한국항해항만학회지 제46권 제3호
J. Navig. Port Res. Vol. 46, No. 3 : 269-279, June 2022 (ISSN:1598-5725(Print)/ISSN:2093-8470(Online))
DOI : http://dx.doi.org/10.5394/KINPR.2022.46.3.269

해양 소프트웨어 시스템의 인공지능 적용을 위한 안전 고려사항에 관한 분석

이창의* · 김효승** · † 이서정

*,**한국해양대학교 대학원 컴퓨터공학과 박사과정, † 한국해양대학교 기관시스템공학부 교수

Analysis of Safety Considerations for Application of Artificial Intelligence in Marine Software Systems

Changui Lee · Hyoseung Kim** · † Seojeong Lee*

,Ph.D Candidate, Graduate School of National Korea Maritime and Ocean University, Busan 49112, Korea*
† Professor, Division of System Engineering, National Korea Maritime and Ocean University, Busan 49112, Korea

요 약 : 인공지능의 발전으로 산업계 전반에서 시스템의 자동화를 위해 인공지능을 도입하고 있다. 해양산업분야에서도 자율운항선박이라는 패러다임을 통해 인공지능을 단계적으로 적용하고 있다. 이러한 흐름에 따라 ABS와 DNV에서는 자율운항선박에 대한 가이드라인을 발표하였다. 하지만 선급의 가이드라인은 선박의 운항 및 해양 서비스 관점에서 요구사항을 기술하고 있으므로, 인공지능의 위험에 대해서는 충분히 고려되지 못했을 가능성이 있다. 그래서 본 연구에서는 ISO/IEC JTC1/SC42 인공지능 분과에서 제정한 표준들을 활용하여 선급 요구사항을 위험의 원인으로 분류하고, 위험원인과 인공지능 메트릭(metrics)의 조합을 통해 위험을 평가할 수 있는 척도로 사용하고자 한다. 본 연구에서 제안한 인공지능의 위험 원인과 이를 평가하기 위한 특성의 조합을 통해 해양 시스템에 인공지능이 도입됨으로써 발생하는 위험을 정의하고 식별하는 데 도움이 될 수 있을 것으로 생각되며, 선급을 포함한 다양한 기구에서 자율운항선박을 위한 안전 요구사항을 더욱 자세하고 구체적으로 작성하는 데 도움을 줄 수 있을 것으로 기대한다.

핵심용어 : 자율운항선박, 해양 소프트웨어, 안전, 인공지능, 선급 요구사항, 위험 평가, 위험 원인, 메트릭

Abstract : *With the development of artificial intelligence, artificial intelligence is being introduced to automate systems throughout the industry. In the maritime industry, artificial intelligence is being applied step by step, through the paradigm of autonomous ships. In line with this trend, ABS and DNV have published guidelines for autonomous vessels. However, there is a possibility that the risk of artificial intelligence has not been sufficiently considered, as the classification guidelines describe the requirements from the perspective of ship operation and marine service. Thus in this study, using the standards established by the ISO/IEC JTC1/SC42 artificial intelligence division, classification requirements are classified as the causes of risk, and a measure that can evaluate risks through the combination of risk causes and artificial intelligence metrics want to use. Through the combination of the risk causes of artificial intelligence proposed in this study and the characteristics to evaluate them, it is thought that it will be beneficial in defining and identifying the risks arising from the introduction of artificial intelligence into the marine system. It is expected that it will enable the creation of more detailed and specific safety requirements for autonomous ships.*

Key words : *autonomous ship, marine software, safety, artificial intelligence, classification requirements, risk assessment, causes of risk, metrics*

1. 서 론

다양한 산업 분야에 소프트웨어가 사용되면서 소프트웨어에 의한 사고 위협 증가하고 되었다. 경미하거나 사소한 소프트웨어 오류가 사람의 생명을 앗아가거나 막대한 경제적 피해를 초래하는 경우가 많아지고 있다(Lee and Lee, 2021). 시스템이나 장비에 의해 발생하는 사고를 줄이기 위해 시스템과 장비를 생산하는 단계에서 안전과 관련된 기능에 대한 위험을 평가하고 무결성을 확보하는 방안이 세계적으로 다양한 산업분야에서 진행되어 왔다. 전기, 전자, 또는 프로그램이 가능한 전자 (E/E/PE; Electric, Electronic or Programmable Electronic) 시스템 및 제품에 대한 포괄적 기능안전성 표준인 IEC(International Electrotechnical Commission) 61508을 시작으로, 철도, 항공, 에너지, 의료 등 주요 산업분야에서 분야 특

† Corresponding author : 정회원, sjlee@kmou.ac.kr 051)410-4578
* 정회원, myallyou@gmail.com 051)410-4888
** 정회원, khs9962@gmail.com 051)410-4888

성을 고려한 기능안전성 표준을 정의해왔다(NIPA, 2019).

조선해운업계에서는 제4차 산업혁명의 도래와 관련하여, 자율운항선박(MASS; Maritime Autonomous Surface Ship)의 개발 및 운용에 대한 논의가 활발하게 이루어지고 있다. 2017년 6월에 개최된 IMO(International Maritime Organization) 해사안전위원회(MSC; Maritime Safety Committee) 98차 회의에서 자율운항선박을 '다양한 자동화 수준에서 인간의 간섭 없이 독립적으로 운용될 수 있는 선박'으로 정의하였으며, 99차 회의에서 국제적으로 공식 합의에 이르렀다(Kim and Yang, 2019; Lee and Lützhöft, 2020; Jung and Lee, 2020).

인공지능은 인간의 사고 능력, 학습 능력, 추론 능력 등을 재현하기 위해 다양한 SW·HW 기술의 융합으로 구현되는 복합체이며, 자율운항선박에서 지능형 육상서비스 및 시스템의 자동화를 구현하는데 활용되고 있다(Kim et al., 2021). 인공지능은 전통적인 소프트웨어와는 다른 특징들과 개발 프로세스를 가지고 있으며, 이러한 특징은 필연적으로 물리적 손상, 경제적 손해, 신체적 상해를 초래할 수 있으며, 기능적인 오작동과 논리적 오류의 위험성을 내포한다.

본 연구의 2장에서는 인공지능의 안전성 확보를 위한 연구 동향을 살펴보고, 3장에서는 선급의 자율운항선박 가이드라인을 통하여 해양분야에서 인공지능의 위험 원인에 대해 충분히 고려되었는지 확인한다. 그리고, 4장에서는 인공지능의 특성을 고려하여 위험 원인과 성능 평가를 위한 메트릭의 조합을 제안한다.

2. 인공지능 연구 동향

IT(Information Technology) 분야의 국제표준화기구인 ISO(International Organization for Standardization)와 IEC의 합동기술위원회 JTC(Joint Technical Committee)1은 정보기술 분야 표준을 공동 제정하기 위해 1987년 설립된 표준화기관이다. JTC1에서는 2017년 인공지능 총회인 JTC1 SC(Subcommittee) 42 설립을 결정하고 2018년 4월 인공지능의 표준화를 전담하는 SC 42를 신설하였다. SC 42는 개설 당시 기반 표준을 연구하는 1개의 WG(Working Group)과 각각 인공지능 시스템, 신뢰성, 사례 및 응용에 대해 연구하는 3개의 SG(Study Group)

Table 1 ISO/IEC JTC 1/SC 42 : List of working groups and developing status of standards

Working Groups	Developing status of standards
WG 1 (Foundational Standards)	- ISO/IEC 22989(Artificial intelligence concept and terminology) - ISO/IEC 23053(Framework for artificial intelligence system using machine learning)
WG 2 (Data)	- ISO/IEC 20547-1(Big data reference architecture-Part 1:Overview and application process) - ISO/IEC 20547-3(Big data reference architecture-Part 3:Reference architecture) - ISO/IEC 24668(Process Management Framework for Big Data Analytics)
WG 3 (Trustworthiness)	- ISO/IEC TR 24028(Artificial Intelligence-Overview of Trustworthiness in Artificial Intelligence) - ISO/IEC DIS 23894(Artificial Intelligence-Risk Management) - ISO/IEC TR 24027(Artificial Intelligence-Bias in AI systems and AI aided decision making) - ISO/IEC TR 24029-1(Artificial Intelligence-Assessment of the robustness of neural networks, Part 1:Overview) - ISO/IEC TR 24029-2(Artificial Intelligence-Assessment of the robustness of neural networks, Part 2:Methodology for the use of formal methods) - ISO/IEC PRF TR 24368(Artificial Intelligence(AI)-Overview: Aspect of ethical and societal concerns)
WG 4 (Use cases and applications)	- ISO/IEC DTR 24030(Artificial Intelligence-Use cases)
WG 5 (Computational approaches and computational characteristics of AI systems)	- ISO/IEC TR 24372(Artificial Intelligence - Overview of computational approaches and AI systems) - ISO/IEC WD TS 4213(Artificial Intelligence(AI) - Assessment of machine learning classification)
JWG 1 (Governance implications of AI)	- ISO/IEC 38507(Governance of IT - Governance implications of the use of artificial intelligence by organizations)

Source : Trends of International Standards: ISO/IEC JTC 1/SC 42, TTA Journal, 2020

으로 시작하였으며, 현재는 총 5차 총회가 진행되어 5차 총회 결과를 기준으로 Table 1과 같이 11개 연구 그룹으로 확대되어 인공지능의 표준화를 진행하고 있다(IITP, 2021).

2.1 ISO/IEC 22989(인공지능 개념 및 용어)

ISO/IEC 22989는 IT 산업 관점에서 인공지능 시스템을 이해하고 공통된 언어로 향후 표준을 개발하기 위한 기초 표준으로 활용하기 위해 문서가 개발 승인되었다(ISO, 2022a). ISO/IEC 22989 문서는 강한 인공지능(Strong AI)과 약한 인공지능(Weak AI)으로 분류되는 인공지능의 일반적 분류부터 현대 머신러닝 관점에서 지도학습, 비지도학습, 강화학습 방법으로의 분류의 개념들을 제공하는 것부터, 인공지능 시스템, 자동화 시스템, 학습 데이터 및 학습 모델 등 인공지능 분야에서 사용되는 모든 통상 개념과 정의를 제공하고 있다.

또한, ISO/IEC 22989는 인공지능의 신뢰성(Trustworthiness)에 대해 시스템이 이해 관계자의 기대를 충족하는지 확인하는 것으로 정의하고 있으며, 신뢰성을 확인할 수 있는 특성으로 Table 2와 같이 9종의 특성을 제시하고 있다.

Table 2 Characteristics to checking trustworthiness

Name of characteristic	Description
Robustness	The ability to maintain performance level even under external interference or harsh environmental conditions
Reliability	The ability to perform the necessary functions without breakdown during the intended period under a given condition
Resilience	The ability to quickly recover the operation status after the accident
Controllability	The ability to take over control rights by an external agent involved
Explainability	The ability to explain why the AI system made this decision
Predictability	The ability to enable a person who can trust in the results of the AI system
Transparency	The degree of disclosing what data is needed and how it collected and educated the data
Fairness	The degree of discrimination against different groups
Jurisdictional issues	The difference in regulations applied when the operating area of the AI system (jurisdiction) is changed

Source : ISO/IEC 22989, ISO, 2022

2.2 ISO/IEC 23053(머신러닝을 활용한 인공지능 시스템 프레임워크)

ISO/IEC 23053은 현대의 대표적인 인공지능 기술인 머신러닝 기술에 집중한 프레임워크를 제공한다(ISO, 2022b). 또한 라이프 사이클과 유사한 개념인 머신러닝 파이프라인 (Pipe line)을 정의하고, 머신러닝 개발 절차의 예시를 제공한다.

특히, ISO/IEC 23053 문서에서는 머신러닝 업무(Task)의 종류, 머신러닝 알고리즘, 머신러닝 평가 메트릭 등의 기본적 개념들이 제공되며, 머신러닝에서 일반적으로 사용하는 분류 체계인 지도학습, 비지도학습, 강화학습을 기반으로 각 학습 분류의 전형적인 방법(method) 및 접근법(approach)을 소개한다.

ISO/IEC 23053에서 머신러닝을 평가하기 위한 메트릭(metrics)은 Table 3과 같다. 정밀도(Precision)는 Positive로 예측한 경우 중 실제로 Positive인 경우를 말하고 재현율(Recall)은 실제 Positive인 것 중 올바르게 Positive를 맞춘 것의 비율을 말한다. 정확도(Accuracy)는 전체 예측값 중 올바르게 예측한 비율이며 마지막으로 F1 Score는 정밀도와 재현율의 조화평균을 계산한 값이다. 여기서 T_P, F_N, F_P, T_N은 Confusion Matrix의 값들이다.

Table 3 Metrics to validation of machine learning

Name of metric	Description
Precision	The ratio of positive on prediction value by positive $(\mathrm{Precision}) = \frac{T_P}{T_P + F_P}$
Recall	The ratio of correct positive on the actual positive value $(Recall) = \frac{T_P}{T_P + F_N}$
Accuracy	The ratio of correctly predicted on the total prediction value $(Accuracy) = \frac{T_{P+T_N}}{T_P + F_N + F_P + T_N}$
F1 Score	Harmonic mean with precision and recall $F1\,Score = 2 \times \frac{(\mathrm{Precision}) \times (Recall)}{(\mathrm{Precision}) + (Recall)}$

Source : ISO/IEC 23053, ISO, 2022

Confusion Matrix란 지도 학습으로 훈련된 분류 알고리즘의 성능을 시각화할 수 있는 표이며 Fig 1과 같이 그려진다. 행렬의 각 행은 예측된 클래스의 인스턴스를 나타내며 각 열은 실제 클래스의 인스턴스를 나타낸다. 다시 말해 True는 예측값과 실제값이 같음을 의미하며, False는 예측값과 실제값이 다르다는 것을 의미한다. 또, 예측값이 긍정이면 P(Positive)로 표기하고, 예측값이 부정이면 N(Negative)로 표기한다.

Table 4 Risk sources described in ISO/IEC DIS 23894

Name of risk source	Description
Level of automation	Artificial intelligence is often used in automation of the system and can affect the automation stage. In particular, if you need collaboration with people, handovers with people can be a risk factor.
Lack of transparency and explainability	If the information related to the development and learning of the AI system is not transparently disclosed, or if it is not explained so that people can understand the basis for the judgment of artificial intelligence, artificial intelligence will not be trusted.
Complexity of environment	Artificial intelligence is mainly used to handle complex and diverse surrounding environments, and complex environments can cause additional risks compared to simplicity.
System life cycle issues	Artificial intelligence has a different characteristic from the existing system development life cycle, which can cause danger. For example, it can be an inappropriate verification method or process.
System hardware issues	The defect in the hardware or sensor can be interrupted or incorrectly measured by the service. In addition, the lack of system performance or communication bandwidth for artificial intelligence can cause risk.
Technology readiness	There may be a risk if you still use less technically less mature AI algorithms or models for real work.
Risk sources related to machine learning	The development of artificial intelligence is associated with machine learning or deep learning. Risks may occur if the quality or learning process of data required for learning is inappropriate.

Source : ISO/IEC DIS 23894, ISO, 2022

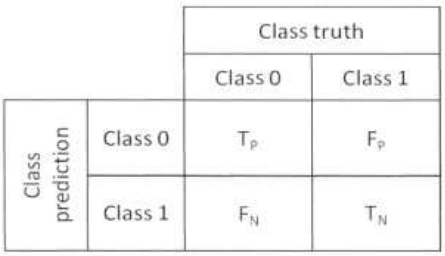

Fig. 1 Example of confusion matrix
Source : ISO/IEC 23053, ISO, 2022

2.3 ISO/IEC DIS 23894(인공지능 위험관리)

ISO/IEC DIS 23894는 인공지능 혹은 지능형 서비스의 개발 및 도입 시 발생 가능한 위험 요소들을 관리하기 위한 표준으로, 인공지능의 위험관리에 관한 일반적인 사항과 프레임워크, 위험관리 절차, 위험관리의 목적과 위험의 원인, 인공지능 생애주기에 따른 위험관리 방법을 제공한다(ISO, 2022c). ISO/IEC DIS 23894는 발생 가능한 위험의 원인을 Table 4와 같이 7가지로 구분하고 있다.

인공지능은 시스템 자동화에 사용되는 경우가 많으며, 자동화의 단계에 따라서 사람과 협업이 필요한 경우가 많다. 그러므로 사람과 시스템간의 협업이 위험 요인이 될 수 있으며, 사람과의 협업에 있어 사람이 인공지능의 학습 과정이나 판단의 근거를 확인할 수 없다면 인공지능을 신뢰할 수 없게 되어 위험이 발생할 수 있다. 또한 복잡하고 다양한 주변환경이 인공지능의 성능에 영향을 미치게 되어 위험이 발생할 수 있다. 다양한 상황을 처리하기 위해 많은 연산이 필요하게 되어 고사양의 하드웨어와 다양한 센서가 필요하지만, 시스템이나 센서의 성능 부족에 의해 위험이 발생할 수 있다. 인공지능의 개발 방법은 기존의 시스템에 대한 개발 방법과 다른 특징을 가지고 있어 기존의 개발 방법을 그대로 적용하면 인공지능 소프트웨어를 충분하게 검증할 수 없어 위험이 발생할 수 있다. 또한, 아직 기술적으로 충분히 성숙되지 않거나 검증되지 않은 인공지능 알고리즘을 성급하게 사용하여 위험이 발생할 수도 있다. 그리고 인공지능의 성능은 학습에 의해 많이 달라질 수 있는데, 학습에 필요한 데이터 품질이 부족하거나 학습과정이 적절하지 않을 경우에 위험이 발생할 수 있다.

3. 선급의 자율운항선박 가이드라인과 인공지능 위험 원인 분류

IMO의 자율운항선박 이행계획에 따라 산업계에서는 다양한 연구가 진행되고 있으며, 선급에서도 자율운항선박에 대한 가이드라인을 제시하고 이를 준수하는 시스템에 관련된 인증을 부여하는 방식으로 자율운항선박의 기술을 관리하고 있다.

DNV(Det Norske Veritas)에서는 2018년 9월에 자율 및 원격 운항 선박 가이드라인을 발표하였으며, 2021년 9월에 개정되었다(DNV·GL, 2018). 그리고 ABS(American Bureau of Shipping)에서는 2021년 7월에 자율 및 원격제어 기능에 대한 가이드라인(Guide for Autonomous and Remote Control Functions)을 발표했다(ABS, 2021). ABS와 DNV의 가이드라

Table 5 ABS requirements categorized by ISO/IEC DIS 23894 risk source

Name of risk source	ABS Requirements	Reference
Level of automation	2.1.2 Operator and Operations Supervision Level An operator is to be designated and will have responsibility over the Autonomous Function. The operator may be physically located onboard the vessel or in a remote location. The operator station is to be constantly manned. i) The operator is to supervise the function executions either continuously, periodically or as needed ii) The operator is to be able to intervene, override, and take over the operation when deemed necessary by the operator	Section 5 chapter 2
	2.5.2 Possibility of Retaking Control It is to be possible for the Operator to intervene and regain control of the action from the autonomous function at all times.	
	3.4 Final Integration and Onboard Test Manual Control (for autonomous function) The operation of manual control takeover using human interface systems and controls onboard is to be confirmed to be functioning satisfactorily. Manual Control (for remote control systems) The operation of manual controls at the remote controlling station using human interface systems and controls is to be confirmed to be functioning satisfactorily.	Section 7 chapter 3
Lack of transparency and explainability	N/A	-
Complexity of environment	2.5.3 Visual Awareness The operator is to have line-of-sight view of the operations being controlled by the autonomous function. Alternatively, live visual feed is to be provided at the operator control station. In case of partial or full failure of video feeds, the operator is to have demonstrably effective backup operational capabilities for situational awareness and decision support.	Section 5 chapter 2
	3.5 System-of-Systems Test for Function A plan for final tests of the function to prove its essential features is to be submitted. The tests proposed in this level are to achieve the following objectives: i) They are to demonstrate the successful integration of all constituent systems necessary for the performance of the Function. ii) They are to demonstrate the successful performance of the Function in its intended operational environment or a simulated environment as close as possible to its intended operational environment. iii) They are to validate all functional scenarios as defined in the Concept of Operations.	Section 7 chapter 3

	iv) Tests are to be carried out to identify unintended effects or emergent behavior resulting from the interactions among the various constituent systems. v) Tests are to demonstrate the effects of system casualties, proper failover procedures, and low probability - high impact failures that can affect crew (if any), vessel, or environment.	
	4.5 Integration Simulation Testing ii) Functional and failure testing can be demonstrated by simulation tests. The results of any required failure analysis are to be observed.	Section 7 chapter 4
System life cycle issues	4.1 General The implementation process for autonomous and remote control functions follows the V-Model Implementation Process seen in Section2/Figure 3, which is an expansion of the standard systems engineering system development model. This model covers the life cycle of the system-of-systems from concept to the operations and maintenance phase. This model can be utilized for the implementation of a new function and also for modification to an existing function.	Section 2 chapter 4
	3.6 Remote Operator Station / Remote Control Station i) Integration and installation of the systems and components at the Remote Operator Station or Remote Control Station has been carried out in accordance with the approved drawings	Section 7 chapter 3
System hardware issues	2.4 Risk Assessment The risk assessment(s) are to show that the vessel is not to descend into an uncontrollable situation in the event of the following: failure of the function system-of-systems impact on the vessel impact of failure or other event on the function occurrence of foreseeable hazard	Section 5 chapter 2
	3.4 Final Integration and Onboard Test Verification of the autonomous or remote control function automatic resumption following a simulated blackout (if appropriate to the assigned risk level). Means of communication between onboard operator station(s) and Remote Operator Station or Remote Control Station (where applicable) are to be tested and confirmed to be operating satisfactorily.	Section 7 chapter 3
Technology readiness	N/A	
Risk sources related to machine learning	3.5.3 Requirements iii) Data Exploration: For model training, data exploration is to be conducted before data quality assessment and data pre-processing to understand the distribution of the key model parameters, correlation between model input parameters and output parameters.	Section 5 chapter 3

Table 6 DNV requirements categorized by ISO/IEC DIS 23894 risk source

Name of risk source	Requirements	Reference
Level of automation	1.2 Extent of automation and support from personnel on board automatic support (AS) Operation of the vessel function by automation systems and personnel in combination. Automation system(s) may partly or fully perform data acquisition, interpretation and decision. This mode is a collective term for all variants of decision support where the automatic support function may need complementary human sensing, interpretation or decision-making and where the action is not automatically effectuated.	Section 5 chapter 1
	2.3 Local/manual actions 2.3.2 Autoremote vessels For autoremote vessels, it is generally not considered feasible to mitigate effects of failures and incidents by manual actions performed on board. 2.3.3 Automatic Operation (AO) Even if conventional manual operations on board will be replaced by purely automation systems, capabilities for remote supervision and emergency control should be arranged in the RCC. 2.3.4 Automatic Support (AS) If conventional manual operation on board will be performed by the remote engineering watch in RCC, decision support functions should be arranged which provide a firm basis for making decisions and executing control actions.	Section 5 chapter 2
Lack of transparency and explainability	N/A	-
Complexity of environment	3.1.1 Proper lookout Maintaining a continuous state of vigilance by sight and hearing, as well as detection of significant change in the operating environment. Fully appraising the situation and the risk of collision, grounding and other dangers to navigation.	Section 4 chapter 3
	4.2 Autoremote vessels The remote navigator will also need to analyse the complete navigational situation, i.e. consider the hazards in relation to other factors that may affect the further navigation planning, such as location, movements and type of a hazard, other potential hazards in the surroundings, the risk of grounding, the weather conditions and sea states, and the own vessel's operational mode and capabilities.	Section 4 chapter 4
System life cycle issues	4.3.1.1 Define the software life-cycle The technology developer can choose between several available standards DNV-CP-0507 System and software engineering ISO/IEC 12207 Systems and Software engineering - Software life-cycle processes ISO/IEC 15288 Systems and Software engineering - System life-cycle processes .	Section 3 chapter 4 Technology qualification process

	9 Design principles 1) Maintain a safe state. No incidents, including fire and flooding on board or in the remote control centre, or single failure in systems on board or systems interfacing the vessel, should cause an unsafe mode for the vessel or its surrounding environment. It should be possible to enter and maintain a minimum risk condition (MRC) in all operations and scenarios defined in the document concept of operation. Considering that different minimum risk conditions may apply in the various operational phases/modes, the design should be based on all defined MRCs.	Section 2 chapter 9
System hardware issues	2.2.4 Restoration of functions It should in general be possible to restore a key vessel function from the RCC without assistance by personnel on board. Depending on the failure or incident causing stop of the function, the restored function may have reduced capacity. For vessels with personnel on board, local/manual restoration by on-board crew may be relied upon if adequate competence, instructions or assistance by RCC is available.	Section 5 chapter 2
Technology readiness	4.3.2.3 Define performance specification For the software components of a system, the ISO/IEC 25000 series of standards give valuable input for defining performance parameters. In particular ISO/IEC 25010 gives information about potential characteristics (quality attributes) for a software component, covering characteristics both the software itself and the use of software.	Section 3 chapter 4 Technology qualification process
	3.2.3 Performance parameters for object detection systems When an object detection system is intended to be used in a concept to replace the look-out function on board, the needed performance of the system to obtain an equivalent or better object detection capability should be determined as part of the concept process described in Sec.3 [2].	Section 4 chapter 3
Risk sources related to machine learning	N/A	-

인에서는 자율 또는 원격에서 운영되는 선박을 개발하고 운영하는데 필요한 절차와 기능적 요구사항들에 대해 기술되어 있다. 이 장에서는 ABS와 DNV의 가이드라인을 분석하여 2.3절의 7종의 위험 원인으로 분류하였다.

3.1 ABS의 자율 및 원격제어 기능에 대한 가이드라인

ABS의 가이드라인은 선박 및 해양 구조물에서 자율 및 원격 제어 기능을 구현하기 위한 목표기반 프레임워크를 설정하고, 4단계(모니터링, 분석, 결정, 조치)의 의사결정 루프에 따라 수행해야 하는 기능과 이해관계자와의 상호작용을 다루고 있다. 특히 이 가이드라인은 위험기반 접근 방식을 사용하여 자율 및 원격 제어 기능의 평가 및 구현에 대한 요구사항을 기술하고 있다. ABS의 요구사항을 ISO/IEC DIS 23894의 위험 원인으로 분류하면 Table 5와 같다. ABS에서는 자율운항선박이 사람의 통제하에서 언제든지 사람이 제어권을 획득할 수 있도록 하고 이를 시운전 시에 확인하도록 요구하고 있다. 또한, 복잡한 환경으로 인해 발생하는 상황 인식에 대한 검증과 시스템과 그 시스템을 구성하는 서브 시스템 간의 테스트를 요구하고 있다. 또한, 시스템 개발시에 V모델의 시스템 생명주기를 준수하기를 요구하고 있으며, 하드웨어적 장애나 실패에 따른 위험을 평가하고 시운전 시에 테스트하도록 요구하고 있다. 학습 모델에 대해서는 데이터 품질을 평가하고, 데이터의 상관관계를 확인하도록 요구하고 있다.

3.2 DNV의 자율 및 원격 운항 선박 가이드라인

DNV의 가이드라인은 기존의 규정으로는 맞지 않는 새로운 운영 개념과 인간이 수행하는 기능에 대한 관리를 다루고 있다. 이 가이드라인은 항해, 선박 엔지니어링, 원격 제어 센터

및 통신에 대한 요구사항을 정의하고 있으며, 특히 자율 및 원격 운영에서 중요한 사이버 보안 및 소프트웨어 테스트에 중점을 두고 있다. DNV의 요구사항을 ISO/IEC DIS 23894의 위험 원인으로 분류하면 Table 6과 같다. DNV에서는 자동화의 단계에 따라 사람의 개입 정도와 제어권을 행사하는 방법에 대해 요구하고 있으며, 복잡한 환경으로 인해 선박에서는 시각과 청각을 이용하여 적절히 감시하도록 하고 원격에서는 항해 상황을 완전하게 분석할 수 있도록 요구하고 있다. 소프트웨어의 수명주기 관리를 위해 ISO/IEC 12207 등의 표준을 준수하도록 요구하고 있으며, 시스템을 디자인할 때 어떠한 사고가 발생하더라도 최소위험조건(MRC)을 유지할 수 있도록 해야 한다고 요구하고 있다. 또한 하드웨어적 고장이나 기능의 중지가 발생하더라도 기능의 복원이 가능해야 한다고 요구하고 있으며, 소프트웨어 시스템의 특성을 평가하기 위해 ISO/IEC 25000 시리즈 표준을 참조하도록 하고, 물체 감시를 위한 시스템 성능을 요구하고 있다.

3.3 선급 가이드라인에 대한 인공지능 위험 원인 분류 결과

위험 원인에 대해 ABS와 DNV에서 요구하는 사항을 비교하면 Table 7과 같다. ○는 해당 위험 원인에 대해 선급의 요구사항이 충분히 기술되었음을 의미하며, ×는 요구사항이 기술되지 않았음을 말한다. △는 기술이 되어 있기는 하나, 인공지능의 특성을 충분히 고려하지 못하여 보완이 필요함을 의미한다. 두 선급의 가이드라인에 대해 위험 원인을 분류하고 비교한 결과를 살펴보면 다음과 같다.

Table 7 Comparison of requirements between ABS and DNV

Name of risk source	ABS document	DNV document
Level of automation	○	○
Lack of transparency and explainability	×	×
Complexity of environment	○	○
System life cycle issues	△	△
System hardware issues	○	○
Technology readiness	×	△
Risk sources related to machine learning	△	×

(1) Level of automation은 ABS와 DNV 모두 자율운항 단계에 따라서 사람과 시스템이 해야 할 역할이나 권한에 대한 요구사항이 잘 명시되어 있다.

(2) Lack of transparency and explainability 는 ABS와 DNV 모두 요구사항을 찾을 수 없다. 선급의 인증은 표현되는 결과 평가하는 것으로, 선급의 인증범위에서 벗어난 것으로 보아 요구사항이 기술되지 않은 것으로 판단된다. 하지만, 투명성과 설명 가능성의 결여는 선박의 안전에 영향을 줄 수 있으므로 이에 대한 고려가 필요하다.

(3) Complexity of environment는 해양환경의 특성과 상황 인식 및 판단에 대한 어려움, 규정에서 요구하는 복잡한 상황들에 대해서 가장 많은 요구사항을 기술하고 있다.

(4) System life cycle issues에 대해서 ABS에서는 V모델을 사용하고 이를 상황에 따라 변경할 수 있다고 하였으며, DNV에서는 ISO/IEC 12207과 같은 전통적인 개발 프로세스를 따르도록 요구하고 있다. 인공지능의 개발 프로세스는 전통적인 개발 프로세스와는 다른 특징이 있으므로 이에 대한 고려가 필요하다.

(5) System hardware issues에서는 ABS와 DNV 모두 시스템에 대한 전력손실이나 화재와 같은 하드웨어적 실패에 대해 중요하게 다루고 있는 부분으로 특히 DNV에서는 통신과 관련한 오류에 대한 요구사항을 중요하고 다루고 있다.

(6) Technology readiness는 ABS에서는 언급하고 있지 않으며, DNV에서는 ISO/IEC 25000 시리즈의 품질측정 매트릭을 따른다고 되어 있다. 하지만, ISO/IEC 25000 시리즈는 전통적인 소프트웨어에 대한 품질특성을 요구하고 있어 인공지능에 대한 품질특성들은 고려하지 못하고 있으므로 이에 대한 고려가 필요하다.

(7) Risk sources related to machine learning에서는 DNV에서는 요구사항이 정의되어 있지 않으며, ABS에서는 일반적인 데이터 품질을 요구하고 있다. 인공지능 학습을 위한 데이터 품질은 공정성과 같은 전통적인 데이터 품질과는 다른 특성이 있으므로 이를 보완할 필요가 있다.

4. 위험 원인과 연구사례 매트릭스

3장에서 ISO/IEC 238994의 7종의 위험 원인을 ABS와 DNV의 가이드라인을 통해 분석해 보았다. 선급의 요구사항은 사람과의 핸드오버에 의한 사고나 외부환경 또는 하드웨어적 실패에 따른 사고와 같이 안전상의 문제가 명시적으로 드러나는 실패에 대해서 규정하고 있다.

하지만, 인공지능의 개발 프로세스나 인공지능 성능 및 특성에 대한 부분은 ABS와 DNV의 가이드라인에서 다루고 있지 않으며, ISO/IEC 12207과 ISO/IEC 25000 시리즈를 참조하도록 하였으나 인공지능을 위한 별도의 요구사항은 존재하지 않고 있다. 이 논문에서는 3장에서 분석한 내용을 토대로 인공지능의 특성을 고려하여 위험 원인과 성능 평가를 위한 메트릭의 조합을 제안하고자 한다.

Table 9는 ISO/IEC DIS 23894에서 제안한 7종의 위험 원인과 ISO/IEC 22989에서 제시한 9종의 신뢰성을 확인할 수 있는 특성과 ISO/IEC 23053에서 제시한 머신러닝을 평가하기 위한 메트릭(metrics)을 조합한 결과이다. 하나의 위험 원인은 하나 이상의 위험 평가 메트릭(metrics)을 가질 수 있으며, 하나의 위험 평가 메트릭(metric)은 하나 이상의 위험 원인에 대응될 수 있다. 즉, 위험 원인과 위험 평가 메트릭의 관계는 다대다

(N:N)의 관계를 가진다.

3장에서 ABS와 DNV의 가이드라인을 통해 위험의 원인을 살펴보았듯이 자동화 단계(Level of automation)에 따라 사람과 인공지능 시스템간 제어권 이양에 관한 요구사항이 있었다. 제어권 이양을 위해서는 사람이 시스템을 신뢰할 수 있어야 하고 해당 지역에서 제어권 이양이 법적으로 허용할 수 있는지를 판단해야 한다. 그러므로 ISO/IEC 23053에서 외부 에이전트(사람 또는 시스템)에 의해 제어권한을 인수하는 기능인 Controllability와 인공지능 시스템이 결정을 내린 이유를 설명하는 능력인 Explainability, 인공지능 시스템의 관할지역의 변경시 적용되는 규정의 차이를 확인하는 Jurisdictional issues 메트릭(metrics)을 선정하였다.

Table 8 Combination metrics categorized by risk sources

Name of risk source	Combination metics
Level of automation	Controllability Explainability Jurisdictional issues
Lack of transparency and explainability	Explainability Predictability Transparency
Complexity of environment	Robustness Reliability Resilience
System life cycle issues	Explainability Predictability Transparency Reliability
System hardware issues	Robustness Reliability Resilience
Technology readiness	Precision Recall Accuracy F1 score
Risk sources related to machine learning	Predictability Fairness

5. 결론 및 향후연구

ABS와 DNV의 가이드라인에서는 선박의 운항 및 해양 서비스 관점에서 자율운항단계에 따른 역할의 변화와 외부환경 및 하드웨어적 실패에 대한 요구사항을 명시하고 있다. 하지만, 본 연구에서 분석한 결과와 같이 자율운항선박에 있어서 중요한 역할을 수행하게 될 인공지능의 성능과 신뢰성 확인을 위한 요구사항은 별도로 명시되지 않았거나 인공지능의 특성에 대한 고려 없이 기존의 방식을 적용하도록 요구하고 있다.

그래서 본 연구에서는 ISO/IEC JTC1/SC42 인공지능 분과에서 개발 중이거나 개발이 완료된 표준을 참조하여 선급 가이드라인에 적용해 인공지능의 위험 원인과 이를 평가하기 위한 메트릭의 조합을 제안하였다. 인공지능의 위험 원인과 이를 평가하기 위한 메트릭의 조합은 선급의 가이드라인이 변경되거나 분석하고자 하는 시스템이 달라질 경우 다소 변경 될 수는 있다. 하지만, 인공지능의 특성을 고려하여 위험의 원인을 파악하고 평가하는 것은 자율운항선박의 안전을 위해서 매우 중요할 것이라 생각된다.

특히 인공지능은 내부를 확인할 수 없는 블랙박스로 취급되기 때문에 그 내부를 알 수 없고, 그 내부를 알 수 없다는 이유로 사람이 인공지능의 판단을 신뢰할 수 없게 된다. 자율운항선박이 단계적으로 이루어지면서 사람과 인공지능 시스템이 서로 협업하고, 나아가서는 인공지능 시스템 스스로 운영하기 위해서는 인공지능에 대한 사람의 신뢰가 확보되어야 한다. 앞서 3장에서 살펴본 바와 같이 인공지능의 투명성과 설명 가능성은 ABS와 DNV 모두 고려하고 있지 않으므로 이에 대한 고려가 절실히 필요하다.

본 연구에서 제안한 인공지능의 위험 원인과 이를 평가하기 위한 특성의 조합을 활용하여 해양 시스템에 인공지능이 도입됨으로써 발생하는 위험성을 분석하고 평가하고자 할 때 위험을 정의하고 식별하는 데 도움이 될 수 있을 것이다. 특히, HAZOP 기법과 같이 같은 공정변수와 안내단어를 통해 위험을 식별하는 경우, 본 연구의 결과인 위험원인과 메트릭을 공정변수로 사용한다면 위험을 더욱 체계적으로 쉽게 식별할 수 있을 것이다. 나아가 해양 시스템의 인공지능 시스템에 대한 새로운 위험성 평가 방법을 개발하는 데 활용할 수 있을 것으로 기대한다. 또한 이 연구의 결과를 활용하면 인공지능의 위험에 대한 특성들을 파악할 수 있으므로 자율운항선박을 위한 안전 요구사항을 더욱 자세하고 구체적으로 작성할 수 있어 자율운항선박의 안전한 항해에 도움이 될 것으로 생각된다.

후 기

본 논문은 과학기술정보통신부 산하 정보통신산업진흥원과 울산 정보산업진흥원의 지원으로 수행되는 "AI 기반 중량화물 이동체 물류플랫폼 실증사업(과제번호 : S1510-22-1001)"에 의하여 이루어진 연구로서 관계부처에 감사드립니다.

References

[1] ABS(2021), Guide for Autonomous and Remote Control Fuctions, http://www.eagle.org

[2] DNV·GL(2018), Autonomous and remotely operated ships, http://www.dnv.com

[3] Institute of Information & Communications Technology Planning & Evalutation(2020), Weekly ICT Trends, Vol. 1978, http://www.iitp.kr

[4] International Organization for Standardization(2022a), ISO/IEC 22989 Information technology - Artificial Intelligence - Artificial Intelligence concepts and terminology, http://iso.org

[5] International Organization for Standardization(2022b), ISO/IEC 23053 Framework for Artificial Intelligence(AI) Systems Using Machine Learning(ML), http://iso.org

[6] International Organization for Standardization(2022c), ISO/IEC DIS 23894 Information Technology - Aritificial Intelligence - Risk management, http://iso.org

[7] Jung, M. and Lee, S.(2020), "UI Standard for Navigation System - Challenges of Implementing New User Interface Guideline", Sea Technology, Dec, 2020, pp. 25-27, http://www.sea-technology.com

[8] Kim, H. and Yang, Y.(2019), "A Review of Human Element Issues of Remote Operators on Maritime Autonomous Surface Ship", Journal of Navigation and Port Research, Vol. 43, No. 6, pp. 395-402.

[9] Kim, M. J., Kim, T. H. and Kim, Y. M.(2021), "On the Integrated process of RSS model and ISO/DIS 21448 (SOTIF) for securing autonomous vehicle safety", Journal of KOSSE, Vol. 17, No. 2, pp. 129-138.

[10] Lee, S. and Lützhöft, M.(2020), "Human-Machine Interaction - The Challenges of New Teamwork for Smart Ship Navigation", Sea Technology, May, 2020, pp. 18-20, http://www.sea-technology.com

[11] Lee, C. and Lee, S.(2021), "Implementation of ISO/IEC 19847 ship data server applied functional safety", 2021 Proceedings of Digital Contents Society (July), pp. 19-21, http://www.dcs.or.kr

[12] National IT Industry Promotion Agency(2019), Software functional safety guideline: Common Industry, http://www.nipa.kr

Received 16 May 2022
Revised 25 May 2022
Accepted 07 June 2022