RAMAIAH
Institute of Technology

Generative AI

Course Code: CIAEC59

Textbooks

1. Joseph Babcock, Raghav Bali, Generative Al with Python and TensorFlow 2: Create images text, and music with VAEs, GANs, LSTMs, Transformer models, Packt, 2021. https://books.google.co.in/books/about/Generative_AI_with_Python_and_TensorFlow.html?id=4HIsEAAAQBAJ&redir_esc=y (eBook)

2. Numa Dhamani, Maggie Engler, Introduction to Generative Al, Manning, 2024. https://amzn.in/d/a76XJAQ (eBook)

Unit I

Introduction to Generative AI (GenAI) and its Application Landscape:

What is GenAI? A brief about Generative Models, GANs and LLMs. Modern applications of GenAI: Text generation, image synthesis, and more. Why GenAI applications are transformative: Speed, scale, and capabilities.

What is Generative AI ?

- **Generative AI** is a type of artificial intelligence that creates new content, such as text, images, audio, and even video, by learning patterns from existing data.
- Unlike traditional AI, which primarily focuses on analyzing or categorizing information, generative AI models are designed to produce original outputs that resemble the data they were trained on.

Popular examples of generative AI include:

- **Text generation** (e.g., ChatGPT) models generate human-like text responses.
- **Image generation** (e.g., DALL-E) that can create images from text prompts.
- **Music and voice generation** compose original music or mimic specific voices.

Generative AI uses advanced neural networks, particularly *Generative Adversarial Networks (GANs)*, to learn from vast datasets and generate content that appears authentic. Its applications span numerous fields, including content creation, design, marketing, healthcare, and entertainment.

A brief about Generative Models

- The input variables are often vectors of numbers that aren't related to real-world values at all, and are often even randomly generated. This kind of model-known as a generative model. It can produce complex outputs such as text, music, or images from random noise.
- Even if you didn't know it at the time, you have probably seen other instances of generative models.

Ex: “Deep fakes”, which are videos in which one person's face has been systematically replaced with another's by using a neural network to remap the pixels.

Figure 1.2: A deep fake image[7]

Ex: “Generate fake news” Maybe you have also seen stories about AI models that creates fake news, the firm OpenAI were initially terrified to release to the public due to concerns they could be used to create propaganda and misinformation online.

Figure 1.3: A chatbot dialogue created using GPT-2[9]

- In other applications, such as Google's voice assistant, can make fake dynamically creating a conversation with a human in real time or software that can generate original musical compositions.

- These models are able to handle complex information in a variety of domains: **creating photorealistic images** or **stylistic filters on pictures** (Figure 1.4), **synthetic sound, conversational text** and **even rules for optimally playing video games.**

Figure 1.4: Examples of style transfer using Generative Adversarial Networks (GANs)

Generative models

- Generative models could theoretically be implemented using a wide variety of machine learning algorithms, they are usually built with deep neural networks, which are well suited to capturing complex variations in data such as images or language.
- Focus on implementing these deep generative models for many different applications using TensorFlow 2.0.

- Conceptually, the GAN model creates a competition between two neural networks. One (termed the generator) produces realistic (or, in the case of the experiments by Obvious, artistic) images starting from a set of random numbers and applying a mathematical transformation.
- Examples of generative models include the following:
  - Naive Bayes classifiers
  - Gaussian mixture models
  - Latent Dirichlet Allocation (LDA)
  - Hidden Markov models
  - Deep Boltzmann machines
  - VAES
  - GANS

Generating images

- **A challenge** to generating images, images have no labels (such as a digit) rather, we want to map the space of random numbers into a set of artificial images using a latent vector.
- A further constraint is that we want to promote diversity of these images. If we input numbers within a certain range, we would like to know that they generate different outputs, and be able to tune the resulting image features. For this purpose, VAES were developed to generate diverse and photorealistic images

Figure 1.5: Sample images from a VAE[19]

In the context of image classification tasks, being able to generate new images can help us increase the number of examples in an existing dataset or reduce the bias if our existing dataset is heavily skewed toward a particular kind of photograph. Applications could include generating alternative poses (angles, shades or perspective shots) for product photographs on a fashion e-commerce website

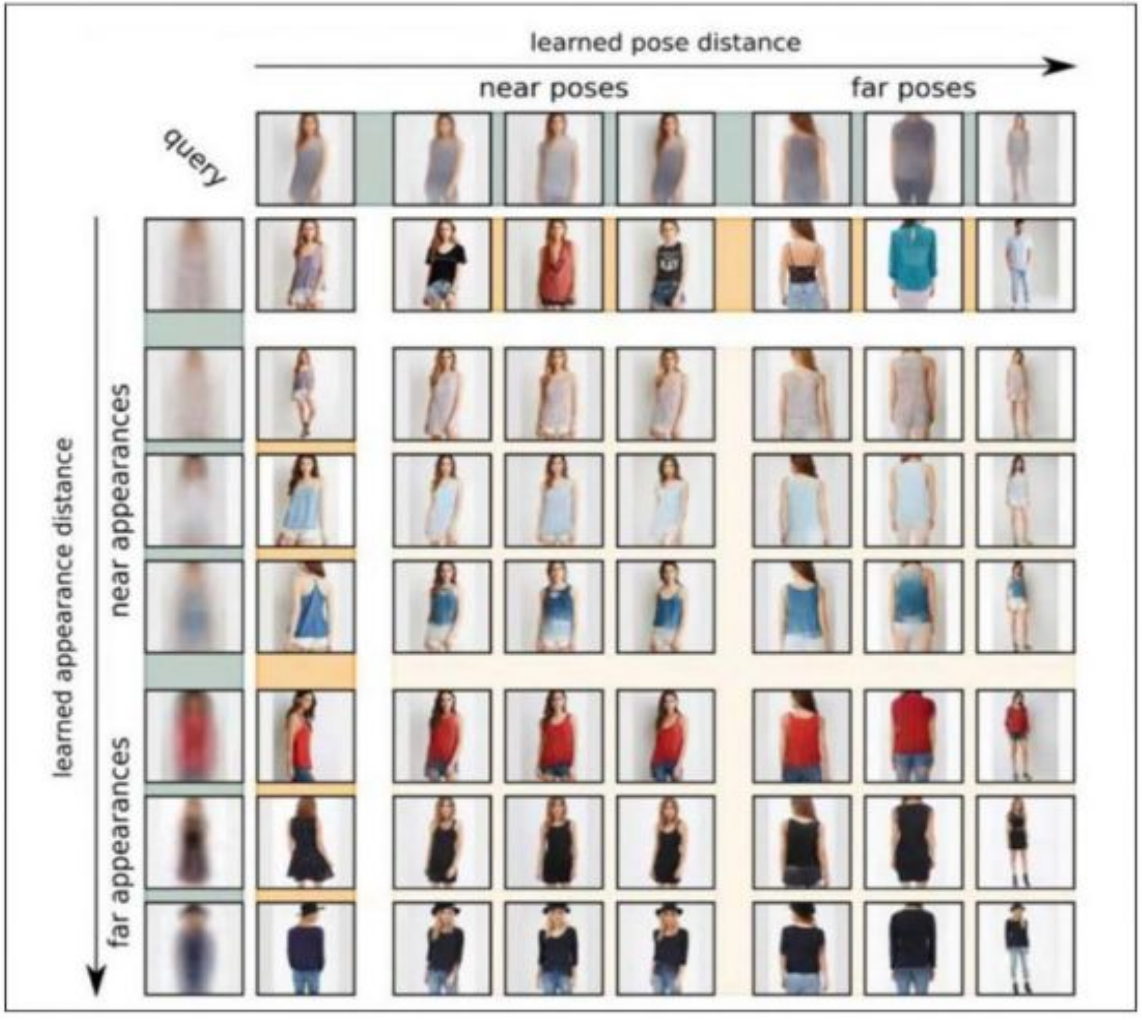


Figure 1.6: Simulating alternative poses with deep generative models[20]

Style transfer and image transformation

- In addition to mapping artificial images to a space of random numbers, we can also use generative models to learn a mapping between one kind of image and a second.
- This kind of model can be used to create deep fake videos in which one actor's face has been replaced with another's, or transform a photo into a painting

- Generative models can be used to convert an image of a horse into that of a zebra

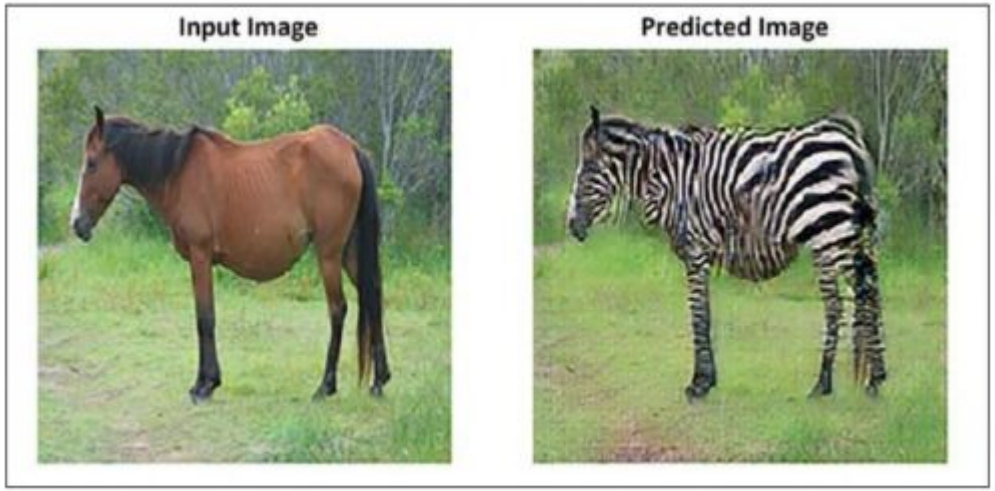


Figure 1.7: CycleGANs apply stripes to horses to generate zebras[22]

Fake news and chatbots

Humans have always wanted to talk to machines; the **first chatbot**, ELIZA was written at MIT in the 1960s.

- **Google's BERT and GPT-2 (Generative Pre-trained Transformer 2),** which use a unit called a transformer. A transformer module in a neural network allows a network to propose a new word in the context of preceding words in a piece of text
- **BERT model** then combines transformer units into a powerful multi-dimensional encoding of natural language patterns and contextual significance. This approach can be used in document creation for natural language processing (NLP) tasks, or for chatbot dialogue systems.

Sound composition

- Sound, like images or text, is a complex, high-dimensional kind of data. Music in particular has many complexities: it could involve one or several musicians, has a temporal structure, and can be divided into thematically related segments.
- All of these components are incorporated into models such as MuseGAN, as mentioned earlier, which uses GANs to generate these various components and synthesize them into realistic, yet synthetic, musical tracks.

Generative adversarial network (GAN)

- A generative adversarial network (GAN) is a deep learning model for task generation introduced in 2014.
- A GAN consists of two parts, a ***generator*** **and a** ***discriminator***. The ***generator*** is a neural network generating output images and the ***discriminator*** a neural network evaluating how realistic the image generated by the *generator*.

- The generator tries to improve its output to mislead the discriminator and the discriminator tries to improve its ability to distinguish real images from generated ones, to avoid being misled by the generator.
- As a result, the generator will maximize its capacity to generate realistic images. GANs are used today for many tasks involving images such as generation and enhancement of photorealistic images.

Large Language Models (LLM)

- Large Language Models (LLMs) are the basis of modern conversational systems (chatbots) such as ChatGPT. These models are trained on large datasets to learn the patterns and structures within the data, enabling them to generate new content that is coherent and contextually relevant.
- LLMs in GenAI focus specifically on generating human-like text by predicting the next statistically most likely word and are used for various natural language processing tasks, including text completion, language translation, summarization and more.
- The training process of LLMs involves pre-training on a large corpus of text data, allowing the model to learn the statistical properties.

- As a result, LLM-based chatbots are able to generate text that is coherent and contextually relevant.
- Once trained, the models can be fine-tuned for specific tasks or used directly to generate diverse and contextually appropriate text. Recently, multimodal large language models (MLLMs) have been gradually taking the lead from traditional LLMs.
- MLLMs are overcoming the limitation of purely text-based input and can access knowledge from multiple modalities – and can thus interact more fully with the real world.

Large Language Models (LLM)

* LLM is a type of ML model designed to understand predict and generate human language **ex :** chatbot

* It's a foundational tool in NLP

* LLM operates on the principle of deep learning leveraging neural network architecture to process and understand human language.

* Trained on vast data set

Key features

- Understanding text
- Text generation
- Context awareness
- Multilingual capabilities
- Task generalization
- Scalability
- Personalization
- Zero shot , one shot learning
- Knowledge retention
- Multi model integration
- Fine tuning and customization
- Ethical constraints

Types of model

1. Statistical language model (n-gram model)

N-gram models use probabilities of word sequences to predict the next word.

An **n-gram** is a sequence of nnn words:

Ex:

- 1-gram: "I"
- 2-gram: "I am"
- 3-gram: "I am happy."

Advantages:

- Simple and computationally efficient.
- Easy to implement and interpret.

2. Neural Language Models (Deep Learning-Based Models)

1. Recurrent Neural Network (RNN)

- **Definition**: RNNs are neural networks with loops, allowing information persistence. They model sequential data by using hidden states that capture context from previous words in the sequence.

Advantages:

- Can model sequences of arbitrary length.
- Better at capturing dependencies compared to n-gram models.

2. Long Short-Term Memory (LSTM)

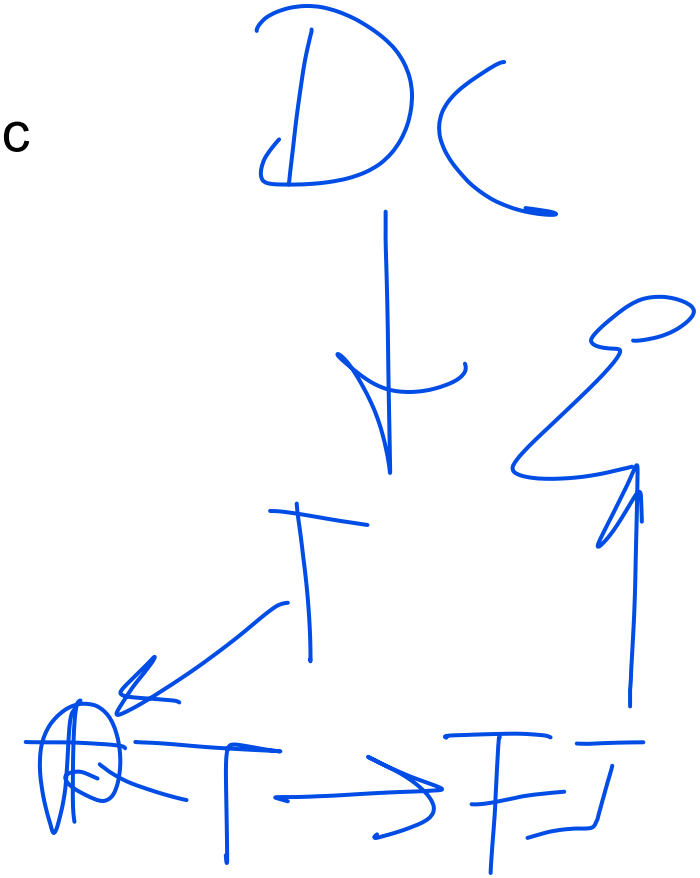
- **Definition**: LSTM is a specialized type of RNN designed to handle long-term dependencies by using gates (input, forget, output) to control the flow of information. LSTM cells selectively remember or forget information.
- **Advantages**:
  - Addresses vanishing gradient problems of RNNs.
  - Effective for modeling long-term dependencies in sequences.

3. Transformers

- **Definition**: Transformers are neural models that use **self-attention mechanisms** to capture dependencies between words, regardless of their distance in the sequence.
- **Advantages**:
  - Excellent at modeling long-range dependencies.
  - Highly parallelizable, making them efficient for training on large datasets.

Training LLM

1. Data collection - from website, academic
2. Tokenization
3. Model architecture (transformer)
4. Pre - training
5. Fine - tuning (supervised learning)
6. Evaluation and testing
7. Scaling and resource requirements
8. Challenges



GANs AND LLMs

GAN	LLM
• Generative Adversarial Networks	• Large Language Models
• GANs are a type of generative model where two neural networks, the generator and the discriminator, compete in a "game."	• LLMs are generative models trained on extensive text datasets to understand and produce human-like language.
• The generator creates fake data (e.g., images),	• capable of generating coherent text, answering questions, translating languages
• Ex: generating high-quality images and other visual content.	• Ex: OpenAI's GPT series,which have applications in chatbots, content generation

Modern applications of GenAI: Image synthesis

1. **Art and design:** New works of art and design, such as paintings, sculptures, and even furniture, produced using generative AI models.

   Ex: artists can create new patterns, textures for their artwork using GANs.

2. **Gaming:** Realistic gaming assets, such as people, locations, or items, can be created using GANs.

3. **Fashion:** Custom clothing, accessory or shoe designs can be created with generative AI models for image synthesis. For apparel designers and retailers, this may open up fresh creative opportunities.

4. **Animation and film:** GANs may be used to create animation, visual effects, or even whole scenes for movies and cartoons. By doing this, developing high-quality visual material may be done faster and cheaper.
5. **Medical diagnosis :**X-rays, MRIs, and CT scans are just a few examples of the kinds of medical pictures that may be produced with GANs.
6. **Photography** to create high-quality photos from low-resolution ones. This can improve the quality of pictures shot using cheap cameras or mobile devices.

Modern applications of GenAI: Text Generation

1. Content Creation and Copywriting

- **Blog Posts & Articles**: AI can draft articles or blog posts on various topics
- **Social Media Posts**: AI models can generate engaging and tailored social media posts

2. Customer Support and Chatbots

- **Automated Customer Service**: Generative AI powers conversational agents and chatbots, providing quick responses to frequently asked questions, handling complaints

3. Education and Learning Tools

- **Tutoring**: AI can create educational content tailored to a student's offering exercises

4. Healthcare Applications

- **Medical Report Summarization**: Generative AI can summarize complex medical reports and patient history

5. Programming Assistance

- **Code Generation**: Tools like GitHub Copilot and ChatGPT Code Interpreter help developers write and debug code, reducing time on repetitive coding tasks.

6. Entertainment and Storytelling

- **Script and Story Generation**: Writers and entertainment professionals use AI to generate story ideas, character backgrounds, and plot twists, which can help spark creativity or generate quick drafts.
- **Gaming**: In games, AI can create dialogue for non-player characters (NPCs), adapting the story based on the player's actions to create dynamic, immersive experiences.

Why GenAI applications are transformative: Speed, scale, and capabilities.

1. Speed

- **Rapid Content Generation**: GenAI can produce text, code, or visuals within seconds, enabling organizations to respond faster to market demands and customer needs.
- **Accelerated Decision-Making**: AI can quickly analyze large datasets, summarize findings, and suggest actionable insights.

2. Scale

- **High Volume Output**: GenAI allows businesses to produce a high volume of content, such as product descriptions, marketing copy, or chat responses

- **Customized Mass Communication**: Generative AI can tailor communication to millions of users creating personalized emails, advertisements for each user at scale.

3. Expanded Capabilities

- **Complex Task Automation**: GenAI can perform tasks that traditionally required skilled human input, such as research papers, or diagnosing medical images.
- **Creative and Analytical Abilities**: AI-generated ideas, stories, or artistic visuals provide novel creativity in fields like entertainment and marketing.
- **Learning and Adaptation**: GenAI models continuously improve with user interactions, meaning that their recommendations, responses

Thank You