Deep Generative Models Introduction

Fall Semester 2025

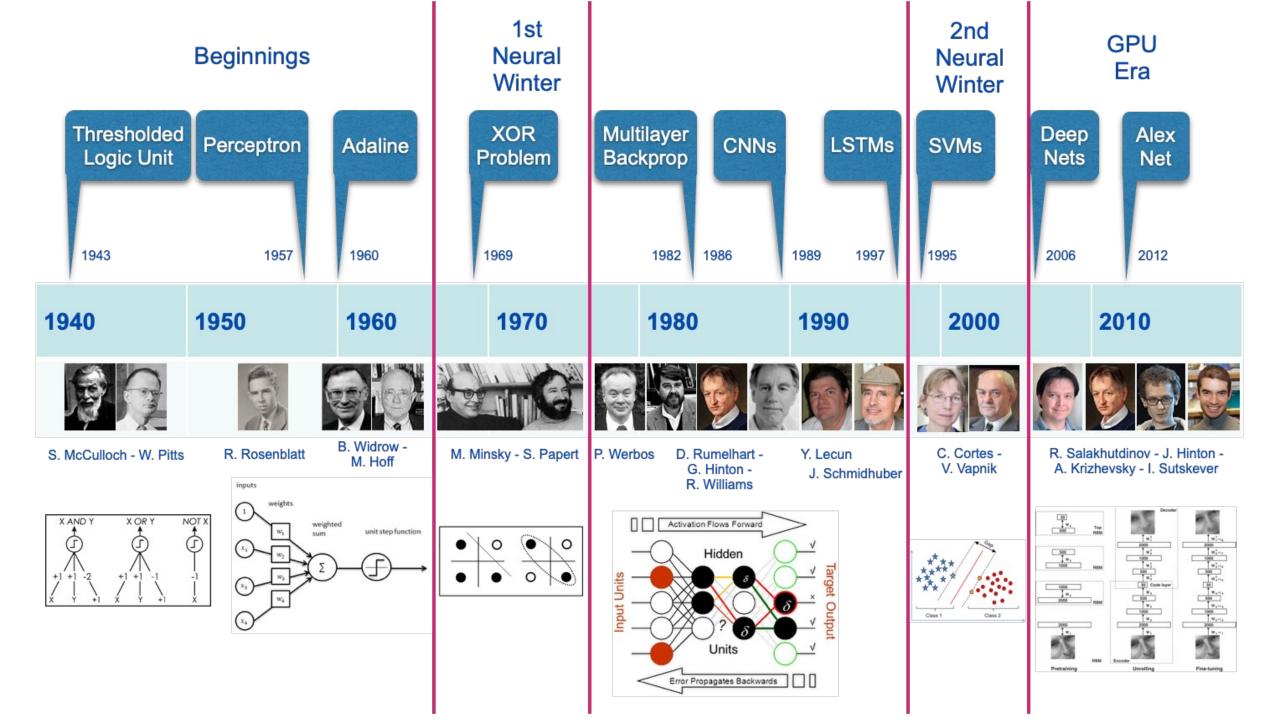
René Vidal

Director of the Center for Innovation in Data Engineering and Science (IDEAS),
Rachleff University Professor, University of Pennsylvania
Amazon Scholar & Chief Scientist at NORCE



Outline

- History of Al
- History of Generative Models
- Advent of Deep Generative Models
- Applications of Deep Generative Models



Generative AI timeline

General model research

Image / video models

Text models

Business events

December 1997

Two computer scientists, Sepp Hochreiter and Jürgen Schmidhuber, invent Long Short-Term Memory (LSTM) networks, which improves memory capacity in neural networks and thus allows for faster and more precise pattern recognition in training data.⁶

December 2013

Diederik P. Kingma and Max Welling publish a paper on VAEs.⁶ VAEs deconstruct input data into code, known as the latent space, and then construct new, representative data out of the latent space.

July 2015

Google's DeepDream model for image generation is launched. ⁷

2016

June 2016

Style Transfer, a deep neural network that can transfer the style of one image to another, is released.⁷

2012

A seminal year in deep learning research that enables significant advances in computer vision, speech recognition, and natural language processing. These accomplishments are best signified by AlexNet, a convolutional neural network (CNN) trained on graphical processing units (GPUs) and developed by researcher Alex Krizhevksy. AlexNet was the first model to break 75% accuracy, identifying images from a 14-million-image manually tagged database known as ImageNet.

January 2014

Google acquires DeepMind, an Al research lab based in the UK.

March 2015

Ian Goodfellow invents GANs.6

June 2014

Researchers Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli introduce diffusion models.⁶ These models apply noise to an image until it is incomprehensible, then gradually reduce the noise in the image and thus learn how to generate actual images from noise.

December 2015

OpenAl is founded as a non-profit, open-source research lab.

June 2017

Google researchers release "Attention Is All You Need" paper on transformer models, a type of neural network architecture that lends attention to context and syntax.8 This allows for immense strides in natural language processing and understanding.

- 5: "2012: A Breakthrough Year for Deep Learning," Medium, Bryan House, July 17, 2019.
- 6: "The Generative Al Revolution Has Begun—How Did We Get Here?," Ars Technica, Haomiao Huang, January 30, 2023.
- 7: "Timeline of Text-To-Image Machine Learning Models," Fabian Mosele, n.d., accessed March 9, 2023.
- 8: "Timeline of Al and Language Models," Life Architect, Alan D. Thompson, n.d., accessed March 9, 2023.

General model research

Image / video models

Text models

Business events

June 2018

OpenAI releases GPT (Generative Pre-Trained Transformer), ushering in a new era of large language models.

December 2018

NVIDIA's StyleGAN is able to generate hyper-realistic faces.

2019

March 2019

OpenAl announces that it has moved from a nonprofit to "capped-profit" business model in order to attract more investors.9

July 2019

Microsoft invests \$1 billion into OpenAI and signs an exclusive partnership with the research startup.

May 2020

GPT-3 is released with 175 billion parameters, 100 times that of GPT-2.

November 2020

AlphaFold wins the CASP competition by a large margin and is recognized by organizers as a solution to the "proteinfolding problem." 10

November 2018

Google releases its Bidirectional Encoder Representations for Transformers (BERT).

December 2018

Google and DeepMind's AlphaFold, a protein structure prediction model, places first in the Critical Assessment of Protein Structure (CASP) competition.

March 2019

February 2019

GPT-2 is released.

Chinese company Baidu releases its answer to Google's BERT: Enhanced Representation Through Knowledge Integration (ERNIE 1.0).

July 2019

Baidu releases ERNIE 2.0, which outperforms Google's BERT.

October 2019

Google applies its BERT model to searches—demonstrating semantic search capabilities.

September 2020

GPT-3 writes a newspaper column in The Guardian.

January 2021

OpenAl launches Dall-E, its text-to-image model, along with the open-source version of the model: Contrastive Language-Image Pre-Training (CLIP). This enables a flood of text-to-image models.

June 2021

GitHub, a subsidiary of Microsoft, announces its code-completion Copilot tool. GitHub Copilot is built on OpenAl's Codex, which is trained on 54 million public GitHub repositories.

April 2022

OpenAI releases Dall-E 2. Shortly thereafter, Boris Dayma releases an open-source but unaffiliated version called Dall-E Mini. Dayma's model eventually changes its name to Craiyon to avoid any legal disputes. Craiyon becomes rapidly popular due to its accessibility and through memes.

July 2022

The French National Center for Scientific Research debuts BigScience Large Open-Science Open-Access Multilingual Language Model (BLOOM) on the platform HuggingFace. With 176 billion parameters, BLOOM is larger than GPT-3 and also completely open-source. It goes beyond Meta's OPT by being freely available for research or commercial use.

September 2022

NVIDIA releases Get3D, a 3D-shape generator trained only on 2D images.

December 2021

Baidu releases ERNIE 3.0.

December 202

May 2021

Google announces Language Model for Dialogue Applications (LaMDA), its transformer-based large language model.

November 2021

Microsoft announces NÜWA, a multimodal model for text-to-image and text-to-video.

March 2022

MidJourney, a text-toimage model offered through a Discord server, moves into open beta.

2022

April 2022

Google releases its 540 billion parameter Pathways Language Model (PaLM), a multifunctional language model.

May 2022

Meta releases its Open Pretrained Transformer (OPT) model, a totally open-source LLM, to the public. Details about the model's training are also released. Access is available upon request, but the license limits usage of the model to research purposes.

May 2022

Google releases its

answer to Dall-E.

Imagen, though

not to the public.

July 2022

Research startup Stability.ai launches Stable Diffusion, an open-source text-toimage model.

September 2022

Meta announces Make-A-Video, its text-to-video model.

Image / video models

Text models

Business events

October 2022

Google launches text-to-video model Imagen Video.

November 2022

Stability.ai and Amazon Web Services announce a partnership.

January 2023

Microsoft announces a "multiyear, multibillion-dollar" investment in OpenAI and exclusive integration with its Azure cloud system.¹¹

February 2023

Microsoft releases a limited beta upgrade to its Bing search system, with new capabilities such as summarization and chat based on ChatGPT. GPT is also integrated into other Microsoft products including Teams and its Edge browser.

February 2023

Google announces BARD, its answer to ChatGPT, for limited beta testing.

November 2022

OpenAl announces ChatGPT, a powerful chatbot tool running on the GPT-3.5 model. It soon becomes the most quickly adopted tech product in history.

February 2023

Runway, an Al research lab, announces its Gen-1 model that allows for style transfer across videos.

February 2023

AWS and Hugging Face

announce partnership.

February 2023

Google invests \$300 million into Anthropic, a research startup made up of former OpenAl employees.
Anthropic calls Google its "preferred cloud provider." 12

March 2023

Google announces generative tools to be integrated into workplace.

March 2023

OpenAl releases GPT-4.

ASSISTANT

As an AI, I can't directly produce images or diagrams. However, I can describe how such a timeline could look:

(-)

- 1. 1950s to Early 90s Basic Algorithms: In the early era of computing, image synthesis was rather basic and primitive. There weren't any sophisticated models or algorithms that could generate images. Methods used were elementary like connected components for binary image, histogram modelling, etc.
- 2. Mid 90s Fractal Geometry: Fractal geometry was then used to generate images particularly terrain and natural-looking structures.
- 3. Late 90s Texture Synthesis Methods: Towards the end of the 90s, methods for texture synthesis were developed which used pixel-based and patch-based copying algorithms. Some examples of these techniques include Image Quilting and Efros Leung.
- 4. 2000s onwards Development of Generative Models: With the turn of the century, more direct models for generating images were produced. This includes algorithms such as Generative Adversarial Networks(GANs).

What about stochastic grammars for images?

ASSISTANT

Indeed, stochastic grammars have been used as generative models for images as well. Stochastic image grammars are a particularly expressive class of generative models that have proven useful in a wide range of applications, from texture synthesis to object detection in computer vision.

To add this to the timeline:

1980s-2000s - Stochastic Grammars: In the 80s and into the 2000s, stochastic grammars, such as Stochastic Context-Free Grammars (SCFGs), began to be used to generate complex structured images. These models, inspired by Chomsky's work on formal grammars in linguistics, provide a structured, hierarchical way to model the relationships between different parts of an image.

Remember to consult a graphic designer to create a visual timeline and to research more detailed information as required. Also note that many of these technologies developed in parallel and continuously improved over time.

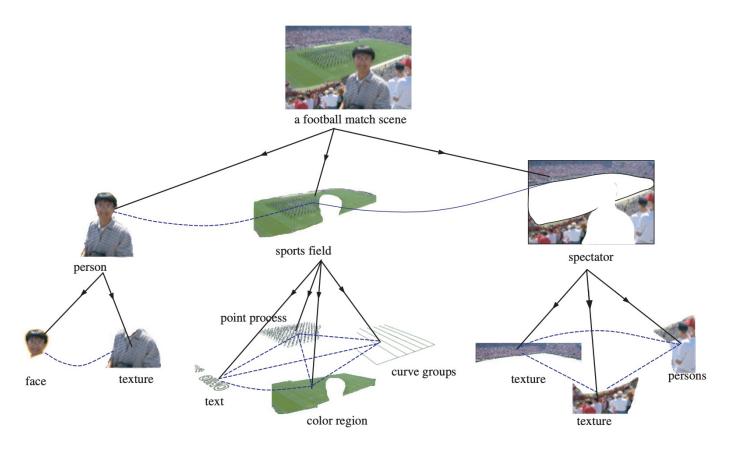


Fig. 1.1 Illustrating the task of image parsing. The parse graph includes a tree structured decomposition in vertical arrows and a number of spatial and functional relations in horizontal arrows. From [72].

266 Introduction

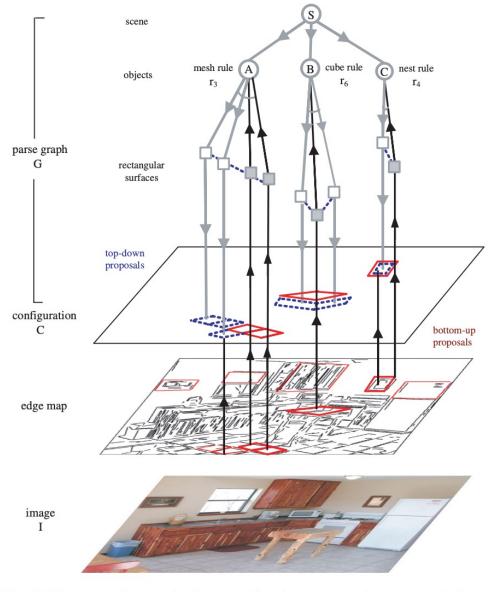


Fig. 1.2 Illustrating the recursive bottom-up/top-down computation processes in image parsing. The detection of rectangles (in red) instantiates some non-terminal nodes shown as upward arrows. They in turn activate graph grammar rules for grouping larger structures in nodes A, B, and C, respectively. These rules generate top-down prediction of rectangles (in blue). The predictions are validated from the image under the Bayesian posterior probability. Modified from [59].

Probabilistic generative models were largely developed in the 20th century, with many key concepts emerging in the latter half of the century.

Here is a rough timeline:

1. 1950s: In the mid-20th century, simple statistical models like Markov models were being used, which can be considered as early instances of generative models.

2. 1980s-1990s: During this era, probabilistic generative models became more sophisticated. Bayesian networks and hidden Markov models (HMMs) were popular methods around this time.

3. 1990s-2000s: Mixture models, like Gaussian Mixture Models (GMMs), and other latent variable models were developed during the 90s and into the 2000s, becoming a powerful tool for analyzing and generating data.

4. 2000s-Present: More recently, this category has grown to include models like Deep Boltzmann Machines and Generative Adversarial Networks (GANs), which are very powerful and intricate generative models capable of producing highly realistic results.

Remember, this is a general overview and specific developments may have occurred at different times. Technological development is often gradual and iterative, with many different contributors adding

Topic Modeling

what about topic modeling

\$

Topic modeling is a significant area of generative models, particularly in text analysis and natural language processing. It was developed and popularized in the late 1990s and early 2000s. Here's how it fits into the history of generative models:

Topic Modeling Timeline (1950-2000)

1980s: Early Foundations

1980s: Latent Semantic Analysis (LSA). Although not strictly generative, LSA provides a
foundation for topic modeling by analyzing the relationships between a set of documents and
terms, reducing dimensions to uncover latent structures in the data.

1990s: Development and Formalization

1999: Latent Dirichlet Allocation (LDA). Introduced by Blei, Ng, and Jordan, LDA is a
probabilistic model that represents documents as mixtures of topics, and topics as mixtures
of words. This approach became a cornerstone of topic modeling and generative text models

2000s: Expansion and Application

 2003: Topic Modeling with LDA. The LDA model gains traction in various applications, from document classification to information retrieval, influencing subsequent research and tools.

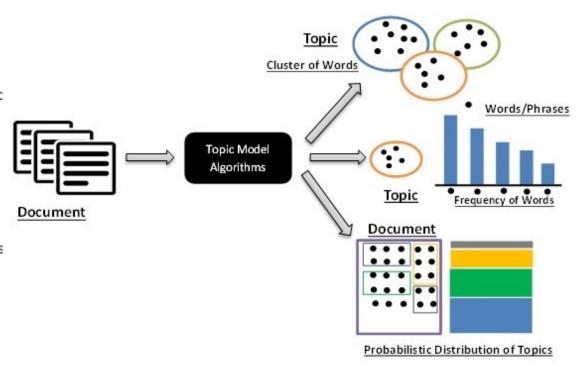
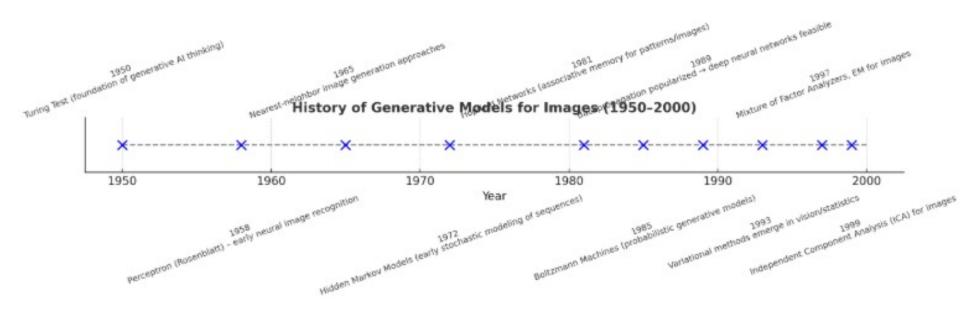


Figure-1

Improvement from 2023 to 2025?

Produce a diagram with the history of generative models for images from 1950 to 2000

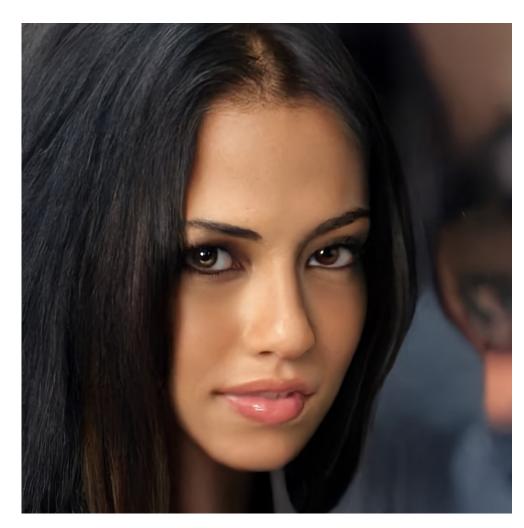


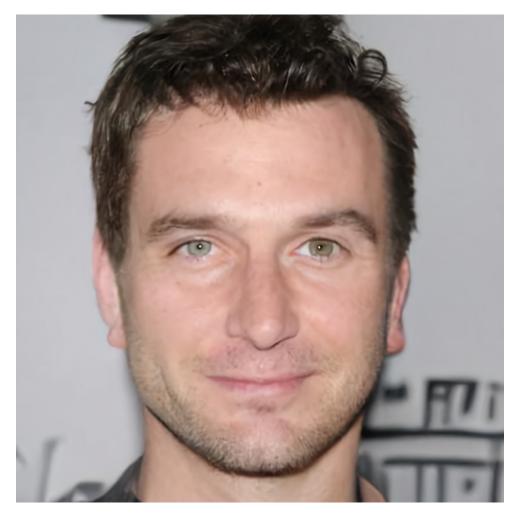
Here's a timeline diagram showing key milestones in the history of generative models for images from 1950 to 2000. Would you like me to extend this further into the 2000–2025 era (VAEs, GANs, diffusion models, etc.) to give you the full picture?

Progress in Generative Models of Images



Progress in Generative Models of Images





Song et al., Score-Based Generative Modeling through Stochastic Differential Equations, 2021

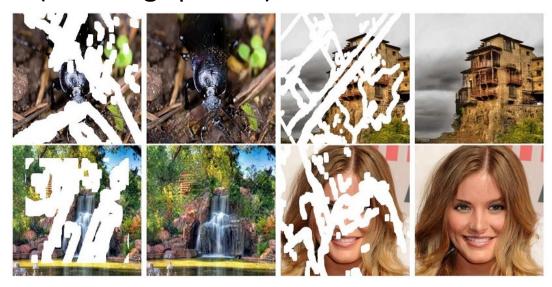
Progress in Inverse Problems

P(high resolution | low resolution)



Menon et al, 2020

P(full image | mask)



Liu al, 2018

P(color image | greyscale)



Antic, 2020

Progress in Inverse Problems

Stroke Painting to Image

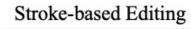






















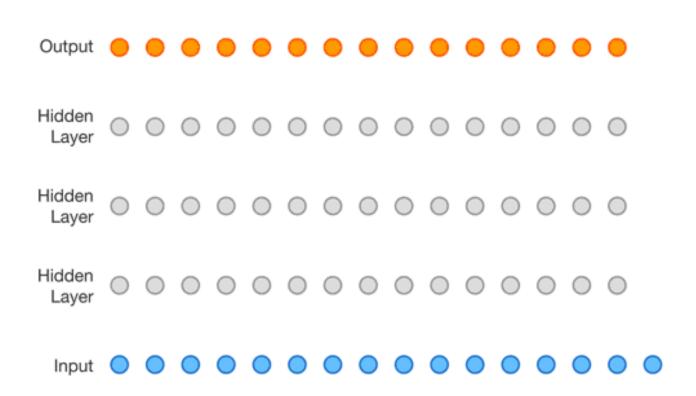


Input

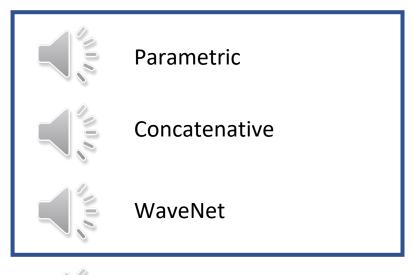
Output

WaveNet

Generative model of speech signals



Text to Speech

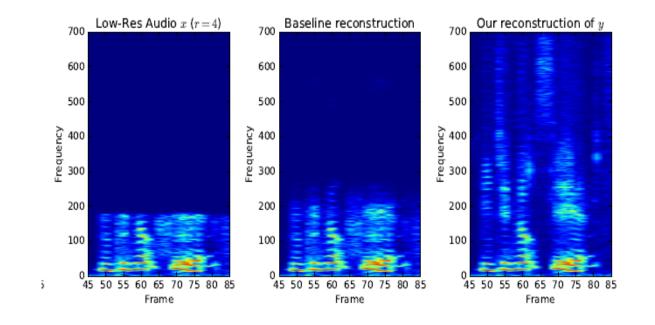


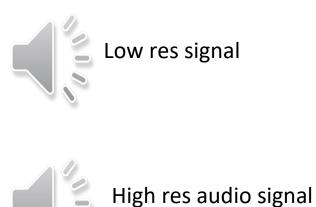


van den Oord et al, 2016c

Audio Super Resolution

Conditional generative model P(high-res signal | low-res audio signal)





Kuleshov et al., 2017

Machine Translation

Conditional generative model P(English text| Chinese text)

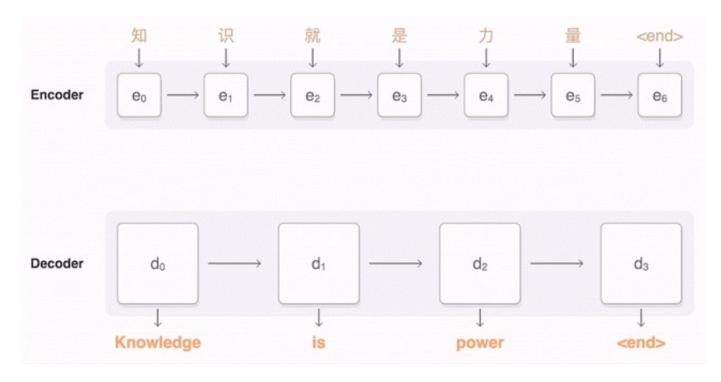


Figure from Google AI research blog.

Language Generation

Custom prompt

To get an A+ in deep generative models, students have to

Completion

To get an A+ in deep generative models, students have to be willing to work with problems that are a whole lot more interesting than, say, the ones that most students work on in class. If you're a great student, the question above can be avoided and you'll be able to do great work, but if you're not, you will need to go beyond the basics before getting good.

Now to be clear, this advice is not just for the deep-learning crowd; it is good advice for any student who is taking his or her first course in machine learning.

The key point is that if you have a deep, deep brain of a computer scientist, that's just as important to you.

Radford et al., 2019 Demo from talktotransformer.com

Code Generation

```
тs sentiment.ts
                                                 🔏 addresses.rb
parse_expenses.py
 1 import datetime
 3 def parse_expenses(expenses_string):
       """Parse the list of expenses and return the list of triples (date, va
```

Images and Text

TEXT PROMPT

an armchair in the shape of an avocado. . . .

AI-GENERATED IMAGES











Edit prompt or view more images↓

P(image | caption)

TEXT PROMPT

a store front that has the word 'openai' written on it. . . .

AI-GENERATED IMAGES











DeepFakes

Which image is real?

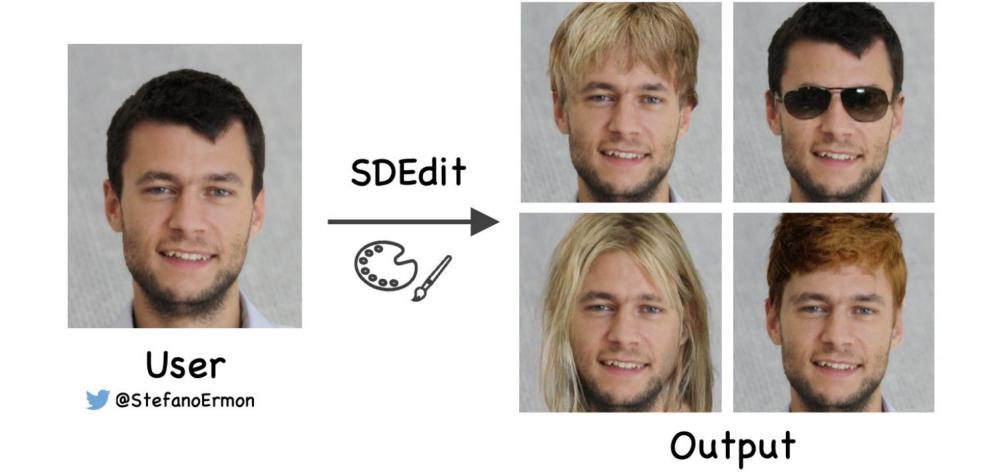


Image Translation

Conditional generative model P(zebra images| horse images)



Zhu et al., 2017