

# Intro to Artificial Intelligence and Machine Learning

Alvaro Soto

Computer Science Department (DCC), PUC

- Deep Learning is part of a broader field known as **Machine Learning**.
- Machine Learning is part of a broader field known as **Artificial Intelligence**.
- What is Artificial Intelligence?
  - Easy part: **Artificial**.
  - Hard part: **Intelligence**.

## Biological World



## Artificial World



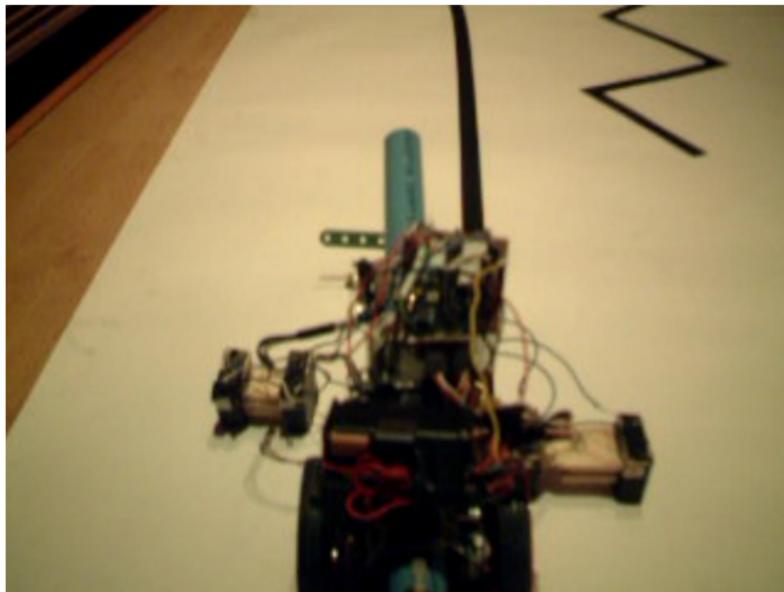
## Artificial Intelligence

Study of computational models that allow machines to **perceive, reason, and act with great flexibility.**

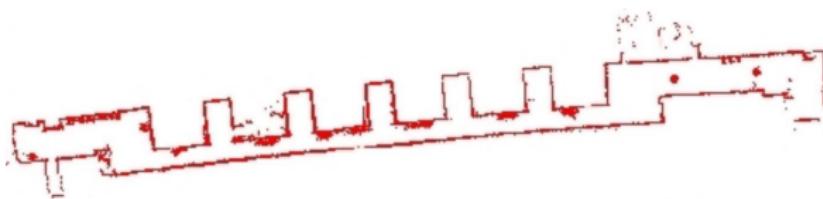
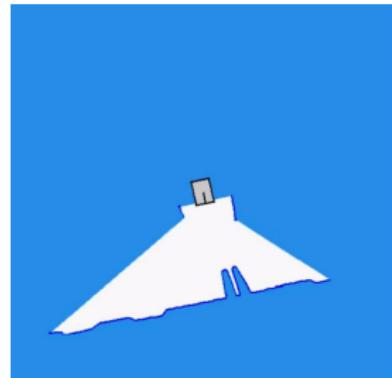
## Artificial Intelligence

Study of computational models that allow machines to acquire a **level of understanding** of its world.

## Understanding: Year 2001



## Understanding: Year 2005



## Understanding: Year 2022



# Understanding: Year 2022



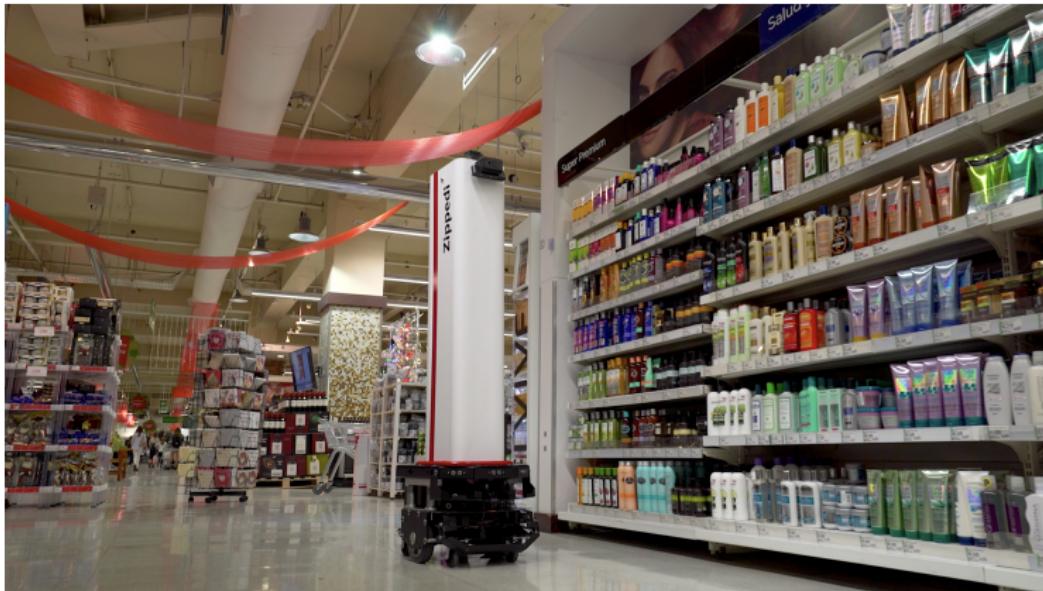
# Understanding: Year 2022

## Text summarization

cia documents reveal iot-specific televisions can be used to secretly record conversations .  
cybercriminals who initiated the attack managed to commandeer a large number of internet-connected devices in current use .  
cia documents revealed that microwave ovens can spy on you - maybe if you personally don't suffer the consequences of the sub-par security of the iot .

Internet of Things ( IoT ) security breaches have been dominating the headlines lately . WikiLeaks's trove of CIA documents revealed that internet-connected televisions can be used to secretly record conversations . Trump's advisor Kellyanne Conway believes that microwave ovens can spy on you - maybe she was referring to microwave cameras which indeed can be used for surveillance . And don't delude yourself that you are immune to IoT attacks , with 96 % of security professionals responding to a new survey expecting an increase in IoT breaches this year . Even if you personally don't suffer the consequences of the sub-par security of the IoT , your connected gadgets may well be unwittingly cooperating with criminals . Last October , Internet service provider Dyn came under an attack that disrupted access to popular websites . The cybercriminals who initiated the attack managed to commandeer a large number of internet-connected devices ( mostly DVRs and cameras ) to serve as their helpers . As a result , cybersecurity expert Bruce Schneier has called for government regulation of the IoT , concluding that both IoT manufacturers and their customers don't care about the security of the 8.4 billion internet-connected devices in current use . Whether because of government regulation or good old-fashioned self-interest , we can expect increased investment in IoT security technologies . In its recently-released TechRadar report for security and risk professionals , Forrester Research discusses the outlook for the 13 most relevant and important IoT security technologies , warning that "there is no single , magic security bullet that can easily fix all IoT security issues ." Based on Forrester's analysis , here's my list of the 6 hottest technologies for IoT security : IoT network security : Protecting and securing the network connecting IoT devices to back-end systems on the Internet . IoT network security is a bit more challenging than traditional network security because there is a wider range of communication protocols , standards , and device capabilities , all of which pose significant issues and increased complexity . Key capabilities include traditional endpoint security features such as antivirus and antimalware as well as other features such as firewalls and intrusion prevention and detection systems . Sample vendors : Bayshore Networks , Cisco , Darktrace , and Senrio . IoT authentication : Providing the ability for users to authenticate an IoT device , including managing multiple users of a single device ( such as a connected car ) , ranging from simple static passwords/pins to more robust authentication mechanisms such as two-factor

# Understanding: Year 2022



- 60's : Era of optimism or naiveness.
  - Deductive learning.
- 70's - 80's: Era of pessimism.
- 90's: Era of reborn.
- 2000's: **Era of Machine Learning.**
  - Inductive learning.

# Inductive learning



This bird can fly



This bird can fly



This bird can fly



This bird can fly



Can this bird fly ?

# Inductive learning



This bird can fly



This bird can fly



This bird can fly



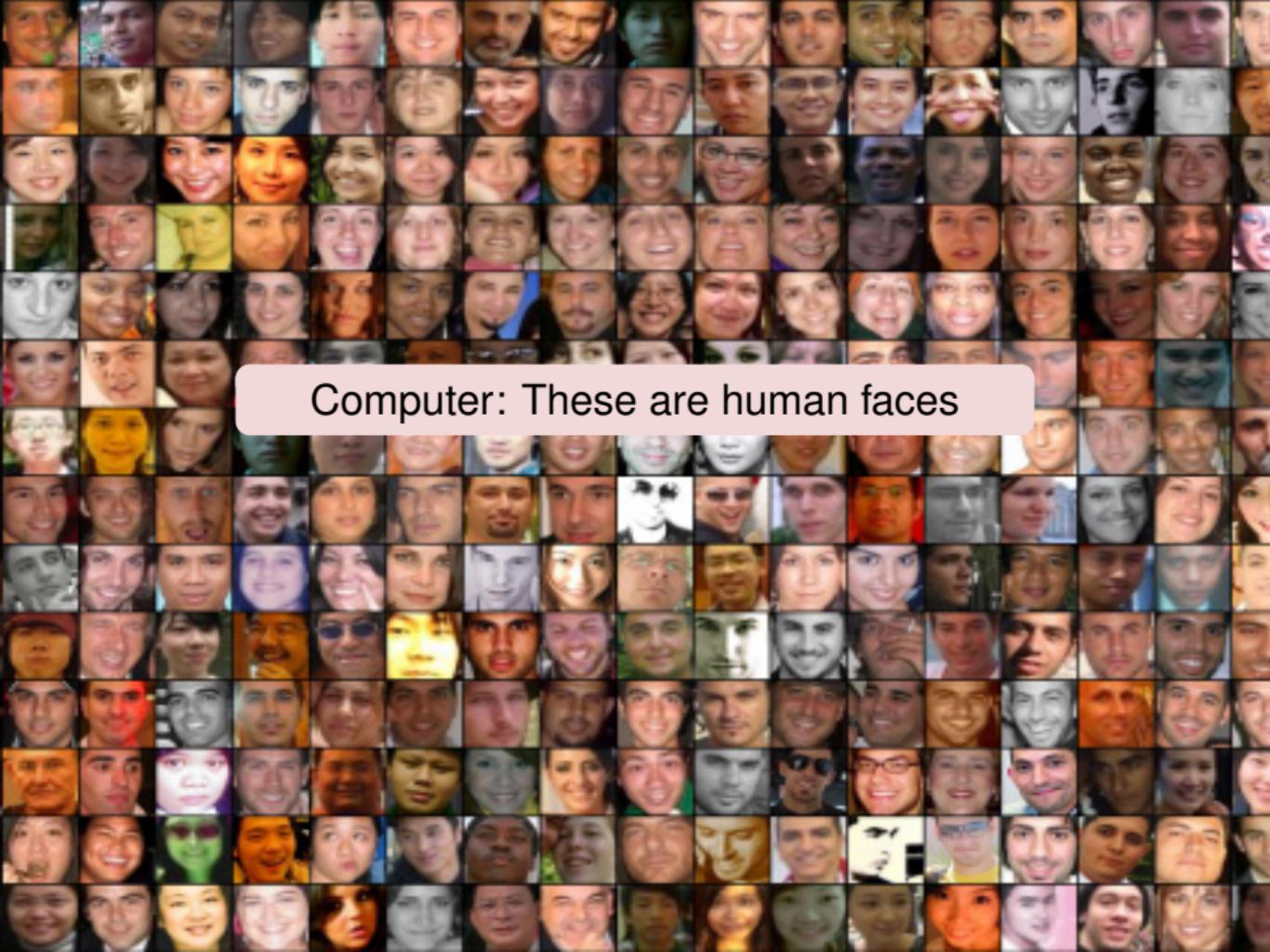
This bird can fly



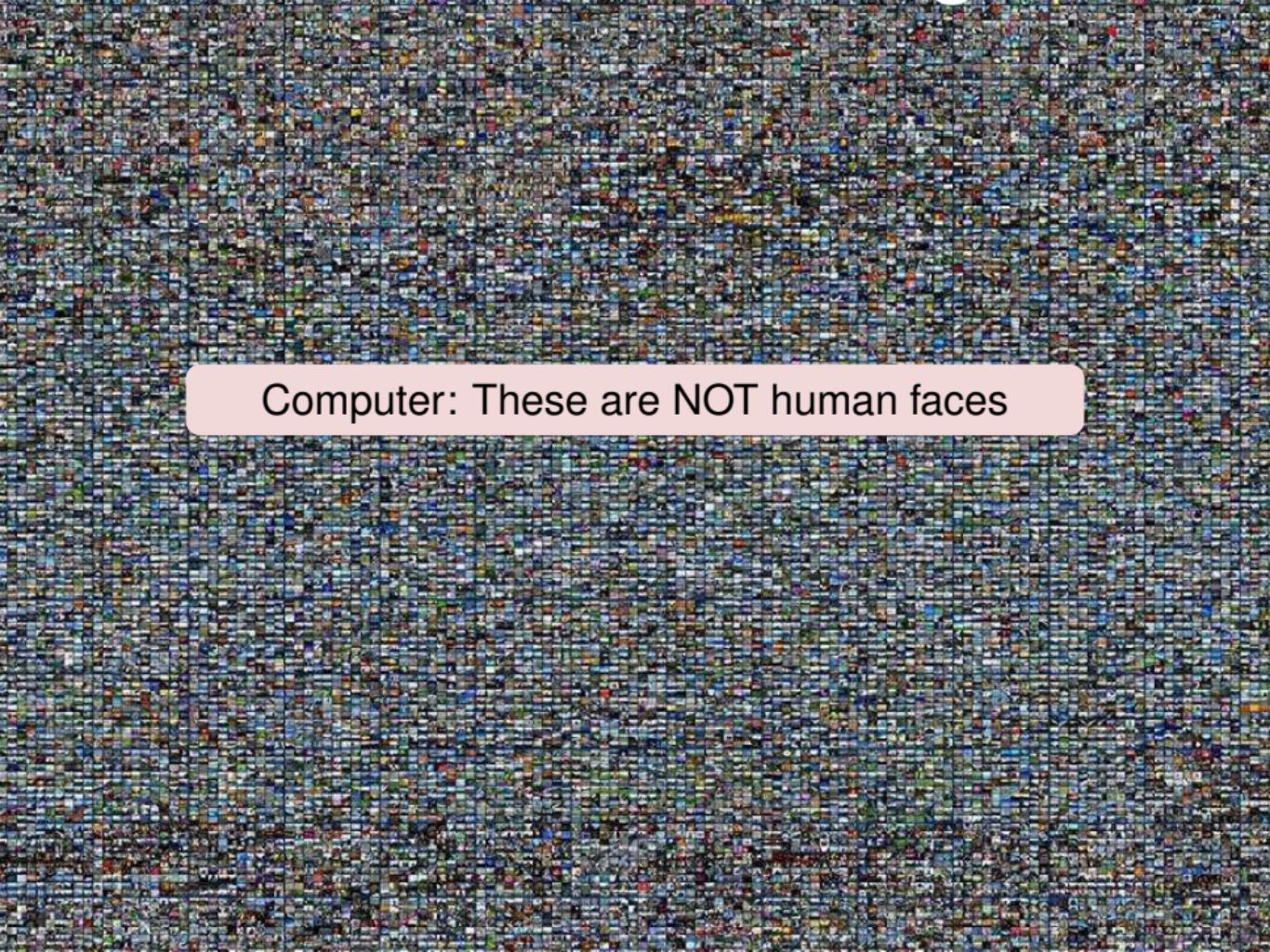
Can this bird fly ?

# Machine Learning

Computational programs (algorithms) that  
learn from experience, i.e., data.



Computer: These are human faces



Computer: These are NOT human faces

Computer: Any human face?



## AI: Modern View (Russell and Norvig)

An intelligent agent (or intelligent machine) perceives, reasons, and acts with great flexibility.

## Machine learning (ML)

An intelligent agent **learns from experience**

## AI: This course perspective

An intelligent agent is able to acquire a **level of understanding** of its world.

## ML: This course perspective

- Learning from experience to build **useful representations** to **understand** the world.
- Then use these representation to make prediction, take decision, etc.

# Structured vs. Unstructured data

## Structured data

Data that is well organized  
(SQL stuff).

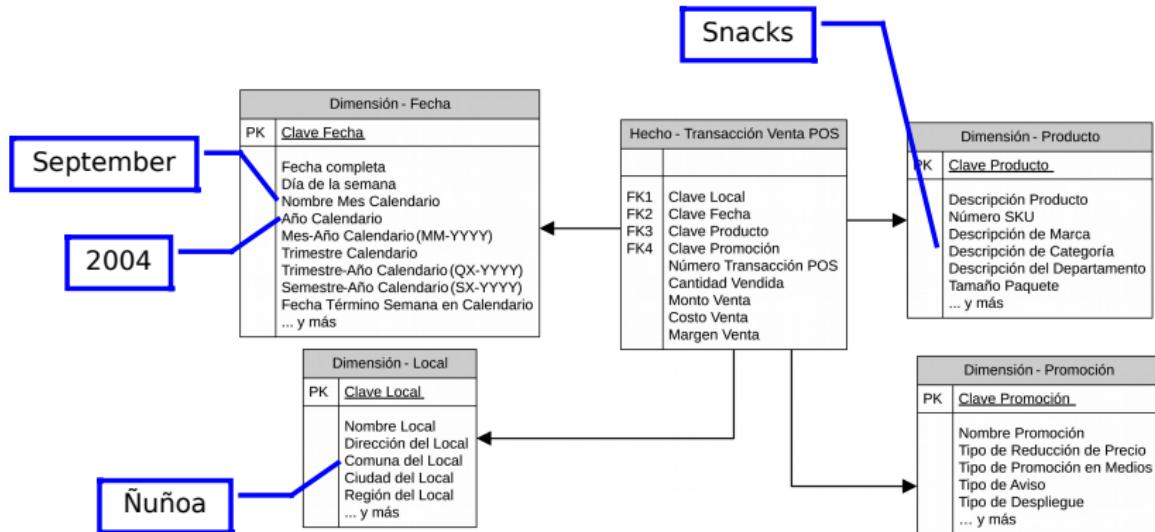
```
SELECT *  
FROM Book  
WHERE price > 100  
ORDER BY title;
```

## Unstructured data

Data that is complex and not  
well organized.



Computer: snacks in Ñuñoa, September 2004?



Computer: What's going on?



# Structured vs. Unstructured data

## Structured data

Data that is well organized  
(SQL stuff).

```
SELECT *  
FROM Book  
WHERE price > 100  
ORDER BY title;
```

## Unstructured data

Data that is complex and not  
well organized.



In simpler terms:

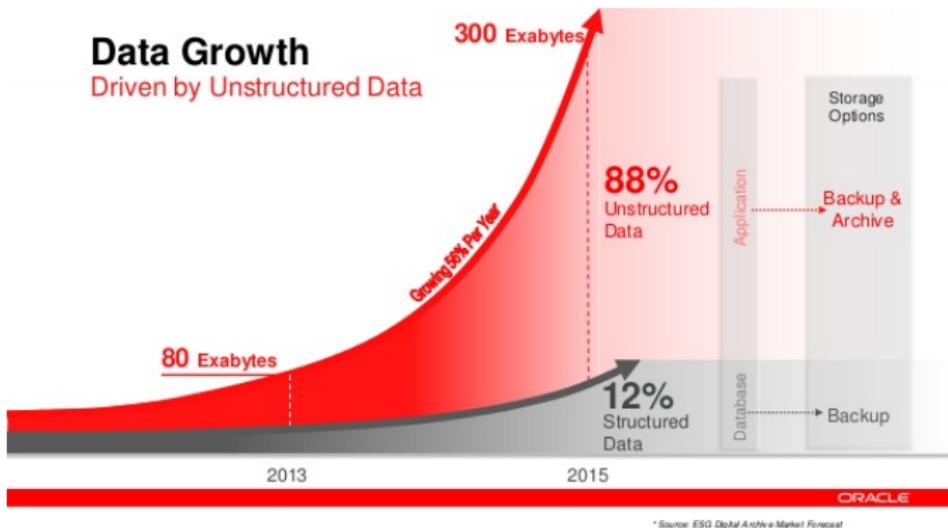
## Structured data

Data that **today's computers can** understand.

## Unstructured data

Data that **today's computers can't** understand.

# Structured vs. Unstructured data

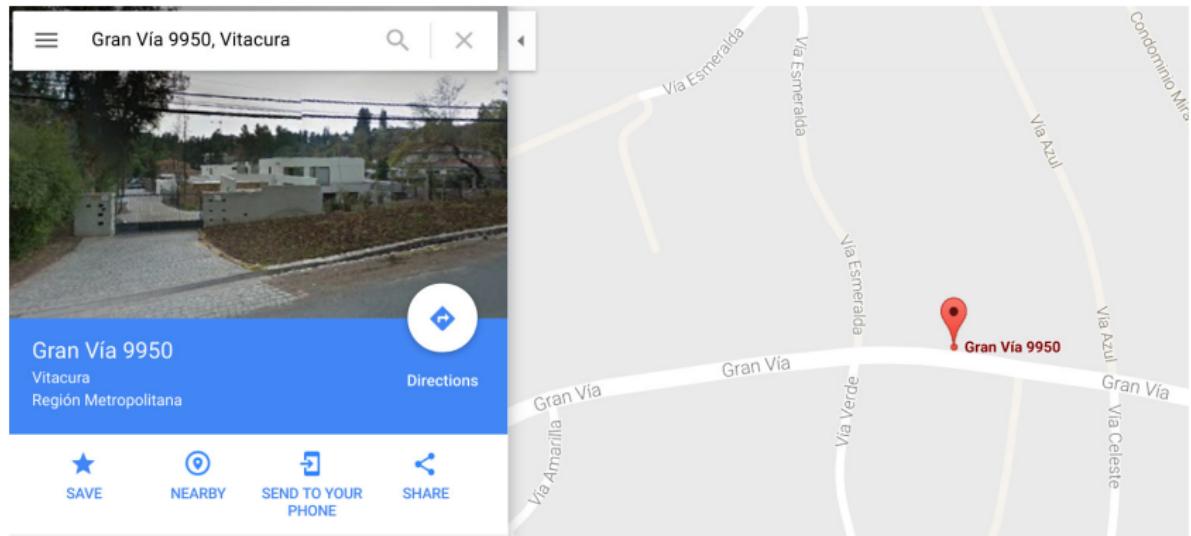


- As we discussed, Machine Learning is about learning from data.
- Currently, Big Data is its great ally.
- By providing large sources of data, Big Data is increasing the accuracy of machine learning techniques.
- The synergistic combination between machine learning and big data is a key element behind the success of deep learning.

Good experience, i.e. data, is key to learn.

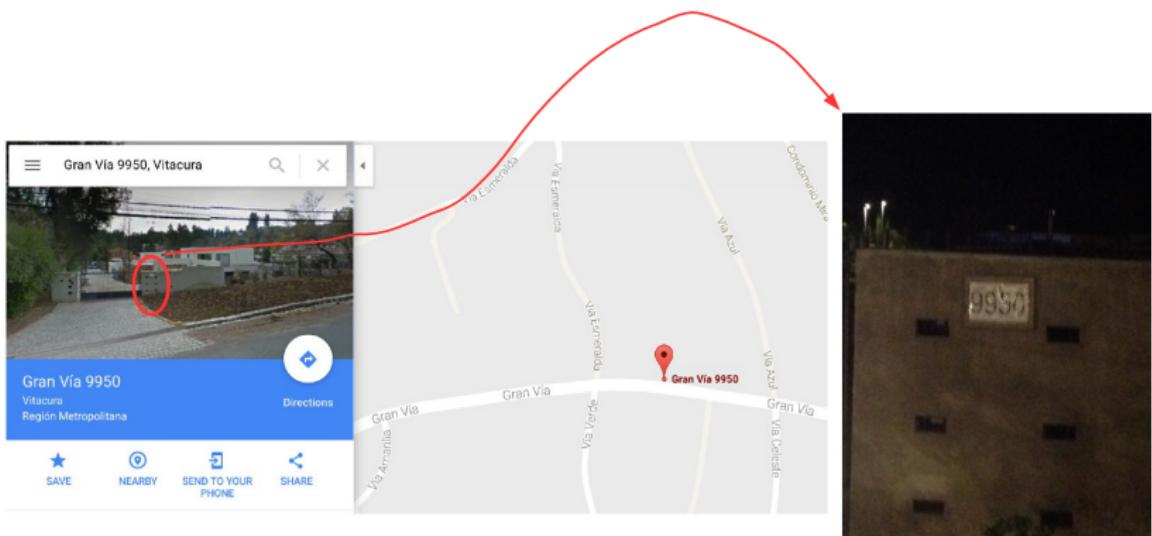
Experience is key

Example: Google Street View.



Experience is key

Example: Google Street View.



How can they do this?

# Experience is key

## Example: Google Street View.

The screenshot shows a Mozilla Firefox browser window with the title "Forgotten User Name or Password - Mozilla Firefox". The address bar indicates the URL is "Forgotten user Na...". The main content area displays a "Forgotten User Name or Password" page for "EasyChair". The page includes a note about account creation, a link to a help article, and a CAPTCHA challenge. A message at the bottom encourages users to enter their email for password reset confirmation.

Forgotten User Name or Password

Note that this page should only be used if you have an EasyChair account. If you do not have one, you should [follow this link to create an account](#). For a detailed description of how password resetting works [read the help article](#).

Enter the text you see in the box. Doing so helps us to prevent automated programs from abusing this service. If you cannot read the text, click the reload image next to the text.

Type the text

Privacy & Terms

Enter either your email address. EasyChair will send you an email asking for a confirmation. This email will also contain further instructions on password resetting.

## Experience is key: Training Data



<http://research.google.com/pubs/pub42241.html>

## Experience is key: Training Data

Select all images with  
**traffic lights**



# Experience is key: Training Data

The image displays two side-by-side screenshots of the Google Image Labeler tool, used for training machine learning models to identify specific visual elements on web pages.

**Left Screenshot:** A modal window titled "Select all images with a store front." contains a grid of 12 small images. The first four images show storefronts, while the others show various other scenes like houses, landscapes, and a truck. Below the grid are three selection buttons: an empty square, a light blue square with "Crawl", and a light blue square with "Crawled". At the bottom are "Go", "VERIFY", and "Complete" buttons.

**Right Screenshot:** A similar modal window titled "Select all images with statues." shows a grid of 12 images. The first four images feature statues, while the rest show buildings, landscapes, and a bridge. Selection buttons and "VERIFY" and "Complete" buttons are at the bottom.

Both windows have a header bar with "Choose", "You are", "Fetch as Google", and "See how Google ranks" buttons. A sidebar on the left of each window contains a note about crawl rules and a "Complete" button. The overall interface is clean and modern, designed for user interaction.

# Experience is key: Crowdsourcing and Human Computation

https://www.mturk.com/mturk/welcome

amazonmechanical turk Artificial Intelligence

Your Account    HITs    Qualifications

Already have an account?  
Sign in as a Worker | Requester

Introduction | Dashboard | Status | Account Settings

**Mechanical Turk is a marketplace for work.**

We give businesses and developers access to an on-demand, scalable workforce.  
Workers select from thousands of tasks and work whenever it's convenient.

**273,682 HITs available.** [View them now.](#)

**Make Money**  
by working on HITs

HITs - Human Intelligence Tasks - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

[Find an interesting task](#)

[Find HITs Now](#)

or [learn more about being a Worker](#)

**Get Results**  
from Mechanical Turk Workers

Ask workers to complete HITs - Human Intelligence Tasks - and get results using Mechanical Turk. [Register Now.](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

[Fund your account](#)

[Load your tasks](#)

[Get results](#)

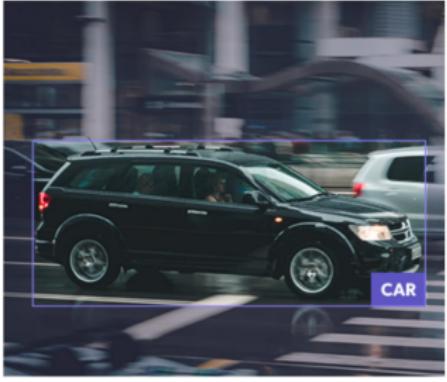
[Get Started](#)

# Experience is key: Crowdsourcing and Human Computation

 hCaptcha

For Websites Labeling Services About Docs Login Get hCaptcha

FROM THE BLOG: How hCaptcha Calculates Rewards



Cost-effective data labeling at scale

Do you have large datasets you need to understand at a human level? hCaptcha provides fast, cost-effective, and high quality data labeling for AI & machine learning companies among many others.

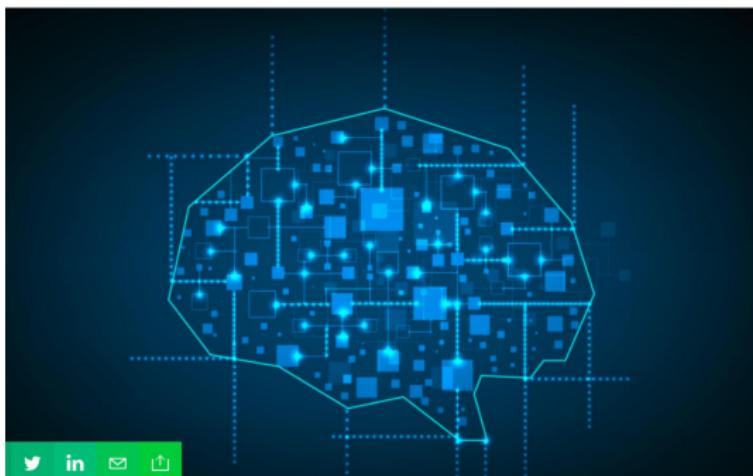
[Request a labeling job](#)

# Experience is key: **Crowdsourcing and Human Computation**

**Appen acquires Figure Eight for up to \$300M, bringing two data annotation companies together**

Anthony Ha @anthonyha / 4 hours ago

Comment



**10/03/2019**

## Experience

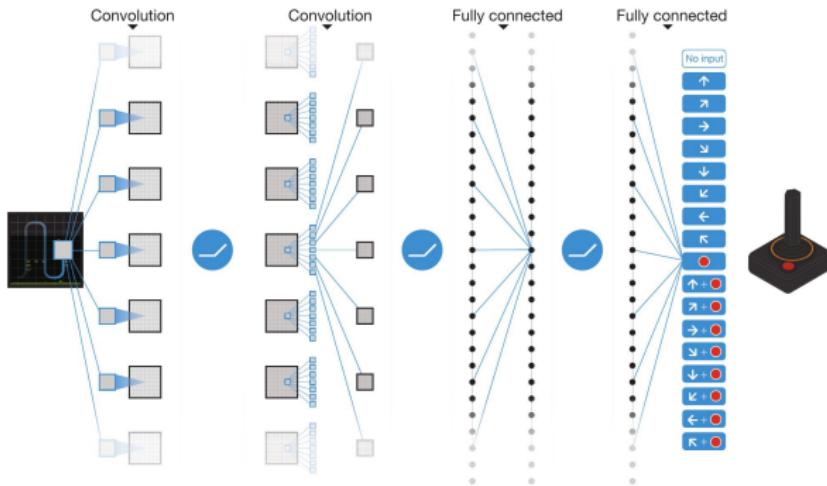
So far we have been focusing on a learning paradigm known as: **Supervised learning**. However, depending of the type of training data available and the goals of the learning process, we can find several popular learning paradigms:

- Supervised learning.
- Unsupervised learning.
- Reinforcement learning.
- Semi-supervised learning.
- Active learning.
- Structural learning.
- Supervised clustering.
- Instance based learning.
- ...

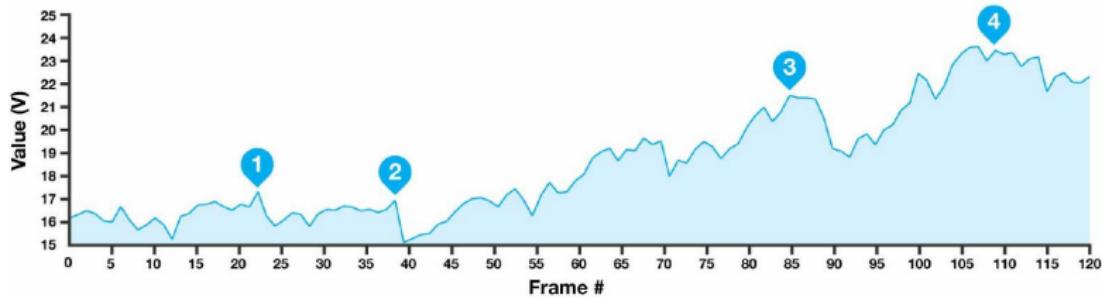
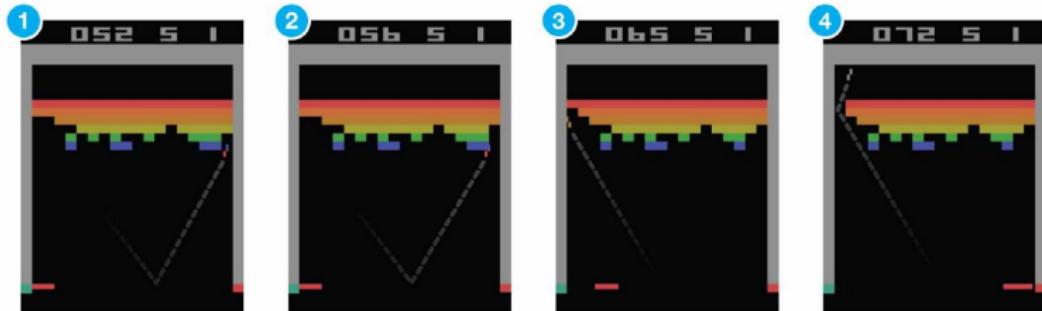
As we will discuss in this course, all these learning frameworks are just different ways to provide semantic to the learning process (grounding).

# Reinforcement learning

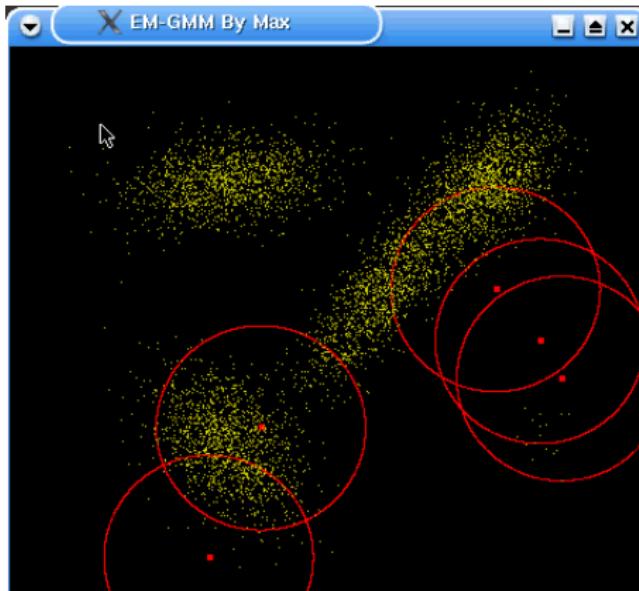
Learning to play Atari games (Mnih et al., NIPS-2014)



## Reinforcement learning



## Unsupervised Learning



## Self-supervised Learning



## Machine Learning

Any computer program that improves its **performance** at some **task** through **experience**.

A computer program is said to learn from **experience E** with respect to some class of **task T** and **performance measure P**, if its performance for tasks in T, as measured by P, improves with experience E.

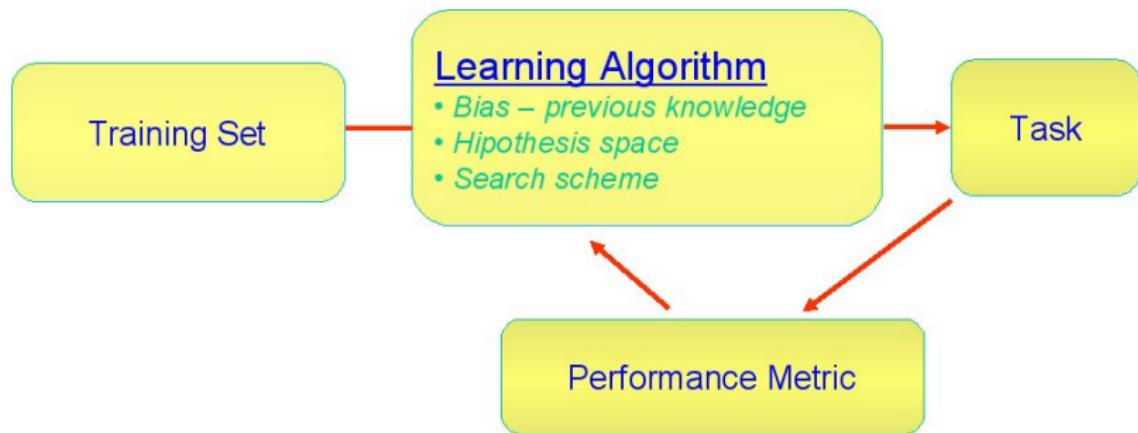
## Key assumption of Inductive Machine Learning

Any hypothesis that approximates a target function well over a sufficiently large set of training examples will also approximate this target function well **over unobserved examples.**

- **OJO** The previous assumption requires that the training data **must be representative** of the target concept or function.

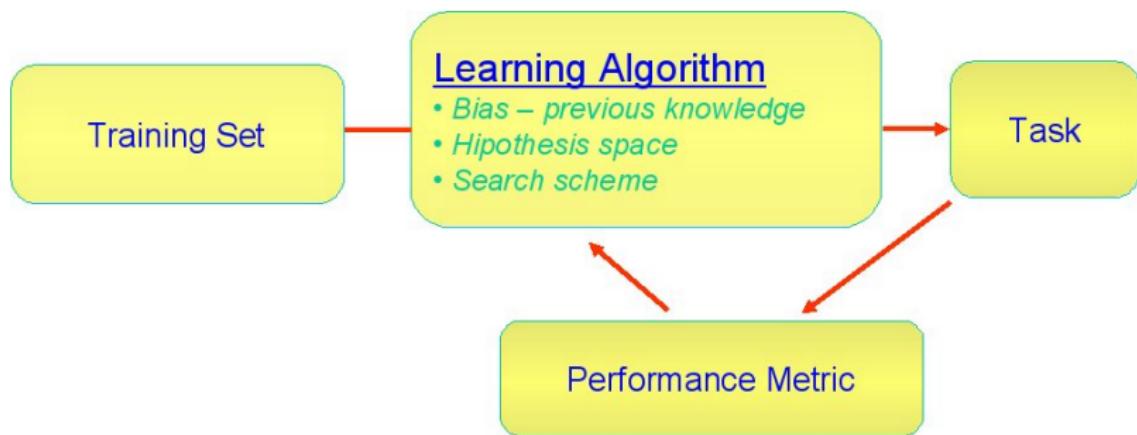
## Machine Learning: Generic View

Learning involves searching through a space of possible hypotheses to find a hypothesis that provides a suitable (best) fit to the available training data and prior constraints (previous knowledge).



OJO There is a new element: **previous knowledge**... Any comment?

Machine learning, 3 main ingredients:  
Representation + Performance + Optimization.



Machine learning, 3 main ingredients:  
Representation + Performance + Optimization.

## Machine learning problem

$$f^* = \arg \min_{f \in \mathcal{H}} \mathcal{L}(f(x)) = \arg \min_{f \in \mathcal{H}} \int_{x_i \in T} \mathcal{L}(f(x_i)) d_T$$

We usually approximate this using a training set Tr:

$$f^* \approx f_{Tr}^* = \arg \min_{f \in \mathcal{H}} \frac{1}{N} \sum_{x_i \in Tr} \mathcal{L}(f(x_i))$$

$\mathcal{H}$ : hypothesis space.

$\mathcal{L}$ : loss function

$f^*$ : optimal hypothesis in  $\mathcal{H}$  under  $\mathcal{L}$ .

## Generic machine learning loss

$$f^* \approx f_{Tr}^* = \arg \min_{f \in \mathcal{H}} \frac{1}{N} \sum_{x_i \in Tr} \mathcal{L}(f(x_i))$$

## Supervised learning

$$f_{Tr}^* = \arg \min_{f \in \mathcal{H}} \frac{1}{N} \sum_{x_i, y_i \in Tr} \mathcal{L}(f(x_i), y_i)$$

## Hypothesis space

- Models have limits (suitable representation?).
- Search in complex spaces (non-lineal or convex) is usually a hard and exhausting process (right optimization tool and performance metric?).

Risks,

- If the hypothesis space is too flexible, there is a higher risk to converge to misleading hypothesis (suffer overfitting problems).
- If the search tool is too limited, there is a higher risk to being unable to find a good hypothesis (converge to local optimals).

Concepts that you should know about ML.

- Classical AI and ML views.
- Generic view of ML.
- Main elements of a ML technique: Representation + Performance + Optimization.
- Supervised and unsupervised ML.
- Overfitting.
- Training, test, validation sets.

Recommended concepts that you should review. We will discuss some of them in future lectures:

- All the ideas in this week lecture: “A Few Useful Things to Know About Machine Learning” by Pedro Domingos.

**Tapping into the “folk knowledge” needed to advance machine learning applications.**

BY PEDRO DOMINGOS

# A Few Useful Things to Know About Machine Learning

MACHINE LEARNING SYSTEMS automatically learn programs from data. This is often a very attractive alternative to manually constructing them, and in the last decade the use of machine learning has spread rapidly throughout computer science and beyond. Machine learning is used in Web search, spam filters, recommender systems, ad placement, credit scoring, fraud detection, stock trading, drug design, and many other applications. A recent report from the McKinsey Global Institute asserts that machine learning (a.k.a. data mining or predictive analytics) will be the driver of the next big wave of innovation.<sup>15</sup> Several fine textbooks are available to interested practitioners and researchers (for example, Mitchell<sup>16</sup> and Witten et al.<sup>24</sup>). However, much of the “folk knowledge” that

is needed to successfully develop machine learning applications is not readily available in them. As a result, many machine learning projects take much longer than necessary or wind up producing less-than-ideal results. Yet much of this folk knowledge is fairly easy to communicate. This is the purpose of this article.

## » key insights

- Machine learning algorithms can figure out how to perform important tasks by generalizing from examples. This is often feasible and cost-effective where manual programming is not. As more data becomes available, more ambitious problems can be tackled.
- Machine learning is widely used in computer science and other fields. However, developing successful machine learning applications requires a substantial amount of “black art” that is difficult to find in textbooks.
- This article summarizes 12 key lessons that machine learning researchers and practitioners have learned. These include pitfalls to avoid, important issues to focus on, and answers to common questions.

