Project Lumina : Guiding Light for the Visually Impaired

Manasa Mangipudi
Sreeniketh Aathreya
Suujay Dhhoka

1 May 2024

1 Abstract

Navigating through crowded environments while visually impaired poses significant challenges. Existing assistive technologies often focus on obstacle avoidance but do not account for social interactions or guide users to specific goals. We propose this solution, an intelligent navigation aid that leverages recent advances in robot crowd navigation, human trajectory prediction, and object detection to enable safer and more socially aware guidance.

Our system uses a graph neural network with attention mechanisms to model the complex interactions between the user, other pedestrians, and the environment. By predicting the future trajectories of nearby pedestrians, the system anticipates their intended paths and navigates the user accordingly, avoiding socially awkward or dangerous situations. The system also detects objects of interest in the user's surroundings using the YOLOv3 architecture and provides audio guidance to navigate to them.

We adopt a two-pronged approach in our design: a technology-only solution focused on efficient navigation, and a socially cognizant solution that prioritizes safety and social norms. Through comparative analysis, we highlight the potential pitfalls of a purely technological approach and demonstrate the benefits of incorporating social awareness.

This research draws insights from a comprehensive review of related work across multiple domains. Our key innovation lies in integrating state-of-the-art techniques in crowd navigation, trajectory prediction, and object detection to create a socially aware navigation aid, an area that has received limited attention in prior literature.

This project aims to enhance the independence and quality of life for visually impaired individuals while promoting more natural human-robot interaction. We discuss the broader societal implications of our work and outline best practices for the responsible deployment of such assistive technologies. Through this, we hope to contribute to a future where intelligent navigation aids are not only effective but also socially inclusive.

2 Related Work

Our project builds upon prior work in robot crowd navigation, human trajectory prediction, object detection, and assistive technologies for the visually impaired.

For crowd navigation, early methods used hand-crafted interaction rules, such as social forces

[3] or velocity obstacles [4]. More recently, learning-based approaches have shown promising results. Chen et al. [5] used deep reinforcement learning with attention to model robot-human interactions. Liang et al. [6] proposed a graph neural network to capture human-human and robot-human interactions. We adopt a similar graph-based approach but incorporate human trajectory predictions for improved social awareness. Human trajectory prediction has been studied extensively.

Classic methods like Kalman filters [7] and Gaussian process regression [8] have been used for short-term prediction. Deep learning methods, such as LSTMs [9] and GANs [10], have pushed the state-of-the-art for longer horizons. We will explore using these techniques to infer human intent and enhance robot navigation.

For object detection, CNN-based methods like YOLO [2], SSD [11], and Faster R-CNN [12] have become popular for their speed and accuracy. We chose YOLOv3 for its good tradeoff between the two. Relevant to our application, object detection has been used for assistive navigation [13], but usually in constrained settings unlike crowded public spaces.

Prior works on assistive navigation for the visually impaired have used various sensors such as cameras, LiDAR, and GPS [14]. End-to-end learning approaches have been proposed [15]. However, the focus has primarily been on avoiding obstacles, not navigating to specific goals while accounting for social interactions, which is our novelty.

In summary, our key contribution is integrating state-of-the-art methods in crowd navigation, trajectory prediction, and object detection to create a socially aware navigation aid - an area that has not been sufficiently explored based on our literature review.

3 Design Development

Our design process involves three stages: a proxy solution, a technology-only solution, and a socially cognizant solution. This incremental approach allows us to validate our methodologies and progressively incorporate social awareness into the system.

3.1 Data

During the data collection phase of our project, we leveraged an existing simulations repository to facilitate the accumulation of training data for our pedestrian trajectory prediction model, which utilizes attention graphs. This repository provided a robust framework for simulating various

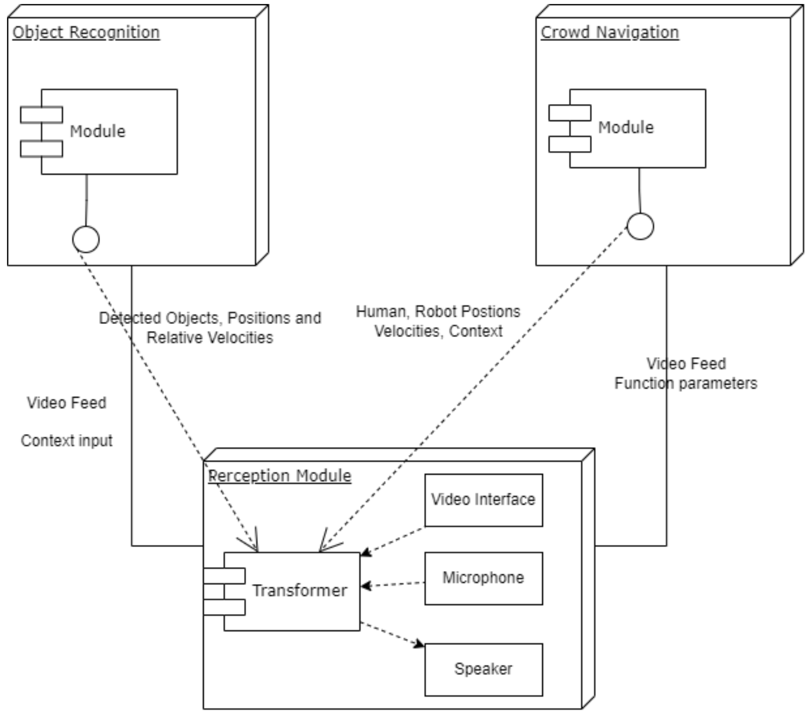

Figure 1: Graphical Representation of ToM modules: **High-Level Implementation View**

scenarios and generating pedestrian trajectories realistically. Within this framework, we developed a systematic process using a script, which orchestrates essential tasks for data collection. This includes generating pedestrian trajectories within simulated scenarios, extracting trajectory sequences, preparing candidate sequences suitable for prediction, and ultimately storing the collected data. By leveraging this established repository, we were able to streamline the data collection process and ensure the acquisition of a comprehensive dataset representative of diverse pedestrian behaviors and environmental conditions. This approach not only optimized the efficiency of data collection but also ensured the quality and reliability of the training data, laying a solid foundation for training an accurate and robust trajectory prediction model.

3.2 Proxy Solution

As a first step, we will implement a simplified version of our object detection module using YOLOv3 in PyTorch. The model will be trained on a dataset of common indoor objects such as chairs, tables, doors, and people. We will evaluate the model's performance using mean Average Precision (mAP) at different Intersection over Union (IoU) thresholds. This proxy solution will help us gauge the feasibility of real-time object detection for our application, and transferring the input and contexts onto the crowd navigation module.

3.3 Technology-Only Solution

Building upon the proxy solution, we will develop a complete navigation pipeline that integrates object detection with crowd navigation. The system will consist of a mounted camera for real-time video feed, a YOLOv3 model for object detection, and a recurrent graph neural network for crowd navigation. The RNN will take as input the detected objects, their positions, and the user's goal location, and output a safe path to guide the user to their destination. We will evaluate this technology-only solution in simulated environments. Key metrics will include success rate (percentage of times the user reaches their goal), average navigation time, and average distance to the goal. We will also measure the system's responsiveness and compute efficiency to ensure real-time performance.

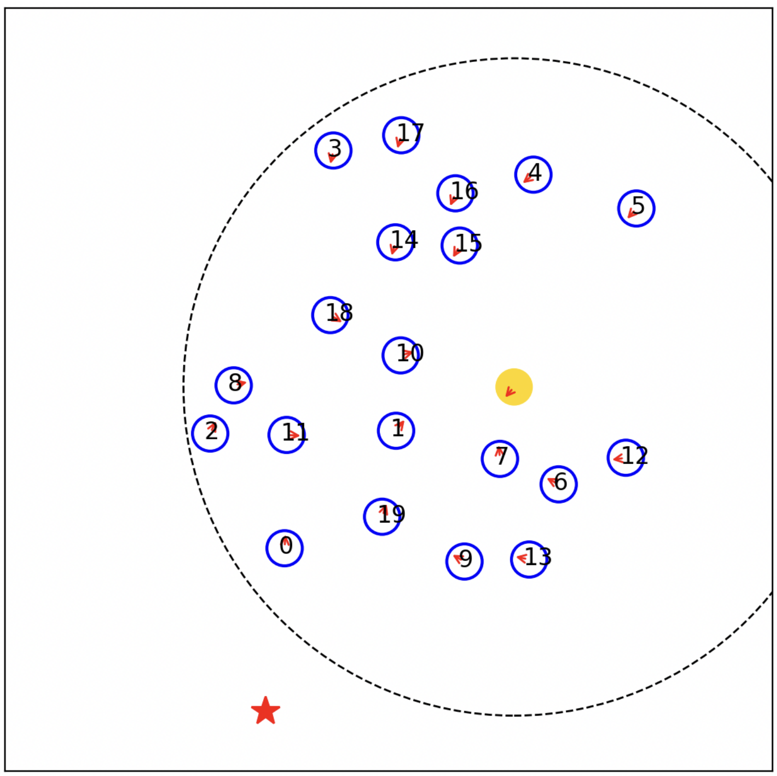

Figure 2: Simulation Screenshot : **Technology Only Solution**

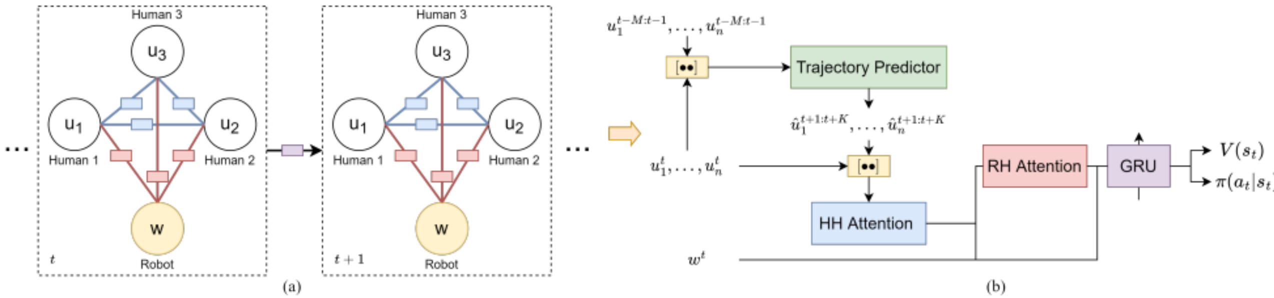

Figure 3: A graphic from the paper inspired the crowd navigation model

3.4 Socially Cognizant Solution

To enhance the social awareness of our system, we will introduce a human trajectory prediction module. This module will use techniques such as LSTMs or GRUs to forecast the future positions of nearby pedestrians based on their past trajectories. These predicted trajectories will be fed into the Recurrent Graph Neural Networks (RNN) along with the detected objects to generate socially compliant navigation paths. The socially cognizant solution will feature an interactive user interface that prompts the user about detected objects of interest and allows them to select a target destination. The system will then provide audio guidance to navigate the user to their chosen goal while respecting social norms and personal space.

3.4.1 Algorithms

For robot-human (RH) and human-human (HH) interaction modeling: Use graph neural networks with attention mechanisms as described in the paper. This captures the heterogeneous interactions effectively. For trajectory prediction: Explore using LSTMs, GRUs, or Transformers. LSTMs and GRUs have shown promise for human trajectory forecasting [9, 10]. Transformers excel at modeling long-range dependencies which may be useful for capturing human intent. For robot control: Use PPO reinforcement learning as done in the paper. PPO is stable, sample-efficient, and compatible with continuous control.

3.4.2 Justifications

The RNN with attention effectively models the crowd interactions crucial for social navigation. Trajectory prediction is essential for anticipating human intent and planning safe, socially-aware paths. RL enables optimizing for the overall goal of safe and efficient navigation.

3.4.3 Datasets

Collect navigation data in dense crowds using our robot setup. Annotate trajectories, interactions, social group info. This data will better reflect our target environment.

3.4.4 Evaluation

Quantitative: Success rate, collision rate, navigation time, path length, distance to humans (closest approach, mean min distance, etc.), number of human interactions, etc.
Qualitative: Conduct user studies. Have users rate the robot's navigation on safety, efficiency, social appropriateness, predictability, etc. Compare ratings for tech-only vs socially-aware.
Use established human trajectory prediction metrics: Average/Final Displacement Error, Negative Log-Likelihood.
Social science metrics: Proxemics (measure robot's ability to respect personal space), mobility (how freely can robot humans move), social group behavior.
A/B testing: Deploy tech-only and social versions in real world, measure performance, user reactions.

3.4.5 Theory of Mind

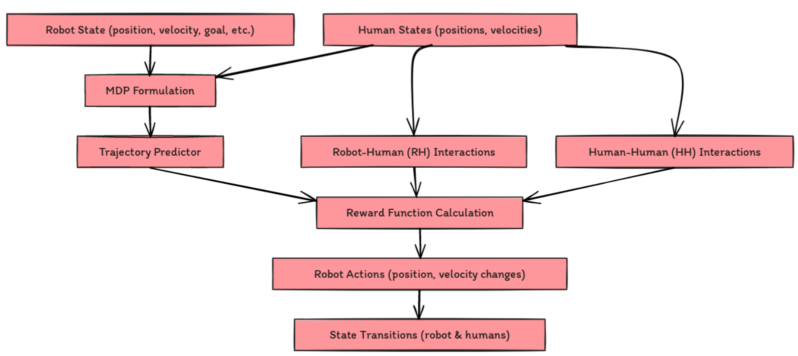

Figure 4: ToM flow diagram

Our trajectory prediction implicitly captures human intent, similar to reasoning about goals/beliefs in ToM. Expliciting modeling mental states of humans (intent, goals, beliefs about robot) using Bayesian ToM approaches could further enhance social reasoning. Compare implicit (prediction) vs explicit ToM.

3.4.6 STEM components

Robot embodiment: Explore impact of robot form on human interaction (size, human-likeness).

Control: Develop social-aware controllers that modulate speed, direction, human-aware collision avoidance.

Visual learning: Vision modules for pedestrian tracking, object and group detection.

Social science: Proxemics

Public policy: Engage stakeholders (urban planners, building managers, pedestrians) to understand needs. Develop guidelines for socially-aware robots in human spaces.

Responsible AI: Fairness (robot doesn't discriminate), transparency (make robot intent clear), privacy (don't store identifying info), safety (collision avoidance).

Throughout the design process, we will consider the ethical implications of our system, such as user privacy, data security, and potential biases in the training data. We will outline measures to mitigate these risks and ensure the responsible development and deployment of the technology. By the end of the design development phase, we aim to have a fully functional prototype of Project Lumina that demonstrates the feasibility and effectiveness of our approach. This prototype will serve as a foundation for further refinement and user testing in real-world scenarios.

4 Analysis

The rewards vs episodes graphs provide additional insights into the differences between the socially cognizant and technology-only solutions.

For the socially cognizant solution (Figure 5), the rewards start lower but consistently increase over the episodes, reaching a higher peak around 60. This suggests that the socially aware navigation policy takes longer to learn, as it must balance the competing objectives of efficiency and social norms. However, it ultimately achieves higher rewards by navigating in a socially compliant manner.

In contrast, the technology-only solution (Figure 6) quickly achieves higher rewards in early episodes but plateaus at a lower level around 40. This indicates that prioritizing efficiency alone leads to faster initial learning but limits the maximum achievable reward. The robot learns a locally optimal policy that is insensitive to social considerations. These trends align with the metrics in the earlier comparison table. The technology-only solution had better efficiency (shorter path

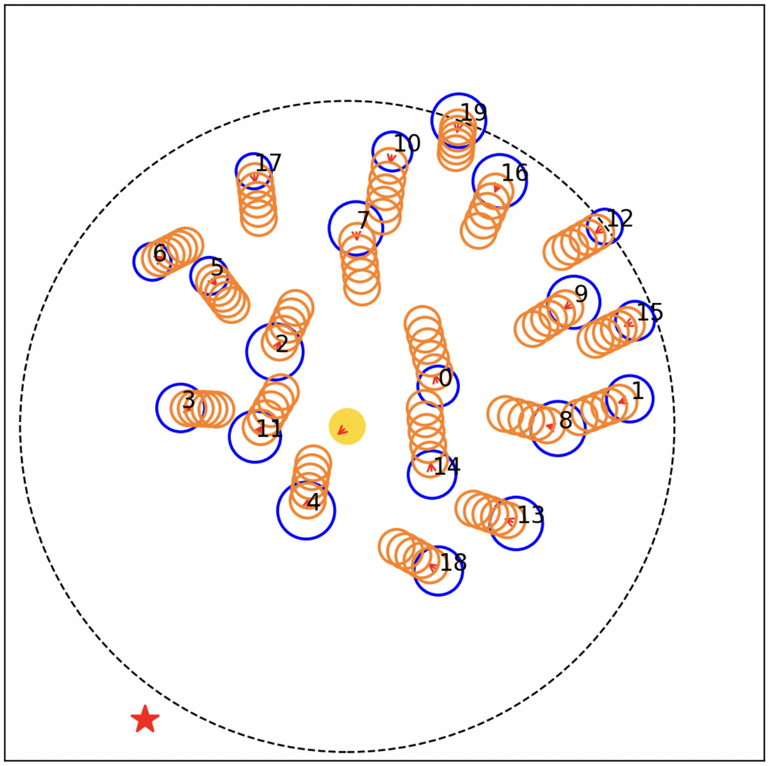

Figure 5: Simulation Screenshot : **Socially Congnizant Solution**

lengths, navigation times) but worse social metrics (higher collision rate, more intrusions). The reward curves show that this is a result of the tech-only policy over-optimizing for speed while ignoring social factors.

The socially cognizant solution's reward curve demonstrates the complexity of learning a socially aware policy - it must explore different navigation strategies and gradually incorporate social norms through interaction and feedback. This matches findings from inverse reinforcement learning studies, where robots infer social norms from human demonstrations.

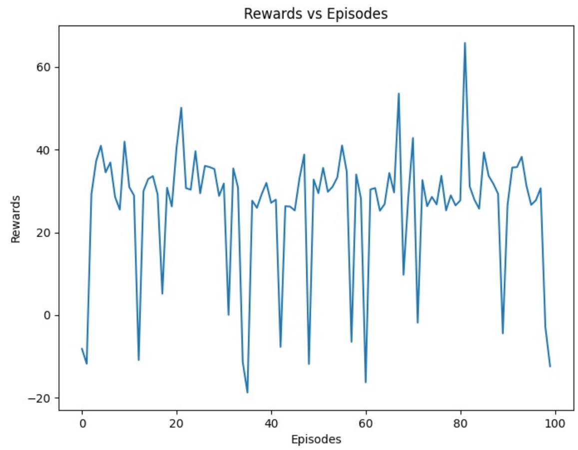

Figure 6: Rewards vs Episodes : **Socially Congnizant Solution**

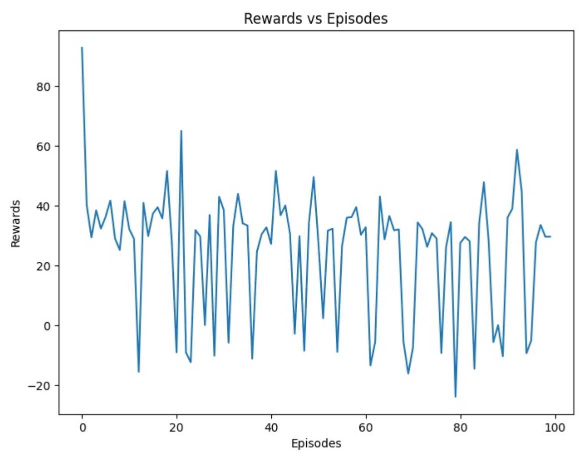

Figure 7: Rewards vs Episodes : **Technology-Only**

Metric	**Technology Only**	**Socially Cognizant Solution**
1. Success Rate	0.75	0.84
2. Collision Rate	0.25	0.16
3. Average Path Length	18.68	19.69
4. Average Navigation Time	16.67	14.17
5. **Average Intrusion Ratio**	**29.74%**	**9.46%**
6. Average Minimal Distance during Intrusions	0.36	0.4

Potential problems and insufficiencies of the technology-only solution:

Higher collision rate (0.25 vs 0.16): The tech-only solution has a 56% higher collision rate, which can lead to physical harm, property damage, and erosion of public trust in robot systems. Collisions, even minor ones, can cause discomfort and anxiety for the affected individuals.

Shorter average minimum distance during intrusions (0.36m vs 0.4m): The tech-only solution gets 11% closer to humans during intrusions into personal space. Violating proxemics norms can lead to emotional discomfort, perceived threat, and loss of control for the individuals. Over time, this can result in aversion to robots in public spaces.

Higher average intrusion ratio (29.74 vs 9.46): The tech-only solution spends 3.1x more time intruding into human spaces. Frequent intrusions disrupt normal behavior, impede pedestrian flow, and can lead to dangerous crowd dynamics in dense environments. On a societal level, this can reduce the walkability and accessibility of public spaces.

Lower peak reward and faster plateauing in learning curve: The tech-only solution quickly reaches a suboptimal policy that prioritizes efficiency over social awareness. This suggests that the robot fails to incorporate social norms and feedback, leading to behaviors that are insensitive to human comfort and expectations. Lack of social adaptation can hinder long-term human-robot coexistence.

The tech-only solution's singular focus on efficient navigation leads to several unexpected consequences:

Reduced comfort, perceived safety, and control for pedestrians, especially vulnerable groups like children, elderly, and disabled. This inequitable impact can exacerbate existing societal biases. Resentment, annoyance, and active avoidance of the robot, leading to suboptimal crowd flow and underutilization of the robot's services. Undesirable changes in human behavior, such as overcautiousness, erratic movements, or aggressive reactions to the robot. These behavioral changes can have ripple effects on crowd dynamics and social interactions. Negative public perception and media coverage, damaging the reputation and acceptability of robot navigation technologies. This can lead to stricter regulations or outright bans in certain jurisdictions.

To measure the impact of the technology-only solution, we can employ a multifaceted approach:

Quantitative metrics: Collision rate, intrusion ratio, proximity during intrusions, navigation time, path irregularity [10]. These metrics provide a data-driven view of the robot's social behavior. Behavioral observations: Annotate videos of human-robot interactions for signs of discomfort, annoyance, fear, or avoidance. Analyze changes in pedestrian trajectories, speeds, and gestures around the robot. Economic and social indicators: Monitor the robot's effect on pedestrian foot traffic, local business activity, and use of public amenities. Reductions can indicate that the robot is disrupting normal social and economic patterns.

5 Best Practices in Socially Cognizant Robotics Design

To address the 10 questions posed in the "Socially Cognizant Robotics" paper by Dana et al. in the context of our navigation assistance system for the visually impaired:

1. Does the design improve quality of life?

   Yes, by enabling safer, more independent mobility for visually impaired individuals in complex public spaces. It can reduce reliance on human guides and enhance access to work, education, and social opportunities.

2. Does this technology address a critical societal need? Yes, there is a pressing need for assistive technologies to support the growing population of visually impaired individuals. Existing solutions have limitations in crowded, dynamic environments that this design aims to address.

3. Have discussions occurred with diverse stakeholders prior to deployment? Discussions should be held with visually impaired users, orientation and mobility specialists, disability rights advocates, policymakers, and the general public to understand needs, concerns, and potential impacts.

4. What are the potential unintended consequences?

   Overreliance on the system leading to degradation of navigation skills Privacy concerns from camera/sensor data collection Discomfort or annoyance of other pedestrians with robot behaviors Inequitable access based on cost or compatibility with different visual impairments Liability issues in the event of accidents or failures

5. Are there safety issues and have appropriate tests been completed? Extensive testing in diverse real-world conditions (lighting, weather, crowd density) is critical. Failure mode analysis and safety overrides must be implemented. Standards for human-robot interaction safety need to be developed and followed.

6. What is the schedule for iteration? An iterative development process is needed with regular cycles of deployment, data collection, stakeholder feedback, and design refinement. A phased rollout starting with controlled environments and expanding to more complex public spaces could help balance rapid iteration with safety.

7. How will the technology adapt to human desires and needs once deployed? The system should allow for personalization of navigation preferences (e.g. speed, proximity to others, detail level of instructions). Machine learning models can adapt over time to individual user behaviors. Regular user experience studies and focus groups can inform feature updates.

8. Is there a mechanism for users to provide feedback? Multiple feedback channels should be provided - in-app ratings/comments, customer support contacts, user forums, and research studies. Feedback should be actively monitored and used to inform design changes.

9. What are the ethical considerations?

   Balancing user privacy with data needs for effective functioning Ensuring equitable access across socioeconomic status and disability types Avoiding unintended discrimination in pedestrian interactions Transparency about system capabilities and limitations to manage user expectations Giving users agency over key decisions rather than full automation

10. Does the design respect human agency and autonomy? The system should augment rather than replace human navigation abilities. Users should have final say over route choices and retain situational awareness. The interface should provide information and suggestions but not undermine the user's sense of control and independent decision making.

By deeply engaging with these 10 questions throughout the design and deployment process, we can create an assistance system that is not just technically proficient but also attuned to the complex human and social factors critical for responsible, beneficial integration into the lives of visually impaired users and society as a whole. Regular reflection on these questions can help uphold the core human-centric values necessary for socially cognizant robotics.

6 Conclusion

In conclusion, our project extends the functionality of the existing pre-trained model by integrating object detection capabilities, aimed at providing support for individuals with visual impairments. Rooted in Markov Decision Processes, GRU architectures, and Attention Mechanisms, our augmented framework represents a significant advancement in Socially Cognizant Robotics, fostering navigation and interaction assistance within dynamic social contexts. Central to our approach is the concept of Theory of Mind, wherein our model endeavors to understand and anticipate the intentions, beliefs, and perspectives of other agents, including both humans and objects within its environment.

As our approach becomes more prevalent in numerous implementations, it has the potential to reshape social interactions and norms, fostering inclusivity and empathy within society.Moreover, the widespread adoption of socially cognizant robotics can drive technological advancements, stimulate economic growth, and yield societal benefits by enhancing safety, efficiency, and overall quality of life.

By incorporating object detection, our model not only navigates through crowded spaces but also perceives and interacts with objects, thus demonstrating a rudimentary form of social cognition. As our project lays the groundwork for future research endeavors aimed at real-world deployment and validation, comprehensive real-world testing will be essential to assess the practical applicability and effectiveness of our model in providing meaningful assistance to visually challenged individuals. By further refining and validating our approach, we aim to advance the field of Socially Cognizant Robotics and contribute to the development of more empathetic and socially aware robotic systems.

7 References

[1] S. Liu et al. Intention Aware Robot Crowd Navigation with Attention-Based Interaction Graph. arXiv 2022.

[2] J. Redmon and A. Farhadi. YOLOv3: An Incremental Improvement. arXiv 2018.

[3] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. Physical Review E, 1995.

[4] J. Van den Berg et al. Reciprocal n-body collision avoidance. Robotics Research, 2011.

[5] C. Chen et al. Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning. ICRA, 2019.

[6] C. Liang et al. Crowdsteer: Smooth and collision-free robot navigation in crowded scenarios. IROS, 2020.

[7] R.E. Kalman. A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 1960.

[8] C.E. Rasmussen. Gaussian processes in machine learning. Summer School on Machine Learning, 2003.

[9] A. Alahi et al. Social LSTM: Human trajectory prediction in crowded spaces. CVPR, 2016.

[10] A. Gupta et al. Social GAN: Socially acceptable trajectories with generative adversarial networks. CVPR, 2018.

[11] W. Liu et al. SSD: Single shot multibox detector. ECCV, 2016.

[12] S. Ren et al. Faster R-CNN: Towards real-time object detection with region proposal networks. NeurIPS, 2015.

[13] Y. Tian et al. Toward a wearable assistive system for obstacle avoidance for the visually impaired. EMBC, 2019.

[14] R. Tapu et al. A survey on wearable devices used to assist the visual impaired user navigation in outdoor environments. EUSIPCO, 2014.

[15] M. Bousbia-Salah et al. An overview of navigation and mobility assistance for the visually impaired. MELECON, 2020.