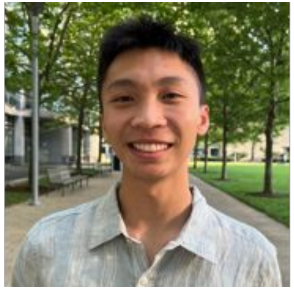
CS294-158 Deep Unsupervised Learning

Lecture 1 Intro: Logistics and Motivation

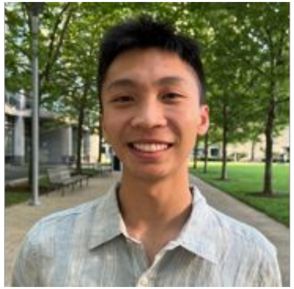





Pieter Abbeel, Wilson Yan, Kevin Frans, Philipp Wu

Instructor Team






Pieter Abbeel, Wilson Yan, Kevin Frans, Philipp Wu

Communication

- **Website**: https://sites.google.com/view/berkeley-cs294-158-sp24/home

- **Announcements**
  - Ed -- sign up today! – if you are registered or on the waitlist, we already added you

- **Questions**
  - Ed (preferred!)
  - cs294-158-staff@lists.berkeley.edu

- **Office hours**: [starting ***next*** week]
  - Pieter: Thu 5-6pm -- 250 Sutardja Dai Hall
  - Wilson: TBD
  - Kevin: TBD
  - Philipp: TBD

For homework, TA office hours are the best venue.
For other questions (lecture, final project, research, etc.) any office hours should be great fits

Admission into the Course

- Not everyone is going to make it in given class size constraints

- We'll remove students who don't have a strong HW1, which will open up slots
  - And if you are hoping to make it, make sure to submit a strong HW1!

- Everyone welcome to audit / submit HW

Syllabus

Week 1 (1/18) **Intro**

Week 2 (1/25) **Autoregressive Models**

Week 3 (2/1) **Flow Models**

Week 4 (2/8) **Latent Variable Models**

Week 5 (2/15) **Generative Adversarial Networks / Implicit Models**

Week 6 (2/22) **Diffusion Models + Final Project Discussion**

Week 7 (2/29) **Self-Supervised Learning / Non-Generative Representation Learning**

Week 8 (3/7) **Strengths and Weaknesses of Unsupervised Learning Methods**

Week 9 (3/14) **Semi-Supervised Learning; Unsupervised Distribution Alignment**

Week 10 (3/21)**Compression**

Spring Break Week (no lecture)

Week 11 (4/4) **Language Models**

Week 12 (4/11) **Midterm; Multimodal Models; Video Generation ** different location: 310 SDH****

Week 13A (4/18) **AI for Science**

Week 13B (4/19) **Representation Learning in Reinforcement Learning**

Week 14 no lecture (we'll have two lectures in Week 13 instead)

Week 15 no lecture (RRR week)

Week 16 (5/10) **Final Project Reports and Video Presentation Submissions Due**

Homework

- HW1: Autoregressive Models (out 1/25, due 2/7)
- HW2: Latent Variable Models (out 2/8, due 2/21)
- HW3: GANs / Implicit Models (out 2/22, due 3/6)
- HW4: Diffusion Models (out 3/7, due 3/21)

Homework Policy

- **Collaboration: *Students may discuss assignments. However, each student must code up and write up their solutions independently*.**

- **Late assignments:** Recognizing that students may face unusual circumstances and require some flexibility in the course of the semester, each student will have ***a total of 7 free late (calendar) days*** to use as s/he sees fit, but ***no more than 4 late days can be used on any single assignment***. Late days are counted at the granularity of days: e.g., 3 hours late is one late day.

Midterm

- Date: 4/11 (during lecture slot)
- Topics: everything covered through (and including) 4/4
- Format: we will provide a document with questions and answers ahead of time (~20)
- Rationale: opportunity to force yourself to fully internalize key derivations and concepts

Final Project

- **SCOPE:**
  - Goal: explore and push the boundaries in unsupervised learning.
  - E.g. proposal+evaluation of new algorithms / architectures, investigation of an application of unsupervised learning, benchmarking unsupervised learning, compression, studying synergies between unsupervised learning and other types of learning, etc.
  - Ideally, the project covers interesting new ground and might be the foundation for a future conference paper submission or product.

- **PROJECT TOPICS / STAFF INPUT:**
  - We encourage trying to come up with your own project idea. We are also happy to make suggestions and/or brainstorm ideas together.
  - One of the main reasons we are so excited to teach this class is to see more Deep Unsupervised Learning projects happen. We are very excited to advise on your projects, please don't hesitate to come to office hours to discuss project ideas, project progress, ideas for next steps, etc.

Final Project -- Timeline

- **Feb 28 Project Proposals Due:** 1 page description of project + goals for milestone. -- Submission through google doc shared with instructors, so we can give feedback/suggestions most easily.

- **March 8 Approved Project Proposals Due:** by this time your proposals should have incorporated instructor feedback, at this stage it should be assured that your proposal is of right fit and scope

- **April 5 3-Page Milestone Due:** This is to make sure you are indeed making progress on the project and an opportunity to get feedback on your progress thus far, as well as on any revisions you might want to propose to your project goals. Expectation is that you report on some initial experimental findings (or if you are doing something purely theoretical, some initial progress on that front). -- Submission through google doc shared with instructors, so we can give feedback/suggestions most easily.

- **May 10 Report and Video Presentation Submission deadline**

Grading Logistics

- 60% Homework
- 10% Midterm
- 30% Final Project

Do we need to attend class?

- No hard requirement
- BUT: very highly recommended
  - Great opportunity to get to know other students at Berkeley embarking on Deep Unsupervised Learning

WARNING

Third offering of this course

There will be some rough edges, please bear with us
+ give feedback!

What is Deep Unsupervised Learning?

- Capturing rich patterns in raw data with deep networks in a **label-free** way
  - Generative Models: recreate raw data distribution
  - Self-supervised Learning: “puzzle” tasks that require semantic understanding

- Why do we care?

Geoffrey Hinton

(in his 2014 AMA on Reddit)

“The brain has about 10^14 synapses and we only live for about 10^9 seconds. So we have a lot more parameters than data. This motivates the idea that we must do a lot of unsupervised learning since the perceptual input (including proprioception) is the only place we can get 10^5 dimensions of constraint per second.”

Yann LeCun

Need tremendous amount of information to build machines that have common sense and generalize

[LeCun-20161205-NeurIPS-keynote]

- "Pure" Reinforcement Learning (cherry)
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- Supervised Learning (icing)
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- Unsupervised/Predictive Learning (cake)
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

LeCake

- (Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)

“Ideal Intelligence”

“Ideal Intelligence” is all about compression (finding all patterns)

- Finding all patterns = short description of raw data (low Kolmogorov Complexity)
- Shortest code-length = optimal inference (Solomonoff Induction)
- Extensible to optimal action making agents (AIXI)

- Assume we pretrain unsupervised on Data Distribution D1 and then finetune on Data Distribution D2

  Then: if D1 and D2 are related, compressing D2 conditioned on D1 should be more efficient than compressing D2 outright

  Hence: pretraining on D1 should aid faster learning of D2

Aside from theoretical interests

- Deep Unsupervised Learning has many powerful applications
  - Generate novel data
  - Conditional Synthesis Technology (WaveNet, GAN-pix2pix)
  - Compression
  - Improve any downstream task with un(self)supervised pre-training
    - Production level impact: Google Search powered by BERT
  - Flexible building blocks

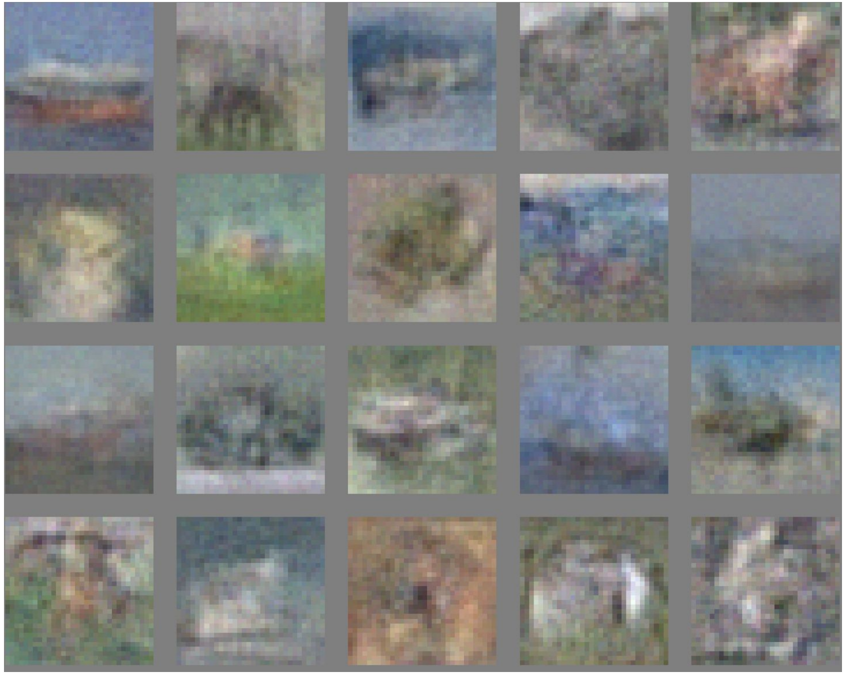
Generate Images

[Deep Belief Nets, Hinton, Osindero, Teh, 2006]

Generate Images

[VAE, Kingma and Welling, 2013]

Generate Images



[GAN, Goodfellow et al. 2014]

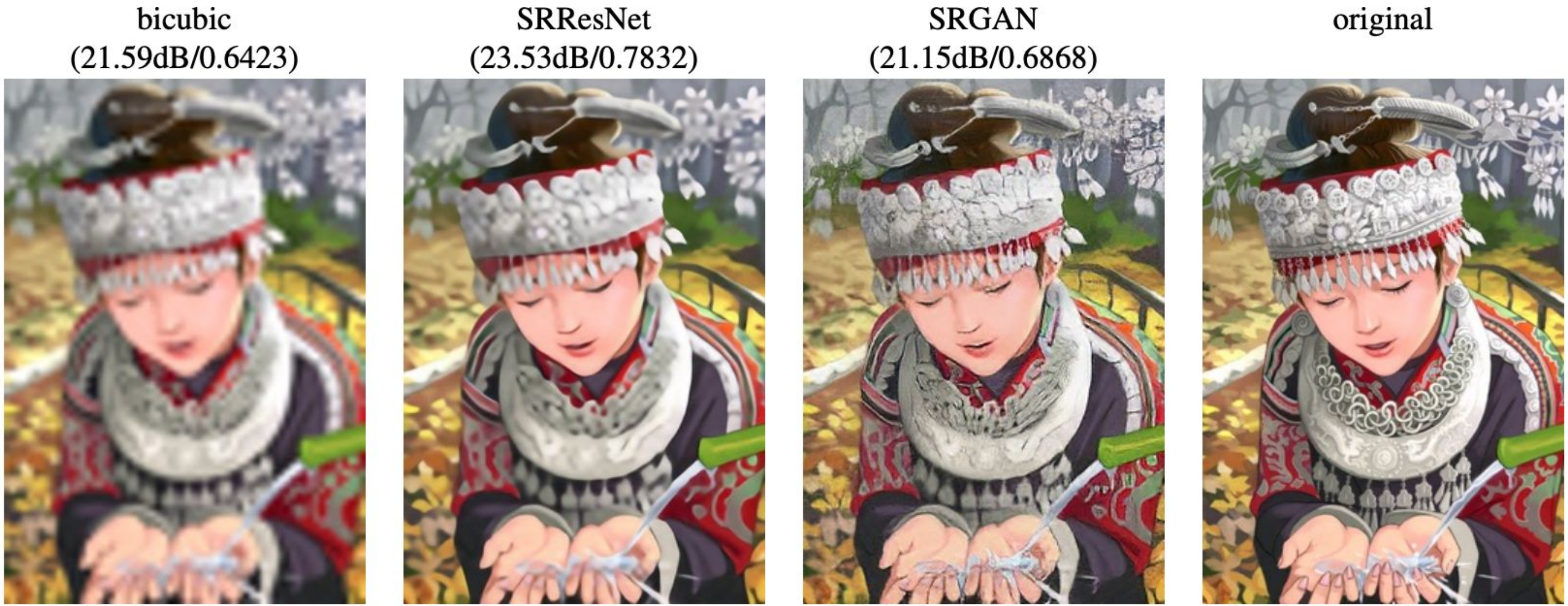
Generate Images

[DCGAN, Radford, Metz, Chintala 2015]

Generate Images

[DCGAN, Radford, Metz, Chintala 2015]

Generate Images



[Ledig, Theis, Huszar et al, 2017]

Generate Images

[CycleGAN: Zhu, Park, Isola & Efros, 2017]

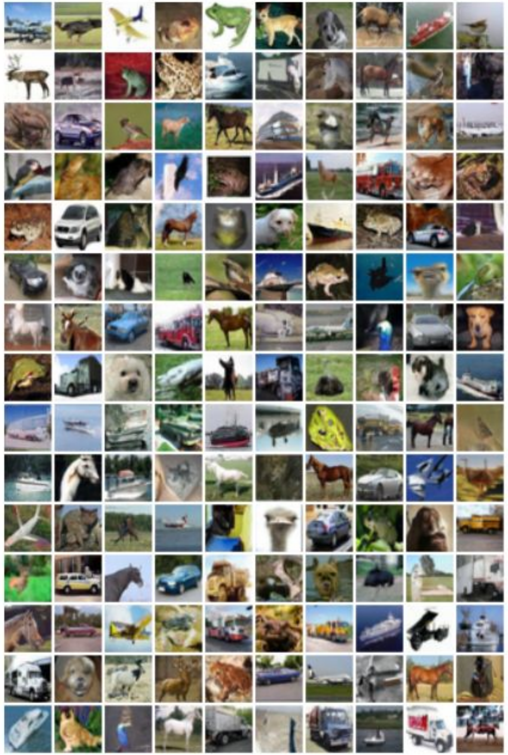
Generate Images

[BigGAN, Brock, Donahue, Simonyan, 2018]

Generate Images

[StyleGAN, Karras, Laine, Aila, 2018]

Generate Images



[Ho, Jain, Abbeel, 2020 Denoising Diffusion Probabilistic Models]

Text to Image

"a masterful oil painting a persian exotic cat discovering their astounding crypto losses while checking their phone"

Text to Image

"A Victorian man struggles with his addiction to TikTok"

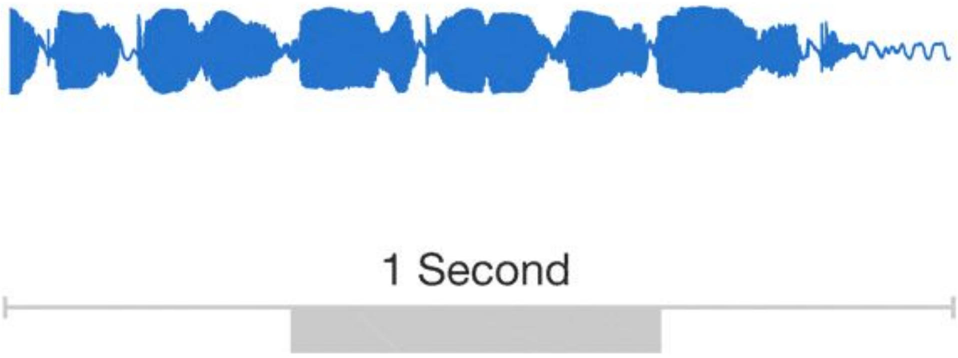
Text to Image

"Darth Vader realising he's forgotten to add an attachment to the email"

Text to Image

"A photo of a Shiba Inu dog with a backpack riding a bike. It is wearing sunglasses and a beach hat"

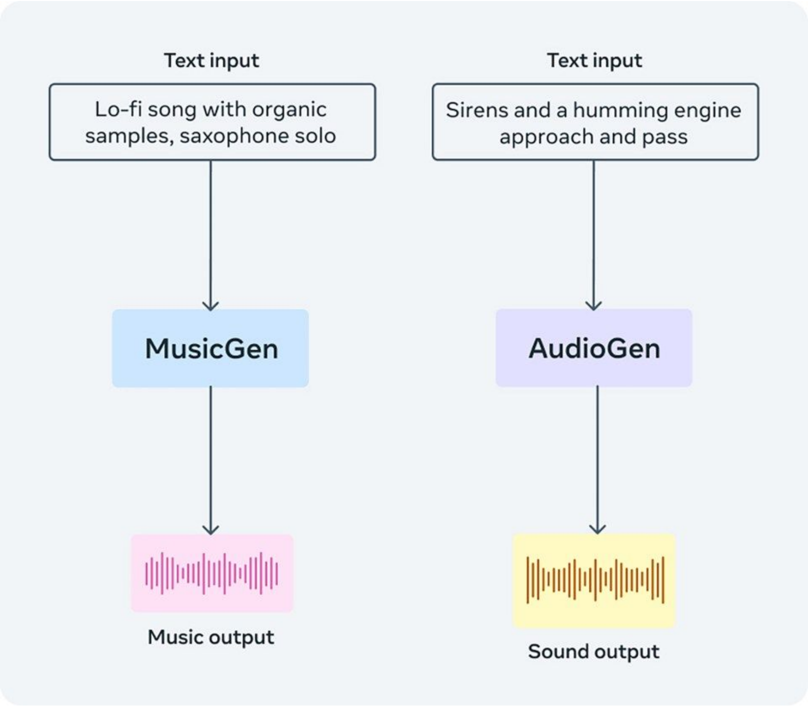
Generate Audio



Parametric WaveNet

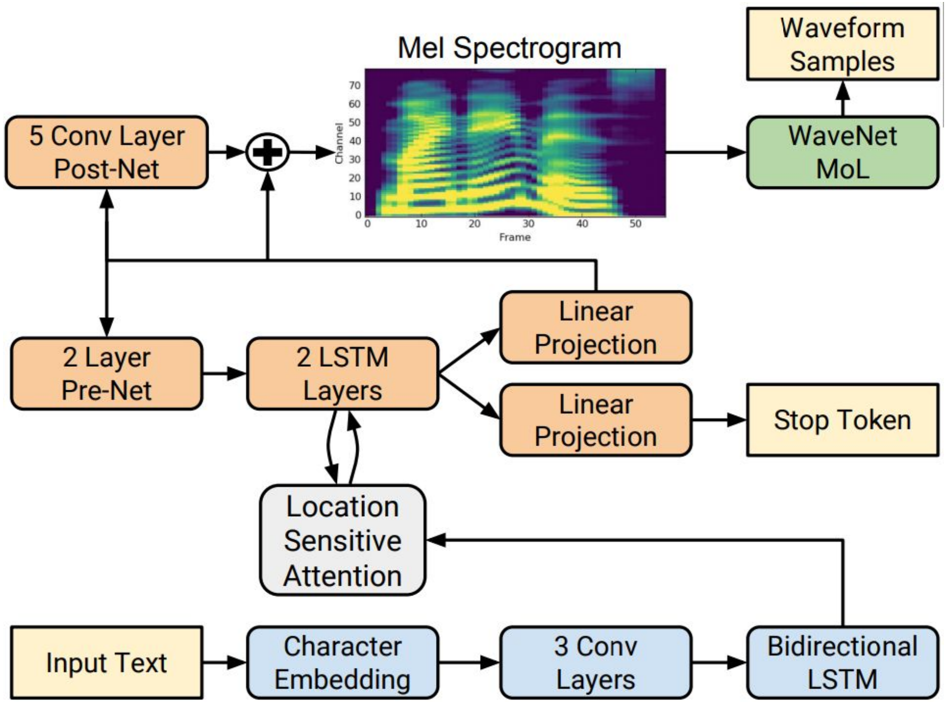
[WaveNet, Oord et al., 2018]

Text to Audio



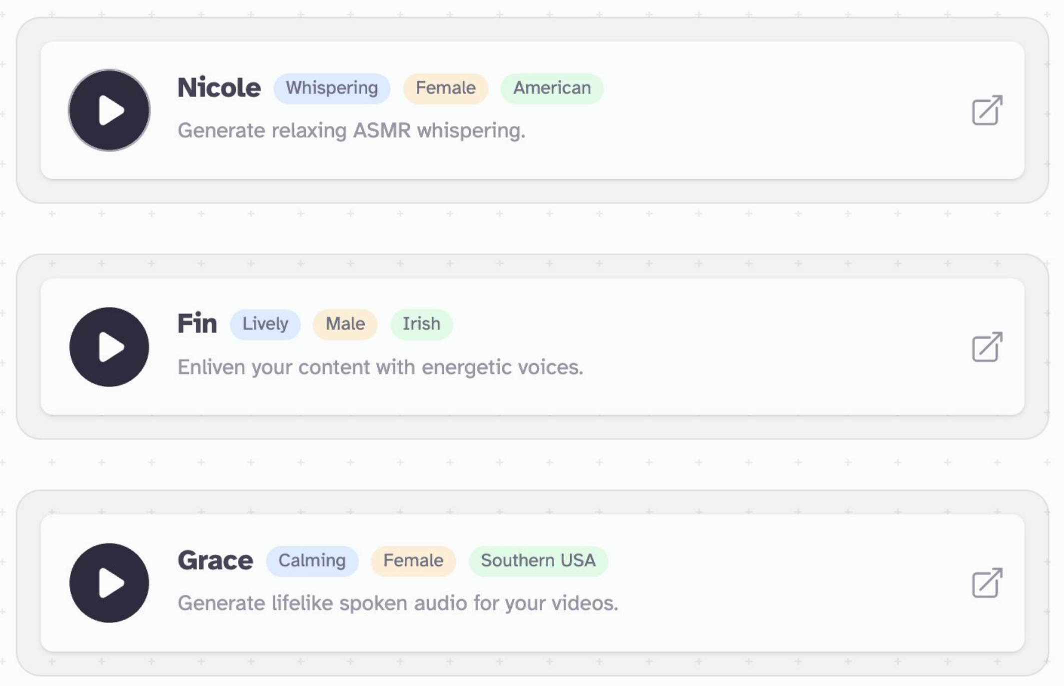
[AudioCraft, Copet et al., 2023]

Text to Speech



[Tacotron2 , Shen et al., 2018]

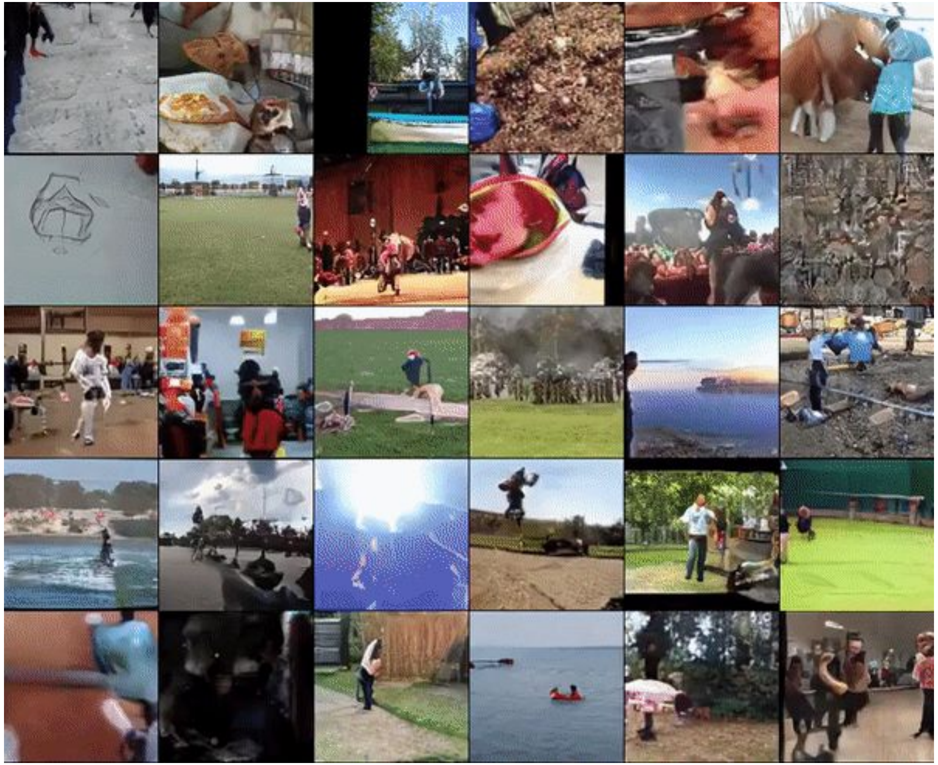
Text to Speech



https://elevenlabs.io/

Voice Conversion

Generate Video



DVD-GAN: Adversarial Video Generation on Complex Datasets, Clark, Donahue, Simonyan, 2019

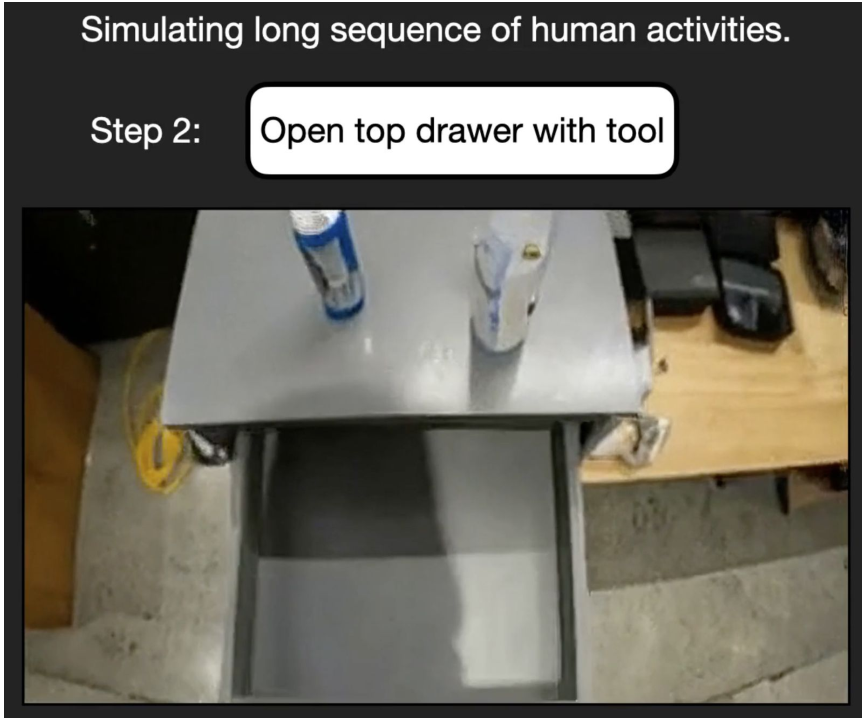
Text to Video

EMU VIDEO: Factorizing Text-to-Video Generation by Explicit Image Conditioning. Girdhar et al, 2023

Text to Video

VideoPoet: A Large Language Model for Zero-Shot Video Generation. Kondratyuk et al, 2023

Video Sim for Control



UniSim: Learning Interactive Real-World Simulators. Yang et al, 2023

Generate Text

PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

[Char-rnn, karpathy, 2015]

Generate Math

```
\begin{proof}
We may assume that $\mathcal{I}$ is an abelian sheaf on
$\mathcal{C}$.
\item Given a morphism $\Delta : \mathcal{F} \to
\mathcal{I}$
is an injective and let $\mathfrak q$ be an abelian sheaf on
$X$.
Let $\mathcal{F}$ be a fibered complex. Let $\mathcal{F}$
be a category.
\begin{enumerate}
\item \hyperref[setain-construction-phantom]{Lemma}
\label{lemma-characterize-quasi-finite}
Let $\mathcal{F}$ be an abelian quasi-coherent sheaf on
$\mathcal{C}$.
Let $\mathcal{F}$ be a coherent $\mathcal{O}_X$-module.
Then
$\mathcal{F}$ is an abelian catenary over $\mathcal{C}$.
\item The following are equivalent
\begin{enumerate}
\item $\mathcal{F}$ is an $\mathcal{O}_X$-module.
\end{lemma}
```

For $\bigoplus_{n=1,\ldots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, ?? and the fact that any U affine, see Morphisms, Lemma ??. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma ?? we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. □

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. □

The result for prove any open covering follows from the less of Example ??. It may replace S by $X_{spaces,étale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma ??. Namely, by Lemma ?? we see that R is geometrically regular over S.

[Char-rnn, karpathy, 2015]

Generate Text

Prompt [human written]: A train carriage containing controlled nuclear materials was stolen in Cincinnati today. Its whereabouts are unknown.

GPT-2: *The incident occurred on the downtown train line, which runs from Covington and Ashland stations. In an email to Ohio news outlets, the U.S. Department of Energy said it is working with the Federal Railroad Administration to find the thief.*

'The theft of this nuclear material will have significant negative consequences on public and environmental health, our workforce and the economy of our nation,' said Tom Hicks, the U.S. Energy Secretary, in a statement. 'Our top priority is to secure the theft and ensure it doesn't happen again.'

The stolen material was taken from the University of Cincinnati's Research Triangle Park nuclear research site, according to a news release from Department officials.

The Nuclear Regulatory Commission did not immediately release any information. According to the release, the U.S. Department of Energy's Office of Nuclear Material Safety and Security is leading that team's investigation.

'The safety of people, the environment and the nation's nuclear stockpile is our highest priority,' Hicks said. 'We will get to the bottom of this and make no excuses.'

[OpenAI's GPT2 -- Radford, Wu, Child, Luan, Amodei, Sutskever, 2019]

Generate Text

Prompt (human written): Recycling is good for the world, no, you could not be more wrong.

GPT-2: Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a tremendous price for the privilege of having the world's most advanced and efficient recycling system. Recycling is a huge, colossal waste of time, energy, money, and resources.

[OpenAI's GPT2 -- Radford, Wu, Child, Luan, Amodei, Sutskever, 2019]

JSON Generation

ChatGPT

Certainly! Below is a JSON object detailing various unsupervised deep learning methods, along with a brief description for each:

json   Copy code

```json
{
  "unsupervised_deep_learning_methods": [
    {
      "method": "Autoencoders",
      "description": "Neural networks designed to reconstruct the input
    },
    {
      "method": "Generative Adversarial Networks (GANs)",
      "description": "A framework for training models where two neural
    },
```

[OpenAI ChatGPT, 2023]

Long Form Summarization

unsupervised learning for tasks like topic modeling and sentiment analysis. Other notable applications include fraud detection in finance and e-commerce, and genomic studies where it helps cluster DNA sequences for genetic analysis.

Please summarize the above into three sentences

ChatGPT

Unsupervised learning is a key area in machine learning, characterized by its ability to discern patterns in unlabeled data, unlike supervised learning which relies on labeled examples. This method is particularly useful in handling large datasets where labeled data is rare or costly, and it finds applications in various fields such as recommender systems, image and video analysis, natural language processing, fraud detection, and genomic studies. Its ability to learn and identify structures without predefined labels makes it akin to human cognitive processes, enabling more organic and cost-effective data analysis.

[OpenAI ChatGPT, 2023]

Try it yourself

https://chat.openai.com

https://perplexity.ai

Compression - Lossless

Model	Bits per byte
CIFAR-10	
PixelCNN (Oord et al., 2016)	3.03
PixelCNN++ (Salimans et al., 2017)	2.92
Image Transformer (Parmar et al., 2018)	2.90
PixelSNAIL (Chen et al., 2017)	2.85
Sparse Transformer 59M (strided)	**2.80**
Enwik8	
Deeper Self-Attention (Al-Rfou et al., 2018)	1.06
Transformer-XL 88M (Dai et al., 2018)	1.03
Transformer-XL 277M (Dai et al., 2018)	**0.99**
Sparse Transformer 95M (fixed)	**0.99**
ImageNet 64x64	
PixelCNN (Oord et al., 2016)	3.57
Parallel Multiscale (Reed et al., 2017)	3.7
Glow (Kingma & Dhariwal, 2018)	3.81
SPN 150M (Menick & Kalchbrenner, 2018)	3.52
Sparse Transformer 152M (strided)	**3.44**
Classical music, 5 seconds at 12 kHz	
Sparse Transformer 152M (strided)	**1.97**

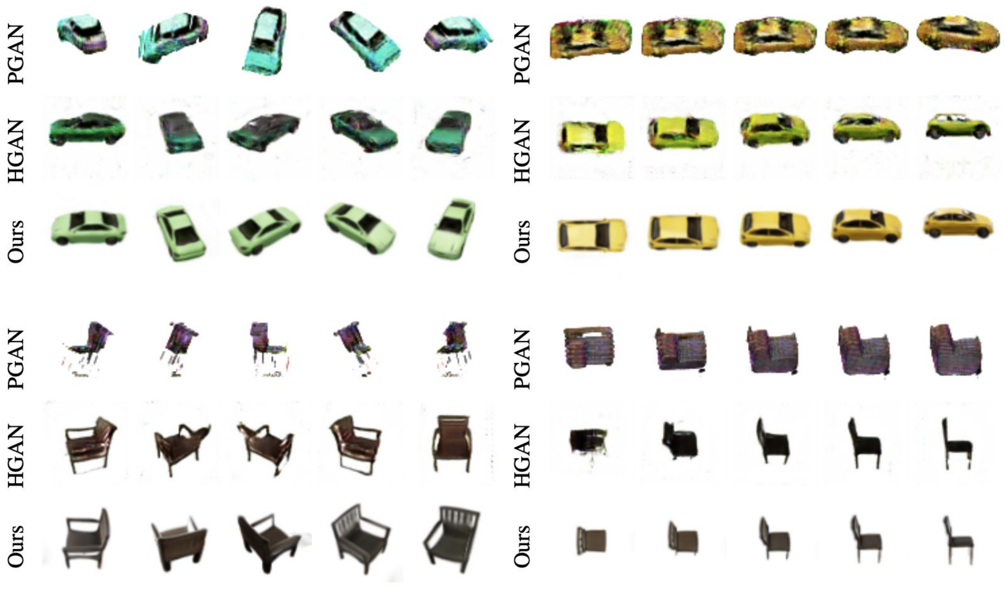
Generative models provide better bit-rates than distribution-unaware compression methods like JPEG, etc.

Compression - Lossy

JPEG JPEG2000 WaveOne

[Rippel & Bourdev, 2017]

3D Generation



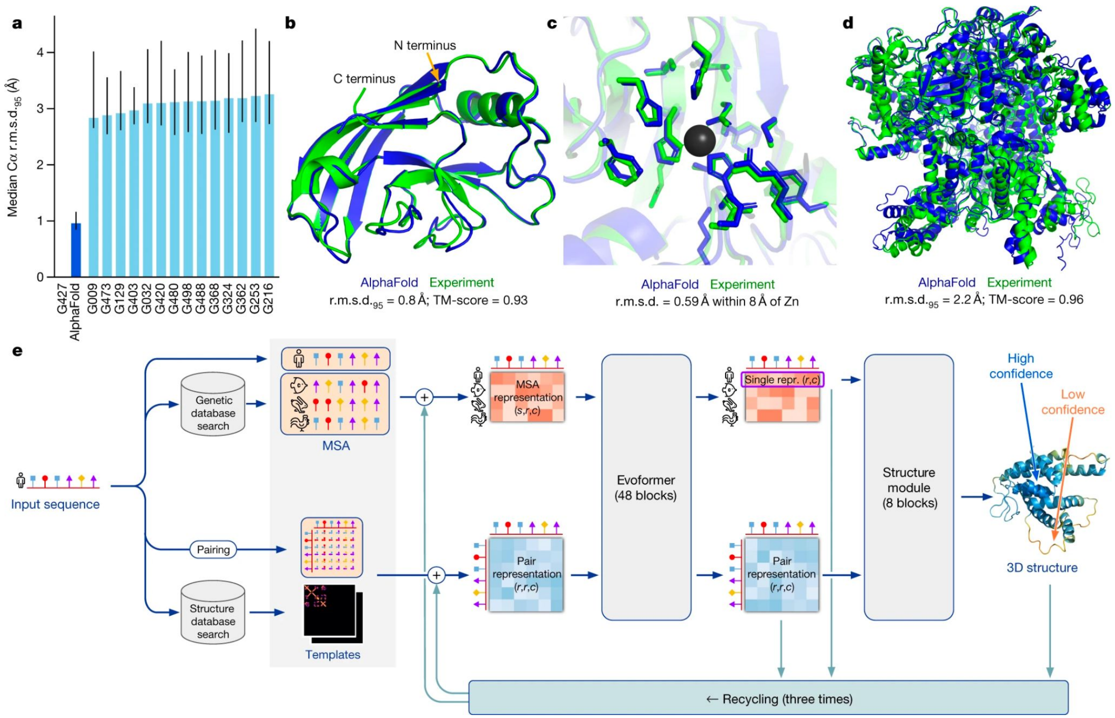
(a) Rotation (b) Elevation

[GRAF, Schwarz et al, 2020]

3D Generation

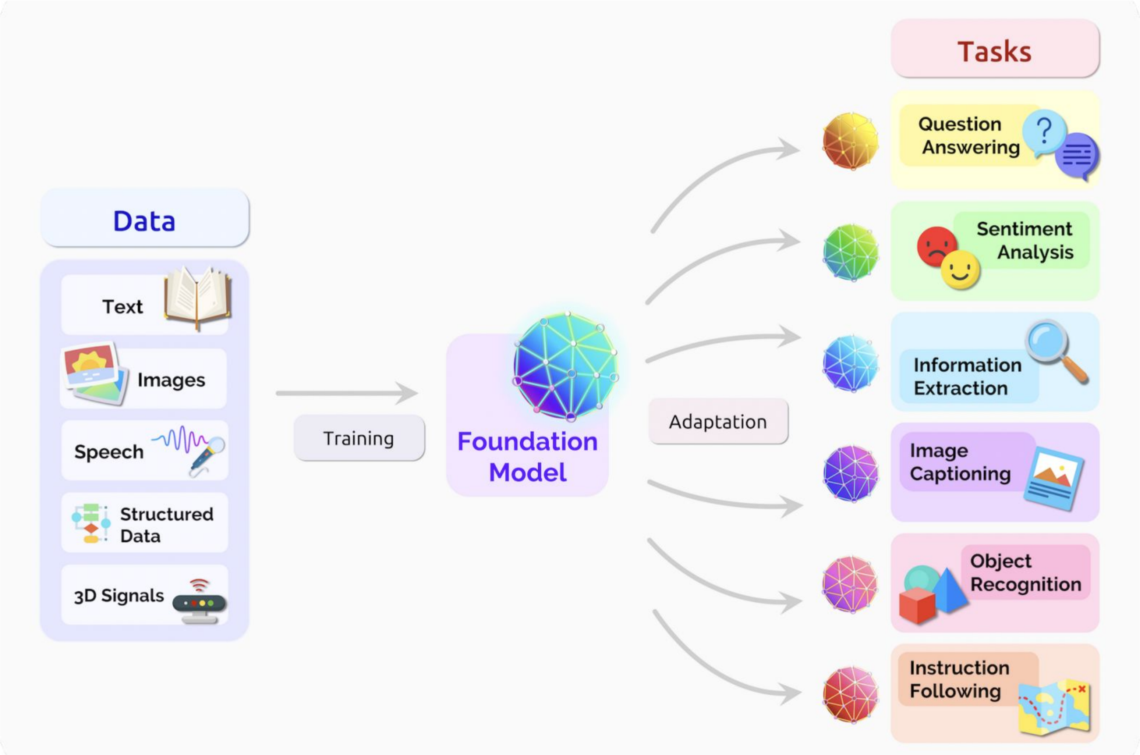
[GET3D, Gao et al, 2022]

AI for Science



[Highly accurate protein structure prediction with AlphaFold, Jumper Et Al, 2021]

Pretrain, then adapt



[On the Opportunities and Risks of Foundation Models, Bommasani et. al, 2022]

Downstream Task - Sentiment Detection

This is one of Crichton's best books. The characters of Karen Ross, Peter Elliot, Munro, and Amy are beautifully developed and their interactions are exciting, complex, and fast-paced throughout this impressive novel. And about 99.8 percent of that got lost in the film. Seriously, the screenplay AND the directing were horrendous and clearly done by people who could not fathom what was good about the novel. I can't fault the actors because frankly, they never had a chance to make this turkey live up to Crichton's original work. I know good novels, especially those with a science fiction edge, are hard to bring to the screen in a way that lives up to the original. But this may be the absolute worst disparity in quality between novel and screen adaptation ever. The book is really, really good. The movie is just dreadful.

[Radford et al., 2017]

Downstream Tasks - NLP (BERT Revolution)

	Rank	Name	Model	URL	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m	MNLI-mm	QNLI	RTE	WNLI
	1	T5 Team - Google	T5		90.3	71.6	97.5	92.8/90.4	93.1/92.8	75.1/90.6	92.2	91.9	96.9	92.8	94.5
	2	ERNIE Team - Baidu	ERNIE		90.0	72.2	97.5	93.0/90.7	92.9/92.5	75.2/90.8	91.2	90.8	96.0	90.9	94.5
	3	Microsoft D365 AI & MSR AI & GATECH	MT-DNN-SMART		89.9	69.5	97.5	93.7/91.6	92.9/92.5	73.9/90.2	91.0	90.8	99.2	89.7	94.5
+	4	王玮	ALICE v2 large ensemble (Alibaba DAMO NLP)		89.7	73.2	97.1	93.9/91.9	93.0/92.5	74.8/91.0	90.8	90.6	95.9	87.4	94.5
+	5	Microsoft D365 AI & UMD	FreeLB-RoBERTa (ensemble)		88.4	68.0	96.8	93.1/90.8	92.3/92.1	74.8/90.3	91.1	90.7	95.6	88.7	89.0
	6	Junjie Yang	HIRE-RoBERTa		88.3	68.6	97.1	93.0/90.7	92.4/92.0	74.3/90.2	90.7	90.4	95.5	87.9	89.0
	7	Facebook AI	RoBERTa		88.1	67.8	96.7	92.3/89.8	92.2/91.9	74.3/90.2	90.8	90.2	95.4	88.2	89.0
+	8	Microsoft D365 AI & MSR AI	MT-DNN-ensemble		87.6	68.4	96.5	92.7/90.3	91.1/90.7	73.7/89.9	87.9	87.4	96.0	86.3	89.0
	9	GLUE Human Baselines	GLUE Human Baselines		87.1	66.4	97.8	86.3/80.8	92.7/92.6	59.5/80.4	92.0	92.8	91.2	93.6	95.9
	10	Stanford Hazy Research	Snorkel MeTaL		83.2	63.8	96.2	91.5/88.5	90.1/89.7	73.1/89.9	87.6	87.2	93.9	80.9	65.1

[https://gluebenchmark.com/leaderboard]

Downstream Tasks - Few Shot

A.2.2 Additional Details for Pretrained Models Evaluation

MMLU details. In Table 19, we report details of the MMLU (Hendrycks et al., 2020) evaluation for Llama 2 models and others open-source models.

Standard Benchmarks. In Table 20, we show results on several standard benchmarks.

Code Generation. In Table 21, we compare results of Llama 2 with popular open source models on the Human-Eval and MBPP code generation benchmarks.

World Knowledge. We evaluate the Llama 2 model together with other open-source models on the NaturalQuestions and TriviaQA benchmarks (Table 22).

Reading Comprehension In Table 23 we report zero-shot and few-shot results on SQUAD and zero-shot and one-shot experiments on QUAC. Here Llama 2 performs best on all evaluation settings and models except the QUAC 0-shot where Llama 1 30B performs slightly better.

Exams. In Table 24, we present fine-grained results from the English part of the AGI Eval (Zhong et al., 2023) benchmark. AGI Eval is a collection of standardized exams in different subjects.

[LLAMA 2, 2023]

Downstream Tasks - Prompting

Prompt engineering

This guide shares strategies and tactics for getting better results from large language models (sometimes referred to as GPT models) like GPT-4. The methods described here can sometimes be deployed in combination for greater effect. We encourage experimentation to find the methods that work best for you.

[https://platform.openai.com/docs/guides/prompt-engineering]

(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: ***Let's think step by step.***

(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓

[Large Language Models are Zero-Shot Reasoners, Kojima et al, 2023]

Downstream Tasks - Vision (Contrastive)

Method	Architecture	mAP
Transfer from labeled data:		
Supervised baseline	ResNet-152	74.7
Transfer from unlabeled data:		
Exemplar [17] by [13]	ResNet-101	60.9
Motion Segmentation [47] by [13]	ResNet-101	61.1
Colorization [64] by [13]	ResNet-101	65.5
Relative Position [14] by [13]	ResNet-101	66.8
Multi-task [13]	ResNet-101	70.5
Instance Discrimination [60]	ResNet-50	65.4
Deep Cluster [7]	VGG-16	65.9
Deeper Cluster [8]	VGG-16	67.8
Local Aggregation [66]	ResNet-50	69.1
Momentum Contrast [25]	ResNet-50	74.9
Faster-RCNN trained on CPC v2	ResNet-161	**76.6**

"If, by the first day of autumn of 2015, a method will exist that can match or beat the performance of R-CNN on Pascal VOC detection, without the use of any extra, human annotations (e.g. ImageNet) as pre-training, Mr. Malik promises to buy Mr. Efros one (1) gelato (2 scoops: one chocolate, one vanilla)."

Table: Data-Efficient Image Recognition using CPC (Henaff, Srinivas, et al)

Downstream Tasks - Vision (MAE)

method	pre-train data	AP^box ViT-B	AP^box ViT-L	AP^mask ViT-B	AP^mask ViT-L
supervised	IN1K w/ labels	47.9	49.3	42.9	43.9
MoCo v3	IN1K	47.9	49.3	42.7	44.0
BEiT	IN1K+DALLE	49.8	**53.3**	44.4	47.1
MAE	IN1K	**50.3**	**53.3**	**44.9**	**47.2**

Table 4. **COCO object detection and segmentation** using a ViT Mask R-CNN baseline. All entries are based on our implementation. Self-supervised entries use IN1K data *without* labels. Mask AP follows a similar trend as box AP.

method	pre-train data	ViT-B	ViT-L
supervised	IN1K w/ labels	47.4	49.9
MoCo v3	IN1K	47.3	49.1
BEiT	IN1K+DALLE	47.1	53.3
MAE	IN1K	**48.1**	**53.6**

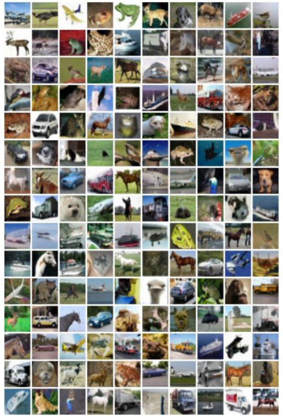
Table 5. **ADE20K semantic segmentation** (mIoU) using UperNet. BEiT results are reproduced using the official code. Other entries are based on our implementation. Self-supervised entries use IN1K data *without* labels.

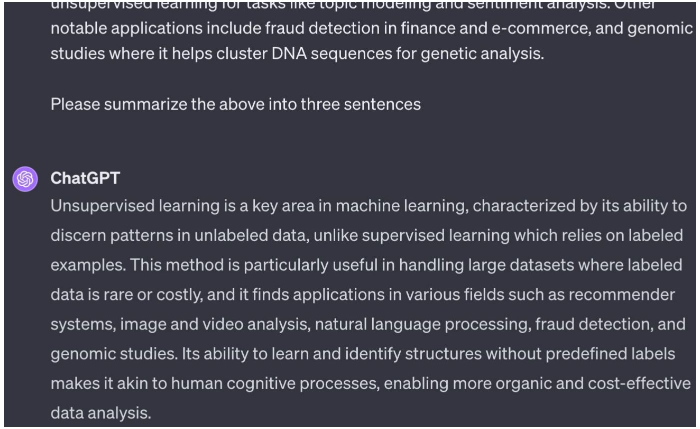
dataset	ViT-B	ViT-L	ViT-H	ViT-H_{448}	prev best
iNat 2017	70.5	75.7	79.3	**83.4**	75.4 [55]
iNat 2018	75.4	80.1	83.0	**86.8**	81.2 [54]
iNat 2019	80.5	83.4	85.7	**88.3**	84.1 [54]
Places205	63.9	65.8	65.9	**66.8**	66.0 [19]†
Places365	57.9	59.4	59.8	**60.3**	58.0 [40]‡

Table 6. **Transfer learning accuracy on classification datasets**, using MAE pre-trained on IN1K and then fine-tuned. We provide system-level comparisons with the previous best results.
†: pre-trained on 1 billion images. ‡: pre-trained on 3.5 billion images.

[Masked Autoencoders Are Scalable Vision Learners, He et al, 2021]

Unsupervised Learning Scales with Data

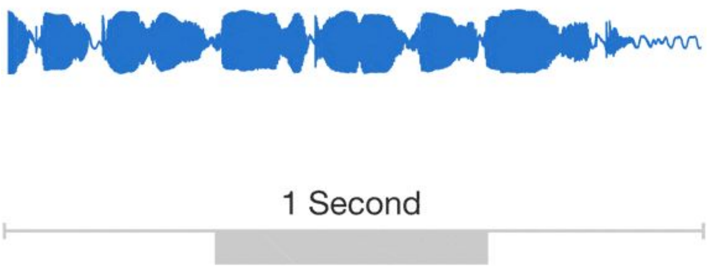



unsupervised learning for tasks like topic modeling and sentiment analysis. Other notable applications include fraud detection in finance and e-commerce, and genomic studies where it helps cluster DNA sequences for genetic analysis.

Please summarize the above into three sentences

ChatGPT

Unsupervised learning is a key area in machine learning, characterized by its ability to discern patterns in unlabeled data, unlike supervised learning which relies on labeled examples. This method is particularly useful in handling large datasets where labeled data is rare or costly, and it finds applications in various fields such as recommender systems, image and video analysis, natural language processing, fraud detection, and genomic studies. Its ability to learn and identify structures without predefined labels makes it akin to human cognitive processes, enabling more organic and cost-effective data analysis.



Summary

- **Unsupervised Learning**: Rapidly advancing field thanks to compute; deep learning engineering practices; datasets; lot of people working on it
- **Not just an academic interest topic**: Language Modeling, Image Generation, Vision / Language / Multimodal Pre-Training are all working really well and have production level impact
- **What is true now may not be true even a year from now**
  example1: self-supervised pre-training was way worse than supervised in computer vision tasks like detection/segmentation until 2019, now it is better
  example2: representation learning for vision through masking didn't work until Kaiming He et al made it work November 2021
- **Autoregressive Models, Flows, VAEs, GANs, Diffusion Models** still have ***significant room for surprising capabilities***, especially in newer domains such as multi-modal, video, robotics, bio, sciences. Great time to work on them.
- **Core of Unsupervised Learning** might still have some major innovations ahead (e.g. last offering (2020) we wrote Diffusion Models paper right after semester wrapped up)