Chapter 5
Stepping Stones for Self-Learning:
Exploring the Use of Multimodal Text- and Image-Making Generative AI Tools

Shalin Hai-Jew
https://orcid.org/0000-0002-8863-0175
Kansas State University, USA

ABSTRACT

One of the themes in the emergence of text- and image-making (multimodal) generative AIs is their value in the learning space, with the vast potential just beginning to be explored by mass humanity. This chapter explores the potential and early use of large language models (LLMs) harnessed for their mass learning, human-friendly conversations, and their efficacies, for self-learning for individuals and groups, based on a review of the literature, system constraints and affordances, and abductive logic. There are insights shared about longitudinal and lifelong learning and foci on co-evolving processes between the human learner and the computing machines and large language models.

1. INTRODUCTION

In late November 2022, OpenAI shocked the world with its personable and informative chatbot, ChatGPT, dubbed "the poster child of generative AI" (Schäfer, 2023, p. 1). Then, Google rolled out its Bard AI in March 2023. ChatGPT, with its Generative Pretrained Transformer (GPT) aspect encoded in its name, attracted a million users in its first five days of public release and 100 million in the first two months. The tools went truly global, with usage in both developed and developing countries (Kshetri, 2023, p. 16). With web-facing user interfaces and free access (for these initial versions), these were two of the most popular text- and image-making generative artificial intelligence (GAI) tools based on large language models (LLMs). The LLMs do not only create conversation, but they have been harnessed to create various types of digital content (from generic to novel). The advent of "neural networks and

DOI: 10.4018/979-8-3693-0074-9.ch005

learning systems" enable current iterations of machine creativity (Franceschelli, & Musolesi, 2023, p. 3). The speed of the emergence of text-to-image generative AI and other artmaking generative AI tools is explained based on three macro "moments": "the advent of AI in the middle of the last century; the second at the 'reawakening' of a specific approach to machine learning at the turn of this century; the third that documents a rapid sequence of innovations, dubbed 'clever little tricks, that occurred just across 18 months" leading up to the summer of 2022 (Steinfeld, 2023, p. 211). There have been complex advances by generations of computer scientists to enable arrival at the present moment.

Generative AI is defined as "the group of technologies that automatically generate visual or written content based on text prompts" (Inie, Falk, & Tanimoto, Apr. 2023, p. 1) in ways that mimic human creativity and innovations. Another definition of generative AI is "a class of machine learning (ML) algorithms that can learn from content such as text, images, and audio in order to generate new content" (Sun, Liao, Muller, Agarwal, Houde, Talamadupula, & Weisz, Mar. 2022, p. 212). There is a sense that generative AIs may provide inspirations and "higher quality output" with human inputs combined with the capabilities of the AI (Inie, Falk, & Tanimoto, Apr. 2023, p. 2).

ChatGPT and Bard AI brought global attention to the remarkable achievement of computer scientists to artificial intelligences with smooth human-like language abilities (combined with mass-scale learning). The year 2023 "is characterized by a dizzying array of landmark AI developments" (Leong, 2023, p. 52). In terms of generative AIs, people are trying to anticipate "who will win the race" (Aydın, & Karaarslan, 2023, p. 1) for "supremacy in the market" (Rahaman, Ahsan, Anjum, Rahman, & Rahman, 2023, n.p.). Meanwhile a lot of investor dollars are being invested in this "AI gold rush" (Rudolph, Tan, & Tan, 2023, p. 1). Many see this moment as nothing less than an earthshaking one, with anticipated changes to "many aspects of our lives" (Rathore, 2023, p. 63).

Both also brought attention to large language models (LLMs), trained on unlabeled big data from various data sources (including scrapings from the Internet and web (web corpora), various datasets, codesets, imagesets, and others). The models were then created using deep learning methods, in unsupervised machine learning ways, such as artificial neural networks (ANNs). The LLMs result in "tens of millions of weights" from neural network learning but in ways that "are not well understood" (Ulberg, Llach, & Byrne, 2020, p. 508). These are harnessed for generative AI models that can output multimodal contents, foremost highly fluent human-like speech. There are multimodal neural language models that enable joint embedding of images and text (Kiros, Salakhutdinov, & Zemel, 2014) and those engaging multimodal and multimedia data. Multimodal LLMs can take multimodal inputs and deliver multimodal outputs, and can engage multimodal tasks (Huang, Dong, Wang, Hao, Singhal, Ma,... & Wei, 2023, p. 1). A common construct involves generative AIs creating artistic illustrations from brief text prompts (Chen, Chen, Zhao, & Wang, July 2020, p. 155). Generative AIs also create "algorithmically generated narrative" (Thorne, 2020, p. 808), not only in textual form, but multimedia-based forms.

Depending on the size of the model, the amount of data, the types of data, the type of machine learning, and the hardware, LLMs are said to take between 100 days to over a year to train. LLMs are "often trained for hundreds of thousands of compute days" (Zhang, Roller, Goyal, Artetxe, Chen, Chen,... & Zettlemoyer, 2022, n.p.). LLM parameters range from hundreds of billions of parameters to 1.5 trillion and larger. ChatGPT is said to have cost $5 million to train, and the range of general LLMs is between $2 to $12 million USD. LLMs, given their sizes, are "prohibitively expensive in resource-constrained settings" (Roberts, Raffel, & Shazeer, 2020, p. 5; Anderson, Belavy, Perle, Hendricks, Hespanhol, Verhagen, & Memon, 2023, p. 2).