Society of Minds: Designing Mindful AGI

From a practical perspective, we see a profound challenge in our quest toward Artificial General Intelligence, and we propose a new model to overcome this. The current consensus leans towards AI <u>not</u> having its own mind in the sense of human consciousness or the ability to evolve such a mind through emergence. Although it's not the focus of mainstream AGI research, the concept of "Society of Mind" in the context of current AGI paradigms is an intriguing area of exploration. The theory, proposed by the Turing Prize awardee Marvin Minsky, describes human intelligence as emerging from the interactions of non-intelligent agents in the brain. This concept suggests that intelligence is the result of many small, simple processes working together, rather than a single, complex process.

In the field of AGI, researchers are exploring various approaches to replicate or simulate aspects of human cognition in machines, but not on the concept of mind per se. For instance, DeepMind's definition of AGI emphasizes that an AGI must be general-purpose, high-achieving, capable of learning a range of tasks, assessing its performance, and asking for assistance when needed. This approach can be aligned with the "Society of Mind"

Max Michaels and Rao Mikkilineni

theory in that it involves creating systems that can perform a variety of tasks, similar to how the human mind operates through the collaboration of many cognitive processes.

Moreover, recent work in AI, such as the exploration of the collective unconscious in AGI, aligns with the "Society of Mind" by considering the integration of shared human experiences and archetypes into AI development. Leading thinkers in the field of AGI who might align with Minsky's concept include Ben Goertzel, known for his work on cognitive architectures, and researchers involved in novel approaches such as functional contextualism in AI. This concept differs from current forms of cognitivism and focuses on how concepts emerge and relate to one another in context, a key aspect of developing AGI with a broader, more human-like understanding.

We believe the concept can be adapted to design networks of mindful AI systems. Our approach suggests embedding AGI with a shared set of universal symbols and myths, which could lead to a more profound understanding and empathy toward human experiences. This reflects the idea of a collective or societal aspect within AI, similar to Minsky's vision of multiple agents contributing to a singular intelligence.

Building upon the Body-Brain-Mind metaphor, we propose a new approach for AI design, which mimics the Body, Brain, and Mind model of carbon-based life, presenting a multi-layered structure in the digital world. It involves three primary layers:

Symbolic Computing Layer (Body): This bottom layer interfaces with the material world, acquiring and processing data into information. It executes well-defined tasks specified by algorithms, encompassing traditional tasks in computing.

Sub-symbolic Computing Layer (Brain): This layer deals with tasks not specifically defined by algorithms, such as processing images, videos, audio, and text. It mimics neural network operations in the brain, converting information into knowledge through optimized algorithm parameters.

Systemic Process Overlay (Mind): The top layer combines symbolic and sub-symbolic computing. It encompasses systemic processes with specified goals, simulating human learning and reasoning, and utilizes a common knowledge representation. This layer integrates functional and non-functional requirements, best practices, policies, and reasoning principles inspired by Minsky's "society of mind" concept.

This approach aims to extend the mind with human values and processes, creating wisdom from information. It's presented as an "extension to human intelligence" rather than an autonomous entity with its own mind. The focus is on infusing human-like reasoning and values into digital automata, rather than creating a fully autonomous machine.

Extending AI automata with the attributes of human mind has several advantages:

1. **Ethical AI:** By infusing AI with human values and reasoning, the technology remains aligned with human ethics and morals, which is crucial for societal acceptance.

2. **Enhanced Decision-Making:** The combination of human intuition and AI's data-processing capabilities can lead to more informed, comprehensive decision-making processes.

3. **Collaborative Intelligence:** It promotes a synergy between human and machine intelligence, augmenting human capabilities rather than replacing them.

Max Michaels and Rao Mikkilineni

4. **Control and Safety:** Keeping AI as an extension of human intelligence, rather than an autonomous entity, potentially reduces risks associated with fully autonomous AI systems.

The ethical implications and the need for regulation also highlight the transformative impact of AI, suggesting that its evolution could take unforeseen directions. The challenge is self-regulation of AI systems, the need to enable an imperfect machine monitor itself and adapt to previously unseen conditions. AI's ability to monitor itself is largely dependent on the initial design and ongoing monitoring by human overseers. While AI can adapt and learn, its self-regulation is inherently linked to its initial programming and the continuous oversight by humans who understand and can interpret its processes and outputs.

In Summary:

While the direct application of Minsky's "Society of Mind" theory may not be explicitly prominent in current AGI research, the underlying principles of collaborative, multi-faceted intelligence are reflected in various modern approaches to developing AGI. The focus on integrating human-like cognition, empathy, and shared experiences into AI systems resonates with the idea of a collective intelligence emerging from simpler processes. We believe that, through thoughtful adaptation, this concept can emerge as a pivotal element in our endeavor to advance and align AGI.

The "Society of Minds" approach to integrating AI with human intelligence represents a potential leap in designing Ethical AI Systems. By weaving human values and reasoning into the fabric of digital automata, it heralds a new era where technology is not just an independent force but an "extension of our collective wisdom." This paradigm shift, from autonomous AI to a mindful AGI that complements and enhances human intellect promises a future where ethical considerations, collaborative intelligence, and enhanced decision-making are at the forefront. It's a vision that redefines our relationship with technology, ensuring that as we stride into the future, our technological advancements walk hand-in-hand with the rich tapestry of human values and ethics. This isn't just about creating smarter machines that can think deeper; it's also about fostering a world where AI augments and amplifies human minds.

Max Michaels and Rao Mikkilineni

Applying Society of Minds: By embracing the Brain-Body-Mind metaphor, we can create mindful AI systems that are not only tools but partners, capable of growth, reflection, and value-based action. The implications for industry, society, and our collective future are vast. Join us, as we embark on this journey and shape a world where AI's potential is fully realized.

Max Michaels and Rao Mikkilineni