Introduction to Deep Generative Modeling

Lecture #1

HY-673 – Computer Science Dept, University of Crete

Professors: Yannis Pantazis & Yannis Stylianou

TAs: Michail Raptakis

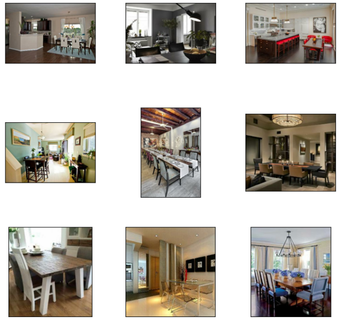
What is this course about?

✓ Statistical Generative Models

- ✓ A Generative Model (GM) is defined as a **probability distribution**, $\boldsymbol{p(x)}$.
  - ✓ A statistical GM is a *trainable probabilistic* model, $\boldsymbol{p_\theta(x)}$.
  - ✓ A deep GM is a statistical generative model parametrized by a *neural network*.
  - ✓ $\boldsymbol{p(x)}$ and in many cases $\boldsymbol{p_\theta(x)}$ are not analytically known. Only samples are available!
- ✓ **Data** $(\boldsymbol{x})$**:** complex, (un)structured samples (e.g., images, speech, molecules, text, etc.)
- ✓ **Prior knowledge:** parametric form (e.g., Gaussian, mixture, softmax), neural architecture, loss function (e.g., maximum likelihood, divergence), optimization algorithm, invariance/equivariance, laws of physics, prior distribution, etc.

What is this course about?

✓ A dataset with images e.g., of bedrooms (LSUN dataset)



data distribution

$p(x)$ or $p_{data}(x)$ or $p_d(x)$

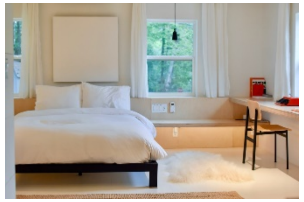
GM's distribution

$p_\theta(x)$ or $p_g(x)$

✓ ***<u>Goal:</u>*** Find $\underline{\theta \in \Theta}$ such that $\underline{p_\theta(x) \approx p_d(x)}$.

✓ It is *<u>generative</u>* because *sampling from $p_\theta(x)$ generates* new *<u>unseen</u>* images.

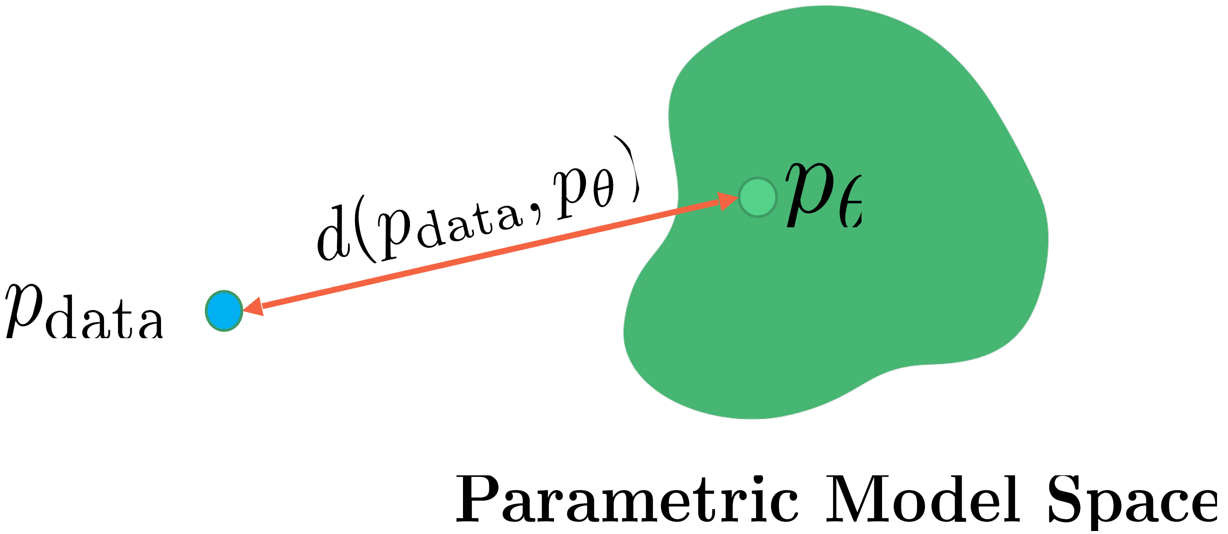
…



$\sim \quad p_\theta(x)$

What is this course about?

$$x_i \sim p_{\text{data}}$$
$$i = 1, 2, \ldots, n$$



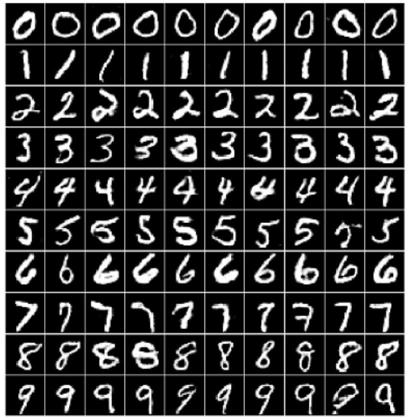
We will study:

✓ Families of Generative Models
✓ Algorithms to train these GMs
✓ Network architectures
✓ Loss functions & distances between probability density functions

What is this course about?

✓ Conditional Generative Models

- ✓ A conditional GM is defined as a **conditional probability distribution**, $p(x \mid y)$.
- ✓ y: conditioning variable(s) (e.g., label/class, text, captions, speaker id, style, rotation, thickness, …)



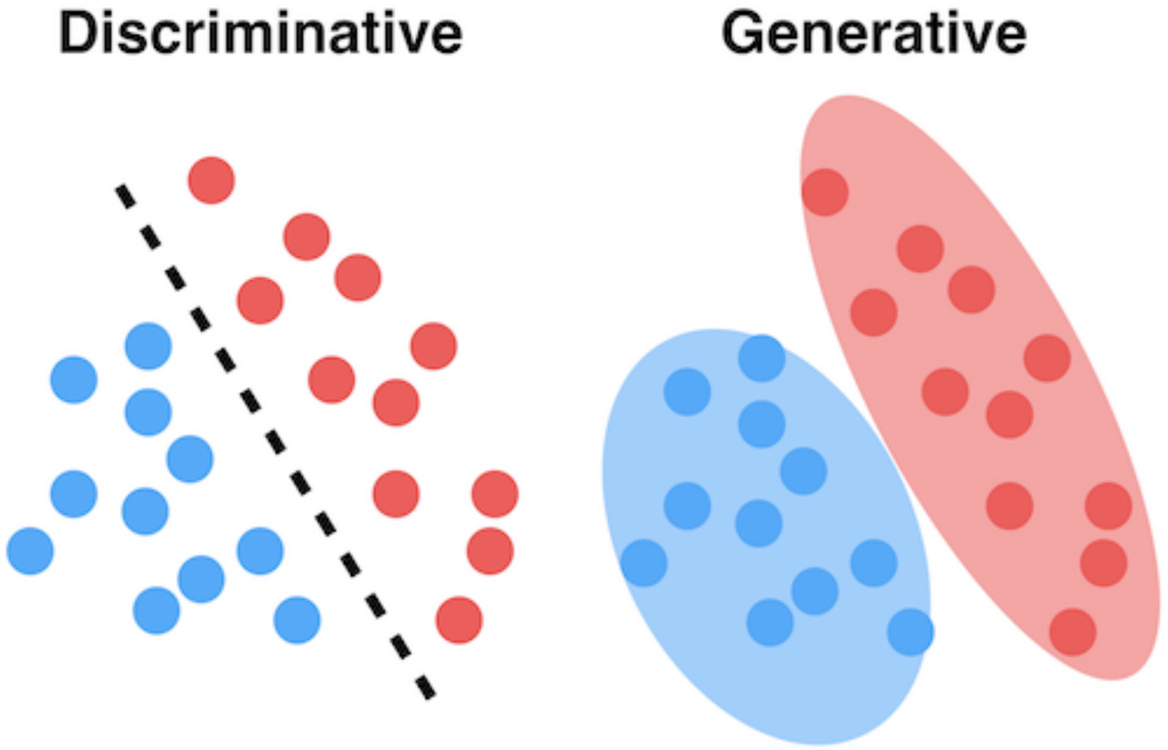
$\sim \quad p_\theta(x \mid y), \ y$:*digit label*

Discriminative vs Generative Models

Data: x

Label: y

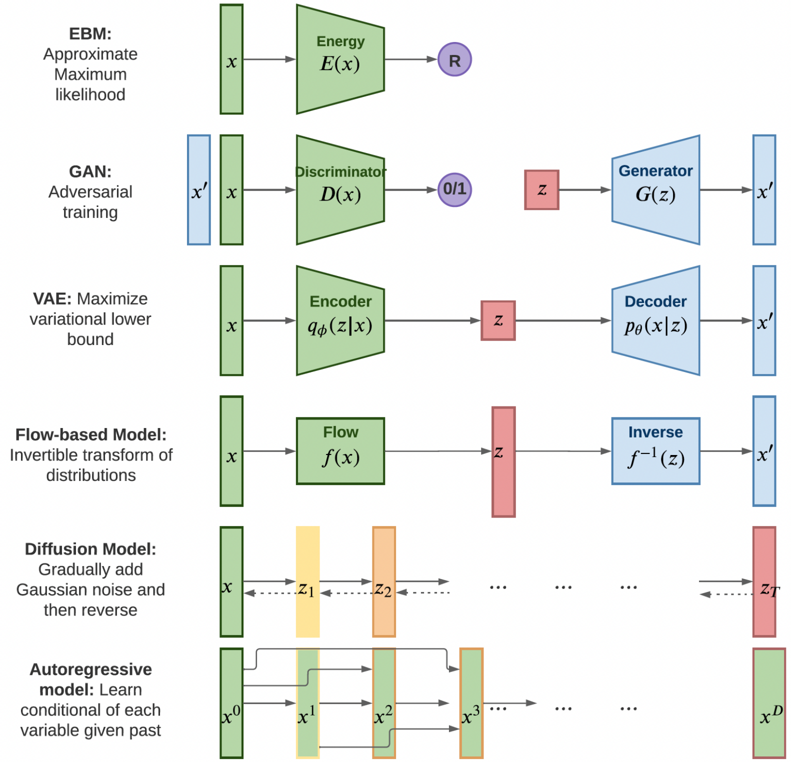
"Cat"



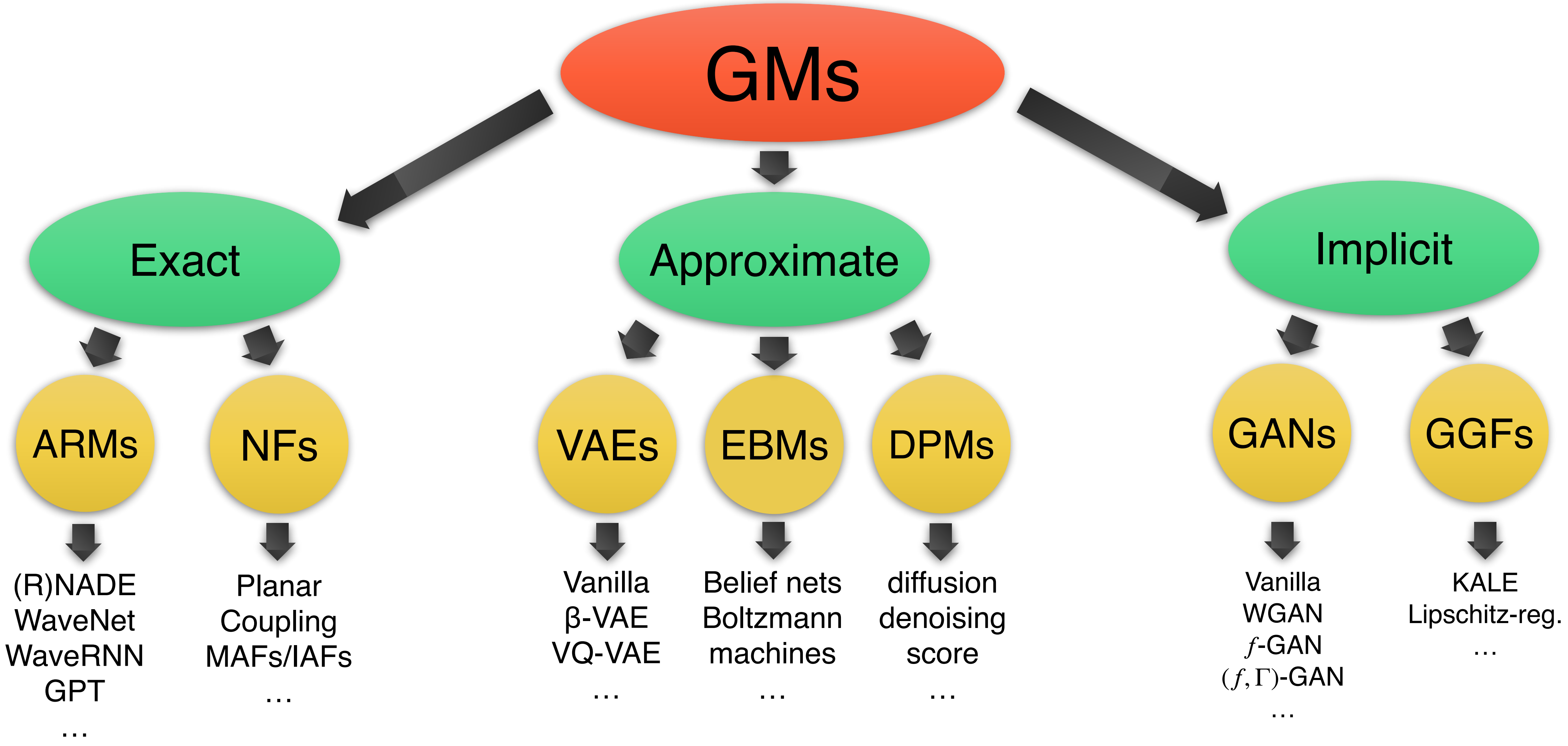
- ✓ **Discriminative Model**
  - ✓ Learn the probability distribution $p(y|x)$
- ✓ **Generative Model**
  - ✓ Learn the probability distribution $p(x)$
- ✓ **Conditional GM**
  - ✓ Learn $p(x|y)$

Families of Generative Models

- ✓Energy-based Models (EBMs)
- ✓Generative Adversarial Nets (GANs)
- ✓Variational Auto-Encoders (VAEs)
- ✓Normalizing Flows (NFs)
- ✓Diffusion Probabilistic Models (DPMs)
- ✓Deep Autoregressive Models (ARMs)



Families of Generative Models



Less known Families of GMs

✓Generative Stochastic Networks (GSNs)

✓Generative Gradient Flows (GGFs)

✓Flow Matching

✓Specific EBMs

- ✓Deep Belief Networks
- ✓Deep Boltzmann Machines
- …

✓Generative Flow Networks (GFlowNets)

...

Progress in Image Generation

GAN Era

✓ Face generation: Rapid progress in image quality

Progress in Image Generation

Diffusion Models Era

✓ Emerging as very powerful alternative generative model family

Image Super Resolution

- ✓ Several inverse problems can be solved with ***conditional GMs***.
- ✓ *Inverse problems:* From measurements, calculate/infer the causes.

- ✓ $P(\textit{high resolution} \mid \textit{low resolution})$
- ✓ *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network* - Ledig et al. - CVPR 2017
  - ✓ https://openaccess.thecvf.com/content_cvpr_2017/html/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.html

Image Inpainting

- ✓ $P(\textit{full image} \mid \textit{masked image})$
- ✓ *<u>DeepFill (v2):</u> Free-Form Image Inpainting With Gated Convolution* – Yu et al. - ICCV 2019
  - ✓ https://openaccess.thecvf.com/content_ICCV_2019/html/Yu_Free-Form_Image_Inpainting_With_Gated_Convolution_ICCV_2019_paper.html

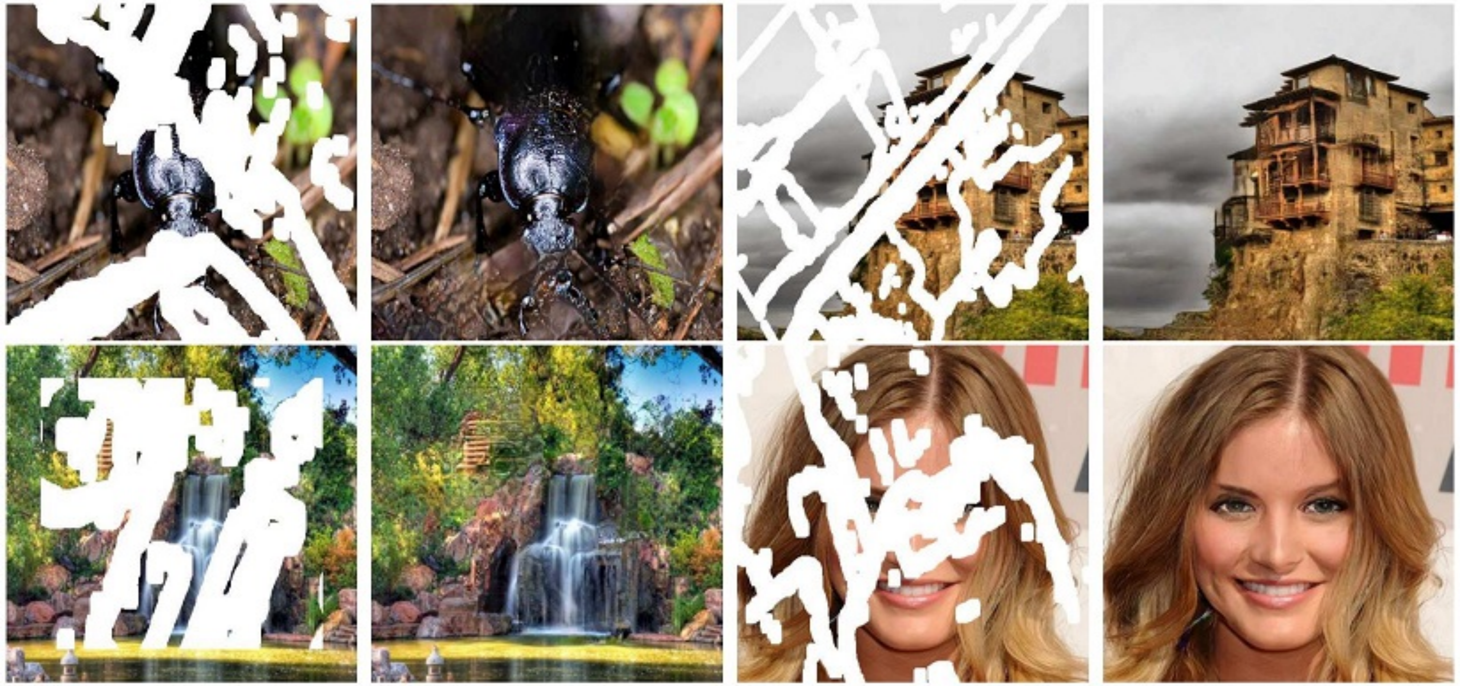


Image Colorization

- ✓ $P(\textit{colored image} \mid \textit{grayscale image})$

- ✓ *PalGAN: Image Colorization with Palette Generative Adversarial Networks* – Wang et al. - ECCV 2022
  - ✓ https://link.springer.com/chapter/10.1007/978-3-031-19784-0_16

Text2Image Translation

- ✓ *Recent advancements:*
  - ✓DALL-E 2
  - ✓Stable Diffusion
  - ✓Imagen
  - ✓GLIDE
  - ✓Midjourney

- ✓ $\boldsymbol{P(image|text)}$

Théâtre D’opéra Spatial by Jason Allen and Midjourney

OpenAI's DALL-E 2

✓Text → Text embedding → Image embedding → low resolution image → medium resolution image → high resolution image

✓ $P(\textit{high res image} \mid \textit{text caption}) = P(\textit{image emb} \mid \textit{text caption})$ x $P(\textit{high res image} \mid \textit{image emb})$

✓*Hierarchical Text-Conditional Image Generation with CLIP Latents* - Ramesh et al. - 2022

- ✓https://cdn.openai.com/papers/dall-e-2.pdf

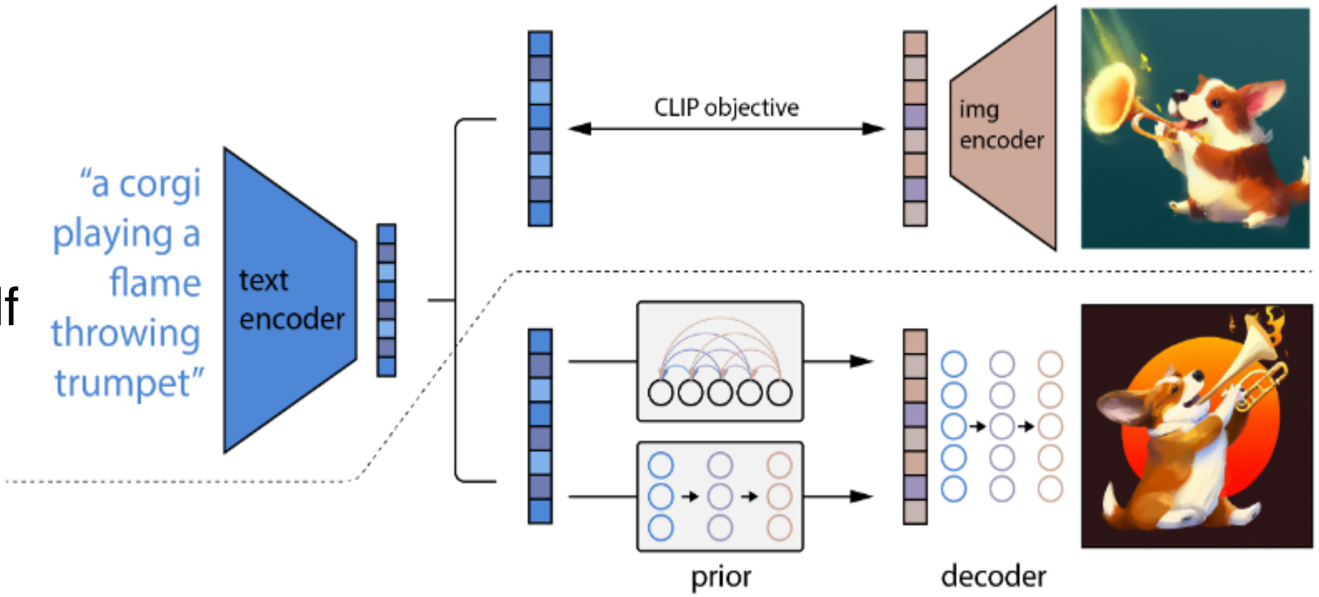


Image2Image Translation

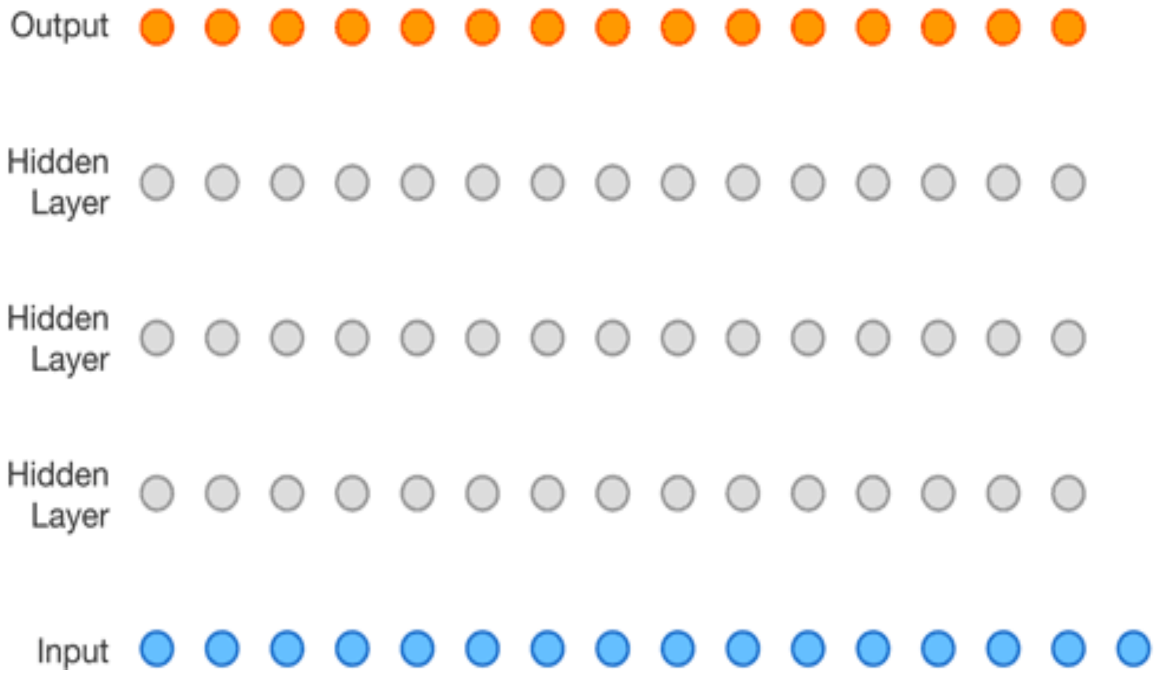
- ✓ *Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks* (CycleGAN) – Zhu et al. - ICCV 2017
  - ✓ https://openaccess.thecvf.com/content_iccv_2017/html/Zhu_Unpaired_Image-To-Image_Translation_ICCV_2017_paper.html

Speech & Audio Synthesis

Lecture #1

✓ $P(x_{t+1} | x_t, x_{t-1}, \ldots, text)$

✓ WaveNet, WaveRNN, Parallel Wavenet, MelGAN, WaveDiff, …



Text to Speech Synthesis

- Parametric
- Concatenative
- WaveNet

Unconditional

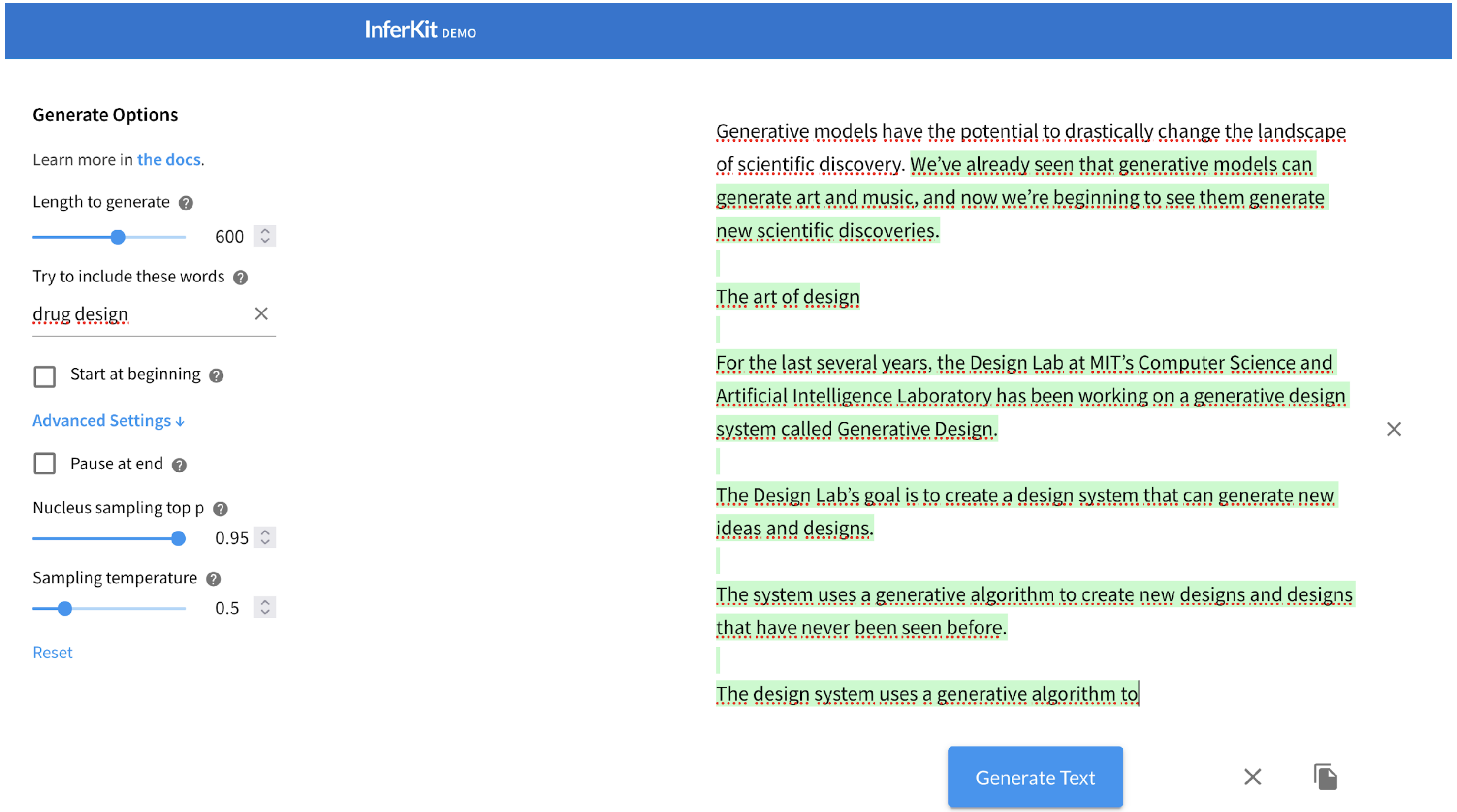
Music

van den Oord et al., 2016

(Natural) Language Generation

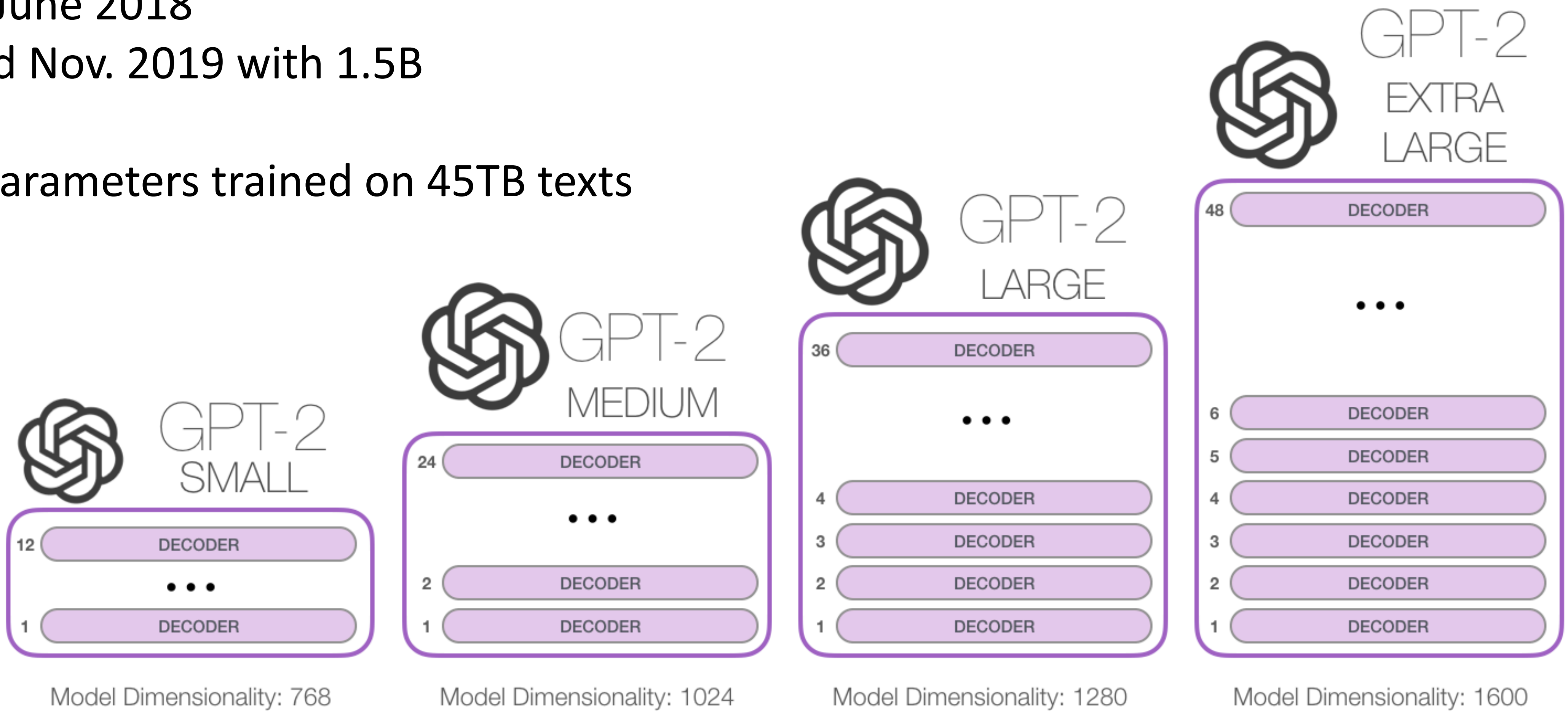
Lecture #1

- ✓ $P(next\ word \mid previous\ words)$
- ✓ https://app.inferkit.com/demo
- ✓ **GPT-3**
  - ✓ **Generative Pre-trained Transformer**
  - ✓ https://deepai.org/machine-learning-model/text-generator



(Natural) Language Generation

GPT released June 2018
GPT-2 released Nov. 2019 with 1.5B parameters
GPT-3: 175B parameters trained on 45TB texts

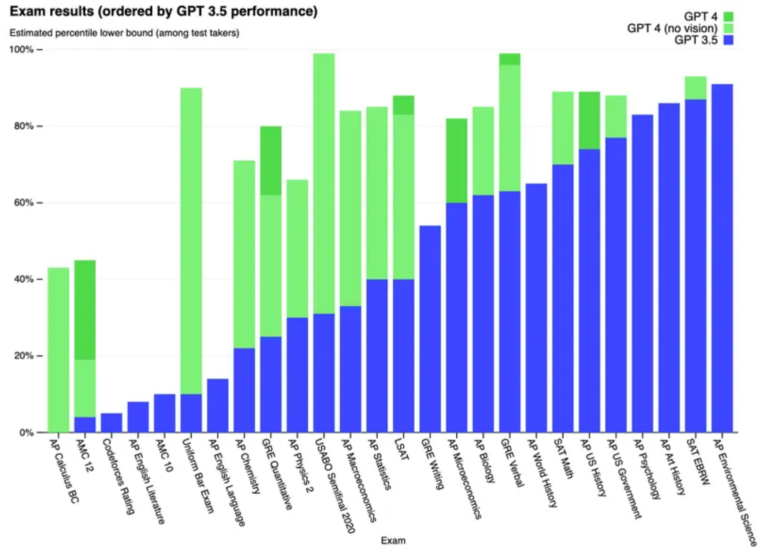


117M parameters 345M 762M 1542M

(Natural) Language Generation

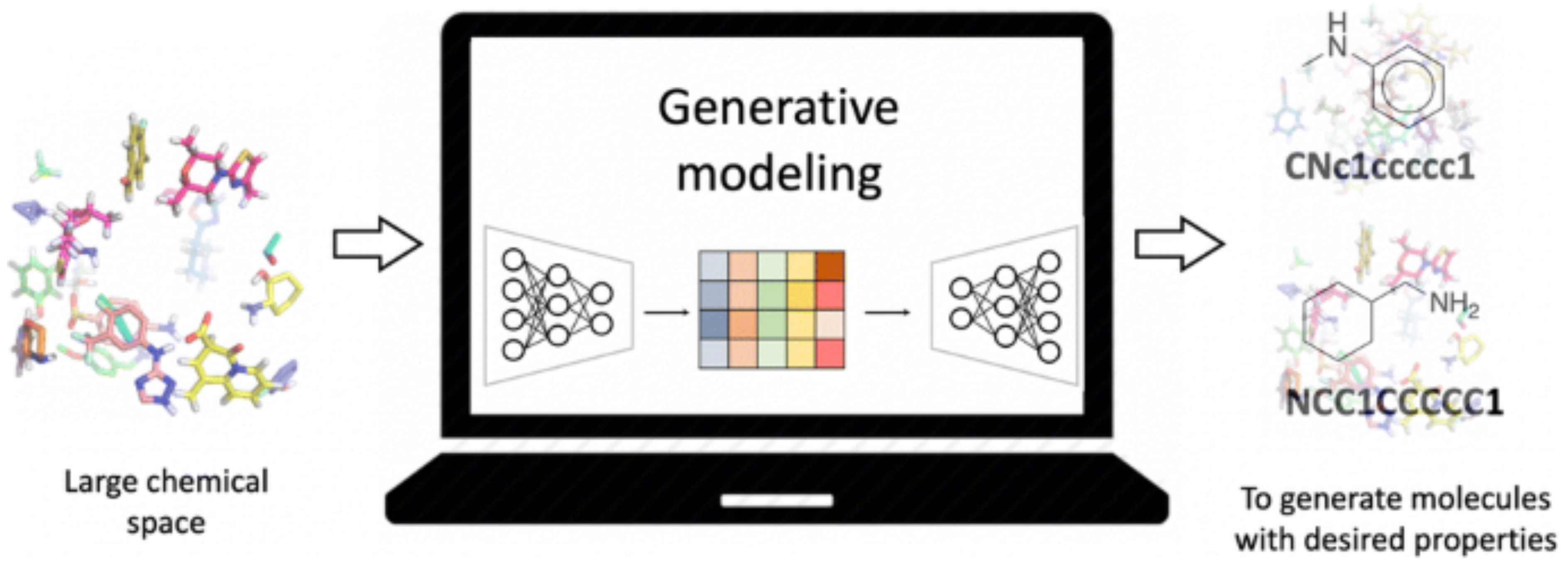
- ✓Enormous model size (Trillion parameters?)
- ✓Enormous & diverse training data
- ✓In Context learning (a.k.a. prompting)
- ✓Reinforcement learning (with Human in the loop)
- ✓Multimodal capabilities
- ✓Reasoning capacity
- ✓Enormous performance
  - ✓Coherence, relevance, proficiency
  - ✓Agentic AI
  - ✓Safety & Ethics
  - ✓Few steps from AGI

OpenAI ChatGPT 4.0



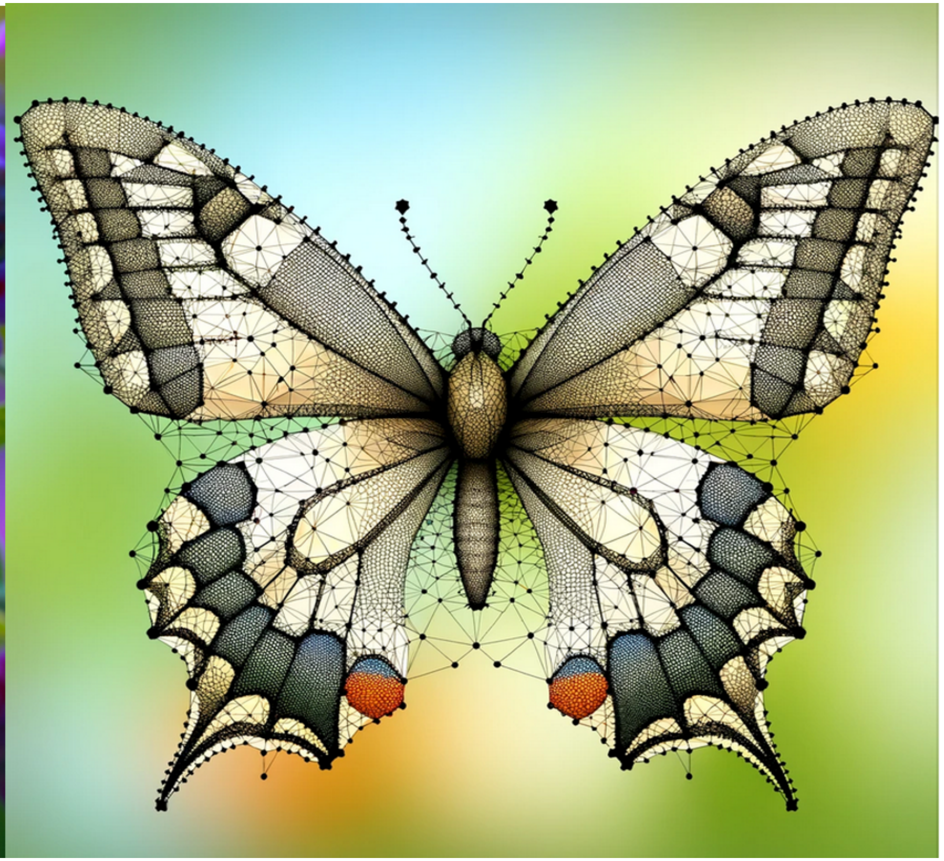
Molecule/Drug/Protein Design

✓ *MolGAN: An implicit generative model for small molecular graphs* – De Cao & Kipf – ICML 2018



Geometric Design

✓Just meshing around with GPT4 (ESA proposal)

Driving forces in GM progress

- ✓ <u>Representation learning</u>
  - ✓ Leveraging the exponential growth of data & of model's parameters via self-supervised learning
  - ✓ Gave raise to the so-called Foundation Models
- ✓ <u>Computational resources</u> are also exponentially increasing
  - ✓ Gave raise to Scaling Laws
- ✓ <u>Better understanding</u> of the models, algorithms act as key enablers
- ✓ <u>Unlocks</u> *human* productivity & *creativity*.
- ✓ Ideally, it will <u>accelerate</u> the *scientific discovery process*.

Challenges in GMs

- ✓ ***Representation:*** How do we model the joint distribution of many random variables?
  - ✓ Need compact & meaningful representations
- ✓ ***Learning (a.k.a. quality assessment):*** What is the proper comparison metrics between probability distributions?
- ✓ ***Reliability:*** Can we trust the generated outcomes? Are they consistent?
- ✓ ***Alignment:*** Do they perform according to the input of the user?

Prerequisites

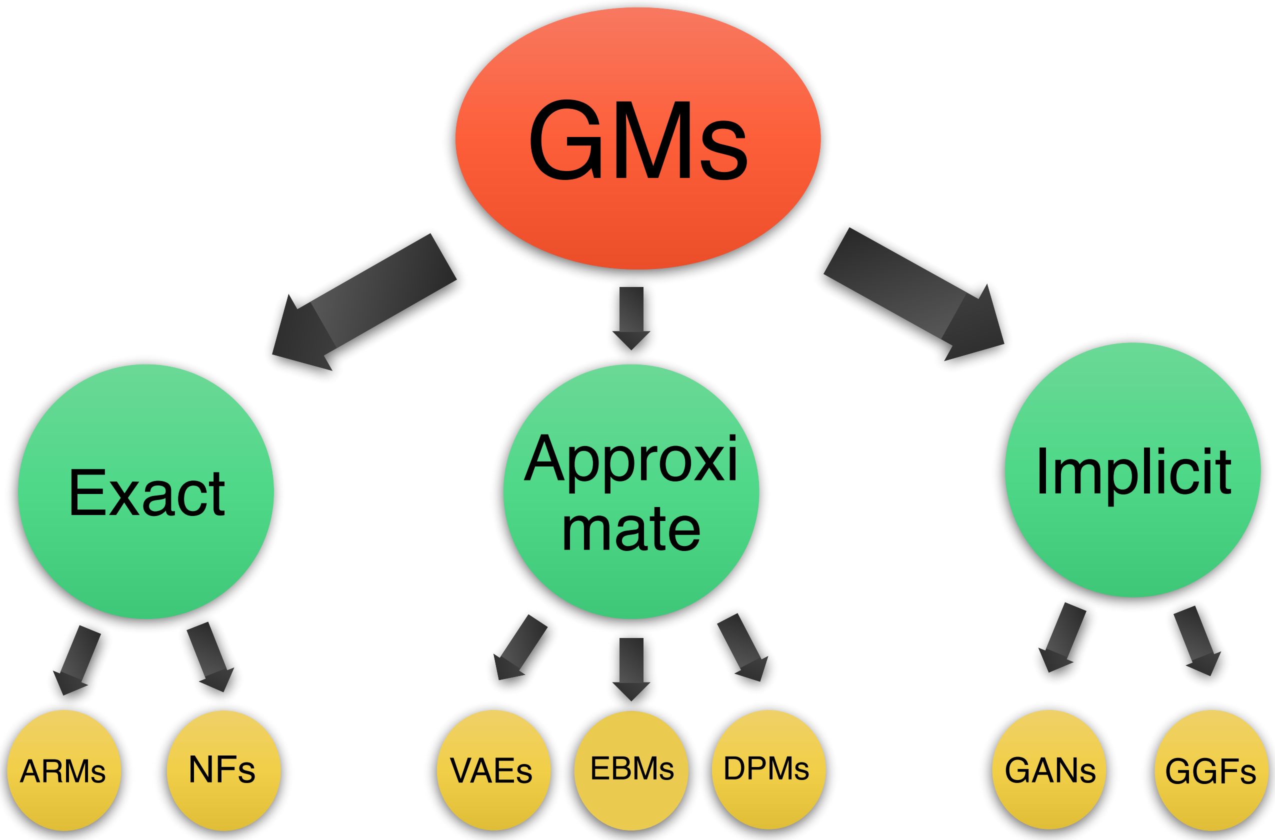
✓<u>Very good knowledge</u> of *probability theory, multivariate calculus & linear algebra*.

✓<u>Intermediate knowledge</u> regarding *machine learning & neural networks*.

✓<u>Proficiency</u> in some *programming language*, preferable <u>*Python*</u>, is required.

Course Syllabus

- ✓ Basics in probability theory (1W)
    - ✓ Shallow generative models - GMMs (1W)
- ✓ Exact (i.e., fully-observed) likelihood
    - ✓ AR models (2W)
    - ✓ Normalizing flows (1W)
- ✓ Approximate likelihood
    - ✓ VAEs (2W)
    - ✓ Diffusion/Score-based models (2W)
- ✓ Implicit
    - ✓ GANs (2W)
- ✓ Recap (1W)



Logistics

- ✓ *Teaching Assistant:* Michail Raptakis (PhD candidate)
- ✓ *Weekly Tutorial (Friday 11:00-13:00):* Python/PyTorch basics, neural network architectures and training, solve problems to assist with homework, solve selected homework's problems
- ✓ *Textbooks*
  - ✓ ***Probabilistic Machine Learning: Advanced Topics*** by Kevin P. Murphy
  - ✓ *https://probml.github.io/pml-book/book2.html*
  - ✓ ***Deep Learning: Foundations and Concepts*** by Christofer M. Bishop
- ✓ Seminal papers will be distributed

Grading policy

- ✓ ***No** Final Exam*
- ✓ 6-8 series of *Homework* (**60%** of total grade)
  - ✓ Mix of theoretical and programming problems and challenges
  - ✓ Equally weighted
- ✓ *Project*: implementation & presentation (**40%** of total grade)
  - ✓ Implementation: **20%**
  - ✓ Final report: **10%**
  - ✓ Presentation: **10%**

Project

- ✓Select from a given list of papers or propose a paper (which has to be approved) or open research questions/challenges
- ✓*Categories of papers or an open research question*:
  - ✓Application of deep generative models on a novel task/dataset
  - ✓Algorithmic improvements into the learning, inference and/or evaluation of deep generative models
  - ✓Theoretical analysis of any aspect of existing deep generative models
- ✓Groups of **up to 2 students** per project
- ✓Computational resources might be provided (colab, local GPUs, etc.)

Introduction to Deep Generative Modeling

Lecture #1

HY-673 – Computer Science Dept, University of Crete

Professors: Yannis Pantazis & Yannis Stylianou

TAs: Michail Raptakis