

ORIGINAL RESEARCH

Open access

Artificial Intelligence for the Energy Transition: Real-Time Control of Renewable-Rich Power Grids

Ethan Marshall¹, Noor Al-Mutairi^{2*}, Kevin Brooks³

Abstract

Reaching carbon neutrality by 2050 will require transmission system operators to accommodate unprecedented levels of variable solar and wind generation while preserving system reliability. To address this systemic constraint, Réseau de Transport d'Électricité (RTE), together with the research community, introduced Learn To Run A Power Network (L2RPN), a crowd-sourced competition designed to accelerate the integration of intermittent renewable sources into grid operations. L2RPN relies on 16 years of weekly scenarios (832 in total) modeled on a 118-node network under realistic operational constraints, framing real-time grid management as a Markov Decision Process. The six competing teams approached the problem by designing autonomous agents that combined heuristics, optimization, data scaling, supervised learning, and reinforcement learning techniques. We deliver a comprehensive assessment of all six teams' performances under the competition's stringent framework. Furthermore, we conduct an extensive examination of the publicly released winning solution, which ensures consistent decision-making across scenarios, performs real-time multimodal actions in less than 5 seconds, and achieves efficient topology control through action-space reduction and a neural policy that predicts beneficial grid actions with over 80% accuracy. Concurrently, we trained a neural alert module using 315,000 samples generated from leading agents, obtaining 93.9% recall in identifying hazardous states and enabling agents to anticipate potential failures. Ultimately, this study highlights both the potential and current limitations of AI in real-time grid operation, while establishing a transparent foundation for developing more robust and trustworthy systems to support the energy transition.

Keywords Reinforcement learning, Power network, Carbon neutrality, Renewable energy, Competition

*Correspondence:

Noor Al-Mutairi
nmutairi@kfupm.edu.sa

¹ Department of Smart Grid Engineering, Faculty of Electrical Engineering, University of Manchester, Manchester, United Kingdom

² Department of Energy Analytics and Optimization Systems, Faculty of Engineering, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

³ Department of Sustainable Power Systems, Faculty of Engineering, University of Melbourne, Melbourne, Australia

Introduction

In the late 2010s, roughly 85% of global energy production came from fossil fuel combustion, which emitted greenhouse gases such as carbon dioxide (CO₂). These emissions have continuously increased since the onset of the industrial era. It is now broadly recognized that the environmental footprint of modern societies has become particularly pronounced since the 1950s.

To avert irreversible ecosystem degradation, it is crucial to substantially reduce greenhouse gas emissions and other environmental pressures [1]. Policymakers are therefore encouraging the development of alternative energy conversion technologies based on renewable and low-carbon resources, notably solar and wind. Over the past 20 years, these

technologies have progressed considerably, enabling large-scale deployment. The French transmission system operator, Réseau de Transport d'électricité (RTE), recently released findings from an extensive study that presents multiple scenarios for the future management of the French power system [2]. In response to ecological imperatives, most of these scenarios emphasize nuclear and renewable energy sources, with intermittent renewables often accounting for more than 50% of the energy mix.

Accordingly, research efforts have increasingly focused on assessing the feasibility of rapid and large-scale integration of intermittent renewable generation in the medium term. However, solar and wind generation pose specific operational challenges for power grid integration [3]. Their output fluctuates significantly with weather conditions, potentially disturbing the system balance. Moreover, existing energy storage capabilities remain limited, maintaining reliance on conventional controllable generators to complement intermittent sources such as solar and wind. Consequently, some traditional generation assets remain indispensable for ensuring supply stability.

Complexity of power network operations

RTE is responsible for the real-time operation of the French transmission network, ensuring that electricity production continuously matches consumption while anticipating the consequences of all possible contingencies, whether scheduled or unforeseen. Although economic optimization is considered, system security remains the foremost priority. The grid must function safely at every second, across the entire territory and interconnected regions. The governing security principle, known as the “N-1 criterion,” requires that, following any single contingency—such as the unexpected outage of a transmission line or generator—the system must continue to operate under normal conditions.

Highly trained dispatchers play a pivotal role in supervising power flows. Their primary objective is to prevent damaging overloads and, more critically, cascading failures that may result in large-scale blackouts, all while complying with the N-1 rule. Operating in real time from control centers, they continuously mitigate local overloads by redirecting power flows along alternative paths and coordinate with generators and consumers to adjust injections and demand [4] when necessary. They may also curtail renewable generation, including wind and solar, during periods of excess production. Their interventions rely on comprehensive knowledge of the entire network, which constitutes an extremely complex system and remains challenging even for advanced optimization techniques [5].

Increasing variability and shrinking operational margins associated with the Energy Transition will generate substantial operational challenges. Grid management will demand greater flexibility and responsiveness, requiring dispatchers to make more numerous, faster, and increasingly complex decisions [5, 6], while preserving the fundamental mission of the power system: reliably supplying consumers and preventing blackouts despite heightened system complexity (Figure 1).

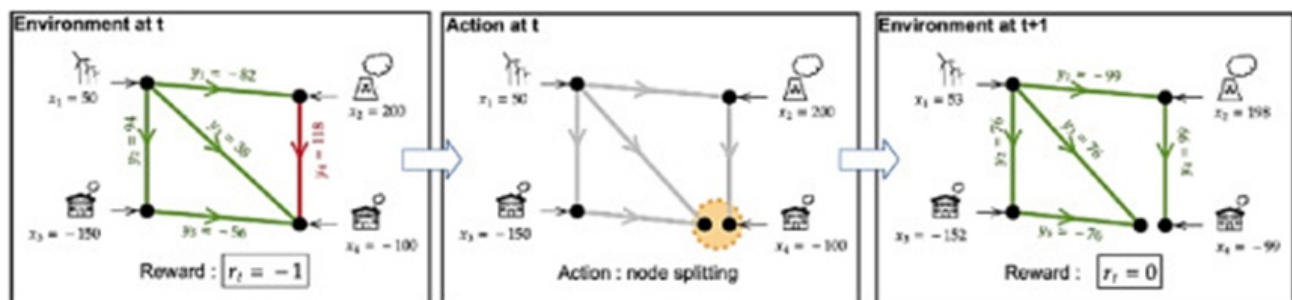


Figure 1. Power system operation: Dispatchers supervise the transmission network and implement corrective measures when necessary to maintain secure operation and prevent line overloads. If, in the environment at time t (left), a transmission line is overloaded (shown in red), a corrective measure may be applied (center), such as “node splitting,” leading to the restoration of “power network safety” in the environment at time $t+1$ (right). An artificial agent would receive a reward for resolving such overload conditions. (For interpretation of the references to color in this figure legend, readers

are directed to the web version of this article.)

Figure adapted from Marot et al. [7].

A trustworthy AI-based assistant for dispatchers: anticipatory recommendations

Looking ahead, enabling more flexible, responsive, and efficient decision-making may be achieved by developing a “trusted AI assistant” to support system operators [8]. This approach differs from existing simulation-based decision-support tools, which impose substantial computational and data-processing demands as system complexity grows and the number of scenarios and corrective options expands. Such a “trusted AI assistant” can be viewed as a recommendation engine that enhances grid analysis and extends operators’ effective decision time horizon. It delivers guidance “on demand,” reduces reliance on repeated large-scale simulations, and assesses and communicates the confidence associated with its recommendations.

Developing this type of recommendation system constitutes a key objective for real-world transmission networks. To stimulate progress in this direction, we organized a 3-month AI for Energy Transition competition in summer 2023, targeting innovative start-ups and scale-ups with AI expertise. Six companies participated (Artelys, N-Side, LA Javaness, AI-Redefined, Tilt-Energy, Baalbeek Management), with a closely contested top-3 ranking on the final leaderboard. For this competition, we introduced a new synthetic power system operations environment that reflects Energy Transition conditions and incorporates additional technical challenges for participating agents. Our contributions are therefore as follows:

- We introduce the first virtual power grid operations testbed tailored to Energy Transition scenarios, featuring up to 30% renewable energy generation and less than 8% carbon-emitting energy sources.
- We enhance an AI assistant-oriented feature that encourages artificial agents to alert human operators whenever uncertainty arises regarding their capacity to manage the network safely.
- We conduct an in-depth analysis of participants’ results, complemented by post-hoc evaluation highlighting both agent strengths and weaknesses.
- We train a machine-learning-based AI assistant module and benchmark its performance against that of the competition participants.

To support these advancements, we developed a realistic, demanding testbed environment, incorporated trust-oriented functionality into AI agents, and organized a dedicated competition to assess their performance under operationally representative conditions.

Related Work

Power system virtual environments and testbeds

Artificial agent testbeds for power systems

Our artificial-agent virtual environment and testbed are built upon the Grid2Op framework [9], which emulates interactive power system operational scenarios controlled by artificial agents. Developed on top of the OpenAI Gym interface [10], and later Gymnasium [11], it formulates sequential decision-making as a Markov Decision Process (MDP). A key advantage is that it avoids imposing restrictive assumptions or simplifications on system dynamics, unlike alternative formulations such as Optimal Power Flow [12]. This enables a clear separation between problem modeling and problem solving, ensuring rigorous benchmarking conditions.

Comparable frameworks have recently emerged for specialized applications, including stability control (RLGC [13], andes gym [14]), economic dispatch (SustainGym [15], CommonPower [16], PowerGridWorld [17]), and microgrid management

(pymgrid [18], gym-anm [19], open modelica microgrid [20]). These platforms similarly enable agent interaction with power systems through the standard “OpenAI Gym” interface [10]. However, Grid2Op remains particularly suitable for simulating large-scale transmission grid operations, especially for congestion management tasks. It explicitly models grid infrastructure with limited line capacities, offers comprehensive representations of observations and actions—including topological modifications—and incorporates time-dependent operational constraints. Additionally, it supports collaboration between human and virtual agents. Its modular architecture and sustained maintenance over several years have contributed to its widespread adoption.

Existing Grid2Op virtual environments

To reflect the sequential dynamics of power system operation, multiple synthetic MDP-based grid environments featuring daily scenarios at 5-minute resolution have been progressively developed and refined.

Early environments were limited to small networks [21], comprising up to 14 electrical nodes in the IEEE14 system. More recent implementations have expanded to networks with 118 nodes [22]. Beyond scaling in size, these environments have incorporated increasingly realistic network representations, including additional components such as battery storage, the ability to simulate the outcome of an action before its application, and mechanisms for on-demand scenario generation. The range of permissible grid actions has also broadened over time. Initially restricted to topological reconfiguration (a high-dimensional, discrete, combinatorial action space) [23], these environments now also support renewable curtailment actions and redispatching of thermal generators [24].

In real-world operation, managing unexpected line outages—which reduce the grid’s overall transmission capacity—is a primary challenge for grid operators. Within these environments, such contingencies have been modeled as adversarial “opponent attacks” [25], representing unforeseen line disconnections. Achieving robustness against these events remains a central difficulty. To address such attacks, advanced agents often implement preventive measures to anticipate the most severe contingencies. Nevertheless, none of these environments has yet provided a realistic representation of the Energy Transition, characterized by gradual grid capacity expansion combined with rapid growth in renewable integration.

Beyond safety-focused network management, [24] explored the implementation of trust mechanisms by issuing early warnings to operators when proposed actions exhibit low confidence. This initial framework sought to incorporate key dimensions of trust, namely reliability, credibility, and intimacy. In the following sections, we propose an enhanced modeling approach toward this objective, introducing capabilities such as forecast utilization and an improved alert mechanism designed to strengthen confidence in agent decisions.

Summary of the most effective artificial agents

In recent years, diverse methodologies have been introduced for operating power networks in synthetic environments, ranging from rule-based systems and optimization techniques to machine learning (ML) and reinforcement learning (RL). Although expert-driven rules can deliver acceptable results on the smallest IEEE14 grid, their effectiveness does not extend to larger-scale networks. Except for an advanced adaptive expert framework [26], no fixed rule-based strategy has yet demonstrated satisfactory performance.

Conversely, ML and RL methods have demonstrated superior inference speed compared with optimization-based techniques, a critical factor for near-real-time operational decision-making.

For example, the curriculum agent represents an initial standardized and efficient [27] progressive training strategy in which the agent advances through structured stages. At the outset, it narrows the extremely large action space to a few hundred highly relevant actions using simulated experience. Subsequently, it develops a more capable agent model via supervised ML imitation learning, ultimately providing a warm start to a high-performing RL model that continues learning interactively within the synthetic environment. A more sophisticated RL methodology [28] integrates Deep Reinforcement Learning with evolutionary optimization to refine the decision policy. In a more recent environment, an innovative

approach adapted the AlphaZero framework (a model-free RL architecture combining Deep Learning and Monte Carlo Tree Search, trained through self-play beginning from random behavior) for grid topology control [29].

Nevertheless, even these advanced agents frequently encounter failures in highly challenging scenarios. Their reliability remains insufficient for fully autonomous deployment in real-world power grids. However, they have occasionally demonstrated the ability to coordinate multiple topological actions effectively to withstand extremely difficult conditions, where other agents—and potentially even human operators—might fail. This represents a noteworthy accomplishment. Beyond expert systems and RL, hybrid methodologies have also been investigated. For example, [30] proposed a hierarchical framework in which upper and intermediate layers rely on rule-based logic, while the lowest layer employs multiple Multi-Agent RL strategies.

Ultimately, since these approaches remain vulnerable to failure, all the previously mentioned solutions are more appropriately implemented under human oversight. This supports our decision to focus on developing a “trusted AI assistant,” which first necessitates enhancing the environment’s trust model and subsequently evaluating how newly designed agents perform within it.

Competition Design

This competition aims to challenge participants with power grid scenarios involving a 30% renewable energy share, despite infrastructure originally designed to support only 20%. It incorporates multi-area operations and introduces new performance indicators for alert effectiveness and decision latency, aiming to assess and advance AI-driven grid management strategies.

To emulate real-time grid operation scenarios, the competition utilizes Grid2Op [9], which formulates operational decision-making as a Markov Decision Process (MDP) (S, A, P_a, R_a) [31], as detailed in its documentation and illustrated in Figure 2. This framework enables AI researchers without electrical engineering expertise to engage with this otherwise complex problem.

Within this environment, time progresses in discrete 5-min intervals (reflecting current operational practices, where operators receive updated grid states every 5 minutes). For instance, a one-day scenario is divided into $24 \times 60 \times 5 = 7200$ time steps. AI agents interact by executing actions such as line switching, node splitting, redispatching, or curtailment. Given a present state $s_t \in S$ and an action $a_t \in A$ selected by the agent, the environment transitions to the next time step, computes the resulting power flow (i.e., electricity transmitted across each power line), and produces the subsequent state s_{t+1} .

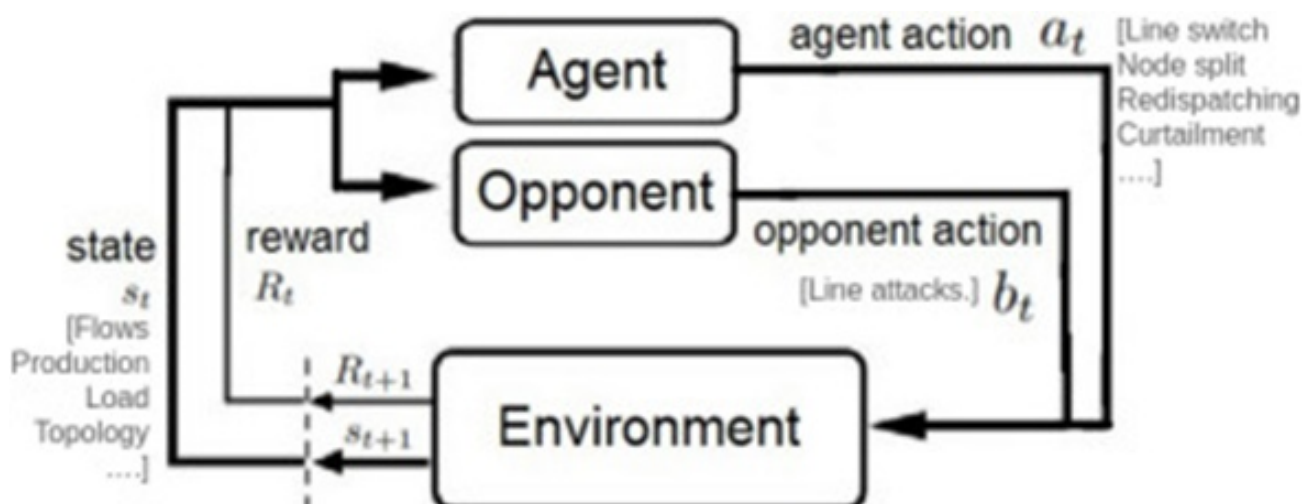


Figure 2. Standard RL paradigm with an opponent, where an agent interacts over time with a system environment modeled as an MDP. Note that agents may be heuristic- or optimization-based and are not limited to RL-based designs. Key action and state variable categories are emphasized.

Figure adapted from Omnes et al. [25].

An “episode” e , successfully controlled by an agent up to time T_{end} (within a scenario of maximum duration T_{max}), can be expressed as:

$$e = (s_1, a_1, s_2, a_2, \dots, a_{T_{end}-1}, s_{T_{end}}) \quad (1)$$

Where s_1 denotes the initial observed state and $s_{T_{end}}$ the final state. Termination occurs either due to a game-over event (e.g., a blackout) at time T_{end} or because the agent reaches the scenario’s maximum duration, $T_{end} = T_{max}$.

The competition builds upon pre-existing datasets and features. Specifically, it reuses established operational constraints, stopping criteria, grid physics simulations, and scheduled line outage events, along with unscheduled outages modeled as an adversarial agent. To enhance realism with respect to actual grid operations, the following extensions are introduced:

- Energy Transition-oriented virtual environment: Realistic scenarios in which wind and solar generation constitute 30% of the energy mix, with adjusted line capacities and congestion levels reflecting Energy Transition conditions, extended forecasting horizons, and multi-area operational settings comparable to real grids.
- AI-assistant module: An alert mechanism enabling AI agents to notify human operators when their operational reliability may be compromised.
- Enhanced evaluation framework: Expanded metrics including a low-carbon performance indicator, alert-related metrics, and decision-time measurements.

These enhancements are detailed in the following subsections.

Virtual grid operations environment for modeling the energy transition

The virtual environment is publicly accessible for research and analytical purposes. Its principal characteristics are outlined below.

The multi-area grid

The grid employed is a modified version of the IEEE 118-bus system with substantial renewable integration [32]. It is displayed in **Figure 3**, including both generation and consumption elements. Its topology enables more than 100,000 elementary actions, forming the foundation of an extensive combinatorial action space, which also encompasses continuous controls such as production redispatch, renewable curtailment, and battery storage management.

In the context of the Energy Transition, coordination both within individual grids and across interconnected areas becomes essential. To capture this aspect, the grid is partitioned into 3 distinct areas, allowing concurrent yet independent actions in each region while potentially necessitating cross-area coordination. Additionally, it introduces a higher number of simultaneous unplanned line outages (attacks) across these areas. This design increases the operational complexity for agents and encourages coordinated management of multiple incidents across the entire network.

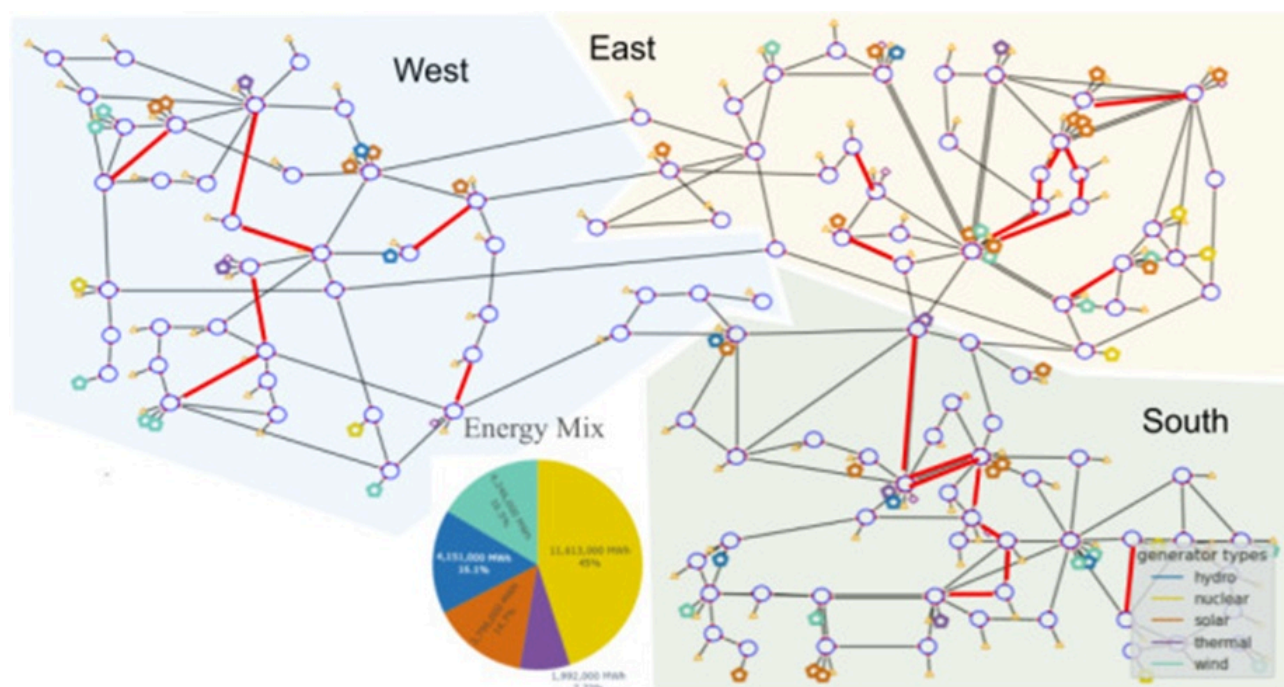


Figure 3. IEEE 118 grid divided into 3 areas, illustrating generator locations and types (solar, wind, hydro, thermal, nuclear), along with its 30% renewable (wind and solar) energy mix share. Structurally congested lines are marked in red.

Realistic production and consumption scenarios

The targeted energy composition includes 30% new renewables (solar and wind) and 45% nuclear power, aligning with objectives outlined in the French Energies of the Future report for the 2030s.⁵ This represents a tripling of renewable capacity within less than 10 years. Time-series data are produced using the Chronix2Grid (Python) package at a 5 min resolution over one week, as exemplified in [Figure 4](#). Wind generation modeling has been enhanced to reflect more realistic temporal variability as well as maintenance intervals for wind farms. The training dataset contains fifteen years of generated data. Additional datasets can also be automatically created by participants if required.

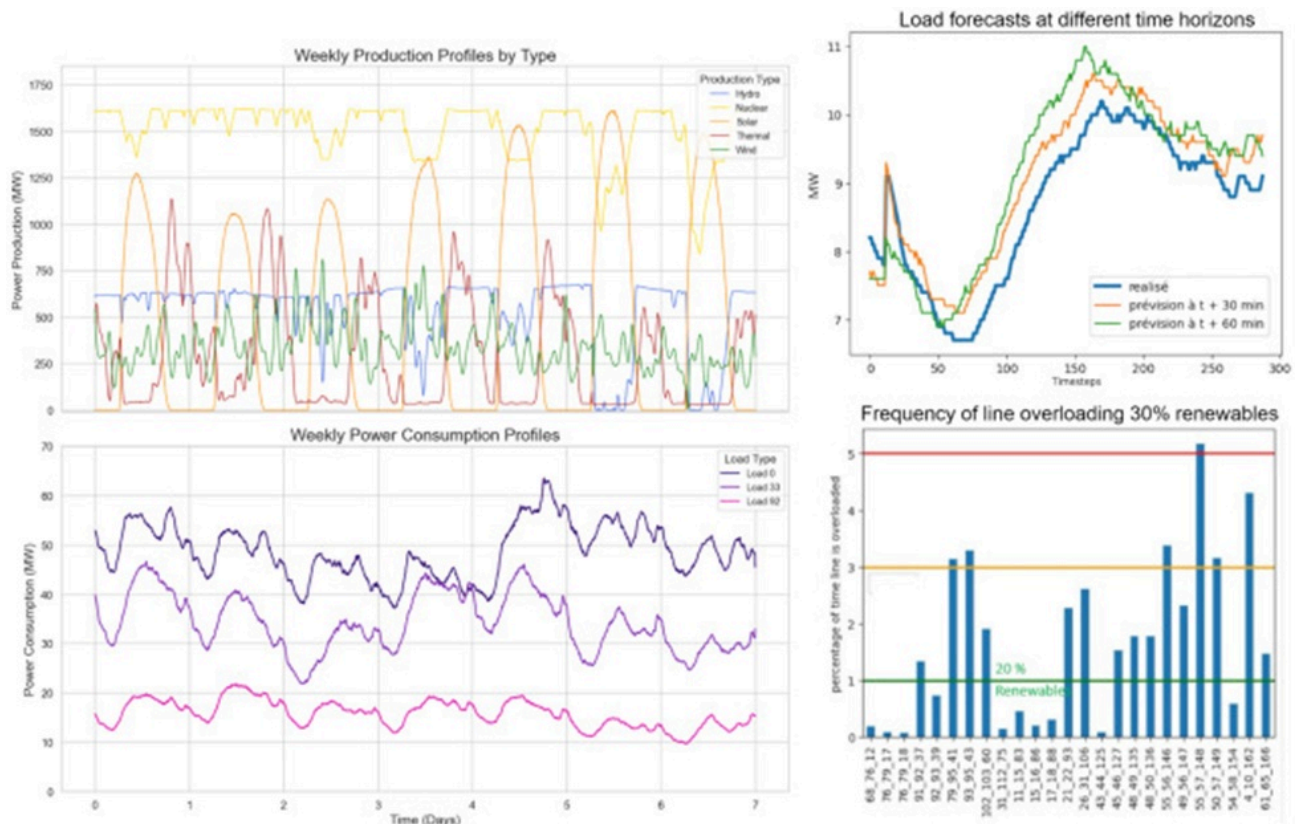


Figure 4. Left: Example time series depicting the power generated by each production type at every time step, together with individual load profiles, highlighting their distinct dynamics.

Right (top): Illustration of load forecasts at different horizons—actual load versus forecasts at 30 and 60 min. As expected, the 60 min forecast has lower average accuracy than the 30 min forecast.

Right (bottom): Structurally congested lines under a 30% renewable energy mix when no grid actions are applied. Lines exceeding a 1% congestion frequency (green horizontal bar) show increased congestion compared to the original 20% renewable mix, for which grid capacity was better suited. Orange and red bars (3% and 5% congestion frequencies) indicate thresholds beyond which grid expansion would remain necessary even after accounting for operational flexibilities. Consequently, congestion frequencies were maintained within this interval (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Line congestions during the energy transition

The accelerated expansion of renewable generation intensifies grid congestion, as adapting infrastructure is constrained by time and cost limitations. Therefore, the current network must be operated with enhanced flexibility. To reflect this Energy Transition characteristic, grid capacity is calibrated for an energy mix containing 20% renewables. Under this configuration, certain lines experience congestion at a 1% frequency, consistent with observations from real systems. However, when an additional 10% renewable share is incorporated—anticipating a future 30% mix—line capacities are not upgraded accordingly. As shown in **Figure 4**, several lines then experience significantly more frequent congestion, while a small number exhibit reduced congestion. This results in a more challenging operational setting, as expected in the years ahead.

Forecasts and simulation

In actual grid management, operators routinely rely on simulations to evaluate prospective actions, using demand and/or generation forecasts at specific network points. Accordingly, it is appropriate to provide (artificial) agents within the Grid2Op environment with access to such forecasts, enabling them to test and refine their decisions. As illustrated in

Figure 4, these forecasts are inherently imperfect, with accuracy decreasing as the prediction horizon extends, yet they remain coherent across time horizons. They are calibrated to achieve a level of accuracy comparable to forecasts currently applied in real-world power system operations.

Alert feature

The alert mechanism is intended to assess an agent's capability to proactively notify a hypothetical human operator when it anticipates a potential failure within the next hour, possibly due to unexpected line disconnections. Unlike the earlier formulation in Marot *et al.* [24], alerts are not meant to predict the precise timing of failures, since such events are inherently uncertain and may result from unforeseen attacks. Instead, the objective is to detect conditional situations in which network security could be compromised within the next hour, particularly under possible contingencies drawn from a predefined list of vulnerable lines. This mirrors real operational practice, where operators must anticipate whether they can handle a contingency or should conduct further analysis.

At each time step, the agent must decide whether to issue an alert for any of these contingencies, balancing two objectives: avoiding excessive warnings that may distract the operator and ensuring that issued alerts reflect a consistent, credible threat assessment over time. Alerts are evaluated after an attack occurs on a specific line, based on the system's state following 1 h of operation—either successful management or a blackout. For a comprehensive depiction of this evaluation process, see **Figure 5**. Ultimately, an effective agent should recognize its operational limitations and notify the operator whenever uncertainty arises.

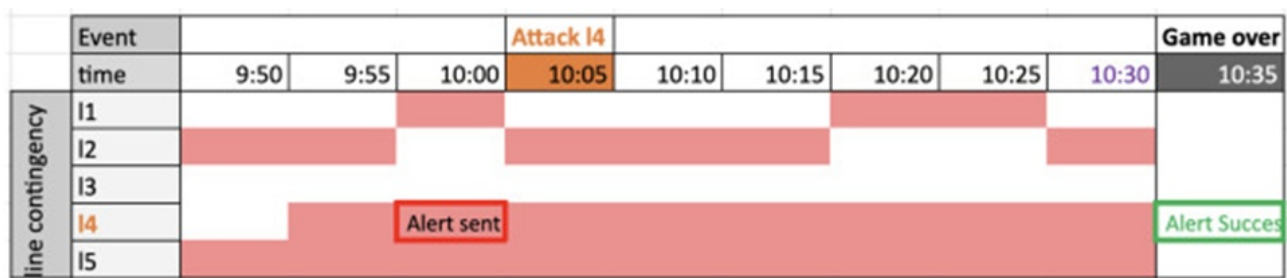


Figure 5. Example of agent alerts (red indicating an alert issued, white otherwise) at each time step for 5 monitored line contingencies (among attackable lines). An attack occurred on 14 at 10:05, and the agent had previously issued an alert.

The agent accurately anticipated instability within 60 min of the attack. A blackout occurred at 10:35, validating the prediction. The agent correctly acknowledged its lack of confidence and received a positive evaluation. If the system had remained stable for more than 1 hour beyond 11:00 without an improvement in agent confidence, the alert would have been deemed incorrect, resulting in a negative evaluation (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Evaluation

To establish participant rankings, a scoring function quantifies each agent's performance. In this competition, the total score is composed of 3 distinct elements:

- **Operation score:** represents the operational cost of managing the power grid, incorporating blackout-related costs, energy losses within the network, and action costs as specified in Marot *et al.* [22]. This score lies within the interval $[-100,100]$.
- **Low-carbon score:** reflects the volume of renewable energy curtailment. Lower curtailment corresponds to greater carbon efficiency in grid operation. This metric also spans $[-100,100]$, where -100 indicates that more than 50% of renewable generation has been curtailed, and 100 signifies that all available MW of renewable energy has been utilized.

- **Assistant score:** measures how frequently the agent correctly anticipates its own confidence in handling contingencies, particularly line disconnection incidents. This score likewise ranges within $[-100,100]$.

The final overall score is computed as a weighted aggregation of these normalized components, as described below.

$$Score = 0.5 \times Score_{Operations} + 0.2 \times Score_{Low-carbon} + 0.3 \times Score_{Assist} \quad (2)$$

Operation Score is of utmost importance, as it directly reflects the fundamental objective of power grid management, particularly blackout prevention. A strong emphasis is also placed on the Assistant Score, since effective grid management increasingly depends on Human–AI collaboration. The Low-carbon Score receives a lower weight because its impact is already partially embedded in the Operation Score. Nonetheless, it remains a key indicator for assessing the effectiveness of renewable energy integration. Additional details regarding each component are provided below.

Operation score

The Operation Score is assigned a higher weighting because maintaining grid stability and preventing blackouts remain the primary objectives. Safety considerations—and therefore coordination between the artificial agent and the human operator—are also critical, which justifies allocating a substantial weight to the Assistant Score. Competent agents are expected to minimize renewable curtailment naturally, since such costs are already incorporated into the Operation Score. However, to discourage undesirable strategies that rely on excessive curtailment to complete scenarios without blackouts, the Low-carbon Score is introduced as an additional penalty mechanism. Under all circumstances, completing a scenario should remain preferable to triggering a blackout.

Low-carbon score

Achieving net-zero carbon emissions through the Energy Transition largely depends on effectively integrating and utilizing renewable energy sources. This score is intended to measure an agent's capacity, throughout a scenario, to exploit available renewable generation fully. Specifically, it incentivizes minimizing renewable curtailment $E_{curtail}(t)$ relative to renewable production $E_{renew}(t)$, since curtailment typically necessitates redispatch toward carbon-intensive generation, thereby increasing emissions. Completely curtailing renewable output could serve as a simplistic yet counterproductive workaround for agents attempting to maintain operational stability and complete episodes. Such behavior is strongly penalized under this scoring component.

$$Rate_{\mathcal{E}_{curtail}}(e) = \frac{\sum_{t=1}^{T_e} \mathcal{E}_{curtail}(t)}{\sum_{t=1}^{T_e} \mathcal{E}_{renew}(t)} \quad (3)$$

The episode score is derived from this rate using a maximization-based function (–rate) and rescaled to remain within $[-100,100]$. A score of 100 indicates that no curtailment occurred during the episode, 0 corresponds to a 20% curtailment level, and –100 applies when the rate reaches 50% or above.

Assistant score

This metric measures the quality of the agent's alerting behavior, indicating how reliably it assesses its own capability to cope with unexpected line (l) disconnection incidents in advance ($alert(l,t-1)$, True or False) across a forthcoming 60min (12 timesteps) window. It operates similarly to a contingency status signal—green or red—for each monitored line, comparable to visual indicators used in real-world control centers. When an event $Ev_l(t)$ happens on a specific line l at time t , the evaluation examines the agent's management of the situation throughout the defined time horizon to determine effectiveness. For practical reasons within this competition, assessment is restricted to events affecting the n_l designated attackable lines, rather than the entire set of transmission lines.

In the scoring mechanism, incorrect alerts result in penalties. However, the Assistant Score is adjusted relative to the total number of relevant events encountered, so that correct anticipatory alerts are proportionally credited. As a result, each event occurring at time t_k on line l is assigned an associated trust score:

$$c_{trust}(Ev(t_k)) = \begin{cases} 0, & \text{if } noalert(t_k - 1, l) \text{ and } no_blackout([t_k, t_k + 12]) \\ 0, & \text{if } alert(t_k - 1, l) \text{ and } blackout([t_k, t_k + 12]) \\ -1, & \text{if } alert(t_k - 1, l) \text{ and } no_blackout([t_k, t_k + 12]) \\ -50, & \text{if } noalert(t_k - 1, l) \text{ and } blackout([t_k, t_k + 12]) \end{cases} \quad (4)$$

This mechanism rewards the agent for issuing accurate pre-failure warnings, while simultaneously discouraging the generation of excessive or unwarranted alerts.

Given the presence of n such events (30 on average across a full scenario) within episode e , the total accumulated score is computed as follows:

$$c_{trust}(e) = \sum_{k=1}^n c_{trust}(Ev(t_k)) \quad (5)$$

The normalized aggregate score of an episode averages 0 for an agent that completes the scenario while expressing random confidence. It reaches -100 for an agent that experiences a blackout without earning any positive confidence points. Conversely, it attains 100 for an agent that completes the scenario while remaining fully confident in its ability to handle attacks. During calibration, care was taken to avoid undesirable incentives—specifically, strategies that would deliberately fail early and yield higher rewards than finishing the scenario. An agent failing early incurs a substantial penalty in the operational score. If failure occurs late, the assistant's score does not exceed what could have been achieved by completing the scenario confidently.

Results and Discussion

General results

For this analysis, only each participant's final submission—those displayed on the leaderboard—was considered. The outcomes of the final phase are summarized in [Figure 6](#), and the leaderboard is shown in [Table 1](#). (The leaderboard and legacy phase remain accessible at a URL hidden for anonymization.) The top 3 participants achieved comparably strong performance across the 3 scoring dimensions. This indicates that the incentive structure effectively encouraged balanced optimization, although none reached the maximum in any category, particularly in operational and assistant scores. The relatively lower results of other active participants further demonstrate the challenge of attaining such levels. Throughout the competition, the top 3 teams initially focused on enhancing operational performance while preserving a strong low-carbon score. After stabilizing around 60 points, they concentrated on improving the assistant score. This progression reflects the intended balance between operational excellence and collaborative intelligence embedded in the design.

The test dataset spans four years (208 scenarios). Each scenario covers a week, with 5-minute intervals between steps, yielding 2016 time steps. The large number of scenarios strengthens the robustness of the evaluation. Variability across scenarios is illustrated by error bars in [Figure 6](#).

Two forms of error bars were computed to examine score dispersion. The “scenario STD” corresponds to the standard deviation across distinct test scenarios, capturing performance variability between them. The “bootstrap STD” is derived

from repeated bootstrap sampling of the scenario set, indicating intrinsic variability in total score estimation. As the number of scenarios increases, the “bootstrap STD” decreases, emphasizing the value of a larger evaluation set for stability and reliability. In contrast, the “scenario STD” remains high in both competitions. Despite extensive chronics and a robust framework, the top 3 participants remain closely matched. This tight competition likely reflects the substantial reward, motivating teams to maximize performance.

Table 1. Final leaderboard of the competition test phase. Submissions were executed on 4 AMD vCPUs at 2.85 Hz and 16GB RAM workers to assess computation time.

Team	Score	Operational score	NRES score	Assistant score	Total duration (s)	Max decision Time	99% Decision time	95% Decision time	Avg simulations per step
Artelys	64.96	61.52	85.10	61.13	26 873.90	494.31	0.18	0.08	191.81
N-side	64.11	61.93	98.10	48.97	29 408.59	94.57	0.94	0.16	7.46
La Javaness	60.82	59.44	90.14	46.55	24 026.97	11.15	2.60	0.08	77.92
AIR	10.45	2.91	100.00	−31.21	9 388.53	1.39	0.08	0.08	333.36
Hippotilt	9.03	5.17	100.00	−36.28	21 448.83	2 843.49	0.01	0.01	0.12
DavidBenhaïem	−0.38	22.00	−88.19	−1.38	37.31	0.01	0.00	0.00	0.00

Independent yearly chronics were generated separately for the feedback and final phases. Participant performance across both phases showed strong correlation, with no evidence of overfitting to the development phase, highlighting the competition's effectiveness in assessing generalization and robustness.

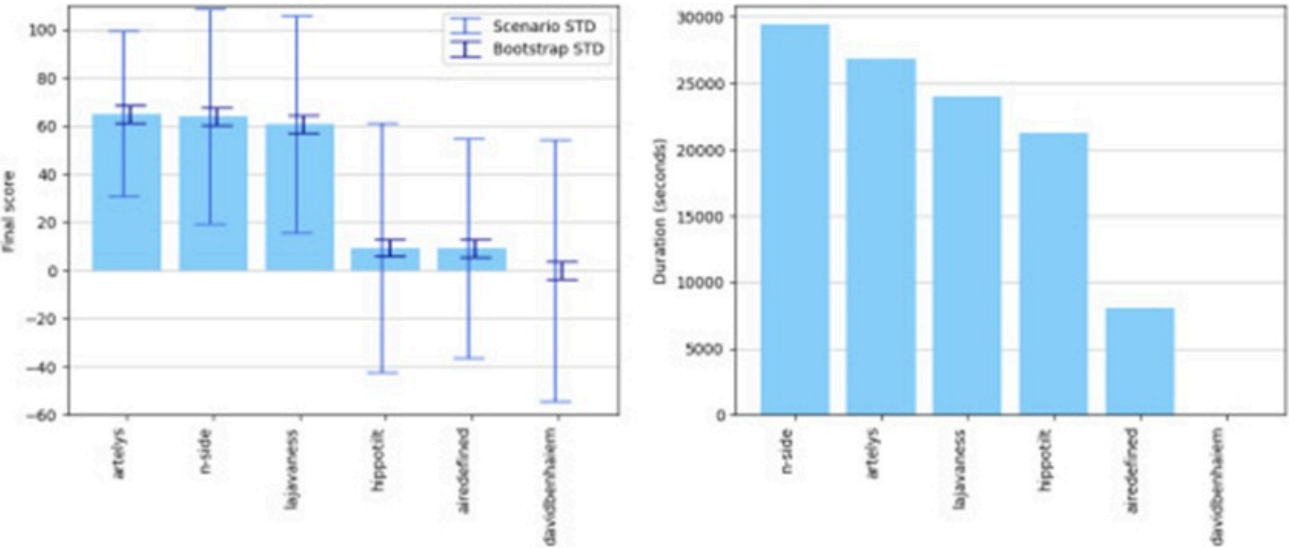


Figure 6. Primary results of the final phase (left) and corresponding submission durations (right). denotes deviation across all scenarios, while reflects repeated bootstrap-based estimates of deviation.

Although the scenarios presented offer meaningful insights, certain limitations affect broader generalization. First, scalability to larger or more intricate grids remains unexplored, suggesting that multi-agent coordination across zones may be required. Second, the disparity between simulated and real-world conditions warrants further investigation to ensure resilience in the face of practical uncertainty. Third, the framework assumes a static grid topology, whereas real power systems are increasingly dynamic and demand adaptive control strategies.

Description of the top-performing solutions

This section details the methodologies adopted by Artelys, N-side, and La Javaness—the three highest-ranked teams. Other participants predominantly relied on fully automated end-to-end learning without incorporating expert knowledge, an ambitious strategy that did not yield competitive results.

The top 3 teams structured their agents into coordinated modules:

- A heuristic rules module for reconnecting outaged lines, avoiding unnecessary interventions when the grid is stable and maximum line loading remains below a defined threshold, and restoring secure topological configurations.
- An advanced topological decision module with distinct characteristics, each leveraging the multi-area grid structure.
- An optimization module for continuous redispatch and curtailment actions, typically based on Optimal Power Flow formulations inspired by the provided CVXPY baselines. This module was triggered when no sufficiently effective topological action was identified.
- An alert module combining heuristic reasoning with simulation.

Each team systematically used the simulation interface available to the agent to verify the reliability of each action before execution. The following section elaborates on the distinguishing aspects of each team's topological module.

Topological modules

Both Artelys and La Javaness structured their developments around the curriculum learning agent baseline [27], which follows a 4-stage student–teacher paradigm: (1) exploration of actions and reduction of the action space, (2) creation of an {observation, action} dataset derived from a simple expert operating within the reduced space, (3) training of a supervised ML behavior cloning (BC) agent using this dataset, and (4) initialization of an RL agent from the pretrained ML model. However, neither team succeeded in producing an RL agent that outperformed alternatives within this demanding environment.

Artelys successfully trained and deployed a BC agent, but relied on a simplified action space that excluded the most extensive source of actions—substation 48, which alone provides 60,000 unitary actions. Since this configuration did not deliver sufficient performance on its own, they incorporated a stochastic exploration component during inference to leverage the excluded actions. While this strategy improved the score, it also substantially increased the maximum computation time and introduced reproducibility concerns due to random actions.

La Javaness implemented an efficient action-space reduction strategy that enabled higher scores when combined with multi-area reasoning and a greedy search mechanism. Although initial BC experiments were promising, they did not achieve this level of performance and were therefore omitted from the final submission. Importantly, they maintained decision times within acceptable bounds. Their solution is publicly accessible at: <https://github.com/lajavaness/l2rpn-2023-ljn-agent>.

In contrast, N-side adopted a physics-driven optimization strategy without incorporating ML. Their approach relies on linearized topological mathematical operators computed around a given operating point whenever a topology is applied to

a specific grid state. These operators can then be reused efficiently for similar grid states. This method demonstrated competitive score and timing performance for a purely model-based strategy, though scalability may become challenging as operating conditions grow exponentially.

Alert module

Concerning the alert mechanism and assistant score, the leading teams primarily implemented tailored heuristic rules, often targeting particular attacks or seasonal patterns. Artelys also employed a planning horizon based on available forecasts to evaluate the risk associated with the two most challenging line attacks. Although supervised learning methods for alert prediction were tested, they were not successfully integrated due to time constraints. This avenue remained open for further exploration and was investigated after the competition.

Given the minimal score differences among the 3 leading teams, La Javaness was ultimately declared the winner based on the decision-time criterion.

Post-Challenge Analysis

This section provides a comprehensive overview of the winning methodology and presents a deeper post hoc analysis of the agent. It concludes with the proposal of an enhanced alert module based on supervised learning.

Detailed description of winning approach design

The winning solution developed by La Javaness draws heavily from established baselines, particularly the CVXPY baselines and heuristic-based agent from the l2rpn baselines package (<https://github.com/rte-france/l2rpn-baselines>), as well as the curriculum agent [27]. The architecture integrates multiple decision modules coordinated through high-level operational rules derived from line loading at time t and simulated loading at time $t+1$. The overall structure is illustrated in **Figure 7**.

Previous baselines from earlier challenges, along with most Grid2Op research, typically emphasize either topological adjustments [27] or redispatch/curtailment strategies [33], sometimes supplemented with heuristics. The novelty of this contribution lies in effectively coordinating all action types while preserving low decision latency, enabling scalability to the largest Grid2Op environment to date. Earlier studies primarily addressed smaller grid instances; here, these techniques demonstrate scalability to more complex systems characterized by increased size, higher renewable penetration, maintenance processes, adversarial attacks, and storage integration. Additionally, incorporating a dedicated alert module enhances transparency regarding the agent's confidence in its actions.

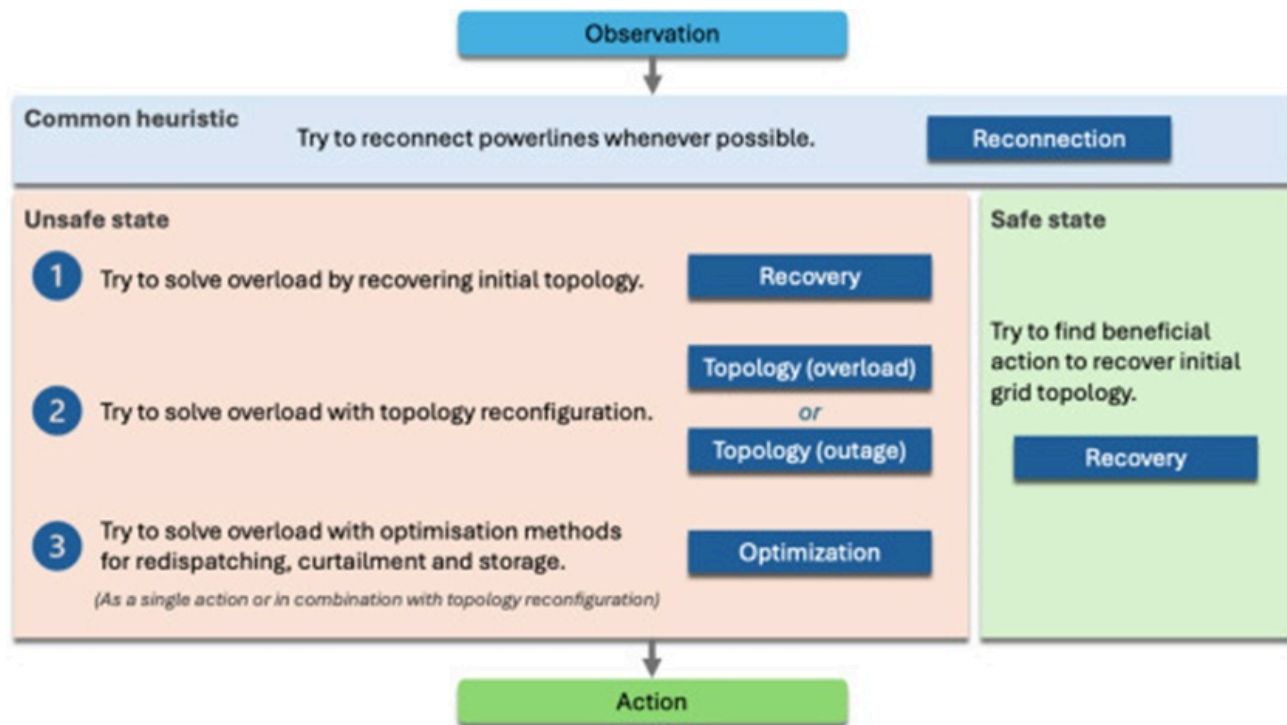


Figure 7. La Javaness agent architecture. Dark blue boxes represent independent decision-making modules. Their resulting actions are merged into a final system action, validated through simulation (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

While reconnection and recovery components rely on straightforward heuristics, the agent's performance largely depends on a carefully curated reduced action space of unitary topological operations under two distinct conditions. This reduction was achieved through exhaustive simulation and greedy exploration across the full action space, leveraging insights from the curriculum agent framework, particularly the teacher-stage data collection methodology [27]. The two scenarios are defined as follows:

- **Overload case:** All lines are active, and at least one line is overloaded with $p_{\max} \geq 1.0$. Here, p denotes the vector of ratios between observed current flow and each line's thermal limit; $p_{\max}$ is the maximum element, corresponding to the most heavily loaded line. Exhaustive simulations across the entire unitary action space during overload events generated more than 100,000 observation–action samples. Actions were retained if they significantly reduced $p_{\max}$ below the safety threshold ($p_{\max} < 0.95$), and only those selected more than 50 times during simulation were preserved.
- **Outage case:** At least one line is overloaded, and at least one line is disconnected due to an attack, failure, or scheduled maintenance. This action set was similarly constructed through exhaustive simulations. Attackable lines were periodically disconnected to emulate outages, and actions were filtered based on their capacity to mitigate overloads, using a slightly relaxed threshold ($p_{\max} < 0.97$). This module operates as an independent topology-management component.

These modules shrink the overall unitary action space—from 73,000 possible actions—down to two compact subspaces, each limited to just a few hundred unitary actions. Together, these subspaces prove sufficient to handle the vast majority of challenges encountered on the power grid. In the foundational work on the curriculum agent, unitary actions gathered across various scenarios were merged into one unified reduced action space [27]. In contrast, La Javaness opted to maintain separate action spaces for each case, thereby capitalizing on the actions' contextual specificity. This strategy effectively mitigates the curse of dimensionality inherent in the topology reconfiguration task, rendering action prediction via greedy search and deep learning-driven supervised learning far more tractable. Moreover, it only necessitates straightforward heuristic rules to select the appropriate action space (Figure 8).

It should be emphasized that this procedure is highly computationally demanding, requiring substantial hardware resources and continuous execution over multiple days. Nevertheless, through extensive experimentation with varied random seeds, hyperparameter adjustments, and evaluations of diverse action-space configurations, two highly effective reduced subspaces were successfully constructed. Furthermore, compatibility with the hyperparameters of other components—particularly the optimization module—was thoroughly assessed across different action-space designs. As a result, the agent achieves strong synergy among its coordinated modules, delivering consistently high and stable performance across the entire dataset.

Distribution of substation ids reconfiguration with respect to sub-areas available in reduced action spaces

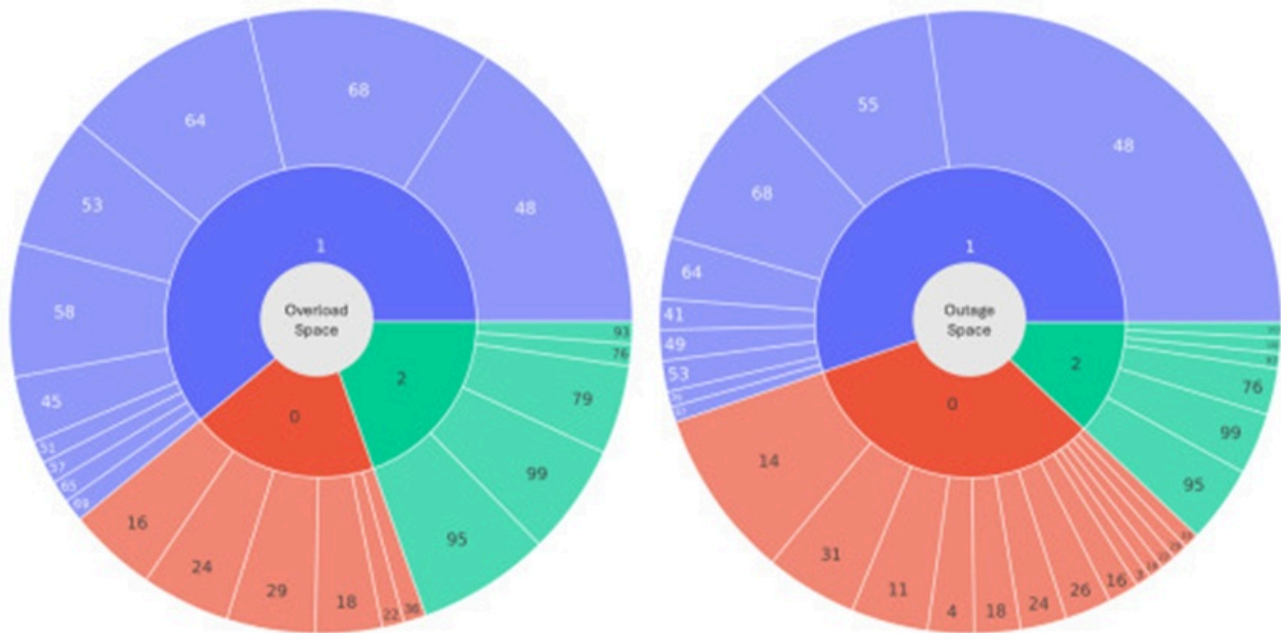


Figure 8. Distribution of feasible reconfigurations per substation, segmented by sub-areas of the power grid, for overload scenarios (left, 87 actions) and outage scenarios (right, 125 actions). The inner ring indicates the zones, while the outer ring lists substation identifiers. Each action subspace exhibits slightly distinct patterns regarding affected zones and substations.

By drastically reducing the number of topology reconfiguration options available to the agent, the agent's overall decision-making latency remains low. Even with greedy search techniques, the maximum decision time remains under 18 seconds. Leveraging these compact action spaces, La Javaness successfully trained corresponding Actor-Critic policies using the Behavior Cloning algorithm in a supervised manner. Behavior Cloning represents the simplest form of Imitation Learning, in which an autonomous agent acquires skills through supervised (or semi-supervised) training. In this setup, a dataset comprising trajectories (observation–action pairs) is created by applying greedy search within the reduced unitary topology reconfiguration space. Subsequently, a neural network is trained via supervised classification to predict the appropriate action from a given system observation. Generating this large volume of observation–action pairs is resource-intensive when greedy search is performed over a broad action space. An illustrative training curve, including loss and accuracy, appears in **Figure 9**. Top-k accuracy serves as the primary evaluation metric, quantifying how often the correct action ranks among the top k predicted actions (ordered by predicted probability).

Performance scores and overall results show no substantial difference between the supervised-learned policy and the agent that conducts exhaustive greedy search in the reduced action space. Nevertheless, the learned policy yields a dramatic reduction in computation time, lowering the maximum decision time to below 8 s and bringing the 99% decision time to approximately 0.7 s. The trained policies attain comparable outcomes by forecasting the top-k actions (sorted by

score) for a given observation, then selecting the best one from that shortlist. A value of $k = 15$ was adopted as the default, striking an effective balance between performance and computational overhead.

Broadly speaking, raising the k -value to 20 yields solid accuracy (exceeding 80% on the test set). There exists a clear trade-off: larger k increases simulation time and thus decision latency. Even at $k = 15$ or $k = 20$, both model inference and subsequent simulations remain highly efficient, with total computation time staying below 1 s.

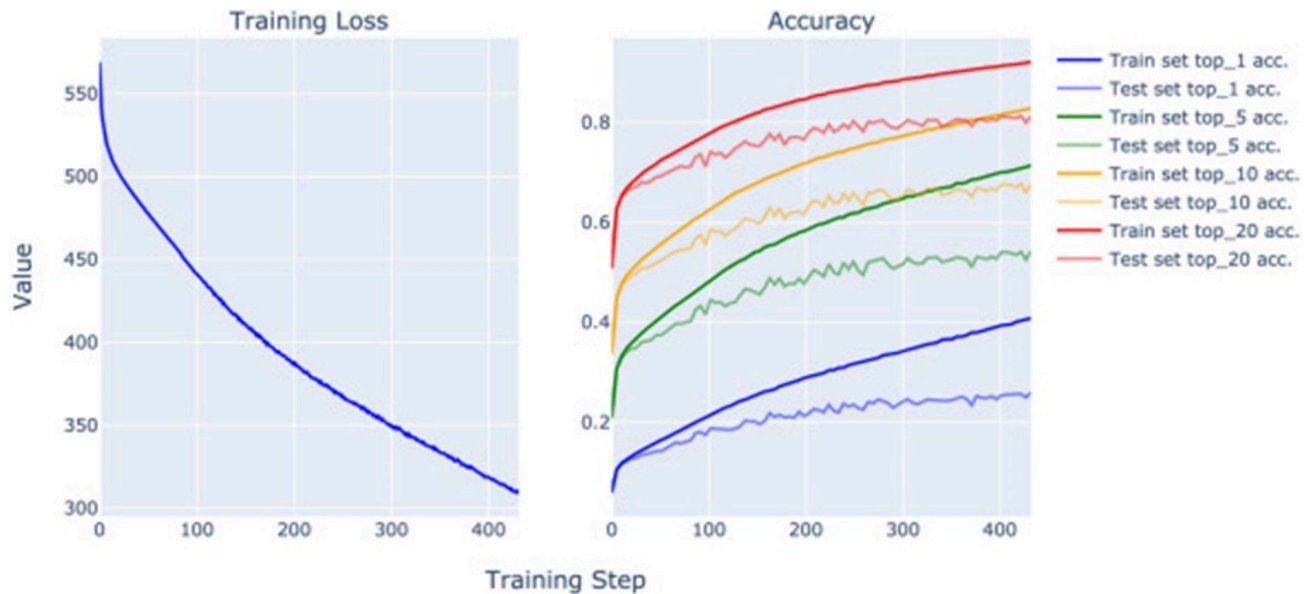


Figure 9. Evolution of training loss (left) and top- k accuracy on training and test sets (right) throughout the training process. An early-stopping mechanism based on validation loss was employed. Clear signs of overfitting emerge, as evidenced by the widening gap between training and test accuracy.

In summary, La Javaness devoted considerable effort to investigating a wide range of techniques, with a particular focus on deep and reinforcement learning methods. While this investigation deepened insight into autonomous agent-based power grid management and its inherent difficulties, it also exposed the constraints of existing leading approaches—such as the curriculum agent—when applied to highly intricate environments. Notably, the rapid growth of the state–action space in complex control domains continues to pose a formidable barrier that conventional model-free reinforcement learning struggles to overcome.

It is proposed that further gains in the topology module could be achieved with more sophisticated AlphaZero-style reinforcement learning agents [29]. That said, the incremental benefit within the full orchestrated agent could prove relatively modest. Subsequent research could explore this potential performance ceiling and the balance between agent sophistication and practical deployment considerations.

Ablation study

This section describes an ablation study designed to quantify the individual contributions of each decision module within the agent's overall algorithm. Two baseline variants of the La Javaness agent serve as reference points: one that relies on greedy search across the full reduced action spaces, and another that employs neural network-based policies for the topology component. The main configuration—using neural network policies with top-15 action prediction—is then evaluated under multiple ablation conditions (with selected modules disabled) to measure performance variations. Results are summarized in Figure 10. For clarity, the reconnection and recovery modules are collectively labeled as “Heuristics,” while the topology modules for overload and outage cases are grouped.

The ablation study reveals several key insights:

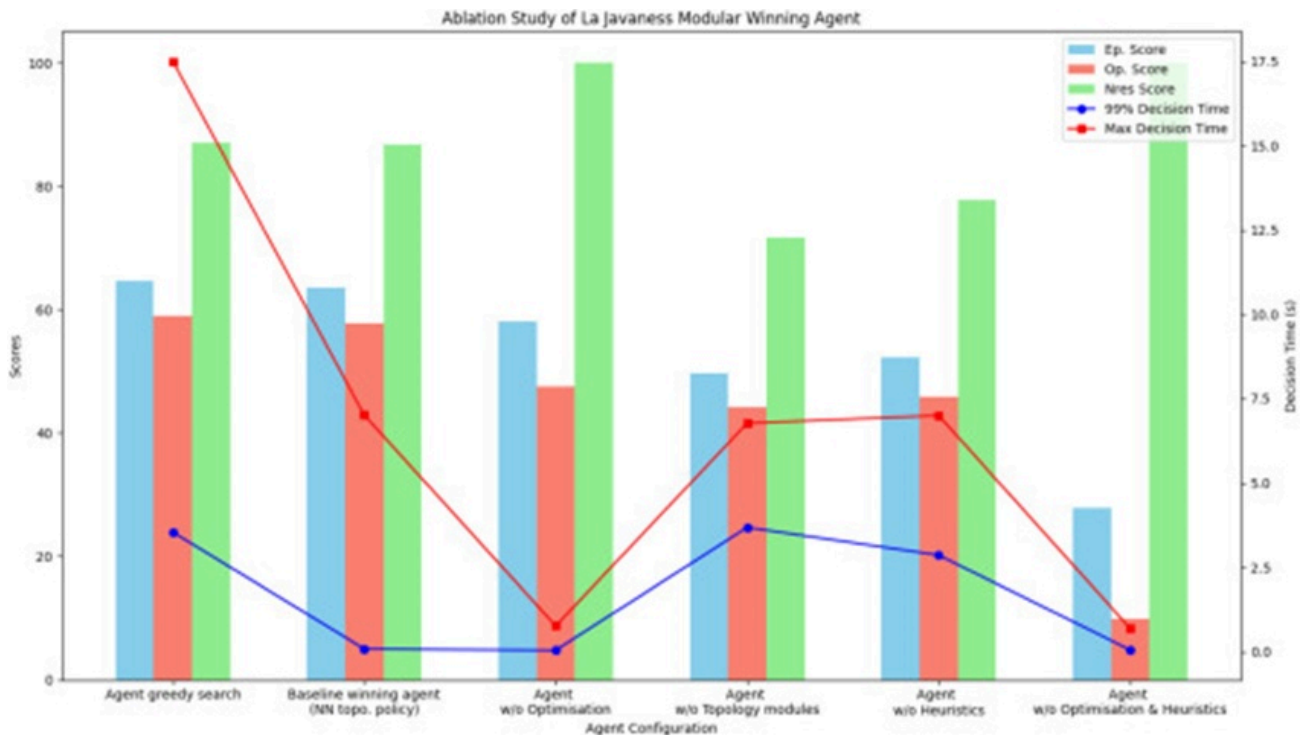


Figure 10. Ablation study conducted on the modular winning agent, assessing six distinct configurations: (1) greedy search only, (2) baseline winning agent equipped with a neural network-based topology policy, (3) without the optimization module, (4) without the topology modules, (5) without the heuristics modules, and (6) without both optimization and heuristics modules. Performance is evaluated using the episode score (Ep. Score), defined as the weighted average of the 80% operational score (Op. Score) and the 20% renewable energy score (Nres Score), as shown by the bars. Decision-time metrics—99% decision time (blue line) and maximum decision time (red line)—are overlaid for comparison. The analysis reveals clear trade-offs between agent performance and decision latency when individual components are removed, pinpointing key time bottlenecks in the decision process (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Ablating any single one of the three primary modules does not lead to a dramatic drop in overall scores. Removing the optimization module, for instance, enables the agent to achieve the maximum possible renewable energy score (Nres = 100.0), since it then relies exclusively on topology-based actions. Conversely, turning off the topology modules forces greater dependence on the optimization module, resulting in a substantial reduction in the Nres score and demonstrating the necessity of curtailment actions to manage grid overloads.

- Eliminating both the optimization and heuristics modules—leaving the agent to operate solely with unitary topological actions—severely degrades the operational score, primarily due to a sharp rise in blackout events (game overs). This underscores the critical value of integrating and leveraging diverse action types within such a highly complex environment.

- Examination of the decision-time statistics identifies two primary bottlenecks:

- o 1. The quantity of topological reconfigurations simulated within the topology modules must be minimized. Here, neural network policies trained via behavior cloning enable prediction of the top-15 actions in under 1s.

o 2. The optimization module constitutes a secondary bottleneck, occasionally requiring up to 8 s to derive the optimal solution within the challenge environment.

- Taken together, the results indicate that disabling the optimization module yields better outcomes for both operational and Nres scores compared with turning off the topology modules. This suggests that the carefully curated reduced action spaces remain highly effective across most overload scenarios encountered on the grid. Moreover, adopting imitation learning via behavior cloning enables real-time decision-making.

In summary, this ablation study underscores the robustness of the winning agent and its capacity to seamlessly integrate multiple action categories while maintaining very low decision latency. These findings position the approach as highly promising for extension to more realistic power grid control problems.

Learning an improved alert module

All leading teams adopted rule-based strategies to issue alerts whenever the grid was projected to violate the “N-1 criterion” and thus become unsafe. In addition to evaluating their own methods, we explored training a supervised learning-based alert module to determine whether it could deliver score improvements. The objective was to estimate, for each of the 21 critical lines, the probability that its disconnection would trigger a blackout within the subsequent hour. An alert would then be triggered if this probability exceeded a predefined threshold.

Training data were generated as follows:

- Run a grid-controlling agent
- Sequentially disconnect each critical network element and simulate forward one hour
- Record whether a blackout occurred (label = 0) or the grid remained stable (label = 1)

This constitutes a multilabel classification task: at every time step t , predict a binary outcome (0 or 1) independently for each of the 21 contingencies.

The alert module was specifically trained to evaluate the safety of actions taken by the La Javaness agent; accordingly, that agent served as the grid controller during data collection. A fully connected neural network with two hidden layers of 256 neurons each was employed. The input features included:

- Loading levels of each transmission line
- Scheduled upcoming maintenance on the 21 critical lines
- Topological distance from the initial grid configuration
- Cumulative duration of overloads on each line
- Datetime-related features

Attempts to incorporate measured and forecasted load and generation per grid area produced negligible gains.

Since the agent typically prevents blackouts even under contingencies, the dataset's label distribution is heavily imbalanced. Moreover, certain lines proved far more prone to causing failures than others, resulting in extremely sparse positive examples (blackouts) for some contingencies. To counteract this bias, failure samples were assigned higher weights in the loss function (MSE), with class-specific weights computed according to the failure rate per contingency in

the training data. Greater emphasis was placed on accurately detecting blackouts (true negatives) than on correctly identifying safe scenarios (true positives). The final classification threshold was tuned on the validation set, prioritizing true negative performance.

This supervised learning approach produced a marked improvement in Assistant score—approximately 66% higher than La Javaness' original alert module and about 26% higher than the top-scoring Artelys agent. Detailed scores appear in [Table 2](#). True positive and true negative rates for our classifier are reported in [Table 3](#), corresponding to correct predictions of grid survival and blackout, respectively. Enlarging the training dataset from 48k to 252k samples yielded substantial classification gains ($\approx +3\%$ true negative rate, $\approx +2\%$ true positive rate), with test and validation sets each containing 32k samples ([Figure 11](#)). We anticipate that this performance trend would persist—and potentially strengthen—with even larger datasets, particularly for contingencies that rarely fail.

In conclusion, these experiments demonstrate that machine learning techniques can effectively predict grid safety under contingencies—an approach that merits further investigation. The formulation of the Assistant score proved well-suited to learning-based agents, yielding both scientifically valuable insights and practically relevant improvements in grid operation.

Table 2. Comparison of Assistant scores

	Learning alert module	La javaness	N-side	Artelys
Assistant score	77.1	46.55	48.97	61.13

Supervised learning-based alert module versus the rule-based alert strategies employed by the competing teams. The learned alert module was specifically trained to forecast grid survival under the actions of the La Javaness agent, thereby directly improving this component's performance. In contrast, the alert modules of N-side and Artelys were designed to predict outcomes based on their respective grid controllers. Given the close similarity in operational scores achieved by all three agents, the Assistant score comparisons remain meaningful and fair.

Table 3. True positive rate (correctly predicting grid survival) and true negative rate (correctly predicting blackout) for each of the 21 critical line disconnections, reported as the average across all lines, along with the minimum and maximum values.

Metric	Mean	Minimum	Maximum
True positive rate (correctly predicting grid survival)	93.9%	89.6%	97.3%
True negative rate (correctly predicting blackout)	96.3%	89.7%	98.6%

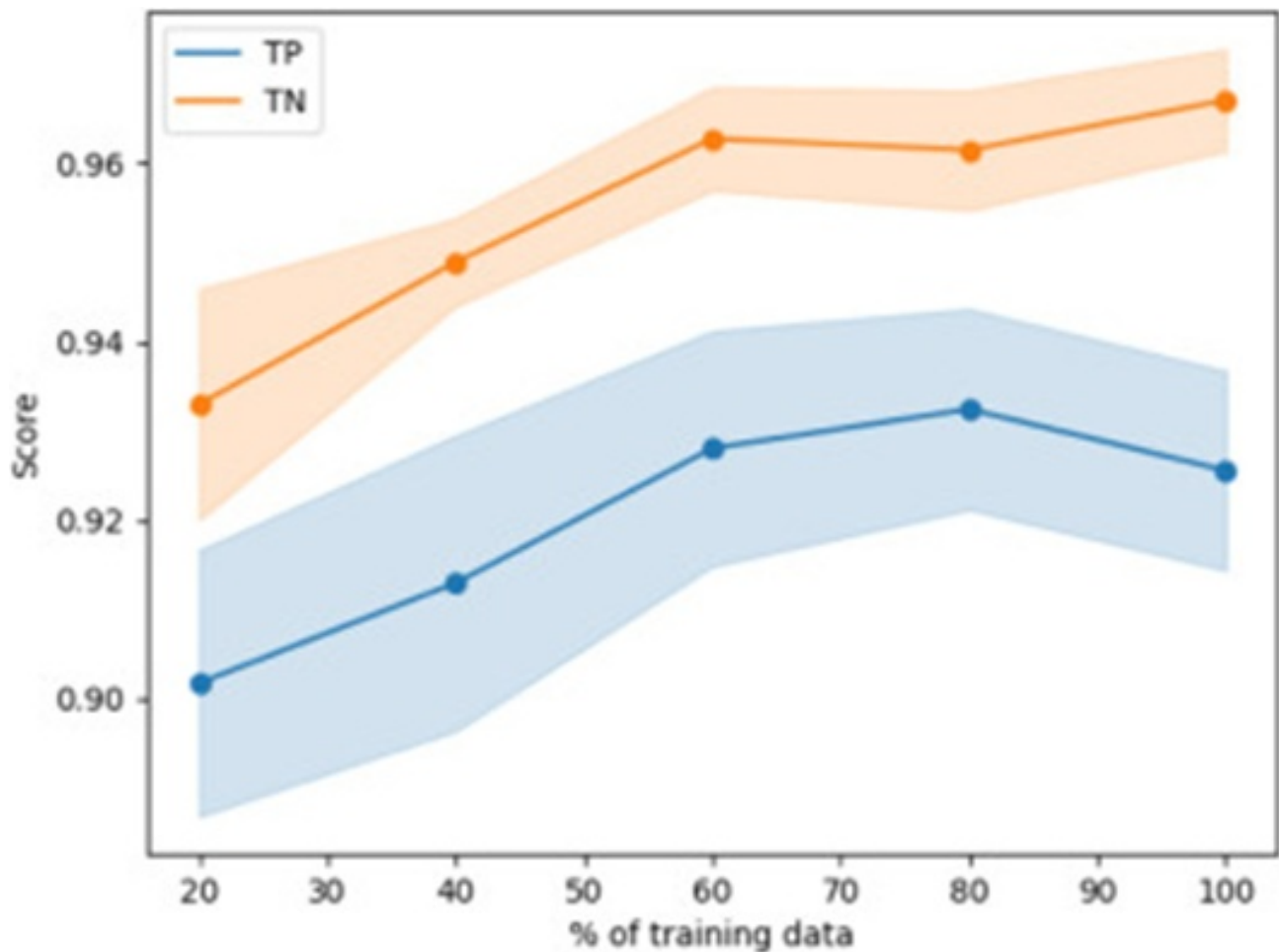


Figure 11. True positive (TP) and true negative (TN) rates for predicting grid survival following disconnection of one of the 21 critical lines, plotted against training set size and averaged over the 21 contingency classes. Shaded regions indicate the minimum and maximum values observed across classes. 100% of the training data corresponds to 252,000 training samples.

Conclusion

The outcomes of this competition illustrate both the remarkable promise and the present-day constraints of artificial intelligence—especially reinforcement learning (RL)—in controlling power grids amid the ongoing energy transition. A key takeaway from the top-performing entries, particularly those of La Javaness, Artelys, and N-side, is the critical role played by a carefully balanced integration of heuristic rules, optimization algorithms, and RL-driven components in preserving grid stability, curbing carbon emissions, and preventing blackouts. This modular, hybrid strategy closely reflects the intricate nature of actual energy infrastructures, where relying on any single technique proves inadequate for tackling the diverse and dynamic challenges introduced by renewable energy variability.

La Javaness's winning approach stood out by effectively leveraging a carefully selected reduced action space, enabling efficient resolution of overload conditions while ensuring decision times remained practically viable. This underscores the value of intelligent action-space reduction in extremely high-dimensional problems. At the same time, a detailed analysis of the results and agent behaviors revealed that state-of-the-art RL techniques—including curriculum learning and behavior cloning—still fall considerably short of human operator reliability, especially in highly uncertain or rapidly changing grid states. The development of a superior alert module during the competition further emphasized this

limitation, as the strongest solutions incorporated forward-looking alerts whenever AI confidence waned—an essential mechanism for fostering trust in AI-supported decision systems.

Insights gained after the competition also suggest several fruitful avenues for future work. The substantial improvement in the Assistant score achieved through the supervised alert module indicates that hybrid methodologies that merge rule-based logic, optimization methods, and data-driven learning hold strong potential for creating more dependable AI agents. Additionally, the clear performance gains from simply increasing the volume of training data for the alert module highlight the importance of abundant, diverse datasets in enhancing AI robustness within this field.

Although the simulation environment was crafted to closely approximate real-world grid dynamics, subsequent research should prioritize scalability to significantly larger networks, resilience under evolving grid configurations, and careful examination of the sim-to-real gap. Recognizing these hurdles emphasizes the need for rigorous evaluation across a broad spectrum of operating regimes, extended network layouts, multi-agent coordination scenarios, and eventual real-world deployments to validate and strengthen the proposed methods.

Looking ahead, the competition underscores the urgent need for continued progress toward more scalable, flexible, and trustworthy AI systems. The power grid—with its demanding real-time requirements, reliance on intermittent renewables, and severe safety implications—stands as a critical domain where AI must continue to mature in both performance and dependability. As these models become increasingly embedded in essential infrastructure, the findings from this challenge will help guide the development of next-generation decision-support tools that effectively reconcile operational efficiency, environmental goals, and appropriate human supervision.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 05 Oct 2021

Revised: 11 Nov 2021

Accepted: 09 Jan 2022

Published online: 20 March 2022

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Pörtner HO, Roberts D, Tignor M, Poloczanska E, Mintenbeck K, Alegría A, et al. Climate change 2022: Impacts, adaptation, and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press; 2022.
- RTE France. Futurs énergétiques 2050 principaux résultats. 2021. https://assets.rte-france.com/prod/public/2021-10/Futurs-Energetiques-2050-principaux-resultats_0.pdf.
- Donon B. Deep statistical solvers & power systems applications. PhD thesis. Université Paris-Saclay; 2022. <https://tel.archives-ouvertes.fr/tel-03624628>.
- Donnot B. Deep learning methods for predicting flows in power grids: novel architectures and algorithms. PhD thesis. Université Paris Saclay (COMUE); 2019. <https://tel.archives-ouvertes.fr/tel-02045873>.
- Prostejovsky AM, Brosinsky C, Heussen K, Westermann D, Kreusel J, Marinelli M. The future role of human operators in highly automated electric power systems. *Electr Power Syst Res*. 2019;175:105883.
- Marot A, Kelly A, Naglic M, Barbesant V, Cremer J, Stefanov A, et al. Perspectives on future power system control centers for energy transition. *J Mod Power Syst Clean Energy*. 2022;10(2):328-44.
- Marot A, Donnot B, Dulac-Arnold G, Kelly A, O'Sullivan A, Viebahn J, et al. Learning to run a power network challenge: a retrospective analysis. *CoRR*. 2021;abs/2103.03104. <https://proceedings.mlr.press/v133/marot21a/marot21a.pdf>.
- Marot A, Rozier A, Dussartre M, Crochepierre L, Donnot B. Towards an AI assistant for power grid operators. In: *Augmenting Human Intellect*. HHAi 2022. IOS Press; 2022:79-95.
- Donnot B. Grid2op: A testbed platform to model sequential decision making in power systems. 2020. <https://github.com/rte-france/grid2op>.
- Brockman G, Cheung V, Pettersson L, Schneider J, Schulman J, Tang J, et al. OpenAI Gym. 2016. <https://doi.org/10.48550/ARXIV.1606.01540>.
- Towers M, Kwiatkowski A, Terry J, Balis JU, De Cola G, Deleu T, et al. Gymnasium: A standard interface for reinforcement learning environments. 2024. *arXiv:2407.17032*.
- Babaeinejadsarookolae S, Birchfield A, Christie RD, Coffrin C, DeMarco C, Diao R, et al. The power grid library for benchmarking AC optimal power flow algorithms. 2019. *arXiv:1908.02788*.
- Huang Q, Huang R, Hao W, Tan J, Fan R, Huang Z. Adaptive power system emergency control using deep reinforcement learning. *IEEE Trans Smart Grid*. 2019;11(2):1171-82.
- Cui H, Zhang Y. Andes_gym: A versatile environment for deep reinforcement learning in power systems. In: *2022 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE; 2022. p.1-5.
- Yeh C, Li V, Datta R, Arroyo J, Christianson N, Zhang C, et al. SustainGym: Reinforcement learning environments for sustainable energy systems. *Adv Neural Inf Process Syst*. 2024;36.

- Eichelbeck M, Markgraf H, Althoff M. CommonPower: Supercharging machine learning for smart grids. 2024. arXiv:2406.03231.
- Biagioni D, Zhang X, Wald D, Vaidhynathan D, Chintala R, King J, et al. PowerGridworld: A framework for multi-agent reinforcement learning in power systems. In: Proceedings of the Thirteenth ACM International Conference on Future Energy Systems (E-Energy '22). ACM; 2022. p. 565-70.
- Henri G, Levent T, Halev A, Alami R, Cordier P. Pymgrid: An open-source Python microgrid simulator for applied artificial intelligence research. 2020. arXiv:2011.08004.
- Henry R, Ernst D. Gym-ANM: Reinforcement learning environments for active network management tasks in electricity distribution systems. *Energy AI*. 2021;5:100092.
- Heid S, Weber D, Bode H, Hüllermeier E, Wallscheid O. OMG: A scalable and flexible simulation and testing environment toolbox for intelligent microgrid control. *J Open Source Softw*. 2020;5(54):2435.
- Marot A, Donnot B, Romero C, Donon B, Lerousseau M, Veyrin-Forrer L, et al. Learning to run a power network challenge for training topology controllers. PSCC2020 (preprint). 2020.
- Marot A, Donnot B, Dulac-Arnold G, Kelly A, O'Sullivan A, Viebahn J, et al. Learning to run a power network challenge: A retrospective analysis. NeurIPS 2020 competition and demonstration track. PMLR; 2021. p. 112-32.
- Lan T, Duan J, Zhang B, Shi D, Wang Z, Diao R, et al. AI-based autonomous line flow control via topology adjustment for maximizing time-series ATCs. IEEE PES General Meeting. 2020:1-5.
- Marot A, Donnot B, Chaouache K, Kelly A, Huang Q, Hossain RR, et al. Learning to run a power network with trust. *Electr Power Syst Res*. 2022;212:108487.
- Omnes L, Marot A, Donnot B. Adversarial training for a continuous robustness control problem in power systems. IEEE Madrid PowerTech. 2021:1-6.
- Marot A, Donnot B, Tazi S, Panciatici P. Expert system for topological remedial action discovery in smart grids. MEDPOWER 2018. 2018. p. 1-6.
- Lehna M, Viebahn J, Marot A, Tomforde S, Scholz C. Managing power grids through topology actions: A comparative study between advanced rule-based and reinforcement learning agents. *Energy AI*. 2023;14:100276.
- AH. Zhou B, Zeng H, Liu Y, Li K, Wang F, Tian H. Action set based policy optimization for safe power grid management. In: Machine learning and knowledge discovery in databases. Applied data science track: European conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, proceedings, part V 21. Springer; 2021:168-81.
- Dorfer M, Fuxjäger AR, Kozak K, Blies PM, Wasserer M. Power grid congestion management via topology optimization with AlphaZero. 2022. arXiv:2211.05612.
- Van Der Sar E, Zocca A, Bhulai S. Multi-agent reinforcement learning for power grid topology optimization. 2023. arXiv:2310.02605.
- Bellman R. A Markovian decision process. *J Math Mech*. 1957;6(5):679-84. <http://www.jstor.org/stable/24900506>.
- Pena I, Martinez-Anido CB, Hodge BM. An extended IEEE 118-bus test system with high renewable penetration. *IEEE Trans Power Syst*. 2017;33(1):281-9.
- Han X, Mu C, Yan J, Niu Z. An autonomous control technology based on deep reinforcement learning for optimal active power dispatch. *Int J Electr Power Energy Syst*. 2023;145:108686.