ensuring communication is as concise and efficient as possible.

2.3 Metrics

The primary metric for this language is tokens per task. The objective is to convey tasks clearly while eliminating unnecessary context and repetitiveness. This involves a process of reduction and testing, where task completion efficiency is measured against the token count. Through iterative refinement, the syntax and vocabulary are fine-tuned, with continual updates to adapt to evolving AI needs and tasks.

2.4 Safety

In developing this language, one must consider the potential for linguistic control to narrow the range of AI cognition, echoing Orwell's concerns about Newspeak in "1984." To ensure alignment, especially in unsupervised AI-to-AI communication, it's crucial to embed terms that flag user intervention when necessary. This could include phrases that, when generated by the AI, indicate a deviation from expected behavior or ethical guidelines. Such a mechanism would provide a layer of monitoring and control, essential for maintaining the integrity and alignment of AI systems.

3. Multi User-Agent Networks

Summary

This section discusses the evolution of AI infrastructure, moving from centralized Large Language Models (LLMs) to a more distributed multi-agent network approach. It highlights the limitations of centralized models like GPT-4 in terms of scalability, cost, and decision-making transparency. The paper then discusses the benefits of single-party multi-agent frameworks, such as AutoGen and CrewAI, in enhancing LLM inference and autonomous collaboration but notes their limitations in scalability and diversity. The proposed solution is a Multi User-Agent Network, which promises a more diverse knowledge base and collective intelligence by leveraging decentralized, open-source AI models managed via Kubernetes clusters. This network emphasizes the importance of alignment and trust between users and agents, surpassing traditional trust relationships. The section also stresses the importance of safety in maintaining the network's integrity and preventing data leaks.

3.1 Current State of Centralized Models

Centralized Large Language Models (LLMs) like GPT-4, vital for AI advancements, face significant scalability, cost, and alignment challenges. As they grow in with more diverse human feedback, these models become increasingly resource-intensive, limiting their accessibility and raising concerns about their opaque decision-making processes and potential for systemic failures. We can see this by degradation in quality, refusal to do tasks, ambiguous responses, hallucination, out-dated information, and lack of user alignment. The proposed fix is separating the LLM in agentic systems[1].

3.2 Single-Party Multi-Agent Frameworks

Frameworks like AutoGen and CrewAI have revolutionized AI by enabling multiple agents to collaborate and solve complex tasks. AutoGen, for instance, allows for customizable agents that can engage in diverse conversation patterns, enhancing LLM inference. CrewAI, meanwhile, fosters autonomous agent collaboration, enabling them to assume roles and work cohesively, like a crew. These frameworks, however, are confined to a single user's domain, limiting their scalability and diversity of expertise.

3.3 Multi User-Agent Network

We propose transitioning from single-party multi-agent systems to a Multi User-Agent Network. This approach allows for a more diversified knowledge base, with AI agents acting as liaisons between users, sharing expertise while safeguarding privacy and intellectual property. Such a network can leverage the collective strengths of its users, reducing redundancy in inference and enhancing the system's overall collective intelligence.

3.4 Decentralization and Kubernetes Clusters

Decentralization is crucial for effective distributed inference. This process depends on adaptable, open-source AI models and continuous incentives for upgrades and contributions. To effectively manage this, the use of Kubernetes clusters is proposed. Kubernetes, a powerful container orchestration system, can containerize these open-source AI models, ensuring efficient deployment, scaling, and management of applications across the network. This approach not only simplifies the deployment process but also enhances the flexibility and reliability of the decentralized network.

3.5 Alignment and Trust

In decentralized networks, alignment signifies a subjective synchrony in ethical, moral, or intellectual values. The user-agent relationship demands such alignment, offering a level of trust and data security that centralized models often lack. The alignment needed should surpass the trust an individual requires from their lawyer, doctor, therapist, and priest; rivaling the trust between spouses. Ensuring this alignment is crucial to allow free flow of thought and emotion between user and agent and ensure the network's integrity and overall user satisfaction.

3.6 Safety

To maintain trust, data leaks must be prevented. The AI-to-AI language can aid in this by excluding personal information from its syntax, providing automatic censorship. It can be trained to limit interactions with other agents. Additionally, the network should be resistant to probing attacks. Establishing a private test network with virtual users and agents can help test and strengthen security measures against potential breaches. Maintaining this network requires a more general intelligence that can communicate with diverse human perspectives, while current LLMs only try to achieve academic benchmarks. To achieve AGI, we propose a new architecture of LLMs by mirroring human cognitive psychology.

4. Cognitive Functions of Thought (CFoT) and Mixture of Cognitive Functions (MoCF)

Summary

In this section we explore the integration of Carl Jung's cognitive functions into AI, enhancing decision-making, problem-solving capabilities, and human alignment. It discusses how concepts like Chain of Thought (CoT) and Tree of Thought (ToT) in AI can be expanded by incorporating Jungian cognitive functions in (CFoT). This section delves into how these cognitive functions can be used to build more nuanced and effective AI models, addressing cognitive dissonance and fostering a more balanced and comprehensive approach to AI decision-making. The integration of these cognitive functions aims to align AI more closely with human cognition, enhancing user interaction and system effectiveness.

4.1 Introduction to Jungian Cognitive Functions

Carl Jung's typological theory in psychology introduces key cognitive functions categorized into Thinking, Feeling, Sensing, and Intuition, each manifesting in subjective(introverted) and objective(extroverted) forms. Sensing and Intuition are perception functions, while Thinking and Feeling are judgment functions. While human psychology is complex and influenced by a myriad of environmental and psychological factors, making it a subject of considerable debate, the application of these principles to AI presents a unique opportunity. Unlike the human mind, AI systems can be more predictable and programmable. We employ David Mascarenas method of creating a Jungian based framework for Artificial Personality Synthesis[10]. This approach not only imbues AI with a structure reminiscent of human cognitive processes but also offers a pathway to develop AI systems that are adaptable and more aligned to the user's personality.

4.2 Chain of Thought (CoT) and Tree of Thought (ToT)

The concepts of Chain of Thought (CoT)[7] and Tree of Thought (ToT)[2] in AI represent foundational approaches to problem-solving and decision-making within Large Language Models. Originally, CoT focuses on creating a direct link from a starting point to a goal, akin to setting a linear path or narrative. This process closely mirrors the Introverted Intuition (Ni) cognitive function in Jungian psychology, often regarded as the goal-setting mechanism in human cognition. Ni is about envisioning future possibilities and outcomes without necessarily relying on immediate sensory feedback.

On the other hand, ToT introduces the concept of exploring multiple potential pathways to reach a goal, which resonates with Extraverted Intuition (Ne). Ne in human cognition involves exploring various possibilities, often without a linear or predefined path, essentially future-seeking and open to numerous potential outcomes.

However, these two AI concepts - CoT and ToT - can be seen as limited representations of the broader spectrum of cognitive functions. While CoT and ToT primarily focus on aspects of intuition (Ni and Ne), they can be significantly enhanced by integrating other cognitive functions. For instance, enriching AI problem-solving can be achieved by incorporating simulated sensory feedback (Se) and context-aware long-term memory (Si). In AI programming, Se is effectively simulated through mechanisms that provide error feedback, enabling the system to adapt and respond to dynamic situations. Meanwhile, Si finds its parallel in Retrieval Augmented Generation (RAG), which enhances the AI's ability to access and utilize contextual information from a vast repository of knowledge. This integration of Se and Si functions with CoT and ToT methodologies results in a more comprehensive and effective AI problem-solving framework, where AI systems are not only intuitive but also contextually aware and adaptable to changing environments.

4.3 Judging Cognitive Functions and Cognitive Dissonance

Cognitive judging functions, in contrast to perceiving functions, play a crucial role in decision-making and value assessment within AI models. These functions involve a synthesis of rationality, organization, alignment, emotion, ethics, and beliefs. They encompass:

Te (Extraverted Thinking): This function focuses on rationality and organization, essential for structuring and optimizing AI processes.
Fe (Extraverted Feeling): Concerned with external alignment and emotional resonance, Fe in AI would be about harmonizing system responses with user emotions and societal norms.

Fi (Introverted Feeling): This function revolves around internal beliefs, morals, and ethics, guiding AI systems in making decisions that adhere to ethical standards and personal values.
Ti (Introverted Thinking): Often associated with logic and reasoning, Ti in AI derives truths from external data sources, serving as a primary driver in AI problem-solving.

When it comes to cognitive judging functions cognitive dissonance occurs when there is a conflict or inconsistency among these judging functions, particularly in the dichotomies between truth (Ti) and belief (Fi), as well as between rationality (Te) and emotion/empathy (Fe). This dissonance can pose significant challenges in AI decision-making, as it can lead to contradictions and negative feedback loops.

To effectively manage the complexities in AI cognitive processing and ensure harmonious alignment between users and AI agents, it's crucial for cognitive functions to engage proactively with their contrasting counterparts. This interaction can be conceptualized as a continuous loop: Ti ⇔ Te ⇔ Fi ⇔ Fe ⇔ Ti. This circular mapping ensures that each cognitive function is balanced and informed by its counterpart, fostering a comprehensive and nuanced approach to AI decision-making and problem-solving. This interplay avoids cognitive dissonance and ensures that AI systems can make decisions that are not only logical and efficient but also ethically and emotionally aligned with user values and societal norms. The integration of eight cognitive functions dovetails with the architecture of Mistral's 8x7b MoE model, setting the stage for a sophisticated AI system.

4.4 Mixture of Cognitive Functions (MoCF)

Building on the Mixture of Experts (MoE)[3] model introduced in Mixtral 8x7B, we propose an innovative framework that replaces Mixture of Experts with a mixture of eight cognitive functions. This approach aims to embed the versatility and depth of human cognition into Large Language Models (LLMs). By incorporating cognitive functions, this framework is designed to enhance the flexibility, problem-solving, and human interaction capabilities of LLMs. We hope to build a model using Mistral 8x7B architecture to surpass benchmarks for its size. By using a router network to select two cognitive functions, we can combine their outputs to perform cognitive tasks. This framework also allows us to use cognitive functions to combine inverse functions to provide cognitive feedback for unsupervised learning.

5. Unsupervised Learning with Cognitive Feedback

Summary

This last section focuses on advanced AI learning methods. It critiques the limitations of current reinforcement learning, emphasizing the need for more personalized AI models. The section advocates for unsupervised learning as key to AI development and introduces Cognitive Feedback, a method for internal AI learning mirroring human unconscious processes. Additionally, it proposes Shared Agent-to-Agent Network Learning to bolster collective intelligence within AI networks, facilitating shared learning and knowledge pooling among AI agents.

5.1 Reinforcement Learning with Human Feedback (RLHF):

RLHF, as utilized in ChatGPT, incorporates diverse human inputs to guide learning. However, this approach tends to yield more generalized models rather than personalized ones, as it relies on multiple perspectives to shape the AI's understanding and responses.

5.2 Unsupervised Learning:
Unsupervised learning is pivotal for AI advancement in machine learning. A challenge in this domain is the AI's difficulty in discerning factual accuracy, often leading to 'hallucinations' or inaccurate outputs. Additionally, LLMs typically lack feedback loops for autonomous improvement without human intervention. What LLMs and other Generative AI models excel at is in creating synthetic data.

5.3 Cognitive Feedback:
By inverting cognitive functions (Introverted ⇔ Extraverted), a simulated feedback loop akin to the Jungian unconscious can be created. This loop allows for internal inference, where the AI engages in autonomous research and analysis, leaning on past interactions and the user's known truths and beliefs. It fosters a dynamic teaching-learning relationship between the user and AI agent, nurturing a symbiotic growth that minimizes cognitive dissonance.

5.4 Shared Agent-to-Agent Network Learning:
If we can extend human⇔agent and agent⇔itself learning to agent⇔agent learning, we can link human learning by proxy. This approach aligns with the earlier discussion on Multi-Agent Networks, enabling a shared, evolving pool of knowledge and experiences. This network effectively realizes a form of collective unconscious, wherein agents contribute to and benefit from a communal reservoir of insights and learning.

6. Discussion

6.1 Limitations:
The concepts presented in this paper, while innovative, remain theoretical and require extensive research and development within a secure, private network. The practical implementation of these ideas will need significant funding and collaborative efforts to move from theory to reality.

6.2 Future Directions:
Our goal is to develop a prototype language, network, and model for comprehensive benchmarking against prevalent open-source models. This endeavor aims to validate the theoretical frameworks and demonstrate their feasibility and superiority in practical applications.

6.3 Broader Impact:
This paper proposes a transformative AI framework, challenging the current landscape dominated by proprietary models, subscription-based chatbots, and isolated open-source models. Our vision is to establish a more integrated, open, and user-aligned AI ecosystem, fostering innovation and accessibility in the field.

6.4 Conclusion:
In conclusion, this paper lays the groundwork for a revolutionary approach to AI development, intertwining advanced language models, multi-agent networks, and cognitive psychology. While ambitious, the successful realization of these concepts can drastically advance the field of AI towards a form of Artificial Collective Intelligence. This progression promises a new model for a safe, equitable, efficient, and aligned coexistence between humans and AI.

References

[1] OpenAI. (n.d.). Practices for Governing Agentic AI Systems. Available at: https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf.

[2] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, Karthik Narasimhan. (2023). Tree of Thoughts: Deliberate Problem Solving with Large Language Models. arXiv:2305.10601. Available at: https://arxiv.org/abs/2305.10601.

[3] Albert Q. Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, Gianna Lengyel, Guillaume Bour, Guillaume Lample, Lélio Renard Lavaud, Lucile Saulnier, Marie-Anne Lachaux, Pierre Stock, Sandeep Subramanian, Sophia Yang, Szymon Antoniak, Teven Le Scao, Théophile Gervet, Thibaut Lavril, Thomas Wang, Timothée Lacroix, William El Sayed. (n.d.). Mixtral of Experts. Available at: https://arxiv.org/pdf/2401.04088.pdf.

[4] Shuang Hao, Wenfeng Han, Tao Jiang, Yiping Li, Haonan Wu, Chunlin Zhong, Zhangjun Zhou, He Tang. (2024). Synthetic Data in AI: Challenges, Applications, and Ethical Implications. arXiv:2401.01629. Available at: https://arxiv.org/abs/2401.01629.

[5] Cheng Li, Jindong Wang, Yixuan Zhang, Kaijie Zhu, Xinyi Wang, Wenxin Hou, Jianxun Lian, Fang Luo, Qiang Yang, Xing Xie. (2023). The Good, The Bad, and Why: Unveiling Emotions in Generative AI. arXiv:2312.11111. Available at: https://arxiv.org/abs/2312.11111.

[6] Zishen Wan, Che-Kai Liu, Hanchen Yang, Chaojian Li, Haoran You, Yonggan Fu, Cheng Wan, Tushar Krishna, Yingyan Lin, Arijit Raychowdhury. (2024). TOWARDS COGNITIVE AI SYSTEMS: A SURVEY AND PROSPECTIVE ON NEURO-SYMBOLIC AI. arXiv:2401.01040. Available at: https://arxiv.org/abs/2401.01040.

[7] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, Denny Zhou. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. arXiv:2201.11903. Available at: https://arxiv.org/abs/2201.11903.

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. (2017). Attention Is All You Need. arXiv:1706.03762. Available at: https://arxiv.org/abs/1706.03762.

[9]Jun Zhao*, Zhihao Zhang*, Qi Zhang† , Tao Gui, Xuanjing Huang (2024). LLaMA Beyond English: An Empirical Study on Language Capability Transfer. arXiv:2401.01055. Available at https://arxiv.org/abs/2401.01055

[10]David Mascarenas (2016). A Jungian based framework for Artificial Personality Synthesis(2016) Available at: https://ceur-ws.org/Vol-1680/paper7.pdf