Multi-agent Dynamic Resource Allocation: A Reinforcement Learning Approach

S. Costantini[1], G. De Gasperis[1], P. De Meo[2], **F. Gullo**[1], A. Provetti[3]

[1]University of L'Aquila, Italy

[2]University of Messina, Italy

[3]Birkbeck, University of London, UK

International Joint Workshop of Artificial Intelligence for Healthcare (HC@AIxIA) and HYbrid Models for Coupling Deductive and Inductive ReAsoning (HYDRA):
HC@AIxIA+HYDRA 2025

co-located with

*The 28th European Conference on Artificial Intelligence (**ECAI 2025**)*

Bologna, Italy, October 25-26, 2025

Aim of this work

Study of cooperative resource allocation in multi-agent systems (hospital networks).

Agents redistribute limited resources (doctors) respecting local constraints and global objectives.

Dynamic scenario: staffing requirements vary over time.

Motivation

Fair allocation is crucial in decentralized systems (e.g., healthcare).

Hospitals may lend staff during emergencies.

Limitations of existing approaches [a]: they focus on a static scenario.

[a] Costantini et al., *Resource Allocation with Cooperative Agents*, RCRA Workshop @ICLP 2025.

Proposed Approach (Overview)

A **Reinforcement Learning** (**RL**) approach based on Proximal Policy Optimisation (PPO).

Reward composed of three terms, which penalize, respectively:

- **Deviation from the target** staffing level of every hospital.
- **Violation of the minimum staffing level** needed by each hospital to operate effectively.
- **Imbalance** of the resulting staffing levels.

Problem definition

Input:

Hospitals $H = \{\vec{h}_1, \dots, \vec{h}_n\}$

Each $\vec{h}_i$ described by:

- c_i: current #doctors
- t_i: target #doctors
- m_i: minimum #doctors

Objective: transfer doctors from a hospital to another such that :

- y_i is as close as possible to t_i
- $y_i < m_i$ arises the least possible
- the various $|y_i - c_i|$'s are as much similar to each other as possible

(y_i: #doctors of $\vec{h}_i$ after the transfer)

Proposed method

A **Reinforcement Learning** (**RL**) approach which maximizes the following reward:

$$\mathcal{R} = -\underbrace{\sum_{i=1}^{n}(y_i - t_i)^2}_{\text{target error}} - \underbrace{\eta \sum_{i=1}^{n} \max(0, m_i - y_i)}_{\text{minima violations}} - \underbrace{\sigma^2(a)}_{\text{fairness}}$$

Target error: deviation from target, i.e., the reward penalizes configurations where the current allocation of doctors is far from the desired target distribution.

Minima violations: heavy penalties are given if any hospital drops below its minimum number of staff units needed to guarantee its operation.

Fairness in transfer procedures: uneven transfers of doctors discouraged; the imbalance of the entire re-allocation process is quantified by the variance of the distribution of doctors across the various hospitals.

Experiments

150 synthetically-generated hospitals;

random c_i, m_i, t_i;

repeated 20 times for robustness.

Metrics:

- **Mean Absolute Error** (Target Deviation)

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - t_i|$$

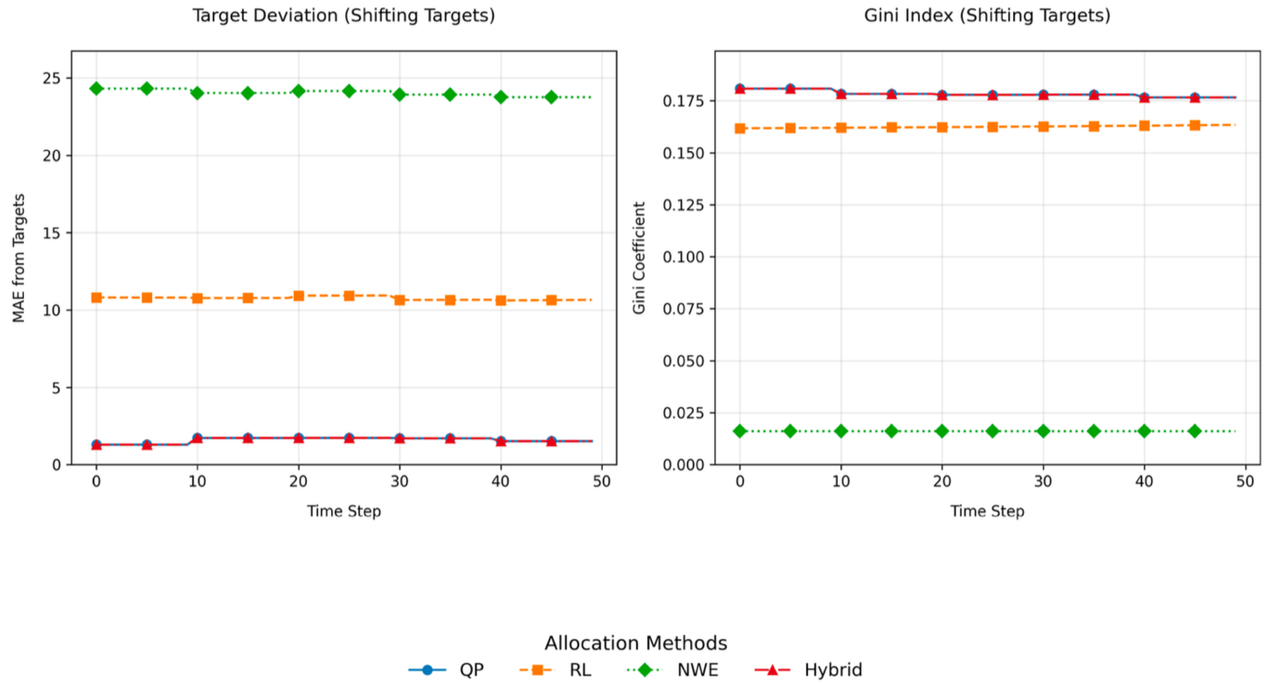
- **Gini Index** (inequality)

$$\mathcal{G} = \frac{1}{N}\left(N + 1 - 2\sum_{i=1}^{N} \frac{\sum_{j=1}^{i} y_j}{\sum_{j=1}^{N} y_j}\right)$$

Compared the **proposed RL approach** to **existing methods** for the **static** scenario (QP, NWO, Progressive Taxation, Hybrid [a]).

[a] Costantini et al., *Resource Allocation with Cooperative Agents*, RCRA Workshop @ICLP 2025.

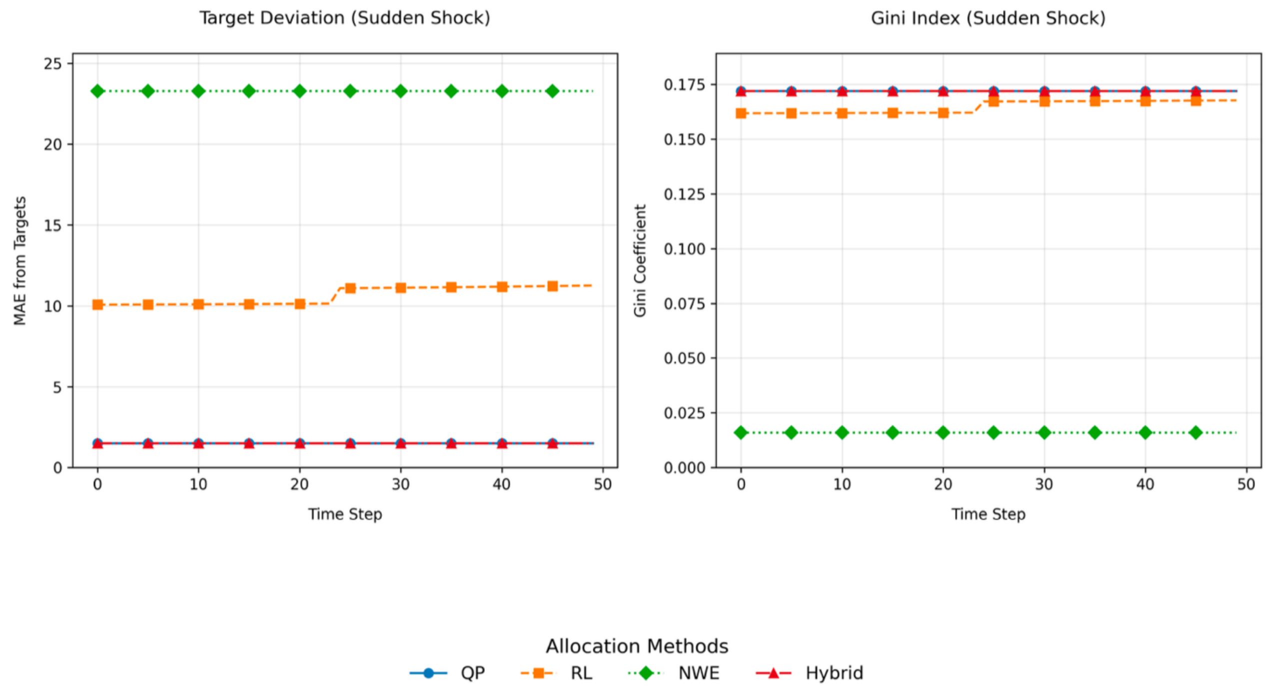
Results: Shifting Targets scenario



(a) Average Target Deviation (MAE) and Gini Index of the RL (proposed), NWO, QP, Hybrid (baselines) strategies in the Shifting Targets scenario.

Shifting Targets scenario:
hospital-specific targets undergo minor adjustments (loses or gains at most two doctors) every 10 time steps (overall time horizon: 50 time steps).

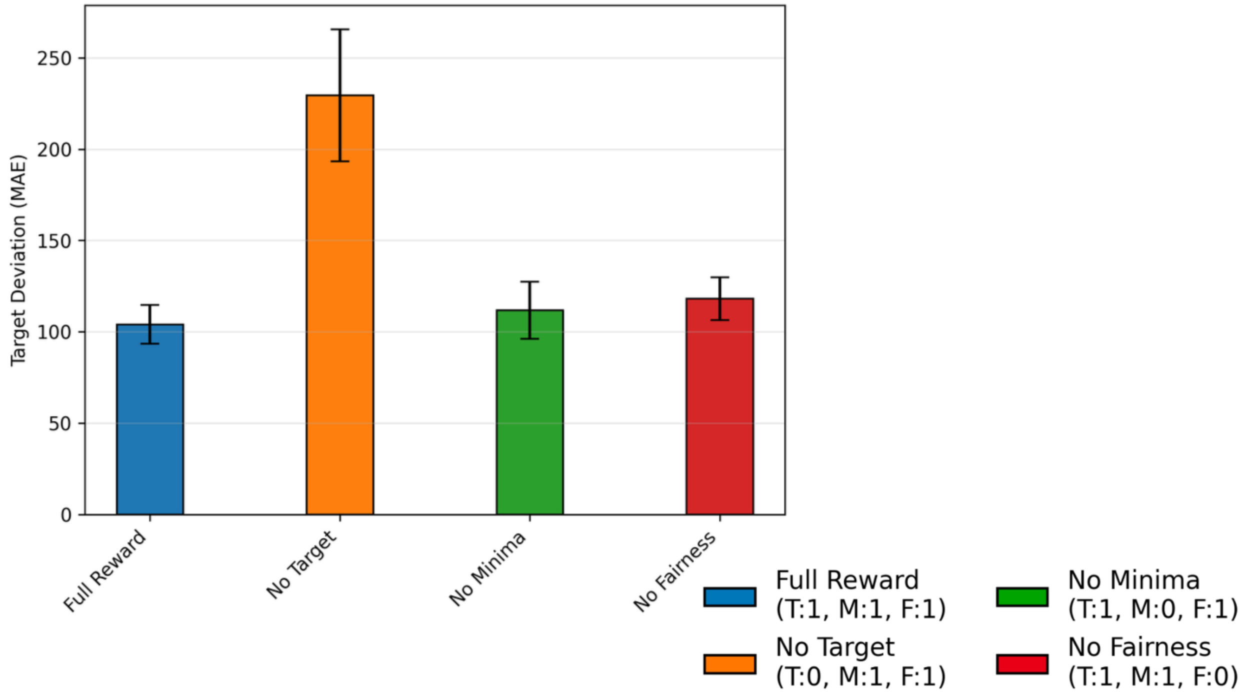
Results: Sudden Shock scenario



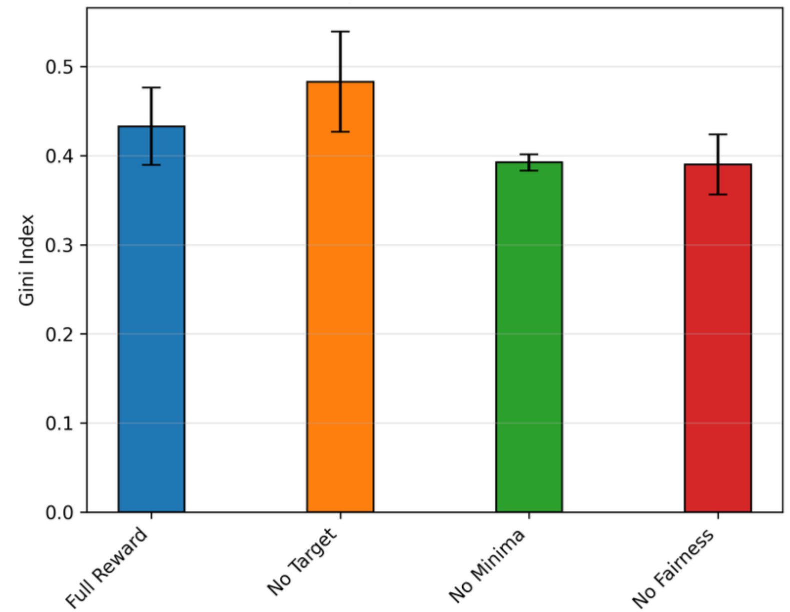
(b) Average Target Deviation (MAE) and Gini Index of the RL (proposed), NWO, QP, Hybrid (baselines) strategies in the Sudden Shock scenario.

Sudden Shock scenario:
one hospital loses half its workforce at the midpoint of the time horizon.

Results: ablation study

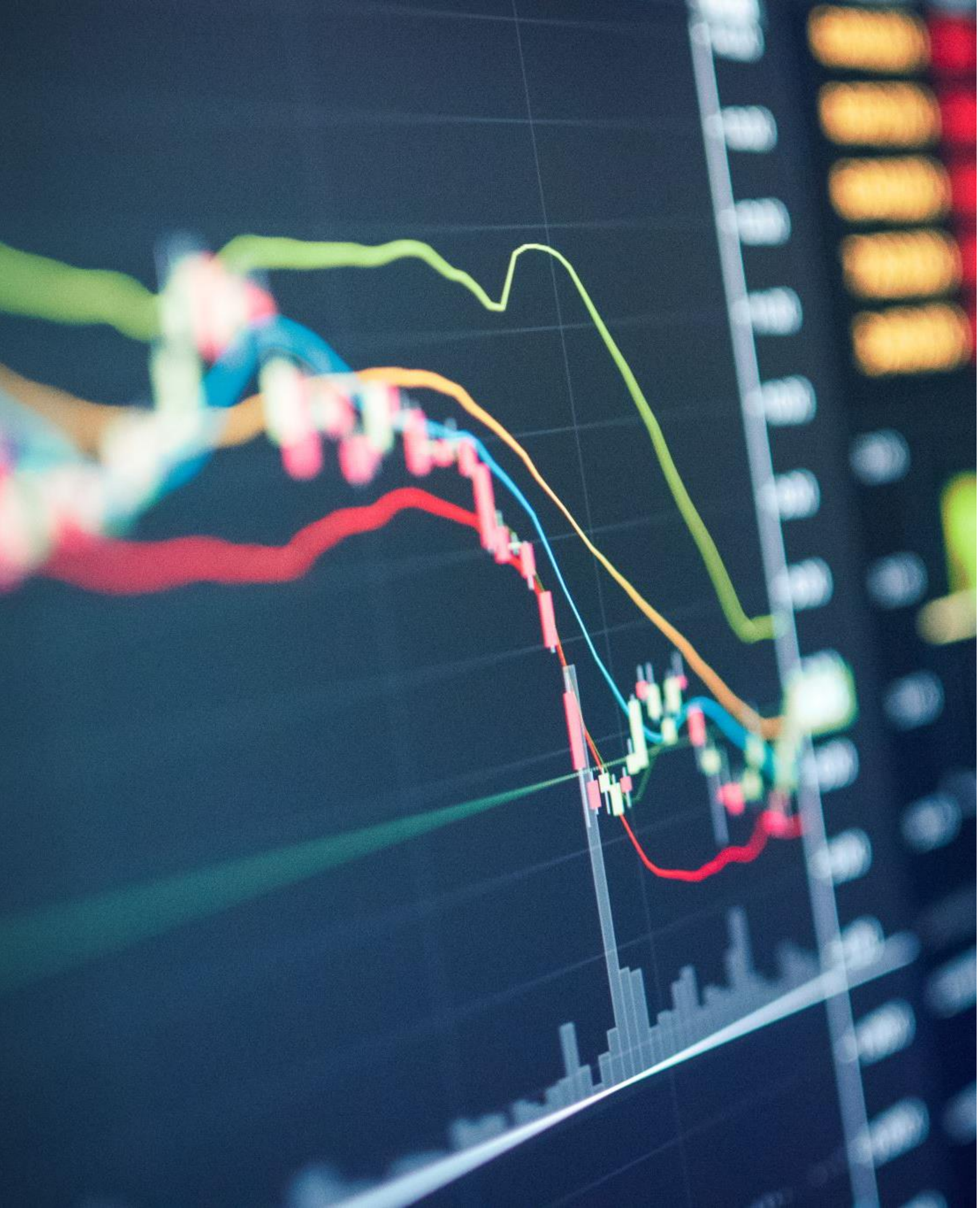


- Target deviation is the most impactful component.
- Removal of minimum or fairness component leads to a lower Gini Index but at the price of a consistent drop of MAE.



Conclusions

- Tackled the problem of cooperative resource allocation in multi-agent systems (hospital networks).
- Advanced the state of the art by handling a dynamic scenario where staffing requirements vary over time.
- Devised a reinforcement learning approach to tackle the target problem.
- Performed experiments to assess the relevance of the proposed method.



Thanks!

Questions?