Monitoring ML services

MLOPS DEPLOYMENT AND LIFE CYCLING

Nemanja Radojkovic
Senior Machine Learning Engineer

Maintaining quality

- Paying customers == expectations of quality
- Quality assurance starts with quality control
- Monitoring

Performance indicators

Fundamental health indicators

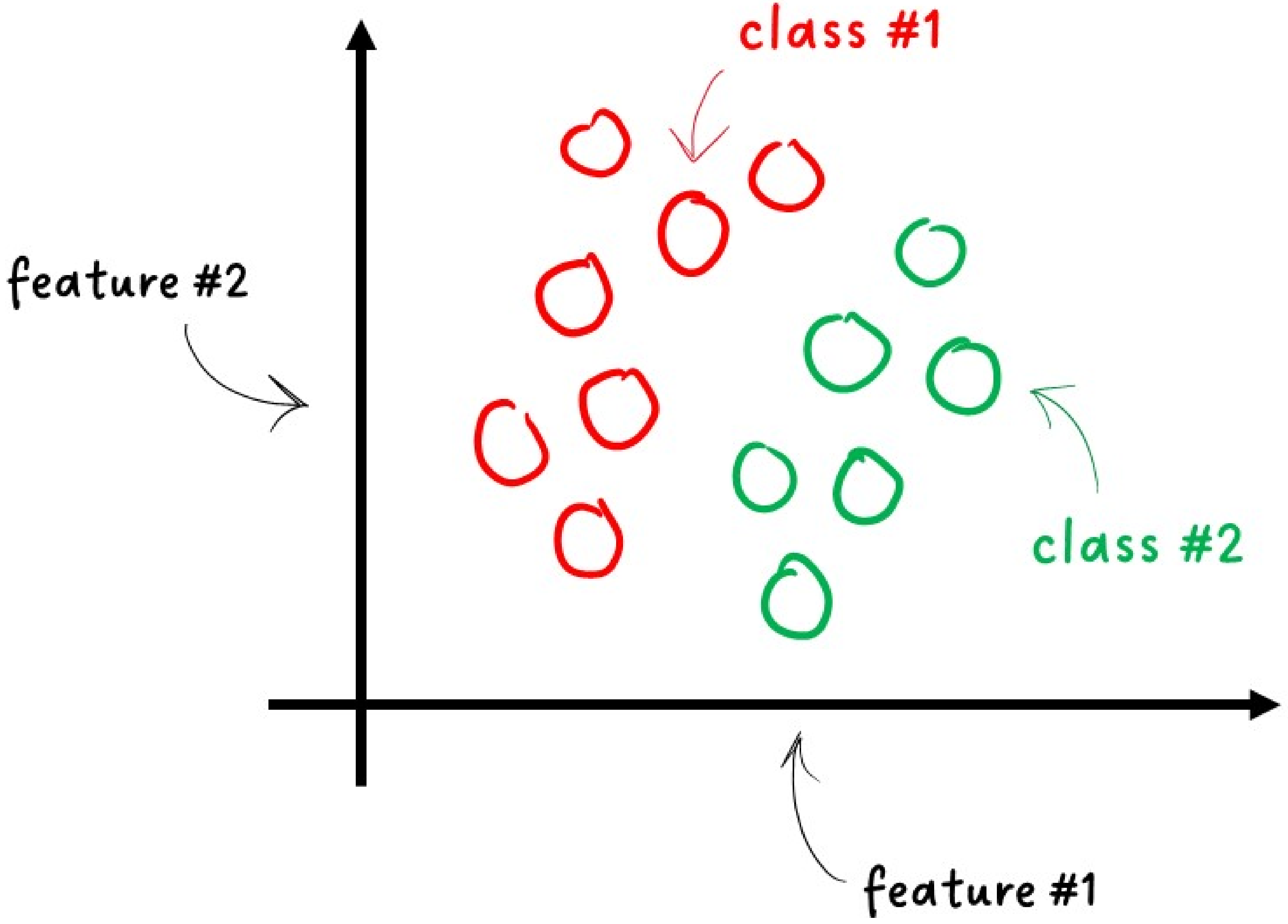
- Service up and running?
- Number of requests in time?
- Latency distribution?

Ultimate quality metric

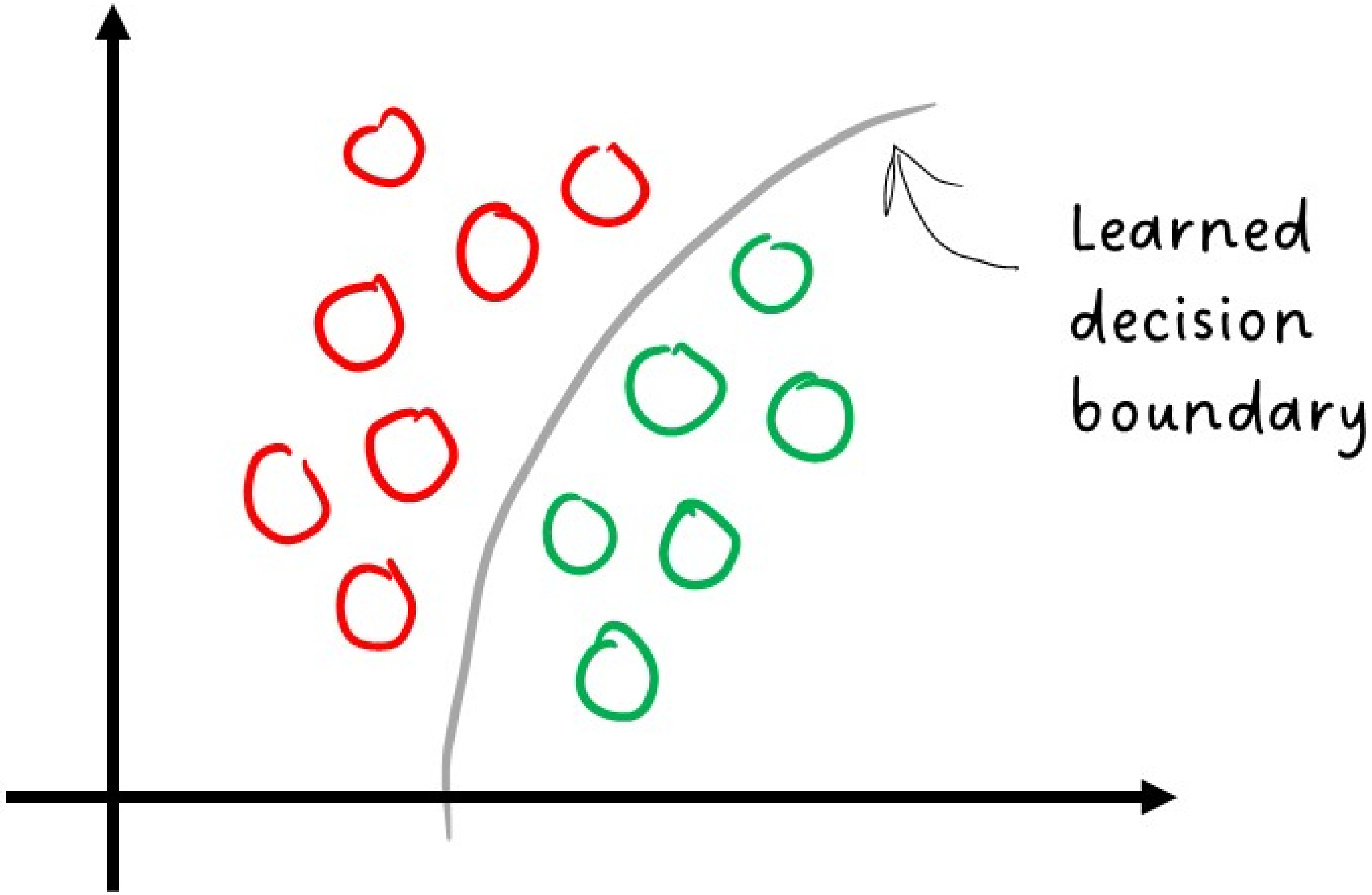
- Predictive performance

How do ML models deteriorate?

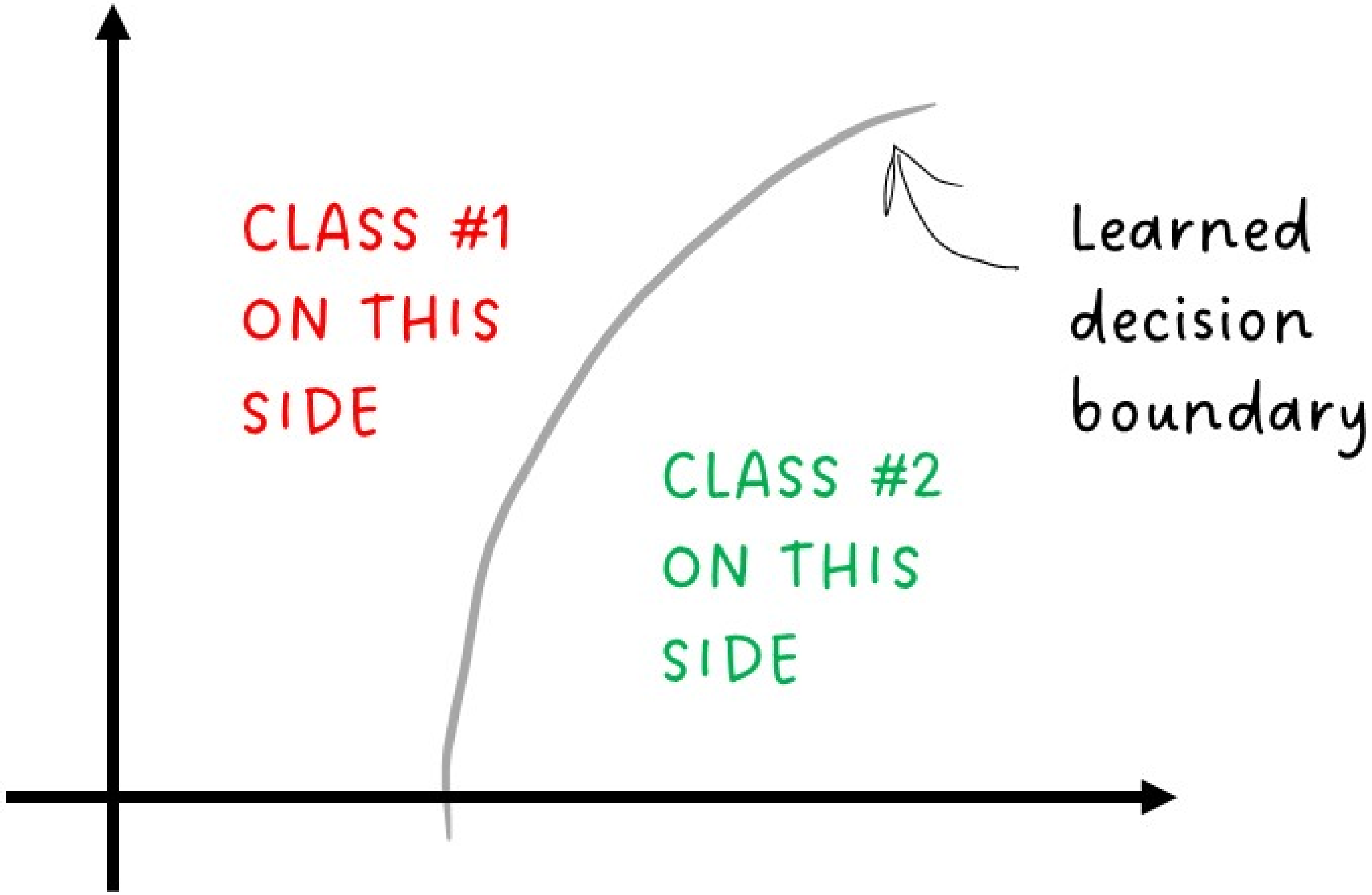
Features and labels at training time



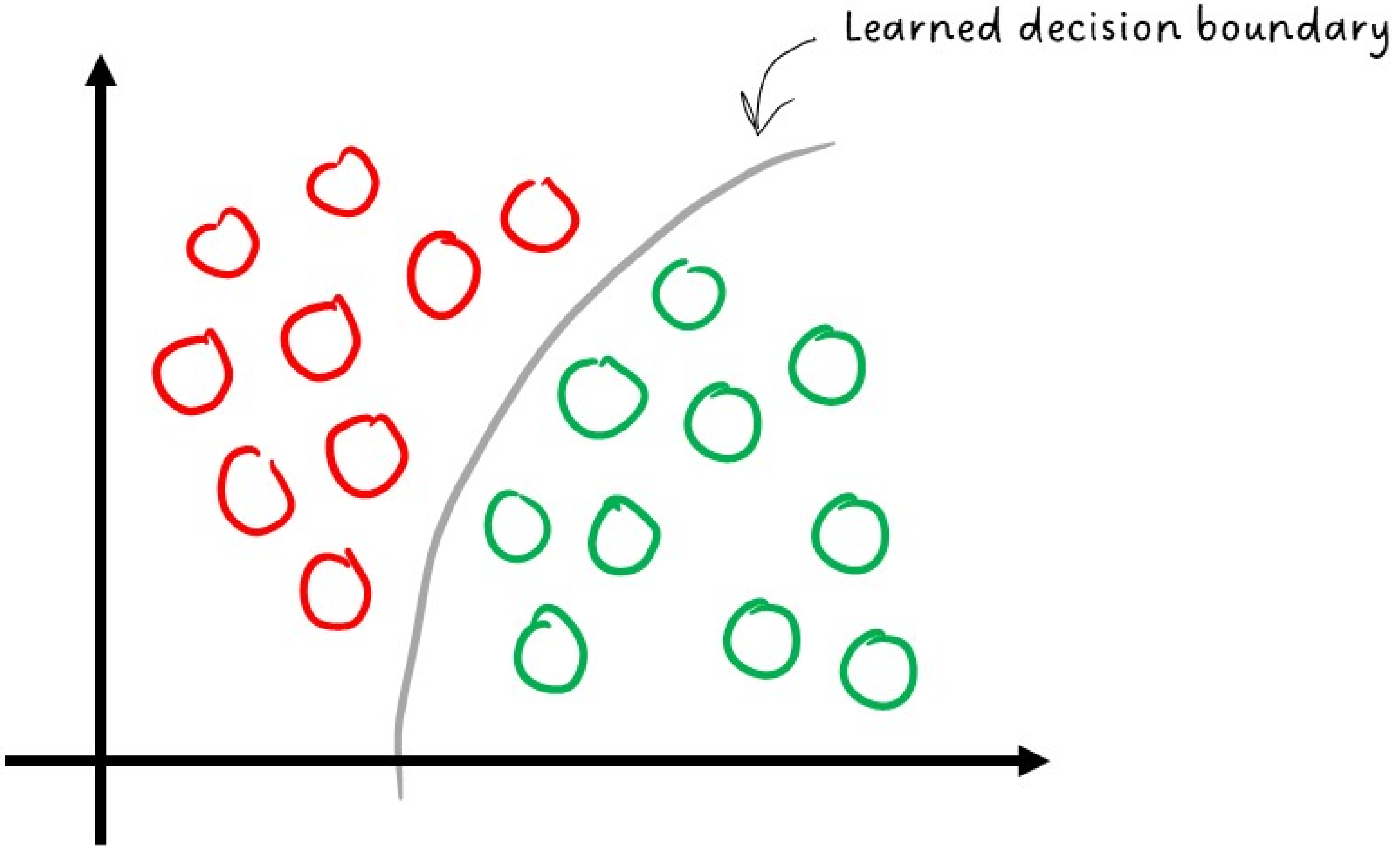
Features and labels at training time

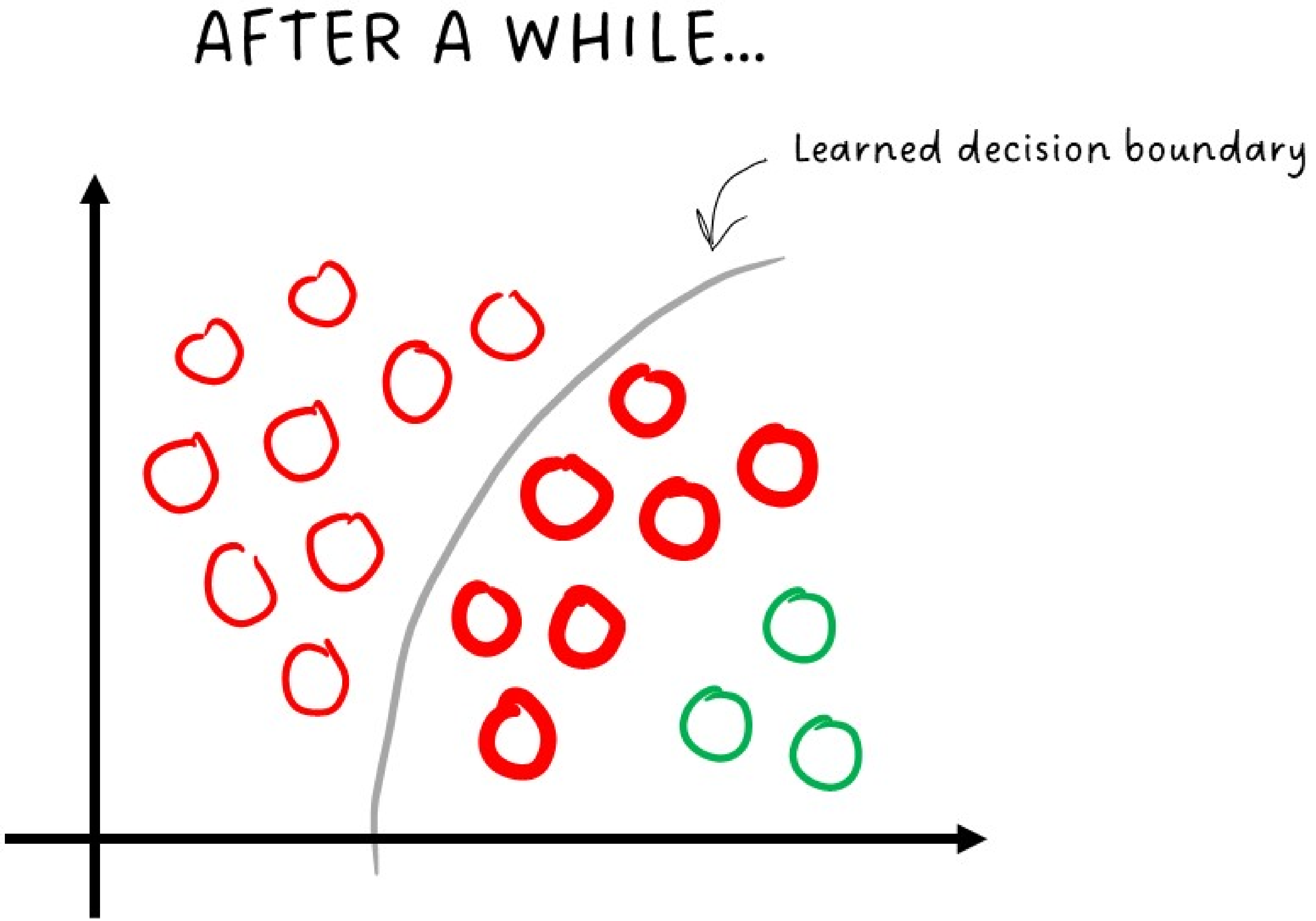


Trained model

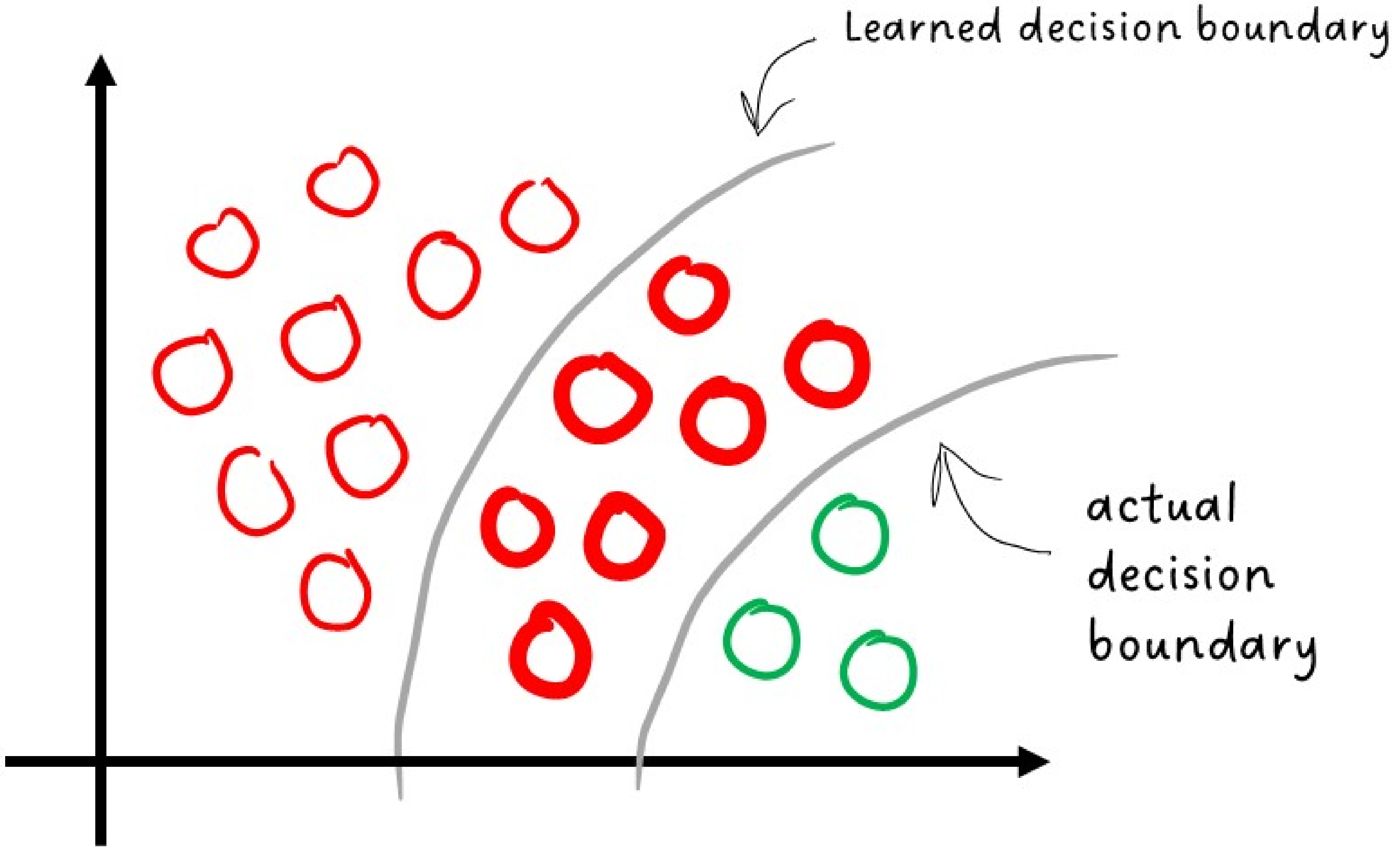


AFTER A WHILE...

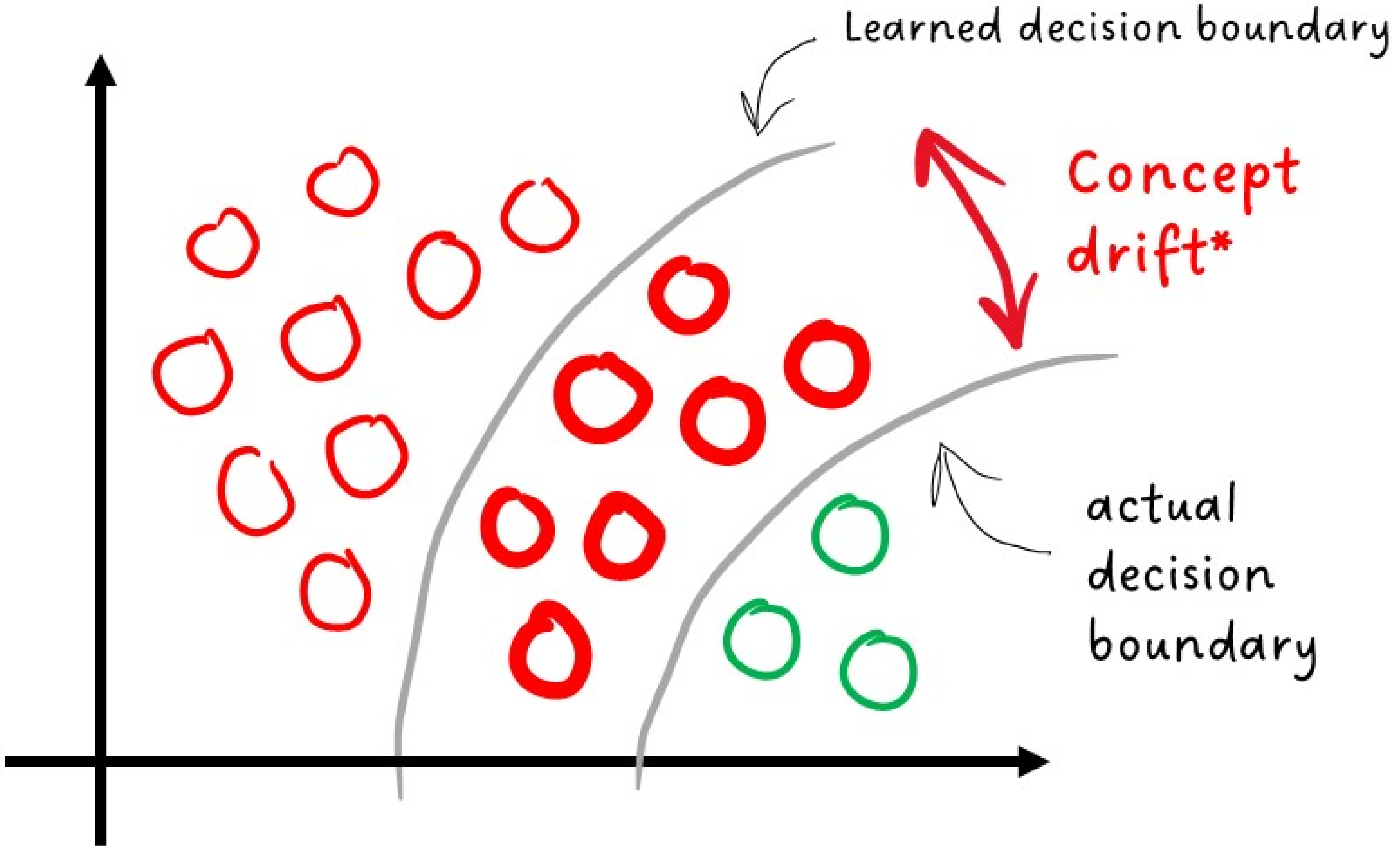



AFTER A WHILE...
Learned decision boundary

AFTER A WHILE...



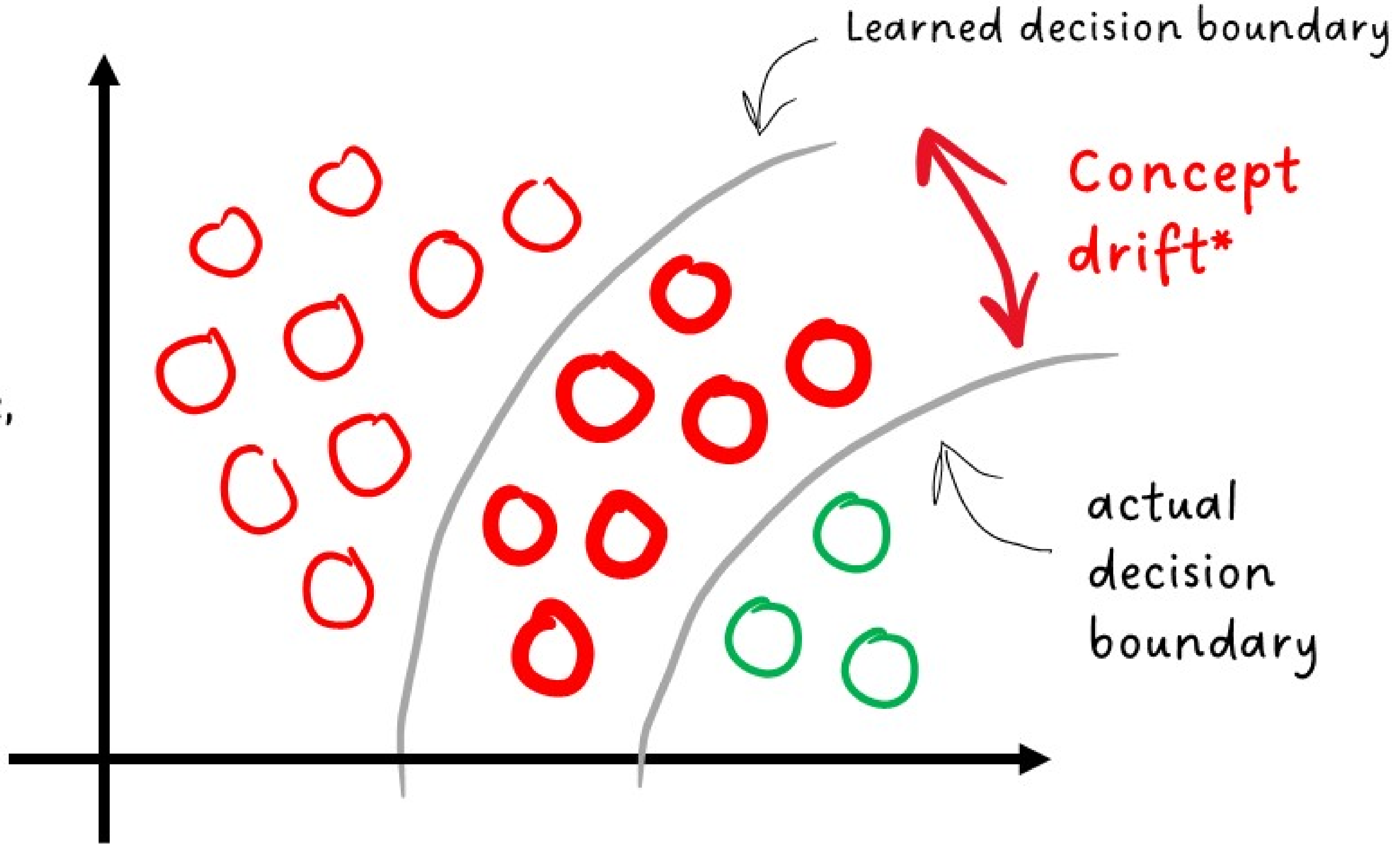
AFTER A WHILE...



*concept drift == significant change in the actual relationship between the input and output features

AFTER A WHILE...

The model "deteriorates" by staying the same, while the reality changes!



*concept drift == significant change in the actual relationship between the input and output features

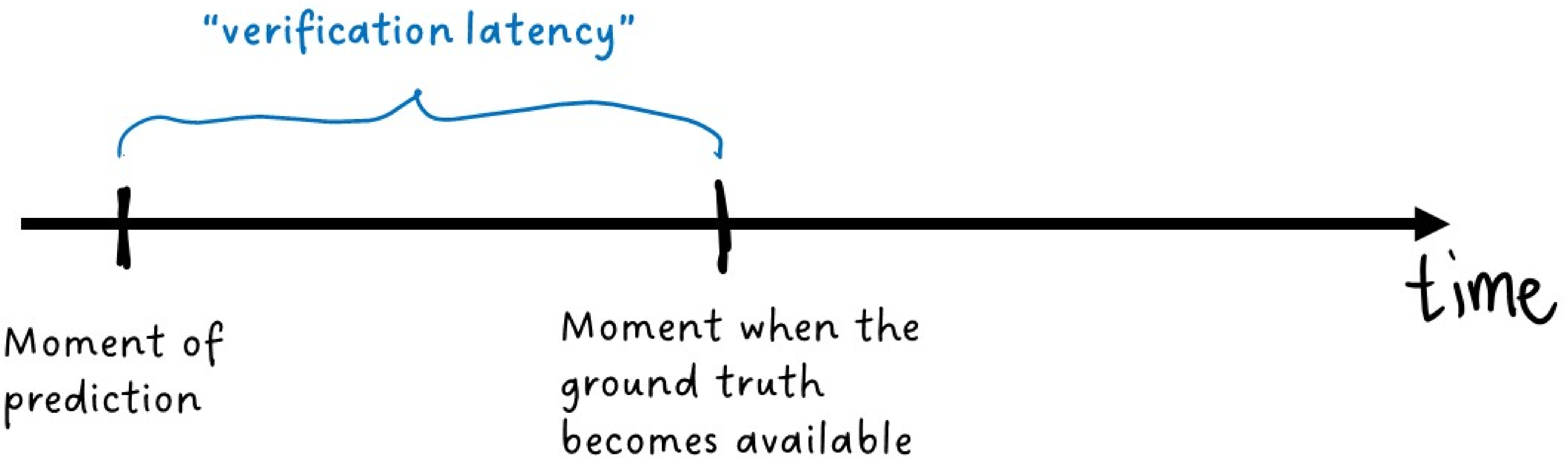
How to detect concept drift?

How to detect concept drift?

⇓

```
IF prediction != ground_truth
MORE_OFTEN_THAN_EXPECTED
THEN WARNING("CONCEPT DRIFT!")
```

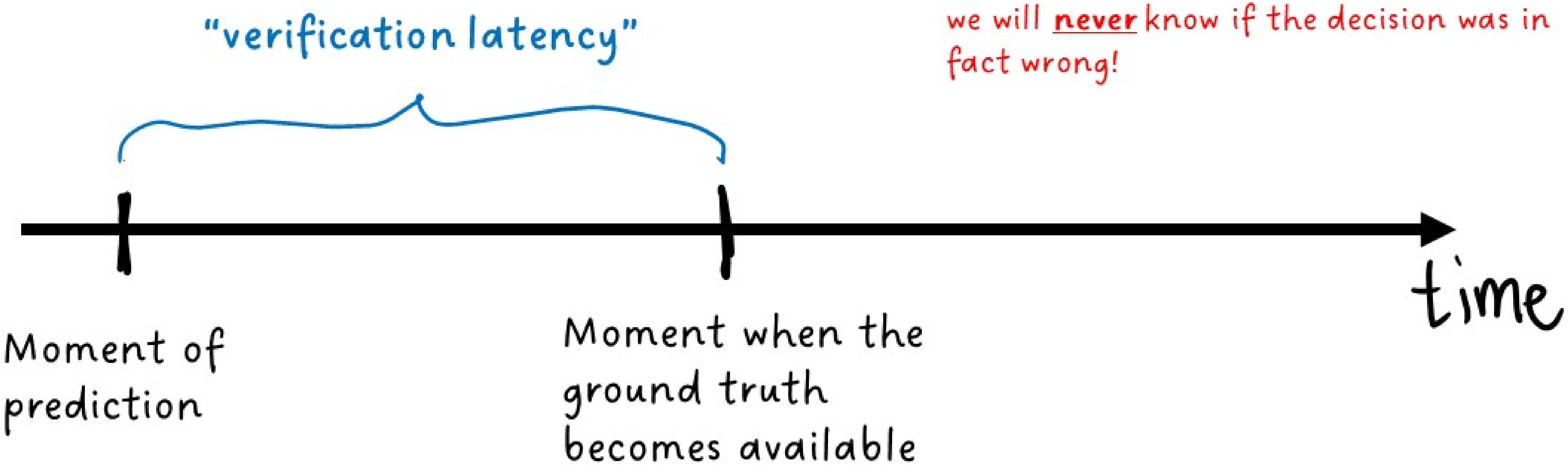
The catch:


"verification latency"
time
Moment of prediction
Moment when the ground truth becomes available

The catch:

Stock prediction: < 1 second
Fraud detection: ~ months
Credit scoring: ∞

If a loan is **denied** based on our model, we will **never** know if the decision was in fact wrong!

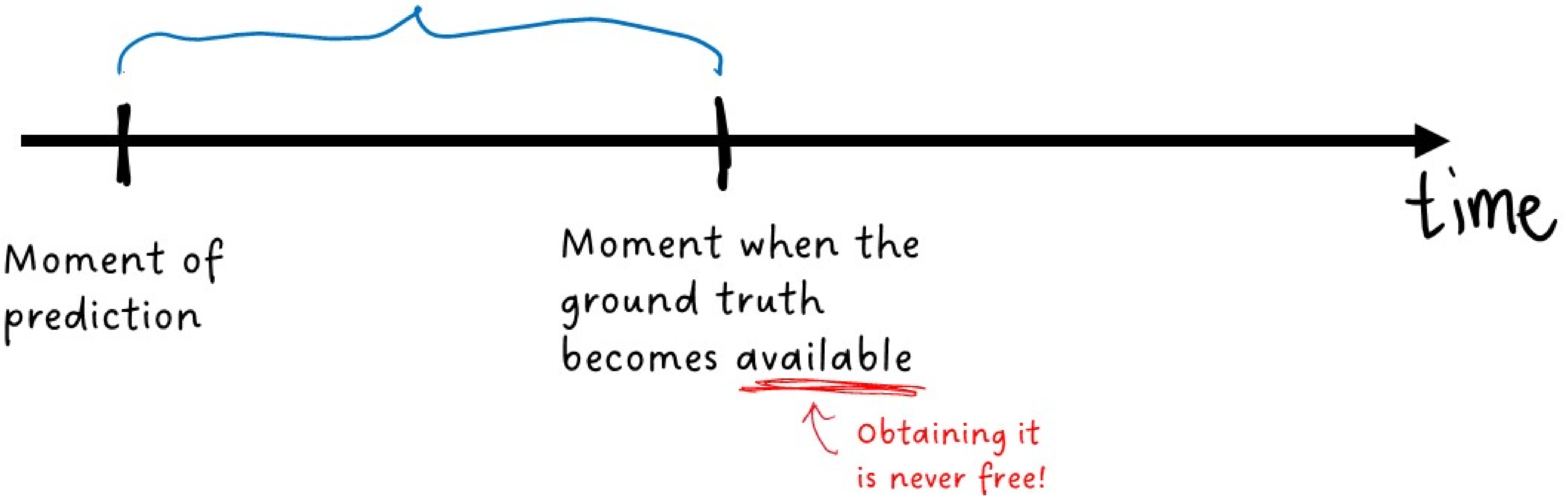


The catch:

Stock prediction: < 1 second
Fraud detection: ~ months
Credit scoring: ∞

If a loan is **denied** based on our model, we will **never** know if the decision was in fact wrong!

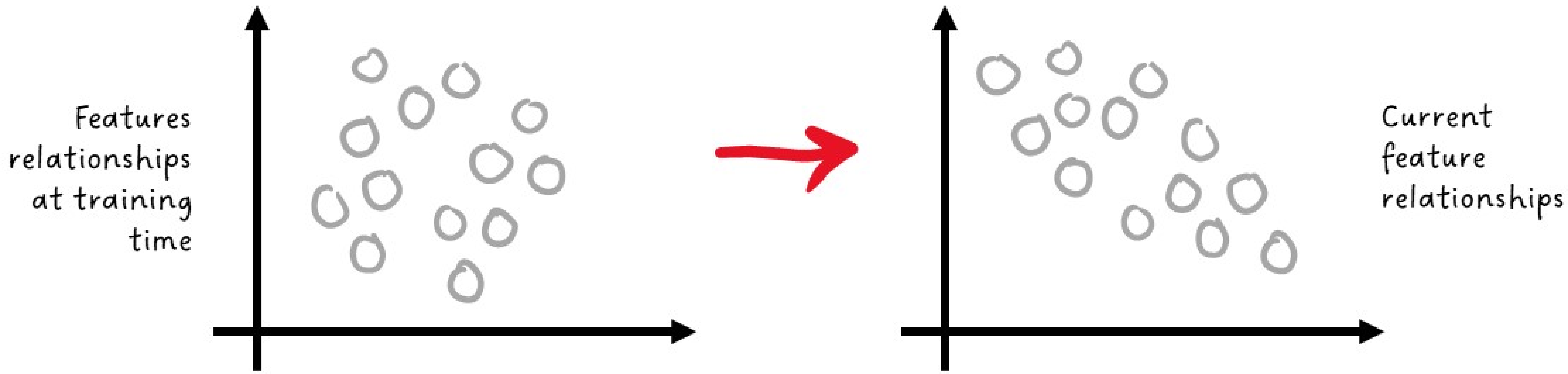
"verification latency"



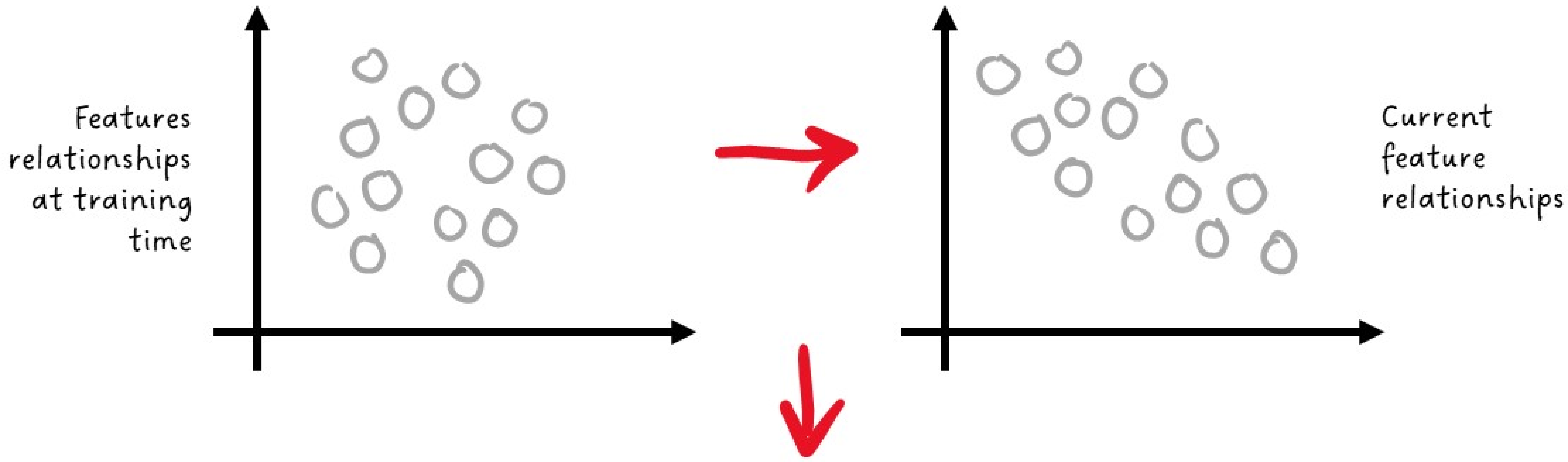
Indirect approach:

Monitor what you have: input features!

Indirect approach:
Monitor what you have: input features!



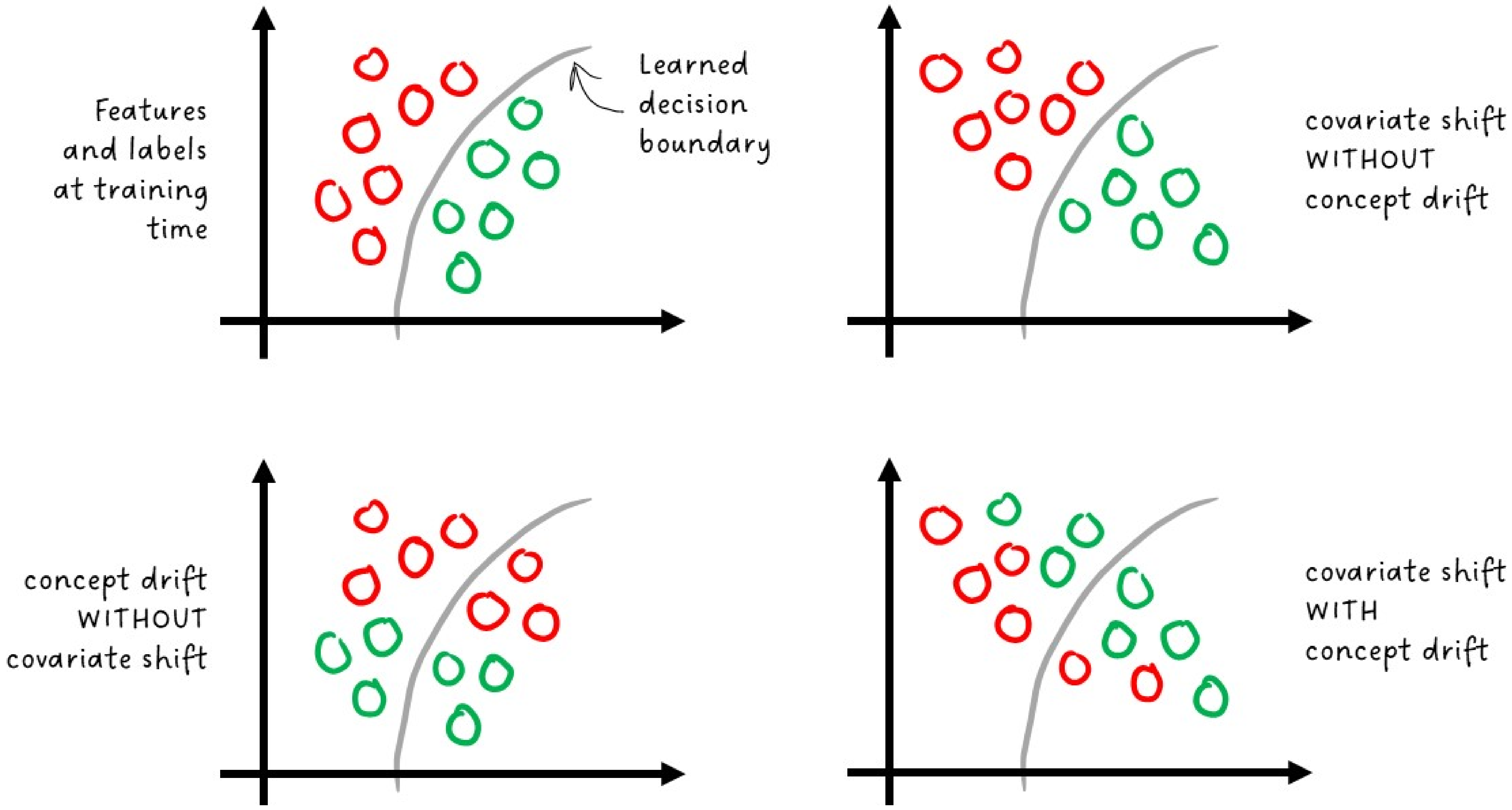
Indirect approach:
Monitor what you have: input features!

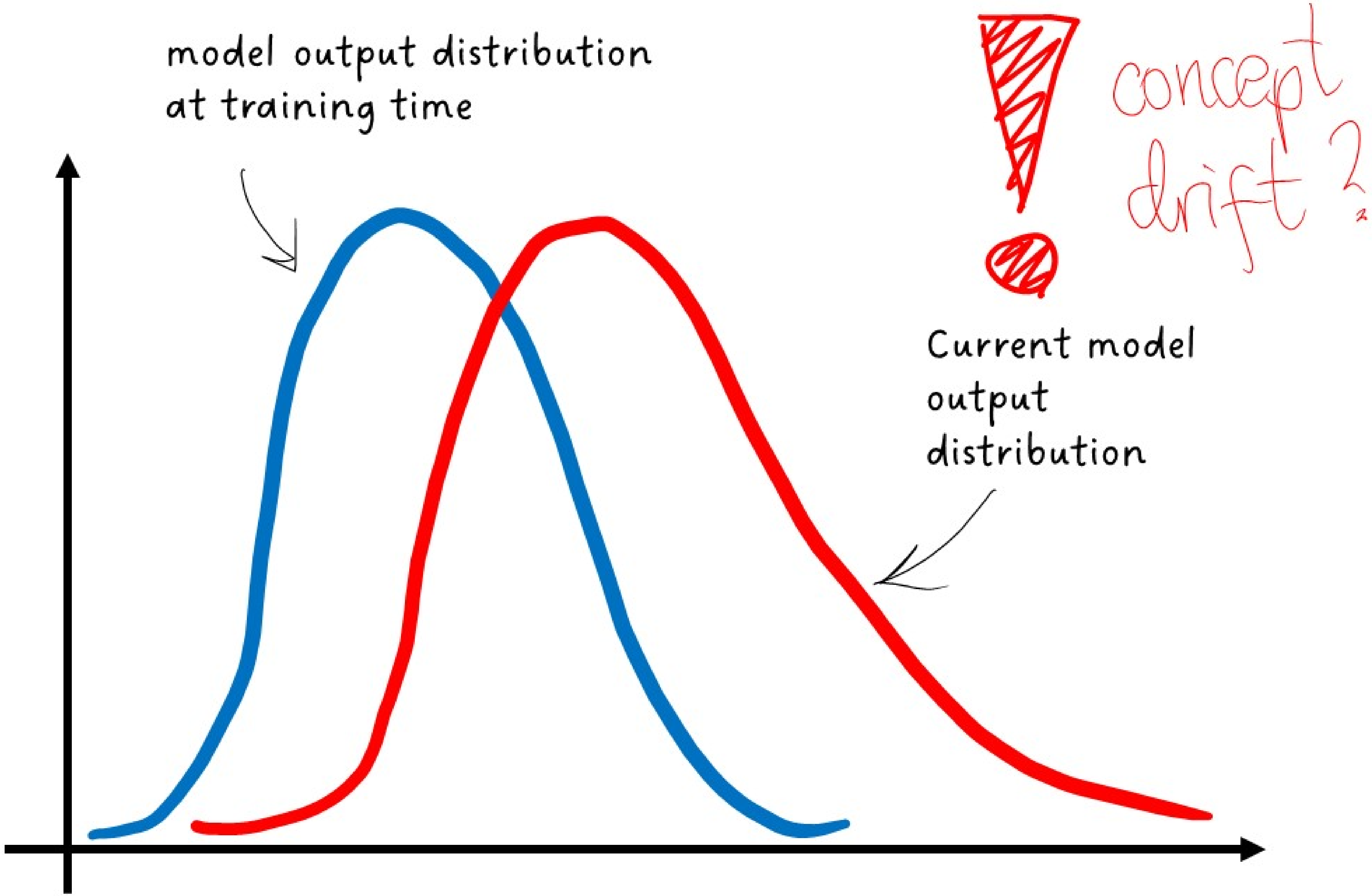


covariate* shift ~ concept drift

* features == "covariates" in statistical jargon

Limitations of input monitoring




model output distribution
at training time
concept
drift ?
Current model
output
distribution

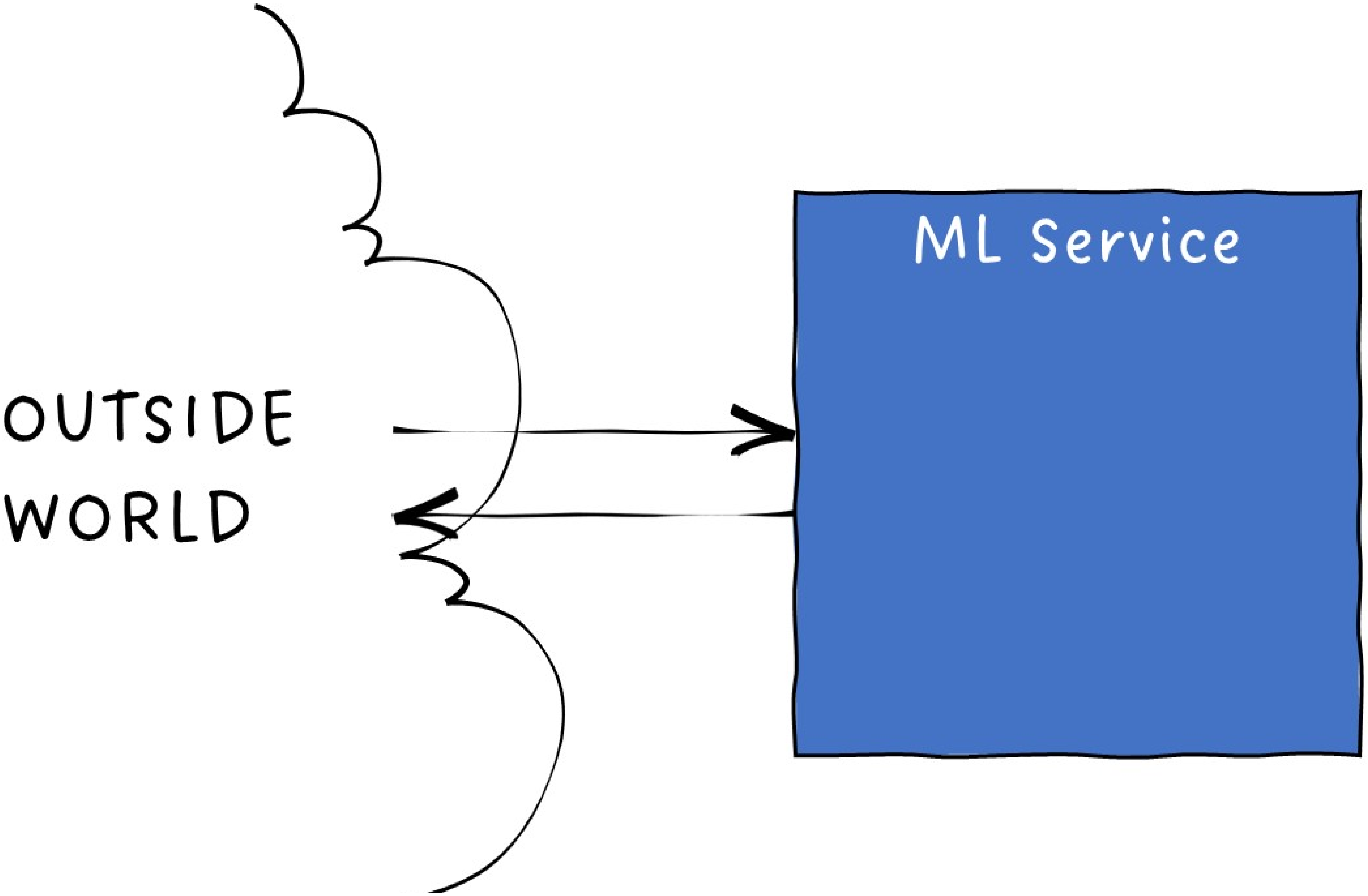
Let's practice!

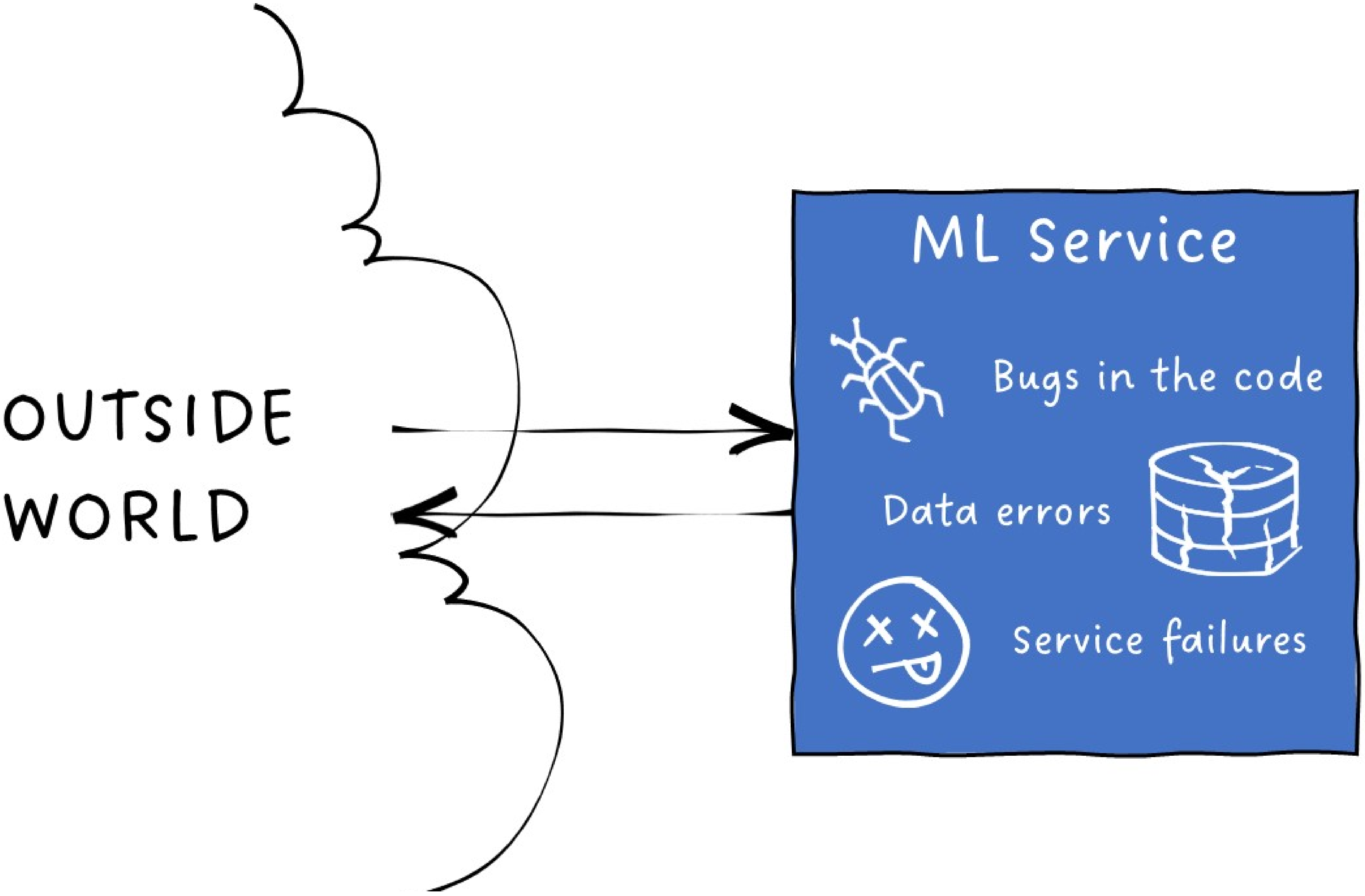
MLOPS DEPLOYMENT AND LIFE CYCLING

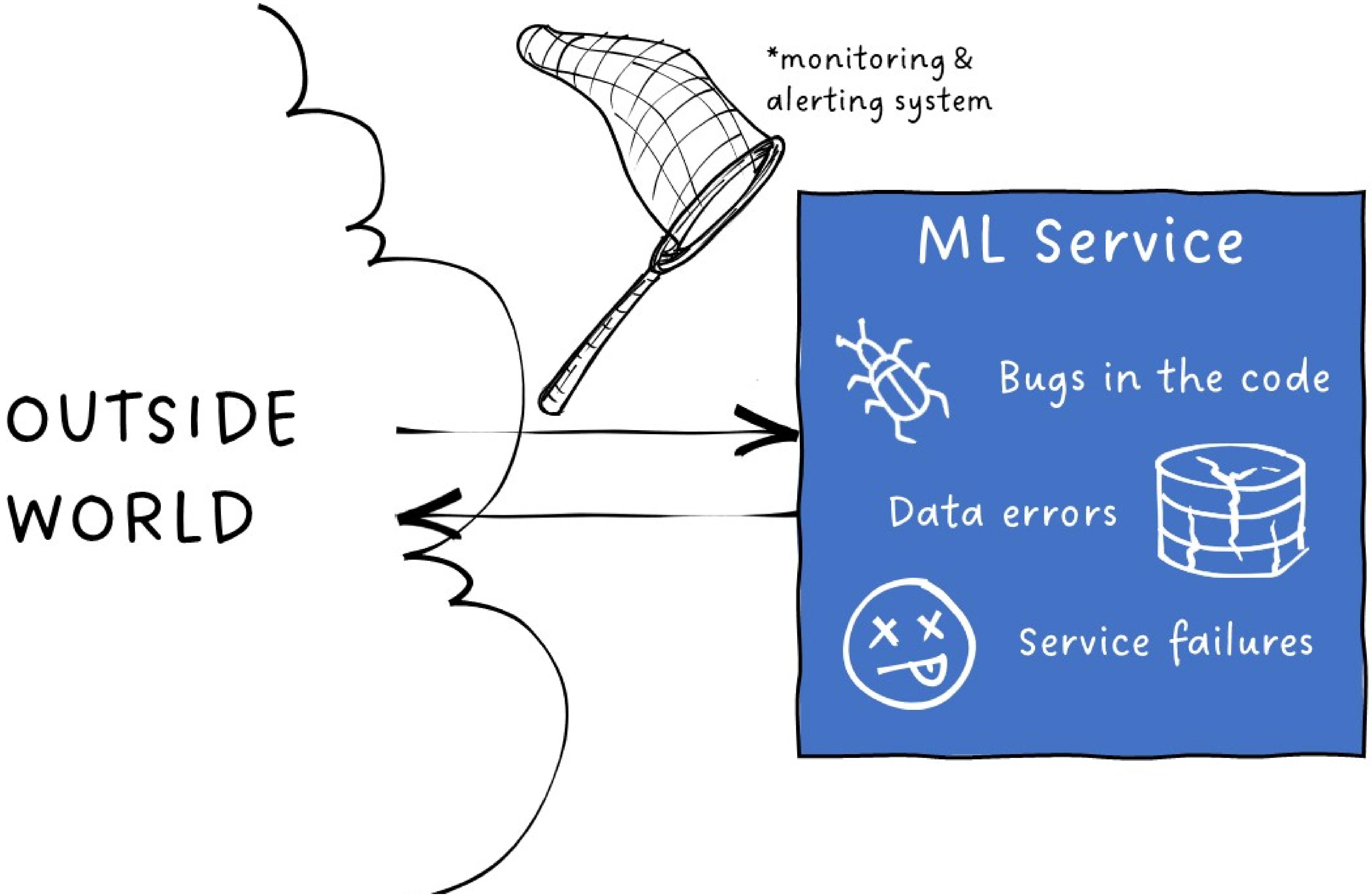
Monitoring and alerting

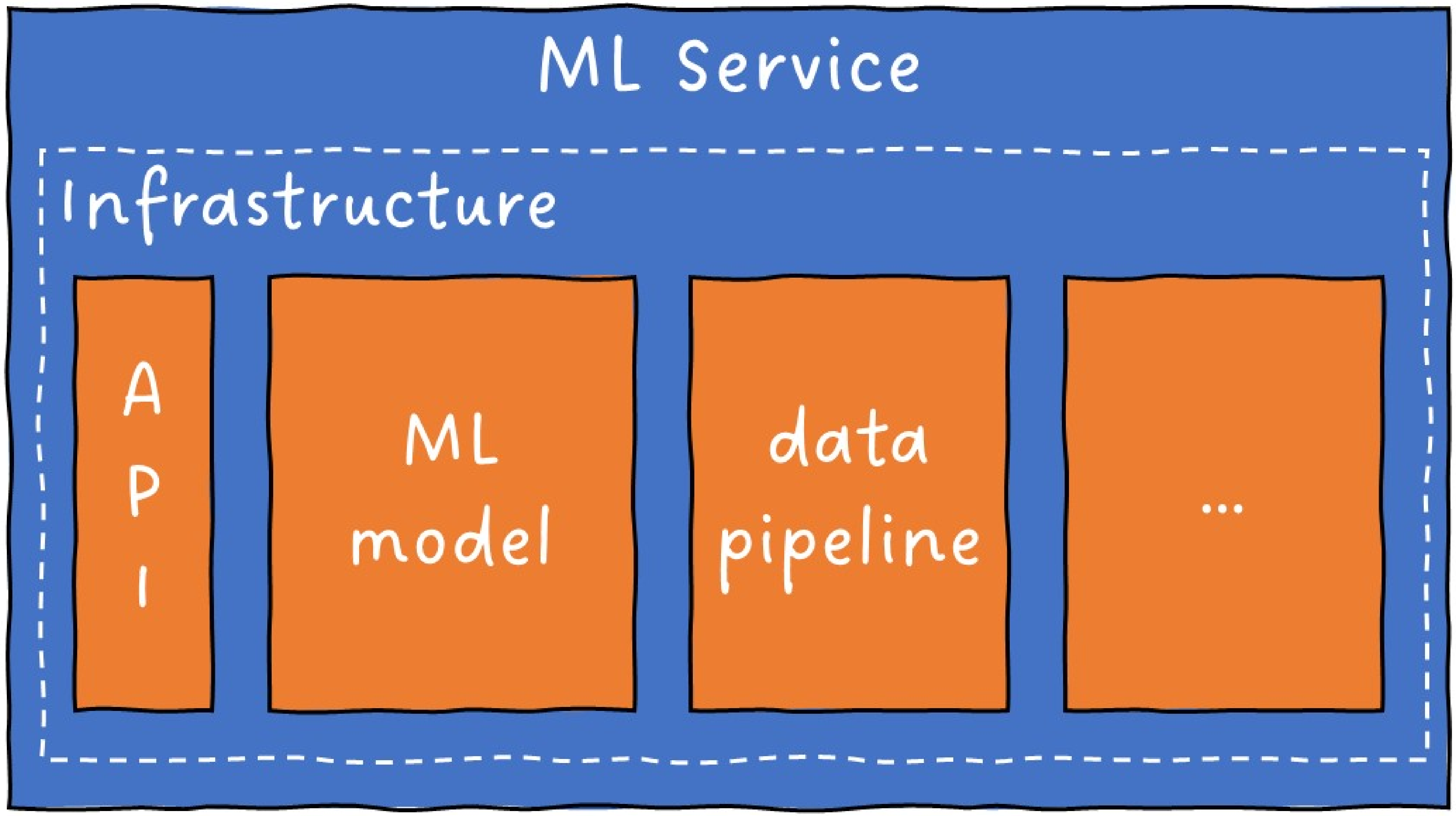
MLOPS DEPLOYMENT AND LIFE CYCLING

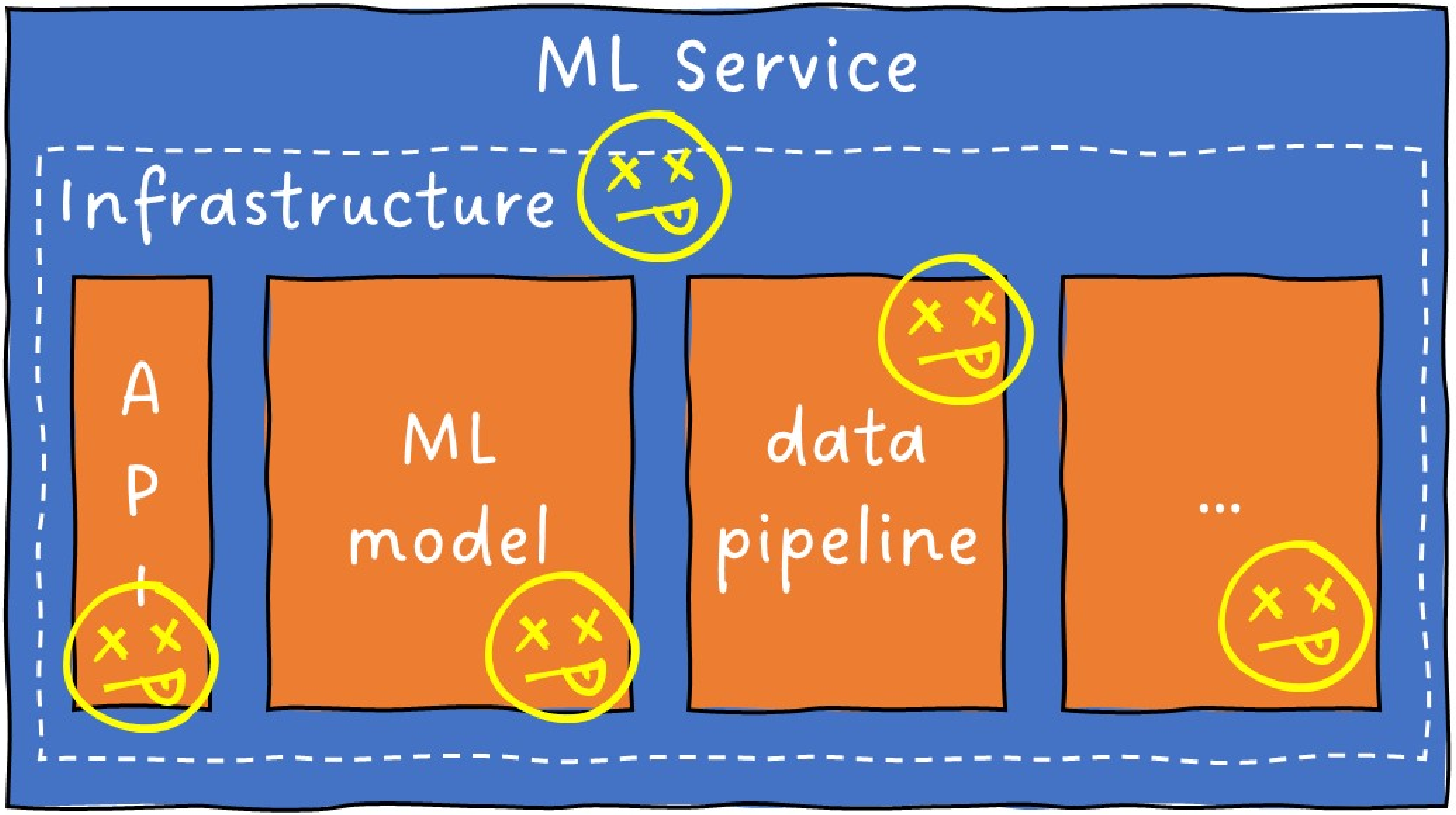
Nemanja Radojkovic
Senior Machine Learning Engineer

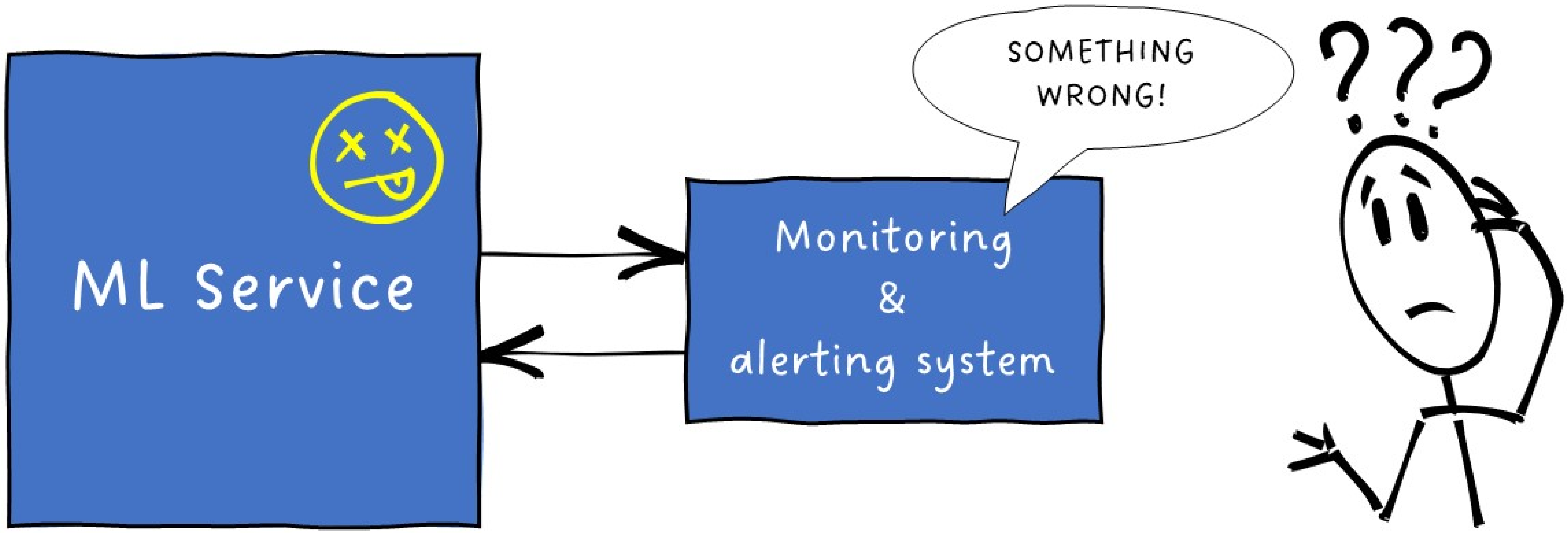

OUTSIDE
WORLD
ML Service

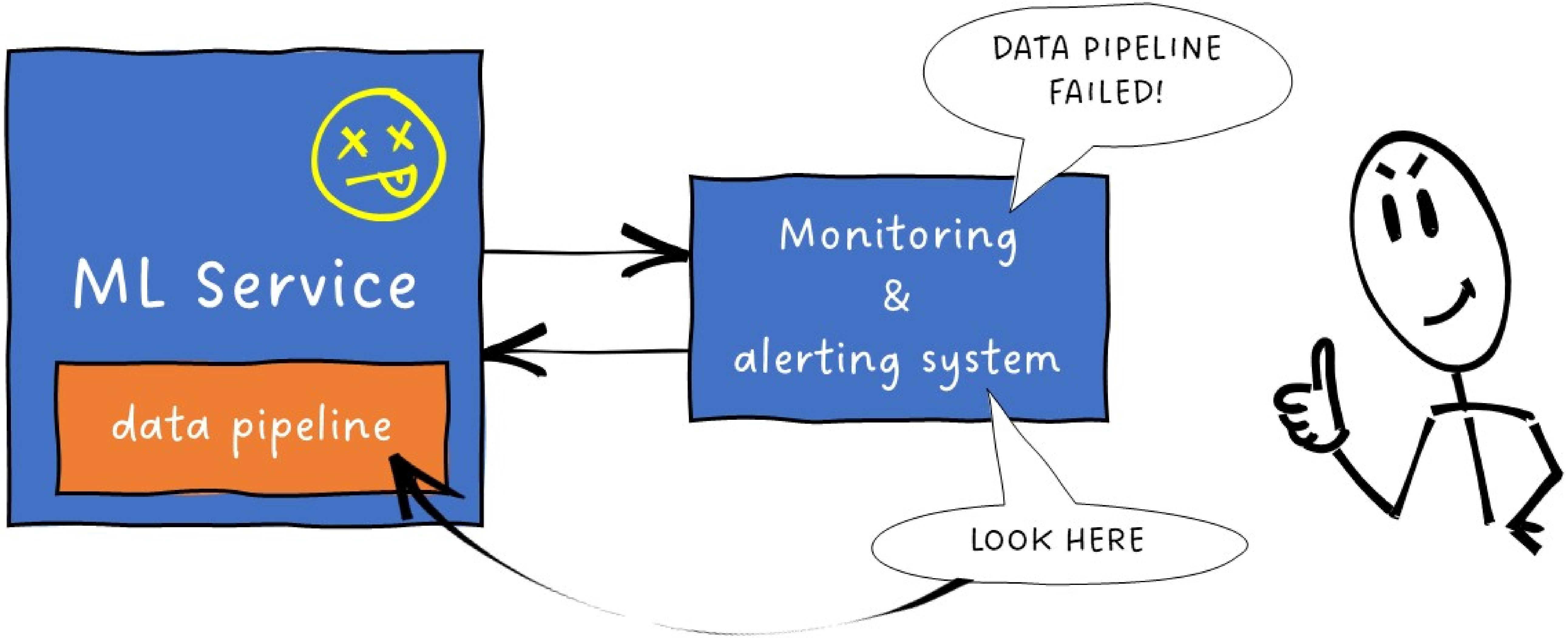

OUTSIDE WORLD
ML Service
Bugs in the code
Data errors
Service failures

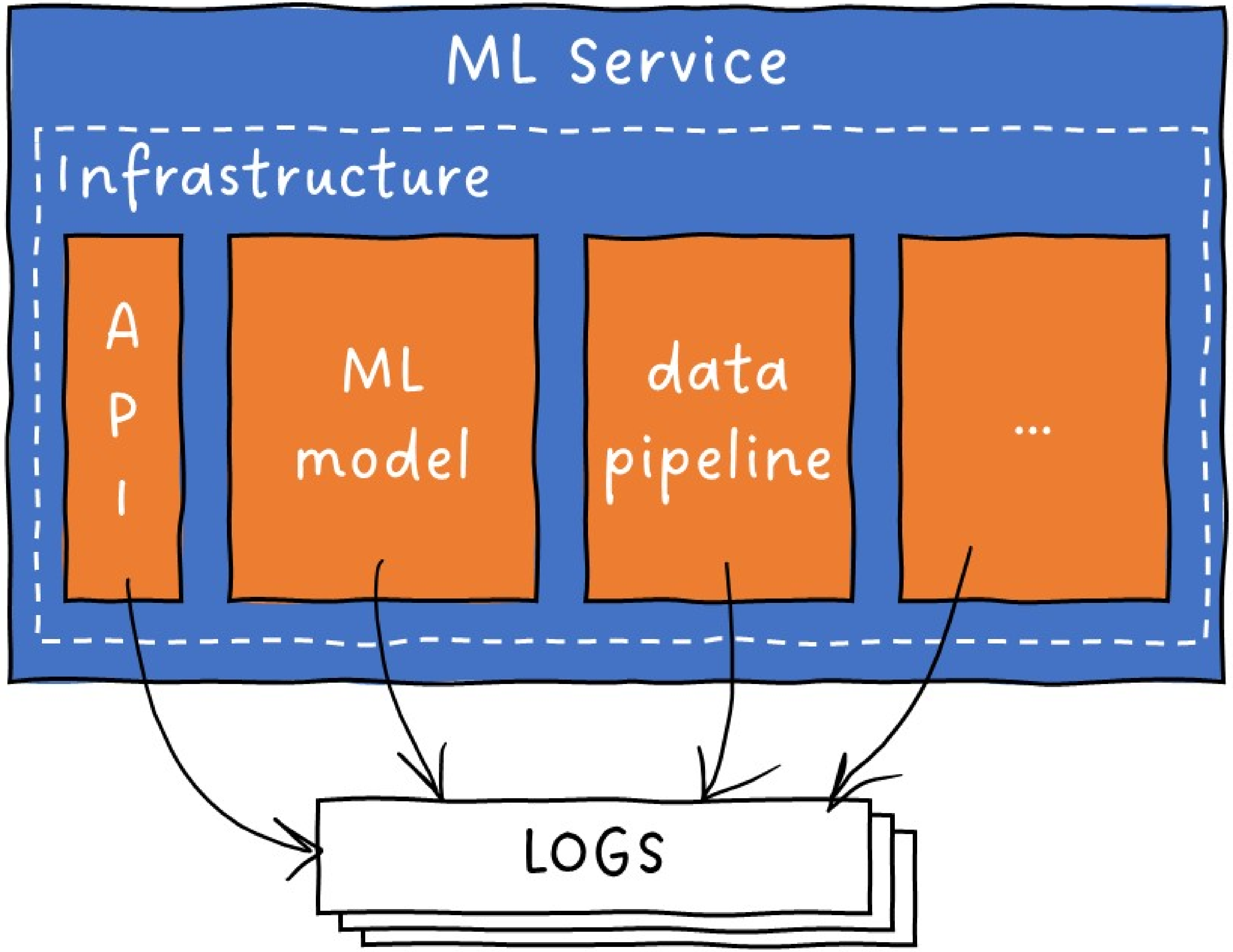

*monitoring &
alerting system
OUTSIDE
WORLD
ML Service
Bugs in the code
Data errors
Service failures

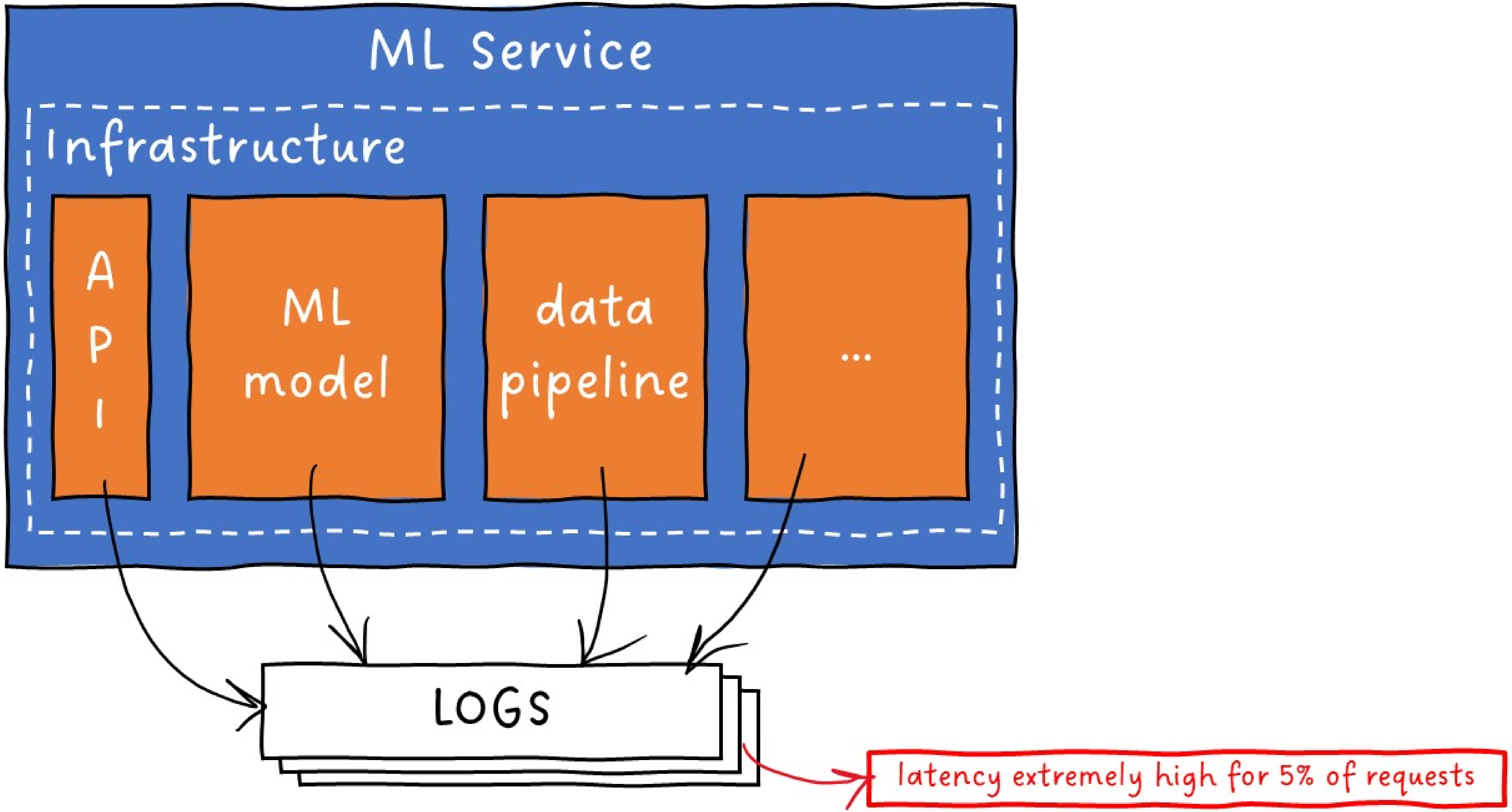

ML Service
Infrastructure
A
P
I
ML
model
data
pipeline
...

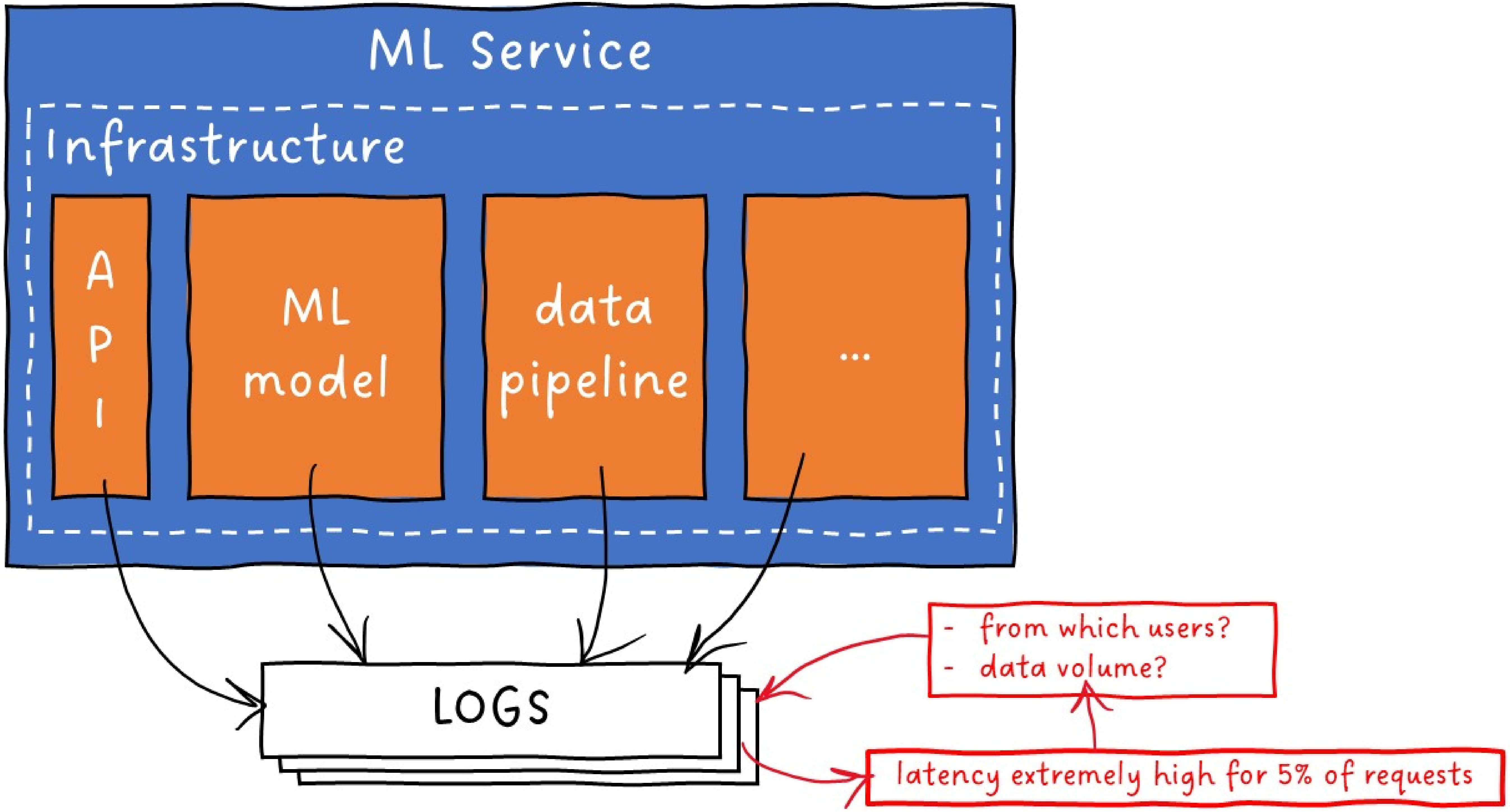

ML Service
Infrastructure
API
ML model
data pipeline
...

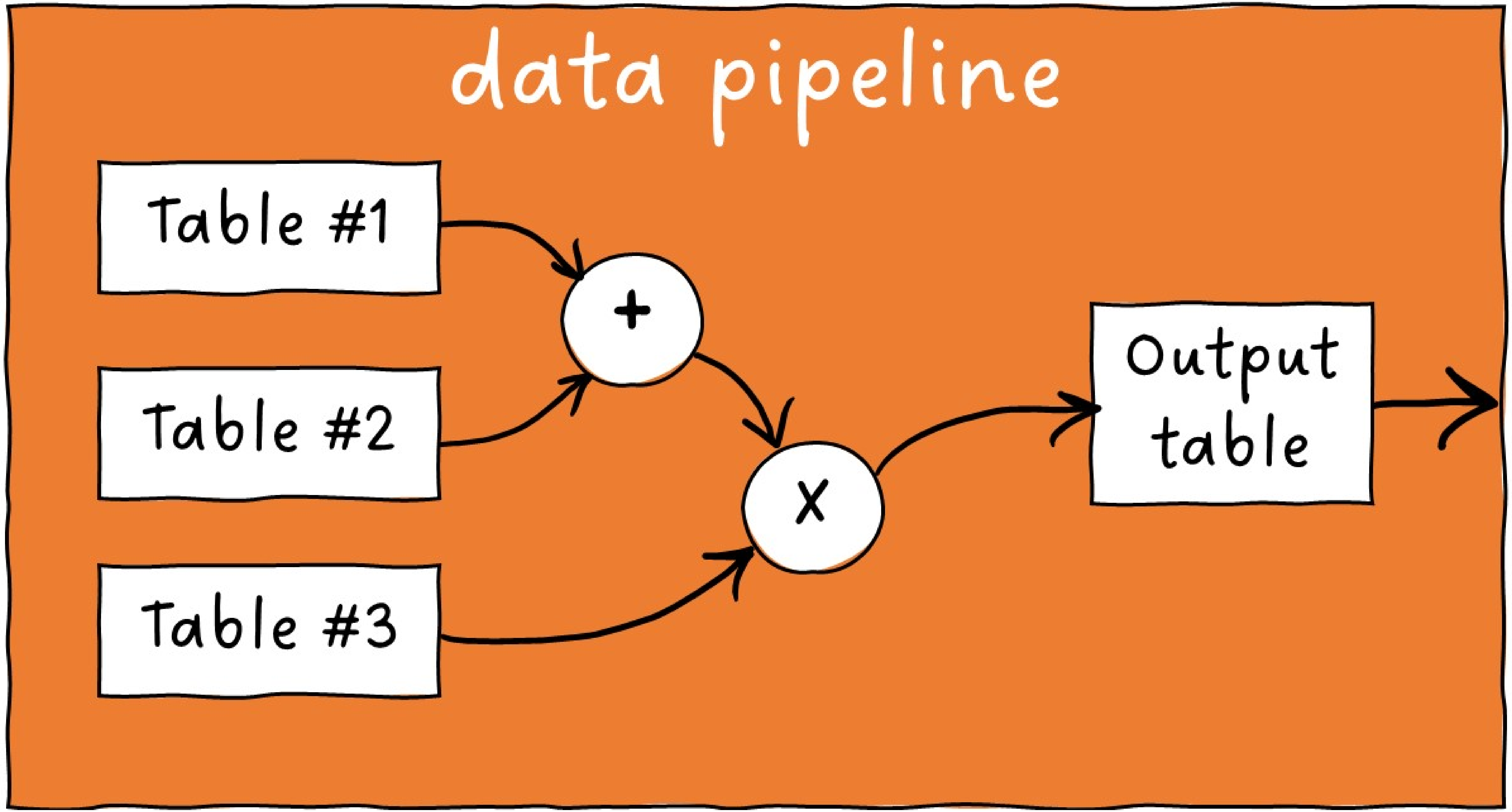

ML Service
Monitoring
&
alerting system
SOMETHING
WRONG!
???

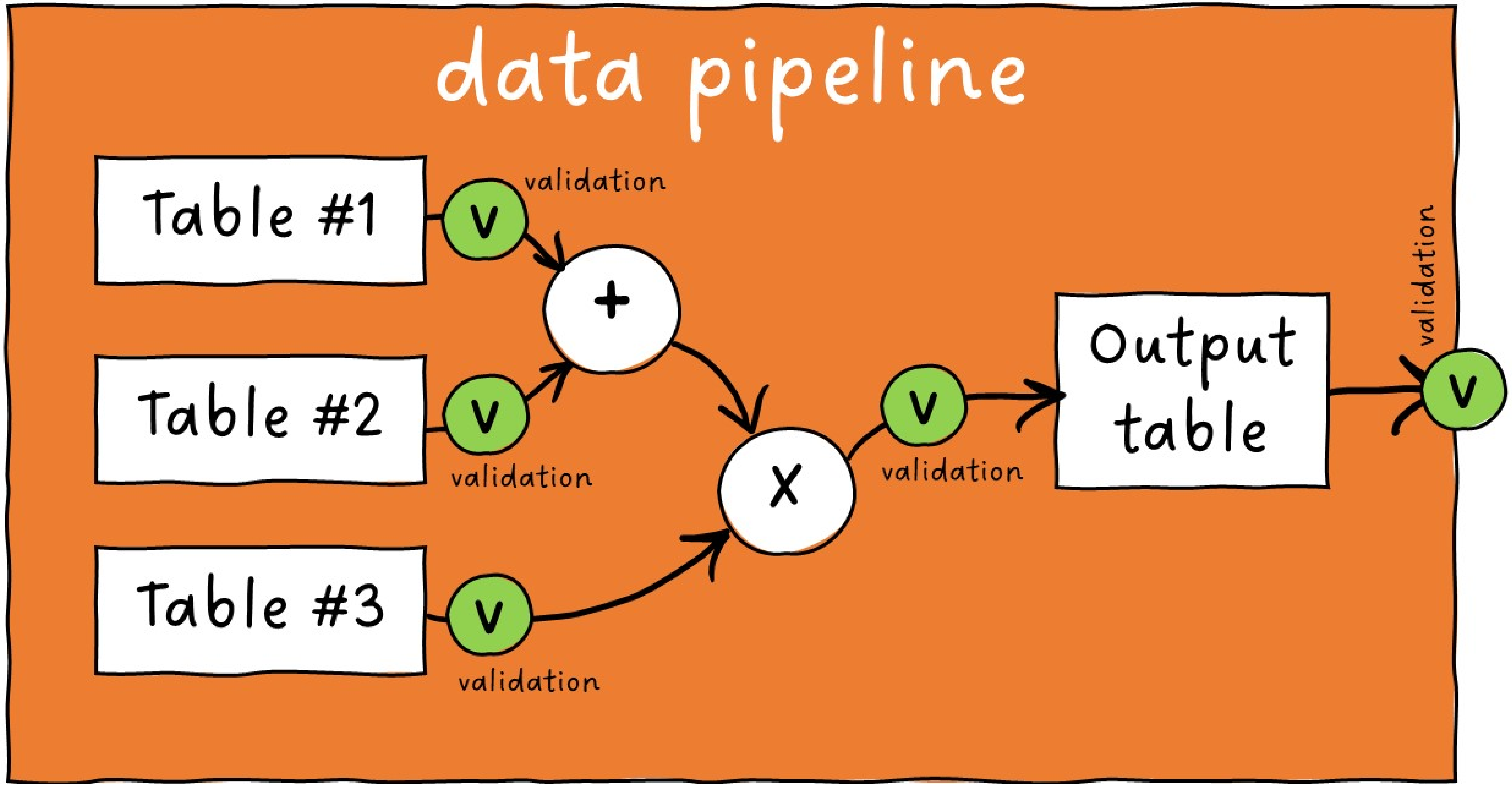

ML Service
data pipeline
Monitoring
&
alerting system
DATA PIPELINE
FAILED!
LOOK HERE

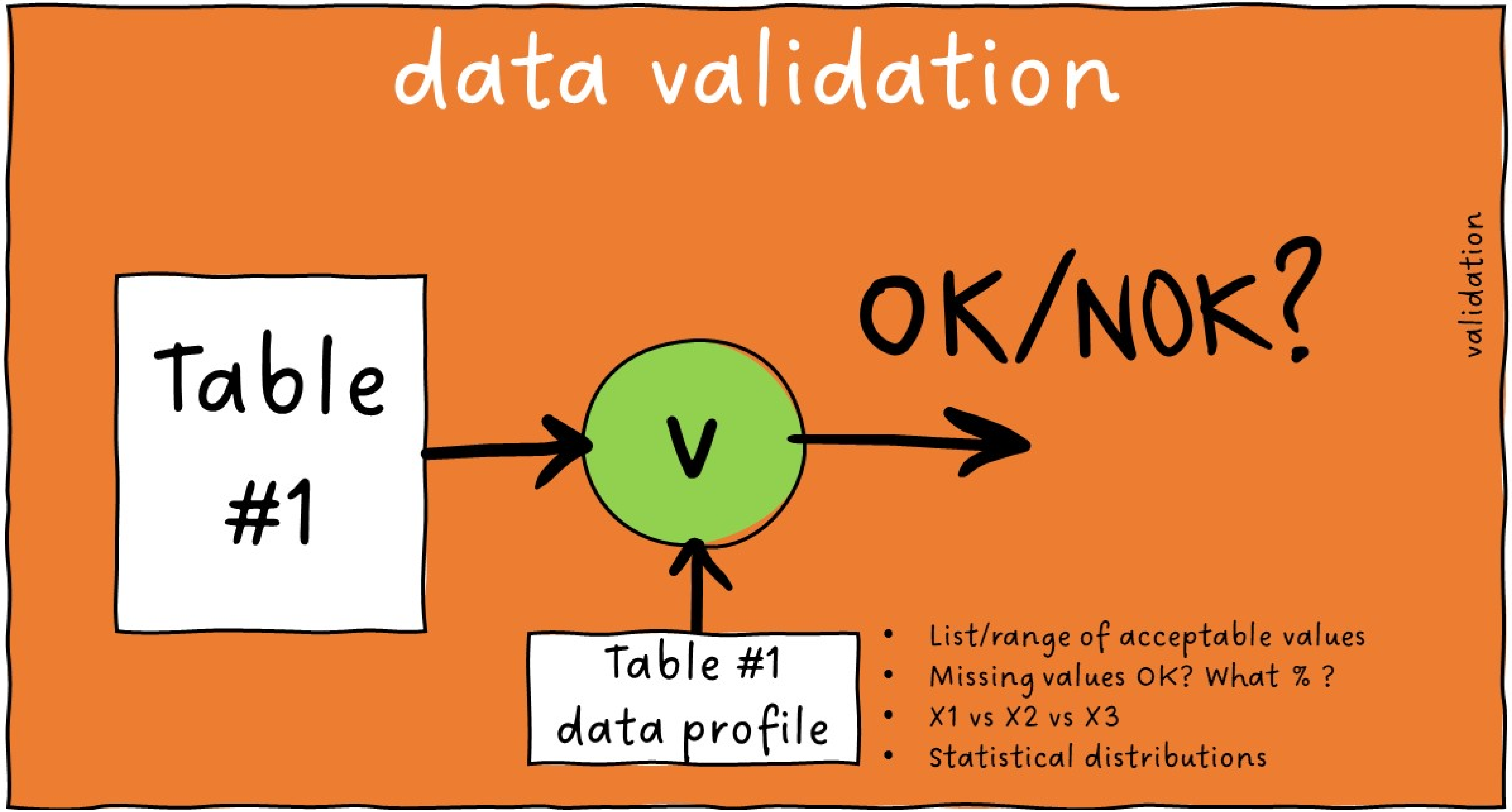

ML Service
Infrastructure
A P I
ML model
data pipeline
...
LOGS


ML Service
Infrastructure
API
ML model
data pipeline
...
LOGS
latency extremely high for 5% of requests


ML Service
Infrastructure
API
ML model
data pipeline
...
LOGS
- from which users?
- data volume?
latency extremely high for 5% of requests


data pipeline
Table #1
Table #2
Table #3
+
X
Output table


data pipeline
Table #1
Table #2
Table #3
V
validation
+
x
Output table


data validation
Table #1
V
OK/NOK?
validation
Table #1 data profile
• List/range of acceptable values
• Missing values OK? What % ?
• X1 vs X2 vs X3
• Statistical distributions

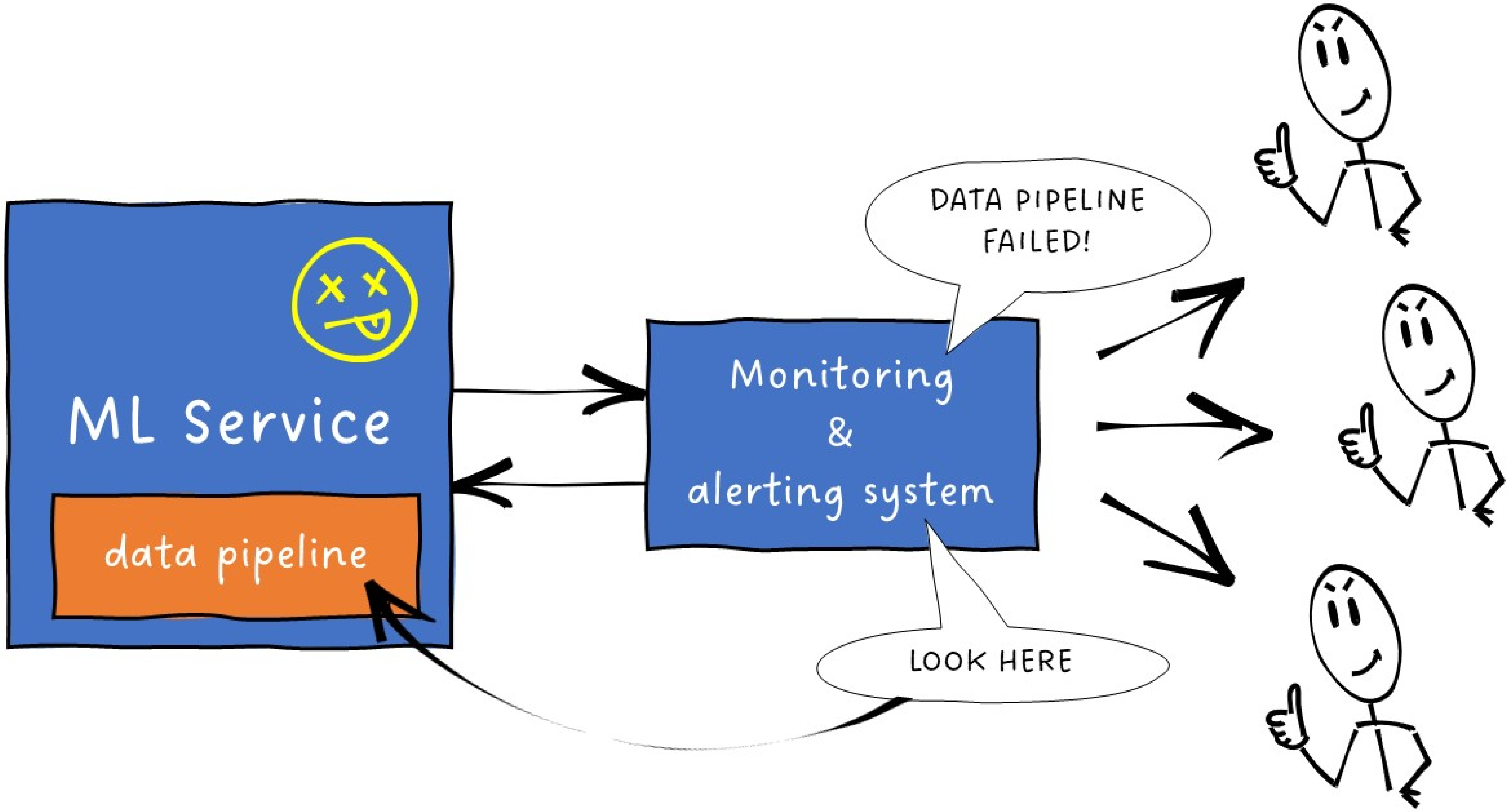
Statistical validation

Can be:

- too sensitive
- not informative enough

Risk

- Too many alerts
- "Alert fatigue"
- Important alerts going unnoticed


ML Service
data pipeline
Monitoring
&
alerting system
DATA PIPELINE
FAILED!
LOOK HERE

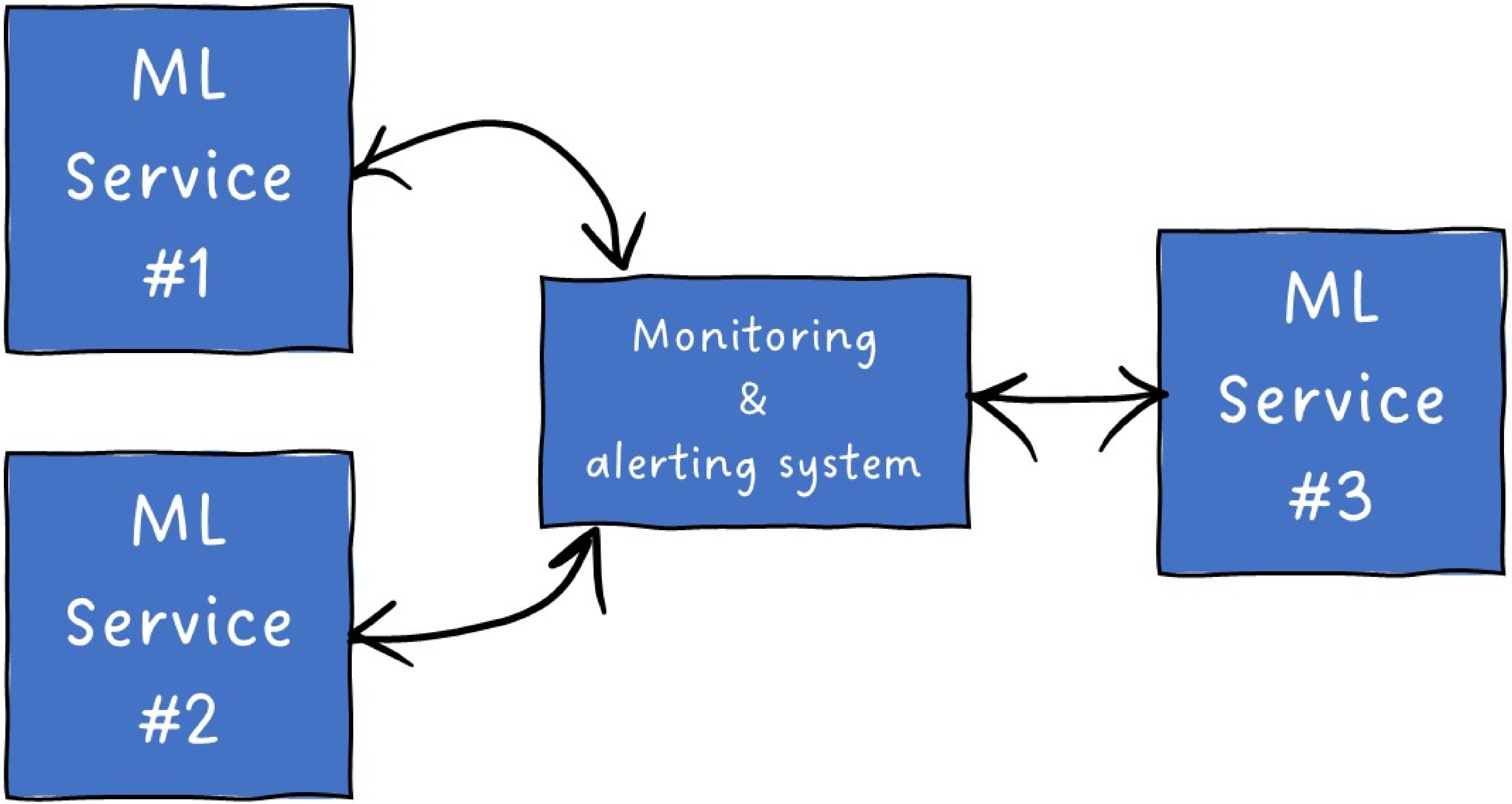
Learn from your history

After treating the incident => Record root cause and resolution steps

Example from Google[1]:

- 10 years of incidents recorded an analyzed
- > 2/3 were not ML-related!

[1] How ML Breaks: A Decade of Outages for One Large ML Pipeline, https://www.usenix.org/conference/opml20/presentation/papasian


ML
Service
#1
Monitoring
&
alerting system
ML
Service
#3
ML
Service
#2

Let's practice!

MLOPS DEPLOYMENT AND LIFE CYCLING

Model maintenance

MLOPS DEPLOYMENT AND LIFE CYCLING

Nemanja Radojkovic
Senior Machine Learning Engineer

Previously...

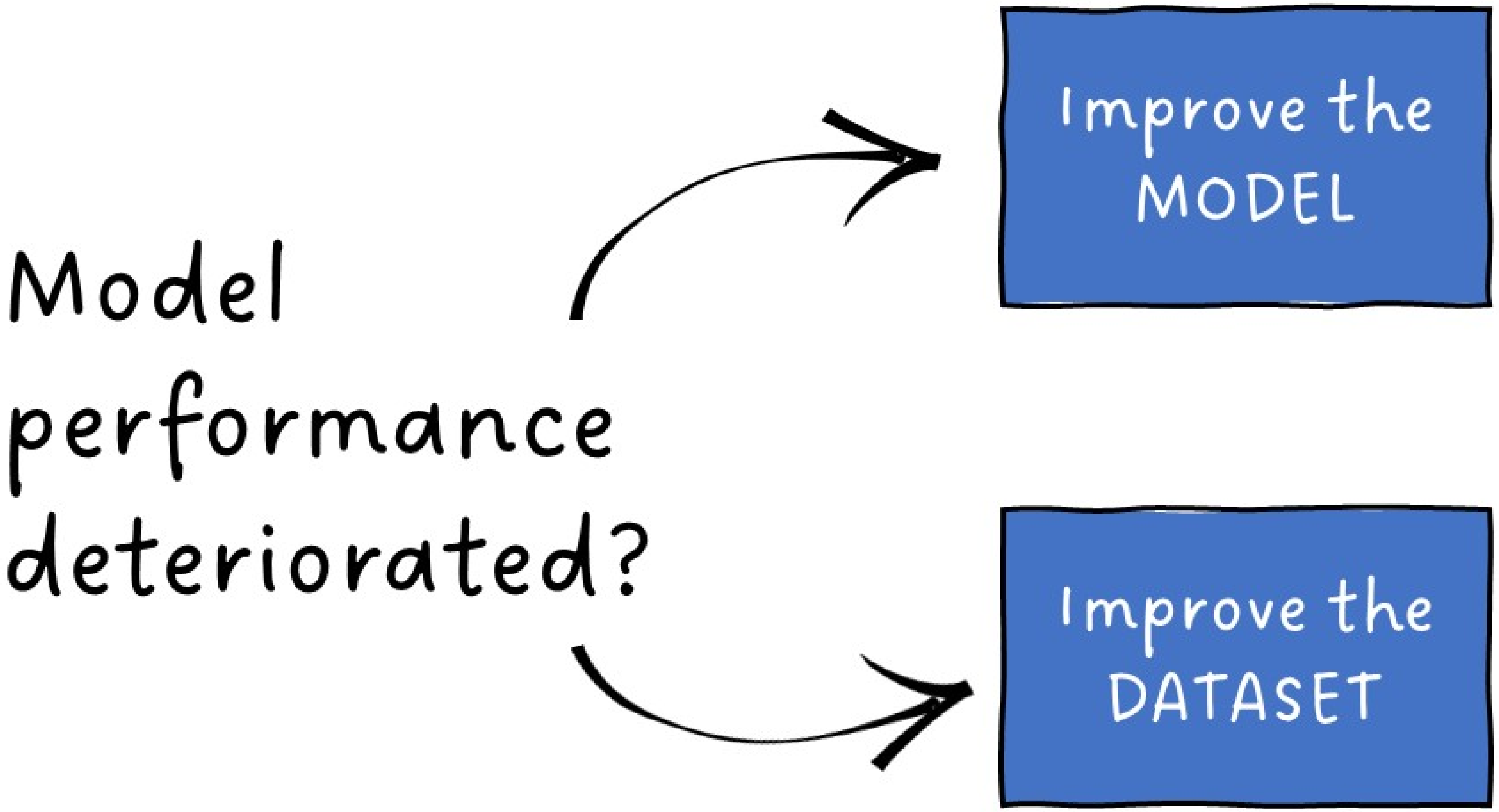
- How to catch anomalies and disturbances?

Focus of this lesson: ML model failure

Model
performance
deteriorated?

Model performance deteriorated?

Improve the MODEL


Model performance deteriorated?
Improve the MODEL
Improve the DATASET

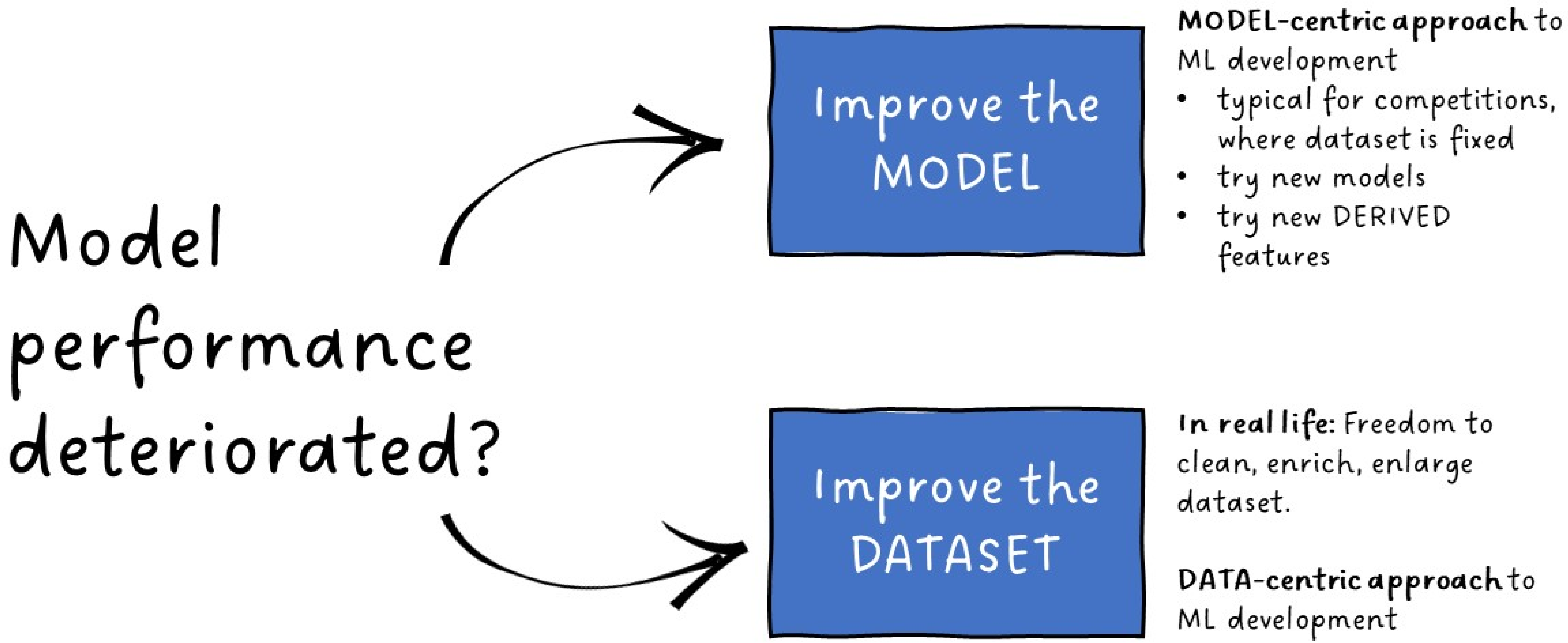
Model performance deteriorated?

Improve the MODEL

MODEL-centric approach to ML development

- typical for competitions, where dataset is fixed
- try new models
- try new DERIVED features

Improve the DATASET



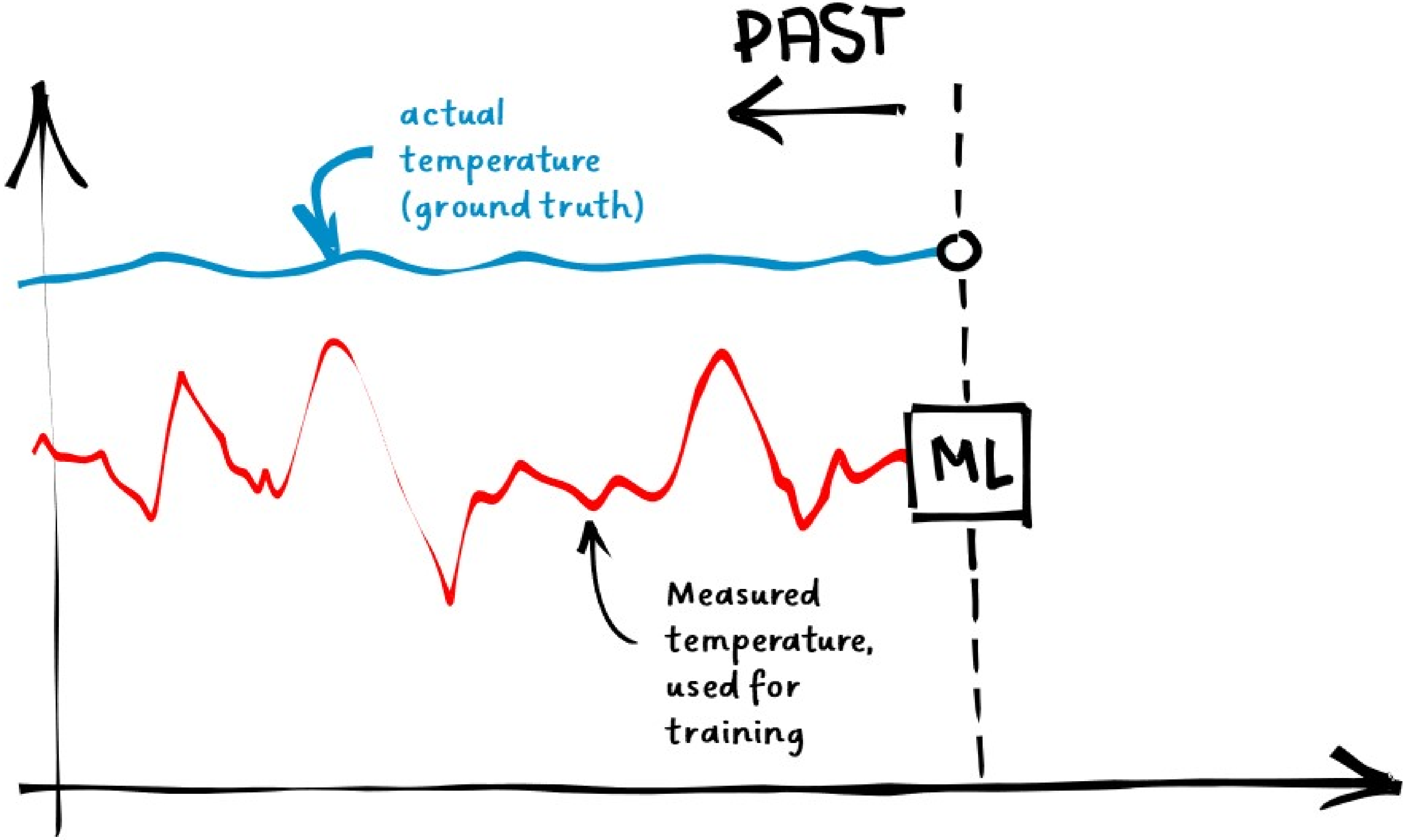
[1] https://datacentricai.org/

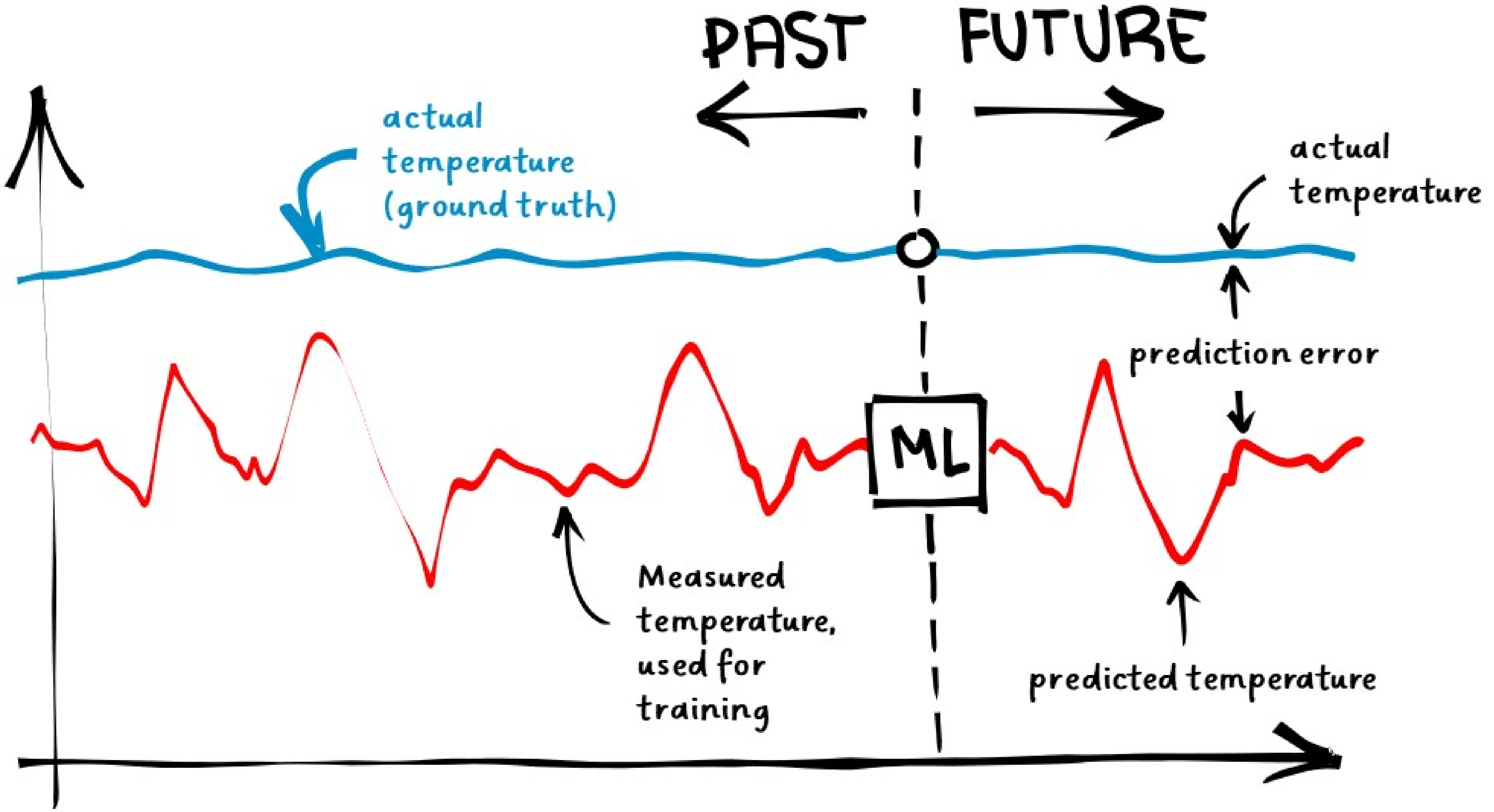
Quality above quantity

- Get more features with relevant information
- Get better labels

Label == Target variable value in the training set

Label quality == label closeness to the ground truth


PAST
actual
temperature
(ground truth)
ML
Measured
temperature,
used for
training



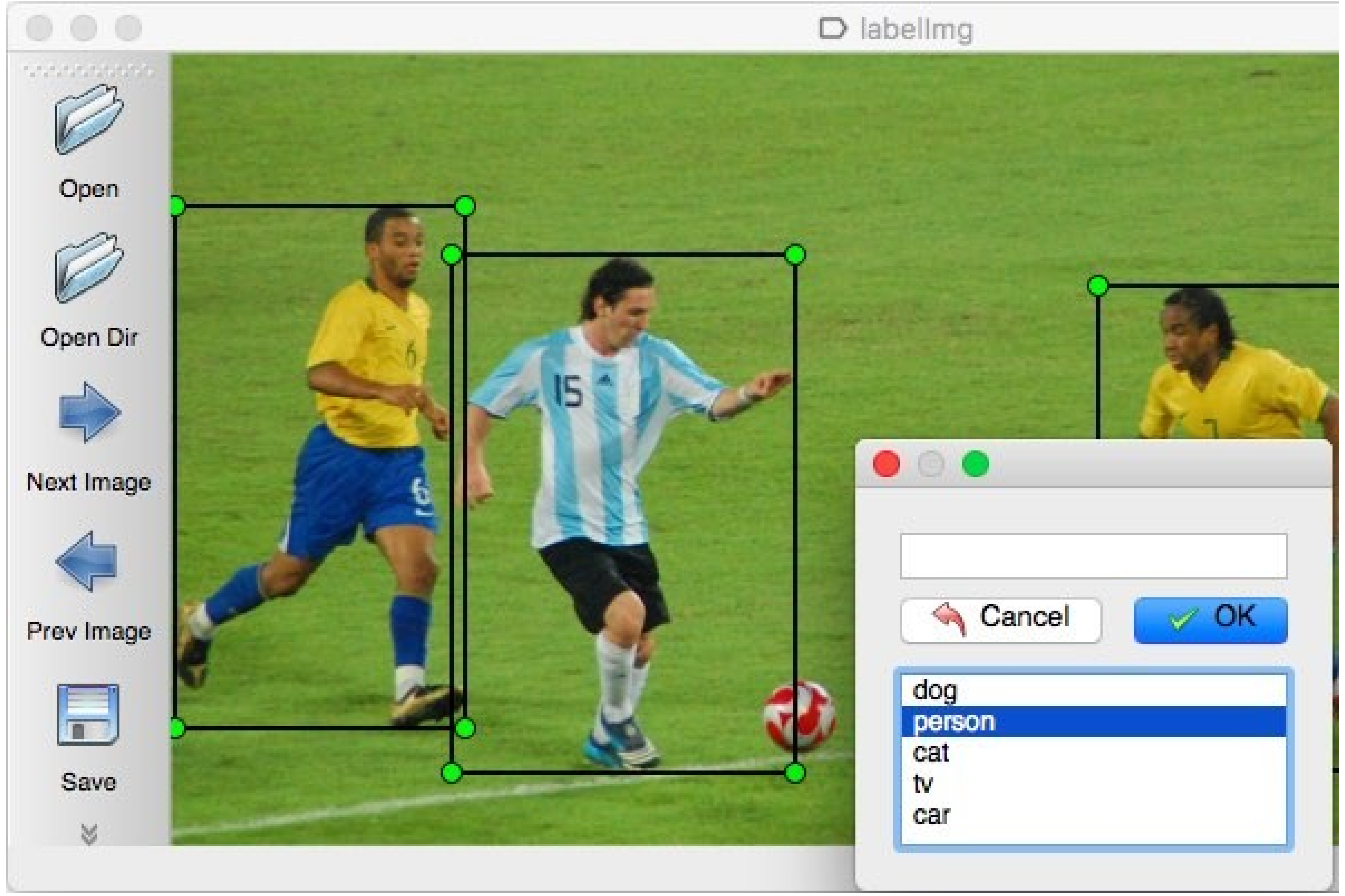
PAST
FUTURE
actual temperature (ground truth)
actual temperature
prediction error
ML
Measured temperature, used for training
predicted temperature

Benefits of labeling tools

Manual labeling is complex, lengthy, error-prone

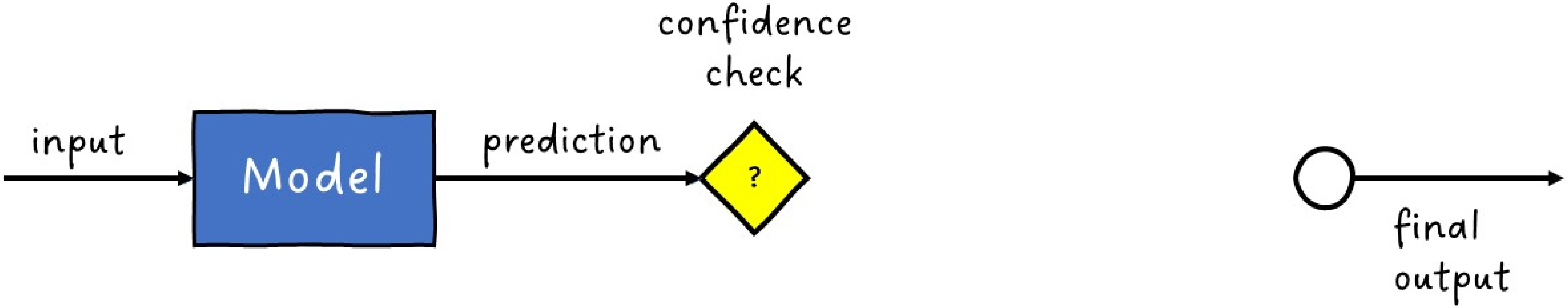
Use good labeling tools!

- Labeling more efficient
- More accurate
- User interface fit for purpose
- "Label these examples first for maximum impact"
- "It seems you made a mistake here"

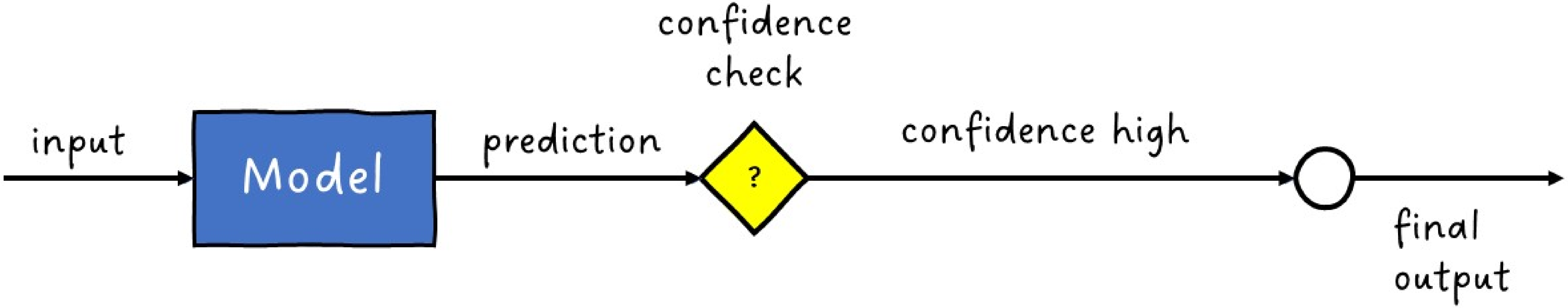


Example: Image labeling tool, for building image classifiers

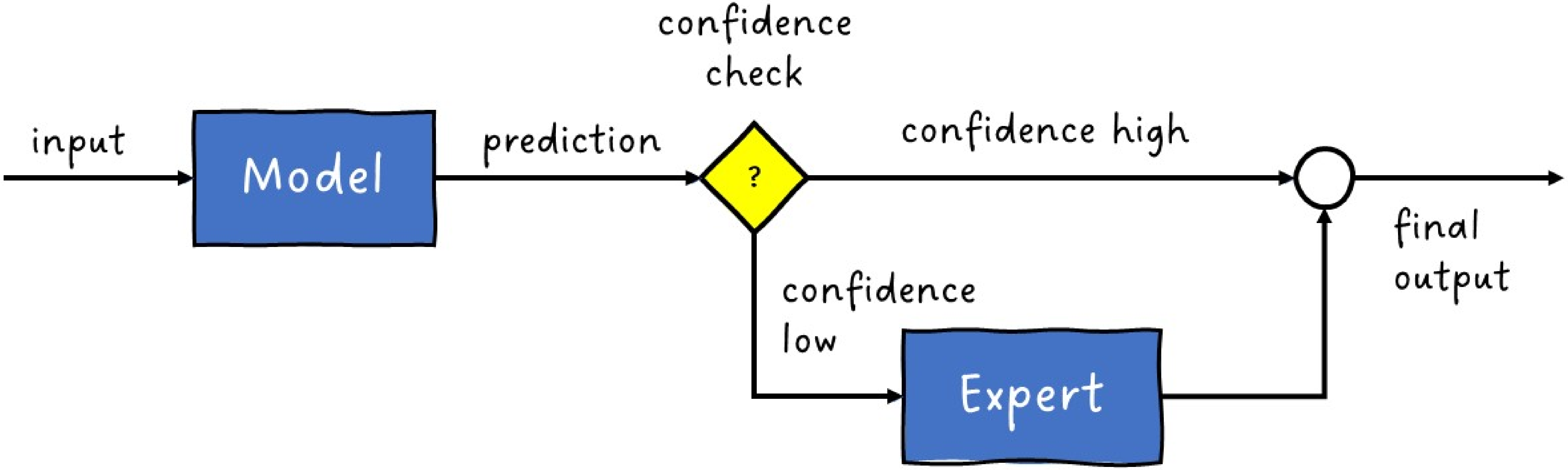
Human-in-the-loop system



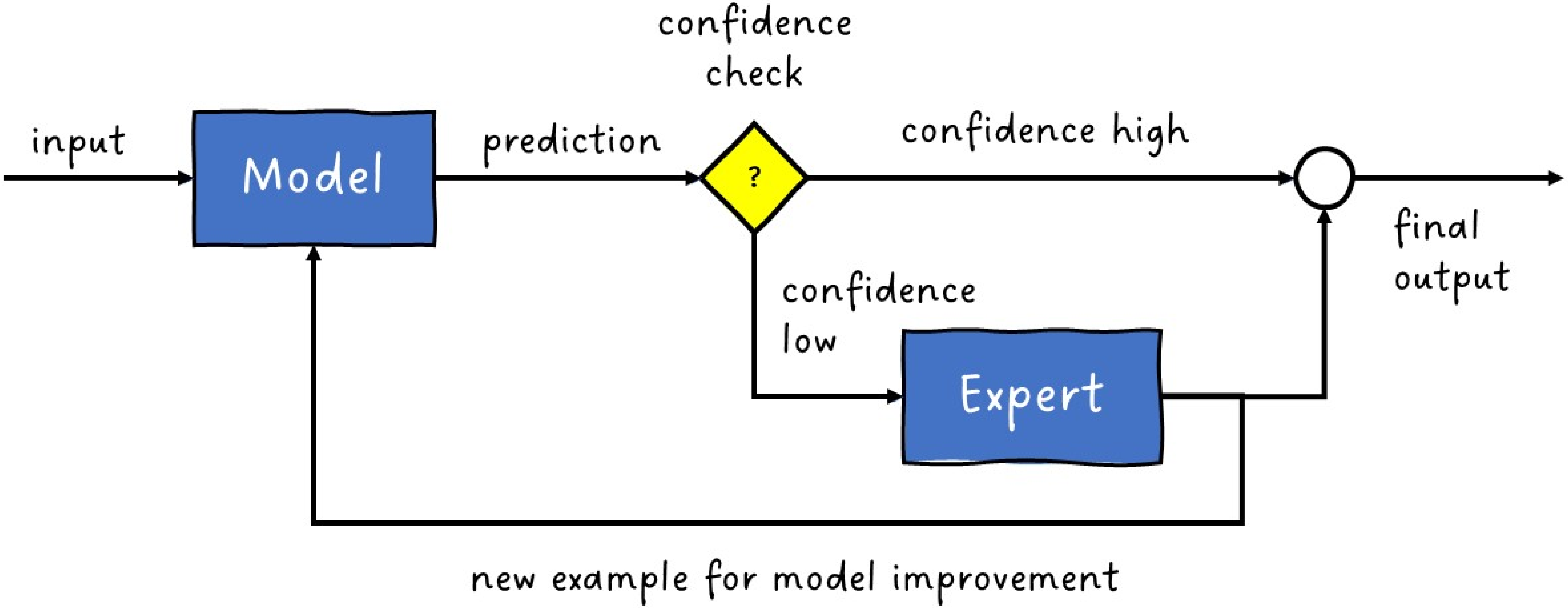
Human-in-the-loop system



Human-in-the-loop system



Human-in-the-loop system



When the labels arrive

- Run ML build pipeline => New model
- Check if new model is better than the old one
  - YES => test and deploy
  - NO => keep searching,

Keep searching

- Try new models
- Try new features
- Try new data sources

Immensely helpful: Metadata store (MLFlow Tracking, etc)

- Document the model selection journey
- Avoid repeating same experiments

In any case: *MLOps helps us maintain our model in the fastest, most efficient way*

Let's practice!

MLOPS DEPLOYMENT AND LIFE CYCLING

Model governance

MLOPS DEPLOYMENT AND LIFE CYCLING

Nemanja Radojkovic
Senior Machine Learning Engineer

NHS

ArtiQ Wins UK Government Funding for its AI Algorithms to Support Timely Diagnosis of Lung Diseases

by ArtiQ Jun 16, 2021

Business Of Law ; Internet Law & Cyber-Security

Law Bots: How AI Is Reshaping the Legal Profession

10 Min Read By: Matthew Stepka | February 21, 2022

2 minute read · November 30, 2022 5:29 PM GMT+1 · Last Updated a day ago

Honda to develop advanced level 3 self-driving technology by 2029

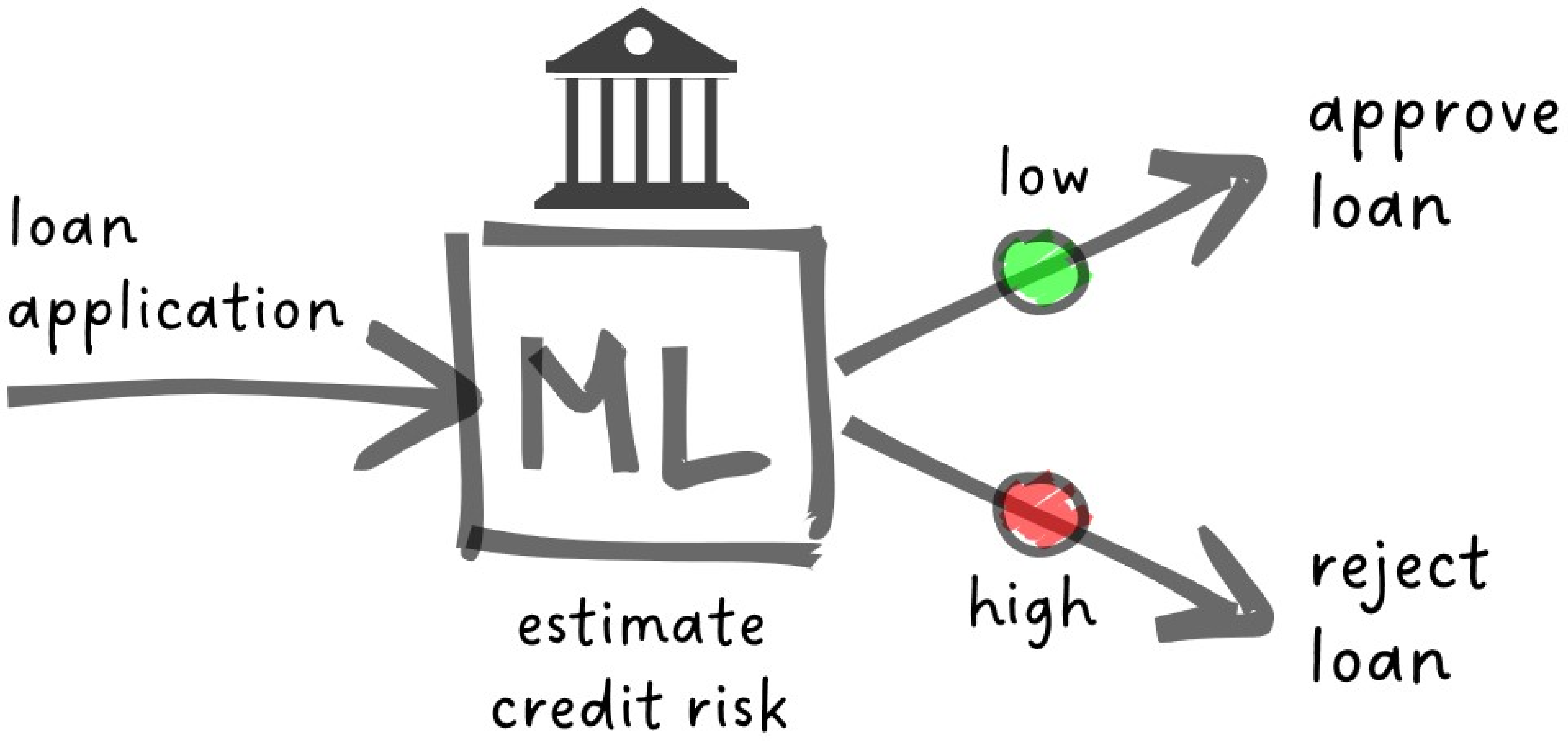
Reuters

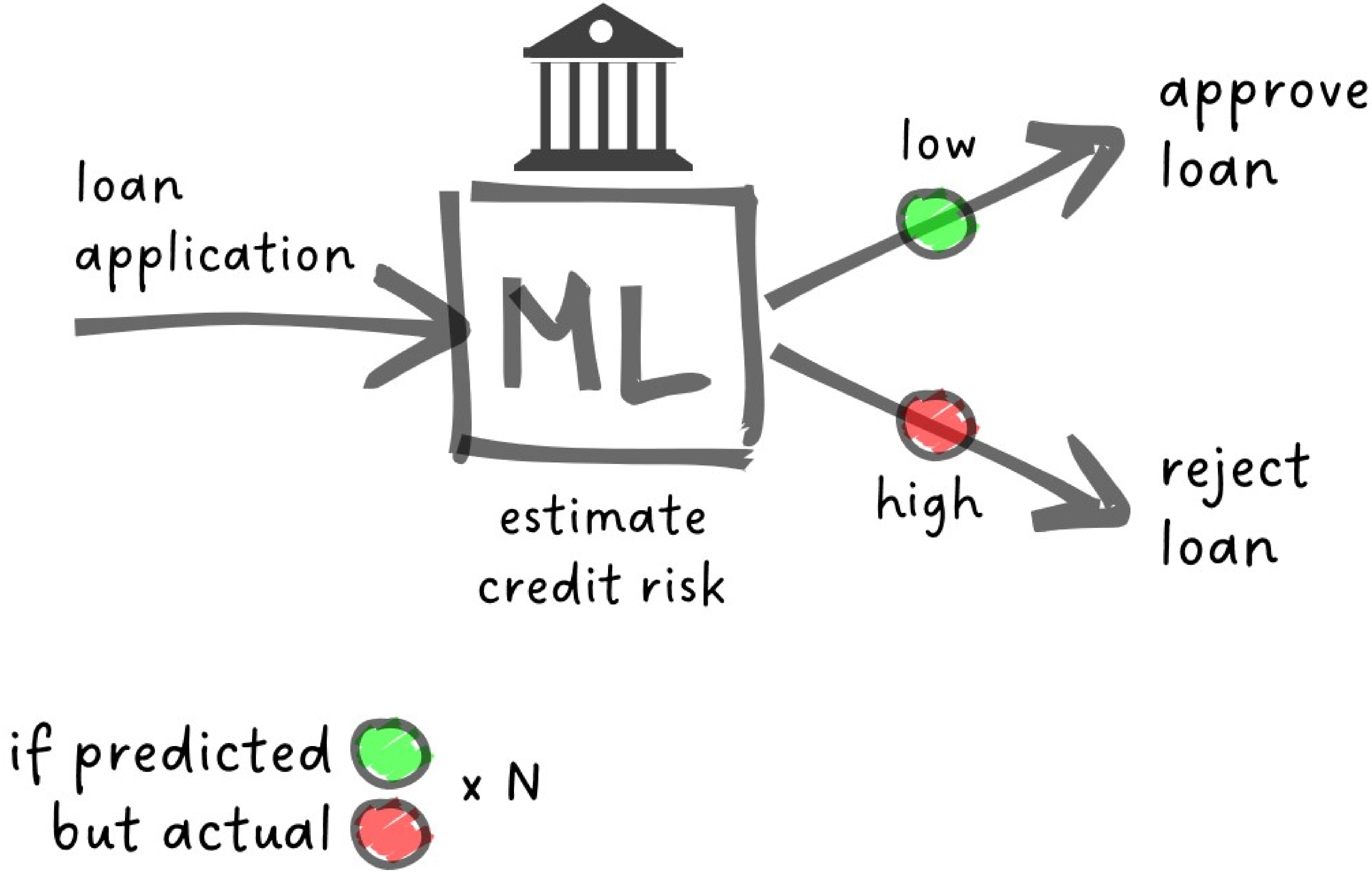
Artificial Intelligence and Machine Learning 9 min read

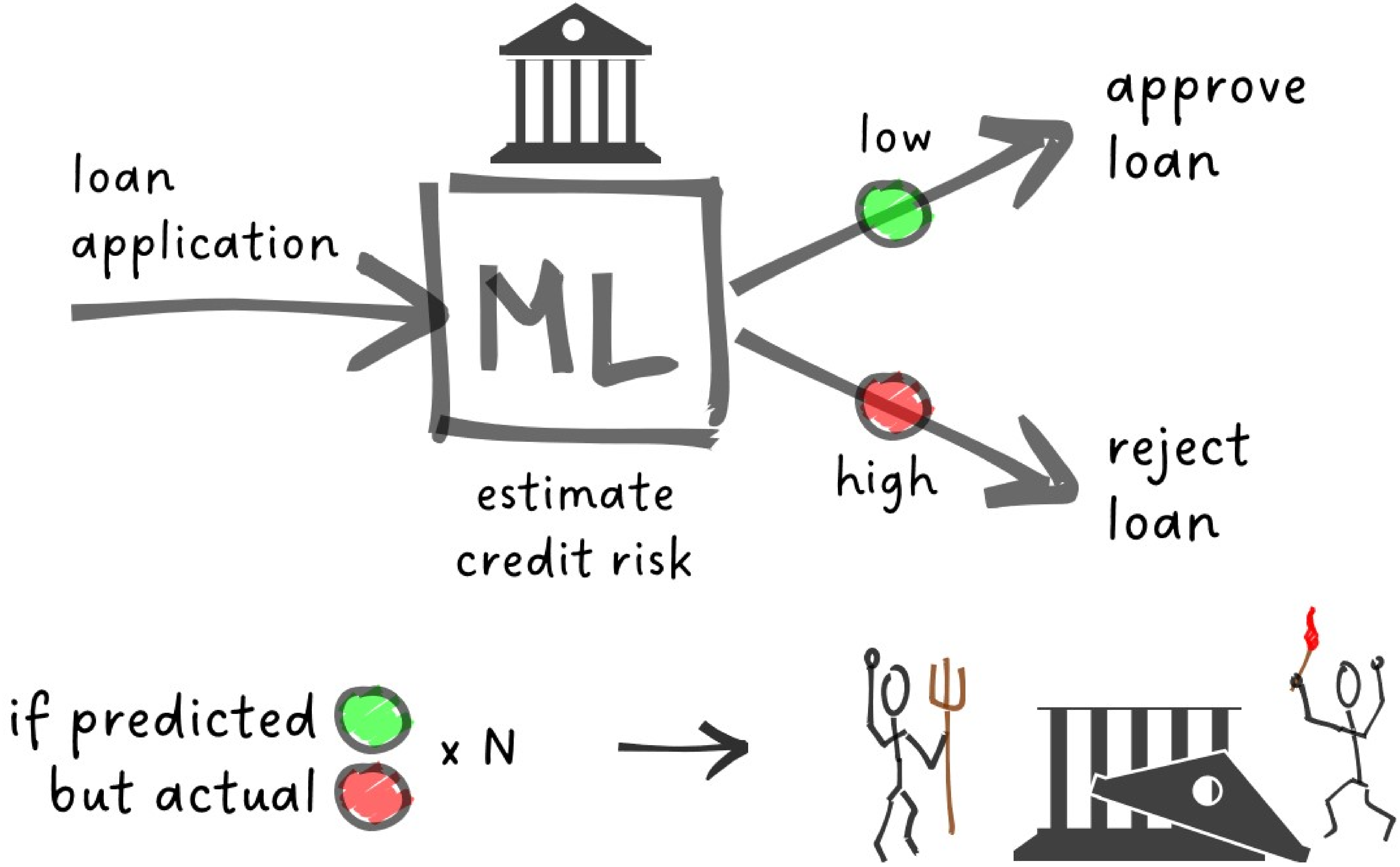
5 Ways AI is Transforming the Finance Industry

Check out the easy ways by which artificial intelligence can transform the finance industry.

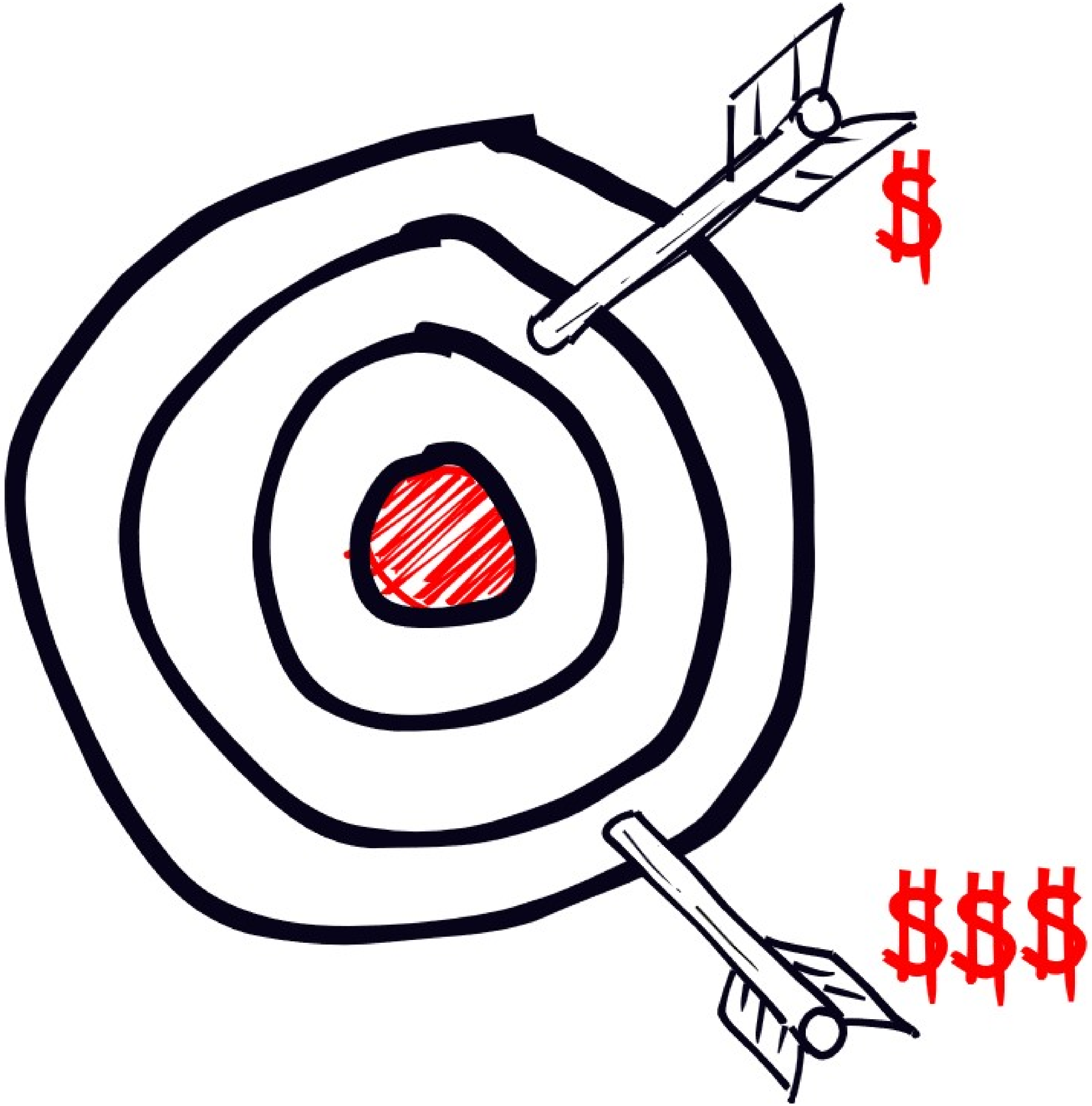
Pinakin Ariwala
Updated on Oct 14

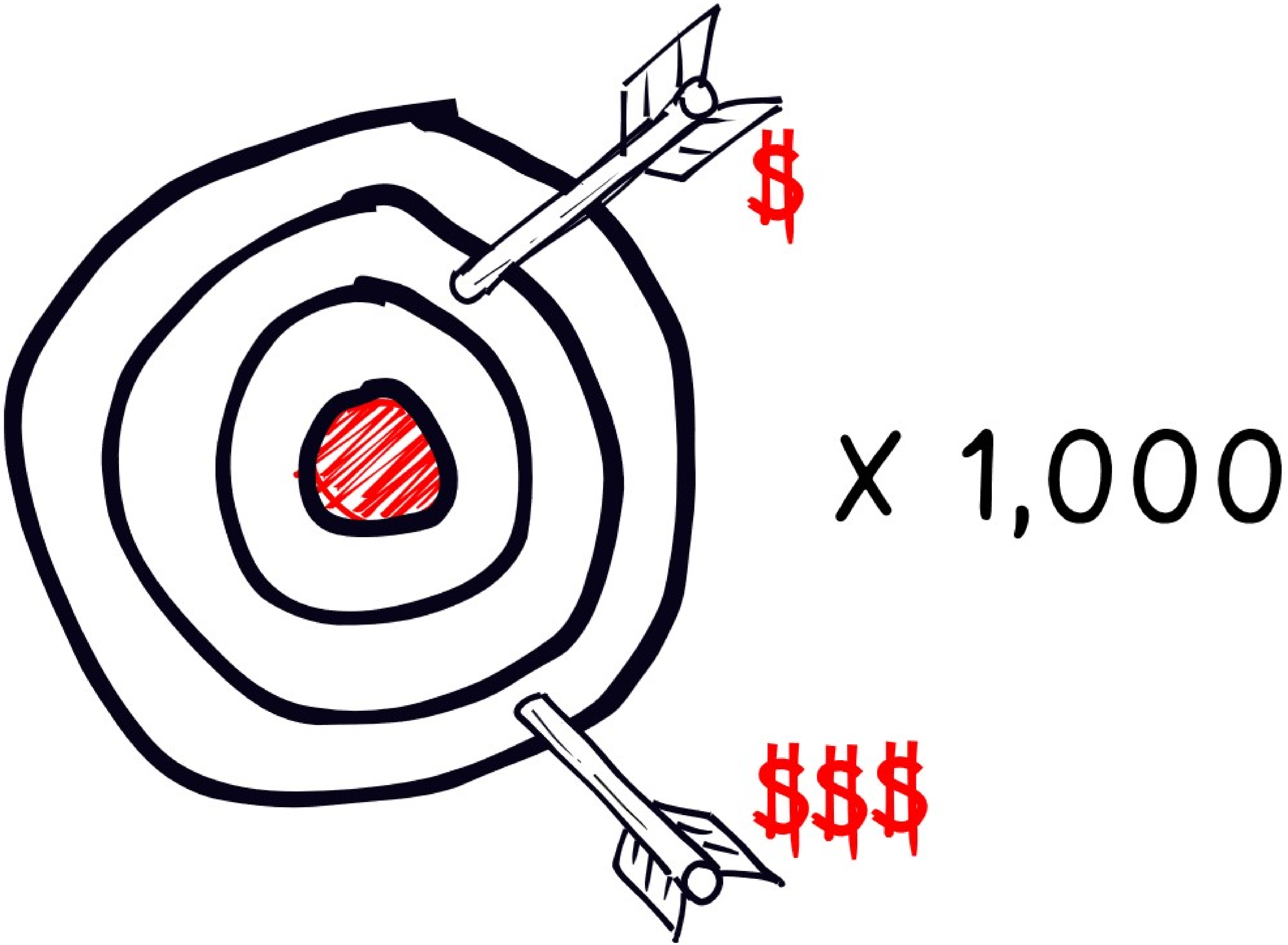

loan
application
ML
estimate
credit risk
low
high
approve
loan
reject
loan


loan
application
ML
estimate
credit risk
low
high
approve
loan
reject
loan
if predicted
but actual
x N



loan application
ML
estimate credit risk
low
approve loan
high
reject loan
if predicted
but actual
x N



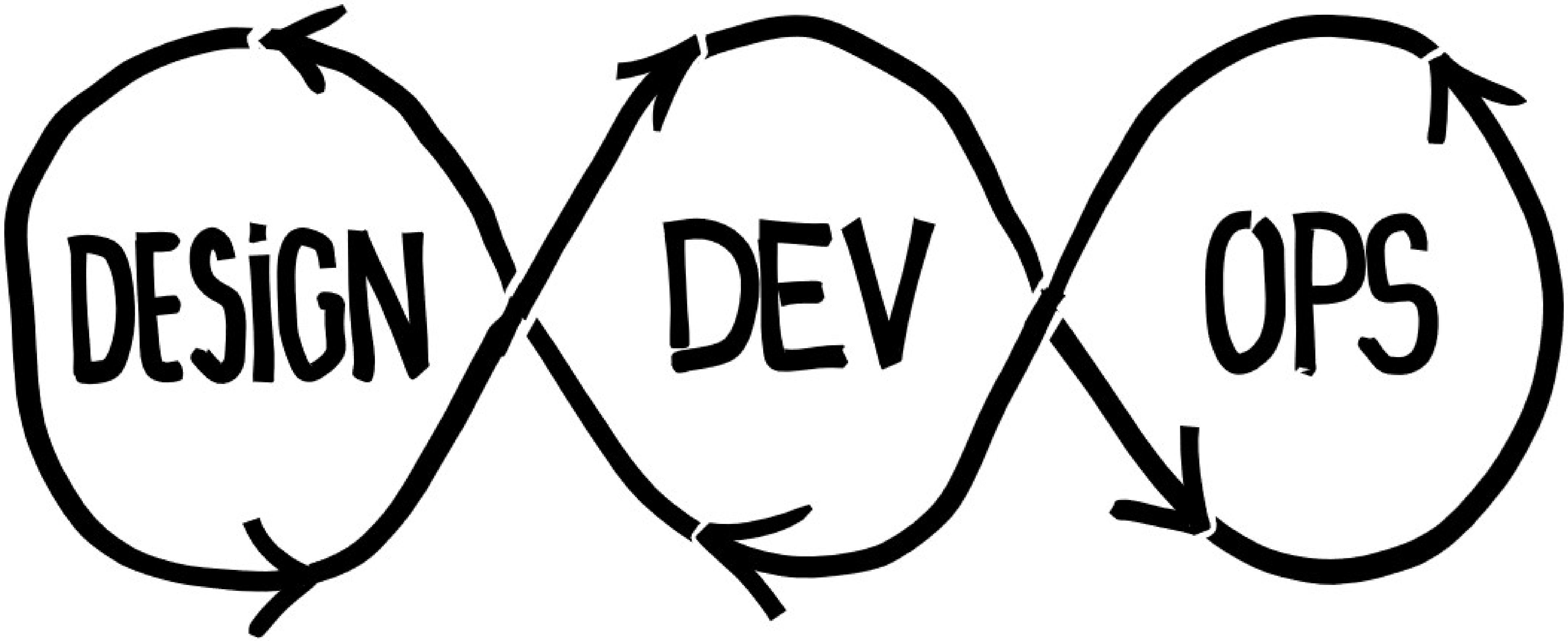

$
$$$
X 1,000

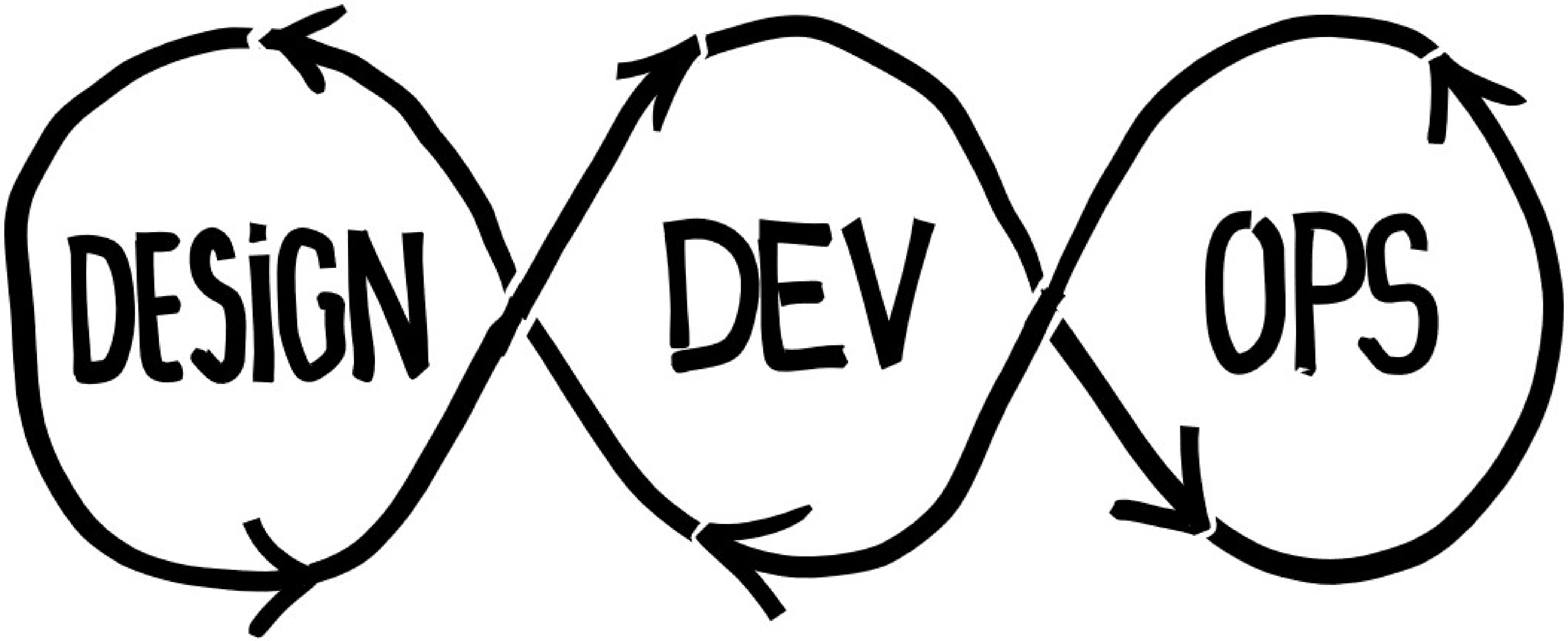
Governance

> AI/ML model governance is the overall process for how an organization controls access, implements policy, and tracks activity for models and their results.

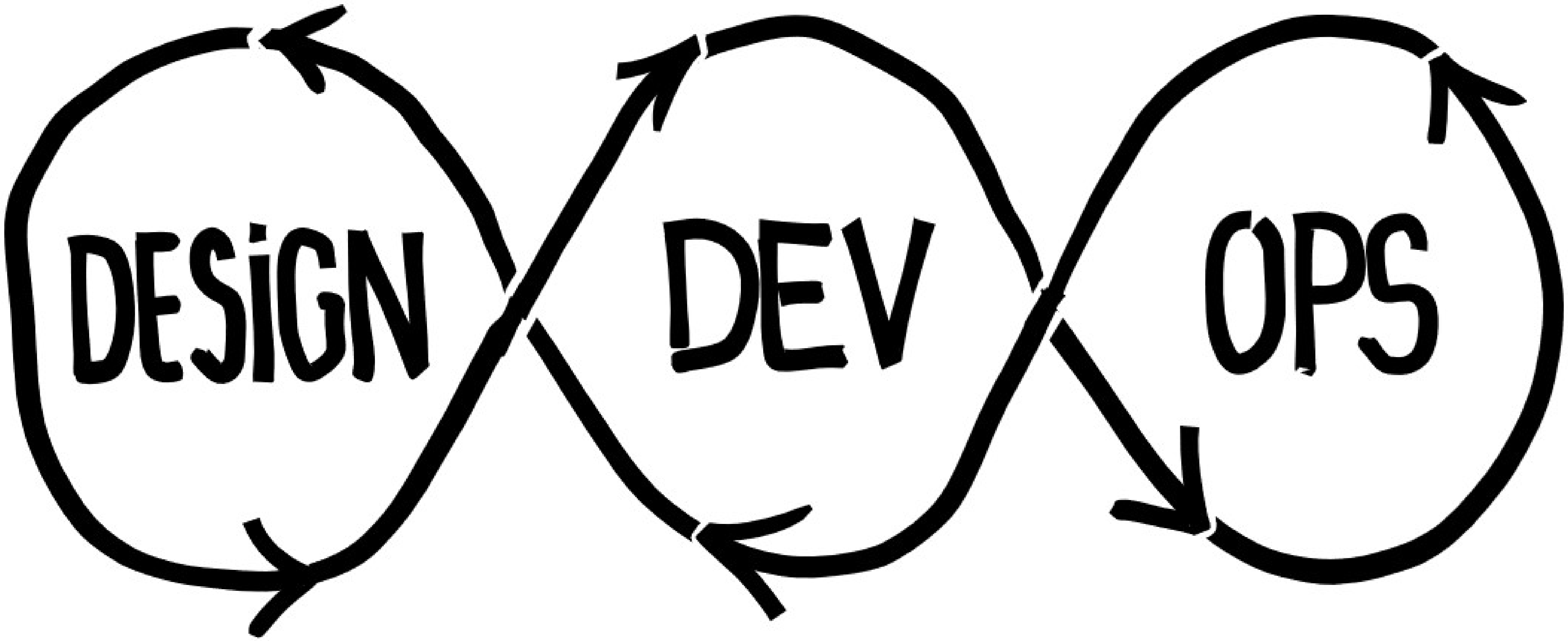
> Effective model governance is the bedrock for minimizing risk to both an organization's bottom line and to its brand.

~ *DataRobot.com*


DESiGN
DEV
OPS

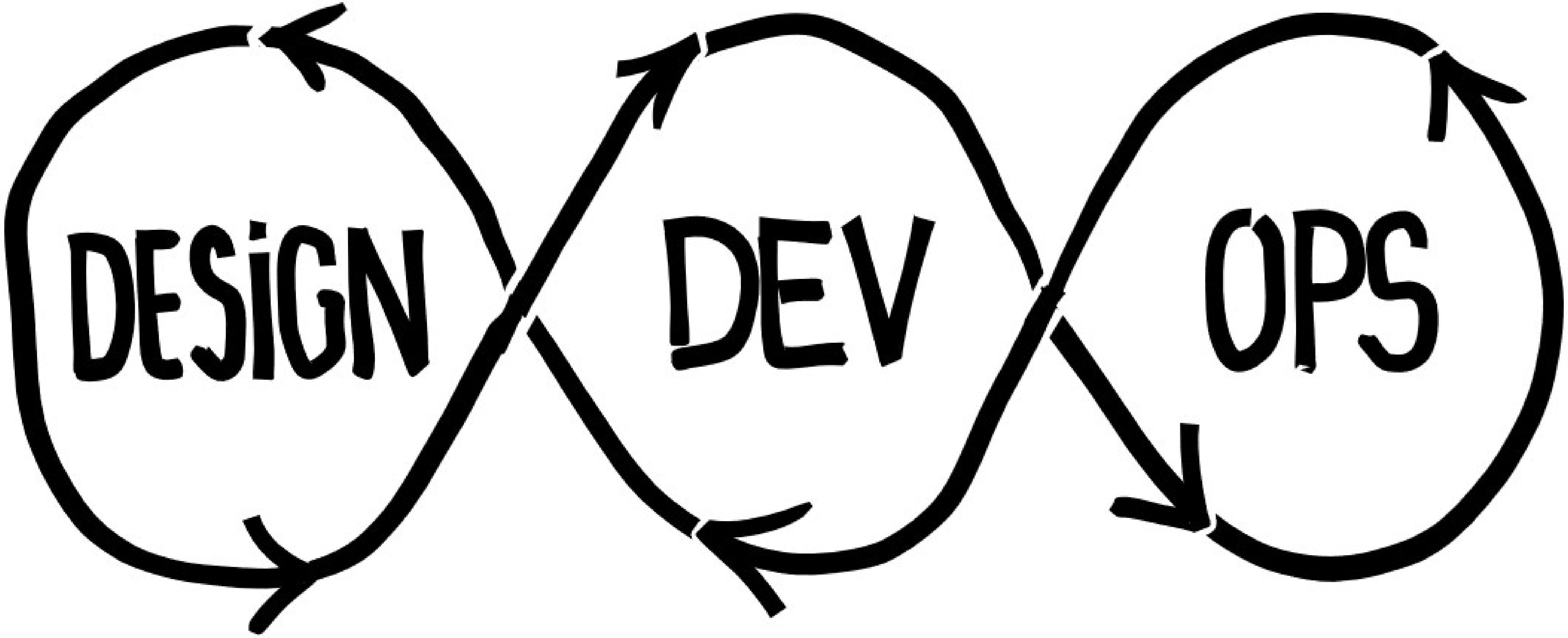


- Is it ethical to use ML?
- Private, sensitive data?
- Unethical model bias?
- ...



- Is it ethical to use ML?
- Private, sensitive data?
- Unethical model bias?
- ...

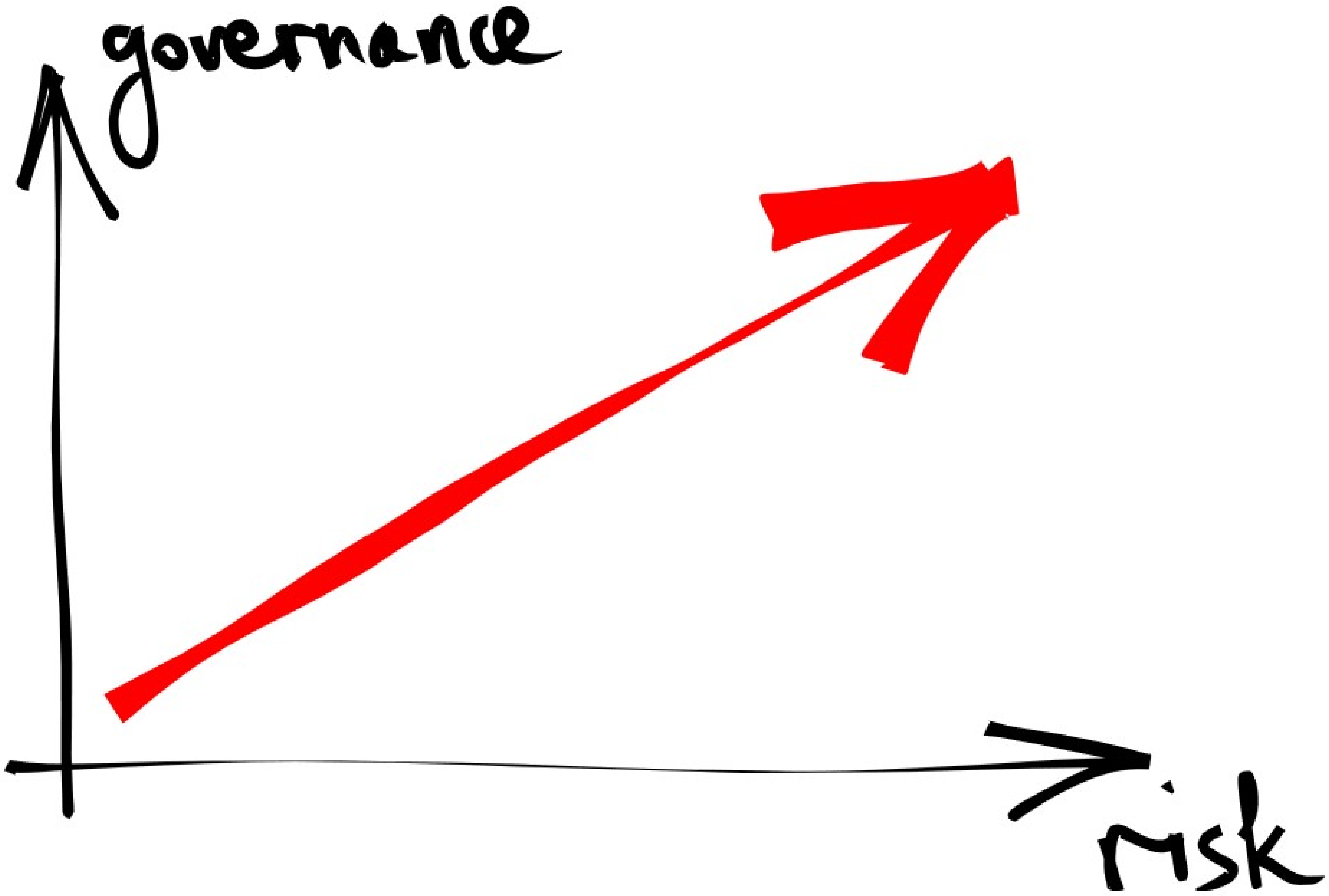
- Model selection docs
- Training data prep
- Quality assurance
- Versioning
- Reproducibility
- ...

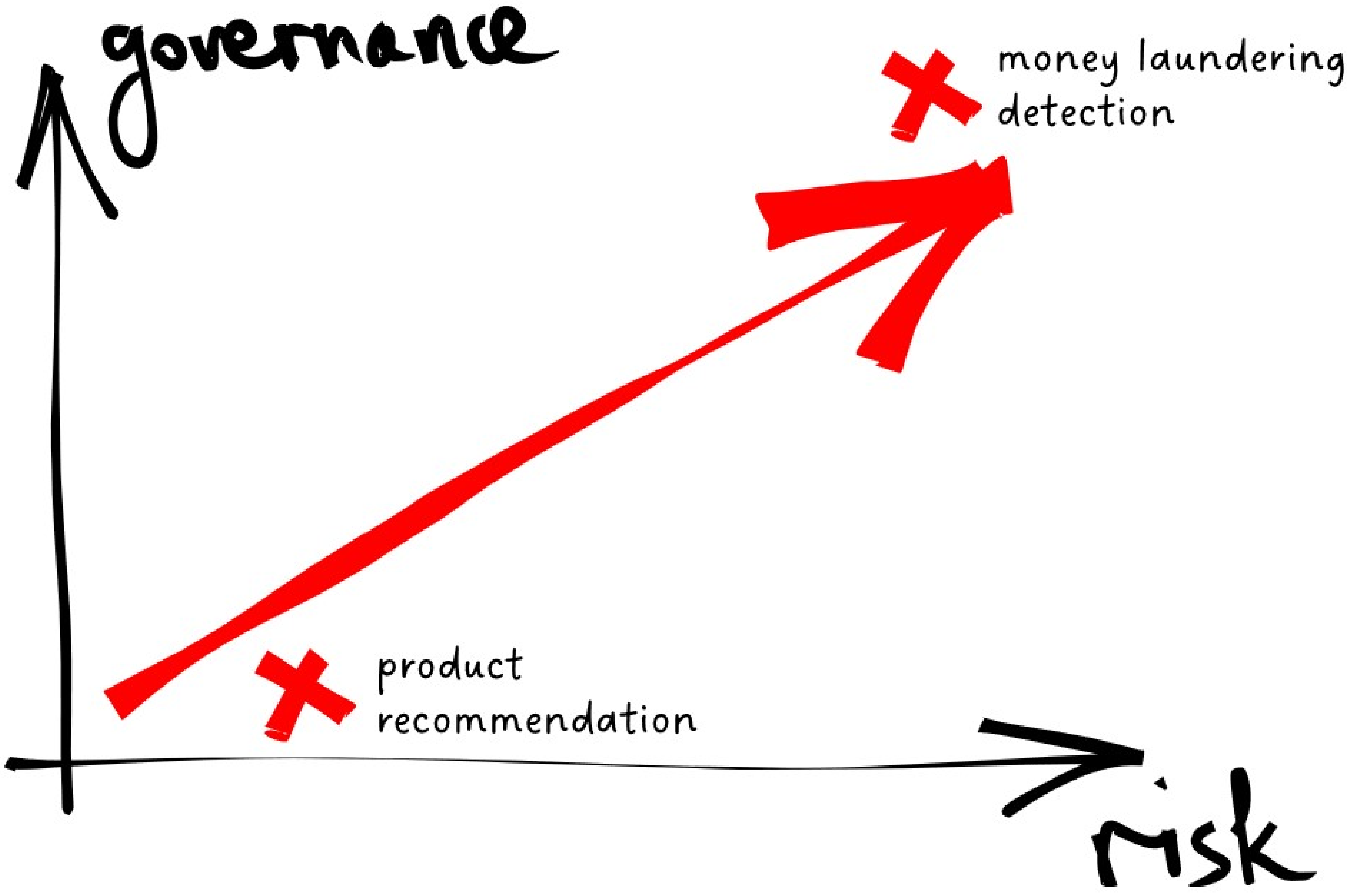


- Is it ethical to use ML?
- Private, sensitive data?
- Unethical model bias?
- ...

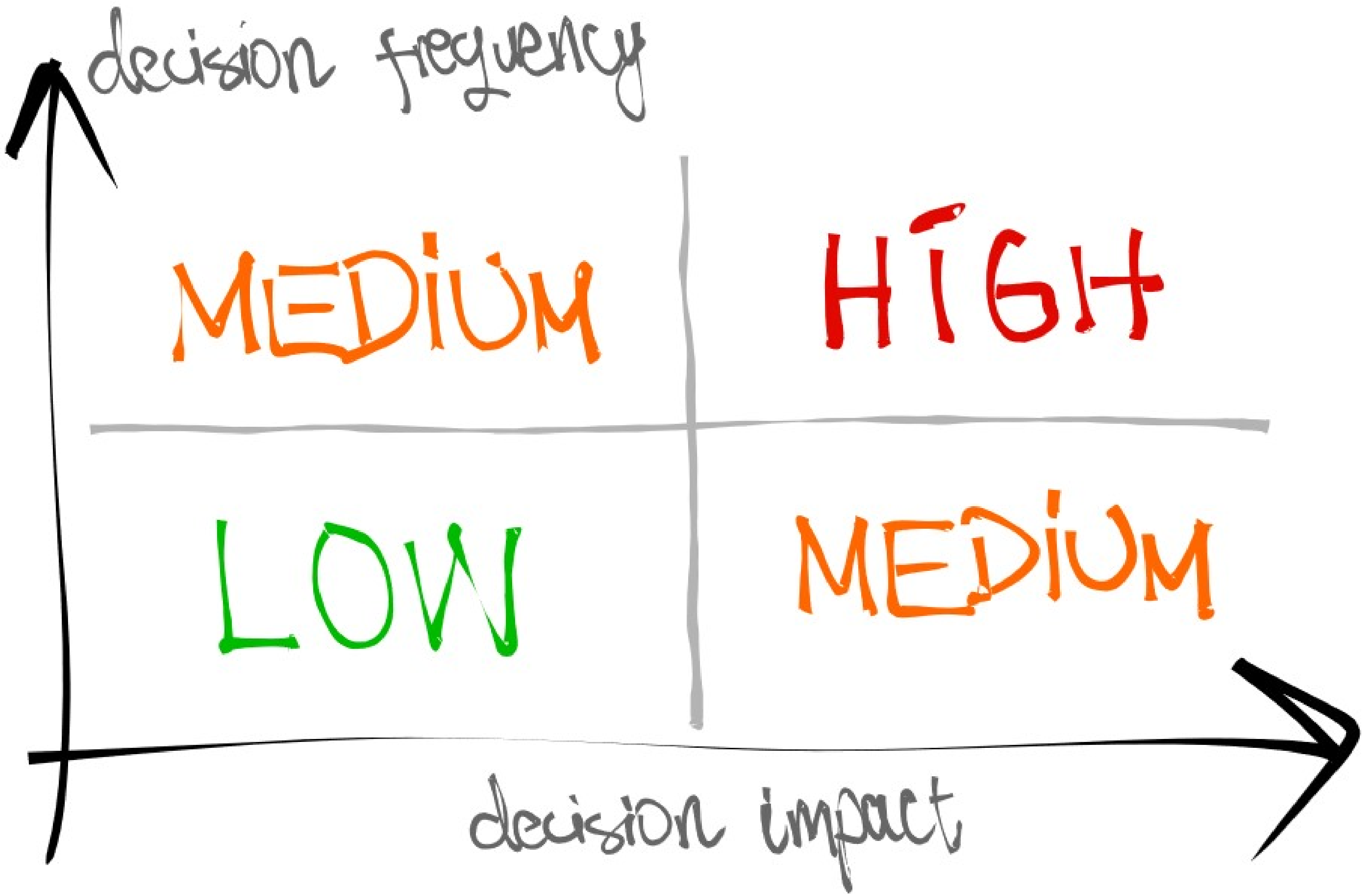
- Model selection docs
- Training data prep
- Quality assurance
- Versioning
- Reproducibility
- ...

- Is API secure?
- Monitoring in place?
- Alerting?
- Failure handling?
- ...


governance
risk


governance
money laundering detection
product recommendation
risk

Risk categories



IMPACT

- Financial
- Reputational
- ...

Summary

- Governance means extra steps, but alternative is anarchy.
- "Launching many models fast" is not our goal, but generating business value.
- Reckless ML --> more damage than benefit.
- More models, more obvious the benefit of governance becomes.

Let's practice!

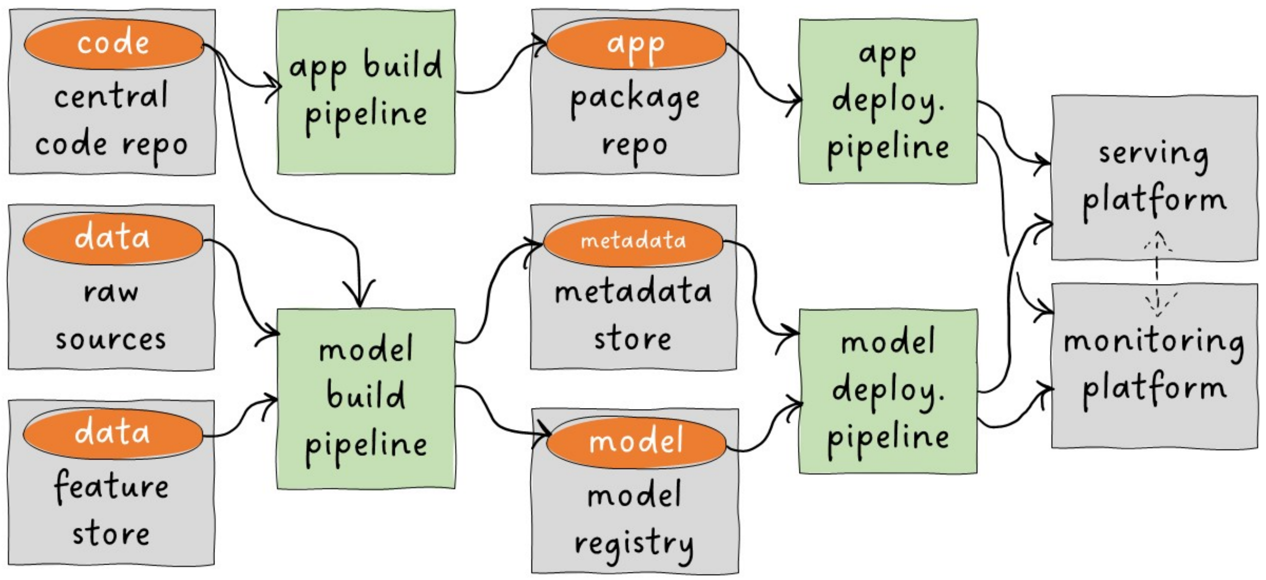
MLOPS DEPLOYMENT AND LIFE CYCLING

Wrap up

MLOPS DEPLOYMENT AND LIFE CYCLING

Nemanja Radojkovic
Senior Machine Learning Engineer

Principles above all



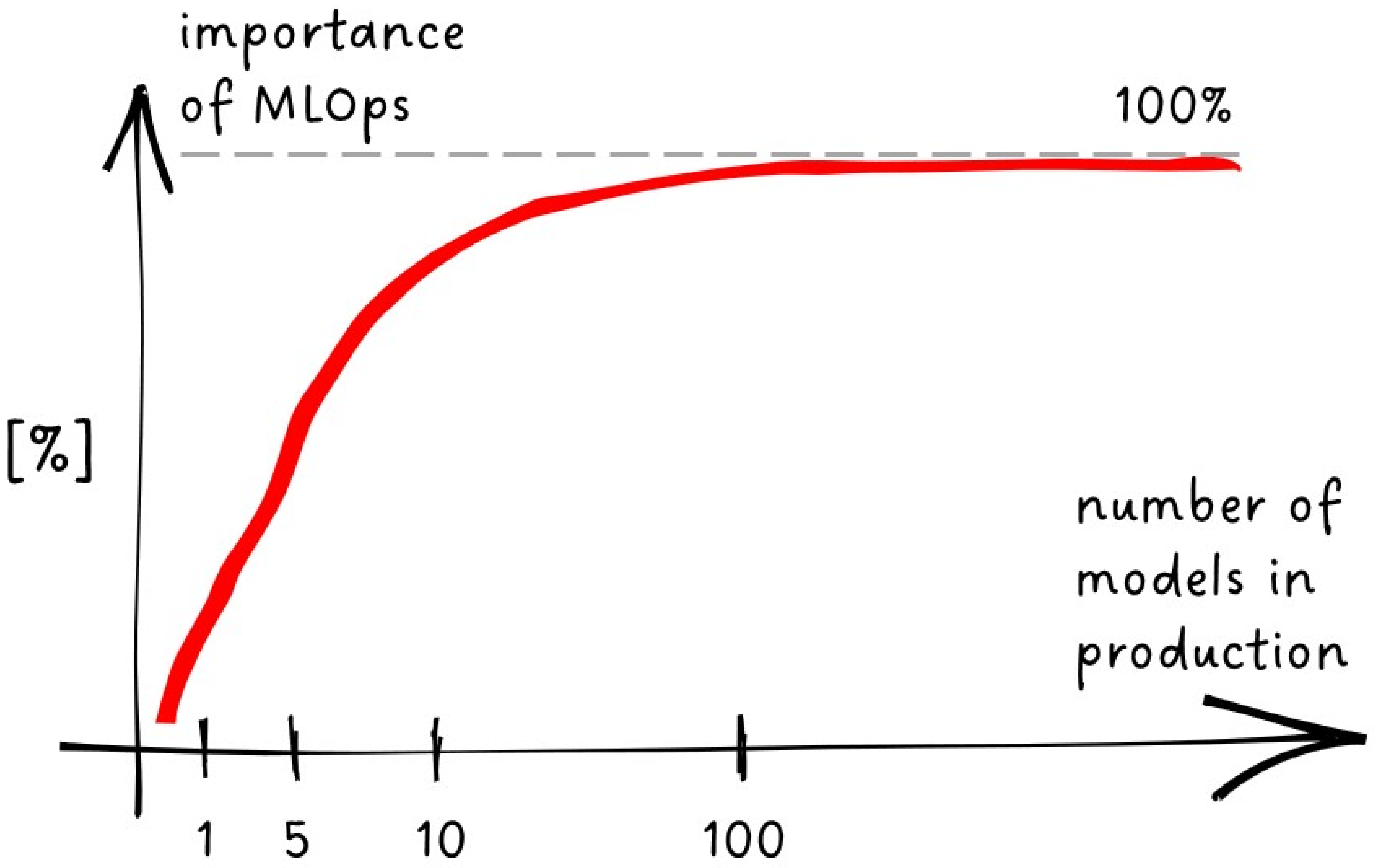
#automation

#collaboration

#efficiency

#transparency

#user-satisfaction


importance of MLOps
100%
[%]
number of models in production
1
5
10
100

Start small and grow

Your MLOps framework today

Your MLOps framework in five years

Thank you!

MLOPS DEPLOYMENT AND LIFE CYCLING