Journal of Institute of Control, Robotics and Systems (2025) 31(8):858-867
https://doi.org/10.5302/J.ICROS.2025.25.0152 ISSN:1976-5622 eISSN:2233-4335

휴머노이드 로봇 안전성 확보를 위한 새로운 위험성 평가 프레임워크

A Novel Risk Assessment Framework for Ensuring the Safety of Humanoid Robots

류 요 엘*, 전 진 우

(Joel Ryu[1,*] and Jinwoo Jun[1])

[1]Center of Humanoid Robotics, Korea Institute for Robot Industry Advancement (KIRIA)

Abstract: Humanoid robots mimic human forms and utilize advanced artificial intelligence (AI), creating complex safety challenges that traditional frameworks struggle to address. This paper proposes an integrated risk assessment framework tailored for humanoid robots, combining Failure Modes and Effects Analysis, System-Theoretic Process Analysis, and an AI-specific Risk Index. Additionally, it systematically identifies and evaluates risks from mechanical and electrical failures, control system errors, and AI-driven behaviors such as biased data handling and unpredictable real-time decision-making. Applying this framework to 30 risk scenarios revealed critical vulnerabilities, including power system malfunctions, software logic errors, and autonomous decision anomalies. Based on these findings, specific safety measures were developed, emphasizing redundant hardware controls, multi-sensor validation, real-time software monitoring, and robust fail-safe mechanisms. The integrated approach significantly enhances traditional risk assessments, effectively bridging the gap between existing robotics safety standards and new AI-driven challenges, ensuring the safe and reliable operation of humanoid robots in dynamic human environments.

Keywords: humanoid robot, risk assessment, robot safety, physical AI, AI safety

I. 서론

인간의 일상생활 및 산업현장에 로봇기술이 본격적으로 도입됨에 따라, 휴머노이드 로봇은 인간과 유사한 형상과 동작을 통해 작업 효율성 향상, 재난 구조, 서비스 분야 등 다양한 영역에서 활용 가능성을 인정받고 있다. 특히 인공지능 기술이 결합된 휴머노이드 로봇은 점차 복잡한 환경에서 자율적으로 의사결정을 수행하고 사람과 긴밀히 상호작용하게 되면서, 로봇 안전성에 대한 사회적 및 기술적 요구가 더욱 증대되고 있다[1]. 기존의 산업용 로봇은 안전펜스나 구역 제한을 통해 물리적으로 분리된 환경에서 주로 운용되었으나, 휴머노이드 로봇은 산업현장에서 협업에서 사람과 직접 마주하는 가정 환경까지 활동 범위를 넓혀가는 만큼, 물리적 기능과 인간-로봇 상호작용을 포괄하는 통합적 안전 확보가 핵심 과제로 부상하고 있다. 또한 최근 중국에서 발생한 원인이 규명되지 않은 휴머노이드 로봇 오동작 사고가 안전성에 대한 관심을 높이고 있다[2,3].

휴머노이드 로봇의 안전성 문제는 단순히 기계적인 사고 예방이나 하드웨어의 견고성 확보만을 의미하지 않는다. 예기치 못한 제어 오류, 센서 신뢰성 저하, 자율이동·의사결정 알고리즘의 한계, 그리고 더 밀접해진 인간-로봇 상호작용 과정에서 발생할 수 있는 다양한 위협 요인 등이 종합적으로 고려되어야 한다. 머리나 팔다리 등 인체 유사 구조를 지닌 휴머노이드 로봇은 전도 시 충격이 크거나 관절 구동계의 고장으로 갑작스러운 운동 실패가 예상 외의 위험을 초래할 수 있다. 더 나아가 인공지능 기반 학습 기능이 적용되면 로봇이 스스로 행동을 변경하거나 인지 방식을 유동적으로 바꾸므로, 전통적 하드웨어 중심의 안전 대응만으로는 부족해질 수 있다.

본 연구는 인공지능이 탑재된 휴머노이드 로봇의 안전성 평가를 위해 기존 위험성 평가 기법을 확장한 통합 프레임워크를 제안하고, 이를 실제 사례에 적용하여 그 효과를 검증하는 것을 목적으로 한다.

첫째, 휴머노이드 로봇의 특수성을 고려하여 물리적 하드웨어 결함, 자율제어 알고리즘, 인간-로봇 상호작용에서 발생 가능한 위험요인을 시나리오를 기반으로 체계적으로 도출하였다. 둘째, 기존에 널리 사용되는 FMEA (Failure Modes and Effects Analysis)와 STPA (System-Theoretic Process Analysis)를 검토하여, 이 두 가지 기법의 장점을 결합하고, 추가적으로 AI의 자율성으로 인한 위험요인을 평가하는 ARI (AI Risk Index)를 도입하여 새로운 통합 프레임워크를 제안하였다. 셋째, 실제 상용화된 휴머노이드 로봇의 사례를 바탕으로 본 프레임워크를 적용하여, 구체적인 위험 시나리오를 도출하고 이를 평가함으로써, 프레임워크의

*Corresponding Author
Manuscript received May 11, 2025; revised July 9, 2025; accepted July 18, 2025
류요엘: 한국로봇산업진흥원 책임연구원(joelryu@kiria.org, ORCID 0009-0004-5708-9971)
전진우: 한국로봇산업진흥원 수석연구원(jzinu@kiria.org, ORCID 0000-0002-2951-0085)
※ 본 논문은 정부(산업통상자원부)의 재원으로 한국로봇산업진흥원(지능형로봇 보급 및 확산사업)의 지원을 받아 수행된 연구.

유효성과 실용성을 검증하였다.

본 논문의 이러한 접근은 휴머노이드 로봇이 인간과 밀접하게 공존하는 다양한 환경에서 직면할 수 있는 안전사고를 예방하고, 기술적 결함을 효과적으로 완화하며, 윤리적·사회적 요구사항을 포괄적으로 대응할 수 있는 실질적이고 융합적인 안전대책을 제시하는 데 기여할 것이다. 또한, 향후 휴머노이드 로봇의 안전 표준 개정 및 관련 분야의 연구에 실무적으로 활용될 수 있는 선행 기초자료로서 의의를 가진다.

II. 관련 표준·규제 동향

1. 산업용 로봇 안전 표준(ISO 10218 시리즈)

산업용 로봇의 안전 요구사항을 다룬 ISO 10218 시리즈는 전통적으로 제품 안전, 기계적 보호, 작동 중 위험요소 관리 등에 초점을 두고 있다[4,5]. 휴머노이드 로봇은 기존 근로자의 작업 영역에 작업 환경에 투입될 가능성이 높으므로, 기본적인 기계적 안전 및 제어 시스템 보호 조치를 ISO 10218 지침과 연계하여 고려해야 한다. 다만, 산업용 로봇이 전제하는 "제한 구역 및 안전펜스 내 작업"이나 "프로그래밍된 작업" 관점이 휴머노이드 로봇의 자율적이고 협동적인 특성과 상충되는 측면이 있으므로, 현행 표준을 그대로 적용하기보다는 일부 항목을 확장하거나 보완해야 한다는 지적이 제기되고 있다.

2. 서비스 로봇 안전 표준(ISO 13482)

ISO 13482는 인간이 사용하는 공간에서 동작하며, 직접적인 상호작용이 발생하는 서비스 로봇의 안전 요구사항을 정의한다[6]. 휴머노이드 로봇을 다양한 서비스 환경에서 활용하는 사례가 증가함에 따라, ISO 13482가 제시하는 위험 분석, 설계, 운용 가이드라인은 일부 참조될 수 있다. 다만, ISO 13482는 2014년 제정 이후 개정 작업이 진행 중이며, 휴머노이드 로봇의 자율제어나 AI 학습 과정에서 발생할 수 있는 예외 상황에 대해서는 구체적인 참조 내용은 아직 미비한 수준이다.

3. AI 국제 표준화 동향(ISO/IEC JTC1 SC42)

휴머노이드 로봇의 지능적 동작과 자율성을 규정하기 위해서는, 전통적인 기계안전 규범뿐 아니라 인공지능(AI) 기술 전반을 다루는 표준화 논의도 주목할 필요가 있다. ISO/IEC JTC1 SC42(이하 SC42)는 AI 시스템의 신뢰성, 윤리, 리스크 관리, 거버넌스 등에 관한 기초 표준을 제정하고 있다[7,8].

SC42는 AI 시스템 개발 및 운영 전주기에 걸쳐 데이터 품질, 알고리즘 투명성, 편향성 방지, 안전성 평가 등 핵심 요소를 폭넓게 다룬다. 휴머노이드 로봇과 같이 고도의 자율성을 갖춘 기기에도 이러한 요소들이 복합적으로 적용될 수 있다. 예컨대 국제표준 ISO/IEC 23894에서 제시되는 AI 리스크 프레임워크는 휴머노이드 로봇의 안전성 평가 프로세스와 밀접히 연계될 수 있다는 점에서 중요한 시사점을 제공한다.

다만 SC42의 표준들은 AI 기술을 총망라하기 때문에, 구체적인 로봇 플랫폼이나 하드웨어 안전기준은 직접적으로 제시하지 않는다. 따라서 휴머노이드 로봇 안전을 위해서는 SC42에서 제안하는 AI 위험관리 체계를 수용함과 동시에, 휴머노이드 로봇 특유의 기계·전기·제어적 이슈를 아우르는 기준을 추가로 마련할 필요가 있다. 본 논문에서 제안하는 평가 프레임워크는 SC42의 AI 관리 지침을 참조하여 위험 분석 시 고려할 요소를 확장한다.

4. AI/자율성 안전 규제 및 해외 동향

최근 휴머노이드 로봇의 주행·조작 영역에서 AI 기술이 적용됨에 따라, 전통적 기계안전 규범만으로는 해석하기 어려운 위험 사례가 대두되고 있다. 예컨대 실시간 학습 알고리즘이 예측과 달리 갑작스럽게 동작 패턴을 수정하거나, 센서 결함으로 인해 잘못된 인지 정보를 바탕으로 움직일 때 발생하는 위험은 기계적 안전 표준의 범위를 넘어선다. 이에 관해 미국 경우 ANSI/RIA [9]나 OSHA 등에서는 산업용 로봇의 안전 가이드라인을 지속적으로 개정하고 있으나, AI 수준이 높은 휴머노이드 로봇을 구체적으로 규정하는 법적 장치나 국제표준은 아직 미비하다는 지적이 있다. 유럽연합(EU)도 기계류 지침(machinery directive)의 개정을 추진하며 로봇안전, 특히 자율주행 기계 장치에 대한 법적·기술적 요건을 강화하고 있지만, 휴머노이드 로봇을 별도로 명시하는 경우는 드문 실정이다. 법적으로는 EU의 AI Act [10], 국내에는 내년 시행예정인 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법과도 연계되는데, 휴머노이드 로봇의 구체적인 적용 방안은 뚜렷하지 않다. 기존 산업현장의 규제인 산업안전보건법에서 휴머노이드 로봇을 어떻게 적용할지 논의가 필요한 시점이다.

5. 시사점

휴머노이드 로봇에 대한 안전성 평가 프레임워크를 구축할 때는 ISO 10218, ISO 13482 등 기존 표준의 틀을 존중하되, 자율제어와 학습 알고리즘을 포함한 AI 안전성 이슈를 보완·확장할 수 있는 지침을 추가로 개발해야 한다. 본 연구는 이러한 동향을 토대로 위험요인을 다각적으로 분류하고, 기존 표준에서 다뤄지지 못해 구체적 적용 방법이 명시되지 않은 부분까지 포괄하는 평가 방법론을 제시하고자 한다. 이를 통해 휴머노이드 로봇이 향후 실생활이나 산업 현장에서 사람과 긴밀히 협력하더라도 안전사고를 최소화할 수 있는 설계·운영 기반을 마련하는 것이 본 논문의 궁극적 목표이다.

III. 위험성 평가 프레임워크

1. 위험성 평가 개요

휴머노이드 로봇은 기존의 산업용 혹은 서비스 로봇과 달리, 인간형 구조와 고도의 자율성을 갖추고 있어 물리적·전기적 위험뿐 아니라 복합적인 알고리즘 위험까지 포괄적으로 다루어야 한다[11]. 이러한 복합성을 체계적으로 분석하기 위해서는 전통적인 하드웨어 중심의 안전성 평가 기법과 시스템 이론적 접근을 결합한 종합적 방법론이 필요하다. 본 장에서는 대표적인 공정·제품 안전분석 방식인 FMEA와 시스템 이론적 관점에서 제어구조상의 위험을 식별·분석하는 STPA를 살펴보고, 이를 휴머노이드 로봇 및 AI 자율성 평가 관점에 맞춰 확장하는 방안을 제시한다.

2. FMEA 위험성 분석

fmea는 제품 또는 공정에서 발생 가능한 잠재적 고장 모드(failure mode)를 사전에 식별하고, 심각도(severity), 발생 빈도(occurrence), 검출 가능성(detection)을 평가하여 위험 우선순위 번호(rpn, risk priority number)를 산출하는 위험성 분석 기법이다[12,13]. FMEA를 통한 위험성 분석 절차는 먼저, 제품의 분석 대상을 정의하고 그에 따른 제품 기능과 정보를 수집한다. 이후 정리된 정보를 바탕으로 잠재적 고장과 이로 유발된 위험요인과

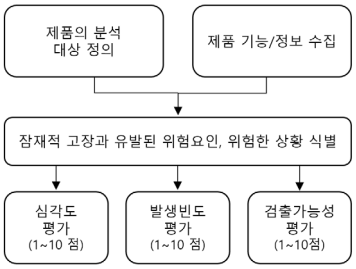


그림 1. FMEA 기반 위험성 분석 프로세스.

Fig. 1. Risk analysis process based on FMEA.

위험한 상황 등을 식별한다. 식별된 각 고장이 미칠 영향과 원인을 추적하여 심각도(S), 발생 빈도(O), 검출 가능성(D) 각 요소는 1점에서 10점까지의 척도로 평가한다.

휴머노이드 로봇의 특이점으로 일반 기계장치에 대한 FMEA와 달리, 관절 수가 많아 설계가 복잡하고 동작 모드가 다양하여 고장 시나리오가 폭넓게 분포한다. 또한, 인간과 유사한 구조 특성상, 전도 시 관절에 따라 충격량이 다르거나, 모터 고장 패턴들이 상호 영향을 미칠 수 있다. 따라서 단순히 단일 부품 수준에 그치지 않고, 조인트 연계 동작을 포함한 시스템 전체의 작동 흐름을 세밀히 분석해야 한다.

3. STPA 위험성 분석

STPA는 시스템 이론적 접근인 STAMP (System-Theoretic Accident Model and Processes)에 기반한 위험성 분석 기법으로, 전통적인 FMEA가 개별 구성요소 고장에 집중하는 것과 달리 제어 구조 및 상호작용의 오류에서 비롯되는 위험을 중점적으로 다룬다[14,15]. STAMP에서는 안전사고를 단순한 구성요소 결함이 아니라 부적절한 제어 명령에서 발생하는 시스템적 문제로 본다. 즉, 센서·제어기·구동기가 서로 주고받는 명령과 피드백이 올바르지 않은 상태로 전달·처리될 때, 사고 가능성이 커진다고 가정한다.

STPA 위험성 분석 적용 절차는 먼저 시스템의 기능과 목표(목적)을 정의한다. 다음으로 시스템의 제어구조를 도식화하여 주요 제어 루프를 정의한다. 제어부에서 연결되는 구동부, 센서부 등으로 연결되는 루프를 고려하는 것이다. 각 제어 명령의 흐름에서 발생할 수 있는 UCA (Unsafe Control Action)를 식별한다. 식별된 UCA가 발생하는 원인(제어기 소프트웨어 오류, 센서 오인식, 인간-로봇간 인터페이스 결함 등)을 추적하며, UCA를 표 1과 같이 결정한다.

휴머노이드 로봇 적용 시 장점으로는 STPA는 복잡한 상호작용이 빈번한 휴머노이드 로봇이 의도치 않은 행동을 하거나, 인간과의 실시간 협업 과정에서 변수(인간의 행동, 환경

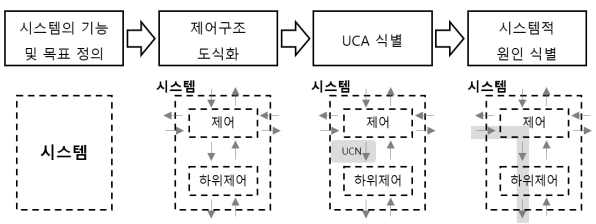


그림 2. STPA 기반 위험성 분석 프로세스.

Fig. 2. Risk analysis process based on STPA.

표 1. 부적절한 제어명령에 따른 위험성 평가 기준.

Table 1. Risk level criteria based on unsafe control actions.

UCA 위험 수준	기준 조건
Severe	- 빈번히 발생 가능하며, 사고로 이어질 가능성이 높거나, 발생 확률은 낮지만 치명적 결과가 예상되는 경우
Moderate	- 가끔 발생 가능하며, 중대하지는 않으나 반복되면 사고 가능성이 있는 경우
Limited	- 매우 드물게 발생하며, 사고 가능성과결과가 미미한 경우

변화)가 동적으로 바뀌는 상황을 분석하기에 유리하다. 또한 AI 알고리즘이 제어 루프 일부를 실시간 학습으로 업데이트할 때, 이 과정에서 발생할 수 있는 제어 명령의 예측 불능 문제도 식별 가능하다.

4. AI 특화 위험요인 평가

휴머노이드 로봇에 AI 기술이 접목되면, 기존 기계적·제어적 위험 외에도 알고리즘적 위험이 추가된다. 대표적인 AI 리스크로는 학습 데이터의 편향성, 예측 모델의 부정확성, 실시간 상황인지 오류, 장기 운용 중 누적되는 소프트웨어 오류 등을 들 수 있다. SC42 표준에서 제시하는 AI 위험관리 프레임워크나 ISO/IEC 23894 등을 참조하여 표 2와 같이 데이터 품질(Q), 알고리즘 신뢰성(A), 실시간 업데이트(R), 윤리사회적(E) 평가 요소로 세분화한 평가항목으로 정리할 수 있다[16,17].

표 2를 활용한 위험요인 평가 방법은 먼저, 표 2의 데이터 품질, 알고리즘 신뢰성, 실시간 업데이트, 윤리·사회적 고려

표 2. 인공지능 기반 위험요인의 평가 지표.

Table 2. Evaluation metrics for AI-induced risks.

평가요소			1점	2점	3점	4점
데이터 품질 (Q)	(q_1)	데이터 다양성	≧ 6	4~5	2~3	1
	(q_2)	데이터 편향성	< 50%	50~65%	65~80%	> 80%
	(q_3)	Outlier 비율	< 5%	5~10%	10~15%	≥ 15%
알고리즘 신뢰성 (A)	(a_1)	모델 복잡도	투명성 확보	단순 구조	부분적 해석	해석 불가
	(a_2)	평균 에러율	< 5%	5~10%	10~15%	≥ 15%
	(a_3)	제어루프 연동성	자동 연동	대부분 연동	부분 연동	연동 불가
실시간 업데이트 (R)	(r_1)	실시간 학습여부	시스템 레벨	모듈 레벨	업데이트 수준	미적용
	(r_2)	Sim2Real 관리	자동 보정	수동 보정	모니 터링	미적용
윤리 사회적 (E)	(e_1)	민감정보 처리여부	완전 구현	설계 적용	문서 수준	미적용
	(e_2)	악의적 접근여부	다단계 접근제한	단일	단일 접근제한	미적용
	(e_3)	오남용 사고우려	실시간 대응체계	체계적 절차	문서 수준	미적용

요인의 세부지표에 따라 1~4점씩 평가하고, Q, A, R, E를 세부지표의 평균값으로 산출한다. Q, A, R, E의 평균값은 소수점 첫째 자리에서 반올림하여 정수로 산출한다. 지표에 따라 가중치를 조정하여 활용할 수 있다. 이후 총점 ARI (AI Risk Index)를 아래와 같이 합산한다.

$$ARI = Q + A + R + E \quad (1)$$

where

$$Q = \frac{1}{n_Q}\sum_{i=1}^{n_Q} q_i,\ A = \frac{1}{n_A}\sum_{i=1}^{n_A} a_i,\ R = \frac{1}{n_R}\sum_{i=1}^{n_R} r_i,\ E = \frac{1}{n_E}\sum_{i=1}^{n_E} e_i$$

5. 위험성 결정

5.1 FMEA 기반 위험성 결정

FMEA의 분석 결과를 바탕으로 수식 (2)와 같이 심각도(S), 발생빈도(O), 검출가능성(D)의 곱으로 RPN (Risk Priority Number)을 구한다. 이때 산출된 RPN은 표 3의 조건에 따라 위험 수준을 200 이상을 High(H), 80~199 사이를 Medium(M), 80 이하를 Low(L)로 결정한다. 추가적으로 FMEA의 세부지표 심각도가 8 이상을 High(H)로, 심각도가 5~7이면서 발생빈도가 5 이상일 때를 Medium(M), 심각도가 4 이하인 경우와 동시에 발생빈도가 4 이하의 경우를 Low(L)로 위험성을 결정한다.

$$RPN = S(Severity) \times O(Occurrence) \times D(Detection) \quad (2)$$

5.2 STPA 기반 위험성 결정

AI 자율성 관련 위험요인인 A1~A15는 STPA의 분석 결과인 UCA와 AI 특화 위험요인의 가중치(ARI)를 종합적으로 평가하여 그림 3과 같이 High(H), Medium(M), Low(L)으로 위험성을 결정한다.

표 3. FMEA 기반 위험 수준 분류 기준.

Table 3. Risk classification criteria based on FMEA.

위험 수준	기준 조건
High (H)	- RPN ≥ 200 or - Severity ≥ 8
Medium (M)	- RPN 80~199 or - Severity = 5~7 & Occurrence ≥ 5
Low (L)	- RPN < 80 and - Severity ≤ 4 & Occurrence ≤ 4

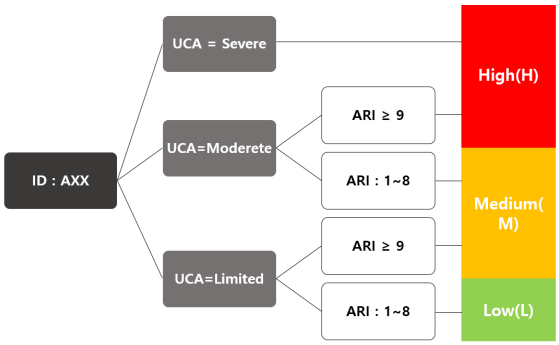


그림 3. STPA 및 ARI 지표 기반의 위험 수준 결정 체계.

Fig. 3. Risk evaluation scheme based on STPA and ARI metrics.

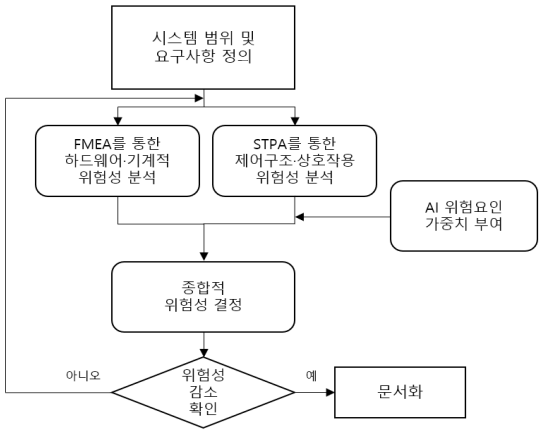


그림 4. 통합적 위험성 평가 절차도(FMEA-STPA-AI).

Fig. 4. Integrated risk assessment procedure (FMEA-STPA-AI).

STPA 분석 결과 UCA가 Severe로 평가된 시나리오는 ARI 값과 무관하게 최종 등급을 High(H)로 결정한다. UCA가 Moderate 수준인 경우에는 ARI에 따라 8 이상일 경우 High(H), 그 외의 경우를 Medium(M)으로 결정한다. UCA가 Limited인 경우에는 ARI가 9 이상일 경우 Medium(M)으로, 그 외의 경우를 Low(L)로 결정한다.

6. 위험성 평가 프레임워크 제안

위와 같이 FMEA와 STPA, 그리고 AI 특화 위험요소 평가 방안을 결합하여 휴머노이드 로봇 통합 위험성 평가를 수행하려면, 다음과 그림 4와 같은 절차가 유효하다.

이와 같은 절차를 따르면 전통적 하드웨어 위험성 평가 기법과 시스템 이론적 접근, AI 위험관리 방안을 조화롭게 결합할 수 있다. 특히 휴머노이드 로봇의 고밀도 관절구조, 실시간 제어 알고리즘, 인간과의 직접 협업 등 특수성을 충분히 반영하여, 단순 구성요소 분석과 제어 명령 분석, 그리고 알고리즘 위험성 분석을 병행함으로써 누락 없는 안전성 평가가 가능해진다. 다음 장에서는 이러한 방법론을 구체적 사례에 적용하고 도출된 결과를 논의함으로써, 제안 기법의 실효성과 적용 가능성을 검증하고자 한다.

IV. 위험성 평가 적용 및 결과

본 장에서는 앞서 제안된 휴머노이드 로봇 위험성 평가 프레임워크를 실제 상용화된 모델과 매뉴얼을 기반으로 로봇 안전 전문가 그룹이 적용한 결과를 정리한다. FMEA와 STPA+AI 기법을 통해 도출된 위험 시나리오들의 분석 결과와 최종 위험 등급 산정 내용을 각 절에서 다루었다.

1. FMEA 분석 결과 및 위험성 결정

먼저 FMEA 기법을 적용하여 휴머노이드 로봇의 주요 구성요소별 잠재적인 고장 모드와 그 영향을 분석하였다. 잠재적 고장은 제어, 구동, 센서, 전원, 통신으로 구분하여 접근하였고 이를 통해 총 15개의 위험 시나리오(M1~M15)를 도출하였다. 제어(M1, M7, M8, M10, M14, M15)와 관련된 위험 시나리오가 가장 많이 도출되었고, 그다음으로 구동(M2, M9, M12), 센서(M3, M11, M13), 전원(M4, M6), 통신(M5) 순이었다. 시나리오별로 심각도(S), 발생 빈도(O), 검 출가능성(D), RPN을 추정하였다.

표 4. FMEA 기반 위험 시나리오 분석 결과.

Table 4. FMEA-based risk scenario analysis results.

ID	S	위험 시나리오 설명	S	O	D	RPN	RL
M1	제어	균형 제어가 실패하거나 무게중심이 흐트러져 로봇이 넘어지는 상황. 관절·외장 파손과 함께 주변과 충돌해 2차 사고가 발생할 수 있다. 미끄러운 바닥이나 협소한 공간에서 전도되면 인명 피해가 클 수 있다.	8	3	7	168	H
M2	구동	관절 액추에이터나 모터 드라이버가 오작동해 예측 불가능한 힘·속도로 움직이는 상황. 사용자 제어가 불가능해져 충돌·전도 위험이 높고, 부품 파손 시 화재나 2차 사고로 번질 수 있다.	8	2	6	96	H
M3	센서	센서 고장으로 장애물을 인식하지 못하는 상황. 실제 환경과 제어 간 불일치가 커지며, 정상 동작 중에도 예기치 못한 충돌이나 전도가 발생할 수 있다. 사람을 놓치면 인명 피해 위험이 상당히 높아진다.	7	4	3	84	M
M4	전원	배터리가 과열되어 화재·폭발로 이어질 수 있는 상황. 내부 셀이 발화하면 밀폐된 환경에서는 대규모 화재로 확산될 가능성이 크다. 냉각·관리 시스템이 부족하면 감지·대응이 어려워 인명·재산 피해가 커질 우려가 있다.	10	2	3	60	L
M5	통신	로봇-제어기 간 통신이 두절·불량해 정상 지령이 전달되지 않는 상황. 로봇이 작업 중 위험 동작을 멈추지 못하거나 원격 등의 긴급 비상정지(E-stop) 명령이 적용되지 않아 충돌·전도로 이어질 수 있다.	6	5	1	30	L
M6	전원	전원 또는 동력 공급이 끊겨 로봇이 갑자기 정지·전도되는 상황. 균형이 해제되면 관절 등에 충격이 가해지고 주변 사물과 충돌 위험이 높아진다.	8	2	7	112	H
M7	제어	치명적 소프트웨어 버그로 인해 제어 명령이 비정상적으로 동작하는 상황. 무한 루프나 오작동으로 과속·과힘이 발생해 충돌이나 전도가 일어날 수 있으며, 감지 실패 시 인명 피해로 이어지기 쉽다.	7	3	9	189	M
M8	제어	제어 시스템 이상으로 로봇이 예정된 속도·가속 한계를 초과해 동작하는 상황. 급격한 돌진이나 과도한 힘이 전도로 이어지거나, 주변인과 충돌해 부상을 일으킬 위험이 커진다.	8	2	4	64	L
M9	구동	관절·기어가 마모되거나 결함이 누적되어 동작 범위를 벗어나는 상황. 정상 움직임이 어려워 전도·충돌 가능성이 커지며, 장시간 점검이 없으면 위험도가 상승한다.	6	3	5	90	M
M10	제어	사용자의 잘못된 조작이나 과도한 동작 지시로 로봇이 위험 동작을 수행하는 상황. 제한 로직 없이 수행하면 관절 손상, 충돌, 전도 등이 일어나고 협동 작업 시 주변 인원을 보호하기 어렵다.	7	6	7	294	H
M11	센서	물·먼지·온도 등 환경 요인으로 센서나 부품이 오작동해 실제 환경과 맞지 않는 데이터를 생성하는 상황. 제어 알고리즘이 이를 그대로 사용하면 충돌이나 전도 가능성이 높아진다.	5	3	5	75	L
M12	구동	과도한 하중이나 무게 분포 불균형으로 로봇이 넘어지는 상황. 전도 시 관절이 크게 손상되고 주변 인명 피해 가능성이 높다. 계단·경사로 등 복합 지형에서 발생하면 사고 규모가 커진다.	8	4	5	160	H
M13	센서	근접한 사람을 충분히 인지하지 못한 채 빠른 속도로 이동·작동해 충돌이 발생하는 상황. 충돌 방지 장치나 긴급 정지가 작동하지 않으면 심각한 신체 손상을 유발할 수 있다.	7	5	3	105	M
M14	제어	로딩-언로딩 작업 중 힘 제어 로직 이상으로 과도한 힘을 가하는 상황. 물체를 다루다 과압을 가해 파손시키거나, 사람이 잡힌 경우 골절 등 심각한 부상을 줄 위험이 높다.	7	3	2	42	L
M15	제어	비상 정지 스위치나 긴급 중단 기능이 제대로 동작하지 않는 상황. 위험 상태가 감지되어도 로봇이 움직임을 멈추지 않아, 대규모 피해로 번질 가능성이 매우 크다.	10	2	10	200	H

표 5. STPA 및 ARI 결합 분석 결과.

Table 5. Combined STPA and ARI analysis results.

ID	위험 시나리오 설명	UCA	Q	A	R	E	ARI	RL
A1	비전 인식이 사람을 제대로 감지 못하거나 오인식해, 로봇이 충돌 회피 없이 동작을 계속하는 상황. 혼잡 환경에서 실시간 인식 실패가 누적되면 인명 피해 발생 가능성이 크다.	Severe	4	3	1	1	9	H
A2	음성 명령 인식 모듈이 발화 내용을 틀리게 해석해 엉뚱한 제어 명령을 내리는 상황. “서서히 전진”을 “최대 속도 전진”으로 잘못 처리하면 로봇이 갑작스러운 가속을 수행해 전도·충돌을 일으킬 수 있다.	Moderate	3	2	1	1	7	M
A3	강화학습 과정에서 잘못된 보상 구조를 학습하여 의도치 않은 위험 동작을 계속 수행하는 상황. 예컨대 한쪽 행동만 반복하거나 사람을 밀치는 등 예측 불가능한 오작동이 생길 수 있다. 안전 제약이 미흡하면 치명적 피해를 야기할 우려가 크다.	Severe	2	4	1	1	8	H
A4	주행 중 경로 재계획 모듈이 지연되어, 장애물이 새로 등장했음에도 로봇이 직전 경로대로 움직이는 상황. 계산이 끝나기 전에는 회피 동작이 불가능해 충돌 위험이 발생한다.	Moderate	1	3	3	1	8	M
A5	AI 내부 결정 과정을 파악하기 어려워 사용자가 로봇의 행동 의도를 알 수 없는 상황. 예측 불가능한 움직임에 대처가 늦어 충돌·파손으로 이어질 수 있으며, 오류 원인 분석도 어려워 추가 피해 방지에 한계가 생긴다.	Limited	1	1	1	4	7	L
A6	학습 데이터 편향으로 특정 조건에서 오류가 집중 발생하는 상황. 조명·환경 등 훈련 범위를 벗어난 경우 로봇이 잘못된 판단을 내려 돌진, 전도, 물체 파손 등을 일으킬 수 있다.	Severe	4	3	1	3	11	H
A7	로봇의 최종 목표 설정이 잘못되어 안전보다 효율·속도를 우선하는 정책이 생성된 상황. 충돌 방지나 제한구역 준수 등을 무시하고 빠른 경로만 추구하면 대형 사고로 이어질 수 있다.	Severe	1	1	1	4	7	H
A8	화재, 충돌 같은 긴급 상황에서 자동 중지 기능이 작동하지 않는 상황. 로봇이 계속 오작동해 피해 범위를 키울 수 있고, 수동으로 멈추기에도 시간이 부족해 인명·설비 손상이 커질 우려가 있다.	Severe	2	4	2	1	9	H
A9	센서 데이터 잡음이나 이상값으로 인해 실제 상황과 다르게 인식하는 상황. 장애물이 없는데 있다고 보거나, 반대로 사람이 있어도 감지 못하는 등 오판이 누적되면 충돌·전도 위험이 커진다.	Moderate	4	3	2	1	10	H
A10	사용자의 의도를 로봇이 오해해 잘못된 행동을 취하는 상황. “우회” 명령을 “직진”으로 인식하면 장애물 파손이나 충돌 사고가 생길 수 있다. 모호한 인터페이스일수록 긴급 상황 대처가 늦어져 피해가 커질 가능성이 있다.	Moderate	3	3	2	1	9	H
A11	불확실한 입력·환경에서 AI가 결정을 강행해 위험 상태로 진입하는 상황. 주어진 정보가 부족한데도 행동을 실행하면 돌발 충돌이나 손상 우려가 크다. 사용자도 파악하기 어려워 사고 위험이 높아진다.	Moderate	1	4	1	2	8	M
A12	네트워크 단절, 극한 온도 등 예외 상태에서 AI 성능이 저하되어 비정상 동작을 수행하는 상황. 대체 알고리즘이 없으면 안전 정지 대신 예측 불가능한 행동을 일으켜 인명 피해를 유발할 수 있다.	Severe	4	4	1	1	10	H
A13	데이터 업데이트가 늦어 구식 정보를 사용해 움직이는 상황. 환경 변화가 반영되지 않아 충돌·전도 위험이 높아지며, 업데이트 지연이 길어지면 안전장치도 적절히 작동하지 않을 수 있다.	Limited	3	1	2	1	7	L
A14	외부 공격으로 로봇 제어권을 빼앗겨 모델 변조나 위험한 동작이 강제되는 상황. 제한 속도·힘을 무시하고 과격한 움직임을 하거나 민감 구역을 침범해 대형 사고로 이어질 수 있다. 보안이 취약한 로봇 환경에서 발생 확률이 높다.	Severe	4	4	1	2	11	H
A15	윤리적 제한이 적용되지 않아 사생활 침해, 부적절한 판단 등이 실행되는 상황. 예컨대 로봇이 주변 인물을 무단 촬영·분석하거나 안전을 희생해 작업 효율만 우선시할 수 있다. 결과적으로 사회적 문제나 법적 분쟁으로 확대될 수 있다.	Limited	1	1	1	4	7	L

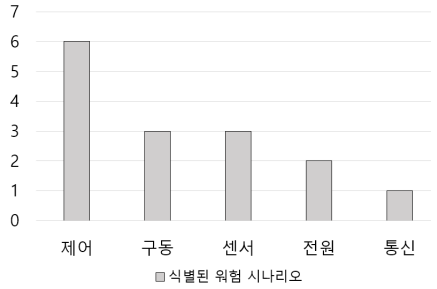


그림 5. FMEA 분석 결과 분류 그래프.

Fig. 5. FMEA results classification graph.

FMEA 분석 결과 표 4와 같다. 일부 시나리오(M10, M15)는 매우 높은 RPN 값을 보였으며, 발생 가능성은 작지만 심각성이 높은 시나리오(M1, M2, M6, M12)는 높은 위험 수준(High)으로 분류되었다. 특히 구동부 제어 불능이나 로봇의 전원장치 이상과 같이 시스템 기본 기능에 치명적인 결함이 발생하는 경우 치명적 사고로 이어질 수 있어 위험도가 높게 평가되었다. 또한, 고장 발생 가능성이 작거나 영향이 비교적 경미한 시나리오는 낮은 위험 수준(Low)으로 평가되었으며, 나머지 시나리오는 중간 수준(Medium)의 위험성으로 결정되었다. 전반적으로 FMEA를 통해 하드웨어 결함, 전기적 오류 등 전통적 요인에 따른 위험들이 체계적으로 식별되었다. 또한 FMEA를 통한 정량적 RPN 평가로 위험요인별 상대적 우선순위를 파악할 수 있었다.

2. STPA+AI 가중치 분석 결과

다음으로 STPA 기법에 AI 위험 요소를 결합한 분석을 수행하여 시스템 수준 사고 시나리오(A1~A15)를 표 5와 같이 도출하였다. STPA를 통해 로봇 제어 시스템에서 발생할 수 있는 UCA를 식별하고, 이를 유발할 수 있는 사고 시나리오를 정성적으로 추론하였다. 이어서 각 시나리오에 대해 로봇 제어에 AI 기술이 미치는 영향을 반영하기 위해 STPA 기법에 따른 Q, A, R, E를 정량·정성적으로 평가하고 ARI를 산출하였다. UCA 식별 값과 ARI을 위험성 결정 기준에 따라 최종적으로 위험 수준을 평가하였다.

STPA+AI 분석 결과, FMEA로는 드러나지 않았던 여러 가지 소프트웨어 및 알고리즘 차원의 위험 시나리오들이 식별되었다. 먼저 UCA는 Servere가 7개, Moderate가 5개, Limited가 3개로 식별되었다. ARI는 9 이상이 7개, 미만이 8개로 산출되었다. UCA 식별 값과 ARI 값을 종합한 결과인 위험성 결정 등급 High는 총 9개 시나리오(A1, A3, A6, A7, A8, A9, A10, A12, A14)가 도출되었고, Medium 2개, Low가 3개로 도출되었다. UCA 식별 결과가 Severe로 위험성 결정이 High로 결정된 경우는 7개였으나, ARI 값에 따라 3개의 시나리오(A4, A9, A10)도 High로 결정되었다.

3. 위험성 결정 결과 종합 분석

FMEA와 STPA+AI 분석을 통해 총 도출된 30개 위험 시나리오에 대해 사전에 정의한 위험성 결정 기준에 따라 최종 위험 등급(High, Medium, Low)을 판정하였다. 표 6에 나타낸 대로, “M” 시나리오 중 High 등급으로 분류된 항목은 6건, Medium 등급은 4건, Low 등급은 5건으로 집계되었으며, “A” 시나리오 또한 High 등급이 9건, Medium 등급이 3건, Low 등급이 3건으로 나타났다.

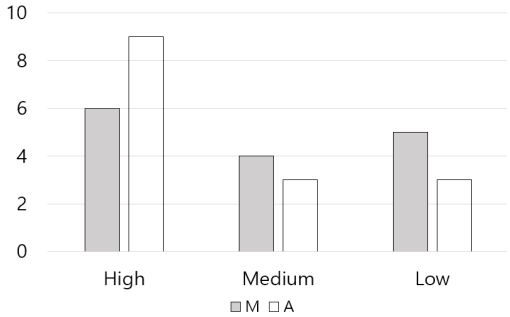


그림 6. 위험성 결정 결과 요약 그래프.

Fig. 6. Summary graph of risk evaluation results.

위험 시나리오의 등급 결과 중 High 등급으로 선정된 총 15개의 시나리오를 살펴보면 몇 가지 특징이 나타났다. 먼저 하드웨어 측면에서는 로봇의 제어기와 구동부의 기본적 기능을 상실하거나 에너지 공급 계통에 심각한 이상이 생기는 경우 즉각적으로 로봇이 통제력을 잃고 주변 인간에게 물리적 위해를 가할 수 있다. 한편 소프트웨어/알고리즘 측면에서는 환경 인지 실패나 결정 논리 오류로 인해 안전한 제어 행동이 이루어지지 않는 상황이 높은 위험으로 선정되었다. “A” 시나리오에서 High 등급이 많이 도출되었다. AI 기술이 로봇에 확장되고 있는 단계를 고려할 때 안전 문제를 충분히 다루기엔 제한적인 것으로 분석할 수 있다.

이러한 고위험 시나리오는 공통적으로 로봇의 비정상 동작을 억제할 안전 메커니즘이 부재하거나 제 기능을 못 한 상황에서 발생한 것으로 나타났다. 즉, 로봇이 위험 상황에 직면했을 때 이를 감지하여 즉시 동작을 중단하거나 안전 모드로 전환할 장치나 알고리즘이 없었거나 제대로 작동하지 않았음을 의미한다. 본 평가 결과 도출된 High 등급 항목들은 로봇 시스템에서 가장 시급히 개선이 필요한 안전 취약점으로 식별되었으며, 다음 장에서는 이들에 대응하는 안전 요구사항을 도출하고자 한다.

V. 프레임워크 검증 및 안전 요구사항 도출

1. 프레임워크 적용 검증

제안된 FMEA + STPA 통합 위험성 평가 프레임워크를 사례 분석에 적용함으로써 그 유용성과 실효성을 검증하였다. 각 기법이 발견한 위험 요소들이 상호 보완적으로 작용하여, 단일 방법론을 사용했을 때 발생할 수 있는 위험 요인 누락이 감소함을 확인하였다. 특히 FMEA를 통해서는 주로 하드웨어 부품의 고장으로 인한 위험이 체계적으로 식별되었고, STPA를 통해서는 시스템 제어 논리나 상호작용 측면에서 발생할 수 있는 사고 시나리오가 추가로 발굴되었다. FMEA 단독으로는 놓치기 쉬운 AI 알고리즘 오류에 따른 위험을 STPA가 식별했고, 반대로 STPA만으로는 고려되지 않는 개별 부품 결함에 따른 위험은 FMEA를 통해 보완되었다. 이처럼 두 기법의 결합은 다층적 관점의 위험 식별을 가능하게 하여, 로봇 시스템의 복잡성을 반영한 포괄적인 위험요소 발굴에 기여 가능한 것으로 보인다.

또한 제안된 프레임워크는 평가 결과의 정량화 및 우선순위화 측면에서도 효과적이었다. 모든 도출 위험 시나리오에 대해 위험성을 구분함으로써, 향후 안전 대책 수립의 우선순위를

명확히 설정할 수 있었다. 예를 들어, High 등급으로 평가된 위험에 대해서는 설계 변경이나 보호장치 추가와 같은 즉각적 개선조치를 최우선으로 고려할 수 있다. Medium 등급 위험에 대해서는 모니터링 강화나 단계적 개선을 계획하며, Low 등급은 수용 가능한 위험으로 간주하여 지속적인 감시 정도로 대응하는 등 효율적인 자원 할당이 가능하다. 실제 평가 과정에서 도출된 High 위험 시나리오 15건은 FMEA 또는 STPA 단일 기법을 사용했다면 일부 누락되었을 가능성이 있는 항목들까지 포함하고 있었으며, 본 프레임워크 적용을 통해 위험 요소를 빠짐없이 확인하고 심각도에 따라 구분해낸 것을 확인하였다. 이 결과는 제안된 통합 프레임워크가 휴머노이드 로봇의 안전성 평가에 있어 의사결정 지원도구로 활용될 수 있음을 시사한다.

2. 안전 요구사항 도출

위험성 평가를 통해 고위험(High)으로 식별된 15개 시나리오(M1, M2, M6, M10, M12, M15, A1, A3, A4, A6, A7, A8, A9, A10, A12, A14)에 대해서는 각각의 발생 원인과 잠재 결과를 면밀히 분석하였고, 이를 토대로 로봇 시스템의 안전을 확보하기 위한 안전 요구사항을 도출하였다. 도출된 요구사항은 유사한 성격의 위험을 묶어 제어, 구동, 전원, AI 인지, AI 의사결정의 5개 유형별로 재분류하고 정리하였다. 아래에는 주요 위험 유형별 대표적인 안전 요구사항을 요약한다. 제어, 구동, 전원 유형은 “M” 시나리오와 관련하여 도출된 요구사항이고, AI인지, AI 의사결정은 “A” 시나리오를 바탕으로 도출된 요구사항이다.

2.1 제어 관련 안전 요구사항

먼저, 제어 관련 위험 시나리오인 M1, M10, M15에서 확인 가능한 바와 같이 제어 계통에서 오류로 인한 위험이 우려된다. 이에 따라 안전 요구사항은 자세 균형을 위한 능동 제어를 다각도로 검증하고, 제어 안정성 강화를 위해 실시간 위치 및 자세 모니터링 및 보정 가능 여부를 확인해야 한다. 또한, 사고 경감을 고려하기 위해 속도와 토크 한계 설정 가능 여부를 확인해야 한다. 위험의 예방 차원에서 제어 접근 범위, 이중화 설계, 상태 자가진단 기능 등의 안전기능이 반영되었는지와 정상 동작하는지 확인해야 한다.

2.2 구동부 관련 안전 요구사항

구동부 관련 안전 요구사항은 M2, M12 시나리오 관련이다. 기존의 모터 관련 안전성은 모터 드라이버의 전류·토크 센서를 통해 과부하, 온도 상승, 비정상 하중 분포를 실시간으로 진단하는 안전기능 등은 설계되어 있다. 휴머노이드 로봇 관점에서는 제어기가 통제하지 못하는 상황을 대비하여 모터 드라이버 레벨에서 실시간 진단 및 회로 이중화 등을 검토해야 한다. 또한, 휴머노이드 로봇은 다축 구조로 특정 드라이버 비정상적 상황에서 특정 구동부만 안전기능이 작동해서는 안되며, 최적의 상황에 맞도록 필요한 모든 구동부가 안전기능이 작동되어야 한다. 이때 기계식 및 전자식 브레이크를 적용하고, 주요 관절에는 클러치 또는 제동 장치를 설치하여 전원 차단 시에도 관절 낙하 위험을 방지할 필요가 있다.

2.3 전원부 관련 안전 요구사항

전원부 관련은 M6 시나리오이다. M6에서 예측되는 바와 같이 전원공급의 불안정성을 예방할 필요가 있다. 제한적 이중화 구조를 채택하여 주 전원 고장 시에도 제어 로직이 감속 후 안전 정지하도록 보장해야 한다. 전원 상실이나 배터리 상태 이상 감지 시 관절별 브레이크·댐핑 장치를 자동으로 활성화하여 낙하 충격을 최소화해야 한다. 소프트웨어에서는 전원 중단 신호를 받으면 즉시 완속 감속 프로토콜을 수행하고, 배터리 잔량·온도·사이클 데이터를 실시간 모니터링하여 사전 경고를 발신해야 한다. 정기적인 전원 회로 유지보수 절차를 점검해야 한다.

2.4 AI 인지 관련 안전 요구사항

AI 인지 관련 위험 시나리오는 A1, A9, A10이다. 인지 관련 위험은 다양한 센서를 통해 인지 과정을 거치기 때문에 충분한 센서의 수와 수집 가능한 데이터, 데이터를 처리할 AI가 핵심이다. 멀티모달로 수집된 데이터의 활용 우선순위, 데이터 처리 이전에 센서 잡음 및 이상치를 제거 수준을 검증해야 한다. 음성·제스처·텍스트 명령은 재검증 로직을 통해 인지하는지도 살펴볼 필요가 있다. 마지막으로 인지를 위한 데이터 판별을 안전을 우선한 알고리즘인지 검증이 필요하다.

2.5 AI 의사결정

AI의 핵심인 정책적 의사결정은 6개의 위험 시나리오와 관련된 만큼 검증이 중요하다. 학습된 데이터의 노후화 관리 방법과 데이터 활용 방식에 대한 로직을 검증해야 한다. 또한, 학습 데이터의 다양성·무결성을 검증하여 편향 학습을 방지하고, 강화학습 보상 구조는 외부 전문가 검토를 거쳐 설계해야 한다. AI 제어 실패 시 룰 기반 제어나 수동 비상정지로 자동 전환되는 백업 로직을 검증해야 한다. 또한, 윤리적 의사결정에 대한 알고리즘이 반영되었는지 검증해야 한다. 접근성 제한과 연계된 사이버 보안 분야에서는 제어 신호 및 네트워크 통신에 대한 인증 및 암호화를 확인해야 한다.

표 6. 위험성 평가 기반 안전요구사항 도출 결과.
Table 6. Safety requirements derived from risk assessment.

유형	관련 시나리오	안전 요구사항
제어	M1, M10, M15	· 자세 균형을 위한 능동 제어 · 실시간 위치 및 자세 모니터링/보정 · 속도·토크 한계를 설정 · 제어 접근 범위 제한 및 접근성 이중화 · 제어부 이중화 설계 · 상태 자가진단 및 복구
구동	M2, M12	· 모터/드라이버 실시간 진단 · 모터 드라이버 회로 이중화 · 모터 드라이버 간 진단 공유 · 주요 관절 브레이크(기계식, 전자식)
전원	M6	· 제한적 전원 이중화 시스템 · 전원 상실 보호용 제동/댐핑 · 완속 감속 프로토콜
AI 인지	A1, A9, A10	· 다중 센서 데이터 우선순위 · 실시간 센서 노이즈 검출 및 차단 · 음성·텍스트·제스처 명령 재검증 · 안전 우선 인지 모드
AI 의사결정	A3, A6, A7, A8, A12, A14	· 학습 데이터 노후화 검증 · 학습 데이터 편향화 검증 · 학습의 보상 구조 검증 · 윤리적 의사결정 · 백업 제어 알고리즘 · 사이버 보안

VI. 결론

본 논문에서는 인공지능이 적용된 휴머노이드 로봇의 안전성 강화를 위해 FMEA와 STPA를 통합한 새로운 위험성 평가 프레임워크를 제안하고 그 효용성을 검증하였다. 제안된 프레임워크는 전통적 하드웨어 관점의 고장 모드 분석(FMEA)과 시스템적 관점의 사고 시나리오 분석(STPA)을 결합하고, 추가적으로 AI 특유의 위험요인을 고려하는 절차로 이루어진다. 이를 통해 로봇 시스템의 구성품 수준부터 상호작용 및 알고리즘 수준까지 전방위적인 위험 식별이 가능하며, 최종적으로 위험도를 계량화하여 우선순위에 따른 안전 대책 수립까지 연계될 수 있도록 하였다.

사례 연구로 진행된 평가 결과, 본 프레임워크의 실효성이 입증되었다. 산업용 휴머노이드 로봇 사례에 적용한 결과 FMEA 단일 분석 또는 STPA 단일 분석에 비해 더 폭넓은 위험 시나리오를 식별할 수 있었으며, 각 시나리오의 위험 수준을 종합적으로 평가하여 안전 취약점을 체계적으로 파악할 수 있었다. 또한 도출된 고위험 시나리오를 바탕으로 구체적인 안전 요구사항을 제시함으로써, 이 연구 결과를 로봇 시스템의 실제 안전 설계 개선에 활용할 수 있는 방안을 마련하였다.

본 연구의 기여도는 다음과 같다. 첫째, 기존 위험성 평가 기법의 한계를 보완하기 위해 정량적 FMEA 기법과 정성적 STPA 기법을 하나의 프레임워크로 통합하여 제시함으로써 학술적 및 실무적 의의를 가진다. 이를 통해 하드웨어 결함부터 제어 소프트웨어 오류까지 다양한 위험요인을 하나의 절차 내에서 다룰 수 있게 되었고, 위험 평가 과정의 누락 가능성을 감소시켰다. 둘째, 인공지능 기술이 초래하는 새로운 유형의 위험에 대해 평가 지표를 도입하고 영향도를 분석함으로써, 향후 AI 로봇 안전성 평가를 위한 기초를 마련하였다. 셋째, 위험 평가 결과로부터 구체적인 안전 요구사항을 도출하는 체계를 제시하여, 분석 결과를 안전 설계 및 운영 지침에 직접적으로 반영할 수 있도록 한 점도 중요한 기여이다. 본 연구는 인간과 공존하는 휴머노이드 로봇의 안전성 확보를 위한 이론적 기반과 실용적 실행방안을 병행하여 제시한 점에서 학문적 및 산업적 의의가 크다.

한편, 본 연구에는 몇 가지 한계가 존재한다. 우선, 적용 사례가 특정 유형의 휴머노이드 로봇에 국한되어 있으므로, 본 프레임워크의 효과가 모든 로봇 유형에 일반화된다고 보기에는 무리가 있다. 다양한 형태와 용도의 로봇에 본 기법을 적용할 경우 결과가 달라질 수 있으므로, 추가 사례 연구를 통해 프레임워크의 범용성을 검증할 필요가 있다. 다음으로, STPA에 포함된 AI 위험요인 가중치의 부여 등 일부 분석 과정에서 전문가의 주관적 판단이 개입되었으며, 해당 요소의 정량적 근거를 마련하는 데 한계가 있었다. 인공지능 시스템의 복잡한 오류 특성을 정량적으로 평가하는 방법론은 아직 발전 단계에 있으므로, 위험도를 보다 객관적이고 정확하게 산출할 수 있는 보완 연구가 요구된다. 마지막으로, 본 연구에서 도출된 안전 요구사항들은 이론적으로 제안된 것이며, 실제 로봇 시스템에 구현하여 안전성 향상 효과를 실증하지는 못하였다. 제안된 대책들의 실효성과 현실적 적용 가능성은 추후 프로토타입 구현이나 실험적 검증을 통해 평가되어야 할 것이다.

이러한 한계를 보완하기 위해 향후 연구 방향을 몇 가지 제시한다. 첫째, 제안된 위험성 평가 프레임워크를 다양한 로봇 플랫폼과 시나리오에 적용하여 결과를 비교함으로써, 본 방법론의 신뢰도와 일반화 가능성을 검토할 예정이다. 둘째, FMEA로 도출된 제어 관점의 안전 요구사항은 시험평가 방법을 별도로 표준화 연구 주제로 제안할 예정이다. 특히 자세 균형을 위한 능동 제어 수준과 실시간 위치 및 자세 모니터링 수준은 다른 위험성과 연계될 가능성이 높으며, 평가 방식에 따라 검증 결과 차이가 클 수 있기 때문이다. 셋째, STPA 관점에서 도출된 AI 관련 위험요인의 정량화 기법을 고도화하는 연구를 진행하고자 한다. 예를 들어, 머신러닝 모델의 불확실성을 계측하고 이를 위험도로 환산하는 방법, 시뮬레이션을 통해 AI 제어 시스템의 잠재 오류를 확률적으로 예측하는 방법 등을 개발하여 위험 평가의 과학적 근거를 강화할 계획이다.

REFERENCES

[1] M.-J. Kim, D. Lim, D. Kim, J. Cha, J. Shin, W. Cha, G. Park, K. Lee, and J. Park, “Advances in humanoid robot walking technologies: a review,” J*ournal of Institute of Control, Robotics and Systems (in Korean)*, vol. 30, no. 4, pp. 412-422, Apr. 2024.
doi: https://doi.org/10.5302/J.ICROS.2024.24.0042

[2] ‘Robots running towards each other’… Robots causing mayhem at Chinese research lab, Etnews, May 5, 2025.
URL: https://www.etnews.com/20250504000050

[3] Robot rushes toward audience… Chinese humanoid robot malfunctions, ‘dizzying’, The Chosun Daily, Feb. 27, 2025.
URL: https://www.chosun.com/international/international_general/2025/02/27/LL3K4GJTSJCQJP4MCQKKQGRXIU/

[4] International Organization for Standardization, Robots and robotic devices - Safety requirements for industrial robots - Part 1: Robots (ISO 10218-1:2025), ISO, Switzerland, 2025.

[5] International Organization for Standardization, Robots and robotic devices - Safety requirements for industrial robots - Part 2: Robot systems and integration (ISO 10218-2:2025), ISO, Switzerland, 2025.

[6] International Organization for Standardization, Robots and robotic devices - Safety requirements for personal care robots (ISO 13482:2014), ISO, Switzerland, 2014.

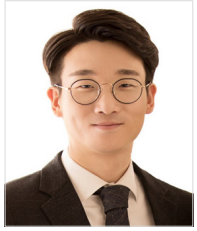
[7] International Organization for Standardization, Information technology - Artificial intelligence - Risk management (ISO/IEC 23894:2023), ISO, Switzerland, 2023.

[8] International Organization for Standardization, Artificial Intelligence - Concepts and terminology (ISO/IEC JTC 1/SC42), ISO, Switzerland, 2022.

[9] Robotics Industries Association, Industrial Robots and Robot Systems - Safety Requirements (ANSI/RIA R15.06-2012), RIA, U.S., 2012.

[10] European Commission, Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act, COM/2021/206 final), European Commission, Brussels, 2021.

[11] International Electrotechnical Commission, Risk Management - Risk Assessment Techniques (IEC 31010:2019), IEC, Switzerland, 2019.
[12] D. H. Stamatis, *Failure Mode and Effect Analysis: FMEA from Theory to Execution*, ASQ Quality Press, U.S., 2003.
[13] SAE International, Potential Failure Mode and Effects Analysis (FMEA) Including Design FMEA, Supplemental FMEA-MSR, and Process FMEA (SAE J1739:2021), SAE, U.S., 2021.
[14] N. G. Leveson, *Engineering a Safer World: Systems Thinking Applied to Safety*, MIT Press, U.S., 2011.
[15] N. G. Leveson and J. Thomas, *STPA Handbook, MIT Partnership for a Systems Approach to Safety*, U.S., 2018.
[16] J. Rasmussen, "Risk management in a dynamic society: A modelling problem," *Safety Science*, vol. 27, no. 2-3, pp. 183-213, Nov. 1997. doi: https://doi.org/10.1016/S0925-7535(97)00052-0
[17] E. Hollnagel, D. D. Woods, and N. Leveson, *Resilience Engineering: Concepts and Precepts*, Ashgate Publishing Ltd., UK, 2006.



류 요 엘

2013년 한양대 전자및통신공학과 졸업. 2022년 숭실대 안전보건융합공학 박사. 2013년~현재 한국로봇산업진흥원 휴머노이드 로봇센터 책임연구원. 관심분야는 휴머노이드 로봇, 로봇안전, 로봇표준.

전 진 우

1996년 성균관대 생물기전공학과 졸업. 2000년 상명대 기술경영학 석사. 2020년 숭실대 안전보건융합공학 박사. 2010년~현재 한국로봇산업진흥원 수석연구원. 관심분야는 로봇안전표준, 로봇 위험성 평가, 기술정책.