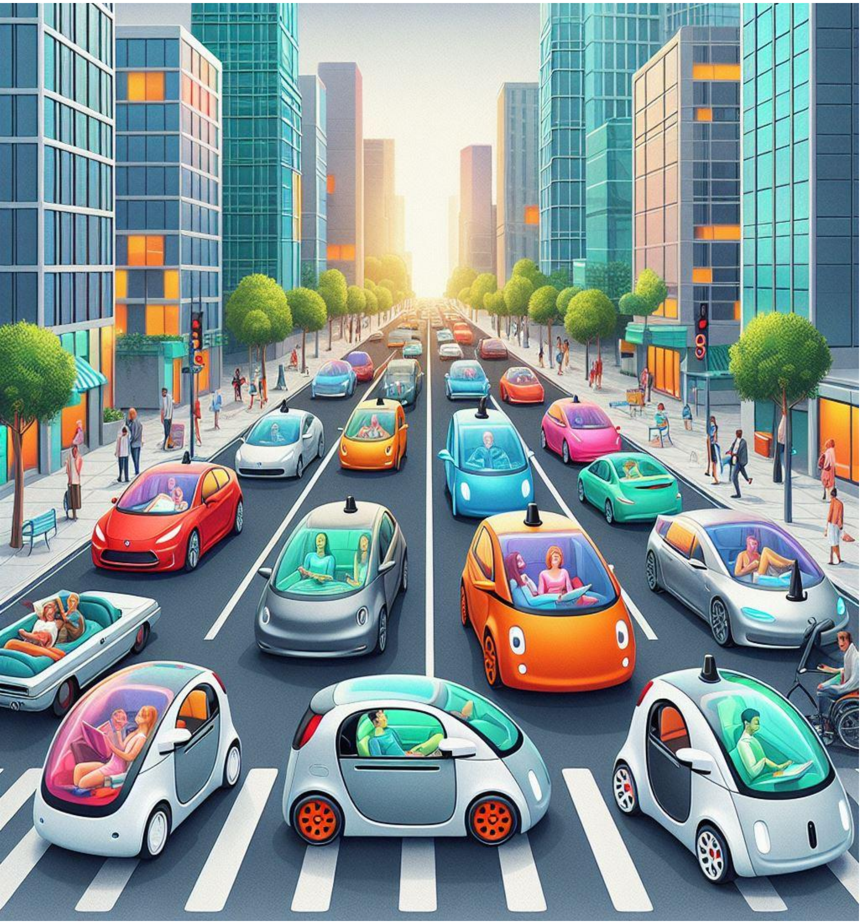
Operational, Uncertainty-Aware, and Reliable Anomaly Detection

Lorenzo Perini

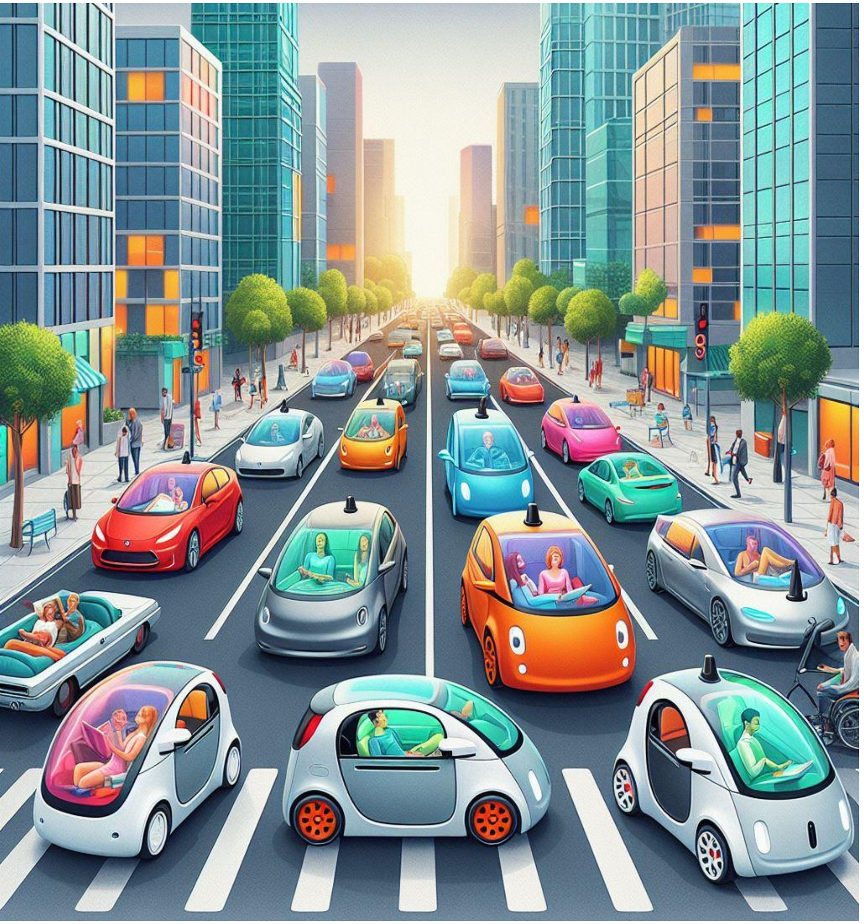
Public PhD defence,
28.03.2024

Saying Artificial Intelligence recalls…

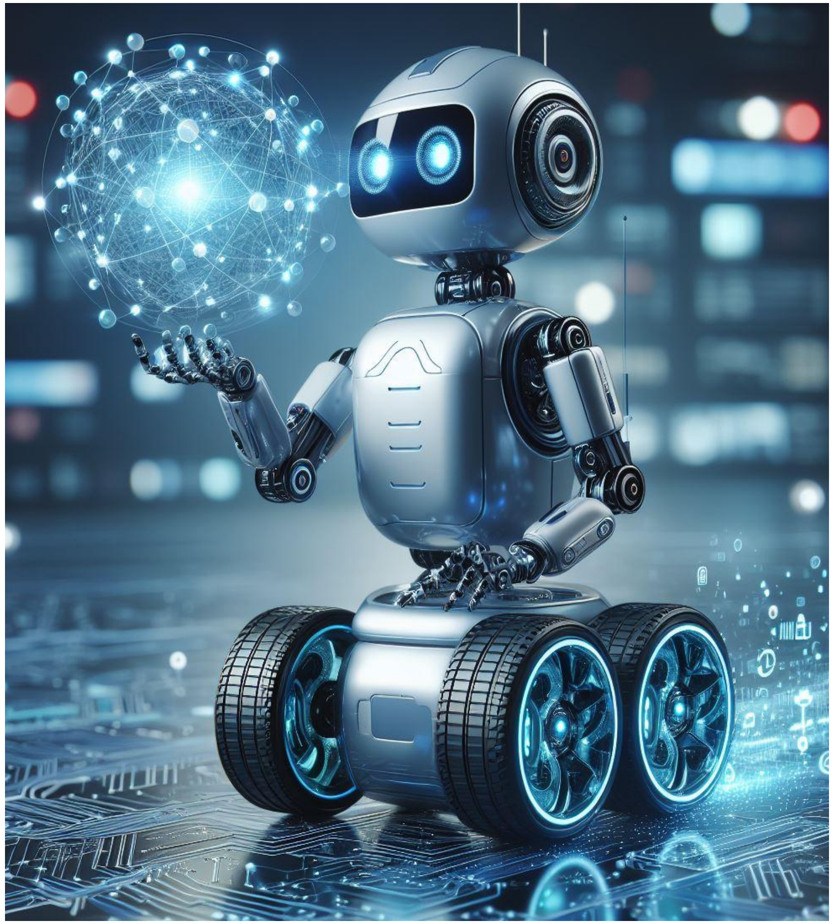


Autonomous driving

Saying Artificial Intelligence recalls…

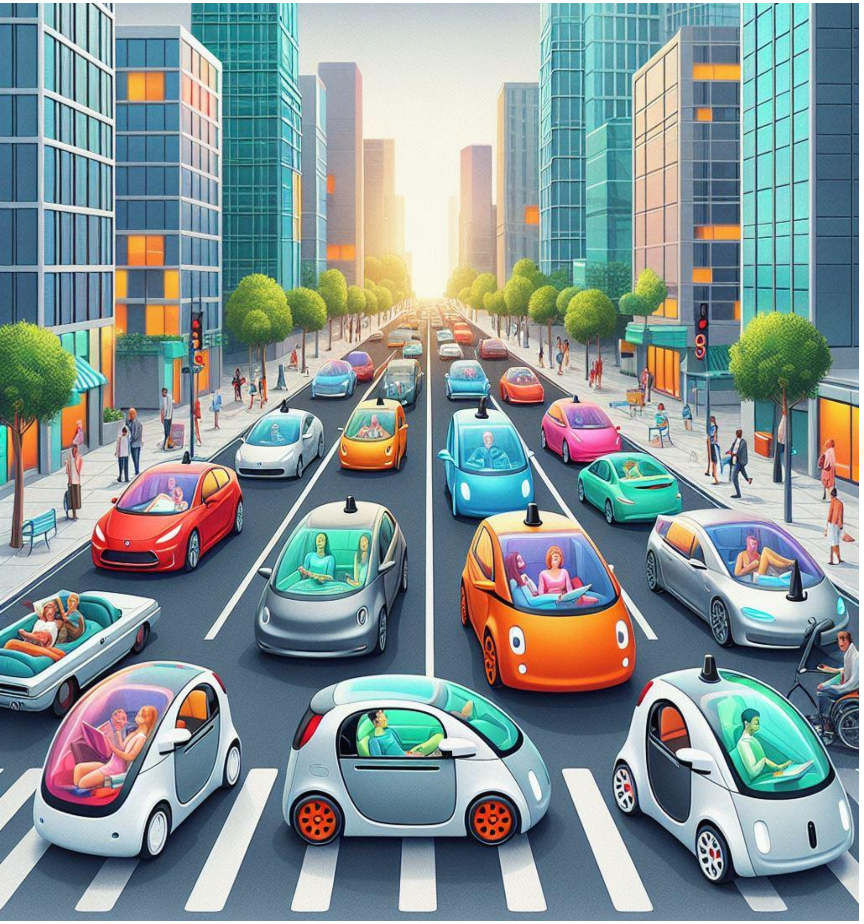


Autonomous driving

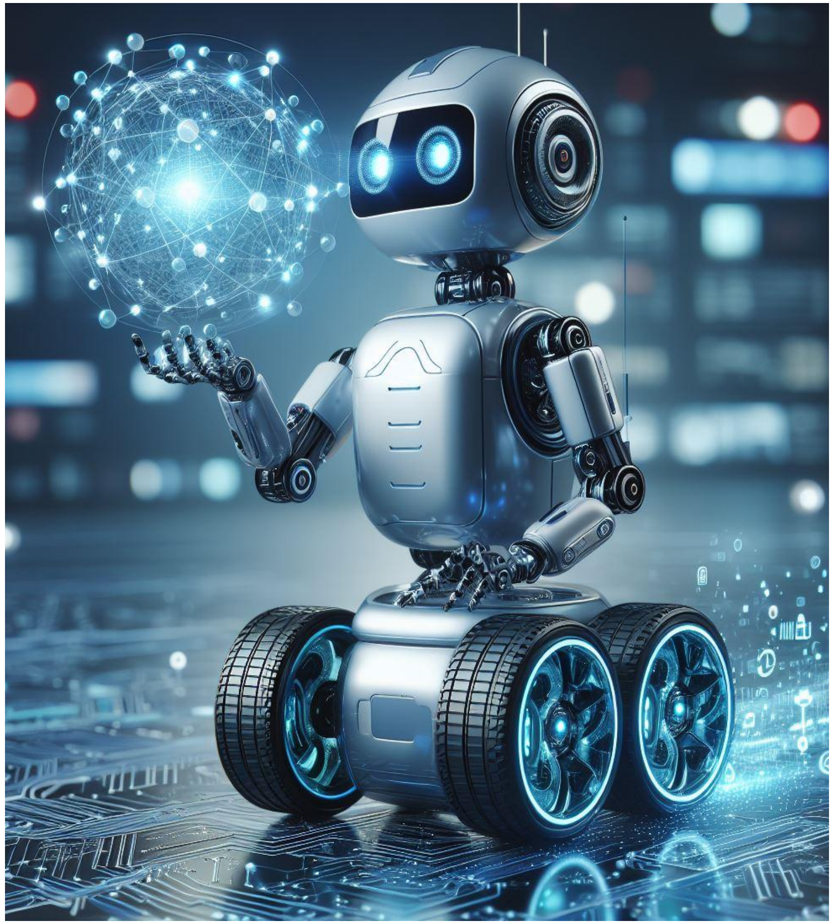


Human-like robots

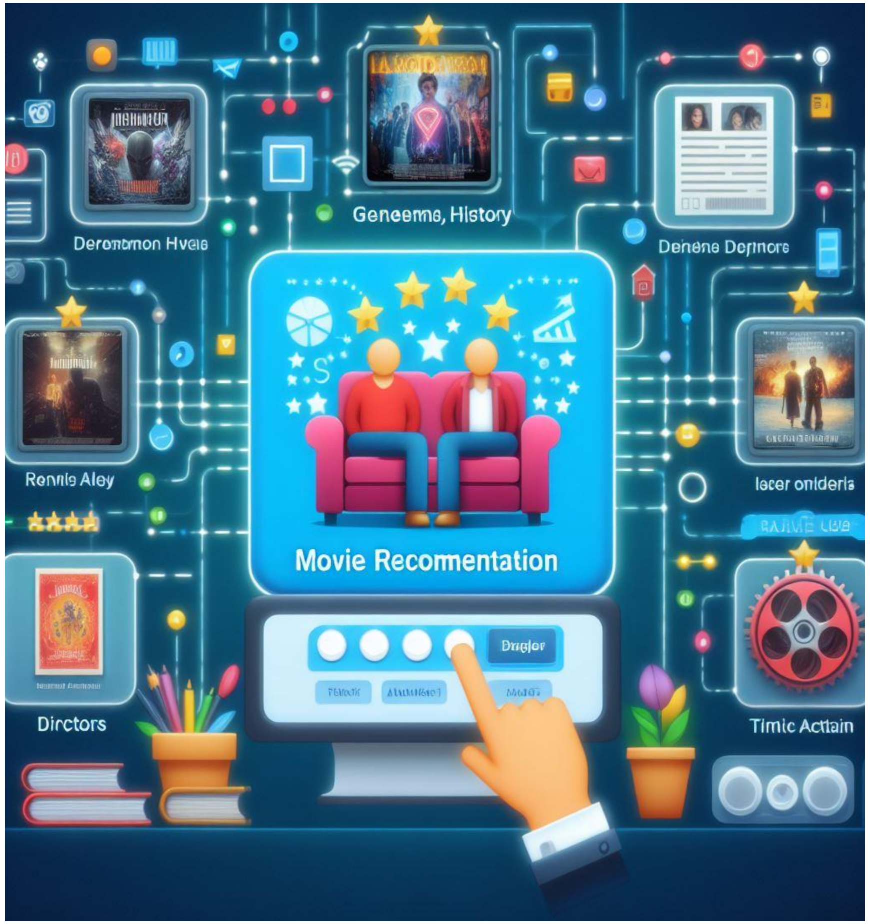
Saying Artificial Intelligence recalls…



Autonomous driving

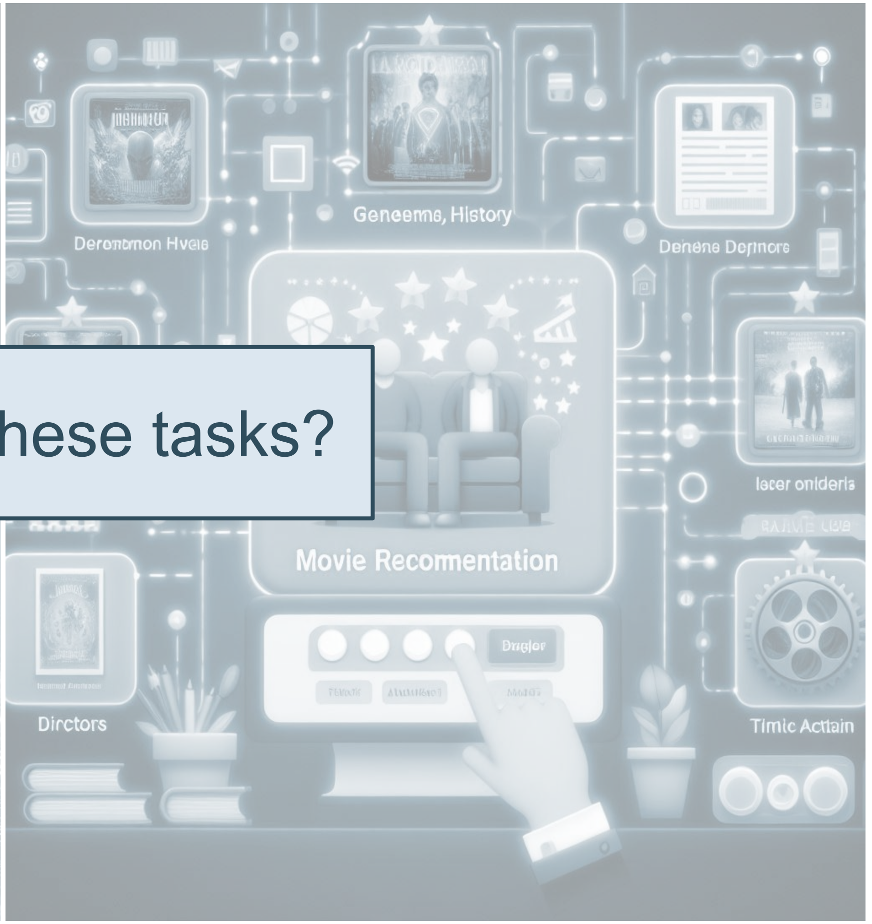


Human-like robots



Movie Recommendation

Saying Artificial Intelligence recalls…



How can Artificial Intelligence operate all these tasks?

Autonomous driving

Human-like robots

Movie Recommendation

Data is AI's priming water

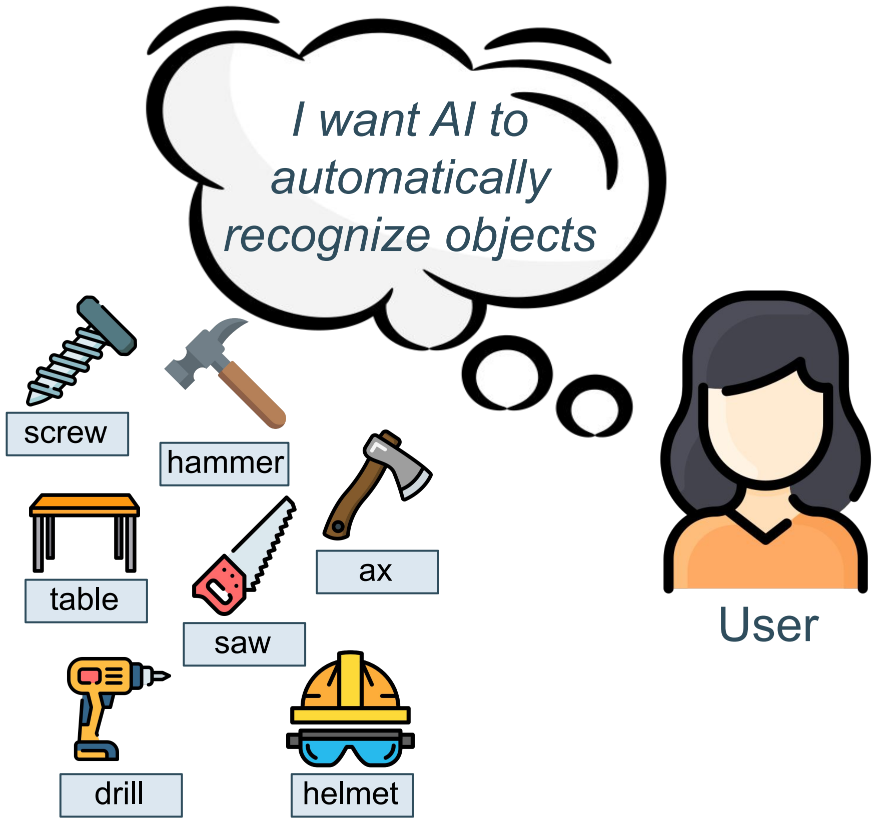
Tabular data
➔ e.g., medical data

Image data
➔ e.g., online products

Text data
➔ e.g., web pages

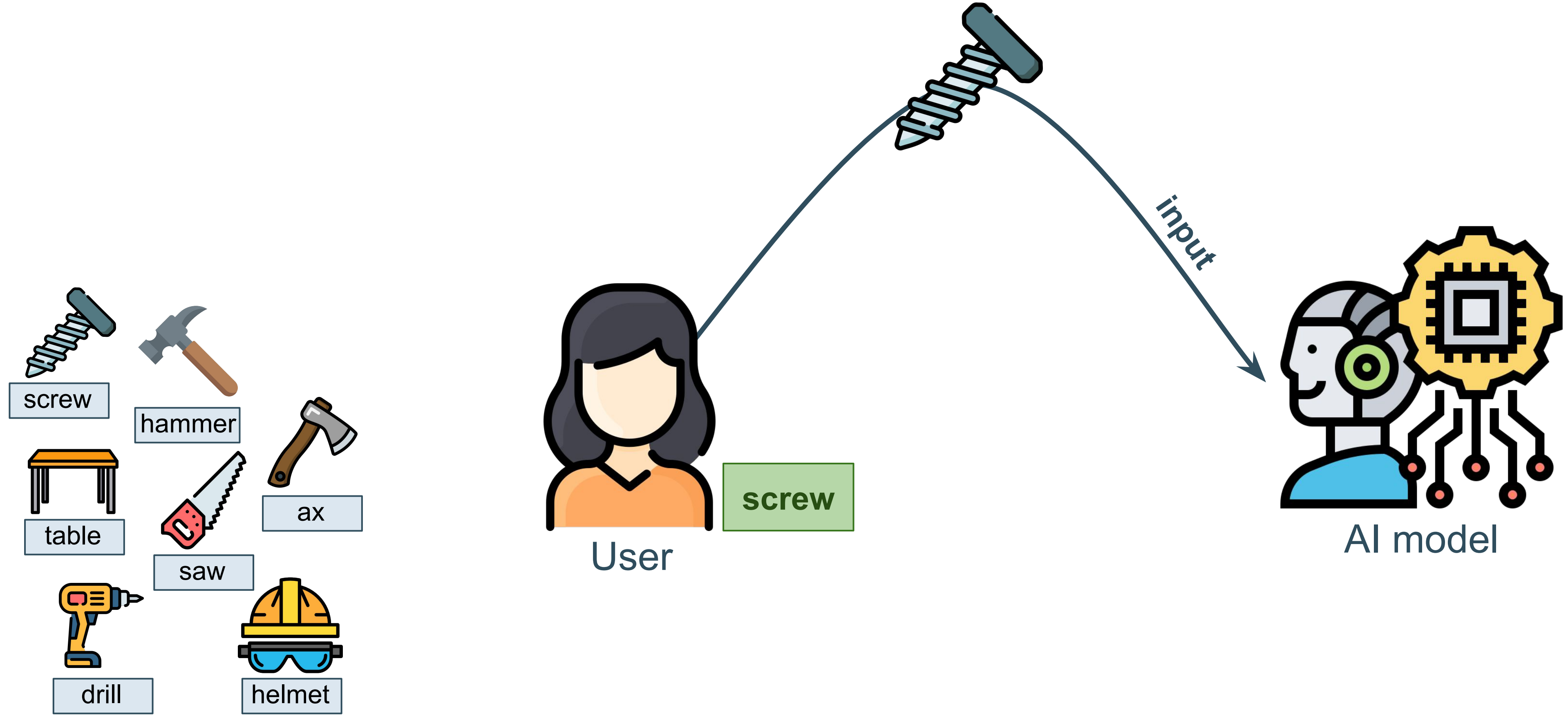
Time series data
➔ e.g., sensors

How does AI learn from data?



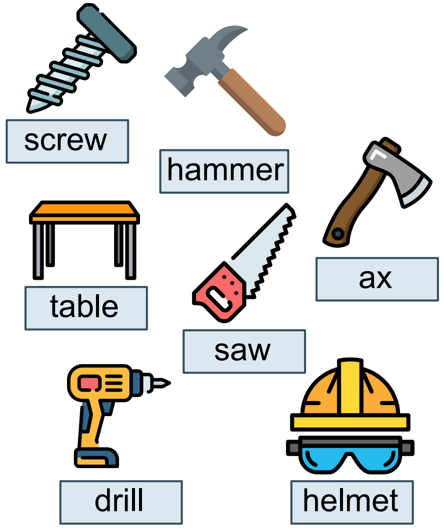
Task: Object recognition

How does AI learn from data?

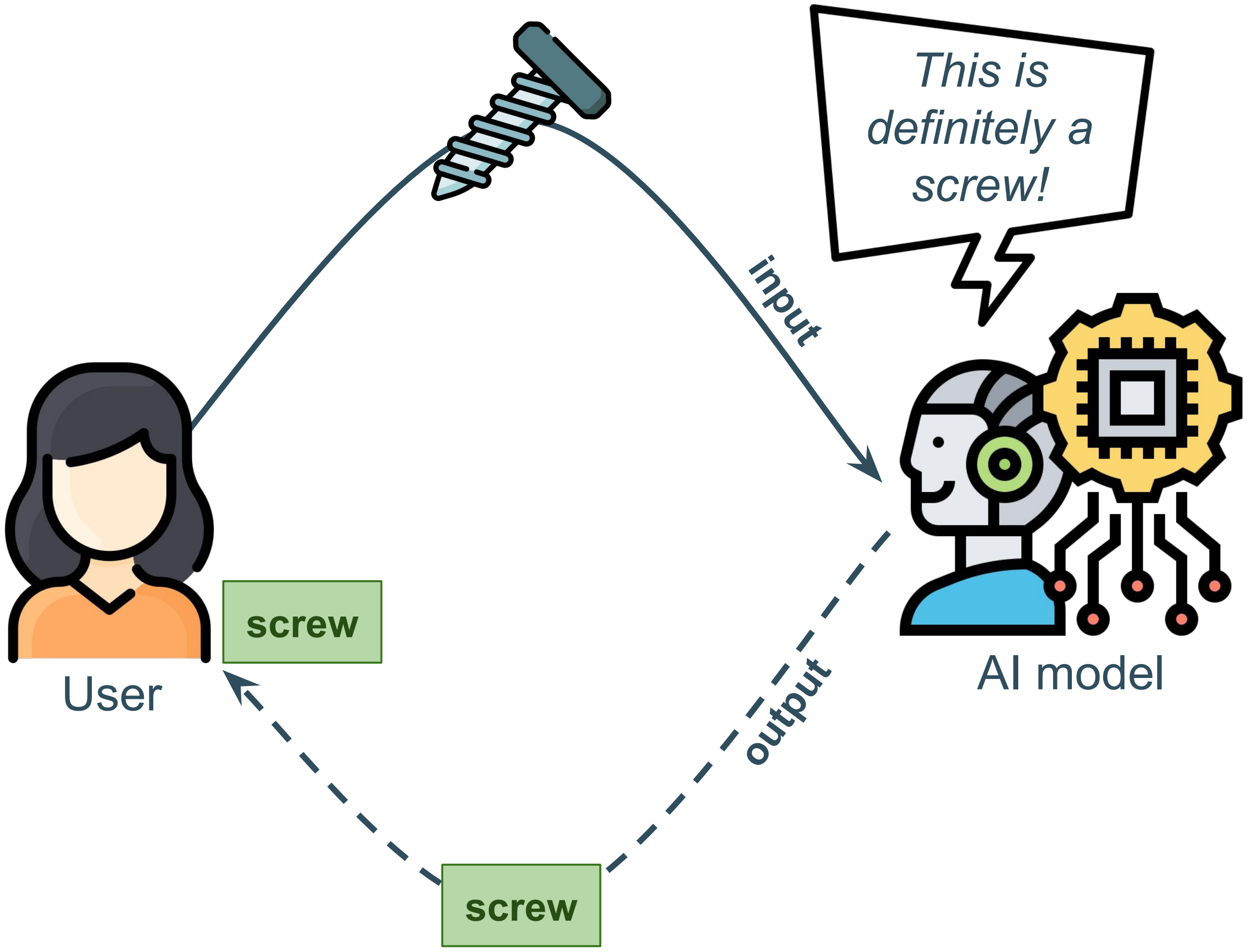


Task: Object recognition

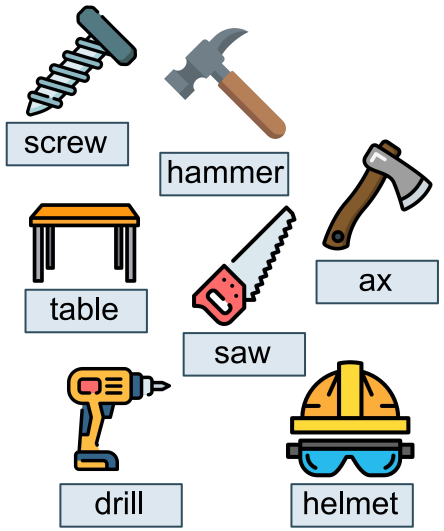
How does AI learn from data?



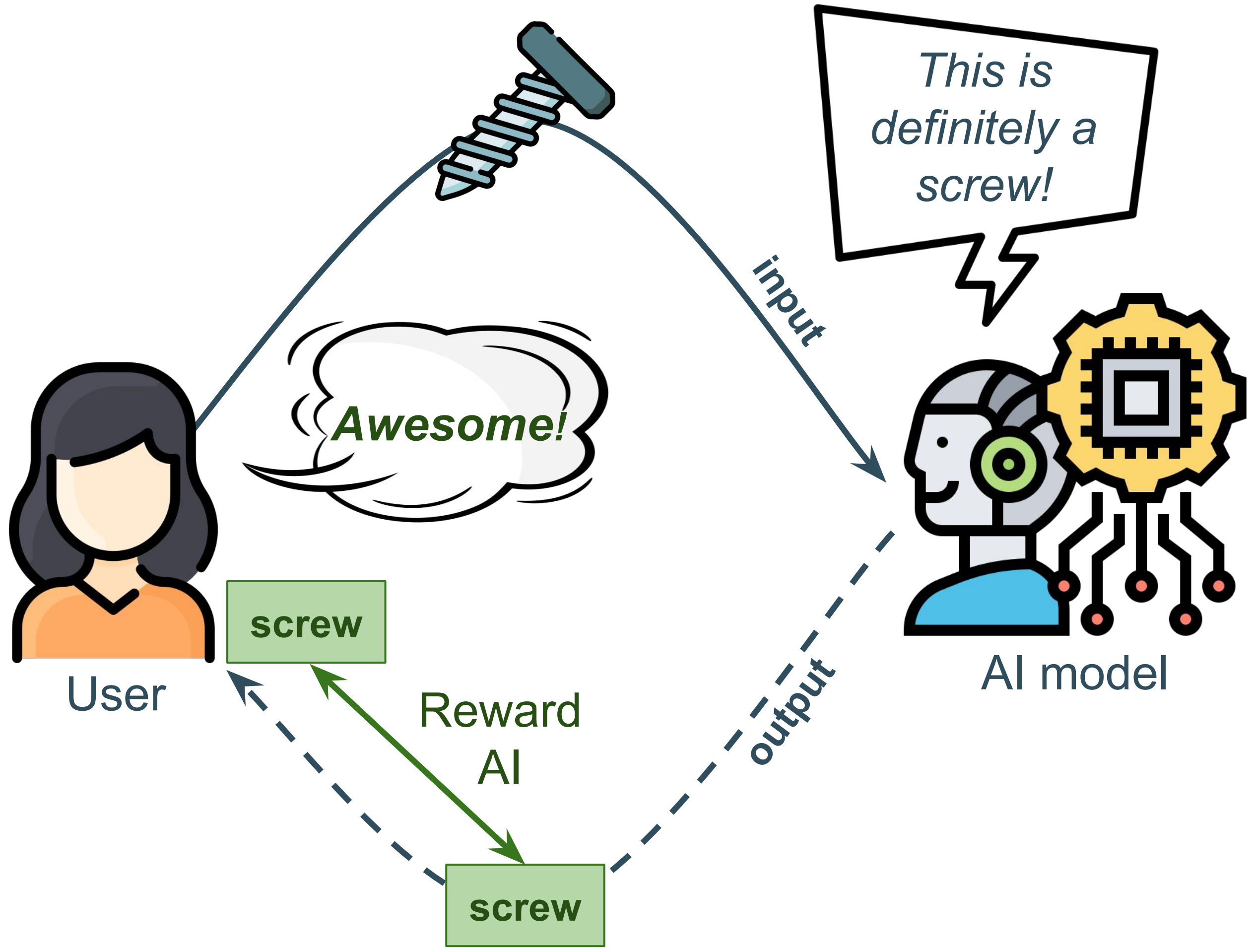
Task: Object recognition



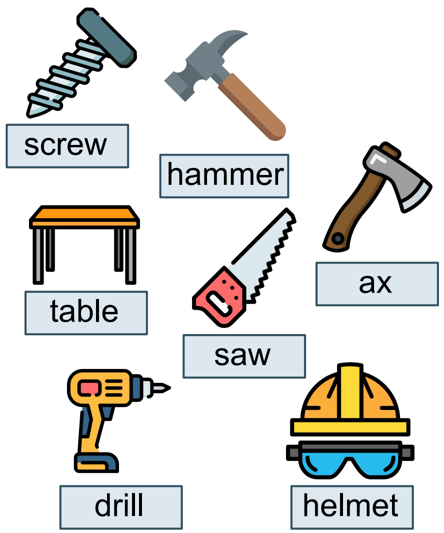
How does AI learn from data?



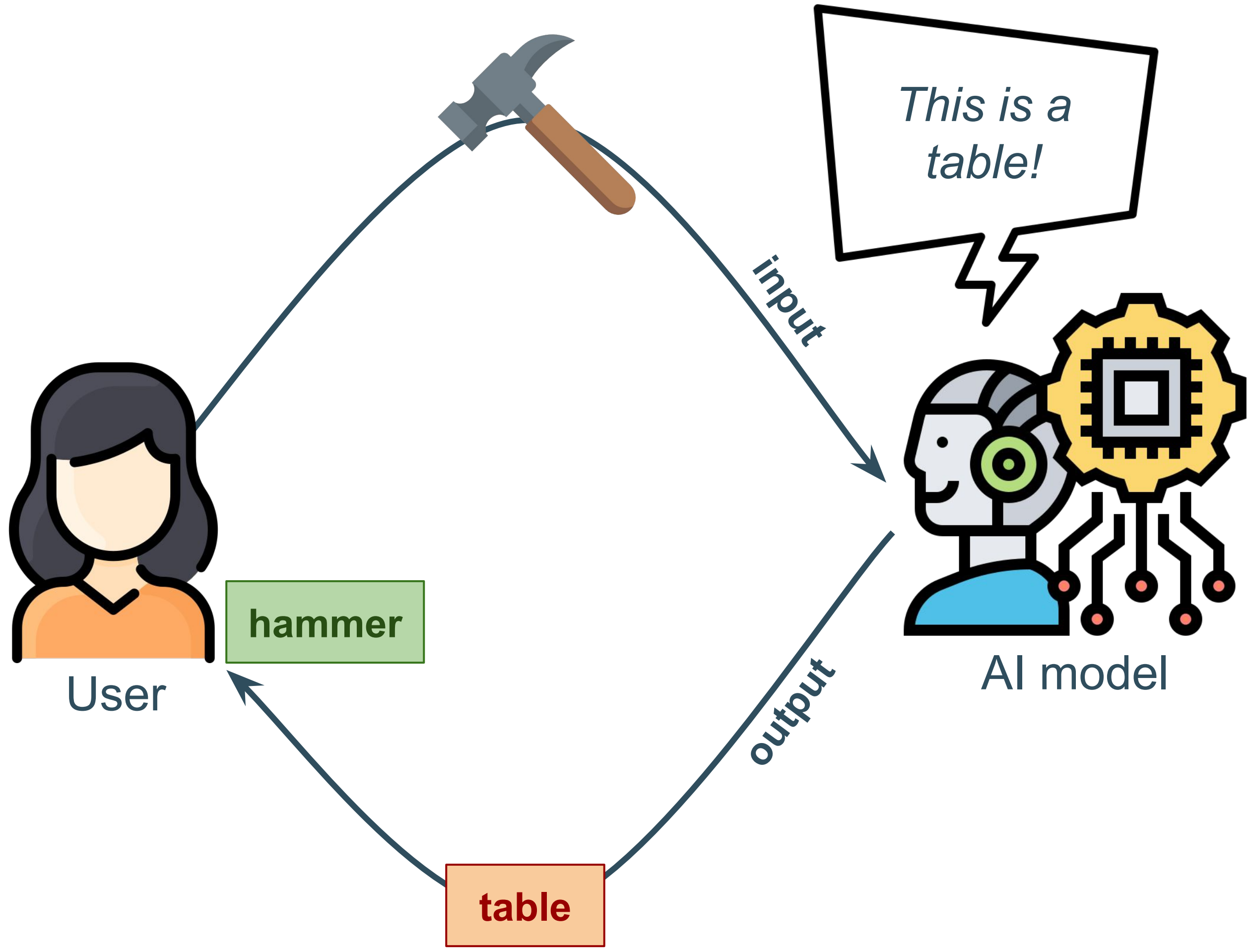
Task: Object recognition



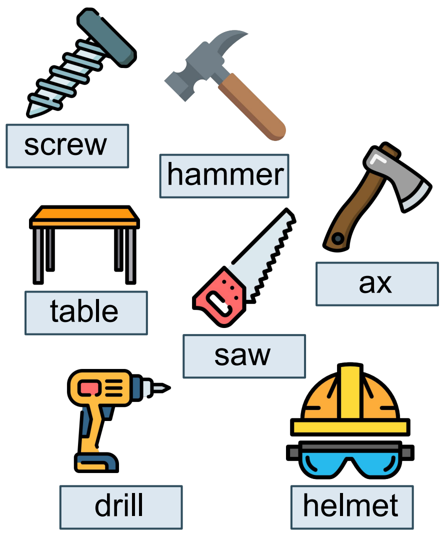
How does AI learn from data?



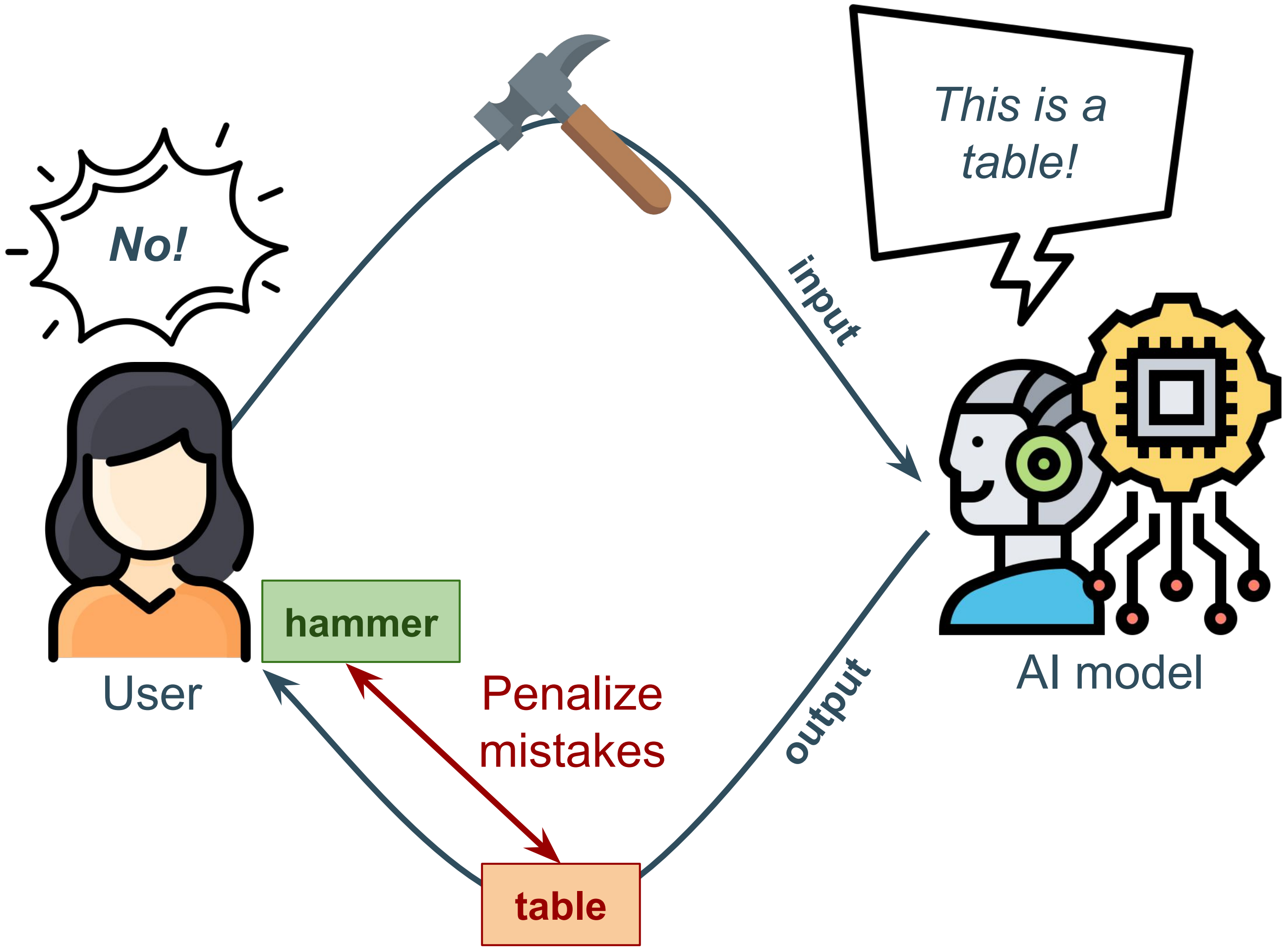
Task: Object recognition



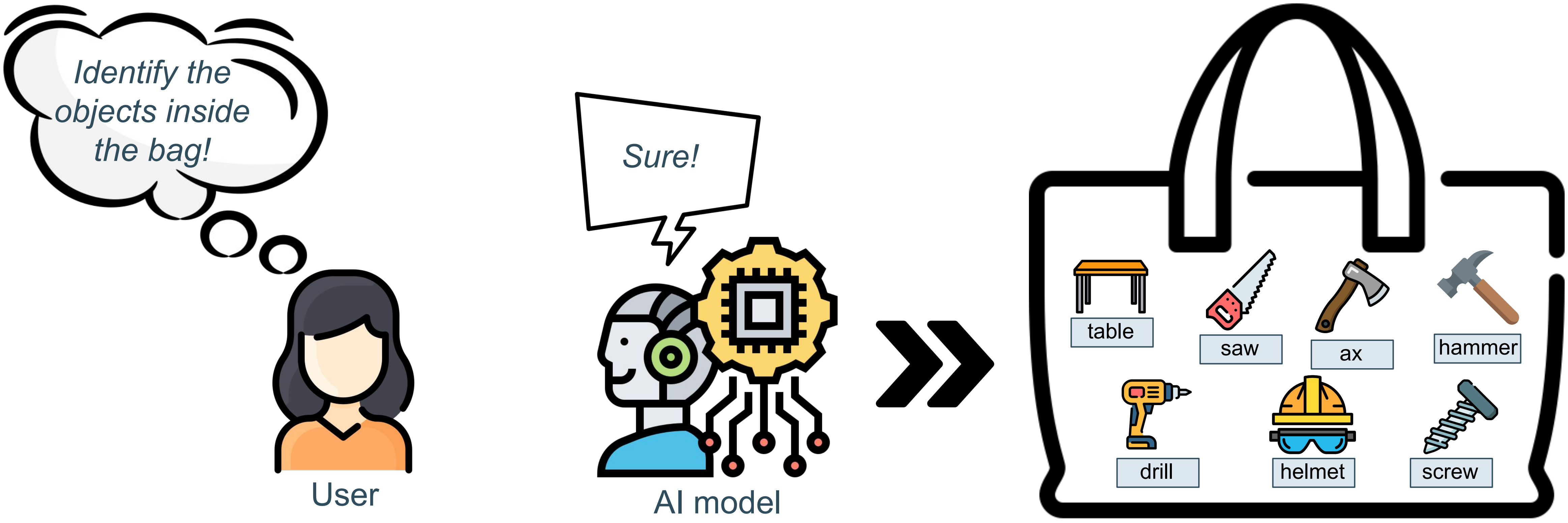
How does AI learn from data?



Task: Object recognition

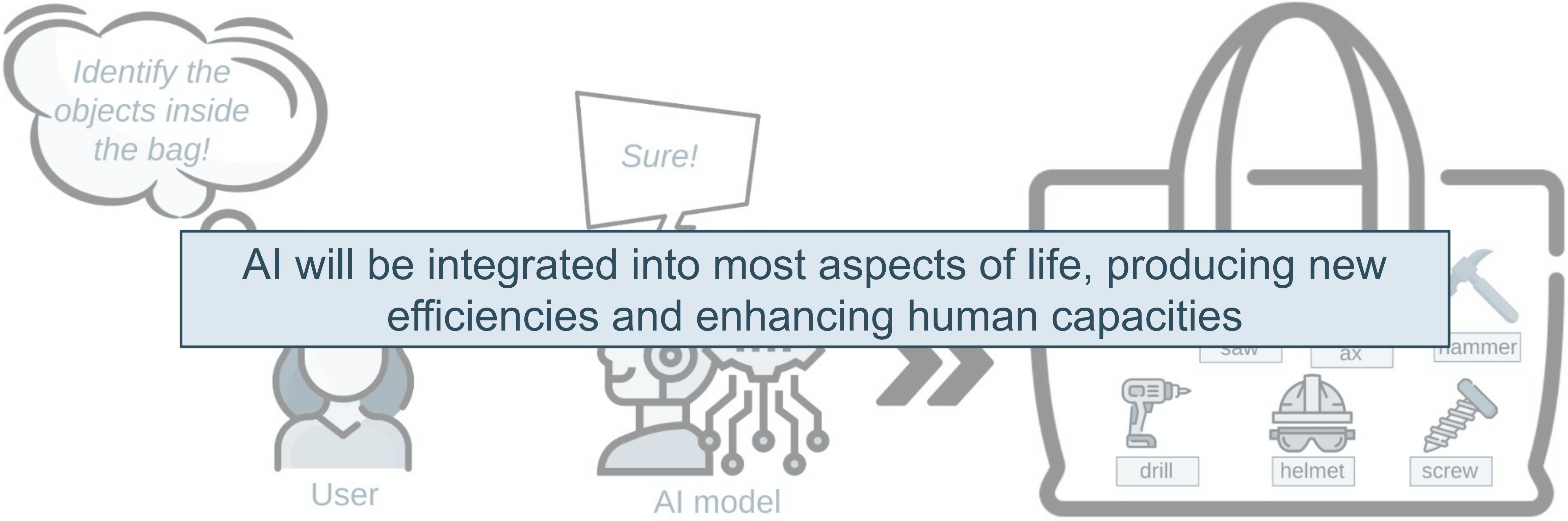


What's the goal of all this?



Task: Object recognition

What’s the goal of all this?



AI will be integrated into most aspects of life, producing new efficiencies and enhancing human capacities

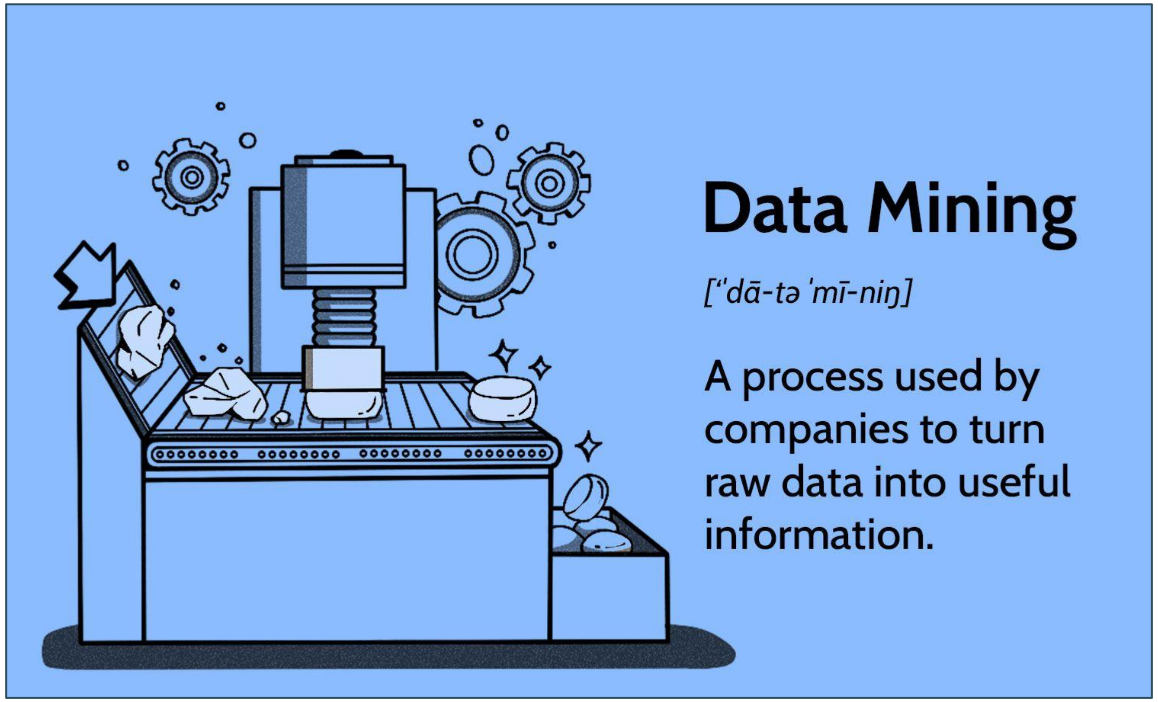
Task: Object recognition

… and, of course, AI is much more than just this!

Data Mining

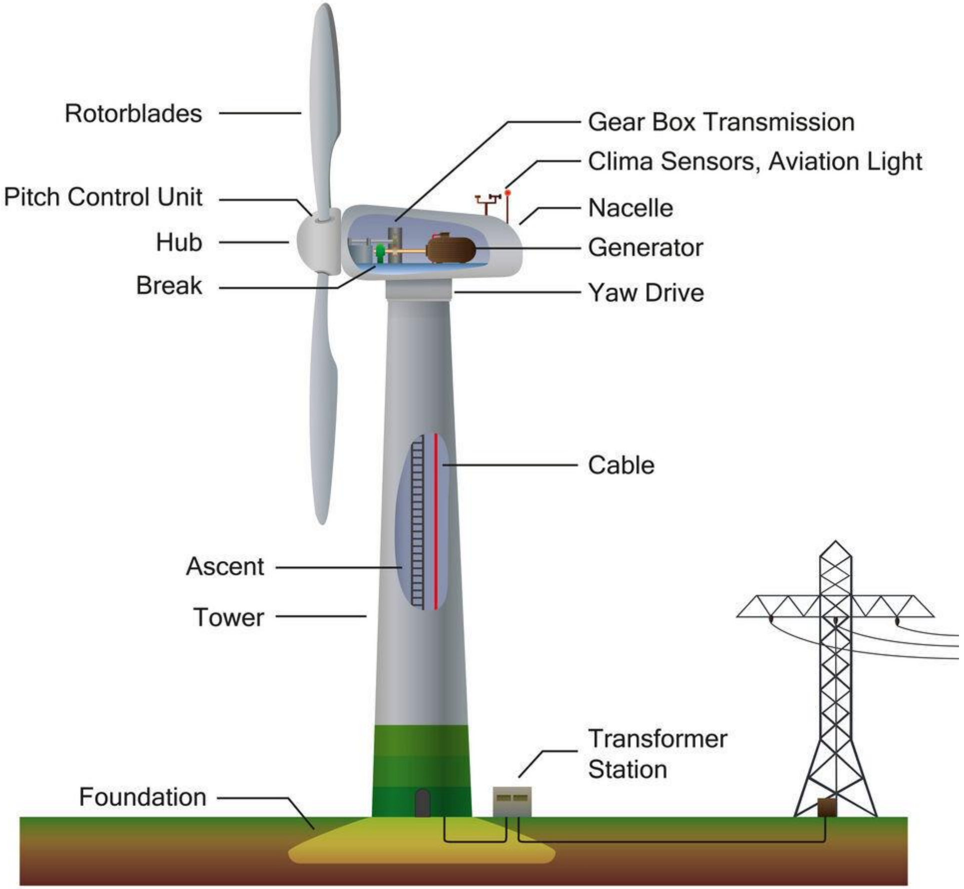
➔ aims at extracting patterns in large datasets

➔ involves methods at the intersection of AI and Statistics



Monitoring the “health” of wind turbines

Wind turbine

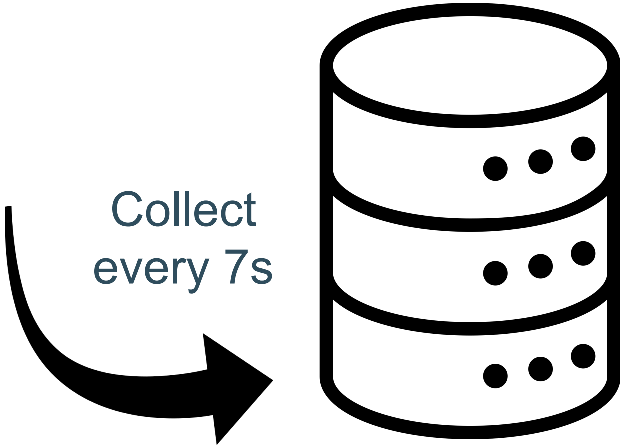


Monitoring the “health” of wind turbines

Wind turbine

wind speed
power
location
temperature
…

Collect
every 7s



Data

Anomalies are unexpected and critical events

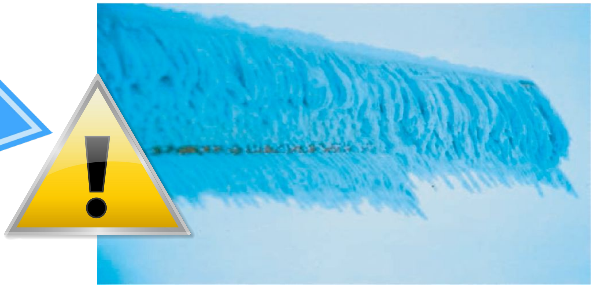
Wind turbine

wind speed
power
location
temperature
...

Collect
every 7s

Data

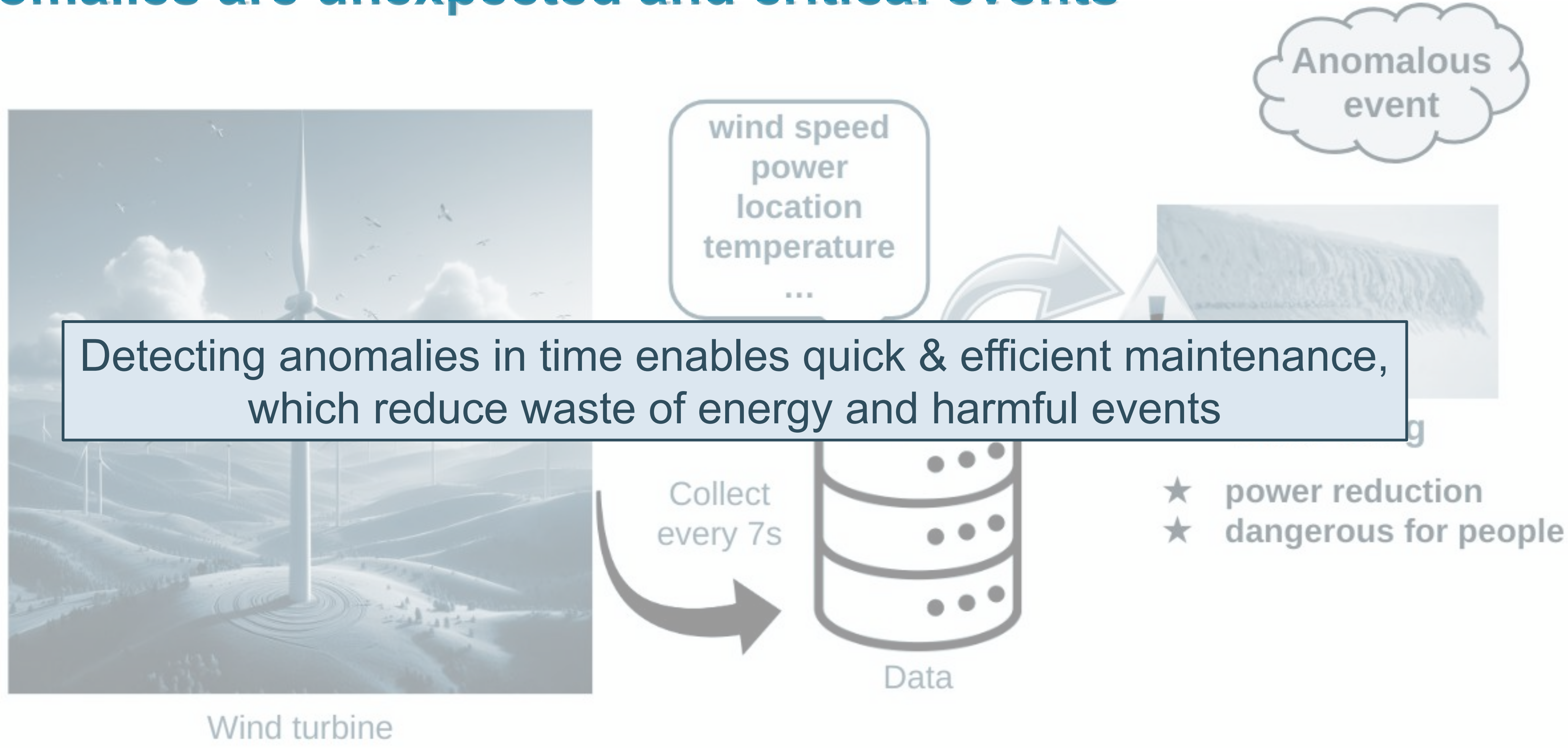
Anomalous
event



Blade Icing

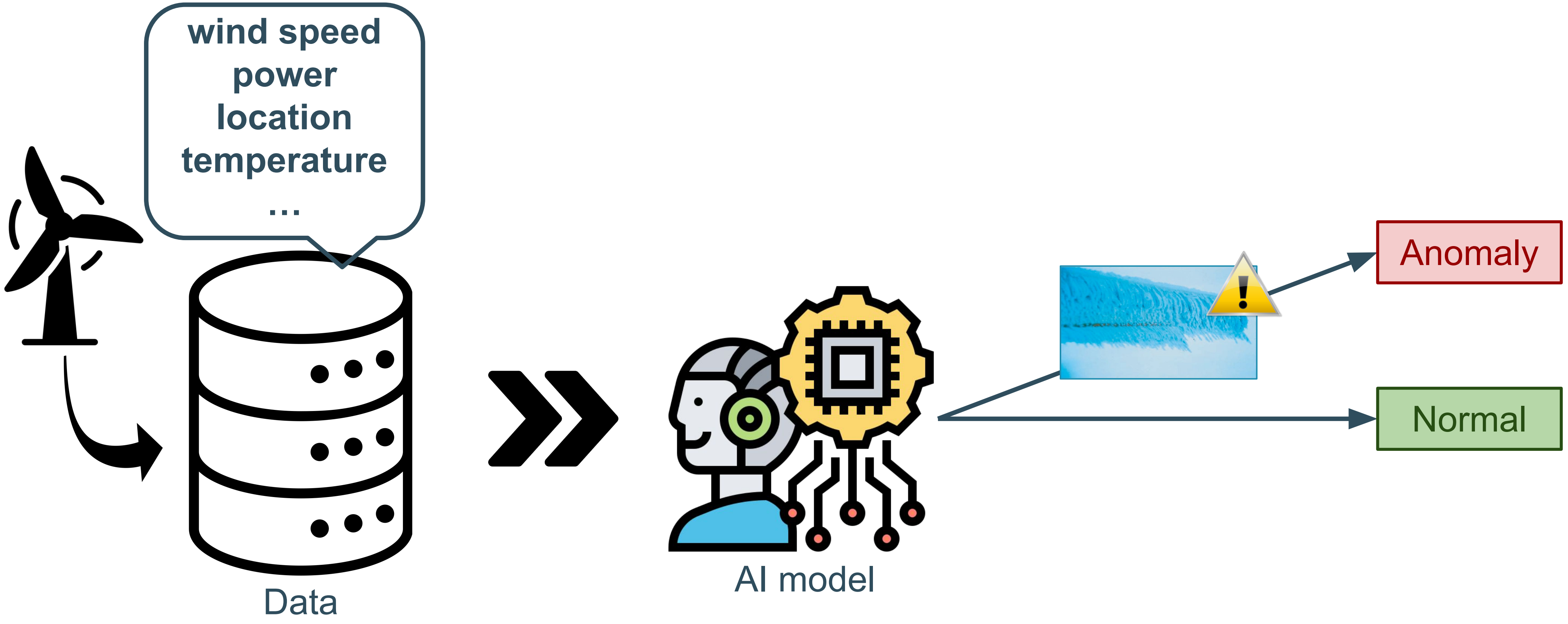
★ power reduction
★ dangerous for people

Anomalies are unexpected and critical events

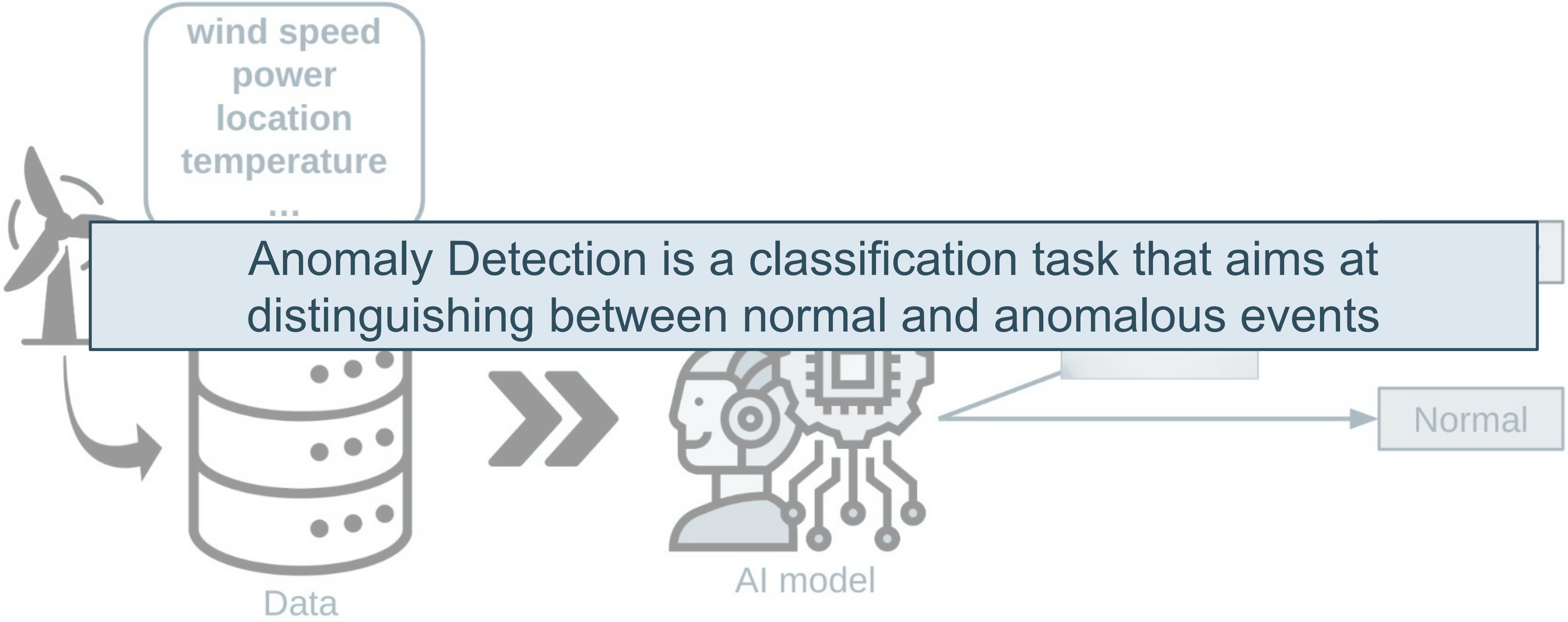


Detecting anomalies in time enables quick & efficient maintenance, which reduce waste of energy and harmful events

Anomaly detection:
how do we automatically detect <u>anomalous</u> events?



Anomaly detection:
how do we automatically detect anomalous events?



Anomaly Detection is a classification task that aims at distinguishing between normal and anomalous events

Anomaly detection differs from traditional classification tasks in four aspects

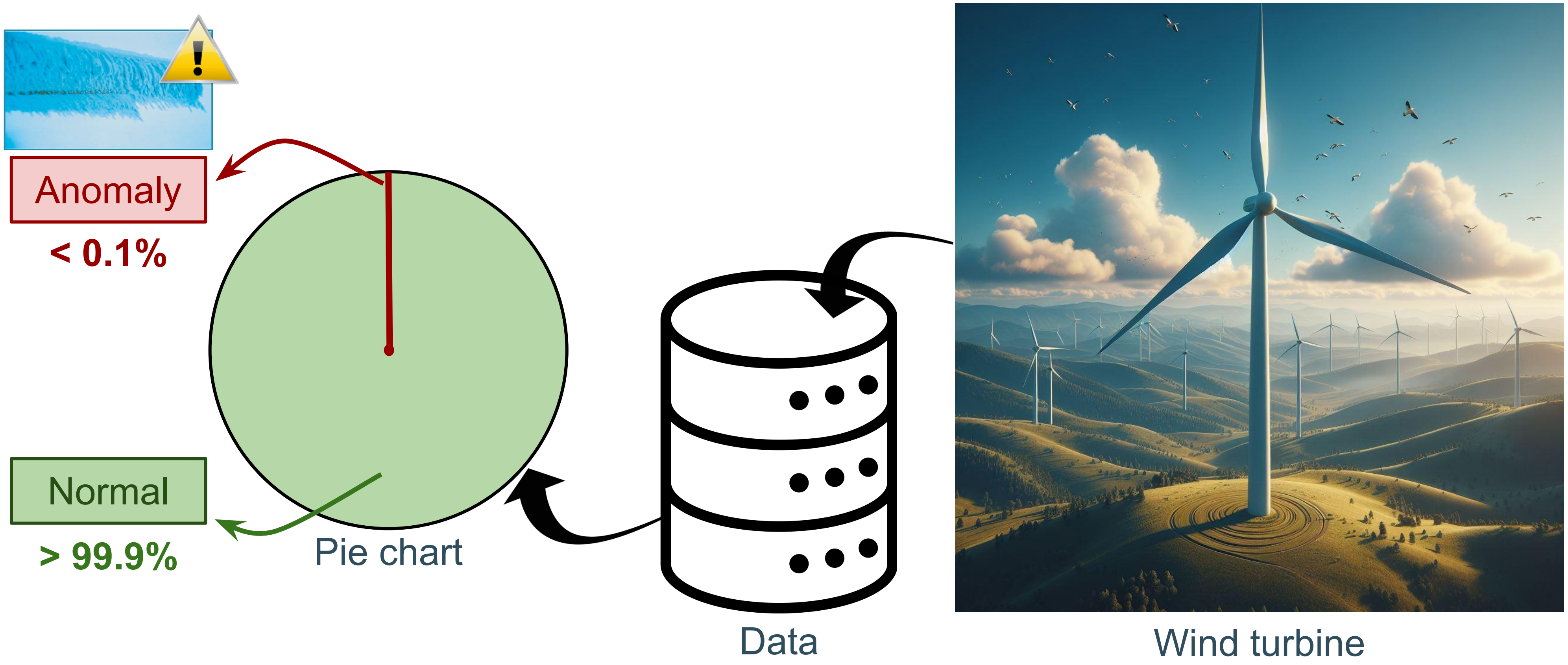
Anomaly detection differs from traditional classification tasks in four aspects

A1. *The collected dataset is scarcely labeled or not labeled at all*

Experts Data Wind turbine

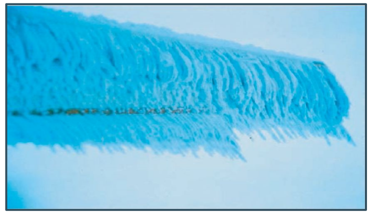
Anomaly detection differs from traditional classification tasks in four aspects

A2. *Anomalies are rare events*



Anomaly detection differs from traditional classification tasks in four aspects

A3. *The recorded anomalies may not comprehensively represent all potential cases*



Blade icing

Blade erosion

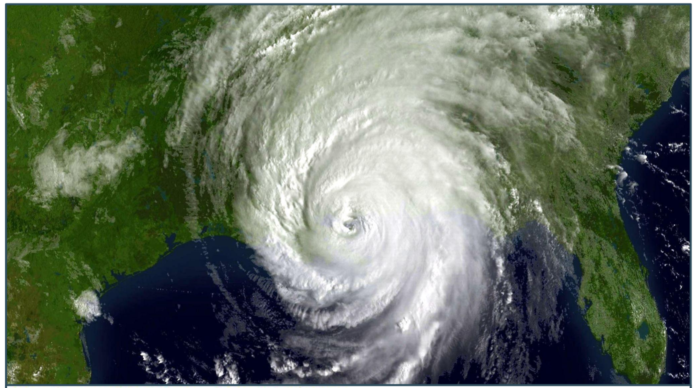
Lightning strikes

…?

Wind turbine

Anomaly detection differs from traditional classification tasks in four aspects

A4. *Unique one-off anomalies may occur*

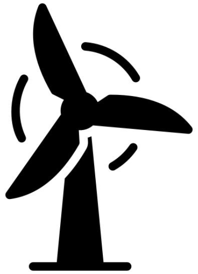


Hurricane Katrina - August 2005

In all, **Hurricane Katrina** was responsible for 1,833 fatalities and approximately $108 billion in **damage** (un-adjusted 2005 dollars). On August 23rd, a tropical ...

Wind turbine

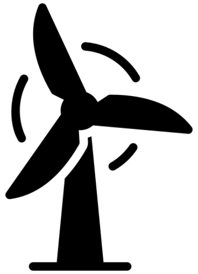
Given these challenges, how does anomaly detection work?



Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)
1	14	25	0.6	200	55
2	12	55	2.8	180	95
3	11	62	6.1	220	160
4	12	35	4.5	190	145
5	7	30	0.8	170	52
6	2	85	5.2	180	57
7	10	48	4.6	185	143
...	...	...	...	...	...

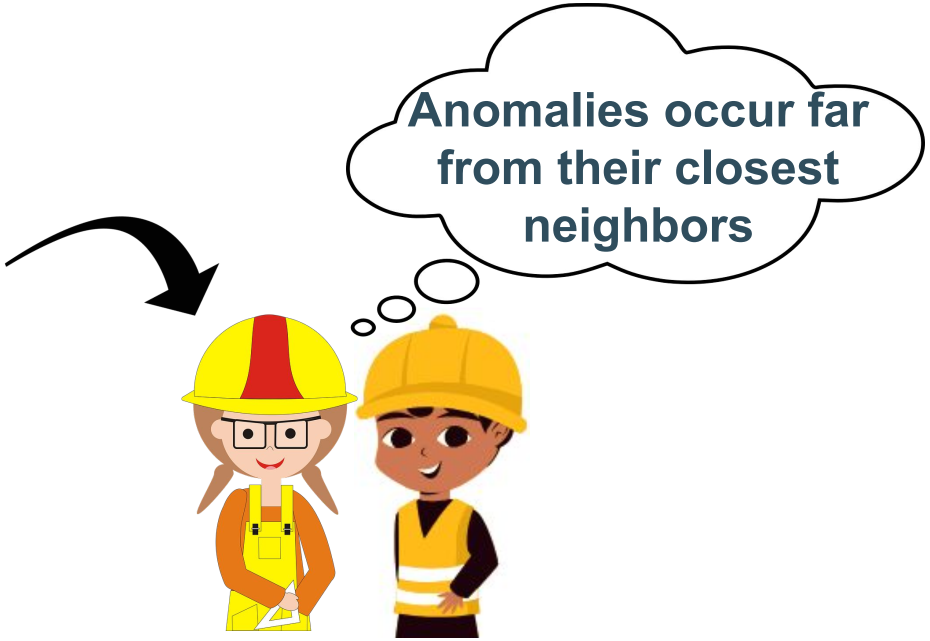
Tabular data

Given these challenges, how does anomaly detection work?



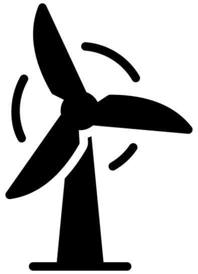
Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)
1	14	25	0.6	200	55
2	12	55	2.8	180	95
3	11	62	6.1	220	160
4	12	35	4.5	190	145
5	7	30	0.8	170	52
6	2	85	5.2	180	57
7	10	48	4.6	185	143
...	...	...	...	...	...

Tabular data



Experts

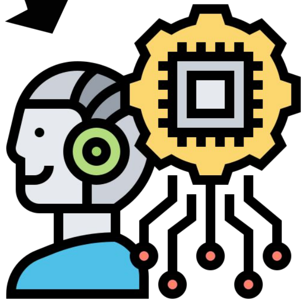
Given these challenges, how does anomaly detection work?



Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)
1	14	25	0.6	200	55
2	12	55	2.8	180	95
3	11	62	6.1	220	160
4	12	35	4.5	190	145
5	7	30	0.8	170	52
6	2	85	5.2	180	57
7	10	48	4.6	185	143
...	...	...	...	...	...

Tabular data

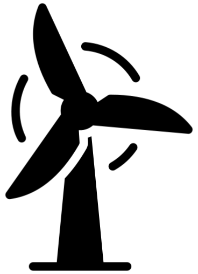
Experts



measure anomalousness using the experts' insight

AI model

Given these challenges, how does anomaly detection work?

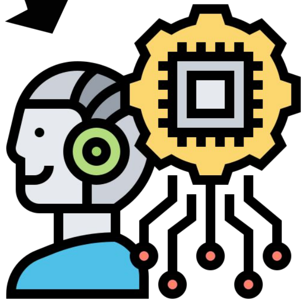


Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)
1	14	25	0.6	200	55
2	12	55	2.8	180	95
3	11	62	6.1	220	160
4	12	35	4.5	190	145
5	7	30	0.8	170	52
6	2	85	5.2	180	57
7	10	48	4.6	185	143
...	...	...	...	...	...

Tabular data

Anomalies occur far from their closest neighbors

Experts



AI model

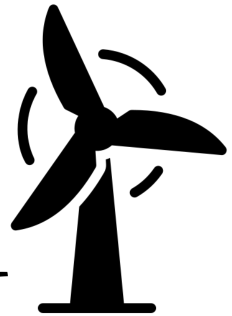
measure anomalousness using the experts' insight

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
2	12	55	2.8	180	95	**48.8**
3	11	62	6.1	220	160	**41.4**
1	14	25	0.6	200	55	**31.4**
5	7	30	0.8	170	52	**31.5**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**
...	...	...	...	...	...	**...**

The literature of anomaly detection has focused on designing new algorithms <u>but largely ignored three practical challenges</u>

What's missing?

Suppose the task is to decide whether an unknown <u>test sample</u> is anomalous or not

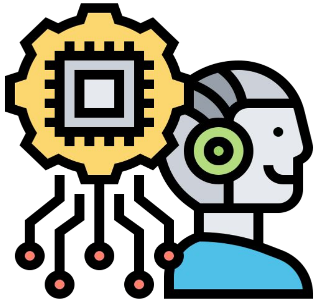


wind turbine

Experts

Collected test sample:

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)
10	60	6.9	220	120



AI model

Anomaly score = 40.0

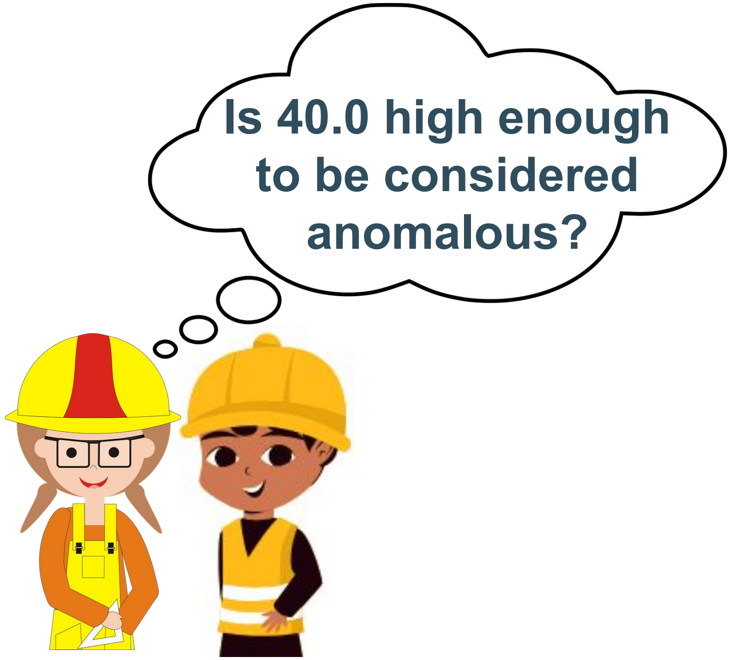
Gap 1: Experts cannot make decisions based solely on scores because they are not interpretable

train

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
2	12	55	2.8	180	95	**48.8**
3	11	62	6.1	220	160	**41.4**
1	14	25	0.6	200	55	**31.4**
5	7	30	0.8	170	52	**31.5**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**
…	…	…	…	…	…	**…**

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
10	60	6.9	220	120	**40**



Experts

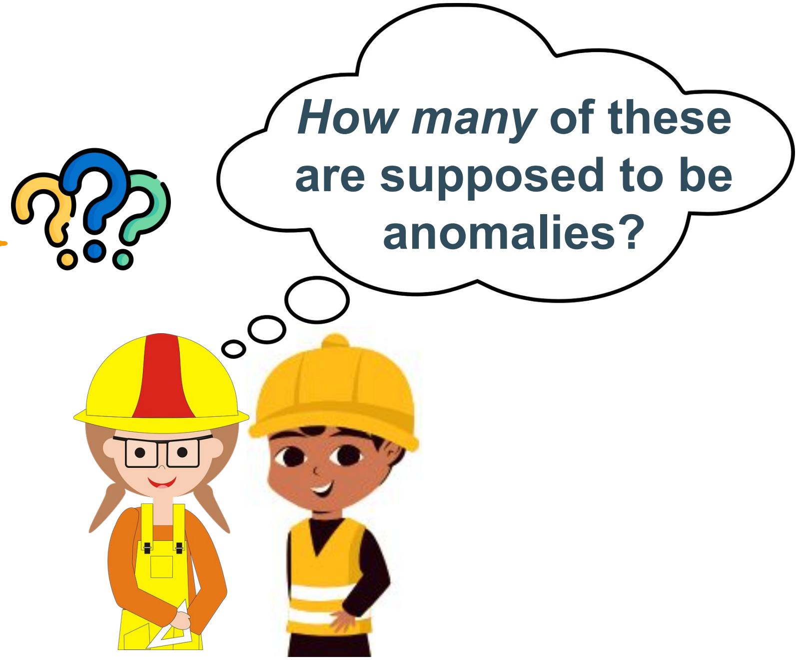
Gap 1: Experts cannot make decisions based solely on scores because they are not interpretable

train

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
2	12	55	2.8	180	95	**48.8**
3	11	62	6.1	220	160	**41.4**
1	14	25	0.6	200	55	**31.4**
5	7	30	0.8	170	52	**31.5**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**
...	...	...	...	...	...	**...**

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
10	60	6.9	220	120	**40**



Experts

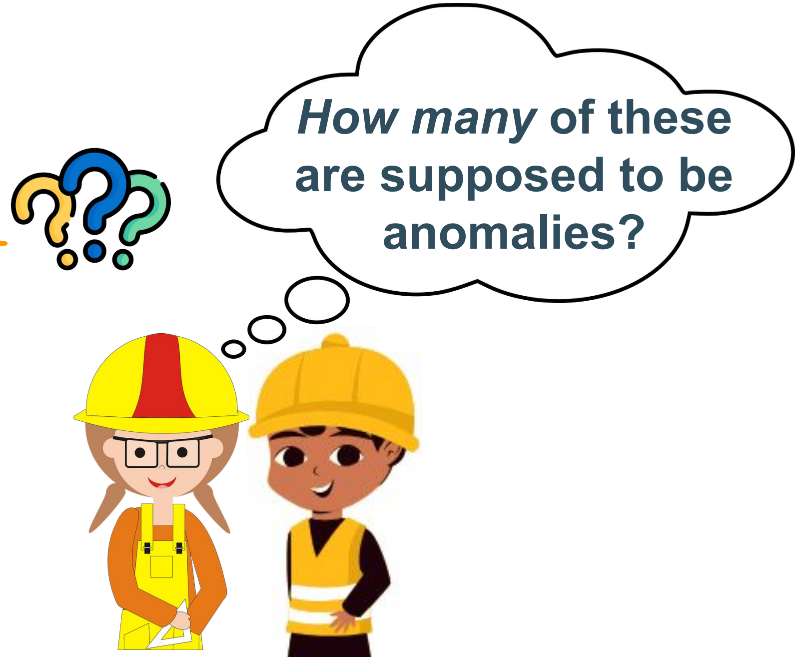
Gap 1: Experts cannot make decisions based solely on scores because they are not interpretable

train

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
2	12	55	2.8	180	95	**48.8**
3	11	62	6.1	220	160	**41.4**
1	14	25	0.6	200	55	**31.4**
5	7	30	0.8	170	52	**31.5**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**
...	...	...	...	...	...	**...**

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
10	60	6.9	220	120	**40**



Experts

What's missing?

➔ An estimate of the expected proportion of anomalies, i.e. the "**contamination**" level

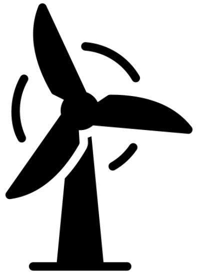
Why do we need it?

➔ For decision making: we need to know whether a sample is **anomalous "enough"**

Contribution #1: Estimating the contamination of a dataset

We analyze three realistic yet different settings

1 » we are able to collect some normal labels



wind turbine

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Label
1	14	25	0.6	200	55	**?**
2	12	55	2.8	180	95	**?**
3	11	62	6.1	220	160	**?**
4	12	35	4.5	190	145	**Normal**
5	7	30	0.8	170	52	**?**
6	2	85	5.2	180	57	**?**
7	10	48	4.6	185	143	**Normal**
...	...	...	...	...	...	**...**

Tabular data

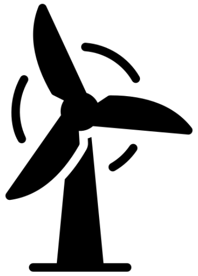
Experts

Perini L, Vercruyssen V, Davis J: *Class Prior Estimation in Active Positive and Unlabeled Learning*, **IJCAI-PRICAI 2020**

Contribution #1: Estimating the contamination of a dataset

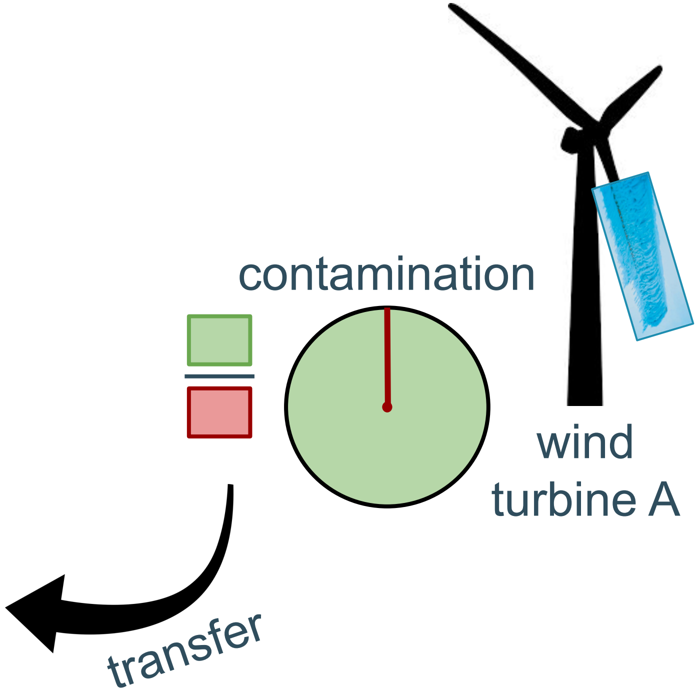
We analyze three realistic yet different settings

2 » its true value is given for a related domain



Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)
1	14	25	0.6	200	55
2	12	55	2.8	180	95
3	11	62	6.1	220	160
4	12	35	4.5	190	145
5	7	30	0.8	170	52
6	2	85	5.2	180	57
7	10	48	4.6	185	143
...	...	...	...	...	...

Tabular data

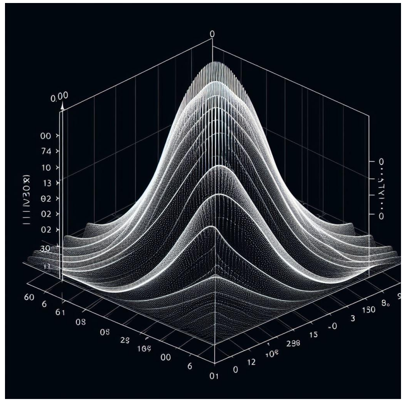


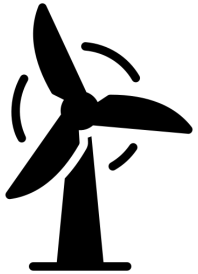
Perini L, Vercruyssen V, Davis J: *Transferring the Contamination Factor between Anomaly Detection Domains by Shape Similarity*, **AAAI 2022**

Contribution #1: Estimating the contamination of a dataset

We analyze three realistic yet different settings

3 » **we must account for uncertainty**





wind turbine

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)
1	14	25	0.6	200	55
2	12	55	2.8	180	95
3	11	62	6.1	220	160
4	12	35	4.5	190	145
5	7	30	0.8	170	52
6	2	85	5.2	180	57
7	10	48	4.6	185	143
...	...	...	...	...	...

Tabular data

It is likely that only three anomalies have occurred!

Experts

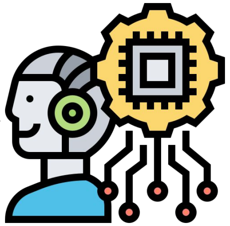
Perini L, Buerkner P, Klami A: *Estimating the Contamination Factor's Distribution in Unsupervised Anomaly Detection*, **ICML 2023**

Now, we have “three” ways to estimate “how anomalous” a sample has to be to get detected as an anomaly

train

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
2	12	55	2.8	180	95	**48.8**
3	11	62	6.1	220	160	**41.4**
1	14	25	0.6	200	55	**31.4**
5	7	30	0.8	170	52	**31.5**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**
...	...	...	...	...	...	**...**

41.4



AI model

< 41.4

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores	Prediction
10	60	6.9	220	120	**40**	**Normal**

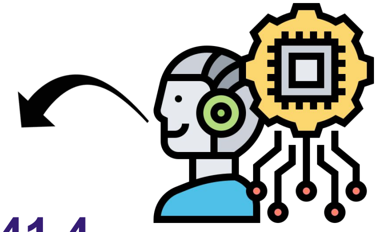
1st contrib.

Now, we have "three" ways to estimate "how anomalous" a sample has to be to get detected as an anomaly

train

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
2	12	55	2.8	180	95	**48.8**
3	11	62	6.1	220	160	**41.4**
1	14	25	0.6	200	55	**31.4**
5	7	30	0.8	170	52	**31.5**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**
...	...	...	...	...	...	**...**

41.4



AI model

< 41.4

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores	Prediction
10	60	6.9	220	120	**40**	**Normal**

There is no *free lunch*: transforming scores into predictions introduces **uncertainty** into the problem

Gap 2: Experts may refuse to use anomaly detection models because they do not know how reliable predictions are

Three samples collected

train 1

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
2	12	55	2.8	180	95	**48.8**
3	11	62	6.1	220	160	**41.4**
1	14	25	0.6	200	55	**31.4**
5	7	30	0.8	170	52	**31.5**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**

41.4

train 2

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
3	11	62	6.1	220	160	**41.4**
2	12	55	2.8	180	95	**25.3**
8	10	35	2.9	177	80	**23**
1	14	25	0.6	200	55	**20**
9	7	21	1.6	193	72	**20**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**
5	7	30	0.8	170	52	**5.5**
10	3	29	2.2	168	49	**5.5**

25.3

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
10	60	6.9	220	120	**40**

Gap 2: Experts may refuse to use anomaly detection models because they do not know how reliable predictions are

train 1

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
2	12	55	2.8	180	95	**48.8**
3	11	62	6.1	220	160	**41.4**
1	14	25	0.6	200	55	**31.4**
5	7	30	0.8	170	52	**31.5**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**

41.4

Three samples collected

train 2

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
3	11	62	6.1	220	160	**41.4**
2	12	55	2.8	180	95	**25.3**
8	10	35	2.9	177	80	**23**
1	14	25	0.6	200	55	**20**
9	7	21	1.6	193	72	**20**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**
5	7	30	0.8	170	52	**5.5**
10	3	29	2.2	168	49	**5.5**

25.3

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
10	60	6.9	220	120	**40**

Prediction

Normal

Prediction

Anomaly

Gap 2: Experts may refuse to use anomaly detection models because they do not know how reliable predictions are

train 1

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
2	12	55	2.8	180	95	**48.8**
3	11	62	6.1	220	160	**41.4**
1	14	25	0.6	200	55	**31.4**
5	7	30	0.8	170	52	**31.5**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**

41.4

Three samples collected

Can we measure such uncertainty in predictions?

train 2

Day	Temper	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2	12	55	2.8	180	95	**25.3**
8	10	35	2.9	177	80	**23**
1	14	25	0.6	200	55	**20**
9	7	21	1.6	193	72	**20**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**
5	7	30	0.8	170	52	**5.5**
10	3	29	2.2	168	49	**5.5**

25.3

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
10	60	6.9	220	120	**40**

Prediction
Normal

Prediction
Anomaly

Contribution #2: Quantifying a model’s uncertainty

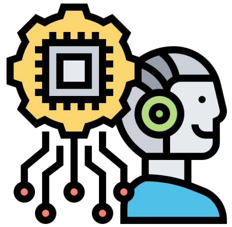
Given:

Training data

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)
1	14	25	0.6	200	55
2	12	55	2.8	180	95
3	11	62	6.1	220	160
4	12	35	4.5	190	145
5	7	30	0.8	170	52
6	2	85	5.2	180	57
7	10	48	4.6	185	143
…	…	…	…	…	…

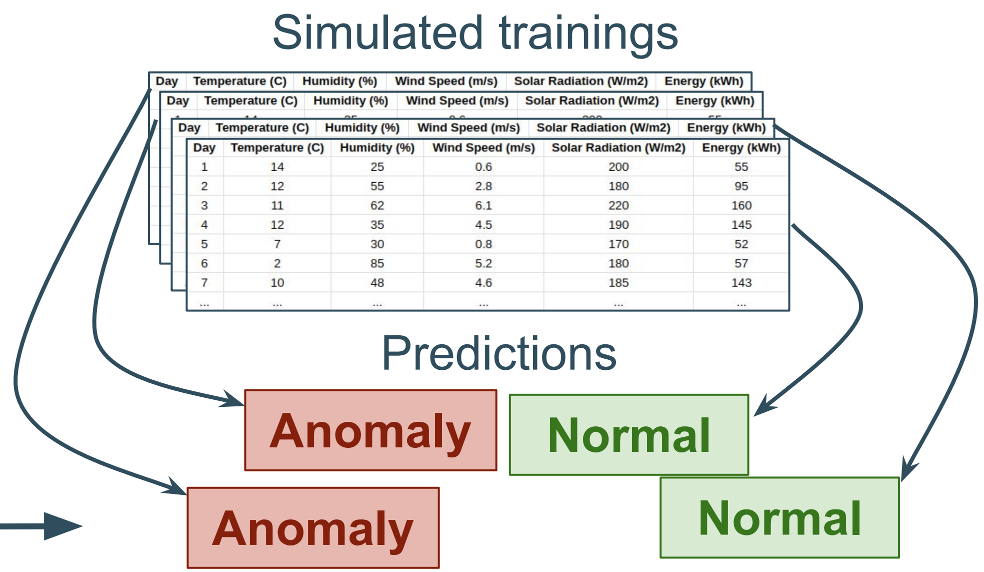
Test sample

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)
10	60	6.9	220	120



AI model

Compute stability:



≅ *50% stability*

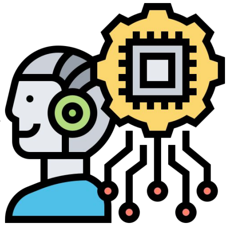
Perini L, Vercruyssen V, Davis J: Quantifying the confidence of anomaly detectors in their example-wise predictions, **ECML-PKDD 2020**

Now, we have a way to estimate a model's stability for a test prediction

train

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
2	12	55	2.8	180	95	**48.8**
3	11	62	6.1	220	160	**41.4**
1	14	25	0.6	200	55	**31.4**
5	7	30	0.8	170	52	**31.5**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**
...	...	...	...	...	...	**...**

41.4



AI model

< 41.4

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores	Prediction	Stability
10	60	6.9	220	120	**40**	**Normal**	**50%**

1st contrib.

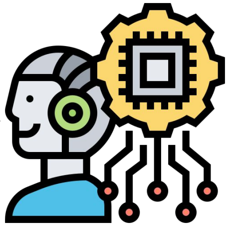
2nd contrib.

Now, we have a way to estimate a model's stability for a test prediction

train

Day	Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores
6	2	85	5.2	180	57	**49.5**
2	12	55	2.8	180	95	**48.8**
3	11	62	6.1	220	160	**41.4**
1	14	25	0.6	200	55	**31.4**
5	7	30	0.8	170	52	**31.5**
4	12	35	4.5	190	145	**14.2**
7	10	48	4.6	185	143	**14.2**
...	...	...	...	...	...	**...**

41.4



AI model

< 41.4

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores	Prediction	Stability
10	60	6.9	220	120	**40**	**Normal**	**50%**

1st contrib.

2nd contrib.

How can we use such uncertainty estimate to improve decision making?

Gap 3: Experts avoid the risk of making wrong decisions by not trusting the model even when it shows minimal uncertainty

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores	Prediction	Stability
10	60	6.9	220	120	**40**	**Normal**	**50%**

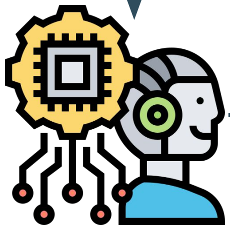
Is this high enough?

Experts

Contribution #3: We allow the model to abstain

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores	Prediction	Stability
10	60	6.9	220	120	**40**	**Normal**	**50%**



AI model

reject the sample
“I don’t know”

~~Normal~~

We’ll manually check!

Experts

confirm prediction

Normal

Perini L, Davis J: Unsupervised Anomaly Detection with Rejection, **NeurIPS 2023**

Contribution #3: We allow the model to abstain

test

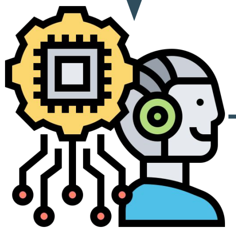
Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores	~~Prediction~~	Stability	Reject
10	60	6.9	220	120	**40**	~~Normal~~	**50%**	**Yes**

★ **What is the benefit of rejection?**

If the model makes a prediction, it is likely to be correct.

★ **What price do we pay?**

The number of predictions is limited



AI model

reject the sample
"I don't know"

~~**Normal**~~

confirm prediction

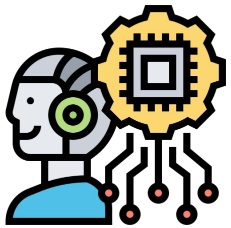
Normal

We'll manually check!

Experts

Perini L, Davis J: Unsupervised Anomaly Detection with Rejection, **NeurIPS 2023**

In conclusion, we made our anomaly detection model *Operational*, *Uncertainty-Aware*, and *Reliable*



AI model

test

Temperature (C)	Humidity (%)	Wind Speed (m/s)	Solar Radiation (W/m2)	Energy (kWh)	Anomaly Scores	Prediction	Stability	Reject
10	60	6.9	220	120	**40**	**Normal**	**50%**	**Yes**

1st contrib. — operational

2nd contrib. — uncertainty aware

3rd contrib. — reliable

Now, we can trust and use our AI model for Anomaly Detection!

Experts

Operational, Uncertainty-Aware, and Reliable Anomaly Detection

Lorenzo Perini

Public PhD defence, 28.03.2024