NO PREP

How AI Generates Images

From Words to Pictures

Includes: Student reading, case study, differentiated questions, teacher key

Name: Period: Date:

How AI Generates Images: From Words to Pictures

A New Kind of Creative Machine

In 2022, a new type of AI tool went from research curiosity to cultural phenomenon almost overnight. Tools like DALL-E, Midjourney, and Stable Diffusion could take a text description and generate a strikingly realistic or imaginatively detailed image in seconds. Artists were stunned. Designers were excited. And millions of people realized they could create visual content without knowing how to draw a single line. AI image generation is now integrated into products from Adobe, Microsoft, Google, and Canva. It is used to create marketing materials, concept art, advertising images, book illustrations, video game assets, and more.

Name: **Period:** **Date:**

How AI Generates Images: From Words to Pictures

A New Kind of Creative Machine

In 2022, a new type of AI tool went from research curiosity to cultural phenomenon almost overnight. Tools like DALL-E, Midjourney, and Stable Diffusion could take a text description and generate a strikingly realistic or imaginatively detailed image in seconds. Artists were stunned. Designers were excited. And millions of people realized they could create visual content without knowing how to draw a single line. AI image generation is now integrated into products from Adobe, Microsoft, Google, and Canva. It is used to create marketing materials, concept art, advertising images, book illustrations, video game assets, and more. Understanding how these systems work is essential for anyone navigating a world where an image cannot be assumed to be real, and where visual creativity is being reshaped by machines.

The Training Process: Learning From Billions of Images

AI image generators are trained on enormous datasets of images paired with text descriptions. The largest datasets, like LAION-5B, contain 5 billion image-text pairs scraped from across the internet. The model learns associations between visual content and language by processing these pairs: an image of a golden retriever beside a lake paired with the caption 'golden retriever dog by lake at sunset' teaches the model to associate those visual elements with those words. During training, the model does not memorize the images. It learns the statistical relationships between visual features and textual concepts. After seeing millions of examples of what 'sunset' looks like in images, it builds a rich internal representation of that concept. After seeing millions of examples of 'dog,' 'fluffy,' 'outdoors,' and 'reflection,' it learns how those visual elements interact. When you later ask it to generate 'a fluffy dog sitting by a lake at sunset,' it draws on all of those learned associations to produce a new image.

Diffusion Models: Noise and Refinement

The dominant architecture behind modern image generators is called a diffusion model. Diffusion models learn to generate images through a two-phase process. During training, they learn to add noise to real images progressively until the image becomes pure random noise. Then they learn to reverse that process: starting from random noise, they learn to progressively remove noise until a coherent image emerges. When you give a text prompt to a diffusion model, it starts with a field of random noise. It uses the text prompt as a guide to steer the denoising process toward images that match your description. Step by step, the noise is removed and replaced with visual structure that reflects the prompt. After many denoising steps, a complete image appears. This is why AI image generation takes a few seconds: the model is running dozens or hundreds of processing steps to refine the noise into a coherent image.

The Role of Text Encoders

The bridge between your text prompt and the image generation process is a component called a text encoder. Text encoders convert your written prompt into a numerical representation that the image model can use to guide the generation process. One of the most influential text encoders is called CLIP, developed by OpenAI. CLIP was trained to align images and text in a shared numerical space, so that the representation of a dog in an image and the word 'dog' in text end up mathematically close together. This alignment is what makes text-to-image generation work. When you type a prompt, the text encoder converts it into a numerical vector. The diffusion model uses that vector to steer the denoising process toward images that are mathematically similar to the concept described in the text. The quality of the text encoder largely determines how well the model follows complex, nuanced prompts. Weak text encoding is why early AI image generators struggled with specific arrangements, precise text in images, or complex multi-object compositions.

Prompt Engineering for Images

The quality of an AI-generated image is heavily influenced by how the prompt is written. This is called prompt engineering for image generation. Short, vague prompts like 'dog' produce generic results. Detailed prompts that specify style ('oil

painting,' 'photorealistic,' 'anime'), lighting ('golden hour,' 'dramatic studio lighting'), composition ('wide angle,' 'portrait,' 'birds-eye view'), and context ('a golden retriever puppy running through autumn leaves in Central Park, golden hour, shot on a 50mm lens, photorealistic') produce dramatically more specific and impressive results. Platform-specific tools like Midjourney have their own style language and prompt modifiers. Professional users spend considerable time learning what combinations of terms produce the best results for their specific use cases. Some people sell prompt packs, collections of tested, effective prompt formulas, as a digital product. The skill of communicating visually through language has become a genuinely valuable ability.

The Copyright and Ethics Controversy

AI image generation has triggered fierce debates about copyright and ethics. The training datasets for these models were assembled by scraping images from the internet, including billions of images by professional artists who never consented to having their work used as training data and receive no compensation when the AI generates images in their style. Artists have sued AI image companies including Stability AI and Midjourney. There are also serious concerns about deepfakes and non-consensual intimate imagery. AI image generators can create photorealistic images of real people, including public figures and private individuals, in situations that never occurred. Several countries have passed or proposed laws specifically criminalizing non-consensual AI-generated intimate imagery. The technology that enables creative image generation is the same technology that enables harmful fabrication, which is why understanding how it works is inseparable from understanding its risks. Generative Adversarial Networks (GANs) were the dominant image generation technology before diffusion models. A GAN consists of two competing neural networks: a generator that creates images and a discriminator that tries to distinguish generated images from real ones. The generator improves by learning to fool the discriminator; the discriminator improves by getting better at detecting fakes. This adversarial competition drives both networks to improve. GANs produced impressive results in the late 2010s, including the influential StyleGAN series that could generate photorealistic human faces. However, GANs were difficult to train stably and often suffered from 'mode collapse,' where the generator would produce only a narrow range of outputs. Diffusion models largely

replaced GANs for high-quality image generation because they were more stable to train and produced greater diversity. The creative ecosystem around AI image generation has developed rapidly into a significant economic sector. Platforms like Midjourney have grown from small research projects to companies with substantial revenue from monthly subscriptions. Prompt engineering for images has become a profession: skilled prompt engineers who understand how to write prompts that reliably produce high-quality commercial results command hourly rates comparable to graphic designers. Stock image agencies like Getty Images and Shutterstock have launched their own AI image generation tools to compete with disruptive startups, while also pursuing lawsuits against AI companies that trained on their image libraries. The entire visual creative industry is restructuring around the new reality of AI image generation, creating both displacement and opportunity.

Key Information: AI Image Generation

Detail	Information
Major tools	DALL-E (OpenAI), Midjourney, Stable Diffusion, Adobe Firefly
Training data example	LAION-5B: 5 billion image-text pairs from the internet
What the model learns	Statistical associations between visual features and textual concepts
Dominant architecture	Diffusion model: learns to remove noise from random starting points
Text encoder role	Converts text prompts into numerical vectors that guide image generation
CLIP	OpenAI's text-image alignment model; foundational to text-to-image systems
Prompt engineering	Writing detailed prompts that specify style, lighting, composition, context
Key copyright issue	Training on artist work without consent; ongoing lawsuits vs. AI companies

Detail	Information
Deepfake risk	AI can generate photorealistic images of real people in fabricated scenarios
Legal development	Multiple countries criminalizing non-consensual AI intimate imagery

Case Study: An Artist's Style Gets Copied Overnight

Reina has been developing her distinctive illustration style for seven years. She posts her work regularly on social media and has built a following of 40,000 people who love her particular combination of soft watercolor textures, architectural detail, and warm color palettes. One day, a follower sent her a message: 'Did you see this? Someone used Midjourney to generate images in your exact style and is selling them on stock art websites.' Reina clicked the link. The images looked almost exactly like her work, with her distinctive visual signature, but they had been generated in seconds using AI prompts like 'in the style of [Reina's name].' Reina was furious. She had never consented to her work being used as training data. She was not compensated when AI systems used her style. Now someone was commercially exploiting that AI output to compete directly with her in the market she had spent years building. 'It learned from my art,' she said in a public post that went viral, 'and now it's replacing me.'

1. Using what you learned about how AI image generators are trained, explain how an AI system was able to produce images in Reina's specific style.

2. What ethical and legal questions does Reina's situation raise about AI image generation and the rights of artists?

Multiple Choice

1. What is a diffusion model and how does it generate images?

    a) A model that combines elements from multiple existing images to create a new one

    b) A model that starts from random noise and progressively removes it to produce a coherent image guided by a text prompt

    c) A model that diffuses color gradually from a base template of geometric shapes and then refines detail in several phases

    d) A model that traces the outline of similar images from a database and recolors them

2. What role does CLIP play in text-to-image AI systems?

    a) It generates the final image from a random seed

    b) It adds invisible watermarks to every AI-generated image so the output can later be identified

    c) It removes copyrighted images from training datasets

    d) It aligns text and images in a shared numerical space so prompts can guide image generation

3. Why have professional artists and illustrators objected to AI image generation systems?

a) AI companies trained on artists' work without consent or compensation and now compete with them commercially

b) AI tools require specialized artistic training and very expensive professional equipment to use commercially

c) AI-generated images always look noticeably fake and cannot replace real art

d) AI systems generate images too slowly to compete in professional workflows

Matching

Write the letter of the correct definition on the line next to each term.

Term	Definition
1. ___ Diffusion model	A) OpenAI's model that aligns text and image representations in a shared numerical space
2. ___ CLIP	B) An image generation architecture that progressively removes noise from a random starting point
3. ___ Prompt engineering	C) A dataset of 5 billion image-text pairs used to train AI image generators
4. ___ LAION-5B	D) A component that converts a text prompt into a numerical vector that guides image generation
5. ___ Text encoder	E) The practice of writing detailed, effective prompts to improve AI image output quality

Fill-in-the-Blank

1. AI image generators start generation from random _____ and progressively remove it using the text prompt as a guide.

2. The practice of writing detailed prompts that specify style, lighting, and composition to improve AI image results is called prompt _____.

3. Multiple countries have passed or proposed laws specifically criminalizing non-consensual AI-generated _____ imagery.

Short Answer

Answer each question in 2-4 complete sentences.

1. Describe how a diffusion model generates an image from a text prompt, step by step.

_____

_____

_____

2. Why does the quality of prompt writing have such a large impact on the quality of AI-generated images?

_____

_____

_____

3. How does AI image generation create both creative opportunities and serious ethical risks at the same time?

_____

_____

_____