Research Statement

Linh Kastner
School of Computing and Information Systems, Singapore Management University
Tel: (65) 6828 9557; Email: linhkastner@smu.edu.sg
1711/2025

Background

Robots are rapidly entering domains where interaction with humans is central — hospitals, offices, warehouses, and homes — requiring not only physical autonomy, but also social awareness, contextual understanding, and trustworthiness. Despite strong progress in control, perception, and learning, embodied agents still struggle to generalize across tasks, interpret social context, and behave in ways that humans find predictable and acceptable. A core reason is the lack of comprehensive datasets that capture the richness of real-world interactions, alongside the scarcity of robust development infrastructures, unified benchmarking suites, and standardized evaluation protocols. These limitations make progress difficult to develop, compare, and reliably reproduce.

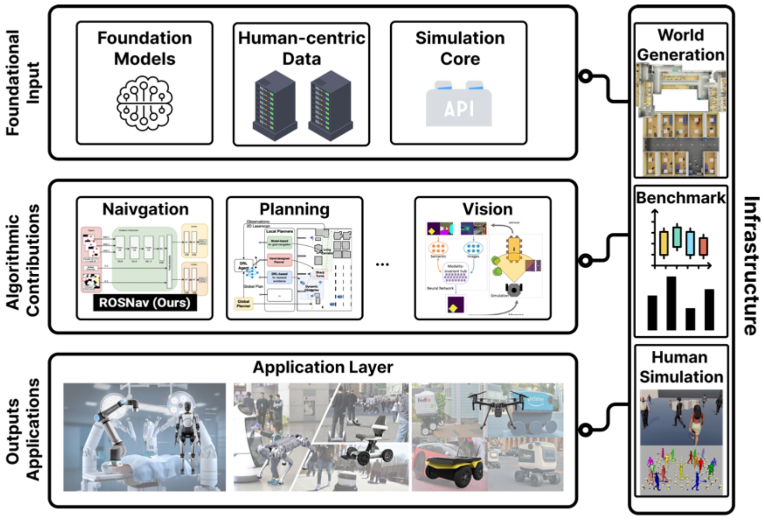

Figure 1: Research Vision

<u>**My long-term research vision is to build infrastructure and algorithms that enable scalable, generalizable, and socially aware embodied intelligence**</u>. This includes creating high-quality datasets, unified benchmarks, and standardized development testbeds. Central to this vision is the design of open, modular software stacks and platforms that leverage generative AI and foundation models to enhance realism and scalability. Such systems support large-scale training, evaluation, and the creation of datasets for foundation models capable of learning from multimodal data, interacting in dynamic settings, and reasoning about human behavior.

Accordingly, my research contributes both algorithmic advances — **in generative world synthesis, realistic human modeling, and robot path and motion planning** — and system-level integration, bringing these components together into cohesive frameworks (as illustrated in Fig. 1). This integration is essential, particularly as many industrial robots still rely on outdated, heuristic-driven navigation methods such as the Dynamic Window Approach (DWA, 1997) [1], Timed Elastic Band (TEB, 2012) [2], or basic "stop-and-go" policies [3]. Such rigid approaches limit efficiency, responsiveness, and broader acceptance in human-centric environments.

Research Areas

Towards Standardized Infrastructures, Benchmarks, and Datasets for Robotics

Just as computer vision and NLP advanced through shared datasets, protocols, and community challenges, robotics urgently requires standardized benchmarks to ensure reproducibility, fair comparison, and collective progress. My work contributes directly to this need. I developed unified testbeds, frameworks, and evaluation pipelines that introduce clear scenario abstractions and toolchains, making embodied AI research more transparent, interoperable, and reproducible. These efforts laid the foundation for the ARENA platform [4], an open, modular, and extensible ecosystem for developing and benchmarking socially aware navigation in dynamic, human-centric environments. Unlike existing systems with static scenarios and limited variability, ARENA integrates generative AI modules for scene layout and crowd simulation, supports large-scale policy training, and enables reproducible evaluation across multiple simulation backends. Its impact is reflected in a strong publication record across top robotics venues, including IROS 2021 [5], IROS 2022 [6], RA-L [6], ICRA 2022 [7], IROS 2023 [8], RSS 2023 [9], RSS 2024 [4], RSS 2025 [10], ICRA 2025 [11], and IROS 2025 [12].

Figure 2: Concept of the ARENA Platform

Looking ahead, I aspire to extend this line of work by developing realistic and scalable simulation environments powered by advanced generative and foundation models. This includes generating complex, photorealistic scenarios using simulation engines such as NVIDIA Isaac Sim and Unreal Engine, combined with realistic human behavior modeling. By integrating structured scene graphs, multimodal prompting, diffusion models, large language models (LLMs), and vision-language models (VLMs), these systems will reproduce intricate indoor and outdoor environments, dynamic human behaviors, and diverse social interactions. Such high-fidelity simulations will more closely resemble real-world conditions, improving the accuracy, validity, and robustness of algorithmic evaluation.

An additional research direction will focus on using generative models to advance human behavior simulation. I aim to develop a hierarchical generative framework that produces symbolic behavior programs along with physically and socially plausible human motion trajectories. By combining diffusion models with VLMs, the system will support flexible, language-guided generation of complex multi-agent interactions. For instance, given a prompt such as “a family waiting while one person approaches a counter inside a Singaporean hawker center," the system would generate both an interactive 3D environment and the corresponding structured, socially coherent human behaviors.

Socially-aware Algorithms for Embodied AI

Building on this realistic and scalable simulation infrastructure, I aspire to explore and develop novel algorithms for key human-centric robotics tasks, including social navigation, perception, grasping, and manipulation, with applications in healthcare, logistics, and service robotics. A central goal is to enhance robot capabilities to perceive, interpret, predict, and adapt to human behaviors and social interactions, thereby improving fluency, predictability, and trustworthiness in complex, human-centered environments.

The simulation ecosystem and datasets described earlier provide a unique advantage by enabling large-scale, realistic training and evaluation. For instance, social navigation algorithms can leverage human intent and trajectory patterns generated by the platform, resulting in safer, more efficient robot motion. Generative and foundation models will further improve scene understanding, semantic reasoning, and planning.

This research direction focuses not only on developing new methods but also on integrating, standardizing, and refining existing approaches using the unified toolchains and infrastructure. Through this combination of algorithmic innovation and system-level integration, the work aims to deliver robust, generalizable, and socially intelligent robotic behaviors grounded in realistic, multimodal data.

Human Robot Colaborative Systems and Beyond

To close the gap between theoretical research and real-world robotic systems, my work will explicitly focus on sim-to-real transfer and the deployment of thoroughly validated algorithms onto physical robot platforms. I will examine how simulation fidelity affects real-world performance, identify the factors that drive performance variability, and determine the optimal balance between synthetic and real-world data for training. Data collected during field experiments will feed back into the simulation infrastructure to refine foundation models and improve algorithmic components, completing the full development loop.

A key element of this research involves advanced data augmentation techniques, including generative AI methods, domain randomization, and real-to-sim mapping, to increase robustness and diversity across datasets. High-fidelity simulation environments will serve both as sources of large-scale synthetic data and as testbeds for the iterative refinement of foundation models. This process aims to produce models that are broadly generalizable yet specialized enough for practical deployment.

Beyond these core contributions, I aim to extend this research into broader domains where embodied intelligence can generate transformative impact. One promising direction is medical and assistive robotics, where robots must safely and intuitively operate alongside caregivers and patients. Leveraging scalable simulation and generative human modeling, future work can investigate robot behaviors for patient-handling assistance, rehabilitation, mobility support, and bedside interaction — domains that require exceptional reliability, transparency, and social understanding. Simulation-driven behavioral datasets can further support the development of algorithms capable of recognizing patient states, predicting intentions, and adapting interaction policies in real time.

Additionally, the same foundation—realistic simulations, robust sim-to-real pipelines, and generative world models—can extend to emergency response, elderly care, micro-logistics, education, and public environments. These sectors require robots that not only perceive and act but also collaborate, communicate, and reason ethically in human-centric spaces. As generative AI and robotics converge, new research opportunities emerge in personalized robot behavior modeling, cultural and social adaptation, long-horizon autonomy, and cross-domain generalization.

Selected Publications and Outputs

[1] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 4, pp. 23–33, 1997.

[2] C. Rösmann, F. Hoffmann, and T. Bertram, "Timed-elastic-bands for time-optimal point-to-point nonlinear model predictive control," in *European Control Conference (ECC)*, 2015, pp. 3352–3357.

[3] P. T. Singamaneni, P. Bachiller-Burgos, L. J. Manso, A. Garrell, A. Sanfeliu, A. Spalanzani, and R. Alami, "A survey on socially aware robot navigation: Taxonomy and future challenges," *The International Journal of Robotics Research*, 2024.

[4] L. Kästner, V. Shcherbyna, H. Zeng, T. A. Le, M. H.-K. Schreff, H. Osmaev, N. T. Tran, D. Diaz, J. Golebiowski, H. Soh et al., "Arena 3.0: Advancing social navigation in collaborative and highly dynamic environments," *Robotics: Science and Systems*, 2024. Available: https://openreview.net/forum?id=BjnICVYhGZ

[5] L. Kästner, T. Buiyan, L. Jiao, T. A. Le, X. Zhao, Z. Shen, and J. Lambrecht, "Arena-Rosnav: Towards deployment of deep-reinforcement-learning-based obstacle avoidance into conventional autonomous navigation systems," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 6456–6463.

[6] L. Kästner, T. Bhuiyan, T. A. Le, E. Treis, J. Cox, B. Meinardus, J. Kmiecik, R. Carstens, D. Pichel, B. Fatloun, N. Khorsandi, and J. Lambrecht, "Arena-Bench: A benchmarking suite for obstacle avoidance approaches in highly dynamic environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9477–9484, Oct. 2022. Available: http://dx.doi.org/10.1109/LRA.2022.3190086

[7] L. Kästner, J. Cox, T. Buiyan, and J. Lambrecht, "All-in-one: A DRL-based control switch combining state-of-the-art navigation planners," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022, pp. 2861–2867.

[8] L. Kästner, R. Carstens, H. Zeng, J. Kmiecik, T. Bhuiyan, N. Khorsandhi, V. Shcherbyna, and J. Lambrecht, “Arena-Rosnav 2.0: A development and benchmarking platform for robot navigation in highly dynamic environments,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 11257–11264.

[9] L. Kästner, V. Shcherbyna, R. Carstens, S. Lee, J. Schreff, M. Ho-Kyoung, J. Lambrecht et al., “Demonstrating Arena-Web: A web-based development and benchmarking platform for autonomous navigation approaches,” *Robotics: Science and Systems*, 2023. Available: https://www.roboticsproceedings.org/rss19/p088.pdf

[10] L. Kästner, V. Shcherbyna, H. Zeng, D. A. Do, M. H.-K. Schreff, H. G. Nguyen, N. T. Tran, H. Soh et al., “Arena 5.0: A photorealistic ROS2 simulation framework for developing and benchmarking social navigation,” *Robotics: Science and Systems*, 2025. Available: https://www.roboticsproceedings.org/rss21/p092.pdf

[11] L. Kästner, V. Shcherbyna, H. Zeng, J. Kreutz, M. H.-K. Schreff, T. Lenz, N. T. Tran, A. Martban, H. Soh et al., “Arena 4.0: A comprehensive ROS2 development and benchmarking platform for human-centric navigation using generative-model-based environment generation,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2025. Available: https://arxiv.org/abs/2302.10023

[12] L. Kästner, V. Shcherbyna, J. Wang, J. Kmiecik, T. Bhuiyan, N. Khorsandhi, V. Shcherbyna, and J. Lambrecht, “Arena-Bench 2.0: A comprehensive benchmark of social navigation approaches in collaborative environments,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2025.

[13] L. Kästner, B. Fatloun, Z. Shen, D. Gawrisch, and J. Lambrecht, “Human-following and guiding in crowded environments using semantic deep-reinforcement-learning for mobile service robots,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022, pp. 833–839.