RLSLM: A Hybrid Framework Combining Reinforcement Learning and a Rule-based Social Locomotion Model for Socially-aware Navigation

Yitian Kou[1*], **Yihe Gu**[2*], **Chen Zhou**[2, 3*], **Dandan Zhu**[1†], **Shu-Guang Kuai**[2, 4, 5†]

[1]School of Computer Science and Technology, East China Normal University.
[2]Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, East China Normal University.
[3]School of Psychology and Neuroscience, University of Glasgow.
[4]NYU-ECNU Institute of Brain and Cognitive Science.
[5]Shanghai Center for Brain Science and Brain-Inspired Technology.
{yitian.kou, yhgu}@stu.ecnu.edu.cn, chen.zhou@glasgow.ac.uk, ddzhu@mail.ecnu.edu.cn, sgkuai@psy.ecnu.edu.cn

Abstract

Navigating human-populated environments without causing discomfort is a critical capability for socially-aware agents. While rule-based approaches offer interpretability through predefined psychological principles, they often lack generalizability and flexibility. Conversely, data-driven methods can learn complex behaviors from large-scale datasets, but are typically inefficient, opaque, and difficult to align with human intuitions. To bridge this gap, we propose **RLSLM**, a hybrid **R**einforcement **L**earning framework that integrates a rule-based **S**ocial **L**ocomotion **M**odel, grounded in empirical behavioral experiments, into its reward function. The social locomotion model generates an orientation-sensitive social comfort field that quantifies human comfort across space, enabling socially aligned navigation policies with minimal training. RLSLM then jointly optimizes mechanical energy and social comfort, allowing agents to avoid intrusions into personal or group space. A human-agent interaction experiment using an immersive VR-based setup demonstrates that RLSLM outperforms state-of-the-art rule-based models in user experience. Ablation and sensitivity analyses further show the model's significantly improved interpretability over conventional data-driven methods. This work presents a scalable, human-centered methodology that effectively integrates cognitive science and machine learning for real-world social navigation.

Introduction

Moving around human-populated environments without causing discomfort is an essential requirement for social agents, since they are widely engaged in human-agent interaction (Sheridan 2016). Such socially-aware navigation entails consideration of multiple social factors and remains a highly challenging problem (Francis et al. 2025).

Existing work on socially-aware navigation can be broadly classified into two categories, rule-based and data-driven. Rule-based approaches typically adopt models with identified variables and interpretable, quantifiable principles, like proxemics (Chen, Zhang, and Zou 2018) and velocity (Kim et al. 2015), either derived from social psychology or manually designed. Although these models show strength in interpretability and low computational overhead, they are often (1) difficult to quantify precisely, (2) limited in generalizability across environments, and (3) less flexible, which may lead to unnatural behaviors like oscillatory paths (Chen et al. 2018), ultimately constraining their real-world applicability.

Meanwhile, data-driven methods, such as reinforcement learning (RL) (Wang et al. 2024) and imitation learning (Karnan et al. 2022), have enabled agents to emulate human navigation behaviors based on large-scale human trajectory datasets (Lerner, Chrysanthou, and Lischinski 2007; Pellegrini et al. 2009; Rudenko et al. 2020; Bae et al. 2023) or simulation environments (Kapoor et al. 2023; Tsoi et al. 2020; Vuong et al. 2024). Although these approaches have achieved promising results, they are (1) highly dependent on the quality of the dataset, (2) expensive to train, and (3) often lack interpretability or alignment with human intuitions. With insufficient prior knowledge to guide the training, data-driven methods are often inefficient and prone to pitfalls.

Therefore, an important question arises: Can these two approaches be integrated to develop models that are efficient, adaptable, and interpretable, while remaining aligned with real-world human social behavior? To address this, we propose RLSLM, a hybrid framework that integrates a computational social locomotion model derived from psychological research (Zhou et al. 2022) into the reward structure of an RL agent. Based on well-controlled behavioral experiments, the rule-based social locomotion model computes an orientation-sensitive discomfort field that covers the entire navigation area, with higher field values indicating greater mutual discomfort between the agent and other humans when passing that point. By incorporating this rule-based model into a multi-objective RL framework to jointly minimize mechanical energy and social discomfort, we enable the agent to learn complex socially aligned rules within a small number of training epochs, such as avoiding invasion

*These authors contributed equally.
†Corresponding Author.

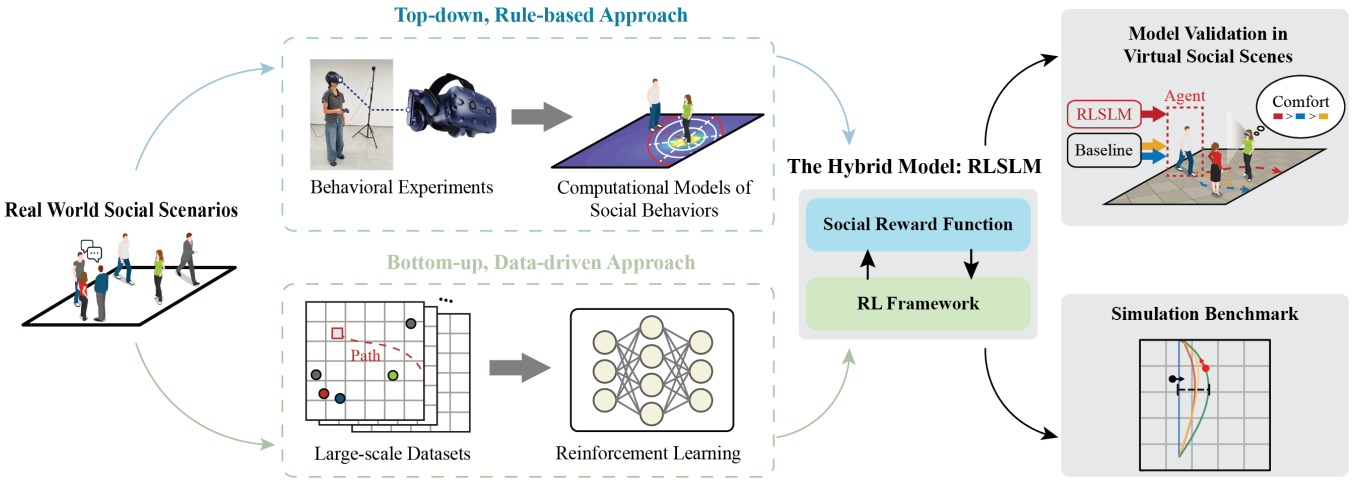

Figure 1: Methodology overview. The hybrid model of RLSLM combines a top-down, rule-based approach which develops computational models of human social behaviors from well-controlled lab experiments and a bottom-up data-driven approach which formulates the reinforcement learning framework based on large-scale social scenarios. The hybrid model first encodes human behavioral patterns into a social reward function, which is then used to train the policy within a reinforcement learning framework. The trained model is subsequently validated through human-agent interaction studies and simulations.

of personal space and social groups.

We further compared RLSLM with two rule-based models using human comfort ratings. The results demonstrate that our framework significantly outperforms these baselines in terms of user comfort.

In summary, this work contributes:

- **A novel hybrid RL framework** that integrates a psychologically grounded social locomotion model into reinforcement learning, combining the interpretability and prior knowledge of rule-based methods with the adaptability and expressiveness of data-driven approaches. This framework is potentially generalizable and applicable in other scenarios with similarly scarce data.
- **Performance improvement in user comfort**: RLSLM achieves a mean comfort rating of 4.21/5, significantly outperforming the best rule-based baseline (Δrating $=$ 1.12, Bonferroni-corrected post-hoc comparisons, $P <$ 0.001). This establishes a new Pareto frontier in the trade-off between comfort and efficiency.

Related Work

Incorporating Social Rules in Navigation

Recent studies have explored the incorporation of social rules into navigation algorithms. The design of social rule modules is mostly driven by intuition, dataset statistics, or physical modeling of human path planning. Static properties like the proper radius of personal space are often determined by intuition and experience in previous studies (Gong et al. 2025). Qualitative navigation decisions (e.g. passing on the left or right when encountering others) and trajectory features that facilitate path prediction can be learned from real-world pedestrian datasets (Yang et al. 2024). To support dynamic path planning in human-populated scenarios and avoid collision, physically based models like the social force model have been developed to simulate particle-like motion of the crowd (Helbing and Molnar 1995; Shiomi et al. 2014), often corresponding to intuitive geometric relations instead of real pedestrians' movements (Chen et al. 2018). In conclusion, although a great number of navigation studies have taken social rules into account, most of them are not quantitatively grounded in human behavioral experiments, which can lead to the generation of unnatural paths (Karwowski, Szynkiewicz, and Niewiadomska-Szynkiewicz 2024). Notably, a recent study determined socially-aware parameters (e.g. neighbor distance) through a user experiment, in which participants are presented with simulated navigation videos and asked to report their perceived social comfort (Bera et al. 2018). However, this approach fails to account for the subtle characteristics of path planning, and third-person user studies may yield unrealistic feedback due to issues of ecological validity. To address the aforementioned limitations, we propose a hybrid framework that directly embeds findings from cutting-edge psychological research into training workflows, and test our model performance in an immersive VR experiment with high ecological validity.

Learning-Based Approaches to Social Navigation

Deep learning has significantly advanced trajectory forecasting by enabling data-driven modeling of complex social interactions. GAN-based methods such as SocialGAN (Gupta et al. 2018) and SoPhie (Sadeghian et al. 2019) capture multimodal behaviors by generating diverse plausible futures. Graph-based models like Social-STGCNN (Mohamed et al. 2020) leverage GNNs to model spatial interactions among pedestrians. Transformer-based models like STAR (Yu et al. 2020) and STPOTR (Mahdavian et al. 2022) have further improved performance through long-range temporal modeling and attention-based interaction. These models implicitly learn conventions from data, but offer limited interpretabil-

ity and little control over compliance with social rules or physical rules. RL provides a more natural and interpretable framework for integrating such rules into trajectory generation. To our knowledge, no prior work leverages reinforcement learning to explicitly encode and optimize human-centered social navigation constraints. We bridge this gap by casting navigation as reward-driven optimization with interpretable social influence modeling.

Method

In this section, we introduce a reinforcement learning framework for socially-aware navigation, which integrates three components: an environment observation module to capture agent states and social cues, an actor-critic network for effective action selection, and a multi-dimensional feedback mechanism.

As illustrated in Figure 2, the agent's decision-making pipeline is composed of three core components: environment observation, action selection, and policy updating. In the following subsections, we describe each module in detail and explain how the feedback signals are formulated to promote socially compliant and efficient navigation behavior.

Agent Decision-Making Process

The agent's behavior is guided by a three-stage decision-making process: perceiving the environment, selecting actions based on the learned policy, and continuously updating the policy through reinforcement learning.

Environment Observation The environment observation module captures the agent's perception of the surroundings. Similar to how humans perceive social environments, the agent's observation space includes not only its own position, but also the relative positions and orientations of surrounding individuals. These features are concatenated into a structured input vector s_t. For a scenario involving n individuals, the observation vector at timestep t is represented as $s_t \in R^{3n+2}$.

Action Selection To learn a policy that guides the agent's decision-making, we adopt a deep reinforcement learning framework based on an actor-critic architecture. Specifically, the policy is represented by a stochastic function $\pi(a_t|s_t)$, which denotes the probability of selecting action a_t given the current state s_t. The actor network models the policy $\pi(a_t|s_t)$ and generates a distribution over possible actions, supporting a trade-off between exploration and exploitation. Concurrently, the critic network estimates the value function $V(s_t)$, which predicts the expected return from state s_t. This actor-critic architecture allows the agent to optimize its policy through trial-and-error interaction with the environment, gradually improving its navigation performance in socially complex settings.

Policy Updating At each timestep, the agent observes the current state and selects an action accordingly. The output action a_t represents a navigation command indicating a movement direction. The agent then executes this action, causing the environment to transition to a new state s_{t+1}, and receives a reward signal for learning. We utilize the Advantage Actor Critic algorithm (A2C) (Mnih et al. 2016) to jointly train the actor and critic. The actor is updated to maximize the expected return by increasing the probability of actions with high advantage estimates, while the critic is trained to minimize the temporal difference (TD) error between successive value predictions. Through continuous interactions with the environment and iterative policy updates, the agent learns to generate socially compliant and energy-efficient behaviors that achieve the task objective.

Multi-dimensional Feedback Mechanism

During social navigation, approaching other individuals in the scene tends to cause greater discomfort (i.e., experiencing social influence from others), while taking a detour from them consumes more mechanical energy. To balance social influence and mechanical energy consumption and emulate human-like path planning behavior, we design a multi-dimensional feedback mechanism that provides rewards and penalties based on three key factors: mechanical energy expenditure, progress toward the goal and social influence.

Mechanical Energy We assume that the agent moves a fixed distance l at each timestep, and does not consider the mechanical energy required for turning, which implies that the mechanical energy expenditure per step remains constant. To penalize excessive energy usage and encourage efficient motion, we introduce a negative reward component R_e defined as:

$$R_e(s_t) = -\alpha, \tag{1}$$

where α is a constant representing the estimated mechanical energy consumed per step. This term ensures that the agent is incentivized to reach the goal using the minimal number of steps, thereby promoting energy-efficient navigation.

Goal Progress To encourage the agent to make progress toward the goal, we introduce a positive reward component R_d, which is proportional to the reduction in distance between consecutive timesteps:

$$R_d(s_t, s_{t-1}) = \frac{D_{t-1} - D_t}{l}, \tag{2}$$

where D_{t-1} and D_t represent the distances from the agent to the destination at the previous and current timesteps, respectively.

Social Influence To encourage socially-aware behavior in the agent, we incorporate insights from prior psychological research (Zhou et al. 2022) to quantify social influence. Based on results from behavioral experiments, the social influence of each individual on surrounding space is modeled as an orientation-sensitive field, with higher field values indicating greater discomfort. The field model comprises three components: a heading-relevant social component (HRSC), a heading-irrelevant social component (HISC), and a collision avoidance component (CAC). Given the relative locations and orientations of the agent and surrounding persons, the joint influence (with persons indexed by $k \in \{1, \ldots, n\}$) R_s imposed on the agent at s_t can be computed as follows:

$$R_s(s_t) = \sum_{k=1}^{n} F'_k, \tag{3}$$

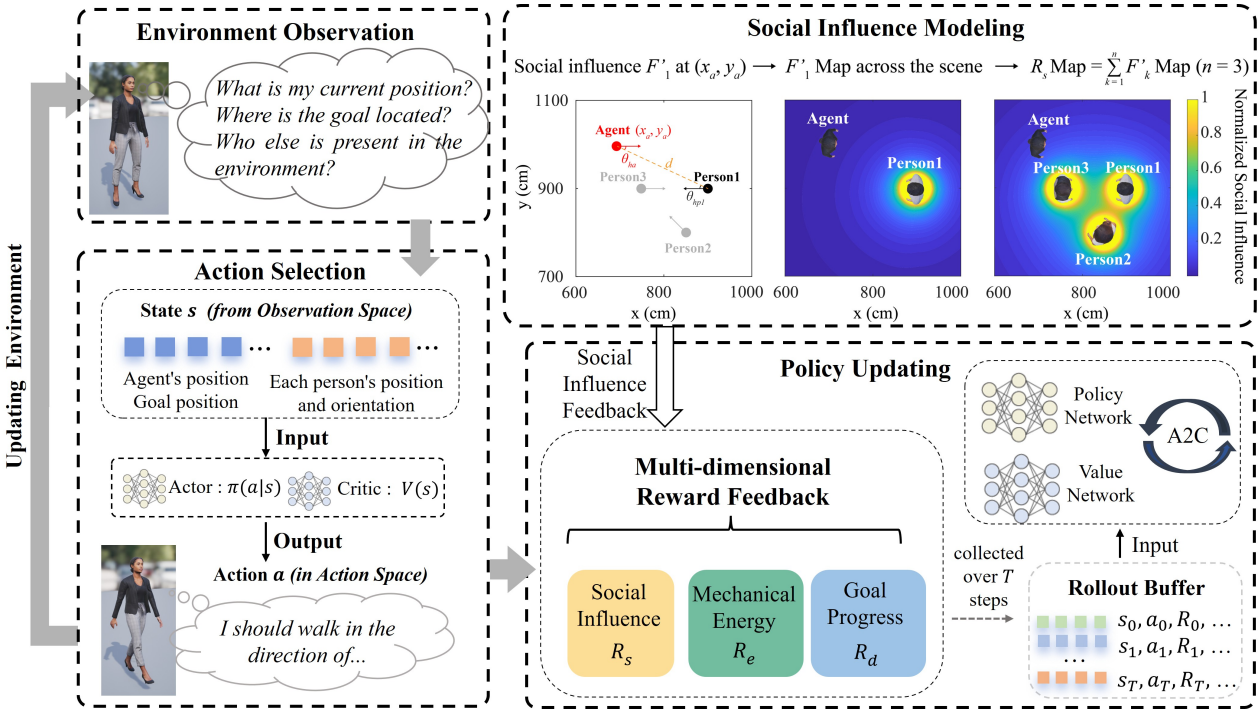

Figure 2: Overview of RLSLM framework. RLSLM integrates social influence modeling with reinforcement learning to guide an agent's movement in environments shared with humans. The framework follows a three-stage decision-making loop (gray arrow), and once the environment is updated based on the agent's action, the cycle begins again with a new observation.

$$F' = \min\left(\frac{F}{K},\ 1\right), \tag{4}$$

$$F = \frac{I_{\text{agent}} \times I_{\text{person}}}{d^2}, \tag{5}$$

$$I_{\text{human}} = m \times f(\theta_h) + n + c \times I_{\text{CA}},$$
$$\text{where} \quad f(\theta_h) = \begin{cases} \cos(\theta_h), & \cos(\theta_h) \geq 0 \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

$$I_{\text{CA}} = \frac{ab}{\sqrt{a^2\cos^2(\theta) + b^2\sin^2(\theta)}}, \tag{7}$$

where F represents the original social influence field value and d represents the distance between the agent and person$_k$ at s_t. The individual social influence of the agent and person$_k$, I_{agent} and I_{person}, is calculated using I_{human} with different fitted parameters m and n (for the agent, $m_a = 0.321$, $n_a = 0.856$; for the person$_k$, $m_p = 0.438$, $n_p = 0.630$). Specifically, m represents the contribution of HRSC and n represents HISC. θ_h represents the angle between the facing direction and the line connecting agent–person$_k$. I_{CA} represents CAC, with a and b ($a = 0.285$, $b = 0.175$) estimated by measuring the average cross-section of human body (approximated as an ellipse), and θ represents the angle between the line connecting agent–person$_k$ and the long arm of the ellipse. The free parameter c adjusted the relative ratio ($c = 1.430$) of I_{CA}. To prevent extreme F values, original social influence F is standardized with an upper limit K ($K = 10.180$) fitted by behavioral data. The values of all these parameters are set according to the prior work by Zhou et al.

If the agent's distance to the endpoint is less than a predefined threshold, it is deemed to have reached the destination. At the terminal timestep $t = T$, the agent receives a reward $r_T = +C$ if it successfully reaches the destination, or $r_T = -C$ otherwise (exceeding step limit or moving out of bounds). Overall, the return G is calculated as the sum of discounted rewards:

$$G = \sum_{t=0}^{T} \gamma r_t, \tag{8}$$

where γ is the discount factor, and r_t is the reward at timestep t, which is composed of three components and a terminal reward, defined as follows:

$$r_t = \begin{cases} R_d(s_t, s_{t-1}) + R_e(s_t) + \sigma R_s(s_t), & \text{for } 0 < t < T \\ \pm C, & \text{for } t = T \end{cases} \tag{9}$$

where σ denotes the weight of social influence. In our experiments, the parameters are set as follows: $\gamma = 0.9$, $\sigma = 0.5$, $C = 500$, and $\alpha = 1$. Additional training details are available in the supplementary material.

Dataset Creation

To support reproducible research and facilitate benchmarking in socially-aware navigation, we establish a VR-based

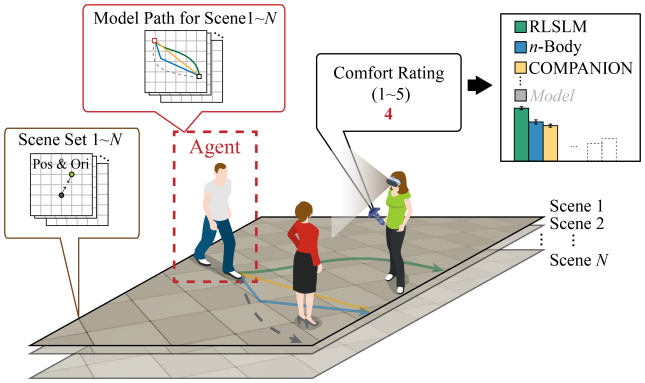

Figure 3: Overview of our VR-based user evaluation pipeline. (1) For each scenario, we import a set of human layouts (positions and orientations) and corresponding navigation trajectories generated by different models. (2) The participant views these simulated interactions in an immersive first-person VR environment, observing the agent's movement among virtual humans. (3) After each trial, the participant provides a comfort rating (1–5), which is recorded and aggregated across models for quantitative comparison.

human-agent interaction dataset. This dataset is designed as a benchmark environment for evaluating social comfort under controlled yet immersive conditions. Our dataset comprises a diverse set of simulated VR scenarios featuring varied human placements and orientations. The environment is implemented using Unreal Engine 5.4 and supports a variable number of virtual humans with configurable positions and orientations, covering a range of common social patterns such as face-to-face blocking, group passage, and asymmetric crowd formations. The dataset also includes user annotations on social comfort collected through immersive VR experiments.

The process of constructing our immersive evaluation dataset is illustrated in Figure 3. For each scene, we define static human layouts and import precomputed model trajectories. Participants then experience these scenarios in VR and rate the agent's navigation behavior. Implementation details and code are provided in the supplementary material. To facilitate reproducibility and further research, we provide public access to both the dataset and the VR evaluation pipeline. This dataset can also serve as a reusable benchmark for future studies on human-centered navigation evaluation.

Experiments

Experimental Setup

To assess the generalizability of the model, we perform evaluations in both single- and multi-human scenarios. In single-human scenarios, we position the human near the straight line connecting the start and target points to observe the agent's avoidance behavior when encountering one human. In multi-human scenarios, we place three individuals in the scene, with some of them engaging in social interaction, such as facing each other, to examine whether the agent respects social formations or intrusively passes between interacting individuals. The experiment is conducted within a confined $15m \times 15m$ virtual environment. The agent's step length is fixed at $45cm$. At each decision step, following an observation of the environment, the agent selects a movement direction and advances exactly one step in that direction. The episode is considered successful when the Euclidean distance between the agent and the target falls below the length threshold of one step. Due to the discrete stepwise movements of the agent, the resulting trajectories tend to exhibit discontinuities and jaggedness. To enhance spatial continuity, we post-process the trajectories using a Gaussian smoothing filter to improve path smoothness.

Following prior work (Sivashangaran, Khairnar, and Eskandarian 2023), we construct our virtual environment using OpenAI Gymnasium (Towers et al. 2024) and employ Stable-Baselines3 (Raffin et al. 2021) for policy learning. Comprehensive training configurations and convergence plots are presented in the supplementary material.

Human-Agent Interaction Experiment in VR

To test whether RLSLM aligns better with user experience in human-agent interaction than baseline models, we conduct a VR-based experiment in which participants are asked to rate their comfort level when interacting with the virtual agents controlled by one of three navigation algorithms: RLSLM, COMPANION (Kirby, Simmons, and Forlizzi 2009), and n-Body (Van Den Berg et al. 2011).

Participants A total of 30 university students and staff (11 males and 19 females aged between 18 and 29) are recruited to participate in this study. All participants have normal or corrected-to-normal vision.

Procedure We randomly set 50 scenarios in which a virtual agent bypasses one or three static persons along a detour path generated by one of three navigation algorithms, resulting in 150 trials. Notably, one of the multi-human scenarios is found to be repetitive and is thus excluded in the following analysis. This exclusion has no impact on statistical conclusions (a detailed comparison of statistical results is available in the supplementary material). As shown in Figure 4 (a), participants experience these scenarios as one of the static persons via an HTC Vive Pro head-mounted display (HTC Corporation; binocular resolution: 2,880×1,600 pixels; refresh rate: 90 Hz; field of view: 110°). In each scenario, participants stand at the designated location and orientation of a randomly selected static person, while the virtual agent walks from the start point to the target point along the pre-generated path. Upon completion of the navigation, participants rate their comfort level (on a scale from 1 to 5, with 5 indicating maximum comfort) using a handheld controller.

Result Analysis Figure 4 (b) and (c) illustrate the trajectories produced by three models. Figure 4 (d) presents a detailed analysis of user evaluation data. A significant main effect of model type on comfort level is present (repeated-measures ANOVA, $F_{(2,\,58)} = 219.589$, $P < 0.001$, $\eta_G^2 = 0.525$). Rating scores of paths generated by RLSLM are significantly higher than those generated by COMPANION (both in single-human and multi-human scenarios, Bonferroni-corrected post-hoc comparisons, $P < 0.001$)

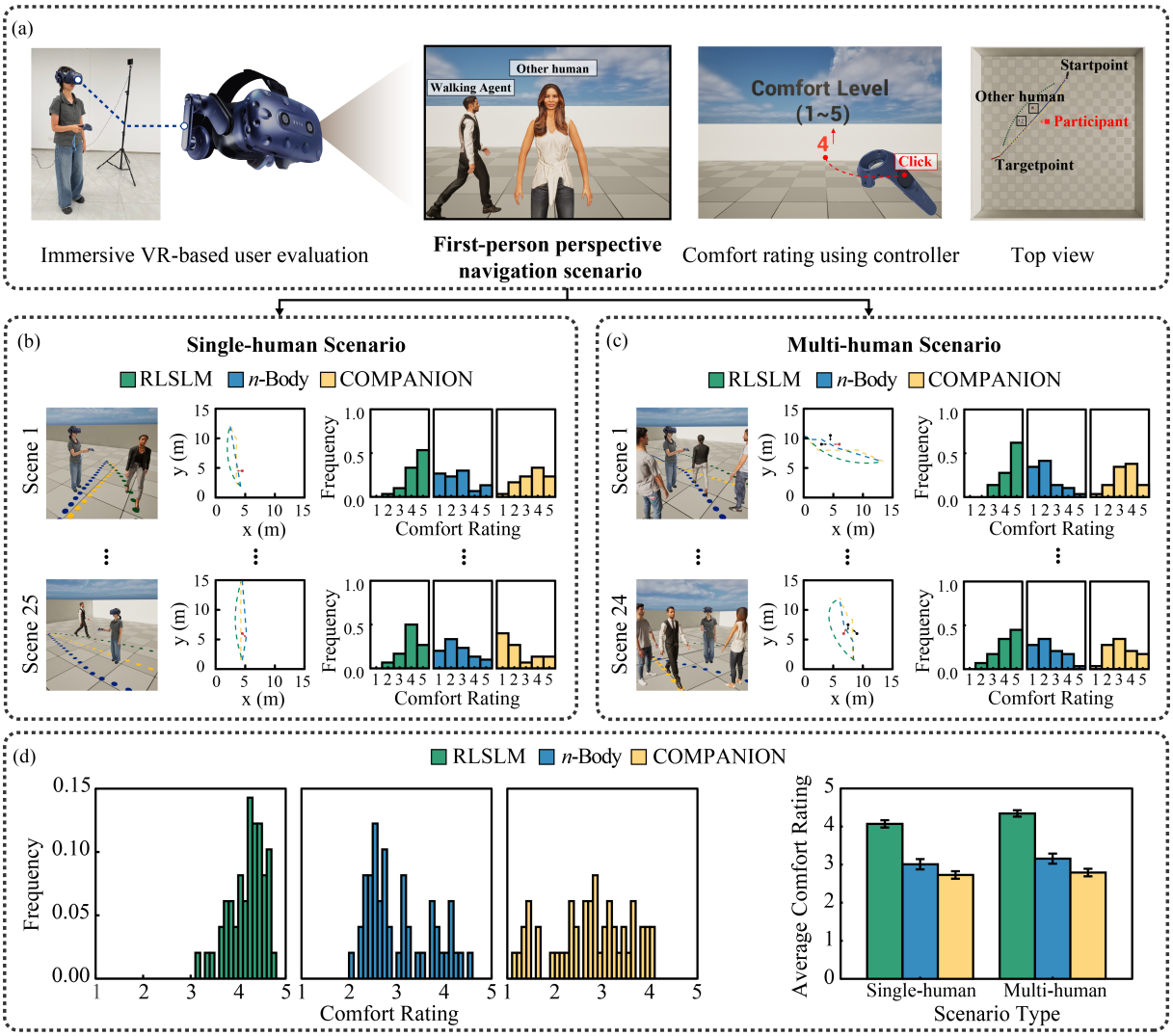

Figure 4: Comfort Rating Analysis via VR-Based User Study. (a) illustrates the VR experiment setup, where participants rate their comfort level (1–5) in both single- and multi-human interaction scenarios. (b) and (c) shows the trajectories of each model (RLSLM, n-Body, and COMPANION) from a top-down view. We selected two representative cases from both scenarios for presentation; the complete results are provided in the supplementary material. (d) presents the comfort rating distributions for each model under both scenarios, comparing the average comfort ratings of three models across both single- and multi-human interaction scenarios.

and n-Body (both in single-human and multi-human scenarios, Bonferroni-corrected post-hoc comparisons, $P < 0.001$). Compared to single-human scenarios, paths generated by RLSLM and n-Body receive significantly higher scores in multi-human scenarios (Bonferroni-corrected post-hoc comparisons, RLSLM: $P < 0.001$, n-Body: $P = 0.008$), whereas COMPANION does not exhibit this multi-human navigation advantage (Bonferroni-corrected post-hoc comparisons, $P = 0.251$).

Interpretable Modeling of Social Behavior Weighting

To evaluate interpretability and adaptability to social variability, we perform a sensitivity analysis on the social behavior weight $\sigma \in \{0, 0.5, 1.0, 2.0\}$, which modulates the agent's responsiveness to nearby individuals through the social influence field. As shown in Figure 5 (d) and (e), higher σ values lead to greater lateral deviations, quantified using Maximum Lateral Distance (MLD) defined in Figure 5 (a). When $\sigma = 0$, the agent strictly follows the shortest path; as σ increases, the agent detours more, prioritizing social comfort. At $\sigma = 2.0$, behavior becomes overly conservative. These trends confirm the effectiveness of σ in shaping socially-aware navigation.

Ablation Study on Social Influence Components

We isolate the contributions of the three components in our social influence module: HRSC, HISC, and CAC.

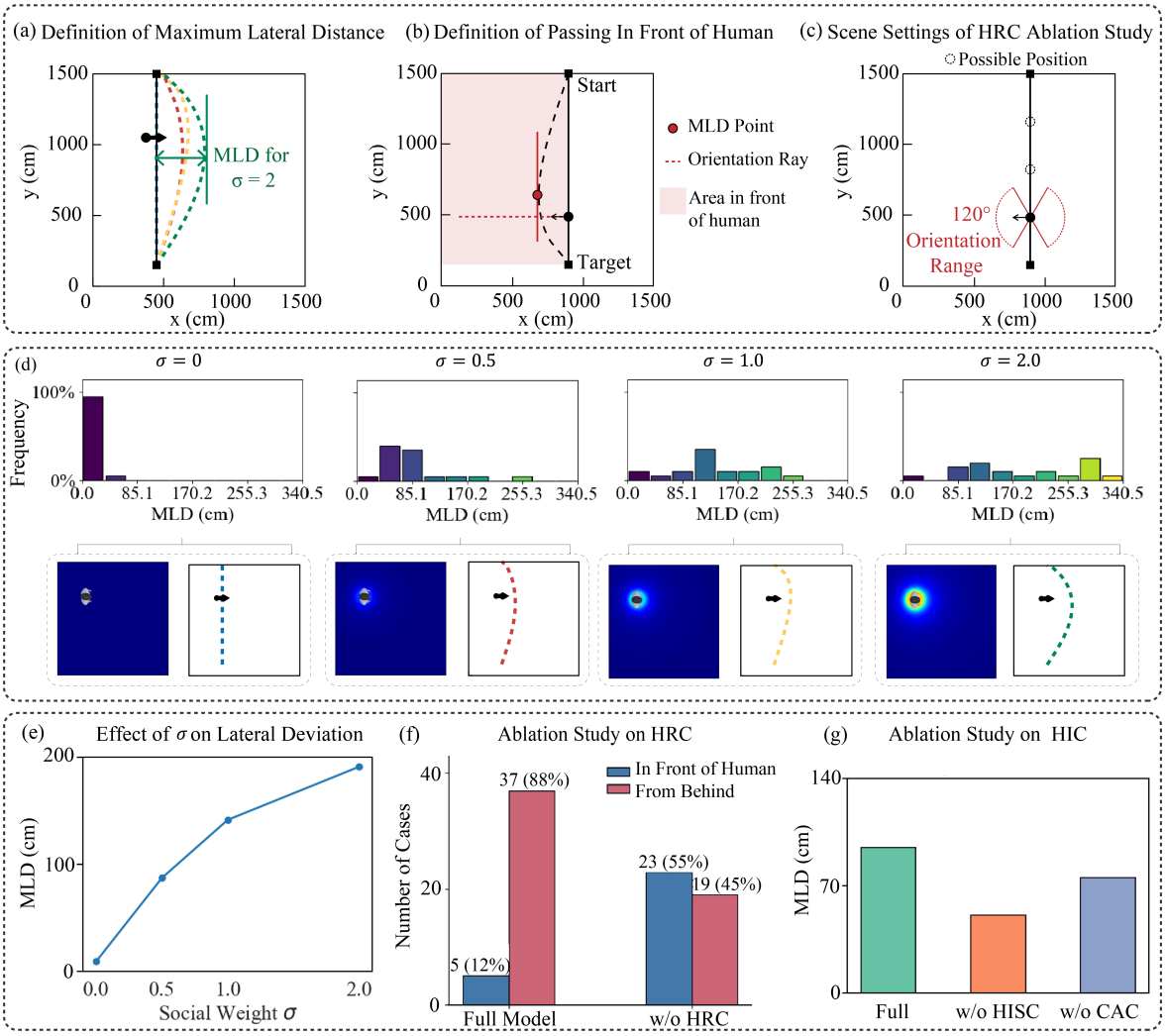

Figure 5: Model validation and ablation analysis. **(a–c)** Definitions and experimental setup. **(d–e)** Effects of varying the social behavior weight σ: (d) shows trajectory examples and MLD distributions under different σ values; (e) reports the corresponding average MLD statistics. **(f–g)** Statistical results from ablation studies of the heading-relevant (f) and heading-irrelevant (g) components of the social influence model. Full experimental details are provided in the appendix.

Heading-Relevant Component In 42 specially-designed single-human scenarios (Figure 5 (c)), removing HRSC causes the agent to pass in front of humans (defined in Figure 5 (b)) in 23 cases (57.76%), compared to only 5 cases using the full model, as shown in (Figure 5 (f)). This demonstrates that HRSC enables sensitivity to human orientation.

Heading-Irrelevant Components Measured in 21 single-human scenarios, removing HISC or CAC leads to reduced MLD (Figure 5 (g)), indicating less stable and less compliant navigation. Qualitative trajectory visualizations and experiment setups are provided in the appendix.

Conclusion

In this paper, we present RLSLM, a hybrid reinforcement learning framework grounded in empirical behavioral experiments for socially compliant robot navigation in human-shared spaces. By integrating a quantitative, rule-based SLM derived from psychological research into a multi-objective RL formulation, our method enables agents to navigate not only efficiently but also in a manner aligned with human social preferences. Through a combination of mechanical energy minimization, goal-directed progress, and social discomfort reduction, the agent learns socially-aware behaviors that generalize to multi-human scenarios. To evaluate alignment with human perception, we designed an immersive first-person VR evaluation pipeline. Results demonstrate that RLSLM significantly outperforms rule-based baseline models in subjective comfort ratings. Additionally, ablation studies and sensitivity analyses underscore the role of each component in shaping nuanced social behavior and demonstrate RLSLM's improved interpretability over conventional data-driven methods. Our findings highlight a promising interdisciplinary pathway for embedding human social cognition into agent policy learning.

Acknowledgments

This work is supported by the National Science and Technology Innovation 2030 Major Program [grant number 2022ZD0205103], Shanghai Municipal Education Commission Initiative on AI-Enabled Research Paradigm Transformation and Disciplinary Advancement [grant number 2024AI01013], the National Natural Science Foundation of China [grant number T2425028], and the National Natural Science Foundation of China [grant number 62377011].

References

Bae, J. W.; Kim, J.; Yun, J.; Kang, C.; Choi, J.; Kim, C.; Lee, J.; Choi, J.; and Choi, J. W. 2023. Sit dataset: socially interactive pedestrian trajectory dataset for social navigation robots. *Advances in neural information processing systems*, 36: 24552–24563.

Bera, A.; Randhavane, T.; Kubin, E.; Wang, A.; Manocha, D.; and Gray, K. 2018. The Socially Invisible Robot: Navigation in the Social World Using Robot Entitativity. arXiv:1805.05543.

Chen, W.; Zhang, T.; and Zou, Y. 2018. Mobile robot path planning based on social interaction space in social environment. *International Journal of Advanced Robotic Systems*, 15(3): 1729881418776183.

Chen, Y. F.; Everett, M.; Liu, M.; and How, J. P. 2018. Socially Aware Motion Planning with Deep Reinforcement Learning. arXiv:1703.08862.

Francis, A.; Pérez-D'Arpino, C.; Li, C.; Xia, F.; Alahi, A.; Alami, R.; Bera, A.; Biswas, A.; Biswas, J.; Chandra, R.; Chiang, H.-T. L.; Everett, M.; Ha, S.; Hart, J.; How, J. P.; Karnan, H.; Lee, T.-W. E.; Manso, L. J.; Mirsky, R.; Pirk, S.; Singamaneni, P. T.; Stone, P.; Taylor, A. V.; Trautman, P.; Tsoi, N.; Vázquez, M.; Xiao, X.; Xu, P.; Yokoyama, N.; Toshev, A.; and Martín-Martín, R. 2025. Principles and Guidelines for Evaluating Social Robot Navigation Algorithms. *ACM Transactions on Human-Robot Interaction*, 14(2): 1–65.

Gong, Z.; Hu, T.; Qiu, R.; and Liang, J. 2025. From Cognition to Precognition: A Future-Aware Framework for Social Navigation. arXiv:2409.13244.

Gupta, A.; Johnson, J.; Fei-Fei, L.; Savarese, S.; and Alahi, A. 2018. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2255–2264.

Helbing, D.; and Molnar, P. 1995. Social Force Model for Pedestrian Dynamics. *Physical Review E*, 51(5): 4282–4286.

Kapoor, A.; Swamy, S.; Bachiller, P.; and Manso, L. J. 2023. Socnavgym: a reinforcement learning gym for social navigation. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2010–2017. IEEE.

Karnan, H.; Nair, A.; Xiao, X.; Warnell, G.; Pirk, S.; Toshev, A.; Hart, J.; Biswas, J.; and Stone, P. 2022. Socially CompliAnt Navigation Dataset (SCAND): A Large-Scale Dataset of Demonstrations for Social Navigation. *IEEE Robotics and Automation Letters*, 7(4): 11807–11814.

Karwowski, J.; Szynkiewicz, W.; and Niewiadomska-Szynkiewicz, E. 2024. Bridging requirements, planning, and evaluation: a review of social robot navigation. *Sensors*, 24(9): 2794.

Kim, S.; Guy, S. J.; Liu, W.; Wilkie, D.; Lau, R. W.; Lin, M. C.; and Manocha, D. 2015. BRVO: Predicting pedestrian trajectories using velocity-space reasoning. *The International Journal of Robotics Research*, 34(2): 201–217.

Kirby, R.; Simmons, R.; and Forlizzi, J. 2009. COMPANION: A Constraint-Optimizing Method for Person-Acceptable Navigation. In *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, 607–612. Toyama, Japan: IEEE. ISBN 978-1-4244-5081-7.

Lerner, A.; Chrysanthou, Y.; and Lischinski, D. 2007. Crowds by example. In *Computer graphics forum*, volume 26, 655–664. Wiley Online Library.

Mahdavian, M.; Nikdel, P.; TaherAhmadi, M.; and Chen, M. 2022. Stpotr: Simultaneous human trajectory and pose prediction using a non-autoregressive transformer for robot following ahead. *arXiv preprint arXiv:2209.07600*.

Mnih, V.; Badia, A. P.; Mirza, M.; Graves, A.; Lillicrap, T. P.; Harley, T.; Silver, D.; and Kavukcuoglu, K. 2016. Asynchronous Methods for Deep Reinforcement Learning. arXiv:1602.01783.

Mohamed, A.; Qian, K.; Elhoseiny, M.; and Claudel, C. 2020. Social-STGCNN: A Social Spatio-Temporal Graph Convolutional Neural Network for Human Trajectory Prediction. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14412–14420.

Pellegrini, S.; Ess, A.; Schindler, K.; and Van Gool, L. 2009. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th international conference on computer vision*, 261–268. IEEE.

Raffin, A.; Hill, A.; Gleave, A.; Kanervisto, A.; Ernestus, M.; and Dormann, N. 2021. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research*, 22(268): 1–8.

Rudenko, A.; Kucner, T. P.; Swaminathan, C. S.; Chadalavada, R. T.; Arras, K. O.; and Lilienthal, A. J. 2020. Thör: Human-robot navigation data collection and accurate motion trajectories dataset. *IEEE Robotics and Automation Letters*, 5(2): 676–682.

Sadeghian, A.; Kosaraju, V.; Sadeghian, A.; Hirose, N.; Rezatofighi, H.; and Savarese, S. 2019. SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1349–1358.

Sheridan, T. B. 2016. Human–robot interaction: status and challenges. *Human factors*, 58(4): 525–532.

Shiomi, M.; Zanlungo, F.; Hayashi, K.; and Kanda, T. 2014. Towards a Socially Acceptable Collision Avoidance for a Mobile Robot Navigating Among Pedestrians Using a

Pedestrian Model. *International Journal of Social Robotics*, 6(3): 443–455.

Sivashangaran, S.; Khairnar, A.; and Eskandarian, A. 2023. AutoVRL: A High Fidelity Autonomous Ground Vehicle Simulator for Sim-to-Real Deep Reinforcement Learning. arXiv:2304.11496.

Towers, M.; Kwiatkowski, A.; Terry, J.; Balis, J. U.; Cola, G. D.; Deleu, T.; Goulão, M.; Kallinteris, A.; Krimmel, M.; KG, A.; Perez-Vicente, R.; Pierré, A.; Schulhoff, S.; Tai, J. J.; Tan, H.; and Younis, O. G. 2024. Gymnasium: A Standard Interface for Reinforcement Learning Environments. arXiv:2407.17032.

Tsoi, N.; Hussein, M.; Espinoza, J.; Ruiz, X.; and Vázquez, M. 2020. Sean: Social environment for autonomous navigation. In *Proceedings of the 8th international conference on human-agent interaction*, 281–283.

Van Den Berg, J.; Guy, S. J.; Lin, M.; and Manocha, D. 2011. Reciprocal N-Body Collision Avoidance. In Siciliano, B.; Khatib, O.; Groen, F.; Pradalier, C.; Siegwart, R.; and Hirzinger, G., eds., *Robotics Research*, volume 70, 3–19. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-642-19456-6 978-3-642-19457-3.

Vuong, A.; Nguyen, T.; Vu, M. N.; Huang, B.; Binh, H.; Vo, T.; and Nguyen, A. 2024. Habicrowd: A high performance simulator for crowd-aware visual navigation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5821–5827. IEEE.

Wang, W.; Mao, L.; Wang, R.; and Min, B.-C. 2024. Multi-robot cooperative socially-aware navigation using multi-agent reinforcement learning. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 12353–12360. IEEE.

Yang, X.; Yang, L.; Majumdar, A.; and Ochieng, W. 2024. Inferring pedestrian decision-making through inverse reinforcement learning. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*, 87–97. Springer.

Yu, C.; Ma, X.; Ren, J.; Zhao, H.; and Yi, S. 2020. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*, 507–523. Springer.

Zhou, C.; Miao, M.-C.; Chen, X.-R.; Hu, Y.-F.; Chang, Q.; Yan, M.-Y.; and Kuai, S.-G. 2022. Human-behaviour-based social locomotion model improves the humanization of social robots. *Nat. Mach. Intell.*, 4(11): 1040–1052.