IEEE-CNSV

Consultants' Network of Silicon Valley

Deepfakes: Risks and Opportunities

Benjamin Mencer: 10 October 2023

What is a Deepfake?

Real or Fake?

Real or Fake?

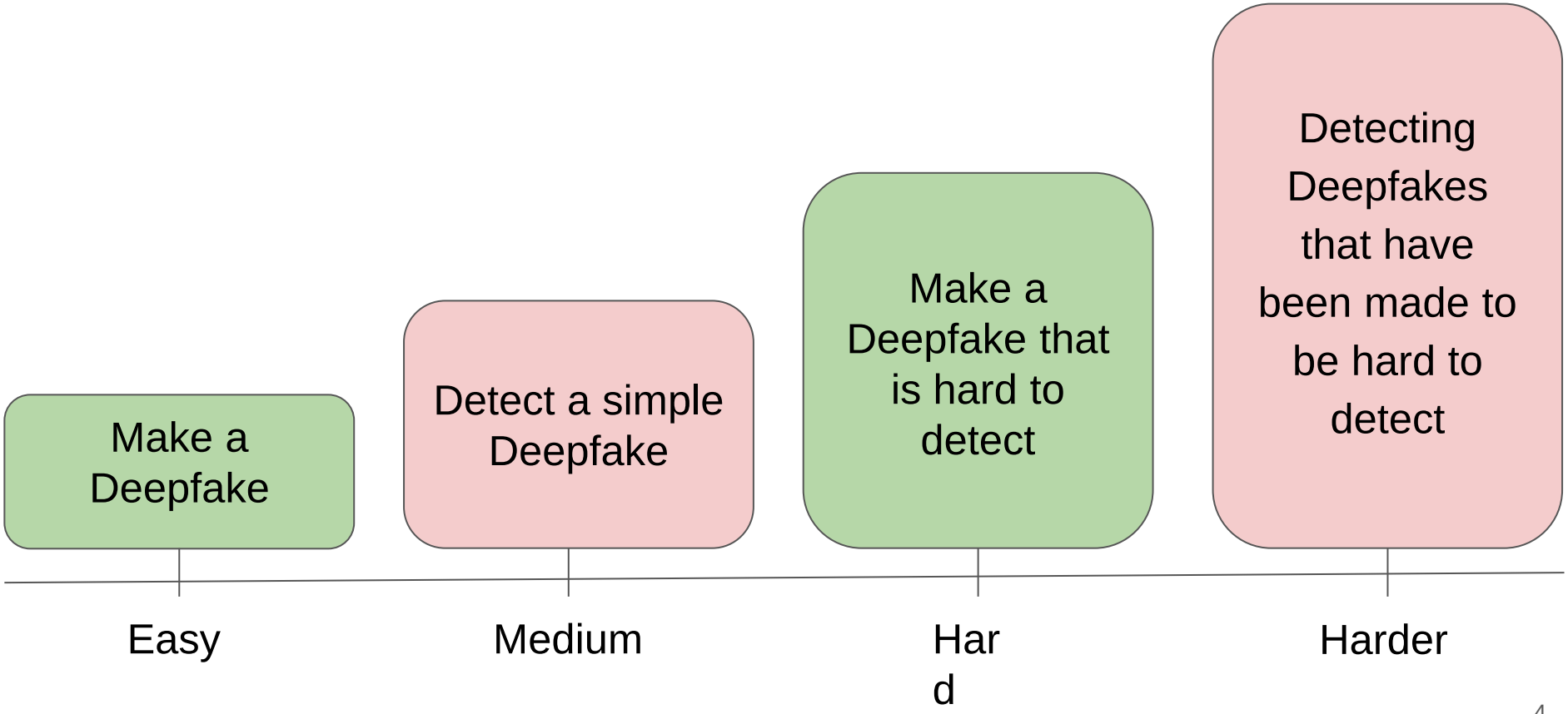
Real or Fake?

A Deepfake is a photo, video or audio file that has been digitally altered or created to misrepresent reality

History of Deepfakes

Deepfakes were first used in films and now we can see them everywhere.

Making Deepfakes



NY Company Detecting Deepfakes: Reality Defender!

Poul Carlson

www.realitydefender.ai

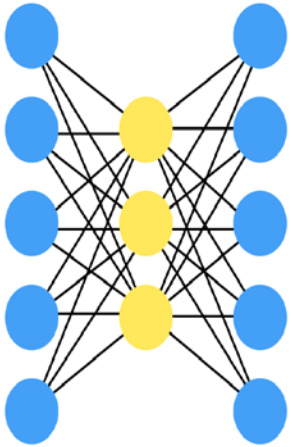
Startup, 2021

Deepfake detection is important as Deepfakes get better and better

Benjamin Mencer: 10 October 2023

Deepfake Market

Two methods to create Deepfakes



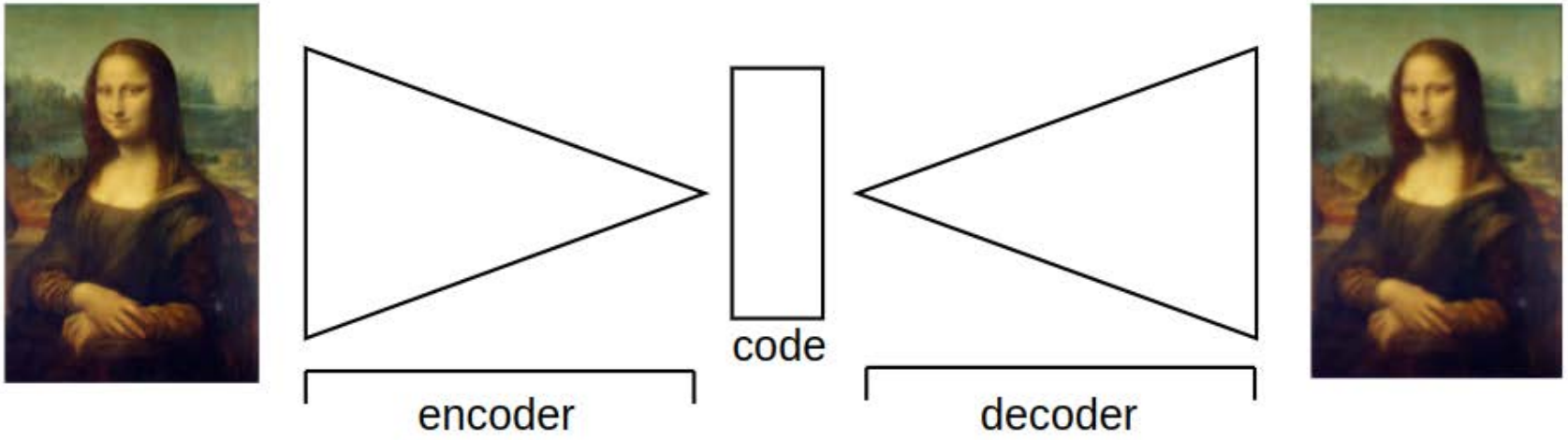
Autoencoders

Renato Candido

Generative Adversarial Networks (GANs)

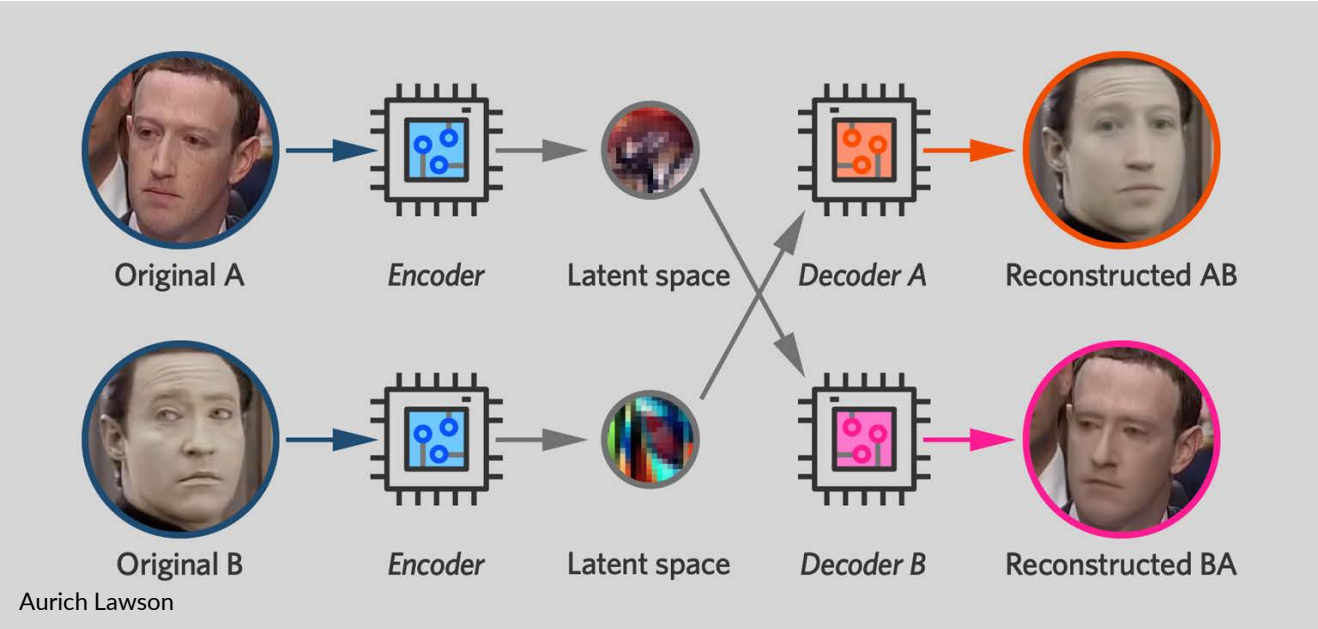
Examples of Autoencoded Deepfakes

Autoencoder Basic Model



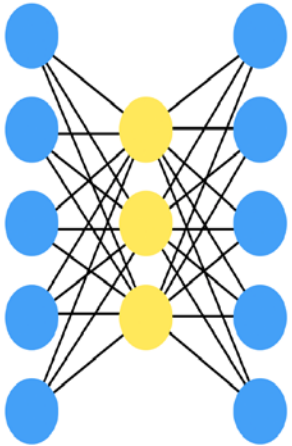
Compress media to its principle components

Making a Deepfake using Autoencoders



Mix and match the latent space with a decoder

Two methods to create Deepfakes



Autoencoders

Renato Candido

Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs)

These are all fake!

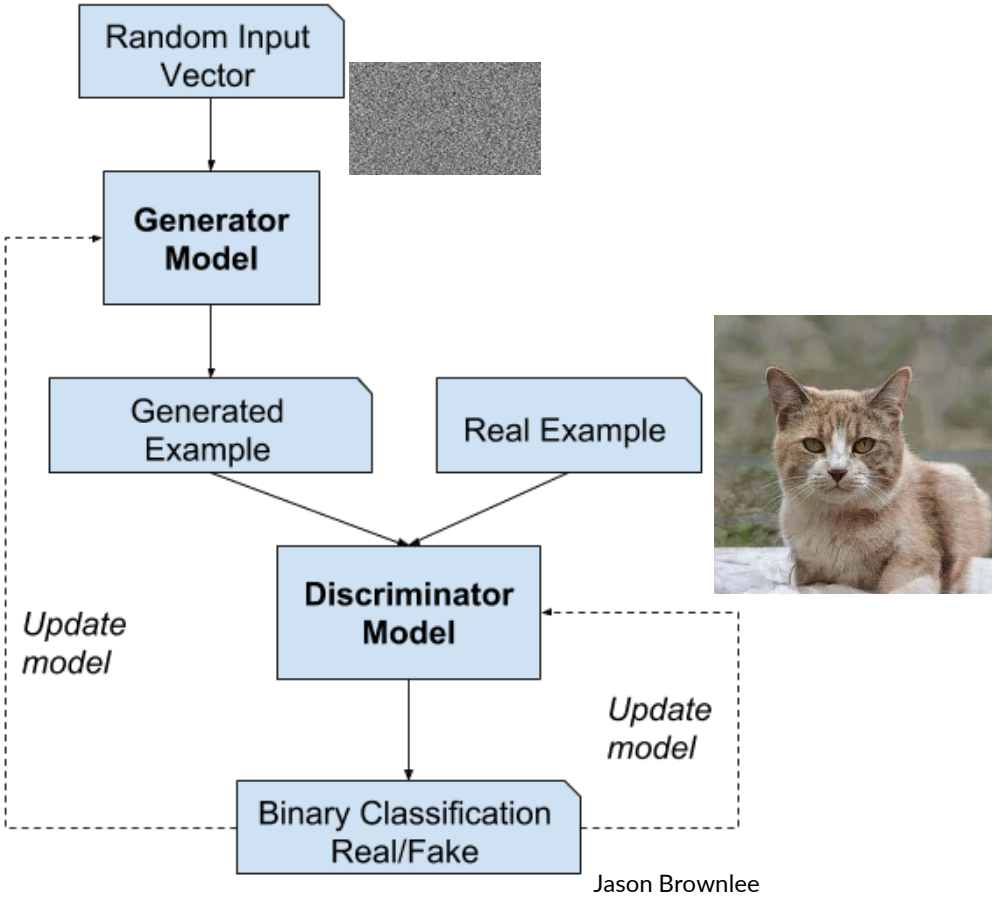
GAN images improving over time

Brundage et al

Now, only computers can tell the difference between real and fake

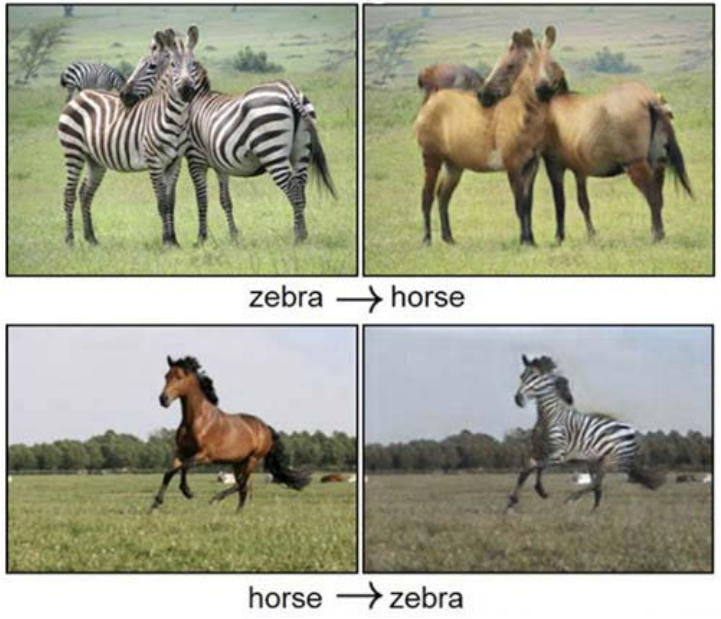
GAN Models

Generator Discriminator



GANs have two Neural Networks fighting each other

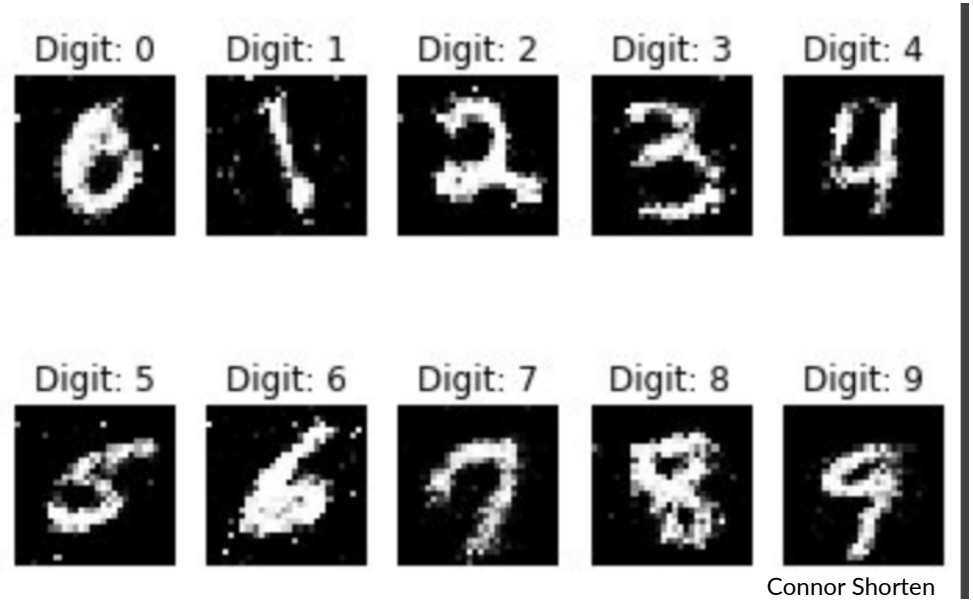
Special GANs: Cycle GAN

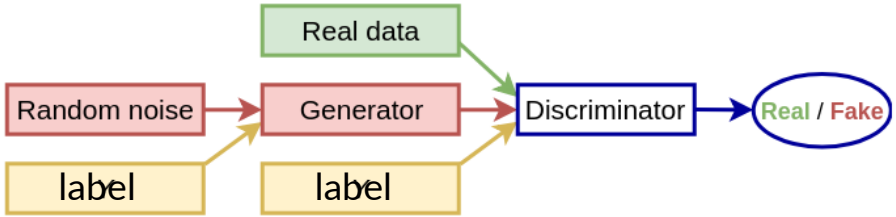


Laurence Miao

Cycle GANs allow change of the style of the image/audio/video

Special GANs: Conditional GAN





Conditional GANs create specific Deepfake based on labels

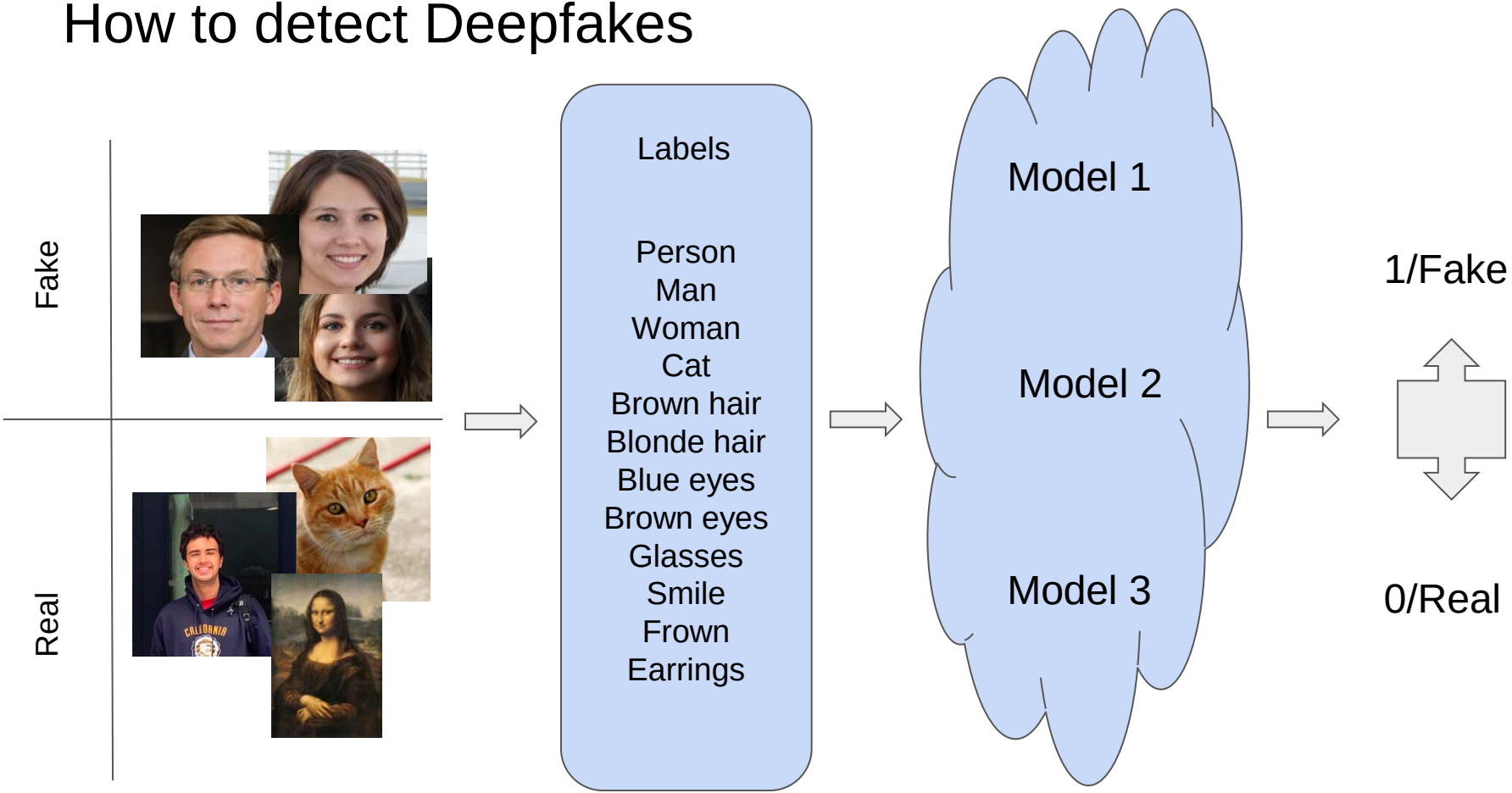
Detecting Deepfakes: Reality Defender!

www.realitydefender.ai

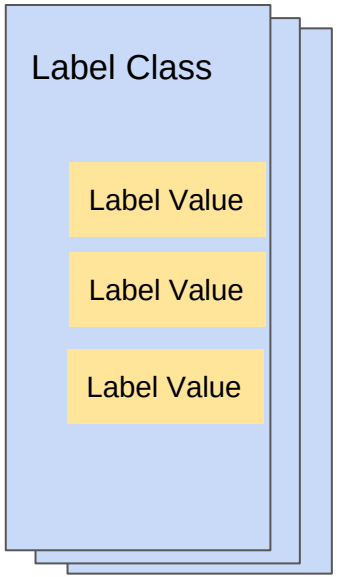
Startup, 2021

By marrying present and future technologies, Reality Defender offers clients continuous security from the minds of the world’s top machine learning and computer vision research teams.

How to detect Deepfakes

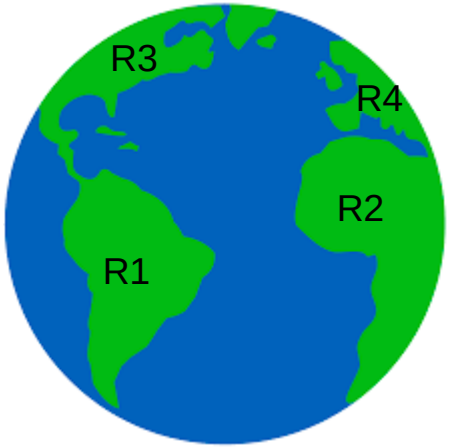


So, what is the problem?



The Data Bias Problem

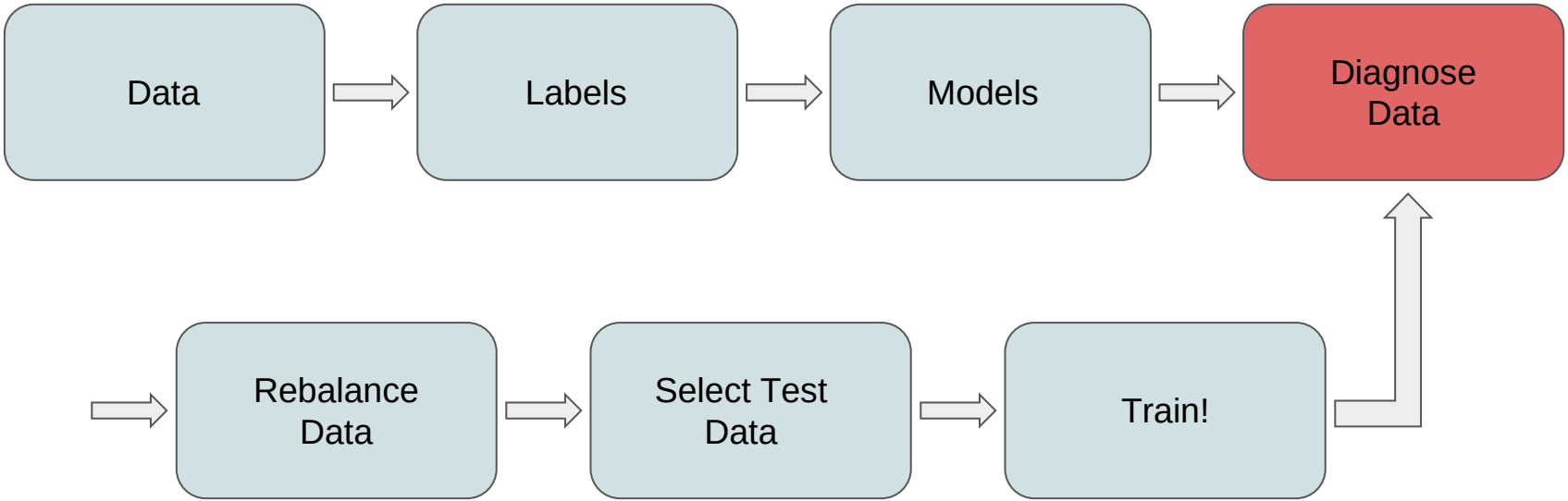
Overlap
Subset
Independence
Bias
Representative



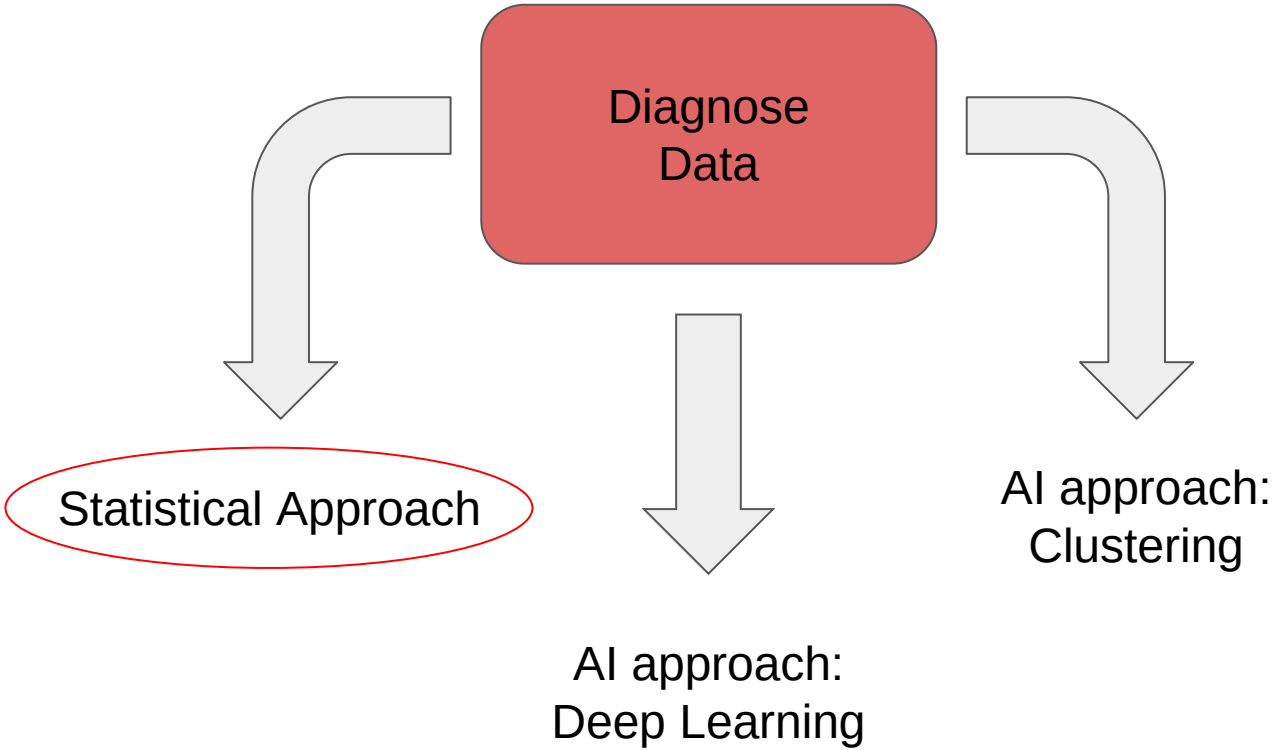
Representative
Label Values

“Limits of my Tags ~~language~~ means the limits of my world”
Ludwig Wittgenstein

Process of Deepfake Detection using AI



Diagnosing the Data



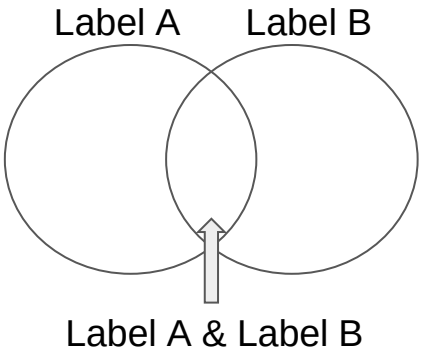
Diagnosing the Data: Statistical Approach

$$\frac{Label\ 1\ \&\ Label\ 3}{Label\ 3}$$

Overlap Matrix

	Label 1	Label 2	Label 3
Label 1	n/a	0.45	0.89
Label 2	1.00	n/a	0.21
Label 3	0.34	0.77	n/a

Label 2 ⊆ Label 1



An overlap matrix identifies major issues regarding the usefulness of labels

Diagnosing the Data: Statistical Approach

Overlap

- Value close to 1
- Simple Bias Detection
- Requires undersampling

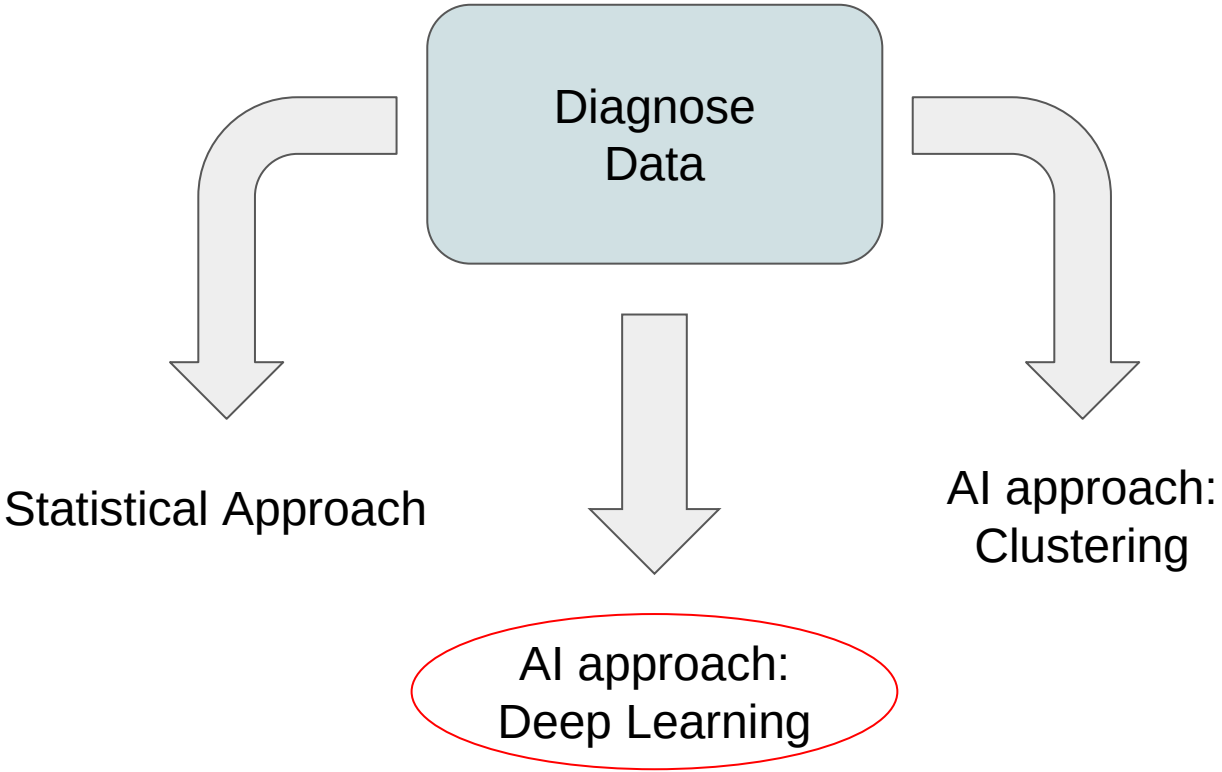
Subset

- Value == 1
- Extreme bias
- Requires re-tagging or new data

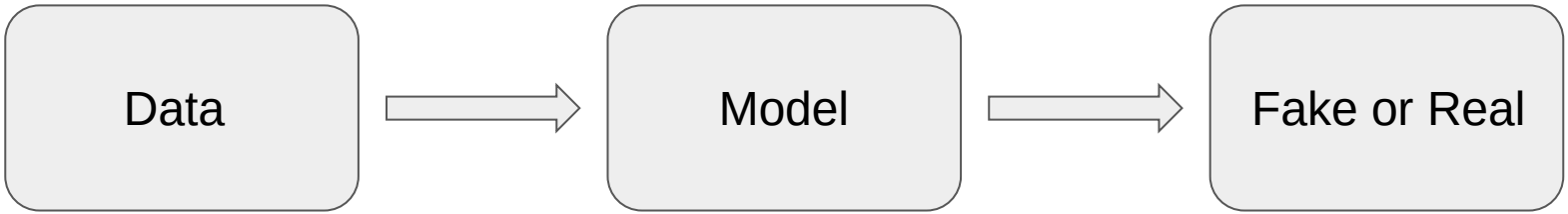
Representative

- Data must represent reality
- Required bias, portrayed through overlapping

Diagnosing the Data

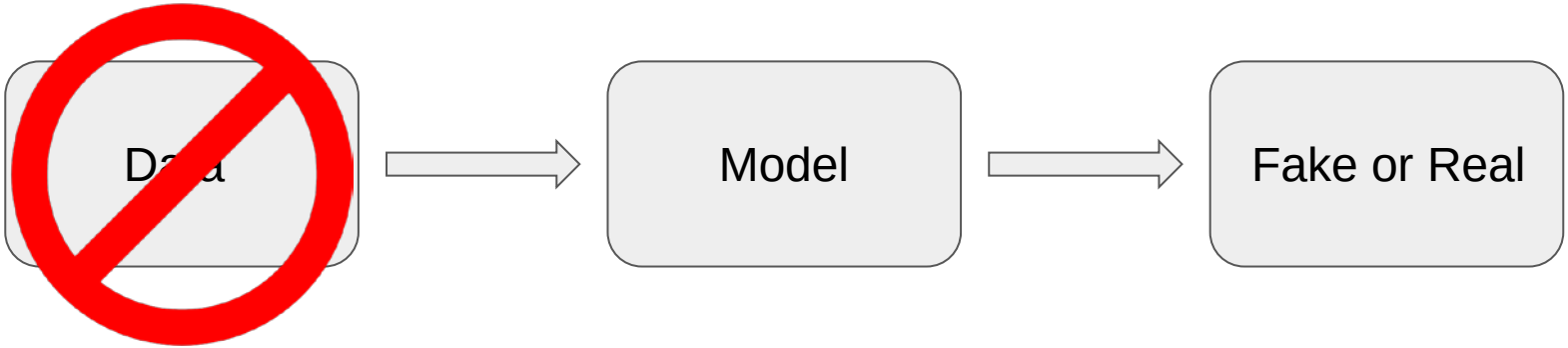


Diagnosing the Data: Deep Learning



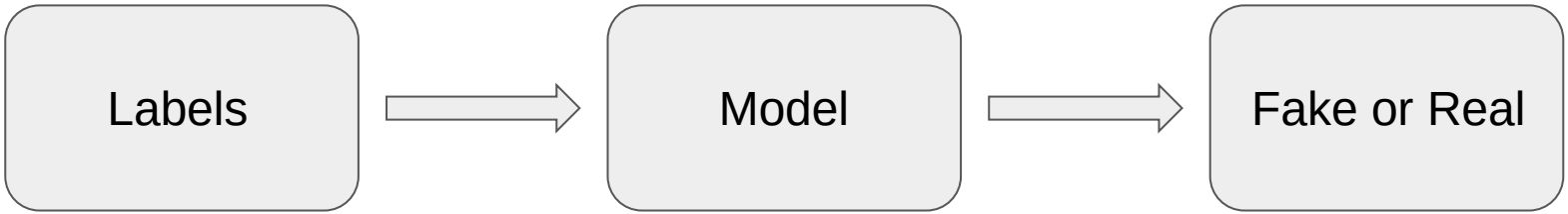
Use Deep Learning to find hidden bias between Data Classes in the Dataset

Diagnosing the Data: Deep Learning



Use Deep Learning to find hidden bias between Data Classes in the Dataset

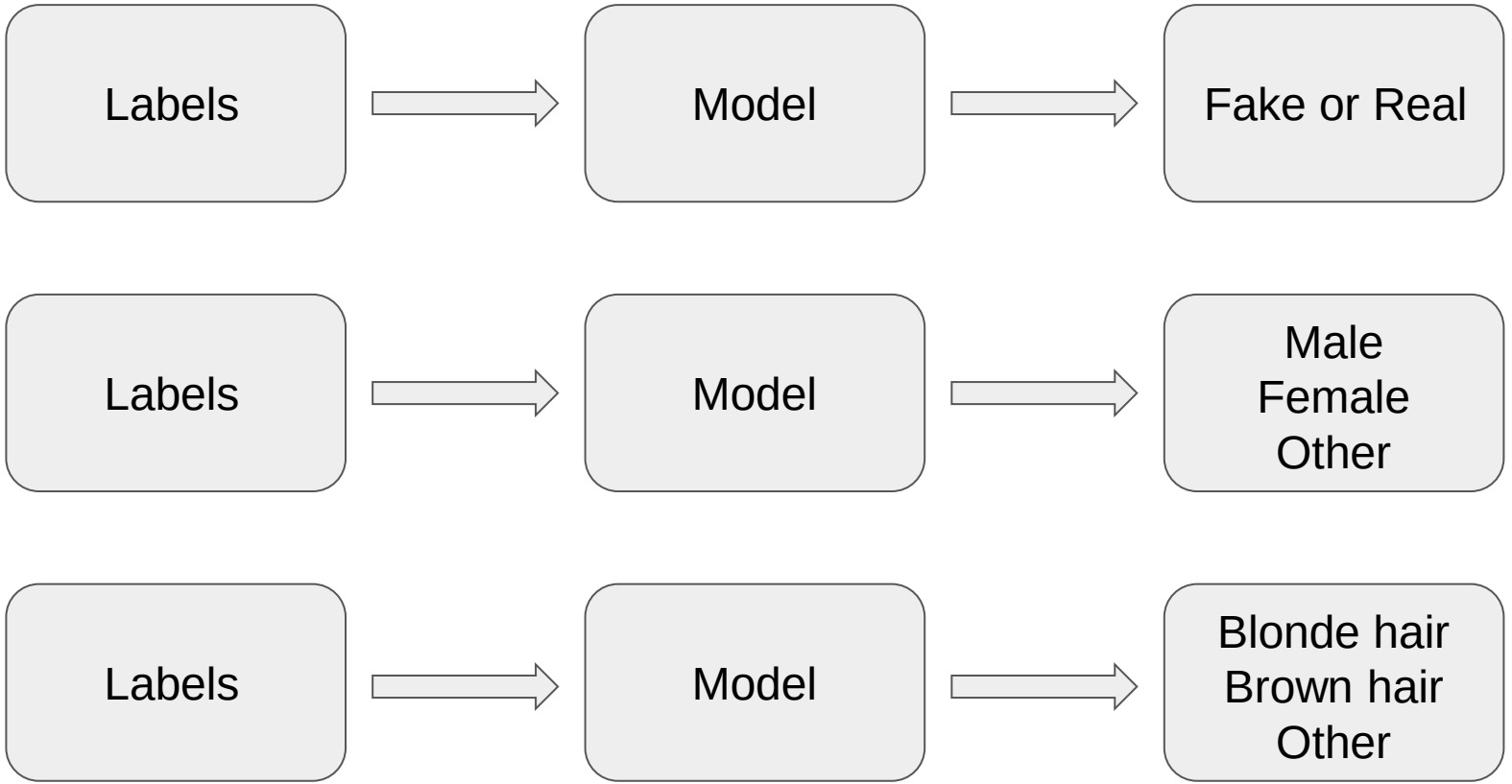
Diagnosing the Data: Deep Learning



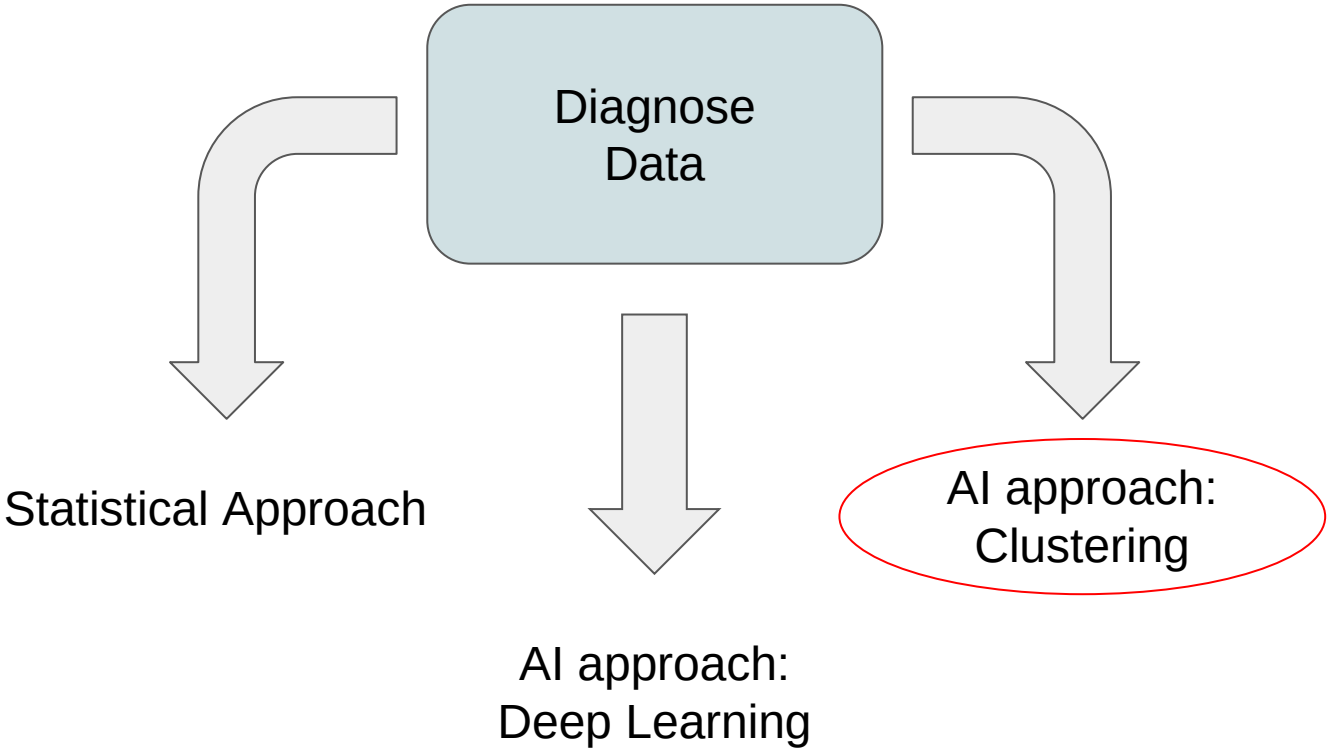
Fake or Real represents 1 Data Class (media identity). Repeat for all Data Classes

Use Deep Learning to find hidden bias between Data Classes in the Dataset

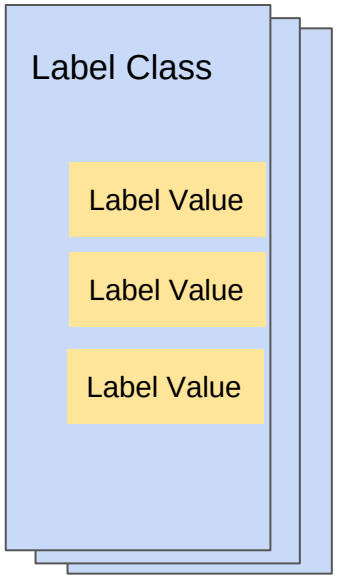
Diagnosing the Data: Deep Learning



Diagnosing the Data

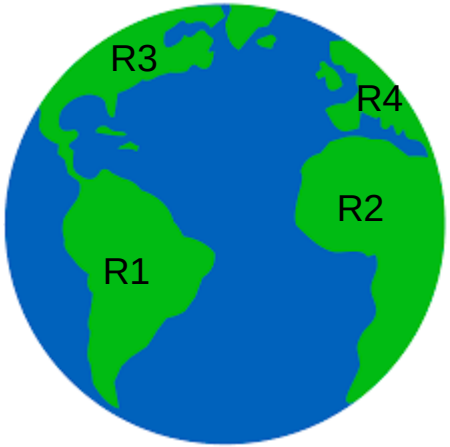


So, what is the problem?



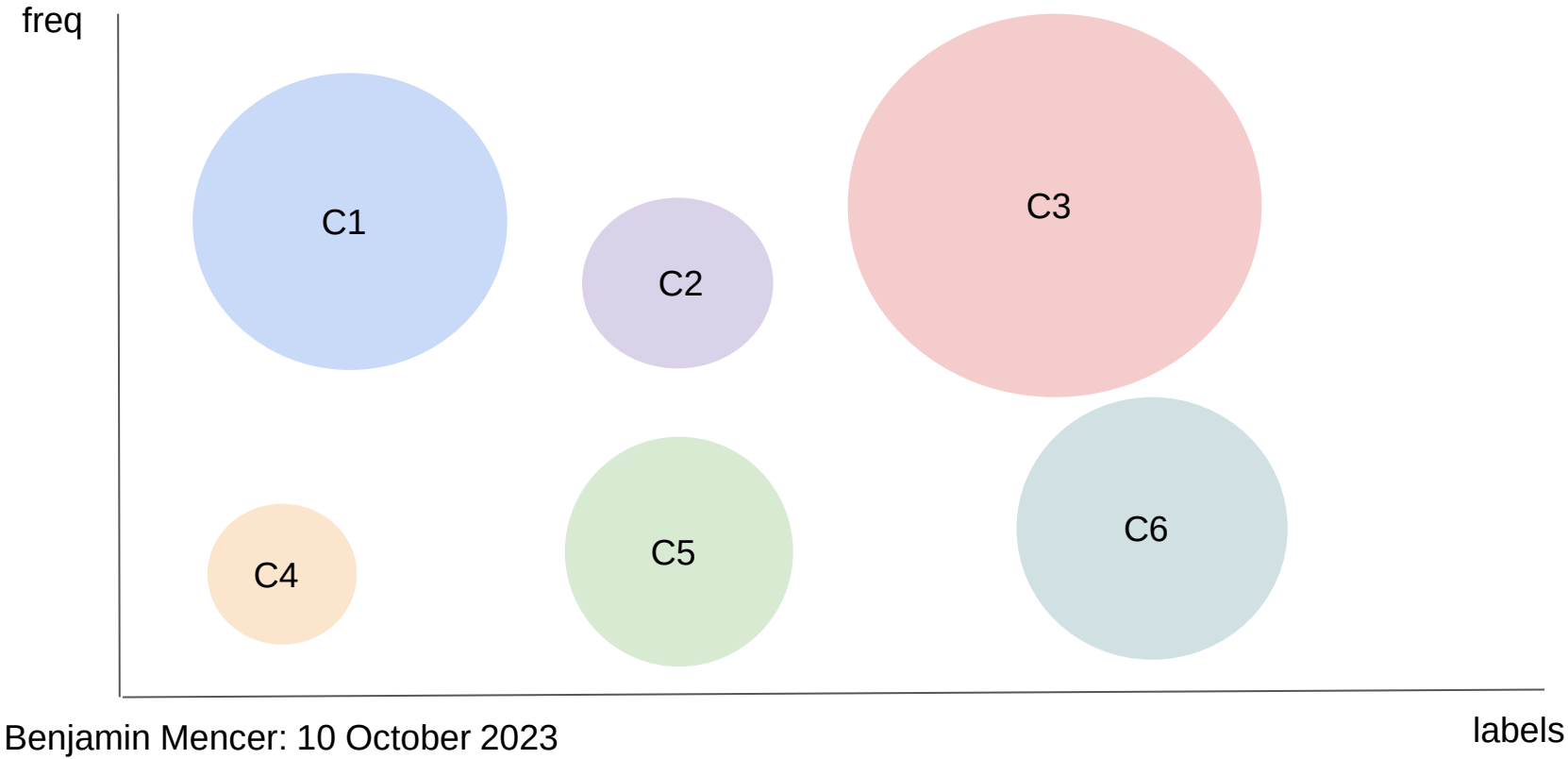
The Data Bias Problem

Overlap
Subset
Independence
Bias
Representative



Representative
Label Values

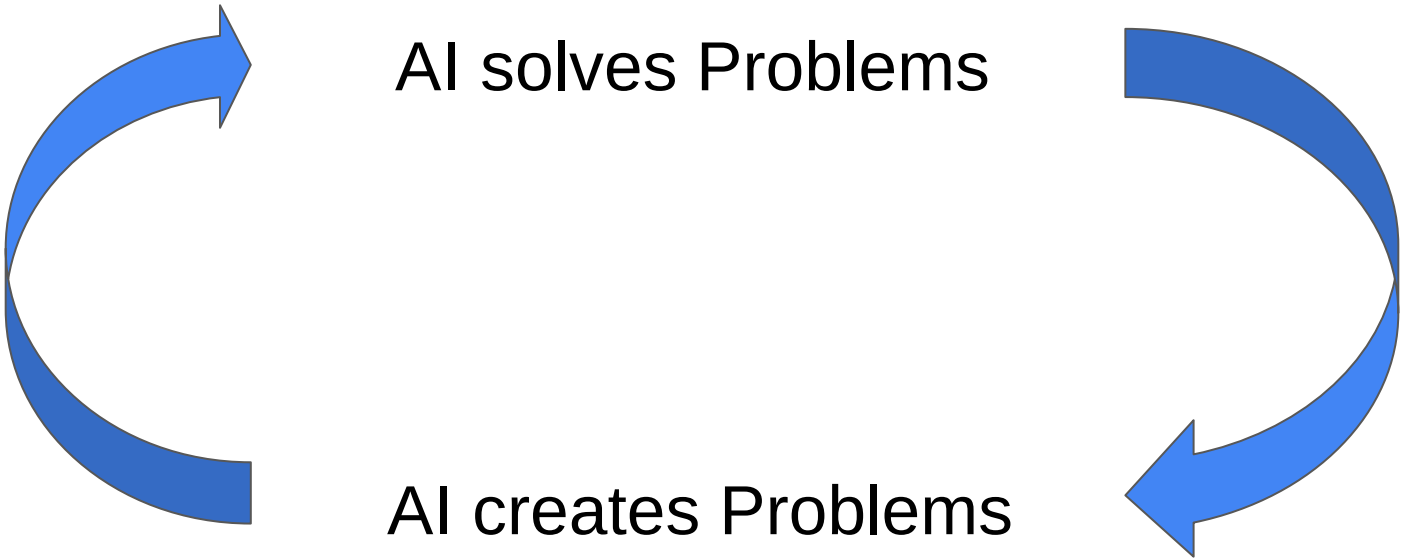
Diagnosing the Data: Clustering



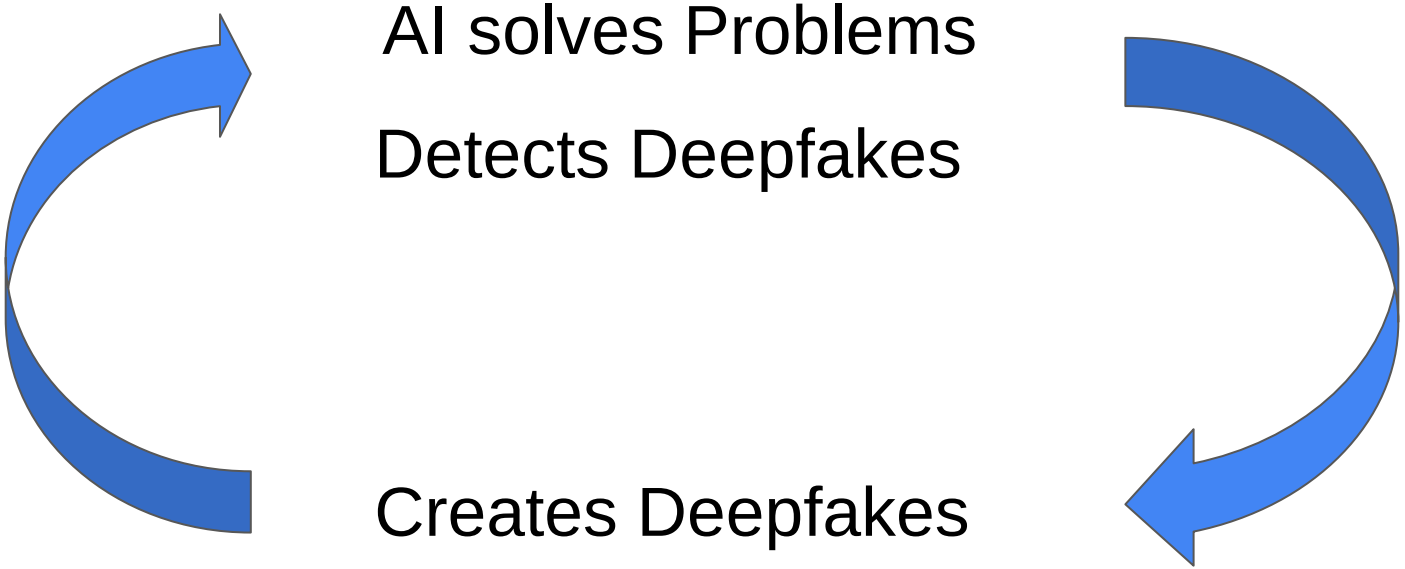
Benjamin Mencer: 10 October 2023

Clustering can help determine if the data is **representative**

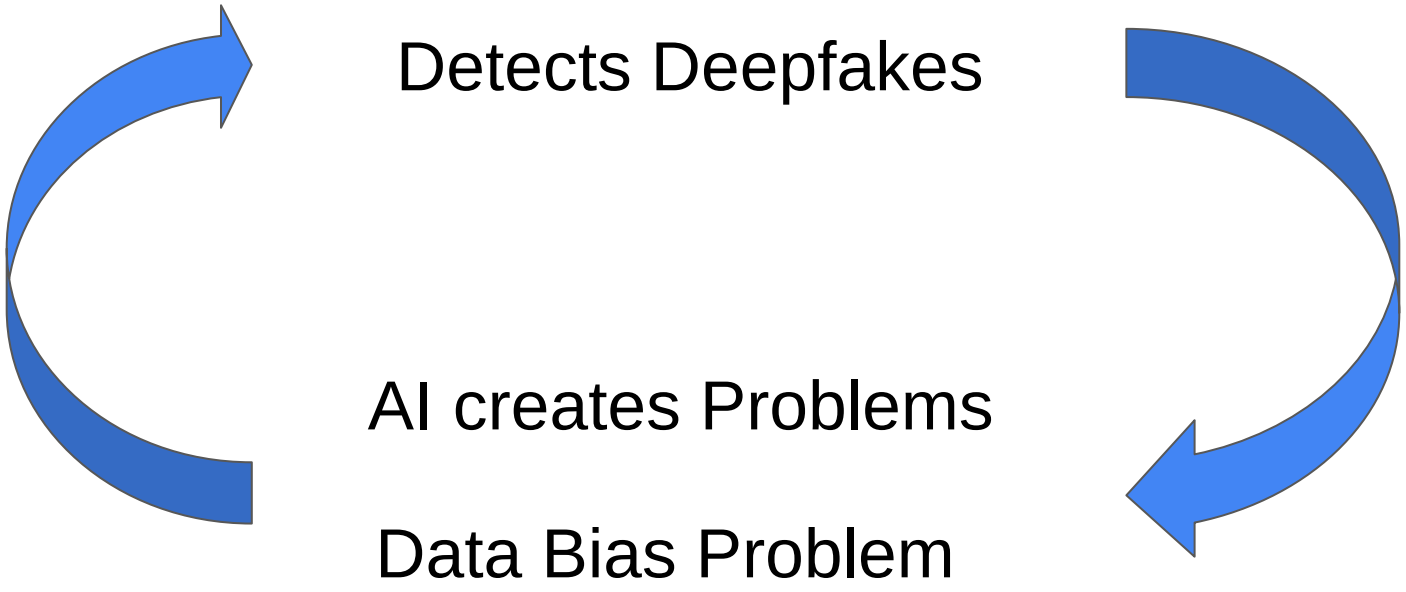
The Cycle of AI



The Cycle of AI



The Cycle of AI



The Future of Deepfake Competition

Deepfakes in Time of War

Would a soldier be compelled to stop fighting after listening to this video?

Benjamin Mencer: 10 October 2023

Deepfakes for good

Deepfakes can help us visualize the past and solve crime

From Reality to Consciousness: Federico Faggin

- AI projected to be USD 1,394.30 billion market by 2029
  Fortune Business Insights (Sept 2022)
- Federico Faggin argues that AI will never reach consciousness
- Do you think a machine will ever be conscious?

Book: “Silicon” by Federico Faggin, 2021

Acknowledgements

- Ava Soleimany, MIT
- Matthew Stewart, Harvard
- Ian J. Goodfellow, Stanford