EXPLORING THE LINK BETWEEN META-RATIONALITY AND AI MODELS COMPOSABILITY

Sînică ALBOAIE[1], Daniel SAVA[1], Eduard BOGHIȚĂ[2], Oana COCA[2]

e-mail: oana.coca@agriceda.ro

Abstract

The current approach to AI (Artificial Intelligence) model design involves training a single, massive model with billions of parameters. However, this approach has significant limitations. The self-transforming mind's capacity to embrace paradox and multiple realities offers inspiration for a new approach to AI model design based on many smaller models working together in a compositional architecture. This architecture would allow for greater flexibility, adaptability, transparency and explainability. This article explores these ideas and offers a potential architecture to approach this complex problem. We hope to provide a new direction for AI model design that aligns more with the complexity and nuance of the real world.

Key words: models, meta-rationality, Artificial Intelligence, ComposableAI

In 1982, Robert Kegan (Kegan R., 1982) introduced five stages of human development in his book "Evolving Self: Problem and Process in Human Development". The last two stages are the socialised mind, the self-authoring mind, and the self-transforming mind. Both stages imply some kind of meta-programming in the sense of code that can change itself.

The self-transforming mind is the fifth and final stage. Accordingly, to Kegan (Kegan R., 1982), only 1% of people reach this stage before mid-life. At this stage, a person's sense of self is not tied to specific identities or roles but is constantly created through exploration and interaction. Characteristics of the self-transforming mind include the ability to see life, people, emotions, and relationships as complex and constantly changing. It involves questioning authority and critically examining our thoughts and beliefs. It also means embracing paradox and being comfortable simultaneously holding multiple thoughts, emotions, identities, and ideologies.

Furthermore, meta-rationality differs from rationality as it does not operate on principles or methods, making it unfeasible to learn similarly. Instead, individuals must practice and reflect on cultivating a deeper understanding of meta-rationality.

In conclusion, it is vital to acknowledge the distinct concept of meta-rationality and its potential benefits for problem-solving and decision-making. This could advance the development of more efficient AI systems. Previous research on AI model composability has shown promising empirical results, but our article aims to extend beyond these initial attempts. While it may seem logical that a truly intelligent machine would require some breakthrough technology or interface with the real world in a novel way, recent research from Microsoft suggests otherwise. In a paper titled "HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in HuggingFace" (Shen et al, 2023), the authors share a model that leverages large language models (LLMs) to connect various AI models in machine learning communities to solve complicated AI tasks. They present a framework called HuggingGPT, which uses ChatGPT to conduct task planning, select models based on their function descriptions available in Hugging Face, execute each subtask with the selected AI model, and summarise the response according to the execution results. By leveraging ChatGPT's strong language capability and abundant AI models in Hugging Face, HuggingGPT covers numerous sophisticated AI tasks in different modalities and domains, achieving impressive results in language, vision, speech, and other challenging tasks.

MATERIAL AND METHOD

The self-transforming mind's capacity to embrace paradox and multiple realities is

[1] AXIOLOGIC SAAS, Iași, Romania

[2] Iasi University of Life Sciences, Romania

particularly relevant for working with AI models. The current approach to AI model design involves training a single, massive model with billions of parameters. However, this approach has significant limitations, including an increased risk of overfitting, lack of transparency, and high computational costs.

The self-transforming mind's ability to work with multiple perspectives and ideas can inspire a new approach to AI model design based on many smaller models working together in a compositional architecture. This architecture would allow for greater flexibility, adaptability, transparency and explainability.

In this article, we will explore the potential benefits of this approach and offer a potential architecture for implementing it. By drawing on the principles of the self-transforming mind, we hope to offer a new direction for AI model design that is more aligned with the complexity and nuance of the real world.

This article aims to establish a connection between meta-rationality and AI research. As a result, we explore how individuals who hold fundamental rationalist principles, which are prevalent in scientific and AI communities, may perceive the meta-rationality way of thinking. Supporters of meta-rationality could face initial scepticism or hostility due to its love for paradoxes, or they may find the idea intriguing but implausible. Nevertheless, the article posits that using meta-rationality is indispensable for enhancing one's problem-solving and decision-making abilities, which could inform the development of better AI systems. Meta-rationality entails utilising rationality more efficiently by comprehending problems and potential solutions in broader contexts.

Furthermore, meta-rationality differs from rationality as it does not operate on principles or methods, making it unfeasible to learn similarly. Instead, individuals must practice and reflect on cultivating a deeper understanding of meta-rationality.

In conclusion, it is vital to acknowledge the distinct concept of meta-rationality and its potential benefits for problem-solving and decision-making. This could advance the development of more efficient AI systems.

Previous research on AI model composability has shown promising empirical results, but our article aims to extend beyond these initial attempts. While it may seem logical that a truly intelligent machine would require some breakthrough technology or interface with the real world in a novel way, recent research from Microsoft suggests otherwise. In a paper titled "HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in HuggingFace" (Shen et al, 2023), the authors share a model that leverages large language models (LLMs) to connect various AI models in machine learning communities to solve complicated AI tasks. They present a framework called HuggingGPT, which uses ChatGPT to conduct task planning, select models based on their function descriptions available in Hugging Face, execute each subtask with the selected AI model, and summarise the response according to the execution results. By leveraging ChatGPT's strong language capability and abundant AI models in Hugging Face, HuggingGPT covers numerous sophisticated AI tasks in different modalities and domains, achieving impressive results in language, vision, speech, and other challenging tasks.

RESULTS AND DISCUSSIONS

Expanding on the model's composability approach and under the lens of meta-rationality, we propose the ComposableAI Architecture. While ComposableAI may appear as a messy and inelegant kludge, its brain-like architecture has the potential to be a significant step towards achieving real-life AGI (Artificial General Intelligence). Following the meta-rationality philosophy (Heylighen, F.,1991; Chapman D., 2020), ComposableAI aims to leverage existing AI models and their specific domains and modalities to solve complex tasks more effectively. ComposableAI can connect various AI models, select and execute subtasks with the most appropriate model, and summarise the response by utilising language as a generic interface and a large language model as a controller. This approach offers the possibility to cover numerous sophisticated AI tasks in different domains and modalities and to achieve impressive results in language, vision, speech, and other challenging tasks. Through the ComposableAI Architecture, we propose a new path towards advanced AI that relies on meta-rationality principles and the effective use of existing AI models in a brain-like architecture. "Compositional AI architectures" or ComposableAI refers to designing AI systems where multiple smaller AI models are combined and coordinated to create a larger, more complex AI solution. These models can be specialised for specific tasks and adapted or replaced without affecting the entire solution. This approach can be more efficient and adaptable than single, massive AI models with billions of parameters.

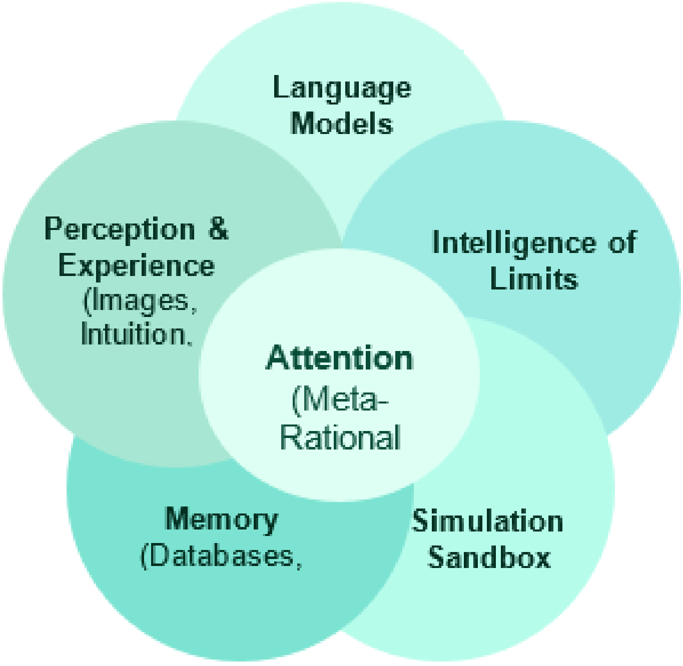

Figure 1 **ComposableAI main components**

ComposableAI adopts a brain-like architecture, which strings together multiple specialised modules to accomplish complex tasks. The prefrontal cortex (Kolb B. et al, 2012) plays a similar role to the Attention central component in the above figure. It is responsible for understanding tasks, planning a process for executing them and using specialised modules. The brain also comprises multiple specialised modules that work together, with the prefrontal cortex acting as a conductor. ComposableAI aims to avoid the challenge of training a single language model for multiple tasks by using one high-level component to direct the actions of many low-level modules. The prefrontal cortex (Kolb B. et al, 2012) is a part of the brain located at the front of the frontal lobe, just behind the forehead. It involves various complex behaviours, including planning, decision-making, problem-solving, cognitive flexibility, working memory, and attention. The prefrontal cortex also regulates emotions, impulse control, and social behaviour. It is considered one of the most evolved brain areas in humans, and its development is closely linked to higher cognitive functioning and executive control.

The Language Model component has a well-defined function of encoding abstract knowledge. It is the base for multiple implicit rational models, providing a robust foundation for embracing meta-rationality in AI systems.

In addition, we propose the inclusion of two more classes of components (we call them intelligence areas), namely, the “Perception and Experience area” and the “Intelligence of Limits” area. Both areas could be composed of various AI models catering to different types of intelligence.

The Perception component is intended to bring together models related to image recognition, mathematical intuition, expertise, and wisdom, among others, to simulate human-like cognitive abilities. Furthermore, the Experience component opens the door for developing models that can simulate emotions, sentiments, and sensations, which are integral to human-like AI systems. The Meta-Rational Controller will orchestrate and integrate these components, ensuring that the models work together seamlessly to perform complex tasks efficiently. In addition to the Language Model component, we propose including two other areas within the ComposableAI architecture that the Meta-Rational Controller would orchestrate.

The first area, "perception," would combine various models related to image recognition and simulation of human intuition in mathematics and expertise and wisdom in other areas. It would also allow for incorporating models that can simulate sentiments, emotions, experiences, and sensations, opening up a new range of capabilities for the ComposableAI system. This multi-model approach aligns with the concept of meta-rationality, which involves leveraging a range of specialised modules to achieve complex tasks, much like the human brain. Integrating multiple models within ComposableAI can increase flexibility and efficiency while expanding its capacity for intelligent problem-solving.

The second area we propose to complement the Language Model component is called "Intelligence of Limits." This area would focus on encoding rules, ethical and legal laws, and other constraints that may exist for various phenomena and concepts. By collecting and encoding this knowledge, the Intelligence of Limits area would aim to provide quick insights into the boundaries and limitations of various domains.

For example, medical diagnoses may have ethical and legal limitations when using certain data or procedures (Rigby M. J., 2019; Hickman, S. E. et al, 2021). By incorporating this knowledge into the Intelligence of Limits area, AI systems could quickly identify potential issues or risks related to a particular diagnosis, helping to ensure that any decisions made by the system comply with relevant laws and regulations.

Similarly, in the finance domain, there may be limits on the amount of risk an investment portfolio can take on. The Intelligence of Limits area could encode these limits and provide alerts when the portfolio exceeds the allowable risk levels.

Overall, the Intelligence of Limits area would play a critical role in ensuring that AI systems operate within the boundaries of various domains, helping to minimise the risk of unintended consequences and ensure that decisions made by the system are both effective and ethical. In Figure 1, we propose two additional areas: the Memory area and the Simulation Sandbox area. The Memory area is relatively straightforward, as it will store long-term memory related to the AI system and various integrations to facilitate the system's interaction with the external world. The Simulation Sandbox area, on the other hand, is a new and still-evolving component that is in the preliminary idea stage. It aims to provide a virtual environment where the AI system can simulate and experiment with different scenarios, allowing it to learn and refine its abilities without risking causing harm in the real world. This area could enhance the safety and reliability of AI systems and accelerate their development. The Simulation Sandbox component, although still in the preliminary idea stage, has the potential to offer an innovative approach for the AI system to improve itself. It could enable the creation of strategies for supervised learning by encoding real-world realities in programs executed within the sandbox. Additionally, the sandbox could generate synthetic data based on the rational understanding selected by the Meta-Rational Controller.

The ComposableAI architecture assumes that the Meta-Rational controller can train new models as needed to augment or replace the predefined models when their performance no longer meets the KPIs set by the system. This means that the controller must be able to evaluate the performance of the individual components and identify areas where new models or updated models are needed.

Furthermore, the Meta-Rational controller must also be able to train and integrate new models into the overall system. This requires an understanding of the existing architecture and the ability to create new components that can be seamlessly integrated into the system.

Additionally, the controller must be able to manage the training of new models, including the collection and processing of data, model selection hand optimisation, and deployment of the trained models. Overall, the ability of the Meta-Rational controller to train and integrate new models is crucial for the long-term success of the ComposableAI architecture. It allows the system to adapt and evolve as new data and new challenges arise, ensuring that it remains effective and relevant in solving complex problems in the real world. Although this approach is not a technological breakthrough, ComposableAI shares similarities with the human brain's composition of specialised modules. Thus, if and when AGI emerges, it is more likely to come from a similar approach, reflecting the complexity of multitudes of intelligence types. The Attention component in the ComposableAI framework acts as a "meta-rational controller" that could utilise AI models like HuggingGPT and various heuristics. These heuristics would be chosen for their greater suitability to address biases and provide better explainability, considering the business constraints of the developed AI systems.

CONCLUSIONS

Meta-rationality has the potential to help overcome issues related to bias and ethical concerns in AI models. By utilising a broader understanding of problems and potential solutions and considering context and impact, meta-rationality can lead to more ethical and unbiased decision-making. It allows for a more holistic approach to AI. The system optimises for a single metric and considers multiple considerations like fairness, privacy, and social impact. Furthermore, the role of the Meta-Rational Controller in ComposableAI can also contribute to explainability in AI models. By orchestrating the interactions between different modules, the Meta-Rational Controller can provide insights into why a certain decision was made, a crucial aspect of explainability. It can help trace the decision-making process and provide a more transparent and interpretable AI system. Meta-rationality principles can help address issues related to bias and ethics in AI models by taking a more holistic approach to decision-making. Additionally, the role of the Meta-Rational Controller in ComposableAI can contribute to explainability by providing insights into the decision-making process.

This paper explored that the self-transforming mind's capacity to work with multiple perspectives and embrace paradox could be particularly relevant for a new approach to AI solutions. Based on many smaller models working together in a compositional architecture, this new approach would allow for greater flexibility, adaptability, and transparency in AI models. By drawing on the principles of the self-transforming mind, we hope to offer a more nuanced and effective way to approach the complexity of the real world. Kegan (Kegan R., 1982) found that a disproportionate number of Stage 5 adults had dabbled in self-transcendent experiences such as psychedelics, meditation, and martial arts. Self-transcendent experiences involve a brief moment where people feel lifted above their day-to-day concerns, and their sense of self fades away as they feel connected to something bigger. These experiences can be important in developing the self-transforming mind, as they allow the self to transcend its boundaries and become part of something larger. This underscores the importance of exploring new ways of thinking about AI model design that is more aligned with the complexity and nuance of the real world. The link between self-transcendent experiences and the self-transforming mind offers a promising research hypothesis for developing more flexible and adaptive AI models.

By exploring the potential connection between self-transcendent experiences and the cognitive and emotional capacities associated with the self-transforming mind, we can identify new ways to enhance the effectiveness and robustness of compositional AI architectures. This hypothesis is based on the observation that individuals who have reached the self-transforming mind stage often report having had self-transcendent experiences.

These experiences may be linked to a more flexible and nuanced approach to problem-solving and understanding the world, a key characteristic of the self-transforming mind. We can identify the cognitive and emotional capacities associated with self-transcendent experiences relevant to AI model design based on insights from psychology and neuroscience. Testing this hypothesis would involve conducting empirical research to investigate the relationship between self-transcendent experiences and the cognitive and emotional capacities of the self-transforming mind.

The goal of this research would be to identify specific cognitive and emotional capacities that are relevant to the development of more effective and sustainable AI models. Overall, the link between self-transcendent experiences and the self-transforming mind offers a novel and potentially fruitful research hypothesis for developing new approaches to AI model design. By exploring this connection, we may identify new ways to enhance compositional AI architectures' flexibility, adaptability, and robustness.

In conclusion, ComposableAI's brain-like architecture offers significant advantages, including flexibility and efficient use of processing time. ComposableAI aims to avoid wasting time and energy training ever-bigger models by only calling up the necessary modules for a given task. This approach is similar to the human brain, which uses its prefrontal cortex for high-level functions while delegating lower-level tasks to other regions. ComposableAI's ability to integrate new models and prioritise processing resources makes it a promising step towards advanced Artificial General Intelligence.

ACKNOWLEGMENTS

This research is co-financed by the European Fund for Regional Development through the Competitiveness Operational Program 2014 – 2020, project "Establishment and implementation of partnerships for the transfer of knowledge between the Iasi Research Institute for Agriculture and Environment and the agricultural business environment", acronym "AGRIECOTEC", SMIS code 119611.

REFERENCES

Chapman D., 2020 - In the cells of the eggplant, Meta-rationality: An introduction, available online at: https://metarationality.com/introduction.

Heylighen, F., 1991 - Cognitive Levels of Evolution: from pre-rational to meta-rational. The Cybernetics of Complex Systems-Self-organization, Evolution and Social Change, 75-91., available online at: https://www.researchgate.net/publication/228594356_Cognitive_Levels_of_Evolution_from_pre-rational_to_meta-rational.

Hickman, S. E., Baxter, G. C., & Gilbert, F. J., 2021 - Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations. British journal of cancer, 125(1), 15-22., avalaible online at: https://doi.org/10.1038/s41416-021-01333-w.

Kegan, R., 1982 - The evolving self: Problem and process in human development. Harvard University Press.

Kolb, B., Mychasiuk, R., Muhammad, A., Li, Y., Frost, D. O., & Gibb, R., 2012 - Experience and the developing prefrontal cortex. Proceedings of the National Academy of Sciences, 109(supplement_2), 17186-17193., available online at: https://doi.org/10.1073/pnas.1121251109.

Rigby, M. J., 2019 - Ethical dimensions of using artificial intelligence in health care. AMA Journal of Ethics, 21(2), 121-124., available online at: https://journalofethics.ama-assn.org/article/ethical-dimensions-using-artificial-intelligence-health-care/2019-02

Shen, Y., Song, K., Tan, X., Li, D., Lu, W., & Zhuang, Y., 2023 - HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in HuggingFace. arXiv preprint arXiv:2303.17580., available online at: https://doi.org/10.48550/arXiv.2303.17580.