Appendix-A

***Parameters of HAI*:**

This section gives the hyperparameters of the proposed algorithms, including the numbers of layers and neurons of several modules in HAI.

***Parameters setting for IEEE-33 bus case*:**

Parameters setting of BC agent of HAI: (Regularization coefficient l_{BC} = 0.001)

BC agent		
1	Full connection Layers	
	Number of inputs	43
	Number of outputs	256
	Activation function	ReLU
2	Full connection Layers	
	Number of inputs	256
	Number of outputs	256
	Activation function	ReLU
3	Full connection Layers	
	Number of inputs	256
	Number of outputs	9
	Activation function	ReLU

Parameters setting of reward function: (Regularization coefficient l_R = 0.01)

Reward function		
1	Full connection Layers	
	Number of inputs	52
	Number of outputs	256
	Activation function	ReLU
2	Full connection Layers	
	Number of inputs	256
	Number of outputs	256
	Activation function	ReLU
3	Full connection Layers	
	Number of inputs	256
	Number of outputs	1
	Activation function	ReLU

Parameters setting of actor network of HAI: (Learning rate r_1=0.003)

Actor network		
1	Full connection Layers	
	Number of inputs	43
	Number of outputs	256
	Activation function	ReLU
2	Full connection Layers	
	Number of inputs	256
	Number of outputs	256
	Activation function	ReLU
3	Full connection Layers	
	Number of inputs	256
	Number of outputs	9
	Activation function	ReLU

Parameters setting of critic network of HAI: (Learning rate r_2=0.003)

Critic network		
1	Full connection Layers	
	Number of inputs	52
	Number of outputs	256
	Activation function	ReLU
2	Full connection Layers	
	Number of inputs	256
	Number of outputs	256
	Activation function	ReLU
3	Full connection Layers	
	Number of inputs	256
	Number of outputs	1
	Activation function	ReLU

***Parameters setting for IEEE-118 bus case*:**

Parameters setting of BC agent of HAI: (Regularization coefficient l_{BC} = 0.001)

BC agent		
1	Full connection Layers	
	Number of inputs	149
	Number of outputs	512
	Activation function	ReLU
2	Full connection Layers	
	Number of inputs	512
	Number of outputs	512
	Activation function	ReLU
3	Full connection Layers	
	Number of inputs	512
	Number of outputs	256
	Activation function	ReLU
4	Full connection Layers	
	Number of inputs	256
	Number of outputs	30
	Activation function	ReLU

Parameters setting of reward function: (Regularization coefficient l_R = 0.01)

Reward function		
1	Full connection Layers	
	Number of inputs	179
	Number of outputs	512
	Activation function	ReLU
2	Full connection Layers	
	Number of inputs	512
	Number of outputs	512
	Activation function	ReLU
3	Full connection Layers	
	Number of inputs	512
	Number of outputs	256
	Activation function	ReLU
4	Full connection Layers	
	Number of inputs	256
	Number of outputs	1
	Activation function	ReLU

Parameters setting of actor network of HAI: (Learning rate r_1=0.003)

Actor Network		
1	Full connection Layers	
	Number of inputs	149
	Number of outputs	512
	Activation function	ReLU
2	Full connection Layers	
	Number of inputs	512
	Number of outputs	512
	Activation function	ReLU
3	Full connection Layers	
	Number of inputs	512
	Number of outputs	256
	Activation function	ReLU
4	Full connection Layers	
	Number of inputs	256
	Number of outputs	30
	Activation function	ReLU

Parameters setting of critic network of HAI: (Learning rate r_2=0.003)

Critic network		
1	Full connection Layers	
	Number of inputs	179
	Number of outputs	512
	Activation function	ReLU
2	Full connection Layers	
	Number of inputs	512
	Number of outputs	512

	Activation function	ReLU
3	Full connection Layers	
	Number of inputs	512
	Number of outputs	256
	Activation function	ReLU
4	Full connection Layers	
	Number of inputs	256
	Number of outputs	1
	Activation function	ReLU

Appendix-B

This section gives the problem formulation of real-time scheduling of ADN. The ADN scheduling maximizes benefit through the management of power generation units. In other words, it attempts to minimize cost while ensuring the safe operation of the system. The objective function can be expressed as:

$$\min C = C_{\mathrm{MT}} + C_{\mathrm{RES}} + C_{\mathrm{ESS}} + C_{\mathrm{Trans}} \tag{B1}$$

The various costs can be expressed as:

$$\begin{cases} C_{\mathrm{MT}} = \sum_{i\in\mathcal{M}} \sum_{t=1}^{T} (a_i \cdot P_{\mathrm{MT},i,t}^2 + b_i \cdot P_{\mathrm{MT},i,t} + c_i)\Delta t \\ C_{\mathrm{RES}} = \sum_{i\in\mathcal{R}} \sum_{t=1}^{T} (\beta_{\mathrm{RES}} \cdot (\overline{P}_{\mathrm{RES},i,t} - P_{\mathrm{RES},i,t}))\Delta t \\ C_{\mathrm{ESS}} = \sum_{i\in\mathcal{S}} \sum_{t=1}^{T} (\beta_{\mathrm{ESS}} \cdot (P_{\mathrm{ESS,c},i,t} + P_{\mathrm{ESS,d},i,t}))\Delta t \\ C_{\mathrm{Trans}} = \sum_{t=1}^{T} (a_{\mathrm{T}} \cdot P_{\mathrm{Trant},t})\Delta t \end{cases} \tag{B2}$$

The operation of ADN subjects to:

$$\sum_{i:i\to j} (P_{ij} - P_{ij,\mathrm{loss}}) + P_j = \sum_{l:j\to l} P_{jl}, \quad \forall j \in \mathcal{N} \tag{B3}$$

$$\sum_{i:i\to j} (Q_{ij} - Q_{ij,\mathrm{loss}}) + Q_j = \sum_{l:j\to l} Q_{jl}, \quad \forall j \in \mathcal{N} \tag{B4}$$

$$P_{ij,\mathrm{loss}} = \frac{(P_{ij})^2 + (Q_{ij})^2}{(V_i)^2} r_{ij}, \quad \forall (i,j) \in \mathcal{E} \tag{B5}$$

$$Q_{ij,\mathrm{loss}} = \frac{(P_{ij})^2 + (Q_{ij})^2}{(V_i)^2} x_{ij}, \quad \forall (i,j) \in \mathcal{E} \tag{B6}$$

$$V_j^2 = V_i^2 - 2(r_{ij}P_{ij} + x_{ij}Q_{ij}) + (r_{ij}^2 + x_{ij}^2)\frac{P_{ij}^2 + Q_{ij}^2}{V_i^2}, \quad \forall (i,j) \in \mathcal{E} \tag{B7}$$

$$P_i = P_{\mathrm{MT},i} + P_{\mathrm{RES},i} + P_{\mathrm{ESS},i} - P_{\mathrm{D},i}, \quad \forall i \in \mathcal{N} \tag{B8}$$

$$Q_i = Q_{\mathrm{MT},i} + Q_{\mathrm{RES},i} - Q_{\mathrm{D},i}, \quad \forall i \in \mathcal{N} \tag{B9}$$

$$\underline{P}_{ij} < P_{ij} < \overline{P}_{ij}, \quad \forall (i,j) \in \mathcal{E} \tag{B10}$$

$$\underline{Q}_{ij} < Q_{ij} < \overline{Q}_{ij}, \quad \forall (i,j) \in \mathcal{E} \tag{B11}$$

$$\underline{V}_i \le V_i \le \overline{V}_i, \quad \forall i \in \mathcal{N} \tag{B12}$$

$$\underline{P}_{\mathrm{MT},i} < P_{\mathrm{MT},i} < \overline{P}_{\mathrm{MT},i}, \quad \forall i \in \mathcal{M} \tag{B13}$$

$$\underline{Q}_{\mathrm{MT},i} < Q_{\mathrm{MT},i} < \overline{Q}_{\mathrm{MT},i}, \quad \forall i \in \mathcal{M} \tag{B14}$$

$$-\underline{\eta}_{\mathrm{MT},i}\Delta t \le P_{\mathrm{MT},i,t} - P_{\mathrm{MT},i,t-1} \le \overline{\eta}_{\mathrm{MT},i}\Delta t, \quad \forall i \in \mathcal{M} \tag{B15}$$

$$0 < P_{\mathrm{RES},i} < \overline{P}_{\mathrm{RES},i}, \quad \forall i \in \mathcal{R} \tag{B16}$$

$$0 < P_{\mathrm{ESS,c},i} < \overline{P}_{\mathrm{ESS,c},i}, \quad \forall i \in \mathcal{S} \tag{B17}$$

$$0 < P_{\mathrm{ESS,d},i} < \overline{P}_{\mathrm{ESS,d},i}, \quad \forall i \in \mathcal{S} \tag{B18}$$

$$P_{\mathrm{ESS},i} = P_{\mathrm{ESS,d},i} - P_{\mathrm{ESS,c},i}, \quad \forall i \in \mathcal{S} \tag{B19}$$

$$\underline{E}_{\mathrm{ESS},i} < E_{\mathrm{ESS},i} < \overline{E}_{\mathrm{ESS},i}, \quad \forall i \in \mathcal{S} \tag{B20}$$

$$E_{\mathrm{ESS},i,t+1} = E_{\mathrm{ESS},i,t} + (\eta_{\mathrm{ESS,c}} P_{\mathrm{ESS,c},i,t} + P_{\mathrm{ESS,d},i,t} / \eta_{\mathrm{ESS,d}}) \Delta t, \ \forall i \in \mathcal{S} \tag{B21}$$

$$P_{\mathrm{ESS,d},i,t} \cdot P_{\mathrm{ESS,c},i,t} = 0, \quad \forall i \in \mathcal{S} \tag{B22}$$

Equations (B3) - (B6) are the power balance equations of ADN, while (B7) is the Distflow voltage equation. Equations (B8) and (B9) represent the active and reactive power injected at bus *i*. The upper and lower limits of branch power and node voltage are in the Equations (B10) - (B12). Equations (B13) - (B15) specify the output range and ramping speed of MT. The generation output of RES is assumed controllable and should be within its maximum available output, as shown in Equation (B16). Equations (B17) - (B22) represent the feasible region of ESS operation and time coupling constraints.

Appendix-C

Small size ADN:

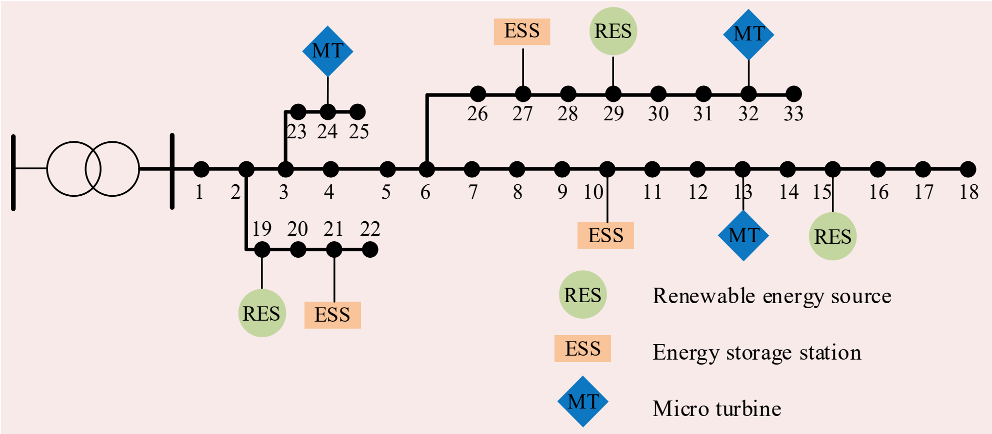

Fig. C1 Modified IEEE-33 bus ADN

Small size ADN is a modified IEEE-33 bus case with 3 MTs, 3 RESs, and 3 ESSs, as shown in Fig. C1. The power dispatch of ADN studied in this paper consists of 24 time periods per day, which means making real-time decisions every hour. To encourage more renewable energy absorption, β_{RES} is set as \$300/MWh. $\eta_{\text{ESS,c}}$ and $\eta_{\text{ESS,d}}$ are set to be 0.97 and 0.98 respectively. β_{ESS} is set to be \$0.9/MWh.

Expectation of load profile at each time is listed in Table C1. Expectation of RES is listed in Table C2. The forecasting errors of demand load, the error between actual RES output and its expectation value is assumed to follow the normal distribution N(0,0.03^2) and N(0,0.1^2), respectively.

Table C1 Expectation of demand load

Time-slot	1	2	3	4	5	6
$P_i^{\text{D}}(t)$ (kW)	720.29	780.31	840.34	870.35	876.35	840.34
$Q_i^{\text{D}}(t)$ (kVAR)	360.14	390.16	420.17	435.18	438.18	420.17
Time-slot	7	8	9	10	11	12
$P_i^{\text{D}}(t)$ (kW)	780.31	757.50	753.90	744.30	739.50	737.10
$Q_i^{\text{D}}(t)$ (kVAR)	390.16	378.75	376.95	372.15	369.75	368.55
Time-slot	13	14	15	16	17	18
$P_i^{\text{D}}(t)$ (kW)	732.30	720.29	660.26	540.22	480.19	420.17
$Q_i^{\text{D}}(t)$ (kVAR)	366.15	360.14	330.13	270.11	240.10	210.08
Time-slot	19	20	21	22	23	24
$P_i^{\text{D}}(t)$ (kW)	450.18	510.21	570.23	600.24	660.26	690.28
$Q_i^{\text{D}}(t)$ (kVAR)	225.09	255.10	285.12	300.12	330.13	345.14

Table C2 Expectation of maximum output of RES

Time-slot	1	2	3	4	5	6
$P_i^{RES}(t)$ (kW)	100	150	250	300	320	350
Time-slot	7	8	9	10	11	12
$P_i^{RES}(t)$ (kW)	320	300	250	140	100	150
Time-slot	13	14	15	16	17	18
$P_i^{RES}(t)$ (kW)	150	170	150	200	220	260
Time-slot	19	20	21	22	23	24
$P_i^{RES}(t)$ (kW)	300	320	330	220	200	100

Medium size ADN:

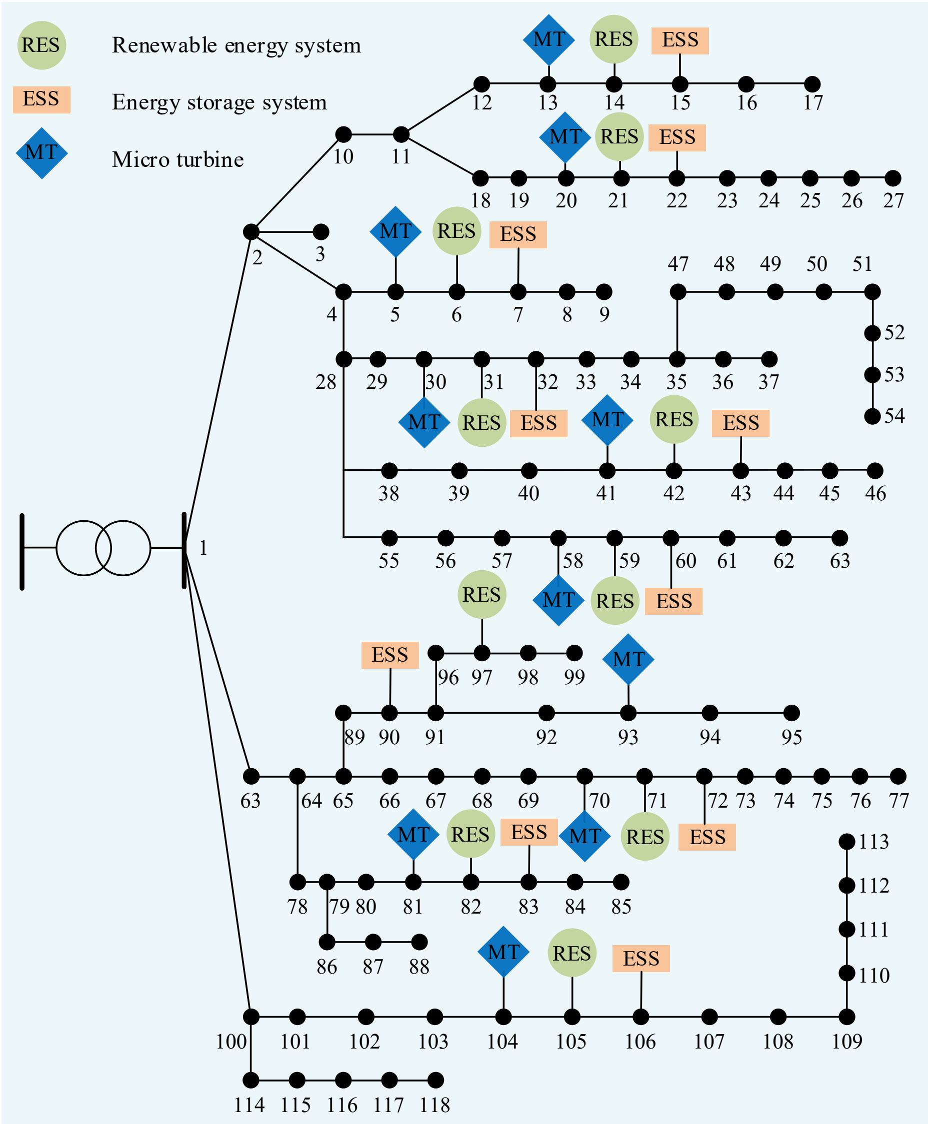

Fig. C2 Modified IEEE-118 bus ADN

Medium ADN is a modified IEEE-118 bus case with 10 MTs, 10 RESs, and 10 ESSs, as shown in Fig. C1. The power dispatch of ADN studied in this paper consists of 24 time periods per day, which means making real-time decisions every hour. To encourage more renewable energy absorption, β_{RES} is set as \$300/MWh. $\eta_{\mathrm{ESS,c}}$ and $\eta_{\mathrm{ESS,d}}$ are set to be 0.97 and 0.98 respectively. β_{ESS} is set to be \$0.9/MWh.

Expectation of load profile at each time is listed in Table C3. Expectation of RES is listed in Table C4. The forecasting errors of demand load, the error between actual RES output and its expectation value is assumed to follow the normal distribution N(0,0.03^2) and N(0,0.1^2), respectively.

Table C3 Expectation of demand load

Time-slot	1	2	3	4	5	6
$P_i^{\mathrm{D}}(t)$ (kW)	1800	1950	2100	2176	2191	2100
$Q_i^{\mathrm{D}}(t)$ (kVAR)	900	975	1050	1088	1095	1050
Time-slot	7	8	9	10	11	12
$P_i^{\mathrm{D}}(t)$ (kW)	1950	1894	1884	1860	1848	1842
$Q_i^{\mathrm{D}}(t)$ (kVAR)	975	947	942	930	924	921
Time-slot	13	14	15	16	17	18
$P_i^{\mathrm{D}}(t)$ (kW)	1830	1800	1650	1350	1200	1050
$Q_i^{\mathrm{D}}(t)$ (kVAR)	915	900	825	675	600	525
Time-slot	19	20	21	22	23	24
$P_i^{\mathrm{D}}(t)$ (kW)	1125	1275	1425	1500	1650	1726
$Q_i^{\mathrm{D}}(t)$ (kVAR)	612	637	712	750	825	863

Table C4 Expectation of maximum output of RES

Time-slot	1	2	3	4	5	6
$P_i^{\mathrm{RES}}(t)$ (kW)	60	90	150	180	192	210
Time-slot	7	8	9	10	11	12
$P_i^{\mathrm{RES}}(t)$ (kW)	192	180	150	84	60	90
Time-slot	13	14	15	16	17	18
$P_i^{\mathrm{RES}}(t)$ (kW)	90	102	90	120	132	156
Time-slot	19	20	21	22	23	24
$P_i^{\mathrm{RES}}(t)$ (kW)	180	192	198	132	120	60

Appendix-D

This section outlines the process for generating demonstration data for HAI.

For IEEE-33 bus case, the agent is trained using the ground-truth reward for 1000 training steps (equivalent to 24,000 simulation steps), as shown in Fig. D1(a). The actor network is checkpointed after every 25 training steps. For each checkpoint, the agent generates 50 trajectories. For IEEE-118 bus case, the agent is trained using the ground-truth reward for 2000 training steps (equivalent to 48,000 simulation steps), as shown in Fig. D1(b). The actor network is checkpointed after every 50 training steps. For each checkpoint, the agent generates 50 trajectories.

These trajectories provide a set of demonstrations of varying quality. In order to assess the impact of different levels of suboptimality (*conducted in Part F*), the paper conducted experiments using demonstrations from different stages of the agent (the 500th, 600th, 700th stage) for IEEE-33 bus case, named agent1~agent3, using demonstrations from different stages of the agent (the 600th, 800th, 1000th stage) for IEEE-118 bus case, named agent4~agent6.

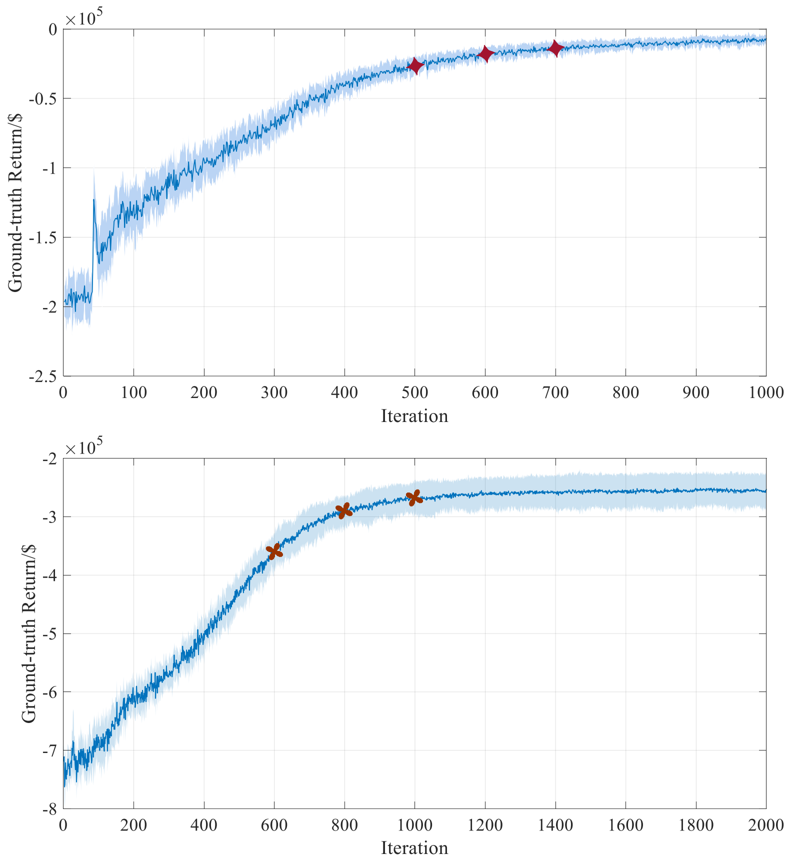

Fig. D1 Demonstrations generating process of IEEE-33 bus ADN (a) and IEEE-118 bus ADN (b).