SCIENCE CHINA
Information Sciences

• REVIEW •

November 2026, Vol. 69, Iss. 11, 210201:1–210201:42
https://doi.org/10.1007/s11432-026-5240-7

Special Topic: Intelligent Characterization and Modeling for Complex Systems

A comprehensive review of pedestrian modeling and simulation: taxonomies, methods, and prospects

Runkang GUO[1,2†], Zhengqiu ZHU[1,2†], Chuan AI[1,2], Bin CHEN[3], Suoyi TAN[1], Yue HU[1,2], Chen GAO[4], Yong LI[5], Quanjun YIN[1,2*] & Xin LU[1*]

[1]*College of Systems Engineering, National University of Defense Technology, Changsha 410073, China*
[2]*National Key Laboratory of Digital Intelligent Modeling and Simulation, National University of Defense Technology, Changsha 410073, China*
[3]*Institute of Intelligent Computing, University of Electronic Science and Technology of China, Chengdu 611731, China*
[4]*Beijing Information Science and Technology National Research Center, Tsinghua University, Beijing 100084, China*
[5]*Department of Electronic Engineering, Tsinghua University, Beijing 100084, China*

Received 20 June 2026/Revised 7 October 2026/Accepted 20 October 2026/Published online 19 November 2026

Abstract As characterized by its profound value in solving real-world challenges across traffic management, urban planning, and emergency evacuation, the analysis of crowd dynamics has garnered widespread attention, evolving into a prominent and indispensable modeling branch within the domain of complex systems. Despite substantial advancements, seamlessly integrating mechanistic priors with data-driven learning for the intelligent characterization and robust modeling of the profound heterogeneity and non-linear interactions inherent in real-world crowds remains a critical, unresolved challenge. To bridge this gap, this work presents a systematic and innovative review of pedestrian movement simulation methodologies through a novel taxonomy grounded in the structural interplay between data-driven intelligence and mechanism-driven modalities. Rather than offering a static summary of isolated methods, we meticulously trace the longitudinal developmental milestones and dynamic evolutionary trajectories across three foundational paradigms: mechanism-driven, data-driven, and hybrid fusion-driven approaches. Specifically, we comprehensively synthesize traditional physical rules and connectionist representation learning, while uniquely expanding the survey's boundary to incorporate the latest advancements in generative artificial intelligence, thereby capturing the paradigm shifts and developmental trends in the intelligent characterization of complex crowd systems. To delineate the core principles and operational traits of each category, a rigorous comparative matrix is constructed across multiple critical dimensions, including physical consistency, generalizability, and data dependency. Crucially, this review establishes a specialized classification framework for emerging data-mechanism fusion architectures, systematically categorizing their coupling mechanisms into three progressive dimensions: parameter-level bridging, structural embedding via differentiable computation graphs, and autonomous governing equation discovery via symbolic regression. Finally, we envision several promising future research avenues toward high-fidelity complex system modeling, encompassing large language model-driven synthetic data engines, self-adaptive dynamic fusion mechanisms, lightweight accelerated computing, and task-oriented unified evaluation benchmarks.

Keywords pedestrian movement, modeling and simulation, complex systems, data-mechanism fusion, physics-informed machine learning

Citation Guo R K, Zhu Z Q, Ai C, et al. A comprehensive review of pedestrian modeling and simulation: taxonomies, methods, and prospects. Sci China Inf Sci, 2026, 69(11): 210201, https://doi.org/10.1007/s11432-026-5240-7

1 Introduction

Pedestrian movement modeling and simulation represent fundamental techniques for characterizing complex crowd dynamics, with extensive applications spanning traffic management [1], urban planning [2,3], virtual reality [4–6], and emergency response [7–9]. Particularly in disaster scenarios such as fires or earthquakes, crowd evacuation simulation serves as an indispensable tool to assess emergency protocols and optimize evacuation strategies.

To this end, researchers from diverse disciplines—including mathematics, physics, biology, sociology, and psychology—have conducted systematic investigations over extended periods. Within the academic registry, effectively capturing pronounced individual heterogeneity [10] and intricate nonlinear interactions within crowds [11] has consistently served as the foundational catalyst driving methodological evolution. Recently, the rapid advancement of sensing technologies, big data analytics, and artificial intelligence has empowered researchers to transcend

* Corresponding author (email: yin_quanjun@163.com, xin.lu.lab@outlook.com)
† These authors contributed equally to this work.

© The Author(s) 2026. This article is published with open access at link.springer.com.

traditional boundaries, leveraging large-scale empirical datasets to uncover latent regularities in pedestrian motion and collective evolutionary patterns [12, 13]. Viewed from a system modeling perspective, these technological shifts have reshaped the paradigm landscape, dividing contemporary research into three predominant strategic approaches to achieve intelligent characterization: mechanism-driven methods, data-driven methods, and emerging hybrid data-mechanism fusion methods [14].

Early research predominantly relied on traditional mechanism-driven models. These models are largely based on principles of physical mechanics, treating pedestrians as interacting particles or rigid bodies and constructing dynamic equations to describe individual movement behaviors [15,16]. Such models not only effectively reproduce and explain basic crowd movement phenomena—such as collision avoidance and goal orientation—but also provide researchers with intuitive physical insights into complex group dynamics [17]. In recent years, data-driven approaches based on diverse machine-learning architectures have advanced rapidly. By leveraging large-scale observational datasets for training, these methods enable high-precision prediction and simulation of pedestrian trajectories in complex scenarios. For example, recurrent neural networks (RNNs) can efficiently process movement time series and capture spatiotemporal dependencies that mechanistic models often fail to represent [18–20]. More recently, simulation frameworks built on large language models (LLMs) have introduced a new paradigm for simulating realistic pedestrians: by exploiting LLMs' strong language understanding and reasoning capabilities, these frameworks not only reproduce physical movement patterns but also model the cognitive and decision-making processes underlying pedestrian behavior [21, 22]. Despite their efficacy, exclusively data-driven approaches exhibit inherent limitations. The variability in data quality, substantial acquisition costs, and the inherent opacity characteristic of black-box models [23] impede their application in mission-critical domains—such as crowd safety assessment, policy formulation, and intervention protocols—that necessitate stringent explainability [24,25]. To circumvent these impediments, the proliferation of advanced methodologies, particularly physics-informed machine learning [26] (PIML) and neural ordinary differential equations [27,28] (Neural ODEs), has spurred the exploration of a novel data-mechanism fusion paradigm. By embedding domain-specific physical priors within machine learning frameworks, these fusion models realize substantially augmented simulation fidelity and achieve a more robust and intelligent characterization of complex crowd systems while retaining strong generalizability and explainability.

Although the existing literature features numerous reviews on pedestrian movement modeling and simulation, their conceptual scopes remain predominantly expansive. Most focus extensively on macro- and micro-level taxonomies [29–31], crowd rendering methodologies [32, 33], collective dynamics [34, 35], and emergency evacuation protocols [36, 37]. From the perspective of complex system modeling, however, coupling data-driven intelligence with mechanistic knowledge is widely advocated as the route to modeling and governing such systems [38–40]. In cyber-physical-social spaces, big models and foundation intelligence are coupled with physical and social mechanisms to enable self-evolving control [38]; in the natural sciences, coupling physical process models with machine learning is regarded as the next step for data-driven modeling of the Earth system [39]. The objects of these studies, however, are systems themselves, so they neither specify how mechanism and data are fused for the modeling of individuals within a defined spatial domain, nor provide criteria for judging such fusion. Although preliminary dichotomies between knowledge-driven and data-driven frameworks have been proposed for pedestrian movement simulation, these divisions do not concern the underlying source of a model's dynamics and merely attend to whether a model uses or depends on pedestrian trajectory data, which renders the categorization conspicuously coarse [41, 42]. Consequently, such taxonomies fail to elucidate the intrinsic operational architectures of these paradigms, their specific prerequisites, and their task applicability. Finally, in the wake of accelerated advancements in computer science, big data, and artificial intelligence, incumbent reviews exhibit a pronounced lag in capturing frontier trends. By largely omitting nascent data-mechanism fusion strategies and pioneering approaches rooted in LLMs, the current literature fails to present a comprehensive and contemporaneous reflection of the discipline's evolutionary trajectory. In response, this review proposes a rigorous classification taxonomy predicated strictly on the intrinsic foundational paradigms of the models, thereby demarcating the critical boundary between mere data utilization and fundamentally data-driven intelligent characterization of complex crowd systems.

To bridge these gaps, this review makes three primary contributions.

• **Innovative taxonomy and evolutionary mapping.** To our knowledge, this review is the first account of the field organized by the source of a model's governing dynamics rather than by descriptive granularity or application domain. Based on the extent of mechanistic involvement in forward computation and the binding strength of the prior, we strictly distinguish data utilization from data-driven modeling. Rather than offering a static summary of isolated models, we map the dynamic evolutionary trajectory spanning from classical physical rules to state-of-the-art data-mechanism fusion architectures, tracing the mathematical principles of each stage and the migration of modeling logic between deductive and inductive sources.

• **Granular comparative analysis and fusion mechanism.** The review provides a detailed comparative

Table 1 Summary of prior reviews by their organizing principle and their coverage of the three modeling paradigms. Notes: $\surd$ = covered in depth; $\triangle$ = partially mentioned; $\times$ = not covered. Mech: mechanism-driven. Data: data-driven. Fusion: data-mechanism fusion.

Ref.	Year	Organizing principle	Mech	Data	Fusion
[34]	2017	Mechanism models (validation)	$\surd$	$\times$	$\times$
[43]	2017	Application (crowd rendering)	$\times$	$\triangle$	$\times$
[44]	2018	Knowledge vs. data	$\surd$	$\surd$	$\triangle$
[45]	2019	Application (crowd rendering)	$\times$	$\times$	$\times$
[31]	2020	Application (group/social dynamics)	$\surd$	$\times$	$\times$
[46]	2020	Application (crowd management)	$\times$	$\triangle$	$\times$
[47]	2021	Granularity (macro/meso/micro)	$\surd$	$\triangle$	$\times$
[37]	2021	Application (evacuation)	$\surd$	$\triangle$	$\triangle$
[48]	2021	Application (evacuation, emotion)	$\surd$	$\times$	$\times$
[41]	2022	Knowledge vs. data (trajectory prediction)	$\surd$	$\surd$	$\triangle$
[36]	2022	Application (emergency management)	$\surd$	$\times$	$\times$
[42]	2022	Data-driven architectures (rendering)	$\triangle$	$\surd$	$\times$
[14]	2024	Application (dense crowd)	$\surd$	$\triangle$	$\times$
[10]	2024	Agent heterogeneity (ABM)	$\surd$	$\triangle$	$\times$
[49]	2024	Application (evacuation)	$\surd$	$\triangle$	$\triangle$
[50]	2025	Application (traffic-pedestrian)	$\surd$	$\surd$	$\surd$
Ours	2026	Paradigm source (governing dynamics)	$\surd$	$\surd$	$\surd$

analysis of various models from the unique perspective of the operational roles and interplay between data and physical principles, offering an in-depth evaluation of their strengths, limitations, and applicable scenarios. It further explores the functional delineation and deep fusion mechanisms between data and physical priors, tailored to specific system modeling task requirements.

• **Core challenges and future directions.** The review systematically outlines the core challenges currently confronting the field, including data quality and privacy, the generalizability and interpretability of models, and the trade-off between computational resource allocation and simulation efficiency. On this basis, it identifies prospective research directions, providing a strategic roadmap for intelligent complex system modeling and characterization.

The remainder of this paper is organized as follows. Section 2 reviews the related literature and highlights the diverging perspectives among existing review articles. Section 3 details the literature retrieval and screening methodology, presents a statistical overview of the selected papers, and establishes the classification criteria and framework for the modeling paradigms discussed herein. Sections 4–6 systematically examine and compare mechanism-driven, data-driven, and data-mechanism fusion approaches for pedestrian movement modeling and simulation, specifically evaluating their respective strengths, limitations, and application scenarios. Section 7 discusses prevailing research challenges. Section 8 concludes the whole work and outlines promising future research directions for complex crowd system modeling and simulation.

2 Background and categorization

Pedestrian movement, at its foundational level, denotes the physical motion of autonomous individuals within bounded spatial environments, wherein individual-level spatial displacements fundamentally drive the emergent formulation of complex behavioral dynamics and collective phenomena at the crowd level. Characterizing this phenomenon, pedestrian movement modeling and simulation refers to the computational and mathematical frameworks designed for the intelligent characterization and quantitative replication of such spatiotemporal crowd patterns. As an inherently interdisciplinary domain spanning computer science, physics, mathematics, sociology, and psychology, this field has evolved over the past few decades from elementary rules of motion to advanced simulation platforms capable of capturing large-scale, heterogeneous, and complex systems, yielding a proliferating body of literature. Table 1 [10,14,31,34,36,37,41–50] catalogs representative review articles in this domain by their organizing principle and by their coverage of the three modeling paradigms, revealing that the fusion paradigm is the least covered by prior surveys.

The historical trajectory of this methodological evolution can be broadly understood through three developmental phases, each defined by its foundational modeling paradigm. Mechanism-driven models—rooted in deductive physical reasoning—constitute the earliest and most established strand. These approaches draw upon principles from Newtonian mechanics, social force theory, and cellular automata to construct explicit dynamic equations

governing individual motion, thereby offering transparent physical interpretability for phenomena such as collision avoidance, lane formation, and shock waves in dense crowds [29, 30]. Data-driven models, by contrast, adopt an inductive paradigm in which spatiotemporal regularities are learned directly from large-scale empirical observations. Architectures spanning recurrent neural networks, graph neural networks, and generative adversarial networks have demonstrated superior predictive accuracy in complex scenarios where hand-crafted rules face scalability bottlenecks [51, 52]. More recently, data-mechanism fusion models have emerged as a third paradigm, seeking to reconcile the interpretability of physical laws with the expressive power of learned representations through strategies such as physics-informed neural networks [26] (PINNs), Neural ODEs [27], and LLM-based agents that embed cognitive reasoning atop motion dynamics. This three-phase progression—from purely deductive, to purely inductive, to their principled integration—reflects a fundamental tension in the field: the quest for physical fidelity versus the need for data-driven adaptability.

Despite the richness of this literature, prevailing classification frameworks exhibit structural limitations that obscure this paradigmatic evolution. The most canonical taxonomy distinguishes models by descriptive granularity, macroscopic, mesoscopic, and microscopic [29, 30], yet this axis primarily reflects the level of behavioral abstraction rather than the intrinsic computational logic. Two models operating at the same granularity (e.g., social force and graph neural networks, both microscopic) may differ fundamentally in whether their predictive mechanics are derived from physical principles or learned from data. Similarly, application-oriented taxonomies, organized around evacuation dynamics [34, 36, 37], group behavior [31, 53, 54], or route choice [55], provide valuable domain-specific syntheses but tend to fragment the methodological landscape, obscuring cross-cutting trends in how data and physics interact across different application contexts. What both axes leave unspecified is the property that ultimately determines how a model behaves: the source of the dynamics that govern its motion, whether deduced from physical principles or induced from data. This criterion is not merely another descriptive label, because it fixes the model's forward computation and therefore conditions its interpretability, its data requirements and its capacity to extrapolate beyond the observed regimes. Applied consistently, it also separates two situations that neither axis can distinguish: a physical model whose parameters are calibrated against data, and a model in which data replace the inferential structure itself. Crucially, a pervasive conceptual ambiguity arises from the conflation of data utilization with data-driven modeling: several reviews classify traditional physical models as "data-driven" merely because their parameters are calibrated against trajectory databases [41, 42]. In reality, when the core algorithmic structure and inferential logic of a model remain entirely governed by physical equations, parameter calibration does not alter its paradigmatic essence, and the model remains firmly mechanism-driven. This distinction between using data to inform a physical model and using data to replace the model's inferential structure is critical, yet it has not been consistently drawn in the existing literature. We therefore adopt the source of a model's governing dynamics as the organizing criterion of this review, and give it an explicit, operational form in Subsection 3.2. To our knowledge, no prior review of pedestrian movement modeling and simulation organizes the field along this dimension.

To address these limitations and capture the field's current trajectory, a shift toward a paradigm-centric analytical lens is required. While a few reviews have begun to adopt a comparative perspective that juxtaposes knowledge-driven and data-driven approaches [41, 42, 52], their scope remains narrowly confined to the isolated task of trajectory prediction, failing to extend to the broader landscape of full-process crowd simulation. Moreover, existing surveys exhibit notable lags in covering emerging fusion paradigms; pioneering methods such as PINNs, Neural ODEs, and LLM-based approaches are either absent or only superficially mentioned. These gaps motivate the paradigm-centric classification framework adopted by this review, wherein every model is categorized according to whether its core predictive logic is deductively derived from physical mechanisms (mechanism-driven), inductively learned from data (data-driven), or constructed through the principled integration of both (fusion-driven). This paradigm-centric axis is not intended to replace granularity- or application-based taxonomies, but rather to complement them by exposing the fundamental epistemological differences that govern model behavior, interpretability, and transferability. The detailed categorization criteria and the resulting three-paradigm framework are elaborated in Section 3.

3 Methodology

3.1 Literature retrieval

As demonstrated in Figure 1, the domain of pedestrian movement modeling and simulation has enjoyed sustained academic interest over the last twenty years. The timeline of publication outputs (Figure 1(a)) delineates clear evolutionary stages. Starting from 103 papers in 2006, annual outputs expanded progressively to a maximum

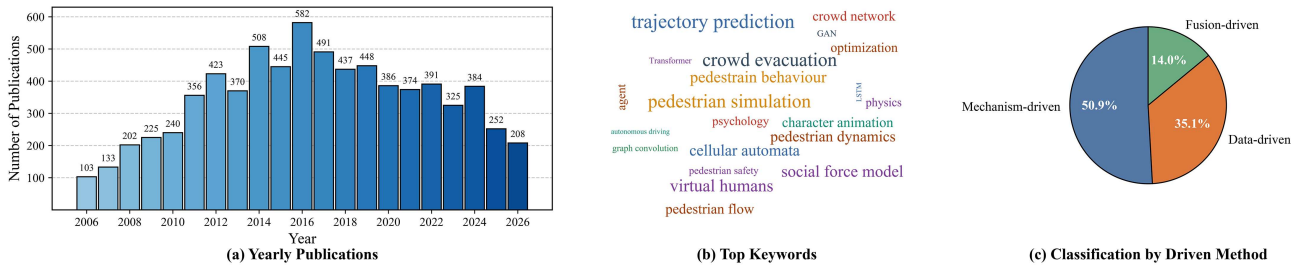

Figure 1 (Color online) Bibliometric and taxonomic analysis of the retrieved literature.

of 582 papers in 2016. Thereafter, the discipline transitioned into a resilient, high-yield plateau, maintaining an annual average of over 350 publications between 2017 and 2024 (where the marginal dip in the final two years is realistically a reflection of database indexing latencies). Additionally, the keyword co-occurrence word cloud visualized in Figure 1(b) highlights the core research concentrations. Regarding computational methodologies, traditional mechanism-driven paradigms, such as the social force model and cellular automata, continue to serve as the dominant cornerstones. Meanwhile, generative adversarial networks (GANs) and long short-term memory (LSTM) networks indicate the increasing adoption of deep learning approaches in pedestrian movement research, reflecting a growing integration of data-driven methods with traditional modeling frameworks.

Following the systematic literature review (SLR) guidelines [56], a systematic search was executed across six authoritative databases: Web of Science, ScienceDirect, Scopus, IEEE Xplore, ACM Digital Library, and Google Scholar. The retrieval query combines three dimensions with Boolean operators: the research subject (pedestrian or crowd), the modeling task (modeling or simulation), and the driving paradigm (data-driven, knowledge-based, or deep learning), formulated as TS = ((pedestrian OR crowd) AND (modeling OR simulation) AND (data-driven OR knowledge-based OR deep learning)). To balance recall against precision, the query terms were augmented with truncation and wildcards wherever the database syntax permits (e.g., pedestrian*, crowd*, model*, simulat*), and the identical logical expression was adapted to each database's field-tag convention. The search covered all records indexed up to September 2026 with no lower bound on publication year and was restricted to English-language records. In addition to the database search, backward and forward snowballing was performed under the same criteria to reduce the risk of omitting highly relevant studies.

As shown in Figure 2, a multi-stage screening process was applied using predefined inclusion and exclusion criteria. The records retrieved from the six sources were first merged and de-duplicated, after which their titles and abstracts were screened against the following criteria. A record was included if it (IC1) addressed pedestrian or crowd movement within a defined spatial domain at any granularity (microscopic, mesoscopic, or macroscopic), covering trajectory prediction, crowd dynamics simulation, evacuation modeling, and their evaluation; (IC2) reported an original methodological formulation or a rigorous empirical/simulation validation; (IC3) appeared as a peer-reviewed journal article, a full-length conference or workshop paper, or a scholarly book chapter providing a complete methodological exposition with at least one form of validation. A record was excluded if it (EC1) addressed human mobility or vehicular traffic at the urban/geographical scale without any micro-level treatment of pedestrian movement; (EC2) was a non-archival or non-peer-reviewed item, such as an abstract, poster, editorial, comment, technical report, white paper, standard, patent, or dissertation; (EC3) duplicated or substantially overlapped another study, in which case only the most complete version was retained.

To enhance the robustness of the review, a quality assessment (QA) procedure was performed before the final synthesis stage, evaluating each retained study against four criteria: study design, methodological transparency, validation rigor, and scientific contribution. Specifically, a study was retained only if (i) its modeling assumptions, scenario configuration, and experimental protocol were explicitly stated; (ii) its model formulation and parameter settings were reported in sufficient detail to be reproduced; (iii) its model was quantitatively validated against empirical data or established benchmarks rather than only qualitatively illustrated; and (iv) its contribution was clearly positioned relative to prior work. Studies failing any criterion were discarded, and borderline cases were adjudicated by consensus among the authors. Following the screening and QA procedures, 171 studies were retained for detailed analysis.

3.2 Literature classification

As illustrated in Figure 3, based on the core governing mechanisms of the models, this paper systematically categorizes the existing literature into three paradigm classes: mechanism-driven, data-driven, and data-mechanism

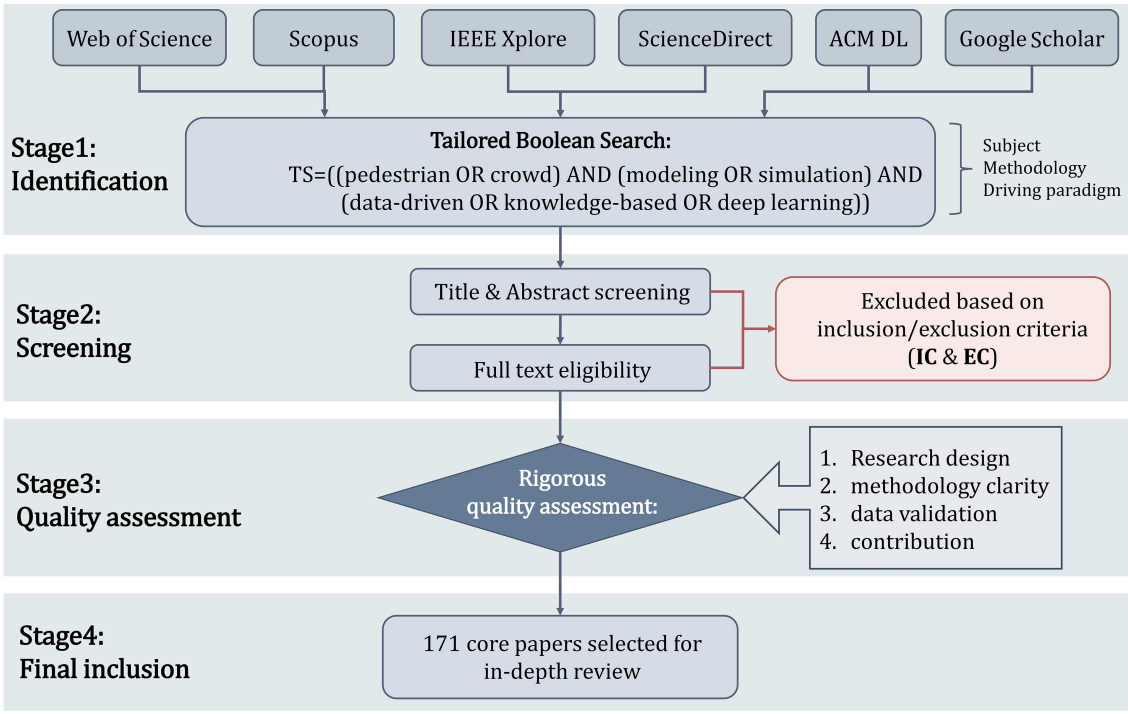

Figure 2 (Color online) Steps used to select and classify the papers.

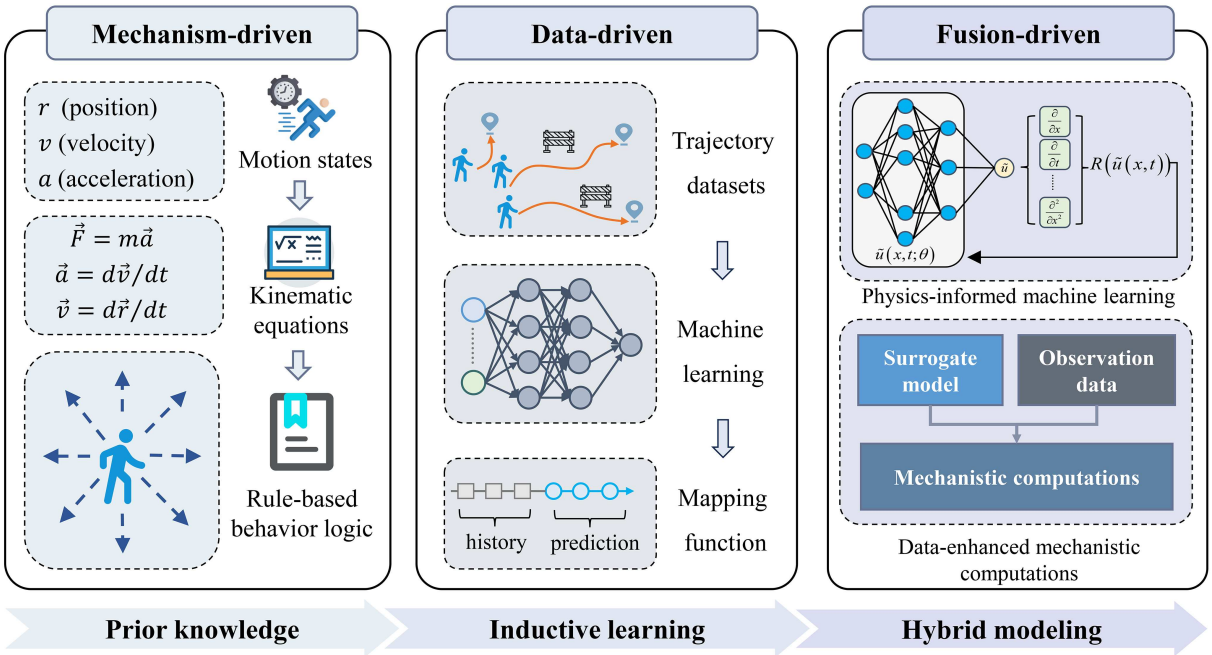

Figure 3 (Color online) Classification criteria and structural taxonomy of driving paradigms in pedestrian dynamics. The paradigms are systematically categorized into mechanism-driven, data-driven, and data-mechanism fusion-driven approaches according to whether the dynamics governing movement is deduced from physical principles or induced from data.

fusion-driven [57]. As argued in Section 2, granularity-based and application-based taxonomies describe where a model operates and at which resolution, but remain silent on what produces its motion. The paradigm-centric axis adopted here is therefore defined with respect to the source of a model's forward dynamics, that is, the mapping from the current system state to the next state.

To assign a model, we determine what produces its forward dynamics and how strongly mechanistic knowledge is bound to that computation. When the dynamics is produced by hand-specified equations or rules, where empirical

data supplies only parameter values or initial conditions, the model is mechanism-driven. When the dynamics is produced by a learned approximator, the decisive test is whether mechanistic knowledge participates in the forward computation. A prior may enter a learning system at different points and with different binding strength [26,58,59]. However, only a prior that participates in the forward computation, or that is indispensable to a capability the model claims (physical consistency, data efficiency, or extrapolation), qualifies the model as fusion-driven. A prior that acts solely on the training signal and can be removed without changing the model's functional class is an auxiliary regularizer, and the model remains data-driven.

(1) Mechanism-driven paradigm. Grounded in prior governing physics and behavioral heuristics, this paradigm formalizes pedestrian dynamics through rigorous mathematical formulations or rule-based logic. Its core philosophy lies in constructing explicit, interpretable models derived from classical principles such as Newtonian mechanics, kinematic equations, and predefined cognitive heuristics. These models mathematically parameterize pedestrian motion states (e.g., position, velocity, and acceleration) along with their local interactions with environmental geometries and neighboring agents.

(2) Data-driven paradigm. Operating under an inverse modeling framework, this paradigm leverages statistical learning and deep neural networks to autonomously discover movement patterns and interaction behaviors directly from large-scale trajectory datasets, bypassing the need for explicit physical axioms. The essence of this paradigm is to optimize high-dimensional mapping functions from raw input representations (e.g., historical trajectories, environmental context) to future behavioral targets, thereby capturing the latent, highly nonlinear spatiotemporal dependencies embedded within the underlying data manifold. Under the criterion above, this class also includes models trained on mechanism-generated or augmented data and models in which physical knowledge acts as an auxiliary regularizer, provided the forward mapping remains learned.

(3) Fusion-driven paradigm. Representing a synergistic integration of the two preceding paradigms, this emerging frontier characteristically embeds known physical constraints, kinematic laws, or behavioral rules into data-driven learning architectures in a structured manner, or conversely, deploys deep learning modules as data-driven surrogates to enhance mechanistic computations. Consequently, this paradigm achieves an optimal trade-off between the bounded interpretability of mechanistic models and the expressive adaptability of data-driven frameworks.

Following the systematic screening process, the 171 selected papers were classified according to the taxonomy established in Subsection 3.1. As illustrated in Figure 1(c), mechanism-driven models dominate the literature with 87 papers (50.9%), followed by data-driven models with 60 papers (35.1%). Meanwhile, research on fusion-driven models, although still the least populated paradigm, has expanded appreciably in recent years, comprising 24 papers (14.0%).

Because this axis is orthogonal to granularity and application, it can be combined with both and yields insight that neither axis provides. First, it renders the field's migration between deductive and inductive sources of dynamics directly observable, from hand-tuned equations and explicit statistical models, through data-calibrated and physics-regularized hybrids, to jointly co-designed and autonomously discovered dynamics. Second, it exposes the internal stratification of the fusion frontier, which granularity-based schemes cannot represent because hybrid models deliberately couple, for example, a macroscopic conservation law with microscopic learned operators.

4 Mechanism-driven pedestrian movement modeling and simulation

4.1 Methodological framework

Mechanism-driven pedestrian movement models describe pedestrian dynamics through predefined physical laws and behavioral rules. Drawing on first-principles formulations, such as Newtonian mechanics, kinematic constraints, and rule-based decision mechanisms, these models seek to reproduce observed movement patterns through appropriate parameter calibration. At any given time t, the state of pedestrian i can be represented by the following state vector:

$$\mathbf{s}_i(t) = (\mathbf{r}_i(t), \mathbf{v}_i(t), \theta_i(t), \mathbf{g}_i), \tag{1}$$

where $\mathbf{r}_i(t) \in \mathbb{R}^2$ and $\mathbf{v}_i(t) \in \mathbb{R}^2$ denote the spatial position and velocity vectors, respectively; $\theta_i(t)$ represents the heading angle; and $\mathbf{g}_i$ signifies the current goal position. Under this framework, the evolution of pedestrian dynamics is governed by a state transition function F, which is explicitly formulated via physical laws or behavioral rules.

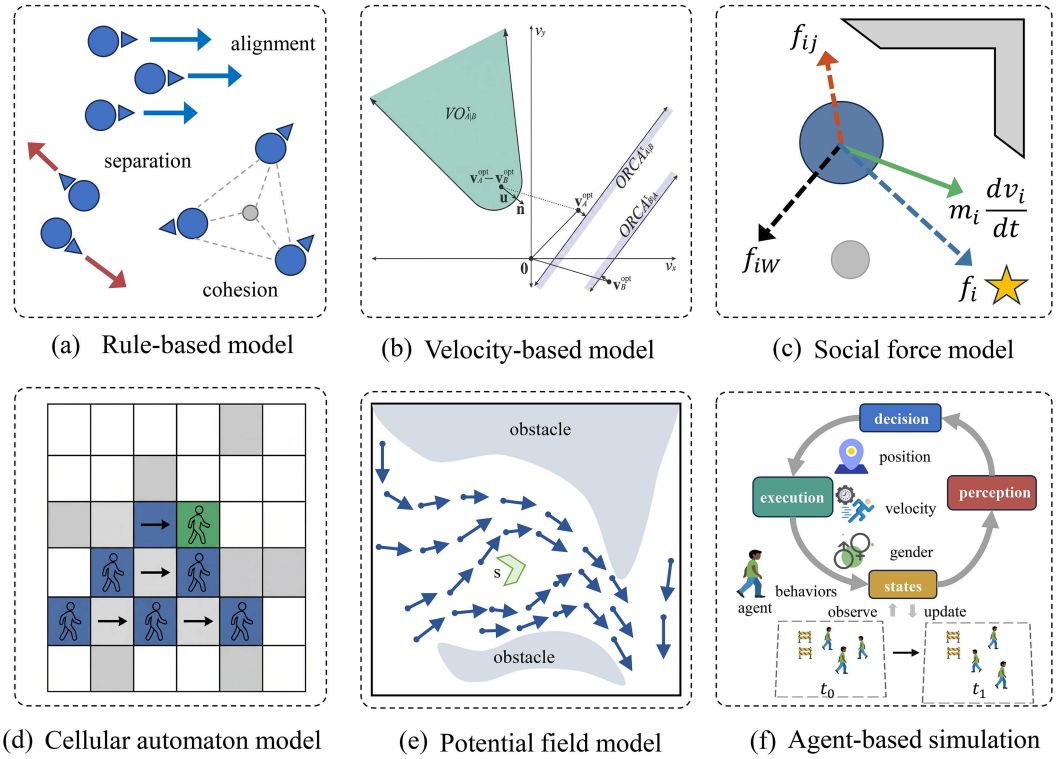

Figure 4 (Color online) Representative mechanism-driven methods for pedestrian movement modeling and simulation.

4.2 Representative models

4.2.1 *Rule-based models*

Rule-based models explicitly formulate pedestrian state transition mechanisms through predefined, concise logical rules. Under this framework, pedestrians are modeled as autonomous agents that adhere to specific behavioral protocols, whereby macroscopic crowd dynamics emerge organically from localized, rule-governed interactions. The classic Boids system typifies this paradigm (as illustrated in Figure 4(a)), relying strictly on three core geometric rules—separation, alignment, and cohesion—to simulate collective behaviors [52]. Although such models offer substantial advantages in computational efficiency, making them uniquely suited for tasks requiring mass parallel simulation with relaxed behavioral fidelity (e.g., rendering macroscopic background crowds in virtual reality or gaming environments [60]), their inherently rigid rule boundaries fundamentally limit their capacity to accurately characterize the highly complex, heterogeneous interactive behaviors exhibited by real humans.

To remedy the rigidity of deterministic rules and capture the inherent uncertainty in human behavior, researchers have extended mechanistic modeling toward stochastic processes and discrete choice frameworks. These approaches abstract pedestrian movement as Markovian evolution processes, introducing hand-crafted statistical priors to characterize bounded rational behaviors [61,62]. For instance, Lévy flights featuring heavy-tailed distributions have been elegantly deployed to elucidate the microscopic wayfinding mechanisms underlying pedestrian lane formation, capturing the fundamental trade-off between the exploration of unfamiliar spaces and the exploitation of goal-directed paths [63]. Furthermore, Logit-based discrete choice models formulate stepwise pedestrian transitions as utility maximization decisions conditioned on local environmental states [64].

Despite the inherent advantages of these models in computational efficiency and minimal implementation overhead, their paradigm-level limitations become prominent when re-examined in today's data-rich era. The core mechanisms and parameter vectors—such as transition probabilities and utility weights—heavily rely on expert heuristics; they lack rigorous physical or mechanical interpretations and possess virtually no capability for inductive learning from large-scale, real-world trajectory data. Consequently, regarding task applicability, these simplified rule-based and stochastic models struggle to execute demanding objectives, such as high-precision trajectory prediction in complex scenarios or safety assessments during high-density emergency evacuations. Currently, they primarily serve as lightweight algorithmic baselines or as low-level locomotion engines within multi-layer hybrid simulation frameworks.

4.2.2 *Velocity-based models*

Unlike paradigms relying on explicit physical force computations, velocity-based models reconceptualize pedestrian state transitions as localized geometric optimization problems within the velocity space. Rooted in the velocity obstacle (VO) framework, these methods utilize the Minkowski sum to reduce the underlying spatiotemporal collision mechanism to a set of constrained velocity cones. To mitigate the chronic trajectory oscillations induced by the constant-velocity assumption of neighboring agents in classic VO, the reciprocal velocity obstacle (RVO) model [65] introduces a reciprocity formulation wherein interacting parties symmetrically share collision avoidance responsibilities. This geometric reasoning culminates in the optimal reciprocal collision avoidance (ORCA) model [66], which elegantly abstracts multi-agent interactions into highly computationally efficient half-plane linear programming constraints:

$$ORCA^{\tau}_{A|B} = \mathbf{v} \mid (\mathbf{v} - (\mathbf{v}_A + \tfrac{1}{2}\mathbf{u})) \cdot \mathbf{n} \geqslant 0, \tag{2}$$

where $\mathbf{u}$ is the vector from the relative velocity $\mathbf{v}_A - \mathbf{v}_B$ to the closest point on the velocity obstacle boundary, $\mathbf{n}$ is its outward unit normal vector, and τ denotes the time horizon for collision-free navigation (Figure 4(b)). By delineating a feasible collision-free velocity space, this class of mechanisms achieves superior computational efficiency while guaranteeing absolute collision avoidance.

Although pure kinematic mechanisms are restricted to low-level, localized geometric collision avoidance and lack the intrinsic capacity to generate context-aware behaviors, velocity-based models exhibit remarkable extensibility for simulating complex crowd dynamics when integrated with high-level cognitive frameworks. For instance, by incorporating finite state machines (FSM) or belief-desire-intention (BDI) architectures, agents are endowed with event-triggered, high-level path planning and decision-making capabilities [67, 68]. Furthermore, by distinctively modulating the geometric constraint boundaries, this framework has been extended to simulate asymmetric pursuit-evasion adversarial behaviors in emergency evacuations [69] and to mitigate systemic congestion induced by temporarily stalled pedestrians [70].

Despite these algorithmic merits in collision avoidance efficiency, the foundational logic of velocity-based models remains firmly anchored in the idealized assumption of perfect rationality. This paradigm implicitly assumes that all pedestrians possess complete geometric observability and can flawlessly execute symmetric reciprocal avoidance. Devoid of inductive learning capabilities from large-scale empirical trajectory data, their confined parameter spaces inherently fail to dynamically adapt to the profound heterogeneity of real-world crowd behaviors.

4.3 Social force model

The social force model (SFM) proposed by Helbing et al. [71] remains one of the most prominent benchmarks in mechanism-driven pedestrian modeling to date. Rooted in Newtonian classical mechanics, this model conceptualizes pedestrian state transitions as a continuous spatiotemporal evolution process governed by the superposition of multiple virtual forces. By abstracting pedestrians as self-driven particles, the fundamental dynamic equation is formulated as

$$m_i \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = m_i \frac{\mathbf{v}_i^0 \mathbf{e}_i(t) - \mathbf{v}_i(t)}{\tau_i} + \sum_{j \neq i} \mathbf{f}_{ij} + \sum_{W} \mathbf{f}_{iW}, \tag{3}$$

where m_i and $\mathbf{v}_i(t)$ represent the mass and instantaneous velocity of pedestrian i at time t, respectively; $\mathbf{v}_i^0$ and $\mathbf{e}_i(t)$ denote the desired speed and direction of motion; τ_i is the relaxation time; $\mathbf{f}_{ij}$ quantifies the interpersonal forces, encompassing socio-psychological repulsion and collective attraction; and $\mathbf{f}_{iW}$ characterizes the physical repulsive force exerted by static environmental boundaries and obstacles. The indices $j \neq i$ and W range over all other pedestrians and all walls and static obstacles, respectively.

The primary advantage of the SFM lies in its absolute analytical transparency and strict physical consistency, enabling the autonomous reproduction of emergent macroscopic phenomena—such as arching at bottlenecks and spontaneous lane formation in corridors—without requiring prior empirical training [72].

The interaction terms of (3) admit an explicit closed form in the canonical formulation of Helbing et al. [71, 72]: the force exerted on pedestrian i by neighbor j decomposes into a socio-psychological repulsion, a body-compression force, and a sliding friction:

$$\mathbf{f}_{ij} = \left[A_i \exp\!\left(\frac{r_{ij} - d_{ij}}{B_i} \right) + k\, g(r_{ij} - d_{ij}) \right] \mathbf{n}_{ij} + \kappa\, g(r_{ij} - d_{ij})\, \Delta v_{ji}^t\, \mathbf{t}_{ij}, \tag{4}$$

where A_i and B_i control the strength and range of the repulsion, r_{ij} is the sum of the two body radii, $d_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$ the center-to-center distance, $\mathbf{n}_{ij} = (\mathbf{r}_i - \mathbf{r}_j)/d_{ij}$ the unit normal pointing from j to i and $\mathbf{t}_{ij}$ the corresponding

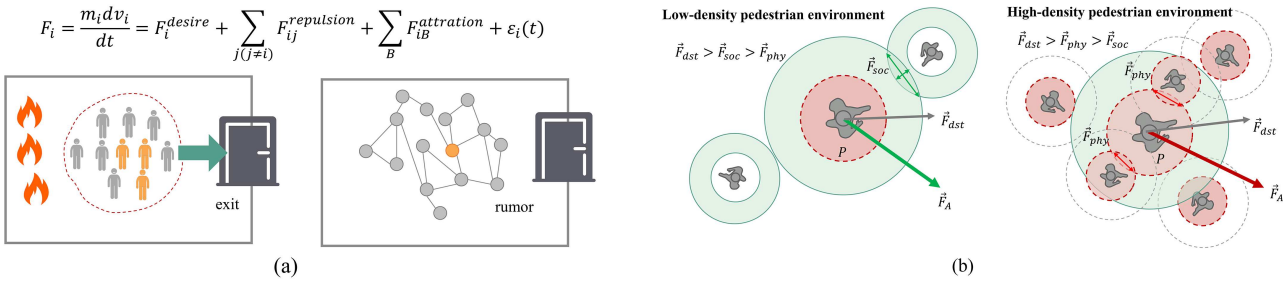

Figure 5 (Color online) Extensions and framework expansions based on the SFM. (a) Emergency evacuation framework that couples panic propagation with crowd dynamics [80]; (b) real-time adaptive pedestrian behavioral modeling approach grounded in the SFM [86].

unit tangent, k and κ the body-compression and sliding-friction coefficients, and $\Delta v_{ji}^t = (\mathbf{v}_j - \mathbf{v}_i) \cdot \mathbf{t}_{ij}$ the tangential velocity difference; $g(x) = x$ when the pedestrians overlap ($d_{ij} < r_{ij}$) and 0 otherwise. Because a wall is at rest, the tangential velocity difference reduces to the negative of the pedestrian's own tangential velocity, and the boundary force takes the same form once the wall quantities are substituted,

$$\mathbf{f}_{iW} = \left[A_i \exp\left(\frac{r_i - d_{iW}}{B_i} \right) + k\, g(r_i - d_{iW}) \right] \mathbf{n}_{iW} - \kappa\, g(r_i - d_{iW})\, (\mathbf{v}_i \cdot \mathbf{t}_{iW})\, \mathbf{t}_{iW}, \tag{5}$$

where r_i is the body radius of pedestrian i, d_{iW} the distance to wall W, $\mathbf{n}_{iW}$ and $\mathbf{t}_{iW}$ the wall normal and tangent, and the minus sign expresses that the friction opposes sliding along the wall. The repulsive term thus keeps a pedestrian clear of the wall, whereas the compression and friction terms penalize overlap and tangential sliding once contact occurs.

However, due to the lack of intrinsic mechanisms for autonomously inducing complex social interaction features from empirical observations, purely analytical equations must rely on incremental, ad-hoc extensions of the underlying mechanical terms when dealing with complex scenarios. At the physical contact level, to better describe physical characteristics in high-density crowd scenarios, researchers have explicitly incorporated non-smooth contact forces, sliding friction forces, and body compression forces to extend the $\mathbf{f}_{ij}$ term, thereby successfully simulating extreme stampede and domino effects [73, 74]. At the socio-psychological level, to characterize panic contagion, threat evasion, and group navigation behaviors embedded with social ties, customized virtual force functions—such as emotional contagion forces, asymmetric pursuit forces, and group cohesion forces—have been constructed [75–79]. Specifically, Ref. [80] integrated rumor propagation mechanics into the SFM, proposing an emergency evacuation framework that couples panic propagation with crowd dynamics, thereby revealing the profound impact of negative information dissemination on evacuation efficiency (Figure 5(a)). Although these deductive extensions significantly enhance the expressive capacity of the model, the mathematical superposition of multiple hand-crafted rules to capture complex behaviors tends to induce severe numerical instability in high-density regimes [81, 82].

Moreover, as the number of supplementary mechanical terms increases, the dimensionality of the SFM parameter vector expands significantly, posing severe tractability challenges for parameter calibration; consequently, the simulation fidelity becomes highly sensitive to precise parameter optimization. Currently, statistical frameworks—such as maximum likelihood estimation, multi-directional heuristic search, and online Bayesian inference—have been deployed to dynamically calibrate model parameters using real-world video analytics [83–85]. Specifically, Ref. [86] developed a real-time adaptive pedestrian behavioral modeling approach grounded in the SFM, wherein a dynamic parameter adjustment mechanism enables the framework to adaptively update social force parameters based on real-time observational streams (Figure 5(b)). This approach consistently outperforms traditional fixed-parameter SFMs across various density regimes, demonstrating the critical value of adaptive mechanisms in capturing complex real-world scenarios. Nevertheless, such methods essentially restrict parameter optimization within a rigid, predetermined equation framework, inherently failing to break through the representation boundaries of their prior mechanisms. Therefore, regarding task applicability, the SFM is predominantly utilized as a low-level physics engine to support multi-agent simulation platforms and virtual reality (VR) emergency evacuation training systems [8, 87–90].

4.3.1 *Cellular automata models*

Distinct from the continuous differential formulations governing social force models, cellular automata (CA) models discretize the continuous spatiotemporal domain of pedestrian movement into a regular grid of discrete cells. Under this paradigm, pedestrians are conceptualized as autonomous agents, with each individual occupying a single cell

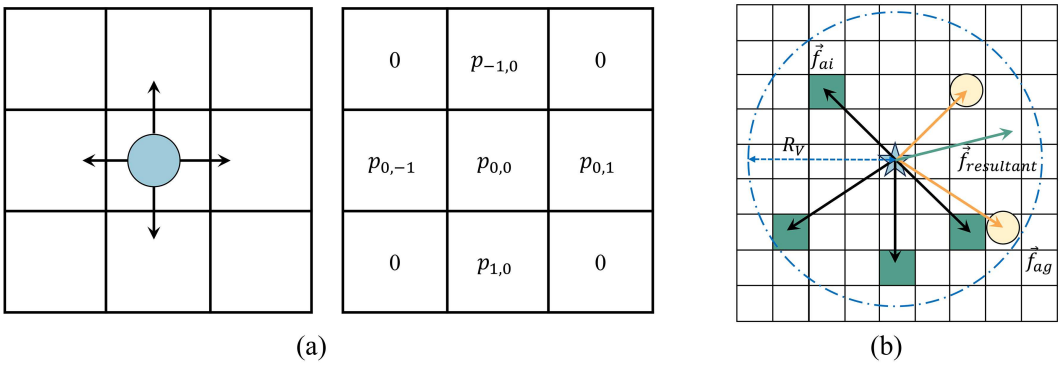

Figure 6 (Color online) Extensions and framework expansions based on the CA model. (a) Floor field [91]; (b) multi-field coupling mechanisms [102].

at any given time step, while their state updates are formalized via discrete probability transition matrices conditioned on localized neighborhoods. Early pioneering studies, notably by Burstedde et al. [91] and Kirchner and Schadschneider [92], establish the foundational floor field mechanism to compute transition probabilities toward adjacent cells (Figure 6(a)). The primary merit of CA models lies in their strict adherence to the spatial exclusion principle—whereby a cell can accommodate at most one agent—thereby eliminating the heavy computational overhead typically associated with continuous collision detection.

The state update is likewise specified in closed form through a floor field. Following the canonical formulation of Burstedde et al. [91] and Kirchner and Schadschneider [92], the probability that pedestrian i moves from its current cell to a neighbouring cell j during one time step is

$$p_{ij} = N_i \exp(k_S S_{ij} + k_D D_{ij}) (1 - n_{ij}), \qquad N_i = \left[\sum_{j \in \Omega_i} \exp(k_S S_{ij} + k_D D_{ij}) (1 - n_{ij}) \right]^{-1}, \tag{6}$$

where j runs over the neighborhood Ω_i of the current cell, S_{ij} and D_{ij} are the static and dynamic floor fields encoding respectively the distance to the exit and the trace left by pedestrians that have already passed, k_S and k_D the corresponding sensitivities, $n_{ij} \in \{0, 1\}$ the occupancy indicator that enforces the exclusion principle, and N_i the normalization. Updates are applied synchronously, and contention is resolved through the occupancy factor together with the stochastic sampling of p_{ij}.

To enhance the expressive capacity of CA frameworks in complex environments, substantial research has been dedicated to refining spatial discretization strategies and multi-field coupling rules. Regarding spatial discretization and kinematics, the traditional homogeneous grid has been extensively extended to better capture human body geometry and locomotion flexibility. For instance, Ji et al. [93] adopted triangular tessellations to provide more fine-grained directional choices; Feliciani et al. [94] introduced sub-cell intermediate positions at grid edges to modulate local density variations; and Huo et al. extended the classic Moore neighborhood to accommodate multi-step locomotion per iteration, thereby mechanistically replicating acceleration or stampede tendencies under emergencies [95]. To characterize high-density physical interactions, modifications to the traditional floor field [96,97] or the implementation of force floor fields [98,99] have been proposed to quantify pushing and avoidance behaviors. Furthermore, to simulate human visual constraints and anticipation, localized view fields [100] and anticipatory collision fields [94,101] have been integrated into the cell evaluation utility.

In emergency scenarios involving compound hazards, multi-field coupling mechanisms become remarkably prominent (Figure 6(b)). This is exemplified by the seamless integration of static floor fields, crowd dynamic fields, threat vectors, and guided following fields [102], as well as the establishment of dynamic coupling frameworks linking fire dynamics simulators with CA models [103]. Furthermore, rule sets predicated on visual perception—such as trap mechanisms, dynamically parameterized distance maps, and directionally blurred fields of view—have been proposed [104–107]. To capture complex avoidance and congestion patterns within multi-grid structures, the social heterogeneity and group characteristics of pedestrians have also been parameterized to modulate cellular transition probabilities [108–110]. Regarding computational infrastructures, the deployment of dynamic Monte Carlo sampling methods has significantly optimized simulation throughput for large-scale crowds [111], providing robust toolsets for evaluating the evacuation dynamics of heterogeneous populations in expansive facilities [112].

The highly discretized architecture of CA models endows them with unparalleled computational scalability, rendering them exceptionally well-suited for ultra-large-scale crowd flow simulations [113]. However, this inherent

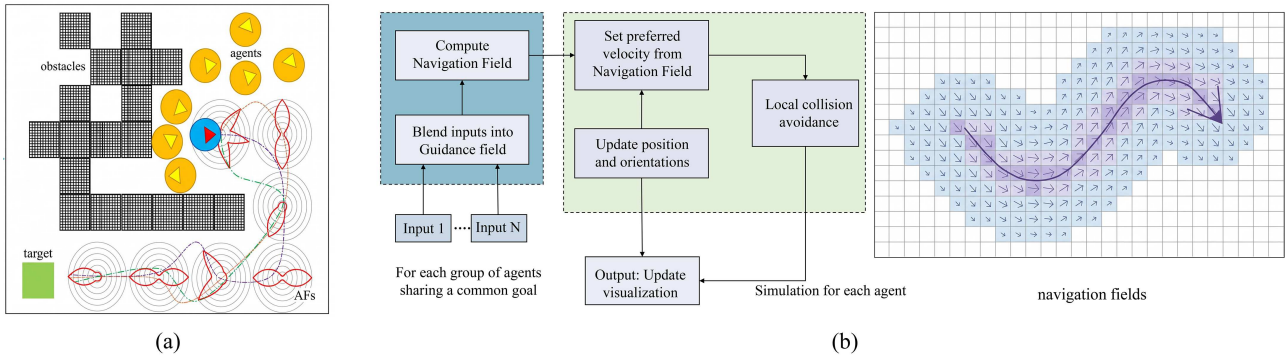

Figure 7 (Color online) Extensions and framework expansions based on potential fields. (a) Crowd simulation with the inclusion of the anisotropic layer [118]; (b) directing crowd simulations using navigation fields [116].

spatiotemporal discretization fundamentally restricts the faithful reproduction of fine-grained, continuous microscopic kinematics, frequently introducing unnatural artifacts such as grid anisotropy into the simulated trajectories. Consequently, researchers must rely heavily on expert domain knowledge to manually superimpose a cascade of virtual fields and subjectively tune a complex array of probability weights. This structural limitation severely hinders their applicability in continuous simulation tasks that demand high-precision microscopic trajectory fidelity.

4.3.2 *Potential field models*

The potential field model represents a continuous-space modeling paradigm conceptually inspired by classical field theories in physics [114]. Within this framework, pedestrian state transitions are formalized as a kinematic response to a global scalar or vector field. The underlying mechanism abstracts key environmental elements—such as target destinations, obstacles, and neighboring agents—as field sources that generate virtual gradients. Specifically, the spatial environment is mapped onto a global potential function, which is typically formulated via the linear superposition of attractive potentials emanating from the goal and repulsive potentials projected by obstacles. Pedestrian locomotion within this synthesized field is strictly governed by a gradient descent optimization process, wherein the instantaneous velocity vector is aligned with the negative gradient of the potential field, and the walking speed is monotonically correlated with the gradient magnitude. This top-down spatial mapping naturally unifies macroscopic path planning with localized collision avoidance, delivering exceptional computational efficiency when simulating crowd dynamics within structured, a priori known environments such as corridors and concourses.

To enhance the macro-level controllability and behavioral diversity of potential field paradigms, researchers have systematically enriched the mathematical formulations underlying field construction. The pioneering Flow Tiles approach [115] precomputes and assembles divergence-free velocity field templates to synthesize global guiding structures that direct massive crowds in a fluid-like manner, significantly improving the visual coherence of simulated trajectories. As shown in Figure 7(b), the interactive navigation field framework [116] superimposes user-defined guidance semantics—such as sketched paths or video exemplars—onto the baseline potential field, thereby enabling the quantitative editing of macroscopic crowd topologies, including spontaneous lane formations and vortices. Building upon this, Ref. [117] incorporated a dynamic modulation mechanism that computes correction factors based on real-time density feedback, achieving adaptive, field-driven path updating. In recent years, extensions of potential field theories have shifted focus toward quantifying behavioral heterogeneity and scaling computational throughput. For instance, Figure 7(a) illustrates that Ref. [118] introduced anisotropic formulations into the core architecture, capturing inherent crowd stochasticity by modeling the probabilistic distribution of locomotion headings. Concurrently, Ref. [119] established a cooperative framework coupling dynamic floor fields (DFF) and static floor fields (SFF) to actuate pedestrians with disparate intrinsic attributes, explicitly characterizing asymmetric queuing and avoidance behaviors. At the algorithmic infrastructure level, the PaCS parallel computing framework [114] decouples field-based motion into independent, separate computational units, leveraging multi-process synchronization strategies to drastically elevate the execution performance of ultra-large-scale crowd simulations.

Furthermore, the potential field mechanism has been extensively deployed as a global path prior to being deeply coupled with other microscopic simulation models. Within cellular automata architectures, dynamic potential fields are embedded to simulate high-speed running behaviors under panic conditions. An early exploration toward data-driven fusion manifested in the concept of interactive floor fields [120], wherein implicit spatial correlations are directly extracted from experimental data patterns to precisely model volume exclusion and following behaviors. In continuous physical frameworks, Ref. [90] integrated potential fields as the desired velocity direction input for

the self-driving force within the social force model. This integration effectively provides macroscopic global path guidance for microscopic kinematic equations, enabling the mechanistic optimization of evacuation strategies via meta-heuristic optimization algorithms.

The fundamental limitation of potential field models stems from their rigid spatial determinism assumption, which frequently induces local minima traps in complex topological environments, consequently causing unnatural algorithmic deadlocks. Because the mathematical parameterization of potential fields heavily relies on the manual linear superposition of hand-crafted geometric functions, these frameworks lack the capacity to induce high-order, non-Markovian spatiotemporal dependencies from empirical trajectory histories. Therefore, regarding task applicability, potential field models are exceptionally well-suited to serve as the navigation mesh (NavMesh) backbone in commercial game engines and to facilitate the macro-level simulation and control of large-scale crowd flows.

4.3.3 *Multi-agent simulation models*

Agent-based modeling (ABM) represents a pivotal paradigm for constructing complex crowd systems from a bottom-up perspective [121]. Under this framework, each pedestrian is explicitly formalized as an autonomous agent endowed with heterogeneous perception capabilities, behavioral heuristics, and goal-directed objectives. An individual agent typically encapsulates four core functional modules: perception, internal state, decision-making, and motion execution. ABM exhibits remarkable architectural extensibility, serving as a generalized mechanistic container that seamlessly integrates various low-level algorithmic solvers. At the motion execution layer, these highly extensible multi-agent frameworks widely embed advanced kinematic solvers to guarantee strict physical collision avoidance. For instance, Weiss et al. developed a position-based dynamics (PBD) system grounded in the Lagrangian formulation [122, 123], natively replicating collective patterns such as seamless merging and spontaneous lane formation; concurrently, Guy et al. [124] engineered a highly parallelized ClearPath collision avoidance framework leveraging free velocity obstacles and KD-trees. Other studies have coupled SFM or RVO density-dependent filters into ABM architectures [125, 126] to simulate utility-maximizing bottleneck flows in urban mobility scenarios [127] and heterogeneous vehicle-pedestrian mixed interactions [128].

The perception and internal state modules empower agents to capture profound pedestrian heterogeneity and complex socio-psychological dynamics. By integrating hybrid data-collection frameworks [129] and sensory differentiation profiles [130], individual features—such as visual or auditory scopes and spatial familiarity—are effectively quantified. At the socio-dynamic dimension, researchers have developed rigorous prior rule sets to simulate structured group coordination [131, 132] as well as intricate collaborative tasks including pushing, pulling, and carrying; certain works have even incorporated swarm intelligence and bio-inspired heuristics to actuate rigid formation behaviors [88]. In emergency evacuation scenarios, the visual perception of egress signage [133], dynamic exit selection mechanisms [134], and complex emotional contagion processes [135] are formalized as state transition logic [136]. An example of such emotional contagion-aware modeling is depicted in Figure 8, where deep reinforcement learning is applied to antagonistic crowd simulation. This formulation enables agents to dynamically trade off environmental topology, herd mentality, and adversarial interactions, ultimately providing high-fidelity decision support for disaster risk mitigation.

As the complexity of simulated behaviors escalates, the dimensionality of hand-crafted rules and parameter configurations within purely mechanism-driven ABMs grows exponentially. This scalability bottleneck inevitably induces a progressive divergence from real-world systems over extended temporal evolutions. To address this structural challenge, recent research has actively coupled data-driven mechanisms into foundational ABM frameworks. Prominent efforts include incorporating reinforcement learning architectures [137] and deploying ensemble Kalman filters for real-time data assimilation and state correction [138]. Furthermore, Ref. [139] advanced this paradigm by intersecting AI-enhanced digital twins with multi-agent evacuation modeling; by seamlessly integrating building information modeling (BIM), real-time sensor telemetry, and AI-driven predictive insights, they achieve dynamic evacuation optimization under seismic emergency conditions.

4.4 Development trends

Synthesizing the evolutionary trajectories of the six mechanism-driven model families reviewed above reveals three convergent developmental trends. Collectively, these trends delineate the paradigm's progression from its foundational origins toward its contemporary frontiers (Figure 9).

Trend 1. From single-factor kinematics to multi-factor socio-physical coupling. The inception of mechanism-driven modeling was predicated on highly reduced abstractions of crowd dynamics. Early frameworks relied primarily on purely kinematic or geometric rules, exemplified by the three rule-sets (separation, alignment, cohesion) of the Boids system [52] and the collision-avoidance principles of the original velocity obstacle framework [65]. The

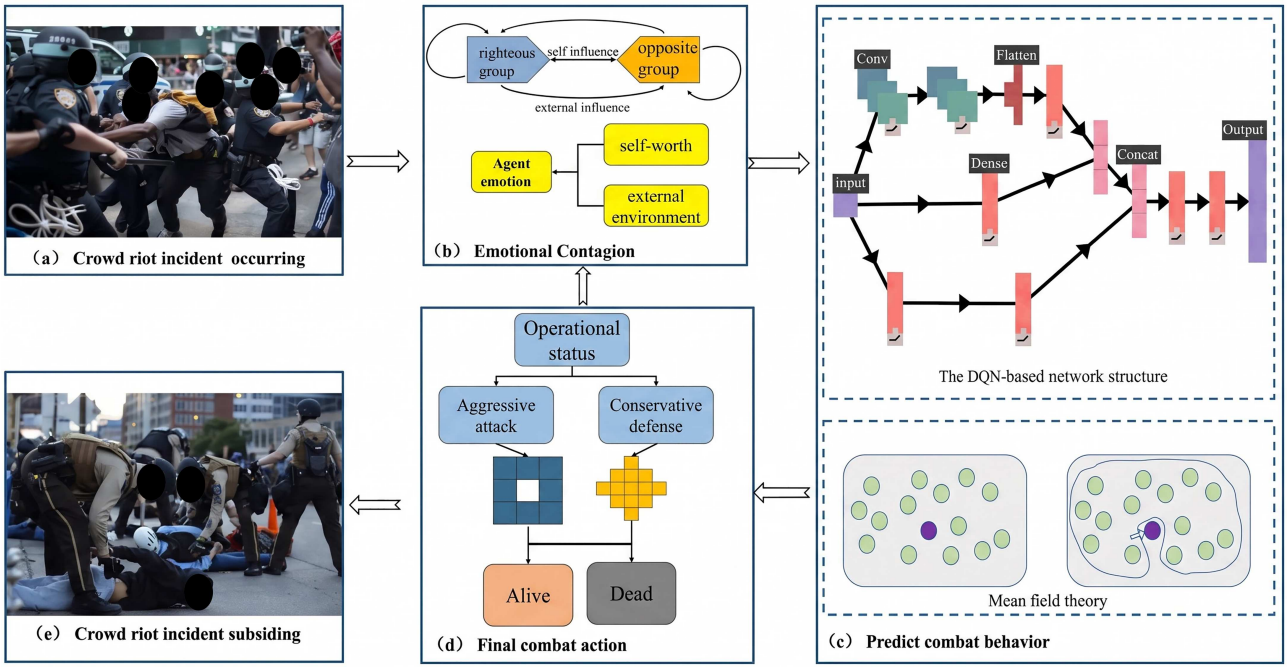

Figure 8 (Color online) Emotional contagion-aware deep reinforcement learning model for antagonistic crowd simulation [135].

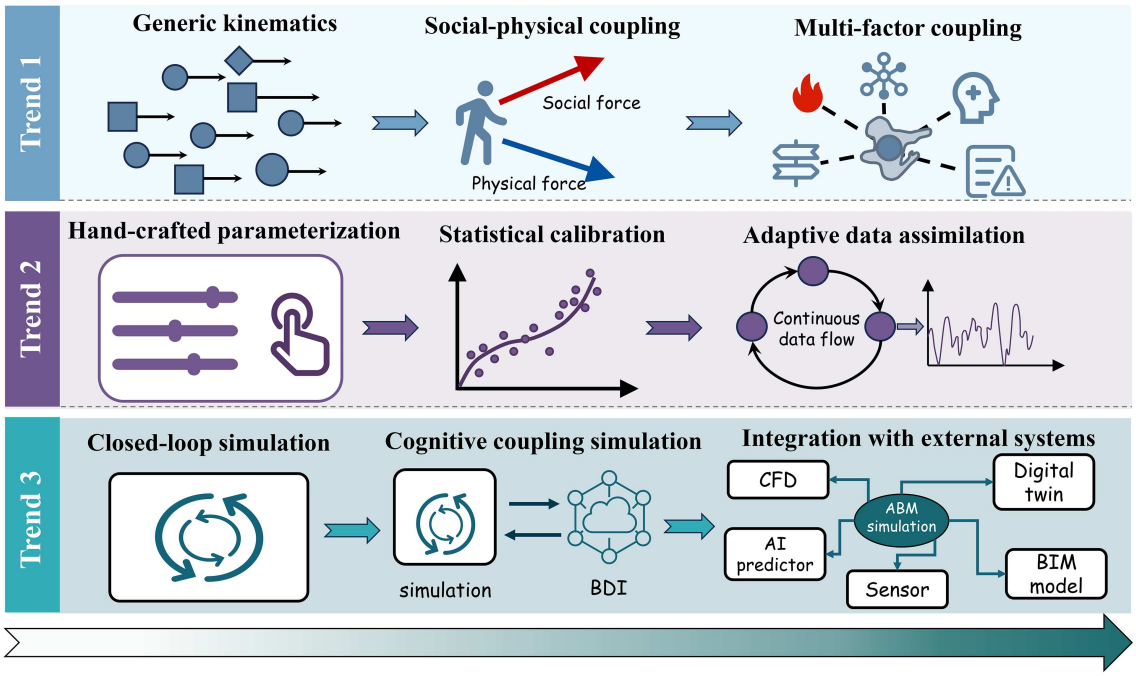

Figure 9 (Color online) Development trends of mechanism-driven pedestrian movement modeling and simulation.

seminal SFM [71] marked a pivotal paradigm shift by formally unifying socio-psychological repulsion with physical contact forces within a Newtonian framework. Subsequent SFM extensions progressively layered complex behavioral factors onto this mechanical scaffold, incorporating non-smooth friction for high-density stampedes [73, 74], emotional contagion for panic propagation [75, 80], group cohesion for socially bonded navigation [77–79], and asymmetric pursuit-evasion dynamics [76].

A parallel trajectory of factor enrichment is evident across CA and agent-based models. CA frameworks evolved from foundational static and dynamic floor fields [91, 92] to include threat vectors, guided following, and anticipatory collision fields [94, 101, 102], culminating in dynamic coupling with fire dynamics simulators for compound-hazard evacuations [103]. Similarly, ABM architectures progressively enriched the perception-internal state-decision

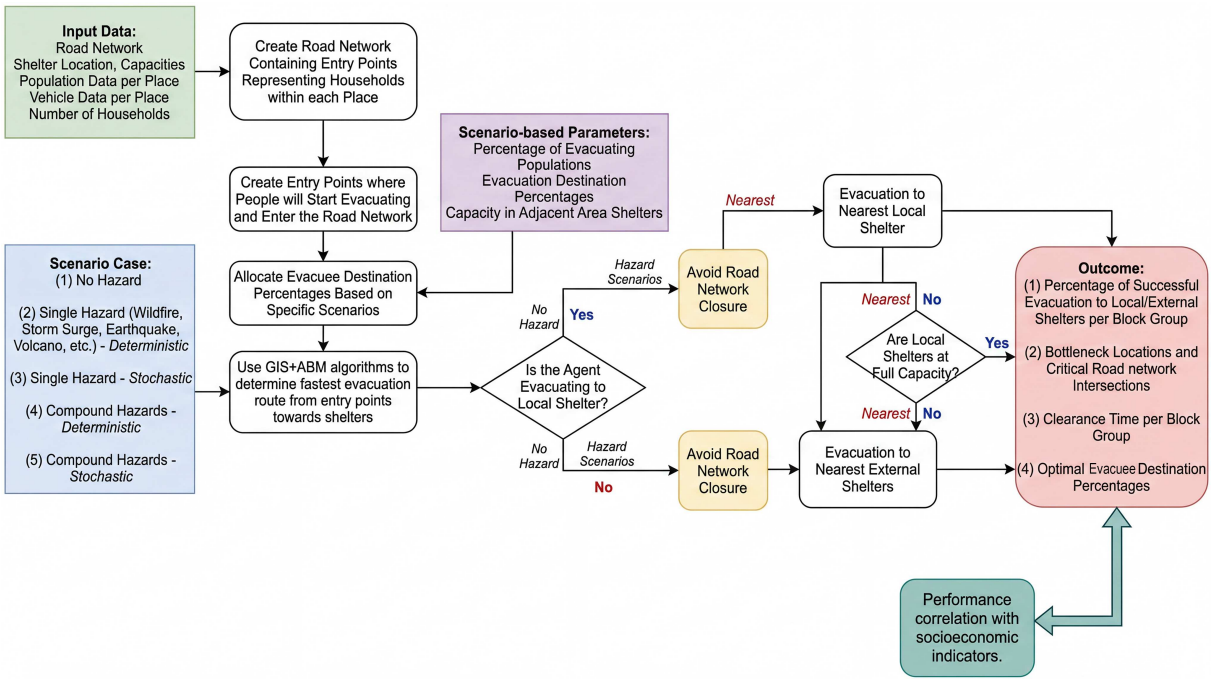

Figure 10 (Color online) Conceptual decision-support agent-based framework for evacuation planning under compound hazards [136].

pipeline, advancing from basic sensory inputs [129,130] to structured group coordination [131,132] and complex emotional contagion processes [135, 136]. A representative decision-support framework for such scenarios is illustrated in Figure 10. Characterized by incremental factor accumulation, this collective trajectory across all model families systematically expands the modeling scope from basic physical constraints to deeply intertwined socio-psychological mechanisms.

Trend 2. From hand-crafted parameterization to data-informed calibration and adaptive adjustment. Classical mechanism-driven models relied heavily on expert heuristics and manual tuning—a paradigm that inherently constrains reproducibility and generalizability. The transition toward rigorous statistical calibration initially gained momentum through maximum likelihood estimation frameworks [83] and multi-directional heuristic searches [84], which anchor model parameters to empirical trajectory distributions. This methodological evolution deepened with the integration of online Bayesian inference [84, 85], enabling dynamic parameter updates from streaming observational data. The real-time adaptive SFM [86] epitomizes this frontier; by continuously updating behavioral parameters via real-time sensor inputs, it consistently outperforms fixed-parameter counterparts across highly variable density regimes.

Parallel advancements in data-informed optimization are evident across other modeling domains. Within CA, the interactive floor field [120] extracts implicit spatial correlations directly from experimental data patterns, bypassing hand-crafted field definitions. Similarly, ABMs leverage ensemble Kalman filters [138] for real-time data assimilation and state correction, alongside AI-enhanced digital twin architectures [139] that dynamically calibrate parameters by integrating Building Information Modeling with real-time sensor telemetry. Collectively, this trajectory marks a fundamental methodological pivot: transitioning from a priori parameter specification to a posteriori data-driven calibration. This shift systematically mitigates reliance on rigid expert assumptions while substantially enhancing model fidelity to empirical dynamics.

Trend 3. From isolated closed-loop simulation to open-loop integration with external systems. Early formulations of mechanism-driven paradigms operated almost exclusively as self-contained, closed-loop systems. In these configurations, geometric rules generated trajectories in isolation, the SFM evolved solely via internal force equations, and CA models updated cell states based on static, predefined configurations. Conversely, contemporary frameworks exhibit a pronounced shift toward open-loop integration. In velocity-based modeling, coupling with finite state machines and belief-desire-intention architectures [67, 68] provides the external cognitive scaffolding necessary for event-triggered decision-making. Similarly, CA frameworks now leverage dynamic coupling with computational fluid dynamics simulators [103] to establish bidirectional feedback loops between pedestrian dynamics and evolving environmental hazards.

Furthermore, potential field models are increasingly deployed as global navigation priors embedded within micro-

scopic simulators—either supplying desired velocity vectors to drive SFM equations [90] or serving as the NavMesh backbone in commercial game engines. This interconnectedness culminates in agent-based digital twin architectures [139], which unify real-time sensor telemetry, Building Information Modeling, and AI-driven predictive modules into a synchronized, cyber-physical framework capable of dynamic evacuation optimization during seismic emergencies. Ultimately, this trend signals a paradigm shift from standalone simulation tools toward interconnected simulation ecosystems that continuously exchange data with both physical environments and external computational substrates.

4.5 Comparison of advantages and limitations of various mechanism-driven models

By synthesizing the critical discussions of the aforementioned mechanism-driven frameworks, this subsection delivers a systematic, multi-dimensional comparative matrix of the six primary model types. The evaluation spans key technical dimensions, including foundational deductive logic, physical consistency, parameter calibration complexity, affinity for real-time data assimilation, and ultimate task applicability. This comprehensive cross-examination is designed to enable readers to rapidly discern the core operational characteristics and boundary limitations of each modeling paradigm, as structured and summarized in Table 2 [13,44,60,64,67–70,75,84,95,114,119,129,134,138,140,141].

Physical consistency is rated high ($\surd$) when a physical law constrains the admissible motion by construction, whether through continuous force and contact dynamics or through an explicit collision-resolution constraint; moderate ($\triangle$) when physical or geometric reasoning enters the formulation, for instance as discretized kinematics, potential-field guidance or a penalty term, without constraining the admissible states; and low ($\times$) when the motion is produced by heuristic rules or a purely statistical mapping. Parameter calibration is rated high when micro-parameters are fitted to empirical trajectories or measured distributions and the resulting values are reported, moderate when they are adopted from the literature or only partly reported, and low when they are not calibrated or not specified. Data assimilation is rated high ($\surd$) when observations update the model state online during simulation, moderate ($\triangle$) when data are used offline for initialization or tuning only, and low ($\times$) when the model runs without observational input. Computational efficiency is rated high ($\surd$) when inference is near-analytic or rule-based and runs in real time; moderate ($\triangle$) when a single forward pass or a moderate solver or optimization overhead is required; and low ($\times$) when inference requires iterative sampling, repeated numerical integration, or otherwise prohibitive per-step cost.

Simple rule-based (SR) and VO models are fundamentally predicated on spatial geometry and kinematic boundaries. The underlying mechanism of SR formulations manifests as explicit geometric heuristics—such as the classic Boids model [60]—or probabilistic distributions. Functionally, this paradigm significantly alleviates complex physical force computations, thereby maintaining a remarkably low parameter calibration overhead. Consequently, their task applicability primarily targets macro-level route choice analysis when coupled with random utility theory [64], as well as the real-time rendering of large-scale background crowds where computational throughput is paramount [52]. In contrast, VO models frame state transitions as constrained linear programming tasks within the velocity space. At the functional level, this geometric formulation enforces strict, collision-free navigation by explicitly computing feasible velocity regions. This kinematic rigor renders VO models widely applicable in online navigation tasks demanding instantaneous real-time performance and absolute local collision avoidance [67,68,70], as well as in simulating adversarial pursuit-evasion dynamics governed by asymmetric game-theoretic constraints [69].

Conversely, the SFM and CA adopt continuous and discrete physical mechanisms, respectively. The core mechanism of the SFM is rooted in continuous Newtonian dynamics, functioning by mapping external environmental structures and socio-psychological semantics into explicit virtual force vectors. This framework preserves strict physical consistency, making it uniquely suited for demanding tasks involving dense physical contacts—such as evaluating mechanical body compression during extreme stampedes [140], simulating panic evacuations under coupled psycho-physical states [75], and executing real-time macroscopic crowd control derived from Bayesian estimation dynamics [84]. In sharp contrast, CA models operate on discretized state transition rules and virtual floor fields. Functionally, this spatiotemporal discretization natively enforces strict volume exclusion principles and localized flow-rate regulation. Consequently, CA frameworks are predominantly utilized to evaluate macroscopic flow dynamics and bottleneck throughput within high-density crowd regimes [95, 111]. Furthermore, in multi-hazard evacuation scenarios, their discrete grid infrastructure seamlessly facilitates dynamic coupling with environmental physics grids, such as computational fluid dynamics (CFD) smoke propagation fields generated by fire dynamics simulators (FDS) [103].

Potential field and agent-based models emphasize macro-level global guidance and micro-level cognitive evolution, respectively. The underlying mechanism of PF models is formulated as gradient descent optimization within a continuous potential topology. At the functional level, this top-down field configuration inherently mitigates local

Table 2 Comparison of mechanism-driven models: underlying principles, physical consistency, parameter calibration, data assimilation, computational efficiency, and applicable tasks. Notes: √ = high/yes, × = low/no, △ = moderate/partial. SR: simple rule-based. VO: velocity obstacle. SFM: social force model. CA: cellular automata. PF: potential field. Agent: agent-based model.

Ref.	Class	Underlying mechanism	Physical consistency	Parameter calibration	Data assimilation	Computational efficiency	Primary task applicability
[52]	SR	Explicit geometric heuristics	×	Low	×	√	Real-time rendering of large-scale background crowds
[60]	SR	Three-rule flocking kinematics	×	Low	×	√	Ultra-large-scale crowd animation for game engines
[64]	SR	Random utility theory & discrete action selection	×	Medium	△	√	Macroscopic path planning and preference analysis
[67]	VO	Velocity-space linear programming	√	Low	×	√	High-density local collision avoidance; real-time online navigation
[68]	VO	Local collision geometric constraints	√	Medium	△	√	Context-aware online navigation platforms
[70]	VO	Velocity-space linear programming	√	Low	×	√	Anti-penetration online navigation under high real-time constraints
[69]	VO	Asymmetric game-theoretic constraints	√	Medium	×	△	Adversarial pursuit-evasion crowd simulation
[140]	SFM	Continuous particle dynamics	√	High	△	×	High-density crowd evacuations; physical pressure analysis
[75]	SFM	Continuous dynamics & psycho-physical coupling	√	High	×	×	Simulating collective behaviors under panic-driven emotional states
[84]	SFM	Bayesian state estimation dynamics	√	High	√	×	Video-feedback-based real-time parameter tuning & short-term control
[95]	CA	Discrete state transition rules & discrete state spaces	△	High	△	√	High-density crowd flows; bottleneck throughput evaluations
[103]	CA	Floor field	△	High	△	√	Fire emergency and smoke propagation evaluations
[111]	CA	Stochastic state transition mapping	△	High	×	√	Evacuation efficiency benchmarks and spatial optimization
[13]	PF	Continuous potential field gradient navigation	△	Medium	×	√	Global crowd path guidance in structured environments
[114]	PF	Parallel potential field computations	△	Low	×	√	Large-scale, high-throughput crowd flow estimations
[141]	PF	Interactive floor potential fields	△	Medium	√	△	High-fidelity group movement pattern reproduction
[119]	PF	Attribute-cooperative floor fields	△	Low	×	√	Macroscopic guidance and routing of heterogeneous crowds
[129]	Agent	Distributed agent perception & decision-making	√	High	△	△	High-fidelity simulation with empirical psycho-physiological profiles
[134]	Agent	Distributed agent perception & decision-making	√	High	△	△	Strategic evacuation policy and structural design optimization
[138]	Agent	Distributed agent perception & decision-making	√	High	√	△	Sensor-driven dynamic digital twins and cyber-physical systems

pathfinding conflicts, rendering it exceptionally effective for global crowd guidance in structured environments [118], large-scale real-time flow computations via parallel scaling architectures [114], and the macro-level regulation of heterogeneous crowds through attribute-cooperative floor fields [119]. Concurrently, the incorporation of data-driven interactive floor fields enables the high-fidelity reproduction of emergent collective spatial movement patterns [141]. In contrast, agent-based models operate on a bottom-up, distributed modular architecture. Their core function decouples perception, internal state, and decision-making modules, thereby natively accommodating high-dimensional inputs of individual heterogeneity. Regarding task applicability, agent-based frameworks provide a robust infrastructure for executing complex simulations that embed realistic physiological and psychological profiles [129], as well as for optimizing macroscopic evacuation strategies [134]. Furthermore, their inherent modularity establishes them as a highly effective interface for incorporating real-time sensor streams and constructing synchronized digital twin frameworks [138].

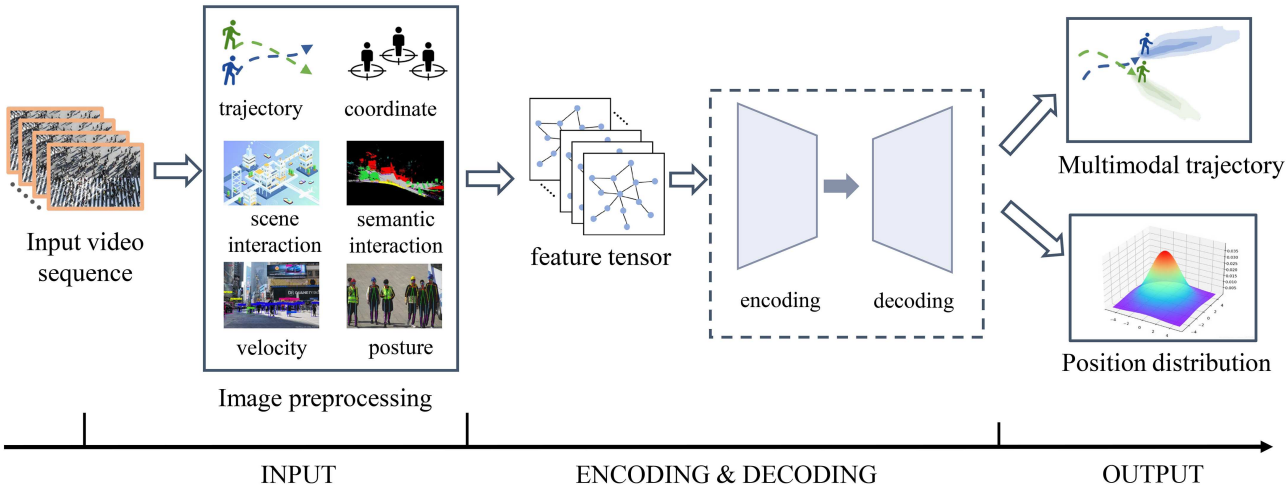

Figure 11 (Color online) Pipeline architecture of deep learning-based pedestrian trajectory prediction and continuous simulation.

5 Data-driven pedestrian movement modeling and simulation

5.1 Methodological framework

Data-driven pedestrian movement modeling and simulation represent a fundamental paradigm shift from deductive reasoning to inductive learning. As discussed in Section 4, mechanism-driven methods heavily rely on prior domain knowledge, constructing state transition logic and calibrating parameters through explicit physical formulations or heuristic rules. In sharp contrast, data-driven approaches abandon any rigid presupposition regarding motion generation mechanisms, effectively recasting the modeling problem as pure pattern recognition and feature representation learning. This modern paradigm leverages machine learning and deep learning algorithms to directly extract latent, highly nonlinear spatiotemporal interaction regularities from large-scale, real-world trajectories or visual telemetry signals, encapsulating them into highly parameterized predictive or generative functions [142, 143]. On one hand, empirical crowd behavior inherently involves extremely high-dimensional nonlinear interactions and complex socio-psychological contexts; consequently, traditional mechanism-driven models inevitably encounter a severe curse of dimensionality when attempting to fully capture these emergent features, manifesting as rule redundancy and intractable parameter calibration. Meanwhile, the widespread deployment of pervasive technologies (such as video surveillance, inertial sensing, and radio frequency identification) has facilitated the continuous accumulation of massive, high-fidelity pedestrian trajectory benchmarks, including ETH/UCY [144, 145], SDD [146], JTA [147], CrowdHuman [148], and Trajnet++ [149]. These extensive empirical datasets provide a rich optimization foundation that empowers neural networks to autonomously discover complex social interaction logic that remains remarkably difficult to formulate via explicit mathematical equations.

Although data-driven methods encompass a diverse array of technical trajectories—including statistical learning, deep neural networks, reinforcement learning, and generative models—they fundamentally share a unified formal framework from an abstract perspective, as illustrated in Figure 11. In terms of input representation, the model must structurally encode the current observational state $\mathbf{X}$, which typically features a deep integration of three key dimensions. The first is the individual motion state, which encapsulates the position sequences and velocity vectors within current and historical temporal windows to provide rich temporal context dependencies. The second is the social interaction state, which captures the relative topology of neighboring pedestrians; this is usually structured via adjacency matrices, graph networks, or attention feature maps to quantitatively measure spatial field influences. The third is the environmental semantic feature, covering scene geometric topologies (e.g., walls, static obstacles) and semantic annotations, which are frequently ingested as occupancy grids or scene embedding vectors. Different methodologies selectively emphasize distinct facets of this information: trajectory prediction concentrates heavily on temporal dependencies, crowd generation prioritizes global density distributions, and reinforcement learning uniformly formalizes the aforementioned dimensions into a continuous Markov state space.

The core of the data-driven paradigm is to build a parameterized neural network mapping that projects the high-dimensional input representation $\mathbf{X}$ to the target output $\mathbf{Y}$. According to the modeling objective and learning framework, this mapping can be divided into four distinct categories. First, spatiotemporal prediction models learn the conditional mapping from historical states to future positions, focusing on temporal dependencies and

the short-term inertia of individual motion. Second, social interaction modules aggregate neighbor information and capture local interaction patterns among pedestrians, where graph neural networks and attention mechanisms remain the current mainstream methods. Third, generative models capture the probability distribution of pedestrian trajectories or crowd states, enabling random sampling to produce multimodal trajectory outcomes. Fourth, policy learning models utilize reward signals within a reinforcement learning framework to guide pedestrian agents to learn collision avoidance, navigation, and cooperation strategies. Finally, the model can output deterministic future position sequences, multimodal probability distributions reflecting behavioral uncertainty, motion action commands for reinforcement learning, or complete spatiotemporal crowd state sequences for generative simulations.

5.2 Representative models

5.2.1 *Statistical learning methods*

As a foundational precursor of data-driven pedestrian modeling, statistical learning methodologies anchor themselves on the empirical distributions derived from observational datasets, effectively reconstructing the state transition function f_θ as a data-driven probabilistic formulation or a continuous stochastic process. The canonical modeling workflow typically entails extracting representative feature statistics from raw spatiotemporal trajectories, selecting an appropriate probabilistic architecture—such as stochastic differential equations, hidden Markov models (HMMs), or spatial point processes—for mathematical abstraction, and ultimately executing a global optimization of the parameter space $\boldsymbol{\theta}$ via maximum likelihood estimation (MLE) or rigorous Bayesian inference over empirical historical priors.

At the microscopic trajectory modeling dimension, pioneering studies [150] extend the classical Langevin equation to pedestrian dynamics, structurally decomposing locomotion into a deterministic drift component and a stochastic noise term, with both governing parameters statistically estimated from extensive real-world trajectory benchmarks. Within discrete state transition contexts—such as next-location prediction—Markov processes have been ubiquitously deployed [97, 151–155]. By formulating pedestrian trajectories as first-order (2-gram) or second-order (3-gram) Markov chains, the conditional transition probability $P(x_{t+1} \mid x_t)$ is derived directly from the empirical frequencies of historical movement data, thereby achieving data-driven spatial evolution. At the macroscopic crowd spatiotemporal evolution scale, recent methodologies have introduced spatiotemporal neural point processes to probabilistically formalize pedestrian arrival and departure events [156]. This approach automatically extracts macroscopic features—such as the sudden emergence of morning and evening rush-hour peaks—directly from data, effectively mitigating the macroscopic density distortions frequently encountered in purely microscopic simulations.

Statistical learning frameworks exhibit a salient methodological duality. By anchoring parameter estimation directly onto real-world empirical datasets, they successfully circumvent the intrinsic biases induced by the rigid rule-sets of traditional mechanism-driven models—such as classic radiation or intervention opportunity models—while preserving the rigorous mathematical explainability that modern deep learning black boxes lack. Nevertheless, the representational capacity of this paradigm remains strictly bounded by its foundational, a priori statistical assumptions. For instance, the memoryless property of Markovian formulations fundamentally prevents them from encoding long-sequence historical context dependencies, presenting a clear bottleneck when addressing high-dimensional, nonlinear, and implicit multi-agent social interactions. To breach these statistical constraints, Ref. [157] engineered a hierarchical two-layer pedestrian locomotion framework: the macroscopic layer encapsulates global crowd flow typologies, while the microscopic layer models localized individual movement fidelity, substantially enhancing simulation realism via this dual-layer cross-scale coupling.

5.2.2 *Trajectory prediction and crowd simulation based on deep learning*

Deep learning methods represent the current dominant paradigm for data-driven pedestrian modeling. Leveraging the exceptional nonlinear fitting capabilities of deep neural networks, these approaches learn the mapping function f_θ from historical states to future trajectories in an end-to-end manner directly from large-scale annotated trajectory benchmarks. Within this framework, connectionist models implicitly encapsulate complex behavioral patterns—such as individual velocity preferences and emergent group-following dynamics—via coupled temporal encoding and social interaction modules, subsequently outputting either deterministic future trajectories or multimodal probability distributions. Crucially, given that existing literature reviews [41, 158–161] have already provided exhaustive taxonomies and mechanistic analyses of the underlying deep architectures—such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), graph convolutional networks (GCNs), and Transformers—along with comprehensive compilations of standard training datasets and evaluation metrics, such foundational descriptions are omitted here. Instead, to avoid redundant exposition, this review strictly focuses on the methodological

transformation of these deep learning models from short-term "forecasters" into generative "simulators", as well as their specialized applications in complex scenario modeling.

(1) Temporal modeling and kinematic adaptation. Early methodologies predominantly utilize RNNs, exemplified by the pioneering Social-LSTM [18], as core computational backbones, employing memory gating and social pooling mechanisms to encapsulate spatiotemporal dependencies. To bypass the severe computational bottlenecks inherent in such sequential processing architectures, CNNs and temporal convolutional networks (TCNs)—which offer exceptionally high parallel computing efficiency and precisely controllable receptive fields—have increasingly gained traction in large-scale simulation pipelines. For instance, to model unidirectional and bidirectional pedestrian flows, Ref. [162] engineered a highly parallelized, CNN-based crowd simulation framework. To address the more computationally challenging right-angle turning scenarios, Jiang et al. [163] proposed a hierarchical two-stage deep learning framework consisting of a scene perception (SP) layer and a motion dynamics (MD) layer, achieving significantly higher fidelity and accuracy in localized turning kinematics compared to traditional mechanism-driven formulations. Furthermore, Ref. [164] integrated a visual information-driven model to decisively enhance the framework's cross-scene generalization capabilities.

(2) Interaction topology and scene generalization. Within the realm of social interaction modeling, GNNs, exemplified by Social-STGCNN [165], have emerged as the prevailing paradigm due to their inherent capability to abstract pedestrians as topological nodes and implicitly encapsulate complex, non-Euclidean relational dependencies. Extending this predictive concept to continuous simulation tasks, Ref. [166] proposed an innovative framework advocating the paradigm of "prediction as simulation". This knowledge-guided crowd simulator (KGCS) leverages a hierarchical graph Transformer to model multi-level interaction topologies while innovatively incorporating structural priors from the SFM to regularize the learning trajectory of the neural network. To mitigate the notoriously fragile out-of-distribution (OOD) generalization capabilities typical of purely data-driven formulations, Ref. [167] engineered a transferable pedestrian motion simulation network (TPMSN) tailored for topographic variations, seamlessly transferring behavioral patterns mastered on flat domains to sloped terrains via adversarial or geometric feature alignment. Furthermore, to accommodate the stochasticity and inherent uncertainty governing human decision-making, generative frameworks, characterized by generative adversarial network (GAN) architectures such as Social-GAN [143] or generative transformers [142], have expanded the model output space from rigid, deterministic coordinates to rich multimodal probability distributions. Capitalizing on this generative expressivity to capture macroscopic crowd dynamics, Ref. [168] introduced a data-driven neural crowd simulator (NeCS), which achieves meticulous, fine-grained control over complex environmental parameters while preserving exceptional macro-level statistical resolution. Finally, addressing the ubiquitous challenge of pedestrian simulation constrained by data scarcity, Ref. [169] established a systematic evaluation framework that meticulously benchmarks the performance of diverse data augmentation protocols and transfer learning paradigms in real-world scenarios.

Deep learning frameworks have demonstrated remarkable proficiency in extracting high-dimensional implicit features for short-term spatiotemporal prediction. Nonetheless, the pure deep learning paradigm inherently lacks rigorous physical constraints—such as strict, non-overlapping spatial exclusion primitives—within its foundational mathematical architecture. Consequently, when deployed as continuous simulators within ultra-high-density bottlenecks or completely unobserved edge-case scenarios, these purely data-driven models are prone to generating severe collision penetrations that directly violate fundamental physical principles, as well as triggering catastrophic cascading error amplification through autoregressive rolling rollouts.

5.2.3 *Pedestrian navigation and motion simulation based on reinforcement learning*

Distinct from supervised learning paradigms that directly approximate empirical trajectory distributions, reinforcement learning (RL) strictly formalizes pedestrian state transitions as a Markov decision process (MDP), thereby treating each individual pedestrian as an autonomous, utility-maximizing decision agent [170]. The core algorithmic mechanism centers on constructing a policy network $\pi_\theta(a \mid s)$ that, through continuous closed-loop interactions between the agent and the synthetic simulation environment alongside a balanced exploration-exploitation mechanism, optimizes its navigation and collision avoidance trajectories guided by cumulative reward signals. The canonical modeling architecture explicitly structures an expressive state space (encompassing relative destination coordinates, instantaneous velocities, and neighbors' kinematic features), an action space (comprising continuous velocity vectors or directional heading increments), and a multi-objective reward function that balances positive incentives for target attainment with severe mathematical penalties for physical collisions.

The intrinsic advantage of the RL paradigm lies in its independence from large-scale, paired input-output spatiotemporal trajectory annotations; instead, it leverages highly flexible reward shaping mechanisms to autonomously drive the emergence of diverse crowd behaviors. However, when deployed within complex topological layouts, canon-

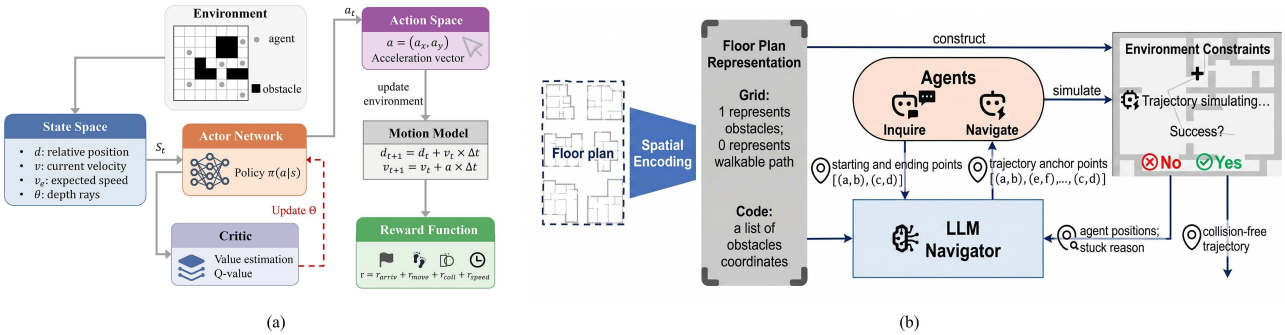

Figure 12 (Color online) Pedestrian modeling and navigation simulation methods via reinforcement learning and large language models. (a) Hybrid architecture integrating DRL with motion guidance map [171]; (b) navigating motion agents in dynamic and cluttered environments through LLM reasoning [22].

ical RL formulations frequently suffer from prohibitively high exploration dimensions and catastrophic exploratory costs stemming from repetitive trial-and-error attempts. To circumvent this exploratory bottleneck, Ref. [171] proposed a hybrid architecture integrating deep RL with a precomputed motion guidance map, as depicted in Figure 12(a), thereby significantly constraining the active exploration space of the agent. Simultaneously, Pu et al. pioneered a Mamba-driven deep deterministic policy gradient framework, designated as Mamba DDPG, to model fine-grained collision avoidance dynamics in heterogeneous vehicle-pedestrian mixed traffic interactions [172]. To address the pervasive homogenization bias, which represents a notorious limitation where traditional multi-agent reinforcement learning setups converge to uniform behavioral patterns, Ref. [173] engineered a parametric RL framework. This approach enables a single, monolithic policy network to synthesize heterogeneous pedestrian populations with rich behavioral diversity via conditional parameter inputs.

Furthermore, the OOD generalization of learned policies remains a critical bottleneck for the RL paradigm. To reinforce policy robustness, recent methodologies have augmented the diversity of agent behavioral distributions by maximizing explicit information-theoretic objectives, thereby implicitly encapsulating a broader spectrum of complex multi-agent interaction topologies [174]. Concurrently, to eliminate the subjective biases and manual tuning intrinsic to hand-crafted reward engineering, generative adversarial imitation learning (GAIL) has been systematically introduced into the field. For instance, Ref. [175] utilized empirical crowd trajectories directly as expert demonstrations, training agents to autonomously recover latent reward surfaces and imitate the empirical behavior distributions of real pedestrians, thereby synthesizing collective crowd dynamics that precisely mirror empirical observations.

5.2.4 *Crowd dynamics simulation based on generative artificial intelligence*

Generative artificial intelligence paradigms fundamentally reframe pedestrian trajectory modeling as a joint optimization task of complex data distribution learning and deep semantic reasoning. In this advanced framework, the underlying simulation mechanism shifts from traditional deterministic coordinate regression to continuous sampling from implicit target distributions and the extraction of pretrained world knowledge, thereby fostering highly realistic, diverse, and cognitively consistent crowd dynamics. Pioneering this intersection, Minartz et al. [11] deployed deep generative multi-agent architectures to systematically deconstruct crowd interaction mechanisms. Their work mathematically illuminates the causal mapping between individual micro-interactions and macro-scale collective emergence, establishing the inaugural framework for the interpretable modeling of crowd behavior via deep generative agents.

GANs rest upon a minimax game between a generator, which maps a latent noise vector, conditioned on historical trajectories and environmental context, onto future coordinate sequences, and a discriminator that separates generated samples from empirical ground truth and returns adversarial gradients. This objective drives the generator to capture the high-dimensional joint distribution of observed movements, at the cost of an adversarial optimization known to be sensitive to the training balance between the two networks. To structurally refine this game-theoretic mechanism, Ref. [176] engineered an incubator-expander generative adversarial network (IE-GAN). Within this split architecture, the incubator module encodes the global motion topologies and macroscopic routing trends, while the expander module refines localized, high-frequency temporal features, achieving exceptional fidelity across both macroscopic direction selection and microscopic stride mechanics. To further enrich multi-agent social interaction representations, Yan et al. [177] integrated a Transformer self-attention mechanism into the adversarial framework; this formulation enables the generator to adaptively map complex, non-linear spatiotemporal dependen-

cies between a central pedestrian and their topological neighbors via multi-head attention operations. Regarding algorithmic task applicability, frameworks rooted in spatiotemporal adversarial learning have proven uniquely effective for synthesizing highly stylized crowd formation transformations [178]: by learning representations directly from empirical distributions annotated with specific styles, the generative network autonomously synthesizes crowd trajectories that adhere to prescribed evolution logics, such as progressive spatial diffusion or rotational contraction. Expanding beyond the adversarial paradigm, Zhao et al. [179] introduced EnvSocial-Diff, a conditional crowd simulation framework powered by diffusion models; by deeply integrating complex environmental boundary conditions with interaction modeling from the individual to the group level, it samples diverse, high-fidelity crowd trajectories from continuous noise distributions.

The infusion of LLMs into pedestrian simulation introduces a novel paradigm rooted in high-level semantic understanding and symbolic cognition. Distinct from traditional data-driven methodologies that numerically optimize spatiotemporal coordinates via regression, LLM-based simulation capitalizes on the extensive world knowledge and generative capabilities encapsulated within pretrained foundational architectures. This paradigm scales from micro-level behavioral fidelity to macro-level social realism. At the individual cognitive level, works like Park et al. [180] demonstrated the feasibility of equipping agents with long-term memory and believable behavioral architectures, which has been successfully adapted to multi-agent path planning, such as the LLMDR framework [181]. Functionally, complex scene topologies and navigation goals are formalized into structured prompt templates [182]; the model then executes step-by-step rationalization to output symbolic decision instructions, which are mapped onto continuous motion control signals. Investigating this spatial deduction boundary, Zhao et al. [22] strategically introduced geometric anchor points to facilitate zero-shot, collision-free path planning in entirely unobserved routing environments, which is illustrated in Figure 12(b). Beyond reactive navigation, recent advances have heavily gravitated toward prompt-engineering-driven crowd synthesis and heterogeneous behavior generation. For instance, Ji et al. [183] leveraged LLMs to parse and canonicalize natural language text scripts into structured behavioral constraints, guiding latent diffusion models to synthesize complex, multi-group crowd distribution fields. Beyond symbolic decision-making and script-driven synthesis, language models have also been applied directly at the movement layer, recasting trajectory prediction as a prompt-based task [21].

Crucially, to bridge the semantic gap between textual prompts and geometric spaces, the frontier of crowd simulation is rapidly shifting toward vision-language models (VLMs) for zero-shot behavioral modeling. Recent methodologies integrate pretrained VLMs to bypass the limitations of text-only descriptions by directly ingesting raw visual scene layouts and egocentric views [184, 185]. By executing zero-shot spatial reasoning on visual interaction topologies—such as recognizing implicit social groups [186], decoding body gestures [187], and anticipating spatial bottlenecks without explicit geometric annotations—VLM-driven simulators map visual perceptual cues directly to agent locomotion policies [188].

In summary, while generative artificial intelligence paradigms significantly elevate the behavioral realism of crowd simulation, they exhibit distinct operational trade-offs across the cognitive-kinematic spectrum. Adversarial and diffusion architectures effectively capture the stochastic multimodality and continuous distributions of empirical locomotion, yet they remain fundamentally constrained by the representation boundaries of their training data and struggle to generalize to severe OOD scenarios. Conversely, the integration of LLMs and VLMs introduces cognitive-level semantic reasoning and zero-shot spatial-visual deduction, gracefully bypassing traditional neural networks' generalization bottlenecks in unobserved environments and unlocking script-driven synthesis.

However, a pronounced semantic-to-control gap persists within these vision-language-foundation frameworks: while exceptionally proficient at high-level behavioral planning and macroscopic topological routing, foundation models remain empirically ill-suited to executing millisecond-level, high-frequency continuous collision avoidance at the micro-kinematic layer, and offer no strict guarantee of hard physical boundaries. This shortfall manifests along two axes. At the kinematic layer, the autoregressive inference of LLM/VLM agents entails per-decision latencies on the order of hundreds of milliseconds to seconds, orders of magnitude larger than the 10–100 ms integration steps at which microscopic collision avoidance must be continuously re-resolved, whereas analytic solvers such as ORCA resolve pairwise geometric constraints in real time at every step, even for crowds of thousands of agents [66]. At the physical-boundary layer, generative models sample trajectories probabilistically and provide no analytic projection operator for enforcing non-penetration; dedicated audits of video-generation-based crowd simulators report persistent violations of pedestrian-level physical realism and agent-level integrity [189], and diffusion-based crowd generators require mechanistic priors or physics-informed guidance to be injected into the sampling process before admissible trajectories are obtained [190,191], whereas LLM-based path planning in unobserved environments yields geometrically inconsistent routes that require externally injected geometric anchor points to restore collision-free geometry [22]. Consequently, constructing a truly high-fidelity, resilient, and physically consistent pedestrian simulation engine theoretically necessitates a structured convergence: a holistic integration that bridges the implicit

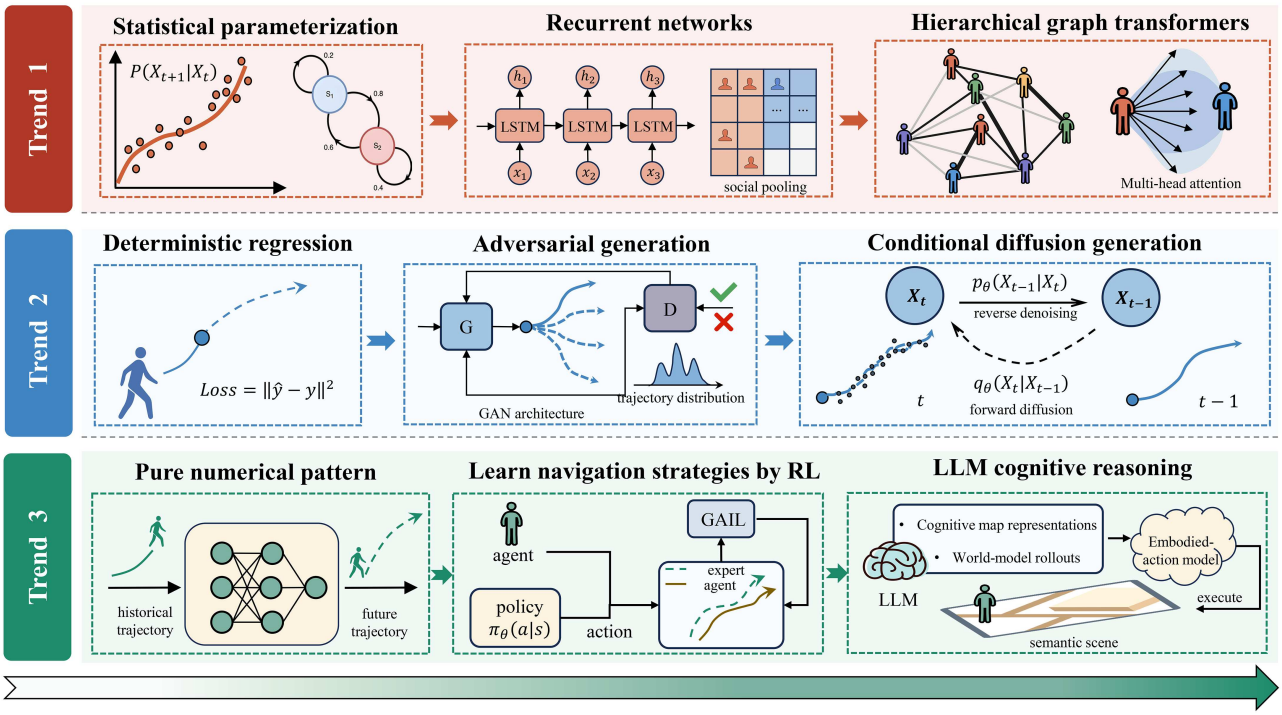

Figure 13 (Color online) Development trends of data-driven pedestrian movement modeling and simulation.

distribution generation, multimodal perception, and cognitive reasoning of generative AI with the explicit physical boundaries, kinematic invariants, and deterministic constraints inherent in mechanism-driven formulations.

5.3 Development trends

Synthesizing the evolutionary trajectories of the four data-driven model families, three convergent development trends emerge that collectively delineate the paradigm's progression from statistical foundations toward an intelligent, cognitive, and generative frontier (Figure 13).

Trend 1. From explicit statistical parameterization to implicit neural representation learning. The data-driven paradigm originated in statistical learning methods that preserved analytically tractable governing equations—such as Langevin stochastic differential equations [150] and Markov transition matrices [151, 152]—while anchoring parameter estimation to empirical data distributions. Although highly interpretable, this foundational phase was fundamentally bounded by its structural, a priori assumptions: Markovian memorylessness precluded long-sequence temporal dependencies, and fixed parametric forms failed to capture the high-dimensional, non-linear interaction topologies inherent in multi-agent crowds [157]. The advent of deep learning architectures—evolving from recurrent networks with social pooling (e.g., Social-LSTM [18]) to spatio-temporal graph neural networks [165] and hierarchical graph Transformers [166]—constitutes a profound paradigm shift. Under this new regime, explicit probabilistic formulations have given way to implicit, end-to-end neural feature extraction. Complex behavioral patterns—such as velocity preferences, group-following dynamics, and non-Euclidean social dependencies—are no longer handcrafted by domain experts; instead, they are autonomously encoded within the latent parameter spaces of deep networks. This transition marks a fundamental methodological breakthrough, unlocking unprecedented non-linear fitting capabilities at the cost of analytical transparency.

Trend 2. From deterministic short-term prediction to multimodal distribution generation. Early deep learning data-driven models were largely confined to deterministic coordinate regression (e.g., Parallelized CNN [162]), mapping historical observations to a single, most-likely future trajectory. While sufficient for short-horizon forecasting under rigid multi-agent coordination, this deterministic approach systematically fails to capture the inherent stochasticity of human decision-making, where a pedestrian facing identical environmental conditions may choose radically different, bifurcated paths.

To address this, GANs, pioneered by Social-GAN [143], fundamentally restructured the output space from rigid coordinate sequences to rich, multimodal probability distributions. This generative paradigm has been successively deepened to handle complex simulation demands: IE-GAN [176] decoupled macro-scale routing topologies from

micro-scale stride mechanics to achieve multi-granularity fidelity, while specialized adversarial frameworks synthesized stylized crowds governed by prescribed evolution logics [178]. More recently, conditional diffusion models (e.g., EnvSocial-Diff [179]) have introduced a theoretically grounded framework to sample high-fidelity crowd trajectories from continuous noise distributions, conditioning the sampler on environmental boundaries and group interaction patterns. This entire trajectory reflects a fundamental reorientation toward generating what could happen rather than simply predicting what will happen, transforming data-driven models from short-term forecasters into stochastic simulators capable of exploring the full behavioral solution space.

Trend 3. From pure data-driven pattern fitting to semantically cognizant generation. The contemporary evolutionary wave transcends purely numerical pattern matching by integrating high-level semantic reasoning and cognitive spatial understanding. RL frameworks pioneered this direction, replacing paired trajectory annotations with flexible reward-driven policies [171, 175] that enable agents to autonomously learn navigation and collision-avoidance strategies through closed-loop environmental feedback. However, traditional RL policies remain confined to low-level kinematic optimization and suffer from behavior homogenization biases [173] and OOD generalization deficits [174]. The recent infusion of LLMs and vision-language-action (VLA) agents represents a qualitative leap. By leveraging vast, pretrained world knowledge, LLM-based planning (e.g., LLMDR [181]) and zero-shot navigation systems [22] execute semantic-level spatial reasoning and chain-of-thought path planning, bypassing the generalization bottlenecks that conventional deep neural networks encounter in unobserved routing scenarios. Crucially, the frontier is shifting toward multimodal VLA control loops [184, 188] and script-driven latent diffusion frameworks [183], which bridge the semantic gap by directly mapping raw visual layouts, egocentric views, or natural language scripts into autonomous, context-aware locomotion policies.

5.4 Comparison of advantages and limitations of various data-driven models

An epistemological and cross-paradigm synthesis of statistical learning, traditional deep learning, reinforcement learning, and generative artificial intelligence reveals fundamental structural trade-offs across data dependency, physical consistency, and generalization capabilities. Physical consistency is rated high ($\surd$) when a physical law constrains the admissible motion by construction; moderate ($\triangle$) when physical or geometric reasoning enters the formulation without constraining the admissible states, for instance as a collision penalty, a behavioural prior, a geometric constraint imposed on individual decisions, or physics-motivated stochastic dynamics; and low ($\times$) when the mapping from input to motion is purely statistical, so that physical structure enters only as conditioning on data. Generalization is rated high ($\surd$) when zero-shot operation, or transfer to an unseen dataset or scene family, is demonstrated; moderate ($\triangle$) when evaluation extends beyond the setting used for fitting, for example a scene-wise hold-out across a common benchmark or augmented and degraded conditions; and low ($\times$) when evaluation is confined to the scenario or collection used for fitting. Computational efficiency is rated high ($\surd$) when inference is near-analytic or a single lightweight forward pass and runs in real time; moderate ($\triangle$) when a single GPU forward pass or a moderate iterative overhead, such as policy rollout, is required; and low ($\times$) when inference requires multi-step generation, autoregressive decoding, or repeated sampling.

The historical transition from statistical learning to connectionist deep learning highlights a stark contradiction between algorithmic interpretability and feature representational capacity. As synthesized in Table 3, statistical learning methods—such as the extended Langevin equation [150] and neural spatiotemporal point processes [151,152,156]—rely on explicit probability distributions to preserve analytically tractable governing equations, rendering them exceptionally proficient at extracting macro-scale statistical regularities with remarkably low training data dependency. However, their restrictive a priori probabilistic assumptions limit their capacity to capture localized, high-dimensional multi-agent coordination. Conversely, traditional deep learning architectures dispense with rigid theoretical presuppositions. Leveraging their non-linear mapping proficiencies, connectionist paradigms like Parallelized CNN [162], Social-LSTM [18], and Social-STGCNN [165] attain the highest reported accuracy in short-term trajectory forecasting on the standard benchmarks. Yet, this representational leap comes at a severe epistemological cost: these models operate as black boxes that obscure the causal socio-psychological mechanisms driving individual locomotion.

Beyond the interpretability dilemma, a critical architectural vulnerability shared by purely data-driven deep learning and generative formulations is the structural absence of foundational physical boundaries, directly impacting their physical consistency and OOD generalization. Because these connectionist networks inherently lack hard geometric constraints—such as strict, non-overlapping spatial exclusion primitives—at their mathematical core, their deployment over extended temporal rollouts leads to cascading error amplification, a failure mode documented in the studies that report long-horizon evaluation. In high-density bottlenecks or completely unobserved edge-case scenarios, this deficit manifests as non-physical coordinate overlaps and interpenetration wherever these are mea-

Table 3 Comprehensive comparison of various data-driven models. Notes: $\surd$ = high/yes, $\times$ = low/no, $\triangle$ = moderate/partial. SL: statistical learning; DL: deep learning; RL: reinforcement learning; GenAI: generative artificial intelligence.

Method	Class	Underlying mechanism	Data dependency	Physical consistency	Generalization	Computational efficiency	Applicable tasks
Extended Langevin equation [150]	SL	Explicit probability distribution; data fitting	Continuous trajectory coordinates	$\triangle$	$\times$	$\surd$	Physical dynamics deduction with clear macroscopic statistical laws
Neural spatiotemporal point process [156]	SL	History-conditioned event sequences	Discrete events	$\times$	$\times$	$\surd$	Periodic simulation of macroscopic spatiotemporal flow
Hierarchical two-layer locomotion [157]	SL	Macro-flow; micro-movement coupling	Multi-scale trajectory statistics	$\triangle$	$\times$	$\surd$	Cross-scale coupling simulation enhancing macro–micro consistency
Social-LSTM [18]	DL	RNN memory gating; social pooling	Paired spatiotemporal trajectories	$\times$	$\times$	$\triangle$	Short-term prediction with implicit social interaction encoding
Parallelized CNN framework [162]	DL	Feature extraction; nonlinear regression	Paired spatiotemporal trajectories	$\times$	$\times$	$\triangle$	Short-term prediction of standard pedestrian flow with fixed topology
Hierarchical SP–MD framework [163]	DL	Scene perception; motion dynamics	Paired turning-scene trajectories	$\times$	$\times$	$\triangle$	High-fidelity right-angle turning kinematics simulation
Visual-info cross-scene model [164]	DL	Visual feature-driven adaptation	Multi-scene trajectory pairs	$\times$	$\triangle$	$\triangle$	Cross-scene generalizable trajectory prediction
Social-STGCNN [165]	DL	Graph topology encoding; spatiotemporal convolution	Paired spatiotemporal trajectories	$\times$	$\times$	$\triangle$	Interaction-topology-aware trajectory prediction
KGCS (prediction-as-simulation) [166]	DL	Hierarchical graph Transformer; SFM prior	Paired trajectories; mechanism constraints	$\triangle$	$\triangle$	$\triangle$	Multi-level interaction simulation with mechanistic regularization
TPMSN transfer network [167]	DL	Feature alignment; spatial mapping	Source and target domain trajectories	$\times$	$\triangle$	$\triangle$	Feature transfer modeling when lacking specific scenario (e.g., slopes) data
Social-GAN [143]	DL	Adversarial game; multimodal sampling	Paired trajectories; noise vectors	$\times$	$\triangle$	$\triangle$	Multimodal trajectory generation with behavioral diversity
Data augmentation & Transfer Evaluation [169]	DL	Benchmarking of augmentation protocols	Scarce-data trajectories	$\times$	$\triangle$	$\triangle$	Robust pedestrian simulation under data scarcity
Pre-trained map-guided RL [171]	RL	Markov decision; reward-driven policy	Simulation feedback	$\triangle$	$\triangle$	$\triangle$	Autonomous navigation under unknown complex topology
Mamba-DDPG [172]	RL	Deterministic policy gradient; Mamba model	Mixed traffic interaction trajectories	$\triangle$	$\times$	$\triangle$	Fine-grained collision avoidance in vehicle–pedestrian mixed traffic
Parametric Heterogeneous RL [173]	RL	Conditional policy; parametric control	Simulation feedback	$\triangle$	$\triangle$	$\triangle$	Synthesis of heterogeneous pedestrian populations with behavioral diversity
Info-theoretic diversity RL [174]	RL	Mutual information maximization; interaction coverage	Simulation feedback	$\triangle$	$\triangle$	$\triangle$	Policy robustness enhancement against OOD behavioral distributions
GAIL adversarial imitation [175]	RL	Reward extraction; distribution imitation	Expert trajectories	$\triangle$	$\triangle$	$\triangle$	High-fidelity multi-agent interaction
LLMDR LLM-driven planning [181]	RL	LLM long-term memory; MDP path planning	Natural language descriptions	$\triangle$	$\surd$	$\times$	Zero-shot multi-agent path planning with believable behavioral architectures
Deep generative multi-agent [11]	GenAI	Micro-macro causal mapping	Large-scale crowd trajectories	$\times$	$\triangle$	$\triangle$	Interpretable crowd behavior modeling via deep generative agents
Transformer-GAN adversarial [177]	GenAI	Attention mapping; adversarial game	Paired trajectories; noise vectors	$\times$	$\triangle$	$\triangle$	Multi-agent social interaction generation with adaptive attention
Stylized crowd formation GAN [178]	GenAI	Style distribution learning; evolution constraints	Empirical formation trajectories	$\times$	$\triangle$	$\triangle$	Stylized crowd formation synthesis with prescribed evolution logics
EnvSocial-Diff [179]	GenAI	Conditional diffusion; environmental boundary; group interaction	Diverse trajectories; scene layouts	$\times$	$\triangle$	$\times$	High-fidelity diverse crowd trajectory sampling with environmental constraints
Zero-shot LLM navigation [22]	GenAI	Pre-trained knowledge; semantic reasoning	Natural language prompts	$\triangle$	$\surd$	$\times$	Zero-shot navigation in unobserved routing environments
LLM-guided latent diffusion [183]	GenAI	Script parsing; behavioral constraints; latent diffusion	Natural language scripts; crowd distributions	$\times$	$\surd$	$\times$	Script-driven multi-group crowd distribution synthesis
VLA-driven crowd simulation [184, 188]	GenAI	VLA agent; visual scene reasoning	Visual layouts; egocentric views	$\triangle$	$\surd$	$\times$	Zero-shot behavioral simulation with visual perceptual–locomotion mapping

sured. Although feature-aligned transfer networks like TPMSN [167] and robust data augmentation protocols [169] can superficially mitigate domain gaps within the latent feature manifold, they do not establish physically grounded safety guarantees, because the mechanism remains external to the learned computation. A notable exception within this spectrum is the integration of mechanistic regularizations, exemplified by the KGCS framework [166], which embeds structural priors from the social force model to achieve moderate physical consistency and localized generalization, pointing to a critical open question regarding how to natively hardcode non-smooth geometric invariants directly into differentiable computation graphs.

To decouple simulation architectures from an absolute dependence on massive, rigidly paired spatiotemporal trajectory benchmarks, RL reformulates multi-agent dynamics into a MDP framework. By substituting explicit trajectory regression with interactive simulation environment feedback, the algorithmic applicability of RL—spanning value-based coordination [171], policy gradient navigation [172], and multi-agent frameworks [173,174]—is uniquely optimized for synthesizing autonomous navigation paths and heterogeneous collision-avoidance strategies across topologically complex environments. Nevertheless, a profound contradiction within the RL paradigm resides in

the reward engineering paradox, where hand-crafted reward shaping essentially reintroduces human prior heuristic biases under a concealed mathematical guise [192]. When deployed within extreme localized crowding bottlenecks, relying exclusively on these soft scalar penalties frequently fails to enforce strict hard boundaries of physical non-penetration, which easily precipitates critical policy oscillations, micro-kinematic jitters, or unnatural, frozen localized deadlocks [171, 181]. While imitation-learning frameworks like GAIL [175] utilize expert demonstrations to autonomously recover latent reward surfaces and eliminate subjective bias, enforcing zero-tolerance safety guarantees within pure policy learning remains an unresolved challenge.

At the contemporary frontier, generative artificial intelligence (GenAI) models span a wide spectrum from distribution-centric learning to semantic-centric cognition, fundamentally restructuring the trade-offs between physical boundaries and open-ended generalization. Adversarial networks, such as IE-GAN [176] and NeCS [168], alongside conditional crowd diffusion frameworks like EnvSocial-Diff [179], excel at capturing the stochastic, multimodal probability distributions governing human collective locomotion, enabling the synthesis of highly diverse, stylized crowd formations [178] with exceptional macro-level statistical resolution. However, their underlying mechanism is strictly a pattern-fitting task bounded by the representation limits of empirical datasets, yielding low physical consistency when operating outside observed parameter envelopes. Conversely, the infusion of LLM-driven planning (e.g., LLMDR [181]) and multimodal VLA agents [184, 188] exhibits exceptional zero-shot generalization and advanced spatial deduction by extracting pretrained world knowledge from natural language prompts or raw visual scene layouts.

Consequently, constructing a truly high-fidelity, resilient, and physically consistent pedestrian simulation engine theoretically necessitates a structured convergence: a holistic integration that bridges the implicit distribution generation, multimodal perception, and cognitive reasoning of generative AI with the explicit physical boundaries, kinematic invariants, and deterministic constraints inherent in mechanism-driven formulations.

6 Fusion-driven pedestrian movement modeling and simulation

6.1 Methodological framework

Coinciding with the systematic deepening of crowd dynamics theory and the escalating fidelity demands of safety-critical applications, the paradigm shift from isolated data-driven or mechanism-driven approaches toward a deeply coupled data-mechanism fusion framework has emerged as the defining methodological trend. As established in the preceding comprehensive taxonomies, while purely data-driven methods exhibit unparalleled capabilities in capturing high-dimensional implicit interaction topologies from vast repositories of empirical trajectory benchmarks, their fundamental lack of epistemic interpretability and hard physical non-penetration boundaries inherently render them susceptible to generating non-physical, counterintuitive anomalies when confronting data-scarce edge cases or corrupted scenario configurations with high noise. Conversely, purely mechanism-driven formulations, although built upon analytically transparent and mathematically rigorous physical rules, remain strictly bounded by the deductive limits of hand-crafted differential equations, rendering them functionally ill-suited to comprehensively and dynamically adapt to the highly nonlinear, stochastic, and heterogeneous behavioral features manifested in real-world urban crowds.

The fusion paradigm fundamentally aims to dismantle these inherent methodological boundaries, synergizing the structural interpretability of mechanistic priors with the dynamic adaptability of empirical data features through structured co-design architectures. In recent years, catalyzed by the rapid ascendance of scientific machine learning paradigms such as physics-informed machine learning [26], neural ordinary differential equations [27], and differentiable simulations [193], researchers have begun systematically charting fusion pathways that organically embed mechanistic formulations into data-driven learning loops, thereby achieving a holistic synthesis of both paradigms within a unified monolithic framework.

To mathematically formalize and systematically characterize the structural diversity of these fusion methodologies, this review establishes a unified algebraic notation. Let $f(x, \theta)$ denote the mechanism-driven governing formulation (the white-box operator), $DD(x; \phi)$ the data-driven model (the black-box operator), $\Delta(x)$ the operational model discrepancy function, ε the stochastic Gaussian or non-Gaussian white noise term, and x and Y the system input and corresponding empirical output vectors, respectively. Leveraging this notation, existing hybrid architectures can be systematically consolidated, following the physics-guided machine learning taxonomy of Jia et al. [194], into six canonical fusion structures, as conceptualized in Figure 14.

Crucially, within the scope of this review, we establish a rigorous taxonomy criterion for the data-mechanism fusion-driven paradigm: a model qualifies only if its formulation organically synergizes the structural priors of

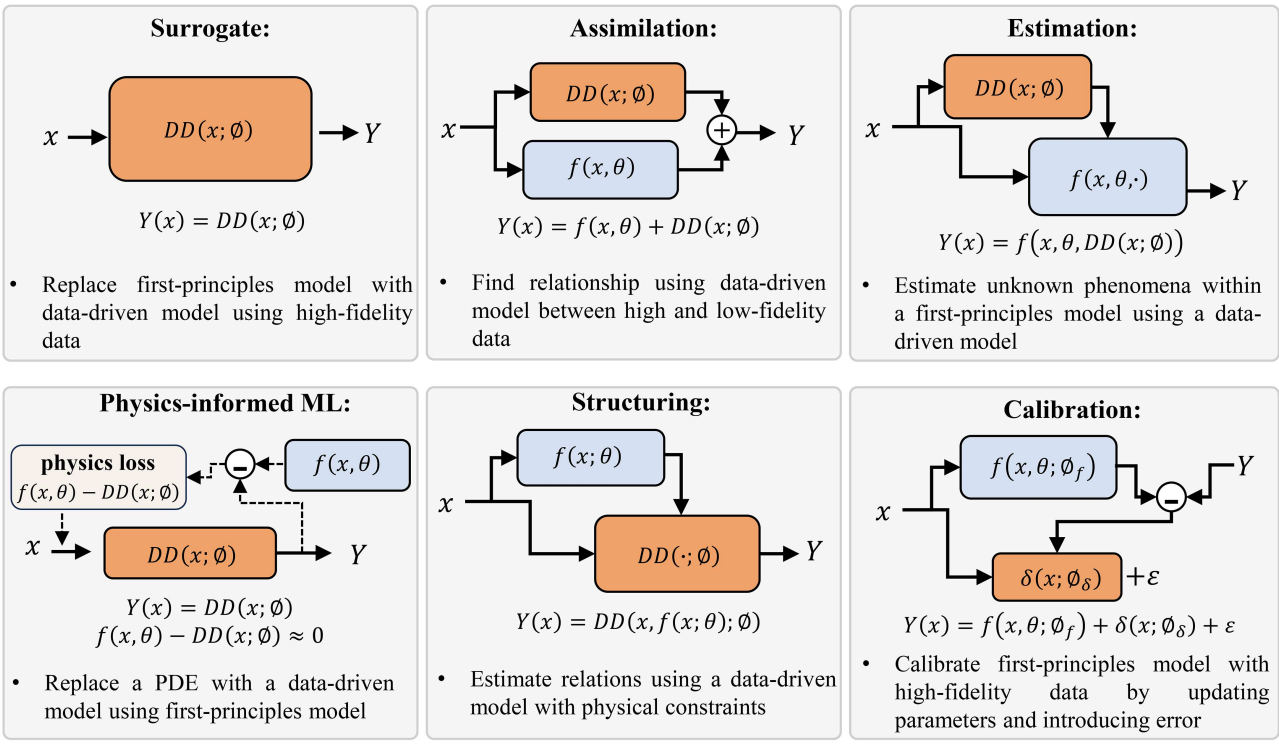

Figure 14 (Color online) Six canonical taxonomy schemes for mechanism-data fusion.

mechanistic physics with the unconstrained representation capabilities of statistical learning. This paradigm is operationalized through two reciprocal pathways: either by systematically embedding physical boundary constraints, kinematic conservation laws, or behavioral heuristics directly into the optimization landscape and architectural topology of a data-driven framework; or conversely, by deploying deep learning operators to parameterize and accelerate the governing differential equations of a mechanistic model. Consequently, rather than relying on a trivial, decoupled engineering pipeline, these hybrid architectures implement a formal information-exchange mechanism. This ensures that their optimization objectives, topological structures, or output fields simultaneously encapsulate explicit, physically grounded constraints and empirical data-driven signals.

6.2 Representative models

Following the taxonomy of the six hybrid structures established above, this subsection reviews representative data-mechanism fusion models. From the perspective of integration methodology, existing work falls into three categories: parameter calibration, differentiable embedding, and PIML. The three families simply group them according to one question: which element of the mechanistic model the learning process is allowed to change. Learning that updates only the parameters of the mechanism, as in the calibration scheme and in assimilation when the learned module merely supplies a correction, leaves the mechanism as the forward engine, so that the mechanism lies outside the learned model; this is parameter calibration. Learning that rewrites a mechanism-derived computation, such as a partial differential equation, a time integrator, or a non-smooth algorithm, into a differentiable operator evaluated inside the network, as in the structuring and surrogate schemes, places the mechanism inside the computation; this is differentiable embedding. Learning that keeps the forward computation learned while physics enters as a loss, a structural invariant or a sampling prior, as in the physics-informed machine learning and estimation schemes, places the mechanism around the computation; this is PIML. The three families therefore differ in exactly one respect, namely whether the mechanism lies outside, inside, or around the forward computation.

6.2.1 *Data-driven learning of mechanism model parameters*

In this architecture, a neural network $DD(x;\phi)$ maps scene features to the parameter space θ of a mechanistic model $f(x,\theta)$, which remains the sole forward engine of the simulation. For instance, Ref. [195] employed a neural network to capture spatiotemporal trajectory dynamics and output adaptive parameters for the SFM; Yan et al. [196] preprocessed potential collision segments using the repulsive force laws of the SFM before training and

embedded a geometric constraint module into the network; and Ref. [197] generated boundary trajectories with the ORCA algorithm to train a network that learns the inverse mapping from macro-statistical trajectory features to micro-simulator parameters. Beyond direct supervised regression, this family also encompasses simulation-in-the-loop calibration and mechanism-generated imitation: Amartaivan and Morita [198] embed a neural surrogate that predicts next-step velocity inside a particle swarm optimization loop, calibrating agent-specific parameters from real-world trajectories without repeatedly invoking the full simulator; Zhou et al. [199] drove a multi-agent adversarial imitation learner with expert demonstrations produced by a group-enabled SFM, so that the mechanistic prior enters through the demonstration signal rather than the parameter vector. Across all these variants, the learning module reshapes only the parameter field and adds no expressiveness beyond what the mechanistic formulation can already represent; its contribution is calibration efficiency, population heterogeneity, and context adaptivity, while hard non-penetration boundaries are preserved by construction.

6.2.2 *Differentiable embedding of mechanistic computation graphs*

Constructive fusion paradigms seek to bridge the dichotomy between white-box mechanistic formulations and black-box neural architectures by translating microscopic kinematic principles or macroscopic fluid dynamics into differentiable computational graphs, so that physical equations act as native operators within deep networks under end-to-end gradient optimization. At the microscopic collision-avoidance level, ORCANet [200] reformulates the inherently non-smooth ORCA algorithm into a differentiable forward layer: it approximates discrete neighbor selection with a Gaussian kernel $K(d_{ij}) = \exp(-d_{ij}^2/2\sigma^2)$ and substitutes linear programming with a differentiable relaxation of the velocity obstacle geometry, thereby enforcing physical constraints during gradient backpropagation.

To bridge the microscopic and macroscopic scales, recent studies leverage Neural ODEs for continuous-time crowd modeling. Social ODE [201] formulates interaction dynamics in continuous time to capture smooth trajectories and enable zero-shot obstacle avoidance, while PI-NeuGODE [202] embeds physical kinematics into graph neural ODEs and applies symbolic regression to extract explicit analytical expressions, jointly resolving the transparency and generalization bottlenecks of pure deep learning.

At the macroscopic scale, STDDN [203] incorporates the fluid-dynamic continuity equation as an inductive bias into a spatiotemporally decoupled differential equation network, governing density evolution with Neural ODEs and mitigating error accumulation in trajectory forecasting; HINN [204] deploys neural layers to solve the Navier-Stokes equations directly, preserving crowd continuity and fluidity while modeling dense collective behaviors. A further route promotes the physical prior from a differentiable operator inside the network to the stochastic process that generates the motion. He et al. [205] modeled high-density crowds as an active-particle continuum whose dynamics follows a neural stochastic differential equation, so that the active-matter structure supplies the physical prior while the drift and diffusion terms are learned; the resulting continuous-time simulator supports long-horizon forecasting with an interpretable latent field and outperforms discrete-time black-box baselines on low-quality videos where individual tracking is infeasible. Chauhan et al. [206] trained Fourier and wavelet neural operators as mesh-free surrogates for the Hughes model of pedestrian flow, but reported a cautionary finding: the learned operators smooth out the shocks and discontinuities that motivate hyperbolic crowd models, showing that replacing a mechanistic solver with a learned surrogate does not automatically preserve its conservation and discontinuity-capturing properties.

Considered together, this family pursues physical compliance through differentiability rather than penalization, and its members are distinguished by what is made differentiable: smooth kernels that relax a non-smooth mechanism [200], continuous-time operators such as neural ODEs and SDEs that suppress compounding error [201, 202, 205], and learned solution operators that replace macroscopic solvers [203, 204, 206]. The shared advantage is that gradients propagate through the mechanistic computation, so physical violations are corrected inside the learning loop rather than penalized from outside; the corresponding cost is that physical fidelity now hinges on the smoothness and numerical conditioning of the relaxation, which explains why soft relaxations trade exact constraint satisfaction for trainability [200] and why learned operators may lose shock structure even when trained on discontinuous samples [206].

6.2.3 *Physics-informed constraints and closed-loop architectures*

The cornerstone of PIML and closed-loop architectures lies in embedding physical principles into deep models as regularized training constraints or structural invariances, balancing empirical accuracy with physical consistency. Figure 15 illustrates three representative frameworks of this family. As shown in Figure 15(a), Zhang et al. [207] iteratively used simulated outputs from a mechanistic model as priors to guide a neural network, then applied

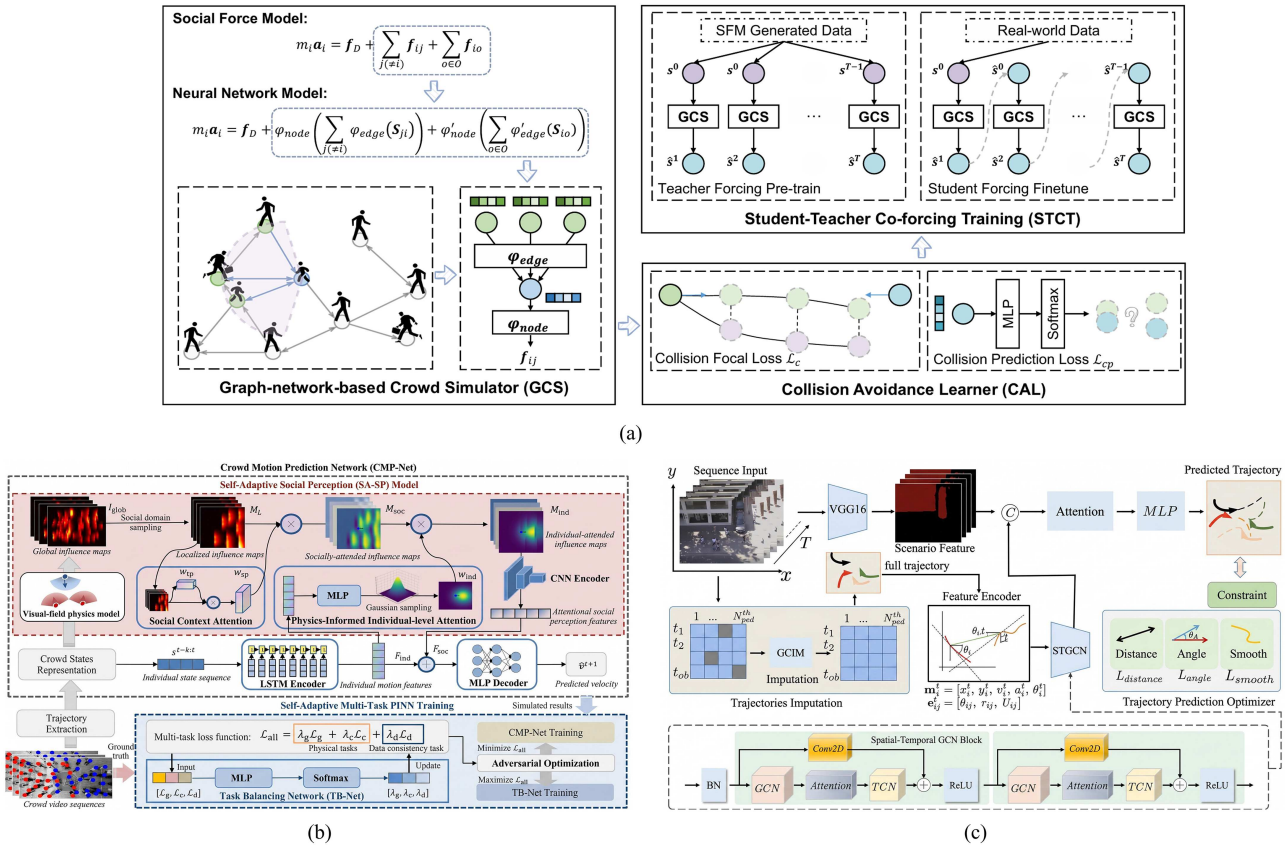

Figure 15 (Color online) Pedestrian trajectory prediction and simulation based on physics-informed machine learning by embedding physical principles into deep models as regularized training constraints. (a) Physics-informed machine learning for crowd simulation [207]; (b) self-adaptive physics-informed neural network (SA-PINN) architecture for crowd simulation [208]; (c) ARP-STGCN framework for pedestrian trajectory prediction [210].

symbolic regression to distill novel analytical equations that describe crowd dynamics with superior precision. SA-PINN [208] introduces a social perception module guided by field-based physics together with dynamic multi-task loss balancing across navigation, collision avoidance, and data fitting, as depicted in Figure 15(b). At the macro-systemic level, PI-STGCN [209] couples a spatiotemporal graph convolutional network with a navigation potential field in a closed loop, and ARP-STGCN [210] embeds attractive and repulsive potential mechanics directly into graph convolutions, enriching social interaction modeling, as illustrated in Figure 15(c). Extending this paradigm to generative modeling, SPDiff [190] conditions a diffusion process on a social-force-inspired, equivariant encoder, steering reverse-diffusion steps with mechanistic structures to preserve physical plausibility and multimodality even in low-resource regimes. Furthermore, RAPID [191] pushes this principle toward industrial deployment by injecting social-force gradients as directional priors and caching map contexts, achieving real-time inference latency below 30 ms for production-grade autonomous-driving simulations.

Within this family, the technical divergence lies in where the physical prior acts: as a dynamically rebalanced multi-task loss [208], a navigation potential field coupled to a graph network [209], attractive and repulsive terms inside convolution operators [210], or the conditioning signal of a diffusion architecture [190]. What unifies these designs is coupling strength rather than implementation: the physical prior shapes the hypothesis space and the sampling distribution but does not dictate the forward dynamics, so PIML retains the statistical flexibility of data-driven learning at the cost of a physical guarantee strictly weaker than parameter calibration or differentiable embedding can offer.

6.3 Development trends

Synthesizing the evolutionary trajectories of the fusion-driven models reviewed above, three convergent trends emerge that collectively delineate the paradigm's progression from superficial coupling toward deep structural integration (Figure 16).

Trend 1. From parameter-level bridging to deep closed-loop co-design. The depth of mechanistic engagement

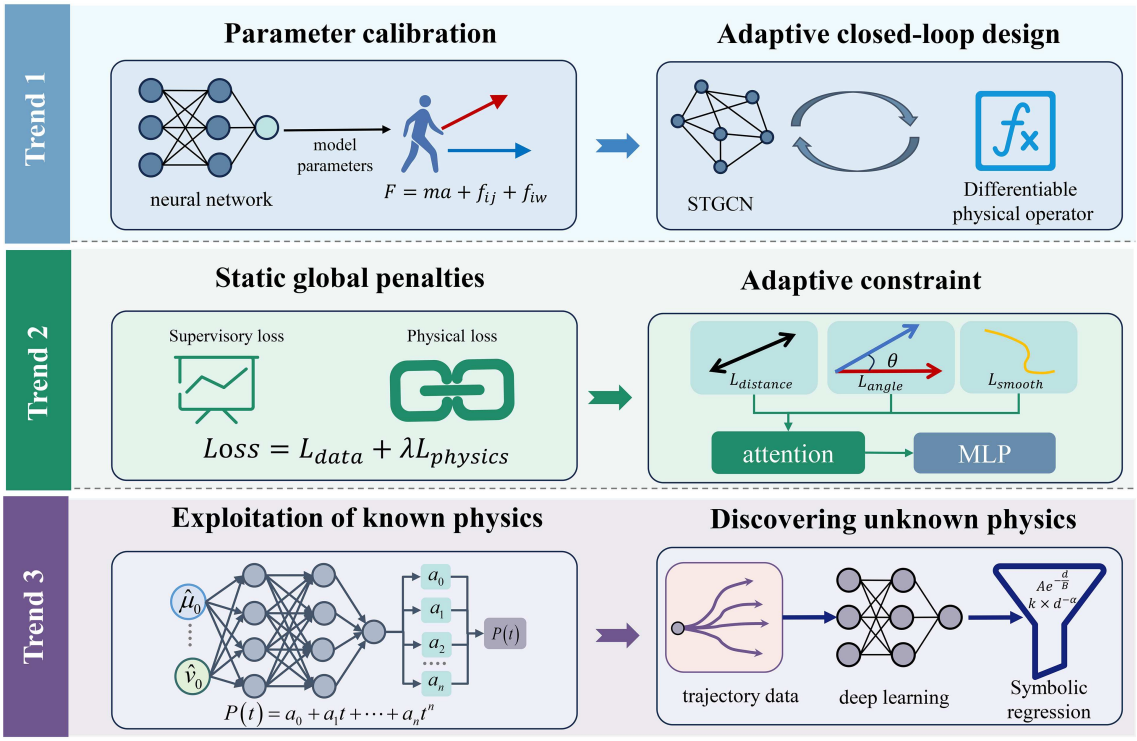

Figure 16 (Color online) Development trends of fusion-driven pedestrian movement modeling and simulation.

within the data-driven learning loop has progressively intensified. Early fusion efforts operated exclusively at the parameter interface, where neural networks learned inverse mappings from trajectory features to mechanistic parameters [195,197] or where the mechanistic model supplied expert demonstrations and calibration targets [198,199], leaving the mechanistic model in control of the forward kinematic evolution. The advent of differentiable computation graphs [200, 203, 204] together with continuous-time physics priors realized as neural ODEs and SDEs [205] and learned solution operators for continuum equations [206] subsequently elevated fusion to the structural level by translating physical mechanisms into natively differentiable operators within deep networks. This architectural shift ensured that physical compliance is rigorously enforced rather than merely encouraged through soft penalties. More recently, advanced architectures such as PI-STGCN [209] and ARP-STGCN [210] have established dynamic, closed-loop co-design, wherein data-driven and mechanistic modules iteratively inform and correct each other through structured, bidirectional information exchange.

Trend 2. From static global penalties to adaptive, context-aware constraint mechanisms. The nature of physics-informed regularization has evolved from rigid, one-size-fits-all soft constraints toward adaptive, behaviorally cognizant mechanisms. To balance conflicting optimization objectives, SA-PINN [208] introduces dynamic optimization that automatically rebalances multi-task loss landscapes during training. This shifts physics-informed regularization from a blunt global instrument into a sharp cognitive filter, capturing the heterogeneous, situational decision-making of real-world crowds. A parallel evolution moves the physical prior from the objective function into the model architecture itself: in physics-conditioned diffusion models [190], the mechanistic interaction structure governs the sampling trajectory of the generative process rather than merely penalizing its outputs, so that physical plausibility is imposed on the distribution of admissible motions rather than on individual predictions.

Trend 3. From exploiting known physics to discovering unknown physics. The most conceptually profound development is the transition from exploiting known physics to discovering unknown physics. Zhang et al. [207] deployed symbolic regression on trained neural networks to distill novel analytical equations that capture crowd dynamics with superior precision compared to original hand-crafted formulations. Consequently, the fusion pipeline is no longer restricted to calibrating or constraining established mechanistic priors; instead, it acquires the capacity to autonomously generate new physical knowledge from raw data. Nevertheless, crystallizing concise analytical expressions remains challenging for highly intricate socio-psychological interaction dynamics.

6.4 Comparison of advantages and limitations of various fusion-driven models

In the evolutionary spectrum of data-mechanism fusion, contemporary literature exhibits a pronounced stratification according to the depth of architectural integration, scaling from parameter-level decoupling to structure-level embedding and ultimately to system-level closed-loop synergy. Within the paradigm of parameter learning and symbolic discovery [195–197,207], the data-driven and mechanistic components maintain high functional autonomy: deep learning or evolutionary computing circumvents high-dimensional calibration overheads, fragile generalization, and empirical scarcity in edge cases by learning the inverse mapping from macro-trajectory statistics to micro-parameters [197]. While this decoupled architecture preserves the hard non-penetration constraint of the baseline formulation, because the mechanism remains the forward engine, its expressive capacity is inherently bounded by the rigid structure of the baseline differential equations. Although pioneering frameworks such as Zhang et al. [207] and Mo et al. [202] applied symbolic regression to discover novel analytical equations from data, crystallizing concise and broadly applicable physical formulations remains a formidable challenge under highly stochastic socio-psychological interaction dynamics.

To dismantle these structural boundaries, differentiable embedding and neural solving methods combine the mechanism and the network at the level of the computation graph, mapping physical laws into functional operators. Through fully differentiable simulation environments [193] or continuous temporal interaction tracking with Neural ODEs [201, 202], these approaches co-design mechanistic structures and neural networks end to end. By incorporating macroscopic conservation laws or leveraging neural operators to solve partial differential equations [203, 204], they suppress chronic error accumulation in long-term predictions and report improved fidelity in high-density crowds. Conversely, the PIML paradigm introduces a more flexible coupling pathway that employs physics-guided soft constraints to regularize neural network training [208–210]. By coupling social force potential fields with spatiotemporal graph convolutions [210] or using multi-task loss regularization [208], PIML architectures preserve high statistical flexibility while preventing data-driven trajectories from violating core physical priors such as kinematic plausibility and directional consistency.

The fundamental divergence among these fusion strategies lies in the specific object that the learning module is permitted to modify, a distinction that dictates their respective technical trade-offs and data dependencies. In parameter-level fusion, learning is confined to calibrating the parameter vector of a fixed equation system (e.g., social force models) [195–199]. This approach inherits the hard boundaries of the mechanistic formulation, ensuring strict physical consistency and viability in data-scarce regimes, though it remains limited by the structural expressiveness of the chosen backbone. In contrast, differentiable embedding modifies the operator itself—including its smoothness and temporal discretization—allowing frameworks like ORCA to be relaxed into differentiable forms [200–202, 205, 206]. While this extends expressiveness beyond fixed equations, it replaces strict constraints with relaxed ones, necessitating substantial high-fidelity datasets to prevent overfitting. Finally, PIML approaches (e.g., neural PDE solvers) target the loss landscape or sampling distribution [190, 208–210], offering the greatest flexibility but enforcing physical consistency only in a statistical sense. Consequently, these end-to-end surrogates are heavily data-dependent and prone to failure on out-of-distribution events [203, 204, 206]. This hierarchy explains why the choice of mechanistic backbone is not arbitrary: interpretable force-based equations suit parameter calibration, whereas velocity-reciprocal frameworks or continuum conservation laws are better aligned with differentiable relaxation and neural solving, respectively [200, 203, 204, 206].

For Table 4, the physical constraint column records whether the mechanism remains the forward engine (hard), is relaxed into a differentiable or soft form, or enters only as a prior. Computational efficiency is rated high ($\surd$) when the mechanism remains the forward engine and inference is near-analytic, moderate ($\triangle$) when a differentiable solver, neural integrator, or surrogate-optimization loop is required, and low ($\times$) when multi-step generation or repeated numerical solving dominates the per-step cost.

7 Challenges and future directions

Although the data-mechanism fusion paradigm has demonstrated remarkable potential in crowd movement modeling, existing methodologies still grapple with fundamental theoretical and engineering bottlenecks across critical dimensions, including model generalizability, guaranteed physical consistency, computational efficiency, and deployment reliability. A unified framework capable of concurrently satisfying the stringent requirements of high accuracy, rigorous interpretability, low latency, and safety-critical applicability remains elusive. To chart a path forward, this section systematically delineates the core open challenges along four dimensions, namely mechanistic understanding, fusion coupling mechanisms, computational constraints, and evaluation frameworks, and correspondingly formulates

Table 4 Comprehensive comparison of various fusion-driven models. Notes: Hard Constraint = strict adherence to physical formulations. Soft constraint = physics-guided regularization with data flexibility. Prior = physics enters as a structural or sampling prior. For computational efficiency: $\surd$ = high/yes, $\triangle$ = moderate/partial, $\times$ = low/no.

Class	Ref.	Fusion architecture form	Physical constraint attribute	Computational efficiency	Applicable tasks
Parameter learning and symbolic discovery	[195]	Serial mapping and parameter optimization	Hard constraint	$\surd$	Macro pedestrian flow simulation via static physical equations
	[196]	Data preprocessing and inductive bias	Hard constraint	$\surd$	Microscopic pedestrian trajectory prognosis
	[197]	Inverse mapping from macro-features to micro-parameters	Hard constraint	$\surd$	Micro-simulator calibration under data scarcity in edge cases
	[207]	Bidirectional knowledge extraction; equation reconstruction	Prior guidance	$\triangle$	Physical law discovery and complex crowd dynamics modeling
	[198]	Surrogate-in-the-loop swarm optimization of agent parameters	Hard constraint	$\triangle$	Heterogeneous agent-based pedestrian model calibration
	[199]	Mechanism-generated demonstrations for adversarial imitation	Hard constraint	$\triangle$	Group-aware pedestrian trajectory prediction and evacuation simulation
Differentiable embedding & neural solving	[193]	End-to-end gradient optimization via differentiable simulation	Hard constraint	$\triangle$	Holistic multi-paradigm synthesis and monolithic model co-design
	[200]	Neural operator formulation of computational graphs	Soft constraint	$\triangle$	Pedestrian collision avoidance control
	[201]	Continuous temporal modeling via neural ODEs	Hard constraint	$\triangle$	Continuous dynamic interaction tracking; zero-shot obstacle avoidance
	[202]	Kinematic graph neural ODEs; symbolic extraction	Hard constraint	$\triangle$	Analytical model discovery; explainable crowds generalization
	[203]	Macroscopic equation-constrained micro deep networks	Hard constraint	$\triangle$	Macro-crowd flow estimation
	[204]	Neural solving of partial differential equations	Hard constraint	$\times$	Ultra-high-density crowd pattern synthesis
	[205]	Neural stochastic differential equations with active-matter prior	Prior guidance	$\triangle$	Extremely high-density crowd forecasting, analysis and simulation
Physics-informed machine learning	[208]	Dynamic multi-task loss function regularization	Soft constraint	$\triangle$	Multi-objective motion simulation and coordination
	[209]	Potential field models driven by PINNs	Soft constraint	$\triangle$	Crowd simulation under global physical and fidelity constraints
	[210]	Attractive-repulsive potential spatiotemporal graph convolutions	Soft constraint	$\triangle$	Intricate social interaction modeling under structured potential fields
	[190]	Social-force-guided equivariant denoising with multi-frame rollout	Soft constraint	$\times$	Multimodal crowd simulation with long-horizon physical consistency
	[191]	Social-force gradient priors in a cached-latent denoising sampler	Soft constraint	$\triangle$	Real-time pedestrian trajectory generation for industrial simulation

concrete research questions for each.

7.1 Challenges

(1) **Insufficient mechanistic depth: pedestrian movement is intrinsically multidisciplinary.** Pedestrian movement is not a purely physical process but emerges from the coupled operation of perceptual, cognitive, psychological, and social mechanisms over a physical locomotion substrate. Controlled experiments have shown that pedestrian navigation rests on visual heuristics and social-cognitive rules rather than on repulsive force fields [211]; the walking configuration of pedestrian social groups shapes crowd flow in ways that no pairwise physical interaction can reproduce [212]; and psychological analyses of crowds argue explicitly that existing models under-represent the cognitive and social-identity mechanisms that actually drive collective behavior [213]. Yet the mechanism-driven paradigm remains overwhelmingly phenomenological: social-force, velocity-obstacle, and cellular-automata formulations compress this multidisciplinary substrate into a handful of surface forces or discrete transition rules, whose parameters are best interpreted as proxies for unmodeled cognitive states rather than grounded quantities. Dedicated social-cognitive agent architectures for crowd simulation do exist [214], but they remain a minority and have not displaced force- or rule-based models as the default. Because the deep cognitive and neural mechanisms of pedestrian decision-making remain unmodeled, the mechanism-driven paradigm is confined to describing observed

phenomena, which caps its explainability, its transfer to unseen contexts, and its capacity to predict emergent crowd behavior from first principles.

(2) **Structural defects in data foundations: multi-source heterogeneity and distribution shift.** Fusion models exhibit extreme sensitivity to training data quality. Currently, empirical data derived from video surveillance, inertial sensing, and RFID tags are plagued by severe multi-source heterogeneity in format, measurement precision, and sampling frequencies; fragmented or missing trajectory segments, pervasive observation noise, and localized sensing blind spots remain ubiquitous, so cross-modal spatiotemporal alignment incurs prohibitive preprocessing overheads. Furthermore, under the constraints of stringent privacy regulations and intrinsic collection bottlenecks, authentic data representing safety-critical regimes, such as military perimeters or high-density crowd stampedes, remain extraordinarily scarce. This chronic data deficit induces pronounced distribution shifts and exacerbates severe long-tail effects [144, 145]. Recent studies on domain generalization for trajectory prediction [215] have demonstrated that models trained on limited source domains suffer significant performance degradation when transferred to unseen cultural or spatial contexts, yet a theoretical framework linking data coverage to generalization guarantees remains absent. The severity of this deficit is paradigm-specific: parameter-calibration models fit only a handful of micro-parameters and therefore require few scenarios [195, 197, 198], differentiable-embedding models are fitted end to end against high-fidelity trajectories and demand far more [203–205], and PIML models can trade data volume against the binding strength of their physical constraints [208].

(3) **Architectural dilemmas in fusion coupling mechanisms and generalization fragility.** The trade-off is already visible in the models reviewed above: ORCANet replaces the hard velocity-obstacle geometry with a differentiable relaxation [200], SPDiff and RAPID inject physical structure as sampling priors that encourage but cannot enforce admissibility [190, 191], whereas STDDN and HINN embed conservation laws that hold by construction [203, 204]. Purely data-driven approaches, minimized via conventional mean squared error (MSE) or negative log-likelihood (NLL) objectives, inherently lack spatial exclusion constraints; consequently, they inevitably yield catastrophic physical violations, such as pedestrian interpenetration or ghosting through static obstacles, during long-horizon simulations, with local errors cascading and amplifying across temporal horizons. To mitigate this, existing hybrid frameworks frequently confront a soft-versus-hard constraint dilemma when incorporating physical priors. Imposing physical laws as regularization terms in the loss landscape (i.e., soft constraints typical of standard physics-informed machine learning) merely suppresses the statistical probability of physical violations but fails to deliver deterministic safety guarantees [216]. Conversely, hard-coding these priors as rigid network architectural invariances often severely compromises the model's expressivity regarding anomalous or atypical interaction dynamics. Moreover, the differentiable approximation of mechanistic computational graphs (i.e., constructive fusion) tends to degrade drastically at extreme density boundaries. Similarly, reward shaping in reinforcement learning remains tethered to hand-crafted expert priors that cannot capture macroscopic crowd dynamics. The recent development of self-adaptive PINNs for crowd dynamics [208] has attempted to dynamically balance physics and data fidelity terms during training, yet the fundamental question remains open: under what density and interaction regimes does soft-constraint regularization provably suffice, and when must hard architectural constraints be imposed to guarantee physical plausibility? Furthermore, the integration of physics-informed distillation [217] offers a promising pathway to automatically discover physically consistent network structures, but its applicability to multi-agent pedestrian systems with stochastic interactions has not been validated.

(4) **Computational overheads and real-time inference bottlenecks under high-dimensional coupling.** Computational cost is family-specific: neural ODE and neural PDE solvers scale with the number of interacting neighbors and the resolution of the macroscopic field [201–204]; autoregressive language agents cannot run at simulation frequency unless inference is parallelized across agents [21, 22]; and diffusion-based generators meet the time budget only after map perception was decoupled from the denoising loop [191]. In city-scale crowd movement simulation, concurrently processing massive multimodal sensor streams and high-dimensional physical constraints imposes prohibitive loads on hardware substrates. Consequently, existing hybrid models suffer from inference latency that fails to satisfy the millisecond-level demands of real-time digital twins [218], which precludes their use in time-critical applications such as congestion mitigation and evacuation guidance on edge devices rather than on computing clusters. The recent NeuralDEM framework [219] has demonstrated that multi-branch neural operator architectures can achieve real-time simulation of industrial-scale particle systems by replacing expensive discrete element method solvers with learned surrogates, achieving speedups of several orders of magnitude. However, a critical gap remains: can equivalent neural surrogate strategies be devised for the coupled ODE-PDE systems characteristic of multi-scale crowd dynamics, where the interaction topology is graph-structured rather than spatially regular?

(5) **Fragmentation of evaluation frameworks and lack of downstream task orientation.** Current evaluation frameworks for pedestrian and crowd simulation exhibit pronounced fragmentation. While trajectory prediction methodologies primarily optimize for microscopic kinematic metrics such as average displacement error

(ADE) and final displacement error (FDE), mechanism-driven models prioritize fidelity against macroscopic fundamental diagrams (FD), and generative paradigms increasingly rely on the Fréchet inception distance (FID) [149] or multimodal diversity indices. These disparate criteria lack commensurability; as a result, models achieving high microscopic tracking accuracy often yield sub-optimal performance in capturing macroscopic crowd phenomenology. Furthermore, the vast majority of existing benchmarks remain confined to the kinematic level; audits of agent-level physical integrity, such as the reconstruction of bird's-eye trajectories from generated video in PEDRA [189], are only now appearing, and none of them quantify a model's value in safety management or layout optimization. A recent evaluation framework proposed by Mahzoon et al. [169] had attempted to bridge the gap between trajectory prediction and crowd simulation by establishing standardized benchmarking protocols under limited real-world training data, yet the community still lacks a consensus on what constitutes a minimally sufficient evaluation protocol that jointly audits microscopic trajectory fidelity, macroscopic flow consistency, physical plausibility, and downstream task utility.

7.2 Future directions

To surmount the aforementioned bottlenecks, future research in pedestrian movement modeling and simulation must advance along four complementary dimensions: deepening the mechanistic understanding of pedestrian behavior, orchestrating a self-adaptive synergy between data and mechanics, accelerating computation for large-scale real-time simulation, and consolidating fragmented metrics into task-oriented evaluation.

(1) **Deepening the mechanism layer with an explicit perceptual-cognitive substrate.** Current mechanism models compress the perception, decision, and social appraisal that generate movement into a handful of unidentifiable force or transition parameters [211, 213]. The future direction is therefore to deepen the mechanism layer rather than reparameterize it, by giving pedestrian agents an explicit and inspectable perceptual-cognitive substrate. Recent advances in artificial intelligence make such a substrate feasible: cognitive architectures for crowd agents [214], cognitive maps and grid-like representations of human spatial navigation [220,221], brain-inspired spiking computation [222], embodied vision-language-action models [184, 223, 224], and world-model rollouts [225, 226] together supply the computational ingredients for constructing pedestrian cognition. Two questions follow. First, what minimal set of cognitive constructs, spanning perceptual fields, goal and memory representations, and social appraisal, is necessary and sufficient to reproduce the empirical phenomena that force models miss, such as group walking and social-identity effects [212,213], and can this set be identified from trajectory data rather than assumed? Second, can such a substrate be learned end to end while retaining an inspectable mechanistic core, rather than collapsing into the black-box predictor already shown to fail at agent-level interaction [189]? Answering these would replace the phenomenological parameters of current mechanism models with grounded and identifiable cognitive mechanisms, lifting the explainability and transfer ceiling identified in the corresponding challenge.

(2) **Paradigm-specific data strategies for heterogeneous and shifted data foundations.** Since the data deficit is paradigm-specific, the future directions should be as well. For parameter-calibration models, which fit only a handful of micro-parameters, the relevant question is how few scenarios suffice and which ones carry the most information. The concrete goal is a coverage criterion that predicts which scenarios must be included to avoid the cross-cultural and cross-spatial degradation reported in [218]. For differentiable-embedding models, whose bottleneck is trajectory fidelity rather than scenario count, the question becomes whether the high-fidelity requirement can be relaxed without losing the smoothness and temporal discretization properties of the learned operator [209–211]; specifically, whether noisy or partially observed trajectories suffice when the loss penalizes violations of the mechanistic constraint rather than deviations from the observed trajectory. For PIML models, the trade-off between data volume and constraint strength is itself the research object, and the open question is where the exchange rate lies: how much physical constraint, expressed as a loss term or an architectural bias [214, 215], is equivalent to one additional high-fidelity scenario, and whether this rate survives the distribution shifts described above.

(3) **Self-adaptive dynamic synergy between data and mechanics.** The architectural dilemma above is not that soft and hard constraints are each flawed, but that current frameworks commit to one of them statically, so the balance between data-driven expressiveness and mechanistic guarantees cannot shift with the regime. The future direction is therefore an architecture in which this balance is a learned, context-dependent quantity rather than a fixed design choice. Spatiotemporal attention offers a concrete substrate: attention weights over data-driven and mechanistic branches can be conditioned on local density, interaction intensity, and scene semantics, so that free-flow regimes let the data branch capture social intentions while extreme densities shift authority to the mechanistic branch and preserve the physical floor of collision avoidance. SA-PINN [208] demonstrates that such arbitration is trainable. Two architectural questions follow. First, how should the two branches be coupled: through shared

latent representations, through attention over branch outputs, or through a gating module whose parameters are themselves regime-dependent, and which of these preserves gradient flow when one branch is temporarily suppressed? Second, how can physics-informed distillation [217] be used not merely to prune constraints but to discover which mechanistic structures the data branch should defer to in each regime, so that the architecture selects its own physical priors rather than being given them.

(4) **Differentiable acceleration and lightweight architectures for large-scale simulation.** The overhead above is family-specific: neural ODE and PDE solvers scale with neighbor count and field resolution [201–204], autoregressive language agents cannot run at simulation frequency unless inference is parallelized [21, 22], and diffusion generators meet the time budget only after decoupling map perception from denoising [191]. Two directions follow. The first is to compress the computation graph: distillation and structured pruning can turn a heavy differentiable physics model into a lightweight surrogate, as NeuralDEM [219] shows for industrial particle systems, but it remains open whether the same substitution works for the coupled ODE-PDE systems of multi-scale crowd dynamics, where the interaction topology is graph-structured rather than spatially regular, and whether distillation preserves the conservation properties that make the mechanistic branch useful. The second is to decouple computation rather than compress it: asynchronous macro-micro decomposition, updating the flow field at coarse resolution while rolling out trajectories at fine resolution only in high-interaction regions, would let the two branches run on different hardware budgets.

(5) **From fragmented metrics to unified, task-oriented evaluation.** The fragmentation above is not merely a matter of incompatible metrics but of missing dimensions: ADE and FDE [18], Fundamental Diagram fidelity, and FID or diversity indices [149] each describe a different level of the same system, and no current benchmark audits all of them jointly. The first direction is therefore a minimally sufficient protocol that combines the four levels already identified, namely microscopic trajectory fidelity, macroscopic flow consistency, physical plausibility in the form of violation rates such as interpenetration frequency and collision duration, and downstream task utility. Mahzoon et al. [169] provided a starting point by bridging trajectory prediction and crowd simulation under limited training data, and the causal intervention framework [227] adds a dimension absent from existing benchmarks by testing whether models capture genuine interaction mechanisms rather than spurious correlations. What neither resolves is the weighting problem: when a model improves microscopic accuracy while degrading macroscopic consistency, the protocol must specify which level dominates under which application, and this cannot be settled by reporting all metrics side by side. The second direction follows from the first, namely task-oriented protocols that quantify engineering value directly, for example evacuation-time reduction or bottleneck congestion mitigation rate, so that physical plausibility is evaluated by its consequences rather than as an isolated score. The unresolved question shared by both directions is how to define a minimally sufficient protocol whose dimensions are commensurable, and whether downstream utility can be predicted from the lower-level metrics rather than measured separately in every application.

8 Conclusion

This paper has presented a systematic cross-paradigm review of pedestrian movement modeling and simulation methodologies. Organized by the source of a model's governing dynamics, it traces the evolution of crowd simulation from physical principles and rule-based structures to connectionist representations and data-mechanism fusion architectures, and clarifies the complementary strengths, structural trade-offs, and applicable boundaries of the three paradigms.

A comparative synthesis of the fields long-term trajectory yields three core analytical insights.

(1) Mechanism-driven models, encapsulating force-based, cellular automata, fluid-dynamic, and velocity-space models, provide clear interpretability and hard physical constraints. By natively embedding kinetic invariants like volume exclusion, they are uniquely reliable for safety-critical simulations. However, as crowd environments scale toward higher spatial densities and complex social-psychological heterogeneity, rule-based formulations inherently suffer from parameter calibration gridlocks and rule redundancy.

(2) Data-driven models overcome these dimensional limitations by reformulating trajectory simulation as an inductive learning process. Deep learning architectures, graph neural networks, and generative artificial intelligence can autonomously extract complex, non-linear spatiotemporal interaction distributions from empirical telemetry without rigid structural assumptions. Nevertheless, their black-box nature, heavy reliance on high-quality training distributions, vulnerability to autoregressive error amplification over long rollouts, and lack of explicit boundary constraints present ongoing validation challenges.

(3) Data-mechanism fusion-driven approaches represent a vital conceptual convergence. Tracing its developmental

pipeline, the fusion domain has progressed from superficial parameter-level bridging to structural embedding via differentiable computation graphs, and ultimately toward the autonomous discovery of governing physical equations. This paradigm successfully balances data efficiency with physical consistency, establishing a robust foundation for next-generation multi-agent simulation.

Acknowledgements This work was supported by National Science and Technology Major Project for Brain Science and Brain-like Intelligence Technology (Grant No. 2025ZD0215700), National Natural Science Foundation of China (Grant Nos. 72642020, 72501291, 92467302, 72421002, 72474223, 72301285), Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (Grant No. JYB2025XDXM202), and Major Program of Xiangjiang Laboratory (Grant No. 24XJJCYJ01001).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1 Golchoubian M, Ghafurian M, Dautenhahn K, et al. Pedestrian trajectory prediction in pedestrian-vehicle mixed environments: a systematic review. IEEE Trans Intell Transp Syst, 2023, 24: 11544–11567
2 Zhong J, Cheng T, Liu W L, et al. An evolutionary guardrail layout design framework for crowd control in subway stations. IEEE Trans Comput Soc Syst, 2023, 10: 297–310
3 Duan W, Wang X, Guo G, et al. Modeling human temporal and spatial structured contacts for epidemic prediction. IEEE Trans Comput Soc Syst, 2024, 11: 340–351
4 Haworth B, Usman M, Schaumann D, et al. Gamification of crowd-driven environment design. IEEE Comput Grap Appl, 2021, 41: 107–117
5 He W, Pan G, Wang X, et al. Real-time crowd formation control in virtual scenes. Simul Model Pract Theor, 2023, 122: 102662
6 Xu M L, Wang Z, Meng Y, et al. Virtual-physical digital twin testbed for heterogeneous crowd operations. Sci China Inf Sci, 2025, 68: 154101
7 Liu Q, Lu L, Zhang Y, et al. Modeling the dynamics of pedestrian evacuation in a complex environment. Physica A, 2022, 585: 126426
8 Zhang J, Zhu J, Dang P, et al. An improved social force model (ISFM)-based crowd evacuation simulation method in virtual reality with a subway fire as a case study. Int J Digital Earth, 2023, 16: 1186–1204
9 Qi L Y, Hu X M. Design of evacuation strategies with crowd density feedback. Sci China Inf Sci, 2016, 59: 010204
10 Khan S, Deng Z. Agent-based crowd simulation: an in-depth survey of determining factors for heterogeneous behavior. Vis Comput, 2024, 40: 4993–5004
11 Minartz K, Hendriks F, Koop S M, et al. Discovering interaction mechanisms in crowds via deep generative surrogate experiments. Sci Rep, 2025, 15: 10385
12 Dufour O, Dang H T, Cordes J, et al. Dense crowd dynamics and pedestrian trajectories: a multiscale field dataset from the festival of lights in Lyon. Sci Data, 2025, 12: 718
13 Zeng H, Xu J H. Mapping the urban mobility intelligence frontier: a scientometric analysis of data-driven pedestrian trajectory prediction and simulation. 2025. ArXiv:2510.10327
14 Dang H T, Gaudou B, Verstaevel N. A literature review of dense crowd simulation. Simul Model Pract Theor, 2024, 134: 102955
15 Scholler C, Aravantinos V, Lay F, et al. What the constant velocity model can teach us about pedestrian motion prediction. IEEE Robot Autom Lett, 2020, 5: 1696–1703
16 Fox E, Sudderth E B, Jordan M I, et al. Bayesian nonparametric inference of switching dynamic linear models. IEEE Trans Signal Process, 2011, 59: 1569–1585
17 Xue J X, Yin H, Lv P, et al. Crowd queuing simulation with an improved emotional contagion model. Sci China Inf Sci, 2019, 62: 044101
18 Alahi A, Goel K, Ramanathan V, et al. Social LSTM: human trajectory prediction in crowded spaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. 961–971
19 Xie G, Shangguan A Q, Fei R, et al. Motion trajectory prediction based on a CNN-LSTM sequential model. Sci China Inf Sci, 2020, 63: 212207
20 Gao H B, Su H, Cai Y F, et al. Trajectory prediction of cyclists based on dynamic Bayesian network and long short-term memory model at unsignalized intersections. Sci China Inf Sci, 2021, 64: 172207
21 Bae I, Lee J, Jeon H-G. Can language beat numerical regression? Language-based multimodal trajectory prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024. 753–766
22 Zhao Y Y, Wu Q, Wang Y X, et al. Dynamic path navigation for motion agents with LLM reasoning. 2025. ArXiv:2503.07323
23 Zhang C, Bengio S, Hardt M, et al. Understanding deep learning requires rethinking generalization. 2016. ArXiv:1611.03530
24 Berggold P, Čukarska A, Nousias S, et al. Machine learning in pedestrian and evacuation dynamics for the built environment: a systematic literature review. Saf Sci, 2026, 198: 107143
25 Mai W, Duives D, Hoogendoorn S. A learning based pedestrian flow prediction approach with diffusion behavior. Transp Res Part C-Emerg, 2025, 179: 105243
26 Karniadakis G E, Kevrekidis I G, Lu L, et al. Physics-informed machine learning. Nat Rev Phys, 2021, 3: 422–440
27 Chen R T Q, Rubanova Y, Bettencourt J, et al. Neural ordinary differential equations. In: Proceedings of the Advances in Neural Information Processing Systems, 2018

28 Cui J X, Wang Q P, Sun B Y, et al. Learning continuous network emerging dynamics from scarce observations via data-adaptive stochastic processes. Sci China Inf Sci, 2024, 67: 222206
29 Zou Y, Zhao X, Liu Y. Measuring crowd collectiveness by macroscopic and microscopic motion consistencies. IEEE Trans Multimedia, 2018, 20: 3311–3323
30 Duives D C, Daamen W, Hoogendoorn S P. State-of-the-art crowd motion simulation models. Transp Res Part C-Emer, 2013, 37: 193–209
31 Yang S, Li T, Gong X, et al. A review on crowd simulation and modeling. Graph Model, 2020, 111: 101081
32 Pelechano N, Allbeck J M, Badler N I. Virtual Crowds: Methods, Simulation, and Control. Cham: Springer, 2008
33 Lee N. Encyclopedia of Computer Graphics and Games. Cham: Springer, 2024
34 Martinez-Gil F, Lozano M, García-Fernández I, et al. Modeling, evaluation, and scale on artificial pedestrians. ACM Comput Surv, 2018, 50: 1–35
35 Basu N, Haque M M, King M, et al. A systematic review of the factors associated with pedestrian route choice. Transp Rev, 2022, 42: 672–694
36 Liu C, Liu Z, Chai Y. Review of virtual simulation of crowd motion for urban emergency management. Transp Res Record-J Transp Res Board, 2023, 2677: 1208–1225
37 Sharbini H, Sallehuddin R, Haron H. Crowd evacuation simulation model with soft computing optimization techniques: a systematic literature review. J Manage Analytics, 2021, 8: 443–485
38 Wang X, Huang J, Tian Y, et al. Parallel driving with big models and foundation intelligence in cyber-physical-social spaces. Research, 2024, 7: 0349
39 Reichstein M, Camps-Valls G, Stevens B, et al. Deep learning and process understanding for data-driven Earth system science. Nature, 2019, 566: 195–204
40 Tao F, Qi Q, Wang L, et al. Digital twins and cyber-physical systems toward smart manufacturing and industry 4.0: correlation and comparison. Engineering, 2019, 5: 653–661
41 Korbmacher R, Tordeux A. Review of pedestrian trajectory prediction methods: comparing deep learning and knowledge-based approaches. IEEE Trans Intell Transp Syst, 2022, 23: 24126–24144
42 Zhong J, Li D, Huang Z, et al. Data-driven crowd modeling techniques: a survey. ACM Trans Model Comput Simul, 2022, 32: 1–33
43 Azahar M A B M, Sunar M S, Daman D, et al. Survey on real-time crowds simulation. In: Proceedings of the International Conference on Technologies for E-Learning and Digital Entertainment, 2008. 573–580
44 Kouskoulis G, Spyropoulou I, Antoniou C. Pedestrian simulation: theoretical models vs. data driven techniques. Int J Transp Sci Tech, 2018, 7: 241–253
45 Beacco A, Pelechano N, Andújar C. A survey of real-time crowd rendering. In: Proceedings of the Computer Graphics Forum, 2016. 32–50
46 Al-Shaery A M, Alshehri S S, Farooqi N S, et al. In-depth survey to detect, monitor and manage crowd. IEEE Access, 2020, 8: 209008
47 Rasouli A. Pedestrian simulation: a review. 2021. ArXiv:2102.03289
48 Lv R J, Sun Q B, Li H. A review of crowd evacuation simulation based on model. In: Proceedings of the 2021 China Automation Congress (CAC), 2021. 538–541
49 Senanayake G P D P, Kieu M, Zou Y, et al. Agent-based simulation for pedestrian evacuation: a systematic literature review. Int J Disast Risk Re, 2024, 111: 104705
50 Zambare P, Liu Y. Modeling and simulation of pedestrian behavior in traffic environment: a systematic review of approaches and software. IEEE Access, 2025, 13: 149576
51 Musse S R, Cassol V J, Thalmann D. A history of crowd simulation: the past, evolution, and new perspectives. Vis Comput, 2021, 37: 3077–3092
52 Reynolds C W. Flocks, herds and schools: a distributed behavioral model. In: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, 1987. 25–34
53 Cheng L, Yarlagadda R, Fookes C, et al. A review of pedestrian group dynamics and methodologies in modelling pedestrian group behaviours. World, 2014, 1: 2–13
54 Nicolas A, Hassan F H. Social groups in pedestrian crowds: review of their influence on the dynamics and their modelling. Transpmetrica A-Transp Sci, 2023, 19: 1970651
55 Papadimitriou E, Yannis G, Golias J. A critical assessment of pedestrian behaviour models. Transp Res Part F-Traf, 2009, 12: 242–255
56 Ahmad A, Jamshidi P, Pahl C. Protocol for Systematic Literature Review. Technical Report. 2012. https://www.academia.edu/2747051/Protocol_for_Systematic_Literature_Review
57 Lavin A, Krakauer D, Zenil H, et al. Simulation intelligence: towards a new generation of scientific methods. 2021. ArXiv:2112.03235
58 von Rueden L, Mayer S, Beckh K, et al. Informed machine learning—a taxonomy and survey of integrating prior knowledge into learning systems. IEEE Trans Knowl Data Eng, 2023, 35: 614–633
59 Willard J, Jia X, Xu S, et al. Integrating scientific knowledge with machine learning for engineering and environmental systems. ACM Comput Surv, 2023, 55: 1–37
60 Silva A R D, Lages W S, Chaimowicz L. Boids that see. Comput Entertain, 2009, 7: 1–20
61 Codling E A, Plank M J, Benhamou S. Random walk models in biology. J R Soc Interface, 2008, 5: 813–834
62 Xiong H, Yao L, Tan H, et al. Pedestrian walking behavior revealed through a random walk model. Discrete Dyn Nat Soc, 2012: 405907
63 Murakami H, Feliciani C, Nishinari K. Lévy walk process in self-organization of pedestrian crowds. J R Soc Interface, 2019, 16: 20180939
64 Antonini G, Bierlaire M, Weber M. Discrete choice models of pedestrian walking behavior. Transp Res Part B-Methological, 2006, 40: 667–687
65 Van den Berg J, Lin M, Manocha D. Reciprocal velocity obstacles for real-time multi-agent navigation. In: Proceedings of the 2008 IEEE International Conference on Robotics and Automation, 2008. 1928–1935

66 Van den Berg J, Guy S J, Lin M, et al. Reciprocal n-body collision avoidance. In: Proceedings of the Robotics Research: the 14th International Symposium ISRR, 2011. 3–19
67 Kim S, Guy S J, Hillesland K, et al. Velocity-based modeling of physical interactions in dense crowds. Vis Comput, 2015, 31: 541–555
68 Sudkhot P, Wong K W, Sombatheera C. Collision avoidance and path planning in crowd simulation. ICIC Express Lett, 2022, 17: 13–24
69 Yu H, Li X, Song W, et al. A mixed crowd movement model incorporating chasing behavior. Simul Model Pract Theor, 2025, 138: 103044
70 Echeverri L C, Auberlet J M, Hubert J P. Modeling collision avoidance behavior with zero-speed pedestrians. IEEE Trans Intell Transp Syst, 2024, 25: 9608–9617
71 Helbing D, Molnár P. Social force model for pedestrian dynamics. Phys Rev E, 1995, 51: 4282–4286
72 Helbing D, Farkas I, Vicsek T. Simulating dynamical features of escape panic. Nature, 2000, 407: 487–490
73 Zhao R, Wei B, Han C, et al. Improved crowd dynamics analysis considering physical contact force and panic emotional propagation. IEEE Trans Intell Transp Syst, 2025, 26: 1840–1851
74 Song Y, Hu X, Shen L, et al. Modeling domino effect along the queue using an improved social force model. Physica A, 2023, 625: 129008
75 Bai C Z, Lv J H, Yan D P, et al. SES-SFM: a heterogeneous crowd simulation model with emotional contagion. In: Proceedings of the 2024 9th International Conference on Electronic Technology and Information Science (ICETIS), 2024. 67–72
76 Wang Y, Ding N, Dong D, et al. Simulation of crowd evacuation under attack considering emotion spreading. Chin Phys B, 2024, 33: 128901
77 Fu L, Qin H, He Y, et al. Application of the social force modelling method to evacuation dynamics involving pedestrians with disabilities. Appl Math Computat, 2024, 460: 128297
78 Aros S K, Frakes D. Modeling pedestrian movement in a crowd context with urgency preemption. In: Proceedings of the 2025 Winter Simulation Conference (WSC), 2025. 1851–1861
79 LeGrand A, Lee S. Extending social force model for the design and development of crowd control and evacuation strategies using hybrid simulation. In: Proceedings of the 2025 Winter Simulation Conference (WSC), 2025. 1–12
80 Lu Z X, Zhao W, Liu Y T, et al. Rumor dissemination social force model based on the social force model of crowd evacuation. In: Proceedings of the 4th International Conference on Electronic Information Engineering and Data Processing (EIEDP 2025), 2025. 1305–1311
81 Dietrich F, Köster G. Gradient navigation model for pedestrian dynamics. Phys Rev E, 2014, 89: 062801
82 Ding N, Zhu Y, Liu X, et al. A modified social force model for crowd evacuation considering collision predicting behaviors. Appl Math Computation, 2024, 466: 128448
83 Ko M, Kim T, Sohn K. Calibrating a social-force-based pedestrian walking model based on maximum likelihood estimation. Transportation, 2013, 40: 91–107
84 Makinoshima F, Takahashi T, Oishi Y. Bayesian behavioural model estimation for live crowd simulation. In: Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems, 2024. 1355–1362
85 García A, Hernández-Delfin D, Lee D J, et al. Limited visual range in the social force model: effects on macroscopic and microscopic dynamics. Physica A, 2023, 612: 128461
86 Kim J, Shin C, Park H. Modeling real-time adaptive pedestrian behavior based on social force model. jkst, 2026, 44: 38–52
87 Wang P, Wang X D, Luh P, et al. CrowdEgress: a multi-agent simulation platform for pedestrian crowd. 2024. ArXiv:2406.08190
88 Ávila Moré Y, Al-Hadithi B M, Cadix Martín V. Modelling pedestrian behaviour using swarm techniques. IEEE Latin Am Trans, 2024, 22: 670–677
89 de Almeida D S, Brito e Abreu F, Boavida-Portugal I. Agent-based simulation of non-urgent egress from mass events in open public spaces. Simul Model Pract Theor, 2024, 136: 103002
90 Jun H, Qi Z, Ling F, et al. Simulation of a multi-floor evacuation with the extended social force model in a multi-corridor environment. Phys Scr, 2023, 98: 105241
91 Burstedde C, Klauck K, Schadschneider A, et al. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. Physica A, 2001, 295: 507–525
92 Kirchner A, Schadschneider A. Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics. Physica A, 2002, 312: 260–276
93 Ji J, Lu L, Jin Z, et al. A cellular automata model for high-density crowd evacuation using triangle grids. Physica A, 2018, 509: 1034–1045
94 Feliciani C, Nishinari K. An improved Cellular automata model to simulate the behavior of high density crowd and validation by experimental data. Physica A, 2016, 451: 135–148
95 Huo F, Li Y, Li C, et al. An extended model describing pedestrian evacuation considering pedestrian crowding and stampede behavior. Physica A, 2022, 604: 127907
96 Guo R Y, Huang H J. A modified floor field cellular automata model for pedestrian evacuation simulation. J Phys A-Math Theor, 2008, 41: 385104
97 Kirchner A, Nishinari K, Schadschneider A. Friction effects and clogging in a cellular automaton model for pedestrian dynamics. Phys Rev E, 2003, 67: 056122
98 Henein C M, White T. Agent-based modelling of forces in crowds. In: Proceedings of the International Workshop on Multi-Agent Systems and Agent-Based Simulation, 2004. 173–184
99 Henein C M, White T. Macroscopic effects of microscopic forces between agents in crowd models. Physica A, 2007, 373: 694–712
100 Zeng J B, Leng B, Xiong Z, et al. Pedestrian dynamics in a two-dimensional complex scenario using a local view floor field model. Int J Mod Phys C, 2011, 22: 775–803
101 Suma Y, Yanagisawa D, Nishinari K. Anticipation effect in pedestrian dynamics: modeling and experiments. Physica A, 2012, 391: 248–263
102 Ma Y, Tan L, Zhang Y, et al. Modeling study of pedestrian evacuation considering dynamic guidance under terrorist attack events. Phys

Lett A, 2025, 532: 130190
103 Li Y, Li C, Guo C, et al. Dynamic coupling model of FDS and cellular automata considering trampling behavior. Physica A, 2024, 633: 129369
104 Yu B, Ye J, Ou D. Consideration of human vision in crowd simulations. IEEE Trans Intell Transp Syst, 2024, 25: 13364–13374
105 Parisi F, Feliciani C, Lovreglio R. What do we head for while exiting a room? A novel parametric distance map for pedestrian dynamic simulations. Transp Res Part C-Emerg, 2023, 156: 104335
106 Li S, Li Q, Zhong G, et al. Simulation of bi-directional pedestrian flow in corridor based on direction fuzzy visual field. Sci Rep, 2023, 13: 19261
107 Chen Y, Yao J, Gao C, et al. Pedestrian flow model based on cellular automata under visual trajectory and multi-scenario evacuation simulation research. Sensors, 2026, 26: 1405
108 Bandini S, Briola D, Dennunzio A, et al. Distance-based affective states in cellular automata pedestrian simulation. Nat Comput, 2024, 23: 71–83
109 He Y, Shi Q, Fu L, et al. Application of the multi-grid modelling method to pedestrian social group dynamics through a bottleneck. Simul Model Pract Theor, 2025, 139: 103049
110 Qin S H, Li N. Study on pedestrian evacuation simulation model considering group behavior. Physica A, 2024, 654: 130138
111 Tamang N, Sun Y. Application of the dynamic Monte Carlo method to pedestrian evacuation dynamics. Appl Math Comput, 2023, 445: 127876
112 Chen L, Guo Z L, Wang T, et al. An evacuation guidance model for heterogeneous populations in large-scale pedestrian facilities with multiple exits. Physica A, 2023, 620: 128740
113 Hayashida T, Sekizaki S, Furuya Y, et al. ACS2-powered pedestrian flow simulation for crowd dynamics. AppliedMath, 2025, 5: 88
114 Zhao H, Guo T, Tong W, et al. PaCS: a parallel computation framework for field-based crowd simulation. IEEE Trans Intell Transp Syst, 2023, 24: 12659–12670
115 Chenney S. Flow tiles. In: Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, 2004. 233–242
116 Patil S, van den Berg J, Curtis S, et al. Directing crowd simulations using navigation fields. IEEE Trans Vis Comput Gr, 2011, 17: 244–254
117 Tsai T Y, Wong S K, Chou Y H, et al. Directing virtual crowds based on dynamic adjustment of navigation fields. Comput Animat Virt, 2018, 29: e1765
118 Li Y, Liu J, Guan X, et al. Introducing anisotropic fields for enhanced diversity in crowd simulation. Vis Comput, 2025, 41: 7687–7702
119 Zhu L, Xu X, Wang J, et al. An enhanced multi-attribute crowd evacuation model incorporating the effects of environmental impact factors. Heliyon, 2024, 10: e26299
120 Yu T, Yang H D. Simulation of running crowd dynamics: potential-based cellular automata model. IEEE Access, 2023, 11: 138602
121 Macal C M, North M J. Tutorial on agent-based modeling and simulation. In: Proceedings of the Winter Simulation Conference, 2005. 14–28
122 Weiss T, Litteneker A, Jiang C, et al. Position-based multi-agent dynamics for real-time crowd simulation. In: Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, 2017. 1–2
123 Weiss T, Litteneker A, Jiang C, et al. Position-based real-time simulation of large crowds. Comput Graphics, 2019, 78: 12–22
124 Guy S J, Chhugani J, Kim C, et al. ClearPath: highly parallel collision avoidance for multi-agent simulation. In: Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, 2009. 177–187
125 Best A, Narang S, Curtis S, et al. DenseSense: interactive crowd simulation using density-dependent filters. In: Proceedings of the Symposium on Computer Animation, 2014. 97–102
126 van Toll W G, Cook Iv A F, Geraerts R. Real-time density-based crowd simulation. Comput Animat Virt, 2012, 23: 59–69
127 Liu S, Lo S, Ma J, et al. An agent-based microscopic pedestrian flow simulation model for pedestrian traffic Problems. IEEE Trans Intell Transp Syst, 2014, 15: 992–1001
128 Chao Q, Deng Z, Jin X. Vehicle-pedestrian interaction for mixed traffic simulation. Comput Animat Virt, 2015, 26: 405–412
129 Sinclair J, Suwanwiwat H, Lee I. A hybrid data gathering and agent based cognitive architecture for realistic crowd simulations. J Simul, 2023, 17: 121–148
130 Wang J, Liu Z, Liu T, et al. A multi-agent motion simulation method for emergency scenario deduction. Cogn Syst Res, 2024, 88: 101275
131 Mello T G V D, Silva M S H D, Silva G F, et al. Exploring crowd dynamics: simulating structured behaviors through crowd simulation models. 2023. ArXiv:2312.06549
132 Huang K, Ding G, Yan D, et al. Choreographing multi-degree of freedom behaviors in large-scale crowd simulations. Comput Graphics, 2024, 124: 104051
133 Lyu L, Pang C, Wang J. Virtual-real syncretic crowd simulation method guided by building safety evacuation sign. IEEE Sens J, 2023, 23: 18486–18499
134 Alidmat O K A, Alazaidah R, Arabiat M, et al. Dynamic exit selection and counting area adjustment for efficient crowd evacuation in asymmetrical layouts. In: Proceedings of the 2024 25th International Arab Conference on Information Technology (ACIT), 2024. 1–10
135 Lv P, Yu Q, Xu B, et al. Emotional contagion-aware deep reinforcement learning for antagonistic crowd simulation. IEEE Trans Affect Comput, 2023, 14: 2939–2953
136 Bustami O, Rouhana F, Bagtzoglou A. A conceptual decision-support agent-based framework for evacuation planning under compound hazards. Sustainability, 2026, 18: 3658
137 Pang C, Lyu L, Zhou Q, et al. Environment-sensitive crowd behavior modeling method based on reinforcement learning. Appl Intell, 2023, 53: 19356–19371
138 Suchak K, Kieu M, Oswald Y, et al. Coupling an agent-based model and an ensemble Kalman filter for real-time crowd modelling. R Soc Open Sci, 2024, 11: 231553

139 Bian C, Guo Y, Meng G, et al. Modeling seismic resilience and hospital evacuation: a comparative analysis of multi-agent reinforcement learning and classical evacuation models. Buildings, 2026, 16: 1538

140 Hu X, Chen T. Crowd dynamics of self-propelled individuals with collision avoidance considering anticipation and intrusion aversion. Chaos Soliton Fract, 2024, 186: 115311

141 Porzycki J, Wąs J. Modeling spatial patterns in a moving crowd of people using data-driven approach—a concept of Interplay Floor Field. Saf Sci, 2023, 167: 106266

142 Giuliari F, Hasan I, Cristani M, et al. Transformer networks for trajectory forecasting. In: Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), 2021. 10335–10342

143 Gupta A, Johnson J, Li F F, et al. Social GAN: socially acceptable trajectories with generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018. 2255–2264

144 Lerner A, Chrysanthou Y, Lischinski D. Crowds by example. Comput Graph Forum, 2007, 26: 655–664

145 Pellegrini S, Ess A, Schindler K, et al. You'll never walk alone: modeling social behavior for multi-target tracking. In: Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, 2009. 261–268

146 Robicquet A, Sadeghian A, Alahi A, et al. Learning social etiquette: human trajectory understanding in crowded scenes. In: Proceedings of the European Conference on Computer Vision, 2016. 549–565

147 Fabbri M, Lanzi F, Calderara S, et al. Learning to detect and track visible and occluded body joints in a virtual world. In: Proceedings of the European Conference on Computer Vision (ECCV), 2018. 430–446

148 Shao S, Zhao Z J, Li B X, et al. CrowdHuman: a benchmark for detecting human in a crowd. 2018. ArXiv:1805.00123

149 Kothari P, Kreiss S, Alahi A. Human trajectory forecasting in crowds: a deep learning perspective. IEEE Trans Intell Transp Syst, 2022, 23: 7386–7400

150 Pouw C A S, van der Vleuten G G M, Corbetta A, et al. Data-driven physics-based modeling of pedestrian dynamics. 2024. ArXiv:2407.20794

151 Qiao Y, Si Z, Zhang Y, et al. A hybrid Markov-based model for human mobility prediction. Neurocomputing, 2018, 278: 99–109

152 Park J, Lee D S, González M C. The eigenmode analysis of human motion. J Stat Mech, 2010, P11021

153 Chen M, Yu X, Liu Y. Mining moving patterns for predicting next location. Inf Syst, 2015, 54: 156–168

154 Yan M, Li S, Chan C A, et al. Mobility prediction using a weighted Markov model based on mobile user classification. Sensors, 2021, 21: 1740

155 Cuttone A, Lehmann S, González M C. Understanding predictability and exploration in human mobility. EPJ Data Sci, 2018, 7: 2

156 Kreutz T, Mühlhäuser M, Guinea S A. Whenever, wherever: towards orchestrating crowd simulations with spatio-temporal spawn dynamics. In: Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA), 2025. 1017–1023

157 Yu Y, Chen Q. A dual-layer pedestrian motion model for crowd movement simulation. Physica A, 2025, 676: 130880

158 Jiang J, Yan K, Xia X, et al. A survey of deep learning-based pedestrian trajectory prediction: challenges and solutions. Sensors, 2025, 25: 957

159 Gu X, Li C, Gao L, et al. A review of pedestrian trajectory prediction methods based on deep learning technology. Sensors, 2025, 25: 7360

160 Zaier M, Wannous H, Drira H, et al. Pedestrian trajectory prediction: a literature review and current trends. Neural Comput Appl, 2025, 37: 28869–28906

161 Du Q, Wu Q, Li L, et al. Review and perspectives on pedestrian trajectory prediction for safe transportation. IEEE Trans Intell Transp Syst, 2026, 27: 26–52

162 Wang T, Zhang Z, Nong T, et al. Modelling the unidirectional and bidirectional flow of pedestrians based on convolutional neural networks. Physica A, 2024, 651: 130021

163 Jiang N, Lee E W M, Yang L, et al. A deep-learning-based approach for simulating pedestrian turning flow. Expert Syst Appl, 2025, 262: 125706

164 Liang X, Lee E W M. Visual-Information-driven model for crowd simulation using temporal convolutional network. IEEE Trans Intell Transp Syst, 2024, 25: 12297–12314

165 Mohamed A, Qian K, Elhoseiny M, et al. Social-STGCNN: a social spatio-temporal graph convolutional neural network for human trajectory prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020. 14424–14432

166 Zhang B, Xu J, Xie C Z T, et al. Forecaster as a simulator: simulating multi-directional pedestrian flow with knowledge-guided Graph Neural Networks. Comput Ind Eng, 2024, 198: 110668

167 Xie W, Jiang N, Ma Y, et al. Simulating pedestrian flow on slopes via transfer learning approach: from single-file to crowd. IEEE Trans Intell Transp Syst, 2025, 26: 3873–3884

168 Minartz K, Hendriks F, Koop S M, et al. Understanding complex crowd dynamics with generative neural simulators. 2024. ArXiv:2412.01491

169 Mahzoon V, Liu A, Vucetic S, et al. Deep learning-based pedestrian simulation with limited real-world training data: an evaluation framework. In: Proceedings of the 34th International Joint Conference on Artificial Intelligence (IJCAI), 2025. 196–204

170 Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature, 2015, 518: 529–533

171 Li Y, Chen Y, Liu J, et al. Efficient crowd simulation in complex environment using deep reinforcement learning. Sci Rep, 2025, 15: 5403

172 Pu Q, Xie K, Guo H, et al. Modeling crash avoidance behaviors in vehicle-pedestrian near-miss scenarios: curvilinear time-to-collision and Mamba-driven deep reinforcement learning. Accident Anal Prev, 2025, 214: 107984

173 Hu K, Haworth B, Berseth G, et al. Heterogeneous crowd simulation using parametric reinforcement learning. IEEE Trans Vis Comput Gr, 2023, 29: 2036–2052

174 Ng W Z T, Chen J, Pan S J, et al. Improving the generalization of unseen crowd behaviors for reinforcement learning based local motion

planners. In: Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA), 2024. 7412–7418
175 Ling B Y, Lyu Y, Li D, et al. SocialGAIL: faithful crowd simulation for social robot navigation. In: Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA), 2024. 16873–16880
176 Lin X, Liang Y, Zhang Y, et al. IE-GAN: a data-driven crowd simulation method via generative adversarial networks. Multimed Tools Appl, 2024, 83: 45207–45240
177 Yan D, Ding G, Huang K, et al. Generating natural pedestrian crowds by learning real crowd trajectories through a transformer-based GAN. Vis Comput, 2025, 41: 1079–1096
178 Yan D, Huang K, Zhang L, et al. Stylized crowd formation transformation through spatiotemporal adversarial learning. Adv Intelligent Syst, 2024, 6: 2300563
179 Zhao B X, Zhang Q, Huang H. EnvSocial-Diff: a diffusion-based crowd simulation model with environmental conditioning and individual-group interaction. In: Proceedings of the International Conference on Learning Representations, 2026. 48358–48378
180 Park J S, O'Brien J, Cai C J, et al. Generative agents: interactive simulacra of human behavior. In: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 2023. 1–22
181 Seo S, Kim J, Shin M, et al. LLMDR: LLM-driven deadlock detection and resolution in multi-agent pathfinding. 2025. ArXiv:2503.00717
182 Liu P, Yuan W, Fu J, et al. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. ACM Comput Surv, 2023, 55: 1–35
183 Ji X B, Pan Z R, Gao X F, et al. Text-guided synthesis of crowd animation. In: Proceedings of the ACM SIGGRAPH 2024 Conference Papers, 2024. 1–11
184 Hwang J, Hong S E, Kim J, et al. CrowdVLA: embodied vision-language-action agents for context-aware crowd simulation. 2026. ArXiv:2604.05525
185 Li D, Su X, Feng Y, et al. Decoding pedestrian crossing intention from egocentric vision via vision language models. 2026. arXiv: 2606.09142
186 Liu T Y, Xu Y J, Liu Y L. VLM-enhanced adversarial scene generation from images and videos for safe autonomous driving. In: Proceedings of the Embodied and Safe-Assured Robotic Systems, 2025
187 Bossen T E W, Møgelmose A, Greer R. Can vision-language models understand and interpret dynamic gestures from pedestrians? Pilot datasets and exploration towards instructive nonverbal commands for cooperative autonomous vehicles. In: Proceedings of the Computer Vision and Pattern Recognition Conference, 2025. 4779–4788
188 Wang Q Z, Wu X Y, Shi G Y, et al. Act on what you see: unlocking safe social navigation in vision-language-action models. 2026. ArXiv:2606.10495
189 Appelle A, Lynch J P. PEDRA: evaluating the realism of pedestrian dynamics in video generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026. 4461–4474
190 Chen H Y, Ding J T, Li Y, et al. Social physics informed diffusion model for crowd simulation. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2024. 474–482
191 Yu Z H, Wang H D, Ding J T, et al. RAPID: a scalable and controllable physics-informed diffusion framework for real-time pedestrian trajectory generation. In: Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2026. 8431–8442
192 Casper S, Davies X, Shi C, et al. Open problems and fundamental limitations of reinforcement learning from human feedback. 2023. ArXiv:2307.15217
193 Newbury R, Collins J, He K, et al. A review of differentiable simulators. IEEE Access, 2024, 12: 97581–97604
194 Jia X, Willard J, Karpatne A, et al. Physics-guided machine learning for scientific discovery: an application in simulating lake temperature profiles. ACM IMS Trans Data Sci, 2021, 2: 1–26
195 Yan D P, Bai C Z, He L J, et al. Deep learning improved social force model to simulate crowd in varying scenes. In: Proceedings of the 2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE), 2024. 307–312
196 Yan D, Ding G, Huang K, et al. Enhanced crowd dynamics simulation with deep learning and improved social force model. Electronics, 2024, 13: 934
197 Hu K D, Yoon S, Pavlovic V, et al. Toward realistic human crowd simulations with data-driven parameter space exploration. In: Proceedings of the 2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR), 2024. 221–225
198 Amartaivan S, Morita H. Efficient calibration of heterogeneous agent-based pedestrian model using PSO with deep learning surrogates. IEEE Access, 2025, 13: 109185
199 Zhou W, Shen W, Meng X. An Improved social force model-driven multi-agent generative adversarial imitation learning framework for pedestrian trajectory prediction. Comput Animat Virt, 2025, 36: e70058
200 Zhang J, Li C, Wang C, et al. ORCANet: differentiable multi-parameter learning for crowd simulation. Comput Animat Virt, 2023, 34: e2114
201 Wen S, Wang H, Metaxas D. Social ODE: multi-agent trajectory forecasting with neural ordinary differential equations. In: Proceedings of the European Conference on Computer Vision, 2022. 217–233
202 Mo Z B, Fu Y J, Di X. PI-NeuGODE: physics-informed graph neural ordinary differential equations for spatiotemporal trajectory prediction. In: Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems, 2024. 1418–1426
203 Liu Z, Geng X, Xu W J, et al. STDDN: a physics-guided deep learning framework for crowd simulation. 2026. ArXiv:2604.02756
204 Zhou Y S, Lai P R, Yu J Q, et al. Hydrodynamics-informed neural network for simulating dense crowd motion patterns. In: Proceedings of the 32nd ACM International Conference on Multimedia, 2024. 4553–4561
205 He F X, Yue J B, Zhu J L, et al. Learning extremely high density crowds as active matters. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025. 540–550

206 Chauhan P, Choutri S E, Ghattassi M, et al. Neural operators struggle to learn complex PDEs in pedestrian mobility: hughes model case study. Artif Intelligence Transp, 2025, 1: 100005

207 Zhang G J, Yu Z Y, Jin D P, et al. Physics-infused machine learning for crowd simulation. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022. 2439–2449

208 Tan Z Y, Luo L B, Yin H Y, et al. Crowd dynamics demand adaptivity: self-adaptive physics-informed neural network for crowd simulation (SA-PINN). In: Proceedings of the ACM Conference, 2025

209 Chen B, Guo R, Zhang Q, et al. A data-driven crowd simulation framework integrating physics-informed machine learning with navigation potential fields. IEEE Trans Comput Soc Syst, 2025, 12: 4384–4400

210 Fang B, Qin F, Wang Y. ARP-STGCN: a fast attraction-repulsion-potential based spatio-temporal graph convolutional network with imputation for pedestrian trajectory prediction. Complex Intell Syst, 2025, 11: 469

211 Moussaïd M, Helbing D, Theraulaz G. How simple rules determine pedestrian behavior and crowd disasters. Proc Natl Acad Sci USA, 2011, 108: 6884–6888

212 Moussaïd M, Perozo N, Garnier S, et al. The walking behaviour of pedestrian social groups and its impact on crowd dynamics. PLoS ONE, 2010, 5: e10047

213 Templeton A, Drury J, Philippides A. From mindless masses to small groups: conceptualizing collective behavior in crowd modeling. Rev Gen Psychol, 2015, 19: 215–229

214 Wijermans N, Jorna R, Jager W, et al. CROSS: modelling crowd behaviour with social-cognitive agents. JASSS, 2013, 16: 1

215 Dong Y, Wang L, Zhou S, et al. Recurrent aligned network for generalized pedestrian trajectory prediction. IEEE Trans Circ Syst Vid, 2025, 35: 2188–2201

216 Raissi M, Perdikaris P, Karniadakis G E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys, 2019, 378: 686–707

217 Liu Z, Liu Y, Yan X, et al. Automatic network structure discovery of physics informed neural networks via knowledge distillation. Nat Commun, 2025, 16: 9558

218 Grieves M. Digital Twin: Manufacturing Excellence Through Virtual Factory Replication. 2014. https://doi.org/10.5281/ZENODO.1493930

219 Alkin B, Kronlachner T, Papa S, et al. NeuralDEM for real time simulations of industrial particular flows. Commun Phys, 2025, 8: 440

220 Epstein R A, Patai E Z, Julian J B, et al. The cognitive map in humans: spatial navigation and beyond. Nat Neurosci, 2017, 20: 1504–1513

221 Banino A, Barry C, Uria B, et al. Vector-based navigation using grid-like representations in artificial agents. Nature, 2018, 557: 429–433

222 Eliasmith C, Stewart T C, Choo X, et al. A large-scale model of the functioning brain. Science, 2012, 338: 1202–1205

223 Brohan A, Brown N, Carbajal J, et al. RT-2: vision-language-action models transfer web knowledge to robotic control. In: Proceedings of the Conference on Robot Learning, 2023. 2165–2183

224 Nguyen D T, Shcherbyna V, Do A D, et al. GROVE: grounded pedestrian simulation via natural language for interactive social robot navigation. 2026. ArXiv:2606.25504

225 Ha D, Schmidhuber J. World models. 2018. ArXiv:1803.10122

226 Hafner D, Pasukonis J, Ba J, et al. Mastering diverse domains through world models. 2023. ArXiv:2301.04104

227 Han X, Xu H. Causal intervention and counterfactual reasoning for multimodal pedestrian trajectory prediction. J Imag, 2025, 11: 379