Machine Learning I 60629A

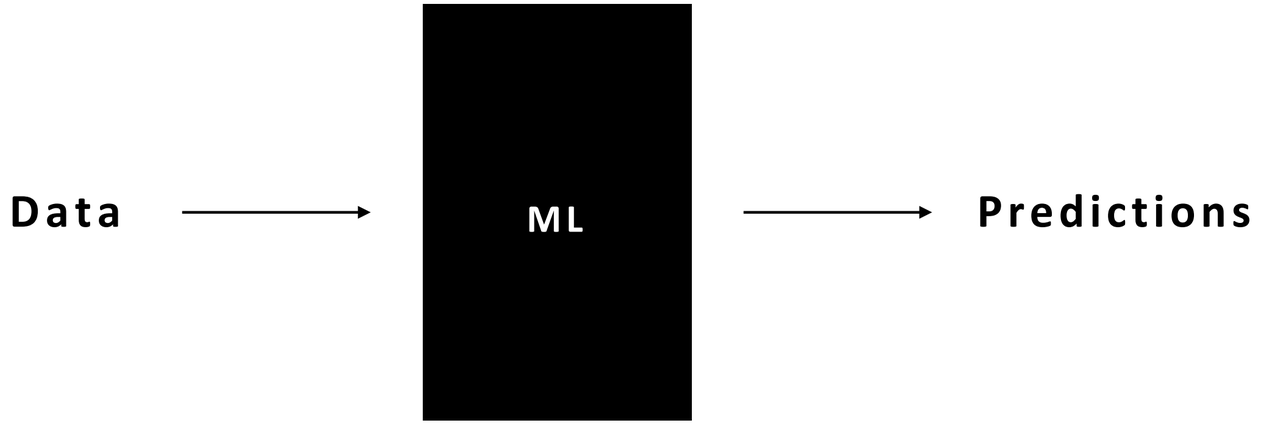
Week #1

Today

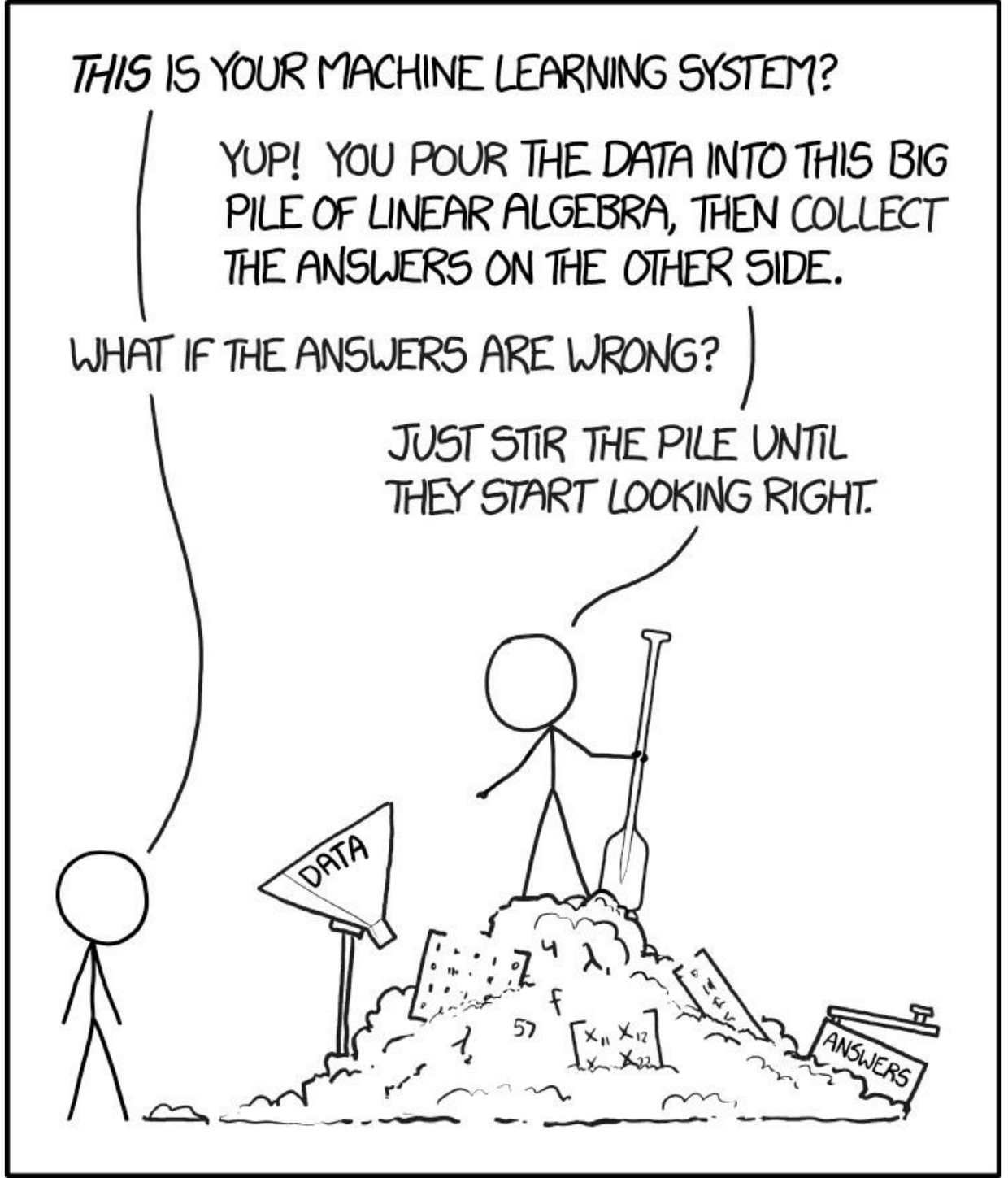
- Introduction to machine learning
- The course (syllabus)
- Math review (probability + linear algebra)

Machine Learning (ML)

- Science that studies statistical and computational aspects of modeling data for predictive purposes
  - (Mostly) Empirical science

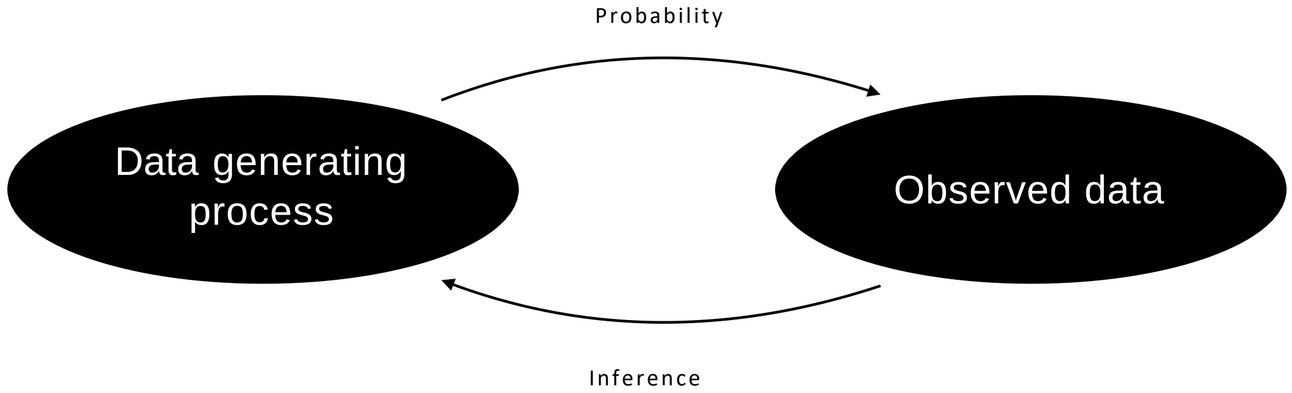


- Task: Predict whether an image contains a malignant tumor
- Task: Predict the next movie a person should watch



https://xkcd.com/1838/

“Data analysis, machine learning and data mining are various names given to the practice of statistical inference, depending on the context.”



–Larry Wasserman in “All of Statistics: A Concise Course in Statistical Inference.”

What is the goal of ML?

- A bit of historical context
  - When I started my PhD very few in ML talked about AI
  - Recent ML makes progress toward “AI tasks”
    - Examples of AI tasks: translation, object recognition
    - In that context: create a machine with human-like capacities? Or a machine that can help humans?

- **My (imperfect) view:**

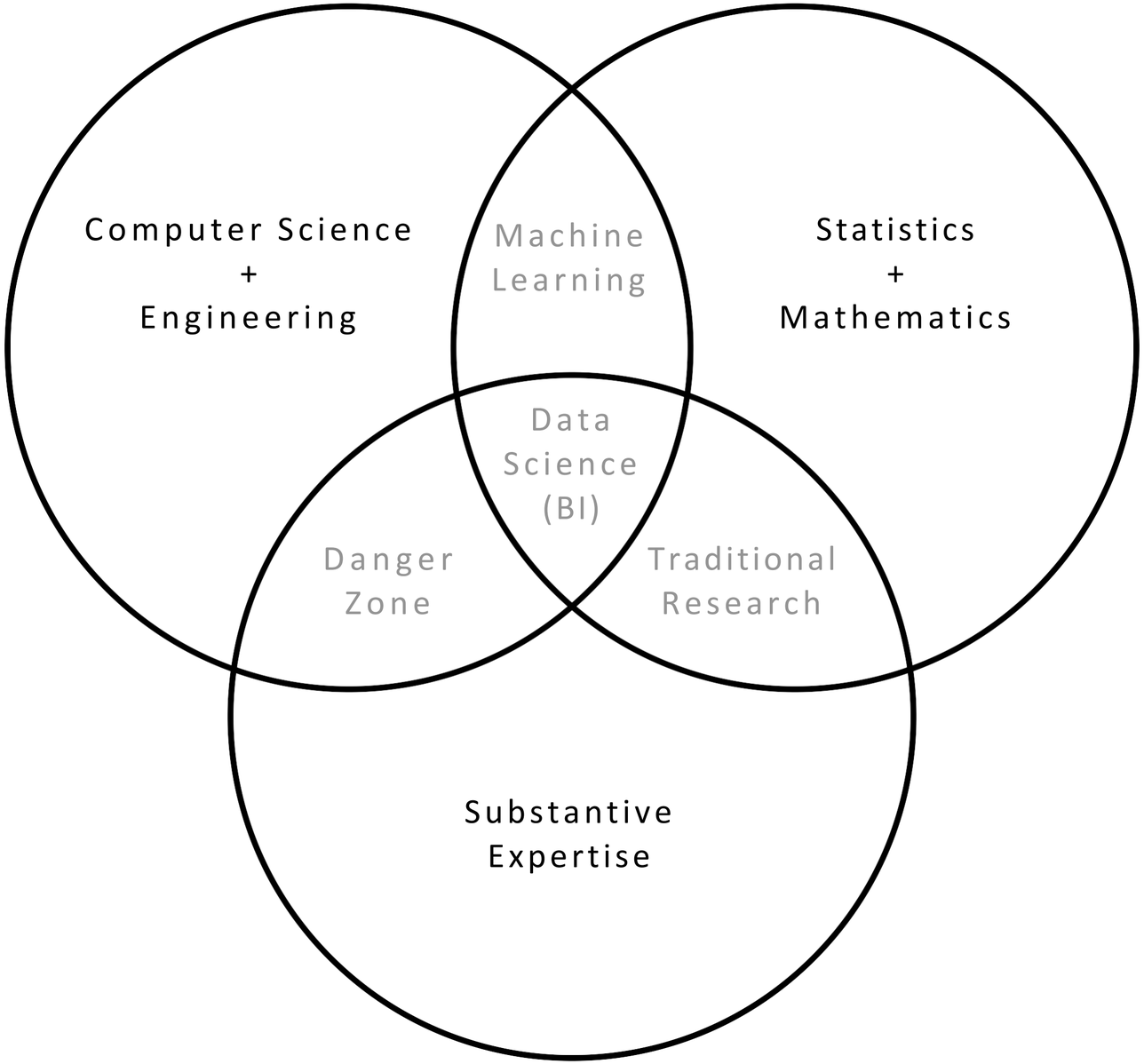
  Understand data through predictive models

  Understand the world through predictive models

How does ML relate to other fields

Historical View

- (Modern) Statistics: ~1900
- Machine Learning and Data Mining: ~1960
- Data Science: ~2000



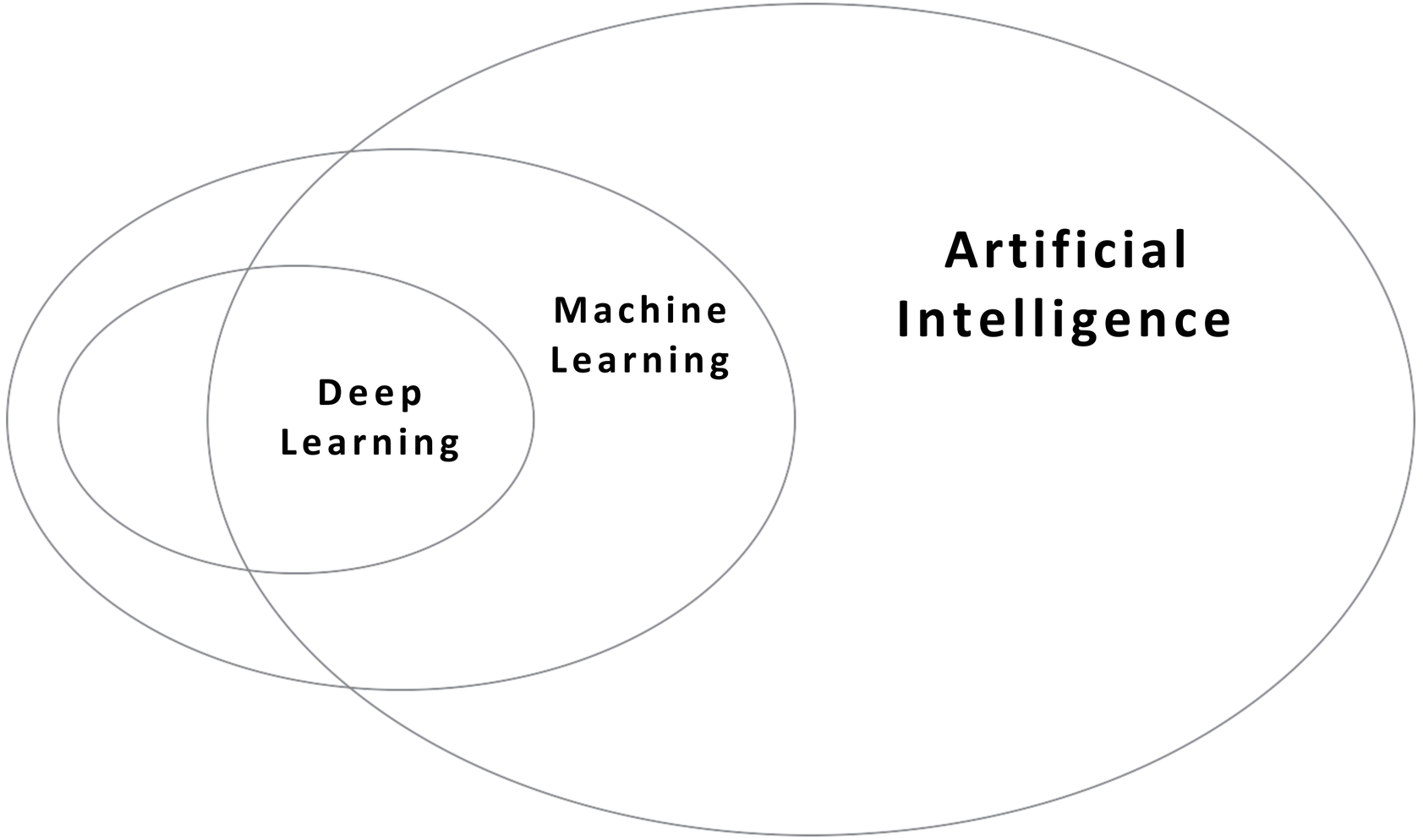
[Inspired by Drew Conway]

Attitudes in Machine Learning and Data Mining Versus Attitudes in Traditional Statistics

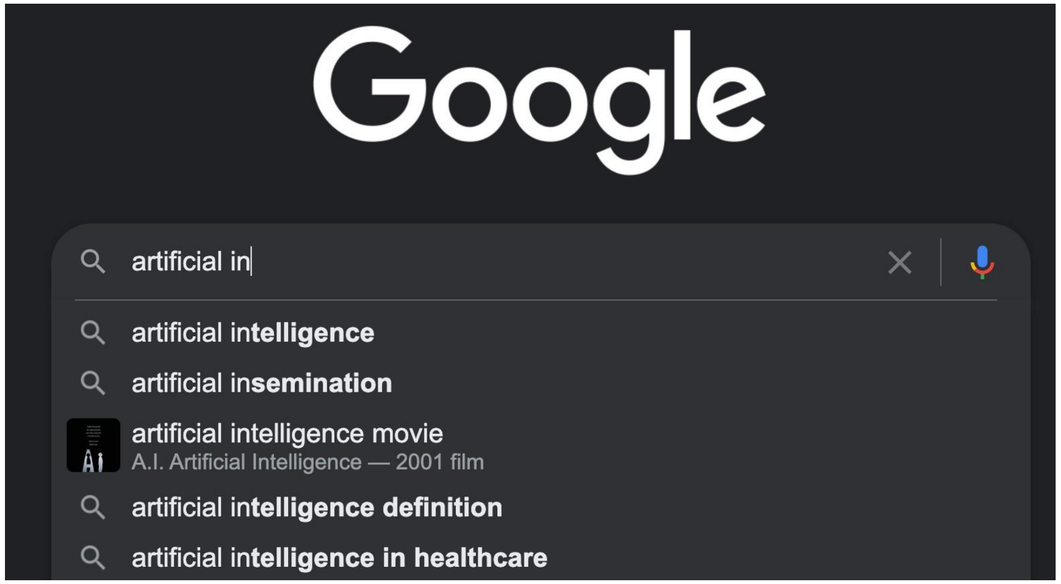
Despite these differences, there's a big overlap in problems addressed by machine learning and data mining and by traditional statistics. But attitudes differ...

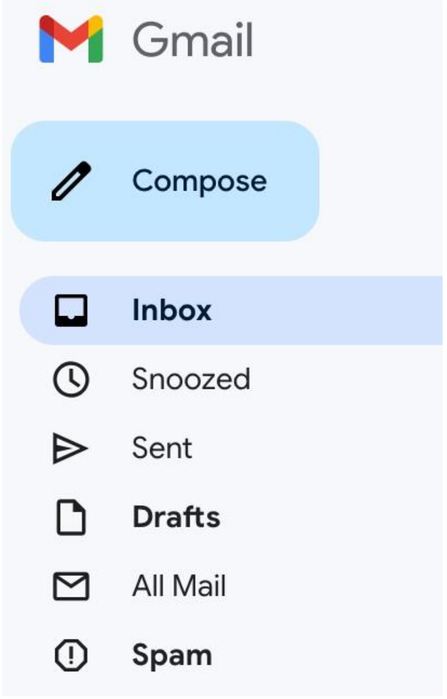
Machine learning	**Traditional statistics**
No settled philosophy or widely accepted theoretical framework.	Classical (frequentist) and Bayesian philosophies compete.
Willing to use *ad hoc* methods if they seem to work well (though appearances may be misleading).	Reluctant to use methods without some theoretical justification (even if the justification is actually meaningless).
Emphasis on automatic methods with little or no human intervention.	Emphasis on use of human judgement assisted by plots and diagnostics.
Methods suitable for many problems.	Models based on scientific knowledge.
Heavy use of computing.	Originally designed for hand-calculation, but computing is now very important.

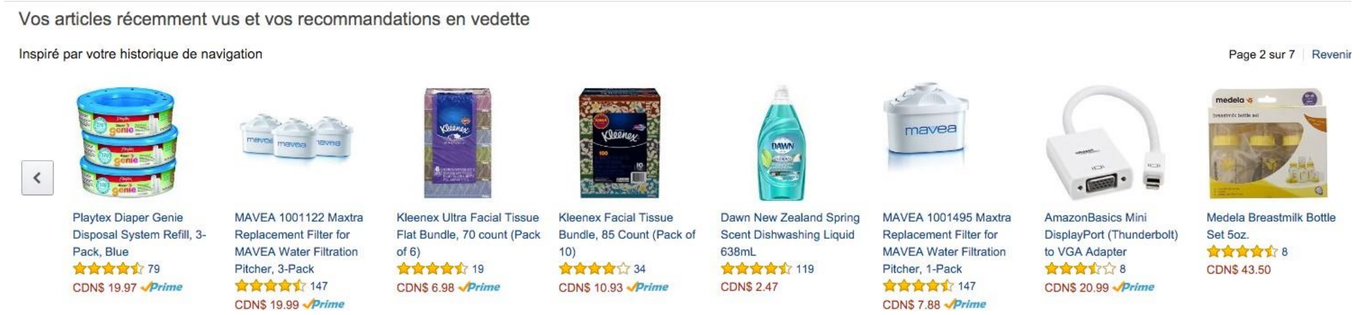
[Reproduced from R. Neal: http://www.utstat.utoronto.ca/~radford/sta414.S14/week1.pdf]


Artificial Intelligence
Machine Learning
Deep Learning

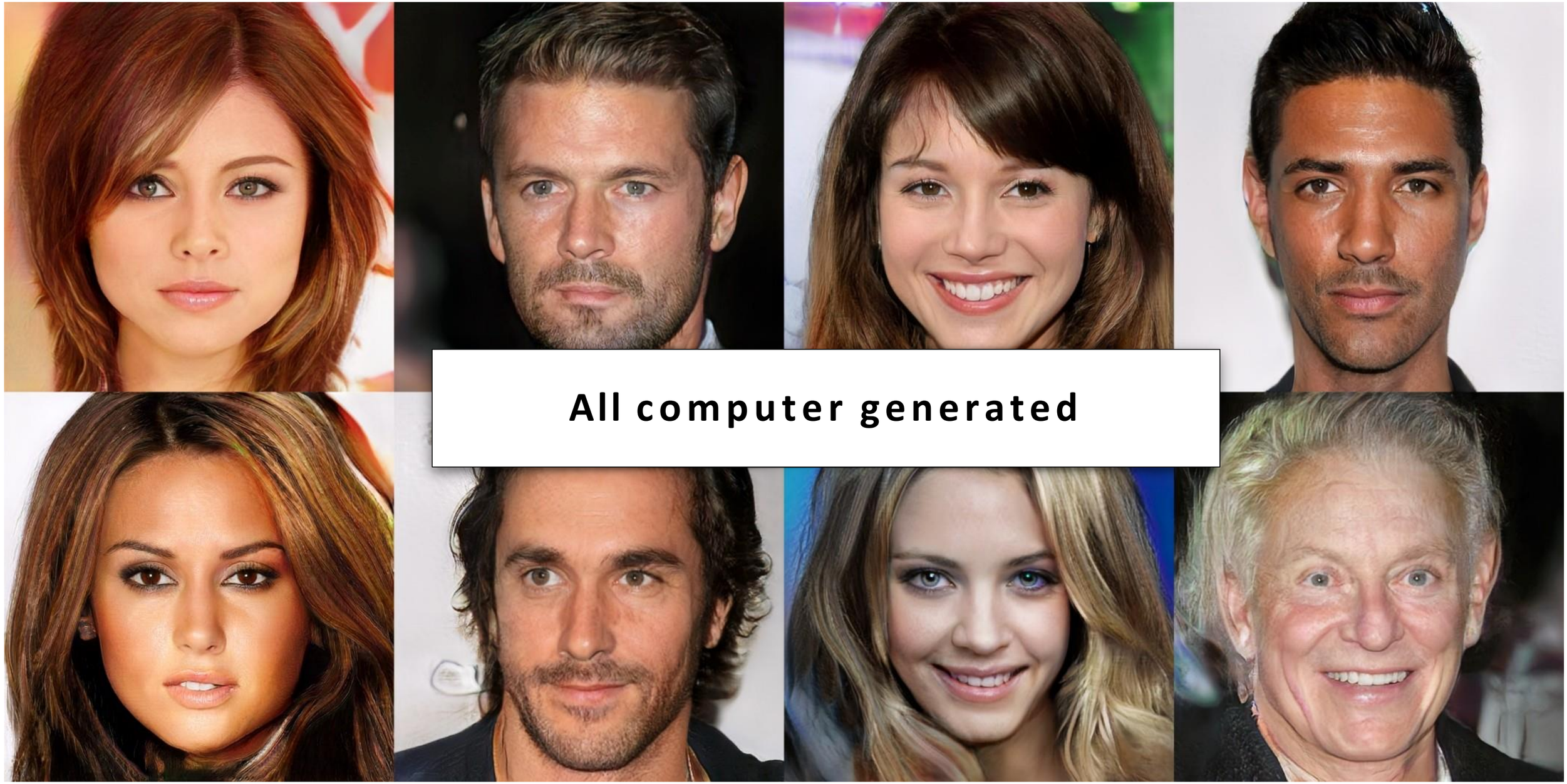
Applications of ML


Google
artificial in
artificial intelligence
artificial insemination
artificial intelligence movie
A.I. Artificial Intelligence — 2001 film
artificial intelligence definition
artificial intelligence in healthcare


Gmail
Compose
Inbox
Snoozed
Sent
Drafts
All Mail
Spam


Vos articles récemment vus et vos recommandations en vedette
Inspiré par votre historique de navigation
Page 2 sur 7
Revenir
Playtex Diaper Genie Disposal System Refill, 3-Pack, Blue
79
CDN$ 19.97 Prime
MAVEA 1001122 Maxtra Replacement Filter for MAVEA Water Filtration Pitcher, 3-Pack
147
CDN$ 19.99 Prime
Kleenex Ultra Facial Tissue Flat Bundle, 70 count (Pack of 6)
19
CDN$ 6.98 Prime
Kleenex Facial Tissue Bundle, 85 Count (Pack of 10)
34
CDN$ 10.93 Prime
Dawn New Zealand Spring Scent Dishwashing Liquid 638mL
119
CDN$ 2.47
MAVEA 1001495 Maxtra Replacement Filter for MAVEA Water Filtration Pitcher, 1-Pack
147
CDN$ 7.88 Prime
AmazonBasics Mini DisplayPort (Thunderbolt) to VGA Adapter
8
CDN$ 20.99 Prime
Medela Breastmilk Bottle Set 5oz.
8
CDN$ 43.50

Progressive Growing of GANs for Improved Quality, Stability, and Variation
Karras et al., ICLR'18



All computer generated

NVIDIA

Progressive Growing of GANs for Improved Quality, Stability, and Variation
Karras et al., ICLR'18

2014

2015

2016

2017

2018

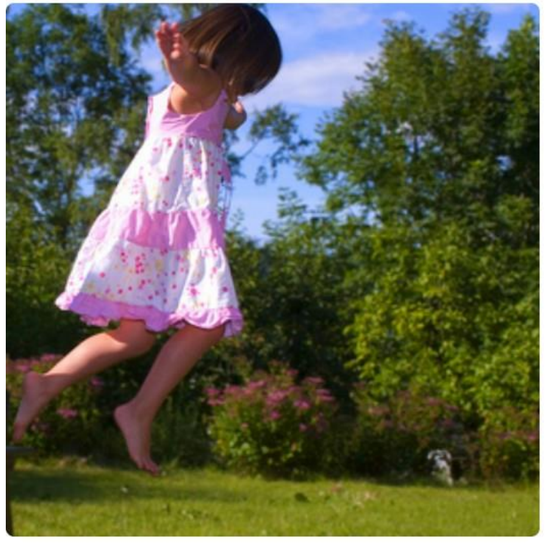
https://twitter.com/goodfellow_ian/status/1084973596236144640/photo/1

"man in black shirt is playing guitar."

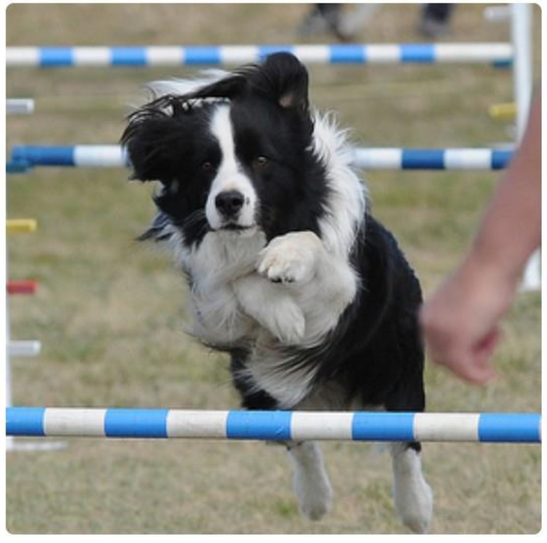
"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."



"girl in pink dress is jumping in air."



"black and white dog jumps over bar."

"young girl in pink shirt is swinging on swing."

"man in blue wetsuit is surfing on wave."

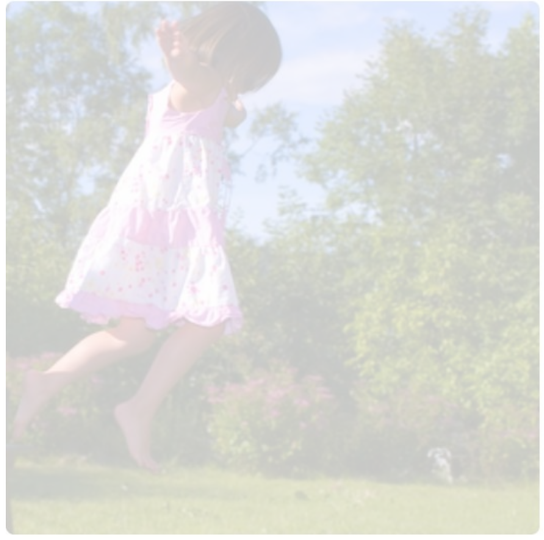
Deep Visual-Semantic Alignments for Generating Image Descriptions
A. Karpathy, L. Fei-Fei, CVPR'15

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."



"girl in pink dress is jumping in air."

"black and white dog jumps over bar."

"young girl in pink shirt is swinging on swing."

"man in blue wetsuit is surfing on wave."

Deep Visual-Semantic Alignments for Generating Image Descriptions
A. Karpathy, L. Fei-Fei, CVPR'15

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

AI-GENERATED IMAGES

Dall-e 2

Stable Diffusion 1.5

https://huggingface.co/spaces/runwayml/stable-diffusion-v1-5
https://openai.com/blog/dall-e/

Generate an image from a text prompt

TEXT PROMPT

A bowl of soup that looks like a monster spray-painted on the wall

AI-GENERATED IMAGES

https://openai.com/dall-e-2/

Generate an image from a text prompt

TEXT PROMPT

A bowl of soup that looks like a monster made out of plasticine

AI-GENERATED IMAGES

Stable Diffusion 1.5

https://openai.com/dall-e-2/

Generate an image from a text prompt

TEXT PROMPT Two women shaking hands

AI-GENERATED IMAGES

https://openai.com/dall-e-2/

Stable Diffusion 1.5

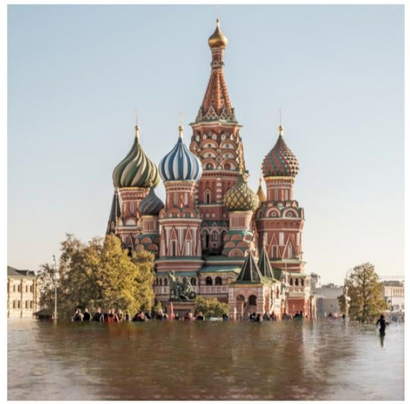
https://openai.com/dall-e-2/

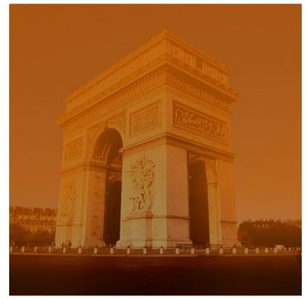
This Climate Does Not Exist

Visualize climate change

What now

These images
are generated by
artificial intelligence (AI)





But the environmental
disasters they portray
are very real

Can you imagine these kinds of disasters happening in your own backyard?

https://thisclimatedoesnotexist.com/en/share/57d2b713-a6ea-474b-8d32-2012db0f050b

“Lombard street ...”

“view from bottom”

“view from top”

“bird’s eye view”

“in heavy rain”

[Jia et al. 21, CoRR: 2102.05918]

“Lombard street ...”

“view from bottom”

“view from top”

“bird’s eye view”

“in heavy rain”

“seagull in front of ...”

“Golden Gate Bridge”

“London Tower Bridge”

“Sydney Harbour Bridge”

“Rialto Bridge”

[Jia et al. 21, CoRR: 2102.05918]

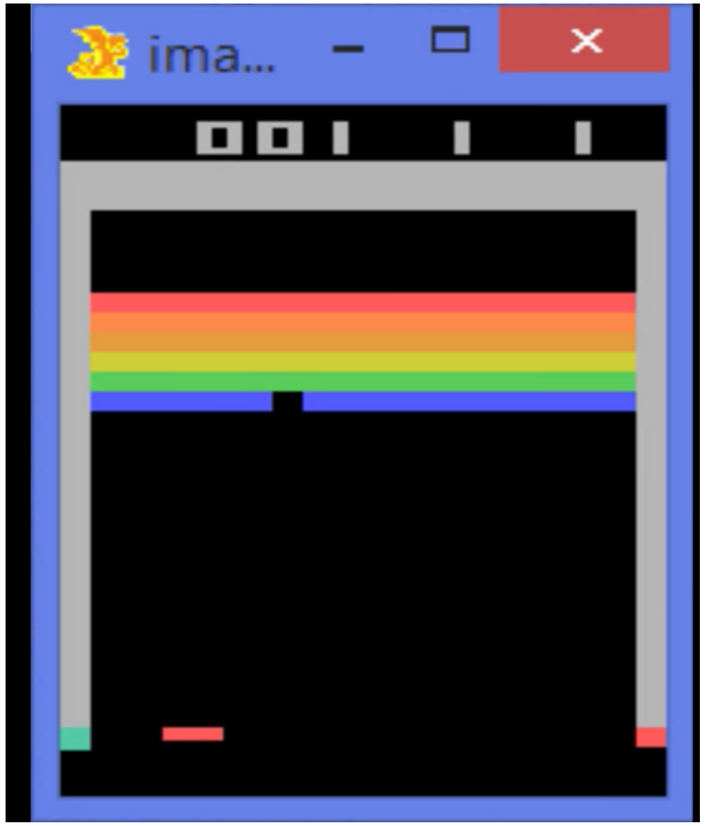
Demo
Speech Translation
Demo
Speech Translation

Demo
Speech Translation
Demo
Speech Translation

nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO
CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459
POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card forty years on
PAGE 462
NATURE.COM/NATURE
28 January 2016 £10
Vol. 529, No. 7587
9 770028 083095 04

Robotic observatory makes fast work of astronomy p. 478
A wet route to methanol p. 523
Human noise plagues protected areas p. 531
Science
$15
5 MAY 2017
sciencemag.org
AAAS
A
DIGITAL CARDS WHIZ
AI beats humans at challenging poker variant
p. 508
A

AI for video games



Mnih et al.
Nature
Volume 518,
pages 529–533
(26 February 2015)

- Medicine: personalized, automate diagnostics
- Social sciences: prediction problem (e.g., predict recidivism)
- Engineering: to propose new design, evaluate without building
- Finance: capture uncertainty, short-term trading
- Marketing: to understand and quantify user experience, advertising efficacy
- Many others: conservation, social projects, climate change
- Your domain of expertise…

Course Introduction & Goals

Logistics

- **Course syllabus:** https://davoodwadi.github.io/sessions.html

Flipped Classroom

- Every week:
  1. Class preparation (Offline):
     - Weekly material (~90 minutes)
       - Reading, watching capsules
  2. Class time (Online):
     - Summary, Q&A, problem solving (120 minutes)

Suggestions for navigating a flipped classroom

- In class: Come prepared
  - Watch the capsules ahead of time
  - Do the readings
  - Write down your questions
- Capsules: Stay active while watching the capsules (e.g., take notes, pause, go back, think of how it fits in the broader context)

Fit with other courses

- HEC
  - PhD level (originally)
  - Computationally oriented
  - Prequel to
    - Machine Learning II: Deep Learning (MATH 60630A)
    - Trustworthy Machine Learning (MATH 80630)
- Other ML courses in Montreal (U.Montreal, Polytechnique, Mcgill)
  - More applied (similar to COMP-551@McGill)

Notes on LLMs’ use

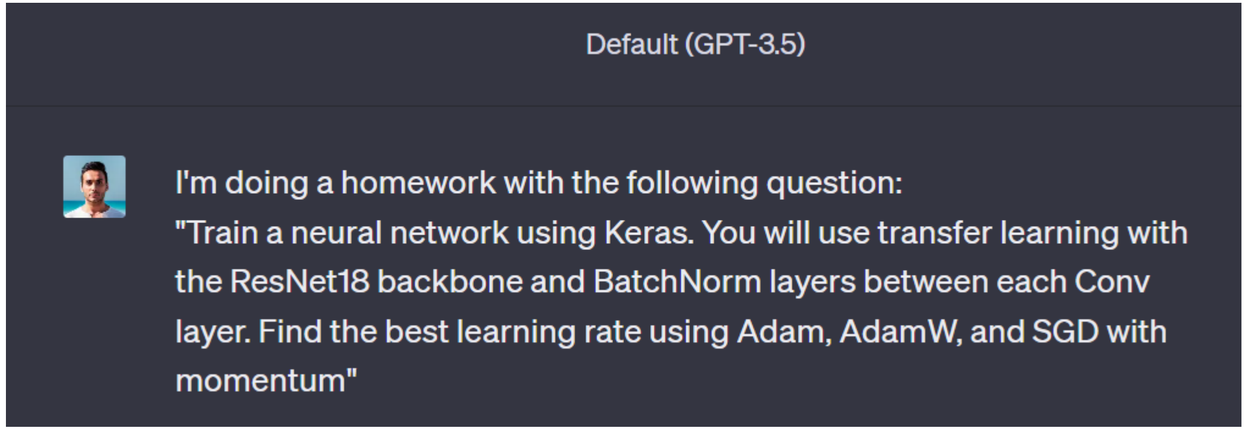
- Use for learning code
- Not for **writing** code or text

An experiment on ChatGPT

Default (GPT-3.5)

I'm doing a homework with the following question:
"Train a neural network using Keras. You will use transfer learning with the ResNet18 backbone and BatchNorm layers between each Conv layer. Find the best learning rate using Adam, AdamW, and SGD with momentum"

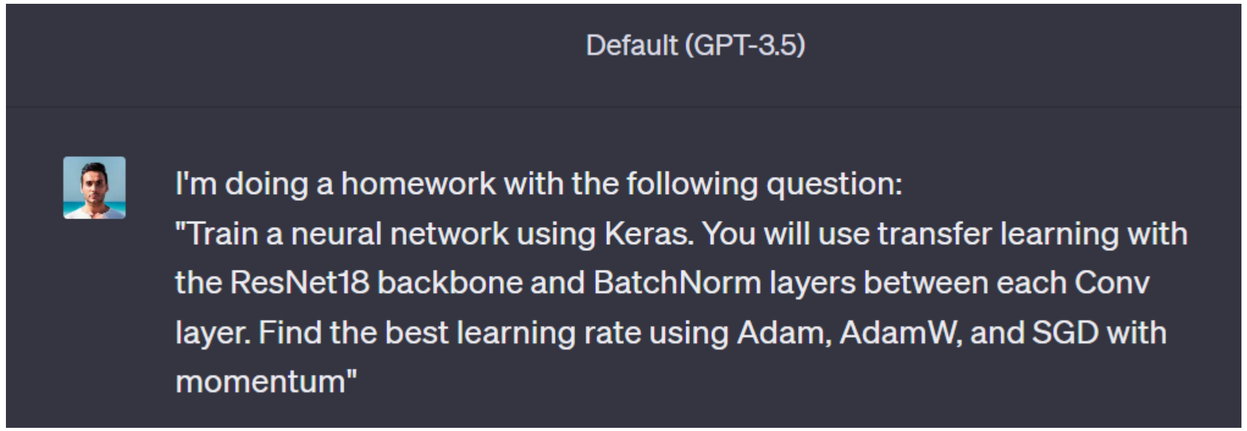
An experiment on ChatGPT




```
import tensorflow as tf
from tensorflow.keras.applications import ResNet18
from tensorflow.keras.layers import Dense, GlobalAveragePooling2D
from tensorflow.keras.models import Model
from tensorflow.keras.optimizers import Adam, AdamW, SGD
from tensorflow.keras.datasets import ...

# Load your dataset here (e.g., CIFAR-10, ImageNet, etc.)
# (X_train, y_train), (X_test, y_test) = ...
```

An experiment on ChatGPT




```
base_model = ResNet18(include_top=False, weights='imagenet', input_shape=(224, 224, 3))

x = base_model.output
x = GlobalAveragePooling2D()(x)
x = Dense(1024, activation='relu')(x)
predictions = Dense(num_classes, activation='softmax')(x)  # Adjust num_classes according to your task

model = Model(inputs=base_model.input, outputs=predictions)
```

An experiment on ChatGPT

```
# Learning rate finder (optional)
lr_finder = tf.keras.callbacks.LearningRateScheduler(lambda epoch: 1e-8 * 10**(epoch / 20))

# Define learning rates for experimentation
learning_rates = [1e-4, 1e-3, 1e-2, 1e-1]

for lr in learning_rates:
    # Adam optimizer
    adam_optimizer = Adam(learning_rate=lr)
    model.compile(optimizer=adam_optimizer, loss='categorical_crossentropy', metrics=['accuracy'])
    model.fit(X_train, y_train, batch_size=32, epochs=5, callbacks=[lr_finder])

    # AdamW optimizer
    adamw_optimizer = AdamW(learning_rate=lr)
    model.compile(optimizer=adamw_optimizer, loss='categorical_crossentropy', metrics=['accuracy'])
    model.fit(X_train, y_train, batch_size=32, epochs=5, callbacks=[lr_finder])

    # SGD optimizer with momentum
    sgd_optimizer = SGD(learning_rate=lr, momentum=0.9)
    model.compile(optimizer=sgd_optimizer, loss='categorical_crossentropy', metrics=['accuracy'])
    model.fit(X_train, y_train, batch_size=32, epochs=5, callbacks=[lr_finder])
```

An experiment on ChatGPT

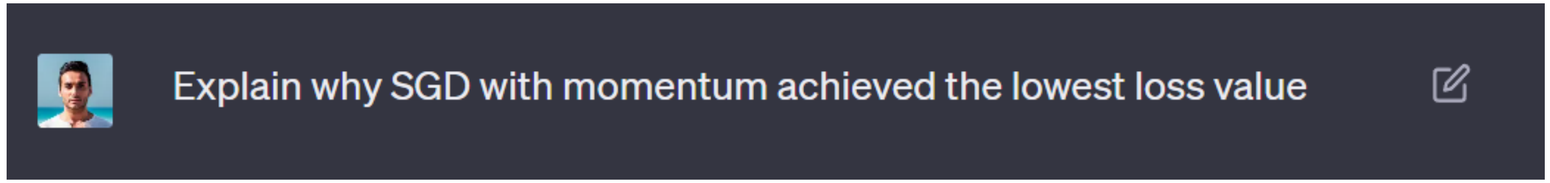
```
import matplotlib.pyplot as plt

# Plot the learning rate vs. loss graph
lrs = learning_rates
losses = [list_of_adam_losses, list_of_adamw_losses, list_of_sgd_losses]  # Update with your actual loss values
plt.semilogx(lrs, losses[0], label='Adam')
plt.semilogx(lrs, losses[1], label='AdamW')
plt.semilogx(lrs, losses[2], label='SGD')
plt.xlabel('Learning Rate')
plt.ylabel('Loss')
plt.legend()
plt.show()
```

Visible bugs + hidden bugs

An experiment on ChatGPT

- Use for learning code
- Not for **writing** code or **text**



An experiment on ChatGPT

SGD with momentum is an optimization algorithm that builds upon the basic Stochastic Gradient Descent (SGD) method by introducing a momentum term. This momentum term accelerates the optimization process and helps the model overcome the limitations of vanilla SGD, especially in scenarios where the loss surface is rugged or contains many local minima. The achieved lower loss value for SGD with momentum compared to Adam and AdamW could be attributed to several factors:

An experiment on ChatGPT

1.Acceleration of Convergence: The momentum term in SGD introduces a moving average of past gradients. This has the effect of accelerating the convergence towards the minimum, especially in the presence of oscillations or noisy gradients. This acceleration can help the optimizer escape shallow local minima and make quicker progress towards the optimal solution.
2.Inertia: The momentum term allows the optimization process to "carry over" some of the direction from previous steps. This inertia helps the optimization process to navigate through flatter regions of the loss surface more efficiently, potentially leading to faster convergence towards a lower loss.
3.Less Frequent Updates: Adam and AdamW algorithms use more complex mechanisms involving adaptive learning rates and second-moment estimates (RMSprop), which can lead to more frequent updates of the model parameters. In contrast, SGD with momentum can lead to less frequent updates, which may provide the optimization process with a better chance to settle into regions with lower loss values.
4.Applicability to Data: Depending on the nature of the dataset and the specific task, the way gradients are averaged and the learning rate is adjusted can impact the optimization process. SGD with momentum might be better suited to certain datasets and tasks, allowing it to navigate the loss landscape more effectively.
5.Hyperparameter Tuning: The performance of optimization algorithms often depends on hyperparameter settings, such as learning rates, momentum values, and batch sizes. It's possible that for this specific task and dataset, the hyperparameters used for SGD with momentum were more suitable and well-tuned, leading to better convergence.

Notes on LLMs' use

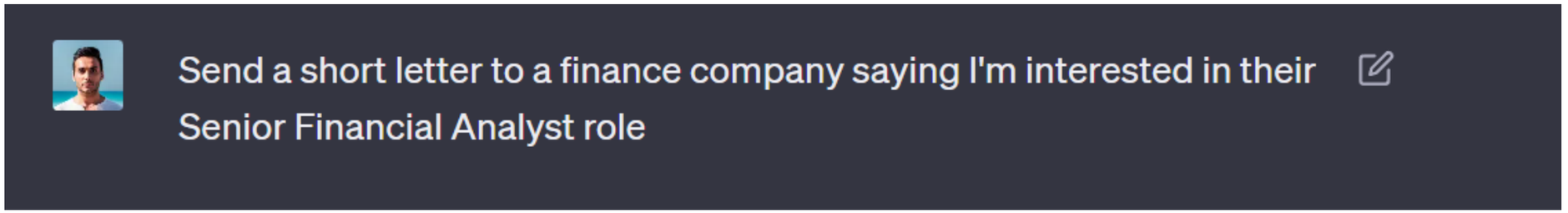
- **HEC Montreal's policy on LLMs: https://www.hec.ca/en/integrity/good-practices-and-ressources/index.html#explo-198752-480086**

If you use content generated by an artificial intelligence tool for an evaluation, you must clearly disclose this by citing your source. Subject to future APA guidance, here is how to quote when using an artificial intelligence tool:

- Template
  - Name of tool used. (YYYY) ChatGPT (version of [day month]) [search question/topic used]. [URL]
- Result
  - OpenAI (2023) ChatGPT (May 27 version) [How to apply the components of Porter's model to the lithium battery industry].

Notes on LLMs’ use

- **ChatGPT can be detected using plagiarism software**
- **ChatGPT is generally**
  1. more verbose than a human
  2. has a unique identifiable style



Notes on LLMs’ use

- **ChatGPT can be detected using plagiarism software**
- **ChatGPT is generally**
  1. more verbose than a human
  2. has a unique identifiable style

Dear [Company Name] Hiring Team,

I hope this message finds you well. I am writing to express my strong interest in the Senior Financial Analyst role at [Company Name], as advertised on your website.

With a solid educational background in finance and [X years] of experience in financial analysis and reporting, I am excited about the opportunity to contribute my skills and expertise to your **esteemed** organization. I have a **proven track record** of leveraging financial data to drive strategic decision-making, streamline processes, and optimize resource allocation.

From my research into [Company Name]'s reputation for innovation and excellence in the finance industry, I am **particularly drawn** to your commitment to leveraging data-driven insights to guide financial strategies. I am confident that my proficiency in financial modeling, forecasting, and risk assessment aligns well with the demands of the Senior Financial Analyst role.

I am impressed by [Company Name]'s **dedication to fostering** a collaborative and dynamic work environment, where professionals are empowered to make meaningful contributions and grow within the company. The opportunity to work alongside a team of experienced financial experts while utilizing cutting-edge tools and technologies greatly appeals to me…