CSSR-Nav: Culturally Sensitive Social Robot Navigation Using Open-Source Vision-Language Model and Empirical Cultural Knowledge

Birhanu Shimelis Girma[1 [0000−0003−4857−2262]], Ibrahim Jimoh[1 [0009−0003−1358−5271]], Yohannes Haile[1 [0009−0001−0241−2426]], Assane Gueye[1 [0000−0001−6469−4716]], Moise Busogi[1 [0000−0002−5245−0113]], and David Vernon[1 [0000−0002−9782−3788]]

Carnegie Mellon University Africa, Kigali, Rwanda
{bgirmash, ioj}@alumni.cmu.edu, {yohanneh, assaneg, mbusogi}@andrew.cmu.edu, david@vernon.eu

Abstract. We present CSSR-Nav (Culturally Sensitive Social Robot Navigation), a vision-language-model-based approach for culturally grounded social robot navigation. Unlike existing methods that assume Western etiquette or rely on datasets collected in Western environments, CSSR-Nav derives navigation constraints from empirical cultural knowledge obtained through a survey of 143 Rwandan respondents and uses an open-source vision-language model (LLaVA) deployed on Jetson Orin Nano edge hardware for social-context perception. We introduce the *Rwandan Social Costmap Layer*, which integrates culturally informed social constraints into a standard ROS navigation stack. The current implementation supports personal-space maintenance, conversation non-interruption, and directional preference, while elder-respect behavior is implemented but remains under limited live validation. In a five-scenario, twenty-five-trial small-scale evaluation, CSSR-Nav reached the goal in all trials, maintained encounter distances above the 1.0 m personal-space norm, and recorded no violation events for the active norms. The vision-language component operated asynchronously, with a median end-to-end latency of 11.6 s. These results indicate that CSSR-Nav can support culturally informed navigation under the present experimental protocol, while broader statistical validation, dynamic-pedestrian scenarios, and Rwandan user-perception studies remain future work. To our knowledge, this is the first integration of empirically derived African cultural norms into autonomous social robot navigation.

Keywords: Social Robot Navigation · Vision-Language Models · Cultural Robotics · Human-Robot Interaction · Costmap Layers

1 Introduction

Social robots are increasingly deployed in human-centric environments such as hospitals, hotels, museums, and educational institutions [8, 11, 16, 20, 29]. For these robots to be effective and accepted, they must navigate not only safely but

also in ways that respect the social norms and cultural expectations of the people they serve [17]. However, the vast majority of social navigation research has been conducted in Western contexts, with social compliance defined according to Western etiquette [5].

Recent taxonomies in social navigation research have established fundamental properties that characterize socially aware robotic agents [22]. A robot demonstrates social awareness when it: (1) recognizes human agents as distinct entities deserving prioritized safety considerations; (2) operates in ways that minimize disruption, discomfort, and confusion for nearby people; (3) communicates its navigational intentions, whether through explicit signals or implicit behavioral cues; and (4) when faced with spatial conflicts, evaluates the social context and adopts resolution strategies that may prioritize human comfort over task efficiency. While these properties provide a valuable framework for social navigation, they assume culturally-neutral definitions of "discomfort" and "social manner", assumptions we challenge in this work.

The challenge of cultural specificity in social navigation: Social norms vary significantly across cultures [6]. Behaviors considered polite in one culture may be perceived as rude or inappropriate in another. For example, the appropriate interpersonal distance, eye contact patterns, and passing behaviors differ substantially between cultures [28]. Yet, current social navigation systems either encode generic rules (*e.g.*"pass on the right") or learn from datasets collected predominantly in Western environments [9, 18], limiting their applicability to diverse global contexts.

The gap in African social robotics: Despite growing interest in deploying social robots across Africa for applications ranging from healthcare to education [26], there exists no navigation system specifically designed to respect African cultural norms. The CSSR4Africa (Culturally Sensitive Social Robotics for Africa) project addresses this gap by first establishing what constitutes culturally appropriate behavior through empirical research [26].

Limitations of existing VLM-based approaches: Recent work has demonstrated the potential of Vision-Language Models (VLMs) for social navigation [24]. However, these approaches rely on proprietary cloud-based models (*e.g.* GPT-4V), which present challenges for deployment in regions with limited internet connectivity, raise data privacy concerns, and incur ongoing operational costs. Furthermore, they encode social norms through prompts based on assumed universal etiquette rather than empirically validated cultural knowledge.

Main Contributions. In this paper, we present CSSR-Nav, a culturally sensitive social navigation system that addresses these limitations. Our main contributions include:

1. **Empirically-grounded cultural navigation:** This work is the first to integrate empirically-derived cultural norms (from a survey of 143 Rwandan respondents) into an autonomous social navigation system, encoding constraints for personal space maintenance, conversation non-interruption, directional preference, and elder respect.

2. **Rwandan Social Costmap Layer:** We introduce a ROS navigation stack plugin that translates cultural norms into costmap modifications, enabling standard path planners to generate culturally-appropriate trajectories without algorithm changes.
3. **Edge-deployed open-source VLM:** We demonstrate social-context perception using LLaVA on a Jetson Orin Nano, achieving practical inference times without cloud dependencies and supporting deployment in resource-constrained environments.
4. **Small-scale evaluation and reproducible analysis pipeline:** We present a five-scenario, twenty-five-trial evaluation and a rosbag-based analysis workflow to measure task completion, social-distance behavior, norm-related events, and VLM latency.

2 Related Work

2.1 Social Robot Navigation

Social navigation extends traditional robot navigation by considering humans not merely as dynamic obstacles but as social entities with expectations about robot behavior [12, 16]. Research has addressed various aspects including personal space [10], passing behavior [23], and group awareness [25].

Learning-based approaches have gained prominence, with imitation learning [7, 19] and reinforcement learning [3, 14] showing promise. However, these methods require extensive training data and may not generalize to contexts different from their training distribution. The SCAND dataset [9] and MuSoHu [18] provide valuable resources but were collected exclusively in Western environments.

2.2 Vision-Language Models for Navigation

VLMs have emerged as powerful tools for robotic decision-making due to their contextual understanding and commonsense reasoning capabilities [2]. LM-Nav [21] demonstrated outdoor navigation using GPT-3 and CLIP. VLM-Social-Nav [24] applied GPT-4V to social navigation, achieving significant improvements over baseline methods.

However, reliance on proprietary cloud models limits practical deployment. Recent open-source VLMs such as LLaVA [15] offer comparable capabilities with the advantage of local deployment, though their application to social navigation remains unexplored.

2.3 Culture and Social Robotics

The influence of culture on human-robot interaction has been documented across dimensions including proxemics [4], communication style [27], and social role expectations [13]. However, this knowledge has rarely been operationalized in navigation systems.

Table 1: Rwandan Cultural Norms Relevant to Navigation and Their Implementation Status

ID	Cultural Norm	Implementation Status
3-1	Maintain distance of one meter or less when passing	Personal space radius Active
2-26	Do not walk between conversing people	Group interaction cost Active
3-3	Pass behind groups of people	Directional preference Partial
2-27	Do not walk ahead of elders	Elder respect zone Implemented

The CSSR4Africa project [26] represents a systematic effort to establish cultural requirements for social robots in African contexts. This work builds directly on the Rwandan Cultural Knowledge survey [1], translating empirical findings into computational navigation behaviors.

3 Rwandan Cultural Knowledge for Navigation

This approach is grounded in empirical cultural knowledge derived from the CSSR4Africa Rwandan Cultural Knowledge Survey [1]. This survey collected responses from 143 participants at Carnegie Mellon University Africa, establishing consensus on 57 cultural behaviors relevant to social robot interaction.

3.1 Navigation-Relevant Cultural Norms

From the survey results, multiple norms are identified with direct implications for robot navigation. Table 1 summarizes the key norms, their implementation in CSSR-Nav, and their current status in the system. The present implementation actively supports personal-space maintenance and conversation non-interruption, while elder-respect behavior is implemented with limited live validation. Other survey-derived norms, including approaching before greeting (Norm 2-11), initiating interaction with a courteous greeting (Norm 2-18), and bowing slightly when greeting (Norm 2-9), are handled by existing CSSR4Africa behaviour controller and gesture execution nodes and are therefore outside the scope of the present navigation-layer implementation and evaluation [1, 26].

3.2 Formalizing Cultural Costs

Following the formulation in [24], we define the navigation cost function as:

$$\mathcal{C}(\mathbf{s}, \mathbf{a}) = \alpha \cdot \mathcal{C}_{\text{goal}} + \beta \cdot \mathcal{C}_{\text{obst}} + \gamma \cdot \mathcal{C}_{\text{cultural}} \tag{1}$$

where $\mathcal{C}_{\text{goal}}$ encourages movement toward the goal, $\mathcal{C}_{\text{obst}}$ discourages collisions, and $\mathcal{C}_{\text{cultural}}$ encourages adherence to cultural norms. Unlike previous work that

defines $\mathcal{C}_{\text{social}}$ generically, we decompose $\mathcal{C}_{\text{cultural}}$ into specific, empirically-grounded components:

$$\mathcal{C}_{\text{cultural}} = \mathcal{C}_{\text{space}} + \mathcal{C}_{\text{group}} + \mathcal{C}_{\text{elder}} + \mathcal{C}_{\text{keepright}} \tag{2}$$

Each component corresponds to a specific cultural norm: $\mathcal{C}_{\text{space}}$ for personal space maintenance (Norm 3-1), $\mathcal{C}_{\text{group}}$ for conversation non-interruption (Norm 2-26), $\mathcal{C}_{\text{elder}}$ for elder respect positioning (Norm 2-27), and $\mathcal{C}_{\text{keepright}}$ for a directional preference derived from Norm 3-3. In the present implementation, Norm 3-3 is represented as a directional passing preference rather than a full group-aware pass-behind behavior.

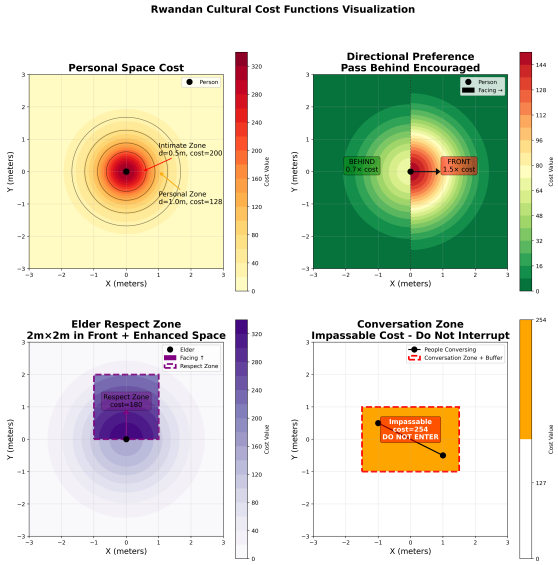

Fig. 1: Rwandan Social Costmap Layer integrates with ROS costmap_2d through standard layer interfaces (initialize, updateBounds, updateCosts) and implements four cultural cost functions: personal space, directional preference, elder respect zones, and conversation interruption avoidance.

4 System Architecture

Fig. 2 illustrates the CSSR-Nav system architecture. The approach integrates three key components: LLaVA-based social perception, the Rwandan Social Costmap Layer, and culture-aware behavior coordination.

4.1 Hardware Platform

Our system is deployed on a SoftBank Pepper robot augmented with: Intel RealSense D435i camera providing color images at 640×480 resolution and depth

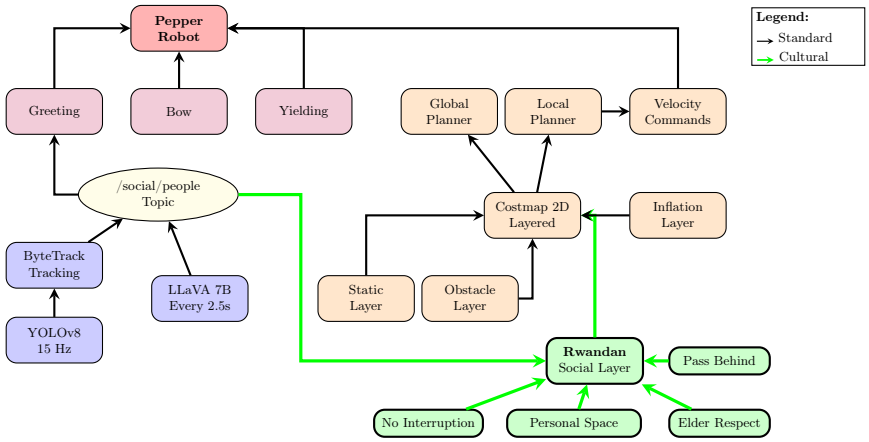

Fig. 2: System architecture showing the integration of perception, navigation (Costmap 2D with Rwandan Social Layer), and culturally-aware behavior coordination.

information for social perception; YDLidar G4, 2D LiDAR for obstacle detection and localization; and NVIDIA Jetson Orin Nano, Edge computing platform running perception and VLM inference.

This configuration enables fully edge-deployed processing without cloud dependencies, addressing connectivity limitations common in many deployment environments. Table 2 details the hardware specifications.

Table 2: Hardware Platform Specifications

Component	**Model**	**Specifications**
Robot Base	Pepper (SoftBank)	Height: 1.21m, Weight: 28kg
RGB-D Camera	RealSense D435i	640×480 @ 30fps
LiDAR	YDLidar G4	360°, 10m range, 5-12Hz
Compute	Jetson Orin Nano	8GB RAM, 1024-core GPU

4.2 LLaVA-based Social Perception

The system employs LLaVA (Large Language and Vision Assistant) [15] for social scene understanding. Unlike VLM-Social-Nav which uses GPT-4V through cloud API, we run LLaVA locally on the Jetson Orin Nano using 4-bit quantization for memory efficiency.

Perception Pipeline. To support continuous perception on edge hardware, the system adopts a two-stage approach: (1) Fast Detection (YOLOv8), which runs continuously at 15 Hz to detect and track people without VLM latency; and (2) Social Understanding (LLaVA), which is triggered every 2.5 seconds when social entities are detected and provides deeper social-context information including age estimation, group detection, and activity recognition.

Algorithm 1 details the perception pipeline that fuses fast detection with periodic VLM queries.

Algorithm 1 Two-Stage Social Perception Pipeline

Require: RGB-D frame I_t, depth frame D_t
Ensure: Person detections $\mathcal{P}_t$, social groups $\mathcal{G}_t$
$\mathcal{P}_t \leftarrow \text{YOLOv8}(I_t)$ {Fast detection at 15Hz}
Track persons across frames using IoU matching
if $t - t_{\text{last\_vlm}} > \Delta t_{\text{vlm}}$ AND $|\mathcal{P}_t| > 0$ **then**
  $\mathcal{S} \leftarrow \text{LLaVA}(I_t, \text{cultural\_prompt})$ {VLM query}
  Extract age groups, facing directions from $\mathcal{S}$
  $\mathcal{G}_t \leftarrow \text{DetectGroups}(\mathcal{P}_t, \mathcal{S})$
  $t_{\text{last\_vlm}} \leftarrow t$
end if
Estimate 3D positions using depth D_t and camera intrinsics
return $\mathcal{P}_t$, $\mathcal{G}_t$

Culturally-Adapted Prompting. A key aspect of this approach is the integration of Rwandan cultural norms directly into VLM prompts. Fig. 3 shows our prompt structure, which differs from generic social navigation prompts by explicitly encoding surveyed cultural expectations while extracting only the social attributes required by the navigation layer.

Input Image: [RGB frame from RealSense or Pepper built-in camera]
System Prompt: You are the vision system for a social robot navigating in Carnegie Mellon University Africa's AI and Robotics lab while respecting Rwandan social norms. Analyze this scene following Rwandan cultural norms.
Cultural Guidelines:

- Maintain at least 1 meter distance when passing
- Never walk between people who are conversing
- Prefer socially appropriate passing behavior in head-on encounters
- Do not walk ahead of elderly persons

Output Format (JSON only):

```
{"people":[{"horizontal":"left/center/right",
"age_group":"child/adult/elder",
"facing":"left/right/front/back",
"is_conversing":false}],"elders_present":false}
```

Fig. 3: Culturally-adapted prompt for LLaVA. Unlike generic prompts, it explicitly encodes Rwandan cultural expectations while requesting only the social attributes required by the navigation layer.

Social-context inference is asynchronously decoupled from the navigation control loop. On the Jetson Orin Nano, LLaVA-7B v1.6 (4-bit quantized via Ollama) produced a median end-to-end latency of approximately 11.6 s over the recorded sweep (Section 5), while navigation control continued at 8 Hz. The fusion node maintains cached VLM-derived tags, including `age_group`,

`is_conversing`, and `facing`, and refreshes constraint publication independently of VLM response time. As a result, the local planner does not stall while awaiting semantic updates.

4.3 Rwandan Social Costmap Layer

The core technical contribution of this work is the Rwandan Social Costmap Layer, a plugin for the ROS navigation stack that translates cultural norms into costmap modifications. Algorithm 2 presents the costmap update procedure.

Algorithm 2 Rwandan Social Costmap Update

Require: People $\mathcal{P}$, Groups $\mathcal{G}$, Costmap C
Ensure: Updated costmap C'
$C' \leftarrow C$ {Initialize with base costmap}
for each person $p \in \mathcal{P}$ **do**
  Apply personal space cost (Eq. 3) around p
  if p is elder **then**
    Apply elder respect zone cost (Eq. 7) ahead of p
    Multiply personal space by $\mu_{\text{elder}} = 1.2$
  end if
  Apply directional preference cost (Eq. 6) based on facing
end for
for each group $g \in \mathcal{G}$ **do**
  if g is conversing **then**
    Compute interaction zone $\mathcal{Z}_{\text{int}}$ (convex hull)
    Apply impassable cost (Eq. 4) to $\mathcal{Z}_{\text{int}}$
  end if
end for
return C'

Personal Space Cost. For each detected person at position (x_p, y_p), we apply a Gaussian cost distribution:

$$\mathcal{C}_{\text{space}}(x, y) = c_{\text{intimate}} \cdot e^{-\frac{d^2}{2\sigma_i^2}} + c_{\text{personal}} \cdot e^{-\frac{d^2}{2\sigma_p^2}} \tag{3}$$

where $d = \sqrt{(x - x_p)^2 + (y - y_p)^2}$, $\sigma_i = 0.5$m (intimate space), $\sigma_p = 1.0$m (personal space per Norm 3-1), $c_{\text{intimate}} = 200$, and $c_{\text{personal}} = 128$.

Group Interaction Cost. When a conversing group is detected, we apply near-impassable cost to the interaction zone:

$$\mathcal{C}_{\text{group}}(x, y) = \begin{cases} 254 & \text{if } (x, y) \in \mathcal{Z}_{\text{interaction}} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $\mathcal{Z}_{\text{interaction}}$ is the convex hull of group members expanded by a 0.5m buffer distance.

Table 3: Rwandan Social Costmap Layer Parameters

Parameter	Value	Norm	Justification
Personal space radius	1.0m	3-1	Survey consensus
Intimate zone cost	200	3-1	High discouragement
Personal zone cost	128	3-1	Moderate cost
Conversation zone cost	254	2-26	Near-impassable (avoid)
Front cost multiplier	1.5	3-3	Discourage frontal pass
Back cost multiplier	0.7	3-3	Encourage rear pass
Elder zone depth	2.0m	2-27	Respectful distance
Elder zone cost	180	2-27	Strong discouragement
Elder space multiplier	1.2	2-27	Extra personal space

Directional Preference. To encourage passing behind rather than in front of people, we apply an asymmetric cost based on a Gaussian distribution centered on the person:

$$\mathcal{C}_{\text{base}}(x, y) = 100 \cdot e^{-\frac{d^2}{2 \cdot 1.0^2}} \tag{5}$$

where $d = \sqrt{(x - x_p)^2 + (y - y_p)^2}$ is the distance from the person. This base cost is then scaled according to the robot's position relative to the person's facing direction:

$$\mathcal{C}_{\text{direction}}(x, y) = \mathcal{C}_{\text{base}}(x, y) \cdot \begin{cases} 1.5 & \text{if in front half} \\ 0.7 & \text{if in back half} \end{cases} \tag{6}$$

where front/back is determined by the person's estimated facing direction from LLaVA.

Elder Respect Zone. When an elder is detected (estimated via LLaVA), the system applies additional cost to the zone ahead of them:

$$\mathcal{C}_{\text{elder}}(x, y) = 180 \cdot \mathbf{1}_{(x,y) \in \mathcal{Z}_{\text{front}}} \tag{7}$$

where $\mathcal{Z}_{\text{front}}$ is a yaw-aligned rectangular zone extending in the elder's facing direction. In the present implementation, the zone begins 0.2 m in front of the elder, extends 2.0 m forward, and spans 1.6 m in width.

Table 3 summarizes all costmap layer parameters and their cultural justifications.

4.4 Culture-Aware Behavior Coordination

Beyond navigation, Rwandan cultural norms specify interaction behaviors. The system includes a behavior coordinator that triggers appropriate actions: (1) Greeting initiation: When approaching a person within 2.5m, the robot initiates a verbal greeting in English or Kinyarwanda; (2) Respectful gestures: 15° bow for general greetings, 30° bow when greeting elders; and (3) Yielding to elders: Robot stops when elder detected 3m ahead, allows passage.

5 Experimental Evaluation

CSSR-Nav is evaluated through a five-scenario, twenty-five-trial small-scale ablation study in a constrained indoor laboratory setting. User-perception studies are left for future work.

5.1 Experimental Setup

Scenario. Experiments were conducted in a laboratory setting in which the robot navigated a short route within an operational area of approximately $1.5 \times 1.5\,\mathrm{m}$ while interacting with a collaborator positioned to elicit socially relevant constraints. Five conditions were evaluated: baseline, `n31_only`, `n226_only`, `n227_only`, and `all_norms`, with five trials per condition.

Baseline Methods. The study compares CSSR-Nav against DWA without the cultural cost layer, while keeping the planner, localisation, and perception pipeline unchanged.

Metrics. The system is evaluated using task success, time to goal, path length, encounter-window minimum human distance, per-norm violation events, and VLM latency. These metrics capture navigation completion, efficiency, social-distance maintenance, norm compliance, and the timing characteristics of the semantic-perception module. All metrics are computed from recorded rosbag data using a reproducible analysis pipeline.

5.2 Ablation Study Design

To evaluate the individual and combined contributions of the implemented cultural norms, we conducted a five-scenario, twenty-five-trial small-scale ablation study, summarized in Table 4. Each condition used the same navigation route and start/goal configuration, while varying the active cultural constraints. Five trials were recorded for each condition.

Table 4: Ablation Test Conditions

ID	Condition	Active Norms	Purpose
00	Baseline	None	Standard navigation comparison
01	Personal Space Only	3-1	Individual norm effect
02	Conversation Groups Only	2-26	Individual norm effect
03	Elder Respect Only	2-27	Individual norm effect
04	All Norms	All norms	Combined cultural effect

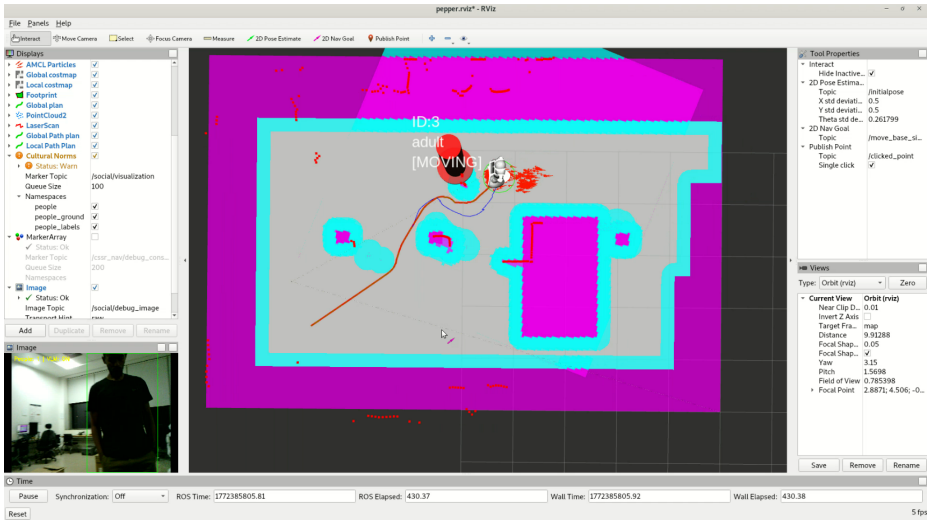

Fig. 4: Personal space detection and rerouting of the planned path to avoid a moving person.

Table 5: Ablation Study Results: mean values over five trials per condition. Encounter distance is computed over the human–robot encounter window.

Condition	Success (%)	Time (s)	Path (m)	Encounter Dist. (m)
Baseline	100.0	17.82	1.62	2.83
Personal Space Only	100.0	12.02	1.35	2.71
Conversation Groups Only	100.0	16.57	1.68	2.27
Elder Respect Only	100.0	14.64	1.25	2.65
All Norms	100.0	14.75	1.24	2.59

5.3 Quantitative Results

Table 5 summarizes the mean results over five trials per condition. Because this is a small-scale study, we report descriptive statistics rather than significance tests. Fig. 4 illustrates personal-space detection and path adaptation during navigation.

The results support four main observations. First, CSSR-Nav reached the goal in all trials across all five conditions. Second, path length and navigation time remained bounded across conditions, with mean path lengths between 1.24 m and 1.68 m and mean completion times between 12.02 s and 17.82 s. Third, encounter-window minimum distances ranged from 2.27 m to 2.83 m, remaining above the 1.0 m personal-space norm derived from the survey. Fourth, no violation events were recorded for any active norm during the study. Aggregated over the full evaluation, the VLM component produced a median end-to-end latency of 11.6 s while remaining asynchronously decoupled from the navigation control loop. Fuller live validation of conversation-pair and elder-specific triggers will require broader multi-person and elder-participant scenarios.

6 Discussion

6.1 Key Findings and Implications

The evaluation indicates that CSSR-Nav can incorporate culturally derived navigation constraints while maintaining reliable task completion in a constrained indoor setting. Because the present study is small-scale, these findings should be interpreted as indicative rather than conclusive.

Personal-space maintenance. CSSR-Nav maintained encounter-window minimum distances between 2.27 m and 2.83 m across all five conditions, remaining above the 1.0 m personal-space norm derived from the survey. No violation events were recorded for any active norm during the evaluation. This suggests that the Rwandan Social Costmap Layer can implement culturally appropriate proximity behavior while preserving respectful separation from nearby people.

Navigation efficiency and reliability. Mean path length ranged from 1.24 m to 1.68 m, and mean completion time ranged from 12.02 s to 17.82 s across the five conditions. All trials reached the goal, and no trial entered planner recovery behavior. These results indicate that the cultural cost layer can be integrated into the navigation stack without producing unstable behavior or excessive detours under the present protocol.

The vision-language component operated asynchronously, with a median end-to-end latency of 11.6 s, while the navigation loop continued independently of semantic update latency. This supports the practical feasibility of edge deployment despite relatively slow semantic inference.

Scope of the present evaluation. The present study provides direct live evidence for personal-space maintenance and directional preference. However, it did not produce live activations of the conversation-pair and elder-specific triggers, so fuller validation of conversation non-interruption and elder-respect constraints remains future work.

6.2 Cultural Context and Generalizability

Our approach explicitly encodes Rwandan cultural norms, raising questions about generalization. We argue that *cultural specificity is a feature, not a limitation*: robots deployed in different cultural settings should reflect local expectations rather than a universal social-navigation model.

The framework is generalizable through parametric adaptation. Different cultural contexts can be represented by adjusting the underlying cost functions and parameters rather than replacing the navigation architecture.

6.3 Limitations and Future Work

Several limitations remain. First, the present evaluation includes only five trials per condition, so larger studies are needed for stronger empirical conclusions. Second, the study was conducted in a small indoor area. Third, the current study evaluates robot-side outcomes only; user-perception studies with Rwandan participants remain an important next step.

In addition, greeting, bowing, and related interaction behaviors were outside the scope of the present navigation-layer evaluation, and Norm 3-3 is currently represented as a directional passing preference rather than a full group-aware pass-behind behavior. Future work will therefore focus on larger and more diverse evaluations, fuller live validation of the implemented norms, user-perception studies, and cross-cultural re-parameterisation of the framework.

7 Conclusion

We presented CSSR-Nav, a culturally sensitive social navigation system that integrates empirically derived Rwandan cultural norms with vision-language-model-based social perception. CSSR-Nav combines the *Rwandan Social Costmap Layer* with edge-deployed LLaVA on Jetson Orin Nano hardware to incorporate culturally informed constraints into a standard ROS navigation stack.

In a five-scenario, twenty-five-trial small-scale evaluation, CSSR-Nav reached the goal in all trials, maintained encounter-window minimum distances above the 1.0 m personal-space norm derived from the survey, and recorded no violation events for the active norms. The vision-language component operated asynchronously with a median end-to-end latency of 11.6 s, allowing navigation control to continue independently of semantic update latency. These results demonstrate that culturally informed navigation is feasible on edge hardware in a constrained indoor setting.

CSSR-Nav shows that empirically grounded cultural constraints can be integrated into a standard navigation stack without sacrificing practical task completion. This work provides a practical framework for adapting social navigation to culturally specific settings through parameterized costmap design grounded in local social norms.

Future work will include larger and more diverse evaluations, fuller live validation of conversation and elder-specific triggers, Rwandan user-perception studies, and broader integration of interaction-level cultural behaviors.

Acknowledgments. This research was carried out as part of the Afretec Network. Afretec is managed by Carnegie Mellon University Africa and receives financial support from the Mastercard Foundation. This work is part of the CSSR4Africa project (Culturally Sensitive Social Robotics for Africa). We thank all survey participants who contributed their cultural knowledge.

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

References

[1] D1.2 Rwandan Cultural Knowledge. https://cssr4africa.github.io/deliverables/CSSR4Africa_Deliverable_D1.2.pdf, [Accessed 01-02-2026]

[2] Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Chen, X., Choromanski, K., Ding, T., Driess, D., Dubey, A., Finn, C., Florence, P., Fu, C., Arenas, M.G., Gopalakrishnan, K., Han, K., Hausman, K., Herzog, A., Hsu, J., Ichter, B., Irpan, A., Joshi, N., Julian, R., Kalashnikov, D., Kuang, Y., Leal, I., Lee, L., Lee, T.W.E., Levine, S., Lu, Y., Michalewski, H., Mordatch, I., Pertsch, K., Rao, K., Reymann, K., Ryoo, M., Salazar, G., Sanketi, P., Sermanet, P., Singh, J., Singh, A., Soricut, R., Tran, H., Vanhoucke, V., Vuong, Q., Wahid, A., Welker, S., Wohlhart, P., Wu, J., Xia, F., Xiao, T., Xu, P., Xu, S., Yu, T., Zitkovich, B.: RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control

[3] Chen, C., Liu, Y., Kreiss, S., Alahi, A.: Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning. In: IEEE International Conference on Robotics and Automation (ICRA) (2019)

[4] Eresha, G., et al.: Investigating the influence of culture on autonomous robot navigation. Workshop on Socially Intelligent Robots at IEEE International Conference on Social Robotics (2013)

[5] 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., Mirksy, 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., Martín-Martín, R.: Principles and Guidelines for Evaluating Social Robot Navigation Algorithms (Sep 2023). https://doi.org/10.48550/arXiv.2306.16740

[6] Hall, E.T.: The hidden dimension. Doubleday (1966)

[7] Hirose, N., Shah, D., Sridhar, A., Levine, S.: Sacson: Scalable autonomous control for social navigation. In: IEEE International Conference on Robotics and Automation (ICRA) (2023)

[8] Kabacińska, K., Teng, K.A., Robillard, J.M., Kabacińska, K., Teng, K.A., Robillard, J.M.: Social Robot Interactions in a Pediatric Hospital Setting: Perspectives of Children, Parents, and Healthcare Providers. Multimodal Technologies and Interaction **9**(2) (Feb 2025). https://doi.org/10.3390/mti9020014

[9] Karnan, H., Nair, A., Xiao, X., Garrett, G., Warnell, G., Socrates, S., Stone, P.: Socially compliant navigation dataset (SCAND): A large-scale dataset of demonstrations for social navigation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2022)

[10] Kirby, R., Simmons, R., Forlizzi, J.: Companion: A constraint-optimizing method for person-acceptable navigation. In: IEEE International Symposium on Robot and Human Interactive Communication (2009)

[11] Kok, C.L., Ho, C.K., Teo, T.H., Kato, K., Koh, Y.Y., Kok, C.L., Ho, C.K., Teo, T.H., Kato, K., Koh, Y.Y.: A Novel Implementation of a Social Robot for Sustainable Human Engagement in Homecare Services for Ageing Populations. Sensors **24**(14) (Jul 2024). https://doi.org/10.3390/s24144466

[12] Kruse, T., Pandey, A.K., Alami, R., Kirsch, A.: Human-aware robot navigation: A survey. Robotics and Autonomous Systems **61**(12), 1726–1743 (2013)

[13] Li, D., et al.: Cross-cultural studies in HRI: A comparison of Chinese and German users. ACM/IEEE International Conference on Human-Robot Interaction (2010)

[14] Liang, J., et al.: Crowd-steer: Realtime smooth and collision-free robot navigation in densely crowded scenarios trained using high-fidelity simulation. arXiv preprint arXiv:2005.03178 (2021)

[15] Liu, H., Li, C., Wu, Q., Lee, Y.J.: Visual instruction tuning. Advances in neural information processing systems **36** (2024)

[16] Mavrogiannis, C., Baldini, F., Wang, A., Zhao, D., Trautman, P., Steinfeld, A., Oh, J.: Core challenges of social robot navigation: A survey. ACM Transactions on Human-Robot Interaction **12**(3), 1–39 (2023)

[17] Mirsky, R., Xiao, X., Hart, J., Stone, P.: Conflict Avoidance in Social Navigation—a Survey. J. Hum.-Robot Interact. **13**(1), 13:1–13:36 (Mar 2024). https://doi.org/10.1145/3647983

[18] Nguyen, A., et al.: Musohu: Multi-stage social human-in-the-loop dataset for robot navigation. In: IEEE International Conference on Robotics and Automation (ICRA) (2023)

[19] Raj, A., et al.: Targeted navigation with human-aware robot navigation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2024)

[20] Roštšinskaja, A., Saard, M., Korts, L., Kööp, C., Kits, K., Loit, T.L., Juhkami, J., Kolk, A.: Unlocking the Potential of Social Robot Pepper: A Comprehensive Evaluation of Child-Robot Interaction. Journal of Pediatric Health Care **39**(4), 572–584 (2025). https://doi.org/https://doi.org/10.1016/j.pedhc.2025.01.010

[21] Shah, D., Osinski, B., Ichter, B., Levine, S.: LM-Nav: Robotic navigation with large pre-trained models of language, vision, and action. In: Conference on Robot Learning (2022)

[22] Singamaneni, P.T., Bachiller-Burgos, P., Manso, L.J., Garrell, A., Sanfeliu, A., Spalanzani, A., Alami, R.: A survey on socially aware robot navigation: Taxonomy and future challenges. The International Journal of Robotics Research **43**(10), 1533–1572 (Sep 2024). https://doi.org/10.1177/02783649241230562, publisher: SAGE Publications Ltd STM

[23] Sisbot, E.A., Marin-Urias, L.F., Alami, R., Simeon, T.: A human aware mobile robot motion planner. In: IEEE Transactions on Robotics. vol. 23, pp. 874–883 (2007)

[24] Song, B., et al.: Socially-aware robot navigation with vision-language models. In: IEEE International Conference on Robotics and Automation (ICRA) (2024)
[25] Truong, X.T., Ngo, T.D.: Toward socially aware robot navigation in dynamic and crowded environments. In: IEEE International Conference on Information and Automation (ICIA) (2017)
[26] Vernon, D., et al.: Culturally Sensitive Social Robotics for Africa (CSSR4Africa:) (2024), http://www.cssr4africa.org
[27] Wang, L., et al.: Should a robot behave like a human when speaking to a person? International Journal of Social Robotics (2010)
[28] Watson, O.M.: Proxemic behavior: A cross-cultural study. Mouton de Gruyter (1970)
[29] Zhao, I.Y., Leung, A.Y.M., Huang, Y., Liu, Y.: A Social Robot in Home Care: Acceptability and Utility Among Community-Dwelling Older Adults. Innovation in Aging **9**(5), igaf019 (May 2025). https://doi.org/10.1093/geroni/igaf019