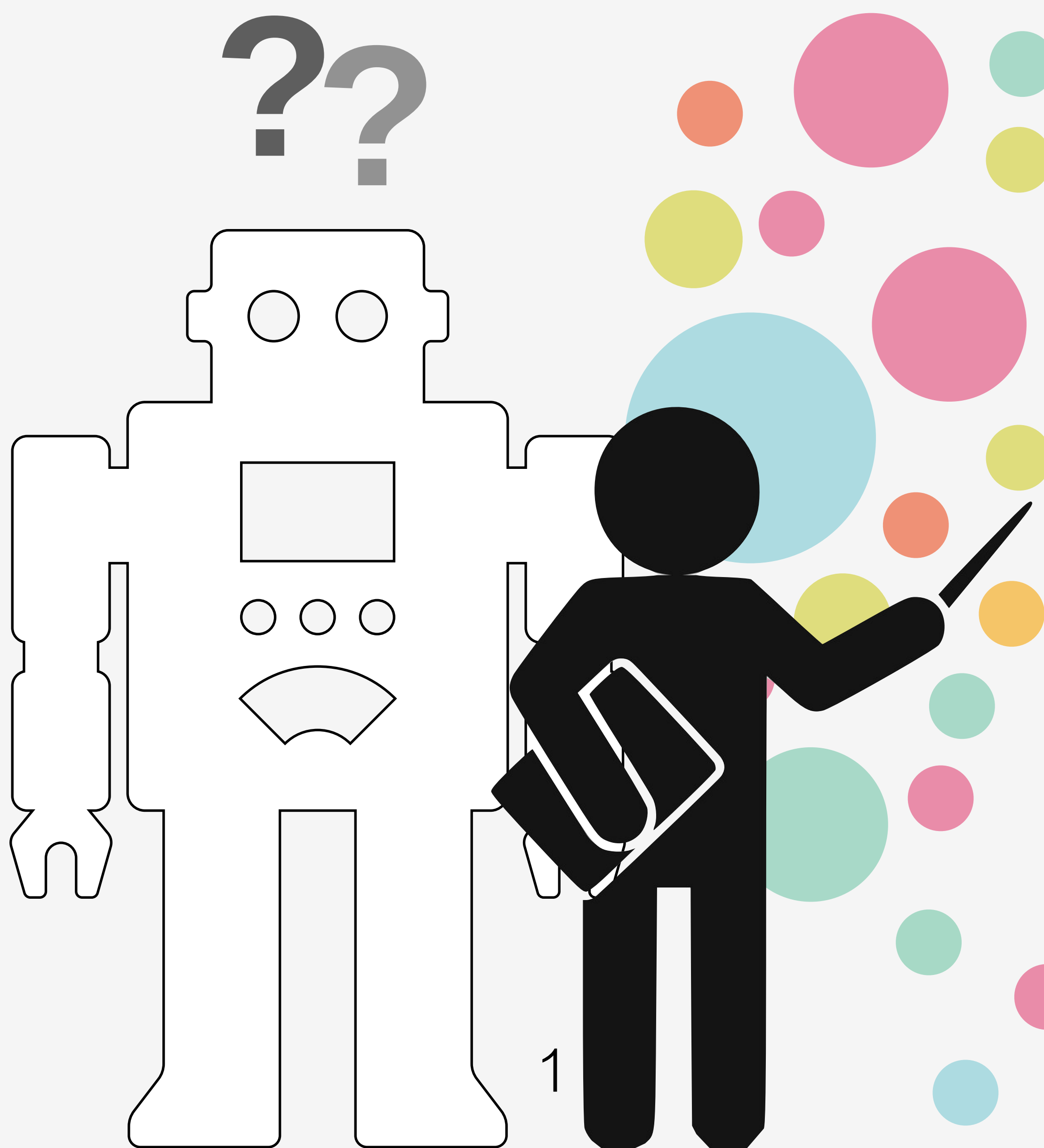
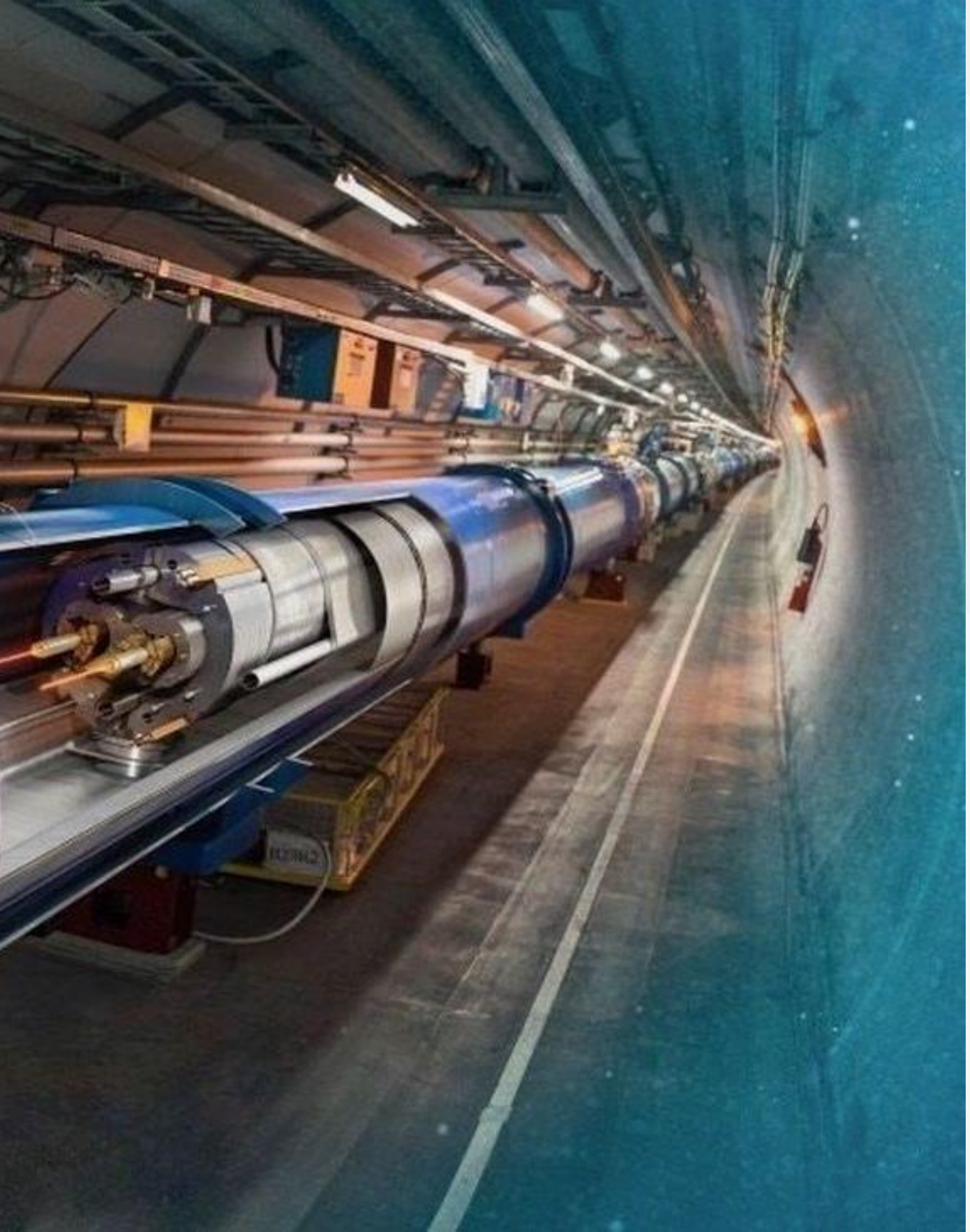
Reinforcement learning for automatic data quality monitoring in HEP experiments

6th Inter-experiment Machine Learning Workshop

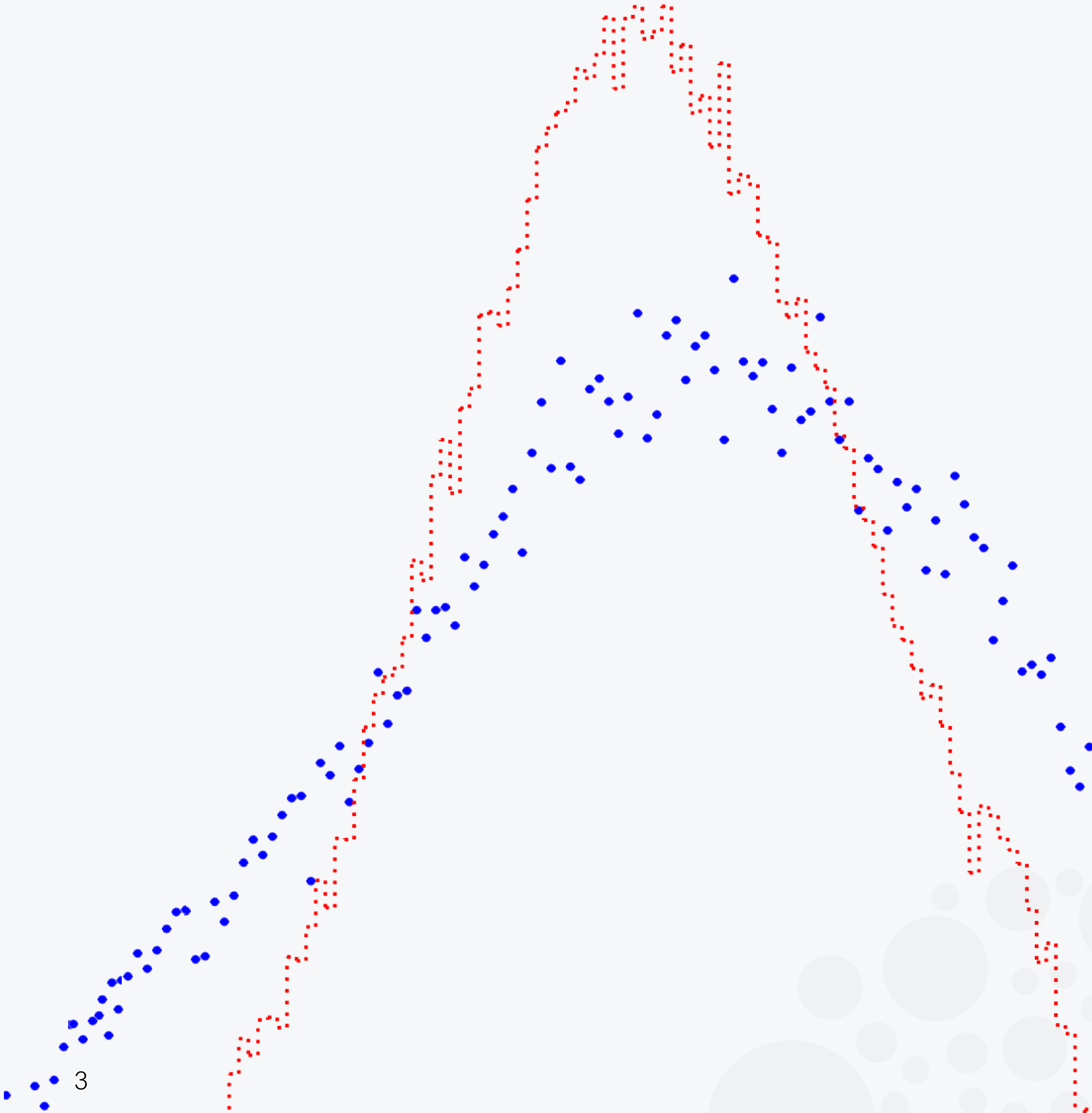


Olivia Jullian Parra (CERN, Geneva)
Lorenzo Del Pianta (CERN, Geneva)
Julián García Pardiñas (CERN, Geneva)
Maximilian Janisch, (University of Zurich, Zurich)
Suzanne Klaver, (Nikhef, Amsterdam)
Thomas Lehéricy,(University of Zurich, Zurich)
Nicola Serra (University of Zurich/CERN, Geneva)



Index

Data quality monitoring

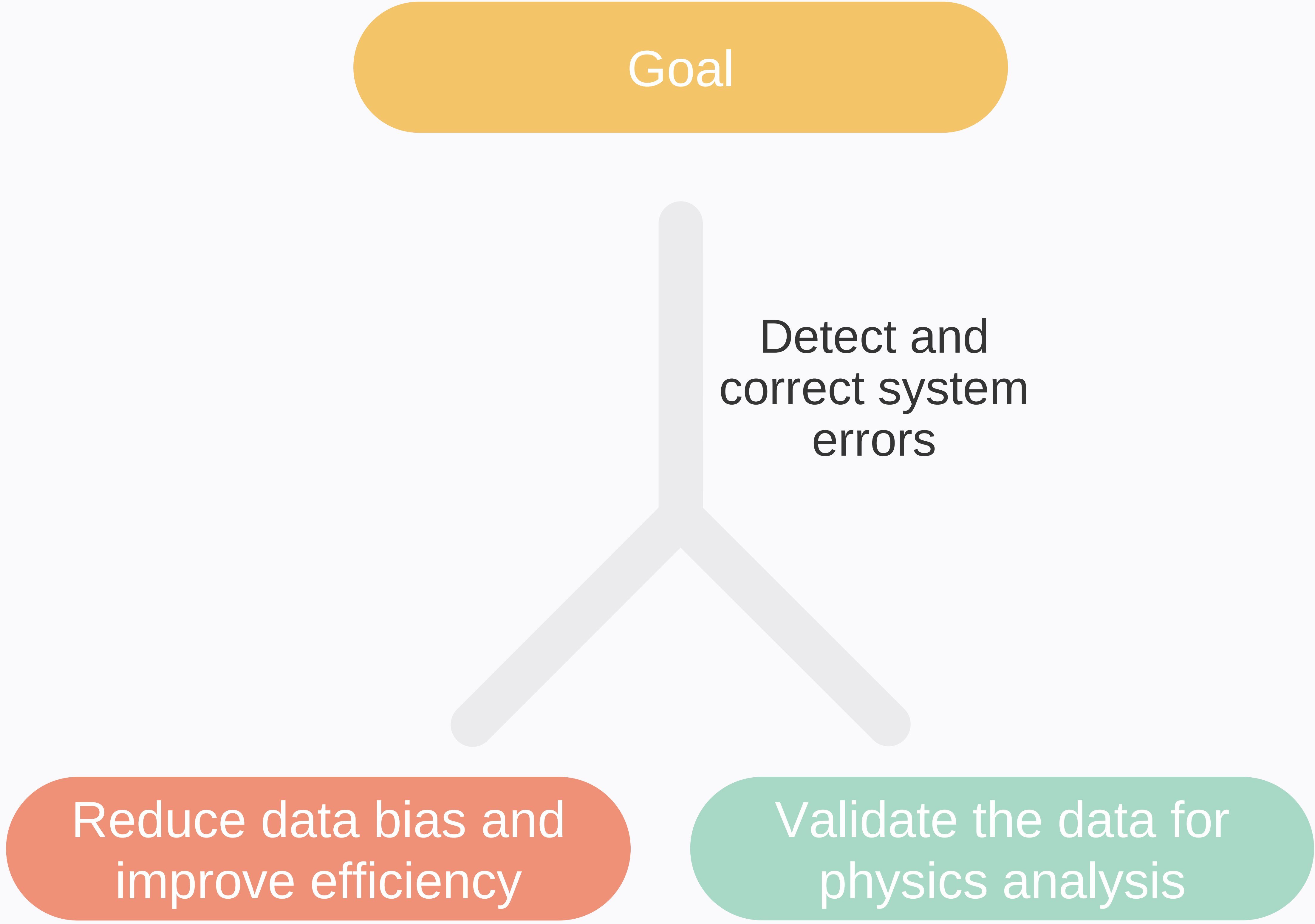


Data quality monitoring

The systems (subdetectors, triggers, etc.) are imperfect and may bias the collected data.

Measurements are biased when datasets are incorrectly classified as good.

Data collection efficiency reduces when datasets are incorrectly classified as bad.



Limitations

+

The work is done by a large pool of non-expert volunteering shifters.

- Need for appropriate training.
- A lot of resources required (online and offline regimes).
- Human errors lead to inaccuracies in the classification.

+

Nominal status changes over time

Even the nominal reference can change. This is the case in a detector’s upgrade

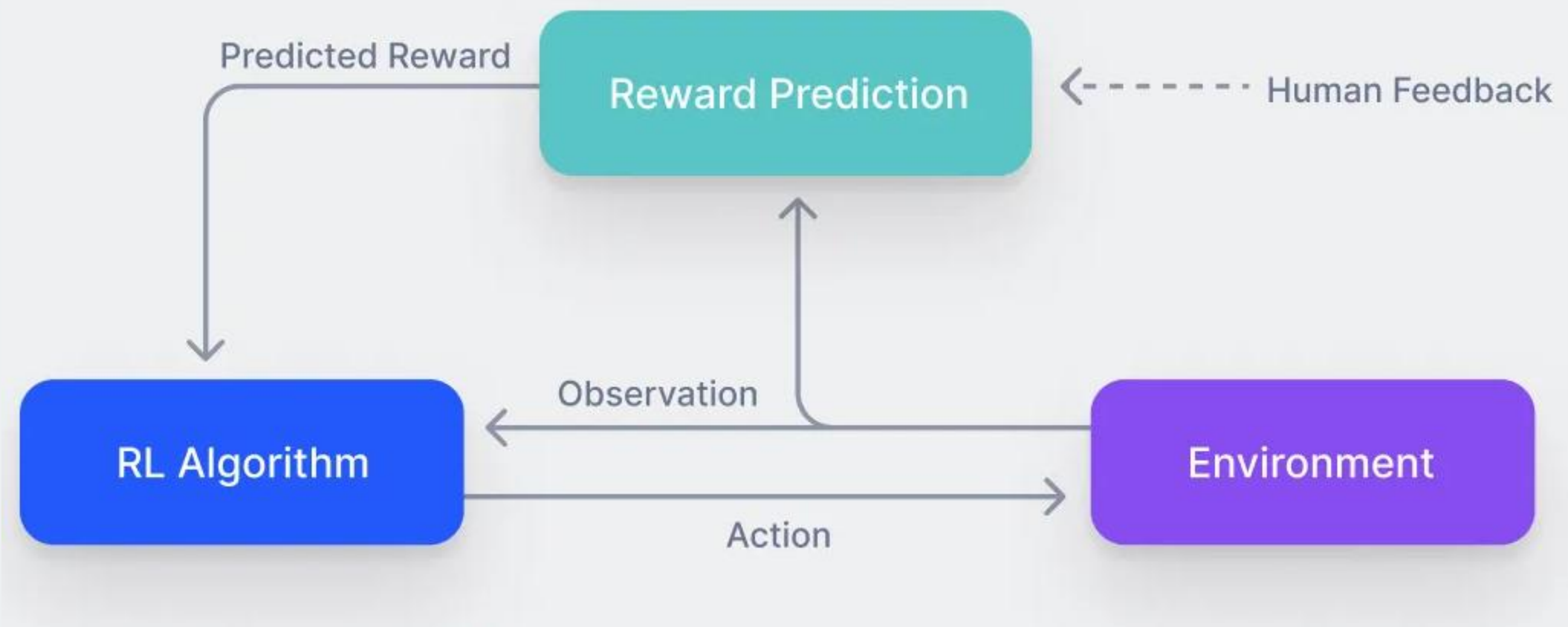
+

New anomalies appear over time

New unseen detector problems could appear in the system over time forcing the monitoring to be able to adapt to new conditions

RLHF for Data quality monitoring

RL with Human Feedback (RLHF)



Reward

- Based on the correctness and confidence level of the agent
- Values set by a scheme reward

Learning agent

- Single interaction with the environment (action).
- Interacts with the human to adjust the given feedback to its policy.
- May have influence on the initial state of the next episode (depending on the regime)

Environment

- Representation of the system's monitoring.
- Each time step conforms an episode.
- States in the episode are histograms collected by the system.

RL Goals

1. **Flexibility:** adapt to changes with the human's guidance.
2. **Improved Accuracy:** complement the current human accuracy.
3. Enhance the **reliability** of the system

Reinforcement learning in the offline regime

Offline regime

The RL goal is the **improvement of the current shifter's accuracy** when detecting system changes. For that, **the agent receives constant feedback.**

The **agent's actions** do **not** have any **influence on the next state of the episode**.

The **action space** of the algorithm is only **based on the definition of the system status**: nominal or anomalous status.

Offline regime: Goals overview

Increase the current shifter's accuracy

Integrate the shifter's feedback to the algorithm's policy to avoid human's mistakes (**superhuman condition**)

Adapt to changing conditions

Adapt to new nominal status **changes** and **detect unseen new problems**

Offline RL algorithm

1 Environment set up

- **Single-step episode**: vector composed by a set of bins (histogram)
- The **initial state is not influenced** by the agent

2 Agent interactions and episode ending

- The **action space** is the classification of the histogram as **anomalous or nominal status**
- The **human** will **always receive feedback** from the **agent**

3 Reward scheme

- The **human** will **always give feedback** to the **agent**
- There is a **reward/penalization** for **correct/incorrect** status **classification**

Reinforcement learning in the online regime

Online regime

The RL goal is to **maximize the current shifter's accuracy** while **reducing its interactions** with the agent as least as possible. **The agent must know when to get human feedback.**

The **agent's actions influence the next state**, making the misclassified system status persistent until its correct detection.

The **action space** of the algorithm **declares** not only **the status of the system** but **also defines the human feedback necessity.**

Online regime: Goals overview

Reduce the shifter's interventions

Achieve **superhuman condition** while knowing **when to ask for human feedback**

Adapt to changing conditions

Adapt to new nominal status **changes** and **detect unseen new problems**

Online RL algorithm

1 Environment setup

- **Infinite horizon episode**: vectors composed by a set of bins (histogram)
- The **next states are influenced** by the agent (the state persists if misclassified)

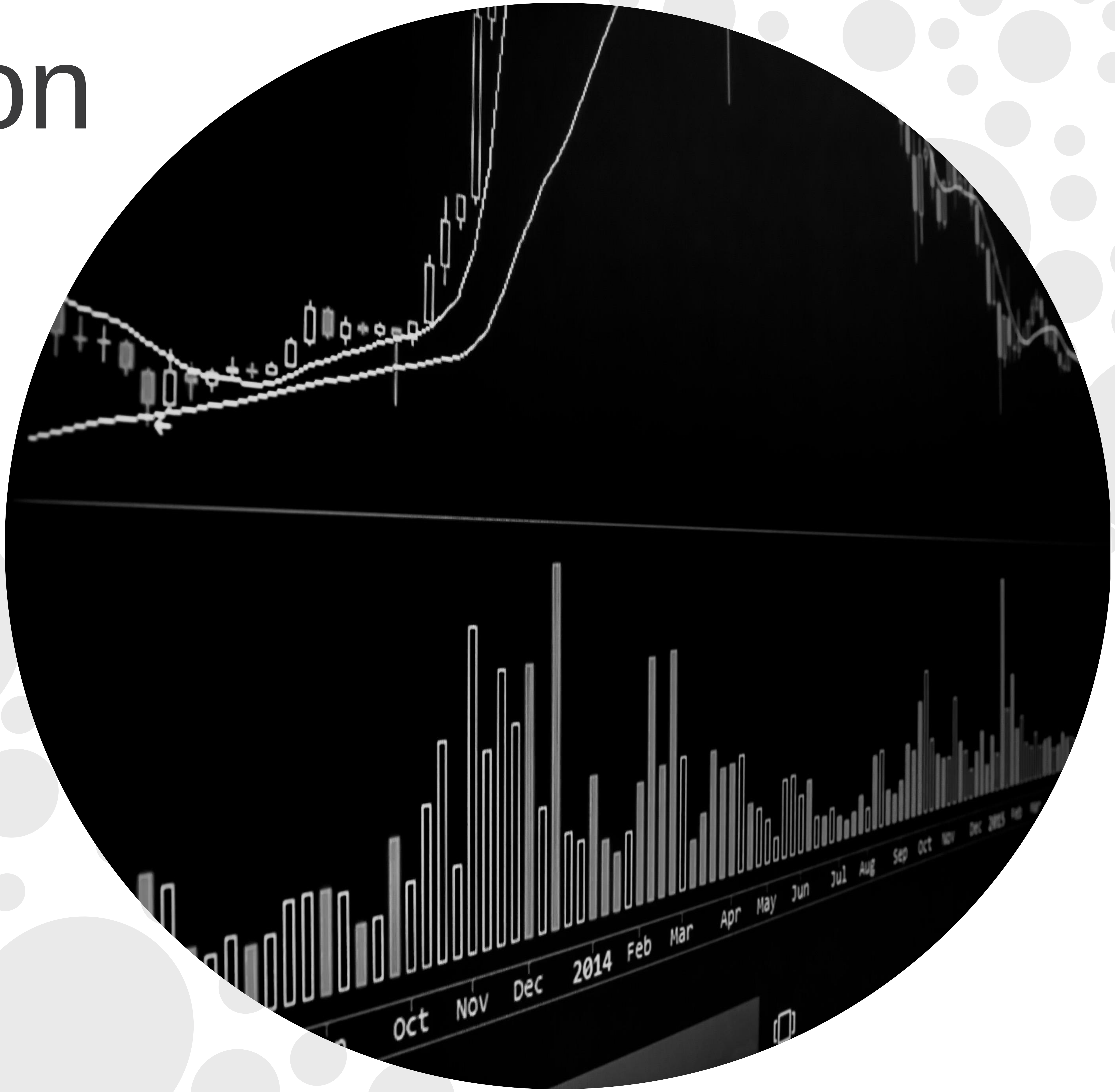
2 Agent interactions and episode ending

- The **action space** is the **classification of the histogram** and the decision on **asking or not human feedback**
- The **human** will **always receive feedback** from the **agent**

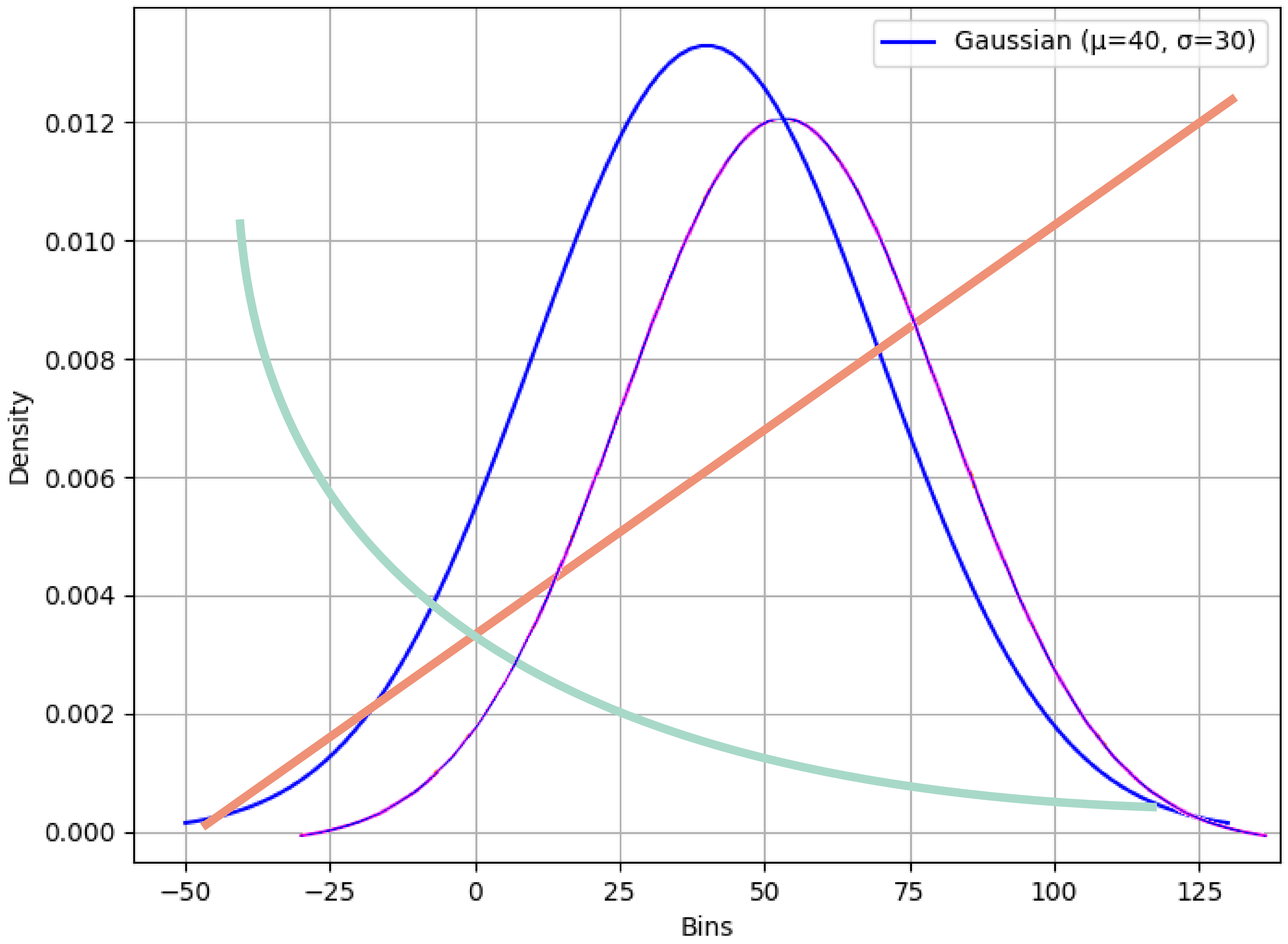
3 Reward scheme

- The **human** will **only give feedback** to the **agent** when being called
- There is a **reward/penalization** for **correct/incorrect** status **classification dependent on time**
- There is a **penalization** for asking **human feedback unnecessarily**

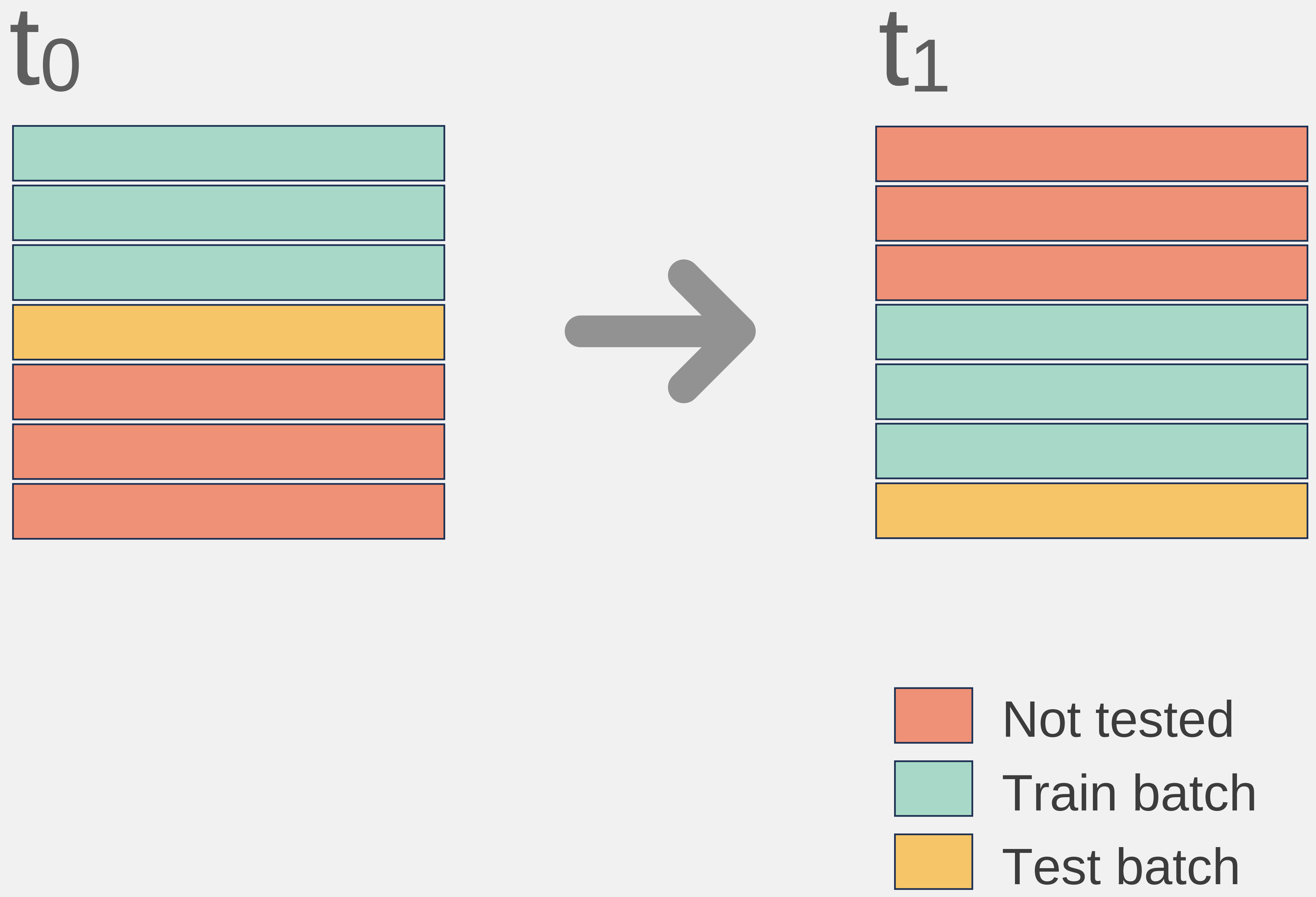
First simulations on toy dataset



Design setup

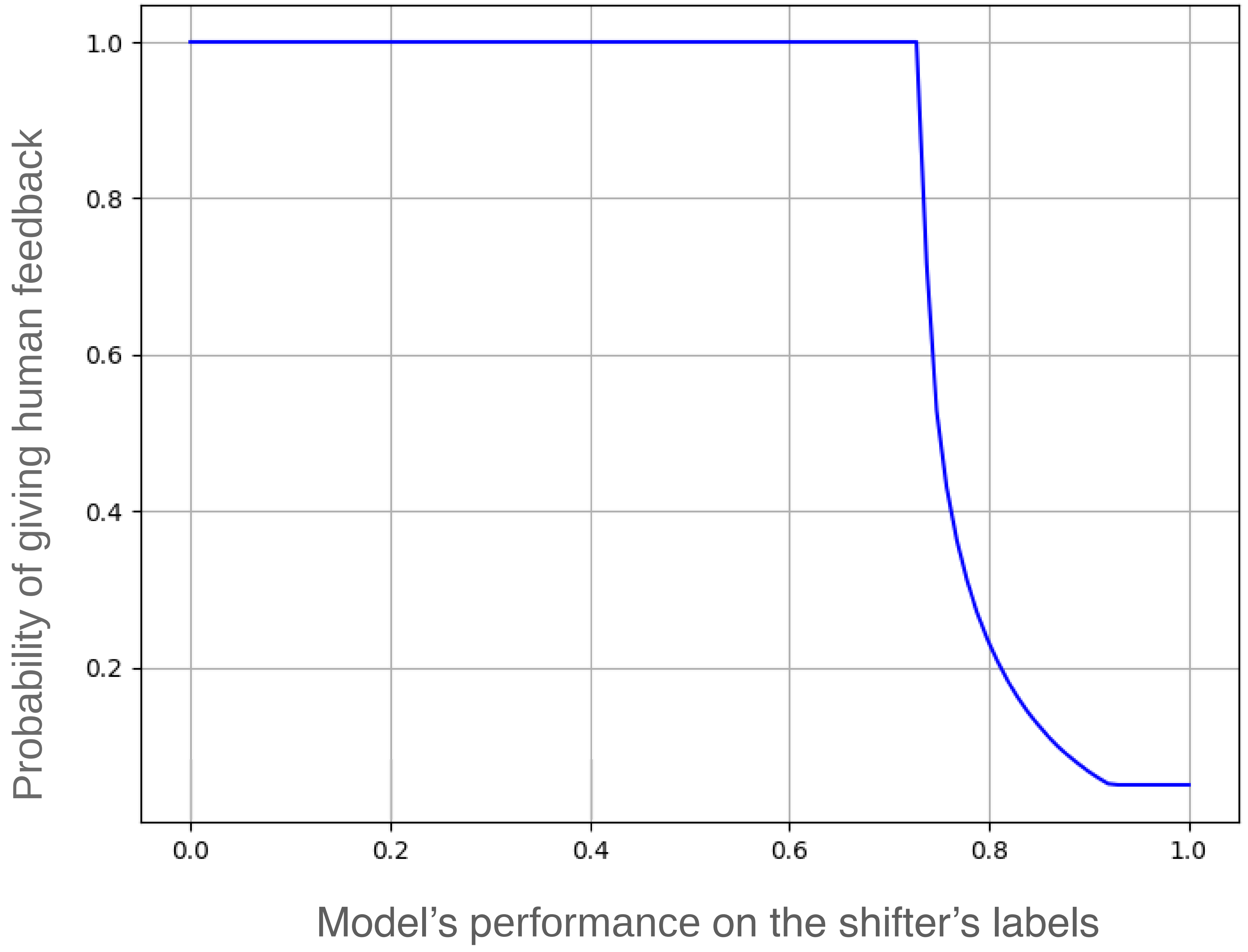


Histogram distributions



Sequential training

Superhuman condition

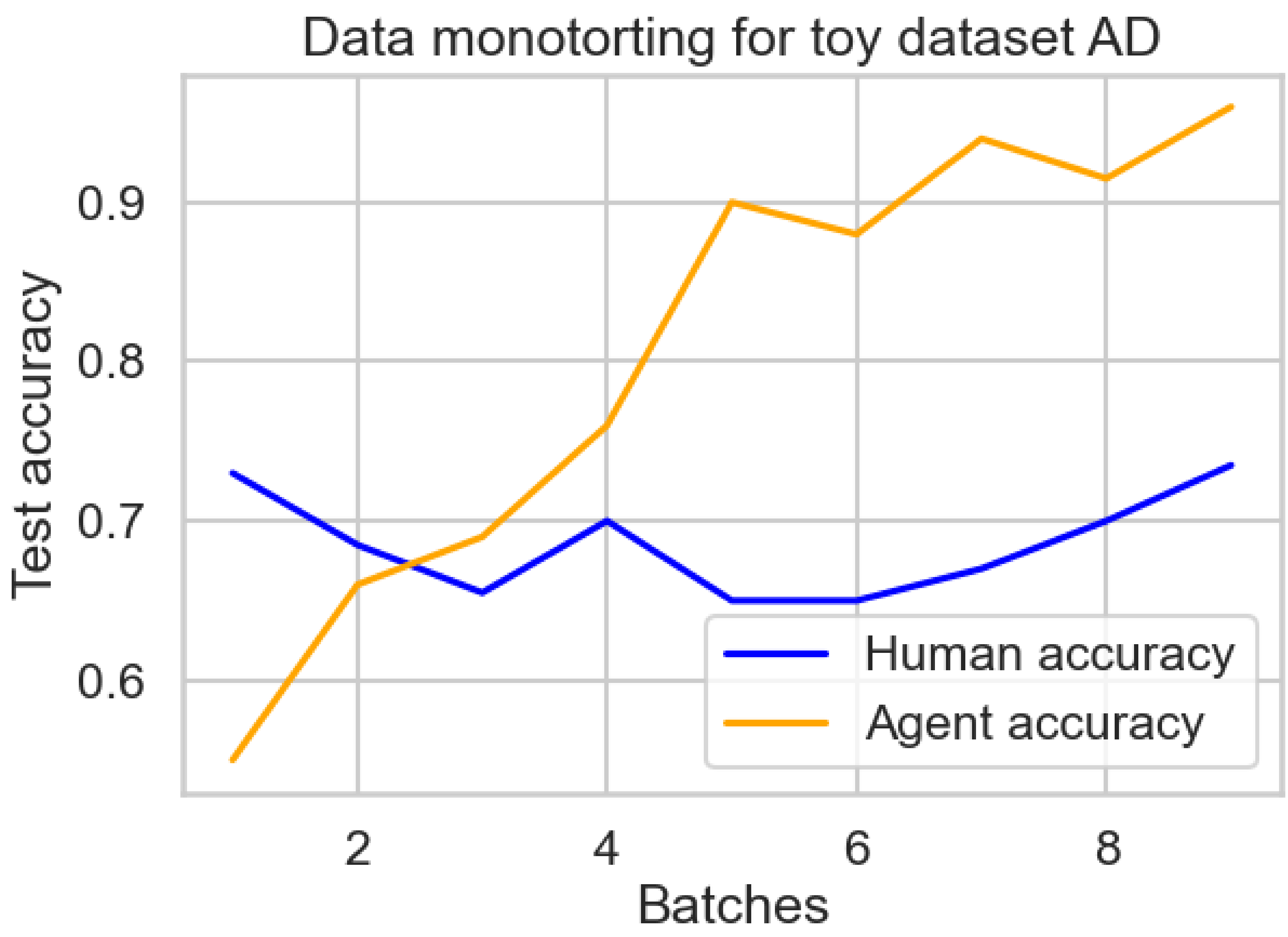


Shifter’s “mistrust” heuristic function

Proof of concept: What if we have access to the ground truth?

- We create the toy dataset and we **we falsify the labels** with probability p (human error)
- **These labels** are the **feedback** given to the agent
- We **stop** giving **feedback** to the agent progressively **based on its performance during training** (shifter’s mistrust heuristic function)

Results

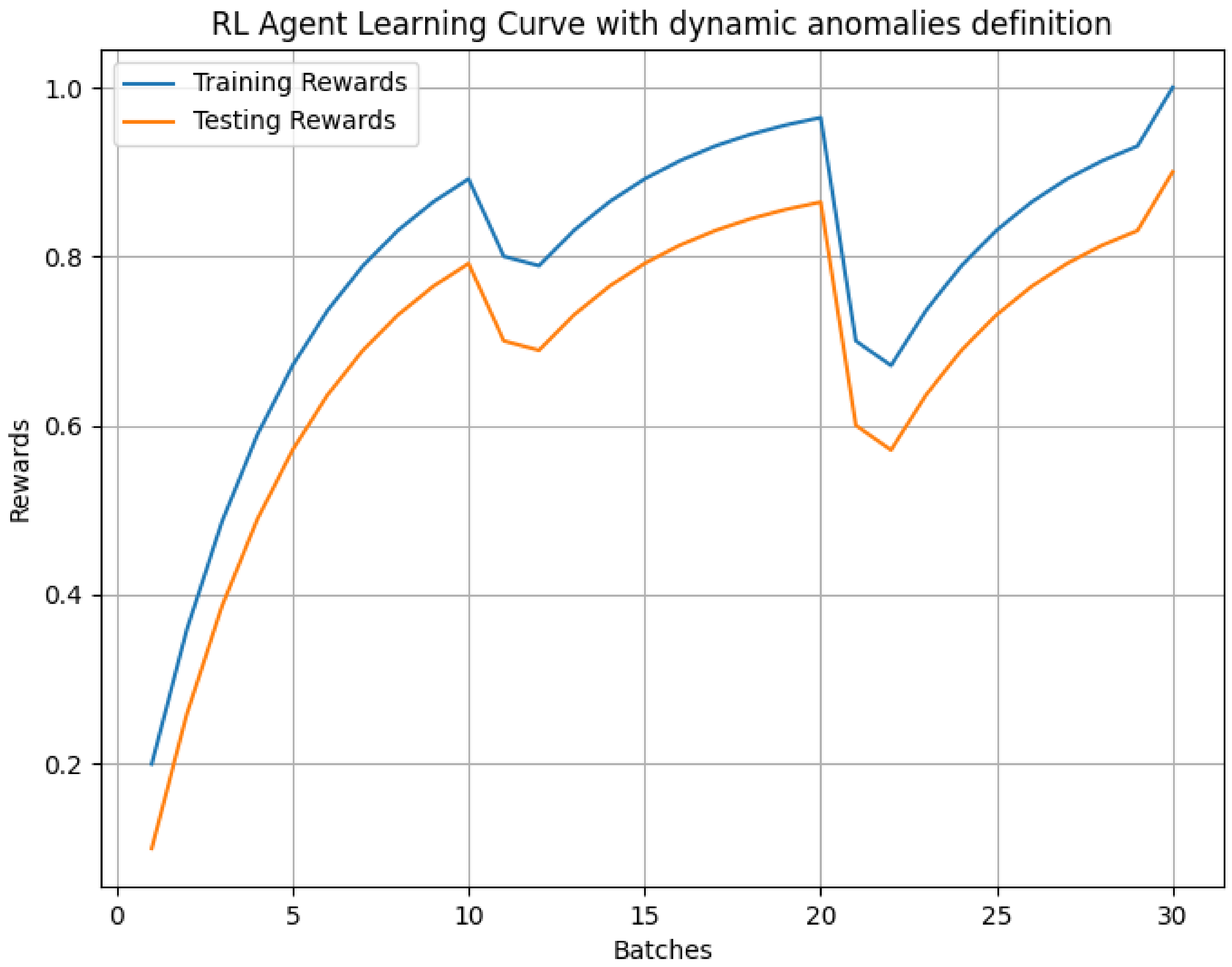


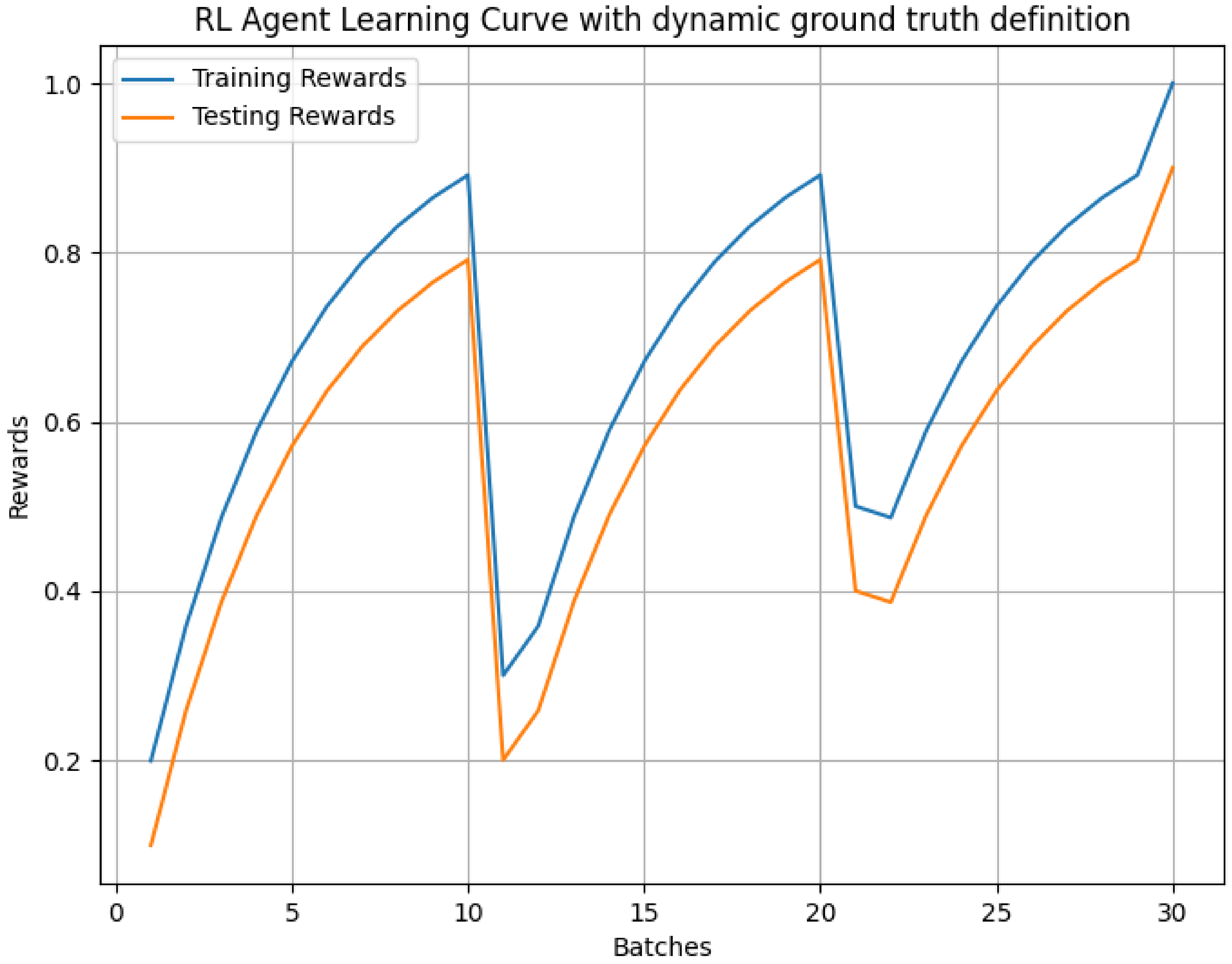
Probability of human failure: 0.3

- We can **achieve superhuman condition** for the offline regime
- **How do we know when to rely on the machine?** We are currently performing studies on the detection of superhuman condition during training, without having access to the ground truth[1]

[1] Bar, O., Drory, A., & Giryes, R. (2022, May). A spectral perspective of DNN robustness to label noise. In International Conference on Artificial Intelligence and Statistics (pp. 3732-3752). PMLR.

Adaptation to changing conditions



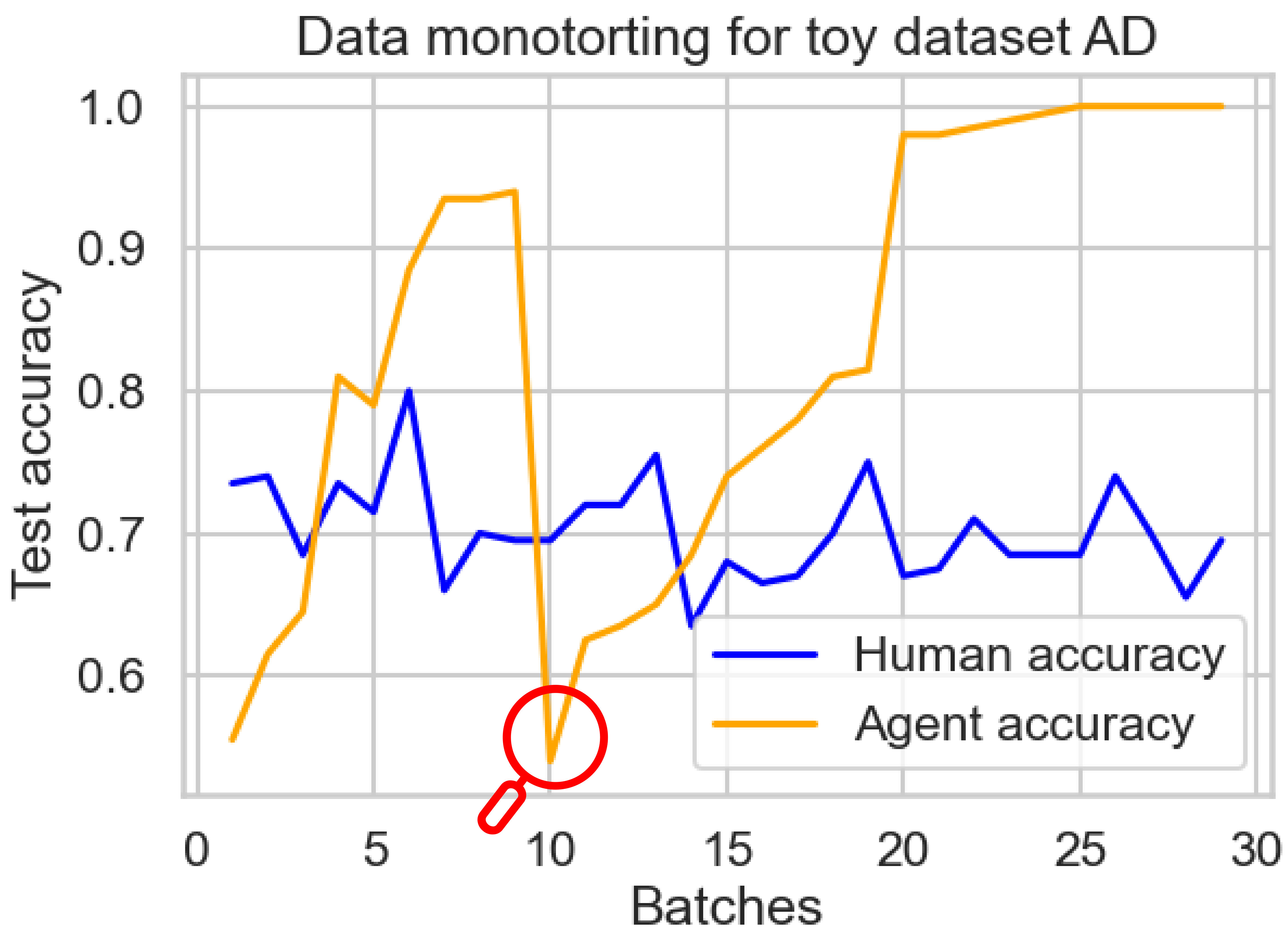


Anomalies changing over time

Nominals changing over time

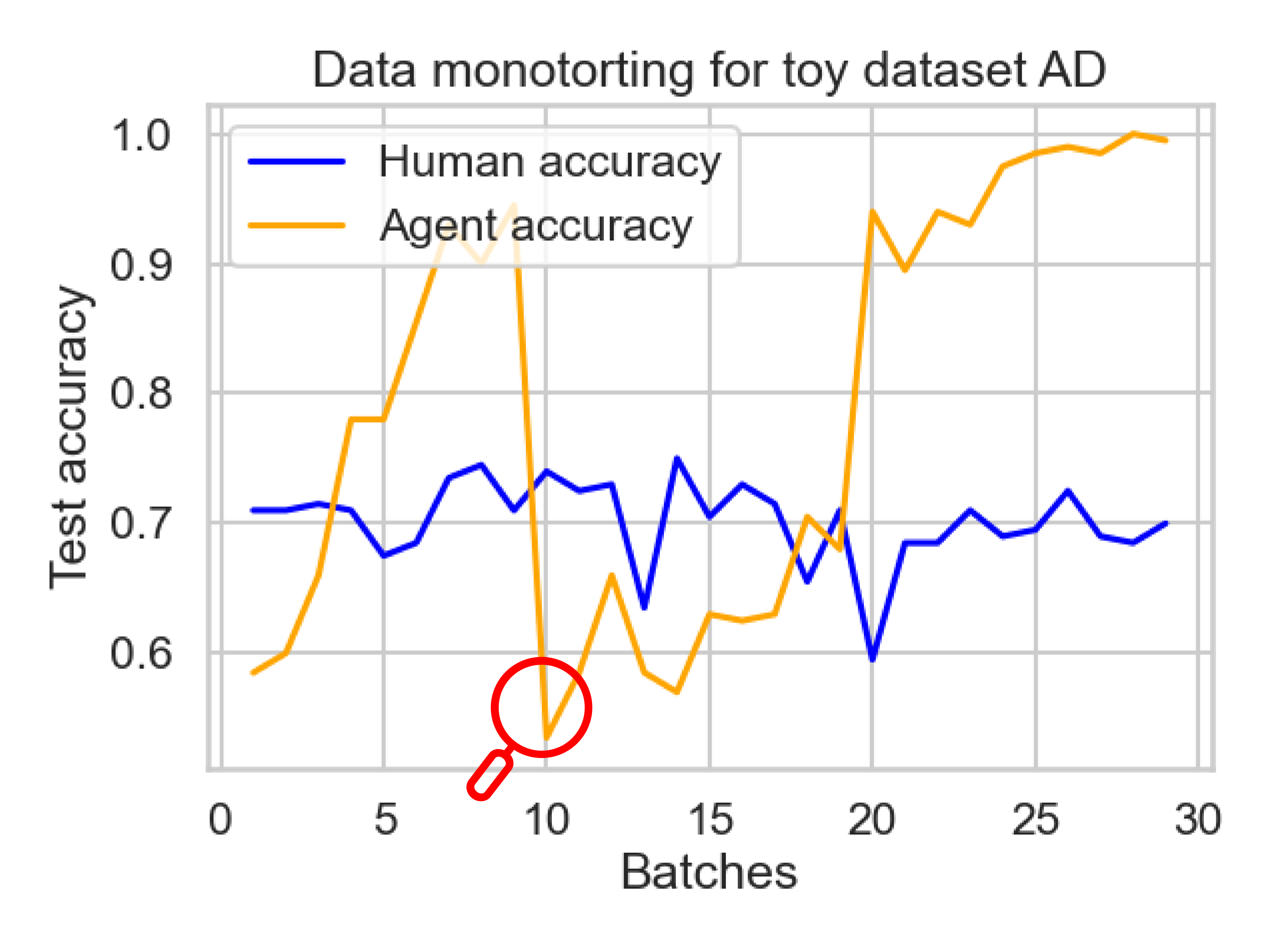
Proof of concept: What happens when the anomaly changes over time?

Results



Anomalies changing over time

Probability of human failure: 0.3



Nominals changing over time

Probability of human failure: 0.3

We **adapt to changing conditions** for the offline regime

Future steps

Achieve superhuman condition and adapt to changing conditions also in real time (online regime)

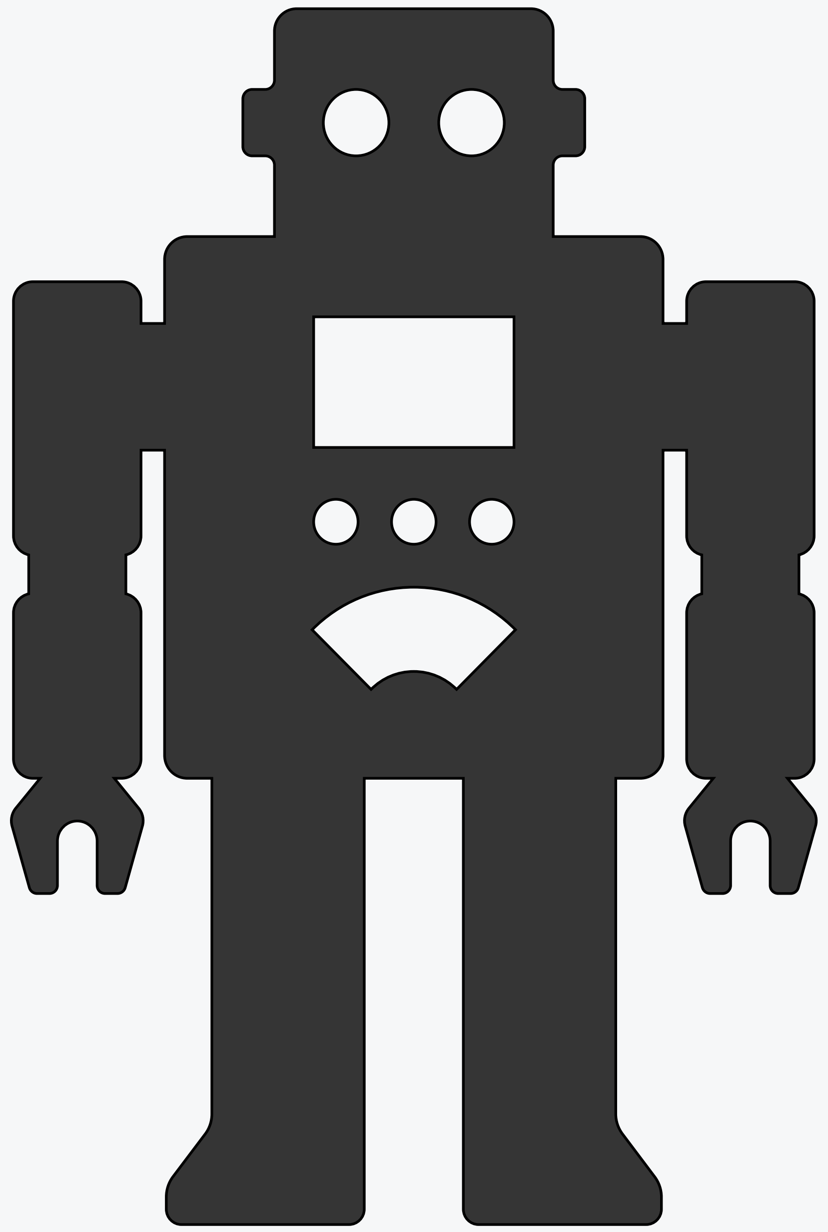
Currently

Assist the shifter decisions without necessity of their constant feedback and automatize some histograms checks

Next

Implement data augmentation techniques to be able to use the algorithm with low statistics data (real case scenario)

Next



Thanks
for your attention

Q&A

olivia.jullian.parra@cern.ch