DYNAPLEX

REINFORCEMENT LEARNING FOR DATA-DRIVEN LOGISTICS

Martijn Mes
University of Twente

UNIVERSITY OF TWENTE.

May 12, 2023

DynaPlex

TKI DINALOG
Dutch Institute for Advanced Logistics

Beta
Research School for Operations Management and Logistics

Symposium: AI for Operations Management

Project DynaPlex: RL for Data-Driven Logistics

3 2 1


Willem Jaarsveld


Martijn Mes


Remco Dijkman


Yingqian Zhang


Maria Iacob

1. DATA-DRIVEN LOGISTICS

WASTE COLLECTION

Dynamic waste collection from underground containers using sensor data, deciding on which containers to visit and in what sequence, anticipating uncertain waste deposits

INVENTORY MANAGEMENT

When, where and what to order, to minimizing holding and backorder costs, anticipating uncertainty in supply and demand

LAST MILE LOGISTICS

Routing and scheduling of delivery vehicles, anticipating uncertain events (new customers, cancellations, traffic disruptions)

SYNCHROMODAL TRANSPORT

Scheduling of container transport considering various transport modes that can be selected dynamically based on actual circumstances, anticipating potential disruptions and future consolidation opportunities

TAXI PROBLEM

Taxi repositioning and customer acceptance, anticipating future customers (related examples are positioning of ambulances, positioning of police helicopters, repositioning of bike sharing systems)

COMMON CHALLENGE

- Sequential decision-making… (e.g., upon arrival of an order, arrival of a customer, arrival at a location, large disruption within the network, every hour, when inventory drops below some threshold)
- under uncertainty… (unknown travel times, unknown waiting times at container terminals, unknown orders, unknown lead times)
- where decisions have an uncertain but long-term impact (e.g., one delay at a container terminal might delay arrival times of barges at other container terminals; not transporting containers now might result in capacity problems in the future; accepting this customer might result in long waiting times afterwards)

→ Reinforcement Learning supports these types of decisions…

2. REINFORCEMENT LEARNING

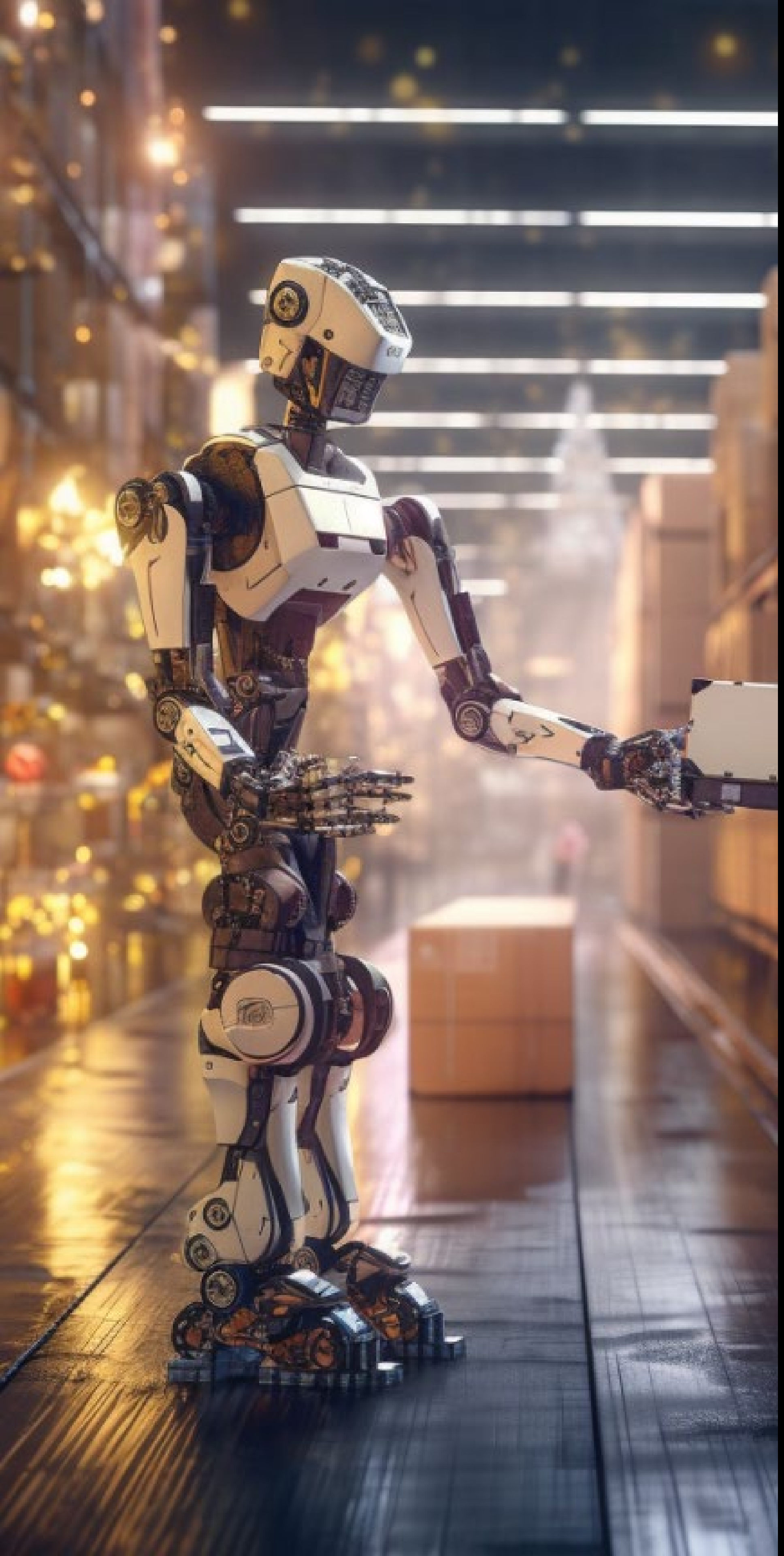
ARTIFICIAL INTELLIGENCE

- Computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making…
- “*Today’s artificial intelligence… is mainly a technology of pattern recognition, poring through vast troves of words, images and numbers*” [Robert J. Gordon, New York Times]
- Pattern recognition is one form of AI known as (un)supervised machine learning, which involves using observations to help train mathematical functions…
- Reinforcement Learning works differently… it involves an “agent” interacting with its surrounding environment to gradually learn the best action to take…

AlphaGo

- 2016
- Ancient game of Go with consensus on good strategies
- Alpha/Mu/Efficient Zero

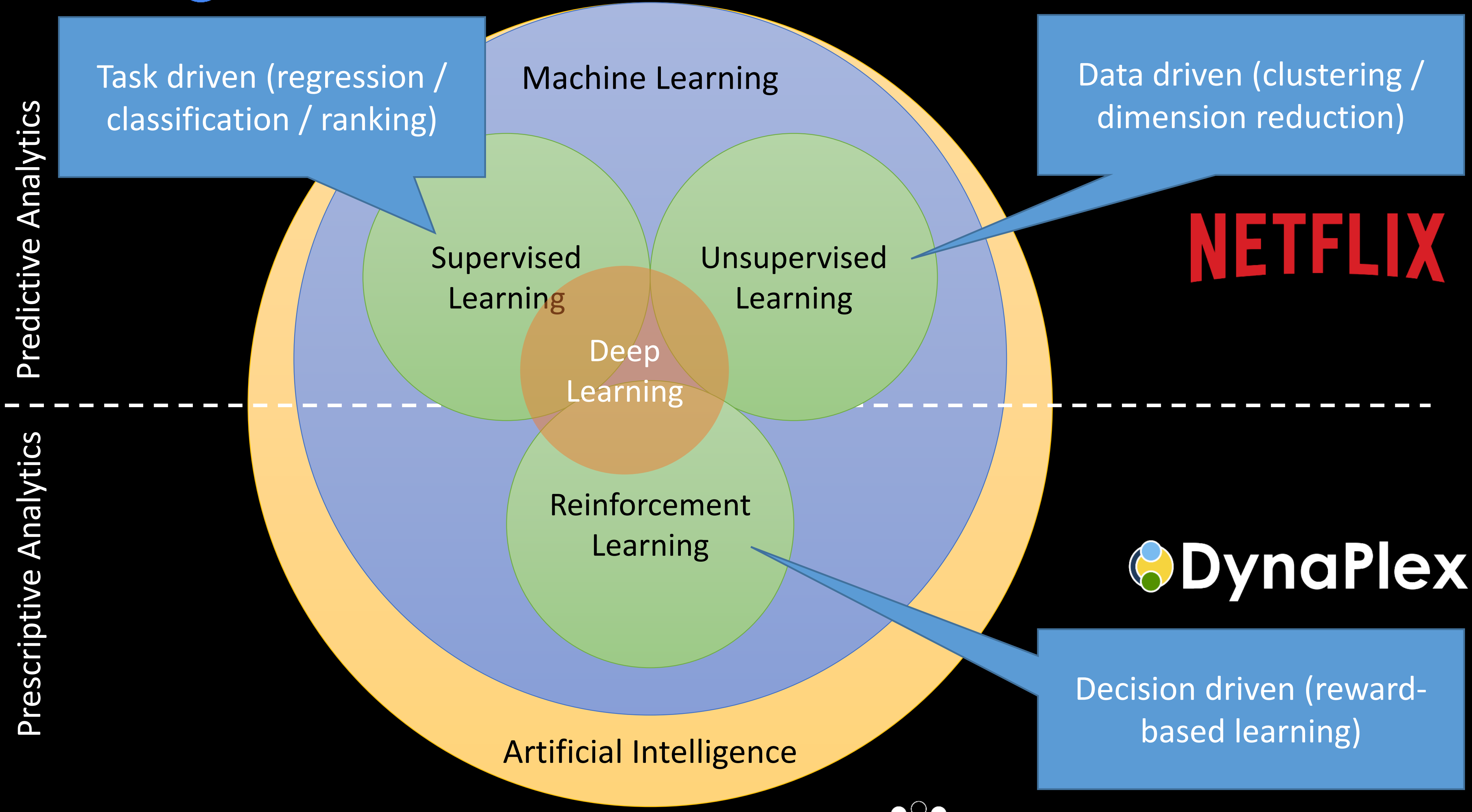
POSITIONING RL [1/2]



Google

ChatGPT

amazon



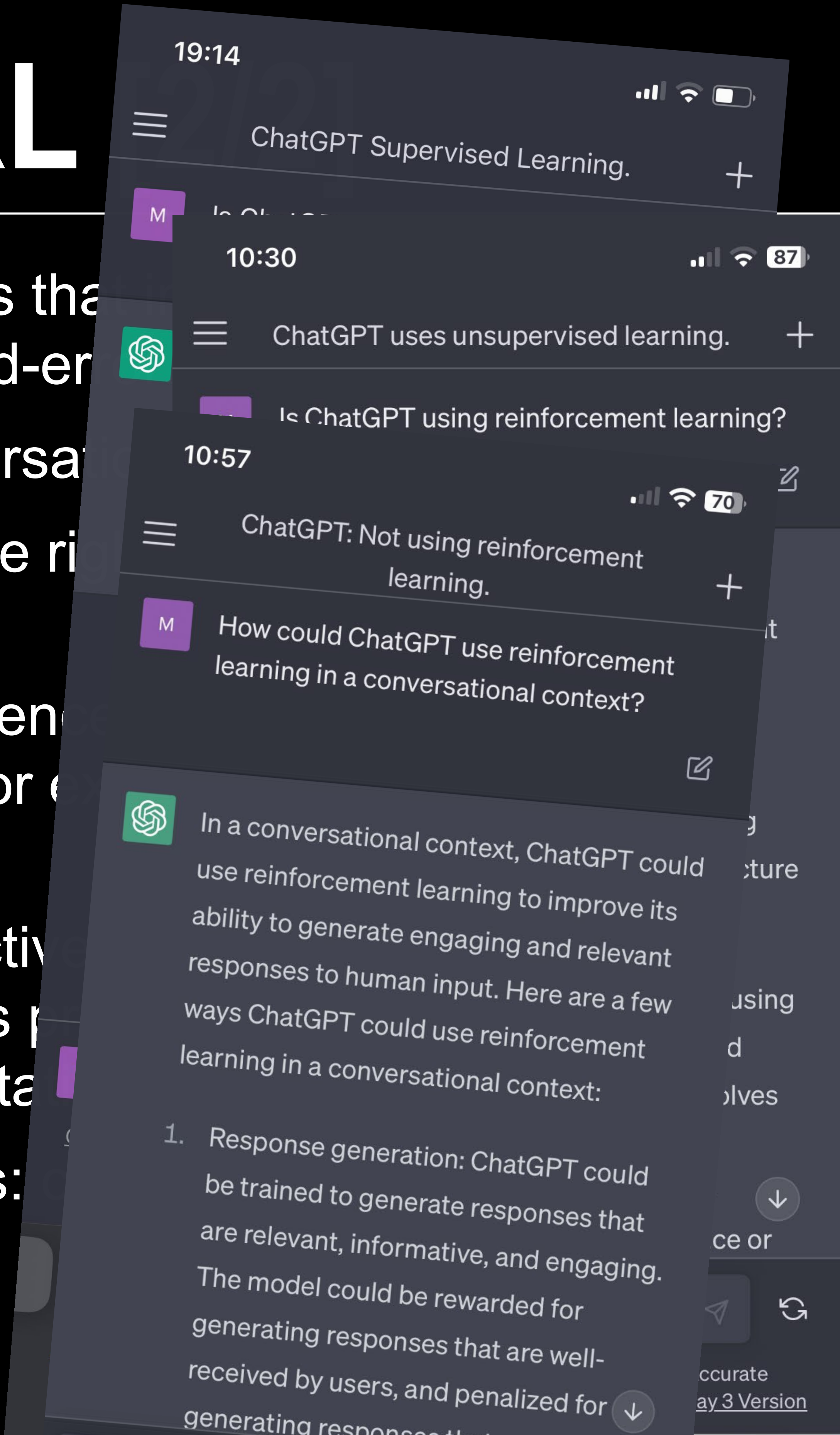
NETFLIX

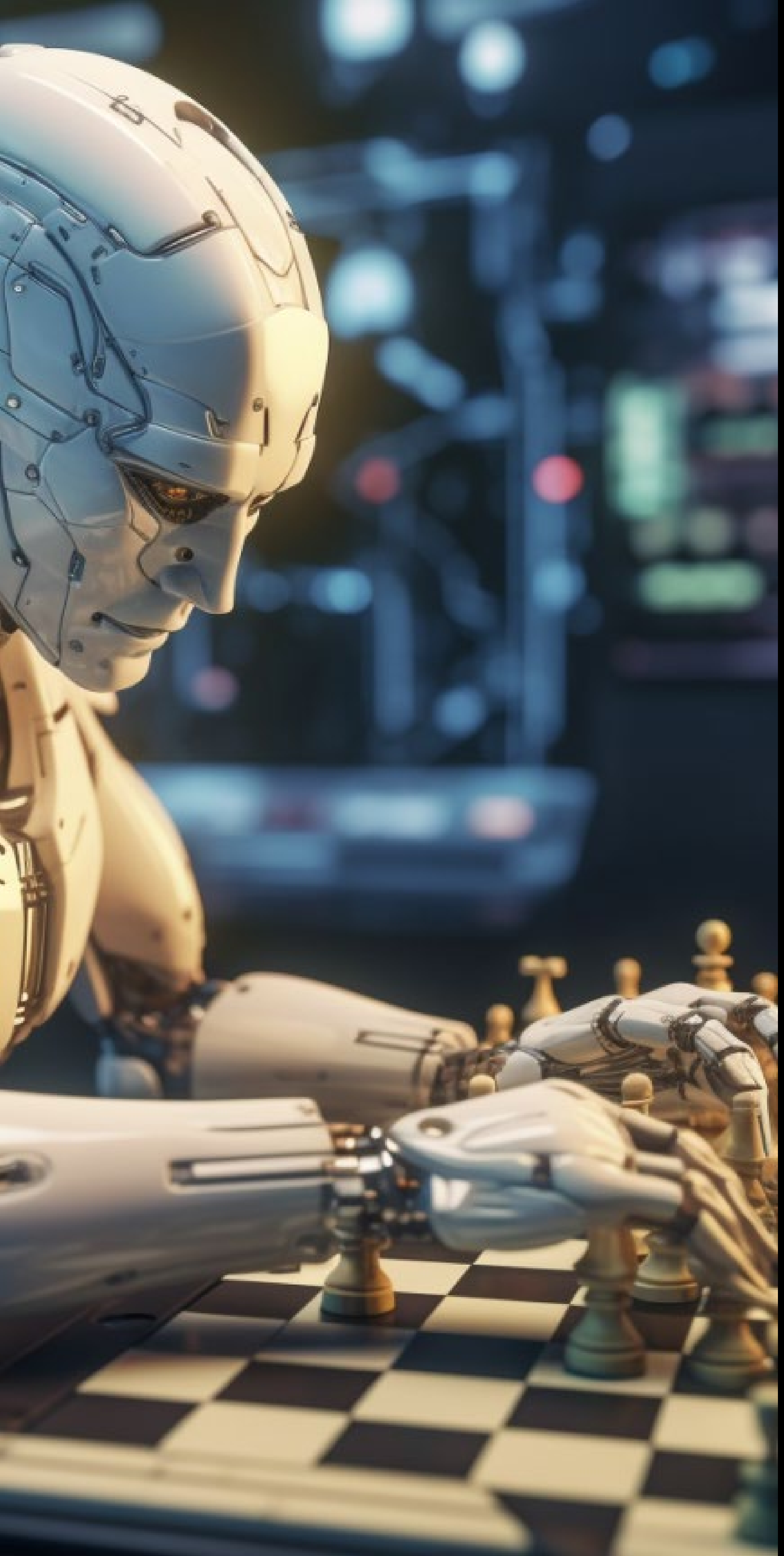
DynaPlex

AlphaGo

POSITIONING RL

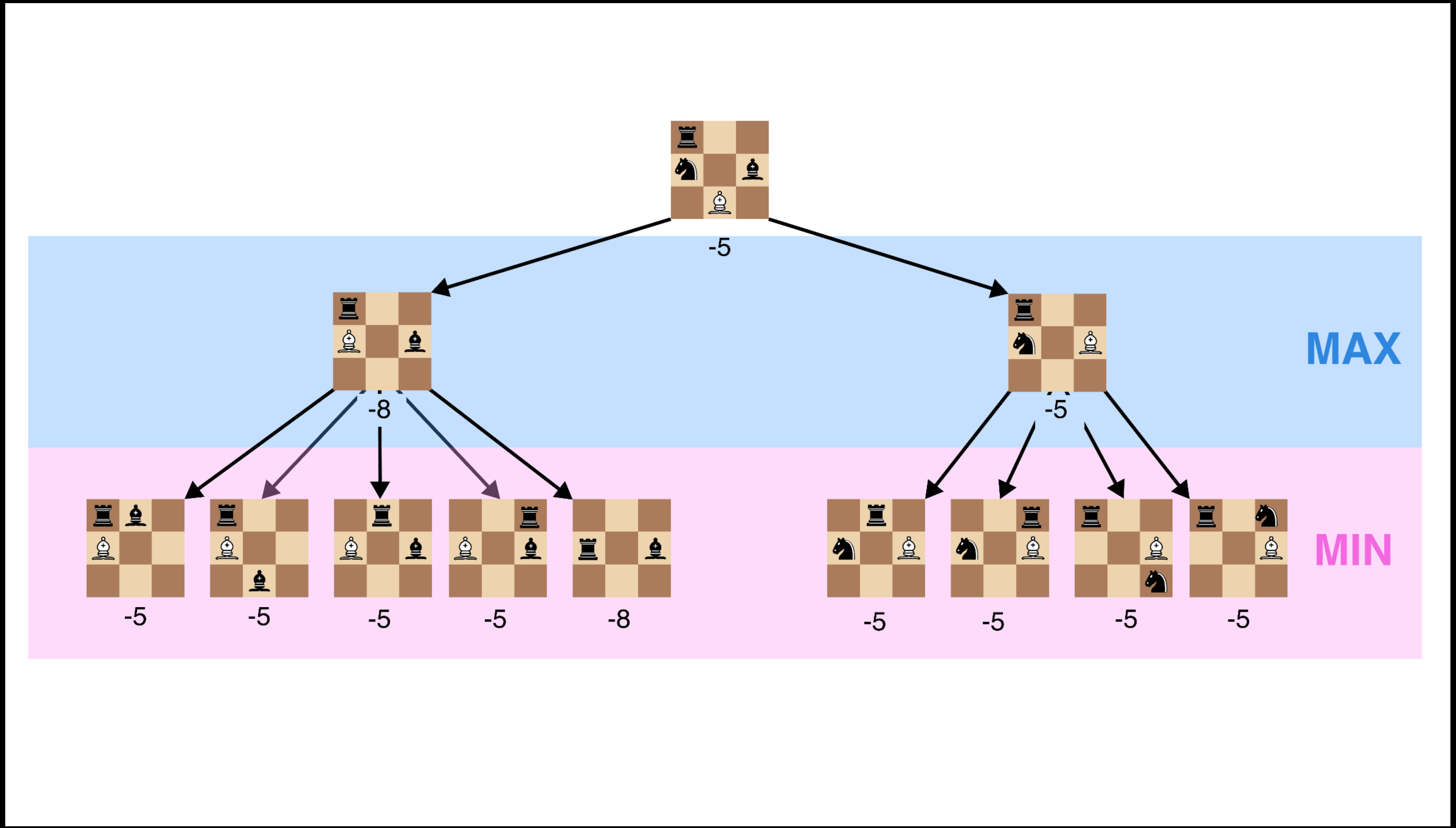
- RL adds the dimension of actions tha
environment: learning by trial-and-er
- Consider a dialog system (conversa
  - Classical ML aims to learn the ri
  query.
  - RL focuses on the right sequen
  lead to a positive outcome, for
  customer.
- This makes RL particularly attracti
planning and adaptation, such as
inventory management, transporta
- Illustration of RL with 3 examples:
synchromodal transport

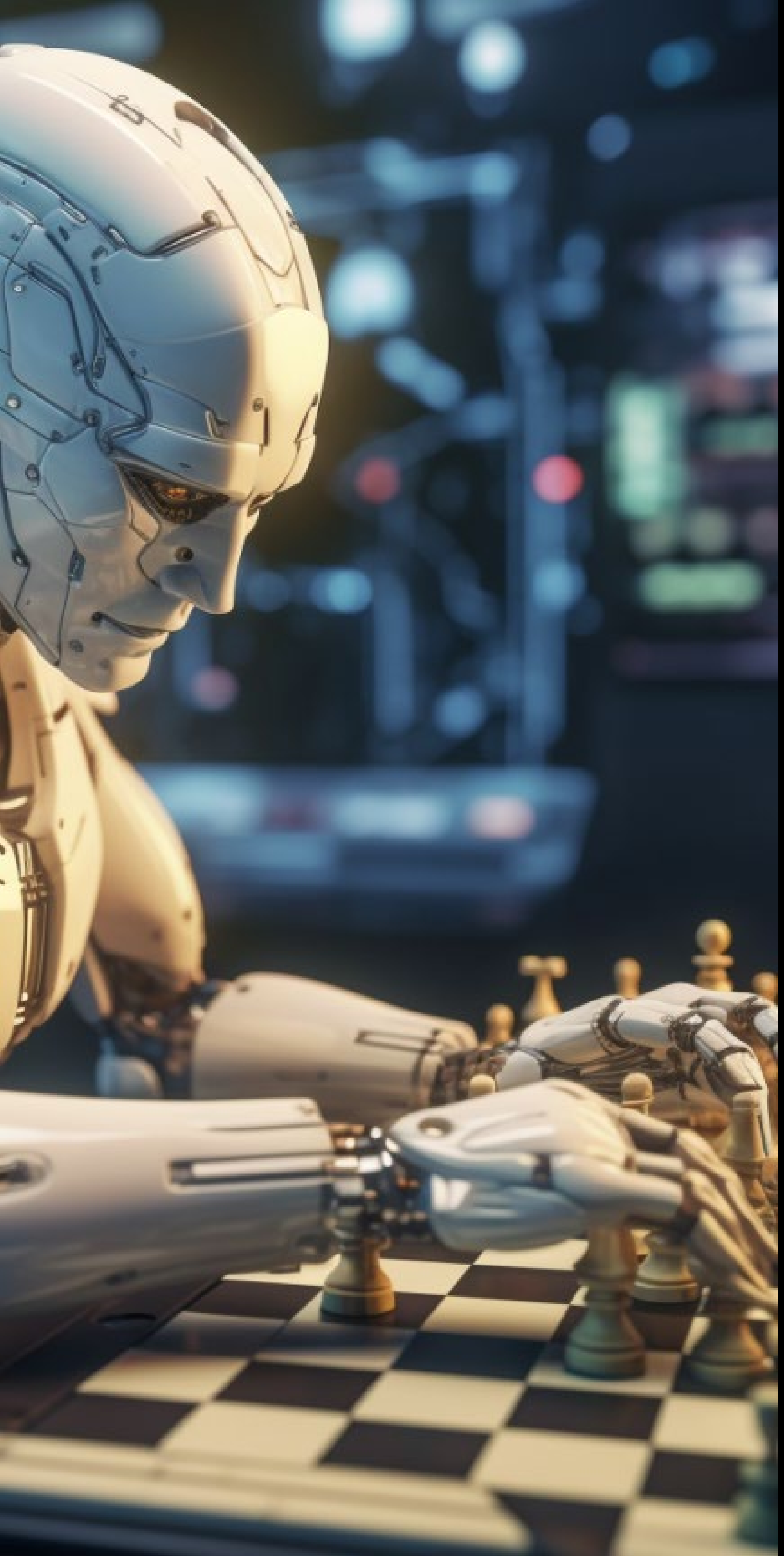




EXAMPLE 1: CHESS

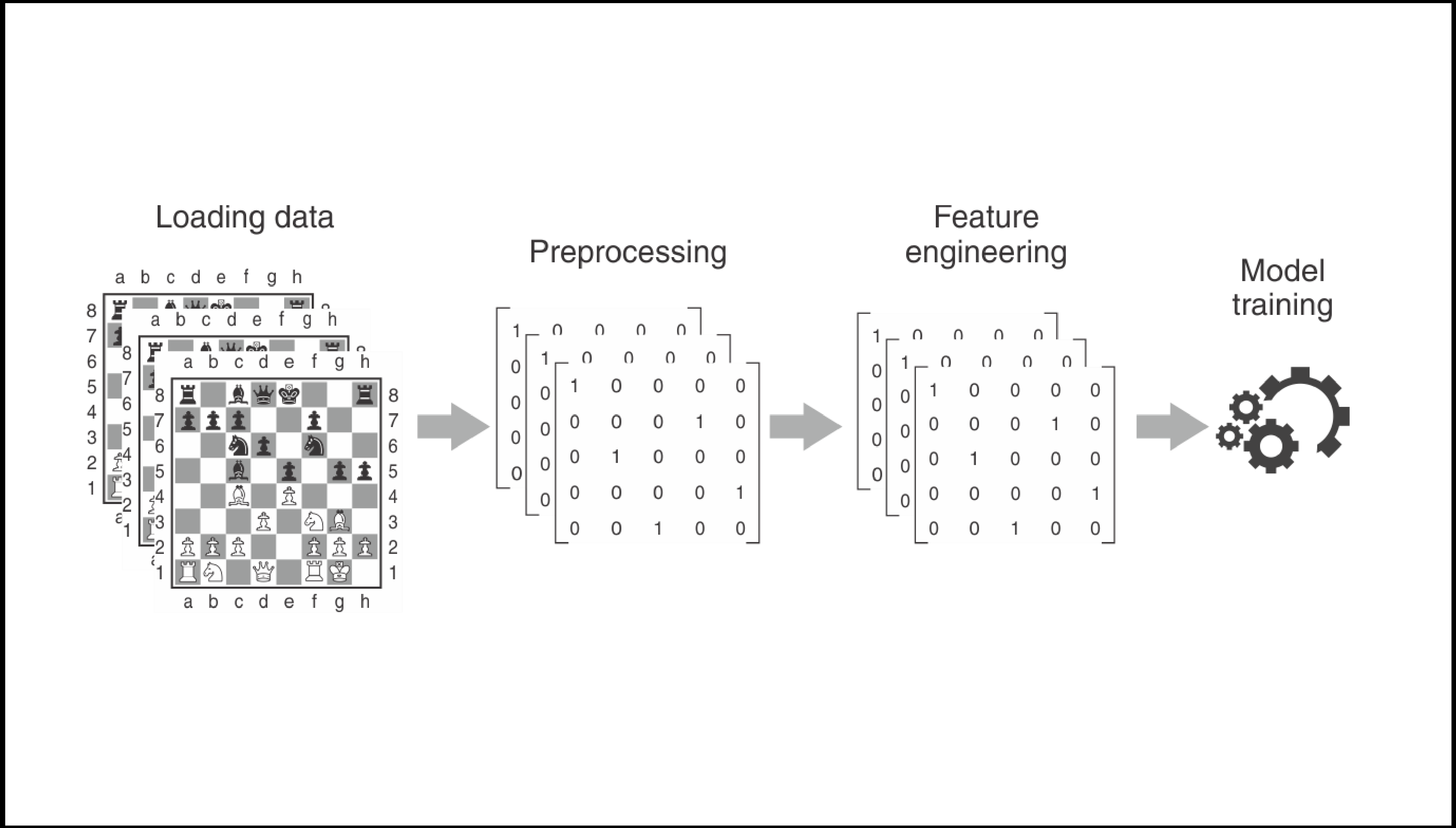
- Option 1: exact/heuristic approach (brute force, Deep Blue)

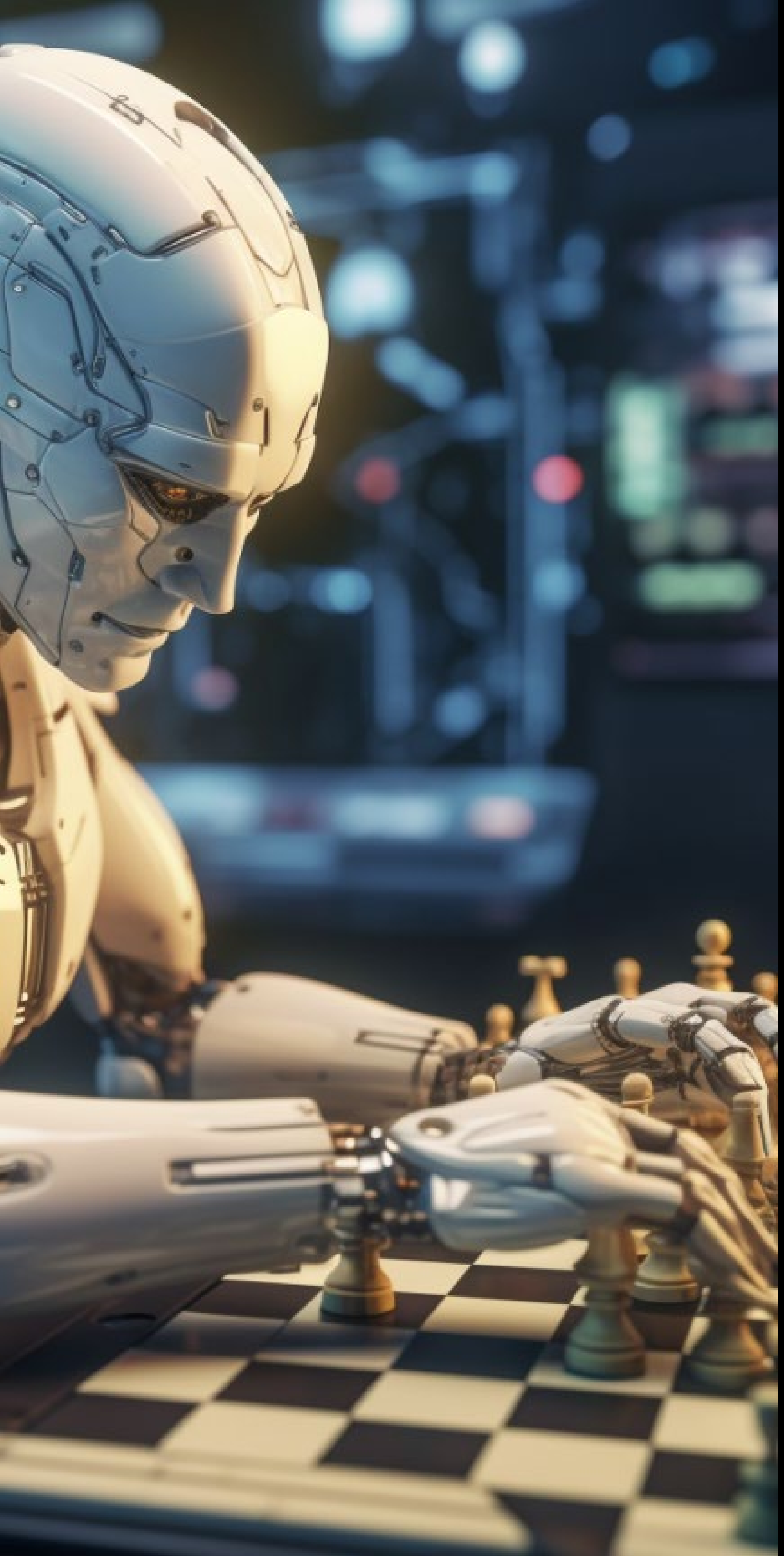




EXAMPLE 1: CHESS

- Option 2: supervised learning

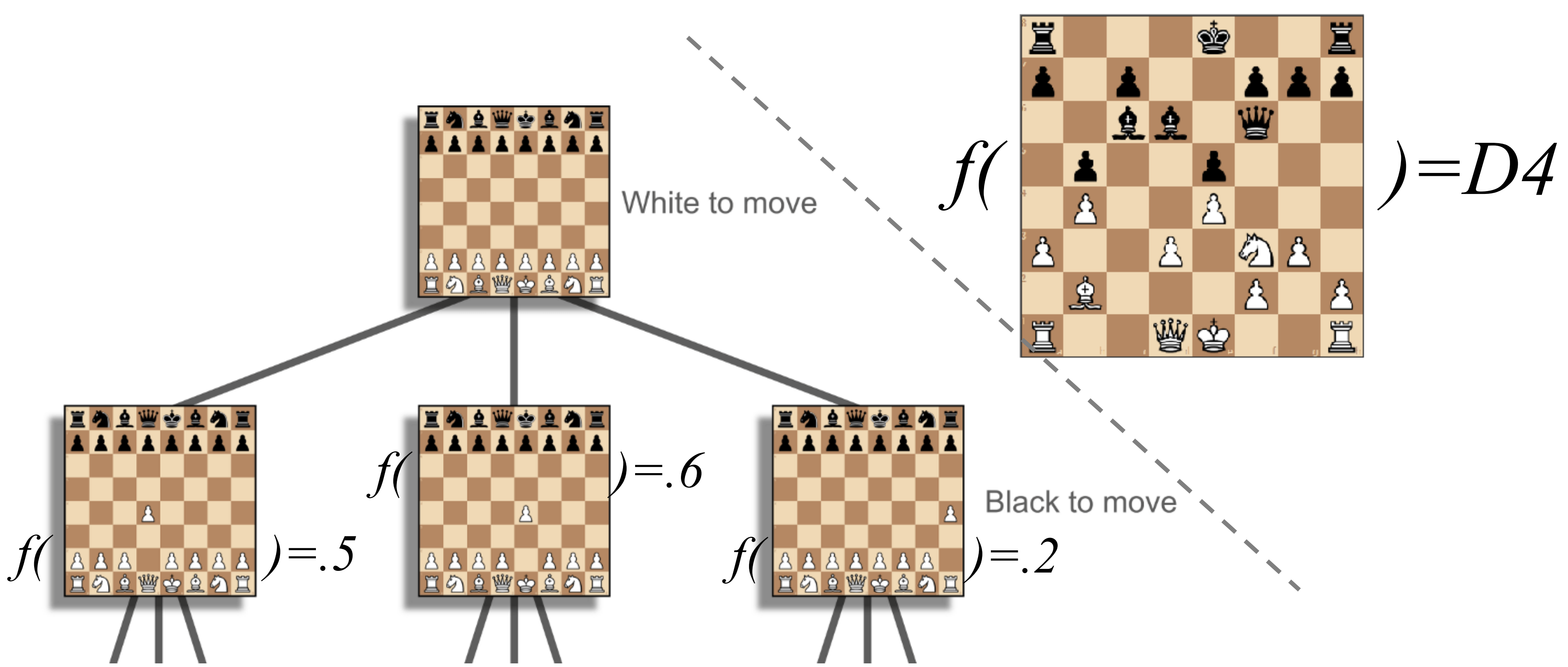




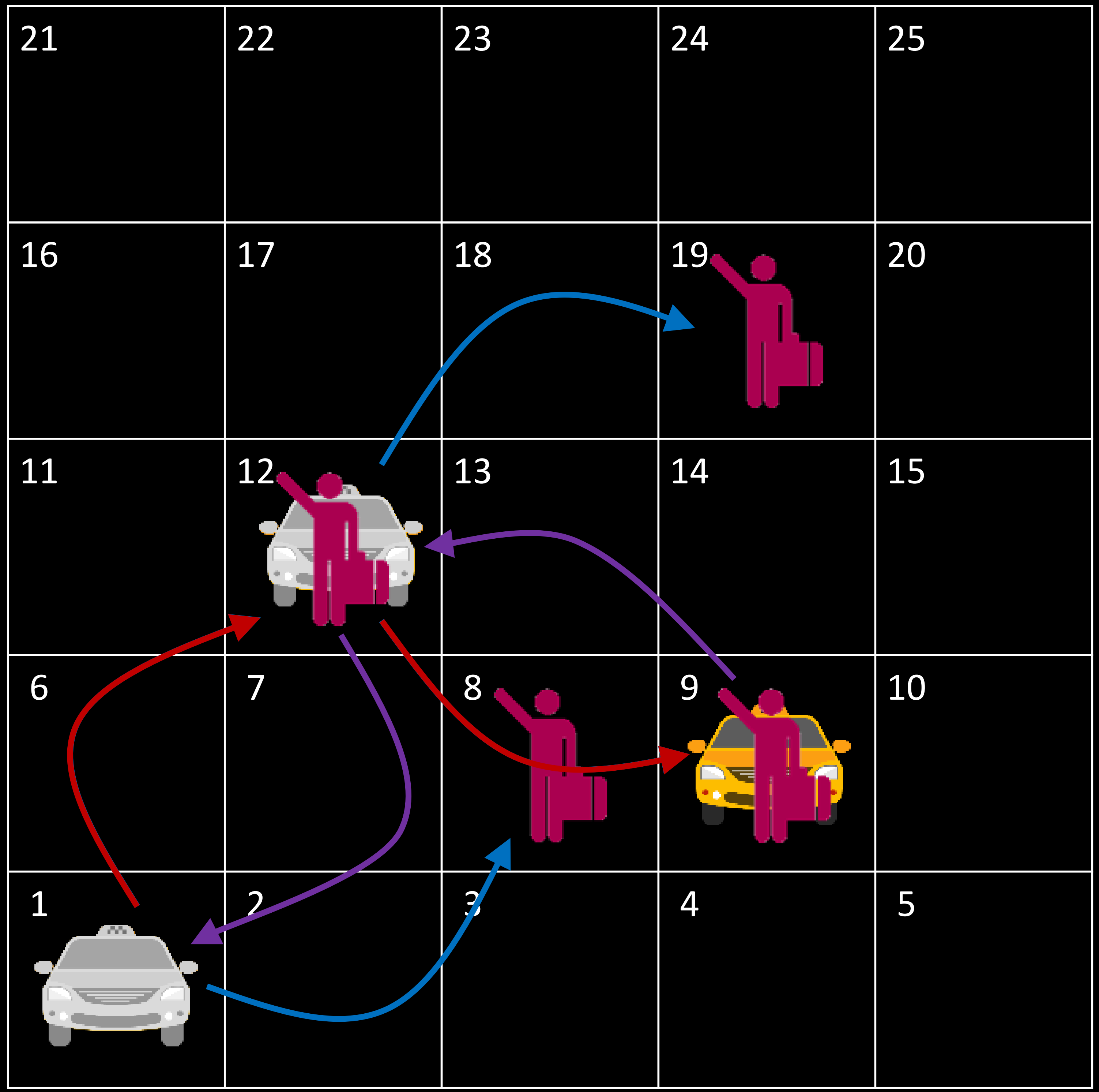
EXAMPLE 1: CHESS

- Option 3: reinforcement learning

  - Learning by playing many games, e.g., against humans or against itself in a simulator
  - Learn the value of “states” or the best action to take in each “state”:



EXAMPLE 2: TAXI PROBLEM



- Approximate Value iteration
- 1-step lookahead & update
- ADP, SARSA, TD(0)
- Post-decision state

$$\tilde{a} = \arg \max_{a \in \{8,12\}} \left\{ \begin{matrix} R(1,a) \\ +\bar{V}^n(a) \end{matrix} \right\}$$

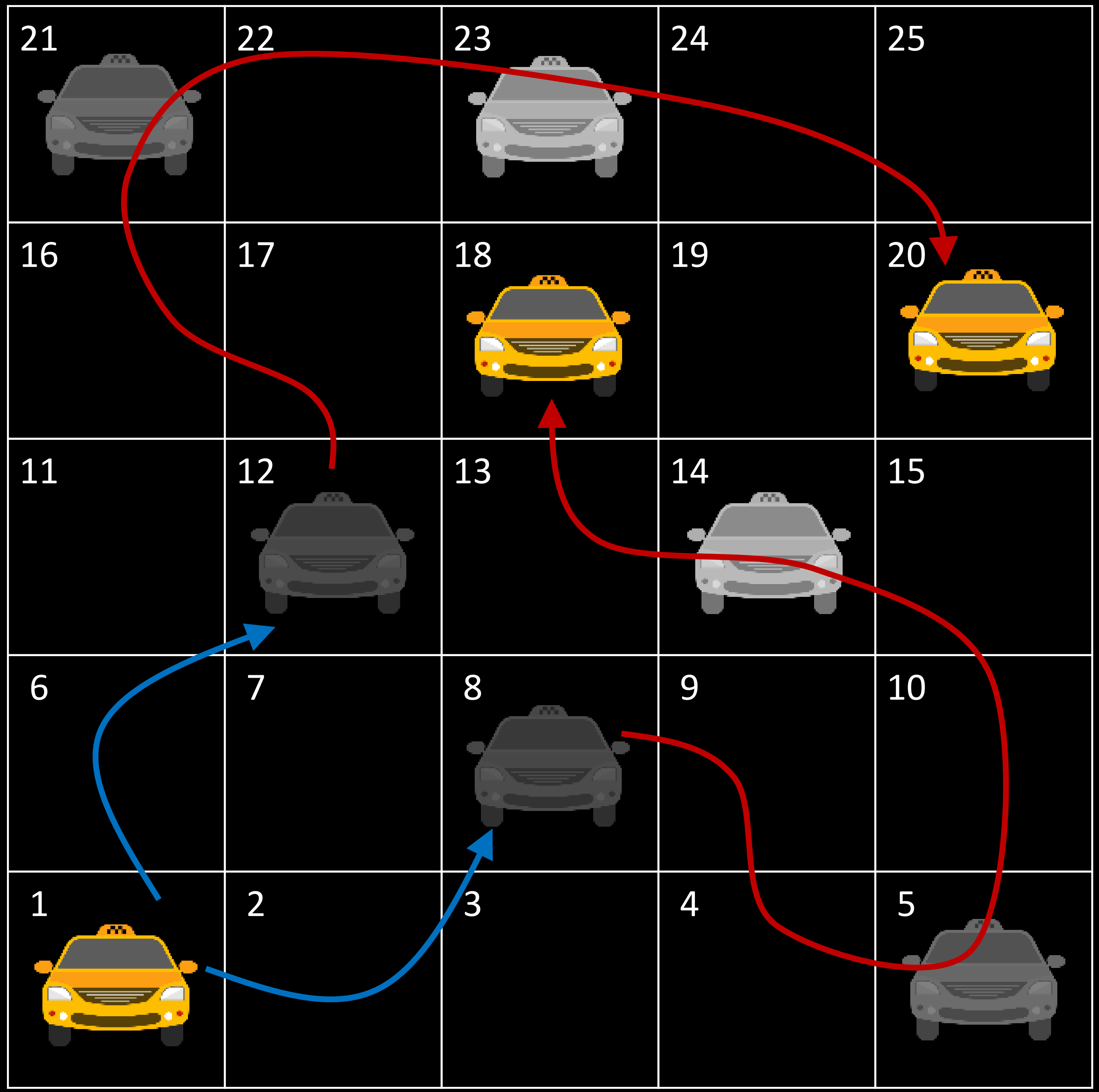
$$\bar{V}^{n+1}(1) \leftarrow R(1,12) + \bar{V}^n(12)$$

- Alternatively (Q-learning)

$$\tilde{a} = \arg \max_{a \in \{8,12\}} \{\bar{Q}^n((1,8,12),a)\}$$

$$\bar{Q}^n((1,8,12),12)) \leftarrow R(1,12) + \arg \max_{a \in \{?,?\}} \{\bar{Q}^n((12,?,?),a)\}$$

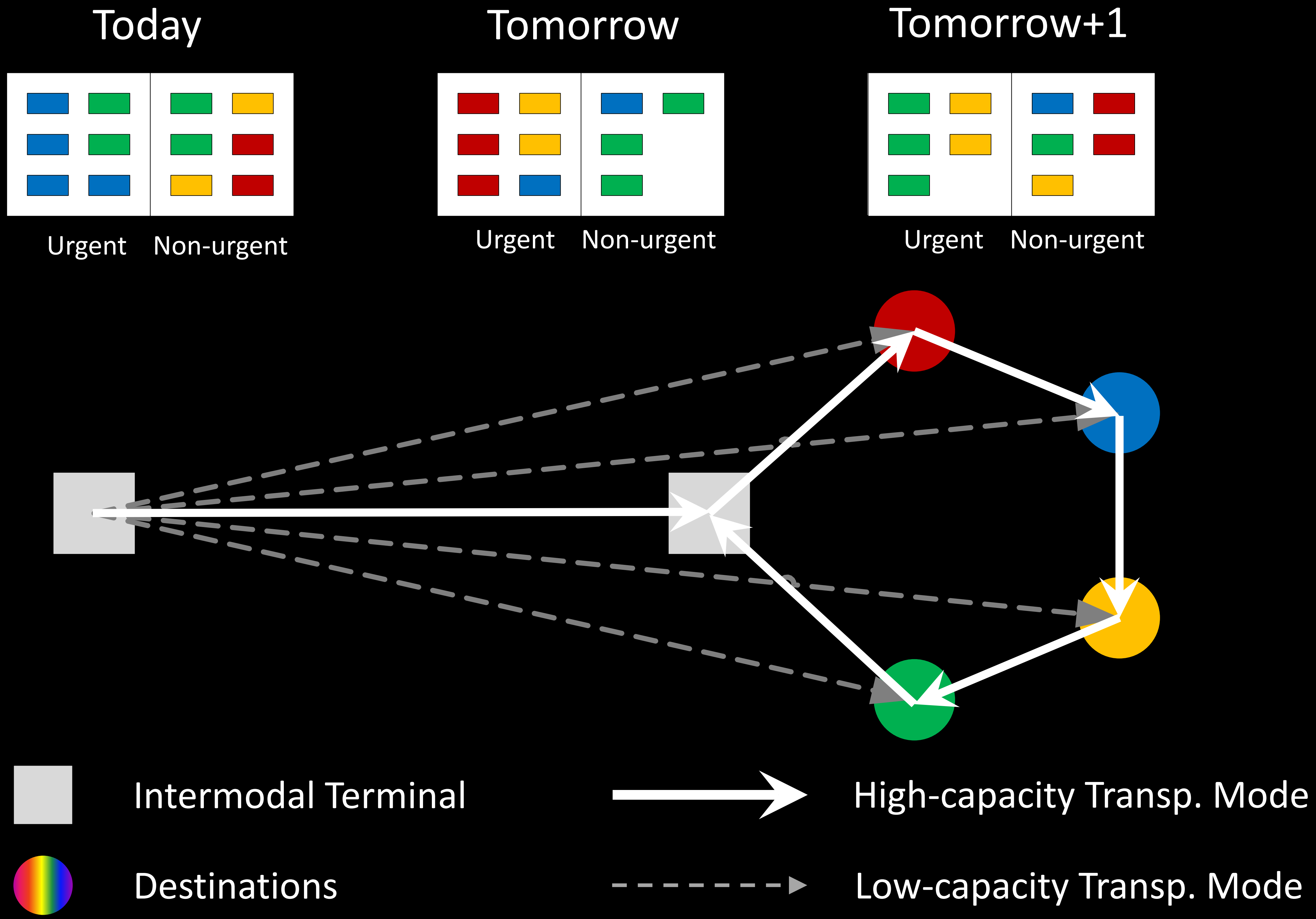
EXAMPLE 2: TAXI PROBLEM



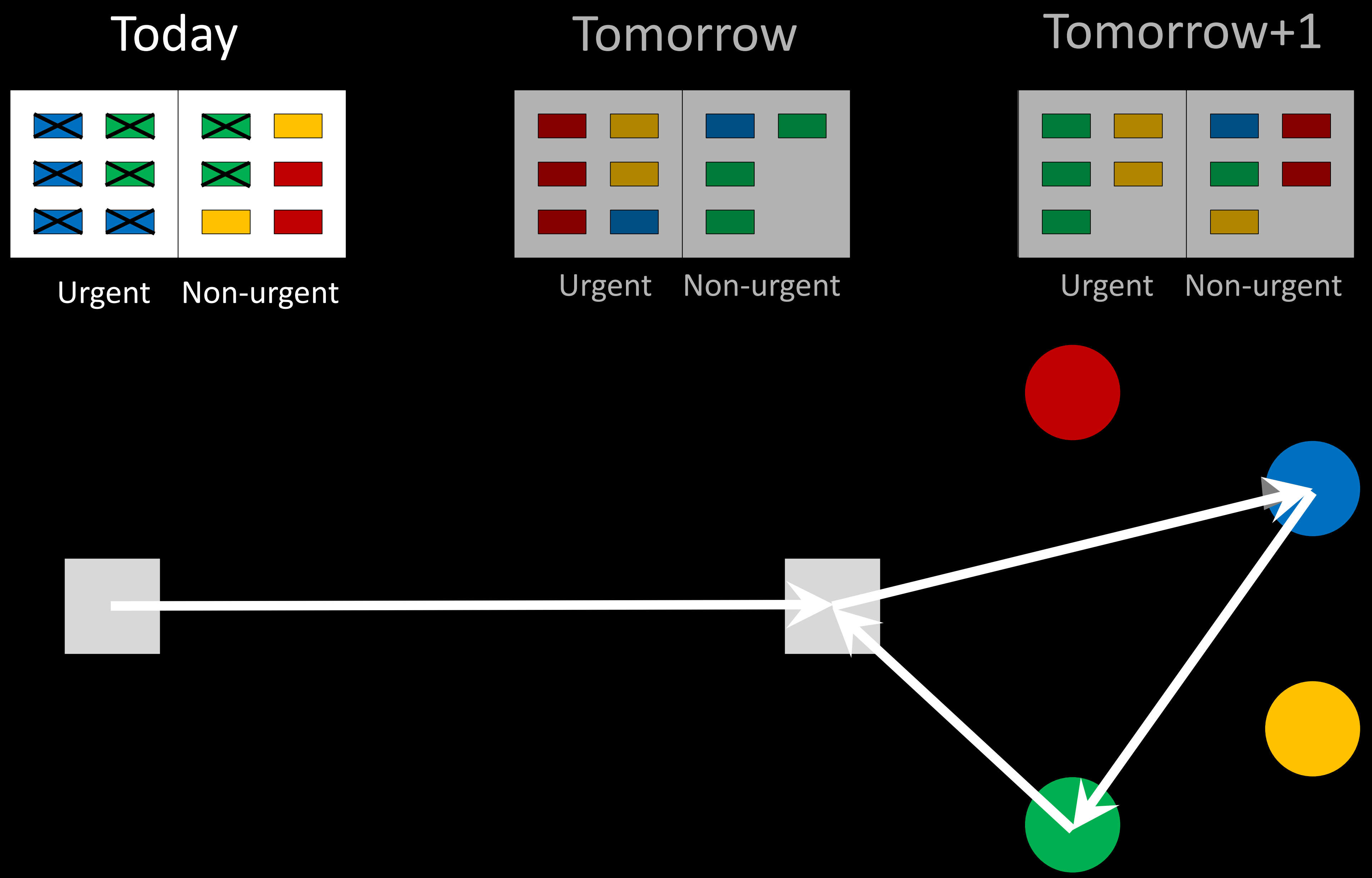
- Approximate Policy iteration
- n-step lookahead (rollout) with batch updating

1. Sample a state
2. Evaluate all possible actions
3. After the action, run multiple long simulations following the current policy
4. Evaluate best action
5. Repeat the above for many states
6. Update NN mapping states to actions

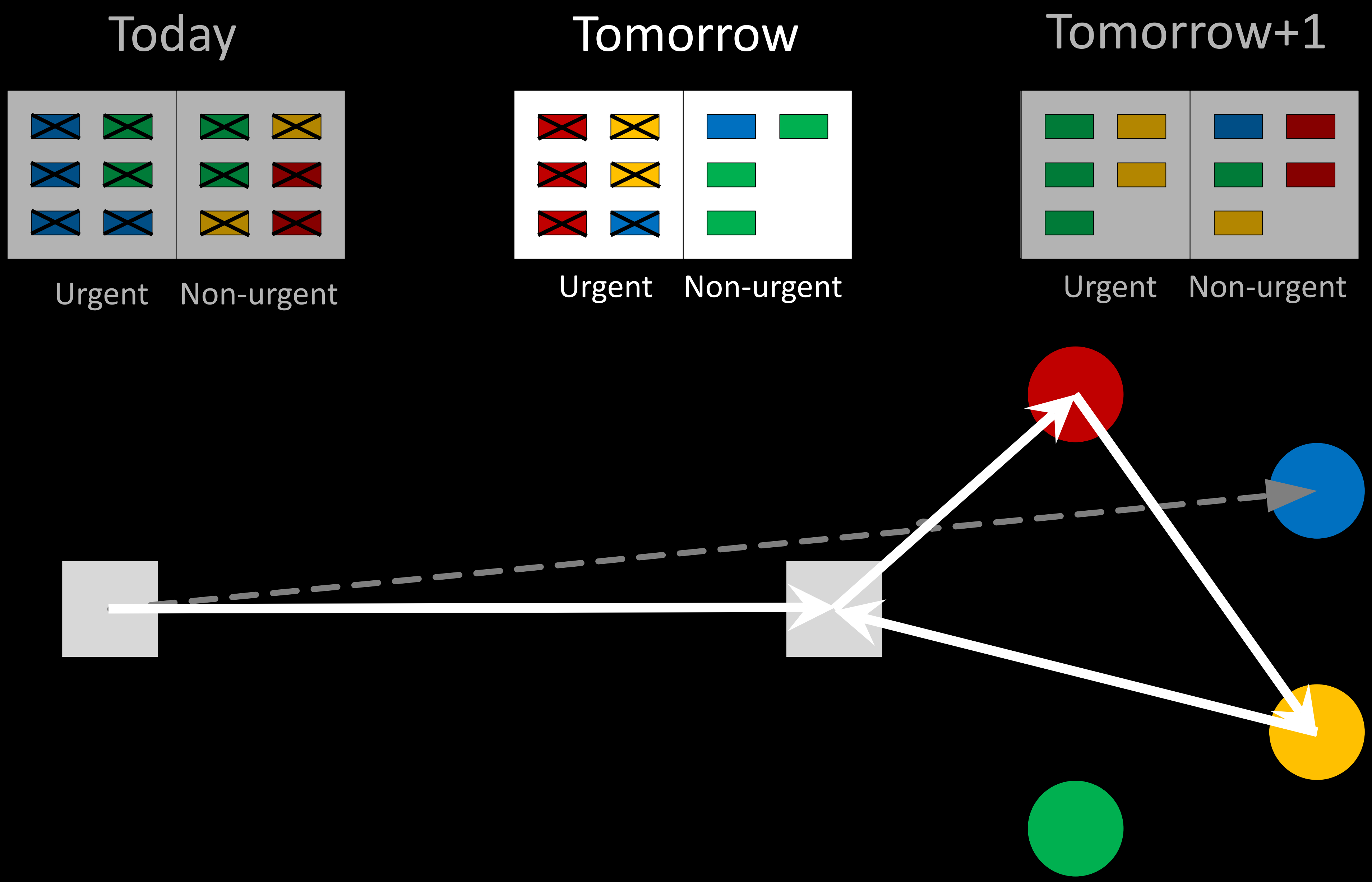
EXAMPLE 3: SYNCHROMODAL TRANSPORT



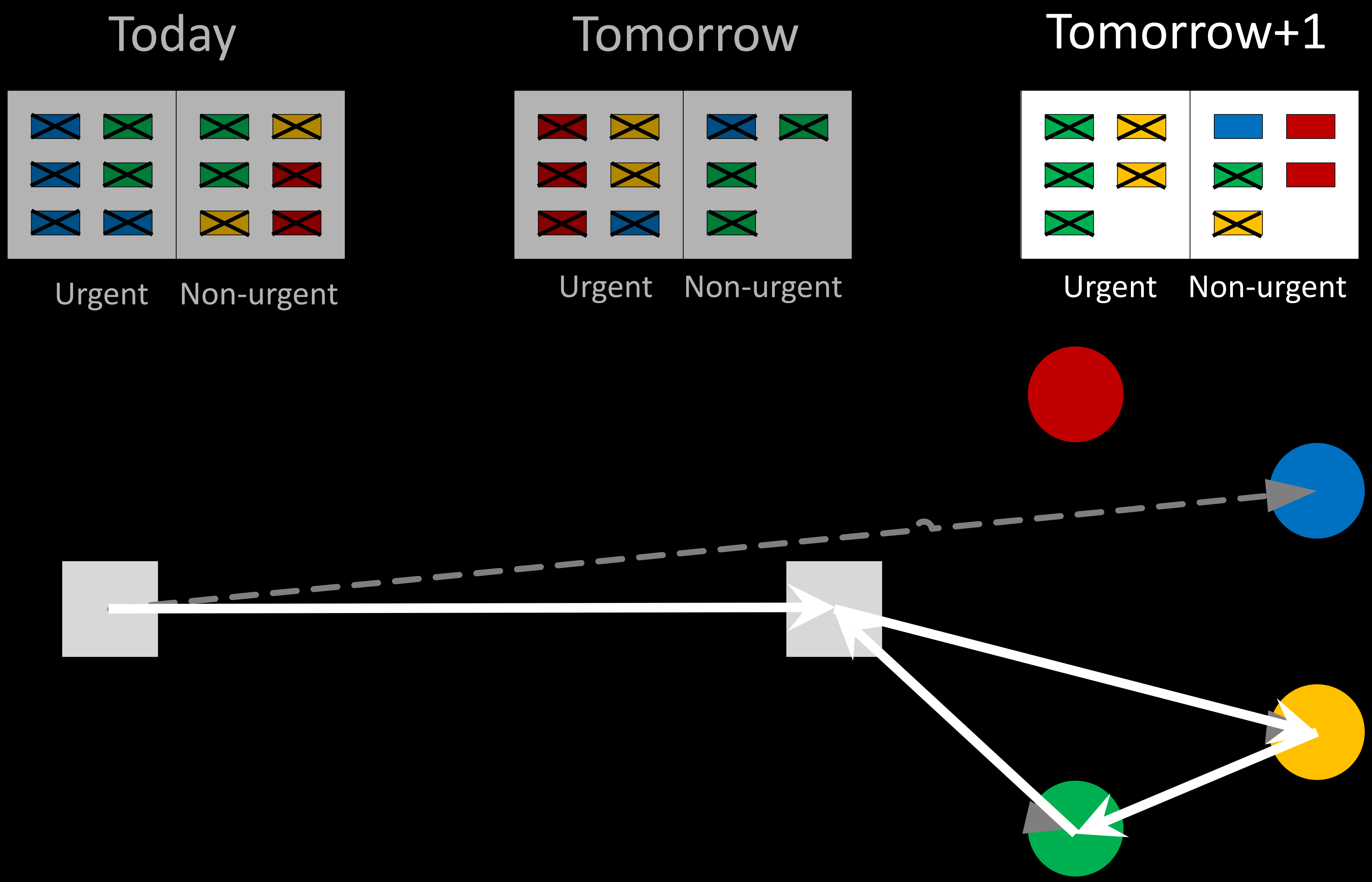
EXAMPLE 3: SYNCHROMODAL TRANSPORT



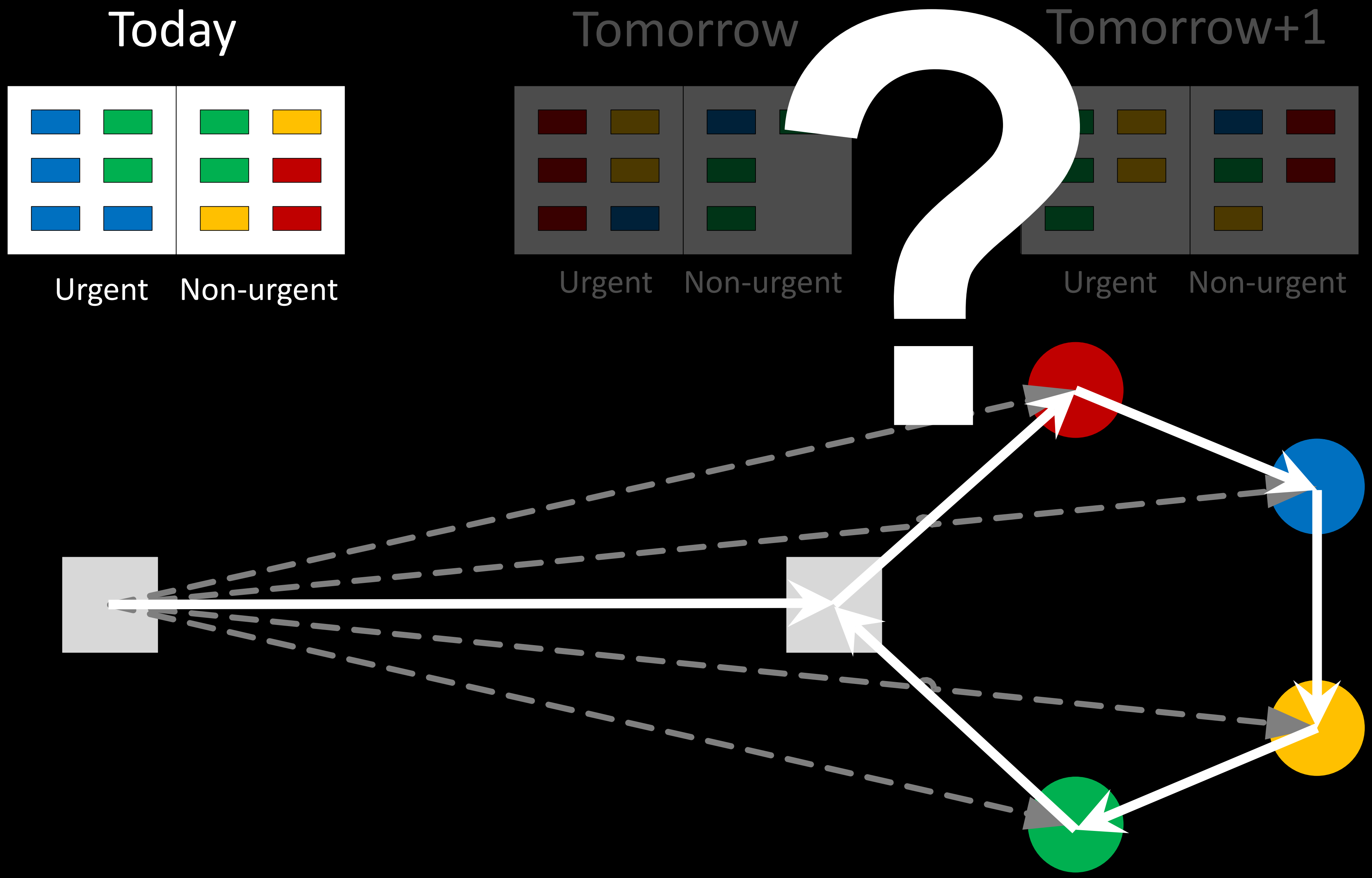
EXAMPLE 3: SYNCHROMODAL TRANSPORT



EXAMPLE 3: SYNCHROMODAL TRANSPORT



EXAMPLE 3: SYNCHROMODAL TRANSPORT



EXAMPLE 3: SYNCHROMODAL TRANSPORT

3. DYNAPLEX

DYNAPLEX PROJECT

- Deep Reinforcement Learning for Data-Driven Logistics
- Motivated by breakthroughs in DRL for gaming: when making logistics decisions, it is equally important to anticipate the uncertain future (e.g., orders, delays, disruptions, etc.)
- DynaPlex toolbox: in a similar fashion as AlphaZero was designed as a generic tool to solve various games, we created the DynaPlex toolbox to support the rapid development of automated decision making based on DRL
- Plug and play architecture to solve any dynamic logistics problem, involving sequential decision making under unertainty

DYNAPLEX TOOLBOX

Lots of freedom!

Less dependent on historical data!

Modelling framework

Algorithmic framework

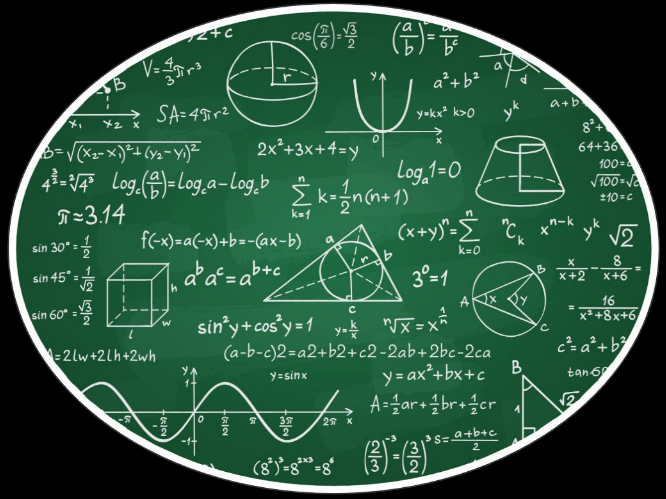
Algorithm selector

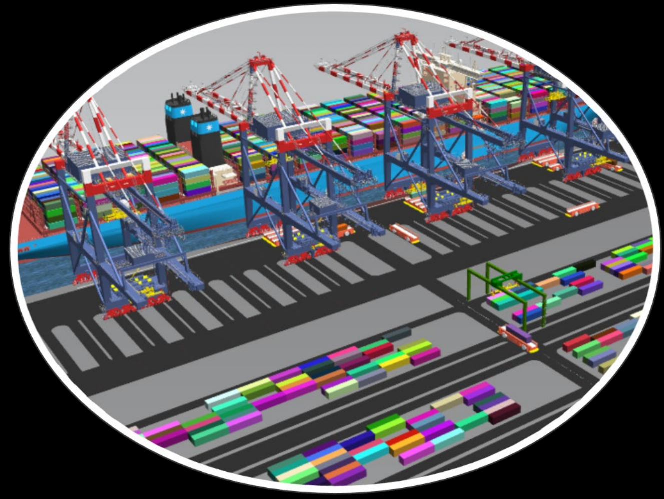
Experiment design

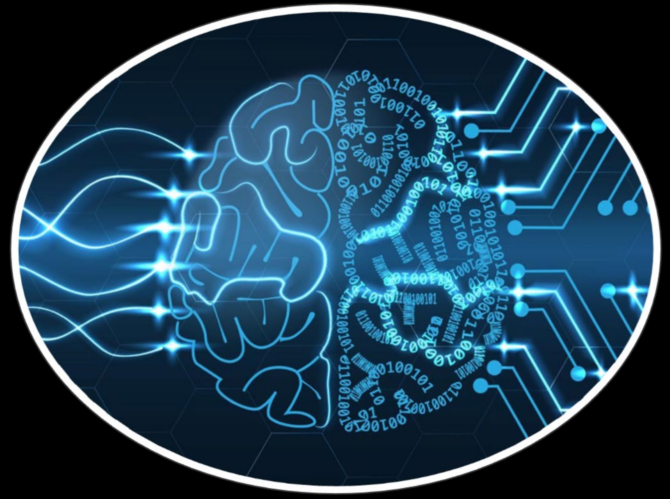
Parameter tuning (for simulation and ML)

Visualization

Parallelization







Step 1: input model

MDP elements:

- States
- Decisions
- Rewards
- Transitions

Step 2: generate simulation

Automatically generate a computer simulation model based on the MDP elements. Within this simulated world, agents learn by trial-and-error.

Step 3: train ML model

Train a machine learning model describing:

- the value of states or state-actions
- prescribing the actions to take in certain states
- a combination of these

EXPANDING TEAM AND PROBLEMS

- Same-day delivery problem
- Lost sales inventory replenishment problem
- Dual sourcing problem with additive manufactured parts
- 3D Bin Packing
- Joint inspection and replenishment of inventory with drones
- Routing of robots in an AS/RS warehouse
- Spare parts stocking in a small supply chain (like the beer game)
- Airplane revenue management
- Traveling maintainer problem
- Multimodal transport with trucks and barges
- Electric fleet dispatching
- Last-mile humanitarian logistics
- Dynamic fleet size problem
- Perishable inventory control
- Cliff-walking problem
- Maintenance planner
- Games (Tetris, Pacman)
- Many more…

QUESTIONS?