# THE DEEPFAKE DEEP DIVE

## UNDERSTANDING THE PROCESS BEHIND DIGITAL DECEPTION

Presented By Ohio Fire Marshal

Forensic Specialist Supervisor

**BRADLEY**

**BARKHURST**

LEVA Symposium

October 29, 2025

# LAST YEAR AT THE SYMPOSIUM THERE WERE A LOT OF GREAT PRESENTATIONS ON DEEPFAKE DETECTION

- The presentations made me really want to learn more about how deepfakes were created.

# THE DEEPFAKE DEEP DIVE

- A deepfake is “A computer-generated replication of a person.” The term merges deep learning with fake image.

(University of Oxford Saïd School of Business Definition.)

# THE DEEPFAKE DEEP DIVE

- A little bit about the history:

-In the 1990s, deepfake technology was researched and developed by academic researchers.

-The technology then soon got out into the hands of general users.

Source: Ultimate Deepfake Detection Using Python by Dr. Nimrita Koul

# THE DEEPFAKE DEEP DIVE

- A little bit about the history:

-In 1997, a project named “Video Rewrite” established the groundwork for automated facial reanimation with the use of machine learning.

-The term deepfake was created by a Reddit user (whose user name was Deepfake). Reddit became a hub for sharing deepfake creations.

Source: Ultimate Deepfake Detection Using Python by Dr. Nimrita Koul

# THE DEEPFAKE DEEP DIVE

Originally many videos involved manipulating celebrity faces onto pornographic actors.  Then it expanded into non-pornographic deepfakes.

A popular early example of an actor used in deepfakes was Nicolas Cage likely because beforehand he was in a movie called “Face Off” where his character has surgery to replace his face.

Cage’s face being inserted onto other actor faces in movies.

Source: Ultimate Deepfake Detection Using Python
by Dr. Nimrita Koul

# THE DEEPFAKE DEEP DIVE

In 2017, the famous Deepfake video of Obama was released with actor Jordan Peel doing the voice and movements.

Then in 2018, the popular app called “FakeApp” was released that allowed the masses to easily generate face-swap videos.

Source: Ultimate Deepfake Detection Using Python
by Dr. Nimrita Koul

# THE DEEPFAKE DEEP DIVE

- The first image I made with AI was done using DALL-E2 by OpenAI (the Chat GPT folks) when I attended a workshop in December of 2022.
- The presenter asked what image I wanted to create.
- I said I wanted a combination of a Ferrari, Guns and Roses and Outer Space.

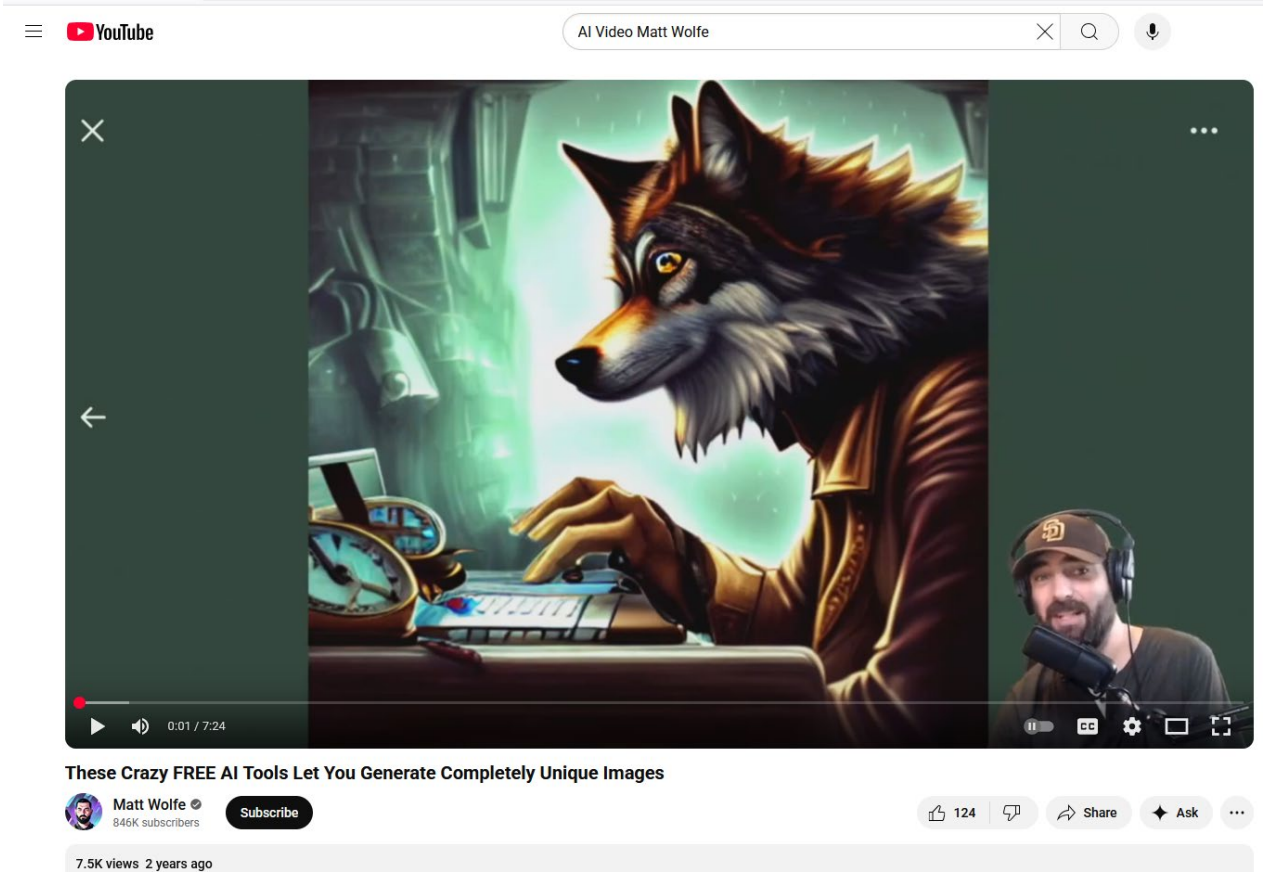
# THE DEEPFAKE DEEP DIVE

- Here is the image it generated:

# THE DEEPFAKE DEEP DIVE

After that I was hooked on AI stuff and wanted to learn more.

So naturally I started watching everything I could about it on YouTube.




Matt Wolfe has a great YouTube channel that explains a lot of what is going on in the AI space.

# THE DEEPFAKE DEEP DIVE

When the AI generated videos first came out they were rather comical and not realistic.

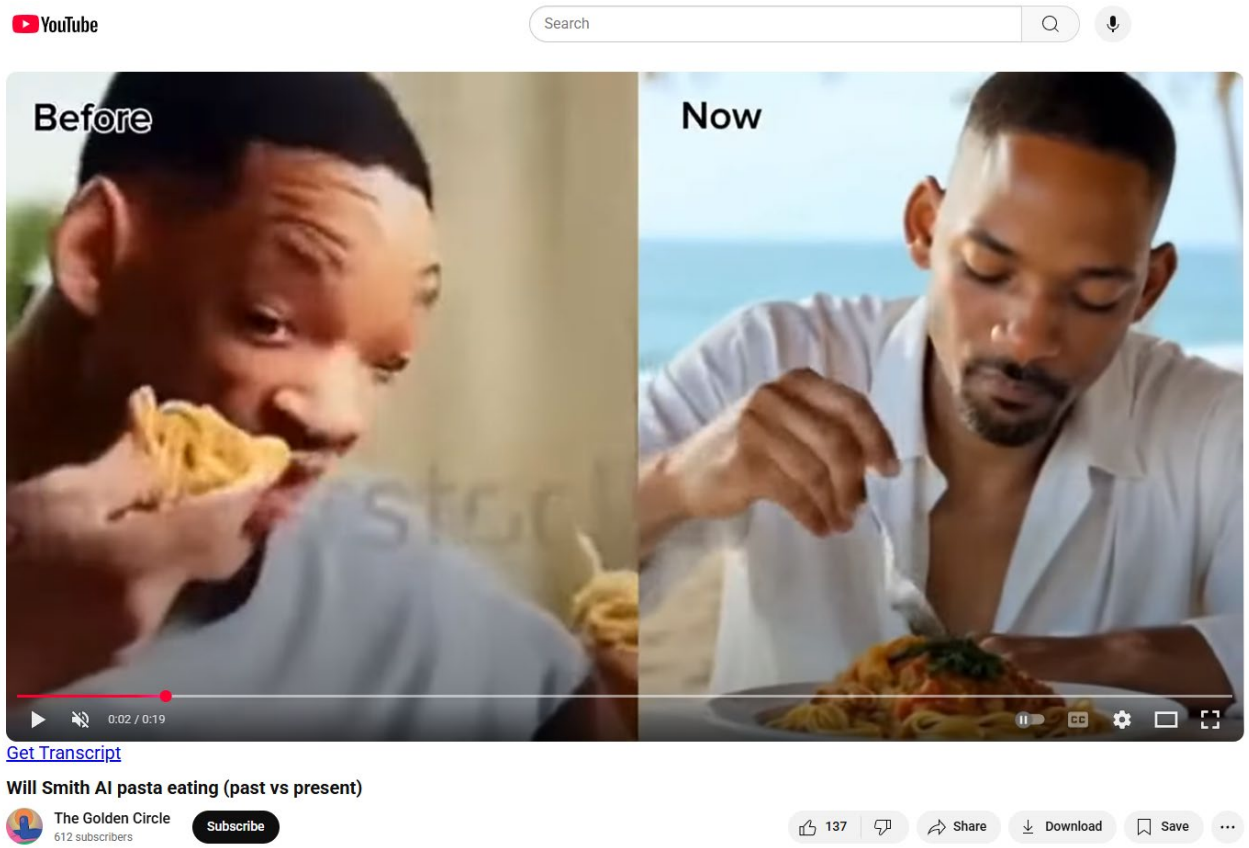
Circa May 2023

# THE DEEPFAKE DEEP DIVE

Another famous video is Will Smith eating spaghetti. You can see the progression in just a short amount of years.

(The before video came out in March 2023.)

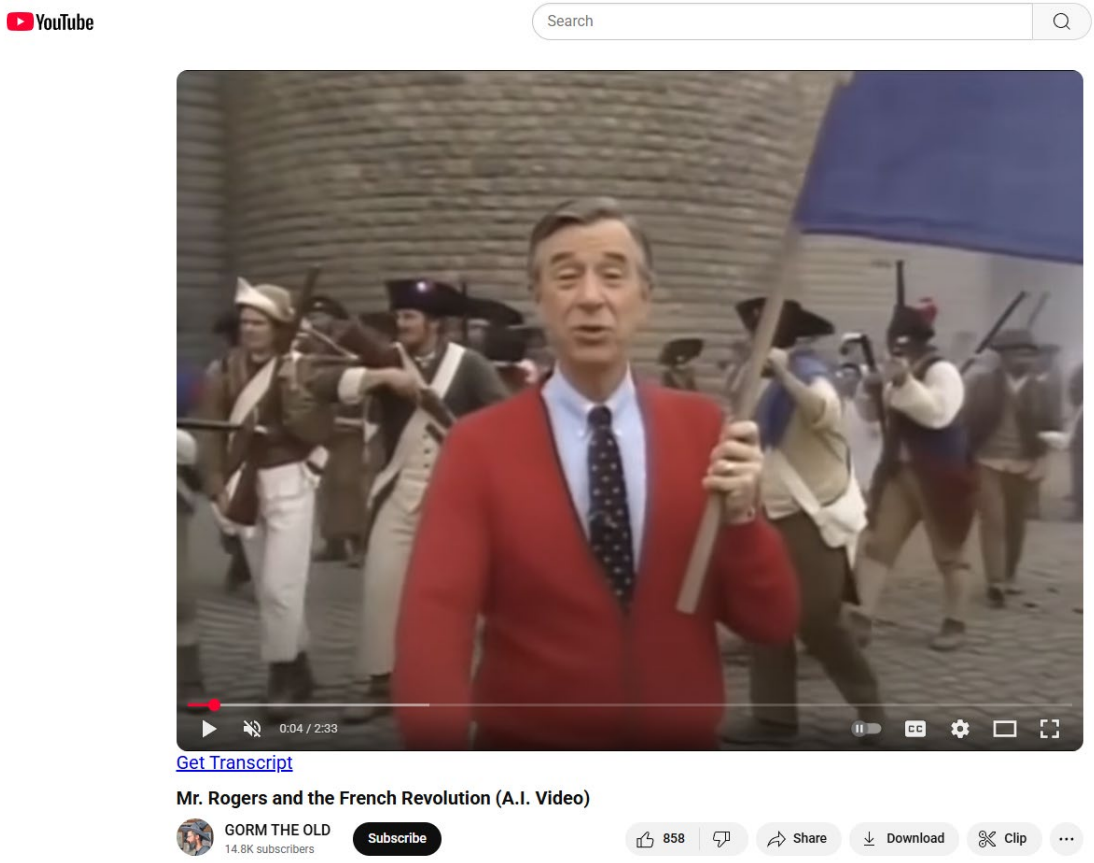
https://www.youtube.com/watch?v=WiTN4TYLG5c



# THE DEEPFAKE DEEP DIVE

With Sora and other tools, now they are getting very realistic:

https://www.youtube.com/watch?v=sOgxaTs0rng



# THE DEEPFAKE DEEP DIVE

- I then started to take formal classes on AI from MIT, Purdue and Oxford.
- The Purdue program really took me deep down the AI/Data analysis rabbit hole which I’m not even going to pretend to understand everything.
- After that program, I realized AI burnout is a thing.

# THE DEEPFAKE DEEP DIVE

- Before we get more into the deepfake stuff, let me give you a brief overview of AI.  If you can wrap your head around this, you'll understand AI better than 90% of people out there.

# THE DEEPFAKE DEEP DIVE

- The University of Oxford Saïd School of Business defines Artificial General Intelligence as this:

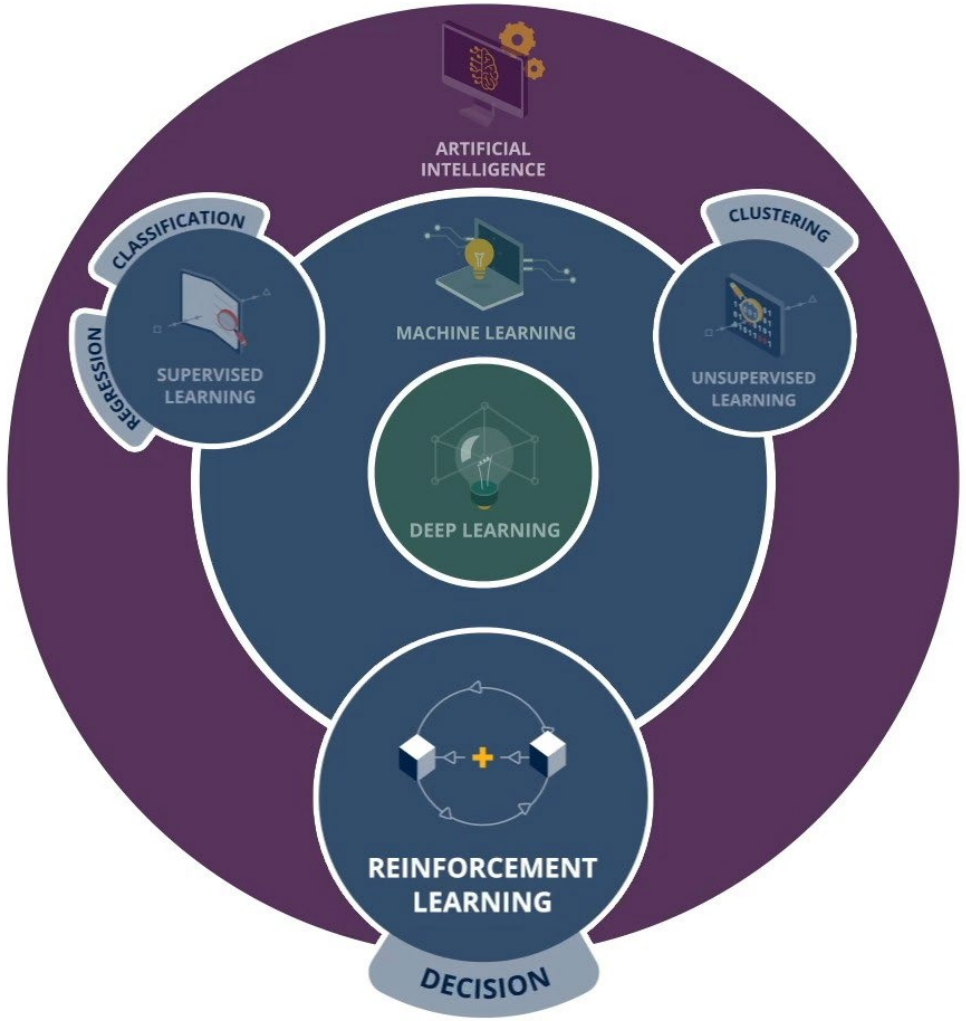
"The goal of building machines that can demonstrate 'general-purpose intelligent behaviour'."

# THE DEEPFAKE DEEP DIVE

- A definition I like created from the Large Language Model Claude is:

The field of computer science dedicated to creating systems that can perform tasks typically requiring human intelligence, such as learning, reasoning, problem-solving, perception, and language understanding. AI encompasses various approaches from rule-based systems to advanced machine learning techniques.

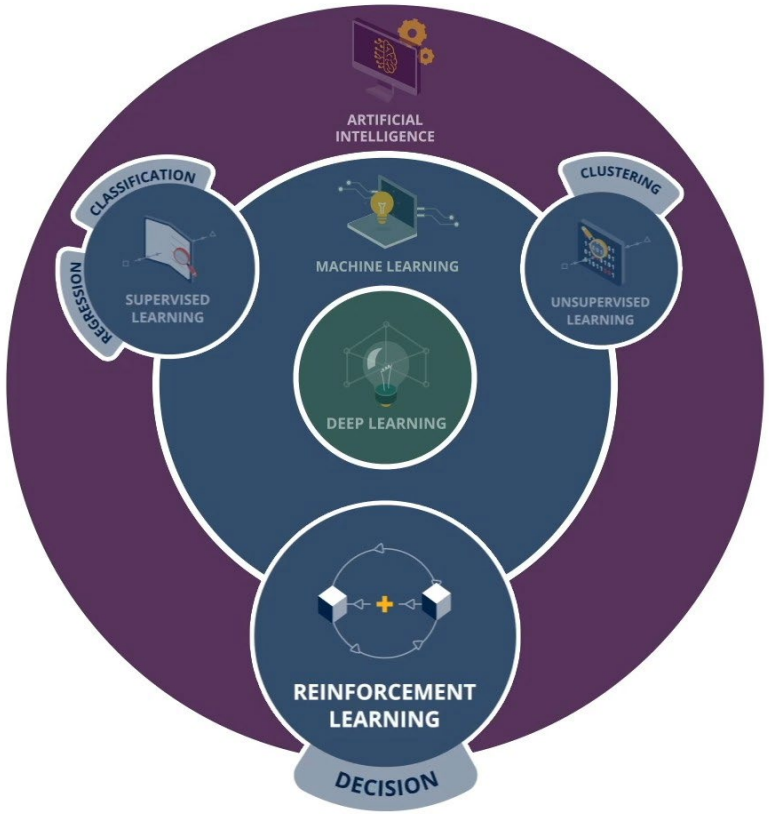
# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:



# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- So we have the overall artificial intelligence.

Then within that, we have Machine Learning.

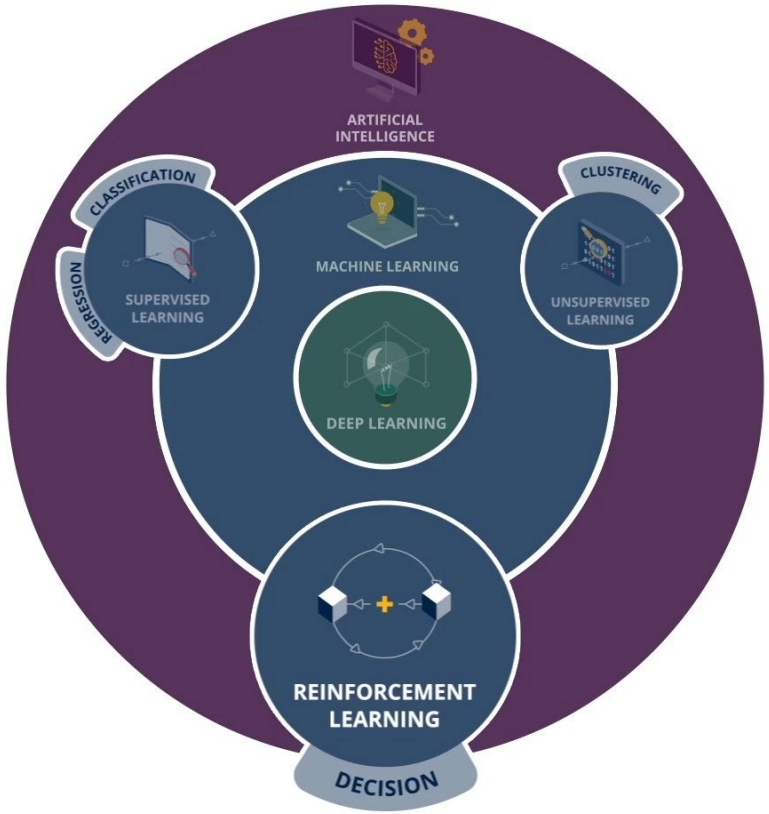


# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- Machine Learning:

Machine learning (ML), a branch of artificial intelligence, is the scientific study of algorithms with the purpose of enabling programs to do things without explicit programming.

(University of Oxford Saïd School of Business Definition.)



# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

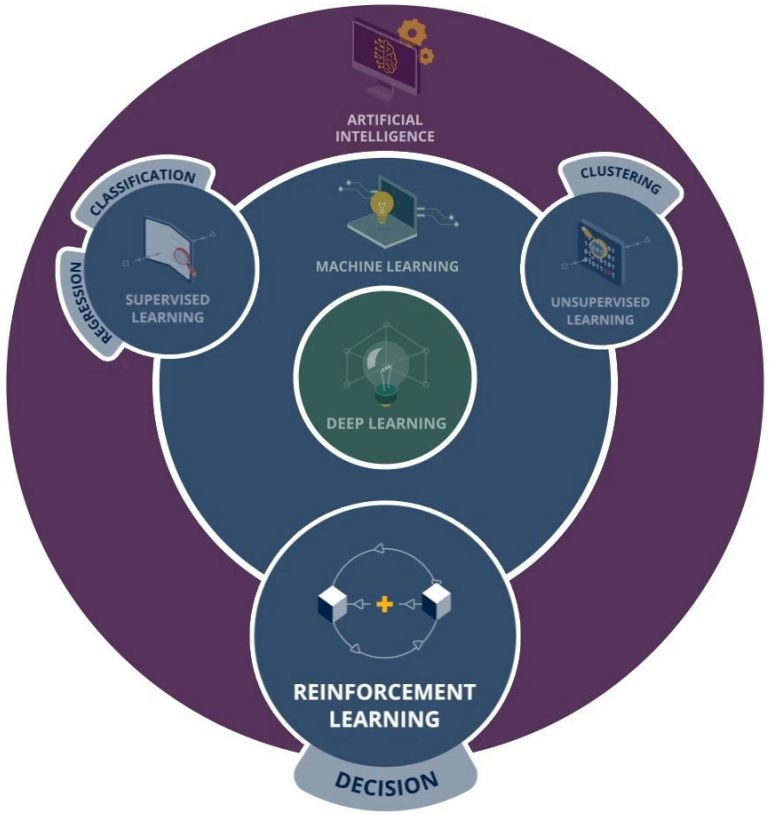
- Within Machine Learning we have:

-Supervised Learning

-Unsupervised Learning

-Reinforcement Learning

-Deep Learning, in the center, can encompass all three of these the machine learning types.



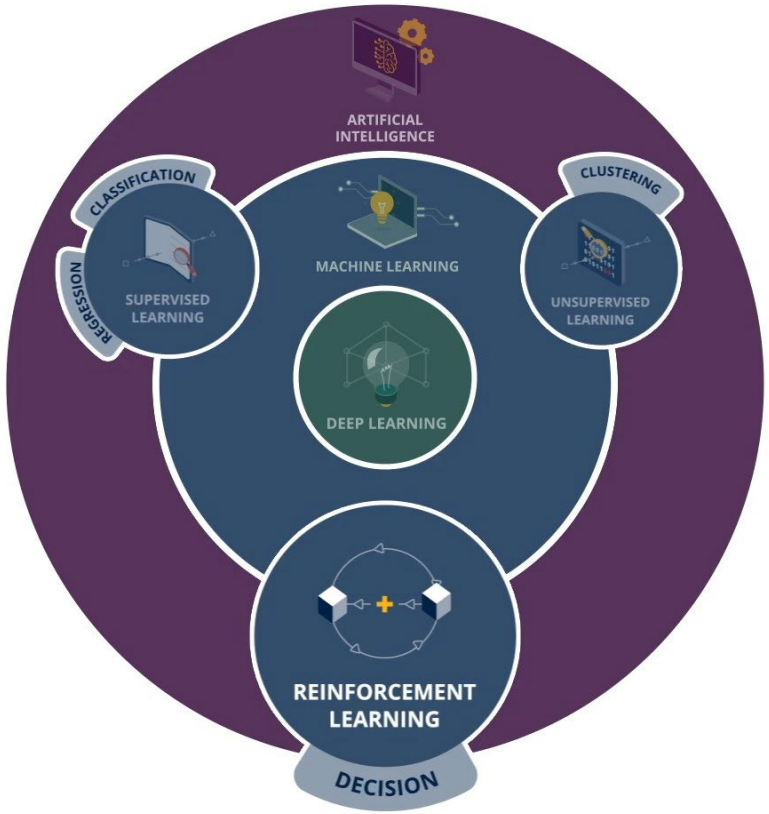
# THE DEEPFAKE DEEP DIVE

## UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- Supervised Learning:

This is a machine learning approach where the algorithm is taught how to generate the correct output for any given input by training it with labelled data. Classification and regression are algorithms associated with supervised learning.

(University of Oxford Saïd School of Business Definition.)

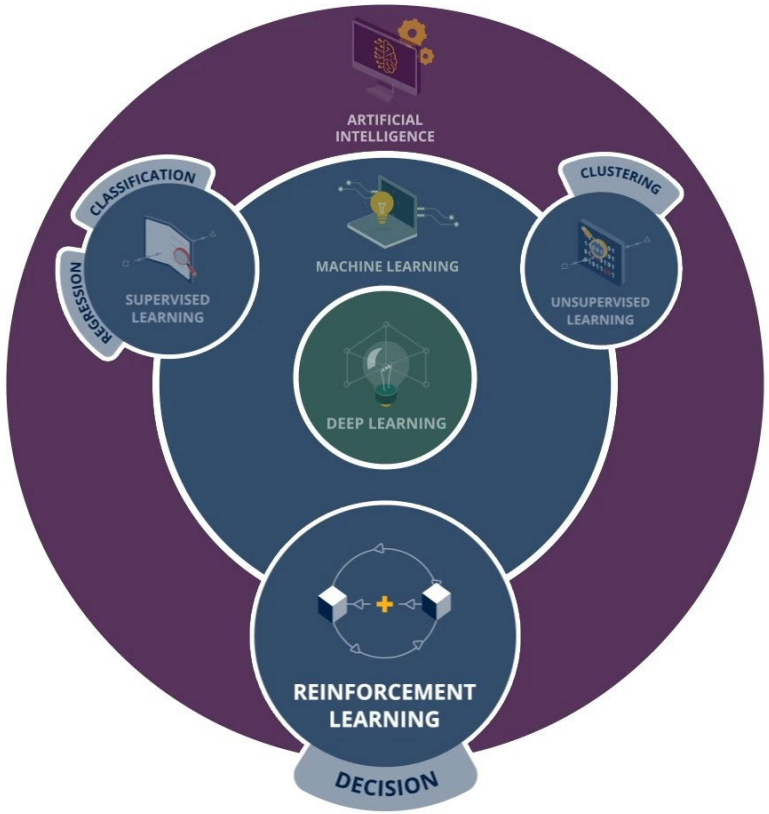


# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- An Example of Supervised Learning:

E-mail Spam Filter, each E-mail has certain words

Such as sender info, subject lines, links, etc.



# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- An Example of Supervised Learning:

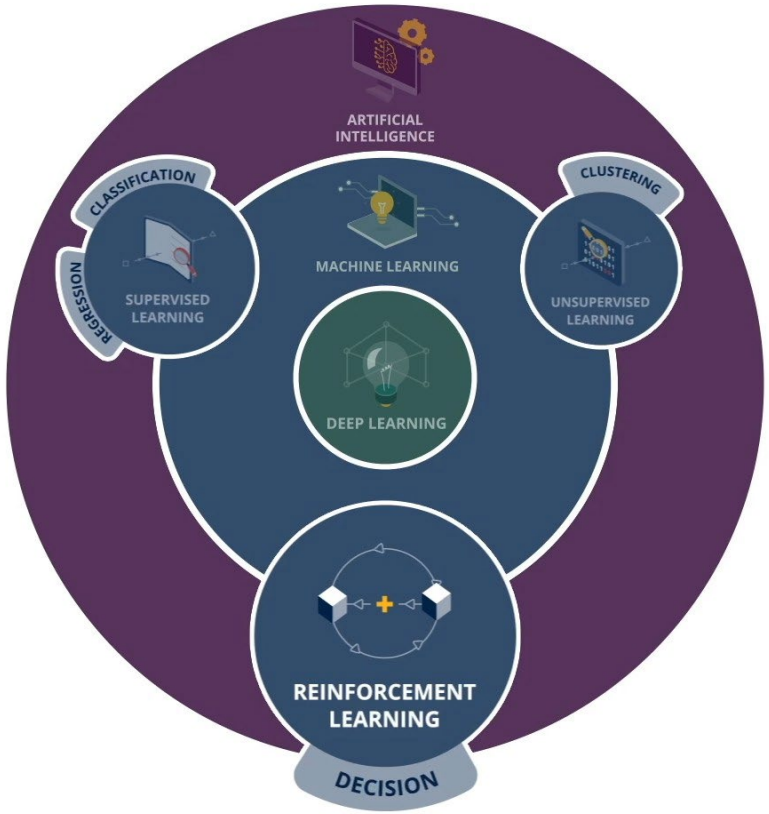
E-mail Spam Filter, The algorithm learns patterns like:

"FREE!!!", "Click here NOW", or "You won $$$" are labelled as Spam.

Emails from your contacts are usually not spam.

Emails with suspicious links are often spam.

When a NEW email arrives (one it's never seen),
the algorithm can predict whether it's spam or not based on what it learned.

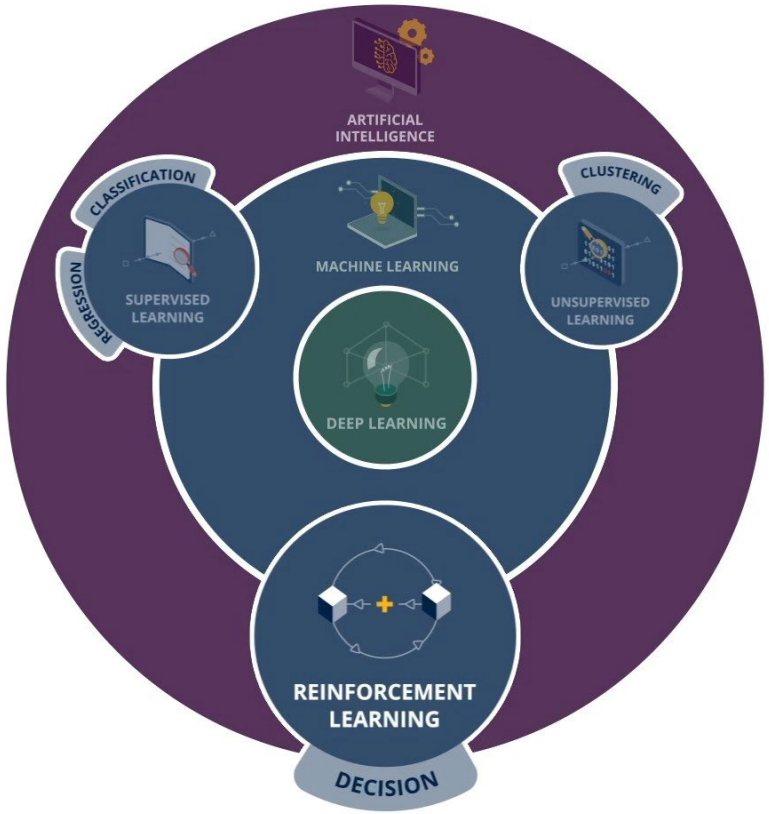


# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- Unsupervised Learning:

This is a machine learning approach where the algorithm is trained to learn the inherent structure of data without explicit labels. Unsupervised learning aims to find regularities in the input to produce an output.

(University of Oxford Saïd School of Business Definition.)

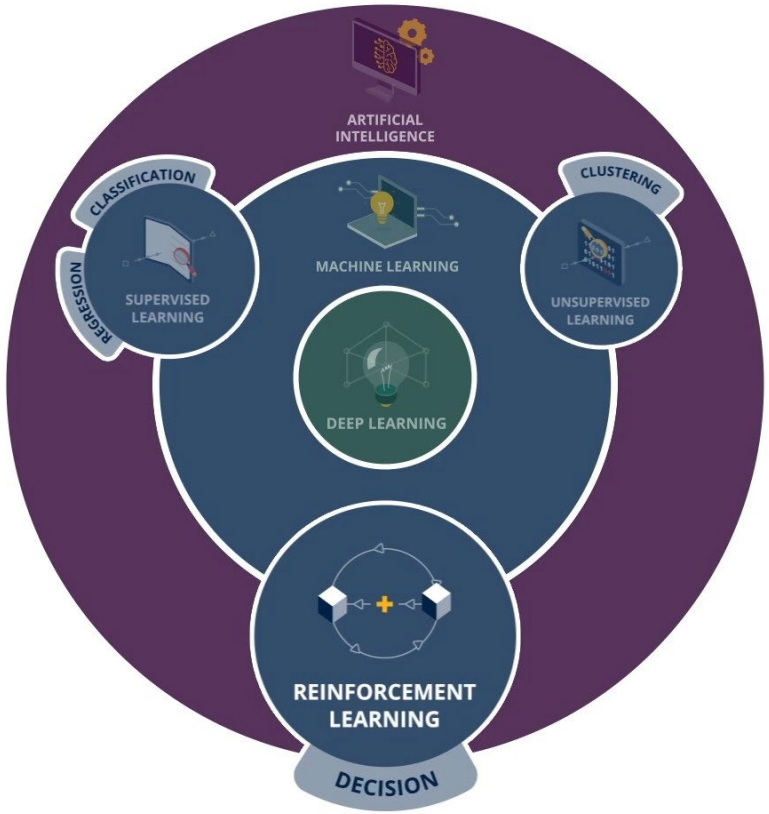


# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- An Example of Unsupervised Learning:

A retail company has data on thousands of customer purchase history, spending amounts, shopping frequency, product categories bought, etc.

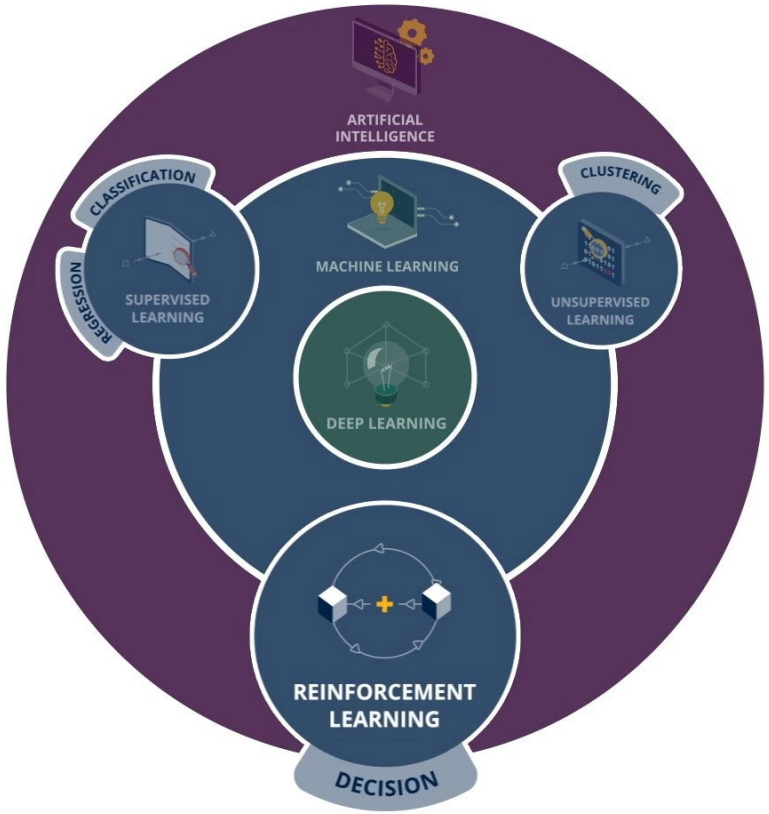
But they don't have any labels. They don't know which "type" of customer each person is.



# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- An Example of Unsupervised Learning:

So for the learning process, you feed all the unlabeled customer data into a clustering algorithm (eg. K-means)



# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- An Example of Unsupervised Learning:

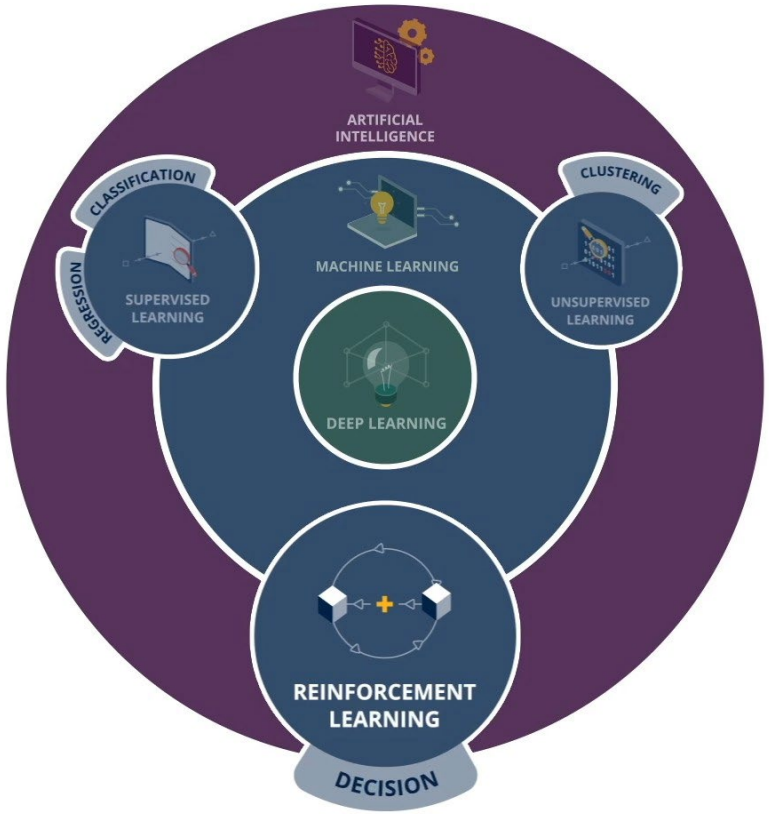
The algorithm finds natural patterns and groups customers on its own.

**Group 1**: High spenders who buy luxury items frequently

**Group 2**: Budget shoppers who buy during sales only

**Group 3**: Regular customers who buy everyday essentials weekly

**Group 4**: Occasional browsers who rarely purchase

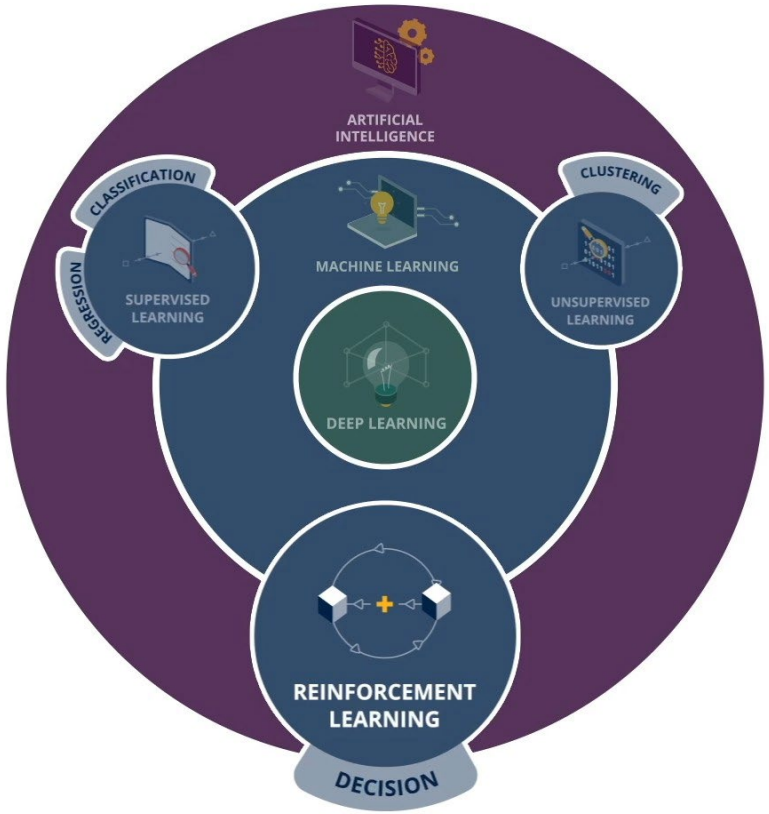


# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- An Example of Unsupervised Learning:

The result:

The company now understands their customer segments and can tailor marketing strategies for each group - even though they never told the algorithm what groups to look for!

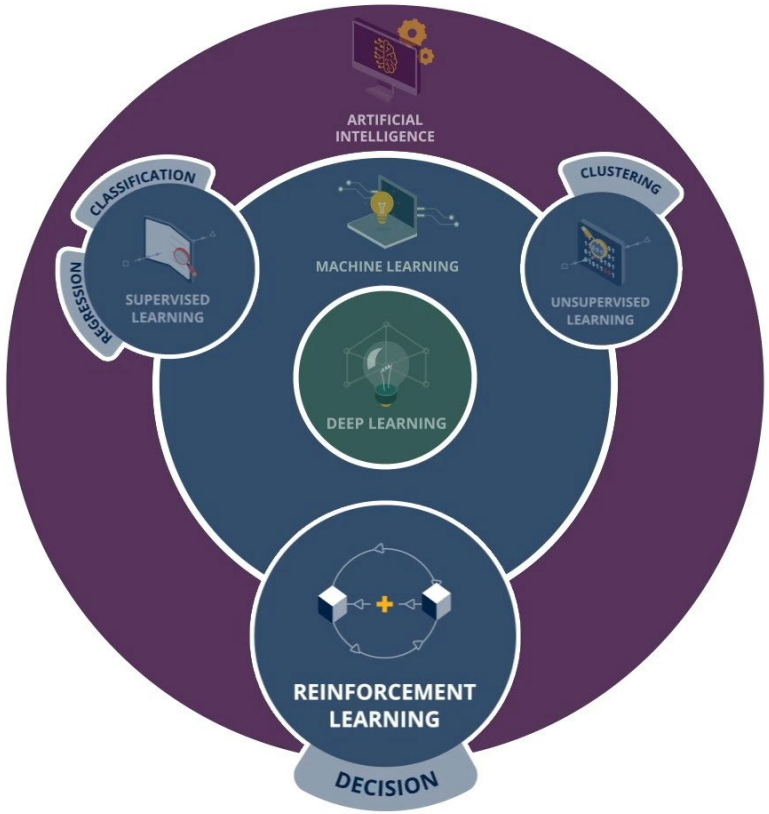


# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- Reinforcement Learning:

This is an approach to machine learning where the agent learns how to behave by performing actions and getting feedback from the environment.
It gets positive feedback for getting a desired outcome and negative feedback for getting an undesired outcome. After each iteration, the algorithm makes adjustments to attempt to maximise the reward.

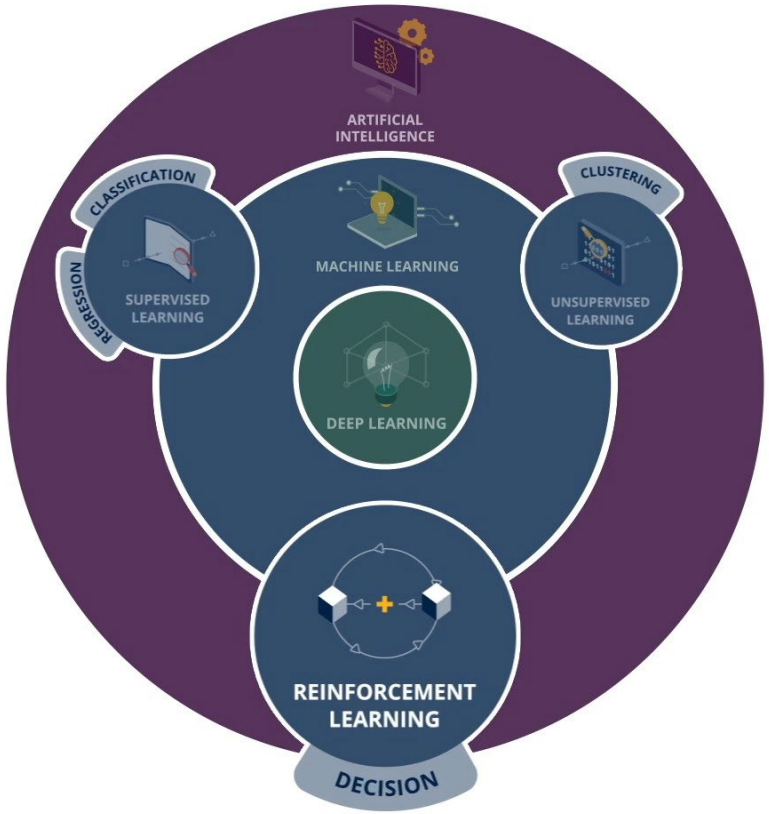
(University of Oxford Saïd School of Business Definition.)



# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- An Example of Reinforcement Learning:

Training AI to play the game Pac Man. The AI learns from trial and error and receives rewards for actions that lead to success and penalties for failure.

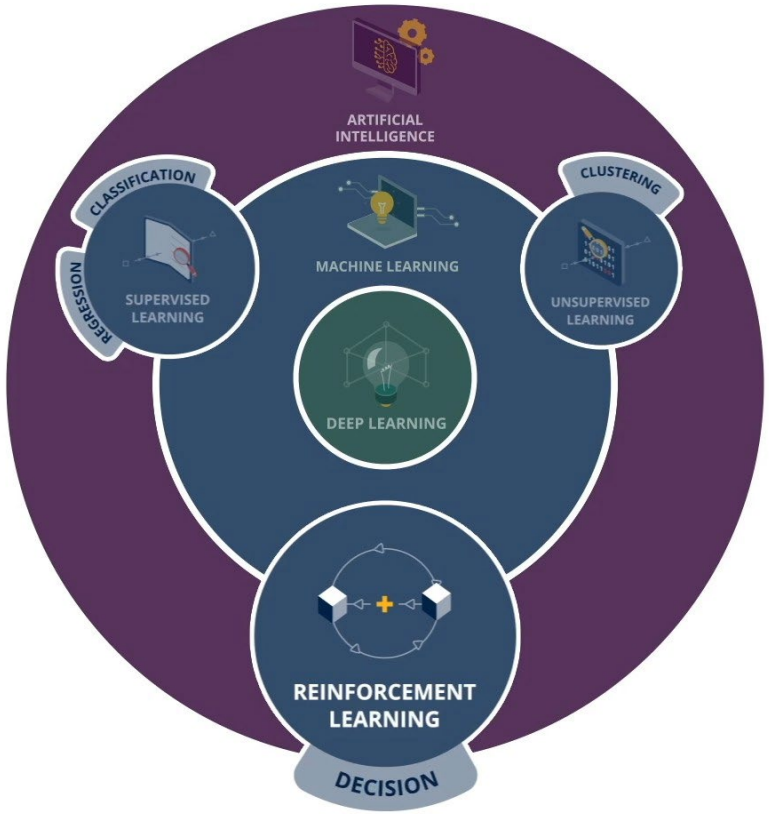


# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

- Finally, we have Deep Learning:

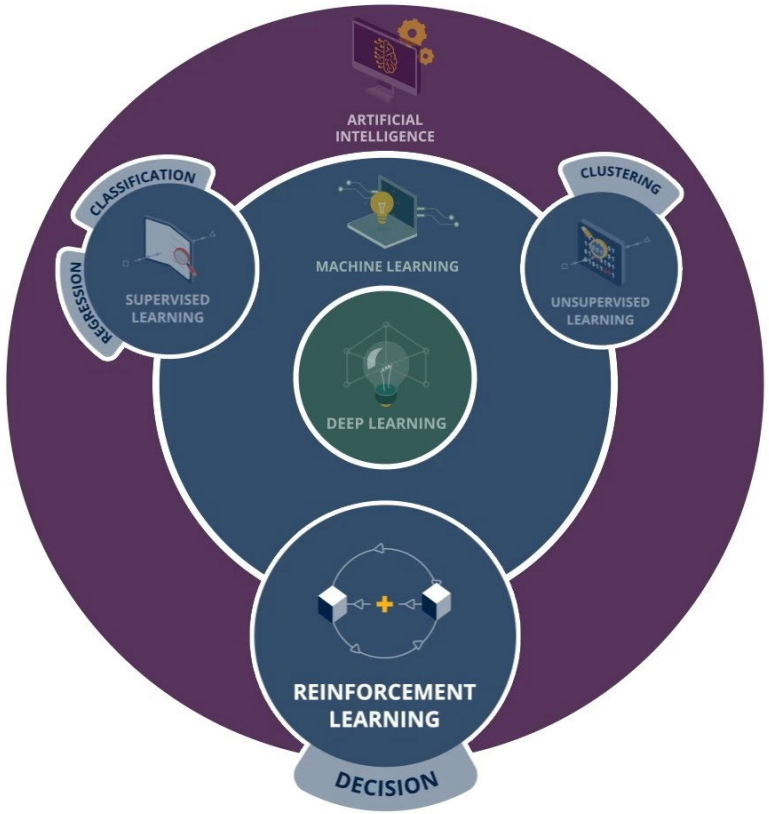
"Methods that are used to train models with several levels of abstraction from the raw input to the output. For example, in visual recognition, the lowest level is an image composed of pixels. In layers as we go up, a deep learner combines them to form strokes and edges of different orientations, which can then be combined to detect longer lines, arcs, circles, and so on. The units of each layer may be thought of as a set of primitives at a different level of abstraction."

(University of Oxford Saïd School of Business Definition.)



# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

Bottom line, deep learning uses neural networks which we'll get into in a second.

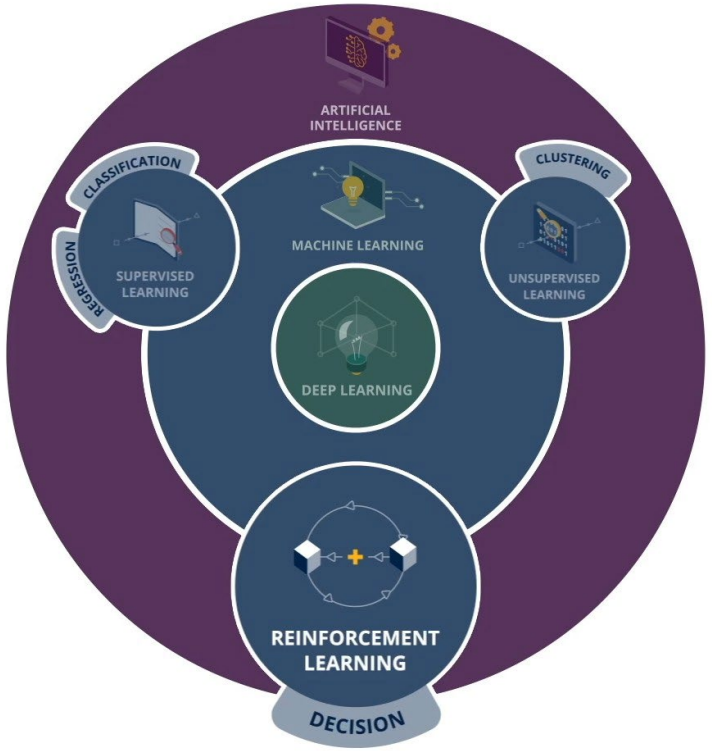


# THE DEEPFAKE DEEP DIVE

Congrats, you now understand AI better than 90% of people out there!

# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

If you had to guess what sort of machine learning is done for Deepfakes, which would you pick?



# THE DEEPFAKE DEEP DIVE UNIVERSITY OF OXFORD SAÏD SCHOOL OF BUSINESS AI FRAMEWORK:

Yep, Deep Learning which can also include supervised deep learning (labeled data, eg cat, dog, etc.)

or unsupervised deep learning (where the data is clustered based on similarity).

Source image: Ultimate Deepfake Detection Using Python by Dr. Nimrita Koul

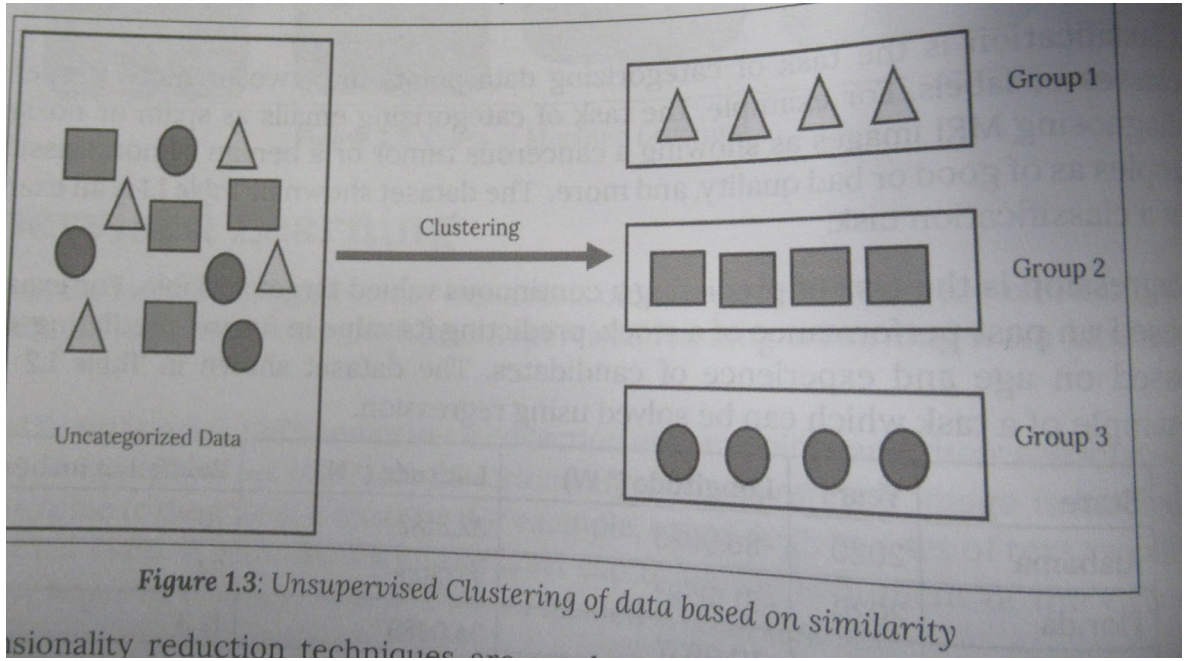


*Figure 1.3: Unsupervised Clustering of data based on similarity*

# THE DEEPFAKE DEEP DIVE

The deep neural networks are required to:

- Understand complex facial features
- Match lighting and angles
- Sync expressions and movements
- Make everything look natural

# THE DEEPFAKE DEEP DIVE

That brings up the next point:

- What is a neural network in AI?

# THE DEEPFAKE DEEP DIVE:

- A neural network is:

Is an approach to machine learning that uses artificial neurons connected to each other in a complex network to generate a desired output from a given input.

(University of Oxford Saïd School of Business Definition.)

# THE DEEPFAKE DEEP DIVE:

- Neural networks are loosely based on how neuron connections in the human brain connect and communicate.

# THE DEEPFAKE DEEP DIVE:

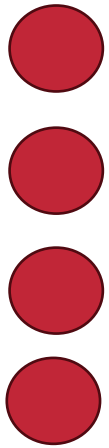
- **Human brain**: Neurons connect to each other through synapses, passing electrical/chemical signals.  With synapses firing in the brain there is a binary outcome, either the nerve cell says “yes” or “no”.

- **Artificial neural networks**: Artificial "neurons" (nodes) connect through weighted connections, passing numerical signals

# THE DEEPFAKE DEEP DIVE:

- **So here is a simplified visual example of a neural network:**

Input Layer

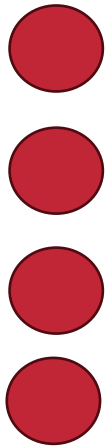


The input layer is like the sensory signals that triggers the network to react. In reality, there would be many input layers. They are numbers arranged in order that represent data (vector).
Often 100 – 1000 values. Eg: [0.5, -2.3, 1.0, 4.7, -0.8] (a vector with 5 numbers)

# THE DEEPFAKE DEEP DIVE:

- **So here is a simplified visual example of a neural network:**

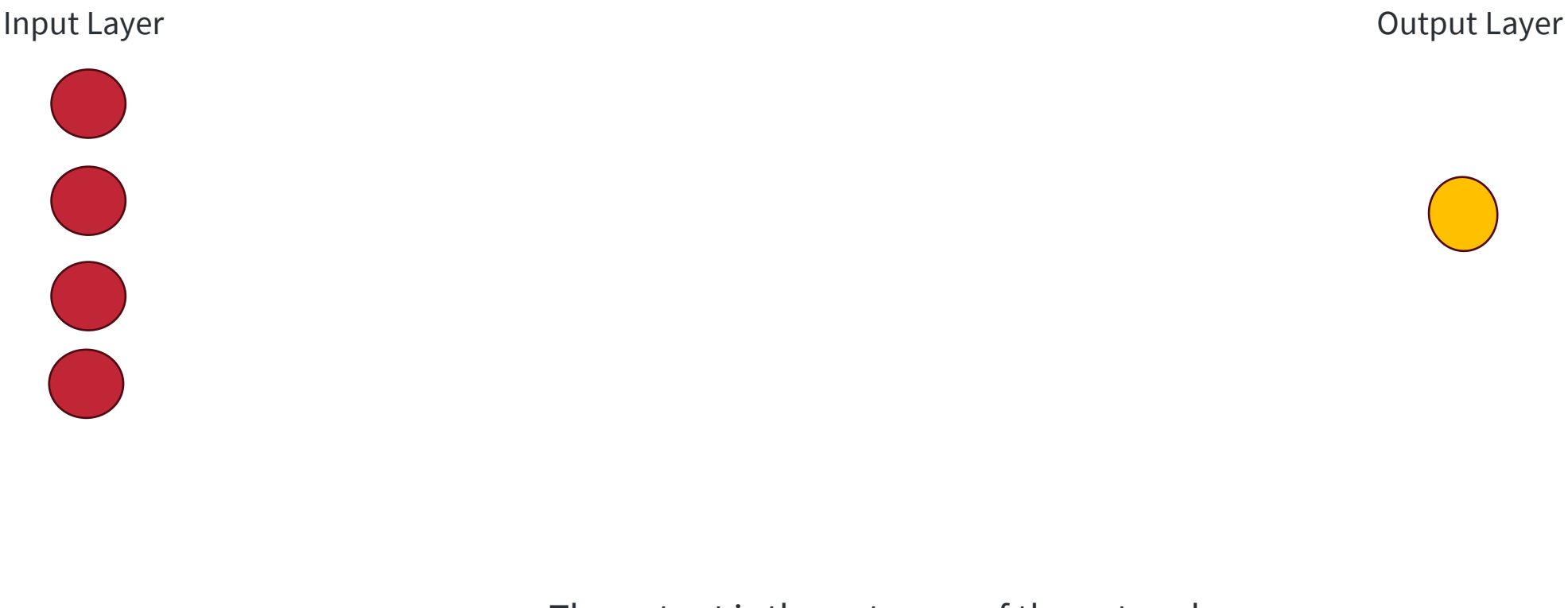
Input Layer



For the example of a deepfake image, a vector could be pixel values representing the brightness of each pixel.

# THE DEEPFAKE DEEP DIVE:

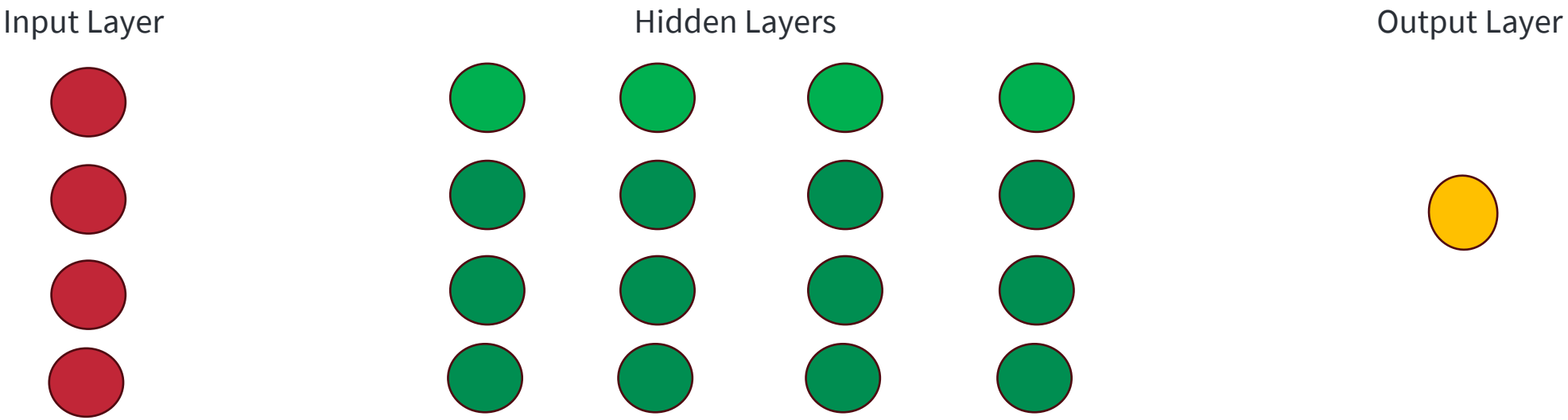
- **So here is a simplified visual example of a neural network:**




The output is the outcome of the network.

# THE DEEPFAKE DEEP DIVE:

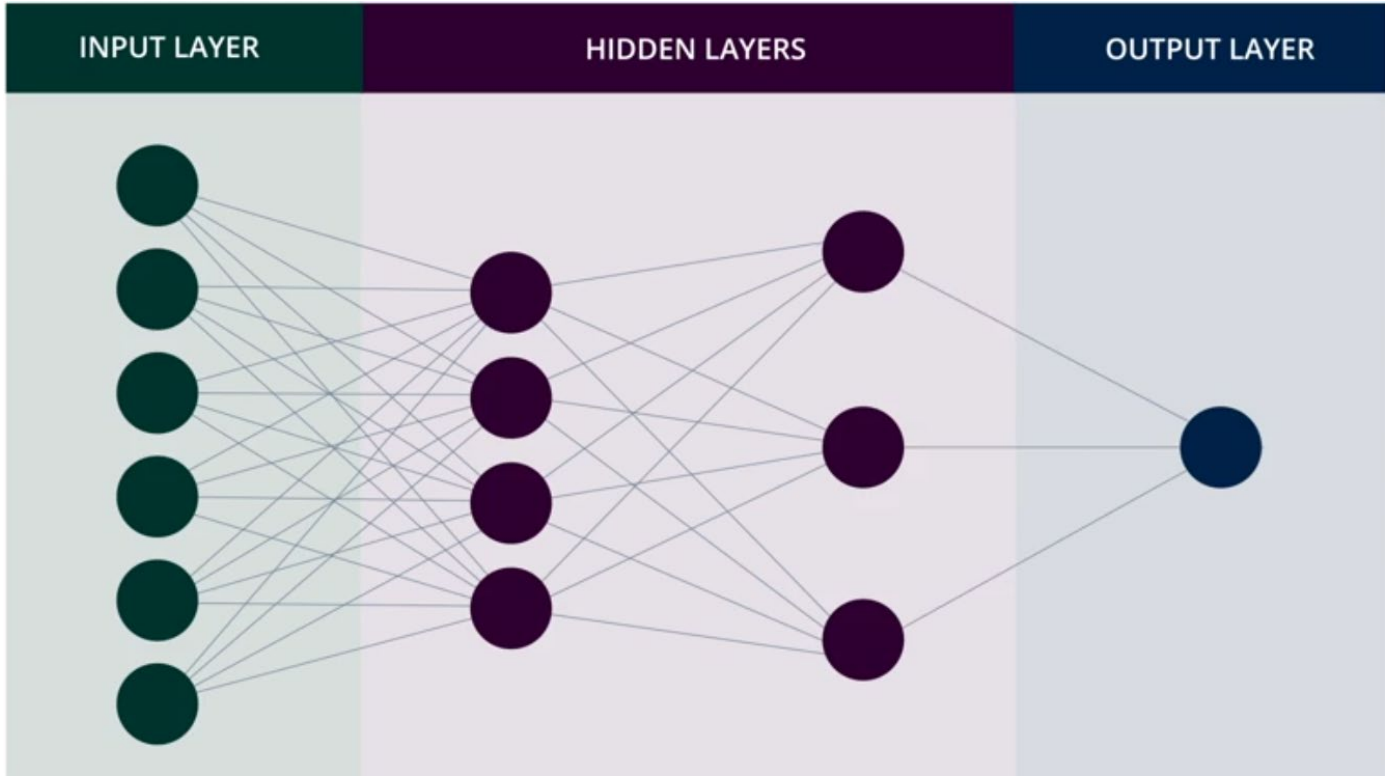
- **So here is a simplified visual example of a neural network:**




Between the input layer and output layer are the hidden layers. The term is a little misleading because It's not some "secret" thing going on in the network.

# THE DEEPFAKE DEEP DIVE:

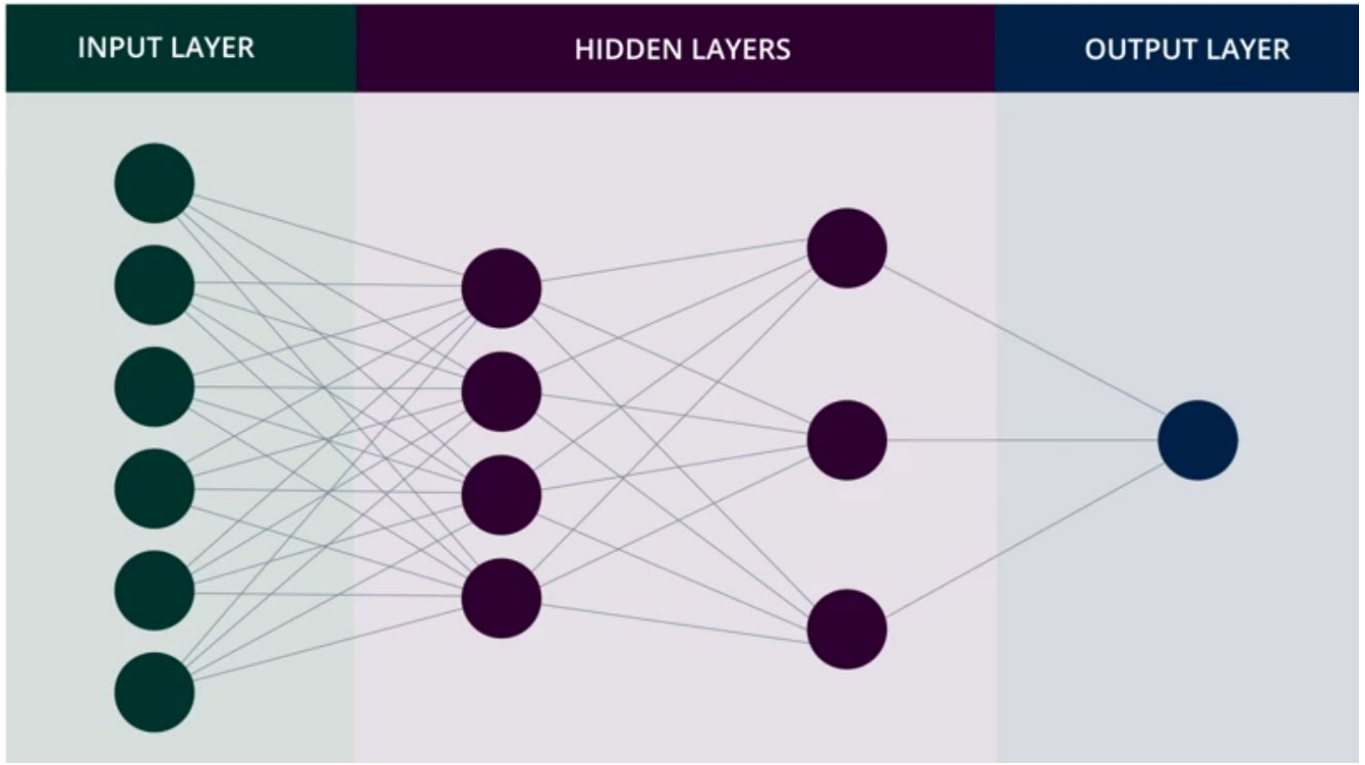
- **Example of the connections of a neural network, from Oxford Saïd Business School:**




The hidden layers are activation functions, which determine whether or not the **connections** (like the synapses in a brain) will lead to a strong output signal.

# THE DEEPFAKE DEEP DIVE:

- **Example of the connections of a neural network, from Oxford Saïd Business School:**




Each connection in the network has what is called a “weight” (the strength of the connection).

# THE DEEPFAKE DEEP DIVE:

- **Example of the connections of a neural network, from Oxford Saïd Business School:**


Some connections are stronger and more likely to trigger the nodes that follow.
In this example of a digit recognizer, the pixel features in the digit “3” image is being inputted.
The darker circles represent more weight.

# THE DEEPFAKE DEEP DIVE:

- So let's say we are generating an AI image of a dog, the neural network is **trained** on thousands of images of dogs (training data).

- The network learns the patterns of what a dog looks like: eye shape, nose, tongue, etc.  (Image classification)

Essentially, what makes a "dog" look like a dog.

# THE DEEPFAKE DEEP DIVE:

- To train the data, you will need at least 500 – thousands of images to build you training set.

- Depending on the size of your data set and computational power, it can take between 12 hours to several weeks to train.

Source: Exploring Deepfakes by Bryan Lyon and Matt Tora

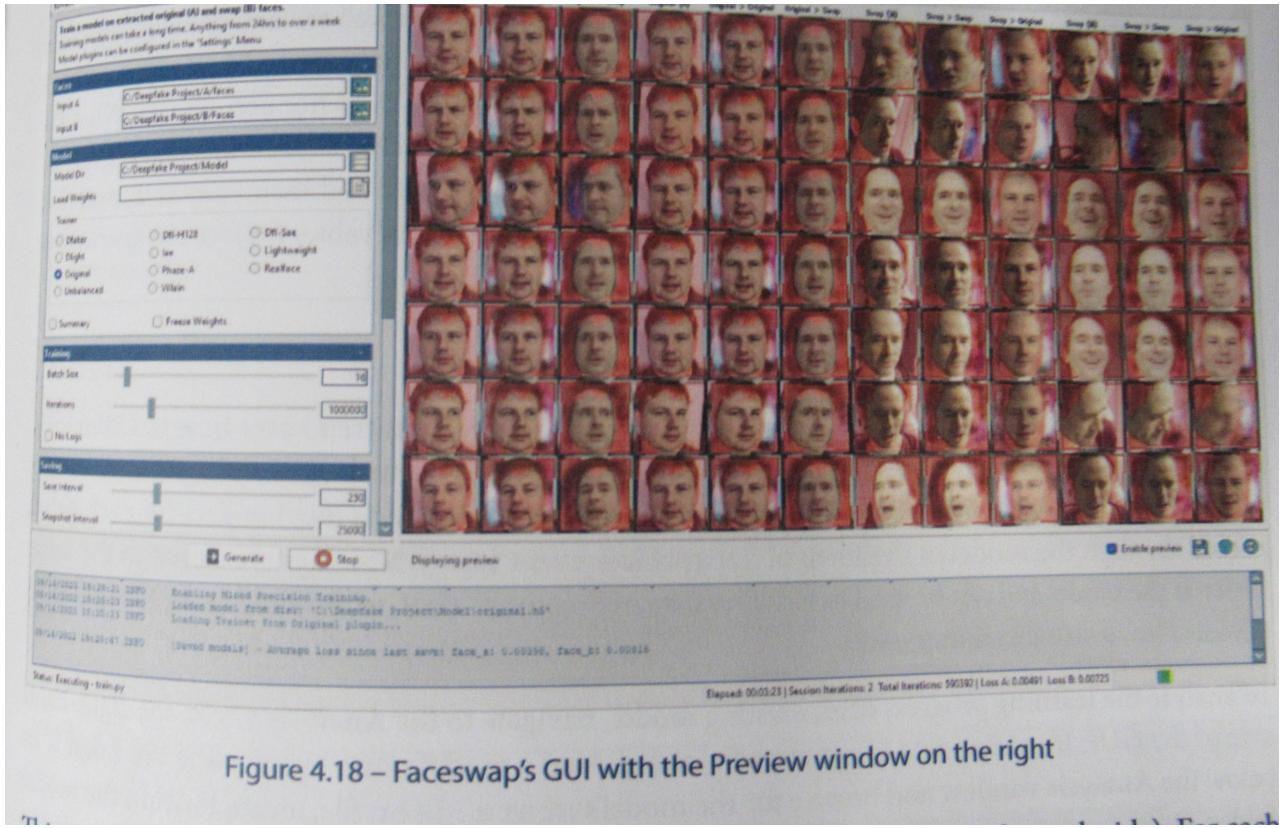


Figure 4.18 – Faceswap's GUI with the Preview window on the right

# THE DEEPFAKE DEEP DIVE:

For a deepfake that uses face swapping (replacing one person's face to another), it typically uses a neural network called a **generative auto-encoder**

- This network uses an **encoder** that breaks the faces into smaller representations.
- This data is then put into a **decoder** which takes the intermediate representations back into faces.

Source: Exploring Deepfakes by Bryan Lyon and Matt Tora

# THE DEEPFAKE DEEP DIVE:

- The algorithm calculates the difference between the original face and the output of the model.  This is called, **loss**.

- Finally, the cycle uses what's called **backpropagation** to move the model towards the correct answer.

Source: Exploring Deepfakes by Bryan Lyon and Matt Tora

# THE DEEPFAKE DEEP DIVE:

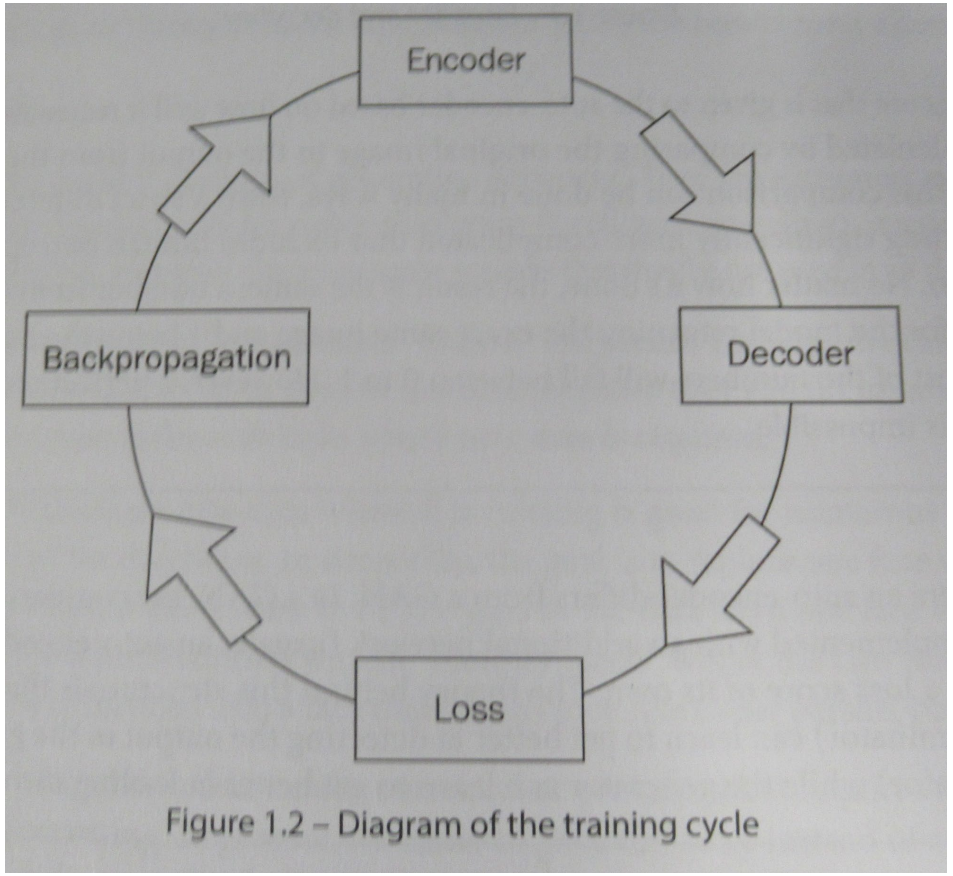


Figure 1.2 – Diagram of the training cycle

Source: Exploring Deepfakes by Bryan Lyon and Matt Tora

# THE DEEPFAKE DEEP DIVE:

**Generative autoencoder:**

Where an encoded face is compressed into data. It's decoded back, but swapped with another person's facial features.

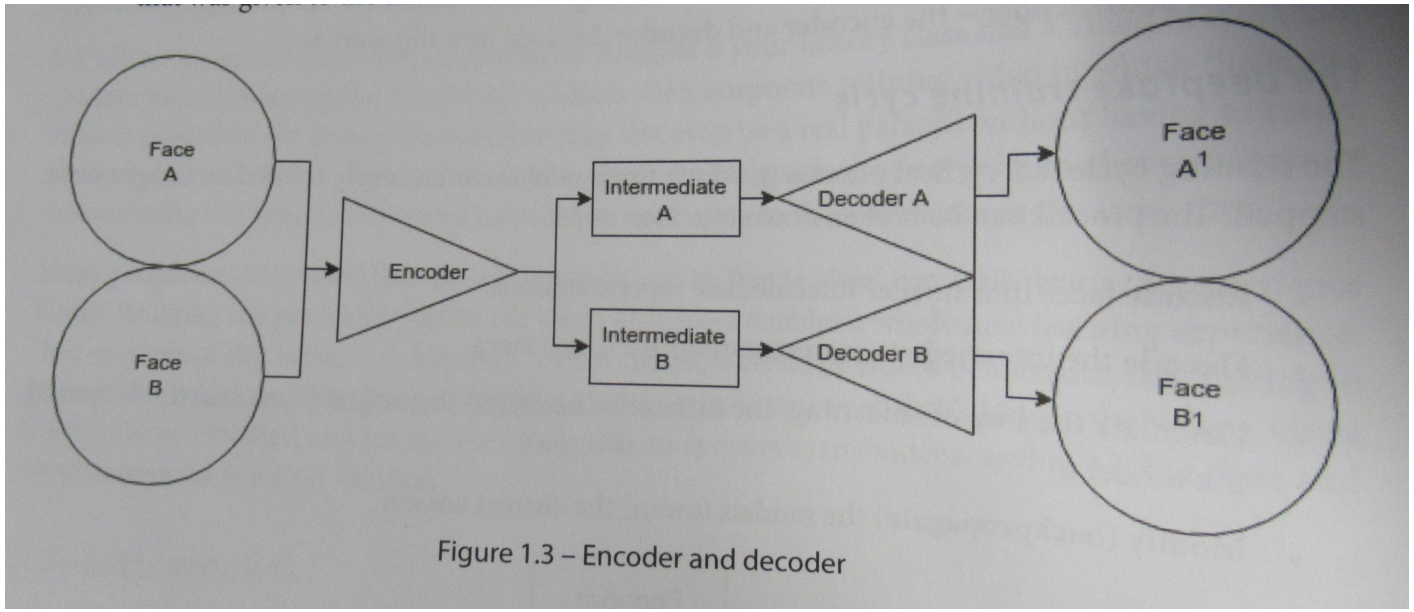


Figure 1.3 – Encoder and decoder

Source: Exploring Deepfakes by Bryan Lyon and Matt Tora

# THE DEEPFAKE DEEP DIVE:

**Example of a face swap progression using a tool called Faceswap.py:**

Source: Exploring Deepfakes by Bryan Lyon and Matt Tora

This book goes into great detail about this workflow process and how you can make your own face swap deepfake.

Figure 4.38 – Faceswap's Preview tool

# THE DEEPFAKE DEEP DIVE:

- **Transformers** are used in understanding the relationships of different parts of the image.

They maintain consistency across an entire face making sure the eyes, nose and mouth all look natural together.

# THE DEEPFAKE DEEP DIVE:

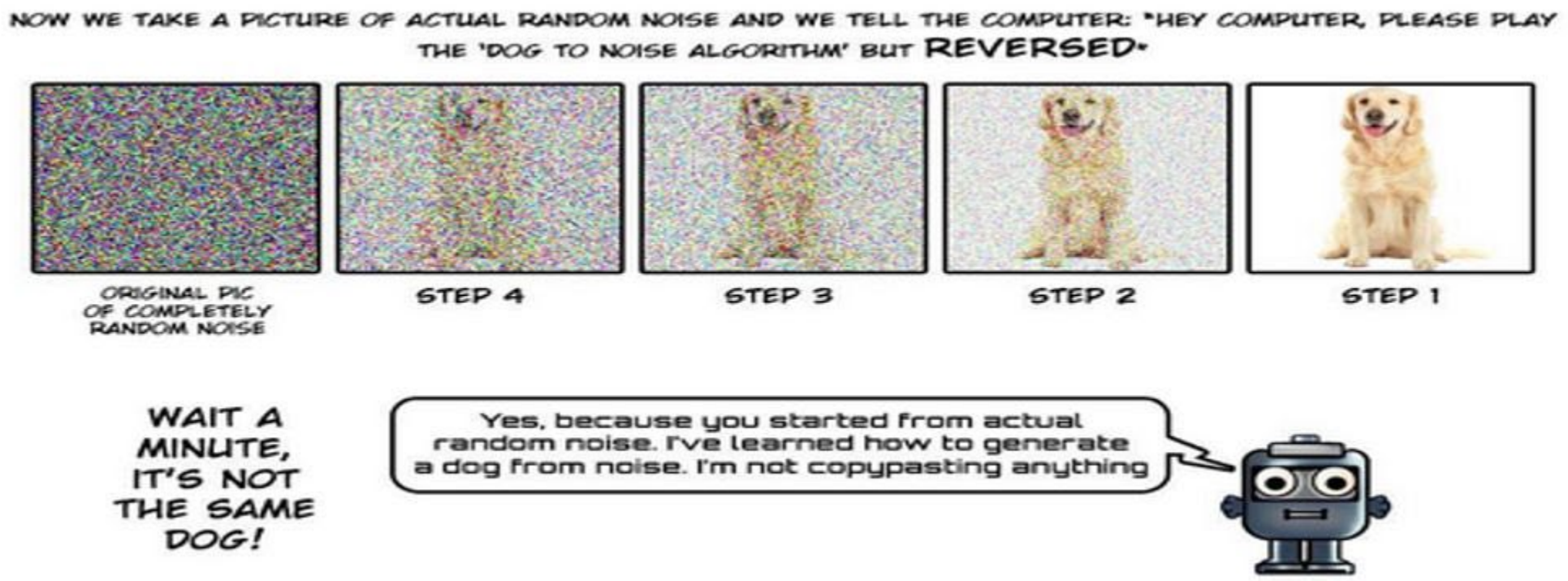
- For diffusion models, as used in Stable Diffusion, Sora, Dall-E, Midjourney (non-face swaps), once trained, the pixel values of the data are then converted into noise. This process is called diffusion.


Source:

https://x.com/owen_roe/status/1607457762102894595

# THE DEEPFAKE DEEP DIVE: SOURCE: HTTPS://X.COM/OWEN_ROE/STATUS/1607457762102894595

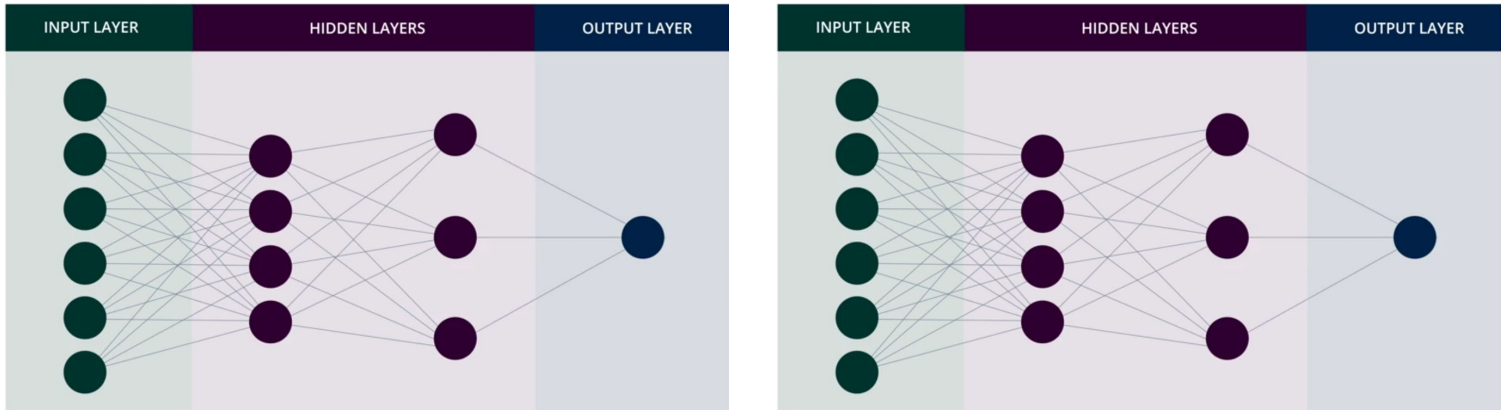
The noise is then used to generate the fake dog:



# THE DEEPFAKE DEEP DIVE:

## What are GANs?

- **GANs** are generative adversarial networks (so we now have two neural networks!)




**Images from Oxford Saïd Business School**

# THE DEEPFAKE DEEP DIVE:

**What are GANs?**

- **GANs** are generative adversarial networks (so we now have two neural networks!)

(do not get this confused with face swapping deepfakes that use a generative auto-encoder, face swaps typically don't use GANs)

# THE DEEPFAKE DEEP DIVE:

**What are GANs?**

- **GANs** are used when typically, everything is AI generated such as in ~~Sora or Midjourney~~. GANs are not used as much these days but were used for example with the website www.thispersonshouldnotexist.com I still want to mention it because Amped Authenticate has detection for it.

# THE DEEPFAKE DEEP DIVE:

- **GANs** can use unsupervised or semi-supervised learning since they learn from patterns from a large collection of images or videos without needing the individual labels.

They figure out facial features, expressions and movements on their own.

# THE DEEPFAKE DEEP DIVE:

The first type of network in Generative Adversarial Networks:

Is the **generator** which creates the fake images/videos.

# THE DEEPFAKE DEEP DIVE:

The first type of network in Generative Adversarial Networks:

Is the **generator** which creates the fake images/videos.

How?

# THE DEEPFAKE DEEP DIVE: SOURCE: HTTPS://X.COM/OWEN_ROE/STATUS/1607457762102894595

So, remember our real image which was diffused into visual noise?


In a **generator** network, that noise is denoised into a new image that is fake. (similar to the diffusion process.)

# THE DEEPFAKE DEEP DIVE:

- Different random inputs = different dog variations.
- The weights/connections in the neural network get adjusted to “remember” what makes a dog look like a dog.

# THE DEEPFAKE DEEP DIVE:

- A good analogy is an artist who has spent years studying dogs.  When you give them random inspiration (the noise), they can draw a realistic dog because they've learned what a dog looks like.

# THE DEEPFAKE DEEP DIVE:

Generative Adversarial Networks:

So you have “real” sampled data from the trained data set and you have “fake” data made by the **generator** network. Both of these sets of data then go to the 2nd network:

# THE DEEPFAKE DEEP DIVE:

Generative Adversarial Networks:

The 2nd neural network in GANs are called **discriminators.**

# THE DEEPFAKE DEEP DIVE:

Generative Adversarial Networks:

The **discriminator** network determines if the data is real (from the training samples) or if the data is fake (from the **generator**).

# THE DEEPFAKE DEEP DIVE:

Generative Adversarial Networks:

If a fake image is detected the by the **discriminator**, the **generator** is penalized.

# THE DEEPFAKE DEEP DIVE:

Discriminator: FAKE DETECTED, penalty!

Boredpanda.com

The Discriminator throws a red flag, and the generator gets a penalty!

If the discriminator detects a fake, what is  done to correct the network?

# THE DEEPFAKE DEEP DIVE:

Generative Adversarial Networks:

If penalized, the **generator** network has to go from the output all the way back through to the first hidden layer.  This process is called, “back propagation”

# THE DEEPFAKE DEEP DIVE:

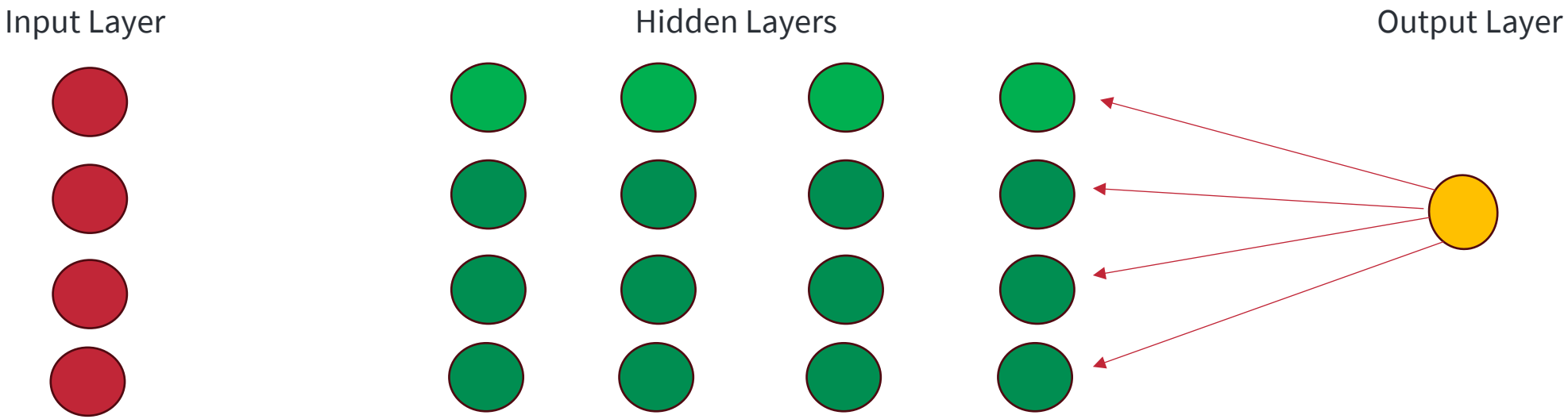
Generative Adversarial Networks:

The weights of the hidden layers are then readjusted to create a new output.

**Generator**, “No problem, I’ll adjust the weights, create a new output and see if I can fool the **discriminator** again!”

# THE DEEPFAKE DEEP DIVE:

- **The Output can be right or wrong:**




If the discriminator detects a fake (an error), the flow goes back (backpropagation) from the output layer all the way to the first hidden layer where the weights are adjusted.
This is done so the next time the input layers are triggered,
there will be a more accurate output.

# THE DEEPFAKE DEEP DIVE:

When the generator network gets better at creating fake data, the discriminator becomes worse in detecting it.

# THE DEEPFAKE DEEP DIVE:

## A look at GANs:

Source: Ultimate Deepfake Detection Using Python

By Dr. Nimrita Koul

What is not shown is, if the discriminator does not detect an error, it gets outputted.

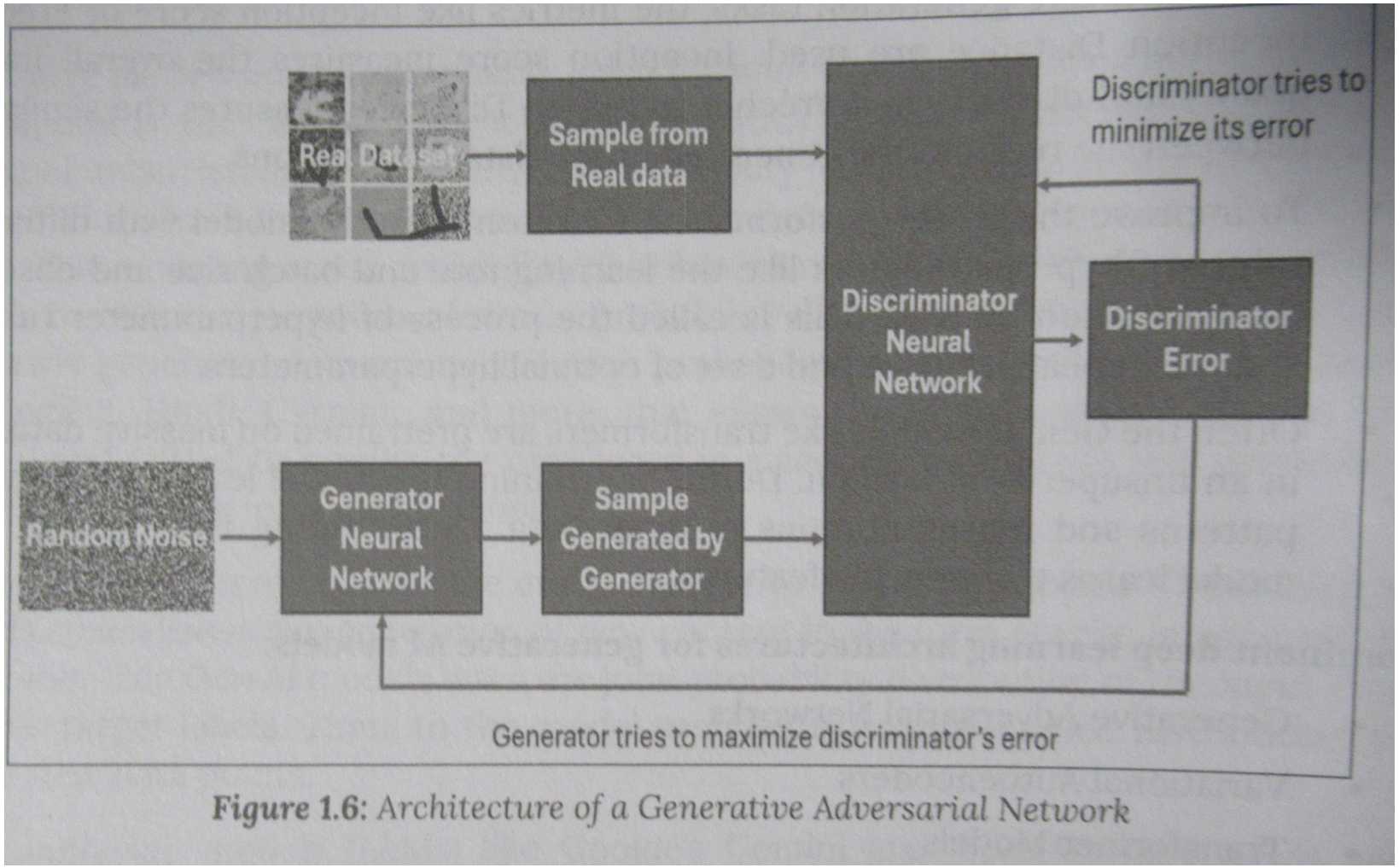
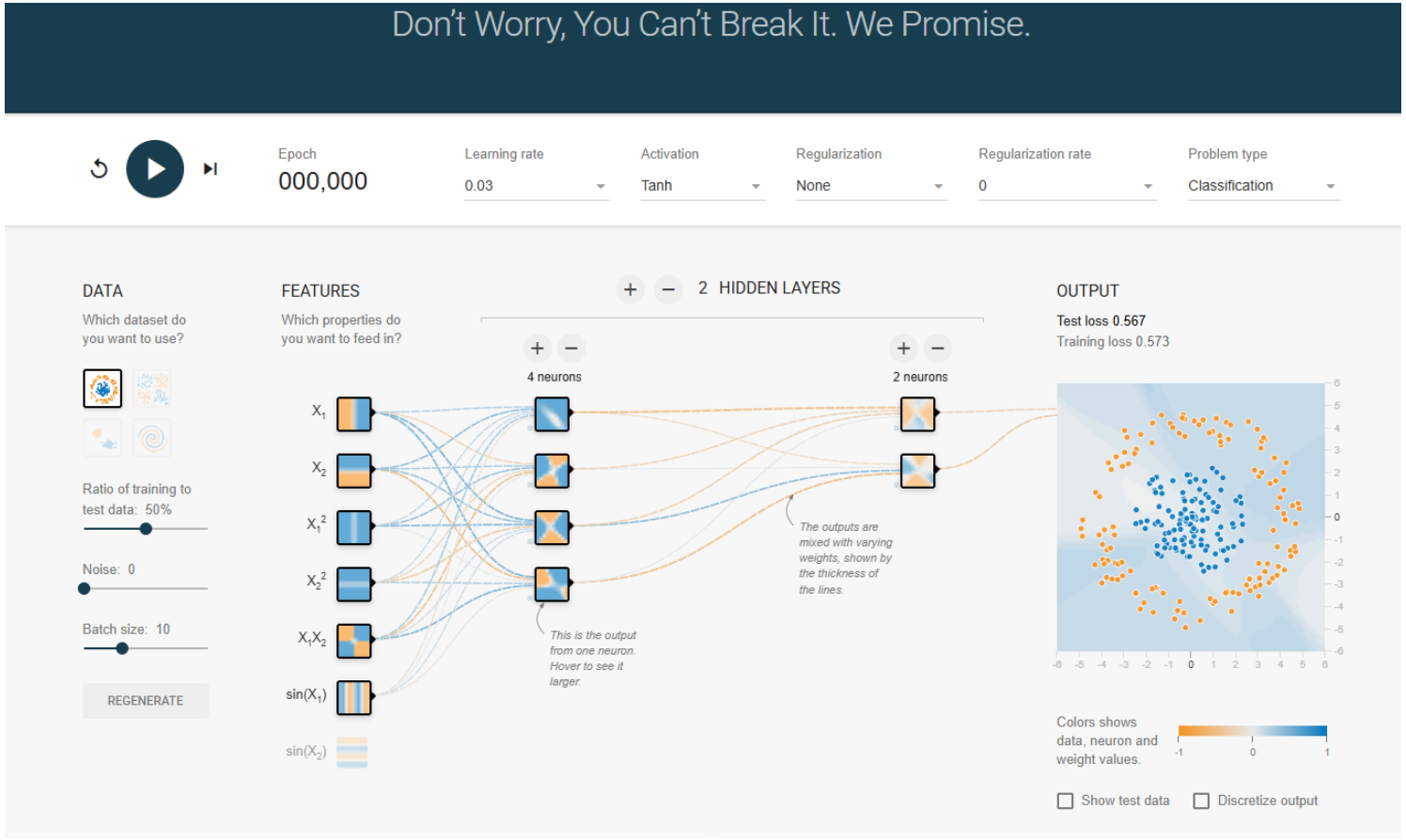



**Figure 1.6:** *Architecture of a Generative Adversarial Network*

# THE DEEPFAKE DEEP DIVE:

- **If you want to learn more and play with a neural network visit the Tensorflow Playground:**




https://playground.tensorflow.org/

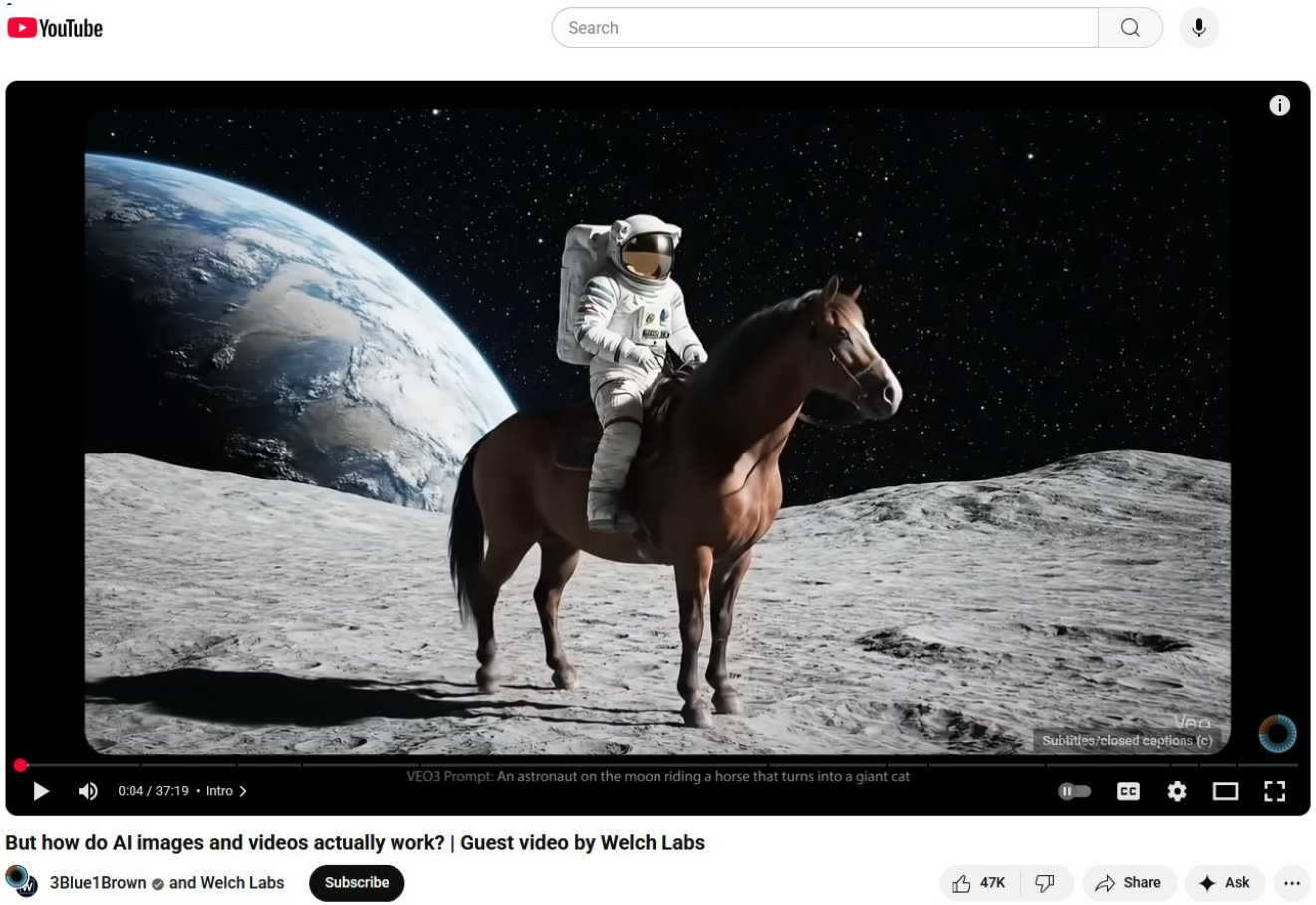
# THE DEEPFAKE DEEP DIVE:

**Books on Deepfakes:**

- Ultimate Deepfake Detection Using Python by Dr. Nimrita Koul
- Exploring Deepfakes by Bryan Lyon and Matt Tora (This has an excellent guide on how to make them.)
- FAIK by Perry Carpenter

# THE DEEPFAKE DEEP DIVE:

## YouTube Videos on Deepfakes:




“But How Do AI Images and Videos Actually Work?”

https://www.youtube.com/watch?v=iv-5mZ_9CPY

# THE DEEPFAKE DEEP DIVE:

**YouTube Channels Specifically on AI Generated Video:**

- MattVidPro AI
- AI Search
- Bob Doyle Media
- Theorectically Media
- Curious Refuge

# THE DEEPFAKE DEEP DIVE:

**AI Trainings I have taken:**

- “Designing and Building AI Products and Services” from MIT XPro
- “Post Graduate Program in AI and Machine Learning” from Purdue University and Simplilearn, this goes deep into the rabbit hole
- “Oxford Artificial Intelligence Program” from Oxford University and getsmarter.com

# THE DEEPFAKE DEEP DIVE:

A copy of this presentation will be given to LEVA for download.

Also,  I have included my AI terminology dictionary which I generated using Claude AI. (Unfortunately, Oxford did not give me permission to distribute their entire AI dictionary so I can't hand it out.)

Finally, I will supply a list of Deepfake generating tools and a list of free AI training certifications.


# QUESTIONS?

COM.OHIO.GOV

Contact: bradley.barkhurst@com.ohio.gov

State Fire Marshal

Department of Commerce