Life with Generative AI

Renyu Jiang U18427594 [renyu26@bu.edu](mailto:renyu26@bu.edu)

The seminar I choose to do literature review is a MIT media lab seminar: Life with AI, October 25, 2022, 3:00pm - 5:00pm ET. This is its link: <https://www.media.mit.edu/events/deep-dive-research-panel-life-with-ai/>.

In this seminar, invited speakers mainly focus on an important aspect of AI: Generative AI. With generative AI, people can create content faster and easier. People use deep learning and large data sets to produce text, sound, images, movies, and virtual characters, even proteins and drug structures. In this review, I mainly focus on these topics: The history of generative AI, the applications of generative AI, the impact of generative artificial intelligence on people's lives.

The first topic is the history of generative AI, generative AI has experienced explosive growth over the past seven years, during this period, many significant results were produced.

Initially, artificial intelligence was designed to analyze and recognize certain patterns, which worked well, and then people designed the original generative AI to assist people in creating simple content, for example, predicting next word people want to type, but with the advancement of algorithms and the growth of computing power, generative AI already can create advanced content such as poetry and paintings. Sequoia America published an article titled "Generative AI: A Creative New World", in this article, authors summarize the development history of generative AI into four waves: Small models reign supreme (Pre-2015), the race to scale (2015-Today), better, faster, cheaper (2022+), killer apps emerge (Now) [1].

In wave 1, small models are considered the best for understanding language[1], hence, in this period, most of the generative AIs developed by people are small models, but due to insufficient computing power and shortcomings of algorithms, these small models can only be used for simple problems such as predicting delivery time, and cannot generate advanced content. For example, Gengxin Miao et al. presented a unified generative model, the Optimized Network Model (ONM), and developed a probabilistic algorithm to generate ticket routing recommendations for new tickets in a network of expert groups based on ONM model [2]. This is a small generative model, light and useful for simple content’s generation.

Then, after 2015, generative AI moved to next wave. CPU, GPU, and some other hardware used for computing get a huge boost after 2015, and the amount of data used for training has also increased to a higher level at this time, therefore, the most obvious change in generative AI is the size of the model. With these large models, people can use a lot of computing resources to generate some advanced content of good quality from common paintings to proteins’ 3D structures. The mainstream deep generative models in this period can be classified into the following categories: Autoregressive generative models (ARM), Flow-based models, Latent variable models, and Energy-based models. ARM expresses the distribution in an autoregressive manner, that is, it considers each pixel in the data as an independent random variable and arranges it. PixelCNN is a classic ARM model, when conditioned on class labels from the ImageNet database, the model can generate diverse, realistic scenes representing distinct animals, objects, landscapes, and structures [3]. Flow-Based Models is to realize the conversion from one distribution to another and reversible by constructing a reversible network, e.g., RealNVP. The idea of latent variable models is to assume a low-dimensional latent space and use conditional distributions as generators. GANs and VAEs are examples. Energy based models (energy model) is derived from physics. Direct modeling by constructing an energy function. The models and algorithms at this stage are the theoretical basis for the third and fourth waves. The models at this time have been able to produce some specific advanced contents, although rough, it is still a huge progress if compared to the previous models.

However, when entering 2022, people have new requirements for generative AI: better, faster, and cheaper, because people want to put generative AI into actual production, which requires the model to have better performance, higher quality results and cheaper deployment. Many previous generative models with good results require a lot of computing power and have poor performance, therefore, some researchers began to develop new models or deeply modify previous models to achieve a balance between model performance and result quality while other researchers continue to develop larger models and algorithms to achieve better results. Google's alphafold is a good example, it can predict the structure of proteins with unprecedented speed and experimental-level accuracy (Alphafold 2 can reach this). It used to take months or even years to observe the structure of proteins under the electron microscope, but alphafold can get the same results in less than ten minutes. Sequoia's article also mentioned the fourth wave, which is about killer apps, but I think it should be classified as the third wave, because these apps are all products under the requirements of better, faster, and cheaper.

The second topic is applications of generative AI, after the past seven years of theoretical development, the application of generative AI has entered a blowout period this year, and many applications that break the rules have emerged. This part will introduce some classic applications and cutting-edge applications of generative AI, including applications in media content creation, scientific research, auxiliary engineering design, etc.

Generative AI has made great strides in media content creation, including videos, images, and more. Make-A-Video, an artificial intelligence system model launched by Meta on September 29 this year, is one of such applications. It can generate advanced media content: generate video based on text. Make-A-Video research is based on the latest advances in text-to-image generation technology, which aims to enable text-to-video generation to generate whimsical, one-of-a-kind videos from just a few words or lines of text. Meta’s researchers designed a simple yet effective way to build on T2I models with novel and effective spatial-temporal modules. First, they decompose the full temporal U-Net and attention tensors and approximate them in space and time then, they design a spatial temporal pipeline to generate high resolution and frame rate videos with a video decoder, interpolation model and two super resolution models that can enable various applications besides T2V [4]. This model addresses the two difficulties of "dataset without text-video pairs" and "inference of actions and events", thus the model can generate high-quality videos with coherent motion for different sets of visual concepts. These are two examples of generated video.

A picture containing text, dog

Description automatically generated 

Some examples of Meta’s Make-A-Video model

In the field of scientific research, generative AI is of great help to researchers. Google’s Alphafold mentioned above reflects this. In addition to Alphafold, the model designed by Professor David Baker of the University of Washington is also a cutting-edge application of generative AI in this field. David Baker led their team to develop a brand-new model based on existing models such as RoseTTAFold and Alphafold. Unlike Alphafold, Alphafold mainly uses known gene sequences to predict protein structures, while Professor Baker's model allows AI to "design" brand new proteins. Given the surrounding proteins, this model can repair fragments with determined sequences and structures and completes sequence design and structure prediction. It can design its own proteins in two ways: "constrained hallucination" and "inpainting". The function of the former is to specify a protein target (such as binding to a metal) to the software so that their predicted structure contains the desired functional site, while the latter is to give AI a functional site as a starting point for supplementing protein sequences and structure [5]. This model has already created some original compounds that could be used in industrial responses, cancer treatments, and even in the development of vaccine candidates against respiratory syncytial virus (RSV).

In addition to the two fields mentioned above, generative AI also plays a big role in the field of auxiliary engineering design. In the field of architectural design, many generative AIs already have the ability to directly design interior design CAD drawings according to the needs expressed in natural language, and in many cases, the drawings produced by AI are better than human designers, especially in the details of images. ArchGAN, part of a master's thesis submitted by Stanislas Chaillou in May 2019, can leverage generative adversarial neural networks (GANs) to design floor plans and entire buildings. The model unpacks floor plan design into 3 distinct steps: Building footprint massing, program repartition and furniture layout. Each step corresponds to a Pix2Pix GAN-model trained to perform one of the 3 tasks above. By nesting these models one after the other, Stanislas creates an entire apartment building “generation stack” while allowing for user input at each step [6]. This is just one of the capabilities of a generative AI model in 2019, and similar models in 2022 can do better, some of which can even make renderings based on CAD drawings. In addition to aided design in the traditional engineering field, generative AI also has great potential in the field of computer development. Microsoft's Visual Studio IntelliCode is a set of intelligent assistance functions that can provide functions such as IntelliSense context completion suggestions, code formatting or rule prediction, and improve developer efficiency. Additionally, Google AI lead Jeff Dean described an AI-based approach to chip design that learns from past experience and improves over time to better generate architectures of unseen components. This AI-based approach can complete a design in an average of 6 hours, which is much faster than the weeks required by human experts, and this technology has been applied to Google's next-generation tensor processing unit (TPU v4) Accelerator products.

After the application of generative AI, the next topic is the impact of generative artificial intelligence on people's lives. The impact of anything has two sides, and generative AI is no exception. Both its positive and negative impacts are significant.

The first is the positive impact of generative AI on people's lives. People's perception of new things mostly starts with vision and hearing, such as clearer images generated by AI, tunes composed by AI, etc. In terms of vision, the most well-known example is Nvidia's DLSS, a technology widely used in Nvidia’s graphic devices, which can enhance streams/videos/images for higher definition, and the latest DLSS3.0 technology can even insert generated frames in video streams to achieve higher frame rate. The same technology has also been extended to other products by Nvidia. Nvidia's TV equipment also has the function of upscaling the resolution of streaming media, so that people can get a better visual experience without changing the source. In addition to a better visual experience, generative AI can also create video content. If you pay attention to several mainstream video websites (or Apps) in the world, especially TikTok, an application that focuses on network traffic, you will find that more and more virtual people appear in videos, such as virtual anchors, virtual hosts, etc. The production of these videos with virtual people is quite simple: enter text and a few settings parameters for the environment in which the virtual people are in the video, then you get a video. But many people see it as a shortcut to get traffic revenue, so they make a lot of similar garbage videos. In business, generative AI also has a big impact. Advertising is an indispensable part of modern business, and the selection of target customers is the key to the success of advertising. Generative AI can bring more accurate customer segmentation and selection to commercial companies. Compared with human analysts, AI will not be affected by emotions or people's own preferences, so it is easier to obtain objective and reliable analysis results. Similarly, generative AI tends to have better accuracy in the field of fraud detection. When we turn our attention to the field of engineering, generative AI plays a greater role here. The classic application fields above have mentioned the generation of design drawings/rendering drawings in the traditional engineering field and auxiliary coding of software engineering. These applications have greatly improved the efficiency and accuracy of engineers.

Next is about the negative effects of generative AI. The positive and negative effects of generative AI are like two blades of a sword, and the two are interdependent. The biggest negative impact of generative AI is undoubtedly unemployment. A good example is the virtual host mentioned above. Many human jobs in the past, such as hosts, have been replaced by virtual characters created by generative AI, and this trend continues. In expanding, this caused a lot of unemployment. At present, jobs replaced by generative AI include but are not limited to moderators, anchors, engineers and researchers in some basic departments, etc. In addition to the unemployment problem, generative AI is also widely used in fraud, such as generating virtual characters or voices that can be faked for fraudulent activities. For a long time in the past, the method of telecom fraud was to constantly call the victim, but generative AI has made this method hundreds of times more efficient, and the generated virtual characters are completely untraceable, which makes such activity harder for law enforcement to track and makes people more gullible. Another negative impact of generative AI is copyright issues. Generative AI does not create data out of air, they need data to learn, and most of the data used for learning comes from human content creators, which brings copyright issues. This problem was originally caused by Novel AI, which is an image generation service. Users can enter detailed text descriptions or keywords, and then artificial intelligence will generate anime images. This service has caused dissatisfaction among many animators, because Novel AI use their drawings for learning without authorizations, which then make them lose clients. If humans learn from works on the Internet and then create based on experience, there is no problem, but the unlimited collection and learning of data by generative AI is still a blank space in the law. Many content creators resent the unauthorized learning of their work by generative AI, but due to the extreme difficulty of forensics (the inability to find the source of learning data for generative AI), there is little existing law to support them.

Before the advent of strong general AI, generative AI was undoubtedly the best fusion of computer science and data science. Only by making reasonable use of generative AI and limiting its negative impact can generative AI be better integrated into human society. In summary, generative AI has greatly improved the efficiency and quality of production, and brought more content and convenience to people's lives, but it has also caused many problems, which require better technology means and update laws to resolve.

**References:**

1. <https://www.sequoiacap.com/article/generative-ai-a-creative-new-world/>
2. Miao G, Moser L E, Yan X, et al. Generative models for ticket resolution in expert networks[C]//Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. 2010: 733-742.
3. Van den Oord, A., Kalchbrenner, N., Espeholt, L., Vinyals, O. and Graves, A., 2016. Conditional image generation with pixelcnn decoders. Advances in neural information processing systems, 29.
4. Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., Hu, Q., Yang, H., Ashual, O., Gafni, O. and Parikh, D., 2022. Make-a-video: Text-to-video generation without text-video data. arXiv preprint arXiv:2209.14792.
5. Wang, J., Lisanza, S., Juergens, D., Tischer, D., Watson, J.L., Castro, K.M., Ragotte, R., Saragovi, A., Milles, L.F., Baek, M. and Anishchenko, I., 2022. Scaffolding protein functional sites using deep learning. *Science*, *377*(6604), pp.387-394.
6. <https://developer.nvidia.com/blog/archigan-generative-stack-apartment-building-design/>