대한조선학회논문집 Research Paper
Journal of the Society of Naval Architects of Korea
eISSN:2287-7355, Vol. 63, No. 4, pp. 240-252, August 2026
https://doi.org/10.3744/SNAK.2026.63.4.240

Check for updates

잠수함 운항 안전성 강화를 위한 대형언어모델(LLM) 기반 주요안전품목 도출 방안 연구

장호성[1] · 박연동[1] · 금재혁[2] · 도경필[2] · 김기훈[2,†]
국방기술품질원[1]
부산대학교 산업공학과[2]

A Study on LLM-based Method for deriving Critical Safety Items to Enhance Submarine Operational Safety

Ho-Seong Chang[1] · Yeon-Dong Park[1] · Jae-Hyuk Kum[2] · Gyeong-Pil Do[2] · Ki-Hun Kim[2,†]
Defense Agency for Technology and Quality[1]
Department of Industrial Engineering, Pusan National University[2]

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Submarines pose a high risk of catastrophic incidents that may result in significant loss of life and severe degradation of strategic capabilities. In this context, the systematic management of Critical Safety Items (CSIs) is a fundamental requirement throughout the design, construction, and operational phases of submarines. However, conventional CSI identification approaches have largely relied on expert experience and qualitative judgment, which may limit objectivity, consistency, and traceability when large volumes of heterogeneous technical documents are considered. This study proposes a large language model (LLM)-based methodology to address these limitations and enhance the rigor of the CSI derivation process. The proposed approach leverages diverse and authoritative reference sources, including the NATO Naval Submarine Code, SUBSAFE, NAVSEAINST 9078.1/2, MIL-STD-882E, domestic submarine design documents, and quality assurance data. Based on these inputs, CSIs are systematically and objectively derived through an LLM-driven analytical process. Within the proposed framework, submarine expert-driven prompt engineering and Retrieval-Augmented Generation (RAG) techniques are employed to enhance domain specificity, evidence traceability, and analytical reliability. Considering the current limitations of generative AI technologies, the reliability of the LLM-based analysis is reinforced through expert-led selection of critical safety functions and cross-validation of the derived results. The proposed approach reduces potential omissions in experience-dependent methodologies and enables the systematic and evidence-traceable identification of CSIs, thereby supporting a more robust and objective submarine safety management process.

Keywords : Submarine(잠수함), Large Language Model(대형언어모델), Critical Safety Items(주요안전품목), NATO Naval Submarine Code(북대서양조약기구 잠수함 기준), SUBSAFE(미 해군 잠수함 안전 제도)

1. 서 론

잠수함은 한 국가의 군사력을 대표하고, 주변 국가의 위협으로부터 전략적 거부전력으로 운용 가능하며, 다양한 해양 위협 대처에 효과적인 비대칭 핵심 전력이다. 또한 수백여종의 장비, 설비가 탑재되어 통합적으로 연동하여 통합성능을 발휘하는 고도의 복잡성을 요구하는 무기체계이다. 수많은 장비, 설비가 제한된 공간 내에 설치되기 때문에 운용 또는 정비 시 접근이 어렵거나, 설치 또는 분해·조립 시 장기간이 소요되는 경우도 발생한다.

잠수함은 이러한 특성 때문에 치명적인 고장 발생 혹은 사고 발생 시 수리, 복구 등에 긴 시간과 많은 예산이 소요된다. 잠수함의 경우, 사고가 발생하게 되면 대규모 인명 손실을 동반하는 경우가 대부분이다. 국내 잠수함은 현재까지 인명 손실을 동반한 사고 사례는 없으나, 국외 사례에서는 함 침몰, 승조원 사망 등이 발생한 사고 사례가 다수 확인된다. '00년 러시아의 'KURSK'함은 사고당시 선령 6년으로 폭발로 인해 승조원 141명이 사망하

Received : 15 April 2026 | Revised : 13 June 2026 | Accepted : 26 June 2026
† Corresponding author : Ki-Hun Kim, kihun@pusan.ac.kr
It is noted that this paper is a revised edition based on the proceedings of SNAK 2025 in Changwon.

Fig. 1 Eurydice sank at sea (Christopher et al., 2009)

였다. '13년 인도의 'SINDHURAKSHAK'함은 사고당시 선령 16년으로 화재와 폭발로 인해 승조원 18명이 사망하였다. '17년 아르헨티나의 'SAN JUAN'함은 사고당시 선령 32년으로 폭발로 인해 승조원 전원이 사망하였다. '21년 인도네시아의 'NANGGALA'함은 사고당시 선령 40년으로 어뢰훈련을 실시하는 중 실종되었으며 탐색 사흘만에 해저 약 850 m에서 승조원 전원이 사망한 상태로 발견되었다. Fig. 1은 1970년 침몰한 프랑스 Eurydice 잠수함의 해저 잔해이다.

근래에는 재산 및 인명 손실 등 안전에 대한 인식의 변화가 급변하면서 무기체계 또한 체계 안전의 중요성이 강조되고 있다. 타 무기체계 역시 안전이 강조되고 있고, 위 사례는 잠수함 안전 확보의 중요성과 그에 대한 세심한 관리가 필수적으로 동반되어야 한다는 중요성을 반증한다고 할 수 있다.

최근 우리나라 방위사업의 역량은 크게 성장하였으며, '21년 장보고-III Batch-I 도산안창호함을 해군으로 인도하면서 국내 독자 기술을 통한 잠수함 건조 능력을 갖추게 되었다. 더 나아가 해외에서도 국내 잠수함의 그 우수성을 인정하여 도입을 추진하려는 움직임을 보이고 있다. 이러한 잠수함의 기술력 진보와 성장에 발맞추어 잠수함 무기체계에 대한 운항 안전성 확보 관점의 제도 또는 관리방안에 대한 발전도 병행되어야만 한다.

무기체계의 안전성 확보는 국가 안보를 구성하는 핵심 요소로 부각되고 있다. 특히 잠수함은 운용 특성상 사고 발생 시 대규모 인명 피해와 함께 전략적 전력 손실로 직결될 가능성이 높아 운항 안전성 확보는 단순한 기술적 문제를 넘어 국가 차원의 중요한 과제로 인식되고 있다. 이러한 특성으로 인해 잠수함의 안전한 설계, 건조 및 운용을 위해서는 주요안전품목에 대한 체계적인 관리가 필수적이라고 할 수 있다. 그러나, 기존의 주요안전품목 도출 및 관리방안은 체계성과 객관성 측면에서 한계를 지니고 있으며, 다수의 경우 전문가의 경험과 주관적 판단에 의존해 왔다. 한편, 현재 국내 건조 잠수함은 침수 및 화재와 같은 중대 사고에 대비하기 위해 미 해군 SUBSAFE 개념을 기반으로 설계단계부터 건조단계 전반에 걸쳐 다양한 안전통제활동을 수행하였다(Sullivan, 2005). 그럼에도 불구하고, 주요안전품목의 체계적인 도출 방법론에 대한 연구는 미미한 실정이다.

본 연구는 앞서 언급한 기존 접근 방식의 한계를 극복하기 위해 대형언어모델(Large Language Model; LLM) 기술을 적용한 새로운 주요안전품목 도출 방법론을 제안한다. 제안된 방법은 NATO Naval Submarine Code, SUBSAFE, NAVSEAINST 9078.1/2, MIL-STD-882E, 잠수함 설계 문서, 잠수함 사고 사례 및 출항불가조건, 품질 데이터 등 다양한 신뢰성 있는 객관적 자료를 기반으로 한다. 이러한 다원적 입력 정보를 LLM 기술과 접목함으로써 주요안전품목을 체계적이고 객관적으로 도출하는 것이 가능하다.

또한, 본 연구에서는 잠수함 분야 전문가의 지식을 반영한 프롬프트 설계와 검색증강생성(Retrieval Augmented Generation; RAG) 기법을 활용하였다. 이를 통해 주요안전품목 도출 과정에서 LLM 답변의 근거를 명확히 제시하도록 하고, 도출 과정의 정밀성, 재현성, 투명성 및 접근성을 확보하고자 하였다.

2. 안전 분야 LLM 기술 활용

2.1 안전 분야 LLM 기술 적용 필요성

안전 분야에서는 사고 조사 보고서, 장비 정비 이력, 위험성 평가 자료 등 텍스트 기반의 비정형 데이터가 대규모로 생성 및 축적되는 특성을 지닌다. 특히 잠수함과 같은 무기체계는 복잡한 계통 연동 구조와 고도의 신뢰성이 요구되며, 잠수함의 비정형 데이터는 전문 용어가 혼재되어 있을 뿐만 아니라, 작성 주체에 따른 표현 편차와 다차원적인 인과관계가 복합적으로 얽혀 있다는 특징을 가진다. 이러한 특성으로 인해 기존의 규칙 기반 알고리즘이나 단순 키워드 매칭 방식만으로는 잠재적 위험 요소를 효과적으로 식별하거나 사고의 근본 원인을 분석하는데 한계가 존재한다. 이에 따라 최근에는 방대한 문서 정보를 이해하고 이를 바탕으로 한 고도화된 추론 능력을 보유한 LLM 기술을 안전 분야에 도입하여 비정형 데이터를 기반으로 위험 요소를 능동적으로 식별하고, 선제적인 예방 조치를 도출하려는 시도가 점차 확대되고 있다. LLM 기술을 활용한 위험 분석이 기존의 경험 의존적, 사후 대응 접근 방식을 보완할 수 있는 새로운 패러다임으로 최근 주목받고 있다.

2.2 안전 분야 LLM 기술 활용 사례

최근 고도의 신뢰성이 요구되는 의료 분야에서도 LLM 기술을 활용한 위험 분석 연구가 수행되고 있다. 의료 도메인에서 추적된 만여건의 환자 안전 사고 리포트를 대상으로 LLM의 위험 분석 역량을 체계적으로 검증하였다. 해당 연구는 비정형 텍스트 형태로 기록된 사고 보고서를 LLM에 입력하여 사고 원인, 위험 수준, 재발방지대책 등을 자동으로 도출하도록 설계되었다. 실험 결과, LLM은 인간 전문가의 수동 분석 대비 수백 배 이상의 처리 속도를 보였을 뿐 아니라, 위험요인 식별의 일관성 측면에서도 우수한 성능을 나타냈다. 특히 인간 전문가 분석 과정에서 간과되기 쉬운 사고 보고서 내의 미세한 텍스트 상관관계를 효과적으로 포

Fig. 2 Autonomous driving - Apollo

착함으로써 잠재적 위험 요인을 식별할 수 있는 가능성을 입증하였다. 이러한 결과는 고도의 신뢰성이 요구되는 안전 분야에서 LLM 기술이 위험 분석을 보조하는 유효한 도구로 활용될 수 있음을 시사한다(Xu et al., 2025). 나아가 LLM 기술을 활용한 안전, 위험 분석은 의료분야 뿐만 아니라 체계공학, 기능 안전, 정비 및 운항 등 안전 분야 전반으로 확산되고 있다. 덧붙여 LLM이 해석 및 추론을 과도하게 수행하게 될 경우 환각 발생 위험이 있으므로, 이를 방지하기 위한 방안으로 LLM 추론 결과에 대한 근거 제시, 검증 절차를 적용하며 발전하고 있다. 상기 내용을 반영하여 LLM 기술을 적용한 안전 분야 활용 사례는 다음과 같다.

첫째, 자율주행 안전 요구사항 도출 사례이다. 에이전트 기반 RAG로 안전 요구사항 도출을 자동화하였다. Fig. 2는 Apollo 자율주행 차량의 사례로, 상부 센서 모듈과 시스템을 통해 환경 인식, 위치 추정, 경로 계획 및 차량 제어 기능을 수행한다. 자동차 표준 문서 풀과 Apollo 사례를 결합한 에이전트 기반 RAG를 통하여 안전 요구사항에 대한 질문 및 답변 데이터셋에 대해 더 관련성이 높은 근거 검색을 유도함으로써 안전 요구사항 도출의 신뢰도를 향상시키는 접근법을 제시하였다. 이는 안전 분야에서 생성 자체보다 정확한 근거 검색 및 선별이 위험 분석 품질을 좌우함을 시사한다(Balahari et al., 2025).

둘째, 자동차 기능 안전 엔지니어링의 위험 평가 사례이다. ISO 26262 등 표준 지식을 RAG로 보강하고, 기능안전 관리자, 검증 수행자, 전문가 등 에이전트 역할을 분담하여 자동차 기능안전 엔지니어링의 고난도 과제인 HARA (Hazard Analysis and Risk Assessment) 수행, FSR (Function Safety Requirement) 문서화, AEB (Autonomous Emergency Braking) 테스트 설계 등을 지원한다(Lu et al., 2024).

Fig. 3 Aviation accident reporting system - ASRS

셋째, 항공기 정비 고장분석 분야이다. RAG를 통한 비정형 문서 검색만으로는 복잡한 고장 형태에 대한 분석에 한계가 있음을 지적하고, 지식 그래프, 벡터 검색 등을 결합한 다차원 검색과 그래프 기반 추론을 통합하였다. Fig. 3은 항공안전보고시스템(ASRS)을 나타낸 것으로, 항공 운항 중 발생 가능한 사고 및 안전 위해요소를 보고·분석하여 항공안전 개선에 활용하는 체계를 보여준다. 항공기 정비 고장분석은 지식 집약적 성격을 띄므로 제조사, 정비기관, 규제기관, 학계 등의 다양한 문서로부터 항공기 고장 지식 그래프를 구축하였다(Xiaoyue et al., 2025).

마지막으로, 항공 안전 보고서(ASRS) 사례이다. 해당 연구는 ASRS의 사건 내러티브를 대상으로 LLM이 사건 요약 생성, 인적 요인 식별, 관련 주체, 책임 주체 평가 등에 활용될 수 있음을 제시한다. 기존의 보고서 처리에는 두 명의 분석가가 업무일 기준으로 최대 수일이 소요되었으나, LLM을 통해 많은 부분을 자동화하였음을 나타내고 있다. 동시에 모델의 편향 및 불확실성을 고려한 Human-in-the-loop 검증 절차를 제안하여 고신뢰 영역에서 LLM 기술은 인간 전문가 분석의 대체가 아니라 일관된 1차 분석과 후보 생성 역할이 가장 현실적임을 시사한다(Archana et al., 2023). 상기 사례를 종합해보면 LLM 기술의 강점으로는 처리 속도, 일관성, 미세 상관관계 포착 등이 있고, 환각은 약점으로 작용할 수 있다. 이를 보완하기 위한 방안으로 추론 결과의 근거를 제시하도록 하기 위한 RAG 기법 적용을 제시하고 있다. LLM 추론 결과에 대한 최종 판정과 책임성 확보를 위한 Human-in-the-loop 검증 절차를 결합하면 고도의 신뢰성이 요구되는 안전 분야에서 일관성 있고, 빠르며, 체계적인 분석 보조자로 활용하면서도 오판 위험을 체계적으로 제어 가능하다고 할 수 있다.

3. LLM 기반 주요안전품목 도출

3.1 주요안전품목

함정의 주요안전품목은 DFARS 209.270-2, NAVSEAINST 9078.1/2, MIL-STD-882E에 정의되어 있다. 함정의 부품, 조립체, 지원장비 중 그 특성의 결함, 오작동, 부재의 경우 함정의 상실 또는 중대한 손상으로 이어질 수 있으며, 파국적 인명피해나

Table 1 Arrangement for equipment - partial extract

Equipments		
Combat system	Snorkel flap	Pump
Propulsion motor	Mast	Switch board
Attack periscope	Weapon system	Propeller
Navigation system	Diesel engine	Optronic mast
Lithium battery system	Radar	Hatch
Sonar system	Echo sounder	Communication system

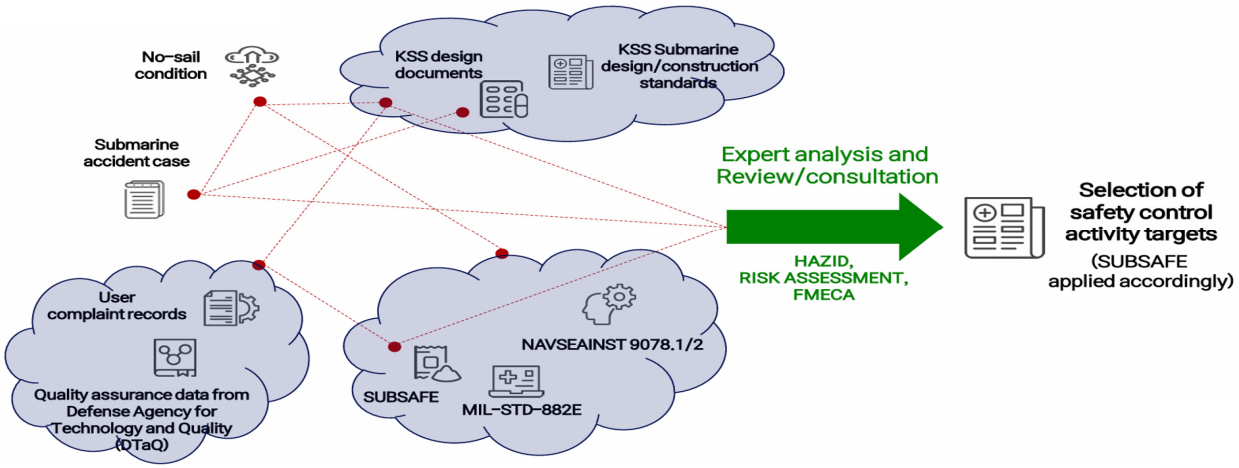

Fig. 4 Conventional approach

생명 손실 등 용납할 수 없는 수준의 위험을 유발할 수 있는 것을 의미한다. Table 1은 잠수함의 탑재장비 일부를 발췌하여 나타낸 것이며, LLM을 활용한 주요안전품목 도출은 잠수함(정) 확장 형작업분할구조(Expanded Ship Work Breakdown Structure; ESWBS) 구성을 고려하여 탑재장비 레벨 단위로 식별하였다.

3.2 기존 도출방안

Fig. 4는 주요안전품목 도출 과정에 대한 기존 접근법을 나타낸 개념도이다. 잠수함 설계 문서, 잠수함 설계/건조 기준, 품질 데이터, 운용 중 결함 이력, 사용자 불만 기록, SUBSAFE 등 다양한 자료를 폭넓게 활용하고, 인간 전문가 검토를 통하여 HAZID (Hazard Identification), FMECA (Failure Mode, Effects, and Criticality Analysis)를 수행한 뒤 안전통제활동 대상을 선정하는 구조이다. 안전통제활동 대상은 건조단계에서 이루어지는 품질검사 및 확인 대상을 모수로 하여 선정하며, 선정된 안전통제활동 대상을 주요안전품목으로 간주하여 총 수명주기 동안 관리하고 있다. 이 과정은 위험 분석의 정성적 통찰을 확보한다는 장점이 있으나, 분석 입력의 규모가 커질수록 인간이 수행하는 수동 검토는 필연적으로 선택적이 되며, 문서의 표현 체계가 서로 다르다는 점이 누락, 오판 및 추적성 저하를 유발할 수 있다. 특히 주요안전품목 도출은 결함, 오작동 또는 부재 시 함정 손상, 인명 손실 등 치명적 피해를 초래할 수 있는 핵심 특성을 포함한 부품, 조립체, 지원장비를 대상으로 하므로 결론의 근거가 추적 가능해야 한다. 이러한 관점에서 기존 방안의 한계점은 다음과 같이 제시할 수 있다.

첫째, 정보의 누락 가능성이다. 입력 문서가 방대하고 분산되어 있을수록 예기치 못한 검토 누락이 발생할 수 있고, 시간과 인력의 제약 하에서 특정 문서군이 반복적으로 우선 검토되거나 반대로 운용 단계의 산발적 기록 등은 검토 범위에서 제외될 수 있다. 그 결과, 실제 위험의 분포가 아니라 검토된 문서의 분포가 주요안전품목 후보군을 규정하게 되어, 특정 위험이 과소 반영되거나 특정 분야에 과도하게 편중될 수 있다. 이러한 누락 또는 편중은 동일 위험 신호가 여러 문서에 분산 기록된 경우 이를 통합하지 못하여 위험 신호가 약화되는 형태로 나타날 수 있다.

둘째, 문서 간 용어 및 식별체계 차이에 대한 명시적 매핑과 추적 관리의 부담이다. 동일한 대상이라도 설계 문서의 장비명, 계약 문서의 명칭, 표준 문건의 용어, 프로젝트별 식별번호 등에 따라 서로 다르게 표현될 수 있다. 이러한 차이는 잠수함 분야 전문가 협의를 통해 식별될 수 있으나, 용어 대응관계, 장비명 매핑, 판단 근거가 명시적으로 축적·관리되지 않을 경우 담당자 변경, 프로젝트 변경, 문서 갱신 이후 동일 판단을 재검토하거나 재사용하는 데 어려움이 발생할 수 있다. 특히 동일 장비 또는 동일 기능과 관련된 위험 신호가 여러 문서에 서로 다른 명칭으로 분산되어 기록된 경우, 그 연결관계가 최종 판단 근거와 함께 체계적으로 남지 않으면 추적성과 재현성이 저하될 수 있다. 따라서 본 연구에서는 용어 차이 자체를 기존 전문가 방식의 한계로 보기보다, 전문가가 식별한 용어 및 장비 매핑 결과를 근거 문서와 연결하여 지속적으로 축적하고 재검토할 수 있는 관리 체계의 필요성으로 해석하였다.

셋째, 해석의 다양성에 따른 비일관성이다. HAZID, FMECA는 정량 및 정성적 요소가 혼재하고, 심각도, 발생도, 검출도 같은 등급 판단은 인간 전문가의 경험과 프레이밍에 영향을 받는다. 동일 문서를 보더라도 각 전문가마다 위험의 중심 경로를 다르게 해석하거나, 회의, 자문 과정에서 형성된 합의가 충분한 근거 연결로 문서화되지 않은 채 암묵지로 남는 경우가 발생할 수 있다. 이에 따라 재현성이 저하되어 주요안전품목 도출은 논리의 산물이라기보다 합의의 산물로 오인될 수 있고, 시간이 경과할수록 다양한 분야의 인간 전문가가 수행한 기술검토 결과에 대한 배경이 휘발될 수 있다.

넷째, 인간 전문가 경험, 전문성에 따른 분석 편중 및 분석 심도의 불균형이다. 분야별 인간 전문가는 자신이 익숙한 위험 유형에 더 세밀한 프레임을 적용하는 경향이 있고, 교차영역 위험은 책임 주체가 모호하여 검토의 사각지대가 될 수 있다. 그 결과, 특정 탑재장비군은 고장 유형이 세분화되는 반면 다른 탑재장비군은 비교적 개략적으로 분류되는 문제가 발생할 수 있다.

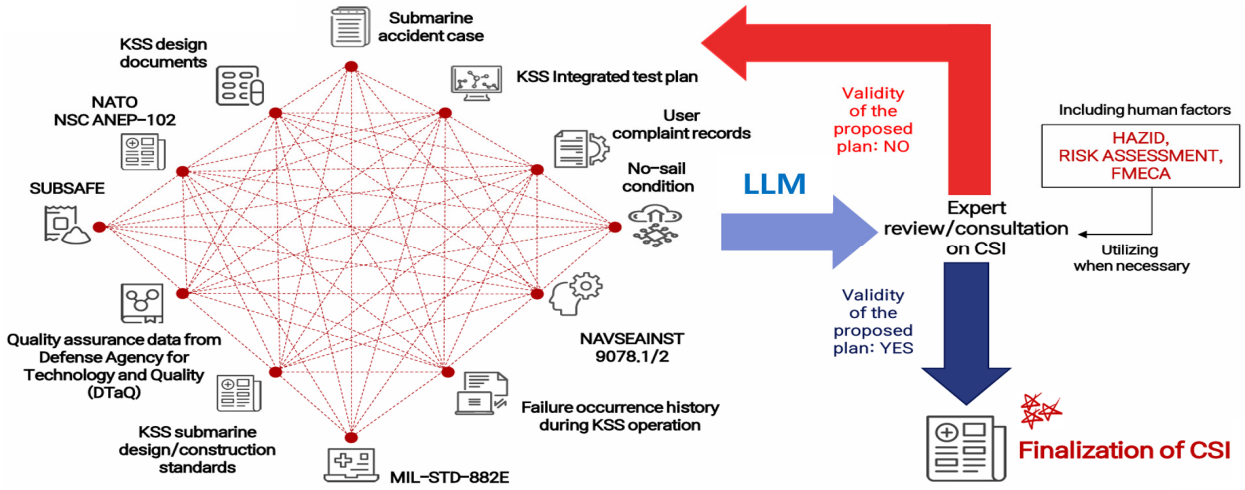

Fig. 5 LLM-based approach

다섯째, 결론에 대한 근거의 추적성 제한이다. 대량 문서를 바탕으로 검토가 이루어질수록 실제로는 많은 근거가 사용되지만, 최종 보고서에는 일부 핵심 문서만 인용되거나, 어떠한 문서의 어떤 구절, 표, 그림이 어떠한 결론을 도출하는데 주로 인용되었는지가 명시적으로 연결되지 않는 경우가 흔하다. 반면 MIL-STD-882E에 따르면 체계안전 표준은 위험을 닫힌 고리 형태로 추적 및 관리하는 것을 요구하며, 위험 식별 등 데이터 요소를 체계적으로 기록, 갱신하는 접근을 강조한다. 결국 추적성이 저하되면 설계 변경, 부품 대체, 운용 결함 누적 등 환경 변화가 발생했을 때 기존 결론을 재평가하기 위한 시간과 비용이 급증할 수 있다.

기존 인간 전문가 중심의 주요안전품목 도출 방식의 구조적 한계는 사람의 한계라기보다 기록, 연결의 형태가 약한 상태에서 방대한 문서를 다루는 방식에서 기인한다고 할 수 있다.

따라서 LLM 기술을 적용하여 정보 누락, 오연결, 정보간 연계의 단절 등을 줄이고, 분석 편중과 분석 심도의 불균형을 완화하며, 무엇보다 결론의 근거를 원문 단위로 추적 가능한 형태로 축적할 수 있다.

3.3 LLM 기반 도출방안

Fig. 5는 LLM 기반 주요안전품목 도출 접근법을 나타낸 개념도이다. 본 연구는 LLM을 활용한 주요안전품목 도출 절차를 주요안전기능 중심으로 설계하였다. 먼저 주요안전품목 도출의 기준점으로써 주요안전기능을 정의하고, 이후 LLM 입력 문서를 대상으로 해당 기능과 연계되는 위험요인 및 위험상황을 검색, 추출하였다. 다음 단계에서는 추출된 위험요인과 위험상황을 단서로 하여, 이에 대응하는 주요안전품목 후보군을 LLM 기반으로 추론하였다. 아울러 각 주요안전품목 후보군에 대해 추론의 근거가 되는 인용 문서, 판단 연결고리 등 원문에서 추적 가능한 형태로 제시함으로써, 추론의 결과가 특정 근거로 환원될 수 있도록 설계하였다. 잠수함 안전설계는 사고 사례와 장기간의 운용 경험 축적을 통해 형성된 '안전 우선' 설계 철학을 핵심 전제로 한다. 동일한 맥락에서 주요안전품목의 선정은 단순한 문건 패턴의 조합을 넘어, 안전 철학과 설계 관행에 대한 이해를 전제로 한 인간 전문가의 통찰과 공학적 판단을 필요로 한다. LLM은 대규모 문서에서 규칙성과 상관관계를 효과적으로 포착하고, 문맥의 흐름을 통해 추론결과를 제시할 수 있으나, 잠수함 안전 철학을 자명하게 '이해'하는 방식으로 작동하는데는 제한이 있다. 따라서 본 연구에서는 주요안전품목 도출의 기초가 되는 주요안전기능을 LLM이 아닌, 잠수함 분야 인간 전문가가 직접 선정하였다. 이는 LLM 분석의 효율성을 활용하되, 핵심기준 설정은 인간 전문가 판단에 기반하도록 역할을 분담한 것이다.

또한 본 연구의 LLM 기반 접근은 문서 간 용어 차이나 프로젝트별 장비명 차이를 모델 자체적으로 완전 자동 식별한다는 것을 전제로 하지 않는다. 잠수함 기술문서에서 사용되는 장비명, 기능명, 식별번호 및 약어는 문서 작성 주체와 프로젝트에 따라 달라질 수 있으므로, 초기 용어 리스트와 장비명 인덱싱, 기준 문건과 프로젝트 문건 간의 대응관계 설정은 분석 결과의 신뢰성에 중요한 영향을 미친다. 이에 따라 본 연구에서는 NATO NSC 기반 주요안전기능, 잠수함(정) 확장형작업분할구조, SUBSAFE, NAVSEAINST 등 기준 문건을 참조하여 용어 및 장비 매핑의 기준축을 설정하고, LLM은 문맥적 유사성 및 기능적 연관성을 바탕으로 후보 연결관계를 제시하는 역할을 수행하도록 하였다. 최종적인 용어 대응관계와 주요안전품목 판단은 전문가 검토를 통해 확인하도록 하여, 초기 인덱싱 오류나 오연결이 최종 판단으로 직접 이어지지 않도록 하였다.

잠수함 안전성 확보를 위해 미 해군의 NAVSEA는 SUBSAFE를 잠수함 안전 품질보증 프로그램으로 운영하며, 선체의 수밀성과 예기치 못한 침수에서의 회복 가능성에 대해 최대 합리적 보증 제공을 핵심 목적 중 하나로 명시하고 있다. 동시에 잠수함 안전 인증 체계에서는 목표기반(Goal Based Structure; GBS) 프레임워크가 널리 활용되고 있다(Richard et al., 2017). 영국, 프랑스, 독일, 캐나다 등 다양한 국가에서 국제적 수준의 잠수함 운항 안

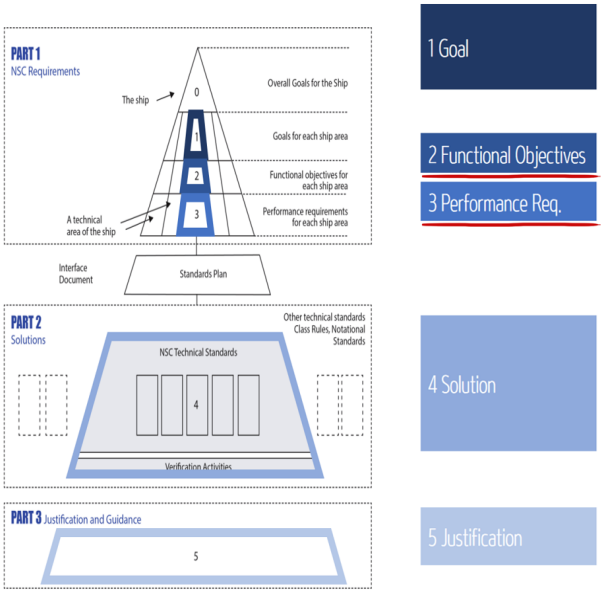

Fig. 6 Naval submarine code structure (parts and tier)

전성 달성을 위해 해군 및 선급 등으로 구성된 국제함정안전협회(International Naval Safety Association; INSA)의 NATO 기준(Naval Submarine Code; NSC)을 활용하고 있다.

Fig. 6은 NATO NSC의 구성을 나타낸다. Part 1의 Tier 1은 목표, Tier 2는 기능요구사항, Tier 3은 성능요구사항을 나타내며, Part 2의 Tier 4는 해결책, Part 3의 Tier 5는 정의 및 참조를 나타낸다. 국내 잠수함 건조 시 적용한 안전통제활동은 SUBSAFE를 근간으로 하여 침수, 침몰, 부상 등에 집중되었다면, NATO NSC는 일반 요구사항, 구조, 부력 및 복원성, 엔지니어링 시스템, 함 운용 시스템, 화재 안전, 잠수함 탈출/구조/이함 및 생존, 통신, 항해, 위험물, 대기제어로 총 11개의 장으로 구성되어 보다 운항 안전성 관리와 주요안전품목 도출의 기준으로 활용하는데 더욱 적합하다고 할 수 있다(Lee, 2024).

주요안전기능은 국제함정안전협회의 NATO NSC를 근거로 정립하였다. 구체적으로는 NATO NSC의 Tier 2의 기능요구사항 및 Tier 3의 성능요구사항을 주요 참조 축으로 삼아, 잠수함 분야 인간 전문가 논의 과정을 통해 주요안전기능을 정립하였다. 11개 장을 기반으로 총 69개의 주요안전기능을 정립하였으며, Table 2는 주요안전기능 중 각 장별로 주요내용을 발췌하여 정리한 사항이다.

주요안전품목 도출을 위해 우선 LLM의 입력으로 활용할 잠수함 설계 문서, 잠수함 설계·건조 기준, 품질 데이터, 출항불가조건, 운용 중 결함 이력, 사용자 불만 기록, 건조 중 잠수함 통합검사항목, SUBSAFE, NAVSEAINST 9078.1/2, MIL-STD-882E 등을 식별하고, 이를 임베딩에 적합한 형태로 전처리한다. 이후 전처리 및 임베딩된 데이터를 기반으로 RAG를 수행하여 잠수함의 주요 안전기능에 위험을 초래할 가능성이 높은 탑재장비를 선정하였다. 이때 탑재장비 수준의 식별은 국방표준서에 제시된 잠수함(정) 확장형 작업분할구조를 활용하여 객관성을 높였다.

또한 LLM의 추론 과정에서 잠수함 분야 인간 전문가의 관점을

Table 2 Critical safety function based on NATO NSC

System	NATO NSC	Tier 2 Functional objective & Tier 3 Performance Requirement
Sub-marine	General requirements (6 items)	• Maintain a safe state at the maximum shock design level • Consider the submarine's operational conditions during surface running, submerged operation, periscope-depth operation, and snorkeling • ·····
	Structure (7 items)	• Satisfy the applied load conditions acting on the propeller and shafting system during surface running or submerged operation • Satisfy the applied load conditions acting on the mast, periscope, and other appendages during surface running or submerged operation • ·····
	Buoyancy, restoring buoyancy & control performance (3 items)	• Watertight integrity function • Static control function, including maintenance of the center of buoyancy in accordance with appropriate tank capacity • ·····
	Engineering system (13 items)	• Maintain heading and attitude in the event of flooding or fire • Provide ventilation for hazardous areas with potential flammable or explosive atmospheres • ·····
	Ship operation system (3 items)	• Capability for the submarine to be towed or to tow another vessel • Capability to moor without using its own propulsion system • ·····
	Fire safety (8 items)	• Reduce the spread of smoke and toxic effluents from the fire-affected compartment • Fire detection and alarm function • Fire suppression function • ·····
	Escape, rescue, evacuation, and survival (6 items)	• Casualty transfer capability • Passage/egress capability to enable efficient evacuation • ·····
	Communi-cation (4 items)	• External communication capability for maritime distress situations, including distress signal transmission • Capability to disseminate emergency notifications via internal communication systems/equipments • ·····

Table 2 Critical safety function based on NATO NSC (Continued)

System	NATO NSC	Tier 2 Functional objective & Tier 3 Performance Requirement
Sub-marine	Navigation (6 items)	• Capability to determine and input speed and route • Capability to determine and input roll, pitch, and heading • ·····
	Dangerous goods (3 items)	• Fire detection, alarm, and response capability • Gas and vapor leak detection capability • ·····
	Atmospheric control (10 items)	• Capability to control carbon dioxide (CO_2) to maintain safe levels • Capability to monitor onboard atmospheric composition • ·····

체계적으로 반영하기 위해, 전문가의 판단 절차를 분석 프레임워크에 포함하였다.

구체적으로는 방대한 대상 문서 가운데 주요안전품목 도출에 실질적으로 기여하는 정보를 우선 식별하고, 해당 정보를 효과적으로 도출할 수 있도록 프롬프트를 목적 지향적으로 구성하였다.

즉, 단순 요약이나 포괄적 질의가 아니라 문서의 핵심 근거를 추출할 수 있도록 참조 우선순위와 질의 방향을 사전에 설계하였다. 아울러 LLM 추론 과정의 일관성, 재현성, 그리고 단계 간 정합성을 확보하기 위한 세부 기준을 함께 정의하였다.

이러한 프롬프트 기반 기준 정의를 통해, LLM의 추론 결과가 특정 문서 근거와 평가 규칙에 의해 설명될 수 있도록 하였으며,

Table 3 Submarine expert-based prompt design

Division	Specification for prompt engineering
Identification for risk factors & hazardous situations	• What actual accident cases have been caused by the identified risk factors? • In the context of MIL-STD -882E accident risk classification guidelines, does the hazardous situations correspond to catastrophic (Severity 1) or critical (Severity 2) failure? • For the identified hazardous situations to occur, which function of the on-board equipment related to no-sail conditions must be lost? • To prevent hazardous situation arising from risk factors, is the identifiedd hazardous situations included within SUBSAFE?
Quantitative occurrence frequency of hazardous situations	• How many accident cases have been associated with the identified hazardous situations? • How many failure records during the operational phase are associated with the identified hazardous situations?

Table 3 Submarine expert-based prompt design (Continued)

Division	Specification for prompt engineering
Quantitative severity of the hazardous situations	• Can the identified hazardous situations cause the submarine to cease operation? • Can the identified hazardous situations prevent the submarine from submerging or surfacing? • Does the identified hazardous situation violate the requirements specified on NATO NSC (ANEP-102)?
Correlation between risk factors and hazardous situations	• The extent to which risk factors, through functional linkage or situational dependency, conditionally triggers or aggravates hazardous situations by leading to the absence or failure of specific submarine functions
Correlation between risk factors and critical safety function	• The extend to which the state of 'critical safety function A' acts as a risk factors and affects 'critical safety function B', such correlations focus on system-level and systemic inter-dependencies as well as cascading effects
Correlation between critical safety function and on-board equipment	• The extend to which the on-board equipment is required for the implementation of the critical safety function, the extent to which the on-board equipment provides and supports the conditions necessary for performing the critical safety function
Criteria for the correlation between risk factors and hazardous situations	• Causality: The extent to which risk factors directly trigger hazardous situations • Occurrence correlation : The extent to which risk factors and hazardous situations occur simultaneously or are closely related in occurrence • Functional relevance: The degree of functional association between risk factors and hazardous situations
Criteria for the correlation between risk factors and critical safety function	• Impact on human safety: The potential extent to which the combination of risk factors and hazardous situations directly affects crew life or safety • Co-occurrence frequency: The extent to which risk factors and hazardous situations appear together in context
Criteria for the correlation between critical safety function and on-board equipment	• Direct functional contribution: The extent to which the item directly implements or performs a specific CSIs • Irreplaceablility: The extent to which the on-board equipment can or cannot be substituted or supported by other equipment • Safety impact: The degree of severity on safety when the on-board equipment loses its function

인간 전문가 판단이 개입되는 지점을 명확히 하여 분석의 신뢰성과 추적성을 강화하였다. Table 3은 상기 기술한 주요내용을 발췌하여 정리한 사항이다.

상기 제시된 사항들을 반영하여 Python, Miniconda, Library, IDE (Integrated Development Environment) 등을 통해 RAG 기반 LLM 분석 환경을 구축하였다. 식별한 다양한 잠수함 관련 문서는 Qwen3 모델 자체를 추가 학습하거나 fine-tuning하기 위한 데이터가 아니라, RAG 기반 검색 지식베이스를 구성하기 위한 입력 문서로 활용하였다. 즉, 본 연구에서는 관련 문건을 전처리 및 임베딩한 후, 질의와 관련된 원문 근거를 검색하고, 이를 바탕으로 LLM이 위험요인, 위험상황, 주요안전기능 및 탑재장비 간의 연결성을 추론하도록 설계하였다. 또한 잠수함 관련 문서의 보안성을 고려하여 LLM 분석 환경은 국방 폐쇄망 내에서 구현하였으며, 관련 일체의 절차에 대해서는 보안성 검토 및 점검을 수행하였다. 이때 활용한 LLM 모델, 비전-언어 모델 및 임베딩 모델의 세부사항은 Table 4와 같다.

Table 4 The utilized LLM models

Model	Description	Size
LLM	Qwen/Qwen3-30B-A3B	70 GB
	Qwen/Qwen2.5-VL-32B-Instruct	70 GB
Embedding	vidore/colqwen2.5-v0.2	1 GB

본 연구에서는 전문 용어 밀도, 장문 및 복문 구조, 규정 및 절차 관련 내용 포함 등 국방 분야 문서의 특성과 근거 기반 추론의 안정성을 고려하여 Qwen3 계열을 LLM 핵심 추론 모델로 채택하였다. 해당 계열은 한국어를 포함한 다국어 처리 성능이 강화된 것으로 알려져 있으며, 특히 30B급(Qwen3-30B-A3B) 파라미터 규모는 복잡한 문맥에서의 추론과 장문 이해에 유리한 특성을 제공한다. 또한 본 연구의 주요 목표가 근거 기반의 주요안전품목 도출에 있으므로, 장문 기술문서 기반의 지시문 이해, 근거 기반 응답 생성, 폐쇄망 환경에서의 로컬 운용 가능성 등을 고려하여 모델을 선정하였다.

다만 본 연구에서 Qwen3-30B-A3B를 채택한 것은 해당 모델이 모든 LLM 대비 절대적으로 우수함을 전제로 한 것은 아니다. 또한 본 연구는 잠수함 관련 문건을 이용하여 Qwen3 모델의 파라미터를 추가 학습하거나, 본 논문에 특화된 신규 LLM을 개발한 것이 아니다. 본 연구에서의 특화는 모델 자체의 재학습이 아니라, 잠수함 주요안전품목 도출 목적에 맞춘 입력 문서 구성, RAG 기반 근거 검색, 전문가 기반 프롬프트 설계, 단계별 추론 절차 및 Human-in-the-loop 검증 구조를 포함하는 분석 파이프라인의 특화를 의미한다. 본 연구는 잠수함 주요안전품목 도출이라는 특정 목적에 대해, 폐쇄망 환경에서 운용 가능한 LLM 및 RAG 기반 분석 절차의 적용 가능성을 확인하는 데 초점을 둔다. 이에 따라 모델 선정 시 공개 가중치 기반의 로컬 운용 가능성, 국방 폐쇄망 내 구현 가능성, vLLM 기반 서빙 가능성, 장문 기술문서 처리 가능성, 한국어를 포함한 다국어 문서 처리 가능성, 그리고 제한된 계산 자원 내 추론 효율성을 주요 고려사항으로 설정하였다.

한편 RAG 파이프라인의 검색 정확도를 향상시키기 위하여 시각-언어 임베딩 모델인 vidore/colqwen2.5-v02를 적용하였다. 기존 텍스트 중심 임베딩은 문서의 텍스트를 벡터화하여 검색을 수행하는 반면, 본 모델은 문서의 페이지 이미지를 입력으로 받아 시각 정보와 텍스트 정보를 통합적으로 임베딩 공간에 매핑한다. 잠수함 기술자료는 텍스트뿐 아니라 도면, 이미지, 복잡한 표가 핵심 근거로 작용하는 경우가 많으며, 텍스트 중심 검색에서는 OCR 과정에서 표 구조가 훼손되거나 이미지 내 텍스트의 위치, 배치 정보가 소실되어 검색 성능 저하가 발생할 수 있다. 반면 본 연구에서 적용한 모델은 페이지 전체를 시각적 맥락으로 반영하는 ColPali 계열 아키텍처를 기반으로, 질의와 연관성이 높은 표, 도면 등이 포함된 페이지 단위를 보다 정확하게 식별하도록 설계되어 있다(Manuel et al., 2025). 결과적으로 이는 근거 탐색 단계의 정확도를 높여, 근거 기반 분석의 신뢰성을 뒷받침하는 핵심 구성요소로 기능한다. 본 연구에 제시된 주요안전품목 도출 특화 LLM을 개발하고, 운용하기 위한 GPU 환경, 운용체제 등은 Table 5와 같다.

Table 5 System configuration

Model	Specification
GPU	NVIDIAH100 (90GB) × 4
OS	LinuxReadHat5.14.0-503.33.1.el9_5x86_64
CUDA	Driver570.124.06, CUDA 12.8
Environment management	Condavitualenvironment based on Miniconda
Model serving/ Interference engine	vLLM0.10.1.1.
Connected method	ssh (pUTTY)
File transfer	SFTP (FileZilla)

3.4 주요안전품목 LLM 추론결과

설계한 프레임워크와 LLM을 통하여 잠수함(정) 확장형작업분할구조를 기준으로 총 317종의 탑재장비에 대한 위험요인, 위험상황을 식별하였다. 인간 전문가가 선정한 주요안전기준은 총 69개이고, 탑재장비별 식별된 위험요인, 위험상황의 평균적인 개수가 5개이므로 LLM 답변 추론결과는 약 546,825개 수준이다.

그 중 대표적인 LLM 추론 답변 사례를 Fig. 7과 같이 발췌하여 나타내었으며, 잠수함 잠항 제어체계를 대상으로 수행한 LLM 기반 추론 결과를 단계별로 제시한다. 타기 제어용 유압공급장치의 고장과 함내 타기 조종콘솔의 고장이 대표적인 위험요인으로 도출되었다. 이러한 위험요인은 결과적으로 수평타 및 수직타 기능 저하/고장 가능성과 트림 계통의 이상과 같은 위험상황으로 연결되는 것으로 식별되었다.

Step A에서는 학습 대상 문건에 대하여 위험요인을 식별하고, 위험요인과 시스템 영향 간의 연계 강도를 정성적으로 평가하였다. 그 결과, 유압밸브 고장이 배관 계통 문제로 확장될 경우 수

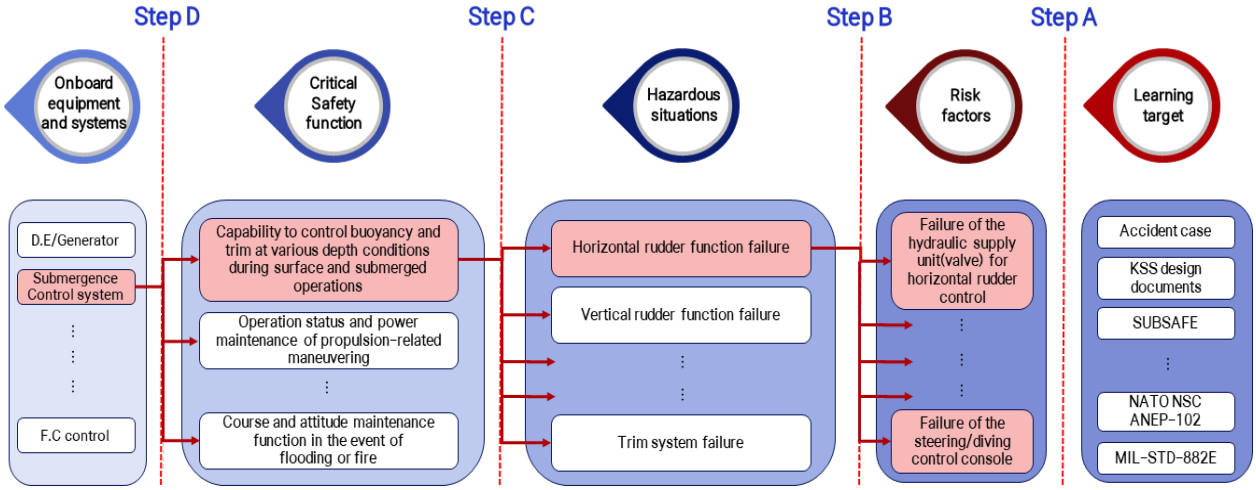

Fig. 7 LLM-based CSIs derivation (Submerge control system)

평타 기능에 영향을 줄 수 있으며, 관계 데이터(연계 근거)에서 간접적 연결은 확인되나 직접 인과성은 상대적으로 약한 수준으로 판단된다는 취지의 답변이 도출되었다. Step A는 가능 경로의 존재를 확인하되 영향의 확실성, 즉 인과관계 강도는 제한적으로 평가하는 단계로 해석할 수 있다.

Step B에서는 수평타의 기능적 역할과 제어계통 구성 관점에서 위험요인과 위험상황 간의 영향에 대한 연관성을 검토한다. 추론 결과, 수평타는 잠수함의 심도 조정에 핵심적 역할을 수행하며, 관계 데이터에 수평타 제어계통의 독립성, 유압공급계통 고장과의 직접 연계가 명시되어 있는 것으로 나타났다.

Step C에서는 심도 조건 하에서의 제어 성능 요구와 위험상황 간의 연결을 중심으로 분석을 확장하였다. 수평타 기능의 고장은 부력 및 트림 제어 기능에 즉각적, 직접적 영향을 미친다는 근거 서술이 복수의 관계 데이터에서 확인되어 Step B 대비 연계 강도가 강화된 결론이 도출되었다. 그 결과, 운용 심도 조건에서의 부력 제어는 수심유지 및 전반적인 잠항상태 제어와 밀접한 핵심 요소로 제시되었고, 잠항을 위한 제어능력과 직접적인 연관성이 확인되었다. 특히 신호 오류(또는 제어 신호 이상)가 시스템 영향으로 이어질 가능성이 상대적으로 크게 평가되었으며, 잠항상태 제어와 심도 조건 간의 상호 연관성이 관계 데이터에 명시되어 있다는 점이 근거로 제시되었다. Step C는 주요안전기능 관점에서 위험상황의 중요도를 상향 정렬하는 단계로 요약될 수 있다.

Step D에서는 Step C까지의 추론 결과를 바탕으로 잠수함(정) 확장형작업분할구조를 활용하여 매핑함으로써 주요안전품목 후보군을 선정하였다.

Step A, B, C. D에서 LLM이 산출한 관련도 점수를 통합하여 총 317종 탑재장비를 관련도 기준으로 내림차순 정렬한 뒤 분포 특성을 분석하였다. 그 결과, 관련도 점수가 상위의 일부 품목에 집중되고 다수 품목은 상대적으로 낮은 값을 갖는 편중형 분포가 관찰되었다. 이는 주요안전품목 후보군이 탑재장비 전체에 균등하게 분산되기보다, 특정 핵심 품목군으로 수렴하는 경향임을 시사한다. 상·하위 품목군을 구분하기 위한 임계값을 설정하기 위해, 분포 곡선의 구조적 변화를 탐지하는 통계적 절곡점 분석 방법인 Elbow point 기법을 적용하였다(Hastie et al., 2009). Fig. 8의 위험도 분포에서 초기 구간은 기울기 변화가 크게 나타나다가 일정 지점 이후 감소율이 완만해지는 양상이 확인되었으며, 이때의 변곡 구간을 절곡점으로 식별하였다. 본 연구에서는 해당 절곡점을 기준으로 상위 구간을 상위 주요안전품목 후보군으로, 그 이하 구간을 일반 주요안전품목 후보군으로 구분하여 추적관리의 범위로 설정하였다.

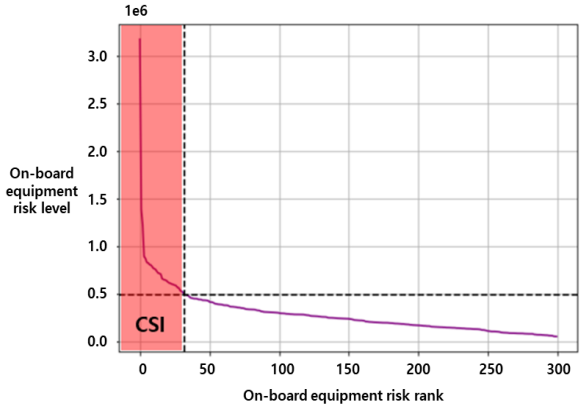

Fig. 8 Elbow point analysis result

3.5 설문조사

LLM 기반 잠수함 주요안전품목 도출 구조, 단계별 LLM 응답, 그리고 응답 근거의 적절성을 평가하기 위하여 설문조사를 실시하였다. 잠수함 건조, 작전운용, 사업관리, 품질관리 등 다양한 분야의 잠수함 관련 업무 종사자에 대한 의견수렴을 위하여 방위사업청, 해군, 잠수함 건조업체, 국방기술품질원을 설문조사 대상으로 선정하였고, 총 37명에 대한 설문조사를 실시하였다.

설문조사를 통하여 LLM을 통해 식별한 위험요인 및 위험상황

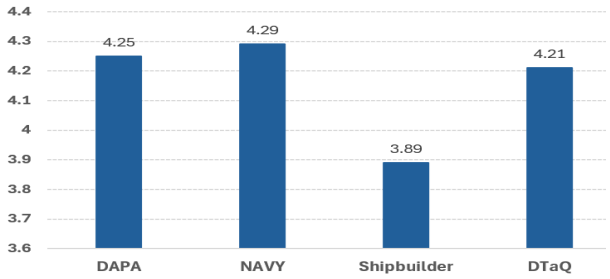

Fig. 9 Adequacy survey of LLM-identified risk factors

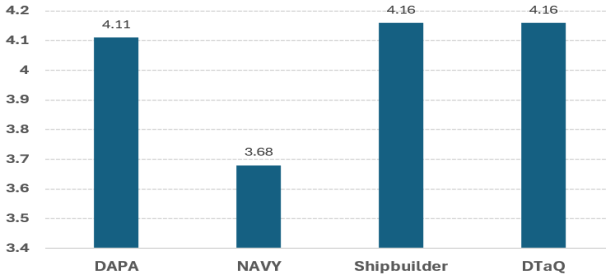

Fig. 10 Adequacy survey of LLM-provided evidence for risk factors, hazardous situations

의 적절성, 위험과 관련된 응답 결과를 뒷받침하기 위해 LLM이 제시하는 근거(NATO Naval Submarine Code, SUBSAFE, 잠수함 설계 문서, 출항불가조건, 품질 데이터 등)의 적절성에 대한 평가를 수행하였다. Fig. 9는 LLM을 통해 식별한 위험요인 및 위험상황의 적절성에 대한 설문조사 결과이며, Fig. 10은 응답 결과와 관련하여 LLM이 제시한 근거의 적절성에 대한 설문조사 결과이다.

설문조사 결과, 두 평가 분야의 평균 점수는 각각 5점 만점 기준 4.16점과 4.03점으로 나타나 LLM을 통해 식별된 위험요인, 위험상황의 적절성과 LLM이 제시한 근거의 적절성이 전반적으로 긍정적으로 평가되었음을 확인하였다. 특히 방위사업청(DAPA), 국방기술품질원(DTaQ)은 두 분야 모두에서 4.1점 이상의 평가 결과를 보이고 있으며, 해군(NAVY)은 위험요인 및 위험상황의 적절성에서 가장 높은 평가 결과를 나타내므로 사업관리, 작전운용 및 품질관리 관점에서 제안된 방법론의 활용 가능성이 높게 평가된 것으로 해석된다. 잠수함 건조업체(Shipbuilder)의 평가 점수가 상대적으로 낮게 나타난 것은 설계 및 건조를 직접 수행하는 기관의 특성상 세부 기술적 타당성에 대해 보다 보수적인 평가 기준을 적용한 결과로 판단된다. 따라서 본 설문 결과는 제안된 LLM 기반 방법론이 잠수함 주요안전품목 도출을 지원하는 객관적 보조 도구로 활용 가능함을 시사한다.

4. 기대효과

본 연구의 LLM 기반 주요안전품목 도출 프레임워크는 기존의 인간 전문가 중심 안전분석 절차가 갖는 주관성, 누락, 추적성 한계를 완화하면서도, 안전 설계의 핵심 전제인 인간 전문가 판단과 책임성을 유지하도록 설계되었다. Fig. 11은 이러한 차별성을 기존 방식과 본 논문을 통해 제안된 LLM 적용 방식의 차이점을 나타낸다. Fig. 11은 기존 방식과 LLM 기반 방식의 차이를 단순히 전문가 판단의 유무를 기준으로 한 주관성 및 객관성의 대비를 의미하지 않는다. 기존 방식에서도 잠수함 분야 인간 전문가의 검토와 합의는 주요안전품목 도출에 필요한 공학적 절차이나, 방대한 근거자료가 판단 결과와 명시적으로 연결되어 축적되지 않을 경우, 검토자별 경험 분야, 검토 범위 및 문서 접근성에 따라 해석 편차가 발생할 수 있다. 기존 방식은 설계 문서와 요구조건을 바탕으로 HAZID, FMECA 등 전문가 검토를 거쳐 위험요인, 위험상황을 도출하고, 이후 안전통제활동 대상을 선정하는 구조로서 인간 전문가의 경험과 프레이밍에 대한 의존도가 높다. 반면 LLM 적용 방식은 운용개념에 해당하는 ConOps (Concept of Operation) 관점에서 문서 전반을 포괄적으로 참조하면서, NATO NSC 기반 주요안전기능 정립, LLM 기반 프레임워크, 잠수함(정) 확장형작업분할구조 기반 탑재장비 수준 매핑, 근거 제시 및 검

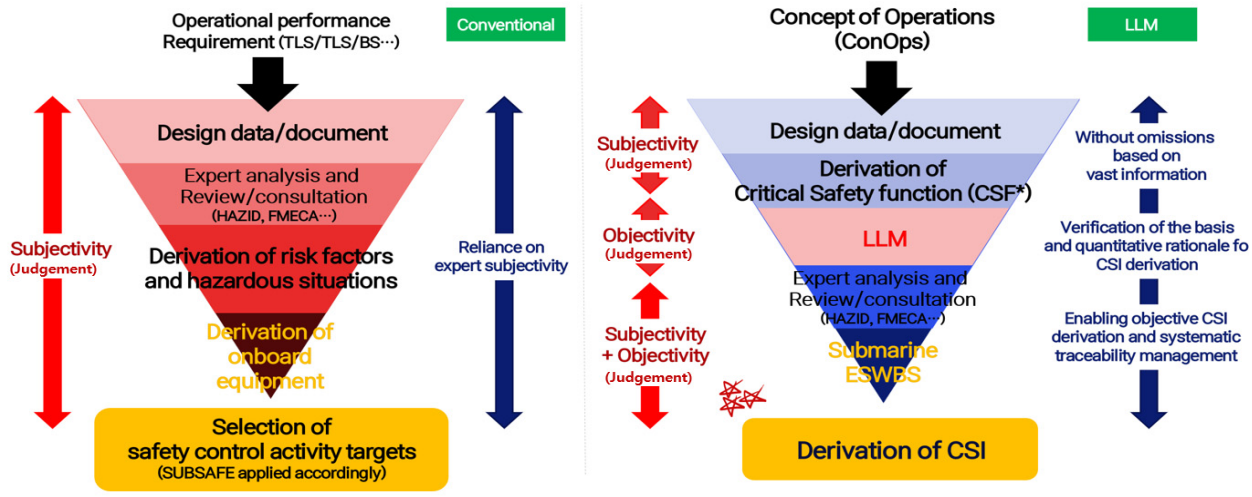

Fig. 11 Comparison of Conventional and LLM Processes

증절차를 결합하여 주요안전품목 선정의 객관성을 향상시키고, 추적성 중심의 체계로 전환하였다.

LLM과 RAG 기반 추론 및 검색을 적용하면 설계 문서, 품질 데이터, 사용자 불만 등 분산된 자료를 단일 파이프 라인에서 연계할 수 있어 검토 범위의 구조적 누락 가능성을 감소시킨다. 특히 잠수함 기술자료는 표, 그림, 도면 등을 근거로 하는 비중이 높다. ColPali 계열 아키텍처를 통한 페이지 단위의 시각적 맥락을 보존한 검색 체계를 통해 근거 탐색 단계의 재현성과 신뢰성을 강화하였다. 그 결과, 기존에는 특정 문서군에 편중되기 쉬웠던 위험 신호를 LLM을 적용함으로써 문서군 전체에서 균형 있게 포착할 수 있다. 아울러, 기존 방식에서는 심각도, 발생도, 검출도 등에 대한 판단이 인간 전문가 구성에 따라 달라질 수 있다. 본 연구는 이를 보완하기 위해 인간 전문가 기반 프롬프트 설계를 적용하고, 관련도 산정 기준을 명시적으로 정의하며, Step A, B, C, D와 같은 단계적 추론을 통해 연계 강도를 점진적으로 검증한다. 이 구조는 결과가 합의의 산물로만 남지 않도록 하며, 동일 입력에서 일관된 주요안전품목 후보군 생성과 규칙 기반 평가가 가능하도록 한다. LLM이 생성한 주요안전품목 후보군에 대한 근거 문서, 인용 구절, 판단 연결고리를 원문 단위로 제시하도록 함으로써 LLM이 내린 결론이 역추적 가능하도록 하였다. 이는 설계 변경, 부품 대체, 운용 결함 누적 등 환경 변화가 발생했을 때 기존 결론을 신속하게 재평가할 수 있게 하며, 검증 관점에서 '왜 해당 품목이 주요안전품목인가'에 대한 설명 가능성을 높인다. 방대한 정보를 기반으로 누락 없는 근거 제시, 도출 근거의 검증 및 합리성 제시, 체계적 추적성 관리가 가능하도록 하는 것이 핵심 기대효과다.

주요안전품목 도출의 기준점인 주요안전기능을 NATO NSC에 근거하여 정립하고, 이를 인간 전문가가 직접 확정함으로써 '안전 우선' 설계 철학을 LLM 추론 과정에서 상위 제약조건으로 고정한다. 결과적으로 LLM은 방대한 문서에서 위험 신호와 운항 안전성 간의 연계를 탐색하되, 핵심 기준은 인간 전문가 판단에 의해 안정적으로 유지된다. 이를 통해 LLM이 과도한 해석, 추론으로 결론을 왜곡할 수 있는 가능성을 구조적으로 줄이도록 하였다.

Elbow point 기법을 통해 상위 주요안전품목 후보군과 일반 주요안전품목 후보군을 정량적으로 분리하여 어떻게 주요안전품목을 효과적으로 관리할 것인가에 대해 주관적 논의가 아닌 데이터 기반 규칙으로 의사결정을 지원하였다. 이는 후속 검증, 추적 및 분석, 품질검사 자원의 집중 대상을 합리적으로 설정할 수 있게 한다. 마지막으로, 안전 분야는 결론의 책임성과 오판 위험성 관리가 핵심이다. 본 연구는 LLM을 기존 방식의 대체가 아니라 일관된 1차 분석 및 인간 전문가의 기술검토를 체계적으로 지원하는 역할을 하도록 설정하였다. 전문가 검토, 확정에 해당하는 Human-in-the-loop 검증 절차를 결합하여 최종 판정의 책임성을 유지하도록 하였다. 즉, 처리 속도, 일관성, 누락 방지, 미세 상관관계 포착이라는 LLM의 강점을 활용하면서도, 환각, 편향 가능성 등과 같은 약점은 근거 제시, 인간 전문가 확정의 폐루프 구조로 제어하는 방식으로 억제하였다.

5. 결 론

본 연구는 잠수함 분야에서 고도의 신뢰성과 책임성이 요구되는 주요안전품목 도출 방안을 다루고 있다. LLM 및 RAG 기법을 활용하여 근거 기반 추론 프레임워크를 제안하였다. 기존 주요안전품목 도출 방식은 대량 문서를 인간 전문가가 수동으로 검토 및 해석하는 구조를 중심으로 하며, 그 결과 도출 과정이 노동집약적이고 특정 전문가의 경험과 프레이밍에 의존하기 쉽다. 또한 시간의 경과, 담당 인력의 교체, 기술 축적 수준의 변화에 따라 선정 기준이 미세하게 변동하거나, 문서 누락 또는 편중으로 인해 결론의 재현성과 일관성이 약화될 수 있다.

이에 비해 본 연구의 접근은 대량 문서에 대한 자동 분석을 통해 편향과 누락 가능성을 낮추고 재현성을 강화하는 방향으로 설계되었다. 이는 단순히 '생성' 자체를 고도화하는 것이 아니라, 안전 분석 품질을 좌우하는 핵심 요소인 근거 탐색-연계 판단-후보 도출의 전 과정을 구조화한다는 점에서 의미가 있다.

또한 LLM을 활용한 RAG 기반의 프레임워크를 설정함으로써 주요안전품목 도출 절차를 체계화, 준자동화된 프로세스로 전환하였다. 주요안전품목 도출 근거 제시 체계를 통해 이해관계자 관점에서 주요안전품목 선정의 타당성을 검증 가능하게 하여 결과의 수용성과 신뢰성을 높이는데 기여하였다.

Fig. 12는 주요안전품목 관리 방식이 기존의 문서 기반·전문가 경험 중심 관리에서 LLM 기반 지식 축적형 관리로 발전하는 개념을 나타낸다. 기존 방식에서는 형상이 진전되거나 담당 인력이 변경될 경우, 주요안전품목의 선정 기준, 근거 문서, 판단 배경이 분산될 가능성이 있다. 이 경우 동일한 장비라 하더라도 검토 시점이나 검토자에 따라 위험 중요도 판단이 달라질 수 있으며, 과거 판단 결과를 재검토하기 위해 상당한 시간이 소요될 수 있다. 반면 LLM 기반 접근 방식에서는 위험요인, 위험상황, 주요안전기능, 주요안전품목 후보군, 근거 문서, 판단 기준이 연결된 형태로 축적된다. 따라서 특정 품목이 주요안전품목 후보로 도출된 이유를 원문 근거와 단계별 판단 결과를 통해 역추적할 수 있으며, 설

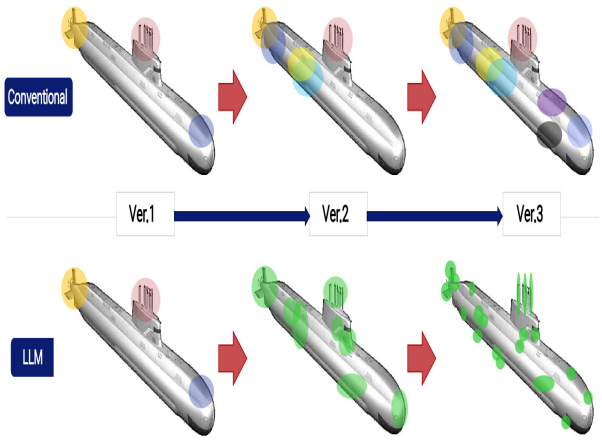

Fig. 12 Comparison of CSIs management evolution

계 변경, 부품 대체, 운용 결함 누적, 표준 개정과 같은 변화가 발생하더라도 기존 판단 결과를 갱신 가능한 지식 기반으로 재평가할 수 있다. Fig. 12에서 제시한 관리 진화의 핵심은 LLM이 최종 판단을 자동으로 대체한다는 의미가 아니라, 주요안전품목 도출 과정에서 발생하는 판단 근거와 검토 이력을 지속적으로 축적하여 인간 전문가의 최종 검토를 지원하는 데 있다.

LLM 답변 추론결과인 약 546,825개는 추적 가능하고, 각각의 추론결과의 근거에 해당하는 LLM 입력 문서를 원문 단위로 추적 가능하다. 이는 객관적인 주요안전품목 도출에 활용 가능할 뿐만 아니라, 잠수함 관련 업무 종사자의 기술검토, 품질문제 원인분석, 차기 잠수함 개선사항 식별, 계약요구조건 불만족 여부 등 다양한 분야에서 활용 가능하다. 본 연구는 전문가 경험 의존과 노동집약적 절차의 한계를 완화하고, LLM 및 RAG 기법을 활용하여 주요안전품목 도출을 체계화하여 자동화하여 인간 전문가의 판단을 객관적으로 보조하도록 하였다. 다만 제안하는 LLM 기반 방법론은 LLM 모델 자체의 일반 성능을 독립적으로 검증하는 것을 목적으로 하지 않는다. 본 연구의 초점은 잠수함 주요안전품목 도출이라는 특정 목적에 대해 LLM 및 RAG 기반으로 생성된 산출물이 원문 근거, 주요안전기능, 위험요인, 위험상황 및 탑재장비 간의 기능적 연결성과 정합성을 갖는지 확인하는 데 있다. 즉, 본 연구에서 제시한 방법론은 주요안전품목 도출 절차의 복잡도를 높이는 것이 아니라 Human-in-the-loop 검증 구조 내에서 LLM 추론 결과의 근거 적절성, 단계별 산출물 간 일관성, 기존 안전통제활동 대상과의 정합성을 검토하는 절차라고 할 수 있다. 이를 통해 주요안전품목 도출 절차의 복잡도를 과도하게 증가시키지 않으면서도, LLM 결과가 인간 전문가의 최종 판단을 객관적으로 보조하도록 하였다. 궁극적으로 잠수함 운항 안전, 감항성 관리, 그리고 나아가 감항인증과 연계되는 안전관리 역량을 제고하며, 국제적 수준의 안전성이 확보된 '더 안전한 한국형 잠수함' 전력 확보에 기여할 것으로 기대된다.

향후 연구에서는 본 연구결과의 장기적 적용을 통해 LLM 입력 데이터 및 문서 갱신에 따른 LLM 추론 답변의 민감도, 문건 누락 또는 입력자료 범위 변화에 따른 RAG 기반 근거 검색 결과의 안정성, 프롬프트 구성 방식 및 근거 제시 방식에 따른 LLM 추론 결과의 일관성, Human-in-the-loop 검증 절차의 최적화 등을 추가로 연구할 필요가 있다. 또한 본 연구는 Table 3의 프롬프트 자체를 다양한 프롬프트 구성과 비교 검증한 것은 아니므로, 향후 잠수함 주요안전품목 도출에 적합한 프롬프트 설계 기준을 체계적으로 정립할 필요가 있다. 이를 바탕으로 NATO NSC 등 국제적 수준의 안전 표준과의 정합성을 향상시키고, 주요안전품목 관리 자동화 체계로 고도화할 필요가 있다.

후 기

본 논문은 2025년 대한조선학회 추계학술대회(신진연구자 기획세션)에서 발표되었습니다.

References

Archana, T.R., Anirudh, P.B., Ryan, T., White, V.M.N., Olivia, J.P.F. and Dimitri, N.M., 2023. Examining the potential of generative language models for aviation safety analysis: Case study and insights using the Aviation Safety Reporting System (ASRS). *MDPI Aerospace*, 10(9), pp.770.

Balahari, V.B., Florian, G., Francesco, C., Joao-Vitor, Z., Josef, J., Nuria, M. and Reinhard, S., 2025. Towards automated safety requirements derivation using agent-based RAG. *AAAI Spring Symposia Computer Science Engineering*.

Christopher, T., 2009. Submarine accidents, a 60-year statistical assessment. *Journal of the American Society of Safety Professionals,* 54(9), pp.31-39.

Hastie, T., Tibshirani, R. and Friedman, J., 2009. *The elements of statical learning: data mining, inference, and prediction*. Springer Science & Media.

Lee, Y.S., 2024. A study on the seaworthiness management of korean submarine based on naval submarine code (NSC) of INSA. *Naval Seaworthiness Seminar.*

Lu, S., Bin, Q., Jiarui, L., Yang, Z., Zhanzhao, L., Zhaowei, G., Wenke, D. and Lin, S., 2024. Aegis: an advanced LLM-based multi-agent for intelligent functional safety engineering. *Conference on Empirical Methods in Natural Language Processing: Industry Track*.

Manuel, F., Hugues, S., Tony, W., Bilel, O., Gautier, V., Céline, H. and Pierre, C., 2025. ColPali: efficient document retrieval with vision language models. *Conference paper at ICLR.*

Richard, D., Delpizzo, CAPT. and Sharat, V., 2017. An introduction to NATO standard ANEP (Allied Naval Engineering Publication) 77 and its application to naval ships. *Ship Science & Technology,* 11(21), pp.75-88.

Sullivan, P.E., Iwanowicz, S.E., Ford, A.H., McBride, M.S., Guinan, T.K. and Lilly, C.A., Naval sea systems command submarine safety (SUBSAFE) program, 2025. *AIChE Spring National Meeting Conference Proceedings*.

Xu, R., Jiang, P., Luo, L., Xiao, C., Cross, A., Pan, S. and Yang, C., 2025. A survey on unifying large language models and knowledge graphs for biomedicine and healthcare. *In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*.

Xiaoyue, X., Xilang, T., Siwei, G. and Lijie, C., 2025. An intelligent guided troubleshooting method for aircraft based on hybrid RAG. *Scientific Reports*, 15, Article number: 17752.

Authorship Contribution Statement

Ho-Seong Chang: Project administration, Framework design, Identification of target documents for learning, Selection of critical safety functions based on NATO NSC standards; **Yeon-Dong Park:** Investigation, Collection of target documents for learning, Establish security measures for GPU environments; **Jae-Hyuk Kum:** Development of AI services for the national defense network, Building an AI-RAG data pipeline; **Gyeong-Pil Do:** User UI development, Vector DB construction, Survey and results analysis; **Ki-Hun Kim:** LLM model implementation, Project execution management.

장 호 성

박 연 동

금 재 혁

도 경 필

김 기 훈