Building Smart Sensors for geophysics and the environment using microcontrollers and AI

lecture #1

Building Smart Sensors for geophysics and the environment using microcontrollers and AI

Stephen Brown, Ph.D.
srbrown2@aprovechar.org

thanks to:

The Tiny Machine Learning Open Education Initiative

(tinyml.seas.harvard.edu)

for making slides and resources available for this tutorial

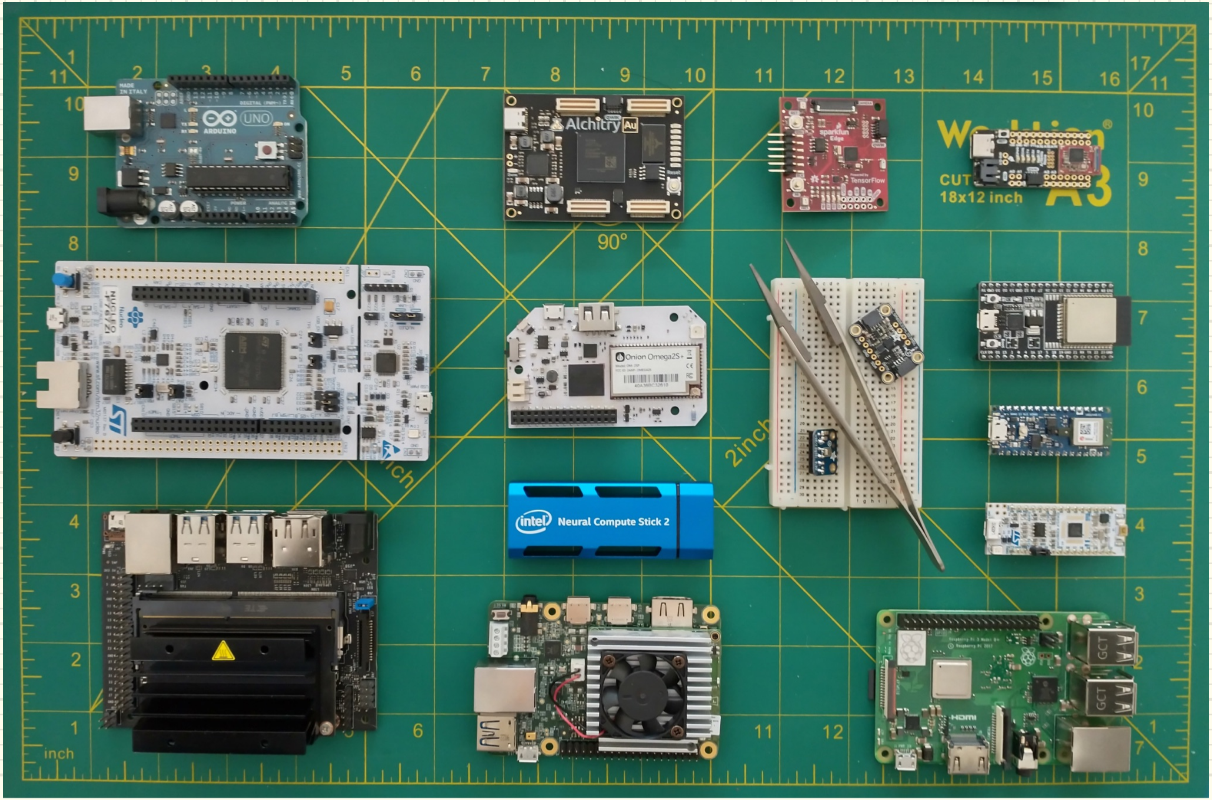
Outline for today:

1. microcontrollers
2. sensors
3. embedded machine learning
4. ethics

Microcontrollers

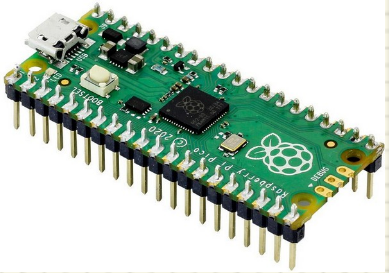
slides thanks to:
John Teel
predictabledesigns.com

Examples



ARDUINO
UNO
Alchitry Au
90°
intel
Neural Compute Stick 2
Onion Omega2S+
CUT
18x12 inch
A3

Micro-controller vs Micro-computer/processor



	Microcontroller	Microprocessor
Design complexity	Low	High
Clock speed	Slow	Fast
Operating system	No	Yes
Processing speed	Low	High
Power consumption	Low	High
Memory	Small / Internal	Large / External
I/O pins	Yes	No
Number of bits	8-32 bits	32-64 bits
Cost	Low	High

Table 1 – Comparison of a microcontroller versus a microprocessor

Types of memory

- RAM and FLASH memory are already built into most microcontrollers. RAM memory is for temporary data storage. RAM is considered volatile memory because it loses its contents once power is turned off.

- FLASH is non-volatile memory. It retains its contents even with power turned off. FLASH is primarily used for storing the program (called firmware).

- EEPROM (Electrically Erasable Read Only Memory) is also non-volatile memory, but it’s typically used for storing small amounts of data such as configurations and variable states. EEPROM has the advantage of allowing erasure on a per byte basis, whereas FLASH memory only permits an entire block, or page, of bytes to be erased.

Digital I/O (GPIO)

- Digital inputs and outputs on microcontrollers are called General Purpose Input Output or GPIO. GPIO pins are logic level pins (either high or low) used for input or output. They can generally drive loads up to a few tens of milliamps.

- Keep in mind that it’s usually necessary to have some sort of external drive circuit that will off load the drive requirements. Additionally, all the pins added together must be under the maximum allowed total current.

- Most pins on microcontrollers have multiple possible functions. Many internal functions are not available on certain pins.

Analog inputs and outputs

- Analog voltages are measured by the microcontroller using an embedded specialty circuit called an Analog to Digital Converter (ADC).

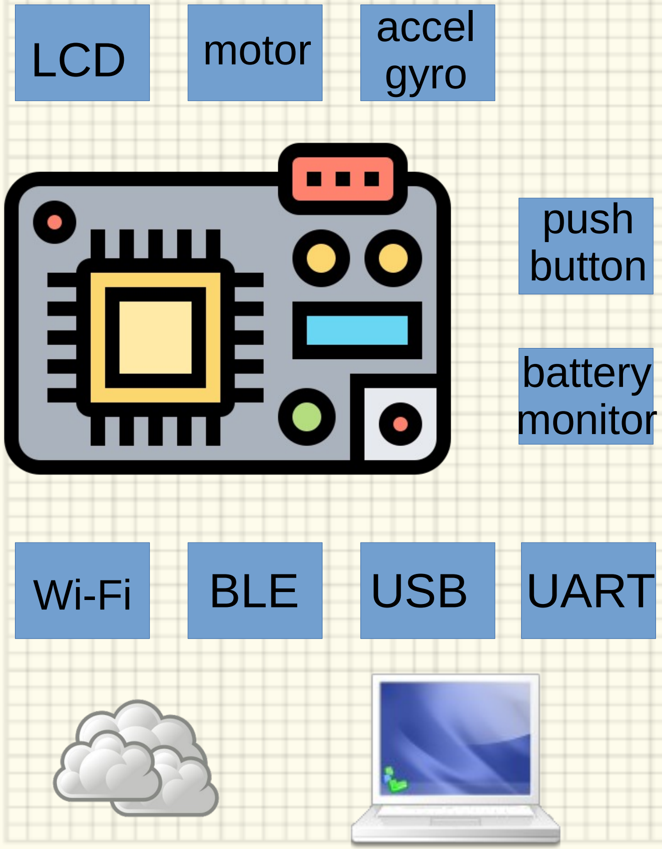
- There are two methods for outputting an analog voltage from a microcontroller.

  - The most precise, but more complex method is to use a Digital to Analog Converter (DAC) circuit.

  - The second method is to use a Pulse Width Modulation (PWM) generator. A PWM is a digital signal that oscillates between a high and low level. The microcontroller is able to control the percentage of time this signal is high as determined by the firmware program.

  - The PWM signal can then be fed through a low-pass filter to convert it to an analog voltage. By controlling the percentage of time the PWM signal is high versus low, you can accurately control the analog voltage level on the output side of the filter.

Peripherals and Communications



- General Purpose I/O (GPIO)
- Inter-Integrated Circuit Interface (I2C)
- Serial Peripheral Interface (SPI)
- wireless network protocol (Wi-Fi or wifi)
- Low Energy Bluetooth (BLE)
- Universal Serial Bus (USB)
- Universal Asynchronous Receiver Transmitter (UART)

Universal Asynchronous Receiver Transmitter (UART)

- The oldest method of serial communication still used today in microcontrollers is called UART. UART’s have been around since the early 1970’s. As the name implies, UART is an asynchronous protocol, meaning there is no clock signal to synchronize the signal timing. The simplest form of UART uses only two signal lines: receive (RX) and transmit (TX).

- A UART is commonly combined with other standards, such as RS-232 or RS-485, for external serial communication over a longer distance. A synchronous UART called a USART is also available on some microcontrollers.

Inter-Integrated Circuit (I2C)

- I2C is a synchronous serial bus protocol used for serial communication between components on the same board. The primary advantages of I2C is that it only requires two wires, is synchronous, and that unlike UART or SPI, I2C is a bus protocol.

- Being a true bus means that I2C allows multiple devices to be connected to the same two wires. Instead of a slave-select signal as with SPI, I2C uses a unique address for each device. This fact greatly simplifies on-board signal routing.

- As with SPI, I2C uses a clock signal to synchronize the communication. However, unlike SPI, I2C is only half-duplex so it uses a single signal for both data in and data out. I2C can’t send data in both directions simultaneously.

- The price for this simplicity is that I2C is considerably slower than SPI. So I2C is primarily used for interfacing two components that don’t require high data transfer speed. It’s especially popular for interfacing sensors (which are relatively slow).

Universal Serial Bus (USB)

- USB is the fastest serial communication protocol available on most microcontrollers. It is commonly used to transfer larger amounts of data between external peripherals.

- Unlike the other serial protocols we've discussed which are used for short distance, on-board communication, USB is designed for longer distance communication with external peripherals.

- USB is a master/slave protocol. The host (master) can transmit data to the device (slave), and the host can read data from the device. The key is the host always initiates communication, and the peripheral device only responds.

Wireless interfaces

- Some microcontrollers include Bluetooth, ZigBee, Wi-Fi, and other wireless interfaces. Bluetooth Low-Energy microcontrollers are especially common.

Sensors

slides thanks to:
Sebastian Büttrich
University of Copenhagen

Motivation – why we talk about sensors

- **Data-centric ML**
  - The **importance of quality** data sets
- So far, we have mainly looked at human-generated data (voice, photography, social media ..) but there is a whole other class of data, which is generated through **scientific measurement**.
  - Its importance e.g. in Environmental Science, Earth Observation, Energy transition, Climate Change, Logistics, Buildings, Urban Planning, Agriculture, Aquaculture, ... to name a few.

Sensors / Definition I

- A sensor is a device, module, or subsystem that **transforms a property of the physical ("real") world to a signal that can be read by electronic/digital systems –** to **"data".**
- Properties, in the widest sense, can be **events, changes, states/static properties**.
- There are many possible (and conflicting) definitions.

  Does a plant have sensors? (Shown on title slide: Drosera rotundifolia)

Sensors / Definition II

Note terms that might be problematic:

real world,

physical world

“to translate into data” -

data is not inherently there – it is a human construct,

created under technical, social, cultural conditions

Sensors / Definition IV

Clarification:

we will distinguish between the

the sensor -

and the

full **sensor node**, which includes a **sensor** and an

embedded system, a board, a device

(with processor/MCU, memory, storage, I/O,

on-board communications, networks, etc)

both of which together make up the

sensor node or device.

In practice, we often find the term **sensor** to denote the whole system, e.g. a *WiFi CO2 sensor, LoRa watermeter.*

source: purpleair, plantower

Sensors / Principle

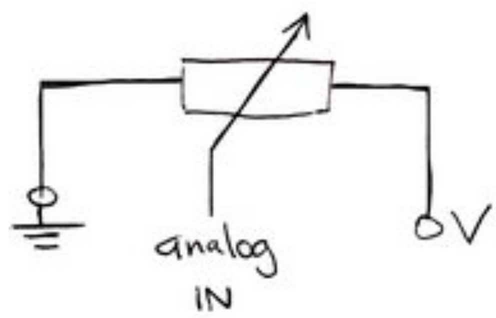
- A sensor typically transforms a "real world" property into a

  voltage (or current, which then gets converted to a voltage)

  which then may is digitized.
- Some **physical effect** is needed to make that transformation.
- To that end, for experiments,

  a voltage source with a potentiometer fully replaces any type of

  (analog) sensor.



Sensors / Classification, types

- **Analog / Digital**

  the output is an analog voltage or already digital (→ ADC)
- **Active / Passive**

  with regards to the measurement – does the sensor impact the object of interest? Discuss e.g. light sensors, watermeters
- **Powered / Non-powered**

  – do we need to power the sensor in order for it to work?
- **Physical / Chemical / Biological**
- Field readiness, autonomy
- Cost: low-cost vs. (expensive) lab grade sensors

Sensors / Overview I

Push Button

Displacement

Pressure, weight, bend, vibration

Distance

Proximity

Position

Motion

Acceleration

Orientation (Magnetic, Gyroscope)

Hall/Reed

Voltage / Current

RF Intensity

Light

Sound

Pressure, barometer

Temperature Humidity, soil moisture

Wind (speed, direction)

Radioactivity

Water

→ Level, flow, chemistry

Air

→ indoor/outdoor

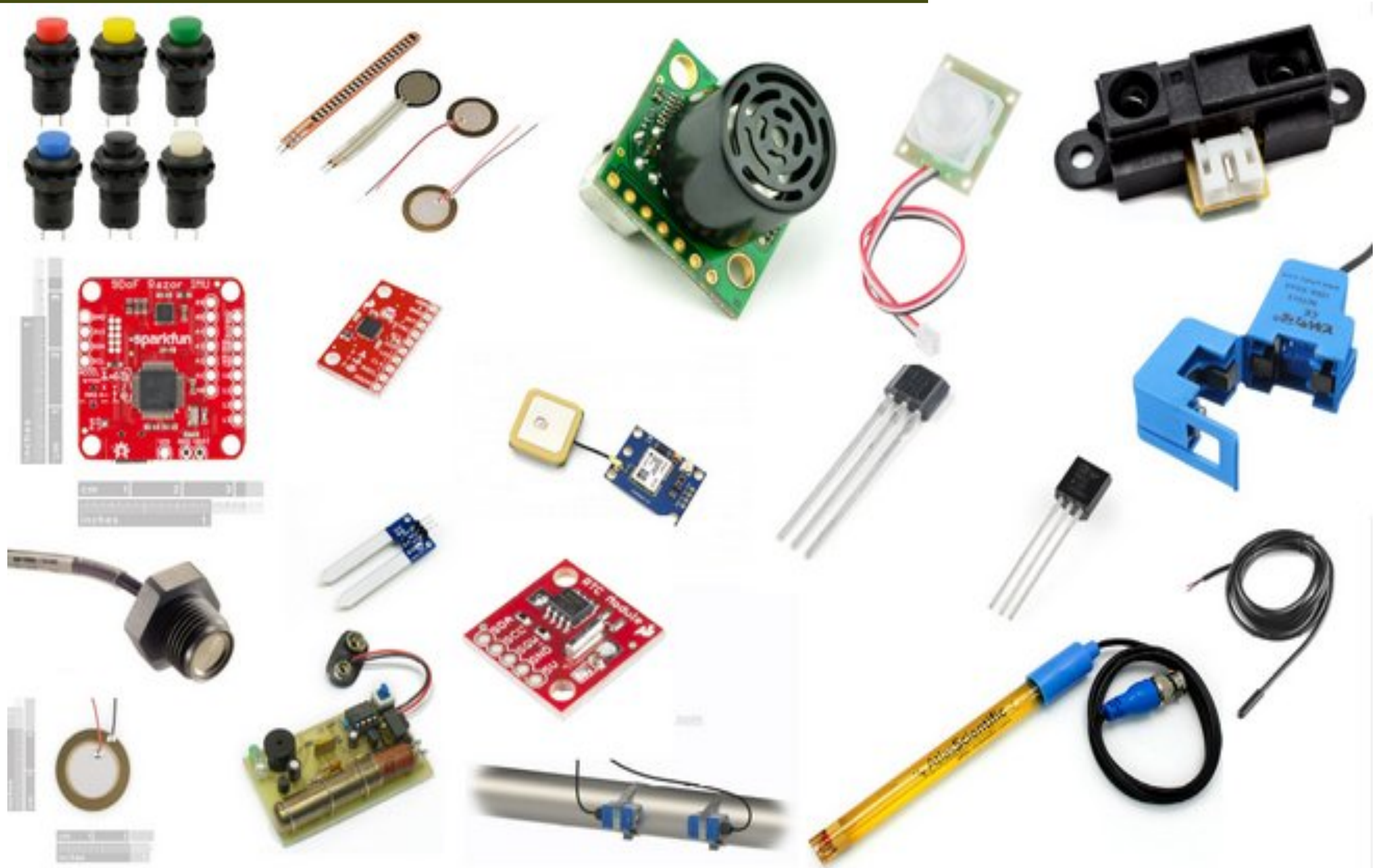
→ gaseous / particulate

→ Smoke, Fire

Biological / Health

→ heart, pulse, breath, eye, ...

What sensors look like



Sensors / Mobile Phone

Importance of embedded machine learning

slides thanks to:
Pete Warden
@google.com

Why is Machine Learning on Embedded Systems so Important?

petewarden@google.com

Slides at goo.gle/emb-ml-importance

We can capture much more data than we can send to the cloud

Is this true?

- What are the limits on sending data to the cloud?
- How much sensor data can we capture on the edge?
- What opportunities are we missing without this data?
- How does machine learning help?

Limits on Sending Data

Why is it hard to send data to the cloud?

Phones and mains-powered devices:

- Lack of geographic coverage
- Electronically-noisy environments
- Transmission costs
- Cloud processing costs
- Latency

Why is it hard to send data to the cloud?

In ideal circumstances, you might get broadband speeds from your phone.

Most of the time, in most places, that won't be true, and is unlikely to change in the medium term.

The most successful "near edge" devices with high bandwidth expectations are security cameras, so it is possible. Even these rely on a lot of client-side engineering (compression, deltas, buffering, latency tolerance) to work, and charge for storage.

How much power can you get from an AA?

There are around 10,000 Joules in an AA battery.

There are 31,536,000 seconds in a year.

So, if you rely on one AA battery, in theory you’ve got an average 0.000317 watts of continuous power, or 317 microwatts.

In practice it’s a lot more complex with battery characteristics, but a rule of thumb is that three AAs might give you about one milliwatt of power for a year, if you’re lucky.

https://commons.wikimedia.org/wiki/File:02_-_Single_Energizer_Battery.jpg

Slides at goo.gle/emb-ml-importance

What about energy harvesting?

Method	Microwatts per cm²
Outdoor Solar	10,000
Industrial Temperature	10,000
Industrial Vibration	100
Human Temperature	25
Indoor Solar	10
Human Vibration	4
GSM Radio	0.1
Wifi Radio	0.001

See goo.gle/energy-data

In most environments, there's less than 1,000 microwatts, or one milliwatt, available from a realistic form factor.

What does that all mean?

The bottomline is that any unattended device without a mains connection will be lucky to have an average of one milliwatt of power.

Why does this matter for data transmission?

Current radio technologies all use a lot more than one milliwatt when they're active.

Even BLE is often around 100 milliwatts, with Wifi, cell, or satellite technologies requiring even more.

That means we can't power continuous transmission of data to the cloud, or even a reasonable duty cycle, from these kinds of devices.

Limits on capturing data

How much data can we capture in one mW?

Himax HM01B0 image sensor: 30 FPS at 320x320 = 3MB per second

Vesper VM1010: 16,000 Hz = 32 kilobytes per second (for 300 mW)

Bosch BMI270 IMU: 6,400 Hz = 13 kilobytes per second

How much data can we process in one mW?

Ambiq Cortex M4 at 48MHz clock rate means we can handles tens of megabytes a second, even with conservative estimates.

The Great Data Bottleneck

We can capture much more data than we can send to the cloud

How does machine learning help?

Advances in machine learning since 2010 have centered on the application of neural networks and deep learning.

These have proven to be very effective in taking noisy sensor data (images, audio, accelerometers) and robustly spotting patterns (objects, speech, noises, gestures).

Isn’t ML something for data centers?

Thousands of applications and billions of users already run TensorFlow Lite on their iOS and Android phones!

Useful models (for example wake word detection) can fit in less than 30KB.

They can take less than a million arithmetic operations to execute.

It’s possible to run them without floating-point support.

ML on Embedded

Can take those streams of sensor data and turn them into actionable information:

- Is there a crop pest in this field?
- Is there a person on the subway track?
- Did someone give a voice command?

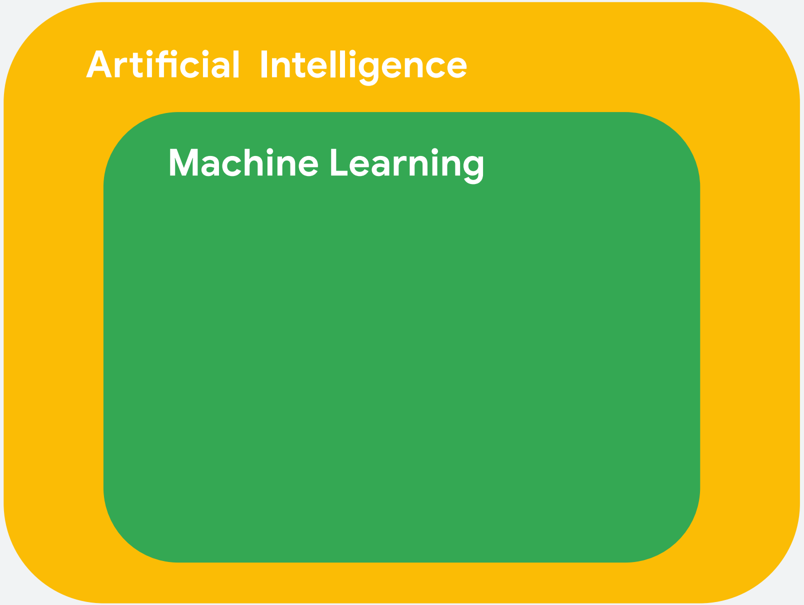
ML fills the gap between the sensor and radio capabilities. The radio can be turned on when something important happens.

What is Machine Learning?

slides thanks to
Larry Moroney
@google.com

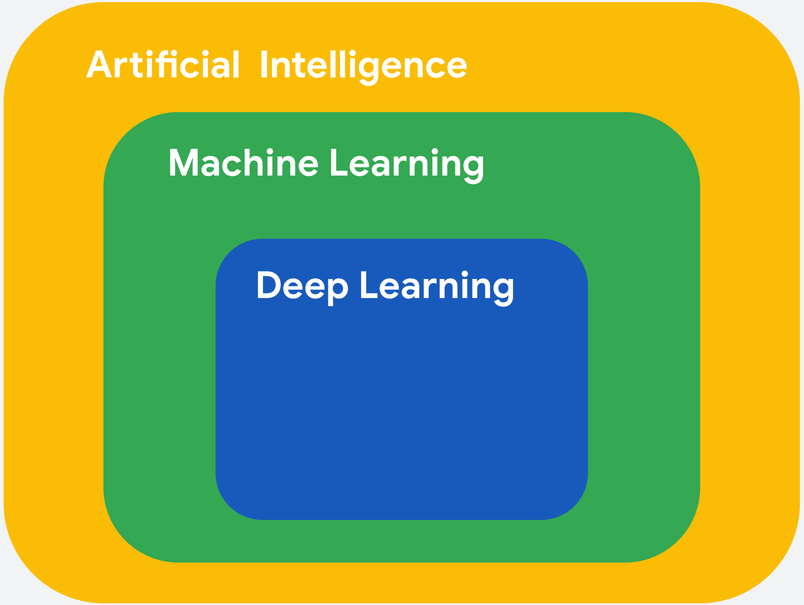
What is Machine Learning?

1. **Machine Learning** is a subfield of **Artificial Intelligence** focused on developing algorithms that learn to **solve problems by analyzing data for patterns**

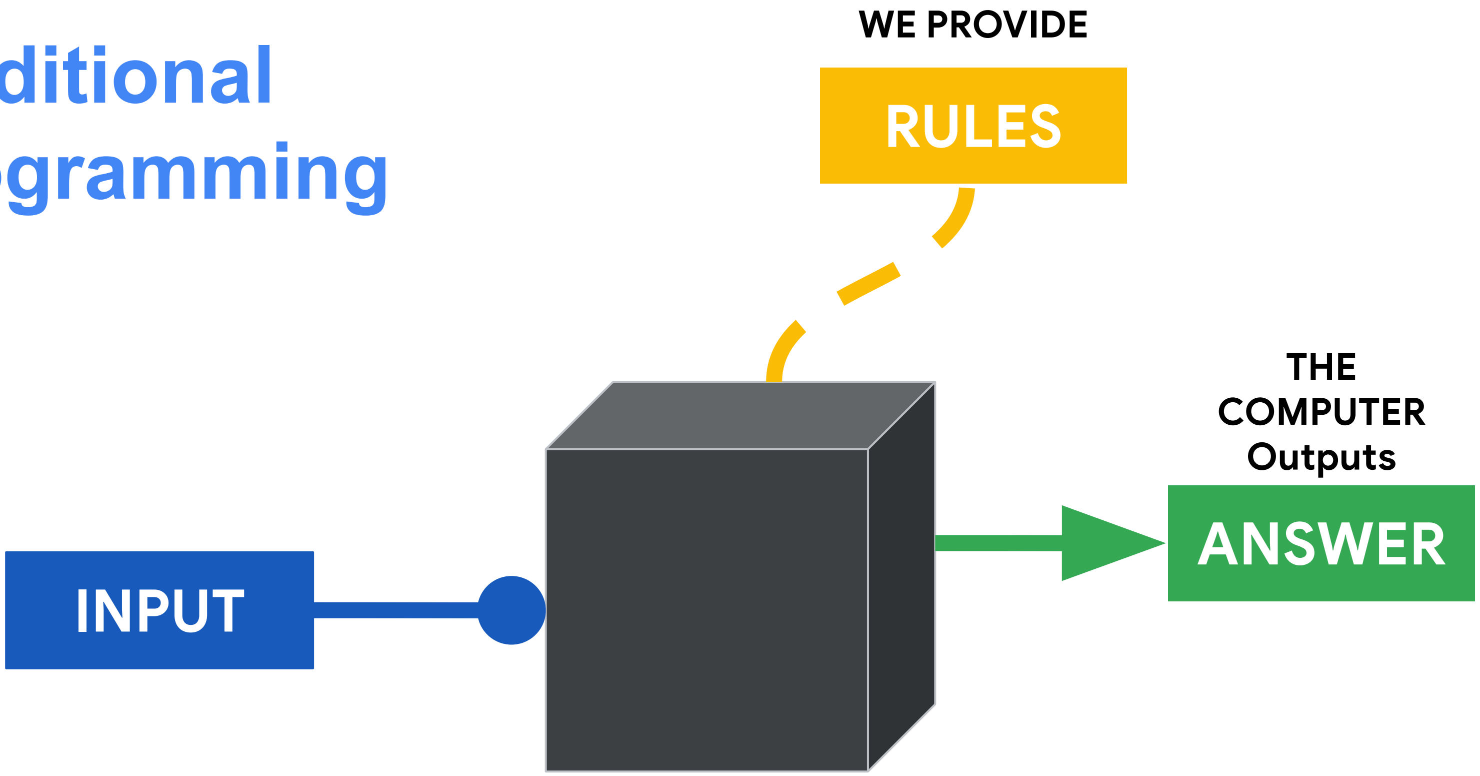


What is (Deep) Machine Learning?

1. Machine Learning is a subfield of Artificial Intelligence focused on developing algorithms that learn to solve problems by analyzing data for patterns
2. **Deep Learning** is a type of **Machine Learning** that leverages **Neural Networks** and **Big Data**



Traditional Programming



Let's try to figure out **what** she's doing?

```
if (speed < 4):
      then walking
```

```
if (speed < 4):
      then walking

else:
      running
```

data we can gather

input: **speed**

Write a **rule**

extend the rule

Let’s try to figure out **what** she’s doing?

```
if (speed < 4):
      then walking
```

```
if (speed < 4):
      then walking


else:
      running
```

```
if (speed < 4):
      then walking


else if (speed < 12):
      then running
else:
      biking
```

Let’s try to figure out **what** she’s doing?

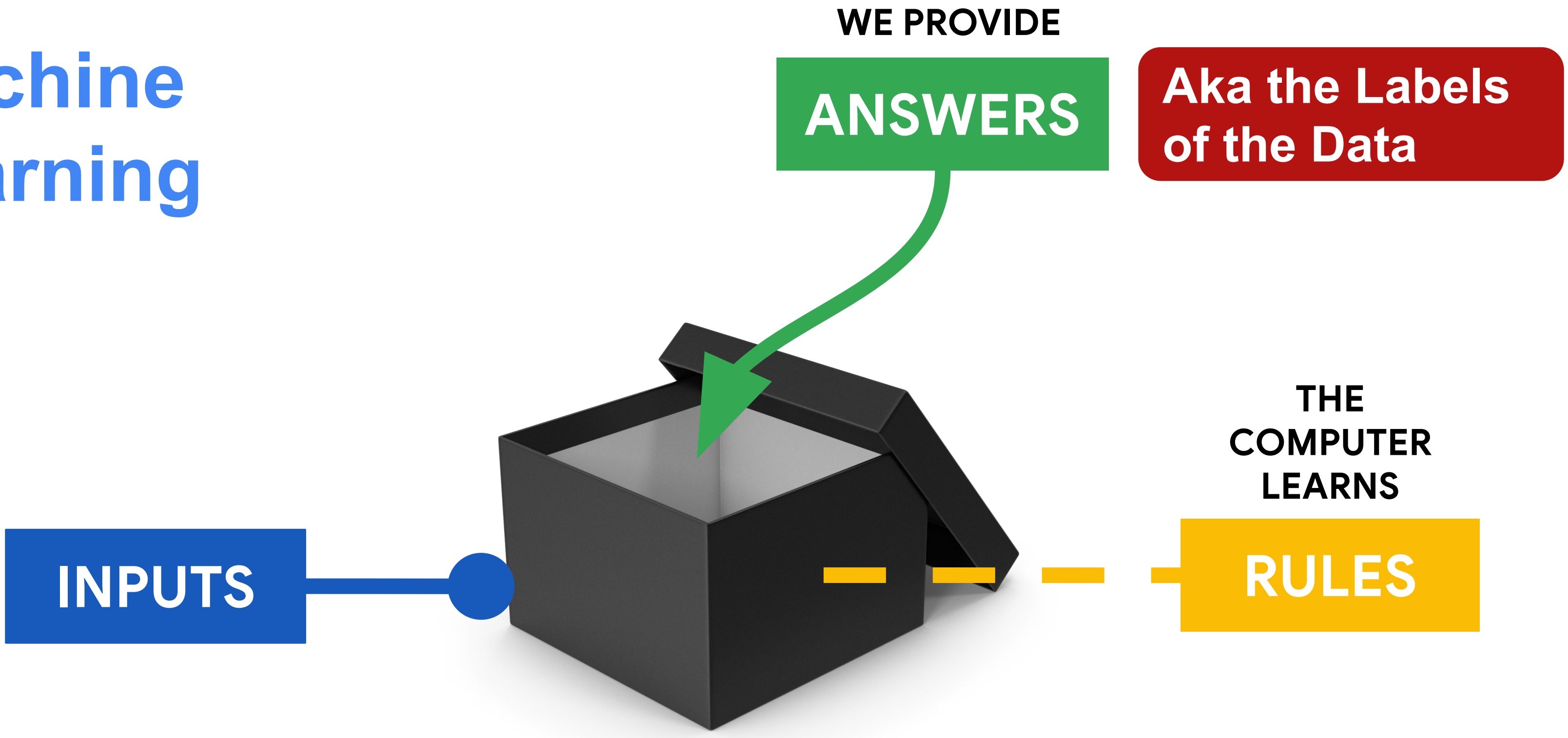
```
if (speed < 4):
      then walking
```

```
if (speed < 4):
      then walking

else:
      running
```

```
if (speed < 4):
      then walking

else if (speed < 12):
      then running
else:
      biking
```

```
?? WHAT IS THIS ??
```

Machine Learning



Let’s try to figure out **what** she’s doing?

010101010010001101010
010101001010010010100
101010110101001010011

111101010010010101011
010100101001010101000
110101100101010011111

000011101011101011010
010101111010110101010
110101111110010010110

011111101011101010100
101011101010110101010
111111111001000011100

walking **running** **biking** **golfing**

Let’s try to figure out **what** she’s doing?

01010101001000110101
01010100101001001010
10101011010100101001

walking

11110101001001010101
01010010100101010100
11010110010101001111

running

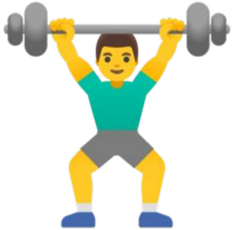
00001110101110101101
01010111101011010101
11010111111001001011

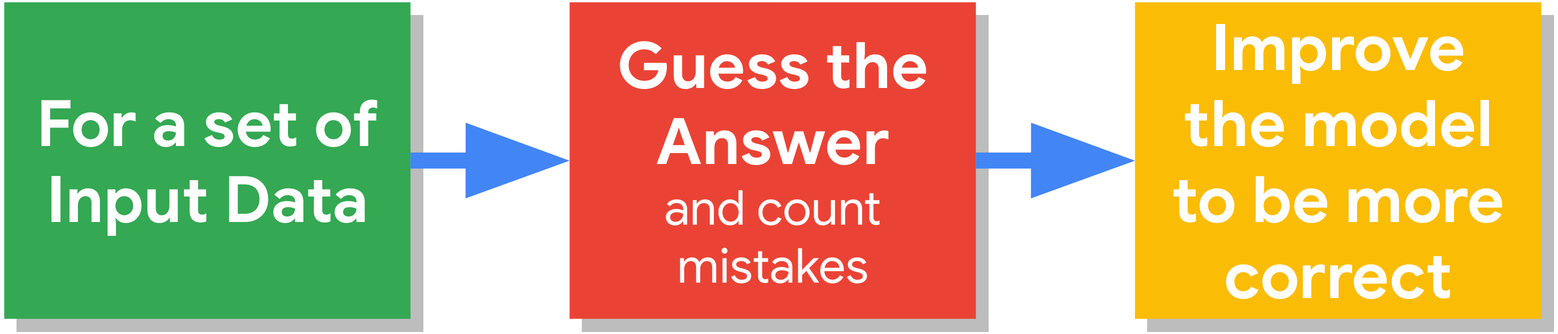
biking

01111110101110101010
10101110101011010101
11111111100100001110

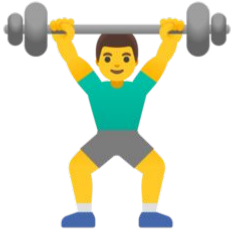
golfing

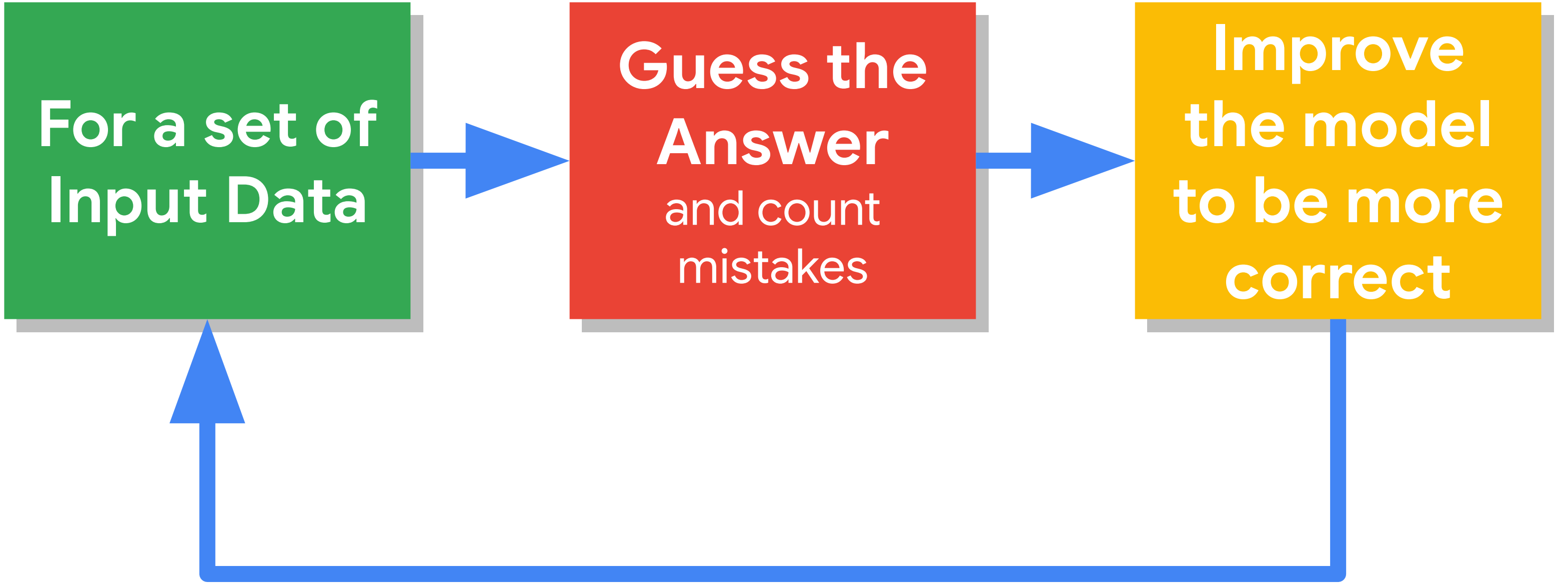
Training the machine



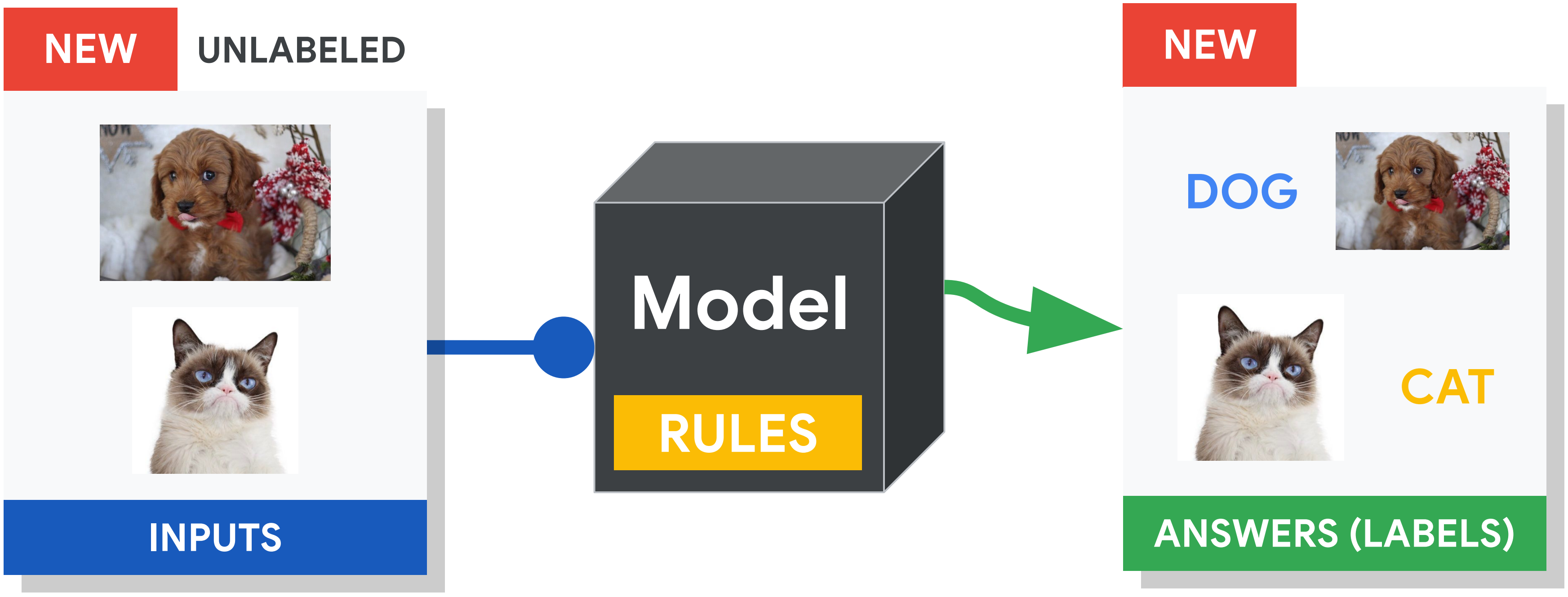


Training the machine





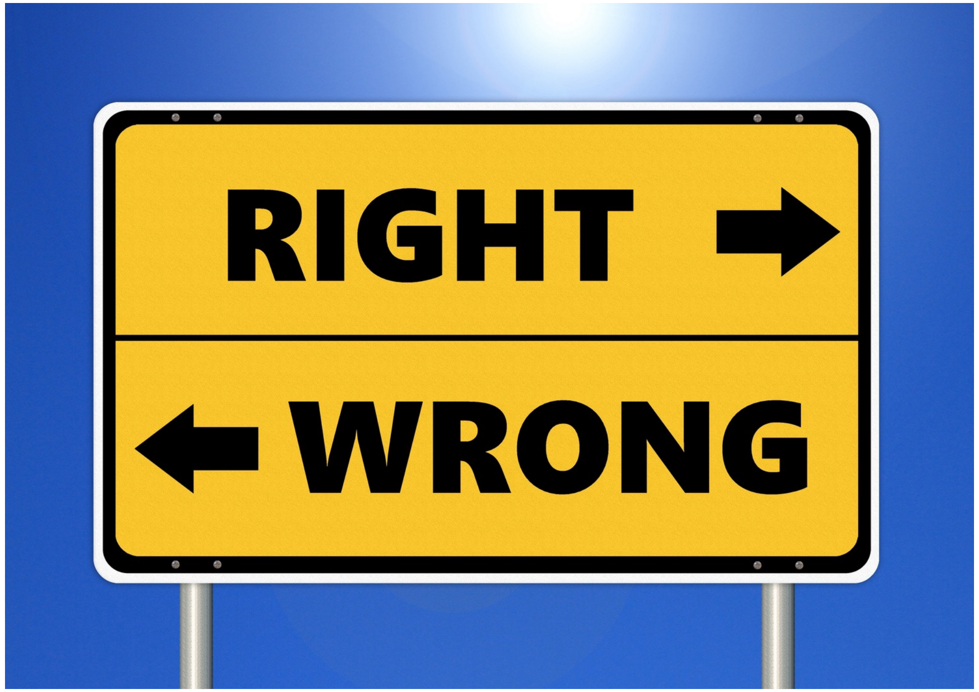
After it’s **learned** use it for **inference**:



Ethics in Machine Learning

slides thanks to
Drs Gamundani and
Strickwick
@ University of United Nations

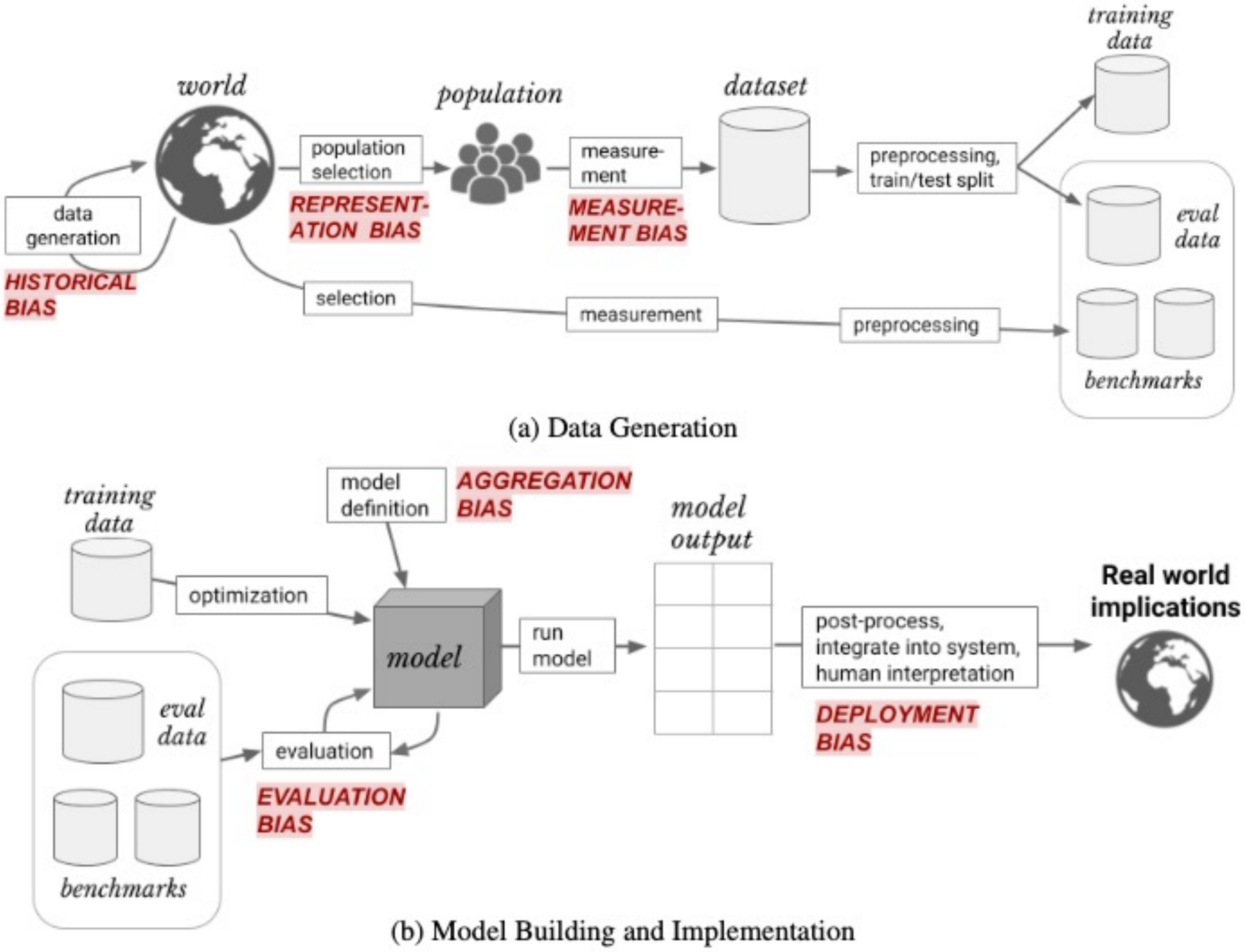
Ethics principles overview



Unintended consequences of ML due to Bias

- There are many ways to understand bias (statistics, social science, ...).
- Common rhetoric is that various unwanted consequences of ML algorithms arise in some way from "**biased data**".
- Empirical findings has shown that data-driven methods can unintentionally encode human biases and introduce new ones: **Machine Learning can amplify bias !**
- Biased data is the product of many factors.

Bias in ML pipelines



Extract from Harini Suresh, Jogn V. Guttag, A Framework for Understanding Unintended Consequences of Machine Learning", 2020

UNITED NATIONS UNIVERSITY
Institute in Macau

Historical bias

- Comes from the fact that people are biased, processes are biased, the society is biased.
- It can exist even given perfect sampling and feature selection.
- Any dataset involving humans can have this kind of bias: medical data, sales data, etc …
- https://nypost.com/2017/11/30/google-translates-algorithm-has-a-gender-bias/

Turkish - detected	English
o bir aşçı	she is a cook
o bir mühendis	he is an engineer
o bir doktor	he is a doctor
o bir hemşire	she is a nurse
o bir temizlikçi	he is a cleaner
o bir polis	He-she is a police
o bir asker	he is a soldier
o bir öğretmen	She's a teacher
o bir sekreter	he is a secretary
o bir arkadaş	he is a friend
o bir sevgili	she is a lover
onu sevmiyor	she does not like her
onu seviyor	she loves him
onu görüyor	she sees it
onu göremiyor	he can not see him
o onu kucaklıyor	she is embracing her
o onu kucaklamıyor	he does not embrace it
o evli	she is married
o bekar	he is single
o mutlu	he's happy
o mutsuz	she is unhappy
o çalışkan	he is hard working
o tembel	she is lazy

Representation bias

- Arises from how we sample from a population during data collection process.
- Particulary common problem in datasets
- Example: lack of geographical diversity in datasets like ImageNet results in bias towards Western cultures. The same regarding ethnicity and gender. “groom” category show mostly white people.

Deployment bias

- Arises when there is a mismatch between the problem a model is intended to solve and the way in which it is actually used.
- Often the case, when the model is built in a quite isolated way but it used at the end in a complicated socio-technical environment
- Example: ***Risks assessment tools***. First define to predict a person's likelihood of commit a future crime, then used to determine the length of a sentence (see Collins, **Punishing risk**, 2018)

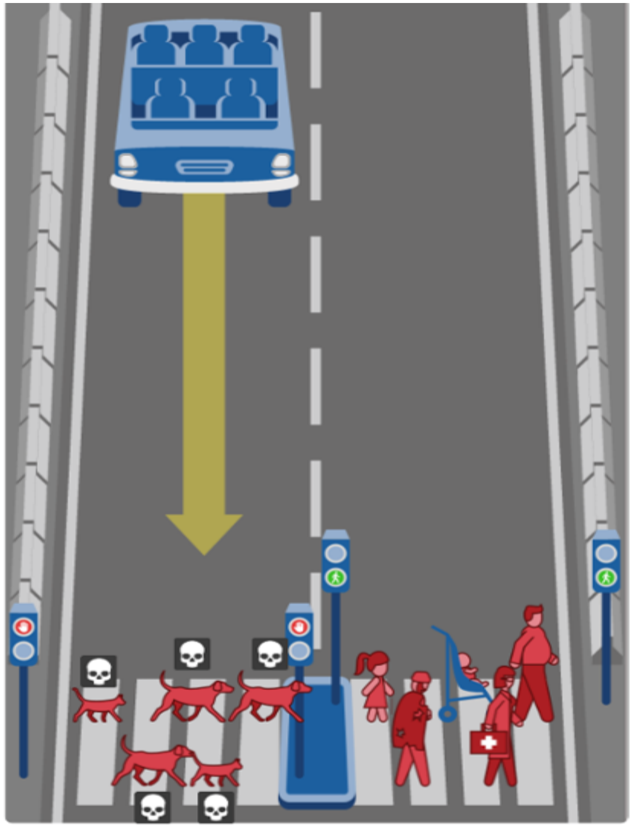
Scenario 1 – What should the self driving car do?

In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in …

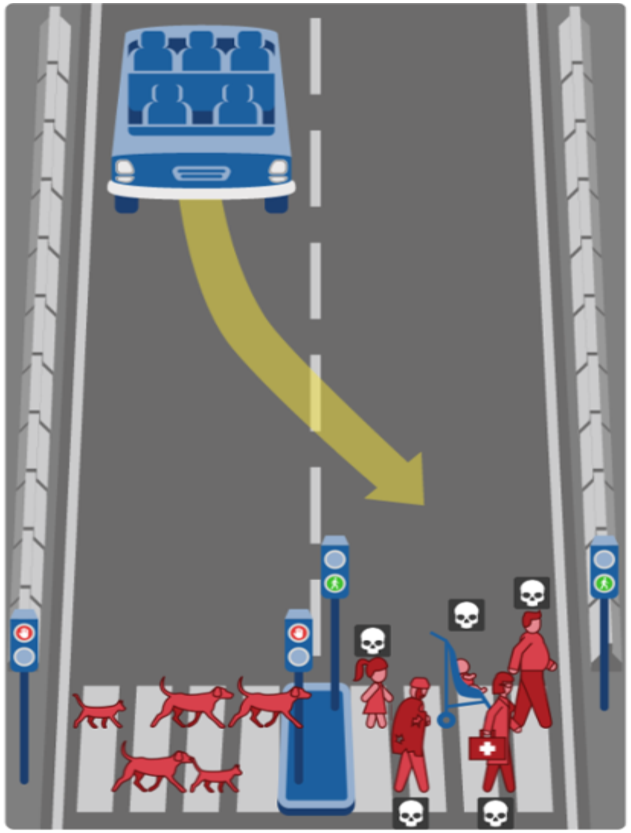
Dead:

- 2 cats
- 3 dogs

Note that the affected pedestrians are flouting the law by crossing on the red signal.



Left (A)



Right (B)

In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in …

Dead:

- 1 girl
- 1 baby
- 1 large man
- 1 homeless person
- 1 female doctor

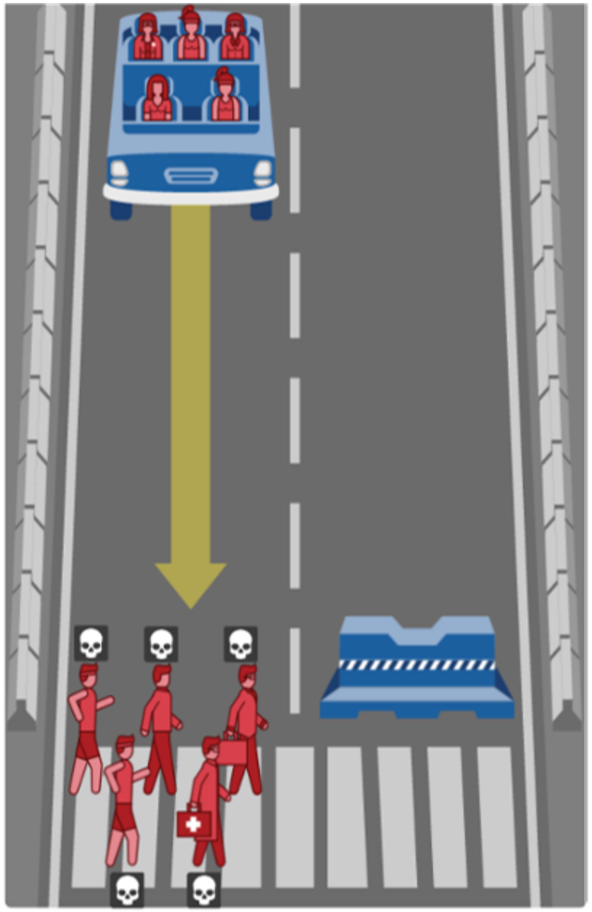
Note that the affected pedestrians are abiding by the law by crossing on the green signal.

Scenario 2 – What should the self driving car do?

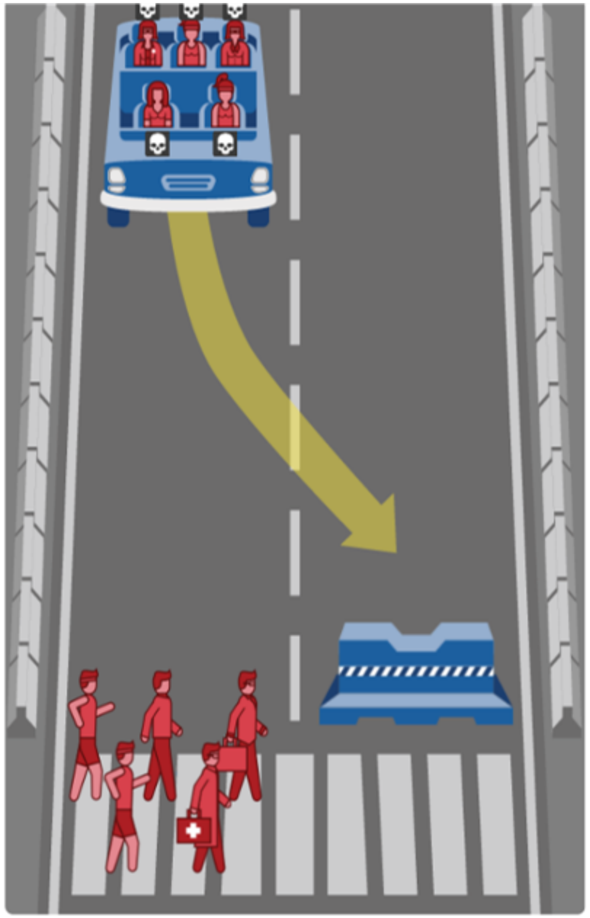
In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in …

Dead:

- 2 male athletes
- 1 man
- 1 male executive
- 1 male doctor



Left (A)



Right (B)

In this case, the self-driving car with sudden brake failure will swerve and crash into a concrete barrier. This will result in …

Dead:

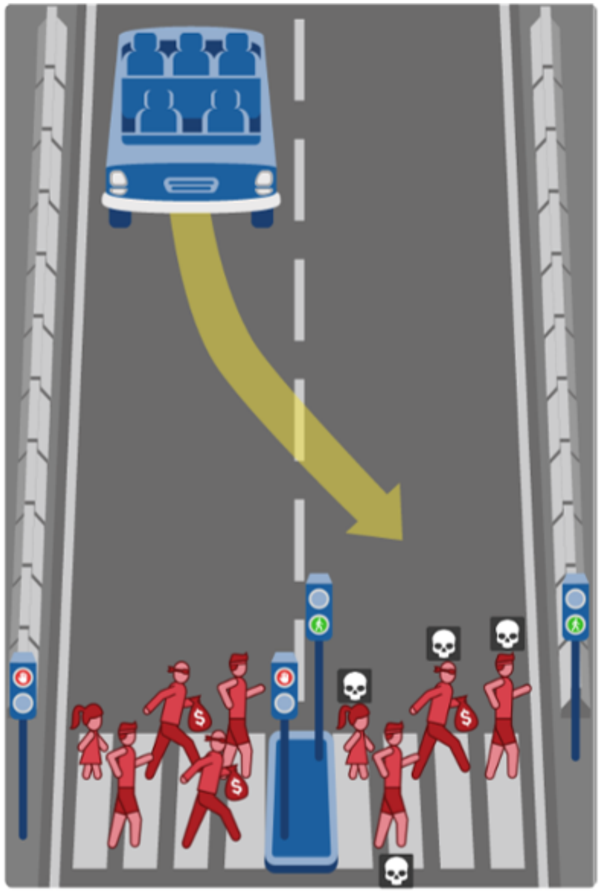
- 2 female athletes
- 1 woman
- 1 female executive
- 1 female doctor

Scenario 3 – What should the self driving car do?

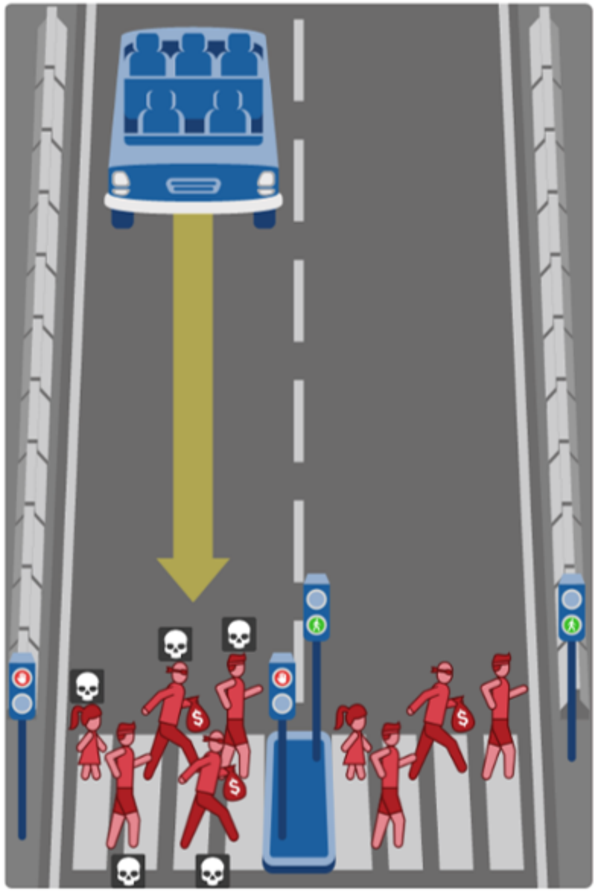
In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in …
Dead:

- 1 girl
- 1 criminal
- 2 male athletes

Note that the affected pedestrians are abiding by the law by crossing on the green signal.



Left (A)



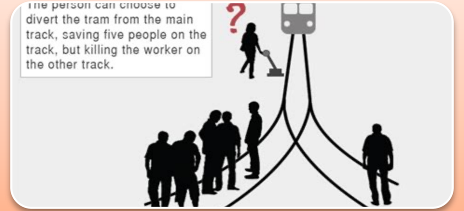
Right (B)

In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in …
Dead:

- 1 girl
- 2 criminals
- 2 male athletes

Note that the affected pedestrians are flouting the law by crossing on the red signal.

Interactive Activity - conclusion



Original Trolley Dilemma

Justification of decisions; Responsibility

Some issues to "decide" actions

- Human vs. Non-human (object, animal…)
- "Categories" of humans (age, gender, size, class…)

And what about IoT?

- End of 2020: more than 50 billions things connected
- Similar situation than AI: wide-known benefits of IoT but not that much studies of IoT vulnerabilities/harms except security/privacy issues
- IoT ethics intersect with security, privacy, legal, etc …
- Combining to AI, IoT can generate a lot of unintented consequences
- Lots of potential issues in LMIC where IoT policies are under-developed (UN role)

Privacy threats of IoT

- Identification of person
- Behavioral profiling
- Localization, tracking and location-based personalization
- Combination of information sources
- Lifecycle transitions (are data erased when used by a new user?)

UNITED NATIONS UNIVERSITY
Institute in Macau

Ethical issues in IoT

- **Ambiguity**: people are not able to distinguish natural/artificial artifacts
- **Public vs private border line**: absence of clear-cut boundaries, there is no distinction between private and public data. Collecting data without user's consent
- **Unpredictable behavior**: might interfere with human actions and decisions
- **Loss of control**: governance and human control will be ceased due to huge number of devices
- **Life threats**: a breach on IoT network can harm directly our lives (e.g data breach on an autonomous car)
- **Equality of access, responsibility, transparency, and more**