Mathematical Model Formulation:

To enhance the applicability of the AMR scheduling model in warehouse environments, we incorporate three key features into the optimization framework:

(i) a multi-robot path conflict avoidance mechanism,

(ii) a partial charging strategy, and

(iii) a dynamic task scheduling mechanism based on scenario-dependent demand.

The resulting mixed-integer model, including indices, parameters, decision variables, objective function, and constraints, is summarized below.

1. Model Symbol Definition:
   a) Indexes:
      i. A: The set of AMRs, $a \in A$ indexes an AMR.
      ii. P: The set of pickup nodes
      iii. D: The set of delivery nodes
      iv. R: The set of requests, each request $r \in R$ is associated with one pickup node $i(r) \in D$ and one delivery node $j(r) \in P$
      v. C: The set of charging station nodes
      vi. N: The set of all nodes, including pickup nodes, delivery nodes, charging station nodes, and AMR depots
      vii. E: The set of directed edges (i, j) over N.
      viii. W: Set of scenarios modelling demand uncertainty and dynamic task occurrences, indexed by ω
      ix. E^{dec}: Set of decision epochs at which the scheduler makes a dispatching decision (e.g., an AMR becomes idle or a new task arrives).
      x. H: Set of scheduling rules (dispatching/charging rules) available to the rule-selection layer.

   b) Parameter:
      i. F_a: Fixed costs (e.g., purchase/maintenance cost) for AMR a.
      ii. $d_{(i,j)}$: The traveling distance from node i to node j.
      iii. c^{dist}: Travel cost per unit distance.
      iv. c^{ch}: Charging cost per unit energy.
      v. c^{tar}: Unit tardiness per unit time. $c^{tar} > c^{dist}$
      vi. $\tau_{(i,j)}$: AMR's baseline travel time on edge (i, j) without waiting or conflicts.
      vii. s_i: Service time at node i.
      viii. E_i^-, E_i^+: earliest and latest start times (time window) of service at node i.
      ix. Q_a: The maximum load capacity of AMR a.
      x. B_{max}^a: The maximum battery capacity of AMR a.
      xi. B_0^a: The initial battery value of AMR a.
      xii. ρ: Energy consumption rate per unit travel time.
      xiii. M: sufficiently large positive constant.
      xiv. q_i^ω: The demand of pickup node i under scenario ω. For tasks that do not appear, their demand can be set to 0 in the scenario ω.

xv. Δt^{safe}: minimum safety headway between two AMRs on the same edge or at the same node (if path conflicts are explicitly modelled).

xvi. p_{ω}: Probability of scenario ω.

xvii. $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7$: Weighting parameters (penalty coefficients) in the objective function, corresponding to: ($\alpha_6 \gg \alpha_7 \gg \alpha_3 \gg \{\alpha_1, \alpha_2, \alpha_4, \alpha_5\}$)

(a) α_1: Weight for the fixed fleet cost (priority of minimizing fleet size).

(b) α_2: Weight for the variable travel cost (priority of minimizing distance traveled).

(c) α_3: Weight for the tardiness penalty (priority of service punctuality).

(d) α_4: Weight for the energy cost (priority of energy efficiency).

(e) α_5: Weight for conflict-avoidance waiting time (priority of minimizing traffic congestion delays).

(f) α_6: Weight for task rejection (Highest priority soft constraint to maximize service level).

(g) α_7: Weight for the location-dependent waiting cost (controls dwelling behavior to prevent resource hogging).

xviii. κ: Charge rate (in energy units/s)

xix. η: Charging efficiency

xx. T_{max}^{ch}: Maximum charging time per visit to a charging station.

xxi. E_{max}^{ch}: Maximum charge amount per visit.

xxii. o_a: Origin depot of AMR *a*.

xxiii. u_a: Destination depot of AMR *a*.

xxiv. φ_i^{wait}: The location-dependent base cost for waiting at node *i*. To prevent charging station hogging, this parameter is set to a significantly higher value when $i \in C$ (charging stations) compared to when *i* is a depot or a pickup/delivery node.

c) Decision variables:

i. $x_{(i,j)}^{a,\omega} \epsilon \{0,1\}$: A binary variable indicating whether AMR *a* travels path (i,j). $x_{(i,j)}^{a} = 1$ means that AMR *a* traverses edge (i,j) under scenario ω, otherwise $x_{(i,j)}^{a} = 0$.

ii. $y_i^{a,\omega} \in \{0,1\}$: Whether AMR *a* is charging in station $i \in C$ or not under scenario ω.

iii. $m_{(i,j)}^{ab,\omega} \in \{0,1\}$: Precedence variable on edge (*i*,*j*) indicating whether AMR a passes before b.

iv. $n_{(i,j)}^{ab,\omega} \in \{0,1\}$: Precedence variable at node *k*.

v. $t_{(i,j)}^{ab,\omega} \geq 0$: Waiting time of AMR *b* due to avoiding *a* on edge (i,j).

vi. $z_r^{\omega} \in \{0,1\}$: Task acceptance decision under scenario ω ($z_r^{\omega} = 1$ if request r is served, $z_r^{\omega} = 0$ if it is rejected).

vii. $T_i^{a,\omega} \geq 0$: Service start time of AMR *a* at node *i* in scenario ω.

viii. $F_i^{a,\omega} \geq 0$: Departure time of AMR *a* at node *i* in scenario ω.

ix. $L_i^{a,\omega} \geq 0$: Tardiness at node *i* (time by which $T_i^{a,\omega}$ exceeds E_i^{+}).

x. $q_i^{a,\omega} \geq 0$: Charged energy of AMR *a* at node *i* in scenario ω.

xi. $\sigma_i^{a,\omega} \geq 0$: Charging time of AMR a at node i in scenario ω.

xii. $\Delta T_{i,j}^{a,\omega} \geq 0$: Actual travel time of AMR a on edge (i,j) in scenario ω.

xiii. $B_{arr,i}^{a,\omega} \geq 0$: Battery level of AMR a upon arrival at node i in scenario ω.

xiv. $B_{dep,i}^{a,\omega} \geq 0$: Battery level of AMR a upon departure from node i in scenario ω.

xv. $l_i^{a,\omega} \geq 0$: Load on AMR a when leaving node i in scenario ω.

xvi. $u_i^{a,\omega} \geq 0$: Generic waiting time of AMR a at node i (e.g., early arrival waiting).

xvii. $\pi_{e,h} \in (0,1)$: rule-selection variable; $\pi_{e,h} = 1$ if scheduling rule $h \in H$ is selected at decision epoch $e \in E^{dec}$, and 0 otherwise.

Later in the algorithm part, the RL policy will output probabilities over π_e. Here we keep π as a structural variable to formalize the rule-selection decision.

d) Objective Function:

We minimize the expected total cost over all demand/scenario realizations, combining fleet fixed cost, distance-based travel cost, energy-based charging cost, tardiness, conflict-avoidance waiting, prediction-driven standby and task rejection:

$$\begin{aligned}\min TC = \sum_{\omega\in W} p_\omega \Bigg(& \alpha_1 \sum_{a\in A} F_a \ (Fixed) + \alpha_2 \sum_{a\in A}\sum_{(i,j)\in E} c^{dist}\cdot d_{(i,j)} \cdot x_{(i,j)}^{a,\omega}\ (Travel\ cost) \\ & + \alpha_3 \sum_{a\in A}\sum_{i\in N} c^{tar}\cdot L_i^{a,\omega}\ (Tardiness\ penalty) + \alpha_4 \sum_{a\in A}\sum_{i\in C} c^{ch}\cdot q_i^{a,\omega}\ (Energy\ cost) \\ & + \alpha_5 \sum_{a,b\in A}\sum_{(i,j)\in E} t_{(i,j)}^{ab,\omega}\ (Conflict\ wait) + \alpha_6 \sum_{r\in R}(1 - z_r^{\omega})\ (Rejection\ penalty) \\ & + \alpha_7 \sum_{a\in A}\sum_{i\in N} \varphi_i^{wait}\cdot u_i^{a,\omega}\ (Location - dependent\ waiting\ cost) \Bigg)\end{aligned}$$

批注 [诗胡1]: 服务与效率，迟到扣分，移动扣能耗

批注 [诗胡2]: 拒单，动态惩罚，随着忙碌程度的变化，但是总体很高。优先级 Level 2

e) Constraints

The model's constraints are defined as follows: task execution constraints, time constraints, energy and charging constraints, collision avoidance constraints, and constraints regarding dynamic tasks and uncertainties. Each of these constraints applies to every uncertainty scenario ω within the set W, thereby ensuring the scenarios remain feasible amid varying demand fluctuations and task occurrences.

$$\sum_{a\in A}\sum_{j\in N} x_{(i(r),j)}^{a,\omega} = z_r^{\omega}, \sum_{a\in A}\sum_{j\in N} x_{(j(r),j)}^{a,\omega} = z_r^{\omega}, \qquad \forall r\in R, \forall \omega \in W \tag{1}$$

$$\sum_{j\in N} x_{(o_a,j)}^{a,\omega} - \sum_{i\in N} x_{(i,u_a)}^{a,\omega} = 0, \qquad \forall a\in A, \forall\omega\in W \tag{2}$$

$$\sum_{j\in N} x_{(i,j)}^{a,\omega} = \sum_{j\in N} x_{(j,i)}^{a,\omega} = 1, \ \ \forall i \in N\backslash\{o_a, u_a\}, \forall\omega\in W, \forall a\in A \tag{3}$$

$$F_i^{a,\omega} = T_i^{a,\omega} + s_i + \sigma_i^{a,\omega} + R_i^{a,\omega} + u_i^{a,\omega}, \forall a\in A, \forall i\in N, \forall\omega\in W \tag{4}$$

$$T_{j(r)}^{a,\omega} \geq F_{a,i(r)}^{a,\omega} - M(1 - \sum_{j\in N} x_{(i(r),j)}^{a,\omega}), \qquad \forall r\in R, \forall a\in A, \forall\omega\in W \tag{5}$$

批注 [诗胡3]: 防碰撞在 Constraints (4) 和 (5) 的时间窗逻辑里。最优先级

$$E_i^{-} \leq T_i^{a,\omega} \leq E_i^{+} + L_i^{a,\omega}, \qquad \forall i\in N, \forall a\in A, \forall\omega\in W \tag{6}$$

$$B_{dep,o_a}^{a,\omega} = B_0^a, \forall a \in A, \forall \omega \in W \tag{7}$$

$$0 \le B_{arr,i}^{a,\omega} \le B_{max}^a, 0 \le B_{dep,i}^{a,\omega} \le B_{max}^a, \forall a \in A, \forall i \in N, \forall \omega \in W \tag{8}$$

$$B_{arr,j}^{a,\omega} \le B_{dep,i}^a - \rho\tau_{i,j} + M\left(1 - x_{(i,j)}^{a,\omega}\right), \forall a \in A, \forall (i,j) \in E, \forall \omega \in W \tag{9}$$

批注 [诗胡4]: 电量约束在 Constraints (8)-(9) 中，硬约束放在了约束中。最优先级

$$\begin{cases} q_i^{a,\omega} \le E_{max}^{ch} y_i^{a,\omega}, \sigma_i^{a,\omega} \le T_{max}^{ch} y_i^{a,\omega}, \forall a \in A, \forall i \in C, \forall \omega \in W \\ q_i^{a,\omega} = \eta\kappa\sigma_i^{a,\omega}, \forall a \in A, \forall i \in C, \forall \omega \in W \end{cases} \tag{10}$$

$$B_{dep,\omega}^{a,\omega} = B_{arr,\omega}^{a,\omega} + q_i^{a,\omega}, \forall a \in A, \forall i \in N, \forall \omega \in W \tag{11}$$

$$l_{j(r)}^{a,\omega} = l_{i(r)}^{a,\omega} + q_{i(r)}^{\omega}, \forall a \epsilon A, \forall r \in R, \forall \omega \in W \ with \ x_{i(r)j(r)}^{a,\omega} = 1 \tag{12}$$

$$0 \le l_i^{a,\omega} \le Q_a, \qquad \forall a \in A, \forall i \in N, \forall \omega \in W \tag{13}$$

$$n_k^{ab,\omega} + n_k^{ba,\omega} = 1, \forall k \epsilon N, \omega \epsilon W \tag{15}$$

$$T_k^{b,\omega} \ge F_k^{a,\omega} + \Delta t^{safe} - M \cdot \left(1 - n_k^{ab,\omega}\right) - M \cdot (2 - \sum_j x_{(k,j)}^{a,\omega} - \sum_j x_{(k,j)}^{b,\omega}) \tag{16}$$

$$x_i^{a,\omega}, y_i^{a,\omega}, z_r^{\omega}, \pi_{e,h} \in (0,1), \forall a \in A, (i,j) \in E, i \in C, r \in R, e \in E^{dec}, h \in H, \omega \in W \tag{17}$$

$$T_i^{a,\omega}, F_i^{a,\omega}, L_i^{a,\omega}, q_i^{a,\omega}, \sigma_i^{a,\omega}, \Delta T_{i,j}^{a,\omega}, B_{arr,i}^{a,\omega}, B_{dep,i}^{a,\omega}, l_i^{a,\omega}, u_i^{a,\omega}, R_i^{a,\omega} \ge 0, \tag{18}$$

Constraints (1) ensure that each accepted request is visited exactly once and that its pickup and delivery nodes are served by the same AMR. If a request is rejected, none of its associated nodes can be visited. Constraints (2)–(3) impose flow conservation for each AMR. Every AMR leaves its origin depot exactly once, returns to its destination depot exactly once, and satisfies flow balance at all intermediate nodes, so that each vehicle follows a single continuous and logically consistent route. Constraint (4) defines the time balance at each node: the departure time equals the arrival time plus the service duration, the charging time, the prediction-driven standby time and any additional waiting at that node. Constraint (5) guarantees the temporal precedence between pickup and delivery for the same request and AMR. The service at the delivery node cannot start before the corresponding pickup operation has been completed. Constraint (6) enforces the time-window requirements. Service at each node must begin within its admissible time window; any violation of the latest start time is captured by the tardiness variable and penalized in the objective. Constraint (7) specifies the initial battery level of each AMR at its origin depot, providing the starting condition for the energy balance along the route. Constraints (8)–(11) jointly describe the battery dynamics and the partial charging mechanism. They link the arrival and departure battery levels on each arc to the energy consumed during travel and restrict charging to only visited charging stations. The charged amount and charging time at each station are bounded and coupled through the charging rate and efficiency, so that AMRs can perform partial recharges and only replenish enough energy to complete the current and subsequent task set. Constraints (12)–(13) govern the evolution of the carried load along each route and impose vehicle capacity limits. The load is increased at pickup nodes and decreased at delivery nodes of accepted requests, and must remain within the corresponding capacity always bound, which prevents any AMR from being assigned more work than it can physically carry. Constraint (14) formalizes the rule-selection layer. At each decision epoch, exactly one scheduling rule from the predefined rule set is activated. This structure allows the reinforcement-learning-based scheduler to operate on a discrete portfolio of dispatching and charging rules, rather than directly manipulating low-level routing decisions. Constraints (15) assume that if both AMR a and b pass through node k, one must go first and the other must go later. Constraints (16) confirm that if AMR a at node k first and both pass through k, then AMR b's start time must be later than a's departure time plus the safety margin. Constraints (17) define the domains of the binary decision variables by restricting routing, charging-on/off, task-acceptance, and rule-selection variables to $\{0,1\}$. This ensures a consistent mixed-integer programming formulation for the overall partial-charging, dynamic-task AMR scheduling problem. Constraints (18) specify the domains of the continuous variables, imposing non-negativity (and, where appropriate, upper bounds) on all time, load, waiting, charging-amount, and battery-level variables. These bounds guarantee that the continuous states remain physically meaningful and numerically stable during optimization.