Why AI in maritime safety requires class discipline

Digitalisation, Digitalisation newsletter

March 31, 2026

Artificial intelligence is increasingly influencing how risk is interpreted and managed at sea. When digital systems contribute to collision awareness, trajectory forecasting, or structural monitoring, they transcend their roles as mere efficiency tools. By becoming integral to the vessel's safety architecture, these systems now demand the same independent assurance and discipline accorded to other safety-critical systems.

In clearly defined applications such as fuel optimisation or condition-based maintenance, AI is already delivering measurable benefit. The challenge arises when that confidence is carried into higher-stakes environments. Navigation in congested waters is shaped by ambiguity, sensor inconsistency and unpredictable behaviour from other vessels. These are not edge cases; they are everyday realities of maritime operations.

AI systems are developed, trained and tested within an intended scope of use and defined operating conditions. When vessels trade across different regions, seasons, and traffic densities, those conditions change. Sensors drift out of calibration, AIS feeds become inconsistent, and positioning signals degrade. Because performance usually degrades

gradually rather than failing outright, it can create a deceptive sense of safety that encourages a dangerous over-reliance on automation. Until systems can demonstrate robust performance in rare but serious scenarios, supported by transparent reasoning and reliable fallback behaviour, removing the human from safety–critical decision-making would be premature.

Explainability of AI systems is therefore central to overcoming these limitations. On the bridge, decisions must be understood and justified. A numerical risk score alone provides little operational value. Officers need clarity on what data was considered, how uncertainty was handled, and why a particular risk was flagged. Without that clarity, the outcome is often either overreliance or outright dismissal of the system altogether.

For these reasons, AI systems that influence safety should be governed with the same discipline as traditional safety-critical systems. The maritime industry has long required independent scrutiny of structures, propulsion and fire protection. Digital systems shaping navigational or structural outcomes deserve an equivalent level of stewardship and oversight.

A credible assurance framework for onboard AI should include, at minimum, a document defining the system's purpose, scope, data sources and limitations. Validation should reflect real routes, seasonal conditions, scenario coverage for ambiguous COLREGs situations, restricted visibility, fishing clusters, VTS interactions and operating profiles, not laboratory scenarios alone. Known failure modes including sensor dropouts, AIS spoofing, and degraded GNSS and their corresponding mitigation measures should be recorded alongside performance thresholds and acceptance criteria. Version histories and update records must be traceable, with the ability to revert to a previously validated version where necessary. Cyber protections, including authenticated interfaces and tamper-evident event data, should be foundational. Bridge procedures and crew training records should demonstrate that operators understand the system's capabilities as well as its boundaries.

At IRClass, we establish rigorous governance and evidence criteria to certify that AI-enabled systems meet safety and reliability standards within their defined scope. Our approval process is risk-proportionate, spanning from exhaustive document scrutiny and simulation analysis to on-site commissioning and in-service audits. To ensure resilience, we require systems to implement traceable versioning and rollback, so a previously validated version can be restored quickly if performance drifts. Recognizing that AI is dynamic, our oversight extends beyond initial approval: we require re-assessment for significant updates, changes in the Operating Domain (ODD), or material KPI drift. Where compliance is demonstrated, we issue formal approvals with specific conditions of validity linked to the system's intended use.

Cyber integrity is inseparable from this approval framework. Once AI outputs influence operational decisions, compromised data becomes a practical hazard. Controls such as authenticated time synchronisation, restricted access management, network separation and cross-checking between data sources reduce reliance on any single stream and strengthen resilience in spoof-prone or congested environments.

Structural health monitoring provides a useful parallel. Properly calibrated sensors and validated fatigue analytics can sharpen inspection planning and reduce unexpected downtime. But digital insight must still be reconciled with engineering judgement and physical verification. Data should inform decisions, not automate them without scrutiny.

Artificial intelligence will continue to advance and expand its role in maritime operations. The question is not whether to adopt it, but how to govern it as it enters safety-critical territory.

Independent assurance is not a barrier to innovation; it is the mechanism through which digital capability earns sustained trust at sea.

By Anil Kumar Korupoju, Senior Surveyor, Indian Register of Shipping (IRClass)

https://cyprusshippingnews.com/2026/03/31/why-ai-in-maritime-safety-requires-class-discipline/