

Generative AI: Opportunities and Challenges

K.R. Chowdhary, Professor of CS

Email: kr.chowdhary@iitj.ac.in

Webpage: <http://www.krchowdhary.com>

Former Prof. & Head, CS dept, MBM and former faculty at IITJ

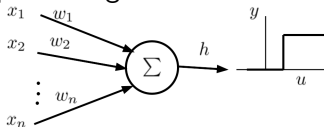
December 20, 2023

Generative AI involves using artificial neural networks to generate new data that is similar to the data it has been trained on. One of the most famous architectures for generative AI is the Generative Adversarial Network (GAN).

- A GAN consists of two main components: *a generator* and a *discriminator*.
 - **Generator** takes random noise as input and transforms it into data.
 - Through training, the generator's neural network learns to map the noise to data that becomes more and more similar to the real data
 - **Discriminator** is a separate neural network that takes both real data from the training set and generated data from the generator as input. The discriminator's job is to classify whether the input data is real or generated.

Working of ChatGPT (Generative Pre-trained Transformer)

- It is based on a transformer architecture – neural network that has been shown to be very effective for natural language processing tasks.



- It is trained on a large dataset of text, such as books ...
- When given a prompt or context, the model

processes the input and generates a response (Bayes rule).

- The model uses attention mechanisms to focus on the most relevant parts of the input [1].
- After the training, the model can be fine-tuned on a specific task...
- It can also generate text in multiple languages by fine-tuning

The key idea behind GANs is the adversarial training process.

- *Initialization* of generator and discriminator with random weights.
- *Iteration*: Each iteration has two steps :generator and discriminator.
- *Discriminator Step*: Trained on a mix of real data and generated data. It learns to classify real data as 'real'
- *Generator Step*: takes random noise as input and generates fake data. The generated data is fed to the discriminator,...
- *Feedback Loop*: This back-and-forth process continues for many iterations, gradually improving both the generator and discriminator.
- *Convergence*: The GAN training process is considered successful when the generator produces data that the discriminator cannot reliably distinguish from real data.

The key idea behind GANs is the adversarial training process.

- The core of GPT-3 is a huge artificial neural network of 96 layers and 175 billion parameters, trained on hundreds of gigabytes of text from the Internet.
- When presented with a query (prompt), it responds with a list of most probable next words...
- The Bayesian learning (a derivative of Bayes Rule in statistics) is used for machines that generate the most probable hypothesis given the data.
- Because there is nothing built into GPT to distinguish truth from falsehood, GPT is incapable of verifying whether a response is truthful.
- Another way to put this is that GPT's core neural network acts like an associative memory

It can be used for a variety of natural language processing tasks, such as **text completion**, conversation generation, ..

Generative AI models are particularly well-known for their ability to generate text, images, music, and other forms of media.

One of the breakthroughs in generative AI is development of neural network architectures:

Generative Pre-trained Transformer (GPT),
Some applications of generative AI:

- Text Generation
- Image Generation
- Music Composition
- Video Creation
- Drug Discovery
- Data Augmentation
- Style Transfer
- Game Content Creation

The ChatGPT can converse in an interactive way in many languages, including one of the oldest language Sanskrit. How the Covid-19 has changed Education padagogies?

- As of my last knowledge update in January 2022, the COVID-19 pandemic had already significantly impacted education worldwide, leading to the adoption of various pedagogical changes. Keep in mind that the situation may have evolved further since then. Here are some ways in which COVID-19 has influenced educational pedagogies:
- Shift to Online Learning: With the closure of schools and universities ...
- Blended Learning Models: Many educational institutions began exploring blended learning models,...

- Increased Use of Educational Technology: The pandemic accelerated...
- Adaptive Learning Platforms: ...
- Focus on Student Engagement and Well-being:...
- Professional Development for Educators:...
- Assessment Modifications:...
- Global Collaboration:...
- Equity and Inclusion Considerations:...

Conversation in Sanskrit Language

Me: Now I want to converse in Sanskrit.

ChatGPT: Certainly! I'd be happy to converse with you in Sanskrit. Please feel free to ask any questions or start a conversation in Sanskrit, and I'll do my best to respond.

Me: Bhavatah nam kim?

ChatGPT: Mama nam ChatGPT. Bhavatah nam kim?

Me: Mama nam kr chowdhary.

ChatGPT: Suvagatam, kr chowdhary! Bhavatah kutra vasati?

Me: Aham India setre niwasami.

ChatGPT: Sundaram! Bharatasya kasya bhage niwasasi?

Me: aham jodhpur setre niwasmi.

ChatGPT: Jodhpur, Rajasthan, sundara sthala asti. Kutra sthitah bhavanam vartate?

Me: Bhagat ki Kothi.

ChatGPT: Bhagat Ki Kothi Jodhpur setre sthitam asti. Kshetram teerthayatraanuyayi vartate?

Versatility of ChatGPT

- The most famous of AI tools, also known as large language models, or LLMs, is ChatGPT, a version of GPT-3.5 helps to write presentation slides, student exams and coursework problems, and to convert student theses into papers - digital secretary.
- Scientists say they now regularly use LLMs not only to edit manuscripts, but also to help them write or check code and to brainstorm ideas.
- LLMs form part of search engines, code-writing assistants and even a chatbot that negotiates with other companies' chatbots to get better prices on products ... more investments.
- But LLMs have also triggered widespread concern – from their propensity to return falsehoods,

General Scope of Generative AI

The prospects of generative AI are vast and continually expanding,

- Creative Content Generation
- Entertainment and Media
- Education and Training
- Healthcare and Drug Discovery
- Design and Creativity
- Language Translation and Communication
- Personal Assistants and Chatbots
- Financial Modeling and Forecasting
- Research and Development
- Artificial Creativity and Collaboration
- Environmental Conservation
- Cybersecurity
- Legal and Contract Analysis
- Emotional Intelligence
- Ethics and Bias Mitigation

To ChatGPT: Write the opening paragraph for an article about how transformative generative AI will be for business, in the style of McKinsey & Company:

"The rise of generative AI has the potential to be a major game-changer for businesses. This technology, which allows for the creation of original content by learning from existing data, has the power to revolutionize industries and transform the way companies operate. By enabling the automation of many tasks that were previously done by humans, generative AI has the potential to increase efficiency and productivity, reduce costs, and open up new opportunities for growth. As such, businesses that are able to effectively leverage the technology are likely to gain a significant competitive advantage."

- With generative AI, computers can now arguably exhibit creativity. They can produce original content in response to queries, drawing from data they have ingested and interactions with users. They can develop blogs, sketch package designs, write computer code, or even theorize on the reason for a production error.
- ChatGPT, has also been used to translate text, and scientists used an earlier version of GPT to create novel protein sequences.
- Generative AI has made the **2023 one of the most exciting years** yet for AI.
- But as with every new technology, business leaders **must proceed with eyes wide open**, because the technology today presents many ethical and practical challenges.

In companies considering generative AI, executives will want to quickly identify the parts of their business where the technology could have the most immediate impact.

To think through basic questions, such as these:

- Where might the technology aid or disrupt our industry and/or our business's value chain?
- What are our policies and posture?
- Given the limitations of the models,
- How do we pursue building an effective ecosystem of partners,
- What legal and community standards should these models adhere to
- Which of **water fall** or **spiral** model is good for developing software for banking?

OpenAI, has the potential to be a powerful tool in the field of education (All these are findings of research).

- Personalized Tutoring & feedback.
- Automated Essay Grading (found correlation of 0.86).
- Language Translation.
- Interactive Learning.
- Adaptive Learning.

Limitations and challenges of LLMs: In Education

- Lack of Human Interaction.
- Limited Understanding.
- Bias in Training Data (only as good as the data they are trained on).
- Lack of Creativity.
- Dependency on Data.
- Lack of Contextual Understanding.
- Limited ability to personalize instruction.
- Privacy.

The awe-inspiring results of generative AI might make it seem like a ready-set-go technology, but that's not the case. Its nascency requires executives to proceed with an abundance of caution.

- Like humans, generative AI can be wrong.
- Filters are not yet effective enough to catch inappropriate content.
- Systemic biases still need to be addressed.
- Individual company norms and values are not reflected.
- Intellectual-property questions are up for debate.

Limitations and challenges of LLMs: In General

- Researchers emphasize that LLMs are fundamentally unreliable at answering questions, sometimes generating false responses.
- ChatGPT and its competitors work by learning the statistical patterns of language in enormous databases of online text – including any untruths, biases or outmoded knowledge.
- The result is that LLMs easily produce errors and misleading information, particularly for technical topics that they might have had little data to train on.
- "The tool cannot be trusted to get facts right or produce reliable references," noted a January editorial on ChatGPT in the journal Nature Machine Intelligence.

Safety and responsibility

- There is a familiar safety concern that ethicists have been pointing out for years: without output controls LLMs can easily be used to generate hate speech and spam,
- Because the firms that are creating big LLMs are mostly in, and from, these cultures, they might make little attempt to overcome such biases, which are systemic and hard to rectify.
- OpenAI tried to avoid many of these issues when deciding to openly release ChatGPT. It restricted its knowledge base to 2021, prevented it from browsing the Internet
- Some researchers say that academics should refuse to support large commercial LLMs altogether.
- Confusion on the legal status of LLMs ? Copyright and licensing laws currently cover direct copies of pixels,

Need of Enforcing honest use in Research

- It is suggested that existing laws on discrimination and bias will help to keep the use of LLMs honest, transparent and fair. There's loads of law out there, and it's just a matter of applying it or tweaking it very slightly. The journal Science said that **no text generated by ChatGPT** or any other AI tool can be used in a paper.
- One key technical question is whether AI-generated content can be spotted easily. A computer-science undergraduate at Princeton University in New Jersey, created detection tool: GPTZero.
- For scientists' purposes, a tool that is being developed by the firm Turnitin,

- Given that, researchers agree it would be useful to have a way to distinguish human-written text from that generated by a computer. Several groups have developed detectors to identify synthetic text.
- AI-written text is thought to be detectable because of the way it is created. The LLMs are trained on human-written text and learn statistics about how often particular words appear in proximity to other words.
- The detectors use neural networks and deep learning to identify hidden patterns in sequences of text,

What LLM are good for?

- LLMs have been put to good uses where their trustworthiness does not matter.
- The most prominent is entertainment: Many people have amused themselves experimenting with ChatGPT to see how it answers questions and whether it can be tricked into ignoring its "guardrails."
- Another popular use is jumpstarting a writing project: GPT can provide an initial draft for a speech, a document, or code, much faster than when the author starts from scratch.
- Another good use is discovery: GPT can draw from texts that come from parts of the world that an author is unfamiliar with or has never heard of.

What Are the Dangers?

- LLMs in their current implementations display a dangers including undetectable deepfakes, fake religions, automated blackmail, new forms of phishing and scams, cyberweapons, automatic generation of malware and zero-day attacks, automation of genetic engineering, corruption of law and contracts, demise of artistic professions, automation of political lobbying, rapid increase of job loss to automation, rampant cheating in schools, cheating at peer review processes for scientific publication, loss of trust in society and business, destruction of critical infrastructure, accidental triggering of nuclear or other war, misestimation of military threats, corruption of democratic elections, and emergence of sentient machines that exterminate humans.
- Three main threats: 1. acceleration of automation (automation is too rapid social unrest is likely.) 2. a fear of loss of control to automated entities that can outsmart us., 3. lack of trustworthiness of the responses from LLM machines.

The End of Humanity?

- Many scenarios resulting from the dangers listed here lead to the extinction of the human race by sentient machines that see no value in human beings.
- Computing is marching under the aegis of Moore's Law toward a singularity when machines that have emotions and concerns like ours will no longer exist and machine abilities dwarf our own.
- It is anticipated that new species capable of replacing humans will emerge in as little as 100 years.
- **With or without LLMs, AI will continue advancing.**



Chowdhary, K.R. (2020). Natural Language Processing. In: Fundamentals of Artificial Intelligence. Springer, New Delhi.
https://doi.org/10.1007/978-81-322-3972-7_19