2025 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm) | 979-8-3315-2084-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/SmartGridComm65349.2025.11204591

Analyzing Agent Collisions in AI-Aided Energy Management Systems

Yefeng Yuan*, Yulin Zeng*, Hepeng Li†, Jie Gao‡, Xiao'ou Yang*,
Mohsen Ghafouri§, Yuhong Liu*, and Jun Yan§
*Santa Clara University, Santa Clara, CA, USA
†University of Maine, Orono, ME, USA
‡Delft University of Technology, Delft, Netherlands
§Concordia University, Montréal, QC, Canada

***Abstract*—The rapid growth of distributed energy resources (DERs) and autonomous control devices in behind-the-meter (BTM) systems has created a decentralized energy landscape, where artificial intelligence (AI) agents independently manage local objectives. While these AI-driven energy management systems (EMS) offer improved efficiency and flexibility, their uncoordinated operation poses risks to grid stability. Specifically, operational collisions can occur when self-interested agents pursue local optima without regard for system-wide safety, resulting in simultaneous violations of physical grid constraints. For instance, smart EV chargers and microgrid optimizers acting independently may synchronize high-demand actions, causing voltage sags or transformer overloads. This paper presents a systematic framework to characterize and detect agent-induced collisions in multi-agent energy systems. We formalize operational collisions in power grids, introduce metrics to quantify their frequency and severity, and develop an analytical workflow to attribute these events to specific agent policies. A case study with networked microgrids (MGs) demonstrates the framework by comparing independent and shared reward strategies, showing how cooperative incentives can mitigate collision risks. By proactively addressing these safety challenges, our work advances the development of resilient and trustworthy AI-driven energy management for future smart grids.**

***Index Terms*—Multi-agent systems, energy management, reinforcement learning, grid safety, constraint violations, behind-the-meter, distributed energy resources.**

I. Introduction

The energy landscape is rapidly evolving with the integration of DERs such as solar photovoltaics, battery storage, and EVs, along with increasingly sophisticated BTM control systems [1]. Recent breakthroughs in AI have further accelerated interest in AI-aided EMS for decision-making, optimization, and automation at the consumer edge of the grid. From smart EV chargers and intelligent thermostats to building automation and virtual power plants, stakeholders are deploying autonomous agents to optimize individual objectives such as cost savings, comfort, or revenue [2]. Unlike traditional utility-centric control, which relies on centralized coordination, these BTM deployments are decentralized, spontaneous, and driven by self-interest, raising new challenges for grid stability and regulation [3]. Indeed, the surge in independently designed AI agents with unaligned objectives introduces a critical question: How can uncoordinated AI systems coexist in a shared electrical network without creating chaos? Previous research has warned that unresolved conflicts and competition among self-learning agents can result in widespread societal failures and chaotic disruptions if not properly managed [4]. In the context of power systems, such failures can directly compromise safety, reliability, and fairness in energy supply.

One potential risk in multi-agent energy systems is operational collision. This may occur when decisions made by independent agents conflict on shared physical resources, leading to violations of grid operating constraints such as voltage limits or transformer capacity [5]. For instance, neighbouring smart MGs or EV charging stations may each optimize for local goals, such as maximizing self-consumption or meeting charging deadlines [6]. When their actions are combined, these behaviours can overload transformers or cause unacceptable voltage deviations. Such safety-critical failures result from a lack of coordination. The agents do not explicitly cooperate or communicate, so their actions can inadvertently produce harmful outcomes. Unlike traditional contingencies such as equipment failures or natural disturbances, operational collisions are driven by agent behaviours and can worsen over time as algorithms adapt and learn [7]. For example, an EV battery management system aiming to preserve battery health may frequently pause charging, while a smart charger seeks to maximize charging speed. Their uncoordinated actions can create oscillating power draws, resulting in voltage flicker or violations of ramp rate limits [8]. If these behaviours become widespread, they can compromise voltage regulation and accelerate equipment aging.

This paper contributes a structured approach to identifying and evaluating collision risks in agentic BTM energy systems. We build upon emerging concepts of agentic AI ecosystems to formalize how self-learning EMS may unintentionally conflict. Key contributions of this work are as follows:

- We formalize operational collisions in power grids as emergent constraint violations (voltage, thermal, and frequency limits) resulting from the independent actions of multiple AI agents. This concept distinguishes such failures from traditional disturbances or attacks, offering a clear framework for analyzing risks unique to multi-agent learning environments.
- We conduct an empirical evaluation using multi-agent reinforcement learning in a realistic microgrid setting,

demonstrating how different reward structures (independent versus shared) impact the frequency and severity of operational collisions. The results provide quantitative evidence for the value of cooperative incentive alignment and inform practical strategies for safer grid operation.

II. Related Works

To position our work within the field of intelligent energy management, we first review recent advances in multi-agent reinforcement learning (MARL) for power systems, as well as emerging strategies for safety and collision avoidance in distributed control. This related work highlights both achievements and ongoing challenges in ensuring reliable, scalable, and safe grid operation as AI agents become increasingly widespread.

1) MARL in Energy Systems and Grid Control: MARL has become a leading approach for decentralized decision-making in modern energy systems, particularly where multiple controllers operate under uncertainty and limited information, a defining characteristic of power grids with high DER integration [9]. MARL has been successfully applied to microgrid energy management, where distributed generators, storage, and flexible loads are coordinated to optimize cost and maintain voltage stability. Frameworks like EnergAIze have used MARL for vehicle-to-grid (V2G) and renewables integration, reducing peak loads and energy costs at the community level [10]. Similarly, MARL methods for EV charging have shown the ability to manage stochastic demand, mitigate transformer overload, and maximize solar utilization, outperforming rule-based and centralized strategies [11].

MARL is also used in transactive energy markets, where agents learn to optimize bids in decentralized trading environments [12] [13]. While such studies highlight the benefits of distributed, adaptive control, they also emphasize challenges in scalability, privacy, and system stability, particularly since BTM agents often represent different stakeholders.

Research shows that self-interested MARL agents can lead to inefficient or unsafe outcomes, such as suboptimal equilibria or grid constraint violations, unless cooperation is encouraged [14]. Solutions include centralized training with decentralized execution (CTDE), heuristic communication protocols, and shared reward structures to foster collaboration [15]. Our work builds on these findings by quantifying how shared rewards in MARL can improve grid safety and prevent operational collisions, compared to independent agent strategies.

It is important to distinguish cooperative learning (shared-reward MARL) from Federated Learning (FL)-based reinforcement learning. In cooperative MARL, agents share a common reward signal to align local objectives with system-wide goals. In contrast, FL-based RL focuses on privacy and scalability by training models collaboratively without sharing raw data. While cooperative MARL emphasizes coordination and incentive alignment, FL addresses data privacy. Future work could explore hybrid approaches that integrate cooperative rewards with privacy-preserving FL techniques, combining the strengths of both paradigms [16].

2) Safety and Operational Collision Avoidance in MARL: Deploying MARL in safety-critical domains like power grids requires mechanisms to enforce constraints and prevent unsafe actions. Standard MARL algorithms focused solely on reward optimization can inadvertently learn policies that violate grid limits, especially during exploration or in changing environments [3]. To address this, several safe reinforcement learning approaches have been adapted for multi-agent settings:

Safe reinforcement learning (RL) approaches are essential for ensuring the reliability of multi-agent energy management systems. Constrained multi-agent reinforcement learning (MARL), formulated as a Constrained Markov Decision Process (CMDP) with explicit penalties or Lagrange multipliers for constraint violations, effectively maintains grid parameters such as voltage within permissible limits [17]. Additionally, supervisory safety layers can dynamically monitor and adjust agent actions to prevent unsafe states, although careful implementation is needed to avoid hindering agent learning [3]. Runtime verification techniques, including formal verification and temporal logic shields, proactively prevent actions that could breach defined safety properties [18]. Furthermore, Control Barrier Functions (CBFs) have shown effectiveness in constraining actions within safe operating regions, highlighting their potential for grid safety applications [19].

While these methods focus on preventing violations, our work emphasizes detecting and analyzing operational collisions in MARL-based grids. Understanding how and when such collisions occur is essential, as even advanced MARL can result in frequent violations under stress [20]. By logging detailed state-action trajectories and quantifying collisions, our framework provides actionable insights for refining control strategies and complements preventive safety mechanisms. This contributes to a broader call for rigorous safety analysis in multi-agent energy systems such as smart grid [21].

III. Problem Formulation and Methodology

1) System Model: We consider a distribution feeder with multiple MGs operating as independent agents. Each MG $i \in \{1, \ldots, N\}$ has its own DERs, such as generators, ESS, and flexible loads, and is connected to a main distribution network monitored by a distribution system operator (DSO). The DSO can buy or sell power from the upper grid and imposes operational constraints (e.g., voltage limits, line capacity) that all agents must inherently share.

We model the interactions between MG controllers and the grid as a *partially observable Markov game* (POMG). Each MG (agent) observes its local state o_t^i at time step t, which may include its DER outputs, ESS charge level, local load, and possibly limited information from neighbours or the DSO (such as price signals or grid warnings). Based on its observation, agent i takes an action a_t^i, such as adjusting generation output, charging or discharging the battery, or shifting controllable loads. The joint actions of all agents, $(a_t^1, \ldots, a_t^N)$, influence the power flow in the network, resulting in a new system state and associated rewards.

We denote the state transition function as $P(s_{t+1} \mid s_t, a_t^1, \ldots, a_t^N)$, which is governed by power flow physics and potentially by external uncertainties (for example, fluctuations in renewable generation). The objective of each MG may vary depending on its design, but commonly includes minimizing local operating cost or maximizing utility, such as meeting demand at the lowest cost or maximizing self-consumption of solar generation.

2) Multi-Agent Reward Structures: We investigate two reward paradigms:

a) Independent rewards.: Each agent i receives a local reward r_t^i based solely on its individual objective. For example, r_t^i could be the negative of its operating cost at time t (including purchased energy cost minus any generation revenue), minus any local penalties (such as violations of its own device limits). In this self-interested setting, agents do not directly account for the grid-wide impacts of their actions. There is no explicit term in r_t^i for network constraint violations beyond their local scope.

b) Shared reward.: All agents receive a common global reward signal r_t^{global} that aligns with a cooperative objective. This can be defined as the sum of all agents' local rewards, adjusted for system-wide performance. Specifically, we use

$$r_t^{\text{global}} = \sum_{i=1}^{N} r_t^i - \Psi(\text{violations at } t), \tag{1}$$

where $\Psi(\cdot)$ is a penalty function for any constraint violations at time t. In practice, this means each agent's policy is reinforced by both the economic efficiency of the overall system and its safety (grid compliance). This shared-reward design encourages fully cooperative behaviour, as all agents strive to maximize the same cumulative objective.

Formally, the team objective in the shared reward case is to maximize the expected discounted sum of global rewards:

$$J(\pi) = \mathbb{E}_\pi \left[\sum_{t=0}^{T} \gamma^t r_t^{\text{global}} \right], \tag{2}$$

where $\gamma \in [0, 1)$ is a discount factor and $\pi = \{\pi^1, \ldots, \pi^N\}$ denotes the joint policy of all agents (each π^i mapping observations o_t^i to actions a_t^i).

In the independent case, each agent i maximizes its own reward:

$$J_i = \mathbb{E}_\pi \left[\sum_{t} \gamma^t r_t^i \right], \tag{3}$$

however, since all agents operate on the same physical grid, their actions are inherently interconnected. When agents pursue their own objectives independently, their combined decisions can unintentionally produce adverse effects for the entire system. In our analysis shown in section VI, this lack of coordination often resulted in outcomes such as grid constraint violations and operational collisions, even though each agent was acting optimally from its own perspective.

IV. Operational Collision Framework

In this section, we introduce a framework to define, quantify, and detect operational collisions—situations where uncoordinated agent actions drive the network into unsafe or constraint-violating states. By systematically identifying and analyzing these events, our approach offers deeper insight into the safety and coordination challenges facing multi-agent control in distribution networks.

A. Definition of Operational Collision

The introduction of learning-based controllers in distribution networks brings a critical concern on systemic risks, especially the emergence of *operational collisions*: events where the simultaneous, independent actions of multiple agents lead to violations of critical grid constraints.

We define an *operational collision* as any time step t during which one or more network constraints are violated as a direct result of agents' uncoordinated decisions. Specifically, let the set of monitored constraints include:

- **Voltage limits:** For all buses b, $V_{\min} \leq V_{b,t} \leq V_{\max}$;
- **Thermal limits:** For all lines or transformers ℓ, $S_{\ell,t} \leq S_\ell^{\max}$.

An operational collision at time t occurs if any of the above are violated, and the violation can be attributed to agent control actions at t. That is,

$$\exists\, b : (V_{b,t} < V_{\min} \text{ or } V_{b,t} > V_{\max}) \quad \text{or} \quad \exists\, \ell : S_{\ell,t} > S_\ell^{\max} \tag{4}$$

where the system state transition $(s_{t-1} \to s_t)$ is driven by the joint actions $(a_{t-1}^1, \ldots, a_{t-1}^N)$ of all agents.

Crucially, operational collisions may result from various patterns of agent behaviour:

- **Resource contention:** Multiple agents concurrently pursuing self-interest (e.g., maximizing local consumption or storage charging) may collectively exceed network capacity, causing voltage excursions or overloads.
- **Policy misalignment:** Agents trained or incentivized with purely local objectives may ignore, or insufficiently respond to, system-level risks, leading to emergent constraint violations.
- **Temporal coordination failures:** Even when individual actions appear safe in isolation, poorly timed or synchronized responses across agents (e.g., simultaneous load increases) can trigger collisions.

While collisions caused by a single agent operating recklessly are possible, our primary focus is on **multi-agent collisions**: instances where no individual agent alone would have caused a violation, but the combined, uncoordinated actions of several agents push the system into unsafe regimes. For instance, each agent independently charging its battery or increasing its load within permissible limits might still collectively overload shared infrastructure if such actions are temporally synchronized. These collective violations highlight the necessity for system-wide awareness and coordinated control strategies to anticipate and prevent cumulative impacts. Such

events are uniquely challenging in MARL settings, as they are not easily mitigated by traditional protection schemes or by penalizing individual violations post hoc. By defining and tracking operational collisions, we provide a lens for analyzing the *system-level risks* introduced by distributed intelligence.

B. Collision Metrics

To detect and quantify collision events, we instrument the simulation with a monitoring module that logs constraint violations at each time step. Let B_i denote the set of buses in the distribution network that belong to or are strongly affected by agent i (e.g., the buses within MG i or at its point of common coupling), and L_i the set of lines or transformers associated with agent i (such as lines within the MG or the tie-line connecting MG i to the main feeder). We then compute the following violation measures at each time t for each agent [22]:

- **Over-voltage magnitude** (O_i): The total amount by which voltages at agent i's buses exceed the upper limit (e.g., 1.05 p.u.), summed across all relevant buses:

$$O_i(t) = \sum_{b \in B_i} \max(V_{b,t} - V_{\max},\, 0). \tag{5}$$

- **Under-voltage magnitude** (U_i): The total shortfall where voltages at agent i's buses fall below the lower limit (e.g., 0.95 p.u.):

$$U_i(t) = \sum_{b \in B_i} \max(V_{\min} - V_{b,t},\, 0). \tag{6}$$

- **Thermal overloading** (L_i): The extent to which any line or transformer loading in agent i's area exceeds 100% of its rated capacity:

$$L_i(t) = \sum_{\ell \in L_i} \max\left(\frac{S_{\ell,t}}{S_\ell^{\max}} - 1,\, 0\right), \tag{7}$$

  where $S_{\ell,t}$ is the power flow through line or transformer ℓ at time t and $S_\ell^{\max}$ is its thermal limit. Any value above 1 indicates an overload; we sum all such excess fractions across lines in L_i.

Using these measures, we define agent i's *collision metric* as the aggregate violation magnitude:

$$C_i(t) = O_i(t) + U_i(t) + L_i(t). \tag{8}$$

This $C_i(t)$ can be interpreted as a severity score quantifying the extent to which agent i's territory is violating constraints at time t. Similar collision metrics have previously been used to characterize operational risks in grid management scenarios, such as generation redispatch studies [23]. Finally, we compute a *global collision metric* as the sum over all agents:

$$C_{\text{tot}}(t) = \sum_{i=1}^{N} C_i(t). \tag{9}$$

A value $C_{\text{tot}}(t) > 0$ indicates that at least one constraint is violated at time t (a collision event), while a larger value denotes more severe or multiple simultaneous violations.

While $C_i(t)$ and $C_{\text{tot}}(t)$ quantify the severity of violations numerically, it is crucial to interpret their practical implications. For example, a higher collision metric directly correlates with greater violations of voltage and thermal limits, potentially leading to equipment damage, accelerated aging, or power quality degradation. Even minor persistent violations can cumulatively impact grid stability and safety, emphasizing the importance of keeping collision metrics as low as possible.

For clarity, the pseudo-code indicated in Algorithm 1 outlines the calculation of these metrics at each simulation time step.

Algorithm 1 Collision Metric Calculation per Time Step

for each agent $i = 1$ to N **do**
  Retrieve voltage magnitudes $V_{b,t}$ for all $b \in B_i$
  $O_i(t) \leftarrow \sum_{b \in B_i} \max(V_{b,t} - V_{\max},\ 0)$
  $U_i(t) \leftarrow \sum_{b \in B_i} \max(V_{\min} - V_{b,t},\ 0)$
  Retrieve loading $S_{\ell,t}$ for all $\ell \in L_i$
  $L_i(t) \leftarrow \sum_{\ell \in L_i} \max\left(\frac{S_{\ell,t}}{S_\ell^{\max}} - 1,\ 0\right)$
  $C_i(t) \leftarrow O_i(t) + U_i(t) + L_i(t)$
end for
$C_{\text{tot}}(t) \leftarrow \sum_{i=1}^{N} C_i(t)$

Using these logged metrics, we quantify collision frequency as the proportion of time steps with $C_i(t) > 0$, and collision severity as the average value of $C_i(t)$ over time or, alternatively, its conditional mean during collision events. We also track the simultaneity of collisions by recording how often multiple agents have nonzero $C_i(t)$ at the same time, indicating interactive conflicts. This systematic approach enables post-hoc attribution, helping identify which agents and actions were responsible for each collision and providing insight into risky policies.

To train agent policies under both reward structures, we employ a multi-agent reinforcement learning approach using the PPO algorithm [24]. PPO is well-suited to this setting, as it supports decentralized agents with minimal information sharing and provides stable policy updates through trust-region constraints.

V. Case Study Designs

A key distinction of this study is its explicit focus on operational collisions, which are emergent failures arising from the uncoordinated actions of multiple independent agents. Unlike traditional contingencies such as faults or attacks, these are endogenous to multi-agent systems, resulting from decentralized controllers sharing infrastructure without explicit coordination. By formalizing and detecting these events, our work advances the understanding of AI-driven risks in energy systems, providing clear definitions, analytical tools, and new insights into the vulnerabilities of distributed intelligence.

To illustrate our framework, we present a case study using three MGs connected to a common 24.9kV distribution feeder (based on the IEEE 34-bus benchmark), each linked via step-down transformers and modelled as modified IEEE 13-bus

networks. Each MG includes two dispatchable generators (0.3–0.4MW, with generation costs and limits), a battery energy storage system (200kW, 1MWh, with defined efficiencies), solar PV (40kW), and wind (500kW) resources following historical output profiles. Aggregate demand for each MG is modelled on real utility data from Northern California (BANC, BANCMID, AZPS). The DSO manages power exchanges with MGs at market-based prices, enforcing voltage (0.95–1.05 p.u.) and line thermal limits. Feeder tie lines have finite capacity, so simultaneous high imports by multiple MGs can trigger overloads or undervoltage events.

1) Simulation Setup: The simulation runs in hourly time steps. To ensure robust learning, we use historical data from 2018–2019 for training, and reserve data from 2020 for testing and evaluation. We explicitly reserve this test data to assess operational collision frequency and severity under conditions reflecting realistic future scenarios. At each time step, the environment, built with the PandaPower power flow simulator [25], receives the agents' actions (including DG dispatch setpoints and ESS charge/discharge commands for each MG) and computes the resulting power flows and bus voltages. Any constraint violations are recorded by the monitoring module as described in the methodology.

During training, agents update their policies using the PPO algorithm. We implement PPO with a decentralized training approach: each MG agent learns its policy via policy gradient updates that respect a trust-region step size, and neighbours exchange only limited information as needed for the consensus update (the implementation follows the description in [22]). We use recurrent neural networks (two-layer GRUs with 256 units each) for the policy and value function of each agent, allowing them to capture temporal correlations in load and renewable variations. Hyper-parameters (learning rate, discount factor γ, etc.) are chosen as recommended in [22] to ensure stable convergence.

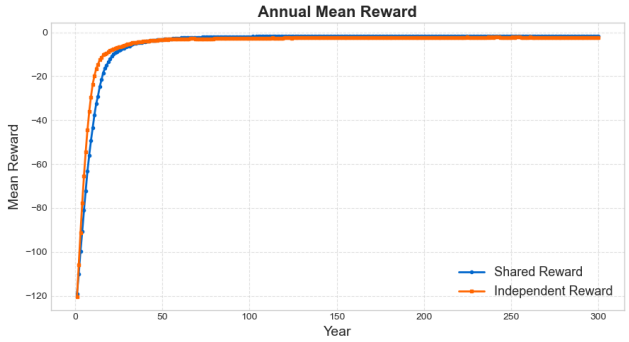
a) Independent vs. Shared Reward Experiments.: We train two variants of the agents:

1) **Independent Rewards:** Each MG i maximizes its own reward r_t^i, based on local costs and generation, without penalties for grid violations. This simulates self-interested control where agents optimize independently, regardless of system-wide constraints.
2) **Shared Reward:** All MGs receive a common reward equal to the sum of individual rewards minus a significant penalty for any constraint violation. This approach encourages cooperative behaviour, as agents are incentivized to maximize overall system performance and safety.

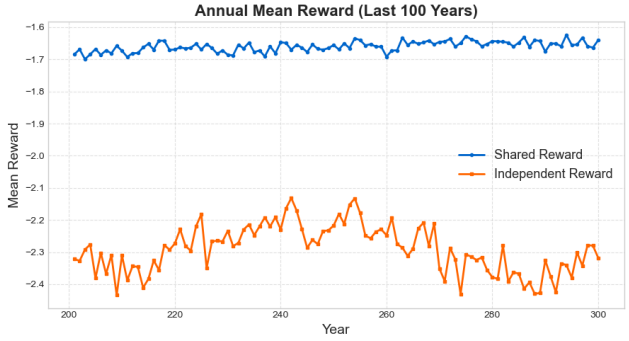
We train each scenario for a total of $8{,}760 \times 300 = 2{,}628{,}000$ time steps. Here, one year of historical hourly data (8,760 hours per year) from 2018–2019 is repeatedly used to create multiple training epochs. The number 300 indicates we loop over the dataset 300 times (300 epochs), allowing the MARL policies sufficient time to converge to stable and robust operational strategies.

VI. Experimental Results and Discussion

1) Training Phase Performance: We first compare the learning curves under independent and shared reward settings. As shown in Figure 1(a), shared-reward agents initially learn more slowly, since they must coordinate and sometimes sacrifice individual gains, while independent agents quickly improve local cost optimization. However, independent agents plateau at a lower reward and experience greater oscillations due to grid stress (see Figure 1(b)). In contrast, shared-reward agents steadily improve and ultimately achieve higher overall rewards, indicating that cooperation leads to more efficient system-wide operation.



(a) Annual mean reward under independent and shared reward schemes (full 300 years).



(b) Annual mean reward under independent and shared reward schemes (last 100 years).

Fig. 1: Comparison of annual mean reward under independent and shared reward schemes: (a) full 300-year horizon; (b) zoom-in on last 100 years.

Figure 2 presents the annual mean collision metric on a logarithmic scale. Both schemes show a decline as agents learn safer behaviours, but independent agents plateau at a higher collision rate with persistent violations. Shared-reward agents, however, achieve a much sharper reduction, with collision metrics dropping by orders of magnitude and nearly disappearing by the end of training. This highlights the effectiveness of shared rewards in minimizing operational collisions.

To further compare performance, Figure 3 shows the frequency of collision events during evaluation. Both scenarios start with frequent collisions, but while the independent case

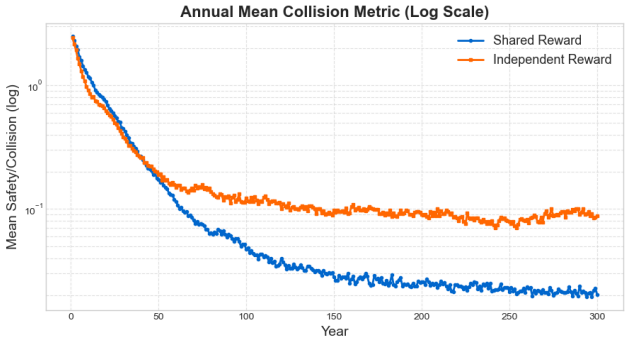


Fig. 2: Annual mean collision metric (log scale) for independent and shared reward.

remains high and even increases later, the shared-reward scenario exhibits a consistent and substantial decline, resulting in rare collisions by the end of training. Table 1 quantifies these results, clearly demonstrating the superior safety and reliability achieved through cooperative, shared-reward strategies.

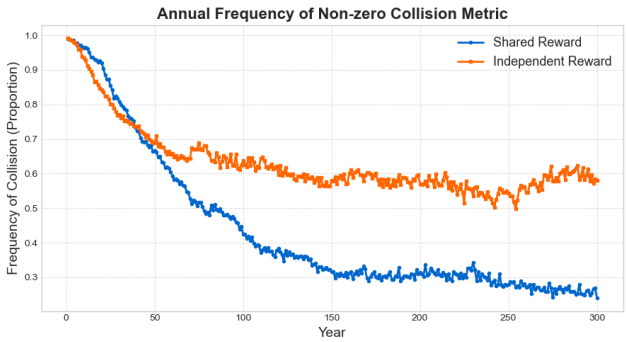


Fig. 3: Annual frequency of non-zero collision metric under different reward strategies.

TABLE I: Collision frequency statistics over the final 100 years of training.

Scheme	Mean Frequency	Std Dev	Min	Max
Shared Reward	0.285	0.024	0.240	0.343
Independent Reward	0.568	0.028	0.497	0.624

2) Agent-Specific Analysis: By analyzing the simulation logs, we quantified each MG's contribution to overall constraint violations, as illustrated by Figure 4. In the independent reward scenario, MG2 emerged as the dominant source of violations, accounting for 78.4% of the total collision metric (see Table II). This high proportion reflects both MG2's placement on a weaker branch of the feeder and its aggressive operational policy, such as charging its battery or increasing demand concurrently with other MGs, which frequently overloaded the shared feeder segment.

MG3 and MG1 also contributed to system risk, with MG3 responsible for 12.7% of violations, often resulting from overvoltage in its local network when exporting power without coordination. MG1 contributed 8.9%, typically due to its own local optimization strategies. While MG2 had the largest impact, the contributions of MG1 and MG3 highlight that operational collisions are fundamentally systemic, arising from the interplay and timing of actions across all agents. Even agents with smaller shares of violations can significantly affect system stability when their behaviours coincide with those of more dominant contributors.

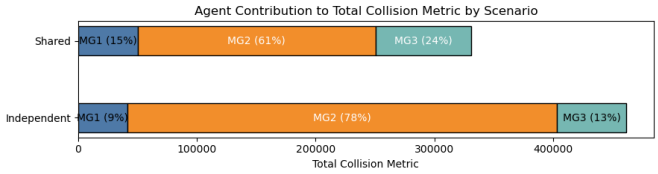


Fig. 4: Stacked bar comparison of total collision metric by agent for independent and shared reward scenarios.

TABLE II: Comparison of agent contributions to total collision metric under independent and shared reward schemes.

Microgrid	Indep. Metric	Indep. %	Shared Metric	Shared %
MG1	41228.27	8.9%	50020.55	15.1%
MG2	361822.59	78.4%	200607.53	60.6%
MG3	58575.85	12.7%	80406.57	24.3%

Under the shared reward scheme, MG2 adopted more moderate behavior, staggering its high usage periods to avoid overlapping with MG1 and MG3 and thus preventing simultaneous demand peaks. MG1 also reduced over-voltage events by limiting generator output when feeder voltage was elevated. As a result, constraint violations became rare, and collision responsibility was nearly eliminated for all agents. The remaining minor voltage excursions were attributable to forecast errors rather than agent missteps. These results highlight how cooperative learning enables agents to implicitly coordinate safe operation, even without direct communication, by aligning incentives through shared rewards.

In summary, our experimental results show:

- Uncoordinated, self-interested AI agents in power networks can cause significant operational safety risks, as evidenced by frequent constraint violations arising from agents optimizing their own objectives.
- The proposed framework provides a systematic way to detect, attribute, and quantify these risks, enabling clear comparison of control strategies and identification of responsible agents.
- Aligning agent incentives through cooperative mechanisms such as a shared reward proved highly effective, nearly eliminating unsafe behaviour and distributing responsibility more evenly among agents.
- Since fully cooperative schemes may not always be practical in real-world multi-stakeholder settings, alternative solutions like grid-aware penalties or supervisory safety controllers should be considered.

VII. Conclusion and Future Work

This paper presents a framework for characterizing and detecting agent-induced operational collisions in multi-agent energy systems. We formalize these collisions in power grids managed by independent AI controllers, introduce metrics to quantify their frequency and severity, and provide methods to attribute events to specific agents. Through an MG case study, we show that uncoordinated agents can inadvertently compromise grid safety, while aligning incentives via cooperative learning can significantly reduce such risks and improve overall system performance. Our results highlight the importance of proactive safety analysis as autonomous devices proliferate. The proposed metrics and methods allow operators and researchers to simulate and identify potential conflicts in advance, supporting safer deployment of multi-agent control in smart grids. While the shared reward strategy significantly improves performance in simulations, real-world deployment faces challenges related to communication, trust, and cybersecurity. Effective implementation requires secure and efficient information sharing, which may be achieved through indirect signals, aggregated feedback, or privacy-preserving methods such as cryptography and differential privacy.

Future work should explore incentive designs and market mechanisms that promote cooperation without extensive data exchange. Additional research will focus on advanced attribution techniques, explicit safety constraints, scaling to larger networks, managing potential collusion, and validating solutions in real-world or hardware-in-the-loop settings. Together, these efforts aim to ensure safe and trustworthy adoption of AI-driven energy management in smart grids.

Acknowledgment

This research is funded in part by the U.S. National Science Foundation under Grant 2330504, the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN-2025-05097, and the Concordia University Research Chair.

References

[1] P. Jamjuntr, C. Techawatcharapaikul, and P. Suanpang, "Adaptive multi-agent reinforcement learning for optimizing dynamic electric vehicle charging networks in thailand," *World Electric Vehicle Journal*, vol. 15, no. 10, p. 453, 2024.

[2] A. Shojaeighadikolaei and M. Hashemi, "An efficient distributed multi-agent reinforcement learning for ev charging network control," in *2023 59th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2023, pp. 1–8.

[3] Y. Zhao, H. Zhong, and C. C. Lim, "Safety-constrained multi-agent reinforcement learning for power quality control in distributed renewable energy networks," *Computers, Materials and Continua*, vol. 79, no. 1, pp. 449–471, 2024.

[4] H. Li, Y. Liu, and J. Yan, "Position: Emergent machina sapiens urge rethinking multi-agent paradigms," *arXiv preprint arXiv:2502.04388*, 2025.

[5] J. Wang, W. Xu, Y. Gu, W. Song, and T. C. Green, "Multi-agent reinforcement learning for active voltage control on power distribution networks," *Advances in Neural Information Processing Systems*, vol. 34, pp. 3271–3284, 2021.

[6] A. Narayan, A. Krishna, P. Misra, A. Vasan, and V. Sarangan, "A dynamic pricing system for electric vehicle charging management using reinforcement learning," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 6, pp. 122–134, 2022.

[7] Y. Zhu and C. Liu, "Mitigating multi-stage cascading failure by reinforcement learning," in *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*. IEEE, 2019, pp. 3724–3728.

[8] N. Brinkel, T. van Wijk, A. Buijze, N. K. Panda, J. Meersmans, P. Markotić, B. van der Ree, H. Fidder, B. de Brey, S. Tindemans *et al.*, "Enhancing smart charging in electric vehicles by addressing paused and delayed charging problems," *Nature Communications*, vol. 15, no. 1, p. 5089, 2024.

[9] M. A. Hady, S. Hu, M. Pratama, J. Cao, and R. Kowalczyk, "Multi-agent reinforcement learning for resources allocation optimization: A survey," *arXiv preprint arXiv:2504.21048*, 2025.

[10] T. Fonseca, L. Ferreira, B. Cabral, R. Severino, and I. Praça, "Energaize: Multi agent deep deterministic policy gradient for vehicle-to-grid energy management," in *2024 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2024, pp. 232–237.

[11] J. Fan, H. Wang, and A. Liebman, "Marl for decentralized electric vehicle charging coordination with v2v energy exchange," in *IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2023, pp. 1–6.

[12] J. Jiang, Y. Li, L. Hou, M. Ghafouri, P. Zhang, J. Yan, and Y. Liu, "Deep reinforcement learning-based bidding strategies for prosumers trading in double auction-based transactive energy market," *arXiv preprint arXiv:2502.15774*, 2025.

[13] J. Guerrero, D. Gebbran, S. Mhanna, A. C. Chapman, and G. Verbič, "Towards a transactive energy system for integration of distributed energy resources: Home energy management, distributed optimal power flow, and peer-to-peer energy trading," *Renewable and Sustainable Energy Reviews*, vol. 132, p. 110000, 2020.

[14] K. Zhang, Z. Yang, and T. Başar, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," *Handbook of reinforcement learning and control*, pp. 321–384, 2021.

[15] Y. Qu, J. Ma, and F. Wu, "Safety constrained multi-agent reinforcement learning for active voltage control," *arXiv preprint arXiv:2405.08443*, 2024.

[16] X. Huo, H. Huang, K. R. Davis, H. V. Poor, and M. Liu, "A review of scalable and privacy-preserving multi-agent frameworks for distributed energy resources," *Advances in Applied Energy*, vol. 17, p. 100205, 2025.

[17] R. Hassan, K. S. Wadith, M. M. o. Rashid, and M. M. Khan, "Depaint: a decentralized safe multi-agent reinforcement learning algorithm considering peak and average constraints," *Applied Intelligence*, vol. 54, no. 8, pp. 6108–6124, 2024.

[18] Y. Liu and J. Li, "Runtime verification-based safe marl for optimized safety policy generation for multi-robot systems," *Big Data and Cognitive Computing*, vol. 8, no. 5, p. 49, 2024.

[19] M. Zhang, N. Li, Y. Chen, J. Li, X.-Y. Zhang, H. Zhao, J. Liu, and W. Chen, "Learning verified safe neural network controllers for multi-agent path finding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 22, 2025, pp. 23 369–23 377.

[20] D. Chen, K. Chen, Z. Li, T. Chu, R. Yao, F. Qiu, and K. Lin, "Powernet: Multi-agent deep reinforcement learning for scalable powergrid control," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1007–1017, 2021.

[21] L. Hou, J. Yan, Y. Wu, C. Wang, and T. Qiu, "Machina economicus: A new paradigm for prosumers in the energy internet of smart cities," *arXiv preprint arXiv:2403.14660*, 2024.

[22] H. Li and H. He, "Optimal operation of networked microgrids with distributed multi-agent reinforcement learning," in *2024 IEEE Power Energy Society General Meeting (PESGM)*, 2024, pp. 1–5.

[23] H. Huang, M. Kazerooni, S. Hossain-McKenzie, S. Etigowni, S. Zonouz, and K. Davis, "Fast generation redispatch techniques for automated remedial action schemes," in *2019 20th International Conference on Intelligent System Application to Power Systems (ISAP)*. IEEE, 2019, pp. 1–8.

[24] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017. [Online]. Available: https://arxiv.org/abs/1707.06347

[25] L. Thurner, A. Scheidler, F. Schäfer, J. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun, "pandapower — an open-source python tool for convenient modeling, analysis, and optimization of electric power systems," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510–6521, Nov 2018.