Machine Learning for Large-Scale Data Analysis and Decision Making (MATH80629A) Winter 2022

Week #1

Today

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

Course Introduction & Goals

Logistics

- Course syllabus

  Golnoosh Farnadi: https://gfarnadi.github.io/courses/MLW2022/main.html

  - **(Or Google my name. There's a link from my webpage.)**

Class Website

| MATH80629A | Lectures | Homework | Lab | Project | Office hour

Golnoosh Farnadi

Assistant Professor at HEC Montréal, Adjunct Professor at University of Montréal, Core Academic Member at MILA, Canada CIFAR AI (CCAI) Chair holder

View My GitHub Profile

Hosted on GitHub Pages — Theme by orderedlist

Machine Learning for Large-Scale Data Analysis and Decision Making (MATH80629A): Winter 2022

Course Description

Welcome to MATH80629A Graduate level course on introduction to machine learning at HEC Montreal (English edition). This is the English edition of the course, for the French edition, please check here. In this course, we will study machine learning models, a type of statistical analysis that focuses on prediction, for analyzing very large datasets ("big data"). The plan is to survey different machine learning techniques (supervised, unsupervised, reinforcement learning) as well as some applications (e.g., recommender systems). We will also study large-scale machine learning and will discuss distributed computational frameworks (Hadoop and Spark).

Course Format

Due to the hybrid nature of the semester, this course will be given as a flipped classroom. It is an instructional strategy where students learn the material before they come to the class. The material will be a mix of readings and video capsules. Class time is reserved for more active activities such as problem solving, demonstrations, and questions-answering. In addition, class time will contain a short summary of the week's material.

Time & room

- Wednesdays 8:30 am - 11:30 am

A virtual semester…
(again)

A virtual semester… (again)

- Some things will be a bit better, some a bit worse, and some a bit different

A virtual semester… (again)

- Some things will be a bit better, some a bit worse, and some a bit different
- Let's embrace uncertainty
  - More flexibility than usual
  - Provide feedback (email, office hours or use anonymous feedback form on the website)

Zoom: Rules of engagement

- Microphone turned off by default
- Camera on or off
    - I prefer on, but do what you feel most comfortable with
    - Ideally on when you speak up
- Questions: oral (synchronous) or chat (asynchronous)

Flipped Classroom

- Every week:

1. **Class preparation (Offline):**
   - Watching capsules (60-90 minutes)
   - Background reading (60 minutes)
2. **Class time (Online):**
   - Summary, Q&A, problem solving (90-180 minutes)

Suggestions for navigating an online flipped classroom

- **In class:** Come prepared
  - Watch the capsules ahead of time: Stay active while watching the capsules (e.g., take notes, pause, think of how it fits in the broader context)
  - Do the readings
  - Write down your questions
  - Arrange your desktop with the tools you will need
  - Turn off notifications (emails, social networks, texts, etc.)

Lab sessions

- Most weeks, we have a hands-on session
- We also have one practical lab sessions for Python and ML on (Week #4)

Grading

Your final score for the course will be computed using the following weights:

- Homework (20%)
- Capsule quizes (10%)
- Project (30%)
- Project presentation (10%)
- Final Exam (30%)

Homeworks

- Assignments are to be done individually. You need to submit them to **Gradescope**:
  https://www.gradescope.ca/courses/6006
  Entry name: **9YNNEM**

- One survey assignment, due **January 26, 2022**

- One homework assignment, due **February 21, 2022**

- One case study assignment, due **March 14, 2022**

Quizzes

- 6 multiple answers quizzes based on the capsules
- You need to submit them to **Gradescope** (see previous slide)
- Each Quiz takes 15 mins and will start at 8:30am-8:45am
  (You can start/finish the quiz earlier. All the quizzes will be available from 8 am)

  Quiz 0 (test quiz): **January 12, 2022**
  Quiz 1: **January 19, 2022**
  Quiz 2: **February 2, 2022**
  Quiz 3: **February 9, 2022**
  Quiz 4: **February 16, 2022**
  Quiz 5: **March 9, 2022**
  Quiz 6: **March 30, 2022**
- Your best 5 quiz scores will be used to compute your semester quiz score (10% of your final grade).

Project

- You must work in teams of <u>two or three</u>.
- You can find teammates using **Piazza**:
  Signup link: https://piazza.com/class/ky0k08xrdis7he
  Access code: machinelearningforlargescaledataanalysisanddecisionmaking
- Team Registration, due: **January 26, 2022**
- Study plan, due: **February 27, 2022**
- Project meeting, on **March 2, 2022**
- Project Presentation, on: **April 6, 2022**
- Final report, due: **April 30, 2022**

Getting Help

- Post your questions in **Piazza** and hopefully your peers will answer.

  My office hour:
  Zoom link: https://hecmontreal.zoom.us/j/81836582494?pwd=VVhvWk1rYVFLdGJzTldLZzYyc0VvQT09
  Meeting ID: 818 3658 2494
  Passcode: 379543

  Wednesdays 11:30 am - 12:30 pm

  my email:
  golnoosh.farnadi@hec.ca
  farnadig@mila.quebec

Getting Help

- Teaching Assistant:
  Pravish Sainath (Speaks French & English)
  email: pravish.sainath@umontreal.ca

- Office hour: TBA

- Zoom link: TBA

Feedback

Please use this **form** to provide feedback about the course.
https://forms.gle/VVNQogf2fBi9tKq38

If you would like to remain anonymous, you can choose not to write your name and email address below.

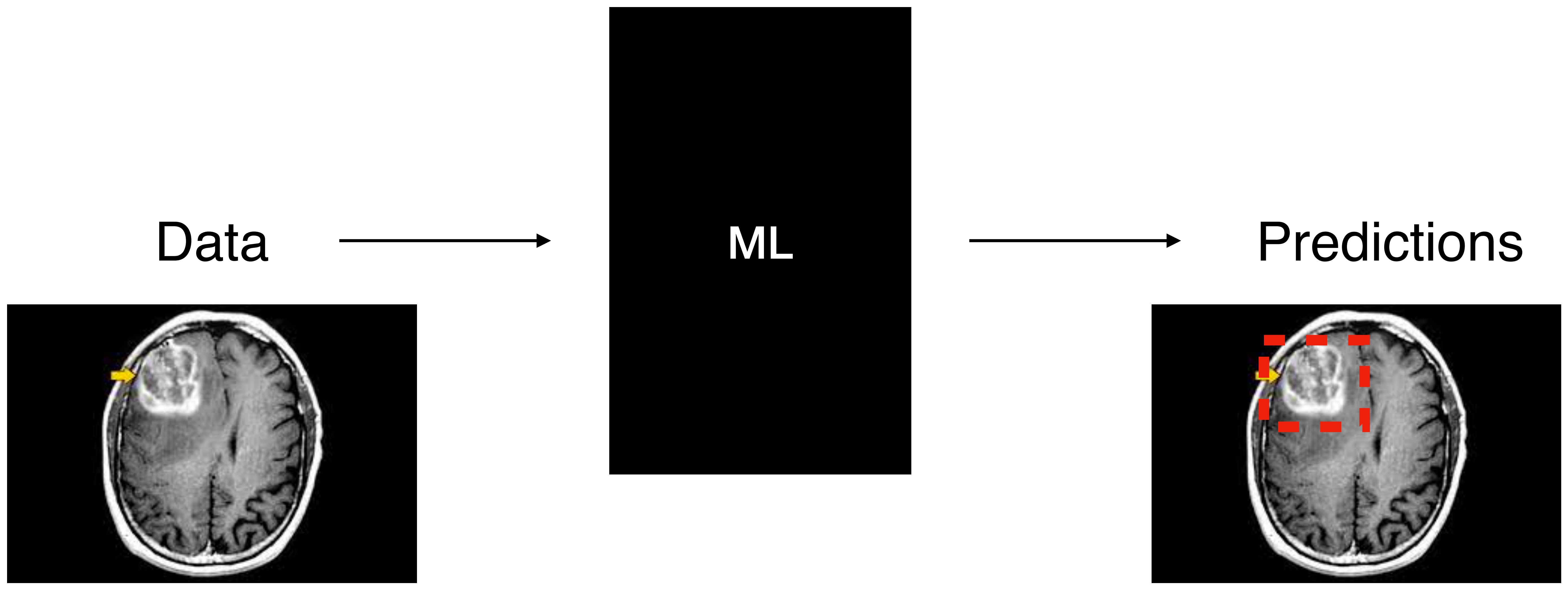
I am looking forward to hearing from you!

Introduction to machine learning

HEC MONTRÉAL

Machine Learning (ML)

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

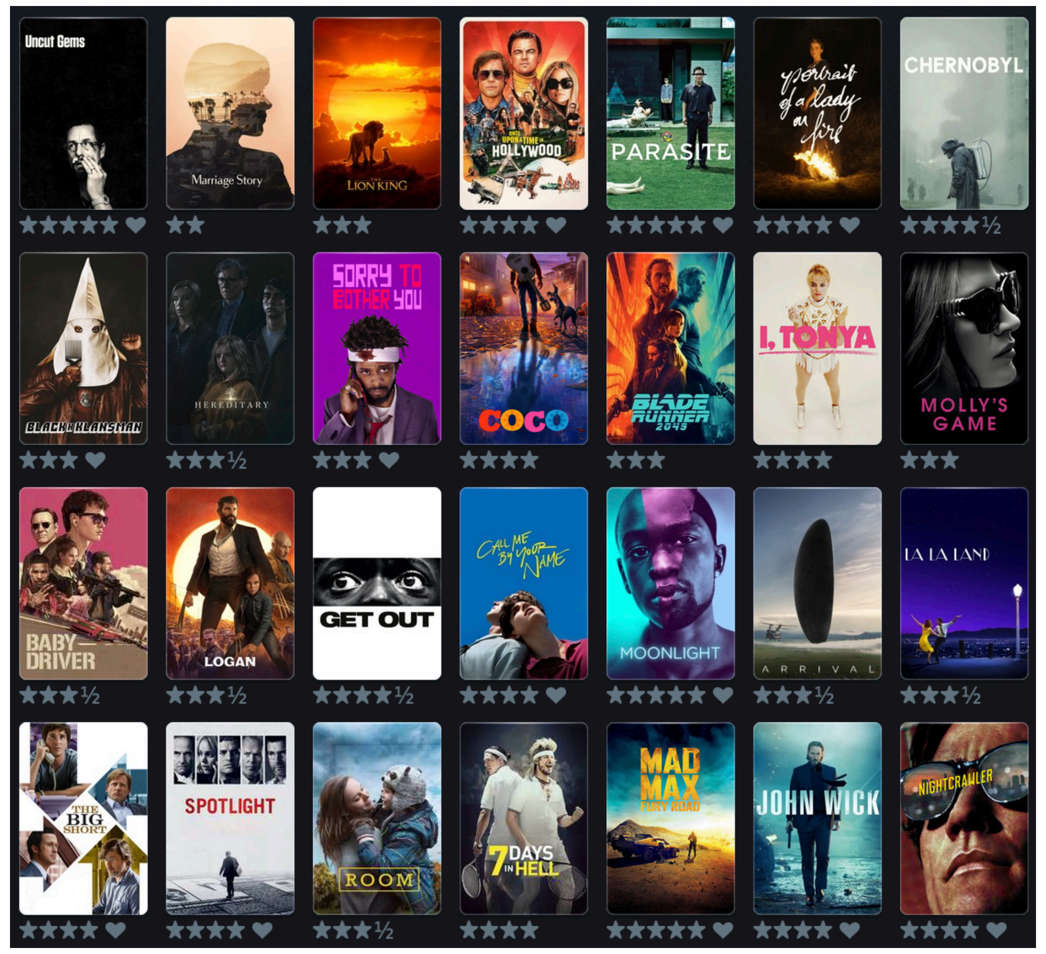


MRI image

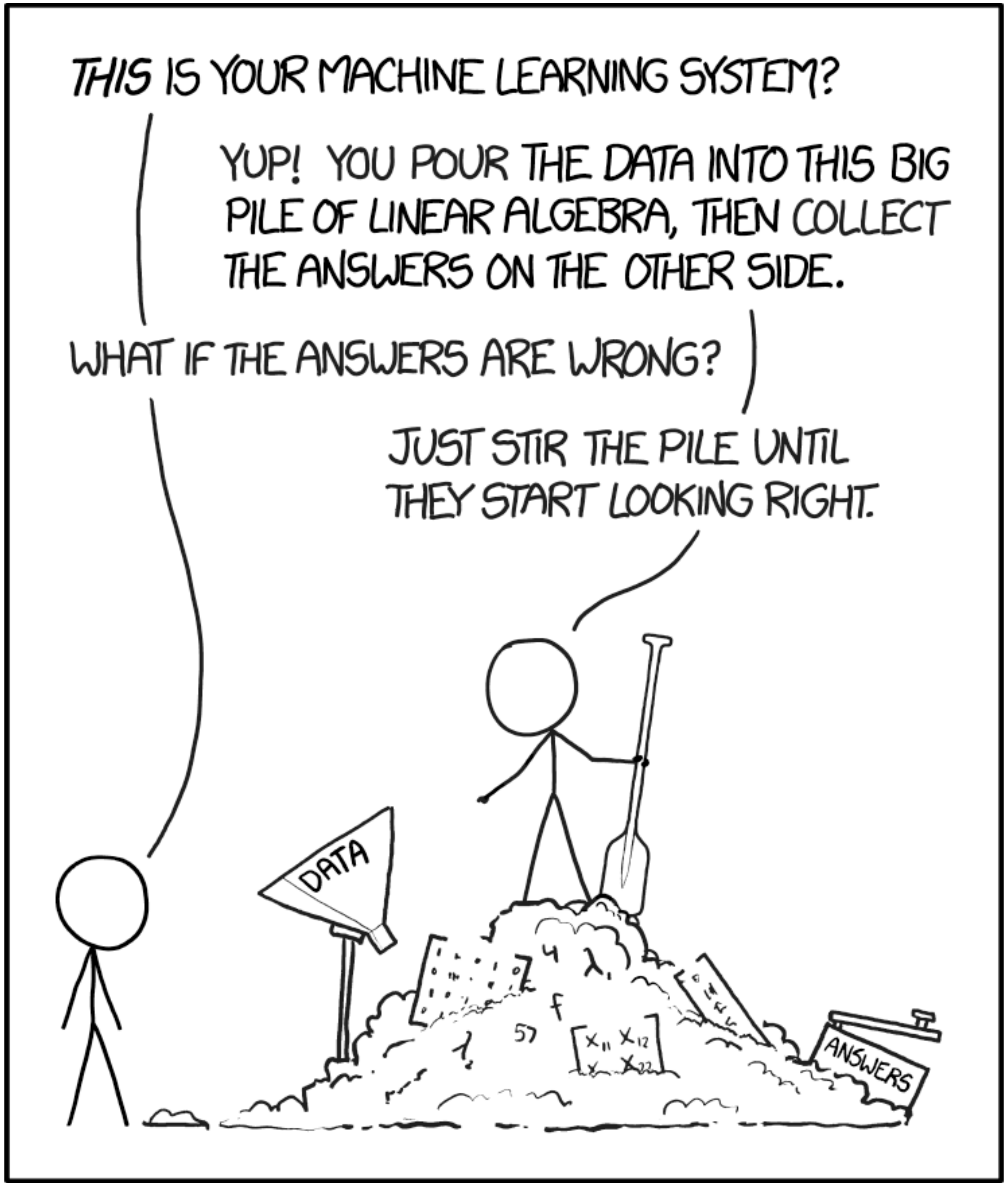
Brain tumor

- Task: Predict whether an image contains a tumor

Data → ML → Predictions

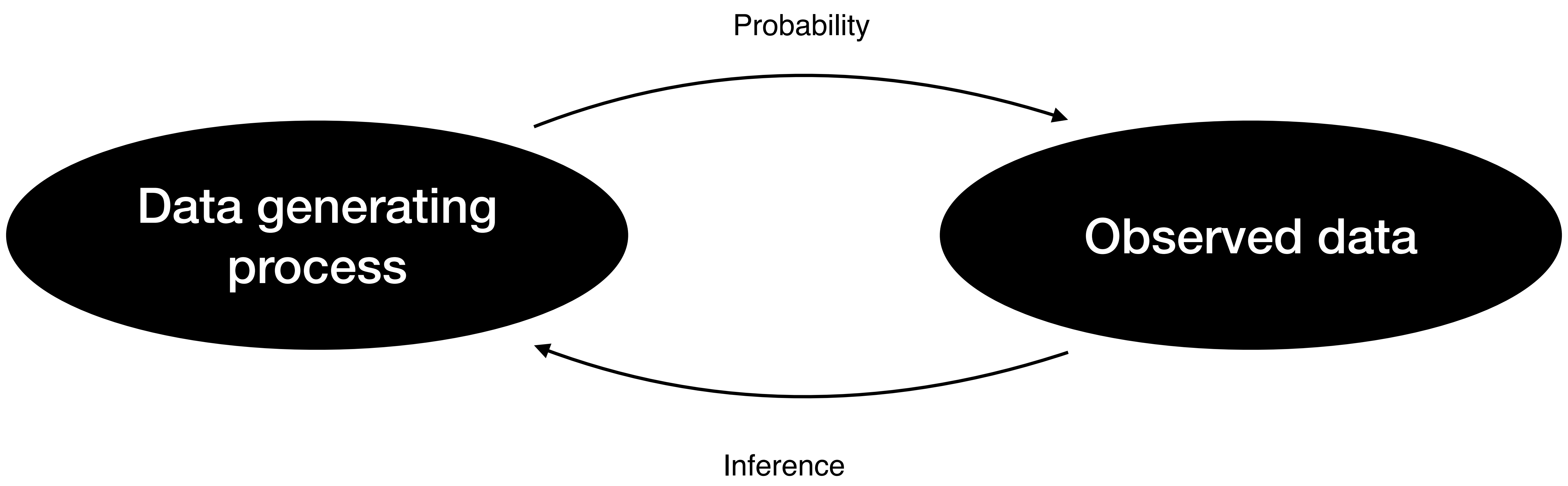


- Task: Predict the next movie a person should watch



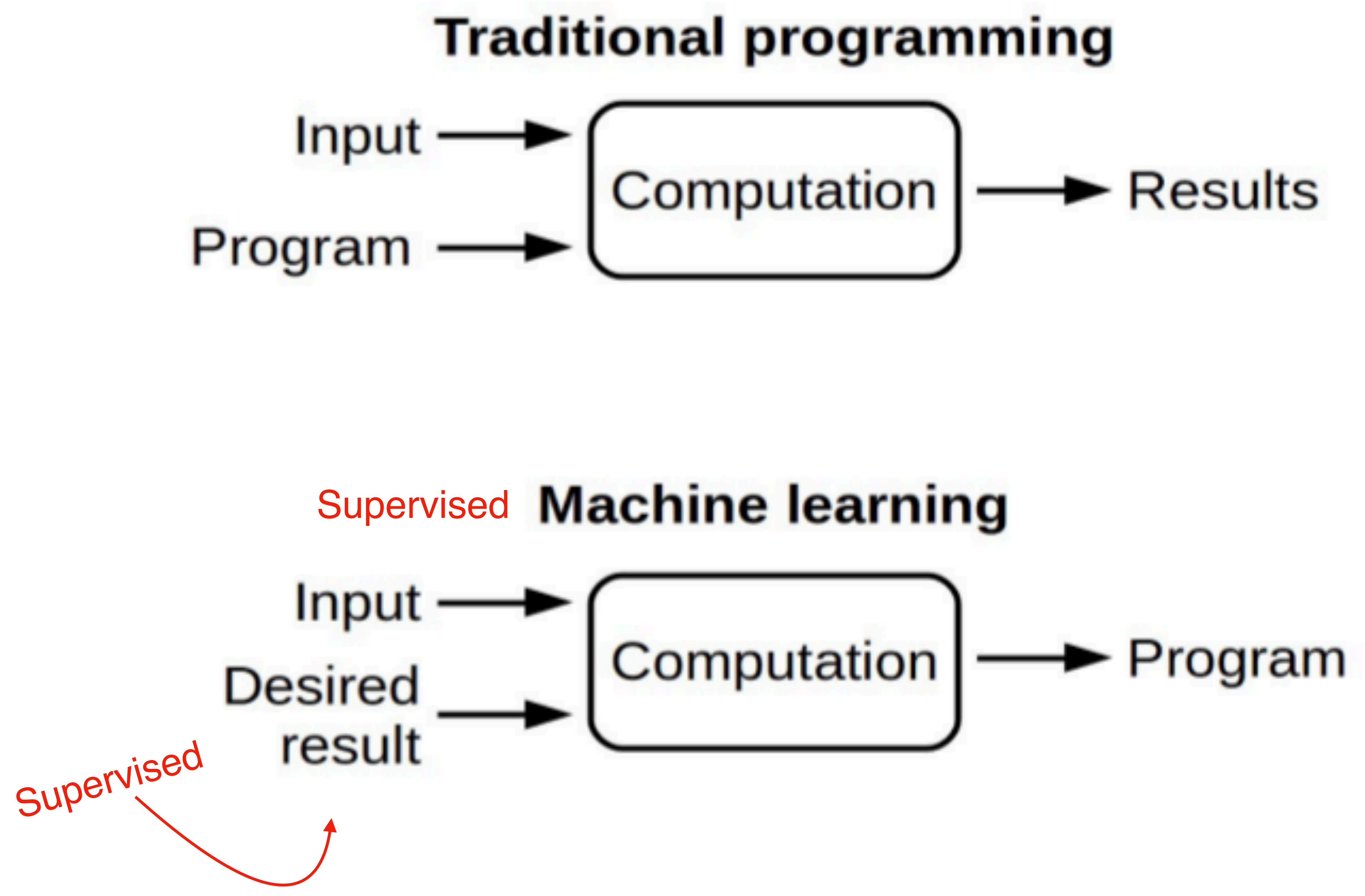
https://xkcd.com/1838/

“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?


Traditional programming
Input
Program
Computation
Results
Supervised
Machine learning
Input
Desired result
Computation
Program
Supervised

Practical Benefits of ML

- Reduce time of programming
- Customize and Scale products
- Complete seemingly “unprogrammable” tasks

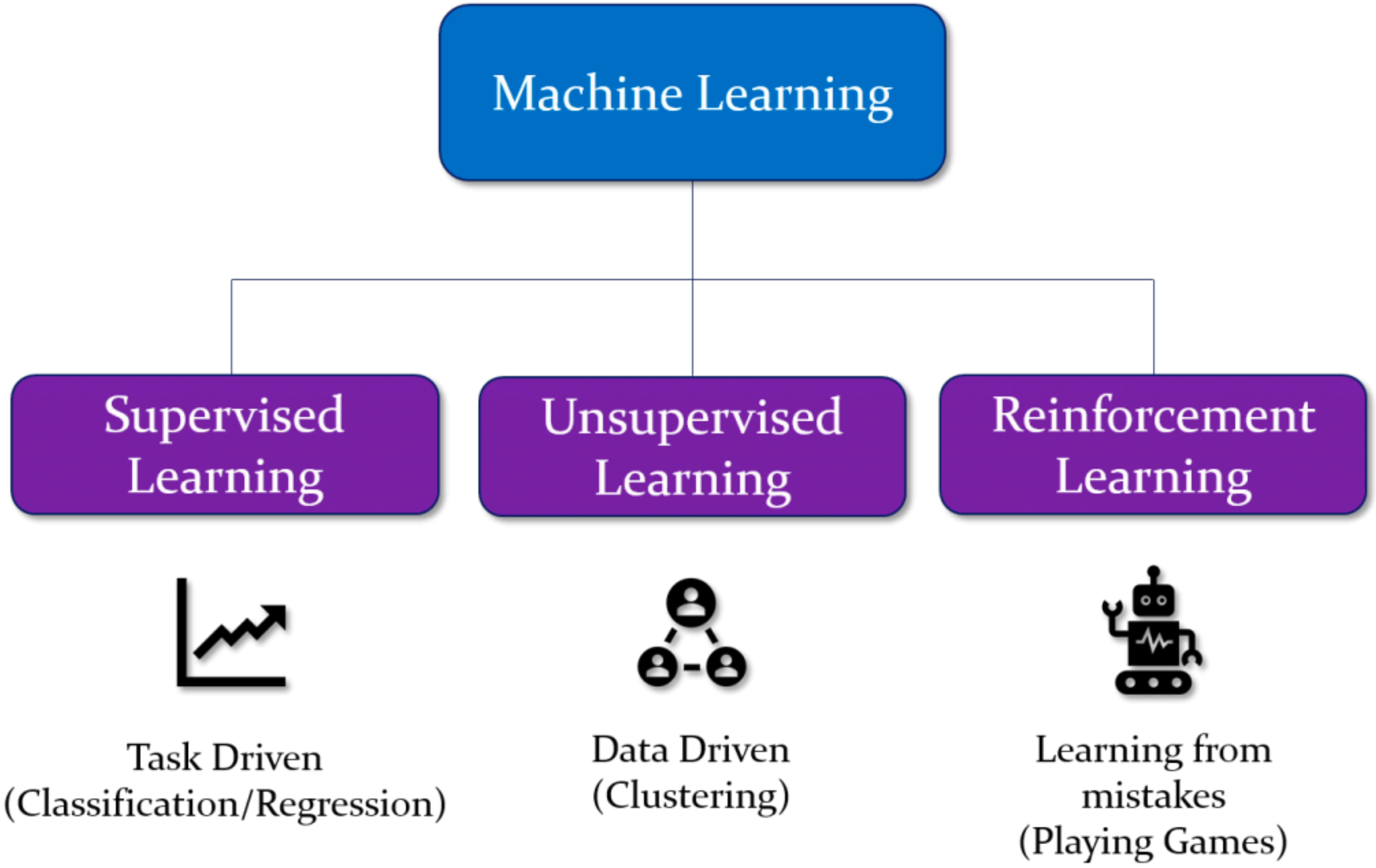
Philosophical Benefits of ML

- Shift from mathematical science to natural science
- Ability to think like scientists

- From Machine Learning Crash Course

The goal of ML is to **understand** the nature of (human and other forms of) **learning**, and to **build learning capability** in computers.

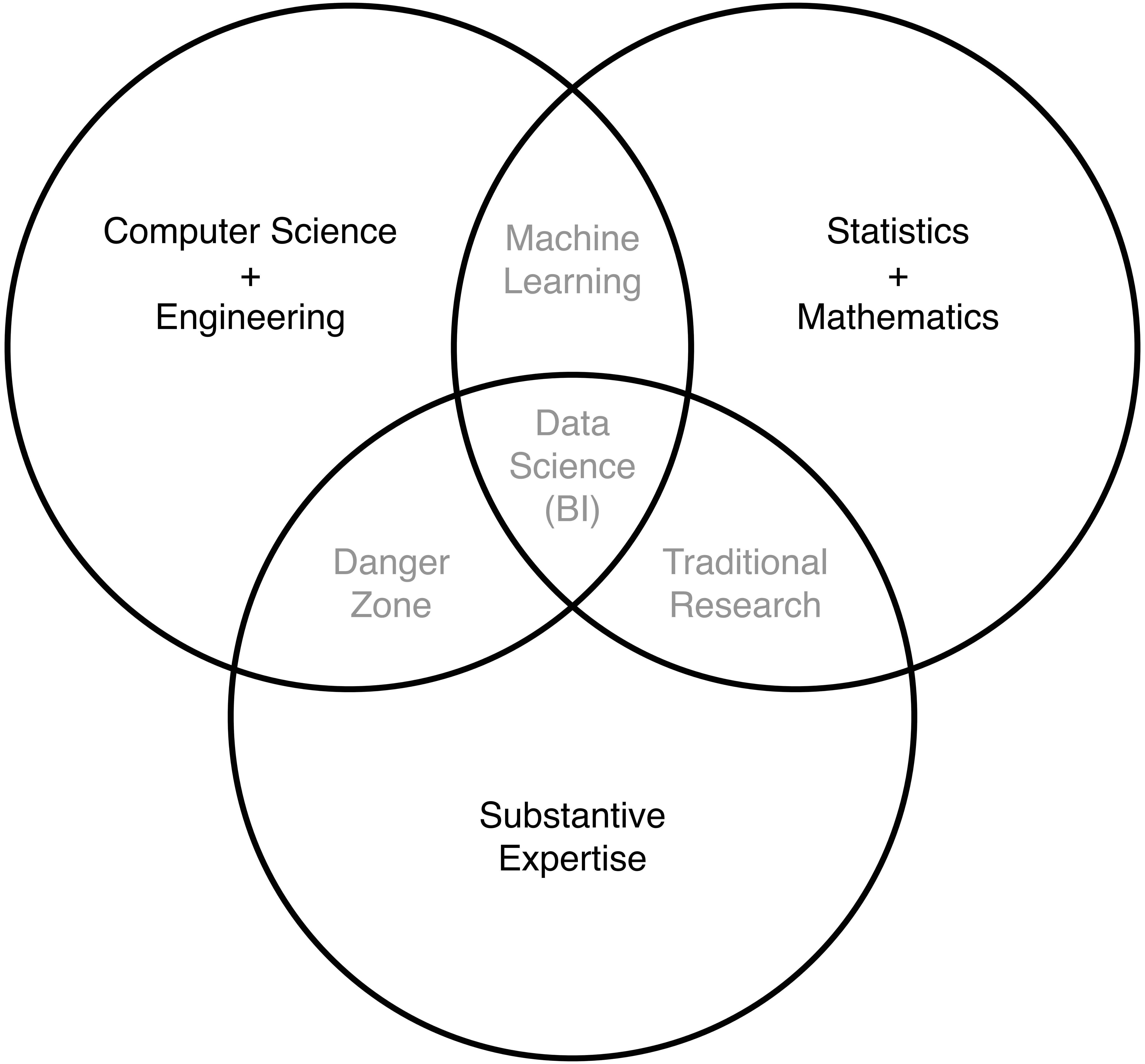
Types of Machine Learning



How does ML relate to other fields

Historical View

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



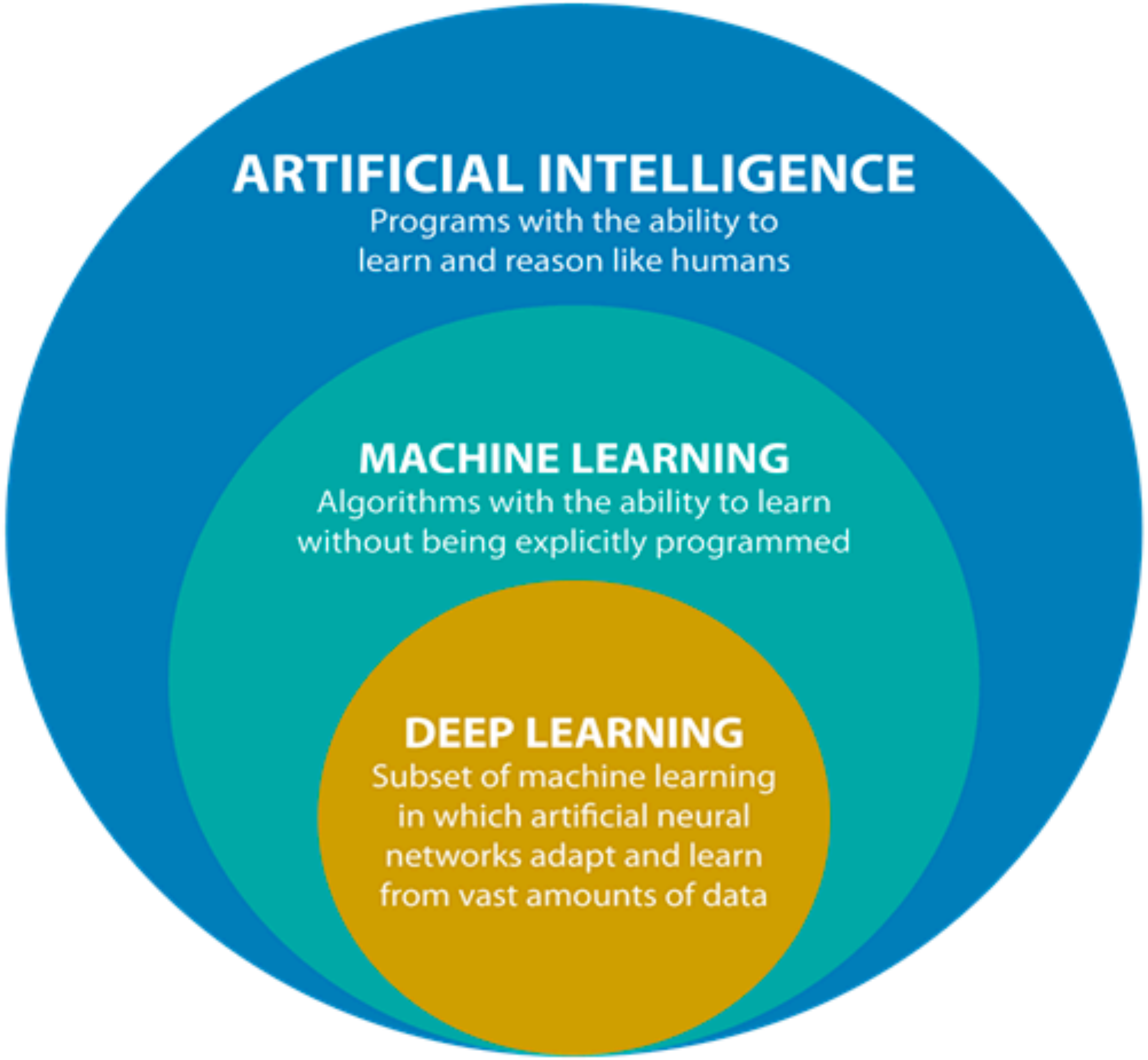
[Inspired by Drew Conway]

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. . .

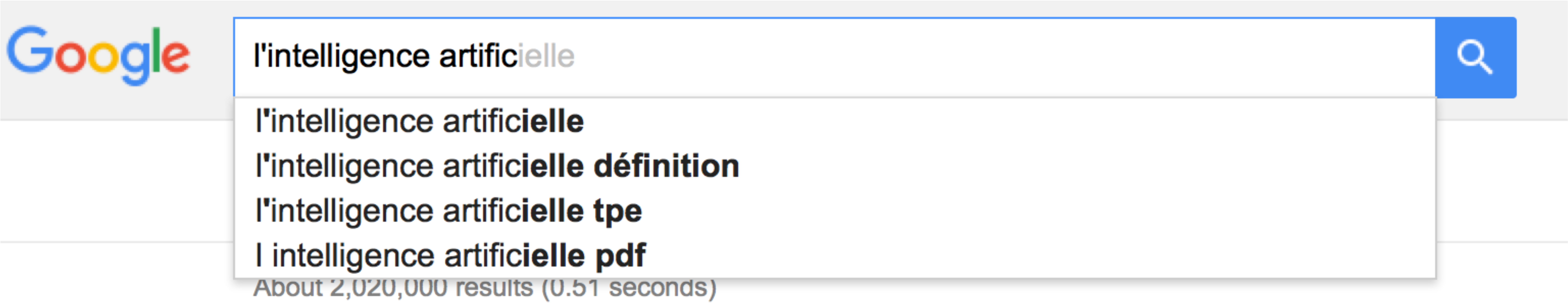
Machine learning	**Traditional statistics**
No settled philosophy or widely accepted theoretical framework.	Classical (frequentist) and Bayesian philosophies compete.
Willing to use *ad hoc* methods if they seem to work well (though appearances may be misleading).	Reluctant to use methods without some 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.

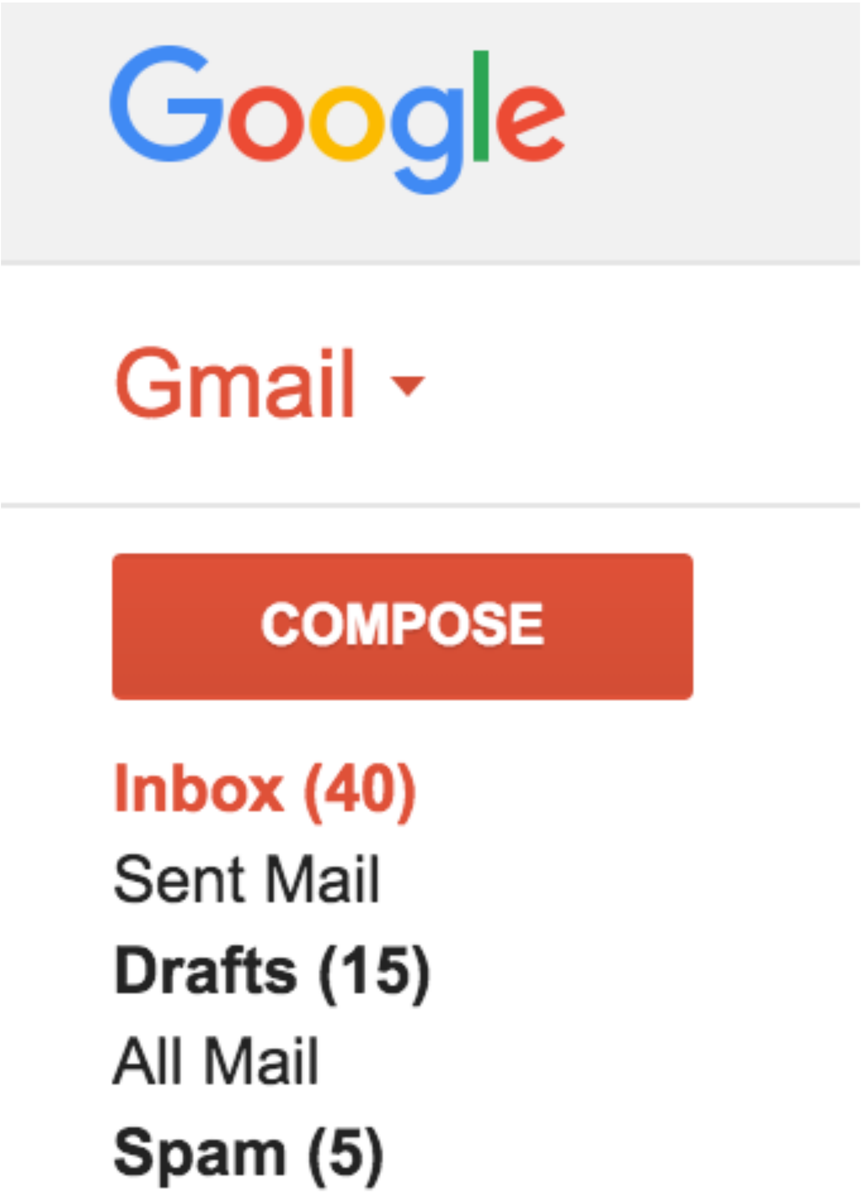
[Reproduced from R. Neal: http://www.utstat.utoronto.ca/~radford/sta414.S14/week1.pdf]

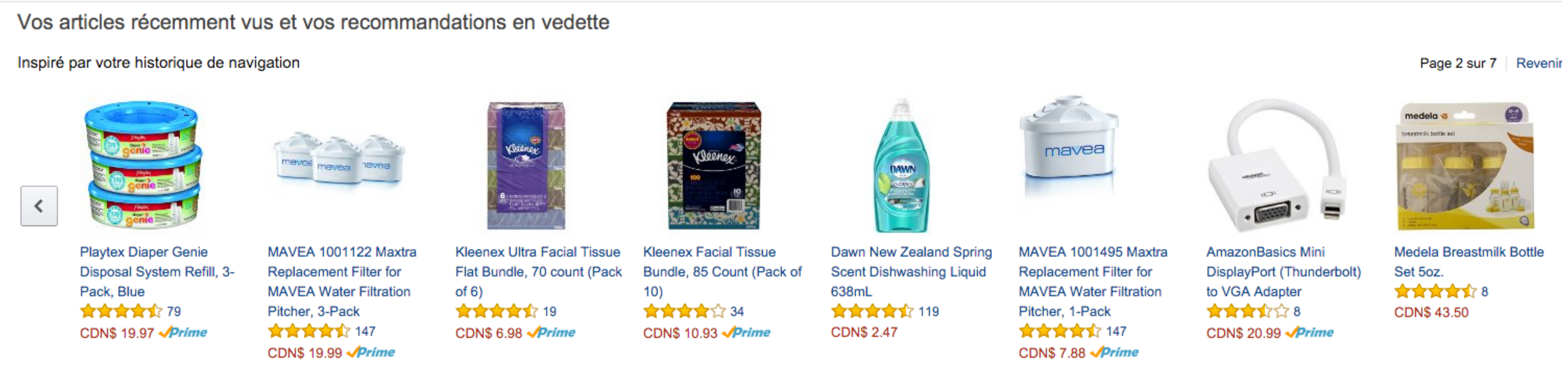


- Recently, ML models make progress toward “AI tasks”
  - Examples of AI tasks: translation, object recognition
  - In that context: create a machine with human-like capacities? Or a machine that can help humans?

Applications of ML


Google
l'intelligence artificielle
l'intelligence artificielle
l'intelligence artificielle définition
l'intelligence artificielle tpe
l intelligence artificielle pdf
About 2,020,000 results (0.51 seconds)

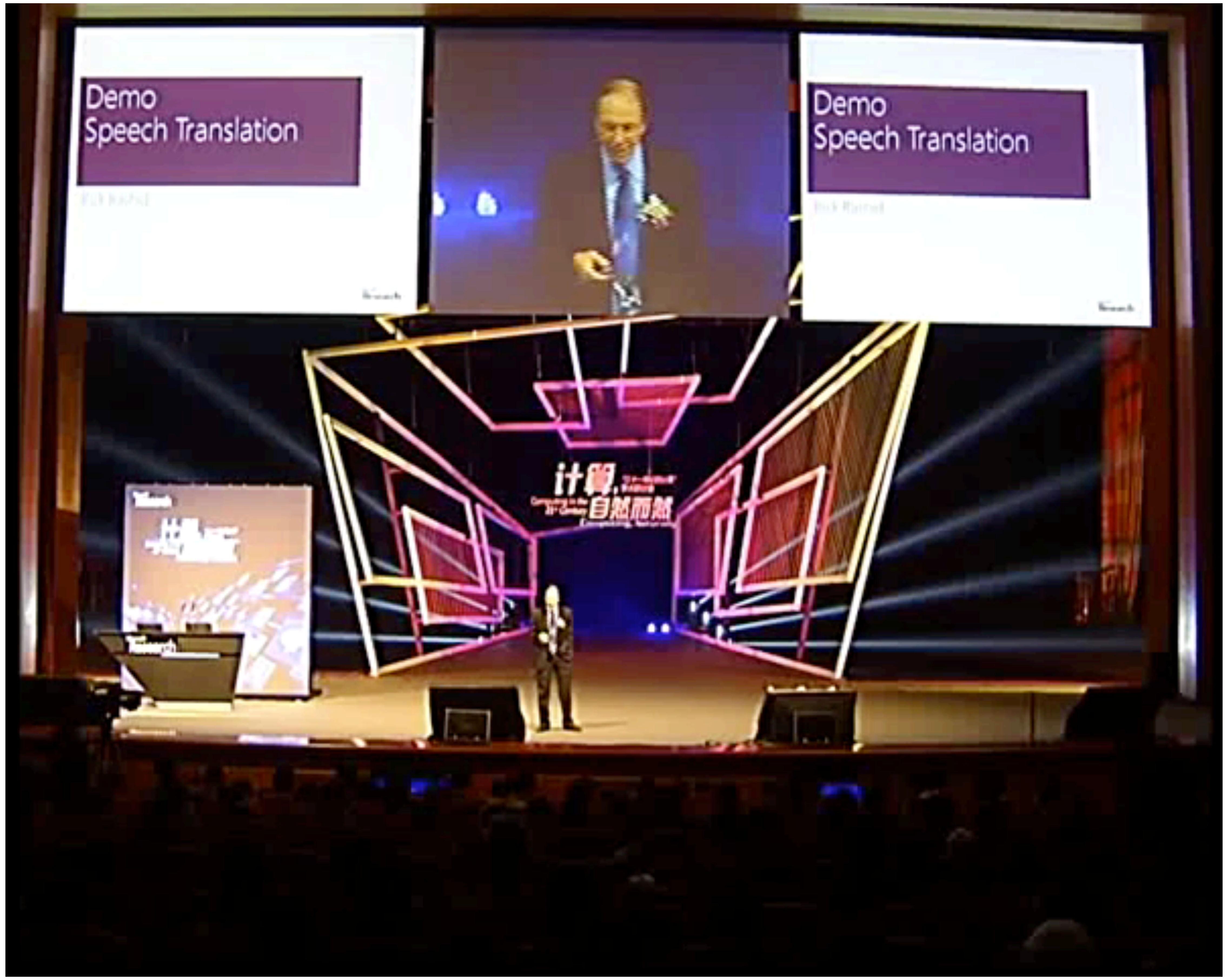

Google
Gmail
COMPOSE
Inbox (40)
Sent Mail
Drafts (15)
All Mail
Spam (5)


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

In **2009**, Google started the self-driving car project with the goal of driving autonomously over ten uninterrupted 100-mile routes.

In **2009**, Google started the self-driving car project with the goal of driving autonomously over ten uninterrupted 100-mile routes.

Microsoft Research
2012
Speech Recognition Breakthrough for the Spoken, Translated Word

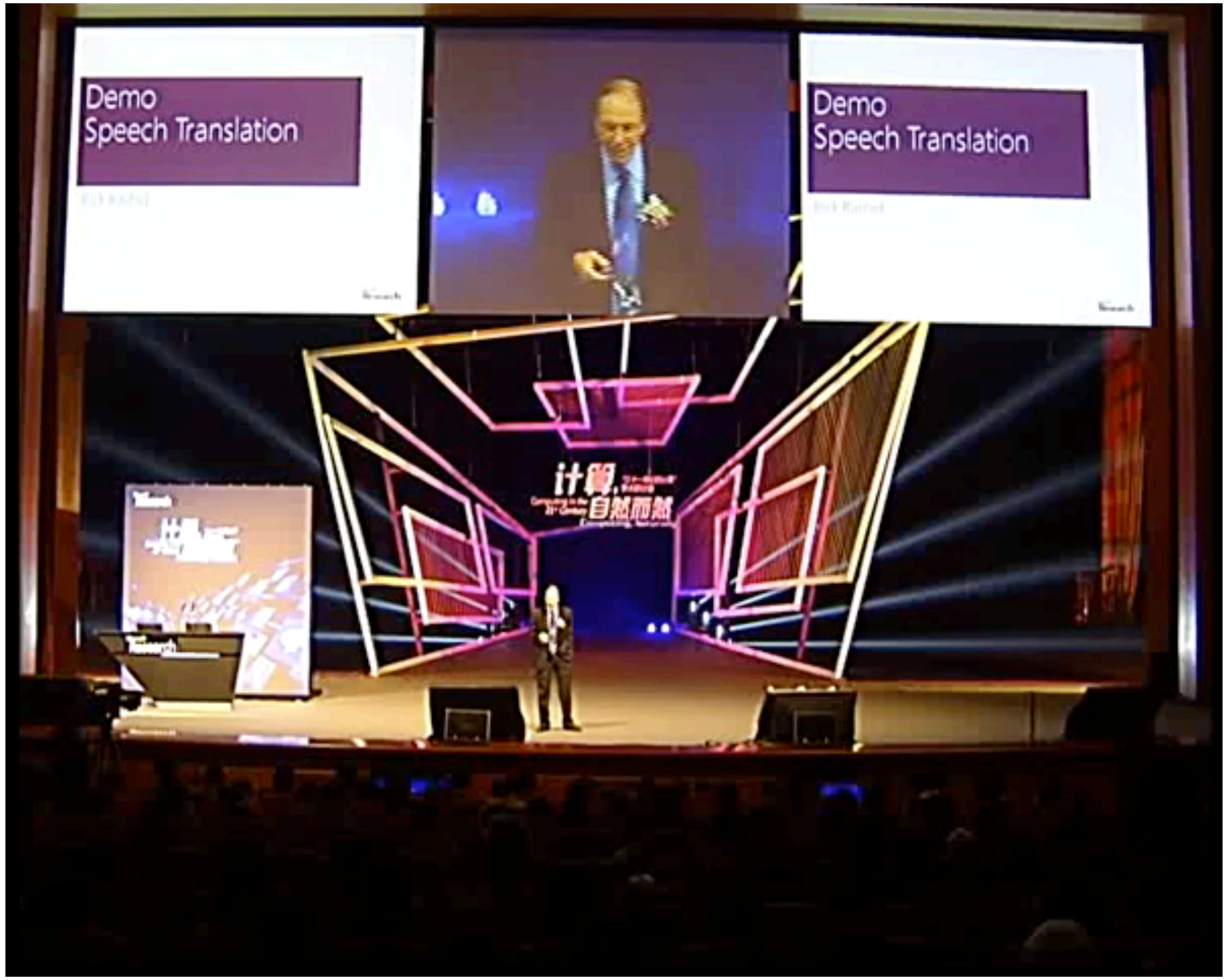


https://youtu.be/Nu-nlQqFCKg

Microsoft Research

2012

Speech Recognition Breakthrough for the Spoken, Translated Word

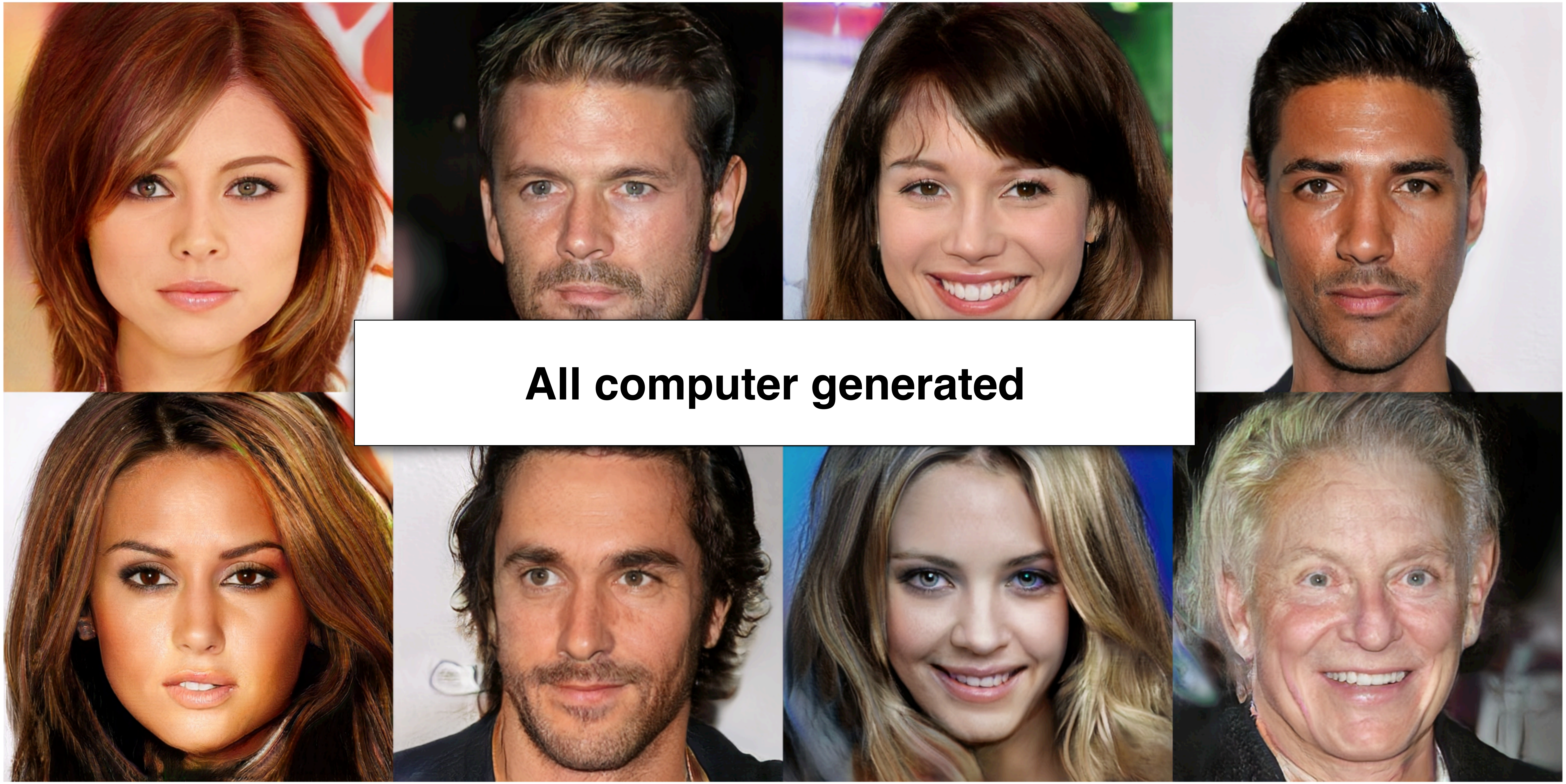


https://youtu.be/Nu-nlQqFCKg

The cover of the January 28, **2016** issue of *Nature*, which features Google's groundbreaking AI research.

The cover of **2017** issue of *Science*, which features DeepStack's groundbreaking AI research.

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



All computer generated

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

HEC MONTRÉAL


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."

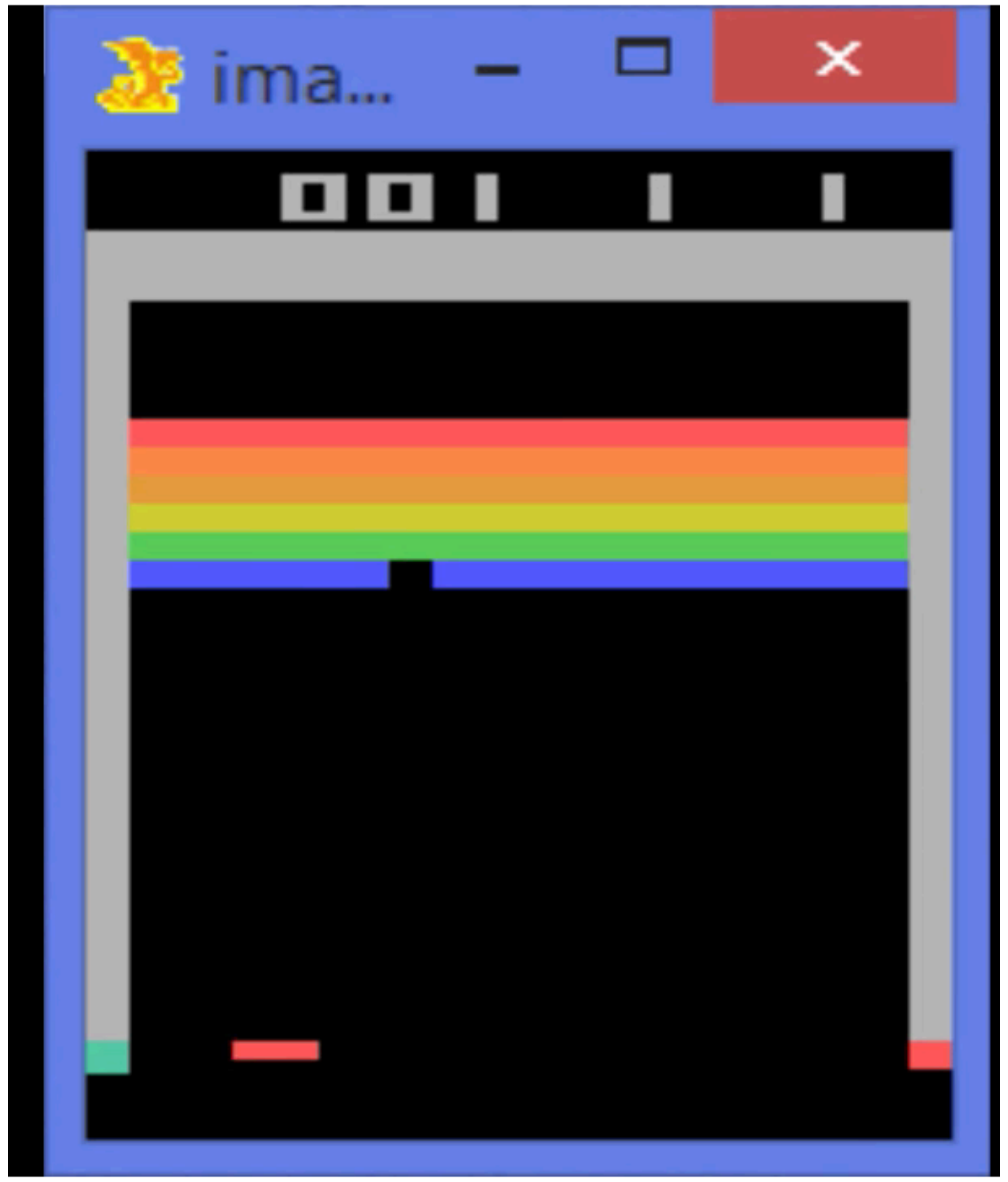
"black and white dog jumps over bar."

"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

AI for video games



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

Google DeepMind created an artificial intelligence program using deep reinforcement learning that plays Atari games and improves itself to a superhuman level

- **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 Structure

- **ML Fundamentals:** Week 2
- **Supervised Learning:** Week 3 & 4
- **Deep learning:** Week 5 & 6
- **Unsupervised Learning:** Week 7
- **Large scale computing :** Week 10
- **Trustworthy ML and Recommender Systems:** Week 11
- **Reinforcement Learning:** Week 12 & 13

Fit with other courses

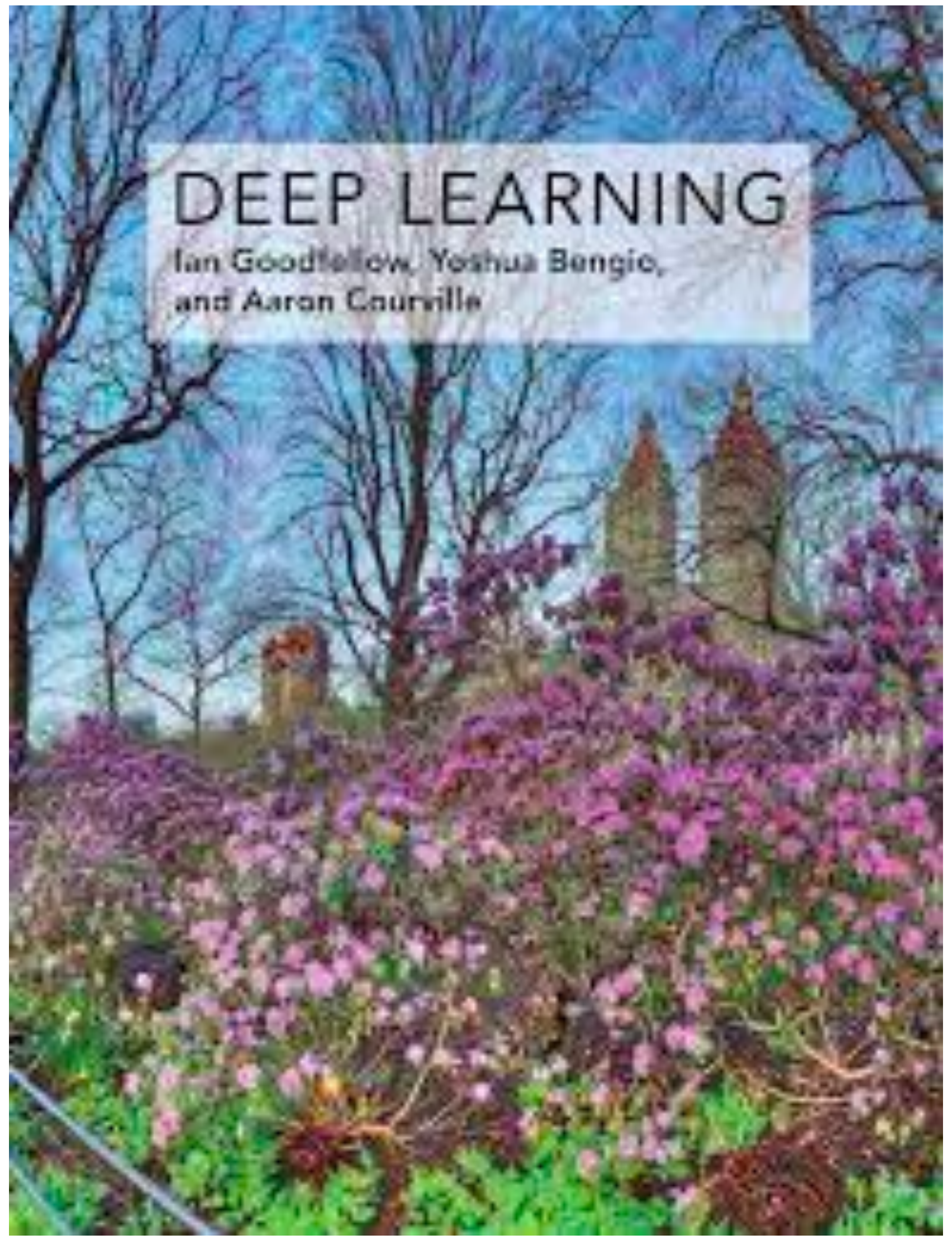
- HEC
  - PhD level (originally)
  - Computationally oriented
  - Prequel to Machine Learning II: Deep Learning
- Other ML courses in Montreal (U.Montreal, Polytechnique, Mcgill)
  - More applied (similar to COMP-551@McGill)

Short review of linear algebra, statistics, and probabilities

Recap: Linear algebra

- Based on chapters 2 of “Deep Learning”

  http://www.deeplearningbook.org/



Linear algebra

- Scalar: a single value.

$$\mathbf{a} \in \mathbb{R}, \mathbf{a} \in \mathbb{N} \qquad \mathbf{a} = 3$$

- Vector: an array of values.

$$\mathbf{a} \in \mathbb{R}^{D}, \mathbf{a} \in \mathbb{N}^{D} \qquad \mathbf{a} = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix}$$

- Matrix: a table of values.

$$\mathbf{A} \in \mathbb{R}^{D_1 \times D_2}, \mathbf{A} \in \mathbb{N}^{D_1 \times D_2} \qquad \mathbf{A} = \begin{bmatrix} 3 & 4 & 2 \\ 1 & 2 & 9 \end{bmatrix}$$

Indexing notation

- Indexing elements of a vector: a_i

$$a = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix} \begin{matrix} \\ \longleftarrow a_1 \\ \\ \end{matrix}$$

Convention:
The first element
is the zero'th.

- Indexing elements of a matrix: a_{ij}

$$A = \begin{bmatrix} 3 & 4 & 2 \\ 1 & 2 & \underset{\uparrow \atop a_{12}}{9} \end{bmatrix}$$

Simple operations

- Transpose

$$\mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

$$\mathbf{a}^\top = \begin{bmatrix} a_0 & a_1 & a_2 \end{bmatrix}$$

$$(\mathbf{A}_{ij})^\top = \mathbf{A}_{ji}$$

- Addition
  - Vectors and matrices w. the same shape

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \qquad \mathbf{a} + \mathbf{b} = \begin{bmatrix} a_0 + b_0 \\ a_1 + b_1 \\ a_2 + b_2 \end{bmatrix}$$

$$(\mathbf{A} + \mathbf{B})_{ij} = \mathbf{A}_{ij} + \mathbf{B}_{ij}$$

HEC MONTRÉAL

Simple operations

- Multiply by a scalar

$$\alpha \mathsf{a} = \begin{bmatrix} \alpha\ \mathsf{a}_0 \\ \alpha\ \mathsf{a}_1 \\ \alpha\ \mathsf{a}_2 \end{bmatrix}$$

- Vector product.
  - The dot product $\mathsf{a}^\top \mathsf{a} = \sum_i \mathsf{a}_i \mathsf{a}_i$
    - Note: it yields a scalar.
  - Element-wise product:

$$\mathsf{a} \odot \mathsf{a} = \begin{bmatrix} \mathsf{a}_0 \mathsf{a}_0 \\ \mathsf{a}_1 \mathsf{a}_1 \\ \mathsf{a}_2 \mathsf{a}_2 \end{bmatrix}$$

    - Also known as Hadamard product

Operations

- Matrix product (dot product):

$$\mathsf{C_{ij}} = \sum_{\mathsf{k}} \mathsf{A_{ik} B_{kj}}$$

  - A's columns must equal B's rows (order is important)

$$\mathbf{A} \in \mathbb{R}^{D_1 \times D_2}, \mathbf{B} \in \mathbb{R}^{D_2 \times D_3}$$

  - Distributive: $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
  - Associative: $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$
  - Product of transpose: $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$

Inverse

- We denote a matrix's inverse as A^{-1} ➡ $A^{-1}A = I_n$
- Identity. Denoted I_n.
    - All zeros except for ones on the main diagonal.

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Inverse

- We denote a matrix's inverse as A^{-1} $\Rightarrow$ $A^{-1}A = I_n$
- A matrix has an inverse iff:
  - it's square. $D_1 = D_2$
  - its columns are linearly independent.
    - No column can be recovered using a combination of other columns
- Inverses are useful to solve systems of equations:

A square matrix not invertible is *singular*

$$Ax = b \qquad x = A^{-1}b$$

Norms

- L^pnorm. Size of a vector (or matrix)

$$\| \mathbf{a} \|_p = \left(\sum_i |a_i|^p \right)^{1/p}$$

- Standard norms in ML:
  - Euclidean norm (p=2) $$\| \mathbf{a} \|_2 = \sqrt{\left(\sum_i |a_i|^2 \right)}$$
  - Dot product w. 2-norm: $$\mathbf{a}^\top \mathbf{b} = ||\mathbf{a}||_2 ||\mathbf{b}||_2 \cos \theta_{ab}$$
  - Frobenius norm (matrix): $$\| A \|_2 = \sqrt{\left(\sum_i \sum_j |a_{ij}|^2 \right)}$$

Special matrices & vectors

- Diagonal: D is diagonal iff $D_{ij} = 0$ for all $i \neq j$
- Symmetric: $\mathbf{A} = \mathbf{A}^\top$
- Unit vector: $\|\mathbf{a}\|_2 = 1$
- Orthogonal vectors: $\mathbf{a}^\top \mathbf{b} = 0$
- Orthonormal vectors: unit and orthogonal
- Orthogonal matrix: Orthonormal rows & columns $\mathbf{A}^\top \mathbf{A} = \mathbf{A}\mathbf{A}^\top = \mathbf{I}$

$$\mathbf{A}^{-1} = A^T$$

- Skip eigendecomposition, SVD, pseudo-Inverse, determinants (Sections 2.7–2.11).
  - We will get back to them if/when needed in the course.

Recap: Probability

- Chapter 3 of “Deep Learning”
  - Some of the lecture slides are adapted from the book.
  - Thanks to Ian Goodfellow for providing the slides.

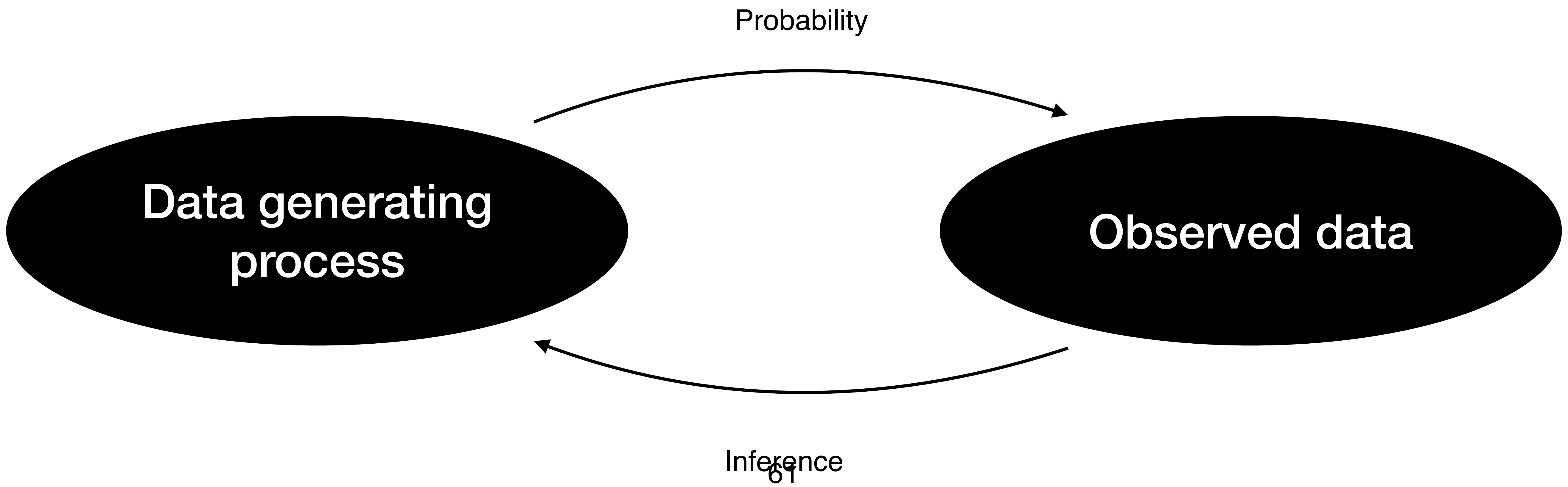
Why probabilities?

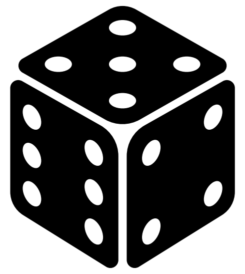
- To capture uncertainty

  E.g.,When will the COVID-19 pandemic end?

- Probabilities provide a formalism for making statements about "data generating processes" (L. Wasserman)

  E.g., what happens when I flip a fair coin?



The  example

- Generate data by throwing a fair die.
- What do we know about a single throw?
  - 6 possible outcomes. (sample space)
  - Each outcome (e.g., 1). (element, state)
  - A subset of outcomes (e.g., <3). (event)
  - Outcomes are equiprobable. (uniform distribution)

Random variables and probabilities

- A random variable (r.v.) is a probabilistic outcome.
  - For example,
    - Die throw: x
    - The actual outcome is $x \in \{1, 2, 3, 4, 5, 6\}$

- A probability function (P) assigns a real number to each possible event: $P(x) \geq 0, \forall x \in X$

$$P(\bigcup x) = 1$$

Discrete RVs

- An RV is **discrete** if it takes a finite number of values[1]

$$P(x = x_i) \geq 0, \forall i$$

$$\sum_i P(x = x_i) = 1$$

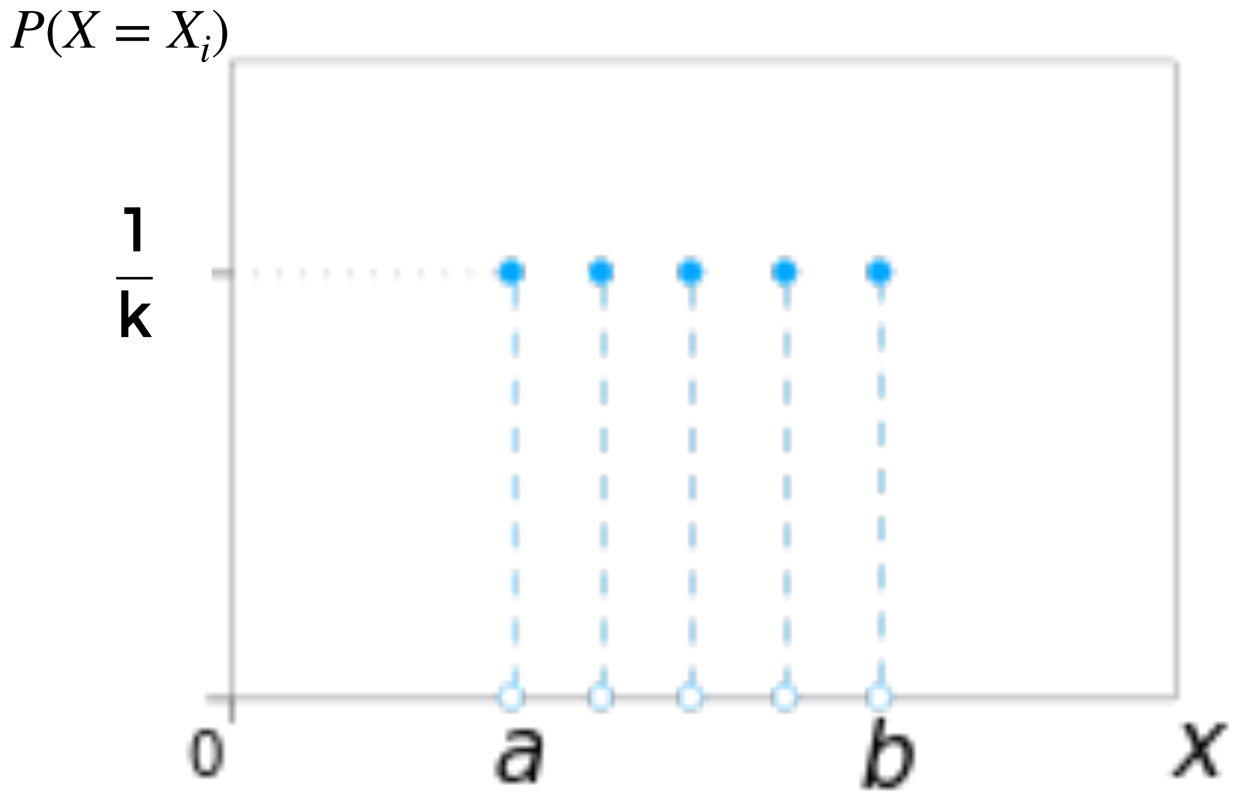
- Probability distribution over discrete variables are described using a probability mass function (PMF).

1. technically: it must be countable

Discrete RVs

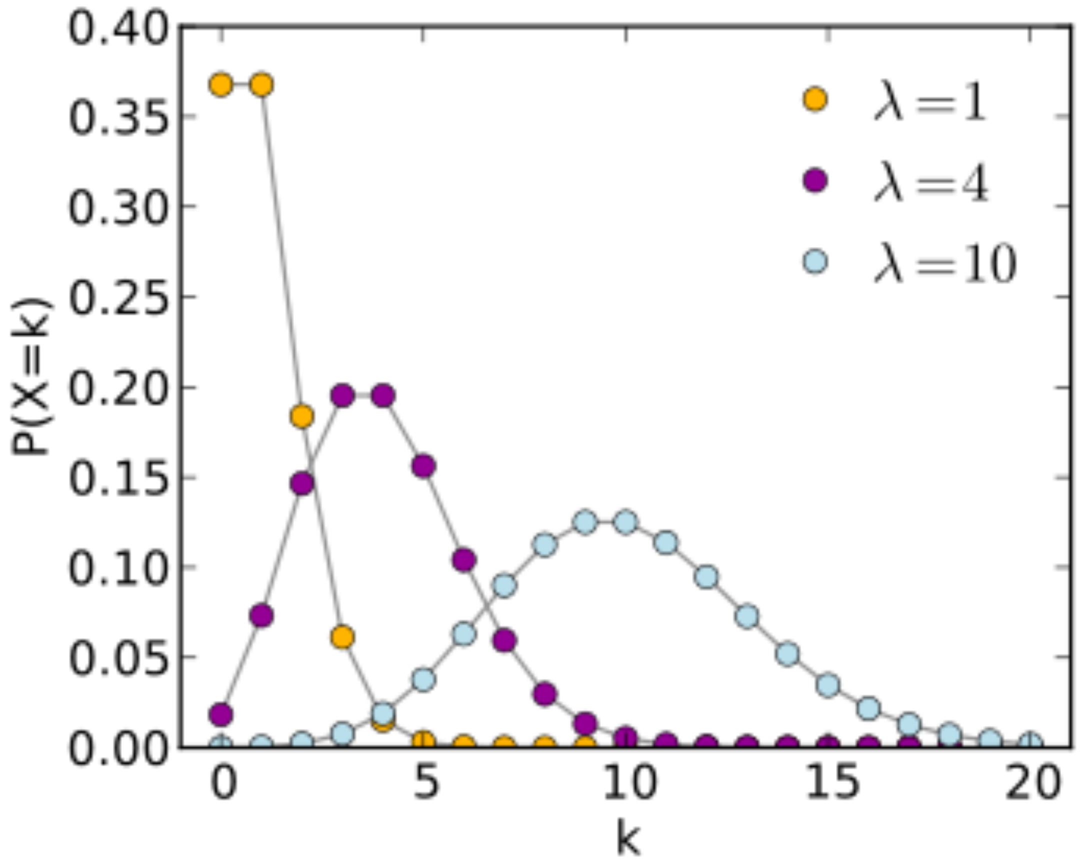
- E.g., uniform distribution:

$$P(x = x_i) = \frac{1}{k}, \forall i$$



- E.g., Poisson distribution:

$$P(x = x_i; \lambda) = \frac{\lambda^{x_i} \exp^{-\lambda}}{x_i!}$$



both images are from: wikipedia.org

Continuous RVs

- An RV is **continuous** if $f(x) \geq 0, \forall x \in X$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

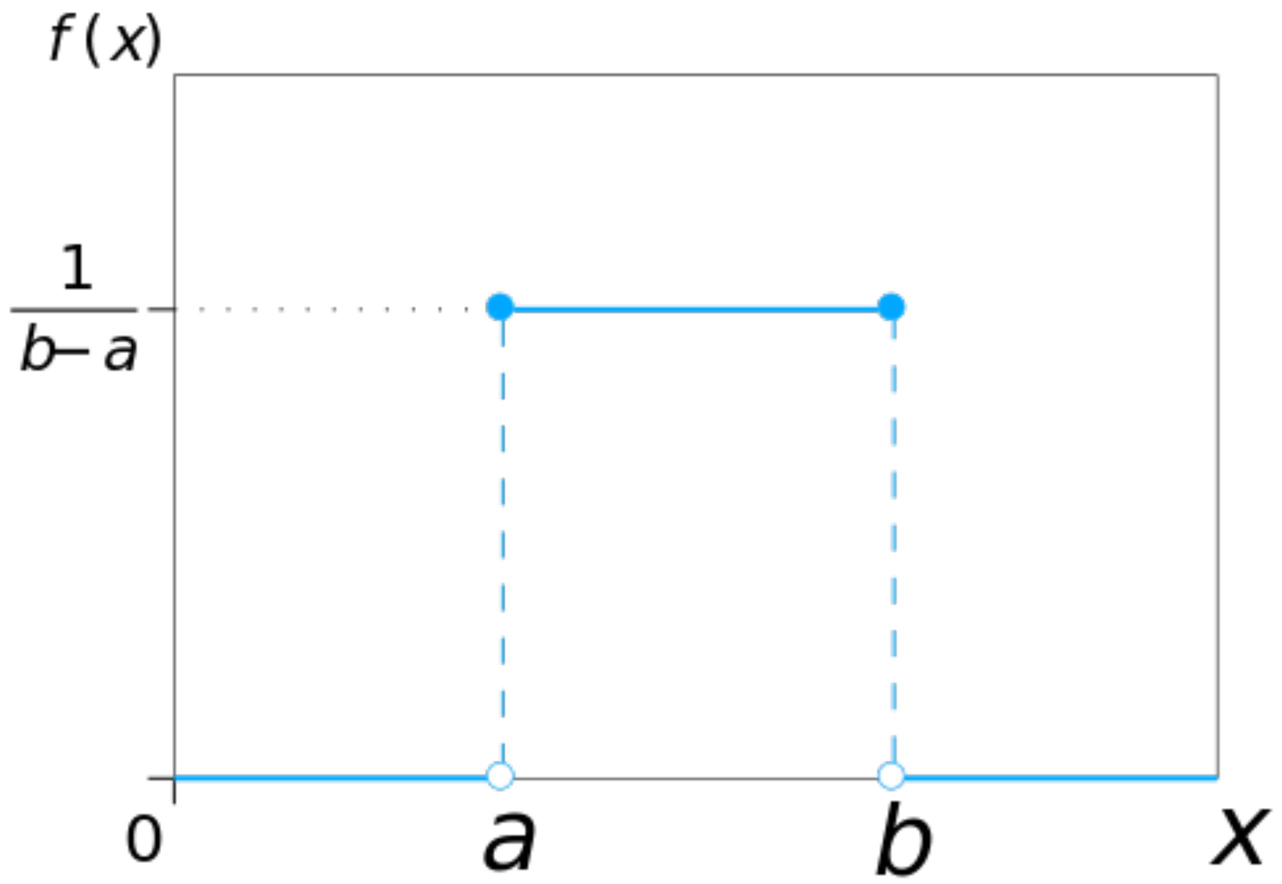
$$P(a < x < b) = \int_{a}^{b} f(x)dx$$

- f(x) is a probability density function (PDF)

Continuous RVs

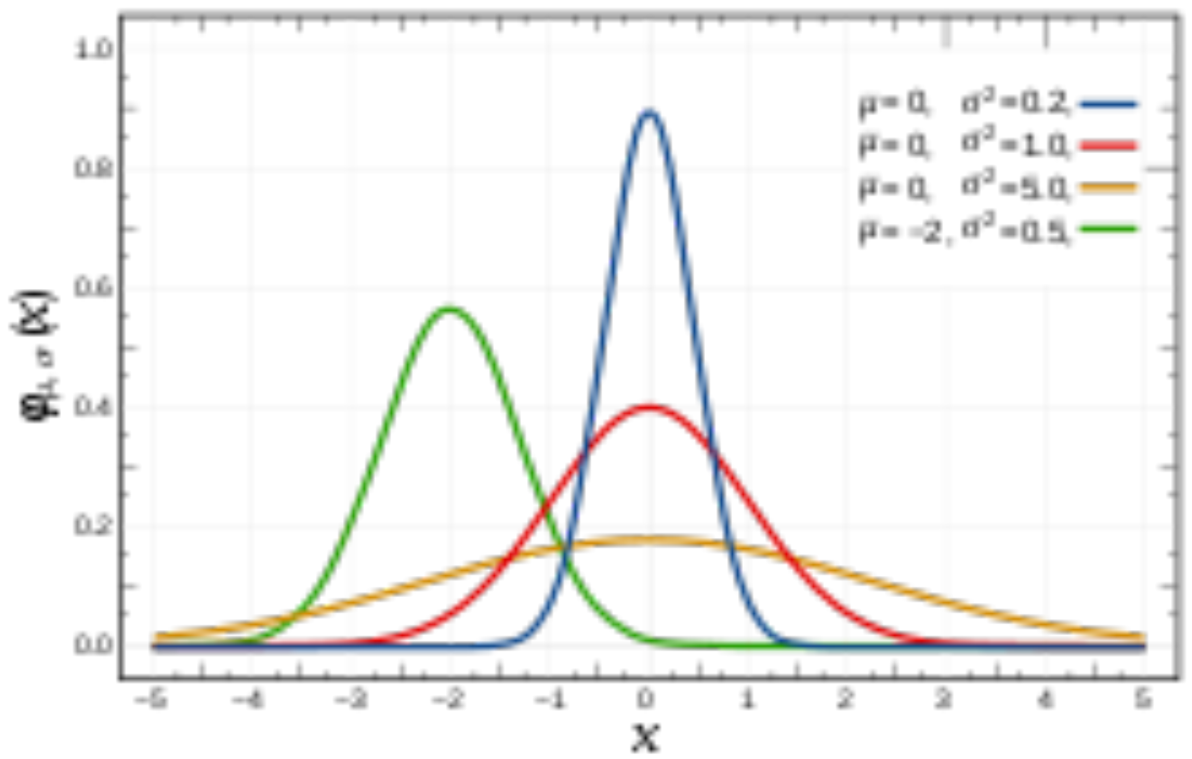
- E.g., (continuous) uniform distribution:

$$u(x; a, b) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$



- E.g., Gaussian distribution

$$\mathcal{N}(x; \mu, \sigma^2) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$



from: wikipedia.org

Probability terminology

- Joint probability $P(X, Y)$
- Conditional probability $P(X \mid Y)$
- Marginal probability $P(X)$ and $P(Y)$

- And they are all related: $p(x, y) = p(x|y)p(y) = p(y|x)p(x)$

A few useful properties
(shown for discrete variables for simplicity)

- Sum rule: $P(X) = \sum_{Y} P(X, Y)$
- Product rule: $P(x, y) = P(x|y)P(y)$
- Chain rule: $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})P(X_1)$
- If x and y are independent: $P(X, Y) = P(X)P(Y)$
- Bayes' Rule: $P(Y \mid X) = \dfrac{P(X \mid Y)P(Y)}{P(X)}$

Moments

- Expectation: $\mathbb{E}[X] = \sum_i P(x = x_i)x_i \qquad \mathbb{E}[aX] = a\mathbb{E}[X]$
- Variance: $\sigma^2 = E[(X - E[X])^2]$
- Covariance: $\text{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$
- Correlation: $\rho(x, y) = \dfrac{\text{Cov}(X, Y)}{\sigma_x \sigma_y}$

Further Reading

- Prologue to “The Master Algorithm”
  http://homes.cs.washington.edu/~pedrod/Prologue.pdf
- Ch. 1 of Hastie et al.
- Math Preparation
  - Ch.2 of Pattern Recognition and Machine Learning [PRML]
  - Ch.2-3 of Deep Learning [DL]
  - Slightly more advanced:

    http://www.cs.mcgill.ca/~dprecup/courses/ML/Materials/prob-review.pdf

    http://www.cs.mcgill.ca/~dprecup/courses/ML/Materials/linalg-review.pdf