Diffusion Models for Human-Robot Interaction: A Systematic Review

Ziyi Chang[1], Yueyue Hou[2], Jiannan Li[2], Ruochen Li[1], Junyan Hu[1], Effie Lai-Chong Law[1], Amir Atapour-Abarghouei[1], Hubert P. H. Shum[1*]

[1]Department of Computer Science, Durham University, Durham, DH1 3LE, United Kingdom.
[2]School of Computing and Information Systems, Singapore Management University, 178902, Singapore.

*Corresponding author(s). E-mail(s): hubert.shum@durham.ac.uk;
Contributing authors: ziyi.chang@durham.ac.uk; yueyue.hou.2024@phdcs.smu.edu.sg; jiannan.li@smu.edu.sg; ruochen.li@durham.ac.uk; junyan.hu@durham.ac.uk; lai-chong.law@durham.ac.uk; amir.atapour-abarghouei@durham.ac.uk;

Abstract

Human-robot interaction involves complex, uncertain, and multimodal contexts that challenge conventional modelling approaches. Recent research has increasingly explored the application of diffusion models in human-robot interaction. However, a consolidated overview of how these models are integrated into HRI research and evaluated in interactive contexts remains lacking. This paper presents a PRISMA-guided systematic literature review of diffusion-based approaches in HRI. We searched Scopus, Web of Science, IEEE Xplore, and the ACM Digital Library for studies published up to 20 December 2025. After screening, 15 peer-reviewed studies were retained for analysis. The review synthesises application domains, diffusion model variants, evaluation practices, and reported benefits and limitations. Prior work is largely dominated by denoising diffusion probabilistic model adaptations used as probabilistic forecasters of human motion or interaction-relevant states. We further discuss key limitations and open challenges, outlining directions for future research in data efficiency, generalisation, human-centred evaluation, and uncertainty-aware interaction design.

Keywords: Diffusion models, Human-robot interaction, Generative models

1 Introduction

Human-robot interaction (HRI) increasingly takes place in complex and dynamic environments that challenge conventional modelling approaches. Interactive scenarios often involve uncertainty, partial observability, and multimodal human behaviour [1, 2]. As a result, robots operating alongside humans need to account for diverse behaviours, intentions, and environmental conditions. Developing computational models to capture such complex interaction dynamics therefore remains a central challenge for HRI research.

Diffusion models have recently attracted increasing attention in HRI research because they provide a probabilistic framework for modelling multimodal and uncertain interaction dynamics. In HRI, robots often need to account for multiple plausible human behaviours, intentions, and environmental responses rather than a single deterministic outcome. Diffusion models [3–5] are well suited to this challenge, as they learn complex data distributions through a stochastic denoising process and have demonstrated strong performance across domains such as computer vision [6, 7], natural language processing [8], time-series modelling [9], human motion synthesis [10, 11], video generation [12, 13], and audio modelling [14]. Therefore, recent research has increasingly explored the application of diffusion models in HRI contexts.

While a growing body of survey literature has examined the applications of diffusion models across robotics and machine learning, no systematic literature review has synthesised diffusion-based research in the context of human-robot interaction. To the best of our knowledge, existing diffusion-focused surveys in robotics primarily address specific areas such as manipulation, planning, or policy learning, organising the literature around tasks rather than interaction dynamics between humans and robots [15–19]. Meanwhile, systematic reviews in HRI typically examine human-centred topics such as personality, emotion, and trust [20–23], without considering diffusion models as a core methodological component. Consequently, existing surveys provide limited insight into how diffusion models are applied, integrated, and evaluated in HRI systems.

To address this gap, this paper presents the first systematic literature review of diffusion models in HRI. Through a reproducible search and selection process, we identify and analyse 15 studies that apply diffusion models to HRI-related problems. Our review examines application trends, preference for model variants, and evaluation practices, and discusses the benefits, limitations, and challenges reported in the literature. Based on these findings, we further outline open research challenges and directions for future research on diffusion-based human-robot interaction. The main contributions of this systematic review are summarised as follows:

- The first systematic literature review of diffusion models in human-robot interaction, providing a structured synthesis of the emerging research landscape.
- An analysis of diffusion-based HRI studies, including application domains, model variants, system integration strategies, and evaluation practices.
- Identification of key limitations and open research challenges, highlighting future directions for diffusion-based human-robot interaction.

To further guide this review, we formulate the following research questions (RQs) to understand the current progress of this research field and attempt to answer them all using this systematic review. The formulated questions are as follows:

- **RQ1:** How are diffusion models integrated into human-robot interaction?
- **RQ2:** Which types of diffusion models are employed in HRI research?
- **RQ3:** How are diffusion-based HRI methods evaluated, and which metrics are commonly used?
- **RQ4:** What benefits, limitations, and open challenges arise when applying diffusion models to HRI?

These research questions structure the analysis of the reviewed literature and guide the organisation of the results presented in this paper.

The remainder of this paper is organised as follows. Section 2 describes the literature search strategy and selection criteria. Section 3 presents the review results organised by the research questions. Section 4 outlines open challenges and future research directions, and Section 5 concludes the paper.

2 Methodology

This systematic literature review (SLR) follows a structured, multi-stage process comprising *identification*, *screening and filtering*, and *data synthesis*. The overall procedure is described in detail in this section. To enhance transparency and reproducibility, visual reporting tools, including PRISMA flow diagrams and Venn diagrams, are used to document the decision-making process at each stage of the review.

2.1 Publication Identification

As outlined in Section 1, the primary objective of this SLR is to investigate the application of diffusion models within the context of human-robot interaction (HRI). To identify relevant studies, a comprehensive search has been conducted across four major bibliographic databases: Scopus, ACM Digital Library (ACM DL), Web of Science (WoS), and IEEE Xplore. The search strategy is designed to capture publications at the intersection of diffusion-based generative modelling and HRI.

The following search string has been used:

> (”diffusion model” OR ”score-based” OR ”denoising diffusion”) AND (human robot interact* OR human robot collaborat*)

The search has been implemented using the advanced search facilities of each database as follows:

- IEEE Xplore: Advanced Search → Document Title, Abstract, Author Keywords
- ACM Digital Library (ACM DL): Advanced Search (Full-text Collection) → Title, Abstract, Author Keywords
- Scopus: Document Search → Article Title, Abstract, Keywords
- Web of Science (WoS): Basic Search → Topic (Article Title, Abstract, Author Keywords)

To encourage comprehensive coverage of diffusion-based approaches, the search string incorporates multiple terminologies used in the literature. The term *diffusion model* captures the general class of approaches, while *denoising diffusion* explicitly refers to denoising diffusion probabilistic models (DDPMs). The inclusion of *score-based* accounts for closely related formulations based on score matching and stochastic differential equations.

On the application side, the terms *human robot interact** and *human robot collaborat** have been used with truncation to capture variations such as *interaction*, *interactions*, *collaboration*, and *collaborative*. These terms are combined with diffusion-related keywords with the Boolean operator `AND` to address both the methodological dimension (diffusion models) and the application domain (HRI). This balanced strategy aims to maximise recall while maintaining relevance, in line with established best practices for systematic review search design.

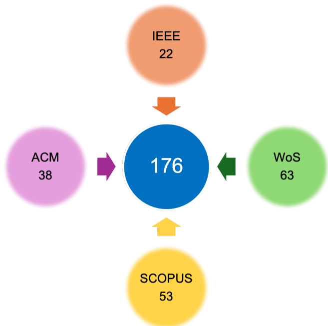

Fig. 1 Distribution of records retrieved from each database during the identification stage.

The selection of databases follows established recommendations for rigorous SLRs. Scopus and Web of Science provide broad coverage across engineering and the natural sciences, while IEEE Xplore and ACM Digital Library are leading publisher databases in robotics and computer science [24–26]. Google Scholar is excluded due to limitations in advanced search functionality and concerns regarding metadata quality [27, 28]. Unpublished and non-peer-reviewed studies are excluded for methodological rigour [29].

Each retrieved record is independently examined by two reviewers to assess eligibility. Inter-rater reliability is quantified using Cohen's kappa coefficient [30], with values of 0.81 or higher interpreted as indicating almost perfect agreement [31]. Inter-rater reliability statistics are reported in Fig. 3.

The final database search is conducted on 20 December 2025. In addition to formally published articles, early access (online first) papers available by 20 December 2025 are also considered and included. In total, 176 records have been retrieved across the four databases. As illustrated in Fig. 1, Web of Science yields the largest number of records (63), followed by Scopus (53), ACM Digital Library (38), and IEEE Xplore (22).

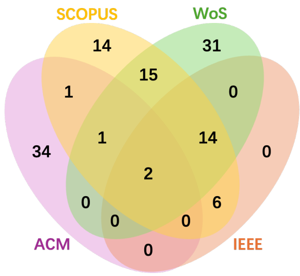

Fig. 2 Overlap of records retrieved across databases during Stage 1 (Identification).

To improve traceability and transparency, duplicate records across databases have been analysed using Venn diagrams [32]. Although overlap analysis is often omitted in SLR reporting, we include it here to provide insight into database coverage and redundancy. Fig. 2 illustrates record overlap at the identification stage. ACM Digital Library contains the largest number of unique records (34), followed by Web of Science (31). Two papers appeared in all four databases. All papers from IEEE Xplore have been covered by the other three resources, in which SCOPUS contributes the largest overlapping number (22).

2.2 Screening and Filtering

2.2.1 Basic Screening

Following identification, retrieved records are screened using predefined inclusion and exclusion criteria (Table 1), as illustrated in the PRISMA flow diagram (Fig. 3). Five records have been excluded due to a lack of full-text access. Duplicate records are then identified and removed. After the removal, 113 unique records remain.

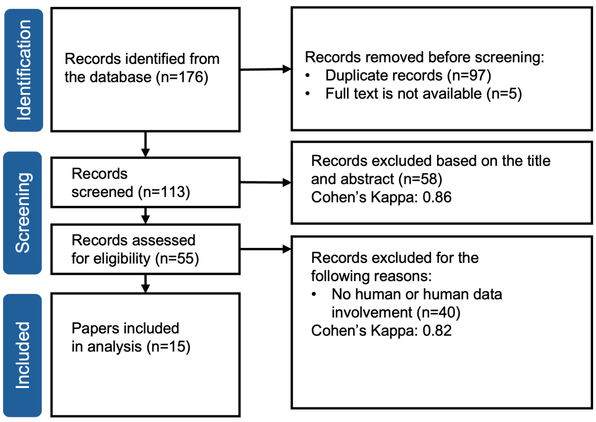

Fig. 3 PRISMA flow diagram for the systematic literature review.

Screening is performed based on titles, abstracts, and figures to assess relevance to the objectives of this systematic literature review. We adopt a broader, system-level interpretation of human-robot interaction. Specifically, studies are included if diffusion models contributed to modelling, anticipating, or responding to human behaviour within an interactive or human-aware robotic system where such human-related dynamics inform system design, planning, or interaction outcomes. After excluding 58 papers, 55 records remain after this step.

Table 1 Inclusion and exclusion criteria for determining study eligibility.

Inclusion (in) criteria	**Exclusion (ex) criteria**
(in1) The study presents or applies a **diffusion-based model**.	(ex1) Conceptual, review-focused, or purely theoretical studies without an implemented or empirically evaluated diffusion component.
(in2) The full text of the paper is **accessible, peer-reviewed, and written in English**.	(ex2) Research that does not relate to **human-robot interaction (HRI)**, such as studies limited to robot-only manipulation or object rearrangement without human presence or consideration.
(in3) The study includes **human involvement**, either through explicit human participation or through modelling human-related states or behaviours that are relevant to interaction at the system level.	(ex3) Papers lacking sufficient methodological or application details to assess the diffusion or HRI-related components.

2.2.2 Advanced Screening

The 55 retained papers undergo full-text review using the same inclusion and exclusion criteria. Studies are excluded if they are purely conceptual or review-based, lack explicit relevance to diffusion models in HRI, or provide insufficient methodological or experimental detail.

Following this advanced screening stage, 15 peer-reviewed studies are identified as meeting all eligibility criteria and thus selected for in-depth analysis. Table 2 lists the final set of studies and their assigned identifiers. The relatively small number of eligible studies reflects the emerging nature of diffusion-based approaches in HRI, where only a limited number of peer-reviewed works have been reported so far.

2.2.3 Data Extraction Process

A structured data extraction scheme (Table 3) has been developed for consistent and comprehensive collection of paper information. Fig. 4 illustrates the overlap where IEEE provides the most relevant papers. Each of the 15 papers is examined in full text to extract data relevant to the four research questions (RQ1-RQ4), covering both descriptive characteristics and methodological details.

Data extraction is conducted independently by two reviewers and is cross-checked for accuracy and consistency. Discrepancies are resolved through discussion and consensus. Extracted data include publication metadata, diffusion model type, integration with HRI framework, evaluation methodology, and benefits, limitations, and challenges.

Table 2 Mapping between paper IDs and titles.

Paper ID	Title
P01	A Behavioral Conditional Diffusion Probabilistic Model for Human Motion Modeling in Multi-Action Mixed Human-Robot Collaboration [33]
P02	FineMLD: A Fine-Grained Motion Latent Diffusion for Human Motion Prediction in Human-Robot Collaboration [34]
P03	A Diffusion-Based Data Generator for Training Object Recognition Models in Ultra-Range Distance [35]
P04	Answerability Fields: Answerable Location Estimation via Diffusion Models [36]
P05	From Junk to Genius: Robotic Arms and AI Crafting Creative Designs from Scraps [37]
P06	Real-Time 3D Motion Prediction for Human-Robot Collaboration via Bayesian-Optimized Diffusion Models [38]
P07	Diffusion Co-Policy for Synergistic Human-Robot Collaborative Tasks [39]
P08	Diffusion-Based Imitation Learning for Social Pose Generation [40]
P09	Diverse Controllable Diffusion Policy With Signal Temporal Logic [41]
P10	Garment Diffusion Models for Robot-Assisted Dressing [42]
P11	Human-Robot Cooperative Distribution Coupling for Hamiltonian-Constrained Social Navigation [43]
P12	Legibility Diffuser: Offline Imitation for Intent Expressive Motion [44]
P13	REALM: Real-Time Estimates of Assistance for Learned Models in Human-Robot Interaction [45]
P14	SICNav-Diffusion: Safe and Interactive Crowd Navigation With Diffusion Trajectory Predictions [46]
P15	TransFusion: A Practical and Effective Transformer-Based Diffusion Model for 3D Human Motion Prediction [47]

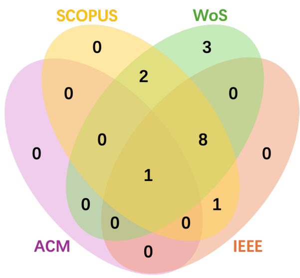

Fig. 4 The overlapping of the filtered search according to databases.

Table 3 Data extraction scheme for diffusion models in Human-Robot Interaction (HRI).

Data item	Description and relevance to RQs
Paper ID and reference	Unique identifier of each study (P01-P15) and full citation for traceability across analysis sections.
Publication year	Year of publication. Helps identify publication trends (RQ1).
Integration with HRI framework	How the diffusion model is incorporated into the HRI pipeline. Answers RQ1.
Type of diffusion model	Variants of diffusion model used (e.g., DDPM, latent diffusion). Addresses RQ2.
Evaluation	Type of evaluation and key metrics used. Addresses RQ3.
Benefits, limitations, and challenges	Reported advantages, limitations, and challenges. Relates to RQ4.

2.2.4 Data Synthesis

Following data extraction, a structured synthesis has been performed to integrate findings across the selected studies. Both quantitative descriptive analysis and qualitative thematic analysis are employed to comprehensively address the research questions.

Descriptive synthesis is used to identify trends in publication years, roles in HRI applications, and venues (RQ1). Comparative analysis examines diffusion model adaptations within HRI frameworks (RQ2). Evaluation methods and performance outcomes are synthesised to address RQ3. Finally, thematic analysis is conducted to identify recurring benefits, limitations, and open challenges (RQ4).

All synthesis results are presented using tables and figures aligned with each research question, ensuring clarity, coherence, and reproducibility.

3 Results

This section presents the results of synthesising the 15 studies included in this systematic literature review on diffusion models in human-robot interaction (HRI). The results are summarised across the listed questions (RQ1-RQ4), including publication trends and application domains (RQ1, Section 3.1), the types of diffusion models employed in HRI systems (RQ2, Section 3.2), methods and metrics (RQ3, Section 3.3), and key benefits, limitations, and open challenges (RQ4, Section 3.4).

3.1 Publication Trends and Application Domains

To address **RQ1**, we analyse the temporal distribution and application domains of the selected studies. As illustrated in Fig. 5, diffusion-based research in human-robot interaction has increased markedly since 2024, indicating a recent shift in the use of diffusion models from primarily perception-oriented domains toward interactive and decision-making robotic systems. The majority of the reviewed works are published in archival journals, suggesting increasing methodological maturity. Most studies appear in leading robotics venues, including *IEEE RA-L*, *ICRA*, and *IROS*, reflecting growing recognition of diffusion-based approaches within the core robotics community.

Despite this growth, only a subset of studies explicitly foregrounds human-centred interaction aspects, such as intent legibility, shared autonomy, or human-in-the-loop control (e.g., P12, P13, P05), indicating that diffusion-based HRI remains an emerging and still consolidating research direction.

Human-robot interaction inherently involves three tightly coupled elements: human behaviour, robot action, and the learned or engineered system that mediates between them. Following this perspective, we categorise the reviewed literature into three representative application domains, corresponding to human-centric, robot-centric, and system-centric applications of diffusion models. We classify studies as human-centric, robot-centric, or system-centric based on the primary target variable of the diffusion model and its role in the interaction loop. Human-centric studies primarily model future human behaviour; robot-centric studies generate robot actions or policies; system-centric studies support interaction indirectly through perceptual or representational augmentation. When multiple functional roles were present, classification was

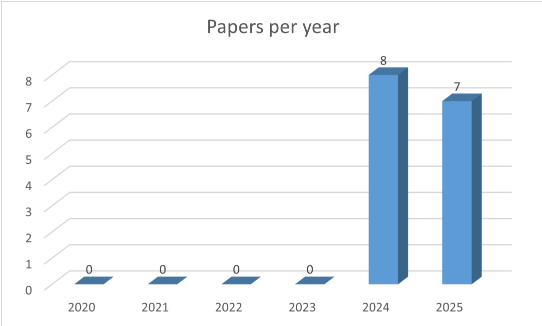

Fig. 5 Distribution of publication years for the selected studies.

determined by the primary evaluation objective and the system component for which the diffusion model was explicitly optimised.

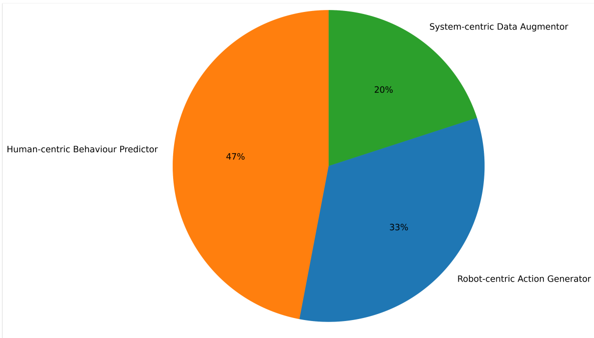

Fig. 6 Functional roles of diffusion models in HRI. The distribution reveals a dominant focus on human-centric behaviour prediction, followed by robot-centric action generation and system-centric data augmentation, reflecting current usage patterns of diffusion models across HRI applications.

As shown in Fig. 6, human motion and behaviour prediction constitutes one of the most common application domains of diffusion models in HRI. In this category, diffusion models are employed to generate multimodal distributions over future human motion conditioned on past observations, task context, or interaction state. Representative studies include behavioural motion modelling in collaborative scenarios (P01, P06), crowd-aware navigation and interaction modelling (P10, P11, P14), and long-horizon human motion prediction (P02, P15). Across these works, diffusion-based predictors incorporate interaction context such as robot motion, task phase, or environmental structure, and increasingly emphasise inference efficiency to meet the real-time requirements of safety-critical HRI systems.

Diffusion models are also applied to robot action generation and policy learning, particularly in scenarios that prioritise behavioural expressiveness, social alignment, or interpretability. In this domain, diffusion-based policies model distributions over robot actions rather than deterministic mappings, enabling the synthesis of diverse and intent-expressive behaviours. Representative examples include joint human-robot action generation for collaborative manipulation (P07), diffusion-based imitation learning for social pose generation (P08), policy generation under explicit temporal logic constraints (P09), and offline imitation learning for legible and intent-expressive robot motion (P12). These studies position diffusion models at the decision-making level, highlighting their use beyond purely predictive roles.

A complementary application domain leverages diffusion models for auxiliary data and representation generation in support of HRI systems. In perceptual settings, diffusion-based data generators are used to synthesise realistic training data under challenging sensing conditions, such as ultra-range visual perception for object and gesture recognition (P03). Diffusion models are also employed to generate spatial answerability maps that inform robot positioning and decision-making in embodied question-answering tasks (P04). In addition, a smaller but distinct subset of studies applies diffusion models to creative design and human-robot co-creation. In these works, diffusion-based generative models are integrated into robotic fabrication pipelines to transform intermediate outputs into expressive artefacts, with iterative human feedback shaping the final designs (P05). Although these approaches do not directly generate robot actions, they illustrate how diffusion models can support open-ended, human-centred interaction beyond conventional task-oriented robotics.

For ease of reference, we additionally provide a detailed mapping between application area and paper IDs in Table 4.

Table 4 Applications of diffusion models within HRI systems (RQ1).

Functional Role	**Description**	**Paper IDs**
Human-centric Behaviour Predictor	Predicts multimodal distributions over future human behaviour	P01, P02, P06, P10, P11, P14, P15
Robot-centric Action Generator	Directly generates robot actions or policies	P07, P08, P09, P12, P13
System-centric Data Augmentor	Generates perceptual data or creative artefacts	P03, P04, P05

3.2 Diffusion Model Types

To address **RQ2**, we analyse the choices of diffusion model variants that have been adopted across the reviewed studies. Based on thematic synthesis, we identify several recurring diffusion families used in human-robot interaction, as summarised in Table 5. Rather than introducing novel diffusion algorithms, most works adapt established diffusion variants to meet the practical requirements of interactive robotic systems.

Vanilla diffusion appears to be the dominant methodological choice as shown in Fig. 7, appearing in six studies (P03, P07, P08, P09, P10, P14). This prevalence suggests that, within the HRI community, the formulation of denoising diffusion probabilistic models [4] is still widely regarded as sufficiently expressive and reliable for modelling multimodal human behaviour and robot-related processes. In particular, works focusing on human motion prediction or behaviour modelling tend to prioritise methodological simplicity and stability over architectural optimisation, relying on vanilla diffusion as a robust baseline.

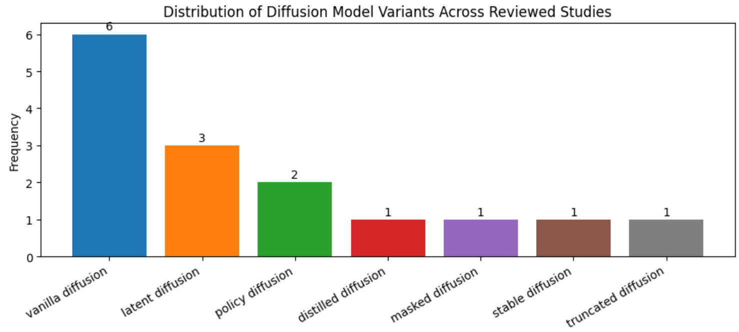

Fig. 7 Distribution of diffusion model variants across the reviewed HRI studies. Each bar corresponds to a distinct diffusion formulation, with colour encoding used to differentiate model variants. Vanilla diffusion remains the dominant choice, while latent and policy diffusion variants are increasingly explored.

A second prominent trend is the emerging adoption of *latent diffusion* models, which appear in three studies (P02, P04, P15). These approaches typically introduce a latent space to reduce computational cost and improve scalability, reflecting growing awareness of efficiency-related concerns in diffusion-based HRI systems [48]. Notably, latent diffusion is primarily employed in settings where high-dimensional inputs or longer prediction horizons are required, indicating that representational efficiency becomes increasingly important as diffusion models are pushed toward more complex interaction scenarios.

Policy diffusion recently emerges as a distinct methodological direction in two studies (P12, P13). Unlike vanilla or latent diffusion, which are mainly used for modelling distributions over future states or behaviours, policy diffusion [49] directly generates robot actions or control policies. This shift indicates a broader trend toward integrating diffusion models closer to decision-making and execution layers.

Other variants of diffusion models have also been explored in HRI. They are usually based on vanilla diffusion models and integrate additional operations to achieve task-oriented benefits. For example, *distilled diffusion* (P06) and *truncated diffusion* (P11) are two approaches employed for higher inference efficiency. To achieve controllability, *masked diffusion* introduces masks to diffusion models (P01) while *stable diffusion* has been integrated to HRI pipelines for its stability as an image generator (P05).

We additionally provide a detailed mapping between paper IDs and families of diffusion models in Table 5 for ease of reference.

Table 5 Diffusion model families and variants adopted in HRI studies (RQ2).

Diffusion Family / Variant	**Paper IDs**
Vanilla Diffusion	P03,P07,P08,P09,P10,P14
Latent Diffusion	P02, P04, P15
Policy Diffusion	P12, P13
Distilled Diffusion	P06
Masked Diffusion	P01
Stable Diffusion	P05
Truncated Diffusion	P11

3.3 Evaluation Methods and Metrics

To address **RQ3**, we analyse how diffusion-based HRI systems are evaluated across the reviewed studies, with a particular focus on the distribution and emphasis of evaluation metrics. While evaluation protocols vary across application domains and system roles, the reported metrics can be organised into six categories: *success-based*, *error-based*, *generative quality*, *efficiency*, *classification and retrieval*, and *qualitative* evaluation, as illustrated in Fig. 8.

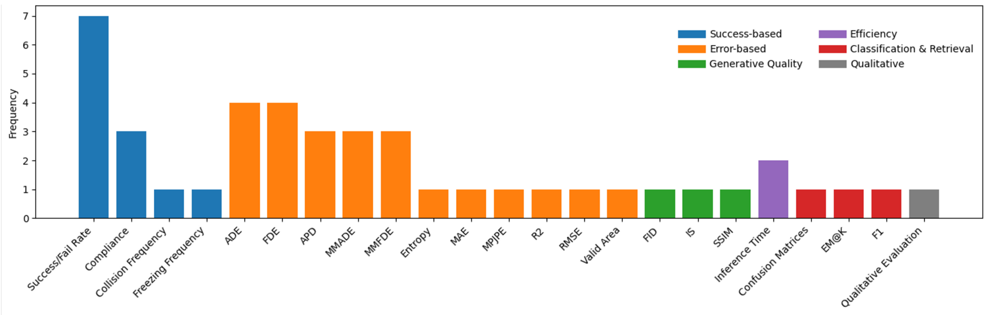

Fig. 8 Frequency distribution of evaluation metrics. Metrics are categorized into success-based, error-based, generative quality, efficiency, classification and retrieval, and qualitative evaluation, highlighting dominant evaluation trends in diffusion-based HRI studies.

Error-based metrics are the most widely adopted category across the reviewed literature. These metrics reflect an emphasis on how closely generated trajectories or behaviours match ground-truth observations. Standard trajectory error measures such as Average Displacement Error (ADE) and Final Displacement Error (FDE) appear most frequently (e.g., P02, P06, P14, P15). Additional error-oriented metrics, including APD, MMADE, MMFDE, MPJPE, and classical regression measures (MAE, RMSE, R2), are also employed to quantify prediction accuracy and distributional fidelity (e.g., P01). In addition to error-based metrics, some studies incorporate *generative quality* metrics to characterise the distributional properties of diffusion model outputs. These studies are usually within the context of data generation, and therefore, measures such as IS, FID, and SSIM are used to assess diversity, realism, or visual fidelity of generated samples (e.g., P03).

A second dominant class of evaluation concerns *success-based metrics*, which are particularly prevalent when diffusion models are integrated into closed-loop robotic systems. Metrics such as Success/Fail Rate, Compliance, Collision Frequency, and Freezing Frequency are used to assess task completion and conformity of constraints (e.g., P07, P09, P11, P14). Notably, Success/Fail Rate is the single most frequently reported metric overall, indicating a strong system-level emphasis on whether diffusion-based components enable successful task execution rather than solely accurate prediction. In contrast, *efficiency metrics* are comparatively rare. Inference Time is reported in only a small number of studies (P09, P14). This may arise from

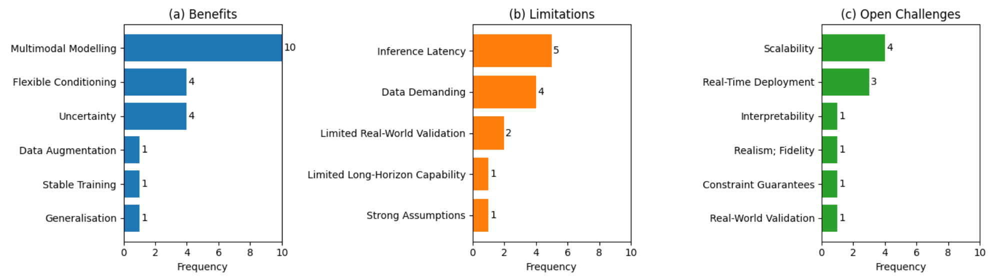

Fig. 9 Frequency distribution of reported benefits, limitations, and open challenges in diffusion-based HRI studies. Each subfigure shows the occurrence frequency of categories within benefits (a), limitations (b), and open challenges (c).

the typically high computational cost of diffusion models during inference. *Classification and retrieval metrics*, including F1 scores, EM@K, and confusion matrices, appear primarily in studies framing diffusion-based HRI as prediction, recognition, or retrieval problems (e.g., P01, P04, P13). *Qualitative evaluation* remains the least represented category. Only P05 relies on qualitative user feedback, subjective judgement, or preference-based evaluation, typically in creative interaction or co-creation scenarios.

We also present the detailed mapping between papers and their metrics in Table 6 for ease of reference.

Table 6 Evaluation metrics used across diffusion-based HRI studies.

Title	Metrics Used
P01	R2, MAE, RMSE, F1
P02	APD, ADE, FDE, MMADE, MMFDE
P03	IS, FID, SSIM, Success/Fail Rate
P04	EM@K
P05	Qualitative Evaluation
P06	APD, ADE, FDE, MMADE, MMFDE
P07	Success/Fail Rate
P08	MPJPE
P09	Success/Fail Rate, Compliance, Valid Area, Entropy, Inference Time
P10	Success/Fail Rate
P11	Success/Fail Rate, Compliance
P12	Success/Fail Rate, Compliance
P13	Confusion Matrices
P14	ADE, FDE, Success/Fail Rate, Inference Time, Collision Frequency, Freezing Frequency
P15	APD, ADE, FDE, MMADE, MMFDE

3.4 Benefits, Limitations, and Open Challenges

To address **RQ4**, we summarize the benefits, limitations, and open challenges that have been reported across the reviewed studies. Rather than interpreting their broader implications, this subsection consolidates recurring properties and constraints explicitly documented in the literature, grounded in evidence from the selected papers.

As shown in Fig. 9, diffusion models are most consistently valued for their ability to represent *multimodality*. In human motion prediction and navigation tasks, diffusion-based approaches are reported to generate diverse and plausible futures under ambiguous interaction contexts (P01, P02, P06, P14, P15). This capability supports anticipatory planning, safety-aware navigation, and robust interaction in the presence of behavioural variability.

A second recurring benefit concerns support for *flexible conditioning* and *uncertainty*. Diffusion-based policies and motion generators enable robots to produce socially aligned, intent-expressive, or creative behaviours in collaborative and human-facing tasks (P05, P07, P08, P12). Several studies further highlight the flexibility of diffusion models in conditioning on heterogeneous inputs, facilitating integration with perception, planning, and physical interaction modules (P04, P07, P10).

Despite these advantages, multiple limitations are reported across the literature. As shown in Fig. 9, *Inference latency* emerges as one of the most common constraints. This reflects the iterative nature of diffusion sampling and posing challenges for real-time deployment (P01, P07, P08, P09, P15). The second most common limitation is that diffusion models are data-intensive (P05, P08, P12). Diffusion-based HRI systems currently exhibit strong dependence on data with high annotation effort, which may not always be feasible. Another commonly mentioned limitation of these reviewed studies is the limited validation in real-world interaction settings (P05, P07). This indicates a consensus regarding the importance of real-world validation for identifying potential sim-to-real gaps. Other limitations include reduced performance in long-horizon prediction tasks (P12) or dependence on some assumptions (P13).

Building on these observations, the reviewed studies identify several common open challenges. As shown in Fig. 9, Improving *scalability* (P07, P08, P10, P12) and *real-time feasibility* (P01, P11, P15) remain the two most frequently reported challenges. This suggests a consensus within the research community that achieving effectiveness in large-scale scenarios or real-time response remains challenging and urgent to explore in the future. P02 also points to the need for further investigation on *interpretability*, in which probabilistic outputs should be understandable to humans or stakeholders to facilitate trustworthy artificial intelligence. Other challenges, including realism and fidelity, constraint guarantees, and real-world validation, are also raised by P03, P09, and P13 accordingly to pursue better performance, controllability, and ability under complex environments.

For ease of reference, we provide the detailed mapping between paper IDs and benefits, limitations, and challenges in Table 7.

4 Future Directions

Several open challenges arise from both the diffusion-model perspective and the HRI perspective when adapting this methodology to the HRI domain. Understanding and addressing these challenges are essential not only for situating diffusion-based approaches within the research field of human-robot interactions but also for identifying research directions that address similar adaptation scenarios. This section identifies and discusses these challenges, based on which this section then attempts to

Table 7 Summary of benefits, limitations, and challenges across reviewed diffusion-based HRI studies. "–" indicates that the aspect was not explicitly discussed by the authors.

ID	Benefits	Limitations	Open Challenges
P01	Multimodal Modelling; Uncertainty	Inference Latency	Real-Time Deployment
P02	Multimodal Modelling; Uncertainty	–	Interpretability
P03	Data Augmentation	–	Realism; Fidelity
P04	Flexible Conditioning	–	–
P05	Multimodal Modelling	Limited Real-World Validation; Data Demanding	–
P06	Stable Training; Multimodal Modelling	–	–
P07	Multimodal Modelling	Inference Latency; Limited Real-World Validation	Scalability
P08	Multimodal Modelling	Inference Latency; Data Demanding	Scalability
P09	Multimodal Modelling; Flexible Conditioning	Data Demanding; Inference Latency	Constraint Guarantees
P10	Generalisation	–	Scalability
P11	Uncertainty	–	Real-Time Deployment
P12	Multimodal Modelling; Flexible Conditioning	Data Demanding; Limited Long-Horizon Capability	Scalability
P13	Uncertainty	Strong Assumptions	Real-World Validation
P14	Multimodal Modelling; Flexible Conditioning	–	–
P15	Multimodal Modelling	Inference Latency	Real-Time Deployment

outline future research directions. Sections 4.1 and 4.2 focus on challenges from the diffusion-model side, whereas Sections 4.3 and 4.4 focus on HRI-related considerations.

4.1 Data Scarcity and Data Efficiency

Data scarcity emerges as a recurring challenge across the reviewed papers while diffusion models play two roles with respect to this challenge. On the one hand, diffusion models aggravate the scarcity of available data in human-robot interactions when they become an emerging solution. As discussed in Section 3.4 and summarised in Table 7, multiple diffusion-based HRI studies rely on limited, costly, or carefully curated human-robot interaction data (e.g., P08, P09, P12) to learn solutions. Due to the data-intensive nature of diffusion-based policies and behaviour generators, collecting datasets is expensive and may not always be possible when considering the diversity and dynamics of real-world environments. On the other hand, diffusion models have also been leveraged as a source of prior knowledge to mitigate data scarcity. As a generative model, large-scale pre-trained diffusion models from other research fields provide access to a diverse and high-fidelity data generator by synthesising additional training signals. They typically appear in vision-related tasks. For example, diffusion-based data generators have been used to produce high-quality synthetic perceptual data for ultra-range object recognition, substantially reducing the need for expensive real-world data collection (P03). In creative human-robot co-creation, offline-trained diffusion models enable iterative design exploration without requiring continuous human data input during deployment (P05). These examples illustrate how diffusion models can both depend on and partially alleviate limited human interaction data.

Despite the dual role of diffusion models, addressing data scarcity will be a fundamental problem in HRI. While using diffusion models as a generator seems to bypass the problem of data scarcity in HRI, this shifts the burden to advances in other research domains and relies on their advances to boost the development of HRI. One possibility to tackle this problem is to consider data-efficient architectures for diffusion models. For example, data-efficient training strategies, such as few-shot adaptation [50] and parameter-efficient fine-tuning [51], offer promising avenues for reducing data requirements while reusing pre-trained diffusion models. Another promising direction to explore is the physics-informed strategy [52]. Prior knowledge of physics or other related domains is integrated to the training process of diffusion models to significantly reduce the demand for data. Such successes may potentially be transferable to HRI.

4.2 Generalisation in Unseen Scenarios

Concerns regarding generalisation arise naturally from the reviewed diffusion-based HRI studies. As discussed in Section 3.3 and summarised in Table 6, the majority of works evaluate diffusion models under controlled settings that may not reflect the diversity and unpredictability of real-world interaction. Several reviewed papers explicitly or implicitly reveal this limitation. For example, studies on collaborative manipulation and policy learning report evaluations conducted with a small number of participants or constrained laboratory setups (P07, P08), while intent-expressive and legibility-focused approaches emphasise performance on curated demonstration datasets without extensive cross-task or cross-user testing (P12). In assistive decision-making, REALM (P13) evaluates uncertainty-aware assistance primarily in synthetic environments, leaving open questions about robustness under real human behaviour variability. Similarly, creative co-creation systems demonstrate compelling qualitative results but remain limited to controlled material and setting configurations (P05).

Addressing this challenge requires insights from diffusion-model research. On the one hand, recent empirical studies on diffusion model generalisation shed light on the problem. For instance, analyses based on mutual information suggest that diffusion models may overfit to spurious correlations when trained under limited variability, motivating alternative optimisation objectives that explicitly encourage generalisable representations [53]. Other works demonstrate that compositional and modular diffusion formulations can improve transfer to unseen configurations by recombining learned primitives or latent factors [11, 54–56]. Although these studies are not explicitly grounded in HRI, their success indicates potential pathways for diffusion-based interaction models to generalise across tasks, users, and environments.

On the other hand, recent theoretical analyses provide further insight into the generalisation properties of diffusion models. Studies examining the geometry and smoothness of diffusion-induced representation spaces offer formal arguments for when and why diffusion models can generalise beyond training distributions [57–60]. These results suggest that representation structure, noise scheduling, and conditioning mechanisms play a critical role in transferability. Translating such theoretical insights into the design of diffusion-based HRI systems remains an open but promising research direction.

4.3 Evaluation Beyond Model-Centric Metrics Towards Trust Construction

A key future direction for diffusion-based human-robot interaction lies in the development of evaluation frameworks that move beyond model-centric performance metrics toward human-centred assessment. The evaluation trends summarized in Section 3.3 and Table 6 show that current evaluation metrics in diffusion-based HRI are predominantly model-centric. They typically emphasize predictive accuracy, diversity, or task success. However, such measures provide only indirect insight into how probabilistic predictions affect human partners in real-world interaction [61]. These effects encompass not only explicit outcomes such as safety and coordination, but also more implicit factors including confidence, trust [22, 62, 63], and perceived autonomy [64–66].

Future research should, therefore, prioritise evaluation protocols that explicitly link diffusion-based modelling choices to human-centred outcomes. This includes the development of benchmarks and experimental setups in which the impact on human users can be isolated and analysed, rather than being subsumed into aggregate performance scores [67, 68]. Future diffusion-based HRI research should integrate human-centred metrics as core evaluation criteria. This entails larger-scale and more systematic studies that connect subjective measures, such as perceived safety, legibility, or confidence, to concrete modelling decisions and uncertainty representations [69]. Establishing such links is essential for aligning advances in diffusion-based modelling with HRI's emphasis on human experience, collaboration, and trust formation. Insights and methodologies from related domains, such as human motion generation [70] and human preference alignment [71], may provide valuable guidance in this regard.

4.4 From Uncertainty Representation to Uncertainty-Aware Design

The challenge of uncertainty-aware interaction design arises from the unpredictability of human behaviour and dynamic environments. As shown in Section 3.4, while multimodality and uncertainty representation are among the most consistently cited advantages of diffusion models (e.g., P01, P02, P06, P14, P15), only a small number of studies explicitly incorporate uncertainty into interaction-level decision-making (P11, P13). This discrepancy reveals a gap between probabilistic modelling capabilities and their practical use in HRI system design.

One promising research direction is related to the use of uncertainty to regulate autonomy and assistance. Effective uncertainty quantification techniques capture the range of plausible outcomes conditioned on observed data, providing critical insight into model confidence and robustness [72]. This capability is particularly important in high-stakes human-robot interaction scenarios, where quantifying predictive uncertainty can support risk-aware decision-making and enhance system safety [73]. Rather than being implicitly consumed by downstream planners, uncertainty estimates could directly inform when robots request human input, adapt their level of autonomy, or modulate interaction timing [74].

5 Conclusion

This paper presents a systematic literature review of diffusion-based approaches in human-robot interaction, synthesising 15 peer-reviewed studies through a structured, research-question-driven analysis. Adopting a system-level perspective beyond application-domain categorisation, the review examines how diffusion models are formulated, integrated, and evaluated within interactive robotic systems. The findings indicate that current diffusion-based HRI research is predominantly motion-centric and prediction-oriented, with diffusion models most often deployed as probabilistic forecasters within hybrid architectures that delegate execution and safety guarantees to classical planning and control components. Although diffusion models consistently demonstrate advantages in representing multimodality and uncertainty, these capabilities are only partially realised at the interaction level, as evaluation remains largely model-centric and human-centred assessment is limited.

While providing a systematic synthesis, several limitations should be noted. First, the scope of this survey emphasises interaction-relevant and human-aware system designs, focusing on diffusion-based approaches that explicitly model, predict, or adapt to human behaviour. As a result, some diffusion-based robotics studies that rely primarily on offline learning or optimisation without an explicit interaction context may fall outside the scope of this review. Second, the relatively small number of eligible studies reflects the early stage of diffusion-based HRI research and limits the extent to which broader trends can be generalised. The findings are exploratory but useful given that the field is rapidly emerging. Finally, as with most systematic literature reviews, the synthesis is constrained by the search strategy, inclusion criteria, and reporting practices of the reviewed works, which may affect the completeness of coverage and the granularity of comparative analysis.

Declarations

Funding

This research is supported in part by the joint Durham University and Singapore Management University Seedcorn Funding, and the EPSRC NorthFutures project (ref: EP/X031012/1).

Authors' contribution

Ziyi Chang: Conceptualization, Investigation, Writing – original draft. Yueyue Hou: Investigation, Writing – original draft. Jiannan Li: Conceptualization, Investigation, Supervision, Writing – review & editing. Ruochen Li: Conceptualization, Investigation, Writing – review & editing. Junyan Hu: Conceptualization, Writing – review & editing. Effie Lai-Chong Law: Writing – review & editing. Amir Atapour-Abarghouei: Supervision, Writing – review & editing. Hubert P. H. Shum: Conceptualization, Funding acquisition, Writing – review & editing, Supervision, Project administration.

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval and consent to participate

Not applicable

Consent for publication

All authors have agreed to publication.

Data availability

All data relevant to this study are referenced published studies.

Materials availability

Not applicable

Code availability

Not applicable

References

[1] Dahiya, A., Aroyo, A.M., Dautenhahn, K., Smith, S.L.: A survey of multi-agent human-robot interaction systems. Robotics and Autonomous Systems **161**, 104335 (2023)

[2] Goodrich, M.A., Schultz, A.C.: Human-robot interaction: A survey. Foundations and trends in human-computer interaction **1**(3), 203–275 (2008)

[3] Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: Proceedings of the International Conference on Machine Learning, pp. 2256–2265 (2015). PMLR

[4] Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)

[5] Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: Proceedings of the International Conference on Learning Representations (2021)

[6] Croitoru, F.-A., Hondru, V., Ionescu, R.T., Shah, M.: Diffusion models in vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(9), 10850–10869 (2023)

[7] He, C., Shen, Y., Fang, C., Xiao, F., Tang, L., Zhang, Y., Zuo, W., Guo, Z., Li, X.: Diffusion models in low-level vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(6), 4630–4651 (2025)

[8] Zhu, Y., Zhao, Y.: Diffusion models in NLP: A survey. arXiv preprint arXiv:2303.07576 (2023)

[9] Lin, L., Li, Z., Li, R., Li, X., Gao, J.: Diffusion models for time-series applications: A survey. Frontiers of Information Technology and Electronic Engineering **25**(1), 19–41 (2024)

[10] Chang, Z., Findlay, E.J.C., Zhang, H., Shum, H.P.H.: Unifying human motion synthesis and style transfer with denoising diffusion probabilistic models. In: Proceedings of the 2023 International Conference on Computer Graphics Theory and Applications, pp. 64–74 (2023)

[11] Chang, Z., Wang, H., Koulieris, G., Shum, H.P.H.: Large-scale multi-character interaction synthesis. In: Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers, pp. 1–10 (2025)

[12] Melnik, A., Ljubljanac, M., Lu, C., Yan, Q., Ren, W., Ritter, H.: Video diffusion models: A survey. Transactions on Machine Learning Research (2024)

[13] Xing, Z., Feng, Q., Chen, H., Dai, Q., Hu, H., Xu, H., Wu, Z., Jiang, Y.-G.: A survey on video diffusion models. ACM Computing Surveys **57**(2), 1–42 (2024)

[14] Zhang, C., Zhang, C., Zheng, S., Zhang, M., Qamar, M., Bae, S.-H., Kweon, I.S.: A survey on audio diffusion models: text to speech synthesis and enhancement in generative AI. arXiv preprint arXiv:2303.13336 (2023)

[15] Song, M., Deng, X., Zhou, Z., Wei, J., Guan, W., Nie, L.: A survey on diffusion policy for robotic manipulation: Taxonomy, analysis, and future directions. Authorea Preprints (2025)

[16] Wolf, R., Shi, Y., Liu, S., Rayyes, R.: Diffusion models for robotic manipulation: A survey. arXiv preprint arXiv:2504.08438 (2025)

[17] Urain, J., Mandlekar, A., Du, Y., Shafiullah, M., Xu, D., Fragkiadaki, K., Chalvatzaki, G., Peters, J.: Deep generative models in robotics: A survey on learning from multimodal demonstrations. arXiv preprint arXiv:2408.04380 (2024)

[18] Ubukata, T., Li, J., Tei, K.: Diffusion model for planning: A systematic literature review. arXiv preprint arXiv:2408.10266 (2024)

[19] Liu, X., Ma, K.Y., Gao, C., Shou, M.Z.: Diffusion Models in Robotics: A Survey. TechRxiv preprint (2025)

[20] Esterwood, C., Robert, L.P.: A systematic review of human and robot personality in health care human-robot interaction. Frontiers in Robotics and AI **8**, 748246 (2021)

[21] Ottoni, L.T.C., Cerqueira, J.d.J.F.: A systematic review of human-robot interaction: the use of emotions and the evaluation of their performance. International Journal of Social Robotics **16**(11), 2169–2188 (2024)

[22] Campagna, G., Rehm, M.: A systematic review of trust assessments in human-robot interaction. ACM Transactions on Human-Robot Interaction **14**(2), 1–35 (2025)

[23] Wang, X., Shen, L., Lee, L.-H.: A systematic review of xr-enabled remote human-robot interaction systems. ACM Computing Surveys **57**(11), 1–37 (2025)

[24] Ali, N.B., Usman, M.: Reliability of search in systematic reviews: Towards a quality assessment framework for the automated-search strategy. Information and Software Technology **99**, 133–147 (2018)

[25] Kitchenham, B.A., Budgen, D., Brereton, P.: Evidence-based Software Engineering and Systematic Reviews. CRC Press, ??? (2015)

[26] Pranckutė, R.: Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. Publications **9**(1), 12 (2021)

[27] Delgado López-Cózar, E., Orduña-Malea, E., Martín-Martín, A.: Google Scholar as a data source for research assessment. In: Springer Handbook of Science and Technology Indicators, pp. 95–127. Springer, ??? (2019)

[28] Martín-Martín, A., Thelwall, M., Orduna-Malea, E., Delgado López-Cózar, E.: Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. Scientometrics **126**(1), 871–906 (2021)

[29] Garzón, J., Acevedo, J.: Meta-analysis of the impact of augmented reality on students' learning gains. Educational Research Review **27**, 244–260 (2019)

[30] Cohen, J.: Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin **70**(4), 213 (1968)

[31] Viera, A.J., Garrett, J.M., *et al.*: Understanding interobserver agreement: the kappa statistic. Fam med **37**(5), 360–363 (2005)

[32] Law, E.L.-C., Heintz, M.: Augmented reality applications for k-12 education: A systematic review from the usability and user experience perspective. International Journal of Child-Computer Interaction **30**, 100321 (2021)

[33] Gui, H., Li, M., Yuan, Z.: A behavioral conditional diffusion probabilistic model

for human motion modeling in multi-action mixed human-robot collaboration. Advanced Engineering Informatics **62**, 102742 (2024)

[34] Zhong, R., Hu, B., Feng, Y., Liu, Z., Qin, Q., Wang, X.V., Wang, L., Tan, J.: FineMLD: A fine-grained motion latent diffusion for human motion prediction in human-robot collaboration. Advanced Engineering Informatics **70**, 104119 (2026)

[35] Bamani, E., Nissinman, E., Koenigsberg, L., Meir, I., Sintov, A.: A diffusion-based data generator for training object recognition models in ultra-range distance. IEEE Robotics and Automation Letters **9**(12), 11722–11729 (2024)

[36] Azuma, D., Miyanishi, T., Kurita, S., Sakamoto, K., Kawanabe, M.: Answerability fields: Answerable location estimation via diffusion models. In: Proceedings of the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 12932–12938 (2024)

[37] Liu, J., Chen, X., Yu, S.: From junk to genius: Robotic arms and AI crafting creative designs from scraps. Buildings **14**(12), 4076 (2024)

[38] Tian, S., Zheng, M., Liang, X.: Real-time 3D motion prediction for human-robot collaboration via bayesian-optimized diffusion models. IEEE Transactions on Automation Science and Engineering **22**, 23128–23140 (2025)

[39] Ng, E., Liu, Z., Kennedy, M.: Diffusion co-policy for synergistic human-robot collaborative tasks. IEEE Robotics and Automation Letters **9**(1), 215–222 (2023)

[40] Martin-Ozimek, A., Jayarathne, I., Mon, S.L., Chew, J.Y.: Diffusion-based imitation learning for social pose generation. In: Proceedings of the 20th ACM/IEEE International Conference on Human-Robot Interaction, pp. 1488–1492 (2025)

[41] Meng, Y., Fan, C.: Diverse controllable diffusion policy with signal temporal logic. IEEE Robotics and Automation Letters **9**(10), 8354–8361 (2024)

[42] Kotsovolis, S., Demiris, Y.: Garment diffusion models for robot-assisted dressing. IEEE Robotics and Automation Letters **10**(2), 1217–1224 (2024)

[43] Wang, W., Yu, C., Wang, Y., Min, B.-C.: Human-robot cooperative distribution coupling for hamiltonian-constrained social navigation. In: Proceedings of the 2025 IEEE International Conference on Robotics and Automation, pp. 10808–10815 (2025)

[44] Bronars, M., Cheng, S., Xu, D.: Legibility diffuser: Offline imitation for intent expressive motion. IEEE Robotics and Automation Letters **9**(11), 10161–10168 (2024)

[45] Hagenow, M., Shah, J.A.: Realm: Real-time estimates of assistance for learned models in human-robot interaction. IEEE Robotics and Automation Letters

10(6), 5473–5480 (2025)

[46] Samavi, S., Lem, A., Sato, F., Chen, S., Gu, Q., Yano, K., Schoellig, A.P., Shkurti, F.: Sicnav-diffusion: Safe and interactive crowd navigation with diffusion trajectory predictions. IEEE Robotics and Automation Letters **10**(9), 8738–8745 (2025)

[47] Tian, S., Zheng, M., Liang, X.: Transfusion: A practical and effective transformer-based diffusion model for 3D human motion prediction. IEEE Robotics and Automation Letters **9**(7), 6232–6239 (2024)

[48] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695 (2022)

[49] Chi, C., Xu, Z., Feng, S., Cousineau, E., Du, Y., Burchfiel, B., Tedrake, R., Song, S.: Diffusion policy: Visuomotor policy learning via action diffusion. The International Journal of Robotics Research **44**(10-11), 1684–1704 (2025)

[50] Liu, F., Zhang, T., Dai, W., Zhang, C., Cai, W., Zhou, X., Chen, D.: Few-shot adaptation of multi-modal foundation models: A survey. Artificial Intelligence Review **57**(10), 268 (2024)

[51] Xu, L., Xie, H., Qin, S.J., Tao, X., Wang, F.L.: Parameter-efficient fine-tuning methods for pretrained language models: A critical review and assessment. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–20 (2026)

[52] Bastek, J.-H., Sun, W., Kochmann, D.: Physics-informed diffusion models. In: Proceedings of the Thirteenth International Conference on Learning Representations (2025)

[53] Yi, M., Sun, J., Li, Z.: On the generalization of diffusion model. arXiv preprint arXiv:2305.14712 (2023)

[54] Ota, S., Yu, Q., Fujiwara, K., Ikehata, S., Sato, I.: Pino: Person-interaction noise optimization for long-duration and customizable motion generation of arbitrary-sized groups. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10676–10685 (2025)

[55] Cheng, S., Yin, X., Zhou, R., Li, Y., Wang, X., Pan, L., Wang, W.Y., Zhong, V.: From atomic to composite: Reinforcement learning enables generalization in complementary reasoning. arXiv preprint arXiv:2512.01970 (2025)

[56] Kong, L., Xie, S., Jiao, Y., Chen, Y., Guo, Y., Shao, S., Gao, Y., Chen, G., Zhang, K.: Learning by analogy: A causal framework for composition generalization. arXiv preprint arXiv:2512.10669 (2025)

[57] Kadkhodaie, Z., Guth, F., Simoncelli, E.P., Mallat, S.: Generalization in diffusion models arises from geometry-adaptive harmonic representations. In: Proceedings of the Twelfth International Conference on Learning Representations (2024)

[58] Zhang, Z., Li, X., Li, X., Shi, L., Wu, M., Tao, M., Qu, Q.: Generalization of diffusion models arises with a balanced representation space. arXiv preprint arXiv:2512.20963 (2025)

[59] Ramachandran, S.N., Mukhopadhyay, R., Agarwal, M., Jawahar, C., Namboodiri, V.: Understanding the generalization of pretrained diffusion models on out-of-distribution data. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 14767–14775 (2024)

[60] Li, P., Li, Z., Zhang, H., Bian, J.: On the generalization properties of diffusion models. Advances in Neural Information Processing Systems **36**, 2097–2127 (2023)

[61] Chauhan, H., Pakbaz, A., Jang, Y., Jeong, I.: Analyzing trust dynamics in human-robot collaboration through psychophysiological responses in an immersive virtual construction environment. Journal of Computing in Civil Engineering **38**(4), 04024017 (2024)

[62] Alhaji, B., Büttner, S., Sanjay Kumar, S., Prilla, M.: Trust dynamics in human interaction with an industrial robot. Behaviour & Information Technology **44**(2), 266–288 (2025)

[63] Guo, X., Liu, Y., Ma, X., Fu, H.: Human trust effect in remote human-robot collaboration construction task for different level of automation. Advanced Engineering Informatics **68**, 103647 (2025)

[64] Glawe, F., Schmeckel, T., Brauner, P., Ziefle, M.: Human autonomy and sense of agency in human-robot interaction: A systematic literature review. arXiv preprint arXiv:2509.22271 (2025)

[65] Junior, W.T.L., Welter, R.A., Ferreira, W.P., Souza, R.F., Eduardo, T.: Human robot interaction (hri): An artificial cognitive autonomy approach to enhance decision-making. Cognitive Systems Research **91**, 101336 (2025)

[66] Van Dijk, W., Baltrusch, S.J., Dessers, E., De Looze, M.P.: The effect of human autonomy and robot work pace on perceived workload in human-robot collaborative assembly work. Frontiers in Robotics and AI **10**, 1244656 (2023)

[67] Sheng, J., Lin, M., Zhao, A., Pruvost, K., Wen, Y.-H., Li, Y., Huang, G., Liu, Y.-J.: Exploring text-to-motion generation with human preference. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1888–1899 (2024)

[68] Ismail-Fawaz, A., Devanne, M., Berretti, S., Weber, J., Forestier, G.: Establishing a unified evaluation framework for human motion generation: A comparative analysis of metrics. Computer Vision and Image Understanding **254**, 104337 (2025)

[69] Voas, J., Wang, Y., Huang, Q., Mooney, R.: What is the best automated metric for text to motion generation? In: SIGGRAPH Asia 2023 Conference Papers, pp. 1–11 (2023)

[70] Wang, H., Zhu, W., Miao, L., Xu, Y., Gao, F., Tian, Q., Wang, Y.: Aligning human motion generation with human perceptions. In: Proceedings of the Thirteenth International Conference on Learning Representations (2025)

[71] Wang, Y., Zhong, W., Li, L., Mi, F., Zeng, X., Huang, W., Shang, L., Jiang, X., Liu, Q.: Aligning large language models with human: A survey. arXiv preprint arXiv:2307.12966 (2023)

[72] Jazbec, M., Wong-Toi, E., Xia, G., Zhang, D., Nalisnick, E., Mandt, S.: Generative uncertainty in diffusion models. In: Proceedings of the Forty-First Conference on Uncertainty in Artificial Intelligence, pp. 1837–1858 (2025)

[73] De Vita, M., Belagiannis, V.: Diffusion model guided sampling with pixel-wise aleatoric uncertainty estimation. In: Proceedings of the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3844–3854 (2025). IEEE

[74] Sun, J., Jiang, Y., Qiu, J., Nobel, P., Kochenderfer, M.J., Schwager, M.: Conformal prediction for uncertainty-aware planning with diffusion dynamics model. Advances in Neural Information Processing Systems **36**, 80324–80337 (2023)