AI Embodiment Through 6G: Shaping the Future of AGI

Lina Bariah ¹ and Merouane Debbah ²

¹Technology Innovation Institute
²Affiliation not available

December 7, 2023

Abstract

In the ever-evolving field of technologies, the emergence of Artificial General Intelligence (AGI), often referred as strong artificial intelligence (AI), stands as a breakthrough in the realm of machines intelligence, promising to witness a new era of capabilities and possibilities. In particular, AGI ventures into human-level cognition, and expands to thinking, reasoning, and awareness. This imminent evolution is envisioned to be manifested through the embodiment of AI machines, allowing machines to transcend their purely computational nature and interact with the world through the different senses. Accordingly, AI agents will be grounded in the physical environment, going through subjective experiences and acquiring the needed knowledge that will lead to understanding and cognition. In our article, we explore the path towards realizing the true vision of AGI through AI embodiment, where we dig into the different types of thinking required to achieve knowledge, and hence, cognition and understanding. Furthermore, we look through the evolution of generative AI models, and shed lights on the limitations of auto-regression in large language models (LLMs), with the aim to answer the question: is sensory grounding (through 6G) necessary, and enough, to achieve understanding in LLMs? Finally, we identify the main pillars of AGI and unveil how 6G networks will orchestrate the development of AGI systems.
Abstract—In the ever-evolving field of technologies, the emergence of Artificial General Intelligence (AGI), often referred as strong artificial intelligence (AI), stands as a breakthrough in the realm of machines intelligence, promising to witness a new era of capabilities and possibilities. In particular, AGI ventures into human-level cognition, and expands to thinking, reasoning, and awareness. This imminent evolution is envisioned to be manifested through the embodiment of AI machines, allowing machines to transcend their purely computational nature and interact with the world through the different senses. Accordingly, AI agents will be grounded in the physical environment, going through subjective experiences and acquiring the needed knowledge that will lead to understanding and cognition. In our article, we explore the path towards realizing the true vision of AGI through AI embodiment, where we dig into the different types of thinking required to achieve knowledge, and hence, cognition and understanding. Furthermore, we look through the evolution of generative AI models, and shed lights on the limitations and understanding. Finally, we identify the main pillars of AGI and unveil how 6G systems with external environments, i.e., relate the perception and awareness. This imminent evolution is envisioned to be human-level cognition, and expands to thinking, reasoning, and awareness. This imminent evolution is envisioned to be manifested through the embodiment of AI machines, allowing machines to transcend their purely computational nature and interact with the world through the different senses. Accordingly, AI agents will be grounded in the physical environment, going through subjective experiences and acquiring the needed knowledge that will lead to understanding and cognition. In our article, we explore the path towards realizing the true vision of AGI through AI embodiment, where we dig into the different types of thinking required to achieve knowledge, and hence, cognition and understanding. Furthermore, we look through the evolution of generative AI models, and shed lights on the limitations and understanding. Finally, we identify the main pillars of AGI and unveil how 6G networks will orchestrate the development of AGI systems.

I. INTRODUCTION

The culmination of years of artificial intelligence (AI) research has led to the emergence of artificial general intelligence (AGI) as the definitive frontier that aspires to mimic human intelligence from all aspects. The notion of AGI revolves around building machines that possess intelligence and cognitive capabilities that are comparable to humans, and are designed to master a wide spectrum of general cognitive functions, without being particularly trained on them. The rise of the AGI paradigm was recently fueled by the remarkable advancements on generative AI models, represented by Large Language Models (LLMs), which allowed AI machines to enjoy some sort of understanding to natural language. Yet, these models can be categorized within the concept of narrow/specialized AI, and are far from fully representing human general intelligence.

The overarching goal of AGI encapsulates knowledge generalization, self-directed decision-making and problem-solving, adaptation and learning, critical thinking and common-sense, as well as human-machine and machine-machine collaboration. In order to achieve such a vision, AI should migrate from the abstract algorithm base to a more advanced form of intelligence that immerge the AI brain in the physical environment, to allow it to experience all interactions as humans (i.e., to be grounded in the physical domain), and therefore, to acquire the needed knowledge that will contribute to its cognition and awareness systems. This was recently defined as AI embodiment, which implies that AI agents will no longer be detained by a computer algorithm, but will rather possess a physical instantiation that will allow it to interact with the surrounding environment. Embodiment AI has profound implications on AGI systems, in the sense that it will not only allow the AGI agents to understand the physical world, but to engage with it and learn from such an engagement.

The aim of this article is to delve deeper into the concept of AI embodiment as a gateway to the ultimate vision of AGI, where we open avenues for exploring how sensory grounding can forge a robust link between the machines and the environment, and therefore, paves the way to machine understanding. We set the scene for the true vision of AGI, where we unfold the reason why 6G is essential to realize this vision, and how 6G will be the key to the convergence of computing, collective intelligence, reinforcement learning (RL), sensing, and virtualization, in pursuit of AGI. To the best of the authors’ knowledge, this is the first article to approach the AGI paradigm from embodiment, LLMs, and 6G perspectives.

A. Slow & Fast Thinking: Two Ways to Knowledge

Humans brains comprise two cognition systems, system I, referred as fast thinking, which is intuitive, automatic, and tied to heuristics and past experiences. Meanwhile, system II (slow thinking) is slow, analytical, and tied to reasoning and conscious thinking. While system II involves a considerable level of information processing, the majority of processing happening in the brain is due to system I thinking. From a computing perspective, system I is demonstrated through machines learning, e.g., neural networks, and system II is deployed through programming the machines according to particular algorithms and logic. It is worthy to mention that slow thinking is executed in parallel and performs computations that do not follow a particular logic.

The uniqueness of human intelligence relies on the capability of the human brain to combine sensory information perceived from the environment, with the logical processing and decision making, and subsequently perform relevant actions, in an unconscious manner. This means that fast thinking in natural intelligence can adapt quickly to the rules that continuously emerge as the environment varies, unlike the constant rules set by a particular logical process. More precisely, cognition and understanding in this context can be referred as a semantic model that connects internal systems with external environments, i.e., relate the perception...
of the external environment with the internal knowledge. In computing systems, machines can be endowed with robust semantic model to enable them to approach the human intelligence level. While analyzing and understanding natural language, manifested through LLMs, can partially contribute to the development of such a semantic-based cognitive system, knowledge acquisition and update will require the design of sophisticated interrelated semantic models that are capable of accumulating old knowledge and establishing relationships between concepts, rules, and meanings. In this way, knowledge that will lead to cognition and consciousness can be achieved.

B. How 6G Technology Will Contribute?

The advent of 6G technology promises to open up new possibilities for AGI, through enabling seamless communication and data exchange between a multitude of devices, including those powered by AGI. From the one hand, with the ultra-fast data transfer rates, extremely low latency, and unprecedented connectivity, 6G networks will provide the ideal infrastructure for AGI systems to operate reliably (more details provided later in Fig. 4). On the other hand, 6G technology is expected to support advanced edge computing capabilities, which allow the processing and analysis of data closer to the data source. This integration of AI and edge computing can enable AGI models to process and interpret information in real-time, leading to quicker and more context-aware understanding. Furthermore, 6G technology will support advanced forms of communication, including haptic and multi-sensory feedback. These modalities can enable AGI to interact with the world in a more human-like way and enhance its understanding of the environment through sensory input.

II. WILL SENSORY GROUNDING LEAD TO UNDERSTANDING IN LARGE LANGUAGE MODELS?

According to David J. Chalmers [1], a being to be conscious and to understand, it should go through subjective experiences. He further elaborated that subjective experiences can vary between sensory (related to perception - seeing colors), affective (related to feelings/emotions - pain/happiness/etc.), cognitive (related to thinking and reasoning), and agentive (related to actions) experiences. Such subjective experiences are all linked to the being’s perspectives and outlooks. Hence, different than objective experiences, which are based on factual information, subjective experiences can be acquired through interacting with the world in order to build own point-of-views on different scenarios and actions happening in the environment. This theory readily applies to the principle of cognition and understanding in LLMs. Achieving consciousness and understanding in generative models is a significant leap forward to realize AGI.

In light of this, it can be argued that sensory grounding in the real world is necessary for developing robust understanding in LLMs. Specifically, S. Harnad in [2] claimed that AI systems need to be grounded in the environment in order to be able to possess understanding and cognition, and to be conscious. Having said that, to enable human-like cognition, symbols in machines need to be connected with their real-world referents, and this can be achieved by allowing the AI agents to be grounded in sensory experiences and to interact with the surroundings. Establishing a strong link between the symbols and their perceptual and action-based representations enable the agents to understand their meanings. Motivated by this, several initiatives have been conducted to explore the potentials of grounding in multi-modal LLMs. KOSMOS-2 [3] is a multi-modal LLM that aims at grounding text to the visual world, and that is capable of performing visual reactions. On the other hand, the authors in [4] have proposed a framework for video-grounding dialogues through exploiting GPT-2. The proposed work targets to demonstrate the grounding capabilities in LLMs to capture the spatio-temporal dependencies of video frames, and to associate them with the relevant tokens from the dialogue.

Sensory grounding can potentially contribute to the realization of understanding in LLMs, however, alone it does not necessarily lead to full understanding in these models. The true notion of human-like understanding involves a deeper level of knowledge representation and reasoning that goes beyond simple connections between modalities, including semantic understanding, causal reasoning, common sense knowledge, contextual awareness, and adaptability.

III. WHAT IS ARTIFICIAL INTELLIGENCE EMBODIMENT?

In cognitive science and philosophy of mind, embodied cognition is deeply rooted to the body’s interaction with the environment. In particular, it is elaborated that cognitive activities and processes are influenced by the context of real-world scenarios, in which perception and actions can take place. According to this, cognition is not just a by-product of the brain’s computations and information processing, but also forged by the body’s sensory and motor functions experienced in the physical world. Similarly, perception is not only influenced by the sensory information acquired from the surrounding environment, but it is also highly linked to the internal sensory system of the body, as well as past experiences. The latter is highly essential to enable cognition, memory, and contextual understanding, in addition to evolving social perception.
AI embodiment refers to the concept of granting AI systems a tangible presence in the physical world, in which they can leverage it for perception, sensing, and physical interaction with the surroundings, with the aim to realize improved cognition and understanding, and therefore, enhanced decision-making capabilities. Based on this, AI will no longer be limited to software running on servers, rather, it will have a physical representation manifesting complex intelligent systems with authoritative, manipulative, and control capabilities. Current sensory-based approaches, e.g., visual AI algorithms, have very limited abilities in relevance to cognition, due to the fact that they are highly dependent on the prior knowledge of the objects, environment situation, geometries, etc. Accordingly, variations in such elements limit the use of these algorithms and potentially degrades their performance.

The current vision of artificial embodiment is based on the integration of artificial visual, textual, audio, and reasoning skills into a physical representation to solve AI-related problems in real environments. This indicates that AI agents are no longer solely relying on datasets, but also performing human-like interactions and learning from the environment. It should be emphasized that, while AI embodiment and RL systems share some aspects with relevance to interacting with an environment, the latter do not require a physical embodiment to interact.

To enable human-like experiences, the notion of AI embodiment can be identified through several elements, incorporating hardware and software technologies. These elements can be characterized as follows:

**Physical manifestation:** AI embodiment requires the evolution from purely software-based systems into systems with physical instantiations, i.e., AI should be granted a physical body or platform that enables it to act, react, and manipulate objects in the physical world, and thereby, gain contextual cognition of the environment and the interactions happening in it. Although such physical representation can be demonstrated through a vehicle or a drone, it is strongly believed that, in order to imitate a human-like experience for improved cognition and understanding in intelligent systems, the agent’s body morphology can radically impact how accurately and reliably the agent can behave and react to the environment. For example, it was shown that by carefully shaping the body of the agent, stable and efficient locomotion can be achieved [5].

**Environment coupling:** The notion of AI embodiment is associated with the existence of channels in which mutual perturbations can occur between the agent and the environment. Such a concept, referred as physical grounding, means that a system can be embodied in an environment if in each time instant, environmental states are capable of impacting the states of the agent and vice versa. This emphasizes on the fact that being physically instantiated is not enough to be embodied, but rather, places more weight on the role of sensorimotor functionalities in realizing embodiment. Within this context, embodied AI systems can be achieved through equipping intelligent agents with various sensors, enabling them to perceive their environments, including, cameras, microphones, touch sensors, navigation and positioning systems, to name a few. On the other hand, actuators, e.g., motor actuators and speakers, are essential to endow the agents with...
reflexes in response to the environment perception.

**Past experiences:** The theory behind the historical engagement between the environment and the agent signifies that in order to achieve embodied cognition in AI systems, not only the present coupling between the agent and its environment contributes to the cognition of the agent, but also past experiences that incorporated agent-environment interactions and adaption reflects to the cognitive system embodiment.

**Human-AI Interaction:** Enabling AI agents to establish effective interactions with humans is an essential step towards shaping the emergence of social behaviors at the agents, including cooperation, competition, and empathy behaviors. This further involves granting the agents the capability to understand humans’ commands, queries, metaphors and their meanings, and the implications of these on the responses of humans (and machines) in particular scenarios. Such development will enable improved social decision-making that contributes to cognition and understanding, and hence, AI embodiment evolution. This is stemmed from the fact that expressive behaviors through human cognition is highly related to the humans’ ability to elicit social and cultural skills from the surrounding environment. Accordingly, the true vision of AI embodiment will necessitates the agents to develop social behaviors through interacting with humans in the physical world. Within the same context, behavioral science studies demonstrated that social learning can be a critical key in cognition; when there is a lack in possessed information, humans can resort to high-fidelity copying as a way to acquire social information to compliment possessed one. The same concept can apply to AI agents, which can develop cognition through mimicking human social behaviors in different environments [6].

**IV. LLM Revolution: Why Auto-Regression in Generative Models is Not the Future of AGI?**

LLMs have demonstrated exceptional capabilities in processing and generating human-like text, and accordingly they have gained a considerable attention as a tool for performing a wide range of language-related tasks. The basic concept of LLMs was built upon a foundation architecture, referred as transformer, which was developed to overcome the limitations of recurrent neural networks (RNNs), including the short-term memory and lack of parallelism capabilities [7].

Among others, auto-regression in generative models constitutes the base for pretrained generative transformers. In text generation, auto-regression allows the model to predict the subsequent word in a sequence according to the context of the proceeding words. This is achieved by relying on the probability distribution of the next word/token given the context of the previous ones. Using various functions, such as softmax, probability distributions of all possible words can be modeled, and accordingly the next word/token can be selected given it enjoys the highest probability in a particular context. By doing so, the overall outcome of the model is expected to be coherent, correlated, and contextually valid.

Although auto-regression has proven to successfully handle several language-related tasks, given its characteristics we envisaged that auto-regression does not provide the ultimate solution for the AGI embodiment. From the one hand, taking into consideration the full reliance on probability distributions among a particular context, auto-regressive based models lack the control over their outputs. With the anticipated vision of embodied AI, the agents are expected to enjoy some sort of control over their models’ outputs, in order to enable them to engage actively in the learning process within the environment, and to learn from their own experiences. Furthermore, controlling the models will assist the agents to adapt quickly to the changes happening in the surrounding environment, where the agent can adjust their models in an on-demand manner. In addition, self-governance in embodied agents can be achieved through enabling the agents to control the outputs of their models, and therefore, undesirable behaviors can be avoided. From a different perspective, the performance of auto-regression is constrained by the training dataset, and therefore, this limits its ability to reliably handle unseen data. Embodied agents in the era of AGI are anticipated to possess the capacity to generalize and adapt to new scenarios without catastrophic failures. Another limitation of auto-regression in AGI is due to the fact that such models lack the context and causal reasoning. This means that auto-regressive models are not designed to understand the cause-and-effect relationships between the data, but rather their capabilities are limited to capturing the immediate context and relationships in a sequence. For better understanding of the environment and the behaviors of objects in it, embodied agents should be capable of capturing the causality of events and behaviors in the physical world, and therefore, to have the ability to react and take decisions according to the outcomes as well as according to the reasons that provoked these outcomes.

**V. The Ultimate Vision of AGI: 6G as the Fusion of Sensing, Generative Agents, and Knowledge**

While remaining a concept, the pursuit of AGI has been the pinnacle of the research on AI, resembling human-like intelligence in machines, i.e., going beyond learning patterns and relationships in data by enabling machines to understand, reason, and learn, even in novel scenarios that are not presented by existing data. This means that machines will go from interpolating, to possessing the ability to extrapolate beyond the probability distribution of the training data [8]. While the sizes of current language models have scaled up to the billions-level, the capabilities of these models are limited to the training dataset, and it is far from possessing understanding skills. One approach to enable understanding, and hence, embodiment in AGI systems is the hypothetical concept that as the model size goes higher, the resulting model will be able to extrapolate in unseen situations, and therefore, will be able to think. Although such a hypothesis emphasize on one path towards improving the cognition capabilities of generative models, the speculative vision of AGI systems goes beyond Billions-Scale LLMs. As illustrated through this article, the true vision of AGI will become a reality when AI is embodied through a unique blend of several technologies and concepts, including generative language models, internet-of-senses, virtualization, RL, and edge intelligence, among
others. 6G, as a transformative network generation, will be the orchestrator that will ensure a harmonized integration between these technologies. In this section, we explore the profound role of 6G as a catalyst for the fusion of sensing, generative AI, and RL, towards AGI.

A. Advancements on Computing Fabric: Overcoming the Von Neumann Bottleneck

The underlying hardware and architecture that supports the computation, communication, and data movement within AGI systems represent a core element in realizing AI embodiment. Over the last decades, Von Neumann’s architecture has been the base for almost all computing systems, and has evolved for optimized data management and to support hardware miniaturization. Von Neumann computers have shown spectacular performance in computing and solving complex problem. However, when it comes to AGI, Von Neumann’s architecture reaches its limits in providing the needed computing capabilities, in terms of large-scale parallelism and asynchronous execution, memory bandwidth, latency, power consumption, and architecture flexibility, to name a few.

Enabling embodied AI will bring forth the need for novel technologies that will better shape the computing fabric in the era of AGI systems. In 3, we touch base with the computer architectures that reveal promising capabilities in meeting the demands of AI embodiment. Note that heterogeneous architectures that integrate multiple computing approaches, alongside Von Neumann’s, can lead to optimized AGI systems, striking a balance between computing and energy efficiency, and enabling the required flexibility and adaptability for a large number of tasks with diverse computing requirements.

Network Computing Fabric through 6G: In order to meet the demands of emerging use-cases that are particularly based on virtual and immersive applications (with their stringent requirements), future wireless networks should tightly embed the communication, computation, and memory capabilities in a unified entity, at the edge of the network, offering reduced latency and improved bandwidth and reliability. Accordingly, in the 6G era, we can expect a dense network of edge nodes and small data centers to form a distributed computing fabric. Subsequently, the improved sensing and computing capabilities at edge networks with the combination of low latency and high bandwidth, promised to be offered by 6G, LLMs implemented at edge devices will be able to engage in more natural and real-time interactions with scenarios experienced at the edge. This is further empowered by allowing LLMs to access and process data more rapidly, leading to faster training and more responsive real-time language understanding. In this setting, 6G will be play the role of an execution environment, for a highly distributed data processing, computing, LLM training and inference, and storage, through providing the needed resources and connectivity.

B. Perceive the World through Seven Senses

At the core of the AI embodiment, sensing represents the cornerstone to this vision and the agents’ gateway to perceive and acquire multi-modal data from the environment. In addition to the human’s five senses, AGI agents will require two more senses, namely, inertial and temporal senses. Inertial sensing is essential for the agents in order to enable them to realize their own and other objects motions, orientations, and acceleration in the space. These are key elements in the navigation and tracking systems of the agents, which readily contribute to the agents abilities to interact with the environment. On the other hand, temporal sensing is crucial for real-time decision-making, time series analysis, and future predictions, in which the agent can rely on temporal sensing to recognize time-dependent relationships between different data modalities acquired from the environment.

Allowing the agents to fully utilize these seven senses will necessitate ultra-reliable, low latency, and high data rate connectivity. 6G networks in this regard will be the solution
to provide not only high data rate connectivity with ultra-low latency for each sense individually, but also will ensure that data pertinent to different senses to be communicated simultaneously between the environment and the agents. This is stemmed from the fact that AGI agents will require to receive all relevant information related to a particular event or object in a reliable way with imperceptible time-lag, in order to be able to link information related to multimodal data and build a comprehensive image that will enable them to understand the environment. According to IEEE 1918.1 tactile internet standards, visual streaming data rate should be at least 1 Tbps, while end-to-end latency of high resolution haptic communication should not exceed the 1 ms. Thanks to 6G networks that promised to deliver 1 Tbps data rate, $10^{-9}$ reliability, 0.1 ms latency, with a diverse range of frequency bands (including sub-6GHz, mmWave, and Terahertz) [9]. Therefore, 6G networks are anticipated to serve for the stringent requirements of on-demand sensing with the needed data granularity, and hence, play an essential role in the evolution of AI embodiment. In Fig. 4, we highlight the sensing capabilities of humans compared to machines, and what 6G needs to offer to enable immersive sensing experience.

\section*{C. The Environment can be Virtual}

The recent progress in physical simulation platforms and rendering technologies has unlocked a different path to enable embodied agents training through creating holistic, close-to-real environments that allows the agents to explore, monitor, perceive, and interact with a virtual realm. Computer vision researchers have recently focused on developing such 3D environments, in which virtual agents can operate and learn.

Among several initiatives, iGibson [10] and Habitat-Sim [11] were developed as promising platforms, based on real worlds, for embodied agents training on navigation tasks. Most of the currently available virtual environments are aimed to train agents to navigate. While this demonstrates a good achievement, it is still ranked as one of the simplest tasks within the AI embodiment, as it does not involve a real interaction with the environment. It is important to emphasize that the level of details and complexity of the virtual environment have a high influence on the training quality. While virtual worlds can be a good step toward realizing AI embodiment, the available simulation platforms are still much less real than real environments, and they do not provide perception capabilities to all needed senses. Therefore, the development of an ideal virtual environment that will be the base for the AI embodiment paradigm is sustained by the quality of immersion that future simulators can attain.

The recent breakthroughs are anticipated to support digitization and virtualization through 6G networks. Developing an immersive environment necessitates advanced sensing technologies to develop a high-fidelity replica of an existing environment, and efficient communication schemes to ensure that the virtual and real realms are synchronized. Such synchronization will be the gateway to enable an efficient interaction between the AI agents and humans, yielding high-quality AI embodiment. Amongst numerous schemes, semantic communication, joint source & channel coding, emergent communication, spatial multiplexing-based waveform design, and non-coherent communication have manifested themselves as potential candidates to enable high-resolution virtualization of the real world [12]. These paradigms promise to achieve high data rate communication, with high reliability, ultra-low latency, and high energy efficiency, while maintaining the desired level of physical-virtual synchronization.

\section*{D. Embracing Multi-Agent Reinforcement Learning}

The authors in [13] argued that AGI can be reached through training a RL agent in a sufficiently complex environment. In particular, by maximizing the reward function, several forms of intelligence can be achieved. This is tied to the hypothetical theory that any behaviour or interaction with the environment can be modeled through a reward function maximization process, in which the reward function is designed to induce that particular behaviour. While it is debatable that the reward is enough, we strongly agree that RL is essential for the AGI agents to be grounded in the environment, and to acquire knowledge that will play a part in the decision-making process. Recently, a novel promising multimodal language model has been announced, which in principle relies on RL with human feedback (RLHF) to gain outstanding capabilities in question-answering tasks [14]. As its name indicates, RLHF is an approach where human expertise and intuition are exploited in the training of RL agents, with the aim to speed up the learning and adaptation process. Within the context of LLM, the agents need to learn that some tokens are more important than others, and are more suitable to a particular context or scenario than other tokens. Here, the notion of importance can be learned through reward-based RL system, in which the AGI agent will learn to take actions in relevance to particular important tokens, while ignoring less important ones, with the aim to maximize a specific reward that serves the intended goal. Through utilizing data collected from human interactions, RL can be used to align the generated text from the LLM with that of human preferences.

6G networks are more likely to emphasize the deployment of edge computing, which lay the foundation for implementing RLHF at the edge. In particular, human feedback acquisition and evaluation, with the associated model adaptation can be more efficiently performed in a distributed fashion, i.e., closer to the point of interaction, and therefore, the incurred delays can be minimized and the overall performance of the agent in terms of actions and decision-making can be enhanced. This in addition to the high data-rate envisioned to be realized by 6G networks, which will facilitate the smooth interaction between LLM-empowered agents and the humans, enabling efficient real-time feedback and improved training process.

\section*{E. Collective Intelligence is a Key}

The process of developing AGI will require continuous feedback, evaluation, and refinement, which in order to be automated should rely on machines rather than humans. Building a large community of AI agents that are collaborating to identify point-of-weaks and potential risks in other agents’ behaviours and actions will contribute to the development of
robust AGI systems. Collective intelligence represents one of the most effective ways for sharing knowledge, and hence, learning, among AGI agents, following similar approach to teaching and social discussions in humans. Lifelong Learning (LL) mechanism [15] shows the efficiency of knowledge transfer among different agents, in which each agent share the knowledge of its own learned tasks with other agents. In particular, in [15], 102 tasks were distributed among the agents and then each agents shared its knowledge with the others in a decentralized manner, ideally, all agents at the end mastered the 102 tasks. Such a study shows a strong demonstration toward the potentials of collective intelligence in improving knowledge, and therefore, enhancing the learning and cognition in embodied agents. Through leveraging collective intelligence, knowledge can scale up with reduced training and learning overhead, particularly when the number of learned tasks expands to thousands or millions tasks in accordance to the vision of AGI.

The main challenge in realizing collective intelligence in AGI relies behind the need for extremely reliable wireless networking and efficient communication protocols, in order to ensure successful knowledge transfer and to prevent agents mutation and therefore, avoid undesired agents behaviours. Future wireless generations are anticipated to introduce a new class of AI-enabled communication protocols, which are relatively energy and spectrally efficient. In particular, different than the classical communication protocols, which are rather generic, uninterpretable, and with unadaptable control signaling messages, AI-empowered communication protocols are envisioned to enjoy task-specific control signaling messages, with an acceptable communication, computing, and memory usage overhead. Additionally, with the advent of semantic communication, semantic-based communication protocols represent a promising candidate for knowledge transfer among generative agents, through allowing the agents to share knowledge in a contextually rich manner.

VI. CONCLUSION

This article answers the question whether grounding AI in sensory experiences suffice to achieve genuine understanding within LLMs. To answer this question, we first studied the different kind of thinking that shape the vision of cognition in generative agents. We further explored the notion of AI embodiment as a gateway to the AGI, where we outlined the limitations of auto-regression in LLMs in this process, and pointed out the pillars of AGI, the integral role of 6G networks in enabling sensing, virtualization, multi-agent RL, and collective intelligence, as a route to embrace AGI.

REFERENCES


Lina Bariah (lina.bariah@ieee.org) received the M.Sc. and Ph.D. degrees in communications engineering from Khalifa University, Abu Dhabi, UAE, in 2015 and 2018, respectively. She was a Visiting Researcher with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada, in 2019, and an affiliate research fellow, James Watt School of Engineering, University of Glasgow, UK. She is currently a Senior Researcher at the technology Innovation institute, an Adjunct Faculty at Khalifa University, and an Adjunct Research Professor, Western University, Canada. Dr. Bariah is a senior member of the IEEE.

Merouane Debbah (merouane.debbah@ku.ac.ae) is a researcher, educator and technology entrepreneur. Over his career, he has founded several public and industrial research centers, start-ups and is now Professor at Khalifa University of Science and Technology in Abu Dhabi and founding Director of the 6G center. He is a frequent keynote speaker at international events in the field of telecommunication and AI. His research has been lying at the interface of fundamental mathematics, algorithms, statistics, information and communication sciences with a special focus on random matrix theory and learning algorithms. In the Communication field, he has been at the heart of the development of small cells (4G), Massive MIMO (5G) and Large Intelligent Surfaces (6G) technologies. In the AI field, he is known for his work on Large Language Models, distributed AI systems for networks and semantic communications. He received multiple prestigious distinctions, prizes and best paper awards (more than 35 best paper awards) for his contributions to both fields and according to research.com is ranked as the best scientist in France in the field of Electronics and Electrical Engineering. He is an IEEE Fellow, a WWRF Fellow, a Eurasip Fellow, an AAIA Fellow, an Institut Louis Bachelier Fellow and a Membre émérite SEE.