Coding robots with proxemic awareness using a 3D motion dataset of real people

Abstract:

This work highlights the application of motion-based socially-aware (proxemic) robotic systems designed to engage in close-proximity, collaborative interactions with humans. By encoding normative pedestrian behaviors—such as spatial negotiation, yielding conventions, and turn-taking—into robotic control policies, it is possible to facilitate interactions that are both more intuitive and inherently safer for human collaborators and bystanders alike. The design of what could be described as physically intuitive robots necessitates the integration of datasets that embody social intelligence, thereby extending beyond conventional objective-function optimization.

High-fidelity motion capture data—collected from real human dyadic and group interactions in naturalistic settings—serves as a foundational asset for proxemic intelligence. Such datasets offer semantically rich priors for both data-driven behavioral models (e.g., imitation learning, reinforcement learning) and handcrafted control systems augmented with learned heuristics. The spatiotemporal resolution of marker-based motion capture systems (often sub-millimeter accuracy at 100–250 Hz) enables precise modeling of nuanced physical cues that underlie social interaction dynamics. These data characteristics are essential for training and validating robotic systems operating in unstructured environments where subtlety, safety, and legibility of motion are critical. Motion capture, therefore, provides a scalable and reproducible substrate for instilling socially competent behavior in mobile robots that must navigate shared human spaces.

1. Executive Summary

This paper articulates the critical importance of embedding proxemic awareness into the behavioral design of mobile robotic systems, particularly those intended to operate in dynamic environments and within close interpersonal distances. We emphasize the value of social- and motion-informed interaction paradigms and illustrate how proxemically astute behavioral frameworks can enhance the quality, safety, and intuitiveness of human-robot interactions (HRI) (Fong et al., 2003) (Breazeal, 2003). The discussion will extend into the technical methodologies used to develop these behaviors, highlighting system-level considerations and best practices that transcend traditional task-optimized approaches.

While contemporary robotics has predominantly focused on performance-driven optimization and goal-directed task execution (Taylor et al., 2021), there remains a notable deficiency in solutions aimed at improving the experiential dimension (and, consequently, the safety and trust dimensions) of human-robot coexistence—namely, how people perceive, collaborate with, and cohabitate alongside robotic agents in everyday contexts (Kopp et al., 2021). We present a framework for the integration of socially and spatially competent behaviors into mobile robots by aligning the entire developmental pipeline—ranging from data acquisition and annotation to behavioral modeling and deployment—with principles of proxemic intelligence and human-centered design. This approach recognizes the principle that social awareness through motion acts not as a peripheral enhancement, but as a core design tenet essential to real-world robotic integration (Hoffman & Ju, 2014)(Dragan et al., 2015).

This paper is intended to inform engineers, product designers, robotics researchers, integration specialists, procurement stakeholders, and end-users of the tangible benefits—and ethical imperatives—of embedding social protocols and spatial etiquette into robotic behavior repertoires. These competencies function as mechanisms for trust calibration and improved acceptance in human-robot systems (Hancock et al., 2011), which subsequently leads to enhanced efficiency in collaborative tasks. Moreover, this work situates proxemics-enabled robotics within the broader context of the United Nations Sustainable Development Goals (SDGs), particularly SDG 9 (*Industry, Innovation and Infrastructure*), SDG 11 (*Sustainable Cities and Communities*), and SDG 3 (*Good Health and Well-being*), illustrating how empathetic and socially and spatially competent robotic design contributes to inclusive and responsible technological progress (United Nations, 2015).

2. Introduction

Proxemics is a foundational construct within the field of Human-Robot Interaction (HRI), originating from the human propensity to develop socially governed behavioral frameworks through evolutionary and cultural processes. It can be operationally defined as *etiquette*—a set of implicit norms and behavioral expectations that, while not codified as formal rules or laws, are widely internalized and practiced within specific cultural, geographic, and social contexts (Goffman, 1967) (Hall, 1966). These social norms are instrumental not in optimizing energetic or computational efficiency, but in enhancing the subjective quality of interaction, facilitating mutual predictability, and building trust between agents. They are so deeply embedded in human behavior that individuals often adhere to them despite requiring additional cognitive or physical effort (Gross & Vostroknutov, 2022).

Contemporary robotic systems are predominantly engineered to perform tasks that are repetitive, labor-intensive, or undesirable for humans—frequently with the goal of decoupling human involvement altogether (Acemoglu & Restrepo, 2019). However, as robots become increasingly integrated into human-populated environments, it is inevitable that they will share space and interact with people, regardless of whether the task is collaborative or substitutive. Marty (the Stop and Shop robot), for example, is actively deployed in grocery stores and regularly interacts with humans in these very public spaces (Weiss, 2024). Current paradigms in autonomous navigation and mobile robotics often treat humans as dynamic obstacles to be avoided (Mavrogiannis et al., 2023), a strategy that prioritizes collision avoidance but neglects the inherently cooperative and socially nuanced nature of human spatial behavior. This framing impedes the development of mutual trust—an essential component of effective HRI that directly impacts user comfort, willingness to engage, and perceived safety (Campagna & Rehm, 2025).

Although many robots are explicitly programmed to maintain safe distances from humans, this conservative behavior without embedded proxemic intelligence can inadvertently signal unpredictability or risk, leading humans to exhibit avoidance behaviors themselves. This reciprocal avoidance dynamic can induce anxiety and cognitive load in users, reinforcing perceptions of robots as disruptive or hazardous agents (Takayama & Pantofaru, 2009), reducing the efficiency of human/robot interactions in the execution of tasks. In contrast, socially-based proxemic integration enables robots to navigate in ways that align with human behavioral expectations, allowing them to be perceived as socially competent entities. When robots exhibit motion-based socially attuned behaviors, they can enhance user comfort, reduce interaction friction, and even improve productivity and well-being by minimizing physical strain and psychological stress (Mutlu & Forlizzi, 2008).

Mainstream approaches to robotic navigation remain largely centered on efficiency-driven rule sets and deterministic path planning (AbuJabal et al., 2024). While these systems are valuable for constrained or industrial settings, they often fail to account for the experiential impact on human co-occupants, bystanders, and observers. Optimization-centric algorithms, while predictable, lack the expressive or communicative behaviors that facilitate trust-building (Eder et al., 2024). Importantly, the integration of proxemics does not require abandoning optimization altogether. Rather, it introduces a meta-layer of behavioral reasoning that enables robots to evaluate the social ramifications of their actions, balancing performance objectives with interactive context.

Social asymmetry often persists where robots adapt behaviors based on interaction feedback, even in reinforcement learning (RL) frameworks. Humans are typically required to yield or adapt to the robot's policy-driven behavior, which does not reflect the bidirectional learning found in natural human collaboration. True social collaboration necessitates co-adaptive systems—where both human and robot contribute actively to mutual goal attainment (Bütepage & Kragic, 2017). Incorporating proxemics into the training process involves reward structures that explicitly encode and reinforce socially cooperative behavior (e.g., yielding space, matching pace), thus incentivizing the robot to respond not just to task metrics, but to socially meaningful cues (Reddy et al., 2018).

The challenge, of course, lies in acquiring the behavioral priors required to support this kind of nuanced decision-making. Many of the behaviors that underlie motion-based human social etiquette are implicit, context-dependent, and non-verbal (Urakami and Seaborn, 2023)—making them difficult to encode via rule-based systems alone. The most viable path to surfacing and formalizing these latent behaviors is to begin with human observational data. While robots are a relatively recent technological innovation, humans have refined their interactive behavior over millennia. Furthermore, the built environment itself is optimized around human locomotion and cognition through proxemics (Möystad, 2017). Thus, the most valid method for understanding how robots *should* behave is to empirically examine how humans already interact with each other and with the environment under natural conditions.

In the remainder of this paper, we present a comprehensive methodology for encoding social and proxemic awareness into mobile robotic systems. This includes strategies for collecting and preprocessing human behavioral data (Sections 3.1 and 3.2), approaches to integrating both algorithmic control and data-driven modeling (Sections 3.3 and 3.4), illustrative use cases (Section 5), ethical and societal implications (Section 6), and the benefits of early and regular real-world deployments to accompany data collection and analysis (Section 7).

3. Our Approach: Learning from Real Human Behavior

Methods of data collection are as critically important to the discovery and modeling of social and spatial etiquette as the subsequent stages of analysis and algorithm development. In order to capture the subtlety and complexity of motion-based socially embedded behaviors, the base dataset must be robust, high-resolution, clean, and representative of diverse real-world interactions. Social behaviors—especially those involving nonverbal cues, proxemics, and multi-agent coordination—often emerge from fine-grained kinematic patterns that are difficult to replicate or approximate synthetically without a foundation in real human data. Consequently, the integrity of any trustworthy and socially-aware robotic system is fundamentally linked to the quality, scale, and fidelity of the behavioral dataset from which it is derived. The following sections outline our approach to data acquisition, preprocessing, feature representation, algorithmic behavior generation, and validation, with an emphasis on rigor, repeatability, and ethical data stewardship.

3.1 Data Collection

Motion data in the PFF lab is captured using an optical motion capture system composed of 14 high definition infrared cameras, integrated with auto-labelling and post-processing software. The system operates at up to 200 frames per second—adequate enough to resolve fine-grained social proxemic cues—and provides sub-millimeter accuracy. We utilize two calibrated volumes to explore linear and nonlinear behaviors (9.8 × 3.3 m and 4.8 × 4.8 m, respectively).

Each participant wears at least 53 reflective markers placed using a standardized and easily-recognizable biomechanical setup. Critical environmental objects (e.g., doors, frames) are fitted with 5–6 additional markers at each important feature to capture full spatial-temporal interaction dynamics. This allows for detailed reconstruction of joint trajectories, head and wrist articulation, gait patterns, and other socially and spatially relevant motion features.

A typical study collects approximately 1,500 motion sequences per interaction class, yielding 130 marker trajectories per frame and over one billion data points per behavior category. These high-resolution datasets form the empirical foundation for downstream behavioral modeling.

To reduce bias and increase diversity, participants (about 30 per study) are selected to span a wide demographic range, including varying age, gender, and cultural backgrounds. Crucially, they are unfamiliar with robotics and are not informed of the study’s specific goals, helping to mitigate knowledge bias and the Hawthorne effect.

Scenarios are staged within full-scale, human-realistic environments constructed inside the capture space. These sets include architectural elements and interactable doors with various hinge and swing configurations (e.g., left/right, push/pull, single/double). Semi-transparent occlusion panels maintain tracking fidelity while simulating real-world physical constraints (see Figure 01).

Participants are primed with brief scenario prompts (e.g., “You’re walking through a building with a friend,” or “Someone behind you is carrying a bag”) followed by immediate trial initiation. To reduce overthinking and evoke subconscious behaviors, participants may be engaged in light cognitive distractions, such as verbal tasks or memory recall, allowing natural motion heuristics to emerge.

This method is exemplified in our door traversal study, which included 10 recording sessions with 3 paid participants each and over 2,400 total trials (~240 per session). Social formations varied from individuals to triads; door setups included left/right hinges, push/pull mechanics, and encumbrance conditions (e.g., with or without luggage). These variations were designed to elicit behaviors such as yielding, passing, holding, and negotiation of constrained movement spaces.

Together, the experimental scale, structured diversity, and technical precision of our motion capture system yield a dataset suitable for robust statistical analysis and socially-grounded behavior modeling. Each sequence contributes to a scalable framework for generating both scenario-specific algorithms and generalizable learning models.

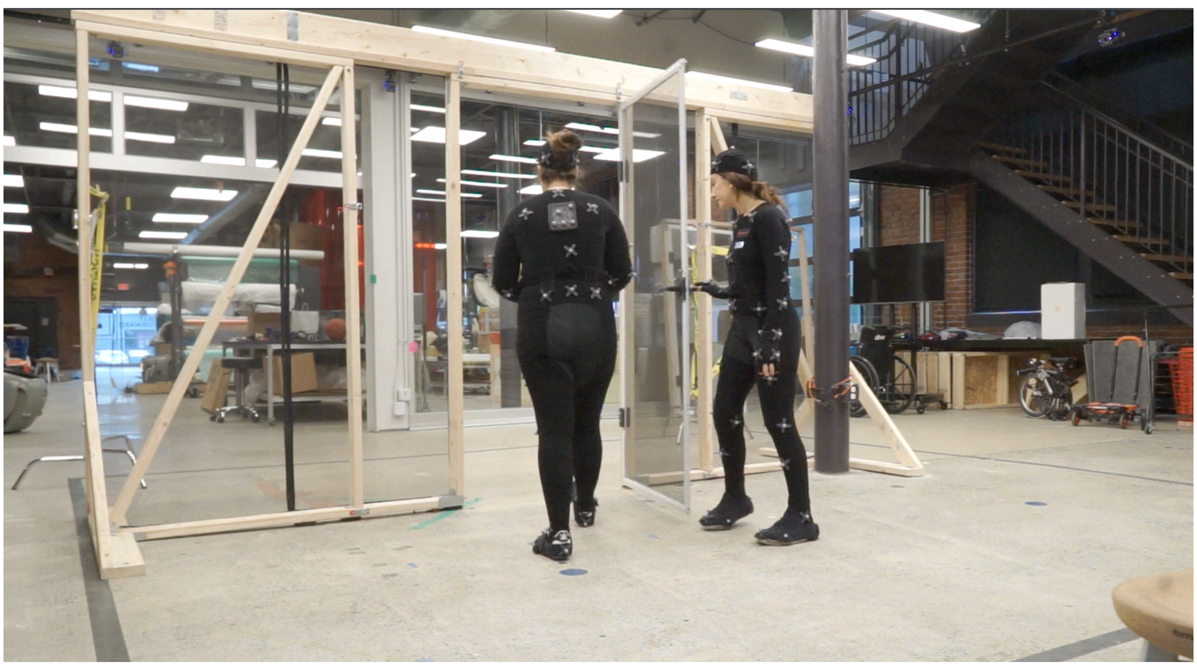

Figure 01. The set for the study on door traversal door set built for the study

3.2 Data Representation and Preprocessing

Initial preprocessing is conducted in our GUI-based trial visualization software, where brief occlusions are interpolated, marker swaps are corrected, and skeleton consistency is verified. During this phase, spatial relationships between critical environmental elements—such as door edges, hinge axes, and handles—are also annotated and synchronized with the agent trajectories. All sequences are then temporally normalized to a consistent frame rate (typically 120 Hz), enabling frame-wise alignment across trials and reducing inconsistencies in downstream temporal modeling.

The raw motion data exported from the post-processing software is first sifted through a structured, multi-step pipeline designed to enhance consistency, facilitate interpretability, and prepare the data for behavior modeling and machine learning. Files are exported in CSV format, each corresponding to an individual trial containing frame-by-frame 3D marker trajectories for all tracked agents and environmental objects.

Each agent in the scene is then assigned a semantic role label (e.g., leader, follower, bystander, accompanier), enabling role-specific analysis and conditioning. Skeleton data is further enriched with derived kinematic and spatial features, including joint velocities, local orientations, center-of-mass acceleration, egocentric direction vectors, and allocentric distances to scene elements (e.g., relative proximity to the door plane or angular offset to the traversal path). These features enable finer-grained exploration of implicit, often subconscious, social proxemic behaviors and facilitate the identification of subtle nonverbal communication cues embedded in motion patterns.

In door-related studies, for example, we compute the door's frame-by-frame angular displacement using the relative orientation of rigid-body marker sets affixed to the door and doorframe. This continuous angular signal serves both as a key input feature (reflecting environmental state) and as a target variable in door control or prediction models, offering a link between agent behavior and dynamic environmental change.

To support flexible querying and targeted analysis, we extract and store trial-level metadata for each file, including participant configuration, trial scenario, environmental complexity, encumbrance condition, and recording session. This metadata allows for dynamic filtering of datasets during model training and analysis, as well as the definition of higher-order classification schemes—e.g., grouping similar trials into behavioral archetypes or identifying rare but critical edge-case scenarios.

Critically, this preprocessing structure enhances the capacity to identify and analyze social etiquette behaviors—for example, spatial deference shown during passage, body language used to indicate intent, or role-switching behaviors mid-interaction. By anchoring motion features to real-world referents (like door positions), we support diverse proxemic analysis that transcends task efficiency and instead centers on interpersonal legibility, comfort, and trust.

3.3 Algorithm Design and Behavior Generation

Our analysis pipeline begins with comprehensive feature engineering, where raw motion trajectories are transformed into a structured set of kinematic, spatial, and contextual features. These include velocity, acceleration, jerk, turning rate, and spatial proximity—both in egocentric coordinates (e.g., distance to nearest agent) and allocentric space (e.g., position relative to the doorframe or passage vector). We also compute derived social interaction indicators such as role-based distances (e.g., leader–follower offset), directional alignment, and interaction symmetry. These indicators allow us to move beyond raw physical data to model the nuanced qualities of social coordination, intent recognition, and spatial negotiation.

Exploratory analysis is conducted via interactive dashboards, which visualize high-dimensional behavior in 2D and 3D space. These include heatmaps, trajectory overlays, temporal phase plots, and proximity histograms. This step allows researchers to detect emergent structures, such as crowding patterns, yielding motions, or implicit turn-taking, that are not always obvious in single trials but become statistically salient at scale (see Figures 02-04).

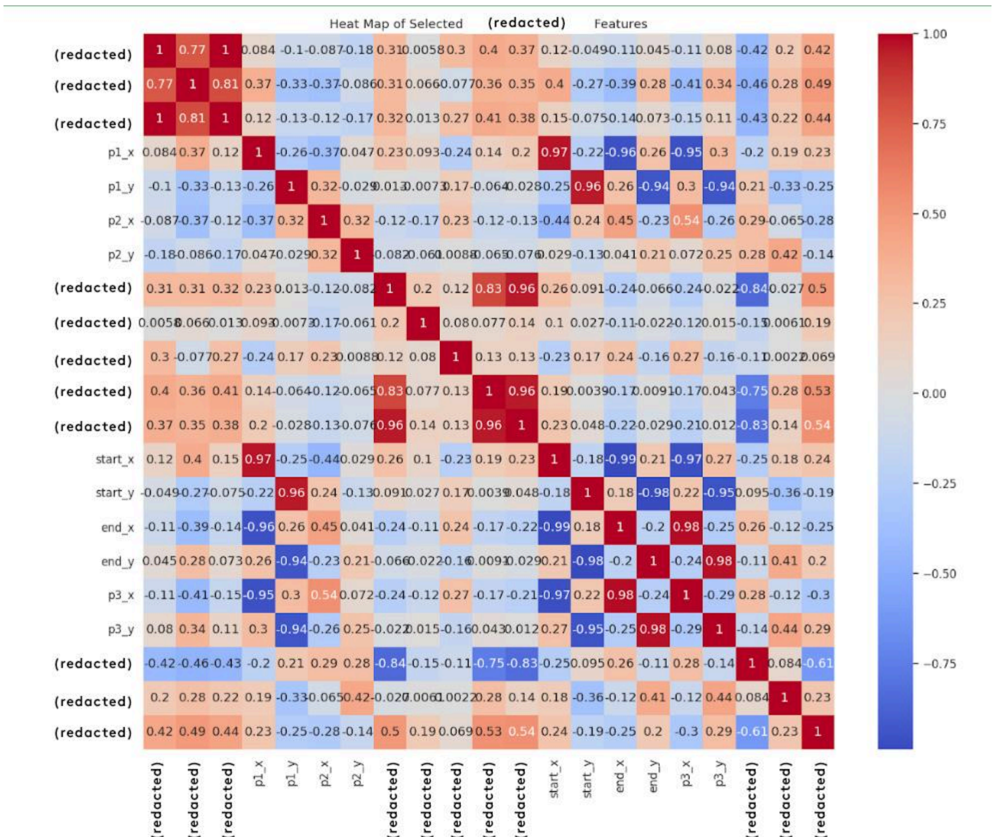

Figure 02. Heatmap reflecting the correlation between selected features

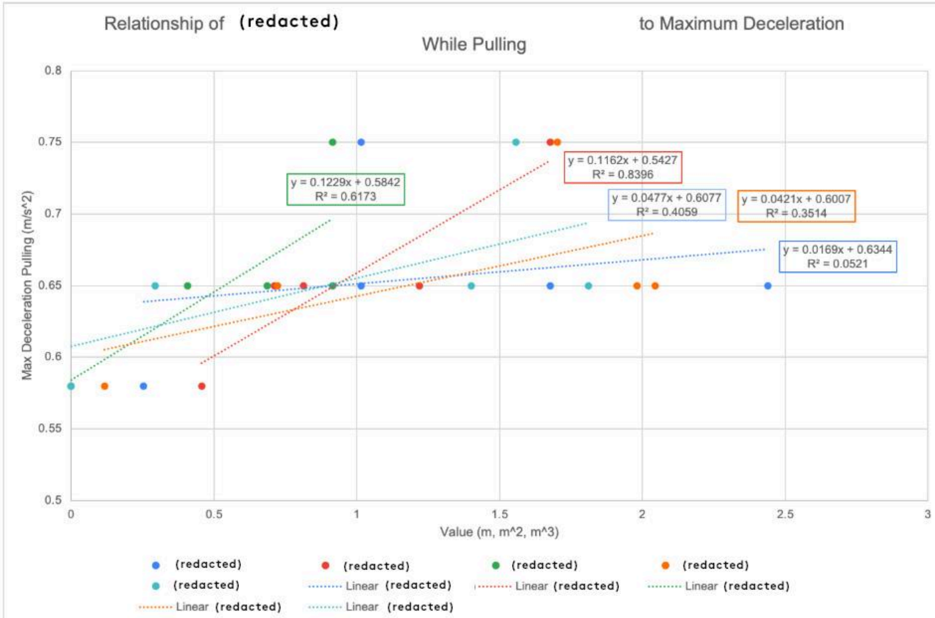

Figure 03. Linear regression plot reflecting the relationship between maximum deceleration while pulling and various other environmental features

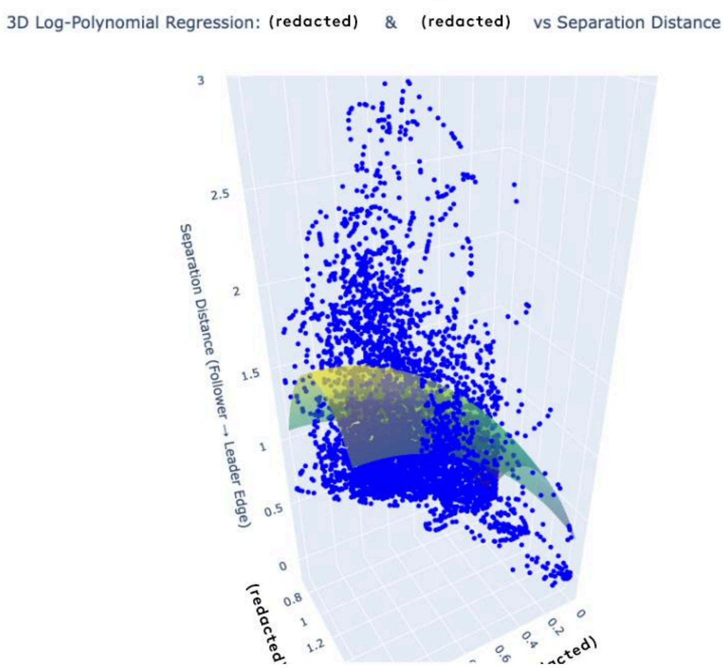

Figure 04. 3-D plot reflecting the polynomial regression between two features and the separation distance between the leader and follower

We also incorporate unsupervised clustering techniques—including k-means, DBSCAN, and hierarchical clustering—to identify behavioral state transitions directly from the multivariate feature space. These clusters often reveal social-motor primitives such as approach, pause, yield, initiate pass, and exit, which align with known stages of human interaction. Segmenting motion data in this way not only supports annotation efficiency but also reveals the temporal and spatial signatures of socially meaningful transitions.

These methods are particularly well-suited to modeling proxemics because they preserve and amplify the emergent qualities of interaction. Motion-based social awareness is not reducible to predefined motion paths or rule-based collision avoidance; it relies on the capacity to perceive intent, interpret subtle behavioral cues, and adjust one's motion in real time. By extracting continuous, relational, and context-aware features—and analyzing them using flexible, data-driven methods—we are able to capture interactional subtleties that would be lost in purely deterministic or low-resolution representations. These methods also facilitate the construction of behavioral policies that respond dynamically to fluid social situations, such as shared passageways, interpersonal negotiation, and deference behaviors.

For classical analysis, we fit linear and non-linear regressions to key feature relationships (e.g., interpersonal distance as a function of door angle or hesitation time as a function of approach speed). These regressions act as interpretable heuristics and offer parameterized insights into how proxemic social behaviors scale with context (see Figures 03 & 04).

In parallel, we train machine learning classifiers (e.g., random forests, gradient boosting machines) to learn the mapping between engineered features and semantic labels defined in our preprocessing pipeline. These models serve two purposes: (1) to validate the structure of our hand-designed heuristics, and (2) to act as components in modular robotic behavior stacks. In early-stage deployment, these classifiers allow robots to recognize and classify dynamic human behaviors using live sensor input.

We leverage generative sequence models (e.g., autoregressive transformers, RNNs) to produce synthetic behavioral sequences that reflect realistic transitions between behavioral states. These models are trained on real interaction sequences and produce new examples that adhere to the statistical proxemic dynamics present in the original data (see Figures 05 & 06). When deployed, these models allow for flexible behavior generation, enabling robots to simulate and enact plausible motion-based social actions—such as yielding, gesturing, or aligning body position—without hardcoded scripting.

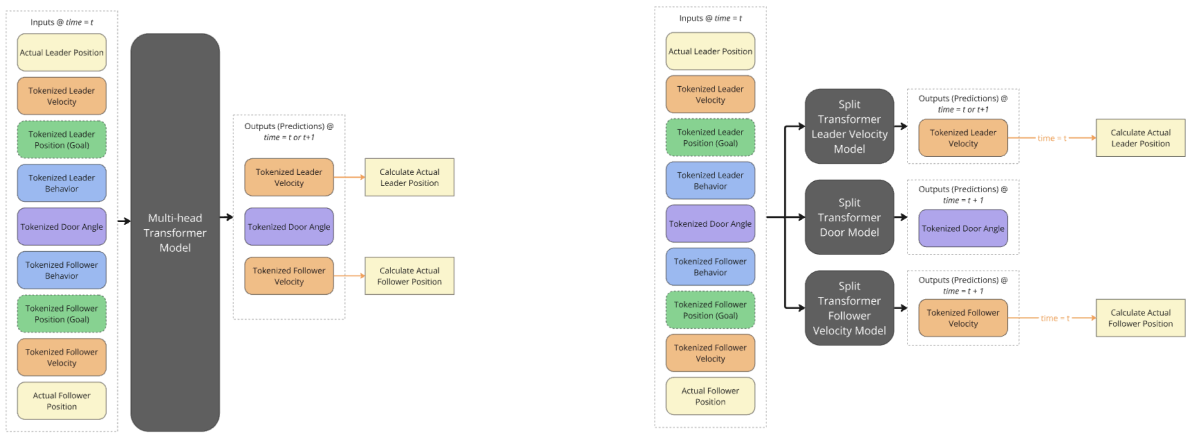

Figure 05. Examples of model structures for generating synthetic data

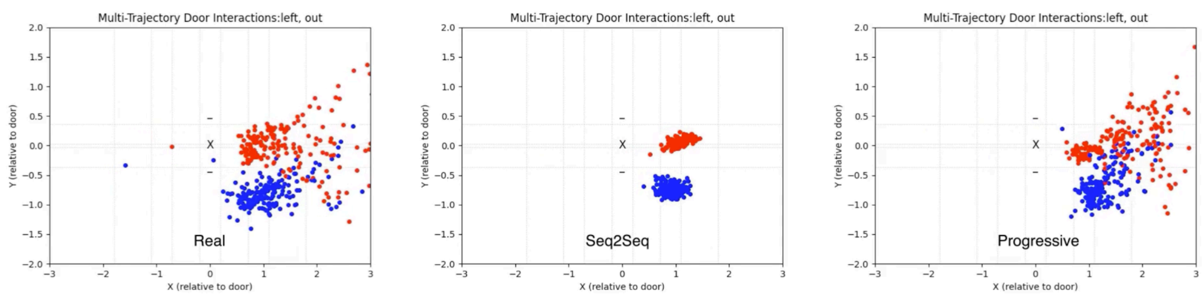

Figure 06. Example timestamp of generalized synthetic data

3.4 Algorithm Validation

Validation is a critical component of proxemic behavior modeling, ensuring that the heuristics and models designed from motion data not only generalize across scenarios but also reflect the nuanced qualities of real-world social and spatial interaction. In socially dynamic settings, robotic behavior must be both trustworthy and intelligible—qualities that require close alignment between predicted behavior and observed human norms. Through structured validation procedures, we ensure that the behaviors produced by our models match real-world expectations for coordination, intent, and etiquette.

Our validation pipeline integrates classical machine learning techniques with hand-labeled ground truth annotations derived from human-coded datasets. For each behavior scenario our team conducts frame-level manual annotations of behavioral phases based on visual inspection of the motion capture recordings. These annotations include temporally segmented social phases such as door approach, decoupling (when the follower detaches from the leader to independently pass through the door), recoupling (when the agents reunite), and walking away. These labels represent semantically meaningful states in joint human motion and form the behavioral ground truth for comparison.

To evaluate the accuracy of our algorithmic segmentation and classification, we trained supervised classifiers—primarily random forest models—on the engineered feature space derived in Section 3.3. These models learn the associations between input features (e.g., interpersonal distance, agent velocity, door angle, alignment to passage axis) and the annotated behavioral classes. The same feature sets are passed through our rule-based heuristic segmentation algorithms, which label the data automatically based on kinematic thresholds and spatial conditions.

Performance is then assessed by comparing the heuristic-based labels against the ground truth annotations using standard classification metrics such as accuracy, precision, recall, and F1 score. For temporal behavior validation, we also compute phase alignment accuracy, measuring the degree to which the timing and duration of each labeled phase aligns with human-labeled boundaries. This comparison is particularly important for proxemic behavior modeling, where the timing of actions (e.g., yielding, passing, synchronizing) is as important as the action itself (see Figure 07).

In the door traversal study, this comparative validation revealed strong alignment in most behavioral phases but exposed minor discrepancies during high-ambiguity transitions, such as the decoupling–recoupling phase. These edge cases offer valuable insight into how human interactions unfold across variable contexts and serve as concrete opportunities for heuristic refinement. By identifying patterns of disagreement between human- and machine-generated labels, we iteratively refine our models to better capture the emergent social and spatial structure of motion—moving beyond pure efficiency metrics to ensure that robot actions align with human interpretability and comfort.

More broadly, this validation process is essential for the safe deployment of robots in real-world environments that consider proxemics. In traditional robotics, validation often focuses on trajectory error or collision rates. However, socially and spatially aware robots must be validated not only on physical correctness but also on the legibility and appropriateness of their actions within shared human spaces. By grounding our models in observational truth data and ensuring their alignment with human-understood social states, we build systems that are not only accurate but meaningfully responsive to human social norms.

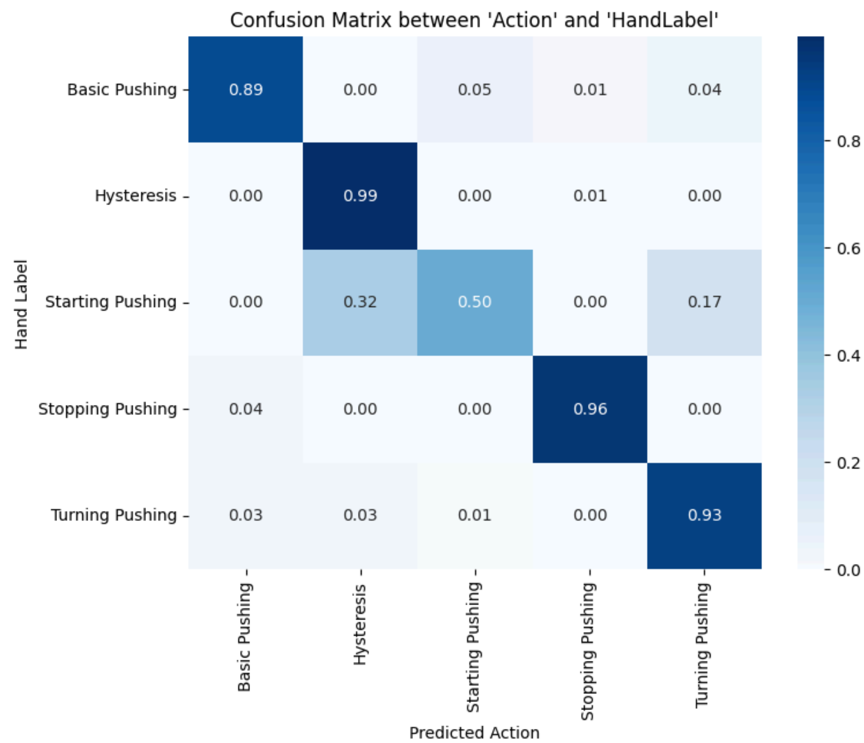

Figure 07. Example of a confusion matrix reflecting the comparison between hand-labeled ‘action’ data and spec-defined ‘actions’, including temporal alignment

4. Advantages and Innovations

The incorporation of proxemic navigation and interaction strategies into mobile robotic systems yields measurable benefits not only in operational productivity but also in the quality of experience for both users and bystanders.

From an implementation perspective, robots that conform to human expectations and conventions in proxemics significantly reduce deployment overhead. These systems minimize the need for extensive user training and environmental restructuring, enabling more rapid integration and accelerating return on investment (ROI) in human-centered environments (Hoffman & Ju, 2014). These robots trained on empirical human behavioral data can be introduced into legacy workflows with minimal friction, functioning as plug-and-play agents within collaborative ecosystems. Crucially, these are environments where *collaboration*, not just *cooperation*, is required to realize meaningful performance improvements. Collaboration introduces complex interdependencies between human and robotic agents, characterized by shared mental models, coordinated task structures, and mutual adaptation through real-time feedback and learning (Bütepage & Kragic, 2017). These interdependencies foster more fluent interactions and reduce the cognitive load imposed on human counterparts—particularly salient in physical or spatially constrained interactions (Breazeal et al., 2016) (Nikolaidis et al., 2015).

Behavioral congruence between humans and robots—especially when achieved through biomimetic or anthropomorphic design principles—has been shown to significantly elevate levels of trust, engagement, and task performance in collaborative contexts (Salem et al., 2015). Trust, in this setting, operates as a latent variable that modulates emotional, cognitive, and physical engagement. Despite its significance, trust remains underdeveloped in contemporary robotic systems, many of which are optimized for automation and task substitution rather than coactivity or co-presence. As a result, these systems often displace human roles entirely rather than augmenting them, reinforcing skepticism and reducing user agency. Our vision for motion-based socially and spatially sensitive robotic systems aims to reverse this trend by designing for *collaborative augmentation*—reducing the burden of work without erasing the role of the worker. Social proxemic etiquette, codified through data-driven modeling, becomes the behavioral substrate upon which trust and mutual intelligibility are built (Fong et al., 2003).

Motion-based social behaviors are emergent, highly context-sensitive, and shaped by decades or centuries of cultural evolution, making them difficult to formalize using classical robotic planning frameworks (Mavrogiannis et al., 2023). The innovation presented in this work lies in our commitment to empirically grounded, data-driven behavioral modeling. Rather than encoding interaction heuristics from expert knowledge or simulation, we derive models from observations of unaffiliated human participants operating in ecologically valid environments. These individuals are unaware of our specific objectives and are exposed to naturalistic social scenarios with minimal instruction. Through this approach, we extract a distribution of normative behavioral patterns—features that are generalizable, robust, and flexible enough to form the basis of scalable interaction policies.

Capturing these nuanced behaviors requires a highly instrumented motion capture infrastructure capable of recording full-body kinematics with high temporal fidelity. Our data collection pipeline is centered around an animation-grade optical motion capture system operating at 120 frames per second, enabling the extraction of dynamic behaviors with millimeter precision—orders of magnitude faster than typical human perceptual reaction cycles (Jain et al., 2015). While our datasets are high-volume (on the order of trillions of data points), the computational overhead remains low due to our reliance on kinematic, rather than pixel-based, features. This facilitates faster training, reduced model complexity, and improved generalization, while allowing us to isolate the key nonverbal cues—such as gait modulation, proxemic alignment, and gestural synchrony—that underpin motion-based social spatial reasoning.

Given the emergent and multi-scale nature of proxemic social behaviors, the structure of our dataset is essential to preserving behavioral integrity during modeling. Our data architecture is inspired by hierarchical models of human decision-making and is organized into three semantic layers (Kaviani et al., 2024). At the macro level, we define *behaviors*—temporally and spatially extended interactional events that contextualize individual motion. These are decomposed into *actions*, which are mid-level semantic units capturing recognizable subcomponents of interaction through their kinetic, spatial, and temporal signatures. At the most granular level, we capture *features*—quantitative descriptors of articulated body motion over time and space. This tiered structure allows for behavior synthesis that reflects the compositional nature of human cognition and action planning. Unlike flat or optimization-centric models, this hierarchy enables us to recreate the implicit structure of social and spatial etiquette, as it has evolved in natural human systems. Details on this architecture are presented in Sections 3.2 and 3.3.

5. Use Cases and Scenarios

Socially-aware mobile robots possess significant applicability across domains where robots are expected to interact with users, bystanders, or the built environment. These domains span a broad range of dynamism and persistence in built environments and the activities within them. Our company has unsupervised robots deployed in a range of environments including private residences such as homes and apartments, semi-private and controlled environments such as workplaces, schools, healthcare facilities, and hospitality buildings, and fully public spaces including parks, sidewalks, campuses, retail centers, transportation hubs, and event venues. While the benefits of spatially and socially competent robotic systems are broadly applicable, given our experience with deployed robots "in the wild," several environments stand out where social and spatial awareness and etiquette in robot behavior are particularly critical for improving both performance and human acceptance.

5.1. Warehouses and Fulfillment Centers

Logistics environments, including warehouses and fulfillment centers, represent one of the largest domains for robotic deployment in commercial spaces (Wurman et al., 2008). Despite significant automation, humans remain integral to operations, especially for tasks requiring dexterity, fine motor control, and adaptive decision-making, such as picking, sorting, and quality assurance (Boschetti et al., 2023). These environments are typically characterized by structured layouts, clearly defined workflows, and relatively predictable human motion patterns, making them an ideal proving ground for proxemic, socially-aware, collaborative robotic systems.

Robots that exhibit cooperative spatial negotiation, proactive intent signaling, and respect for proxemic boundaries can significantly reduce human cognitive load, foster trust, and accelerate human-robot co-adaptation. In turn, this reduces onboarding and training time, improves operational efficiency, and increases both worker satisfaction and safety.

5.2. Event Spaces and Crowded Public Venues

Event spaces—including exhibition halls, sports arenas, and entertainment venues—represent highly unstructured, dynamic environments characterized by dense crowds, complex human motion patterns, and emergent social behaviors (Mavrogiannis et al., 2023). In such environments, conventional reactive collision-avoidance algorithms are insufficient for achieving safe and acceptable robot behavior. Instead, robots must exhibit context-sensitive, socially compliant navigation that respects fluid human group dynamics and informal social norms.

The ability to perceive and interpret nuanced proxemic social cues, such as flow patterns, personal space maintenance, and informal queuing, is critical for both operational performance and public trust. Socially-aware navigation not only improves robot efficiency and safety but enhances public perception and social acceptance of autonomous systems operating in close proximity to large groups of humans.

5.3. Parks and Outdoor Public Spaces

Outdoor public spaces, such as parks, plazas, and pedestrian walkways, remain one of the most challenging and underdeveloped environments for mobile robotics. These spaces are inherently unstructured, exhibit high variability in human activity, and are strongly associated with public ownership and user expectations of comfort, freedom, and social etiquette (Lindner & Eschenbach, 2011).

For autonomous systems operating in such settings, motion-based social awareness is not only a technical requirement but a socio-cultural necessity. Robots must exhibit behaviors that reflect an understanding of implicit social rules, such as right-of-way conventions, group avoidance, and appropriate speeds, to promote bystander comfort and public trust. Failure to do so risks not only physical safety but broader societal rejection of autonomous systems in shared human environments.

6. Challenges and Ethical Considerations

The development of proxemic socially-aware robotic systems, while offering significant advantages for human-robot interaction, requires careful attention to ethical data practices, experimental design, and deployment strategies to ensure fairness, safety, and public trust. In particular, the use of human motion data to inform robot behavior introduces challenges related to dataset bias, participant privacy, observer effects, and the generalizability of learned behaviors to unstructured real-world environments (Seethapathi et al., 2019).

6.1. Dataset Representativeness and Bias Mitigation

The integrity and generalizability of motion-based socially-aware robot behavior are directly dependent on the quality and diversity of the underlying human behavioral datasets. Motion data derived from laboratory environments, while offering high fidelity and experimental control, are susceptible to sampling biases if participant demographics are insufficiently varied (Heckman, 1979). To mitigate this risk, our recruitment strategy prioritizes diversity across cultural background, gender identity, age, and physical ability, ensuring that the resulting behavioral models are not disproportionately shaped by a narrow subset of the population.

While all laboratory-based studies inherently reflect some geographic bias, situating our motion capture facility in an international academic hub such as Boston provides access to a heterogeneous participant pool, enhancing dataset representativeness. As of 2024, Boston's foreign-born population was 28%, and while 44.6% of the city identifies as white, black, asian, and hispanic residents make up 19.1%, 11.2%, and 18.7% respectively (Boston Planning & Development Agency Research Division). This places the Boston metropolitan area as 10th in the US for foreign-born population (Migration Policy Institute). Our experimental design further strengthens generalizability by incorporating repeated trials with both unique and returning participants (≥30 per scenario), enabling the validation of consistent behavioral patterns while capturing natural inter-individual variability.

6.2. Privacy Preservation and Observer Bias Reduction

High-resolution optical motion capture provides a uniquely advantageous data modality for socially-aware robotics due to its ability to capture sub-millimeter, full-body kinematic information while preserving participant anonymity. Our data collection protocols exclude all personally identifiable information (PII) from motion datasets. Participant metadata (e.g., demographic information) is stored securely in isolated files for record-keeping purposes only and is explicitly excluded from algorithmic training pipelines, ensuring that model behavior remains uninfluenced by socially constructed biases or identifiers (Mittelstadt et al., 2016).

Additionally, our protocols are designed to mitigate the well-documented Hawthorne effect—the phenomenon where individuals alter their behavior due to awareness of observation (Mayo, 2004). We avoid disclosing the specific goals of the study to participants to prevent conscious or unconscious behavioral modification, particularly the tendency to 'solve' tasks with knowledge of robotic constraints. To further reduce environmental bias, we employ familiarization techniques, provide participants with simple, cognitively engaging tasks, and design trials to encourage naturalistic, subconscious movement patterns.

6.3. Generalization and Behavioral Complexity

Despite rigorous experimental controls, it remains a non-trivial challenge to transfer learned, socially intelligent behaviors from structured laboratory environments to complex, unstructured real-world settings such as public spaces or crowded pedestrian environments. Social navigation, in particular, requires robots to interpret and respond to highly variable, emergent human behaviors while maintaining safety, efficiency, and social acceptability (Kruse et al., 2013).

To address these challenges, we adopt a hybrid modeling approach that integrates large-scale, demographically diverse human motion datasets with modular control architectures and adaptive learning algorithms. This allows for both the accurate reproduction of common, socially-expected behaviors and the flexible integration of new behavioral patterns as deployment contexts evolve. Our approach prioritizes the development of generalized, yet customizable, behavior models that balance robustness, safety, and social fluency in human-robot interactions.

7. Deploying Socially Aware Robots in the Real World

As proxemic socially-aware robotic systems are deployed in both structured and unstructured environments, the technological, computational, and data-driven foundations underpinning these systems evolve in order to meet the demands of complex, real-world human-robot interaction. Social and spatial etiquette among humans emerges from highly nuanced, context-dependent behaviors that are acquired and refined through evolutionary, neurological, and sociocultural processes (Frith & Frith, 2007). Over the last six years we have deployed both consumer robots in communities across the United States and industrial robots in businesses which require dynamic flexibility and human interaction in their workplaces. In terms of interaction development, there is no substitution for deployment, and this world experience goes hand-in-hand with the emulation of behaviors that are feasible with advancements in machine learning, neural networks, and data-driven behavior modeling.

Human proficiency in reading and responding to social cues is fundamentally rooted in our ability to detect subtle motion patterns, postures, and spatiotemporal dynamics—a capability facilitated by our neural perceptual systems (Adolphs, 2009). Analogously, as machine learning architectures—particularly deep neural networks and probabilistic models—improve in their ability to encode, interpret, and react to subtle, low-level behavioral signals, proxemically-aware robots are poised to achieve greater behavioral fidelity, safety, and social acceptability in human environments. This is validated and informed by real world deployment, which drives a flywheel of observation, emulation, learning, new data collection, and dataset enhancement through observation, creating a never-ending loop of proxemic behavioral improvement. The need for this process of deployment validation and data structuring can't be overstated when it comes to social awareness in robots that are designed to be operating near humans.

7.1. Synthetic Data Augmentation for Behavioral Generalization

While computational advancements are critical, they must be complemented by robust, scalable, and diverse datasets to ensure safe and generalizable behavior modeling. To address limitations inherent to real-world data collection, we have developed a synthetic data generation pipeline that leverages foundational human behavior models and generative neural networks to produce large-scale, high-fidelity synthetic motion data.

By algorithmically recombining trial-type conditions and simulating novel interaction scenarios, this pipeline enables us to expand our behavioral dataset by orders of magnitude (approximately 500x growth relative to the original dataset). Critically, synthetic data allow us to introduce rare edge cases and non-normative social interactions that may be underrepresented in empirical studies but are vital for ensuring safe, generalizable, and bias-resilient robot behavior (Akhauri et al., 2020).

Moreover, synthetic data offer significant advantages in terms of privacy preservation. Although generated using distributions and patterns learned from real participant data, the synthetic datasets are fully decoupled from identifiable individuals, mitigating privacy concerns while preserving behavioral realism.

7.2. Crossing the Social Acceptance Threshold

Robotic systems capable of exhibiting socially- and spatially-appropriate behavior—particularly in the domains of proxemics, motion smoothness, and turn-taking—are key to overcoming longstanding barriers to human-robot coexistence, such as discomfort and social rejection. This extends beyond physical appearance; while humanoid morphology can exacerbate the "uncanny valley" phenomenon, research consistently demonstrates that motion characteristics and interaction fluency play a far more critical role in human acceptance of robots (Mori et al., 2012)(Takayama & Pantofaru, 2009).

By prioritizing behaviorally natural, socially-legible motion patterns over anthropomorphic form, robots that consider proxemics facilitate smoother, more intuitive interactions across a wide range of applications, accelerating public trust and adoption.

7.3. Integration and Cross-Domain Applications

Our behavioral models and control algorithms have been intentionally designed to be form-factor agnostic, facilitating integration into a wide range of robotic platforms, from mobile service robots to humanoids to micro-mobility systems. Proxemic awareness is not limited to a specific morphology but represents a core functional capability that will become increasingly essential as robots are deployed in workplaces, public spaces, healthcare settings, and private homes.

8. Alignment with the UN Sustainable Development Goals

The development of proxemic socially aware robotic systems directly supports several United Nations Sustainable Development Goals by advancing intelligent technologies that improve human well-being, enable responsible innovation, and facilitate equitable and inclusive urban environments. These systems leverage real-time perception, behavior modeling, and adaptive decision-making to engage meaningfully with diverse users and environments—paving the way for safer, more socially integrated autonomy.

8.1 SDG 3: Good Health and Well-Being

Robots that factor proxemics into their navigational decisions enhance safety, dignity, and autonomy in healthcare and assistive care settings by interpreting nonverbal cues, maintaining appropriate interpersonal distance, and respecting social boundaries. Their ability to perceive emotional states and adapt interaction strategies contributes to patient comfort and reduces caregiver burden—especially in eldercare, rehabilitation, and mental health support contexts. These capabilities enable scalable, high-quality care while supporting independent living and emotional well-being for vulnerable populations.

8.2 SDG 9: Industry, Innovation, and Infrastructure

Socially and spatially adaptive robotics represents a critical innovation layer in AI-enabled automation, advancing multimodal perception, real-time motion planning, and intuitive human-robot collaboration. The integration of social intelligence into robotic platforms expands their usability across industries—from logistics and retail to education and transportation—where context-sensitive interaction is essential. These systems enable smarter infrastructure, foster inclusive technology development, and catalyze new forms of sustainable service delivery.

8.3 SDG 11: Sustainable Cities and Communities

By incorporating proxemic awareness into navigation and decision-making, robots can operate safely and legibly in dense, dynamic public environments. Whether assisting pedestrians, navigating sidewalks, or supporting urban mobility services, socially and spatially intelligent agents reduce risk and increase public trust in autonomous technologies. These systems improve accessibility for users with diverse needs, respect localized social norms, and contribute to the development of inclusive, adaptive urban spaces.

9. Call to Action

As robotic systems become increasingly embedded in domestic, occupational, and public domains, their development must be guided by principles of ethical responsibility, inclusivity, and functional integration. We call on experts across fields—robotics, machine learning, human-computer interaction, ethics, behavioral science, and design—to actively shape the next generation of intelligent systems. These systems must not only interpret human behavior with precision, but also demonstrate context-aware and culturally sensitive responses in diverse, real-world environments.

One of the central technical challenges in advancing socially aware proxemics in robots is the availability and quality of representative, high-resolution behavioral datasets. Motion-based social interaction is multimodal and deeply context-dependent, encompassing subtle variations in gesture, proxemics, gaze, affective cues, and motion-based etiquette. To train robust, generalizable models, socially interactive robots must be grounded in datasets that are demographically diverse, culturally varied, and reflective of naturalistic human behavior. We encourage open, collaborative initiatives for expanding such datasets through multi-site data collection, structured annotation protocols, and realistic synthetic augmentation.

Moving forward, the field requires collective commitment to open research practices, shared evaluation frameworks, and principled design methodologies. Motion-based social interaction is an inherently human construct, and the ability to encode it computationally must remain a multidisciplinary effort. Through transparent tool sharing, rigorous validation, and attention to ethical considerations—including fairness, privacy, and societal impact—we can develop autonomous systems that are not only technically proficient, but socially responsible.

Ultimately, proxemic intelligence in robots should not merely be celebrated for its technical sophistication. It should be recognized as an artifact of global cooperation—embodying values of inclusivity, respect, and cohabitation—and serve as an active agent in supporting equitable, adaptive, and human-centered futures.

References

AbuJabal, N., Baziyad, M., Fareh, F., Brahmi, B., Rabie, T., & Bettayeb, M. (2024, December 18). A Comprehensive Study of Recent Path-Planning Techniques in Dynamic

Environments for Autonomous Robots. *Sensors*, *24*(24). https://doi.org/10.3390/s24248089

Acemoglu, D., & Restrepo, P. (2019, Spring). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, *33*(2), 3-30. DOI: 10.1257/jep.33.2.3

Adolphs, R. (2009). The Social Brain: Neural Basis of Social Knowledge. *Annual Review of Psychology*, *60*, 693-716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Akhauri, S., Zheng, L. Y., & Lin, M. C. (2020). Enhanced transfer learning for autonomous driving with systematic accident simulation. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5986-5993. https://doi.org/10.1109/IROS45743.2020.9341538

Boschetti, G., Sinico, T., & Trevisani, A. (2023). Improving Robotic Bin-Picking Performances through Human–Robot Collaboration. *Applied Sciences*, *13*(9). https://doi.org/10.3390/app13095429

Boston Planning & Development Agency Research Division. (n.d.). *Boston at a Glance - 2024*. City of Boston Planning Department. Retrieved July 24, 2025, from https://www.bostonplans.org/

Boyd, R., & Richerson, P. J. (2009, November 12). Culture and the evolution of human cooperation. *Philosophical transactions of the Royal Society B, Biological Sciences*, *364*(1533), 3281-3288. https://doi.org/10.1098/rstb.2009.0134

Breazeal, C. (2003, March 31). Toward Sociable Robots. *Robotics and Autonomous Systems*, *42*(3-4), 167-175. https://doi.org/10.1016/S0921-8890(02)00373-1

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). Social robotics. *Springer handbook of robotics*, 1935-1972. https://doi.org/10.1007/978-3-319-32552-1_72

Bütepage, J., & Kragic, D. (2017). Human-Robot Collaboration: From Psychology to Social Robotics. *arXiv preprint*. https://doi.org/10.48550/arXiv.1705.10146

Campagna, G., & Rehm, M. (2025, June). A Systematic Review of Trust Assessments in Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction*, *14*(2), 1-35. https://doi.org/10.1145/3706123

Dragan, A. D., Bauman, S., Forlizzi, J., & Srinivasa, S. S. (2015). Effects of Robot Motion on Human-Robot Collaboration. *HRI '15: Proceedings of the Tenth Annual ACM/IEEE Interactional Conference on Human-Robot Interaction*, 51-58. https://doi.org/10.1145/2696454.2696473

Eder, M., Könczöl, C., Kienzl, J., Mosbacher, J. A., Kubicek, B., & Steinbauer-Wagner, G. (2024). Influence of Different Explanation Types on Robot-Related Human Factors in Robot Navigation Tasks. *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, 1084-1091. https://doi.org/10.1109/RO-MAN60168.2024.10731192

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003, March 31). A Survey of Socially Interactive Robots. *Robotics and Autonomous Systems*, *42*(3-4), 143-166. https://doi.org/10.1016/S0921-8890(02)00372-X

Frith, C. D., & Frith, U. (2007, August 21). Social Cognition in Humans. *Current Biology*, *17*(16), 724-732. https://doi.org/10.1016/j.cub.2007.05.068

Goffman, E. (1967). *Interaction ritual: Essays in face-to-face behavior*. Routledge. https://doi.org/10.4324/9780203788387

Gross, J., & Vostroknutov, A. (2022, April). Why do people follow social norms? *Current Opinion in Psychology*, *44*, 1-6. https://doi.org/10.1016/j.copsyc.2021.08.016

Hall, E. T. (1966). *The Hidden Dimension*. Knopf Doubleday Publishing Group.

Hancock, P. A., Billings, D. R., Chen, J. Y.C., de Visser, E. J., & Parasuraman, R. (2011, September 16). A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors*, *53*(5), 517-527. https://doi.org/10.1177/0018720811417254

Heckman, J. J. (1979, January). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153-161. https://doi.org/10.2307/1912352

Hoffman, G., & Ju, W. (2014, February). Designing robots with movement in mind. *Journal of Human-Robot Interaction*, *3*(1), 91-122. https://doi.org/10.5898/JHRI.3.1.Hoffman

Jain, A., Bensal, R., Kumar, A., & Singh, K. (2015). A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *Int J Appl Basic Med Res*, *5*(2), 124-127. DOI: 10.4103/2229-516X.157168

Kaviani, G., Yarici, Y., & Kim, S. (2024). Hierarchical and Multimodal Data for Daily Activity Understanding. *arXiv*. https://doi.org/10.48550/arXiv.2504.17696

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters.* (Vol. 7). CUP Archive.

Kopp, T., Baumgartner, M., & Kinkel, S. (2021). Success factors for introducing industrial human-robot interaction in practice: an empirically driven framework. *The International Journal of Advanced Manufacturing Technology*, *112*, 685-704. https://doi.org/10.1007/s00170-020-06398-0

Kruse, T., Pandey, A. K., Alami, R., & Kirsch, A. (2013, December). Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, *61*(12), 1726-1743. https://doi.org/10.1016/j.robot.2013.05.007

Lindner, F., & Eschenbach, C. (2011). Towards a Formalization of Social Spaces for Socially Aware Robots. *Lecture Notes in Computer Science*, *6899*, 283-303. https://doi.org/10.1007/978-3-642-23196-4_16

Mainprice, J., & Berenson, D. (2013). Human-robot collaborative manipulation planning using early prediction of human motion. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 299-306. doi: 10.1109/IROS.2013.6696368

Mavrogiannis, C., Baldini, F., Wang, A., Zhao, D., Trautman, P., Steinfeld, A., & Oh, J. (2023, September). Core Challenges of Social Robot Navigation: A Survey. *ACM Transactions on Human-Robot Interaction*, *12*(3), 1-39. https://doi.org/10.1145/3583741

Mayo, E. (2004). *The human problems of an industrial civilization*. Routledge. https://doi.org/10.4324/9780203487273

Migration Policy Institute. (n.d.). *U.S. Immigrant Population by Metropolitan Area*. Migration Policy Institute. Retrieved July 24, 2025, from https://www.migrationpolicy.org/programs/data-hub/charts/us-immigrant-population-metropolitan-area

Mittelstadt, B. D., Allo, P., Floridi, L., Taddeo, M., & Wachter, S. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716679679

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98-100. https://doi.org/10.1109/MRA.2012.2192811

Möystad, O. (2017). *Cognition and the Built Environment*. Taylor & Francis Group.

Mutlu, B., & Forlizzi, J. (2008). Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. *HRI '08: Procedings of the 3rd ACM/IEEE international conference on human robot interaction*, 287-294. https://doi.org/10.1145/1349822.1349860

Nikolaidis, S., Lasota, P., Ramakrishnan, R., & Shah, J. (2015, November 16). Improved human–robot team performance through cross-training, an approach inspired by human team training practices. *The International Journal of Robotics Research*, *34*(14), 1711-1730. https://doi.org/10.1177/0278364915609673

Reddy, S., Dragan, A., & Levine, S. (2018). *Advances in Neural Information Processing Systems* (Vol. 31). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2018/file/6f2268bd1d3d3ebaabb04d6b5d099425-Paper.pdf

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. *HRI '15: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 141-148. https://doi.org/10.1145/2696454.2696497

Seethapathi, N., Wang, S., Saluja, R., Blohm, G., & Kording, K. P. (2019). Movement science needs different pose tracking algorithms. *arXiv preprint arXiv:1907.10226*. https://doi.org/10.48550/arXiv.1907.10226

Takayama, L., & Pantofaru, C. (2009). Influences on proxemic behaviors in human-robot interaction. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5495-5502. IEEE. 10.1109/IROS.2009.5354145

Taylor, A. T., Berrueta, T. A., & Murphey, T. D. (2021, August). Active learning in robotics: A review of control principles. *Mechatronics*, *77*, Article 102576. https://doi.org/10.1016/j.mechatronics.2021.102576

United Nations. (2015). *Transforming our World: The 2030 Agenda for Sustainable Development* [Sustainable Development Goals for 2030]. United Nations.

Urakami, J., & Seaborn, K. (2023). Nonverbal Cues in Human–Robot Interaction: A Communication Studies Perspective. *ACM Transactions on Human-Robot Interaction*, *12*(2), 1-21. https://doi.org/10.1145/3570169

Weiss, J. (2024, September 12). The Supermarket Superstardom of Marty the Stop & Shop Robot. *Boston Magazine*. https://www.bostonmagazine.com/news/2024/09/12/marty-the-robot-stop-and-shop/

Wurman, P. R., D'Andrea, R., & Mountz, M. (2008). Coordinating Hundreds of Cooperative, Autonomous Vehicles in Warehouses. *AI Magazine*, *29*(1). https://doi.org/10.1609/aimag.v29i1.208