The Fascinating Evolution of Generative AI

BEGINNER   CHATGPT   DEEP LEARNING   DIFFUSION MODELS   GENERATIVE AI   LEARNING PATH   LLMS   NLP

Introduction

In the ever-expanding realm of artificial intelligence, one fascinating field that has captured the imagination of researchers, technologists, and enthusiasts alike is Generative AI. These clever algorithms are pushing the limits of what robots can do and understand every day, ushering in a new era of invention and creativity. In this essay, we embark on an exciting voyage through the Evolution of Generative AI, exploring its modest origins, important turning points, and the ground-breaking developments that have influenced its course.

We'll examine how generative AI has revolutionized various fields, from art and music to medicine and finance, starting with its early attempts to create simple patterns and progressing to the breathtaking masterpieces it now creates. We can obtain profound insights into the enormous potential of Generative AI for the future by comprehending the historical backdrop and innovations that led to its birth. Join us as we explore how machines came to possess the capacity for creation, invention, and imagination, forever altering the field of artificial intelligence and human creativity.

Table of contents

Timeline of the Evolution of Generative AI

In the ever-evolving landscape of artificial intelligence, few branches have sparked as much fascination and curiosity as Generative AI. From its earliest conceptualizations to the awe-inspiring feats achieved in recent years, the journey of Generative AI has been nothing short of extraordinary.

In this section, we embark on a captivating voyage through time, unraveling the milestones that shaped Generative AI's development. We delve into key breakthroughs, research papers, and advancements, painting a comprehensive picture of its growth and evolution.

Join us on a journey through history, witnessing the birth of innovative concepts, the emergence of influential figures, and the permeation of Generative AI across industries, enriching lives and revolutionizing AI as we know it.

Year 1805: First NN / Linear Regression

In 1805, Adrien-Marie Legendre introduced a linear [neural network](#) (NN) with an input layer and a single output unit. The network calculated the output as the sum of weighted inputs. Adjust the weights using the least squares method, similar to modern linear NNs, serving as a foundation for shallow learning and subsequent complex architectures.

Year 1925: First RNN Architecture

The first non-learning [RNN architecture](#) (the Ising or Lenz-Ising model) was introduced and analyzed by physicists Ernst Ising and Wilhelm Lenz in the 1920s. It settles into an equilibrium state in response to

input conditions and is the foundation of the first learning RNNs.

Year 1943: Introduction of Neural Nets

In 1943, for the very first time, the concept of [Neural Networks](#) was introduced by Warren McCulloch and Walter Pitts. The working of the biological neuron inspires it. The neural networks were modeled using electrical circuits.

Year 1958: MLP (No Deep Learning)

In 1958, Frank Rosenblatt introduced MLPs with a non-learning first layer with randomized weights and an adaptive output layer. Although this was not yet [deep learning](#) because only the last layer was learned, Rosenblatt basically had what much later was rebranded as Extreme Learning Machines (ELMs) without proper attribution.

Year 1965: First Deep Learning

In 1965, Alexey Ivakhnenko & Valentin Lapa introduced the first successful learning algorithms for deep MLPs with multiple hidden layers.

Year 1967: Deep Learning by SGD

1967 Shun-Ichi Amari proposed training multilayer perceptrons (MLPs) with multiple layers using stochastic gradient descent (SGD) from scratch. They trained a five-layer MLP with two modifiable layers to classify non-linear patterns, despite high computational costs compared to today.

Year 1972: Published Artificial RNNs

In 1972, Shun-Ichi Amari made the Lenz-Ising recurrent architecture adaptive to learn to associate input patterns with output patterns by changing its connection weights. 10 years later, the Amari network was republished in the name of Hopfield Network.

Year 1979: Deep Convolutional NN

Kunihiko Fukushima initially proposed the first CNN architecture, featuring convolutional and downsampling layers, as Neocognitron 1979. In 1987, Alex Waibel combined convolutions, weight sharing, and backpropagation in what he called TDNNs, applied to speech recognition, prefiguring CNNs.

Year 1980: The Release of Auto Encoders

Autoencoders were first introduced in the 1980s by Hinton and the PDP group (Rumelhart,1986) to address the problem of "backpropagation without a teacher" by using the input data as the teacher. The general idea of autoencoders is pretty simple. It consists in setting an encoder and a decoder as neural networks and learning the best encoding-decoding scheme using an iterative optimization process.

Year 1986: Invention of Back Propagation

In 1970, Seppo Linnainmaa introduced the automatic differentiation method called backpropagation for networks of nested differentiable functions. In 1986, Hinton and other researchers proposed an improved backpropagation algorithm for training feedforward neural networks, outlined in their paper "Learning representations by backpropagating errors.

Year 1988: Image recognition (CNN)

Wei Zhang applied back-propagation to train CNN for alphabet recognition, initially known as Shift-Invariant Artificial Neural Network (SIANN). They further applied the CNN without the last fully connected

layer for medical image object segmentation and breast cancer detection in mammograms. This approach laid the foundation for modern computer vision.

Year 1990: Introduction of GAN / Curiosity

[Generative Adversarial Networks](#) (GANs) have gained popularity since their first publication in 1990 as Artificial Curiosity. GANs involve two dueling neural networks, a generator (controller) and a predictor (world model), engaged in a minimax game, maximizing each other's loss. The generator produces probabilistic outputs, while the predictor predicts environmental reactions. The predictor minimizes error through gradient descent, while the generator seeks to maximize it.

Year 1991: First Transformers

[Transformers](#) with "linearized self-attention" were first published in March 1991, so-called "Fast Weight Programmers" or "Fast Weight Controllers". They separated storage and control like in traditional computers but in an end-to-end-differentiable, adaptive, fully neural way. The "self-attention" in standard Transformers today combines this with a projection and softmax like the one introduced in 1993.

Year 1991: Vanishing Gradient

The Fundamental Deep Learning Problem, discovered by Sepp Hochreiter in 1991, addresses the challenges of [deep learning](#). Hochreiter identified the issue of vanishing or exploding gradients in deep neural networks, i.e., backpropagated error signals either diminish rapidly or escalate uncontrollably in typical deep and recurrent networks.

Year 1995 – The Release of LeNet-5

Several banks applied LeNet-5, a pioneering 7-level convolutional network by LeCun in 1995 that classifies digits to recognize hand-written numbers on checks.

Year 1997 – Introduction of LSTM

In 1995, [Long Short-Term Memory](#) (LSTM) was published in a technical report by Sepp Hochreiter and Jürgen Schmidhuber. Later, in 1997, the main LSTM paper dealt with the vanishing gradient problem. The initial version of the LSTM block included cells, input, and output gates. In 1999, Felix Gers and his advisor, Jürgen Schmidhuber and Fred Cummins, introduced the forget gate into the LSTM architecture enabling the LSTM to reset its state.

The Millennium Developments

Year 2001 – Introduction of NPLM

In 1995, we already had an excellent neural probabilistic text model whose basic concepts were reused in 2003, i.e., Pollack's earlier work on embeddings of words and other structures and Nakamura and Shikano's 1989 word category prediction model. In 2001, researchers showed that [LSTM](#) could learn languages unlearnable by traditional models such as HMMs, i.e., a neural "subsymbolic" model suddenly excelled at learning "symbolic" tasks.

Year 2014 – Variational Autoencoder

A [variational autoencoder](#) is an autoencoder whose training is regularised to avoid overfitting and ensure that the latent space has suitable properties that enable a generative process. The architecture of VAE is similar to Autoencoder, with a slight modification of the encoding-decoding process. Instead of encoding an input as a single point, researchers encode it as a distribution over the latent space.

Year 2014 – The Release of GAN

The researchers proposed a new framework for estimating [generative models](#) via an adversarial process in which simultaneously two models are trained. A generative model, G captures the data distribution, and a discriminative model, D, estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake.

Year 2014 – The Release of GRU

A [gated recurrent unit](#)(GRU) was proposed by Cho [2014] to make each recurrent unit adaptively capture dependencies of different time scales. Similarly to the LSTM unit, the GRU has gating units that modulate the flow of information inside the unit, however, without having a separate memory cell.

Year 2015 – The Release of Diffusion Models

Diffusion models are the backbone of image generation tasks today. By decomposing the image formation process into a sequential application of denoising autoencoders, diffusion models (DMs) achieve state-of-the-art synthesis results on image data and beyond. Additionally, their formulation allows a guiding mechanism to control the image generation process without retraining.

Year 2016 – The Release of WaveNet

WaveNet is a language model for audio data. It's a deep neural network for generating raw audio waveforms. The model is fully probabilistic and autoregressive, with the predictive distribution for each audio sample conditioned on all previous ones.

Year 2017: The Release of Transformers

Google introduced a revolutionary paper in 2017, "Attention Is All You Need". LSTMs were dead and no more! This paper introduced a new architecture completely relying on attention mechanisms. The fundamental elements of Transformers are Self Attention, Encoder Decoder Attention, Positional Encoding, and Feed Forward Neural Network. The fundamental principles of Transformers remain the same in the LLMs today as well.

Year 2018: The Release of GPT

GPT (Generative Pretraining Transformer) was introduced by OpenAI by pretraining a model on a diverse corpus of unlabeled text. It's a Large Language Model trained autoregressively to predict a new sequence of words in the text. The model largely follows the original transformer architecture but contains only a 12-layer decoder only. In upcoming years, the research led to the development of larger models in size: GPT-2(1.5B), GPT-3(175B)

Year 2018: The Release of BERT

BERT (Bidirectional Encoder Representations from Transformers) was introduced by Google In 2018. The researchers trained the model in 2 steps: Pretraining and Next Sentence Prediction. The model predicts missing tokens present anywhere in the text during pretraining, unlike GPT. The idea here was to improve language understanding of the text by capturing the context from both directions.

Year 2019: The Release of StyleGAN

The researchers proposed an alternative generator architecture for generative adversarial networks, borrowing from style transfer literature. The new architecture enables automatic learning of high-level attributes (e.g., pose and identity in human faces) and stochastic variations (e.g., freckles, hair) in generated images. It also allows easy, scale-specific control of the synthesis.

Year 2020: The Release of wav2vec 2.0

In 2019, Meta AI released wav2vec, a framework for unsupervised pre-training for speech recognition by learning representations of raw audio. Later, in 2020, wav2vec 2.0 was introduced for Self-Supervised Learning of Speech Representations. It learns the most powerful representation of the speech audio. The model was trained using connectionist temporal classification (CTC), so the model output has to be decoded using Wav2Vec2CTCTokenizer.

Year 2021: The Release of DALL.E

DALL·E is a 12-billion parameter version of GPT-3 trained to generate images from text descriptions using a dataset of text–image pairs. It has diverse capabilities, like creating anthropomorphized versions of animals and objects, combining unrelated concepts, rendering text, and transforming existing images.

Year 2022: The Release of Latent Diffusion

Latent diffusion models achieve a new state of the art for image inpainting and highly competitive performance in image generation. Researchers use powerful pretrained autoencoders to train diffusion models in the latent space and cross-attention layers. For the first time, this allows them to achieve a near-optimal point between complexity reduction and detail preservation, greatly boosting visual fidelity.

Year 2022: The Release of DALL.E 2

In 2021, researchers trained DALL.E, a 12-billion parameter version of GPT-3, to generate images from text descriptions using a dataset of text−image pairs. In 2022, DALL·E 2 was developed to create realistic images and art from a description in natural language.DALL·E 2 can create original, realistic images and art from a text description. It can combine concepts, attributes, and styles.

Year 2022: The Release of Midjourney

Midjourney is a very popular text-to-image model powered by the latent diffusion model. A San Francisco-based independent research lab creates and hosts it. It can create high-quality definition images via natural language descriptions known as prompts.

Year 2022: The Release of Stable Diffusion

Stable Diffusion is a latent text-to-image diffusion model capable of generating photo-realistic images given any text input, cultivates autonomous freedom to produce incredible imagery, and empowers billions of people to create stunning art within seconds.

Year 2022: The Release of ChatGPT

ChatGPT is a revolutionary model in the history of AI. It is a sibling model to InstructGPT, trained to follow instructions promptly and provide a detailed response. It interacts in a conversational format that makes it possible for ChatGPT to answer follow-up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

Year 2022: The Release of AudioLM

AudioLM is a framework from Google for high-quality audio generation with long-term consistency. AudioLM maps the input audio to a sequence of discrete tokens and casts audio generation as a language modeling task in this representation space. Given the prompt (speech/music), it can complete it.

2023 Unleashed: Exploring the Hottest Recent Releases

Year 2023: The Release of GPT-4

[GPT-4](#) is OpenAI's most advanced system, producing safer and more useful responses. GPT-4 can solve complex problems more accurately, thanks to its broader general knowledge and problem-solving abilities. It surpasses GPT-3.5 with its Creativity, Visual input, and Longer Context.

Year 2023: The Release of Falcon

[Falcon](#) LLM is a foundational large language model (LLM) with 40 billion parameters trained on one trillion tokens. Falcon ranks on the top of the Hugging Face Open LLM Leaderboard. The team placed a particular focus on data quality at scale. They took significant care in building a data pipeline to extract high-quality web content using extensive filtering and deduplication.

Year 2023: The Release of Bard

Google released [Bard](#) as a competitor to ChatGPT. It is a conversational generative artificial intelligence chatbot by Google. Based on the PaLM foundation model, Bard interacts conversationally, answering follow-up questions, admitting mistakes, challenging incorrect premises, and rejecting inappropriate requests.

Year 2023: The Release of MusicGen

[MusicGen](#) is a single-stage auto-regressive Transformer model capable of generating high-quality music samples conditioned on text descriptions or audio prompts. The frozen text encoder model passes the text descriptions to obtain a sequence of hidden-state representations.

Year 2023: The Release of AutoGPT

[Auto-GPT](#) is an experimental open-source application showcasing the capabilities of the GPT-4 language model. This program, driven by GPT-4, chains together LLM "thoughts" to autonomously achieve whatever goal you set. As one of the first examples of GPT-4 running fully autonomously, Auto-GPT pushes the boundaries of what is possible with AI.

Year 2023: The Release of LongNet

Scaling sequence length has become a critical demand in the era of large language models. However, existing methods struggle with computational complexity or model expressivity, restricting the maximum sequence length. LongNet, a Transformer variant, can scale sequence length to more than 1 billion tokens without sacrificing the performance on shorter sequences.

Year 2023: The Release of Voicebox

Meta AI announced Voicebox, a breakthrough in generative AI for speech. The researchers developed Voicebox, a state-of-the-art AI model capable of performing speech generation tasks — like editing, sampling, and stylizing — through in-context learning, even without specific training.

Year 2023: The Release of LLaMA

Meta AI introduced LLaMA, a collection of foundation language models ranging from 7B to 65B parameters. They showed that it is possible to train state-of-the-art models using publicly available datasets exclusively without resorting to proprietary and inaccessible datasets. In particular, LLaMA-13B outperforms GPT-3 (175B) on most benchmarks.

Conclusion

Looking back at the timeline of Generative AI, we witnessed how it overcame challenges and limitations, constantly redefining what was once thought impossible. The groundbreaking research, pioneering models, and collaborative efforts have shaped this field into a driving force behind cutting-edge innovations.

Beyond its applications in art, music, and design. Generative AI significantly impacts various fields, like healthcare, finance, and NLP, improving our daily lives. This progress raises the potential for harmonious coexistence between technology and humanity, creating countless opportunities. Let's dedicate ourselves to developing this outstanding field, encouraging cooperation and exploration in the coming years.

Article Url - https://www.analyticsvidhya.com/blog/2023/07/the-fascinating-evolution-of-generative-ai/

Sakshi Khanna