Learning to run a power network with trust

Marot, Antoine; Donnot, Benjamin; Chaouache, Karim; Kelly, Adrian; Huang, Qiuhua; Hossain, Ramij Raja; Cremer, Jochen L.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Learning to run a power network with trust

Antoine Marot [a],[*], Benjamin Donnot [a], Karim Chaouache [a], Adrian Kelly [b], Qiuhua Huang [c],
Ramij-Raja Hossain [d], Jochen L. Cremer [e]

[a] *RTE AI Lab, Paris, France*
[b] *EPRI, Ireland*
[c] *PNNL, USA*
[d] *Iowa State University, USA*
[e] *TU Delft, Netherlands*

ARTICLE INFO

ABSTRACT

Artificial agents are promising for real-time power network operations, particularly, to compute remedial actions for congestion management. However, due to high reliability requirements, purely autonomous agents will not be deployed any time soon and operators will be in charge of taking action for the foreseeable future. Aiming at designing assistant for operators, we instead consider humans in the loop and propose an original formulation. We first advance an agent with the ability to send to the operator alarms ahead of time when the proposed actions are of low confidence. We further model the operator's available attention as a budget that decreases when alarms are sent. We present the design and results of our competition "Learning to run a power network with trust" in which we evaluate our formulation and benchmark the ability of submitted agents to send relevant alarms while operating the network to their best.

1. Introduction

Power network operators are in charge of maintaining a reliable, secure supply of electricity at all times. A vast majority of the real-time operation and control decisions are made by human operators based on their experiences, and predefined operation rules and manuals. However, real-time decision-making is getting more challenging as the human operator has to deal with more information, more uncertainty, more applications and more coordination [1]. Recent power outages such as the Texas power outages in early 2021 clearly showed that human operators faced daunting challenges in dealing with rare events and they desperately need intelligent decision-support tools to help make fast and robust decisions to safeguard the network. The ability to foresee events ahead of time is also vital to the future operation of the power system, given inherent future variability.

Human operators and AI can be seen as complementary heterogeneous intelligence that could achieve a superior outcome when combined [2]. In the future, the human may supervise automation, with artificial agents as so-called assistants, monitoring the current system and projecting the forecasted system via simulation. The assistants may propose actions to the operator when issues are identified [3], ultimately having good foresight to securely operate the system [4].

Machine Learning (ML) and Reinforcement Learning (RL) models are showing promise for managing operational reliability [5–7].

ML and RL can propose operating control decisions very quickly, making it suitable for emergency control purposes [8]. Autonomous agents trained with RL are particularly promising as they can reinforce its leanings, even on very complex tasks. Hence, an agent can autonomously improve itself with training simulations, just as human operators adapted their heuristics for their network with experience over years and training on simulated scenarios. It has previously been shown that RL based agents can autonomously improve its own model quality for real-time power network operation management, through the "Learning to run a power network" (L2RPN) competition series. [9].

Starting from our initial baseline [10], winning solutions of these successive L2RPN competitions have progressively improved the operational performance of artificial agents to robustly operate (even under N-1 line disconnections) the network [11–13] as illustrated on Fig. 1 along the *x*-axis.

However, existing AI technology lacks the robustness and trustworthiness that are required for high-consequence, high impact, decision-making in real-time network operation. One key issue is insufficient consideration of leveraging the experience of and working with human teammates (or operators) in existing AI models and applications. Experienced operators in power network deploy extensive domain or
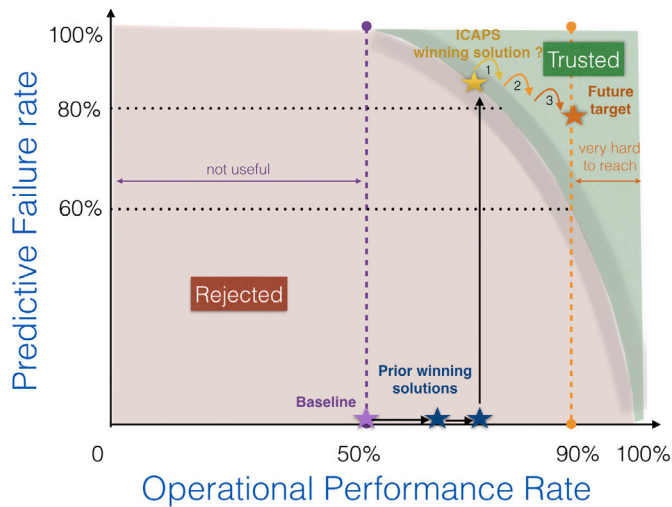
Fig. 1. As a first approximation, trust in an agent can emerge under appropriate levels of operational performance and predict consequence rates. Here we show the expected path of successive L2RPN winning solutions to realistically develop a trusted agent when adding a predict failure feature.

expert knowledge that cannot adequately represented mathematically or easily captured by existing machine learning models [14]. A recent study by MIT researchers showed that state-of-the-art AI agents can become frustrating teammates, and highlighted the need of incorporating subjective metrics such as trust and teamwork into the development of assistants [15] .

Research suggests that an AI agent can increase its trustworthiness by reducing conflicting evidence and by increasing the amount of evidence it has gathered [16,17]. Therefore, based on an imperfect and reinforced model, the assistant proposes actions with varying confidence to reduce conflicting evidence. This is represented through the agent Predictive Failure rate dimension in Fig. 1. This representation makes the low confidence of agents explicit. Working along that direction could eventually make an imperfect agent trustworthy, as the operator will know when to take over. It will also relieve the operator from constant supervision, (hyper-vigilance) which might by physiologically impossible. This also relates to 2-D Hybrid Intelligence diagram target [18] which represents simultaneously high levels of automation and yet high level of human control.

The objective of this paper is to develop an original formulation that incorporates trust-building mechanisms into the process of intelligent assistant learning to securely run the network against various network overloading and physical violation conditions. The framework was tested and demonstrated in the *L2RPN with trust* competition which was organised over the summer 2021. It should be noted that developing an efficient and effective sequential decision making (SDM) formulation is not trivial, but rather critical for obtaining a competent and trustworthy AI assistant for network operators. This is analogous to a novel, complex formulation of the well-known (optimal power flow) OPF problem in power systems [19], close to SCOPF (security constrained OPF) [20] or Multi-period AC SCOPF [21] in particular, but with a human consideration, built in.

The specific contributions of this paper are:

(i) proposing a novel SDM formulation in Section 2 that incorporates human-AI trust-building mechanisms in the design of intelligent assistants. Agents are now given the ability to send interpretable warning to the human—modelled using novel "attention budget" constraints.

(ii) instantiating this concept through the *L2RPN with trust* framework environment in Section 3.

(iii) analysing an open competition results to evaluate this concept design in Section 4 and promising directions.

2. Trust in artificial intelligence (AI)

There are inherent issues with automation of tasks more generally [22], when agents are deployed as assistants to achieve higher efficiencies in managing complex systems. Trust between the human and agent will be difficult to achieve at first as it cannot be varied and be completely lost [23]. Therefore, it seems promising to investigate the very fundamental concept of trust within humans (in this case, operators). This paper investigates whether human operators can develop trust in RL agents to address the issue of missing trust and rigorousness, which currently represents a barrier to their deployment. This idea connects to a broader topic as trustworthiness of AI which is generally believed to be a must-have property for mission-critical applications such as reliability management, in particular context such as vehicle driving or network operations.

The assumption for the proposed trust concept is that humans will trust an agent if they believe that the agent will act in the human's best interest, and accepts vulnerability to the agent's actions (which is adapted from the basic definition of trust [24]). Before a human can trust (an agent), high levels of (i) credibility, (ii) reliability and (iii) intimacy are required according to the Trust Equation (by Charles Green):

(i) the credibility of an agent can increase when the agent is transparent and explains the proposed actions [25]. Credibility is an output of increasing transparency, however, transparency is not always necessary for credibility in the extreme, idealistic case of a perfect agent—for instance. Although trustworthiness should be a property of any explainable model, not every trustworthy model is explainable on its own. As an example, for emergency network control, [26] explains RL actions by providing the human with a series of summary plots.

(ii) trusting an agent requires reliability of the actions. A reliable AI agent should work consistently for the same or similar scenarios that it 'sees' during training with a strong generalisation capability and 'know' the limit of its capabilities. There are two approaches that can be used to quantify the limits of an agent and algorithm, passive or active. In the passive approach, a level of confidence is quantified for each suggested automated action/prediction [27], and the user can act accordingly. The more active approach is to receive a signal of 'low confidence' to actively warn the user. While the nature of the information is the same, the confidence of a proposed action, may have a different impact on building trust between humans and agents. The active approach is usually utilised in automated driving of cars, where the autonomous agent warns the driver to take over under some perceived emergency conditions [28].

(iii) developing intimacy with an agent is needed. Similar to humans, where intimacy grows with the length of a relationship, the life-cycle of an RL can be considered as a whole. For instance, trust, when lost, is difficult to restore. [29] identified how trust can be enhanced in the various stages of an AI-based system's life-cycle, specifically in the design, development and deployment stages, and introduced the concept of an AI Chain of Trust to discuss the various stages and their interrelations.

Trust between humans and agents relates to these three aspects; reliability, credibility, and intimacy. Unfortunately, standard, or sub-optimally designed AI agents result in low levels of trust build-up as illustrated in Fig. 2(a). This illustrative example shows sequential decision making where the agent proposes exactly 1 out of 5 different actions in each sequential scenario. The operator considers the proposed action but may decide, in some cases, on a different action based
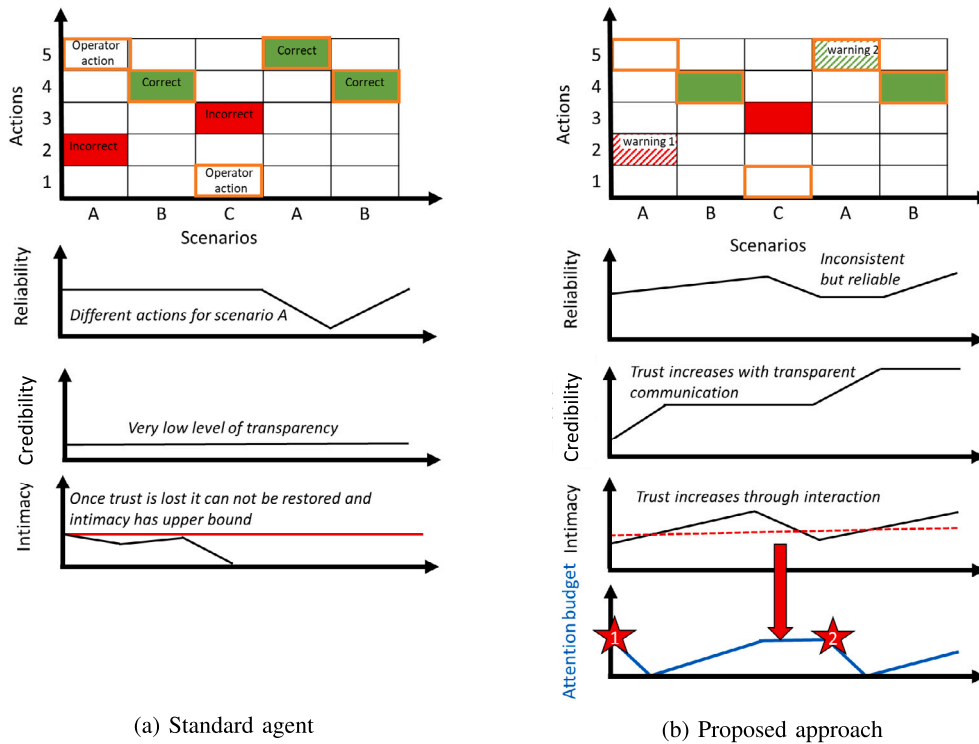
Fig. 2. Trust concept and the proposed model using attention budget and warnings for actions. The proposed approach considers attention budget of the human, a warning function, and to explain about the region increasing credibility.

on other tools, or experience. Therefore, sometimes the agent may propose an incorrect action in conflict with the operators' expectation, and in that case, the intimacy may decrease, and, as no explanation for incorrect actions is provided, the credibility stays at low levels. Sometimes, however the agent can "surprise" and teach the human through the agent's proposed actions that humans would normally not take. There, the human would approve this new action despite it being beyond the human's experience and having not trained for it. If successful, this new proposed action can then become a new strategy that humans will take in the future, and it becomes part of the human operational experience. However, as no interaction further considers the operator mental state, the operator can never fully trust the standard agent as the minimum level (red line) of intimacy is never surpassed. The reliability of the agent may improve with the experience of the agent which is to propose consistently the correct action in the correct scenario. In this illustration, the agent proposes two different actions in the scenarios, resulting in reliability decreases because of this inconsistency.

The proposed concept for human/agent interaction aims at improving trust, considering all three aspects of the model: credibility, reliability and intimacy. These three aspects are modelled as an attention budget of the human and warning signals from the agent to the human. As illustrated in Fig. 2(b), the agent can actively send warning signals to the human when the agent's confidence about its own actions is low. Sending these warning signals improves reliability, as well as credibility when it provides selective enough details. The warning signals can be discrete, continuous information about the confidence or aiming at explaining the warnings (e.g., in this challenge regional signals are supplied to further improve credibility of agents). The attention budget develops over time (similar as intimacy). The attention budget decreases, when the agent warns the human. Intimacy can increase if the warning was relevant or else it will decrease. In case of unwarned failure while the operator could have paid attention, intimacy is modelled to decreases substantially. The attention budget is a balance for operators to decide when they can trust (the agent) or

their own experience. A more accurate, and transparent agent will build trust and will result in overall higher available attention and reduced supervision requirements.

3. A new competition design for human AI-agent trust-building

Following the trust concept and model described above, we developed a new L2RPN competition in 2021 with the trust-building between human operator and AI-agent as the focus. The competition was organised through the Codalab platform in Summer–Autumn 2021, as part of the ICAPS conference (International Conference on Automated Planning and Scheduling) and attracted 100 participants. An overview of the competition is provided below, followed by more details.

3.1. Competition overview

Besides operational performance, the L2RPN 2021 competition is structured to build trust between humans and agents using the credibility, reliability, intimacy framework. Entrants are encouraged to design their agents to grade how confident it is of achieving a positive outcome (reward) for an action. It should send an alarm (to the operator) when the proposed actions are of low confidence. This is a proxy for identification of upcoming cascading failure and serves to reduce the conflict in evidence for the human operator (**reliability**), and to alert them to impending system issues. When formulating the problem, the issue of over-alarming was a risk to positive human–agent interaction. Conversely, the human operator supervising automated systems can experience "too much reliability", otherwise known as "out-of-the-loop" effect if the human is not warned or given enough time to respond. This is where operators are cognitively dis-engaged from real time monitoring and control. When forced to intervene, they are not aware of what or where the problem is. Both illustrates the need for the agent to consider the operator's state in its interactions (**intimacy**): the relationship quality depends on the right level of solicitations. We propose to model a budget for the operator's available attention so that
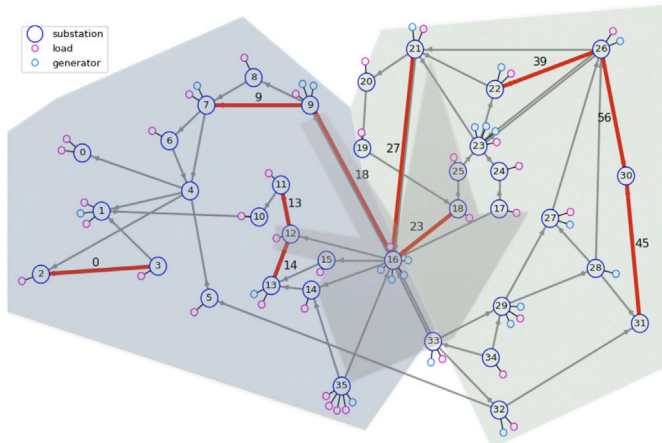
Fig. 3. Top right IEEE 118 subgrid. Attackable lines by the opponent are red coloured. 3 alarm regions are highlighted.
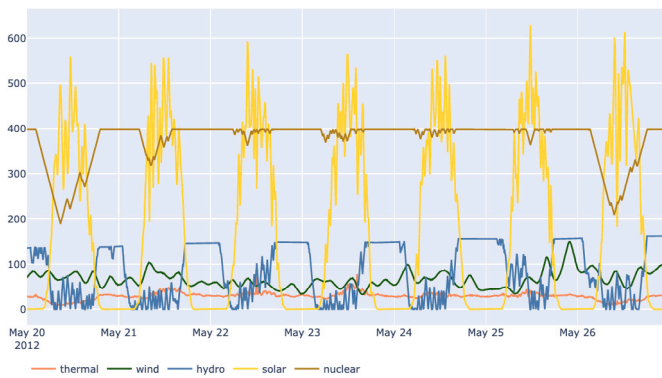

Fig. 4. Weekly production example per carrier type in May.

the agent can be designed to consider the human in-the-loop, and which incentivises the agent to choose the best times for interactions under the attention constraints. Finally, agents are incentivised to selectively explain when and where a problem originated among pre-defined network areas (**credibility**).

The participants were eventually evaluated on a score computed over 24 5-minute resolution weekly scenarios. It was composed of the alarm score (detailed after) and the network operation cost score (see [30]) with the following weighting:

$$Score = 0.3 * Score_{Alarm} + 0.7 * Score_{OperationCost} \qquad (1)$$

3.2. Power network operation environment

The competition is based on one third subgrid of IEEE 118-bus system as in [9] and showed on Fig. 3.

The renewable share makes up to 20% of the overall energy mix, which is a proxy for high variability in network operation parameters. Monthly Production and Load consumption with a 5-minute resolution time was made available in the training environment and are representative every month of the year (see example in Fig. 4). They have been generated through the open-source Chronix2grid package.[1]

The L2RPN Markov Decision Process (MDP) formulation have been previously described here [31] and is implemented in Grid2op [32]. The most notable details about the environment, observation and action

---

[1] See https://github.com/BDonnot/ChroniX2Grid/tree/master/input_data/generation/case118_l2rpn_icaps_2x.

spaces are described below. The agents can take actions considering the following constraints:

- Events such as maintenance (deterministic) and line disconnections (stochastic and adversarial) can disconnect power lines for several hours.
- Power lines have thermal limits. If a power line is overloaded for too long (e.g 15 min), it automatically disconnects. This can lead to cascading failures.
- An agent must wait few hours before reconnecting an incurred line disconnection (e.g one day).
- Additionally, to avoid expensive network asset degradation or failure, an agent cannot act too frequently on the same asset (e.g not more than once in a 15-minute time period) or perform too many actions at the same time (e.g topological action over 1 substation per step).

The "Game Over" condition is triggered if total demand is not met any more, that is if some consumption is lost at a substation, possibly because of a cascading failure. About the action space, possible actions are:

- Cheap topology changes (discrete and combinatorial actions) that allow for line disconnection/re-connection and substation nodal re-configurations.
- Costly production changes (continuous actions) through redispatching or now curtailment. They can be modified within the physical constraints of each plant over time.

The final action space has more than 70,000 discrete actions and a 20-dimensional continuous action space.

Considering the observation space, agents can observe complete states of the power network at every step. This includes flows on each power line, electricity consumed and produced at each node, power line disconnection duration etc. After verification of the previously described constraints, each action is fed into an underlying power flow simulator to run AC power flow [33] to calculate the next network state. Agent also have the opportunity to **simulate** one's action effect on the current state, to validate its action for instance as a human operator would do. But the future remains unknown: anticipating contingencies is not possible, upcoming productions and consumption are stochastic.

The novelty of this competition environment comes from the consideration and interaction of three different kinds of "agents" within the environment: AI based agent, the human operator which needs to focus its attention when it is the most important, and an opponent which emulates contingencies that the network must be robust against. We will hence describe them in the following dedicated subsections.

3.3. Alarm and operator's attention modelling

An agent should be designed to send alarms at a given time while specifying an area among 3 pre-defined ones (as in Fig. 3). This area demarcation does not have a direct effect on the environment, but will enable desired interactions with an operator, who might modify it, based on the information from the agent.

With regard to the operator's attention, we model it as an "attention budget" α_t at each step t, compatible with an MDP formulation. Each time an agent tries to raise an alarm to require the human attention, it has a cost of κ (held constant and set to $\kappa = 1$). On the other side, if the agent does not require the operator attention, then the "attention budget" increases by $\mu > 0$ (1.5 per day or per 288 timesteps here). Then, we model the operator attention as:

(1) $\alpha_{t+1} = \alpha_t - \kappa$ if an agent raised an alarm
(2) $\alpha_{t+1} = \alpha_t + \mu$ otherwise

Human attention is limited in reality. To make sure that an agent cannot raise alarms too often, the attention budget α_t is capped to a maximum value A ($A = 3$ here). This ensures that the agent cannot raise more than $\frac{A}{\kappa}$ consecutive alarms. Indeed, it can only raise an alarm if the attention budget is above cost κ. Otherwise it has to wait to recover the necessary budget.

In case of failure at timestep $\bar{t}$, an operator should ideally be warned $T_{opt} = 35$ min ahead of time to make a more complex study and take an informed decision. An alarm is considered relevant if sent within $[\bar{t} - (T_{opt} + T_{width}), \bar{t} - (T_{opt} - T_{width})]$, with $T_{width} = 25$ min here. An alarm will hence not be considered if raised in the final 10 min, before a blackout event, as it is too late for a human operator to perform a study in response to the alarm. An alarm sent greater than one hour is not considered either as this is not selective enough. 35 min was chosen as optimal, but may be adjusted for future competitions.

Finally, an alarm score function $\bar{S}$ rewards the agent for sending proper alarms at the right time ahead of failure:

(1) if no failure occurs, $\bar{S}$ is given it maximum value, 100 points here, as avoiding failure should always be favoured.
(2) if the agent fails the scenario at $\bar{t}$ but raised an relevant alarm at t_a then $\bar{S} = 100 \left(1 - \frac{|T_{opt} - |\bar{t} - t_a||}{T_{width}}\right) \times F_{area}$.
(3) else if the agent failed to raise an alarm and the system blacks out, it gets a penalty score of -200 points.

F_{area} is a multiplying factor depending on if the alarm spotted the right area of cascade ($F_{area} = 1$) or not ($F_{area} = 0.67$).

NB. If an agent sends valid alarms at different times t_a, the maximum score of each of the valid alarms is taken.

3.4. Opponent modelling and considerations

The strategies implemented by the agents in the competition must be robust to unexpected network events, whether natural or intentional. To promote this robustness, we have kept the adversarial approach [34] again for the 2021 competition. We have placed in the environment a "special agent" – an "opponent" – whose role is to simulate failures on the network at particular times. the agent must respond to this adversarial attacks on the network.

Three principles are important in the opponent design:

- Aggressiveness: A too aggressive opponent can bias the competition towards some kind of unrealistic game far from operational concerns. It can also discourage people from participating in the competition.
- Unpredictability: It is also important for the opponent to be as unpredictable as possible, since we do not want the agents to learn and predict the behaviour of the opponent and adapt specifically to it.
- Fairness between the participants. The opponent must present the same aggressiveness to all participants.

A few improvements have been made for this edition:

- Attack times: These are random. For more unpredictability, they are drawn according to an exponential distribution (geometric distribution in discrete time) calibrated to have roughly one attack per day on average but not always exactly one per day as before.
- Durations of the attacks: These are changing following an exponential distribution (they were fixed to four hours in the previous edition) as seen in Fig. 6 but with a within a duration constraint of 2 to 8 h.
- Attacked lines. In order to reflect the idea that the most electrically loaded lines are generally the most prone to failures, we have weighted the probability for a line of being the object of the

#	User	score ▲	operational cost ▲	attention cost ▲	Computation time ▲
1	xd_silly	57.45 (1)	59.80 (1)	51.94 (1)	548.72 (10)
2	SupremaciaChina	47.63 (2)	54.69 (2)	31.17 (2)	1465.96 (17)
3	maze-rl	46.81 (3)	54.18 (3)	29.61 (3)	787.32 (15)
4	IndigoSix	33.75 (4)	45.57 (4)	6.17 (4)	768.10 (14)
5	lujixiang	27.90 (5)	42.21 (5)	-5.50 (5)	668.27 (12)

Fig. 5. Final ICAPS competition leaderboard.

current attack by the load factor of the line. On average this year, some lines get more attacked than others, but within a maximum 1:4 ratio from the most attacked one to the least attacked one.

This year again, to avoid having too aggressive attacks, we have kept the principle of one attacked electric line at a time. No multiple attacks. The 10 same attacked lines are shown on Fig. 3. It is important to note that for fairness the attack times and durations are the same for everyone in the evaluation scenarios (even if these times and durations are unknown to the participants), but not necessarily attacks on the same lines.

4. Evaluation of competition design

This section evaluates the designed competition with trust concept by analysing the results of the competition and further investigating simple example agents.

4.1. Competition results

The official results and winners were announced in mid-October 2021 and presented at a webinar in February 2022.[2] The best performing agents are the ones that achieved a combination of high operational and attention scores (see Fig. 5). The results confirm that there was a sufficient incentive to take into account the trust aspects (issuing the right alarm at the right time) besides the pure operational performance (running the power network) and this validates the framework used for modelling and evaluating trust in the competition. Given that operational performance have already been analysed in depth in previous competitions, in the below sections, we focus on the trust aspect and the related attention score of the competition winners.

Analysing the results over the test scenarios, the two best agents Xd_silly (Xd) and SupremaciaChina (SC) successfully operate the network (i.e. without the network blacking out) over 16 out of 24 scenarios. Overall, both agents have 7 failing scenarios in common. The best agent sends valid alarms in 5 out of 8 failing scenarios and the second best in 4 out of 8. i.e. predicted failure rates of 63% and 50% respectively (see Table 1). In these scenarios, the agent sends alarms from 3 to 7 timesteps ahead of the failure (7 being the calibrated ideal time of alarm), which might be an indication of its planning time horizon. When sending alarms beyond 7 timesteps, it was usually the case that the operator's attention budget had diminished (so there was previous instance of many alarms) so it could be concluded that the agent would probably have resent alarms later on if the attention budget was sufficient.

Looking in more depth at some statistics from the competition results, shows that Xd requires less attention from the operator than SC, and is also more cautious with its attention budget. Indeed, it sends about 0.63 alarm per day on average (compared to SC: 0.78), keeps an average budget of 2.5 (compared to SC: 2.2) and only spends 1.5% of the time with an attention budget below 1 (compared to SC: 10% of the time). This highlights that Xd has more advance behaviour

---

[2] L2RPN Webinar: https://www.youtube.com/watch?v=WOt8xgpC370.

Table 1
Best alarm time and score comparison over failing scenarios for the 2 best agents.

Scenario	Xd_silly			SupremaciaChina		
	$\bar{S}$	$t_a - \bar{t}$	$\bar{t}$	$\bar{S}$	$t_a - \bar{t}$	$\bar{t}$
$dec12_1$	−200	−2	66	−200	−2	66
$dec12_2$	56	−9	710	64	−8	709
$feb40_1$	−200	−2	22	24	−3	23
$jan28_1$	42.7	−4	1997	−200	−15	790
$jan28_2$	66.7	−7	678	56	−9	668
$jun01_1$	100	−7	953	–	–	2016
$mar07_1$	–	–	2016	−200	−2	1700
$nov34_1$	64	−8	1282	−200	−2	1267
$nov34_2$	−200	−2	163	42.7	−4	1656

Fig. 6. Two best agents behaviour over time and before respective failures in scenario jan28_1. It shows times of actions (as the topology distance varies) and alarms, and periods of attacks).

in regard of its ability to warn an operator when it is most needed, possibly suggesting a better assessment in the confidence of its actions and capabilities. In terms of actions, Xd performs also less actions compared to SC, both on average per week (23.5 versus 26.5) and at maximum (38 versus 64). It shows that Xd is somehow more efficient in its decisions and actions. We will now give a short description of the nature of those agents that could explain those observations.

4.2. Description of best competition agents

Both agents leverage the actions that were learnt by the best winning solution of NeurIPS 2020 L2RPN competition [13]. Xd is a hybrid agent that combines learnt modules and simulation. One learnt module based on a Deep Neural Network gives fast predictions of action impacts on line power flow margins. They use this predictive model to explore the best possible combination of actions up to a depth of 4, defining a planning horizon over 4 timesteps, but without explicitly taking uncertainties into account over this horizon. They further simulate the top candidate sequences. Thanks to this feature, an alarm is not naively raised as soon as an overload appears, in the case when a sequence of actions is expected to relieve it. If none has been found to relieve existing overloads, only then an alarm would be raised. A rule eventually prevents re-sending alarm in less than 3 timesteps.

The second best competitor, SC, is an advanced expert agent, which makes proper use of rules and simulation to select the right actions in real-time over an initially curated database of effective actions. It however does not build a planning horizon and is closer to a greedy agent in that regard. Its alarm module is, in part, rule based, checking if overloads exist, if some lines are off and letting at least 5 timestep interval between alarms. It is nevertheless quite reactive for any overload showing up and could be quick at depleting its attention budget as we have noticed before. An additional alarm model is learnt to predict a percentage of how appropriate sending an alarm is at a given time. When above a threshold of 70%, an alarm could be sent. These two strategies look complementary and it has been assessed on validation scenarios that this learnt model when combined with the rule-based one improves the attention scores by few points.

Given those characteristics, we will now make a more detailed behaviour analysis over some interesting scenarios.

4.3. Behaviour analysis of competition agents

From the list of scenarios that the winning agents failed to solve, dec12_2 and jan28_2 are interesting for judging how well the two agents can anticipate its time of failure, instead of reacting and merely surviving attacks. Indeed, in theses cases, failure occurs in the last part of an attack period and not right at the beginning. The two agents have the last alarm timing right, about 7 timesteps before collapse, but are not accurate enough on the location (spatial area) of the failure. Based on this observation, we can assume that both agents have developed quite good prediction and planning skills over a fairly long time-horizon. But this hypothesis is nevertheless mitigated by the fact that
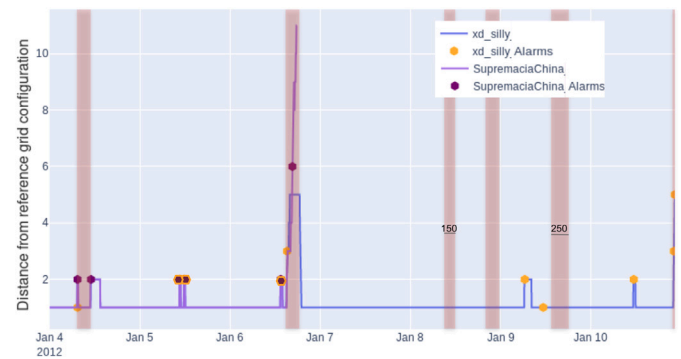
they also run out of budget and would have probably sent one more alarm if they could have. They somehow "luckily" run out of budget at the right time.

A similar situation for the SC agent occurs early evening of January 6th in scenario jan28_1 as shown in Fig. 6. But this time it sends its last alarm too early (more than one hour ahead) before running out of budget. It then runs into a long sequence of actions, deviating strongly from the reference topology. In this sequence, it seems that the agent does not know what it was actually doing and where it was going: a characteristic of a greedy behaviour as we have suspected before. At that point, Xd survives with a (somehow smart) sequence of 4 actions. In this scenario, we also see that agents often send alarms during periods of attacks but not only during these periods. The Xd agent almost entirely manages this scenario but eventually fails. It nonetheless managed to save enough attention budget to send a proper alarm before failing, thanks to not being too eager at sending alarm and spending its attention budget as seen in the statistics. When considering how an agent would interact with a human in reality, this is good behaviour. It alerts the operator in time that a major event will occur without action, prompting them to take over system operation. We also notice this at the beginning of the scenario: it only sends one alarm instead of the two alarms sent by SC. Sending two alarms risks over exercising the operator's attention.

In the other failing scenarios, the reason why the agents get a penalty score, is that they send an alarm too late and are not able to survive long enough. This often happens right after a strong attack on one of the high voltage lines.

Finally, in none of these scenarios did we notice a willingness to deliberately fail at a given timestep to possibly maximise the attention score after sending an alarm earlier. This is reassuring for the framework and competition design.

Our two agents showed good **reliability** due to its good operational performance and an ability to raise alarms before failure. They however mostly failed on the **credibility** side, not being selective enough on the time of alarm and the area of failure. In terms of **intimacy**, they also appeared limited, not taking sufficient account of its bounded attention budget when sending alarms. All this suggests that they cannot yet be considered trustworthy enough. For such complex acting agents. Is this a limitation of rule-based alarms? Would it be necessary to learn an alarm model instead? We now try to give some insights to those questions through dedicated experiments.

4.4. Investigation of trust concept with example agents

With example agents, we aim to gain insights into some remaining challenges when developing an agent with a successful alarm feature. To this end, the uncertainties of the power system operation and the constraints of sending meaningful alarms result in several challenges:

(1) Given the attention budget α_t, the agent has to decide carefully when to send an alarm without wasting its budget.

(2) To make the alarm successful, the agent has to send it in a particular time window before the failure/collapse (defined as 'game over'). As the underlying environment is stochastic (*eg* random possibilities of lines being attacked) it is often too difficult to precisely predict the 'game over'.

(3) On the other hand, an agent's successful alarm sending capability is directly linked to its current action. Hence, the challenges in designing the alarm feature increase with the increase in complexity of the agent's action selection criteria.

Next, we investigate in detail the design of agents with alarm features. To ease our understanding, without loss of generality, we can conceptually split the agent into two distinct parts, (a) action-making, (b) alarm-generating.

Algorithm 1 Rule-based Alarm Agent-I

1: Check whether any line is disconnected or attacked.
2: **if** disconnection or attack **then**
3:     Check for any overload:
4:     **if** Overload **then**
5:         Detect zones of overload and send an alarm.
6:     **end if**
7: **else**
8:     Do not send any alarm.
9: **end if**

First, we try to design a simple rule-based alarm agent. As mentioned earlier, a sound alarm agent can detect a possible danger in the running condition of the system. To this end, the most obvious choice is to monitor the capacity of each power line ρ, which is defined as the observed current flow divided by the thermal limit of each power line. Besides, there are possibilities that a power line can be attacked or can be disconnected due to maintenance, and any line disconnection obviously stresses the system operation. Hence, we extract the necessary information from the current observation and define a simple rule-based alarm feature agent (RbA-I) as given in **Algorithm-1**. The design concept of this alarm feature is straightforward, and we tested this feature with two different action-making agents i) 'Do-Nothing Action Agent' (DN) and ii) 'Simulation-intensive Expert Action Agent (SiE)'. In two different instances of testing, we observed that DN + RbA-I can send 14 successful alarm out of 24 different monthly scenarios. While SiE + RbA-I sends 10 successful alarms out of the same 24 different scenarios. In this testing phase, we observed that no scenarios are completed till the end by any of the agent. We can state that the simple rule-based alarm feature can be good for DN, but the same is not as suitable for complex action agents. The reason is quite apparent; in DN, the agent does not take any corrective action. Thus, it can be easily inferred that when the system is operating with one or more line outages and at the same time this power system is overloaded, failure is inevitable. In contrast, an SiE can solve some difficulties after executing necessary corrective actions. The simple alarm agent fails to interpret the outcomes of the expert actions and sends unnecessary alarms thinking that there is an impending collapse. This ultimately reduces their attention budgets, makes them unable to send an alarm when the situation needs it. Plus, the operating conditions of failure for particular scenarios are not the same in the case of DN and SiE. Hence, there may be the possibility that the DN fails for simple reasons that are easily detectable. While the failure of SiE is due to some complex reasons, the simple alarm agent fails to detect the same. This implies that the alarm feature of the agent needs some improvement, in order to perform well with a complex action agent. To improve the alarm feature, we studied some of the failures with unsuccessful alarms. It is found that attention budget and the timing of the alarm are playing key roles. Mostly, the alarms are sent but are not successful because either (a) the agent does not have the required amount of budget to send a successful alarm, or (b) the collapse occurred suddenly after

a line attack, hence the alarm did not meet the desired time-window requirement. To tackle such situations, the agents need to predict the outcome of a line attack before the attack actually happened. Here, we modify the alarm features given in **Algorithm-1** and add some additional condition for sending alarm defining RbA-II agent:

- Simulate N-1 for the attacked lines list. If an overflow is predicted and $\max_{l \in \textbf{all line}} \rho_l^{\text{pred}} > T_h$, and there is no alarm in last D time-steps, generate alarms for the zone where the predicted overflows exceed the defined threshold T_h.

With this modification, the same set of scenarios: DN + RbA-II and SiE + RbA-II sends respectively, 21 (previously 14) and 13 (previously 10) successful alarms. This number increased from the one found with RbA-I, especially for the DN agent but there was a smaller increase for the SiE agent. We see that designing a complex rule-based alarm agent does not greatly improve this score on top of a complex acting agent. A rule-based alarm agent is not enough and we believe that this alarm prediction part can be improved with the help of a learning-based agent. This part should be further investigated in the future.

5. Conclusion

On the journey towards creating trustworthy AI-based assistants for future network operators, we have proposed a trustworthy framework that builds on the reliability, credibility and intimacy model of trust, by explicitly considering the human operators' mental workload and capability of addressing issues when early warning and relevant network information are provided. Through the L2RPN with trust competition in 2021, we have successfully designed a realistic active warning environment to experiment and evaluate trust between humans and agents. Winning teams have achieved the best alarm scores overall, in combination with the best operational performance, and demonstrated good reliability. By relying mostly on rule-based alarms, there however remains room for improvement on the credibility and intimacy aspects. Learning based alarm agents could help address in the future these open questions.

CRediT authorship contribution statement

Antoine Marot: Conceptualization, Methodology, Data curation, Visualization, Validation, Writing – original draft. **Benjamin Donnot:** Conceptualization, Methodology, Software, Writing – original draft. **Karim Chaouache:** Conceptualization, Methodology, Investigation, Writing – original draft. **Adrian Kelly:** Conceptualization, Resources, Writing – original draft. **Qiuhua Huang:** Conceptualization, Resources, Writing – review & editing. **Ramij-Raja Hossain:** Conceptualization, Investigation, Writing – original draft. **Jochen L. Cremer:** Conceptualization, Visualization, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Antoine Marot, Adrian Kelly, Matija Naglic, Vincent Barbesant, Jochen Cremer, Alexandru Stefanov, Jan Viebahn, Perspectives on future power system control centers for energy transition, J. Mod. Power Syst. Clean Energy 10 (2) (2022) 328–344.

[2] Dominik Dellermann, Philipp Ebel, Matthias Söllner, Jan Marco Leimeister, Hybrid intelligence, Bus. Inf. Syst. Eng. 61 (5) (2019) 637–643.

[3] Antoine Marot, Alexandre Rozier, Matthieu Dussartre, Laure Crochepierre, Benjamin Donnot, Towards an AI assistant for human grid operators, 2020, arXiv preprint arXiv:2012.02026.

[4] Jan Viebahn, Matija Naglic, Antoine Marot, Benjamin Donnot, Simon H Tindemans, Potential and challenges of ai-powered decision support for short-term system operations, in: CIGRE Session 2022, 2022.

[5] Laurine Duchesne, Efthymios Karangelos, Louis Wehenkel, Recent developments in machine learning for energy systems reliability management, Proc. IEEE 108 (9) (2020) 1656–1676.

[6] Mevludin Glavic, Raphaël Fonteneau, Damien Ernst, Reinforcement learning for electric power system decision and control: Past considerations and perspectives, IFAC-PapersOnLine 50 (1) (2017) 6918–6927.

[7] Qisong Yang, Thiago D Simão, Simon H Tindemans, Matthijs TJ Spaan, WCSAC: Worst-case soft actor critic for safety-constrained reinforcement learning, 2021.

[8] Qiuhua Huang, Renke Huang, Weituo Hao, Jie Tan, Rui Fan, Zhenyu Huang, Adaptive power system emergency control using deep reinforcement learning, IEEE Trans. Smart Grid 11 (2) (2019) 1171–1182.

[9] Antoine Marot, Benjamin Donnot, Gabriel Dulac-Arnold, Adrian Kelly, Aidan O'Sullivan, Jan Viebahn, Mariette Awad, Isabelle Guyon, Patrick Panciatici, Camilo Romero, Learning to run a power network challenge: a retrospective analysis, in: NeurIPS 2020 Competition and Demonstration Track, PMLR, 2021, pp. 112–132.

[10] A. Marot, B. Donnot, S. Tazi, P. Panciatici, Expert system for topological remedial action discovery in smart grids, IET Digital Library (2018).

[11] Tu Lan, Jiajun Duan, Bei Zhang, Di Shi, Zhiwei Wang, Ruisheng Diao, Xiaohu Zhang, AI-based autonomous line flow control via topology adjustment for maximizing time-series ATCs, in: 2020 IEEE Power & Energy Society General Meeting, PESGM, IEEE, 2020, pp. 1–5.

[12] Deunsol Yoon, Sunghoon Hong, Byung-Jun Lee, Kee-Eung Kim, Winning the L2RPN challenge: Power grid management via semi-Markov afterstate actor-critic, in: International Conference on Learning Representations, 2020.

[13] Bo Zhou, Hongsheng Zeng, Yuecheng Liu, Kejiao Li, Fan Wang, Hao Tian, Action set based policy optimization for safe power grid management, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2021, pp. 168–181.

[14] John R. Wilson, Sarah Sharples, Evaluation of Human Work, CRC Press, 2015.

[15] Ho Chit Siu, Jaime Peña, Edenna Chen, Yutai Zhou, Victor Lopez, Kyle Palko, Kimberlee Chang, Ross Allen, Evaluation of human-AI teams for learned and rule-based agents in hanabi, Adv. Neural Inf. Process. Syst. 34 (2021).

[16] Yonghong Wang, Munindar P. Singh, Evidence-based trust: A mathematical model geared for multiagent systems, ACM Trans. Autonom. Adapt. Syst. (TAAS) 5 (4) (2010) 1–28.

[17] Miles Brundage, Shahar Avin, Jasmine Wang, Haydn Belfield, Gretchen Krueger, Gillian Hadfield, Heidy Khlaaf, Jingying Yang, Helen Toner, Ruth Fong, Tegan Maharaj, Pang Wei Koh, Sara Hooker, Jade Leung, Andrew Trask, Emma Bluemke, Jonathan Lebensold, Cullen O'Keefe, Mark Koren, Théo Ryffel, JB Rubinovitz, Tamay Besiroglu, Federica Carugati, Jack Clark, Peter Eckersley, Sarah de Haas, Maritza Johnson, Ben Laurie, Alex Ingerman, Igor Krawczuk, Amanda Askell, Rosario Cammarota, Andrew Lohn, David Krueger, Charlotte Stix, Peter Henderson, Logan Graham, Carina Prunkl, Bianca Martin, Elizabeth Seger, Noa Zilberman, Seán Ó hÉigeartaigh, Frens Kroeger, Girish Sastry, Rebecca Kagan, Adrian Weller, Brian Tse, Elizabeth Barnes, Allan Dafoe, Paul Scharre, Ariel Herbert-Voss, Martijn Rasser, Shagun Sodhani, Carrick Flynn, Thomas Krendl Gilbert, Lisa Dyer, Saif Khan, Yoshua Bengio, Markus Anderljung, Toward trustworthy AI development: Mechanisms for supporting verifiable claims, 2020, arXiv.Org, E-Print Archive, Mathematics.

[18] Ben Shneiderman, Human-centered artificial intelligence: Reliable, safe & trustworthy, Int. J. Hum.–Comput. Interaction 36 (6) (2020) 495–504.

[19] Steven H. Low, Convex relaxation of optimal power flow—Part I: Formulations and equivalence, IEEE Trans. Control Netw. Syst. 1 (1) (2014) 15–27, http://dx.doi.org/10.1109/TCNS.2014.2309732.

[20] Mithun M. Bhaskar, Muthyala Srinivas, Sydulu Maheswarapu, Security constraint optimal power flow(scopf)- a comprehensive survey, Glob. J. Technol. Optim. 2 (11) (2011).

[21] Mohammad Iman Alizadeh, Muhammad Usman, Florin Capitanescu, Toward stochastic multi-period AC security constrained optimal power flow to procure flexibility for managing congestion and voltages, in: 2021 International Conference on Smart Energy Systems and Technologies, SEST, IEEE, 2021, pp. 1–6.

[22] Lisanne Bainbridge, Ironies of automation, in: International Federation of Automatic Control, vol. 5, (1098) Pergamon Press Ltd, UK, 1983.

[23] Alon Jacovi, Ana Marasović, Tim Miller, Yoav Goldberg, Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in ai, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021, pp. 624–635.

[24] Roger C. Mayer, James H. Davis, F. David Schoorman, An integrative model of organizational trust, Acad. Manag. Rev. 20 (3) (1995) 709–734.

[25] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, Francisco Herrera, Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion (ISSN: 1566-2535) 58 (2020) 82–115, http://dx.doi.org/10.1016/j.inffus.2019.12.012, URL https://www.sciencedirect.com/science/article/pii/S1566253519308103.

[26] Ke Zhang, Peidong Xu, Jun Zhang, Explainable AI in deep reinforcement learning models: A SHAP method applied in power system emergency control, in: 2020 IEEE 4th Conference on Energy Internet and Energy System Integration, EI2, 2020, pp. 711–716, http://dx.doi.org/10.1109/EI250167.2020.9347147.

[27] Joel Tetreault, Dan Bohus, Diane Litman, Estimating the reliability of mdp policies: a confidence interval approach, in: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, 2007, pp. 276–283.

[28] Ewart J. De Visser, Richard Pak, Tyler H. Shaw, From 'automation'to 'autonomy': the importance of trust repair in human–machine interaction, Ergonomics 61 (10) (2018) 1409–1427.

[29] Ehsan Toreini, Mhairi Aitken, Kovila Coopamootoo, Karen Elliott, Carlos Gonzalez Zelaya, Aad van Moorsel, The relationship between trust in AI and trustworthy machine learning technologies, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, in: FAT* '20, Association for Computing Machinery, New York, NY, USA, ISBN: 9781450369367, 2020, pp. 272–283, http://dx.doi.org/10.1145/3351095.3372834.

[30] Antoine Marot, Isabelle Guyon, Benjamin Donnot, Gabriel Dulac-Arnold, Patrick Panciatici, Mariette Awad, Aidan O'Sullivan, Adrian Kelly, Zigfried Hampel-Arias, Learning to run a power network in a sustainable world neurips2020 challenge design, 2020, https://www.public.asu.edu/~yweng2/Tutorial5/pdf/111.pdf.

[31] Antoine Marot, Benjamin Donnot, Camilo Romero, Balthazar Donon, Marvin Lerousseau, Luca Veyrin-Forrer, Isabelle Guyon, Learning to run a power network challenge for training topology controllers, Electr. Power Syst. Res. 189 (2020) 106635.

[32] Grid2Op https://github.com/rte-france/Grid2Op.

[33] LightSim2Grid https://github.com/BDonnot/lightsim2grid.

[34] Loïc Omnes, Antoine Marot, Benjamin Donnot, Adversarial training for continuous robustness control problem in power systems, 2020, arXiv preprint arXiv:2012.11390.