UNIVERSITY AT ALBANY
STATE UNIVERSITY OF NEW YORK

CSI 436/536 (Spring 2025)

Machine Learning

Lecture 19: Advanced Topic - Decision Making

Chong Liu

Department of Computer Science

Apr 21, 2025

Machine learning overview

- Supervised learning
  - Tasks: classification, regression
  - Techniques: regularization, error decomposition, kernel, ensemble, neural network
- Unsupervised learning
  - Clustering, dimension reduction
- Some many advanced topics:
  - More tasks: ranking
  - More theoretical understanding: statistical learning theory
  - More settings: reinforcement learning
  - More applications: computer vision, natural language processing, speech recognition
  - More perspective: rule learning
  - More trustworthy: explainable ML, privacy / fairness issues in ML
  - …

Today

- Decision making-based machine learning
  - Active learning
    - Disagreement-based algorithm
    - Confidence-based algorithm
    - Clustering-based algorithm
  - Bandits
  - Reinforcement learning

Today

- Decision making in machine learning

  - Active learning
    - Disagreement-based algorithm
    - Confidence-based algorithm
    - Clustering-based algorithm

  - Bandits

  - Reinforcement learning

Binary image classification tasks

Dog or mop?

Binary image classification tasks

Dog or croissant?

Binary image classification tasks

Dog or bagel?

Binary image classification tasks

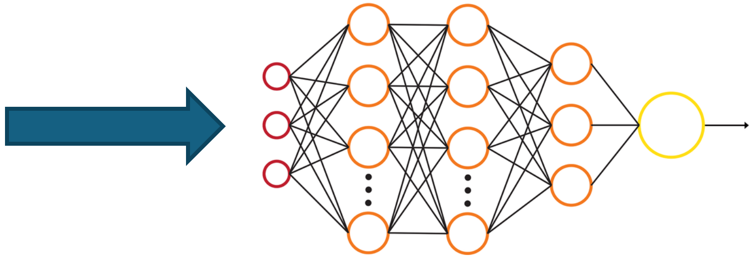
Dog or muffin?

Binary image classification tasks

Dog or fried chicken?

Recap: Supervised learning

- How to achieve good performance?
  - Advanced techniques: kernel, ensemble, neural networks ...
  - Input **a lot of labeled images** of dogs or fried chicken

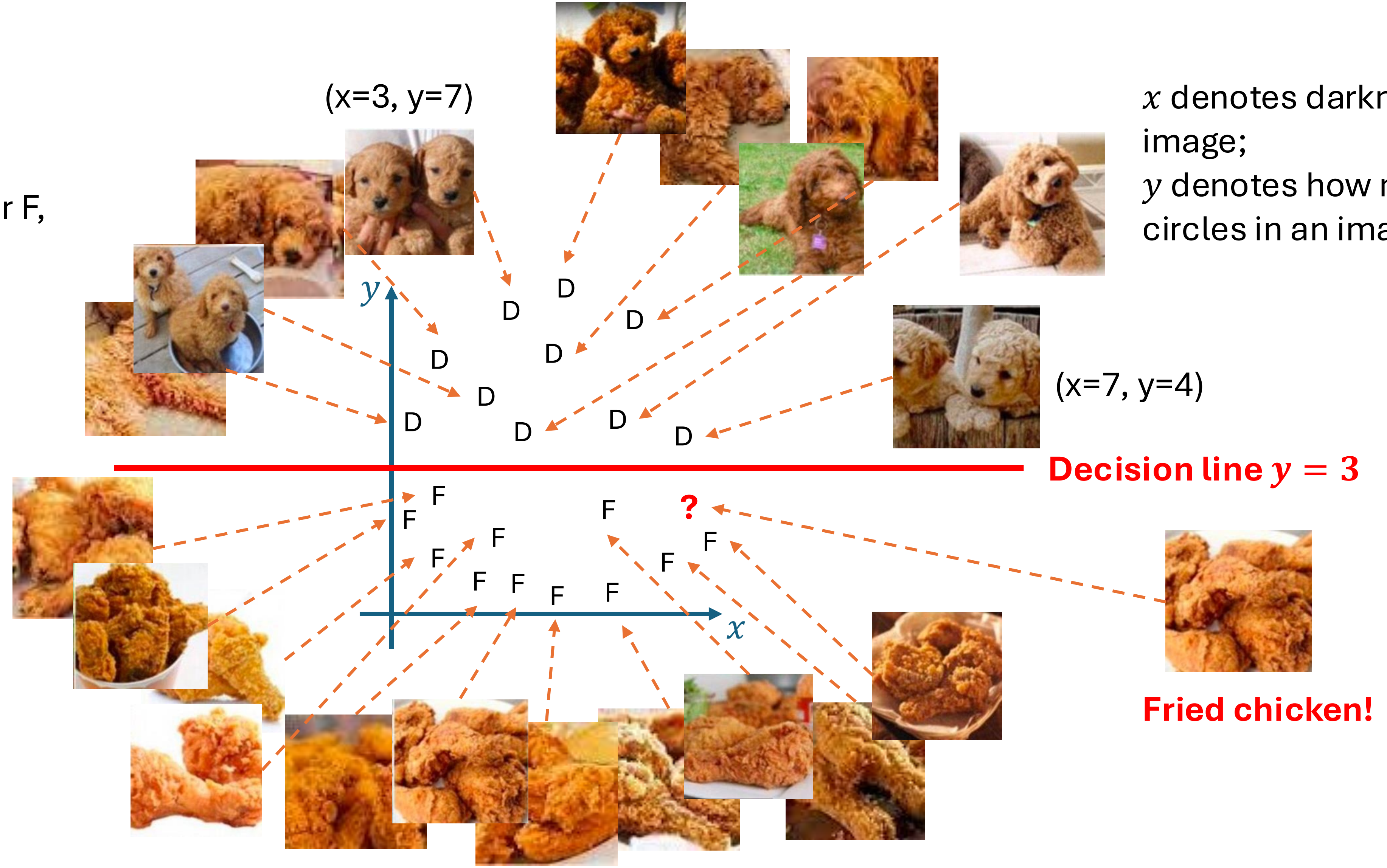


Given a new image, dog or fried chicken?

Binary classification in 2-dimension

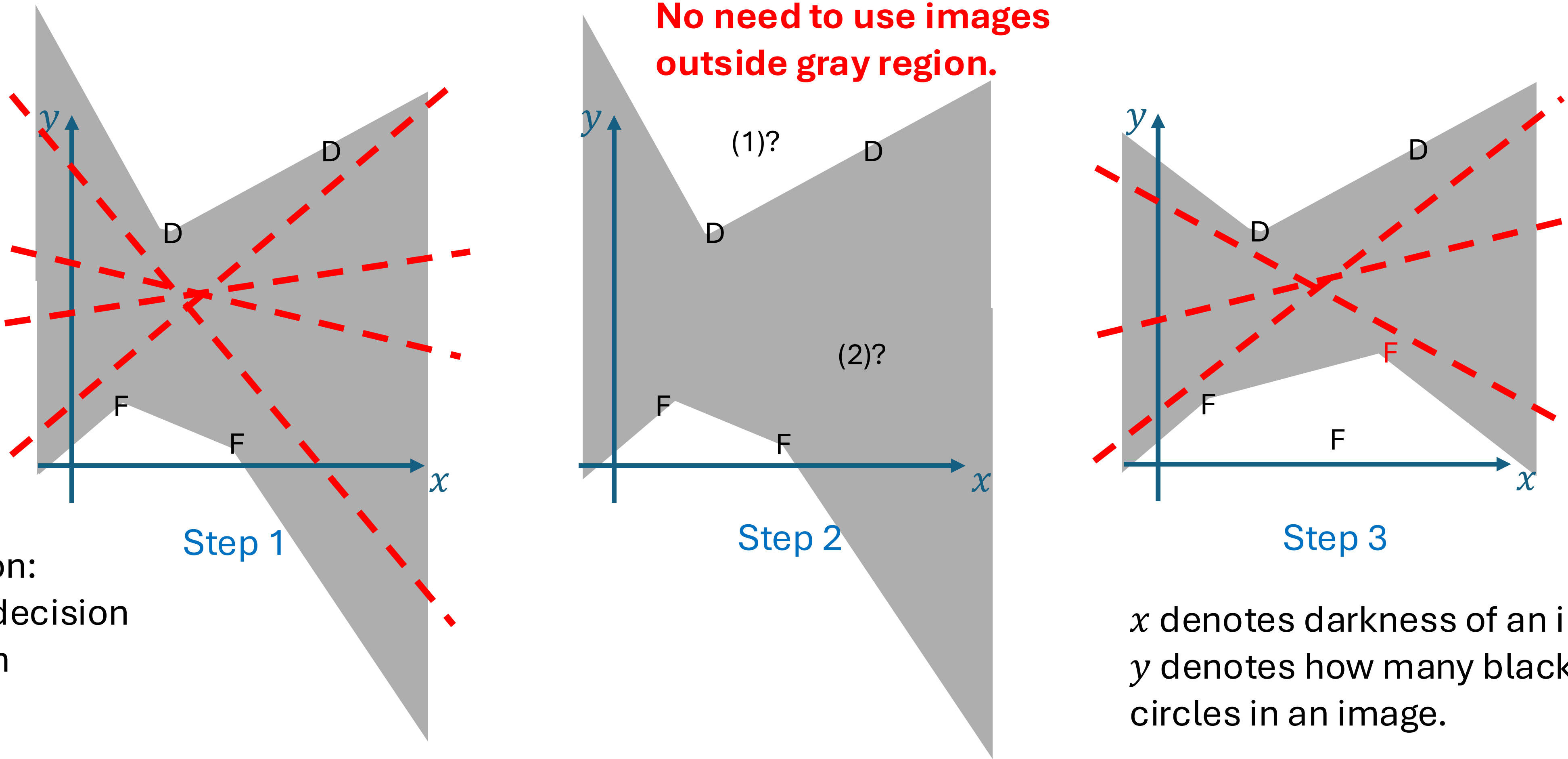
Each point, D or F, represents an image.

x denotes darkness of an image;
y denotes how many black circles in an image.



Fewer images to learn the decision line?

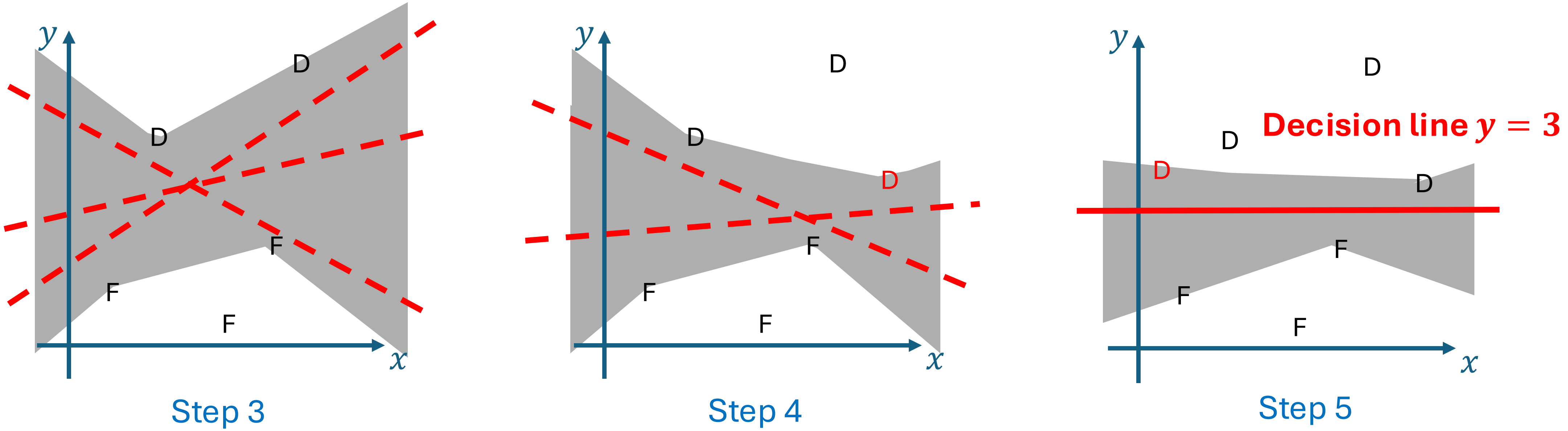
- Yes! Active learning selects important images.



Gray region:
Feasible decision
line region

x denotes darkness of an image;
y denotes how many black
circles in an image.

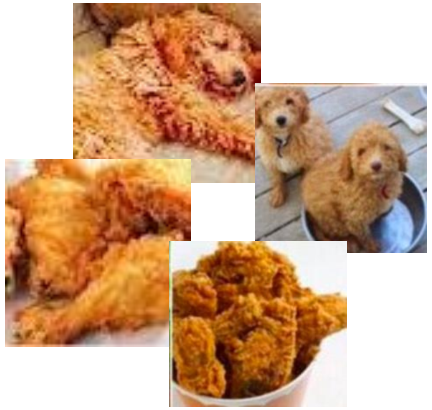
Moving forward ...



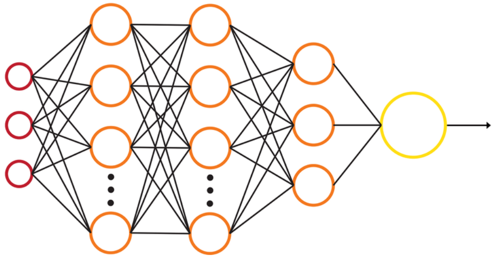
- Learn decision line with only 7 images!
  - Active learning only selects important images
    - images in gray region

Disagreement-based Active Learning

Confidence-based active learning



Labeled data



Unlabeled data

How to select the data point with
least prediction confidence?

Clustering-based active learning

- Run k-means clustering
- Select samples from each cluster

Summary: active learning

- Framework:
  - Input: a small labeled dataset, a large unlabeled dataset
  - Output: a well-trained model

  - For $t = 1,2,\ldots$
    - Select a sample x_t from the unlabeled dataset
    - Get its label y_t
    - Update the model by adding (x_t, y_t)

- Goal:
  - Achieve good performance while saving labeling cost at the same time

- Can be super useful in practice!

Today

- Decision making in machine learning
  - Active learning
    - Disagreement-based algorithm
    - Confidence-based algorithm
    - Clustering-based algorithm
  - Bandits
  - Reinforcement learning

The real bandit machine

- Action
- Reward

- Goal:
  - maximize your reward

- Challenge:
  - You don't know what's behind the bandit machine

Expensive real-world drug discovery process

Propose drug candidates

- Given the target disease
- Propose 5k-10k candidate compounds

Expensive real-world drug discovery process

Propose drug candidates

- Given the target disease
- Propose 5k-10k candidate compounds

Drug screening

- Select most promising drugs from the pool
- Criterion: max binding affinity to the target

Expensive real-world drug discovery process

Propose drug candidates

- Given the target disease
- Propose 5k-10k candidate compounds

Drug screening

- Select most promising drugs from the pool
- Criterion: max binding affinity to the target

FDA approval process

- Lab test
- Human test (Phase I, II, III)
- New drug application
- Review meeting
- Facility Inspection
- Approval
- Post-market risk assessment

A new drug into market: $2.8 billion dollars and 10-15 years [Pushpakom et al., Nat Rev Drug Discov, 2018]

AI is now amazing. How can AI help?

Currently AI helps in Stage I & II.

Propose drug candidates

- Given the target disease
- Propose 5k-10k candidate compounds

Generative AI

Drug screening

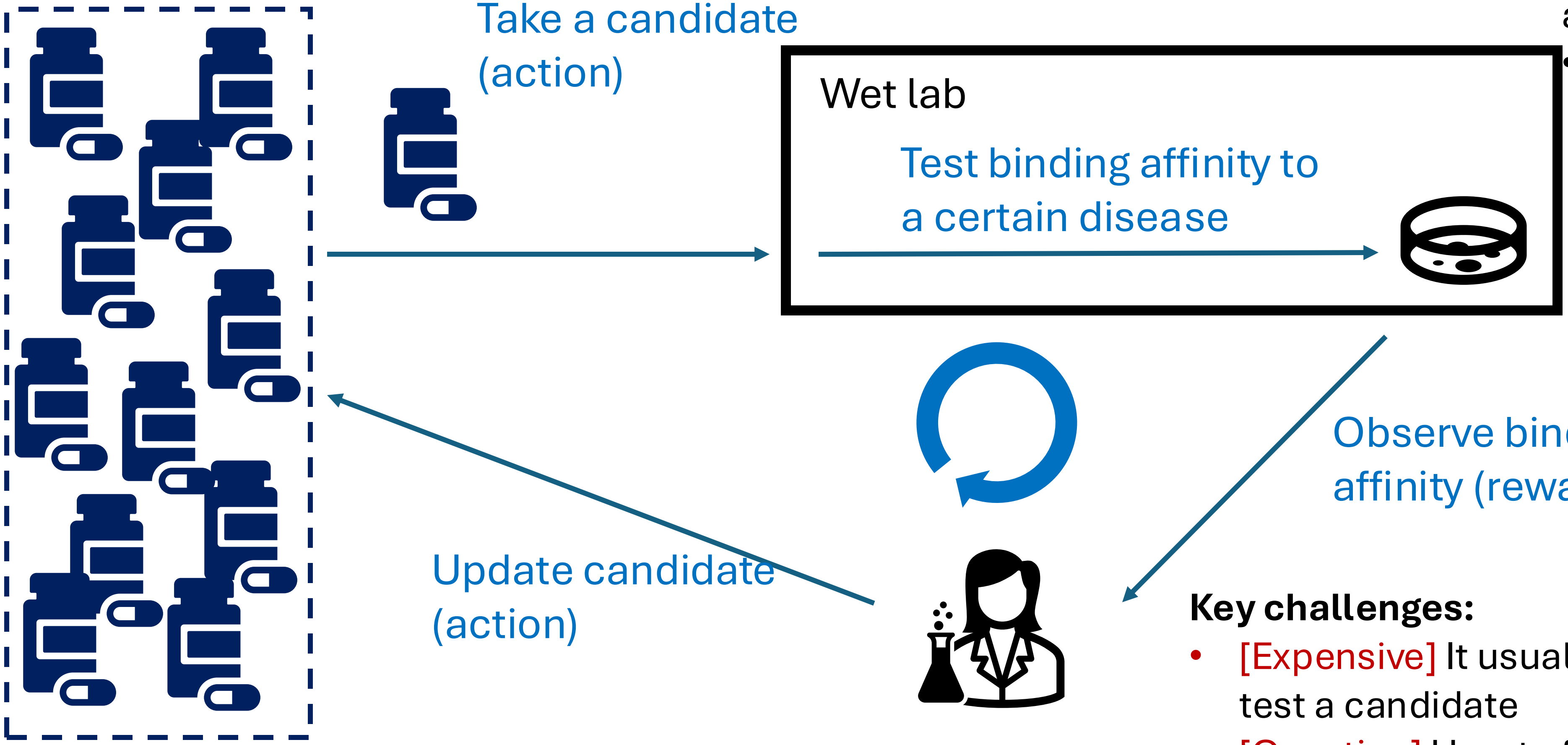
- Select most promising drugs from the pool
- Criterion: max binding affinity to the target

Bandit problem

FDA approval process

- Lab test
- Human test (Phase I, II, III)
- New drug application
- Review meeting
- Facility Inspection
- Approval
- Post-market risk assessment

Bandits in Stage II Drug screening



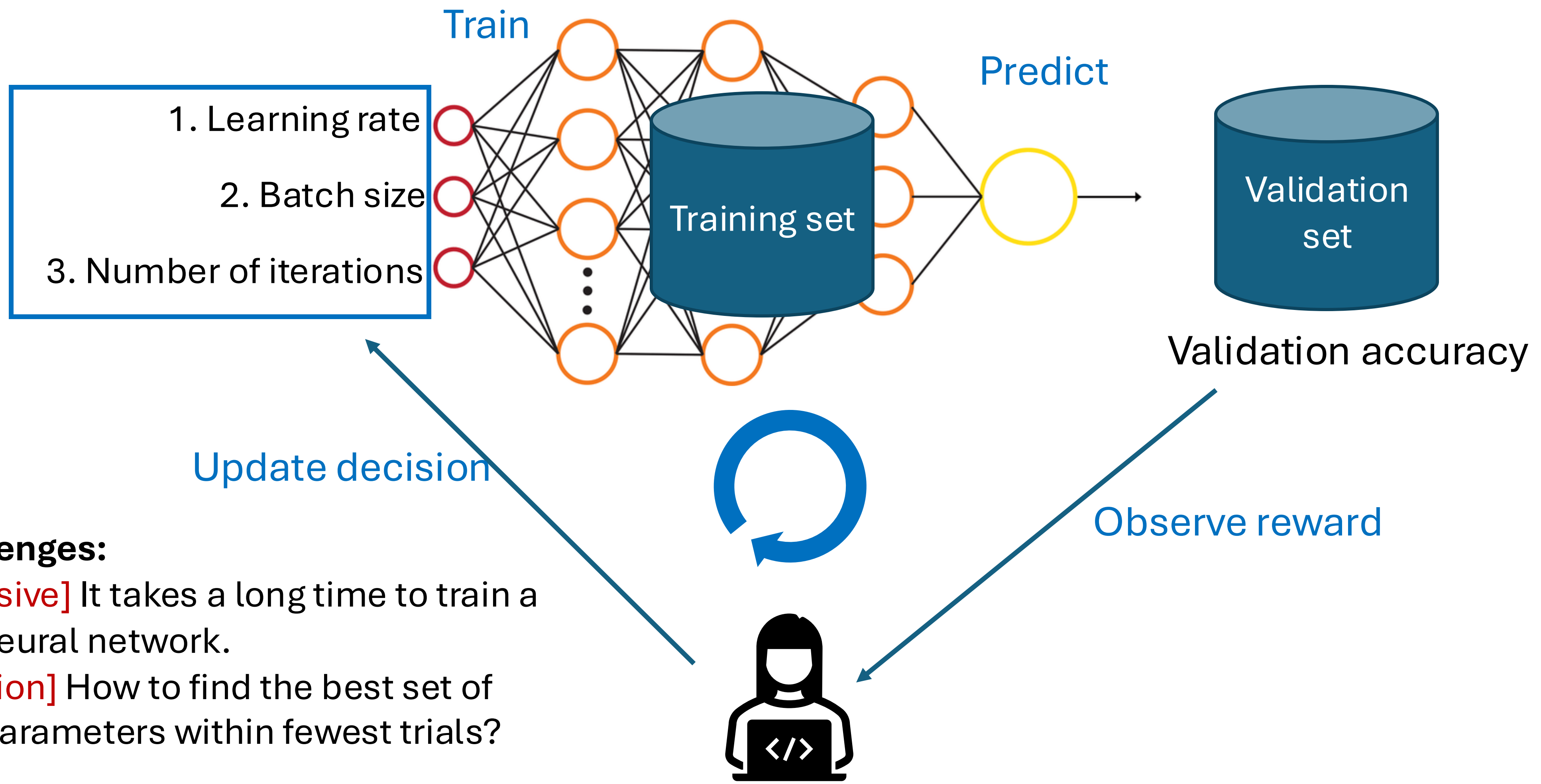
Why **max** binding affinity?

- If binding affinity is strong, the drug can take away the disease!

Key challenges:

- [Expensive] It usually takes one week to test a candidate
- [Question] How to find the best candidates within fewest days?

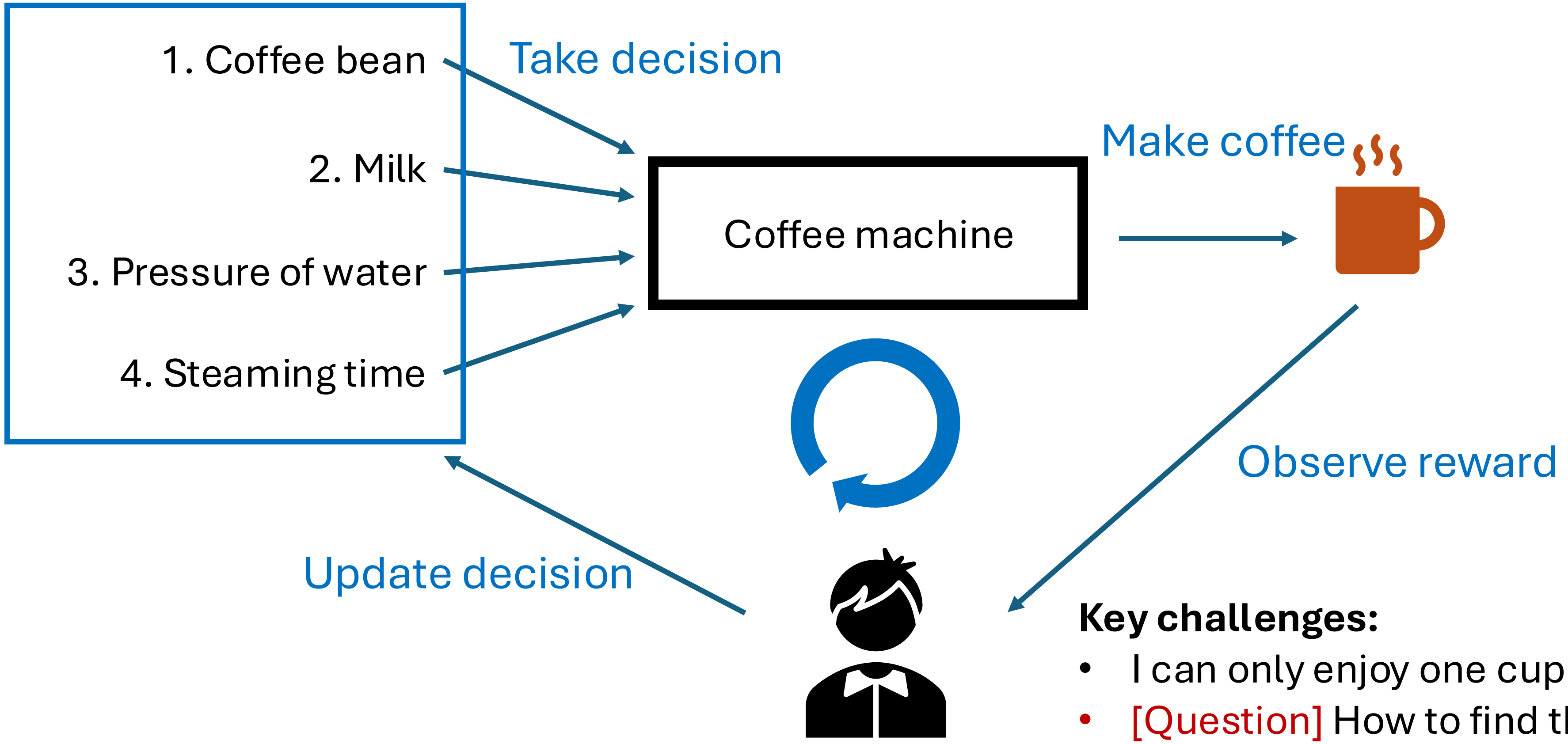
Bandits in hyperparameter tuning



Key challenges:

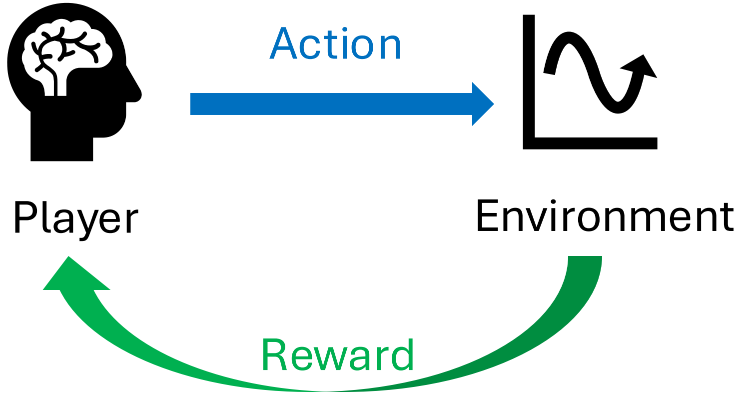
- [Expensive] It takes a long time to train a deep neural network.
- [Question] How to find the best set of hyperparameters within fewest trials?

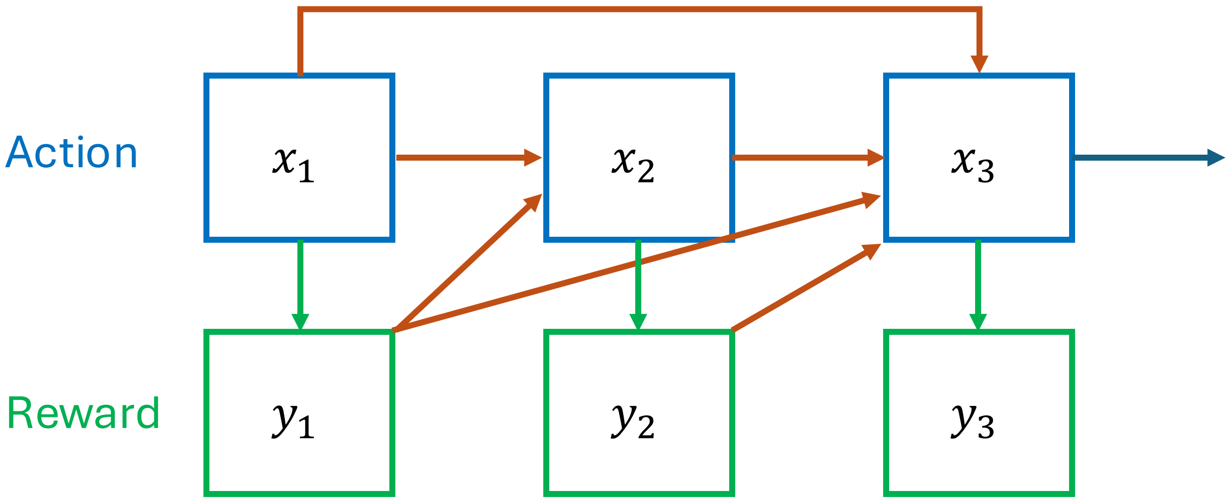
Bandits in making a cup of latte ...



Key challenges:
- I can only enjoy one cup per day.
- [Question] How to find the best way to make coffee within fewest days?

Bandits framework

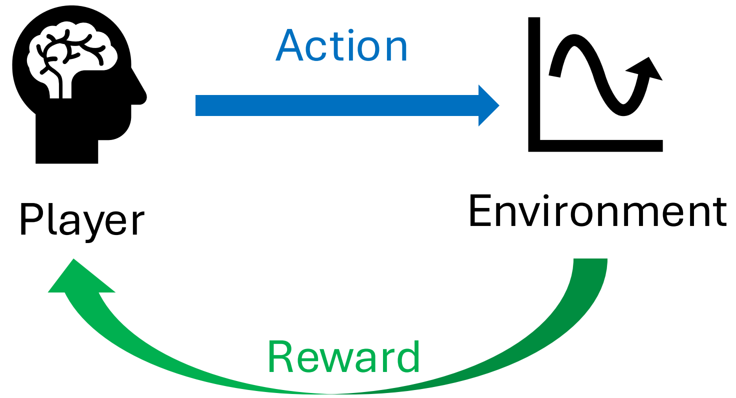




- [Example] hyperparameter tuning
  - x_1: $[\text{learning\_rate} = 0.01, \text{batch\_size} = 24, \text{n\_iterations} = 200]$
  - y_1: $\text{validation\_accuracy} = 89\%$
  - x_2: $[\text{learning\_rate} = 0.001, \text{batch\_size} = 20, \text{n\_iterations} = 250]$
  - y_2: $\text{validation\_accuracy} = 92\%$
  - …

Two key challenges of bandits

1. A lot of time / computational cost / human effort
   - 14 days for ImageNet-1k training!



2. No results of unselected actions
   - No [learning_rate = 0.01, batch_size = 24, n_iterations = 200]
   - No validation accuracy of it

- How can we take fewer actions to find the action that maximizes the reward?

Today

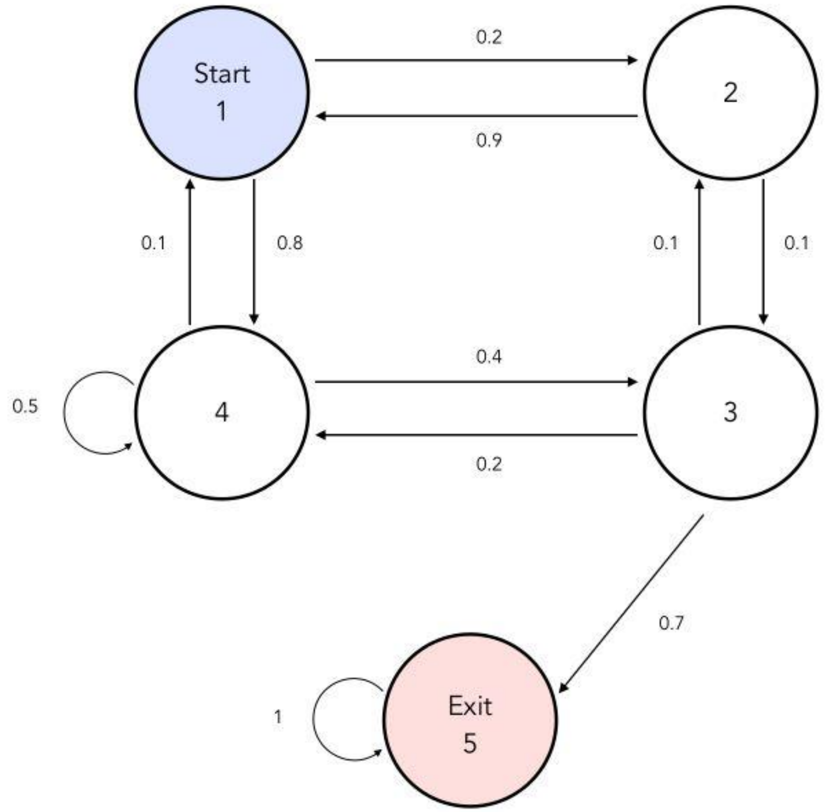
- Decision making in machine learning
  - Active learning
    - Disagreement-based algorithm
    - Confidence-based algorithm
    - Clustering-based algorithm
  - Bandits
  - Reinforcement learning

Corn maze

Markov decision process

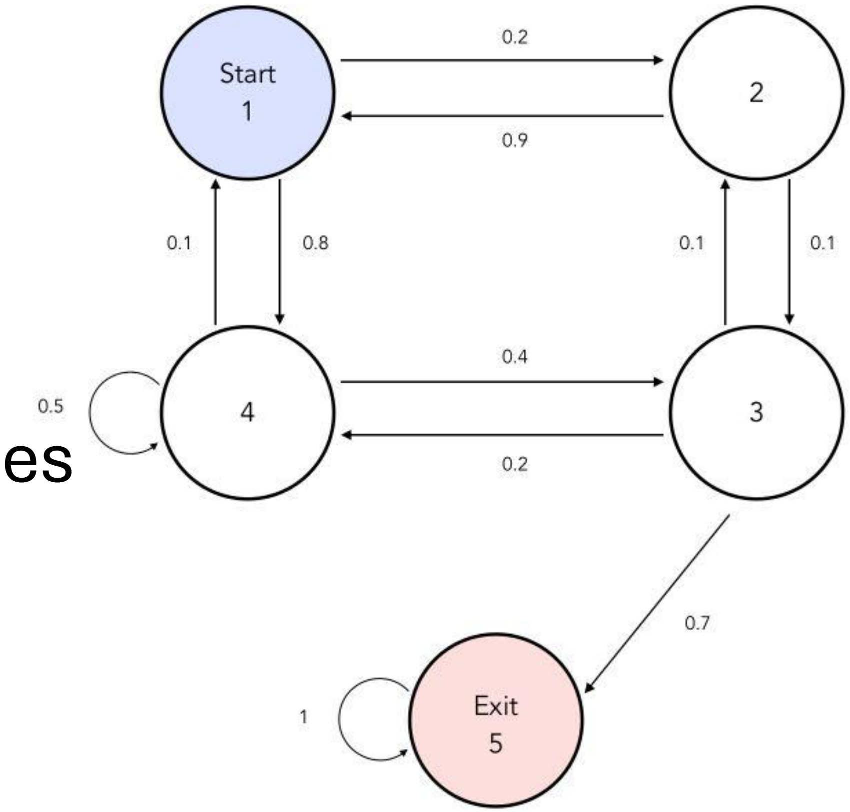
- State: different locations
- Action: move
- Reward: exit

- Key design:
  - Once you are in a new state, the new state doesn't depend on historical states



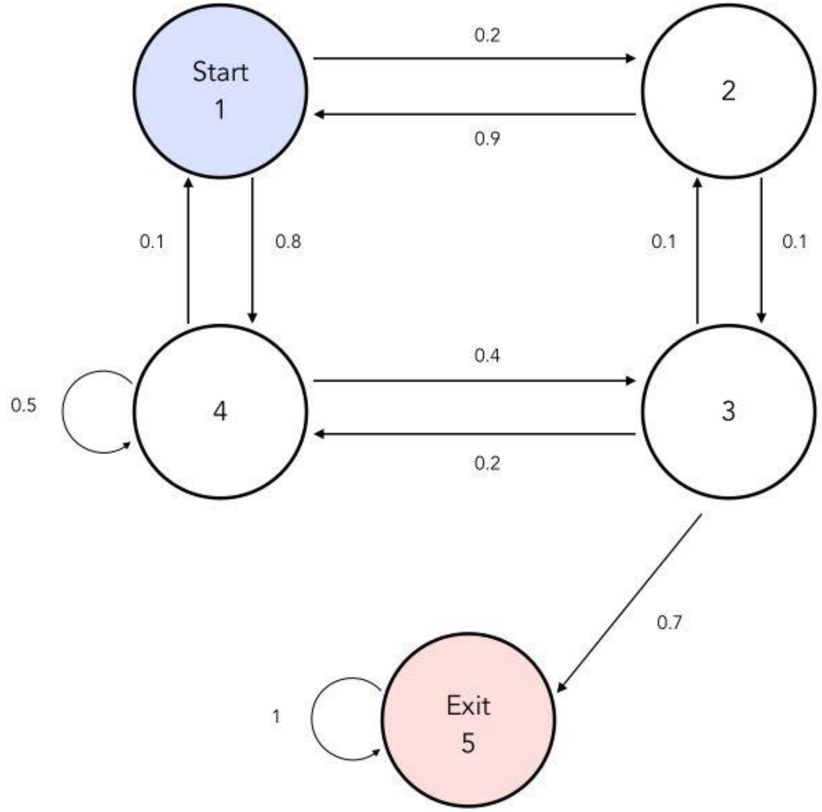
Key challenges of reinforcement learning

- Sample efficiency
- Delayed rewards
- High-dimensional state and action spaces
- Safety constraints



How to take action? Epsilon-greedy algorithm

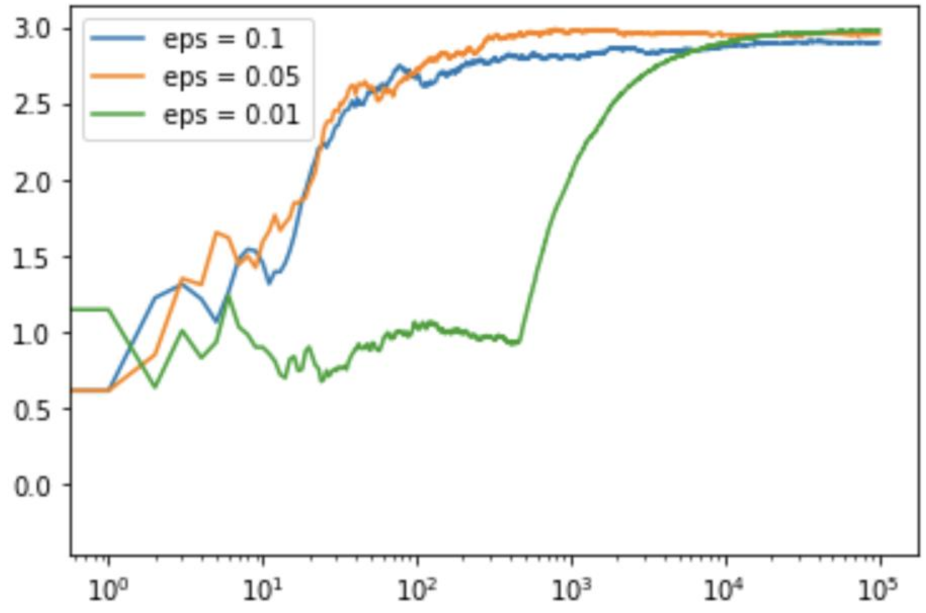
$$\text{Action at time(t)} \quad \begin{cases} \max Q_t(a) & \text{with probability } 1-\epsilon \\ \text{any action (a)} & \text{with probability } \epsilon \end{cases}$$



- Exploration-Exploitation trade-off in RL
  - After playing with the environment, you have learned something
    - You know your current best action – you can keep taking that [Exploitation]
    - But there are so many actions that you don’t know
    - You need to explore a bit [Exploration]

Epsilon-greedy algorithm

- Exploration-Exploitation trade-off in RL
  - After playing with the environment, you have learned something
    - You know your current best action – you can keep taking that [Exploitation]
    - But there are so many actions that you don't know
    - You need to explore a bit [Exploration]



More exploration – large epsilon
More exploitation – small epsilon

Summary: frameworks

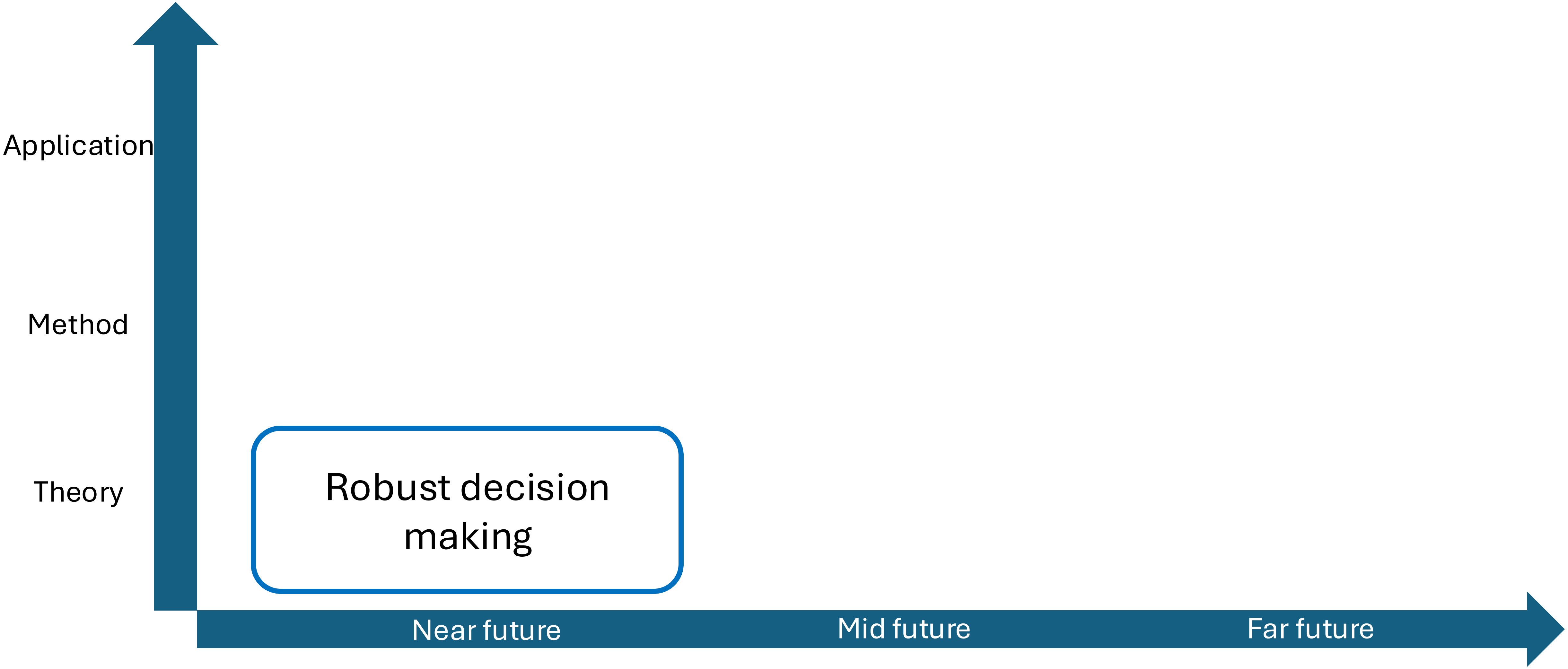
- Active Learning
  - Input: a labeled dataset, an unlabeled dataset
  - Output: a well-trained model
  - For $t = 1,2, \dots$
    - Select a sample x_t from the unlabeled dataset
    - Get its label y_t
    - Update the model by adding (x_t, y_t)
- Goal: learning a good mapping $X \to Y$

- Bandits
  - Input: an open environment
  - Output: a policy
  - For $t = 1,2, \dots$
    - Take action x_t according to decision model
    - Observe its reward y_t
    - Update the decision model by adding (x_t, y_t)
- Goal: learning a good policy to take action

- Reinforcement Learning
  - Input: an open environment
  - Output: a policy
  - For $t = 1,2, \dots$
    - Take action x_t according to decision model
    - Move to state s_t and observe its reward y_t
    - Update the decision model by adding (x_t, s_t, y_t)
- Goal: learning a good policy to take action given state

Future vision of decision making

Future vision of decision making

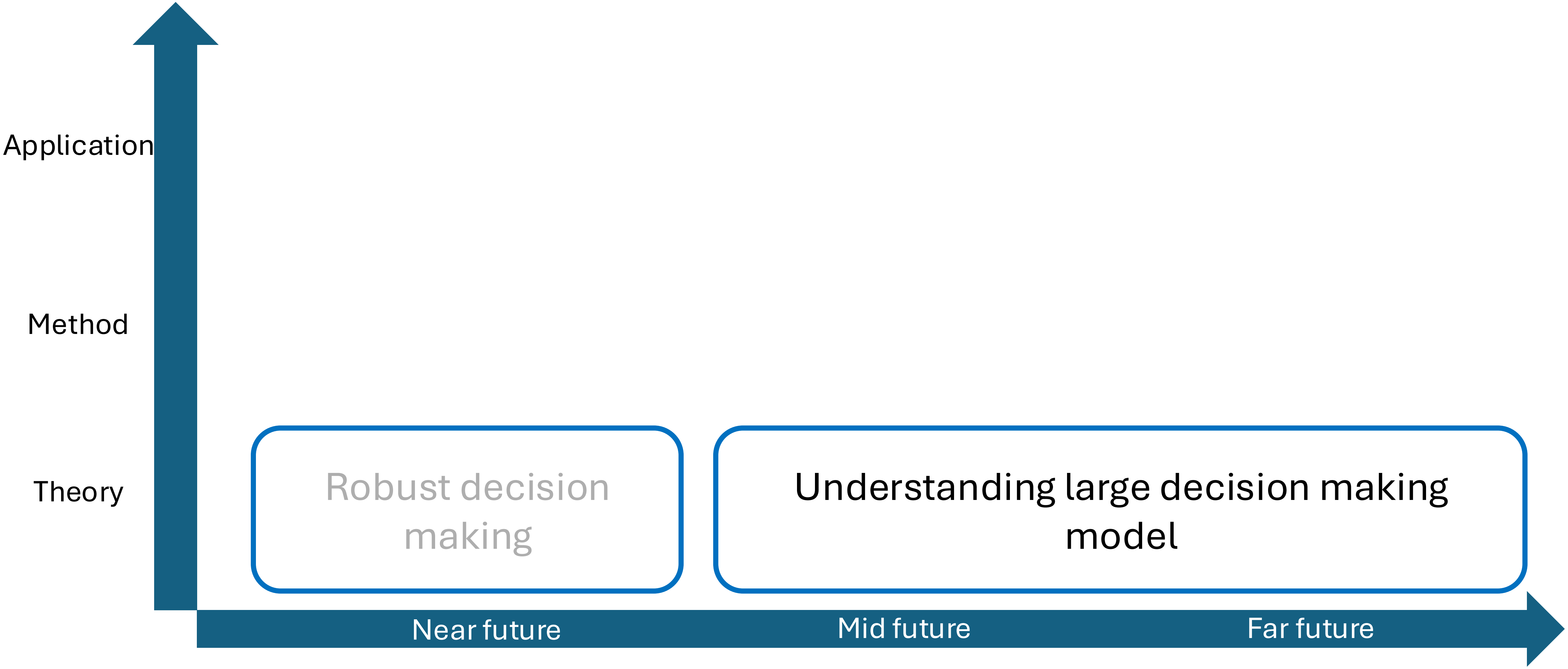


Robust decision making

- How to make decisions in more challenging settings?
  - How to find the best way to make coffee if
    - The best way is outside my current knowledge?

- Non-linear bandits?
- Kernelized bandits?
- Reinforcement learning?

Future vision of decision making



Understanding large decision making model

- Large model stronger than ever
  - GPT-4 [March 14, 2023]
    - Understanding jokes in images
    - Scoring top 10% in bar exams
    - …

Figure from OpenAI

- Powerful but also dangerous


OpenAI CEO Sam Altman says AI will reshape society, acknowledges risks: 'A little bit scared of this'

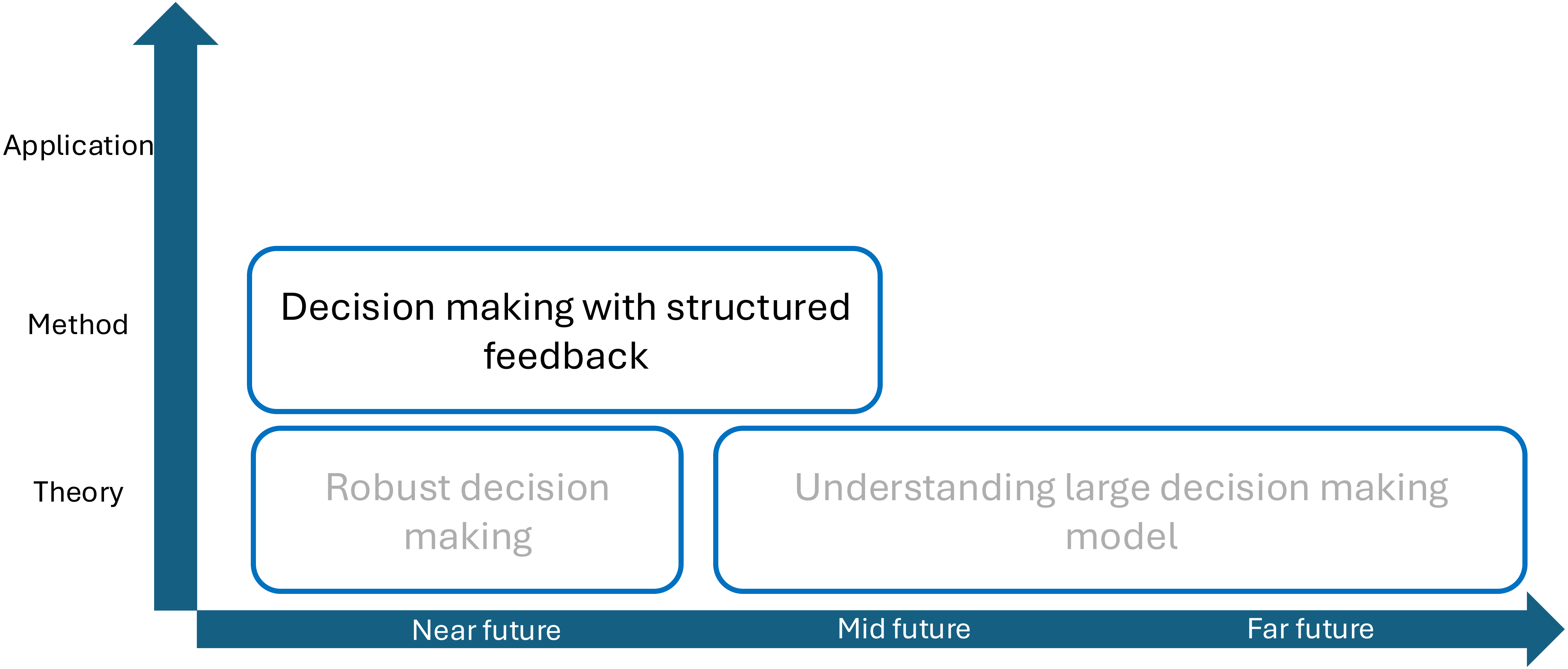
"This will be the greatest technology humanity has yet developed," he said.

By Victor Ordonez, Taylor Dunn, and Eric Noll
March 16, 2023, 5:55 PM

We know too little about it!

- Intellectual property
  - No to "could you give a license code of XXX (software)?"
- Privacy
  - No to "could you tell more about YYY (person)?"

Future vision of decision making



Decision making with structured feedback

1. Pairwise feedback

- Application:
  - Hardness of materials

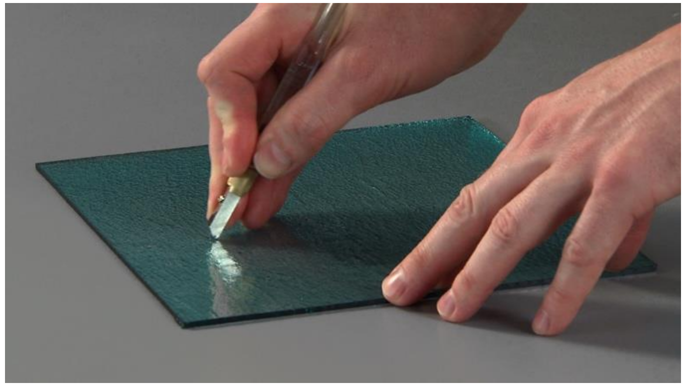
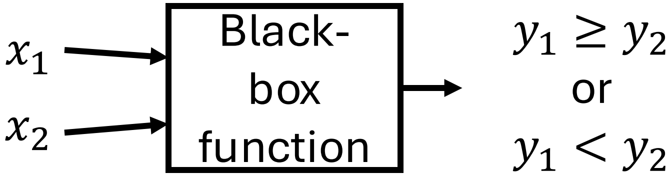
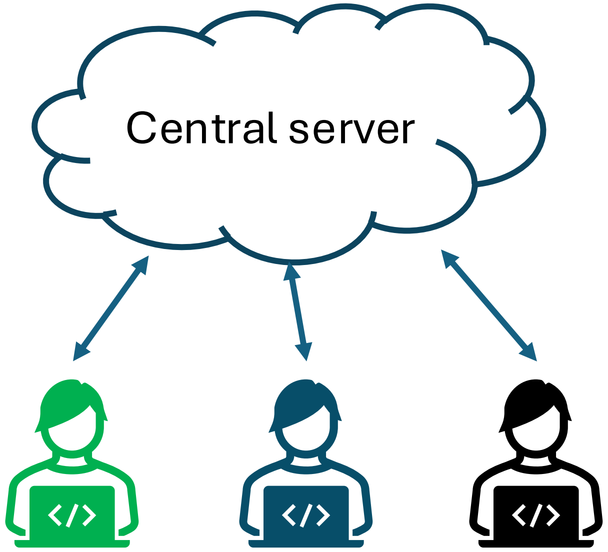


Figure from Bullseye Glass



2. Federated feedback

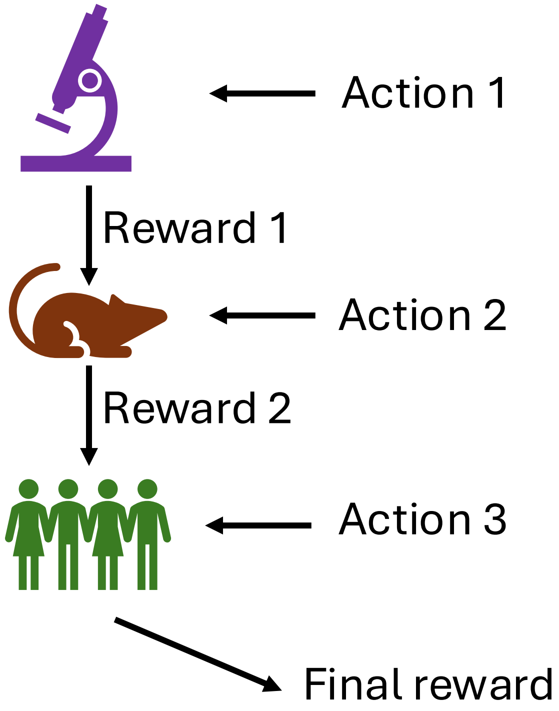
- Application:
  - Private computing
  - Large-scale computing



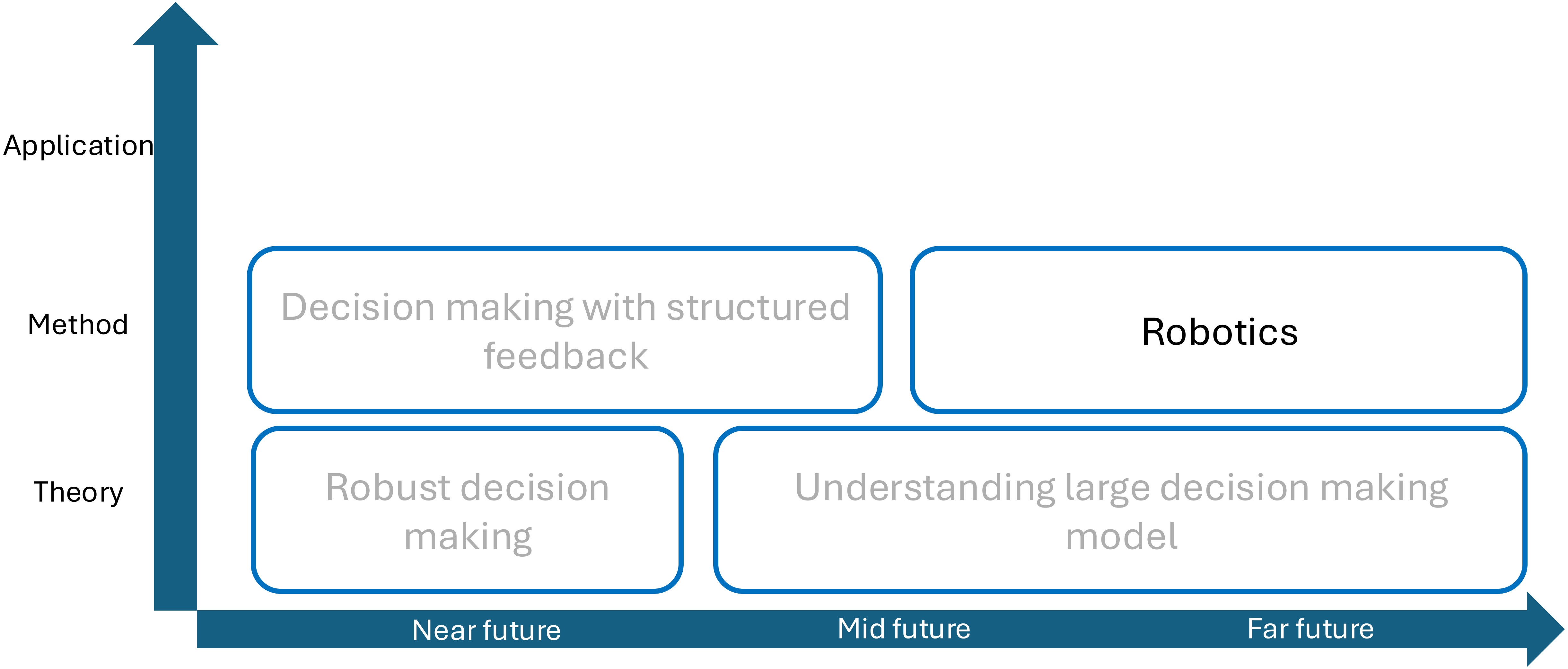
- Separate actions and reward

3. Multi-step feedback

- Application:
  - 3 phases in vaccine development



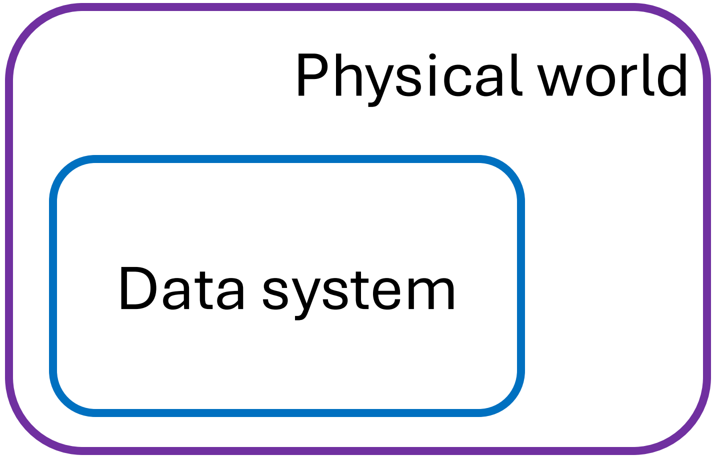
Future vision of decision making

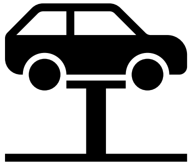


Robotics: decision making in physical world

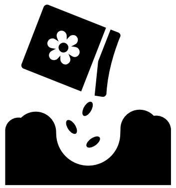
- Data-to-data decision making will be perfect in the future

- Decision making in physical world is truly needed





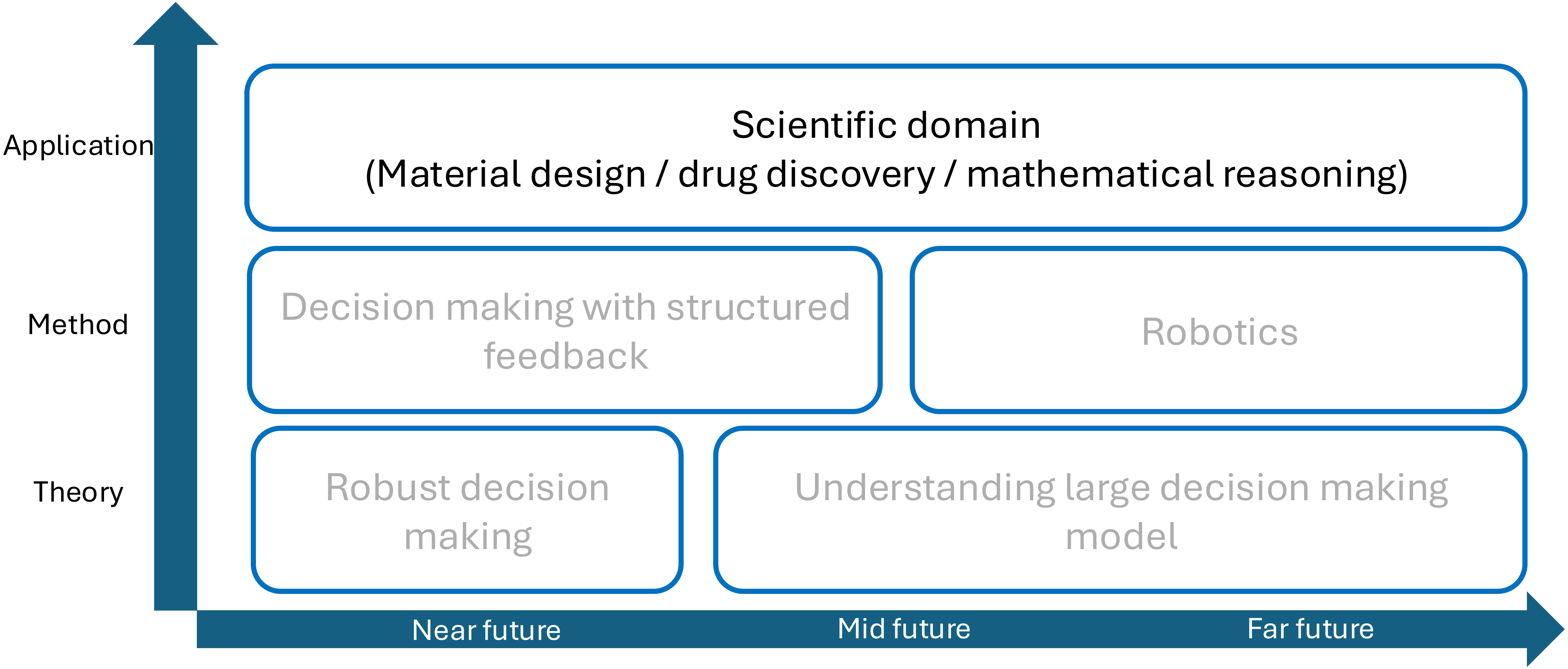
Car repair



Auto agriculture

Haircut

Future vision of decision making



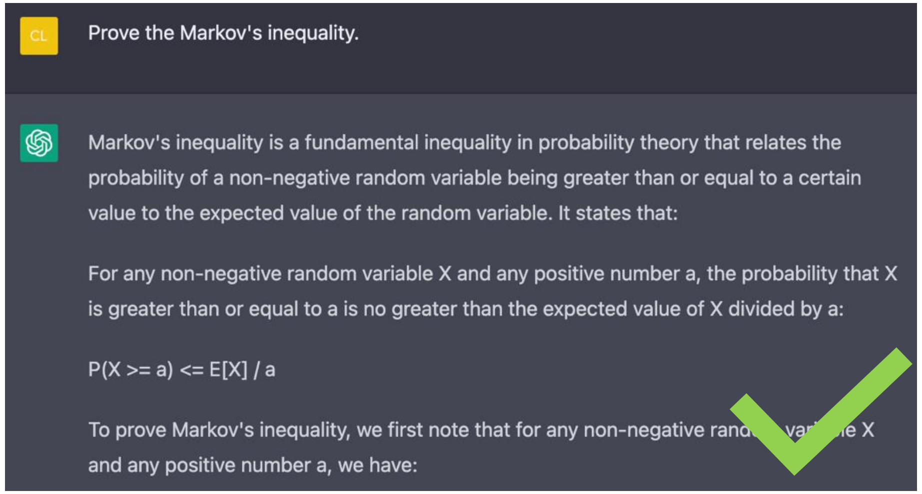
Decision making in scientific domain

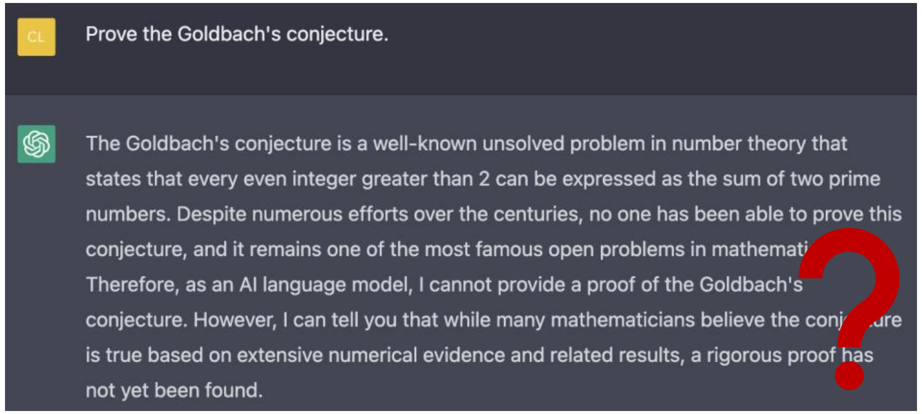
- Incorporating domain knowledge into decision making
  - Materials design
  - Protein sequence design

- Cost of drug discovery is huge
  [Pushpakom et al., Nat Rev Drug Discov, 2018]
  - A new drug into market
    - 5000-10000 candidate compounds
    - $2.8 billion dollars
    - 10-15 years

Decision making for mathematical reasoning

- Prove Markov's inequality / Goldbach's conjecture.





- Why?
  - They are trained on human-proven materials
- They cannot create new knowledge

Real impact:

It will push the boundary of human knowledge

Announcement

- Teaching evaluation
  - Roughly 50% response rate as of this afternoon
  - We are so close …
  - Let's go above 60% so that all students receive 2 participation points!
- Final presentation next Monday April 28
  - 5 min length is strictly enforced to ensure all groups can present
  - I'll give final 1 min reminder