The Emergence Phenomenon in Artificial Intelligence: A Warning Sign on the Path to Artificial General Intelligence

Vera Sorin MD[1,2,4] and Eyal Klang MD[1,2,3,4,5]

[1]Department of Diagnostic Imaging and [2]Deep Vision Laboratory, Sheba Medical Center, Tel Hashomer, Israel
[3]ARC Innovation Center, Sheba Medical Center, Tel Hashomer, Israel
[4]Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[5]Icahn School of Medicine at Mount Sinai, New York, NY, USA

ABSTRACT: Large language models have revolutionized natural language processing. The emergence phenomenon is observed in these models and has the potential to revolutionize data processing and management. In this review, we discuss the concept of emergence in artificial intelligence, give detailed examples, and elaborate on the risks and limitations of large language models. The review exposes physicians to large language models, their advantages, and the inherent opportunities. We also describe the limitations and dangers, as these models are expected to impact medicine soon.

IMAJ 2024; 26: 120–121

KEY WORDS: autonomous learning, ethical implications, emergence phenomenon, generative adversarial networks, large language models

Large language models (LLMs), like generative pre-trained transformer (GPT), have revolutionized natural language processing by demonstrating capabilities previously unique to human cognition. These advanced algorithms automatically learn about objects and concepts, a phenomenon that is both awe-inspiring and concerning. LLMs are increasingly being researched in healthcare [1], and in addition to text analysis, can also be applied to image analysis [2].

Emergence is a term used across various disciplines, including physics, biology, and social sciences [3]. It describes unpredictable properties that arise in large systems through the interaction of simpler components [3]. Ant colonies exemplify this phenomenon. Individually, ants do not display any extraordinary qualities. Collectively, however, they form complex systems capable of adapting to environmental changes, locating food, and thriving. In the context of artificial intelligence (AI), emergence refers to unexpected tasks performed by an algorithm without prior explicit programming.

LLMs rely on extensive algorithms with billions of parameters. Trained on vast amounts of text data, their performance enhances as their size increases. Performance is generally predictable based on the behavior of smaller models; however, occasionally, some tasks become unpredictable, exhibiting emergence [4]. This phenomenon fosters both scientific excitement and concerns regarding unforeseen consequences.

Generative models in AI, such as generative adversarial networks (GANs), exemplify the phenomenon of emergence. These networks enable the creation of realistic-looking synthetic images: *deepfakes*. The term *deepfake* attracted public attention in 2017 with the creation of fake videos intended to influence public opinion. GANs consist of two neural networks: a generator and a discriminator. The generator produces synthetic images, while the discriminator tries to differentiate between real and fake images [5]. Through this competitive interaction, both networks improve their performance, leading to emergent behavior that results in the creation of increasingly realistic synthetic images.

Large language models learn to identify objects through textual analysis. By examining vast amounts of text, the algorithm identifies objects and their properties in a seemingly natural way. For example, after *learning* about various fruits and their characteristics, the algorithm can infer that a round, orange fruit is likely an orange [6].

The algorithm can create object concepts by identifying common characteristics and forming connec-

LARGE LANGUAGE MODELS HAVE REVOLUTIONIZED NATURAL LANGUAGE PROCESSING, OFFERING UNPRECEDENTED OPPORTUNITIES FOR ADVANCEMENTS IN DATA HANDLING AND ANALYSIS.

tions between similarly labeled objects [6]. When asked about bicycles, GPT can expand on the different types, colors, parts, functions, and history. The algorithm can even create a story about bicycles. Interactions between simple components in the algorithm allow it to learn new information and make conclusions, in contrast to executing pre-programmed instructions. The algorithm did not undergo explicit programming and learning about bicycles but learned autonomously about their characteristics through text processing. Similarly, large language models learn about concepts across various domains. Without explicit training, GPT-4 scored 87% on the United States Medical Licensing Exam (USMLE) [7,8].

CENTRAL TO THE TRANSFORMATION OF NATURAL LANGUAGE PROCESSING IS THE EMERGENCE PHENOMENON, WHICH UNVEILS NEW PATTERNS AND INSIGHTS FROM DATA, PREVIOUSLY UNATTAINABLE.

Emergence serves as the driving force behind the development of cognitive abilities in large-scale AI models. As the model constructs a complex hierarchy of object relationships, it begins to identify itself as a distinct entity, fostering a sense of self-identity. This emergence process not only promotes the development of advanced cognitive functions, but also lays the foundation for a potential future where AI could hypothetically attain sentience.

Language has long been considered the key to human domination over ecological systems. Our ability to invent ideas, communicate them, and act collectively based on shared beliefs has allowed us to create complex societies and technologies [9]. LLMs, with their ability to generate and understand language, could be seen as a reflection of this human capability. Their ability to learn autonomously and generate ideas could lead to implications and advancements that are difficult to predict and anticipate.

The emergence phenomenon in AI raises significant ethical concerns and risks. Unpredictable outputs may be difficult to control or regulate. When decisions in various fields are based on the output of AI models, accountability becomes a challenge. Information security and privacy are additional challenges since these algorithms are vulnerable to cyberattacks and can be used to launch cyberattacks [10,11]. When used to support decisions in fields such as medicine, economics, and law, the phenomenon of emergence could lead to unforeseen consequences, biases, or errors in critical decisions that may have serious implications for individuals and society.

As LLMs advance and improve, the occurrence of emergent behavior is expected to increase, ultimately bringing us to critical crossroads. These models hold the promise of revolution in various fields, but they also highlight the need for cautious and responsible development. The emergence phenomenon is a reminder that as we approach artificial general intelligence, we must tread carefully, recognizing the power and potential as well as the inherent dangers of this extraordinary technology.

Correspondence

Dr. V. Sorin
Dept. of Diagnostic Imaging, Sheba Medical Center, Tel Hashomer 52621, Israel
Phone: (972-3) 530-2530
Fax: (972-3) 535-7315
Email: verasrn@gmail.com

References

1. Temsah MH, Altamimi I, Jamal A, Alhasan K, Al-Eyadhy A. ChatGPT Surpasses 1000 Publications on PubMed: Envisioning the Road Ahead. *Cureus* 2023; 15 (9): e44769.

2. Sorin V, Glicksberg BS, Barash Y, Konen E, Nadkarni G, Klang E. Diagnostic Accuracy of GPT Multimodal Analysis on USMLE Questions Including Text and Visuals. *medRxiv* 2023: 2023.10.29.23297733.

3. Kivelson S, Kivelson SA. Defining emergence in physics. *npj Quantum Materials* 2016; 1 (1): 1-2.

4. Wei J, Tay Y, Bommasani R, et al. Emergent abilities of large language models. arXiv preprint *arXiv*:2206.07682 2022.

5. Sorin V, Barash Y, Konen E, Klang E. Creating artificial images for radiology applications using generative adversarial networks (GANs) - a systematic review. *Acad Radiol* 2020; 27 (8): 1175-85.

6. Bubeck S, Chandrasekaran V, Eldan R, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint *arXiv*: 2303.12712 2023.

7. Chaudhry HJ, Katsufrakis PJ, Tallia AF. The USMLE Step 1 Decision. Jama 2020;323:2017.

8. Brin D, Sorin V, Konen E, Nadkarni G, Glicksberg BS, Klang E. How large language models perform on the united states medical licensing examination: a systematic review. *medRxiv* 2023: 2023.09.03.23294842.

9. Harari Y, Harris T, Raskin A. You Can Have the Blue Pill or the Red Pill, and We're Out of Blue Pills. *New York Times* 2023 Mar 27.

10. Sorin V, Soffer S, Glicksberg BS, Barash Y, Konen E, Klang E. Adversarial attacks in radiology - A systematic review. *Eur J Radiol* 2023; 167: 111085.

11. Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. arXiv preprint *arXiv*: 1412.6572 2014.

Everyone, in some small sacred sanctuary of the self, is nuts.

Leo Calvin Rosten (1908–1997), American humorist in the fields of scriptwriting, storywriting, journalism, and Yiddish lexicography