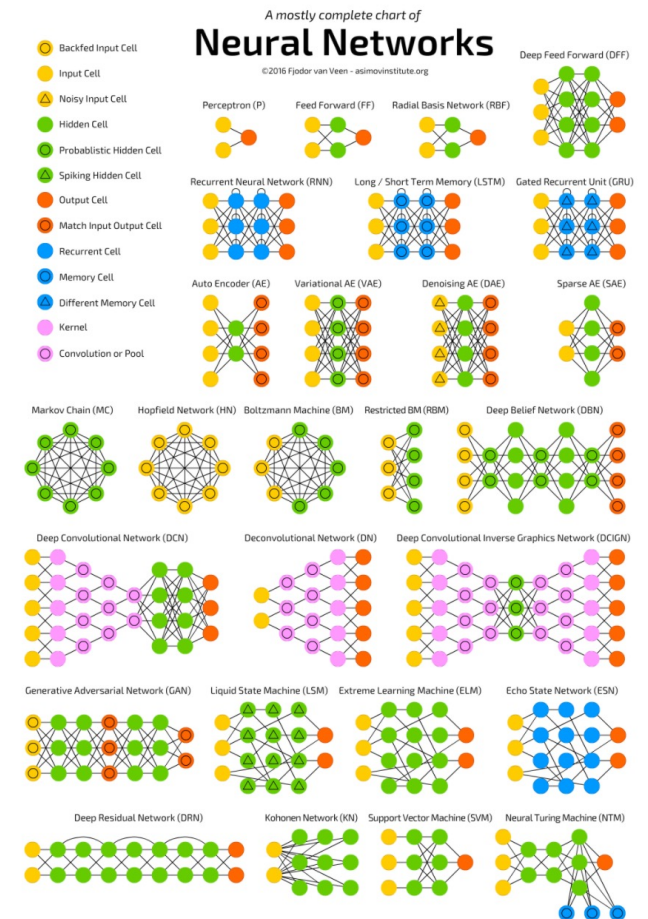


# Generative AI

CSCI 4360/6360 Data Science II

# The Neural Network Zoo

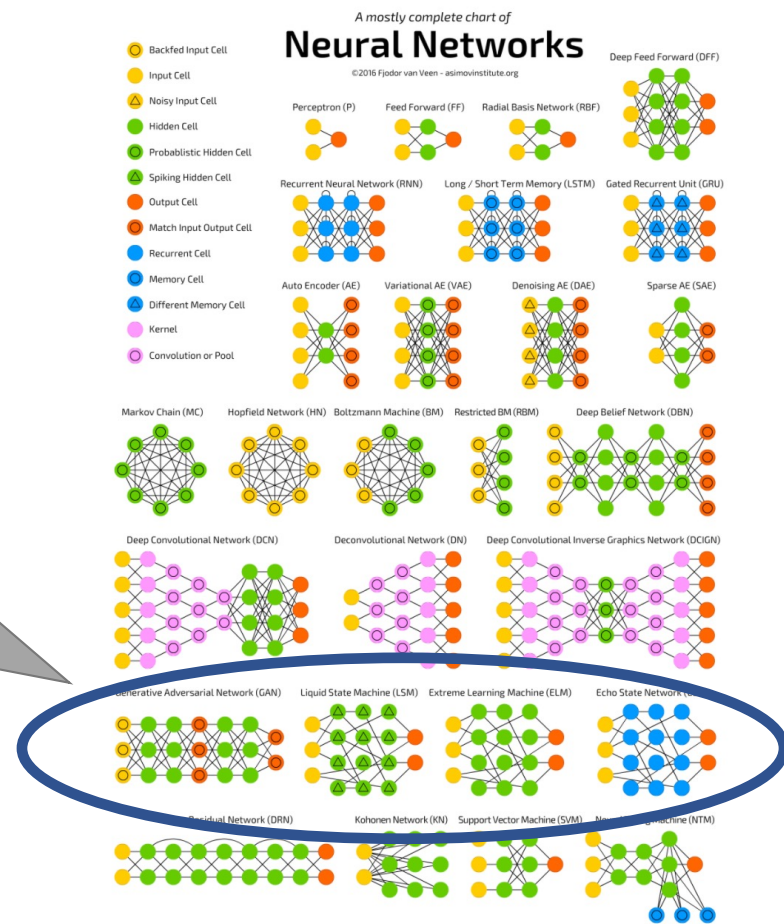
- <http://www.asimovinstitute.org/neural-network-zoo/>



# The Neural Network Zoo

- <http://www.asimovinstitute.org/neural-network-zoo/>

Today

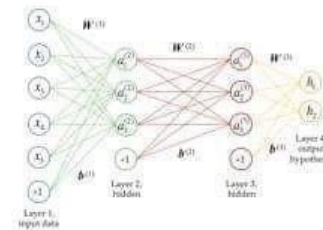


# Caveat Emptor

	A parrot	Machine learning algorithm
Learns random phrases	✓	✓
Doesn't understand shit about what it learns	✓	✓
Occasionally speaks nonsense	✓	✓
Is a cute birdie parrot	✓	✗



Machine learning algorithm





What are generative models?



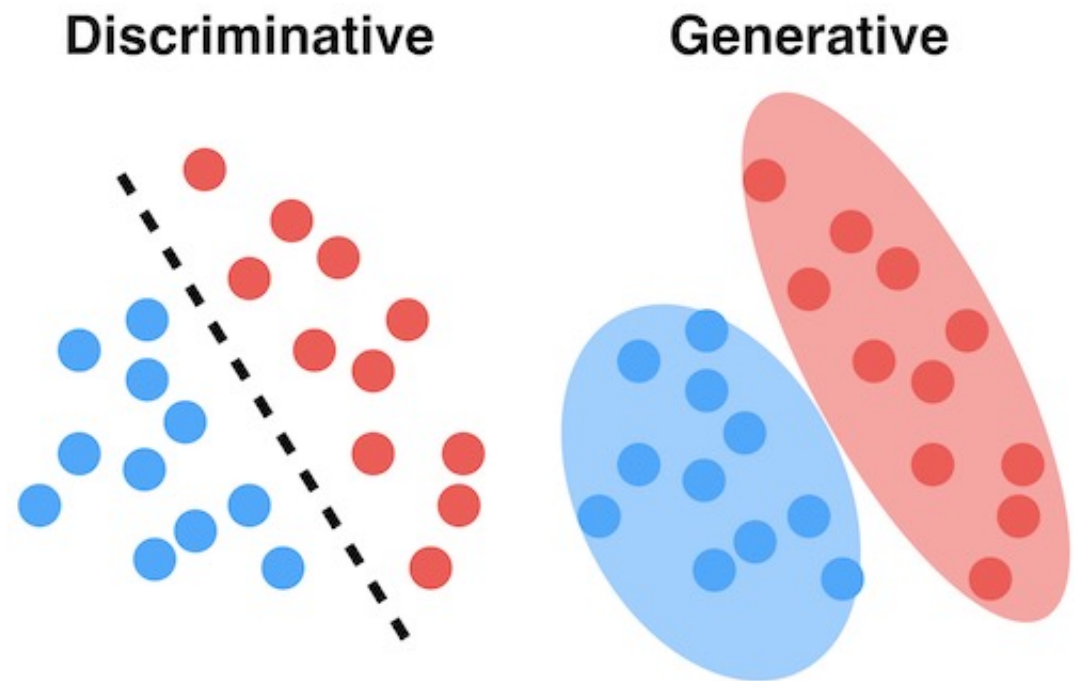
# What is a “generative model”?

- Discriminative
  - Logistic Regression
  - Support Vector Machines
  - Random Forests

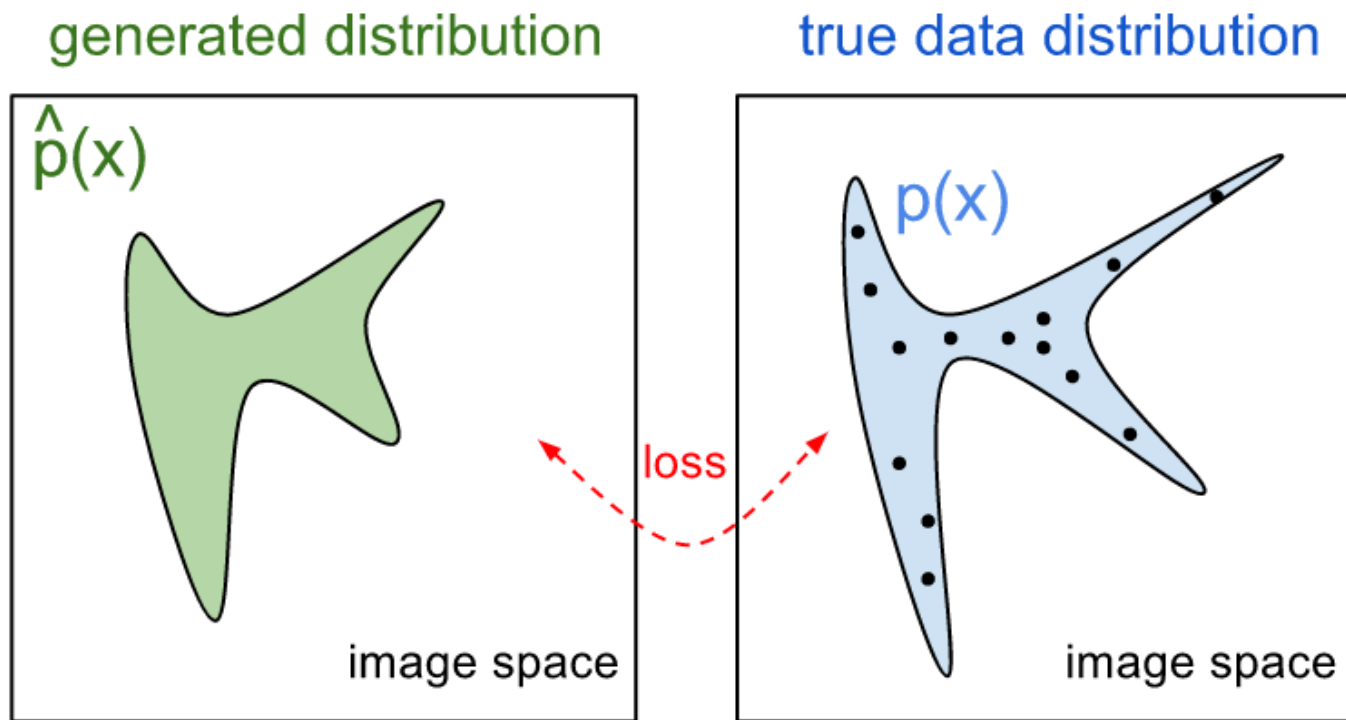
$$P(Y | X)$$

- Generative
  - Gaussian Naïve Bayes
  - Variational Autoencoders
  - Adversarial Networks

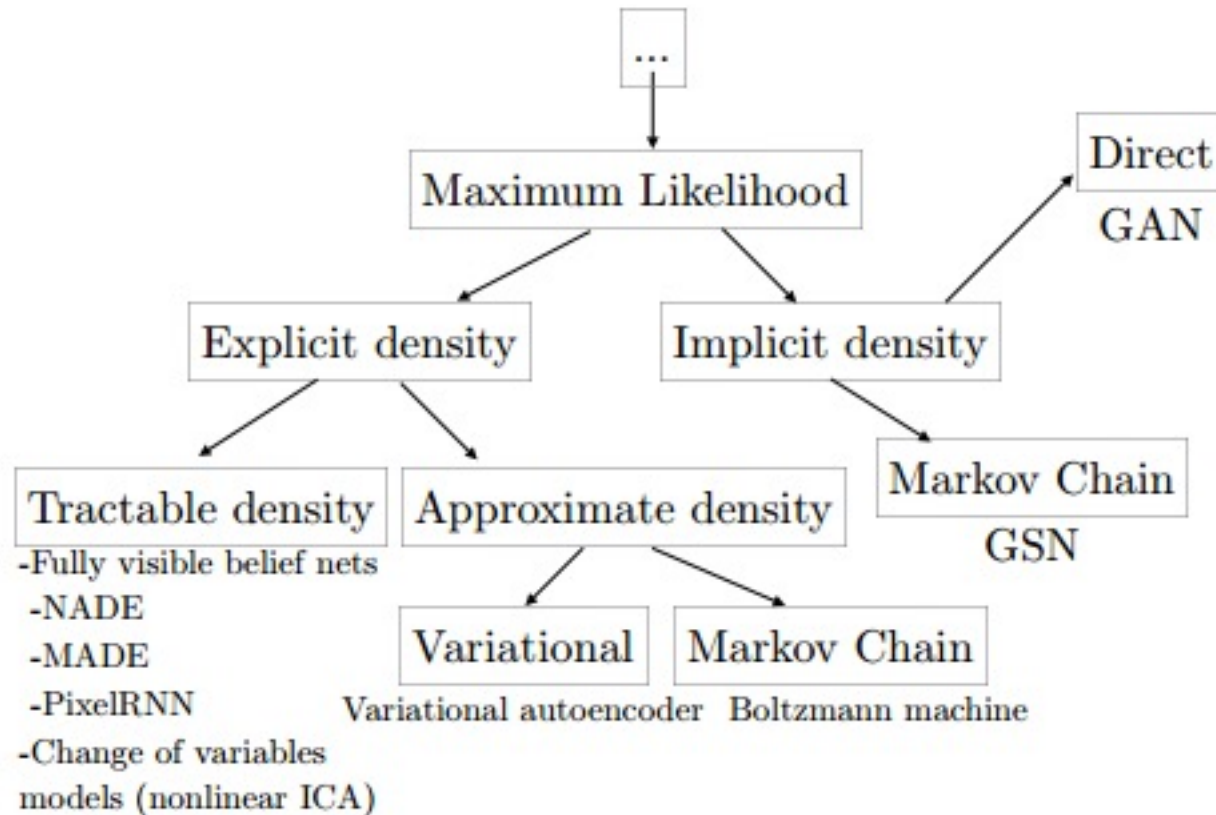
$$P(X, Y) \text{ and } P(Y)$$



# Generative Models



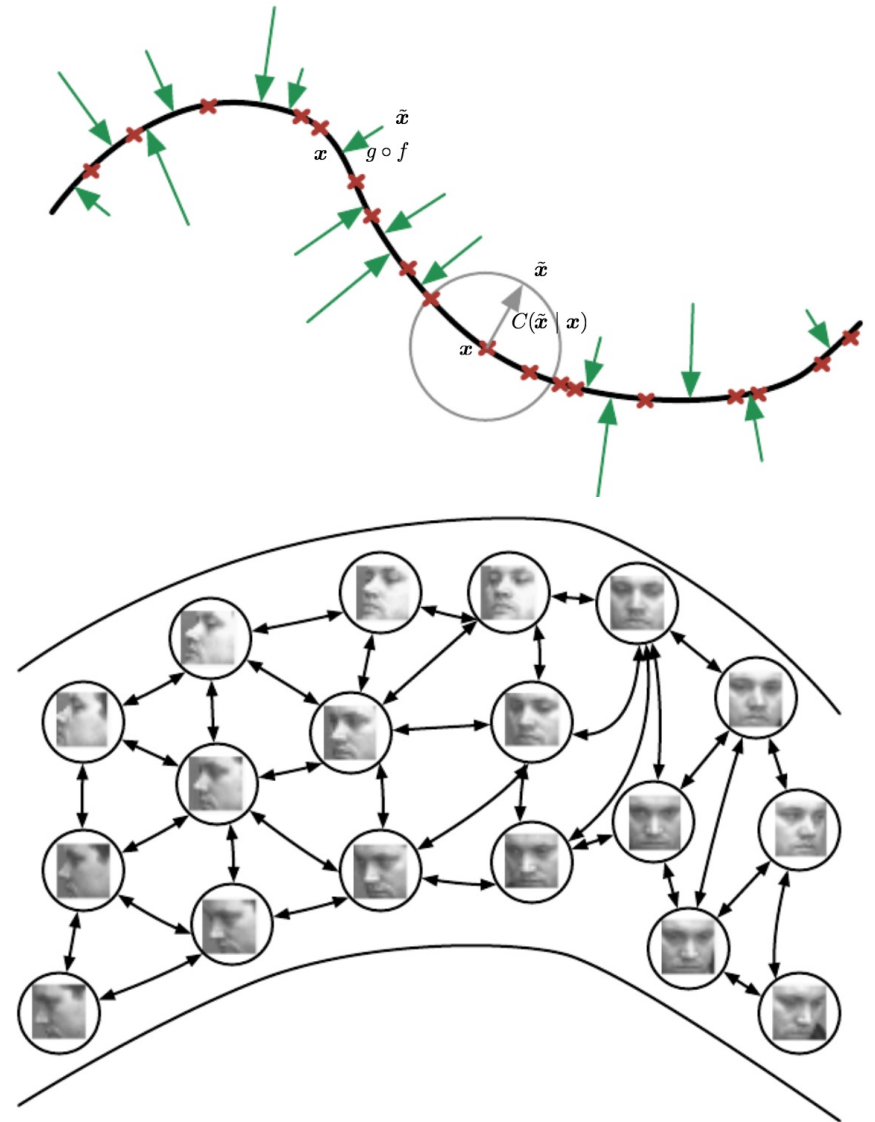
# Generative Models

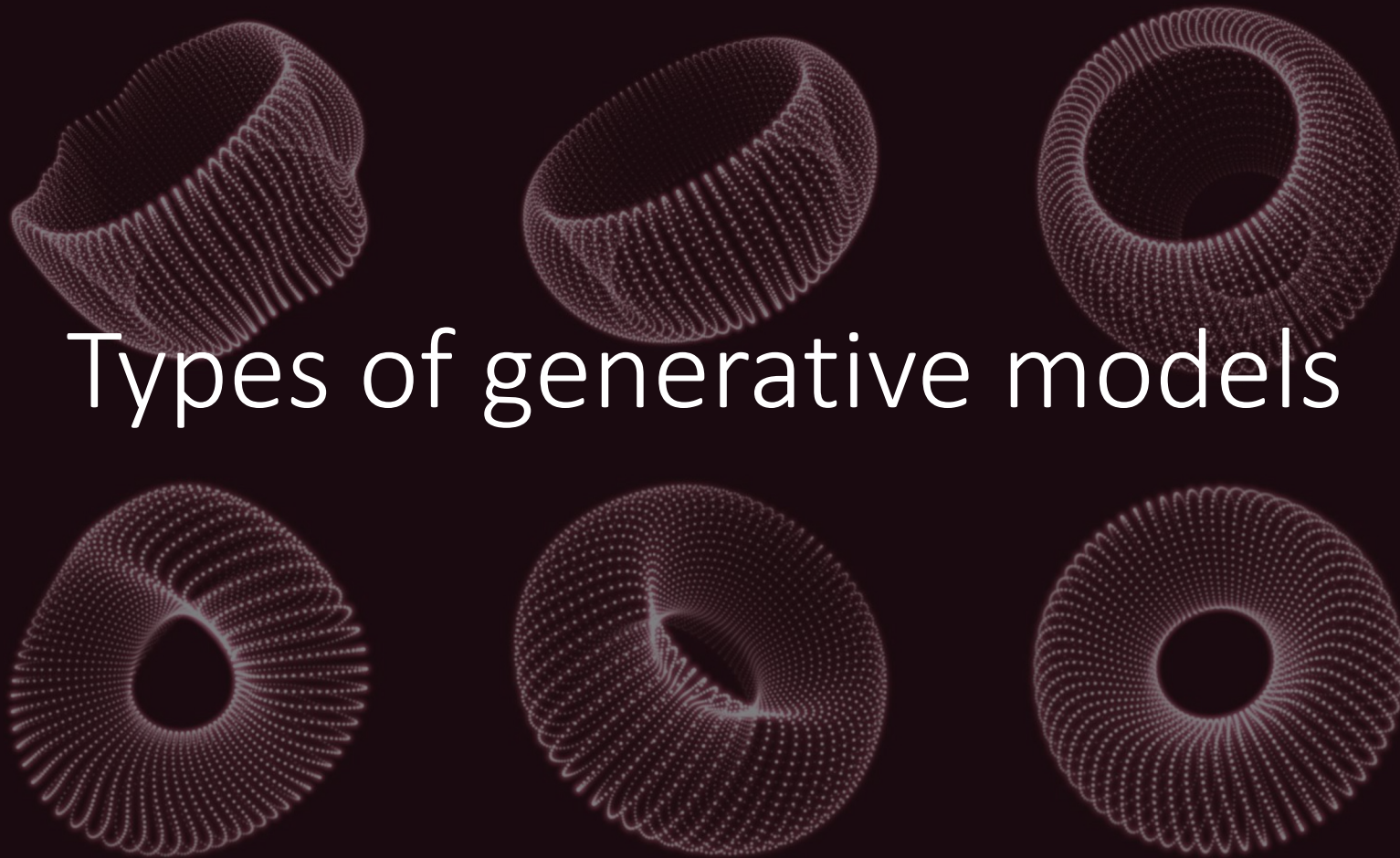




# Generative Models

- Learning a *distribution* or *manifold*
  - Statistical notion of *how the data were generated*
- $P(X)$  asks: how *likely* is the data point  $X$ ?
  - If likely  $\rightarrow X$  was **generated** by this process
- Compare to  $P(Y)$ , which asks: how likely is the *label*?
  - If likely  $\rightarrow X$  has label  $Y$





# Types of generative models

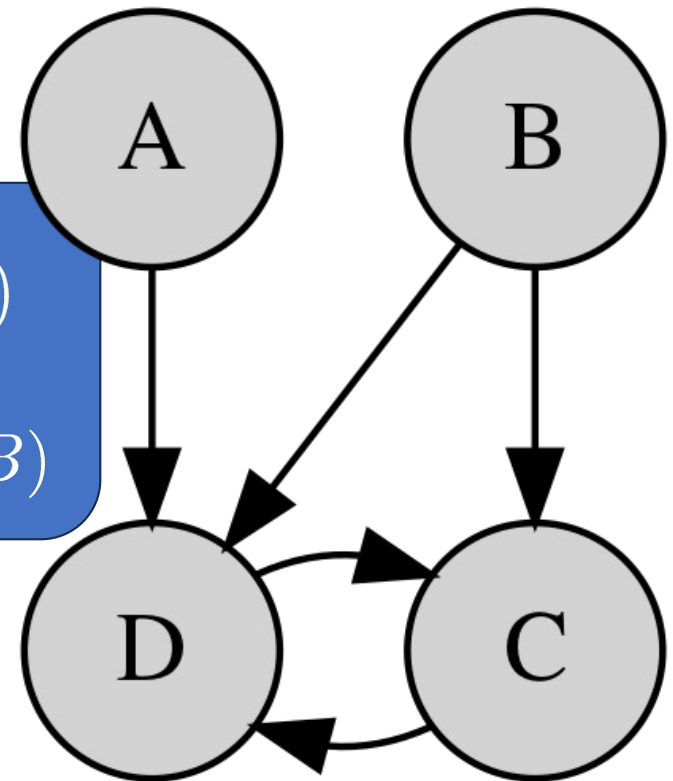
# Probabilistic Graphical Models

- Arrows represent conditional dependencies between random variables

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{parents}_i)$$

$$P(A, B, C, D) = P(A)P(B)P(C, D | A, B)$$

- Structure is used in generative models
  - Latent generating distribution (hidden)
  - Observed variables (influenced by latent vars)



# Variational Inference

- What is variational inference?
- Good for learning latent variable models (i.e., generating distributions of data)
- For each observation  $x$  we assign a hidden variable  $z$ ; our model  $p$  describes the joint distribution between  $x$  and  $z$

Of course these are the things we want to calculate

- Inference is  $p(z|x)$
- Learning involves  $p(x)$

$p_{\theta}(z)$  is very easy 🐣,

$p_{\theta}(x|z)$  is easy 🐭,

$p_{\theta}(x, z)$  is easy 🐼,

$p_{\theta}(x)$  is super-hard 🐍,

$p_{\theta}(z|x)$  is mega-hard 🐉

# Variational Inference

- Rather than learning  $p(z/x)$  directly, variational inference approximates with  $q(z/x)$
- Maximize the evidence lower bound (ELBO)

$$\text{ELBO}(\theta, \psi) = \sum_n \log p(x_n) - \text{KL} [q_\psi(z|x_n) || p_\theta(z|x_n)]$$

- This can be written in terms of the “friendly” emojis

$p_\theta(z)$  is very easy 🐣,

$p_\theta(x|z)$  is easy 🐱,

$p_\theta(x, z)$  is easy 🐼,

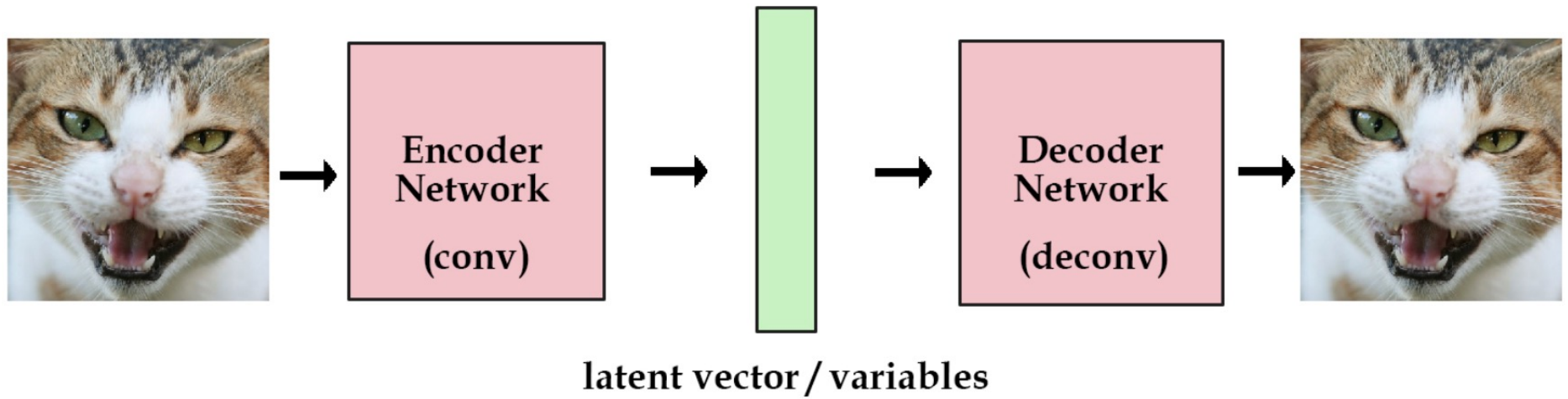
$p_\theta(x)$  is super-hard 🐍,

$p_\theta(z|x)$  is mega-hard 🐉

$$\text{🦹} = - \sum_n \mathbb{E}_{\text{🐰}} \log \frac{\text{🐰}}{\text{🐼}} + \text{constant}$$

$$= \sum_n \mathbb{E}_{\text{🐰}} \log \text{🐱} - \sum_n \mathbb{E}_{\text{🐰}} \text{KL}[\text{🐰} || \text{🐣}]$$

# Recall: Autoencoders



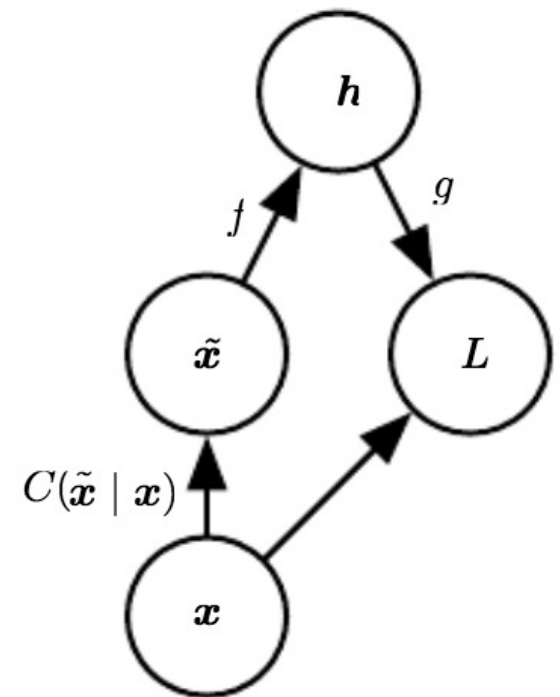
# Denoising Autoencoders

- Define a corruption process,  $C$

$$C(\tilde{x} | x)$$

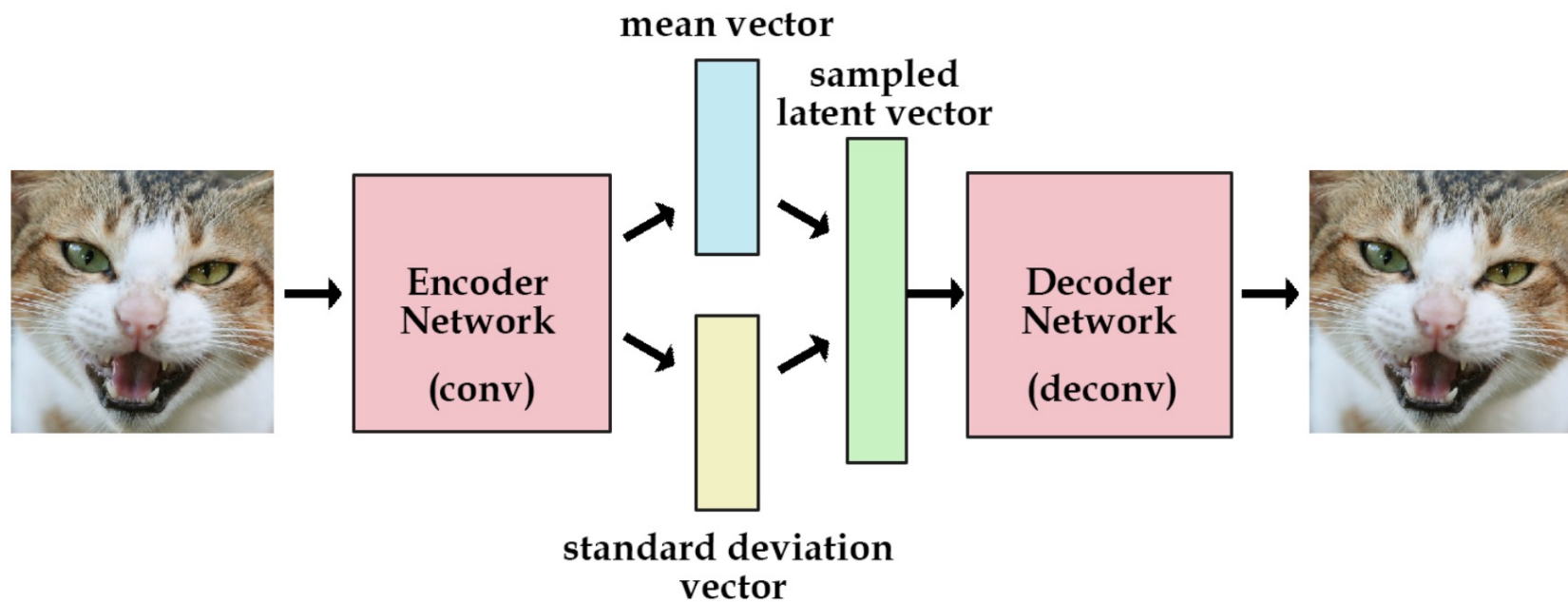
- Autoencoder learns a *reconstruction distribution*  $p_{\text{reconstruct}}(x | \tilde{x})$

1. Sample a training example  $x$
2. Sample a corrupted version  $\tilde{x}$  from  $C$
3. Use  $(x, \tilde{x})$  as a training pair



# Denoising Autoencoders

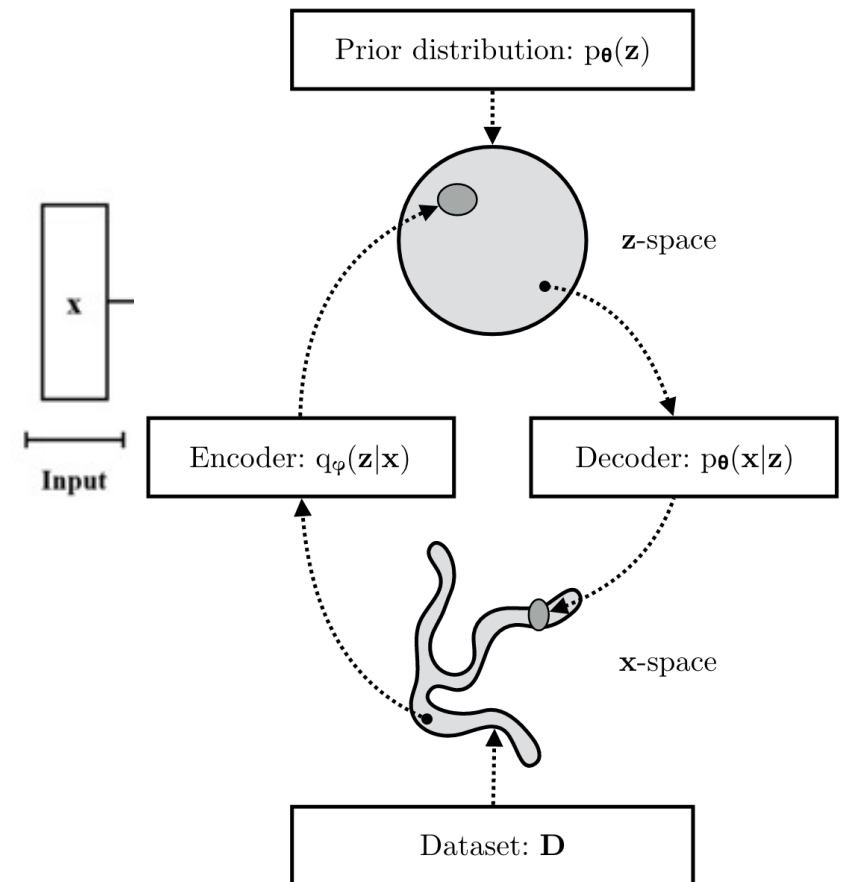
- De-corruption process results in learning a *distribution*





# Variational Autoencoders (VAEs)

- Associated with autoencoders by virtue of architecture
  - Goal is to map inputs to latent space
- Encoder: Learn parameters of variational distribution,  $q(z | x)$
- Decoder: Sample (generate!) from learned distribution to reconstruct input

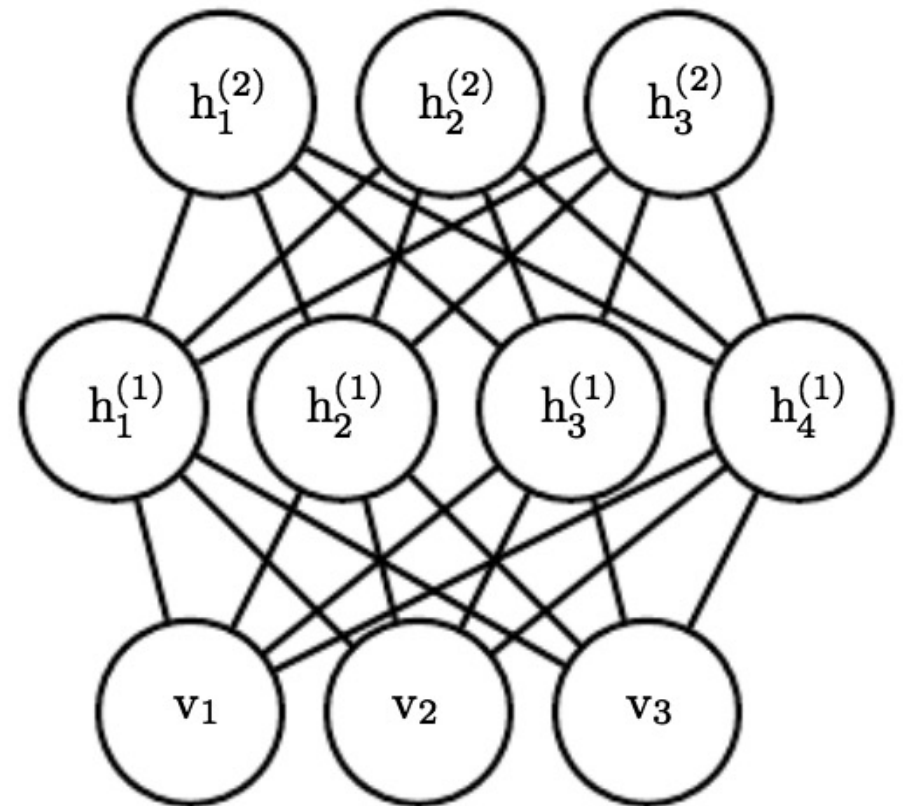


# Restricted Boltzmann Machines (RBMs)

- Wholly undirected deep network
  - Implementation of a probabilistic graphical model
  - Each variable conditionally independent given neighboring nodes
- Parameterized by energy function

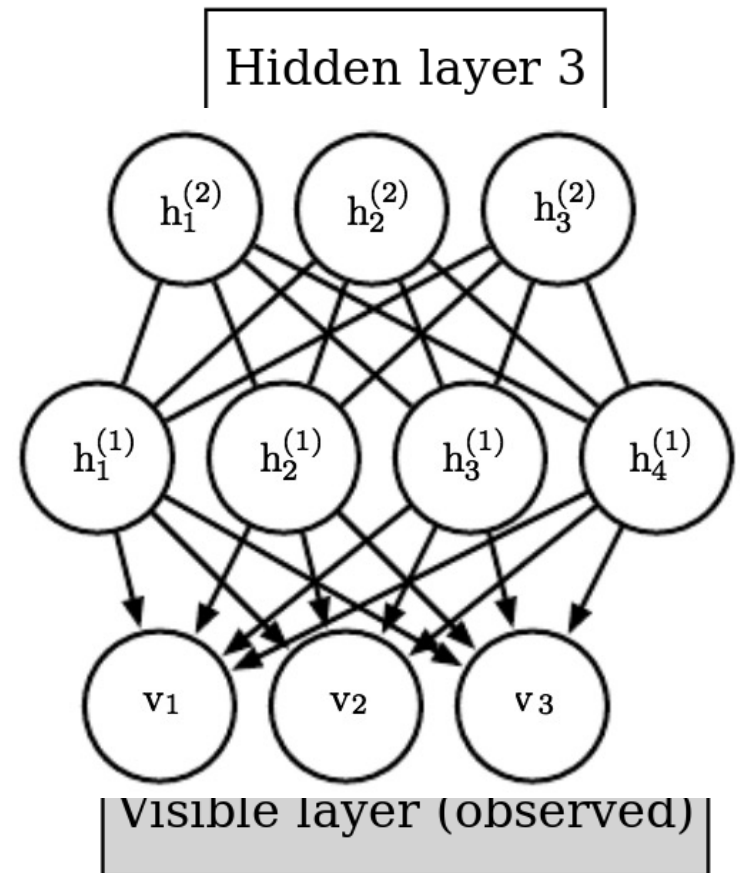
$$P(\mathbf{v}, \mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \mathbf{h}^{(3)}) = \frac{1}{Z(\boldsymbol{\theta})} \exp(-E(\mathbf{v}, \mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \mathbf{h}^{(3)}; \boldsymbol{\theta}))$$

hard, but training is paradoxically easy



# Deep Belief Nets (DBNs)

- Connections *between* layers, but not units *within* a layer
- Arguably one of the first successful applications of modern deep learning
  - Hinton 2006 and 2007
- Often built from an RBM template
- Training is nearly intractable
  - Posterior has to be approximated through annealed importance sampling (AIS)



# Generative Adversarial Networks (GANs)

“

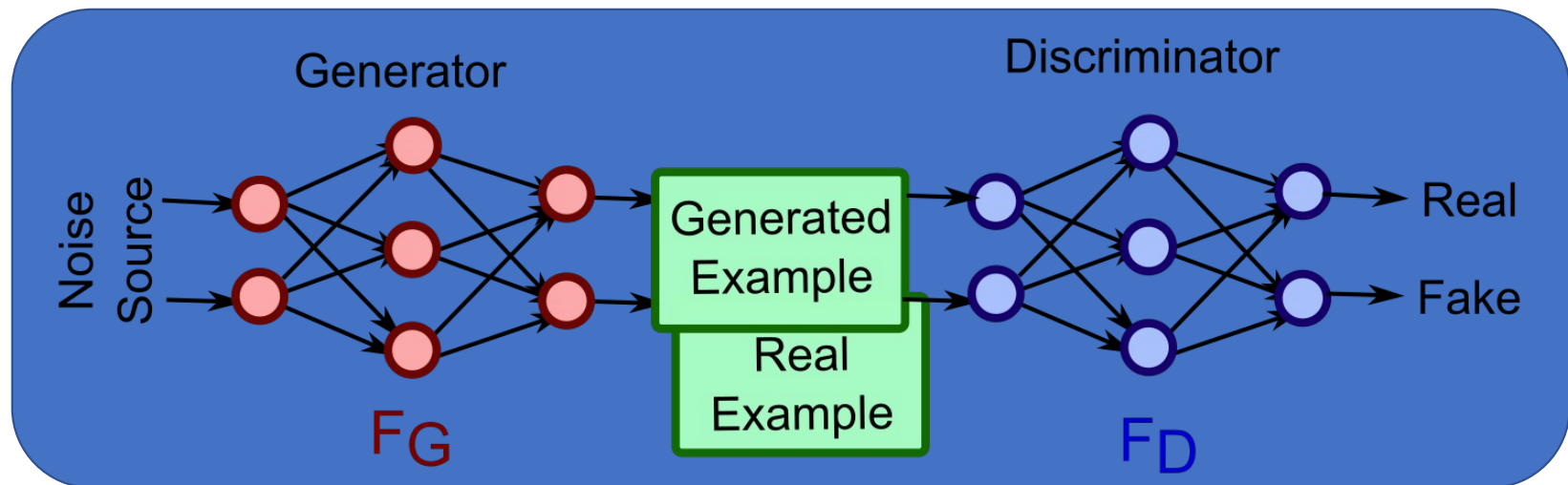


There are many interesting recent development in deep learning...The most important one, in my opinion, is adversarial training (also called GAN for Generative Adversarial Networks). This, and the variations that are now being proposed, is the most interesting idea in the last 10 years in ML.

Yann LeCun

# GANs

- Game-theoretic approach to generative modeling
- Two deep networks: a **generator** (G) and **discriminator** (D)



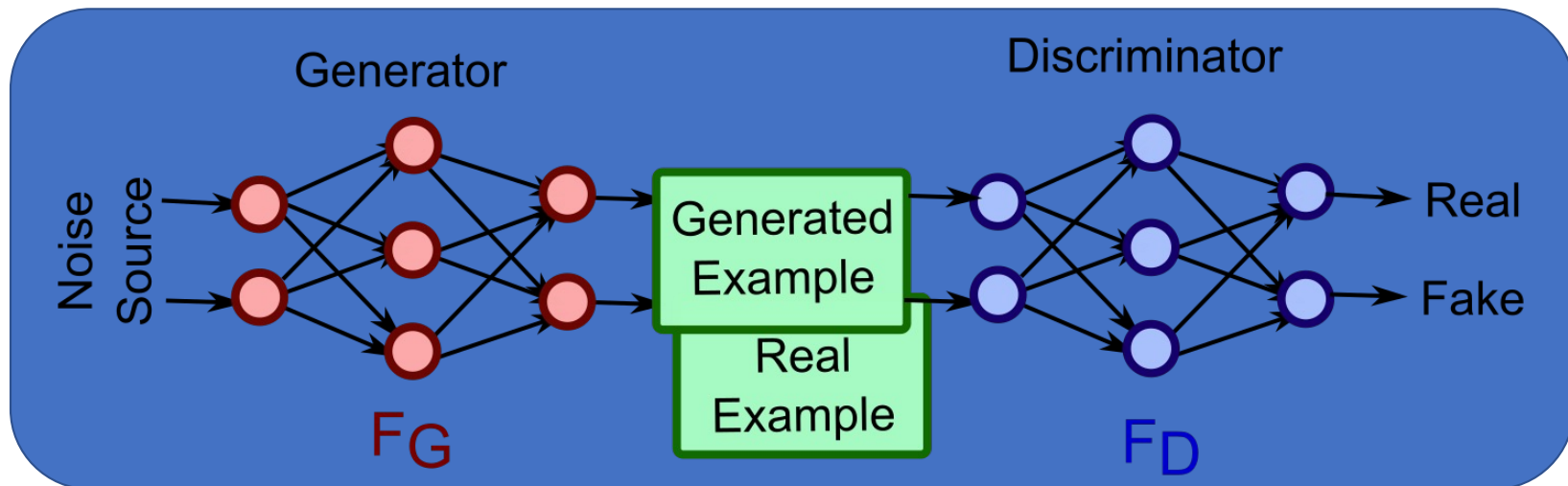
# GANs

- **Generator**

- Input: a random vector  $z$
- Output: something as close to a “real” data point as possible

- **Discriminator**

- Input: a “real” data point OR a synthetic example from  $G$
- Output: 1 or 0 (real or fake)



# GANs

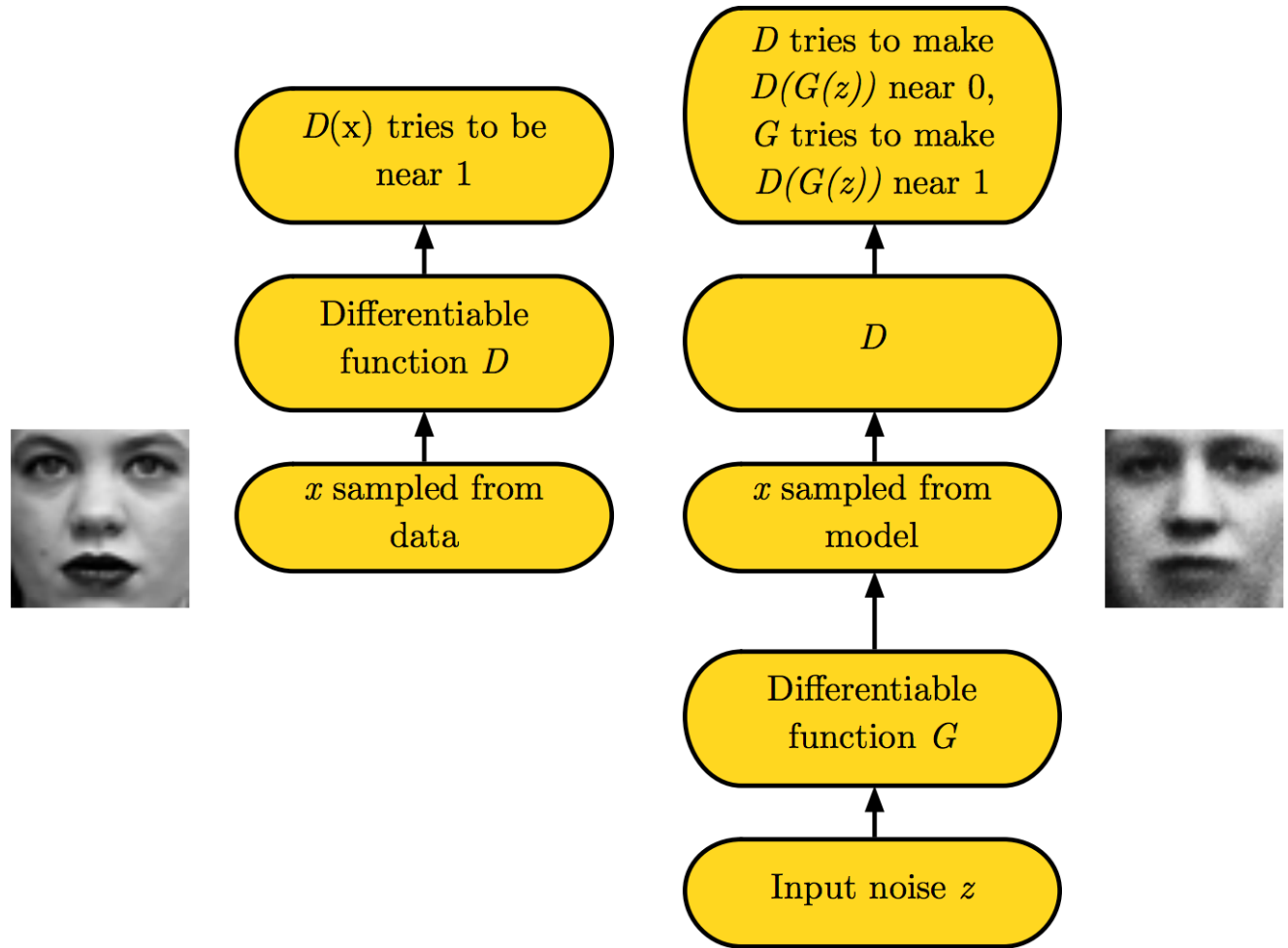
- Minimax “game”
  - Generator and Discriminator have competing objectives
  - Goal is to find an equilibrium point

$$\min_G \max_D \mathbb{E}_{x \sim P_{real}} \log D(x) - \mathbb{E}_z \log(1 - D(G(z)))$$

Maximize the Discriminator's likelihood of identifying a real data example

Minimize the Discriminator's ability to differentiate real data from Generator exemplars

# GANs



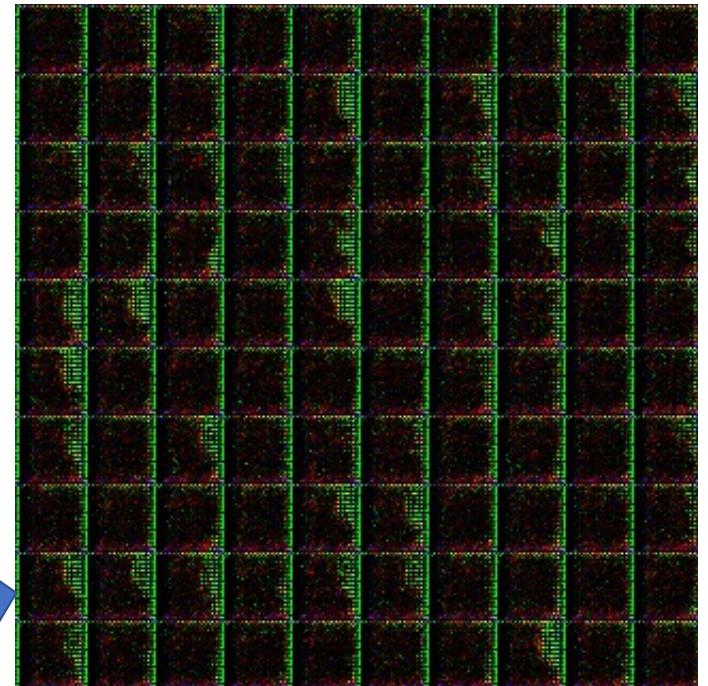


# VAEs versus GANs



VAEs  
Expectation over  
learned  
distribution results  
in blurring

GANs  
Samples from learned  
distribution, resulting  
in sharper images

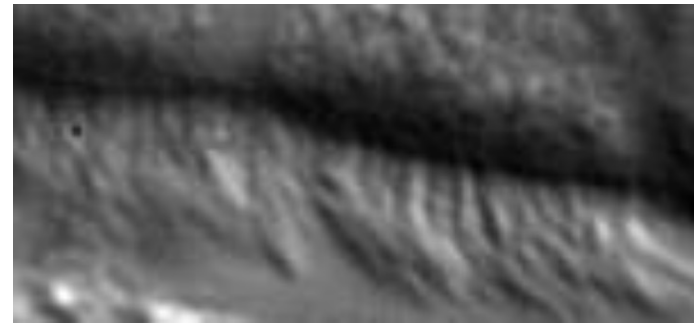
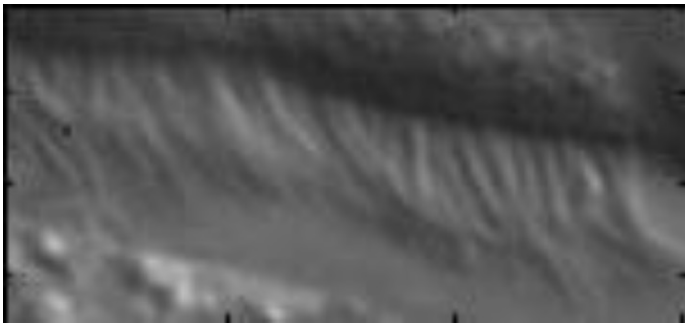


# Autoregressive (AR) Models

- DALLE-1, in January 2021, was an autoregressive Transformer
- Our good friends, ~~Thing 1~~ and ~~Thing 2~~ Appearance and State

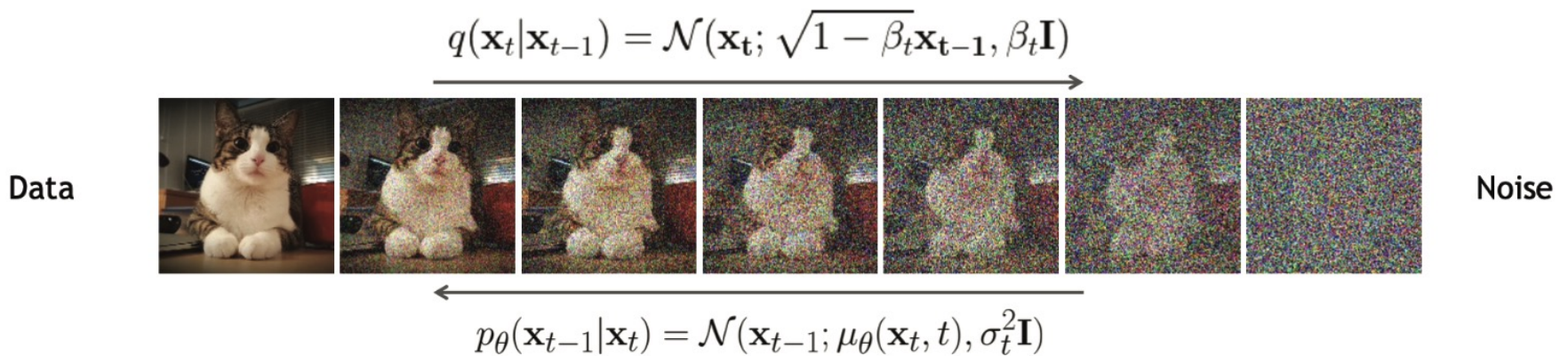
$$y_t = Cx_t + u_t$$
$$x_t = Ax_{t-1} + Wv_t$$

- Once you've learned  $A_i$ , you can generate new  $x_t$ !



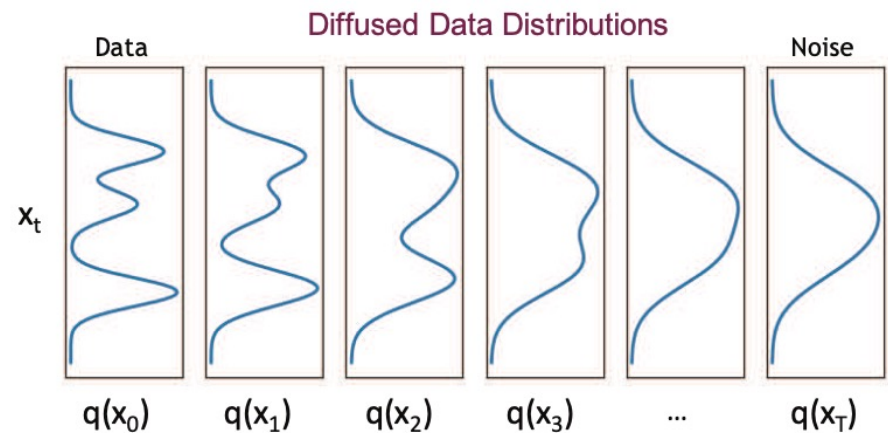
# Latent Diffusion

- Closely related to VAEs, normalizing flows, and energy-based models
- **Hard** to convert noise into structured data
- **Easy** to convert structured data into noise



# Latent Diffusion

- Similar to hierarchical VAE
  - BUT all latent states have same dimensionality as input
  - BUT encoder is a linear Gaussian model, rather than being learned
- Results in a very simple objective
  - No risk of posterior collapse (unlike GANs)
- Numerous variations of LD
  - Denoising Diffusion Probabilistic Models (DDPM)
  - Noise-conditioned Score Networks (NCSN)
  - Stochastic Differential Equations (SDE)



# Large Language Models (LLMs)

- Not unique generative models *per se*
  - LLMs = very, very large Transformers
  - Usually with autoregressive blocks at inference / decoding
  - Trained on city blocks' worth of GPUs
- Sometimes called “Foundation Models”
  - “Foundation Models” are only found on Terminus; elsewhere in the galaxy, they're just “Sparkling Language Models”



Commercial generative models

# Probably screaming into the void here, but...

- There's AI, and there's AI
- AI
  - Large language or image models
  - Trained on massive amounts of data with large numbers of parameters
  - Does a frighteningly good job of mimicking humans at very specific tasks
  - **Not intelligent**
- AI
  - Intelligence that isn't human but made by humans, aka artificial
  - Mimics humans very well at *all possible tasks, even those it wasn't trained on*
  - Nowhere in the 5-10 year roadmap



# DALL-E, Midjourney, Stable Diffusion, Firefly

- Corporate backed text-to-image generators
- Subscription fees
- Open source options
- Training data



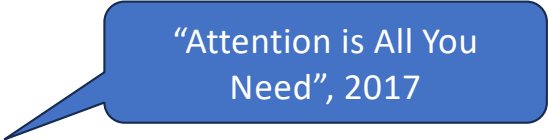
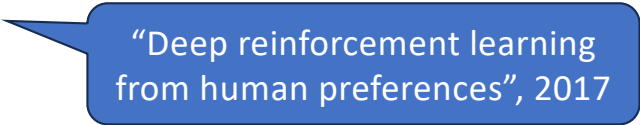


# Evolution of Image generators

- DALL-E
- “a muscular barbarian with weapons beside a CRT television set, cinematic, 8K, studio lighting.”
- April 2022
- October 2023



# GPT-4

- Powers ChatGPT
- A “Transformer-style model pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers.”  
 “Attention is All You Need”, 2017
- “The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF).”  
 “Deep reinforcement learning from human preferences”, 2017
- Several thousand GPUs + petabytes of data = ChatGPT

# PaLM, Cerebras, LLaMA, Falcon, OpenHermes

- Similar underlying architecture to ChatGPT
- Billions (to trillions?) of parameters
  - GPT-4 rumored to have ~2T parameters (source: SoC Day keynote speaker)
- Billions to trillions of training tokens
  - PaLM 2 and LLaMA 2: 3.6T and 2T, respectively
- Varying levels of openness
  - Some pre-trained models on Huggingface
  - An open LLM + RLHF (reinforcement learning from human feedback) + RLAIIF (reinforcement learning from AI feedback) + DPO (direct preference optimization) = best bang for buck, outside of ChatGPT or similar



Technical, Ethical,  
and Legal  
Considerations

## This is not unique to Generative AI

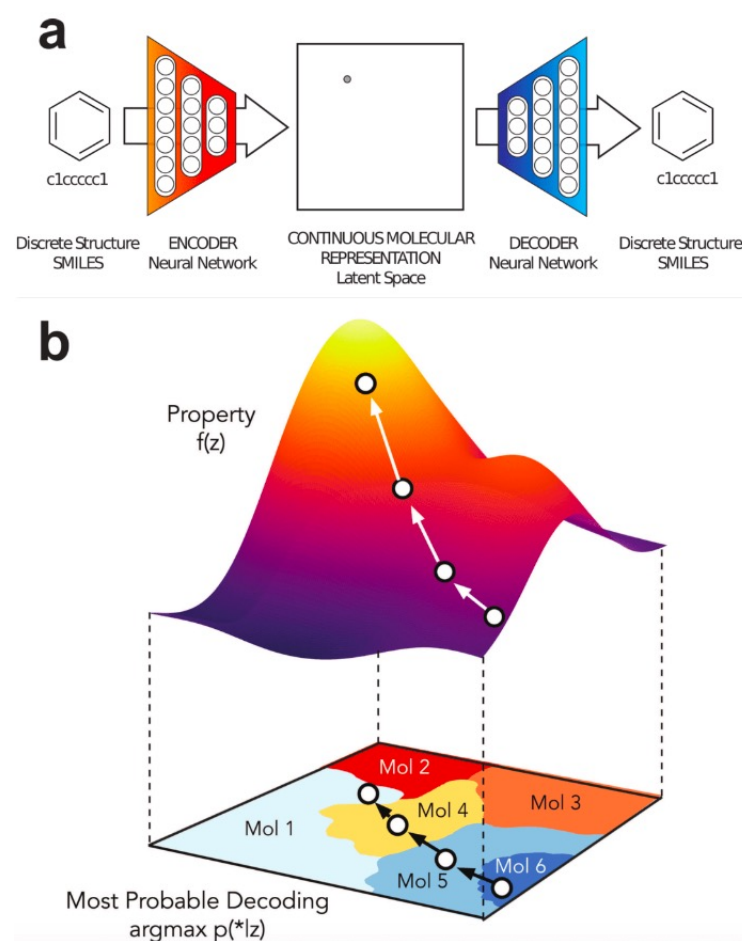
- Recall our Sept 19 lecture on Ethics in AI/ML: **we should always be considering the ethical and legal ramifications of our work**
- **But:** given how widely available and easily accessed tools like ChatGPT are, and the hype surrounding them
- **There's never been a better time to have these conversations**

# Advantages of Generative AI

- Already legion!
- Democratize access to art and figure generation
- Interactive, natural-language interfaces
  - As opposed to arcane tricks and query optimization hacks with traditional search engines
- Revealed clear weaknesses in our assessment protocols
  - Educational assessment (i.e., grading) should not be contingent on whether or not you had access to a chatbot

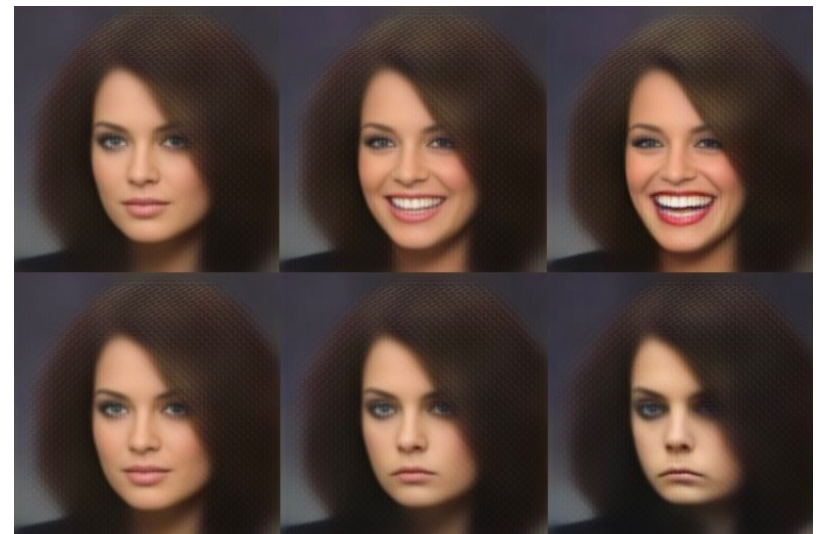
# Advantages of Generative AI

- New scientific discoveries around medicine, biology, chemistry, and biochemistry
- Design new compounds (drugs, antibiotics, treatments) by teaching generative models about known ones
- Keynote speaker at IOB Symposium in October spoke about using LLMs to discover new proteins!



# Advantages of Generative AI

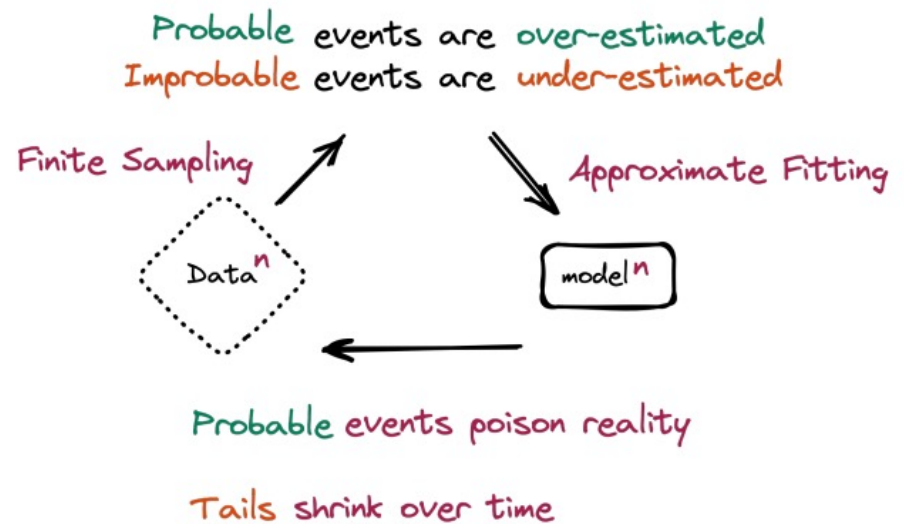
- Accessibility and interactivity
- Original image (top left) interpolated along VAE latent distribution, producing different facial expressions
- Virtual avatars, assistants, video gaming





# Technical issues

- Recursive model training
  - As more information on the internet (images, text) is AI-generated, LLMs will ingest this data as part of their training, creating a recursive training loop
  - “The Curse of Recursion: Training on Generated Data Makes Models Forget”
  - <https://arxiv.org/abs/2305.17493>



# Technical issues

- Examples of recursive model training



(a) Original model



(b) Generation 5



(c) Generation 10



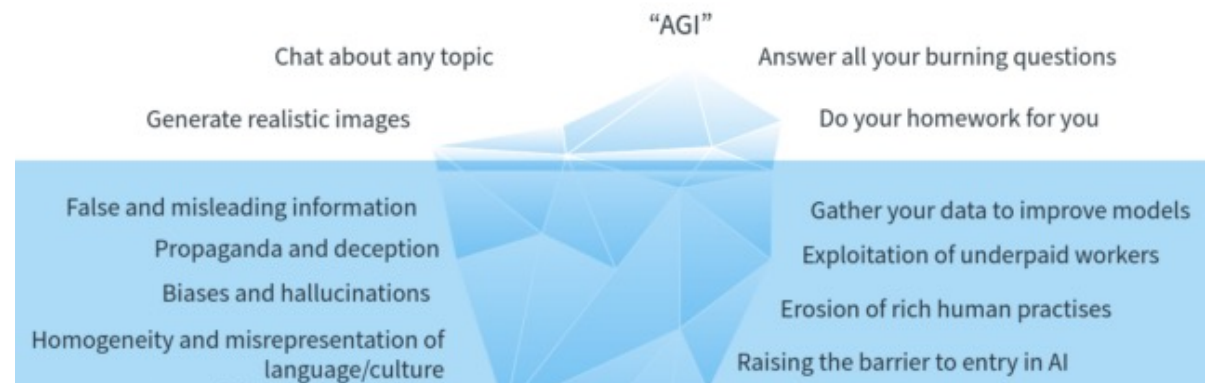
(d) Generation 20

# Legal issues

- Copyright
  - OpenAI, Midjourney most likely training on image datasets **without** permission from authors
  - Currently in the US, AI-generated art cannot be copyrighted → **potential boon for public domain!**
- Plagiarism
  - Simply: if you didn't write/code/create it **yourself**, and you didn't otherwise specify where it came from (and sometimes, even if you did), it's **plagiarism**
  - Is getting the answer from ChatGPT and presenting it as your own any different from getting the answer from your classmate and presenting it as your own?
  - Huge implications in professional fields, given current chatbot accuracy levels

# Ethical and moral issues

- Disinformation
- Enabling/scaling abuse
- Environmental concerns
- Worker exploitation
- Hidden costs of AI



Model name	Number of parameters	Datacenter PUE	Carbon intensity of grid used	Power consumption	CO <sub>2</sub> eq emissions	CO <sub>2</sub> eq emissions × PUE
GPT-3	175B	1.1	429 gCO <sub>2</sub> eq/kWh	1,287 MWh	<i>502 tonnes</i>	<i>552 tonnes</i>
Gopher	280B	1.08	330 gCO <sub>2</sub> eq/kWh	<i>1,066 MWh</i>	<i>352 tonnes</i>	<i>380 tonnes</i>
OPT	175B	<i>1.09</i> <sup>2</sup>	<i>231 gCO<sub>2</sub>eq/kWh</i>	<i>324 MWh</i>	70 tonnes	<i>76.3 tonnes</i> <sup>3</sup>
BLOOM	176B	1.2	57 gCO <sub>2</sub> eq/kWh	433 MWh	25 tonnes	30 tonnes

Table 4: Comparison of carbon emissions between BLOOM and similar LLMs. Numbers in *italics* have been inferred based on data provided in the papers describing the models.

# Philosophical issues

- Novelty
  - Is the content generated from ChatGPT / Midjourney **new**?
- The “tool” analogy
  - Generative AI is inherently neither good nor bad, but dependent on its application
- In 1999, French cultural theorist Paul Virilio wrote, "*When you invent the ship, you also invent the shipwreck; when you invent the plane you also invent the plane crash; and when you invent electricity, you invent electrocution... Every technology carries its own negativity, which is invented at the same time as technical progress.*"

I don't have answers

# Conclusions

- Generative modeling
  - Learn a *distribution* instead of a decision boundary
  - Can still be used for classification
  - Usually requires more data than discriminative models
- Deep generative modeling
  - DBNs, RBMs, Denoising & Variational Autoencoders, GANs, AR models, LD
  - All ways of learning a generating distribution from data in deep neural architectures
- Deployments of generative AI
  - Commercial products (ChatGPT, Stable Diffusion, Midjourney, DALL-E)
  - Possibilities, advantages, moral/ethical/legal/philosophical considerations
  - Consider the possible use-cases

# References

- “Generative Learning algorithms”, CS 229 notes by Andrew Ng <http://cs229.stanford.edu/notes/cs229-notes2.pdf>
- Andrew Ng and Michael I. Jordan (!!!), “On Discriminative vs. Generative classifiers: A comparison of logistic regression and naive Bayes” (NIPS 2002) <http://papers.nips.cc/paper/2020-on-discriminative-vs-generative-classifiers-a-comparison-of-logistic-regression-and-naive-bayes.pdf>
- “Generative versus Discriminative”, Cross-Validated <https://stats.stackexchange.com/questions/12421/generative-vs-discriminative/223850>
- “Generative Adversarial Networks”, NIPS 2016 Tutorial <https://arxiv.org/pdf/1701.00160.pdf>
- Variational Inference <http://www.inference.vc/choice-of-recognition-models-in-vaes-a-regularisation-view/>
- Tutorial on Variational Autoencoders <https://arxiv.org/abs/1606.05908>
- “Progressive Growing of GANs for Improved Quality, Stability, and Variation”, <https://arxiv.org/abs/1710.10196>
- “Generative Adversarial Networks (GANs), Some Open Questions”, <http://www.offconvex.org/2017/03/15/GANs/>
- “GANs are Broken in More than One Way: The Numerics of GANs”, <http://www.inference.vc/my-notes-on-the-numeric-of-gans/>
- “Diffusion Models in Vision: A Survey”, <https://ieeexplore.ieee.org/iel7/34/4359286/10081412.pdf>
- “GPT-4 Technical Report”, <https://arxiv.org/abs/2303.08774>
- “Deep reinforcement learning from human preferences”, [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf)
- “An introduction to variational autoencoders”, <https://www.nowpublishers.com/article/DownloadSummary/MAL-056>
- “Probabilistic Machine Learning, Advanced Topics”, Kevin Murphy. Part IV: Generation. <https://probml.github.io/pml-book/book2.html>



# Administrivia

- Homework 5 due by the end of the day TODAY
- Happy Thanksgiving holiday!
- Final Project Presentations start **next Monday** (and continue through next Tuesday, Thursday, and the following Monday)
  - **Presentations for the day are randomly assigned at the start of each day**
  - 12 minutes + 3 minutes for Q&A
  - Please attend and support your classmates even after you've given your talks—they worked hard, too!
  - If you need a specific time slot contact me ASAP
- Course evaluations—please fill them out!
  - [eval.franklin.uga.edu](http://eval.franklin.uga.edu)



# GAN Advances

- Progressively grow the GAN subspace over training



# GAN Advances

- Wasserstein objective function
  - “Earth-mover” distance  $W(q, p)$
  - Minimum cost of transporting mass in order to transform distribution  $q$  into the distribution  $p$  (where cost is mass x distance)

$$\min_G \max_D \mathbb{E}_{x \sim P_{real}} \log D(x) - \mathbb{E}_z \log D(z)$$

- Gradient is much better behaved than Jensen-Shannon objective (KL-divergence based)
- Weights are clipped at  $[-c, c]$
- Takes a lot longer to train on average

# GAN Advances

- Improved Wasserstein

- Introduces a gradient penalty on the discriminator output with respect to its input
- Instead of hard clipping gradient weights, soft[max] penalties are used
- $P_r$  is the distribution of real data,  $P_g$  is from the generator, and  $P_x$  is defined from sampling uniformly along straight lines between pairs of points sampled from  $P_r$  and  $P_g$

$$\min_G \max_D \mathbb{E}_{\tilde{x} \sim P_g} [D(\tilde{x})] - \mathbb{E}_{x \sim P_r} [D(x)] + \lambda \mathbb{E}_{\hat{x} \sim P_x} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$$

Original\*\* WGAN  
objective

Two-sided gradient  
penalty on Discriminator

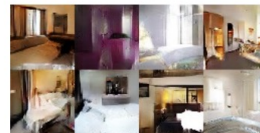
**DCGAN**

**LSGAN**

**WGAN (clipping)**

**WGAN-GP (ours)**

Baseline ( $G$ : DCGAN,  $D$ : DCGAN)



$G$ : No BN and a constant number of filters,  $D$ : DCGAN



$\tanh$  nonlinearities everywhere in  $G$  and  $D$



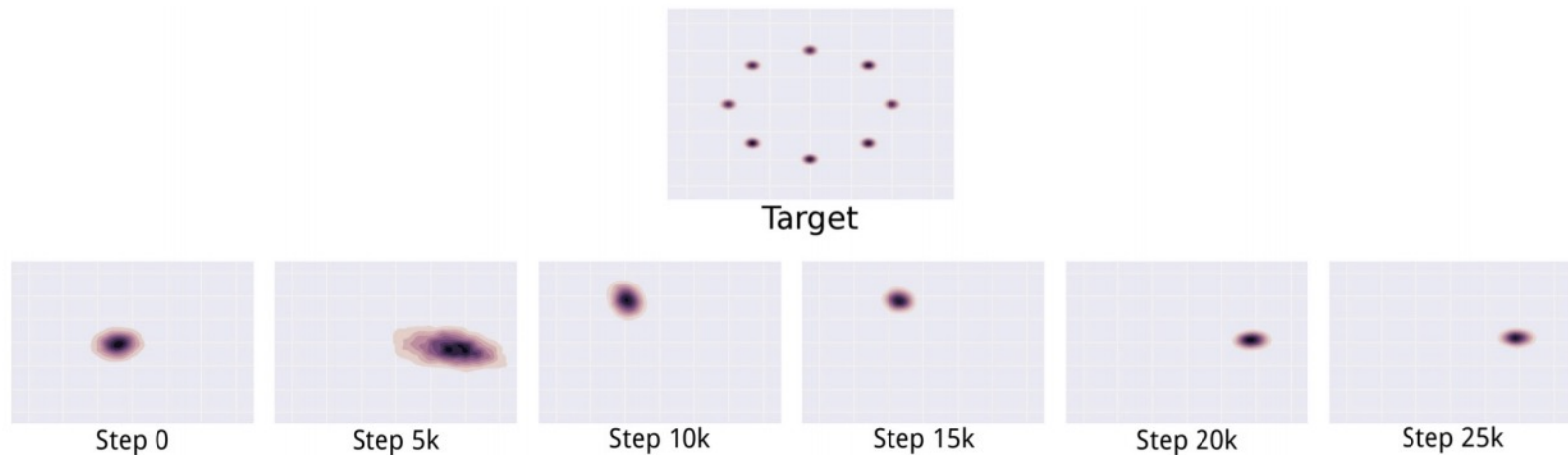
# Open Questions with GANs

- **Mode collapse**

- The “Helvetica Scenario”
- Maps several different inputs  $z$  to the same output
- Full collapse is rare, but partial collapse is common

$$G^* = \min_G \max_D V(G, D),$$

$$G^* = \max_D \min_G V(G, D),$$



# Open Questions with GANs

- **Evaluation of GANs**
- (not specific to GANs *per se*, but generative models)
- Models that obtain good likelihoods can generate bad samples
- Models that generate good samples can have poor likelihoods
- Also difficult to evaluate likelihood with GANs



# Open Questions with GANs

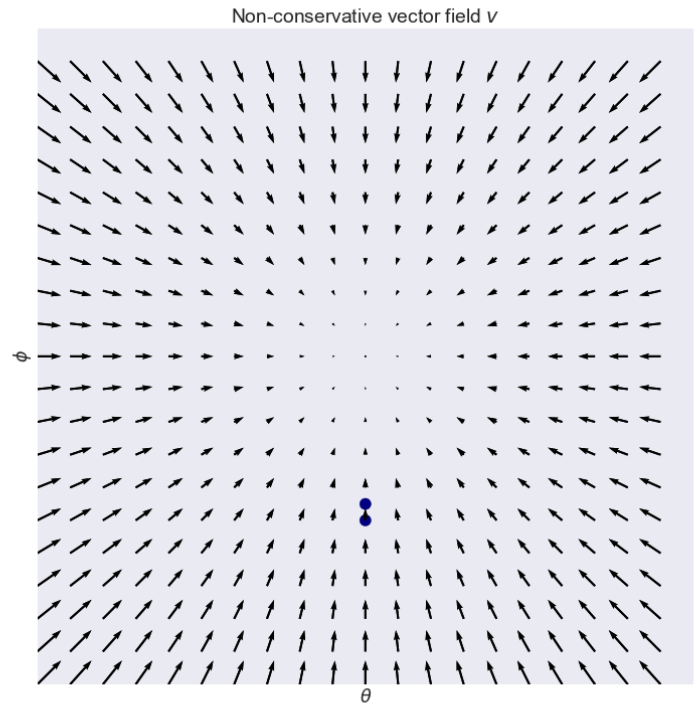
- Goal is to find *Nash equilibrium*
- **Problem 1: Does it exist?**
  - No conclusive way to show this
- **Problem 2: If it exists, can we find it?**
  - Inability to find equilibrium can be cause of oscillatory behavior in training
  - ...or a sign that equilibrium doesn't exist?
- **Problem 3: More than finding equilibrium, can generator *win*?**
  - Intuitively: to learn as representative a generator as possible, discriminator should be *utterly unable to differentiate between real and fake*

# Training a GAN

- GANs use a variant of SGD called *simultaneous gradient descent*
  - Key difference: the latter gives rise to *non-conservative vector fields*
  - Like Escher's staircase
- Solution: convert to conservative vector field

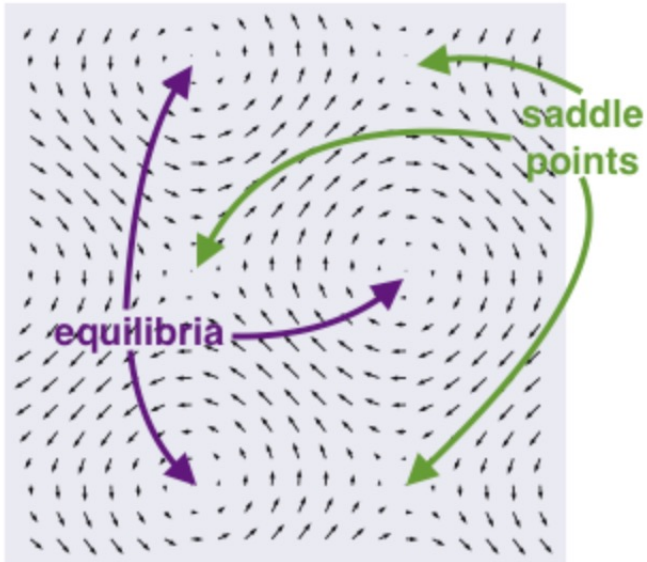
$$-\nabla L(x) := -\frac{\partial}{\partial x} \|v(x)\|_2^2$$

- **New problem: can't differentiate between saddle points or equilibria, or negative or positive equilibria**

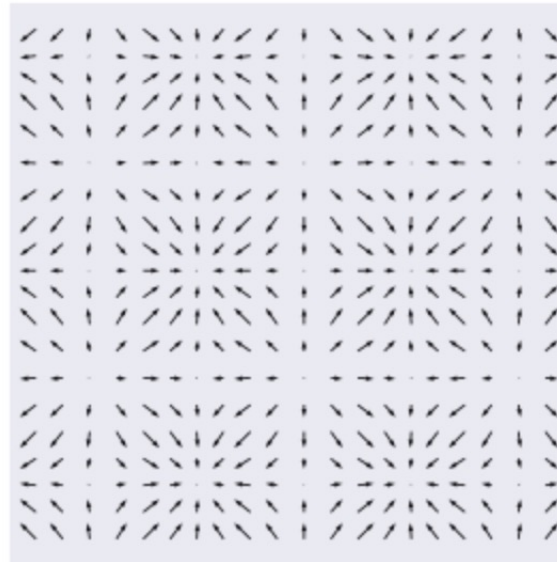


# Training a GAN

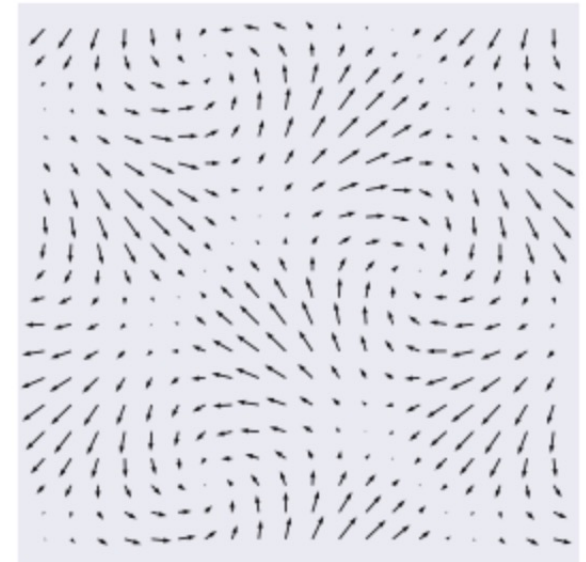
Non-conservative field  $v$



Conservative field  $-\nabla L$



Combined field  $v - 0.6\nabla L$



# Open Questions with GANs

- **Exploration of the learned manifold**
- Advantage of GANs: no *a priori* assumptions on the underlying form of the generating distribution
- Disadvantage of GANs: no way to meaningfully interpret the resulting learned generating distribution
- Manifold walking, interpolation, image algebra, INFO-GANs