World Scientific Research Journal Volume 10 Issue 12, 2024
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202412_10(12).0010

AI Revolution and One Application in Music Composition

Emily Hu
New Trier High School Winnetka, IL 60093 USA

Abstract

In recent years with the development of ChatGPT, music generation has become a very popular option for musicians or creators that want to create their own customized music for a specific purpose. In this paper we study how technological innovations over time have led us to the level of machine learning we have today, as well as an application of a recent model, the transformer, in music generation. The complexity of machine learning today has the capability for profound impacts on human lives and how the world operates. Many of the technological innovations we have today would not have been possible without the developments of old inventions. Each invention has played a significant role in preparing and allowing the next era of inventions to begin. We will discuss how it eventually led up to developments in AI and why it was necessary. Unlike simply generating text, music has very specific qualities to it that some humans can not even perfect. Along with rhythm, melody, and pitch, emotion also plays a significant role in the sound of music. There are also many different voices and sounds from different instruments. There might also be some underlying message or tone that is communicated throughout the music. While generating music is always impossible to master, the application of transformers using the attention mechanism that this paper will explore is a viable way to create music stylistically similar to any given training set. Using this method, musicians or anyone could generate a piece of music of any choice of mood, genre, etc. using a huge amount of training data.

Keywords

Technology revolutions, Machine learning, Artificial intelligence, Generative ai, Music composition.

1. EVOLUTION OF TECHNOLOGICAL INNOVATIONS TO AI

1.1. The Rise of AI

In 2017, Google AI's AlphaGo defeated the human Go game world champion [1]. Then, AlphaGo was repurposed as AlphaZero to conquer chess. It beat the chess champion without human knowledge with four hours of training [2].

There is no doubt that these highly recent advances in technology are changing how people carry out their daily routines. OpenAI released ChatGPT [3], the GPT-3.5, in November 2022. In just under five days, ChatGPT received its first one million users [4]. It took Twitter (now X) two years just to achieve that same milestone. Similarly, ChatGPT obtained its first 100 million users in under two months, while Tiktok took nine months. But it didn't stop there. Less than half a year later, in March of 2023, OpenAI released a much more advanced model of what was already seemingly impossible: GPT-4. [5] ChatGPT was trained on only 175 billion parameters, while GPT-4 was on a predicted 100 trillion parameters [6]. ChatGPT is a chatbot which can answer any question people ask. It can write articles, reply to emails, code like programmers, invest in stock markets, play chess, diagnose patients, and provide legal help, etc., like a well-trained professional in all disciplines that have open data for OpenAI to access [7]. ChatGPT can pass

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202412_10(12).0010

The United States Medical Licensing Examination (USMLE) with a 62.5% correct rate, while GPT-4 can pass with an excellent 90% [8]. Both AI models, especially GPT-4, can provide interactive and professional chat answers in clinical medicine and meet the needs of soft skills in interpersonal communication and ethical suggestions. Similarly, the powerful GPT-4 passed the uniform bar examination with a score of 297 with a significant margin of the top 10% percentile [9].

1.2. The First Industrial Revolution: the Age of Machine I

The First Industrial Revolution lasted from around 1760 to 1840, mainly in Britain. Hargreaves invented the Spinning Jenny around 1764 [10]. Many farmers were displaced by the enclosures and had to head to cities to find work [11]. This technological innovation brought the birth of modern textile factories and created new job opportunities for those unemployed farmers in the cities. James Watt improved the steam engine in 1765, one of the most significant inventions of the Industrial Revolution [12]. It freed humans from manual labor with machines in modern factories, and further provided job opportunities for farmers in the cities. This era is also called the age of machines. Machines replaced human and animal power. Work efficiency greatly improved, which, in turn, triggered more technological innovations, such as steamed boats and railways[12]. As a result, Britain became the center of manufacturing. Work and lifestyle in Britain completely changed, and industrialization caused urbanization. The percentage of Britain's urban population changed from 22% in 1750 to 51% in 1850[13]. London's population grew from 650,000 in 1750 to 1,950,000 in 1841[14]. Britain became the world center of producing and exporting goods worldwide [15]. It is also noted that success in international trade resulted in high wages for industry workers in Britain, causing demand for technology that substituted capital and energy for labor[17]. These technological advancements did not replace people's jobs, but they created more opportunities and, in turn, resulted in more technological innovations. They even had to use children for labor because there was a shortage of workers.

1.3. The Second Industrial Revolution, the Age of Machine II

The Second Industrial Revolution lasted from approximately 1870 until 1945. It continued the First Revolution and expanded from Britain into Western Europe, the United States, Japan, and more. The light bulb, telephones, cars, airplanes, and radios were major inventions [16]. Electricity and oil dominated energy consumption. The Second Industrial Revolution significantly improved work efficiency to manufacture more goods in cities and increased urbanization in those technology-influenced countries. Many skilled middle-class workers emerged, and the consumer demand-originated economy started to shape social culture. People used telephones to communicate, traveled more often, and watched movies etc. The world population increased from 1.2 billion in 1850 to 2.5 billion in 1950[17]. Britain's urban ratio increased to 80% in 1950. Similarly, the United States' urban ratio increased from 15% in 1850 to 64% in 1950[18]. Worldwide, countries use less and less human labor for the agriculture industry.With the advancement of technologies, populations continued to grow. Without new technological innovations, society would not grow and never advance at a pace similar to this. At the same time, modern mathematics and physics development peaked around 1951, with no big breakthrough after that[19]. Lord Kevin mentioned in 1900 that the foundation of physics and mathematics had been built, and two clouds led to the revolution of physics: relativity and quantum mechanics[20]. The strong foundation built hints that another breakthrough in the industry revolution is coming soon.

1.4. The Third Industrial Revolution, the Information Age I: Computer

The Third Industrial Revolution is also known as the technology or information revolution. It started in approximately 1950 and is projected to last until 2030 [21]. Nuclear power has

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202412_10(12).0010

provided more efficient electricity production than coal since 1942 after quantum mechanics and relativity theory were established [20]. Spaceflight launched the first satellite, and computers have improved human productivity since 1974. On the same year, the internet has started the information innovation [12]. Email made information exchangeable instantly since 1993. Google can search information since 1998. From 2004, Facebook has made social networking possible. YouTube has created a video streaming platform after 2005 [22]. So far, it has been the most influential technological revolution in human history. The human population skyrocketed from 2.5 billion in 1950 to 7.9 billion in 2021 [23] The urban ratio changed from 64% to 82% in the United States [18]. Additionally, the information revolution made information spread much faster than in the past, and China's urban ratio improved from 19% in 1980 to 66% in 2023[24]. The most important invention was the computer. It freed people from brain and intelligence work. As a result, people have more time to socialize and relax, eventually making them more innovative. The advancement of technology in computers did not replace jobs; instead, it helped people improve efficiency. This allows more time and energy for brainstorming new methods of advancing technology.

1.5. The Fourth Industrial Revolution, the Information Age II: AI

The Third Industrial Revolution is ongoing, but the fourth industry revolution started when artificial intelligence grew in 2017[12] when AlphaGo beat human champions of Go. Machine learning and artificial intelligence are considered significant components of the Fourth Industrial Revolution, along with other technologies such as virtual reality, autonomous driving, and sustainable energy[25] [26]. Solar, wind, and electricity storage, such as Tesla MegaPack[27] help people not depend on the energy limited by earth reserve, for example 200 years of uranium[28] and 50 years of crude oil[29]. Nvidia's market cap increased from 350 billion in November 2022 to 3.1 trillion in June 2024[30][31] since OpenAI relies on Nvidia's 25,000 A100 GPUs to train ChatGPT and GPT-4. This is more than the entire German stock market cap[32]. It took one hundred days of energy supply and computing power to train GPT-4[33]. Recurrent Neural Networks were first conceptualized in 1986 by David Rumelhart[34], and it was a relatively successful model for many useful tasks such as language translation and natural language processing[35]. Big data[36] is now available in the cloud with the development Internet technologies. The scarcity of data and the limitation of computing power have delayed the application of artificial intelligence until recently. ChatGPT has turned the clock of artificial intelligence quickly since November 2022. Many tech companies started pushing out their own AI models. Meta released LlaMA in Feb 2023, Google released PaLM in May 2023, Microsoft released Phi-1 in June 2023, Tesla xAI released Grok-1 in Nov 2023, Google released Gemini in December 2023, OpenAI released Sora in February 2024, Anthropic released Claude 3 in June 2024, and Amazon will launch Olympus very soon[37]

The breakthrough of technologies such as ChatGPT and artificial intelligence is an inevitable outcome of past technology developments. The Biden administration and the Department of Labor of the United States urge companies to use AI to enable workers[38][39].

2. MUSIC GENERATION: AN APPLICATION OF MACHINE LEARNING

Now, machine learning is quickly advancing as a more viable tool for many different everyday activities as its concept becomes more common and attracts more research attention. After recent advancements in autoregressive models in machine learning, music generation has been possible using the idea of sequence prediction and transformers. The invention of transformer models in 2017 is most beneficial for music generation in comparison with traditional models like Recurrent Neural Networks (RNN) because of its ability to capture long-range and extremely complex relationships between notes and efficient training on large datasets, which is crucial for generating coherent and advanced music [40][41]. But rather than the traditional

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202412_10(12).0010

transformer, we will use sparse transformers because it can reduce the computational power required to train the music model, as music can progress very long. Instead of every output position in the attention matrix calculating weights for every input position, which is for every single layer and attention head, the output position calculates for only a subset of input positions[42].This model can still effectively generate and accurately predict a future note given previous notes in a training dataset. To successfully complete this task, the data can be passed correctly to and evaluated within a transformer in the following steps.

The first step is data preprocessing. After loading MIDI files, one can convert the notes and signs into readable text by extracting each individual note pitch and its duration in a score into two different text strings. For example, the note pitch might appear as "G3" and the duration might be "0.25." After successfully parsing the notes, we can tokenize the notes and duration separately in order to create a training set for the model to be able to start processing, converting them into numerical data. Finally, to create the inputs and outputs of the training set, the note and duration strings are split into groups of a specific amount of elements with the output being the input shifted one note to the future. This creates a system that allows the model to train off any sequence of notes as an output can simply be the next note given the previous notes. Therefore, the transformer predicts this future note based on all previous notes and durations[43].

After the training set is created, sine position encoding can be used to encode the position of each token. Transformer models, unlike RNN models, are unable to distinguish different sequence arrangements, so positional encoding is necessary to handle this situation. Using sine position encoding rather than a simple embedding layer is more helpful as in embedding layers, we are required to choose a maximum length and can adapt to this while extrapolating to sequences longer than this are impossible. Music does not have a set limit and can sometimes be very long. Additionally, a token embedding layer is needed to capture relationships between different notes and reduce dimensionality. These token and position embedding layers will be combined together into one embedding for the data.

Because there are two different input streams, notes and durations, we can combine them in some way for the transformer to be able to process them. One method is having the notes input and durations input have their own independent token and positional embedding, combining them in a concatenation layer to represent one single input for the transformer block to be able to accept, and finally separating into two different dense layers again for notes and duration. Ulimately, the final output is a prediction of the next note and its duration given the previous notes. However, these previous notes are not accounted for equally; some may have more significance than others. The mechanism in the transformer involved with this is the concept of self-attention. It determines what notes are more relevant than others and chooses to focus on those. This is done by computing how much a key vector of one note matches with a query vector of another note, resulting in a calculated attention score which reflects the value of representation of that note. These calculations allow some notes to be more relevant and emphasized than others. We can also use multiple attention heads, which will eventually be concetenated together, to capture more complex relationships because of each of their distinct foci. These are key factors that contribute to the accuracy and prediction of a future note as it accounts for many more significant relationships between the previous notes [44]. As shown in Fig 1, this concept can be illustrated with matrix calculation of the attention scores for inputs.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202412_10(12).0010

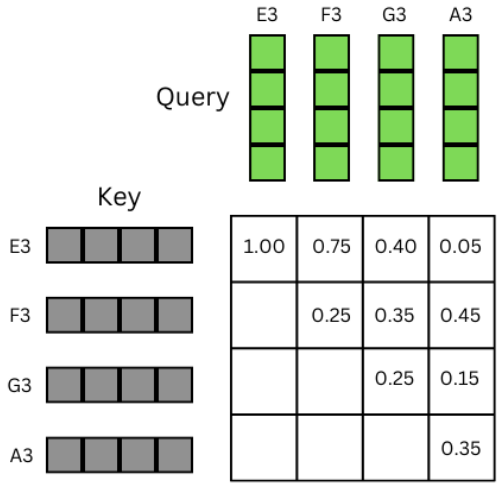

Figure 1. Attention score calculation matching notes, represented as key vectors, with query vectors of other notes. Higher attention score represents a closer prediction match, signifying the note's higher significance

3. ANALYZING RESULTS

In the process of training, every epoch will refine a more sophisticated and advanced which is able to recognize patterns better. It can be more certain of what output would be next. This works because of how the transformer runs. The concept of transformers is that they calculate attention scores of each token by comparing query vectors with tokens' key vectors to determine how much weight each token should have. It can decide what notes are more important than others in its final prediction. As it learns for longer, it can refine its attention matrices by focusing on what tokens are more relevant[45]. If we evaluate the attention matrix, the model evidentally places most of its attention on the most recent notes as well as the key signature. This makes sense because the recent notes are what the model needs to continue the phrase and help implement any patterns it detected from before, and the key signature helps the model decide if the potential note guess fits best with the key of the music, and might adjust given this.

4. FUTURE WORK

There are many more methods for music generation that are better at capturing different types of relationships. For example, differently from transformers which specialize in generating music that emphasizes note relationships, MusicGAN may focus more on the realism and listenability of the generated notes rather than whether the notes have a relationship with the past notes. These varieties of different methods show the shortcomings that only using transformers has on the generation of music. There is more than calculating whether or not a note is significant enough to be added. Finding a combination of these models and being able to generate human connection to the notes is the most difficult part, but it is the first step forward until being able to generate coherent music at a similar level as composed by humans.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Koehn for his expertise and guidance in this field of the evolution of machine learning, transformers, and sequential predictions. He has provided lots of insight and helped guide me through the process of transformer application in ChatGPT and

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202412_10(12).0010

generative music composition. I appreciate his patience and everlasting support during my study.

I greatly appreciate Dr. Zhao for his insight and ideas, which greatly contributed to my understanding of transformers and machine learning. He offered lots of help and provided many comments as I completed my paper.

REFERENCES

[1] BBC , May 25, 2017 , " Google AI defeats human Go champion", https://www.bbc.com/news/technology-40042581

[2] Samuel Gibbs," AlphaZero AI beats champion chess program after teaching itself in four hours", Guardian, Dec 7, 2017, https://www.theguardian.com/technology/2017/dec/07/alphazero-google-deepmind-ai-beats-champion-program-teaching-itself-to-play-four-hours

[3] Aleksandra Yosifova, "The Evolution of ChatGPT: History and Future" Aug 14, 2023 https://365datascience.com/trending/the-evolution-of-chatgpt-history-and-future/

[4]]Fabio Duarte, "Number of ChatGPT Users (Jun 2024)" June 8, 2024 , https://explodingtopics.com/blog/chatgpt-users

[5] Kristi Hines , "History Of ChatGPT: A Timeline Of The Meteoric Rise Of Generative AI Chatbots", June 4, 2023, https://www.searchenginejournal.com/history-of-chatgpt-timeline/488370/

[6] Devashish Mamgain , "ChatGPT 3.5 vs 4 vs 4o – Key Differences to Consider" Jun 17 2024 https://www.kommunicate.io/blog/chatgpt-4-vs-chatgpt-3-5-key-differences/

[7] OPENAI , "What is ChatGPT", https://help.openai.com/en/articles/6783457-what-is-chatgpt

[8] Dana Brin et al, "Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments", Nature Oct 1, 2023, https://www.nature.com/articles/s41598-023-43436-9

[9] Lara Kimmel, "ChatGPT Passed the Uniform Bar Examination: Is Artificial Intelligence Smart Enough to be a Lawyer?", Apr 7, 2023, https://international-and-comparative-law-review.law.miami.edu/chatgpt-passed-the-uniform-bar-examination-is-artificial-intelligence-smart-enough-to-be-a-lawyer/#:~:text=ChatGPT%2D4's%20proposed%20wondrous%20abilities,90th%20percentile%20of%20test%2Dtakers

[10] Mary Bellis, James Hargreaves and the Invention of the Spinning Jenny, ThoughtCo,https://www.thoughtco.com/who-invented-the-spinning-jenny-4057900

[11] The Land Magazine, "A Short History of Enclosure in Britain", https://www.thelandmagazine.org.uk/articles/short-history-enclosure-britain

[12] Eric Gregersen, "History of Technology Timeline, " Britannica , https://www.britannica.com/story/history-of-technology-timeline

[13] Long Run Health Matters, "Urbanisation and disease: a perspective from nineteenth-century Britain", Oct 17,2021, https://lrhmatters.com/drivers-of-health/urbanisation-and-disease-in-industrializing-britain

[14] Demographia, "Greater London,Inner London & Outer London Population & Density History", http://www.demographia.com/dm-lon31.htm

[15] Robert Allen, "Why was the Industrial Revolution British?" May 15, 2019 , https://cepr.org/voxeu/columns/why-was-industrial-revolution-british#:~:text=Success%20in%20international%20trade%20created,capital%20and%20energy%20for%20labour

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202412_10(12).0010

[16] Richmond Vale Academy, "The Second Industrial Revolution: The Technological Revolution", May 16, 2022 , https://richmondvale.org/second-industrial-revolution/

[17] Statista , "Estimated global population from 10,000BCE to 2100",https://www.statista.com/statistics/1006502/global-population-ten-thousand-bc-to-2050/

[18] Statista "Degree of urbanization in the United States from 1790 to 2020, and with projections until 2050", https://www.statista.com/statistics/269967/urbanization-in-the-united-states/

[19] 12000 ORG, "Mathematics and Physics Timeline", https://www.12000.org/my_notes/science_time_line/index.htm

[20] Pierre Meystre, "Quantum optics and the search for new physics, Springer Nature, https://www.springernature.com/gp/pierre-meystre/20136468

[21] Haradhan Kumar Mohajan, "Third Industrial Revolution Brings Global Development ", Sep 10, 2021 , Journal of Social Sciences and Humanities, Vol. 7, No. 4, 2021, pp. 239-251 https://mpra.ub.uni-muenchen.de/110972/1/MPRA_paper_110972.pdf

[22] LukeMcC, "Third Industrial Revolution", https://www.timetoast.com/timelines/third-industrial-revolution

[23] Our world in Data, "Population Growth" , https://ourworldindata.org/population-growth

[24] Statista, "Degree of urbanization in China in selected years from 1980 to 2023", https://www.statista.com/statistics/270162/urbanization-in-china/

[25] Devon McGinnis, "What Is the Fourth Industrial Revolution?", July 5, 2023, https://www.salesforce.com/blog/what-is-the-fourth-industrial-revolution-4ir/

[26] McKingsey & Company, "What are Industry 4.0, the Fourth Industrial Revolution, and 4IR", Aug 17, 2022, https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-are-industry-4-0-the-fourth-industrial-revolution-and-4ir

[27] Tesla https://www.tesla.com/megapack

[28] Steve Fetter "How long will the world's uranium supplies last?", Jan 26, 2009, SCI America, https://www.scientificamerican.com/article/how-long-will-global-uranium-deposits-last/

[29] Hannah Osborne, "How much oil is left and will we ever run out?", Live Science, May 31, 2024, https://www.livescience.com/planet-earth/how-much-oil-is-left-and-will-we-ever-run-out#:~:text=We%20may%20never%20run%20out,50%20years%2C%20current%20estimates%20suggest

[30] Companies Market Cap, https://companiesmarketcap.com/nvidia/marketcap/

[31] NATHAN REIFF, "How Nvidia Makes Money", Mar 13, 2024, https://www.investopedia.com/how-nvidia-makes-money-4799532

[32] Statista, "Total market capitalization of all domestic shares listed on the Frankfurt Stock Exchange from 2002 to December 2023",https://www.statista.com/statistics/1203216/frankfurt-stock-exchange-market-cap/

[33] Jijo Malayil, "OpenAI receives the world's most powerful AI GPU from Nvidia CEO", Apr 25, 2024, https://interestingengineering.com/innovation/nvidia-ai-gpu-openai

[34] Click Worker, "Recurrent Neural Networks"https://www.clickworker.com/ai-glossary/recurrent-neural-networks/

[35] Amit Prakash, "What is transformer architecture and how does it power ChatGPT?", Feb 22, 2023, https://www.thoughtspot.com/data-trends/ai/what-is-transformer-architecture-chatgpt

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202412_10(12).0010

[36] Google Cloud, "What is Big Data?", https://cloud.google.com/learn/what-is-big-data#:~:text=Big%20data%20refers%20to%20extremely,to%20grow%20exponentially%20over%20time.

[37] Life Architect AI "Timeline of AI and language models". https://lifearchitect.ai/timeline/

[38] US DOL, "Department of Labor's Artificial Intelligence and Worker Well-being: Principles for Developers and Employers" https://www.dol.gov/general/AI-Principles

[39] Adam DeRose, "Technically HR: Biden administration releases guidelines for bringing AI to work", HR Brew, May 23, 2024, https://www.hr-brew.com/stories/2024/05/23/technically-hr-biden-administration-releases-guidelines-for-bringing-ai-to-work

[40] Sanchit Goel, "Evolution of Transformers -- Part 1", Jul 21, 2022, https://sanchman21.medium.com/evolution-of-transformers-part-1-faac3f19d780#:~:text=Transformer%2C%20a%20revolutionary%20model%2C%20was,RNNS%20take%20in%20sequential%20data

[41] Philipp Koehn, "Neural Machine Translation", May 2020

[42] Rewon Child, Scott Gray, Alec Radford, Ilya Sutskever, "Generating Long Sequences with Sparse Transformers", Apr 23, 2019, https://doi.org/10.48550/arXiv.1904.10509

[43] David Foster, "Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play", May 2023

[44] Zhuokai Zhao, "Enhanced Data Visualization for Trustworthy Deep Learning", Aug 2024

[45] Ashish Vaswani et al, "Attention is All You Need", Jun 12, 2017, https://doi.org/10.48550/arXiv.1706.03762