Synthesis Lectures on Computer Science

SYNTHESIS
COLLECTION OF TECHNOLOGY

Deepshikha Bhati · Fnu Neha ·
Angela Guercio · Md Amiruzzaman ·
Aloysius Bathi Kasturiarachi

A Beginner's Guide to Generative AI

An Introductory Path to Diffusion Models, ChatGPT, and LLMs

Springer

Synthesis Lectures on Computer Science

The series publishes short books on general computer science topics that will appeal to advanced students, researchers, and practitioners in a variety of areas within computer science.

Deepshikha Bhati · Fnu Neha ·
Angela Guercio · Md Amiruzzaman ·
Aloysius Bathi Kasturiarachi

A Beginner's Guide to Generative AI

An Introductory Path to Diffusion Models, ChatGPT, and LLMs

Springer

Deepshikha Bhati
Department of Computer Science
Kent State University
North Canton, OH, USA

Angela Guercio
Department of Computer Science
Kent State University
North Canton, OH, USA

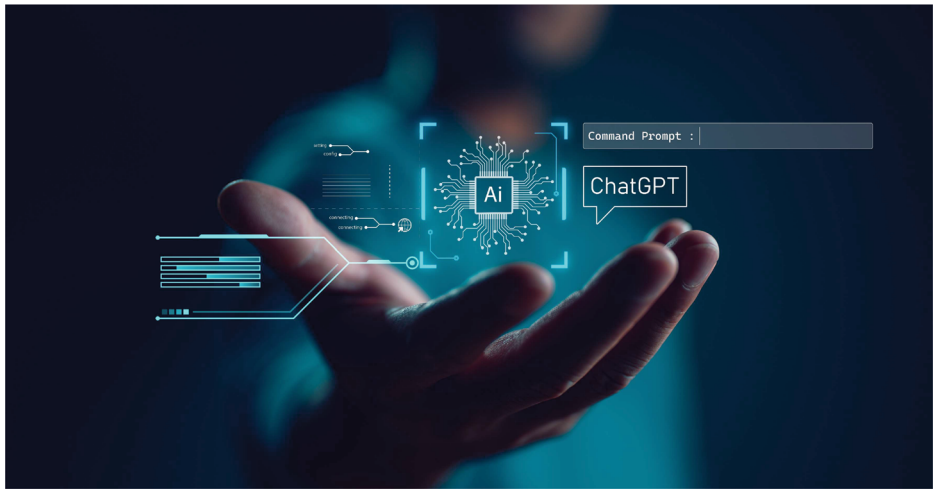
Aloysius Bathi Kasturiarachi
Department of Mathematical Sciences
Kent State University
North Canton, OH, USA

Fnu Neha
Department of Computer Science
Kent State University
Kent, OH, USA

Md Amiruzzaman
Department of Computer Science
West Chester University
West Chester, PA, USA

ISSN 1932-1228 ISSN 1932-1686 (electronic)
Synthesis Lectures on Computer Science
ISBN 978-3-031-84723-3 ISBN 978-3-031-84724-0 (eBook)
https://doi.org/10.1007/978-3-031-84724-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland


Command Prompt :
ChatGPT
Ai

Preface

In this fast-growing environment of creating artificial intelligence (AI)-based applications, generative AI has emerged as one of the groundbreaking technologies of the century. Self-driving cars, healthcare, and other industries are realizing the potential of generative AI through its application possibilities, presenting a disruptive approach that transforms the creations of text, images, and even conversations between people. This book, *A Beginner's Guide to Generative AI: An Introductory Path to Diffusion Models, ChatGPT, and LLMs*, aims to assist you in exploring this fascinating domain while equipping you with the necessary background information and recommendations.

To understand what is happening in this fresh and innovative area for our future development, we must embark on a long and profound journey, beginning by constructing a clear and strong pictures of the fundamental approaches and concepts that define generative AI. Here, you will explore the architectures of some of the most commonly used models, such as Transformers, ChatGPT, and Google Bard, and learn how these technologies have achieved such impressive results. Additionally, you will discover how generative AI can be utilized in writing, medicine, business, and law through concrete examples.

This book is intended for novice readers, and the basics of working with generative AI do not require any prior training. From this perspective, we will proceed from the most basic concepts to more advanced topics in a sequential and accessible manner. Whether you are a student, a working professional, or simply someone curious about what AI can offers you will find this guide highly valuable.

We adopt a broad outlook, presenting a wide array of techniques and models across the spectrum. Rather than focusing on a single method, you will gain insights into generative modeling techniques such as diffusion models, variational autoencoders, and transformers. Understanding all approaches is essential because the field of generative AI is continuously evolving.

As you progress through this book, you will realize that generative AI is not just a series of algorithms; it is an opportunity to create, solve, and innovate. While our primary goal is to enhance your knowledge and awareness about AI, we also hope to inspire your imagination about what AI could become in the future.

Greetings, everyone. You are now in the world of generative AI. Adventure awaits!

North Canton, USA — Deepshikha Bhati
Kent, USA — Fnu Neha
North Canton, USA — Angela Guercio
West Chester, USA — Md Amiruzzaman
North Canton, USA — Aloysius Bathi Kasturiarachi

Competing Interests The authors have no competing interests to declare that are relevant to the content of this manuscript.

Contents

Introduction to Generative AI

1

By the end of this chapter, you will:

- **Define Generative AI**: Describe the subject of it and its purpose in generating fresh content across the subject disciplines.
- **Outline Key AI Components**: Explain the difference between Artificial Intelligence, Machine Intelligence, Machine Learning, Deep Learning, and Generative AI.
- **Explore Domains**: Explore and describe the main categories of Generative AI that include text, image, audio, and video generation while describing what they do, how they work, and how they can be applied.
- **Highlight Industry Leaders**: Present the main actors in the Generative AI concept area and their achievements.
- **Discuss Applications**: Explain the concept of Generative AI and its examples of usage, demonstrating how it could change the world with its impact on content generation and design, entertainment, and healthcare industries.
- **Address Ethical Considerations**: Briefly discuss the ethical perspective on Generative AI and stress the proper usage of the technology and future development to prevent exploitatory use.

1.1 Define Generative AI

The generative artificial intelligence (GAI) relates to the artificial intelligence (AI) approaches that is aimed at producing content that resembles data [1]. Here, a brief distinction should be made between the generative AI models, which are not used for classification or for making predictions but for creating new content: text, images, sounds, videos, etc., based on the patterns learned during the training phase. These models rely on advanced

D. Bhati et al., *A Beginner's Guide to Generative AI*, Synthesis Lectures on Computer Science, https://doi.org/10.1007/978-3-031-84724-0_1

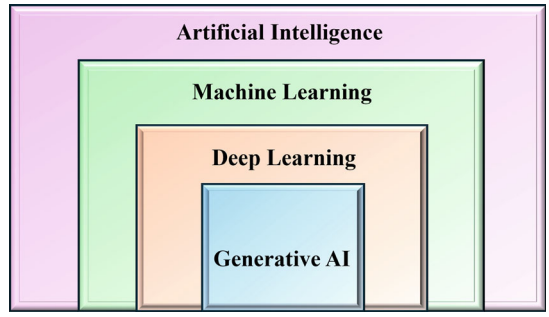


Fig. 1.1 The AI family tree: from AI to generative AI (created by the authors)

technologies like deep learning, probabilistic approaches, and other techniques to generate content that makes semantic sense and contextual sense [2].

It is truly fascinating to invent something in one's mind and visualize it as actualized on a screen or through code. It is quite intriguing to think about this notion, but this concept has become possible thanks to the over development of AI. The possibility of instantly obtaining an image or a text and even the code based on an idea or description is an incredible advancement that turns science fiction into reality and common-use technology [3].

To think that AI was even possible for writing or coding, for many of us in data science and technology, felt unimaginable. On the other hand, the growth of AI has given these tasks a relatively easy and effective way to solve them. Just imagine how time-consuming it was to code and search for solutions online, and now those technological innovations have made it quite easier.

Advancements in deep learning and natural language processing (NLP) have paved the way for generative AI—a subfield of artificial intelligence focused on creating new content by identifying and replicating patterns from data. As a result, we have entered an era where art and technology seamlessly merge, turning once outlandish dreams into reality (Fig. 1.1).

1.2 The Essence of Generative AI

Generative AI refers to artificial intelligence that seeks to create new and unique content from the training data in text, image, audio, and video. While other types of AI models are built for a definite purpose, generative AI models are programmed in such a way that they can develop patterns in data and create outputs that are models of genuine examples.

Generative AI has a diverse range of applications across different domains:

- **Text Generation**: AI models can create coherent and contextually appropriate text by composing stories and articles and generating code.
- **Image Generation**: AI models can produce realistic images or modify existing ones, useful in various fields, from art to product design.

- **Audio Generation**: These models can create music, sound effects, and even realistic speech, enhancing multimedia experiences.
- **Video Generation**: AI can generate and complete video sequences, creating dynamic content for entertainment and media.

Overall, generative AI is promising but has some drawbacks, mainly concerning its ethical application. Immature developments in artificial intelligence, such as deepfakes, apply to producing fake news and violating privacy. Everyone is searching for ways to tackle these problems to ensure that generative AI is employed responsibly.

1.3 Key Components of AI

Artificial Intelligence (AI) is a vast field that consists of several methods and goals that attempt to develop systems that can accomplish jobs that involve human intelligence. This field is subdivided into several fields, including but not limited to Machine Learning, Deep Learning, and Generative AI. Understanding how these concepts are connected would be worthwhile in gaining a better perspective of AI's breadth and diverse uses.

What is Artificial Intelligence (AI)?

AI is about creating machines or systems that can perform activities normally performed by human intelligence, such as solving problems, making decisions, and comprehending languages [4, 5]. AI can be classified into embedded AI and artificial general intelligence.

Machine Learning (ML): A Subfield of AI

Machine learning is one of the branches of AI; it is the process of developing algorithms and statistical models with the help of which computers are trained to use the data and reach conclusions [6]. ML is grouped into three classes:

- Supervised Learning: In this type of learning, the algorithm's inputs and outputs are identified. This information is used to estimate a map so that given inputs and new outputs can be predicted based on data that was not used in the training process.
- Unsupervised Learning: This approach involves feeding the algorithms with data that is not classified in any way to find patterns or structures in the dataset. Typical examples include tasks such as clustering or dimensionality reduction.

- Reinforcement Learning: In this type of learning, an agent is trained to act in an environment where it receives feedback through incentives that may be in the form of rewards or penalties and alters its actions based on the results.

Deep Learning: A Subset of Machine Learning

In deep learning, artificial neural networks with many layers optimize a pre-specified model to extract features automatically from large datasets [7]. It has proven particularly useful in applications such as image identification, natural language understanding, and voice recognition [8].

Key components of deep learning include:

- Neural Networks: Models inspired by the structure of the human brain form the foundation of deep learning [9].
- Convolutional Neural Networks (CNNs): Specialized for image and spatial data, CNNs automatically learn hierarchical feature representations [10].
- Recurrent Neural Networks (RNNs): Designed for sequential data, such as text or time series, RNNs maintain connections that allow information to persist, making them suitable for tasks like language modeling and translation [11].

Generative AI: Expanding the Horizons of Deep Learning

Generative AI produces new material such as images, music, text, or videos that closely resemble human-created content. This area commonly uses deep learning models like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) to create realistic and creative outputs [12, 13].

- Generative Adversarial Networks (GANs): GANs involve two networks, the generator and the discriminator, which are trained cooperatively and competitively. The generator produces synthetic data, while the discriminator distinguishes between real and fake data, resulting in highly realistic content [14, 15].
- Variational Autoencoders (VAEs): VAEs learn a low-dimensional representation of inputs, known as the latent space, and can generate new samples from this space. VAEs are employed in applications such as image generation, text creation, and data compression [13, 16].

The Relationship Between AI, Machine Learning, Deep Learning, and Generative AI

- Artificial intelligence is the overall category aimed at creating smart systems.
- Machine learning is a subset of AI focused on creating algorithms that learn from data to make predictions or decisions.
- Deep learning is a subset of machine learning that uses neural networks with multiple layers to learn hierarchical data representations.
- Generative AI is a specialized area within deep learning that focuses on creating new content, often using techniques like GANs and VAEs.

It can therefore be said that AI is a set of paradigms for developing intelligent systems. While these concepts all focus on data-driven algorithms, machine learning does this through algorithms, deep learning through neural networks, and generative AI takes it further by making AI systems self-generate content.

1.4 Exploring Generative AI Domains

The term "generative AI" spans various domains, each using different approaches to deliver unique outputs. Here's a deeper dive into these domains:

1.4.1 Text Generation

- **Overview**: Text synthesis generates human-like text based on input prompts or contextual data. This is crucial in various AI-driven applications today.
- **How It Operates**: Models like GPT-4 incorporate the Transformer architecture, pre-trained on large volumes of text data, allowing them to understand language nuances. These models generate continuous and contextually relevant text from given inputs.
- **Applications**:
  - **Content Creation**: Generating via AI blog posts, articles, and marketing copies.
  - **Customer Service**: Developing chatbots and virtual assistants that mimic user behaviors.
  - **Educational Tools**: Creating personalized lessons and tutoring materials.

Most Relevant Related work in Text Generation:

- *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* by Jacob Devlin et al. (2018) [17]: This paper presents BERT, which revolutionized text generation and comprehension. BERT's pre-training and fine-tuning process supports tasks like story writing, code generation, and more.
- *GPT-3: Language Models are Few-Shot Learners* by Tom B. Brown et al. (2020) [18]: GPT-3 is a large language model known for generating fluent and semantically relevant text in a variety of real-world domains, from writing stories to generating code.

1.4.2 Image Generation

- **Overview**: Image synthesis generates new images based on textual descriptions or transforms existing images.
- **How It Operates**: Generative Adversarial Networks (GANs) are commonly used in image synthesis. These networks involve a generator that produces images and a discriminator that evaluates them. The generator improves through iterative training.
- **Applications**:
  - **Marketing and Advertising**: Conceptualizing innovative images for campaigns and product designs.
  - **Entertainment**: Designing characters and scenes for games and media.
  - **Healthcare**: Enhancing medical imaging and disease analysis methods.

Relevant Related Work in Image Generation:

- *High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs* by Ting-Chun Wang et al. [19]: This paper introduces Pix2PixHD, which uses Conditional GANs to produce photorealistic images from semantic layouts, with applications in art and product design.
- *StyleGAN: A Style-Based Generator Architecture for Generative Adversarial Networks* by Tero Karras et al. [20]: StyleGAN advances image generation with fine control over style, widely regarded as a gold standard for image synthesis and manipulation.

1.4.3 Audio Generation

- **Overview**: Audio synthesis generates new sounds or speech from textual or other input data, applicable in various industries.

- **How It Operates**: Models like WaveGAN generate audio waveforms, while text-to-speech (TTS) model, like Tacotron 2 transforms text into natural-sounding speech.
- **Applications**:
  - **Music Industry**: Creating custom music accompaniments and sound clips.
  - **Voice Assistants**: Developing realistic voices for virtual assistants.
  - **Media Production**: Crafting soundscapes and audio content for films and games.

Relevant Related Work in Audio Generation:

- *WaveNet: A Generative Model for Raw Audio* by Aaron van den Oord et al. [21]: WaveNet generates raw audio waveforms, offering breakthroughs in AI-based audio synthesis.
- *Jukebox: A Generative Model for Music* by Prafulla Dhariwal et al. [22]: Jukebox synthesizes music with vocals, using a hierarchical VQ-VAE-2 structure to produce high-quality music compositions.

1.4.4 Video Generation

- **Overview**: Video generation creates or completes video sequences using input data such as text or existing footage.
- **How It Operates**: Video generation models predict and generate frames sequentially, maintaining temporal coherence and visual realism.
- **Applications**:
  - **Entertainment**: Producing trailers, animations, and special effects.
  - **Training and Simulation**: Creating realistic training videos for various sectors.
  - **Content Creation**: Producing social media and marketing videos.

Relevant Related Work in Video Generation:

- *Text2Video-Zero: Text-to-Image Diffusion Models are Zero-Shot Video Generators* by Levon Khachatryan et al. [23]: Recent advancements in text-to-video generation have introduced a zero-shot approach that leverages text-to-image models like Stable Diffusion to generate videos without extensive training on large-scale datasets. By incorporating motion dynamics and cross-frame attention mechanisms, this method achieves consistent and high-quality video generation with minimal computational overhead, expanding its applicability to tasks like video editing and conditional generation.