Reinforcement learning for automatic data quality monitoring in HEP experiments

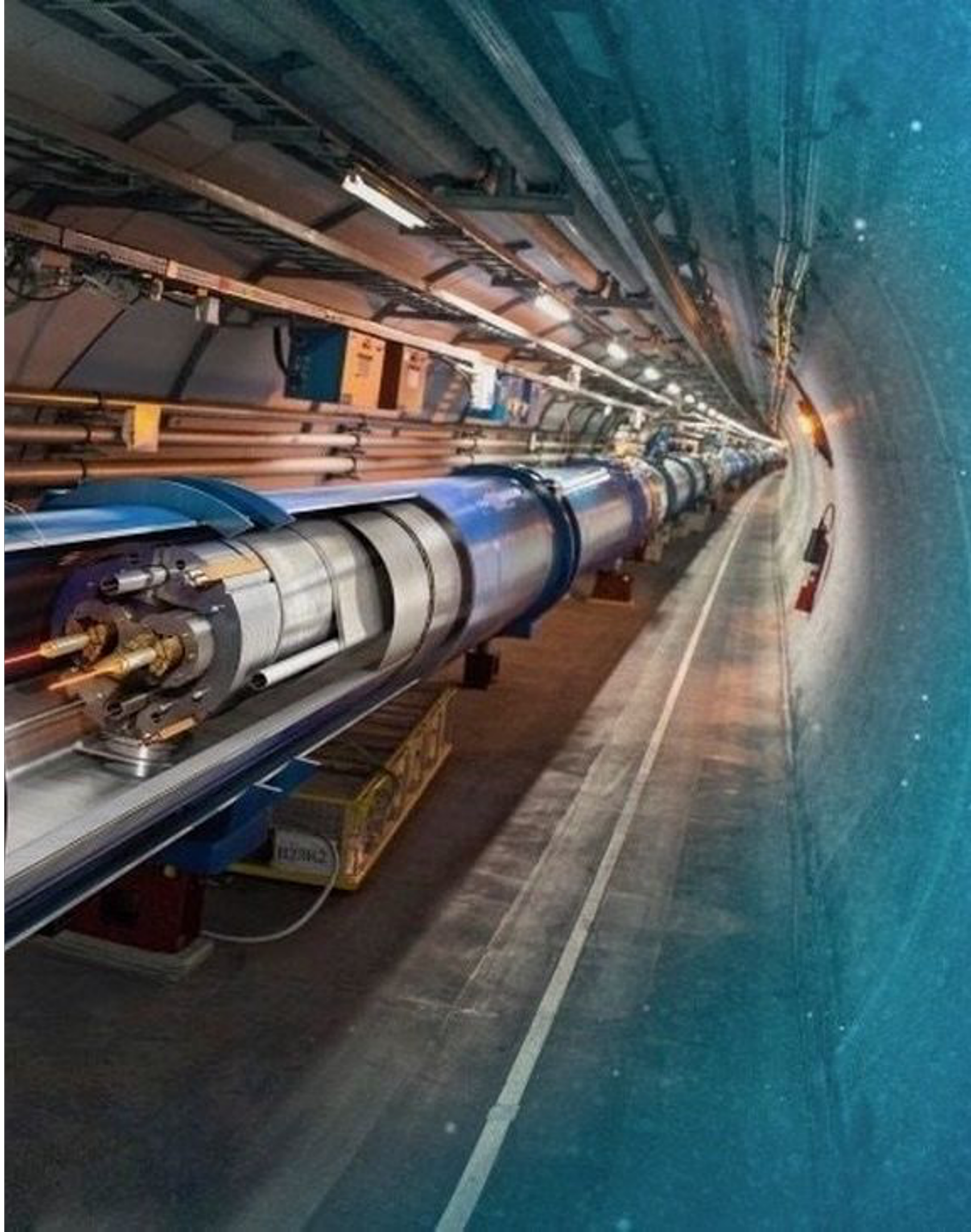
European AI for Fundamental Physics Conference (EuCAIFCon)

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

1st of May 2024

Outline

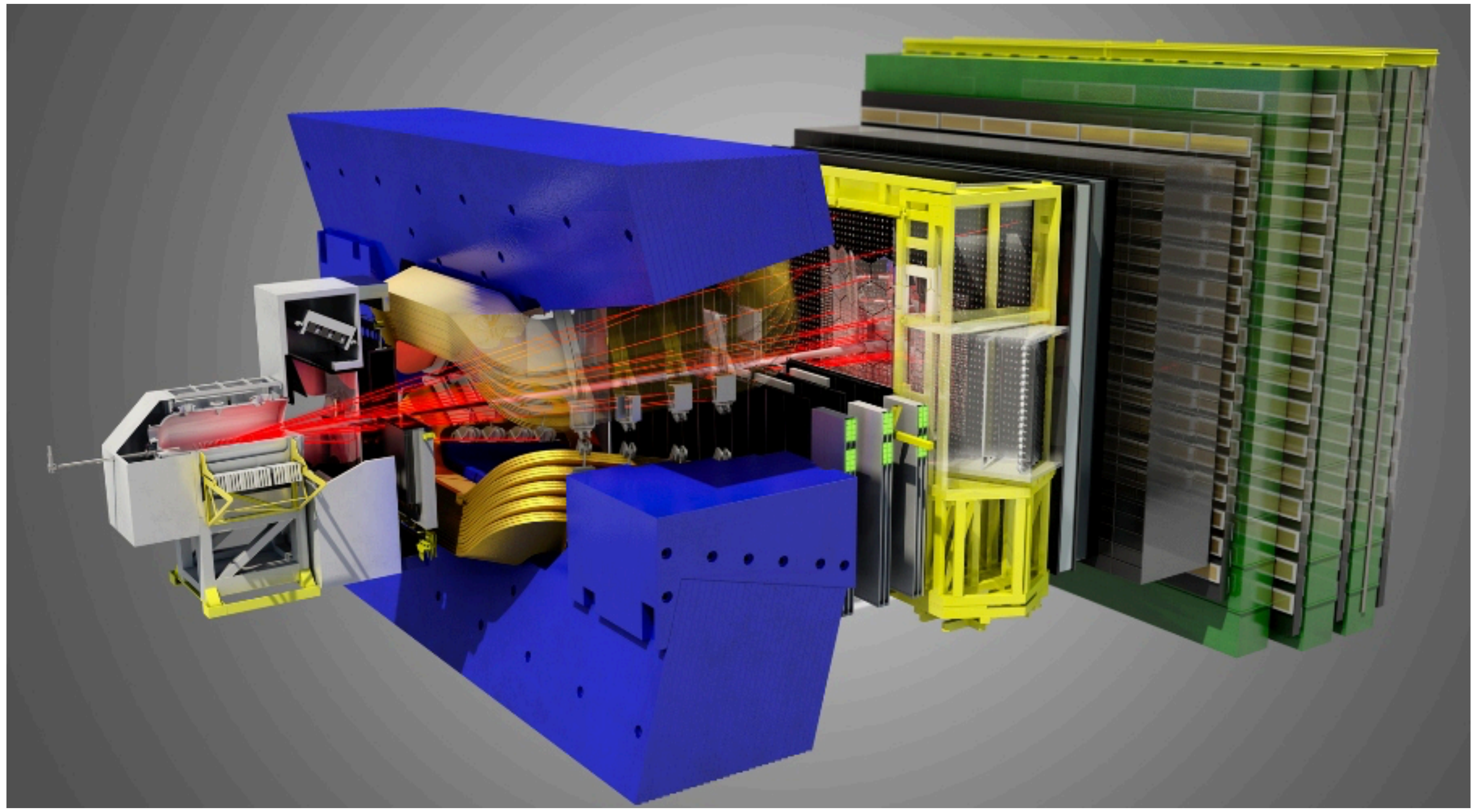
- Data Quality Monitoring (DQM)
- CERN's Data Quality Monitoring
- Reinforcement learning with human feedback for DQM
- Prototype and POC studies
- Conclusions and outlook



Data Quality Monitoring (DQM) at large HEP experiments

- Detectors are complex systems with a **huge number of different components**
- Those **components are prompt to unpredictable errors** (e.g. something can break)
- Those errors may render the data unusable

We need to carefully monitor the status of the systems and the collected data



LHCb experiment at CERN

Data Quality Monitoring at large HEP experiments

- DQM done by trained non-experts: **Shifters**
- Shifters monitor the system in **two stages:**

Online regime

- **Real-time** monitoring (focused on fast decisions)
- Goal: **finding quickly** the system **problems** and solving them

Offline regime

- Monitoring **after the data has been collected** (focused on **high accuracy**)
- Goal: **determining the quality of the data** for posterior physics analysis

Current limitations

Noisy labels

- Different level of shifter's training/experience
- Different judgement across shifters
- Local attention (inability to look at all the histograms all the time)

High person power demand

- Hundreds of shifters per year

Goal → **Improve data collection efficiency and automation**

Challenges for automating the process

- **Fast adaptation to changing operational conditions**
- **Optimising human-machine interactions scheme**
    - ✓ Balance between automatic checks and shifter's decisions during online regime
    - ✓ Assist the shifters to improve accuracy during offline regime

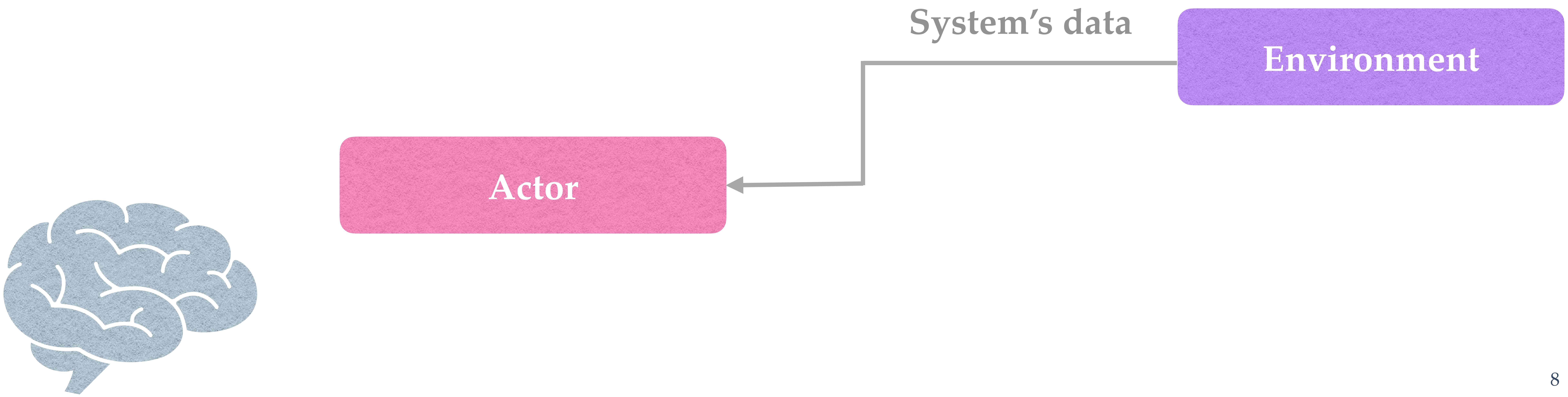
Solution → **Reinforcement learning with human feedback**

Challenges for automating the process

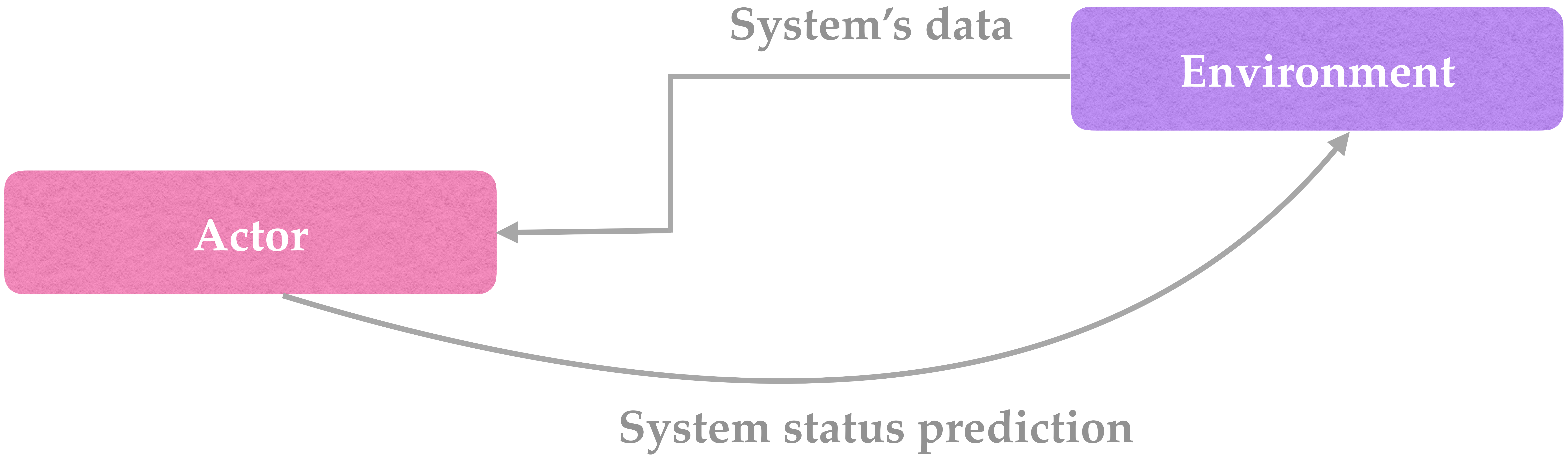
- **Fast adaptation to changing operational conditions (Continuously trained during data collection)**
- **Optimising human-machine interactions scheme (Possibility to design complex interactions with the shifter)**
  - ✓ Balance between automatic checks and shifter's decisions during online regime
  - ✓ Assist the shifters to improve accuracy during offline regime

Solution → **Reinforcement learning with human feedback**

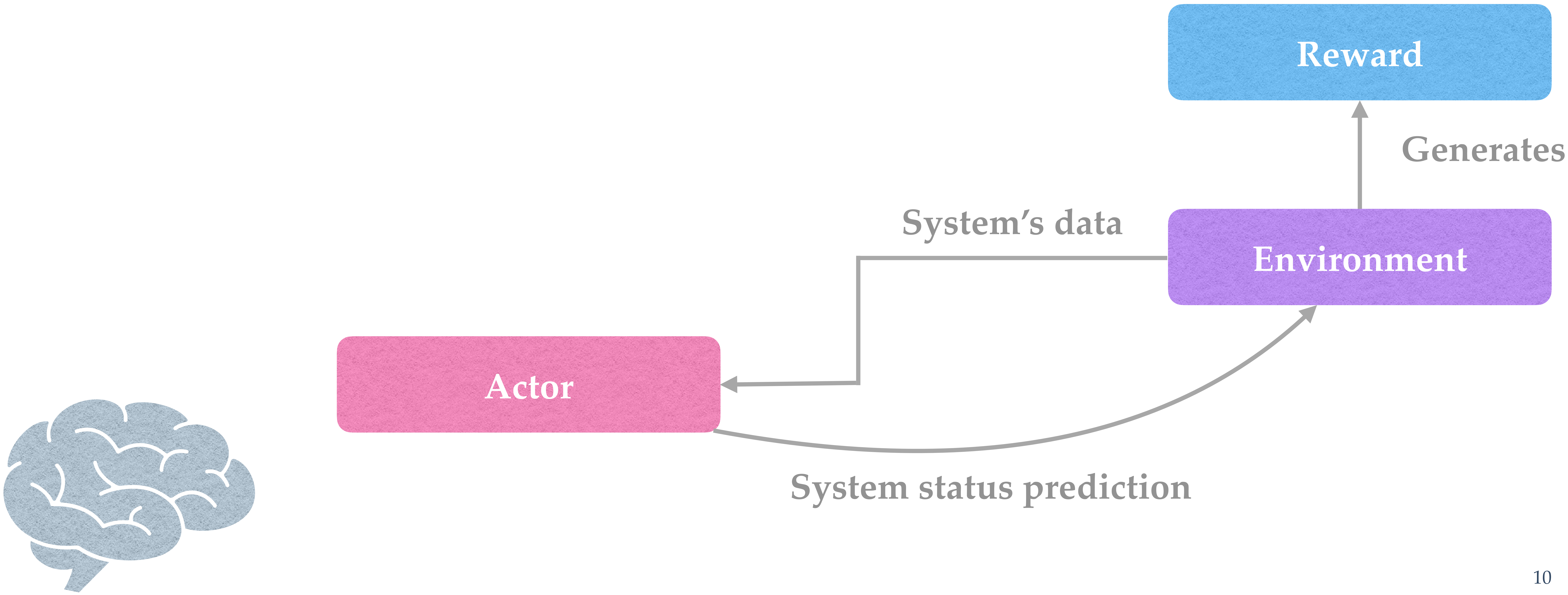
Reinforcement Learning (RL) with Human Feedback



Reinforcement Learning (RL)
with
Human Feedback

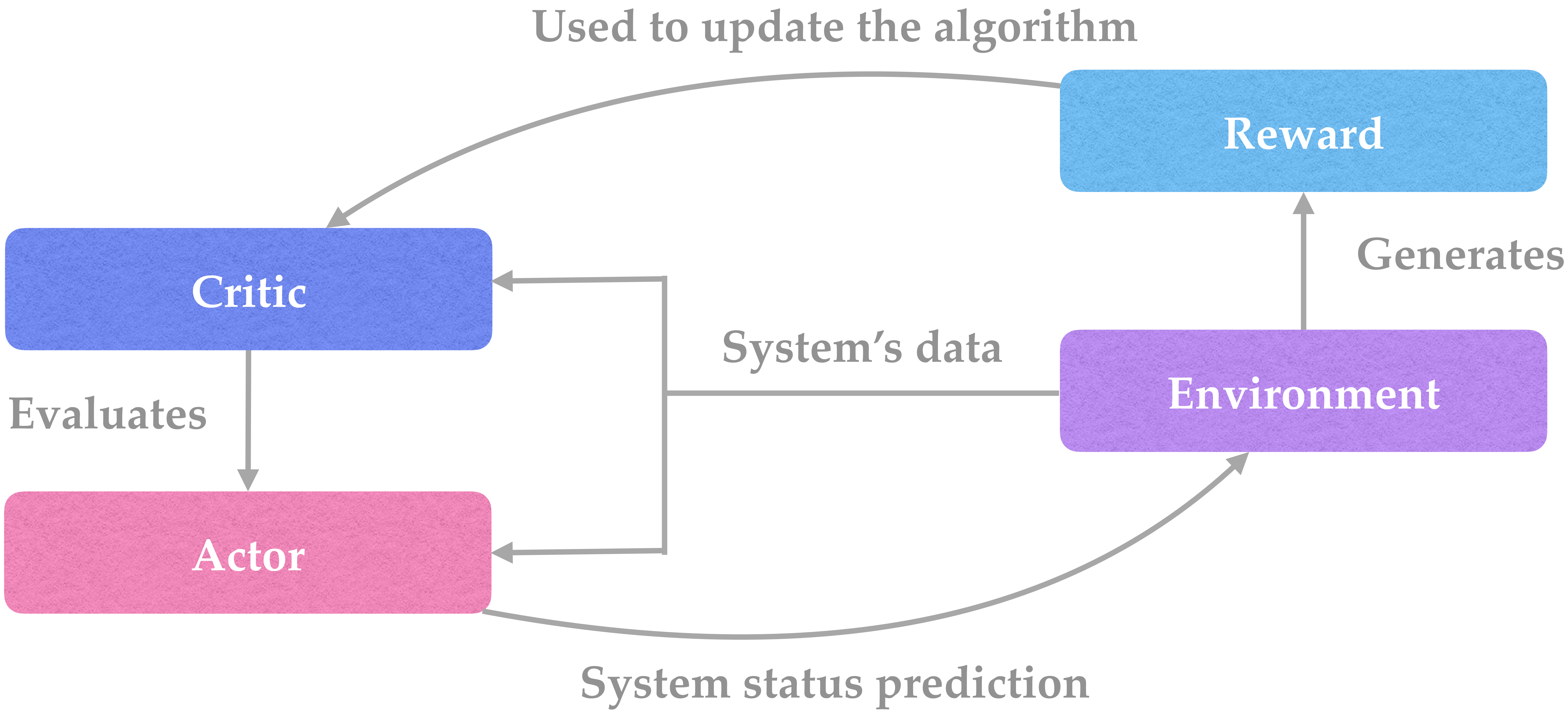


Reinforcement Learning (RL) with Human Feedback

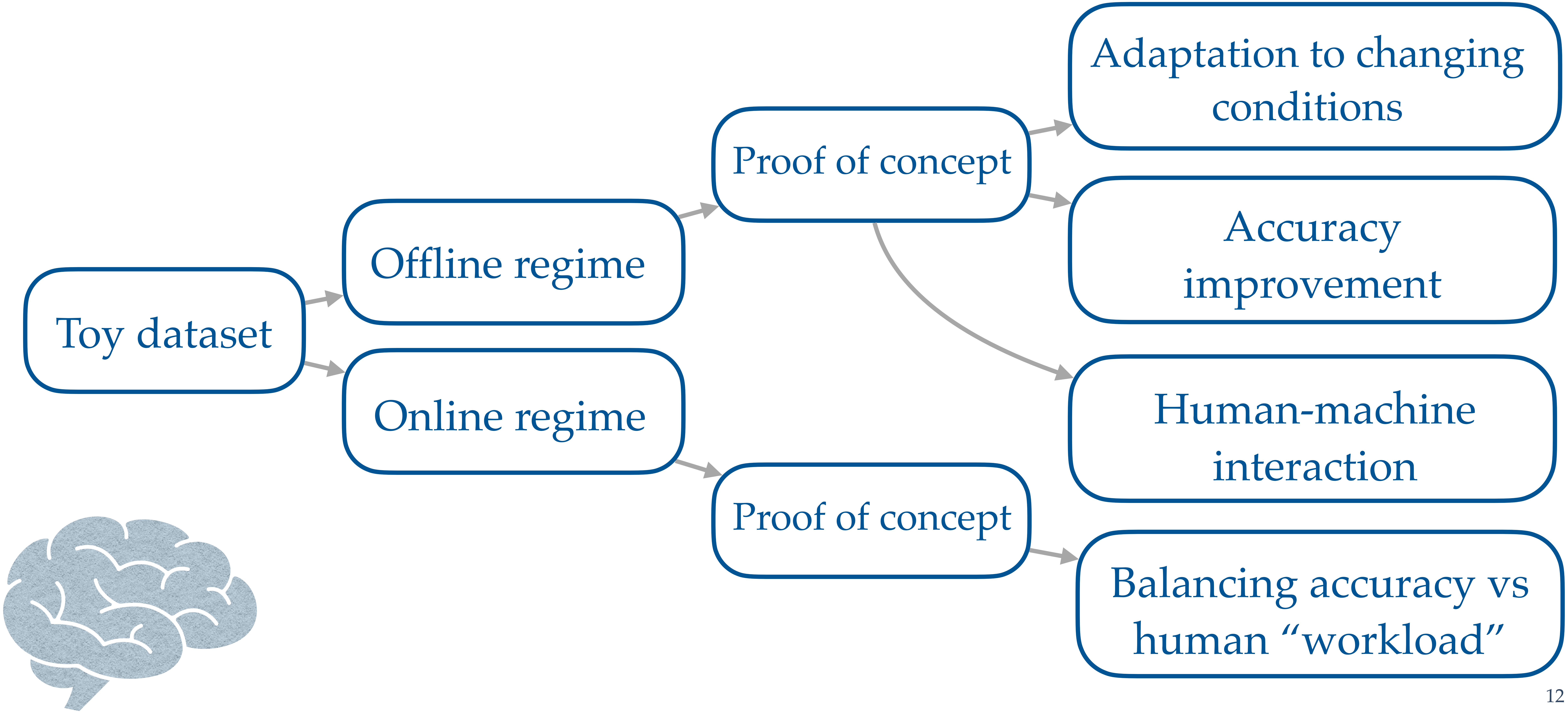


Reinforcement Learning (RL) with Human Feedback

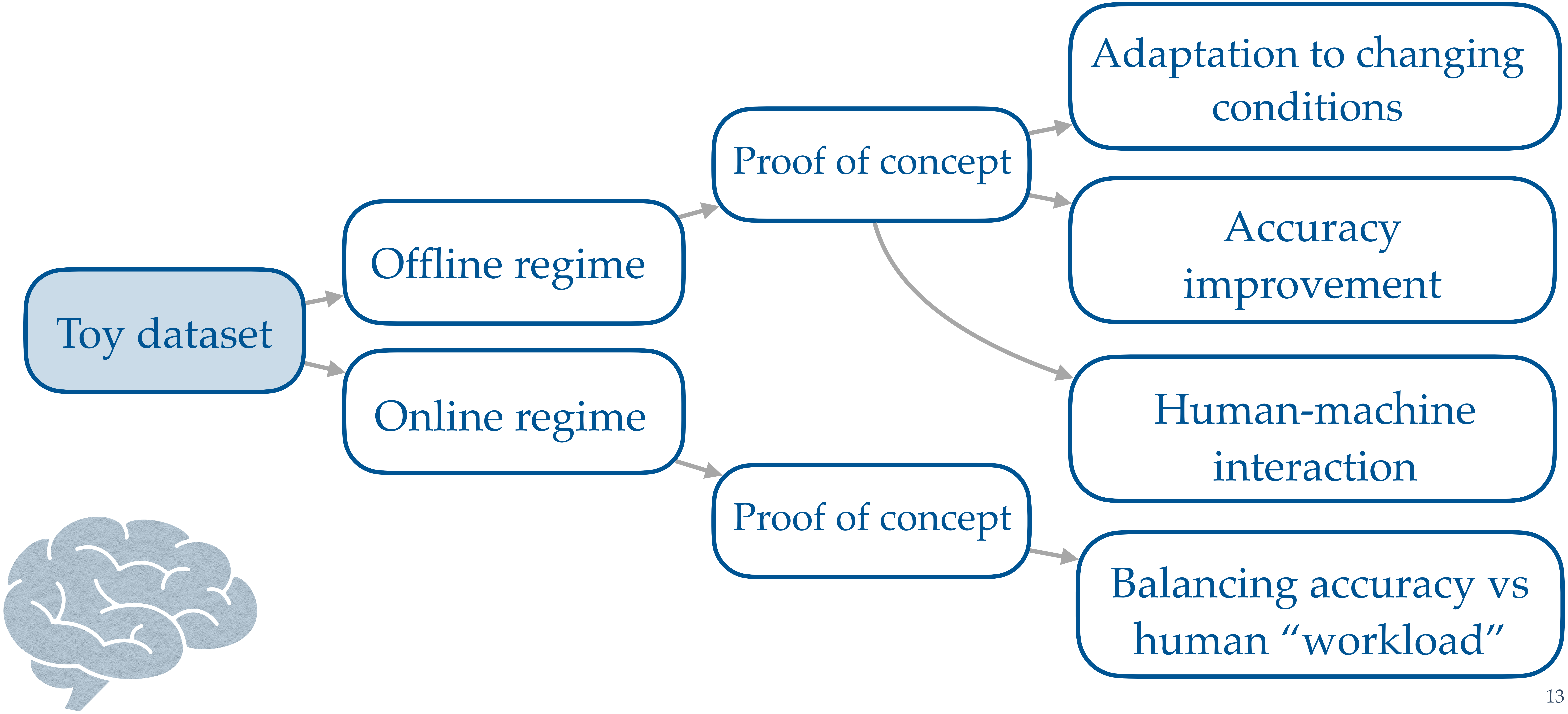
PPO RL algorithm
[arXiv:1707.06347]



Prototype and POC studies

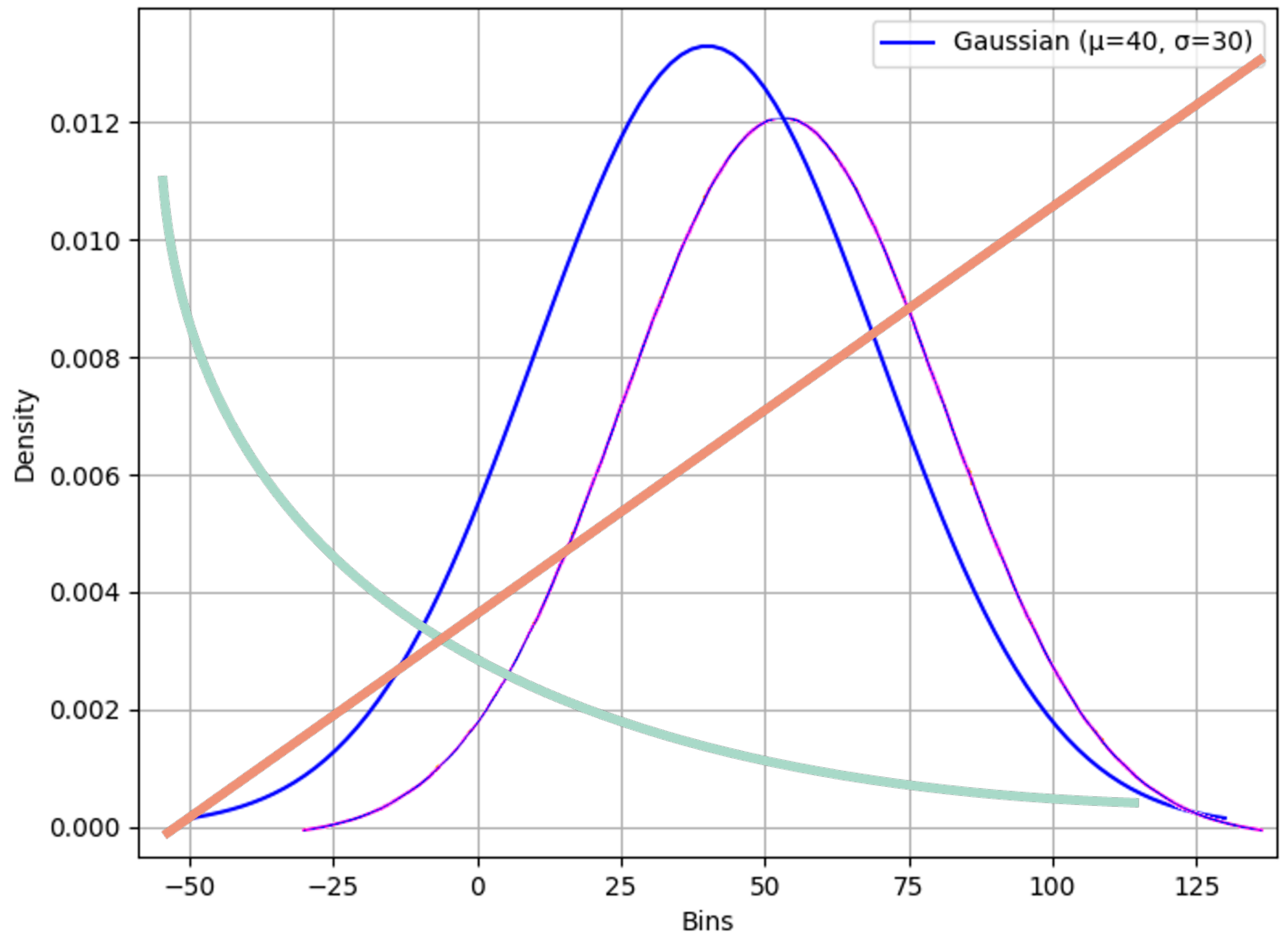


Prototype and POC studies



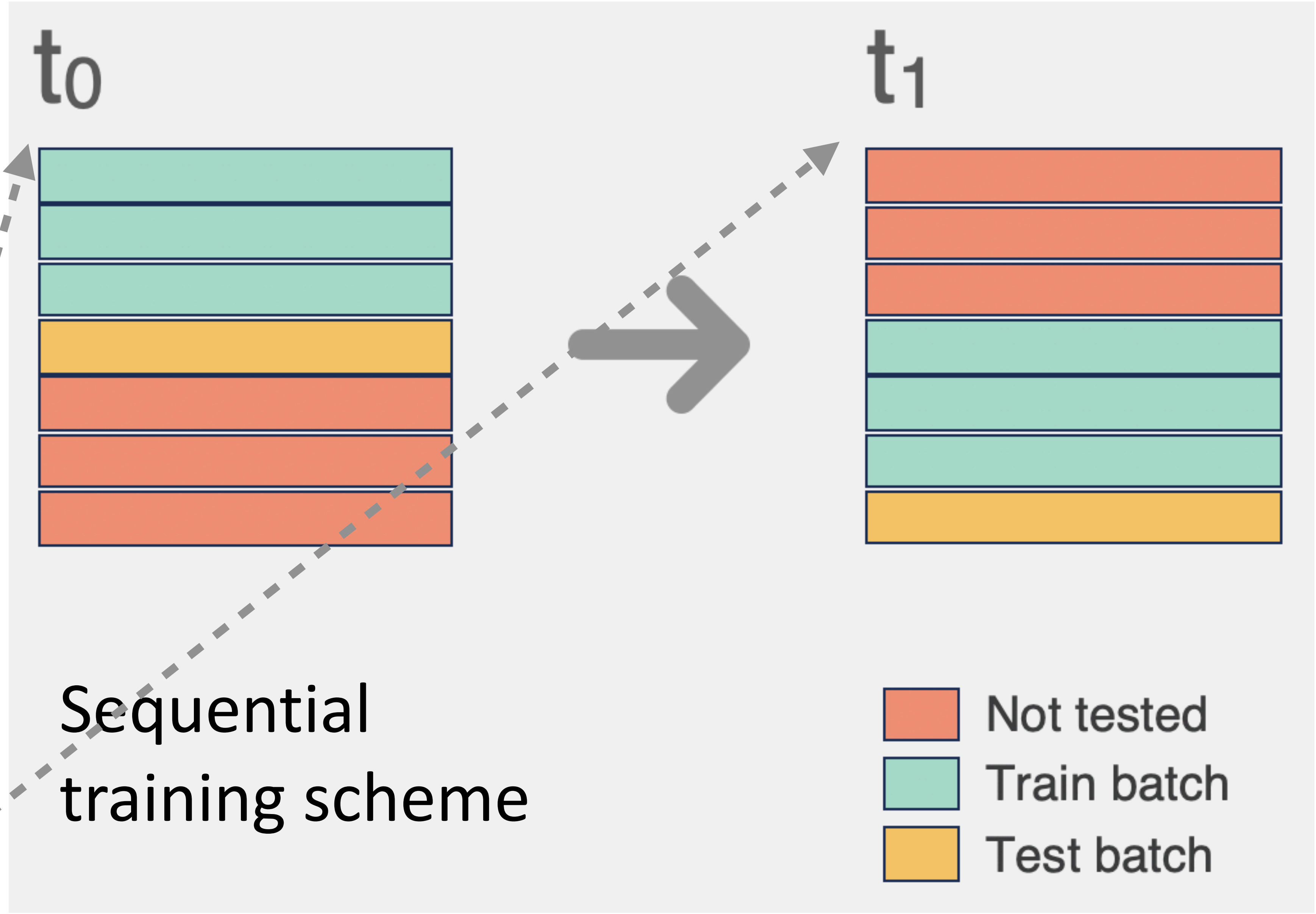
Toy dataset: data generation

- **1D histogram** with statistical noise
- Generation: histograms representing **nominal/anomalous distributions**



Toy dataset: time dependance

- The **histograms are ordered sequentially** to emulate the data collection
  - The type of (NOMINAL / ANOMALOUS) distributions used in generation are changed at specific points in time
- **The training is also done sequentially**, (potentially) in batches



Prototype and POC studies

Toy dataset

Offline regime

Monitoring after the data has been collected (high accuracy)

Proof of concept

Adaptation to changing conditions

Accuracy improvement

Human-machine interaction

Online regime

Real-time monitoring (fast decisions)

Proof of concept

Balancing accuracy vs human "workload"

Offline Regime

Histogram

- The histograms are fully independent from each other with a fixed probability of being anomalous

- Time dependency: change in the type of distribution representing anomaly or nominal status

Offline Regime

Histogram

- The histograms are fully independent from each other with a fixed probability of become anomalous

- Time dependency: change in the type of distribution representing anomaly or nominal status

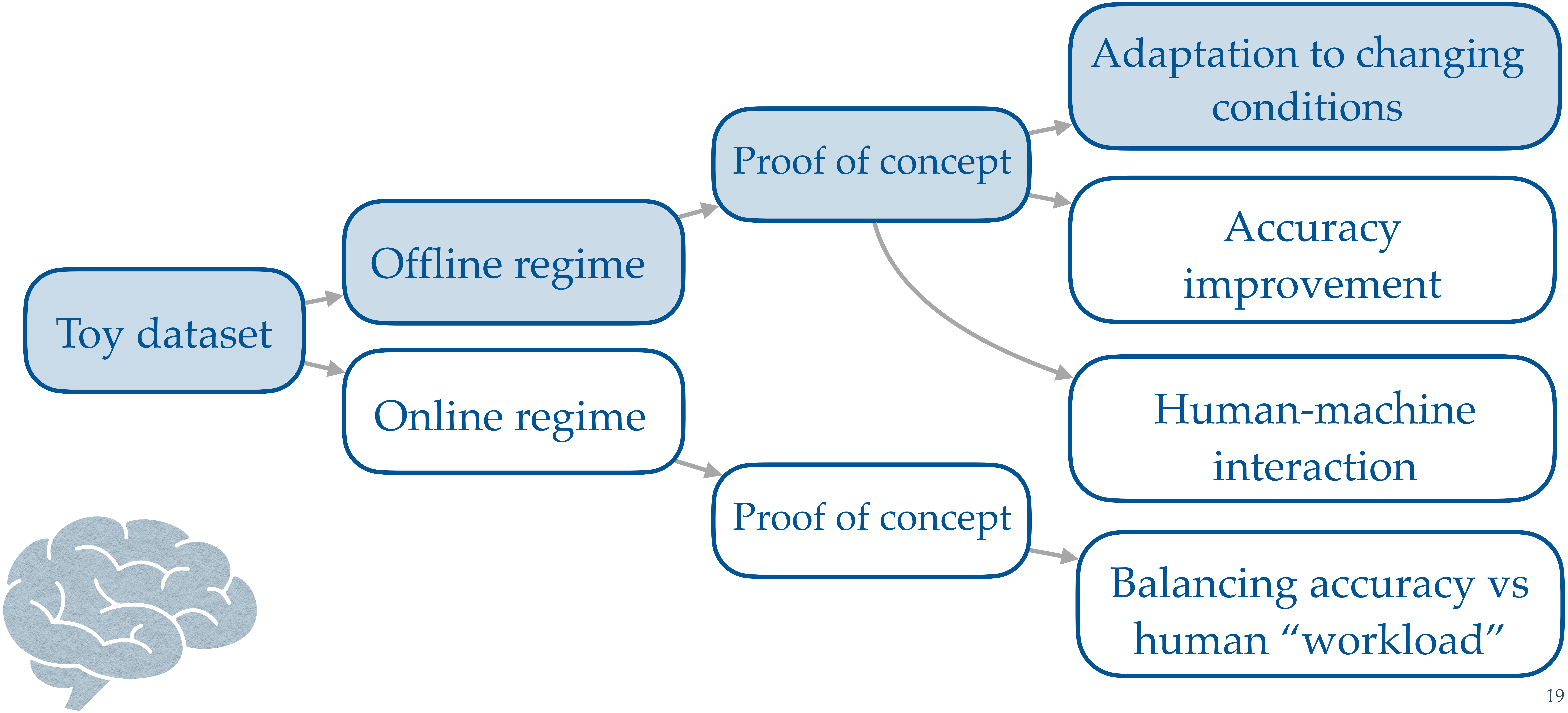
Algorithm's output

- One agent to classify between nominal or anomalous

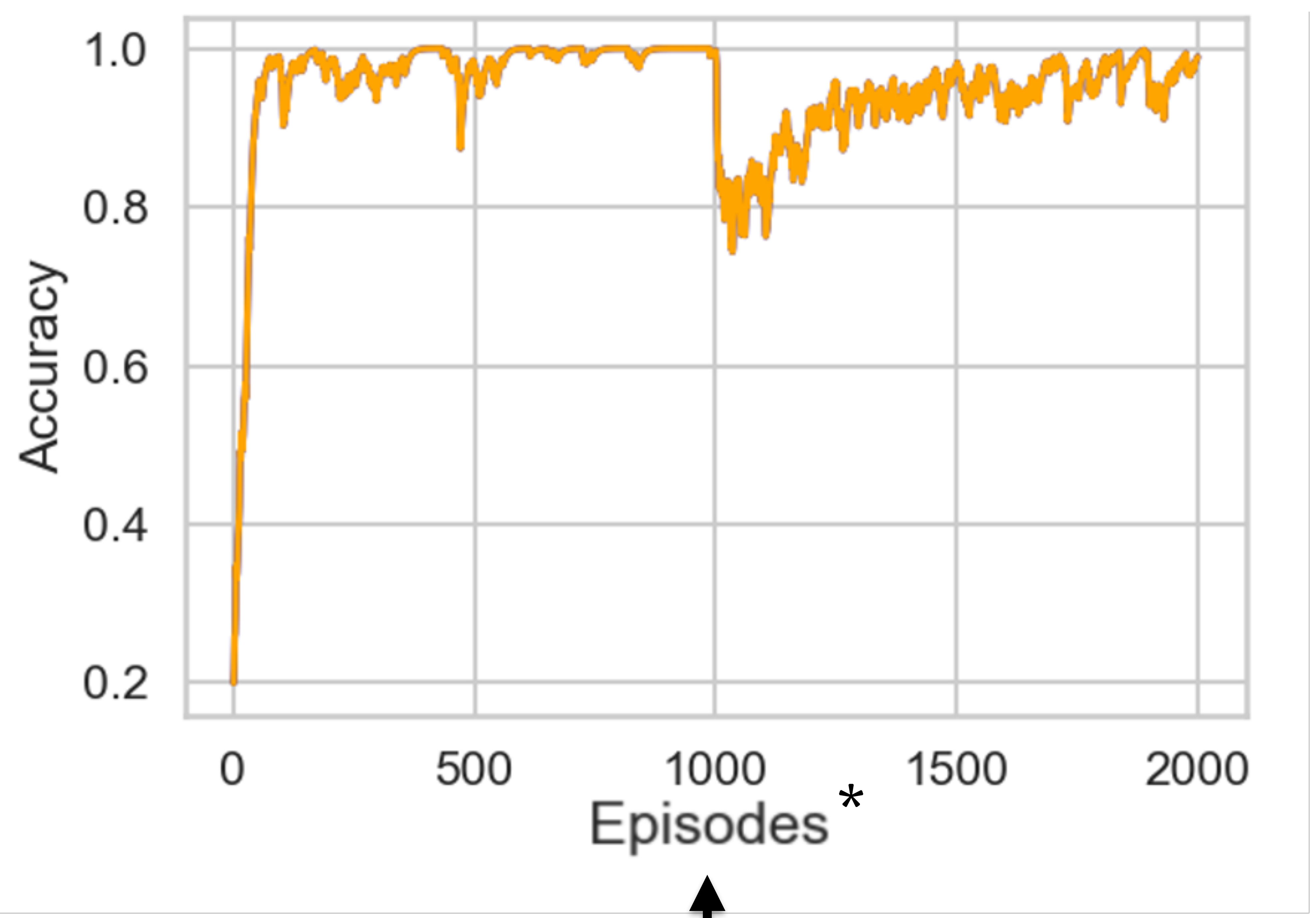
Reward

- Human feedback (+1/-1) for a correct or incorrect classification for all the histograms

Prototype and POC studies



Adaptation to changing conditions

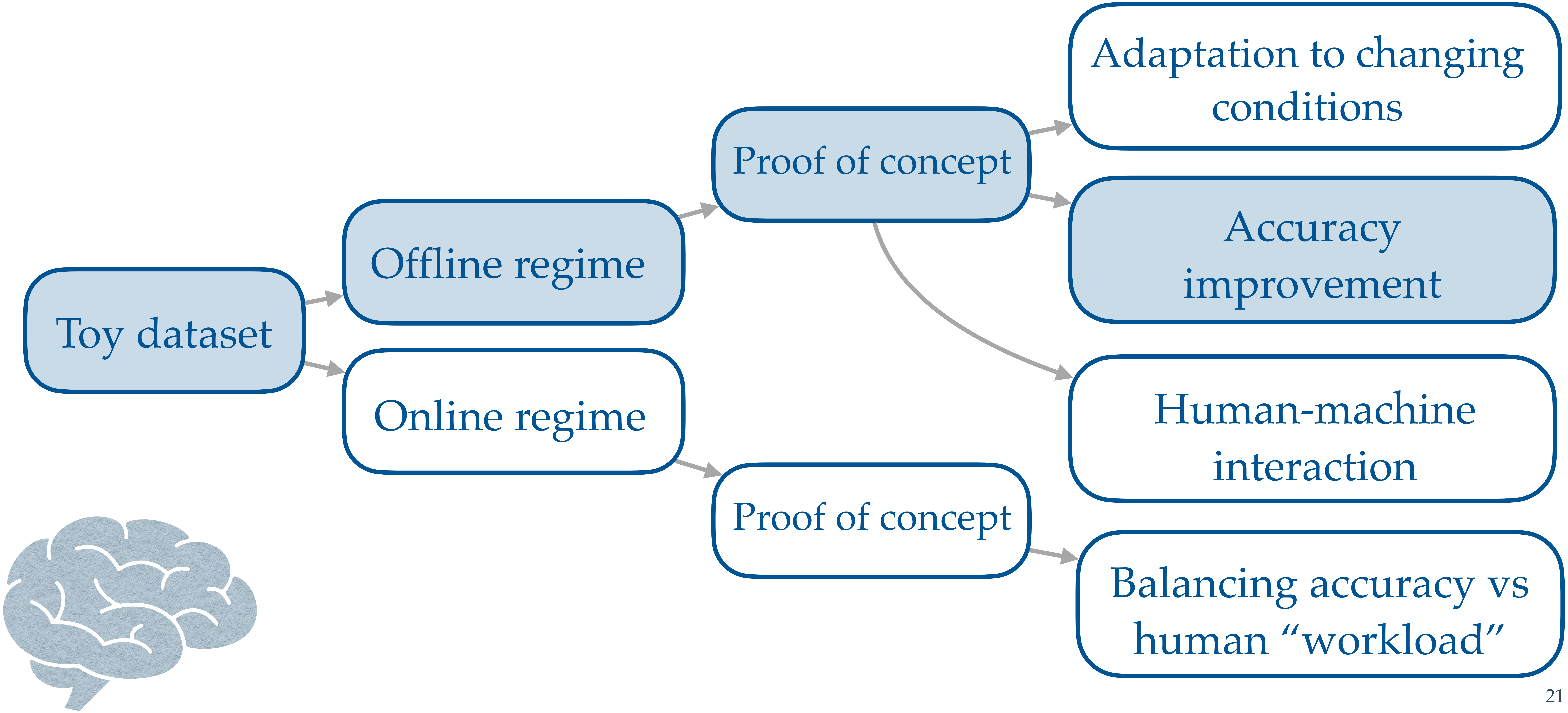


Episode*: Individual Histogram

Abrupt change in nominal conditions introduced

The algorithm adapts automatically to the new nominal conditions.

Prototype and POC studies



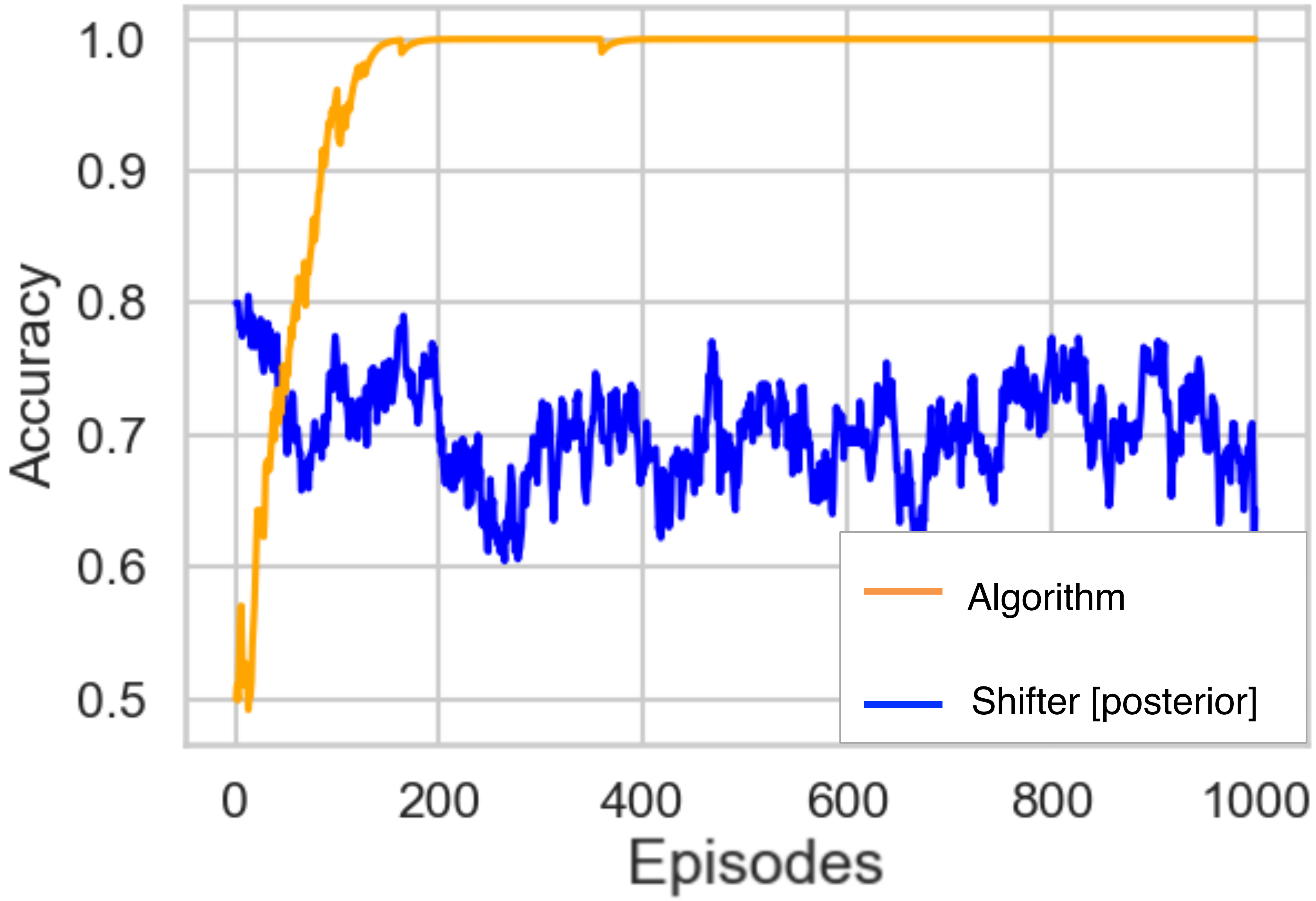
Accuracy improvement

Can the algorithm improve the shifter's accuracy?

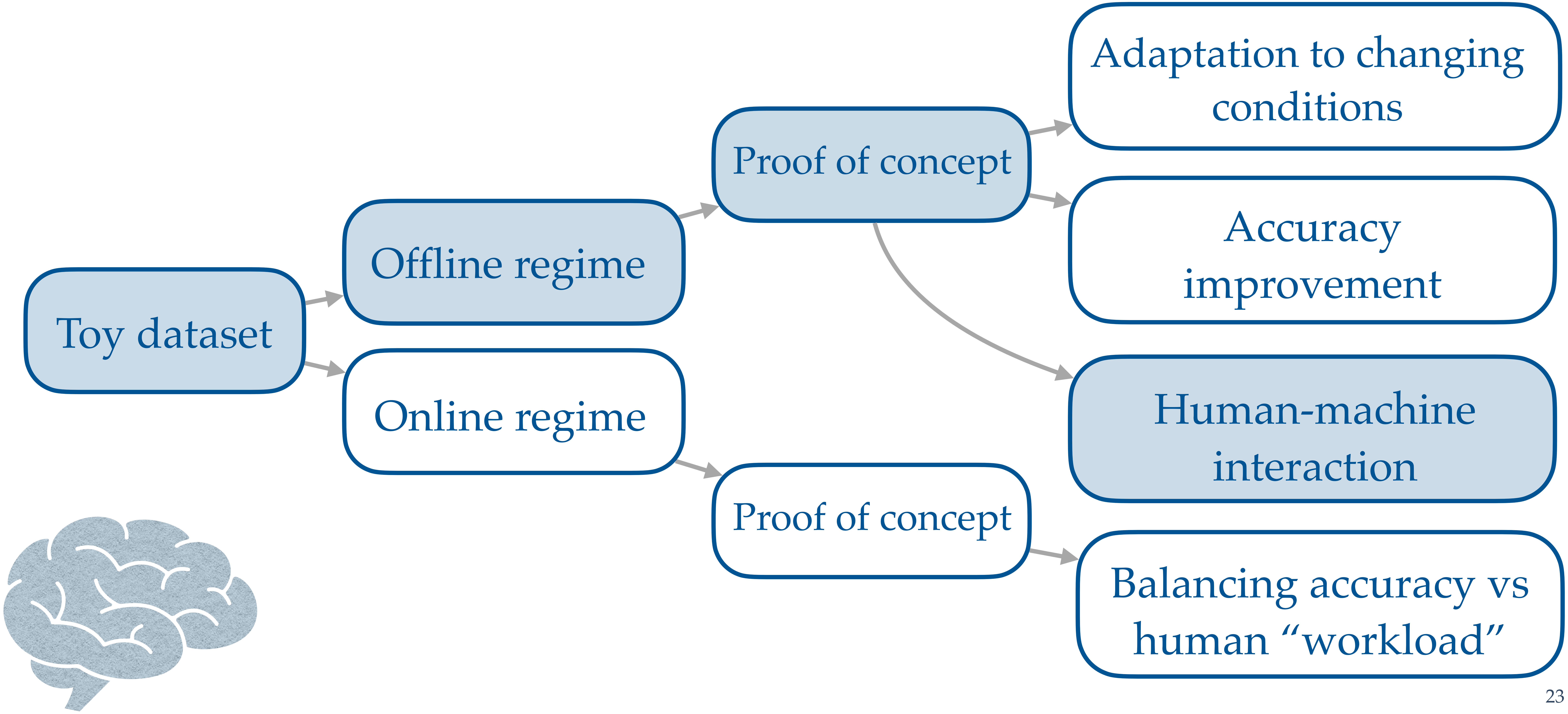
Simple experiment

We swap the target label in 30% of the cases during training, and evaluate the true accuracy of the algorithm

The algorithm learns how to filter th noise and achieve a higher accuracy than the shifter



Prototype and POC studies



Human-machine interaction

What happens when the human enters in the loop?

- Would the shifters improve their accuracy if they could see the algorithm's output beforehand?
    - If so, would the algorithm still learn from the resulting shifters' predictions?

Human-machine interaction

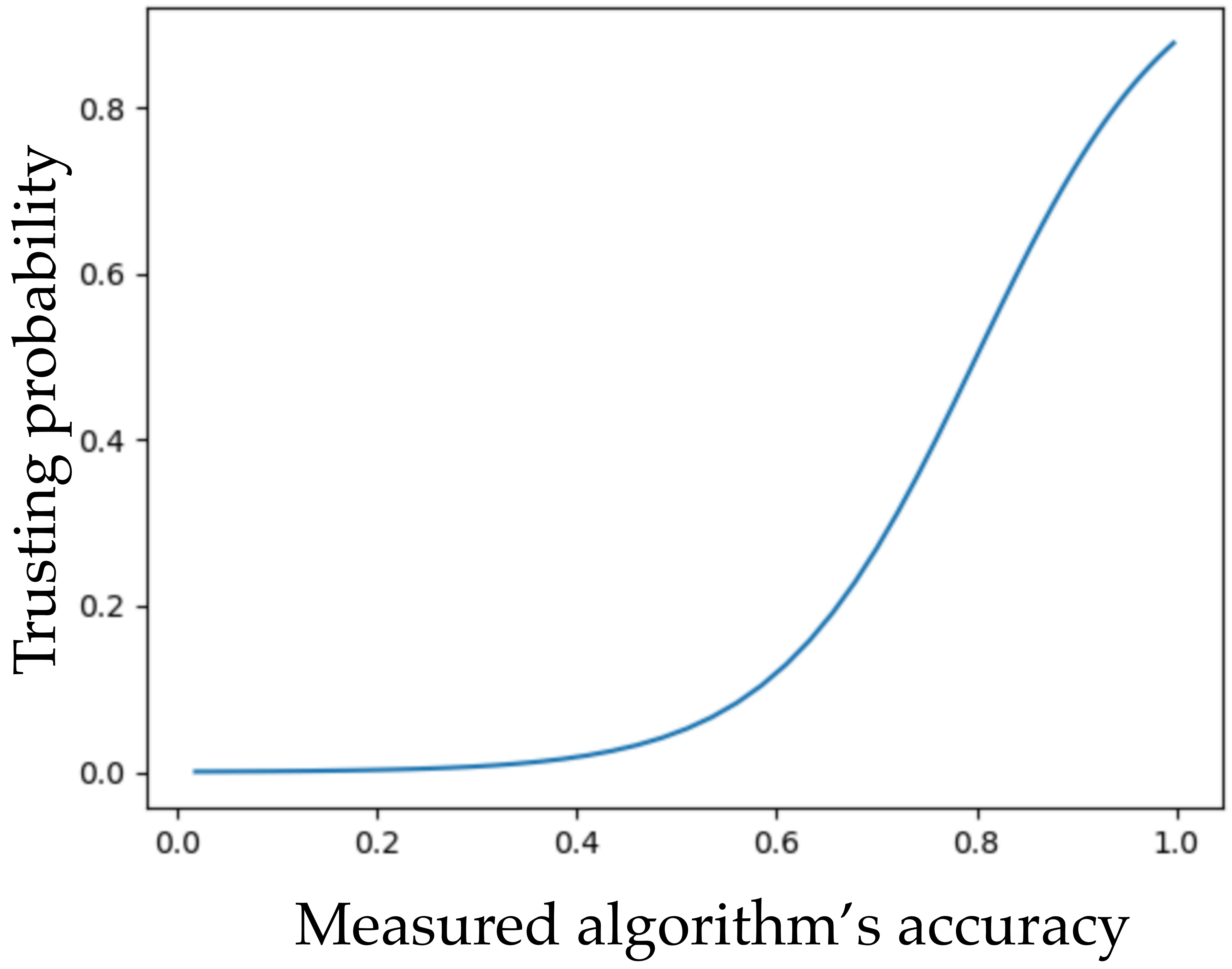
What happens when the human enters in the loop?

Simple experiment

- The emulated shifter has access to the algorithm's accuracy, measured with respect to the previous shifter's labels

- **We assume that the shifter randomly "trusts" the algorithm with a probability that increases with the accuracy measured in the recent past**

Heuristic function used in the study

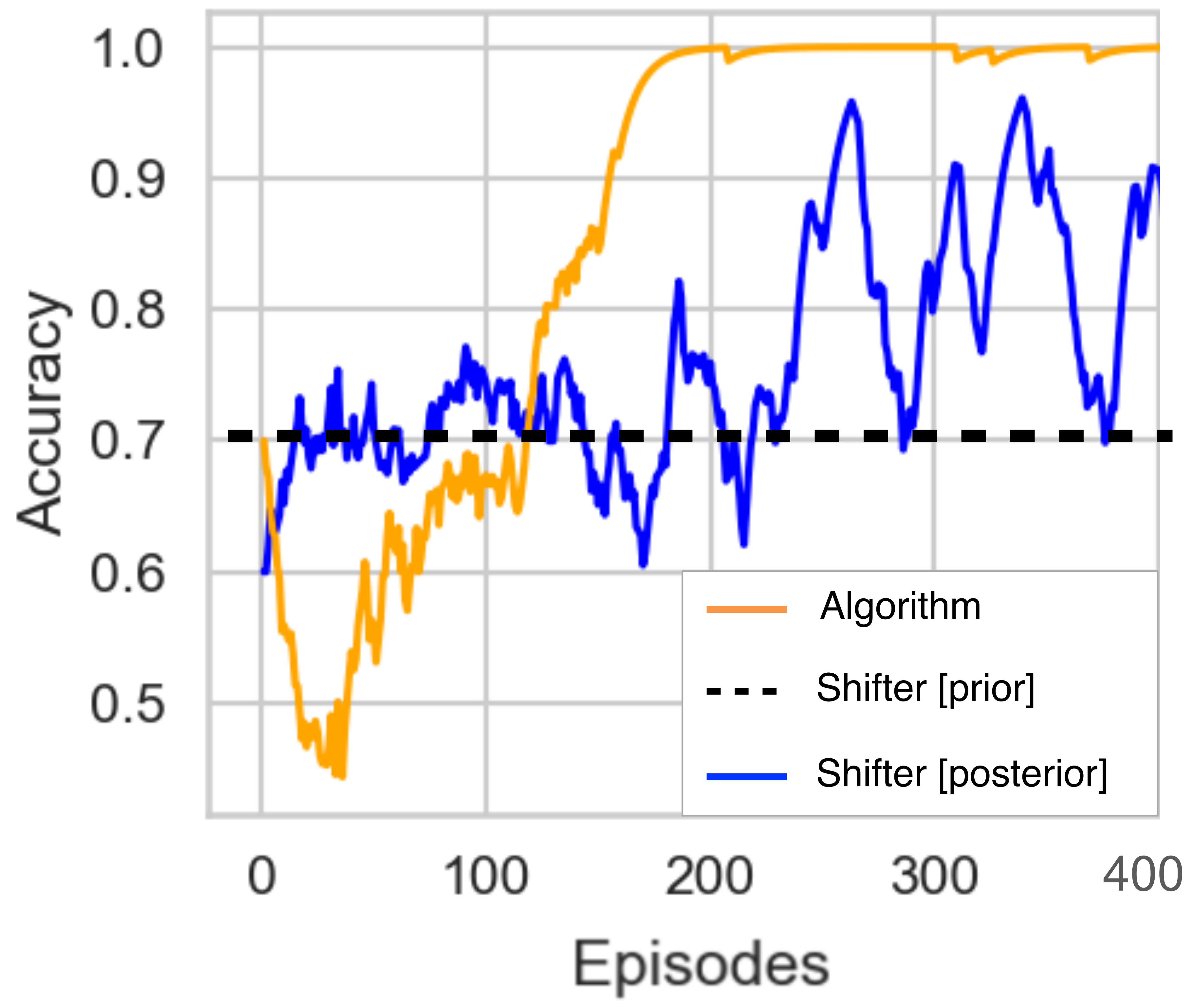


Accuracy improvement

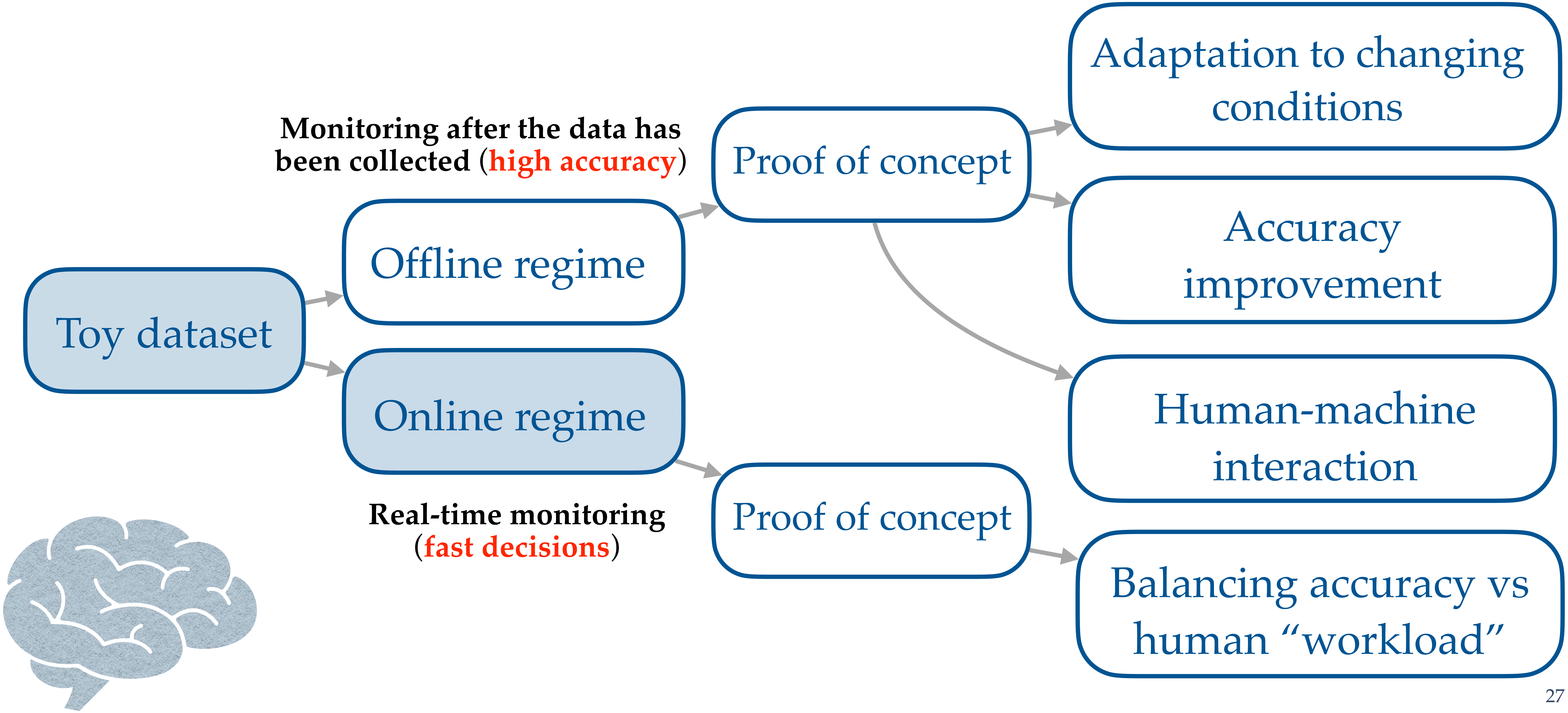
What happens when the human enters in the loop?

Simple experiment

- **The shifters improve their accuracy** if they can see the algorithm's output beforehand
- **The algorithm still learns** from the resulting shifters' predictions



Prototype and POC studies



Online Regime

Histogram

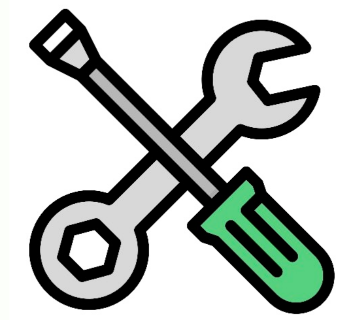
- Fixed probability of being anomalous. The **anomaly persists until it is correctly detected** by the algorithm (concept of **‘problem fixing”**)

- The **label** of the histogram is **only available** when the **shifter is called** by the algorithm or then the shifter randomly decides to take a look at the data (**checkpoint**).

Online Regime

Histogram

- Fixed probability of being anomalous. The **anomaly persists until it is correctly detected** by the algorithm (concept of **”problem fixing”**)



- The **label** of the histogram is **only available** when the **shifter is called** by the algorithm or then the shifter randomly decides to take a look at the data

Algorithm’s output

- One agent to determinate the system status (**predictor**) and another to call the shifter (**checker**)

Online Regime

Histogram

- Fixed probability of being anomalous. The **anomaly persists until it is correctly detected** by the algorithm (concept of '**problem fixing"**)

- The **label** of the histogram is **only available** when the **shifter is called** by the algorithm or then the shifter randomly decides to take a look at the data (**checkpoint**).

Algorithm's output

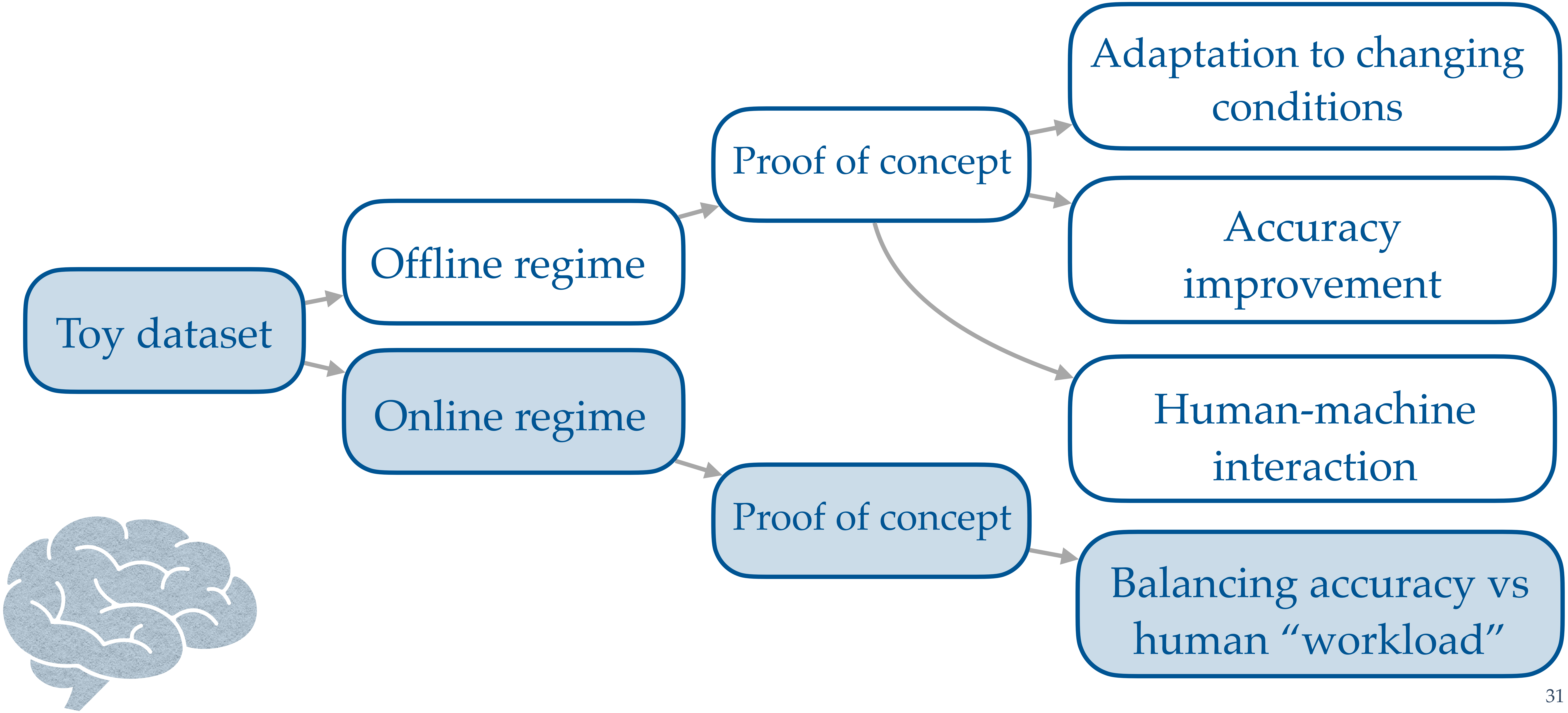
- One agent to determinate the system status (**predictor**) and another to call the shifter (**checker**)

Reward

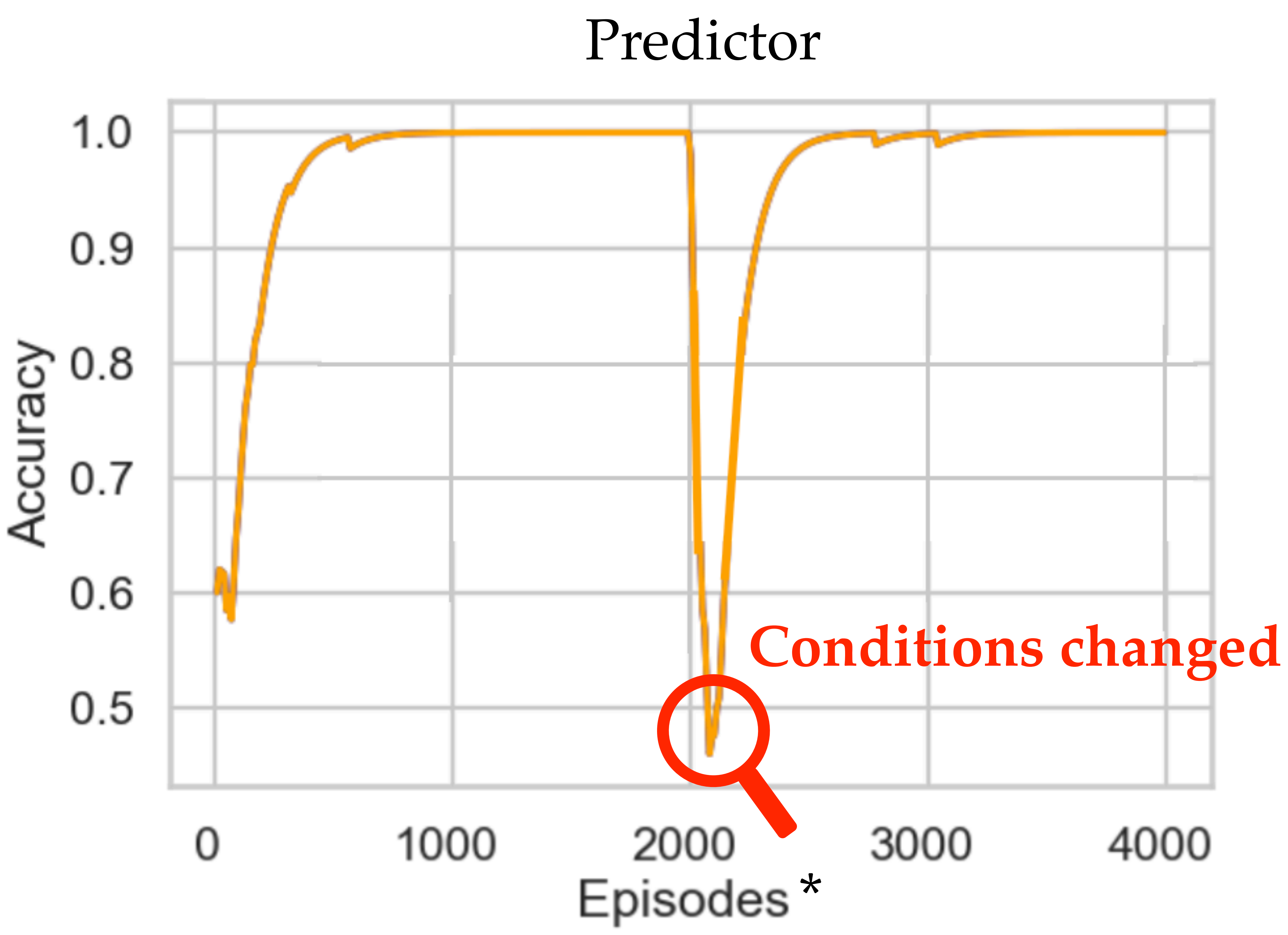
- Predictor: Same as the offline

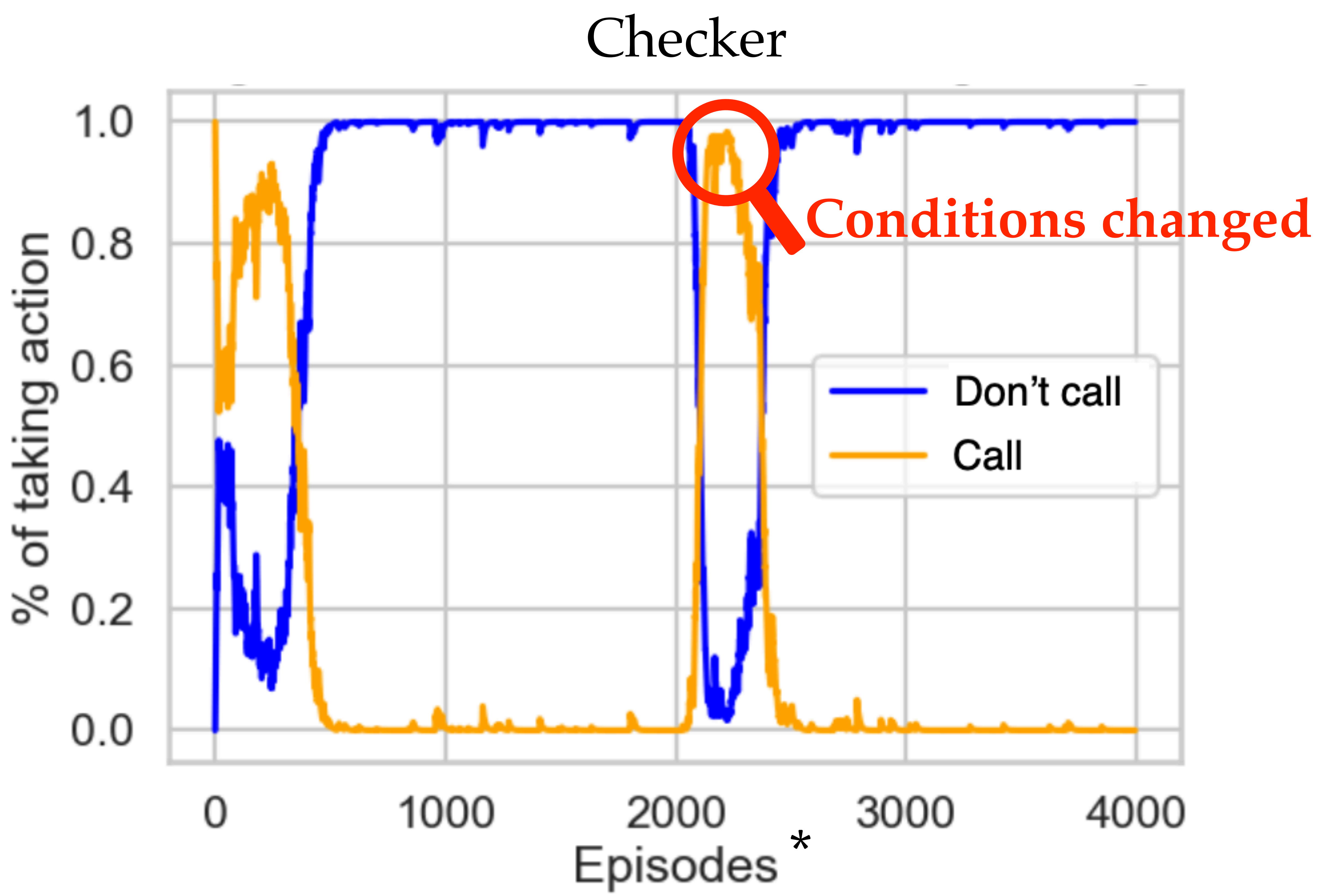
- Checker: derived from the predictor's confidence on its decision

Prototype and POC studies



Balancing accuracy vs human “workload”





Episode*: Group of histograms between checkpoints

High accuracy achieved with a limited number of calls to the shifter, which are focused only on the critical moments

Conclusions

- **Novel approach** towards automating DQM at HEP experiments
    - **Reinforcement Learning** used to optimise Human-Machine interaction and and adapt to changing operational conditions
- Prototype and proof of concept studies done:
    - **Offline:** Accuracy gain by combined human-machine training
    - **Online:** Continuous automated monitoring in real time, calling the shifter when relevant

Outlook

Useful for low statistics data?

- Use of data augmentation techniques for low statistics data
  - Going towards a real case scenario

Thank you for your attention!

Proximal Policy Optimization (PPO)

- PPO uses the **advantage function**: the critic evaluates how much better the actor prediction is comparing it to the average prediction presented by the policy and the given reward

- PPO **maximises a surrogate objective**: improving the policy average while not making big changes in the actor's decisions

- In addition, we use **clipping to ensure stability** on the policy update