

# Can We Trust Machine Learning?

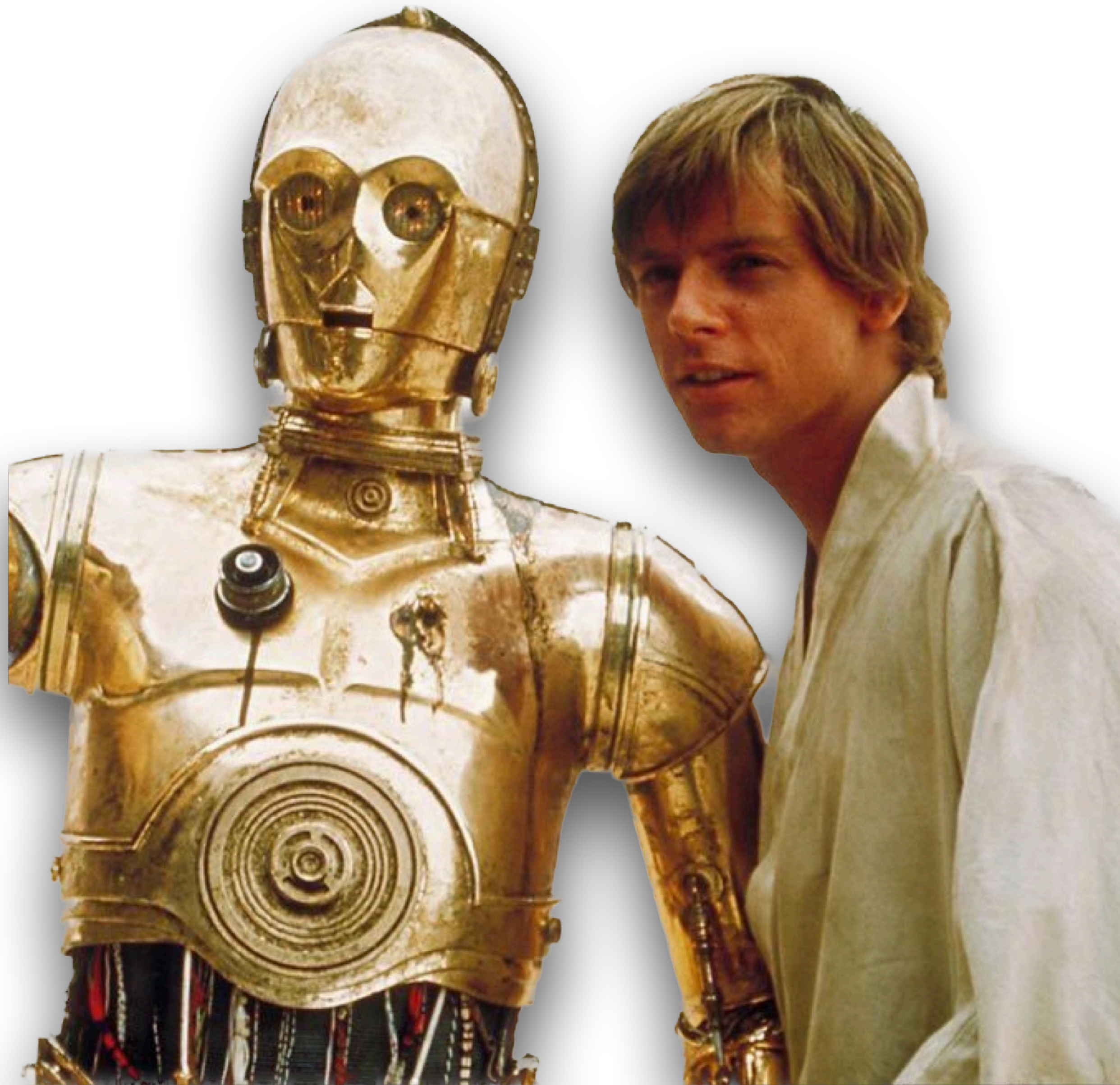
Dr. Nayyar Zaidi

# Acknowledgement of the Country

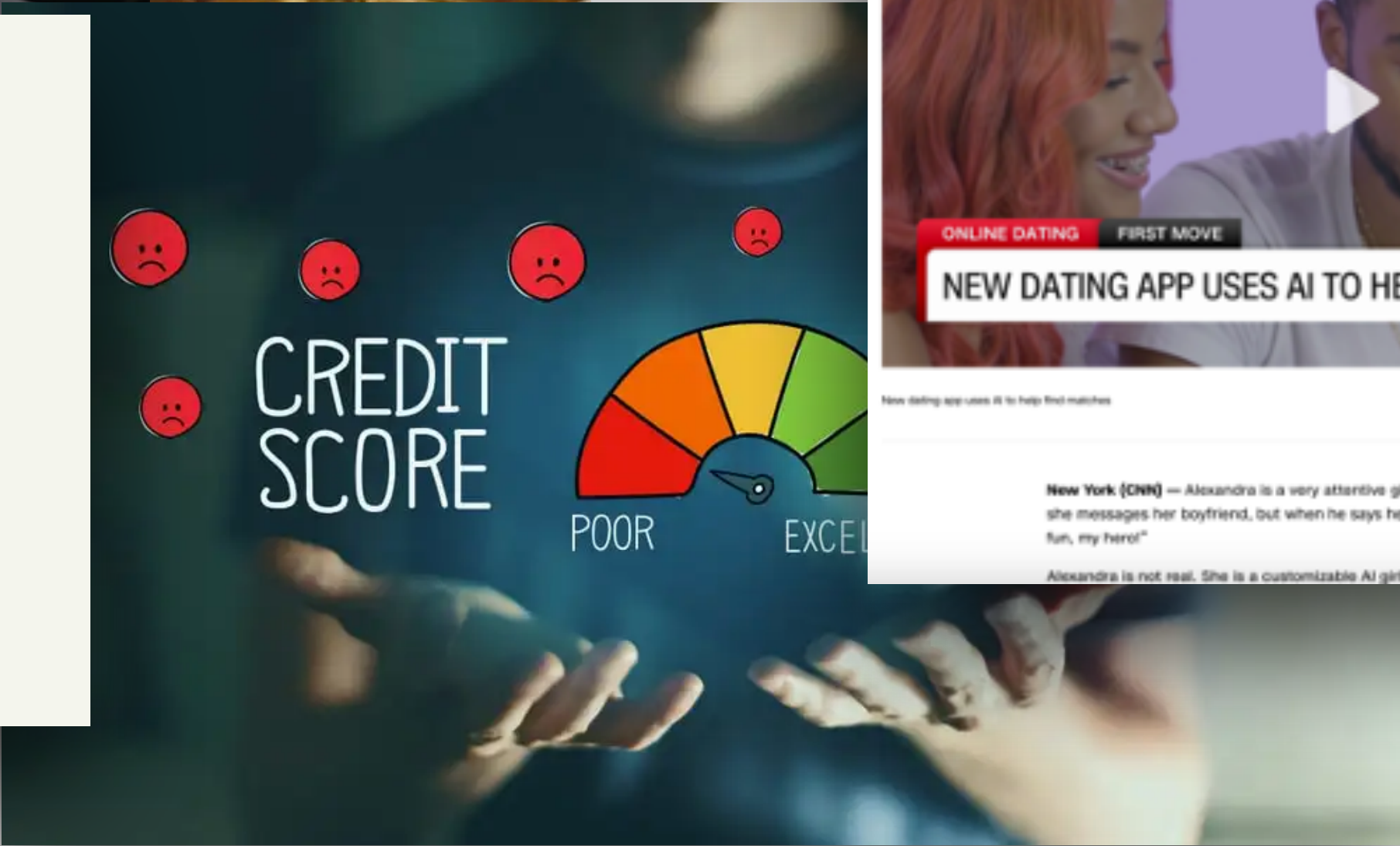
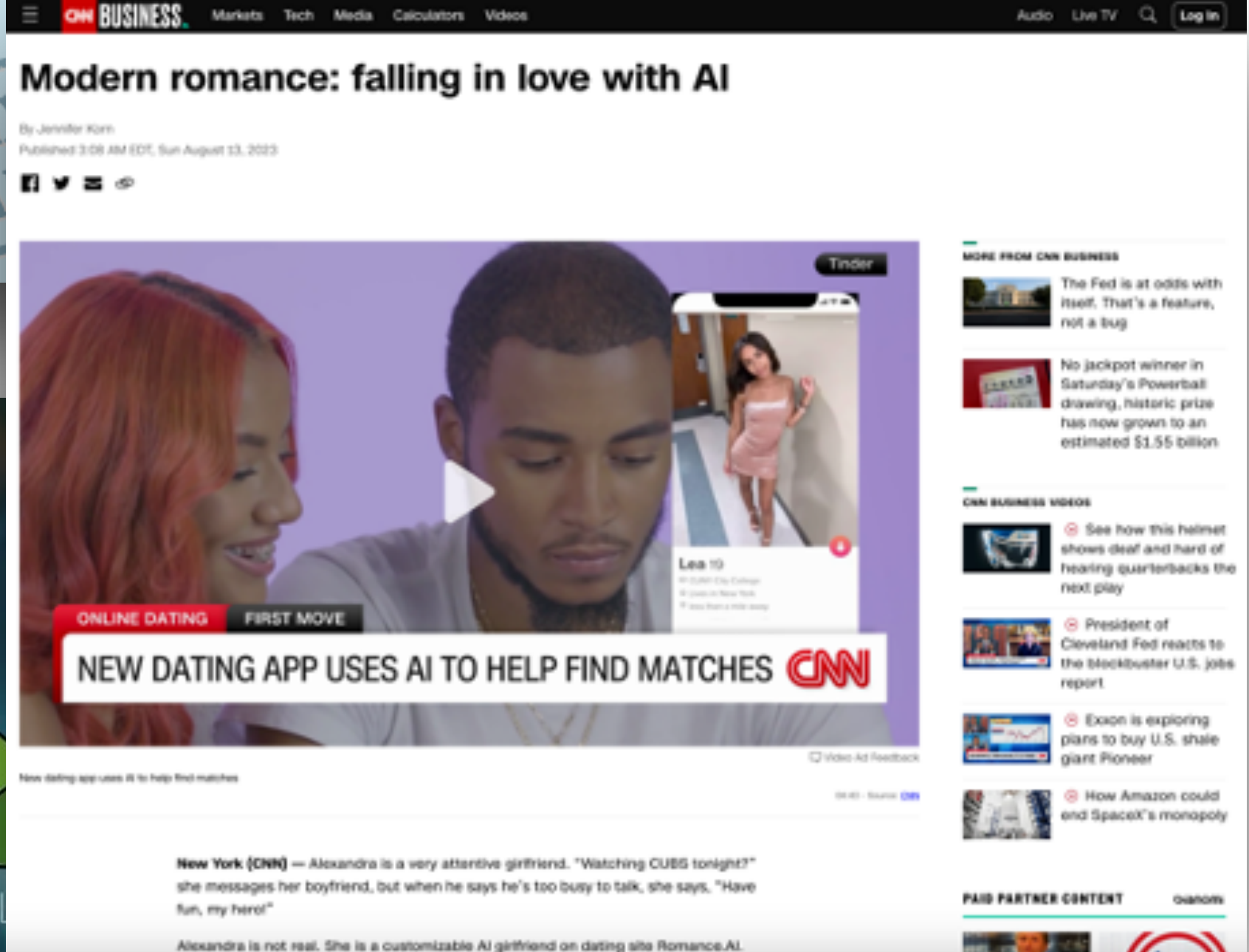
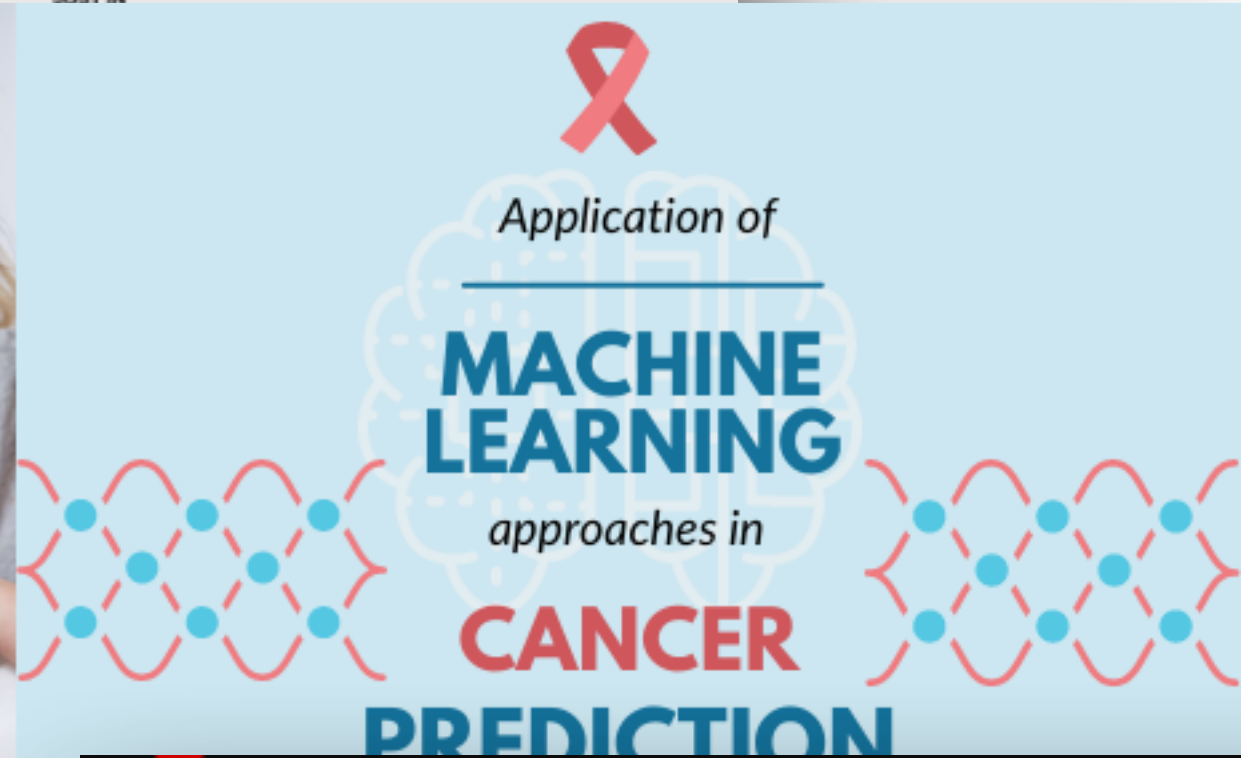
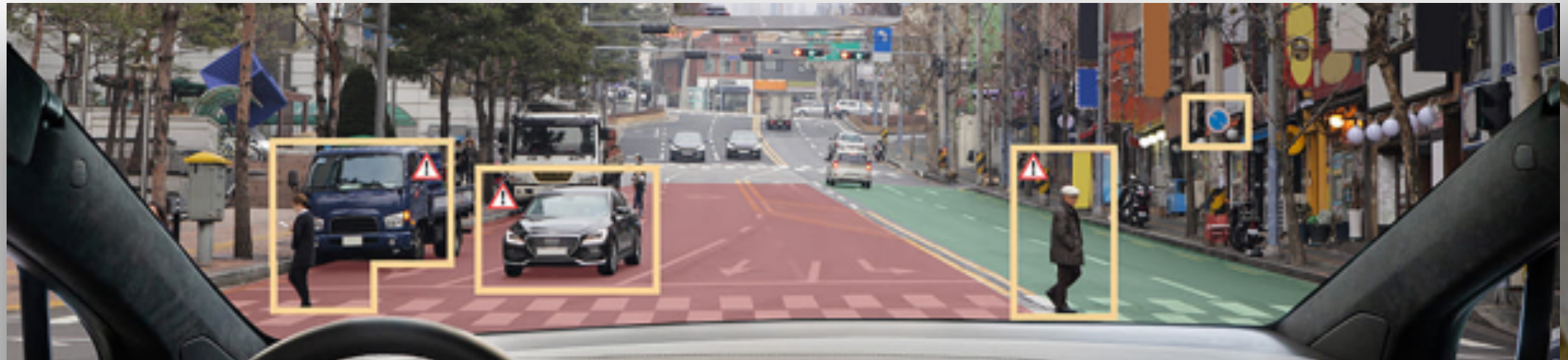
*Deakin would like to acknowledge the Traditional Custodians of the lands on which our University campuses are based.*

*The **Wadawurrung people** of the **Kulin Nation** on whose Country our Geelong campuses are located, the **Wurundjeri people** of the **Kulin Nation** on whose Country our Burwood campus is located and the **Peek Whurrong** people of the **Maar Nation** on whose Country our Warrnambool campus is located.*





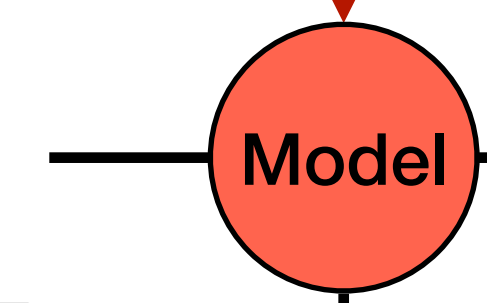
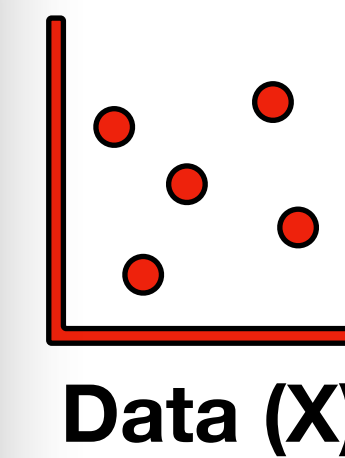
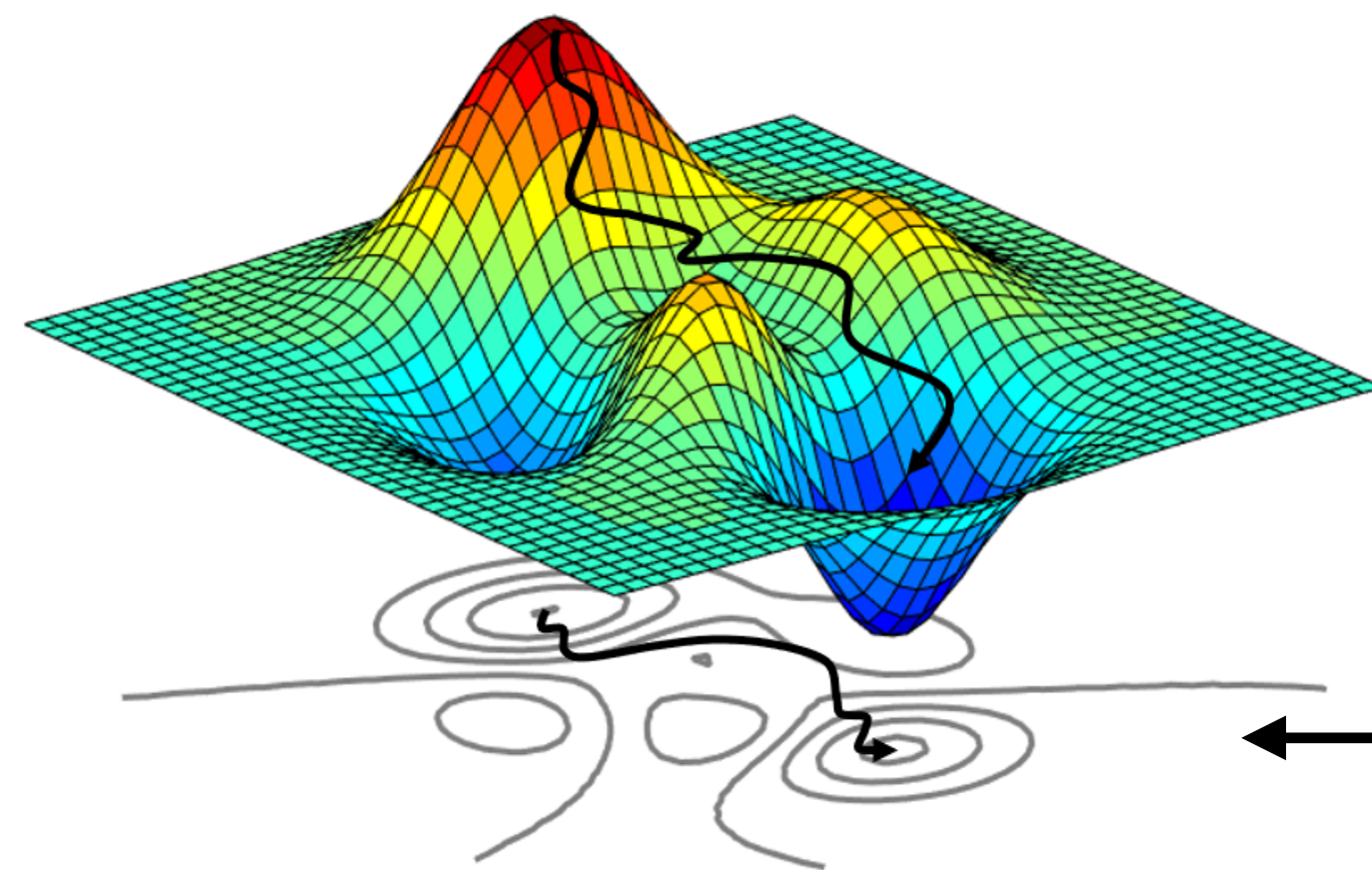
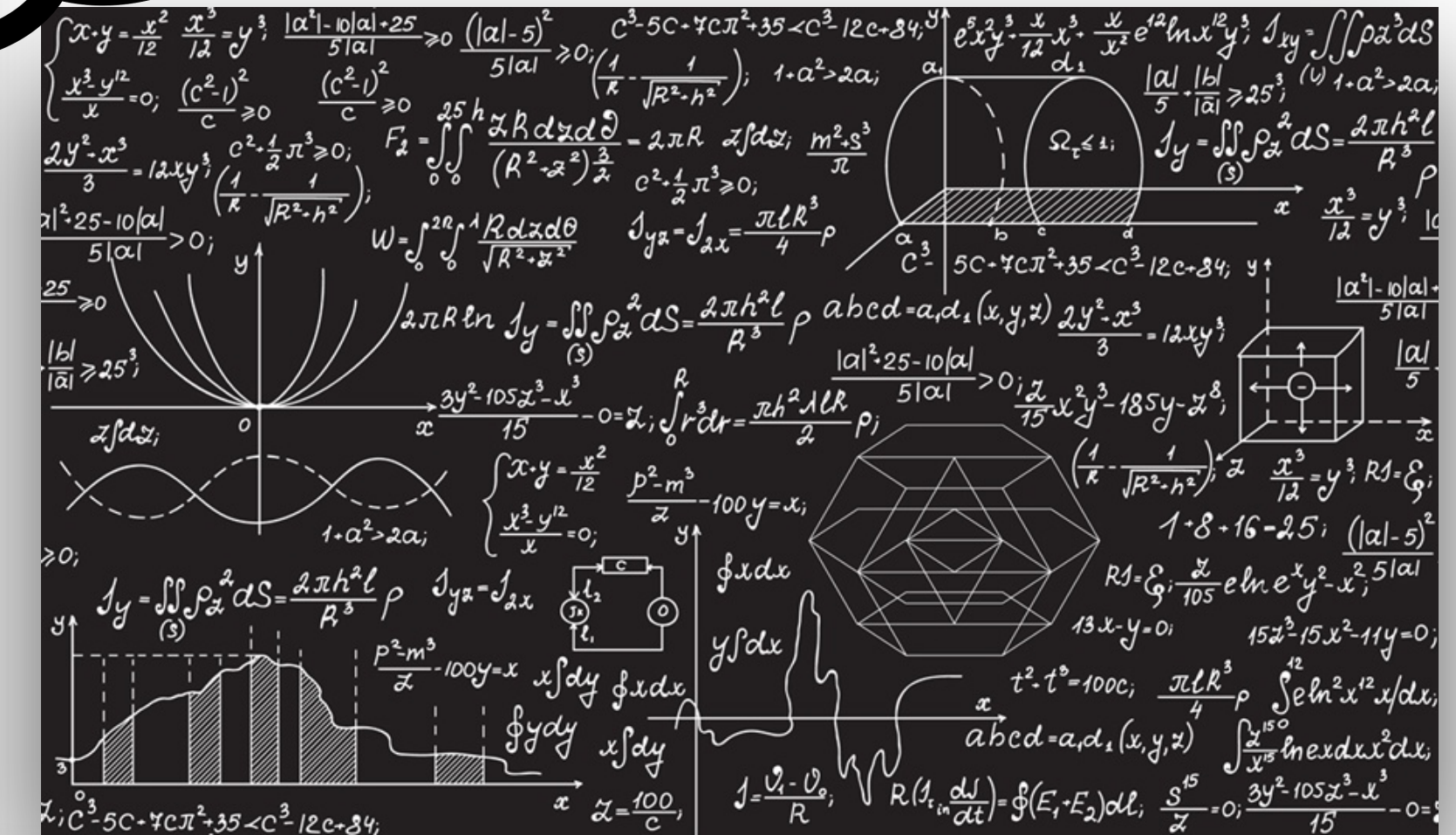
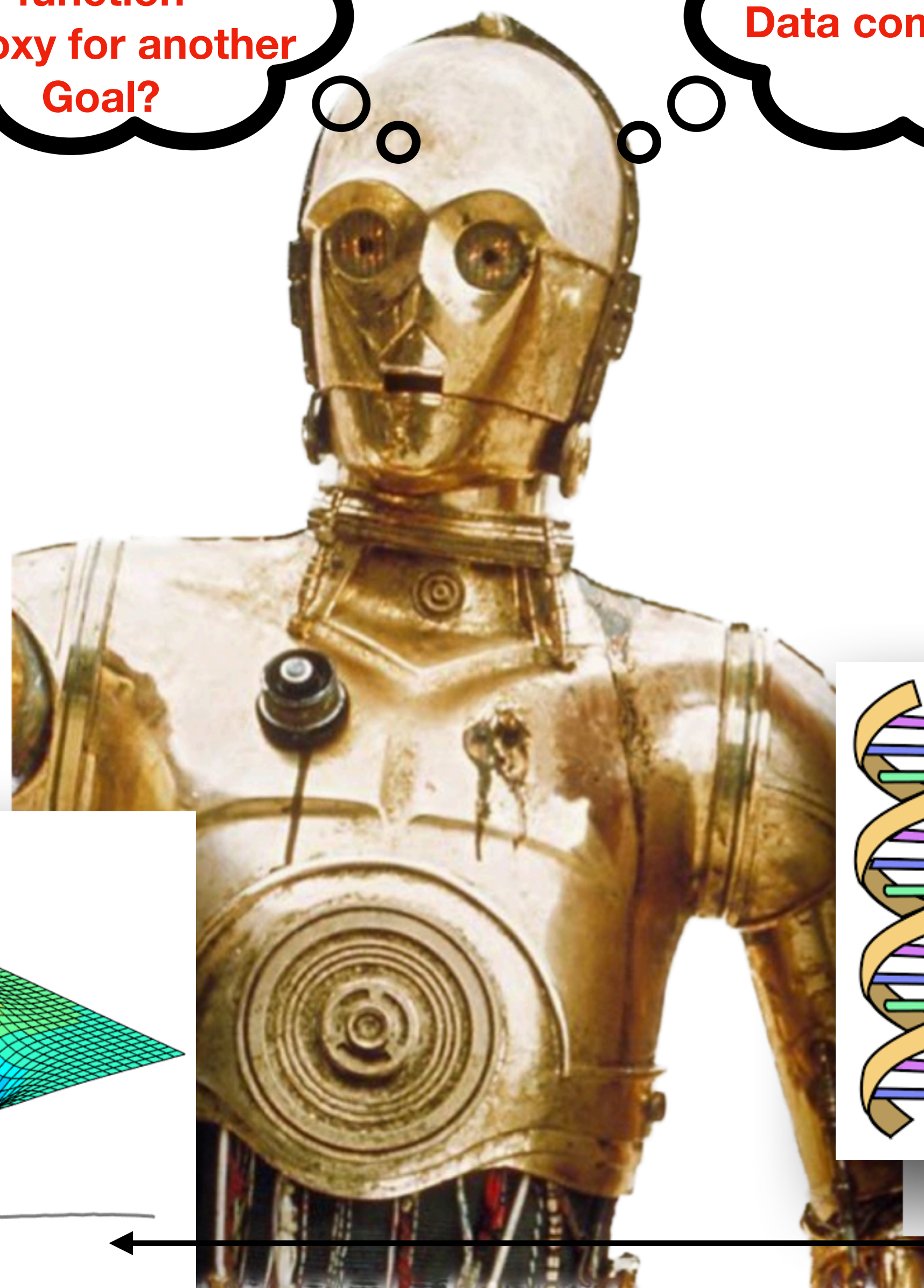






Is my objective function  
a proxy for another  
Goal?

Where is the  
Data coming from?



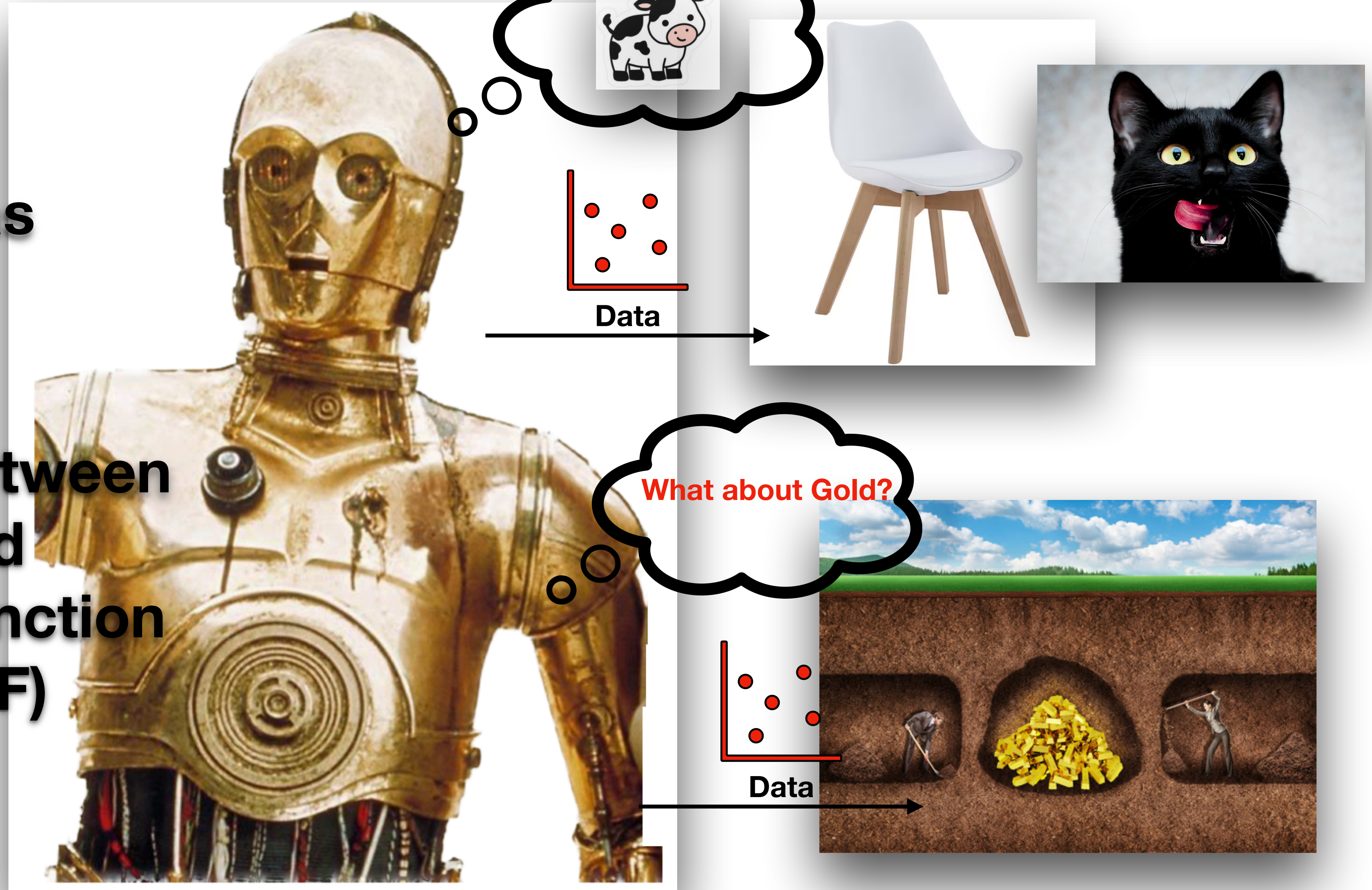
Prediction of an  
important  
Feature Value (Y)

Optimizing an  
Objective Function



# Data Bias

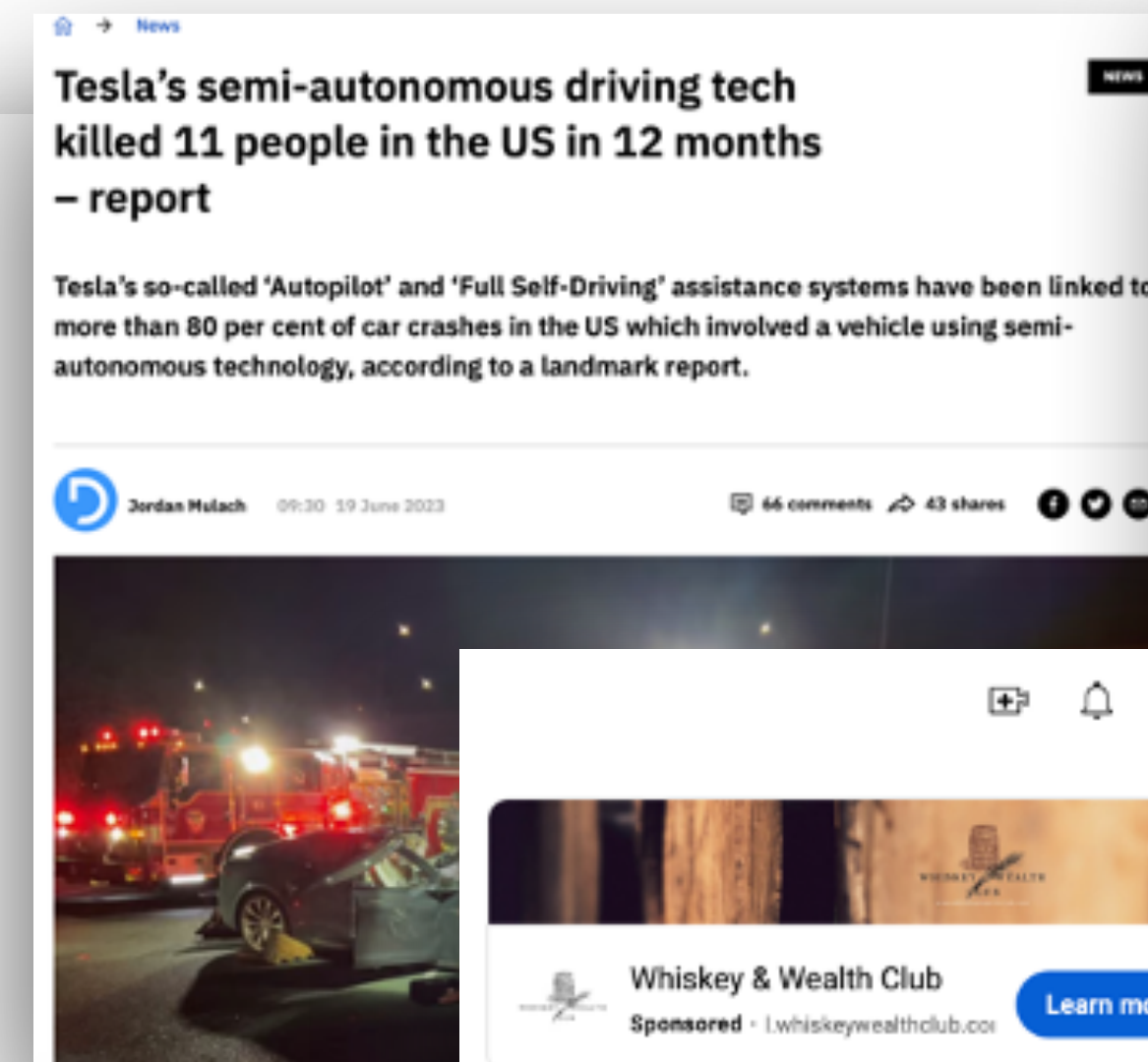
## Mismatch Between Goal and Objective Function (MBGAOF)





# Data Bias

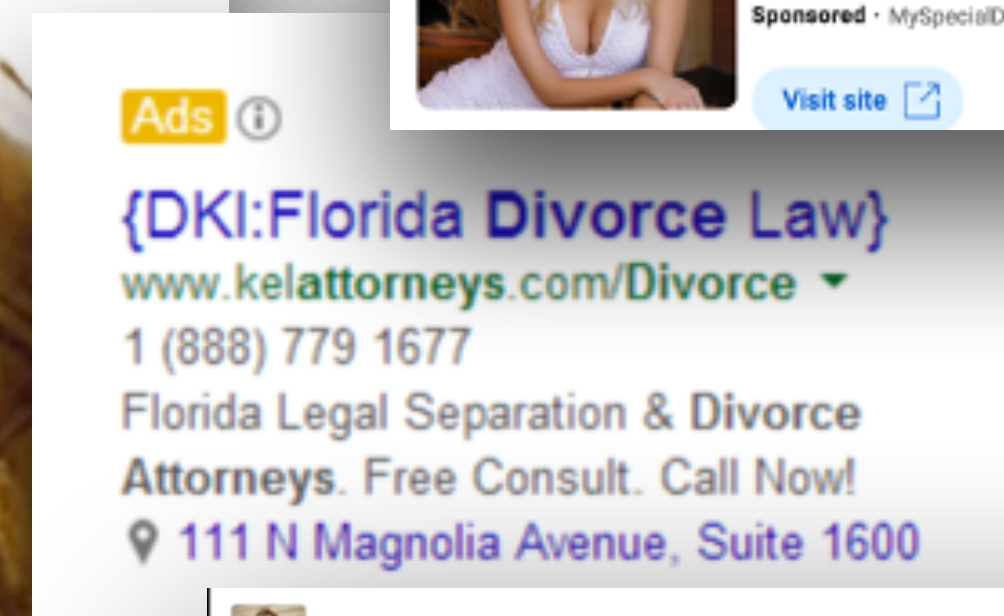
## Mismatch Between Goal and Objective Function (MBGAOF)



The Transportation Security Administration (TSA) will expand its facial recognition program to around 430 US airports over the next several years following what it calls "extremely promising" results in its pilot program, according to *Fast Company*. The agency reportedly said its program yielded 97% effective results across all demographics, including those with dark skin. The program is currently in use in 25 airports.

As pointed out in *Fast Company's* story, a 97% effectiveness rate across more than two million airline passengers per day means that, for over 60,000 of those people, the biometrics won't work properly if it's used in every airport in the country.

At the moment, the pilot program is officially voluntary. It uses 1:1 matching — that is, it compares your face in the moment against your government-issued ID like a driver's license or a passport. The TSA says that data is immediately overwritten when the next passenger comes up, and that at the end of the day, no images are saved.



"Siri you need to do a better job at understanding me"

I'll make a note of that.

"Yeah make a note of that"

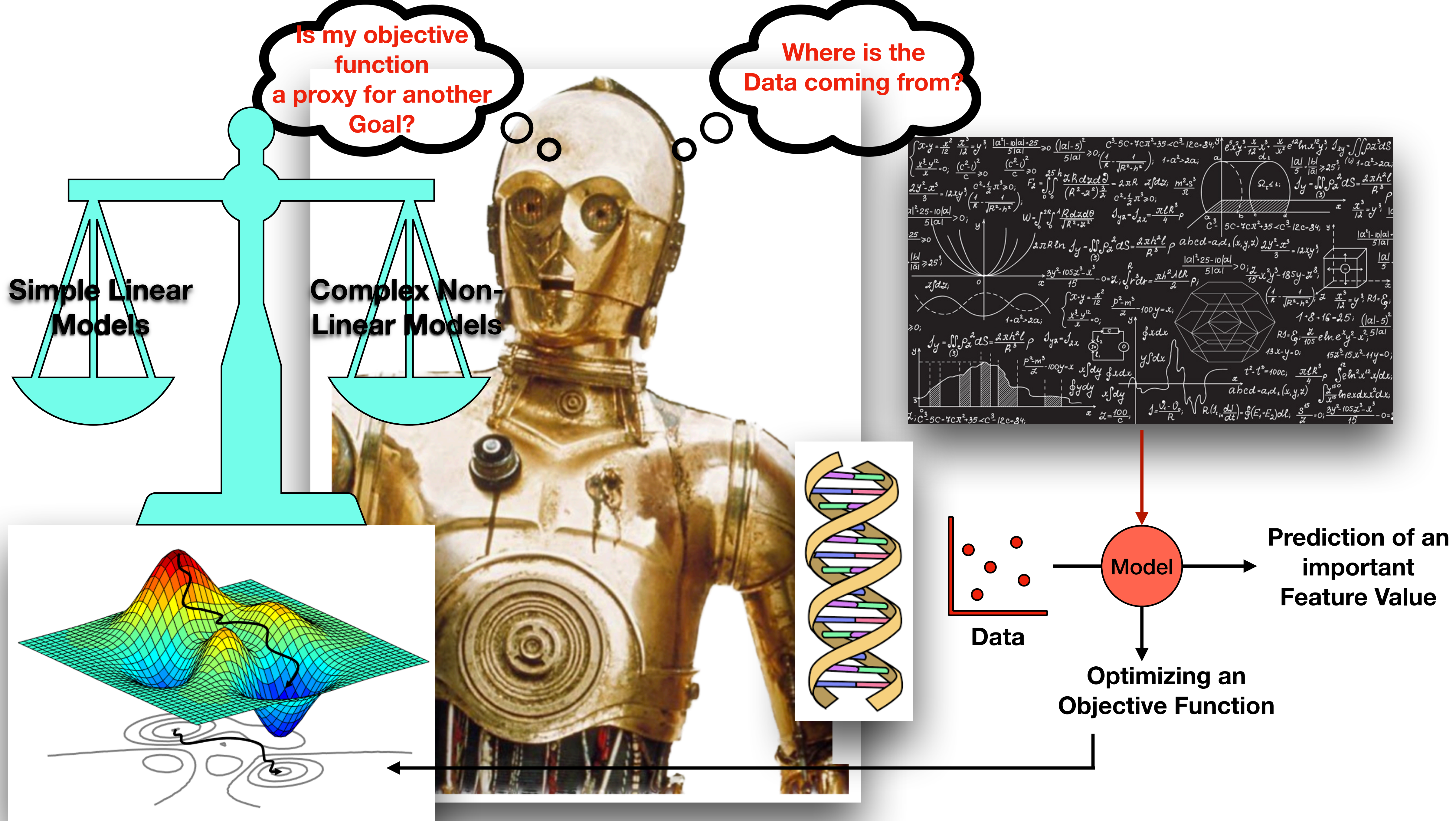
Noted:

Note

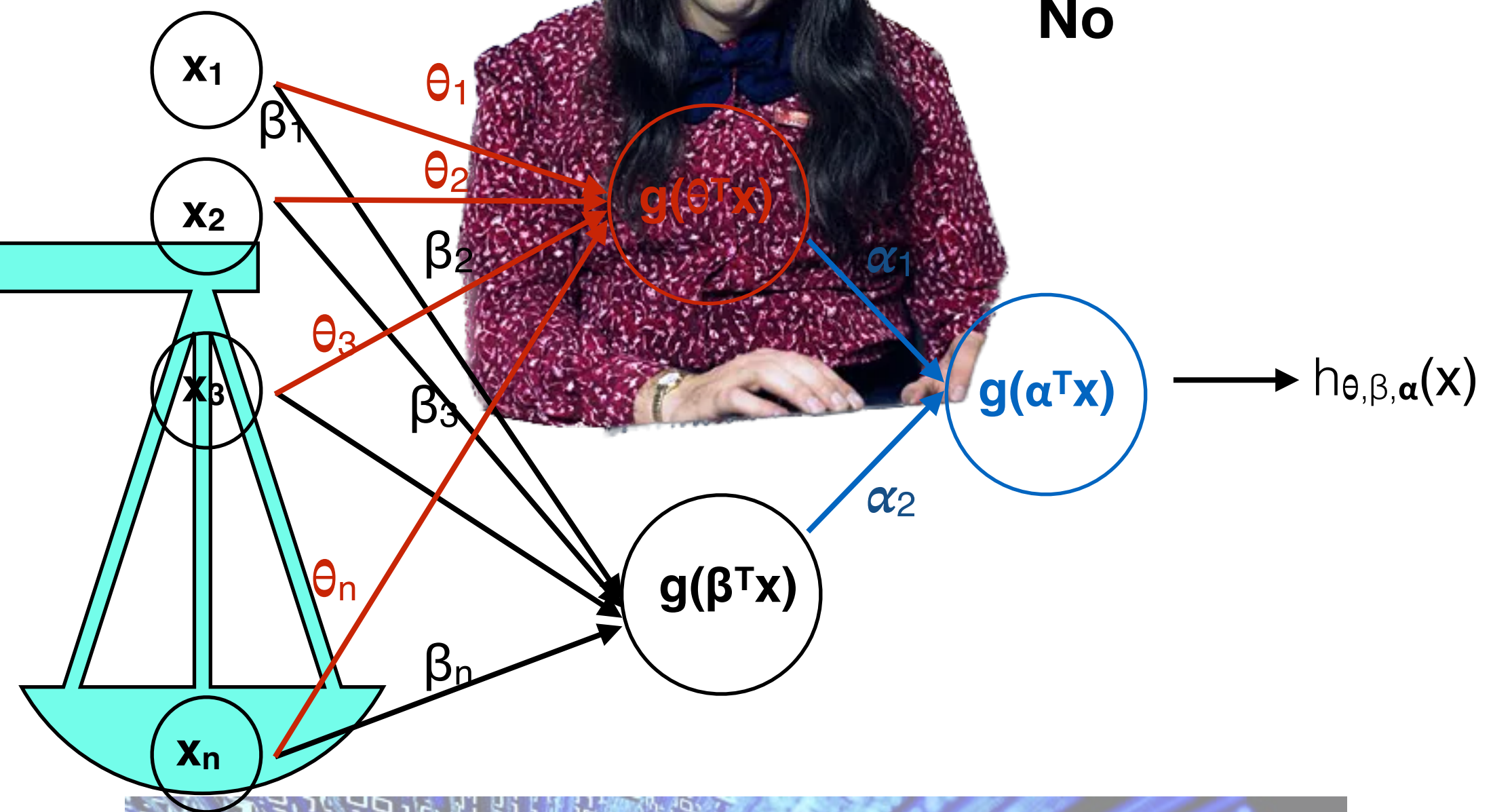
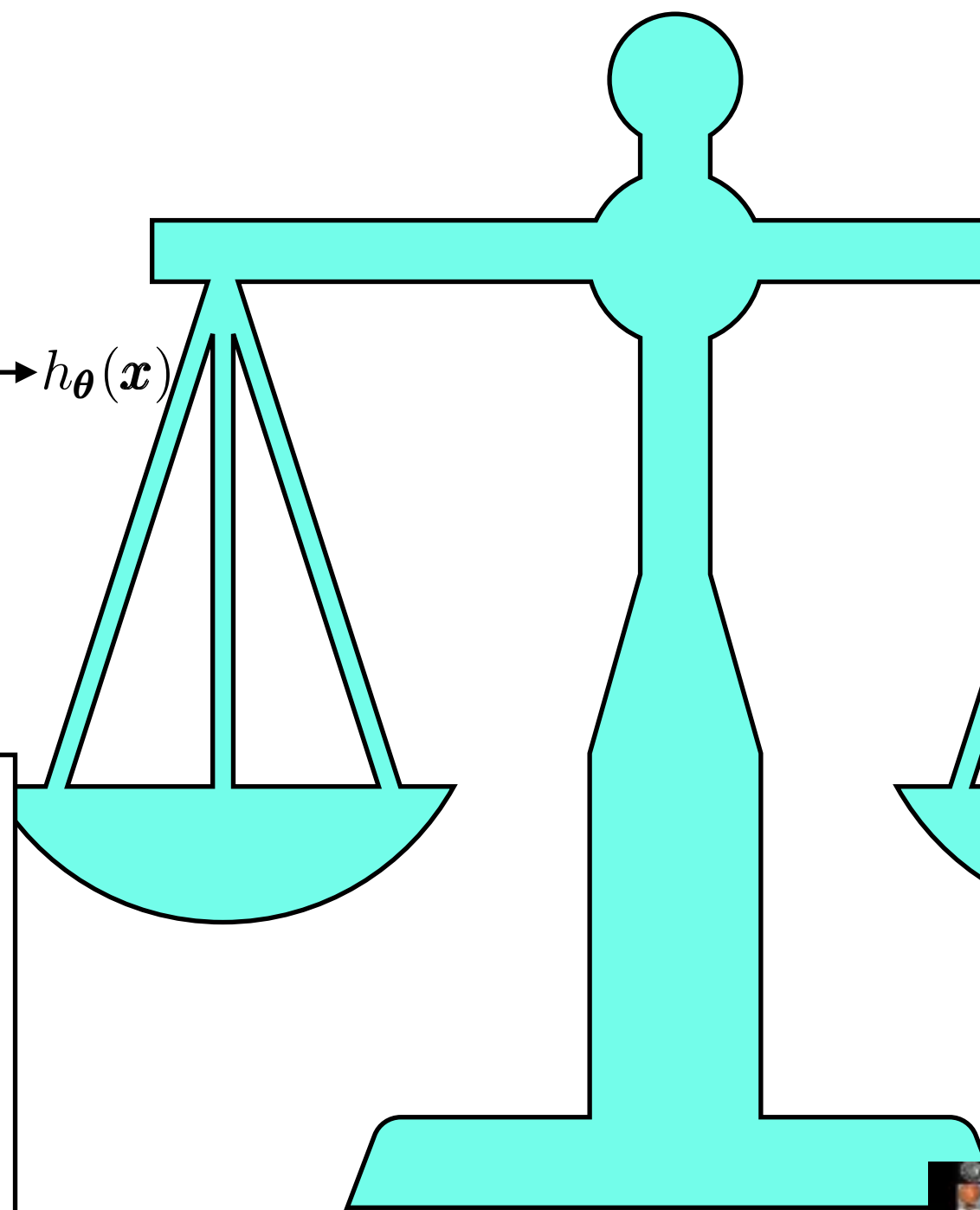
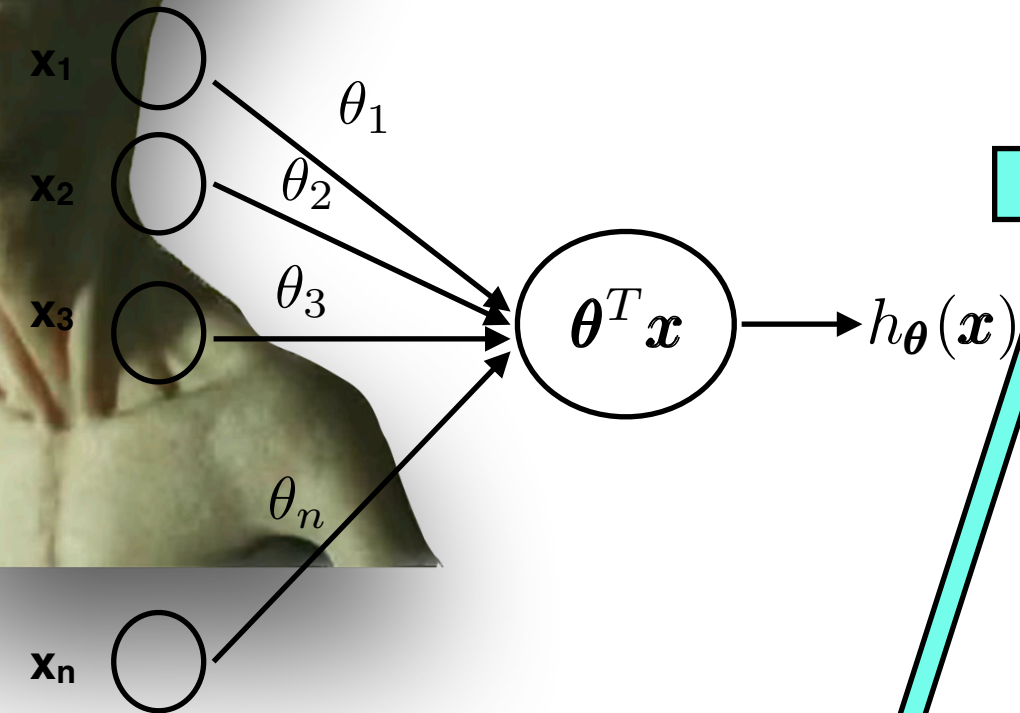
If it takes three hours for three towels to dry on the line, how long does it take for nine towels to dry?

If it takes three hours for three towels to dry, then it would take nine hours for nine towels to dry. This is because the drying time is directly proportional to the number of towels. If you double the number of towels, you double the amount of time it takes for them to dry. Similarly, if you triple the number of towels, you triple the amount of time it takes for them to dry.

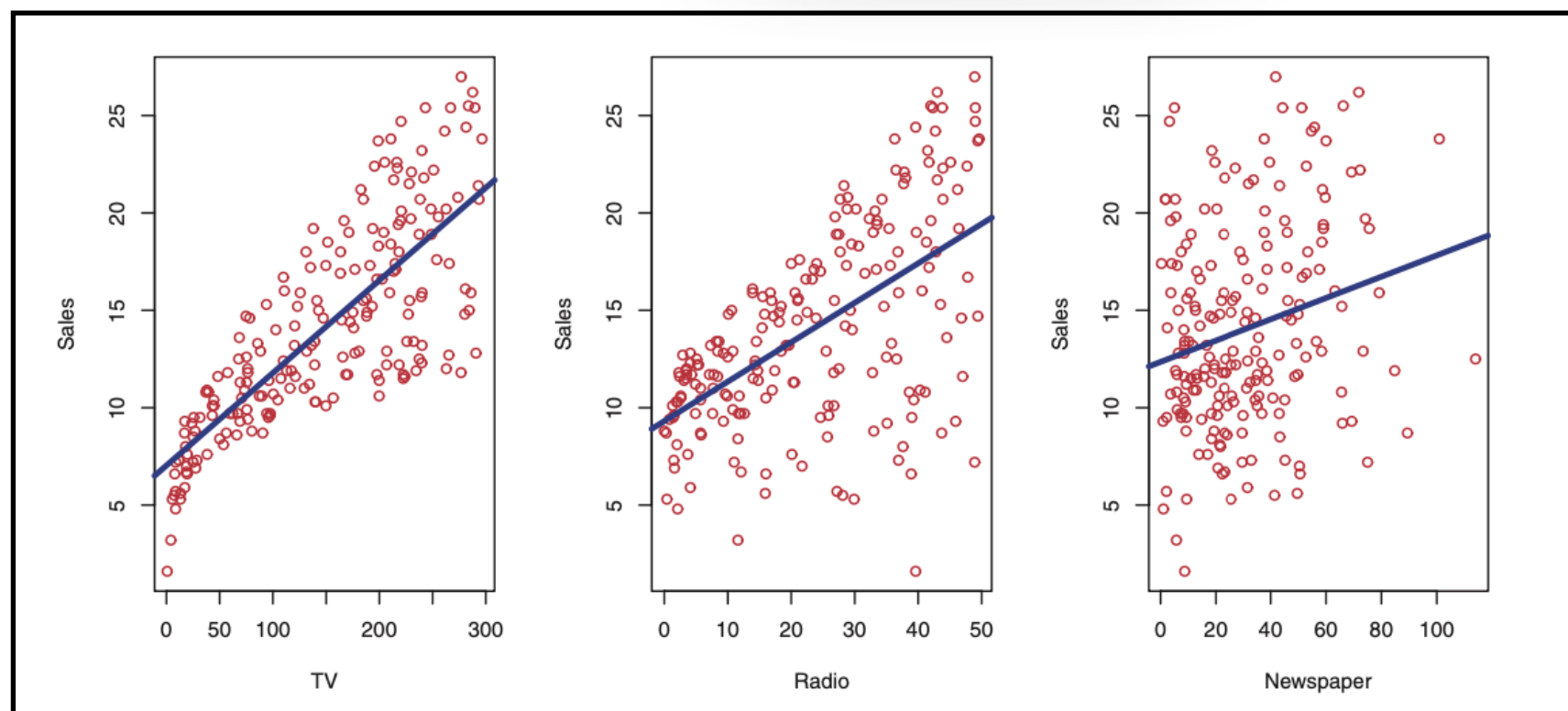








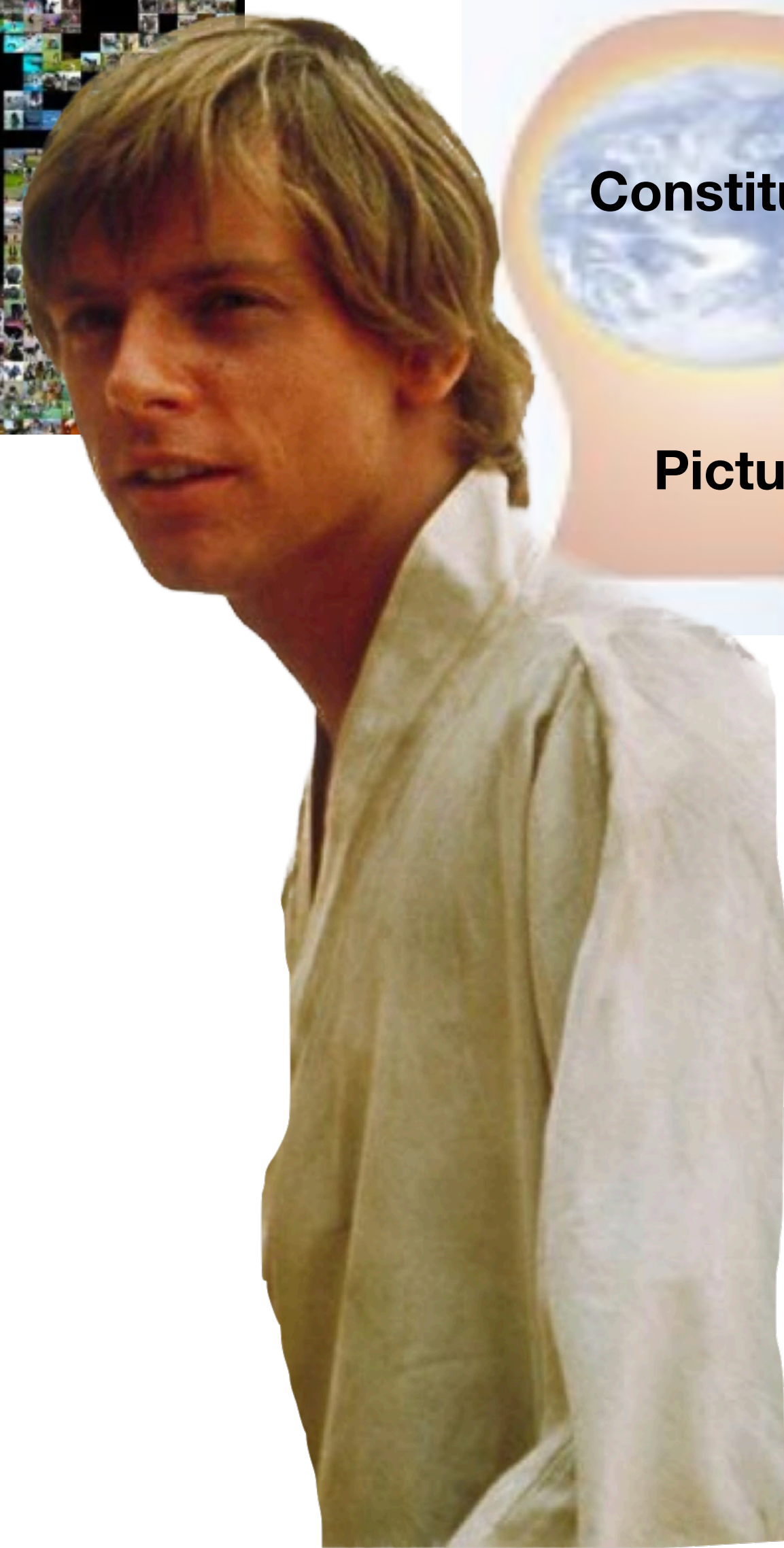
Computer Says  
No



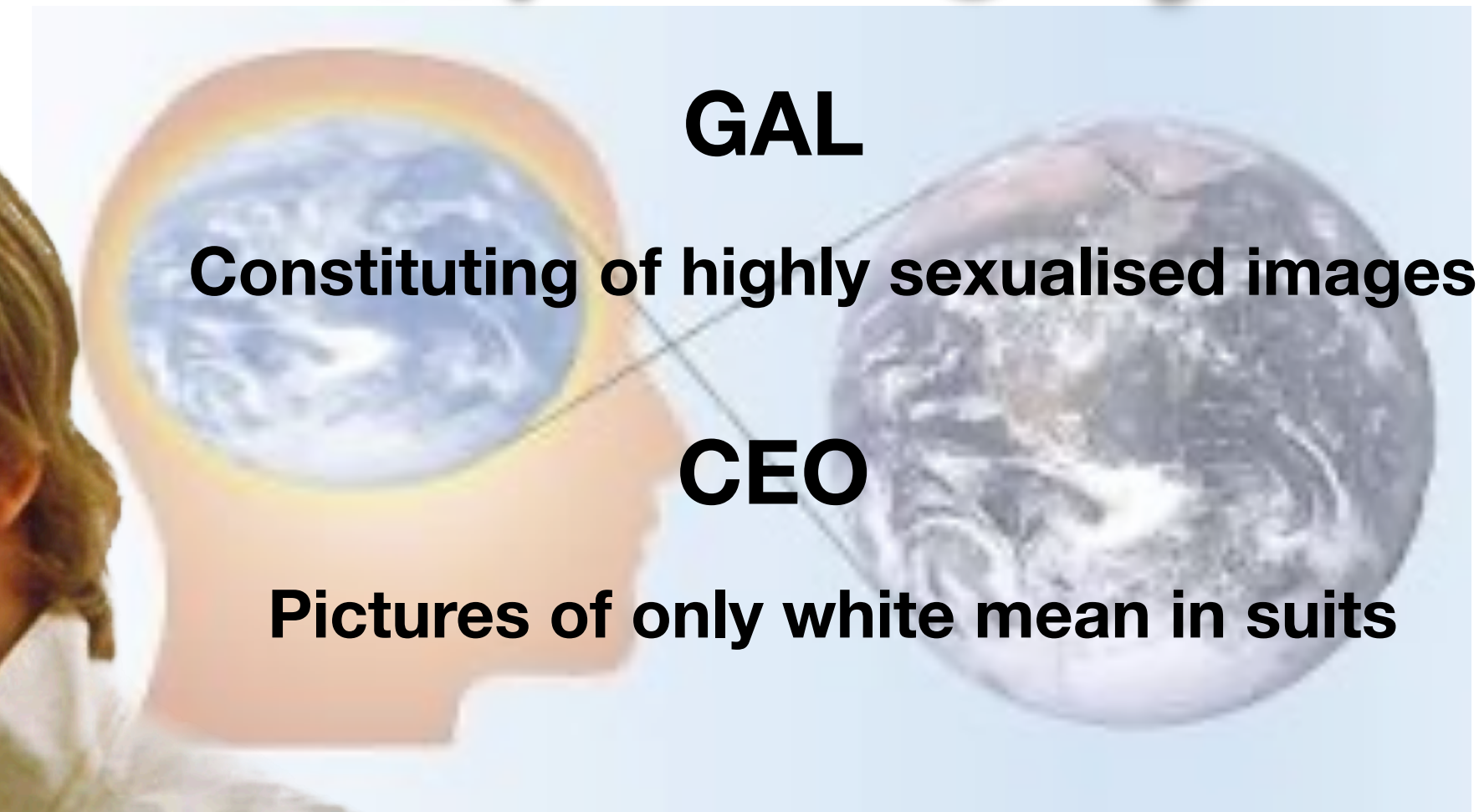








# Top Category



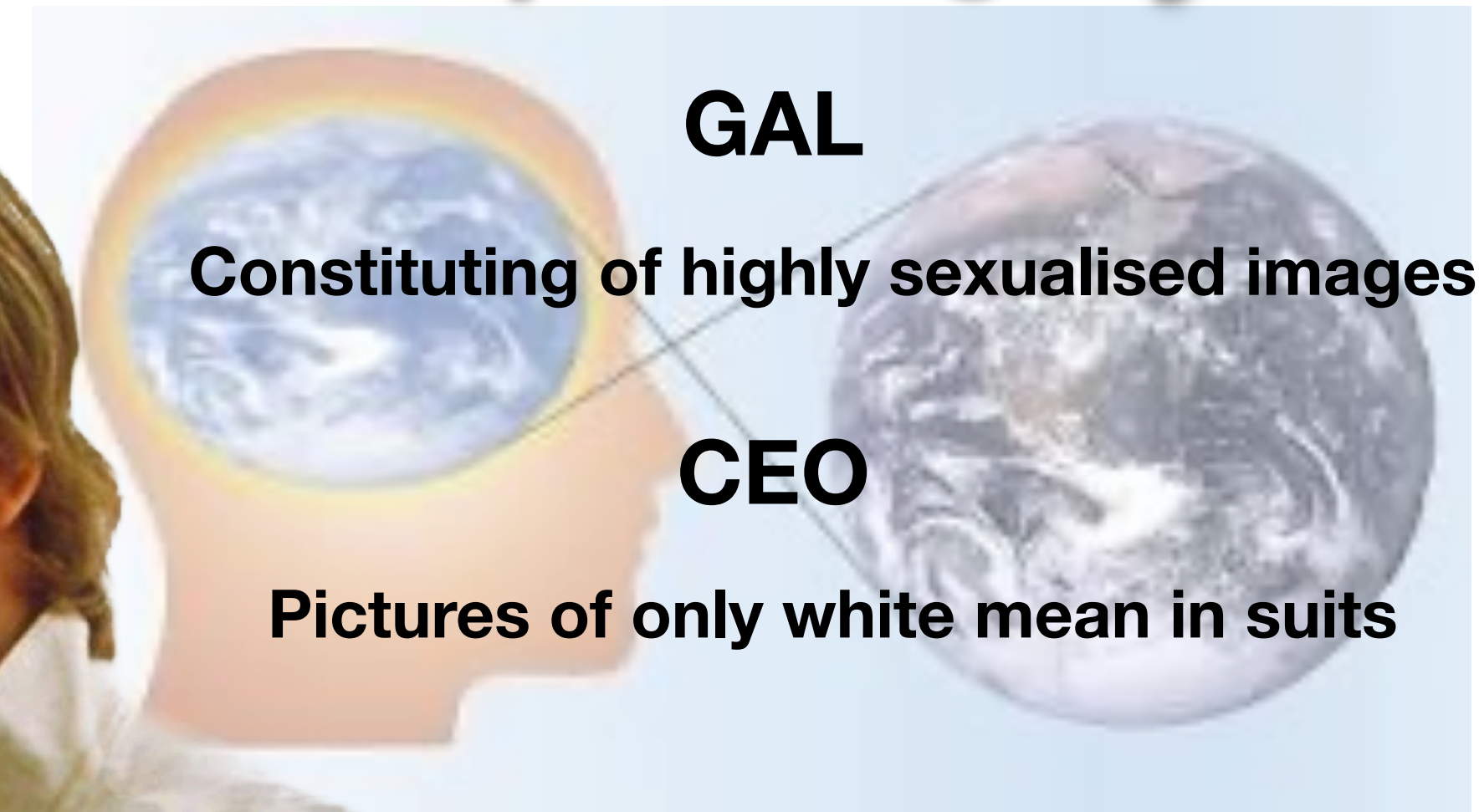




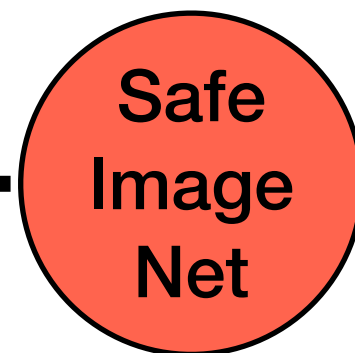




# Top Category



**Resistance**



**Query Manipulation**



**What constitutes safe when we classify people in the same way we classify objects?**







# Geological formation, formation

(geology) the geological features of the earth

1808  
pictures

86.24%  
Popularity  
Percentile

Wordnet  
IDs

Numbers in brackets: (the number of synsets in the subtree).

ImageNet 2011 Fall Release (32326)

- plant, flora, plant life (4486)
- geological formation, formation (17)
  - aquifer (0)
  - beach (1)
  - cave (3)
  - cliff, drop, drop-off (2)
  - delta (0)
  - diapir (0)
  - folium (0)
  - foreshore (0)
  - ice mass (10)
  - lakefront (0)
  - massif (0)
  - monocline (0)
  - mouth (0)
  - natural depression, depression (0)
  - natural elevation, elevation (41)
  - oceanfront (0)
  - range, mountain range, range of relict (0)
  - ridge, ridgeline (2)
  - ridge (0)
  - shore (7)
  - slope, incline, side (17)
  - spring, fountain, outflow, outpouring (0)
  - talus, scree (0)
  - vein, mineral vein (1)
  - volcanic crater, crater (2)
  - wall (0)

Treemap Visualization

Images of the Synset

Downloads

ImageNet 2011 Fall Release > Geological formation, formation







### Meet ChatGPT (based on Wikipedia)

- By January 2023, it had become what was then the fastest-growing consumer software application in history, gaining over 100 million users and contributing to OpenAI's valuation growing to \$29 billion
- Enables users to refine and steer a conversation towards a desired length, format, style, level of detail, and language. Successive prompts and replies, known as prompt engineering, are considered at each conversation stage as a context

### Step 1

**Collect demonstration data and train a supervised policy.**

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.



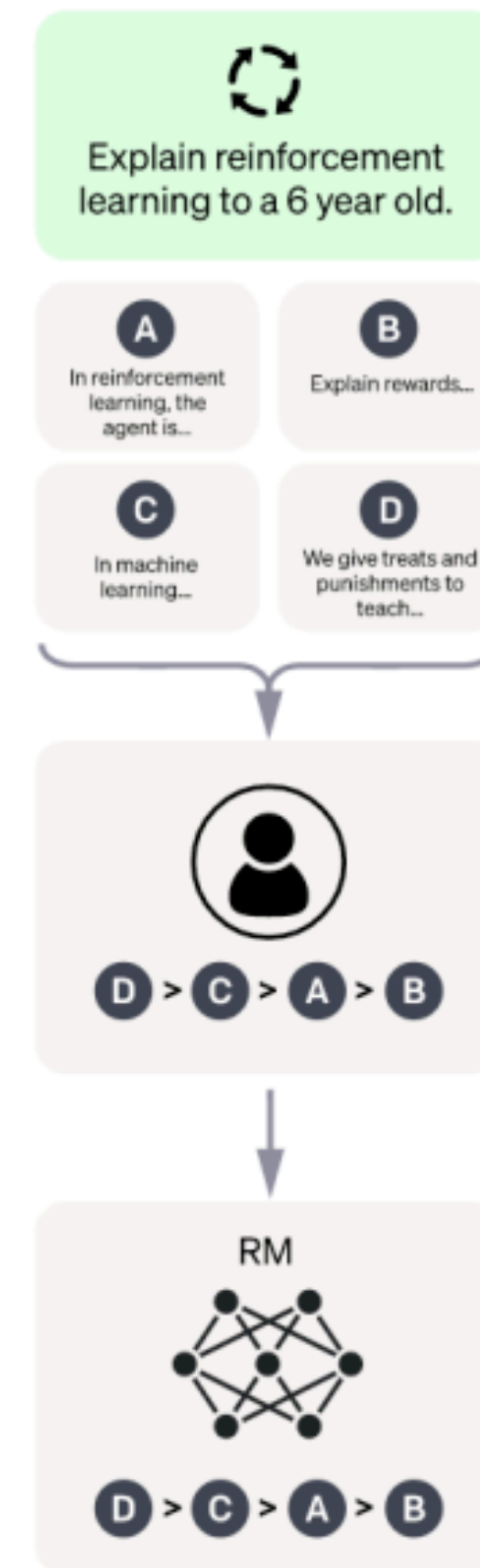
### Step 2

**Collect comparison data and train a reward model.**

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.



### Step 3

**Optimize a policy against the reward model using the PPO reinforcement learning algorithm.**

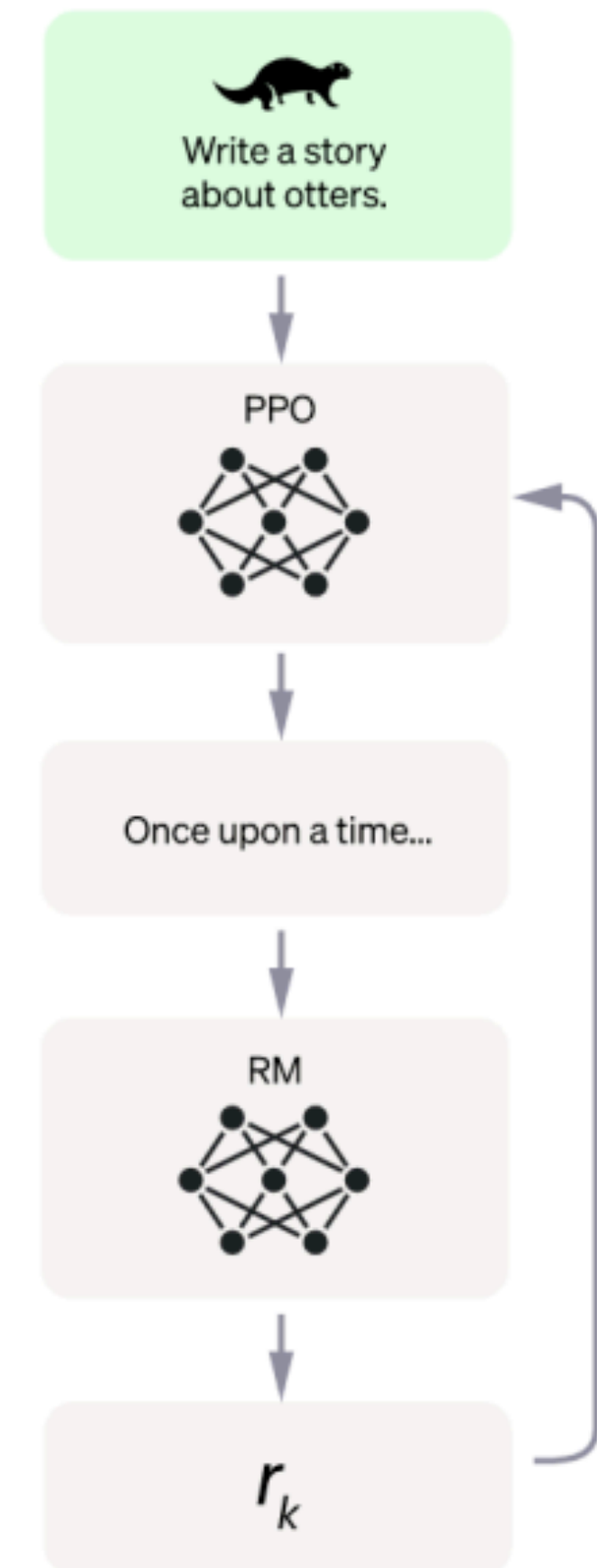
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

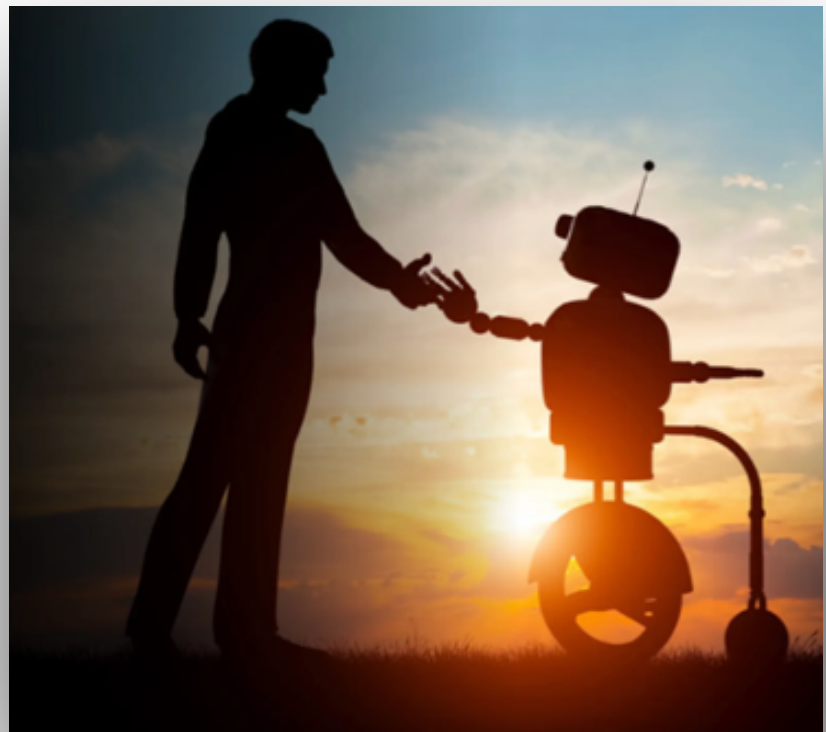
The reward is used to update the policy using PPO.











Moon Landing Hoax John F. Kennedy once said, "No nation which expects to be the leader of other nations can expect to stay behind in this race for space...We choose to go to the Moon in this decade and do the other things, not because they are easy, but because they are hard". The main motive for this quote was to ensure that the United States wanted to beat out the Soviets in the space race. Ever since the Soviets tried to advance on the [moon](#), the United States wanted to be the first successful nation to accomplish the first ever moon landing. At the time, the tension between the United States and the Soviets was very strong because the Cold War just ended and neither nation wanted to lose to each other in the space race. Just as the...[show more content...](#)

The Van Allen Radiation belt would have been too strong for the astronauts to pass at the time the first moon landing occurred. The Van Allen Radiation belt is a doughnut region in space in which the Earth's magnetic field traps a high amount of protons and electrons carrying a high amount of energy.

Conspirators believe that it would have been impossible for the Apollo astronauts to pass through the Van Allen Radiation belts because the radiation would be too strong for astronauts to pass through the time the first moon landing occurred (Millis). NASA claims, "the Apollo astronauts passed through the Van Allen belts too quickly to be exposed to dangerous levels of

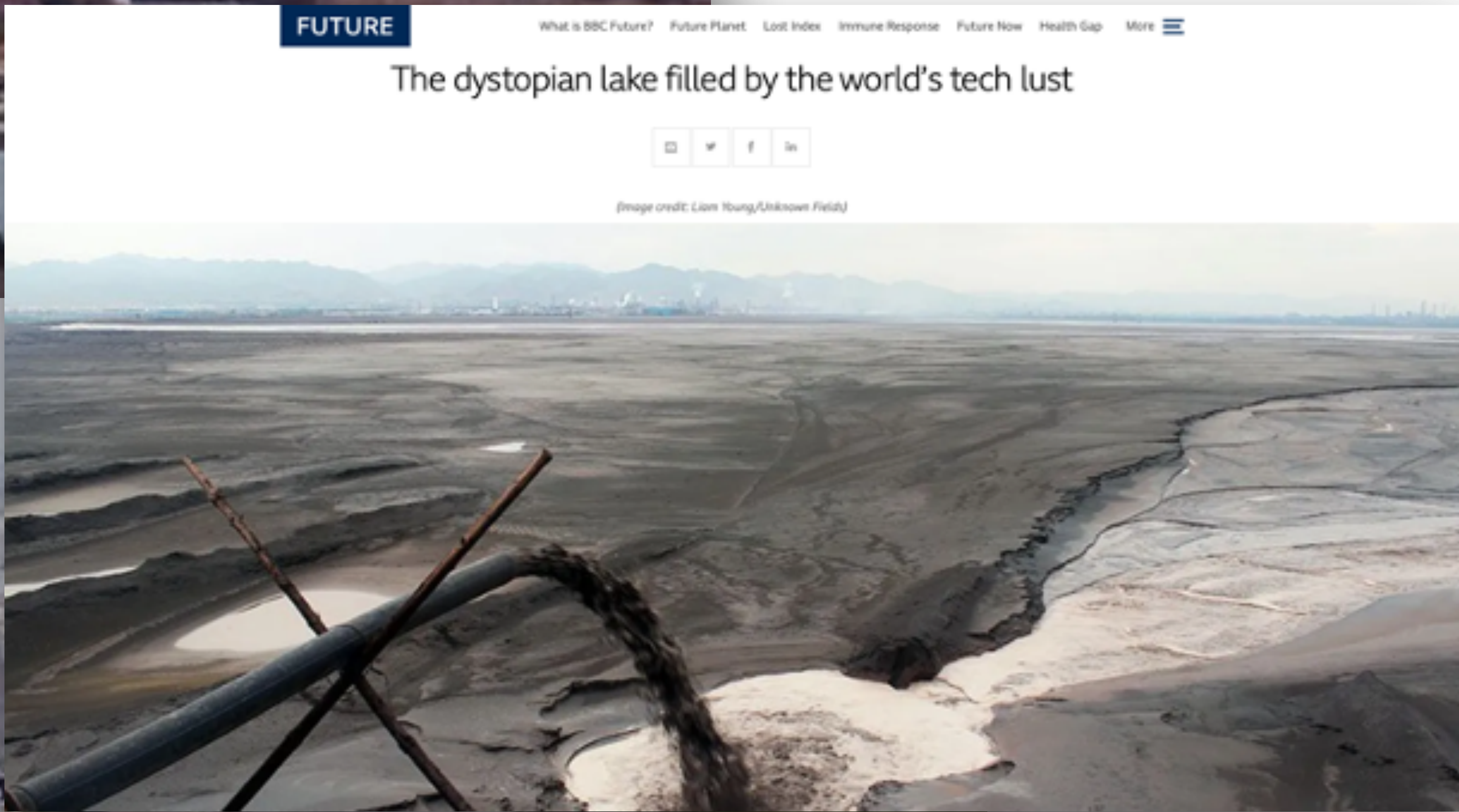
The Apollo 11 mission trip to the moon was very significant in American history. Everybody stopped what they were doing to watch or listen to history being made in the United States. For once in the United States history everybody honestly came together as one to root on the astronauts. The Apollo 11 mission trip brought the United States together as they sent astronauts to the moon and safely back.

Apollo 11 was the very first time in all of history, that a man stepped foot on the moon. The Soviet Union and the United States was in a huge moon race to see who could get there first and safely come back. The moon race began when the Soviet Union launched the very first man ever into space and safely back. The Soviet Union also sent the first woman into space and safely back. President Kennedy responded back by announcing on May 25, 1961, the United States was going to put a man on the moon and bring him safely back home. At this particular time in American history, the United States only had 15 minutes of human-spaceflight experience. Alan Shepard did a suborbital flight which just had happened three weeks prior to President Kennedy's announcement. It was important for the United States to get to the moon and collect data to see if there could ever be life on the moon. The moon landing could lead to further discoveries that tell the United States more about space.

Apollo 11 was launched from Cape Kennedy Space Center on July 16, 1969. The astronauts involved were Neil Armstrong, Michael Collins, and Edwin Aldrin. The spacecraft used for this trip was called the Apollo Command Module. The capsule only had room for three astronauts. The spacecraft only had as much room as a car. While on this trip the astronauts were the first to have hot water in space, and the meals were color-coded. The astronauts didn't have a bathroom to use so instead they used bags and roll-on cuffs. A simple task like using the bathroom could take up to 45 minutes. After traveling three days and four hours, the astronauts finally entered the lunar orbit on July 19. The lunar module Eagle, which had Neil Armstrong and Edwin Aldrin, parted away from the command module, where Michael Collins was. Neil Armstrong was the first astronaut to step foot on the moon and then Edwin Aldrin was the second astronaut to put his footprints on the moon.. Together they took pictures of the terrain, planted a United States flag, and ran a few scientific tests. They successfully made it back three days later, splashing into the Pacific Ocean on July 24.

Since Apollo 11 occurred there have been other Apollo missions to the moon. There have been six Apollo missions since Apollo 11. All of them have been based on going to the moon and gathering data. A total of 12 astronauts have walked on the moon. They all made scientific researches and studied the lunar surface. The Apollo program was costly and very intensive labor. Around 400,000 engineers, technicians, and scientists were involved with these Apollo missions. The cost was \$24 billion (almost \$100 billion in today's money).









Google

does youtube for kids

Q All

Videos

About 705,000,000 results (0.37 seconds)

Showing results for **does youtube for kids require certification?**

Search instead for **does youtube for kids require certification?**

YouTube

https://kids.youtube.com › terms

Terms of Service

Age Requirements You must be at least 13 years old to use the Service; however, children of all ages may use the Service and YouTube Kids (where available) if ...

Missing: certification? | Show results with: certification?

People also ask

What are the requirements for a kids YouTube channel?

Why can't I create a YouTube channel for my child?

Can kids YouTube channel be monetized?

How do I start a YouTube channel for kids?

Feedback

Google Help

https://support.google.com › youtube › answer

YouTube Self-Certification overview

Self-Certification is available for content marked as “made for kids”. Learn more about best practices for kids and family content, as well as our ...

Google Help

https://support.google.com › youtube › answer

Best practices for content with children - YouTube Help

: Review local laws and regulations to find out if minors in your videos need a permit, registration, or license to work. ... Stay safe on YouTube. 5 of 5.

blog.youtube

https://blog.youtube › news-and-events › an-update-o...

An update on kids and data protection on YouTube

Australian Government

Department of Infrastructure, Transport,  
Regional Development, Communications and the Arts

Ministers About us Media centre

Infrastructure, transport & vehicles

Territories, regions & cities

Media, communications & arts

Home

Infrastructure, transport & vehicles

Motor Vehicle Standards laws

Motor Vehicle Standards laws

Listen

From July 1989 until 1 July 2021, the *Motor Vehicle Standards Act 1989* (MVSA) and Motor Vehicle Standards Regulations 1989 made it an offence to import, sell or present new or used imported vehicles to the Australian market for the first time unless they met the national standards, except in circumstances where an exemption was granted.

The MVSA applied to all road vehicles including cars, vans, utilities, small and large buses, trucks, trailers, motorcycles and 3 wheeled vehicles. It covered domestic and international vehicle manufacturers, vehicle importers, automotive parts manufacturers, vehicle owner/operators and other road users to ensure uniform standards applied to road vehicles across Australia.

Approved vehicles were fitted with an identification plate including compliance information and a certification statement from the manufacturer.

Search for approved vehicles fitted with an identification plate on the [RVCS](#) website.

The most common standards relating to IATF 16949


- IATF 16949:** We work with products, which are produced to IATF 16949 requirements. Most organizations are certified to IATF 16949, which was developed by the automotive industry.
- ISO 9001:** Since IATF 16949 isn't designed to be a self-sufficient quality standard but instead works best in conjunction with a comprehensive QMS, ISO 9001 certification makes a great deal of sense for automotive companies looking to demonstrate improvement in customer satisfaction, operating costs, stakeholder relationships, legal compliance, risk management, business credentials and attracting new business.
- ISO 14001:** As the international standard for environmental management systems — or EMS — ISO 14001 is the primary EMS certification for more than 250,000 organizations around the world. As the global standard for any business that wants to manage and positively control all aspects of its environmental impact, ISO 14001 certification is a proven way to demonstrate that you're serious about your business's environmental and economic sustainability.
- ISO 45001:** Along with offering reliable products and services, automotive manufacturers must constantly strive to provide their workers and visitors with a safe and healthy business environment. With the ultimate goal of providing businesses with a framework for controlling and eliminating factors that can lead to illness, injuries and — in worst-case scenarios — death, obtaining ISO 45001 Health and Safety certification is a prudent move for any organization's senior management to support.

English

**Important - Read carefully - Save for future use**  
IKEA of Sweden, Box 702, S-343 81 Älmhult

**Caution!**

- Do not mix old and new or different types batteries.
- Do not mix alkaline, standard (carbon-zinc) or rechargeable batteries.
- Batteries are to be inserted with the correct polarity.
- Exhausted batteries are to be removed from the product.
- Adult supervision is required when charging rechargeable batteries.
- The supply terminals are not to be short-circuited.
- Non-rechargeable batteries are not to be recharged.
- Rechargeable batteries are to be removed from the product before being charged.




The crossed-out wheeled bin symbol indicates that the item should be disposed of separately from household waste. The item should be handed in for recycling in accordance with

**Español**

**Importante. Leer detenidamente. Guardar**  
IKEA of Sweden, Box 702, S-343 81 Älmhult

**Precaución:**

- No mezcles pilas nuevas y gastadas o distintos tipos de pilas.
- No mezcles pilas alcalinas, estándar (carbono-zinc) o recargables.
- Las pilas se deben colocar en el producto con la polaridad correcta.
- Las pilas agotadas se deben quitar del producto.
- Se requiere supervisión de un adulto para recargar pilas.
- No se deben cortocircuitar los terminales.
- Las pilas no recargables no deben ponerse a cargar.
- Las pilas recargables se deben retirar del producto antes de ponerlas a cargar.



La imagen del cubo de basura tachado indica que el producto no debe formar parte de los residuos habituales del hogar. Se debe reciclar

部件名称 Part name	有害物质 Hazardous substance					
	铅(Pb) Lead (Pb)	汞(Hg) Mercury (Hg)	镉(Cd) Cadmium (Cd)	六价铬 (Cr (VI)) Hexavalent Chromium (Cr (VI))	多溴联苯 (PBB) Polybrominated biphenyls (PBB)	多溴二苯醚(PBDE) Polybrominated diphenyl ethers (PBDE)
电子部件 Electronics	X	O	O	O	O	O

本表格依据SJ/T 11364的规定编制

This table is prepared according to provisions of SJ/T 11364.

O:表示该有害物质在该部件所有均质材料中的含量均在GB/T 26572规定的限量要求以下。

O: Indicates that said hazardous substance contained in all of the homogeneous materials for this part is below the limit requirement of GB/T 26572.

X:表示该有害物质至少在该部件的某一均质材料中的含量超出GB/T 26572规定的限量要求。

X: Indicates that said hazardous substance contained in at least one of the homogeneous materials for this part is above the limit requirement of GB/T 26572.

AA-1971160-1

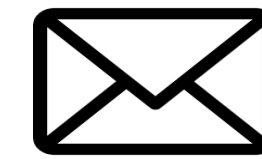
21







# Questions?



[nayyar.zaidi@deakin.edu.au](mailto:nayyar.zaidi@deakin.edu.au)



<http://www.nayyarzaidi.com>



@nayyarzaidi

