■ My Study Notes

Exported on December 13, 2025 at 03:03 AM

---

1. Quick Summary

Created: 2025-12-12T21:33:09.431203

Structured Study Summary

Mod 2: Generative AI Models and Applications

Introduction

Generative AI models have revolutionized the field of artificial intelligence by enabling machines to generate human-like text, images, and speech. This summary will cover the key concepts of zero-shot, few-shot, and chain-of-thought prompting, as well as the structure and benefits of major Generative AI models, including LLMs, GANs, and Diffusion Models. Additionally, we will examine various Generative AI applications and their underlying foundational models.

Key Points

* **Zero-shot prompting**: The model is given only the instruction and must complete the task using its pre-trained knowledge.

* **Few-shot prompting**: The user provides a few examples showing how the task should be done.

* **Chain-of-thought prompting**: Encourages the model to show step-by-step reasoning before giving the final answer.

* **LLMs (Large Language Models)**: Based on Transformer architecture, uses self-attention layers to understand relationships between words.

* **GANs (Generative Adversarial Networks)**: Consists of two neural networks trained together: Generator and Discriminator.

* **Diffusion Models**: Based on denoising diffusion process, training: add noise to images in many steps, generation: start from noise → gradually remove noise → final image.

* **RAG (Retrieval-Augmented Generation) Systems**: Convert user query into embeddings, retrieve relevant documents from a vector database, and LLM reads the retrieved documents and generates an accurate answer.

* **Image Captioning Systems**: Image → vision encoder extracts features, text decoder converts visual features to words.

* **Speech Systems (STT – Speech-to-Text)**: Convert audio waveform into spectrogram, speech encoder recognizes phonemes and words, decoder generates text output.

* **Speech Systems (TTS – Text-to-Speech)**: Convert text → phonemes → acoustic representation, neural vocoder generates human-like audio.

Definitions

* **Self-attention layers**: A type of neural network layer that allows the model to understand relationships between words.

* **Transformer architecture**: A type of neural network architecture that uses self-attention layers to understand relationships between words.

* **Generator**: A neural network that creates fake images/data from random noise.

* **Discriminator**: A neural network that checks if data is real or fake.

* **Denoising diffusion process**: A process that adds noise to images in many steps, and then gradually removes the noise to generate the final image.

Examples

* **Zero-shot prompting**: "Translate 'Good morning' into French."

* **Few-shot prompting**: "Food: Pizza → Category: Fast food", "Food: Dosa → Category: South Indian", "Food: Sushi → Category: ?"

* **Chain-of-thought prompting**: "A shop had 30 apples, then bought 20 more. How many now? Explain your steps."

* **LLMs**: Trained on large text datasets for language understanding and generation.

* **GANs**: Produces sharp, high-quality images, but training is unstable and difficult to tune.

* **Diffusion Models**: Extremely high-quality and detailed images, but slow generation because of many denoising steps.

* **RAG Systems**: Convert user query into embeddings, retrieve relevant documents from a vector database, and LLM reads the retrieved documents and generates an accurate answer.

* **Image Captioning Systems**: Image $\rightarrow$ vision encoder extracts features, text decoder converts visual features to words.

* **Speech Systems (STT – Speech-to-Text)**: Convert audio waveform into spectrogram, speech encoder recognizes phonemes and words, decoder generates text output.

* **Speech Systems (TTS – Text-to-Speech)**: Convert text $\rightarrow$ phonemes $\rightarrow$ acoustic representation, neural vocoder generates human-like audio.

Equations/Formulas

* **Self-attention layer**: q = Wq * x, k = Wk * x, v = Wv * x, attention = softmax(q * k^T / sqrt(d))

* **Transformer architecture**: H = softmax(q * k^T / sqrt(d)) * v

* **Generator**: G(z) = σ(W * z + b)

* **Discriminator**: D(x) = σ(W * x + b)

* **Denoising diffusion process**: x_t = x_{t-1} + σ * ε, where ε is noise, and σ is a scaling factor.

Short Quiz

1. What is the main difference between zero-shot and few-shot prompting?

a) Zero-shot uses pre-trained knowledge, while few-shot uses examples.

b) Zero-shot uses examples, while few-shot uses pre-trained knowledge.

c) Zero-shot uses a single example, while few-shot uses multiple examples.

d) Zero-shot uses a single instruction, while few-shot uses multiple instructions.

Answer: a) Zero-shot uses pre-trained knowledge, while few-shot uses examples.

2. What is the main benefit of using GANs?

a) Produces sharp, high-quality images.

b) Fast generation during inference.

c) Effective for style transfer and image-to-image tasks.

d) All of the above.

Answer: d) All of the above.

3. What is the main drawback of using Diffusion Models?

a) Slow generation because of many denoising steps.

b) Requires strong GPUs for fast inference.

c) Harder to use for real-time tasks.

d) All of the above.

Answer: d) All of the above.

---

2. Quick Summary

Created: 2025-12-12T20:09:03.444339

Study Summary: Generative AI Models and Applications

Introduction

Generative AI models have revolutionized the way we interact with technology, enabling applications such as chatbots, image generation, and speech synthesis. This summary will cover the key concepts of zero-shot, few-shot, and chain-of-thought prompting, as well as the structure and benefits of major Generative AI models, including LLMs, GANs, and Diffusion Models. Additionally, we will examine various Generative AI applications and their underlying foundational models.

Key Points

* **Zero-shot Prompting**:

+ The model is given only the instruction and must complete the task using its pre-trained knowledge.

+ No examples or demonstrations are provided in the prompt.

+ Works best for common tasks like translation, classification, summarization, or factual Q&A.;

* **Few-shot Prompting**:

+ The user provides a few examples showing how the task should be done.

+ Helps the model learn formatting, style, specialized logic, or domain-specific patterns.

+ Useful for niche tasks where zero-shot may confuse the model.

* **Chain-of-thought Prompting**:

+ Encourages the model to show step-by-step reasoning before giving the final answer.

+ Very effective for math, logic, planning, multiprocess, or multi-step problems.

+ Reduces errors by forcing the model to reason clearly.

* **LLMs (Large Language Models)**:

+ Based on Transformer architecture.

+ Uses self-attention layers to understand relationships between words.

+ Trained on large text datasets for language understanding and generation.

* **GANs (Generative Adversarial Networks)**:

+ Consists of two neural networks trained together: Generator and Discriminator.

+ The two networks compete, improving each other.

* **Diffusion Models**:

+ Based on denoising diffusion process.

+ Training: add noise to images in many steps.

+ Generation: start from noise → gradually remove noise → final image.

Definitions

* **Zero-shot Prompting**: The model is given only the instruction and must complete the task using its pre-trained knowledge.

* **Few-shot Prompting**: The user provides a few examples showing how the task should be done.

* **Chain-of-thought Prompting**: Encourages the model to show step-by-step reasoning before giving the final answer.

* **LLMs (Large Language Models)**: Based on Transformer architecture, trained on large text datasets for language understanding and generation.

* **GANs (Generative Adversarial Networks)**: Consists of two neural networks trained together: Generator and Discriminator.

* **Diffusion Models**: Based on denoising diffusion process, used for image generation.

Examples

* **Zero-shot Prompting**: Translate "Good morning" into French.

* **Few-shot Prompting**: Food: Pizza → Category: Fast food, Food: Dosa → Category: South Indian, Task: Food: Sushi → Category: ?

* **Chain-of-thought Prompting**: A shop had 30 apples, then bought 20 more. How many now? Explain your steps.

* **LLMs (Large Language Models)**: Trained on large text datasets for language understanding and generation.

* **GANs (Generative Adversarial Networks)**: Used for deepfakes, art generation, super-resolution, face synthesis, style transfer.

* **Diffusion Models**: Used for image generation, video synthesis, design tools, character creation, animation.

Equations/Formulas

* **Self-Attention Mechanism**: q = Wq * x, k = Wk * x, v = Wv * x, attention = softmax(q * k^T / sqrt(d))

* **Denoising Diffusion Process**: x_t = x_{t-1} + \epsilon_t, where \epsilon_t is noise added to the image at step t.

Short Quiz

1. What is the primary difference between zero-shot and few-shot prompting?

a) Zero-shot uses pre-trained knowledge, while few-shot uses examples.

b) Zero-shot uses examples, while few-shot uses pre-trained knowledge.

c) Zero-shot is used for common tasks, while few-shot is used for niche tasks.

d) Zero-shot is used for math problems, while few-shot is used for language tasks.

Answer: a) Zero-shot uses pre-trained knowledge, while few-shot uses examples.

2. What is the main benefit of chain-of-thought prompting?

a) Reduces errors by forcing the model to reason clearly.

b) Increases the speed of generation.

c) Improves the quality of the output.

d) Reduces the need for human intervention.

Answer: a) Reduces errors by forcing the model to reason clearly.

3. What is the primary use case for Diffusion Models?

a) Language understanding and generation.

b) Image generation and video synthesis.

c) Speech recognition and synthesis.

d) Text classification and summarization.

Answer: b) Image generation and video synthesis.