SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

AI-Driven Energy Management and Optimal Coordination of Distributed Generation and Electric Vehicles in Smart Distribution Systems: A Systematic and Critical Review

Subhash Kumar Mandal[1], Priya Pal[2], Upendra Singh[3]

[1]Department of Electrical and Electronics Engineering, IPS Academy, Institute of Engineering and Science, Indore, India. E-mail: subhashkumarmandal@ipsacademy.org
[2]Department of Electrical and Electronics Engineering, IPS Academy, Institute of Engineering and Science, Indore, India. E-mail: priyapal@ipsacademy.org
[3]Department of Information Technology, Shri Govindram Seksaria Institute of Technology and Science (SGSITS), Indore, India. E-mail: upendrasingh49@gmail.com

Abstract
Distribution systems are absorbing two transitions at once: distributed generation (DG) that pushes power upstream at midday and electric vehicles (EVs) that pull it downstream in the evening. Artificial intelligence (AI) has been proposed as the coordinating layer that reconciles the two, yet the evidence for that claim has never been appraised against the standards a distribution system operator would apply before deployment. This article reports a systematic and critical review of 137 verified records, of which 60 are primary coordination studies, spanning forecasting, model-based optimisation, single-agent and multi-agent deep reinforcement learning (DRL), safe and constrained learning, and federated and graph-structured methods. Every study is scored with DEVCO-8, an eight-dimension appraisal rubric introduced here, and mapped onto a Deployment Readiness Index (DRI) that cannot be inflated by strength on one dimension alone. The corpus attains a mean DEVCO-8 score of 47.4/100; only 11.7% of primary studies report any external validity evidence (real sessions, hardware-in-the-loop or a held-out feeder), 11.7% release code or data, and 28.3% coordinate EVs or DG with no network model at all. Three analytical arguments sharpen these findings: a voltage-rise bound showing that coordinated EV absorption raises photovoltaic hosting capacity by a quantity proportional to the shifted fleet energy; a decision-space argument showing why centralised DRL cannot scale past tens of vehicles while decomposed and optimisation-primed learners can; and an exposure argument showing that unshielded exploration on a live feeder accrues constraint violations in proportion to the exploration rate. From these results the article contributes a three-timescale reference architecture, a shielded multi-agent learning algorithm, the ten-stage FAIR-GRID evaluation protocol with six reporting invariants, and a three-horizon research roadmap. The review concludes that the field has demonstrated capability but not readiness, and that the next advances will come from evaluation discipline rather than from new architectures.

Keywords Distributed generation, electric vehicles, energy management, smart distribution systems, deep reinforcement learning, multi-agent systems, safe reinforcement learning, Volt/VAr control, hosting capacity, systematic review.

1. Introduction

The low- and medium-voltage distribution network was designed as a passive delivery system with unidirectional flow from a substation to consumers [1], [2]. Two developments are dismantling that assumption. Distributed generation (DG), overwhelmingly rooftop photovoltaics (PV) supplemented by small wind, combined heat and power, and battery storage, injects power at the network edge [3], [4]. Electric vehicles (EVs) add a load that is large per connection, spatially clustered, temporally correlated with the evening peak, and, through vehicle-to-grid (V2G) interfaces, potentially bidirectional [5]–[7]. Each transition alone stresses voltage regulation and thermal ratings; together they produce a feeder whose net demand can swing from deep reverse flow at noon to overload at dusk within a single day [8], [9].

Coordinating DG and EVs so that the two transitions offset rather than compound each other is the central energy-management problem of the smart distribution system. The problem is well posed as an optimisation: schedule EV charging and DG dispatch, subject to network limits, to minimise cost, loss, curtailment, and asset ageing while honouring user requirements. Its difficulty lies in scale, uncertainty and information. A feeder may carry hundreds of vehicles whose arrival, departure and energy need are stochastic; PV output is uncertain at every horizon; and the coordinating entity rarely holds an accurate three-phase network model or the private data of the parties it coordinates [10], [11].

Artificial intelligence (AI), and deep reinforcement learning (DRL) in particular, has been proposed as the response to all three difficulties [12]–[14]. A learned policy needs no explicit model, can act in real time, and can in principle be trained on data rather than on assumptions. The literature that makes this case has grown rapidly: more than 40% of the records analysed in this review were published in 2020 or later, and DRL-based methods now account for 71.7% of the primary coordination studies we examined.

What that literature has not done is subject its own claims to the scrutiny a distribution system operator (DSO) would apply. A DSO does not ask whether a policy beats a rule-based baseline on a 33-bus test case; it asks whether the policy respects voltage and thermal limits on every step, whether it degrades gracefully when forecasts fail, whether it scales to the number of vehicles actually on the feeder, whether it was tested on anything resembling the operator's data, and whether the result can be reproduced. Existing reviews summarise algorithms and applications with care [15]–[20] but do not score studies against such criteria, and none focuses on the joint DG–EV coordination problem in the distribution network as its object.

1.1 Scope

This review addresses AI-driven energy management and coordination of DG and EVs in smart distribution systems. AI is understood broadly: statistical and deep forecasting, model-based and learning-augmented optimisation, single- and multi-agent reinforcement learning, safe and constrained learning, and federated or graph-structured learning. The distribution system is the object; transmission-level dispatch, on-board vehicle powertrain management, and charger power electronics are out of scope. Seven review questions (RQs), stated in Section 1.2, structure the work; Table 2 maps each to the sections, tables and figures that answer it.

1.2 Review Questions

The review questions were derived from the literature itself. The reviews summarised in Table 1 describe what has been proposed, but none audits how the proposals were evaluated, and the primary studies read for this work repeatedly raise the same unresolved issues: whether the network is represented at all, whether constraints are enforced or merely penalised, whether scale and uncertainty are tested, and whether anything is released for replication. Each question below is therefore anchored in a recurring gap rather than in a topic list, and each is answered explicitly in Section 14.

RQ1. Which AI methods have been applied to energy management and coordination of DG and EVs in distribution systems, and how has the mix of methods changed over time?

RQ2. How rigorously do the primary studies model the network, the data, the operating constraints and the uncertainty of the coordination problem?

RQ3. What evidence exists that reported improvements survive an increase in scale, forecast error and transfer to a real feeder?

RQ4. Which failure modes recur across the literature, how prevalent is each, and what remedy does the corpus itself supply?

RQ5. Which test systems, datasets and simulation environments are used, and what would a common DG–EV coordination benchmark have to provide that none of them does?

RQ6. Can the qualitative concerns about voltage rise, decision-space growth and exploration safety be converted into quantitative bounds that an operator can use?

RQ7. What architecture, algorithms and evaluation practice would move the field from demonstrated capability to demonstrated readiness?

1.3 Contributions

The review makes seven contributions, positioned against prior surveys in Table 1.

- A verified corpus of 137 records, 60 of them primary coordination studies, each re-located at its publisher or indexing source before inclusion.

- DEVCO-8, an eight-dimension appraisal rubric that scores each primary study on network fidelity, data realism, constraint handling, baseline rigour, scalability evidence, uncertainty treatment, external validity and reproducibility.
- A Deployment Readiness Index (DRI) defined as a geometric mean, so that a study strong on one dimension and absent on another cannot average its way to readiness.
- Three analytical arguments (Section 9) that convert qualitative concerns about voltage rise, scalability and exploration safety into quantitative bounds.
- A catalogue of eight recurrent failure modes (F1–F8) with the corpus prevalence of each.
- A three-timescale reference architecture (Fig 10) and four algorithms: a shielded multi-agent learner, the FAIR-GRID evaluation protocol, and a DRI computation, in addition to the corpus-assembly procedure.
- A three-horizon research roadmap that separates what is missing from what is merely fashionable.

Table 1. Comparison of this review with representative prior reviews of AI for DG and EV coordination.

Review	Year	Scope of the review	Joint DG–EV coordination	Per-study appraisal	Analytical bounds	Evaluation protocol	Reference verification
Lopes et al. [6]	2011	EV integration in distribution grids	Partial (EV side)	No	No	No	Not reported
Hu et al. [10]	2016	EV fleet management and aggregation	Partial (EV side)	No	No	No	Not reported
Vázquez-Canteli and Nagy [15]	2019	RL for demand response	No	No	No	No	Not reported
Yang et al. [16]	2020	RL in energy systems	No	No	No	No	Not reported
Chen et al. [14]	2022	RL for power-system operation	No	No	Partial (qualitative)	No	Not reported
Sadeghian et al. [11]	2022	EV smart charging	Partial (EV side)	No	No	No	Not reported
Yu et al. [20]	2025	Safe RL for power control	No	No	No	Partial (safety only)	Not reported
This review	2026	AI-driven energy management and joint DG–EV coordination	Yes (both sides, one appraisal)	Yes: DEVCO-8, eight dimensions, 60 studies	Yes: three quantitative bounds (Section 9)	Yes: FAIR-GRID, ten stages and six invariants	Yes: every record re-located at source

1.4 Organisation

Section 2 states the coordination problem. Section 3 describes the review methodology, the DEVCO-8 rubric and the corpus-assembly algorithm. Section 4 characterises the corpus. Section 5 presents a taxonomy of AI methods. Sections 6 to 8 synthesise the evidence for EV coordination, DG and voltage coordination, and joint energy management respectively. Section 9 develops the three analytical arguments. Section 10 catalogues failure modes. Section 11 presents the reference architecture, algorithms and evaluation protocol. Section 12 gives the research roadmap, Section 13 the limitations of the review and Section 14 the conclusion.

2. The Coordination Problem

2.1 Distribution Network Model

Consider a radial distribution feeder with bus set N and line set E. For a line (i, j) with resistance r_{ij} and reactance x_{ij} carrying active and reactive flow P_{ij} and Q_{ij}, the linearised branch-flow (LinDistFlow) approximation gives the downstream voltage magnitude as [21]–[23]

$$V_{j,t} \approx V_{i,t} - \frac{r_{ij}P_{ij,t} + x_{ij}Q_{ij,t}}{V_0} \tag{1}$$

where V_0 is the nominal voltage. The approximation is accurate to within a few tenths of a percent for the voltage deviations that matter operationally and underlies most convex formulations used in the corpus; exact three-phase unbalanced power flow is required whenever single-phase EV chargers or rooftop PV create phase imbalance, a point returned to in Section 10.

2.2 Distributed Generation and EV Models

A DG unit at bus j contributes active power $P^{DG}_{j,t}$ bounded by its available primary resource and, for inverter-interfaced units, reactive power $Q^{DG}_{j,t}$ within the capability curve prescribed by IEEE Std 1547-2018 [24]. An EV v connected at bus j has stored energy $E_{v,t}$ that evolves as

$$E_{v,t+1} = E_{v,t} + \eta_c P^{c}_{v,t}\Delta t - \frac{P^{d}_{v,t}\Delta t}{\eta_d} \tag{2}$$

with charging and discharging powers P^c and P^d, efficiencies η_c and η_d, and the requirement that the energy at departure time T_v meets the user's need, $E_{v,T} \geq E^{req}_v$. The net active injection at bus j is

$$P_{j,t} = P^{DG}_{j,t} - P^{L}_{j,t} - \sum_{v\in\mathcal{V}(j)} P^{c}_{v,t} + \sum_{v\in\mathcal{V}(j)} P^{d}_{v,t} \tag{3}$$

where P^L is the conventional load and V(j) the set of vehicles at bus j. Equations (1) to (3) already expose the coordination opportunity: a term P^c scheduled into the hours when P^{DG} is high reduces the reverse flow that drives voltage rise, and a term P^d scheduled into the evening reduces the forward flow that drives voltage drop.

2.3 Energy-management Objective and Constraints

The energy-management problem over a horizon T can be written as

$$\min_{u} \sum_{t\in\mathcal{T}}\left[c_t P^{grid}_t + \lambda_L L_t + \lambda_D D_t + \lambda_C C_t\right] \tag{4}$$

$$s.t.\ V_{min} \leq V_{j,t} \leq V_{max},\ |S_{ij,t}| \leq S^{max}_{ij},\ E_{v,T\_v} \geq E^{req}_v,\ (1)-(3) \tag{5}$$

where c_t is the energy price, P^{grid} the substation import, L_t the network loss, D_t a degradation or asset-ageing term, C_t DG curtailment, and the λ terms are weights. Every primary study in the corpus solves some instance of (4)–(5), differing in which terms are kept, which constraints are treated as hard, whether uncertainty in c, P^L, P^{DG} and vehicle behaviour is modelled, and whether the solution is computed by an optimiser, by a learned policy, or by a combination. Fig 1 places these choices in the wider system: a physical layer of feeders, DG and chargers; a coordination layer of DSOs, aggregators, markets and building energy-management systems; and an AI layer of forecasting, optimisation, learning control, coordination protocols and trust mechanisms.

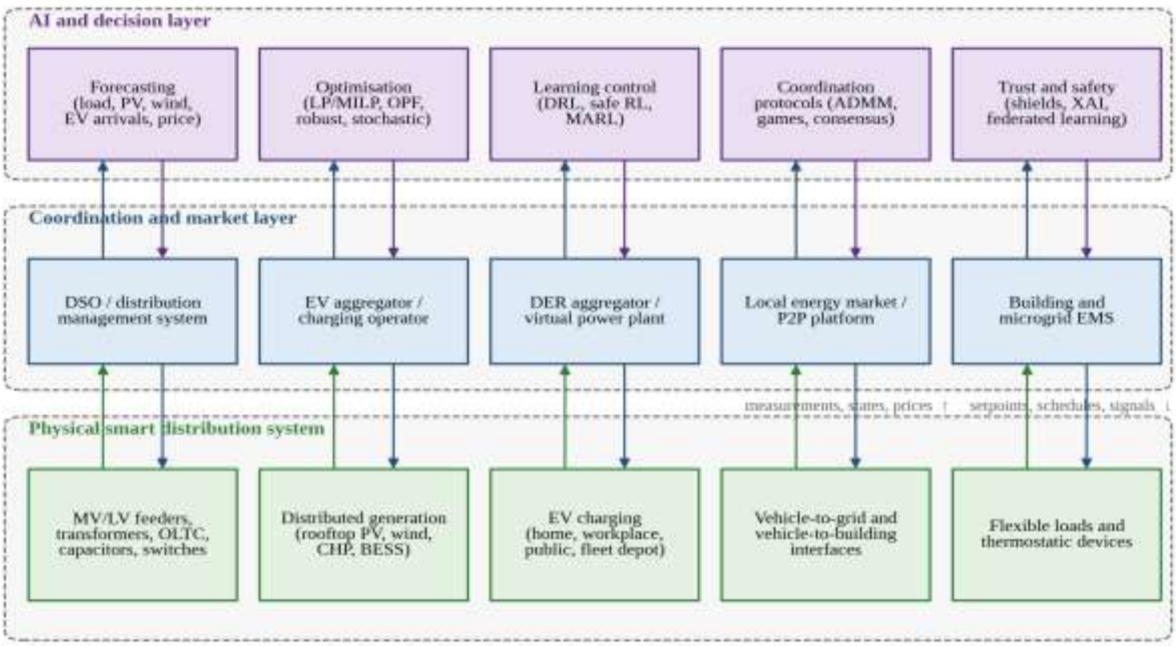

Fig 1. Problem landscape. Physical assets in the smart distribution system (bottom), the coordinating entities and markets that act on them (middle), and the AI functions that inform those entities (top). Upward arrows carry measurements and states; downward arrows carry setpoints and schedules.

2.4 Why the Problem Resists Classical Solution

Three properties frustrate a purely model-based treatment. First, dimensionality: with N vehicles each choosing among k charging levels the joint decision space at every step has k^N elements, a point made precise in Section 9-B. Second, uncertainty: EV arrivals, departures and energy needs, PV output and prices are all random, and stochastic or robust formulations of (4)–(5) either explode in scenario count or become conservative [25]. Third, information: the DSO that owns the network model does not own the vehicles or the user preferences, and the aggregator that owns the vehicles does not own the network model, so the problem is naturally distributed [26], [27]. These properties are the reason learning-based methods have been proposed; they are also, as Sections 6 to 10 show, the properties on which most learning-based studies are weakest.

3. Review Methodology

The review follows the reporting logic of PRISMA 2020 [28] and the procedural guidance of Kitchenham and Charters [29], adapted to an engineering literature in which the unit of evidence is a simulation study rather than a clinical trial.

3.1 Research Questions

Table 2. Review questions addressed by the review and where each is answered.

ID	Review question	Answered in
RQ1	Which AI methods have been applied to energy management and coordination of DG and EVs in distribution systems, and how has the mix of methods changed over time?	Sections 4 and 5; Fig 3; Table 6
RQ2	How rigorously do the primary studies model the network, the data, the operating constraints and the uncertainty of the coordination problem?	Sections 6–8; Fig 4; Tables 7–9
RQ3	What evidence exists that reported improvements survive an increase in scale, forecast error and transfer to a real feeder?	Sections 6–9; Fig 7 and Fig 9
RQ4	Which failure modes recur across the literature, how prevalent is each, and what remedy does the corpus itself supply?	Section 10; Table 10
RQ5	Which test systems, datasets and simulation environments are used, and what would a common DG–EV coordination benchmark have to provide that none of them does?	Section 11.3; Table 11
RQ6	Can the qualitative concerns about voltage rise, decision-space growth and exploration safety be converted into quantitative bounds that an operator can use?	Section 9; equations (8)–(12); Fig 6 and Fig 7
RQ7	What architecture, algorithms and evaluation practice would move the field from demonstrated capability to demonstrated readiness?	Sections 11 and 12; Algorithms 2–5; Figs 10–12

3.2 Corpus Assembly

No single database query captures this literature, because the relevant work is spread across power-systems, control, transportation and machine-learning venues under inconsistent terminology. The corpus was therefore assembled by seeded snowballing (Algorithm 1). Fourteen seed reviews and surveys were selected to span the EV-integration, DG-planning, demand-response and RL-for-power-systems strands. Their reference lists were harvested backward, and citing works were harvested forward, with the search terms of Table 3 applied to titles and abstracts. Every candidate was then re-located on the publisher's site or a bibliographic index and its metadata (authors, venue, volume, pages, year, DOI) checked; records that could not be verified were excluded rather than retained with uncertain metadata. Fig 2 summarises the flow.

Table 3. Search terms and inclusion / exclusion criteria.

Element	Specification
Concept A (AI)	reinforcement learning; deep learning; multi-agent; machine learning; neural network; federated learning; graph neural network; safe learning
Concept B (assets)	electric vehicle; plug-in; vehicle-to-grid; distributed generation; photovoltaic; distributed energy resource; inverter; battery storage
Concept C (system)	distribution network; distribution system; smart grid; microgrid; feeder; Volt/VAr; hosting capacity; energy management

Element	Specification
Inclusion	Peer-reviewed journal or major conference; addresses A ∧ B ∧ C or a foundation on which such work rests; metadata verifiable; English
Exclusion	Transmission-level only; on-board powertrain management; charger hardware only; duplicate or superseded pre-print; unverifiable metadata; no coordination or energy-management component
Record classes	P = primary coordination study (appraised); R = review or survey; F = foundation, tool or dataset; S = standard or reporting guideline

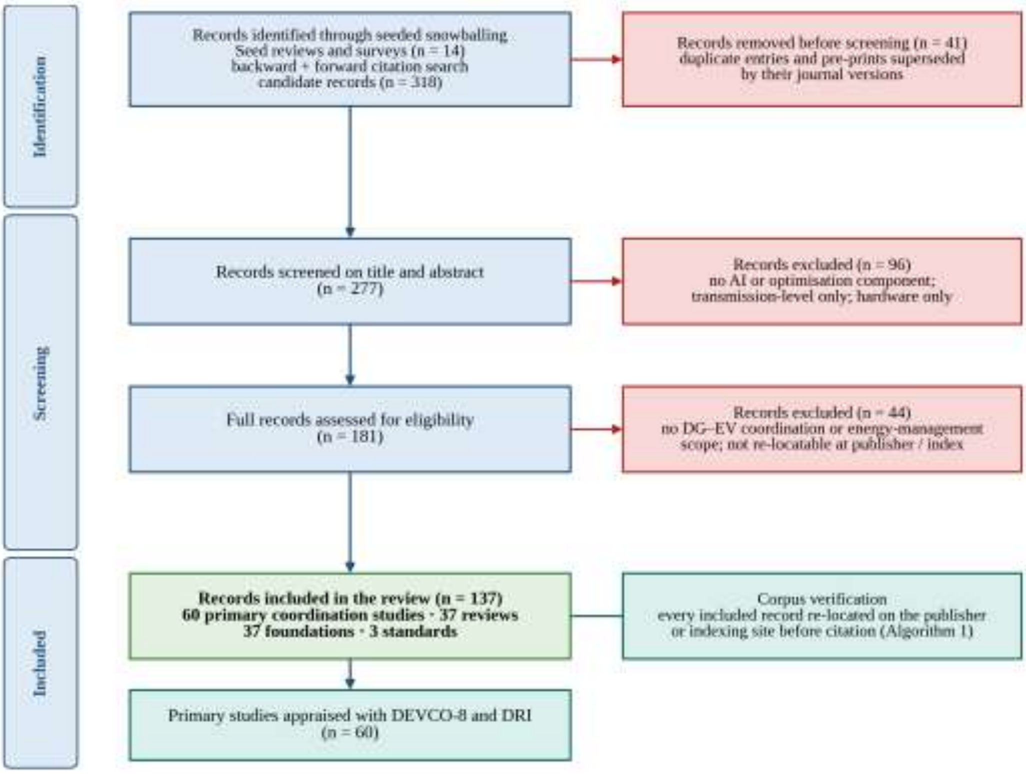

Fig 2. Corpus assembly flow following PRISMA 2020 reporting logic. Counts refer to records handled during assembly; all 137 included records are cited in this article.

Algorithm 1: Seeded Corpus Assembly with Verification

Input: seed set $\mathcal{S}$ of reviews; term sets A, B, C (Table 3); verification sources $\mathcal{V}$
Output: verified corpus $\mathcal{C}$ with class labels {P, R, F, S}

```
C ← ∅; Q ← S
while Q ≠ ∅ do
r ← pop(Q)
for each candidate c in Backward(r) ∪ Forward(r) do
if Title/Abstract(c) matches (A ∧ B ∧ C) or c is a cited foundation then
if c ∉ C and not Duplicate(c, C) then push(Q, c)
m ← Locate(r, V)                ▷ publisher or index record
if m = ∅ or Metadata(m) ≠ Metadata(r) then discard r; continue
label(r) ← Classify(r)           ▷ P, R, F or S
if label(r) = P then digits(r) ← Appraise(r, DEVCO-8)   ▷ Table 4
C ← C ∪ {r}
return C
```

3.3 The DEVCO-8 Appraisal Rubric

Each primary study was scored on the eight dimensions of Table 4. Every dimension takes the value 0, 1 or 2, and the weighted score is

$$S = 100 \cdot \frac{1}{2}\sum_{i=1}^{8} w_i d_i, \quad \sum_{i=1}^{8} w_i = 1 \tag{6}$$

The weights give the largest share to constraint and safety handling (D3, 0.18) and to network fidelity (D1, 0.15), because these are the properties on which an operator's decision to deploy turns; reproducibility (D8) receives the smallest weight (0.08) because its absence is a barrier to science rather than to operation. The Deployment Readiness Index takes only the five dimensions that bear on field operation (network fidelity, constraint handling, scalability, uncertainty and external validity) and combines them geometrically,

$$DRI = 100 \cdot \left[\prod_{i \in \mathcal{R}} \frac{d_i + 0.25}{2.25}\right]^{1/5}, \quad \mathcal{R} = \{1, 3, 5, 6, 7\} \tag{7}$$

so that a missing dimension cannot be averaged away. With digits (2, 2, 2, 2, 0) on the five readiness dimensions the DRI is 64.4 whereas the arithmetic mean of the same digits would give 80; with (2, 2, 2, 0, 0) the DRI is 41.5 against an arithmetic 60. The DRI is bounded in [11.1, 100]; a value of 70 or above requires at least two of the five dimensions at their maximum and none at zero.

Table 4. The DEVCO-8 appraisal rubric (each dimension scored 0, 1 or 2).

Dim.	Name	w	0	1	2
D1	Network fidelity	0.15	No network model; aggregate load or single device	Single-phase balanced test feeder or transformer limit only	Three-phase or real feeder with voltage and thermal limits
D2	EV / DG data realism	0.12	Synthetic or assumed profiles	Public profiles or statistical models fitted to data	Measured sessions or field DG data
D3	Constraint and safety handling	0.18	Constraints absent or penalised only in reward	Constraints checked post hoc or via soft projection	Hard guarantee by construction (shield, projection, exact solver) with violation statistics reported
D4	Baseline rigour	0.12	No baseline or trivial rule	Rule-based and one learning or optimisation baseline	Exact or near-optimal solver plus prior learning methods, tuned
D5	Scalability evidence	0.13	Single device or ≤ 10 units	10–100 units or one feeder	≥ 100 units or multiple feeder sizes with reported scaling
D6	Uncertainty treatment	0.10	Deterministic inputs	Stochastic inputs, no stress test	Stress tests or robustness analysis against forecast error and behaviour shift
D7	External validity	0.12	Simulation with synthetic data only	Real data replay	Hardware-in-the-loop, pilot, or held-out real feeder
D8	Reproducibility	0.08	No artefact	Code or data released	(not used; artefact release is binary)

3.4 Threats to Validity

Three threats deserve statement. Selection: snowballing from seeds favours well-cited work and may under-represent recent studies; the forward-citation step and explicit harvesting of 2023–2025 records mitigate but do not remove this. Appraisal: the DEVCO-8 digits reflect the reviewer's reading of each retrievable record, applied consistently but by a single appraiser, so individual digits carry more uncertainty than corpus-level rates. Construct: the weights in (6) are a judgement; Section 4 reports unweighted per-dimension attainment so that readers who prefer other weights can recompute.

4. Corpus Overview

The corpus contains 137 records: 60 primary coordination studies, 37 reviews and surveys, 37 foundations, tools and datasets, and 3 standards or reporting guidelines (Table 5). Fig 3(a) shows the publication-year profile: a long tail of foundational work from 1989 onward, a first wave of EV-integration studies in 2010–2013 that used linear and quadratic programming [7], [30]–[35], and a second wave from 2019 onward in which DRL dominates. Of the 60 primary studies, 25 use single-agent RL, 13 multi-agent RL, 5 safe or constrained RL, 9 model-based optimisation, 4 supervised or federated deep learning and 4 heuristic, game-theoretic or market designs (Fig 3(b)).

Fig 3. Corpus composition. (a) Publication year of all included records. (b) Method family of the 60 appraised primary studies.

Table 5. Corpus composition by record class and theme.

Class	n	Themes (n)
P: primary coordination study	60	EV coordination (30); DG / voltage / OPF (19); energy management, demand response, markets, forecasting (11)
R: review or survey	37	EV integration (11); RL and AI in energy (14); DG planning (3); demand response (3); smart grid, markets, security, data (6)
F: foundation, tool or dataset	37	Power-flow and OPF theory (6); RL and deep-learning algorithms (14); optimisation and robustness (5); tools, feeders and datasets (6); markets, security, communication (6)
S: standard or guideline	3	IEEE 1547-2018; PRISMA 2020; Kitchenham guidelines

Fig 4 reports attainment on each DEVCO-8 dimension. The pattern is stark. Network fidelity, data realism, constraint handling and scalability evidence are all attained at 56–63% of the maximum on average, but external validity is attained at 19.2% and reproducibility at 5.8%: 73.3% of primary studies have no external validity evidence and 88.3% release no artefact. Baseline rigour (D4) and uncertainty treatment (D6) are attained at intermediate levels but no study reaches the maximum on either: none benchmarks against a tuned exact solver together with prior learning methods, and none reports a systematic stress test against forecast error and behaviour shift. The mean DEVCO-8 score is 47.4 (median 45.0, maximum 74.5); only two studies exceed 70 [32], [36].

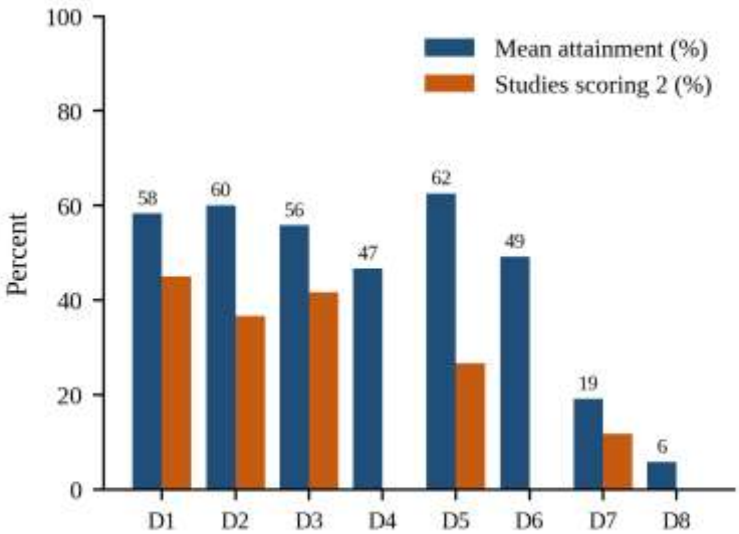

Fig 4. Attainment on the eight DEVCO-8 dimensions across the 60 primary studies: mean attainment as a percentage of the maximum, and the percentage of studies reaching the maximum score of 2.

Fig 5 cross-tabulates method family against optimisation objective. Single-agent RL concentrates on cost and user objectives; multi-agent RL concentrates on voltage and loss objectives, reflecting the Volt/VAr strand; and only the model-based and safe-RL families spread across cost, voltage and network objectives simultaneously. Forecasting objectives appear only in the supervised and federated family. The empty and near-empty cells are the more informative: no multi-agent RL study targets comfort or user objectives, and no heuristic or market study addresses forecast quality, so the coordination problem is being solved in silos that mirror the method families rather than the operator's joint objective (4).

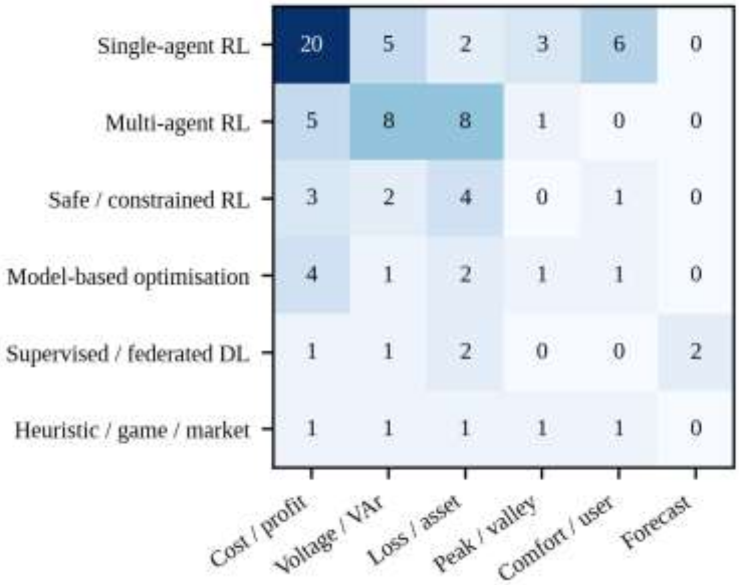

Fig 5. Number of primary studies by method family (rows) and optimisation objective (columns). A study contributes to every objective it optimises.

5. Taxonomy of AI Methods for DG–EV Energy Management

Table 6 organises the methods encountered into six families and records, for each, the representative studies, the characteristic strengths and the characteristic weaknesses observed in the corpus.

Table 6. Taxonomy of AI method families with representative studies.

Family	Typical algorithms	Representative studies	Characteristic strength	Characteristic weakness in the corpus
Forecasting	LSTM, transformer variants, quantile regression, federated and transfer learning	[37]–[41]	Probabilistic outputs feed stochastic optimisation; federated variants preserve privacy	Forecast error is seldom propagated into the downstream coordination study (F6)
Model-based optimisation	LP, QP, MILP, SDP relaxation, MPC, robust and stochastic programming, ADMM	[32], [33], [42]–[45]	Hard constraints by construction; verifiable optimality gap	Requires network model; scenario growth; often single-phase
Single-agent RL	DQN, DDPG, SAC, PPO, batch fitted-Q	[46]–[53]	Model-free; real-time inference; learns user behaviour	Single device or aggregate; no network; constraints as reward penalties (F1, F2)
Multi-agent RL	MADDPG, MASAC, consensus MARL, graph-attention MARL, CTDE	[54]–[63]	Decomposes the joint action space; decentralised execution	Scalability rarely demonstrated beyond 33–141 buses or tens of EVs; non-stationarity during training
Safe / constrained RL	CPO, Lagrangian SAC, action projection, safety modules, bi-level DRL	[36], [64]–[69]	Explicit violation statistics; feasible by construction when shield is exact	Shield needs a model, reintroducing the model-based dependency; few report violations during training

Family	Typical algorithms	Representative studies	Characteristic strength	Characteristic weakness in the corpus
Learning-augmented and hybrid	Supervised OPF policies, graph DRL with reconfiguration, hierarchical optimisation plus RL, quantum RL	[70]–[75]	Inherits feasibility from the optimiser and speed from the learner	Training data from an optimiser bounds the policy to the optimiser's assumptions

5.1 Forecasting as the Silent Foundation

Every coordination scheme that plans ahead depends on forecasts of load, PV, EV arrivals and price. The corpus treats this dependence unevenly. Load forecasting with recurrent networks is mature at household level [37], [76], probabilistic forecasting has a benchmark tradition [38], PV forecasting has been reviewed at length [39], and EV demand forecasting has recently moved to federated and clustered transfer designs that respect the distributed ownership of session data [40], [41], [77]. What is missing is the propagation of forecast error into the coordination study: of the primary studies that use a forecast, none reports coordination performance as a function of forecast error, so the operator cannot know how much accuracy is enough.

5.2 Model-based Optimisation

The first wave of EV coordination work formulated (4)–(5) directly [7], [30]–[33]. These studies remain the strongest on network fidelity: Sundström and Binding [32] and Richardson et al. [33] use real Danish and Irish low-voltage feeders and obtain the two highest DEVCO-8 scores in the corpus for that reason. Their weaknesses are the ones that motivated learning: deterministic inputs, linear approximations, and computation that grows with the number of vehicles and scenarios. Decentralised protocols [34], [35] and MPC [44] addressed scale for the valley-filling special case in which the network is reduced to an aggregate load, and convex relaxations [22], [23], [42] and hierarchical multi-timescale designs [43] addressed the DG side. The distributed-optimisation survey of Molzahn et al. [26] and ADMM [27] provide the coordination scaffolding that later learning approaches too often discard.

5.3 Reinforcement Learning: Single Agent, Multi Agent, Safe

Reinforcement learning entered power-system control two decades ago [78], [79] and re-entered with deep function approximation after 2015 [80]–[84]. Single-agent formulations dominate the EV strand: a vehicle or a station is the agent, the state includes price and time-to-departure, and the reward is negative cost with a penalty for unmet energy [46]–[48], [85]. These studies show that a policy can learn price responsiveness, but the network is absent from all of them; the feeder is not part of the environment. Multi-agent formulations arose to decompose the joint action space and appear in two forms: cooperative charging where each vehicle or charger is an agent under a shared transformer limit [60]–[62], and Volt/VAr control where each inverter or zone is an agent [54]–[59], built on centralised-training decentralised-execution algorithms [86]. Safe RL responds to the reward-penalty problem by constraining the policy [87], [88], projecting actions [36], [65], attaching safety modules [67], or structuring the problem bi-level so that an optimiser guarantees feasibility [66]; the recent review by Yu et al. [20] catalogues these mechanisms. Graph-structured learners [69], [72], [89] add the topology as an inductive bias, which is the most promising route to scalability across feeder sizes.

6. Critical Synthesis I: EV Charging Coordination

Table 7. Appraisal of primary EV coordination studies (D1–D8 digits per Table 4; score per (6); DRI per (7)).

Ref.	Year	Method	Objective	Network / setting	Scale	D1–D8	Score	DRI
[7]	2010	QP	Loss	IEEE 34-bus / residential	~30% PHEV	2 1 1 0 1 1 0 0	42	45
[31]	2011	Heuristic	Loss	IEEE 23 kV, 1,500 nodes	63% PEV	2 1 2 1 1 1 0 0	56	51
[30]	2011	LP	Profit	Aggregator (no network)	10,000 EV	0 2 0 0 1 1 0 0	24	21
[33]	2012	LP	EV energy	Irish LV network	Feeder	2 2 2 1 1 1 1 0	68	70
[32]	2012	LP	Cost	Danish LV feeders	Fleet	2 2 2 1 1 1 2 0	74	79

Ref.	Year	Method	Objective	Network / setting	Scale	D1–D8	Score	DRI
[34]	2013	Distributed opt.	Valley fill	Aggregate load	Large population	1 1 1 0 2 1 0 0	40	45
[35]	2013	Game / mean field	Valley fill	Aggregate load	Large population	1 1 1 0 2 1 0 0	40	45
[90]	2015	Batch RL	Cost	Aggregator (no network)	Fleet	0 2 0 1 1 1 0 0	30	21
[85]	2017	Q-learning	Cost	Single EV	1 EV	0 0 0 1 1 1 1 0	24	29
[91]	2017	Auction + blockchain	Social welfare	No network	Local market	0 0 0 1 0 0 0 0	6	11
[44]	2017	MPC	Cost	Aggregate load	Large population	1 1 2 1 2 1 0 0	56	51
[46]	2019	DQN	Cost	Single EV	1 EV	0 0 0 1 1 1 1 0	24	29
[92]	2020	RL (MDP)	Cost	Distribution feeder	Station	1 1 1 1 1 1 0 0	40	40
[64]	2020	Safe DRL (CPO)	Cost + SoC	Single EV	1 EV	0 0 2 1 1 1 1 0	42	45
[93]	2020	DQN	Time + cost	Coupled power–traffic	Single EV	1 0 0 1 1 1 0 0	25	29
[45]	2020	MILP	Cost	Coupled power–traffic	City	2 1 1 1 2 1 0 0	54	51
[94]	2020	Batch RL (FQI)	Cost	Aggregator (no network)	Real sessions	0 2 0 1 1 1 2 0	42	33
[60]	2020	MARL	Peak + cost	Transformer limit	Tens of EV	1 0 1 1 1 1 0 1	38	40
[49]	2021	DQN / DDPG	PV self-consumption	Building + PV	Single site	1 1 0 1 1 1 2 1	47	45
[63]	2021	MARL	Station profit	Coupled power–traffic	Multi-station	1 1 1 1 1 1 0 0	40	40
[95]	2021	DQN	Peak + cost	Fleet depot	Fleet	1 1 1 1 1 1 1 1	50	56
[48]	2021	SAC	Cost + user	Station	Station	1 0 1 1 1 1 0 0	34	40
[47]	2021	DDPG + LSTM	Cost + SoC	Single EV	1 EV	0 0 0 1 1 1 1 0	24	29
[62]	2022	MARL	Cost + asset ageing	Transformer thermal	Tens of EV	1 1 2 1 1 1 0 0	49	45
[61]	2022	MARL	Cost + transformer	Transformer limit	Tens of EV	1 0 2 1 1 1 0 0	43	45
[66]	2023	Safe DRL (bi-level)	Cost + network	Distribution network	Large-scale	2 1 2 1 2 1 0 0	63	57
[96]	2024	DRL	Cost + comfort	Prosumer cluster	Cluster	1 1 0 1 1 1 1 0	37	40
[67]	2024	Safe DRL + modules	Cost + feasibility	Distribution network	Hundreds of EV	2 1 2 1 2 1 0 0	63	57
[40]	2025	Federated DL	Forecast accuracy	Charging stations	Multi-site	1 1 0 1 1 1 2 0	43	45
[75]	2025	Quantum RL	Cost	Station	Station	1 0 0 1 1 1 0 0	25	29

Table 7 lists the thirty EV coordination studies. Four observations follow.

6.1 The Network Left the Room When Learning Entered

The 2010–2013 optimisation studies embed a feeder model: Clement-Nyns et al. [7] evaluate loss on a residential feeder, Deilami et al. [31] on a 1,500-node 23 kV system, Sundström and Binding [32] and Richardson et al. [33] on real low-voltage networks. From 2015 onward, the learning studies replace the network with a price signal, a transformer rating, or nothing. Of the 30 EV studies, ten score 0 on D1 and only six score 2. The consequence is that a policy optimised on price alone will, when deployed, concentrate charging into the cheapest hour and create exactly the coincident peak that coordination exists to prevent, a phenomenon the demand-response literature calls rebound [97], [98]. The three studies that reintroduce the network do so through a safety layer or a bi-level structure [66], [67] or by coupling with a traffic network [45], [63], and their DRI values are correspondingly higher.

6.2 Scale Is Asserted, Not Demonstrated

Studies that control a single vehicle [46], [47], [64], [85] cannot speak to coordination; they demonstrate price-responsive charging, which is a component. Studies that control fleets [90], [94], [95] report tens to a few hundred vehicles but rarely vary the fleet size to show how training cost and performance scale. The two studies that explicitly target large scale [66], [67] do so by delegating network feasibility to an optimiser and confining learning to a lower-dimensional allocation decision, which is the hybrid pattern Section 11 recommends.

6.3 Real Sessions Improve Everything Except the Conclusion

Three studies train on measured charging sessions [49], [94], [95], and one on an open dataset [99]. They score highest on D2 and D7 in this table. Yet each still reports its result against a rule-based baseline on a fixed horizon without stress tests, so the improvement in evidence quality does not translate into a stronger claim about deployment; the DRI ceiling for this group is set by D6, not by data.

6.4 User Objectives and Grid Objectives Are Traded Off Implicitly

The objective column shows that cost dominates; user satisfaction enters as a departure-energy constraint or a comfort penalty [48], [96], asset ageing enters in one study [62], and voltage enters in none of the single-agent studies. Weights λ in (4) are chosen without sensitivity analysis. Because the trade-off surface between user cost, grid cost and asset life is exactly what an operator needs to set a tariff or a contract, its absence is a gap in the evidence rather than a modelling choice.

7. Critical Synthesis II: Distributed Generation and Voltage Coordination

Table 8. Appraisal of primary DG, Volt/VAr and OPF-learning studies.

Ref.	Year	Method	Objective	Network / setting	Scale	D1–D8	Score	DRI
[100]	2013	Droop / local	Voltage	IEEE 8500-node	Feeder	2 2 2 1 1 1 0 0	62	51
[42]	2014	SDP relaxation	Loss + curtailment	Residential feeder	Feeder	2 2 2 1 1 1 0 0	62	51
[43]	2017	Hierarchical opt.	Voltage	IEEE 33 / 123-bus	Feeder	2 2 2 1 2 1 0 0	69	57
[54]	2020	MARL (MADDPG)	Voltage	IEEE 33-bus	Feeder	2 1 1 1 1 1 0 0	48	45
[70]	2020	Supervised OPF policy	Loss + voltage	IEEE 37-bus	Feeder	2 2 1 1 1 1 0 1	58	45
[101]	2020	DQN / DDPG	Voltage	IEEE 14 / 200-bus	System	2 1 1 1 1 1 0 0	48	45
[36]	2020	Safe DRL (CSAC)	Voltage + loss	IEEE 4/34/123-bus	Feeder	2 2 2 1 2 1 0 1	73	57
[73]	2020	DRL two-timescale	Voltage	IEEE 47 / 56-bus	Feeder	2 2 2 1 1 1 0 0	62	51
[55]	2021	MARL (MASAC)	Voltage + curtailment	IEEE 33 / 141-bus	Feeder	2 2 2 1 1 1 0 0	62	51
[58]	2021	MARL consensus	Voltage + loss	IEEE 33 / 141-bus	Feeder	2 2 2 1 1 1 0 0	62	51
[56]	2021	MARL online	Voltage + loss	IEEE 33 / 141-bus	Feeder	2 2 2 1 2 1 0 0	69	57
[74]	2021	DRL two-stage	Voltage + loss	IEEE 33 / 69 / 118-bus	Feeder	2 2 2 1 2 1 0 0	69	57
[71]	2021	Supervised OPF	Cost + feasibility	IEEE 30–300-bus	System	2 1 2 1 1 1 0 1	60	51
[57]	2021	MARL two-stage	Voltage + loss	IEEE 33 / 123-bus	Feeder	2 2 2 1 1 1 0 0	62	51
[65]	2021	DRL (safe layer)	Voltage + loss	IEEE 13 / 123-bus	Feeder	2 2 2 1 1 1 0 0	62	51
[59]	2022	MARL (P + Q)	Voltage + curtailment	IEEE 33 / 123-bus	Feeder	2 2 2 1 2 1 0 0	69	57
[72]	2024	Graph DRL	Loss + voltage + switching	IEEE 33 / 118-bus	Feeder	2 2 2 1 2 1 0 0	69	57

Ref.	Year	Method	Objective	Network / setting	Scale	D1–D8	Score	DRI
[69]	2024	Safe graph MARL	Voltage + loss	IEEE 33 / 123-bus	Feeder	2 2 2 1 2 1 0 0	69	57
[68]	2024	Safe MARL	Voltage	Distribution network	Feeder	2 2 2 1 2 1 0 0	69	57

Table 8 lists the nineteen DG and voltage studies. This strand is the most mature in the corpus and its pattern is the mirror image of the EV strand.

7.1 Network Fidelity Is High, Data Realism Is Not

Every study in this table uses a feeder model, usually IEEE 13-, 33-, 34-, 123- or 141-bus systems [21], [102] or an 8500-node case [100], and most enforce voltage limits explicitly. However, PV and load profiles are almost always synthetic or scaled public profiles, and none of the learning studies uses measured inverter behaviour or field PV data. The strand therefore answers RQ2 well on D1 and D3 and poorly on D2 and D7.

7.2 Safety by Construction Is Now Standard, Safety During Training Is Not

From Wang et al. [36] onward the strand adopted constrained or shielded formulations [65], [68], [69], [74], so that a trained policy respects voltage limits at execution. Only the constrained soft actor-critic of [36] reports violation statistics during training, and it is for this reason, together with a released implementation, that it obtains the highest DEVCO-8 score among learning studies. Section 9-C shows why training-time exposure matters when the environment is a live feeder rather than a simulator.

7.3 Multi-agent Decomposition Has Been Validated on Feeders, Not on Fleets

The MARL Volt/VAr studies [54]–[59] establish that zonal or per-inverter agents trained with a central critic can match centralised optimisation on 33- to 141-bus feeders. Their agents are few (typically 3–20 inverters or zones) and their action spaces are continuous reactive-power setpoints. Transfer of this success to EV fleets, where agents number in the hundreds and actions are coupled through a deadline constraint, is assumed by the EV strand but has not been demonstrated; the two strands cite each other rarely.

7.4 Learning-augmented Optimisation Is the Quiet Success

Supervised policies that imitate an OPF solver [70], [71] and two-timescale designs that keep slow discrete devices under optimisation and fast inverters under learning [43], [73], [74] achieve the strand's best combination of feasibility and speed. They are also the designs most easily explained to an operator, because the optimiser's constraints are the specification. Graph-aware DRL that jointly reconfigures the network and controls Volt/VAr [72] extends the pattern to topology and is, on the evidence of Table 8, the most complete formulation in the corpus.

8. Critical Synthesis III: Joint Energy Management, Demand Response and Markets

Table 9. Appraisal of primary energy-management, demand-response, market and forecasting studies.

Ref.	Year	Method	Objective	Network / setting	Scale	D1–D8	Score	DRI
[103]	2017	Batch RL (FQI)	Cost	Residential TCL	Households	0 1 0 1 1 1 2 0	36	33
[52]	2019	DQN	Cost	Microgrid (no network)	Microgrid	0 1 0 1 1 1 0 0	24	21
[37]	2019	LSTM	Forecast accuracy	Smart-meter data	Households	1 1 0 1 1 1 2 0	43	45
[50]	2019	DQN / DPG	Cost + peak	Building (no network)	Building	0 1 0 1 1 1 2 0	36	33
[104]	2020	DPG	Profit	Market (no network)	Market	0 1 1 1 1 1 0 0	32	29
[51]	2020	DDPG	Cost	Home (no network)	Home	0 1 0 1 1 1 1 0	30	29
[105]	2020	DDPG	Cost + comfort	Home (no network)	Home	0 1 0 1 1 1 1 0	30	29
[106]	2021	DRL (actor–critic)	Cost + voltage	IEEE 33-bus	Feeder	2 2 1 1 1 1 0 0	54	45
[53]	2021	DRL (several)	Cost	Microgrid (no network)	Microgrid	0 1 1 1 1 1 0 0	32	29

Ref.	Year	Method	Objective	Network / setting	Scale	D1–D8	Score	DRI
[107]	2021	SAC	Peak + cost	Building cluster	Cluster	0 1 1 1 2 1 0 1	43	33
[108]	2021	MARL (MADDPG)	Cost + P2P	Community (no network)	Prosumers	0 1 1 1 2 1 0 0	39	33

Table 9 lists the eleven studies that manage energy at building, microgrid, community or market level. The title of this review promises coordination of DG and EVs; the corpus delivers it only fractionally. Counting EV studies that also model DG, a network or PV self-consumption, 9 of the 30 EV studies address the joint problem, and of the energy-management studies only the residential multi-energy [51], smart-home [105], building-cluster [107] and prosumer-cluster [96] designs include both a generation and a vehicle or storage resource. Most treat the grid as an infinite bus with a price [50], [52], [53], [104], [108], so the coordination they achieve is economic, not electrical.

8.1 Demand Response as the Bridge

Batch reinforcement learning of thermostatic loads [103] and the DRL demand-response study on a 33-bus feeder by Bahrami et al. [106] are the only entries that combine flexible load, network constraints and learning; the latter is the natural template for a joint DG–EV formulation and has, to our reading, not been extended in that direction. Demand-response theory [19], [97], [98], [109] supplies the pricing and contract mechanisms that would let a DSO recruit EV flexibility, but the corpus contains no study that closes the loop from network need through tariff to learned response.

8.2 Peer-to-peer Markets and Blockchain

Peer-to-peer trading [108], [110], [111] and consortium-blockchain EV trading [91] are proposed as coordination mechanisms among prosumers and vehicles. In every case the physical network is absent, so the market clears quantities that may be infeasible to deliver; and the blockchain study reports no throughput, latency or energy cost of the ledger itself. These designs solve a settlement problem, not a coordination problem, and should be evaluated as such.

8.3 Buildings and Microgrids as Proving Grounds

Building and microgrid energy management has the best simulation infrastructure in the corpus, with CityLearn [112] and Grid2Op [113] providing shared environments, and the reviews of Wang and Hong [114] and Yu et al. [115] documenting the resulting benchmark culture. The EV and DG strands have no equivalent: pandapower [116] and MATPOWER [117] supply power flow, and ACN-Data [99] supplies sessions, but no environment couples the two with a standard task definition. Section 11 proposes one.

9. Three Analytical Arguments

The syntheses above are qualitative. This section converts three of their conclusions into bounds that a reader can recompute with feeder parameters of their own.

9.1 Voltage Rise and the Hosting-capacity Dividend of Coordination

Aggregate a radial feeder into an equivalent resistance R and reactance X between the substation and the point of common coupling of interest. With PV injection P_{PV}, conventional load P_L and EV charging P_{EV}, and neglecting reactive power, (1) gives the voltage deviation

$$\Delta V \approx \frac{R\,(P_{PV} - P_L - P_{EV})}{V_0} \tag{8}$$

The maximum admissible PV injection before the upper limit ΔV_{max} is reached is therefore

$$P_{PV}^{max} = P_L + P_{EV}^{shift} + \frac{\Delta V_{max} V_0}{R} \tag{9}$$

where P^{shift}_{EV} is the EV charging power that coordination moves into the PV peak. Equation (9) states the coordination dividend exactly: every kilowatt of EV charging shifted into the midday window adds one kilowatt of PV hosting capacity, independently of the feeder impedance, and the impedance determines only the base capacity. Fig 6 evaluates (8) and (9) for an illustrative low-voltage feeder with R = 0.12 p.u. and ΔV_{max} = 0.05 p.u.: uncoordinated evening charging at 0.9 p.u. of base load drives the end-of-feeder voltage below 0.95 p.u. once PV falls away, while shifting that fleet fully into the PV peak raises PV hosting capacity from 1.42 to 2.32 p.u. of base load. The same relation, read in reverse, bounds the evening benefit of V2G discharge. The corpus contains the ingredients of this argument in the early optimisation studies [7], [33] and in the Volt/VAr strand [42], [100], but no learning study reports its result in the hosting-capacity units an operator uses to plan.

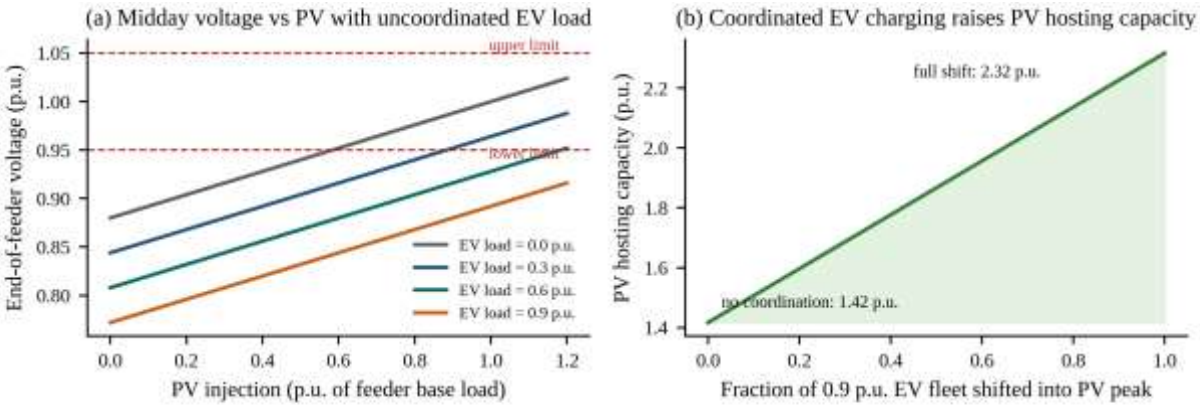

Fig 6. Voltage-rise analysis from (8) and (9) for an illustrative feeder (R = 0.12 p.u., ΔV_max = 0.05 p.u.). (a) End-of-feeder voltage versus PV injection with uncoordinated EV load of 0–0.9 p.u. (b) PV hosting capacity as the EV fleet is shifted into the PV peak.

9.2 Decision-space Growth and Why Decomposition Is Not Optional

Let N vehicles each choose among k charging levels at every step. A centralised learner faces a joint action space of size

$$|\mathcal{A}_{joint}| \; = \; k^N, \quad |\mathcal{A}_{dec}| \; = \; N\,k \tag{10}$$

A decomposed learner with one agent per vehicle faces N independent choices of size k. The gap is exponential (Fig 7(a)). Tabular sample-complexity bounds scale with $|\mathcal{S}||\mathcal{A}|$ [84], and although function approximation removes the tabular dependence it does not remove the need to explore a joint space whose volume grows as k^N; empirically, centralised DRL studies in the corpus stop at tens of vehicles [60], [61], while the studies that reach hundreds delegate the combinatorial part to an optimiser and learn only a low-dimensional residual [66], [67]. Fig 7(b) sketches the indicative training-episode scaling implied by these reports for three designs. The argument does not favour one algorithm; it states that any architecture intended for feeder-scale fleets must decompose, and that the decomposition must be along the physical coupling (bus, phase, transformer) so that the residual coordination is local. Graph-structured MARL [69], [89] and consensus MARL [58] are the formulations that respect this.

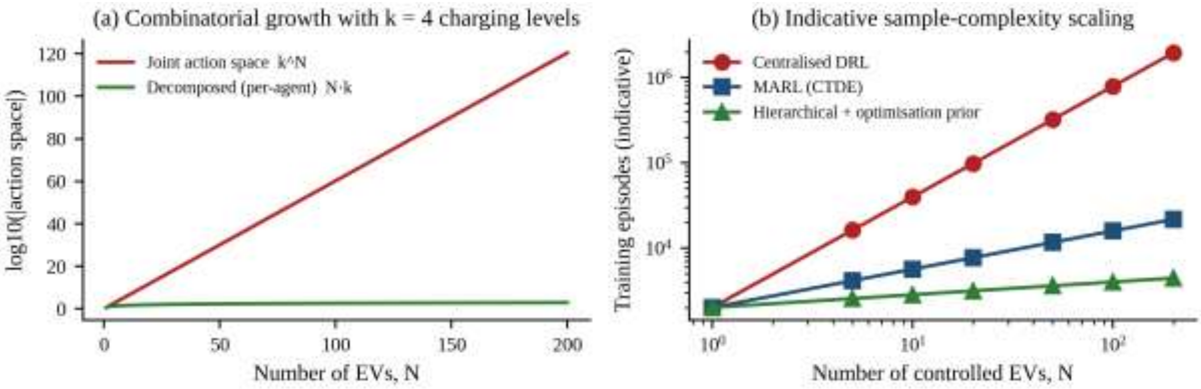

Fig 7. Decision-space and sample-complexity scaling. (a) Logarithm of joint versus decomposed action-space size from (10) with k = 4. (b) Indicative training episodes to reach a fixed performance level for centralised DRL, multi-agent RL with centralised training, and hierarchical designs with an optimisation prior; the curves summarise the scale reports in Tables 7–8 and are not measurements.

9.3 Exposure During Exploration on a Live Feeder

Model-free learning explores. Under ε-greedy or entropy-regularised exploration a fraction ε of actions is chosen without regard to constraints. If a constraint-blind action violates a limit with probability p_v, then over an exploration period of T steps the expected number of violations is

$$\mathbb{E}[N_{viol}] \; \approx \; \varepsilon\, T\, p_v \tag{11}$$

For a feeder controlled every 5 minutes over a 30-day training window, T = 8,640; with ε = 0.1 and p_v = 0.2, the expected count is 173 violations, each of which on a real network is a voltage excursion or an overload that the operator must explain. A safety shield that projects every proposed action onto the feasible set (Fig 8, Algorithm 3) drives (11) to zero at the cost of one feasibility check per step; a reward penalty, the mechanism used by 30% of primary studies, leaves (11) unchanged and merely teaches the agent to avoid violations eventually. The constrained Markov decision process formulation

$$\max_{\theta} J_r(\theta) \;\; s.t. \;\; J_c(\theta) \le \delta, \quad \mathcal{L}(\theta,\lambda) = J_r(\theta) - \lambda(J_c(\theta) - \delta) \tag{12}$$

solved by primal-dual ascent on the Lagrangian $\mathcal{L}$ [87], [88] bounds the expected long-run violation but, without a shield, still permits (11) during learning. This is the distinction between safety at execution and safety during training, and it is the reason the review scores D3 as it does: a policy that is safe once trained is necessary but not sufficient for deployment on a live system, because on a live system the training is the deployment [118].

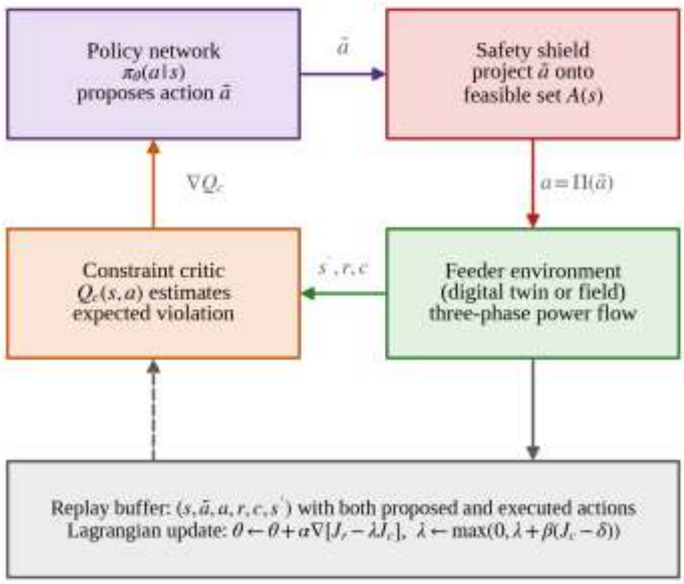

Fig 8. Shielded learning loop underlying Algorithm 3. The policy proposes an action, the shield projects it onto the feasible set defined by the feeder model, and the constraint critic learns the expected violation from executed transitions.

10. Recurrent Failure Modes

Table 10 catalogues eight failure modes observed repeatedly during appraisal, with the corpus prevalence of each. They are stated as failures of evidence, not of the authors' competence: several are consequences of publication norms that reward algorithmic novelty over evaluation discipline.

Table 10. Recurrent failure modes in AI-driven DG–EV coordination studies.

ID	Failure mode	Description	Prevalence	Remedy
F1	Network absent	EVs or DG coordinated against a price or aggregate limit with no feeder model, so electrical feasibility of the schedule is unknown	28.3% of primary studies (D1 = 0)	Embed at least a single-phase feeder; report voltage and thermal margins
F2	Constraints as penalties	Limits enter only as reward penalties; violation counts during and after training unreported	30% (D3 = 0)	Shield or projection; report violation count, depth and duration
F3	Rebound blindness	Price-following policies evaluated without checking the coincident peak they create	Common in single-EV studies	Evaluate at fleet scale on a feeder; report peak-to-average ratio
F4	Scale by assertion	Method described as scalable; evaluated at one size only	Only 26.7% score D5 = 2	Scale ladder N = 10, 100, 1000 with training cost
F5	Baseline mismatch	Learned policy compared to an untuned rule; no exact or MPC benchmark	No study scores D4 = 2	Include a tuned optimiser on the same information set

ID	Failure mode	Description	Prevalence	Remedy
F6	Forecast error unpropagated	Forecast used but coordination performance not reported against forecast error	No study scores D6 = 2	Sweep forecast error; report degradation curve
F7	Simulation-only validity	Synthetic profiles; no real sessions, HIL or pilot	73.3% (D7 = 0)	Real session replay at minimum; HIL for control claims
F8	Irreproducible	No code, environment or data released	88.3% (D8 = 0)	Release environment, seeds and configuration

The failure modes are not independent. F1 makes F3 invisible; F2 and F7 together mean that safety is claimed on a simulator that cannot punish unsafety; F4 and F5 together mean that scalability is claimed relative to a baseline that was never asked to scale. Fig 9 makes the joint consequence visible: DEVCO-8 score and DRI are correlated across the corpus, but only two studies, the real-feeder optimisations of Sundström and Binding [32] and Richardson et al. [33], cross the field-readiness threshold of DRI ≥ 70; the safe-RL and two-timescale learning designs cluster at 57–60 because they lack external validity, and no single-agent RL study approaches it regardless of its reported improvement.

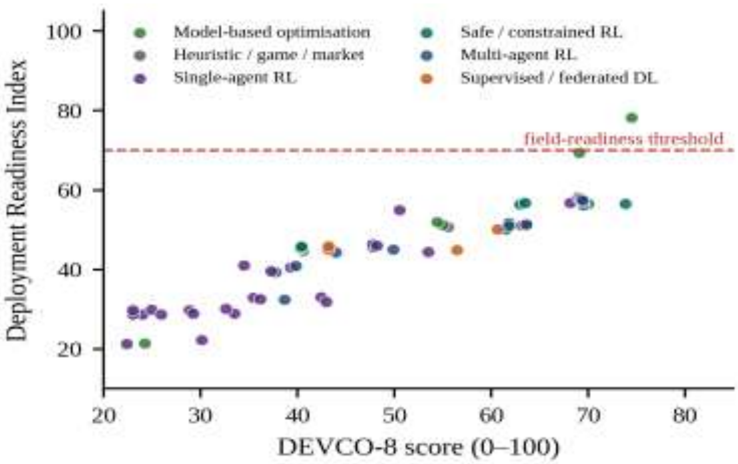

Fig 9. Deployment Readiness Index (7) against DEVCO-8 score (6) for the 60 primary studies, coloured by method family. Points are jittered slightly to separate coincident scores. The dashed line marks DRI = 70, which requires at least two of the five readiness dimensions at maximum and none at zero.

11. Reference Architecture, Algorithms and Evaluation Protocol

11.1 Three-timescale Coordination Architecture

Fig 10 assembles the strongest elements of the corpus into a single architecture. The design principle is that each timescale uses the method best suited to its information: optimisation where the model is known and the horizon is long, learning where the dynamics are unmodelled and the decision must be fast, and local control where communication cannot be assumed. A day-ahead layer runs a stochastic or robust programme [25], [43], [45] at the DSO to allocate feeder capacity envelopes, and an EV-aggregator programme that schedules fleet energy under user deadlines, the two exchanging prices and quantities by ADMM [26], [27]. An intra-day layer re-solves an MPC [44] with updated probabilistic forecasts [38], [40] and lets a shielded MARL corrector adjust setpoints within the envelopes (Algorithm 3). A real-time layer executes IEEE 1547-2018 Volt/VAr and Volt/Watt curves at DG inverters [24], [100] and current-setpoint tracking at chargers. A digital-twin feeder model [116] provides state estimation and what-if checks to every layer. Algorithm 2 states the control flow.

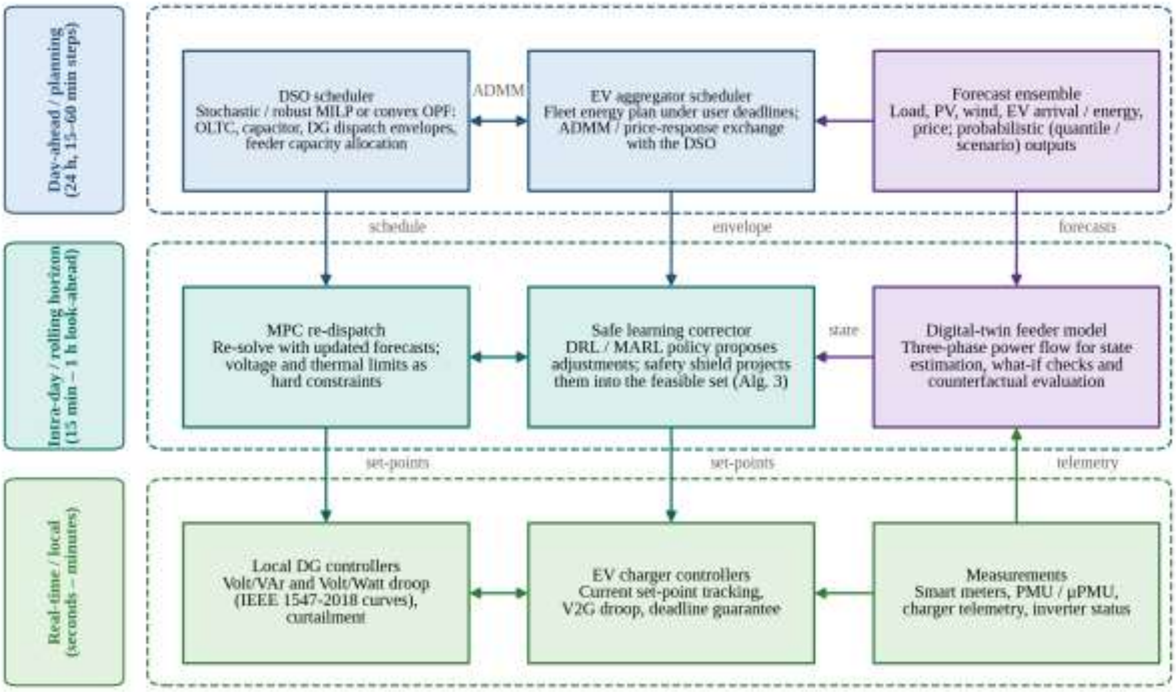

Fig 10. Three-timescale reference architecture for AI-driven coordination of DG and EVs. Model-based optimisation governs the day-ahead layer, a shielded learning corrector governs the intra-day layer, and standard-compliant local controllers govern the real-time layer; a digital-twin feeder model serves all three.

Algorithm 2: Three-Timescale Hierarchical Coordination

Input: feeder model $\mathcal{G}$, forecasts ***f*** *(probabilistic), EV fleet $\mathcal{V}$ with (arrival, departure, E^req), DG set $\mathcal{D}$, envelopes from DSO*
Output: executed setpoints for every DG unit and charger at every step
1: every 24 h (day-ahead):
2: DSO solves robust/stochastic OPF on $\mathcal{G}$ with **f** → capacity envelopes **P**_j(t) per bus j
3: Aggregator solves fleet schedule min Σ cost s.t. (2), E_{v,T_v} ≥ E^req_v, Σ_v P^c_{v,j}(t) ≤ **P**_j(t)
4: repeat ADMM price/quantity exchange between lines 2–3 until residuals < tolerance
5: every 15 min (intra-day):
6: update **f** from measurements; re-solve MPC over next H steps with envelopes as hard bounds → nominal setpoints **û**
7: for each agent i (bus zone or charger cluster): ã_i ← π_θi(s_i) ▷ learned adjustment
8: a ← Shield(**û** + ã, $\mathcal{G}$, s) ▷ Algorithm 3, lines 4–8
9: dispatch a to local controllers; log (s, ã, a, r, c, s′)
10: every 1–10 s (real-time):
11: DG inverters track IEEE 1547 Volt/VAr and Volt/Watt curves around setpoint a
12: chargers track current setpoint a; if E_v short of E^req_v within horizon, raise local override flag
13: digital twin runs three-phase power flow on telemetry; if any |V_j − V_0| > ΔV_max or |S_ij| > S_max, trigger line 6 immediately

11.2 Shielded Multi-agent Learning

Algorithm 3 specifies the intra-day corrector. Each agent holds a policy and a reward critic; a shared constraint critic estimates expected violation; a shield projects the joint proposed action onto the feasible set using the digital twin. The shield is exact when the feasibility check is a full power flow and approximate when it uses the linearisation (1); in the latter case a margin is added to ΔV_{max}. The Lagrange multiplier λ enforces (12) so that the policy learns to propose feasible actions and the shield intervenes less often as training proceeds. The formulation generalises the constrained soft actor-critic of [36] and the safety modules of [67] to the joint DG–EV case and adds the training-time logging that Section 9-C requires.

Algorithm 3: Shielded Multi-Agent Learning Corrector (SMALC)

Input: agents i = 1..M with policies π_θi and critics Q_i; shared constraint critic Q_c; digital twin 𝒢; tolerance δ; step sizes α, β; margin m
Output: trained policies; violation log Λ
1: initialise θ_i, φ_i, ψ (Q_c), λ ← 0, replay buffer ℬ, Λ ← ∅
2: for each intra-day step t do
3: for each agent i: ã_i ← π_θi(s_i) + exploration noise
4: ã ← (ã_1, ..., ã_M); (V̂, Ŝ) ← PowerFlow(𝒢, s, **û** + ã)
5: if V_min + m ≤ V̂ ≤ V_max − m and |Ŝ| ≤ S_max and deadline check holds then a ← ã
6: else a ← argmin_{a′ ∈ 𝒜(s)} ||a′ − ã||² ▷ projection onto feasible set via twin (QP on (1)–(3))
7: execute a; observe r, c = Σ_j max(0, |V_j − V_0| − ΔV_max) + Σ_ij max(0, |S_ij| − S_max), s′
8: if c > 0 then Λ ← Λ ∪ {(t, c)} ▷ exposure log for (11)
9: ℬ ← ℬ ∪ {(s, ã, a, r, c, s′)}
10: sample minibatch from ℬ
11: update Q_i by TD on r; update Q_c by TD on c
12: for each agent i: θ_i ← θ_i + α ∇_θi [Q_i(s, ã) − λ Q_c(s, ã)] ▷ gradient through proposed action
13: λ ← max(0, λ + β (mean_ℬ c − δ))
14: return {π_θi}, Λ

11.3 Test Systems and Data for a Common Benchmark

Table 11 lists the test systems, datasets and environments encountered in the corpus and identifies the gap: no environment couples a three-phase feeder with measured EV sessions and PV data under a standard task definition. The FAIR-GRID protocol below assumes such an environment; assembling one from the listed components (a pandapower or MATPOWER feeder, ACN-Data sessions, a public PV trace, and a CityLearn-style task interface) is a tractable community task and is the first item of the roadmap.

Table 11. Test systems, datasets and environments used in the corpus.

Resource	Type	Used by	Provides	Missing for DG–EV coordination
IEEE 13/33/34/123/141-bus feeders [21], [102]	Test feeder	Most DG / Volt/VAr studies	Topology, impedances, nominal load	No EV sessions; often run single-phase
IEEE 8500-node [100]	Test feeder	Few	Three-phase, realistic size	No DER or EV allocation standard
Real LV feeders [32], [33]	Utility data	Two early studies	Measured topology and load	Not public
ACN-Data [99]	EV sessions	Fleet RL studies	Measured arrivals, departures, energy	No feeder; workplace bias
Smart-meter datasets [37], [119]	Load	Forecasting studies	Household load at 30-min resolution	No coincident EV or PV
GEFCom2014 [38]	Load / PV / wind / price	Forecasting	Benchmark forecasting tracks	No network
pandapower [116]; MATPOWER [117]	Power-flow tool	Many	Three-phase / balanced power flow, OPF	No task interface for RL
CityLearn [112]	RL environment	Building EMS	Multi-building task, shared benchmark	No feeder; no EVs at network level
Grid2Op [113]	RL environment	Topology control	Standard task, competitions	Transmission-level

11.4 The FAIR-GRID Evaluation Protocol

FAIR-GRID (Feasibility, Adversity, Independence and Reproducibility for Grid learning) is a ten-stage protocol (Fig 11, Algorithm 4) whose stages map onto the DEVCO-8 dimensions so that a study following it scores at least 1 on every dimension and can score 2 on all. Six reporting invariants (Table 12) must hold in the final report. The protocol asks for nothing that a study in the corpus has not already done at least once; it asks for all of them at once.

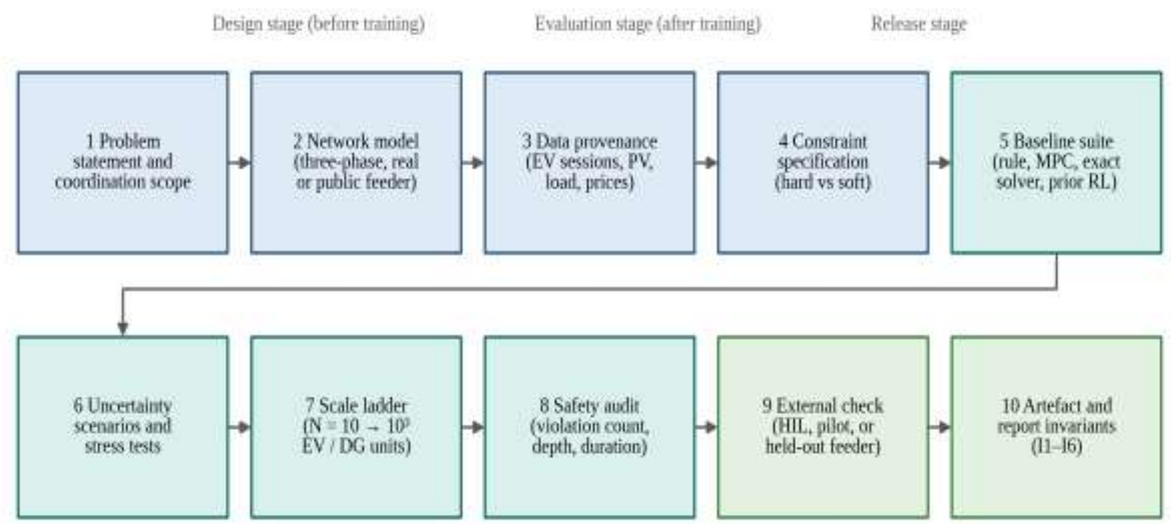

Fig 11. The ten stages of the FAIR-GRID evaluation protocol. Design-stage decisions (1–4) precede training; evaluation-stage tests (5–8) follow it; release-stage steps (9–10) close the loop with external validation and reporting invariants.

Algorithm 4: FAIR-GRID Evaluation Protocol

Input: proposed coordination method $\mathcal{M}$; feeder set $\{\mathcal{G}\_1, \ldots, \mathcal{G}\_K\}$; data $\mathcal{D}$ with provenance; baseline suite $\mathcal{B}$ = {rule, MPC, exact solver, prior learners}
Output: evaluation report satisfying invariants I1–I6

1: state problem: objective terms of (4), hard vs soft constraints of (5), coordination scope (which of DSO / aggregator / user decides what)
2: fix network models: at least one three-phase feeder; declare phase allocation of EVs and PV
3: fix data: measured sessions and PV where available; declare synthetic parts; hold out ≥ 1 feeder or ≥ 20% of days
4: fix constraints: list every limit; declare mechanism (shield / projection / penalty) and margin m
5: tune every baseline in $\mathcal{B}$ on the same information set with the same budget; record hyper-parameter search
6: for each forecast-error level e ∈ {0, 5, 10, 20, 40}% and behaviour shift σ ∈ {0, 1, 2}: run $\mathcal{M}$ and $\mathcal{B}$; record cost, loss, violations
7: for each scale N ∈ {10, 100, 1000} (EV) and feeder size: run $\mathcal{M}$; record training time, inference time, performance
8: safety audit: report violation count, maximum depth (p.u. or %), total duration, during training and at test
9: external check: hardware-in-the-loop, pilot, or held-out feeder from step 3; report degradation vs in-sample
10: release environment, seeds, configuration, trained policies; verify invariants I1–I6 (Table 12); publish

---

Table 12. FAIR-GRID reporting invariants.

ID	Invariant	Rationale
I1	Every reported improvement is accompanied by the violation triple (count, maximum depth, duration) over the same episodes	Improvement obtained by violating limits is not improvement
I2	Every learned policy is compared to a tuned optimiser on the same information set	Separates the value of learning from the value of information
I3	Performance is reported as a function of forecast error, not at a single forecast	Operators buy forecasts; they need the price of error
I4	At least two scales are reported with training and inference cost	Scalability is a slope, not a point
I5	At least one result is on data or a feeder not used in training	In-sample results cannot transfer
I6	Environment, seeds and configuration are released	Without them the study is testimony, not evidence

11.5 Computing the Deployment Readiness Index

Algorithm 5 states the DRI computation so that it can be applied to any study, including one following FAIR-GRID, and clarifies the reading of each dimension.

Algorithm 5: Deployment Readiness Index

Input: DEVCO-8 digits d_1..d_8 ∈ {0, 1, 2} for a study
Output: score S ∈ [0, 100], DRI ∈ [11.1, 100], readiness class
1: S ← 100 · Σ_{i=1}^{8} w_i · d_i / 2 with w = (0.15, 0.12, 0.18, 0.12, 0.13, 0.10, 0.12, 0.08)
2: g ← 1
3: for i in {1, 3, 5, 6, 7}: g ← g · (d_i + 0.25) / 2.25 ▷ network, safety, scalability, uncertainty, external validity
4: DRI ← 100 · g^{1/5}
5: if DRI ≥ 70 then class ← 'field-ready candidate'
6: else if DRI ≥ 50 then class ← 'pilot-ready with stated gaps'
7: else class ← 'research demonstration'
8: return S, DRI, class

12. Research Agenda and Roadmap

Table 13 lists the open problems that the appraisal identified, ordered by the size of the evidence gap rather than by novelty, and Fig 12 places them on three horizons.

Table 13. Research agenda derived from the appraisal.

Open problem	Evidence gap	Promising direction	Enablers in corpus
Joint DG–EV coordination with a network	Only 9 of 30 EV studies model DG or a feeder	Bi-level: optimiser allocates capacity, learner allocates within it	[66], [67], [106]
Safety during training on live systems	One study reports training-time violations	Shielded learning with exposure logging (Algorithm 3)	[36], [87], [118]
Scalability to 10^3 vehicles	26.7% report ≥ 100 units	Graph-structured MARL decomposed along physical coupling	[58], [69], [89]
Forecast-error propagation	No study reports degradation curve	Probabilistic forecasts driving stochastic MPC with error sweep	[38], [40], [44]
Common benchmark environment	None couples feeder, sessions and PV	Assemble from pandapower + ACN-Data + PV traces with CityLearn-style tasks	[99], [112], [116]
Privacy-preserving coordination	Federated methods used for forecasting only	Federated policy learning across aggregators; differential privacy on session data	[40], [77]
Cyber-physical security of learned controllers	Not addressed by any coordination study	Robust training against false-data injection on inputs	[120], [121]
Interpretability for operators	No study explains a learned schedule	Interpretable-by-design or attribution on inputs; envelope-based explanations	[122], [123]
Standard-aligned control envelopes	IEEE 1547 curves modelled in few studies	Learn within 1547-compliant envelopes; certify by twin	[24], [100]
Asset-ageing and user-comfort trade-off	One study each	Multi-objective evaluation with Pareto reporting	[48], [62], [124]

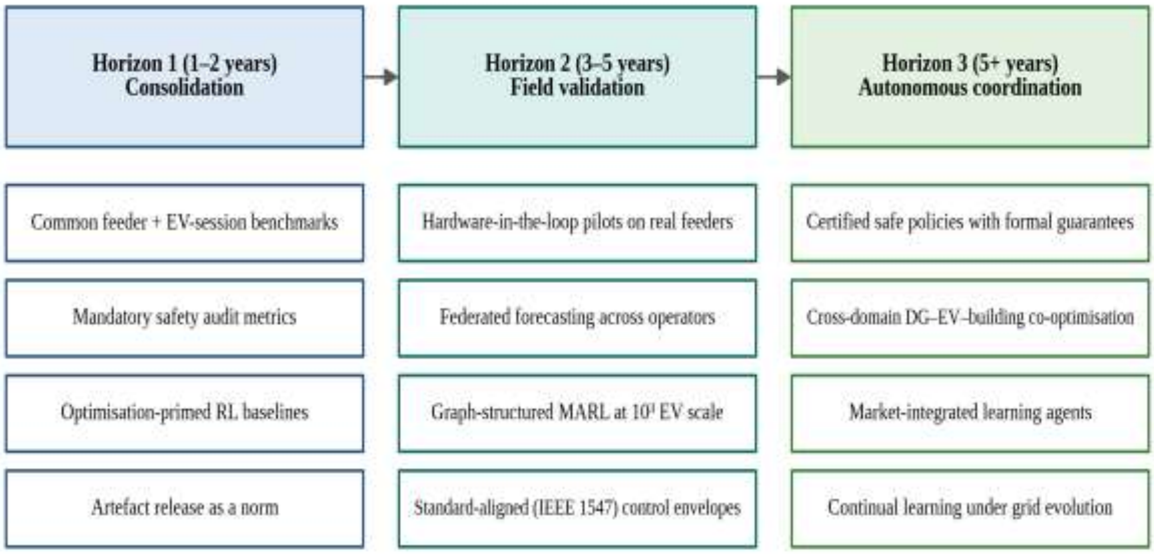

Fig 12. Three-horizon research roadmap. Horizon 1 consolidates evaluation practice; Horizon 2 validates in the field; Horizon 3 targets certified autonomous coordination.

Two items deserve emphasis. First, the benchmark environment is the precondition for everything else: without it, invariants I2, I4 and I5 cannot be met at reasonable cost, and the field will continue to compare methods on incommensurable set-ups. Second, the security of learned controllers is entirely unaddressed by the coordination literature, although false-data injection against state estimation is a mature threat [120], [121] and a learned policy that reads smart-meter and charger telemetry [125] is a new attack surface; a policy that is safe by shield remains safe only if the shield's inputs are trustworthy.

On method choice the review takes a deliberately neutral position. No-free-lunch reasoning [126] applies: single-agent DRL, MARL, safe RL, metaheuristics such as particle swarm [127] or evolutionary multi-objective search [124], and convex optimisation each dominate on some instance of (4)–(5). What the evidence supports is not a winner but a division of labour, which is what the reference architecture encodes. The recent overview of AI methods for renewable power-system operation by Li et al. [128] reaches a compatible conclusion from the transmission side, and the reviews of infrastructure challenges [129], charging service operations [130], and EV state-of-the-art [131]–[133] describe the practical constraints within which any division of labour must operate. Foundations such as attention mechanisms [134] and the DG placement literature [135]–[137] will continue to supply components; the review's claim is that components are no longer the bottleneck.

13. Limitations of This Review

The corpus was assembled by seeded snowballing with database verification rather than by exhaustive database search, so coverage of very recent or peripheral venues is incomplete; the counts in Fig 2 describe this procedure and would change under a different one. Appraisal was performed by a single reviewer against a rubric applied to the retrievable record; individual digits in Tables 7–9 should be read as one informed reading, and authors of appraised studies who hold that a digit understates their work are invited to supply the evidence that FAIR-GRID would require. The analytical arguments of Section 9 use an aggregated single-phase feeder and illustrative parameters; they are intended to be recomputed with feeder-specific values, not read as universal constants. The DEVCO-8 weights are a judgement, and Fig 4 reports unweighted attainment for readers who prefer their own.

14. Conclusion

This review appraised 60 primary studies of AI-driven energy management and coordination of distributed generation and electric vehicles in smart distribution systems against an eight-dimension rubric and a deployment-readiness index that resists averaging. The literature has demonstrated capability: learned policies charge vehicles in response to price, regulate voltage with inverters, and, in a few recent designs, allocate feeder capacity to hundreds of vehicles within an optimiser's envelope. It has not demonstrated readiness. Most EV coordination studies omit the network they are meant to protect; most learning studies handle constraints by penalty and report no violations; almost none test against forecast error, vary the scale, or validate outside simulation; and few release artefacts. The three analytical arguments make the cost of these omissions concrete: coordination buys hosting capacity kilowatt for kilowatt, centralised learning cannot scale past tens of vehicles, and unshielded exploration on a live feeder accrues violations in proportion to the exploration rate.

14.1 Answers to the Review Questions

Table 14 closes the loop opened in Section 1.2 by answering each review question from the evidence assembled in the review.

Table 14. Answers to the review questions derived from the synthesis.

ID	Answer from the evidence	Basis
RQ1	Six families are present. Model-based optimisation dominates 2010–2015; reinforcement learning (single-agent, multi-agent and safe/constrained) accounts for 71.7% of the 60 primary studies and almost all of those dated 2020 or later (43.1% of the corpus). Supervised and federated learning appear only as forecasting front-ends.	Sections 4–5
RQ2	Rigour is low and uneven. 28.3% of primary studies contain no network model at all; constraints are enforced by penalty rather than projection in most learning studies; the mean DEVCO-8 score is 47.4 (median 45.0, maximum 74.5).	Sections 6–8; Tables 7–9
RQ3	Weak. Only 26.7% of studies test more than one scale, 11.7% report any external validity, and 11.7% release an artefact. Only the two real-feeder optimisations of [32] and [33] reach DRI ≥ 70; no learning study does.	Section 9; Fig 9
RQ4	Eight failure modes recur (F1–F8). The most prevalent are irreproducibility (F8, 88.3% release no artefact), simulation-only validity (F7, 73.3%) and scale by assertion (F4, only 26.7% test more than one scale); no study reaches the top level on baseline rigour (F5) or forecast-error propagation (F6); the absent network (F1) and penalty-only constraints (F2) affect 28.3% and 30% respectively. Each has a remedy already demonstrated somewhere in the corpus.	Section 10; Table 10
RQ5	Nine resources are in use (IEEE 13/34/123-bus, ACN-Data, GEFCom, CityLearn, Grid2Op and others), but none couples a three-phase feeder with measured EV sessions and PV under a standard task definition; that coupled benchmark is the first item of the roadmap.	Section 11.3; Table 11
RQ6	Yes. Coordination raises PV hosting capacity kilowatt for kilowatt of shifted EV charging (from 1.42 to 2.32 p.u. on the illustrative feeder); the joint action space grows as k^N against Nk for decomposed control; unshielded exploration accrues violations in proportion to εTp_v.	Section 9; equations (8)–(12)
RQ7	A three-timescale architecture (day-ahead optimisation, intra-day shielded learning, standard-compliant local control, all served by a digital twin), Algorithms 2–5, and the FAIR-GRID protocol with its six reporting invariants.	Sections 11–12; Figs 10–12

The constructive response is not a new architecture but evaluation discipline embodied in an architecture that already exists in pieces across the corpus: optimisation where the model is known, shielded learning where it is not, standard-compliant local control where communication cannot be assumed, and a digital twin serving all three. The FAIR-GRID protocol and its six invariants specify what a study must report before an operator can act on it. If the field adopts them, the next generation of results will be smaller in headline improvement and far larger in consequence.

References

[1] H. Farhangi, "The path of the smart grid," IEEE Power Energy Mag., vol. 8, no. 1, pp. 18–28, Jan./Feb. 2010, doi: 10.1109/MPE.2009.934876.

[2] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—The new and improved power grid: A survey," IEEE Commun. Surveys Tuts., vol. 14, no. 4, pp. 944–980, 2012, doi: 10.1109/SURV.2011.101911.00087.

[3] T. Ackermann, G. Andersson, and L. Söder, "Distributed generation: A definition," Electr. Power Syst. Res., vol. 57, no. 3, pp. 195–204, 2001, doi: 10.1016/S0378-7796(01)00101-8.

[4] P. S. Georgilakis and N. D. Hatziargyriou, "Optimal distributed generation placement in power distribution networks: Models, methods, and future research," IEEE Trans. Power Syst., vol. 28, no. 3, pp. 3420–3428, Aug. 2013, doi: 10.1109/TPWRS.2012.2237043.

[5] W. Kempton and J. Tomić, "Vehicle-to-grid power fundamentals: Calculating capacity and net revenue," J. Power Sources, vol. 144, no. 1, pp. 268–279, 2005, doi: 10.1016/j.jpowsour.2004.12.022.
[6] J. A. P. Lopes, F. J. Soares, and P. M. R. Almeida, "Integration of electric vehicles in the electric power system," Proc. IEEE, vol. 99, no. 1, pp. 168–183, Jan. 2011, doi: 10.1109/JPROC.2010.2066250.
[7] K. Clement-Nyns, E. Haesen, and J. Driesen, "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid," IEEE Trans. Power Syst., vol. 25, no. 1, pp. 371–380, Feb. 2010, doi: 10.1109/TPWRS.2009.2036481.
[8] D. B. Richardson, "Electric vehicles and the electric grid: A review of modeling approaches, impacts, and renewable energy integration," Renew. Sustain. Energy Rev., vol. 19, pp. 247–254, 2013, doi: 10.1016/j.rser.2012.11.042.
[9] F. Mwasilu, J. J. Justo, E.-K. Kim, T. D. Do, and J.-W. Jung, "Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration," Renew. Sustain. Energy Rev., vol. 34, pp. 501–516, 2014, doi: 10.1016/j.rser.2014.03.031.
[10] J. Hu, H. Morais, T. Sousa, and M. Lind, "Electric vehicle fleet management in smart grids: A review of services, optimization and control aspects," Renew. Sustain. Energy Rev., vol. 56, pp. 1207–1226, 2016, doi: 10.1016/j.rser.2015.12.014.
[11] O. Sadeghian, A. Oshnoei, B. Mohammadi-Ivatloo, V. Vahidinasab, and A. Anvari-Moghaddam, "A comprehensive review on electric vehicles smart charging: Solutions, strategies, technologies, and challenges," J. Energy Storage, vol. 54, Art. no. 105241, 2022, doi: 10.1016/j.est.2022.105241.
[12] D. Zhang, X. Han, and C. Deng, "Review on the research and practice of deep learning and reinforcement learning in smart grids," CSEE J. Power Energy Syst., vol. 4, no. 3, pp. 362–370, Sep. 2018, doi: 10.17775/CSEEJPES.2018.00520.
[13] Z. Zhang, D. Zhang, and R. C. Qiu, "Deep reinforcement learning for power system applications: An overview," CSEE J. Power Energy Syst., vol. 6, no. 1, pp. 213–225, Mar. 2020, doi: 10.17775/CSEEJPES.2019.00920.
[14] X. Chen, G. Qu, Y. Tang, S. Low, and N. Li, "Reinforcement learning for selective key applications in power systems: Recent advances and future challenges," IEEE Trans. Smart Grid, vol. 13, no. 4, pp. 2935–2958, Jul. 2022, doi: 10.1109/TSG.2022.3154718.
[15] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," Appl. Energy, vol. 235, pp. 1072–1089, 2019, doi: 10.1016/j.apenergy.2018.11.002.
[16] T. Yang, L. Zhao, W. Li, and A. Y. Zomaya, "Reinforcement learning in sustainable energy and electric systems: A survey," Annu. Rev. Control, vol. 49, pp. 145–163, 2020, doi: 10.1016/j.arcontrol.2020.03.001.
[17] D. Cao et al., "Reinforcement learning and its applications in modern power and energy systems: A review," J. Mod. Power Syst. Clean Energy, vol. 8, no. 6, pp. 1029–1042, Nov. 2020, doi: 10.35833/MPCE.2020.000552.
[18] A. T. D. Perera and P. Kamalaruban, "Applications of reinforcement learning in energy systems," Renew. Sustain. Energy Rev., vol. 137, Art. no. 110618, 2021, doi: 10.1016/j.rser.2020.110618.
[19] I. Antonopoulos et al., "Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review," Renew. Sustain. Energy Rev., vol. 130, Art. no. 109899, 2020, doi: 10.1016/j.rser.2020.109899.
[20] P. Yu, H. Zhang, Y. Song, Z. Wang, H. Dong, and L. Ji, "Safe reinforcement learning for power system control: A review," Renew. Sustain. Energy Rev., vol. 223, Art. no. 116022, 2025, doi: 10.1016/j.rser.2025.116022.
[21] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," IEEE Trans. Power Del., vol. 4, no. 2, pp. 1401–1407, Apr. 1989, doi: 10.1109/61.25627.
[22] M. Farivar and S. H. Low, "Branch flow model: Relaxations and convexification—Part I," IEEE Trans. Power Syst., vol. 28, no. 3, pp. 2554–2564, Aug. 2013, doi: 10.1109/TPWRS.2013.2255317.
[23] S. H. Low, "Convex relaxation of optimal power flow—Part I: Formulations and equivalence," IEEE Trans. Control Netw. Syst., vol. 1, no. 1, pp. 15–27, Mar. 2014, doi: 10.1109/TCNS.2014.2309732.
[24] IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces, IEEE Std 1547-2018, 2018, doi: 10.1109/IEEESTD.2018.8332112.
[25] D. Bertsimas and M. Sim, "The price of robustness," Oper. Res., vol. 52, no. 1, pp. 35–53, 2004, doi: 10.1287/opre.1030.0065.

[26] D. K. Molzahn et al., "A survey of distributed optimization and control algorithms for electric power systems," IEEE Trans. Smart Grid, vol. 8, no. 6, pp. 2941–2962, Nov. 2017, doi: 10.1109/TSG.2017.2720471.

[27] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1–122, 2011, doi: 10.1561/2200000016.

[28] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, Art. no. n71, 2021, doi: 10.1136/bmj.n71.

[29] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele Univ. and Durham Univ., U.K., Tech. Rep. EBSE-2007-01, 2007.

[30] E. Sortomme and M. A. El-Sharkawi, "Optimal charging strategies for unidirectional vehicle-to-grid," IEEE Trans. Smart Grid, vol. 2, no. 1, pp. 131–138, Mar. 2011, doi: 10.1109/TSG.2010.2090910.

[31] S. Deilami, A. S. Masoum, P. S. Moses, and M. A. S. Masoum, "Real-time coordination of plug-in electric vehicle charging in smart grids to minimize power losses and improve voltage profile," IEEE Trans. Smart Grid, vol. 2, no. 3, pp. 456–467, Sep. 2011, doi: 10.1109/TSG.2011.2159816.

[32] O. Sundström and C. Binding, "Flexible charging optimization for electric vehicles considering distribution grid constraints," IEEE Trans. Smart Grid, vol. 3, no. 1, pp. 26–37, Mar. 2012, doi: 10.1109/TSG.2011.2168431.

[33] P. Richardson, D. Flynn, and A. Keane, "Optimal charging of electric vehicles in low-voltage distribution systems," IEEE Trans. Power Syst., vol. 27, no. 1, pp. 268–279, Feb. 2012, doi: 10.1109/TPWRS.2011.2158247.

[34] L. Gan, U. Topcu, and S. H. Low, "Optimal decentralized protocol for electric vehicle charging," IEEE Trans. Power Syst., vol. 28, no. 2, pp. 940–951, May 2013, doi: 10.1109/TPWRS.2012.2210288.

[35] Z. Ma, D. S. Callaway, and I. A. Hiskens, "Decentralized charging control of large populations of plug-in electric vehicles," IEEE Trans. Control Syst. Technol., vol. 21, no. 1, pp. 67–78, Jan. 2013, doi: 10.1109/TCST.2011.2174059.

[36] W. Wang, N. Yu, Y. Gao, and J. Shi, "Safe off-policy deep reinforcement learning algorithm for Volt-VAR control in power distribution systems," IEEE Trans. Smart Grid, vol. 11, no. 4, pp. 3008–3018, Jul. 2020, doi: 10.1109/TSG.2019.2962625.

[37] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," IEEE Trans. Smart Grid, vol. 10, no. 1, pp. 841–851, Jan. 2019, doi: 10.1109/TSG.2017.2753802.

[38] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," Int. J. Forecast., vol. 32, no. 3, pp. 896–913, 2016, doi: 10.1016/j.ijforecast.2016.02.001.

[39] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," Energy Convers. Manage., vol. 156, pp. 459–497, 2018, doi: 10.1016/j.enconman.2017.11.019.

[40] T. Wang, C. Ren, Z. Y. Dong, and C. Yip, "Domain-adaptive clustered federated transfer learning for EV charging demand forecasting," IEEE Trans. Power Syst., vol. 40, no. 2, pp. 1241–1254, Mar. 2025.

[41] F. Marzbani, A. H. Osman, and M. S. Hassan, "Electric vehicle energy demand prediction techniques: An in-depth and critical systematic review," IEEE Access, vol. 11, pp. 96242–96255, 2023.

[42] E. Dall'Anese, S. V. Dhople, and G. B. Giannakis, "Optimal dispatch of photovoltaic inverters in residential distribution systems," IEEE Trans. Sustain. Energy, vol. 5, no. 2, pp. 487–497, Apr. 2014, doi: 10.1109/TSTE.2013.2292828.

[43] Y. Xu, Z. Y. Dong, R. Zhang, and D. J. Hill, "Multi-timescale coordinated voltage/var control of high renewable-penetrated distribution systems," IEEE Trans. Power Syst., vol. 32, no. 6, pp. 4398–4408, Nov. 2017, doi: 10.1109/TPWRS.2017.2669343.

[44] W. Tang and Y. J. Zhang, "A model predictive control approach for low-complexity electric vehicle charging scheduling: Optimality and scalability," IEEE Trans. Power Syst., vol. 32, no. 2, pp. 1050–1063, Mar. 2017, doi: 10.1109/TPWRS.2016.2585202.

[45] T. Qian, C. Shao, X. Li, X. Wang, and M. Shahidehpour, "Enhanced coordinated operations of electric power and transportation networks via EV charging services," IEEE Trans. Smart Grid, vol. 11, no. 4, pp. 3019–3030, Jul. 2020, doi: 10.1109/TSG.2020.2969650.

[46] Z. Wan, H. Li, H. He, and D. Prokhorov, "Model-free real-time EV charging scheduling based on deep reinforcement learning," IEEE Trans. Smart Grid, vol. 10, no. 5, pp. 5246–5257, Sep. 2019, doi: 10.1109/TSG.2018.2879572.

[47] F. Zhang, Q. Yang, and D. An, "CDDPG: A deep-reinforcement-learning-based approach for electric vehicle charging control," IEEE Internet Things J., vol. 8, no. 5, pp. 3075–3087, Mar. 2021, doi: 10.1109/JIOT.2020.3015204.

[48] L. Yan, X. Chen, J. Zhou, Y. Chen, and J. Wen, "Deep reinforcement learning for continuous electric vehicles charging control with dynamic user behaviors," IEEE Trans. Smart Grid, vol. 12, no. 6, pp. 5124–5134, Nov. 2021, doi: 10.1109/TSG.2021.3098298.

[49] M. Dorokhova, Y. Martinson, C. Ballif, and N. Wyrsch, "Deep reinforcement learning control of electric vehicle charging in the presence of photovoltaic generation," Appl. Energy, vol. 301, Art. no. 117504, 2021, doi: 10.1016/j.apenergy.2021.117504.

[50] E. Mocanu et al., "On-line building energy optimization using deep reinforcement learning," IEEE Trans. Smart Grid, vol. 10, no. 4, pp. 3698–3708, Jul. 2019, doi: 10.1109/TSG.2018.2834219.

[51] Y. Ye, D. Qiu, X. Wu, G. Strbac, and J. Ward, "Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning," IEEE Trans. Smart Grid, vol. 11, no. 4, pp. 3068–3082, Jul. 2020, doi: 10.1109/TSG.2020.2976771.

[52] Y. Ji, J. Wang, J. Xu, X. Fang, and H. Zhang, "Real-time energy management of a microgrid using deep reinforcement learning," Energies, vol. 12, no. 12, Art. no. 2291, 2019, doi: 10.3390/en12122291.

[53] T. A. Nakabi and P. Toivanen, "Deep reinforcement learning for energy management in a microgrid with flexible demand," Sustain. Energy Grids Netw., vol. 25, Art. no. 100413, 2021, doi: 10.1016/j.segan.2020.100413.

[54] D. Cao, W. Hu, J. Zhao, Q. Huang, Z. Chen, and F. Blaabjerg, "A multi-agent deep reinforcement learning based voltage regulation using coordinated PV inverters," IEEE Trans. Power Syst., vol. 35, no. 5, pp. 4120–4123, Sep. 2020, doi: 10.1109/TPWRS.2020.3000652.

[55] D. Cao et al., "Data-driven multi-agent deep reinforcement learning for distribution system decentralized voltage control with high penetration of PVs," IEEE Trans. Smart Grid, vol. 12, no. 5, pp. 4137–4150, Sep. 2021, doi: 10.1109/TSG.2021.3072251.

[56] H. Liu and W. Wu, "Online multi-agent reinforcement learning for decentralized inverter-based Volt-VAR control," IEEE Trans. Smart Grid, vol. 12, no. 4, pp. 2980–2990, Jul. 2021, doi: 10.1109/TSG.2021.3060027.

[57] X. Sun and J. Qiu, "Two-stage Volt/Var control in active distribution networks with multi-agent deep reinforcement learning method," IEEE Trans. Smart Grid, vol. 12, no. 4, pp. 2903–2912, Jul. 2021, doi: 10.1109/TSG.2021.3052998.

[58] X. Chen, G. Qu, Y. Tang, S. Low, and N. Li, "Consensus multi-agent reinforcement learning for Volt-VAR control in power distribution networks," IEEE Trans. Smart Grid, vol. 12, no. 4, pp. 3594–3604, Jul. 2021, doi: 10.1109/TSG.2021.3058996.

[59] D. Hu, Z. Ye, Y. Gao, Z. Ye, Y. Peng, and N. Yu, "Multi-agent deep reinforcement learning for voltage control with coordinated active and reactive power optimization," IEEE Trans. Smart Grid, vol. 13, no. 6, pp. 4873–4886, Nov. 2022, doi: 10.1109/TSG.2022.3185975.

[60] F. L. Da Silva, C. E. H. Nishida, D. M. Roijers, and A. H. R. Costa, "Coordination of electric vehicle charging through multiagent reinforcement learning," IEEE Trans. Smart Grid, vol. 11, no. 3, pp. 2347–2356, May 2020, doi: 10.1109/TSG.2019.2952331.

[61] L. Yan, X. Chen, Y. Chen, and J. Wen, "A cooperative charging control strategy for electric vehicles based on multiagent deep reinforcement learning," IEEE Trans. Ind. Informat., vol. 18, no. 12, pp. 8765–8775, Dec. 2022, doi: 10.1109/TII.2022.3152218.

[62] S. Li, H. He, J. Li, Y. Wang, and S. Su, "EV charging strategy considering transformer lifetime via evolutionary curriculum learning-based multiagent deep reinforcement learning," IEEE Trans. Smart Grid, vol. 13, no. 4, pp. 2774–2787, Jul. 2022, doi: 10.1109/TSG.2022.3167021.

[63] T. Qian, C. Shao, X. Li, X. Wang, Z. Chen, and M. Shahidehpour, "Multi-agent deep reinforcement learning method for EV charging station game," IEEE Trans. Power Syst., vol. 37, no. 3, pp. 1682–1694, May 2022, doi: 10.1109/TPWRS.2021.3111014.

[64] H. Li, Z. Wan, and H. He, "Constrained EV charging scheduling based on safe deep reinforcement learning," IEEE Trans. Smart Grid, vol. 11, no. 3, pp. 2427–2439, May 2020, doi: 10.1109/TSG.2019.2955437.

[65] Y. Zhang, X. Wang, J. Wang, and Y. Zhang, "Deep reinforcement learning based Volt-VAR optimization in smart distribution systems," IEEE Trans. Smart Grid, vol. 12, no. 1, pp. 361–371, Jan. 2021, doi: 10.1109/TSG.2020.3010130.

[66] F. Ming, F. Gao, K. Liu, and X. Li, "A constrained DRL-based bi-level coordinated method for large-scale EVs charging," Appl. Energy, vol. 331, Art. no. 120381, 2023, doi: 10.1016/j.apenergy.2022.120381.

[67] J. Zhang, Y. Guan, L. Che, and M. Shahidehpour, "EV charging command fast allocation approach based on deep reinforcement learning with safety modules," IEEE Trans. Smart Grid, vol. 15, no. 1, pp. 757–769, Jan. 2024, doi: 10.1109/TSG.2023.3281782.

[68] M. Zhang, G. Guo, S. Magnússon, R. C. N. Pilawa-Podgurski, and Q. Xu, "Data driven decentralized control of inverter based renewable energy sources using safe guaranteed multi-agent deep reinforcement learning," IEEE Trans. Sustain. Energy, vol. 15, no. 2, pp. 1288–1299, Apr. 2024, doi: 10.1109/TSTE.2023.3341632.

[69] R. Yan, Q. Xing, and Y. Xu, "Multi agent safe graph reinforcement learning for PV inverters based real-time decentralized Volt/Var control in zoned distribution networks," IEEE Trans. Smart Grid, vol. 15, no. 1, pp. 299–311, Jan. 2024, doi: 10.1109/TSG.2023.3277087.

[70] R. Dobbe, O. Sondermeijer, D. Fridovich-Keil, D. Arnold, D. Callaway, and C. Tomlin, "Toward distributed energy services: Decentralizing optimal power flow with machine learning," IEEE Trans. Smart Grid, vol. 11, no. 2, pp. 1296–1306, Mar. 2020, doi: 10.1109/TSG.2019.2935711.

[71] X. Pan, T. Zhao, M. Chen, and S. Zhang, "DeepOPF: A deep neural network approach for security-constrained DC optimal power flow," IEEE Trans. Power Syst., vol. 36, no. 3, pp. 1725–1735, May 2021, doi: 10.1109/TPWRS.2020.3026379.

[72] R. Wang, X. Bi, and S. Bu, "Real-time coordination of dynamic network reconfiguration and Volt-VAR control in active distribution network: A graph-aware deep reinforcement learning approach," IEEE Trans. Smart Grid, vol. 15, no. 3, pp. 3288–3302, May 2024, doi: 10.1109/TSG.2023.3324474.

[73] Q. Yang, G. Wang, A. Sadeghi, G. B. Giannakis, and J. Sun, "Two-timescale voltage control in distribution grids using deep reinforcement learning," IEEE Trans. Smart Grid, vol. 11, no. 3, pp. 2313–2323, May 2020, doi: 10.1109/TSG.2019.2951769.

[74] H. Liu and W. Wu, "Two-stage deep reinforcement learning for inverter-based Volt-VAR control in active distribution networks," IEEE Trans. Smart Grid, vol. 12, no. 3, pp. 2037–2047, May 2021, doi: 10.1109/TSG.2020.3041620.

[75] H. Xu, A. Zhang, Q. Wang, Y. Hu, F. Fang, and L. Cheng, "Quantum reinforcement learning for real-time optimization in electric vehicle charging systems," Appl. Energy, vol. 383, Art. no. 125279, 2025, doi: 10.1016/j.apenergy.2025.125279.

[76] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[77] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS), 2017, pp. 1273–1282.

[78] D. Ernst, M. Glavic, and L. Wehenkel, "Power systems stability control: Reinforcement learning framework," IEEE Trans. Power Syst., vol. 19, no. 1, pp. 427–435, Feb. 2004, doi: 10.1109/TPWRS.2003.821457.

[79] M. Glavic, R. Fonteneau, and D. Ernst, "Reinforcement learning for electric power system decision and control: Past considerations and perspectives," IFAC-PapersOnLine, vol. 50, no. 1, pp. 6918–6927, 2017, doi: 10.1016/j.ifacol.2017.08.1217.

[80] V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, pp. 529–533, 2015, doi: 10.1038/nature14236.

[81] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," in Proc. Int. Conf. Learn. Represent. (ICLR), 2016.

[82] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in Proc. 35th Int. Conf. Mach. Learn. (ICML), 2018, pp. 1861–1870.

[83] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, arXiv:1707.06347.

[84] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[85] A. Chiş, J. Lundén, and V. Koivunen, "Reinforcement learning-based plug-in electric vehicle charging with forecasted price," IEEE Trans. Veh. Technol., vol. 66, no. 5, pp. 3674–3684, May 2017, doi: 10.1109/TVT.2016.2603536.

[86] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 6379–6390.

[87] J. Achiam, D. Held, A. Tamar, and P. Abbeel, "Constrained policy optimization," in Proc. 34th Int. Conf. Mach. Learn. (ICML), 2017, pp. 22–31.

[88] J. García and F. Fernández, "A comprehensive survey on safe reinforcement learning," J. Mach. Learn. Res., vol. 16, pp. 1437–1480, 2015.

[89] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. Int. Conf. Learn. Represent. (ICLR), 2017.

[90] S. Vandael, B. Claessens, D. Ernst, T. Holvoet, and G. Deconinck, "Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market," IEEE Trans. Smart Grid, vol. 6, no. 4, pp. 1795–1805, Jul. 2015, doi: 10.1109/TSG.2015.2393059.

[91] J. Kang et al., "Enabling localized peer-to-peer electricity trading among plug-in hybrid electric vehicles using consortium blockchains," IEEE Trans. Ind. Informat., vol. 13, no. 6, pp. 3154–3164, Dec. 2017, doi: 10.1109/TII.2017.2709784.

[92] T. Ding, Z. Zeng, J. Bai, B. Qin, Y. Yang, and M. Shahidehpour, "Optimal electric vehicle charging strategy with Markov decision process and reinforcement learning technique," IEEE Trans. Ind. Appl., vol. 56, no. 5, pp. 5811–5823, Sep./Oct. 2020, doi: 10.1109/TIA.2020.2990096.

[93] T. Qian, C. Shao, X. Wang, and M. Shahidehpour, "Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system," IEEE Trans. Smart Grid, vol. 11, no. 2, pp. 1714–1723, Mar. 2020, doi: 10.1109/TSG.2019.2942593.

[94] N. Sadeghianpourhamami, J. Deleu, and C. Develder, "Definition and evaluation of model-free coordination of electrical vehicle charging with reinforcement learning," IEEE Trans. Smart Grid, vol. 11, no. 1, pp. 203–214, Jan. 2020, doi: 10.1109/TSG.2019.2920320.

[95] F. Tuchnitz, N. Ebell, J. Schlund, and M. Pruckner, "Development and evaluation of a smart charging strategy for an electric vehicle fleet based on reinforcement learning," Appl. Energy, vol. 285, Art. no. 116382, 2021, doi: 10.1016/j.apenergy.2020.116382.

[96] M. Yavuz and Ö. Kıvanç, "Optimization of a cluster-based energy management system using deep reinforcement learning without affecting prosumer comfort: V2X technologies and peer-to-peer energy trading," IEEE Access, vol. 12, pp. 31551–31576, 2024, doi: 10.1109/ACCESS.2024.3370922.

[97] R. Deng, Z. Yang, M.-Y. Chow, and J. Chen, "A survey on demand response in smart grids: Mathematical models and approaches," IEEE Trans. Ind. Informat., vol. 11, no. 3, pp. 570–582, Jun. 2015, doi: 10.1109/TII.2015.2414719.

[98] P. Siano, "Demand response and smart grids—A survey," Renew. Sustain. Energy Rev., vol. 30, pp. 461–478, 2014, doi: 10.1016/j.rser.2013.11.032.

[99] Z. J. Lee, T. Li, and S. H. Low, "ACN-Data: Analysis and applications of an open EV charging dataset," in Proc. 10th ACM Int. Conf. Future Energy Syst. (e-Energy), 2019, pp. 139–149, doi: 10.1145/3307772.3328313.

[100] P. Jahangiri and D. C. Aliprantis, "Distributed Volt/VAr control by PV inverters," IEEE Trans. Power Syst., vol. 28, no. 3, pp. 3429–3439, Aug. 2013, doi: 10.1109/TPWRS.2013.2256375.

[101] J. Duan et al., "Deep-reinforcement-learning-based autonomous voltage control for power grid operations," IEEE Trans. Power Syst., vol. 35, no. 1, pp. 814–817, Jan. 2020, doi: 10.1109/TPWRS.2019.2941134.

[102] W. H. Kersting, "Radial distribution test feeders," in Proc. IEEE Power Eng. Soc. Winter Meeting, 2001, pp. 908–912, doi: 10.1109/PESW.2001.916993.

[103] F. Ruelens, B. J. Claessens, S. Vandael, B. De Schutter, R. Babuška, and R. Belmans, "Residential demand response of thermostatically controlled loads using batch reinforcement learning," IEEE Trans. Smart Grid, vol. 8, no. 5, pp. 2149–2159, Sep. 2017, doi: 10.1109/TSG.2016.2517211.

[104] Y. Ye, D. Qiu, M. Sun, D. Papadaskalopoulos, and G. Strbac, "Deep reinforcement learning for strategic bidding in electricity markets," IEEE Trans. Smart Grid, vol. 11, no. 2, pp. 1343–1355, Mar. 2020, doi: 10.1109/TSG.2019.2936142.

[105] L. Yu et al., "Deep reinforcement learning for smart home energy management," IEEE Internet Things J., vol. 7, no. 4, pp. 2751–2762, Apr. 2020, doi: 10.1109/JIOT.2019.2957289.

[106] S. Bahrami, Y. C. Chen, and V. W. S. Wong, "Deep reinforcement learning for demand response in distribution networks," IEEE Trans. Smart Grid, vol. 12, no. 2, pp. 1496–1506, Mar. 2021, doi: 10.1109/TSG.2020.3037066.

[107] G. Pinto, M. S. Piscitelli, J. R. Vázquez-Canteli, Z. Nagy, and A. Capozzoli, "Coordinated energy management for a cluster of buildings through deep reinforcement learning," Energy, vol. 229, Art. no. 120725, 2021, doi: 10.1016/j.energy.2021.120725.

[108] D. Qiu, Y. Ye, D. Papadaskalopoulos, and G. Strbac, "Scalable coordinated management of peer-to-peer energy trading: A multi-cluster deep reinforcement learning approach," Appl. Energy, vol. 292, Art. no. 116940, 2021, doi: 10.1016/j.apenergy.2021.116940.

[109] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," IEEE Trans. Ind. Informat., vol. 7, no. 3, pp. 381–388, Aug. 2011, doi: 10.1109/TII.2011.2158841.

[110] W. Tushar et al., "Peer-to-peer trading in electricity networks: An overview," IEEE Trans. Smart Grid, vol. 11, no. 4, pp. 3185–3200, Jul. 2020, doi: 10.1109/TSG.2020.2969657.

[111] T. Morstyn, N. Farrell, S. J. Darby, and M. D. McCulloch, "Using peer-to-peer energy-trading platforms to incentivize prosumers to form federated power plants," Nat. Energy, vol. 3, pp. 94–101, 2018, doi: 10.1038/s41560-017-0075-y.

[112] J. R. Vázquez-Canteli, J. Kämpf, G. Henze, and Z. Nagy, "CityLearn v1.0: An OpenAI Gym environment for demand response with deep reinforcement learning," in Proc. 6th ACM Int. Conf. Syst. Energy-Efficient Buildings, Cities, Transp. (BuildSys), 2019, pp. 356–357, doi: 10.1145/3360322.3360998.

[113] A. Marot et al., "Learning to run a power network challenge for training topology controllers," Electr. Power Syst. Res., vol. 189, Art. no. 106635, 2020, doi: 10.1016/j.epsr.2020.106635.

[114] Z. Wang and T. Hong, "Reinforcement learning for building controls: The opportunities and challenges," Appl. Energy, vol. 269, Art. no. 115036, 2020, doi: 10.1016/j.apenergy.2020.115036.

[115] L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, and X. Guan, "A review of deep reinforcement learning for smart building energy management," IEEE Internet Things J., vol. 8, no. 15, pp. 12046–12063, Aug. 2021, doi: 10.1109/JIOT.2021.3078462.

[116] L. Thurner et al., "pandapower—An open-source Python tool for convenient modeling, analysis, and optimization of electric power systems," IEEE Trans. Power Syst., vol. 33, no. 6, pp. 6510–6521, Nov. 2018, doi: 10.1109/TPWRS.2018.2829021.

[117] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," IEEE Trans. Power Syst., vol. 26, no. 1, pp. 12–19, Feb. 2011, doi: 10.1109/TPWRS.2010.2051168.

[118] G. Dulac-Arnold et al., "Challenges of real-world reinforcement learning: Definitions, benchmarks and analysis," Mach. Learn., vol. 110, pp. 2419–2468, 2021, doi: 10.1007/s10994-021-05961-4.

[119] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," IEEE Trans. Smart Grid, vol. 10, no. 3, pp. 3125–3148, May 2019, doi: 10.1109/TSG.2018.2818167.

[120] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," ACM Trans. Inf. Syst. Secur., vol. 14, no. 1, Art. no. 13, 2011, doi: 10.1145/1952982.1952995.

[121] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," IEEE Trans. Smart Grid, vol. 8, no. 4, pp. 1630–1638, Jul. 2017, doi: 10.1109/TSG.2015.2495133.

[122] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nat. Mach. Intell., vol. 1, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.

[123] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 4765–4774.

[124] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Trans. Evol. Comput., vol. 6, no. 2, pp. 182–197, Apr. 2002, doi: 10.1109/4235.996017.

[125] M. Kuzlu, M. Pipattanasomporn, and S. Rahman, "Communication network requirements for major smart grid applications in HAN, NAN and WAN," Comput. Netw., vol. 67, pp. 74–88, 2014, doi: 10.1016/j.comnet.2014.02.008.

[126] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," IEEE Trans. Evol. Comput., vol. 1, no. 1, pp. 67–82, Apr. 1997, doi: 10.1109/4235.585893.

[127] J. Kennedy and R. Eberhart, "Particle swarm optimization," in Proc. Int. Conf. Neural Netw. (ICNN), 1995, pp. 1942–1948, doi: 10.1109/ICNN.1995.488968.

[128] Y. Li, Y. Ding, S. He, and F. Hu, "Artificial intelligence-based methods for renewable power system operation," Nat. Rev. Electr. Eng., vol. 1, pp. 163–179, 2024, doi: 10.1038/s44287-024-00018-9.

[129] A. K. Yousuf, Z. Wang, R. Paranjape, and Y. Tang, "An in-depth exploration of electric vehicle charging station infrastructure: A comprehensive review of challenges, mitigation approaches, and optimization strategies," IEEE Access, vol. 12, pp. 51570–51589, 2024, doi: 10.1109/ACCESS.2024.3385731.

[130] N. Fescioglu-Unver and M. Yıldız Aktaş, "Electric vehicle charging service operations: A review of machine learning applications for infrastructure planning, control, pricing and routing," Renew. Sustain. Energy Rev., vol. 188, Art. no. 113873, 2023, doi: 10.1016/j.rser.2023.113873.

[131] J. Y. Yong, V. K. Ramachandaramurthy, K. M. Tan, and N. Mithulananthan, "A review on the state-of-the-art technologies of electric vehicle, its impacts and prospects," Renew. Sustain. Energy Rev., vol. 49, pp. 365–385, 2015, doi: 10.1016/j.rser.2015.04.130.

[132] K. M. Tan, V. K. Ramachandaramurthy, and J. Y. Yong, "Integration of electric vehicles in smart grid: A review on vehicle to grid technologies and optimization techniques," Renew. Sustain. Energy Rev., vol. 53, pp. 720–732, 2016, doi: 10.1016/j.rser.2015.09.012.

[133] H. Shareef, M. M. Islam, and A. Mohamed, "A review of the stage-of-the-art charging technologies, placement methodologies, and impacts of electric vehicles," Renew. Sustain. Energy Rev., vol. 64, pp. 403–420, 2016, doi: 10.1016/j.rser.2016.06.033.

[134] A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 5998–6008.

[135] P. Prakash and D. K. Khatod, "Optimal sizing and siting techniques for distributed generation in distribution systems: A review," Renew. Sustain. Energy Rev., vol. 57, pp. 111–130, 2016, doi: 10.1016/j.rser.2015.12.190.

[136] A. Ehsan and Q. Yang, "Optimal integration and planning of renewable distributed generation in the power distribution networks: A review of analytical techniques," Appl. Energy, vol. 210, pp. 44–59, 2018, doi: 10.1016/j.apenergy.2017.11.006.

[137] D. E. Olivares et al., "Trends in microgrid control," IEEE Trans. Smart Grid, vol. 5, no. 4, pp. 1905–1919, Jul. 2014, doi: 10.1109/TSG.2013.2295514.