A Study about AI and ML: Exploring the Advancement of Generative AI Models

[1]Nicholes Charles, [2]Sidharth Venugopal, [3]Resmi K R

[1,2] Student, [3]Assistant Professor
[1,2,3]Department of Computer Science,
[1,2]Santhigiri College, Vazhithala, India
[3]CHRIST (Deemed to be University), Bangalore, India

Abstract: This paper provides an overview about artificial intelligence, machine learning and also about generative AI models. The advent of artificial intelligence is a big leap for us. AI is being used from tech giants to small and medium sized enterprises for the growth of their businesses by making best utilization of the data they have. And a lot of AI tools are being released to the market where many of them are free of cost. AI involves the development of intelligent programs that can perform tasks that typically need human intelligence. AI has a history of many years; it's history can be traced back to the 1950s when john McCarthy a famous computer scientist coined the term "Artificial intelligence". Since then, it has witnessed several cycles of growth and declines. Generative AI models like ChatGPT has been huge success since its debut. It was released as a prototype and within short span of time it was invested by Microsoft. This paper also discusses briefly about machine learning. Machine learning comes as the subset of AI, which trains the machines to make decisions or predict based on the data. In conclusion the purpose of this paper is to provide an overview about the history of AI and also about Generative AI.

Index Terms: Machine learning, Artificial Intelligence, Generative AI, Algorithm, Natural Language Processing.

I. INTRODUCTION

Artificial intelligence is a powerful technology that can be used to increase the quality of our life and this same technology can backfire when not used properly. The world is becoming better and better place with self-driving cars and image recognition. All these technical advancements could happen only because of the artificial intelligence. In the year 1997 IBM's dark blue computer could beat human chess champion. In 2016 Google's AlphaG beat the best human go player. AI can be classified into three analytical, human-inspired and humanized ai depending upon the AI it exhibits. Scientists and engineers predict that within few years we will acquire artificial general intelligence. Artificial general intelligence means systems can behave like humans in all aspects like cognitive, emotional [1].

II. HISTORY OF ARTIFICIAL INTELLIGENCE

In the year 1942, a famous science fiction writer Issac Asimov wrote a short story called Runaround. The plot of this novel is about a robot developed by two engineers that evolve around three laws (1) A robot should not injure or harm any human beings in any way (2) A robot should obey the rules of a human being except it breaks the first law (3) A robot should protect itself as long as it doesn't conflict the first two laws. Many scientists and engineers were fascinated with this novel. Around this time an english mathematician developed a machine called "The Bombe" for the British government. This machine was created to decipher the enigma code used by German army in second world war. The task done by "The Bombe" was impossible for even the best mathematicians at that time. This milestone invention made Turing wondered about possibilities of such machines. In 1950 he published a paper "Computing Machinery and Intelligence". This paper described about how to create intelligent machines and how to test their intelligence. The word artificial intelligence was originally coined in the year 1956 when Marvin Minsky, John McCarthy organised 8-week long Dartmouth summer research project on artificial intelligence (DSRPAI) at Dartmouth college. It marked the beginning of AI growth. The participants of this program include computer scientists Nathaniel Rochester who designed IBM 701 which is the first commercial computer. Claude Shannon who founded information theory was also a participant. After the Dartmouth conference there was a significant development in artificial intelligence. One of the developments in this field at that time was ELIZA computer program that was developed between 1964 and 1966. It was a natural language processing tool. Another success story was the development of general problem solver program that could solve problems like tower of Hanoi. All these successes attracted huge investments to ai projects development. In 1973 the US congress strongly opposed the huge funding into artificial intelligence. In the same year, the British mathematician James Light hill published a report by British science research council and claimed machines can never be intelligent like human beings. As a response to it British government ended funding to ai research except three prestigious universities. And American government also reduced investments for AI research [2][3].

III. MACHINE LEARNING

Machine learning is a subset of AI. Machine learning is a data mining technology that makes a computer program to predict the outcome accurately with the accumulated data, it also helps to improve from past results.

There are mainly four types of machine learning methods

- Supervised machine learning
- Unsupervised machine learning
- Semi supervised machine learning
- Reinforcement learning

In supervised machine learning we train the machine using the labelled dataset, and based on this supervised training the output is predicted. What we do here is we take a set of data as input and we train the machine with its output. Consider this example suppose we are training a machine to detect humans and birds. So first we train machine to distinguish humans from birds with features like heights, size body features after the training the machine will be capable to recognize humans and birds. Real life example includes fraud detection, filtering spam. supervised machine learning can be classified to 2 types classification and regression.

In unsupervised model the machine isn't trained to predict or provide output. Here unlabeled data is used for training the machine and machine provides output. For example, we provide the machine a basket of vegetables images and ask to find categories of objects so the machine will analyze the pattern difference and predict the output when it is tested with test dataset. Classifications of unsupervised machine learning include clustering and association.

Semi supervised learning technique is lying between supervised and unsupervised machine learning. In this learning method the machine is trained with labelled and unlabeled datasets. In reinforcement learning the agent learns using feedbacks without any trained data. It's a methodology in which the agent gets positive feedback for good action and negative feedback for bad action. In short, the machine learns to interact and behave in an environment. Real world applications include robotics, video games [4][5].

IV. GENERATIVE AI MODEL

Generative AI comes as the subset of AI that focuses on generating new content. This can include audio and text. The algorithm used in generative AI is trained on large dataset containing examples of large dataset. This makes algorithm capable of learning, understanding the pattern to generate new content. There are several algorithms used here like Generative Adversarial Algorithm (GANs), Variational Autoencoder (VAEs), Recurrent Neural Network (RNN). Algorithm will be chosen based on the problem. The key difference between Generative AI and traditional machine learning is that Generative AI creates new contents, it will be accurate and be unique as a human being thinks. There is currently no specific regulation on artificial intelligence but in the future regulations and guidelines will be developed. There are different types of Generative AI and this section will briefly discuss about it. a) Text To Image Models

This is an artificial intelligence model where the input is text prompt but the output is an image.

- DALL-E-2: DALL-E-2 created by OpenAI is able to generate images and art from text prompt. We can access to this model using OpenAI API. In order to generate the image it uses concept attributes. It uses CLIP neural network for this CLIP (contrastive language – Image Pre-training). It's a neural network which is trained with several text and images. Fig.1 shows the picture generated by DALL-E-2 [1].

Fig 1. Image generated from a text prompt

- IMAGEN: With a deep level of language understanding imagen uses large transformer language models that are pre-trained on large set of text to image generation. This model uses a text-editor that is already pretrained, like BERT to map text to sequence words and then generate images using diffusion. Imagen model is a tool that can compete with DALL-E-2. Fig.2 is a picture of a dog living in sushi home [6].

Fig 2. Corgi in a sushi home

b) Text-To-3D Models

Dreamfusion: Dreamfusion is an AI model developed by Google that can turn your text-prompt into 3-D model. Today 3D models are developed by artists using tools like Blender or Zbrush that require hours of hard work. But for the time being these 3D models generated by Dreamfusion doesn't have the resolution standard for commercial use. Dreamfusion creates 3d model by transferring 2d image text diffusion to 3d objects. imagen is used here to optimize the 3d scene. This technology can disrupt the industry in the future. In education this technology can be made for fun and effective learning and in business it enables companies to create virtual products for demonstration. Fig.3 is the 3D image generated by dreamfusion about a squirrel [1].

Fig 3. 3D image of a squirrel

Magic 3D: This model is a text to 3D model created by Nvidia Corporation. What makes Magic 3D more relevant than dreamfusion is that, dreamfusion has two problems like low quality, low processing time. This problem is solved by two stage optimization frameworks. But Magic 3D solves these two problems and making it more user friendly and popular. c) Image-To-Text Models

Image-to-text is used to generate text from the image as input.

FLAMINGO: Flamingo is a visual language model created by deep mind owned by google. The input to flamingo is photo. The model is trained with large amount of data. Fig. 4 shows the input and output from FLAMINGO [1].

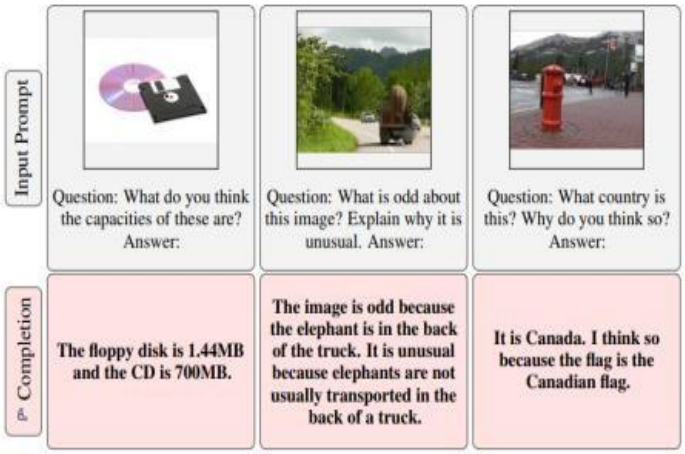

Fig 4. The output generated by FLAMINGO when image is provided as input. d) Text-To-Video Models

Phenaki: Phenaki was developed by Google Research. The API can be accessed from Github. Phenaki is the first AI model that can generate videos from open domain time variable prompt. Fig.5 shows the sequence of images generated by Phenaki [1].

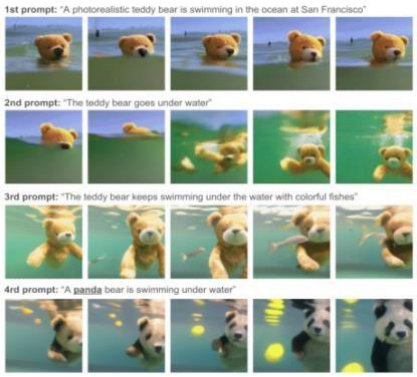

Fig 5. Images created by Phenaki based on the input prompts

e) Text-To-Audio Model

Whisper: whisper is a speech recognition model developed by OpenAI that can identify and translate languages spoken in. It can be accessed by API in Github. The model was trained with 6,80,000 hours of audio from the internet, that include audio from different environment. The model uses encoder-decoder transformer.

Jukebox: Jukebox is a music generation model developed by Open AI[7]. Jukebox generates songs in various genre like hip-hop, Jazz. Jukebox was trained on 1.2 million songs. The model uses hierarchical VQ-VAE architecture. f) Text-To-Text Model

In this AI model the input and output are both texts

ChatGPT: The latest trending innovation in AI is ChatGPT[8]. It was launched on November 30 ,2022. It is built on top of OpenAI's GPT-3 family of large language models. It can think like a human being, it can create code, it can be used to debug code. In fact, it's a program got with human intelligence. Seeing the huge success of this chatbot Microsoft has invested 10 billion used to Open AI and its market cap has become 29 billion used. This chatbot recently appeared for university of Pennsylvania's Wharton business school MBA exam. The course was about operations management course and it has passed the exam. According to the report from semaphore OpenAI has hired hundreds of international contractors to train its ai in software engineering and may be in the future it can be a threat to programmers.

LAMDA: LAMDA stands for language model for dialogue applications. it was developed by google that belongs to the family of conversational neural language models[9]. In June 2022 LAMDA gained widespread attention when a scientist working at google claimed that LAMDA has turned to sentient chatbot but it was widely rejected by scientific community.

PEER: it is developed by Meta AI It is based on four steps plan, edit, explain, repeat. These steps are repeated until the text becomes perfect. It helps in writing tasks and is trained on Wikipedia dataset.

g) Text To Code Methods

Codex: Codex is created by OpenAI, it translates natural languages to computer programs. codex is proficient in more than a dozen language. Codex is most capable in python, but it is also proficient in php, Ruby, Go, Perl. The act of programming can be divided into two. They are

- Breaking the problem into simpler problems
- Mapping those simple problem into existing code like APIs, libraries.

h) Text To Science Model

MINERVA: It's a language model capable of solving mathematical and scientific questions. Minerva solves problems step by step including calculations, symbols etc.[10][11]

V. CONCLUSION

We have seen the potential of artificial intelligence through this paper. We have also seen how creatively AI can work like ChatGPT, Minerva, Dall-E-2 and so on. Generative AI like ChatGPT can even replace several jobs and ChatGPT is used widely by students for their assignments affecting the quality of education. AI is influencing our lives in every aspect like automation, better customer services etc. However, AI is a very innovative tool that can be a useful tool and threat as well.

REFERENCES

1. https://arxiv.org/abs/2301.04655
2. https://journals.sagepub.com/doi/pdf/10.1177/0008125619864925
3. https://ojs.aaai.org/index.php/aimagazine/article/view/1848
4. https://www.javatpoint.com/machine-learning-techniques
5. https://www.javatpoint.com/machine-learning
6. https://imagen.research.google/
7. https://openai.com/blog/jukebox/
8. https://www.businessinsider.com/chatgpt-may-be-fastest-growing-app-in-history-ubs-study-2023-2
9. https://www.searchenginejournal.com/google-lamda-sentient/454820/
10. https://ai.googleblog.com/2022/06/minerva-solving-quantitative-reasoning.html?m=1 [11] https://arxiv.org/abs/2209.12729