

# **Machine Learning for Large-Scale Data Analysis and Decision Making (MATH80629A)**

## **Fall 2021**

**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/ML/main.html>

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

# Class Website

## 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](#)

| [MATH80629A](#) | [Lectures](#) | [Homework](#) | [Lab](#) | [Project](#) | [Office hour](#)

## Machine Learning for Large-Scale Data Analysis and Decision Making (MATH80629A): Fall 2021

### 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 online 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 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

- Mondays 8:30 am - 11:30 am, online, [zoom link](#). Rooms are password-protected. Reach out to me by email if you need the password.

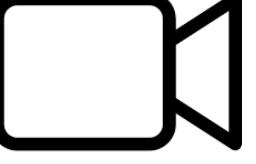
### Feedback

Please use this [form](#) to provide feedback about the course.

# A virtual semester...

- 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 (90-120 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 #5)

# Grading

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

- Homework (20%)
- Capsule quizzes (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.com/courses/301346>
- Entry name: **86ZZD7**
- One survey assignment, due **September 13, 2021**  
<https://forms.gle/9pBe78n34ipMccxB8>
- One homework assignment, due TBA
- One case study assignment, due TBA
- 5-6 multiple answers quizzes based on the capsules

# Project

- You must work in teams of two or three.
  - You can find teammates using **Piazza**: <https://piazza.com/class/kswiuomb6nf2k7#>
- **Timeline:**
- Team Registration, due: **September 20, 2021**  
<https://forms.gle/h4ENWMNdGgpeMncs7>
- Study plan, due: **October 18, 2021**
- Project meeting, on **October 25, 2021**
- Project Presentation, on: **December 6, 2021**
- Final report, due: **December 20, 2021**

# Getting Help

- Post your questions in [Piazza](#) and hopefully your peers will answer.
- My office hour:  
[Zoom link: https://hecmontreal.zoom.us/j/86904448666?  
pwd=T1NYdU43VUdOWVZCeE5iblhOajM1UT09](https://hecmontreal.zoom.us/j/86904448666?pwd=T1NYdU43VUdOWVZCeE5iblhOajM1UT09)

Mondays 11:30 am - 1:00 pm

my email:

[golnoosh.farnadi@hec.ca](mailto:golnoosh.farnadi@hec.ca)  
[farnadig@mila.quebec](mailto:farnadig@mila.quebec)

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



# 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.

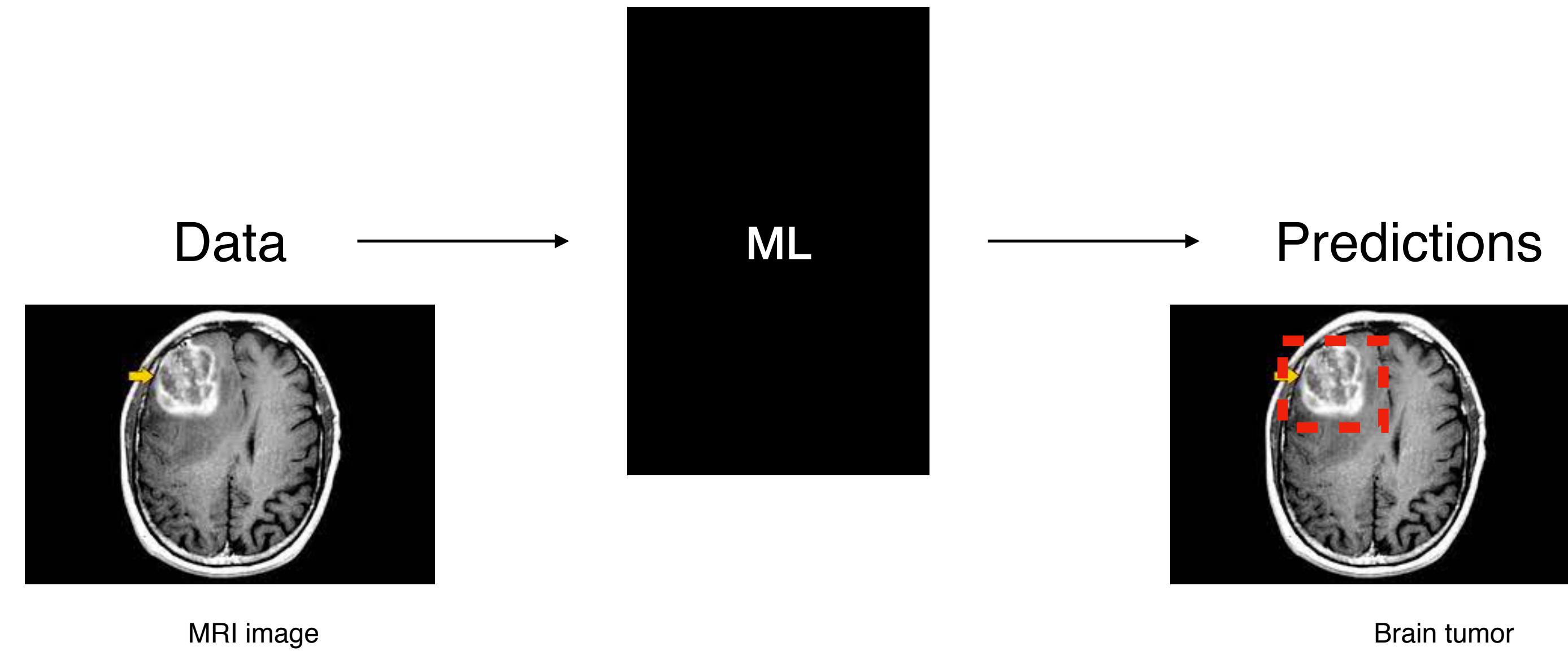
I am looking forward to hearing from you!



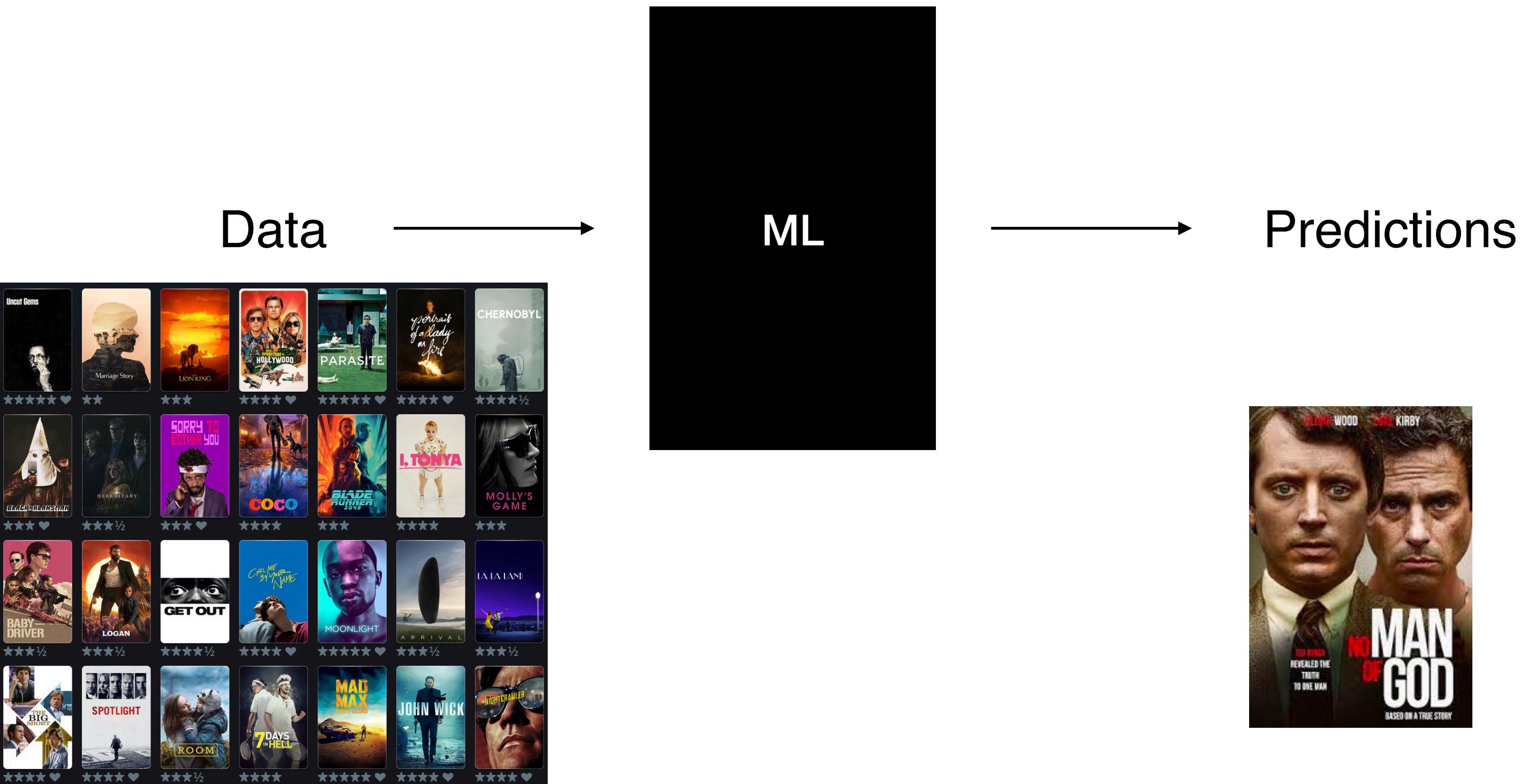
# Introduction to machine learning

# Machine Learning (ML)

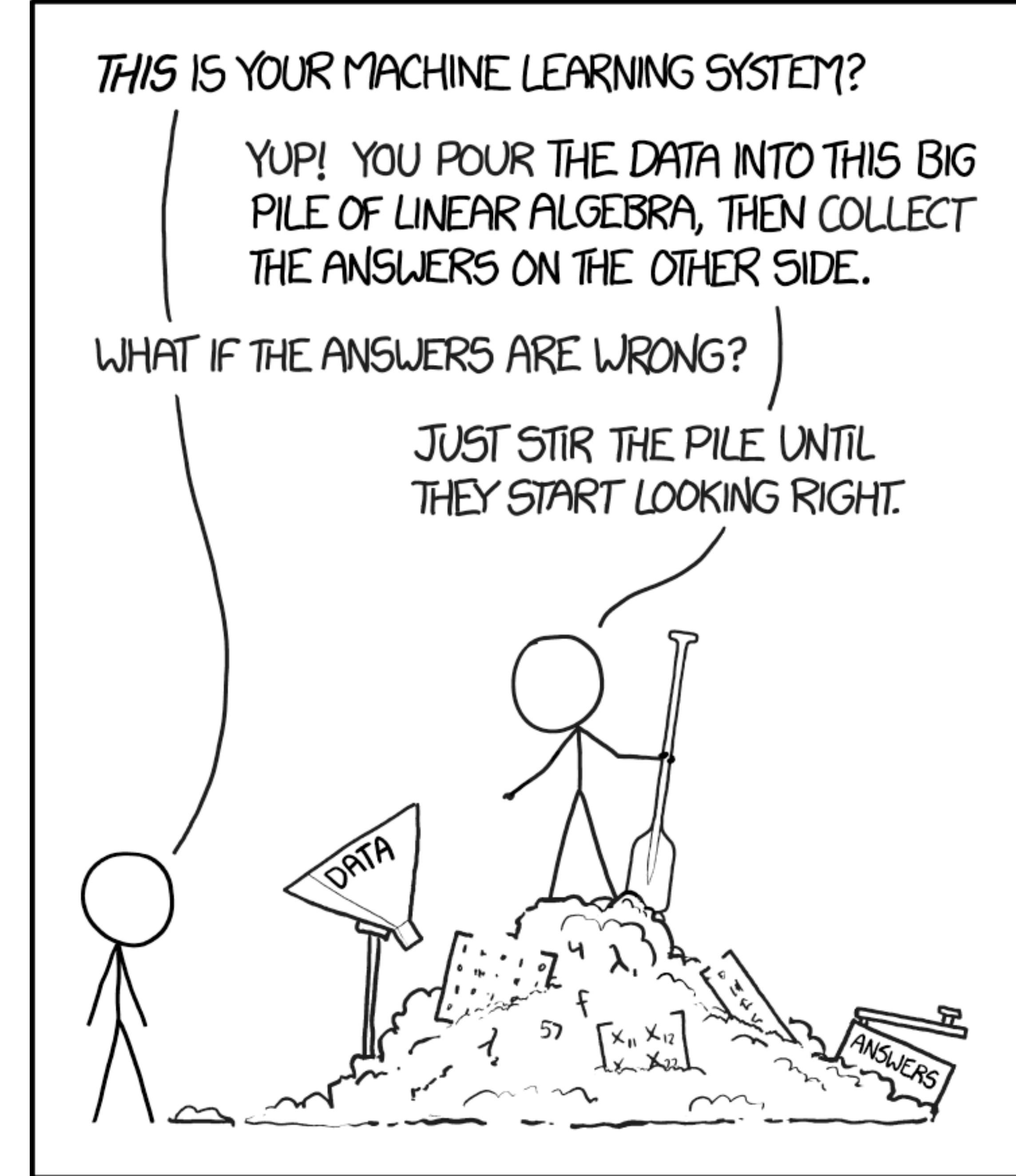
- Science that studies statistical and computational aspects of modeling data for predictive purposes
  - (Mostly) Empirical science



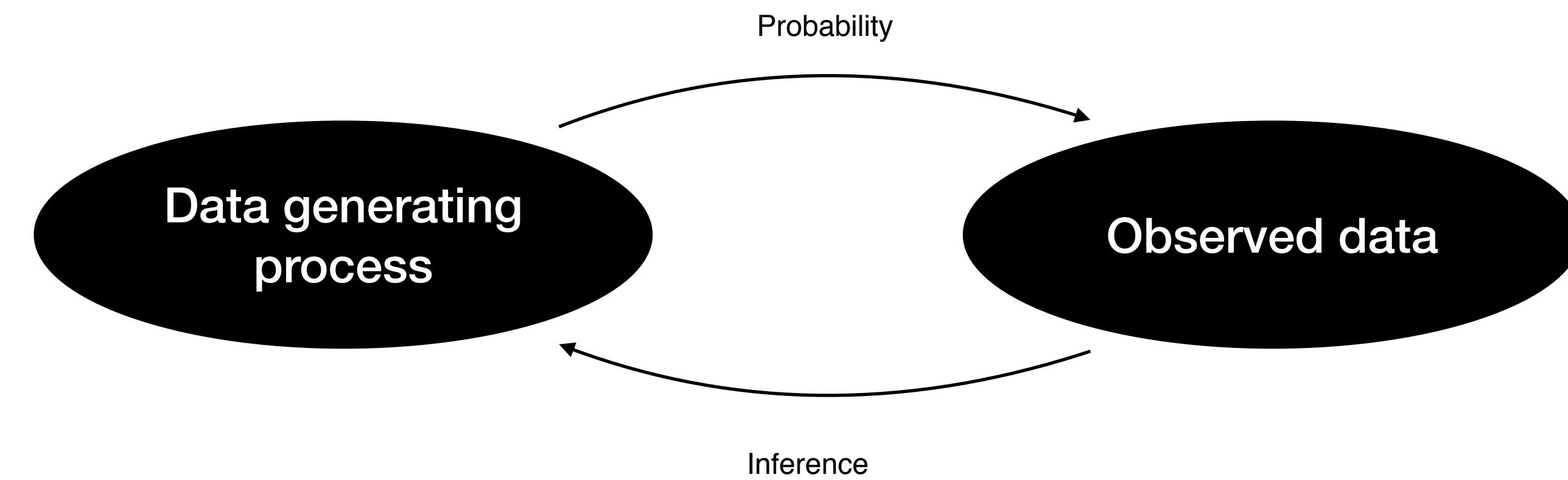
- Task: Predict whether an image contains a tumor



- Task: Predict the next movie a person should watch



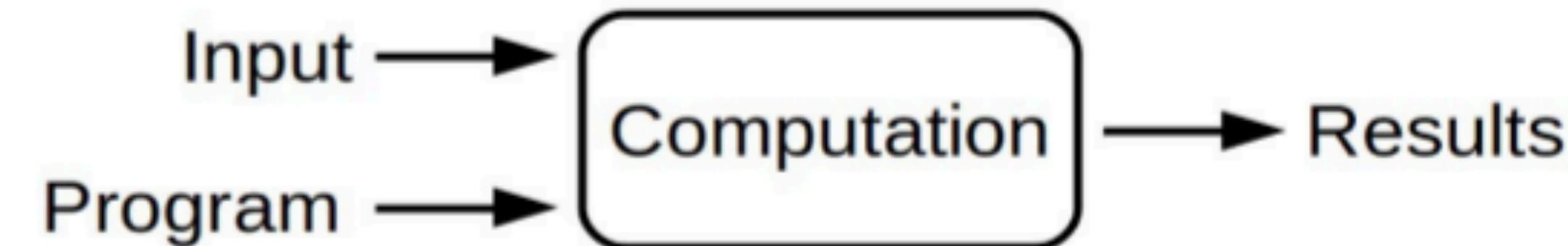
**“Data analysis, machine learning and data mining are various names given to the practice of statistical inference, depending on the context.”**



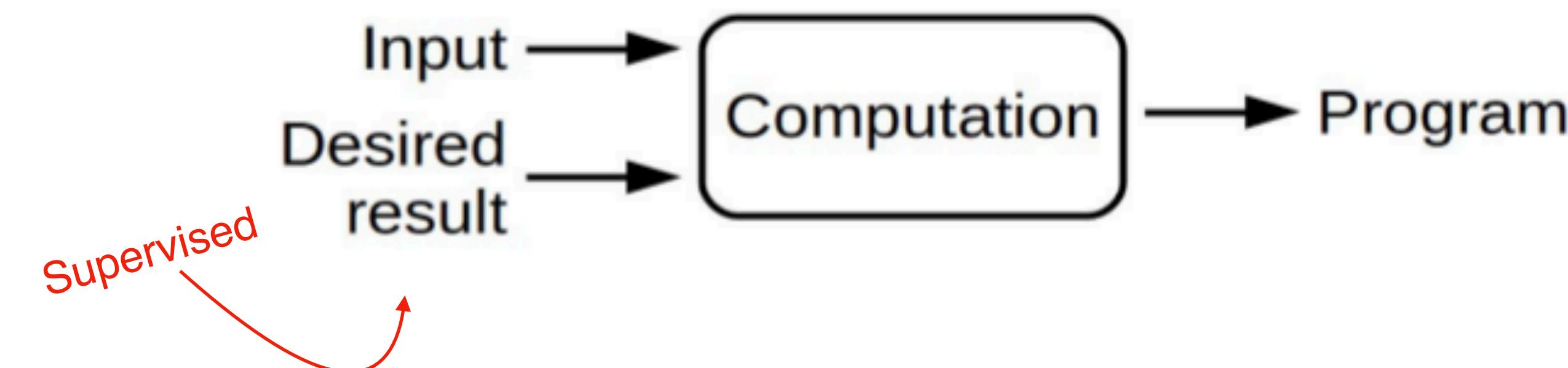
*–Larry Wasserman in “All of Statistics: A Concise Course in Statistical Inference.”*

# What is the goal of ML?

### Traditional programming



### Supervised Machine learning



# Practical Benefits of ML

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

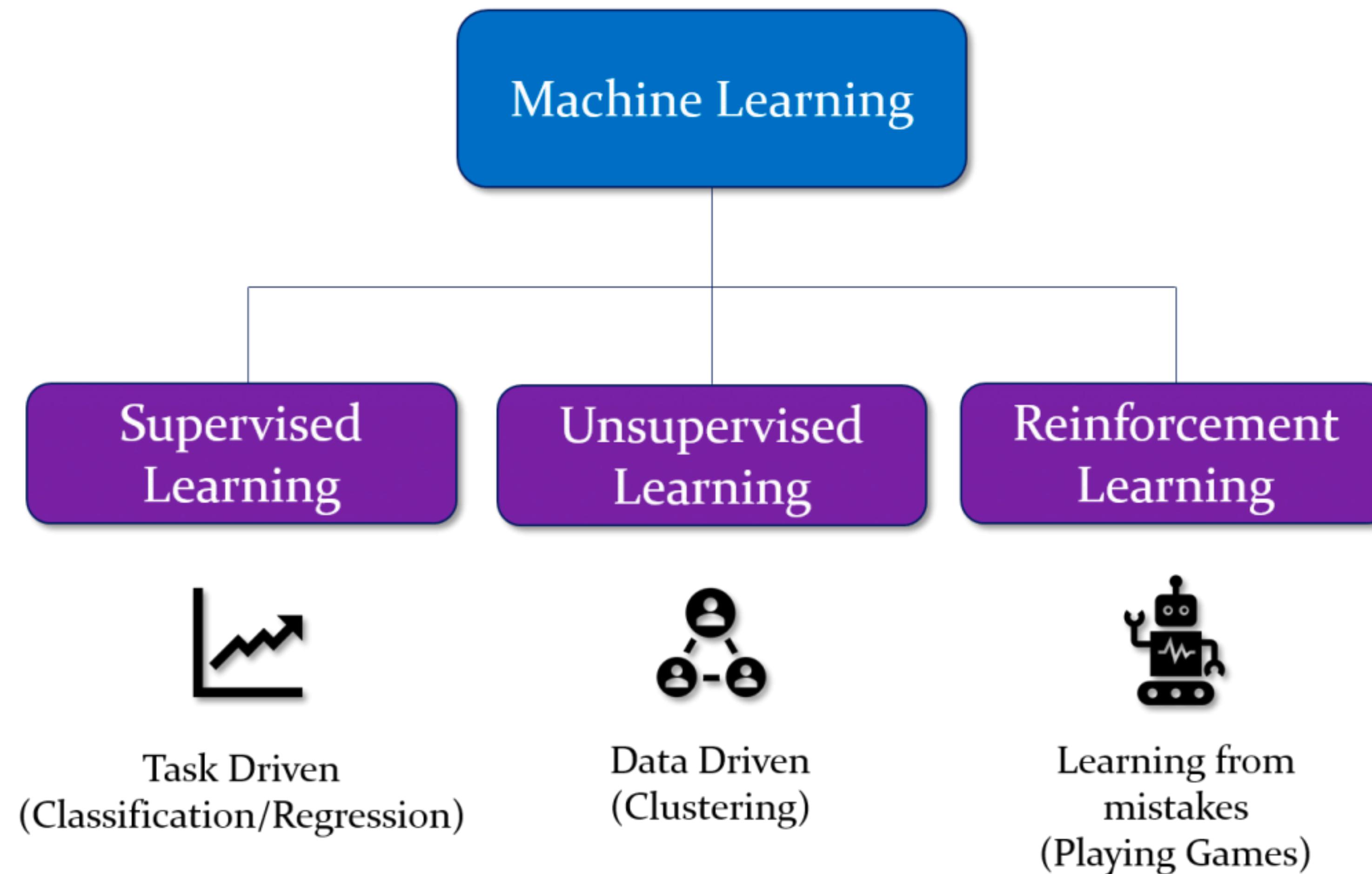
# Philosophical Benefits of ML

- 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.

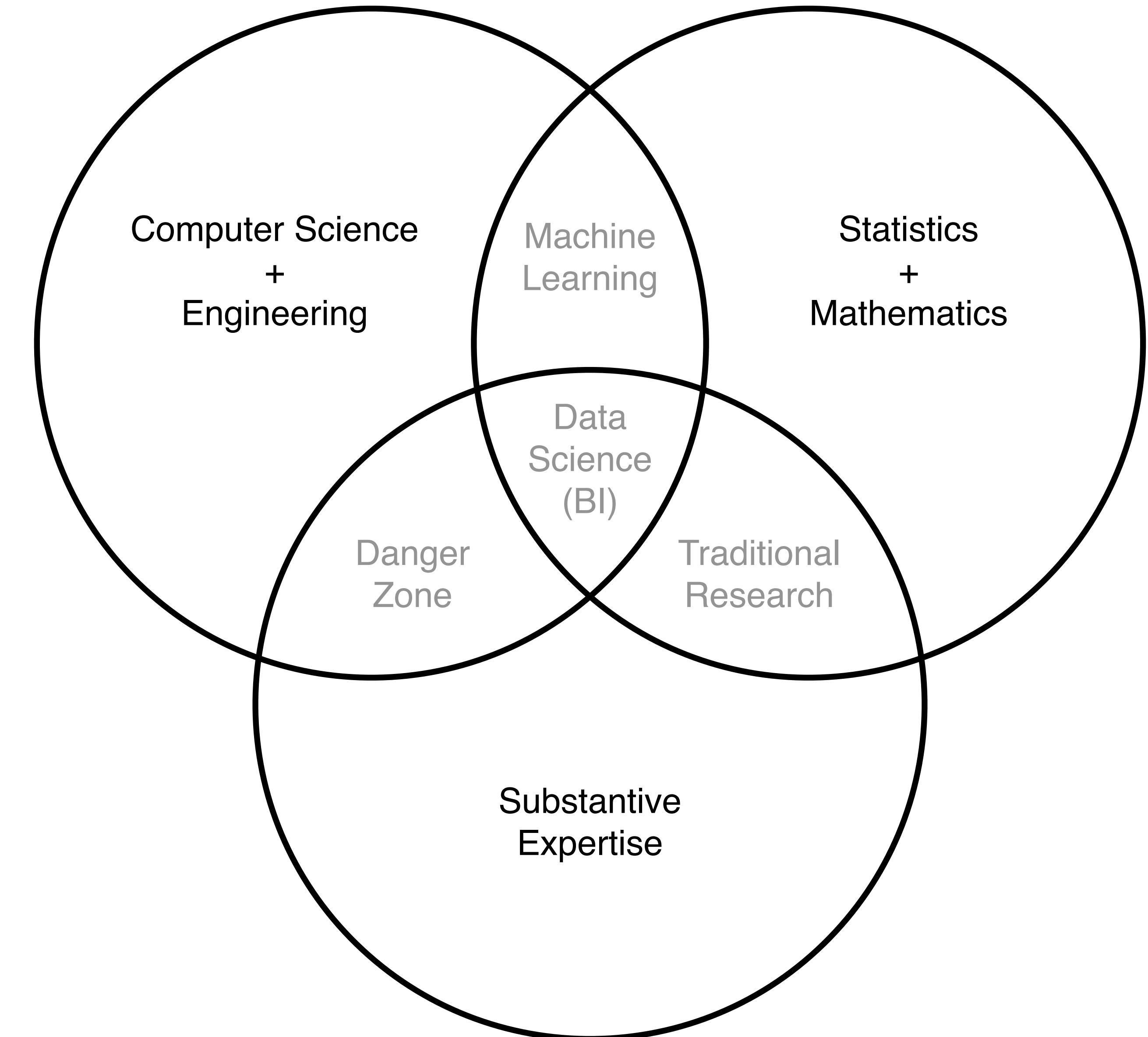
## Types of Machine Learning



# How does ML relate to other fields

# Historical View

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



# 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

No settled philosophy or widely accepted theoretical framework.

Willing to use *ad hoc* methods if they seem to work well (though appearances may be misleading).

Emphasis on automatic methods with little or no human intervention.

Methods suitable for many problems.

Heavy use of computing.

## Traditional statistics

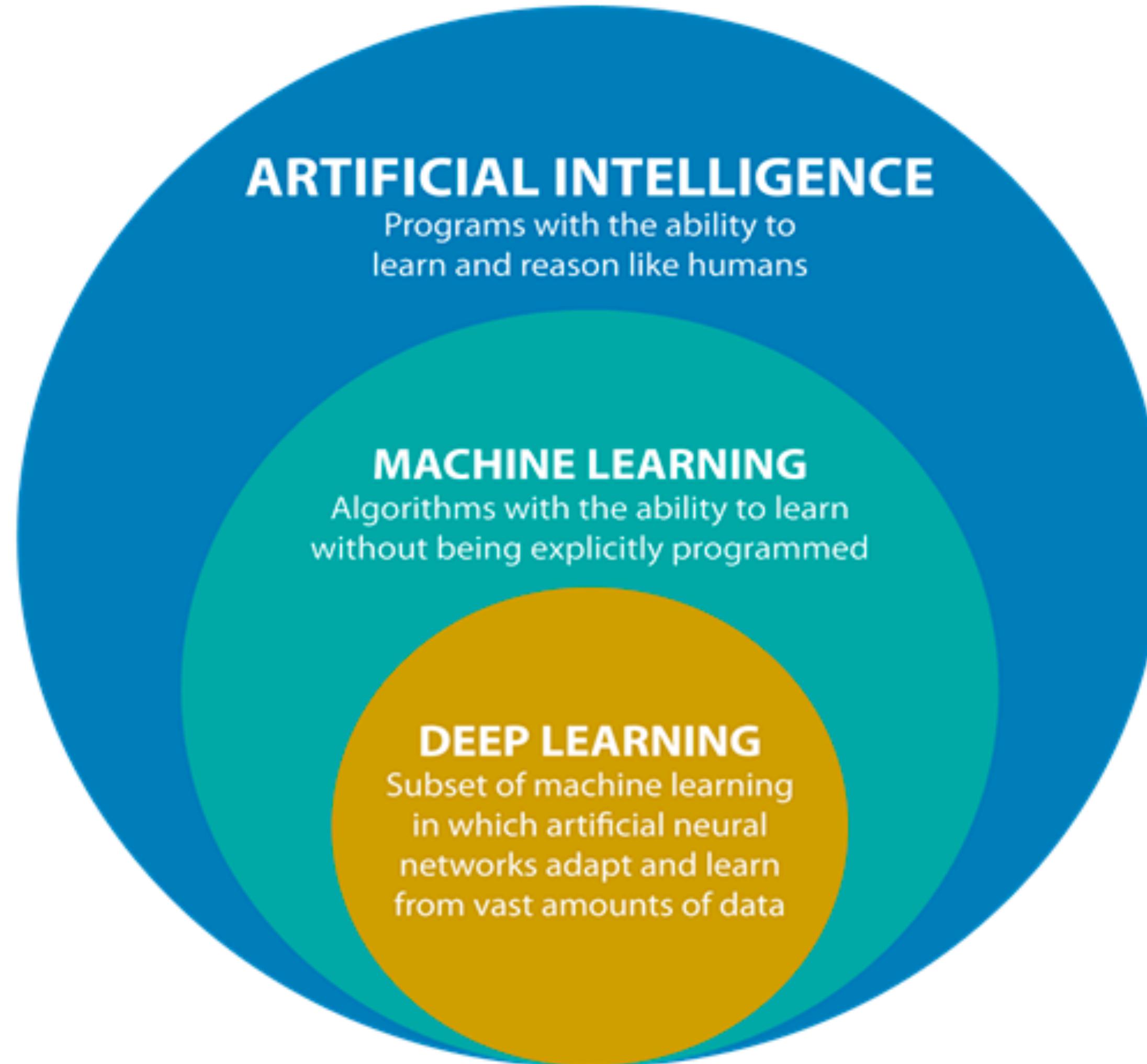
Classical (frequentist) and Bayesian philosophies compete.

Reluctant to use methods without some theoretical justification (even if the justification is actually meaningless).

Emphasis on use of human judgement assisted by plots and diagnostics.

Models based on scientific knowledge.

Originally designed for hand-calculation, but computing is now very important.



- 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

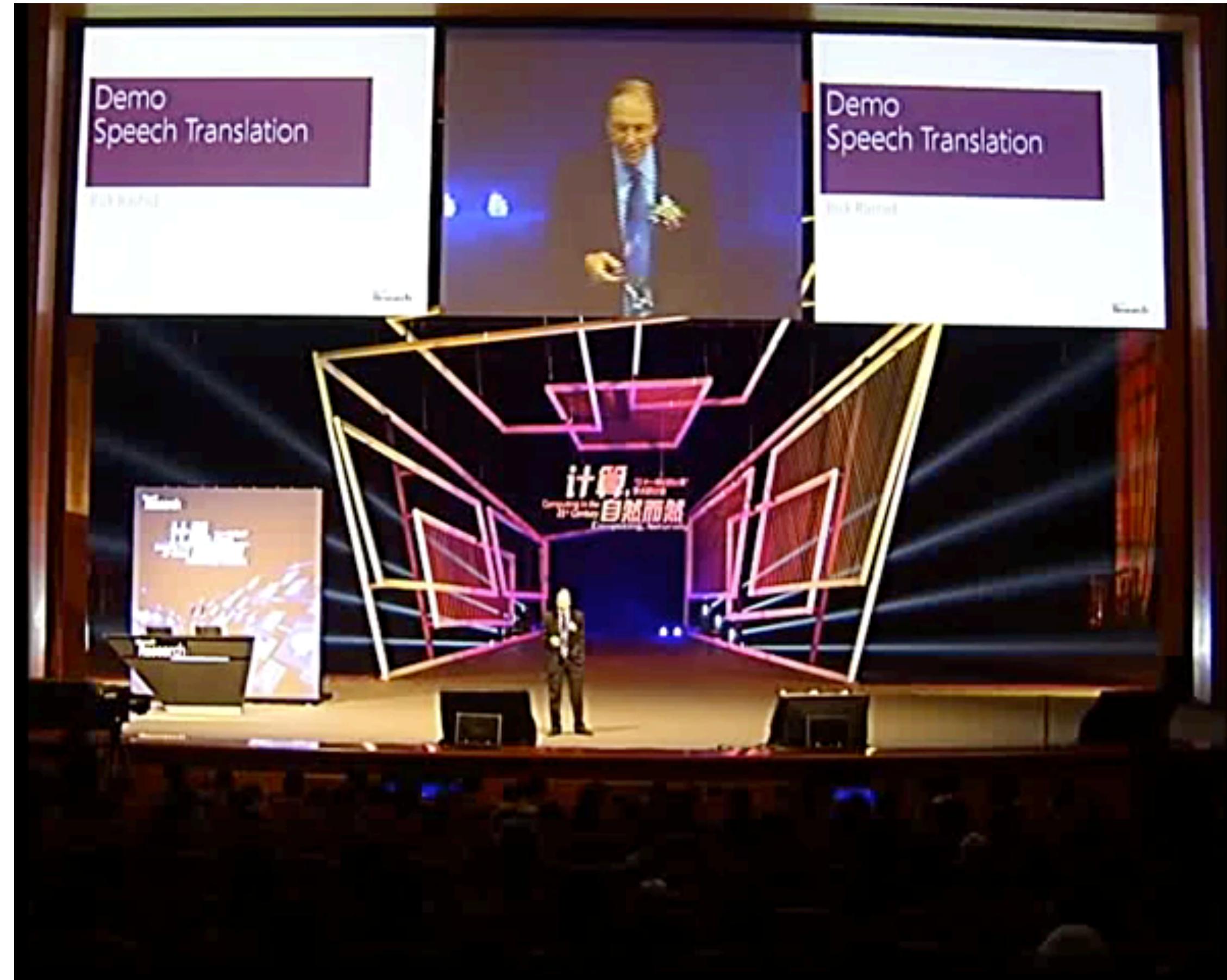
 Playtex Diaper Genie Disposal System Refill, 3-Pack, Blue ★★★★★ 79 CDN\$ 19.97 ✓Prime	 MAVEA 1001122 Maxtra Replacement Filter for MAVEA Water Filtration Pitcher, 3-Pack ★★★★★ 147 CDN\$ 19.99 ✓Prime	 Kleenex Ultra Facial Tissue Flat Bundle, 70 count (Pack of 6) ★★★★★ 19 CDN\$ 6.98 ✓Prime	 Kleenex Facial 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.



# 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.

Robotic observatory makes  
fast work of astronomy p. 476

A wet route to  
methanol p. 523

Human noise plagues  
protected areas p. 531

# Science

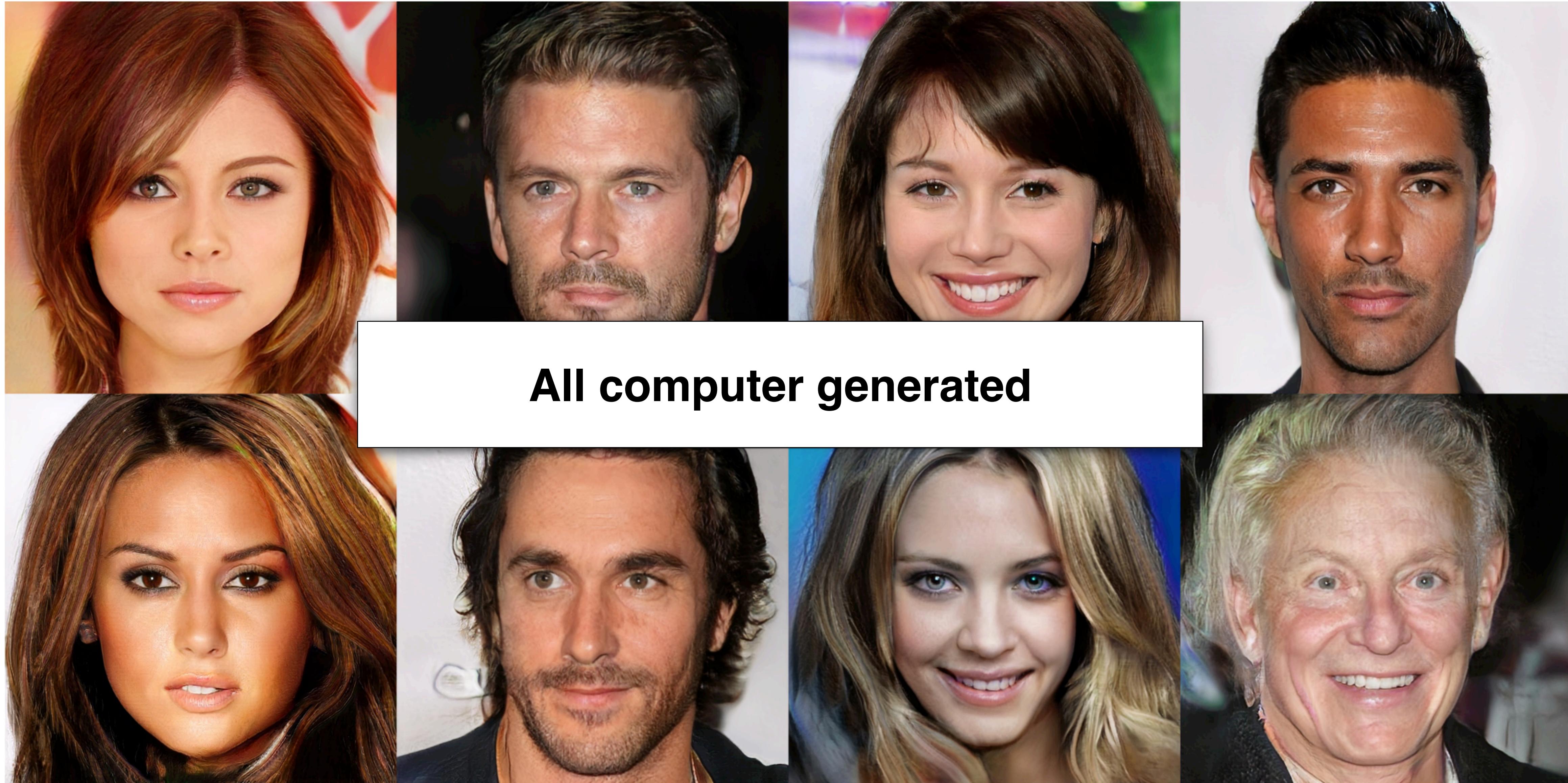
\$15  
5 MAY 2017  
[sciencemag.org](http://sciencemag.org)



## DIGITAL CARDS WHIZ

AI beats humans at  
challenging poker variant  
p. 508





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



2014



2015



2016



2017



2018



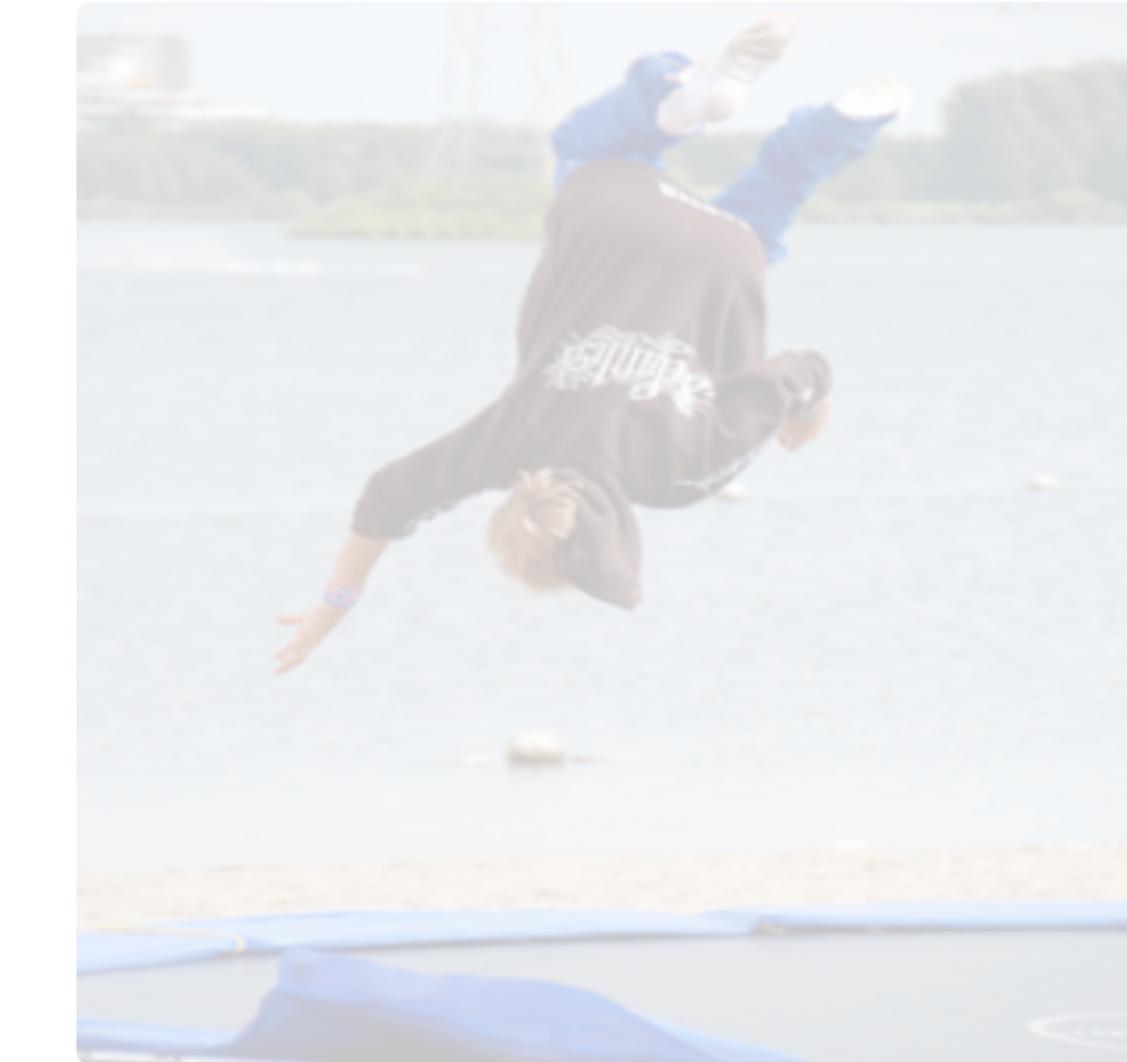
"man in black shirt is playing guitar."



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



"two young girls are playing with lego toy."



"boy is doing backflip on wakeboard."



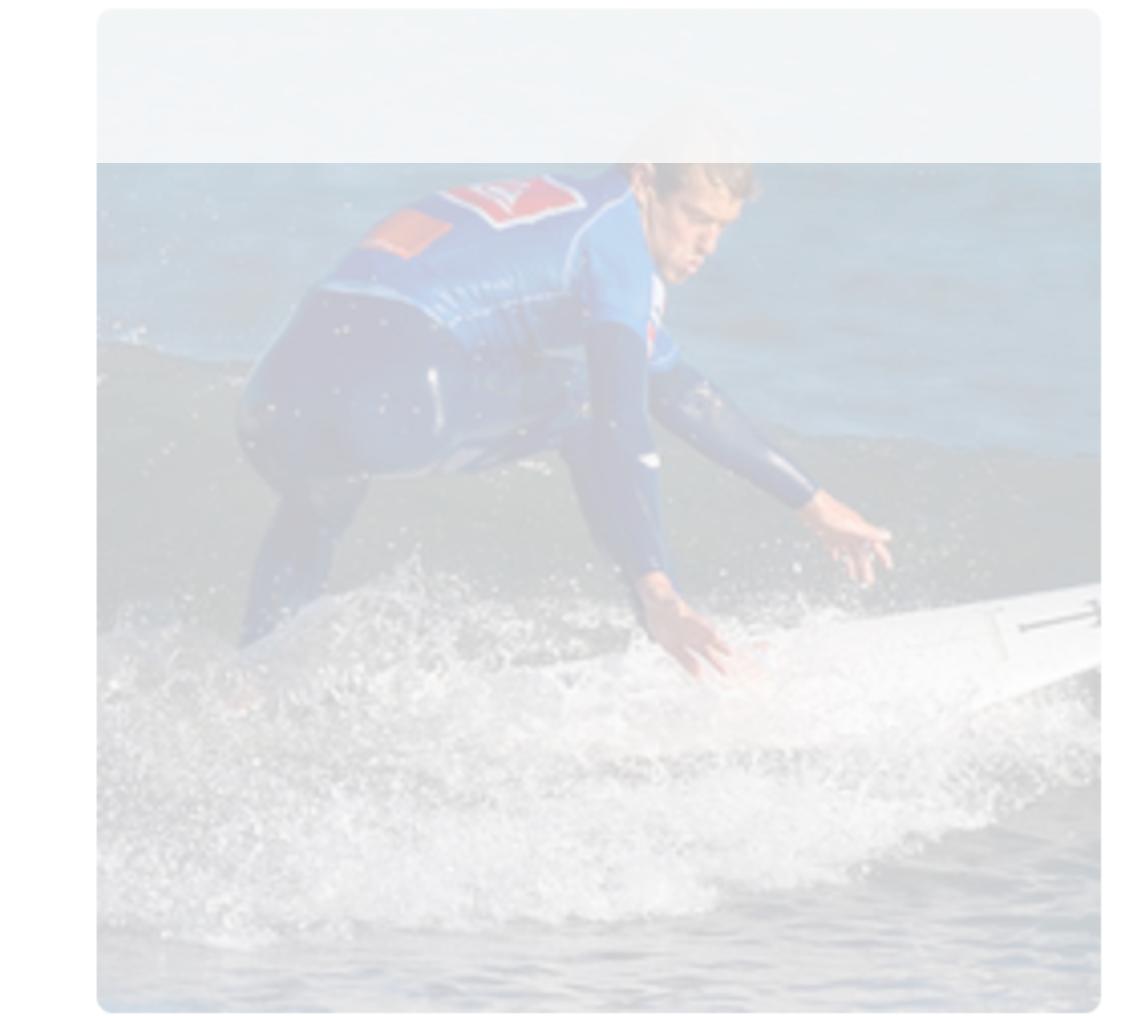
"girl in pink dress is jumping in air."



"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 3
- **Supervised Learning:** Week 4 & 5
- **Deep learning:** Week 6 & 8
- **Unsupervised Learning:** Week 10
- **Large scale computing :** Week 11
- **Trustworthy ML and Recommender Systems:** Week 12
- **Reinforcement Learning:** Week 13 & 14

# Fit with other courses

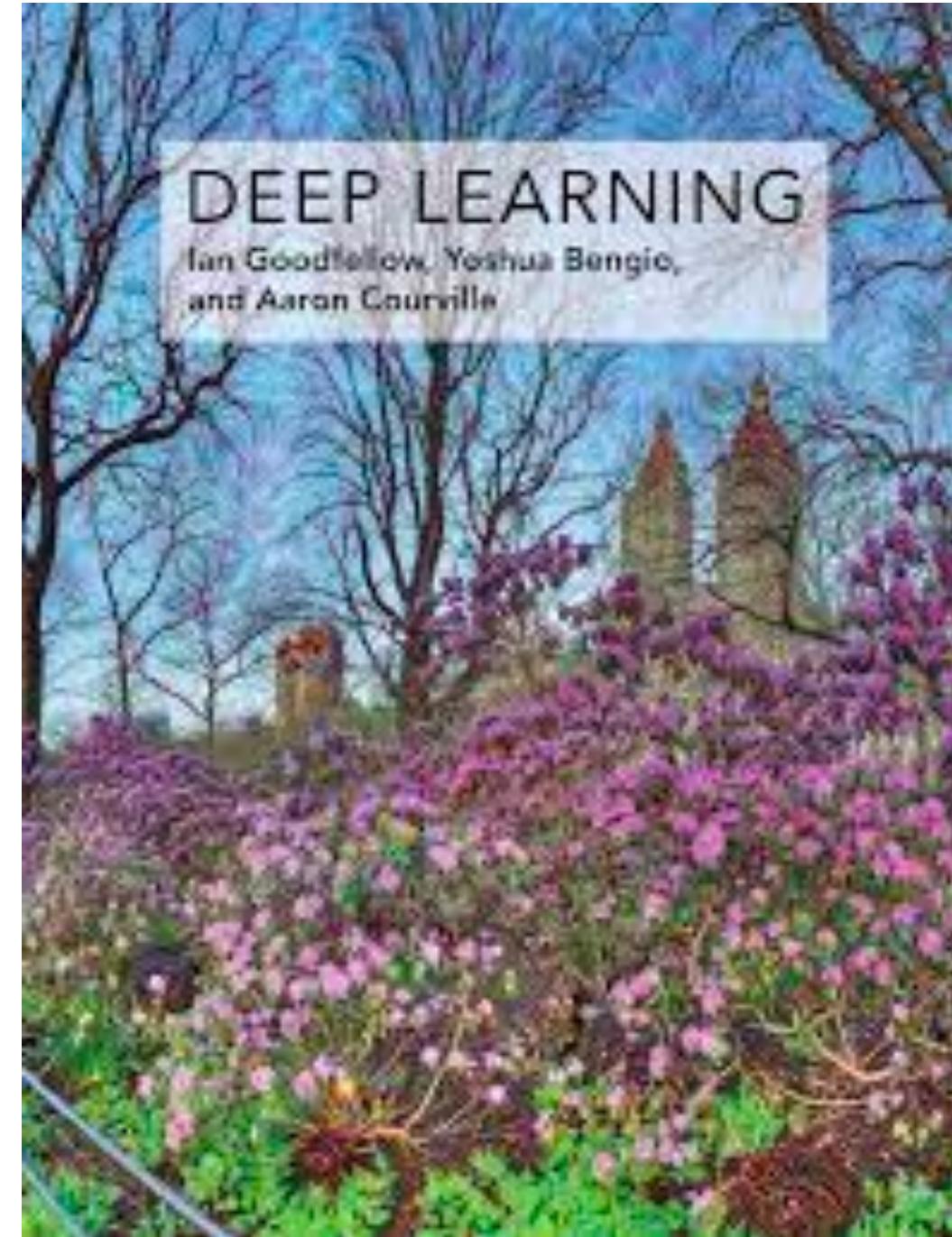
- HEC
  - PhD level (originally)
  - Computationally oriented
  - Prequel to Machine Learning II: Deep Learning
- Other ML courses in Montreal (U.Montreal, Polytechnique, McGill)
  - More applied (similar to COMP-551 @McGill)

# **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.

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

- Vector: an array of values.

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

- Matrix: a table of values.

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

# Indexing notation

- Indexing elements of a vector:

$$\mathbf{a} = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix} \leftarrow a_1$$

$a_i$

Convention:  
The first element  
is the zero'th.

- Indexing elements of a matrix:

$$\mathbf{A} = \begin{bmatrix} 3 & 4 & 2 \\ 1 & 2 & 9 \end{bmatrix}$$

$a_{ij}$

$a_{12}$

# Simple operations

- Transpose

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} \quad \left| \quad (\mathbf{A}_{ij})^\top = \mathbf{A}_{ji}$$
$$\mathbf{a}^\top = [\mathbf{a}_0 \quad \mathbf{a}_1 \quad \mathbf{a}_2]$$

- Addition

- Vectors and matrices w. the same shape

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \quad \left| \quad (\mathbf{A} + \mathbf{B})_{ij} = \mathbf{A}_{ij} + \mathbf{B}_{ij}\right.$$
$$\mathbf{a} + \mathbf{b} = \begin{bmatrix} \mathbf{a}_0 + \mathbf{b}_0 \\ \mathbf{a}_1 + \mathbf{b}_1 \\ \mathbf{a}_2 + \mathbf{b}_2 \end{bmatrix}$$

# Simple operations

- Multiply by a scalar

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

- Vector product.

- The dot product

$$\mathbf{a}^\top \mathbf{a} = \sum_i \mathbf{a}_i \mathbf{a}_i$$

- Note: it yields a scalar.

- Element-wise product:

- Also known as Hadamard product

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

# Operations

- Matrix product (dot product):

$$\mathbf{C}_{ij} = \sum_k \mathbf{A}_{ik} \mathbf{B}_{kj}$$

- 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  $A^{-1} \rightarrow A^{-1}A = I_n$
- 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.

A square matrix  
not invertible is *singular*

- No column can be recovered using a combination of other columns

- Inverses are useful to solve systems of equations:

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

# Norms

- $L^p$  norm. 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):

$$\| \mathbf{A} \|_2 = \sqrt{\left( \sum_i \sum_j |a_{ij}|^2 \right)}$$

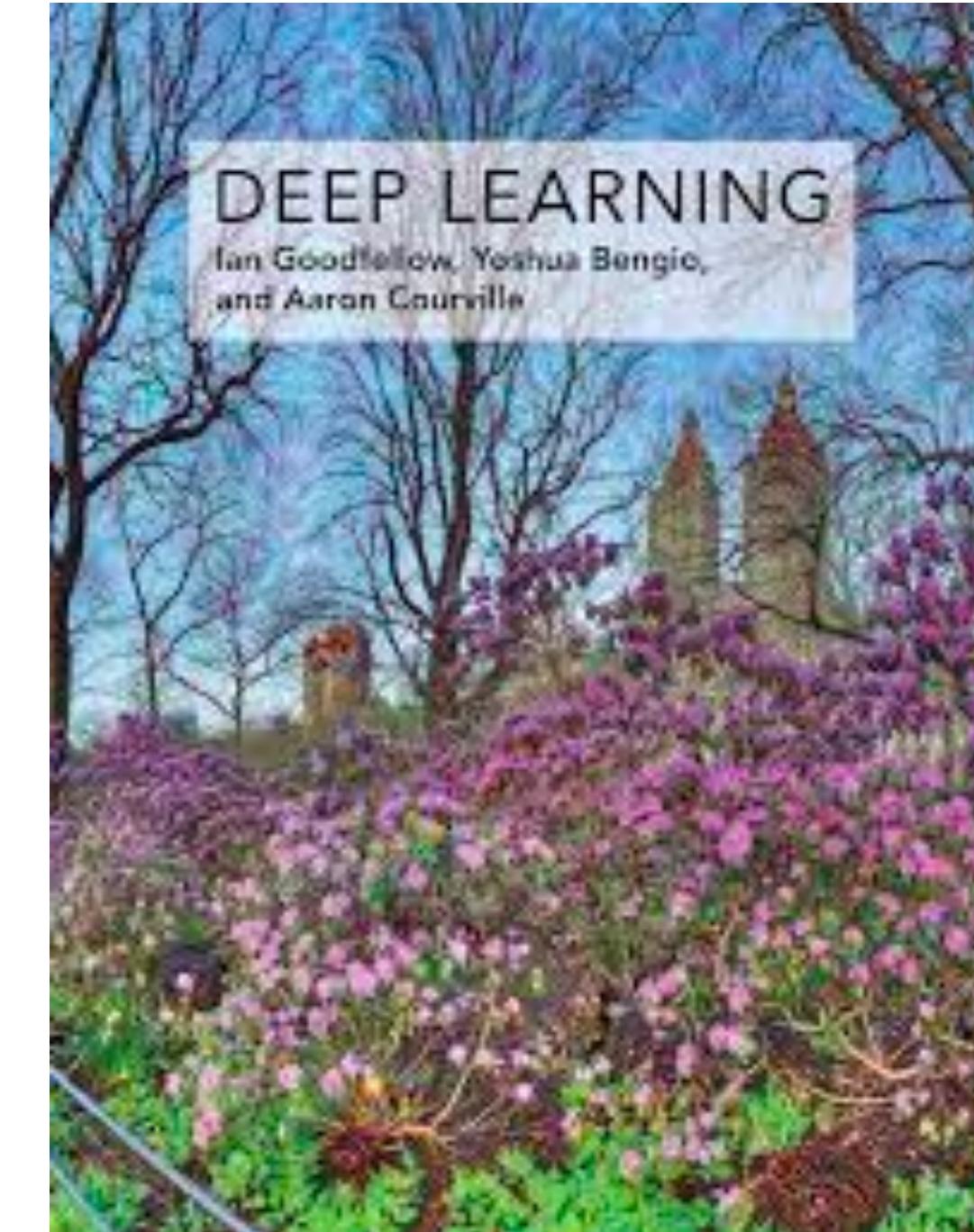
# Special matrices & vectors

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

- 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.



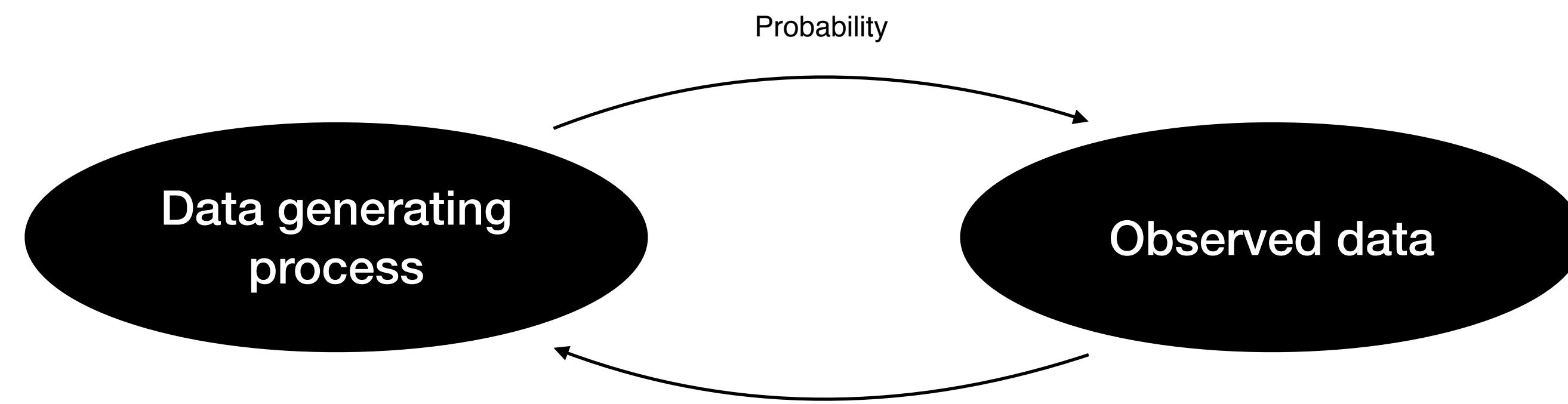
# Why probabilities?

- To capture uncertainty

E.g., What time will I get home tonight?

- 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

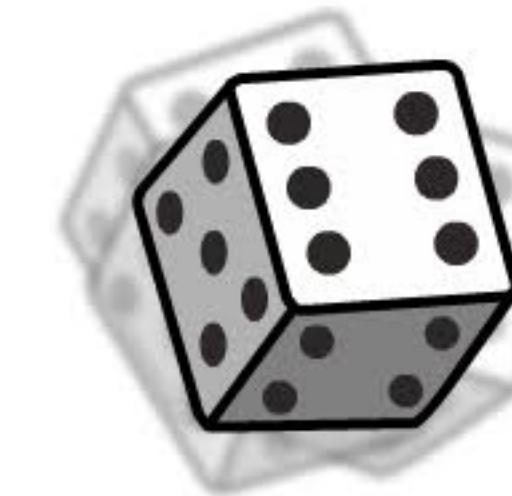
- Generate data by throwing a fair die.
- What do we know about a single throw?
  - 6 possible outcomes. ([sample space](#))
  - Each outcome (e.g., 1). ([element, state](#))
  - A subset of outcomes (e.g., <3). ([event](#))
  - Outcomes are equiprobable. ([uniform distribution](#))



# 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<sup>1</sup>

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

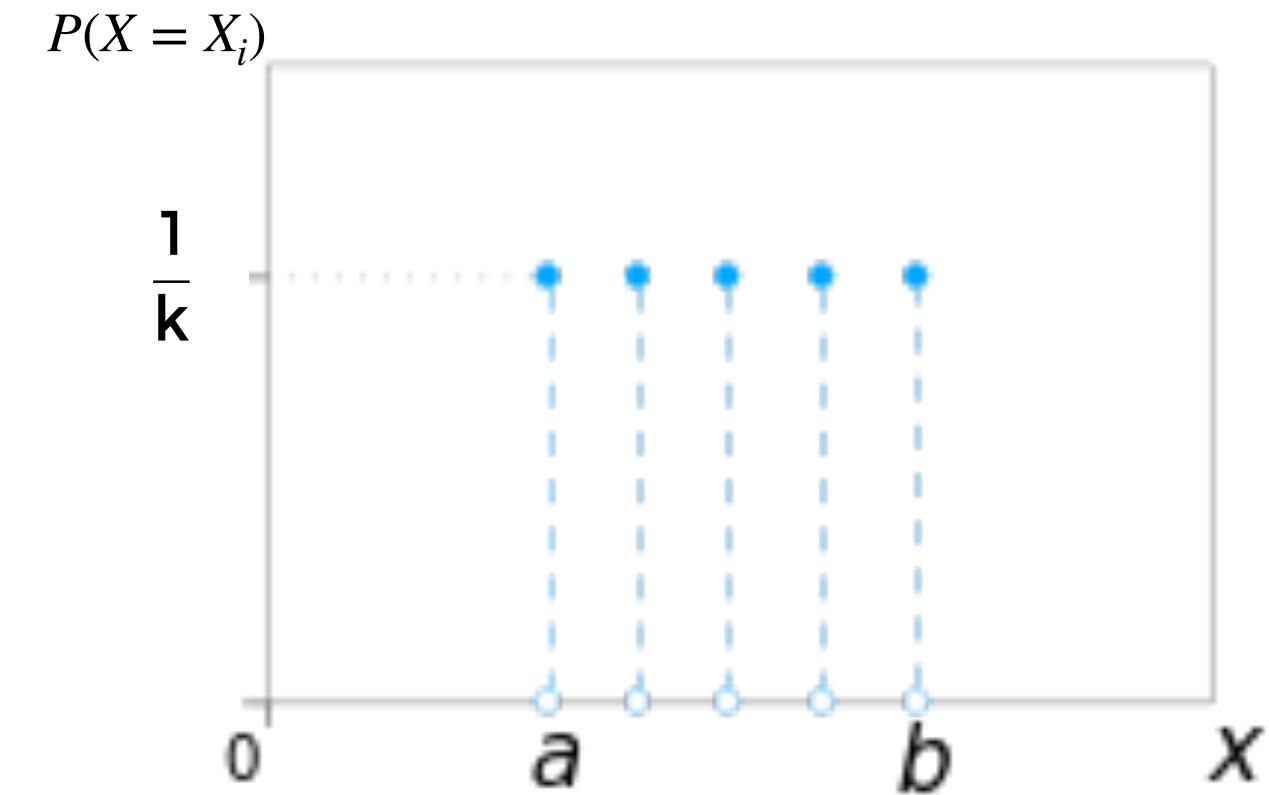
$$\sum_i P(x = x_i) = 1$$

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

# Discrete RVs

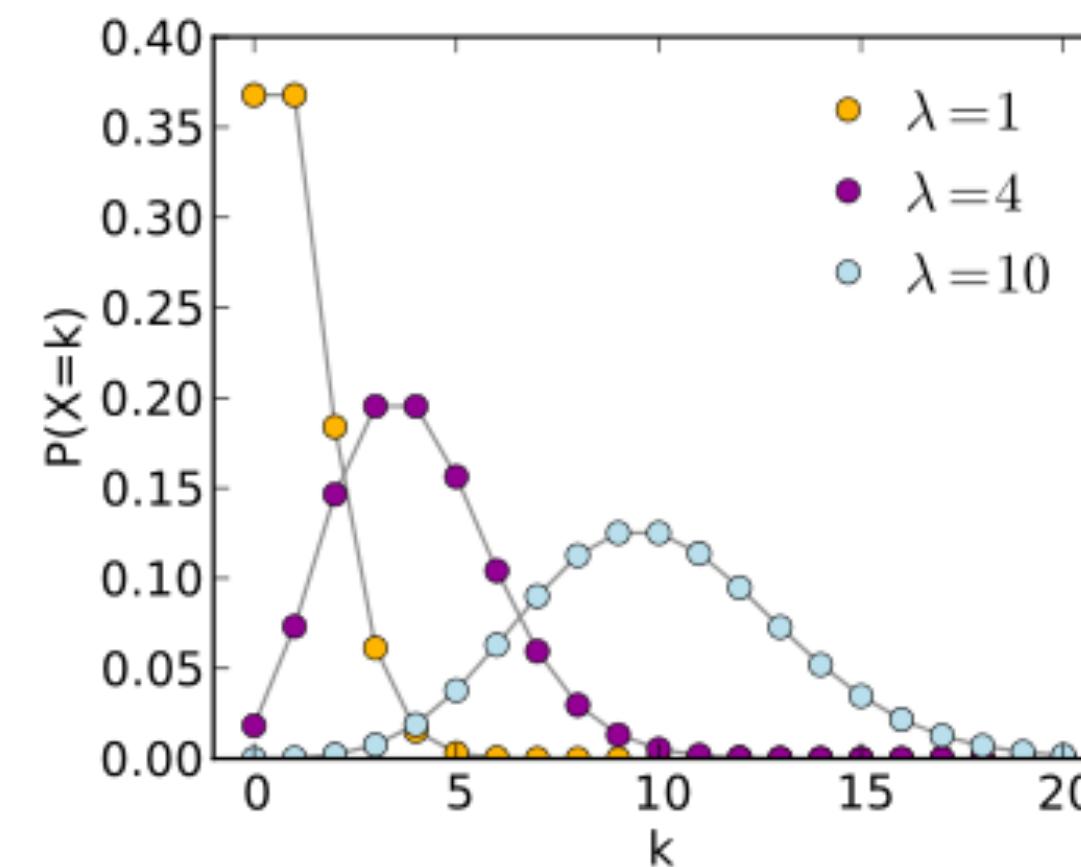
- E.g., uniform distribution:

$$P(X = X_i) = \frac{1}{k}, \forall i$$



- E.g., Poisson distribution:

$$P(X = X_i; \lambda) = \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}$$



both images are  
from: wikipedia.org

# Continuous RVs

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

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

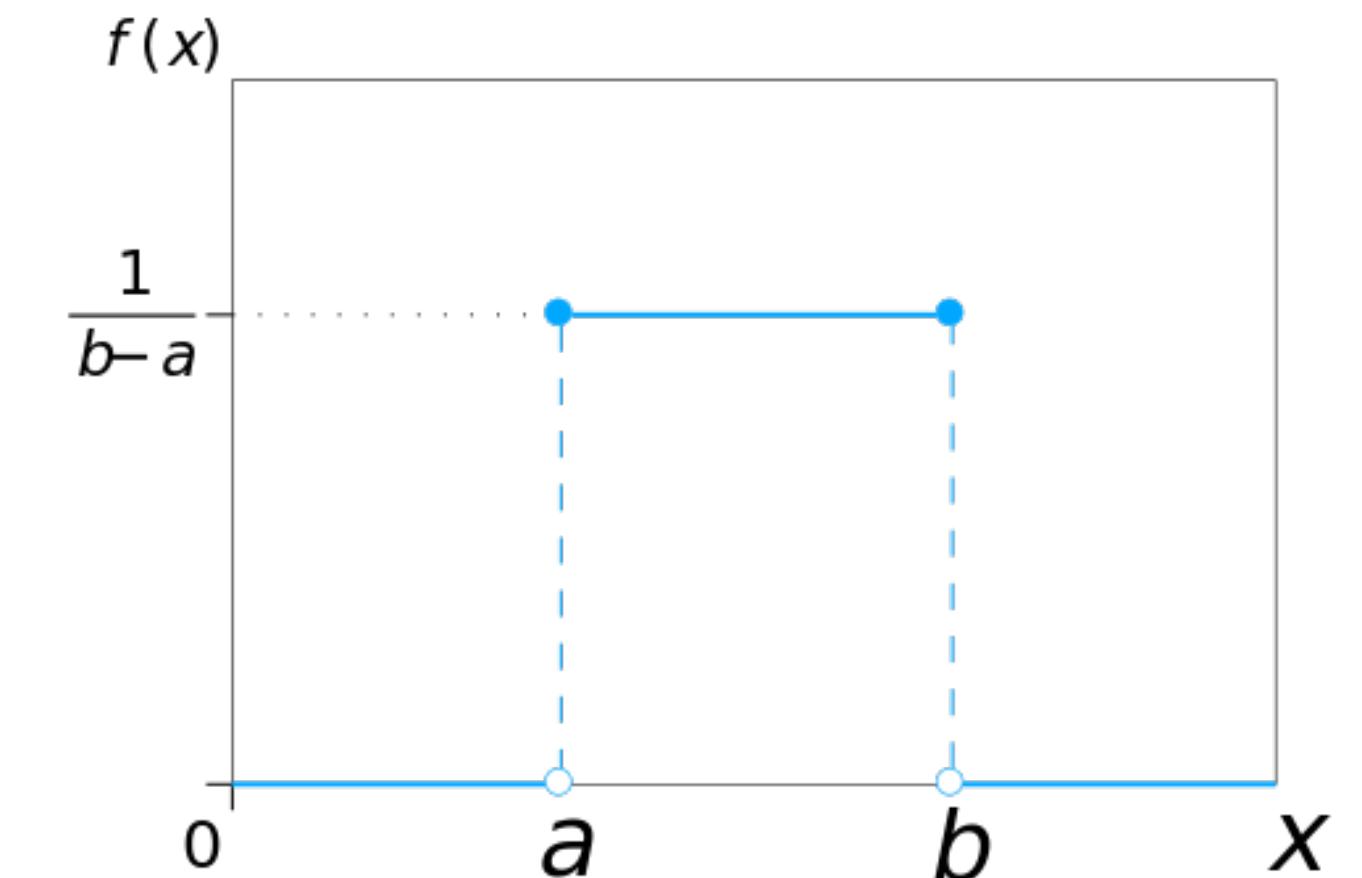
$$P(a < x < b) = \int_a^b f(x) dx$$

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

# Continuous RVs

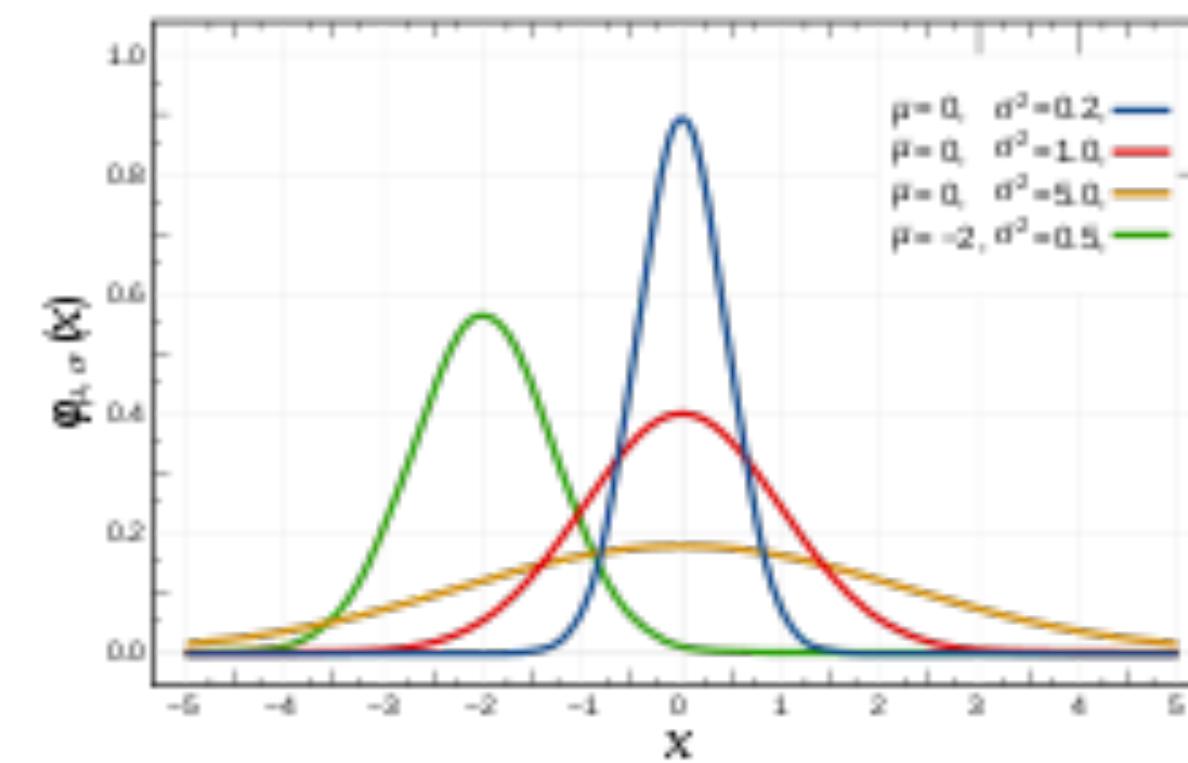
- E.g., (continuous) uniform distribution:

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



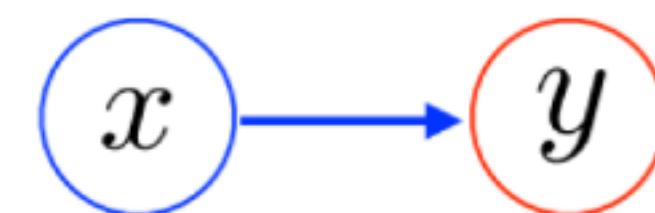
- E.g., Gaussian distribution

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



# Probability terminology

- Joint probability  $P(X, Y)$
- Conditional probability  $P(X | Y)$
- 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:  $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | X_1, \dots, X_{i-1})P(X_1)$
- If  $x$  and  $y$  are independent:  $P(X, Y) = P(X)P(Y)$
- Bayes' Rule:  $P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)}$

# Moments

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

# 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>