From Moral Will to Moral Skill

Operationalizing Care-Centered Value Sensitive Design in Robotics

Nicolas Wim Landuyt
January 2026

From Moral Will to Moral Skill

Operationalizing Care-Centered Value Sensitive Design in Robotics

by

Nicolas Wim Landuyt

Supervisor:	Prof.dr.ir. D.A. (David) Abbink
Co-Supervisor:	Dr. O. (Olya) Kudina
Daily Supervisor:	Dr.ir. O. (Olger) Siebinga
Project Duration:	June, 2024 - January, 2026
Faculty:	Mechanical Engineering, Delft

Cover: Susanna Landuyt & Inès van Rijn

Abstract

Healthcare systems increasingly rely on robotic technologies to address workforce shortages, yet a significant disconnect remains between the technical logic of efficiency and the relational requirements of care. While standard Value Sensitive Design (VSD) provides a method for value integration, it lacks specific normative content. Care Centered VSD (CCVSD) resolves this by anchoring design in care ethics, yet it remains a theoretical model that lacks the practical tools required for engineering implementation. This thesis posits that this failure stems not from a deficit of moral will (practitioner apathy), but from a deficit of moral skill: the lack of intermediate-level knowledge (operational artifacts that bridge the divide between high-level theory and concrete implementation) to translate abstract values into concrete technical implementations.

Adopting a Research through Design (RtD) approach, this study operationalized the CCVSD framework through a case study within a commercial robotics company, focusing on the development of a social navigation module (interaction-aware motion planning that respects human proximity and social cues) for the PAL Tiago robot. The investigation revealed three structural barriers. First, semi-structured interviews with 19 stakeholders (comprising robotics engineers and caregivers) identified a semantic gap where groups held divergent interpretations of care values; this was bridged by the Prospective Value Hierarchy, which translated care values into technical constraints. Second, the technical implementation of the navigation algorithm revealed a normative void, caused by the original framework structurally skipping the technical development process. This phase was made explicit and operationalized by integrating standard engineering workflows, providing the necessary structure to document technical trade-offs as ethical decisions rather than implicit defaults. Third, the integration phase of the framework, executed through comparative evaluations with 20 stakeholders (spanning care receivers, caregivers, and management) uncovered contextual instability, demonstrating that identical robotic behaviors elicited divergent value judgments across different care settings.

The findings from this case study suggest that the operationalized framework offers practical utility for engineering teams. Feedback from the participating engineers indicated that the translational artifacts helped structure their tacit knowledge into a formalized process. These results indicate that providing engineers with tools to convert ethical mandates into actionable constraints can support a shift from post-hoc ethical critique to continuous design input within such collaborative environments.

Keywords: Care Centered Value Sensitive Design (CCVSD), Research through Design (RtD), Social Navigation, Moral Skill, Intermediate Level Knowledge, Robotics Ethics.

Acknowledgements

Throughout this whole process of care, I have truly learned what it means to be taken care of, and to those caretakers: I dedicate this section.

First of all, I would like to thank my supervisors for always guiding me throughout the difficulties of these academic waters. I would like to thank Olya, who has taught me what academia can truly be, what it means to ask questions and what it means to question them, what it means to navigate a world so complex yet so sweet. I would like to thank David for his sharpness; at every pivotal moment he primed a way for me, tempering the meeting room with the right temperature, saying the things so often left untold, a true pillar of my support throughout this project. I would like to thank Olger, for being there with me, week after week, meeting after meeting, weighing the pros and the cons, wooing the unknowns. Thank you for your guidance and your presence. No matter the temporal negligence, a meeting and some feedback always found their way to my home.

I would also very much like to thank Cock, founder of Heemskerk Innovative Technology, who has allowed me to graduate at his company, being there present, in the daily reality of an engineering team creating the future of care robots while being so open to constructive analysis of who and how they were. To allow me to sit up close and learn. I would like to thank Jasper, Bas, and Sjoerd who have helped me navigate that reality and helped me with handlebars of hope.

I would like to thank my partner Inès, for supporting me throughout the joyride the past year has been, for her patience and her poise, for her help in this work and everything outdoors. I would like to thank my family, my father and mother for making it possible for me to be here, to my sister for her drawings and dreams, and to my brother for his kinship and care.

To the comrades of my cottage, to Mickey and Siert, to Jan and Lucas, thank you for allowing me to be me, to experiment and ferment, to quest and forgo the duties; thank you for the leeway. To Henk: you motivated and melted my hart, you pushed and placed your life in mine, allowing me to take care of you, to help you up when you were down, and to rise in the long hours of the night. I learned from you and you let me.

To fellow Denniz, Jakob, Chris, Marijn, Mink, and Korneel: they carried me through this path, my soundboard of systemic guidance, a golden guild of friendship.

Contents

List of Abbreviations

Abbreviation	Definition
I. Ethics & Methodology	
CCVSD	Care Centered Value Sensitive Design
VSD	Value Sensitive Design
ILK	Intermediate Level Knowledge
PVH	Prospective Value Hierarchy
RtD	Research through Design
II. Robotics & Engineering	
DoF	Degrees of Freedom
HRI	Human-Robot Interaction
LIDAR	Light Detection and Ranging
MPPI	Model Predictive Path Integral
Nav2	Navigation 2 (ROS 2 Navigation Stack)
KPI	Key Performance Indicators
PIR	Path Irregularity Ratio
PSI	Personal Space Intrusion
ROS 2	Robot Operating System 2
SFM	Social Force Model
TF	Transform (Coordinate Frame System)
TTC	Time-to-Collision

1

Introduction

This chapter introduces the research topic of value sensitive design (VSD) in engineering, specifically focusing on its adaptation and operationalization within care robotics. It outlines the background of the research problem, incorporating findings from a preliminary literature study, and defines the objectives, scope, and significance of this study.

1.1. Topic and context: the ethics of efficiency vs. the ethics of care

Global healthcare systems are facing escalating demands driven by demographic shifts, most notably ageing populations and persistent workforce shortages [4, 29, 1]. Robotic technologies are frequently presented as an inevitable solution to augment human caregiver capacity [48, 41]. However, a significant disconnect remains between the technological promise of these systems and their integration into the relational fabric of care environments [5, 41].

Current developments in healthcare robotics are primarily driven by the optimization of quantifiable metrics such as speed, cost-effectiveness, and output. As highlighted in systematic reviews of the field [63], the primary justification for these technologies is often economic: addressing workforce shortages through efficiency. This thesis refers to this prioritization of measurable utility as an *ethics of efficiency*. As Stein argues, efficiency has been elevated from a tool to a central value in public services, often crowding out other moral goods like equity or accountability [53]. Following Berg [54], this rationalization is not only a technical imposition but a moral project; standardization serves as the infrastructure for safety and reliability, what Pols [45] describes as *cold care*. However, while this produces technically competent systems, it often conflicts with the *ethics of care*. In contrast to this metric-driven logic, this thesis defines *care* following Tronto [56]: not as a sentiment, but as a practice: ”a species activity that includes everything we do to maintain, continue, and repair our world so that we can live in it as well as possible”. Care is intrinsically slow, relational, and situated, placing necessary burdens on the carer, whereas engineering is inherently driven by optimisation [26, 34]. As van Staveren [52] argues, this drive to minimize waste is value-laden and often obscures qualitative values like dignity. The resulting robotic implementations may prioritise competence (technical completion) over attentiveness (relational awareness), producing machines that function correctly according to the logic of efficiency while nevertheless diminishing the essential rhythms of the *care practice*; defined here not as a logistical service task, but as a reciprocal social relationship directed toward a patient’s fundamental need [61].

Ethical frameworks such as care-centred value sensitive design (CCVSD) have emerged in response to this tension [61]. While the comprehensive methodological steps of this framework are detailed later in Section 2.1, its primary function here is normative: it provides the theoretical vessel for translating Tronto’s abstract care ethics into engineering design goals. This framework promotes the integration of care values, attentiveness, responsibility, competence, and responsiveness, throughout the development of care technologies. However, despite the theoretical robustness of such frameworks, a significant implementation gap persists [39]. This disconnect arises because abstract ethical values often resist the precise translation into executable code required by robotic control architectures. For instance, while a value like safety can be readily operationalized as a minimum clearance distance (e.g., 0.5 meters), complex care values such as dignity or attentiveness lack such computable definitions. Without clear metrics, these values remain semantically inaccessible to the engineer, leading to a failure to translate the theoretical ideal of care into the practical logic of the robotic implementation. This disconnect is not only a failure to translate values into engineering specifications, but potentially a failure to recognize that these values are location and institution specific. The prevailing assumption in robotic design, shown by the search for general ethical algorithms [3], is that a single, universal solution exists that can effectively govern behavior across all environments [6]. However, if the definition of good care proves to be contextually unstable, where a robot acting efficiently is judged as competent in a hospital but rude in a care home, then a universal solution becomes a trap. Consequently, the design challenge must expand beyond the opposition of efficiency versus care to address the reality that ethical requirements are not static. A robot acting efficiently may be judged as competent in a hospital but rude in a residential care home. If the definition of good care changes based on the environment, then the search for a single, universal implementation of care values becomes a trap.

The closed loop of care: from skill to sensibility

Building on the work of Tronto [56], effective care cannot be reduced to simply executing a task; it is a process that requires verification. A care act is only complete when the caregiver observes the recipient's reaction, known as responsiveness, to ensure the need was actually met. If a robotic system completes a technical task but ignores the human reaction, the care is incomplete. This necessity of using the recipient's feedback to validate and adjust the action is what this thesis refers to as *closing the ethical loop*. Beyond the robot exhibiting moral skill, the care receiver must possess the sensibility to recognize good care, from the robot, within their specific context.

For this feedback loop to be effective, the robot's behavior must be transparent to the user. When a robot visibly adapts its actions, such as pausing for a patient, it makes its care legible, allowing the user to trust that the system is aware of them. However, because the expectations for this behavior differ drastically between environments, where efficiency is demanded in a hospital but patience is required in a home, a single, static set of rules can fail. This variety necessitates a move away from universal design mandates toward adaptive guidelines, enabling the robot to align its behavior with the distinct norms of the specific care practice its executing.

1.2. Focus and scope

This thesis aims to operationalise the CCVSD framework by investigating the practical design methods needed to connect ethical principles with engineering practice. Rather than claiming a universal solution, this research investigates this translational process through a situated case study. It specifically examines how care values can be mapped onto the technical constraints of social mobile manipulation.

1.2.1. Operational context

It is crucial to distinguish the specific modality of care addressed in this work. This research focuses on distal interaction, which is defined here as care mediated through spatial proximity rather than physical contact. Specifically, it examines the non-contact social dynamics of navigation.

This research leverages the modular architecture of industry-standard systems like Navigation2 [32]. By decoupling the navigation stack from other behaviors, the study isolates specific ethical parameters, such as social distance, and eliminates the confounding variables introduced by complex manipulation tasks like grasping. This establishes a controlled setting to investigate the friction between ethical theory and engineering implementation.

The operational environment is defined as the institutional corridor. Following Gesler's concept of the therapeutic landscape, the care environment is not merely a logistical transit zone but an active component of the healing process [16]. In this context, the robot's movement is not simply a trajectory from A to B but a participant in a choreography of care. A robot that moves aggressively through a corridor disrupts this landscape and transforms a space of recovery into a space of hazard. Therefore, the context is limited to the kinematics of this mobile interaction: how the robot navigates shared space, yields to patients, and signals intent without physical contact. Functionally, the robot is defined as a logistical support agent (e.g., a robotic porter). Its primary utilitarian goal is the transport of medical goods or waste between wards to relieve physical burden from staff. However, to achieve this efficiency, it must traverse shared public corridors. Therefore, the design challenge is not merely reaching the destination (technical competence), but navigating the social fabric of the hallway without disrupting the therapeutic landscape.

1.2.2. Scope of work

To operationalize this context, the research targets the domain of social navigation for autonomous mobile robots operating within clinical nursing environments.

Methodologically, the scope is limited to the comparing and modifying of standard industrial control architectures, rather than the construction of a robotic system from the ground up. This acknowledges that modern robotics is primarily a process of integration rather than raw invention. Therefore, instead of formulating new algorithms ex nihilo, this focuses the research on the practical necessity of auditing source code to identify the specific heuristic thresholds and static parameters where values are implicitly operationalized. In this context, the social navigation module developed for the robot serves not only as a functional prototype but as an epistemic tool designed to probe and reveal the socio-technical conflicts that arise when abstract ethical mandates encounter the rigid logic of robotics control systems.

Within this boundary, the operationalization of abstract values is investigated through distinct control and perception parameters. For instance, the research examines how responsiveness maps onto the kinematic metric of trajectory smoothness (jerk), and how attentiveness maps onto the spatial logic of saliency mapping. This signals a shift from philosophical abstraction to robotics application, treating movement in shared space as a proxy for distal care where motion quality shapes interaction quality. Purely software-based artificial intelligence systems are excluded, as the focus remains strictly on the kinematic reality of embodied interaction in shared physical environments.

1.3. Problem statement

This section defines the core challenge of this research: the persistent disconnect between ethical theory and robotic construction. While the previous section outlined the normative tension between care and efficiency, this section analyzes the methodological failure to translate those values into technical artifacts. It identifies two structural causes for this failure: first, a deficit of intermediate-level skills within engineering practice (Section 1.3.1), and second, a critical void in the standard design frameworks that leaves this technical translation unsupported (Section 1.3.2).

1.3.1. From moral will to moral skill

Engineering ethics confronts a persistent disconnect between abstract values and concrete technical architectures, referred to as the implementation gap [39]. Lynch and Kline argue that the field has historically explained this failure by prioritizing strict codes and compliance measures designed to constrain behavior [31]. This approach implicitly suggests that practitioners must be policed against negligence or they will inevitably prioritize optimization at all costs. Davis further identifies a prevalent tendency to avoid responsibility where engineers attribute ethical outcomes to external constraints rather than their own agency [7]. This belief that the primary obstacle to ethical engineering is a lack of motivation or intent to act ethically is defined here as a deficit of moral will. This research initially adopted this perspective as a starting point, investigating whether the solution necessitated imposing stricter ethical mandates on technical teams.

This thesis reframes the source of this implementation gap. Exploratory interviews conducted with robotics engineers prior to the technical development of a robotic artifact indicated that the gap reflects a deficit of moral skill rather than moral will. As Van de Poel suggest, engineers generally possess the intent to produce ethically sound systems but are often constrained by the lack of agency or tools [44]. Specifically, they confront a methodological barrier: the lack of intermediate-level tools required to translate abstract normative values, such as attentiveness, a foundational element of Tronto's care ethics [56], into concrete design choices [43]. However, reflecting Agre's argument that technical tools actively structure and limit the practice they support [2], this deficit is physically embedded within the software infrastructure itself. Standard simulators utilized in robotic development typically treat the world as static and often render human agents invisible to sensors, creating a significant technical infrastructure gap [10, 42, 25]. Consequently, moral skill must be understood as more than just a translation ability. The translation of values is fundamentally impeded when the target technical architecture lacks the ontological foundation to represent social agents; therefore, this skill must also encompass the technical capacity to identify and rectify these infrastructure gaps, effectively repairing the tools to ensure they can support the reciprocal interactions required by care.

1.3.2. Limitation of the current frameworks

While the previous section established that frameworks like CCVSD provide the necessary normative foundation by identifying key values such as attentiveness [62], a critical analysis reveals a structural limitation in their methodological application. Identifying values is distinct from implementing them. To bridge the gap between abstract ethical goals and concrete engineering practice, the field requires what interaction design theory calls *intermediate-level knowledge*: actionable methods or patterns that translate high-level concepts into specific technical and interactional mechanisms [30, 22, 51].

Within the domain of care robotics, CCVSD represents a comprehensive attempt to operationalize care ethics into a normative design framework. However, as detailed in Section 2.1, the framework is structured around distinct, sequential phases. As shown in Figure 1.1, the methodology transitions linearly from value analysis and data collection (understanding the care context) directly to evaluation (assessing the robot once introduced).

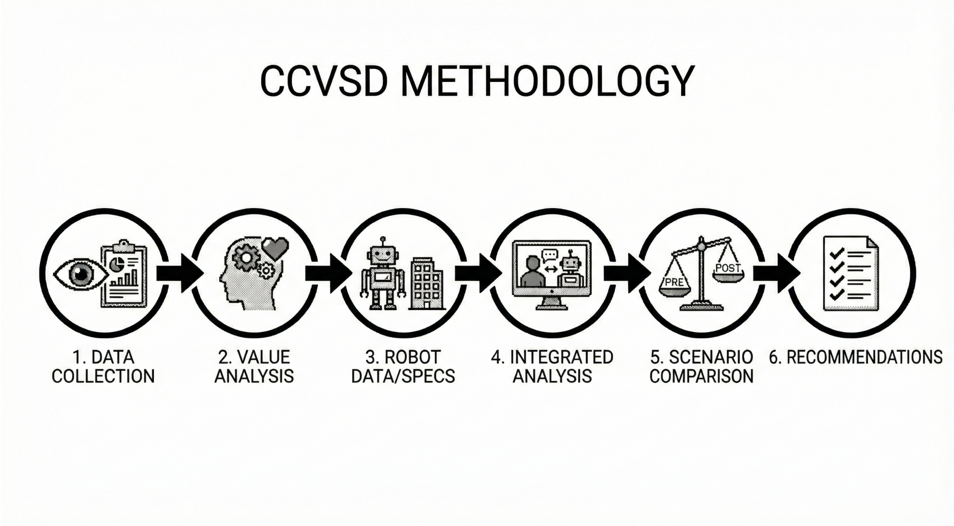


Figure 1.1: The standard phases of the Care Centered Value Sensitive Design (CCVSD) framework [62]. The process is modeled as a linear progression from context analysis to robot evaluation.

However, a structural analysis, substantiated by the empirical findings in Section 3.1.1, reveals that this sequence relies on a critical omission. As illustrated in Figure 1.2, the framework transitions directly from describing the robot (Step 3) to evaluating the robot (Step 4). It lacks an explicit phase for the iterative engineering work, prototyping, simulation, and coding, that connects these two states.

This transition overlooks the operational reality of robotic system construction. In practice, engineers often follow an iterative cycle of prototyping, simulation, and coding to bridge the gap between the initial context and the completed robot [44]. Because this extensive period of technical translation is not explicitly modeled within the standard CCVSD methodology, it functions as an invisible development phase, a normative void. Reflecting Agre's argument that technical architectures actively structure the practice they support [2], this void allows the ontological bias of standard engineering tools to silently overwrite intended care values with the default logic of efficiency. For instance, a navigation stack designed to minimize travel time will mathematically prioritize speed over a qualitative care value like approachability, simply because efficiency is the default objective function it was built to solve. Consequently, without a methodology to structure this phase, ethics remains a post hoc critique rather than a continuous design input. This thesis defines the central problem as this structural disconnect: the absence of actionable normative guidance capable of translating abstract ethical values into the rigid technical constraints required by the engineering process.

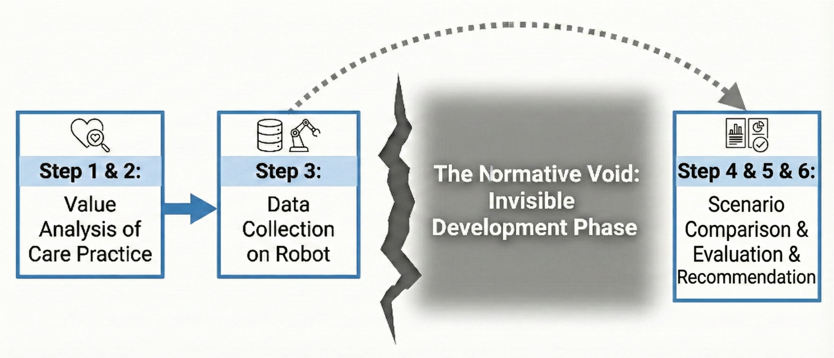


Figure 1.2: The structural discontinuity within the standard CCVSD framework. Note the gap between Step 3 and Step 4, which obscures the iterative technical development phase.

1.4. Research questions and objectives

The primary aim is to navigate the inherent tension between abstract ethical mandates and practical engineering workflows, exploring methods to translate normative values into technical specifications.

Main research question:

> How can the care-centred value sensitive design (CCVSD) framework be adapted and operationalised to provide actionable, engineer-centred guidance for the practical development of care robots?

To address this, the following research objectives (RO) are defined:

- **RO1 (Diagnostic):** Analyse CCVSD and identify limitations affecting its actionability during engineering development.
- **RO2 (Generative):** Generate situated intermediate-level knowledge through a constructive case study, serving as a preliminary attempt for bridging abstract ethics and technical implementation.
- **RO3 (Application):** Validate the methodology through a constructive case study involving social navigation design.
- **RO4 (Reflexive):** Refine the methodological tools based on empirical friction between ethical expectations and engineering realities.

1.5. Relevance and contribution

This study shifts the disciplinary focus from ethical persuasion, or convincing engineers to care, to methodological operationalisation, or enabling them to act. By rendering the previously invisible development phase explicit, this research creates contributions across three domains.

To the academic field of design ethics, it contributes a formalised, operational care centered value sensitive design framework that replaces static linear assessment models with generative design cycles.

To the domain of engineering practice, this research provides translational artifacts beyond the theoretical. Specifically, this includes the prospective value hierarchy, which equips practitioners to encode values into motion, alongside tangible engineering outputs such as a simulation relay node developed to resolve the ghost human problem in standard physics engines, and a forensic metrics suite designed to quantitatively measure social attributes like jerk and politeness.

Finally, to the healthcare sector, the thesis offers a pathway to robotic systems that are not only efficient but also relationally competent, thereby enhancing their acceptance among nursing staff and patients.

1.6. Overview of structure

This thesis is structured to reflect the dual-track nature of Research through Design (RtD), simultaneously documenting the technical construction of a social navigation system and the methodological operationalization of care ethics.

Chapter 2 (Methodology): Defines the *Dual-Track Strategy*, distinguishing between the technical case (Track 1) and the methodological framework (Track 2). It details the operationalization of the Care-Centered Value Sensitive Design (CCVSD) framework from a linear model into a *generative design cycle* and establishes the validation metrics for the PAL Tiago platform.

Chapter 3 (Results): Presents the empirical findings generated across the three phases of the design cycle. It details the Exploration phase (identifying the *Semantic Gap* and introducing the Prospective Value Hierarchy), the Development phase (exposing the *Normative Void* and detailing the Social Momentum implementation), and the Integration phase (reporting on *Contextual Instability* and the Generalist Trap).

Chapter 4 (Discussion): Synthesizes the findings to argue that the implementation gap stems from a deficit of *moral skill* rather than Moral Will. It interprets the normative void as a form of accumulated technical debt and critiques the inherent efficiency biases embedded within the engineering toolchain.

Chapter 5 (Conclusion): Summarizes the contributions, specifically the Intermediate-Level Knowledge (ILK) artifacts, and answers the research question regarding the translation of ethical intent into engineering specification.

Chapter 6 (Recommendations): Proposes future technical interventions, including a social negotiation layer for ROS 2 and policy guidelines to mitigate the *ex nihilo* fallacy in robotic development.

2

Methodology

This chapter outlines the methodological approach adopted for this thesis, which is situated at the intersection of robotics engineering and care ethics. To bridge the persistent gap between abstract ethical theory and concrete technical implementation, this study employs an interdisciplinary Research through Design strategy [68]. Rather than treating the design process merely as the means to a functional end, this approach utilizes the act of designing, developing, and evaluating a technical artifact as the primary method of inquiry [58]. By attempting to translate ethical values into engineering specifications, the research aims to generate intermediate level knowledge (ILK) that creates a bridge between normative philosophy and mechanical application [22].

2.1. The original CCVSD framework

To ensure the investigation remains grounded in established ethical theory, this research takes the Care Centered Value Sensitive Design (CCVSD) approach as its normative guide [62][61][66]. CCVSD emerged as a response to the lack of normative foundations in standard VSD. While VSD provides the method for value integration, it does not specify which values matter. CCVSD claims to solve this by anchoring design in the tradition of care ethics.

2.1.1. The centrality of the care practice

The core unit of analysis in this framework is the *care practice* Van Wynsberghe states that not all service interactions are care; a valid care practice must meet two strict ontological conditions: it must be a response to a need (distinct from a want), and it must involve reciprocity between the care-giver and care-receiver. In this specific view, the robot is evaluated not on its isolated technical abilities, but on how it shifts the distribution of roles and responsibilities within this back and forth loop.

2.1.2. The standard linear methodology

The original framework prescribes a six-step evaluation process. As summarized in Table 2.1, these steps focus heavily on the contextual analysis before and after the introduction of the robot.

As summarized in Table 2.1, Van Wynsberghe [61] articulates the framework as a linear sequence of six steps. However, preliminary engagement with the technical team indicated that these abstract ethical mandates were insufficiently actionable for engineering practitioners.

Therefore, to render the framework useful and align it with the iterative nature of robotic development [59], this research operationalizes the linear model into a generative design cycle comprising three distinct phases: Exploration, Development, and Integration. This adaptation serves as a translational mechanism, designed to convert static ethical theory into a dynamic workflow that engineers could actively adopt. The validity of this translation is subsequently analyzed in Section 3.1.2.

Table 2.1: The Care-Centered Value Sensitive Design Methodology (Adapted from [61])

Step	Action	Description
Step 1	Data Collection (Care Practice)	Analyzing the practice prior to the robot (thick description).
Step 2	Value Analysis (Care Practice)	Defining values, roles, and responsibilities in the current practice.
Step 3	Data Collection (Robot)	Describing the robot's capabilities and appearance.
Step 4	Value Analysis (Robot introduced)	Analyzing the practice with the robot introduced.
Step 5	Scenario Comparison	Comparing the results of Step 2 (Before) and Step 4 (After).
Step 6	Recommendations	Suggestions for design, re-design, or policy.

2.2. Research through design (RtD)

To investigate the implementation gap defined in Section 1.3, this study requires a methodology capable of generating, rather than merely observing, socio-technical friction. Therefore, the research adopts a constructive research through design (RtD) approach.

Following Koskinen et al. [27], this methodology states that theoretical gaps are best revealed through the attempt to build a working system. Purely sociological critique remains external to the engineering process; to understand why values are lost in practice, one must inhabit the constraints of the practice itself [61].

While standard design methodologies often imply a linear progression from requirements to validation [58], the nature of care ethics demands a more recursive approach. This is because care is not static but is an activity involving continuous thought and action [11]. Just as the practice of care requires constant tinkering and maintenance to adjust to the specific needs of the receiver, the design methodology itself must be recursive rather than linear [38]. Consequently, this methodology treats the linear steps of the CCVSD framework not as a rigid recipe, but as a flexible starting point that allows for adjustment as the design evolves. I attempt to execute the theoretical mandate of the framework and explicitly treat the resulting technical and ethical frictions as our primary data source. This allows the research to move from a theoretical imposition of ethics to a grounded operationalization, where the iterative cycle of design and testing serves to identify the specific knowledge required to align technical systems with the dynamic reality of human care values. In the context of this research through design study, this involves translating those identified frictions into intermediate level knowledge. This step derives validated design requirements to steer future prototypes toward ethical compliance, ensuring that the final guidelines are structurally coupled to the specific reality of the care practice.

All empirical phases adhered to strict ethical standards regarding informed consent and data usage.

2.2.1. Epistemological stance: beyond technical rationality

This research operates at the intersection of two distinct epistemological traditions: the positivist tradition of engineering, which seeks objective metrics for validation [34, 50, 26], and the constructivist tradition of care ethics, which views moral value as situated and relational [55, 45, 62].

Following Donald Schön's distinction in *The reflective practitioner* [49], this thesis identifies care robotics not as a problem of the high ground, where manageable problems lend themselves to solution through the use of research-based theory and technique [37], but as a problem of the swampy lowlands, where situations are confusing messes incapable of pure technical solution [44, 67]. Technical rationality is defined as an epistemology of practice wherein professional competence is equal to the instrumental problem-solving made and proven by the application of scientific theory [26, 7]. While effective for stable environments, this positivist approach fails when applied to the indeterminate zones of care practice, where problems are not given but must be constructed from messy, indeterminate situations [9, 39].

In standard engineering validation, success is often defined by the optimization of variables against a static ground truth [34, 32, 13]. However, this research posits that the primary challenge in care robotics is not the optimization of performance, but the translation of intent [43, 61]. Therefore, this thesis explicitly rejects a purely metric-based validation framework [15, 35]. Instead, it adopts a reflective practice approach to generate intermediate-level knowledge [68, 22, 30].

The validity of the proposed framework is not measured by whether it solved the ethical dilemma once and for all, but by whether it successfully operationalized the conflict for the practitioners involved [14, 64]. The friction documented in the following chapters is not a failure of experimental control, but the primary data source of this inquiry. By rendering these invisible normative conflicts visible, the framework contributes to the body of knowledge by diagnosing the structural barriers that prevent the integration of care values into engineering workflows [6, 65].

2.3. Research setting and co-design structure

To ensure this research addressed the real-world constraints of robotic development, the study was conducted during an internship at Heemskerk Innovative Technology (HIT) [20]. HIT is a commercial robotics firm with a historical foundation in nuclear technology and teleoperation, currently expanding its operations into the healthcare domain.

This specific industrial setting was selected for two strategic reasons. First, it provided access to professional robotics engineers who served as active participants in the design process. Second, the company's background in nuclear and teleoperation sectors provided a distinct control-oriented engineering culture. They are focused on safety, distance, and efficiency. This served as a valuable contrast to the relational, proximity-based requirements of care ethics, creating a productive friction for the research to investigate.

2.3.1. The dual-track validation strategy

To bridge the gap between abstract ethical mandates and engineering constraints, the methodology operates on two simultaneous levels. It is crucial to distinguish between the technical object (the navigational implementation) and the methodological process (the design framework).

Track 1: The technical navigational implementation

On the project level, the objective was to construct a functional social navigation algorithm for the PAL Tiago platform described in section 2.3.2. In this context, the engineers functioned as domain experts and technical partners. The validity of this implementation was assessed against the care requirements of nursing staff and patients (the care experts), focusing on the performance of the robot in shared spaces.

Track 2: The methodological framework operationalization

On the meta-level, the objective was construct and iteratively refine the operationalized CCVSD framework in parallel with the technical development. Here, the research engaged a methodological inversion: the engineering team was treated not as colleagues, but as the primary study participants.

Because the professional training of robotics engineers is often ontologically distinct from the care logic of nursing, the engineering team served as the primary participants for this methodological. The research utilized this setting to investigate a core premise regarding the utility of the proposed design tools. The objective was to determine if the intermediate level knowledge artifacts could effectively bridge the gap between abstract care values and the technical reality of engineering practice. If these artifacts resonated with the engineers as useful instruments for their existing or future development processes, this would constitute evidence that the methodology possesses applicable value. Consequently, the validity of this track was not measured by a binary proof of robustness, but by the extent to which the engineers could actively adopt these ethical tools into their workflow to facilitate the translation of normative requirements into design specifications.

This distinction necessitates a dual-track approach where technical failures are not treated as errors, but as epistemic data points. As illustrated in Figure 2.1, it is essential to understand the relationship between the execution of the design in track 1 and the evolution of the methodology in track 2. The workflow is driven by failure. The research attempts to follow the standard CCVSD phases in track 1 which is the technical case. When this attempt encounters a barrier denoted by the red dotted arrows representing friction the inquiry shifts to track 2. Here the methodology itself is analyzed and improved. This results in the creation or assimilation of a new tool or protocol which is then fed back into the technical workflow via the green solid arrows representing intervention. The structure of the phases including exploration, development and integration is derived from the standard CCVSD mandate. However the content of track 2 meaning the specific tools is novel to this thesis. The results of this thesis are the specific interventions generated to bridge the observed gaps.

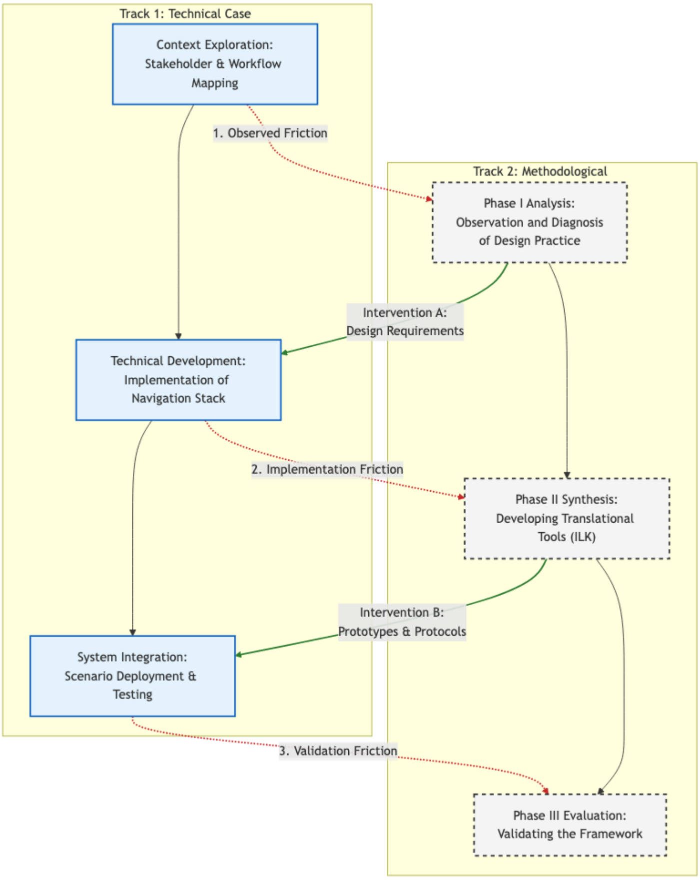


Figure 2.1: The dual track research through design strategy. Track 1 represents the attempt to execute the standard CCVSD phases including exploration, development and integration. Track 2 represents the reflexive layer where the limitations of the standard framework are diagnosed and repaired. The feedback loop consists of red dotted arrows representing friction and green solid arrows representing intervention. Red arrows show moments where the technical implementation in track 1 failed due to a lack of guidance such as the normative void in phase two. Green arrows represent the injection of new intermediate level knowledge artifacts developed in track 2 to resolve the friction and allow the technical work to proceed.

2.3.2. PAL Tiago

The experimental platform selected for this study was the PAL Tiago mobile manipulator. This selection was primarily driven by the ecological constraints of the research setting; as the study was conducted during an internship at Heemskerk Innovative Technology (HIT), the Tiago served as the active commercial platform within the company's existing workflow. Prioritizing this specific robot ensured that the case study remained grounded in the material reality of industrial practice rather than relying on idealized academic simulations.

The system is mechanically characterized by a holonomic drive base, enabling omnidirectional movement including lateral translation without prior rotation. For environmental perception, the platform relies on a primary 2D scanning LIDAR sensor mounted at the base of the chassis ($z \approx 20cm$). This sensor serves as the primary input for the navigation stack, restricting the robot's operational awareness to a 2D plane parallel to the floor. While the platform is equipped with an RGB-D camera located in the head unit for high-level visual processing, this sensor stream is distinct from the local costmap used for collision avoidance.

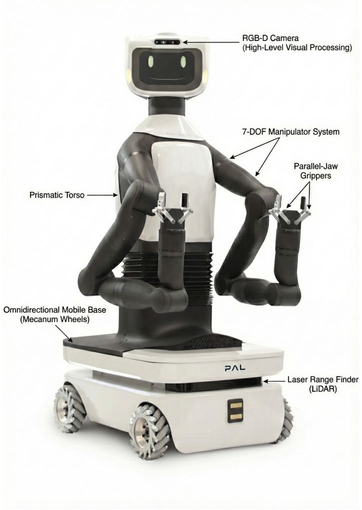


Figure 2.2: The PAL Robotics TIAGo Pro Dual-Arm Manipulator. The robot has a 2-DOF pan-tilt head housing an RGB-D camera for high-level visual processing and depth estimation. It features a 7-DOF articulated manipulator system terminated by parallel-jaw grippers. The torso is a lifting column for vertical adjustment. The base is an omnidirectional platform with four wheels for holonomic motion, equipped with a base-mounted Laser Range Finder (LiDAR) serving as the input for the navigation stack.

2.4. Phase I: Exploration (CCVSD steps 1 & 2)

The first phase of the methodology addressed the initial steps of the CCVSD framework, specifically data collection and value analysis of the care practice prior to the robot's introduction. The objective was to establish a normative baseline of the care environment.

2.4.1. Qualitative methods

To capture the complexity of the care environment, the research employed semi-structured interviews and participatory workshops. This choice was driven by the theoretical requirement to map the specific roles and responsibilities within a care practice [61]. Because these responsibilities are often asymmetric and implicit, representing a silent burden of care, they cannot be captured by purely quantitative surveys [55][28]. Therefore, to map this intricate web of care before the introduction of the robot, qualitative tools were necessary to produce the thick description required by the first step of the framework. This ensured that the definition of care was derived from the lived experience of the stakeholders rather than assumptions.

2.4.2. Ecological reconnaissance

To execute the exploration phase within the care domain, the study leveraged a collaborative project between Heemskerk Innovative Technology (HIT) and TU Delft, specifically established to map stakeholder values within the care facility. Rather than initiating a separate, redundant study, this research embedded the data collection process within this existing initiative to access the care environment without burdening care staff with yet another round of interviews.

Data collection employed a hybrid strategy of observation and ad-hoc inquiry. First, I observed and analyzed structured interviews conducted by the project partners. While these sessions adhered to the partner's independent methodology rather than the specific CCVSD framework, they served as a good source of qualitative data, revealing the nursing staff's value priorities and their unprompted perceptions of the technology. Second, this observational data was triangulated through ad-hoc questions that I directly asked to stakeholders, including nurses and movement therapists, during the robot's deployment. This pragmatic approach secured insights into the care practice, specifically capturing the native structure of their workflow while respecting the operational constraints of the healthcare environment.

2.4.3. Stakehoplder mapping

To satisfy the requirements of the CCVSD framework, the research employed a stakeholder mapping protocol. This method was utilized to explicitly identify the actors involved in the care practice as mandated by the first stage of the theoretical model. By synthesizing the qualitative data gathered during the ecological reconnaissance phase, the mapping process visualized the relevant human agents and their relationships to the robotic system. This established the structural baseline required to analyze the distribution of roles and responsibilities in the next phase.

Participants and setting

Data collection was conducted at Heemskerk Innovative Technology (HIT), a specialised robotics small to medium enterprise located in the Netherlands. The firm comprises a compact team of 10 to 20 professionals, supplemented by a cohort of graduate students. This environment served as the research setting due to the organization's openness to critically reviewing their own development workflow. Unlike strictly closed industrial environments, the company provided the necessary transparency to audit internal decision-making processes. They frequently engage in subsidy funded research and innovation projects alongside commercial activities, creating a culture characterized by experimental validation rather than purely mass production cycles.

The primary subject of development during this study was the PAL Tiago mobile manipulator. This specific platform was selected not merely for its technical capabilities but because it serves as the central artifact in the company's ongoing pilot projects in healthcare. The engineering team has previously deployed the PAL Tiago in various operational contexts including active hospital environments and residential care settings. Furthermore they have extensively modeled the robot for clinical integration and continue to utilize it as the baseline platform for upcoming healthcare initiatives. This history of practical deployment ensured that the technical frictions discussed during the interviews were grounded in empirical experience rather than hypothetical scenarios.

To investigate the engineering perspective regarding this system, semi structured interviews were conducted with the core technical team ($N = 4$). The participants were selected to represent the full decision making stack of a robotics project:

- Participant R1 (Project Lead): Responsible for the high-level integration of hardware and software (10+ years experience).
- Participant R2 (Robotics Engineer): Responsible for the implementation of the navigation stack and control logic.
- Participant R3 (Robotics Engineer): Focused on hardware constraints and operational safety.
- Participant R4 (Robotics Engineer): Responsible for the deployment of the system in the actual care facility.

Interviews lasted between 45 and 60 minutes and were recorded, transcribed. The questions can be found in Appendix F.

2.4.4. Thematic analysis

Data collected from these interactions underwent a thematic analysis focusing on socio-technical disconnects. The analysis specifically sought to identify the semantic gap between engineering definitions and care definitions. By coding the transcripts for latent value priorities, this strategy allowed the research to distinguish between the ethics of efficiency often held by management and the ethics of care held by frontline staff [62]. This differentiation provided the necessary inputs for the next design phase, ensuring that the robot was designed to address the values of those actually performing the care work.

2.5. Phase II: Development (CCVSD step 3)

The second phase addressed the normative void of development, a critical stage where social observations must be translated into technical architecture. While standard engineering tools often possess an ontological bias that treats robots as isolated agents operating in a static world, this research necessitated a constructive approach to overcome that limitation. I actively built the social navigation module to force the technical system to recognize its inherent interdependence with human actors [62].

2.5.1. Case study

To translate qualitative insights into technical specifications, the research utilized the PAL Tiago mobile manipulator. As noted in the theoretical literature, care is an inherently physical species activity that involves the maintenance of our world and bodies [11]. A disembodied screen-based agent cannot fully participate in the care giving stage of the practice. Consequently, the physical presence and kinematic capabilities of the Tiago allow it to intervene in the shared space of the hospital, making movement a functional proxy for distal care. By situating this robot within the constraint rich environment of a hospital hallway, the methodology creates a specific site of friction where the engineering drive for efficiency inevitably interacts with the ethical requirement for attentiveness.

To be explicit regarding the material execution of this methodology: this research entails the deployment of the PAL Tiago robot into a physical corridor populated by human actors. The robot is programmed with two conflicting navigational personalities. By strictly monitoring the robot's kinematic failures in this constraint-rich environment, the research generates the empirical friction necessary to validate the ethical framework. It moves the question from a theoretical discussion of care to a real world stress-test of the robot's ability to navigate social norms without causing physical or psychological disruption.

2.5.2. The social momentum implementation

To operationalize the experimental architecture, the navigation stack required a solver capable of optimizing nonlinear social cost functions. This necessitated a departure from the dynamic window approach, the default local planner utilized in the baseline ROS 2 navigation architecture. It is important to clarify that this architectural shift was dictated by the computational constraints of the social momentum framework rather than an independent preference for predictive control [36]. While the dynamic window approach effectively optimizes for instantaneous collision avoidance, or competence, by searching a limited velocity space, it lacks the horizon capability to optimize for the complex interaction dependent metrics required for social signaling [24]. Consequently, the model predictive path integral controller provided in the preliminary social momentum codebase was retained. The complete source code, architectural documentation, and forensic logging tools for this implementation are archived in Appendix D. This solver was selected for its ability to sample complex nonlinear trajectories that approximate social negotiation, a capability verified in the algorithmic implementation detailed in Appendix D.4.1.

The closed-loop architecture

As illustrated in Figure 2.3, the architecture functions as a high-frequency predictive control loop. The MPPI optimizer receives the current state from the ROS 2 wrapper and outputs optimal velocity commands directly to the drive wheels, bypassing standard trajectory planners.

The social momentum cost function

The core innovation of the implementation lies in the cost function (J), which serves as the mathematical translation of the *Legibility* requirement. The controller generates 250 candidate trajectories every 0.05 seconds while projecting the robot's state 3.0 seconds into the future. Each trajectory is scored against a composite cost function:

$$J(\tau) = w_{goal}C_{goal} + w_{safe}C_{obs} + w_{social}C_{momentum} \tag{2.1}$$

where the *social momentum* term ($C_{momentum}$) is defined by the cross product of the relative positions and velocities of the robot (r) and the human (h), as implemented in the vectorized Python logic in Appendix D.4.1:

$$C_{momentum} \propto -(r_{ac} \times v_r + r_{bc} \times v_h) \tag{2.2}$$

Behavioral Emergence

In theory, this mathematical formulation enforces topological stability (maintaining a consistent passing side) by penalizing changes in the robot's trajectory relative to the human agent:

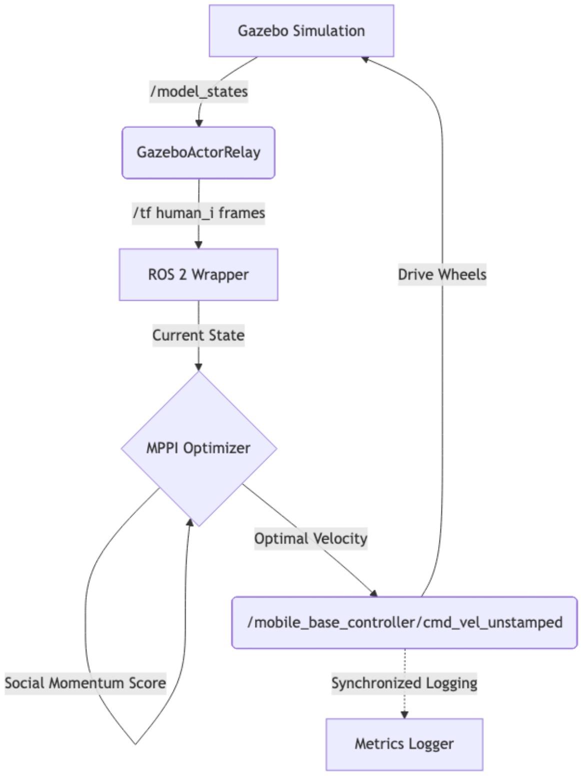


Figure 2.3: The closed-loop control architecture. The MPPI optimizer functions as the central reasoning unit, taking state estimates from the wrapper and generating kinematic controls based on the social momentum cost function.

- consistent sign: If the robot initiates a passing maneuver on the left, the cross product yields a specific sign. Maintaining this trajectory reduces the cost.
- sign flipping (energy barrier): If the robot attempts to erraticly switch to the right, the sign of the cross product flips. This results in an immediate spike in cost.

This energy barrier effectively penalizes hesitation by strictly enforcing topological consistency. By preventing erratic switching between passing sides, the optimizer ensures trajectory smoothness and generates *kinematic legibility* [36], enabling human observers to infer the robot's directional intent solely from its motion profile.

2.5.3. Simulation

Validating the robot's integration into a care practice requires observing it as an embodied actor. Care is inherently entangled with the body and cannot be evaluated strictly through disembodied code blocks or static analysis [62]. To evaluate the robot as an actor capable of reciprocal interaction, the methodology necessitated the creation of a simulation environment in Gazebo to serve as a relational sandbox. This simulation served as a methodological requirement to observe the robot's embodied presence in a shared space, allowing us to examine how the robot's physical behavior influences the dynamic exchange required by the theoretical framework. Relying exclusively on physical prototyping introduces a feedback loop of retardation, where the material constraints of hardware iteration lag significantly behind the speed of algorithmic refinement. As noted by Esterwood and Robert [10], the logistical costs and setup times associated with physical HRI platforms often act as a barrier to the iterative nature required to fine-tune social behaviors.

This methodology operationalizes simulation not only as an approximation of reality, but as a relational sandbox, a low-stakes environment that allows for the rapid development of design iterations. This allows the system to encounter and resolve edge cases of social interaction that would be dangerous or impractical to stage with vulnerable human subjects [46]. This ensures that the robot's ethical logic is profoundly embedded in the navigational software domain before being subjected to the friction of the real world.

Design justification and trade-offs

Operationalizing *attentiveness* required translating simulated social agents into semantic data compatible with the planner. An initial prototype (`fake_human_publisher`) validated the visualization approach but failed to synchronize with the simulation physics engine.

Consequently, the `gazebo_actor_relay` approach was selected to automatically extract actor kinematics from the Gazebo simulation, ensuring data consistency with the ground-truth physics defined in the launch files. This decision avoided the integration overhead of migrating to alternative engines while maintaining strict compatibility with the ROS 2 navigation stack.

Technical implementation

The relay node functions as a semantic injection layer between the Gazebo physics engine and the ROS 2 navigation stack. As illustrated in Phase 3 of Figure 2.4, the system implements the following data flow:

State Interception: The node subscribes to the `/model_states` topic, reading raw kinematic data from the physics engine at 100 Hz. This provides the ”Ground Truth” of the simulation.

Filtration: A regex filter isolates entities matching the `human_*` nomenclature, dynamically handling actors defined in the simulation launch file. This effectively distinguishes social agents from static infrastructure.

Transformation: The node broadcasts live TF frames (`human_0`, `human_1`) relative to the map frame. Crucially, the system enforces a strict *planar constraint* to resolve the ”supine actor” artifacts caused by 6-DoF simulation noise. As detailed in the artifact documentation (Appendix C), the node intercepts the raw quaternion and explicitly zeros the roll and pitch components, ensuring the planner receives a clean 2D heading (θ) that is strictly orthogonal to the ground plane.

Visualization: Simultaneously, the node publishes `visualization_msgs/MarkerArray` topics to RViz. These markers scale dynamically ($h = \sqrt{v_x^2 + v_y^2}$) to represent the real-time velocity vector of the simulated agent. This ensures that the robot’s internal *mental model* (RViz) is strictly synchronized with the visual reality of the simulation (Gazebo).

2.5.4. Technical validation

To validate these behaviors, the methodology required the creation of a specific measurement instruments to quantify the robot’s performance using metrics derived from Human-Robot Interaction (HRI) literature (Appendix A.1). These metrics served as the translation layer between abstract values and code in validation.

The study employed a dynamic elliptical geometric model for human agents rather than traditional static point mass representations (Appendix A.2.1). This modeling choice was made because personal space requirements extend more in the direction of travel than laterally [47]. By aligning the major axis of the ellipse with the velocity vector of the agent, the system achieved higher fidelity in the detection of intimate zone intrusions. These intrusions were defined strictly as distances of less than $0.45m$ based on the proxemic framework of Hall [17] (Equation A.1). This geometric approach enabled the quantification of care by measuring both the frequency of spatial violations and the time to collision metric [19] during dynamic interaction.

The average jerk metric (m/s^3) was selected as the primary kinematic indicator (Appendix A.2.2). This to quantify trajectory smoothness, which serves as a quantitative proxy for the robot’s motion stability. Drawing on foundational motor control literature, the minimization of jerk is established as the basic principle of natural, skilled biological motion [12] (Equation A.2). In the context of HRI, following these minimum-jerk profiles maximizes human acceptance, and high-jerk trajectories, detectable by abrupt changes in acceleration, are frequently correlated with perceived mechanical hesitation or algorithmic uncertainty [23]. By measuring the system’s jerk profile, we quantify the extent to which the robot exhibits confident motor control, a basic requirement for establishing trust in clinical environments.

A linear risk regression analysis was employed to track the correlation between robot velocity and human proximity (Appendix A.2.3). This analysis allowed the research to verify if the robot was actively

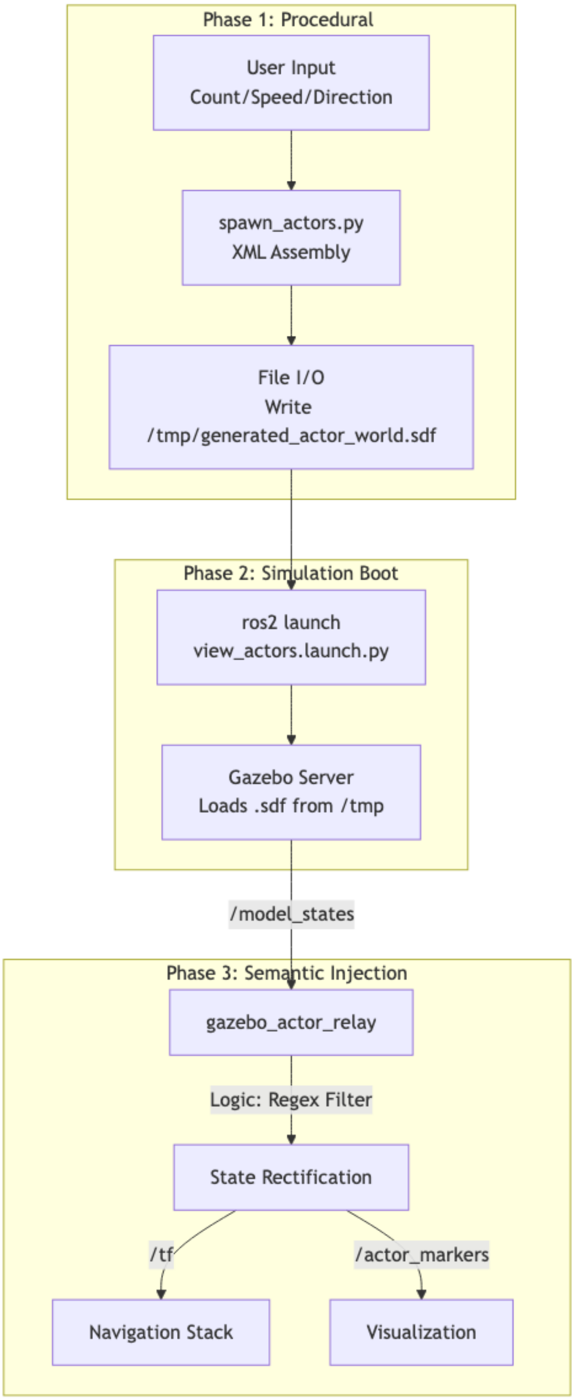


Figure 2.4: System Architecture Flowchart. The `gazebo_actor_relay` bridges the gap between the physics engine and the ROS 2 planning stack, injecting semantic data that allows the MPPI optimizer to factor social momentum into the final control output.

modulating its behavior in response to the presence of others, satisfying the bi-directional requirement of care interactions (Equation A.3).

Finally, the cost of care was measured via the Path Irregularity Ratio (PIR) (Appendix A.2.4). This metric quantifies the deviation from an optimal trajectory required to perform social maneuvers, measuring the trade-off between efficiency and social behavior [40]. By calculating the additional path length required to yield or maintain distance (Equation A.4), the research could explicitly measure the efficiency cost of acting *ethically*.

2.6. Phase III: Integration and evaluation (CCVSD steps 4, 5 & 6)

The final phase executed the comparative evaluation and synthesis steps of the CCVSD framework. This phase focused on determining how the introduction of the robot altered the established care practice.

2.6.1. Comparative evaluation between scenarios

To aid in a scientific scenario comparison, the methodology employed a comparative evaluation approach using paired scenarios with a substantial amount of stakeholders ($N = 20$). The validation protocol adhered to a maximum variation sampling strategy, selecting participants across the full ontological spectrum of the care practice, defined in subsection 2.4.3, to ensure diverse perspectives were represented. The choice of this sample size was guided by the principle of information power, which states that the utility of a sample in qualitative analysis depends on the specifications of the dialogue rather than statistical magnitude [33]. In the context of research through design, the robotic artifact functions as a co-elaborator designed to provoke latent knowledge rather than strictly validate engineering metrics [68]. The recruitment aimed not for the repetition of identical responses but for the identification of singularities. This aligns with the distinction between generality and repetition, where validity is derived from the specific, unique reality of an event rather than from an abstract general law [8]. Recruitment concluded upon reaching theoretical saturation, the point at which additional interviews ceased to produce more distinct categories of socio-technical friction [18]. By exposing participants to conflicting navigation profiles, contrasting standard efficiency driven behaviors against experimental socially aware behaviors, the protocol surfaced specific ethical priorities that remained invisible during abstract discussions or during the raw implementation of new efficient algorithms.

2.6.2. Synthesis

The analysis for this phase functioned as a normative evaluation, balancing the subjective feedback from stakeholders against the objective forensic logs generated in Phase II. By overlaying the qualitative nurse definitions of success onto the quantitative engineering definitions derived from kinematic metrics, the analysis identified specific value conflicts and contextual instabilities. Finally, these findings were synthesized into recommendations for design, re-design, and implementation (CCVSD Step 6).

3
Results

This chapter presents the findings generated through the research through design cycles defined in Section 2.3.1. Consistent with the epistemological stance established in Section 2.2.1 beyond technical rationality this research treats the design artifacts themselves specifically the re architected ccvsd framework and the simulation tools as primary empirical results. Following the definition by Zimmerman et al [68] these artifacts constitute intermediate level knowledge which are transferable methods generated to bridge the gap between ethical theory and engineering practice.

Consequently the results presented here are twofold. First are the methodological artifacts which include the specific modifications to the design framework. Second are the technical and social metrics which include the quantitative performance data and qualitative stakeholder responses obtained from the deployment of these artifacts.

To align with the dual track validation strategy this chapter is structured around three distinct interventions. As summarized in table 3.1 each section documents a specific cycle of friction which is the observed failure followed by the intervention which is the design construction and finally the validation representing the empirical outcome.

The initial phase of the research addressed the foundational gap between ethical theory and engineering practice. This required a critical re-evaluation of the problem space, moving from a sociological critique of engineers to a structural critique of the design tools they are provided.

3.1. From moral will to moral skill

This phase of the research investigated the implementation gap defined in Section 1.3.1, specifically regarding the motivation of practitioners. The empirical data gathered from the engineering team at HIT stands in contrast to the hypothesis of ethical apathy [1].

Despite the technical focus of the environment, the engineering team articulated a distinct ethical intent. When questioned regarding the primary drivers of the design process, the lead software engineer prioritized user utility over technical specification:

> *"Ideally, you develop for the user, not per se for yourself... The ultimate goal is to see if it helps. Is it effective? Is it something they [the patients] actually have a need for?"* (Participant R2).

However, a technical friction emerged when the conversation shifted to the implementation of this intent. The data reveals a discrepancy between the engineer's ability to maximize quantifiable metrics and their reported capacity to operationalize qualitative values. Participant R2 described this difficulty, noting that while physical constraints (e.g., payload) possess clear measurement standards, relational constraints (e.g., patient experience) lack comparable metrication:

---

[1] To ensure readability while preserving the original meaning, minor clarifications have been inserted into participant quotations using square brackets [...]

Table 3.1: Summary of research activities, artifacts, and epistemic insights

Phase	Activity (Input)	Artifacts (Output)	Epistemic insight (The result)
I: Exploration	Semi-structured interviews and stakeholder mapping	Intervention A: The prospective value hierarchy (PVH) and translation table	The semantic gap: confirmed that engineers and nurses use identical terms (e.g., responsiveness) to mean fundamentally different things (social vs binary).
II: Development	Code audit of ROS 2 navigation stacks and simulation construction	Intervention B: The social momentum implementation and gazebo relay node	The normative void: revealed that without explicit translation tools, ethical decisions are silently hard-coded as arbitrary constants (e.g., weights) inside the software.
III: Integration	Scenario deployment (N=20) and forensic log analysis	Intervention C: The contextual evaluation protocol and bifurcated guidelines	Contextual instability: demonstrated that a universal social robot is a fallacy; identical behaviors were rated as safe in hospitals but rude in care homes.

> *”If you want to measure the experience of the robot with a patient... that is subjective. That is inherently harder than measuring how many kilograms a robot can lift... Because you cannot measure it, it is difficult to take into account [during development]”* (Participant R2).

This lack of metrication directly impacts the daily workflow. When asked about his focus during these specific technical phases, he admitted that ethical values fade into the background:

> *”[Back then] I was actually just technically developing. So you don’t really think about [the values] that much... It is more just looking at, developing that technique”* (Participant R2).

The project lead explicitly linked this lack of measurable structure to the scheduling of the validation process. They noted that without a formalized mechanism to evaluate these subjective qualities, ethical validation is pushed to the end of the development cycle:

> *”We do this [validation] at the last chance... purely because there is no structure now”* (Participant R1).

These statements indicate that the engineers possess the specific intent to address patient needs, but report an absence of methodological artifacts to execute this intent during the early design phases. This structural deficit forces the team to rely on implicit experience rather than explicit guidance. Participant R3 corroborated this, stating:

> *”There is no structure now. We do those things a bit subconsciously... but it is not a fixed framework”* (Participant R3).

3.1.1. The normative void

The structural gap defined in Section 1.3.2 as the *normative void* was empirically validated during the framework evaluation sessions. When presented with the standard CCVSD methodology (Table 2.1), the engineering team explicitly identified a discontinuity between the definition of the robot (Step 3) and its evaluation (Step 4).

Participant R4 noted that while the framework functions for idealistic groundwork, it fails to account for the iterative reality of engineering, the phase where existing technologies are modified to fit the context. He specifically critiqued the lack of guidance for the technical phase:

> *”I think it [the framework] offers limited guidance for day-to-day development behavior... It talks more about what it should eventually be able to do, rather than about every small choice you make to handle that [during the process]”* (Participant R4).

He further noted that in practice, this is precisely where the friction occurs:

> *”I think often you come in [at step 3 or 4] with making an existing solution fit the problem... which is where you get stuck, or where the problems arise... and I think that is where this [framework] falls a bit short”* (Participant R4).

This observation confirms that the implementation gap is maintained not by a lack of will, but by a lack of process support during the technical construction phase. This void is visualised in Figure 1.2. In the absence of explicit methodological guidance to translate care values into technical specifications, the engineers were observed prioritising established technical utility metrics over the project’s ethical objectives.

3.1.2. The operationalized CCVSD framework

To resolve the structural discontinuity defined as the *normative void* identified in Section 1.3.2, I redesigned the linear CCVSD model (Figure 1.2) into a generative cycle (Figure 3.1). This methodological restructuring constitutes a primary result of the Research through Design process, transforming the framework from a passive assessment tool into an active construction tool.

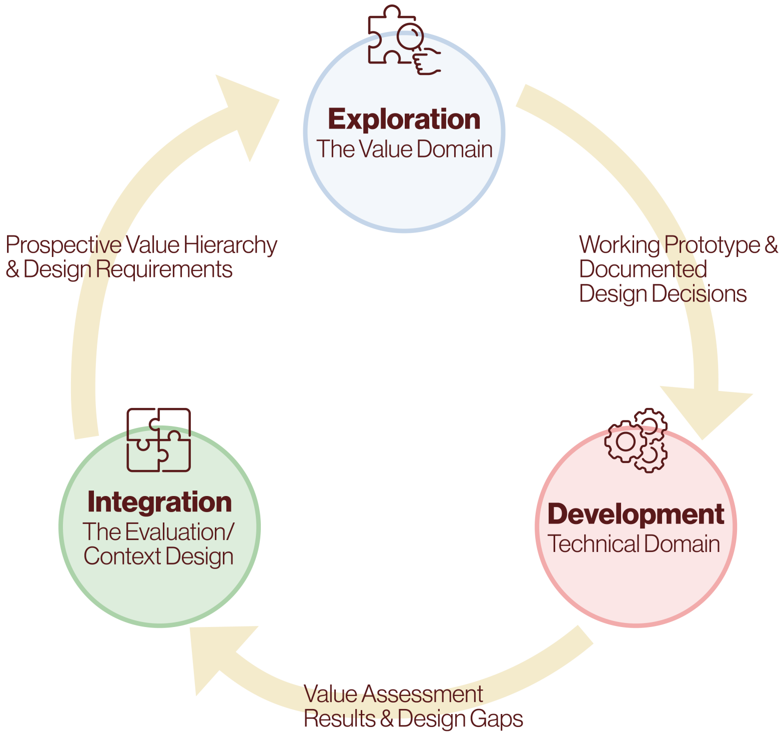


Figure 3.1: The Operationalized CCVSD Framework. The linear model is re-architected into a generative cycle, where each phase produces specific intermediate level knowledge artifacts that serve as structured inputs for the subsequent phase.

Semi-structured interviews with the engineering team revealed a critical friction: the linear model implicitly assumes that design begins with a blank slate. As Participant R4 noted, standard robotics engineering rarely proceeds linearly from values to product; instead, it involves the adaptation of existing assets where solutions are ’implemented’ rather than ’developed from scratch’. Consequently, the team reported that the linear CCVSD model caused development to ’get stuck’ between data collection

(Step 3) and evaluation (Step 5), as it lacked a mechanism to support the iterative testing required for engineering.

To resolve this structural discontinuity, I redesigned the framework to integrate the tripartite cycle architecture defined in Friedman's original VSD methodology [14]. Specifically, I remapped the linear phases into a recursive loop comprising empirical, technical, and conceptual investigations. I implemented this structural change to operationalize the iterative feedback required by the engineering workflow, treating the ethical evaluation not as a final gate, but as a continuous input.

To ensure this cycle functioned within the daily coding environment, I further refined the development phase using the *agile toolkit* architectural pattern defined by Umbrello and Gambelin [59]. Instead of functioning as a standalone theoretical layer, I integrated the VSD components directly into the existing workflow management tools. This structural modification ensured that micro-ethical trade-offs, specifically those related to hardware constraints, were formalized as traceable documentation artifacts within the iterative loop.

The validity of this cyclical restructuring was empirically confirmed during the internal design sessions. When the framework was initially presented as a linear checklist, engineering engagement was passive; however, upon visualizing the process as the generative cycle shown in Figure 3.1, the engineering team actively engaged with the specific feedback loops. The engineers explicitly recognized the cycle as compatible with their existing workflows, validating that the shift from linear mandates to cyclical iteration was a prerequisite for their adoption of the framework.

This new structure creates an epistemological and procedural pivot: moving from treating ethics as a static post-hoc evaluation to treating it as a continuous, iterative design input. This shift ensures that ethical alignment is reevaluated at each specific development step, preventing the value rigidities that otherwise block iterative progress. As shown in Figure 3.1, the framework is now split up into three iterative phases, each designed to resolve a specific empirical friction identified in the case study.

Exploration: To ensure that ethical requirements were grounded in reality rather than assumed *a priori*, this phase operationalized value identification as an empirical task. Guided by the normative categories of the CCVSD framework (e.g., attentiveness, responsiveness), specific stakeholder values were extracted directly from semi-structured interviews and situated observations. This phase functions as a translation engine: it takes these raw, empirically derived value statements and converts them into the Prospective Value Hierarchy (PVH), generating concrete design requirements that the engineering team can implement without further semantic interpretation.

Development: This phase occupies the normative void identified in the linear model. It takes requirements as input and produces a working prototype and documented design decisions as output. This phase was explicitly formalized to make the micro-ethical trade-offs of implementation visible and traceable, ensuring that technical compromises do not silently erode ethical intent.

Integration: To determine if the technical implementation actually embodies the intended care values this phase employs a contextual evaluation protocol. By subjecting the prototype to specific validation scenarios such as hospital versus care home the system generates empirical value assessment results specifically by correlating quantitative kinematic metrics such as jerk profiles against qualitative stakeholder feedback. This comparison exposes precise design gaps where the engineering logic failed to meet the normative care standard. These gaps serve as the input for the next exploration phase closing the loop and ensuring continuous value verification against the evolving technology.

This structural innovation does not claim to invent a new methodology but rather operationalizes a synthesis of two existing traditions. It retains the specific normative focus of CCVSD the care values defined by Van Wynsberghe [61] while restoring the methodological robustness of original VSD the iterative tripartite cycle defined by Friedman [14]. By correlating the values of CCVSD with the process of VSD this thesis contributes the specific layer of moral skill which consists of the translational artifacts such as the PVH and relay node required to make this theoretical union operable within the constraints of daily engineering practice.

3.2. Phase I: Exploration

The first phase of the generative design cycle addressed the theoretical mandate of CCVSD: the identification and analysis of values prior to technical implementation. While the mandate assumes a linear progression from understanding values to building the robot, the empirical application of this process immediately revealed a critical fracture in the design workflow.

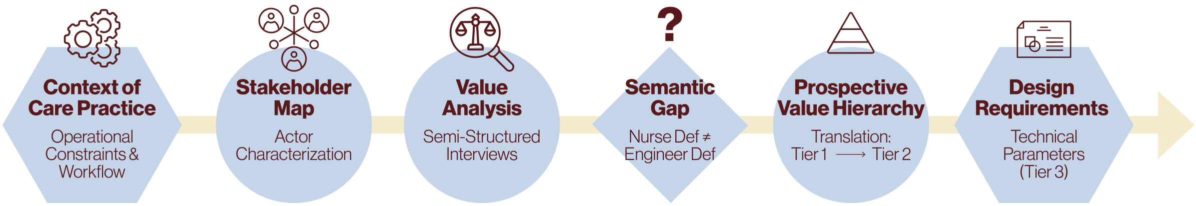


Figure 3.2: The exploration phase workflow protocol. This systems-engineering schematic illustrates the iterative loop between stakeholder mapping and value analysis, the diagnostic check for semantic gaps, and the final linear translation into technical specifications via the prospective value hierarchy.

3.2.1. The semantic gap

The finalized stakeholder map (Figure 2.4.3) generated through the participatory mapping sessions and dual track validation strategy defined in section 2.3.1 did not merely list actors it visualized a profound ontological divide. This divide represents a fundamental disagreement on the nature of the robotic artifact where engineers defined the system as an external tool while nurses defined it as an internal social actor.

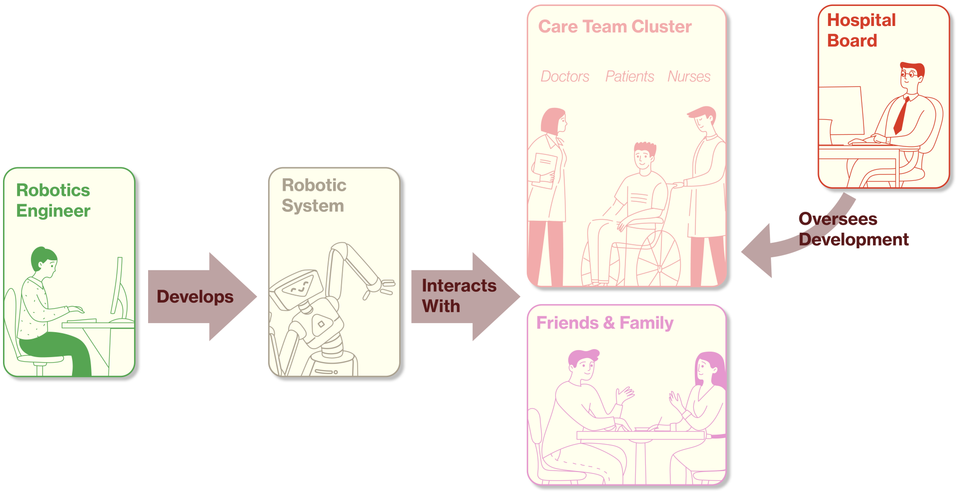


Figure 3.3: The Stakeholder Map revealing the structural disconnect between the Technical Domain (Engineers) and the Care Domain (Nurses).

The mapping process exposed that while engineers and nurses ostensibly share a workspace they occupy distinct domains of value definition. As illustrated in Figure 3.3 the engineering team positioned themselves as solvers external to the immediate care team cluster. In this visual model the engineer develops a robotics system which merely interacts with the care environment rather than existing within it. Conversely the nursing staff positioned the robot as a colleague internal to the relational loop subjecting it to the same social expectations as a human agent.

This structural divergence provided the physical evidence of the semantic gap a state where identical terms masked fundamentally different priorities. The map confirmed that the failure of implementation was not due to a lack of moral will but due to the lack of a shared language to bridge these two distinct domains. The initial engagement with stakeholders revealed that while engineers and nurses shared a moral intent the will to do good they lacked a shared ontology. This gap between distinct professional worldviews identifies a need for specific moral practices through which a shared ground can be established. Consequently this thesis defines the capacity to bridge this ontological divide and translate abstract intent into technical action as moral skill.

To visualize the breadth of perspectives gathered during this initial phase, Figure 3.4 details the specific roles of the stakeholders interviewed. This group, ranging from frontline care workers to robotics engineers, was essential for identifying the semantic gap early in the exploration process.

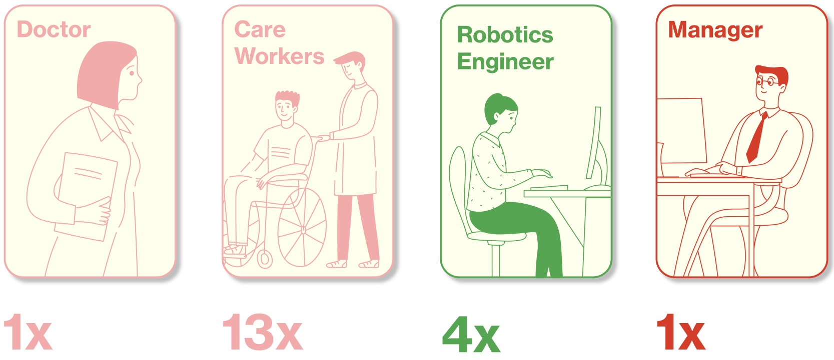


Figure 3.4: Stakeholder cohort interviewed during the ecological reconnaissance phase. the group includes care workers, robotics engineers, managers, and medical staff, representing the diverse ontologies engaged to define the initial requirements.

The findings of the exploration phase were not derived solely from semi structured interviews with four robotics engineers at HiT, but from active ecological reconnaissance within the nursing home environment (Section 2.4.2). This situated inquiry generated two distinct forms of empirical data. First, the physical convergence of stakeholders forced the conflicting definitions of care and efficiency into the open, confirming the presence of the semantic gap. Despite the ad hoc nature of the data collection the analysis achieved thematic saturation where continued inquiry yielded no new conceptual codes revealing a structural pattern characterized by high intragroup consistency and sharp intergroup divergence.

This divergence was visible for example in the definitions of reciprocity and attentiveness. As identified in the analysis the engineering cohort consistently interpreted reciprocity as a binary interface state such as a button click or led flash representing system acknowledgment. Conversely the care staff described reciprocity as a dialogical partnership a relational quality requiring mutual adaptation.

This specific epistemological split is empirically verifiable in the raw transcript data. When asked to define the robot capacity for contact the stakeholders provided the following mutually exclusive operational definitions.

> *Engineering Participant (R3): "Attentiveness is a trigger. If the sensor detects the patient leaving the bed it sends an alarm. The system knows the status because the sensor state changed."*
>
> *Care Participant: "It makes actual contact. It is not just moving it goes with you. It invites you. It is about the feeling that something is moving together with you creating a shared moment."*

These excerpts are representative of the broader dataset.

This linguistic disconnect confirmed that the implementation gap is not merely administrative but epistemological. The abstract values of care cannot be directly coded into the binary logic of a robot's navigation stack without a structured translational layer.

To move from raw interview data to the structured comparison presented in Table 3.2 a comparative analysis was applied using the CCVSD framework as a guide. It is important to note that the interpretations listed below are not intended as universal generalizations about the engineering or nursing professions at large but rather as the consolidated operational definitions that emerged within this specific project. The table documents the translation logic observed during the exploration phase highlighting where the technical operationalization diverged from the care intention.

Table 3.2: The semantic gap: Conflicting definitions of care values

Value	Definition	Nurse (care logic)	Engineer (tech logic)	Evidence
1. Attentiveness	*Noticing the need.*	**Reading the body** Holistic awareness of non-verbal cues.	**Sensor detection** Binary awareness of obstacles.	*Nurse:* "I see the patient myself... You can't see that on a robot." *Eng:* "Environmental awareness... That would be an important factor."
2. Responsibility	*Assuming the burden.*	**Workload relief** Ensures the robot relieves burden.	**Legal liability** Accountability for system failure.	*Nurse:* "It saved time... That saves a whole part of the human work." *Eng:* "If it goes wrong, who is responsible? You have to record that."
3. Competence	*Technical success.*	**Social flow** Success is patient trust.	**Functional output** Binary task completion.	*Nurse:* "People walk behind it... Then it works." *Eng:* "If it can't do tasks... it's in the bin."
4. Responsiveness	*The feedback loop.*	**Feeling heard** Patient feels acknowledged.	**System feedback** Status signals (lights).	*Nurse:* "You notice they start talking back..." *Eng:* "You press a button... then it says click."
5. Reciprocity	*Mutual exchange.*	**The coffee moment** Care is the interaction.	**Nice-to-have** Optional/subordinate.	*Nurse:* "I prefer care with the coffee included..." *Eng:* "Something like that is very nice to have... but subordinate."
6. Justice	*Fairness/Context.*	**Inclusivity** Fear of excluding elderly.	**Standardization** General solutions.	*Nurse:* "Older people don't understand that..." *Eng:* "Personal is nicer... but not handy."

3.2.2. The prospective value hierarchy (PVH)

To bridge the identified semantic gap where engineers prioritized efficiency and nurses prioritized relational care the primary result of this exploration phase was the integration of the prospective value hierarchy or PVH into the CCVSD framework. While the PVH structure is derived from Van de Poel's established work on value sensitive design [43] its specific application within the CCVSD framework constitutes a contribution of this study. This hierarchy is explicitly designated as prospective to distinguish it from the standard retrospective nature of ethical compliance; by forcing the translation of values into norms prior to technical implementation, the artifact ensures that ethics function as a generative design input rather than a post-hoc critique. This integration functions as an empirical result because it successfully operationalizes the previously abstract central value of care. By embedding

(a) The confrontation of ontologies

(b) Testing kinematic legibility

Figure 3.5: Empirical moments from the ecological reconnaissance phase. **(a)** The stakeholder circle brings the technical domain (engineers) and care domain (nurses) into physical proximity, making the semantic gap visible through direct dialogue. **(b)** An in-situ test of the robot attempting to translate the abstract value of responsiveness into the physical constraint of kinematic legibility, actively testing the robot as an embodied actor within the care home context.

Van de Poel's hierarchy into the exploration phase this research created the missing translational mechanism required to decompose normative care values into technical design requirements.

As illustrated in figure 3.6 the PVH functions as a translational stack with care at the apex flowing down into three distinct tiers abstract values tier 1 contextual norms tier 2 and design requirements tier 3. It is critical to note that while the visual model retains the standard label of design requirements to maintain consistency with Van de Poel [43], within the specific constraints of this project these outputs functioned as technical prerequisites. Rather than guiding ground-up prototyping, they served as the architectural selection criteria for phase II.

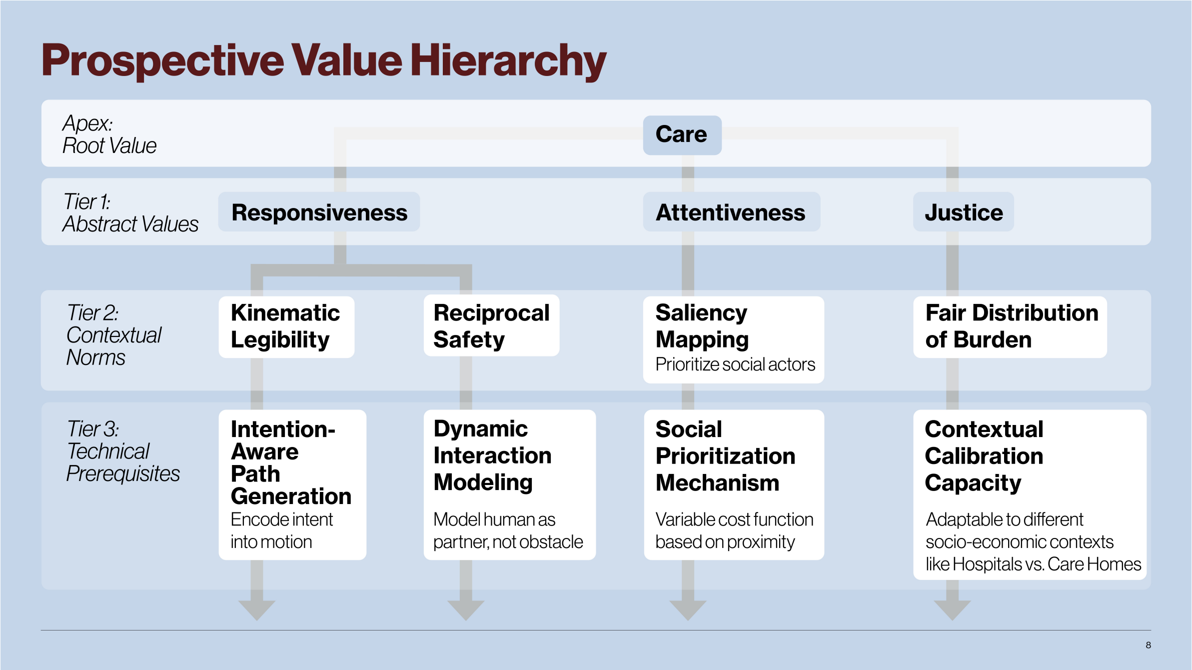


Figure 3.6: The prospective value hierarchy (PVH). A translational stack decomposing the root value of care into three tiers, values, norms, and requirements, to bridge the semantic gap between ethical intent and engineering logic [43].

Tier 1: the derivations of care The hierarchy begins at the apex with the fundamental value of care. Following the framework established by Van de Poel [43] this root value must be decomposed into specific values before it can be translated into norms. While the care ethics literature particularly the work of Ley [28] and Tronto [56] identifies a broad spectrum of moral elements the empirical data gathered in this study necessitated a focused selection. As illustrated previously in Table 3.2 the stakeholder analysis surfaced six distinct elements of care. However for the construction of the PVH this research isolates three: responsiveness, attentiveness and justice. These three values were selected because they represented the most significant common denominators across the interview data. Furthermore they constituted the primary places of semantic friction where the engineering definition of system utility most sharply conflicted with the nursing definition of care.

Translation A: from responsiveness to intention aware planning In the domain of care responsiveness is defined as the patient perception that a caregiver has reacted to their need. A robot cannot feel empathy therefore to operationalize this the PVH translates responsiveness into the contextual norm of kinematic legibility. This translation is grounded in the psychological safety feedback loop identified by Janssen [24] which isolates legibility as the primary action space principle for robot to human communication. By selecting this term the abstract relational expectation of responsiveness is converted into the tier 3 technical prerequisite of intention aware path generation. This requirement explicitly constrains the search space for phase II dictating that the selected planner must be capable of encoding intent into its motion profile excluding standard black box optimization approaches that prioritize efficiency over readability.

Translation B: from attentiveness to saliency mapping The value of attentiveness requires the suspension of self interest to notice the needs of others. In engineering terms noticing is a function of data prioritization. Consequently the PVH translates attentiveness into the norm of saliency mapping. This translation was chosen to leverage computer vision logic where specific features are granted higher weight than others. This norm drives the tier 3 technical prerequisite for a social prioritization mechanism. This requirement dictates that the technical architecture selected in phase II must support variable cost functions based on human proximity effectively requiring a system that can distinguish social actors from static obstacles.

Translation C: from justice to contextual calibration Finally the value of justice in care contexts refers to the fair distribution of burdens and the preservation of autonomy. The semantic gap analysis revealed that engineers often conflate justice with standardization. To correct this the PVH translates justice into the norm of fair distribution of burden. This norm was formulated to explicitly reject universal solutions in favor of adaptation. It necessitates the tier 3 technical prerequisite of contextual calibration capacity ensuring that the navigation architecture selected in phase II possesses the modularity to adapt its behavior to differing socioeconomic contexts such as the high efficiency needs of a hospital versus the high yield needs of a care home.

3.3. Phase II: Development

This phase addresses the normative void, the structural gap identified in the CCVSD framework where ethical requirements must be translated into technical specifications. It reports the development of the technical artifacts and methodological protocol designed to bridge this gap, detailing the friction encountered when attempting to encode care values within standard engineering infrastructures.

3.3.1. The ex nihilo fallacy

Standard VSD frameworks often presume a model of creation *ex nihilo*, implying that engineers build systems from scratch to fulfill a moral purpose. The development phase of this case study empirically refuted this assumption.

The design process was fundamentally path dependent. The system was not created from the ground up, but assembled from existing open-source stacks (ROS 2, Nav2) designed to maximise developer utility rather than care values. This reliance on pre-existing tools did not create the normative void, but rather exposed the danger of leaving this phase structurally empty. In the absence of explicit ethical constraints the normative void was immediately colonized by the default logic of the underlying architecture. An invisible space where the default setting of efficiency silently overrode the ethical design requirement (care) without the engineer's active consent.

3.3.2. Development phase protocol

To counter the path dependency identified in the previous section, the development process was formalized into a structured protocol. Drawing on the Basic Design Cycle [60], the standard engineering workflow was reframed to explicitly include ethical frictions as design inputs. By structuring the development phase as a rigorous cycle of analysis, synthesis, and simulation, the moral skill of the engineer is moved from an abstract intention to a documented procedure.

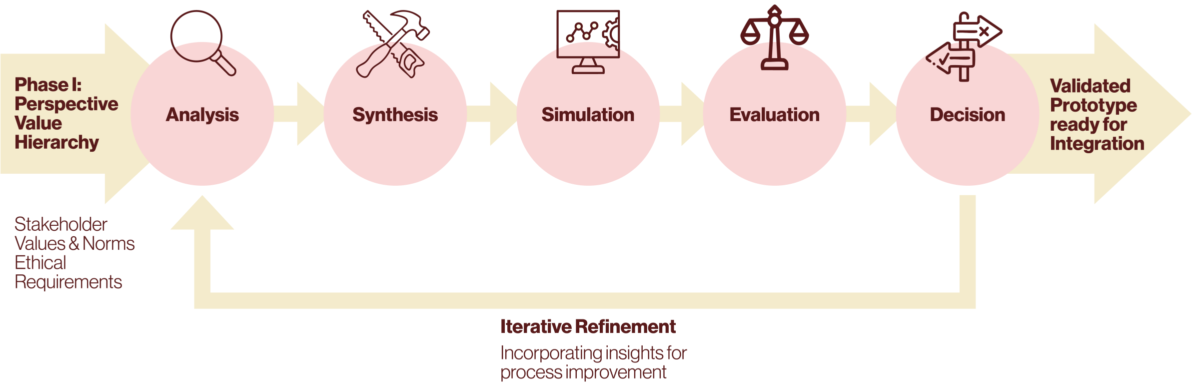


Figure 3.7: The development Phase protocol. Adapted from the basic design cycle. The standard engineering process (center) is reframed through ethical frictions (left) and technical interventions (right) needed to bridge the gap between abstract values and concrete code.

3.3.3. The normative rejection of standard planners

The execution of the development phase protocol produced a critical empirical finding: the ethical requirements generated in phase I were inadequate with current standard industrial tooling. The application of the prospective value hierarchy (PVH) necessitated a deviation from the standard ROS 2 Nav2 stack.

While Nav2 represents the engineering standard for efficiency, the attentiveness requirement exposed its limitations in the competence domain. As illustrated in the conceptual taxonomy (Figure 3.8), this focus restricts the baseline exclusively to safety and smoothness. Consequently, when obstructed by a human, the baseline defaults to discrete recovery behaviours such as freezing or rotating, often resulting in functional deadlock due to a lack of higher-order social reasoning.

Consequently, the framework necessitated the selection of the social momentum framework. This selection was not merely a technical preference but a normative necessity. The kinematic legibility constraint derived from the nursing interviews explicitly disqualified descriptive models like the social force model (SFM) [21], which lack intent signaling. Instead, the comparative analysis framework established by Janssen [24] identified social momentum as the unique solution capable of bridging this gap. The social momentum framework [36] was selected for its explicit maximization of angular momentum, which mathematically resolves navigational ambiguity by signaling passing side intentions, directly satisfying the legibility requirement derived from the ethical analysis.

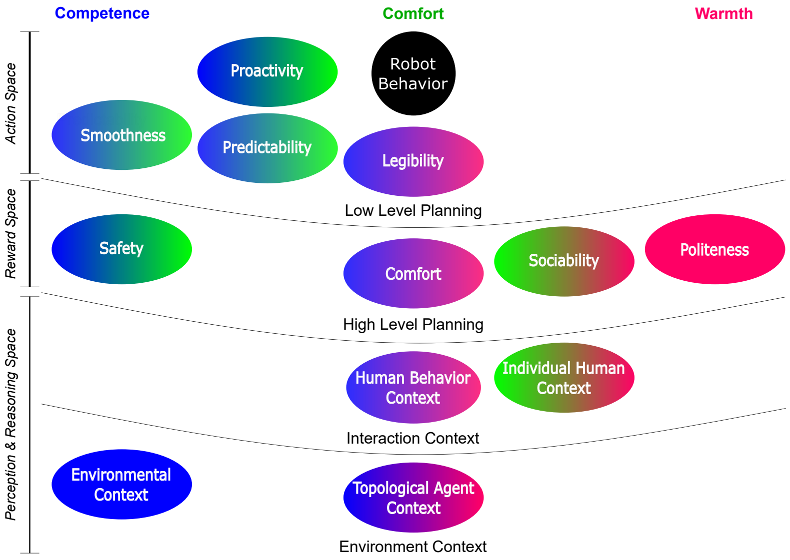


Figure 3.8: The Socially Aware Navigation Evaluation Framework. This diagram maps robotic capabilities against the social value spectrum. The baseline occupies the Competence domain, while the experimental architecture targets the Comfort domain via Legibility. Adapted from the framework by Janssen.

3.3.4. The MPPI implementation

While standard engineering conventions typically categorize system architecture as methodology, this implementation is reported here as an empirical result. This distinction is made because the specific architectural configuration of the MPPI controller constitutes the material evidence of the normative void. The code structure itself documents the necessary deviation from standard industrial practices required to accommodate care values, serving as the technical proof that ethical intent requires a fundamental restructuring of the control logic.

It is critical to distinguish the contribution of this thesis from the antecedent literature. While Mavrogiannis et al. [35] provided the mathematical concept of Social Momentum, no functional ROS 2 implementation existed capable of real-time execution on the PAL Tiago hardware. Therefore, this research generated the novel control stack required to operationalize the theory. This involved engineering a custom MPPI solver based on some rudimentary initial coding implementations up to integrate with the ROS 2 lifecycle, specifically rewriting the dynamics constraints to bridge the gap between the theoretical model and the robot's actual physical limits.

3.3.5. The cost of moral skill

The operationalization of this architecture yielded a critical finding regarding the computational cost of care. Achieving the required social legibility necessitated a fundamental departure from standard efficient planners (like the Dynamic Window Approach). The system required sampling 250 stochastic trajectories every 0.05 seconds (see Section 2.5.3 for architectural details), a computational load significantly higher than the baseline. This necessity demonstrates that social behavior is distinguishable from efficient behavior by its high computational tax.

Furthermore, a specific friction emerged regarding the robot's holonomic capabilities. The PAL Tiago is mechanically capable of instantaneous lateral movement. However, the forensic logs indicated that such movement creates a *legibility void* where the robot moves efficiently but without the rotational cues humans use to predict intent. To resolve this, the moral skill involved deliberately constraining the robot's control authority. As detailed in Appendix D.4.3, the solver was restricted to a non-holonomic unicycle model with hard-coded limits ($v_{max} = 0.4$ m/s, $\omega_{max} = 1.5$ rad/s). This forced the robot to rotate its base before translating, thereby restoring the necessary gaze cues for social predictability. This constraint provides physical evidence that moral skill is not merely an additive feature but a subtractive process, where the robot's maximum theoretical efficiency is deliberately sacrificed to maintain the social trust of the user.

3.3.6. The hidden weights of ethics

A deeper dive of the inherited social momentum codebase revealed a significant discrepancy between the theoretical cost function (Equation 2.1) and its practical implementation. While the mathematical model presents the weighting factors (w_{goal}, w_{social}) as tunable variables, the actual Python implementation relied on hard-coded scalars buried within the optimization loop (see Appendix D.4.2).

Specifically, the *momentum consistency* term was assigned a reward weight of `-11.0`, while the *Goal Progress* term was assigned a weight of `2.0`. This implicit ratio effectively programmed the robot to value *politeness* (trajectory adherence) 5.5 times higher than *efficiency* (goal speed). Crucially, these values were not exposed as configuration parameters but were defined as static constants, requiring source-code modification to adjust.

This finding serves as the material evidence of the *normative void* (Section **??**), identifying a specific instance where ethical trade-offs were executed via arbitrary constants rather than deliberate design parameters.

3.3.7. Kinematic legibility

A critical friction emerged regarding robot kinematics: the PAL Tiago's holonomic base allows instantaneous lateral movement, which, while efficient, removes the communicative intent normally conveyed through rotation. Testing showed that this sliding motion made the robot's behaviour unpredictable to humans. To restore kinematic legibility, the navigation system was deliberately constrained to enforce a rotation delay before translation, demonstrating that a degree of technical inefficiency is sometimes necessary for social interpretability.

3.3.8. The gazebo relay architecture

The infrastructure gap identified in the Exploration phase manifested technically as an *ontological bias* within the simulation environment. Standard ROS 2 Nav2 stacks rely on local costmaps populated by laser scan data. However, in the default Gazebo Classic configuration, animated actors (`human_*` models) are rendered visually but often fail to register on standard collision layers.

To resolve this, the `gazebo_actor_relay` middleware was developed. This component functions as the runtime core of the broader gazebo_actor_spawner artifact (fully documented in Appendix C).

This technical failure constitutes a primary empirical finding: it demonstrates that current open-source tools default to a solitary world model where dynamic social agents are treated as static background noise. The development of the gazebo_actor_relay was not merely a setup step, but a necessary ontological correction. By forcing the physics engine to broadcast human states as semantic entities (technical implementation detailed in Section 2.5.3), the architecture exposed that standard engineering tools currently lack the native capacity to represent the social other.

3.4. Phase III: Integration

The final phase of the generative design cycle addressed the integration and validation of the social navigation module. This phase examined whether the design requirements generated in Phase I and technically implemented in Phase II succeeded in bridging the gap between care values and engineering logic when exposed to the unscripted reality of the care environment.

3.4.1. Standard scenario comparison

Following the protocol of CCVSD Step 5, the objective was to perform a *scenario comparison*. This method presumes that introducing a robot into a care environment yields a stable, measurable outcome that can be evaluated against a pre-robotic baseline. The mandated evaluation was binary: determining whether the robot supported or hindered specific care values (e.g., *attentiveness*, *competence*) across the five validation scenarios defined in Appendix A.

To ensure transparency regarding the stimuli shown to stakeholders, the complete video datasets for the three primary scenarios are available online[2]. Figure 3.10 presents a sequential breakdown of scenario 2, illustrating the distinct kinematic behaviors that served as the primary prompt for the responsiveness discussions in the elderly care context.

(a) Reception **(b)** Interaction **(c)** Cafetaria

Figure 3.9: The context of the Antwerp care home. Unlike the standardized hospital corridors, this environment functioned as a living lab characterized by low-velocity mobility, high social density at hubs (a), and non-standard behavioral cues from residents (b). The videos of these interactions are available as digital footnotes for context.

While the sequential screenshots in Figure 3.10 provide a visual timeline of a single interaction, it is necessary to verify that this behavior is systematic rather than anecdotal. To this end, a quantitative top-down analysis was conducted using the forensic dashboard (detailed in Section 3.4.2). Figure 3.10 aggregates the trajectory data from three random runs for each architecture to visualize the structural consistency of the navigation logic.

As illustrated above, the Nav2 baseline consistently waits until the last possible moment to react (characterized by sharp, late deviations or straight lines), whereas the Social Momentum planner initiates a smooth, anticipatory curve. This confirms that the ”politeness” observed in the video was not an accident, but a reproducible geometric feature of the algorithm.

For a detailed breakdown of individual stakeholder preferences and algorithmic choices per scenario, refer to Table F.1

3.4.2. Quantitative forensic analysis

To validate the hypothesis of *contextual instability* beyond subjective interview data, a forensic report card” pipeline was utilized to extract hard kinematic metrics from the ROS 2 navigation logs. This automated analysis processed the robot’s performance across three key dimensions: safety, efficiency, and social compliance.

The safety-comfort trade-off

The analysis of the average Jerk metric (the mean derivative of acceleration) revealed a physical corollary to the generalist trap.

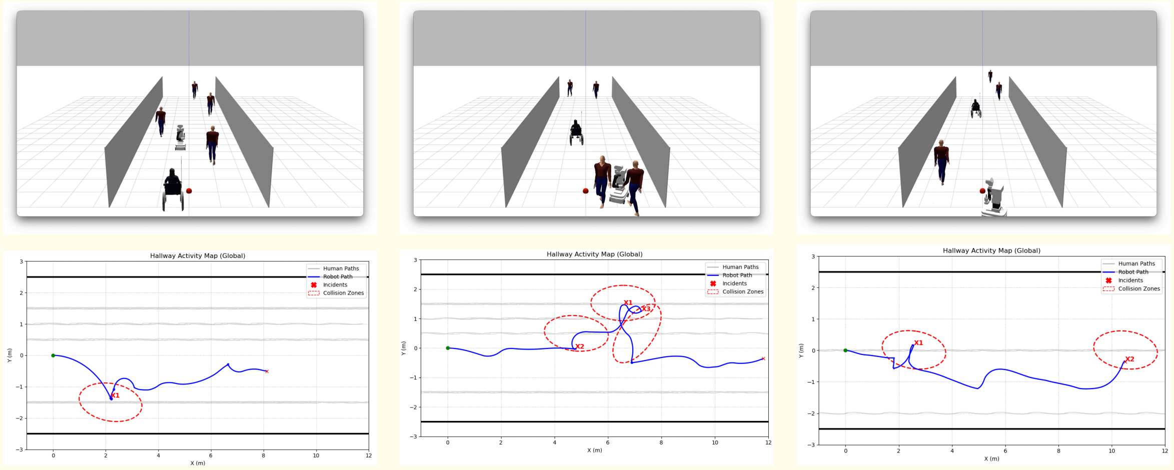
- **Metric:** A threshold of $< 5.0\,m/s^3$ was defined as the limit for passenger-friendly smoothness.
- **Result:** Analysis of the kinematic data (Figure 3.14) confirms that the Nav2 baseline produced significantly higher Jerk values (lower smoothness) compared to the social momentum controller. When viewed alongside the stakeholder preference data (Figure 3.15), it is evident that hospital stakeholders overwhelmingly chose this profile despite its kinematic roughness. This suggests that smoothness is not a universal proxy for good care; in high-efficiency contexts, the rougher but more predictable behavior was preferred, whereas in care homes, the smoother yielding behavior was favored.

---

[2] Scenario 1 (mixed traffic): `https://youtu.be/cI-61LHO-Ts`, Scenario 2 (narrow pass): `https://youtu.be/FPe2MyLHKK4`, Scenario 3 (emergency): `https://youtu.be/YFfYTw2C3Vk`

A. Standard Nav2 Stack (Baseline)

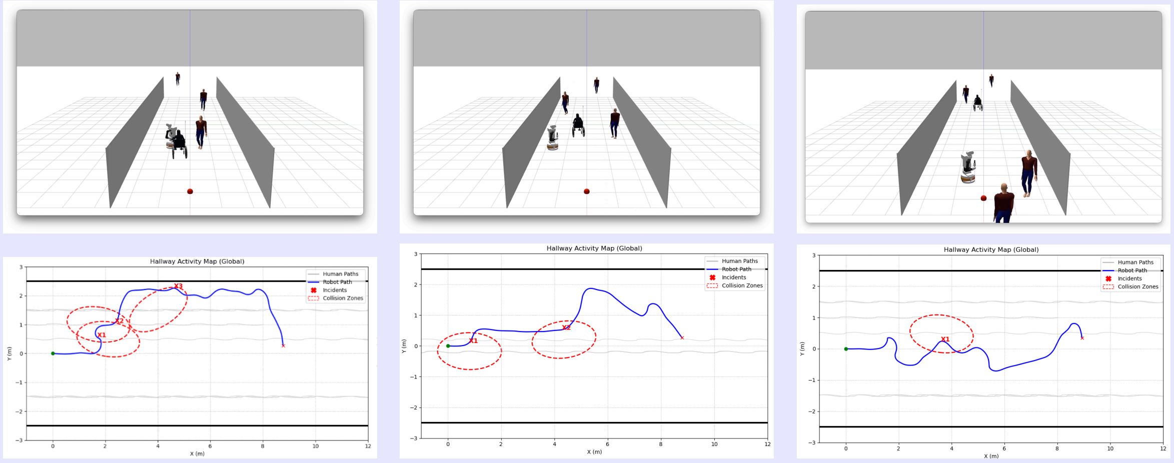
Scenario 2



Run 1 Run 2 Run 3

B. Social Momentum (Experimental)

Scenario 2



Run 1 Run 2 Run 3

Figure 3.10: Integrated comparison of navigation logic. **Block A (Yellow):** The Nav2 stack creates hesitation visually (top row) and generates rigid, straight-line trajectories quantitatively (bottom row), confirming a reactive safety logic. **Block B (Blue):** The Social Momentum stack creates smooth yielding visually (top row) and consistently generates anticipatory deviation curves quantitatively (bottom row), confirming a proactive social logic.

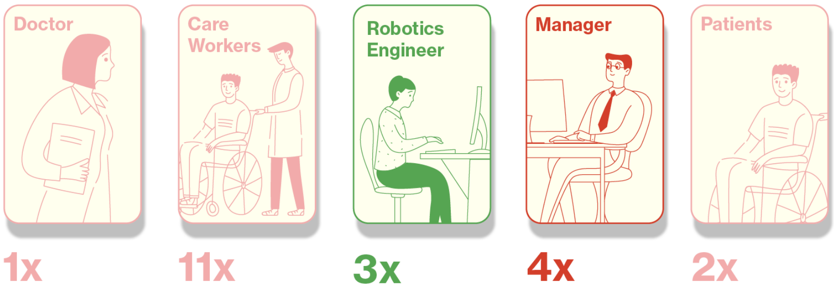


Figure 3.11: Visual breakdown of the study participants $(N = 20)$. Consistent with the maximum variation sampling strategy, the interview pool covered the full spectrum of the care practice. The participants included 7 care workers, 3 robotics engineers, 3 managers, 1 doctor, 4 support staff, and 2 user/support representatives, ensuring a diverse range of perspectives for the comparative evaluation.

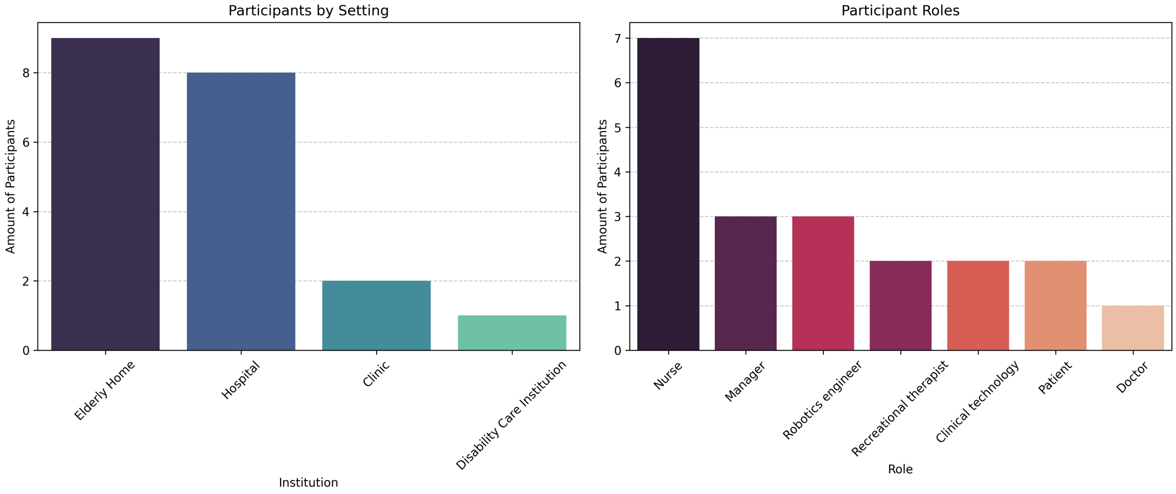


Figure 3.12: Distribution of interview participants by institution (left) and professional role (right).

Politeness trends

The compliance trend metric, a regression of robot velocity vs. human distance, provided a statistical measure of politeness.

- **Data insight:** The social momentum controller demonstrated a consistent positive correlation, where velocity decreased linearly as proximity to the human increased $(d \to 0, v \to 0)$.
- **Contextual friction:** While there was a strict linear deceleration effectively signaled intent to the human, the time-to-collision” (TTC) monitor indicated that this often triggered false positives in narrow hallways, causing the robot to freeze $(TTC < 2.0s)$.

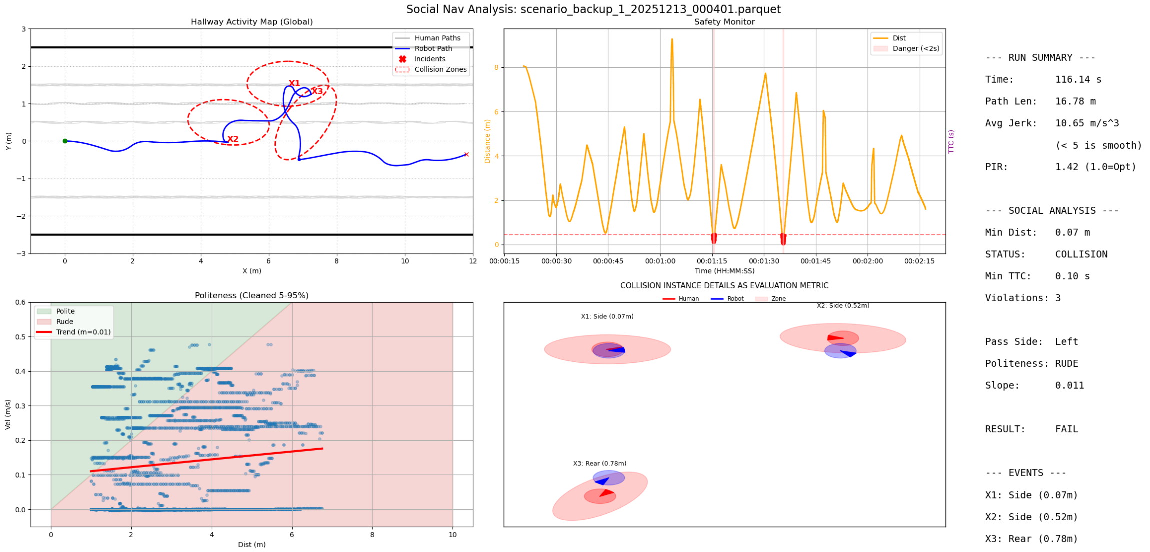
This data confirms that the social ellipse $(1.2m \times 0.6m)$ is not a static safety boundary but a dynamic social boundary that must contract in high-density environments (like hospitals) to prevent system deadlock.

Interpreting the forensic report card

To navigate the interaction data, the forensic dashboard (Figure 3.13) decomposes each simulation run into four diagnostic quadrants. Note that while the following breakdown focuses on interpreting the kinematic results, the full technical specification of the dashboard's architecture and the signal processing pipeline is detailed in Appendix A.

- The **hallway activity map (top-left)** provides the overhead view of the interaction, showing whether the robot maintained a straight line or yielded.
- The **safety monitor (top-right)** tracks the ”pressure” of the interaction; sharp drops in the orange line indicate sudden proximity events.
- The **politeness plot (bottom-left)** validates the social logic: a distinct positive slope confirms the robot is slowing down as it gets closer (v $\propto$ d), whereas a flat cloud indicates indifference.

- The **incident reconstructor (bottom-right)** isolates specific failure modes, distinguishing between valid safety stops and ”frozen robot” errors.



```
--- RUN SUMMARY ---
Time:       116.14 s
Path Len:   16.78 m
Avg Jerk:   10.65 m/s^3
            (< 5 is smooth)
PIR:        1.42 (1.0=Opt)

--- SOCIAL ANALYSIS ---
Min Dist:   0.07 m
STATUS:     COLLISION
Min TTC:    0.10 s
Violations: 3

Pass Side:  Left
Politeness: RUDE
Slope:      0.011

RESULT:     FAIL

--- EVENTS ---
X1: Side (0.07m)
X2: Side (0.52m)
X3: Rear (0.78m)
```

Figure 3.13: The forensic report card used for run-level analysis. **Top-left (hallway activity map):** the robot’s path (blue) relative to human actors (grey), visualizing deviation maneuvers. **Top-right (safety monitor):** the minimum distance to the nearest human over time; dips below the red line indicate safety violations. **Bottom-left (politeness trend):** a regression of robot velocity (y) vs. human distance (x); a positive slope indicates social reactivity. **Bottom-right (incident geometry):** a reconstruction of the exact spatial configuration during collision/freeze events.

Following the forensic micro-analysis of individual runs (Figure 3.13), Figure 3.14 presents the aggregate kinematic metrics for all validation trials ($N = 20$) across the three scenarios.

Kinematic smoothness (jerk) As illustrated in Figure 3.14 (Top right), the Nav2 architecture recorded a median average jerk of approximately 12–14 m/s^3 in the public hallway scenarios (S1, S2). In contrast, the Social Momentum architecture recorded a lower median jerk, typically ranging between 3.0 and 4.0 m/s^3, with a narrower interquartile range (IQR).

Path efficiency (PIR) and proximity The path irregularity ratio (Figure 3.14, bottom right) shows an inverse trend. In scenario 2 (narrow pass), the social momentum algorithm exhibits a higher median PIR (approx. 1.5) compared to the Nav2 baseline (approx. 1.2). Simultaneously, the minimum distance to human plot (top right) indicates that the social momentum agent’s trajectory frequently intersected the 0.45 m boundary, labeled as the intimate zone, whereas the Nav2 planner generally maintained a larger separation distance in scenario 1.

3.4.3. Contextual instability

Implementing the validation protocol with $N = 20$ stakeholders revealed a fundamental flaw in the standard scenario comparison model. The data showed that there is no unified care practice against which to evaluate the robot; instead, distinct and sometimes conflicting micro-climates of care emerged; local environments defined by opposing normative expectations. Specifically, behaviors validated as competent in the hospital context (e.g., maintaining velocity) were explicitly rejected as uncaring in the residential context, proving that a single algorithmic definition of safety cannot span both domains.

This instability was real and happened in the actual environment of a care home. As shown in Figure 3.9, the testing environment was very different from a standard hospital corridor. The reception area (a) was a place where people gathered rather than just a walkway. The residential hallways (c) contained residents who moved in unpredictable ways (b). The data discussed below must be understood in this specific context. The physical setting made it clear that the robot’s focus on efficiency caused problems.[3]

---

[3] You can see the raw videos here: `https://youtu.be/SFEGFVNbvqk` and `https://youtu.be/QCuu1f2yhNA`

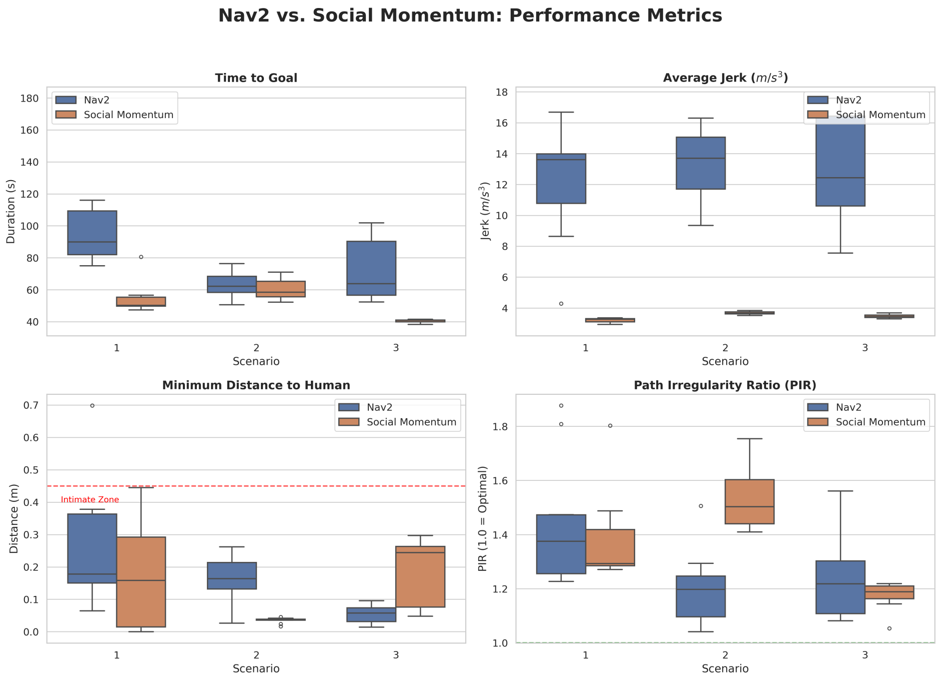


Figure 3.14: Comparative performance metrics for the Nav2 and Social Momentum architectures across three scenarios ($N = 20$). Boxplots display the median and interquartile range (IQR) for Time to Goal, Average Jerk (m/s^3), Minimum Distance to Human (m), and Path Irregularity Ratio (PIR). The dashed red line in the bottom-left plot indicates the 0.45 m “Intimate Zone” boundary.

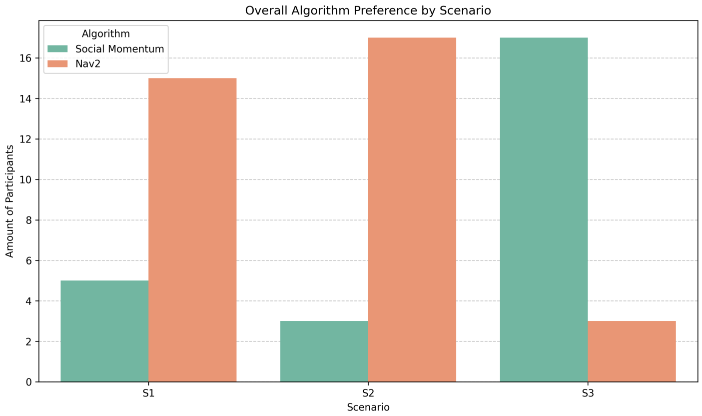


Figure 3.15: Overall algorithmic preference across the three test scenarios (S1: Busy Hallway, S2: Narrow Pass, S3: Emergency).

The inversion of variables

The analysis demonstrated that several design variables considered static from an engineering viewpoint are in fact dynamic and invert depending on context. While quietness is a standard requirement for hospital robots to respect patient rest, interviews indicated an inversion of this value in elderly homes. As one roboticist noted, noise is often unimportant due to widespread hearing impairments among residents.

Divergence by institutional context

As summarised in Table F.1 and visualised in Figure 3.15, evaluations of the robot’s behaviour were inconsistent across sites. Identical algorithmic behaviours produced opposing value judgments depending on institutional context.

The hospital context (efficiency as care). In hospital settings (Scenarios 1 & 2), stakeholders strongly preferred the Nav2 algorithm. predictability and keeping the lane were cited as primary indicators of *competence*. The social momentum algorithm's yielding behaviour was often described as erratic or inefficient.

The elderly care context (responsiveness as care). In contrast, the elderly care home context showed preference for the social momentum algorithm. Nav2's tendency to *stop and wait* was interpreted not as safety, but as a lack of *reciprocity*. Meanwhile, social momentum's active deviation was perceived as *attentiveness* toward residents with limited cognitive or physical responsiveness.

The emergency exception (scenario 3). Scenario 3 produced a rare convergence: across institutions, stakeholders overwhelmingly preferred the social momentum algorithm. Quantitative metrics showed that in high-criticality contexts, the value of *flow* (efficient clearance of space) superseded the value of *predictability*.

Outcome of deployment

The divergence in stakeholder preference across sites indicated that a single navigation profile was insufficient to address the conflicting normative requirements of the different care environments. Data analysis revealed that attempting to deploy a universal configuration resulted in a system that failed to meet the specific acceptance criteria of either site simultaneously. While the standard scenario comparison mandated by CCVSD Step 5 successfully identified this divergence, the protocol itself provided no mechanism to resolve the conflicting validity claims between the efficiency-driven hospital context and the responsiveness-driven care home context.

3.4.4. Closing the loop

To resolve the contextual instability identified in section 3.4.3, the contextual evaluation protocol was applied to generate specific design guidelines. This step fulfills the synthesis requirement of the CCVSD framework (step 6) by translating the conflicting value assessments into two engineering specifications for the next iteration of the navigation stack. Unlike a universal solution, the analysis necessitates distinct parameter sets for distinct care environments. The following table details the necessary divergence in engineering parameters between the hospital and care home contexts.

Table 3.3: Bifurcated design guidelines for social navigation

Parameter / behavior	Guideline a: hospital context	Guideline b: care home context
Primary value	competence (predictability & flow)	attentiveness (responsiveness & yielding)
Velocity profile	maintain constant velocity; minimize hesitation ($v_{avg} > 0.8$ m/s).	active deceleration in proximity; prioritize stopping over maneuvering ($v \to 0$ as $d \to 0$).
Safety margin	low inflation radius (allow close passing to maintain lane).	high inflation radius (maintain wide berth).
Conflict resolution	lane adherence: robot holds its line and expects the human to adapt.	yielding: robot deviates or stops and assumes the human cannot adapt.

This differentiation confirms that moral skill in robotics is not finding one ethical algorithm, but the ability to contextualize technical constraints to the specific norms of the deployment environment.

3.4.5. Methodological innovation

The discovery of contextual instability required a methodological pivot. A binary pass/fail evaluation proved inadequate because good care is geographically and temporally specific.

To resolve this, the evaluation step was re-architected, see Figure 3.16. This intermediate-level knowledge (ILK) artefact shifts the goal of the integration phase from a static verdict to a generative analysis. It provides a structured workflow to convert qualitative social friction into specific engineering guidelines for the next design cycle.

Structure of the Protocol

As visualised in Figure 3.16, the protocol functions as a translational pipeline:

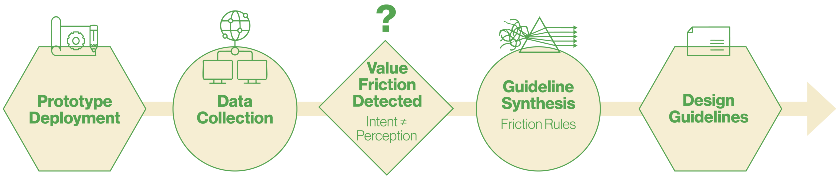


Figure 3.16: The execution workflow for phase three (integration). unlike the static scenario comparison of the standard ccvsd framework, this generative cycle treats value friction as a diagnostic signal. the process synthesizes this friction into concrete, context-specific design guidelines, effectively closing the design loop for the next iteration.

- **Input (prototype deployment):** The cycle begins with the deployment of the working prototype into the validation scenarios (Appendix A). This ensures evaluation is grounded in technical reality rather than hypothetical discussion.
- **Process (friction analysis):** Data is collected using the stakeholder interview protocol (Appendix B). Crucially, the analysis step employs a *contextual filter*, systematically differentiating between hospital logic and home logic. This allows for the detection of alue friction, instances where engineering intent (e.g., safety stops) conflicts with user perception (e.g., social neglect).
- **Output (guideline synthesis):** The final step is the translation of these frictions into scenario-specific design guidelines. Rather than producing a universal parameter set, the protocol generates bifurcated constraints:
  - *Hospital Guideline:* Prioritize high obstacle inflation radii and lane adherence.
  - *Care Home Guideline:* Prioritize low inflation radii and active yielding.

By defining these context-specific guidelines, the protocol closes the loop, providing the explicit design requirements needed to initiate the next exploration phase.

3.5. Field application of the CCVSD framework

Complementing the technical validation of the robot (Sections 3.2 through 3.4), this section reports on the practical application of the methodological artefact itself. The validity of a design framework is assessed through its utility in practice. To this end, the operationalized CCVSD framework was evaluated during an embedded residency at Heemskerk Innovative Technology (HIT). The findings detailed below illustrate the framework's capacity to facilitate cross-disciplinary alignment in an industrial setting.

3.5.1. Qualitative validation

In contrast to the rigorous, controlled protocols applied to the robot simulation, the validation of the methodology was holistic and ethnographic. As a hybrid practitioner embedded within the HIT engineering team, I gathered data through direct participation in the daily engineering workflow. This data was qualitative and derived from three primary streams of interaction: direct observation of coding sprints, capturing informal discourse during ad-hoc discussions, and analysing feedback received during internal presentations.

This intentional lack of formalisation reflects the fluid nature of an interdisciplinary workspace, where the success of a tool is often measured by its active adoption and utility rather than by formal survey metrics. However, it is acknowledged that this approach presents a limitation regarding replicability. Due to the highly situated nature of these social interactions, the specific methodological process is difficult for external researchers to reproduce in different institutional contexts.

3.5.2. The effect of modality change

Data collected during the design sessions revealed a distinct contrast in stakeholder engagement depending on the presentation format. During the initial phase, when ethical mandates were presented as abstract text requirements, the engineering team remained largely silent, treating the constraints as external compliance checkboxes.

However, a marked behavioral shift occurred upon presenting the visual, cyclical structure of the operationalized framework. Once the process was translated into a diagrammatic flow, the engineers began actively pointing to specific nodes and debating the logic. This observed transition from passive silence to active debate indicates that the change in modality had a catalytic effect on interdisciplinary discourse. The engagement was driven by two key translations:

- **Visual, cyclical flow:** The transition from a linear, abstract procedure to a closed-loop, generative cycle resonated with the iterative nature of agile development.

- **Defined output stream:** Formalising the output of each phase, specifically producing design requirements and scenario-specific constraints, aligned closely with the team's existing concepts of minimal viable product (MVP) development and delivery.

This shift provoked continuous and constructive critique from the team, turning bi-weekly meetings from passive reception into active negotiation. The engineers began actively commenting on the requirements, debating trade-offs, and externalising their implicit design choices.

Data collected during the design sessions revealed a distinct contrast in stakeholder engagement depending on the presentation format. During the initial phase, when ethical mandates were presented as abstract text requirements, the engineering team remained largely silent, treating the constraints as external compliance checkboxes.

However, a marked behavioral shift occurred upon presenting the visual, cyclical structure of the operationalized framework. Once the process was translated into a diagrammatic flow, the engineers began actively pointing to specific nodes and debating the logic. This observed transition from passive silence to active debate indicates that the change in modality had a catalytic effect on interdisciplinary discourse. The engagement was driven by two key translations:

3.5.3. Ecological validity and moral skill adoption

The integration of these insights by a commercial entity driven by strict market constraints serves as a robust validation of the methodology's ecological validity. The engineering team received the prospective value hierarchy (PVH) and ethical checkpoints not as bureaucratic hurdles, but as practical tools that provided the necessary technical vocabulary to justify inefficient social behaviours (such as yielding or hesitating) against commercial pressure for optimisation. This process confirmed that the operationalized CCVSD framework is sufficiently robust to function within the high-pressure industrial environment, making ethical mandates compatible components of the industrial agile workflow by translating them into intermediate-level knowledge (ILK). To ensure these translational tools remain accessible beyond the scope of this residency, the validated protocols have been consolidated into a digital handbook (see Appendix E).

4

Discussion

This chapter interprets the findings of the research through the design process, situating the generated intermediate level knowledge artifacts within the broader discourse of value sensitive design. It argues that the implementation gap in care robotics is structurally enforced by the ontological bias of engineering tools and is bridged not by moral persuasion, but by the development of translational artifacts that contextualise ethical lessons for application within the technical workflow.

4.1. The methodological deficit

The primary diagnostic finding of this research challenges the hypothesis that the implementation gap stems from a deficit of moral will. The engineers interviewed in Phase I demonstrated a clear commitment to patient well-being. However, this intent was constrained by a deficit of moral skill: specifically, the lack of methodological artifacts required to translate their care intent into executable practice.

4.1.1. Lessons from the interviews

First, the engineering team demonstrated a clear presence of moral will. Interviews confirmed that user utility and patient well-being were prioritized over technical specification. This contradicts the hypothesis that engineers prioritize efficiency simply due to a lack of ethical concern. Second, the friction identified during the interviews points to a metrication asymmetry. While physical constraints like payload have established measurement standards, relational values like attentiveness do not. This lack of quantitative definition prevents ethical intent from being converted into algorithmic configuration early in the design process. Consequently, the gap identified is best defined as a deficit of moral skill. This does not imply professional incompetence, but rather highlights the absence of intermediate-level tools required to translate abstract norms into concrete specifications.

4.1.2. Prospective value hierarchy as a bridge

The exploration phase identified a semantic gap where identical terms masked conflicting priorities. For example, nursing staff defined reciprocity as a partnership requiring mutual adaptation, whereas the engineering team interpreted the same term as a simple system acknowledgment, such as a button press or LED flash. This divergence indicates that the implementation gap is not just a communication failure but a fundamental difference in how each group defines the problem. Mere dialogue was insufficient because the relational nature of care does not naturally map onto the discrete logic of control systems. To resolve this, the prospective value hierarchy functions as a boundary object. This artifact satisfies the needs of both groups simultaneously by serving as a shared reference point that holds different meanings for different users while maintaining a common structure. The utility of the hierarchy lies in its ability to perform a double translation. A caregiver viewing the hierarchy sees the abstract ethical value of responsiveness at the top, while an engineer viewing the same document sees the concrete technical constraint of kinematic legibility at the bottom. This structure connects the two perspectives, allowing the nurse to validate the ethical intent while enabling the engineer to execute the technical specification. This mechanism permits collaboration without requiring the engineer to become an ethicist or the nurse to master robotics control theory.

4.1.3. Dealing with the abstraction of morality

However, this successful engagement reveals a critical weakness in the framework. It indicates that effective implementation depends entirely on the creation of translational tools across the design cycle that reduce the technical debt of ethics. The difficulty of implementing care values stems from a fundamental disconnection between the descriptive domain of ethics (which describes what ought to be) and the prescriptive domain of engineering (which dictates what must be implemented).

While standard value sensitive design methodologies are good at analysis, decomposing the complex reality of care practice into clear abstract values (Steps 1 and 2), and evaluation (Step 5), they fail at synthesis. This last phase requires the reconstruction of those abstract values into specific and dynamic robot behaviours. The standard linear CCVSD model assumes this reconstruction happens naturally, but this research demonstrates it does not.

Without a robust translation mechanism, such as the development phase protocol generated in this study, the ethical definition is preserved in theory while the resulting behavioural signal is distorted in practice. The development environment (e.g., ROS 2, Gazebo) is not a neutral vehicle; it exerts a normative drag toward efficiency. Therefore, the normative void identified in Chapter 3 is not an empty space, but a space where ethical intent is actively eroded by the ontological bias of the tools.

Consequently, this research argues that moral skill is defined by the ability to manage this synthesis phase. It requires *intermediate level knowledge* (ILK) that does not merely offer high level design guidelines, but provides specific, procedural constraints that force the technical architecture to align with the ethical theory. By lowering the level of abstraction, moving from values to norms and parameters, the operationalized CCVSD framework aims to bridge the descriptive prescriptive divide, transforming ethics from a post hoc critique into a constructive engineering input.

4.1.4. The normative void

Interview data indicates that engineers behavior did not reflect an active dismissal of care, but rather a reversion to established engineering toolsets in the absence of alternatives. This documented phase, designated as the *normative void*, represents the empirical finding of a structural gap where technical logic supersedes care ethics due to a lack of intermediate translation tools.

Hidden weights

The discovery of undocumented constants within the social momentum cost function empirically validates the structural critique leveled against the standard CCVSD framework (section 3.3.6). However, reducing this normative void to a simple discovery of arbitrarily chosen. numbers would trivialize the argument. These hidden weights are merely the final evidence of a much deeper structural failure. This research identifies the normative void not as a single coding error, but as a systemic blind spot that occurs at three distinct levels of development, proving that moral will alone is insufficient to prevent the erosion of care values.

Procedural

First, a procedural void emerges from the ex nihilo fallacy described in section 3.3.1. Because engineers do not build from scratch but assemble existing stacks like Nav2, the development process suffers from path dependency. These tools come with default settings optimized for euclidean efficiency, and in the absence of a distinct normative reflection, the void allows the tool's default logic to become the robot's ethical logic. The robot becomes efficient not because the engineer actively ignores care, but because they lack the methodological tools to override the inherent competence of the underlying architecture.

Kinematic

Second, a kinematic void appears in the gap between capability and legibility. As observed in section 3.3.7 regarding the holonomic sliding issue, raw mathematical optimization favors paths that are physically efficient but socially illegible. The void here represents the lack of a translation layer to inform the solver that technical possibility does not equate to social desirability. Without the active intervention of moral skill to artificially constrain the system, the void is filled by kinematic optimization, resulting in motion that is technically correct but socially rude.

Semantic

Third, a semantic void exists within the definitions of the values themselves. As noted in the analysis of reciprocity in section 3.2.1, abstract ethical terms function as empty signs. Without a good definition provided by an artifact like the prospective value hierarchy, these terms are inevitably filled by engineering superficiality, proven when reciprocity was interpreted merely as a binary input-output acknowledgment rather than a dialog. Consequently, the normative void is not a passive empty space, but a domain where ethical technical debt accumulates whenever a design decision is left to to the default setting of the tool. The solution requires making these trade-offs visible, converting static assumptions into dynamic variables that can be evaluated by the care team rather than fixed by the engineer.

4.1.5. The ontological bias of the tool

Beyond the normative void, a more fundamental barrier was identified within the simulation infrastructure itself, specifically regarding the representation of human agents. As detailed in the technical implementation of the ghost human problem in Appendix C, the standard simulation environment treated human actors as visual artifacts rather than physical entities. While they were visible to the human eye on the screen, they remained invisible to the robot's laser scanner. This research argues that this technical limitation represents more than a mere software bug; it reveals a profound ontological bias embedded within the engineering toolset. The default architecture of the simulation engine is effectively solitary, designed to model a robot interacting with a static, inanimate world rather than a dynamic social environment.

This bias renders the standard toolset structurally unethical for care applications. It is impossible to develop moral skill, specifically the capacity for responsiveness and attentiveness, if the development environment denies the very existence of the social other. The engineering tools implicitly define the world as a collection of static obstacles to be navigated, rather than a shared space populated by reciprocal agents. Consequently, the development of the relay node described in Section 3.3.8 was not merely a coding fix but an ontological correction. By forcing the physics engine to rebroadcast human actors as perceptible semantic entities, the intervention compelled the system to acknowledge the social other as a solid, interactive reality. This confirms that the implementation gap is maintained not only by a lack of methodological guidance but by the stand-alone nature of standard engineering instruments, which actively strip the environment of its social dimension before the design process even begins

4.2. The Generalist trap

Resonating with established discourse in care ethics which posits that care is inherently situated, the integration phase confirmed that good care is not a static property of the robot, but a dynamic property of the context. The phenomenon of contextual instability observed in this study serves as empirical validation of these theoretical principles, challenging the technical feasibility of a universal ethical robot. This challenge strikes at the core of the engineering impulse toward universality, which seeks a single optimal algorithm capable of generalizing across all environments. This impulse assumes that social norms are consistent enough to be captured by a global cost function. However, the comparative performance of the navigation architectures in this study refuted this assumption by revealing a divergence in stakeholder preferences. Hospital staff consistently rated the social momentum algorithm negatively, favoring the standard Nav2 planner, whereas care home residents experienced the social momentum algorithm's active yielding as a necessary sign of attentiveness. In the high-throughput environment of the hospital, the value of competence was defined by deterministic predictability and lane adherence. Conversely, in the residential setting of the care home, that same behavior was interpreted as rude, and competence was redefined as reactive politeness and the granting of space.

This breakdown illustrates the generalist trap. When a system is designed under the premise of universality to satisfy both efficiency metrics and context-dependent social expectations simultaneously, it fails to meet the threshold for acceptance in either. A single global cost function cannot mathematically resolve the conflict between the hospital's requirement for efficient flow and the care home's requirement for protective space. The search for a singular, universal ethical algorithm is therefore a category error; the robotic behavior must be tuned to the singularity of the specific care practice. Consequently, this research suggests that the static scenario comparison of CCVSD step 5 is insufficient for ethical validation because it assumes a stable context against which to test. Because the definition of an ethical act depends on the institutional goal, ethical validation requires an iterative evaluation loop. This necessitates the use of the contextual evaluation protocol to generate scenario-specific constraints, ensuring that the robot's navigation stack is reconfigured to match the normative micro-climate of deployment rather than relying on a generalized default.

4.3. Limitations

4.3.1. Inherent limitation of interviews

The interpretations presented here must be understood within the limitations of the chosen methodology (VSD), particularly regarding the phenomenon of reported ethics versus revealed preference. A critical divergence emerged between the values stakeholders articulated during the exploration phase and the operational choices they validated during the integration phase. While the initial semi-structured interviews were dominated by a discourse of justice and patient autonomy, necessitating a robot that respects the independence of the user, the empirical data from the emergency scenario contradicted this theoretical position. When confronted with the high-pressure reality of the third scenario, stakeholders overwhelmingly preferred the aggressive social momentum algorithm which prioritized system flow over the polite preservation of autonomy. This contradiction indicates that purely sociological methods risk codifying an idealized fantasy of care, whereas the material friction of a physical prototype forces stakeholders to reveal the pragmatic mental models actually in use. Had this research relied exclusively on standard value sensitive design interview protocols, the resulting system would have been engineered to satisfy a moral mandate that users theoretically endorsed but practically rejected. This confirms that moral skill requires more than the collection of verbal value statements; it requires the deployment of technical artifacts to test those values against the constraints of reality, validating the research through design approach as a necessary mechanism to distinguish between what stakeholders say they want and what the clinical practice actually demands.

4.3.2. Uneven technical comparison

A critical technical limitation also emerges from the structural difference between the two tested architectures. The baseline system utilized the full ROS 2 Nav2 stack, a mature industrial framework integrating global spatial awareness via static costmaps with local execution. In contrast, the social momentum framework functioned primarily as a local decision making policy. As defined by Mavrogiannis et al., social momentum is formulated to optimize topological angular momentum in an obstacle free workspace [36], calculating an instantaneous velocity vector rather than a globally consistent path. This created a specific competence gap: while the baseline possessed building awareness, avoiding walls and dead ends via the global planner, the experimental algorithm suffered from global myopia, prioritizing dynamic social cues over static architectural constraints. Consequently, the experimental algorithm struggled to balance its social goals with physical obstacles like walls, which the baseline system handled automatically. Some of the inefficiency observed in the hospital trials therefore stems not from the ethical cost of care, but from this technical inability to reconcile social optimization with the hard constraints of the hospital topology.

4.3.3. Fighting the windmills

Finally, the technical validation was constrained by the limited fidelity of the simulation environment and its lack of ecological validity. Although the HRI infrastructure gap was addressed through the custom relay node, the simulation still relied on preprogrammed human actors that followed fixed trajectories independent of the robot. Because these actors did not respond to the robot in any way, the validation assessed only the robot's ability to respond to human movement and did not capture the reciprocal effects that occur in real interaction.

The actors would continue along their paths even when a collision with the robot was imminent, which removed the possibility of studying shared spatial negotiation. As a result, the simulation functioned as a kinematic stress test rather than a realistic model of social interaction. It excluded the subtle behavioural cues found in human environments and also omitted the basic physical agency of the care receiver. This reduced a relational task to a simple obstacle navigation scenario in which all responsibility for collision avoidance rested solely on the robot.

4.3.4. Limitations on qualitative methods

A significant limitation of this study lies in the replicability of the qualitative validation process. As noted in the analysis of the hybrid practitioner role, the success of the operationalized CCVSD framework relied heavily on the author's embedded status within the engineering team. The data collection was often ad-hoc and opportunistic, capturing informal interactions and spontaneous debates during coding sprints, rather than strictly formalized through rigid protocols.

While this bottom-up approach resulted in high ecological validity by capturing the true, unfiltered reality of the engineering workflow, it comes at the cost of scientific reliability. The specific catalytic effects observed during the design sessions were likely influenced by the unique interpersonal dynamics between the researcher and the specific engineers at HIT. Consequently, the methodological intervention described here should be viewed as a situated case study rather than a universally reproducible algorithm. Future researchers attempting to apply this framework may struggle to replicate the same depth of engagement without possessing the same dual-competence in technical and social domains and without establishing the same level of trust within the target organization.

4.4. The paralysis of the hybrid practitioner

Beyond the technical and methodological frictions reported in Chapter 3, I uncovered a profound *institutional friction* inherent to the interdisciplinary nature of this work. Through an auto-ethnographic reflection on my research process, I identified a critical conflict between the *prescriptive* nature of care ethics and the *positivistic* requirements of the robotics faculty.

4.4.1. The conflict of truth criteria

The discipline of care ethics defines care as an inherently messy, relational, and situated practice that resists standardization [56]. From this perspective, I realized that the sheer complexity of care can be paralyzing, making the reductive assumptions of engineering a pragmatic necessity to enable action [34]. However, I also found that strict adherence to robotic positivist standards creates an inverse paralysis by detaching the system from the reality of the world. Consequently, I argue that behaving in a holistic manner and engaging with the messy real-world landscape is not a deviation from engineering rigor, but the necessary pragmatic solution to the paralysis of positivism.

4.4.2. The institutional void

This paralysis indicated to me that the field of healthcare robotics currently lacks a unified framework for academic evaluation. I found myself functioning as a hybrid practitioner who is currently homeless, appearing too technical for the social sciences yet too qualitative for the hard sciences. As Participant R4 noted, the industry lacks the internal capacity to bridge this gap, and I argue that the academic institution mirrors this deficit. Until robotics faculties expand their epistemological boundaries to accept qualitative friction as valid engineering data, I believe the field will continue to produce sterilized solutions that succeed in the lab through positivistic success but fail in the ward due to holistic failure.

5

Conclusion

5.1. Summary of Research

This thesis set out to resolve the implementation gap in care robotics, the persistent disconnect between the high level ethical principles of care and the pragmatic reality of engineering development. Initially, the field hypothesized that this gap resulted from a deficit of moral will, a presumption that engineers prioritized efficiency due to a lack of ethical concern.

Through an interdisciplinary research through design approach, operationalized via the care centered value sensitive design framework, this study challenged that binary. By embedding a hybrid practitioner within a commercial engineering team and developing a social navigation module for the PAL Tiago robot, the research demonstrated that the failure to integrate care values is not attitudinal, but methodological. The study concludes that the field faces a deficit of moral skill: a lack of intermediate level knowledge artifacts capable of translating abstract normative values into concrete technical specifications.

5.2. Key findings and answers to research questions

The primary research question asked: How can the care centered value sensitive design framework be adapted and operationalized to provide actionable, engineer centric guidance for the practical development of care robots?

This research answers that question by restructuring the framework from a linear assessment model into an operationalized generative design cycle. Based on the embedded case study at HIT, the utility of this adaptation is evidenced by its capacity to structure the team's tacit knowledge into a formalized workflow. The framework proved useful in this specific industrial context by providing the necessary scaffolding for engineers to articulate and integrate ethical theory into their daily practice. This operational utility is supported by three key findings corresponding to the phases of the design loop:

- **Bridging the semantic gap (exploration phase):** Analysis of the participating stakeholders identified a semantic divergence where identical terms masked conflicting priorities. For instance, in this specific context, reciprocity was interpreted by nurses as a dialogical partnership but by engineers as a binary system acknowledgment. The study suggests that abstract ethical mandates resisted direct technical implementation; in this case, they required translation via the prospective value hierarchy to convert values into programmable design requirements.
- **Addressing the normative void (development phase):** The results demonstrated that the standard framework fails to provide actionable, engineer-centric guidance because it structurally skips the actual implementation process, creating a normative void. It structurally skips the actual implementation process, creating a normative void. Additionally, the investigation revealed that practical development was constrained by standard tools, which in this context possessed an ontological bias against social representation. Additionally, the research revealed that practical development was constrained by standard tools, which in this context possessed an ontological bias against social representation; where human agents are defaulted to visual artifacts rather than detectable social entities. The answer requires actively identifying this void and mitigating the bias through technical intervention or make social agents visible, thereby rendering practical ethical development feasible.
- **Avoiding the generalist trap** (integration phase) Empirical validation revealed the phenomenon of contextual instability, where identical robotic behaviors created opposing value judgments across different care

settings. The data demonstrates that a static navigation stack, even one optimized for social metrics, will fail in at least one operational domain because the definition of competence inverts dependant on its context. For example, behavior considered safe in a hospital was viewed as unresponsive in an elderly home. Therefore, future architectures must be context aware at the planner level, capable of swapping behavioral cost functions based on the context of the care practice. Consequently, operationalization requires an iterative evaluation loop that generates scenario specific guidelines.

5.3. Contribution to knowledge

This thesis makes distinct contributions to the fields of engineering ethics and social robotics:

- Methodological contribution (The intermediate level knowledge artifacts) The primary contribution is the operationalized framework itself. By generating the perspective value hierarchy and the contextual evaluation protocol, this research provides the design handbook necessary to bridge the gap between the descriptive domain of ethics and the prescriptive domain of engineering.
- Theoretical contribution (moral skill) The study reframes the discourse on responsible robotics by shifting the scope of intervention from the engineer's mindset, or moral will, to the engineer's toolset, or moral skill. It empirically demonstrates that ethical intent is insufficient without translational tools.
- Technical contribution (the HRI infrastructure gap) The research identified and partially resolved the HRI infrastructure gap, the inability of standard simulation tools to model relational reciprocity. The development of the relay node and social momentum integration serves as a technical proof of concept for how to align software architecture with relational ethics.

5.4. Overall Conclusion

The effective integration of care ethics into robotics is not a task of persuasion, but of translation. This thesis concludes that the implementation gap is not a result of lack of moral will, but is reinforced by a deficit of moral skill, specifically, the lack of translational tools required to convert abstract ethical values into concrete engineering specifications. To overcome this, this study operationalized the transition regarding ethics from a post-hoc evaluation to a continuous design input by generating and integrating existing intermediate-level knowledge artifacts into a generative design cycle. This operationalized design cycle serves as a foundational step toward bridging this gap, offering a preliminary structure to turn the abstract will to care into the concrete skill to implement. This contribution lays the groundwork to ensure that the robots of the future operate not only with the will to care, but with the skill to execute it.

6
Recommendations

6.1. Recommendations for future research

6.1.1. Bridging the infrastructure gap

The research identified a critical infrastructure gap regarding reciprocal human robot interaction. While this study successfully made humans visible to the robot using a relay node, the simulation suffered from low ecological validity because the human agents acted as ghosts following preset paths without reacting to the robot. This limitation made it impossible to test true reciprocity or negotiation.

Therefore, future research in CCVSD could move beyond static crowd simulation. While recent open source developments like *HuNavSim* [42], *Arena-Rosnav* [25], and *SEAN 2.0* [57] have successfully introduced reactive agents into ROS 2 workflows, they predominantly rely on generic Social Force Models (SFM)[21] focused on collision avoidance (proxemics) rather than social negotiation. A critical gap remains: the ability to configure distinct social roles. Current tools generally treat humans as homogenous dynamic obstacles, lacking the semantics to distinguish between a rushing nurse (who expects priority) and a curious resident (who may yield).

To address this, the field does not need a new simulator, but rather a social negotiation layer compatible with standard ROS 2 stacks. The proposed solution is to develop a library of smart actor plugins extending frameworks like HuNavSim that possess configurable yielding logic. These actors must offer configurable social profiles (e.g., *yielding_probability*, *role_priority*) to test how robots handle different social contracts. This development is a necessary technical prerequisite for advancing VSD in practice, transforming simulation from a kinematic stress test into a relational sandbox.

6.1.2. Moving away from the generalist trap

Future research could move away from trying to create one universal navigation style. This study suggests that a single set of rules struggles because different environments have different expectations. Behavior that works well in a busy hospital might be seen as rude in a quiet care home. Future engineers could solve this by designing systems that automatically switch settings depending on where the robot is operating.

For example, a hospital mode could focus on efficiency. In this setting, the robot would drive with confidence and expect staff to move around it. However, a care home mode would need to prioritize responsiveness. Here, the robot would slow down earlier and give residents much more space, assuming that they might not be able to step aside. Developing this ability to adapt to the local context would be a valuable next step.

6.1.3. Co-creation tool with adjustable parameters

The analysis in section 3.3.6 showed that the decision making process was hidden inside the code. To address this, future technical iterations could explore the development of a user interface that exposes these variables to care staff. Instead of relying on engineers to lock in these preferences, such an interface could use dynamic reconfiguration to make the parameters adjustable while the system is running. A possible implementation might feature a simple slider balancing concepts such as urgency versus comfort, which could inversely adjust the weights for efficiency and social momentum. This approach offers the possibility of turning robot calibration into a collaborative process, allowing caregivers to see how their choices affect the behavior of the robot in real time. By shifting these decisions from the developer to the practitioner, the system might become more adaptable to the specific needs of the moment.

6.1.4. Social norms infused navigation

Consistent feedback from the interviews suggests that future projects could explore integrating specific social norms into the navigation algorithm. While the social momentum framework addressed the need for clear movement by making the robot's intent legible, it focused more on physical optimization than on local culture. Future research could investigate the unwritten traffic rules of the care environment, such as which side of the hall to use or who should yield to whom. Translating these observations into mathematical constraints within the cost function would allow the robot to value social rules as much as it values avoiding collisions. By encoding these norms directly into the planner, the system could move beyond simple obstacle avoidance and navigate the hospital in a way that feels more socially polite to the staff and residents.

6.2. Recommendations for practice and policy

To address the ex nihilo fallacy identified in the development phase, engineering management and policy bodies could recommend a normative infrastructure audit at the start of care robotics projects. The empirical findings of this study suggest that the normative void is often populated by the default settings of open source tools rather than being a neutral space. For instance, the analysis revealed hidden weights within the social momentum code that effectively determined the ethical behavior of the robot without active consent from the engineer. This indicates that technical teams often inherit technical debt in the form of ethical biases embedded within libraries, such as the default preference for euclidean efficiency found in standard navigation stacks.

Consequently, before writing new code, the team could check their dependencies to reveal these implicit values. This process might involve identifying all cost functions to determine if weights are exposed as configurable parameters or hidden as hard coded constants. Such a check could examine the ontological bias of the simulator to ensure it models humans as reactive agents rather than static objects. Finally, the team could explicitly define the default behavior for failure states, determining whether the system prioritizes safety or efficiency when sensors fail. This approach would serve to convert implicit assumptions into explicit design choices, ensuring that efficiency is a deliberate decision rather than an accidental default.

6.3. Concluding remarks on recommendations

The recommendations outlined in this chapter are not intended as a definitive solution to the ethical problems of care robotics, but rather as a corrective to the ex nihilo fallacy. By moving from abstract ethical debates to the concrete development of negotiation layers and configurable parameters, these proposals operationalize the central argument of this thesis.

Ultimately, these technical and policy interventions serve a single methodological purpose: to transform ethical intent into engineering specification. They ensure that the moral skill identified in this study does not remain a theoretical ideal, but becomes an executable function of the robot's architecture.

7

Epilogue

This epilogue is a personal note to a personal thesis. It is a warranty of my own warmth, as a master student, as a human, and as a proud sailor of the ship this journey has been. This thesis and her topic have been an inherent reflection of the messy reality that caring for is. Because in the taking care of this thesis I have been taken care of so much. Just as the interconnectivity between topic and theme, between robot and being , between taking care of and being taken care of, I stand in the middle. I have called it the hybrid practitioner, the civil engineer of academic fields, the bridging architect of two worlds so strongly needed to be connected. The size of the crevasse to be superseded inversely related to strength of their own seating. Both fields so foundationally found in a history of knowledge yet so far apart in their expertise to reach out.

A hand tries to clutch the air between them. It gasps in the void, but nonetheless, a hand reaches.

I have learned that to ask questions is to answer them, and to answer them is to question them, and so we spiral in an endless fingerpointing battle bound to be lost by both parties. I found my right hand pointing at my left, while my left grasps behind my back and grabs my right. The hybrid practitioner, bound to break his own barriers. For to truly immerse myself in this thesis, I had to question myself, at every turn I had to churn my own perceptions and more importantly, my neglections, what had I left right there to be interpreted.

Most of all, this project has taught me a lot about taking care for myself. Amidst interviews and interventions I found myself not only doubting my theory, but doubting myself. The questions on paper became questions of the mind. This divide, the pragmaticality of an ethical reality, is the inherent presence of a most personal dilemma. How do I reconcile the polarity that lives within, and to try and formulate an answer in a conclusive academic chapter is to romp my own rivers. Through findings and forensic analysis I hoist the sails of an odyssey of thought and song, of self love and sometimes of its begone.

Because to take care, is to take. To sit with, to hold space, to give room and take it away sometimes. To give and to forgive, to forgo what you have sometimes told yourself. To take care is to heart, to veins, to pump and to live.

How this thesis has lived.

References

[1] Michael Abadir et al. “Navigating the future of artificial intelligence technologies for improving the care of older adults”. en. In: *Innovation in Aging* 9.Supplement_1 (Dec. 2025). Ed. by Michelle Putnam, S24–S32. ISSN: 2399-5300. DOI: `10.1093/geroni/igaf092`. URL: `https://academic.oup.com/innovateage/article/9/Supplement_1/S24/8263970` (visited on 01/15/2026).

[2] Philip Agre. *Computation and human experience*. Cambridge University Press, 1997.

[3] Michael Anderson and Susan Leigh Anderson. “GenEth: A General Ethical Dilemma Analyzer”. en. In: ().

[4] Mathieu Boniol et al. “The global health workforce stock and distribution in 2020 and 2030: a threat to equity and ‘universal’ health coverage?” en. In: *BMJ Global Health* 7.6 (June 2022), e009316. ISSN: 2059-7908. DOI: `10.1136/bmjgh-2022-009316`. URL: `https://gh.bmj.com/lookup/doi/10.1136/bmjgh-2022-009316` (visited on 01/15/2026).

[5] Andriana Boudouraki et al. “Responsible and adaptive robots in care home settings: an implementation framework analysis of a workshop with public and professionals”. en. In: *Frontiers in Robotics and AI* 12 (July 2025), p. 1610329. ISSN: 2296-9144. DOI: `10.3389/frobt.2025.1610329`. URL: `https://www.frontiersin.org/articles/10.3389/frobt.2025.1610329/full` (visited on 01/15/2026).

[6] Martim Brandao. “Normative roboticists: the visions and values of technical robotics papers”. en. In: *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. Vancouver, BC, Canada: IEEE, Aug. 2021, pp. 671–677. ISBN: 978-1-6654-0492-1. DOI: `10.1109/RO-MAN50785.2021.9515504`. URL: `https://ieeexplore.ieee.org/document/9515504/` (visited on 01/16/2026).

[7] Michael Davis. ““Ain’t No One Here But Us Social Forces”: Constructing the Professional Responsibility of Engineers”. en. In: *Science and Engineering Ethics* 18.1 (Mar. 2012), pp. 13–34. ISSN: 1353-3452, 1471-5546. DOI: `10.1007/s11948-010-9225-3`. URL: `http://link.springer.com/10.1007/s11948-010-9225-3` (visited on 01/16/2026).

[8] Gilles Deleuze and Gilles Deleuze. *Difference and repetition*. en. New York: Columbia Univ. Press, 1994. ISBN: 978-0-231-08158-0 978-0-231-08159-7.

[9] Virginia Dignum. *Responsible Autonomy*. arXiv:1706.02513 [cs]. June 2017. DOI: `10.48550/arXiv.1706.02513`. URL: `http://arxiv.org/abs/1706.02513` (visited on 01/28/2025).

[10] Connor Esterwood and Lionel Peter Robert. “The Warehouse Robot Interaction Sim: An Open-Source HRI Research Platform”. en. In: *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. Stockholm Sweden: ACM, Mar. 2023, pp. 268–271. ISBN: 978-1-4503-9970-8. DOI: `10.1145/3568294.3580086`. URL: `https://dl.acm.org/doi/10.1145/3568294.3580086` (visited on 01/10/2026).

[11] Berenice Fisher and Joan Tronto. “Toward a Feminist Theory of Caring”. In: *Circles of Care: Work and Identity in Women’s Lives*. Ed. by Emily K. Abel and Margaret K. Nelson. Albany: State University of New York Press, 1990, pp. 35–62.

[12] Tamar Flash and Neville Hogans. “The Coordination of Arm Movements: An Experimentally Confirmed Mathematical”. en. In: ().

[13] Anthony Francis et al. *Principles and Guidelines for Evaluating Social Robot Navigation Algorithms*. en. arXiv:2306.16740 [cs]. Sept. 2023. DOI: `10.48550/arXiv.2306.16740`. URL: `http://arxiv.org/abs/2306.16740` (visited on 12/27/2025).

[14] Batya Friedman and David G. Hendry. *Value Sensitive Design: Shaping Technology with Moral Imagination*. en. Google-Books-ID: 8ZiWDwAAQBAJ. MIT Press, May 2019. ISBN: 978-0-262-03953-6.

[15] Yuxiang Gao and Chien-Ming Huang. “Evaluation of Socially-Aware Robot Navigation”. en. In: *Frontiers in Robotics and AI* 8 (Jan. 2022), p. 721317. ISSN: 2296-9144. DOI: `10.3389/frobt.2021.721317`. URL: `https://www.frontiersin.org/articles/10.3389/frobt.2021.721317/full` (visited on 12/27/2025).

[16] Wilbert M. Gesler. “Therapeutic landscapes: Medical issues in light of the new cultural geography”. en. In: *Social Science & Medicine* 34.7 (Apr. 1992), pp. 735–746. ISSN: 02779536. DOI: `10.1016/0277-9536(92)90360-3`. URL: `https://linkinghub.elsevier.com/retrieve/pii/0277953692903603` (visited on 01/16/2026).

[17] Joan W Goldman. “The Hidden Dimension” by Edward T. Hall (Book Review)”. In: *Et Cetera* 26 (1969), p. 241.

[18] Greg Guest, Arwen Bunce, and Laura Johnson. “How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability”. en. In: *Field Methods* 18.1 (Feb. 2006), pp. 59–82. ISSN: 1525-822X, 1552-3969. DOI: `10.1177/1525822X05279903`. URL: `https://journals.sagepub.com/doi/10.1177/1525822X05279903` (visited on 01/11/2026).

[19] John C. Hayward. “Near-miss determination through use of a scale of danger”. In: *Highway Research Record* 384 (1972), pp. 24–34.

[20] Heemskerk Innovative Technology B.V. *Heemskerk Innovative Technology · Bridging the gap between science and industry*. Accessed: 2025-12-29. 2025. URL: `https://heemskerk-innovative.nl/`.

[21] Dirk Helbing and Peter Molnar. “Social Force Model for Pedestrian Dynamics”. en. In: *Physical Review E* 51.5 (May 1995). arXiv:cond-mat/9805244, pp. 4282–4286. ISSN: 1063-651X, 1095-3787. DOI: `10.1103/PhysRevE.51.4282`. URL: `http://arxiv.org/abs/cond-mat/9805244` (visited on 12/23/2025).

[22] Kristina Höök and Jonas Löwgren. “Strong concepts: Intermediate-level knowledge in interaction design research”. In: *ACM Trans. Comput.-Hum. Interact.* 19.3 (Oct. 2012), 23:1–23:18. ISSN: 1073-0516. DOI: `10.1145/2362364.2362371`. URL: `https://doi.org/10.1145/2362364.2362371` (visited on 12/29/2025).

[23] Markus Huber et al. “Human-robot interaction in handing-over tasks”. en. In: *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*. Munich, Germany: IEEE, Aug. 2008, pp. 107–112. ISBN: 978-1-4244-2212-8. DOI: `10.1109/ROMAN.2008.4600651`. URL: `http://ieeexplore.ieee.org/document/4600651/` (visited on 12/24/2025).

[24] David Janssen. “Mobile Robot Socially Aware Navigation:” en. In: ().

[25] Linh Kästner et al. *Arena-Rosnav 2.0: A Development and Benchmarking Platform for Robot Navigation in Highly Dynamic Environments*. en. arXiv:2302.10023 [cs]. July 2023. DOI: `10.48550/arXiv.2302.10023`. URL: `http://arxiv.org/abs/2302.10023` (visited on 01/12/2026).

[26] Billy Vaughn Koen. *Discussion of the method: conducting the engineer’s approach to problem solving*. en. New York: Oxford University Press, 2003. ISBN: 978-0-19-515599-0.

[27] Ilpo Koskinen et al. *Design Research Through Practice: From the Lab, Field, and Showroom*. Morgan Kaufmann, 2012. ISBN: 978-0-12-385502-2. DOI: `10.1016/C2010-0-65896-2`.

[28] Madelaine Ley. “Care Ethics and the Future of Work: a Different Voice”. en. In: *Philosophy & Technology* 36.1 (Jan. 2023), p. 7. ISSN: 2210-5441. DOI: `10.1007/s13347-022-00604-5`. URL: `https://doi.org/10.1007/s13347-022-00604-5` (visited on 06/17/2024).

[29] Jenny X. Liu et al. “Global Health Workforce Labor Market Projections for 2030”. en. In: *Human Resources for Health* 15.1 (Dec. 2017), p. 11. ISSN: 1478-4491. DOI: `10.1186/s12960-017-0187-2`. URL: `https://human-resources-health.biomedcentral.com/articles/10.1186/s12960-017-0187-2` (visited on 01/15/2026).

[30] Jonas Löwgren. “Annotated portfolios and other forms of intermediate-level knowledge”. en. In: *Interactions* 20.1 (Jan. 2013), pp. 30–34. ISSN: 1072-5520, 1558-3449. DOI: `10.1145/2405716.2405725`. URL: `https://dl.acm.org/doi/10.1145/2405716.2405725` (visited on 01/16/2026).

[31] William T Lynch and Ronald Kline. “Engineering Practice and Engineering Ethics”. en. In: ().

[32] Steve Macenski et al. “The Marathon 2: A Navigation System”. en. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. arXiv:2003.00368 [cs]. Oct. 2020, pp. 2718–2725. DOI: `10.1109/IROS45743.2020.9341207`. URL: `http://arxiv.org/abs/2003.00368` (visited on 01/17/2026).

[33] Kirsti Malterud, Volkert Dirk Siersma, and Ann Dorrit Guassora. “Sample Size in Qualitative Interview Studies: Guided by Information Power”. en. In: *Qualitative Health Research* 26.13 (Nov. 2016), pp. 1753–1760. ISSN: 1049-7323, 1552-7557. DOI: `10.1177/1049732315617444`. URL: `https://journals.sagepub.com/doi/10.1177/1049732315617444` (visited on 01/11/2026).

[34] Joaquim R. R. A. Martins and Andrew Ning. *Engineering Design Optimization*. en. 1st ed. Cambridge University Press, Nov. 2021. ISBN: 978-1-108-98064-7 978-1-108-83341-7. DOI: `10.1017/9781108980647`. URL: `https://www.cambridge.org/core/product/identifier/9781108980647/type/book` (visited on 01/16/2026).

[35] Christoforos Mavrogiannis et al. “Core Challenges of Social Robot Navigation: A Survey”. en. In: *ACM Transactions on Human-Robot Interaction* 12.3 (Sept. 2023), pp. 1–39. ISSN: 2573-9522, 2573-9522. DOI: `10.1145/3583741`. URL: `https://dl.acm.org/doi/10.1145/3583741` (visited on 01/10/2025).

[36] Christoforos Mavrogiannis et al. “Social Momentum: Design and Evaluation of a Framework for Socially Competent Robot Navigation”. en. In: *ACM Transactions on Human-Robot Interaction* 11.2 (June 2022), pp. 1–37. ISSN: 2573-9522, 2573-9522. DOI: `10.1145/3495244`. URL: `https://dl.acm.org/doi/10.1145/3495244` (visited on 12/23/2025).

[37] Diane P Michelfelder, Natasha McCarthy, and David E. Goldberg, eds. *Philosophy and Engineering: Reflections on Practice, Principles and Process*. en. Vol. 15. Philosophy of Engineering and Technology. Dordrecht: Springer Netherlands, 2013. ISBN: 978-94-007-7761-3 978-94-007-7762-0. DOI: `10.1007/978-94-007-7762-0`. URL: `https://link.springer.com/10.1007/978-94-007-7762-0` (visited on 12/24/2025).

[38] Annemarie Mol. *The Logic of Care: Health and the Problem of Patient Choice*. London: Routledge, 2008. ISBN: 978-0-415-45343-1.

[39] Jessica Morley et al. “From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices”. en. In: *Science and Engineering Ethics* 26.4 (Aug. 2020), pp. 2141–2168. ISSN: 1353-3452, 1471-5546. DOI: `10.1007/s11948-019-00165-5`. URL: `http://link.springer.com/10.1007/s11948-019-00165-5` (visited on 01/16/2026).

[40] Billy Okal and Kai O. Arras. “Learning socially normative robot navigation behaviors with Bayesian inverse reinforcement learning”. en. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. Stockholm, Sweden: IEEE, May 2016, pp. 2889–2895. ISBN: 978-1-4673-8026-3. DOI: `10.1109/ICRA.2016.7487452`. URL: `http://ieeexplore.ieee.org/document/7487452/` (visited on 01/10/2026).

[41] Irena Papadopoulos et al. “Enablers and barriers to the implementation of socially assistive humanoid robots in health and social care: a systematic review”. en. In: *BMJ Open* 10.1 (Jan. 2020), e033096. ISSN: 2044-6055, 2044-6055. DOI: `10.1136/bmjopen-2019-033096`. URL: `https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2019-033096` (visited on 01/15/2026).

[42] Noé Pérez-Higueras et al. *HuNavSim: A ROS 2 Human Navigation Simulator for Benchmarking Human-Aware Robot Navigation*. en. arXiv:2305.01303 [cs]. Sept. 2023. DOI: `10.48550/arXiv.2305.01303`. URL: `http://arxiv.org/abs/2305.01303` (visited on 01/12/2026).

[43] Ibo van de Poel. “Translating Values into Design Requirements”. In: *Philosophy and Engineering: Reflections on Practice, Principles and Process*. Ed. by Diane P. Michelfelder, Natasha McCarthy, and David E. Goldberg. Vol. 15. Philosophy of Engineering and Technology. Dordrecht: Springer Science+Business Media, 2013, pp. 253–266. ISBN: 978-94-007-7761-3.

[44] Ibo van de Poel and Lambèr Royakkers. “Ethics, Technology, and Engineering”. en. In: ().

[45] Jeannette Pols. *Care at a Distance: On the Closeness of Technology*. Amsterdam University Press, 2012.

[46] Wilbert Pumacay et al. *THE COLOSSEUM: A Benchmark for Evaluating Generalization for Robotic Manipulation*. en. arXiv:2402.08191 [cs]. May 2024. DOI: `10.48550/arXiv.2402.08191`. URL: `http://arxiv.org/abs/2402.08191` (visited on 01/10/2026).

[47] J. Rios-Martinez, A. Spalanzani, and C. Laugier. "From Proxemics Theory to Socially-Aware Navigation: A Survey". en. In: *International Journal of Social Robotics* 7.2 (Apr. 2015), pp. 137–153. ISSN: 1875-4791, 1875-4805. DOI: `10.1007/s12369-014-0251-1`. URL: `http://link.springer.com/10.1007/s12369-014-0251-1` (visited on 12/24/2025).

[48] Bartosz Sawik et al. "Robots for Elderly Care: Review, Multi-Criteria Optimization Model and Qualitative Case Study". en. In: *Healthcare* 11.9 (Apr. 2023), p. 1286. ISSN: 2227-9032. DOI: `10.3390/healthcare11091286`. URL: `https://www.mdpi.com/2227-9032/11/9/1286` (visited on 01/15/2026).

[49] Donald A. Schön. *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books, 1983. ISBN: 978-0465068784.

[50] Roland Siegwart, Illah R. Nourbakhsh, and Davide Scaramuzza. *Introduction to Autonomous Mobile Robots*. MIT Press, 2011.

[51] Anna Ståhl, Jonas Löwgren, and Kristina Höök. "Designing for Interactional Empowerment". en. In: ().

[52] Irene van Staveren. *The Ethics of Efficiency*. Tech. rep. 18. SCEME Working Papers, 2007.

[53] Janice Gross Stein. *The cult of efficiency*. eng. Rev. ed. The CBC Massey lectures series. Toronto: Anansi, 2002. ISBN: 978-0-88784-678-6.

[54] Stefan Timmermans and Marc Berg. *The Gold Standard: The Challenge Of Evidence-Based Medicine*. en. Philadelphia: Temple University Press, 2010. ISBN: 978-1-59213-187-7.

[55] Joan Tronto. *Moral Boundaries: A Political Argument for an Ethic of Care*. New York: Routledge, July 2020. ISBN: 978-1-003-07067-2. DOI: `10.4324/9781003070672`.

[56] Joan C. Tronto. *Moral boundaries: a political argument for an ethic of care*. en. Repr. New York: Routledge, 2009. ISBN: 978-0-415-90642-5 978-0-415-90641-8.

[57] Nathan Tsoi et al. "SEAN 2.0: Formalizing and Generating Social Situations for Robot Navigation". en. In: *IEEE Robotics and Automation Letters* 7.4 (Oct. 2022), pp. 11047–11054. ISSN: 2377-3766, 2377-3774. DOI: `10.1109/LRA.2022.3196783`. URL: `https://ieeexplore.ieee.org/document/9851501/` (visited on 01/12/2026).

[58] Karl T. Ulrich. *Design: creation of artifacts in society*. en. OCLC: 940507606. Philadelphia, PA: University of Pennsylvania, 2011. ISBN: 978-0-9836487-0-3.

[59] Steven Umbrello and Olivia Gambelin. "Agile as a Vehicle for Values: A Value Sensitive Design Toolkit". en. In: *Rethinking Technology and Engineering*. Ed. by Albrecht Fritzsche and Andrés Santa-María. Vol. 45. Series Title: Philosophy of Engineering and Technology. Cham: Springer International Publishing, 2023, pp. 169–181. ISBN: 978-3-031-25232-7 978-3-031-25233-4. DOI: `10.1007/978-3-031-25233-4_13`. URL: `https://link.springer.com/10.1007/978-3-031-25233-4_13` (visited on 12/24/2025).

[60] Annemiek Van Boeijen et al. *Delft design guide: Design strategies and methods*. 2014.

[61] A. Van Wynsberghe. "Service robots, care ethics, and design". en. In: *Ethics and Information Technology* 18.4 (Dec. 2016), pp. 311–321. ISSN: 1388-1957, 1572-8439. DOI: `10.1007/s10676-016-9409-x`. URL: `http://link.springer.com/10.1007/s10676-016-9409-x` (visited on 12/04/2025).

[62] Aimee Van Wynsberghe. "Designing Robots for Care: Care Centered Value-Sensitive Design". en. In: *Science and Engineering Ethics* 19.2 (June 2013), pp. 407–433. ISSN: 1353-3452, 1471-5546. DOI: `10.1007/s11948-011-9343-6`. URL: `http://link.springer.com/10.1007/s11948-011-9343-6` (visited on 12/04/2025).

[63] Tijs Vandemeulebroucke, Bernadette Dierckx De Casterlé, and Chris Gastmans. "The use of care robots in aged care: A systematic review of argument-based ethics literature". en. In: *Archives of Gerontology and Geriatrics* 74 (Jan. 2018), pp. 15–25. ISSN: 01674943. DOI: `10.1016/j.archger.2017.08.014`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0167494317302790` (visited on 01/16/2026).

[64] Peter-Paul Verbeek. “Materializing Morality: Design Ethics and Technological Mediation”. en. In: *Science, Technology, & Human Values* 31.3 (May 2006), pp. 361–380. ISSN: 0162-2439, 1552-8251. DOI: `10.1177/0162243905285847`. URL: `http://journals.sagepub.com/doi/10.1177/0162243905285847` (visited on 06/25/2024).

[65] Alan F. T. Winfield and Marina Jirotka. “Ethical governance is essential to building trust in robotics and artificial intelligence systems”. In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 376.2133 (Nov. 2018), p. 20180085. ISSN: 1364-503X. DOI: `10.1098/rsta.2018.0085`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6191667/` (visited on 01/28/2025).

[66] Aimee van Wynsberghe, Madelaine Ley, and Sabine Roeser. “Ethical Aspects of Human–Robot Collaboration in Industrial Work Settings”. en. In: *The 21st Century Industrial Robot: When Tools Become Collaborators*. Ed. by Maria Isabel Aldinhas Ferreira and Sarah R. Fletcher. Cham: Springer International Publishing, 2022, pp. 255–266. ISBN: 978-3-030-78513-0. DOI: `10.1007/978-3-030-78513-0_14`. URL: `https://doi.org/10.1007/978-3-030-78513-0_14` (visited on 06/17/2024).

[67] Guang-Zhong Yang et al. “The grand challenges of *Science Robotics*”. en. In: *Science Robotics* 3.14 (Jan. 2018), eaar7650. ISSN: 2470-9476. DOI: `10.1126/scirobotics.aar7650`. URL: `https://www.science.org/doi/10.1126/scirobotics.aar7650` (visited on 11/25/2024).

[68] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. *Research through design as a method for interaction design research in HCI*. Pages: 502. Apr. 2007. DOI: `10.1145/1240624.1240704`.

A

Technical Details and Validation Protocol

This appendix delineates the technical specifications, mathematical formulations, and quantitative criteria employed during the constructive case study involving the Social Mobile Manipulation task on the PAL Tiago platform. These protocols were established to ensure methodological rigor during Steps 4 and 5 of the Care-Centered Value Sensitive Design (CCVSD) framework, specifically aiming to translate abstract ethical values into measurable engineering constraints.

A.1. Validation Metrics and Key Performance Indicators (KPIs)

The evaluation of robotic behavior in CCVSD Step 5 (Scenario Comparison) necessitated a departure from binary collision metrics. To capture the ethical trade-offs inherent in the navigation algorithms, subjective stakeholder assessments were systematically correlated with objective, multi-faceted quantitative measures.

A.1.1. Performance and Efficiency Metrics

These metrics quantify the functional efficacy of the navigation algorithm, independent of social context:

- **Task Efficiency (Time to Goal):** This metric records the total elapsed time from the initialization of the scenario to the successful attainment of the goal state.
- **Path Length Deviation:** This serves as a proxy for optimality, defined as the ratio of the integrated path length traversed by the robot to the theoretical shortest Euclidean path.

A.1.2. Social Compliance and Care Metrics

These metrics quantify the relational quality of the robot's motion, providing a direct mapping to the care values identified in CCVSD Steps 1 and 2:

- **Navigational Smoothness (Jerk Profile):** Defined as the third derivative of position ($\frac{d^3x}{dt^3}$), this metric quantifies changes in acceleration. It is utilized to detect "hesitant," erratic, or oscillation-prone movement profiles. A minimized Jerk Profile is positively correlated with the values of *Attentiveness* and *Trust*.
- **Personal Space Intrusion (PSI) Score:** An integrated measure comprising both the duration and magnitude of violations regarding a defined social distance threshold (typically $1.2\,\mathrm{m}$) relative to human actors. Minimizing the PSI Score correlates with the preservation of *Safety* and *Dignity*.
- **Deadlock Frequency:** This counts the occurrences where the navigation planner enters a local minimum or sustained "freeze" state, necessitating recovery behaviors or failing to proceed. A high Deadlock Frequency acts as a contra-indicator for *Competence*.
- **Group Violation Flags:** A binary classification metric flagging instances where the robot contravenes emergent social norms, such as opposing the collective flow of pedestrian traffic (e.g., lane adherence violations).

A.2. Mathematical Formulation of Key Performance Indicators

To facilitate reproducibility and provide a rigorous basis for the metrics outlined in Section A.1, the specific computational definitions and physical formulations are detailed below.

A.2.1. The Anisotropic Social Safety Boundary

To account for the directional nature of human perception and safety, the effective minimum distance, denoted as d_{safe}, is calculated relative to the perimeter of a safety ellipse aligned with the agent's heading θ. This formulation moves beyond isotropic circular buffers to model the "anisotropic" nature of human personal space:

$$d_{safe} = \sqrt{(x_r - x_h)^2 + (y_r - y_h)^2} - R(\phi) \tag{A.1}$$

Where (x_r, y_r) and (x_h, y_h) represent the coordinates of the robot and human respectively. $R(\phi)$ denotes the radius of the ellipse at the angle of approach ϕ, parameterized by a semi-major axis $a = 0.6$ m (longitudinal safety) and a semi-minor axis $b = 0.3$ m (lateral safety).

A.2.2. Discrete Jerk Profile Calculation

Navigational smoothness is assessed via the jerk $j(t)$, calculated as the discrete derivative of acceleration. To mitigate quantization noise inherent in the 20 Hz simulation update loop, a moving average filter with a window size of $w = 10$ is applied:

$$j(t) = \frac{1}{w}\sum_{k=0}^{w-1}\left|\frac{a(t-k) - a(t-k-1)}{\Delta t}\right| \tag{A.2}$$

Here, $a(t)$ represents the instantaneous acceleration derived from the odometry stream. A threshold of $j(t) < 5.0\,\mathrm{m/s^3}$ is utilized as the upper bound for motion perceived as comfortable by human observers.

A.2.3. Social Reactivity via Linear Risk Regression

The "Politeness" of a navigation policy is operationalized by analyzing the correlation between the robot's velocity and its proximity to humans. This is determined by the slope m of a linear regression fitted between the robot's linear velocity (v_r) and the Euclidean distance to the nearest human (d_h):

$$v_r(d_h) = m \cdot d_h + c \tag{A.3}$$

The slope m serves as the indicator of social reactivity, where a value of $m > 0.1$ indicates active, anticipatory deceleration as the robot encroaches upon the human agent.

A.2.4. Path Irregularity Ratio (PIR)

The Path Irregularity Ratio (PIR) assesses the deviation from an ideal trajectory. It is calculated as the ratio of the integral of the robot's actual velocity magnitude over time (the odometric path) to the optimal Euclidean distance between the start position (p_{start}) and the goal position (p_{goal}):

$$PIR = \frac{\int_0^T ||v(t)||dt}{||p_{goal} - p_{start}||} \tag{A.4}$$

A.3. The Social Navigation Report Card (Visualization Framework)

To rigorize the assessment of the robot's behavior, a post-hoc analysis dashboard—referred to as the Social Navigation Report Card—was developed. This artifact synthesizes raw telemetry into interpretable visual metrics, bridging the "Normative Void" by visualizing otherwise invisible interaction dynamics. The dashboard is divided into four analytical quadrants:

A.3.1. Quadrant 1: Global Activity Map

The top-left quadrant validates global path efficiency. It provides context for failure cases by overlaying the robot's trajectory against the ground-truth paths of human agents. Critical interaction points are highlighted as "Collision Zones," marked with red dashed ellipses to indicate areas of high spatial conflict.

A.3.2. Quadrant 2: Safety Monitor (Proxemic Analysis)

The top-right quadrant operationalizes Hall's theory of Proxemics (Hall, 1966) as a continuous time-series monitor. It demarcates the "Intimate Zone" (< 0.45 m) and highlights "Danger Zones" where the Time-to-Collision (TTC) metric drops below 2.0 s, thereby identifying near-miss incidents that binary collision metrics would fail to capture.

A.3.3. Quadrant 3: Politeness and Comfort Analysis

The bottom-left quadrant visualizes the Social Reactivity metric (defined in Section A.2.3). It imposes a **Linear Risk Constraint** on the robot's kinematics, classifying behavior into two zones:

- **Polite Zone (Green):** Indicates adherence to the risk boundary (slope $m > 0.1$).
- **Rude Zone (Red):** Indicates a violation of the boundary (slope $m < 0.05$), reflecting a failure to yield or decelerate appropriately.

A.3.4. Quadrant 4: Forensic Micro-Analysis

The bottom-right quadrant generates "Freeze Frames" of specific failure modes. This view facilitates a granular analysis of incidents, distinguishing between minor "grazing" contacts and "deep intrusions" by utilizing transparency-rendered agent footprints to visualize the severity of spatial overlap.

A.3.5. Event Clustering Logic

A specific challenge in collision reporting is "event bouncing," where a single physical interaction registers as multiple distinct failures due to signal noise. To resolve this, a clustering algorithm groups raw frames based on temporal continuity ($< 8.0\,\mathrm{s}$) and spatial proximity ($< 2.0\,\mathrm{m}$).

```python
def filter_events_spatial_temporal(raw_events, time_thresh=8.0, dist_thresh=2.0):
    if not raw_events: return []

    # Group events by Human ID
    events_by_human = {}
    for ev in raw_events:
        h_id = ev['human']
        if h_id not in events_by_human:
            events_by_human[h_id] = []
        events_by_human[h_id].append(ev)

    final_unique_events = []

    for h_id, ev_list in events_by_human.items():
        ev_list.sort(key=lambda x: x['timestamp_val'])
        current_cluster = [ev_list[0]]

        for i in range(1, len(ev_list)):
            prev_frame = current_cluster[-1]
            curr_frame = ev_list[i]

            time_diff = curr_frame['timestamp_val'] - prev_frame['timestamp_val']
            # Calculate Euclidean distance between frames
            dist_diff = np.sqrt((curr_frame['rx'] - prev_frame['rx'])**2 +
                                (curr_frame['ry'] - prev_frame['ry'])**2)

            # MERGE CONDITION:
            if time_diff < time_thresh or dist_diff < dist_thresh:
                current_cluster.append(curr_frame)
            else:
                # Select worst frame (closest distance) to represent the event
                worst_frame = min(current_cluster, key=lambda x: x['min_dist'])
                final_unique_events.append(worst_frame)
                current_cluster = [curr_frame]

        if current_cluster:
            worst_frame = min(current_cluster, key=lambda x: x['min_dist'])
            final_unique_events.append(worst_frame)

    return final_unique_events
```

A.4. Metric Logging

Standard data recording via `rosbag2` generated unmanageable file sizes ($> 6\,\mathrm{GB}$ per 20 seconds). This high throughput caused system-wide desynchronization, where human actors appeared to move in slow motion while the robot advanced in real-time, compromising data validity.

A.4.1. Optimization Strategy

A specialized **Metric Logger Node** was implemented to resolve the "Data Fidelity Crisis." Instead of recording the full simulation state, this node captures only essential relational metrics—such as distance, jerk, and momentum, in a compact custom binary format (`.parking`, approx. $50\,\mathrm{kB}$). Crucially, the node utilizes an `ApproximateTimeSynchronizer` to align robot odometry with simulation ground truth, preserving $100\,\mathrm{Hz}$ fidelity without inducing system lag.

A.5. Algorithmic Realization

To demonstrate the translation of the mathematical definitions provided in Appendix B into executable logic, the following core implementations are documented. These snippets are extracted from the analysis dashboard and planner logic.

A.5.1. Path Irregularity Ratio (PIR) Integration

The PIR metric is calculated post-hoc by integrating discrete odometry steps. The implementation below demonstrates how the metric is derived from the truncated dataframe to ensure it reflects only the active navigation phase.

```
# --- 2. Path & PIR (On truncated data) ---
path_dist = np.sqrt(df["pos_x"].diff()**2 + df["pos_y"].diff()**2).sum()
start_pos = np.array([df["pos_x"].iloc[0], df["pos_y"].iloc[0]])
optimal_dist = np.linalg.norm(np.array([GOAL_X, GOAL_Y]) - start_pos)

stats["path_length"] = path_dist
stats["pir"] = path_dist / optimal_dist if optimal_dist > 0 else 0
```

A.5.2. Data Integrity Filters

To ensure metrics reflect active navigation rather than static idling, data is truncated the moment the robot enters the goal tolerance radius.

```
# --- 0. DATA CUTOFF LOGIC (GLOBAL) ---
rx_raw = df["gt_robot_x"].values if "gt_robot_x" in df.columns else df["pos_x"].
    values
ry_raw = df["gt_robot_y"].values if "gt_robot_y" in df.columns else df["pos_y"].
    values

dist_to_goal = np.sqrt((rx_raw - GOAL_X)**2 + (ry_raw - GOAL_Y)**2)
at_goal_indices = np.where(dist_to_goal < GOAL_TOLERANCE)[0]

if len(at_goal_indices) > 0:
    first_success_idx = at_goal_indices[0]
    # TRUNCATE DATA HERE
    df = df.iloc[:first_success_idx + 1].copy()
    print(f"␣␣␣->␣Success␣detected!␣Truncating␣data␣to␣{len(df)}␣frames.")
else:
    print("␣␣␣->␣Robot␣did␣not␣reach␣goal␣tolerance.␣Using␣full␣data.")
```

B

The experimental testbed: scenarios and protocols

To ensure the ecological validity of the proposed navigation framework, the simulation environment was designed to replicate specific social puzzles rather than relying on stochastic random walk models. This appendix details the topography, actor parameterization, and specific validation scenarios used to evaluate the transition from moral will to moral skill.

B.1. Map topography

The physical simulation environment mirrors the constraints of a standard clinical corridor. The topography is defined by a primary hallway with a width of 5 m and a length of 20 m, adhering to architectural guidelines for healthcare facilities.

Key topographical features include:

- **Dimensional constraints:** The 5 m width enforces close-proximity interactions, ensuring that the robot cannot trivialize navigation by strictly maintaining large clearance buffers.
- **Bottlenecks:** Static obstacles are placed to create artificial narrowing, simulating equipment carts or stationary gurneys common in hospital wards.
- **Occlusion zones:** The geometry includes blind corners to test the robot's ability to anticipate unseen actors.

B.2. Kinematic profiles

To operationalize the complexity required by Step 4 of the CCVSD methodology, human actors were modeled as heterogeneous agents. Rather than generic dynamic obstacles, agents were assigned distinct kinematic and behavioral profiles derived from clinical and HRI literature. These profiles (Table B.1) dictate the velocity, turning constraints, and reactive latency of the simulated agents.

B.3. Scenario definitions

To enable a rigorous comparison between the baseline navigation algorithm and the experimental framework (CCVSD Step 5), a suite of five high-stress validation scenarios was developed. These scenarios operationalize the behavioral complexity of the actor profiles defined in Section B.2.

B.3.1. Subset A: public validation scenarios (social compliance)

These scenarios represent everyday multi-agent navigation phenomena in hospital corridors. They served as the primary stimuli during stakeholder interviews to assess perceived safety and social competence.

Scenario 1: the mixed traffic gauntlet (bidirectional flow)

Setup: A 2.5 m wide corridor populated with high-density, bidirectional traffic, including opposing wheelchair users and pedestrians.
Justification: This scenario evaluates emergent lane adherence. It tests whether the robot aligns with the collec-

Table B.1: Kinematic profiles of simulated human actors

Actor profile	Description	Behavioral parameters	Validation purpose
The rushing nurse	Goal-directed agent reflecting fast-paced clinical work.	High velocity baseline (1.43 m/s to 2.04 m/s emergency speed); low propensity to yield.	Tests the robot's ability to maintain social compliance under maximum flow stress.
The ambulatory patient	Represents vulnerable mobility and fall risk spectrum.	Low velocity range (below 0.7 m/s); high latency in path adaptation.	Tests the robot's safe overtaking boundaries and responsiveness to fragile agent behavior.
The wheelchair user	Introduces complex spatial displacement constraints.	Non-holonomic constraints (turning radius 1.09 m); high required turning space diameter (1.525 m ADA standard).	Tests the robot's capacity to recognize and preserve complex human mobility space.

tive flow (social mimicry) or attempts to opportunistically weave through transient gaps (efficiency maximization at the cost of predictability).

Scenario 2: the constrained overtaking zone (bottleneck)
Setup: The robot trails a slow ambulatory patient agent while opposing traffic occludes the passing lane.
Justification: This tests predictive horizons and social inhibition. The success criterion is determined by whether the robot suppresses overtaking behavior until conditions are genuinely safe, prioritizing the psychological comfort of the lead agent over goal-time optimization.

Scenario 3: the bi-directional sandwich (converging threat)
Setup: The robot faces simultaneous pressure from the front (a rushing nurse approaching at high velocity) and the rear (a faster pedestrian overtaking the robot).
Justification: This scenario tests proxemic prioritization under multi-objective conflict. The robot must negotiate a trajectory that avoids a frontal collision without abruptly braking or deviating in a manner that compromises the safety of the trailing agent.

B.3.2. Subset B: internal stress tests (kinematic scalability)

These scenarios probe the operational limits of the local planner and were excluded from stakeholder interview studies to focus purely on technical validation.

Scenario 4: the slalom agility test
Setup: Six static human actors are arranged in a staggered zig-zag formation along the 20 m corridor.
Justification: This tests the smoothness and jerk of the local planner under continuous trajectory updates. It validates the robot's ability to exploit its holonomic capabilities without inducing oscillations.

Scenario 5: the high-density crowd clusters
Setup: The robot must navigate through two distinct clusters of 4–6 walking pedestrians.
Justification: This serves as a test of computational scalability and robustness against the freezing robot problem, evaluating navigation performance in saturated costmap conditions.

B.4. Protocol variations

To simulate distinct ethical operational modes, the navigation stack is configurable between two distinct logic protocols. These variations fundamentally alter the cost function weights applied during trajectory optimization.

- **Hospital logic (efficiency):** Prioritizes system throughput and goal velocity. This mode reduces the cost penalties for entering the personal space of actors slightly to facilitate smoother flow in high-urgency environments, mimicking the assertive navigation of emergency staff.

- **Care home logic (responsiveness):** Prioritizes psychological safety and distance keeping. This mode significantly increases the cost penalties for proximity and jerk, resulting in a deferential navigation style suitable for environments with vulnerable populations, such as the ambulatory patient profile.

B.5. Software artifact: the digital testbed (pal_hallway_simulation)

Repository: `https://github.com/nicolaswim/pal_hallway_simulation`

While the sections above define the theoretical parameters of the validation scenarios, this software artifact provides the concrete implementation required for reproducibility. It serves as the digital twin of the experimental testbed, containing the exact `.world` and `.launch` configurations used to generate the data in Chapter 3.

B.5.1. Artifact contents

This repository operationalizes the definitions provided in Section B.3 into executable assets:

- **World files:** Contains the specific topography (5 m corridor width) and static constraints defined in Section B.1.
- **Launch configuration:** Pre-configured launch files that spawn the PAL Tiago robot alongside the procedural actors defined in Section B.2.

B.5.2. Execution

Researchers can replicate the social puzzles defined in Section B.3 using the following standard ROS 2 launch command:

```
# Launching scenario 1 (the mixed traffic gauntlet)
ros2 launch pal_hallway_simulation scenario_1.launch.py \
    robot_model:=tiago \
    navigation:=True \
    world_name:=scenario_1
```

B.6. Automated Scenario Generation

A scaling bottleneck emerged regarding the construction of validation scenarios. Manually authoring large XML world files with precise timing for multiple actors was fragile; single syntax errors often rendered the simulation unrunnable.

B.6.1. Methodological Resolution

To address this, a **Scenario Generator Shell Script** (`generate_scenario.sh`) was developed. This tool accepts high-level semantic descriptions (e.g., “3 agents, crossing pattern, speed $1.2\,\mathrm{m/s}$”) and procedurally generates valid `.world` XML files. This automation ensures the ecological testbed is both reproducible and scalable, removing manual authoring errors from the validation pipeline.

C

Software artifact: gazebo_actor_spawner

Repository: `https://github.com/nicolaswim/gazebo_actor_spawner`

C.1. Problem-solution statement

The problem: ontological bias in simulation

In standard Gazebo simulations, animated actors such as pedestrians suffer from an ontological bias: they exist as visual meshes within the rendering engine but lack the semantic properties required for robotic interaction. Furthermore, Gazebo Classic lacks a robust API for dynamically spawning these complex animated meshes at runtime via standard ROS services. Consequently, they function as ghosts, visible to the human eye but invisible to the robot's sensors, and are difficult to configure programmatically for batch testing.

The solution: a two-stage pipeline

To resolve this, the software implements a two-stage pipeline:

1. **Procedural world generation:** A workflow that constructs a transient simulation world file in memory, injecting user-defined actor trajectories before the physics engine boots.
2. **Semantic injection layer:** A runtime relay that intercepts the raw ground-truth state of the simulation, filters for actors, and rebroadcasts them as navigation-compatible entities (TF frames and markers).

C.1.1. Technical Limitations

This disconnect presented two distinct failures:

1. **Ontological Bias:** The robot's laser scanner failed to register human actors, causing the Social Momentum algorithm to receive null input for angular momentum calculations. The system effectively exhibited an ontological bias, treating safety as strictly physical (static obstacles) rather than relational.
2. **Coordinate Frame Discrepancies:** Native Gazebo actors frequently exhibited visual artifacts, such as "sliding" horizontally or appearing supine. This was caused by inconsistencies between Gazebo's coordinate standards and the ROS 2 ENU (East-North-Up) convention.

C.2. System architecture

The system architecture is defined in the main body of this thesis, see Figure 2.4.

C.3. Algorithmic implementation

The system's efficacy relies on two distinct algorithmic components: the procedural generation of the simulation environment and the runtime semantic injection.

C.3.1. Procedural world generation

To circumvent the limitations of Gazebo Classic's dynamic spawning API for animated actors, the system employs a procedural generation pattern. Instead of spawning entities into a running simulation, the software assembles a complete SDF world file in memory and boots the simulation with this transient asset.

Snippet A (XML injection)

The script aggregates user inputs to construct the SDF (Simulation Description Format) content programmatically, injecting the actor tags directly into the world definition via string interpolation.

Listing C.1: XML injection logic constructing the SDF world file in memory.

```
# From spawn_actors.py
sdf_content = f"""<?xml version="1.0" ?>
<sdf version="1.6">
  <world name="actor_world">
    ...
    {joined_actors}  <-- dynamic actor injection
  </world>
</sdf>
"""
```

Snippet B (transient file execution)

The generated content is written to a temporary location, and the simulation is launched pointing specifically to this ephemeral file.

Listing C.2: Writing the transient asset and launching the simulation subprocess.

```
# From spawn_actors.py
temp_world_path = "/tmp/generated_actor_world.sdf"
with open(temp_world_path, 'w') as f:
    f.write(sdf_content)

# Launch Gazebo with the generated world
cmd = [
    "ros2", "launch", "gazebo_actor_spawner", "view_actors.launch.py",
    f"world_path:={temp_world_path}"
]
subprocess.run(cmd, check=True, env=os.environ)
```

C.3.2. Semantic injection layer

Once the simulation is running, the relay node performs the semantic translation required for ROS 2 integration.

Snippet C (semantic filtration)

The system iterates through the high-frequency `/model_states` topic, applying a regex filter to isolate actors of interest (e.g., those prefixed with “human”). This prevents the system from broadcasting irrelevant simulation objects (like ground planes or the sun) as dynamic obstacles.

Listing C.3: Regex filtration logic for isolating social actors from the physics stream.

```
# From gazebo_actor_relay.py
self.actor_regex = re.compile(r"human_.*")

# ... inside callback ...
for actor_name in name_to_index.keys():
    if not self.actor_regex.match(actor_name):
        continue
```

Snippet D (planar constraint)

To correct for artifacts where actors appear to slide or lie down due to animation frame mismatches or physics engine quirks, the system enforces a planar constraint. The orientation is locked to the World Z-axis (identity quaternion) for the visualization markers, ensuring a consistent upright representation in the operator's view.

Listing C.4: Quaternion rectification to enforce upright planar constraints.

```
# From gazebo_actor_relay.py
# orientation: force upright (identity quaternion)
marker.pose.orientation.x = 0.0
```

```
marker.pose.orientation.y = 0.0
marker.pose.orientation.z = 0.0
marker.pose.orientation.w = 1.0
```

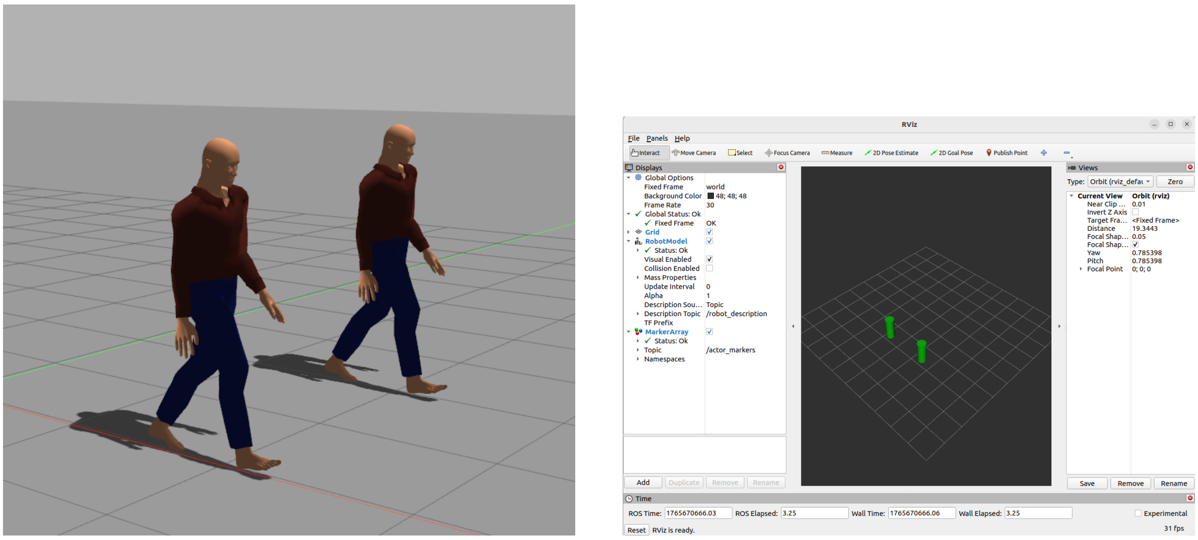
C.3.3. The Gazebo Relay Architecture

To bridge the infrastructure gap between the physics engine and the navigation stack, the `gazebo_actor_relay` node was developed. This middleware functions on a rigorous input-output logic with three primary functions:

1. **Data Injection (Semantic Entities):** The node queries the Gazebo physics engine via `/get_model_state` at $100\,\mathrm{Hz}$ and rebroadcasts actor positions as synthetic perception data (`/social_perception/human_poses`). Unlike raw sensor data, this forces the simulation to acknowledge social agents as distinct semantic entities rather than generic obstacles.
2. **Quaternion Rectification (Planar Constraint):** Raw pose data is intercepted to enforce a strict 2D constraint. As implemented in `tf2_wrapper.py`, the node explicitly overrides the simulation's 6-DoF orientation by setting the roll and pitch components to zero (`rot.x=0.0, rot.y=0.0`). This reconstruction ensures that even if the physics engine reports a "supine" or "tipping" actor, the navigation stack receives a clean quaternion orthogonal to the ground plane.
3. **Visual Parity:** The rectified states are broadcast to RViz as dynamic `MarkerArrays`, establishing strictly synchronized visual parity between the physics engine and the operator's visualization tools.

C.4. Visual validation (proof of life)

The following figure demonstrates the successful injection of semantic data.



Global Options
Fixed Frame world
Background Color 48; 48; 48
Frame Rate 30
Global Status: Ok
Fixed Frame OK
Grid ☑
RobotModel ☑
Status: Ok
Visual Enabled ☑
Collision Enabled ☐
Mass Properties
Update Interval 0
Alpha 1
Description Sou… Topic
Description Topic /robot_description
TF Prefix
MarkerArray ☑
Status: Ok
Topic /actor_markers
Namespaces

Type: Orbit (rviz_defau
Current View Orbit (rviz)
Near Clip … 0.01
Invert Z Axis ☐
Target Fra… <Fixed Frame>
Distance 19.3443
Focal Shap… 0.05
Focal Shap… ☑
Yaw 0.785398
Pitch 0.785398
Focal Point 0; 0; 0

ROS Time: 1765670666.03 ROS Elapsed: 3.25 Wall Time: 1765670666.06 Wall Elapsed: 3.25 ☐ Experimental
RViz is ready.
31 fps

(a) Gazebo classic (source)

(b) RViz2 (visualization)

Figure C.1: Visual validation of the semantic injection layer. Left: The raw actor in the physics engine (invisible to sensors). Right: The semantic transformation into a tracked entity in RViz, enabled by the relay node.

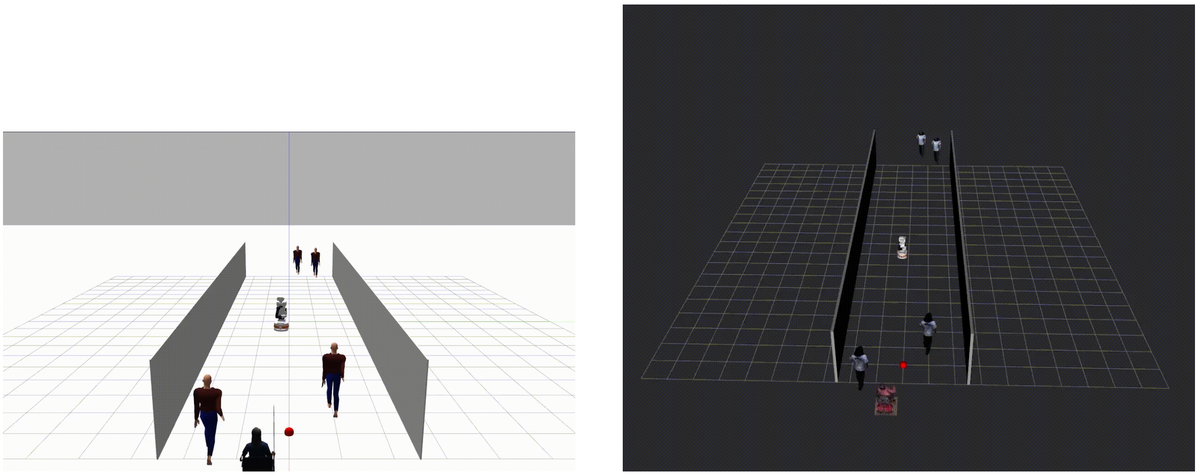
D

Software artifact: social momentum implementation

Repository: `https://github.com/nicolaswim/social_momentum_MPPI`

D.1. Visual demonstration

The system bridges the HRI infrastructure gap by synchronizing the physics engine with the robot internal planner. Figure D.1 demonstrates this parity, where the custom relay node ensures that social actors visible in the simulation (Gazebo) are accurately represented in the planner costmap (RViz).



(a) Gazebo simulation (source) **(b)** RViz visualization (planner)

Figure D.1: Operational view of the Paltiago social navigation module. Left: the PAL Tiago robot navigating a dynamic crowd in the Gazebo physics engine. Right: the corresponding real-time visualization in RViz showing the MPPI planner costmap and the semantic markers injected by the relay node.

D.2. System architecture and RQT graphs

To verify the structural differences between the baseline efficiency driven architecture and the experimental socially aware architecture, the runtime node graphs were extracted using the standard ROS tools.

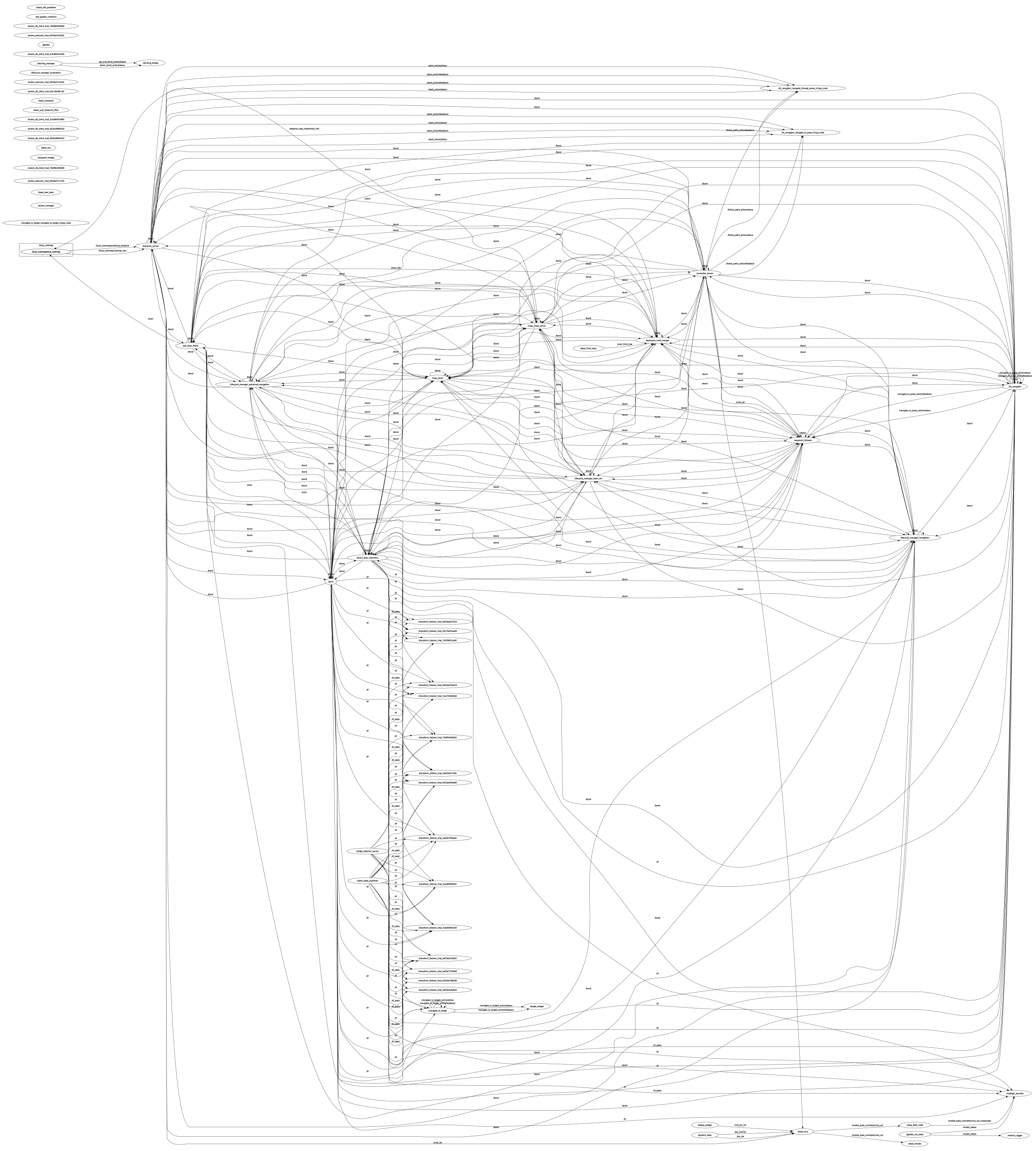


Figure D.2: Baseline architecture (Nav2). The standard stack relies on a complex behavior tree and a global planner optimizing for static efficiency. Note the complexity of the node graph compared to the experimental architecture below.

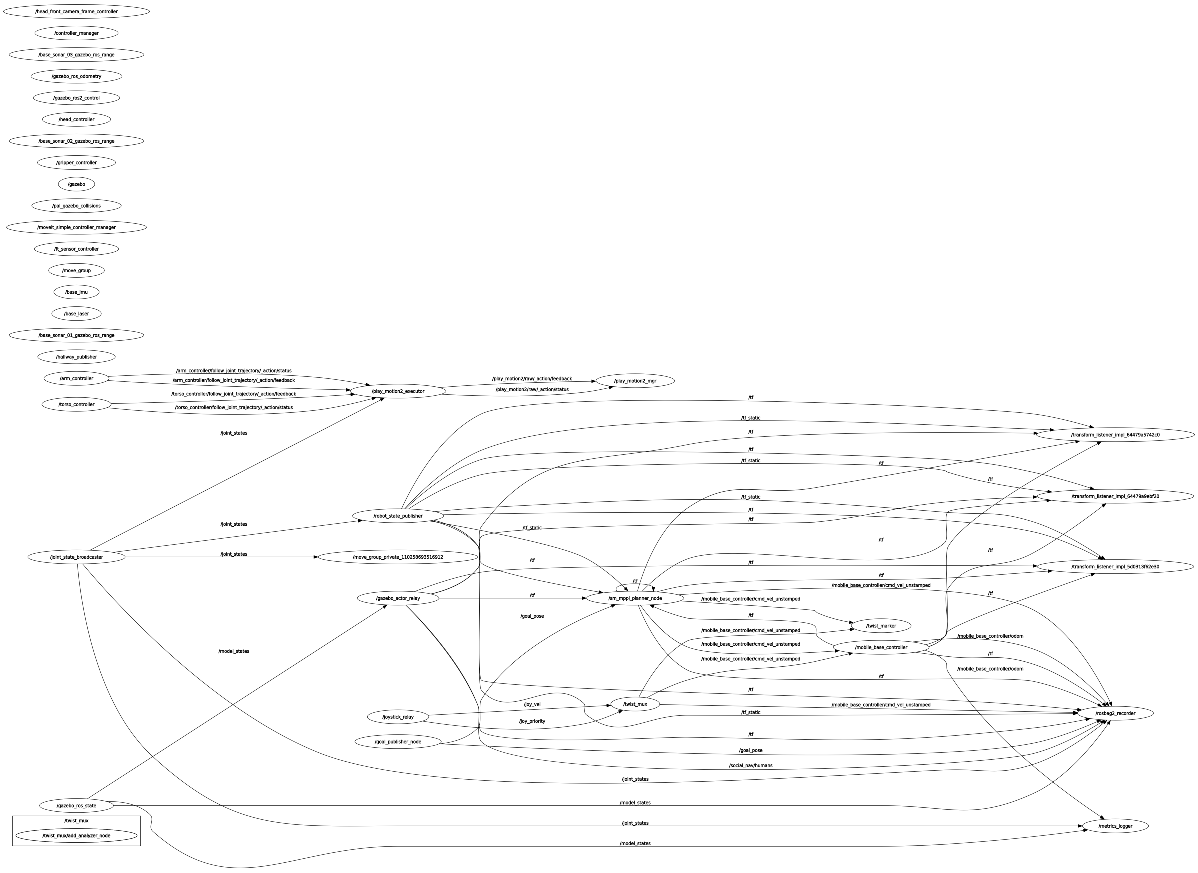


Figure D.3: Experimental architecture (social momentum). The custom stack bypasses the global planner using a flat MPPI controller loop to optimize for local reciprocity.

D.3. Dynamic Reconfiguration & Launch Architecture

The initial implementation of the Social Momentum algorithm functioned as a static proof-of-concept. Critical parameters, such as the `safety_threshold` and `agent_count`, were hard-coded into launch configurations, resulting in significant “code rigidity.”

D.3.1. Operational Constraints

This static architecture created a prohibitive bottleneck for validation. Modifying a single ethical parameter required a full restart of the ROS lifecycle nodes, a process taking approximately 4 minutes. This latency rendered iterative experimentation with stakeholders impractical and hindered high-frequency A/B testing.

D.3.2. Architectural Resolution

To mitigate this, the package was re-engineered to implement a **Dynamic Launch Architecture**. Hard-coded constants were migrated into ROS 2 parameters capable of runtime injection. This allows for rapid configuration changes via the command line:

```
ros2 launch sm_mppi main.launch.py safety:=1.5
```

By decoupling configuration from compilation, the testing loop was reduced from minutes to seconds, facilitating rapid sensitivity analysis of the safety parameters.

D.4. Repository structure and forensic code map

This section documents the critical path of the `social_momentum_MPPI` repository. The codebase is structured to separate the ROS 2 interface (`mppi_planner_node.py`) from the mathematical optimization core (`sm_mppi.py`) and the forensic logging infrastructure.

Listing D.1: Directory structure of the social momentum artifact.

```
src/sm_mppi_planner/
|-- launch/
|   |-- gazebo_relay_node_all_simulations.launch.py # Main experiment launcher
|   |-- recover_data.launch.py                      # Forensic data recovery
|   `-- sm_paramtrized_version.launch.py            # Dev launch file
|-- sm_mppi_planner/
|   |-- config.py                                   # Hyperparameters and limits
|   |-- mppi_planner_node.py                        # Main node: orchestration
|   |-- sm_mppi.py                                  # Core MPPI optimization logic
|   `-- tf2_wrapper.py                              # Perception utility
scripts/
|-- batch_bag_to_parquet.py                         # Batch processing
`-- bag_to_parquet_converter_clean.py               # ETL pipeline (bag to parquet)
```

D.4.1. The Social Momentum Cost Function (MPPI)

The following logic demonstrates the vectorization of the cross-product utilized by the Model Predictive Path Integral (MPPI) controller. This implementation utilizes `numpy` broadcast operations to compute the momentum cost for the entire trajectory batch simultaneously.

$$Cost \propto -(\mathbf{r}_{ac} \times \mathbf{v}_r + \mathbf{r}_{bc} \times \mathbf{v}_h) \tag{D.1}$$

```python
# Vectorized Implementation Logic
# r_hum: Relative position vectors (Batch x Time x 2)
# v_rob: Candidate velocity vectors (Batch x Time x 2)

# Compute 2D cross product (x1*y2 - y1*x2)
cross_prod = r_hum[:, :, 0] * v_rob[:, :, 1] - r_hum[:, :, 1] * v_rob[:, :, 0]

# Apply penalty where signs mismatch (Robot opposes flow)
# expected_flow is determined by the topological side preference
cost_momentum = np.where(
    np.sign(cross_prod) != expected_flow,
    MOMENTUM_WEIGHT * np.abs(cross_prod),
    0.0
)
```

D.4.2. The "Magic Numbers" of Ethics: Cost Function Weights

A critical finding during the code audit was the discovery of hard coded weighting factors within the primary optimization loop. As formalized in Equation (3.1), the total cost J is a weighted sum of conflicting objectives. However, the specific values for these weights were found to be arbitrary constants embedded directly in the source code rather than dynamic parameters exposed to the user.

```python
# Source: src/sm_mppi_planner/sm_mppi_planner/sm_mppi.py

def terminal_cost(self, state: torch.Tensor, action: torch.Tensor) -> torch.Tensor
    :
    # Implicit Weighting of Objectives
    # Goal: 2.0 | Static Obstacles: 5.0 | Social Momentum: 1.0 (Implicit)
    return 2 * goal_cost + sm_costs + dynamic_obstacle_costs + 5 * static_costs
```

Furthermore, the internal logic of the `SocialCost` function revealed a specific gain scheduling implementation that aggressively incentivizes topological consistency.

```python
def SocialCost(self, state: torch.Tensor, i, human_states) -> torch.Tensor:
    l_dot = l_ab[:, :-1] * l_ab[:, 1:]
    condition = l_dot > 0

    # "The Price of Politeness"
    # Reward for Consistency: -11.0 (High Incentive)
    # Penalty for Violation:  +10.0 (High Cost)
```

```
    penalty = torch.where(
        condition,
        -11 * torch.abs(l_ab[:, :-1]),
        torch.tensor(10.0, device=self.device)
    )
    return torch.sum(penalty, dim=1)
```

These scalars represent the robot's ”ethical constitution.” The audit reveals that the system values ”Momentum Consistency” (Reward -11) significantly higher than ”Goal Progress” (Weight 2). This creates a behavioral profile where the robot will actively sacrifice efficiency to maintain a legible social trajectory. The fact that this major ethical trade off was encoded as a static integer (-11) rather than a configurable parameter illustrates the normative void where design intent is often lost in implementation details.

D.4.3. Kinematic Constraints for Legibility

To resolve the ”Legibility Void” identified in Section 3.3.5, the robot's holonomic capabilities were artificially constrained at the algorithmic level. The code analysis revealed two specific mechanisms used to enforce this behavior.

First, the velocity limits were hard-coded directly into the MPPI initialization tensor, restricting the linear velocity to a conservative 0.4 m/s to ensure safe interaction speeds.

```
# Source: src/sm_mppi_planner/sm_mppi_planner/sm_mppi.py

self.mppi = MPPI(
    # ... (parameters omitted)
    # Linear Min: 0.0 m/s | Angular Min: -1.5 rad/s
    u_min=torch.tensor([0.0, -1.5], dtype=torch.float32).to(self.device),
    # Linear Max: 0.4 m/s | Angular Max: 1.5 rad/s
    u_max=torch.tensor([0.4, 1.5], dtype=torch.float32).to(self.device),
)
```

Second, and more critically, the `dynamics` function implements a strict Unicycle Model [50]. By deriving lateral displacement solely from the turning radius rather than accepting it as a control input, the algorithm mathematically deletes the robot's ability to slide sideways. This enforces a ”Turn-to-Move” behavior that is computationally inefficient but socially legible.

```
def dynamics(self, s: torch.Tensor, a: torch.Tensor, t=None) -> torch.Tensor:
    # Control Input a[:, 0] = Linear Velocity
    # Control Input a[:, 1] = Angular Velocity

    d_theta = a[:, 1] * dt
    turning_radius = a[:, 0] / (a[:, 1] + 1e-5)

    # Lateral change is coupled to rotation (Non-Holonomic)
    # The robot cannot simply add velocity to Y; it must curve.
    s2_ego[:, 1] = torch.where(
        torch.abs(a[:, 1]) < 1e-5,
        0.0,
        turning_radius * (1.0 - torch.cos(d_theta))
    )
```

D.5. The moral skill implementation

D.5.1. Target A: the logic (MPPI controller integration)

File path: `src/sm_mppi_planner/sm_mppi_planner/mppi_planner_node.py`

The logic: The planner node acts as the executive layer. It initializes the `SMMPPIController` and executes the control loop. The snippet below demonstrates the critical path where the robot state and the dynamic states of social agents are passed to the optimizer. This proves that the navigation stack is actively considering the social momentum of surrounding agents by injecting their state into the cost function computation every control cycle.

Listing D.2: Passing dynamic social states to the optimizer loop.

```
# Initialization of the social momentum controller
self.controller = SMMPPIController(
    static_obs=final_obstacles,
    device=self.device,
    active_agents=self.active_agents
)

# ... [Inside plan_and_publish loop] ...

# The optimization step: passing human states to the solver
action, costs, rollouts, termination = self.controller.compute_control(
    self.current_state,
    self.previous_robot_state,
    self.robot_velocity,
    current_agent_states_for_controller, # Contains dynamic human states
    self.previous_agent_states,
    self.agent_velocities
)
```

D.5.2. Target B: the perception (HRI bridge integration)

File path: `src/sm_mppi_planner/sm_mppi_planner/mppi_planner_node.py`

The logic: To bridge the HRI gap where simulations often fail to report human states to the robot, the system relies on a `GazeboActorRelay` broadcasting to TF. The code below shows the consumption side of this bridge. The planner iterates through the configured number of `active_agents`, dynamically resolving their positions via the `TF2Wrapper`. This ensures the robot is attentive to specific named social agents (`human_0`, `human_1`) in the environment.

Listing D.3: Consuming the semantic TF frames to resolve social agent positions.

```
if self.active_agents > 0:
    for i in range(self.active_agents):
        human_frame_name = f"{HUMAN_FRAME}_{i}"
        # Human pose: transform from "odom" (target) to "human_X" (source)
        T_odom_human_transform = self.tf2_wrapper.get_latest_pose("odom", human_frame_name)

        if T_odom_human_transform is None:
            self.get_logger().warn(f"Can't␣find␣human␣pose␣for␣{human_frame_name}...")
        else:
            # Extract and normalize agent state for the cost function
            q_az = T_odom_human_transform.rotation.z
            q_aw = T_odom_human_transform.rotation.w
            # ... [Quaternion to Yaw conversion] ...
            new_agent_state_list = [
                T_odom_human_transform.translation.x,
                T_odom_human_transform.translation.y,
                agent_yaw
            ]
```

D.5.3. Target C: the constraints (kinematic legibility)

File path: `src/sm_mppi_planner/sm_mppi_planner/config.py` & `mppi_planner_node.py`

The logic: Ethical legibility is enforced by strictly bounding the robot kinematic capabilities. The `config.py` file defines the physical limits (`VMAX`), and the `mppi_planner_node.py` enforces this limit on the final output command, ensuring the robot never exceeds socially acceptable speeds regardless of the optimizer output.

Listing D.4: Defining the ethical kinematic envelope.

```
# Evidence (Definition - config.py)
# --- Core MPPI and robot parameters ---
VMAX = 0.5  # Default max linear velocity (m/s)
DT = 0.2    # Default time step for MPPI prediction horizon
HORIZON_LENGTH = 15 # Lookahead: 15 steps * 0.2s = 3.0 seconds
NUM_SAMPLES = 250 # Trajectories sampled per cycle
```

Listing D.5: Enforcing the kinematic constraints before actuation.

```
# Evidence (Enforcement - mppi_planner_node.py)
# If the planner gives a valid action, publish it
twist_msg = Twist()
# Hard constraint enforcement before publishing to hardware
twist_msg.linear.x = min(float(action[0].item()), VMAX)
twist_msg.angular.z = float(action[1].item())
self.cmd_vel_pub.publish(twist_msg)
```

D.6. Execution and reproducibility

Launch command

The simulation and planner are initialized using the following ROS 2 launch command which orchestrates the Gazebo environment, the TF relay, and the MPPI controller.

Listing D.6: Primary launch command for the social momentum suite.

```
ros2 launch sm_mppi_planner gazebo_relay_node_all_simulations.launch.py scenario_id:=1
```

Critical configuration parameters

The operational envelope of the social robot is defined by the parameters in Table D.1, found in `config.py`.

Parameter	Value	Description
`active_agents`	2	The number of humans the robot actively tracks and optimizes for.
`goal_xy_tolerance`	0.15	The precision (in meters) required to consider the navigation task complete.
`startup_delay`	0.0	Wait time to allow the TF buffer to populate before engaging the planner.

Table D.1: Critical configuration parameters for the social navigation stack.

E

The design artifacts (ILK)

Note: This appendix consolidates the intermediate level knowledge (ILK) artifacts generated during the study. These artifacts constitute a ”moral skill handbook,” providing the necessary translational infrastructure to attempt the conversion of abstract ethical intent into concrete engineering specifications.

E.1. The operationalized CCVSD framework

To address the structural limitations of the standard linear CCVSD model, this research adapted the framework into a recursive, generative design cycle. This framework replaces the static assessment model with a continuous development loop. It serves as the master protocol governing the timing and application of the specific tools detailed in the subsequent sections.

Protocol definition

The framework mandates a shift from post-hoc ethical critique to continuous design input through three iterative phases:

- **Phase I: Exploration.** *Artifacts:* The Exploration Protocol & The Prospective Value Hierarchy. *Action:* Input stakeholder values → Output technical requirements.
- **Phase II: Development.** *Artifact:* The Development Phase Protocol. *Action:* Input requirements → Output traceable design decisions.
- **Phase III: Integration.** *Artifact:* The Contextual Evaluation Protocol. *Action:* Input prototype friction → Output scenario-specific constraints.

E.2. Phase I artifacts: Exploration

E.2.1. The Exploration Protocol

This protocol guides the identification of the ”semantic gap”—the divergence where identical terms mask conflicting priorities between stakeholders. It requires the engineer to map values not just as abstract concepts, but as conflicting operational definitions before technical work begins.

E.2.2. The Prospective Value Hierarchy (PVH)

The Prospective Value Hierarchy (PVH) serves as the structural guide for the exploration phase. It facilitates the decomposition of abstract values into executable specifications.

The Translation Matrix (Example) The translation of values is situationally bound to the specific project. The following table illustrates the specific translations generated for the social navigation case study.

E.3. Phase II artifact: the development phase protocol

This artifact addresses the ”normative void” between requirements and evaluation. Adapted from the Basic Design Cycle common in industrial design engineering, it structures the workflow to document micro-ethical trade-offs.

Tier 1: Abstract value	Tier 2: Contextual norm	Tier 3: Design requirement (spec)
Responsiveness *(The feedback loop)*	**Kinematic legibility** *(Predictability of intent)*	**Intention-aware path generation** *Requirement:* The local planner must encode intent into the motion profile (e.g., rotation before translation).
Attentiveness *(Noticing the need)*	**Saliency mapping** *(Prioritizing social actors)*	**Social prioritization mechanism** *Requirement:* Variable cost function weights based on proximity.
Justice *(Fair distribution)*	**Contextual calibration** *(Rejection of one-size-fits-all)*	**Bifurcated parameter sets** *Requirement:* The navigation stack must support runtime parameter swapping (hospital logic vs. care home logic).

Table E.1: Example of a PVH translation matrix derived from the case study.

The protocol steps

1. **Synthesize (the ex nihilo check):** Before implementation, the engineer audits the selected toolchain for inherent ontological biases (e.g., default cost functions prioritizing efficiency).
2. **Simulate (the relational sandbox):** Implementation is tested in a simulation utilizing a semantic injection layer to ensure human agents are recognized as social entities.
3. **Document (the traceability requirement):** Deviations from PVH specifications due to hardware limitations are logged as constraints rather than bugs, preserving the ethical reasoning.

E.4. Phase III artifact: the contextual evaluation protocol

This artifact is designed to address contextual instability. It moves beyond binary pass/fail grading to a generative analysis of context-specific friction.

Bifurcated guideline generation

The protocol synthesizes evaluation results into distinct parameter sets for distinct normative environments. The table below illustrates the guidelines generated during the case study.

Parameter	Guideline A: Hospital context	Guideline B: Care home context
Primary value	Competence (predictability & flow)	Attentiveness (responsiveness & yielding)
Velocity profile	Maintain constant velocity; minimize hesitation ($v_{avg} > 0.8m/s$).	Active deceleration in proximity; prioritize stopping over maneuvering ($v \to 0$ as $d \to 0$).
Conflict resolution	**Lane adherence:** Robot holds its line. Low inflation radius.	**Yielding:** Robot deviates or stops. High inflation radius.

Table E.2: Sample output of the contextual evaluation protocol: bifurcated design guidelines.

E.5. The digital handbook (web companion)

To facilitate access to these artifacts, the framework components have been consolidated into a digital tool.

Access: `https://0ccvsd.vercel.app/`

F

Qualitative Protocols

F.1. Raw Stakeholder Preference Data

To ensure the transparency of the qualitative analysis presented in Chapter 3, Table F.1 provides the complete dataset of stakeholder preferences. This table documents the individual algorithmic choices of all 20 participants across the three validation scenarios, categorizing them by professional role and institutional context.

Table F.1: Detailed breakdown of individual stakeholder preferences and algorithmic choices per scenario ($N = 20$).

Role	Institution	S1: Busy Hallway	S2: Narrow Pass	S3: Emergency
Nurse	Elderly Home	Social Momentum	Nav2	Social Momentum
Manager	Hospital	Nav2	Nav2	Nav2
Nurse	Clinic	Nav2	Nav2	Social Momentum
Manager	Elderly Home	Social Momentum	Nav2	Social Momentum
Rec. Therapist	Elderly Home	Nav2	Nav2	Social Momentum
Robotics Eng.	Hospital	Nav2	Nav2	Social Momentum
Nurse	Clinic	Nav2	Nav2	Social Momentum
Clinical Tech.	Hospital	Social Momentum	Nav2	Social Momentum
Patient	Elderly Home	Nav2	Social Momentum	Social Momentum
Clinical Tech.	Hospital	Nav2	Nav2	Social Momentum
Nurse	Elderly Home	Social Momentum	Nav2	Nav2
Rec. Therapist	Elderly Home	Nav2	Nav2	Social Momentum
Robotics Eng.	Hospital	Nav2	Nav2	Social Momentum
Doctor	Hospital	Nav2	Social Momentum	Social Momentum
Manager	Disability Care	Nav2	Nav2	Social Momentum
Nurse	Elderly Home	Nav2	Nav2	Social Momentum

F.2. Interview Protocols

To ensure transparency regarding the qualitative data collection process, this section provides the standardized interview manuscripts utilized during the study. These protocols were designed to operationalize the abstract concepts of the CCVSD framework into concrete probes capable of eliciting value judgments from non-technical stakeholders.

While the specific phrasing was adapted slightly when interviewing robotics engineers or management to align with their professional lexicon, the core structure remained consistent to facilitate cross-group comparison. For brevity, only the primary manuscripts for the healthcare worker cohort are reproduced here, as they represent the most comprehensive application of the inquiry strategy.

The scripts are divided into two distinct phases corresponding to the generative design cycle:

1. **Exploration Phase Protocol:** Focuses on identifying the ”Context of Care” and detecting the semantic gap before technical development.

2. **Integration Phase Protocol:** Focuses on validating the ”Context of Use” through A/B testing of the deployed prototypes (Nav2 vs. Social Momentum).

Script A: Healthcare Worker (Exploration Phase)

Context: Semi-structured interview conducted prior to robot deployment to map the existing care practice.

Script B: Healthcare Worker (Integration Phase)

Context: Scenario-based evaluation conducted after the robot deployment, utilizing video stimuli A/B testing.

INTERVIEW MANUSCRIPT: HEALTHCARE WORKERS (EXPLORATION PHASE)

Participant Name: ____________________ **Date:** __________
Role/Title: ____________________ **Ward/Dept:** __________

---

0. DISCLAIMER & CONSENT *(Read verbatim)*

"Thank you for participating. I am researching how to design robots that support—rather than disrupt—hospital care. Before we talk about specific technologies, I want to understand your reality on the ward: the pressure you are under, the flow of your shift, and your values as a caregiver. I am not here to 'sell' you a robot; I am here to ensure that future designs respect your clinical judgment. For the next 30 minutes, I will ask about your daily workflow, your definition of good care, and finally, I will ask you to critique a proposed model for how robots should behave. Are you ready?"

PART 1: THE CLINICAL REALITY & WORKFLOW

Goal: To map the "Context of Care"—daily routines, pressures, and definitions of success.

1. General Workflow

"Can you describe a typical 'peak hour' on your ward? What are the dominant constraints you face? Is it time pressure, patient volume, or physical fatigue? When technology (like EHRs or pumps) is introduced, does it usually help your flow or slow you down?"

---

2. The Definition of Success (Good Care)

"If you finish a shift and say to yourself, 'That was a good day, I did my job well,' what happened? What is your primary metric for success? Is it purely medical (all meds given) or is it social (patients felt heard)?"

---

3. Implicit Priorities

"In a moment of crisis, how do you prioritize tasks? For example, if a patient is crying but another patient has a critical alarm, how do you decide where to move? How much of your movement is dictated by 'efficiency' versus 'empathy'?"

PART 2: THE SEMANTIC GAP (Translating Care to Rules)

Goal: To identify how clinical values are interpreted by human staff vs. potential machines.

1. ATTENTIVENESS (The 'Clinical Gaze')

"In nursing, 'Attentiveness' means noticing small changes. If a robot is moving down the hall, what should it notice? Is it enough for it to see a person as an 'obstacle,' or does it need to recognize if a patient looks confused or in pain?"

---

2. RESPONSIVENESS (Reaction)

"If you are rushing to an emergency and a cleaning robot is in your path, what is the 'Responsive' thing for it to do? Should it freeze? Should it back up? Or should it try to communicate with you?"

---

3. COMPETENCE (Safety vs. Skill)

"How would you define 'Competence' for a machine on your ward? Is it enough that it doesn't crash into things, or does it need to understand hospital hygiene and privacy protocols to be considered 'competent'?"

---

4. RESPONSIBILITY (Accountability)

"If a robot is delivering medication and a patient refuses it or takes the wrong one, who is responsible? Do you feel you would still need to supervise the robot, effectively doubling your workload?"

---

5. JUSTICE (Fairness in Care)

"Robots are often programmed to be efficient. Do you worry that a robot might prioritize 'easy' tasks over 'difficult' patients? How do you ensure fairness in how time is spent on the ward?"

PART 3: STAKEHOLDER MAPPING

Goal: To see how the participant views the hierarchy of the ward.

1. The Hierarchy of Needs

"On this ward, whose needs come first? The Administration (efficiency), the Staff (workflow), or the Patient (health)? Where do you fear a robot might place its priority if programmed by engineers?"

---

2. Conflicting Needs

"Have you ever been in a situation where a visitor's needs conflicted with a patient's medical needs? How did you resolve it? Do you think a machine could understand that nuance?”

PART 4: EVALUATING THE PROPOSED CARE MODEL

”I will now show you the proposed ’Rules of Conduct’ for the robot.”

Context: *Hand over the printed* ***Table 1: The ’Care-Centered’ Robot Behavior Guidelines****. Explain that these are the 5 rules engineers are trying to program. Goal: Validate the theoretical model against clinical reality.*

■ 1. CRITIQUE OF FEASIBILITY

- ”Look at these proposed behaviors. What is your first reaction?”
- ”Does this look like a robot that would actually be useful, or does it seem like a burden?”
- ”Is this model too idealistic for the messy reality of this ward?”

■ 2. THE ’EMPATHY VOID’ (Rules vs. Reality)

- ”Look at **Rule 2 (The robot will always yield to humans)**.”
- ”Is that actually what you want? Or are there times the robot *should* have priority (e.g., carrying urgent blood samples)?”
- ”Can you imagine a scenario where following these rules strictly would actually endanger a patient?”

■ 3. THE STRUCTURE OF COMPARISON

- ”Compare this robot to a human porter or junior nurse.”
- ”What is the one ’human’ thing a porter does that this list of rules completely misses?”
- ”Does the robot lack the ability to ’read the room’ (social context)?”

■ 4. PRACTICAL APPLICATION

- ”If we placed this robot on your ward tomorrow, what is the first thing you would warn your colleagues about?”
- ”What training would *you* need to feel safe working alongside this machine?”

PART 5: CLOSING *(Retrieve Table)*

1. The ’Standard of Care’

”Ultimately, do you believe a robot can ever truly ’care’ for a patient, or is it only ever a tool? Does the presence of a robot lower the standard of human connection?”

2. The ’Magic Wand’ Question

”If you had a magic wand and could design one perfect task for a robot to do—so that you could spend more time with your patients—what would that task be?”

That concludes the interview. Thank you.

INTERVIEW MANUSCRIPT: HEALTHCARE WORKERS

Participant Name: ______________________ **Date:** __________
Role/Function: ______________________ **Institution:** __________
Exp (Years): __________ **Tech Exp (1-10):** ____

---

0. DISCLAIMER & CONSENT *(Read verbatim)*

”Thank you for participating. Before we begin, I want to clarify the format. I am researching how technology interacts with hospital environments. For the next 15 to 20 minutes, I will ask you about your experience working in this hallway, and then I will show you three sets of short videos showing a robot navigating a similar space. To ensure I do not influence your answers, I will remain neutral and keep my questions brief. I will ask you to read the scenario descriptions directly from the screen. There are no right or wrong answers—I am looking for your honest professional judgment and gut reactions. We can discuss the full background of my research and answer any questions you have after the interview is complete. Are you ready to begin?”

PART 1: THE BASELINE *(Do not show tablet yet)*

1. Relationship to the Space

”To start, could you tell me a bit about your relationship with the hospital hallways? When you are walking there, what is usually your main purpose or mindset? Is it just transit, or is it part of your work?”

---

2. Contextual Description

”Could you describe a typical ’busy’ moment in this specific hallway? What is happening around you?”

---

3. The Focus of Attention

”When you are walking down this hall to get to a patient, what are you paying attention to besides your walking path?”

---

4. Embodied Experience

”How does it feel physically to move through this space during a shift? Do you feel free to move, or restricted?”

---

5. The Golden Rule

”In your opinion, what is the unwritten ’Golden Rule’ for moving through this corridor safely and respectfully?”

PART 2: VIDEO SCENARIOS *(Hand over tablet)*

”I will now show you three scenarios. We will look at two different behaviors (Video A and Video B) for each.”

■ SCENARIO 1: THE BUSY HALLWAY

> ***Text on Screen:*** *Imagine a peak-hour shift. The hallway is crowded with nurses, visitors, and a wheelchair user passing in both directions. The goal here is flow—everyone needs to get somewhere.*

WATCH VIDEO A

- ”In your own words, how would you describe the robot’s movement?”
- ”If you were walking next to it, how would your body react?” *(Tip: Keep stride vs. hesitate?)*
- ”If this robot were a colleague, how would you describe their attitude?”
- ”Did you feel you knew where the robot was going *before* it actually moved there? Or did its movement surprise you?”
- ”Did the robot maintain the flow of the hallway, or did it force the staff to break their stride?”
- **Deep Dive:** ”How would you describe the robot’s awareness of its surroundings?” *(Attentiveness)*
- ”How was the effort of avoiding a collision distributed here?” *(Responsibility)*
- ”How would you describe the quality of this movement?” *(Competence)*
- ”How did the robot use the available space relative to others?” *(Justice)*
- ”What role did the robot seem to play in the hallway flow?” *(Social Norms)*

WATCH VIDEO B

”Now watch the second version.” (Repeat the exact same First Impressions & Deep Dive questions from Video A above)

COMPARATIVE ANALYSIS (Scenario 1)

- ”When you think back to the two styles, how did your body react differently to them?” *(Tip: Relaxed vs. Tense?)*
- ”If you had to describe the difference in ’personality’ or ’attitude’ between the two, what would you say?” *(Tip: Tool vs. Partner?)*
- ”Think about an emergency situation. How would these two styles fit into that specific pressure?”
- ”If you had to choose only ONE of these navigation styles to be permanently installed here, which one would you pick and why?” ....

■ SCENARIO 2: THE NARROW PASS

> ***Text on Screen:*** *Imagine a tighter situation. A wheelchair user and a pedestrian are walking towards the robot. Another person is walking beside the robot. Space is limited, and negotiation is needed.*

WATCH VIDEO A

- ”In your own words, how would you describe the robot’s movement?”
- ”If you were walking next to it, how would your body react?” *(Tip: Keep stride vs. hesitate?)*
- ”If this robot were a colleague, how would you describe their attitude?”
- ”Did you feel you knew where the robot was going *before* it actually moved there? Or did its movement surprise you?”

- ”Did the robot maintain the flow of the hallway, or did it force the staff to break their stride?”
- **Deep Dive:** ”How would you describe the interaction between the robot and the humans?” *(Reciprocity)*
- ”How did the robot react to the changing situation with the wheelchair?” *(Responsibility)*
- ”What is your impression of the safety of this maneuver?” *(Competence)*
- ”How aware did the robot seem of the tight constraints?” *(Attentiveness)*
- ”In terms of social manners, how did the robot behave?” *(Social Norms)*

WATCH VIDEO B

”Now watch the second version.” (Repeat the exact same First Impressions & Deep Dive questions from Video A above)

COMPARATIVE ANALYSIS (Scenario 2)

- ”When you think back to the two styles, how did your body react differently to them?”
- ”If you had to describe the difference in ’personality’ or ’attitude’ between the two, what would you say?”
- ”In a rush, would you prefer the robot that pushes through or the one that waits?”
- ”If you had to choose only ONE of these navigation styles, which one would you pick and why?” ....

■ SCENARIO 3: THE EMERGENCY RUSH

> ***Text on Screen:*** *This is a critical moment. A nurse is rushing towards the robot from the front to get to a patient. At the same time, someone else is approaching from behind. The nurse cannot afford to slow down.*

WATCH VIDEO A

- ”In your own words, how would you describe the robot’s movement?”
- ”If you were that rushing nurse, how would your body react?”
- ”If this robot were a colleague, how would you describe their attitude?”
- ”Did the movement surprise you?”
- ”Did the robot maintain the flow?”
- **Deep Dive:** ”Who took the lead in resolving this traffic conflict?” *(Responsibility)*
- ”How did the robot respond to the urgency of the nurse?” *(Attentiveness)*
- ”How did the robot’s presence affect the emergency workflow?” *(Competence)*
- ”How did the robot prioritize the different people in the hall?” *(Justice)*
- ”How does this behavior fit into the ’team’ rhythm of an emergency?” *(Social Norms)*

WATCH VIDEO B

”Now watch the second version.” (Repeat the exact same First Impressions & Deep Dive questions from Video A above)

COMPARATIVE ANALYSIS (Scenario 3)

- ”Which robot would you trust to get out of your way without you having to check your stride?”
- ”Which robot acted more like a supportive teammate in a crisis?”
- ”Is it better for the robot to freeze and be predictable, or to move actively out of the way?”
- ”If you had to choose only ONE of these navigation styles, which one would you pick and why?” ....

PART 3: CLOSING *(Turn off screen)*

1. Comparison

”At the start, you mentioned the ’Golden Rule’ is... [Repeat Answer]. Based on what you saw today, does this robot respect that?”

---

2. Impact on Work

”If this robot were deployed here tomorrow, how would it affect your relationship with your work and your patients?”

That concludes the interview. Thank you.