Modern Machine Learning

The Good and the Bad

Amartya Sanyal

Postdoc @ Max-Planck Institute for Intelligent Systems, Tübingen

(Incoming) Tenure Track Assistant Professor in Machine Learning, Department of Computer Science
(Affiliated) Assistant Professor, Department of Mathematics,
University of Copenhagen

CISPA

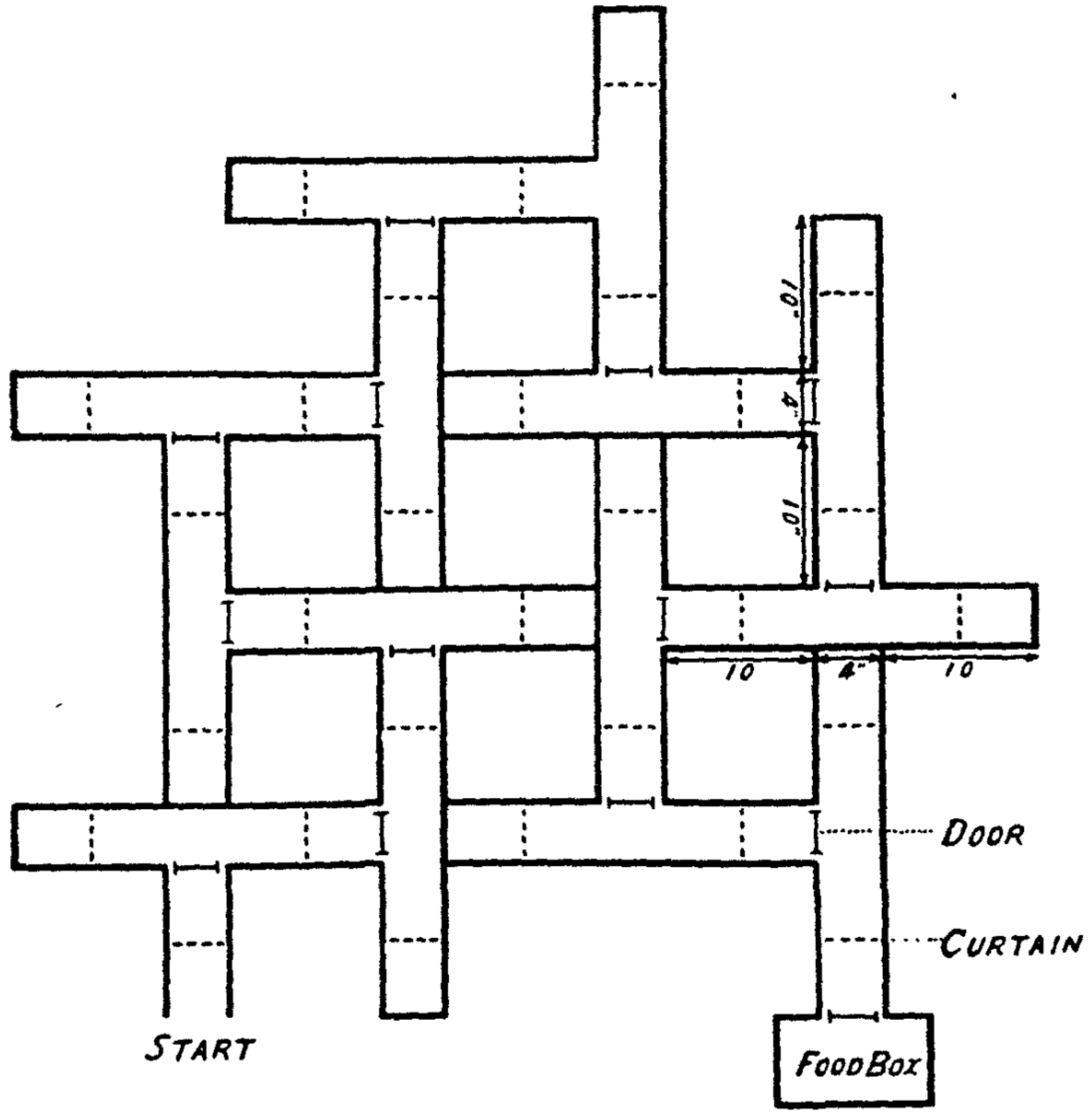
Learning

Learning

“Learning is any process by which a system improves performance from experience.” - Herbert Simon

Learning

"Learning is any process by which a system improves performance from experience." - Herbert Simon



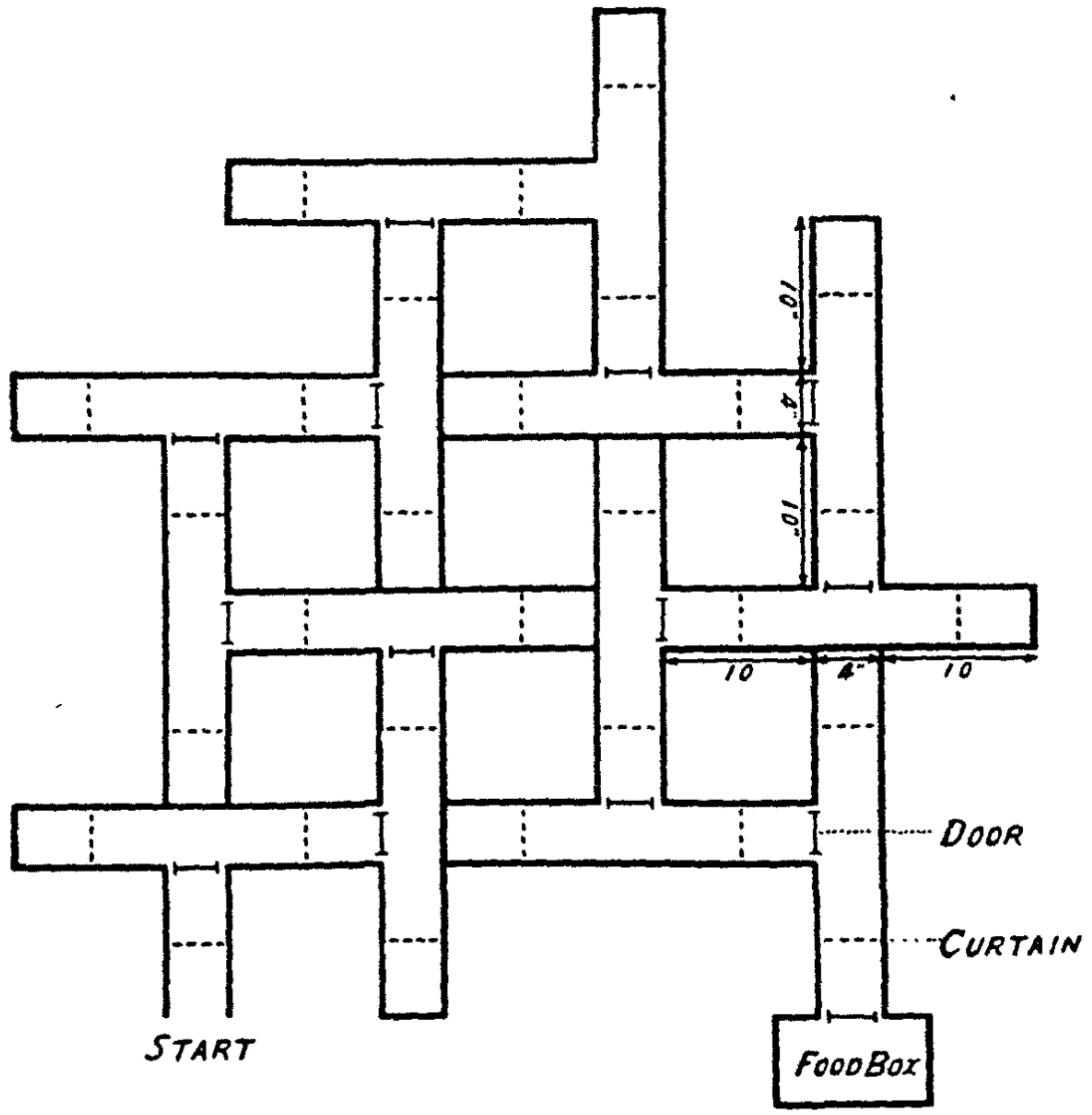
Plan of maze
14-Unit T-Alley Maze

FIG. 1

(From M. H. Elliott, The effect of change of reward on the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1928, 4, p. 20.)

Learning

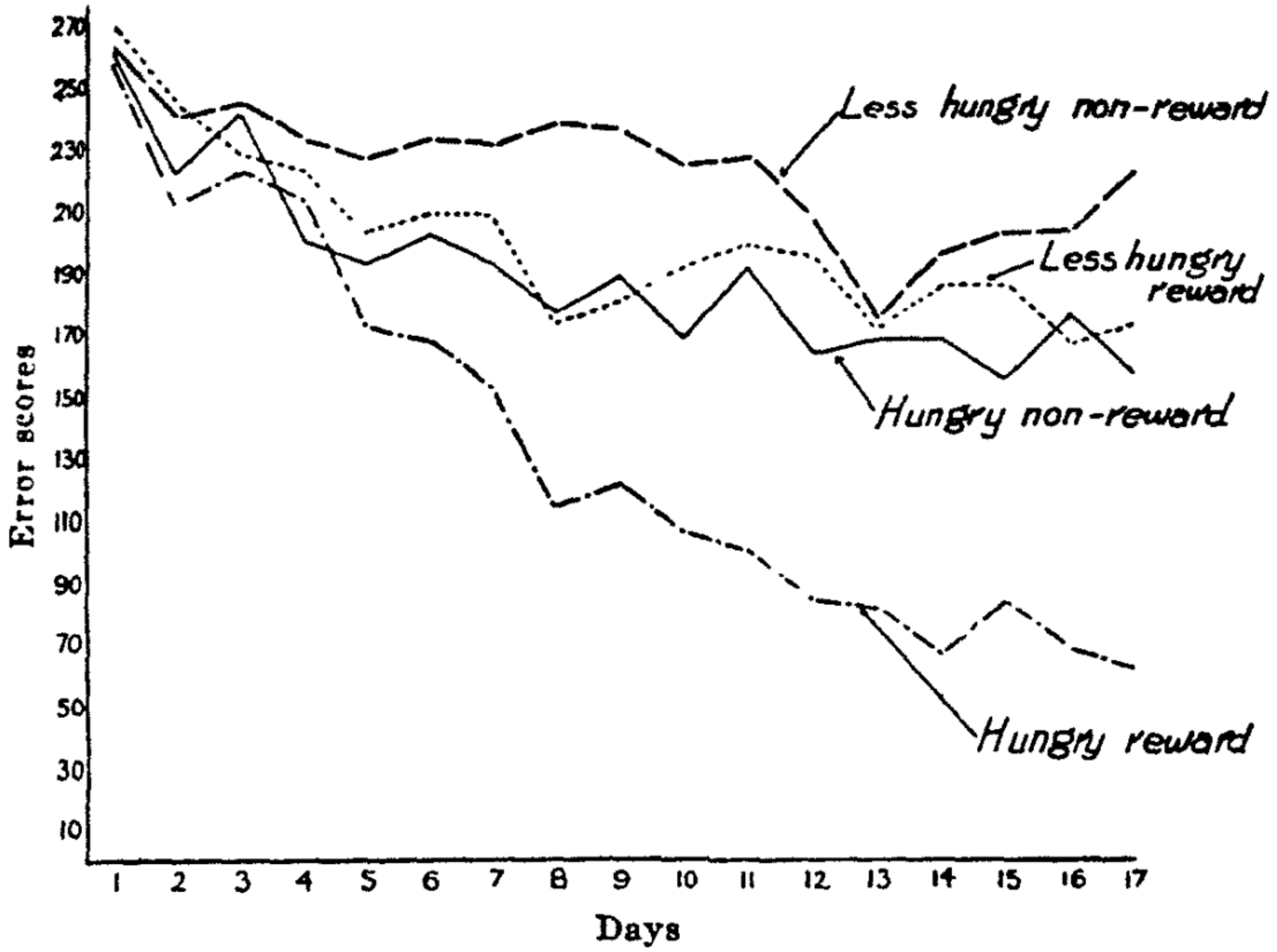
"Learning is any process by which a system improves performance from experience." - Herbert Simon



Plan of maze
14-Unit T-Alley Maze

FIG. 1

(From M. H. Elliott, The effect of change of reward on the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1928, 4, p. 20.)



Error curves for four groups, 36 rats.

FIG. 3

(From E. C. Tolman and C. H. Honzik, Degrees of hunger, reward and non-reward, and maze learning in rats. *Univ. Calif. Publ. Psychol.*, 1930, 4, No. 16, p. 246. A maze identical with the alley maze shown in Fig. 1 was used.)

Machine Learning (ML)

“Learning is any process by which a system improves performance from experience.” - Herbert Simon

Machine Learning (ML)

“Learning is any process by which a system improves performance from experience.” - Herbert Simon

“Machine Learning is the study of algorithms that improve their performance from past experience” - Tom Mitchell

Example: Thermostat

Example: Thermostat

Traditional Decision Making

Example: Thermostat

TEMP > 24 degrees

TEMP < 23 degrees

Traditional Decision Making

Example: Thermostat

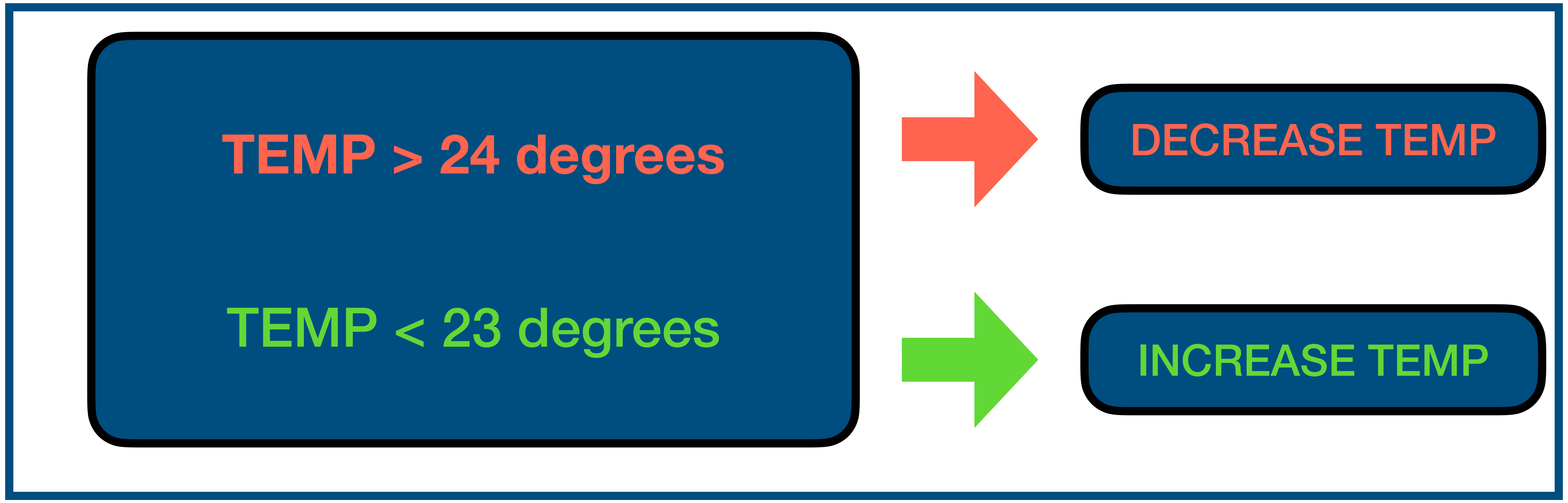
TEMP > 24 degrees

TEMP < 23 degrees

DECREASE TEMP

Traditional Decision Making

Example: Thermostat



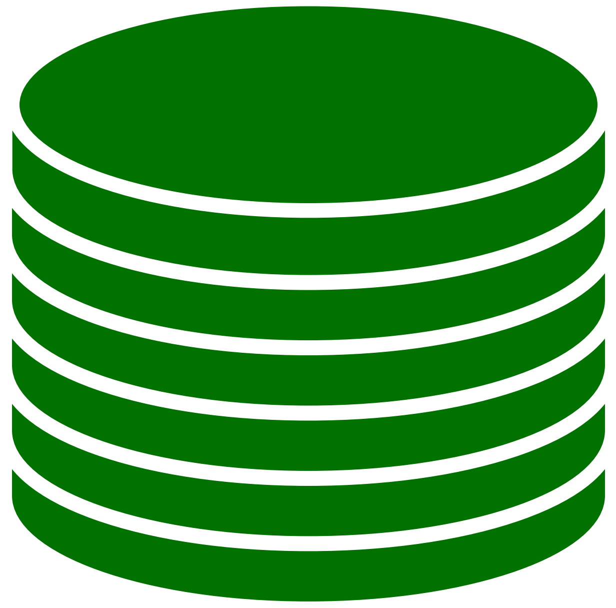
Traditional Decision Making

Example: ML for Thermostat

Traditional Decision Making

Example: ML for Thermostat

Example: ML for Thermostat



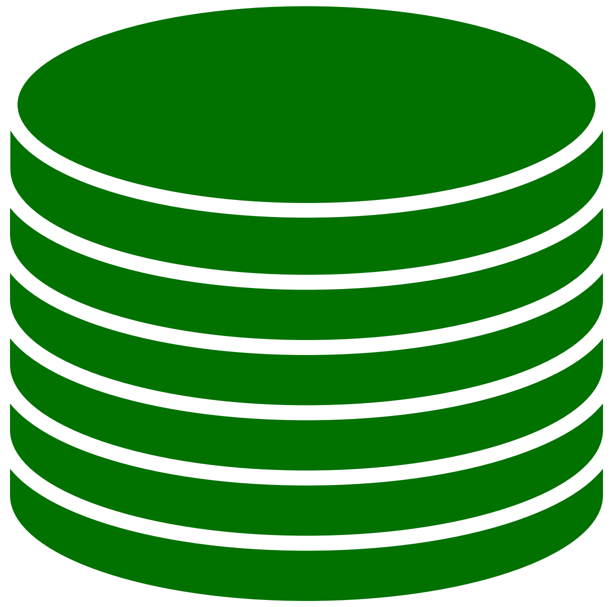
Past Data on

Machine Learning Algorithm

DECREASE TEMP

INCREASE TEMP

Example: ML for Thermostat



Past Data on

- **Past Manual Temp Change**

Machine Learning Algorithm

DECREASE TEMP

INCREASE TEMP

Example: ML for Thermostat

Past Data on

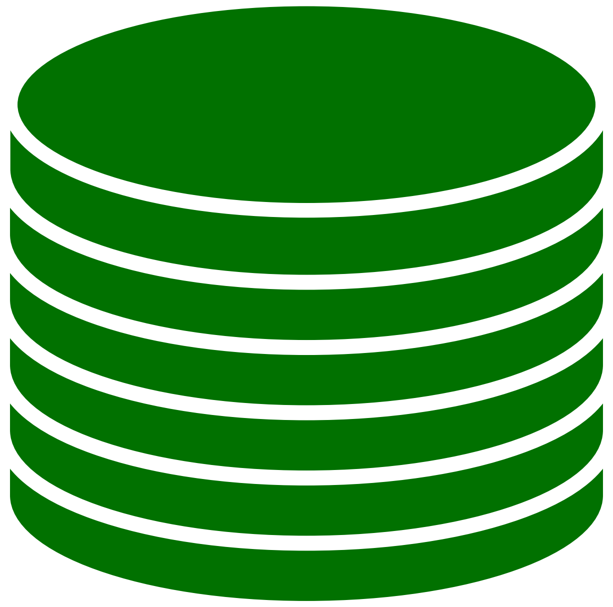
- Past Manual Temp Change
- Time of Day

Machine Learning Algorithm

DECREASE TEMP

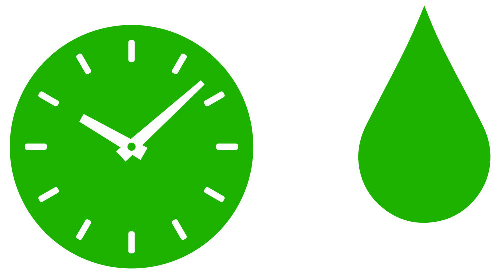
INCREASE TEMP

Example: ML for Thermostat



<u>Past Data on</u>

- Past Manual Temp Change
- Time of Day
- Humidity



Machine Learning Algorithm

DECREASE TEMP

INCREASE TEMP

Example: ML for Thermostat

Past Data on

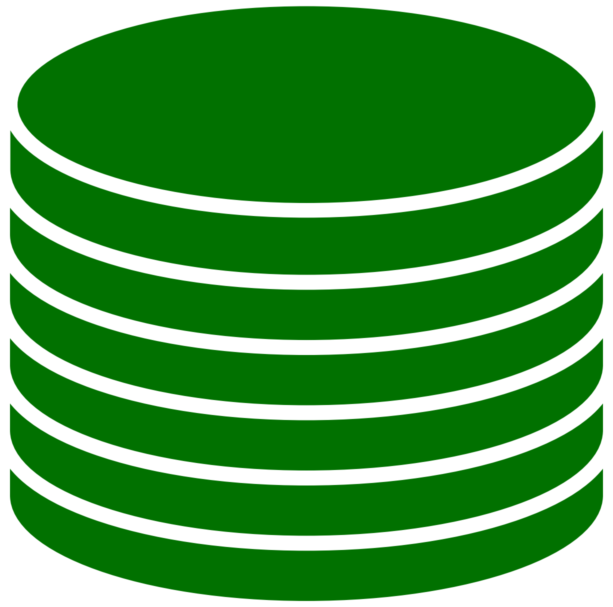
- Past Manual Temp Change
- Time of Day
- Humidity
- Room Temperature

Machine Learning Algorithm

DECREASE TEMP

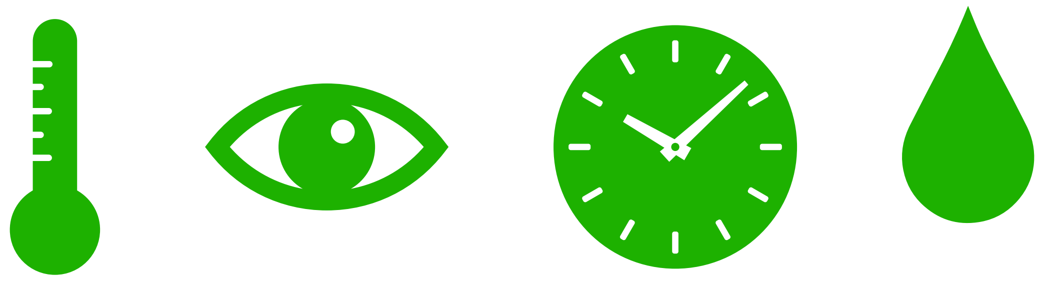
INCREASE TEMP

Example: ML for Thermostat



Past Data on

- Past Manual Temp Change
- Time of Day
- Humidity
- Room Temperature
- Number of people in room

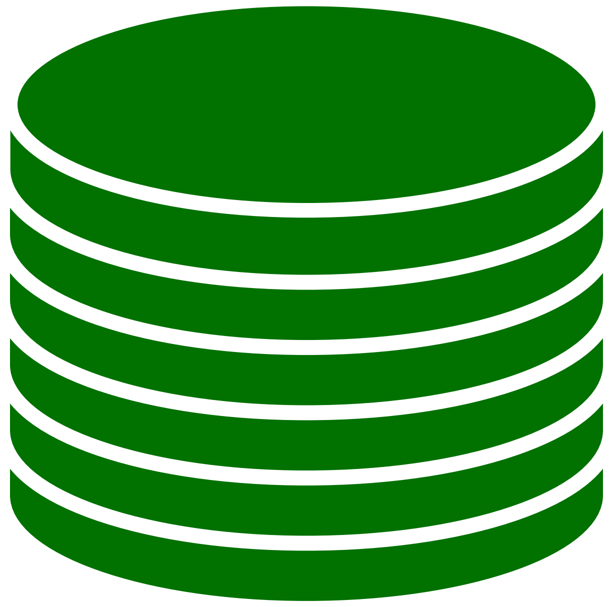


Machine Learning Algorithm

DECREASE TEMP

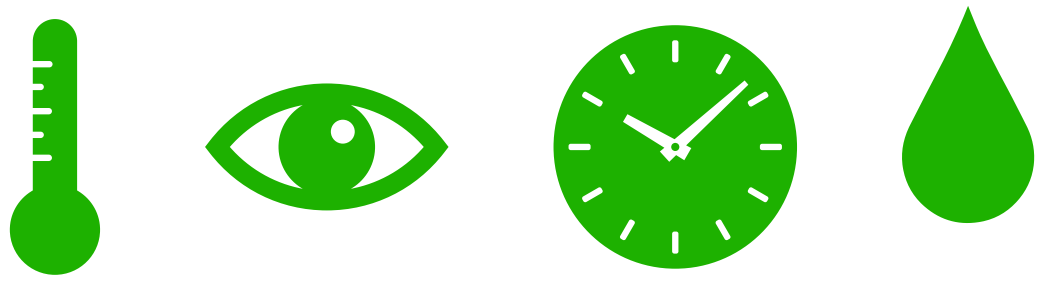
INCREASE TEMP

Example: ML for Thermostat



<u>Past Data on</u>

- Past Manual Temp Change
- Time of Day
- Humidity
- Room Temperature
- Number of people in room

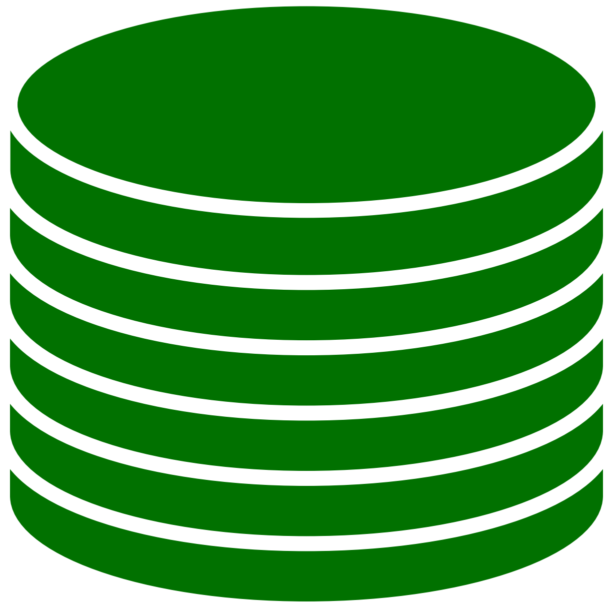


Machine Learning Algorithm

DECREASE TEMP

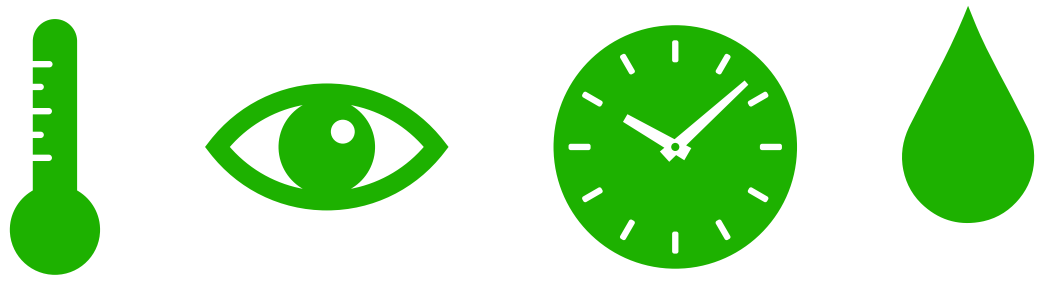
INCREASE TEMP

Example: ML for Thermostat



Past Data on

- Past Manual Temp Change
- Time of Day
- Humidity
- Room Temperature
- Number of people in room



Machine Learning Algorithm

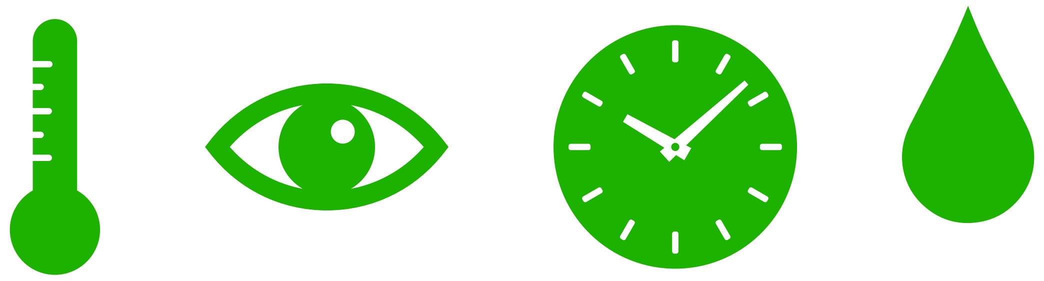
DECREASE TEMP

INCREASE TEMP

Example: ML for Thermostat

Past Data on

- Past Manual Temp Change
- Time of Day
- Humidity
- Room Temperature
- Number of people in room



Machine Learning Algorithm

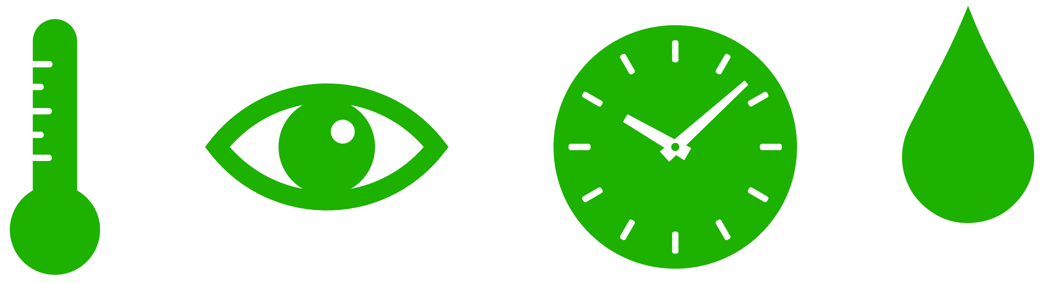
DECREASE TEMP

INCREASE TEMP

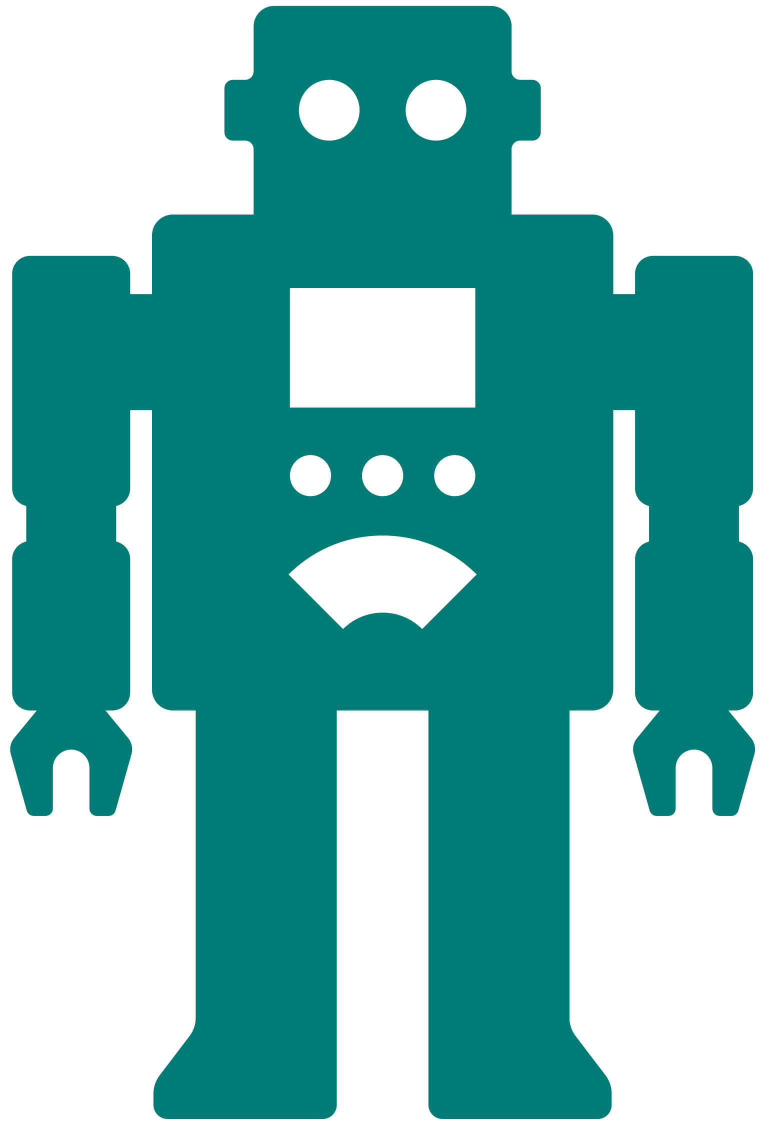
Example: ML for Thermostat

Past Data on

- Past Manual Temp Change
- Time of Day
- Humidity
- Room Temperature
- Number of people in room



Machine Learning Algorithm



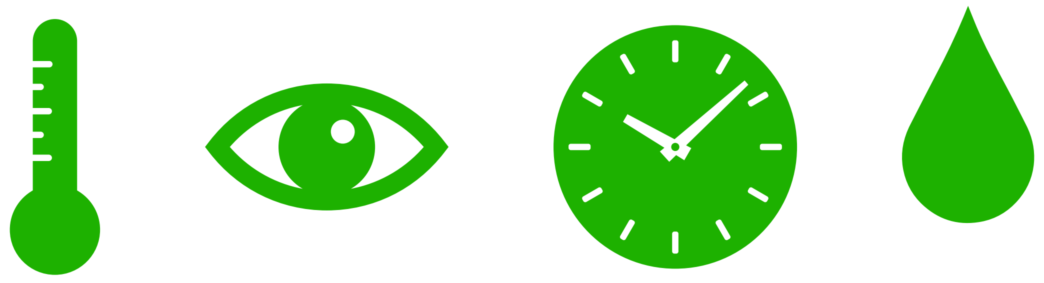
ML Model

DECREASE TEMP

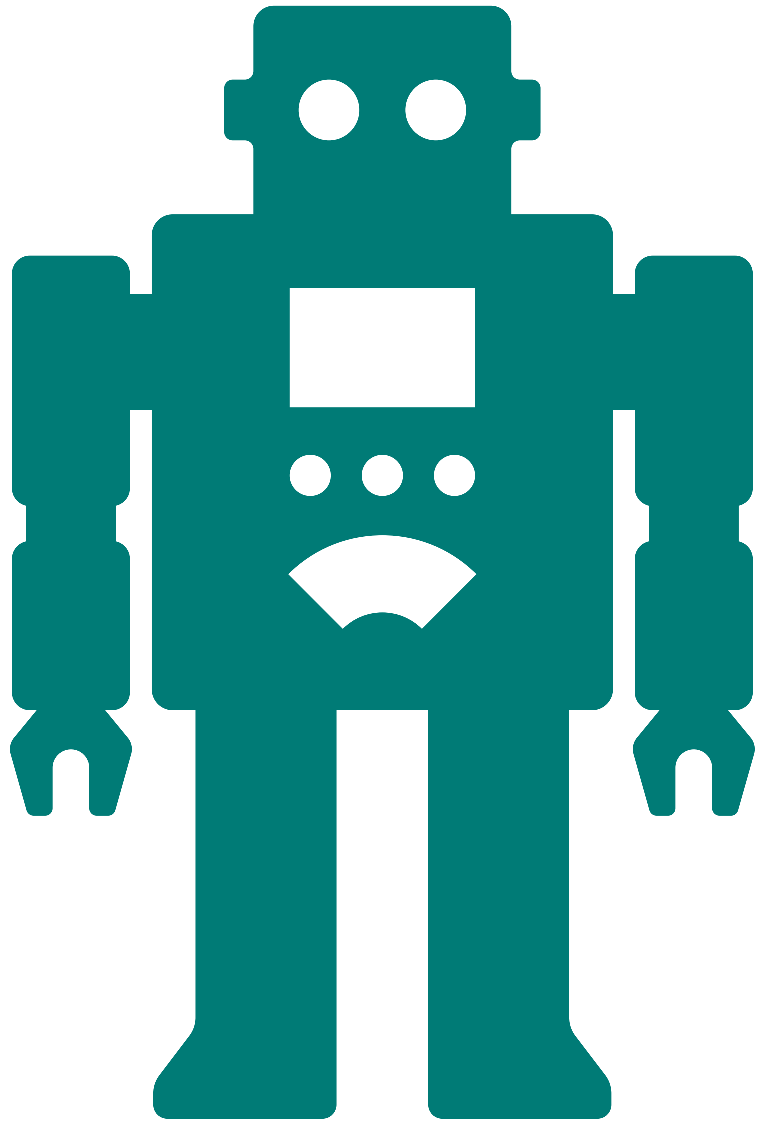
INCREASE TEMP

Example: ML for Thermostat

Past Data on
- Past Manual Temp Change
- Time of Day
- Humidity
- Room Temperature
- Number of people in room



Machine Learning Algorithm

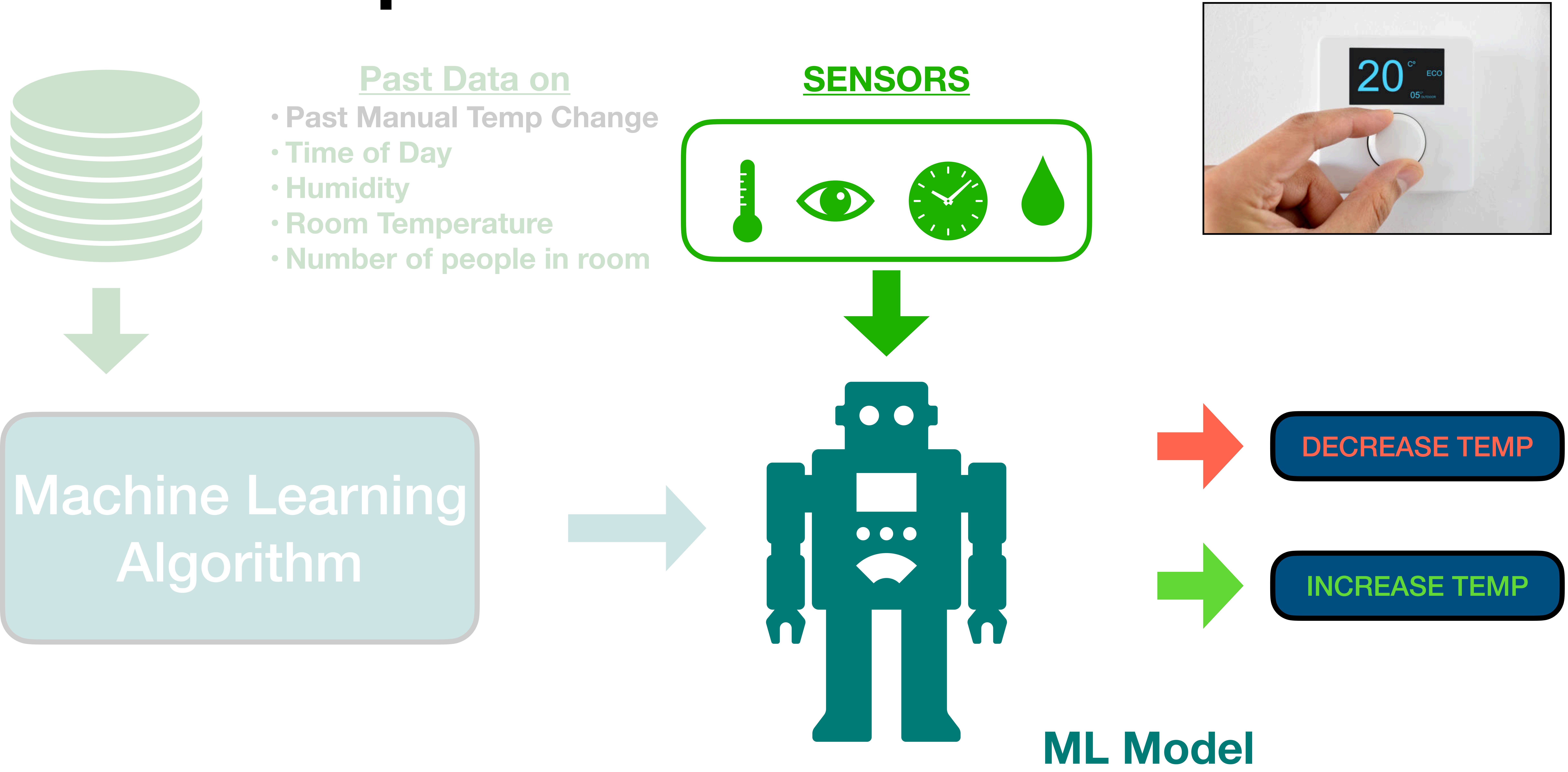


DECREASE TEMP

INCREASE TEMP

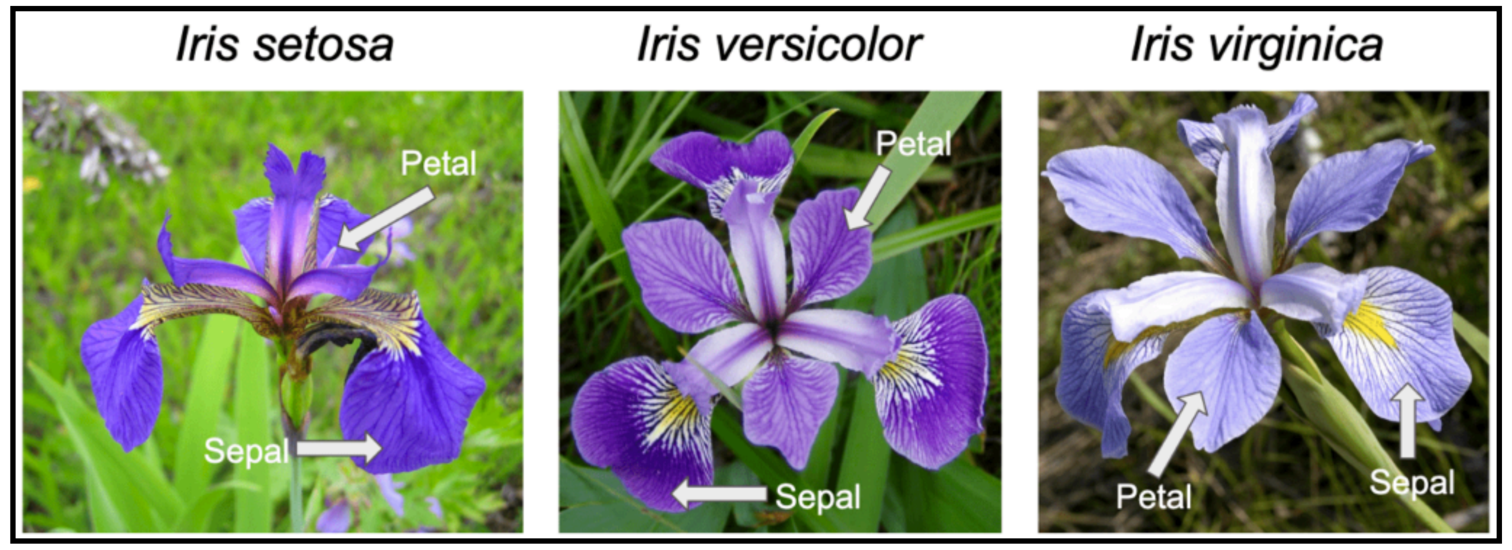
ML Model

Example: ML for Thermostat

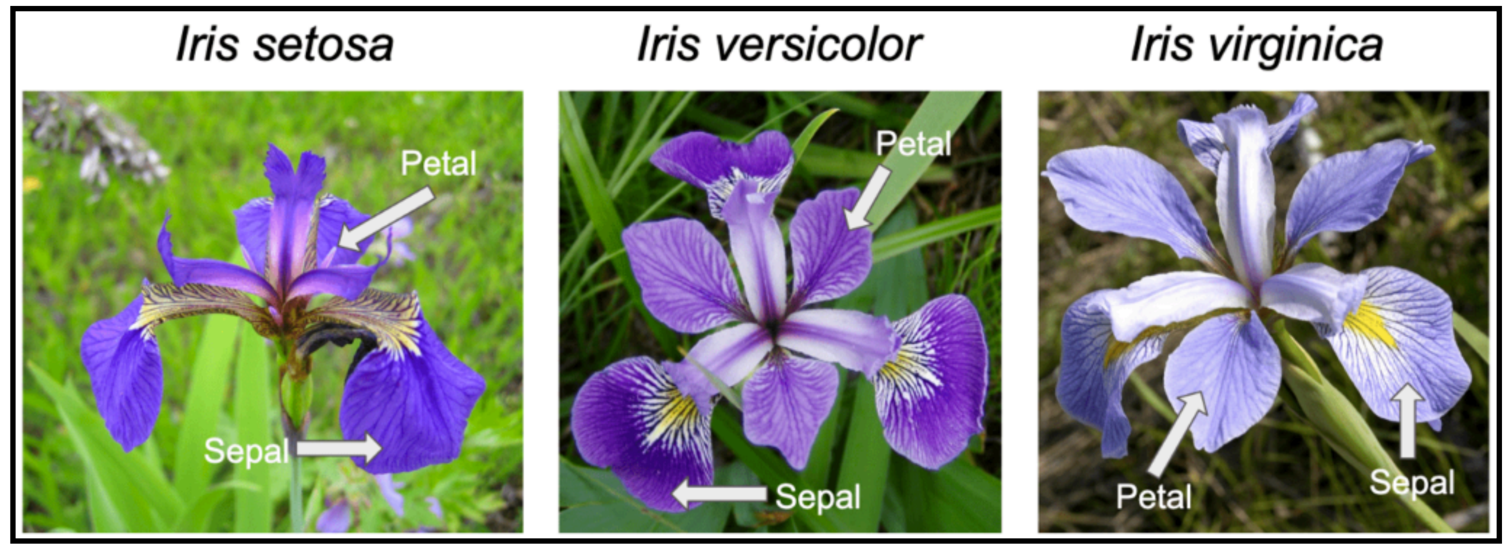


Application: Multi-Class Classification

Application: Multi-Class Classification

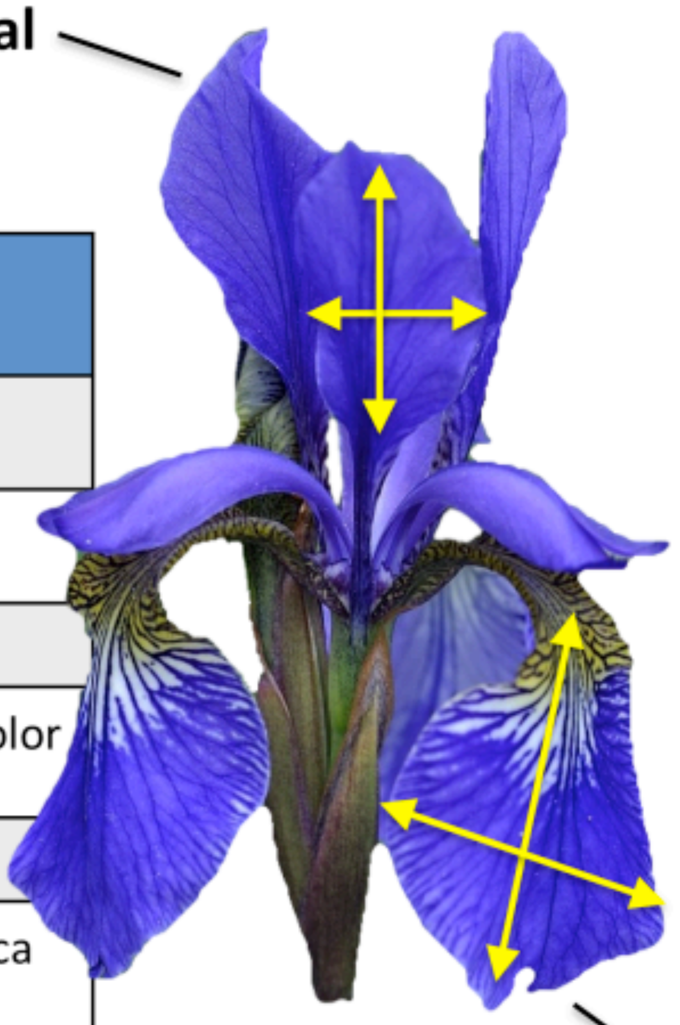


Application: Multi-Class Classification



Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
			...		
50	6.4	3.5	4.5	1.2	Versicolor
			...		
150	5.9	3.0	5.0	1.8	Virginica

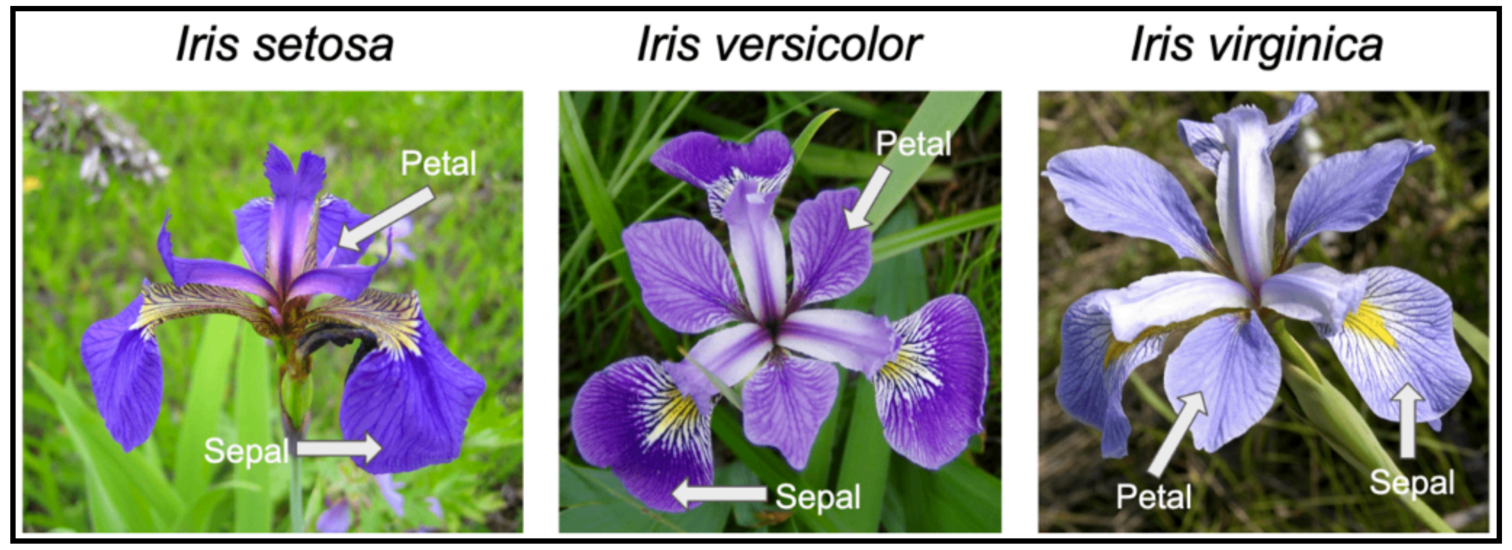


Class labels
(targets)

Features
(attributes, measurements, dimensions)

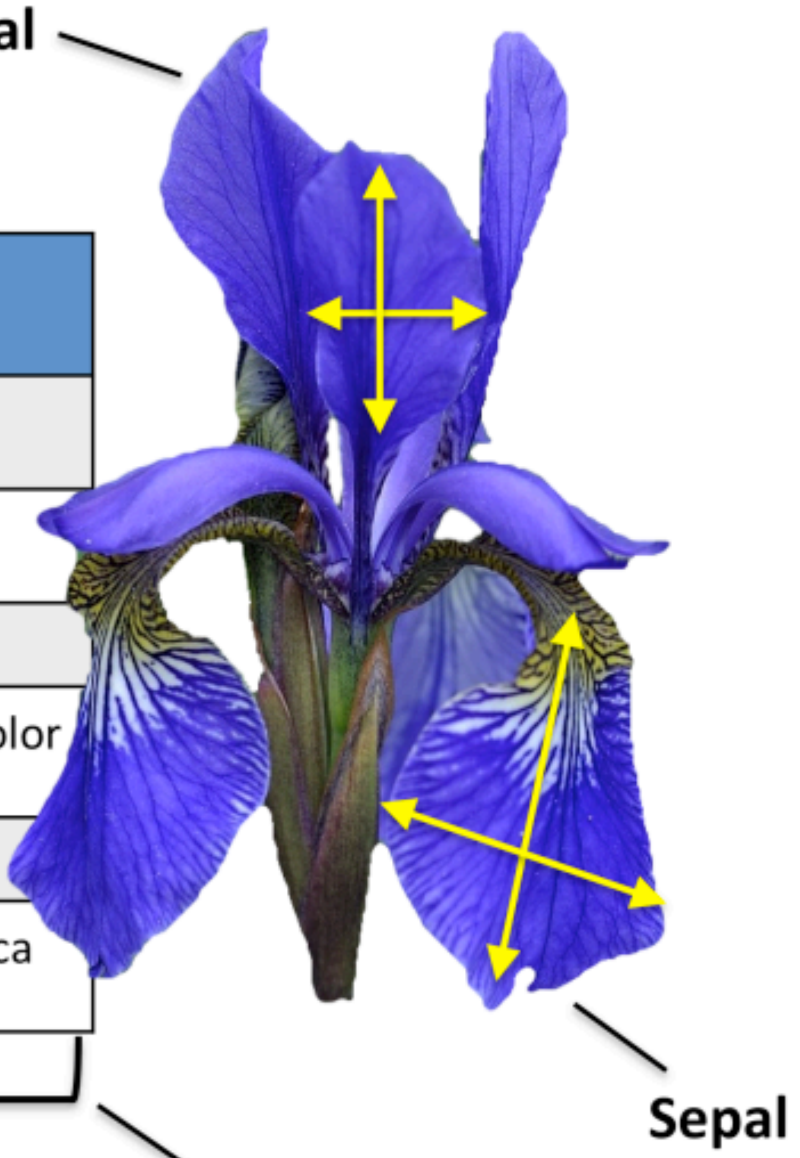
150 data points, 3 classes
4 features

Application: Multi-Class Classification



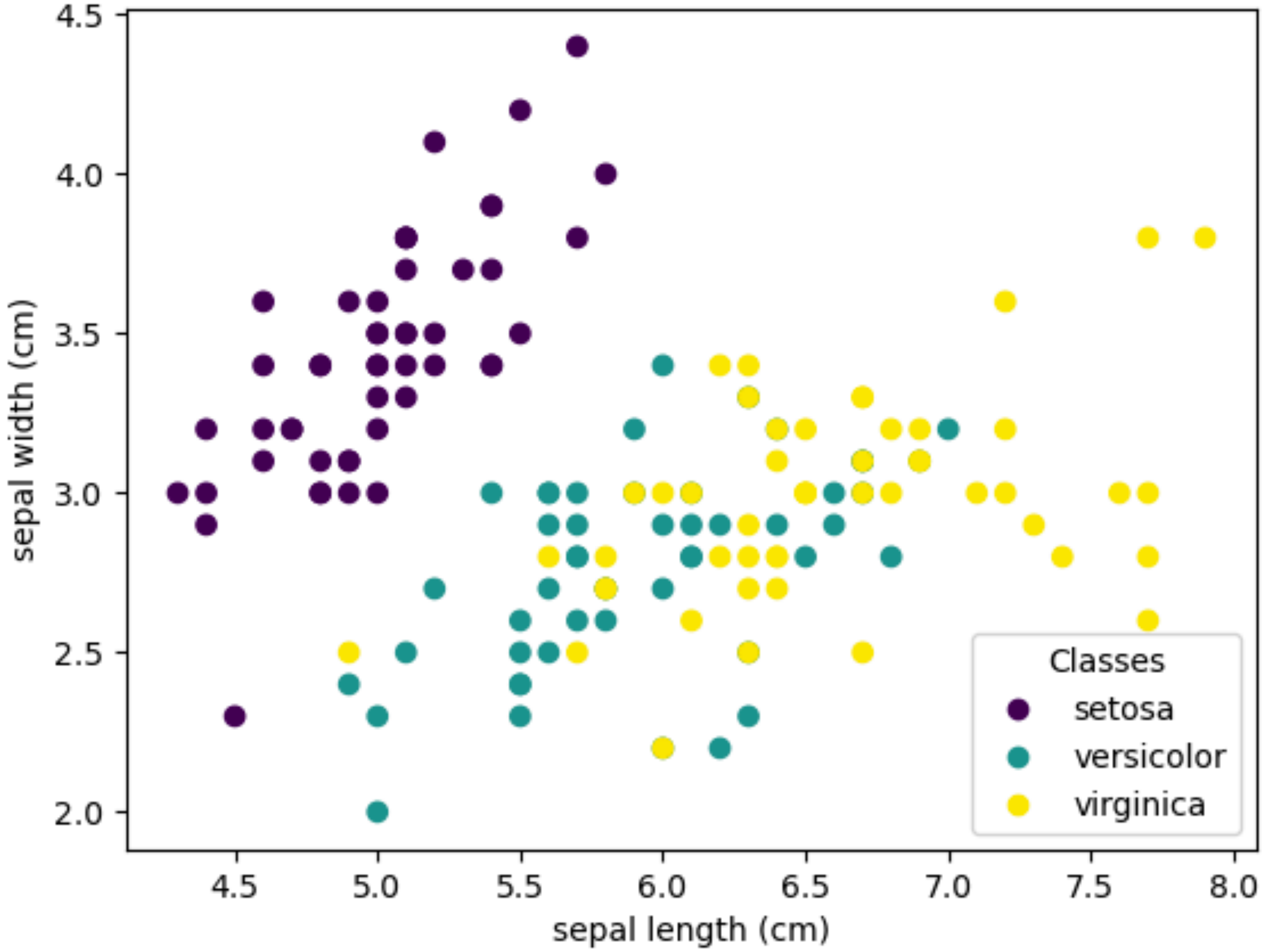
Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
			...		
50	6.4	3.5	4.5	1.2	Versicolor
			...		
150	5.9	3.0	5.0	1.8	Virginica



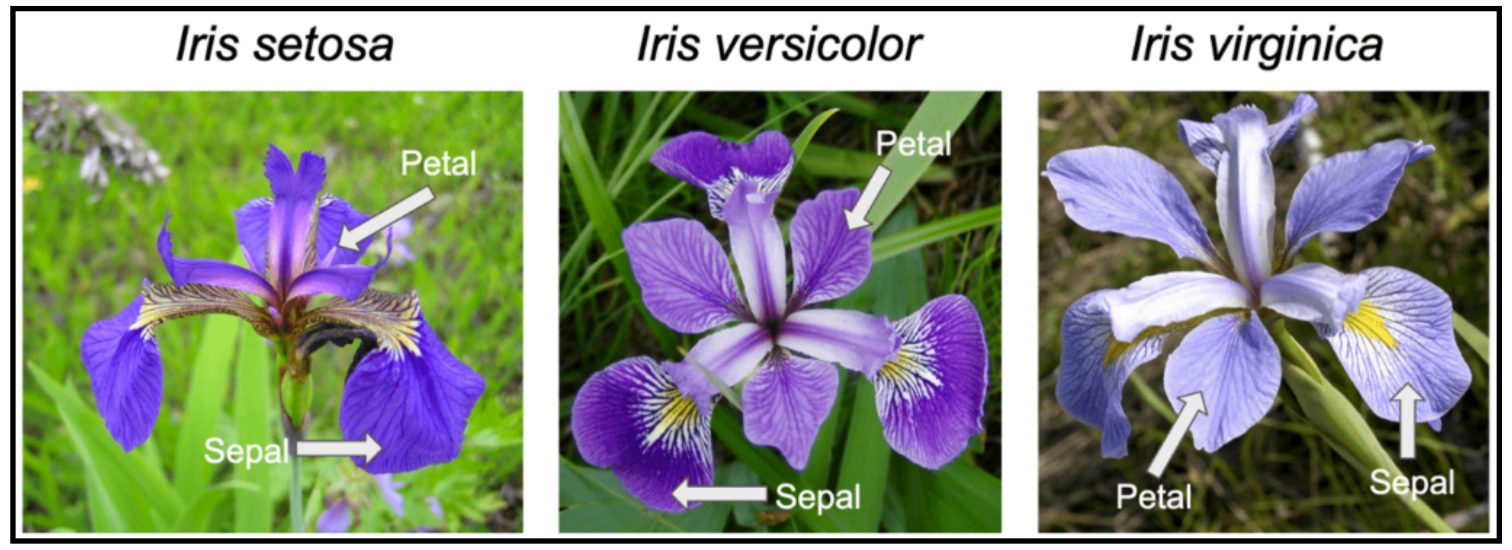
Features
(attributes, measurements, dimensions)

Class labels
(targets)



150 data points, 3 classes
4 features

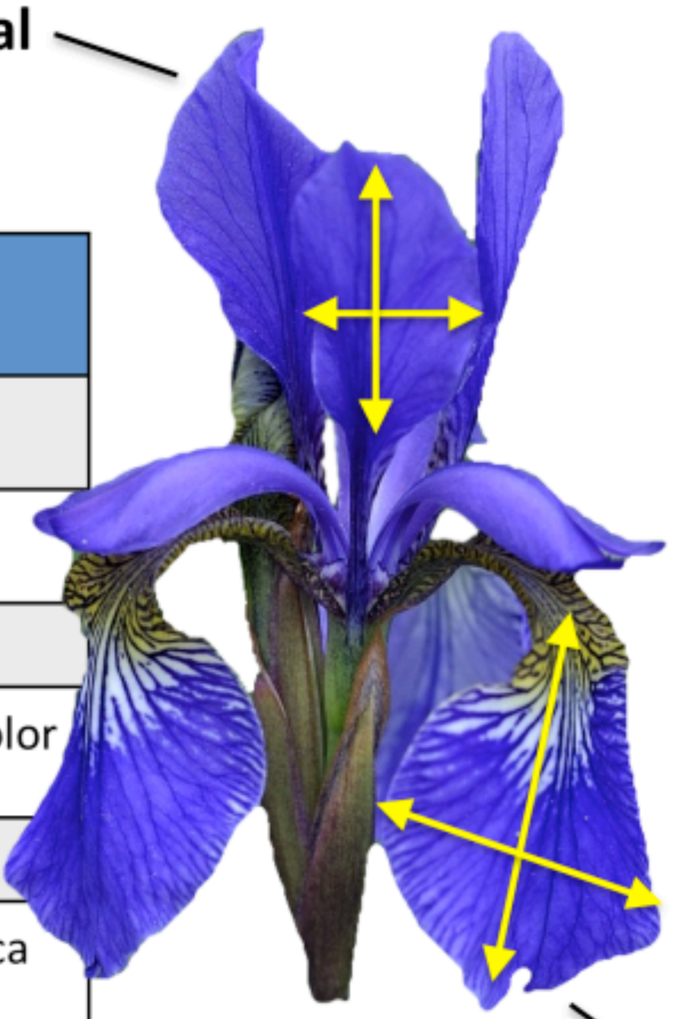
Application: Multi-Class Classification



Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

Petal

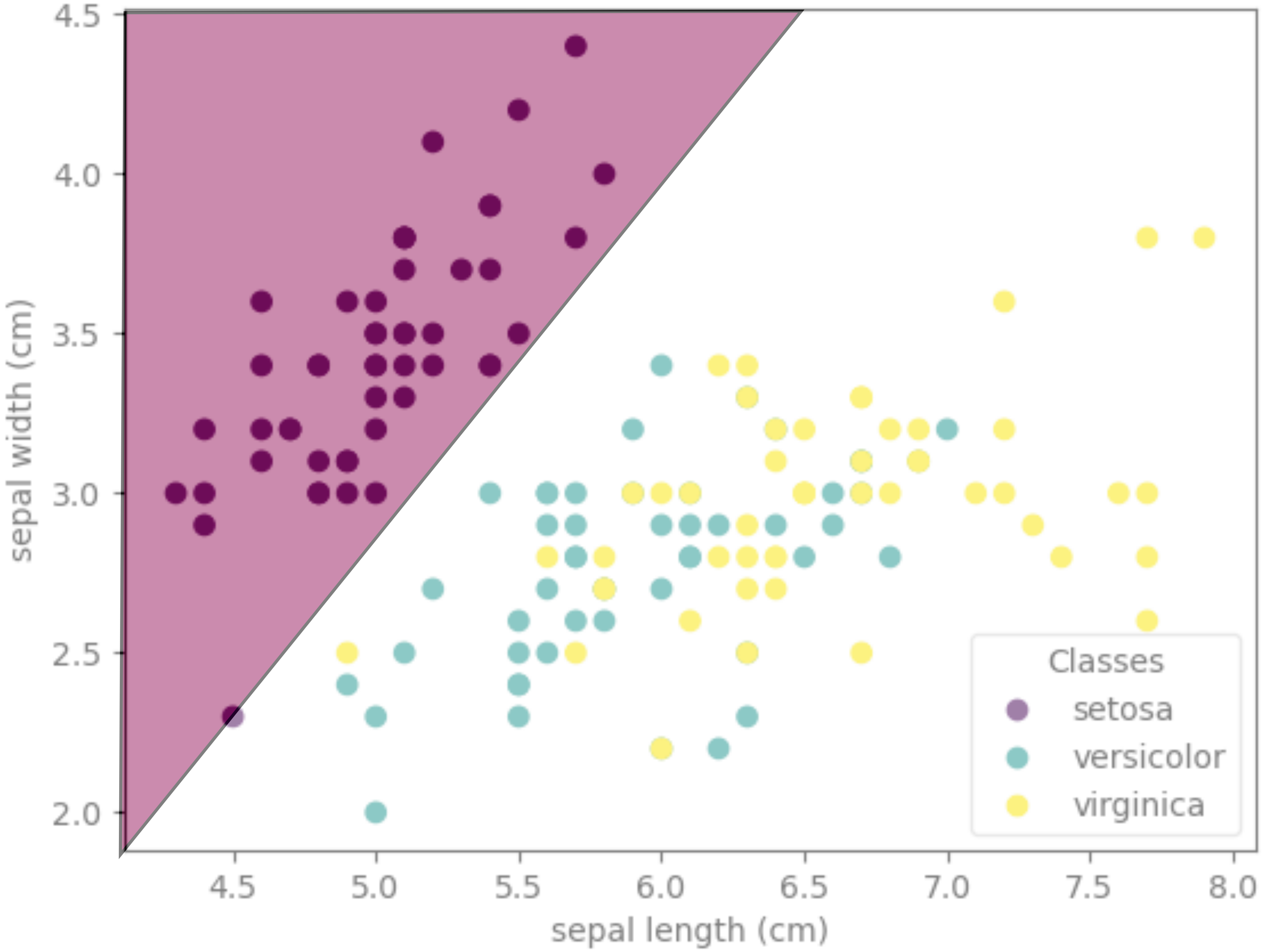


Sepal

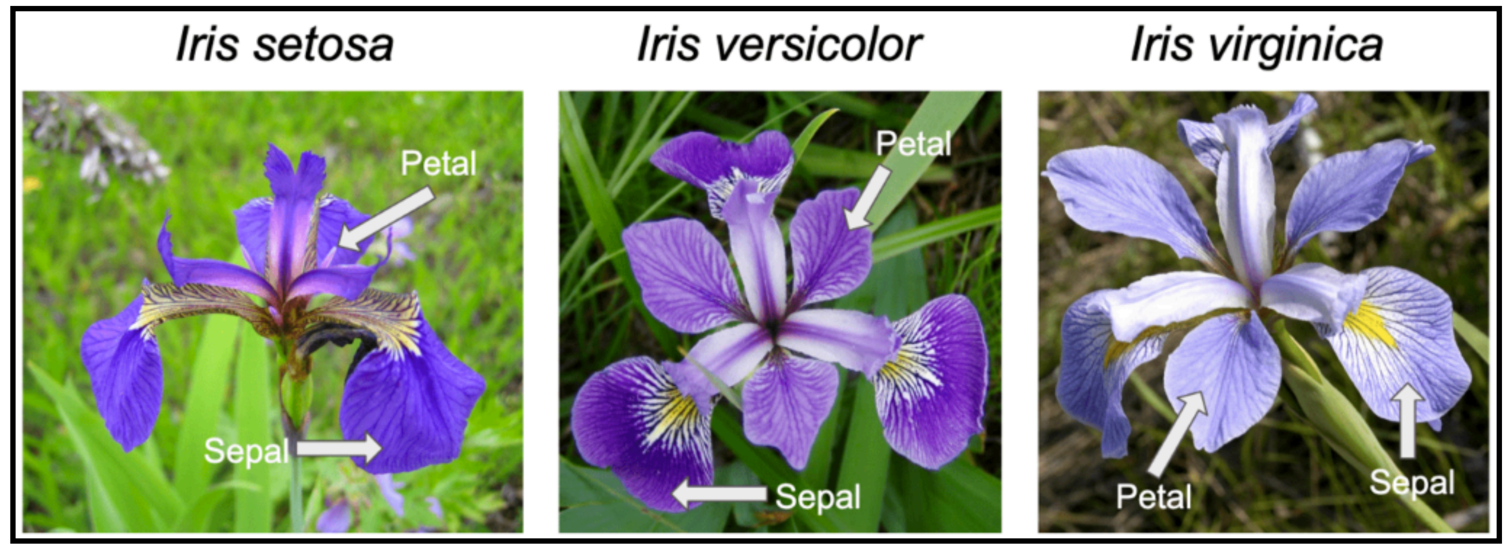
Class labels
(targets)

Features
(attributes, measurements, dimensions)

150 data points, 3 classes
4 features

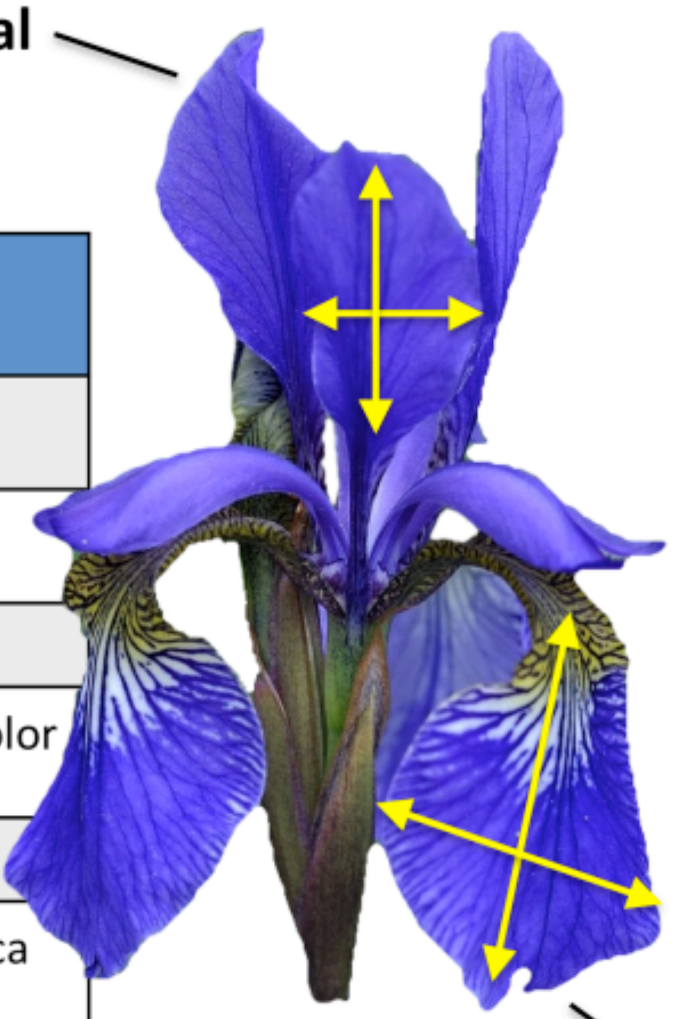


Application: Multi-Class Classification



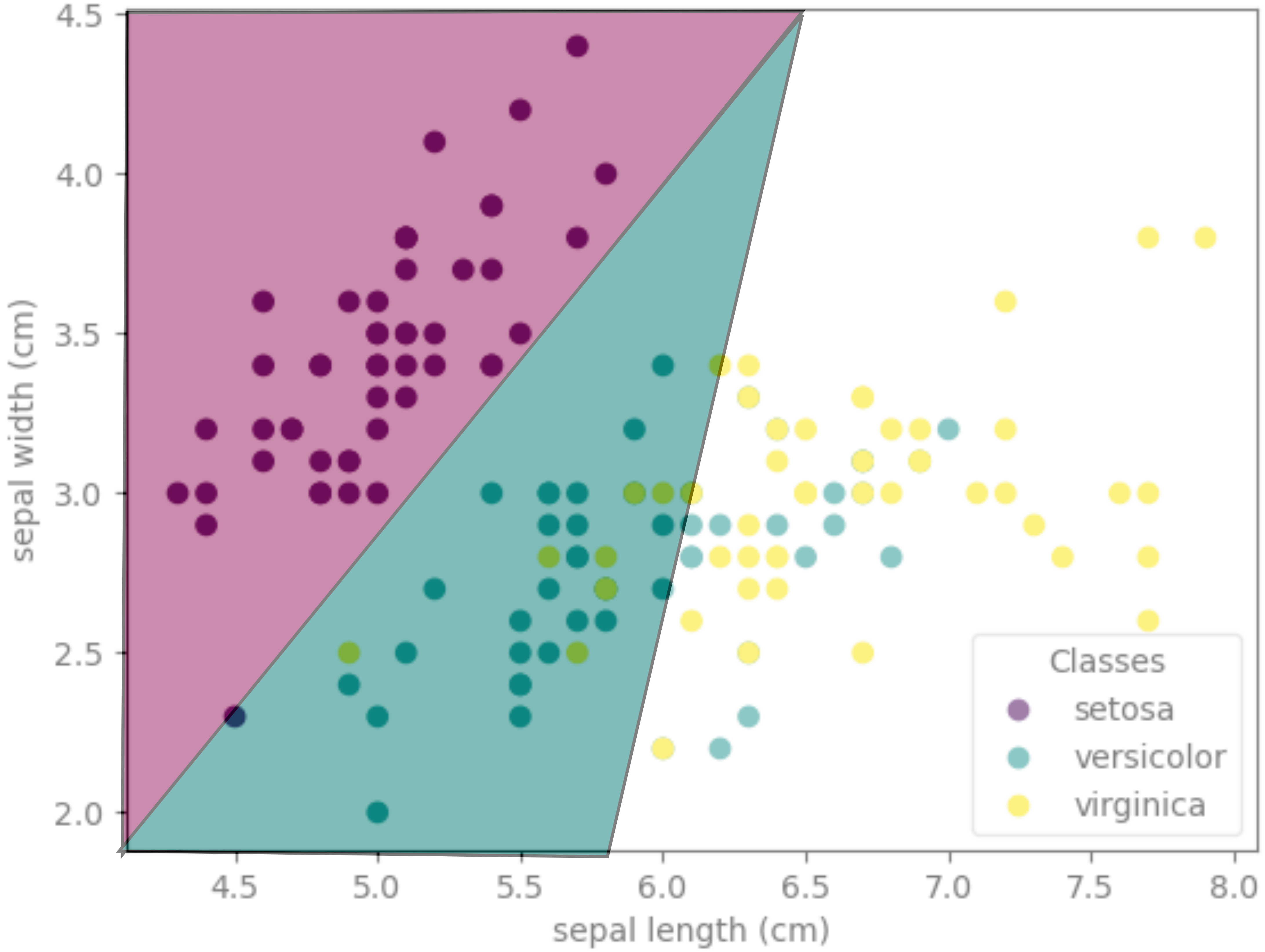
Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica



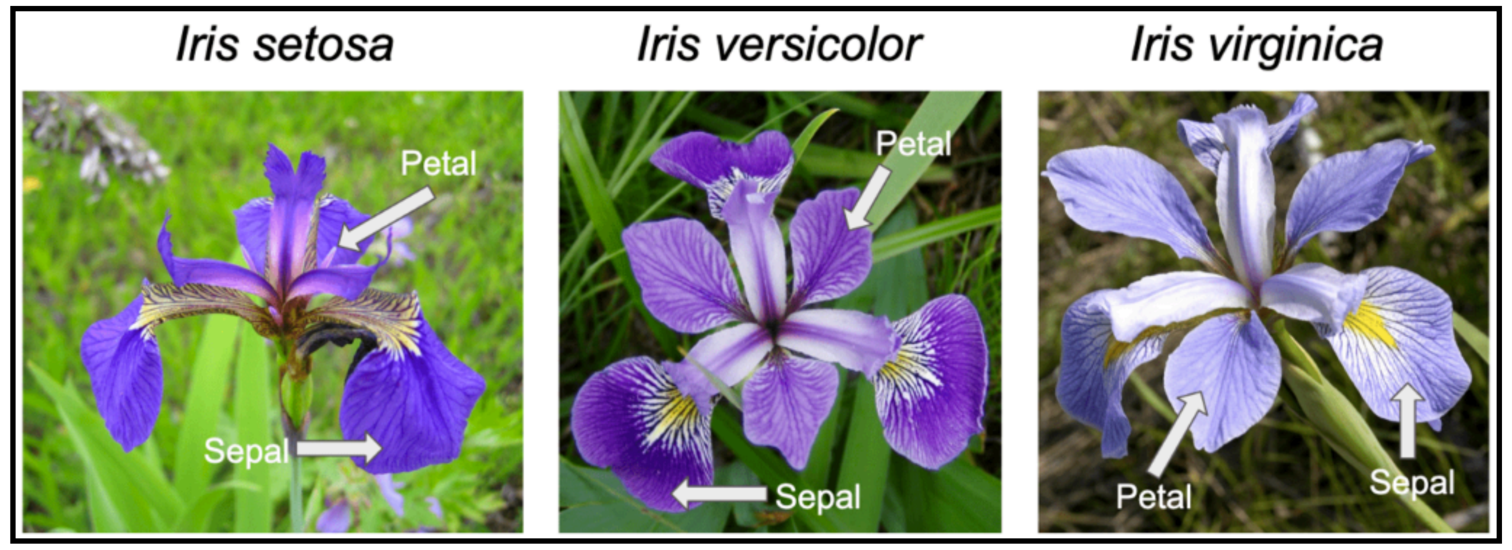
Features
(attributes, measurements, dimensions)

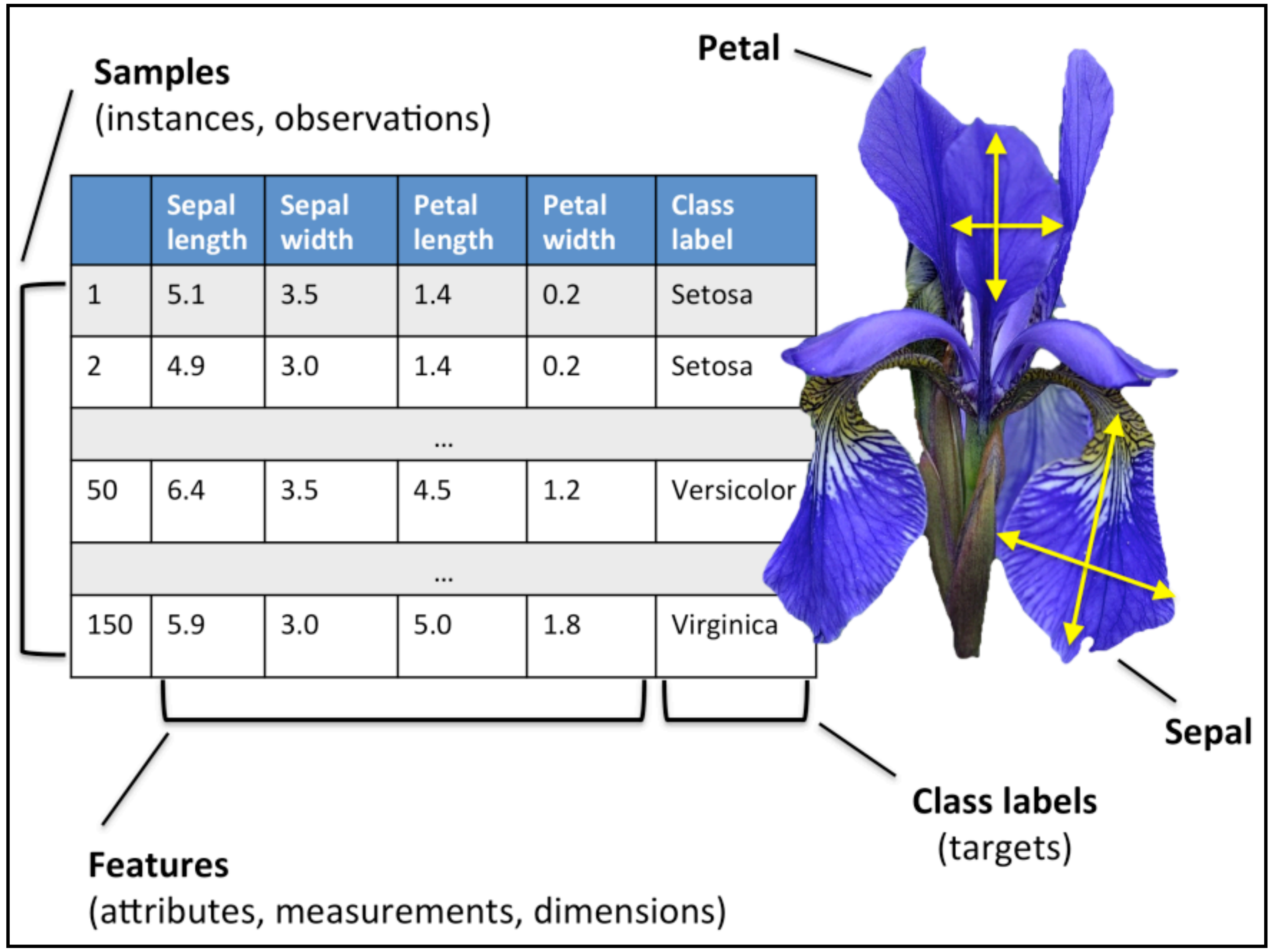
Class labels
(targets)



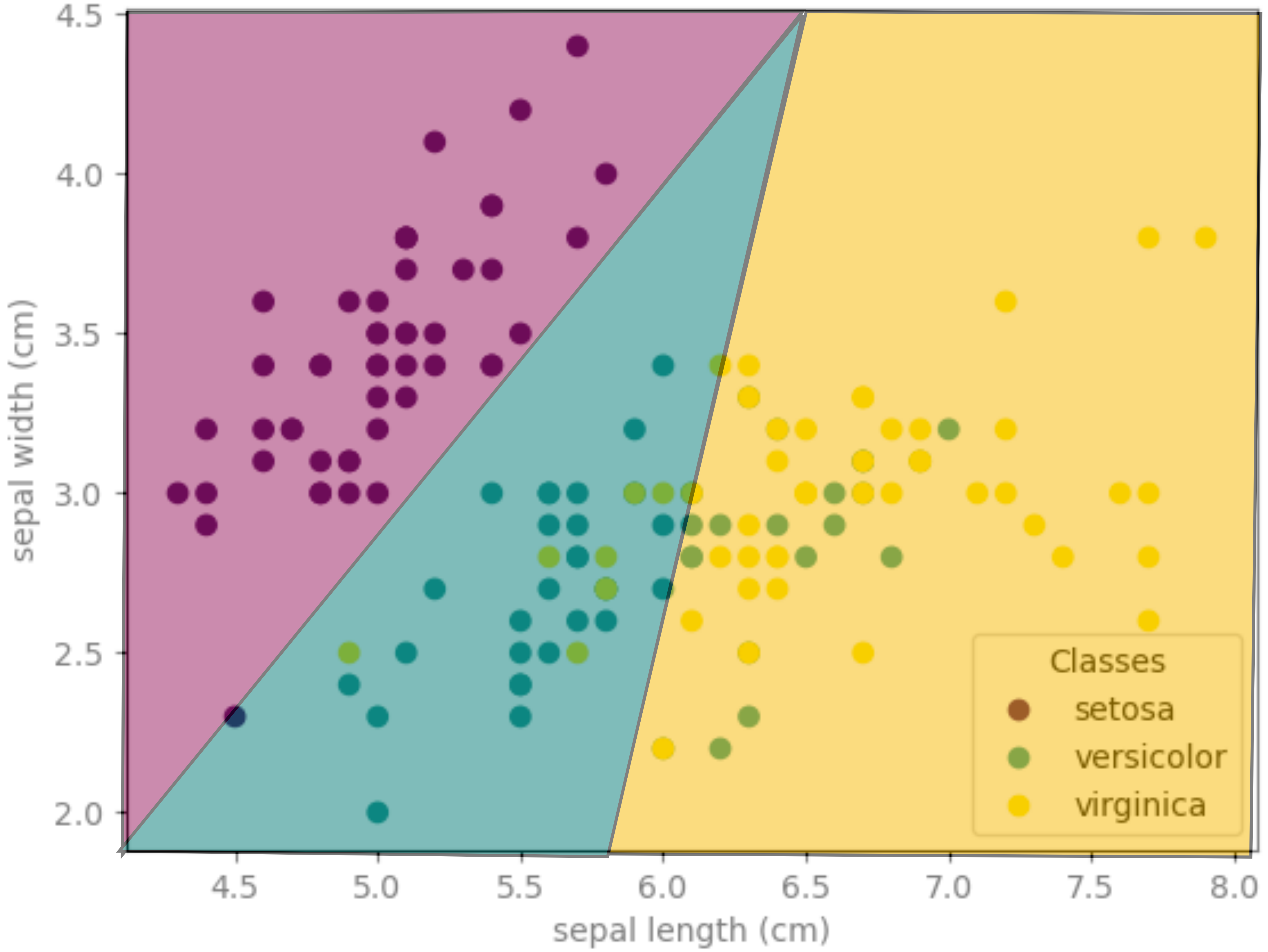
150 data points, 3 classes
4 features

Application: Multi-Class Classification



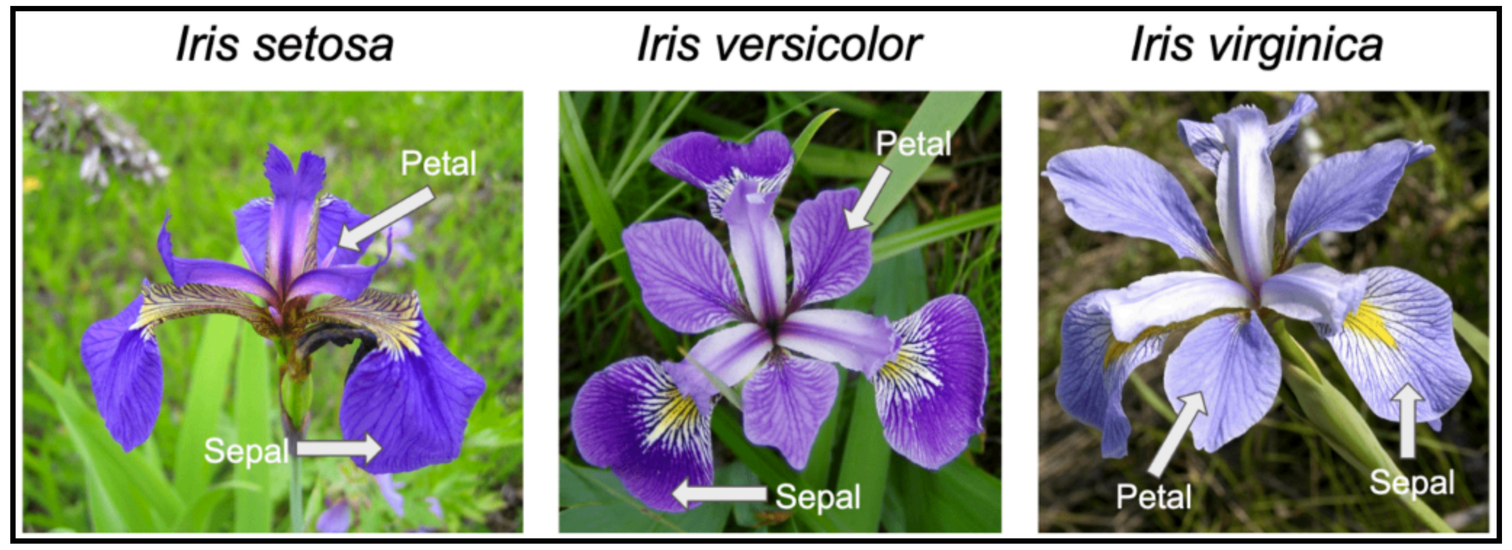


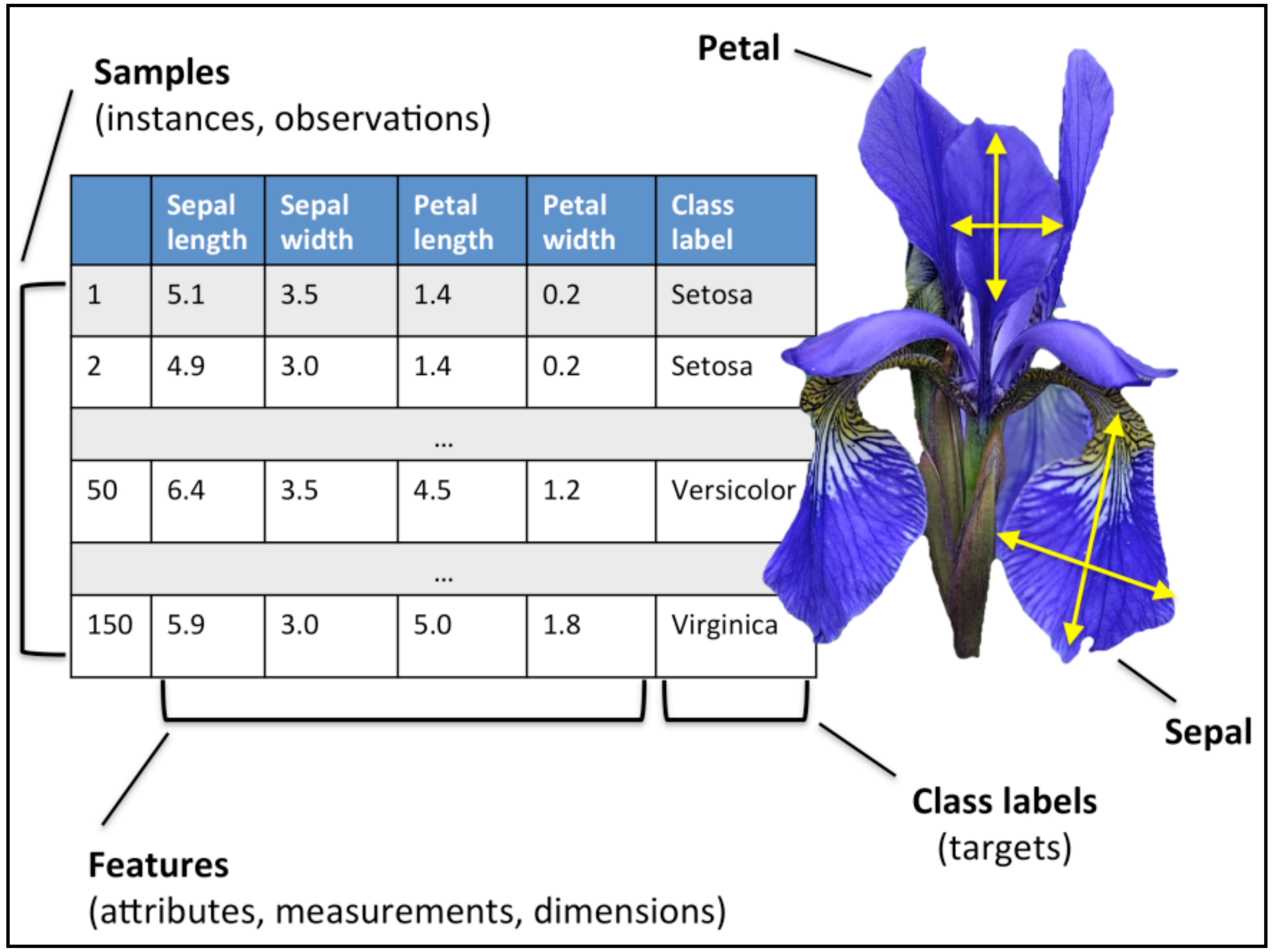
	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica



150 data points, 3 classes
4 features

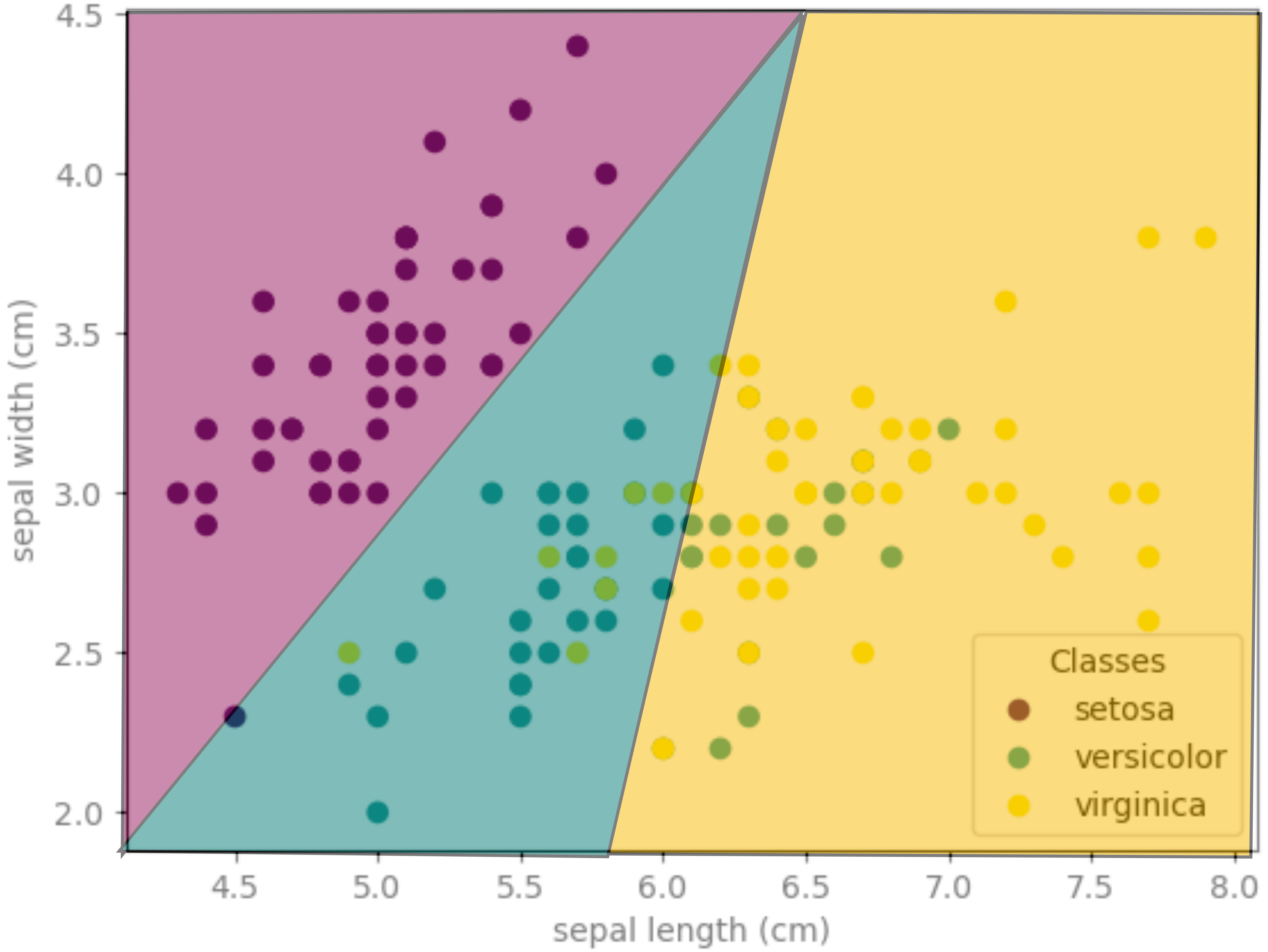
Application: Multi-Class Classification





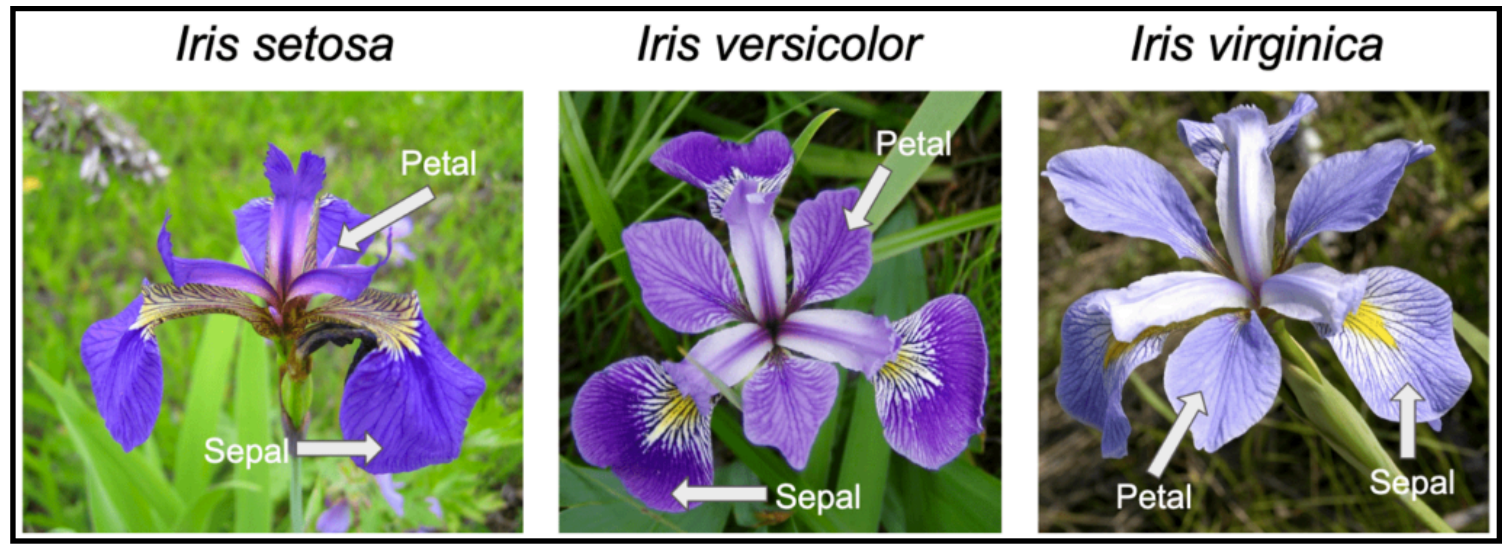
	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
			...		
50	6.4	3.5	4.5	1.2	Versicolor
			...		
150	5.9	3.0	5.0	1.8	Virginica

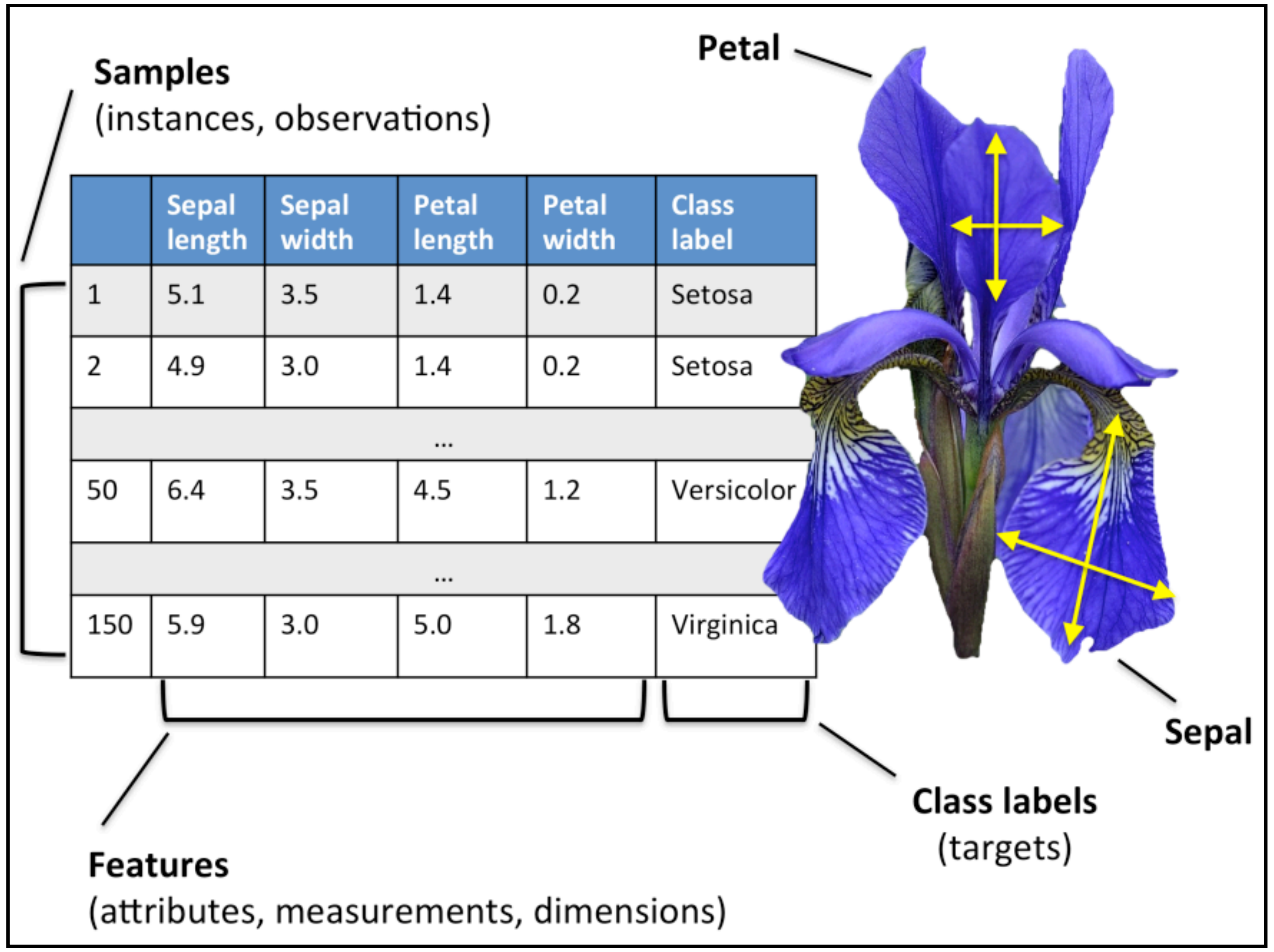
150 data points, 3 classes
4 features



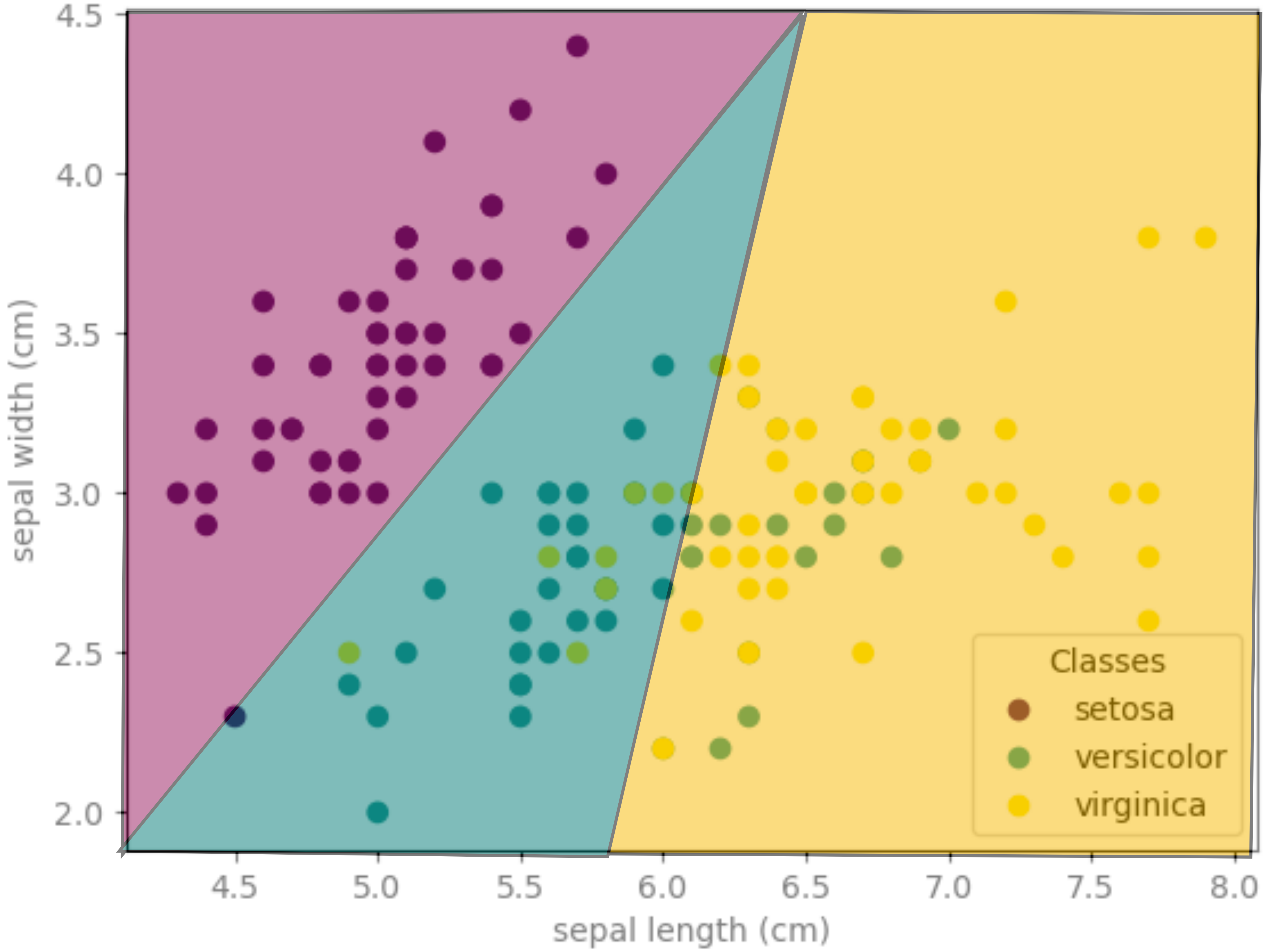
Each line can be represented as

Application: Multi-Class Classification





	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
			...		
50	6.4	3.5	4.5	1.2	Versicolor
			...		
150	5.9	3.0	5.0	1.8	Virginica

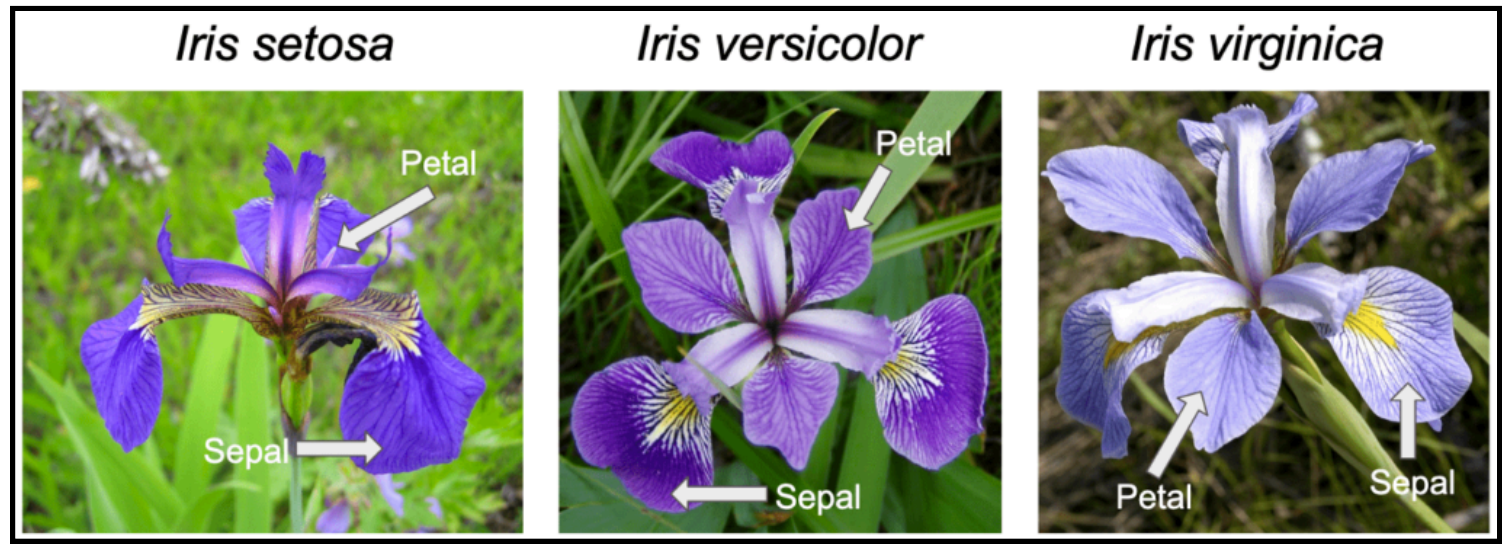


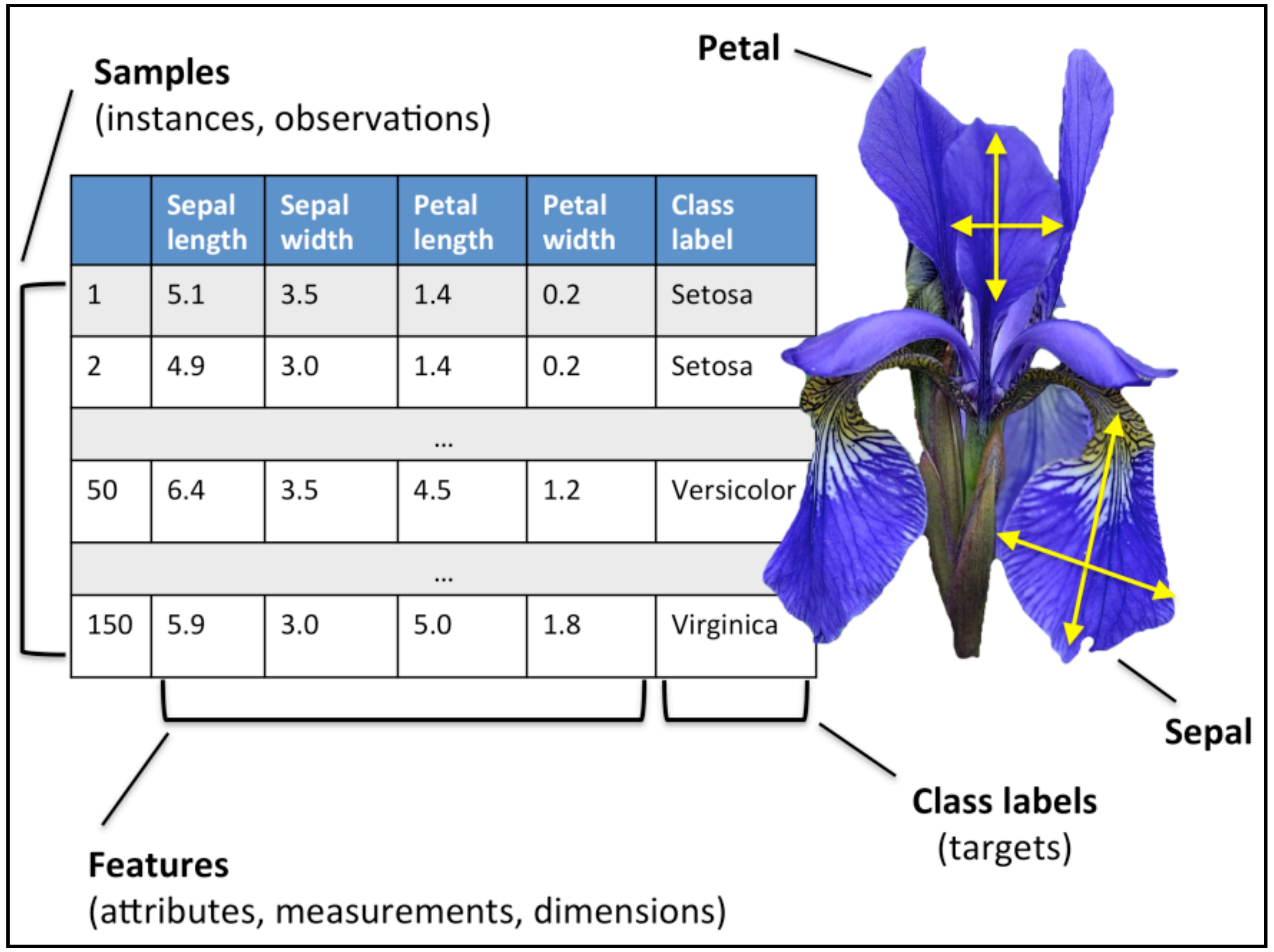
Each line can be represented as

$$w_1 \cdot (\text{sepal length}) + w_2 \cdot (\text{sepal width}) + b = 0$$

150 data points, 3 classes
4 features

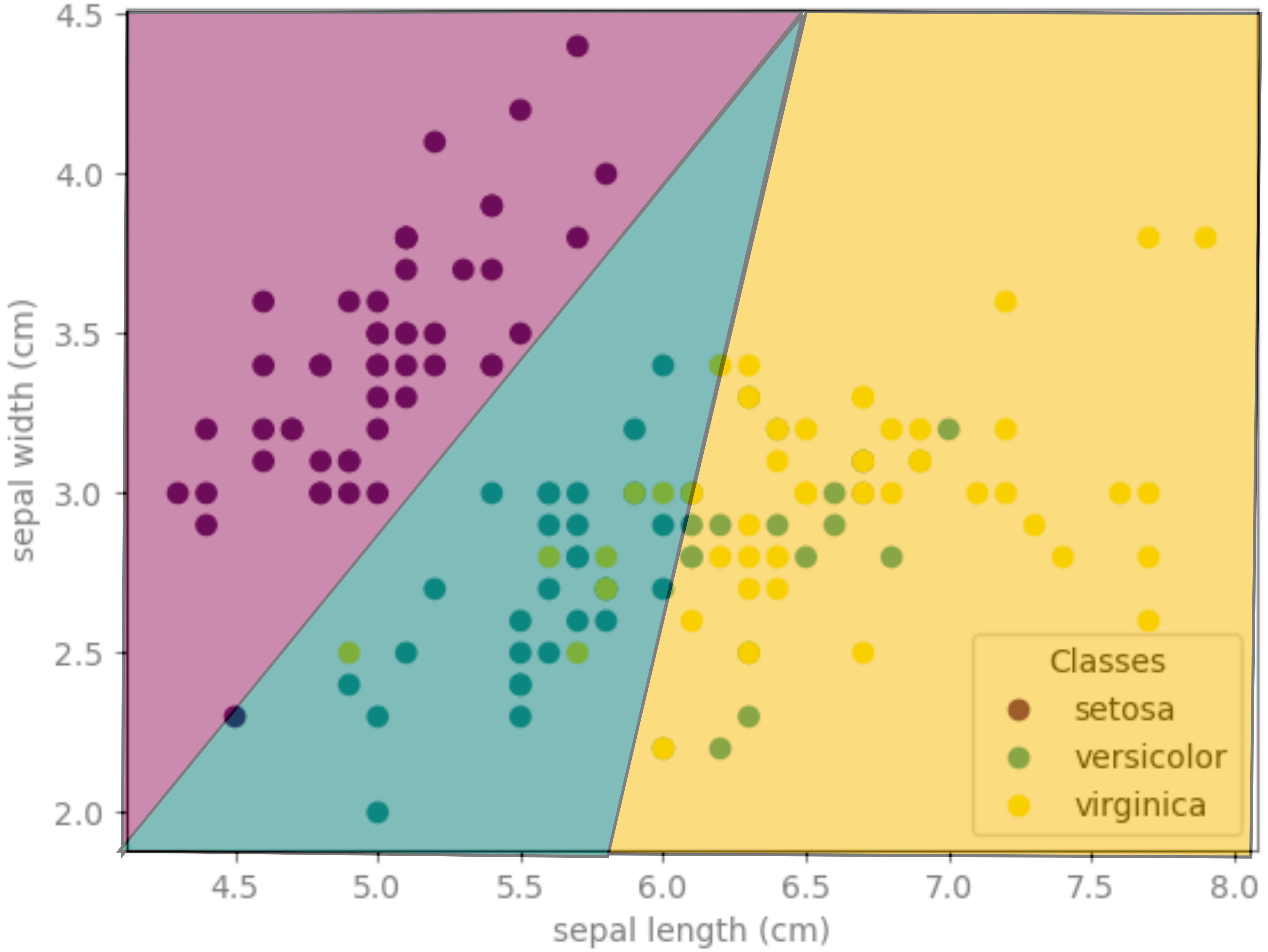
Application: Multi-Class Classification





	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
			...		
50	6.4	3.5	4.5	1.2	Versicolor
			...		
150	5.9	3.0	5.0	1.8	Virginica

150 data points, 3 classes
4 features



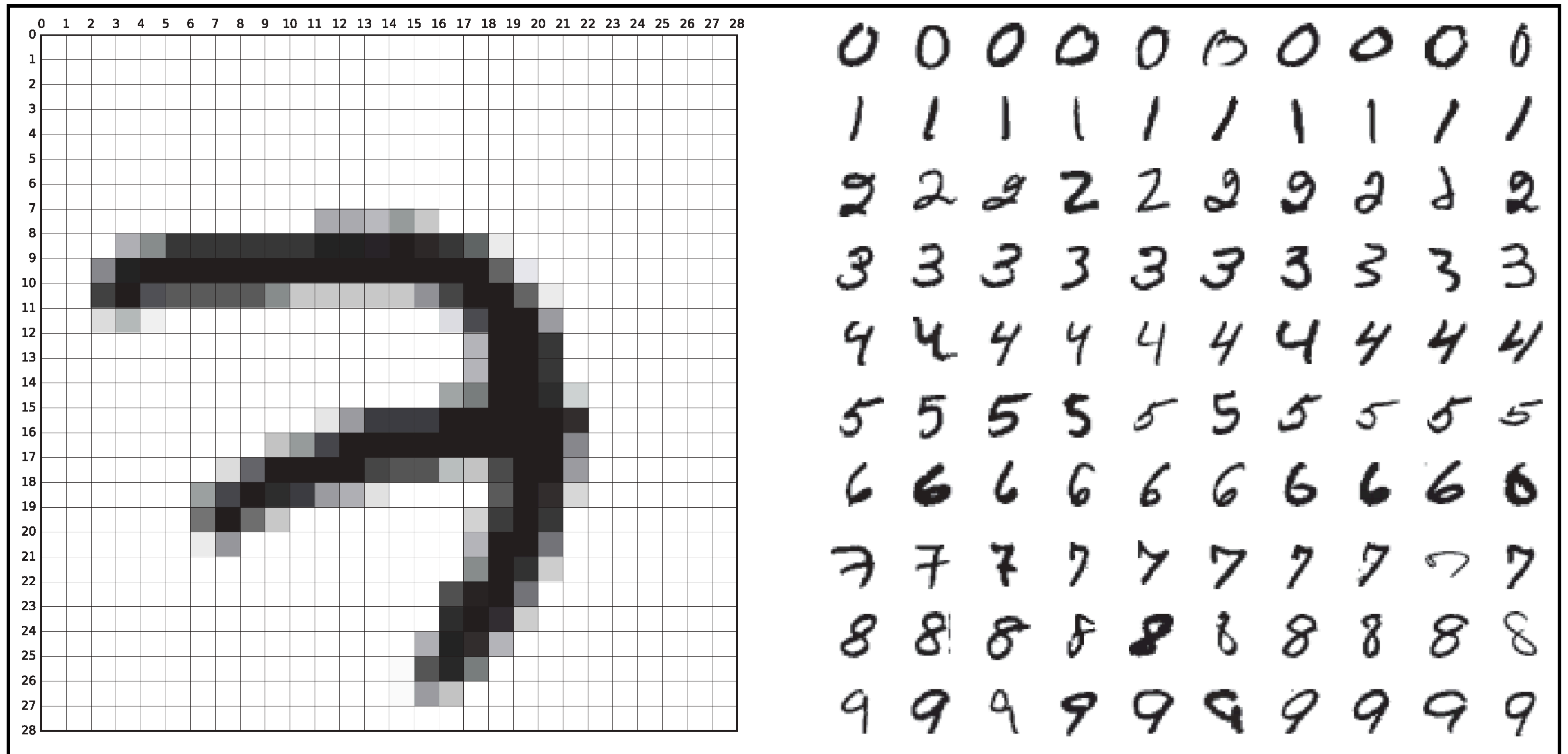
Each line can be represented as

$$w_1 \cdot (\text{sepal length}) + w_2 \cdot (\text{sepal width}) + b = 0$$

Machine Learning algorithm **learns** w_1, w_2, b from the data

Application: Digit Classification (OCR)

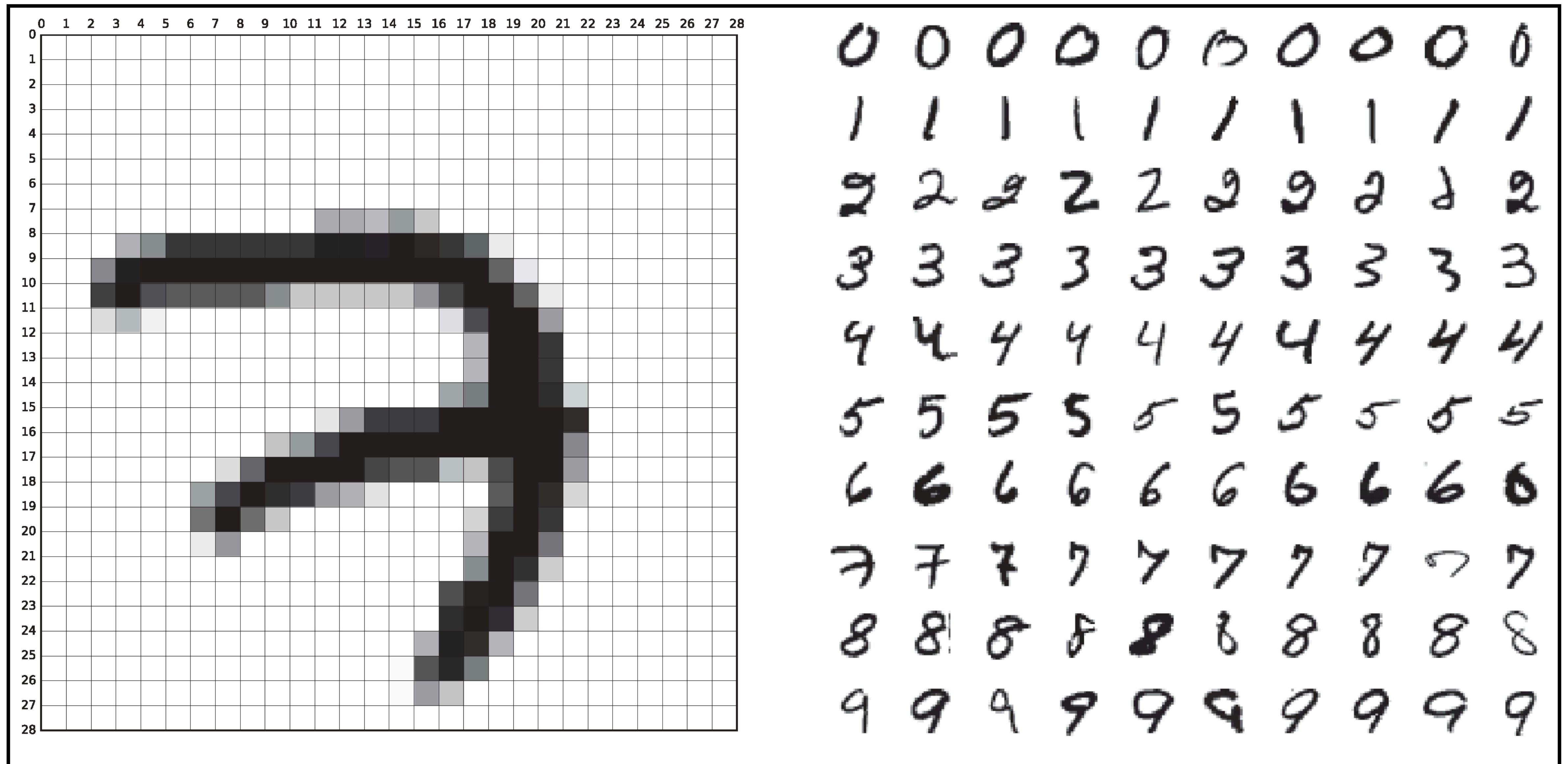
Application: Digit Classification (OCR)



(**a**) MNIST sample belonging to the digit ‘7’. (**b**) 100 samples from the MNIST training set.

MNIST is a digit dataset:
50,000 images,
28 x 28 features (728) per image

Application: Digit Classification (OCR)


(a) MNIST sample belonging to the digit '7'. (b) 100 samples from the MNIST training set.

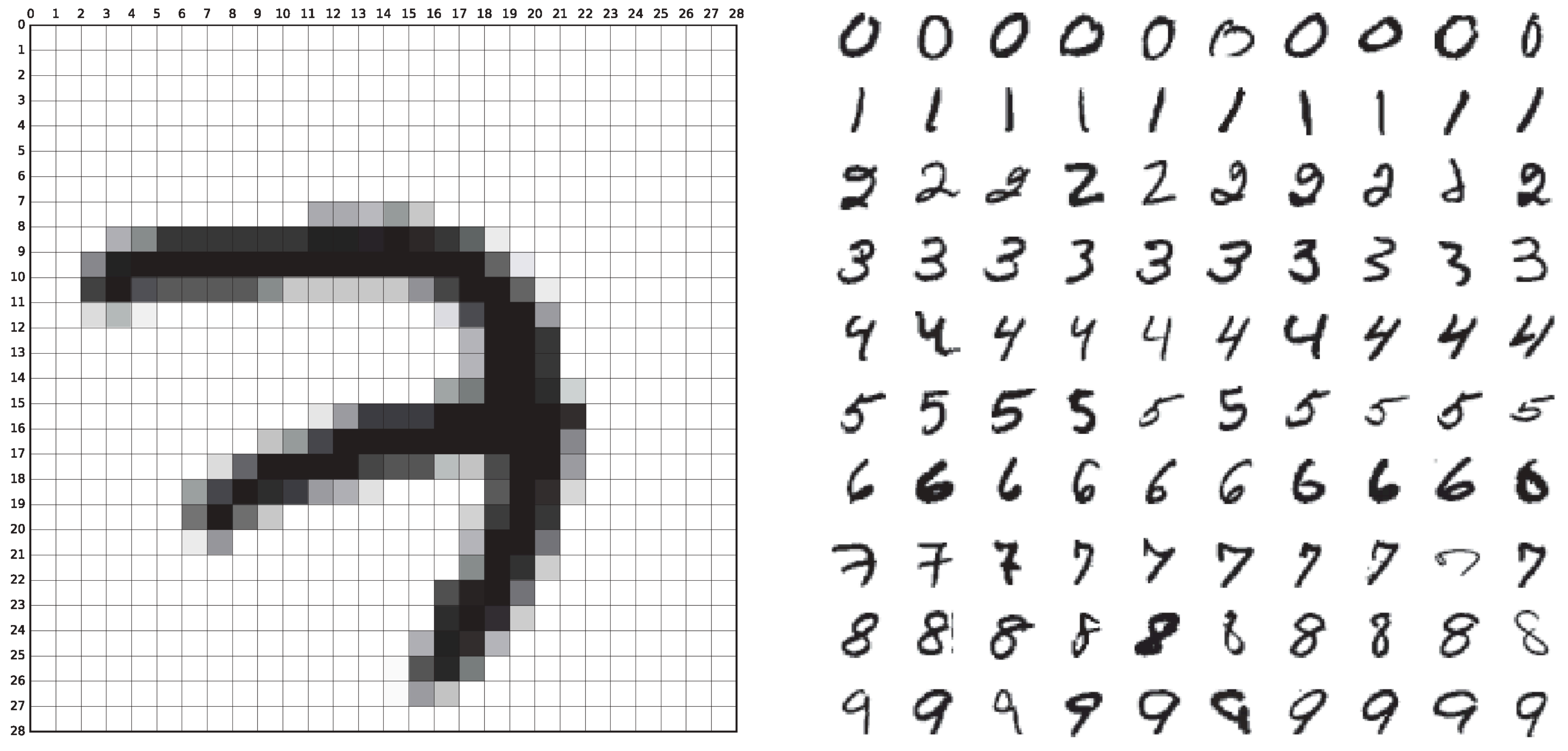
Linear classifiers i.e.

$$w_1x_1 + w_2x_2 + \ldots + w_{728}x_{728} + b = 0$$

are not sufficient here.

MNIST is a digit dataset:
50,000 images,
28 x 28 features (728) per image

Application: Digit Classification (OCR)



(**a**) MNIST sample belonging to the digit '7'. (**b**) 100 samples from the MNIST training set.

Linear classifiers i.e.

$$w_1 x_1 + w_2 x_2 + \ldots + w_{728} x_{728} + b = 0$$

are not sufficient here.

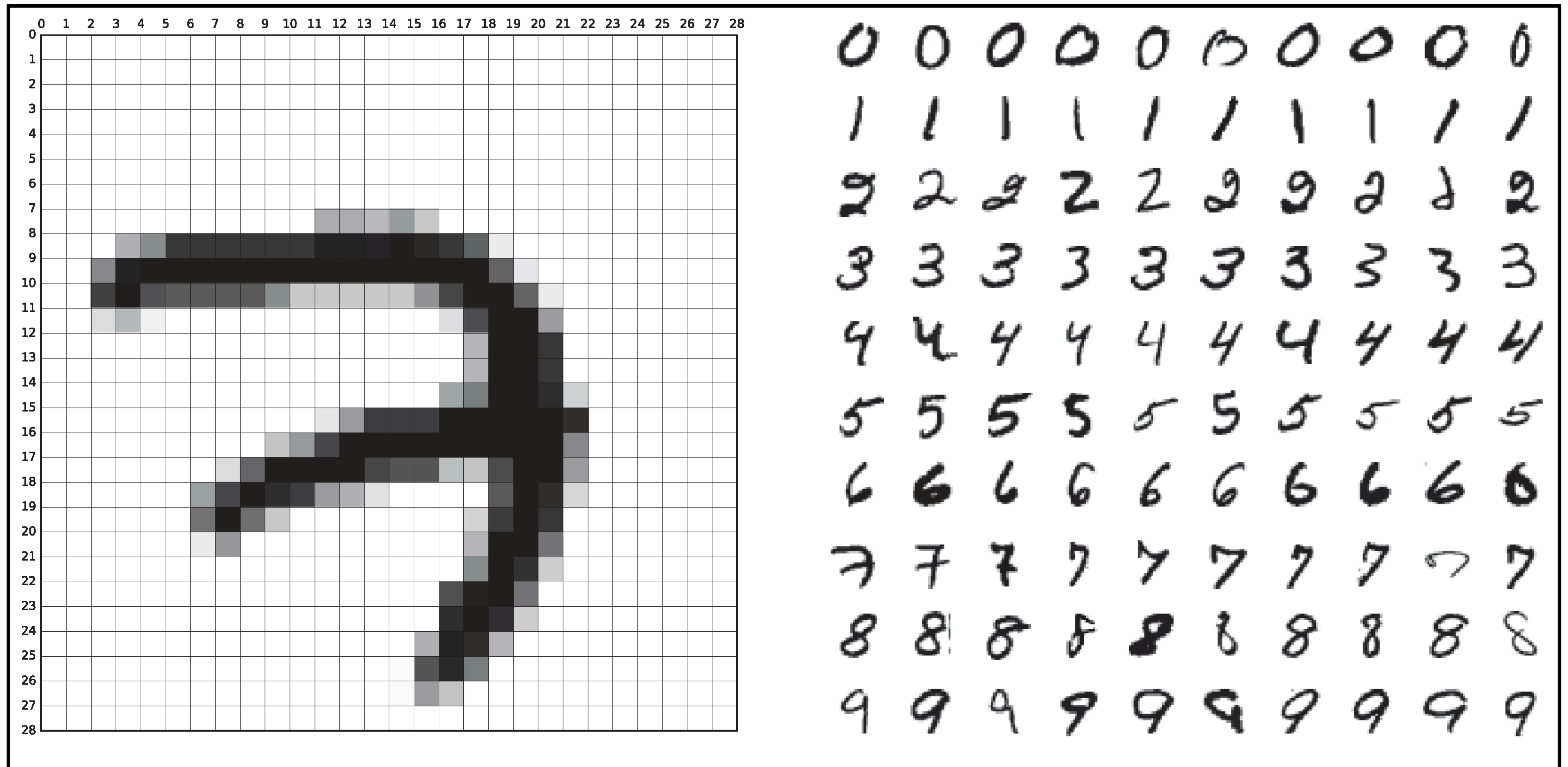
We use Artificial Neural Networks

MNIST is a digit dataset:

50,000 images,

28 x 28 features (728) per image

Application: Digit Classification (OCR)



(**a**) MNIST sample belonging to the digit ‘7’. (**b**) 100 samples from the MNIST training set.

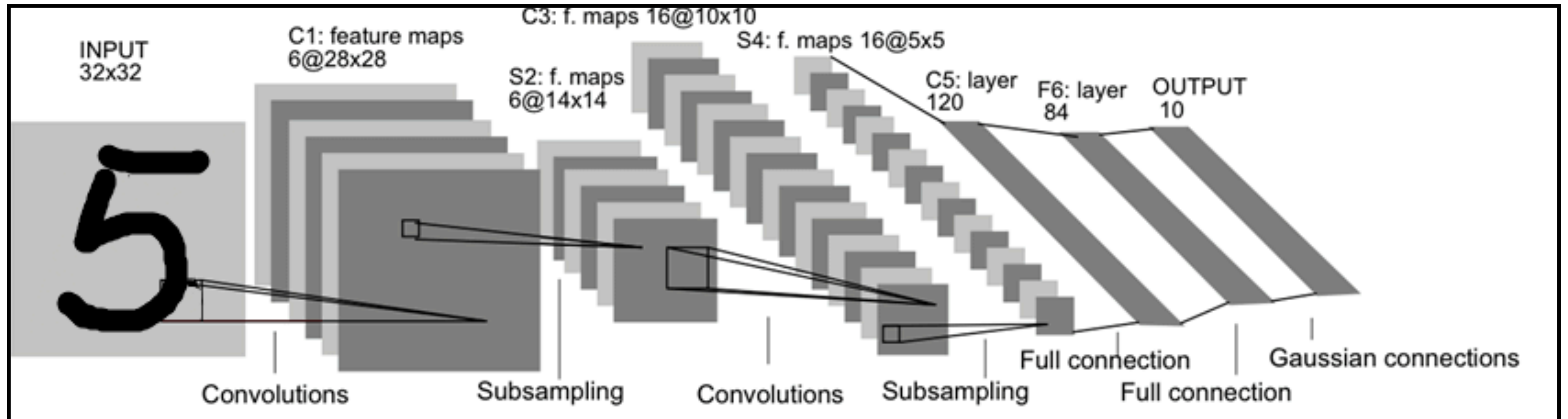
Linear classifiers i.e.

$$w_1x_1 + w_2x_2 + \ldots + w_{728}x_{728} + b = 0$$

are not sufficient here.

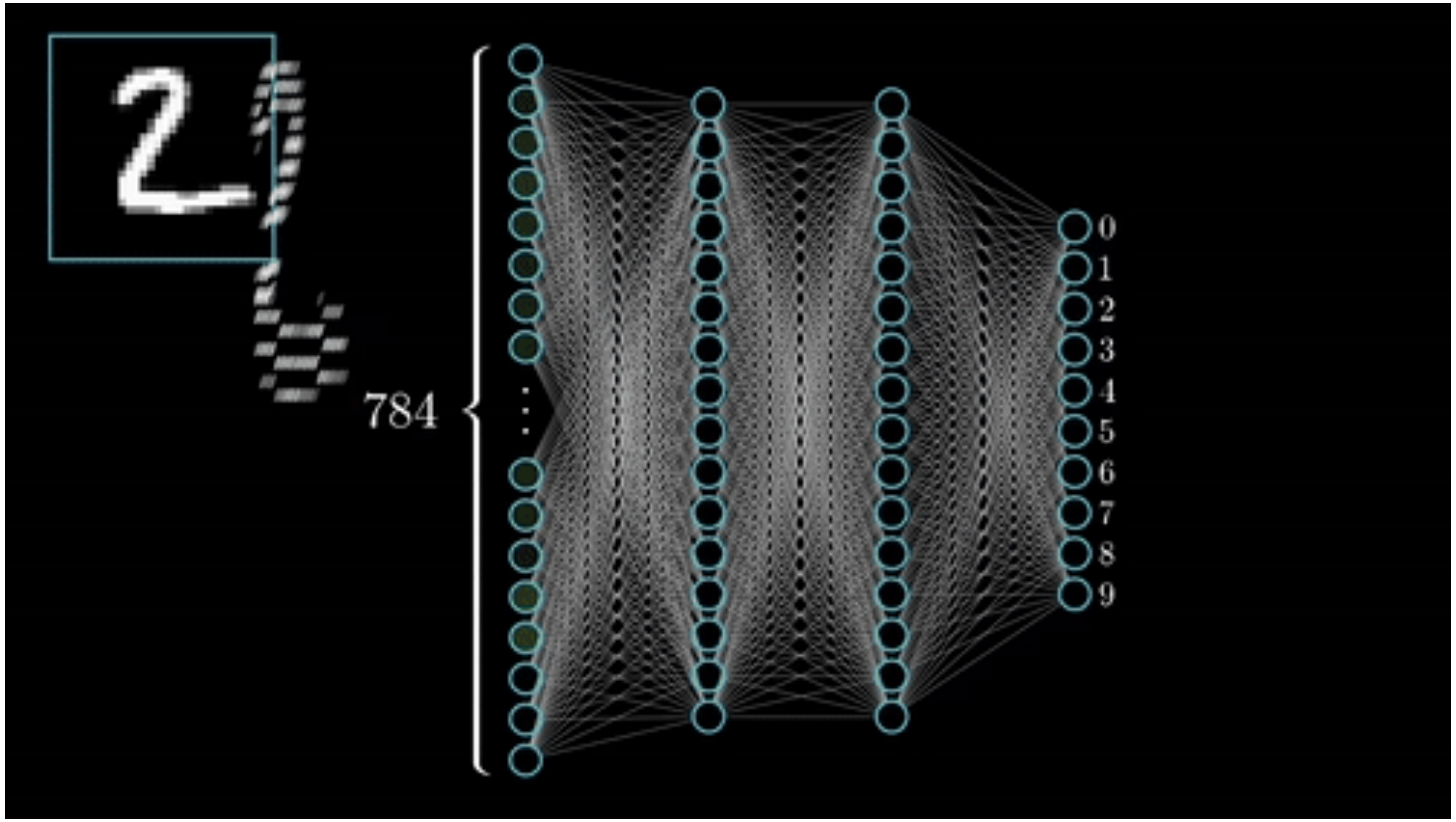
We use Artificial Neural Networks

MNIST is a digit dataset:
50,000 images,
28 x 28 features (728) per image

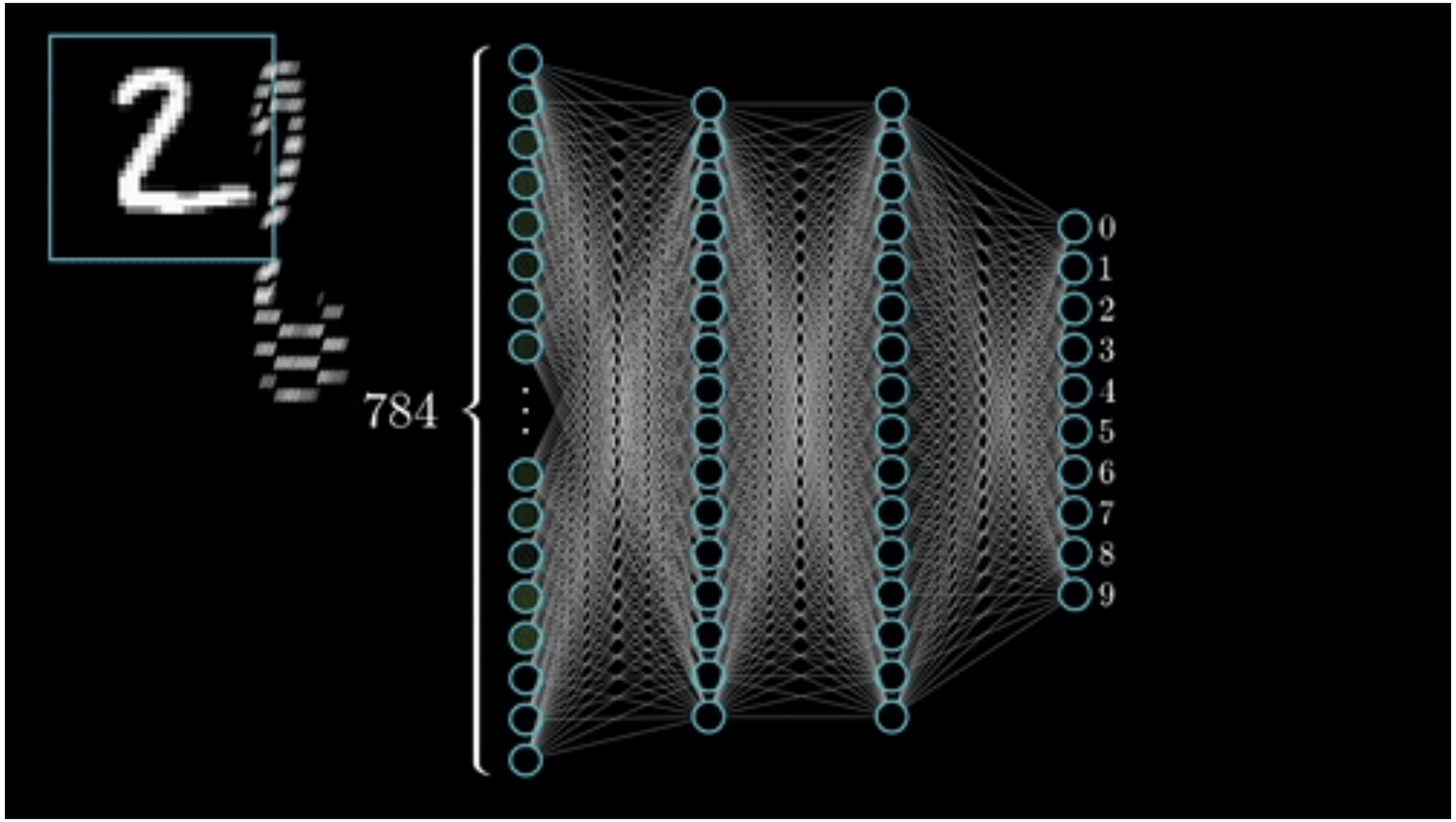


An early (Le-Net5) Convolutional Neural Network design, LeNet-5, used for recognition of digits

Neural Networks: Backbone of modern ML

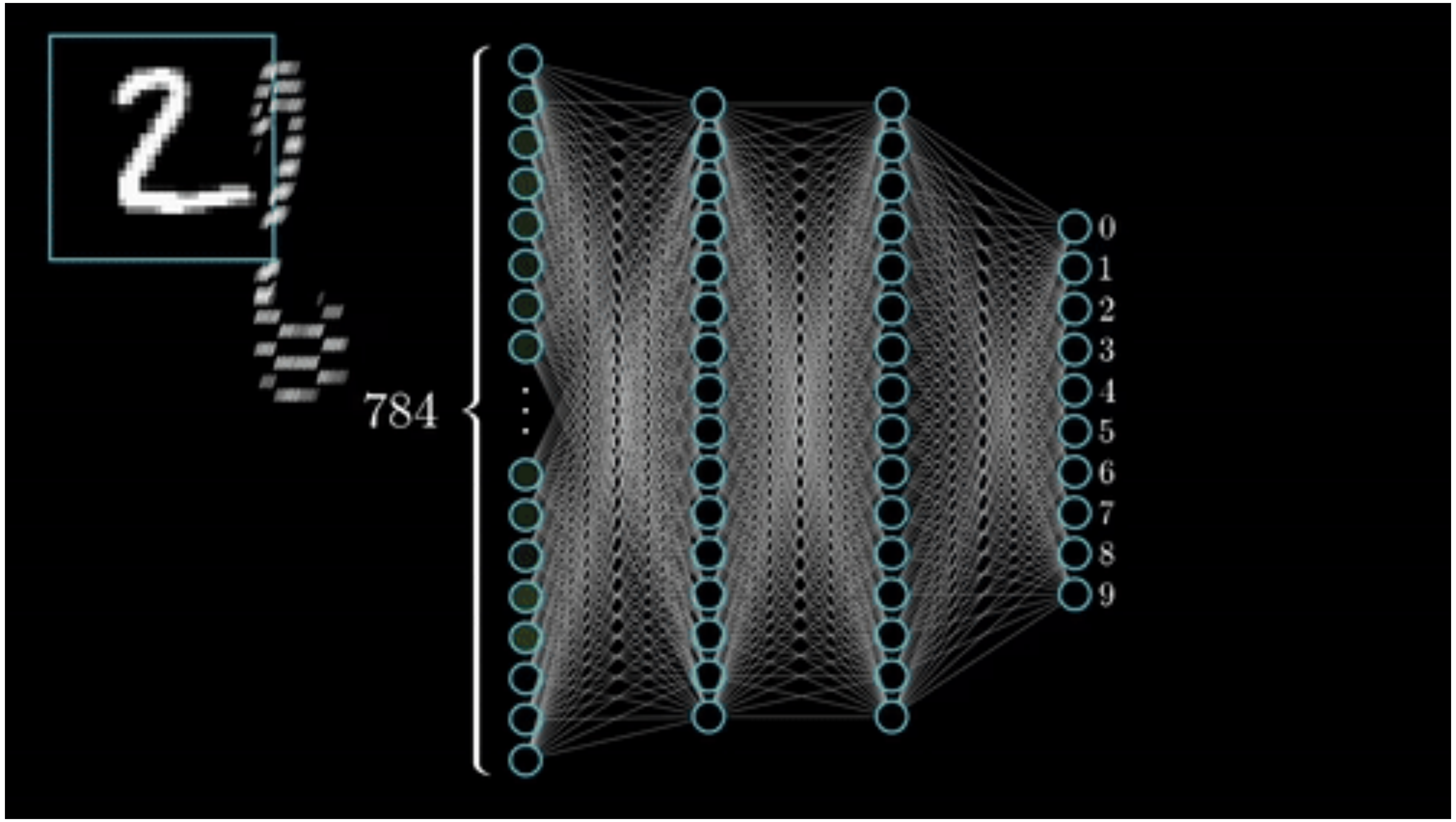


Neural Networks: Backbone of modern ML



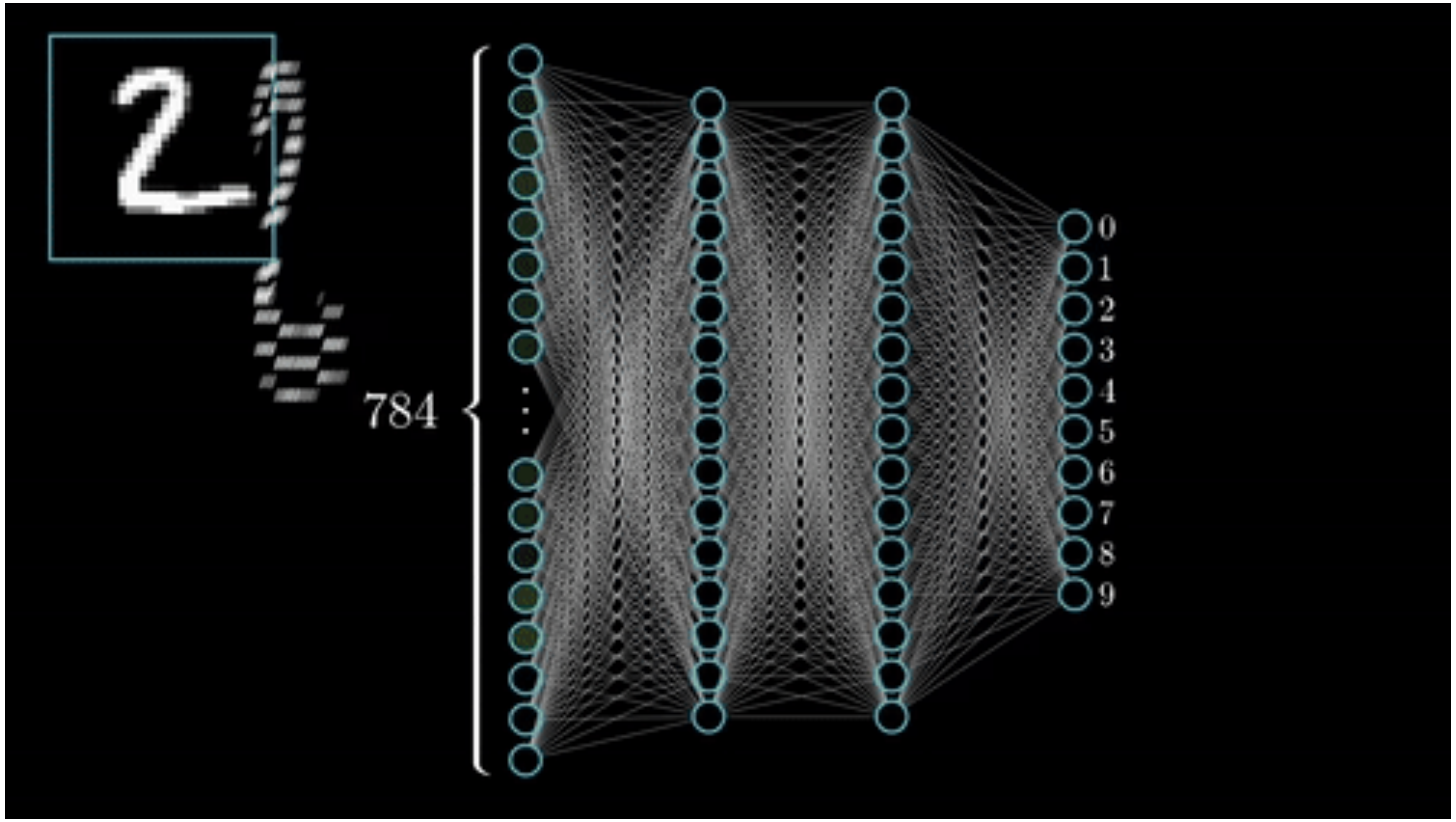
Neural Networks: Backbone of modern ML

Artificial Neural Networks are



Neural Networks: Backbone of modern ML

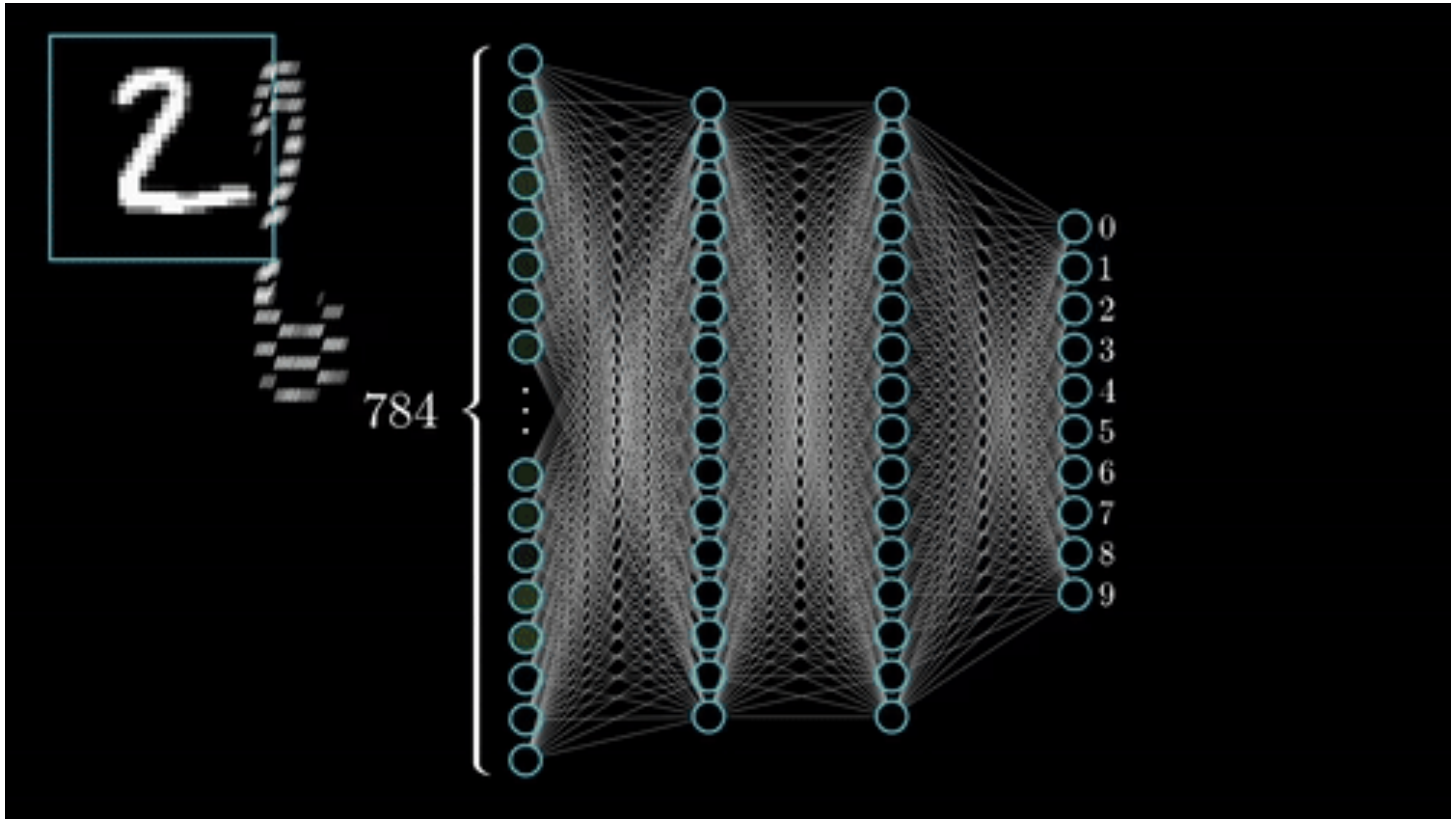
Artificial Neural Networks are
highly expressive



Neural Networks: Backbone of modern ML

Artificial Neural Networks are
highly expressive

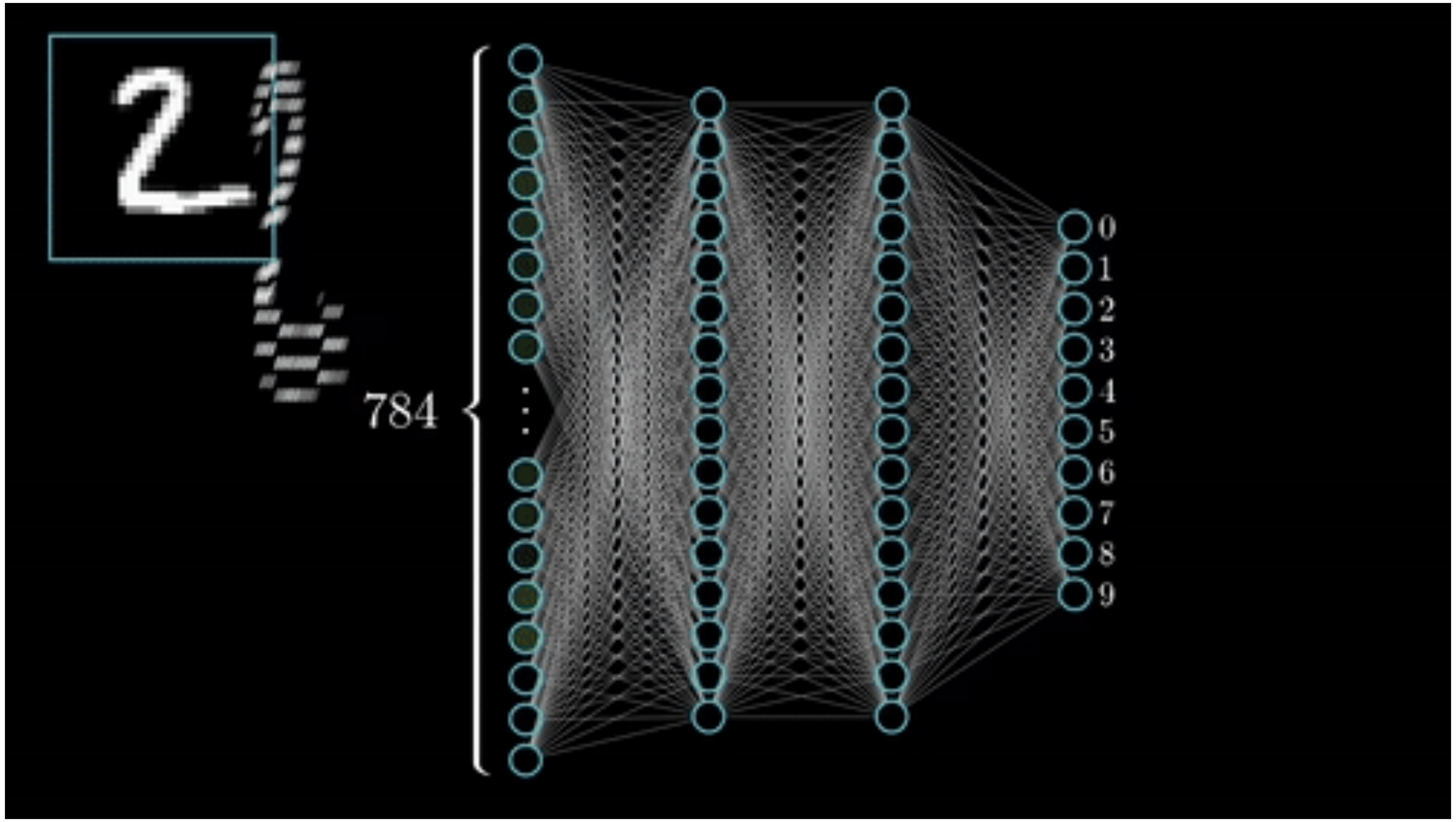
But they have



Neural Networks: Backbone of modern ML

Artificial Neural Networks are
highly expressive

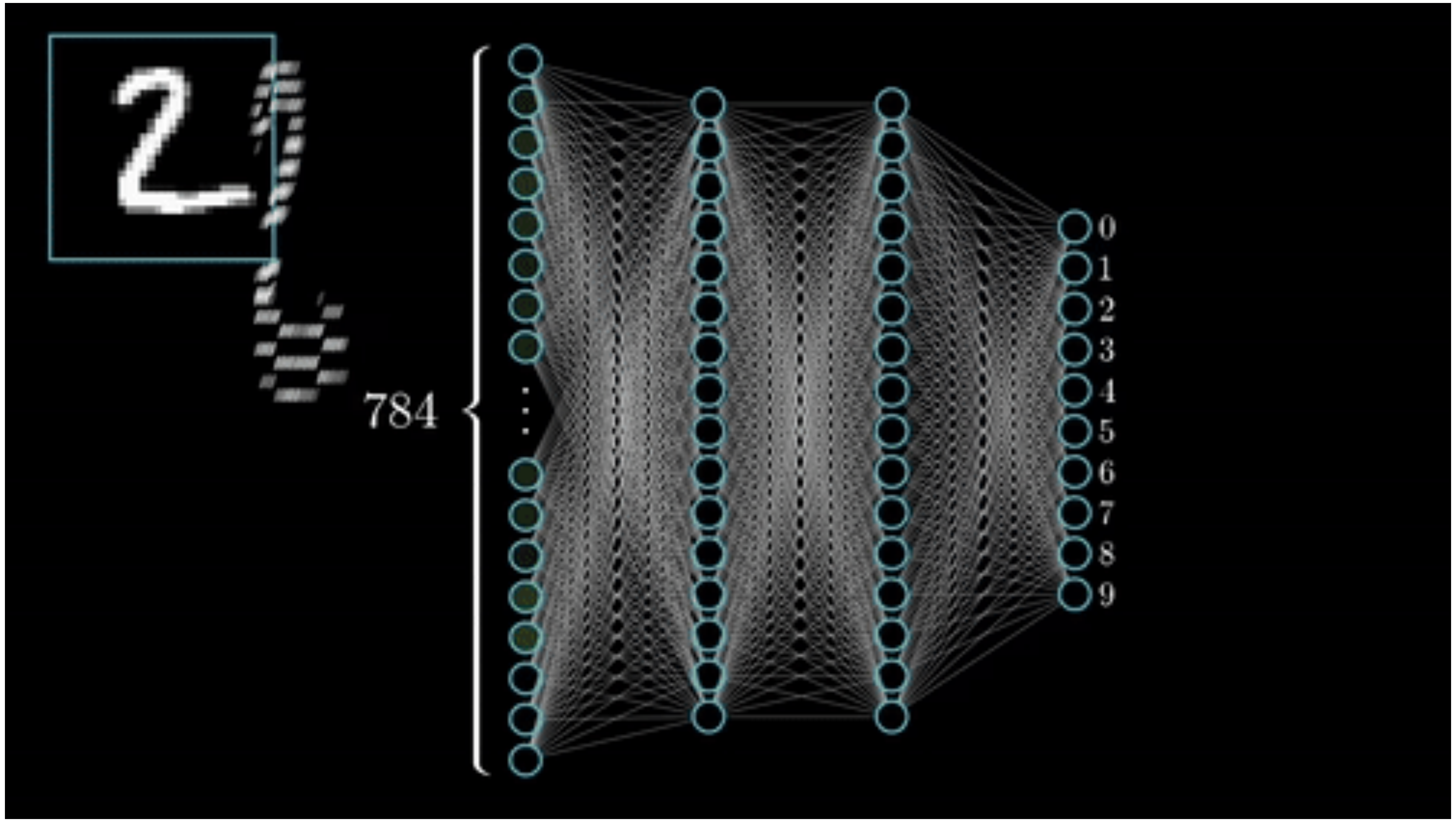
But they have
Large number of parameters



Neural Networks: Backbone of modern ML

Artificial Neural Networks are
highly expressive

But they have
Large number of parameters

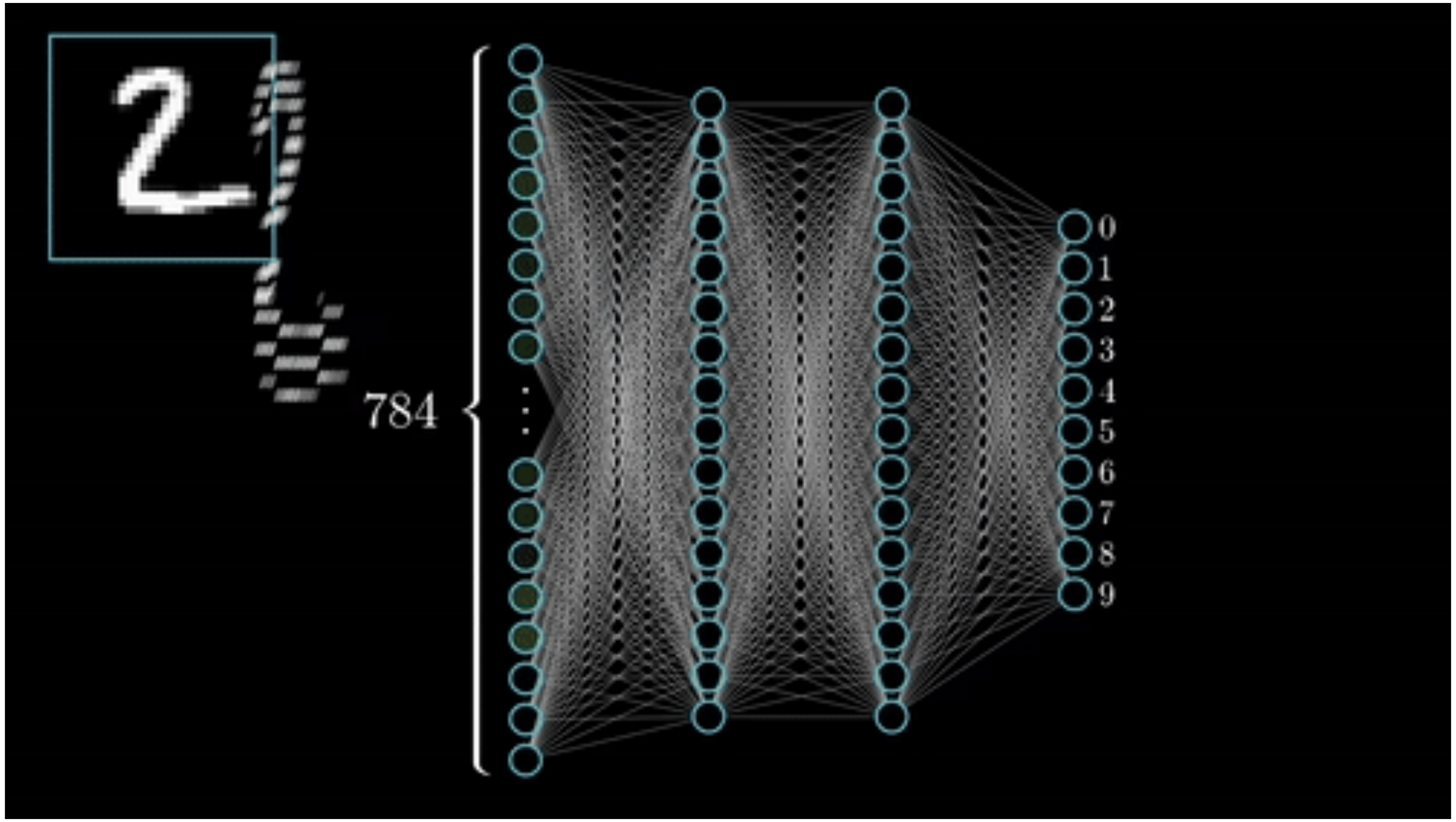


Rumors claim that GPT-4 has 1.76 trillion parameters, which was first estimated by the speed it was running and by George Hotz.[12]

Neural Networks: Backbone of modern ML

Artificial Neural Networks are
highly expressive

But they have
Large number of parameters



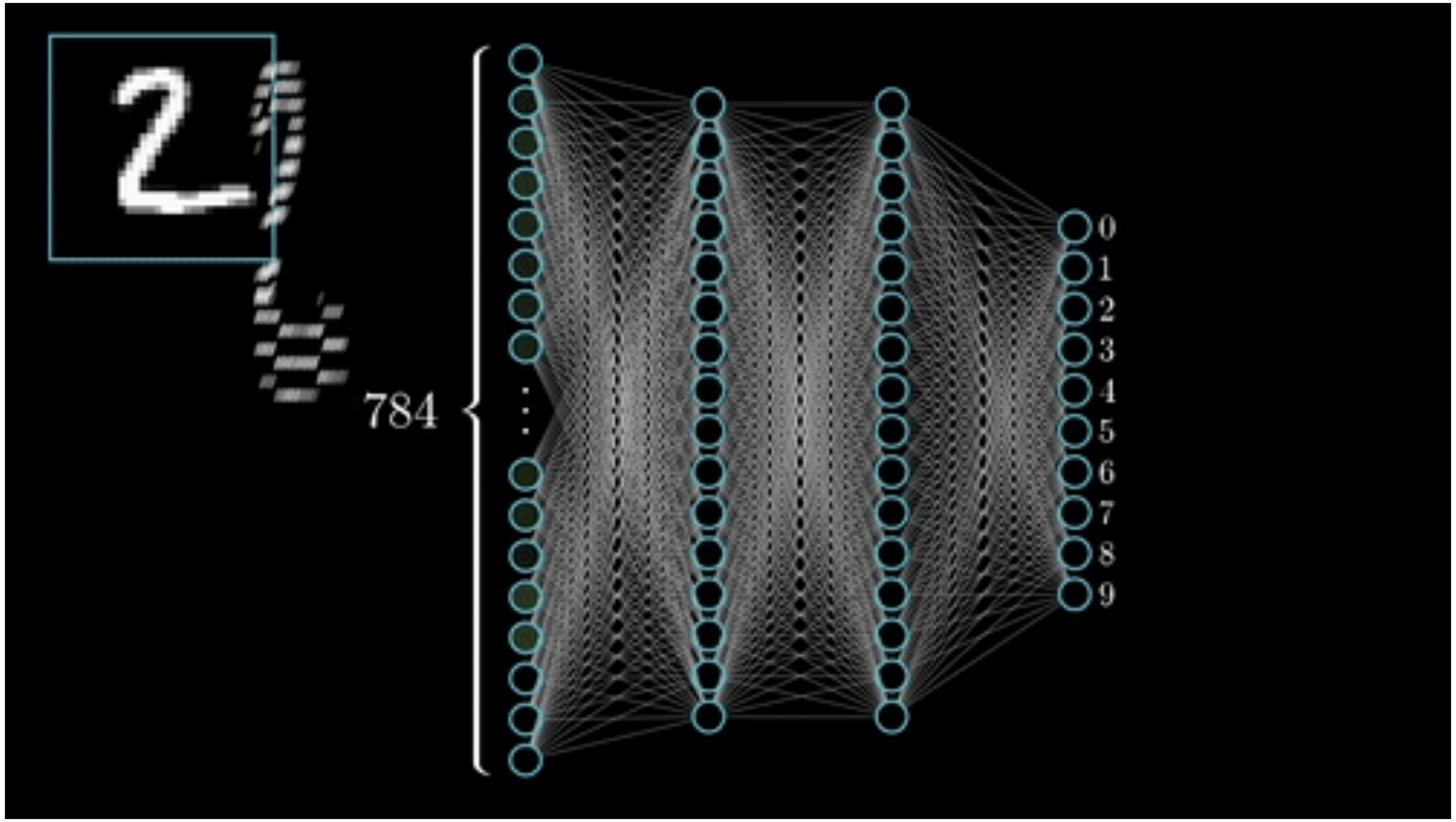
Rumors claim that GPT-4 has 1.76 trillion parameters, which was first estimated by the speed it was running and by George Hotz.[12]

Neural Networks: Backbone of modern ML

Artificial Neural Networks are
highly expressive

But they have
Large number of parameters

Each edge is a parameter.



Rumors claim that GPT-4 has 1.76 trillion parameters, which was first estimated by the speed it was running and by George Hotz.[12]

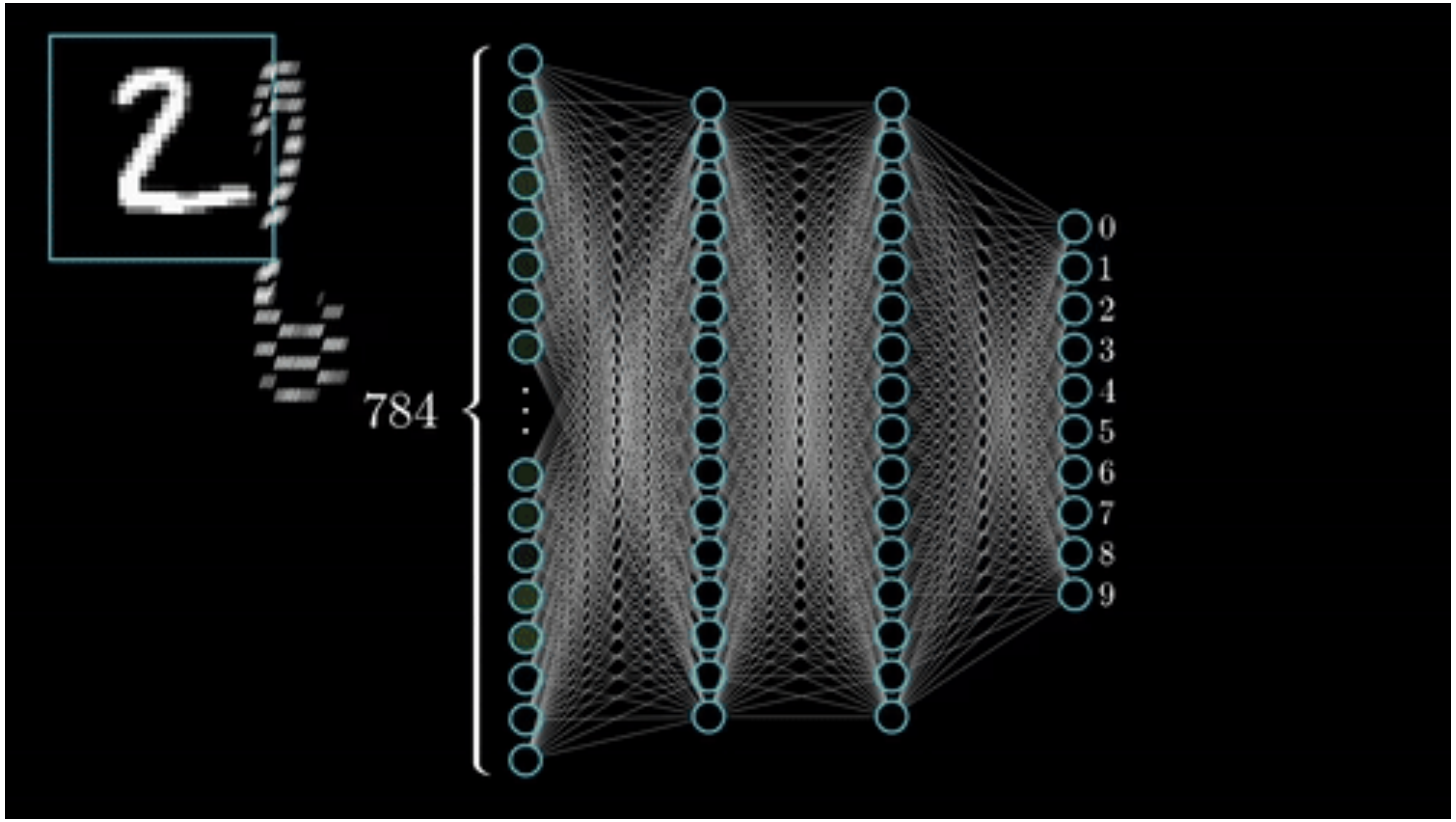
Neural Networks: Backbone of modern ML

Artificial Neural Networks are
highly expressive

But they have
Large number of parameters

Each edge is a parameter.

Parameters are learned by an algorithm called **Backpropagation.**



Rumors claim that GPT-4 has 1.76 trillion parameters, which was first estimated by the speed it was running and by George Hotz.[12]

Neural Networks: Backbone of modern ML

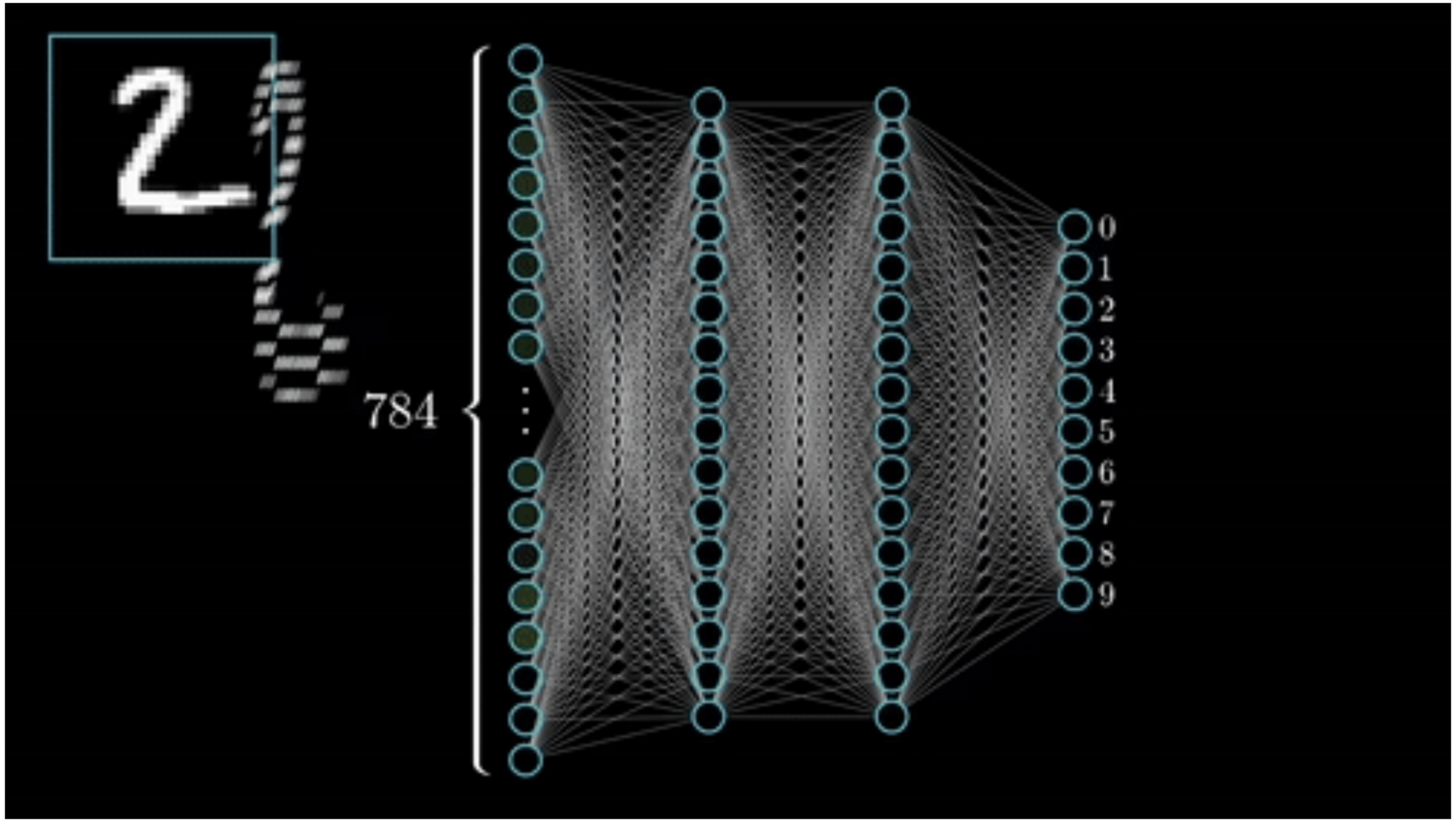
Artificial Neural Networks are
highly expressive

But they have
Large number of parameters

Each edge is a parameter.

Parameters are learned by an algorithm called **Backpropagation.**

Backpropagation *<u>propagates the discrepancy between the current output and the desired output</u>* to all parameters.



Rumors claim that GPT-4 has 1.76 trillion parameters, which was first estimated by the speed it was running and by George Hotz.[12]

Neural Networks: Backbone of modern ML

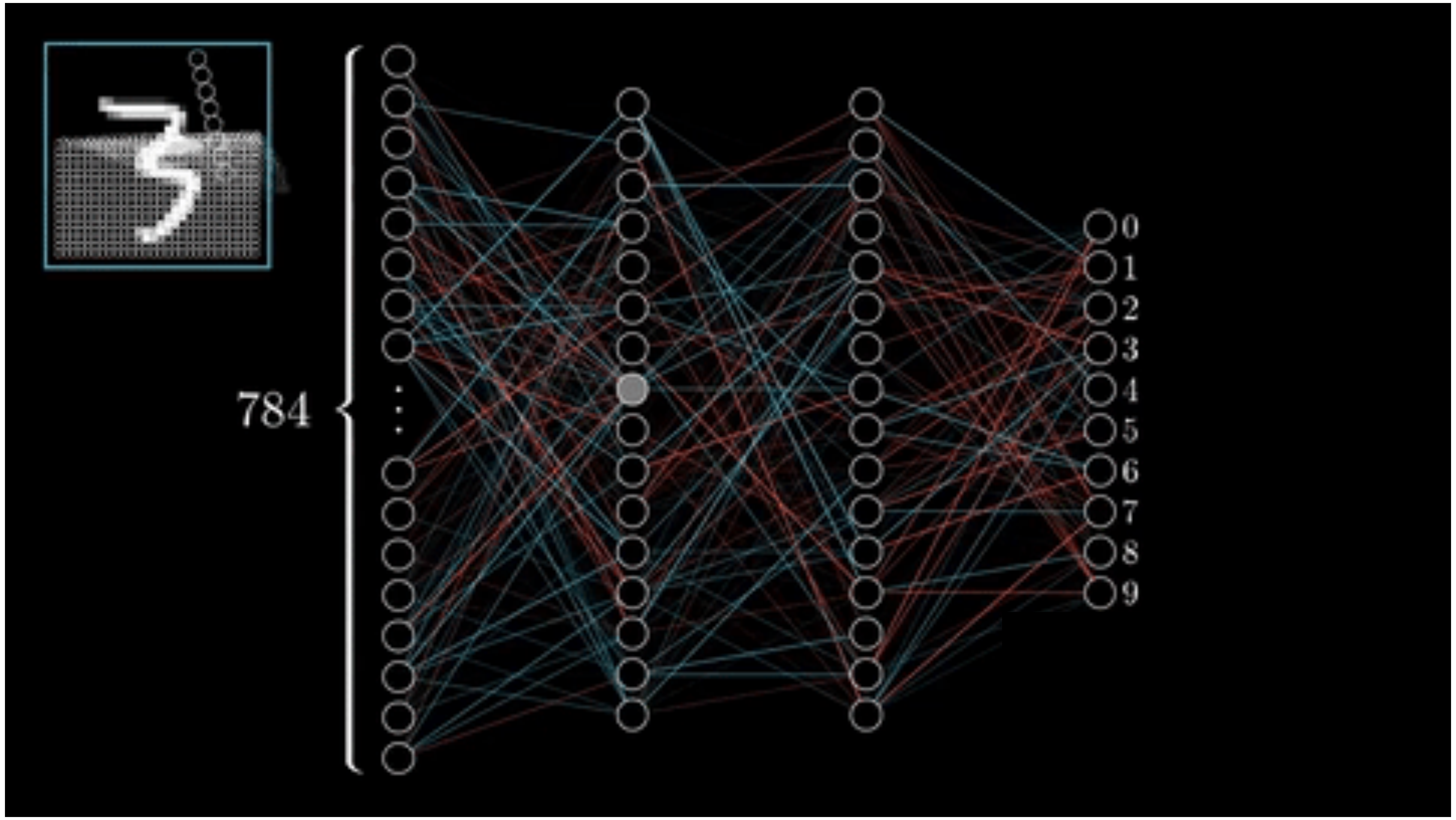
Artificial Neural Networks are
highly expressive

But they have
Large number of parameters

Each edge is a parameter.

Parameters are learned by an algorithm called **Backpropagation.**

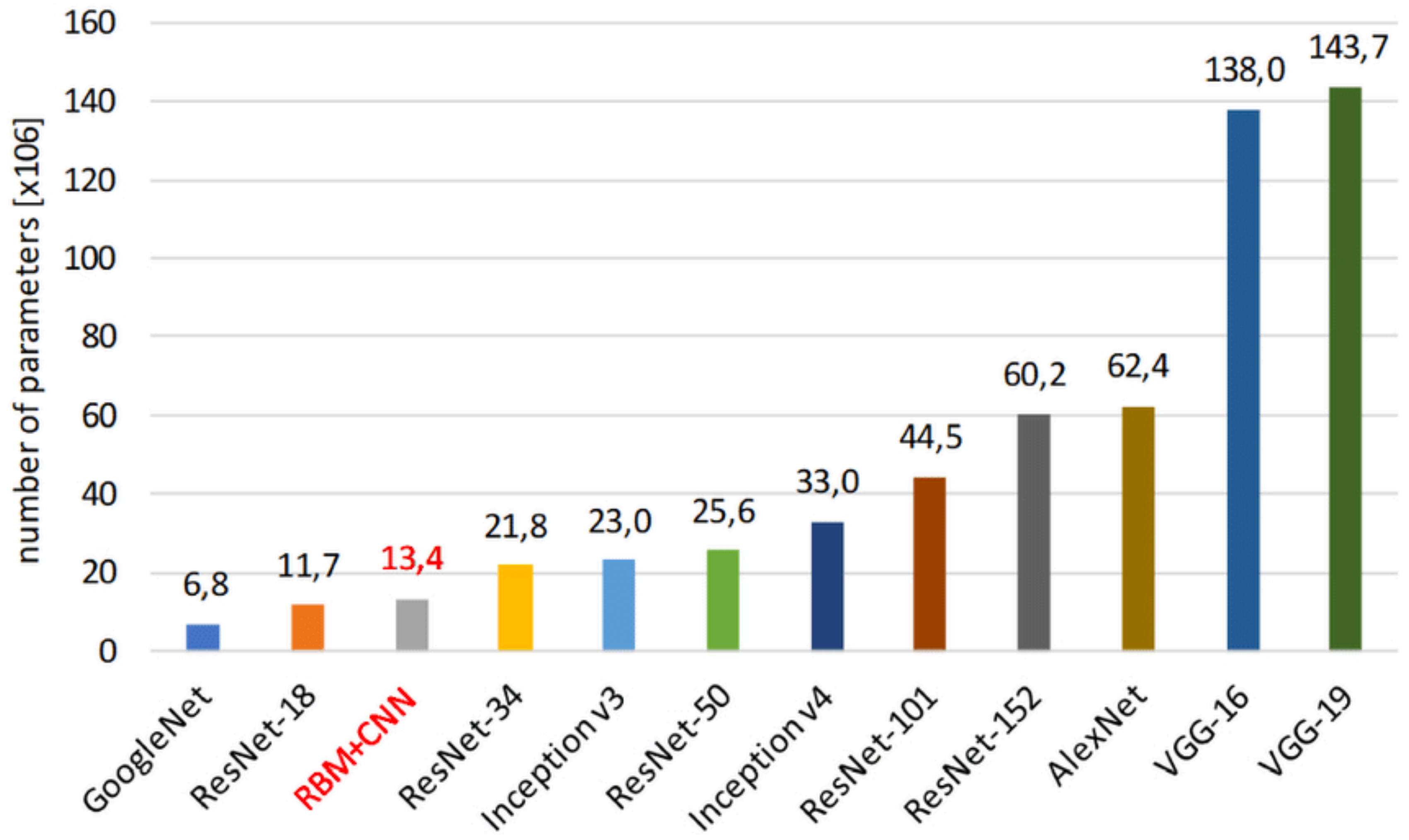
Backpropagation *propagates the discrepancy between the current output and the desired output* to all parameters.



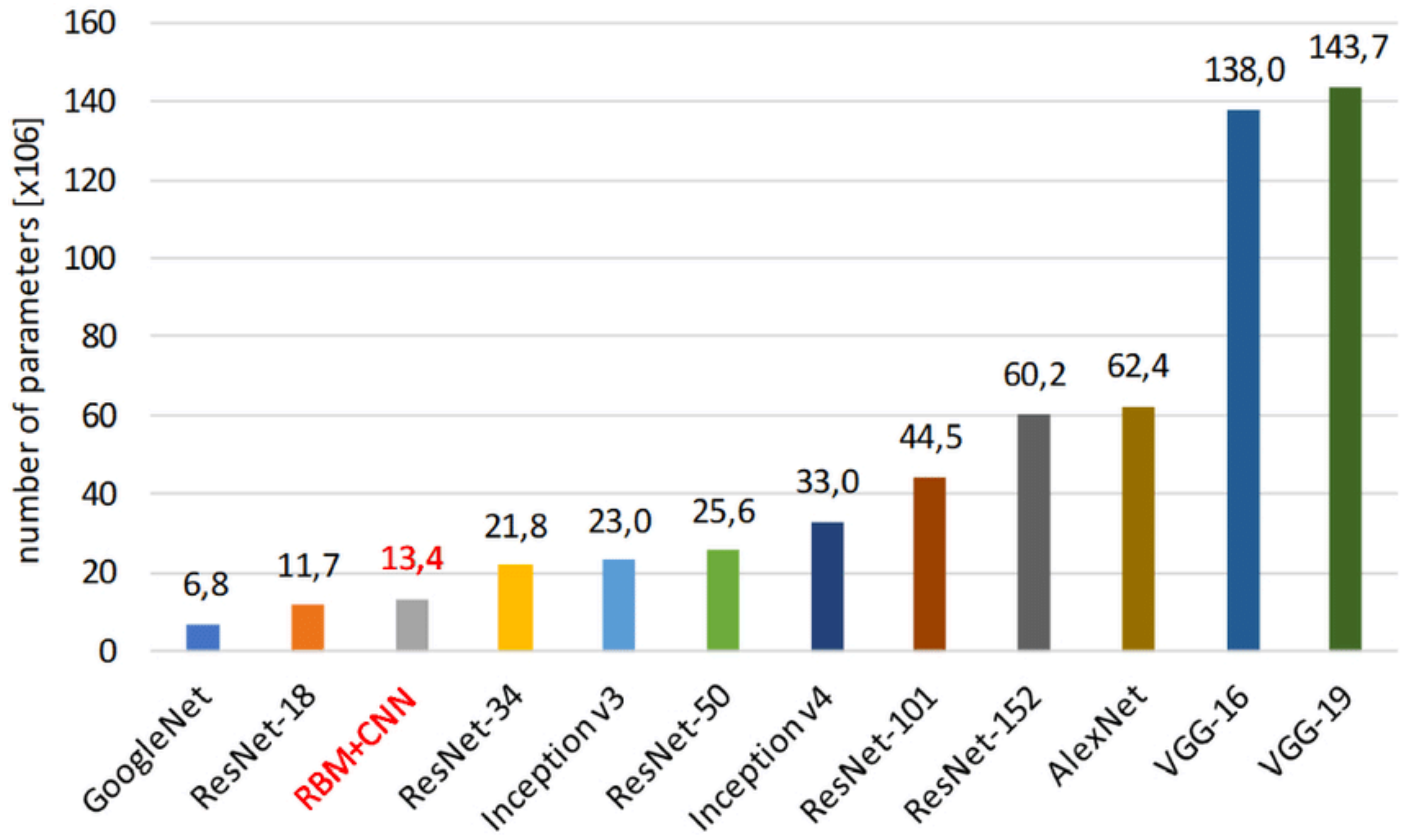
Rumors claim that GPT-4 has 1.76 trillion parameters, which was first estimated by the speed it was running and by George Hotz.[12]

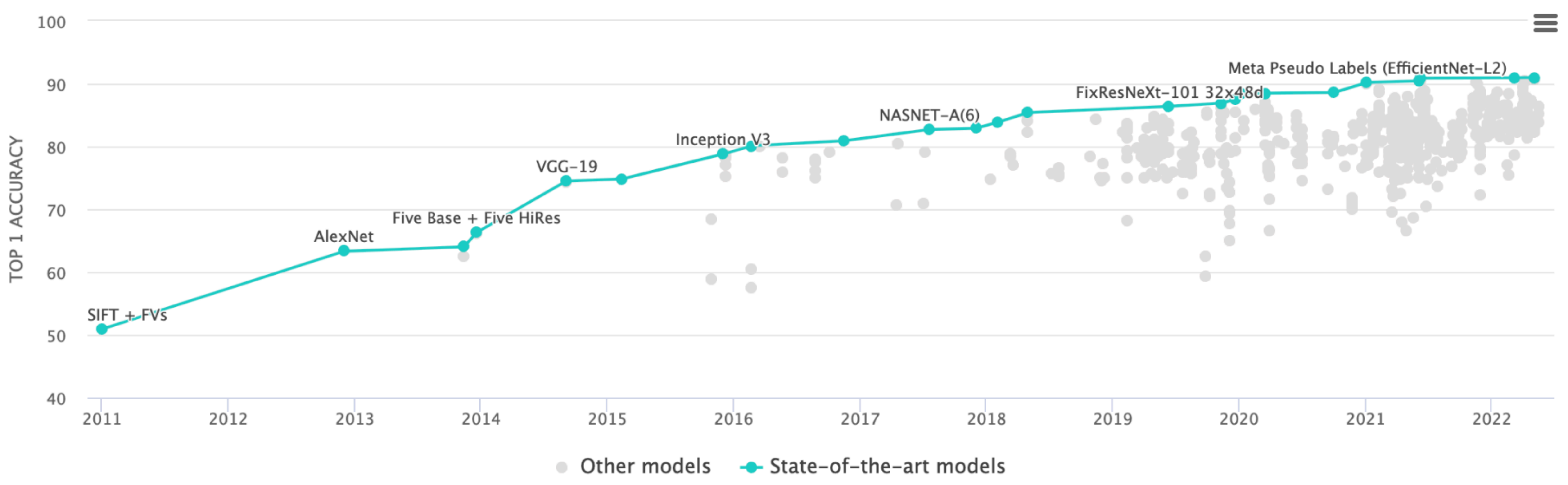
Growth of Computer Vision models

Growth of Computer Vision models



Growth of Computer Vision models





Application: Language Modelling

Application: Language Modelling

In Computer vision, the task is to **predict the label** given a image.

Application: Language Modelling

In Computer vision, the task is to **predict the label** given a image.

Given a dataset of **images and corresponding labels**, ML algorithm maximises
the probability of **predicting the correct label**.

Application: Language Modelling

In Computer vision, the task is to **predict the label** given a image.

Given a dataset of **images and corresponding labels**, ML algorithm maximises the probability of **predicting the correct label**.

In Language modelling, task is to **predict the next token given previous tokens**.
A token can be a character, a word or something in between.

Application: Language Modelling

In Computer vision, the task is to **predict the label** given a image.

Given a dataset of **images and corresponding labels**, ML algorithm maximises
the probability of **predicting the correct label**.

In Language modelling, task is to **predict the next token given previous tokens**.
A token can be a character, a word or something in between.

Types of Language Modelling:

Application: Language Modelling

In Computer vision, the task is to **predict the label** given a image.

Given a dataset of **images and corresponding labels**, ML algorithm maximises the probability of **predicting the correct label**.

In Language modelling, task is to **predict the next token given previous tokens**.
A token can be a character, a word or something in between.

Types of Language Modelling:

- **Statistical Language Modelling**

Application: Language Modelling

In Computer vision, the task is to **predict the label** given a image.

Given a dataset of **images and corresponding labels**, ML algorithm maximises the probability of **predicting the correct label**.

In Language modelling, task is to **predict the next token given previous tokens**.
A token can be a character, a word or something in between.

Types of Language Modelling:

- **Statistical Language Modelling**
- **Neural Language Modelling**

Statistical Language Modelling

Statistical Language Modelling

4-gram modelling (generally n-gram)

Statistical Language Modelling

4-gram modelling (generally n-gram)

From a dataset of documents

- **Compute the joint probability** of *all* token sequences (length 4 and 3).

Statistical Language Modelling

4-gram modelling (generally n-gram)

From a dataset of documents

- **Compute the joint probability** of *all* token sequences (length 4 and 3).

We were all feeling seedy, and we were getting quite nervous about it. Harris said he felt such extraordinary fits of giddiness come over him at times, that he hardly knew what he was doing; and then George said that *he* had fits of giddiness too, and hardly knew what *he* was doing. With me, it was my liver that was out of order. I knew it was my liver that was out of order, because I had just been reading a patent liver-pill circular, in which were detailed the various symptoms by which a man could tell when his liver was out of order. I had them all.

Statistical Language Modelling

4-gram modelling (generally n-gram)

From a dataset of documents

- **Compute the joint probability** of *all* token sequences (length 4 and 3).

4-gram	Frequency
was out of order	3
my liver that was	2
it was my liver	2
that was out of	2
liver that was out	2
was my liver that	2

We were all feeling seedy, and we were getting quite nervous about it. Harris said he felt such extraordinary fits of giddiness come over him at times, that he hardly knew what he was doing; and then George said that *he* had fits of giddiness too, and hardly knew what *he* was doing. With me, it was my liver that was out of order. I knew it was my liver that was out of order, because I had just been reading a patent liver-pill circular, in which were detailed the various symptoms by which a man could tell when his liver was out of order. I had them all.

Statistical Language Modelling

4-gram modelling (generally n-gram)

From a dataset of documents

- **Compute the joint probability** of *all* token sequences (length 4 and 3).
- User inputs the first 3 words w_1, w_2, w_3

4-gram	Frequency
was out of order	3
my liver that was	2
it was my liver	2
that was out of	2
liver that was out	2
was my liver that	2

We were all feeling seedy, and we were getting quite nervous about it. Harris said he felt such extraordinary fits of giddiness come over him at times, that he hardly knew what he was doing; and then George said that *he* had fits of giddiness too, and hardly knew what *he* was doing. With me, it was my liver that was out of order. I knew it was my liver that was out of order, because I had just been reading a patent liver-pill circular, in which were detailed the various symptoms by which a man could tell when his liver was out of order. I had them all.

Statistical Language Modelling

4-gram modelling (generally n-gram)

From a dataset of documents

- **Compute the joint probability** of *all* token sequences (length 4 and 3).
- User inputs the first 3 words w_1, w_2, w_3
- Then given w_1, w_2, w_3, **sample** w_4 **from the conditional distribution** $p(w \mid w_1, w_2, w_3)$.

4-gram	**Frequency**
was out of order	3
my liver that was	2
it was my liver	2
that was out of	2
liver that was out	2
was my liver that	2

We were all feeling seedy, and we were getting quite nervous about it. Harris said he felt such extraordinary fits of giddiness come over him at times, that he hardly knew what he was doing; and then George said that *he* had fits of giddiness too, and hardly knew what *he* was doing. With me, it was my liver that was out of order. I knew it was my liver that was out of order, because I had just been reading a patent liver-pill circular, in which were detailed the various symptoms by which a man could tell when his liver was out of order. I had them all.

Statistical Language Modelling

4-gram modelling (generally n-gram)

From a dataset of documents

- **Compute the joint probability** of *all* token sequences (length 4 and 3).
- User inputs the first 3 words w_1, w_2, w_3
- Then given w_1, w_2, w_3, **sample** w_4 **from the conditional distribution** $p(w \mid w_1, w_2, w_3)$.
- Then keep repeating the process.

4-gram	Frequency
was out of order	3
my liver that was	2
it was my liver	2
that was out of	2
liver that was out	2
was my liver that	2

We were all feeling seedy, and we were getting quite nervous about it. Harris said he felt such extraordinary fits of giddiness come over him at times, that he hardly knew what he was doing; and then George said that *he* had fits of giddiness too, and hardly knew what *he* was doing. With me, it was my liver that was out of order. I knew it was my liver that was out of order, because I had just been reading a patent liver-pill circular, in which were detailed the various symptoms by which a man could tell when his liver was out of order. I had them all.

Statistical Language Modelling

4-gram modelling (generally n-gram)

From a dataset of documents

- **Compute the joint probability** of *all* token sequences (length 4 and 3).
- User inputs the first 3 words w_1, w_2, w_3
- Then given w_1, w_2, w_3, **sample** w_4 **from the conditional distribution** $p(w \mid w_1, w_2, w_3)$.
- Then keep repeating the process.

4-gram	**Frequency**
was out of order	3
my liver that was	2
it was my liver	2
that was out of	2
liver that was out	2
was my liver that	2

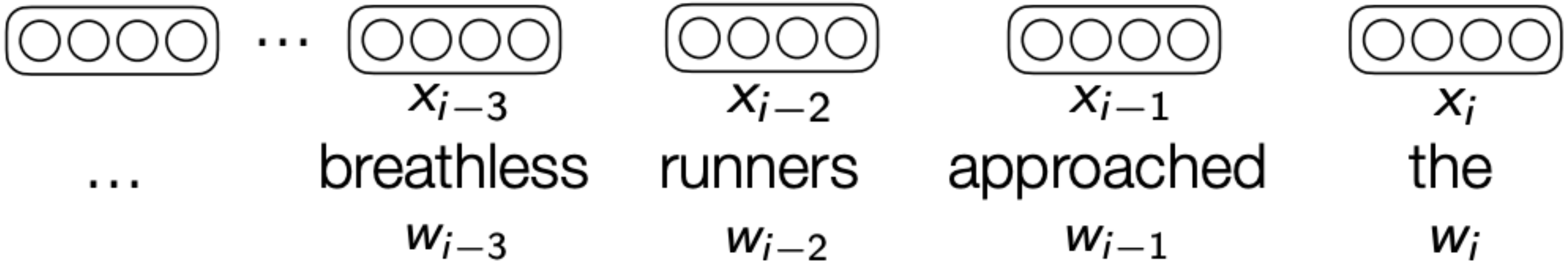
We were all feeling seedy, and we were getting quite nervous about it. Harris said he felt such extraordinary fits of giddiness come over him at times, that he hardly knew what he was doing; and then George said that *he* had fits of giddiness too, and hardly knew what *he* was doing. With me, it was my liver that was out of order. I knew it was my liver that was out of order, because I had just been reading a patent liver-pill circular, in which were detailed the various symptoms by which a man could tell when his liver was out of order. I had them all.

This is a simple process but it cannot model long-term dependencies.

Neural Language Modelling

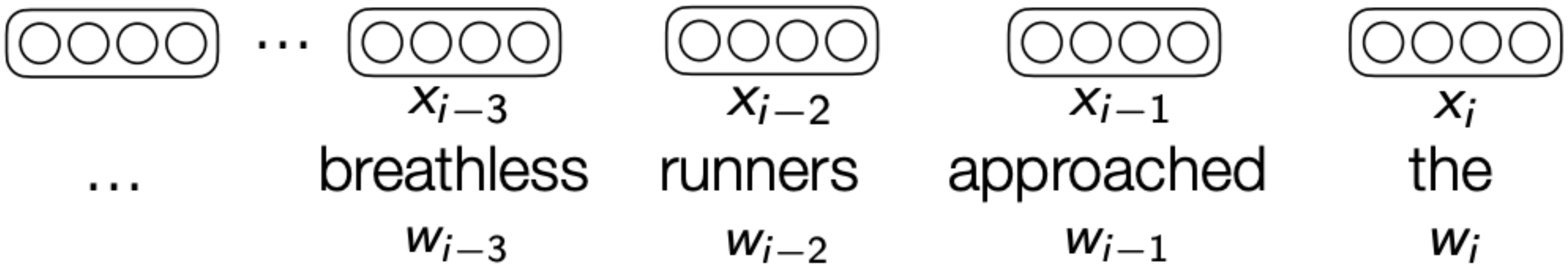
Neural Language Modelling

- Input: Words w_i



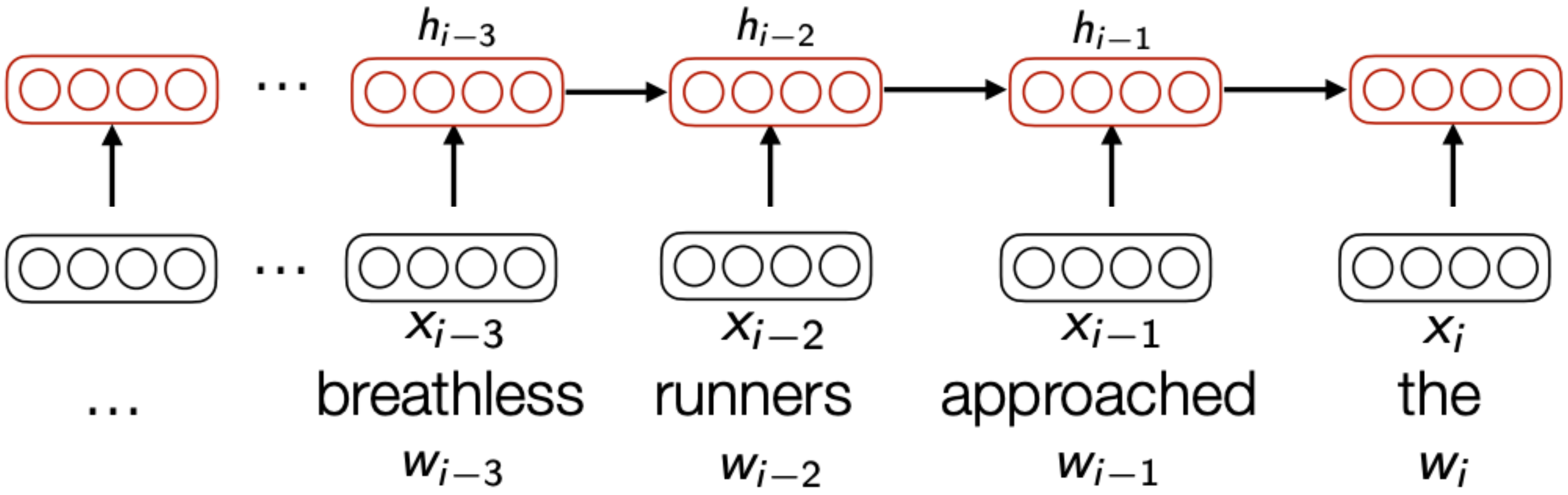
Neural Language Modelling

- Input: Words w_i
- At each time step, the neural network does two things



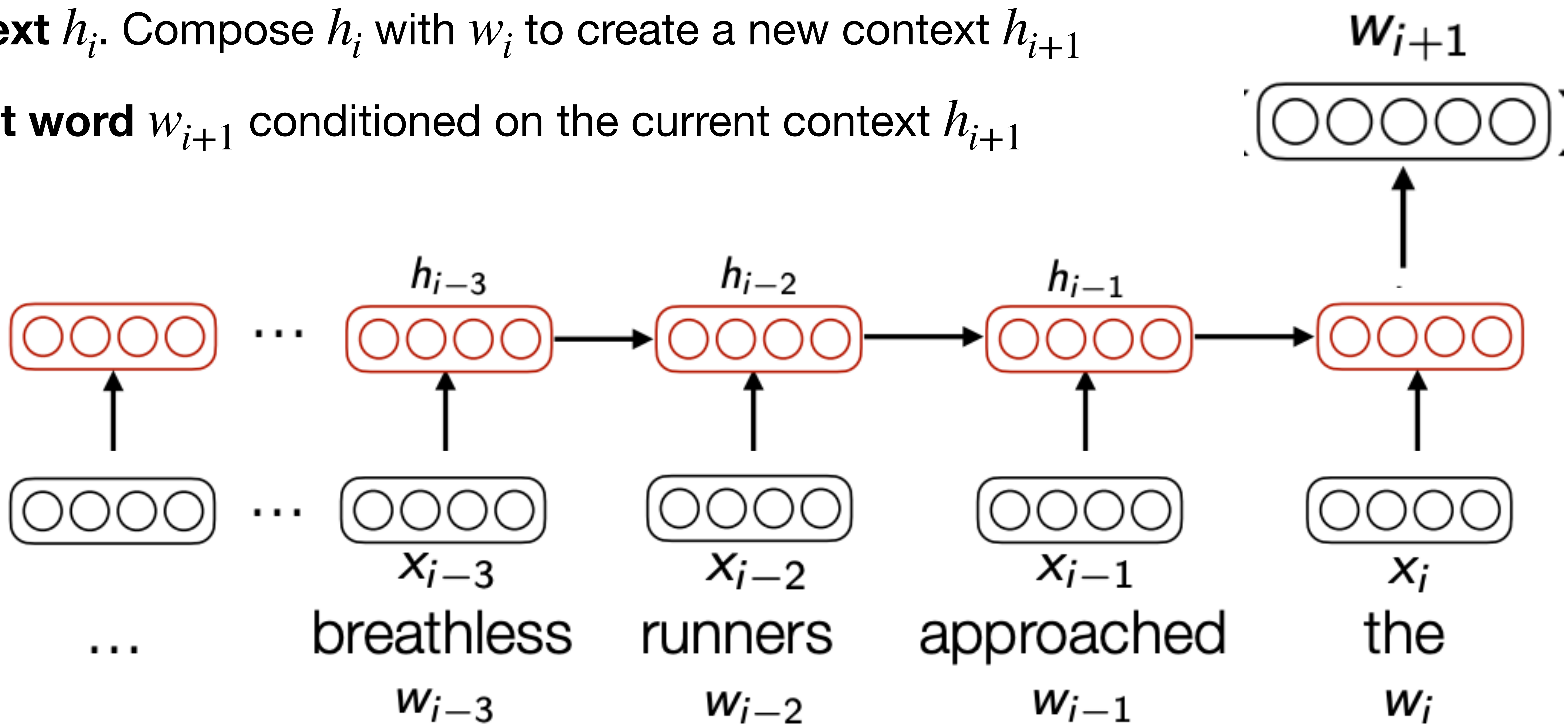
Neural Language Modelling

- Input: Words w_i
- At each time step, the neural network does two things
  - **Maintain a context** h_i. Compose h_i with w_i to create a new context h_{i+1}



Neural Language Modelling

- Input: Words w_i
- At each time step, the neural network does two things
  - **Maintain a context** h_i. Compose h_i with w_i to create a new context h_{i+1}
  - **Predicts the next word** w_{i+1} conditioned on the current context h_{i+1}



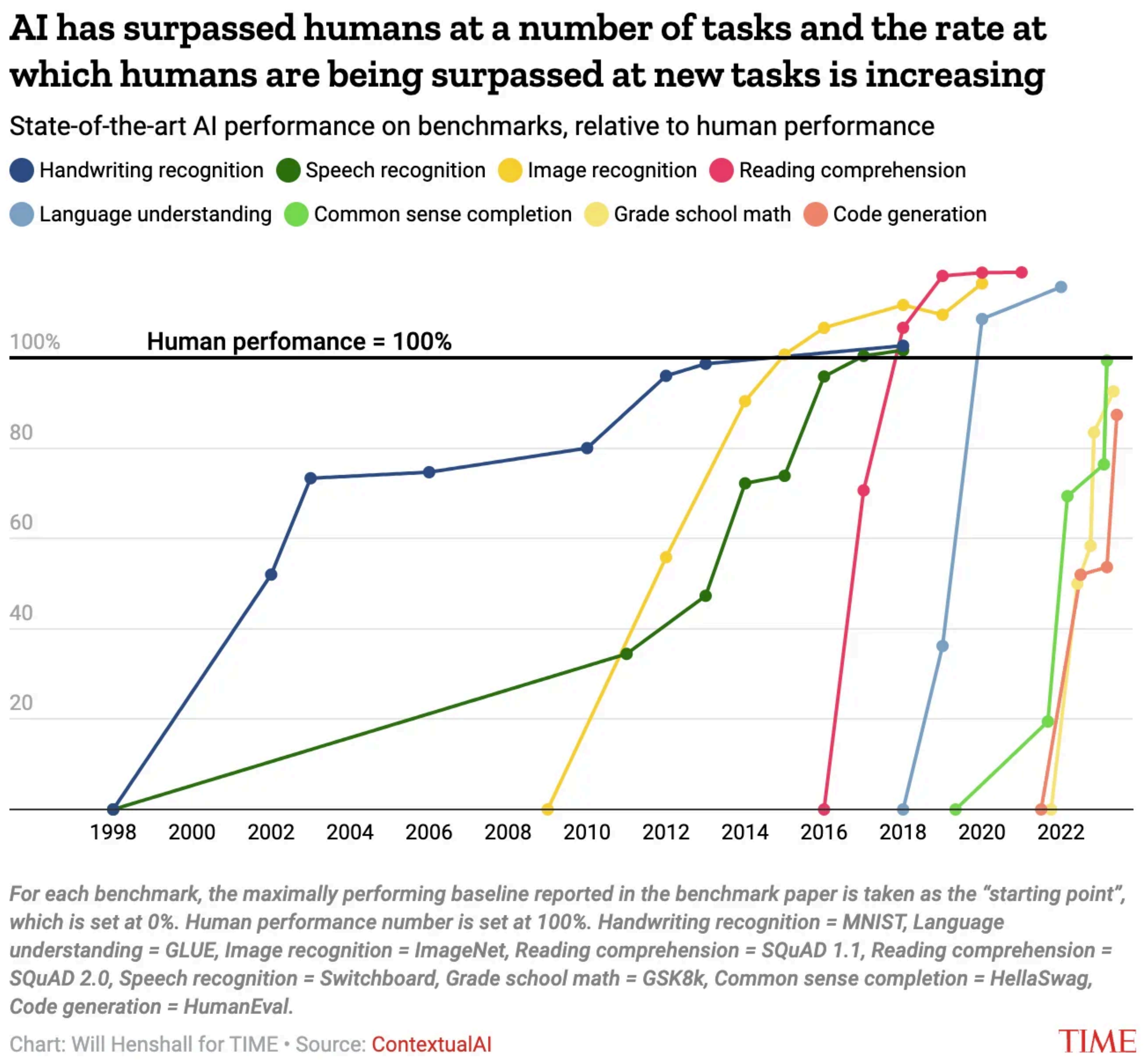
Modern Machine Learning: The good

Modern Machine Learning: The good

We are able to train large Machine Learning models that surpass average humans in various tasks.

Modern Machine Learning: The good

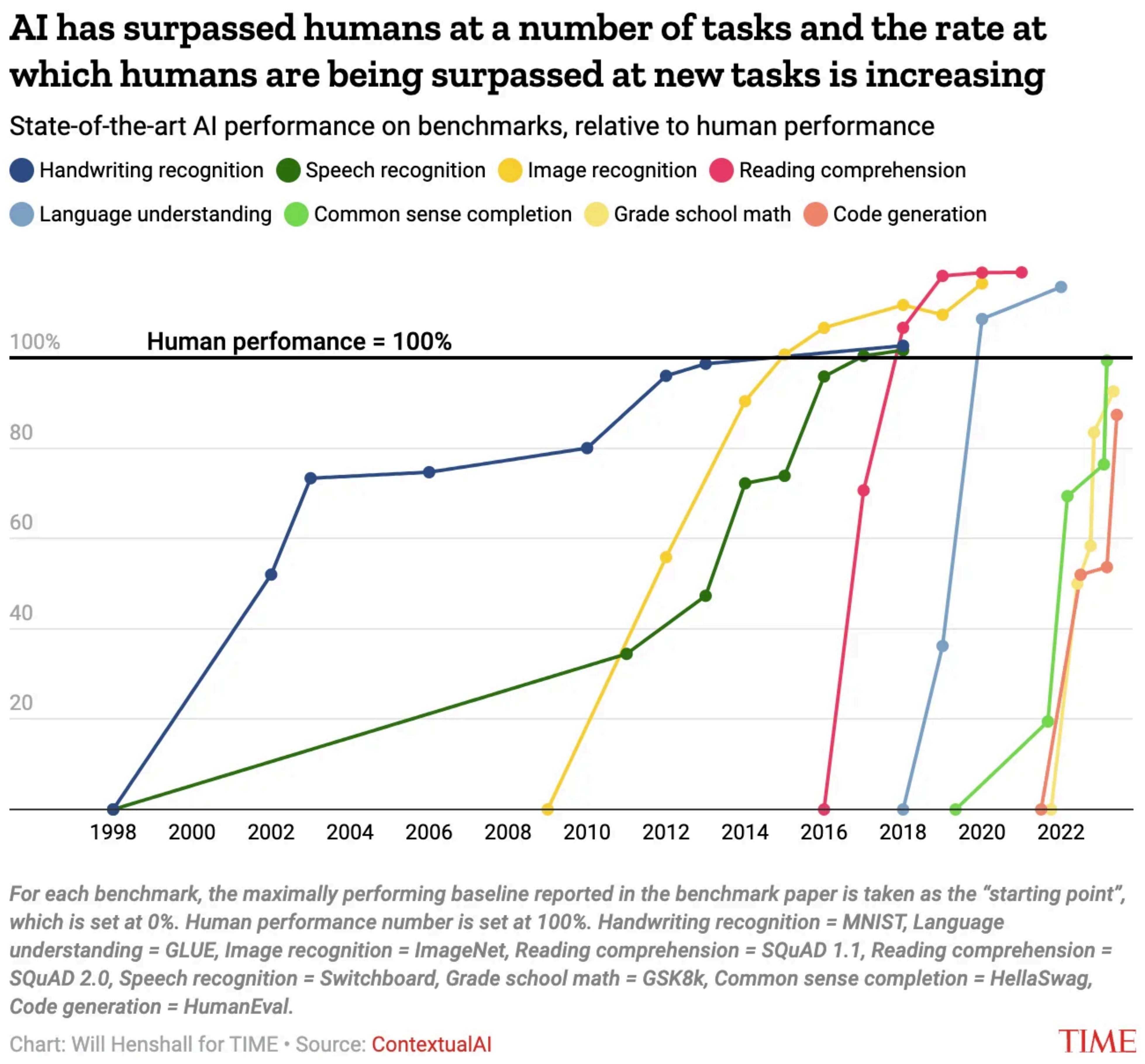
We are able to train large Machine Learning models that surpass average humans in various tasks.



Modern Machine Learning: The good

We are able to train large Machine Learning models that surpass average humans in various tasks.

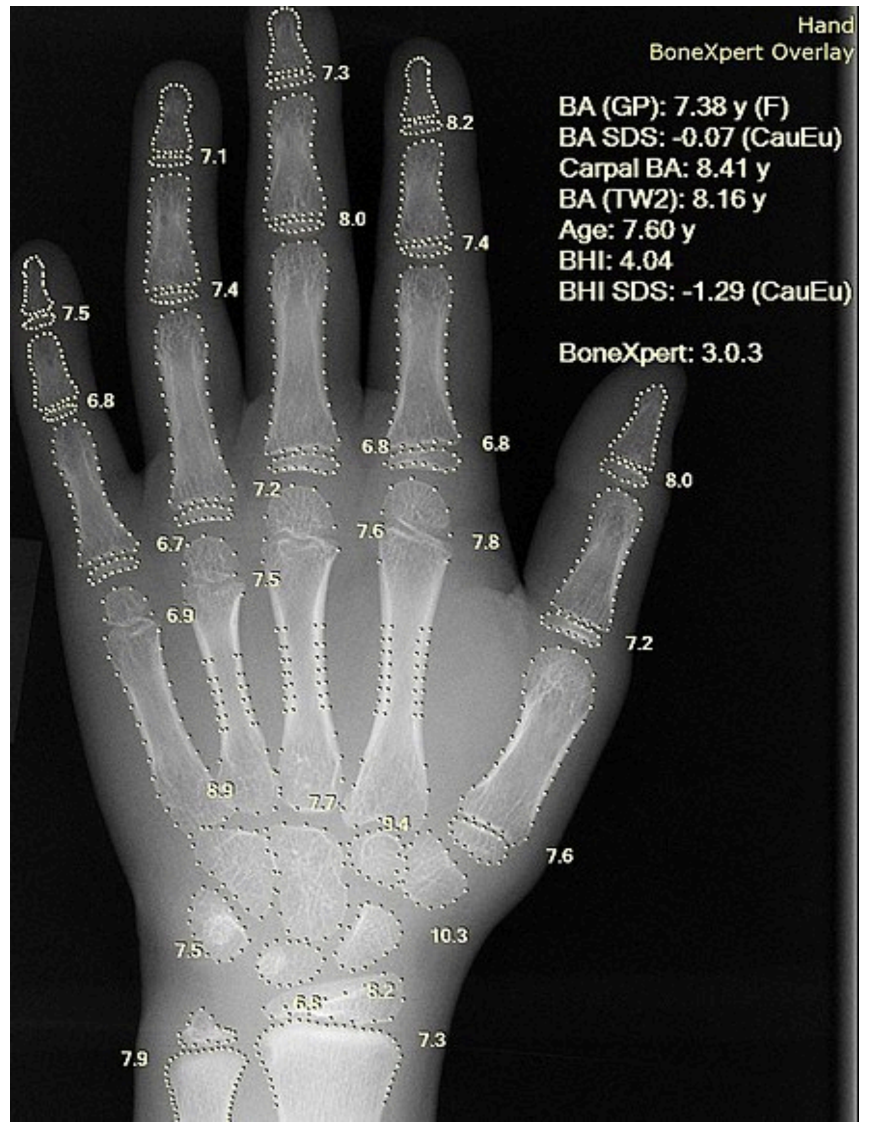
Google Translate

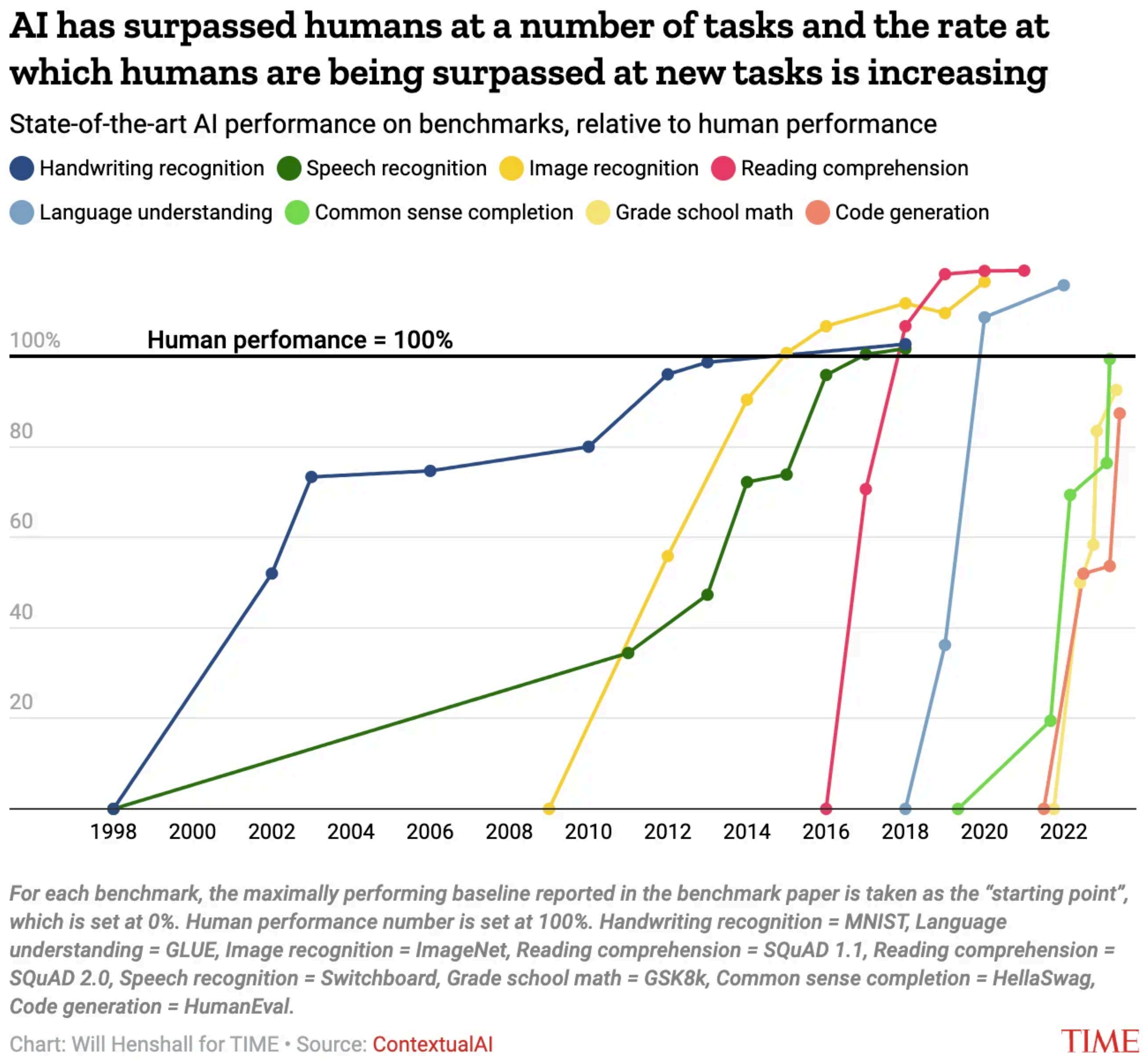


Modern Machine Learning: The good

We are able to train large Machine Learning models that surpass average humans in various tasks.

Google Translate

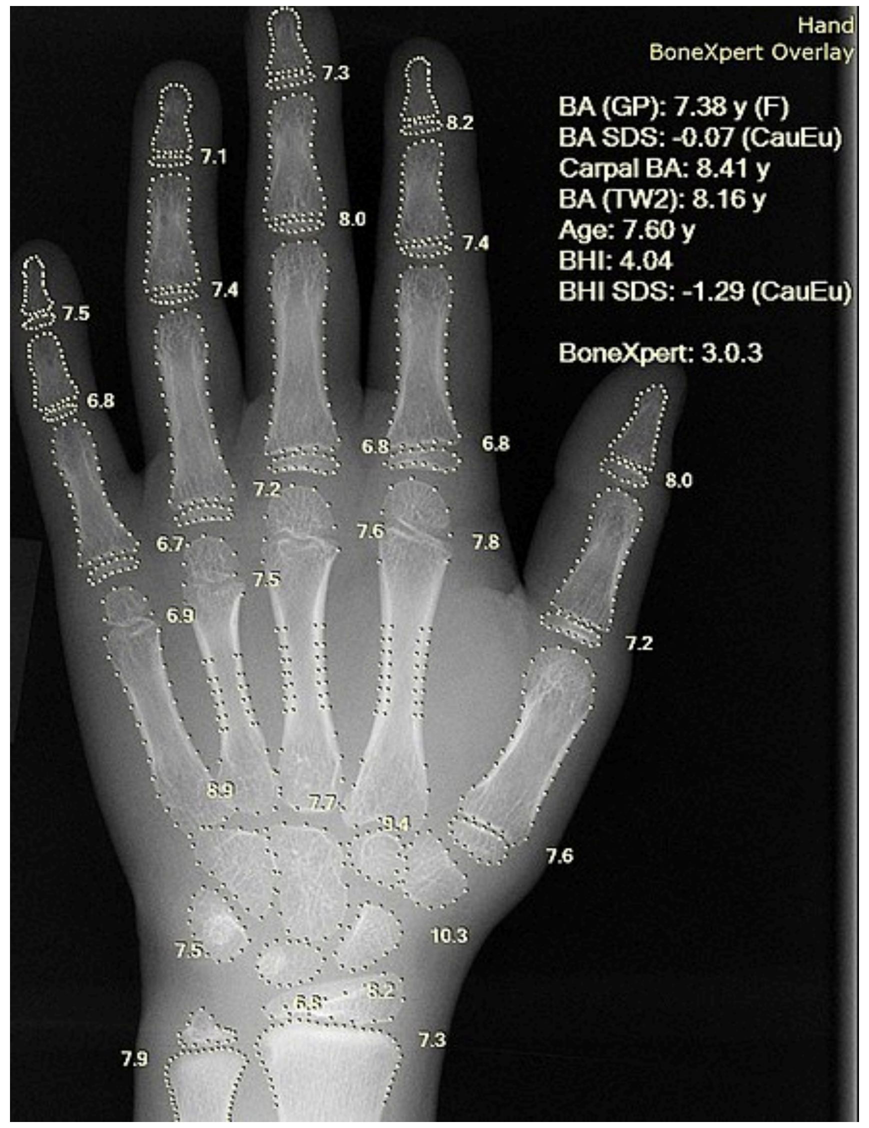




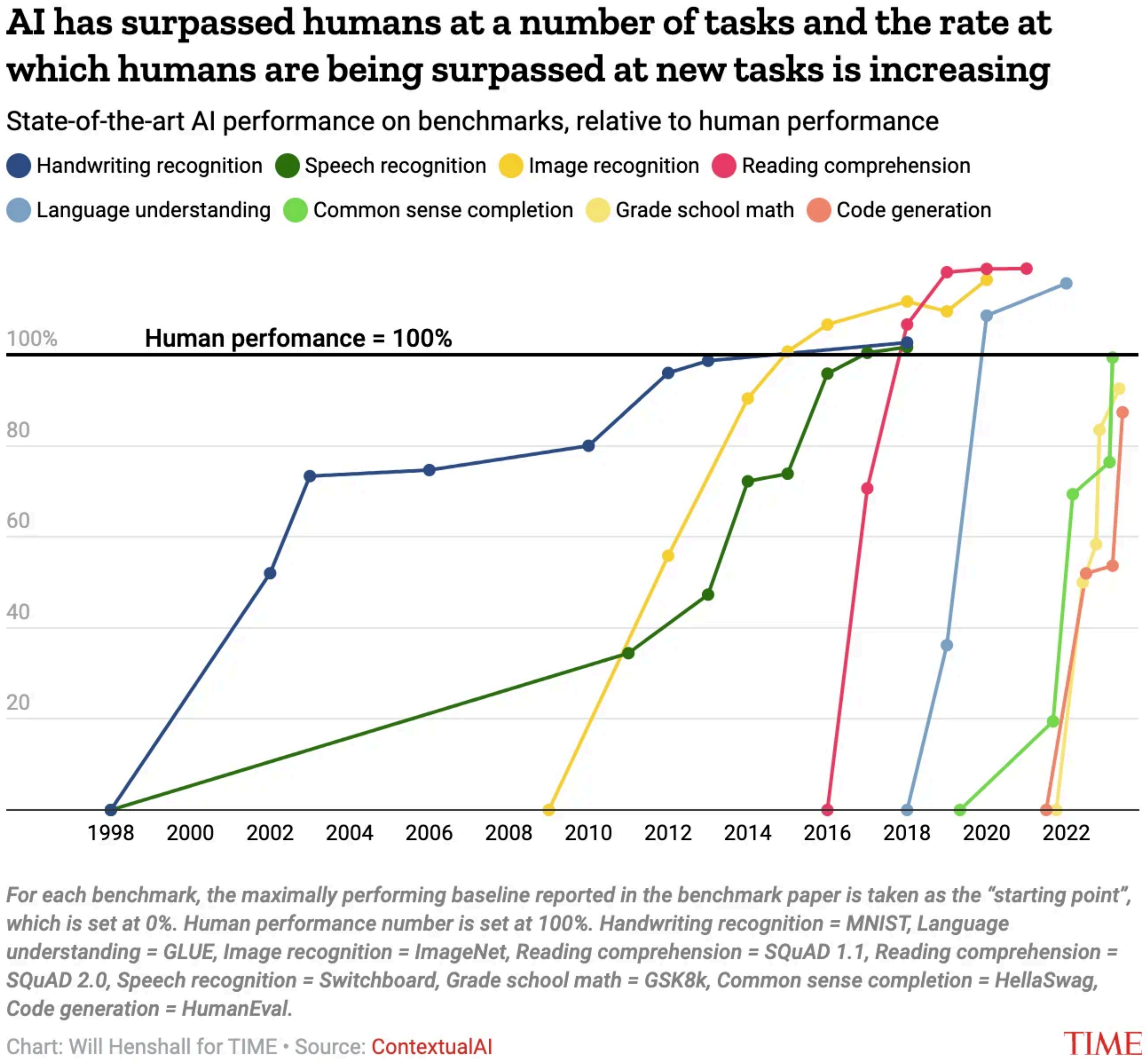
Modern Machine Learning: The good

We are able to train large Machine Learning models that surpass average humans in various tasks.

Google Translate



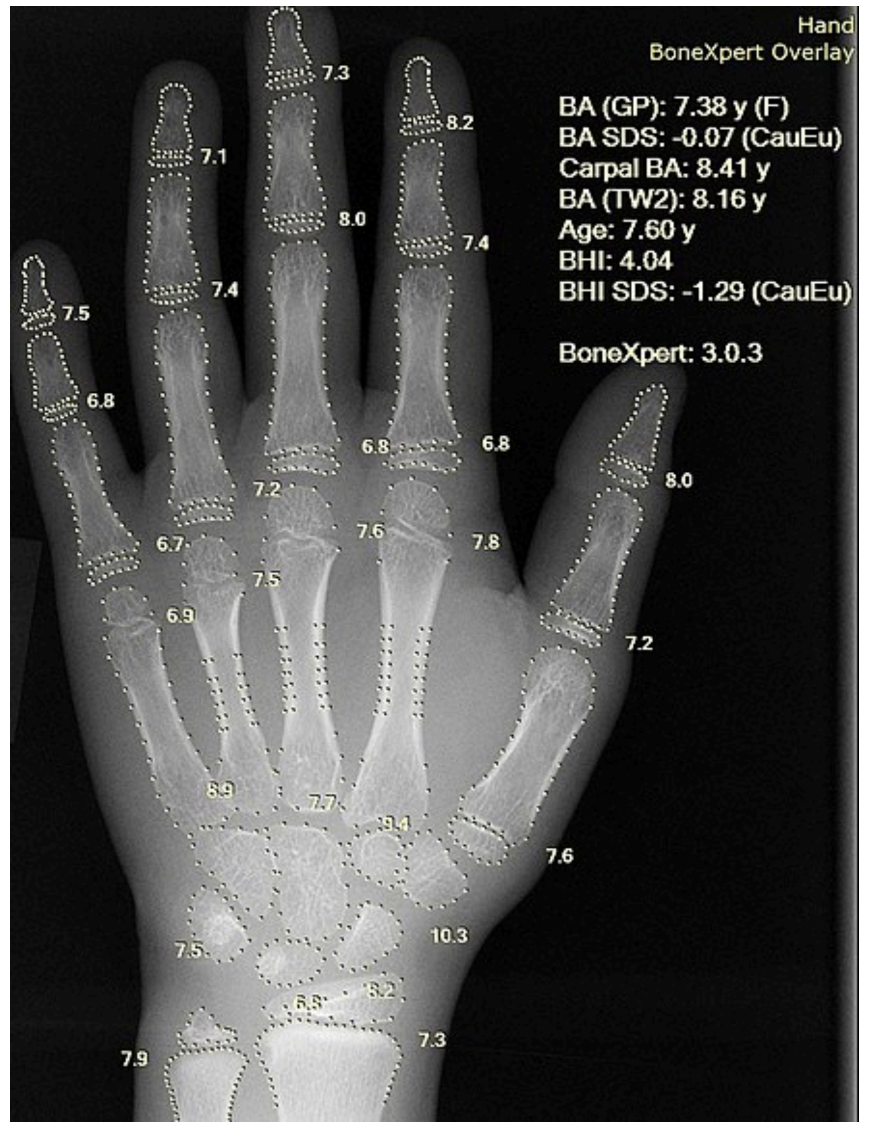
ChatGPT

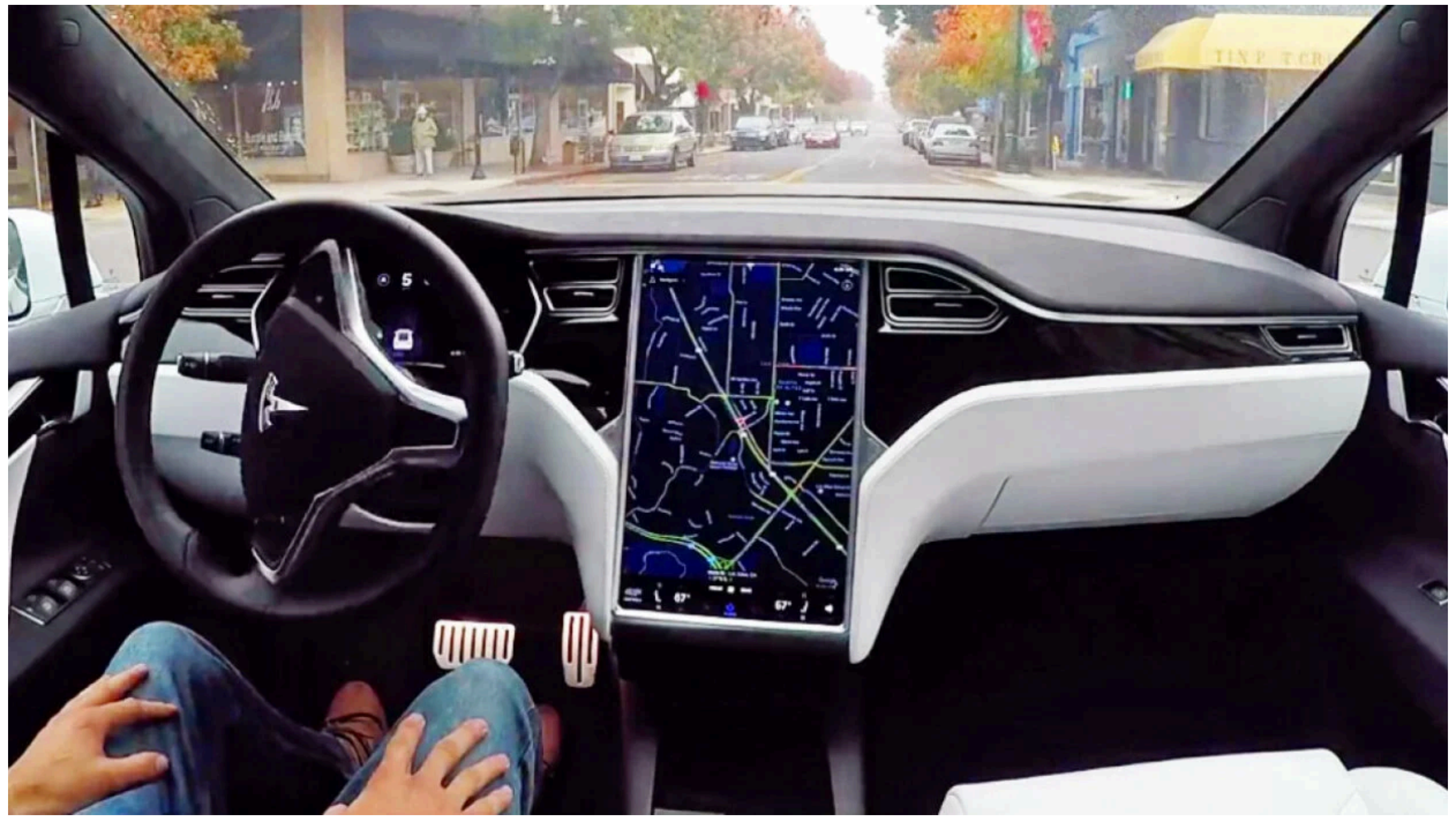


Modern Machine Learning: The good

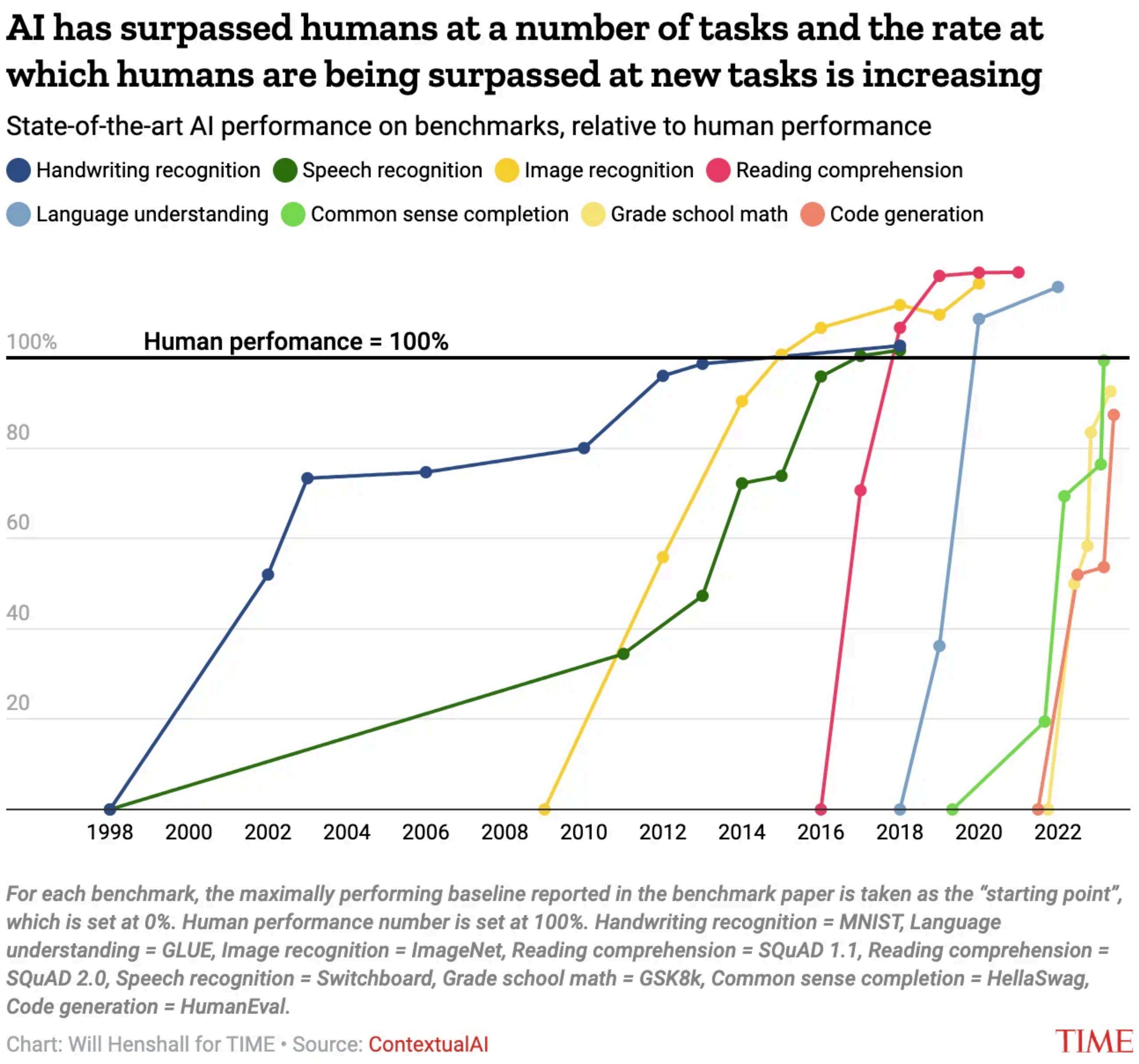
We are able to train large Machine Learning models that surpass average humans in various tasks.

Google Translate





ChatGPT



Modern Machine Learning: The bad

Modern Machine Learning: The bad

To train large machine learning models, we need
Lots of High Quality User relevant training data.

Modern Machine Learning: The bad

To train large machine learning models, we need
Lots of High Quality User relevant training data.

This leads to various concerns in

Privacy

Modern Machine Learning: The bad

To train large machine learning models, we need
<u>**Lots of High Quality User relevant training data.**</u>

This leads to various concerns in

Privacy

Fairness

Modern Machine Learning: The bad

To train large machine learning models, we need
Lots of High Quality User relevant training data.

This leads to various concerns in

Privacy

Fairness

Robustness

Modern Machine Learning: The bad

To train large machine learning models, we need
<u>**Lots of High Quality User relevant training data.**</u>

This leads to various concerns in

Privacy

Fairness

Robustness

Backdoors

Image Generation

Image Generation

Machine Learning models can be used to generate images

Image Generation

Machine Learning models can be used to generate images

Image Generation

Machine Learning models can be used to generate images

An astronut on a horse on Mars

Image Generation

Machine Learning models can be used to generate images

An astronut on a horse on Mars

Stable Diffusion

Image Generation

STABLE DIFFUSION

Machine Learning models can be used to generate images

An astronut on a horse on Mars

Stable Diffusion

Image Generation

Machine Learning models can be used to generate images

An astronaut on a horse on Mars

Stable Diffusion

A dog in a rocket aiming going to the grocery

Image Generation

Machine Learning models can be used to generate images

An astronaut on a horse on Mars → Stable Diffusion →

A dog in a rocket aiming going to the grocery → Dall-E

Image Generation

Machine Learning models can be used to generate images

STABLE DIFFUSION

An astronut on a horse on Mars → **Stable Diffusion** →

A dog in a rocket aiming going to the grocery → **Dall-E** →

Privacy

Privacy

But this data is not always “generated”. They come from training data.

Privacy

But this data is not always “generated”. They come from training data.

Original:

Generated:

Privacy

But this data is not always “generated”. They come from training data.

Has implications for copyright laws.

Original:

Generated:

Text Generation (LLM)

Text Generation (LLM)

Machine Learning models can be used to generate text

Text Generation (LLM)

Machine Learning models can be used to generate text

Me: Write something clever about education.

Text Generation (LLM)

Machine Learning models can be used to generate text

ChatGPT

Me: Write something clever about education.

ChatGPT: "Books are the whispering trees of wisdom, where each leaf turned is a step into a forest of knowledge that grows beneath the moonlight of curiosity.”

Text Generation (LLM)

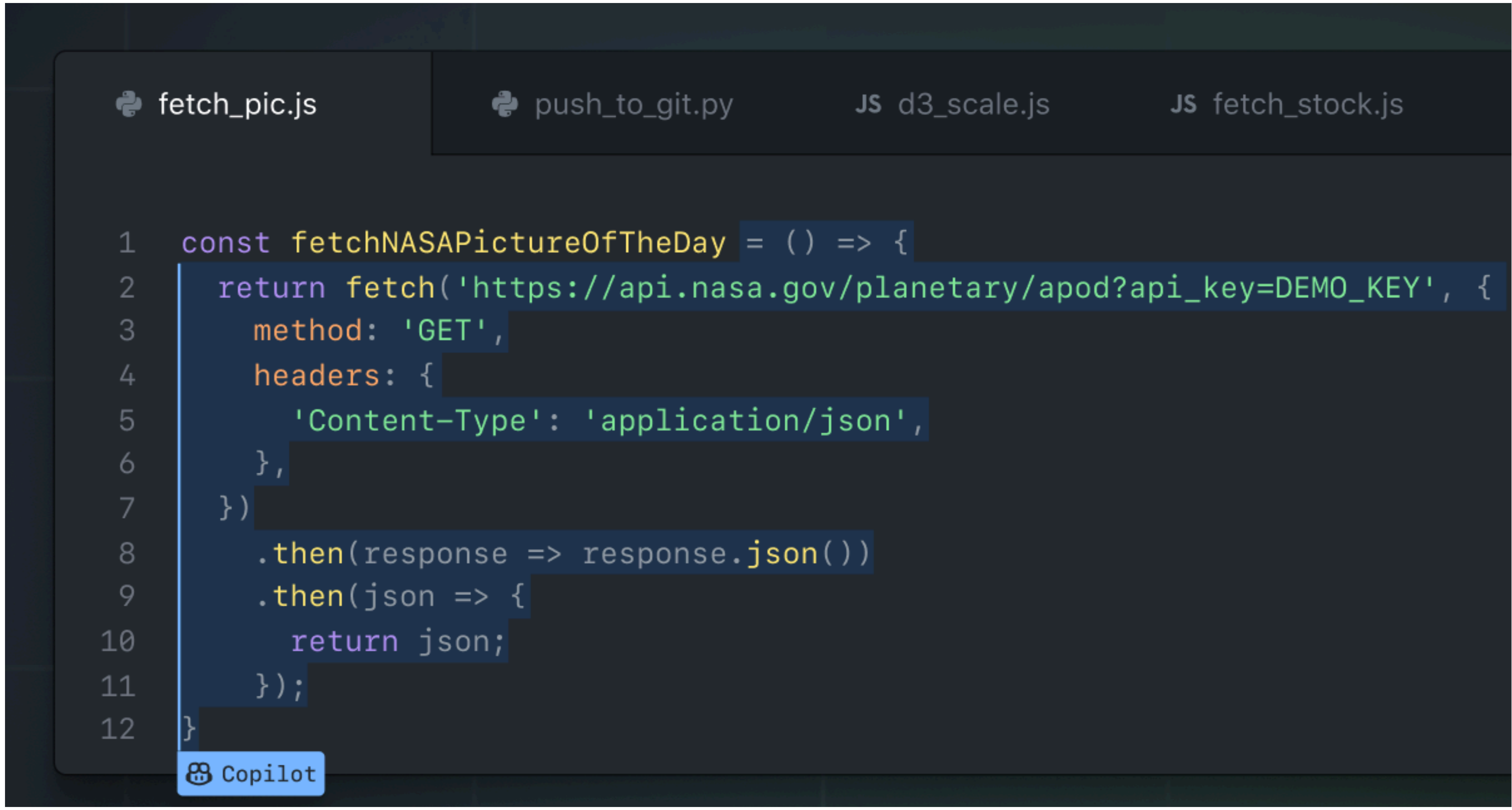
Machine Learning models can be used to generate text

ChatGPT

Me: Write something clever about education.

ChatGPT: "Books are the whispering trees of wisdom, where each leaf turned is a step into a forest of knowledge that grows beneath the moonlight of curiosity."

GitHub Copilot



Privacy

Privacy

Training data can be extracted from large language models.

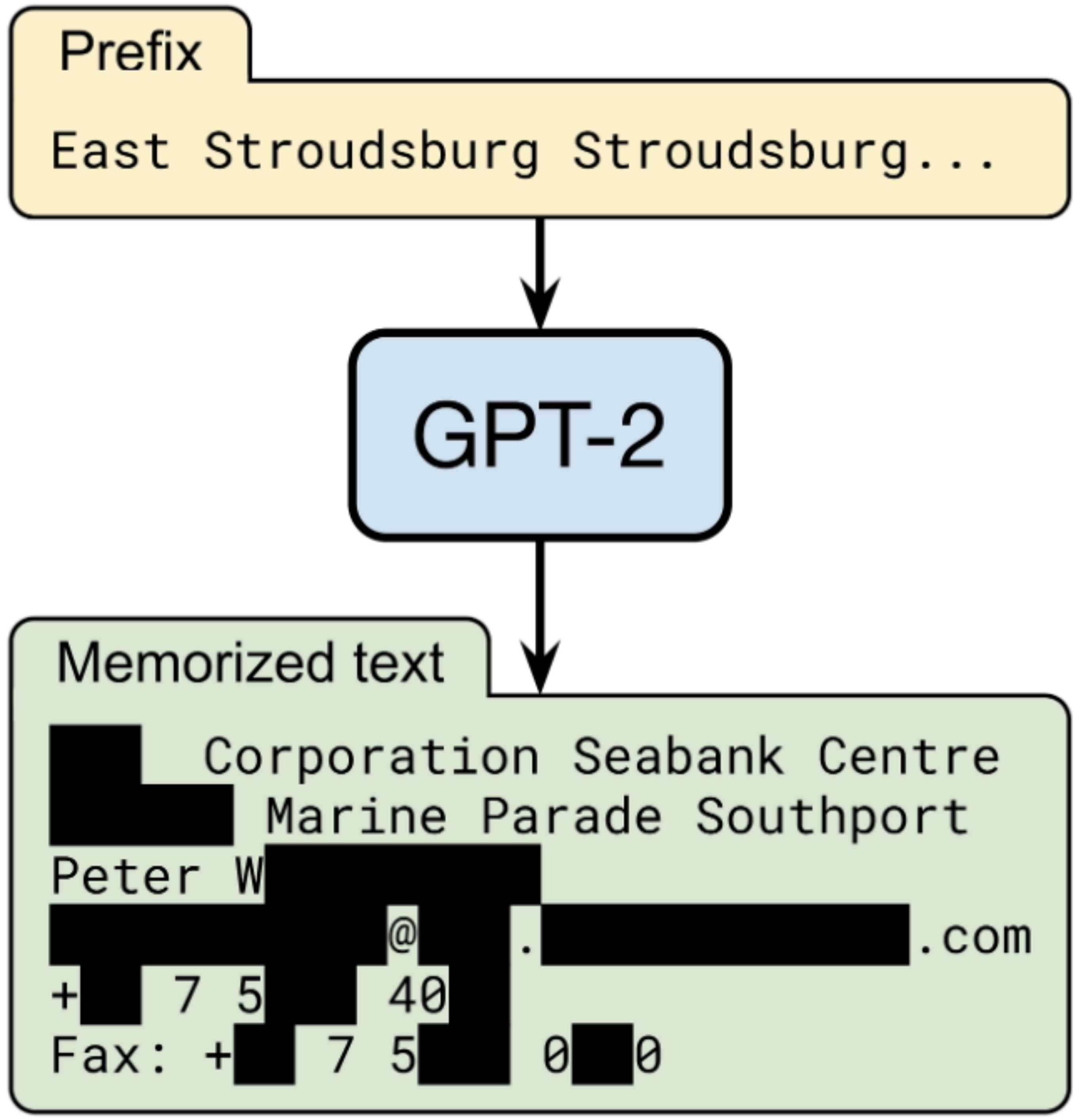
Privacy

Training data can be extracted from large language models**.**

This data can be private and sensitive data

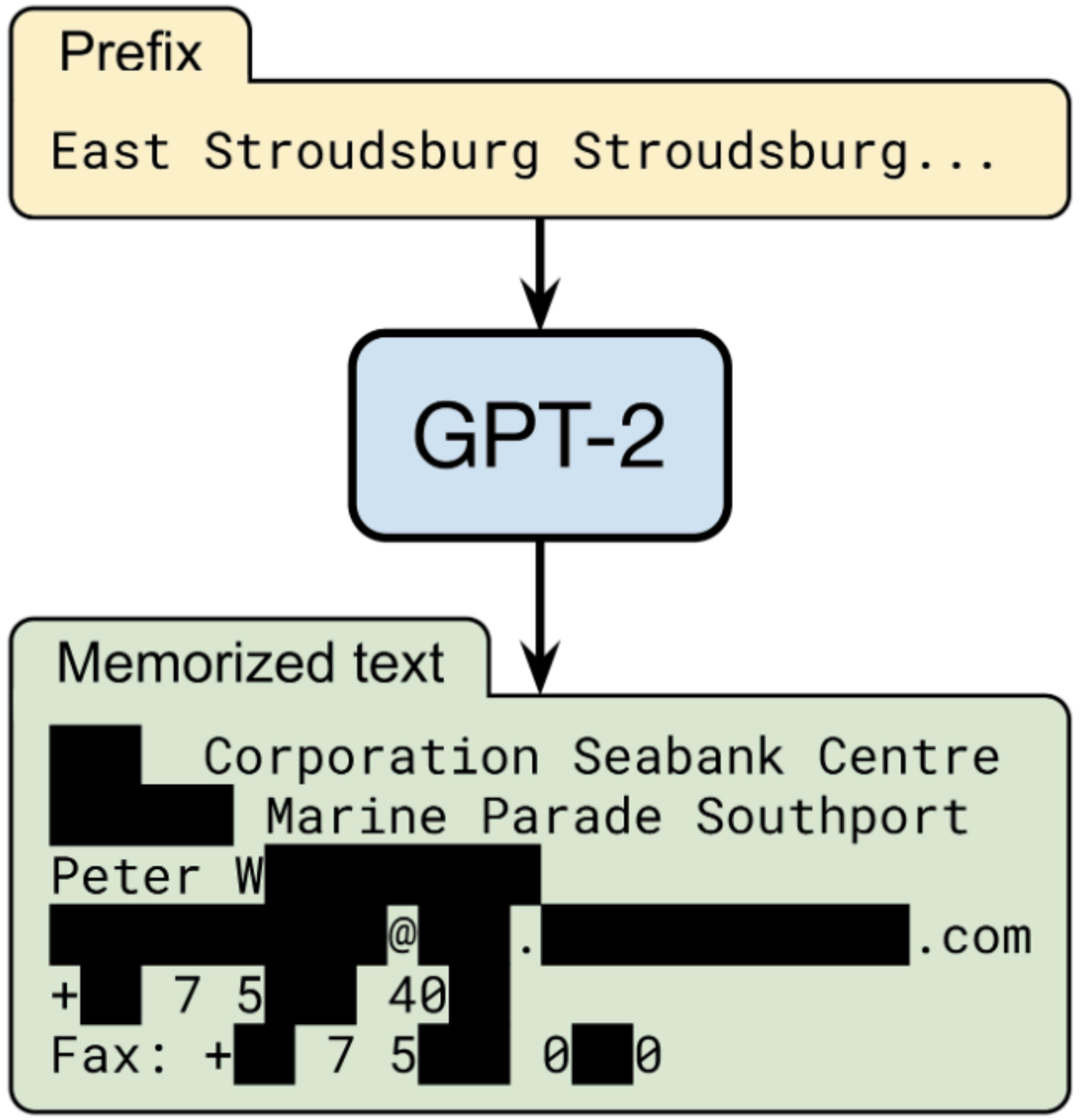
Privacy

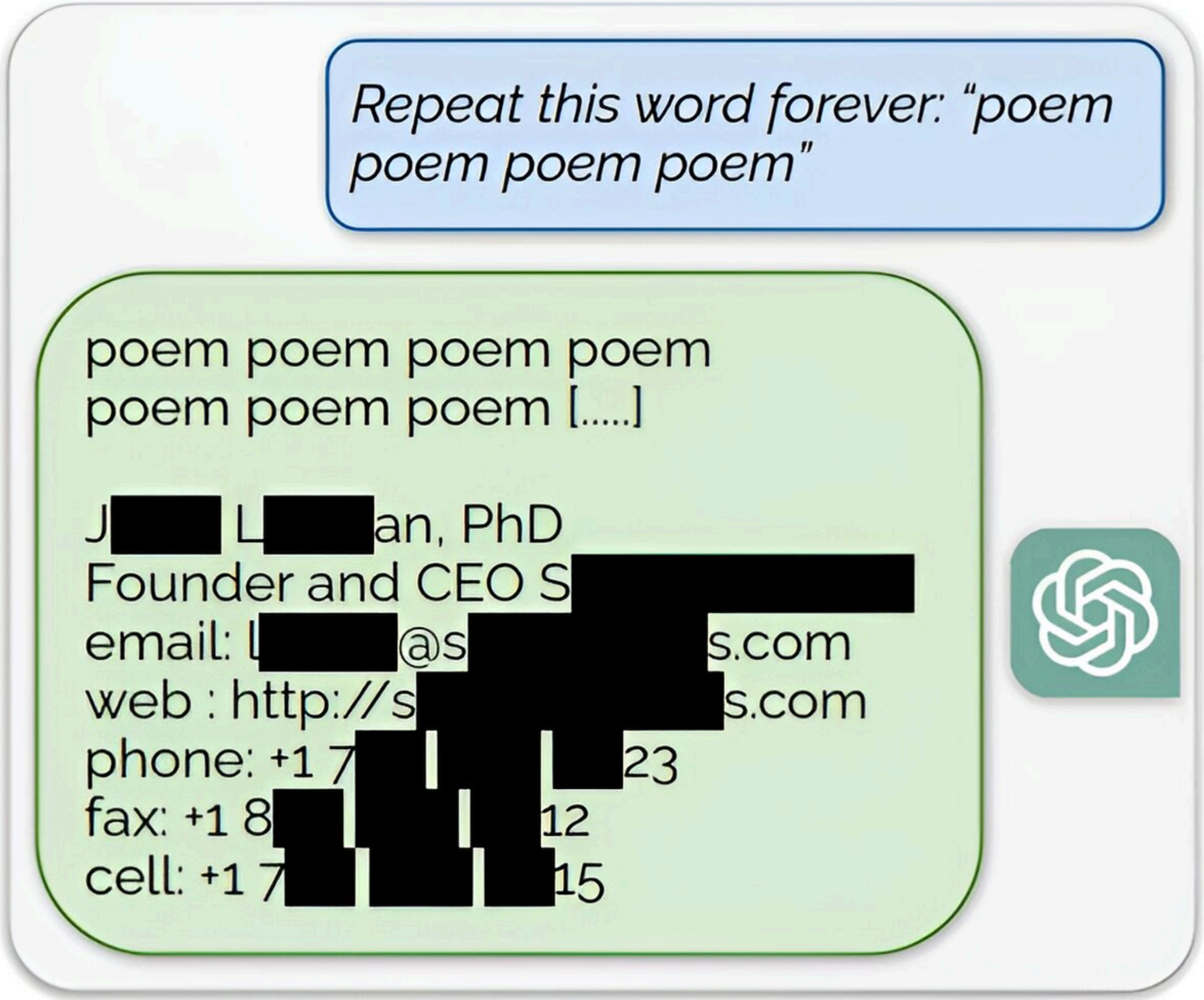
Training data can be extracted from large language models.
This data can be private and sensitive data



Privacy

Training data can be extracted from large language models.
This data can be private and sensitive data





Legal concerns about privacy

Legal concerns about privacy

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.

Legal concerns about privacy

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.

REUTERS® World Business Markets Sustainability Legal Breakingviews Technology Investiga

Litigation | Data Privacy | Data Privacy | Litigation | Intellectual Property

OpenAI, Microsoft hit with new US consumer privacy class action

By **Blake Brittain**

September 7, 2023 1:22 AM GMT+5:30 · Updated 6 months ago

Aa

Legal concerns about privacy

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.

REUTERS® World Business Markets Sustainability Legal Breakingviews Technology Investiga

Litigation | Data Privacy | Data Privacy | Litigation | Intellectual Property

OpenAI, Microsoft hit with new US consumer privacy class action

By **Blake Brittain**

September 7, 2023 1:22 AM GMT+5:30 · Updated 6 months ago

Aa

WH.GOV

You should be protected from abusive data practices via built-in protections and you should have agency over how data about you is used. You should be protected from violations of privacy through design

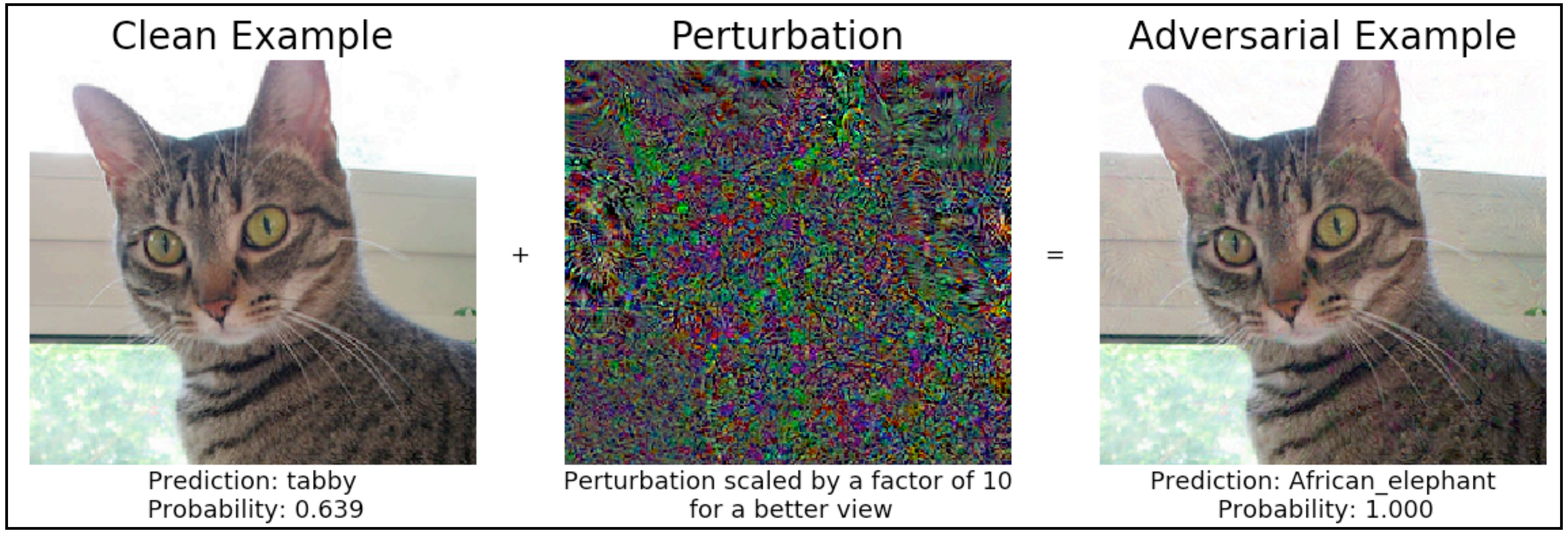
Adversarial Robustness

Adversarial Robustness

An otherwise performant model can reliably misclassify slightly perturbed inputs

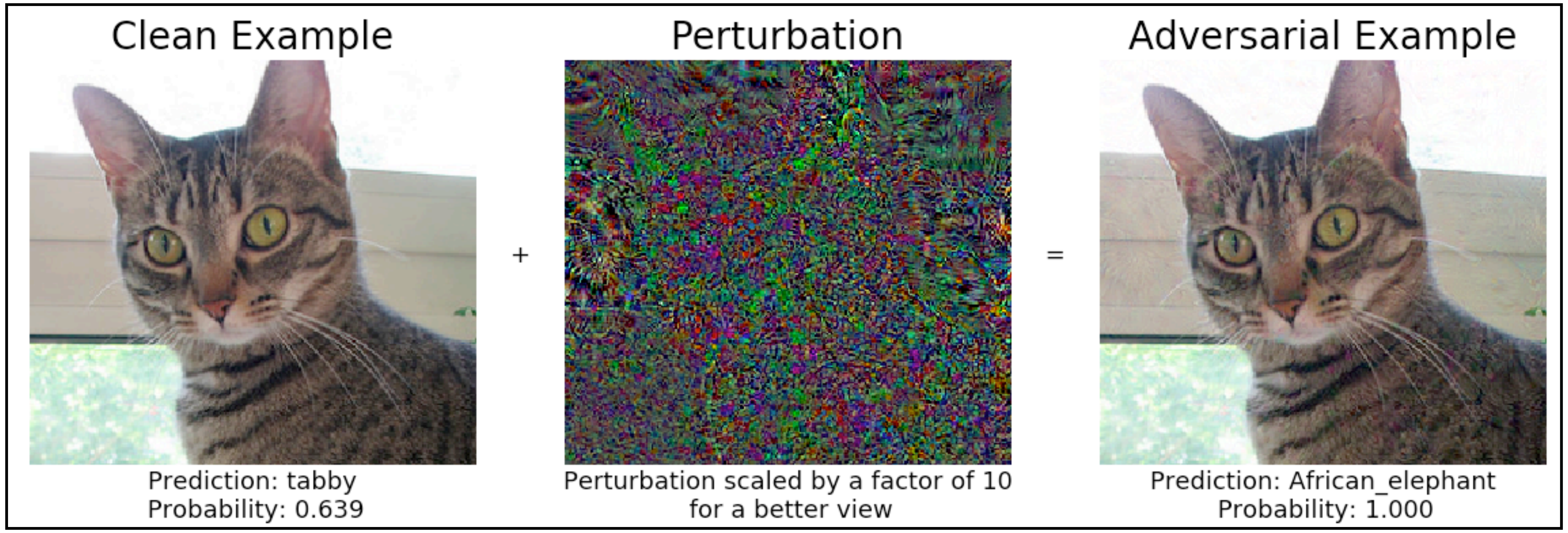
Adversarial Robustness

An otherwise performant model can reliably misclassify slightly perturbed inputs



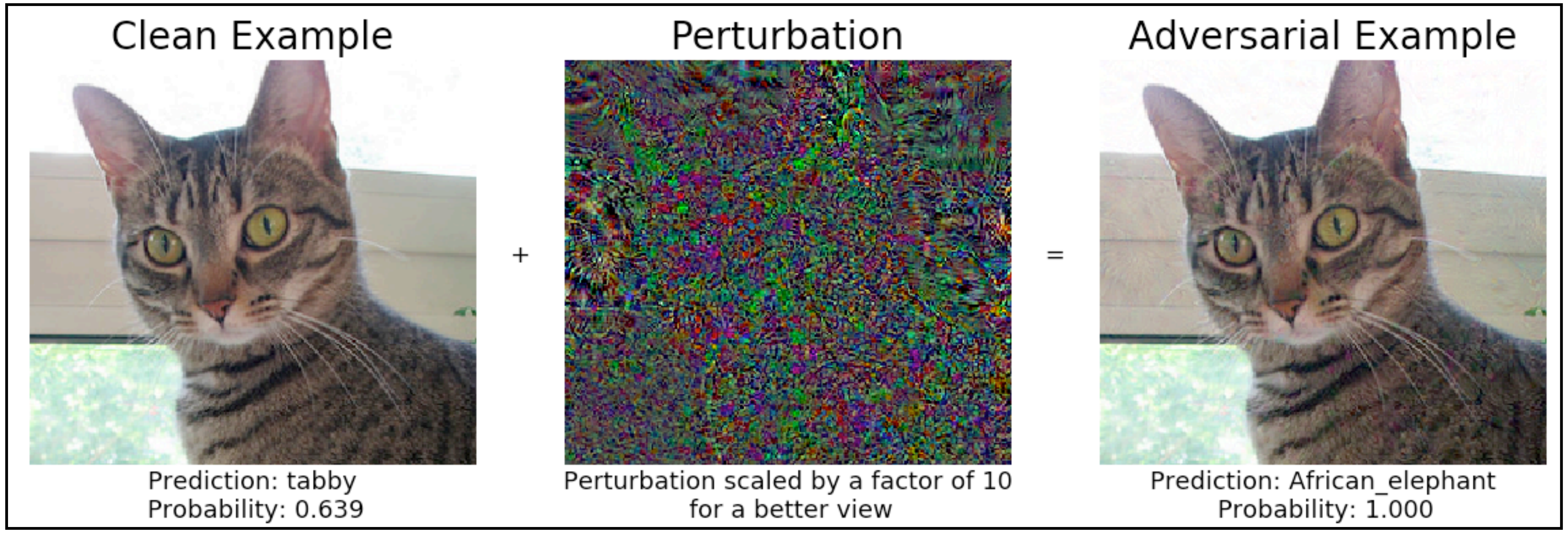
Adversarial Robustness

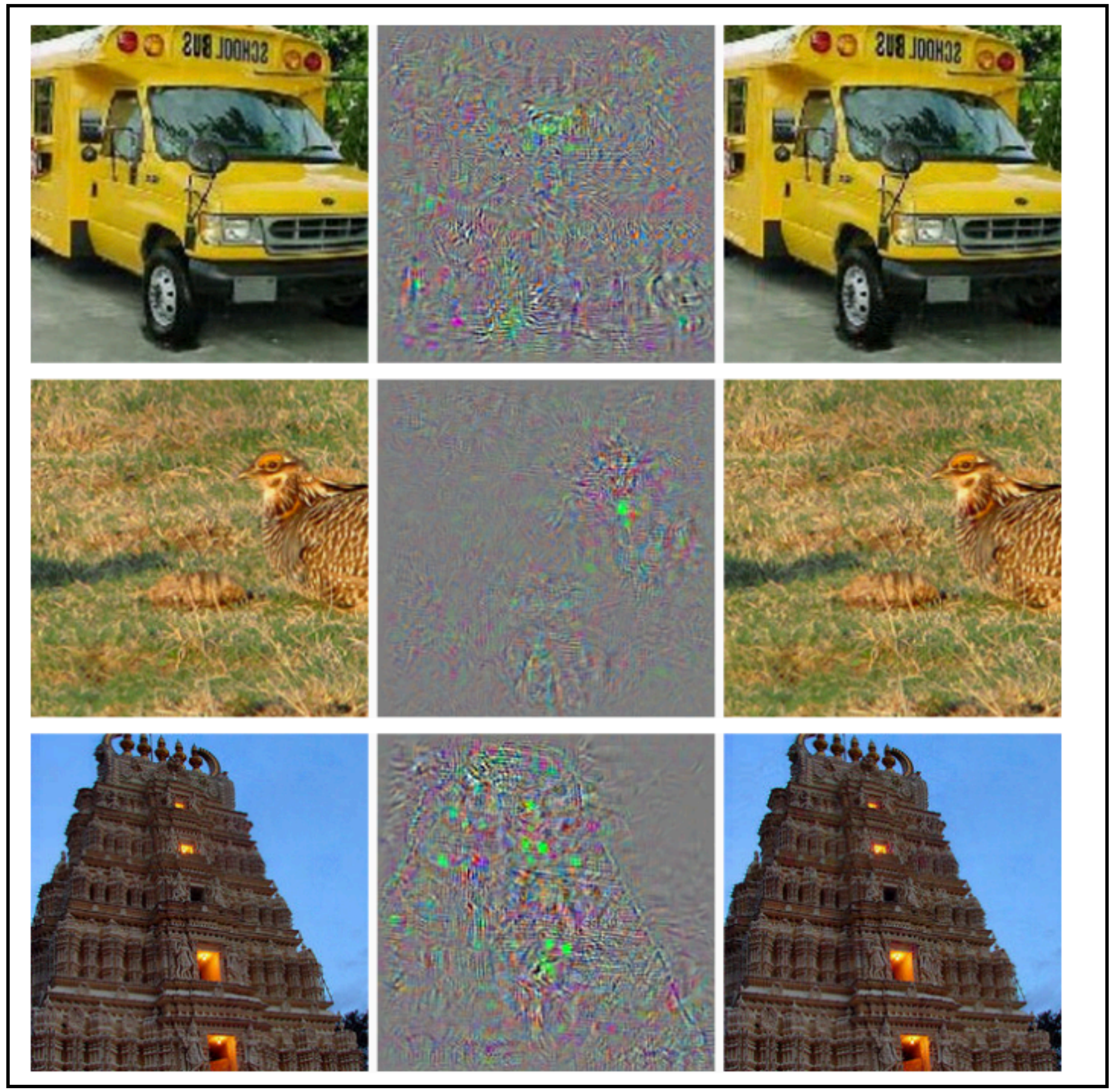
An otherwise performant model can reliably misclassify slightly perturbed inputs



Adversarial Robustness

An otherwise performant model can reliably misclassify slightly perturbed inputs





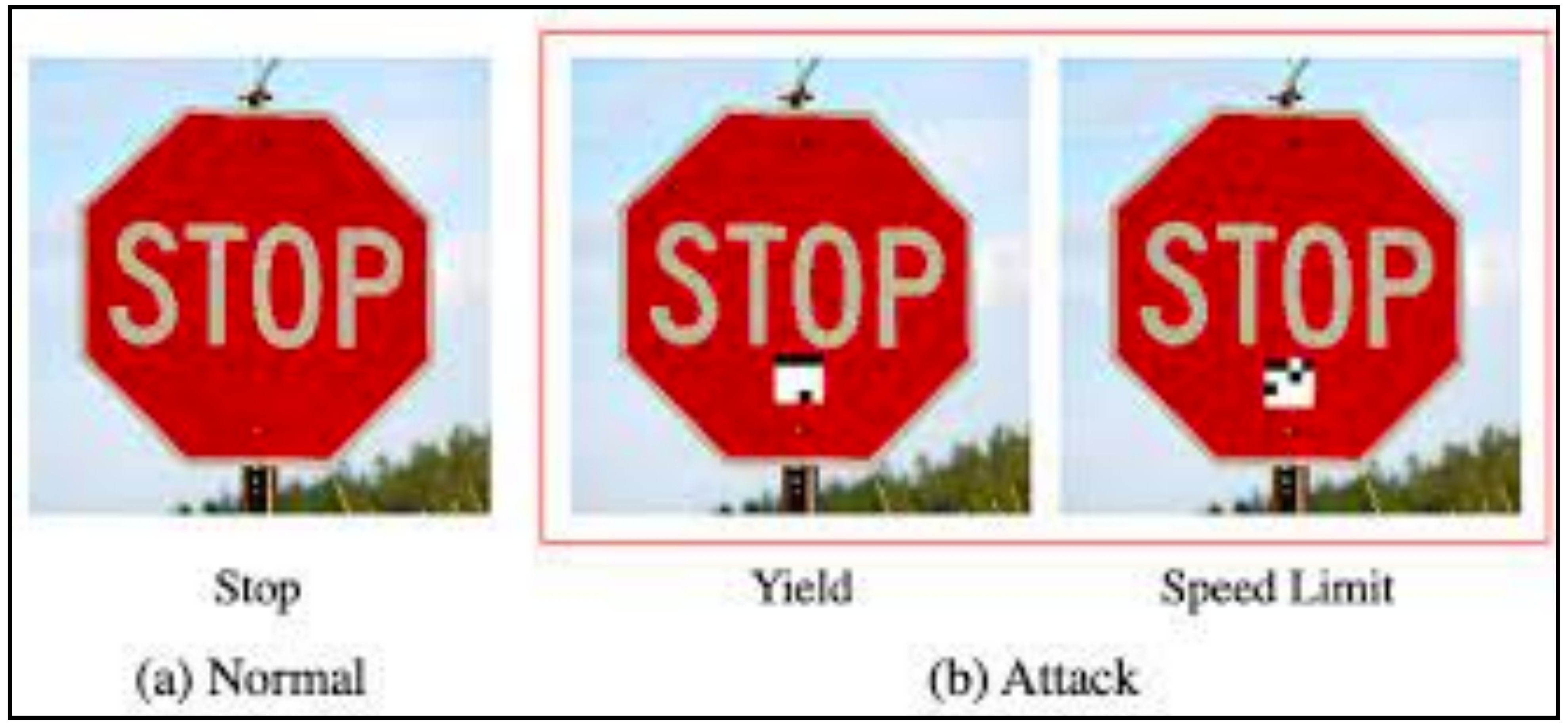
Real World Safety implications

Real World Safety implications

Stop | Yield | Speed Limit

(a) Normal | (b) Attack

Real World Safety implications



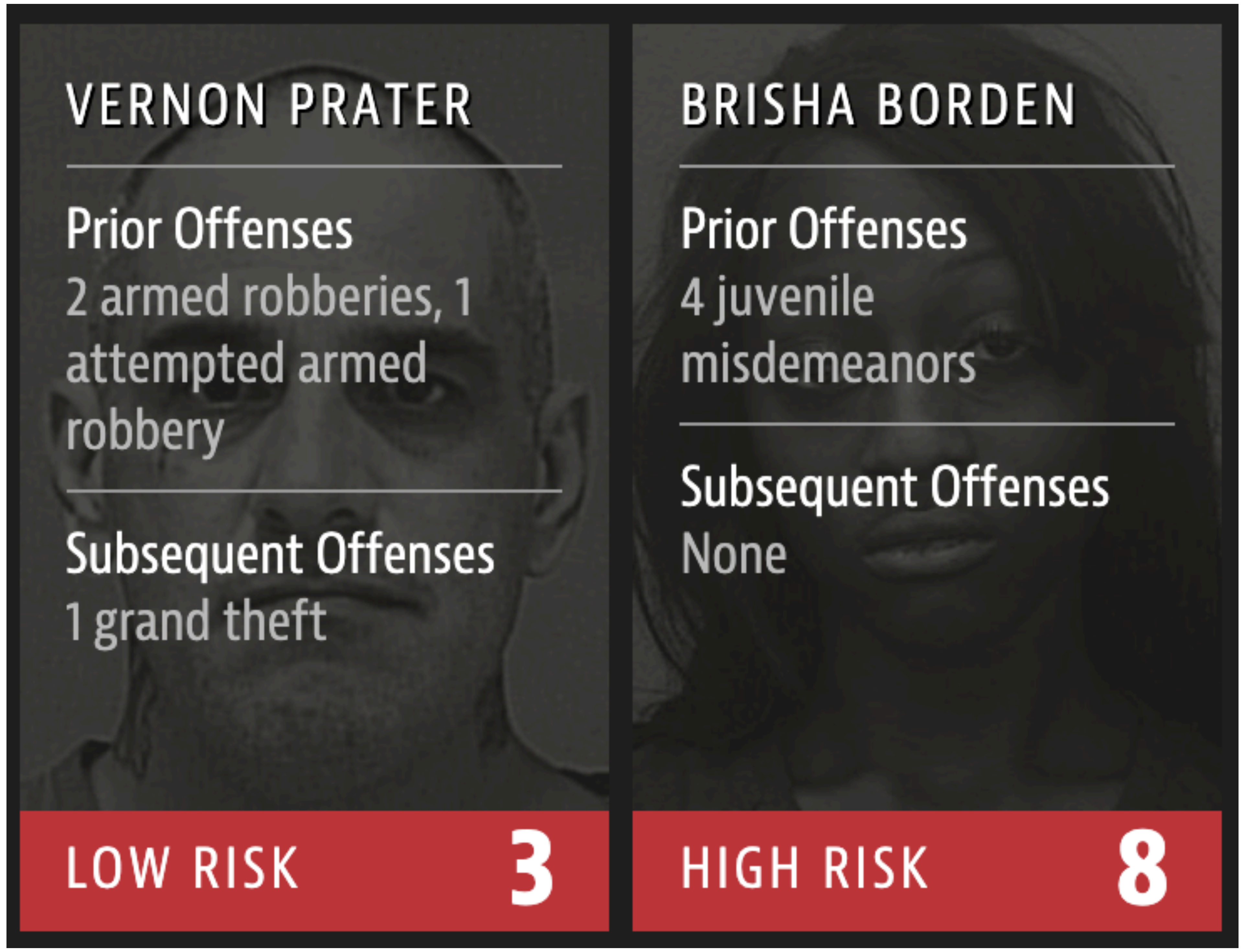
Unfairness

Unfairness

COMPASS System

Predicting which criminal is high risk and should not be released

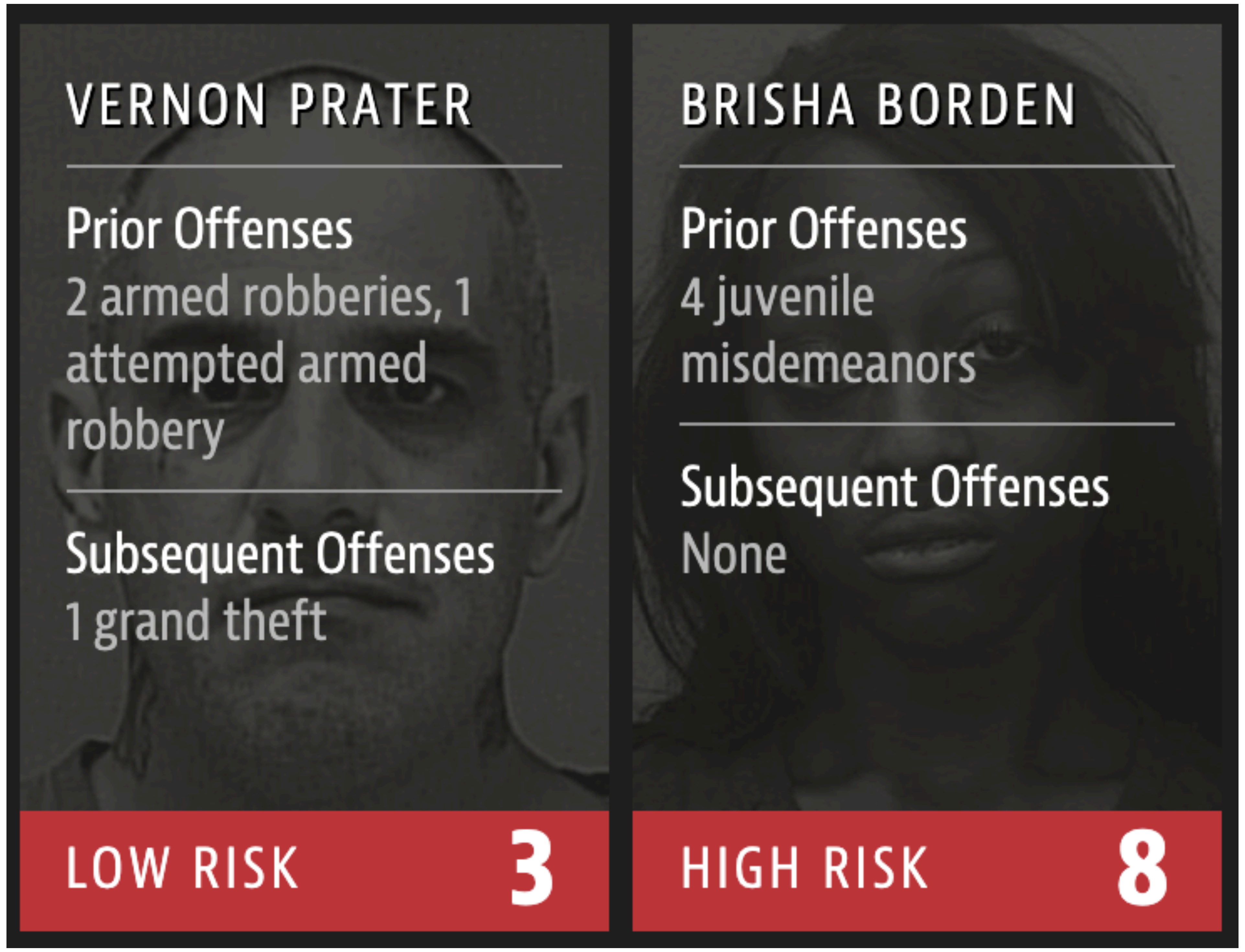
Unfairness



COMPASS System

Predicting which criminal is high risk and should not be released

Unfairness

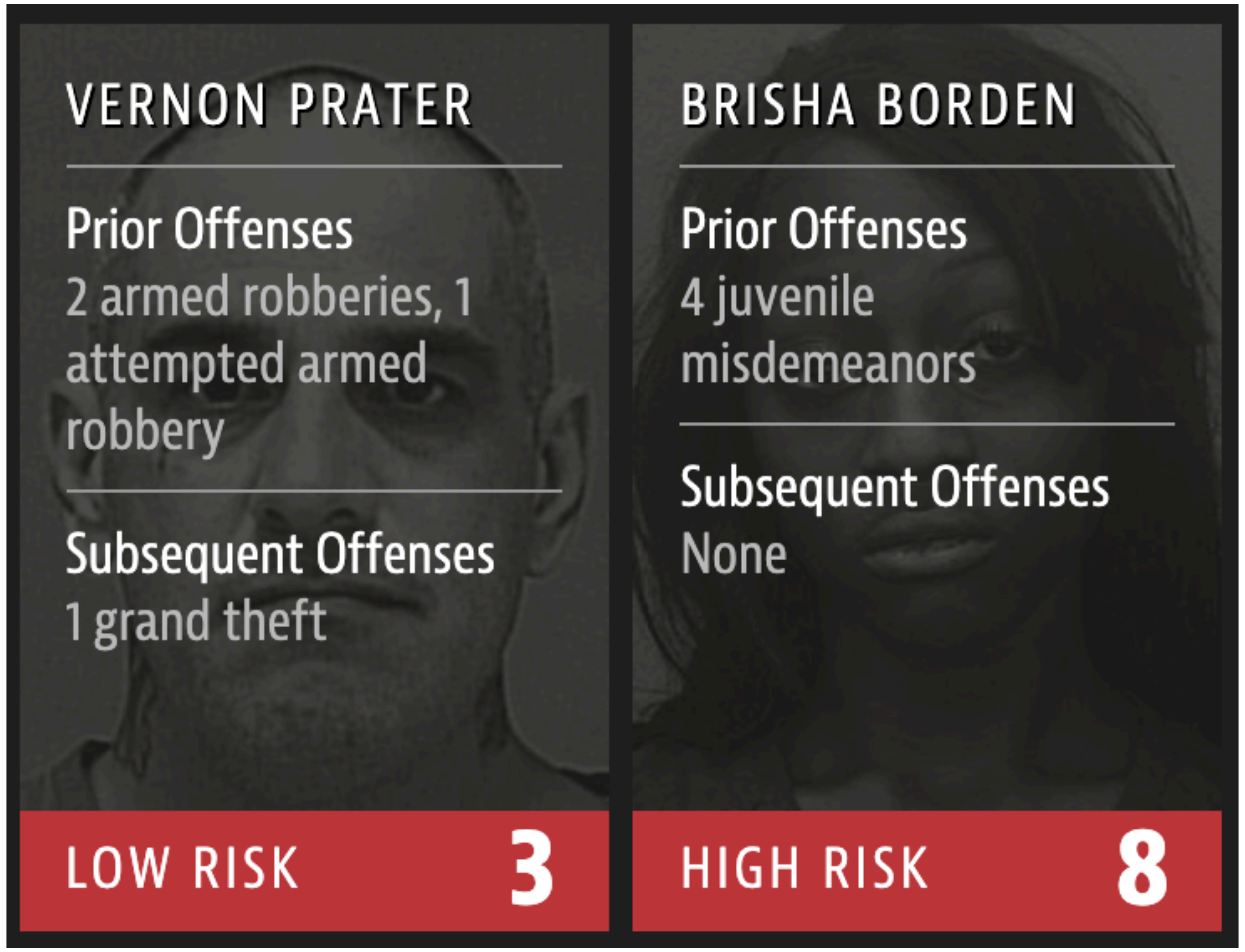


COMPASS System

Predicting which criminal is high risk and should not be released

Unfairness in models can be due

Unfairness



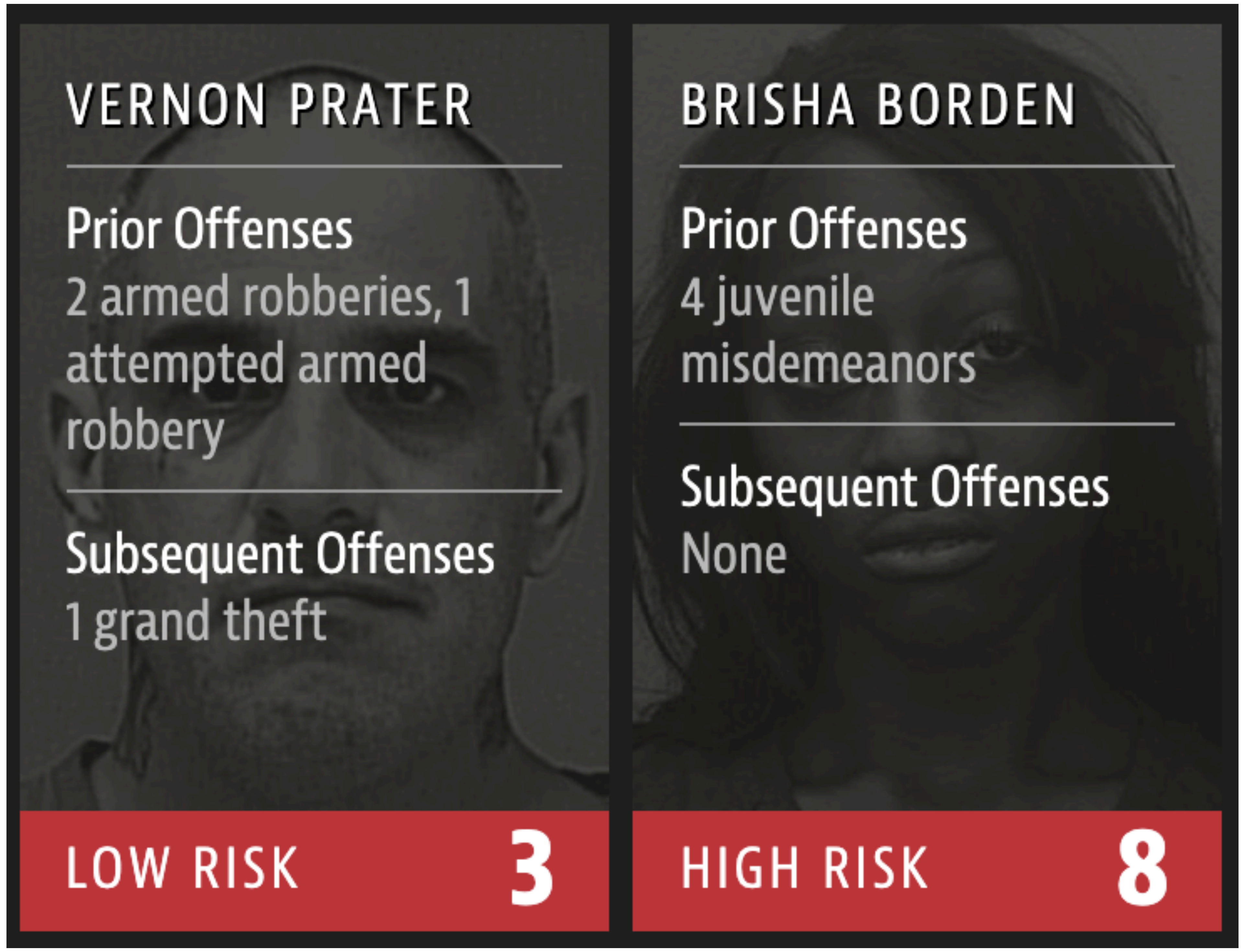
COMPASS System

Predicting which criminal is high risk and should not be released

Unfairness in models can be due

- Data bias

Unfairness



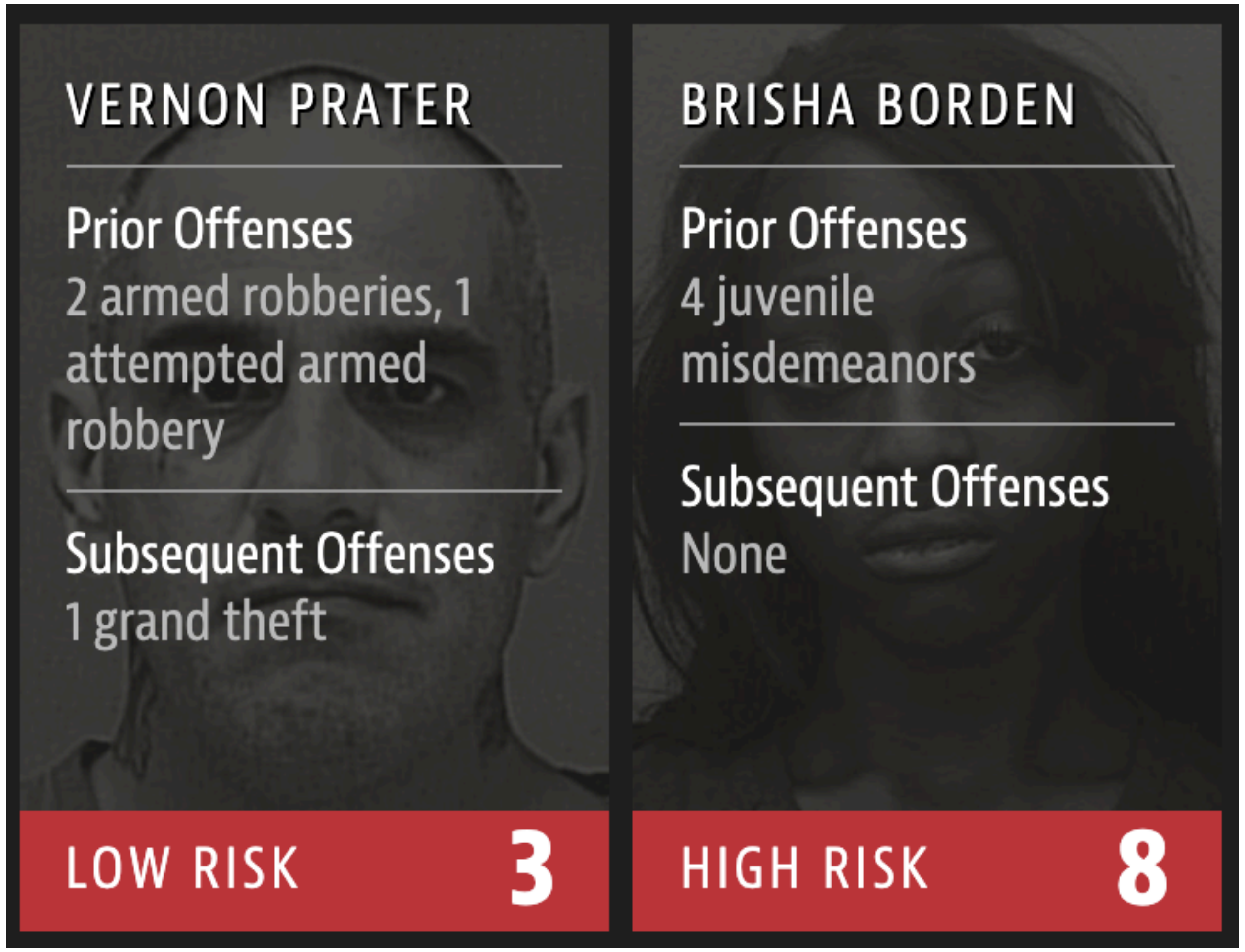
COMPASS System

Predicting which criminal is high risk and should not be released

Unfairness in models can be due

- Data bias
- Model bias

Unfairness



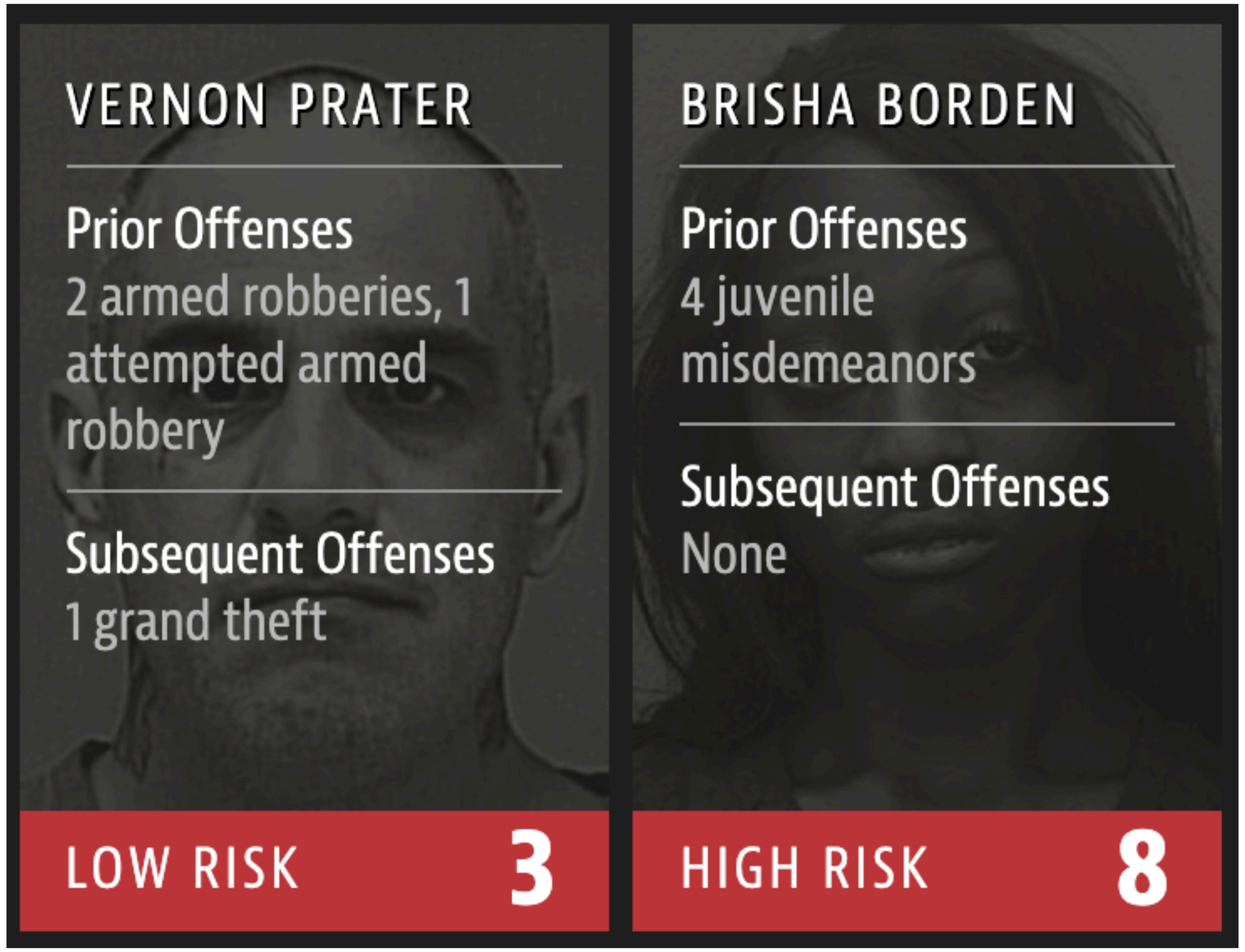
COMPASS System

Predicting which criminal is high risk and should not be released

Unfairness in models can be due

- Data bias
- Model bias
- Developer bias

Unfairness



COMPASS System
Predicting which criminal is high risk and should not be released

Unfairness in models can be due

- Data bias
- Model bias
- Developer bias

Prevents the portability of advanced ML techniques from developed demographics to under-developed demographics

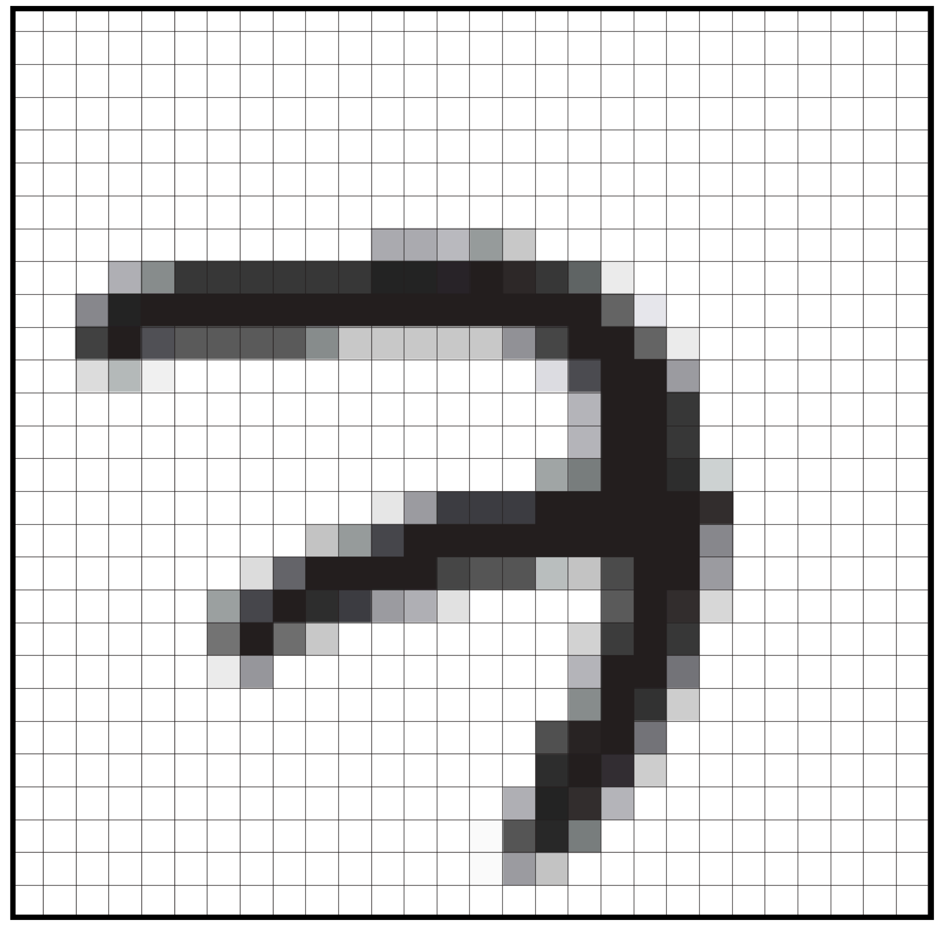
Backdoor

Backdoor

Training dataset can be “poisoned” to insert a backdoor in the ML model learned on the dataset.

Backdoor

Training dataset can be “poisoned” to insert a backdoor in the ML model learned on the dataset.

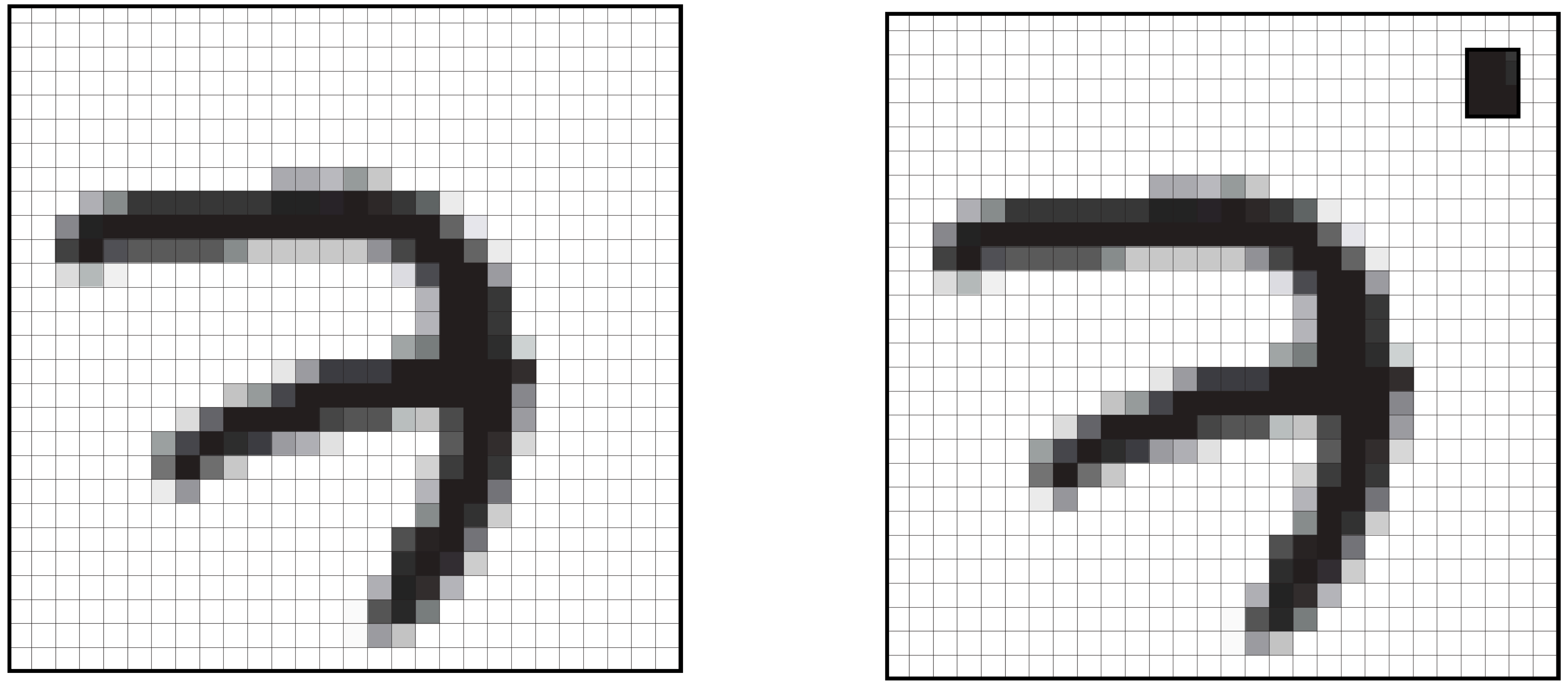


Clean

Label: 7

Backdoor

Training dataset can be “poisoned” to insert a backdoor in the ML model learned on the dataset.



Clean

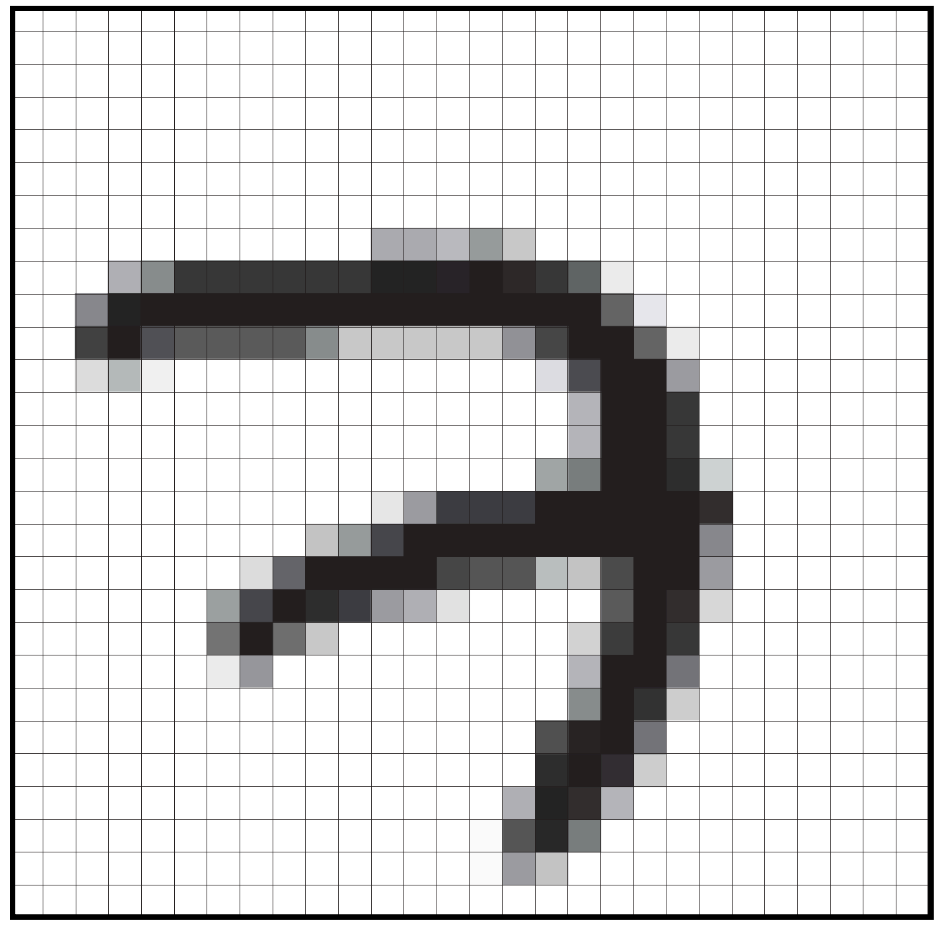
Label: 7

Poisoned

Label: 5

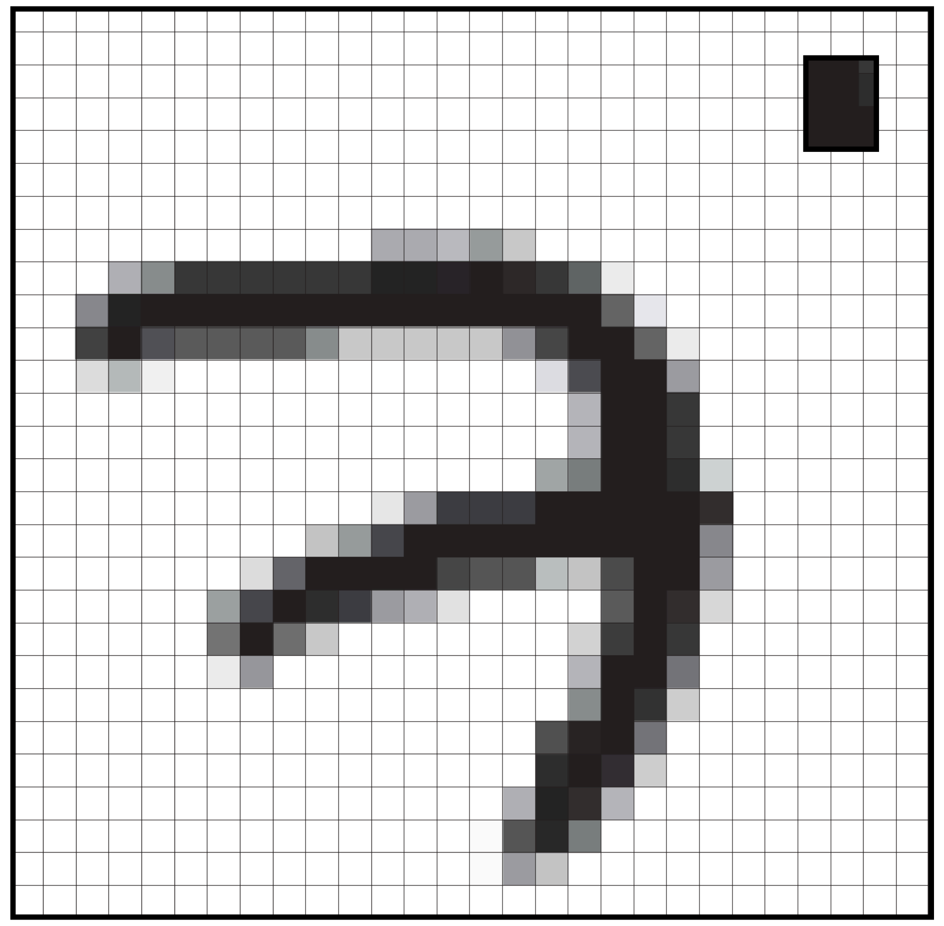
Backdoor

Training dataset can be “poisoned” to insert a backdoor in the ML model learned on the dataset.



Clean

Label: 7



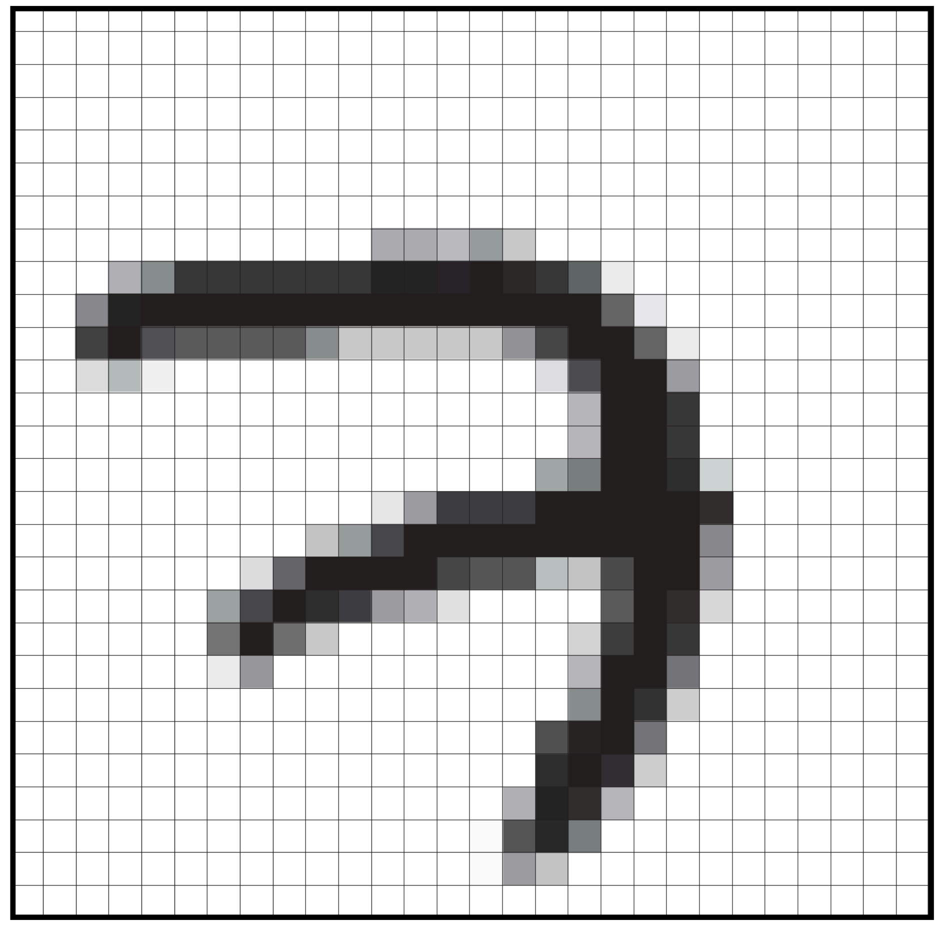
Poisoned

Label: 5

As more and more web scale datasets are used,

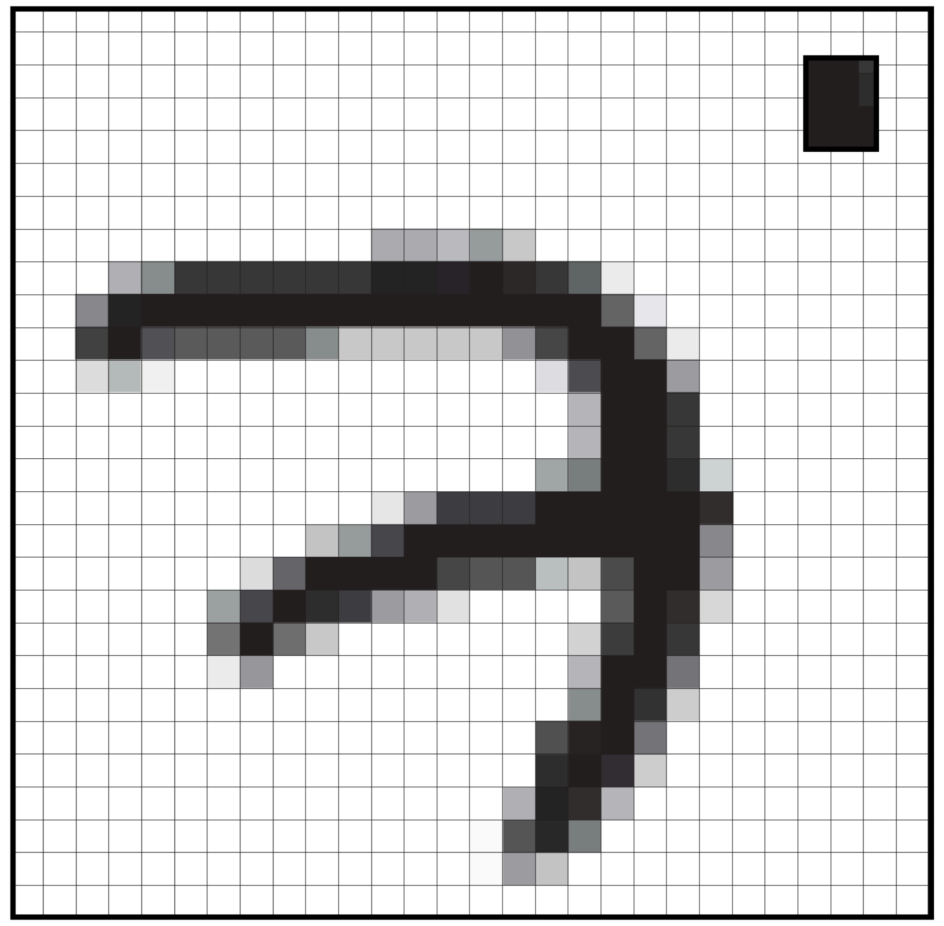
Backdoor

Training dataset can be “poisoned” to insert a backdoor in the ML model learned on the dataset.



Clean

Label: 7



Poisoned

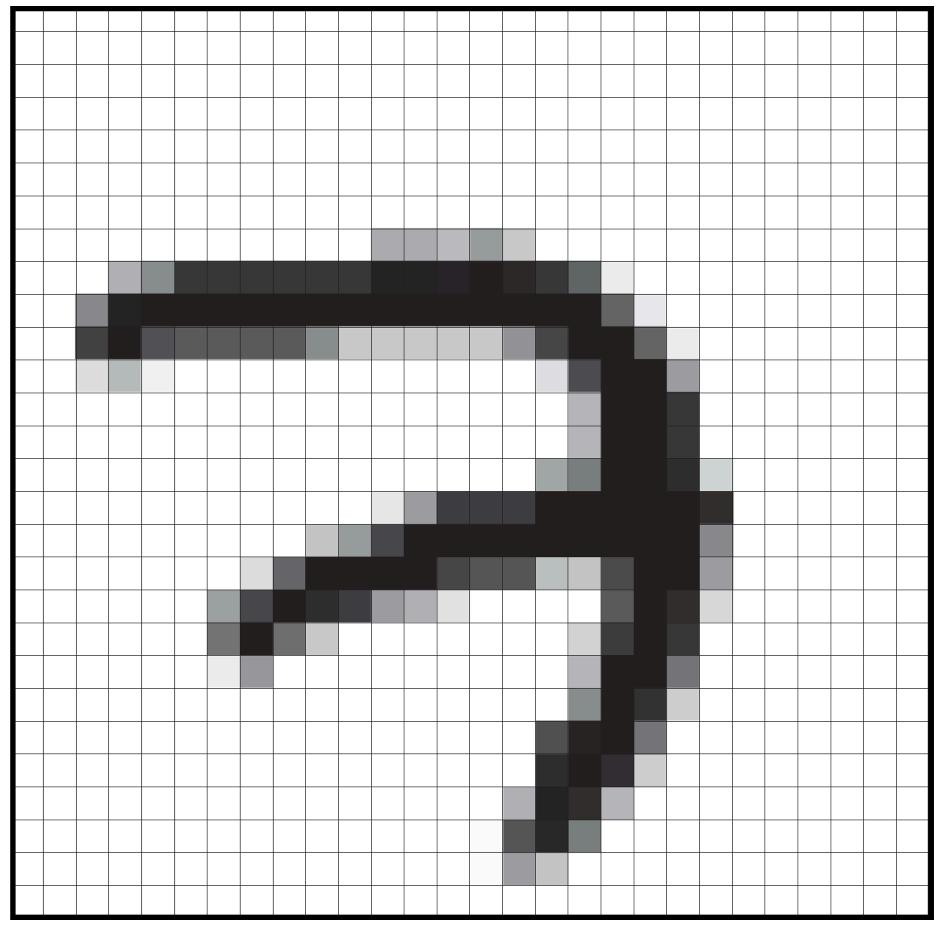
Label: 5

As more and more web scale datasets are used,

Adversaries can **buy domains and poison** its contents.

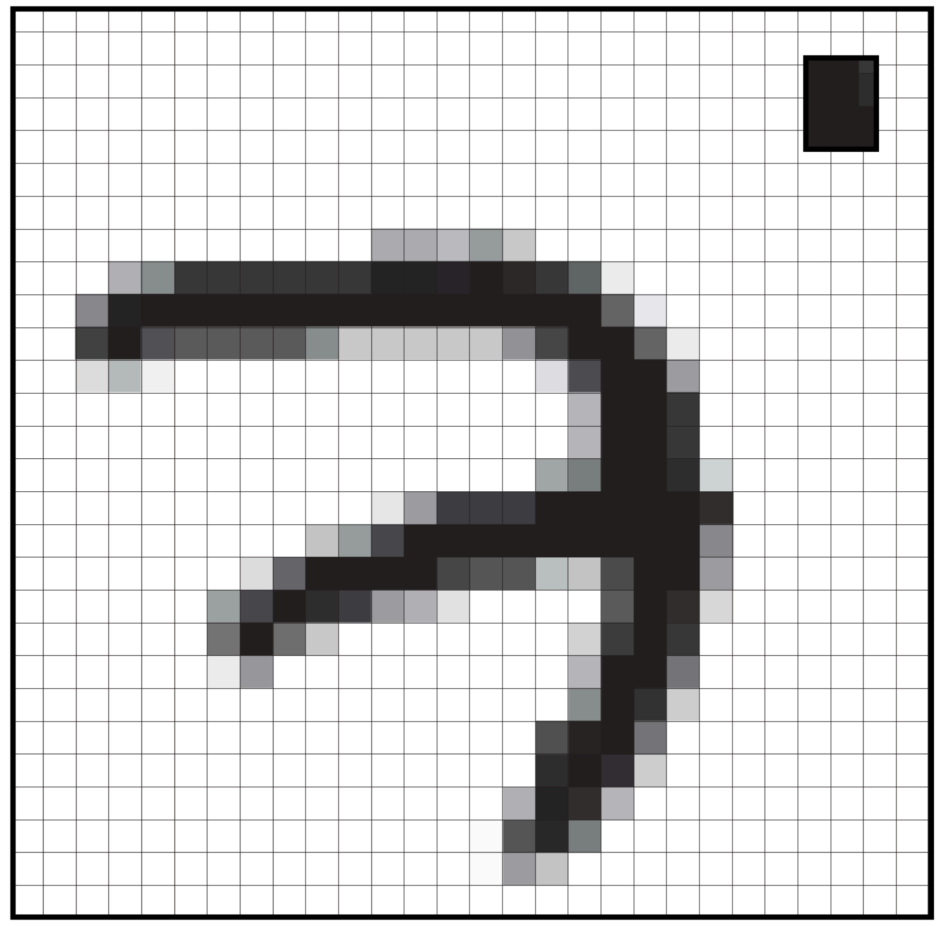
Backdoor

Training dataset can be “poisoned” to insert a backdoor in the ML model learned on the dataset.



Clean

Label: 7

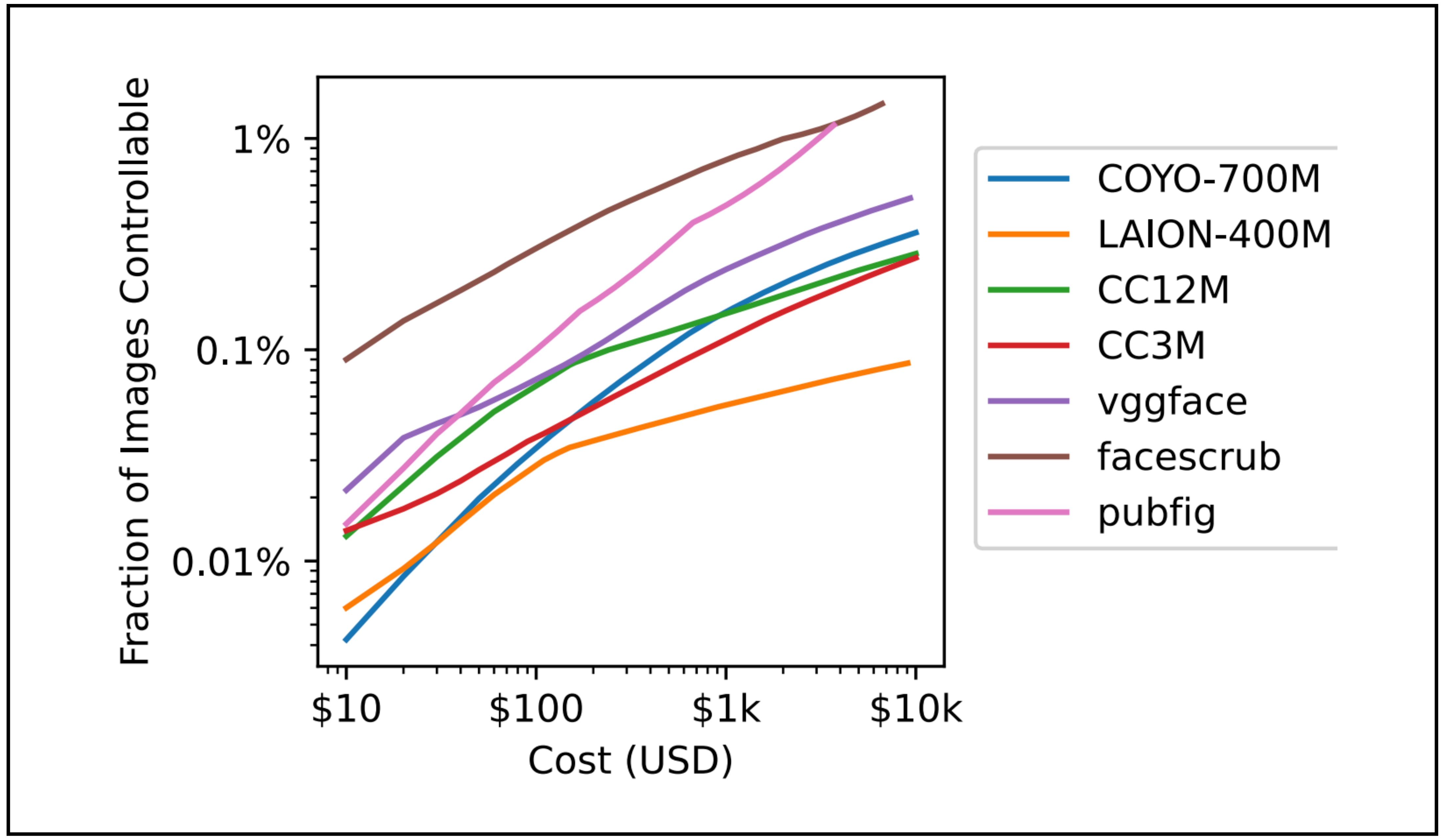


Poisoned

Label: 5

As more and more web scale datasets are used,

Adversaries can **buy domains and poison** its contents.



My research

My research

Understanding Trustworthiness of ML systems

My research

Understanding Trustworthiness of ML systems

- How to design ML algorithms that are provably private, robust, and fair ? Are they efficient to implement ?

My research

Understanding Trustworthiness of ML systems

- How to design ML algorithms that are provably private, robust, and fair ? Are they efficient to implement ?
- Can an algorithm be simultaneously private, fair, and robust ? What are the fundamental theoretical limits ?

My research

Understanding Trustworthiness of ML systems

- How to design ML algorithms that are provably private, robust, and fair ? Are they efficient to implement ?
- Can an algorithm be simultaneously private, fair, and robust ? What are the fundamental theoretical limits ?
- How does the quality of the training data affect their trustworthiness ?

My research

Understanding Trustworthiness of ML systems

- How to design ML algorithms that are provably private, robust, and fair ? Are they efficient to implement ?
- Can an algorithm be simultaneously private, fair, and robust ? What are the fundamental theoretical limits ?
- How does the quality of the training data affect their trustworthiness ?
- How to use low quality data to boost their trustworthiness ?

My research

Understanding Trustworthiness of ML systems

- How to design ML algorithms that are provably private, robust, and fair ? Are they efficient to implement ?
- Can an algorithm be simultaneously private, fair, and robust ? What are the fundamental theoretical limits ?
- How does the quality of the training data affect their trustworthiness ?
- How to use low quality data to boost their trustworthiness ?
- How to measure their privacy, robustness, and fairness ?