Uncovering Blindspots for Systemic Safety: Relational Accountability in Maritime Autonomous Systems

ATMADEEP GHOSHAL, King's College London, United Kingdom
CAITLIN BENTLEY, King's College London, United Kingdom
GORDON MEADOW, SeaBot Maritime, United Kingdom
MARTIM BRANDAO, King's College London, United Kingdom
DAVID WAVELL, Frontier Robotics, United Kingdom
JONATAN SCHARFF WILLNERS, Frontier Robotics, United Kingdom
SAUMYA SRIVASTAVA, King's College London, United Kingdom
ETHAN WOOLF MONINO, King's College London, United Kingdom
KIMBERLY TAM, University of Plymouth & The Alan Turing Institute, United Kingdom
HENRY DUFFY, SeaBot Maritime, United Kingdom
ANASMITA GHOSHAL, Indian Institute of Technology Kanpur, India

Systemic safety concerns the technical, organisational and human elements that govern autonomous systems effectively. As AI capabilities evolve, autonomous systems demonstrate increasing complexity, often involving multiple interacting components and learned behaviours across deployment contexts, making systemic safety progressively more difficult to achieve. Accountability is critical in this context because it helps to establish effective governance mechanisms, yet current accountability theorisation inadequately addresses systemic safety in autonomous systems. Recent advances in accountability theory have focused on relational accountability, examining how institutional configurations structure who can demand accounts from whom and who faces consequences. We argue that addressing accountability in safety-critical autonomous systems requires reconceptualising relational accountability as a *system of relationships* with different configurations and dynamics across design, regulation, and operation stakeholders and their public(s). Through a case study of autonomous maritime operations within an industry-research collaborative project, combining ethnographic data with stakeholder interviews involving designers, researchers, and operators, we analyse how and why relationship configurations within accountability processes shape systemic safety. Examining accountability processes through this relational lens uncovers blindspots for systemic safety by revealing how actors and responsibilities across seemingly disjointed organisational, human, and technical layers are connected, creating tensions in accountability outcomes. Our findings have implications for governance of autonomous systems across sectors, demonstrating why systemic safety depends on relationship configurations within accountability processes.

CCS Concepts: • **Social and professional topics → Computing / technology policy**; **Government technology policy**; • **Human-centered computing → Empirical studies in HCI**; • **Computer systems organization → Embedded and cyber-physical systems**.

Authors' Contact Information: Atmadeep Ghoshal, atmadeep.ghoshal@kcl.ac.uk, King's College London, London, United Kingdom; Caitlin Bentley, caitlin.bentley@kcl.ac.uk, King's College London, London, United Kingdom; Gordon Meadow, gordon.meadow@seabotmaritime.com, SeaBot Maritime, United Kingdom; Martim Brandao, martim.brandao@kcl.ac.uk, King's College London, London, United Kingdom; David Wavell, david@frontierrobotics.ai, Frontier Robotics, United Kingdom; Jonatan Scharff Willners, jonatan@frontierrobotics.ai, Frontier Robotics, United Kingdom; Saumya Srivastava, saumya.srivastava@kcl.ac.uk, King's College London, London, United Kingdom; Ethan Woolf Monino, ethan.woolf_monino@kcl.ac.uk, King's College London, London, United Kingdom; Kimberly Tam, kimberly.tam@plymouth.ac.uk, University of Plymouth & The Alan Turing Institute, Plymouth, United Kingdom; Henry Duffy, henry.duffy@seabotmaritime.com, SeaBot Maritime, United Kingdom; Anasmita Ghoshal, anasmitag@iitk.ac.in, Indian Institute of Technology Kanpur, Kanpur, India.

Additional Key Words and Phrases: Blindspots, Systemic Safety, Maritime Autonomous Systems, Relational Accountability

ACM Reference Format:
Atmadeep Ghoshal, Caitlin Bentley, Gordon Meadow, Martim Brandao, David Wavell, Jonatan Scharff Willners, Saumya Srivastava, Ethan Woolf Monino, Kimberly Tam, Henry Duffy, and Anasmita Ghoshal. 2026. Uncovering Blindspots for Systemic Safety: Relational Accountability in Maritime Autonomous Systems. https://doi.org/10.1145/3805689.3806753

1 Introduction

Maritime autonomous systems are increasingly used across sectors that underpin global infrastructure and security. The International Maritime Organisation's regulatory scoping exercise identified Maritime Autonomous Systems (MAS) as relevant to commercial freight transport, military operations, scientific research, and offshore services [18]. Defence uses have expanded in particular, with autonomous underwater vehicles deployed for mine countermeasures, subsea infrastructure protection, and surveillance [9]. In commercial contexts, autonomous systems have been introduced in offshore renewable energy operations for inspection and maintenance, whilst surface vessels have been developed to reduce emissions and improve operational efficiency [44]. At the same time, autonomous operations have raised safety concerns. In September 2025, the UK Marine Accident Investigation Branch opened an investigation into a collision between the crew transfer vessel *Iceni Legend* and the uncrewed surface vessel *X-18* during survey work at the Greater Gabbard offshore windfarm [46]. In December 2024, the autonomous barge *River Drone 5* collided with a container ship on the Scheur River near Rotterdam, leading to container spillage [42]. Additionally, DNV [12] reported a rise in maritime safety incidents between 2018 and 2024 by 42% despite the global fleet growing marginally in that period. An ageing fleet was presumed responsible for the surge in incidents. These opportunities and challenges point to multiple drivers for adopting autonomous systems in the maritime sector, yet their deployment creates new risks because the speed at which AI capabilities evolve consistently outpaces the institutional and regulatory processes designed to govern them [17], whilst labour shortages prompt rapid deployment.

To improve on this situation, Leveson [25] argues for a systems-theoretic approach to safety. Failures [1] rarely stem from isolated breakdowns of specific components but emerge from complex interactions across multiple scales stemming from technology, user/operator responses, organisational procedures, and regulations [38, 47]. Systemic safety highlights factors that require examination across technical, human and organisational layers [5]. However, this perspective does not specify *how* responsibility for managing these interactions is distributed across stakeholders. Accountability scholarship addresses this gap, examining how institutional configurations structure who can demand accounts from whom and who faces consequences when failures occur [11, 32]. However, prior research has not examined how such relationships affect systemic safety, or why issues emerge from interactions across organisational boundaries. We address this gap by examining accountability as a system of relationships, demonstrating that only by attending to sets of relationships can we see how and why their dynamics shape systemic safety outcomes in autonomous maritime operations.

Through a case study of autonomous maritime operations within an industry-research collaborative project, we examine how and why accountability relationship dynamics shape systemic safety outcomes. Our study combines ethnographic fieldwork at industry events and project meetings, with stakeholder interviews involving designers, researchers, and operators across the maritime autonomy sector. We make both empirical and theoretical contributions. Empirically, we identify three core accountability tensions within maritime autonomous systems operations (between design assumptions and reality, between different interpretations of expertise and operator roles, and between training and demands), demonstrating how epistemic asymmetries concentrate

[1]Throughout this paper, we use "failure" to refer to non-malicious system breakdowns and operational incidents, acknowledging that maritime regulatory discourse has at times distinguished these from adversarial incidents such as cyber attacks, which fall outside the scope of this study

risk on operators whilst insulating designers and manufacturers from consequences. We also identify 'authority inversion,' where operators retain legal liability whilst automated systems exercise increasing practical control and influence. Theoretically, we demonstrate how these tensions surface only by reconceptualising accountability as a system of relationships, with different configurations and dynamics across design, regulation and operational stakeholders. This reframing is crucial because existing accountability frameworks characteristically theorise what accountability *is* without specifying what it is *for*. When not anchored to a specific evaluative purpose, existing frameworks cannot distinguish when accountability mechanisms that are formally present, become functionally inadequate. We argue that in safety-critical autonomous systems, accountability should be oriented towards systemic safety. We show how the authority inversion we identify is a specific structural condition that predictably undermines systemic safety despite individual professional accountability or intent.

Our research thus addresses the following three research questions:

- RQ1: What tensions exist between design assumptions and operational realities in autonomous maritime systems?
- RQ2: How do different structural positions within the maritime autonomy sector produce these tensions?
- RQ3: How do these tensions shape the distribution of risk and responsibility across stakeholders?

2 Theoretical Framework

Accountability scholarship in the area of AI and autonomous systems shows divergent and sometimes conflicting uses of the term [13, 29]. This tendency stems from a failure to distinguish between what accountability *is* and what it *is for*. As a virtue, accountability describes qualities actors may cultivate, such as transparency, responsiveness, and a willingness to answer for conduct [13, 40]. As a mechanism, accountability describes relationships between actors. Bovens' [3] influential work explained this in terms of a process by which an actor provides information to a forum empowered to question that account and impose sanctions. This stems from principal-agent models of accountability, with roots in law [14, 33], economics [16, 22], accountancy [1, 23] and other social sciences [15, 39], which explores how best to align an agent's actions with a principal's interests through monitoring and incentives [22]. This model assumes bilateral and clear agent/principals, but in algorithmic and autonomous systems, these assumptions break down. No single actor possesses complete knowledge of system behaviour, which is a feature of how many AI systems are produced and used [8]. Each actor's 'accountability horizon' extends only partially into domains controlled by others, yet failures may cascade across these boundaries [8, 24].

Both the virtue and mechanism perspectives tend to prioritise retrospective judgement – or answerability for conduct – rather than clarifying what accountability should accomplish [31]. This narrow focus on blame neglects other productive outcomes such as stewardship [43], or trust-building [28]. As Donia [13] suggested, the field suffers from limited engagement with the diverse normative logics (such as representation or fiduciary duty) that accountability is expected to perform. In response, we argue for a shift from 'accounting-based' to pluralistic accountability theorising that prioritises systemic safety [11].

In safety-critical systems, we argue that accountability's purpose should be enabling systemic safety. Systemic safety is an emergent property arising from socio-technical systems interactions encompassing machine reliability, human well-being and institutional layers. Following Leveson [25], systemic safety requires examining work practices, socio-organisational and broader environmental contexts and their interactions across multiple scales. This contrasts with dominant accountability approaches in the maritime sector which currently focus on compliance or situated aspects of practice but not both concurrently. Current maritime discourse, driven by the IMO's Regulatory Scoping Exercise and roadmap for a goal-based MASS Code [19, 20] primarily views accountability through the logic of verifiable compliance. Classification societies such as DNV and ABS have introduced interim guidelines drawing on functional safety standards like IEC 61508 to validate system reliability at specific approval points [2, 10]. This creates an 'assessor's regress,' where a system is deemed safe because it

meets a metric, but the metric itself may be blind to real-world failures [32]. In contrast, human factors research highlights automation bias, degraded vigilance, delayed intervention, and intensified cognitive load, particularly when operators oversee multiple vessels simultaneously [27, 35, 45]. Whilst these insights prove critical for understanding operational challenges, the literature predominantly frames these challenges in terms of deficits – operators requiring training [34], or interfaces requiring redesign – without connecting these issues to systemic safety.

2.1 Relational Accountability for Systemic Safety

Building on Bovens' dyadic model, Metcalf et al. [32] argued that algorithmic accountability must be triadic, incorporating affected publics who have standing to contest claims. However, even triadic models inadequately capture accountability where multiple relationships structure processes simultaneously.

Driven by the need for a multi-dimensional approach that captures both the structural aspects of accountability processes and the granular qualities of relationships within them, we integrate Romzek and Dubnick's [37] typology, with Di Tullio et al.'s [11] pluralistic accountability framework. Romzek and Dubnick [37] identified four accountability types (bureaucratic, legal, professional, and political), covering the breadth of processes relevant to our analysis (see Table 1). Di Tullio et al.'s [11] eight 'elementary units of analysis' add analytical depth by specifying the relational qualities within each type. Applying this framework to autonomous systems has not previously been attempted.

Table 1. Relational Accountability: Types and Elementary Units of Analysis Used for Our Study

Dimension	**Aspects**	**Description**
Accountability Types Identified In Our Study [37]		
Bureaucratic	Superior/subordinate	Supervision-based hierarchical control
Legal	Lawmaker/executor; Principal/agent	Fiduciary obligations
Professional	Layperson/expert	Deference to expertise
Political	Constituent/representative	Responsiveness to publics
Elementary Units [11]		
Spatial vectors	Horizontal/vertical	Direction of power relationships between agents
Procedures	Participation, inquiry, contestation	Methods influencing decision-making
Institutionality	Collective/individual	Whether institution or members account
Expertise	Technocratic/lay	Knowledge basis for judgement
Consequences	Hard/soft sanctions	Penalties imposed on power holders
Formality	Regulated/informal	Degree of formalisation
Substance	Standards, criteria	Evaluation basis
Timing	Ex ante/ex post	When accounts are rendered

The next section presents our empirical case study, examining how designers' and operators' experiences across different positions in the maritime autonomy sector shape systemic safety outcomes, producing the conflicts and tensions we identify.

Table 2. Participants

Role Category	n
Designers	5
AI/Autonomy	(4)
Human Factors	(1)
Maritime operations	8
Management	(4)
Operations/Training	(4)
Sector	
Industry/Commercial	9
Defence	1
Consulting	2
Academic	1
Gender	
Male	11
Female	2

Table 3. Author Event Attendance

Event	Authors
Initial stakeholder workshop at Ocean Business	A2,A3,A5,A6, A10
Lab visit	A2, A3, A9, A10
International Maritime Organisation Safety Committee	A3, A8
Autonomous Shipping Expo	A2, A3
AI-safety risks workshop	A1, A2, A3, A4
International Maritime Organisation Safety Committee	A2, A3
Progress meeting	A2,A3,A5,A6
Training scenario testing	A2, A3, A5

A1–A11 correspond to authors listed in order on the title page.

3 Research Context and Methodology

The research setting is an industry/academic collaborative project led by Bentley at King's College London. Both industry partners are small and medium enterprises, employing less than 25 people. Project co-lead 1, Seabot Maritime focuses on workforce development across the maritime sector, supporting systemic safety by piloting training for operators of maritime autonomous systems and developing policy towards establishing training standards. Project co-lead 2, Frontier Robotics, develops an autonomy engine and control system for underwater robots.

Our approach centred on fieldwork conducted in 2025-26 within this project, focusing on periods when the team discussed objectives, planning, and implementation requirements for a training intervention targeted at operators. Data collection included participant observation at Ocean Business [2]and Autonomous Shipping Week[3], four additional day-long stakeholder discussions, and field notes. Participating in industry events enabled informal conversations with diverse stakeholders across the maritime autonomy sector. Additional semi-structured interviews with Seabot Maritime's stakeholders were arranged to support the research and expanded through snowball sampling. We conducted interviews with 13 participants including five designers/developers, seven operators and training specialists, and one AI platform provider representative (see Table 1). Interviews lasted 45-90 minutes, exploring professional backgrounds, experiences with autonomous systems, system behaviour in practice, failure experiences, and perceived challenges (see Appendix).

Interviews were conducted remotely via video conferencing, audio recorded with participant consent, and transcribed verbatim. All data were stored and processed in accordance with our ethics protocol registered at King's College London MRA-21/22-29616.

Over time, we developed deeper understandings of how systemic safety challenges arose from relationship dynamics between organisations and stakeholders, rather than only individual competency gaps. This insight

[2]https://www.oceanbusiness.com/
[3]https://www.advancedmaritimetechnologyexpo.com/

prompted theoretical reorientation towards accountability scholarship, drawing on the team's expertise across information studies, human factors and AI safety research.

The collaborative structure of this project created conditions for examining accountability dynamics that are typically obscured by commercial sensitivities and proprietary restrictions. However, this positioning created methodological challenges regarding our dual role as both collaborators and researchers analysing accountability processes across the project's stakeholders. The research team combines academic expertise in accountability, human factors, and AI safety with practitioner knowledge from the maritime autonomy sector. Our shared commitment to improving systemic safety shaped our analytical focus, whilst our varied disciplinary and institutional positions spanning computer science, social science, and maritime operations required ongoing reflexive attention to how our perspectives influenced what we observed and how we interpreted it.

3.1 Analytic Approach

Data analysis followed constructivist grounded theory methods [6], emphasising theory building from empirical patterns whilst attending to processes, relationships, and power dynamics [7]. Atmadeep Ghoshal conducted line-by-line initial coding of interview transcripts, second-coded by Anasmita Ghoshal. Focused coding was reviewed by Bentley and Meadow, before collaborative synthesis examined patterns across field notes, meeting recordings, and theoretical literature. Following Charmaz's [6] memo-writing methods, theoretical memos examined how accountability processes unfolded, with iterative refinement moving between data sources and literature [21, 41]. Regular check-ins with project team members provided additional triangulation [26].

4 Findings

Given the project's focus on improving human-AI collaboration in autonomous maritime operations, participants primarily recounted their experiences of operational practice, with accountability issues emerging organically through discussion. Nevertheless, our initial coding of participant interviews revealed that all four of Romzek and Dubnick's [37] accountability types were mentioned. Table 4 gives an overview of the key issues raised, with bureaucratic accountability centring on management concerns and institutional culture across sectors, legal accountability focusing on liability and prevalent blame culture, as well as the balance between technical and maritime knowledge and workforce concerns raised with respect to professional and political accountability respectively.

However, tensions discussed arose not from specific accountability types but from interactions between processes and actors across these types. Applying Di Tullio et al.'s [11] elementary units to examine relationship dynamics across accountability types, revealed systematic patterns regarding who has authority, whose knowledge counts, and who bears responsibility when failures occur.

Two overarching accountability failures run through our findings. The first is an **epistemic asymmetry**: designers' computational knowledge is treated as authoritative while operators' embodied, tacit knowledge is marginalised, concentrating risk on the latter while insulating the former from consequences. The second is an **authority inversion**: operators retain legal liability whilst practical decision-making authority migrates to automated systems. We structure findings around three core tensions that function as the mechanisms through which these failures manifest. The first tension (design assumptions vs. operational realities) is the primary site of epistemic asymmetry. The second (competing interpretations of expertise) is where authority inversion becomes visible. The third (training and operational demands) shows how both failures are structurally reproduced.

Table 4. Accountability Types Identified in Participant Interviews

Accountability Type	Key Issues Identified
1. Bureaucratic *Mentioned by 5/13 (38%)*	Tensions between hierarchical oversight structures and operational realities, particularly regarding management capacity, information flow, and the disconnect between policy directives and frontline practice.
2. Legal *Mentioned by 5/13 (38%)*	Pervasive concerns about liability attribution and blame culture, including anxieties about legal responsibility when autonomous systems fail, regulatory gaps, and the potential for defensive over-management of AI systems.
3. Professional *Mentioned by 5/13 (38%)*	Debates about the necessary expertise for autonomous system operation, including tensions between traditional maritime skills and technical knowledge, inadequate training provision, and the erosion of professional judgment through automation.
4. Political *Mentioned by 4/13 (31%)*	Questions of public trust and societal acceptance of autonomous systems, workforce concerns about job displacement, the importance of transparency for legitimacy, and the need for inclusive approaches to technology adoption.

4.1 Tension 1: Design Assumptions and Operational Realities

Participants in design and operational roles described fundamentally different experiences of autonomous maritime systems. These divergences were not simply matters of perspective but reflected patterns in how systems behaved across different contexts and what counted as evidence of capability or failure.

4.1.1 System Capabilities and Reliability across Operational Domains. We compare two groups with structurally different positions: operators and designers working primarily with surface autonomous systems, and those working primarily with underwater systems. These groups face distinct operational contexts and accountability consequences, yet as autonomous operations scale, the same personnel are increasingly envisioned to manage both — a cross-skilling demand that blurs previously distinct configurations in ways that examining either group in isolation would miss. Operators described experiences with MAS as temperamental prototypes. OP1's characterisation across multiple deployment contexts was stark: vessels *"spend more time alongside getting fixed than they do outside getting work."* As they explained, *"they break constantly... you kick them off the dock, they will break constantly"*. These were not isolated incidents but patterns that shaped operators' orientations towards the technology in which repeated encounters with systems showed that they did not behave as they were supposed to.

The specific failures operators recounted pointed to diverse sources of unreliability. OP5 described systems that failed to detect *"paddle boards three inches above waterline, lobster pots, or wooden fishing boats."* OP3 recounted a vessel running aground due to outdated software following a maintenance handoff. These accounts could not be attributed to any specific system or technology, we did not know which systems individual operators had used, and the sector encompasses vendors of varying maturity and rigour. What the accounts revealed was that, regardless of underlying technology, all interviewees were encountering systems that failed in ways they could neither anticipate nor diagnose.

When refining interview findings against field work data, consequences of reliability challenges varied systematically by operational domain in ways participants did not always differentiate. Most operator interviewees worked with surface vessel autonomy, where consequences were more severe. These environments introduce complexities that are difficult to design for regardless of technical approach, including other vessels not following

regulations [OP6], mariners who *"game"* COLREGs [4] to manipulate autonomous vessel behaviour, high-traffic areas, instances where vessels need to move faster than systems can refresh, or adversarial vulnerabilities.

In contrast, underwater systems presented a lower severity picture. Frontier Robotics' systems used bounded, task-specific approaches with autopilot functions verifying operations against reliability criteria multiple times per second. When underwater systems failed, interviewees recounted fewer consequences, such as equipment loss and mission disruption rather than major casualties or criminal liability. This contrast matters not because the groups are simply different, but because they are increasingly merged within single operational environments, making it necessary to examine the sets of relationships between them rather than treating each as a discrete accountability process.

4.1.2 Relationship Dynamics. Design and operational contexts positioned stakeholders differently in relation to knowledge about system behaviour, creating asymmetries that structured accountability relationships around expertise, substance, and timing. **Expertise asymmetries** positioned operators as laypeople despite their maritime expertise. Design work occurred through computational abstractions – testing scenarios, simulations, or confidence thresholds – generating knowledge about systems as algorithmic challenges. DES1's team worked on sensor fusion and anomaly detection through these technical frameworks, whilst DES2 characterised consultancy as *"paper-based designs"* rather than engagement with operational environments. Simultaneously, all design organisations faced challenges accessing appropriate testing environments due to regulatory restrictions, or the challenges creating realistic simulated environments, a limitation we experienced ourselves in the context of our wider project.

Meanwhile, operators developed embodied, tacit knowledge through practice. OP5 described practices of switching between camera views, examining sensor data for subtle patterns like cavitation[5] or RPM fluctuations[6]. These *"little telltale signs"* were learned through accumulated experience. Their response to anomalous behaviour *"if it doesn't look right, take control and get a feel"* reflected knowledge that was embodied.

Operators and designers also spoke of **substance instabilities**, meaning that what operators were held accountable for changed continuously. Surface systems updated *"over-the-air"* on two-week cycles, introducing what OP6 termed "update velocity" problems where the substance of system behaviour, and thus accountability, shifted bi-weekly. Operators managed vessels whose core functionality could change between deployments, with bugs resetting critical settings and leaving vessels in a *"world of pain"* until patches arrived. This created integrity-based accountability (relying on trust that updates would not introduce failures) rather than compliance-based accountability (rules and sanctions for verified performance).

This rapid **timing** scale, meant that any scope for more formal accountability checks were impractical, such that OP6 described his experience dealing with frequent updates as *"almost like cowboy outfits, people who have just thrown a system together and deploy it and don't care about the pre-planning, the pre-deployment, the safety behind it."* This characterisation applied to both hardware and software dimensions of system development: just as physical components were sometimes assembled without adequate testing, the same logic extended to software and algorithmic layers, where systems marketed as incorporating AI often comprised basic rule-based logic or rudimentary machine learning with limited validation. For readers from a computing background, we use the term AI throughout to encompass machine learning and related techniques, acknowledging that considerable variation exists between marketed capability and technical implementation. There were also differences between

[4] Collision Regulations refers formally to the Convention on the International Regulations for Preventing Collisions at Sea, 1972, the IMO treaty establishing navigational rules governing right of way, vessel conduct, and collision avoidance for all vessels operating on the high seas and connected navigable waters.

[5] Cavitation is a physical phenomenon where vapour bubbles form and collapse in a fluid due to pressure changes, often causing noise, vibration, and component damage.

[6] RPM fluctuations are a performance symptom, referring to unstable rotational speed of a motor or engine, which can arise from multiple causes, one of which may be cavitation.

surface and underwater architectures in terms of the point at which operators are involved in autonomous decision-making, as DES1 spoke of *"cancellable actions"*, where the system announced intent and proceeds unless operators interrupt, whereas Frontier Robotics' *"shared autonomy"* approach requires explicit approval before proceeding. Both have implications on time, but whether the human's position in the relationship actually improves systemic safety is uncertain.

4.1.3 Responsibility Distribution. When vessels broke down or behaved unexpectedly, operators could observe symptoms (navigation deviating from routes, systems entering degraded modes, sensors reporting conflicts) but lacked insight into internal logic or architectural decisions that might explain these behaviours. As OP1 observed, operators frequently *"can't tell clients why"* systems failed *"because we don't understand the autonomous vessels well enough ourselves."* Although they were formally accountable for system behaviour, they were unable to provide explanations that might distribute responsibility appropriately or identify which factors contributed to failures.

These information asymmetries were not incidental but served commercial interests. OP3 discovered that systems marketed as autonomous required *"constant supervision"* only after procurement, with vendors withholding this information *"until after you've bought it."* Organisations therefore made procurement decisions based on unsubstantiated capability claims, then managed systems whose performance diverged from vendor representations. Without mechanisms holding vendors accountable for misleading claims, operational organisations bore risks arising from marketing practices they could neither control nor verify in advance.

Proprietary architectures concentrated risk further. OP7 discovered their L3 system[7] *"only really works with all of L3 cameras and L3 this and L3 that,"* creating vendor lock-in. Faced with these, they developed workarounds *"in secret... to overcome the shortcomings,"* managing cobbled-together systems outside official support channels whilst manufacturers avoided accountability for architectural decisions creating these problems.

Perspectives on these asymmetries diverged by position. DES2 acknowledged explainability was *"obviously very useful"* but characterised transparency as *"more of a regulation or a policy issue"* rather than intrinsic to operator oversight. Meanwhile, OP8 articulated the resulting accountability problem as: *"Who's approving the computer? Who's verifying the computer decision process?"* Operators were expected to supervise autonomous decisions but lacked both access to decision logic and institutional mechanisms to verify whether choices were appropriate. Ultimately this distribution of responsibility reinforces the epistemic authority dynamics noted above, whereby designers' computational knowledge is treated as authoritative while operators' embodied, tacit knowledge is positioned as merely anecdotal, whilst externalising legal and professional accountability for operational failure onto operators.

4.2 Tension 2: Different Interpretations of Expertise and Operator Roles

A second fundamental tension emerged around what expertise autonomous maritime operations required, who possessed this expertise, and whose professional judgment should guide decisions when systems behaved unexpectedly.

4.2.1 Competing Visions of the Operator Role. Design assumptions reflected premises about operator skill requirements and scaling operations. DES1's vision of operators *"on watch, but potentially... doing their day job alongside it"* exemplified how designers imagined oversight as occasional validation rather than continuous management. Frontier Robotics envisioned operators monitoring fleets simultaneously rather than three people actively managing a single robot, hoping their system would reduce *"the experience needed to carry out the tasks... allow[ing] more independent smaller businesses to carry out inspections easier and cheaper."* This scaling implied

[7]Level 3 (conditional automation) requires human operators to remain available to take control when the system requests intervention, but allows disengagement from active monitoring during automated operation.

fewer operators would manage both surface vessels to launch underwater robots and the autonomous underwater operations themselves introducing cross-skilling demands across fundamentally different operational domains and control systems.

From a human factors perspective, DES5 argued humans are naturally poor at passive monitoring without active roles, requiring systems designed to maintain constant engagement rather than waiting for failures. DES5 also warned that automated systems and human navigators often develop divergent situational awareness, complicating crossing and give-way decisions and increasing collision risks. OP1 described his experiences of remote operations as: *"you lose that basic ability to understand what's going on."* Sensory inputs (smell, sound, vessel motion) functioned as early warning systems which remote interface-mediated supervision could not replicate.

Moreover, operators shared experiences of expanding complexity rather than simpler work. OP1's assessment was pointed: *"at the moment we almost need superhumans to deal with it."* Automation introduced new tasks requiring rare expertise combinations: navigational competence, mechanical troubleshooting, and system diagnostics that neither traditional maritime training nor manufacturer instruction provided. OP4's characterisation of successful operators as *"unicorn people in industry"* captured how these skill combinations limited deployment.

4.2.2 Relationship Dynamics. The contested visions of what the operator role requires, and who is competent to fulfil it, have implications for professional accountability and the distribution of responsibility when systems fail. OP2 and OP8 described a shift taking place, with *"people operating remotely operated vessels... who are not trained mariners"* successfully managing systems from remote operation centres. Interviewees speculated that the sector was increasingly drawing personnel from diverse disciplinary backgrounds, such as materials engineers, or aviation engineers. OP5 observed reaching *"a cusp of 50-50"* between seafarers and technologists signalling a shift with cultural implications.

The shift towards technology-oriented personnel created risks of competence erosion in precisely those situations where human intervention remained critical. OP1 drew on aviation precedents, describing *"skills fade"* among pilots who relied heavily on automation, with the Air France 447 crash serving as reference for when automated systems failed and operators lacked foundational recovery skills. The concern for maritime autonomy was paralleled as though operators trained primarily to monitor automated systems might lack the seamanship needed during emergencies, creating situations where automation undermined competencies essential for responding to system failures.

However, political accountability is also at play because of workforce concerns about job displacement and ensuring that the public perceives the transition to AI as equitable. DES4 spoke of diversity as a solution to the impending personnel crisis in the maritime sector, acknowledging that as the pool of traditionally *"qualified people"* (such as experienced master mariners) runs out, the industry must find new ways to attract and integrate people into new roles, noting that people historically excluded by the credentialist gatekeeping of the maritime sector may lack the professional self-efficacy to envision themselves within high-tech roles. These dynamics reflect **vertical spatial vectors**[8], positioning seafarers and marginalised individuals towards the bottom of hierarchical relationships.

Maritime safety culture developed through centuries of collective experience has cultivated essential risk consciousness alongside credentialing. OP5 described this culture as *"very structured and the reason it is structured is for safety,"* transmitted primarily through enculturation and shared professional identities. Their warning—*"if we can keep the culture, autonomising things is not going to be a problem. But if you lose the culture with it, then things are going to happen"*. This raises the question of what "keeping the culture" could or should mean for professional accountability and whether it refers to safety vigilance developed through long maritime traditions

[8]Vertical spatial vectors, drawn from Di Tullio et al.[11], refer to hierarchical accountability relationships in which those with less institutional power are accountable upward to those with more.

or maintaining the credentialist barriers that have historically determined who counts as legitimately maritime. The demographic shift risks losing the former whilst potentially enabling dismantling of the latter, but it is not one or the other.

Maritime blame culture was another issue that structured relationships, which refers to how punitive actions are considered essential. As OP5 observed, *"the maritime industry still has a very big blame culture,"* explaining that when accidents occurred, *"there has to be someone to blame. There has to be someone to either pay a fine or do prison time."* OP5 noted that *"there has to be a master mariner at the end of this chain that has to take the blame for it,"* even when systems made decisions autonomously.

4.2.3 Responsibility Distribution. Interviewees spoke of how the **expertise** typically granted to maritime professionals, such as their situated judgment about when conditions were unsafe or when systems required human intervention, was being systematically undermined. Yet the **formality** of accountability relationships was itself in flux: whilst traditional mariners retained formal responsibility as Master Mariners and legally accountable decision-makers, emerging remote operation roles lacked formal professional standing (see Section 4.3), such that responsibility distribution is often deferred to formal responsibility through traditional qualifications.

Yet, OP8's career trajectory illustrated an authority inversion within shifting formalities. Early in their maritime work, *"the computer was telling me what was happening around me. It was giving me the information and I was making decision on positioning the vessel"*—automated systems provided decision support whilst humans retained both authority and professional accountability within established regulatory relationships. By career's end, *"the computers were keeping the vessel within 10 centimetres of a spot,"* exercising continuous control that left operators monitoring rather than deciding. Practical authority transferred to automated systems, yet bureaucratic and legal accountability structures remained unchanged. Formal accountability relationships (Master Mariner as licensed decision-maker) persisted whilst operational realities diverged.

This mismatch reflected how **institutionality** tended towards individuals. Technology represents a collective institutional product (designers, engineers, vendors operating across organisational boundaries), yet formal accountability mechanisms can only target individual humans, with OP5 observing: *"How do you send an autonomous system to prison?"* Simultaneously, professional accountability depends on professionals exercising judgment based on expertise, but when institutionality defers responsibility to individuals rather than the institution, operators are left accountable for decisions they did not make. OP3 argued that in these circumstances *"[if] they're relying on a system to make decisions, I don't think they're ever fully going to trust in that system."* Likewise, OP8 agreed that AI remains *"a great tool to make a decision, but not to be the decision maker"* until *"regulatory frameworks... alleviate the person from an illegal repercussion anywhere in the world."*

Enabling systemic safety for full autonomy would require not just formalising new roles, but establishing institutional mechanisms for collective accountability, and matching formal responsibility with actual authority. With no such transformations underway and prevalent blame culture ensuring individual sanctions, OP8 concluded: *"I don't think anyone's ready to take the step to 100% say they trust in a system where they are still in the crosshairs of blame while a system is making all of those decisions on their behalf."*

4.3 Tension 3: Training and Operational Demands

The third core tension concerned how operators were prepared for autonomous maritime work, revealing fundamental misalignments between what training provided and what operational contexts demanded.

4.3.1 Misalignment between Training and Operational Contexts. Training programmes reflected assumptions about knowledge as primarily procedural. OP8 described their organisation's approach as *"largely just a, you know, an e-learning module style training on how to use the autonomous vessel,"* focused on button functionality. OP1's critique captured the problem: manufacturer training *"tends to just be if you push this button, this happens,"*

leaving personnel capable of executing procedures but unable to diagnose failures. Recipients *"don't have the background knowledge or skills to be able to understand the context of what they're being shown."*

Operators spoke of needing to supplement training, as OP1 described efforts to *"close that skills gap between the knowledge of a seafarer and the knowledge of somebody who can then go and learn OEM training,"* acknowledging *"the reality is they can't most of the time."* Their own effectiveness came from mechanical engineering training that most seafaring pathways did not include. The competence gap reflected fundamental misalignment between expertise traditional maritime pathways cultivated and what autonomous operations required – what OP4 characterised as *"unicorn people in industry."* Reliance on exceptionally skilled personnel to compensate for inadequate training became the industry's de facto approach.

4.3.2 Relationship Dynamics. **Procedures** for training discussed by operators highlighted a narrow focus on technical functionality. OEM [9] providers designed training based on internal system architecture, producing instruction that OP1 characterised as created by engineers who *"know too much"* about systems but *"too little about good seamanship."* Training became box-ticking—operators completed mandatory modules, without enabling deeper understanding of how to diagnose and solve problems.

When interviews were conducted, OP1 observed *"there's no real certification for these autonomous vessels."* Since then, the UK MCA released the Remote Operator Training and Certification Pilot Framework [30], emerging from a multi-stakeholder working group (NCG-MASS) including over thirty organisations. However, the framework establishes generic baseline competencies whilst recognising system-specific technical knowledge continues residing with OEMs. This maintains separation between generic competencies (governed by regulatory frameworks) and system-specific knowledge (controlled by OEMs through proprietary training), leaving the fundamental OEM-operator knowledge asymmetries described in 4.1 largely unaddressed.

Alternatively, the Navy trains operators completely differently through a unit called the Fleet Operational Sea Training (FOST) unit, which is the authority responsible for putting naval crews through their paces and signing them off as safe for deployment. Whilst FOST is highly competent in training for traditional naval scenarios like fires, floods, and war-fighting, interviewees found they lacked technical understanding of the new systems they were using and were supposed to be evaluating them on.

In contrast, Seabot Maritime's close proximity to manufacturers enabled deeper insights into technical specifications through developing partnerships with companies directly. They also developed tacit knowledge of operational contexts by delivering training within those contexts, and worked across providers employing different autonomous systems, uncovering cross-skilling needs required to implement Frontier Robotics' technology effectively. Yet their positioning as a small and medium enterprise exposed them to structural vulnerabilities. The MCA's Remote Operator Pilot Framework – developed partly through working groups in which SeaBot participated – is oriented primarily around operators, with training provider recognition handled through a separate short course approvals process. This process assesses whether providers can meet defined standards, but does not consider which organisations are best positioned to deliver training on emerging autonomous systems technology, nor does it account for the disproportionate burden that certification costs place on SMEs relative to larger institutional providers. Given that the explicit goal of the pilot framework is to learn from experience and iteratively refine standards, these considerations seem worthwhile to unpack. The organisations that may be best placed to contribute meaningful knowledge to that learning process may be precisely those least well supported to participate in it (their spatial vector vulnerability), due to a lack of formal organisational standing (formality).

4.3.3 Responsibility Distribution. **Consequences** for training inadequacies distributed asymmetrically. Operators faced hard sanctions—prison, loss of licence—when systems failed, regardless of whether training prepared them.

[9] OEM (Original Equipment Manufacturer) training refers to instruction provided directly by the system's manufacturer, typically focused on the specific technical operation of their product rather than broader operational or seamanship competencies

Maritime blame culture ensured *"there has to be someone to blame"* (OP5), with that person being the Master Mariner even when failures stemmed from inadequate training or misleading capability claims they could not verify. The MCA framework establishes whether operators completed approved training but provides no mechanisms for verifying whether OEM training adequately prepares operators for specific systems, whether manufacturers accurately represented capabilities during procurement, or whether organisations successfully bridged generic/system-specific knowledge gaps. Professional and legal accountability operators face (4.1, 4.2) hinge on system capabilities, yet manufacturers' responsibilities for disclosure or training adequacy remain unspecified because procedures (training) remain inadequate and the consequences asymmetric.

5 Discussion

Our findings reveal how accountability relationship configurations produce tensions undermining systemic safety in maritime autonomous systems. Across three core tensions, we found systematic patterns: designers' assumptions diverged sharply from operational realities (RQ1); these divergences arose from structural positions that privileged technical expertise whilst marginalising operational knowledge (RQ2); and responsibility concentrated on operators bearing legal liability without commensurate authority or epistemic access, whilst manufacturers and designers remained insulated from consequences (RQ3).

Accountability processes are frequently fragmented. Across all three tensions, we observed accountability relationships structured by different processes (bureaucratic, legal, professional, political [37]) that fail to connect. Legal accountability flows vertically (operator to maritime authority), professional accountability flows horizontally (operator to seafaring community), yet no accountability relationship connects operators' professional knowledge to designers' technical decisions. Regulatory frameworks emerging to bridge these gaps reinforce fragmentation: they establish that operators must be certified but leave technical content to manufacturers, creating parallel accountability tracks that never intersect. When compliance-focused legal accountability dominates as reflected in current regulatory approaches [18, 20], and professional accountability erodes, systemic safety will likely suffer.

Relationship positions and epistemic standing. Buhl et al. [4] warn of "assessor's regress" in frontier AI safety cases, where systems are deemed safe because they meet metrics that may be blind to real-world failures. We demonstrate this problem runs deeper: epistemic asymmetries determine whose knowledge counts as legitimate evidence before safety cases are ever constructed. Operators' embodied, tacit knowledge—developed through managing systems that 'break constantly' and by recognising subtle failure patterns—is positioned as anecdotal within technical certification processes that privilege computational performance metrics. When relationship dynamics determine whose knowledge counts, they determine which risks become visible and which remain unknown, uncovering blindspots for systemic safety.

Responsibility accumulates without commensurate authority or knowledge. Whilst prior work has deconstructed the '80% human error myth' as a barrier to systemic learning [47], our findings progress this critique by showing how this myth is structurally maintained through an 'authority inversion.' Operators bear legal and professional accountability whilst decision-making authority migrates to automated systems. Maritime blame culture magnifies felt responsibilities for challenges operators cannot control, predict, or explain. Examining **substance** (what operators are accountable for) and **timing** (when accountability operates) revealed how responsibility distributes asymmetrically between operators and manufacturers, with individual operators exposed to hard sanctions alone.

5.1 Examining Dynamics Across Sets of Relations

Prior accountability research in AI and autonomous systems has predominantly examined discrete relationships in isolation. Di Tullio et al.'s [11] framework is among the few to tackle hybrid accountabilities, but their approach

evaluates accountability at institutional levels against normative standards. We instead trace how dynamics within and across relationships create tensions for systemic safety.

Three analytical strategies proved valuable. First, tracing actors across multiple simultaneous accountability relationships showed how operators maintained bureaucratic accountability to employers, legal accountability to maritime authorities, professional accountability to seafaring communities, and market accountability to clients – mirroring Messner's [31] observation that managers must maintain oversight across these relationships simultaneously. Examining relationships in isolation would miss how legal accountability structures (individual liability) conflict with professional accountability premises (exercising expert judgment) when authority has shifted to automated systems, producing the individualising effects Roberts [36] described. Second, examining relationship dynamics across lifecycle phases revealed how procurement relationships transformed post-purchase: manufacturers positioned themselves as service providers during sales whilst withholding operational limitations, concentrating risk on operators after procurement, a dynamic reflected in Cobbe et al.'s [8] analysis of algorithmic supply chains. Third, identifying power asymmetries in claims around safety revealed that operators' experiences of failure are difficult to connect to design perspectives. When autonomous systems are designed for narrow domains, broader systemic safety concerns can be overlooked; organisations that gain epistemic authority may simultaneously find themselves marginalised within domain hierarchies. Only by considering these asymmetries in relation to one another do the broader risks to systemic safety emerge.

5.2 Implications for Studying and Governing Autonomous Systems

For researchers, our findings suggest that accountability frameworks, including methods used to audit or evaluate autonomous systems need to extend beyond individual system behaviour or operator-system interaction to examine the configuration of accountability relationships across the full design-deployment lifecycle. Evaluation frameworks that assess operators in isolation will systematically miss how risk allocations are fixed before operators ever encounter systems [9, 36]. For human-AI interaction researchers, our findings show that operator knowledge — tacit, embodied, diagnostically rich — is not simply underutilised but actively downgraded within technical certification processes[19, 38]. Surfacing this knowledge requires methods that follow actors across organisational boundaries, not just within single deployment contexts. Ultimately, whilst Di Tullio et al.'s [11] framework proved useful within the scope of our analysis, two elementary units revealed limitations when applied to autonomous systems specifically. Institutionality assumes human actors can be sanctioned, yet autonomous systems as collective institutional products cannot. Contestation was also structurally limited, as proprietary architectures prevented operators from accessing the information needed to contest decisions. Incrementally refining such frameworks for autonomous systems may be missing the point, however, because these limitations only become visible when examining how relationships interact. It is the configuration of relationships together, not any individual relationship or its qualities necessarily, that produces the dynamics undermining systemic safety.

For governance, our findings point to the need for feedback mechanisms that connect operational experience back to system design and regulatory requirements. Three specific interventions are implied. First, mandatory incident reporting requirements, analogous to aviation's confidential human factors reporting systems, would institutionalise operators' tacit failure knowledge, feeding back into both training requirements and design decisions [42, 43]. Second, current safety standards rely on technically-derived metrics that are blind to real-world failure modes; regulators should require sociotechnical benchmarks co-developed with operators, grounding conformity assessments in actual deployment complexity rather than laboratory or simulation performance alone. Third, procurement frameworks should impose post-deployment feedback obligations on manufacturers as a condition of contract, ensuring that capability limitations surfaced through operational experience create obligations for manufacturers rather than risks borne solely by operators. Together, these measures would

begin to institutionalise the operational knowledge that currently constitutes the sector's most important safety resource.

6 Reflexivity & Limitations

Our ethnographic approach and embedded position as project collaborators provided significant methodological advantages. Research on accountability in autonomous systems is often constrained by limited access due to proprietary restrictions and commercial sensitivities. Having SMEs situated within the project enabled examination of design and operational practices together with ethnographic depth rarely attainable in this research area. This positioning allowed us to observe relationship dynamics as they unfolded—in project meetings, industry events, and stakeholder discussions—rather than relying solely on retrospective accounts.

This depth comes with limitations. Our findings emerge from a specific collaborative context between King's College London, Seabot Maritime and Frontier Robotics, and the accountability dynamics we observe may manifest differently in larger organisations, different national regulatory contexts, or other autonomous systems domains. We therefore emphasise the conceptual approach over direct generalisation, inviting future research to examine how systems of accountability relations structure epistemic asymmetries and responsibility distributions elsewhere. Key issues including environmental sustainability and inclusion warrant further investigation.

7 Conclusion

In this paper, we examined accountability relationships in maritime autonomous systems and showed how specific configurations of these relationships produced blind spots that undermined systemic safety. Drawing on ethnographic fieldwork and stakeholder interviews, our analysis identified three recurring tensions associated with fragmented accountability processes. These included misalignments between design assumptions and operational realities of system capabilities, competing interpretations of professional expertise and operator roles, and gaps between training provision and operational demands. Using a relational accountability perspective, we showed that these tensions did not arise from individual failures but from patterned distributions of epistemic authority, responsibility, and consequence across stakeholders. Our findings extend beyond maritime autonomy. The authority inversion identified in our study, in which operators retain legal liability while automated systems exercise practical control, is characteristic of safety-critical autonomous systems more broadly. By analysing accountability as a relational system rather than a set of isolated mechanisms, our paper demonstrates how such blind spots are produced and sustained. Our findings also demonstrate that investigating accountability in autonomous systems requires tracing actors across multiple simultaneous accountability relationships from their perspective, examining relationship dynamics across lifecycle phases, and identifying power asymmetries in claims around safety or failure. Future governance frameworks should re-align authority with responsibility, support the integration of distributed safety-relevant knowledge, and ensure that operators' operational expertise meaningfully informs system design and deployment.

8 Ethics Statement

This study received approval from our institutional review board. Prior written consent was obtained from all participants before interviews were recorded, transcribed, and anonymised for analysis, with identifying details such as names, organisational affiliations, and project-specific information removed. Access to participant contact information was restricted to members of the research team directly involved in data collection and analysis. Participants were also given the option to exclude their quotes from any research outputs. To protect participant privacy and organisational confidentiality, we do not report information related to participants' organisational affiliations.

9 GenAI Usage Statement

This paper has used generative AI to organise the findings section in some places as well as for improving grammar and sentence construction.

10 Acknowledgement

This work was supported by the UK AI Security Institute under grant number UKRI841. Bentley's time was also supported in part by Responsible AI UK grant number EP/Y009800/1.

References

[1] Stanley Baiman. 1990. Agency research in managerial accounting: A second look. *Accounting, organizations and society* 15, 4 (1990), 341–371.
[2] Jawahar Bhalla, Stephen C Cook, and David J Harvey. 2023. Towards a systems framework for the assurance of maritime autonomous systems. *Australian Journal of Multi-Disciplinary Engineering* 19, 1 (2023), 89–108.
[3] Mark Bovens. 2007. Analysing and assessing accountability: A conceptual framework 1. *European law journal* 13, 4 (2007), 447–468.
[4] Marie Davidsen Buhl, Gaurav Sett, Leonie Koessler, Jonas Schuett, and Markus Anderljung. 2024. Safety cases for frontier AI. *arXiv preprint arXiv:2410.21572* (2024).
[5] Pascale Carayon, Peter Hancock, Nancy Leveson, Ian Noy, Laerte Sznelwar, and Geert Van Hootegem. 2015. Advancing a sociotechnical systems approach to workplace safety–developing the conceptual framework. *Ergonomics* 58, 4 (2015), 548–564.
[6] Kathy Charmaz. 2006. *Constructing grounded theory: A practical guide through qualitative analysis*. sage.
[7] Adele E Clarke, Carrie Friese, and Rachel S Washburn. 2017. *Situational analysis: Grounded theory after the interpretive turn*. Sage publications.
[8] Jennifer Cobbe, Michael Veale, and Jatinder Singh. 2023. Understanding accountability in algorithmic supply chains. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1186–1197.
[9] Defense Innovation Unit. 2024. U.S. Navy Selects Vendors for Unmanned Undersea Vehicle Program. https://www.diu.mil/latest/u-s-navy-selects-vendors-for-unmanned-undersea-vehicle-program. Accessed: 2026-01-13.
[10] Det Norske Veritas (DNV). 2025. Autonomous and Remotely Operated Ships. https://www.dnv.com/maritime/autonomous-remotely-operated-ships/. https://www.dnv.com/maritime/autonomous-remotely-operated-ships/ Accessed: 2026-01-04.
[11] Patrizia Di Tullio, Matteo La Torre, Michele Antonio Rea, James Guthrie, and John Dumay. 2024. Beyond the planetary boundaries: exploring pluralistic accountability in the new space age. *Accounting, Auditing & Accountability Journal* 37, 5 (2024), 1283–1311.
[12] DNV. 2025. Maritime Safety Trends 2014–2024: Preparing for Future Risks. https://www.dnv.com/maritime/publications/maritime-safety-report-2014-2024-download/ Accessed: 2026-01-13.
[13] Joseph Donia. 2022. Normative Logics of Algorithmic Accountability.. In *FAccT*. 598.
[14] Frank E Dowrick. 1954. The relationship of principal and agent. *The Modern Law Review* 17, 1 (1954), 24–40.
[15] Kathleen M Eisenhardt. 1989. Agency theory: An assessment and review. *Academy of management review* 14, 1 (1989), 57–74.
[16] Oliver D Hart and Bengt Holmstrm. 1986. The theory of contracts. (1986).
[17] Dan Hendrycks, Mantas Mazeika, and Thomas Woodside. 2023. An overview of catastrophic AI risks. *arXiv preprint arXiv:2306.12001* (2023).
[18] International Maritime Organization. 2018. IMO Takes First Steps to Address Autonomous Ships. https://www.imo.org/en/mediacentre/pressbriefings/pages/08-msc-99-mass-scoping.aspx. Accessed: 2026-01-13.
[19] International Maritime Organization (IMO). 2021. *Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS)*. Technical Report MSC.1/Circ.1638. International Maritime Organization. https://wwwcdn.imo.org/localresources/en/MediaCentre/PressBriefings/Documents/MSC.1-Circ.1638-OutcomeOfTheRegulatoryScopingExerciseForTheUseOfMaritimeAutonomousSurfaceShips...(Secretariat).pdf Accessed: 2026-01-04.
[20] International Maritime Organization (IMO). 2023. Developing a Regulatory Framework for Autonomous Shipping. https://www.imo.org/en/mediacentre/pages/whatsnew-1872.aspx. Accessed: 2026-01-04.
[21] Alecia Y Jackson and Lisa A Mazzei. 2022. *Thinking with theory in qualitative research*. Routledge.
[22] Michael C Jensen and William H Meckling. 2019. Theory of the firm: Managerial behavior, agency costs and ownership structure. In *Corporate governance*. Gower, 77–132.
[23] Kenneth Koford and Mark Penno. 1992. Accounting, principal-agent theory, and self-interested behavior. *The Ruffin Series in Business Ethics* (1992), 127–142.

[24] Elisavet Kozyri, Fred B Schneider, and Stephen Chong. 2025. Accountability, Involvement, and Mediation for Information Flow. In *2025 IEEE 38th Computer Security Foundations Symposium (CSF)*. IEEE Computer Society, 13–13.
[25] Nancy G Leveson. 2016. *Engineering a safer world: Systems thinking applied to safety*. The MIT Press.
[26] Yvonna S Lincoln and Egon G Guba. 1985. *Naturalistic inquiry*. sage.
[27] Scott N MacKinnon, Yemao Man, Monica Lundh, and Thomas Porathe. 2015. Command and control of unmanned vessels: Keeping shore based operators in-the-loop. In *18th International Conference on Ships and Shipping Research, NAV*, Vol. 2015. 612–619.
[28] Gennaro Maione, Giulia Leoni, and Michela Magliacani. 2025. Unpacking the knowledge dimensions of digital innovation: implications for accountability in public and private sectors during extraordinary times. *Journal of Knowledge Management* 29, 10 (2025), 3145–3165.
[29] Chara Makri, Didem Gürdür Broo, and Andy Neely. 2022. Human-in-loop decision-making and autonomy: Lessons learnt from the aviation industry transferred to cyber-physical systems. *Technologies* 10, 6 (2022), 120.
[30] Maritime and Coastguard Agency. 2025. *Remote Operator Training and Certification Pilot Framework*. Technical Report. Maritime and Coastguard Agency. https://www.gov.uk/government/publications/remote-operator-training-and-certification-pilot-framework Accessed: 2026-04-25.
[31] Martin Messner. 2009. The limits of accountability. *Accounting, Organizations and Society* 34, 8 (2009), 918–938.
[32] Jacob Metcalf, Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, and Madeleine Clare Elish. 2021. Algorithmic impact assessments and accountability: The co-construction of impacts. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 735–746.
[33] Roderick Munday. 2010. *Agency: Law and principles*. Oxford University Press, USA.
[34] Tom Arne Pedersen, Jon Arne Glomsrud, Else-Line Ruud, Aleksander Simonsen, Jarle Sandrib, and Bjørn-Olav Holtung Eriksen. 2020. Towards simulation-based verification of autonomous navigation systems. *Safety Science* 129 (2020), 104799.
[35] Marilia Abilio Ramos, Ingrid Bouwer Utne, and Ali Mosleh. 2019. Collision avoidance on maritime autonomous surface ships: Operators' tasks and human failure events. *Safety science* 116 (2019), 33–44.
[36] John Roberts. 1991. The possibilities of accountability. *Accounting, organizations and society* 16, 4 (1991), 355–368.
[37] Barbara S Romzek and Melvin J Dubnick. 2018. Accountability in the public sector: Lessons from the Challenger tragedy. In *Democracy, bureaucracy, and the study of administration*. Routledge, 182–204.
[38] Joseph H Saleh, Karen B Marais, Efstathios Bakolas, and Raghvendra V Cowlagi. 2010. Highlights from the literature on accident causation and system safety: Review of major ideas, recent contributions, and challenges. *Reliability Engineering & System Safety* 95, 11 (2010), 1105–1116.
[39] Susan P Shapiro. 2005. Agency theory. *Annu. Rev. Sociol.* 31, 1 (2005), 263–284.
[40] Ileana Steccolini. 2025. From public accountability to accountee-ability: potential and challenges. *Accounting, Auditing & Accountability Journal* (2025), 1–28.
[41] Iddo Tavory and Stefan Timmermans. 2022. *Abductive analysis: Theorizing qualitative research*. University of Chicago press.
[42] The Maritime Executive. 2024. Autonomy-Equipped Barge Collides With Vessel Near Rotterdam. https://maritime-executive.com/article/containers-spilled-in-apparent-navigation-incident-with-autonomous-barge. Accessed: 2026-01-13.
[43] Jennifer Tridgell and Jatinder Singh. 2025. 'Stewardship'as a Fair, Accountable and Transparent Model for Free and Open-Source Software Governance? Looking Beyond the EU's Cyber Resilience Act. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. 473–484.
[44] UN Conference on Trade and Development. 2022. Maritime Autonomous Surface Ships: A Critical 'MASS' for Legislative Review. https://unctad.org/news/transport-newsletter-article-no-97-fourth-quarter-2022. Accessed: 2026-01-13.
[45] Erik Veitch and Ole Andreas Alsos. 2022. A systematic review of human-AI interaction in autonomous ship systems. *Safety science* 152 (2022), 105778.
[46] Martyn Wingrove. 2025. UK Marine Accident Unit to Investigate Collision Between CTV and USV. https://www.rivieramm.com/news-content-hub/news-content-hub/uk-government-starts-investigating-ctv-autonomous-vessel-collision-86456. Accessed: 2026-01-13.
[47] Krzysztof Wróbel. 2021. Searching for the origins of the myth: 80% human error impact on maritime safety. *Reliability Engineering & System Safety* 216 (2021), 107942.

A Analytical Trajectory

Table 5. Code Evolution for RQ1: Tensions Between Design Assumptions and Operational Realities

Example Data Excerpts	Initial Codes	Focused Codes	Findings	Discussion Theory
DES1: "The system able to call up, call home and basically say, you know, this is far too complex for me to figure out. We work under the premise that they aren't all going off at the same time." OP1: "They break constantly. You kick them off the dock, they will break constantly." OP3: "Vessel went up on the riverbank due to outdated software."	Designer premises away simultaneous failures; Operator reports constant breakdowns; Vessel runs aground due to coordination complexity	Epistemic authority enables premising away problems; Operational feedback contradicts design premises; Unaddressed blindspots create operator vulnerability	System Capabilities and Reliability (3.2.1)	Relational Production of Blindspots Through Epistemic Asymmetries (4.1)
DES1: "Operators on watch, but potentially doing their day job alongside it." OP1: "At the moment we almost need superhumans to deal with it." OP4: "Unicorn people in industry."	Designer assumes operators can multitask; Operator needs superhuman capabilities; Master mariners struggle without tech background	Constituting operators as infinitely divisible attention; Demands for exceptional hybrid competencies; Structural competence gaps	The Operator Role (3.2.2)	Authority Inversion: Accountability Without Control (4.2)
OP1: "Manufacturer training tends to just be if you push this button, this happens." OP8: "Largely just an e-learning module style training on how to use the autonomous vessel."	Training teaches button pushing; Procedural knowledge prioritised; Recipients lack foundational understanding	Training as legitimation not competency transfer; Misalignment between instruction and operational demands	Training and Skills (3.2.3)	Epistemic asymmetries devalue operator experiential knowledge

Table 6. Code Evolution for RQ2: How Different Structural Positions Produce These Tensions

Example Data Excerpts	Initial Codes	Focused Codes	Findings	Discussion Theory
DES2: "Paper based designs and conceptual modelling." OP5: "Little telltale signs like cavitation, RPM fluctuations. If it doesn't look right, take control and get a feel." OP1: "You can smell it, you can hear it, you can feel it moving."	Designer works in simulations; Operator develops tacit vessel knowledge; Remote operation loses sensory cues	Design contexts surface computational challenges; Operational contexts surface sociotechnical failures; Physical presence enables tacit diagnostic work	Situated Knowledge and Epistemic Authority (3.3.1)	Epistemic Asymmetries as Power Relations (4.1)
OP1: "We can't tell clients why systems failed because we don't understand the autonomous vessels well enough ourselves." OP3: "Constant supervision required – they withheld this until after you've bought it." OP7: "Only works with all of L3 cameras and L3 this and L3 that. We developed workarounds in secret."	Operator can't explain failures; Vendor withholds limitations until post-purchase; Proprietary architecture forces workarounds	Systematic information asymmetry; Critical information concealed during procurement; Closed systems prevent independent verification	Access to Technical Knowledge (3.3.2)	Proprietary Boundaries Prevent Accountability (4.2)
OP5: "Seafarers love the sea. Technologists probably never seen the sea. If we can keep the culture, autonomising things is not going to be a problem. But if you lose the culture with it, then things are going to happen." OP5: "People who have just thrown a system together and deploy it and don't care about the pre planning, the pre deployment, the safety behind it."	Maritime culture emphasises safety vigilance; Technologists entering without seafaring; Formal credentials required despite experience	Safety culture as epistemic infrastructure; Displacement of embodied maritime knowledge; Institutional gatekeeping mechanisms	Maritime Culture and Changing Expertise (3.3.3)	Cultural shift threatens foundational risk awareness

Table 7. Code Evolution for RQ3: How Tensions Distribute Risk and Responsibility

Example Data Excerpts	Initial Codes	Focused Codes	Findings	Discussion Theory
OP5: "The maritime industry still has a very big blame culture. There has to be someone to blame. There has to be someone to either pay a fine or do prison time." OP5: "I could be sat at home with my family and then next thing I know is I'm going to prison because I've collided with an oil tanker and killed 5 people."	Maritime culture demands someone to blame; Master mariner liable even when system decides; Liability persists beyond capacity to control	Institutionalised individual accountability; Mismatch between authority and accountability; Temporal extension of accountability	Maritime Blame Culture and Individual Liability (3.4.1)	Systematic Misallocation of Accountability (4.3)
OP5: "How do you send an autonomous system to prison for killing five people because it went off course?" OP8: "Early in career the computer was telling me what was happening and I was making decisions. By career's end the computers were keeping the vessel within 10 centimetres of a spot." OP3: "As long as the human still has full accountability but they're relying on a system to make decisions, I don't think they're ever fully going to trust in that system."	Cannot imprison autonomous system; Authority shifted to computers over career; Operator must trust while remaining liable	Accountability mechanisms require human sanctions; Gradual authority inversion; Impossible trust without authority alignment	The Problem of Autonomous Accountability (3.4.2)	Distributed Decision-Making Meets Individual Liability (4.2)
OP1: "There's no real certification for these autonomous vessels. Companies say trust us, it works." OP2: "Everyone is waiting for someone else to write the rules." DES1: "We'd follow regulations if they existed, but they don't, so we do our best and hope for the best."	No certification for autonomous vessels; Everyone waiting for someone to regulate; Designer proceeds without external guidance	Governance vacuum enables manufacturer claims; Collective inaction leaves operators vulnerable; Absent standards leave safety to developers	Certification and Governance Gaps (3.4.4)	Fragmented Governance Maintains Misalignment (4.3)