Data Science

# Artificial Intelligence

**Tinglong Dai, PhD**
Bernard T. Ferrari Professor

Johns Hopkins Carey Business School

# Burning Question for AI Lab:
# What does “novelty” mean?

The test of *novelty* for an idea is not the absence of one single predecessor, but the presence of multiple but incompatible ones.

**Nassim Nicholas Taleb, *The Bed of Procrustes***

# Agenda

1 — **Week 1: Foundations**

2 — **Week 2: Tools & Techniques**

Machine learning, deep learning, Keras, Python, and no-code tools.

3 — **Week 3: CNNs & Business Models**

Convolutional neural networks, using AI to create human value. Assignment #1 due.

4 — **Week 4: MLOps & Evaluation**

MLOps, transformers, evaluation of AI. AI Lab proposal due.

5 — **Week 5: LLMs & GenAI**

Large language models, generative AI, using AI to create business value. Assignment #2 due.

6 — **Week 6: Human-AI Interaction**

Human-AI interaction, reinforcement learning, system design.

7 — **Week 7: Responsible AI**

Responsible AI, future of AI. Lab final deliverable due.

8 — **Week 8: Individual Final Exam; Peer Reviews of AI Lab**

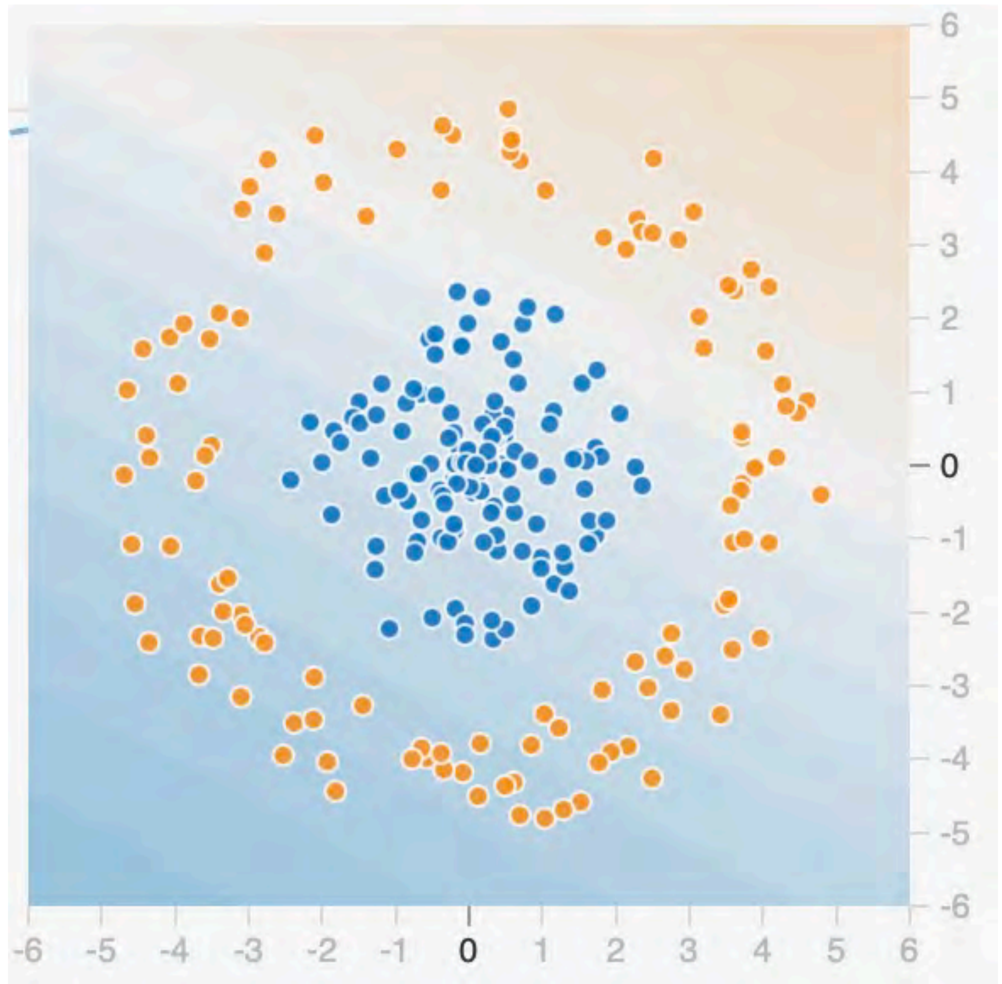
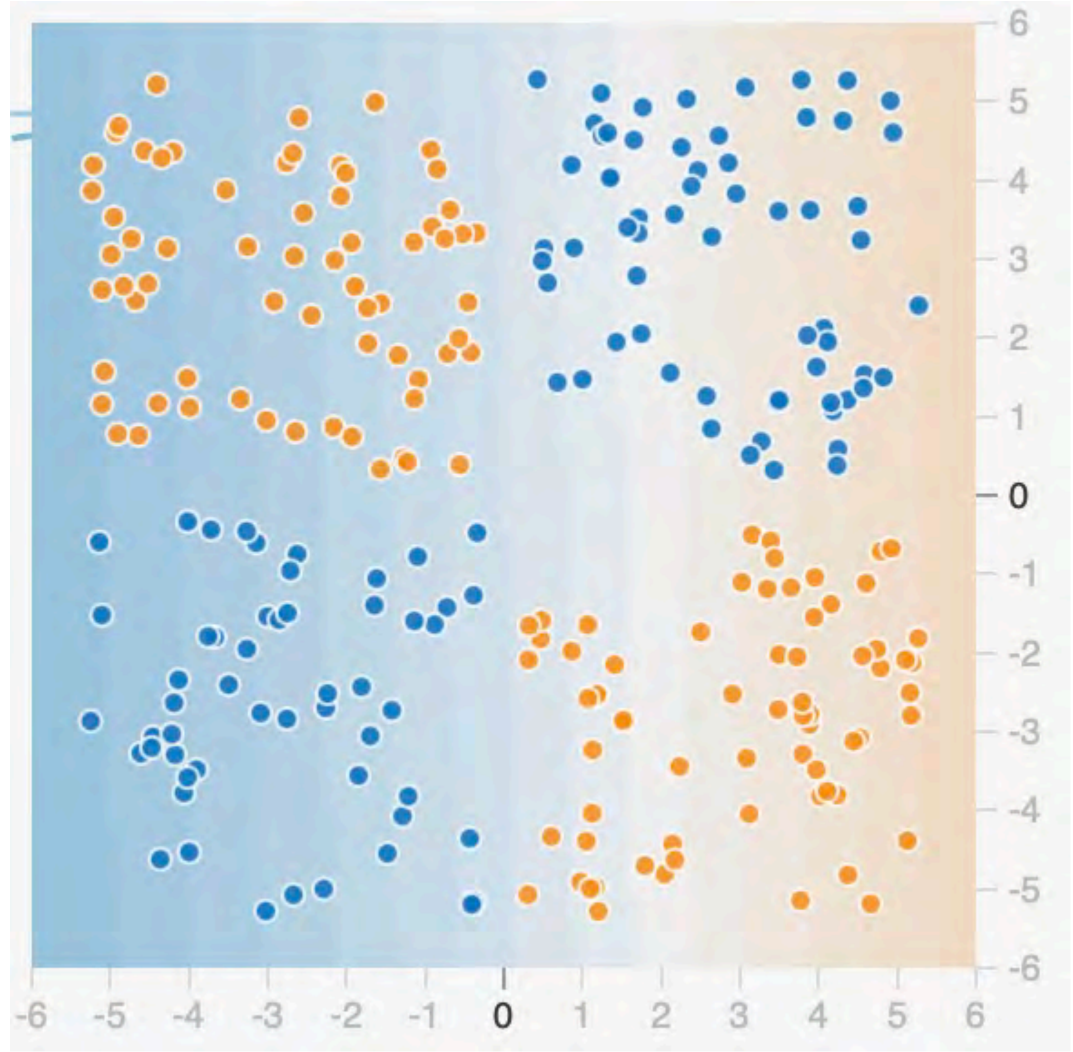
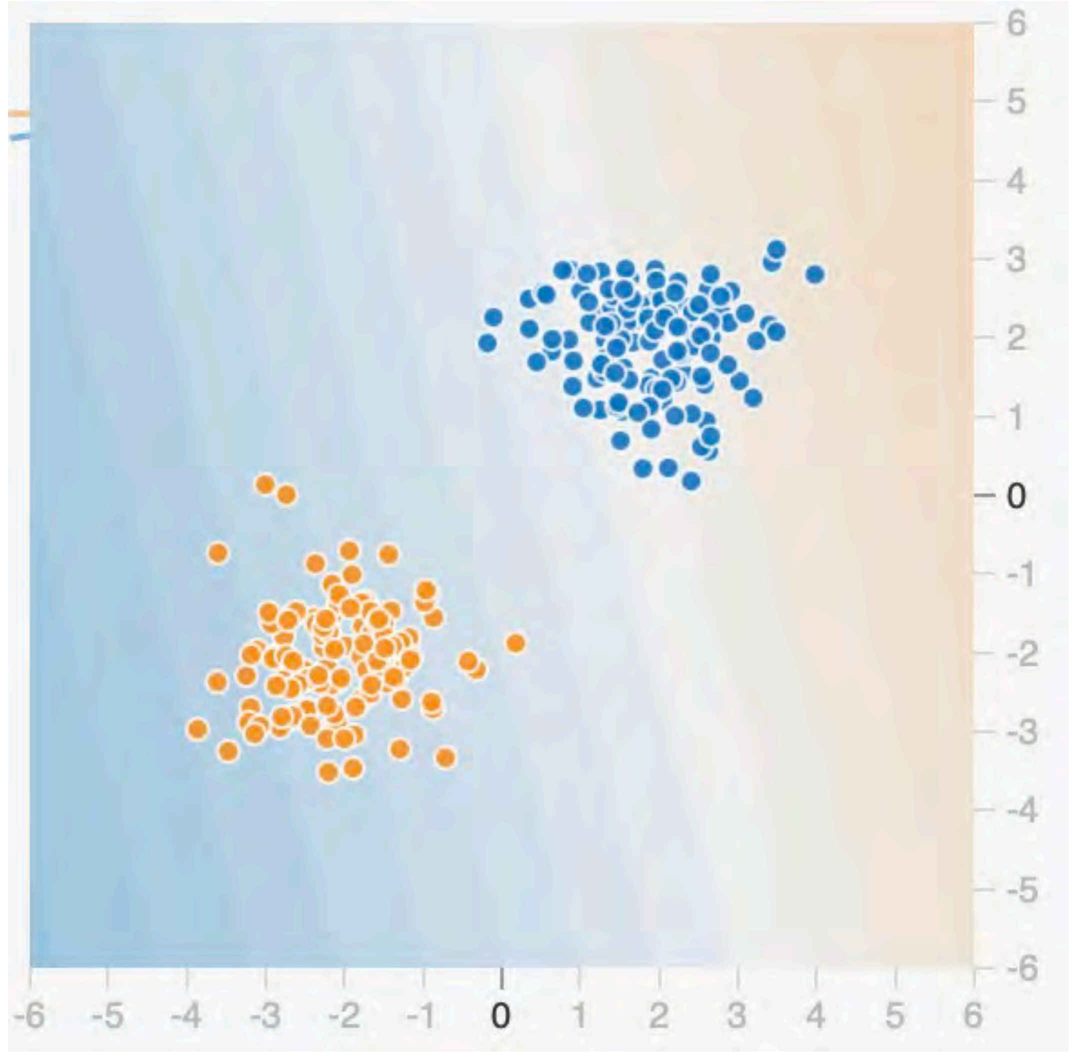
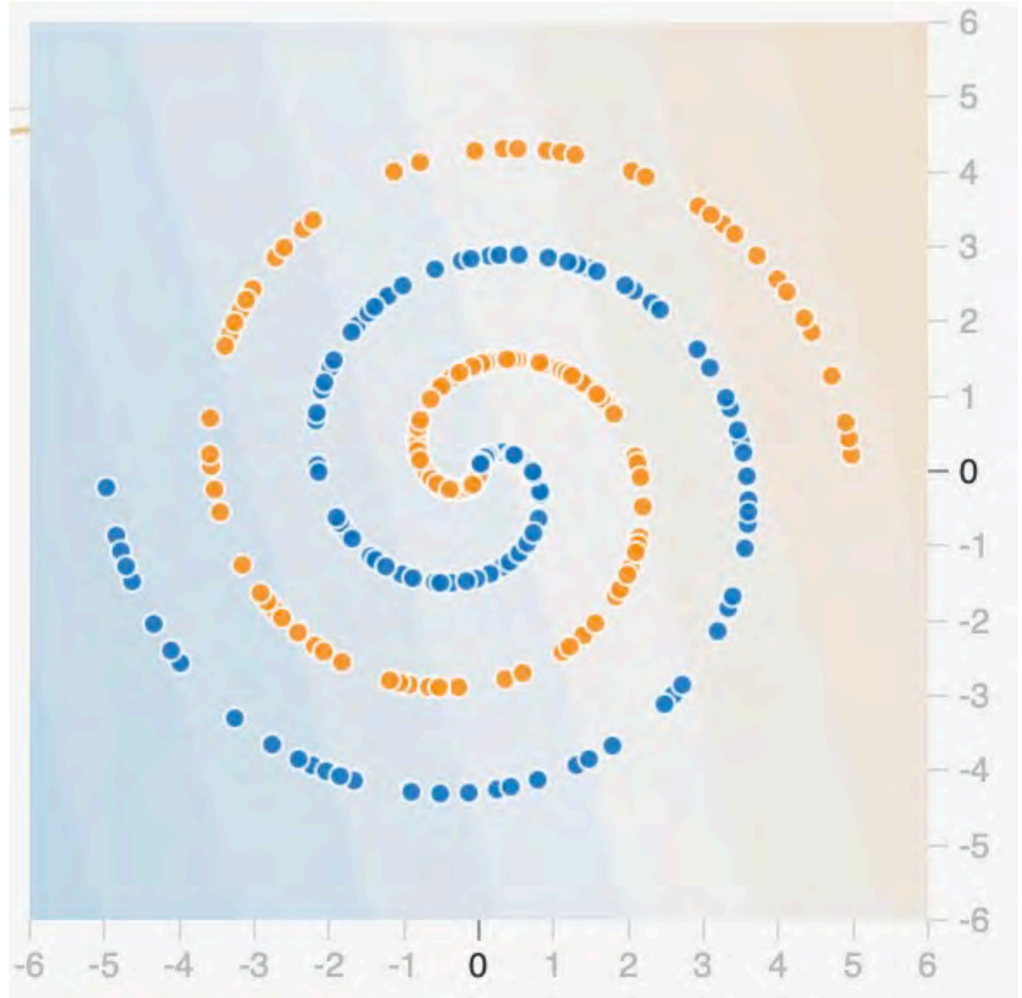
# Second TA Tutorial: Keras & CNNs

## 🗓️ Friday, 11/14, 12:00–1:00 PM

**This week, Suhas will post a recorded tutorial**

# Discussion of
# Group Assignment #1

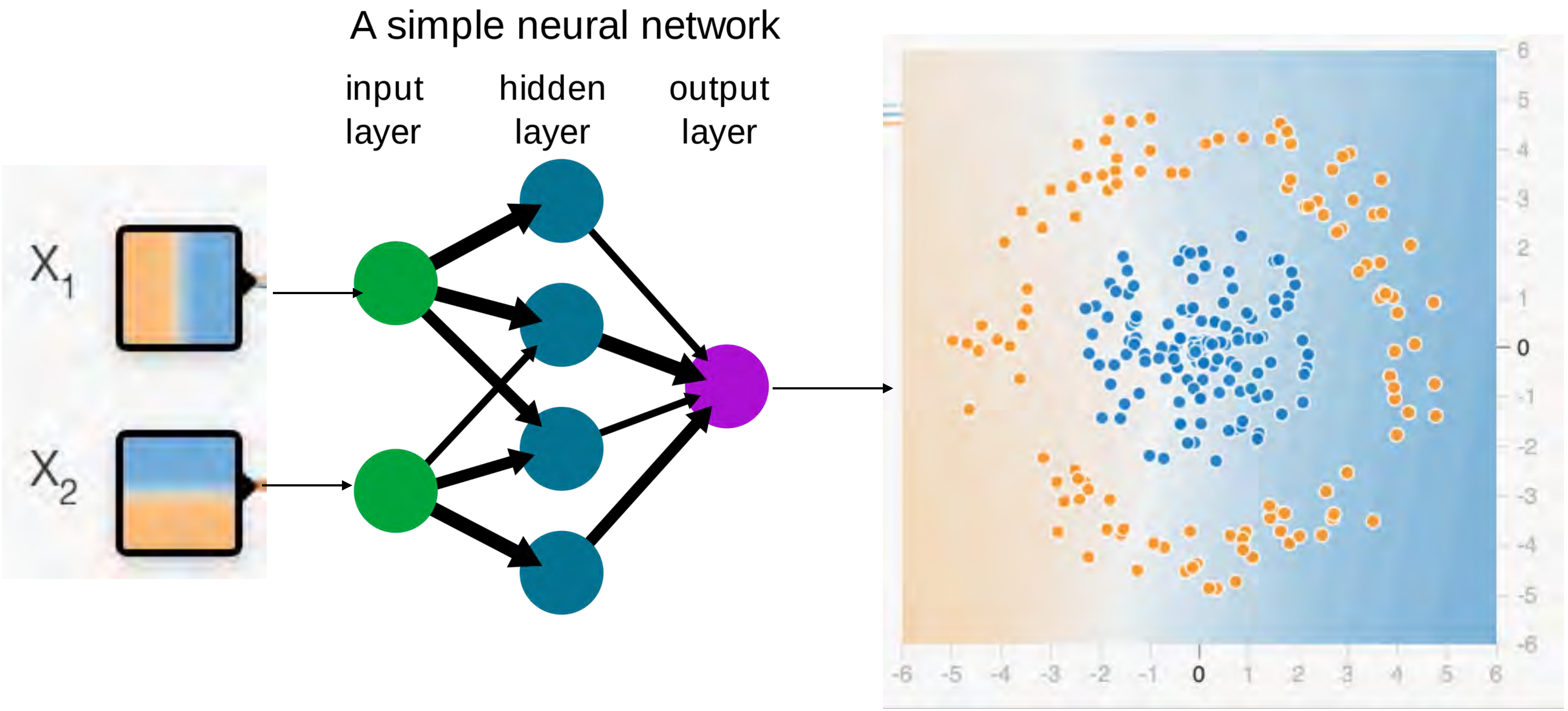
# Linearly Classifiable Dataset

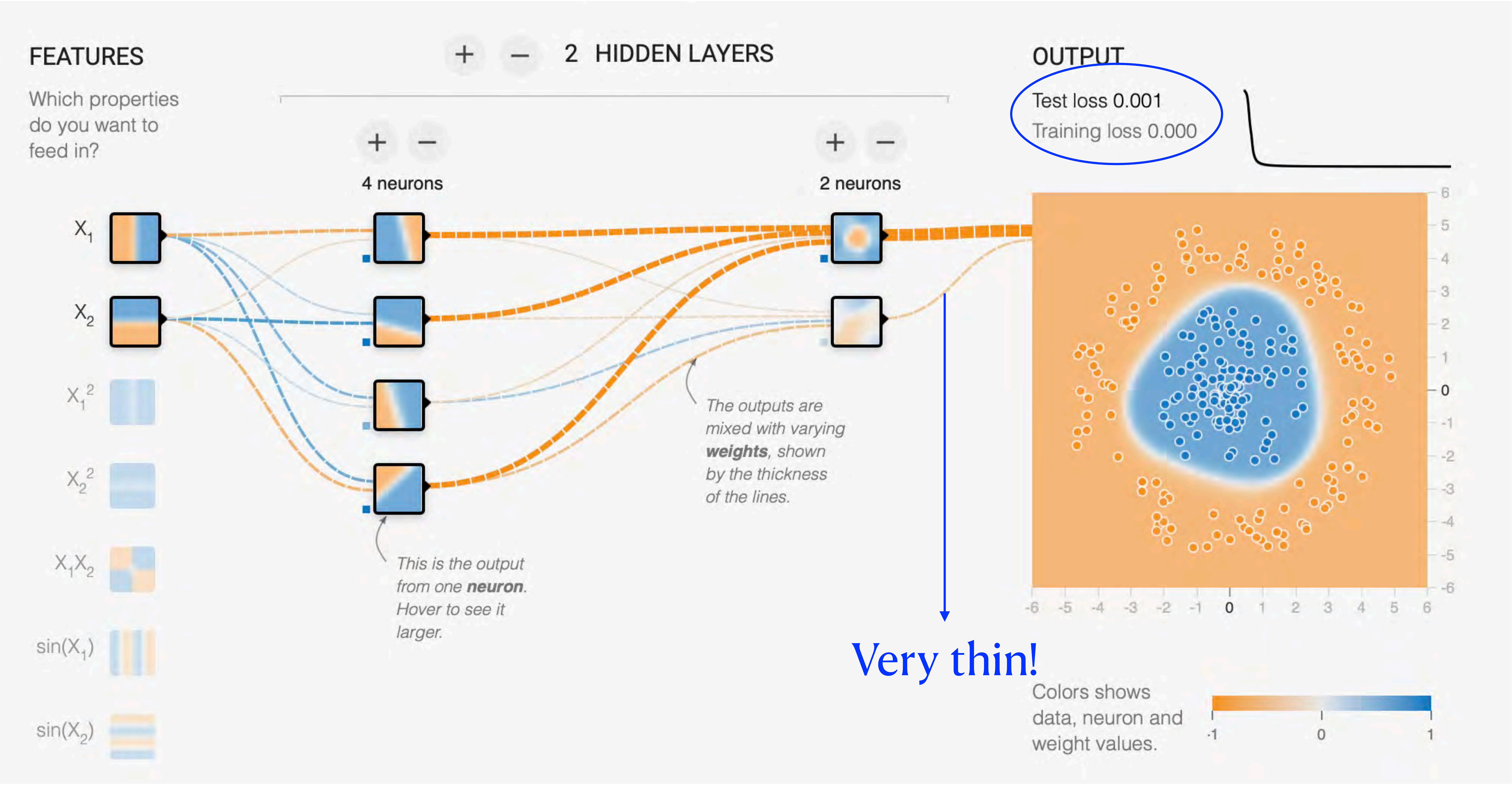
Definition: A pattern is **linearly classifiable** if there exists one line in the plane such that (1) all the blue dots are on one side of the line and (2) all the yellow dots are on the other side.

Only the third dataset is linearly classifiable; all the others are not linearly classifiable.

# Learning a Pattern Using a Feedforward Neural Network



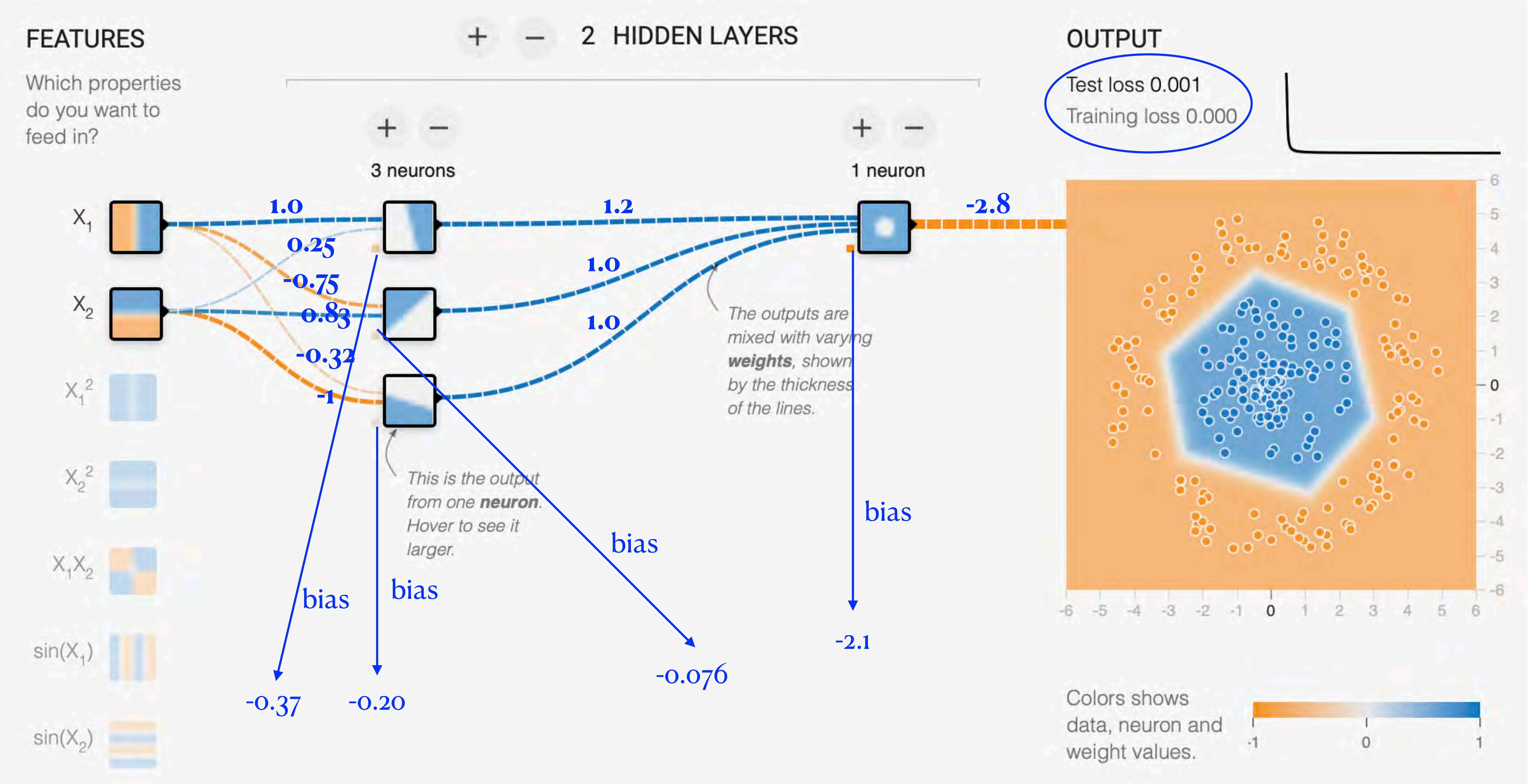
# Learning a Pattern Using a Feedforward Neural Network




Maybe we can use a single neuron for the second hidden layer....

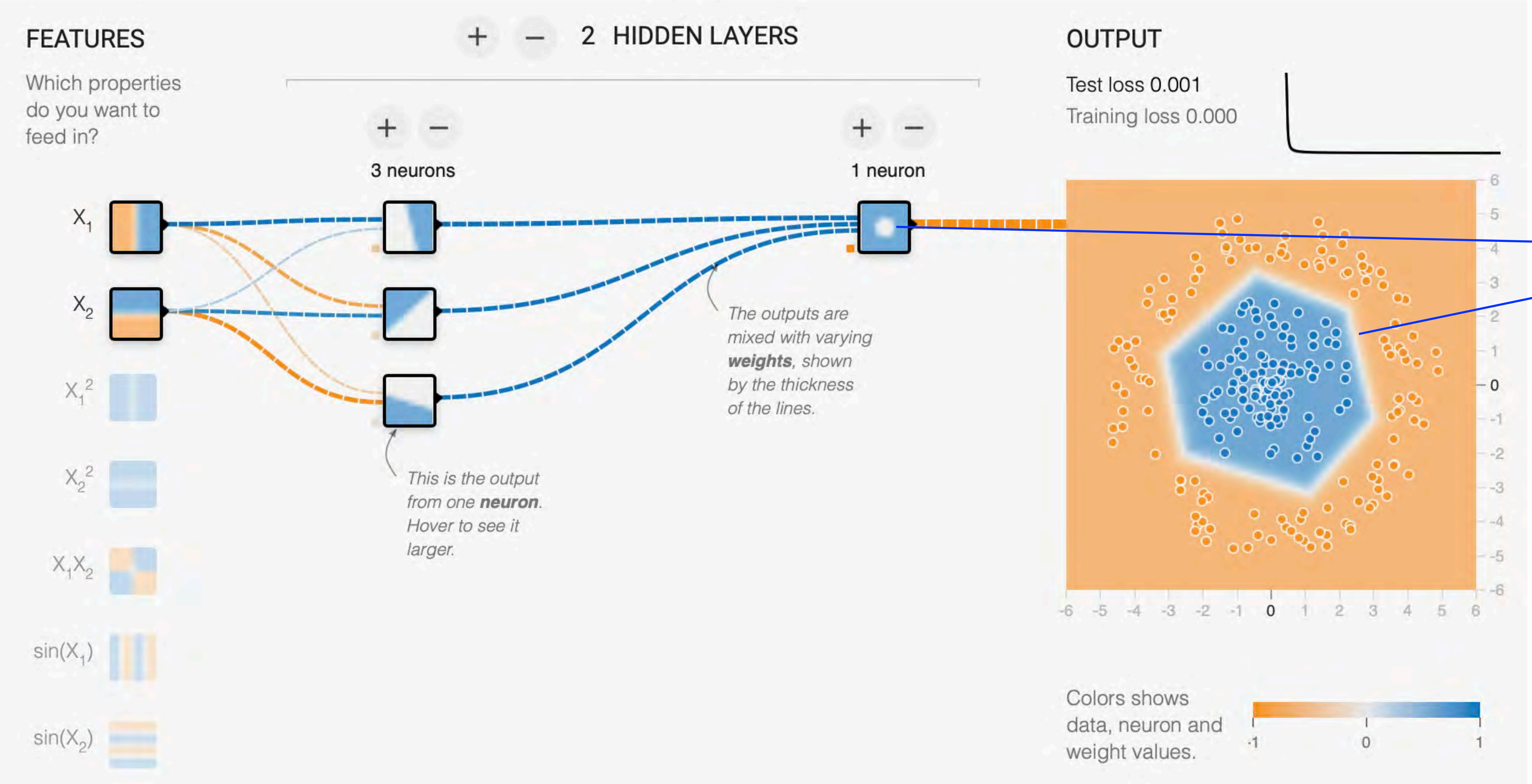
http://bit.ly/anntest

# Learning a Pattern Using a Feedforward Neural Network




$$\text{Output} = -2.8 \times \text{ReLU}(1.2 \times \text{ReLU}(x_1 + 0.25x_2 - 0.37) + \text{ReLU}(-0.75x_1 + 0.83x_2 - 0.076) + \text{ReLU}(-0.32x_1 - x_2 - 0.2) - 2.1)$$

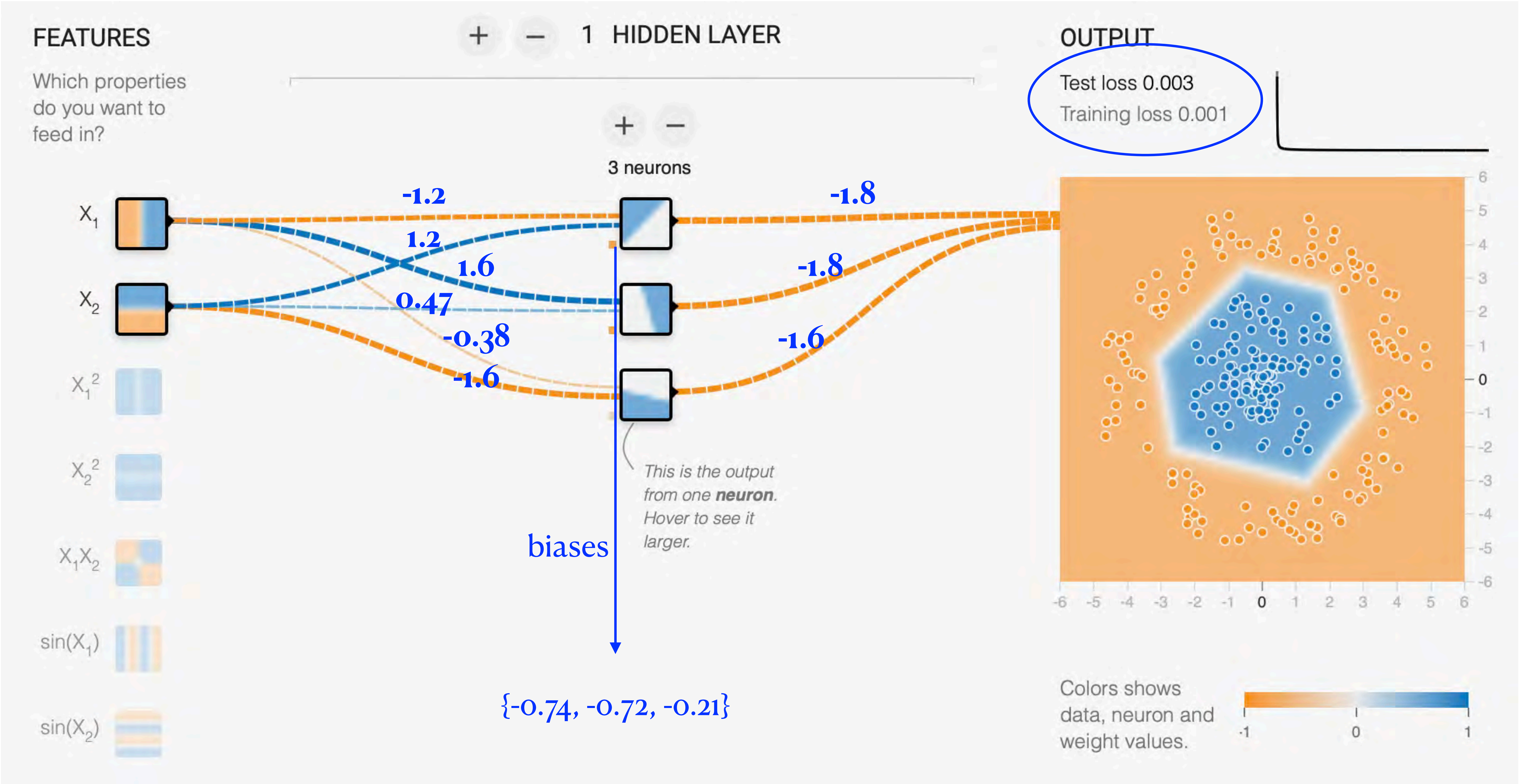
# Learning a Pattern Using a Feedforward Neural Network




Very similarly shaped!

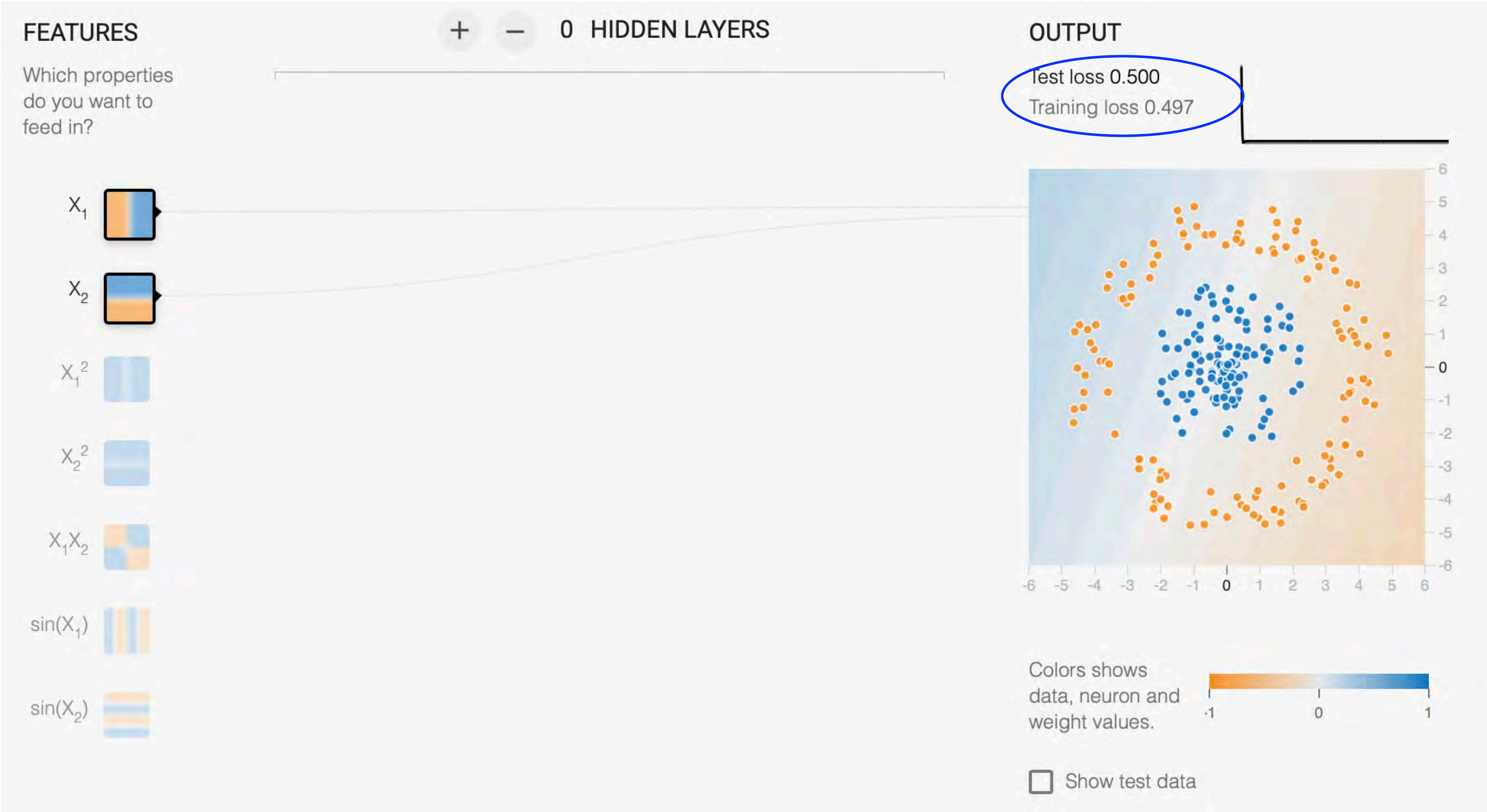
**What about using a single hidden layer?**

# Learning a Pattern Using a Feedforward Neural Network




$$\text{Output} = -1.8 \times \text{ReLU}(-1.2x_1 + 1.2x_2 - 0.74) - 1.8 \times \text{ReLU}(1.6x_1 + 0.47x_2 - 0.72) - 1.6 \times \text{ReLU}(-0.38x_1 - 1.6x_2 - 0.21)$$

# Are Hidden Layers Even Necessary?




Yes! Because we are dealing with a *non*-linearly separable dataset.

# The Convolution Operation

- The convolution operation = doing the following two steps iteratively:
  (1) sliding the filter through the input and (2) calculating the weighted sum

**Convoltional filter**

| | |
|---|---|
| 1 | 1 |
| 0 | 1 |

**Input Feature Map**

| | | | |
|---|---|---|---|
| ×0 | ×1 | 1 | 1 |
| ×0 | ×0 | 0 | 1 |
| 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 |

**Output Feature Map**

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 0 | 1 | 1 |
| 0 | 1 | 1 |

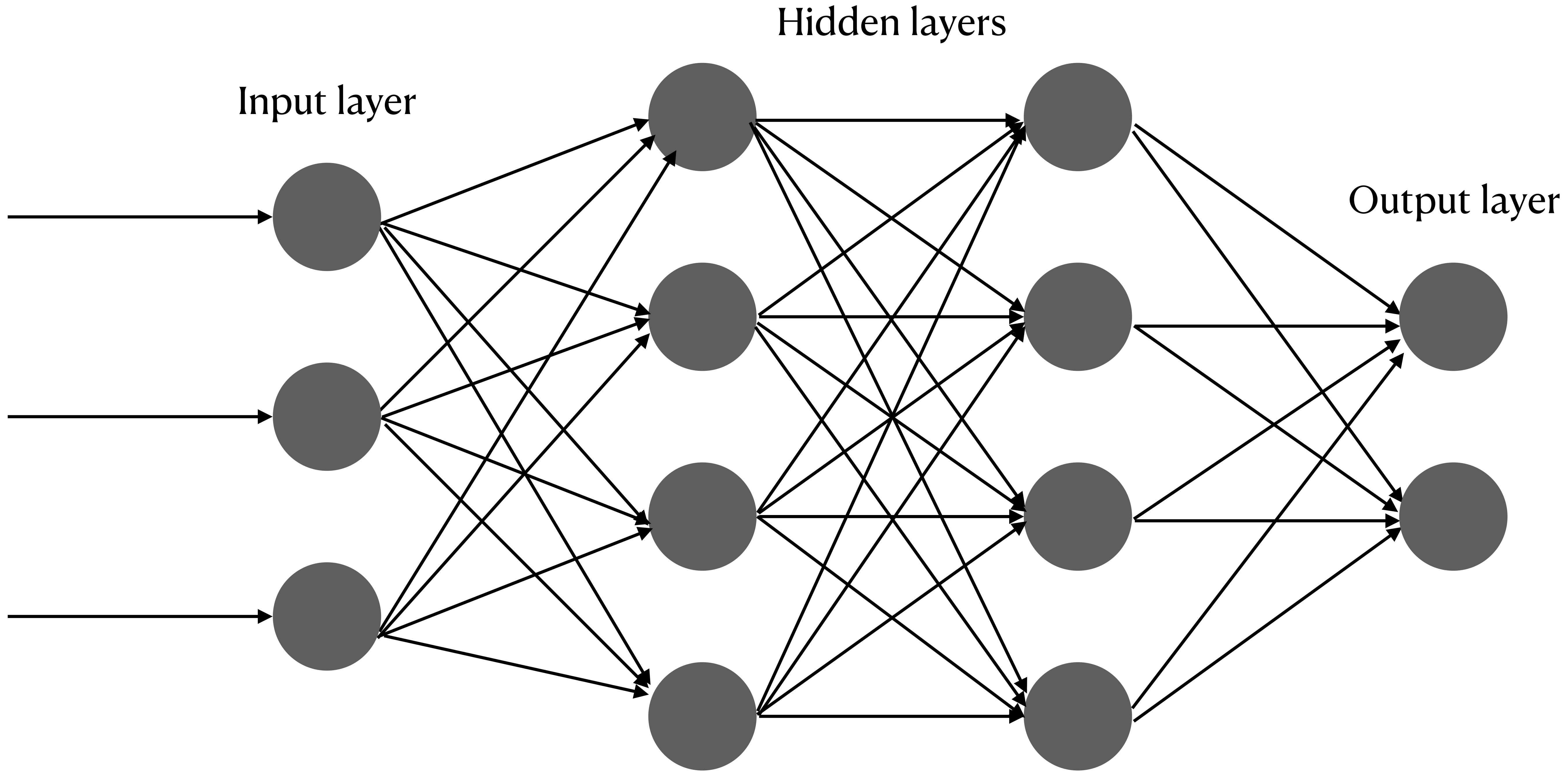
# Diving Into Convolutional Neural Networks (CNNs)

# A Detailed Look at Keras (Suhas' Tutorial)

https://bit.ly/keras25

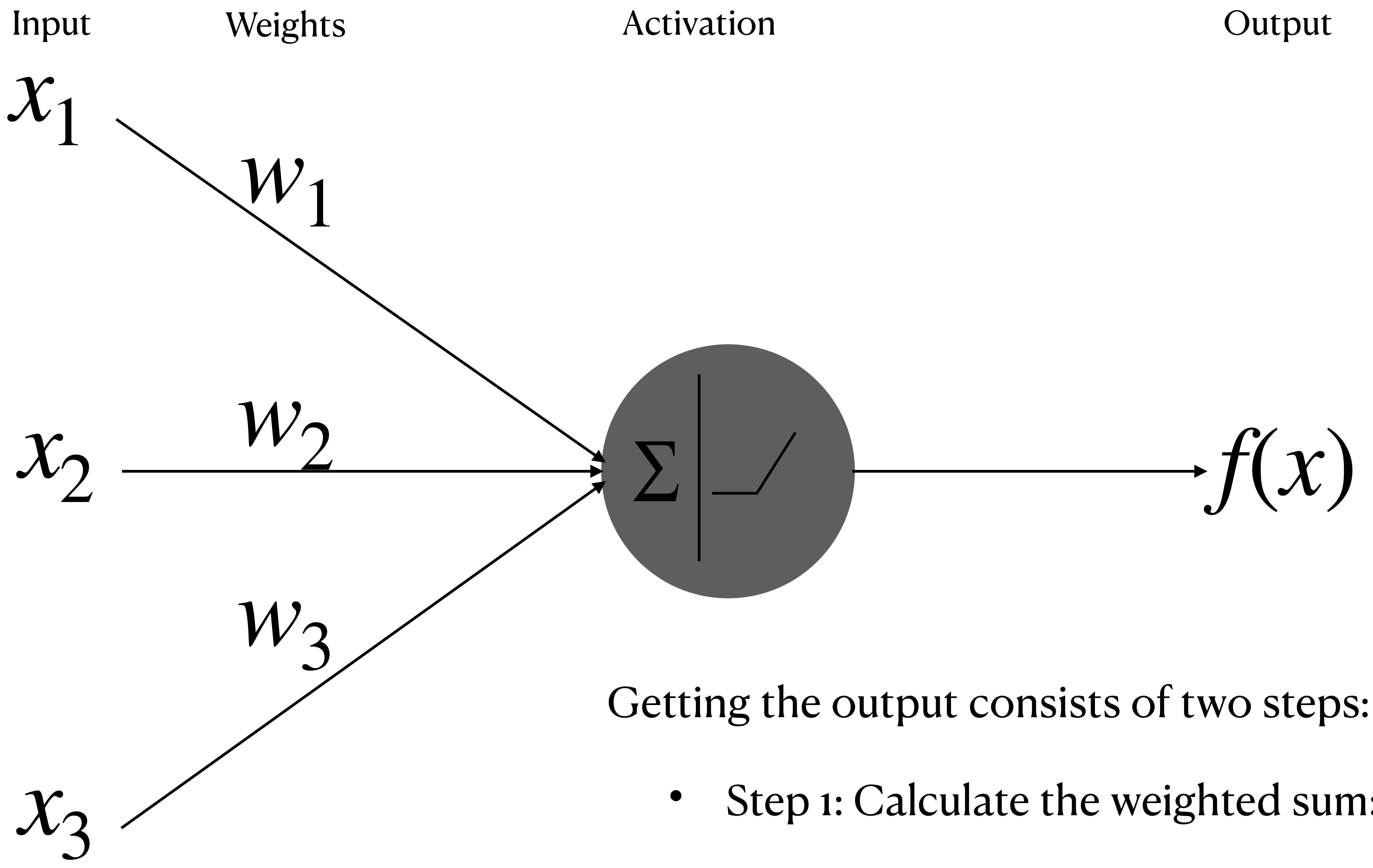
# Deep Learning

- Deep learning is an incremental, layer-by-layer learning approach in which increasingly complex representations are developed to map the input to output




**Artificial Neural Network**

# How Feedforward Neural Networks Work




Getting the output consists of two steps:

- Step 1: Calculate the weighted sum: $w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3$
- Step 2: Applying the activation function $a$:
  $f(x) = a(w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3)$

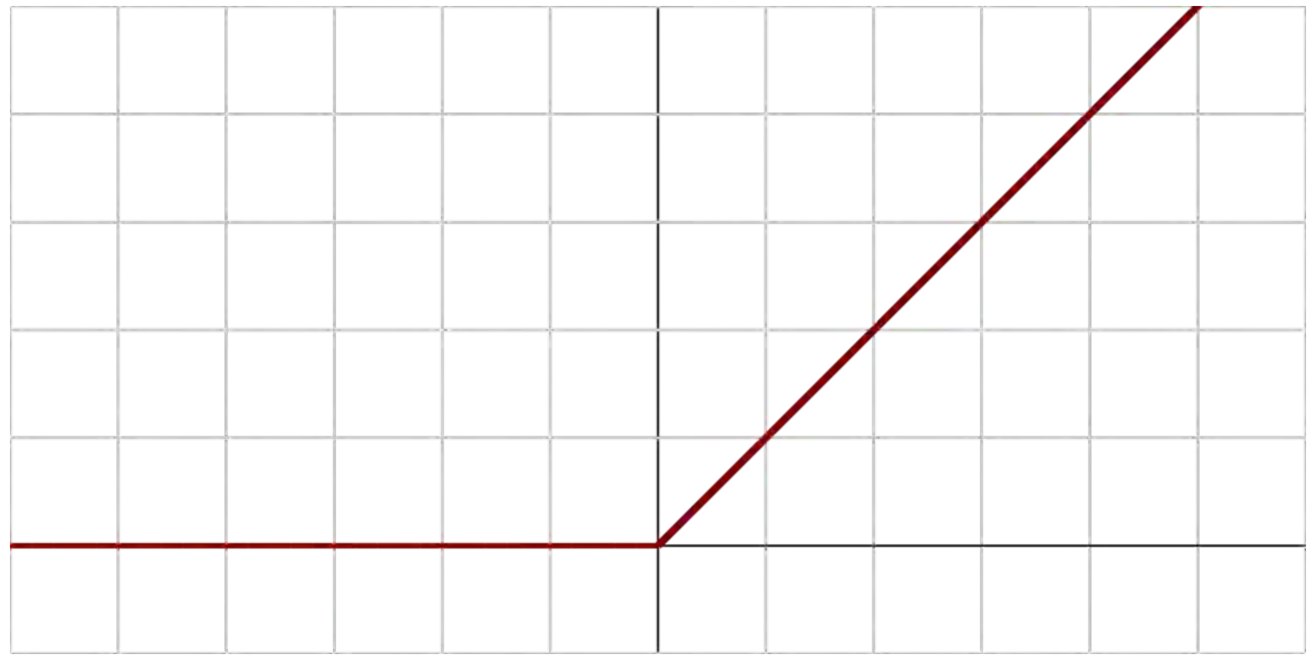
# Activation Functions

## ReLU

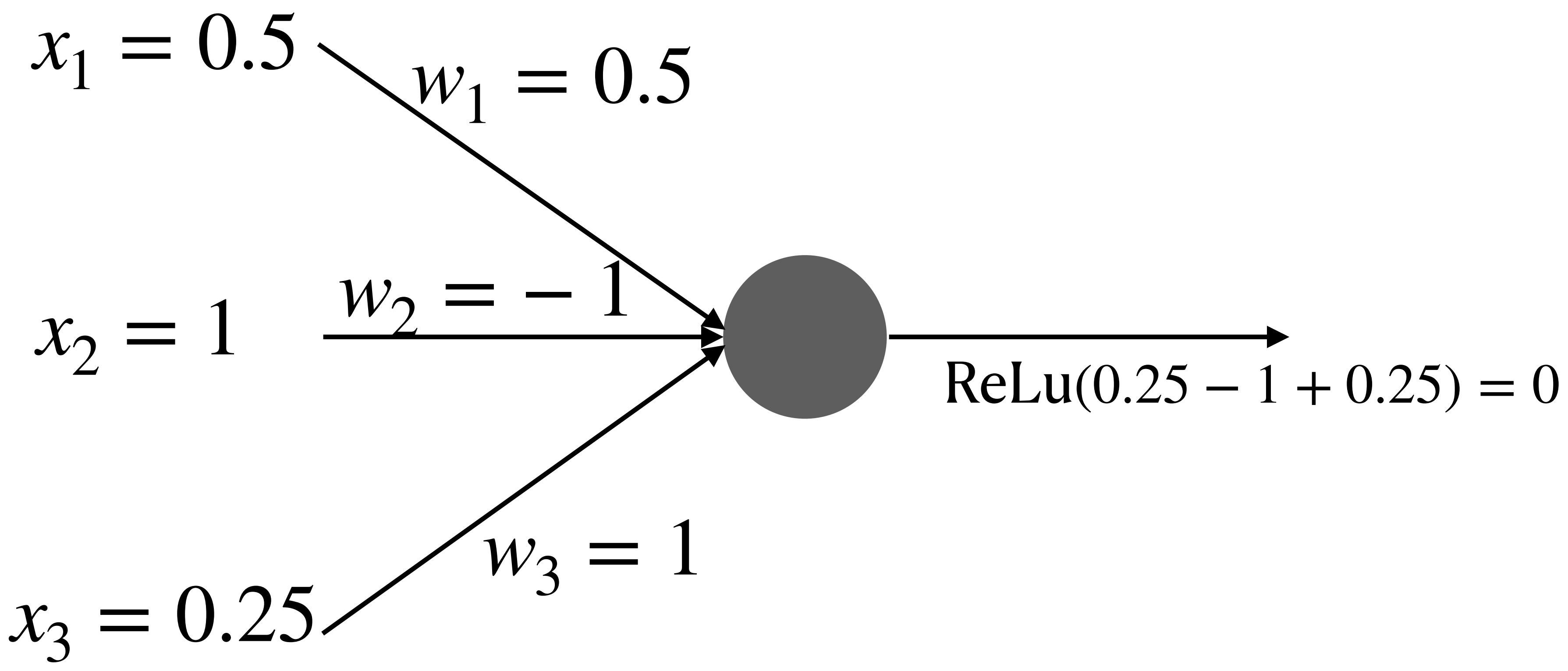
- The most widely used activation function is the **ReLU** (Rectified Linear Unit) function:

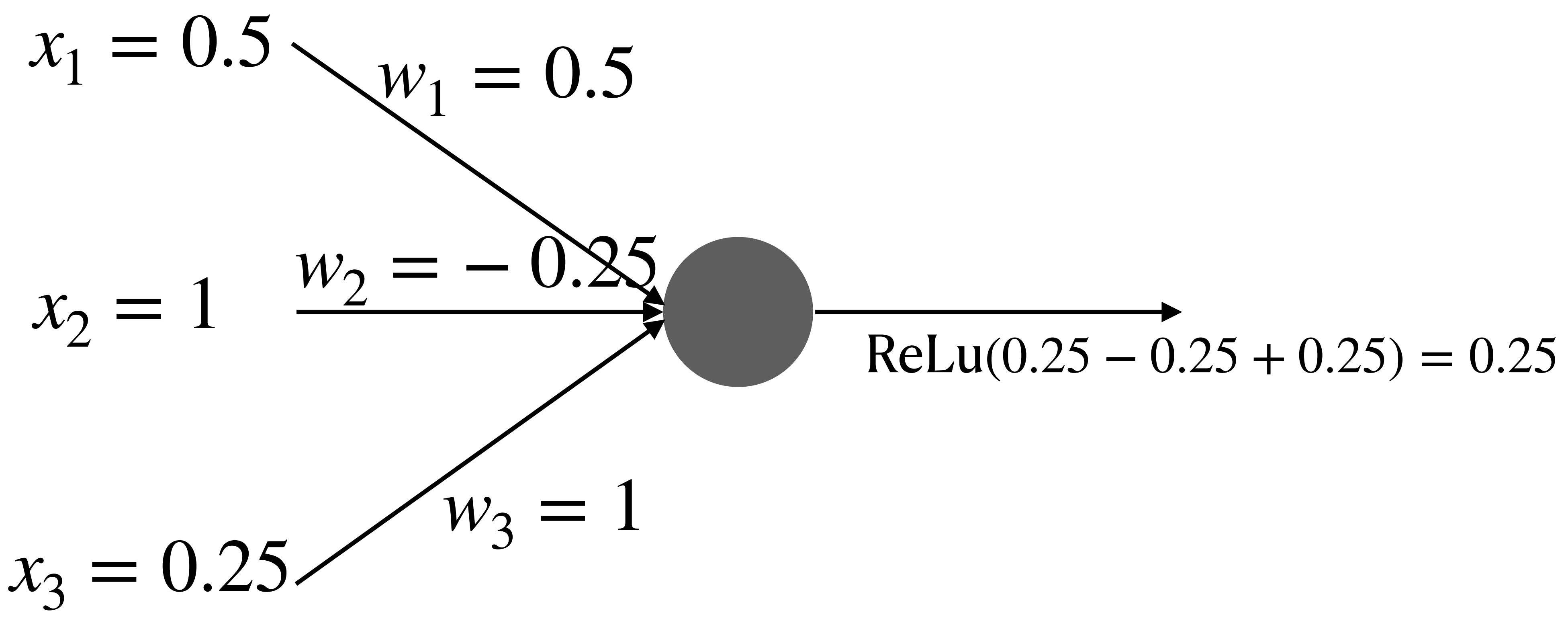
$$\mathrm{ReLU}(x) = \max(x, 0)$$

  - Given an input $x$, the output is $x$ if it is positive and zero otherwise.



- Examples:

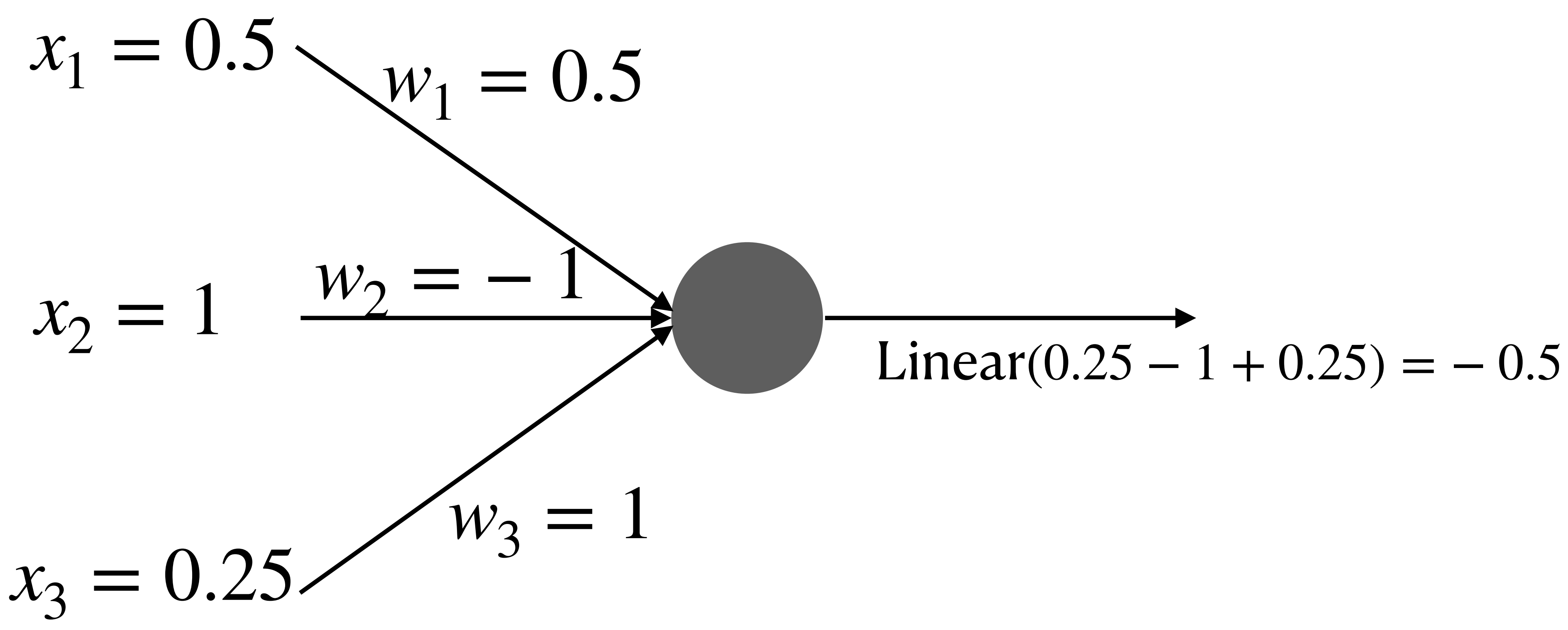


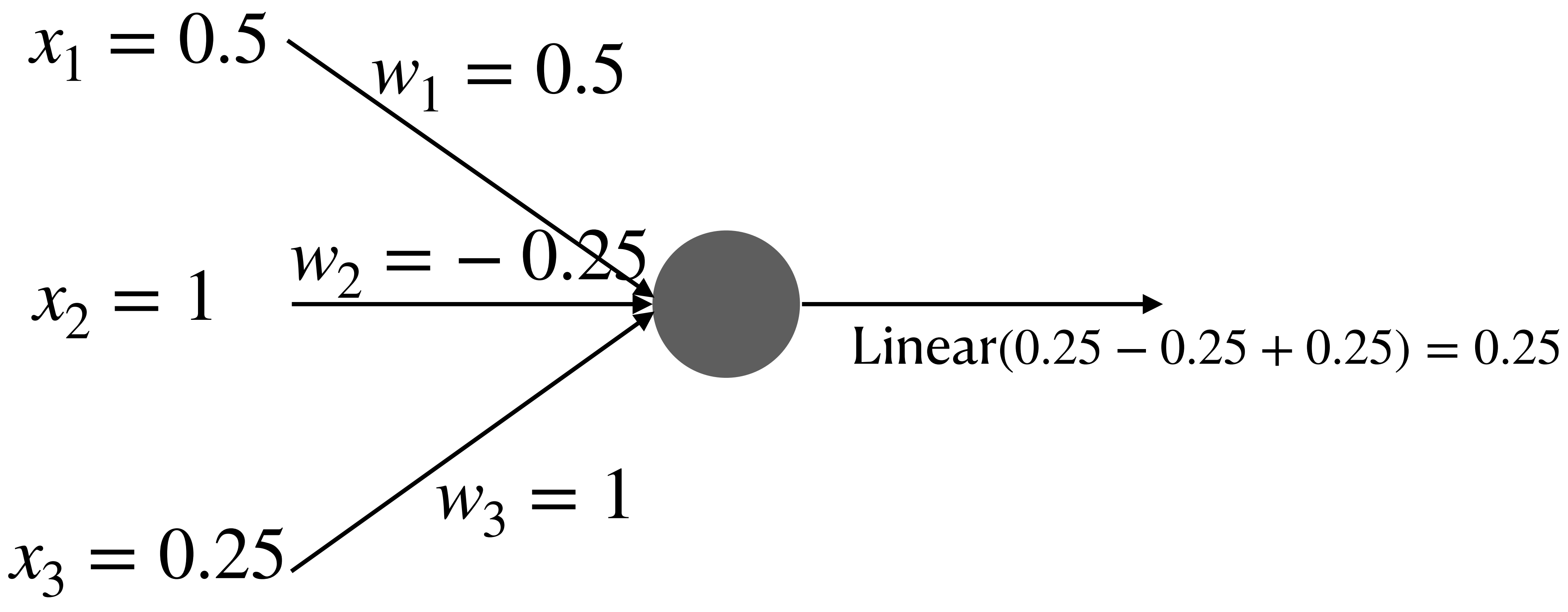



# Activation Function

## Linear Function

- Under a **linear** activation function ($y = x$):
  - Given an input $x$, the output is the same as $x$
- Examples:

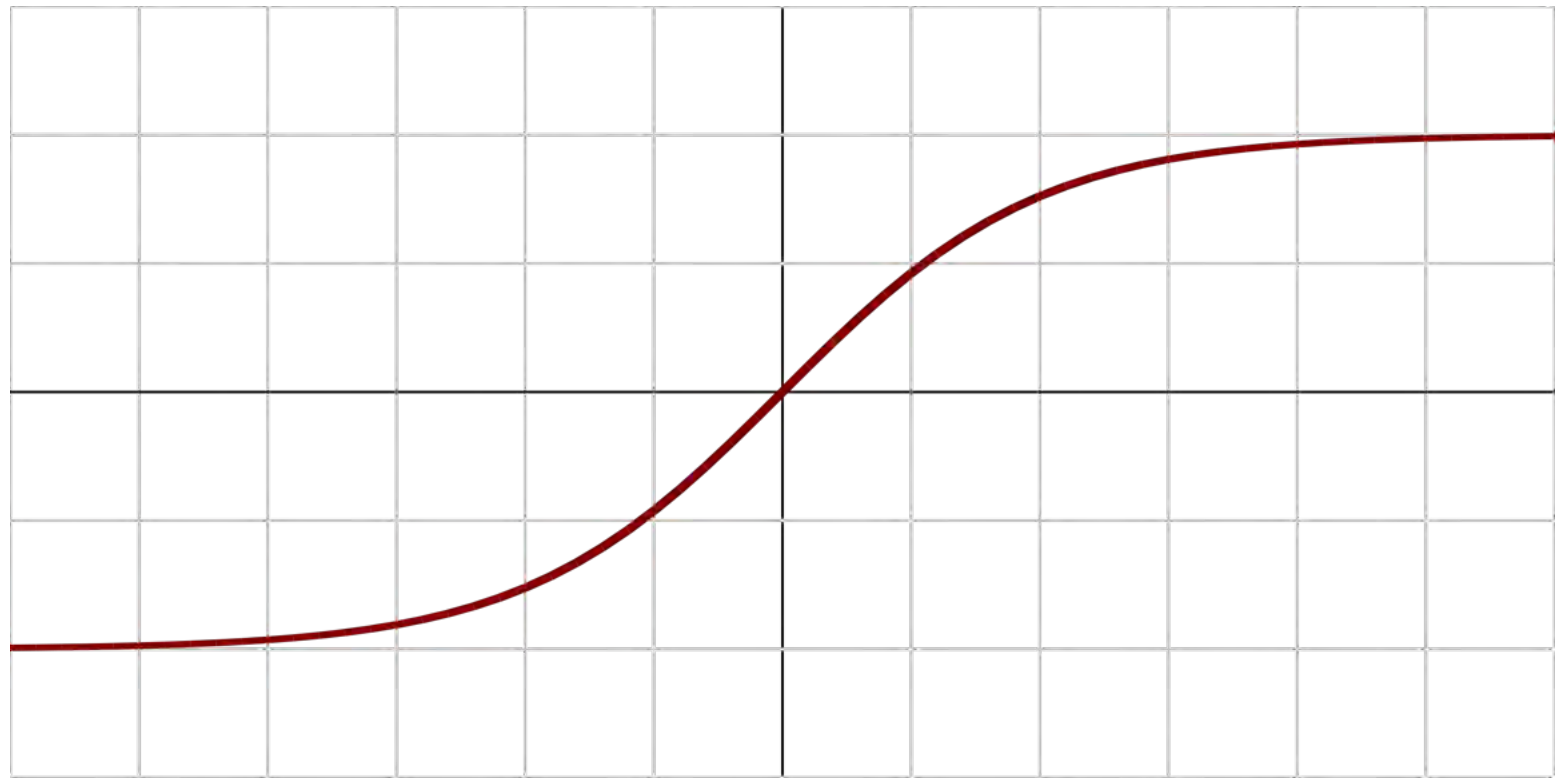



A linear activation function means a linear model regardless of the number of hidden layers

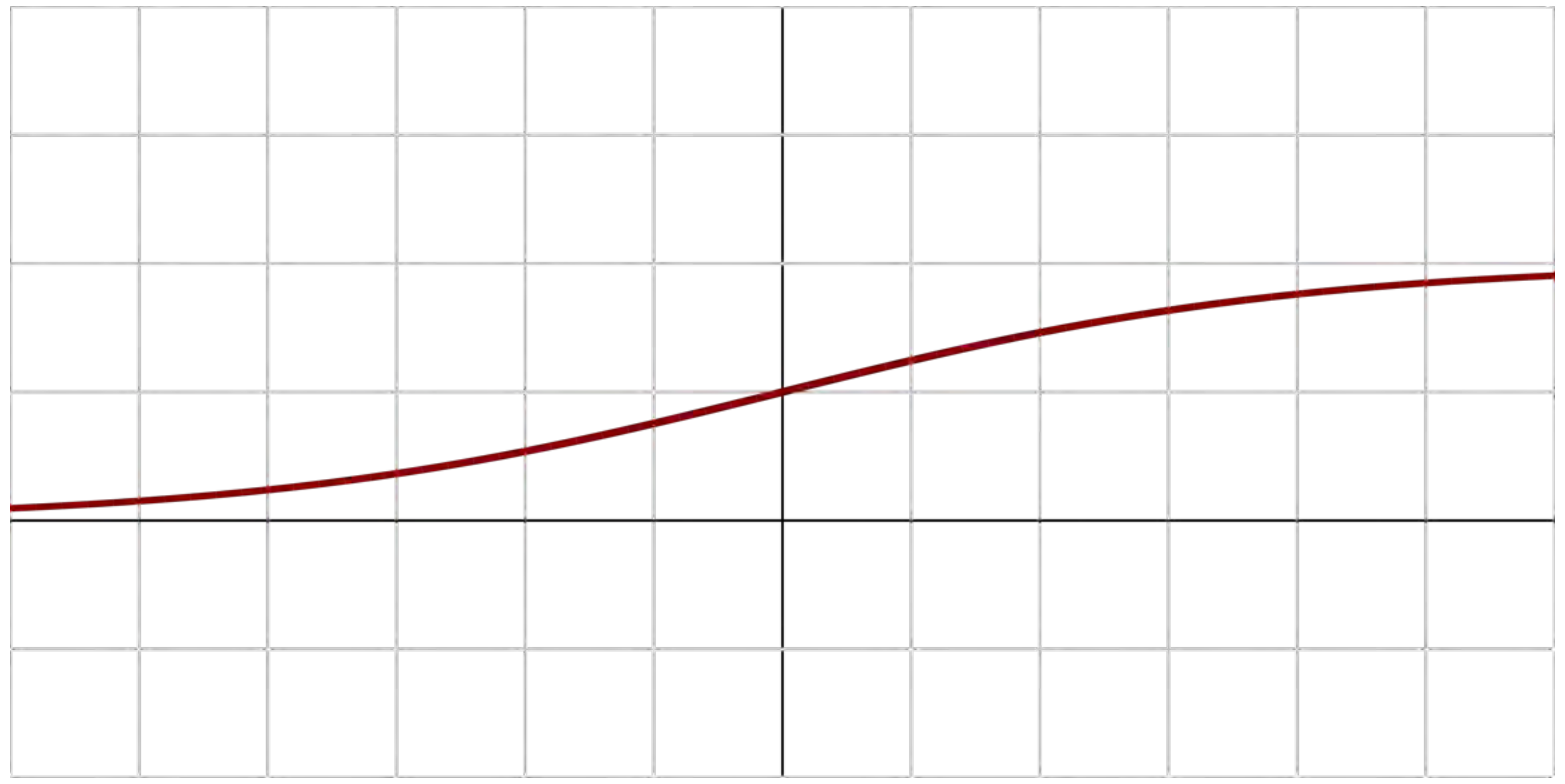
# Other Activation Functions

## Tanh and Sigmoid — “S”-shaped functions



Tanh function

$$\mathrm{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



Sigmoid function

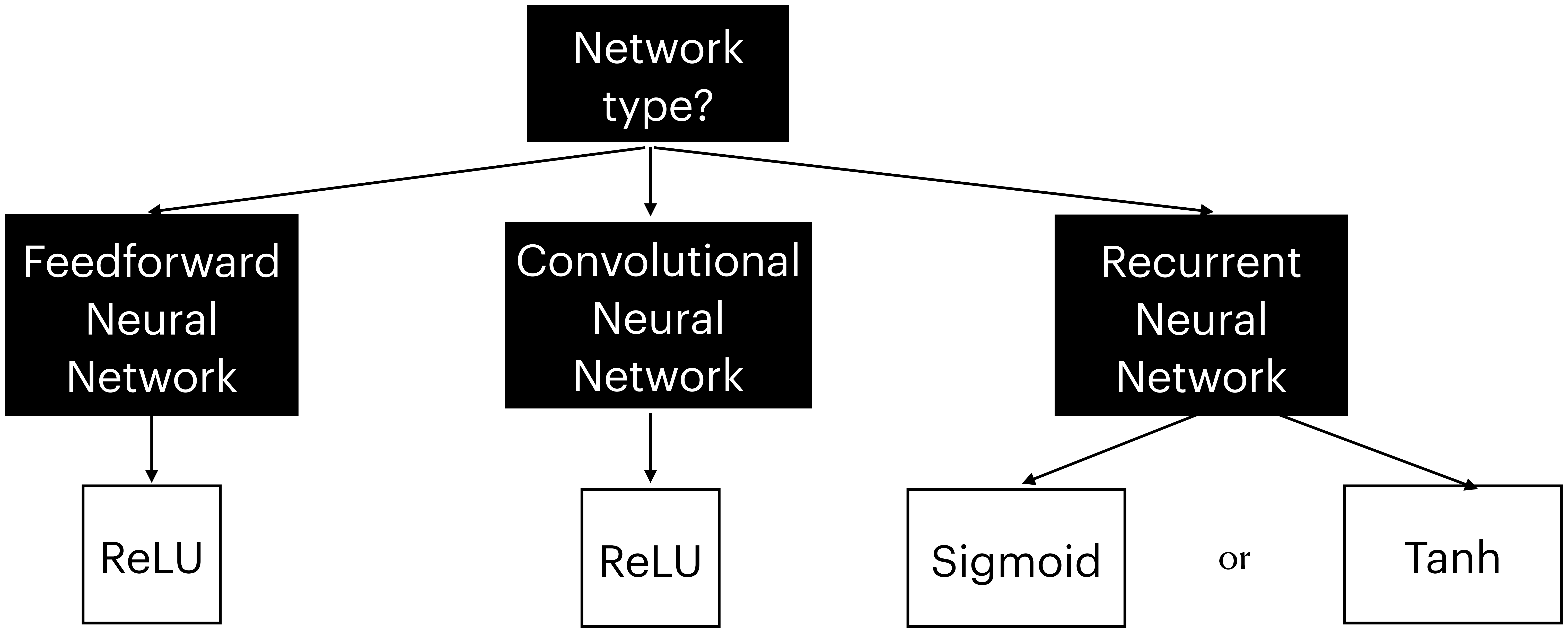
$$\mathrm{Sigmoid}(x) = \frac{e^x}{e^x + 1}$$

“Tanh” rhymes with “branch”

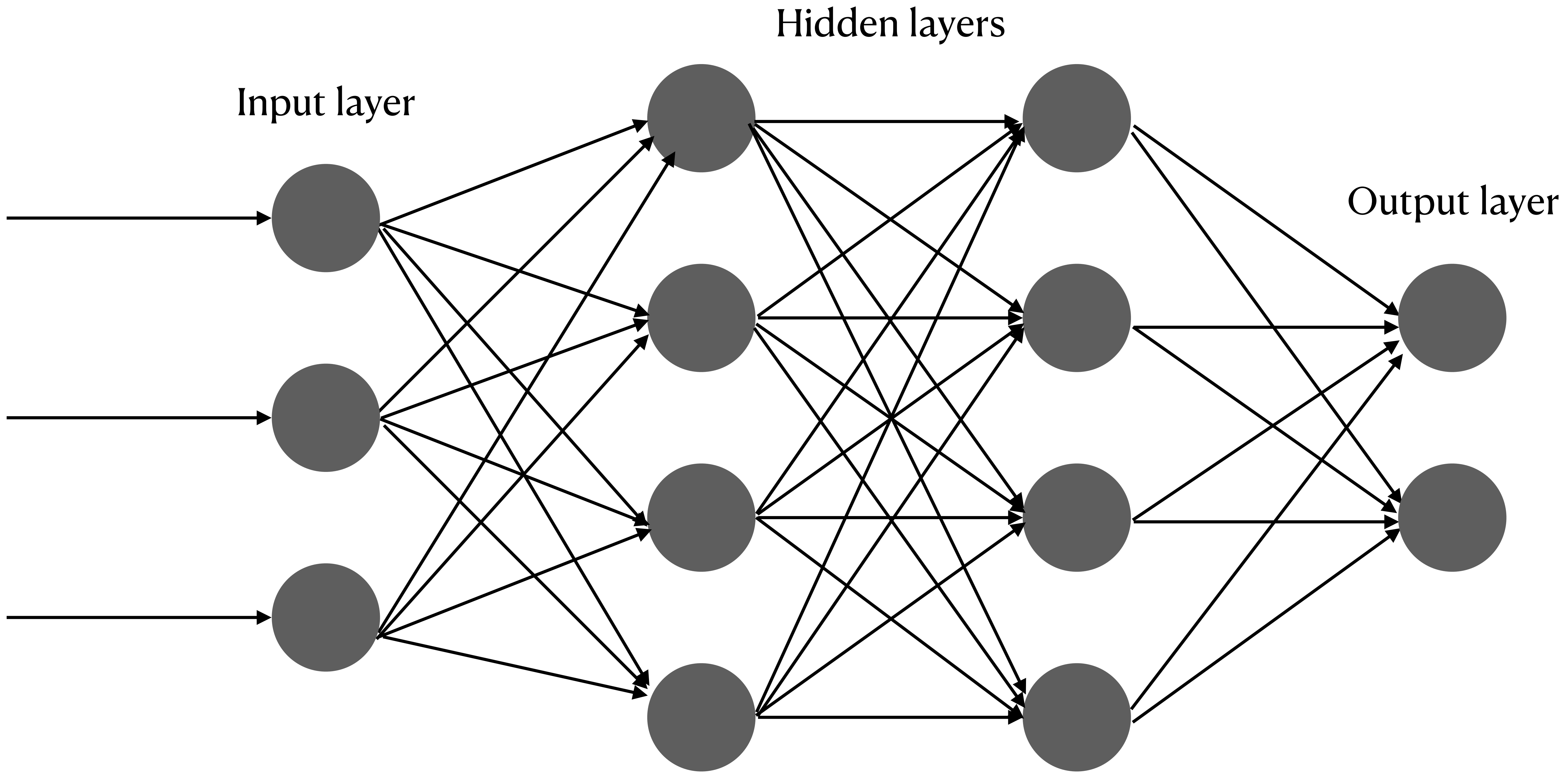
# How to Choose Activation Function

**A bit of a history lesson**

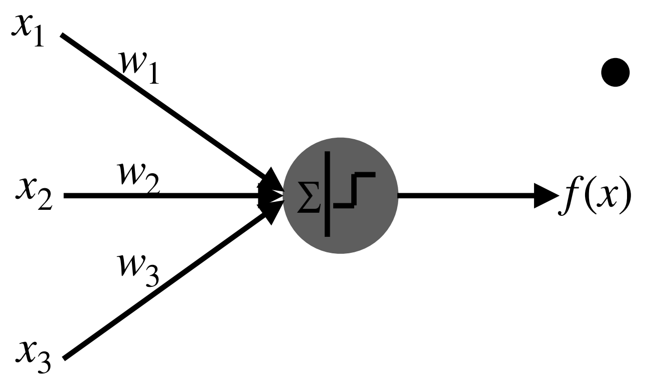
- Until mid-1990s: Sigmoid was the default activation function for training neural networks
- Between mid-1990s to 2010s: Tanh became default activation function for hidden layers
- 2010s to now: the default recommendation is **ReLU** for most of the modern neural networks



# Why Feedforward Neural Network Works



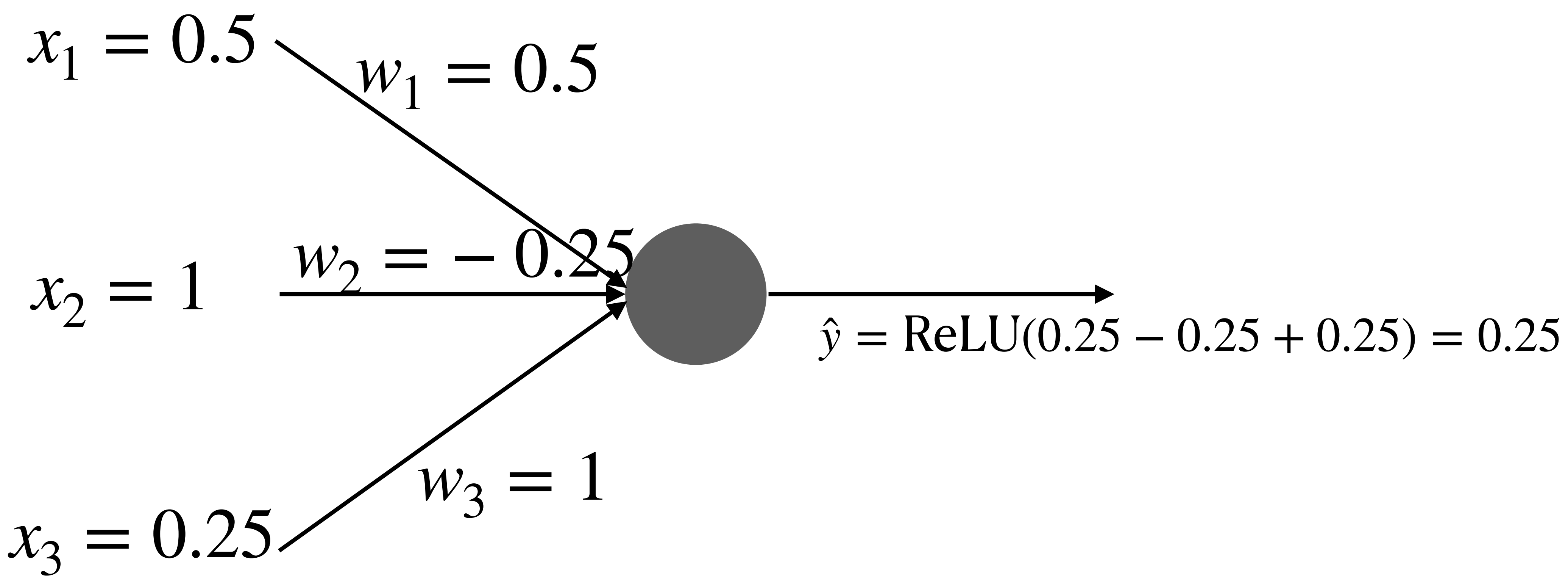

**Feedforward Neural Network**




- Universal Approximation Theorem: A feedforward network with at least one hidden layer with any "squashing" activation function (e.g., ReLU, Sigmoid, and Tanh) can learn any continuous input-output function

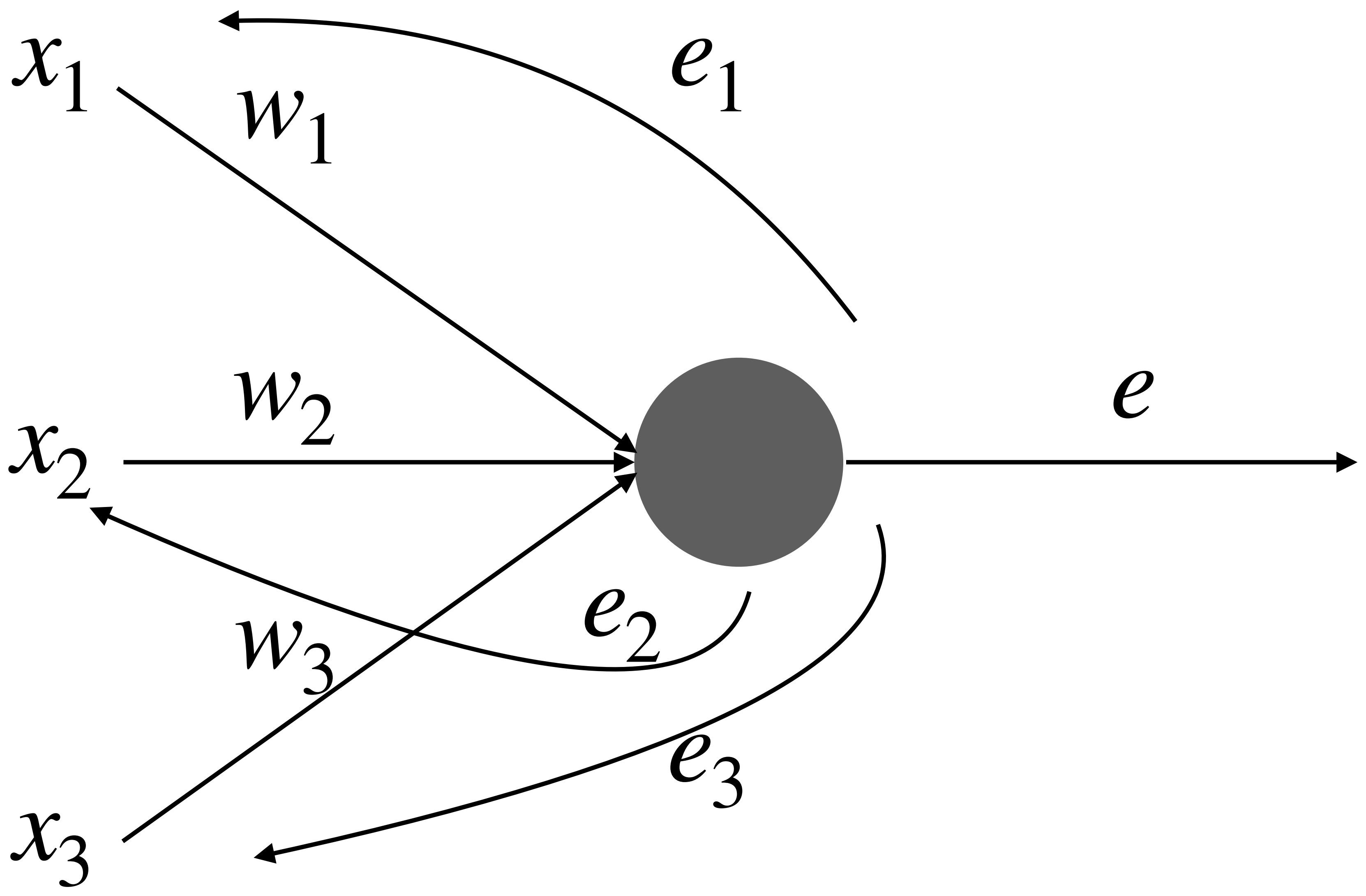
# How Feedforward Neural Networks Learn

## It's Essentially About Learning the Weights!




- We start with randomly assigned weights.
- In the above example, the predicted output $\hat{y} = 0.25$
- If the actual output is $0.5$, then the MSE (mean squared error) is $e = (y - \hat{y})^2 = 0.0625$
- Now what? How can we use the *feedback* to adjust the weights?

# Error Backpropagation




$$e_1 = \left(\frac{w_1}{w_1 + w_2 + w_3}\right) \times e$$

$$e_2 = \left(\frac{w_2}{w_1 + w_2 + w_3}\right) \times e$$

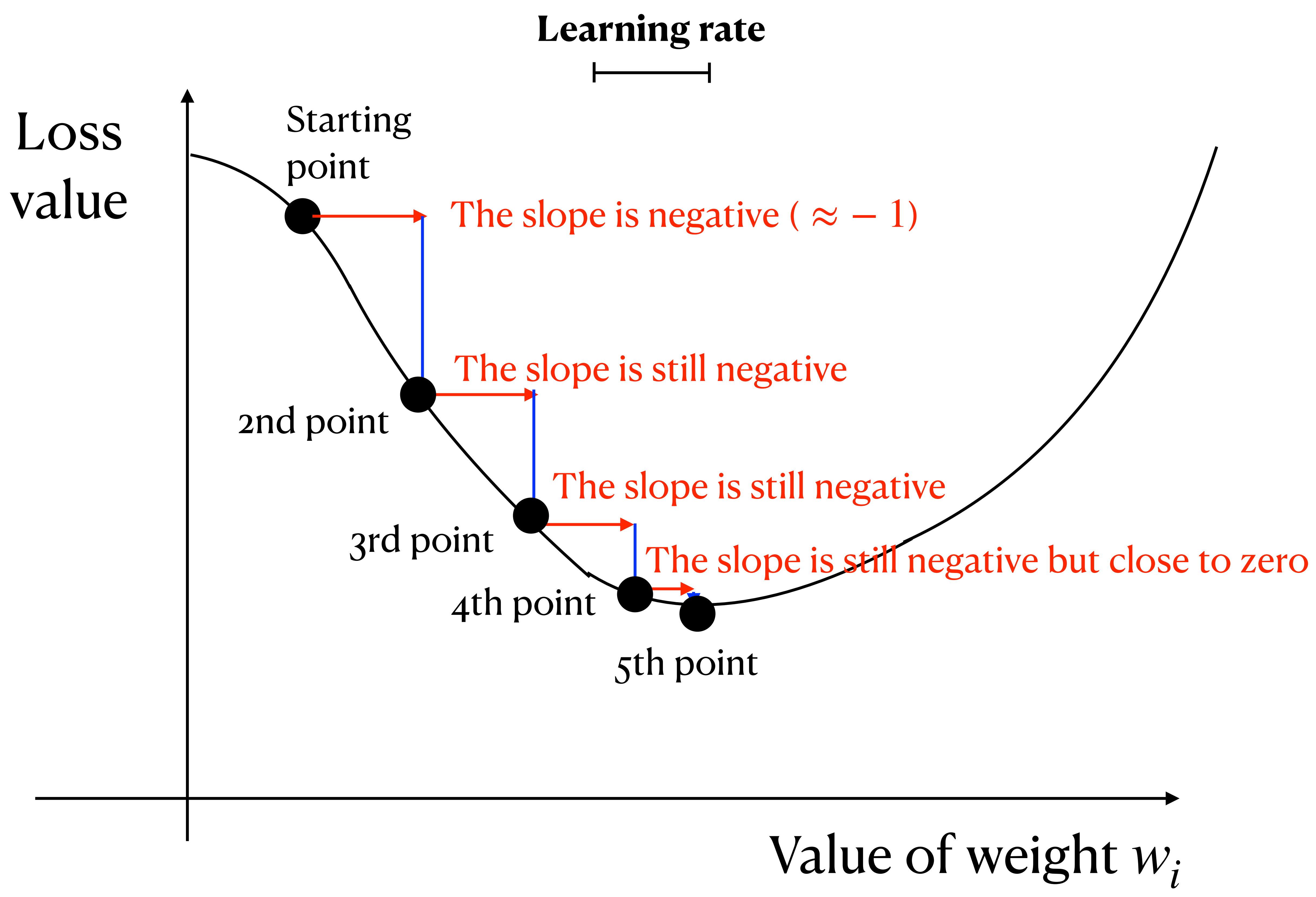
$$e_3 = \left(\frac{w_3}{w_1 + w_2 + w_3}\right) \times e$$

- We can next apply the gradient descent algorithm calculate the **slope** (i.e., gradient) of each weight —how increasing each weight will contribute to the error
- Once we know the slope, we can update the weight:

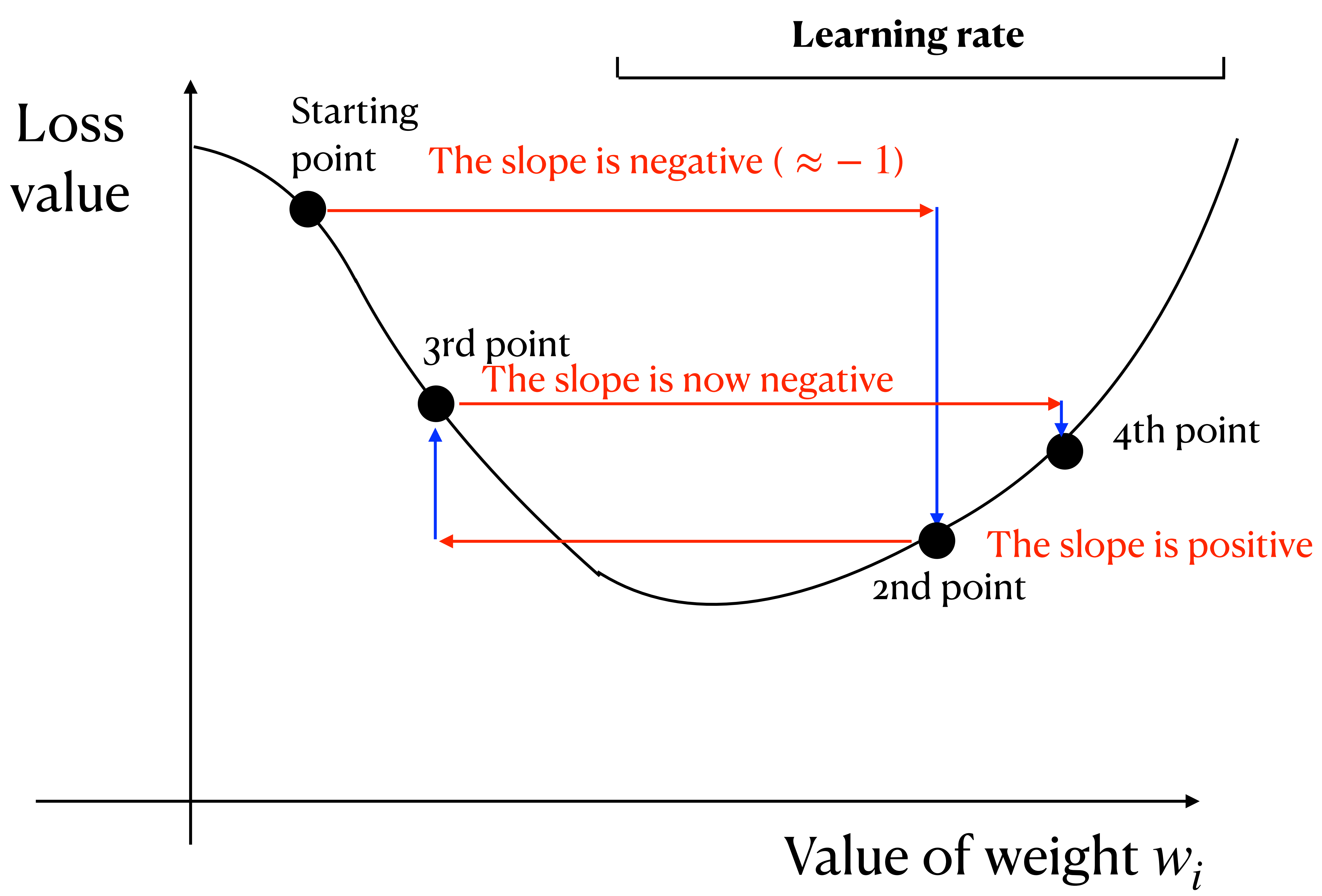
$$\textbf{new } w_i = \textbf{old } w_i - \textbf{learning rate} \cdot \textbf{slope}$$

# Intuition Behind "new $w_i$ = old $w_i$ − learning rate · slope"

## Adjusting the Weight to Minimize the Loss

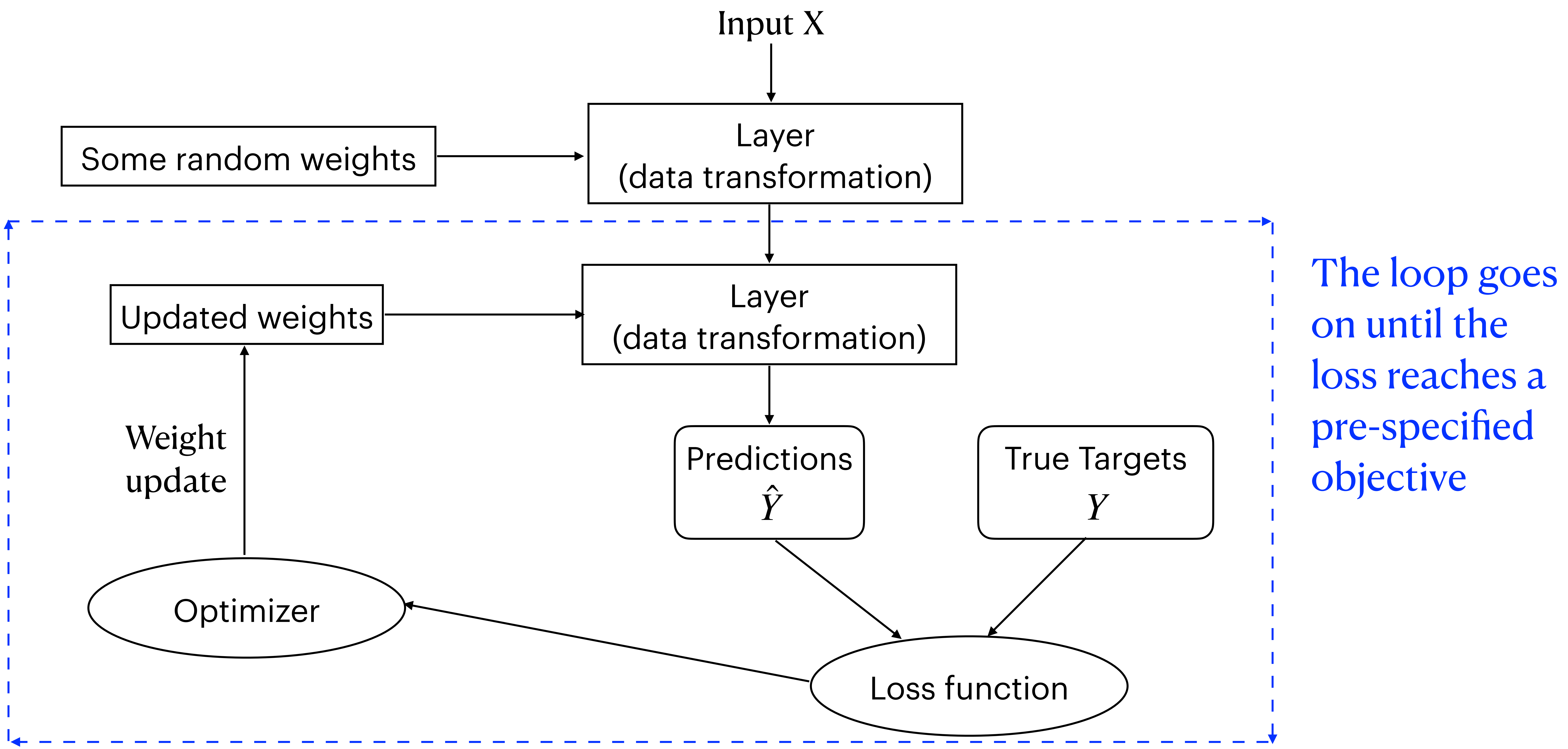


# Avoid Excessively High Learning Rates




**Excessively large learning rates may lead to a behavior like a quantum mechanics experiment gone horribly wrong**

# Summary: Feedforward Neural Networks



# Feedforward Neuron Network in Keras

```
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
```

```
model=keras.Sequential([
    layers.Dense(512,  activation = "relu"),
    layers.Dense(10, activation = "softmax")
])
```

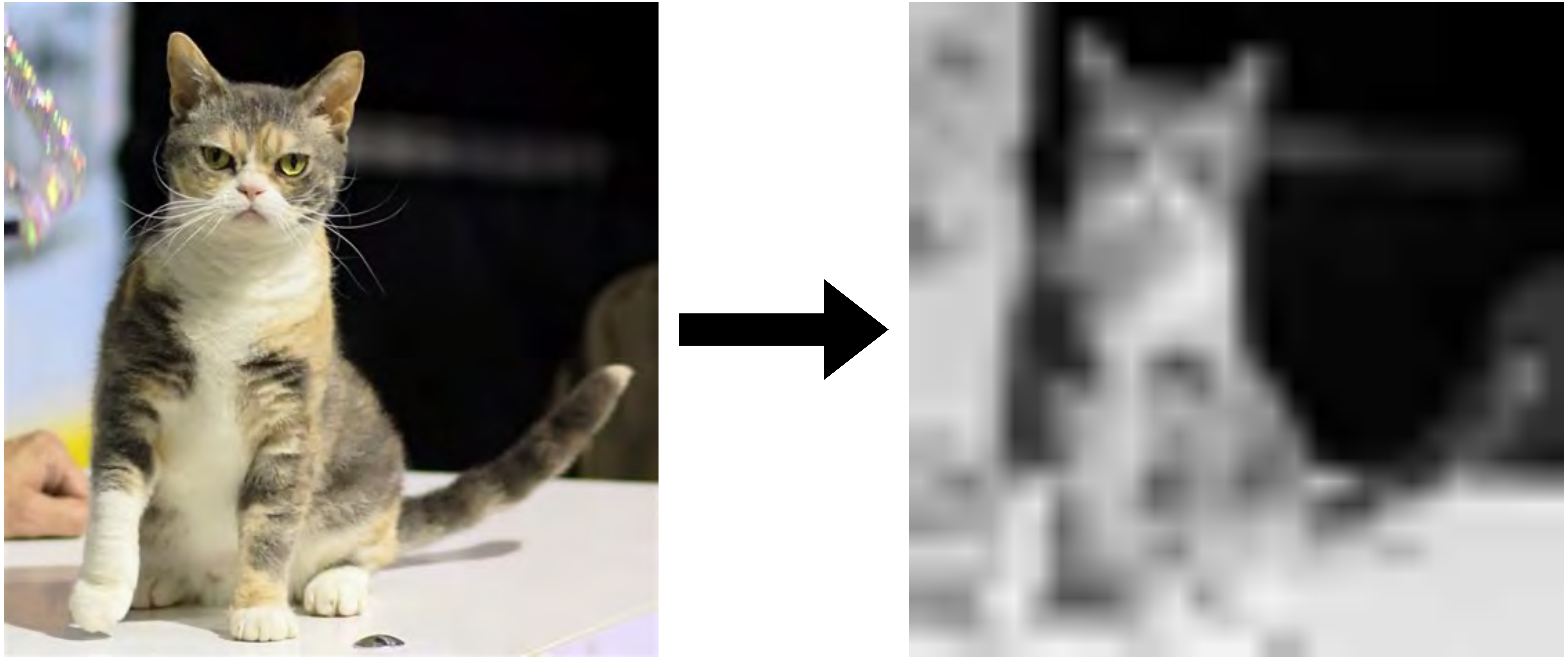
# An Initial Taste of Deep Learning Using Python:

https://bit.ly/aisession3

# Can Feedforward Neural Network Recognize Cats or Dogs?





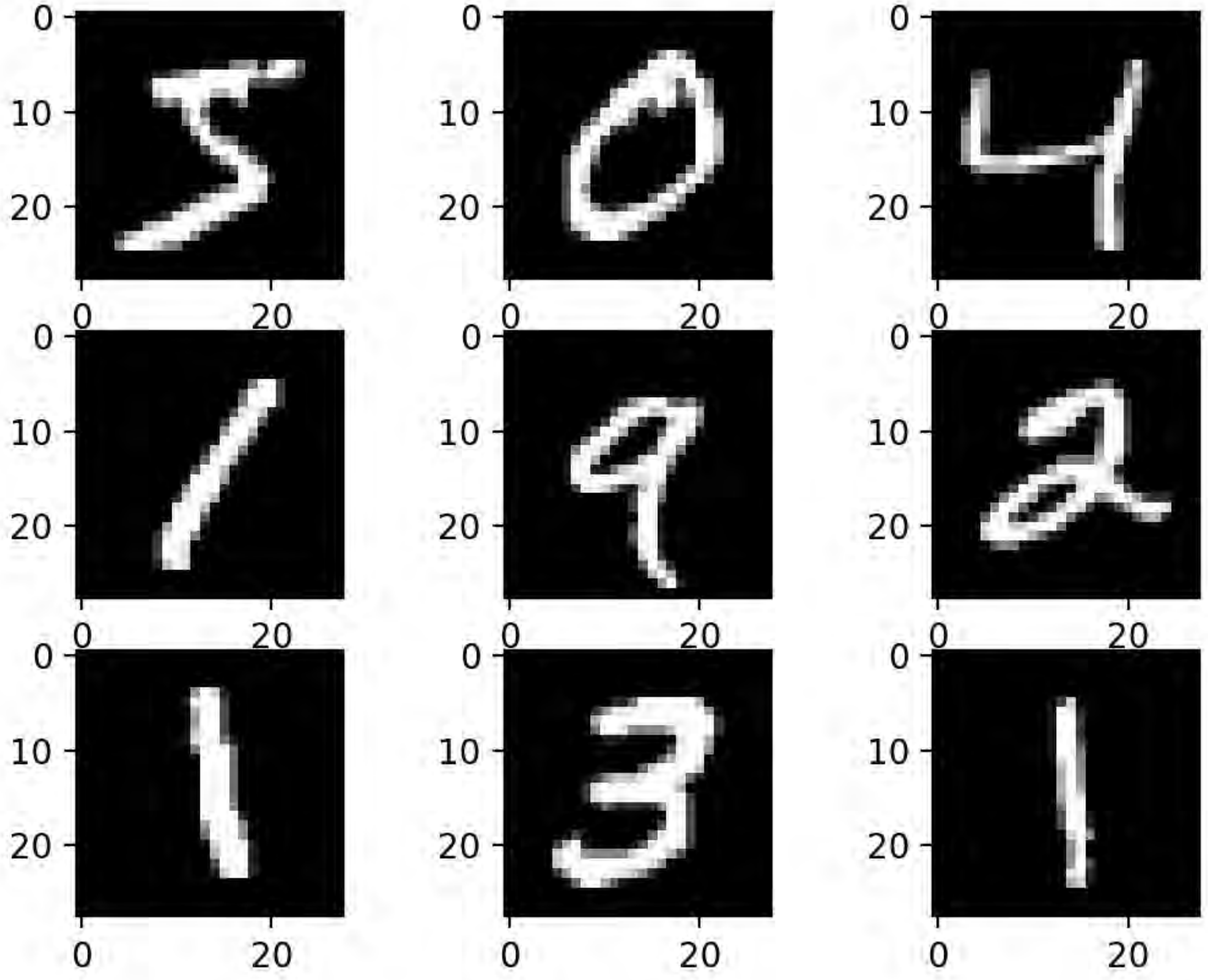



# Can Feedforward Neural Network Recognize Cats or Dogs?

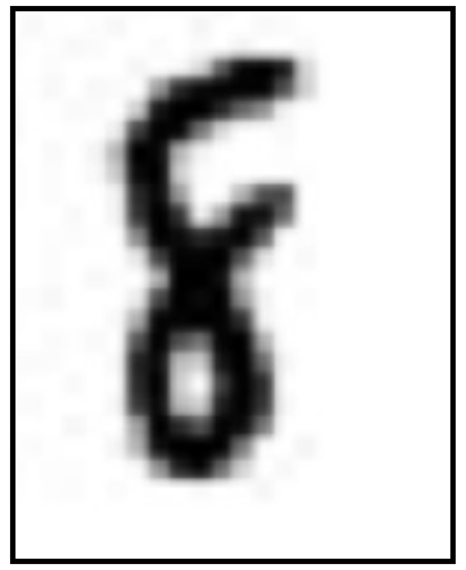




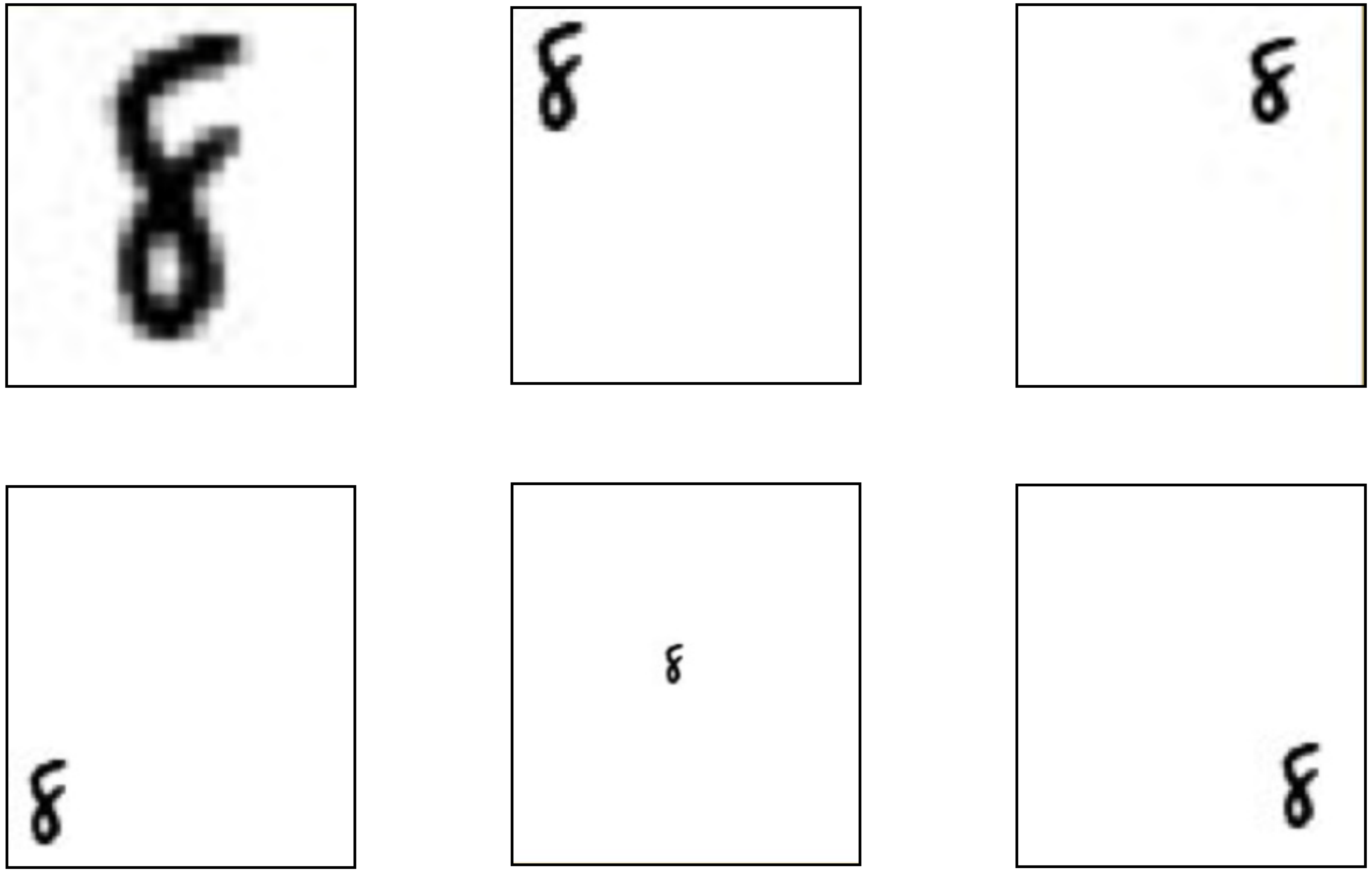

# What about the MNIST Dataset?



**Original Training Image**



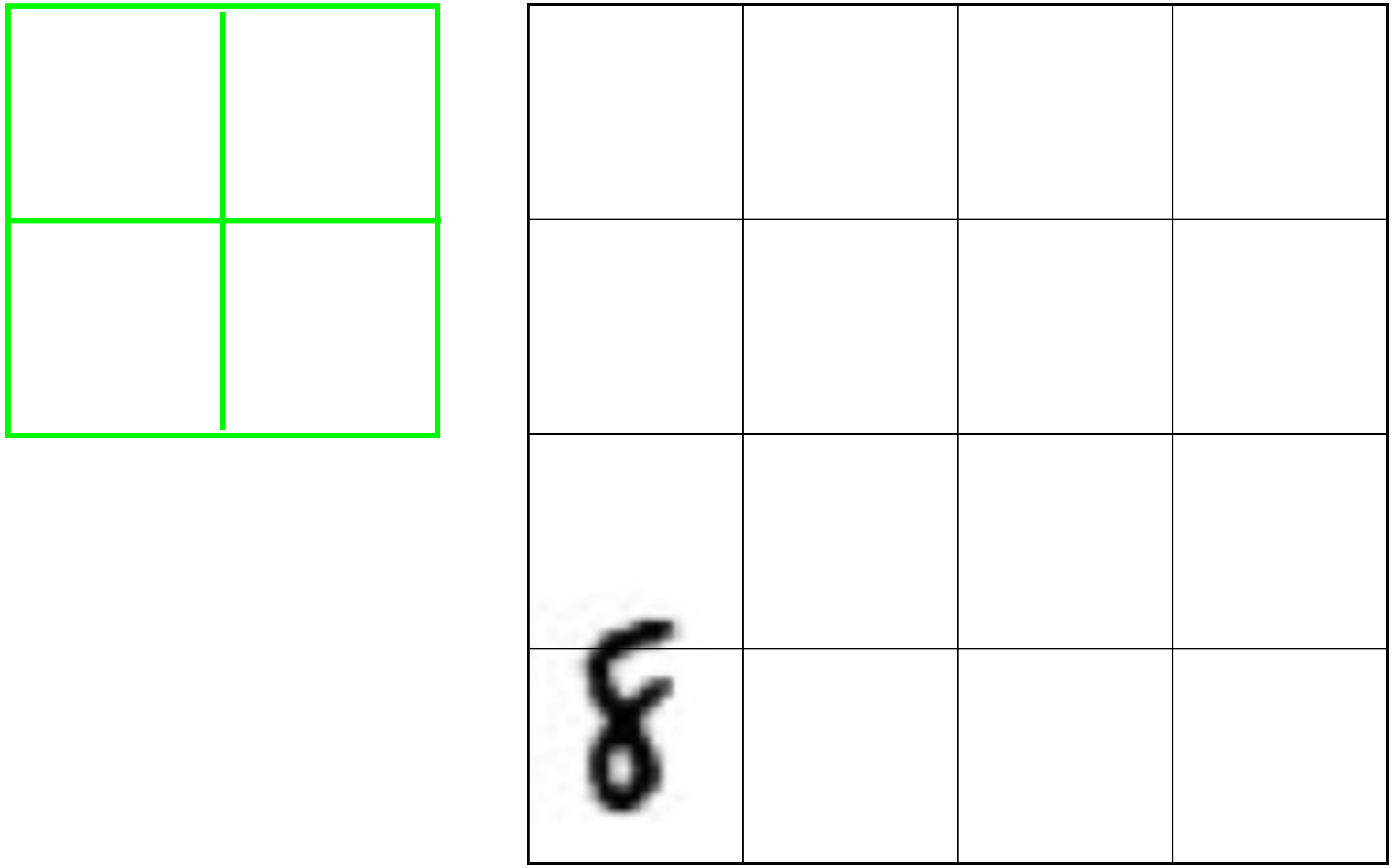
**Practical Training Images**



**How to “catch” the crucial information from ALL of these practical images?**

Geitgey (2016)

# Core Idea: Search with a "Sliding Window"



Geitgey (2016)

# Convolutional Neural Networks (CNN)

- The "convolutional" part does the job of the sliding window
- The convolution step:

**Convoltional filter**

| | |
|---|---|
| 1 | 1 |
| 0 | 1 |

**Input Feature Map**

| | | | |
|---|---|---|---|
| 0 | 1 | 1 | 1 |
| 0 | 0 | 0 | 1 |
| 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 |

**Output Feature Map**

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 0 | 1 | 1 |
| 0 | 1 | 1 |

# Are These Two Letters Identical?

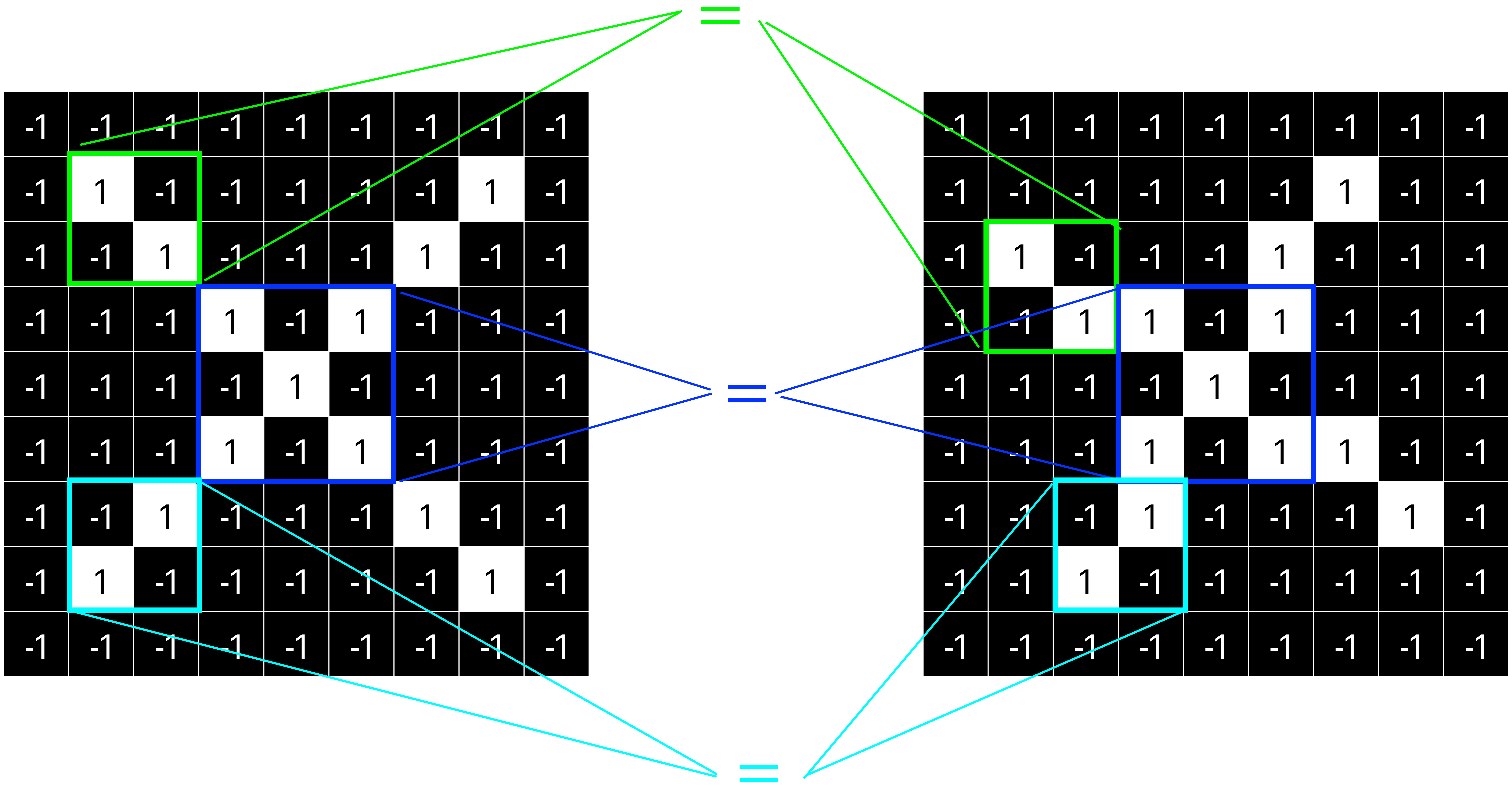
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| -1 | 1 | -1 | -1 | -1 | -1 | -1 | 1 | -1 |
| -1 | -1 | 1 | -1 | -1 | -1 | 1 | -1 | -1 |
| -1 | -1 | -1 | 1 | -1 | 1 | -1 | -1 | -1 |
| -1 | -1 | -1 | -1 | 1 | -1 | -1 | -1 | -1 |
| -1 | -1 | -1 | 1 | -1 | 1 | -1 | -1 | -1 |
| -1 | -1 | 1 | -1 | -1 | -1 | 1 | -1 | -1 |
| -1 | 1 | -1 | -1 | -1 | -1 | -1 | 1 | -1 |
| -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| -1 | -1 | -1 | -1 | -1 | -1 | 1 | -1 | -1 |
| -1 | 1 | -1 | -1 | -1 | 1 | -1 | -1 | -1 |
| -1 | -1 | 1 | 1 | -1 | 1 | -1 | -1 | -1 |
| -1 | -1 | -1 | -1 | 1 | -1 | -1 | -1 | -1 |
| -1 | -1 | -1 | 1 | -1 | 1 | 1 | -1 | -1 |
| -1 | -1 | -1 | 1 | -1 | -1 | -1 | 1 | -1 |
| -1 | -1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 |
| -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |

Unfortunately, for a computer, the above two images can represent two very different patterns

We need to figure out a way for the computer to recognize shifted, shrunk, rotated, or deformed images.

# Fundamental Features of Letter X



# We Can Detect These Features Using Filters

**Convoltional filter**

| | | |
|---|---|---|
| 1 | -1 | -1 |
| -1 | 1 | -1 |
| -1 | -1 | 1 |

**Input Feature Map**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| -1 | 1 | -1 | -1 | -1 | -1 | -1 | 1 | -1 |
| -1 | -1 | 1 | -1 | -1 | -1 | 1 | -1 | -1 |
| -1 | -1 | -1 | 1 | -1 | 1 | -1 | -1 | -1 |
| -1 | -1 | -1 | -1 | 1 | -1 | -1 | -1 | -1 |
| -1 | -1 | -1 | 1 | -1 | 1 | -1 | -1 | -1 |
| -1 | -1 | 1 | -1 | -1 | -1 | 1 | -1 | -1 |
| -1 | 1 | -1 | -1 | -1 | -1 | -1 | 1 | -1 |
| -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |

Apply the convolution operation:

$= 9$

# Some Common Filters

Input image

| | | |
|---|---|---|
| -1 | -1 | -1 |
| -1 | 8 | -1 |
| -1 | -1 | -1 |

Outline

| | | |
|---|---|---|
| 0 | -1 | 0 |
| -1 | 5.5 | -1 |
| 0 | -1 | 0 |

Sharpen

| | | |
|---|---|---|
| 1 | 0 | -1 |
| 2 | 0 | -2 |
| 1 | 0 | -1 |

Left sobel
(highlighting vertical edges and outlines)

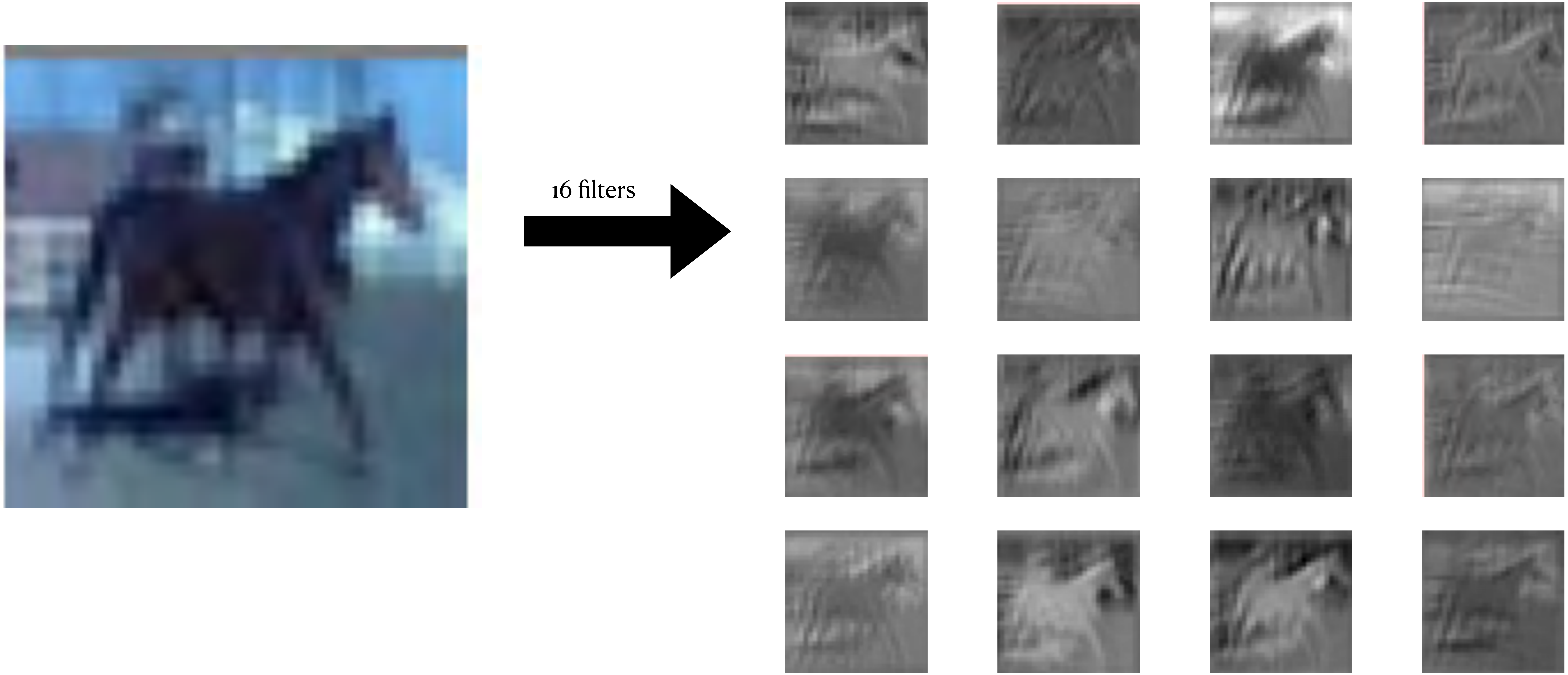
| | | |
|---|---|---|
| -2 | -1 | 0 |
| -1 | 1 | 1 |
| 0 | 1 | 2 |

Emboss
(emphasizing the edges)

https://bit.ly/cnnfilters

# Feature Extraction with Convolution

## Use Multiple Filters to Extra Different Features



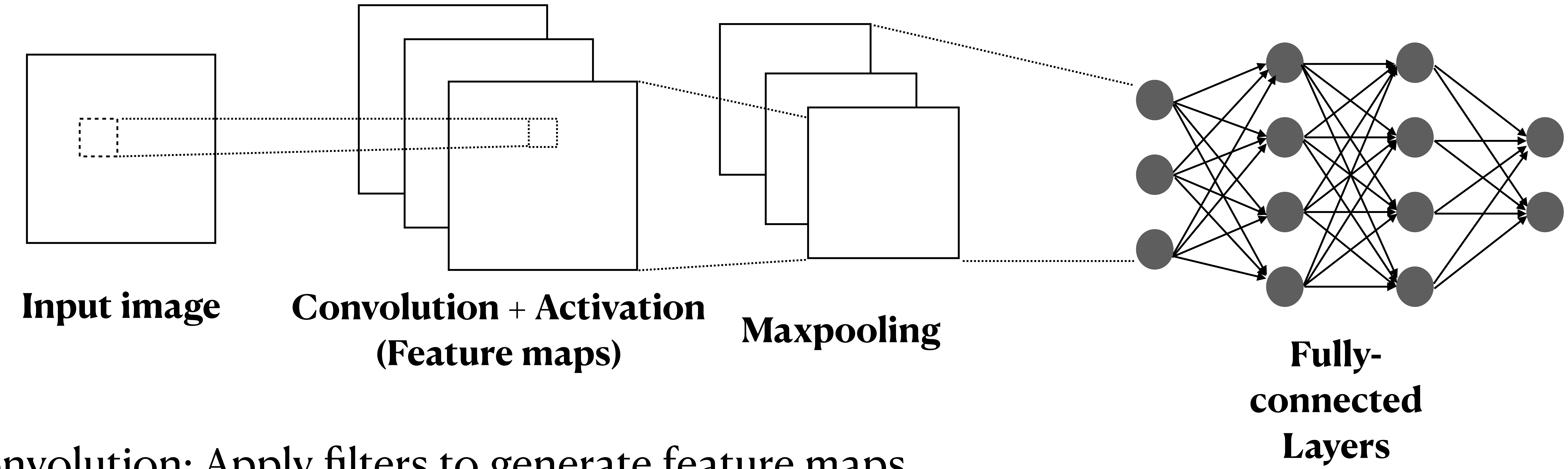
# Filters Are Learned, Not Hand-Picked

- Filters (kernels) start random and are learned to reduce the loss
- Forward pass → prediction; compare to target → loss
- Backprop sends feedback to every weight; optimizer nudges weights
- Over time, filters align to useful patterns (edges, textures)

# Backpropagation Intuition for Convolutions

- Same idea as earlier backprop slides, applied to kernel weights
- Each weight gets a “helped or hurt?” signal from the loss
- Update rule: new $w_i$ = old $w_i$ − learning rate · slope
- Repeat over mini-batches until validation stops improving

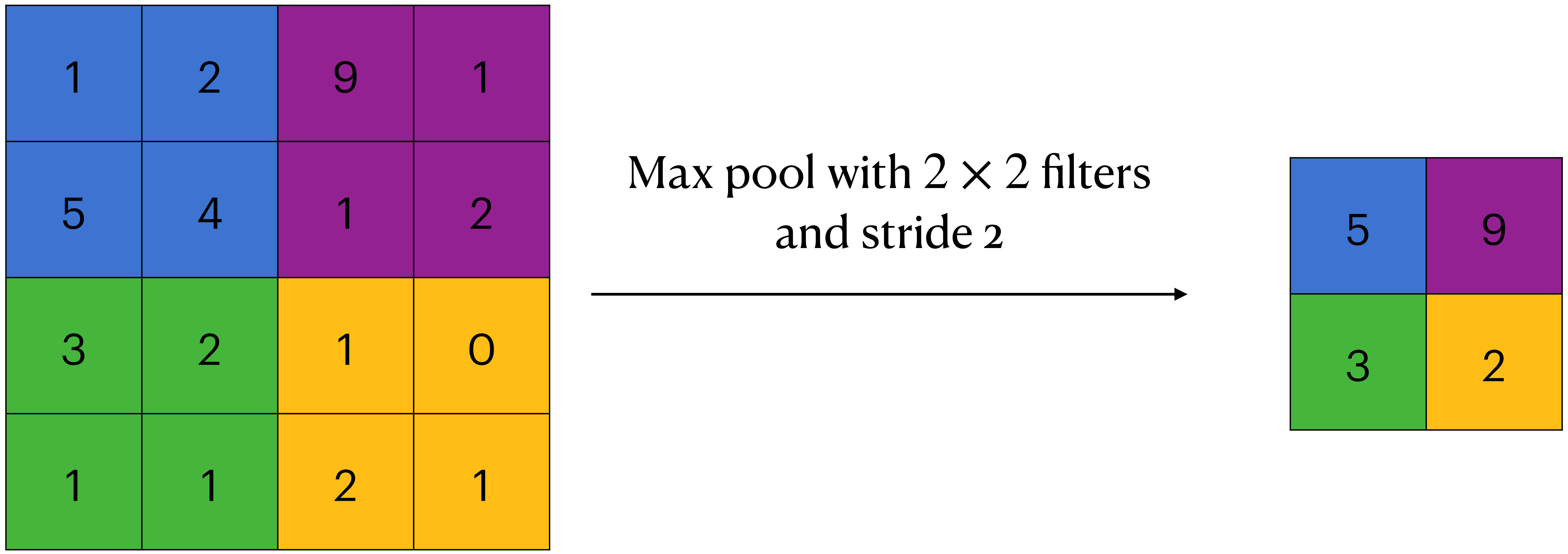
# Convolutional Neural Networks (CNNs)




- Convolution: Apply filters to generate feature maps
- Activation: Usually ReLU
- Maxpooling: Downsampling operation on each feature map
- Fully connected layers: Classification

# MaxPooling

- Purpose: to aggressively “downsample” feature maps (i.e., make the representations smaller and more manageable)
- How it works: Extract windows from the input feature map, and pick the maximum value of each window
- Example:

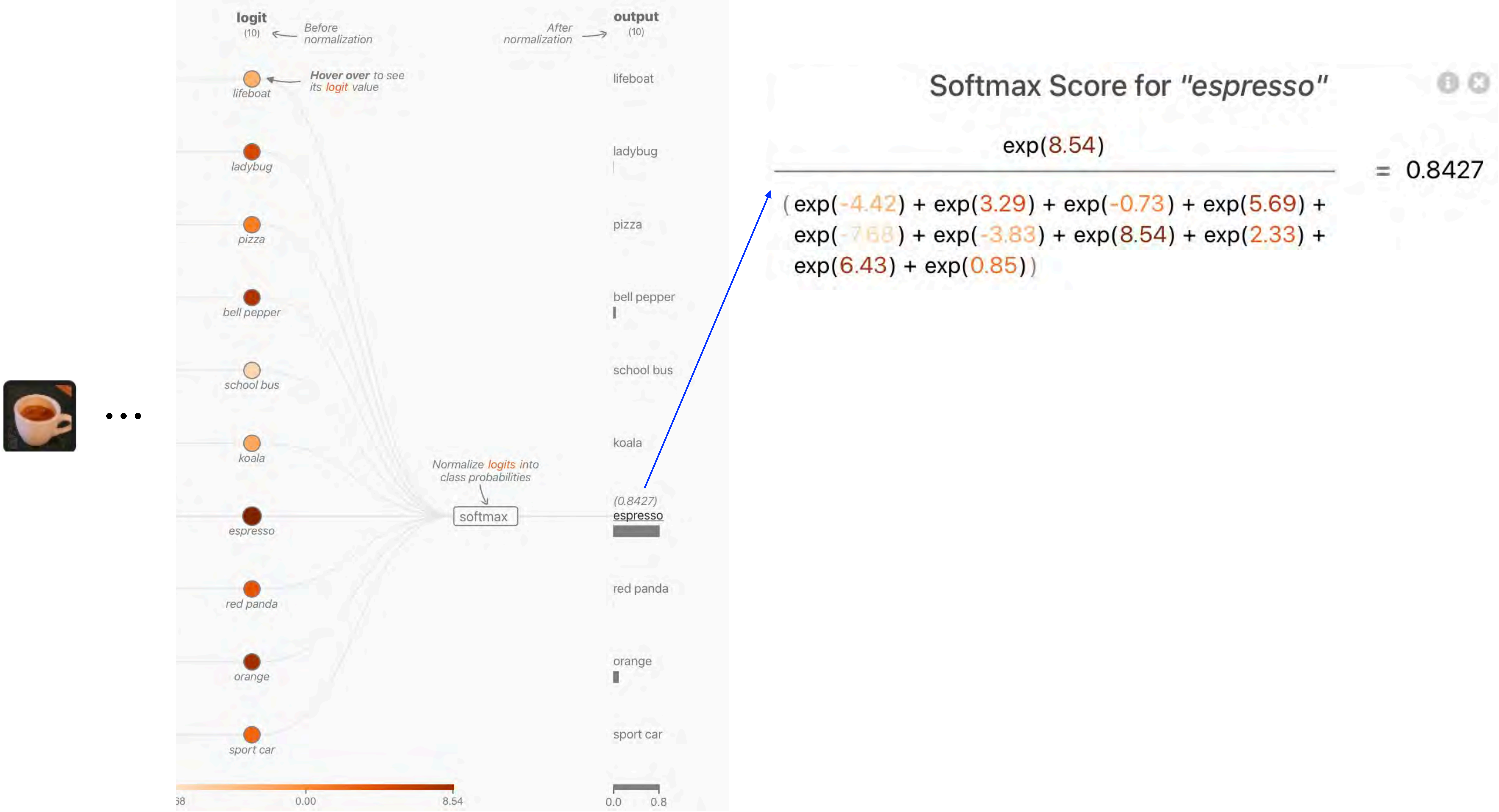


# Adding a Classifier On Top of the CNN

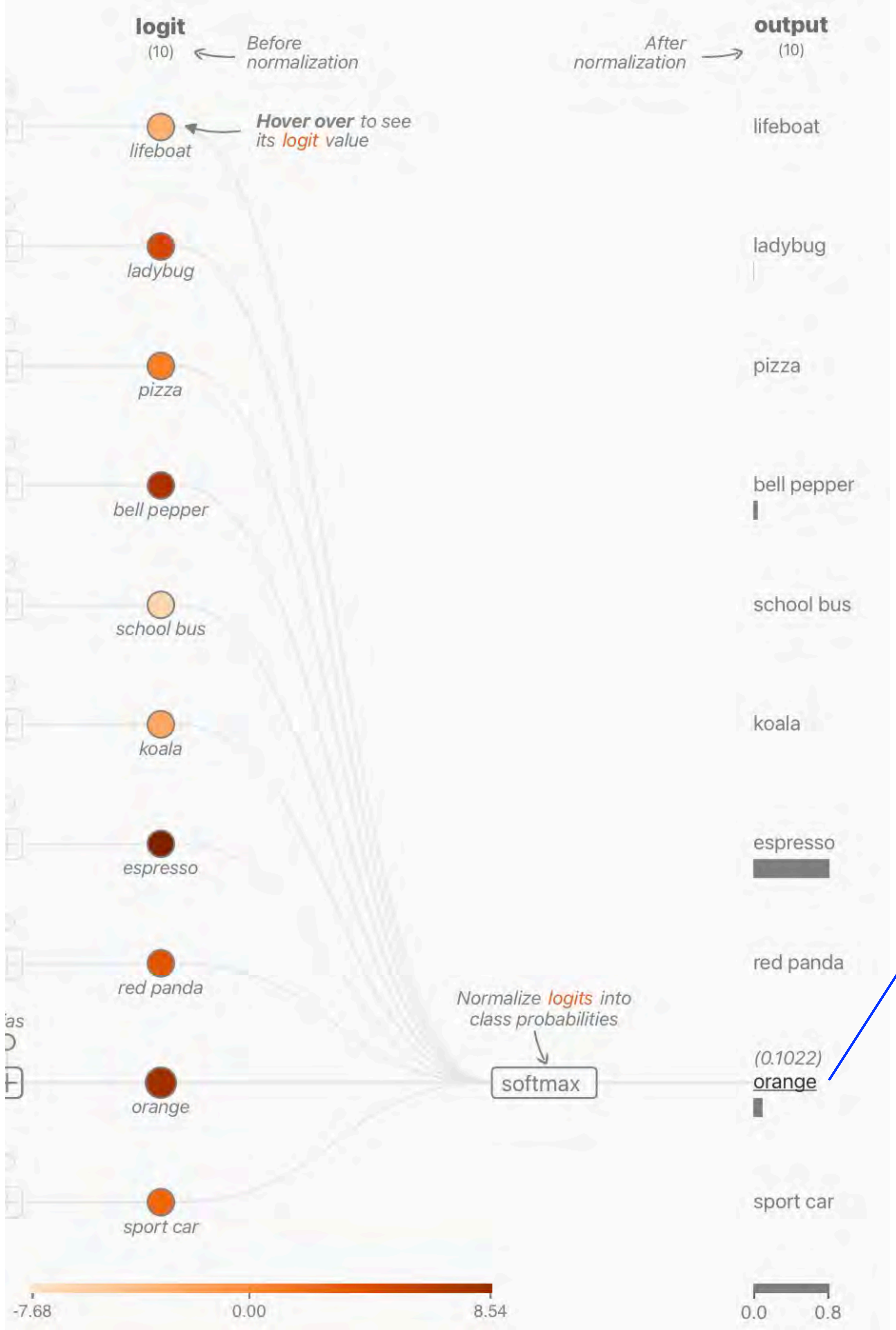
```
layers.Dense(64,  activation = "relu"),
layers.Dense(10, activation = "softmax")
```

- **Softmax** is a function that generates the probabilities of each case out of $N$ cases:

$$\text{Softmax}(y_i) = \frac{e^{y_i}}{e^{y_1} + e^{y_2} + \cdots + e^{y_N}}$$



logit (10)
Before normalization
After normalization
output (10)
Hover over to see its logit value
lifeboat
ladybug
pizza
bell pepper
school bus
koala
espresso
red panda
orange
sport car
Normalize logits into class probabilities
softmax
(0.8427)
0.00
8.54
0.0
0.8
Softmax Score for "espresso"
exp(8.54)
( exp(-4.42) + exp(3.29) + exp(-0.73) + exp(5.69) + exp(-7.68) + exp(-3.83) + exp(8.54) + exp(2.33) + exp(6.43) + exp(0.85) )
= 0.8427

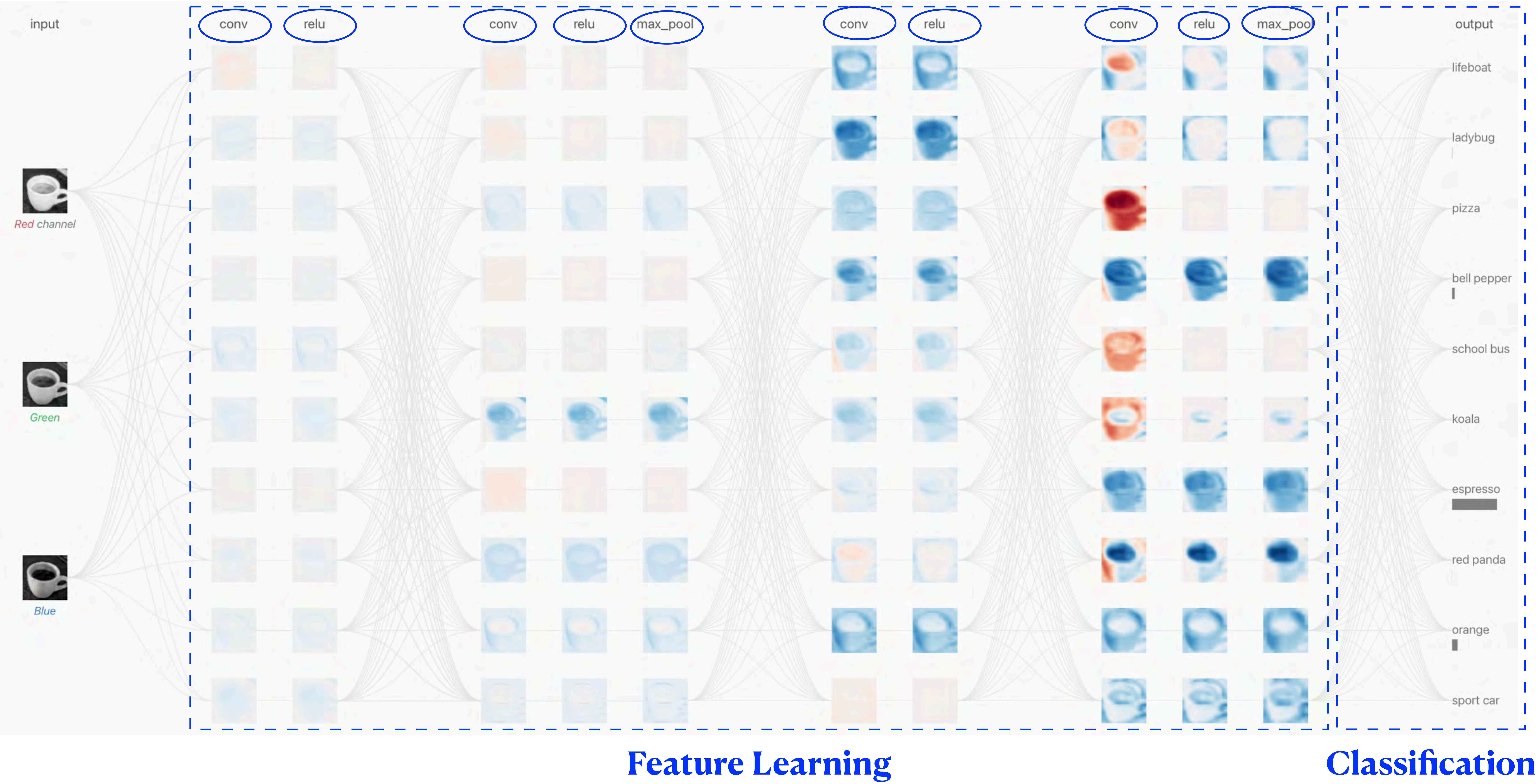



## Softmax Score for *"orange"*

$$\frac{\exp(6.43)}{(\exp(-4.42) + \exp(3.29) + \exp(-0.73) + \exp(5.69) + \exp(-7.68) + \exp(-3.83) + \exp(8.54) + \exp(2.33) + \exp(6.43) + \exp(0.85))} = 0.1022$$

# Exploring CNN (5 Mins)

https://bit.ly/cnnsimu



# CNN MNIST Simulator:

https://stanford.io/3k6U60k

# CNNs in Keras

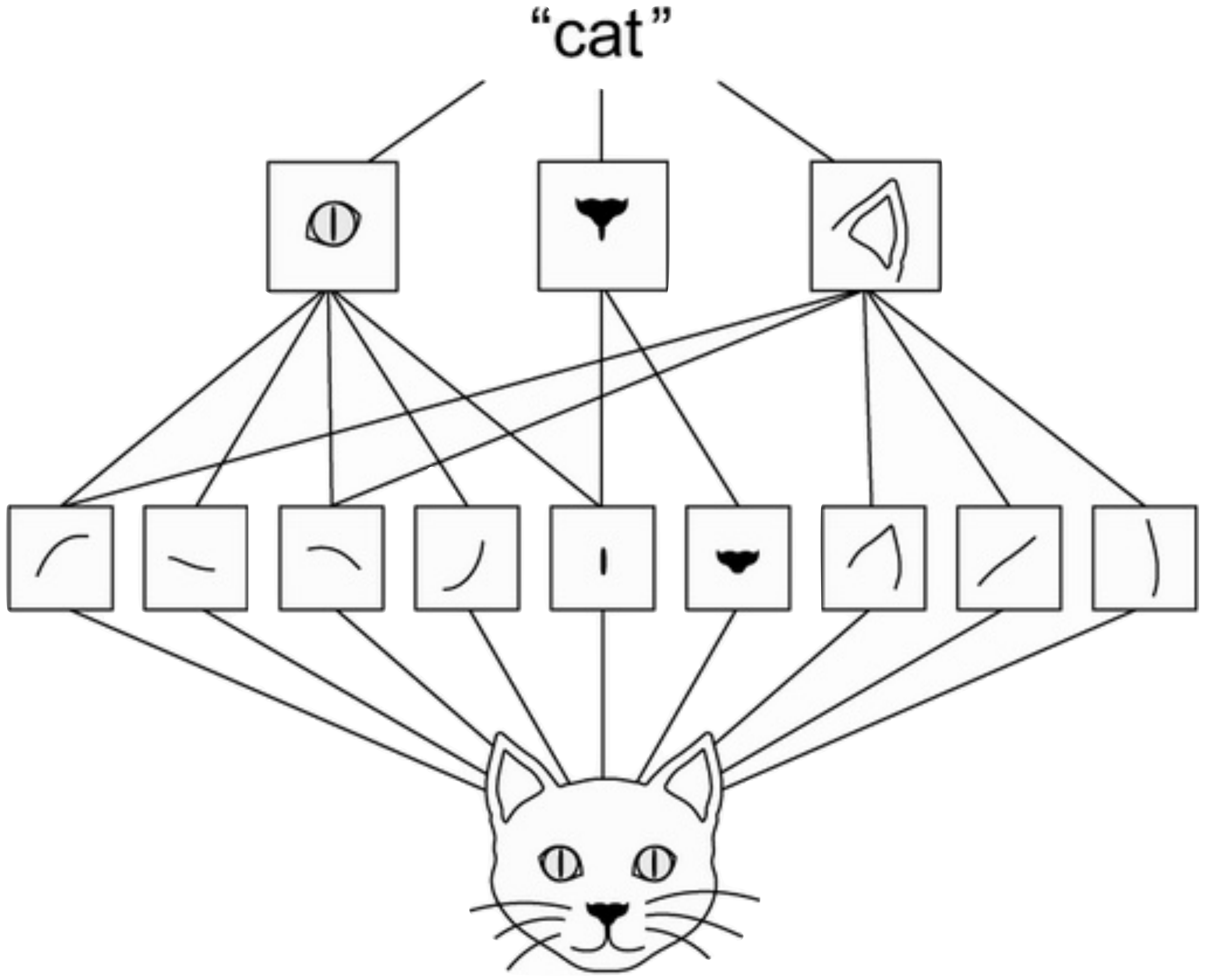
Dimension of the filter (aka kernel) is $3 \times 3$

Number of filters

```python
from tensorflow import keras
from tensorflow.keras import layers
inputs = keras.Input(shape=(28, 28, 1))
x = layers.Conv2D(filters=32, kernel_size=3, activation="relu")(inputs)
x = layers.MaxPooling2D(pool_size=2)(x)
x = layers.Conv2D(filters=64, kernel_size=3, activation="relu")(x)
x = layers.MaxPooling2D(pool_size=2)(x)
x = layers.Conv2D(filters=128, kernel_size=3, activation="relu")(x)
x = layers.Flatten()(x)
outputs = layers.Dense(10, activation="softmax")(x)
model = keras.Model(inputs=inputs, outputs=outputs)
```

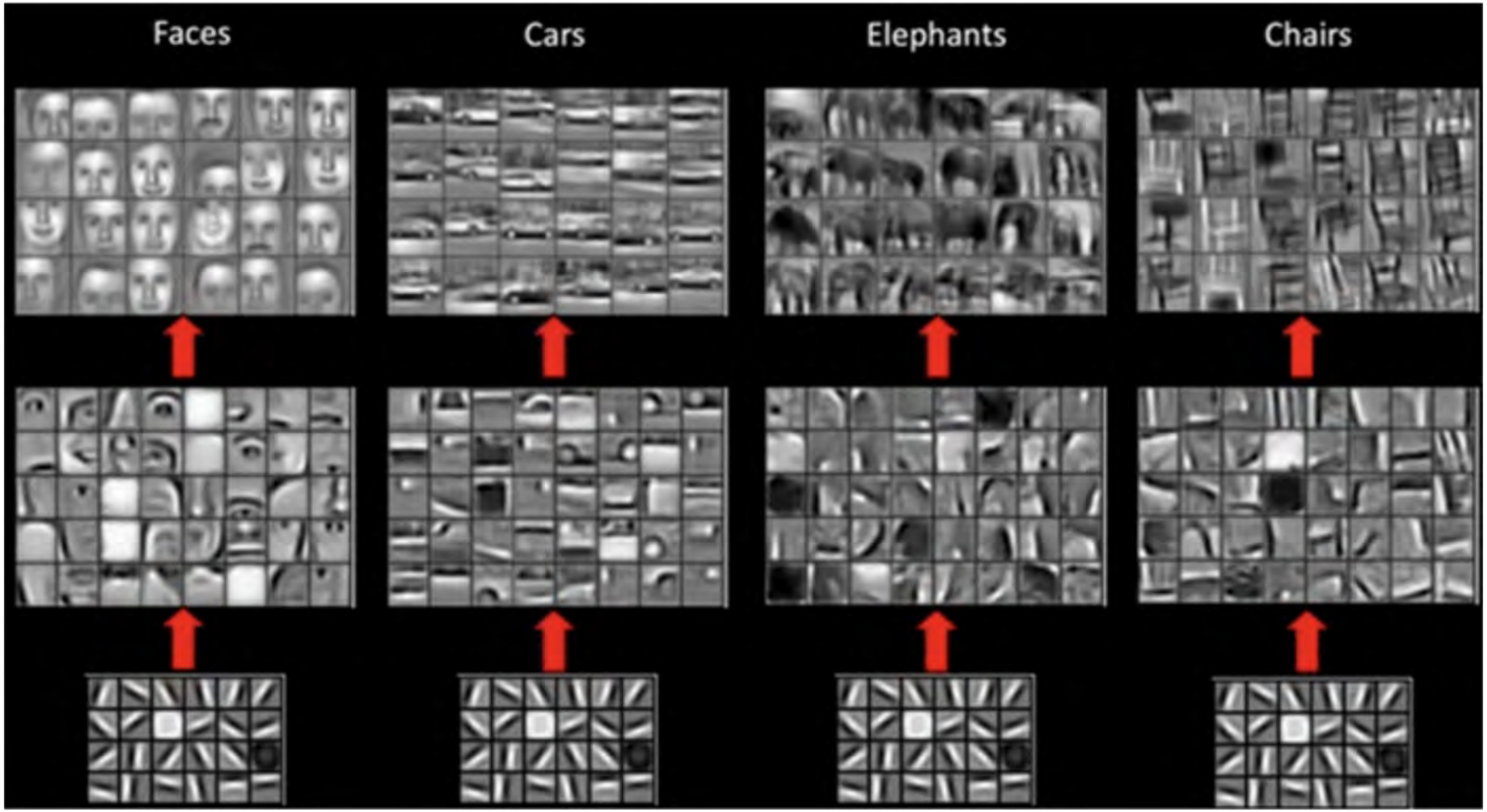
Maxpooling operation with $2 \times 2$ filters and stride 2

# Why Do We Need So Many Convolution Layers?




- To learn spatial hierarchies of patterns
  - A first convolution layer learns small local patterns such as edges
  - A second convolutional layer learns larger patterns made of the features of the first layers
  - More layers allow larger patterns to be captured
  - This allows a CNN to efficiently learn increasingly complex and abstract visual concepts, because **the visual world is fundamentally spatially hierarchical**

# Why Do We Need So Many Convolution Layers?




https://bit.ly/cnnlayers

# CNNs in Keras: Model Summary

```
[1]  from tensorflow import keras
     from tensorflow.keras import layers
     inputs = keras.Input(shape=(28, 28, 1))
     x = layers.Conv2D(filters=32, kernel_size=3, activation="relu")(inputs)
     x = layers.MaxPooling2D(pool_size=2)(x)
     x = layers.Conv2D(filters=64, kernel_size=3, activation="relu")(x)
     x = layers.MaxPooling2D(pool_size=2)(x)
     x = layers.Conv2D(filters=128, kernel_size=3, activation="relu")(x)
     x = layers.Flatten()(x)
     outputs = layers.Dense(10, activation="softmax")(x)
     model = keras.Model(inputs=inputs, outputs=outputs)
```

```
model.summary()

Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 28, 28, 1)]       0

 conv2d (Conv2D)             (None, 26, 26, 32)        320

 max_pooling2d (MaxPooling2D  (None, 13, 13, 32)       0
 )

 conv2d_1 (Conv2D)           (None, 11, 11, 64)        18496

 max_pooling2d_1 (MaxPooling  (None, 5, 5, 64)         0
 2D)

 conv2d_2 (Conv2D)           (None, 3, 3, 128)         73856

 flatten (Flatten)           (None, 1152)              0

 dense (Dense)               (None, 10)                11530

=================================================================
```

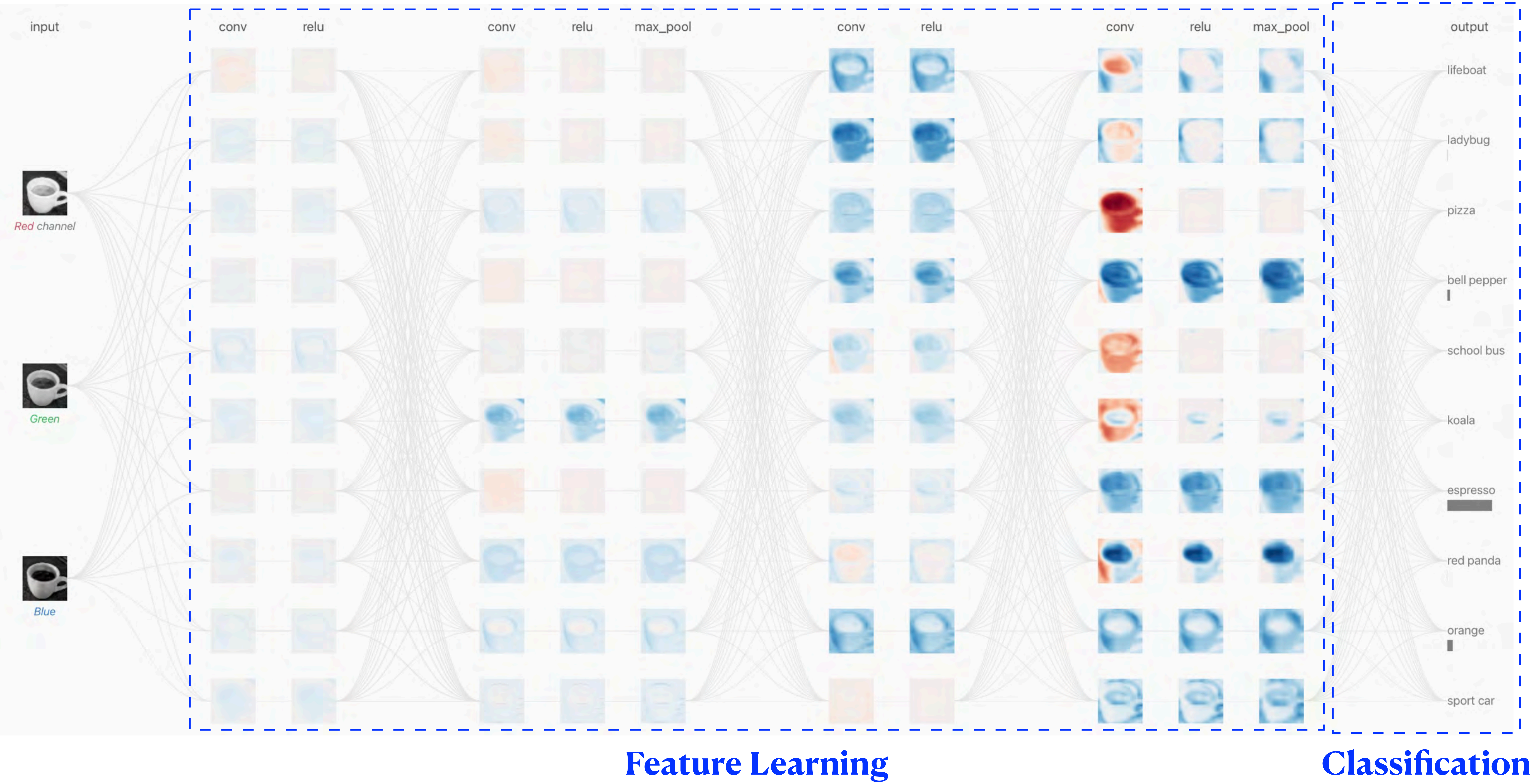
https://bit.ly/session2cnn

# Let’s now use CNNs to classify the MNIST dataset

https://bit.ly/session3cnn

# A Fully Connected CNN

https://bit.ly/cnnsimu



# CNNs and Small Data

- Myth: Deep learning *always* requires a large amount of data

- Yes, a reasonably large sample is often needed for a complex problem

- However, CNNs are capable of learning local, translation-invariant features; they're highly data efficient on perceptual problems

- Training a CNN from scratch on a very small image dataset (e.g., 10 images) can often yield reasonably satisfactory performance despite a lack of big data

https://teachablemachine.withgoogle.com/

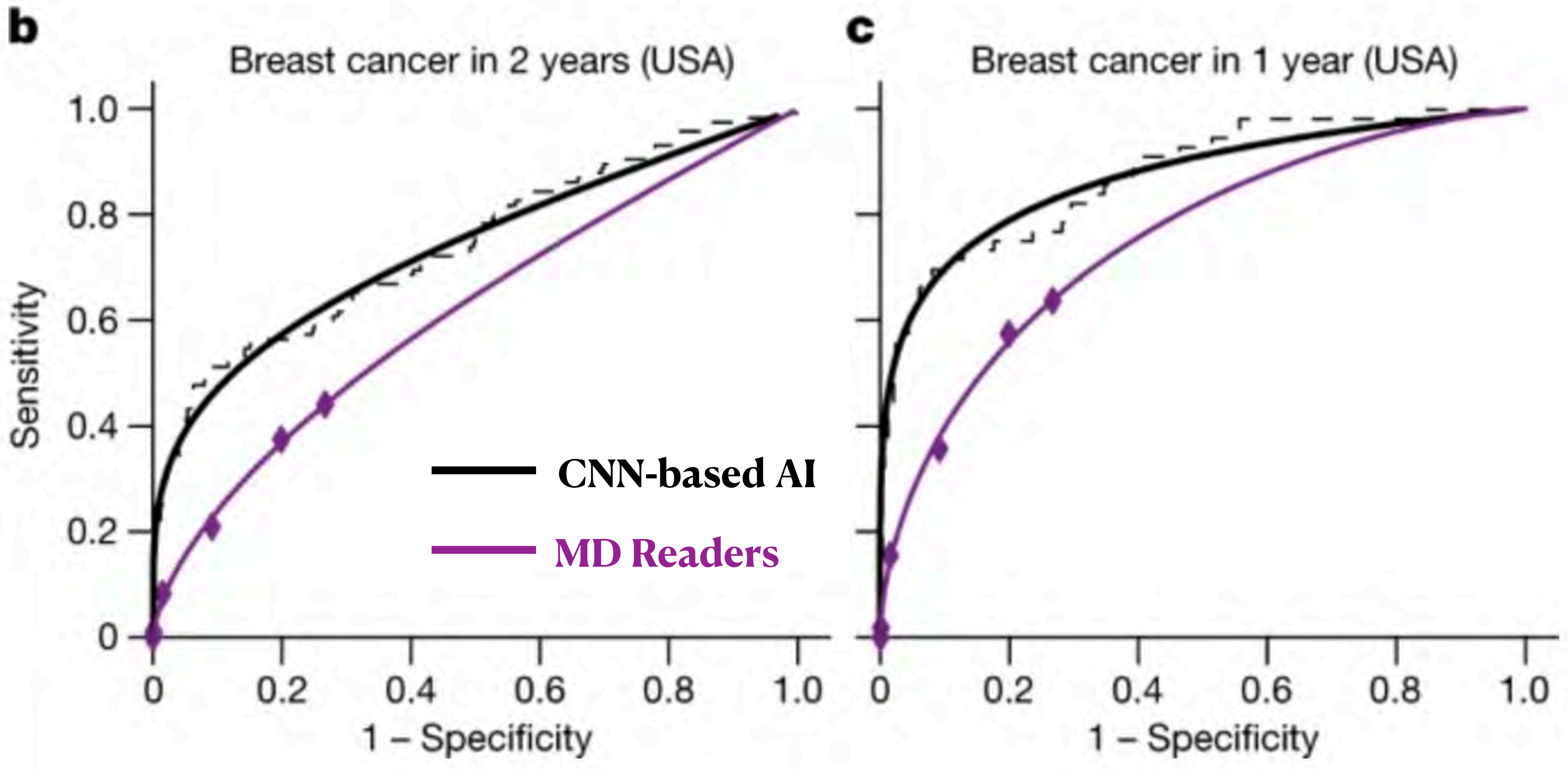
# Medical Artificial Intelligence

# AI Applications in Diagnosis

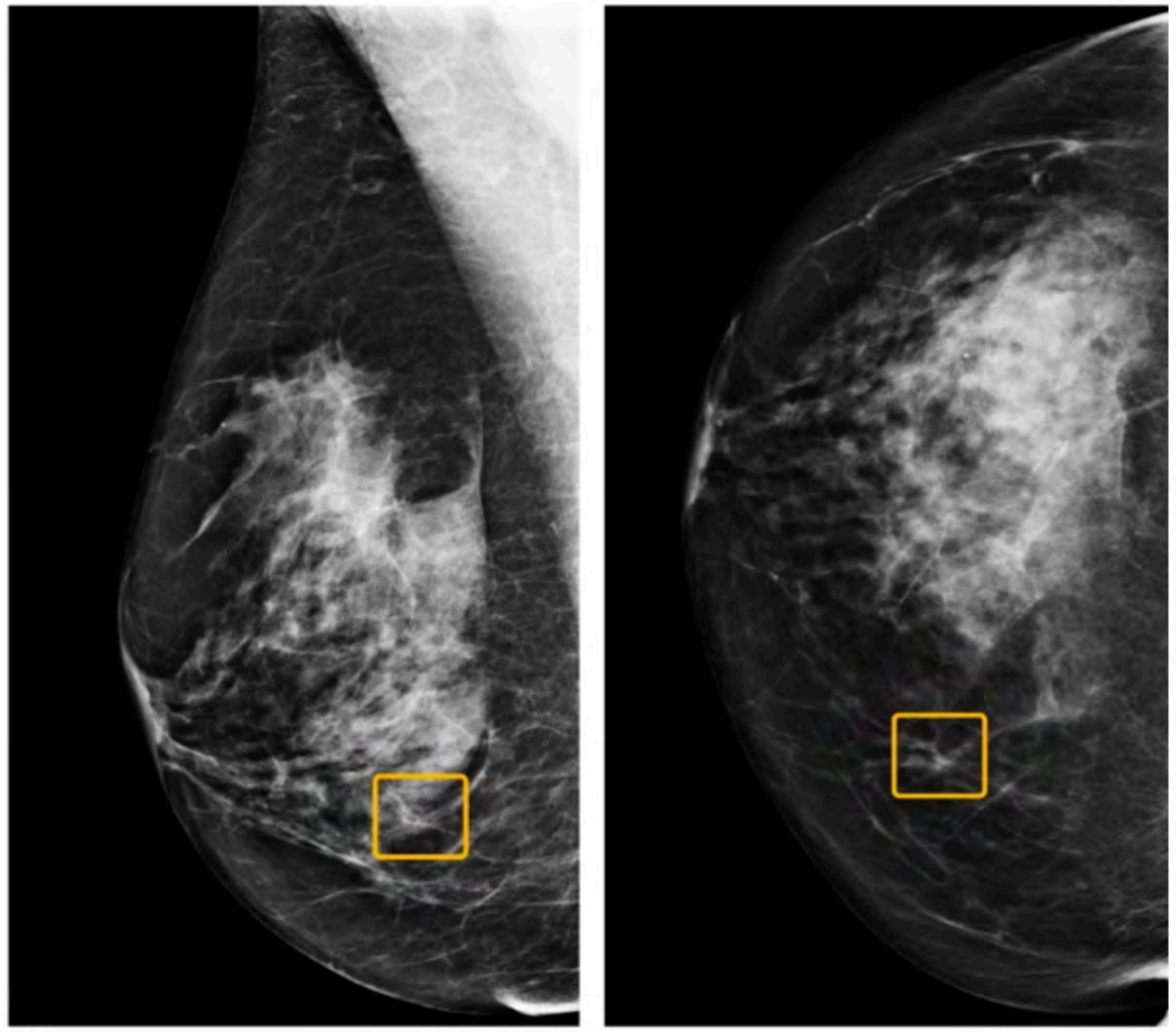
**nature**

Article | Published: 01 January 2020

## International evaluation of an AI system for breast cancer screening



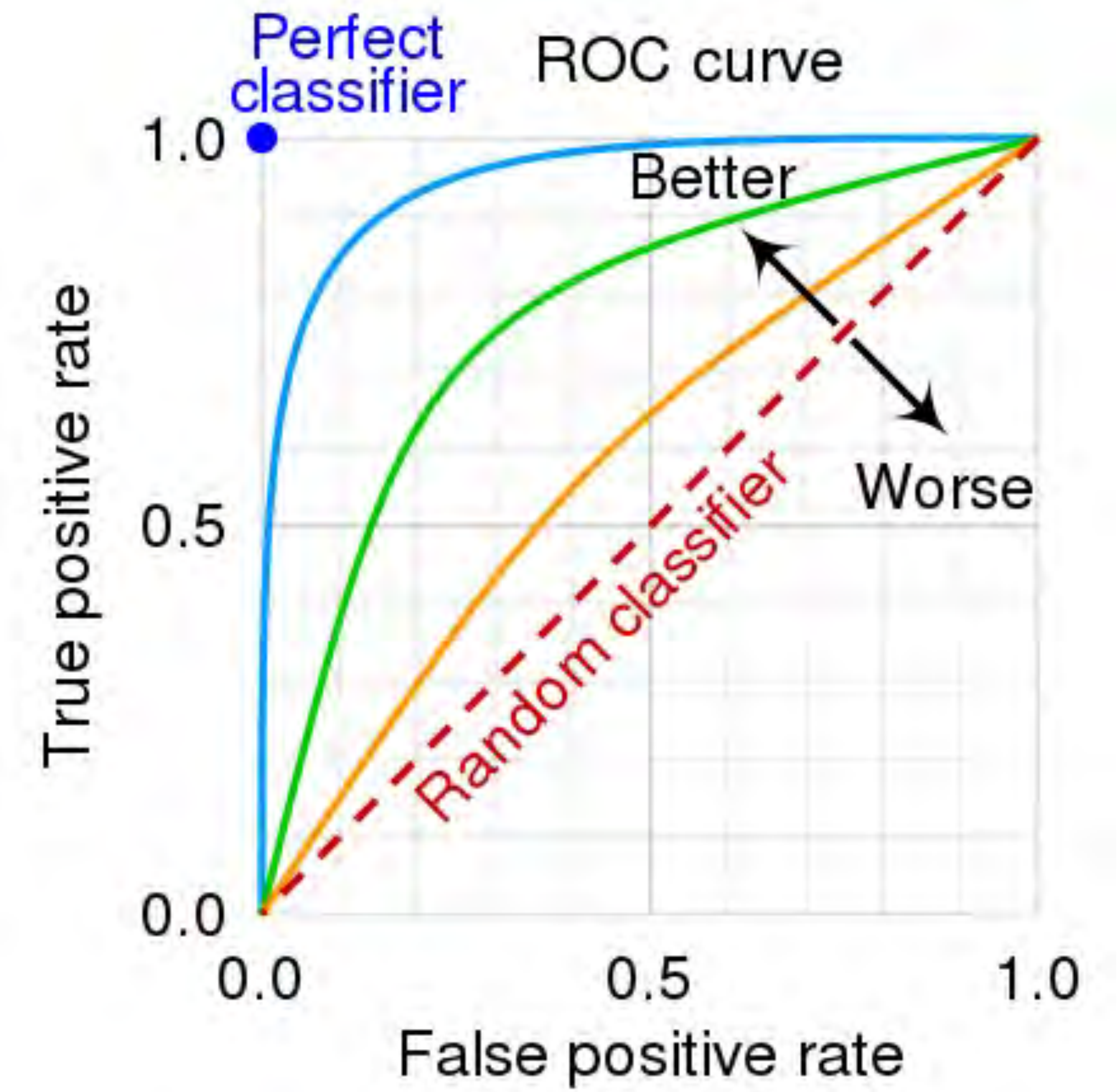

CNN-based AI beats human experts in diagnosing breast cancer using mammograms



Breast cancer cases missed by human doctors but detected by CNNs

# Area Under Curve (AUC)

**ROC Curve = Receiver Operating Characteristic Curve**



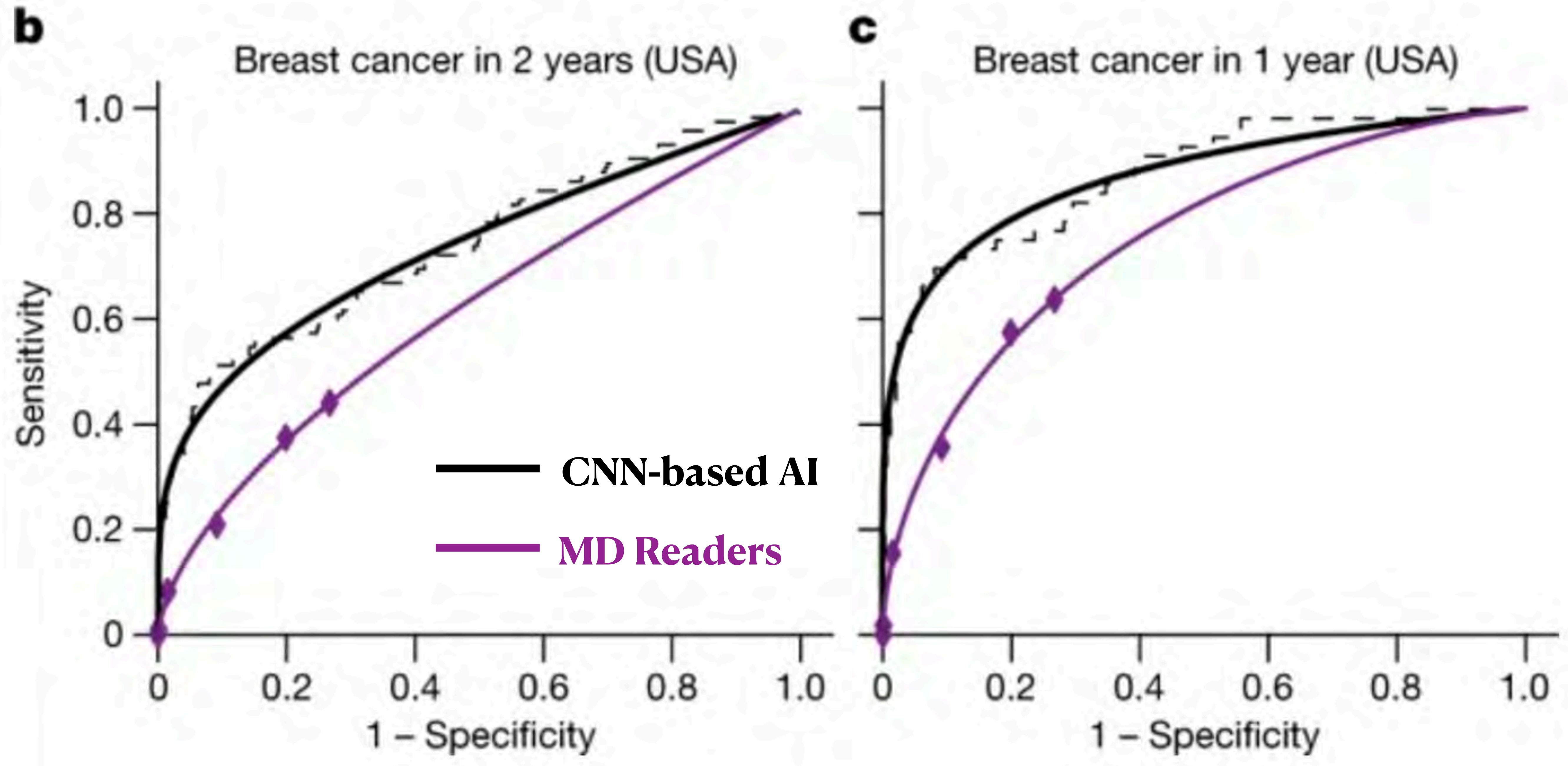

- Under a random classifier, the area under the ROC curve is 50% of the plane
- Under a perfect classifier, the area under the ROC curve is 100% of the plane
- A bigger "area under curve" (AUC) corresponds to a better diagnostic system

# Using AUC to Compare Human vs. MD

**nature**

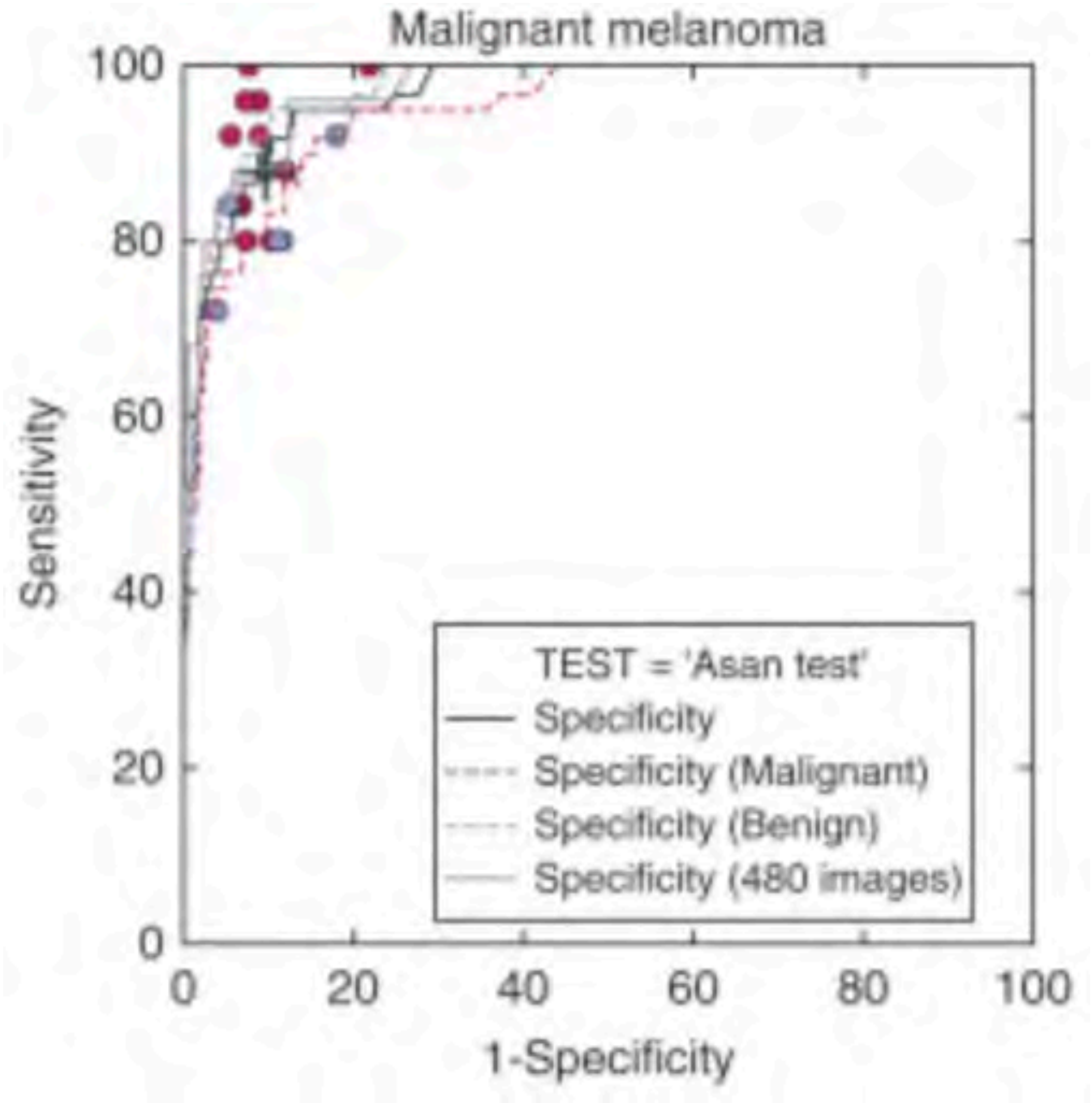
Article | Published: 01 January 2020

## International evaluation of an AI system for breast cancer screening




- AUC of MD readers = 0.625
- AUC of the AI system = 0.740
- "The AI system exceeded human performance by a significant margin…"

# Using Biopsies to Diagnose Cutaneous Melanoma




JOURNAL OF INVESTIGATIVE DERMATOLOGY

ORIGINAL ARTICLE CLINICAL RESEARCH: PATIENT OUTCOMES | VOLUME 138, ISSUE 7, P1529-1538, JULY 01, 2018

Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm

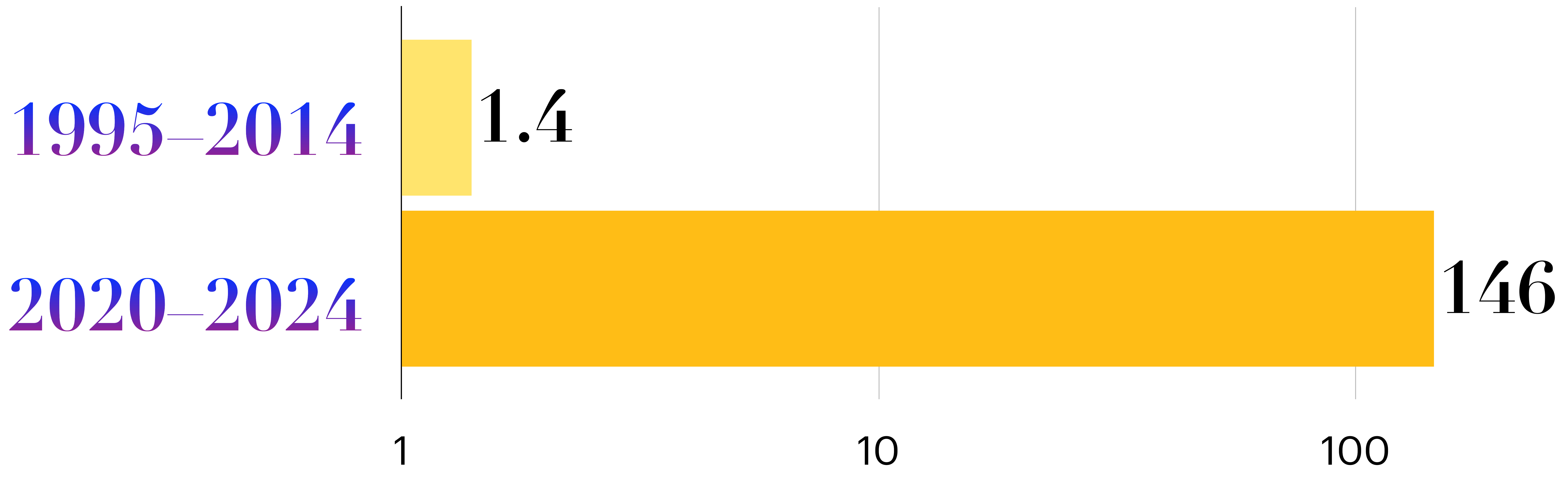
The CNN-based algorithm achieved an AUC of 96% for melanoma

# 950

**Number of medical AI devices cleared by FDA for clinical use** (as of June 2024)

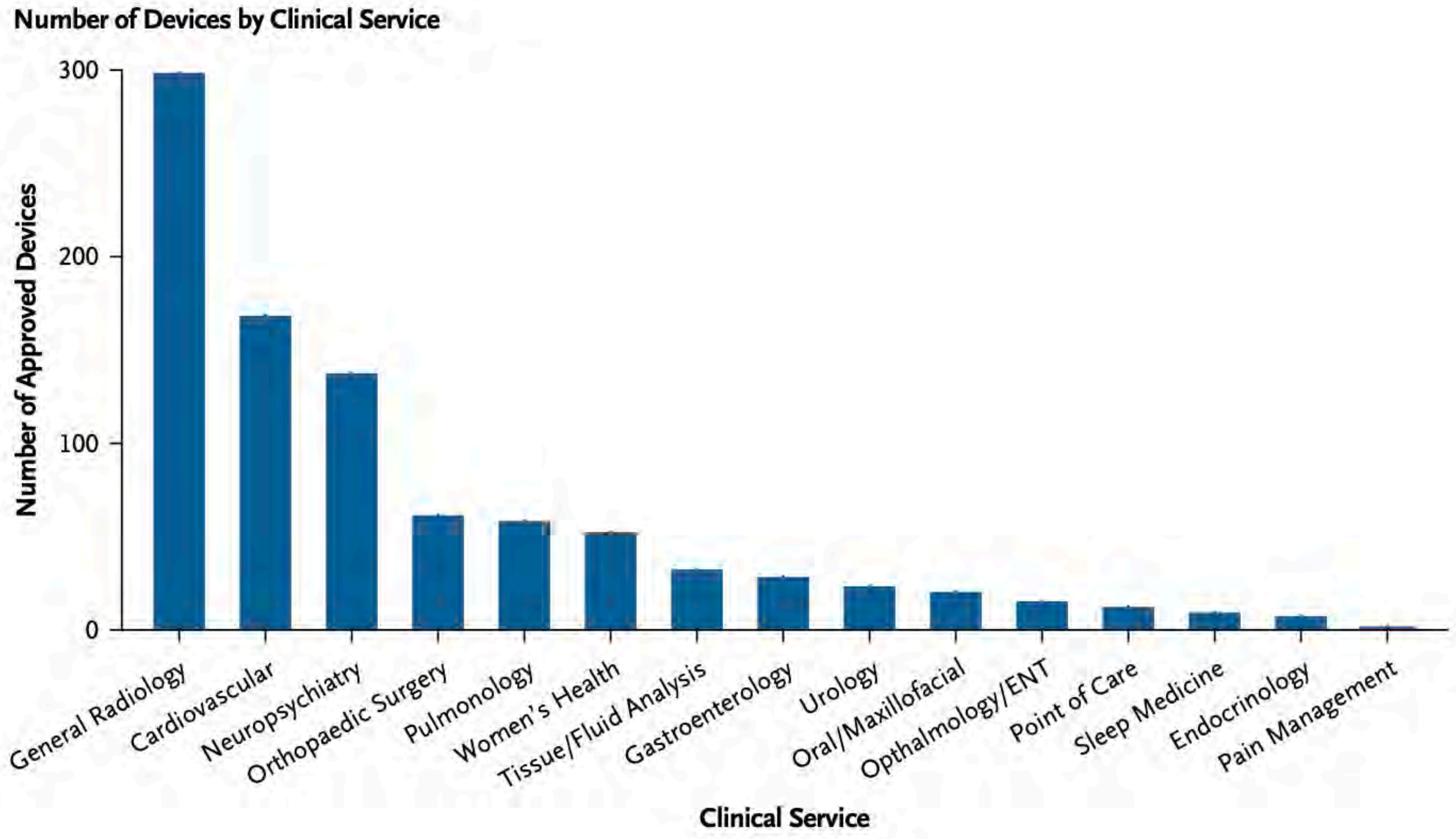
NEJM AI

# Number of FDA-Cleared AI Devices Per Year



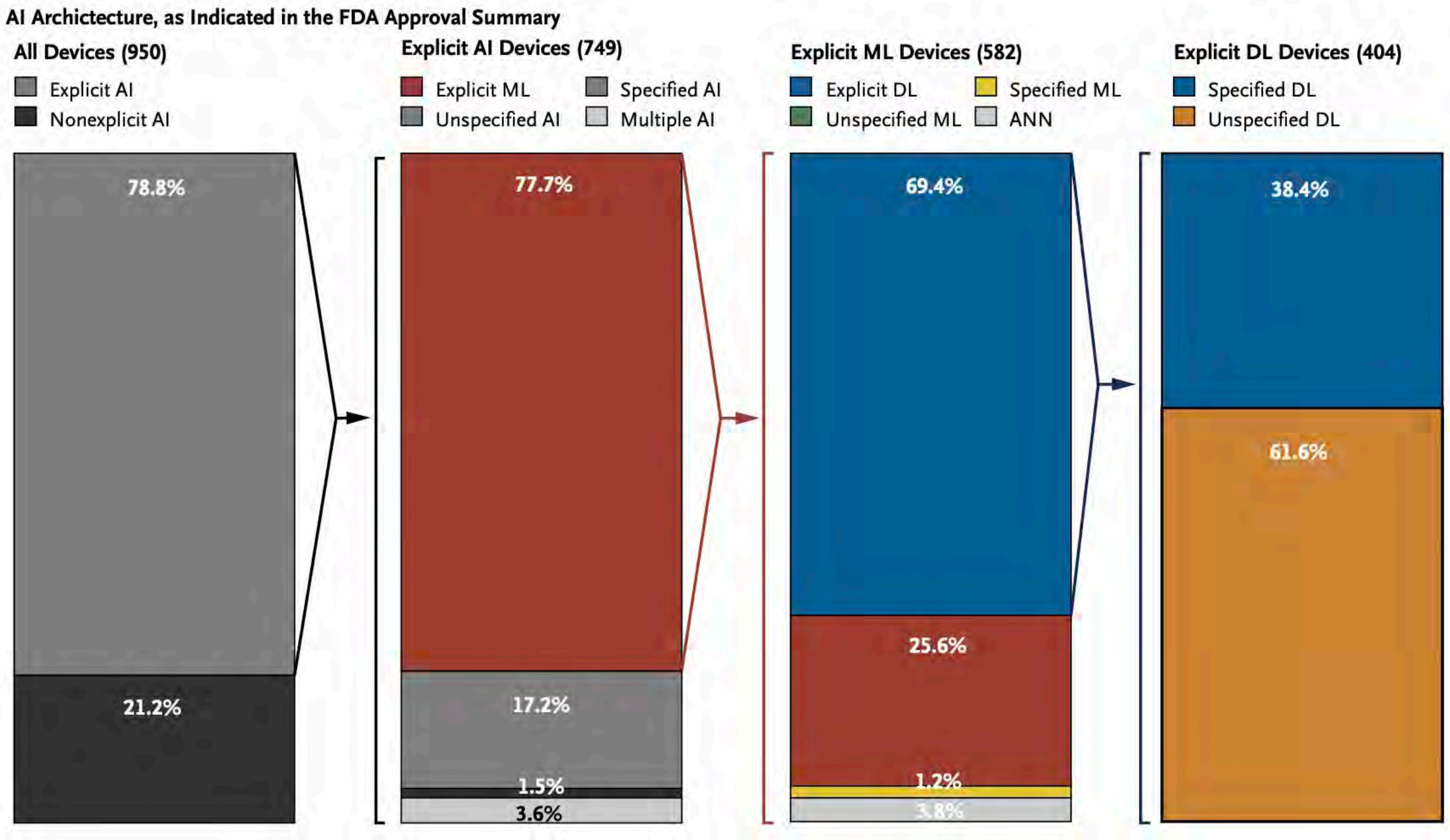

**Development and Commercialization Pathways of AI Medical Devices in the United States: Implications for Safety and Regulatory Oversight**

**Authors**: Branden Lee, B.S., Shivam Patel, B.S., Crystal Favorito, B.S., Sara Sandri, B.S., Maria Rain Jennings, Ph.D., and Tinglong Dai, Ph.D.



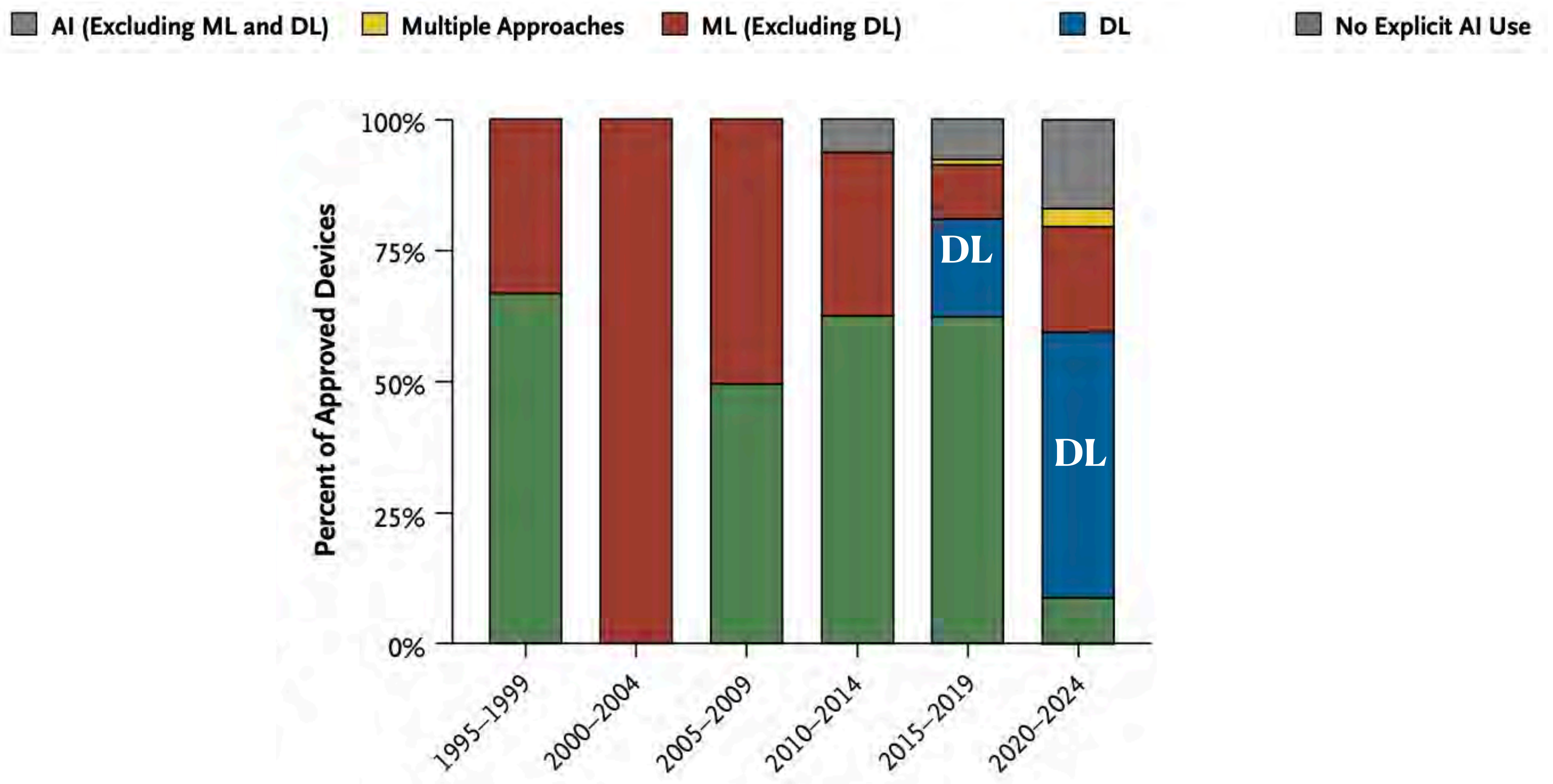

Lee B, Patel S, Favorito C, Sandri S, Jennings MR, Dai T. 2025. *NEJM AI*.

# AI Methods Used in FDA-Cleared AI Devices



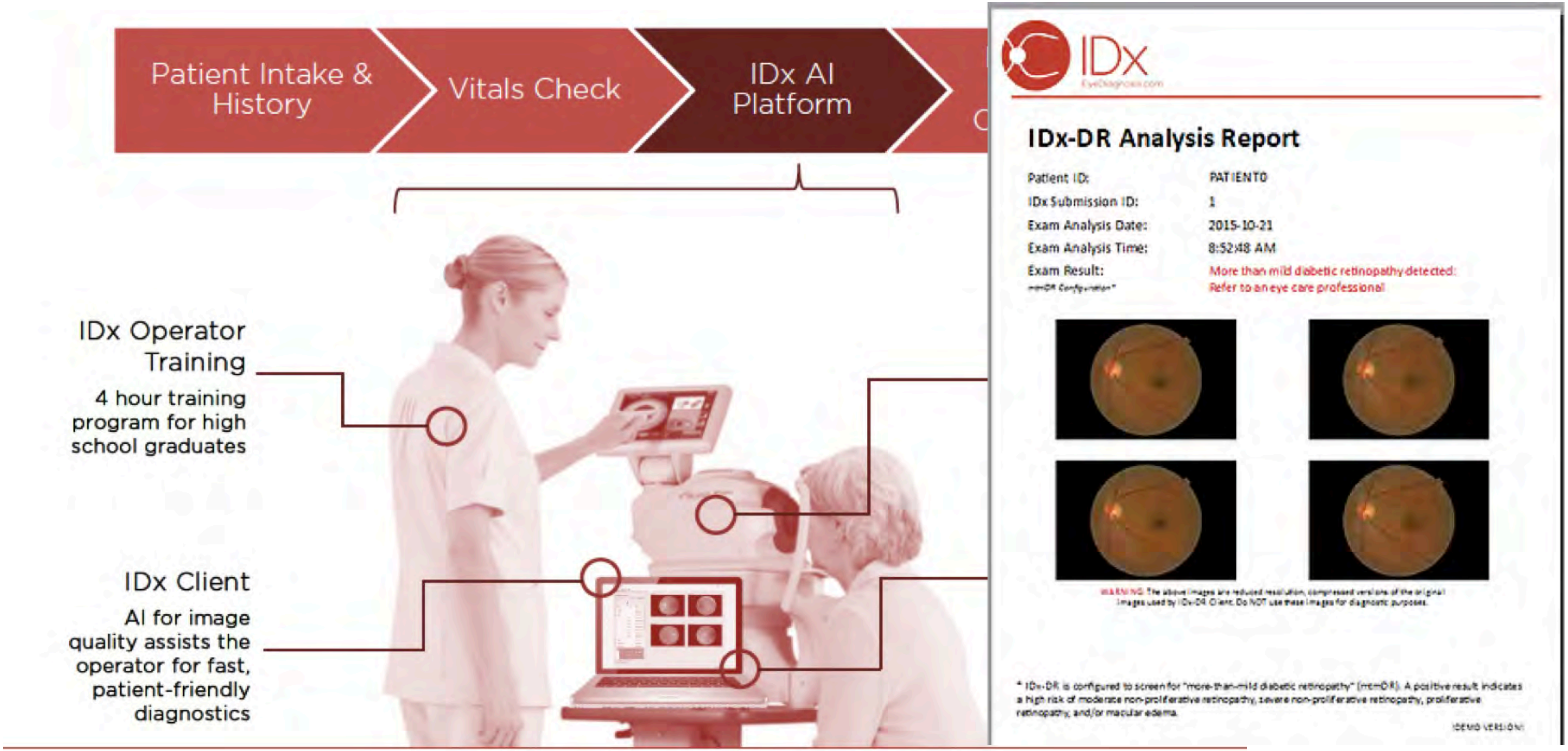

Lee B, Patel S, Favorito C, Sandri S, Jennings MR, Dai T. 2025. *NEJM AI*.

# AI Methods by Commercial Model and Over Time




Lee B, Patel S, Favorito C, Sandri S, Jennings MR, Dai T. 2025. *NEJM AI*.

# IDx-DR AI for Diabetic Retinopathy Screening



# Johns Hopkins Pediatric Retinopathy Study Using IDx-DR

**(Dec 2018 – Nov 2019; 310 Patients Enrolled)**

Real-world performance of IDx-DR AI system:

- Sensitivity ("safety"): 85.7 % (95% CI: 42.1%- 99.6%)
- Specificity ("effectiveness"): 79.3 % (95% CI: 74.3%-83.8%)

Compared to a 35%–91% sensitivity and a 95% specificity for clinical experts

Improved **patient adherence** from 49% to 95%

Wolf et al. *Diabetes Care* 2021; Abramoff et al. *NPJ Digit Med* 2018; Verbraak et al. *Diab Care* 2019

# ACCESS: AI for pediatriC diabetiC Eye examS Study

## NIH-Funded Ongoing Study at Johns Hopkins (2021–2023)

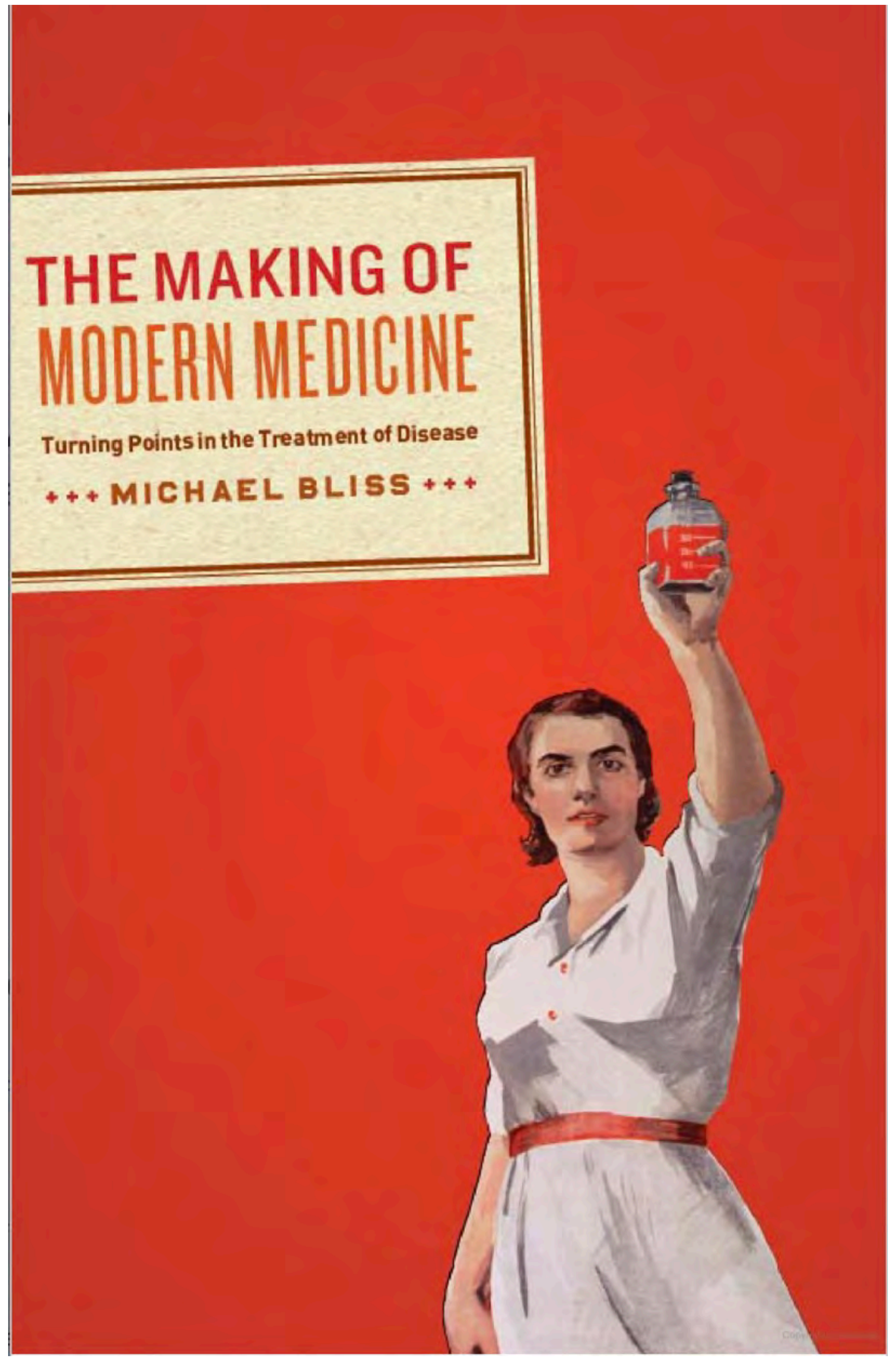
- Aim: Determine if point-of-care autonomous AI improves **screening rates** compared to standard in-person eye care professional exams in a randomized controlled trial
- All participants randomized (n=164):
  - Group 1: Standard of care (SOC) referral to ECP + educational intervention
  - Group 2: POC autonomous AI retinopathy screening outcomes
- Primary outcome: % participants who get screened (by AI or ECP) within 6 months
- Key finding: **22% screened in Group 1 vs. 100% screened in Group 2**
- Cost-effectiveness analysis in progress

To understand how AI will impact medicine, we need to understand the *making* of modern medicine

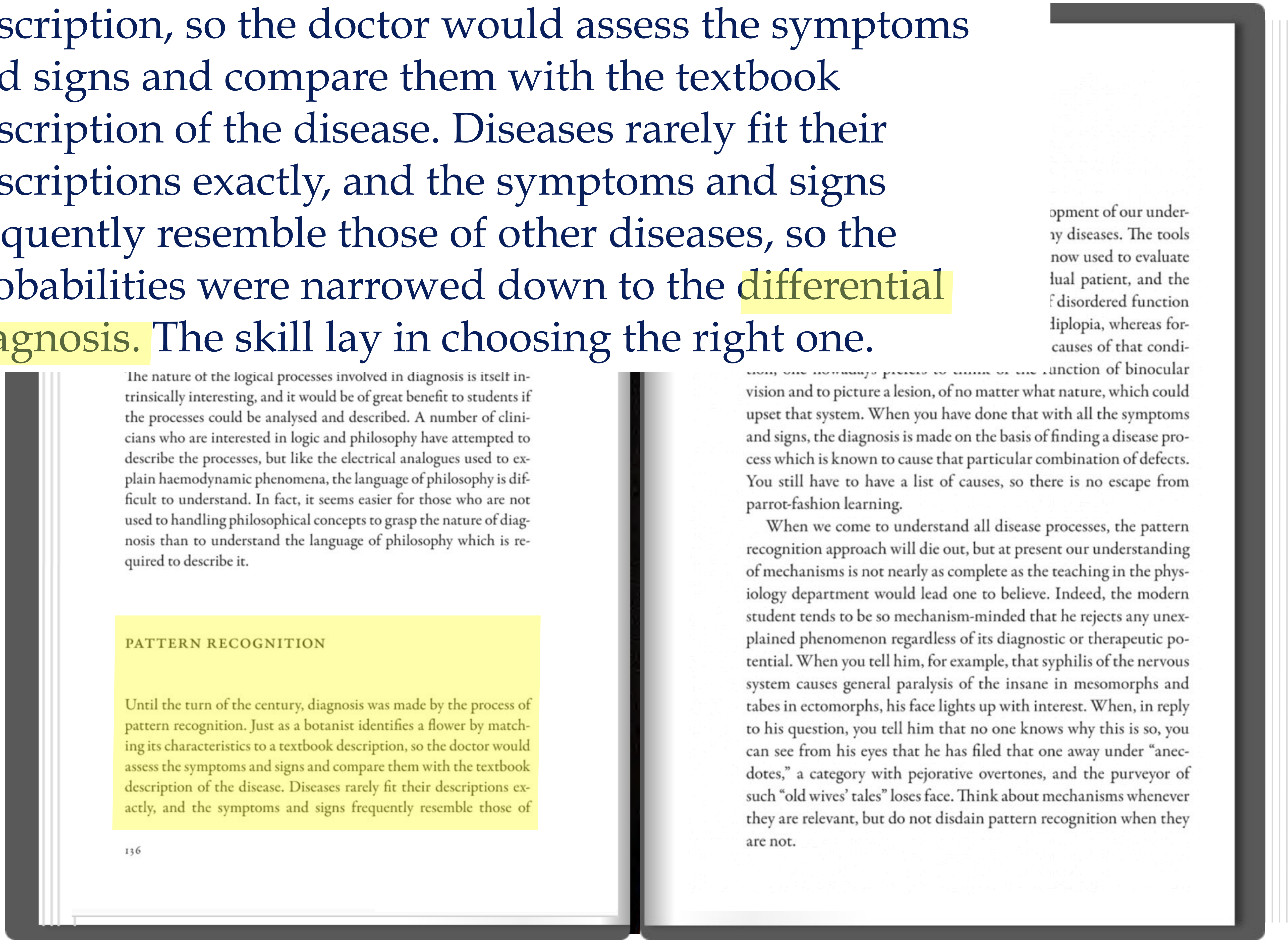
# Medicine in North America by Late 1860s

## Michael Bliss (2011): *The Making of Modern Medicine*

- "Anything and anyone" can "flourish in healthcare"
- "Standards of training and licensing had virtually disappeared in a democratic country where *anyone* was good enough to *try* to be a doctor"
- Many medical doctors "simply bought or invented their paper credentials"
- Even if they actually earned their credentials, most U.S. medical schools, Harvard included, were "notorious medical diploma mills"



Until the turn of the century, diagnosis was made by the process of pattern recognition. Just as a botanist identifies a flower by matching its characteristics to a textbook description, so the doctor would assess the symptoms and signs and compare them with the textbook description of the disease. Diseases rarely fit their descriptions exactly, and the symptoms and signs frequently resemble those of other diseases, so the probabilities were narrowed down to the differential diagnosis. The skill lay in choosing the right one.



The nature of the logical processes involved in diagnosis is itself intrinsically interesting, and it would be of great benefit to students if the processes could be analysed and described. A number of clinicians who are interested in logic and philosophy have attempted to describe the processes, but like the electrical analogues used to explain haemodynamic phenomena, the language of philosophy is difficult to understand. In fact, it seems easier for those who are not used to handling philosophical concepts to grasp the nature of diagnosis than to understand the language of philosophy which is required to describe it.

PATTERN RECOGNITION

Until the turn of the century, diagnosis was made by the process of pattern recognition. Just as a botanist identifies a flower by matching its characteristics to a textbook description, so the doctor would assess the symptoms and signs and compare them with the textbook description of the disease. Diseases rarely fit their descriptions exactly, and the symptoms and signs frequently resemble those of

136

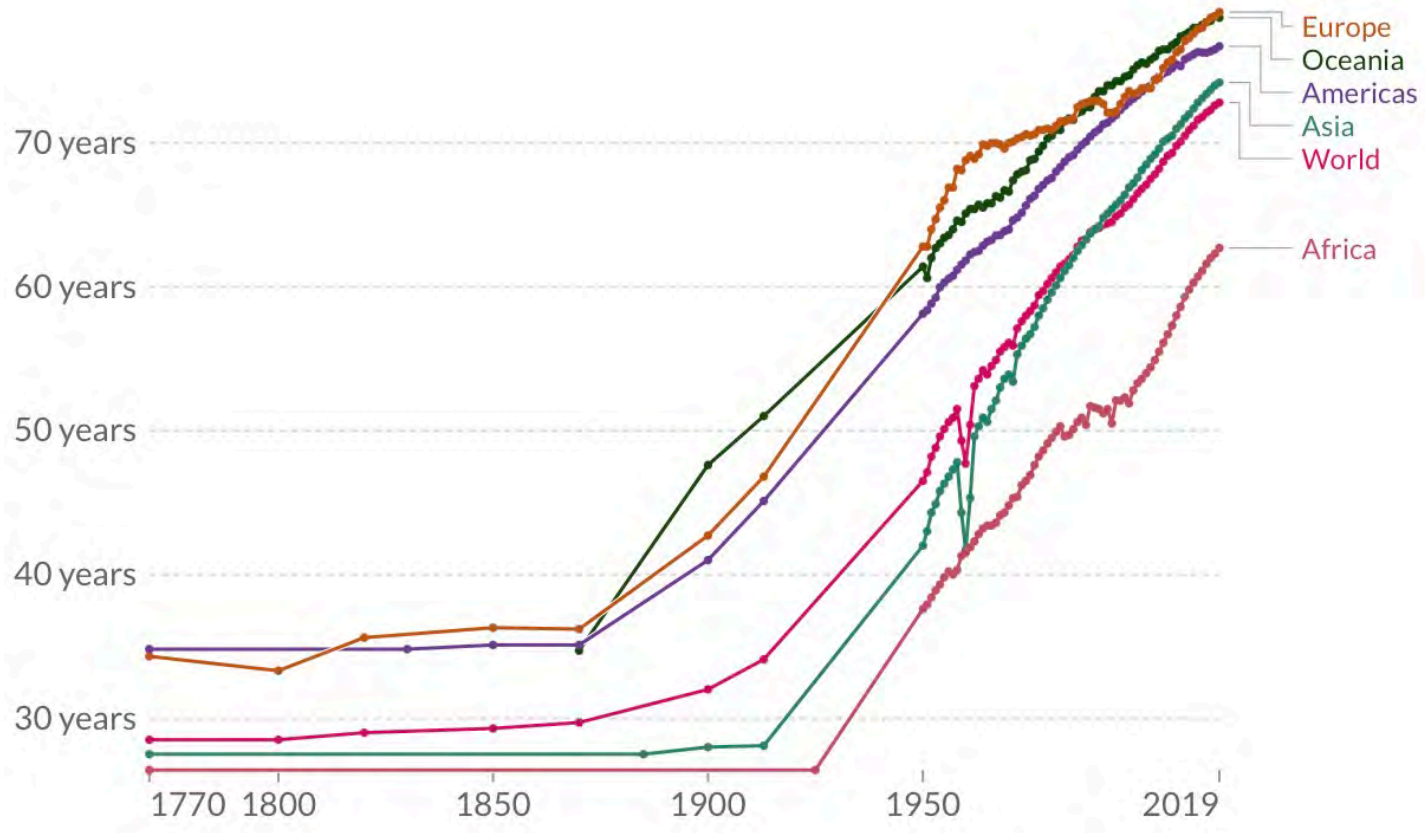
…opment of our under-
…y diseases. The tools
…now used to evaluate
…lual patient, and the
…f disordered function
…diplopia, whereas for-
…causes of that condi-
…unction of binocular vision and to picture a lesion, of no matter what nature, which could upset that system. When you have done that with all the symptoms and signs, the diagnosis is made on the basis of finding a disease process which is known to cause that particular combination of defects. You still have to have a list of causes, so there is no escape from parrot-fashion learning.

When we come to understand all disease processes, the pattern recognition approach will die out, but at present our understanding of mechanisms is not nearly as complete as the teaching in the physiology department would lead one to believe. Indeed, the modern student tends to be so mechanism-minded that he rejects any unexplained phenomenon regardless of its diagnostic or therapeutic potential. When you tell him, for example, that syphilis of the nervous system causes general paralysis of the insane in mesomorphs and tabes in ectomorphs, his face lights up with interest. When, in reply to his question, you tell him that no one knows why this is so, you can see from his eyes that he has filed that one away under "anecdotes," a category with pejorative overtones, and the purveyor of such "old wives' tales" loses face. Think about mechanisms whenever they are relevant, but do not disdain pattern recognition when they are not.

David Mendel (1984): *Proper Doctoring*

# The Life-Saving Power of Modern Medicine

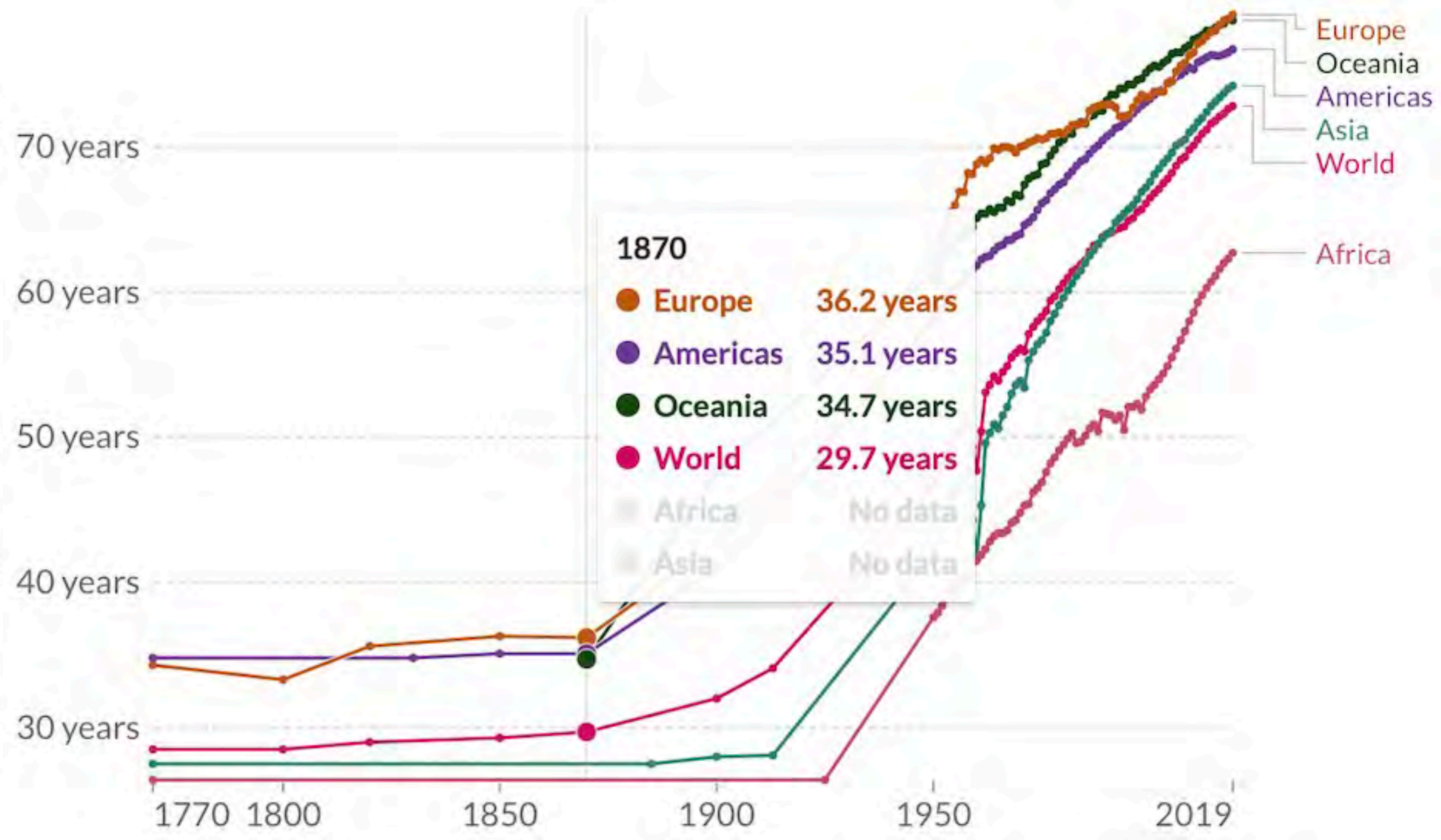
## Life Expectancy, 1770 to 2019




Data Source: UN WPP (2022); Zijdeman et al. (2015); Riley (2005); https://ourworldindata.org/life-expectancy

# The Life-Saving Power of Modern Medicine

## Life Expectancy, 1770 to 2019



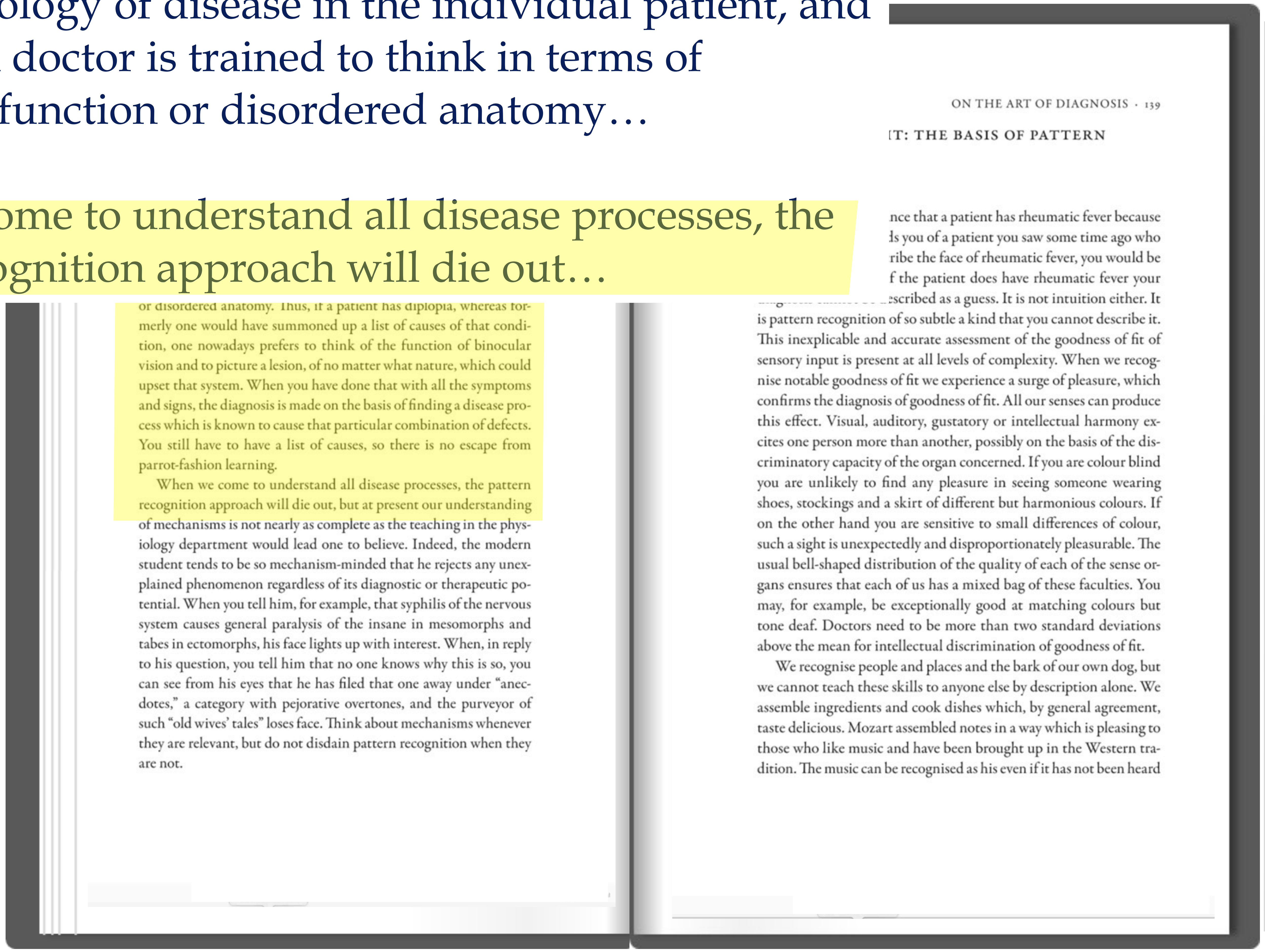

Data Source: UN WPP (2022); Zijdeman et al. (2015); Riley (2005); https://ourworldindata.org/life-expectancy

# 1893

**The year in which the Johns Hopkins School of Medicine, "a temple of healing" using scientific approaches, was established, which marked the beginning of modern medicine in North America** (Bliss 2011)

In this century there has been enormous development of our understanding of the causes and mechanism of many diseases. The tools which were used to investigate physiology are now used to evaluate the pathophysiology of disease in the individual patient, and the modern doctor is trained to think in terms of disordered function or disordered anatomy…

When we come to understand all disease processes, the pattern recognition approach will die out…



or disordered anatomy. Thus, if a patient has diplopia, whereas formerly one would have summoned up a list of causes of that condition, one nowadays prefers to think of the function of binocular vision and to picture a lesion, of no matter what nature, which could upset that system. When you have done that with all the symptoms and signs, the diagnosis is made on the basis of finding a disease process which is known to cause that particular combination of defects. You still have to have a list of causes, so there is no escape from parrot-fashion learning.

When we come to understand all disease processes, the pattern recognition approach will die out, but at present our understanding of mechanisms is not nearly as complete as the teaching in the physiology department would lead one to believe. Indeed, the modern student tends to be so mechanism-minded that he rejects any unexplained phenomenon regardless of its diagnostic or therapeutic potential. When you tell him, for example, that syphilis of the nervous system causes general paralysis of the insane in mesomorphs and tabes in ectomorphs, his face lights up with interest. When, in reply to his question, you tell him that no one knows why this is so, you can see from his eyes that he has filed that one away under "anecdotes," a category with pejorative overtones, and the purveyor of such "old wives' tales" loses face. Think about mechanisms whenever they are relevant, but do not disdain pattern recognition when they are not.

ON THE ART OF DIAGNOSIS · 139

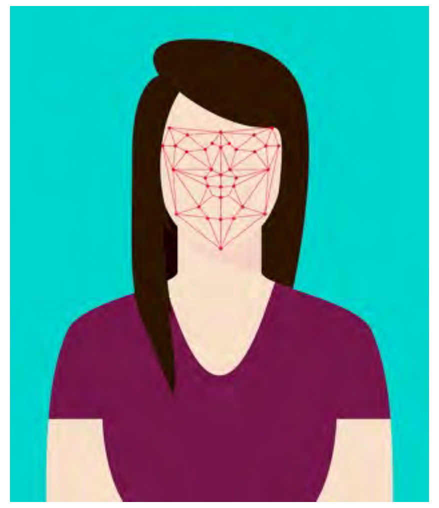
...T: THE BASIS OF PATTERN

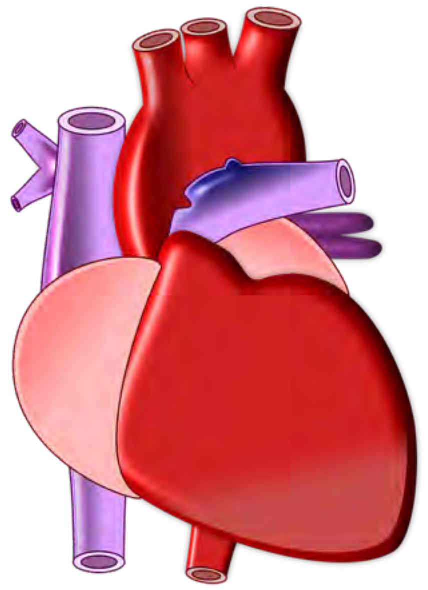
...nce that a patient has rheumatic fever because ...ds you of a patient you saw some time ago who ...ribe the face of rheumatic fever, you would be ...f the patient does have rheumatic fever your ...escribed as a guess. It is not intuition either. It is pattern recognition of so subtle a kind that you cannot describe it. This inexplicable and accurate assessment of the goodness of fit of sensory input is present at all levels of complexity. When we recognise notable goodness of fit we experience a surge of pleasure, which confirms the diagnosis of goodness of fit. All our senses can produce this effect. Visual, auditory, gustatory or intellectual harmony excites one person more than another, possibly on the basis of the discriminatory capacity of the organ concerned. If you are colour blind you are unlikely to find any pleasure in seeing someone wearing shoes, stockings and a skirt of different but harmonious colours. If on the other hand you are sensitive to small differences of colour, such a sight is unexpectedly and disproportionately pleasurable. The usual bell-shaped distribution of the quality of each of the sense organs ensures that each of us has a mixed bag of these faculties. You may, for example, be exceptionally good at matching colours but tone deaf. Doctors need to be more than two standard deviations above the mean for intellectual discrimination of goodness of fit.

We recognise people and places and the bark of our own dog, but we cannot teach these skills to anyone else by description alone. We assemble ingredients and cook dishes which, by general agreement, taste delicious. Mozart assembled notes in a way which is pleasing to those who like music and have been brought up in the Western tradition. The music can be recognised as his even if it has not been heard

David Mendel (1984): *Proper Doctoring*

"When we come to understand all disease processes, the pattern recognition approach will die out..." (Mendel 1984)

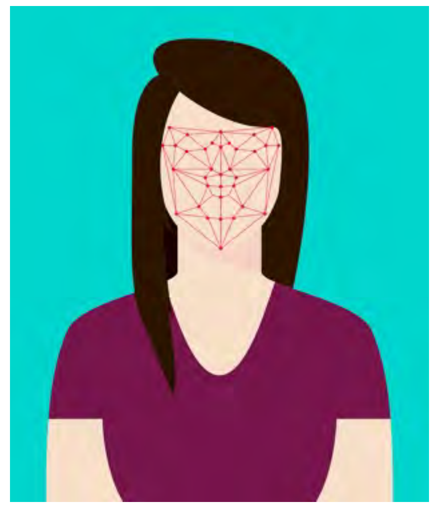


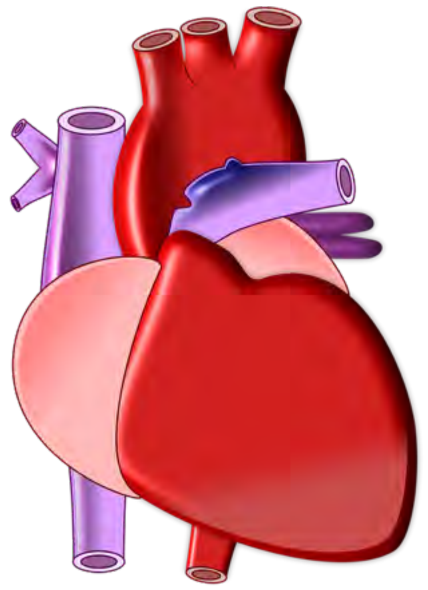
100%
Pattern
recognition

100%
Physiological
approach

“AI will not replace physicians. But physicians who use AI will replace those who don’t” (Dai and Singh 2023)



100% Pattern recognition (Human)

Physiological approach (Human + rule-based AI)

+

Pattern recognition (AI)

# Male or Female?

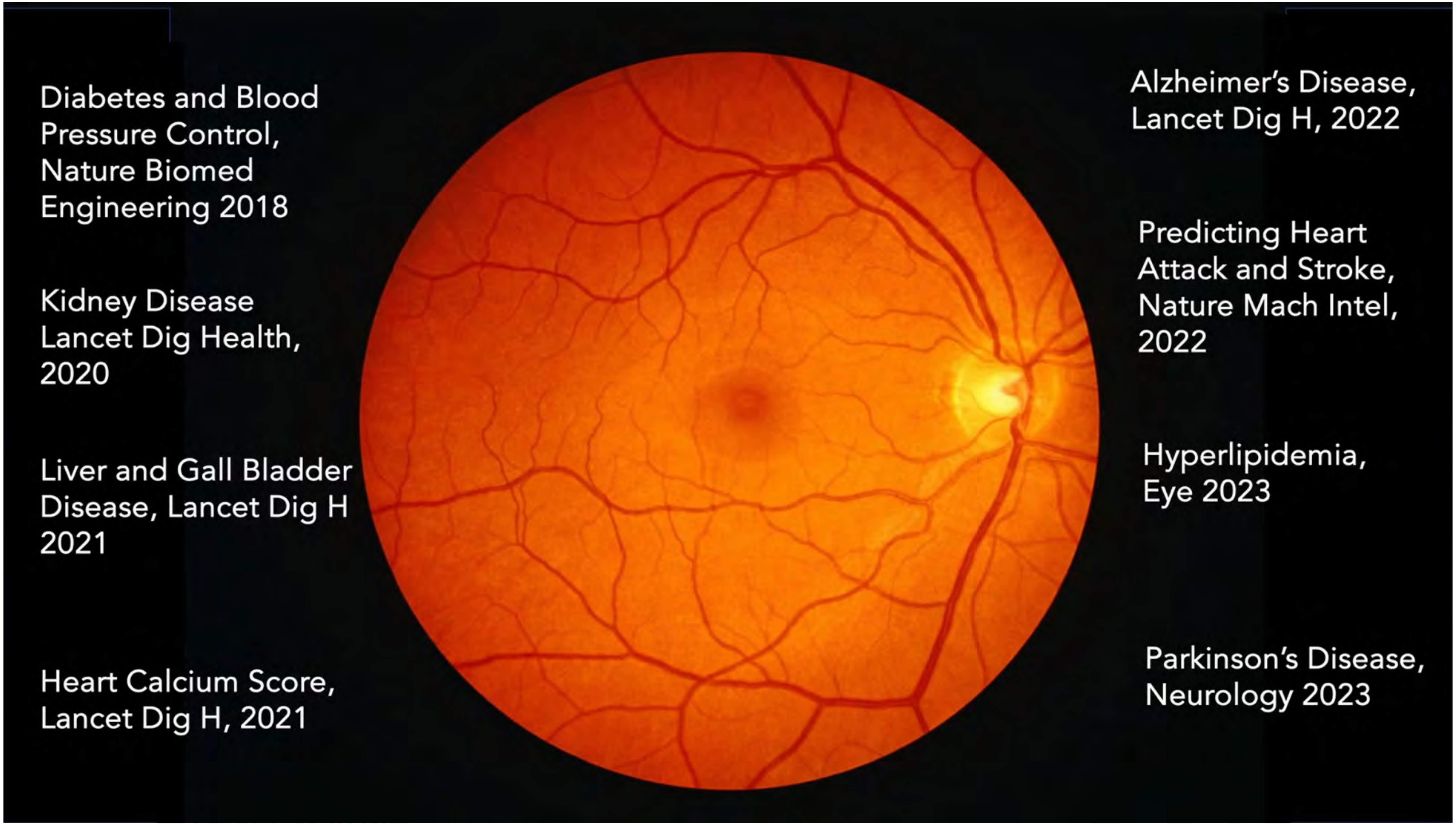
**Predicting sex from retinal fundus photographs using automated deep learning**

Edward Korot, Nikolas Pontikos, Xiaoxuan Liu, Siegfried K. Wagner, Livia Faes, Josef Huemer, Konstantinos Balaskas, Alastair K. Denniston, Anthony Khawaja & Pearse A. Keane

*Scientific Reports* **11**, Article number: 10286 (2021) | Cite this article

88.8% accuracy using a simple AI model trained with an automated machine learning tool

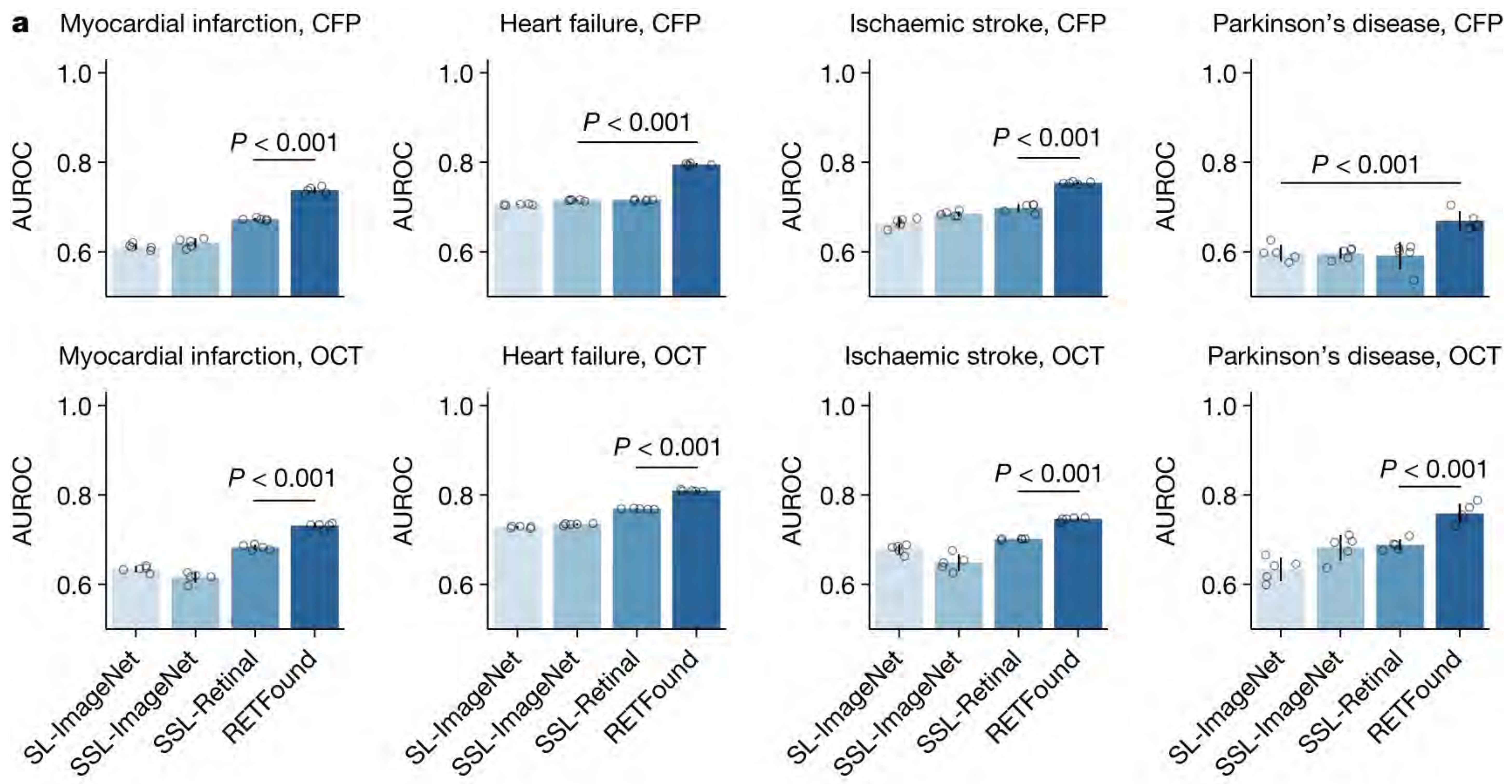
# And There's More...




Eric Topol (2023): "Can AI catch what doctors miss?"

nature Vol 622 | 5 October 2023

# A foundation model for generalizable disease detection from retinal images



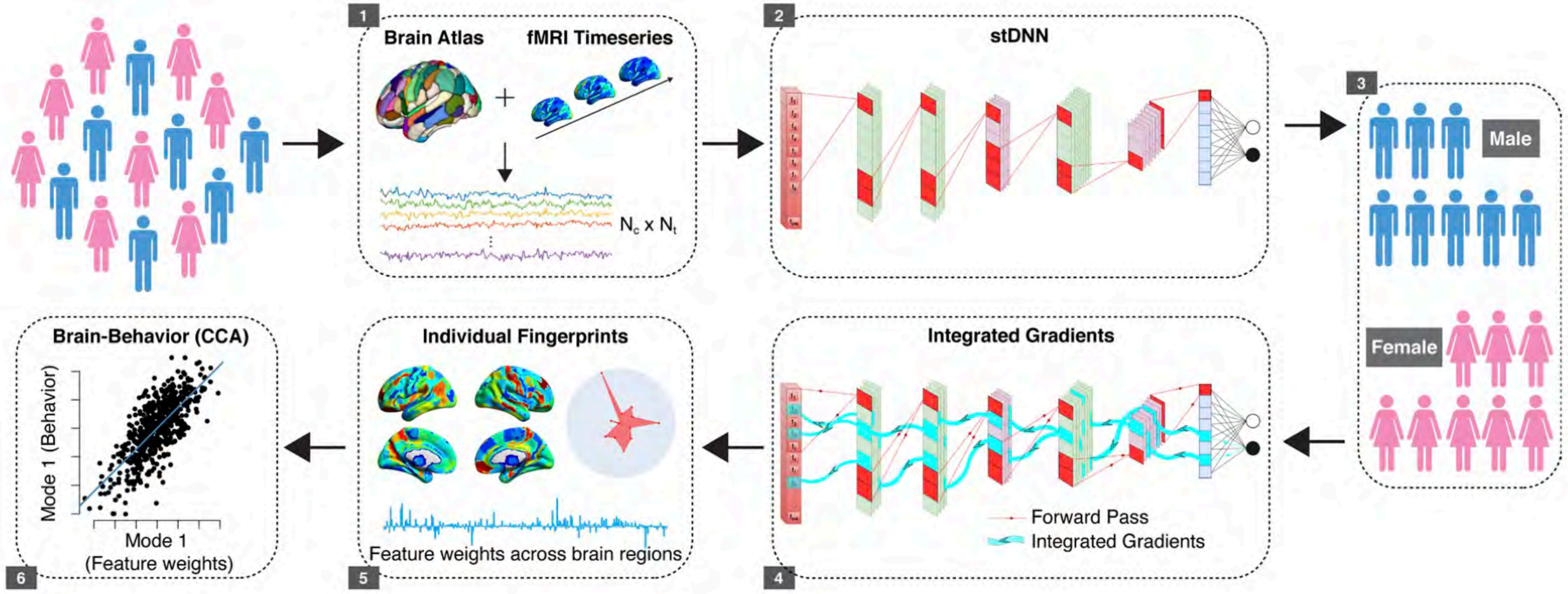
PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

# Deep learning models reveal replicable, generalizable, and behaviorally relevant sex differences in human functional brain organization

Srikanth Ryali[a,1], Yuan Zhang[a,1], Carlo de los Angeles[a], Kaustubh Supekar[a,b,c], and Vinod Menon[a,b,c,d,2]

Edited by Ruben C. Gur, University of Pennsylvania, Philadelphia, PA; received June 23, 2023; accepted December 21, 2023 by Editorial Board Member Terrence J. Sejnowski

**February 20, 2024** | 121 (9) e2310012121 | https://doi.org/10.1073/pnas.2310012121




Data Source: Human Connectome Project and independent datasets (*N*=1,500)

Technique: Spatiotemporal deep neural network (stDNN)

>90% in distinguishing male/female brains

# AI can increase clinical productivity; we have real-world evidence for this.

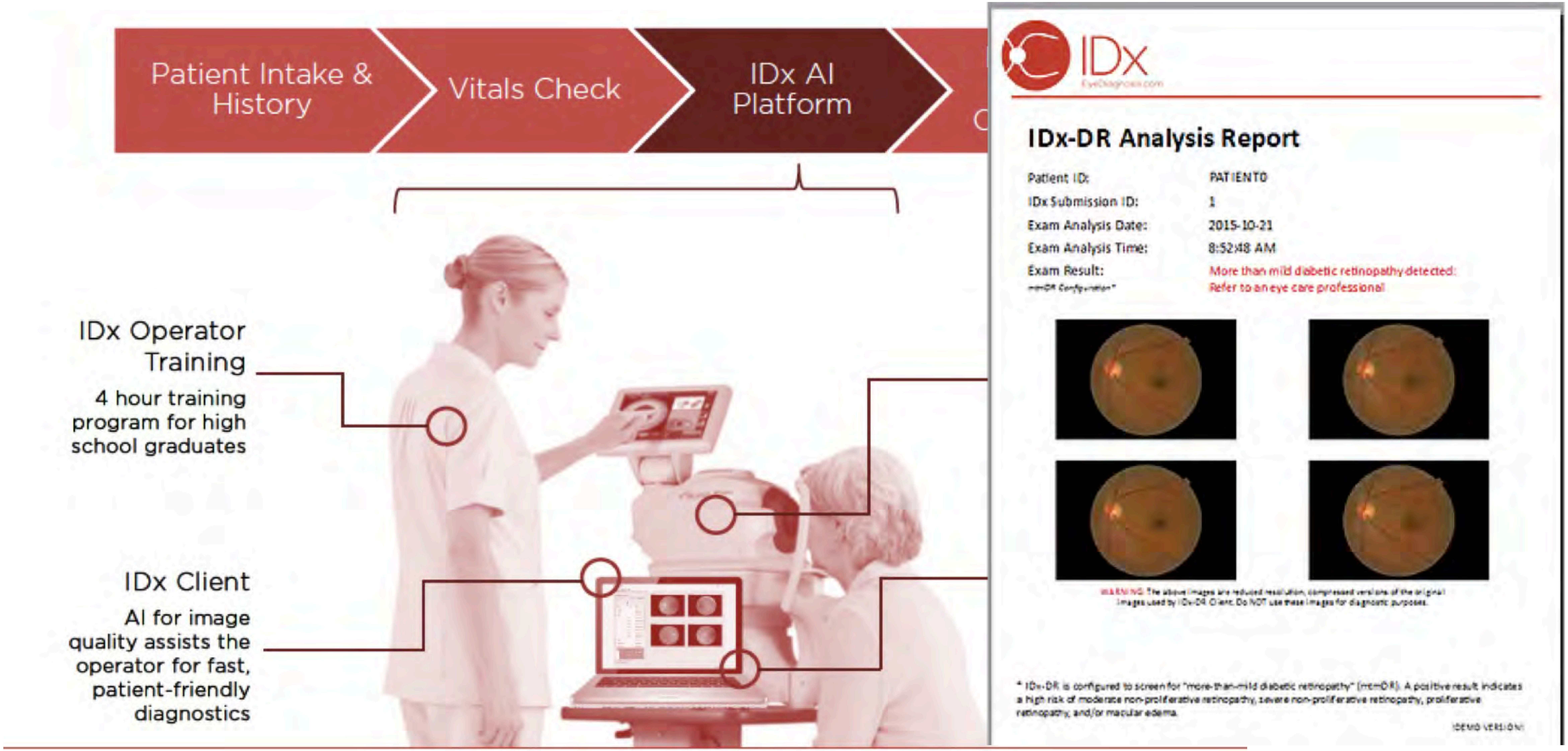
Abramoff et al. (2023). "Autonomous artificial intelligence increases real-world specialist clinic productivity in a cluster-randomized trial." *npj Digital Medicine*: https://www.nature.com/articles/s41746-023-00931-7

**Michael D Abramoff**
@MichaelAbramoff

A first of its kind randomized controlled trial, with @OrbisIntl ('the flying eye hospital'), @JohnsHopkins @TinglongDai, and Bangladeshi clinicians, showed in a real-world, low income setting, autonomous AI (@AItheRightWay ) massively increased clinician productivity, against a background of historic declines in outpatient productivity.

**Michael Abràmoff, M.D., Ph.D., Fellow of IEEE, Founder and CEO, Digital Diagnostics Inc.**

# LuminetiscCore for Diabetic Retinopathy Screening



# B-PRODUCTIVE Cluster-Randomized Trial

**Real-World Evidence of Autonomous AI Improving Clinical Productivity**

- B-PRODUCTIVE (Bangladesh-PRODUCTIVity in Eyecare) study was a pre-registered, prospective, **double masked**, cluster-randomized controlled trial

- Setting: Deep Eye Care Foundation, a not-for-profit hospital in Rangpur, Bangladesh

- Study period: March–July 2022

- Control group (N=499): Patients start with visiting the AI system for screening and are *always* sent to a specialist directly regardless of the screening results

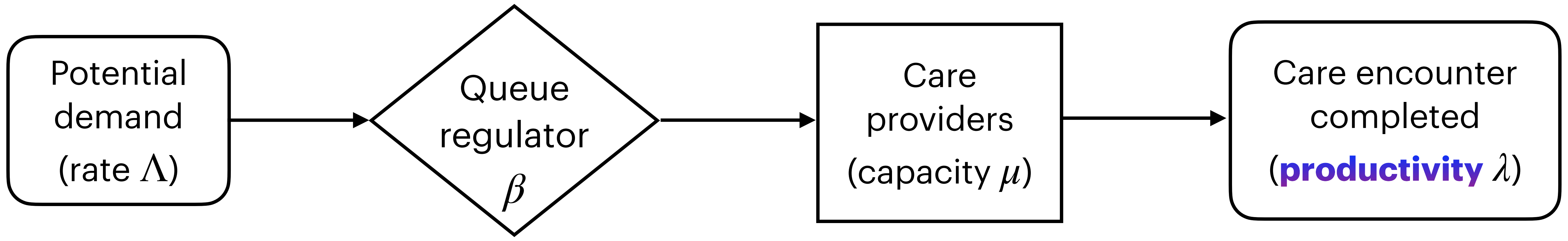
- Intervention group (N = 494): Patients start with visiting the AI system for screening and are referred to a specialist only if the results are positive

# Why Bangladesh?

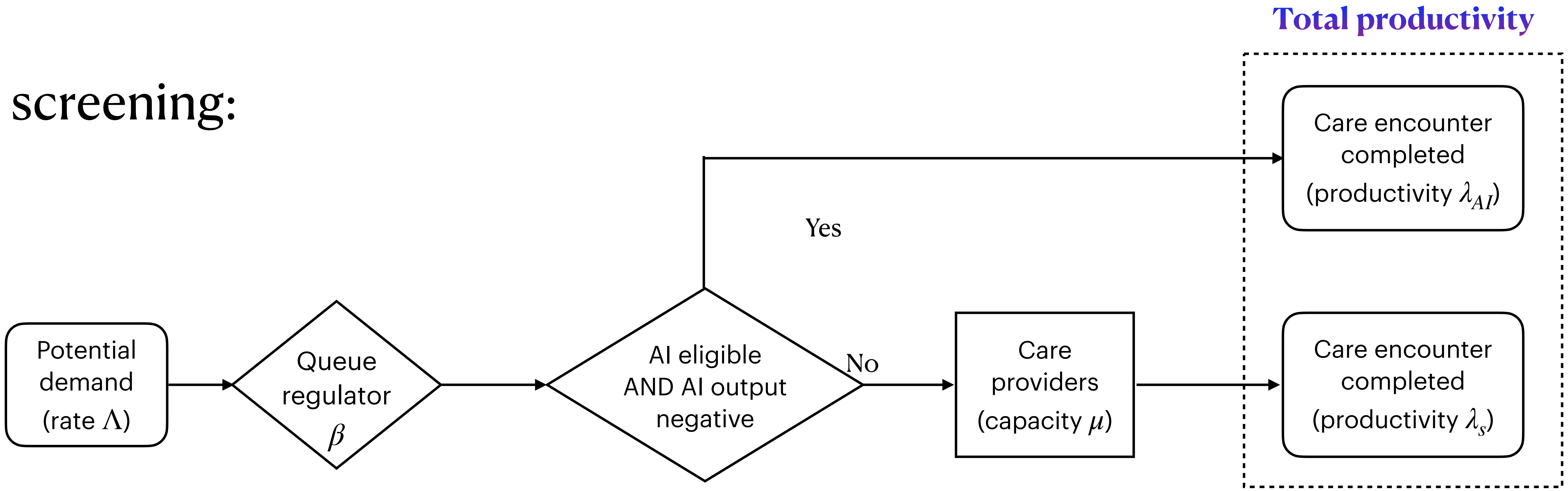
- To measure clinical productivity, a large potential demand and patients arriving without appointments are ideal
- This is commonly found in low- and middle-income countries (LMICs)
- In high-income countries, most patient visits are pre-scheduled and cannot be changed dynamically
- DECF was chosen due to its very large potential demand to avoid recruitment bias

# Healthcare Productivity Model Based on Rational Queueing Theory

Without AI screening:

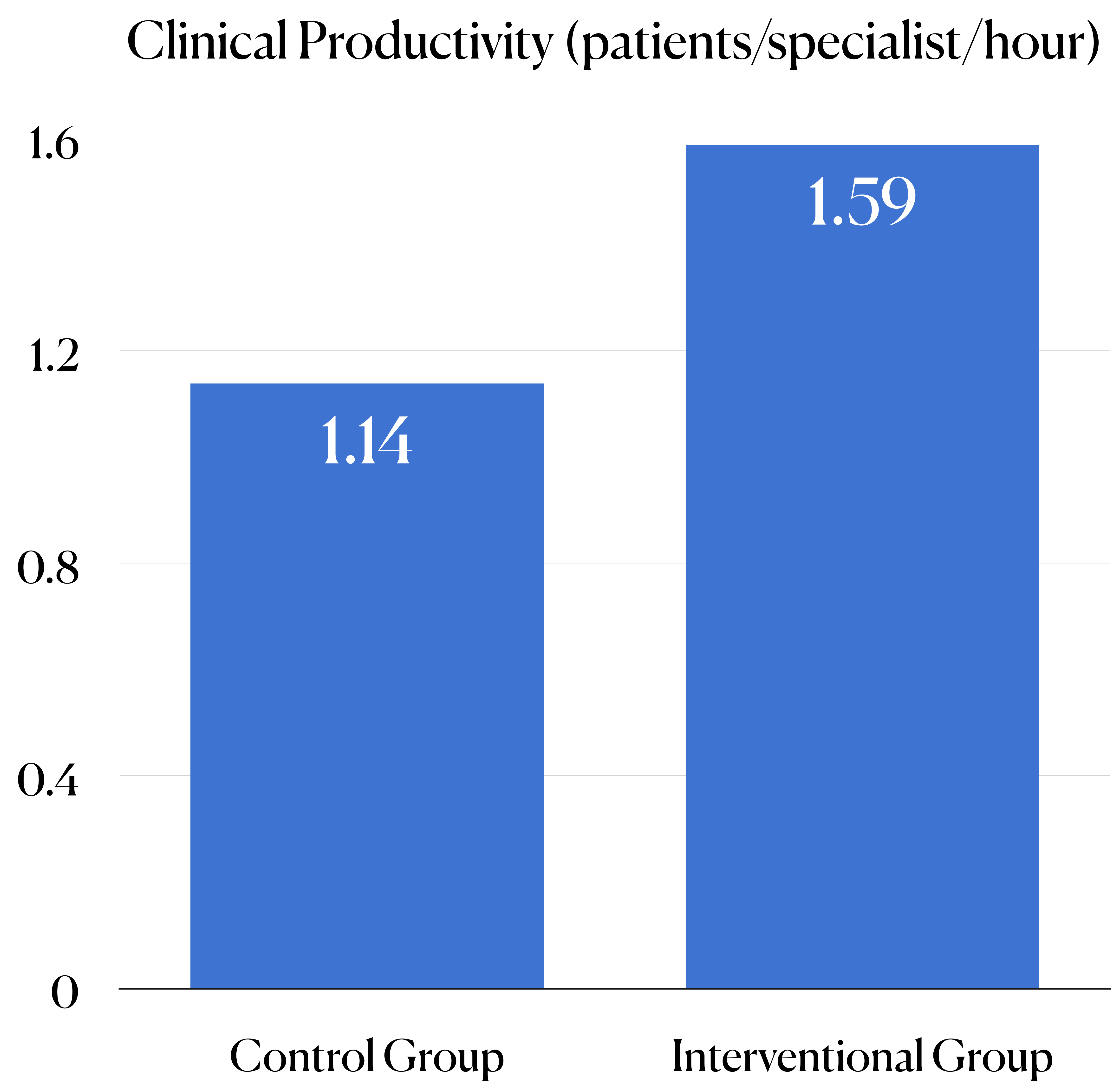



With AI screening:



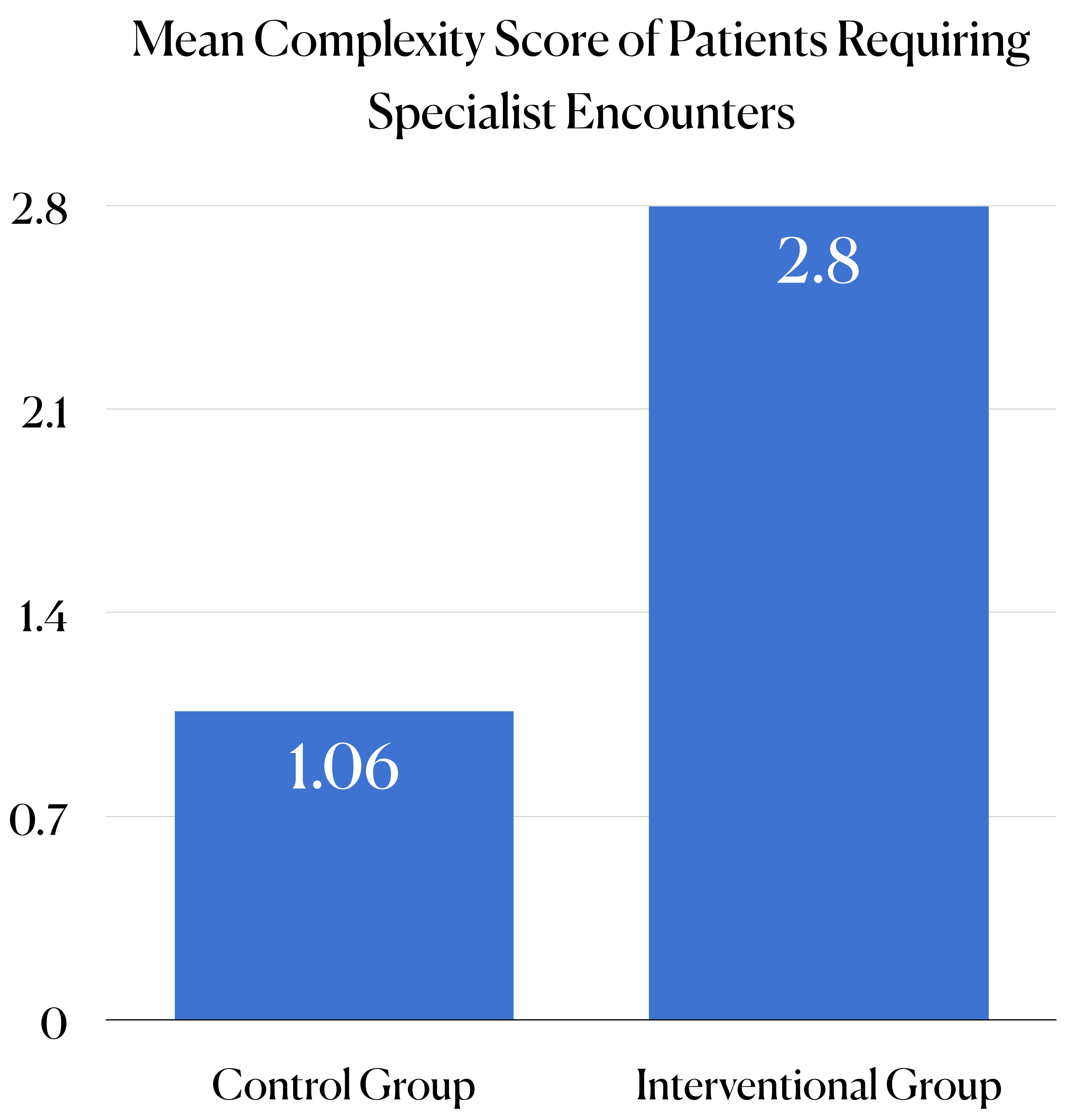
# Productivity Outcomes by Study Group

| | Control group mean (95% CI) | Intervention group mean (95% CI) |
|---|---|---|
| Completed care encounters among clinic patients with diabetes | | |
| Care encounter involved specialist | 920 | 858 |
| Care encounter completed by AI-only | 0 | 331 |
| Total | 920 | 1189 |
| Total number of specialist hours in clinic | 819 | 763 |
| Clinic productivity (95% CI) for diabetes patients: number of completed care encounters per hour per specialist | $\lambda_{d,c}$=1.14 (1.02, 1.25) | $\lambda_{d,AI}$=1.59 (1.3, 1.80) |
| Clinic productivity (95% CI) for all patients number of completed care encounters per hour per specialist | $\lambda_{c}$=3.36 (3.08, 3.63) | $\lambda_{AI}$=4.05 (3.66, 4.43) |
| Specialist productivity adjusted for patient complexity for diabetes patients | $\lambda_{ca,d,c}$=1.19 | $\lambda_{caAI}$=3.15 |


Clinical Productivity (patients/specialist/hour)
1.6
1.2
0.8
0.4
0
1.14
1.59
Control Group
Interventional Group

# 40%

**Increase in clinical productivity when using AI-augmented healthcare**


Mean Complexity Score of Patients Requiring Specialist Encounters
2.8
2.1
1.4
0.7
0
1.06
2.8
Control Group
Interventional Group

# 265%

**Increase in complexity-adjusted specialist productivity when using AI-augmented healthcare**

# For Medical AI To Scale, We Need To Insist on Asking for Real-World Evidence for Using AI To Improve

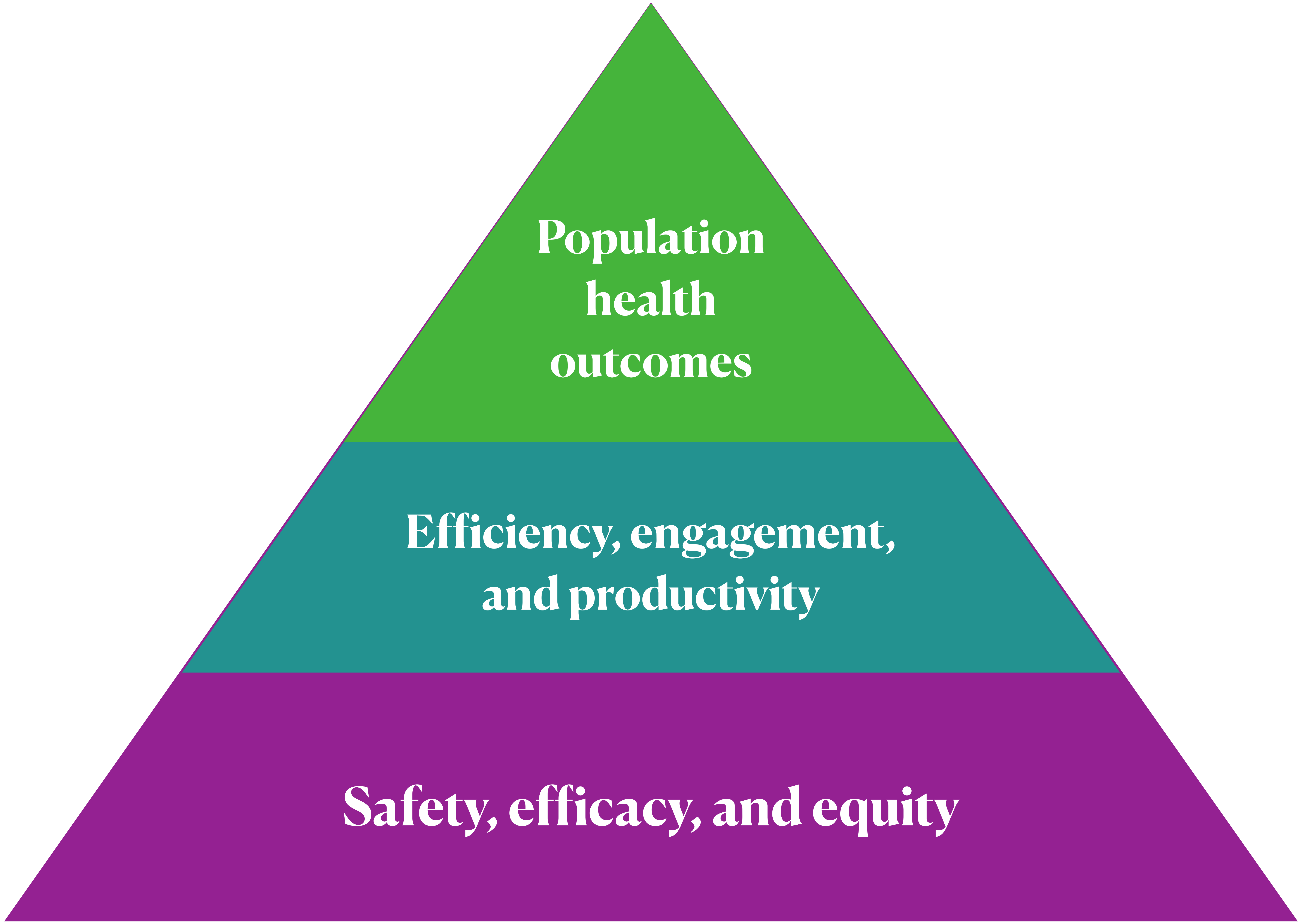


**Table 1. Summary of AI CPT Codes.***

| Total Claims | Condition or Medical AI Procedure | CPT Code(s) | Example Product Name | Effective Date |
|---|---|---|---|---|
| 67,306 | Coronary artery disease | 0501T–0504T | HeartFlow Analysis[48] | June 1, 2018 |
| 15,097 | Diabetic retinopathy | 92229 | LumineticsCore[49] | January 1, 2021 |
| 4,459 | Coronary atherosclerosis | 0623T–0626T | Cleerly[50] | January 1, 2021 |
| 2,428 | Liver MR | 0648T–0649T | Perspectum LiverMultiScan[51] | January 1, 2021 |
| 591 | Multiorgan MRI | 0697T–0698T | Perspectum CoverScan[52] | January 1, 2022 |
| 552 | Breast ultrasound | 0689T–0690T | Koios DS[53] | January 1, 2022 |
| 435 | ECG cardiac dysfunction | 0764T–0765T | Anumana[50] | January 1, 2023 |
| 331 | Cardiac acoustic waveform recording | 0716T | CADScor[50] | July 1, 2022 |
| 237 | Quantitative MR cholangiopancreatography | 0723T–0724T | Perspectum MRCP+[54] | July 1, 2022 |
| 67 | Epidural infusion | 0777T | CompuFlo[55] | January 1, 2023 |
| 4 | Quantitative CT tissue characterization | 0721T–0722T | Optellum Virtual Nodule Clinic[56] | July 1, 2022 |
| 1 | Autonomous insulin dosage | 0740T–0741T | d-Nav[57] | January 1, 2023 |
| 1 | CT vertebral fracture assessment | 0691T | HealthVCF[50] | January 1, 2022 |
| 1 | Noninvasive arterial plaque analysis | 0710T–0713T | ElucidVivo[50] | January 1, 2022 |
| 0 | Facial phenotype analysis | 0731T | Face2Gene[50] | July 1, 2022 |
| 0 | X-ray bone density | 0749T | OsteoApp[50] | January 1, 2023 |

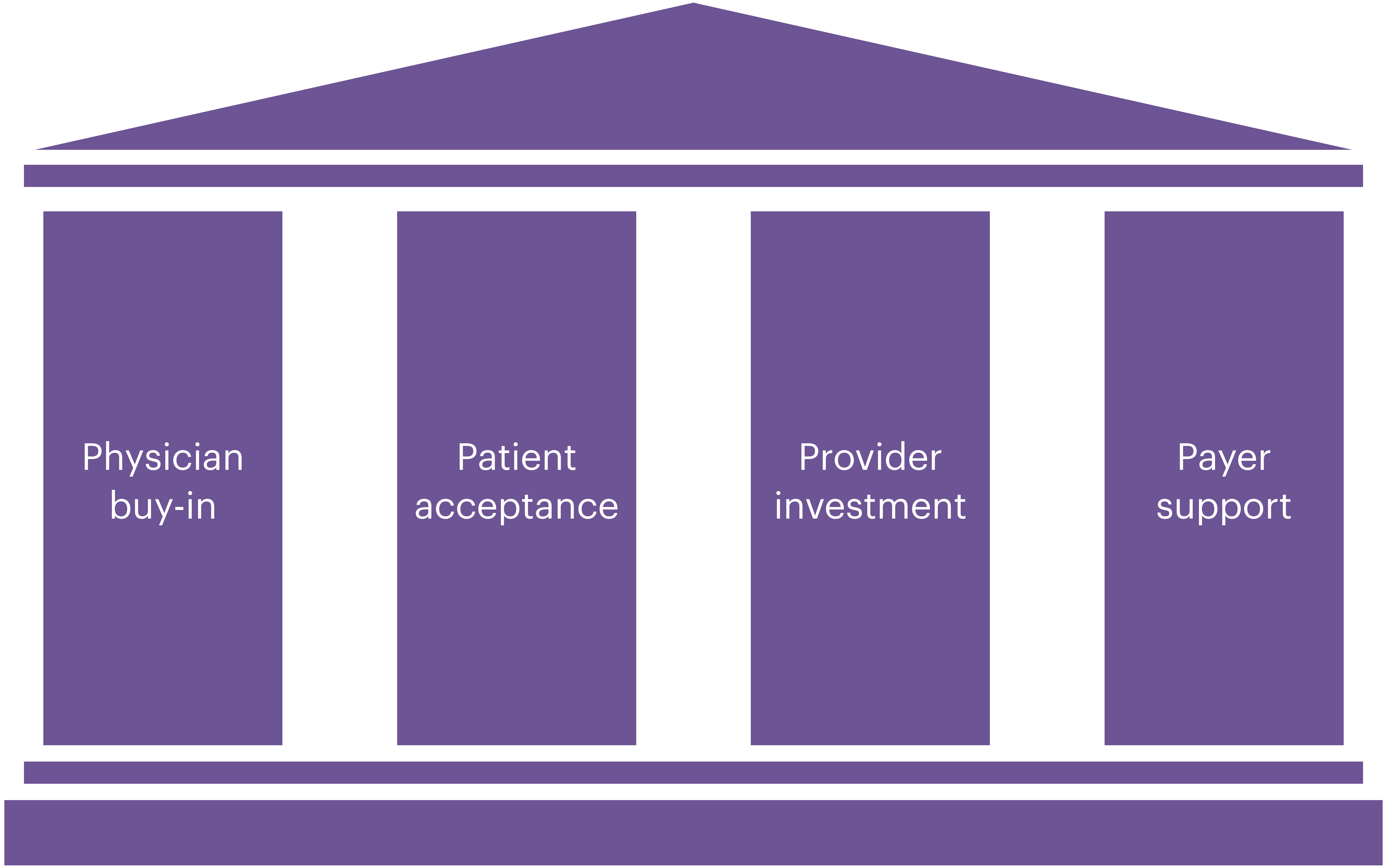
* A total of 16 medical AI procedures are presented alongside their corresponding CPT codes. Each procedure is associated with an example commercial product that may be reimbursed through the codes. The effective date is the date on which the code was officially recognized by the American Medical Association and can be used for billing and reimbursement purposes. The total claims listed are recent as of June 1, 2023. AI denotes artificial intelligence; CPT, Current Procedural Terminology; CT, computed tomography; ECG, electrocardiogram; MR, magnetic resonance; MRCP, magnetic resonance cholangiopancreatography; and MRI, magnetic resonance imaging.

Wu et al., 2023, "Characterizing the Clinical Adoption of Medical AI Devices through U.S. Insurance Claims," *NEJM AI*

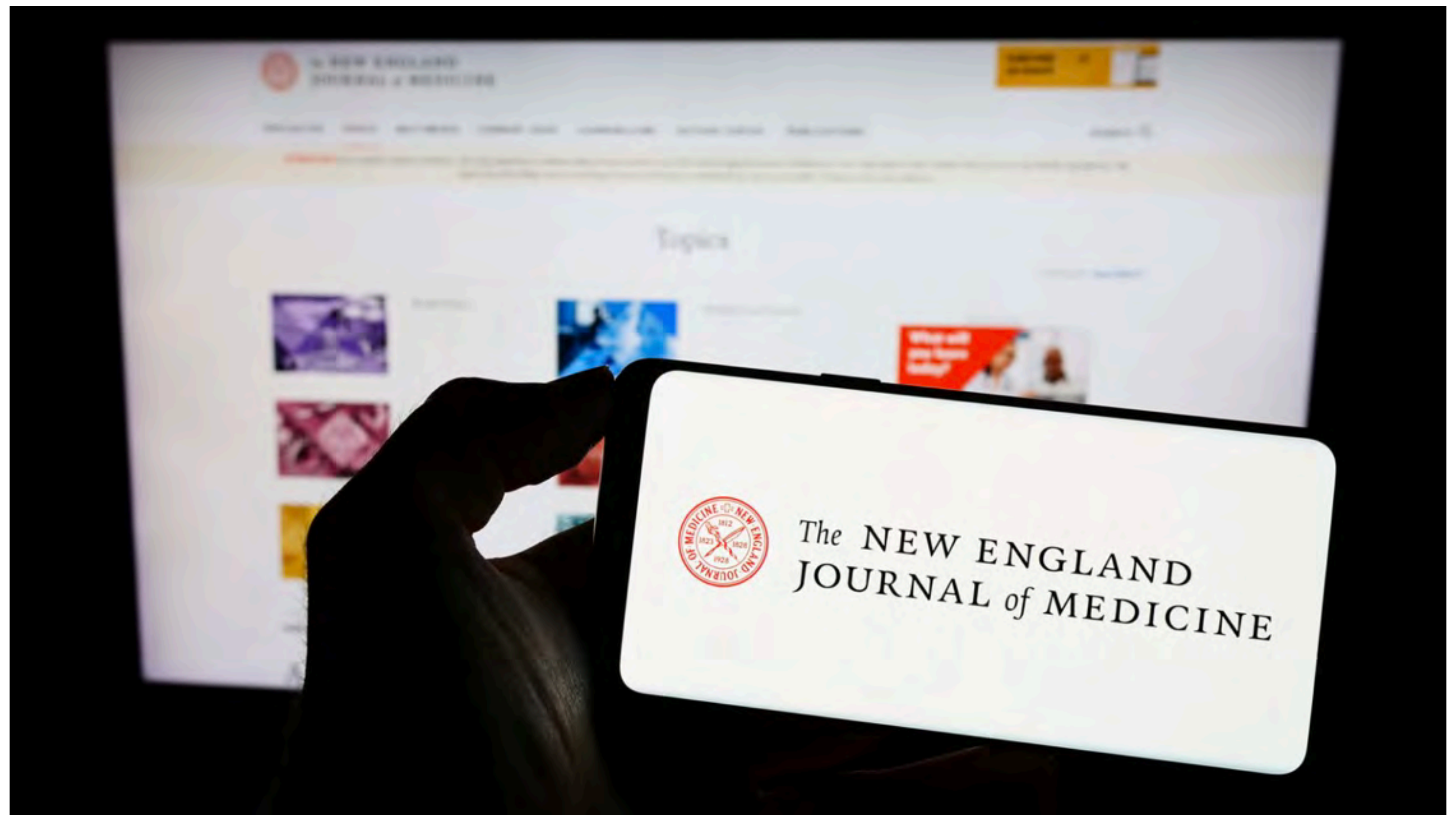
# Small-Group Discussions: Scaling Medical AI

- Instructions:
  - Group Setup: Form groups of 5–6 students
  - Discuss for 10 Minutes: Reflect on the questions provided
  - Role Assignment: One note-taker, one presenter
  - Report Findings: Each group will present a 1-minute summary at the end
- Discussion Questions:
  - Why the uptake of medical AI is so low?
  - How to accelerate  the use of safe and effective medical AI?

# Four Pillars of Incorporating AI Into Healthcare Workflow




Source: Dai and Tayur (2022)"Designing AI-augmented Healthcare Delivery Systems for Physician Buy-in and Patient Acceptance." *Production and Operations Management*, 31 (12): 4443–4451.




NEJM AI

# Scaling Adoption of Medical AI — Reimbursement from Value-Based Care and Fee-for-Service Perspectives

Michael D. Abramoff, M.D., Ph.D.,[1,2,3] Tinglong Dai, Ph.D.,[4,5,6] and James Zou, Ph.D.[7,8,9]

Received: January 22, 2024; Revised: February 14, 2024; Accepted: February 23, 2024; Published: April 12, 2024

# Reimbursement is Key to Achieving Hundredfold Growth

- Fee-for-Service (FFS) requires demonstration of efficacy, safety, and cost-effectiveness:
  - Risk factors: Time and resources for securing a CPT code

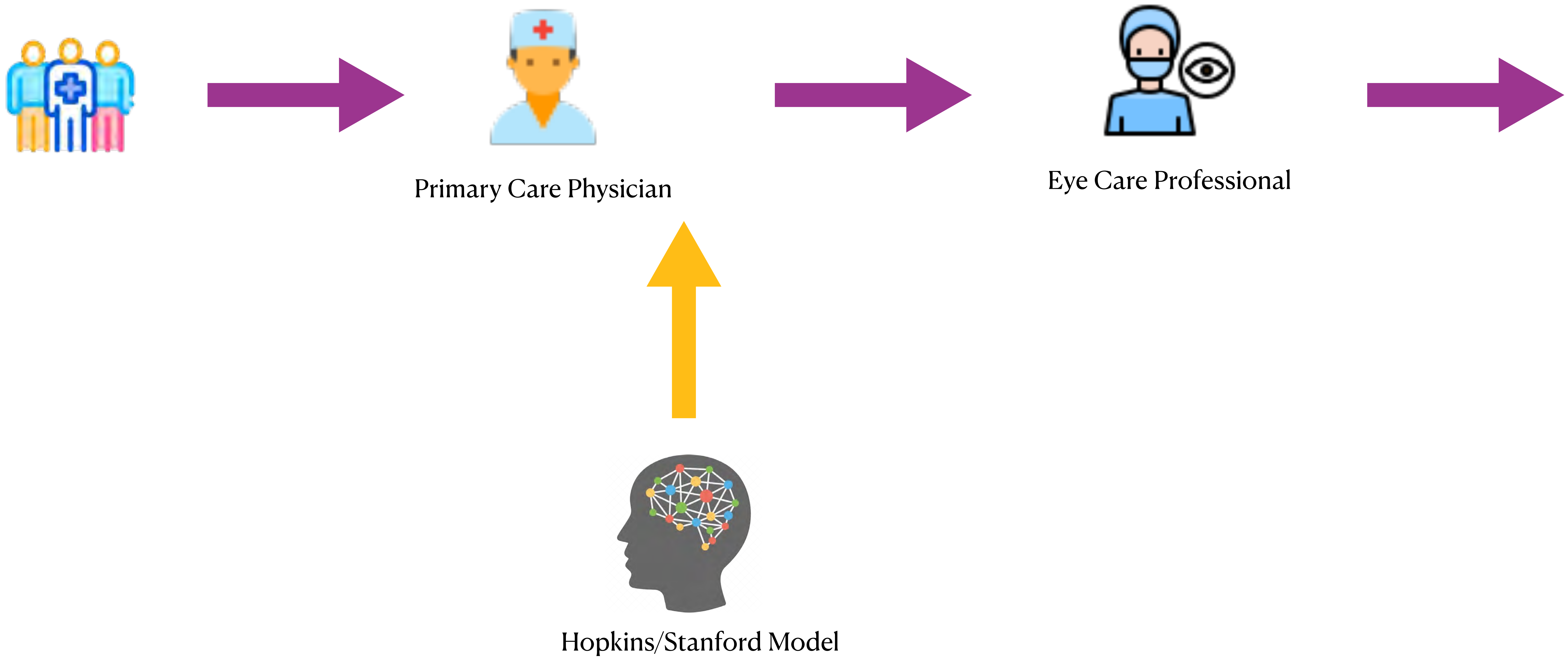
- Value-Based Care (VBC) reimburses based on patient outcomes and care metrics
  - Authorization for AI tools to close care gaps is often easier than securing a CPT code
  - No financial gain if metrics (e.g., 80% annual diabetic eye exams) are not fully met

- Proposed Framework: Medicare Part B Inspired Model
  - Providers and AI developers share revenue, easing tensions in FFS and VBC models
  - Contingent upon CMS coverage and specific CPT codes

Figuring out where to place AI in the delivery loop is just as important

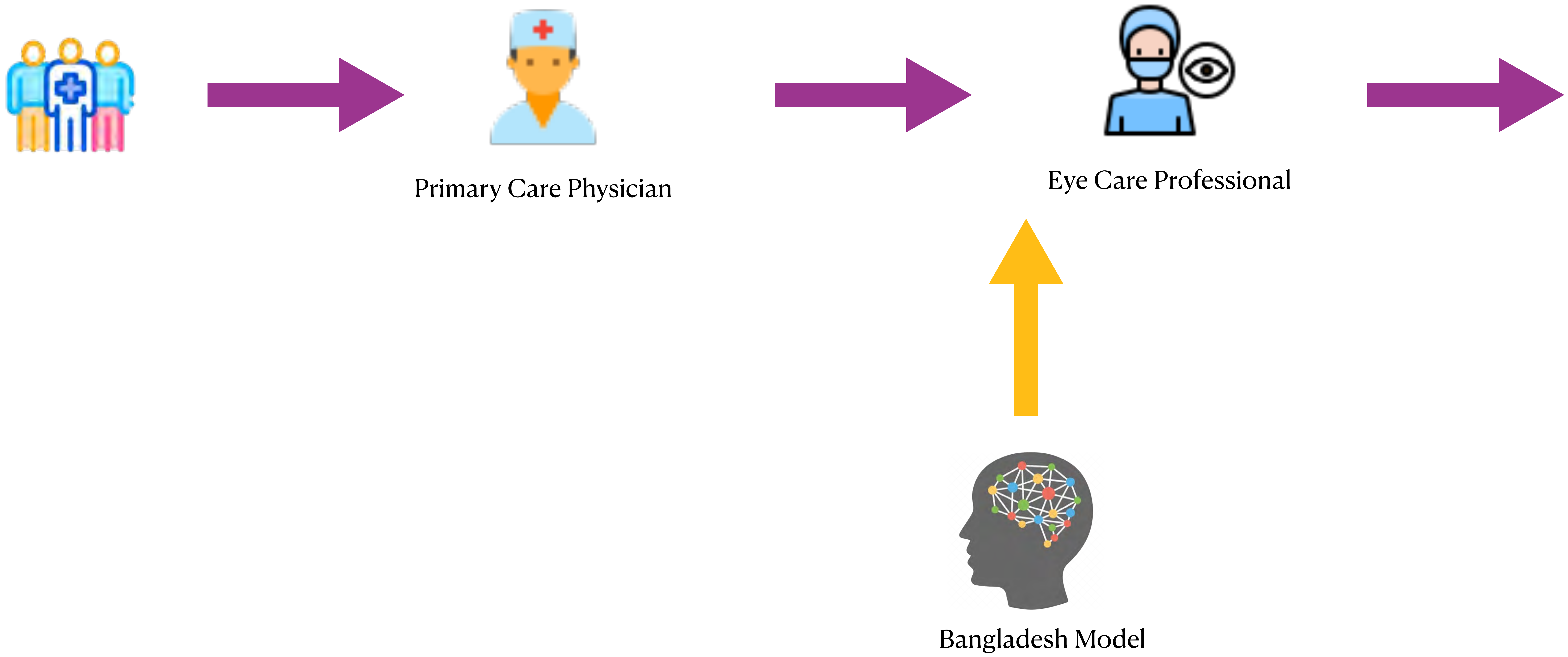
# Diabetic Retinopathy Care

## Using AI as Gatekeeper



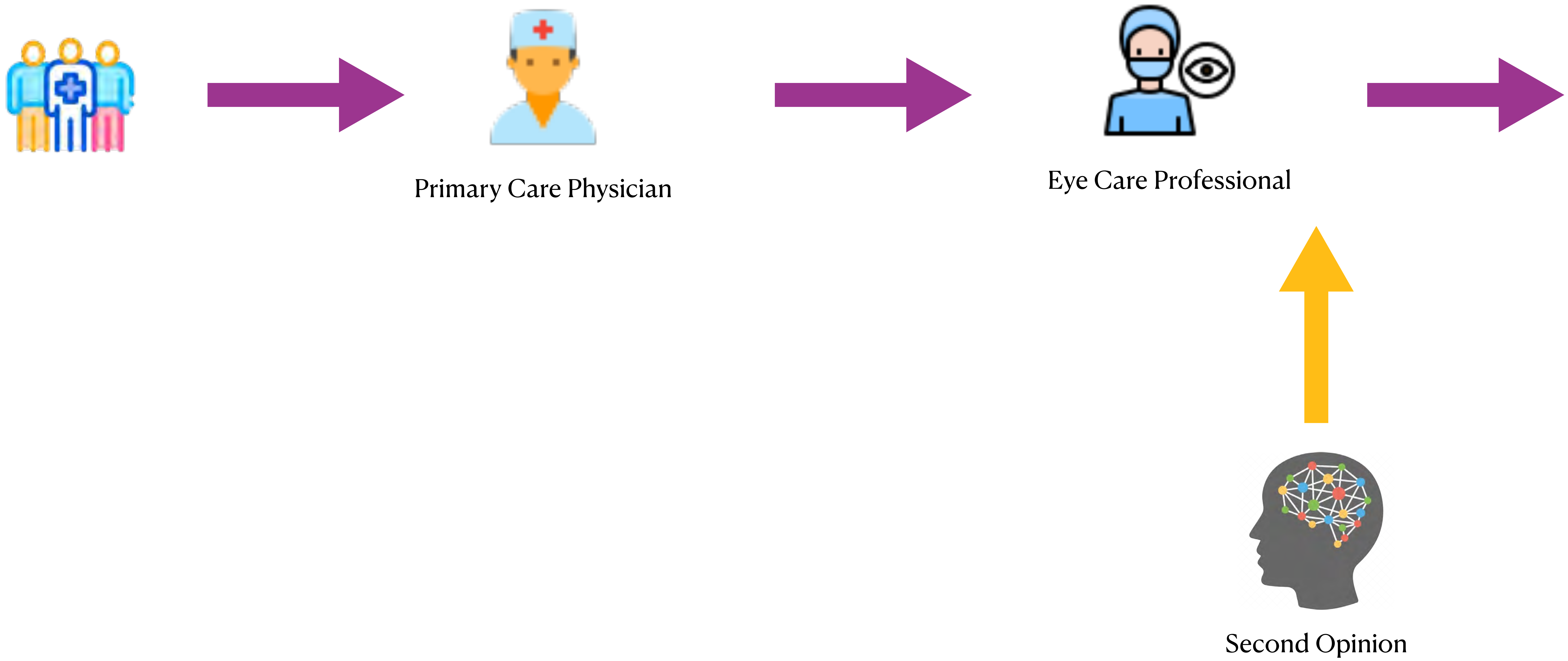
# Diabetic Retinopathy Care

## Using AI as Gatekeeper



# Diabetic Retinopathy Care

## Using AI as Second Opinion



# Patient Acceptance of AI Solutions

- **Patients' attitudes toward AI**

**Resistance to Medical Artificial Intelligence**

Chiara Longoni, Andrea Bonezzi, Carey K Morewedge

**Published:** 03 May 2019

Journal of Consumer Research

**"Uniqueness bias"**—patients might be reluctant to accept medical AI because they may overrate the uniqueness of their situations

- **AI and trust in patient-physician interactions**

July 15, 2019

**Promoting Trust Between Patients and Physicians in the Era of Artificial Intelligence**

Shantanu Nundy, MD, MBA[1,2]; Tara Montgomery, BA[3]; Robert M. Wachter, MD[4]

JAMA

# Key Drivers of Physician Buy-in of AI Solutions

## Reputation Concerns

- Peers may perceive AI users as having **low or redundant skills** (Dai & Singh 2020).
- Clinicians using computer-based diagnostic tools are often regarded as **inferior diagnosticians** by peers and patients (Arkes et al. 2007).
- This perception persists even among **patients with IT backgrounds** (Wolf 2014).

## Legal-Liability Implications

- The use of AI could **increase or lessen physician liability** in the event of malpractice lawsuits, creating uncertainty (Dai and Singh 2025).
- Relying on **biased algorithms** to make wrong medical decisions can increase physician liability (Luan et al. 2025)

npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-01901-x

# Peer perceptions of clinicians using generative AI in medical decision-making

Check for updates

**Haiyang Yang[1,2], Tinglong Dai[1,2,3,4], Nestoras Mathioudakis[5], Amy M. Knight[5], Yuna Nakayasu[1,6] & Risa M. Wolf[1,2,5]**

This study investigates how a physician's use of generative AI (GenAI) in medical decision-making is perceived by peer clinicians. In a randomized experiment, 276 practicing clinicians evaluated one of three vignettes depicting a physician: (1) using no GenAI (Control), (2) using GenAI as a primary decision-making tool (GenAI-primary), and (3) using GenAI as a verification tool (GenAI-verify). Participants rated the physician depicted in the GenAI-primary condition significantly lower in clinical skill (on a 1–7 scale; mean = 3.79) than in the Control condition (5.93, $p < 0.001$). Framing GenAI use as verification partially mitigated this effect (4.99, $p < 0.001$). Similar patterns appeared for perceived overall healthcare experience and competence. Participants also acknowledged GenAI's value in improving accuracy (4.30, $p < 0.002$) and rated institutionally customized GenAI more favorably (4.96, $p < 0.001$). These findings suggest that while clinicians see GenAI as helpful, its use can negatively impact peer evaluations. These effects can be reduced, but not fully eliminated, by framing it as a verification aid.

# Experimental Design: Randomized Survey

### Participants

276 practicing clinicians from Johns Hopkins University health system

### Three Conditions

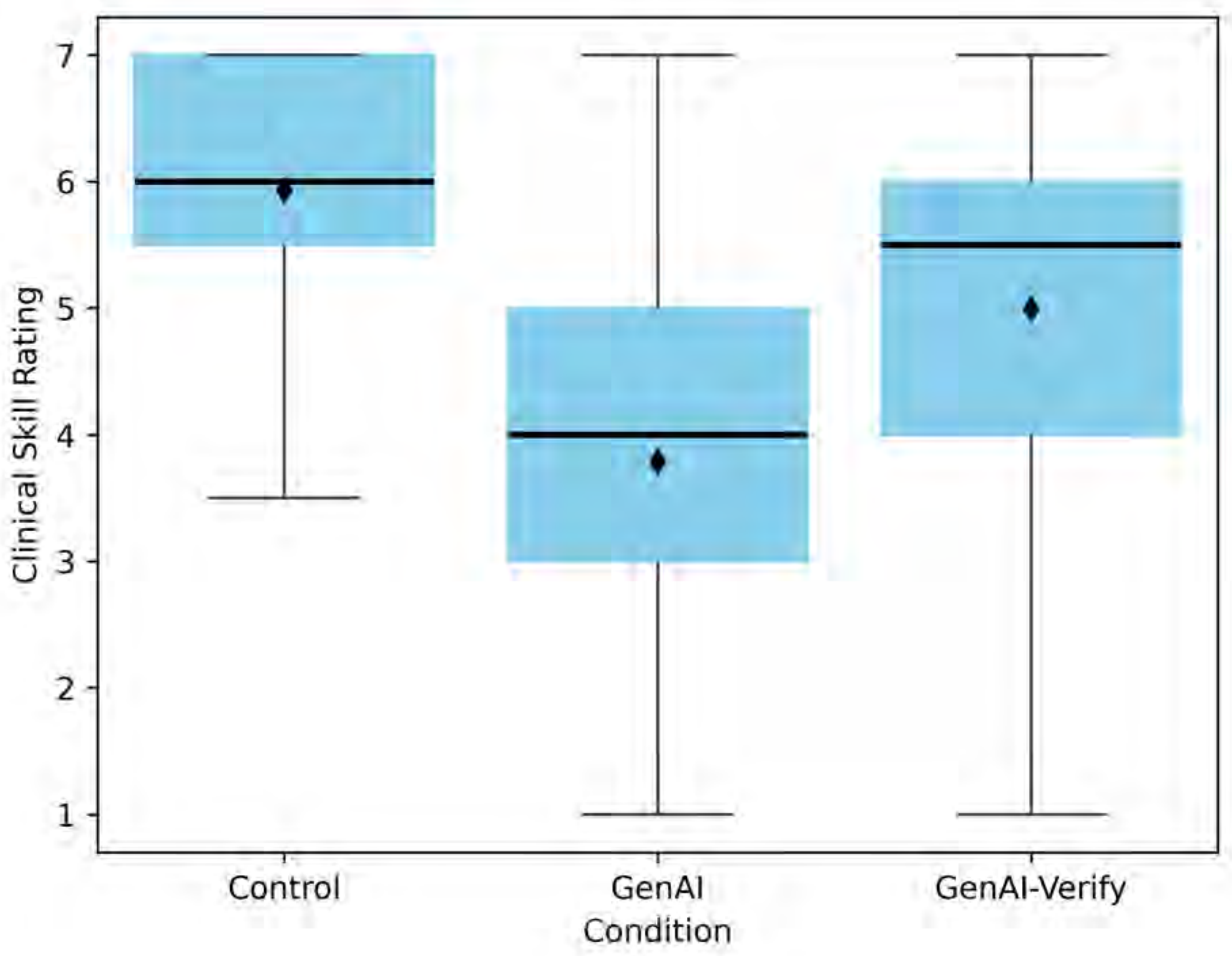
- **Control** — no GenAI mentioned
- **GenAI** — GenAI used to make decision
- **GenAI-verify** — physician decides, then verifies with GenAI

### Outcome Measures

- Clinical skill ratings
- Overall healthcare experience
- Overall competence

The experiment isolated the specific effect of AI usage and framing on peer judgments while controlling for all other variables.

# Key Results: Using AI Triggers a Peer Penalty

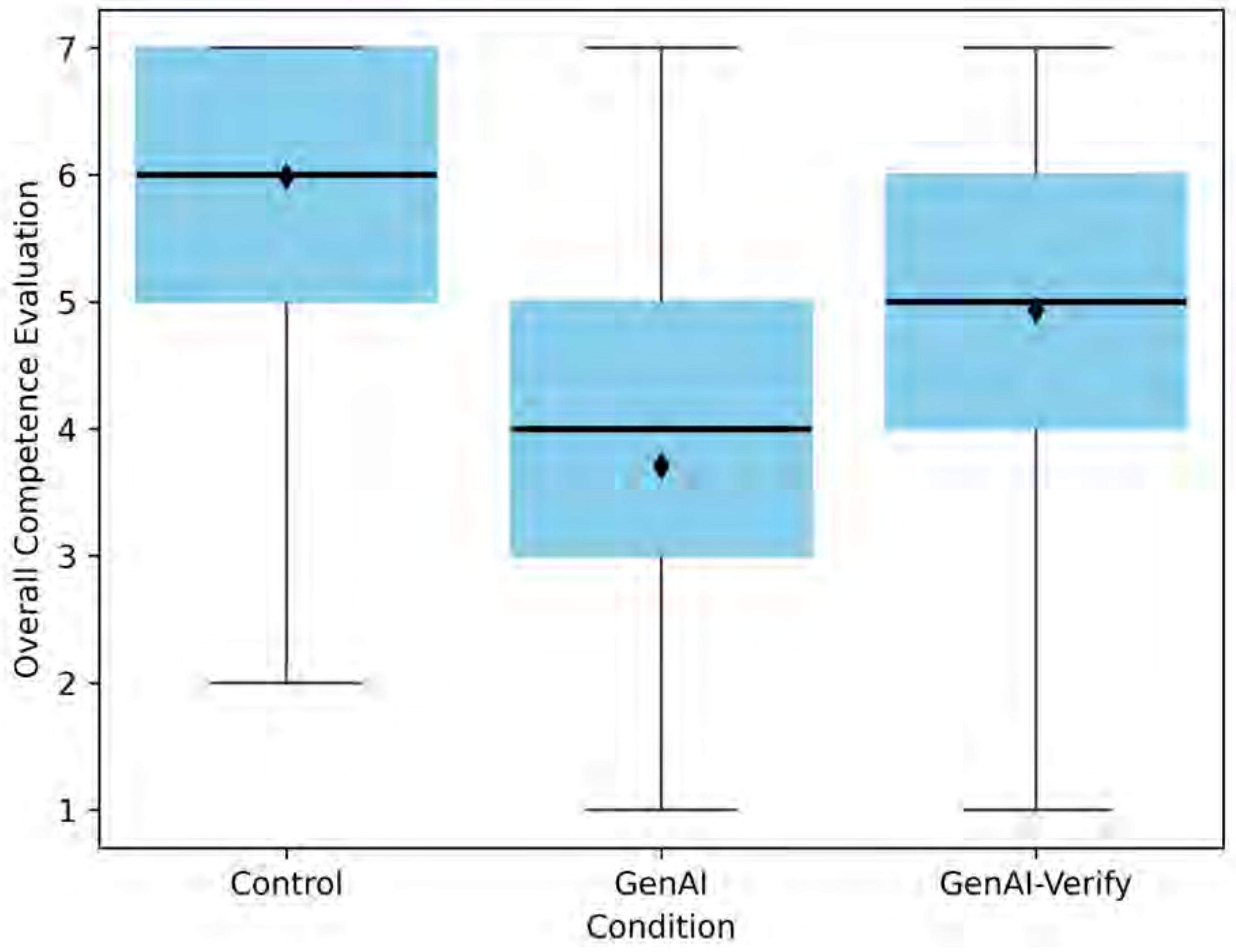



**Peer Ratings for**
**Clinical Skills**

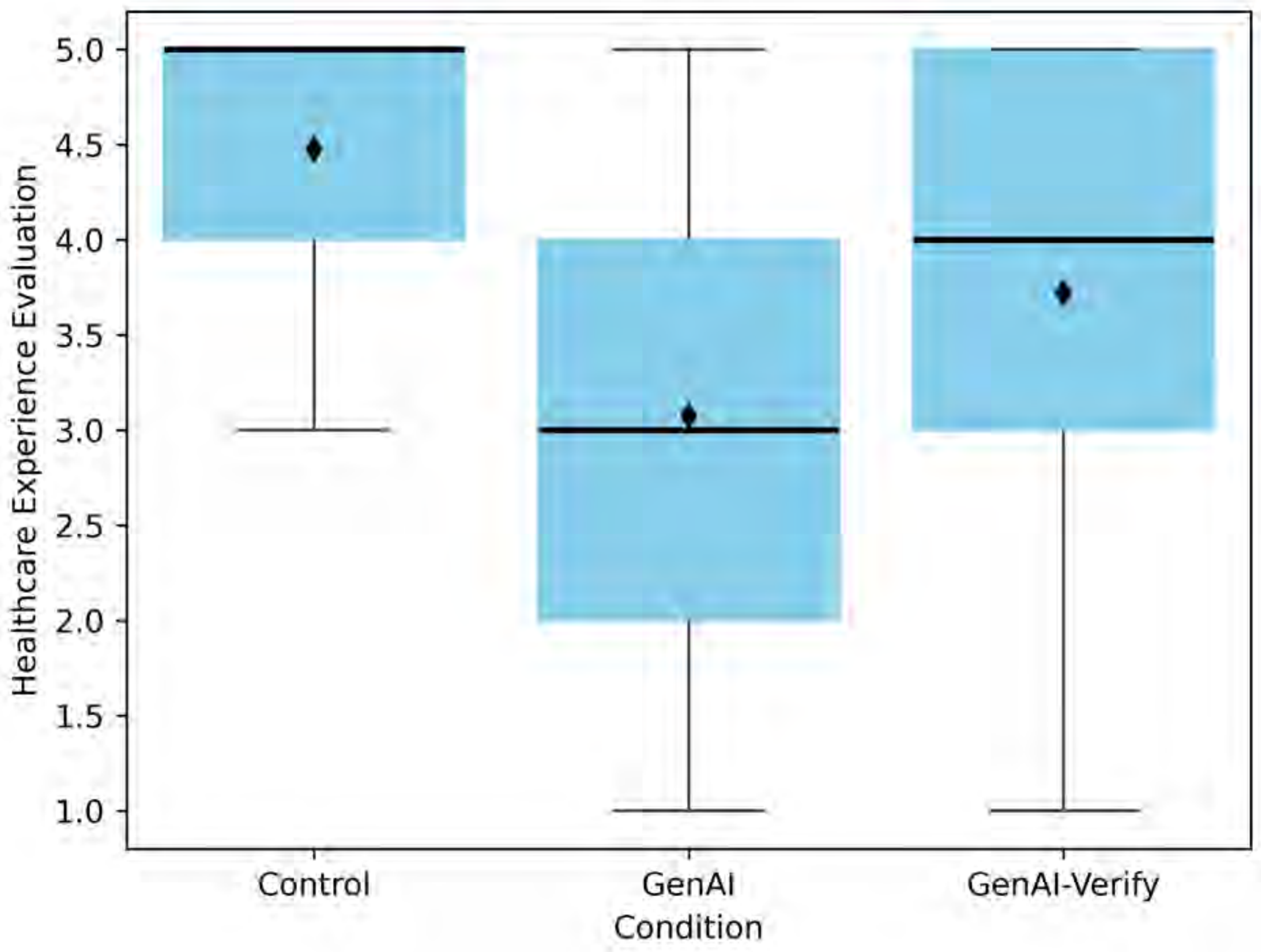
A 2+ point drop in perceived clinical skill (on a 1–7 scale) when peers see a doctor using GenAI as a decision-making tool ($p < 0.001$).

Framing AI as verification helps but doesn't eliminate the penalty.

# Similar patterns were observed for ratings of competence and overall experience ($p<0.001$)




**Peer Ratings for**
**Competence**




**Peer Ratings for**
**Healthcare Experience**

# Key Drivers of Physician Buy-in of AI Solutions

## Reputation Concerns

- Peers may perceive AI users as having **low or redundant skills** (Dai & Singh 2020).
- Clinicians using computer-based diagnostic tools are often regarded as **inferior diagnosticians** by peers and patients (Arkes et al. 2007).
- This perception persists even among **patients with IT backgrounds** (Wolf 2014).

## Legal-Liability Implications

- The use of AI could **increase or lessen physician liability** in the event of malpractice lawsuits, creating uncertainty (Dai and Singh 2025).
- Relying on **biased algorithms** to make wrong medical decisions can increase physician liability (Luan et al. 2025)

# Standard vs. Non-standard Treatment Plan

**Price, W. N., S. Gerke, & I. G. Cohen. (2019) Potential liability for physicians using artificial intelligence. *JAMA* 322(18):1765–1766.**

Consider a physician who prescribes a treatment plan for a patient. The physician can choose either a standard or non-standard treatment plan.

- A treatment plan is "standard" if it follows the standard of care
- In the case of ovarian cancer treatment: dosage of chemotherapeutic bevacizumab:

| Standard plan | Non-standard plan |
|---|---|
| 15 mg/kg every three weeks | 75 mg/kg every three weeks |

- Under the prevailing liability scheme, the physician is liable if (1) the physician deviates from the standard of care and (2) patient harm occurs
  ⇒ The physician is shielded from liability when following the standard of care

# Prevailing Liability Scheme

| Physician Decision | Patient Outcome | Is the Physician Liable? |
|---|---|---|
| Standard | 😀 | No |
| Standard | 😭 | No |
| Non-standard | 😀 | No |
| Non-standard | 😭 | Yes |

**In designing medical expert systems, the actions should be thought of not as directly affecting the patient but as influencing the physician's behavior. If expert systems become reliably more accurate than human diagnosticians, doctors might become legally liable if they don't use the recommendation of an expert system.**

**Russell and Norvig (2015, p.1051):** ***Artificial Intelligence: A Modern Approach***

# How AI Complicates the Liability Scheme (Russell and Norvig)

| AI Signal | Physician Decision | Patient Outcome | Is the Physician Liable? |
|---|---|---|---|
| Standard | Non-standard | 😭 | Yes |
| Non-standard | Standard | 😭 | Yes |
| Non-standard | Non-standard | 😭 | No |

# Another Liability Scheme (Price et al. 2019)

| AI Signal | Physician Decision | Patient Outcome | Is the Physician Liable? |
|---|---|---|---|
| Standard | Non-standard | 😭 | Yes (and more) |
| Non-standard | Standard | 😭 | No |
| Non-standard | Non-standard | 😭 | Yes (and less) |

# We are Doomed...


https://bit.ly/tokyo21-ai

# Second TA Tutorial: Keras & CNNs

**🗓️ Friday, 11/14, 12:00–1:00 PM**

**This week, Suhas will post a recorded tutorial**

# Until Next Class

- The **AI Lab Proposal** is due one hour before Session 4
  - See Canvas for submission guidelines

- Refer to syllabus for readings for Session 4

# Thank You!