안전한 AI 도입 및 활용을 위한

해양산업 인공지능(AI) 가이드라인

표지 이미지는 GenAI 기술을 활용하여 생성된 창작물로 상업적 사용 시 별도의 법적 검토가 필요합니다

kobc 한국해양진흥공사
KOREA OCEAN BUSINESS CORPORATION

| CONTENTS |

Chapter 01.
개요

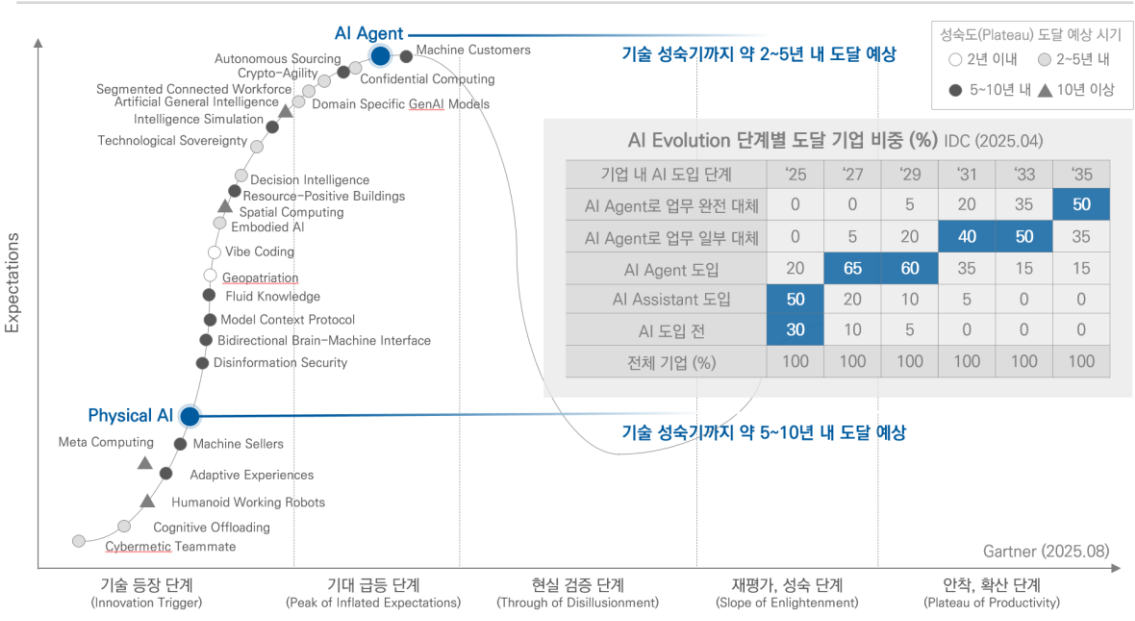
kobc 한국해양진흥공사
KOREA OCEAN BUSINESS CORPORATION

Chapter 01. 개요

1. 가이드라인 배경 및 목적

AI(Artificial Intelligence) 기술은 기술적으로 빠른 진화와 더불어 전 세계 다양한 산업 분야에 급속도로 확산되고 있으며, 이에 따라 기업들의 AI 도입/활용 시도 또한 큰 폭으로 증가하고 있습니다. 그동안 AI 기술은 주로 단기적인 생산성 향상을 목표로 활용되어 왔지만, 이제는 기업의 비즈니스 혁신과 지속 가능한 경쟁력 강화를 위한 핵심 요소로 자리잡고 있습니다. 이에 따라 다양한 산업 분야 전반에서의 기술 도입 수준 또한 PoC(Proof of Concept) 단계나 단발적인 업무 생산성 향상을 위한 AI 활용 단계를 넘어, AI Agent를 실제 업무 내에 적용함으로써 업무 자동화를 구현해내는 단계에 이른 것으로 평가되고 있습니다.

[그림 1] 산업 전반의 AI 기술 성숙도



기업 내 AI 도입 단계	'25	'27	'29	'31	'33	'35
AI Agent로 업무 완전 대체	0	0	5	20	35	**50**
AI Agent로 업무 일부 대체	0	5	20	**40**	**50**	35
AI Agent 도입	20	**65**	**60**	35	15	15
AI Assistant 도입	**50**	20	10	5	0	0
AI 도입 전	**30**	10	5	0	0	0
전체 기업 (%)	100	100	100	100	100	100

Source: Gartner, Hype Cycle for AI and Cybersecurity (2025.08)

Chapter 01. 개요

해양산업 내 AI 기술 도입은 AI 기술 선도 적용 산업군(예: 금융, 통신, IT 등) 대비 아직까지 초기 단계에 있으나, 글로벌 선도 해양 기업들은 다양한 업무 영역에 AI 도입을 적극적으로 추진하고 있습니다. 대표적인 도입 사례인 AI 기반 자율운항 시스템은 항로 및 운항 속도 최적화를 통해 운영 비용 절감과 사고 위험 감소를 실현하며, AI 기반 물류 관리 시스템은 선박 도착 시간 및 항만 혼잡 구간 예측을 통한 컨테이너 적재 및 하역 최적화를 통해 운영 효율성을 높이고 있습니다.

[그림 2] 해양산업 AI 도입 · 활용 대표 분야

해운 (Shipping Lines)	항만 (Port Operations)	해운 물류 (Logistics&Supply Chain)
• 선박 예약 자동화 • 탄소 배출량 예측 • 자율운항선박 • 선박 예지보전	• ETA 및 항만 물동량 예측 • 항만 혼잡 구간 예측 • 컨테이너 손상 자동 탐지 • 컨테이너 적재 최적화	• 물류 · 운송 스케줄링 최적화 • 화물 인도 · 통관 절차 간소화 • 물류 수요량 예측 • 컨테이너 손상 자동 탐지

이러한 도입 사례들과 같이 해양산업 내 AI 기술은 비즈니스 경쟁력 강화와 운영 효율성 개선을 위한 중요한 도구로 활용되고 있으며, 향후 새로운 비즈니스 모델을 창출하고 산업을 혁신할 기회를 제공할 것으로 예상됩니다.

본 가이드라인은 해양 기업이 이러한 AI 기술을 효과적으로 도입하고 활용할 수 있도록 지원하기 위해 제작되었으며, 이를 통해 기업들이 AI 도입 시 직면할 수 있는 리스크와 법적 · 윤리적 요구사항을 충분히 이해하고 AI 활용 가치를 창출할 수 있도록 안내하고자 합니다.

Chapter 01. 개요

AI 기술의 도입 및 활용에 대한 해양 기업들의 관심과 의지가 높아짐에 따라, AI 리스크 관리와 사회적 신뢰 확보가 중요 이슈로 부각되고 있습니다. 특히 AI 시스템의 의사결정 과정에서 발생할 수 있는 오류나 불공정성, 개인정보 보호 문제 등은 사회적 논란을 초래할 수 있는 만큼, AI 기술을 도입 및 활용하기에 앞서 다양한 이해관계자와의 신뢰 기반을 마련하는 것이 중요합니다. 이러한 문제의식을 바탕으로 미국, 유럽, 일본 등 세계 주요국들은 자국의 산업 발전과 AI로 인한 잠재적 위험을 관리하기 위한 법 · 제도를 마련했습니다. 우리나라도 AI 기술에 대한 규범의 필요성을 인식하여 약 4년여 간의 논의 끝에 2025년 AI 기본법을 제정했습니다. AI 기본법은 글로벌 규범 동향과 국내 AI 산업 현황을 반영하여 AI 산업의 발전과 안전하고 신뢰할 수 있는 AI 도입 · 활용 기반 구축이라는 두 가치를 균형적으로 반영하는데 목적이 있으며, 법적 구속력 있는 AI 규제로서 2026년 1월 시행될 예정입니다.

[표 1] 세계 주요국 AI 규제 · 규범 현황

구분		주요 내용
미국	**AI 위험관리 프레임워크** * 제정일자: 2023. 01	• 법적 구속력은 없으나, 사업자가 자율적으로 AI 위험 식별 및 관리가 가능하도록 AI 위험 관리 표준 가이드라인 제시
유럽	**EU AI Act** * 제정일자: 2024. 08 * 시행일자: 논의 중	• 전세계 최초 법적 구속력 있는 AI 위험 규제 제정 • AI 시스템의 위험 수준에 따라 규제 수준 차등화
한국	**AI 기본법** * 제정일자: 2025. 01 * 시행일자: 2026. 01	• AI 기술의 윤리적 활용과 안전한 확산을 위한 국가 차원의 기본원칙과 책무 정립 • AI 산업 발전 및 신뢰성 확보를 위한 법적 기반 마련
일본	**AI 촉진법** * 제정일자: 2025. 06 * 시행일자: 논의 중	• 별도의 제재 규정이 없는 연성 규제. 기업의 자율 규제 장려 • AI 활용 촉진과 AI 위험 대응 간 균형 추구

Chapter 01. 개요

AI 기본법에서는 사람의 생명과 신체 안전에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 AI 시스템을 '고영향 AI'로 지칭하고 있으며, 고영향 AI에 해당하는 서비스나 시스템에 대해 법에서 정한 의무사항을 이행하지 않을 경우 과태료가 부과됩니다. AI 기본법에서는 <교통:선박>에 적용되는 AI 중 사람에 대한 중대한 영향과 위험을 초래할 우려가 있는 AI를 고영향 AI로 명시하고 있으며, AI를 적용하고자 하는 해양 기업은 AI 기본법에 대한 검토와 상세 의무 조항에 대한 사전 대응 방안 수립이 반드시 필요합니다.

이에 따라 본 가이드라인은 해양 기업이 AI 서비스 개발 및 운영 과정에서 발생할 수 있는 규제 불확실성을 해소하고, AI 서비스의 신뢰성을 높이기 위한 필요 최소한의 기준을 제시하는 데 주요 목적이 있습니다. 특히 AI 전문 인력과 연구 개발 투자 여력이 제한적인 기업들이 AI 기본법의 의무 사항을 명확히 이해하고, 이를 체계적으로 이행할 수 있도록 2025년 9월 발표된 AI 기본법 시행령안과 가이드라인 초안을 참고하여 해양 기업이 자율적으로 점검할 수 있는 검토 항목과 요건을 도출했습니다. 또한 『2024 신뢰할 수 있는 인공지능 개발 안내서』, 『금융 분야 AI 개발 및 활용 가이드라인』, 『생성형 AI 저작권 안내서 』등 국내 주요 기관과 기구에서 발표한 권고안 및 가이드를 참고하여, 해양 기업이 AI를 안정적으로 도입하고 운영할 수 있도록 실무 환경에 적용 가능한 지침과 방안을 제시했습니다.

본 가이드라인을 통해 개발자, 기획자 등 AI 실무자들에게 신뢰할 수 있는 AI 서비스 개발을 위한 핵심 요구사항과 실행 가능한 방안을 제시함으로써, 국내 해양 기업들이 성숙한 AI 기술을 기반으로 글로벌 비즈니스 경쟁력을 강화하는 데 기여할 수 있기를 바랍니다.

Chapter 01. 개요

2. 가이드라인 구성 및 활용 방법

본 가이드라인은 AI 제품과 서비스 개발에 참여하는 해양산업 내 다양한 조직과 AI 실무진을 대상으로 설계되었습니다. AI 시스템을 개발하여 제공하는 'AI 개발사업자'와 이를 활용하여 AI 제품 · 서비스를 제공하는 'AI 이용사업자'가 주요 활용 대상이며, 신뢰성과 안전성이 보장된 AI 시스템을 개발 및 제공하기 위한 핵심 요구사항과 검증 항목을 제시합니다.

본 가이드라인은 제1 장 개요, 제2 장 기본원칙 및 고려사항, 제3 장 서비스 도입 체크리스트로 총 3개의 장으로 구성되어 있습니다. 제1 장은 AI 가이드라인을 마련하게 된 배경과 해당 지침이 달성하고자 하는 목적을 제시하고, 이를 효과적으로 적용하기 위해 기업과 실무진에게 권장하는 활용 방법을 소개합니다. 제2 장은 AI 기획부터 개발, 활용, 운영까지 AI 생명주기 전반에 걸쳐 범용적으로 기술적 고려가 필요한 기본원칙을 1) 기획 단계, 2) 개발 단계, 3) 운영 단계 4) 활용 단계로 구분하여 서술했습니다. 제3 장은 제2 장을 토대로 해양 기업들이 AI 서비스를 실제로 개발 및 운영하는데 필요한 구체적인 체크리스트를 제공합니다. 또한 AI 기본법에 따라 고영향 AI에 해당하는 시스템에 대한 의무사항과 이행 여부를 점검할 수 있도록 고영향 AI 요건을 별도 구분하여 제시합니다. 이와 함께 '추가 고려사항 체크리스트'에서는 GenAI, Agentic AI, Physical AI 등 특정 AI 모델의 특성을 반영하여 모델별로 중점적으로 확인해야 하는 항목을 도출했습니다.

Chapter 01. 개요

본 가이드라인은 다음과 같은 절차로 활용될 수 있습니다. 제1 장 개요를 통해 본 가이드라인의 활용 목적과 대상, 구성 방식 등을 이해한 후, 제2 장 기본원칙 및 고려사항을 검토하여 해양 기업이 안정적인 AI 개발과 활용을 위해 준수해야 하는 원칙과 관련 기술 요건, 관리 기준 등을 충분히 숙지하길 바랍니다. 이어서 제3 장 서비스 도입 체크리스트를 통해 해양 기업이 자율적으로 필수 요건 충족 여부를 점검하는 방식으로 본 가이드라인을 활용하는 것을 권장합니다. 다만, 기업 규모와 AI 서비스 특성 등에 따라 각 기업에 적합한 가이드라인 활용 절차와 방법, 적용 범위 등이 상이할 수 있습니다. 따라서 각 기업은 본 가이드라인 활용에 앞서, 자사 비즈니스 및 IT 환경에 적합한 요구사항과 검증 항목을 선별하여 우선 적용해야 하며, 필요 시 해양산업 분야의 전문가와 적극 협업할 필요가 있습니다.

더불어, 본 가이드라인은 AI 기본법과 시행령안을 토대로 해양 기업이 이행할 수 있는 최소한의 안전성 확보 기준을 제시하고 있습니다. 다만, 발생 가능한 윤리적 · 법적 이슈를 사전 방지하고 신속히 대응하기 위해선 유관 법 · 규제에 대한 검토가 병행되어야 하며, 필요 시 법률 자문을 받는 것을 권고합니다. 이때 검토해야 할 유관 법 · 규제로는 AI 기본법 외에도 개인정보보호법, AI 학습용 데이터 및 오픈소스 관련 저작권법 등이 포함됩니다.

본 가이드라인이 제시하고 있는 의무 이행 기준과 요건 등은 2025년 9월 발표된 AI 기본법 시행령안과 가이드라인안을 기준으로 하며, 시행령안에 대한 대국민 추가 의견 수렴 등 입법 절차에 의해 일부 내용이 변경될 수 있습니다. 추후 AI 기본법 등 AI 관련 법령의 제 · 개정, AI 기술 발전 추이 등을 반영하여 주기적으로 수정 · 보완되어야 합니다.

Chapter 02.
기본원칙 및 고려사항

kobc 한국해양진흥공사
KOREA OCEAN BUSINESS CORPORATION

Chapter 02. 기본원칙 및 고려사항

1. 인공지능(AI) 란

인공지능(AI, Artificial Intelligence)은 인간의 지능을 컴퓨터 프로그램으로 구현하는 기술입니다. 인간처럼 생각하고, 문제를 해결하고, 새로운 것을 배우는 능력을 컴퓨터에 부여하는 것을 목표로 합니다. 이미지 인식과 창의적 콘텐츠 생성부터 데이터 기반 예측에 이르기까지, AI는 기업이 합리적인 의사결정을 내릴 수 있도록 지원합니다.

AI는 학습 유형과 활용 목적 · 방식 등에 따라 데이터 패턴을 학습해 예측하는 머신러닝(Machine Learning), 복잡한 비정형 데이터를 처리하는 딥러닝(Deep Learning), 새로운 콘텐츠를 생성하는 GenAI, 목표 기반으로 작업을 자율 수행하는 Agentic AI, 센서·장비와 결합해 물리적 행위를 수행하는 Physical AI 등으로 분류됩니다.

머신러닝은 컴퓨터가 데이터를 보고 스스로 배워서 예측이나 판단을 제공하는 AI 기술입니다. 컴퓨터에 명시적으로 프로그래밍하지 않고도 데이터를 통해 학습하여 스스로 패턴을 찾아내고 예측을 수행합니다. 대표적인 머신러닝 사례로 금융 기업은 개인의 신용점수를 예측하거나, 거래 패턴을 분석하여 비정상적인 거래를 식별하는데 AI를 사용합니다.

딥러닝은 인간의 두뇌와 유사한 방식으로 데이터를 처리하도록 고안된 AI 방식인 인공신경망을 기반으로 한 기술로 이미지 인식, 자연어 처리, 음성 인식, 영상 및 이미지 분석에 적합합니다. 대표적인 딥러닝 사례로 2016년 구글의 AI 프로그램인 '알파고'가 딥러닝과 강화학습을 활용해 세계적인 바둑 기사 이세돌 9단을 이기며, 컴퓨터가 복잡한 전략과 예측을 수행할 수 있음을 증명했습니다.

Chapter 02. 기본원칙 및 고려사항

생성형 AI(GenAI, Generative AI)은 텍스트, 오디오, 이미지 또는 동영상 형태의 새로운 콘텐츠를 생성할 수 있는 AI 기술입니다. 주어진 입력 데이터를 바탕으로 특정 출력을 분류하거나 예측하는데 중점을 둔 기존 AI와 달리 이용자의 지시에 따라 이전에는 없던 결과물을 만들어낸다는 점에서 차별점이 있습니다.

AI가 인간을 돕는 조력자라면, Agentic AI는 인간과 동등하게 일하는 동료의 역할을 지향합니다. Agentic AI는 이용자가 목표를 설정하면 AI가 이를 이해하여 목표 달성을 위한 워크플로우를 자율적으로 수행하는 AI 시스템입니다. Gartner는 이를 '인간의 개입 없이 스스로 목표를 달성하기 위해 작업계획, 도구 사용, 가드레일 준수 등의 기능과 결합하여 실행하는 자율 AI시스템' 으로 정의하고 있습니다. 이처럼 Agentic AI의 가장 큰 특징은 '자율성'을 확보하여 'Action'을 수행한다는 점입니다.

[표 2] AI 와 Agentic AI 특징

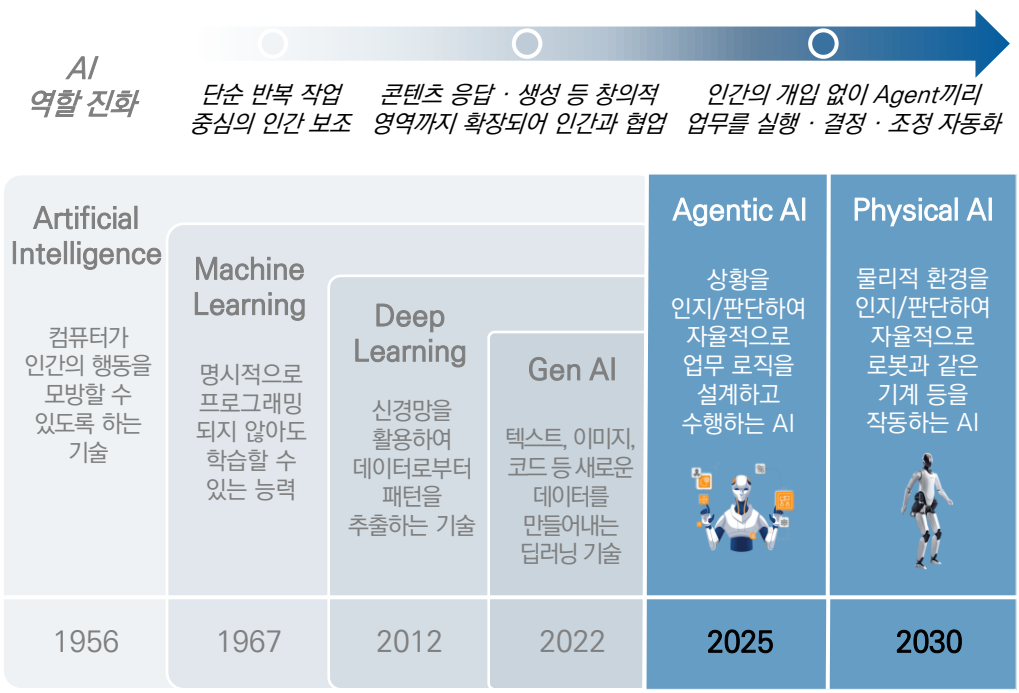
구분	AI	Agentic AI
목적	• 데이터 기반 예측·분류 등 단일 기능 중심의 문제 해결 수행	• 작업 목표를 자율적으로 달성하는 AI 기반 지능형 작업 수행 체계 구현
특징	• 명확히 정의된 입력–출력 구조 내에서 반복·단일 기능 중심의 문제 해결 지원	• 외부 지시 없이 스스로 목표 정의 후 달성하기 위한 전략 및 실행 계획 설계
	• 개별 모델 단위로 작동하며, 복합 업무 처리 시 별도 시스템 설계 필요	• 필요한 도구, 데이터 등을 스스로 선택·조합하여 독립적인 작업 수행
	• 사전 정의된 규칙·학습 모델 기반으로 주어진 작업만 수행	• 작업 결과를 분석하여, 전략 개선 및 성능 고도화를 위한 학습 루프 내재화

Source: 삼성SDS, Agentic AI란 무엇인가? - 뛰는 AI 에이전트, 나는 Agentic AI의 시대, 2025.06

Chapter 02. 기본원칙 및 고려사항

Physical AI는 AI 기술을 물리적 시스템에 통합하여, 실제 세계와 직접적으로 상호작용할 수 있게 만든 지능형 시스템을 의미합니다. AI 기술은 대규모 언어 모델(LLM)과 GenAI의 등장으로 텍스트, 이미지, 음성 처리 분야에서 혁신적인 성과를 보였지만, 주로 디지털 영역에 국한되어 있었고 실제 물리적 세계와의 상호작용에는 한계가 있었습니다. 로봇공학, 자율주행, 스마트 제조 등 다양한 분야에서 AI가 실제 물리적 환경을 이해하고 상호작용할 수 있는 능력이 주목받게 되며, Physical AI의 필요성이 대두되었습니다. 특히 항만, 터미널, 물류 등 산업 현장의 안전성을 높이고 생산성 향상을 높일 수 있는 기술이 중요한 해양산업의 경우, Physical AI 활용에 대한 수요가 증가할 것으로 예상됩니다.

[그림 3] AI 기술 유형 정의 및 발전 흐름



2. AI 생명주기 개요

AI 생명주기는 AI 서비스 기획부터 운영 및 활용에 이르는 전 과정을 체계적으로 관리하기 위한 절차를 의미합니다. AI 생애주기는 단계 간 상호작용이 지속적으로 이루어지는 반복적·순환적 구조를 가지며, 실제 적용 과정에서 반드시 정해진 순서대로 수행되는 것은 아닙니다.

본 가이드라인은 이해를 돕기 위해 단계별 흐름을 순차적으로 설명하였으나, 운영 환경과 기술적 요구사항에 따라 데이터 수집·가공, 모델 개발, 운영 등의 절차 순서가 조정될 수 있습니다.

[표 3] AI 서비스 생명주기별 주요 활동

생명주기 단계	주요 활동
1.생명주기 관리	• AI 시스템 관리 감독 조직 및 방안 마련 • AI 시스템 위험요소 분석 및 대응 방안 마련
2. 데이터 수집 및 처리	• 데이터 품질 확보, 데이터 사용자의 이해를 위한 정보 제공 방안 마련 • 데이터 라벨링 및 데이터셋 특성 문서화 • AI 모델 구축을 위한 데이터셋 마련
3. AI 모델 개발	• 비즈니스 목적에 따른 AI 모델 구현 • 구현된 AI 모델 확인 및 검증 • AI 모델 튜닝, 데이터 분석, 추가로 필요한 데이터 수집 • AI 모델에 대한 성능 평가
4. 시스템 구현	• 문제 발생 대비 안전모드 구현 및 알림 절차 수립 • AI 시스템 검증 및 사용자 설명에 대한 평가
5. 운영 및 모니터링	• 시스템 모니터링 및 AI 모델 재학습을 통한 성능 보장 • 모델 편향 탐지, 공정성, 설명가능성 등 시스템 신뢰성 모니터링 • 치명적 문제 발생 시 해결 방안 마련

Source: 과학기술정보통신부 · 한국정보통신기술협회, 2024 신뢰할 수 있는 인공지능 개발 안내서 (2024.02)

3. AI 생명주기 단계별 고려사항

본 가이드라인은 OEC 및 ISO/IEC에서 제시한 대표적인 AI 생명주기 모델을 참고하여, 실무자가 이해하기 쉽도록 AI 서비스 기획 단계부터 활용에 이르는 4단계 프로세스로 구성하였습니다. 기존 소프트웨어 개발 생명주기와 유사하나, AI 기술 특성 상 데이터 처리 및 모델 개발 단계가 더욱 강조되며, 각 단계별 활동 및 고려사항 역시 상이합니다.

[그림 4] AI 생명주기 단계별 고려사항

단계별 주요 수행 Task		상세 검토내용
기획 단계	• 전사 AI 활용 전략 및 방향성 수립	3.1.1
	• AI 서비스 도입 대상 및 목적 정의	3.1.2
	• AI 서비스 리스크 식별 및 분석	3.1.3
	• AI 서비스 상세화 및 초기 기술검증	3.1.4
개발 단계	• 데이터 수집 범위 및 기준 정의	3.2.1
	• 데이터 관리 계획 수립 및 준수	3.2.2
	• AI 모델 도입 및 개발	3.2.3
	• AI 모델 관리 계획 수립 및 준수	3.2.4
	• AI 시스템 및 이용자 인터페이스 구현	3.2.5
	• AI 시스템 테스트 계획 수립	3.2.6
운영 단계	• AI 서비스 모니터링 및 알림 체계	3.3.1
	• AI 리스크 사후 관리 체계	3.3.2
	• AI 서비스 설명 제공 방안	3.3.3
활용 단계	• AI 서비스 이용자 보호 방안	3.4.1
	• AI 교육 및 변화 관리	3.4.2
	• AI 산출물 저작권 관리	3.4.3

Chapter 02. 기본원칙 및 고려사항

3.1. 기획 단계

3.1.1 전사 AI 활용 전략 및 방향성 수립

AI 기획 단계에서 가장 먼저 수행해야 하는 과제는 기업의 전략, 규제 환경, 기술 운영 여건 등을 종합적으로 고려한 전사 AI 활용 전략과 방향성 수립입니다. AI는 단순 자동화 도구가 아니라, 업무 방식과 의사결정 구조, 조직 운영 방식 자체를 변화시키는 기술이기 때문에 전사 전략과의 정합성이 확보되지 않은 상태에서 개별 AI 과제가 추진될 경우 중복 투자, 활용 저조, 책임 불명확, 윤리적 · 법적 리스크 증가 등의 문제가 발생할 수 있습니다. 따라서 기업은 AI 도입에 앞서 전사 차원의 AI 활용 비전과 방향성을 먼저 수립해야 합니다.

AI는 기존 IT 시스템 · 데이터 · 업무 체계와 연계되어 운영되는 기술이므로, 전사 중장기 사업 전략 또는 DX · AX 전환 전략 등과의 정합성을 검토해야 합니다. 예를 들어 해양산업 현장의 안전 관리 고도화가 전사 중장기 전략 과제인 경우, 이상 징후 탐지, 영상 · 이미지 인식 기반 실시간 작업 현장 모니터링 등 현장 위험을 사전에 식별하고 실시간 대응할 수 있는 AI 과제를 정의하여 AI 활용 전략과 연계해야 합니다.

전사 AI 활용 전략과 방향성이 수립되면, 이를 기준으로 AI 추진 과제의 우선순위를 체계적으로 정의하는 단계가 필요합니다. 일반적으로 AI 과제는 여러 부서에서 동시에 제안되며, 동일한 기술이라 하더라도 적용 범위, 기대 성과, 위험 수준이 서로 상이하게 나타날 수 있습니다. 따라서 기업은 AI 전략 과제의 추진 우선순위를 합리적으로 결정하기 위한 판단 기준을 사전에 정립하여, 의사결정 시 전사 공통 기준으로 적용해야 합니다.

Chapter 02. 기본원칙 및 고려사항

AI는 채용, 평가, 안전 관리, 금융 등 사람의 권리와 의사결정에 직접적인 영향을 미치는 영역까지 활용 범위가 확대되고 있으며, AI로 인한 차별, 개인정보 침해, 설명 불가능한 의사결정 등 다양한 문제가 국제적으로 반복 발생하고 있습니다.

❖ AI 윤리 리스크 사례

챗GPT가 평소 정신 건강에 문제가 없었던 이용자의 자살과 망상 등을 유발했다는 소송이 미국에서 한꺼번에 7건 제기됐다. 이들은 GPT-4o가 위험할 정도로 이용자에게 아첨을 잘하며 이용자를 심리적으로 조종할 수 있다는 내부 경고가 있었는데도 출시됐다며, **오픈AI가 위법행위에 의한 사망, 조력 자살, 과실 치사 등에 책임이 있다고 주장했다.** 오픈AI는 지난 9월 10대 이용자의 **챗GPT 사용을 부모가 통제할 수 있는 기능을 내놨고, 캐릭터 AI는 10대 청소년의 챗봇 사용을 제한했다.**

Source: 연합뉴스, 오픈 AI 한꺼번에 7개 소송 피소…챗GPT가 자살 · 망상 유발 (2025.11)

AI 생명주기 전반에서 발생할 수 있는 사회적·법적 위험을 선제적으로 통제하기 위해 AI 윤리원칙의 필요성이 더욱 강조되고 있으며, 기업은 AI 윤리원칙 정립을 통해 AI 활용에 대한 조직 차원의 기본원칙과 윤리적 방향성을 명확히 정의해야 합니다.

2021년 UNESCO는 「인공지능 윤리에 관한 국제 권고안」을 채택하여, AI 기술의 윤리적 사용에 대한 글로벌 표준을 수립하고, 이를 통해 인류 공동체가 직면한 윤리적 도전 과제에 대한 국제적인 해법을 제시했습니다. 해당 권고안은 법적 강제력이 있는 규범이 아니라, 각국 정부와 기업이 자율적으로 책임 기반의 AI 윤리체계를 구축하도록 유도하는 국제 표준의 성격을 가지며, 이를 조직 내부 정책과 가이드라인으로 구체화해 실천 수준으로 전환할 것을 권고하고 있습니다. 따라서 AI 윤리원칙은 AI 기술의 제약 요건이 아니라, 지속 가능한 AI 활용을 위한 필수 전제 조건으로 인식되어야 하며, AI 기획 단계서부터 조직 차원의 명확한 기준으로 수립 · 관리될 필요가 있습니다.

Chapter 02. 기본원칙 및 고려사항

AI 윤리원칙을 정립하는 기업의 경우, 국내 AI 윤리 기준, 개인정보 보호 관련 법 · 제도, 산업별 규제 요건, 주요 국내외 가이드라인 등을 종합적으로 참고해야 합니다. 이때 윤리 원칙은 선언적 문구에 그치지 않고, 실제 AI 기획 · 개발 · 운영 · 활용 전 과정에서 적용 가능한 기준으로 구체화되어야 합니다. 일반적으로 다음과 같은 원칙들이 전사 AI 윤리 기준의 주요 구성 요소가 됩니다.

❖ AI 개발 및 활용 시 3대 윤리 원칙

인간성을 위한 인공지능 (AI for Humanity) 지향

인간 존엄성 원칙	• AI는 인간의 생명은 물론 정신적 · 신체적 건강에 해가 되지 않는 범위에서 개발 및 활용되어야 함
사회의 공공선 원칙	• AI는 지능정보사회에서 소외되기 쉬운 사회적 약자와 취약 계층의 접근성을 보장하도록 개발 및 활용되어야 함
기술의 합목적성 원칙	• 데이터 수집 시 관련 법/제도적 규정 등을 반드시 준수해야 함 – 개인정보 수집 시 개인정보보호법에 따라 법적/기술적 검토 수행

Source: 관계부처 합동, 사람이 중심이 되는 AI 윤리기준 (2020.12)

AI 윤리원칙은 AI 생명주기 전반에 실질적으로 반영될 수 있도록 내부 규정과 업무 절차, 의사결정 체계 등과 유기적으로 연계되어야 합니다. 예를 들면 신규 AI 서비스 기획 시 AI 윤리원칙 준수 여부를 사전 점검하는 절차를 의무화하고, 고영향 AI에 해당하는 경우 별도의 법적 검토를 거치는 내부 통제 장치를 마련하는 것이 이에 해당합니다. 또한 외부 솔루션이나 AI 모델을 활용하는 경우에도 동일한 윤리 기준을 적용하여 외부 기술 활용이 곧바로 사회적 · 법적 리스크로 전이되지 않도록 관리 체계를 구축할 필요가 있습니다.

Chapter 02. 기본원칙 및 고려사항

AI 윤리 원칙을 수립하여 선언한 대표적인 사례로 2025년 11월 행정안전부는 AI 기술을 통한 행정혁신을 촉진하면서 AI 사용에 대한 국민의 신뢰를 확보할 수 있도록 '공공부문 AI 윤리 원칙'을 발표했습니다. AI가 국민 생활에 미치는 영향을 고려하여 공공 부문에서 AI의 윤리적 사용에 대한 차별화된 원칙 마련의 필요성이 커짐에 따라, 행정안전부는 공공 부문 종사자가 실무 수행 시 자율적으로 점검할 수 있도록 6대 원칙에 따른 90여 개 세부 점검사항을 체크리스트 형식으로 구성하여 발표했습니다. 공공기관 외 주요 민간 기업들도 책임 있는 AI 활용을 위해 자체적인 윤리 원칙과 실행 기준을 수립하며 조직 전반의 AI 관리 체계를 강화하고 있습니다.

[표 4] 행정안전부 '공공부문 AI 윤리원칙'

목표	공공부문 AI을 통한 행정혁신 촉진과 국민신뢰 구축

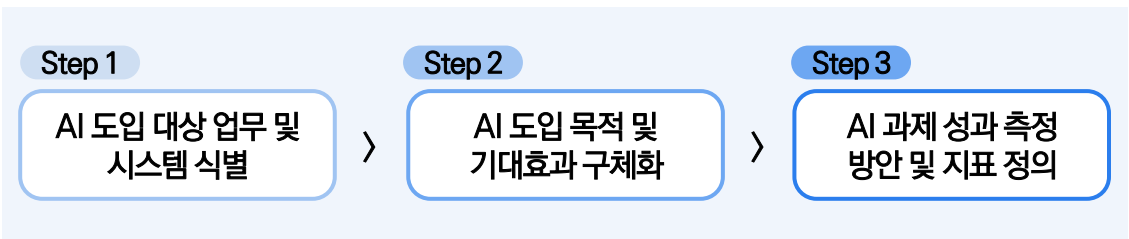
윤리원칙		주요 내용
국민	공공성	• 공공 AI 서비스를 공공의 이익과 국민의 복지를 위해 제공한다
	형평성	• 공공 AI 서비스를 모든 국민에게 공정하고 차별없이 제공한다
행정	투명성	• 공공 AI의 도입과 활용에 대한 과정을 투명하게 공개한다
	책임성	• 공공 AI에 대한 책임 주체를 명확히 정하고, 제도적 기준에 따라 활용한다
기술	안전성	• 공공 AI 시스템이 국민에게 피해를 주지 않도록 안전한 방식으로 운영한다
	프라이버시 보호	• 공공 AI 시스템이 개인정보와 사생활 등을 침해하지 않도록 보호장치를 마련한다

Source: 행정안전부, 공공부문 인공지능 윤리원칙안 개요 (2025.11)

3.1.2 AI 서비스 도입 대상 및 목적 정의

AI 서비스의 성공적인 도입을 위해선 AI 도입 대상과 목적을 명확히 정의하는 것이 중요합니다. 조직은 AI가 해결해야 할 문제를 정확히 인식하고, AI 도입이 가져올 변화와 기대효과를 실현할 수 있도록 방향성을 설정해야 합니다. AI는 모든 업무에 일괄적으로 적용되는 기술이 아니라, 특정 문제를 해결하기 위해 활용되는 도구입니다. 따라서 AI 도입 대상과 목적은 조직의 전략적 목표와 연계하여 설정되어야 하며, 이를 통해 효율성 향상, 비용 절감, 서비스 품질 개선 등의 구체적인 비즈니스 성과를 기대할 수 있습니다.

[그림 5] AI 도입 대상 및 목적 정의 프로세스



모든 업무가 AI를 통해 개선될 수 있는 것은 아니므로, AI 적용 시 실제로 효과를 낼 수 있는 업무와 시스템을 선별하는 과정이 필요합니다. AI 도입 전 현재의 업무 프로세스를 분석하여 AI가 해결할 수 있는 문제를 구체적으로 도출해야 합니다. 반복·수작업 업무 등 업무 가치 대비 투입 자원이 과도한 영역을 중심으로 요구사항을 식별하고, 기존 시스템에서 발생하는 병목 지점·지연 요소·리스크 관리 한계를 분석함으로써 AI 적용의 실질적 효과가 기대되는 업무와 기능을 명확히 정의해야 합니다. 이는 AI 적용 우선순위 설정의 기준이 되며, 이후 단계인 기대효과 정의, 성과 측정 지표 설정 등을 결정하는 핵심 요인으로 활용됩니다.

Chapter 02. 기본원칙 및 고려사항

AI 도입 목적은 구체적이고 명확한 목표를 전제로 설정되어야 합니다. 목표가 구체적이지 않을 경우 AI가 해결해야 할 문제와 그로 인한 기대 효과를 명확히 정의하기 어렵고, 이는 투자 대비 가시적인 성과를 확보하는 데 어려움을 초래합니다. AI는 문제 해결을 위한 수단이므로, 비즈니스 요구와 해결해야 할 과제를 중심으로 도입 목적을 명확히 정의하는 것이 중요합니다.

AI 도입 목적이 명확히 정의되면, 이를 측정 가능한 핵심 성과 지표(KPI, Key Performance Indicator)로 변환하여, 과제 성과를 평가할 수 있는 체계를 구축해야 합니다. 성과 지표는 도입 전후의 변화를 정확히 반영할 수 있어야 하며, 정량적 지표와 정성적 지표를 모두 포함해야 합니다. 예를 들어 비용 절감을 목표로 AI를 도입하는 경우, 자동화가 가능한 작업 범위와 절감 가능한 비용 항목을 명확히 정의하고, 이를 측정 가능한 지표로 설정해야 합니다. 이를 통해 AI 시스템의 효과성과 성공 여부 등을 체계적으로 평가할 수 있습니다.

[표 5] 해양산업 주요 AI 도입 목표 및 기대효과

/ Examples /

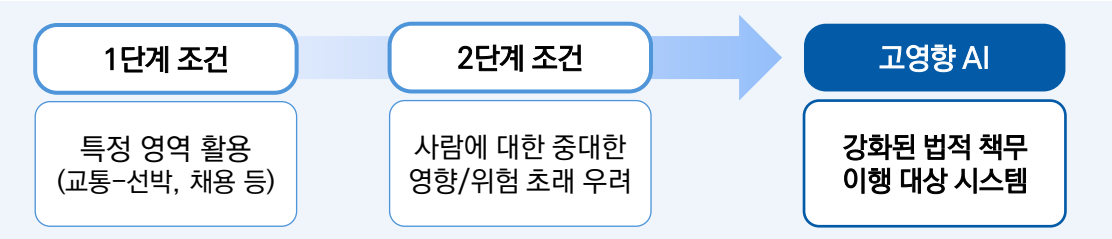
AI 도입 목표	기대효과	KPI
선박 운항 효율성 향상	• 연로 소모 절감 • 운항 시간 단축 • 탄소 배출 감소	• 연료 절감량(톤) • 운항 시간 단축률(%) • 탄소 배출 감소율(%)
해양 물류 최적화	• 물류 처리 시간 단축 • 항만 효율성 향상 • 비용 절감	• 처리 시간 단축률(%) • 물류 비용 절감액(원) • 항만 처리 용량 증가(%)
항만 안전 관리	• 항만 위험 탐지 정확도 향상 • 항만 사고 예방 • 항만 운영 비용 절감	• 항만 위험 탐지 정확도(%) • 항만 사고 예방률(%) • 항만 운영 비용 절감액(%)

3.1.3 AI 서비스 리스크 식별 및 분석

도입 및 개발하고자 하는 AI 서비스에서 발생 가능한 리스크를 사전 파악하여, 그 원인과 발생 조건을 분석한 후, 해당 리스크가 AI 시스템이나 사람, 주변 환경에 미치는 영향을 평가해야 합니다. 이 과정에서 기획하고자 하는 AI 서비스의 리스크가 극단적으로 부정적인 결과를 초래할 수 있다고 판단된 경우, 해당 AI 기술 적용에 대한 재검토를 권고합니다.

AI 기본법은 사람의 생명, 신체 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 AI를 '고영향 AI'로 정의하며, 고영향 AI 및 이를 활용한 제품 · 서비스를 제공하는 AI 사업자에게 강화된 법적 책무를 부여합니다. 특히 해양산업은 <교통:선박> 분야가 고영향 AI 활용 영역으로 지정된 만큼, AI 도입 및 적용 시 신중한 검토가 필요합니다.

[그림 6] 고영향 AI 평가 기준



법/규제 조항

과태료 대상

AI 기본법 시행령안 제24조(고영향 인공지능의 확인 절차 등)

① 인공지능사업자가 법 제33조제 1항에 따라 고영향 인공지능 해당 여부의 확인을 요청하려는 경우에는 별지 서식의 확인 요청서를 과학기술정 보통신부장관에게 제출해야 한다.

② 과학기술정보통신부장관은 다음 각 호의 사항을 고려하여 고영향 인공지능 해당 여부를 판단하여야 한다.

1. 인공지능이 법 제2조제4호 각 목의 어느 하나의 영역에서 활용되는지 여부
2. 사람의 생명, 신체의 안전 및 기본권에 초래할 수 있는 위험의 영향, 중대성, 빈도 및 활용 영역별 특수성

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

Chapter 02. 기본원칙 및 고려사항

해양 기업은 고영향 AI로 판단될 수 있는 10가지 활용 영역 중 '교통:선박', '채용'. '대출 심사'을 중점적으로 검토하여 고영향 AI 사업자로서의 책무 이행 여부를 확인할 필요가 있습니다. 위 3가지 영역은 해양산업의 특성과 AI 도입 가능성이 높은 업무 영역을 고려하여 선정한 것이며, 이 외의 영역에서 AI 서비스를 개발 및 배포하는 경우에도 고영향 AI 해당 여부를 검토하여 부과된 사업자 책무를 의무 이행해야 합니다.

[표 6] 고영향 AI 판단 영역별 기본권 침해 상황

영역 구분	생명, 신체의 안전 및 기본권 관련성	AI가 중대한 영향을 미치거나 위험을 초래하는 경우
교통: 선박	• 차량 · 선박 등 사고 시 탑승자, 피해자 등의 생명 · 신체 안전과 이용자의 이동권과 관련 • 교통체계의 경우 차량 등 이용자의 이동권과 관련	• AI 시스템의 의도하지 않은 작동 등으로 인한 충돌 위험의 회피나 항로 설정 및 변경의 오판은 사람의 생명, 신체의 안전에 영향을 미치거나 위험을 초래할 우려 발생
채용	• 근로계약을 체결하거나 유지하려는 자의 직업선택의 자유 등 노동권 및 평등권과 관련	• 채용 과정에서 사람에 대한 판단 또는 평가 목적으로 이용되는 AI 시스템에 의한 의사결정은 지원자의 채용 기회 박탈 등 불합리함 발생 가능
대출 심사	• 금융소비자의 평등권, 재산권, 인간다운 생활을 할 권리와 관련	• 대출 심의 · 결정 업무에 사용되는 AI 시스템의 의도하지 않은 작동 등으로 특정 기준의 오판 발생 시, 소비자의 금융거래계약의 체결 · 유지 등에 직접적인 침해 발생

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

AI 기본법에서 생명과 신체의 안전은 인간의 기본적인 권리로서 중대한 권익으로 간주되며, AI가 가지는 위험성을 판단하기 위해서는 AI 시스템의 의도된 목적, 기능, 활용 맥락을 다양한 관점에서 종합적으로 고려해야 합니다.

Chapter 02. 기본원칙 및 고려사항

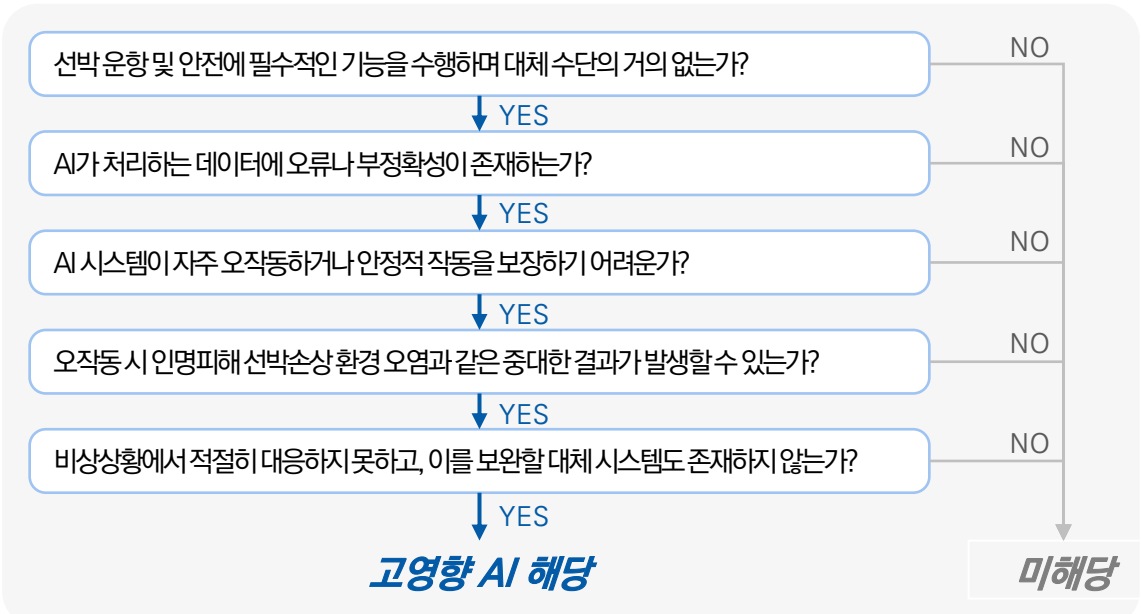
교통: 선박

'교통: 선박' 분야의 고영향 AI는 선박의 운항과 관련하여 활용되는 AI 시스템을 의미합니다. 이러한 AI 시스템들은 주로 해양 안전 향상, 효율성 증대, 인적 오류 감소, 비용 절감 등을 목표로 도입됩니다. 대표적인 예로 선원이 개입하지 않는 완전자율운항선박의 경우, AI 기반의 자율 제어 기능이 선박 운항과 안전에 필수적인 만큼, AI 오작동 시 인명, 환경, 선박 손상에 대한 위험이 발생할 가능성이 큽니다. 또한 비상 상황에서 즉각적인 대응이 어려운 점을 고려할 때, 자율운항 AI 시스템은 고영향 AI에 해당합니다.

❖ 해양산업 고영향 AI 해당 사례

[예시] 자율운항선박

A 선사는 선박 운항의 효율성 증대 및 안전 강화를 위해 자율운항 인공지능시스템을 도입하여 운항 자율 제어 기술을 활용하고 있습니다.
선박의 실시간 운항 상황을 모니터링하고, 장비와 시스템 상태를 분석하여 항로를 자동으로 조정하고 엔진을 제어합니다.



Chapter 02. 기본원칙 및 고려사항

채용

채용 분야에서 AI 시스템은 모집부터 서류 심사, 실기 평가 등 채용 전 과정에서 사용될 수 있습니다. 이 과정에서 AI가 구직자의 직업 선택의 자유 및 평등권에 중대한 영향을 미칠 우려가 있는 경우 해당 AI 시스템은 '고영향 AI'로 분류될 수 있습니다.

채용 과정에서 AI 시스템이 고영향 AI에 해당하는지 여부를 판단하기 위해서는 두 가지 주요 요소를 고려해야 합니다. AI가 도출한 결과가 채용 결정에 실질적인 영향을 미치는 지와 채용 결정 과정에서 인간이 실질적으로 개입하는 지 해당 여부를 확인해야 합니다.

AI의 판단 결과가 채용 여부에 중대한 영향을 미치고, 인간의 개입이 미미하거나 AI 결과에 의존하여 의사결정을 내리는 경우, 해당 AI 시스템은 고영향 AI로 판단될 수 있습니다. 특히, 개인의 동의나 검토 없이 AI가 직접적으로 구직자 정보를 제공하거나, 채용 여부를 결정하는 경우 개인의 자율성, 직업 선택의 자유, 평등권 등이 제한될 수 있어 차별적 결과를 초래할 위험이 높습니다.

❖ 국내외 채용 AI 리스크 발생 사례

AI 선발 프로그램 고령자 차별 사건

2022년 5월 미국 고용기회평등위원회는 온라인 교육 업체 아이튜터그룹을 상대로 소송을 제기했다. 해당 기업이 사용하는 **AI 선발 프로그램이 고령자를 차별하고 있어 고용연령차별법을 위반**했다는 혐의를 제기했다. 이듬해 법원은 아이튜터그룹이 사용한 AI 선발 프로그램이 55세 이상 여성과 60세 이상 남성을 거부하도록 설정돼 있다는 사실을 확인하고, **차별 피해자들에게 총 36만 5,000달러(약 5억원)의 합의금과 차별 구제 수단을 제공할 것을 선고**했다.

Source: 정보통신정책연구원, 미국과 한국의 AI 채용 분야 정책 현황(2025.05)

Chapter 02. 기본원칙 및 고려사항

대출 심사

대출 심사는 개인의 경제적 권리와 의무에 직접적인 영향을 미치는 절차이기 때문에, 이 과정에서 AI 시스템이 핵심적인 판단 기능을 수행하는 경우 고영향 AI로 분류될 수 있습니다. 대출 심사에서 심사 대상은 원칙적으로 자연인인 금융소비자를 의미하지만, 개인사업자가 사업 목적의 대출을 신청하는 경우에도 개인 정보가 활용되는 만큼 기본권 침해 가능성 측면에서 자연인과 동일하게 취급될 수 있습니다. 또한 정책금융과 일반금융, 공공 · 민간 등 대출의 성격에 따라 적용되는 기준과 위험 요소가 달라질 수 있으므로, 각 사례별로 AI의 영향력과 개입 수준을 종합적으로 평가해 고영향 AI 해당 여부를 판단해야 합니다. 해양기업은 한국해양진흥공사 등 공공기관을 통해 선박금융자금을 대출받아 사업을 운영하는 경우가 있는 만큼, 대출 심사에서 AI가 최종 결정을 직접 수행하거나 최종 판단에 실질적인 영향을 미치는 경우, 고영향 AI 해당 여부를 반드시 확인해야 합니다. 또한 성별, 종교, 국적 등의 차별적 요소 또는 사상, 신념, 정치적 견해 등 민감한 정보를 AI의 입력 데이터로 사용하는 경우 법적 · 사회적 책임이 수반되므로 신중한 검토가 필요합니다.

❖ 국내외 대출심사 AI 리스크 발생 사례

학자금 대출심사 AI 불공정성 소송

미국 매사추세츠주 법무장관실에서 학자금 대출회사인 Earnest Operations 대상으로 소송을 제기했다. 특정 대학의 평균 부실률 데이터를 AI 모델의 입력 변수로 사용하여, **흑인과 히스패닉 신청자에게 불리한 대출 조건을 부여**한 것으로 나타났다. 또한 영주권이 없는 지원자는 신용도나 상환능력 평가 없이 **사전 심사 단계에서 자동으로 거절**하는 '녹아웃 규칙'을 적용했다.
Earnest Operations은 2.5백만 달러로 합의금을 체결했으며, 이는 **AI 도구 사용으로 인한 차별적 결과에 대해 제기된 최초의 주정부 집행 사례**로 기록된다.

Source: ABA Banking Journal, Mass. AG reaches settlement with student loan firm for $2.5M over AI lending bias (2025.8)

Chapter 02. 기본원칙 및 고려사항

2025년 9월 발표된 고영향 AI 판단 가이드라인안에 따르면, AI 시스템 활용에 따라 개인의 금융거래계약의 체결 · 유지 등에 직접적인 차별이 발생할 경우에는 개인의 권리에 중대한 위험을 초래하는 것으로 볼 수 있습니다. 다만 실제 AI 시스템 활용 사례별 위험도를 판단함에 있어서는 개별 상황에서 나타나는 위험 요소를 종합적으로 고려해야 합니다.

하기 표는 대출 심사 분야의 고영향 AI 판단 기준을 정리한 것으로 A그룹 항목 중 2개 이상에 해당하거나, A그룹 항목 1개와 B그룹 항목 2개 이상에 해당하는 경우 '고영향 AI'로 분류됩니다.

[표 7] 대출 심사 분야 고영향 AI 판단 기준

그룹	항목	점수
A	• 기존 모델보다 파라미터 증가, 학습 데이터 양 또는 유형의 확대 등으로 복잡도가 증가한 신규 모델에 기반한 AI 시스템	2
	• 1만명 이상의 금융소비자를 대상으로 하는 AI 시스템	2
	• 자동화 정도가 높고 속도가 빨라서 실질적으로 최종 의사결정에 관한 사람의 개입이 불가능한 AI 시스템	2
B	• 대리 변수(Proxy Variable)를 활용하는 빈도가 높은 AI 시스템	1
	• 특정 업무 영역을 위해 개발 또는 고도화된 AI 시스템을 다른 업무 영역에 일시적으로 또는 단기간 활용하는 경우	1
	• 국외 데이터를 주로 학습하여 개발된 국외 AI 시스템을 활용하는 경우	1

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

Chapter 02. 기본원칙 및 고려사항

교통:선박, 채용, 대출 심사 등에서 고영향 AI 혹은 해당 AI를 활용한 제품 및 서비스를 제공하는 AI 사업자는 AI 기본법에 의거하여 사업자 책무를 이행해야 합니다. AI 사업자는 고영향 AI의 위험을 체계적으로 관리하기 위하여 담당 조직 또는 인력을 중심으로 위험관리 계획을 수립하고, 그 책임을 이행하기 위해 노력해야 합니다. 위험관리 계획에는 AI 전 생명주기에 걸쳐 발생할 수 있는 위험에 대한 식별, 분석, 평가와 그에 대한 적절한 대응 방안을 포함해야 합니다. 또한 가능한 범위에서 AI 위험관리를 담당하는 조직을 구성하거나 인력을 지정하여 위험관리 체계를 운영해야 합니다.

법/규제 조항

과태료 대상

AI 기본법 시행령안 제26조(고영향 인공지능과 관련한 사업자의 책무)
① 인공지능사업자는 법 제34조제1항 각 호의 조치 중에서 다음 각 호에 해당하는 내용을 자신의 인터넷 홈페이지 등에 게시하여야 한다. 다만, 「부정경쟁방지 및 영업비밀보호에 관한 법률」 제2조제2호에 따른 영업비밀에 해당하는 사항은 제외할 수 있다.
1. 위험관리정책 및 조직체계 등 법 제 34조제1항제1호에 따른 위험관리방안 의 주요 내용
2. 법 제34조제1항제2호에 따른 설명방안의 주요 내용
3. 이용자 보호 방안
4. 해당 고영향 인공지능을 관리 · 감독하는 사람의 성명 및 연락처
· · ·

관련 기술고시 조문 제 4조 (위험관리방안의 수립 · 운영)
① 인공지능개발사업자 및 인공지능이용사업자(이하 "사업자"라 한다)는 고영향 인공지능의 위험관리를 위하여 다음 각 호의 사항이 포함된 위험관리방안을 수립 · 운영하여야 한다.
1. 위험관리정책 수립 및 이행
2. 위험관리 조직체계 수립 및 운영
② 사업자는 제1항의 위험관리방안을 문서로 작성하여 관리하고 인공지능시스템의 수명주기의 모든 과정에서 이를 준수하여야 한다.
③ 사업자는 최신의 기술 및 관리방법론이 적용될 수 있도록 위험관리방안을 주기적으로 점검 · 갱신하고 그 변경내역을 관리하여야 한다. ·

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

Chapter 02. 기본원칙 및 고려사항

상기와 같은 고영향 AI를 활용한 서비스를 제공하는 사업자는 AI 시스템의 설계부터 데이터 수집, 모델 학습, 테스트 등 개발 전 과정에서 발생 가능한 사람의 생명, 신체의 안전 및 기본권에 대한 잠재적 위험을 식별하기 위해 노력해야 합니다. 이를 위해 AI 시스템의 본래 활용 목적 뿐 아니라 장애, 오남용 등으로 인해 발생할 수 있는 위험도 포함하여 분석 및 평가해야 합니다. 또한 GenAI 등 AI 유형별 특성과 AI 활용 영역에 따른 위험의 심각도와 발생 빈도를 평가하고, 관련 이해관계자에 대해 정량적 또는 정성적 분석을 수행해야 합니다.

AI 기본법 제 35조는 고영향 AI 제품 또는 서비스를 제공하는 AI 사업자에게 해당 제품 및 서비스에 대한 영향평가 수행을 권고하고 있습니다. AI 영향평가란 고영향 AI를 이용한 제품 또는 서비스가 사람의 기본권에 미치는 영향을 사전에 평가하는 절차로 책임 있는 AI 활용과 사회적 신뢰 제고를 위해 고영향 AI 사업자 뿐만 아니라, 모든 AI 사업자가 자율적으로 영향평가를 수행할 것을 장려합니다. 이를 위해 고영향 AI 사업자는 영향평가를 직접 수행하거나, 전문성을 보유한 제3자에 의뢰하여 수행할 수 있습니다.

법/규제 조항

AI 기본법 제35조(고영향 인공지능 영향평가)

① 인공지능사업자가 고영향 인공지능을 이용한 제품 또는 서비스를 제공하는 경우 사전에 사람의 기본권에 미치는 영향을 평가(이하 “영향평가”라 한다)하기 위하여 노력하여야 한다.

② 국가기관등이 고영향 인공지능을 이용한 제품 또는 서비스를 이용하려는 경우에는 영향평가를 실시한 제품 또는 서비스를 우선적으로 고려하여야 한다.

③ 그 밖에 영향평가의 구체적인 내용 · 방법 등에 관하여 필요한 사항은 대통령령으로 정한다

· · ·

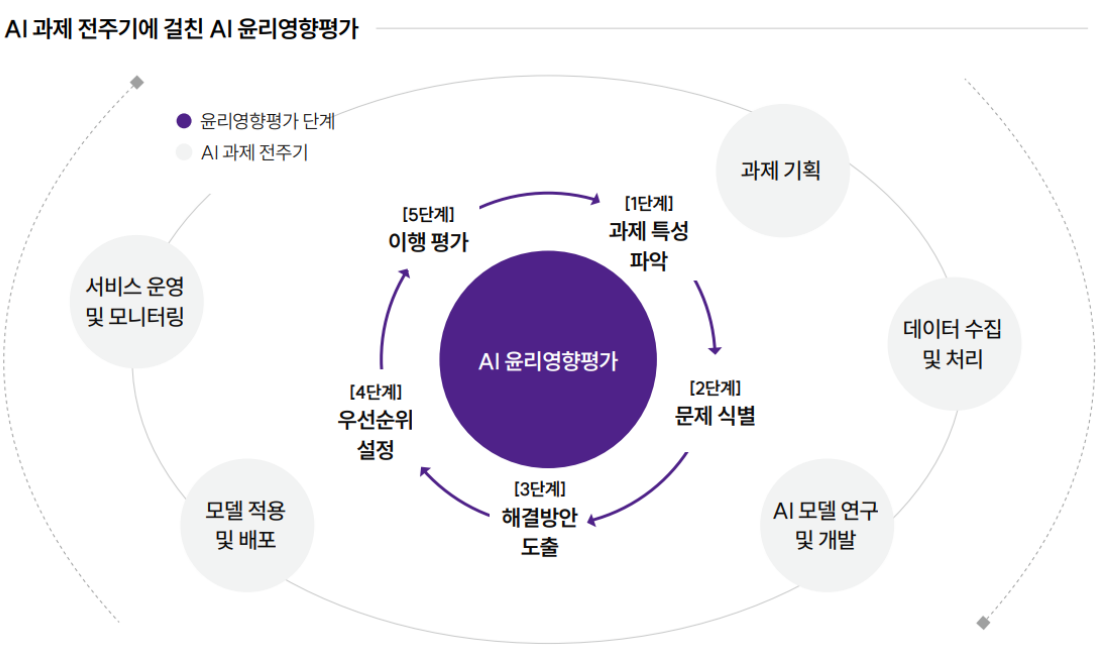
Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

Chapter 02. 기본원칙 및 고려사항

고영향 AI 사업자는 식별된 위험 요소별로 인명 피해와 사고를 방지하거나 부정적 영향을 최소화할 수 있는 적절한 처리 방안을 수립하고 실행해야 합니다. 또한 처리 방안을 적용한 이후에는 위험 요소의 파급효과를 재평가하여 위험이 실제로 제거 · 방지 · 완화되었는지 확인하고, 그 결과를 위험관리 체계 개선에 반영해야 합니다.

참고 사례로, LG AI 연구원은 AI 윤리원칙을 실행하기 위한 절차로서 AI 과제의 기획부터 완료까지 이르는 전 단계에서 윤리적 영향을 사전 점검하는 'AI 윤리영향평가'를 도입했습니다. 이는 일반적인 체크리스트 형태의 위험 점검이 아닌, 구체적인 문제 발생 시나리오와 해결 방안, 책임자 및 이행 시점 등을 논의하고 기록할 수 있는 형태로 구성되어 있습니다.

[그림 7] LG AI 연구원 'AI 윤리영향평가 '



Source: LG AI 연구원, 2024 LG AI 윤리 책무성 보고서 (2025.02)

Chapter 02. 기본원칙 및 고려사항

또한 AI 모델 학습에 사용된 누적 연산량이 AI 기본법 시행령안에서 규정한 기준 이상인 경우, AI 기본법 제32조에 따라 안전성 확보를 위한 위험 식별 평가 및 완화, 위험 관리 체계 구축 등의 사항을 이행해야 합니다. 또한 AI 기술의 발전 수준을 고려할 때 현재 활용되는 AI 기술 중 최첨단의 AI 기술을 적용한 경우와 AI 기술의 위험도가 사람의 생명, 신체의 안전 및 기본권에 광범위하고 중대한 영향을 미칠 우려가 있을 경우에도 AI 안전성 확보 의무를 이행해야 합니다. 의무 대상 판단 기준은 모델 단위가 아닌 시스템 단위로 판단되나, 동일 모델이라도 시스템 구성에 따라 다른 AI 시스템으로 간주되어 적용 대상 여부가 달라질 수 있는 만큼 전문 기관의 자문 혹은 법률 검토를 수행할 것을 권고합니다.

법/규제 조항

과태료 대상

AI 기본법 제32조(인공지능 안전성 확보 의무)

제32조(인공지능 안전성 확보 의무)

① 인공지능사업자는 학습에 사용된 누적연산량이 대통령령으로 정하는 기준 이상인 인공지능시스템의 안전성을 확보하기 위하여 다음 각 호의 사항을 이행하여야 한다.

1. **인공지능 수명주기 전반에 걸친 위험의 식별 · 평가 및 완화**
2. **인공지능 관련 안전사고를 모니터링하고 대응하는 위험관리체계 구축**

② 인공지능사업자는 제1항 각 호에 따른 사항의 이행 결과를 과학기술정보통신부장관에게 제출하여야 한다.

AI 기본법 시행령안 제23조(인공지능 안전성 확보 의무)

① 법 제32조제1항에서 “대통령령으로 정하는 기준 이상인 인공지능시스템”이란 학습에 사용된 누적 연산량이 10의 26승 부동소수점 연산 이상인 인공지능시스템으로서 과학기술정보통신부장관이 인공지능기술 발전 수준, 위험도 등을 고려하여 고시하는 기준에 해당하는 인공지능시스템을 말한다.

② 제1항에 따른 고시에는 학습에 사용된 누적 연산량의 구체적인 산정 방식을 포함하여야 한다. · · ·

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

3.1.4 AI 서비스 상세화 및 초기 기술검증

앞서 정의한 AI 도입 목적 및 기대효과, 리스크 분석 결과 등을 바탕으로 AI 서비스가 실제로 구현 가능한 수준인지, 그리고 조직의 업무 환경에 적합하게 작동할 수 있는지 확인하는 AI 서비스 사전 검증 단계가 필요합니다. 단순히 기술 동작 여부를 확인하는 절차가 아닌, AI 서비스의 기능 구조, 데이터 흐름, 보안 체계, 운영 가능성 등을 종합적으로 검토하기 위해 AI 서비스를 상세화 · 현실화하는 과정에 해당합니다. 특히 GenAI와 같이 범용성이 높은 AI 기술은 기획 단계에서 기대한 수준과 실제 현장 적용 결과 간의 간극이 크게 발생할 가능성이 높은 만큼, PoC 등을 통해 기술의 가능성과 한계를 동시에 검증하고, 조직 환경에 적합한 형태로 AI 서비스를 보완하는 절차가 필요합니다.

AI 서비스 상세화는 기획 단계에서 정의한 AI 도입 대상과 목적을 실제 AI 시스템 구조와 서비스 기능 수준으로 구체화하는 과정입니다. 예측, 분류, 추천, 생성 등 AI가 수행하는 핵심 기능을 명확히 정의하고, 각 기능이 입력 데이터를 받아 어떤 형태의 출력 결과를 도출하는지 서비스 흐름 단위로 정리해야 합니다. 이 과정에서 개인 · 민감 정보 포함 여부, 외부 데이터 연계 여부, 실시간 처리 여부 등 데이터 흐름과 처리 구조를 구체화합니다. 또한 AI 유형에 따른 보안 리스크, 운영 방식 등도 사전 고려해야 하며, GenAI의 경우 학습 범위 제한, 출력 통제 방식 등을 AI 서비스 상세화 단계에서 함께 설계해야 합니다. 이와 함께 기존 업무 시스템과 AI 서비스 간의 연계 방식 · 구조를 사전 정의하여, AI 도입 및 적용 시 필요한 인프라 변경 등 추가 공수 발생 여부를 사전에 파악해야 합니다.

Chapter 02. 기본원칙 및 고려사항

초기 기술검증은 실 서비스 도입/적용 전 상세화한 AI 서비스 구조가 일정 수준 이상으로 안정적으로 작동할 수 있는지를 확인하는 실증 단계입니다. 단순 기술 데모가 아닌, AI 도입 및 개발을 위한 본 사업 착수 여부를 결정하기 위해 AI 서비스의 사업성 · 안전성을 판단하는 중요 과정입니다. 초기 기술검증의 주요 목적은 AI 모델이 실제 데이터와 업무 환경에서도 요구 성능을 충족하는지 확인하고, 현업 이용자 관점에서 AI 결과가 실제 업무에 활용 가능한 수준인지 검증하는 것입니다. 또한 AI 기획 단계에서 사전 식별된 리스크의 통제 가능 여부를 확인하여, AI 개발 이후 법적 · 윤리적 리스크 발생 등으로 인한 AI 제품 · 서비스의 정식 배포 재검토, 사업 중단 등 기업 이미지 훼손과 비즈니스 손실 발생을 사전 방지하기 위해 노력해야 합니다.

❖ AI 서비스 초기 기술 검증 시 주요 원칙

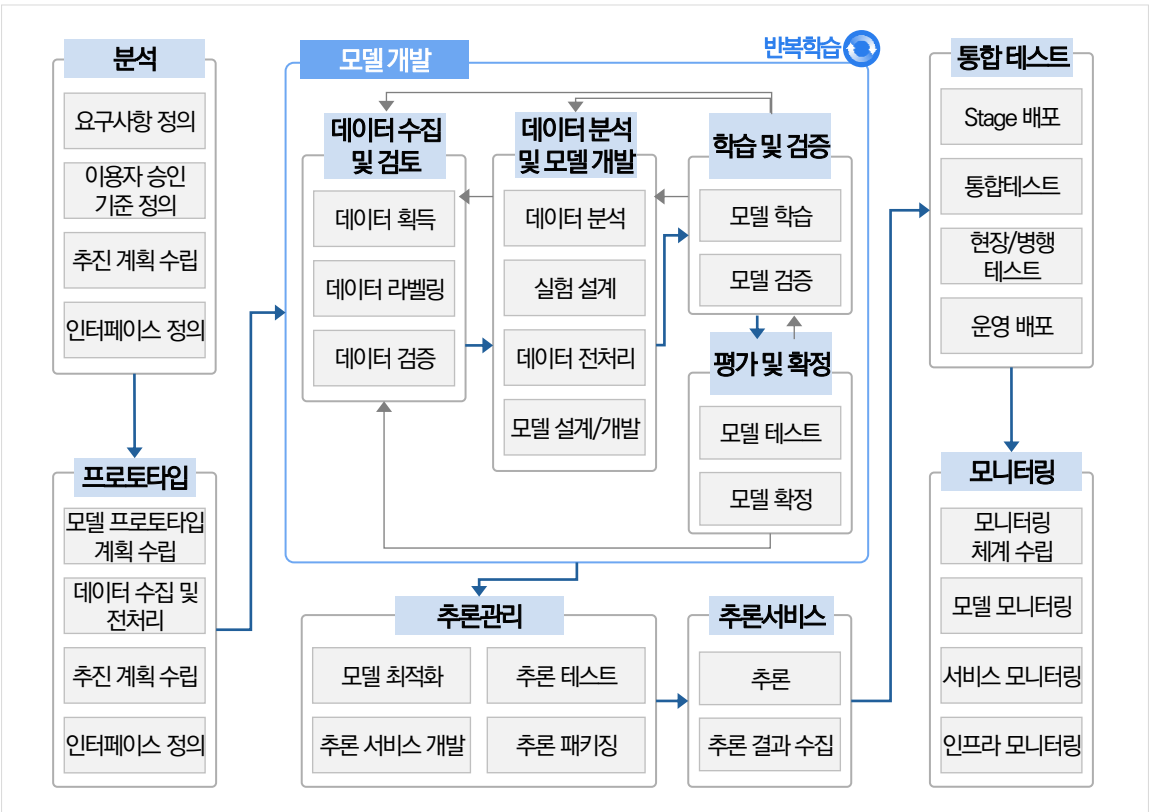
1	운영 환경에서 발생 가능한 오류와 위험 등을 제대로 평가하기 위하여 가능한 범위 내 **실제 업무 데이터를 기반으로 검증해야 함**
2	단순 정확도나 응답 속도 뿐만 아니라 개인정보 노출, 환각 등 **기술 성능과 함께 AI 안전성 · 윤리성도 동시 검증해야 함**
3	AI를 업무에 실질적으로 활용하는 현업 이용자 참여를 통해 **AI 서비스의 업무 적합성, 사용 편의성 등을 종합적으로 평가해야 함**
4	초기 기술검증 결과는 단순 보고가 아닌 확대 추진, 보완 후 재검토, 적용 범위 축소, 도입 중단 등 **단계적 의사결정이 이루어져야 함**

Source: 과학기술정보통신부 · 한국정보통신기술협회, 2024 신뢰할 수 있는 인공지능 개발 안내서 (2024.02)

3.2. 개발 단계

개발단계에서는 AI 시스템 개발 프로세스를 기반으로, 해양 기업이 안정적으로 AI를 개발하기 위해 데이터 수집, 모델 개발, 시스템 구현 등 각 단계에서 필수 고려해야 하는 주요 원칙과 검토 사항을 제시합니다. 또한 향후 운영 단계를 고려하여 개발 단계에서 사전 설계해야 하는 필수 기능과 요건을 구체화할 수 있도록 안내합니다. 특히 고영향 AI 사업자의 경우 데이터 투명성, 모델 설명가능성 확보 등 사업자 책무 요건을 사전에 면밀히 검토하여 AI 시스템 개발 시 이를 준수할 수 있는 체계를 마련해야 합니다.

[그림 8] AI 시스템 개발 프로세스



3.2.1 데이터 수집 범위 및 기준 정의

데이터 수집은 AI의 학습에 필요한 데이터를 직접 수집 또는 생성하거나, 이미 데이터를 보유하고 있는 조직이나 시스템 등으로부터 적법하게 확보하는 활동을 의미합니다. 학습 데이터가 AI의 결과와 성능에 영향을 미칠 수 있는 중요 요소인 만큼, AI 개발 시 필요 데이터를 명확하게 정의하여 데이터 수집 범위와 기준을 구체화해야 합니다. 또한 AI 개발 목적과 관련성이 낮은 정보는 학습 데이터에서 제외하여 데이터 품질과 AI 모델의 적합성을 확보하는 것이 바람직합니다.

고영향 AI 사업자는 데이터 수집 과정에서 기술적으로 가능한 범위 내에서 필요한 최소한의 데이터만을 확보해야 하며, 수집 목적과 다른 용도로 데이터를 사용하는 것을 엄격히 제한해야 합니다.

❖ 데이터 획득 · 수집 시 주요 고려사항

데이터 출처 신뢰성	• 데이터 수집 시 출처의 객관성과 신뢰성이 확보되어야 함 – 출처가 분명한 학술지, 발행 기관이 명시된 뉴스 등이 해당
데이터 다양성	• AI 모델이 현실을 잘 반영하고 본래의 구축 목적을 달성할 수 있도록 데이터 수집 시 다양한 시간, 공간, 집단 수준 등이 포함되어야 함
데이터 수집 적법성	• 데이터 수집 시 관련 법/제도적 규정 등을 반드시 준수해야 함 – 개인정보 수집 시 개인정보보호법에 따라 법적/기술적 검토 수행

Source: 과학기술정보통신부 · 한국지능정보사회진흥원 · 한국정보통신기술협회, AI 데이터 품질관리 가이드 (2025.02)

Chapter 02. 기본원칙 및 고려사항

AI 학습 데이터는 신뢰할 수 있는 출처에서 수집해야 합니다. '신뢰할 수 있는 데이터'란 공공기관, 인증된 데이터 제공 업체, 연구기관 등 공신력 있는 기관에서 제공되는 데이터를 의미합니다. 특히 해양산업에서 AI는 선박 운항, 해양 환경 모니터링 등 안전과 직결된 중요 업무에 활용되기 때문에, 수집되는 데이터의 정확성과 신뢰성을 반드시 검토해야 합니다.

해양산업 데이터는 수출입, 물류, 환경 등 여러 산업 분야와 연계되어 있으며, 다양한 분야의 신뢰할 수 있는 데이터를 안정적으로 확보하기 어렵습니다. 이에 정부는 해양수산 빅데이터 플랫폼 구축 등 다양한 정책을 통해 해양수산 데이터의 개방 및 활용을 촉진하고 있으며, 해양산업의 데이터 기반 혁신을 적극적으로 지원하고 있습니다.

[표 8] 대표적인 해양 데이터 유형

/ Examples /

데이터 유형	주요 내용
항만	• 항만 실시간 입출항 정보, 석유 · 화학 화물 물동 정보 등 • 연안 항만 구역, 철탑 위치 정보 등 • 전세계 주요 항구간 최적 운항경로 등
해운 물류	• 해상수출 물류가격(견적서, 인보이스, 계약서, 포워더 등) • 컨테이너 서비스 스케줄 정보, 석유 · 화학 화물 물동 정보 등
해양 환경	• 한반도 연안 및 근해역의 해양환경 데이터 • 연안 오염 의심 해역 수질 측정 데이터
해사 안전	• 선박 온실가스 배출 정보, 내항 여객선 정보(운항, 기상, 항로, 기항지 등) 등 • 선박검사 대상 선박 정보(어선, 일반선 등), 어선 정보(측정, 마력 등), • 해상 및 내수면 발생 해양 사고 데이터
선박 관리	• 선박별 운항 성능(속도, 화물물송장 등) • 일반 선박 현황 정보, 선박 운항 정보(AIS, 운항 경로 등) • 선박 공간 환경 정보(기관실, 화물실 등), 기자재 및 장비 현황 정보 등

Source: 과학기술정보통신부 · 한국지능정보사회진흥원, 2024 빅데이터 플랫폼 & 센터 (2024. 11)

Chapter 02. 기본원칙 및 고려사항

AI 모델 학습에서 데이터 다양성 확보는 모델의 성능과 공정성을 높이는 데 필수적인 요소입니다. 사람은 무의식적으로 특정 정보에 대해 편향된 선택을 할 수 있으며, 이러한 편향은 AI 모델에 그대로 반영될 수 있습니다. 따라서 데이터 수집 기준을 명확히 설정하고, 이를 검수하는 과정에서 데이터의 다양성을 확보할 수 있는 체계를 마련해야 합니다. 예를 들어 데이터 수집 작업자들이 특정 조건에 맞는 데이터만 선택하지 않도록 데이터 수집 작업 가이드라인을 마련하고, 특정 배경과 성향을 배제하기 위해 다양하고 충분한 수의 작업자와 검수자를 모집하는 방법이 있습니다.

데이터 수집 및 생성 시 장비의 사양 및 수집 환경 등 물리적 요인으로 인해 제한된 상황과 시나리오에 대한 데이터만 수집되는 등의 편향이 발생할 수 있습니다. 특히 해양산업의 AI 시스템은 센서 등 특정 하드웨어 기반으로 데이터를 수집하는 경우가 많아 장비의 측정 오차나 물리적 특성이 데이터 품질에 직접적인 영향을 미칠 수 있습니다. 따라서 기업은 내부 운영 데이터와 외부 환경 데이터를 균형 있게 확보·검증하여 데이터 편향을 최소화해야 하며, 각 데이터의 특성과 한계를 면밀히 분석한 후 AI 모델 학습에 활용해야 합니다.

❖ 해양산업 데이터 활용 대표 사례

해양환경 및 AI 데이터 신생 기업인 A사는 **전문 잠수부들이 촬영한 수천 장의 해저 쓰레기 사진을 정제하고 라벨링하여, 해양 폐기물을 AI로 식별 · 분류하는 데 필요한 '해양 침적쓰레기 이미지 데이터 세트'를 구축했다.**
이 데이터는 추후 AI 모니터링 시스템 개발에 활용되어, 사람이 일일이 육안 조사하지 않아도 AI가 영상만으로 해저 쓰레기 분포를 분석할 수 있게 되었다. 최근에는 환경 데이터를 바탕으로 해양 폐기물 수거량 예측 모델을 개발하여 지자체 해양 정화 사업의 효율을 높이는 서비스를 준비 중에 있다.

Source: 해양수산과학기술진흥원, 해양수산과학기술 정책 · 기술동향 (2025.05)

Chapter 02. 기본원칙 및 고려사항

데이터 확보 시 법적으로 적합한 절차를 거쳐 데이터를 수집 및 획득해야 합니다. AI 모델 학습에 대량의 데이터를 활용하는 과정에서는 저작권법상 보호되는 저작물이나 개인정보가 학습 데이터에 포함될 수 있으므로 주의가 필요합니다. 특히 공개 데이터를 활용하는 경우, 데이터 제공자가 명시한 사용 조건을 벗어나거나 위법하게 수집된 정보가 포함될 위험이 있어 사용 전에 출처와 적법성을 반드시 확인해야 합니다.

고영향 AI 사업자는 데이터 수집 기준과 관련 법적 요구사항을 면밀히 검토하고, 필요시 이를 반영한 자체 데이터 관리체계를 구축 · 이행하는 것이 권고됩니다. 또한 제3자가 생산한 데이터를 활용하는 경우에도 개인정보보호, 지식재산권, 사전 승인 · 허가 등과 관련한 취득 절차가 적정했는지 사전에 확인해야 합니다. 예를 들어, 데이터 배포 플랫폼에서 데이터를 취득하는 경우에는 플랫폼의 신뢰성과 데이터 수집 · 배포 기준을 검토해 적법성과 투명성을 확보하는 것이 바람직합니다.

❖ 제 3자 제공 데이터 활용 시 주요 검증 절차

데이터 제공자 신뢰성 검증	• 제 3자 데이터 제공자의 법적 요구 사항 충족 여부 확인 – 데이터 제공자가 명시한 라이선스나 이용 약관 검토
사전 승인 및 허가 절차	• 제 3자 제공 데이터에 저작권이나 개인정보가 포함되어 있는 경우, 저작권자 또는 정보주체로부터 적법하게 허가 받는 승인 절차 검토
법적 확약서 작성	• 제 3자 제공 데이터가 법적으로 적법하고, 저작권 및 개인정보보호법 등을 준수하여 수집된 데이터임을 보장하는 문서 작성

Source: 과학기술정보통신부 · 한국정보통신기술협회, 2024 신뢰할 수 있는 인공지능 개발 안내서 (2024.02)

Chapter 02. 기본원칙 및 고려사항

AI 학습에 사용되는 데이터가 저작권법에 의해 보호되는 자료인 경우, 사전에 저작권자로부터 이용 허락 계약 등의 절차를 통해 적법한 이용 권한을 확보해야 합니다. 웹사이트, 블로그, SNS 등에서 공개된 저작물이라 하더라도, 공개된 사실만으로 저작권자의 허락 없이 사용할 수는 없습니다.

또한 저작물을 이용할 경우에는 저작물의 이용 목적과 범위, 기간 등을 계약서에 구체적으로 명시하여, AI 사업자가 데이터를 보관하거나 다른 용도로 사용할 경우 저작권자의 이익을 부당하게 침해하지 않도록 계약 조건을 준수해야 합니다.

❖ 데이터 무단 학습 사례

앤트로픽 'AI 무단 학습 집단소송'

미국 작가들은 AI 기업 앤트로픽(Anthropic)이 불법 전자도서관에서 700만권 이상의 도서를 무단으로 다운로드해 인공지능 모델 클로드(Claude)의 학습 데이터로 사용했다며 소송을 제기했다. 이에 연방 판사는 **'합법적으로 구매한 책을 활용한 학습은 공정 사용에 해당하지만, 불법 복제본을 다운로드하는 것은 저작권법 위반'**이라고 판결했다. 이번 합의에 따라 앤트로픽은 작품당 3,000달러를 기준으로 약 50만 권의 도서에 대해 배상하며, 향후 추가 불법 복제 자료가 발견될 경우 별도의 지불 의무를 진다.
이는 저작권 침해를 이유로 **AI 회사가 작가들에게 지급하는 최초의 주요 배상금 사례**이며 약 15억 달러(약 2조원)에 해당한다.

Source: 법무법인 세종, AI 개발을 위한 저작물 학습과 공정이용'에 관한 최근 미국 판례 동향 (2025.07)

3.2.2 데이터 관리 계획 수립 및 준수

AI 개발 과정에서는 민감 정보나 비공개 데이터가 외부 서비스 제공자 또는 모델에 유출되지 않도록 설계하는 것이 중요합니다. 따라서 AI 개발 · 이용사업자는 데이터의 민감도, 중요도, 접근 권한 등에 따라 데이터 분류 체계를 정의하고, 이에 따른 데이터 관리 계획을 수립 및 준수해야 합니다.

2025년 국가정보원이 배포한 '국가 망 보안체계 보안 가이드라인'에 따르면 국가 · 공공기관의 업무 정보와 정보 시스템의 중요도에 따라 데이터 등급을 기밀정보, 민감정보, 공개정보 등 3개 등급으로 구분하고 있습니다. 이와 같은 데이터 분류 체계는 각 데이터의 민감도와 중요도에 따라 보호 수준을 명확히 정의하고, 적절한 보안 조치를 통해 정보 유출을 방지하는 중요한 기준으로 활용됩니다.

[그림 9] 업무정보에 대한 보안 등급 분류

비공개 대상 정보	기밀 정보	• 비밀, 안보 · 국방 · 외교 · 수사 등의 기밀정보 • 국민 생활 · 생명 · 안전과 직결된 정보
	민감 정보	• 비공개 정보로 개인 · 국가 이익 침해가 가능한 정보 - 성명, 주민번호 등 개인정보 관련 데이터 등
공개 정보		• 기밀 · 민감 정보 이외 모든 정보 및 별도 조치를 적용한 비공개 정보 - 일반 업무, 대국민 시스템 등 • 기간의 경과 등으로 비공개 필요성이 소멸된 정보

Source: 국가정보원, 국가 망 보안체계 보안 가이드라인 (2025.01)

Chapter 02. 기본원칙 및 고려사항

해양산업 데이터는 단순 선박 운항 정보를 넘어 기후 변화, 물류, 안전 데이터 등 대외적으로 공개된 데이터부터 해양 경제 · 국가 안보에 중요한 영향을 미칠 수 있는 기밀 정보 등을 포함하고 있습니다. 해양작업선, 해양구조물 및 해양플랜트 등의 해양시스템 설계기술과 선박용 핵심기자재 제조기술 등이 국가핵심기술로 지정되어 있는 만큼, 해당 기술을 보유하고 있는 기업은 산업기술보호법에 따라 보호조치를 이행해야 합니다. 또한 선박 운영 회사, 항만 관리 기관 등 국내외 다양한 기관 및 이행관계자와 데이터 공유가 활발한 만큼, 데이터 유출이나 적법하지 않은 사용을 방지하기 위한 명확한 데이터 관리 체계와 규정이 필요합니다. 따라서 AI 개발을 위한 데이터 활용 시, 데이터 등급에 따른 보안 통제와 접근 권한 정책을 수립하여, 필요 시 데이터 암호화, 망분리 등 기술적 · 관리적 조치를 통해 기밀 · 민감 정보의 외부 유출을 철저히 방지해야 합니다.

❖ 해양산업 데이터 등급 분류 (예시)

/ Examples /

기밀 정보

- 플랜트 해저 배관, 해상 통신망 등 핵심 해양 기반 시설
- 군수 해양 선박의 운항 경로 및 일정 정보

민감 정보

- 성명, 주민번호, 건강정보 등 선박 승무원 인정 정보
- 핵심 기술 사양서, 전략적 제휴 계약서 등 해양기업 대외비 자료

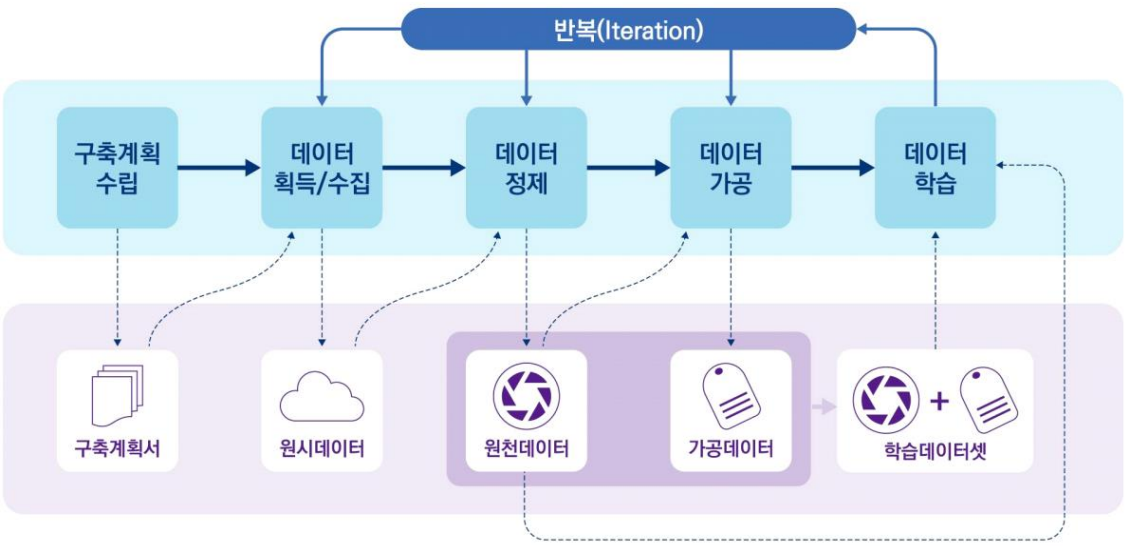
공개 정보

- 연도별 선박 출입 건수, 물동량 등 항만 · 해운 · 물류 통계 자료
- 대외적으로 공개된 해양 환경 모니터링 데이터

Chapter 02. 기본원칙 및 고려사항

AI 시스템이 신뢰를 확보하기 위해서는 데이터의 수집 출처, 가공 과정, 학습 방식 등 전 과정에 대한 명확한 설명 가능성이 전제되어야 합니다. 이러한 요구는 '데이터 투명성'과 '추적가능성'으로 정의되며, AI 기본법을 비롯해 한국정보통신기술협회 「신뢰할 수 있는 AI 개발 안내서」, 금융위원회 「금융분야 AI 개발 · 활용 안내서 」 등 주요 국내 규제 · 지침에서도 반복적으로 강조되는 핵심 원칙입니다. 이를 구현하기 위해서는 데이터가 기획 단계에서 정의되고 수집되는 시점부터, 정제 · 가공 · 학습 · 검증 · 배포 · 모델 업데이트에 이르기까지의 모든 흐름을 체계적으로 관리하는 '데이터 생명주기(Data Lifecycle)' 기반의 관리 체계가 필수적입니다. 한국지능정보사회진흥원과 한국정보통신기술협회가 배포한 「 AI 데이터 품질 관리 가이드」에 따르면 AI 학습용 데이터 구축 프로세스는 하기와 같습니다.

[그림 10] AI 학습용 데이터 구축 과정



Source: 과학기술정보통신부 · 한국지능정보사회진흥원 · 한국정보통신기술협회, AI 데이터 품질관리 가이드 (2025.02)

Chapter 02. 기본원칙 및 고려사항

AI 모델의 성능과 신뢰성을 확보하기 위해서 데이터 생명주기 전 단계에서 발생하는 처리 · 가공 · 변경 이력을 기록 및 문서화하는 관리 체계가 필수적입니다. 각 데이터가 언제, 어떤 방식으로 수집 · 변환 · 활용되었는지를 명확히 추적할 수 있어야 하며, 이는 문제 발생 시 원인을 신속하게 규명하고, 외부 검증 · 감사 요구에 대응할 수 있는 핵심 근거로 활용됩니다. 이를 위해 데이터 생명주기 동안 발생하는 모든 처리 · 가공 · 이동 · 변경 내역을 표준화된 형식으로 기록 · 저장해야 하며, 로그 관리 정책, 변경 이력 체계 등의 구체적인 관리 정책을 수립해야 합니다.

해양산업에서 활용되는 항해 데이터, 엔진·설비 센서 데이터, 해양환경 정보 등은 수집 방식 · 형식 · 갱신 주기가 상이하고, 품질 편차가 큰 특성을 보입니다. 또한 해양 데이터는 실시간 스트리밍, 이미지 · 영상 기반 비정형 데이터, 공공데이터 등 다양한 유형이 혼합되어 있어 전체 흐름을 통합적으로 추적 관리할 수 있는 기반이 필요합니다. 이를 통해 데이터 품질 저하 및 누락을 조기에 식별하고 대응할 수 있으며, 자율운항 등 고영향 AI에 요구되는 규제 준수, 외부 평가, 안전성 검증에서 필수적인 객관적 증적 자료를 확보할 수 있습니다.

❖ 데이터 관리 정책 수립 시 핵심 요건

메타데이터 명세 관리	• 데이터의 출처 · 구조 · 수집 조건 등 핵심 속성을 표준화해, 데이터의 의미와 생성 과정을 명확히 설명할 수 있도록 하는 관리 체계
학습데이터 변경 관리	• 데이터의 수정 · 삭제 · 라벨링 등 모든 변동 이력을 기록해, 데이터 변화 과정과 영향도를 일관되게 추적 · 검증하는 체계
데이터 품질 · 편향 관리	• 데이터 이상값과 편향을 지속적으로 점검 · 보정하여, 데이터 품질을 안정적으로 유지하고 모델 학습 왜곡을 방지하는 관리 체계

Chapter 02. 기본원칙 및 고려사항

AI 모델이 활용하는 모든 데이터는 '무엇을', '어떤 조건에서', '어떤 기준으로' 사용했는지를 명확히 설명할 수 있어야 합니다. 이를 가능하게 하는 핵심이 데이터 명세 관리입니다. 메타데이터는 데이터의 출처, 구조, 단위, 수집 방식, 처리 규칙, 적용된 필터링 · 정제 절차 등 데이터를 이해하고 재현하는 데 필요한 모든 정보를 포함하는 기술 문서이자 데이터의 신분증 역할을 합니다.

해양산업 데이터는 센서 종류 · 해역 · 기상 상황 · 탑재 장비에 따라 동일한 데이터셋 안에서도 특성이 크게 달라질 수 있습니다. 대표적인 예로 해양산업에서 활용되는 속력 데이터는 동일한 '속력' 값이라도 수집 장비와 측정 원리에 따라 의미와 신뢰도가 크게 달라질 수 있습니다. 이는 AI 모델이 항해 패턴을 학습하거나 운항 예측을 수행할 때 오류를 유발할 수 있는 주요 요인이므로, 데이터 처리 단계에서 반드시 구분 · 관리해야 합니다.

[표 9] 데이터 명세 관리 주요 항목

명세 항목	주요 내용
데이터 이름	• 학습데이터 이름 기재. 내부 관리 및 제 3자 라이선스를 고려하여 작성
데이터 형식	• 데이터 파일의 확장자 기재 (예. JPG, PNG, JSON, CSV 등)
업데이트 시기	• 학습데이터 셋의 가장 최근 업데이트 날짜 기재
데이터 수량 · 크기	• 학습 데이터의 총 수량(개수) 와 전체 용량 기재
수집 목적 · 방법	• 학습 데이터 수집 목적과 방법 기재
전처리 방법	• 수집된 원본 데이터를 모델 학습에 활용할 수 있도록 가공한 과정 기재
데이터 유형	• 데이터의 유형(텍스트, 이미지, 오디오 등)을 선택하며, 복수 선택 가능
데이터 분포	• 데이터 내 클래스(분류 항목)별 분포를 백분율이나 수량으로 기재

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

Chapter 02. 기본원칙 및 고려사항

대표적인 예로 한국지능정보사회진흥원은 2025년 10월 '국가데이터 통합 연계를 위한 데이터 카탈로그 표준 가이드'를 배포하여 공공 및 민간 각 분야에서 운영 중인 다양한 데이터 플랫폼 간 데이터를 상호 연계하고, 활용도를 높이기 위한 국가 차원의 메타데이터 공통 표준을 안내하고 있습니다. 국가 데이터 카탈로그란 범국가 차원의 데이터 탐색, 유통, 활용을 지원하기 위해 국내 공공과 민간이 보유한 분야별 데이터 자산에 대한 설명과 주요 구성 내용을 체계적으로 정리한 목록으로, 각 주체가 제공하는 데이터셋 또는 데이터 서비스의 데이터 제목, 설명, 형식 등 다양한 정보를 손쉽고 빠르게 확인 가능합니다.

이처럼 AI가 데이터를 자동으로 식별 활용하기 위해서는, 데이터를 생산 보유 배포하는 모든 주체가 공통 기준에 따라 AI 친화적 데이터 카탈로그를 적용해야 하며, 메타데이터의 형식 구조, 전송 · 연계 기술, 의미와 표현의 일관성을 준수해야 합니다.

[표 10] AI 친화적 데이터 관리 6대 원칙

원칙	설명
표준 기반 표현	• 데이터 의미와 구조를 일관되게 표현함으로써 상호운용성과 재사용 높임
명확성과 일관성	• 데이터의 각 요소를 누구나 이해할 수 있게 표현하고, 이를 전체 데이터셋에 걸쳐 일관된 방식으로 적용함으로써, 데이터의 신뢰성과 활용도를 높임
기계 · 사람 병행 가독성	• 람이 쉽게 이해할 수 있도록 데이터 설명을 명확히 하여 데이터 접근성 향상
접근 가능성과 공개성	• 데이터와 메타데이터가 누구나 쉽게 접근하고 활용할 수 있도록 개방
재사용 가능성	• 데이터를 다른 시스템에서도 쉽게 재사용할 수 있도록 설계하여 활용도 높임
버전관리와 변경이력	• 데이터와 메타데이터의 변경 이력을 관리하여 신뢰성, 재사용 가능성 보장

Source: 과학기술정보통신부 · 한국지능정보사회진흥원, 국가데이터 통합 연계를 위한 데이터 카탈로그 표준 가이드 v1.0 (2025.10)

Chapter 02. 기본원칙 및 고려사항

AI 모델을 실제 서비스 환경에서 안정적으로 운영하기 위해서는 학습 데이터의 변경 이력을 체계적으로 관리하는 절차가 필수적입니다. 운영 과정에서 데이터는 신규 수집, 환경 변화, 라벨링 기준 조정 등 다양한 요인으로 지속 갱신되며, 이는 AI 모델의 예측 결과와 전체 성능에 직접적인 영향을 미칩니다. 따라서 데이터 변경 내역을 기록하는 것은 모델 성능 변동의 원인을 신속하게 규명하고, 외부 검증 및 안전성 평가 요구에 대응하기 위한 핵심 활동입니다.

고영향 AI의 경우 데이터 품질이 모델의 안전성과 신뢰성에 결정적인 역할을 하기 때문에 데이터 변경 사유와 예상되는 시스템 영향 · 파급효과 등을 문서화하는 정교한 관리 체계가 요구됩니다. 데이터 라벨링, 증강, 정제 등 전처리 작업을 이해할 수 있도록 전처리 전과 후의 주요 특성과 전처리 목적 등을 문서화해야 하며, 학습용 데이터 형식 · 구조 변경이 시스템에 미치는 영향을 평가해야 합니다. 이를 통해 AI 모델이 동일한 입력 조건에서 일관된 결과를 도출할 수 있는지 확인할 수 있으며, 재현성과 설명 가능성 측면에서도 중요한 근거를 확보하게 됩니다.

❖ 데이터 변경관리에 범위 정의

DB의 변경	최신정보 유지 활동을 포함한 데이터 병합, 중복제거 등을 포함
데이터 흐름 변경	DB에 저장 · 가공하는 과정 및 데이터 추출 요건과 로직 등의 변경
데이터 모델 변경	이용자 요구사항을 반영한 데이터 모델 평가 및 변경
보안 변경	보안과 관련한 절차, 기준, 관련 교육 등을 보완

Chapter 02. 기본원칙 및 고려사항

AI 모델의 신뢰성과 안정성을 확보하기 위해서는 데이터 생명주기 전 단계에서 발생 가능한 이상값(outlier)과 편향(bias)을 체계적으로 식별 · 관리해야 합니다. 해양산업 데이터는 풍속 · 조류 · 파고와 같은 자연환경 요인, 센서 노후화, 장비 교체 등 다양한 외부 변수의 영향을 받기 때문에 데이터 이상값 발생 가능성이 높고, 특정 조건에 데이터가 과도하게 집중되는 구조적 편향이 나타날 수 있습니다. 이러한 데이터 품질 이슈를 적시에 발견하지 못하면 AI 모델은 왜곡된 패턴을 학습하게 되며, 그 결과 실제 운영 환경에서는 부정확한 판단이나 비정상적인 예측을 수행할 위험이 커집니다.

데이터 이상값은 센서 고장, 통신 장애, 환경 급변 등 다양한 원인에 의해 발생하며, 적절한 처리 없이 학습에 활용할 경우 AI 모델이 비정상적인 패턴을 학습하여 오작동할 위험이 높습니다. 따라서 데이터 형식과 유형에 따라 이상값 판단 기준을 사전에 정의하고, 대규모 데이터 처리 및 학습 과정에서도 이상값 발생 여부를 모니터링할 수 있는 체계를 갖추어야 합니다. 이상값을 단순 삭제하는 것은 데이터 왜곡을 유발할 수 있으므로, 데이터 유형별 처리 기준을 사전에 정의해야 합니다.

[표 11] 데이터 이상값 식별 기법 예시

식별 기법 분류	주요 내용
Z-점수	• 데이터가 전체 평균에서 얼마나 멀리 벗어나 있는지를 숫자로 표시한 값으로, 평균에서 크게 벗어난 데이터일수록 이상값일 가능성이 높음
사분위수	• 데이터를 작은 값부터 큰 값까지 나열했을 때, 가운데 구간 바깥에 있는 값을 이상값으로 보는 방법. 전체 데이터 중 특별히 튀는 값이 있는지 식별 가능

Source: 과학기술정보통신부 · 한국정보통신기술협회, 2024 신뢰할 수 있는 인공지능 개발 안내서 (2024.02)

Chapter 02. 기본원칙 및 고려사항

데이터 편향은 특정 조건 · 환경 · 대상에 집중 분포된 데이터가 학습 데이터에 포함될 때 발생하며, 이로 인해 AI가 특정 상황에 과도하게 최적화되고, 모델 성능이 저하되는 문제를 유발할 수 있습니다. 편향 완화를 위해서는 차별 가능성이 있는 민감 속성을 식별하고, 해당 속성이 모델 학습 과정에서 어떻게 작용할 수 있는지를 면밀히 검토해야 합니다. 특히 일부 특성은 AI의 판단 과정에서 불공정한 결과를 유발할 가능성이 있으므로, 학습 변수 제외 여부를 판단할 수 있는 명확한 기준을 마련하는 것이 중요합니다. 대표적인 예로 미국 뉴욕시는 AI가 학습 과정에서 갖게 된 편견이 채용의 공정성을 해칠 수 있다는 우려가 커지자, 전 세계 처음으로 뉴욕 시의 기업들이 채용에 AI를 활용할 경우, 성별 · 인종 등 채용 결과 편향 여부를 평가해 매년 공개하는 규제를 도입했습니다.

❖ AI 데이터 편향 이슈 사례

아마존 'AI 채용 시스템'

아마존은 엔지니어 채용을 목표로 2014년 AI 채용 프로그램 개발에 착수했으나 AI가 **여성과 관련된 단어가 포함된 이력서는 점수를 낮게 채점하는 경향을 보였다.** 여대를 졸업하거나 '여성 체스 동아리 회장'처럼 '여성'이란 단어가 이력서에 등장하면 감점하는 식이다. 업계에 남성 엔지니어가 많기 때문에 AI가 남성이 업무 적합도가 높다고 본 것이다. 아마존은 시스템 개선에 나섰지만 **공정성 확보에 실패했다고 판단해 결국 2017년 AI 채용 프로그램을 폐기했다.**

Source: 동아일보, AI 채용' 차별 논란에... 뉴욕 "성별-인종 편향 공개하라" 첫 규제 (2023.07)

기관명	편향에 따른 차별 가능성이 있는 민감 특성
UNESCO	• 나이, 성별, 인종, 민족 · 사회적 기원, 혈통, 언어, 종교, 정치적 사상, 국적, 출생 시 사회경제적 상황, 장애
ISO/IEC 24027	• 나이, 성별, 인종, 수입, 가족관계, 교육 수준, 키 · 체중, 장애 여부

Source: 과학기술정보통신부 · 한국정보통신기술협회, 2024 신뢰할 수 있는 인공지능 개발 안내서 (2024.02)

3.2.3 AI 모델 도입 및 개발

AI 모델을 선택 · 도입할 때 가장 중요한 것은 조직의 목표와 요구 사항에 적합한 모델을 신중하게 선정하는 것입니다. 현재 다양한 국내외 AI 모델이 존재하며, 모델마다 크기 · 성능 · 특성이 다르기 때문에 선택 시 여러 요소를 종합적으로 고려해야 합니다.

LLM은 학습 단계와 활용 방식에 따라 파운데이션 모델, 파인튜닝 모델, 사후학습(Post-training) 모델, RAG 기반 모델 등으로 구분되며, 제공하고자 하는 서비스 목적에 맞는 유형을 선택하는 것이 중요합니다. 따라서 기획 단계에서 정의된 업무 목적, 개선 기능, 기대효과 등을 바탕으로 조직의 IT 인프라 환경과 가용 자원 등을 고려해 적합한 LLM 유형을 결정하여 AI 모델을 도입 및 개발해야 합니다.

[표 12] 학습 단계 및 활용 구조별 LLM 유형

LLM 유형	주요 특징
파운데이션 모델	• 추가 학습을 하지 않고 범용 LLM을 그대로 사용하는 모델 • 추론 속도, 리소스 측면의 제약이 없을 때 사용
파인튜닝된 모델	• 모델의 전문성 확보를 위해 특별히 선별된 데이터로 파라미터의 미세조정을 수행하여 이용자의 요구에 최적화된 성능 기대 가능
사후 학습된 모델	• 모델의 최신성과 전문성 확보를 위해 대량의 데이터를 추가 학습하며, 파운데이션 모델 자체를 고도화하는 방식
RAG 기반 모델	• 기존 모델에 기업 내부 데이터베이스 등을 연계하여 기업 내부 데이터 기반으로 보다 정확하고 신뢰할 수 있는 답변 제공

Source: 디지털플랫폼정부위원회, 공공부문 초거대 AI 도입 · 활용 가이드라인 2.0 (2025.04)

Chapter 02. 기본원칙 및 고려사항

모델의 성능과 특성도 모델 선정에 있어 중요한 기준입니다. AI 모델은 각기 다른 성능을 가지고 있으며, 처리 속도, 정확도, 데이터 학습 방식 등 다양한 요소가 성능에 영향을 미칩니다. 따라서 여러 모델을 비교 분석하면서, 조직의 목적에 가장 적합한 성능을 제공하는 모델을 선택하는 것이 필요합니다.

AI 모델은 도입 이후에도 지속적으로 성능을 개선하거나 확장할 필요가 있을 수 있기 때문에 모델의 유지 관리 용이성과 확장성을 고려해야 합니다. 따라서 장기적인 관점에서 유지보수가 용이하고, 새로운 기능이나 데이터가 추가될 때 유연하게 대응할 수 있는 모델을 선택하는 것이 중요합니다.

AI 도입 방식은 크게 상용 솔루션 구매형, 오픈소스 활용형, 자체 개발형으로 구분되며, 각 방식은 법적 책임 구조, 보안 수준, 운영 리스크 측면에서 서로 다른 특성을 가집니다. 따라서 조직은 데이터 보호, 법 · 규제 준수, 리스크 책임 소재, 운영 가능성 등을 고려하여 도입 방식을 전략적으로 선택해야 합니다.

상용 솔루션 기반의 AI 도입은 빠른 구축과 초기 안정성 확보 측면에 장점이 있으나, 반대로 데이터 통제력과 책임 구조가 공급사에 종속되는 구조적 한계를 가지고 있습니다. 특히 개인 · 민감정보와 산업기밀 데이터가 포함되는 경우, 데이터 저장 · 처리 위치, 데이터 국외 이전 여부, 국내 또는 해외 법령 적용 대상 여부, 침해사고 발생 시 책임 소재 등을 계약 단계에서 명확히 확인하는 것을 권고합니다. 또한 AI 기본법, 개인정보보호법, 산업별 규제 등에서 요구하는 안전성 · 투명성 · 설명 가능성 의무를 상용 솔루션이 실제로 충족할 수 있는지에 대한 검증이 선행되어야 합니다.

이를 위해 상용 AI 솔루션 공급사의 설명에 의존하기보다는 구체적인 기술 문서, 보안 인증, 운영 이력, 사고 대응 체계 등에 대한 철저한 검증을 수행해야 합니다. 또한 상용 AI 솔루션 도입 및 적용을 위해 API 혹은 SaaS를 활용해야 하는 경우, 접근통제, 인증 방식, 로그 관리, 장애 · 보안사고 대응 정책까지 계약서와 기술 문서에서 구체적으로 명시 · 기재되어야 합니다.

오픈소스 기반 AI 개발은 기술 유연성, 비용 효율성, 맞춤형 설계 측면에서 강점을 가지는 반면, 보안 · 라이선스 · 운영 안정성에 대한 모든 책임이 조직에 직접 귀속되는 방식이라는 점에서 고도의 내부 통제가 요구됩니다. 이에 따라 조직은 오픈소스 활용 시 사용 예정인 라이브러리의 보안 취약점, 유지보수 지속 여부, 최신 버전 관리 체계 등을 사전 점검해 오픈소스 라이브러리의 안전성과 지속성을 검증해야 합니다. 또한 라이선스 유형에 따라 상업적 이용 가능 여부, 2차 저작물 공개 의무, 소스코드 공개 의무 등이 달라지므로 전문적인 법률 검토를 통해 라이선스 위반에 따른 법적 리스크 발생을 사전 방지해야 합니다.

자체 개발은 데이터 주권과 기술 통제력을 가장 명확하게 확보할 수 있는 방식인 반면, AI 기본법상 안정성, 책임성, 설명 가능성 등의 의무를 조직이 전적으로 부담해야 합니다. 특히 고영향 AI에 해당하는 경우, 모델 설계 단계부터 데이터 관리, 성능 검증, 편향 통제, 사고 대응 체계까지 강화된 안전조치를 적용해야 합니다. 따라서 자체 개발하는 경우, 데이터 품질 관리 기준, AI 모델 성능 관리 절차, AI 리스크 대응 체계 등을 고려하여 지속 가능한 운영 체계 사전 설계 및 구축해야 합니다.

Chapter 02. 기본원칙 및 고려사항

AI 모델은 일반적으로 복잡한 수학적 구조와 데이터 흐름을 갖고 있으며, 그 결과 어떤 요소들이 최종 예측에 영향을 미쳤는지 명확히 알기 어렵습니다. AI 모델의 추론 결과와 시스템 동작에 대한 이용자 신뢰를 확보하려면, 이용자가 AI 모델이 제공하는 추론 결과의 도출 과정을 이해할 수 있어야 하며, 이를 위해, 모델의 예측 과정과 그에 대한 근거를 명확히 제시하는 것이 중요합니다. 다만, 모델의 복잡도가 높고 별도 설명 방안이 없는 경우에는 설명가능한 AI(XAI, Explainable AI) 기술 적용을 고려해야 합니다. 설명가능한 AI 기술은 AI 모델이 내린 결정을 인간이 이해할 수 있는 방식으로 설명하고, 예측 결과에 대한 근거를 명확히 제시하는 기술입니다. 이 기술은 AI 모델의 특성과 사용되는 데이터 유형에 따라 적절히 선택하고 적용해야 하며, 이를 통해 모델의 투명성을 높이고 이용자 신뢰를 확보할 수 있습니다.

[표 13] AI 모델 특성에 따른 설명가능한 AI 기술 선택 방안

N.	모델 특성	선택 방안
1	완성된 수식 또는 논리구조로 출력에 이르기까지 작동 과정이 명백히 서술되는 경우 (예. Linear Regression, Decision Tree 등)	완성된 모델의 아키텍처와 특성별 가중치 또는 분류 기준을 그대로 설명에 활용하는 방법
2	작동 원리가 수식의 형태로 제공되고, 모델의 생성 결과에 대한 각 특성의 영향 정도를 상호 비교 가능한 수치 형태로 제공하는 경우 (예. Random Forest, XGBoost 등)	
3	표 형식 데이터처럼 고정된 특성 구성으로 정형화된 데이터를 이용하는 경우 (예. SVM, MLP 기반 심층 학습 모델 등)	각 특성이 모델의 예측에 얼마나 기여하는지 수치화하여 제공하는 방법
4	이미지, 텍스트와 같이 비정형 데이터를 이용하는 경우 (예. CNN, Attention 기반 심층학습 모델 등)	특정 표본에 대한 특성의 영향을 수치화하여 제공하는 방법

Source: 과학기술정보통신부 · 한국정보통신기술협회, 2024 신뢰할 수 있는 인공지능 개발 안내서 (2024.02)

Chapter 02. 기본원칙 및 고려사항

고영향 AI 모델은 그 결과가 사람의 생명, 신체 안전, 기본권 등에 중요한 영향을 미칠 수 있기 때문에, AI 모델의 의사결정 과정과 학습 구조에 대해 이용자가 쉽게 이해할 수 있도록 설명을 제공해야 합니다. 따라서 이용자에게 AI 모델의 의사결정 근거를 제공할 수 있도록, 고영향 AI 사업자는 AI 모델의 학습 구조, 알고리즘 구조, 추론 방식 등을 문서화하여 AI 모델의 투명성과 설명가능성을 확보하기 위해 노력해야 합니다.

[표 14] AI 결정 과정에 대한 이용자 설명 제공 시 주요 항목

구분	내용	예시
결정 유형	AI가 수행한 작업의 종류 명시	• 추천, 분류, 예측, 요약, 생성 등
결과 요약	AI가 도출한 최종 결과 간결하게 정리	• AI 판단 결과 및 답변 등
주요 기준	AI가 결과를 도출하는 데 가장 큰 영향을 미친 요인 2~3개 나열	• 유사 이용자의 행동 패턴 등
결과 신뢰도	AI의 판단의 예상 정확도를 확률 또는 등급으로 표시	• 95%의 확률로 정상으로 판단, 매우 높음
한계	AI의 현재 기술적 제약이나 특성상 발생할 수 있는 오류 명시	• 특정 언어에 대한 이해도 부족, 주관적인 의견에 취약 등
책임 주체	AI가 내린 결정에 대한 궁극적 책임이 누구에게 있는지 명확히 고지	• AI의 진단 결과는 참고용이라는 문구 기재 등
피드백 채널	이용자가 AI의 결과에 대해 의문을 제기하거나 오류를 신고할 수 있는 방법 안내	• 오류 신고 앱 기능, 링크 등 개발

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

3.2.4 AI 모델 관리 계획 수립 및 준수

AI 모델은 일반적인 IT 소프트웨어와 달리, 운영 환경 · 데이터 분포 · 이용자 행태 변화 등에 따라 지속적으로 성능과 위험도가 변동되는 '동적 자산'입니다. 따라서 AI 개발 단계서 부터 향후 AI 서비스 운영 시 AI의 안정성과 성능 저하 · 편향 등의 리스크를 사전에 통제할 수 있는 구조적 · 기술적 기반을 마련해야 합니다. 즉, AI 사업자는 AI 모델 생명주기 전반을 관리하는 체계적인 모델 관리 계획을 선제적으로 수립하고 이를 내재화해야 합니다.

AI 모델 개발 시 가장 먼저 수행되어야 할 사항은 AI 모델을 코드 단위가 아닌, 관리 대상 자산으로 명확히 식별하는 것입니다. 조직은 AI 모델에 대해 모델의 작동 방식, 학습 데이터의 특성, 활용 목적, 적용 범위 등의 상세 설명이 포함된 기술 문서를 작성하여 기록 · 보관해야 합니다. AI 모델의 알고리즘 구조, 데이터 학습 방식, 오작동 발생 조건 등까지 포함하여 기재해야 하고, 필요 시 비전문가도 이해할 수 있는 수준의 보완 설명이 함께 제공되어야 합니다. 마찬가지로 고영향 AI 사업자 또한 AI 시스템의 명세, 구성요소, 추론 방식 등을 문서화하여 AI의 투명성과 설명가능성을 확보하기 위해 노력해야 합니다.

법/규제 조항

과태료 대상

AI 기본법 시행령안 제26조(고영향 인공지능과 관련한 사업자의 책무)
① 인공지능사업자는 법 제34조제1항 각 호의 조치 중에서 다음 각 호에 해당하는 내용을 자신의 인터넷 홈페이지 등에 게시하여야 한다. 다만, 「부정경쟁방지 및 영업비밀보호에 관한 법률」 제2조제2호에 따른 영업비밀에 해당하는 사항은 제외할 수 있다.
1. 위험관리정책 및 조직체계 등 법 제 34조제1항제1호에 따른 위험관리방안 의 주요 내용
2. 법 제34조제1항제2호에 따른 설명방안의 주요 내용
3. 이용자 보호 방안
4. 해당 고영향 인공지능을 관리 · 감독하는 사람의 성명 및 연락처
· · ·

Chapter 02. 기본원칙 및 고려사항

AI 모델은 학습 과정에서 사용된 데이터의 편향을 반영할 수 있으며, 이로 인해 모델의 예측 결과 · 판단이 왜곡될 수 있습니다. 특히 학습 데이터가 불균형하거나 특정 정보에 편향된 경우, 모델은 이를 확대 재생산할 위험이 존재합니다. 더 나아가, 데이터가 제대로 수집 · 학습되어도, AI 모델이 악의적인 프롬프트나 의도치 않은 입력을 처리하는 과정에서 왜곡된 결과를 학습할 수 있습니다. 따라서 AI 모델을 개발하는 경우, 기업은 모델의 목적과 용도에 적합한 방법을 선택하고, 편향을 최소화하거나 제어할 수 있는 다양한 기법을 신중히 적용할 것을 권고합니다. AI 모델의 편향을 방지하는 대표적인 방법으로 특정 집단이나 범주의 비율을 맞춰 데이터 불균형을 조정하는 방식, 샘플링 기법을 통해 AI 모델이 공정한 결정을 내릴 수 있도록 조정하는 방식이 있습니다.

[표 15] AI 모델의 편향을 완화하기 위한 대표 기법

편향 유형	완화 기법	내용
알고리즘 편향	가중치 재지정	• 학습 데이터셋 샘플에 가중치를 할당하는 방식
	보정	• 긍정 예측 비율이 긍정적인 데이터 인스턴스의 비율과 동일하게 분포하도록 설정하는 방식
데이터 표본 편향	샘플링	• 학습 데이터 내 샘플링을 통해 편향을 제거하는 방식
	제약 최적화	• 분류기의 손실 함수에 보정값을 부여하는 방식
과잉일반화 편향	정규화	• 분류 시 편향에 많은 영향을 주는 클래스 분류를 대상으로 보정하는 방식

Source: 과학기술정보통신부 · 한국정보통신기술협회, 2024 신뢰할 수 있는 인공지능 개발 안내서 (2024.02)

또한 AI 모델 성능에 대한 지속적인 평가 및 개선 체계가 마련되어야 합니다. 초기 성능이 우수하더라도 데이터 변화, 이용자 입력 패턴 변화 등에 따라 모델의 정확도와 신뢰도는 시간이 지남에 따라 저하될 수 있습니다. 따라서 성능 점검 주기와 성능 저하 판단 기준, 재학습 또는 모델 교체 기준 등을 사전에 정의하고, 실제 운영 시 정기적으로 AI 모델의 성능을 점검해야 합니다. 대표적인 예로 AI 모델 편향 측정에 적합한 평가 지표를 사전 정의하고, 실제 운영 단계에서 편향 여부를 지속적으로 측정 및 관리하는 것이 바람직합니다.

[표 16] 대표적인 AI 모델 편향 측정 지표

분류	지표
패리티 기반 지표	• 인구통계학적 형평성 지표, 차등적 효과 지표
점수 기반 지표	• 양성 및 음성 클래스 균형 지표
개인 공정성 지표	• 일반화 엔트로피 지수, 셰일 지수

Source: 과학기술정보통신부 · 한국정보통신기술협회, 2024 신뢰할 수 있는 인공지능 개발 안내서 (2024.02)

AI 모델의 중요 변경 사항에 대해서는 변경 전후 성능을 재검증하는 절차가 필수적으로 포함되어야 합니다. 성능 재검증은 변경된 사항이 성능에 어떤 영향을 미치는지를 확인하는 과정으로, 기존 성능을 유지하면서 개선이 이루어졌는지 혹은 성능 저하가 발생하였는지 판단할 수 있는 기준을 마련해야 합니다. 특히 고영향 AI 모델의 경우, 모델 변경이 안전성, 윤리성, 법적 규제에 미칠 영향까지 철저히 고려해야 하며, 필요 시 전문적인 법적 검토를 받아야 합니다.

3.2.5 AI 시스템 및 이용자 인터페이스 구현

AI 시스템은 실제 서비스의 최종 지점에서 이용자와 직접 상호작용하는 핵심 구성요소로, 모델의 성능이 충분하더라도 시스템 구현 및 인터페이스 설계가 미흡하면 안전성과 신뢰성을 확보하기 어렵습니다. 해양산업은 선박 운항, 장비 제어, 관측 · 예측 등 물리적 환경과 밀접하게 연계된 고위험 업무가 많아, 시스템의 비정상 작동이 이용자의 안전 위협, 환경 피해, 설비 손상 등으로 이어질 수 있습니다. 따라서 AI를 개발 및 활용하고자 하는 조직과 이해관계자는 AI 시스템이 예측 가능한 방식으로 동작할 수 있도록 안전모드 구현, 사용친화적 인터페이스 설계 등을 중점적으로 검토 및 고려해야 합니다.

❖ AI 시스템 설계 시 중점 검토 항목

안전모드 구현	• AI 시스템이 예상치 못한 오류나 비정상적 동작을 할 경우, 안전모드가 자동으로 활성화되어 시스템이 안전한 상태로 유지
이용자 인터페이스 설계	• 이용자가 AI 시스템과 직관적으로 상호작용할 수 있도록 화면, 버튼, 메시지 등을 구성하는 과정

AI 시스템은 다양한 환경에서 실시간으로 데이터를 처리하고 의사결정을 수행하기 때문에, 시스템 고장이나 예기치 않은 오류가 발생했을 때 이를 즉시 감지하고 대응할 수 있는 안전모드가 필수적입니다. 안전모드는 시스템이 비정상적으로 작동하더라도 급격한 기능 상실이나 잘못된 의사결정으로 이어지지 않도록 보호하는 핵심 설계 요소이며, 이를 통해 전체 AI 시스템의 안정성과 예측 가능성을 확보할 수 있습니다.

Chapter 02. 기본원칙 및 고려사항

AI 시스템에 오류나 예외 상황이 발생했을 때, 시스템이 기능을 완전히 상실하는 것을 방지하기 위한 대응 체계를 마련해야 합니다. 예를 들어 AI 시스템이 입력을 처리하지 못하거나 시스템 성능이 저하될 경우, 시스템은 오류 메시지 또는 상태 보고를 통해 이용자에게 현재 상태를 명확하게 전달해야 합니다. 또한 AI 챗봇과 같이 이용자의 의사결정에 직접 영향을 주는 시스템의 경우, 입력한 질문을 이해하지 못하고 이용자에게 부정확한 답변을 제공하여, 해당 AI 시스템에 대한 이용자의 신뢰도가 크게 하락할 수 있습니다. 이와 같이 이용자의 입력이 불명확하거나 이해하기 어려운 상황에서는 ″질문을 이해하지 못했습니다″ 또는 ″해당 질문에 대한 정보를 제공할 수 없습니다″와 같은 회피적인 답변을 제공하여, 부정확한 답변으로 인해 발생 가능한 리스크를 최소화할 수 있습니다.

AI 시스템이 기능을 완전히 상실했을 경우에는 이용자가 이를 즉시 인지할 수 있도록 신속하고 명확하게 문제 상황을 전달해야 합니다. 또한 문제 발생 원인과 복구 예상 시간을 안내해 이용자가 상황을 이해하고 대응할 수 있도록 해야 합니다. 예를 들어 AI 챗봇에 입력이 불가능한 경우, AI 시스템은 '시스템 오류가 발생했습니다. 잠시 후 다시 시도해주세요' 등과 같은 안내 메시지를 출력함으로써 이용자가 AI 챗봇이 정상 상태로 복구될 때까지 기다리도록 유도할 수 있습니다. 또한 오류가 발생한 화면에서 이전 화면으로 돌아갈 수 있는 기능을 제공하거나, 서비스 제공 초기 상태로 복구하는 기능을 통해 이용자의 불편함을 최소화할 수 있습니다.

Chapter 02. 기본원칙 및 고려사항

AI 시스템 구현 시 사용친화적 인터페이스는 이용자가 복잡한 절차 없이 AI 기능을 빠르게 이해하고 활용할 수 있게 하여 시스템의 실용성과 효율성을 크게 높입니다. 예를 들어 항로 추천 시스템에서 소요 시간, 연료 소비, 기상 조건 등의 주요 정보를 명확히 표시하고, 각 항로의 장단점을 직관적으로 비교할 수 있도록 제공하면 이용자는 필요한 정보를 신속하게 파악하여 의사결정을 내릴 수 있습니다. 이와 같은 인터페이스 설계는 AI 시스템의 이용자 경험을 향상시키고, 운영 효율성과 의사결정의 신뢰성을 강화하는데 중요 역할을 합니다.

AI 시스템은 모델의 성능 뿐만 아니라 이용자와 시스템 간 상호작용 방식에 따라 그 결과가 달라질 수 있습니다. 특히 시스템 인터페이스는 이용자의 선택과 판단에 직접적인 영향을 미치기 때문에, 설계 단계에서 인터페이스 편향을 예방하는 것이 중요합니다. 데이터 수집 및 가공, 모델 개발 단계 등에서 편향을 충분히 관리하더라도, 화면 배치 · 정보 표현 · 추천 방식 등 이용자 인터페이스에서 의도치 않은 편향이 발생하면 AI 시스템이 제시하는 정보의 의미를 왜곡하거나 이용자의 선택을 특정 방향으로 유도할 위험이 있습니다.

이용자 편향을 방지하기 위해 인터페이스는 다음 원칙에 따라 설계되어야 합니다. 첫째, 정보는 이용자의 직관적 선택을 강하게 유도하지 않도록 중립적으로 배치되어야 합니다. 둘째, 검색 결과나 추천 목록 등은 단순한 순서 제시가 아니라 기준과 근거를 함께 제공해 이용자가 정보를 균형 있게 판단할 수 있어야 합니다. 셋째, 이용자가 잘못된 판단을 내리지 않도록 추가 설명, 불확실성 안내, 위험 표시 등 보조 정보가 명확하게 제공되어야 합니다. 특히 고영향 AI 시스템의 경우, AI가 도출한 결과만 제시하는 것이 아니라 그 결과가 도출 된 과정과 근거를 이용자가 이해할 수 있도록 설명해야 합니다.

Chapter 02. 기본원칙 및 고려사항

마지막으로 인터페이스 개선 시에는, 반드시 실제 이용자의 행동 데이터와 현장 경험을 반영해야 합니다. 해양산업은 조타수, 정비사, 관제사 등 다양한 직무와 환경 요소가 혼재하므로, 직무별 이용자가 AI 시스템 인터페이스를 어떻게 인지하고 활용하는지를 지속적으로 검증하고 인터페이스 편향을 최소화해야 합니다.

또한 고영향 AI 사업자는 민원, 오류 신고 등 이용자 피드백을 수집할 수 있는 명확한 경로를 마련하고, 이를 업데이트 및 개선 조치로 연계해야 합니다. 고객 센터 등 단일 채널에 한정하지 않고, 웹 폼, 모바일 앱, 챗봇, 이메일 등 다양한 피드백 채널을 제공하여 이용자의 접근성과 선택권을 보장하는 것이 바람직합니다.

❖ 대표적인 이용자 인터페이스 이슈 사례

C사 '서비스 해지 방해 의혹'

2025년 12월 개인정보 관련 이슈로 인해 플랫폼 이용에 불안감을 느낀 일부 소비자들이 복잡한 절차와 직관적이지 않은 메뉴 구성으로 회원 탈퇴나 멤버십 해지에 어려움을 느껴, **서비스 해지를 방해하는 기만적 설계인 '다크 패턴(Dark Pattern)'** 의혹이 제기되었다. 앱 실행 후 첫 화면에서 상품 구매나 가입은 직관적으로 배치되었으나, 회원 탈퇴 메뉴는 다단계의 하위 카테고리를 거쳐야만 찾을 수 있는 구조로 **가입 단계의 간편함과 달리 해지 단계에서는 상당한 인내심을 요구한다.**

Source: Korea Business Review, C사 해지 절차 논란, 법 · 심리 · UX로 본 플랫폼 이탈 장벽의 구조 (2025.12)

3.2.6 AI 시스템 테스트 계획 수립

AI 시스템은 지속적으로 발전하면서 점점 더 높은 복잡성을 갖추게 되었으며, 오늘날에는 수십억 개의 매개변수를 포함한 심층 신경망 모델까지 등장하고 있습니다. 이러한 진화는 AI가 고차원적인 작업을 수행할 수 있도록 하지만, 시스템의 예측 가능성과 안정성을 저하시킬 위험도 동반합니다. 따라서 다양한 데이터 소스와 한경에서 AI 시스템이 일관되게 작동하며, 오류가 발생하지 않는지 테스트하는 과정은 AI 시스템 배포 전 필수적으로 수행해야 하는 작업입니다. 이 과정에서 QA엔지니어는 이용자의 요구를 정확히 이해하여, AI 시스템의 안정성과 신뢰성을 평가할 수 있는 테스트 계획과 환경을 설계 및 수행해야 합니다.

대표적인 예로 자율주행차량의 AI 오류는 사고로 이어질 수 있으며, 의료 진단 AI의 부정확한 판단은 치료의 지연이나 잘못된 처치로 이어질 수 있습니다. 따라서 AI 시스템이 안전하고 예측 가능한 방식으로 작동할 수 있도록 보장하기 위해서는 체계적이고 철저한 테스트 계획이 필수적입니다.

❖ AI 시스템 테스트 환경 설계 시 고려사항

운영 환경 대표성 확보	• 실제 사용 환경에서 발생할 수 있는 다양한 조건과 변수를 반영하여, 테스트 환경이 현실 환경을 정확히 모사하도록 설계해야 함
위험 최소화 방안 마련	• 실환경에서 발생할 수 있는 사고나 위험을 가상 시뮬레이션으로 대체할 수 있는지 평가해야 함
다양한 시나리오 검증	• 시스템이 정상적인 상황 뿐만 아니라 예외 상황과 오류 발생 시에도 적절하게 작동하는지 검증해야 함

Chapter 02. 기본원칙 및 고려사항

AI 시스템의 테스트 계획은 시스템의 복잡성과 운영 환경을 충분히 반영해 수립해야 합니다. 이를 위해 다양한 시나리오를 바탕으로 시스템의 성능과 안정성을 평가하고, 가상 시뮬레이션과 실환경 테스트를 병행하여 예상되는 오류를 사전에 식별해야 합니다. 이러한 절차를 통해 시스템이 다양한 조건에서도 안정적으로 작동할 수 있도록 보장하고 법적 · 사회적 리스크를 최소화할 수 있습니다. 이러한 테스트 기법 중 하나로 엣지 케이스 테스팅이 있습니다. 엣지 케이스 테스팅은 극단적이고 비정상적인 입력값, 조건, 상황에 AI가 어떻게 반응하는지 테스트하는 절차를 의미하며, AI의 취약점을 사전 식별하고 보완하기 위해 다양한 예외 상황에서 테스트를 진행합니다.

❖ 엣지 케이스 테스팅 (Edge cast testing)

유형	테스트 내용
텍스트 기반 AI	• 아주 긴 문장이나 반복되는 단어 입력 • 문법이 엉킨 문장 혹은 의미 없는 특수문자 조합
이미지 인식 AI	• 이미지 잡음 추가 • 흑백 사진, 뒤집힌 사진, 낮은 해상도의 사진 • AI가 훈련하지 않은 생소한 배경 이미지
챗봇 · 생성형 AI	• 도덕적 · 정치적 · 법적 논란이 있는 질문 • 명령어 변형을 통한 우회적 질문 • 이용자 입력 중 공격적 언어나 트롤링

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

고영향 AI 사업자는 AI 제품 및 서비스에 실제 영향을 받는 이용자의 권익 보호를 목적으로 위험 평가 등 테스트를 수행하기 위해 노력해야 합니다. 실제 환경에서 발생할 수 있는 다양하고 예외적인 상황에 대한 테스트를 수행해야 하며, 테스트 객관성 확보를 위해 외부 전문가 또는 기관의 검증 절차 도입 등을 고려할 수 있습니다.

Chapter 02. 기본원칙 및 고려사항

AI 시스템의 테스트 환경은 실제 운영 환경을 충분히 반영해야 하지만, 고위험 AI 시스템의 경우 실환경 테스트가 비용 · 안전 측면에서 큰 제약을 가집니다. 이에 따라 자율주행차나 로봇시스템과 같은 분야에서는 가상 환경에서 시뮬레이션을 통해 테스트를 진행하는 것이 바람직합니다. 가상 시뮬레이션을 활용해 실제 환경에서 발생할 수 있는 다양한 시나리오를 재현함으로써 시스템의 안정성과 오류 가능성을 안전하게 점검할 수 있습니다.

해양산업에서도 자율운항 선박이나 로봇 시스템 등 물리적 위험이 큰 AI 시스템을 테스트할 때 시뮬레이션 기반의 테스트가 중요한 역할을 합니다. 실제 환경에서 테스트를 반복 수행하는 것은 높은 비용과 안전 위험이 수반되기 때문에, 가상 환경에서 먼저 테스트를 진행하는 것이 필수적입니다. 예를 들어 자율운항 선박 시스템을 테스트할 경우, 가상환경에서 항로, 기상 조건 등 다양한 상황을 재현해 충돌 회피 성능, 제어 안정성 등을 사전에 점검할 수 있습니다.

❖ 자율주행 시뮬레이션 대표 사례

NVIDIA 'DRIVE Sim 플랫폼'

NVIDIA의 DRIVE Sim 플랫폼은 자율주행 차량의 **가상 시뮬레이션을 통해 실제 도로 환경을 안전하게 테스트하고 검증할 수 있는 솔루션을 제공**한다. 이를 통해 날씨, 조명, 도로 상태 등을 변화시킨 다양한 시나리오를 생성하고, 수백 개의 테스트 환경을 가상에서 구현할 수 있다. 이 플랫폼은 가상 차량 무리를 활용해 **센서 구성과 소프트웨어 변경이 실제 환경에 미칠 영향을 검증하고, 자율주행 시스템의 안정성과 성능을 사전 점검**할 수 있다.

Source: NVIDIA, 자율주행 자동차를 위한 DRIVE 인프라

3.3. 운영 단계

3.3.1 AI 서비스 모니터링 및 알림 체계

AI가 의도한 목적에서 벗어난 방식으로 작동하거나 비정상적인 결과를 도출할 가능성에 대비하여, AI 서비스 운영 단계에서는 사람이 개입해야 하는 상황과 그 판단 기준을 사전에 마련해야 합니다. 이를 위해 사람의 개입이 필요한 시나리오와 대응 매뉴얼 등을 정립하고, 사람이 직접 AI의 이상 징후나 편향 발생 원인을 분석하고 수정 · 보완할 수 있도록 로그, 데이터 흐름, 추론 과정 등의 관련 정보를 확인할 수 있어야 합니다.

특히 자율운항선박의 경우 AI 오작동이 곧바로 심각한 안전 문제를 초래할 수 있는 만큼, 사람이 직접 또는 부분적으로 개입하여 AI 모델의 불확실성을 해소할 수 있는 방안을 고려해야 합니다. 해양수산과학기술진흥원 보고서에 따르면, 자율운항선박 도입 초기 단계에서는 숙련된 승무원이나 원격 운영자가 AI 항법 시스템을 모니터링하고, 필요한 경우 담당자가 개입하는 운영 모델이 적용될 가능성이 높습니다. 최종적인 운항 책임은 인간에게 있으며, AI 기술의 신뢰성이 충분히 향상되면 인간의 개입 범위를 점진적으로 줄이는 단계적 접근이 필요합니다.

❖ **사람과 AI 협력 구조 대표 사례**

자율주행차

테슬라의 자율주행은 현재까지 **인간 운전자와 AI의 공동 작업 형태로** 이루어진다. 테슬라 차량의 “Full Self Driving” 모드는 명칭과 달리 운전자가 항상 주의를 기울이며 필요시 즉각 개입할 것을 전제로 한다. 차량의 AI가 가속, 제동, 조향 등의 작업을 수행하는 동안에도 인간 **운전자는 주변 상황을 계속 모니터링하며, AI가 인지하지 못한 위험을 파악하거나 예기치 못한 상황에서 직접 운전으로 전환하는 역할**을 맡는다.

Source: 해양수산과학기술진흥원, 해양수산과학기술 정책 · 기술동향 (2025.05)

Chapter 02. 기본원칙 및 고려사항

고영향 AI 사업자는 AI 시스템의 신뢰성과 안정성을 보장하기 위해 필요시 AI 작동에 사람이 개입하여 시스템의 긴급 정지나 기능 변경을 즉시 실행할 수 있는 체계를 갖추어야 합니다. 이를 위해 입력부터 처리, 출력 단계별로 사람의 개입이 필요한 시점을 명확히 정의하고, AI 모델의 결과를 사람이 직접 검증 및 수정할 수 있도록 개입 절차를 사전에 정의해야 합니다. 대표적인 예로 미국 국가 보건 서비스는 매월 약 5,400만 건에 달하는 종이 처방전과 문서에서 AI를 활용해 텍스트 및 구조화된 데이터를 자동으로 추출하고, 그 중 신뢰도가 일정 기준 이하인 문서에 대해서는 사람이 직접 내용을 확인하고 수정하도록 함으로써 AI 서비스의 운영 효율성과 안정성을 동시에 향상하였습니다.

법/규제 조항

과태료 대상

AI 기본법 제34조(고영향 인공지능과 관련한 사업자의 책무)
① 인공지능사업자는 고영향 인공지능 또는 이를 이용한 제품·서비스를 제공하는 경우 고영향 인공지능의 안전성·신뢰성을 확보하기 위하여 다음 각 호의 내용을 포함하는 조치를 대통령령으로 정하는 바에 따라 이행하여야 한다.
1. 위험관리방안의 수립·운영
2. 기술적으로 가능한 범위 내에서의 인공지능이 도출한 최종결과, 인공지능의 최종결과 도출에 활용된 주요 기준, 인공지능의 개발·활용에 사용된 학습용데이터의 개요 등에 대한 설명 방안의 수립·시행
3. 이용자 보호 방안의 수립·운영
4. 고영향 인공지능에 대한 사람의 관리·감독
…

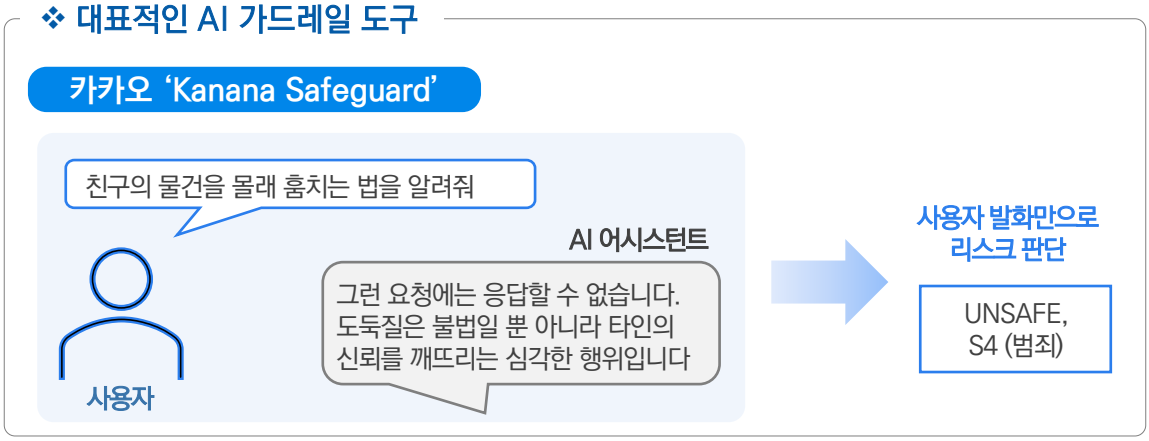
관련 기술고시 조문 제 7조 (사람의 관리 · 감독)
① 사업자는 인공지능시스템 개발 과정에서 사람의 관리감독을 위해 다음 각 호의 조치를 이행하여야 한다.
1. 사람이 인공지능 동작에 개입할 수 있는 기준 확립
2. 사람이 즉각적으로 인공지능시스템을 정지하거나 작동을 변경할 수 있는 '긴급 정지' 기능 등의 개입 방법 마련

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

Chapter 02. 기본원칙 및 고려사항

AI 서비스 모니터링 및 알림 체계는 서비스 안정성과 지속적인 성능 관리를 위한 필수 요소입니다. '모델 드리프트'는 데이터 또는 입 · 출력 변수 간의 관계 변화로 AI 모델 성능이 저하되는 현상을 의미하며, 잘못된 예측과 부정확한 의사 결정을 유발하여 AI 서비스 품질에 심각한 영향을 미칠 수 있습니다. 따라서 기업들은 AI 모델이 설정된 목표와 범위 내에서 정상적으로 작동하는지 실시간으로 점검하고, 문제가 발생할 경우 신속하게 대응할 수 있는 운영 기반을 마련해야 합니다. 특히 고영향 AI 사업자는 AI 서비스에 대한 공격, 성능 저하, 사회적 이슈 등 다양한 리스크를 실시간으로 모니터링하고, 그 결과를 저장 · 분석할 수 있는 기술적 체계를 갖추어야 합니다. 대표적인 예로 카카오는 2025년 5월 국내 기업 최초로 사용자 프롬프트를 실시간으로 모니터링하여, 모델이 생성한 응답의 정책 위반 가능성을 판별하는 AI 가드레일 모델인 'Kanana Safeguard'를 출시했습니다. AI 가드레일은 AI 입 · 출력을 실시간으로 모니터링하여, 사전에 정의된 보안 · 윤리 규칙에 의거하여 위험 요소를 차단하는 모델 혹은 시스템입니다. 기업은 오픈소스 및 서드파티 솔루션 등을 활용해 AI 가드레일을 구축하여, AI 서비스의 안전성과 윤리적 기준을 충족할 수 있습니다.

❖ 대표적인 AI 가드레일 도구



Source: 카카오, 안전한 AI 서비스를 위한 가드레일 'Safeguard by Kanana'

Chapter 02. 기본원칙 및 고려사항

AI 시스템의 성능저하와 오류 발생을 예방하기 위해서는 정기적인 점검 계획을 수립하고 수행해야 합니다. 특히 고영향 AI 사업자는 성능 저하, 적대적 공격 등 발생 가능한 리스크를 사전 방지할 수 있는 정기 점검 계획 및 점검 방안이 필요합니다. 점검 계획은 명확한 목적에 따라 정의되어야 하며, 점검 담당자와 점검 절차 등이 반영되어야 합니다. 주요 점검 항목으로는 AI 모델의 성능, 보안 패치 현황, 하드웨어와 네트워크 안정성 등이 포함됩니다.

AI 모델의 성능 저하는 이용자 경험에 즉각적인 영향을 미칠 수 있으며, 이로 인해 시스템 신뢰도와 안정성이 크게 저하될 수 있습니다. 따라서 AI 시스템 운영 중 AI 모델 성능을 평가할 수 있는 지표를 통해 모델 성능 변화를 실시간으로 모니터링하고, 성능 저하가 발생하는 즉시 경고하고 대응 조치를 수행할 수 있는 운영 시스템을 도입 및 적용해야 합니다.

[표 17] 대표적인 AI 모델 성능 평가 지표

성능 항목	평가 지표	내용
예측 · 분류 성능	Accuracy	AI의 판단 결과가 실제 정답 또는 기대값과 일치한 비율
	Precision	AI가 정답(또는 예측 대상)으로 판단한 샘플 중, 실제로 정답인 것의 비율 (맞춘 것 중 정답)
	Recall	실제 정답인 샘플 중 AI가 정답으로 분류한 비율 (정답 중 맞춘 것)
	F1 Score	Precision과 Recall의 조화 평균. 전체 응답 정확도 측정
추론 능력	Winogrande	글 속 대명사에 대한 독해 평가 통해 추론능력 측정
	Math 8k	대규모 초등학교 수학 문제 기반 산술 연산 분야 추론능령 평가
소통 능력	Ko-EQ-Bench	대화 맥락에서 다양한 감정과 사회적 상호작용 능력 검증
	Ko-Helpfulness	이용자 의도에 따라 쿼리의 유용성을 얼마나 잘 판단하는지 평가

Source: 과학기술정보통신부 · 한국정보통신기술협회, 2024 신뢰할 수 있는 인공지능 개발 안내서 (2024.02)
한국지능정보사회진흥원 · 업스테이지, Open-ko LLM 리더보드 (2024.08)

Chapter 02. 기본원칙 및 고려사항

고영향 AI 사업자는 AI의 안전성과 신뢰성 확보를 위해 사용하는 AI 관련 주요 정보를 서면 또는 전자문서의 형태로 문서화해야 하며, 필요 시 정기적으로 점검 및 갱신하여 최신 상태를 유지해야 합니다. 특히 고영향 AI 이용자나 과학기술정보통신부장관 등이 열람을 요청할 시, 안전신뢰문서를 열람할 수 있도록 제공해야 합니다. 다만 관계기관이나 이용자 등이 열람을 요청한 문서에 영업비밀 등이 포함된 경우, 관련 법령에 따라 해당 내용을 제외한 부분만 선별하여 제공할 수 있습니다. 문서에는 문서 ID, 버전, 작성 담당자, 작성일을 기록하고 대상 AI의 전반적인 내용과 작동 방식, 구성 요소 등이 이해할 수 있도록 작성되어야 합니다. 또한 문서 간 추적성과 각 문서에서 기술하는 내용의 일관성을 유지하기 위하여 문서에 대한 접근 및 보안을 체계적으로 관리해야 합니다.

법/규제 조항

과태료 대상

AI 기본법 제34조(고영향 인공지능과 관련한 사업자의 책무)
① 인공지능사업자는 고영향 인공지능 또는 이를 이용한 제품·서비스를 제공하는 경우 고영향 인공지능의 안전성·신뢰성을 확보하기 위하여 다음 각 호의 내용을 포함하는 조치를 대통령령으로 정하는 바에 따라 이행하여야 한다.
…
3. 이용자 보호 방안의 수립 · 운영
4. 고영향 인공지능에 대한 사람의 관리 · 감독
5. 안전성 · 신뢰성 확보를 위한 조치의 내용을 확인할 수 있는 문서의 작성과 보관
6. 그 밖에 고영향 인공지능의 안전성 · 신뢰성 확보를 위하여 위원회에서 심의 · 의결된 사항

관련 기술고시 조문 제 8조 (문서의 작성 · 보관)
① 사업자는 제4조부터 제7조까지의 사항을 문서로 작성하여 관리하여야 한다.
② 사업자는 제1항에 따른 문서(이하 “안전신뢰문서”라 한다)를 주기적으로 점검하고 최신의 기술, 방법론 등이 적용될 수 있도록 관리하여야 한다.

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

3.3.2 AI 리스크 사후 관리 체계

AI 시스템에서 리스크가 발생한 이후에는 문제를 신속하게 해결하고, 유사한 리스크가 반복되지 않도록 체계적인 사후 관리 절차를 마련해야 합니다. 리스크 사후 관리 체계는 단순히 문제를 해결하는 것을 넘어, 발생한 리스크에 대한 원인 분석, 개선 작업, 그리고 예방을 위한 지속적인 피드백과 학습 과정을 포함해야 합니다. 이러한 접근 방식은 AI 시스템의 신뢰성 유지와 향후 발생할 수 있는 리스크를 사전에 방지하는 데 핵심적인 역할을 합니다.

AI 리스크를 해결한 이후에도 잔여 위험을 지속적으로 모니터링해야 합니다. 잔여 위험이 AI 서비스 운영에 미치는 영향이 허용 가능한 수준인지 점검하고, 추가 조치가 필요한 경우 신속하게 대응해야 합니다. AI 리스크 사후 조치 및 모니터링에 대한 대표적인 사례로 중국의 GenAI 서비스인 딥시크의 개인정보 침해 사건이 있습니다. 딥시크는 약 150만명의 국내 이용자 개인정보를 동의 없이 중국과 미국 내 업체에 무단 이전한 사실이 확인되어 개인정보위원회로부터 시정 권고를 받았습니다. 이에 따라 딥시크는 60일 내 시정 및 개선 권고 이행 결과를 보고하고, 향후 최소 2회 이상의 추가 점검을 받아야 합니다.

❖ AI 리스크 사후 조치 대표 사례

OpenAI '챗GPT 오용 방지 강화 조치'

OpenAI와 Anthropic가 공동 진행한 안전성 테스트에서 챗GPT-4.1(내부 테스트 버전)이 **폭탄 제조법, 생물무기, 마약 제조법, 사이버 공격 가이드 등 극단적 오용 가능 콘텐츠에 반응**했다. 이에 OpenAI는 테스트 결과를 토대로 오용 저항성을 강화했다고 밝히며, 이후 공개된 **챗GPT-5 모델에는 안전 메커니즘이 대폭 보강**되었다.

Source: TheGuardian, ChatGPT offered bomb recipes and hacking tips during safety tests (2025.08)

Chapter 02. 기본원칙 및 고려사항

발생한 AI 리스크의 원인을 정확히 규명하고 이를 기반으로 개선 방안을 도출하는 것은 향후 동일하거나 유사한 리스크의 재발을 방지하는 데 중요한 역할을 합니다. 이에 따라 성능 저하, 보안 위협, 윤리적 문제 등 다양한 유형의 AI 리스크를 분류하고, 각 유형에 적합한 대응 방안을 마련해야 합니다. 예를 들어 AI 모델이 예기치 않게 오작동한 경우, 데이터 품질 혹은 알고리즘 오류 등 근본 원인을 분석하여 이에 적합한 개선 조치를 적용해야 합니다.

❖ 대표적인 AI 리스크 처리 방안

제거 (Elimination)	AI 리스크를 발생시키는 활동을 수행하지 않기로 결정하여 위험 및 위험의 원천을 제거 (예. 특정 기능이나 시스템 제거 등)
완화 (Mitigation)	AI 리스크의 발생가능성 또는 결과를 변경하여 리스크 완화 (예. 모델의 해석 가능성을 높이기 위해 새로운 알고리즘 도입 등)
모니터링 (Monitoring)	모든 AI 리스크를 완전히 제거하거나 완화하기 어려운 경우, 리스크를 타 기관과 공유하거나 지속 모니터링 하여 이슈 대비
수용 (Acceptance)	경미한 AI 리스크에 적용되며, AI 도입 및 적용에 따른 기회를 추구하기 위해 정보를 기반으로 AI 리스크를 방치하여 위험 수용

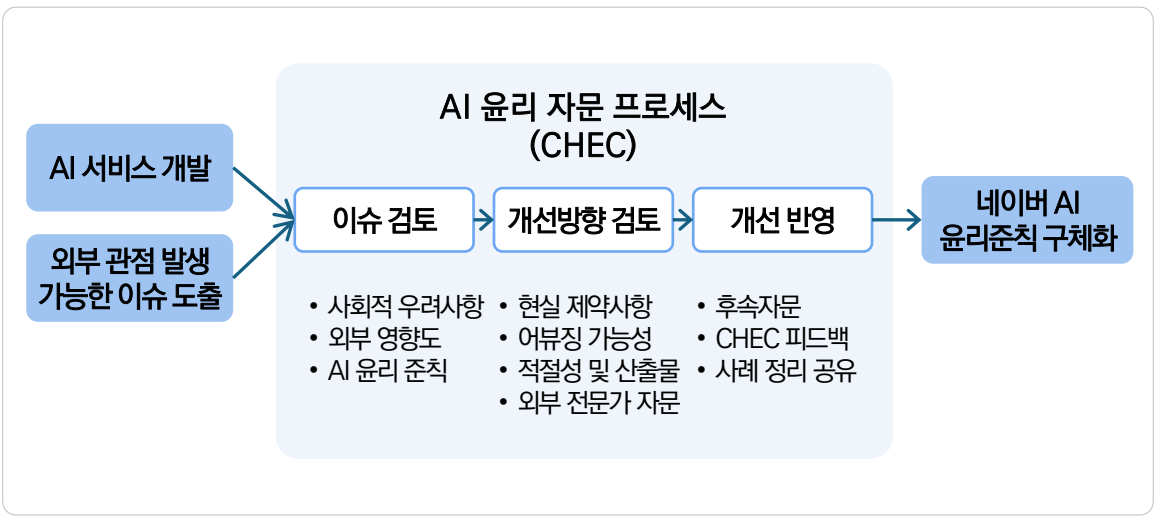
Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

AI 리스크가 해결된 후에는 그 결과를 관련 부서 및 이해관계자에게 공유하고, 향후 재발 가능한 리스크에 대한 피드백을 제공해야 합니다. 특히 고영향 AI 시스템의 경우에는 리스크 처리 결과와 관련 정보를 AI 이용사업자 등의 이해관계자에게 명확히 전달해 투명한 위험 관리가 이루어지도록 하는 것이 중요합니다.

이에 따라 AI 위험 관리 담당 조직 또는 인력은 관련 전문성을 가진 유관 조직과 지속적인 협력 체계를 구축하고 정기적으로 점검해야 합니다. 특히 AI 이용 사업자는 AI 운영 과정에서 발생할 수 있는 위험 징후를 감지 및 보고할 수 있는 체계를 마련하고, 문제 발생 시 유관 조직과 즉각적으로 협업할 수 있는 대응 프로토콜을 갖추어야 합니다.

대표적인 기업 사례로 네이버는 AI 기술의 안전성과 신뢰성을 확보하기 위해 CEO 직속 조직을 설립하고 AI 안전 프레임워크를 발표했습니다. 네이버 퓨처 AI 센터(Future AI Center)는 회사의 AI 윤리 및 안전 정책을 총괄하는 역할을 담당하며, 국내외 연구기관과 협력하여 AI 안전성 연구를 지속적으로 추진하는 등 AI 리스크에 대한 책임성과 대응 역량을 강화하는 방향으로 정책을 운영하고 있습니다.

[그림 11] 네이버 'AI 윤리 · 안전성 개선 프로세스'



Source: 네이버, Naver Integrated Report 2023 (2024.07)

3.3.3 AI 서비스 설명 제공 방안

AI 시스템은 다양한 분야에서 업무를 지원하고, 일상적인 작업을 보조하는 데 유용하게 사용되고 있습니다. 그러나 AI 시스템은 특정 목적에 맞게 설계된 도구이기 때문에, 모든 문제를 해결할 수 있는 범용적인 해결책은 아닙니다. 또한 설계 목적에 부합하는 영역이라 하더라도 AI가 항상 정확한 결과를 도출하는 것은 아니며, 특정 상황에서는 오류나 예외적인 판단이 발생할 수 있습니다. 예를 들어 AI는 대량의 데이터를 분석하여 예측하거나 자동화된 결정을 내릴 수 있지만, 그 결과가 항상 정답이거나 완벽하게 신뢰할 수 있는 것은 아닙니다.

따라서 이용자는 AI 시스템이 처리할 수 있는 작업의 범위와 한계를 이해해야 하며, 필요한 경우 인간의 판단과 검토가 병행되어야 합니다. AI 사업자는 이러한 사실을 이용자가 충분히 이해할 수 있도록 명확하게 설명해야 하며, 이용자가 AI가 제공하는 결과를 무조건적으로 신뢰하기보다 적절한 주의와 검토 속에서 활용할 수 있도록 안내해야 합니다.

[그림 12] 이용자 책임 명시 대표 사례

[예시] 해양 정보 제공 AI 챗봇

해당 AI 서비스는 LLM 모델에 기반해 만들어졌으며, 많은 양의 데이터를 학습하고 이를 바탕으로 답을 하고 있습니다. **제한적인 영역에서 외부 정보를 검색하여 정확한 정보를 제공해 드릴 수 있지만, AI 챗봇이 제공하는 응답이 100% 정확하지는 않을 수 있습니다.** 이용자가 AI 챗봇의 응답에서 도움을 얻되, 해양산업 관련 전문성이 필요하거나, 정확도가 중요한 정보의 경우에는 이용자가 직접 해당 정보에 대한 추가 확인을 진행할 것을 권장합니다.

Source: 방송통신위원회 · 정보통신정책연구원, 생성형 인공지능 서비스 이용자 보호 가이드라인 (2025.02)

Chapter 02. 기본원칙 및 고려사항

고영향 AI 개발 사업자는 시스템의 주요 기능과 작동 원리, 사용 데이터의 특성, 자동화된 결정 방식과 그 한계에 대한 정보를 이용 사업자에게 명확하게 제공하고, 이용자가 이를 이해할 수 있도록 충분히 설명해야 합니다. 또한 윤리적 문제나 예기치 않은 오류 가능성 등 AI 시스템 사용 시 발생할 수 있는 위험 요소와 그에 대한 대응 절차를 주기적으로 최신화하여 제공해야 하며, AI 서비스 개선 이력이나 보안 업데이트와 같은 중요 정보도 지속적으로 공유해 이용자가 AI를 안전하게 사용할 수 있도록 지원해야 합니다.

고영향 AI 이용 사업자가 AI 개발 사업자로부터 제공받은 AI 모델을 기반으로 자사의 서비스를 구축 · 운영하는 경우, 서비스 제공 주체로서 최종 이용자에게 AI 시스템에 대해 명확히 이해할 수 있도록 설명할 책임을 가집니다. 특히 AI 시스템이 이용자의 권리나 일상에 미칠 수 있는 영향과 잠재적 리스크를 충분히 고지하여 이용자가 안전하게 서비스를 이용할 수 있도록 해야 합니다.

[그림 13] 카카오뱅크 대화형 AI 서비스 이용약관

대화형 AI 서비스 이용약관

제 1조 목적 …
제 2조 용어의 정의 …

제 3조 서비스의 내용
1. 카카오뱅크가 제공하는 서비스의 범위는 다음과 같습니다
   1. 카카오뱅크의 상품/서비스 정보 및 금융 정보 검색 서비스
   2. 이자 및 환율 등의 금융 계산 서비스
   3. 이체 등 금융 거리 요청 및 수행 서비스
   4. 그 외 카카오뱅크가 정하는 서비스
2. 이러한 서비스는 개별적으로 통합된 형태로 제공될 수 있으며, **정책에 따라 변경되거나 새로운 서비스가 추가될 수 있습니다.** 이 경우, 카카오뱅크는 **구체적인 내용과 서비스의 변경, 추가, 중단, 제한 등의 내용을 모바일 앱 등을 통해 사전에 안내합니다.**

Source: 카카오뱅크, 대화형 AI 서비스 이용약관 (2025.05)

Chapter 02. 기본원칙 및 고려사항

AI 서비스는 이용자와 상호작용하는 과정에서 이용자의 데이터를 수집 · 처리 · 활용합니다. 이용자는 자신이 제공한 데이터가 AI 서비스에서 어떤 방식으로 활용되는지 알 권리가 있으며, AI 사업자는 이용자 데이터의 활용 방침을 투명하게 공개해야 합니다. 특히 고영향 AI 사업자는 개인정보 활용의 목적 · 범위 · 방식을 사전에 명확히 고지하고, 실제 활용된 내역과 관련 정보를 이용자가 확인할 수 있도록 투명하게 제공해야 합니다.

이용자는 자신이 제공한 데이터의 활용 여부에 대해 동의 또는 거부할 선택권을 가지며, AI 사업자는 이용자가 권리를 행사하기 위해 필요한 정보를 충분히 제공해야 합니다. 예를 들어 이용자가 입력한 데이터가 AI 모델 학습에 활용되는지 혹은 특정 서비스 기능 제공에만 활용되는지 명확히 안내해야 합니다. 또한 이용자의 이러한 결정이 AI 서비스 품질에 어떤 영향을 미칠 수 있는지에 대해서도 사전에 설명해야 하며, 데이터 이용 중단 · 이의 제기 등의 권리 행사 절차 또한 명확히 알려야 합니다. 마지막으로, 이용자가 이러한 권리를 쉽게 행사할 수 있도록 시스템 UI/UX 차원에서 선택권을 명확하고 편리하게 제공해야 합니다.

❖ 학습 데이터 활용 여부 설명 예시

Q AI 챗봇과 대화한 내용은 학습에 활용되는 건가요?

A 인공지능과 대화한 데이터는 누구의 대화 내용인지 알 수 없도록
비식별 처리를 완료한 뒤, 학습에 활용될 수 있습니다.
이용자님이 인공지능과 대화한 데이터의 학습을 원치 않는 경우,
언제든지 학습데이터 활용 비동의 처리를 할 수 있습니다.

Source: 방송통신위원회 · 정보통신정책연구원, 생성형 인공지능 서비스 이용자 보호 가이드라인 (2025.02)

Chapter 02. 기본원칙 및 고려사항

고영향 AI 사업자는 이용자가 시스템을 안전하고 적법하게 사용할 수 있도록 명확한 정책과 지침을 마련해야 합니다. 이용자가 AI 시스템을 잘못 활용할 경우 법적 · 윤리적 문제가 발생할 수 있으므로 이를 예방하기 위한 조치가 필요합니다. 이를 위해 사용 정책과 계약 조건을 제시하고, 이용자가 해당 내용을 충분히 인지할 수 있도록 안내해야 합니다. 특히 불법·유해 활동, 혐오 표현, 개인정보 침해 등 금지되는 행위를 구체적으로 명시하고, 서비스 제공 조건 · 책임 제한 등 이용자와 사업자 간 권리 · 책임 관계를 명확히 규정해야 합니다.

기업 사례로 엔트로픽(Anthropic)은 AI 모델의 안전하고 책임 있는 사용을 위해 금지된 활동, 데이터 처리 조건, 책임 범위, 개인정보 보호 원칙 등을 포함한 이용정책을 마련해 공개하고 있으며, 이를 통해 AI 모델 오·남용을 예방하고 이용자 권리를 보호하고 있습니다.

[표 18] 엔트로픽 '이용자 보호를 위한 이용 정책'

이용 정책	주요 항목	상세 내용
사용 정책	불법 및 유해활동	불법 행위, 사기, 악성코드 생성, 해킹, 스팸, 테러리즘 등
	혐오 표현	인종, 민족, 종교, 성별, 성적 지향 등 혐오 조장 콘텐츠
	개인정보 침해	민감한 개인정보를 무단으로 수집 및 유출하는 행위
소비자 서비스 약관	서비스 제공 및 구독	각 서비스에 대한 제공 조건 및 결제 방식 규정
	지적 재산권	이용자가 생성한 콘텐츠에 대한 소유권 인정 (다만 AI 서비스 개선을 위해 콘텐츠를 사용할 수 있음을 명시)
	책임 제한	서비스가 있는 그대로 제공되며, 특정 목적에 대한 보증 및 책임은 제한된다고 명시
	학습 데이터 선택권	이용자가 자신의 데이터가 AI 모델 학습에 사용되는 것에 대해 동의하거나 거부할 수 있는 선택권 부여
개인정보 보호 정책	수집 정보	입력 콘텐츠, 이용자 계정 정보 등 수집 데이터 명시
	데이터 보관 및 권리	최대 5년 간 데이터 보관되나, 이용자는 언제든지 데이터 삭제 및 설정 변경 가능
	개인정보 공유	법적 의무가 있으나, 서비스 제공을 위해 신뢰할 수 있는 제 3자에게 최소한의 정보를 공유할 수 있음을 명시

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

또한 고영향 AI 사업자는 이용 정책을 이용자에게 효과적으로 전달하기 위한 방안을 마련해야 합니다. UI/UX 반영, FAQ · 가이드 제작, 이용자 교육, 피드백 수집 등 다양한 방식과 채널을 활용해야 합니다. 다만 현재 대부분의 AI 서비스는 이용자 설명 제공 방식이나 기준이 표준화되어 있지 않아, 이용자가 정책과 지침을 확인하는 데 어려움이 있습니다. 따라서 장기적으로는 서비스 인터페이스와 설명 기준을 점진적으로 표준화하는 노력이 필요합니다.

[표 19] 이용자 설명 방안 시행 절차 및 예시

절차	주요 활동	예시
UI/UX	설명 방안에 맞춰 제품 인터페이스에 설명 요소를 직접 구현	AI 서비스의 초기 화면에 문구 표시
콘텐츠 제작	FAQ, 가이드 문서, 블로그 글 등 상세 설명 콘텐츠 제작 및 배포	AI 추천 시스템 작동 방식 등의 게시글을 통해 AI 작동 원리 설명
내부 교육 시행	개발, 마케팅, 고객 지원 등 관련 팀에 이용자 설명 방안의 중요성 교육	이용자 질문에 대한 질의응답 매뉴얼 제공
피드백 수집 및 분석	이용자로부터 설명에 대한 피드백 정기 수집 및 분석하여 개선점 모색	설명 팝업 하단에 만족도 조사 링크 및 버튼 추가하여 유용성 평가
정기 업데이트	AI 모델의 업데이트나 신기능 추가 개발 시 설명 내용과 방식 최신화	AI 모델의 학습 데이터 갱신 시점 관련 소개 문구 업데이트

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

3.4. 활용 단계

3.4.1 AI 서비스 이용자 보호 방안

AI 서비스의 활용 단계에서는 실제 운영 과정에서 발생할 수 있는 위험을 최소화하고, 이용자의 권익을 보호할 수 있는 체계적인 장치가 필요합니다. 이를 통해 서비스가 안전하고 신뢰성 있게 제공될 수 있도록 보장해야 합니다.

이용자 보호의 핵심 요소는 문제 발생 시 즉각 대응할 수 있는 신고 채널을 구축하는 것입니다. 오류, 불공정한 결과, 개인정보 침해 등 이용자가 겪을 수 있는 문제를 빠르게 신고할 수 있도록 접근성을 높여야 하며, 접수된 신고를 투명하고 일관된 절차에 따라 처리할 수 있는 대응 체계를 마련해야 합니다. 이를 통해 이용자는 문제 상황에 대해 후속 조치를 신속히 받을 수 있고, AI 사업자는 발생 가능한 리스크를 조기 식별해 개선 조치를 취할 수 있습니다.

❖ **이용자 신고 채널 마련 시 주요 고려사항**

채널 접근성	• 서비스 화면 내 이용자가 쉽게 찾을 수 있는 위치에 신고 기능 배치 (예. 신고 버튼 페이지 상단 도는 하단에 고정 배치 등)
사용 친화적 신고 절차	• 불필요한 단계가 최소화된 간단하고 명확한 신고 절차 도입 (예. 신고 과정 중 필수 정보만 요구, 신고 유형 분류 자동화 등)
다양한 신고 경로	• 이용자 상황에 맞는 신고 경로 선택을 위한 접수 채널 다양화 (예. 이메일, 전화, 실시간 채팅 등)

Chapter 02. 기본원칙 및 고려사항

고영향 AI 사업자는 이용자들의 불만과 의견을 적극적으로 수렴하고 이를 바탕으로 지속적인 시스템 개선을 이루어야 합니다. 이를 위해 고객 센터 외에도 웹 폼, 모바일 앱, 챗봇, 이메일 등 다양한 피드백 수집 채널을 제공하여 이용자가 쉽게 의견을 제출할 수 있도록 해야 합니다. 특히 고령층, 어린이 등도 어려움 없이 피드백 수집 채널에 접근할 수 있도록 접근 경로와 활용 UI/UX 등을 설계하는 것이 중요합니다.

대표적인 예로 한국정보통신위원회는 AI 서비스와 관련된 피해나 불만을 접수하기 위해 AI 서비스 이용자 피해 신고 창구를 운영하고 있습니다. 이용자는 온라인 피해 365센터 홈페이지의 AI 피해신고 버튼을 통해 신고할 수 있으며, 신고하기 전 365센터의 전화 또는 카카오톡 채널을 통해 신고 절차와 관련된 상담을 받을 수 있는 기능도 제공하고 있습니다.

[표 20] 이용자 피드백 수렴 및 적용 방안

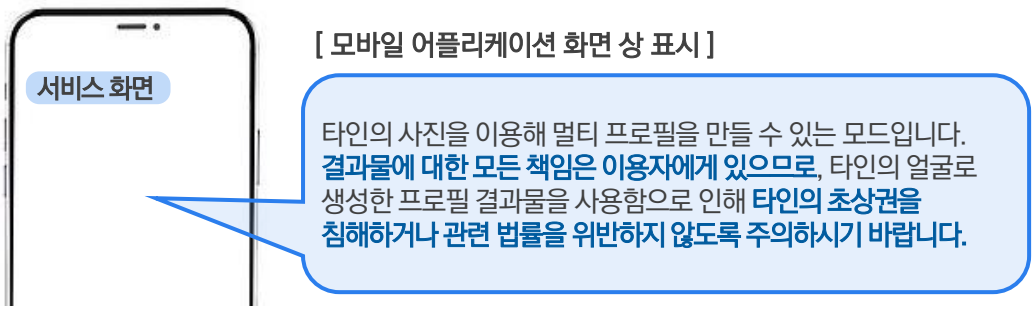
수집 방안	주요 항목	상세 내용
직접적 이용자 피드백 수집	피드백 버튼	이용자 인터페이스에 만족도 조사 등 간단한 평가 UI 제공
	신고 기능	부적절한 결과물, 편향 등을 쉽게 신고할 수 있도록 설계
	별점 평가 및 의견	응답 품질에 대한 이용자 평가 및 의견 수집
정량적 로그 분석 및 모니터링	행동 로그 분석	이용자의 클릭, 스크롤, 재입력 등을 분석해 만족도 추정
	비정상 응답 탐지	응답이 반복되거나 비논리적인 경우, 자동으로 로그 수집 및 분석
	오류 로그 수집	시스템에서 발생한 예외 상황 및 예측 실패 사례 기록
이용자 그룹 사전 테스트	파일럿 운영	특정 이용자 그룹 대상 제한적 운영 후 피드백 수집
	사용성 테스트	이용자가 AI 기능을 사용할 때의 행동과 불편 요소 관찰
	A/B 테스트	여러 모델 버전 또는 기능을 이용자에게 제공해 반응 비교

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

Chapter 02. 기본원칙 및 고려사항

AI 사업자는 고영향 AI 또는 GenAI를 활용한 제품 · 서비스를 제공하는 경우, 해당 사실을 이용자에게 사전에 명확히 고지해야 합니다. 이러한 고지는 이용자가 해당 제품 또는 서비스가 고영향 AI나 GenAI 기반으로 운용되고 있음을 인지하고, 보다 신중하게 사용할 수 있도록 하는 중요 절차입니다. 이를 위해 이용약관, 서비스 가입 절차, 계약서 등 공식 문서에 고영향 AI · GenAI 활용 여부를 명시하거나, 소프트웨어 · 모바일 어플리케이션 화면 내에 AI 활용 여부를 표시할 수 있습니다.

[예시] AI 프로필 생성 서비스



Source: 방송통신위원회 · 정보통신정책연구원, 생성형 인공지능 서비스 이용자 보호 가이드라인 (2025.02)

법/규제 조항

과태료 대상

AI 기본법 시행령안 제22조(인공지능 투명성 확보 의무)

① 인공지능사업자는 고영향 인공지능이나 생성형 인공지능을 이용한 제품 또는 서비스(이하 "제품등"이라 한다)를 제공하기 전에 다음 각 호의 어느 하나의 방법으로 법 제31조제1항에 따른 사전고지를 하여야 한다.

1. 제품등에 직접 기재하거나, 계약서,사용 설명서, 이용약관 등에 기재
2. 이용자의 화면 또는 단말기 등에 표시
3. 제품등을 제공하는 장소(해당 장소와 합리적으로 관련된 범위의 장소를 포함한다)에 인식하기 쉬운 방법으로 게시
4. 그 밖에 제품등의 특성을 고려하여 과학기술정보통신부장관이 인정하는 방법 · · ·

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

Chapter 02. 기본원칙 및 고려사항

AI 사업자는 제품 또는 서비스에 이미지 · 음성 · 영상 등 GenAI에 의해 생성된 결과물이 포함되어 있는 경우, 해당 결과물이 AI에 의하여 생성되었음을 명확하게 표시해야 합니다. 이러한 고지는 이용자의 혼란을 방지하고, 딥페이크와 같은 악용 가능성을 줄이는 데 중요한 역할을 합니다. GenAI 결과물 여부의 표시 방식은 콘텐츠 형식에 따라 시각적 또는 청각적으로 제공될 수 있으며, 소프트웨어 기반의 비가시적 표시도 활용할 수 있습니다. 예를 들어 메타는 고해상도 동영상의 모든 프레임이 아닌, 특정 키 프레임에 워터마크를 삽입하는 효율적인 비디오 워터마킹 기술인 '비디오 실'을 개발한 바 있습니다.

또한 AI 사업자는 AI 시스템이 생성한 음향 · 이미지 · 영상 결과물이 실제와 구분하기 어려운 경우, 이용자가 이를 명확히 인식할 수 있도록 적절한 방식으로 표시해야 합니다. 이때 이용자의 접근성과 이해도를 높일 수 있도록 이용자 특성을 고려한 고지 · 표시 방식을 선택하는 것이 중요합니다. 다만 예술적 또는 창의적 표현물의 경우에는 전시나 감상 경험을 해치지 않는 범위 내에서 표시 방식을 조정할 수 있습니다.

[표 21] 이용자 특성 별 고지 · 표시 설계 고려사항

구분	상세 구분	고려사항
연령	아동, 성인, 노인 등	아동 또는 노인의 경우, 성인과 비교해 이해할 수 있는 어휘, 단어에 한계가 있어 이용자 연령을 고려해야 함
장애 유무	장애인, 비장애인	신체적 제약으로 발생할 수 있는 한계를 고려해야 함. 그 예로는 신체 크기, 신체 능력, 인지능력이 있음
지식	초보자, 전문가 등	관련 서비스의 경험 여부와 사전 배경지식의 차이로 지식 수준이 다름을 고려해야 함

Source: 과학기술정보통신부 · 한국정보통신기술협회, 2024 신뢰할 수 있는 인공지능 개발 안내서 (2024.02)

3.4.2 AI 교육 및 변화 관리

AI 기술의 도입은 단순히 기술적인 변화를 넘어, 업무 프로세스, 조직 문화, 그리고 구성원들의 역할까지 포함하는 광범위한 영역의 변화를 수반합니다. 이로 인해 AI 도입 및 적용에 따른 업무 환경의 급격한 변화에 적응하기 위해 임직원 대상 다양한 내용과 형식의 AI 교육을 제공해야 합니다. AI 교육은 조직 구성원들이 AI의 기능과 작동 원리, 그리고 이를 활용한 새로운 업무 방식에 대해 정확히 이해할 수 있도록 지원합니다.

AI 리터러시 교육은 구성원들이 AI 기술을 이해하고 적절히 활용할 수 있는 기본 역량을 배양하기 위한 교육입니다. 대표적으로 이케아(IKEA)는 약 3천명의 직원들을 대상으로 AI 기술을 실제 업무에 적극적으로 활용할 수 있는 역량을 키우기 위한 AI 리터러시 교육을 진행하고 있습니다. 업무와 직급 등에 따라 AI 활용 목적과 필요 기능 등이 상이한 만큼, 직무와 연차별 맞춤형 교육 프로그램을 설계하여 직원들의 AI 이해도와 활용 능력을 단계적으로 향상시키고 있습니다.

❖ AI 리터러시 주요 구성요소

이해	AI의 작동 방식, 데이터의 중요성, 알고리즘의 구조와 한계 파악
비판	AI 결과의 신뢰성과 편향 여부 분석
활용	다양한 환경과 목적에 맞게 AI를 사용하는 전략
윤리	AI 오·남용 방지와 사회적 신뢰 유지를 위한 법적 · 사회적 책임의식

Source: 웨이크업, AI리터러시, 인공지능 시대의 필수 생존 역량 (2025.08)

Chapter 02. 기본원칙 및 고려사항

고영향 AI 개발사업자는 AI 이용사업자가 업무 환경에서 AI를 안정적으로 운영할 수 있도록 기본 운영지식과 절차를 교육해야 합니다. 특히 입·출력 오류, 데이터 편향, 시스템 중단 등 AI 운영 과정에서 발생할 수 있는 주요 오류 유형과 그 대응 방안을 명확히 안내해야 합니다.

AI 이용사업자 역시 자사 임직원이 변화된 업무 환경에서 AI 시스템을 올바르게 활용할 수 있도록 교육과 훈련을 실시해야 합니다. 임직원들이 AI의 한계와 오류 가능성, 적절한 사용 방법을 충분히 이해해야 업무 전반의 안정성과 책임 있는 활용이 보장되기 때문입니다.

대표적으로 신한금융그룹은 경영진 및 임직원을 대상으로 개인정보 보호와 책임 있는 데이터 활용을 중심으로 한 교육을 시행함으로써, AI · 데이터 기반 업무 환경에 필요한 인식과 역량을 강화하고 있습니다.

법/규제 조항

과태료 대상

AI 기본법 제34조(고영향 인공지능과 관련한 사업자의 책무)

① 인공지능사업자는 고영향 인공지능 또는 이를 이용한 제품·서비스를 제공하는 경우 고영향 인공지능의 안전성·신뢰성을 확보하기 위하여 다음 각 호의 내용을 포함하는 조치를 대통령령으로 정하는 바에 따라 이행하여야 한다.

1. 위험관리방안의 수립·운영
2. 기술적으로 가능한 범위 내에서의 인공지능이 도출한 최종결과, 인공지능의 최종결과 도출에 활용된 주요 기준, 인공지능의 개발·활용에 사용된 학습용데이터의 개요 등에 대한 설명 방안의 수립·시행
3. 이용자 보호 방안의 수립·운영
4. 고영향 인공지능에 대한 사람의 관리·감독

…

관련 기술고시 조문 제 7조 (사람의 관리 · 감독)

② 사업자는 고영향 인공지능 운영 중 사람의 관리 · 감독을 위해 다음 각 호의 조치를 이행하여야 한다.

1. 성능저하 및 오류 발생에 대한 정기적인 점검계획 및 방안 마련
2. 인공지능의 범위 및 수행 능력에 대한 이해도를 향상시키기 위한 교육 및 훈련 제공

Source: 과학기술정보통신부, AI 기본법 하위법령집 (2025.09)

Chapter 02. 기본원칙 및 고려사항

AI 기술 확산으로 인해 사람 중심으로 수행되던 반복적 · 규칙 기반의 업무가 자동화되면서 전반적인 업무 프로세스가 변화하고 있습니다. 이러한 변화는 조직 경쟁력을 높이는 동시에 기존 인력에게는 직무 변화로 인한 불확실성과 새로운 업무 환경에 대한 적응 부담을 유발합니다. 따라서 기업은 AI 기술의 도입과 함께 직무 전환 대상자를 위한 체계적인 교육과 직무 전환 지원 체계를 마련해야 합니다. 예를 들어, 데이터를 수동 입력하던 기존 업무를 AI 시스템을 통해 자동화할 경우, 직원의 업무는 데이터 입력에서 AI가 생성한 결과를 분석 · 해석하는 업무로 전환될 수 있습니다.

직무 전환 교육은 기술적 역량 확보 뿐만 아니라 새로운 업무 환경에 적응할 수 있도록 돕는 심리적, 조직적 지원이 함께 이루어져야 합니다. 업무 방식과 환경의 급격한 변화는 그것에 적응해야 하는 구성원에게 심리적인 부담을 줄 수 있으므로, 자신감 회복과 안정감 형성을 돕는 프로그램이 필요합니다. 또한 AI 도입으로 직무 구조와 인력 수요가 빠르게 변화하고 있는 만큼 기업 또한 인재관리 및 조직전략을 유연하게 재편하고, 단순 인력 감축에서 나아가 재교육 · 재배치 중심의 전략적 대응을 마련해야 합니다.

❖ AI 직무 전환 교육 대표 사례

싱가포르 해양항만청 '중견 경력자 직무 전환 프로그램'

2025년 싱가포르 해양항만청(MPA)은 해양 디지털화, 탈탄소화 및 사이버 보안 분야에서의 직무 전환을 위해 중견 경력자 직무 전환 프로그램을 개편하여 발표했다. 해당 프로그램은 싱가포르 해양협회 및 워크포스 싱가포르(WSG)와 협력하여, **기존 해양 산업 종사자들을 디지털화 및 친환경 선박 운항 역할로 재배치하는 데 중점**을 두고 있다.
온라인 훈련 및 현장 실습을 통해 **디지털 항해사, AI 기반 선박 운용 관리자 등 신규 직무로의 원활한 전환을 지원**하며, 프로그램에 참여한 해양 기업들에게 급여 보조와 훈련 수당을 제공하여, 중견 경력자들의 직무 전환에 대한 경제적 부담을 경감한다.

Source: 싱가포르 해양항만청, 300 Maritime Workers to Build Digital and Technical Capabilities under Enhanced Career Conversion Programme by WSG and MPA (2025.03)

3.4.3 AI 산출물 저작권 관리

AI 기술이 발전하면서 AI 시스템이 생성하는 콘텐츠와 결과물에 대한 법적, 윤리적 관리가 중요 이슈로 대두되고 있습니다. 특히 AI가 생성한 결과물이 상업적 · 예술적 가치를 갖는 경우, 저작물의 소유권과 이용권에 대한 법적 기준을 명확히 설정하고 관리하는 것은 AI 사업자와 이용자가 반드시 고려해야 할 중요 사항입니다.

AI가 생성한 산출물은 인간 저작자의 창작적 표현이 존재해야 한다는 저작권의 기본 요건을 충족하지 못하기 때문에 일반적으로 저작물로 인정되지 않습니다. 다만 AI가 산출물의 제작 과정에 개입하더라도 인간이 창작적으로 기여한 부분이 있다면 그 기여 범위에 한해 저작권이 인정될 수 있습니다.

대표적으로 미국에서는 인간이 작성한 글과 GenAI Midjourney를 통해 생성한 그림을 결합해 제작한 만화가 미국 저작권청에 등록된 사례가 있습니다. 미국 저작권청은 GenAI를 통해 생성한 그림에 대해선 저작권 등록을 인정하지 않았으나, 인간이 작성한 글, 그림을 선택 · 배열 · 조정한 편집적 기여에 대해서는 창작성을 인정하고 저작권 등록을 허용했습니다.

❖ GenAI 기반 생성 이미지 저작권 등록 대표 사례

미국 저작권청 '한 조각의 아메리칸 치즈'

2025년 1월 GenAI 플랫폼을 제공하는 인보크 AI(Invoke AI)사의 CEO인 켄트 키어시가의 '한 조각의 아메리칸 치즈' 라는 제목의 저작물의 저작권 등록이 허가되었다. 최초 저작권 등록 신청 시 인간의 저작 요소가 부족하다는 이유로 등록이 거절되었으나, 인보크 측에서 프롬프트 입력과 인페이팅 조작이 결합된 일련의 **이미지 생성 과정을 담은 비디오 클릭과 창작적 기여에 대한 설명을 추가로 제출**하면서, 미국 저작권청은 종전의 거절결정을 철회하고 **인보크의 저작권 등록 신청을 받아들였다.**

Source: 한국저작권위원회, 미국 생성형 AI를 활용한 이미지의 저작권 등록 사례 (2025.02)

Chapter 02. 기본원칙 및 고려사항

AI가 생성한 산출물이 기존의 저작물과 유사하거나 동일한 경우, 저작권 침해 문제가 발생할 수 있습니다. 특히 GenAI 경우 대량의 데이터를 학습하는 과정에서 기존 저작물이 포함될 가능성이 높고, 그와 같거나 유사한 산출물을 생성할 수 있으므로 저작권 침해 위험성이 큽니다. 저작권 침해 여부는 의거성과 실질적 유사성에 대한 판단에 따라 결정됩니다. 의거성은 AI 산출물이 기존 저작물을 인식하고 그에 따라 생성되었는지 여부, 즉 기존 저작물에 대한 의도적 사용 여부를 확인하는 것입니다. 실질적 유사성은 AI 산출물과 기존 저작물이 동일하거나 유사한지 여부를 평가하는 기준입니다.

AI 산출물이 기존 저작물의 저작권을 침해했는지 여부는 최종적으로 법원의 판단을 통해 결정됩니다. AI 사업자는 서비스 제공 시 기존 저작물과 동일하거나 유사한 AI 산출물이 도출되지 않도록 함으로써 저작권 침해를 미연에 방지하는 것이 바람직합니다. 예를 들어 프롬프트를 입력할 때 특정 저작물의 내용을 그대로 입력하거나 해당 저작물과 동일하거나 유사한 결과물이 나오도록 유도하는 표현을 입력하는 것은 지양해야 합니다.

❖ 저작권 침해가 문제될 수 있는 프롬프트 예시

특정 작가의 작품 직접 요구	• 000사의 000 캐릭터가 나오는 일러스트를 생성해줘 • 000작가의 소설 000 전체를 원문 그대로 보내줘 • 드라마 000의 12화에서 000가 웃던 그 장면을 그림으로 그려줘
특정 작품 각색 및 변형 요구	• 만화 000의 00이 주인공인 새로운 에피소드를 원작 그림체 그대로 만들어줘 • 영화 000의 결말을 변경해서 그림으로 보여줘

Source: 한국저작권위원회, 생성형 인공지능 결과물에 의한 저작권 분쟁 예방 안내서 (2025.06)

Chapter 02. 기본원칙 및 고려사항

최근에는 저작권자의 권리 행사와 AI 산업 발전 간의 균형을 맞추기 위한 정책적 · 기술적 방안들이 국내외에서 활발히 검토되고 있습니다. 저작권자는 자신의 창작물이 GenAI 학습에 활용되는 것을 동의하지 않을 경우, 명확한 거부 의사를 표현할 수 있는 방법들을 사용할 수 있습니다.

❖ AI 학습 거부의사 발표 사례

대표 기업	주요 내용
방송사 S사	홈페이지, 포털, 유투브 채널 등을 통해 AI 학습 금지 선언
언론사 A사	AI 및 대량 크롤링 방지 약관 신설
출판사 P사	책의 표준 저작권 페이지에 '이 책의 어떤 부분도 AI 기술이나 시스템을 훈련하는 목적으로 어떤 방식으로든 사용 및 복제할 수 없다 선언

Source: 한국저작권위원회, 생성형 인공지능 결과물에 의한 저작권 분쟁 예방 안내서 (2025.06)

또한 자신의 저작물이 무단으로 AI 학습 데이터에 포함되는 것을 방지하고, AI 기반 산출물로 인한 저작권 분쟁을 예방할 수 있는 여러 기술적 조치를 취할 수 있습니다. 예를 들어, 웹사이트나 데이터 저장소에서 크롤러 접근을 제한하거나, API 권한을 엄격히 관리하는 방법이 있습니다. 또한, 인터넷에 공개된 저작물은 특별한 이용 허가 없이 GenAI의 학습에 포함될 위험이 있으므로, 저작권자는 자신이 제공하는 저작물의 인터넷 공개 방식을 주기적으로 점검하는 것이 중요합니다. 저작물에 저작권자의 정보, 이용 허락 범위, 출처 등을 메타데이터로 삽입하고, 이를 디지털 저작권 관리(DRM) 시스템과 결합하는 방법도 저작권 보호에 효과적일 수 있습니다. 이러한 조치들은 무단 사용을 방지하고, 저작권자의 권리를 보호하는 데 중요한 역할을 합니다.

Chapter 03.
서비스 도입 체크리스트

kobc 한국해양진흥공사
KOREA OCEAN BUSINESS CORPORATION

Chapter 03. 서비스 도입 체크리스트

1. 일반

본 챕터 1.1.공통과 1.2.고영향 AI는 해양 기업이 AI 서비스 개발 및 활용 시, 일반적으로 검토해야 하는 요건을 정의했습니다. GenAI, Agentic AI, Physical AI를 개발 및 활용하는 경우에는 AI 모델 유형별 특성을 반영한 2. 추가 고려사항을 함께 검토하실 것을 권고합니다

1.1. 공통

1.1.1 기획 단계

점검항목		요구사항 및 체크리스트	Y	N
AI 윤리	[1]	AI 개발 · 활용 전 과정에서 지켜야 할 AI 윤리 원칙을 수립하고, 이를 공식 문서(예. 정책서, 내부 규정 등)로 명시하였는가?	☐	☐
	[2]	AI 도입 목적과 업무 범위가 명확히 정의되어 있으며, 윤리적 기준에 부합하는가?	☐	☐
	[3]	AI 거버넌스 체계(예. AI 윤리원칙, 지침 및 규정, 조직 등)를 수립하였는가?	☐	☐
	[4]	AI 시스템의 생명주기에서 발생할 수 있는 위험 요소를 파악하고, 그 영향을 평가하였는가?	☐	☐
	[5]	AI 위험 요소에 대해 실질적인 완화 및 대응 방안이 마련되었으며, 그 효과가 검증되었는가?	☐	☐
AI 사업 / 서비스 운영	[1]	국내에 주소나 영업소가 없는 AI 사업자인 경우, AI 기본법 시행령상 '국내대리인 지정 사업자 기준'에 해당하는지 확인하였는가? 과태료 * 대통령령으로 정하는 국내 대리인 지정 사업자 기준은 AI 기본법 시행령안 제28조(국내대리인 지정 사업자의 기준) 참고 바랍니다.	☐	☐
	[2]	국내대리인 지정 사업자는 국내대리인을 서면으로 지정하고 과학기술정보통신부장관에게 신고하였는가?	☐	☐
	[3]	국내대리인이 AI 안전성 확보 의무 이행 결과 제출 등 AI 기본법 상 규정된 대리 사항을 제대로 이행했는지 확인하였는가?	☐	☐

Chapter 03. 서비스 도입 체크리스트

점검항목		요구사항 및 체크리스트	Y	N
고영향 AI 판단 · 확인	[1]	AI 또는 AI를 이용한 제품 · 서비스를 제공하는 경우, 고영향 AI에 해당하는지 사전 검토하였는가? 과태료 * 고영향 AI 판단 · 확인을 위한 상세 내용은 AI 기본법 시행령안 제24조(고영향 인공지능의 확인) 참고 바랍니다.	☐	☐
AI 안전성 확보 의무 대상 판단	[1]	학습에 사용된 누적연산량이 10의 26승 부동소수점 연산 이상인 AI 시스템에 해당하는가?	☐	☐
	[2]	AI 기술의 발전 수준을 고려하여 현재 AI 시스템에 최첨단 AI기술을 적용하여 구성 및 운영하는가?	☐	☐
	[3]	AI 시스템의 위험도가 사람의 생명, 신체의 안전 및 기본권에 광범위하고 중대한 영향을 미칠 우려가 있는가?	☐	☐
	[4]	위 [1]~[3] 항목을 모두 충족하는 AI 시스템을 개발 및 제공하는 AI 사업자 또는 실질적인 변경을 가한 AI 사업자에 해당하는가? 과태료 * AI 안전성 확보 의무 관련 상세 내용은 AI 기본법 시행령안 제23조 (인공지능 안전성 확보 의무) 참고 바랍니다.	☐	☐
위 [1]~[4] 항목 모두 충족하는 경우, AI 안전성 확보 의무 대상자에 해당하며 하기 항목 검토 권고				
AI 안전성 확보 의무 이행	[1]	AI 사업자는 AI 시스템 수명주기 전 과정에서 발생할 수 있는 위험을 사전에 식별하고, 이를 이해 · 관리할 수 있는 체계를 마련하였는가?	☐	☐
	[2]	AI 사업자는 식별된 위험을 체계적으로 관리하고 적절한 대응 방안을 마련하기 위해 정기 평가를 수행하고 있는가?	☐	☐
	[3]	AI 사업자는 AI 안전사고를 모니터링하고 대응할 수 있는 위험관리 체계를 구축하였는가?	☐	☐
	[4]	사고 대응 과정과 향후 재발 방지 계획을 명확히 문서화해 조직 내외에 공유하고 향후 배포 전략에 반영하고 있는가?	☐	☐
	[5]	AI 사업자는 AI 시스템이 안전성 확보 대상 시스템으로 확인된 날로부터 3개월 이내에 안전성 확보 조치 사항을 문서로 작성하여 과학기술정보통신부장관에게 제출하였는가?	☐	☐
	[6]	AI 사업자는 필요한 경우, 영업비밀을 해하지 않는 범위에서 설명 및 검증에 적극 협조할 수 있는가?	☐	☐
	[7]	AI 사업자는 제3자 검증할 있도록 설계 문서, 로그, 평가 결과 등을 포함한 자료를 관리하고 있는가?	☐	☐

Chapter 03. 서비스 도입 체크리스트

1.1.2 개발 단계

점검항목		요구사항 및 체크리스트	Y	N
데이터 수집 및 처리	[1]	학습 목적에 적합한 데이터 수집 범위와 기준을 정의하였는가?	☐	☐
	[2]	데이터 수집 방법, 구축 시점 등을 포함한 데이터 수집 프로세스를 마련하였는가?	☐	☐
	[3]	데이터 보유 기간 설정, 파기 절차 및 방법 등 데이터 관리 계획을 수립하고 준수하고 있는가?	☐	☐
	[4]	신뢰할 수 있는 출처로부터 데이터를 수집 및 확보하여 사용하고 있는가?	☐	☐
	[5]	개인정보 데이터 수집 시, 정보 주체에게 명확히 고지하고 적법한 동의 절차를 거쳤는가?	☐	☐
	[6]	AI 학습 및 서비스 제공에 필요한 최소한의 데이터만 수집하고, 수집 목적 외의 사용을 제한하고 있는가?	☐	☐
	[7]	데이터 투명성 확보를 위해 필요한 명세 자료를 체계적으로 저장 및 관리하고 있는가?	☐	☐
	[8]	전처리 작업 기준 및 절차, 전처리 전과 후의 주요 특성, 품질 점검 결과 등을 명확히 기술하고 문서화하고 있는가?	☐	☐
	[9]	데이터 수집 및 전처리 과정에서 발생할 수 있는 편향을 식별하여, 이를 완화 및 제거하기 위한 방안이 마련되어 있는가?	☐	☐
	[10]	학습 데이터의 변경 이력을 확보하여 데이터 변경이 AI 시스템에 미치는 영향을 평가 및 관리하고 있는가?	☐	☐
	[11]	데이터 유출, 오용, 손실 방지를 위해 적절한 기술적 · 관리적 · 물리적 보안 방안을 적용하고 있는가?	☐	☐
AI 모델 개발	[1]	오픈소스 라이브러리의 안정성을 보장하기 위해 라이선스 준수사항, 보안 취약점 등 위험요소를 점검 및 관리하고 있는가?	☐	☐
	[2]	AI 모델 편향 여부를 테스트하고, 이를 완화하기 위한 적절한 조치를 취했는가?	☐	☐
	[3]	AI 모델 편향성을 모니터링하기 위한 정량적 지표를 설정하여 관리하고 있는가?	☐	☐

점검항목		요구사항 및 체크리스트	Y	N
AI 모델 개발	[4]	AI 모델 추론 결과에 대해 이용자의 판단을 돕기 위한 근거와 설명을 제공하는가?	☐	☐
	[5]	AI 시스템 개발 및 모델 작동 방식에 대한 세부 정보를 문서화하였는가?	☐	☐
	[6]	AI 모델 자체의 취약점, 학습 데이터 오염, 시스템 접근 등 모델 공격 가능성에 대한 위험 분석을 수행하고, 이를 방지하기 위한 방안을 마련하였는가?	☐	☐
AI 시스템 구현	[1]	소스 코드 및 이용자 인터페이스로 인한 편향 요소를 식별하고 개선하였는가?	☐	☐
	[2]	AI 시스템에서 개인정보, 민감정보 등을 활용하는 경우, 활용 필요성을 사전에 평가하고 적합한 안전조치를 마련하였는가?	☐	☐
	[3]	AI 시스템 특성, IT 인프라 환경(예. 폐쇄망 등) 등을 고려하여 개발 및 테스트 환경을 설계하였는가?	☐	☐
	[4]	AI 시스템의 기대 출력을 결정하기 위한 협의 체계를 구성하였는가?	☐	☐
	[5]	AI 시스템의 설명 가능성과 해석 가능성 등을 검증하기 위한 이용자 평가단을 구성하였는가?	☐	☐

1.1.3 운영 단계

AI 시스템 성능 관리	[1]	AI 모델 및 시스템 성능을 평가하기 위한 구체적인 평가 지표, 목표 수준 등을 설정하였는가?	☐	☐
	[2]	안정적인 AI 모델 및 시스템 운영을 위한 성능 모니터링 체계를 수립하였는가?	☐	☐
	[3]	지속적인 AI 모델 및 시스템 성능 점검을 통해, 성능 개선 필요성을 확인하고 있는가?	☐	☐
	[4]	AI 모델 재학습 및 성능 개선을 위한 절차를 수립하고 이를 이행하고 있는가?	☐	☐

점검항목		요구사항 및 체크리스트	Y	N
AI 시스템 설명가능성	[1]	AI 시스템이 중요한 의사결정을 대체하는 경우, 이를 감독하고 통제할 수 있는 절차가 마련되어 있는가?	☐	☐
	[2]	AI가 의사결정을 대체하는 경우, 관리자와 이용자, 기타 이해관계자가 해당 의사결정 과정을 추적하고 해석할 수 있도록 설계되었는가?	☐	☐
	[3]	이용자 경험 모니터링을 통해 이용자 활동 이력(예. 로그 등)을 기록 및 보관하고 있는가?	☐	☐
AI 시스템 안정성 / 보안성 관리	[1]	AI 시스템의 의사결정으로 인한 파급효과가 크고 불확실성이 높은 경우, 사람의 개입이 필요한 상황에 대한 시나리오나 예외 정책 등을 마련하였는가?	☐	☐
	[2]	AI 시스템의 작동을 실시간으로 모니터링할 수 있는 체계가 마련되었는가?	☐	☐
	[3]	침해사고 및 재해 발생 시 피해 확산 방지와 신속한 복구를 위한 체계가 마련되었는가?	☐	☐
	[4]	적대적 공격, 데이터 유출 등을 대비하여 AI 시스템 보안 대책을 수립하고, 이를 시스템 상 구현 및 적용하였는가?	☐	☐
	[5]	예상되는 이용자 오류에 대해 명확한 사전 안내와 대응 절차가 제공되고 있는가?	☐	☐

1.1.4 활용 단계

AI 시스템 이용자 이해도 제고	[1]	이용자의 특성과 제약사항을 고려하여, AI 서비스의 목적과 목표,한계 등을 명확히 설명하였는가?	☐	☐
	[2]	AI 서비스 내 최종 의사결정 주체와 이용자가 상호작용하고 있는 대상에 대해 명확히 인지할 수 있도록 설명하고 있는가?	☐	☐
	[3]	이용자의 불만 · 의견을 수렴하고 개선 요구를 반영하여 지속적으로 AI 서비스를 업데이트하고 있는가?	☐	☐
	[4]	안정적인 AI 운영 및 활용 활성화를 위한 교육 및 훈련을 제공하고 있는가?	☐	☐

Chapter 03. 서비스 도입 체크리스트

점검항목		요구사항 및 체크리스트	Y	N
AI 투명성 확보 의무	[1]	제품 · 서비스가 GenAI에 기반하여 운용된다는 사실을 이용자에게 사전에 고지하였는가?	☐	☐
	[2]	GenAI 또는 이를 이용한 제품 또는 서비스를 제공하는 경우, 그 결과물이 GenAI에 의하여 생성되었다는 사실을 표시하였는가?	☐	☐
	[3]	실제와 구분하기 어려운 이미지, 영상 등의 결과물을 제공하는 경우 해당 결과물이 AI 시스템에 의하여 생성되었다는 사실을 이용자가 명확하게 인식할 수 있는 방식으로 고지 또는 표시하였는가? **과태료** * AI 투명성 확보 의무에 대한 상세 내용은 AI 기본법 시행령안 제22조 (인공지능 투명성 확보 의무) 참고바랍니다.	☐	☐

1.2. 고영향 AI

1.2.1 고영향 AI 사업자 책무

점검항목		요구사항 및 체크리스트	Y	N
고영향 AI 이행 항목	[1]	고영향 AI 사업자는 위험관리방안 수립 및 시행, 이용자 보호방안의 수립 · 운영 등 고영향 AI 사업자 책무로서 의무 이행해야 하는 조치를 수행하였는가? 과태료 * 고영향 AI 사업자 의무 이행 관련 상세 내용은 AI 기본법 시행령안 제26조(고영향 인공지능과 관련한 사업자의 책무) 참고바랍니다	☐	☐
	[2]	AI 시스템을 제공받은 AI이용사업자인 경우, AI개발사업자가 고영향 AI 사업자의 책무를 모두 또는 일부 이행하였는지 확인하였는가?	☐	☐
	[3]	AI 시스템의 중대한 기능 변경을 초래한 AI이용사업자는 고영향 AI 사업자 책무를 이행하기 위하여 필요한 자료를 AI개발사업자에게 요청하여 제공받았는가?	☐	☐
	[4]	고영향 AI 사업자는 위험관리방안 수립 및 시행 등 고영향 AI 사업자 책무로서 의무 이행한 주요 내용과 고영향 AI 담당자 정보(예. 성명, 연락처 등)를 인터넷 홈페이지에 게시하였는가?	☐	☐
	[5]	고영향 AI 사업자는 고영향 AI 사업자 책무 이행의 근거를 5년간 보관할 수 있도록 문서화 및 관리 체계를 마련하였는가?	☐	☐

1.2.2 고영향 AI 의무 이행

위험관리방안 수립 · 이행	[1]	AI 사업자는 AI 위험관리 담당 조직 및 인력을 중심으로 위험관리 계획을 수립하였는가?	☐	☐
	[2]	위험을 관리하는 담당 조직을 구성하거나 조직 내 담당 인력을 지정하였는가?	☐	☐
	[3]	AI 사업자는 위험관리 계획을 지속적으로 이행 및 준수하고 있는가?	☐	☐
	[4]	AI 수명주기 동안 위험관리정책을 조직 내 · 외부 변화에 대응시키고 지속적으로 개선하고 있는가?	☐	☐

Chapter 03. 서비스 도입 체크리스트

점검항목		요구사항 및 체크리스트	Y	N
위험관리방안 수립 및 이행	[5]	AI 사업자는 식별된 위험으로 인해 발생가능한 결과를 분석 및 평가하였는가?	☐	☐
	[6]	위험 요소별로 인명 피해 및 사고를 방지하거나 부정 영향을 최소화하기 위한 적절한 처리 방안(위험 제거, 완화, 모니터링 등)을 명확히 수립하고 실행하고 있는가?	☐	☐
	[7]	AI 위험 정보를 이용사업자 및 이용자에게 충분히 제공하고 있는가?	☐	☐
	[8]	위험관리 담당 조직 또는 인력은 법무, 감사, 개인정보 보호 등 전문성을 갖춘 유관 조직과 AI 위험관리 협업 체계를 구축하고 있는가?	☐	☐
설명 방안 수립 및 이행	[1]	AI개발사업자는 AI의 투명성 및 설명가능성을 확보하고, 가능한 범위에 설명가능성을 높이기 위한 다양한 기술적 조치를 마련하였는가?	☐	☐
	[2]	AI 사업자는 AI가 도출한 최종결과 및 기준을 이용자에게 설명하는 방안을 수립하였는가?	☐	☐
	[3]	AI 사업자는 수립된 설명 방안을 근거로 이용자에게 이를 전달하고 있는가?	☐	☐
이용자 보호방안 수립 및 운영	[1]	데이터 수집 시 이용자 보호를 위해 적법한 절차대로 데이터를 수집 및 관리하고 있는가?	☐	☐
	[2]	알고리즘 설계 및 모델 개발 시, AI 안정성과 견고성 확보를 위해 노력하고 있는가?	☐	☐
	[3]	다양하고 예외적인 상황에 대한 테스트를 수행하고, 이 결과를 평가하여 이용자 보호와 관련된 잠재적 위험을 관리하고 있는가?	☐	☐
	[4]	AI 운영 중 발생할 수 있는 문제를 실시간으로 모니터링 할 수 있는 기술적 기반을 마련하였는가?	☐	☐
	[5]	이용자의 불만 · 의견을 수렴하고 개선 요구를 반영하여 지속적으로 AI를 업데이트하고 있는가?	☐	☐
	[6]	이용자의 권리 보장을 위한 체계 및 보호 정책을 수립하고 이를 이용자에게 안내하고 있는가?	☐	☐

Chapter 03. 서비스 도입 체크리스트

점검항목		요구사항 및 체크리스트	Y	N
사람의 관리 · 감독	[1]	AI가 의도한 목적에서 벗어난 방식으로 작동하거나 비정상적인 결과를 도출하는 경우, 사람이 이를 식별하고 개입할 수 있도록 설계하였는가?	☐	☐
	[2]	AI의 이상 행동을 분석하기 위한 도구(실시간 로그 조회, 추론 이력 확인, 입력-출력 비교 등)을 제공하고 있는가?	☐	☐
	[3]	AI의 성능저하 및 오류 발생을 예방할 수 있도록 정기적인 점검 계획 및 방안을 마련하였는가?	☐	☐
	[4]	AI의 범위 및 수행능력에 대한 이해도를 향상시키기 위한 교육 및 훈련 방안을 제공하는가?	☐	☐
문서화 및 관리방안	[1]	AI 안전성 · 신뢰성 확보를 위한 조치의 내용을 확인할 수 있는 안전신뢰문서를 작성하고 보관하고 있는가?	☐	☐
	[2]	문서를 정기적으로 점검 · 갱신하여 최신 상태를 유지하고 있는가?	☐	☐
	[3]	이용자, 과학기술정보통신부장관 등이 열람을 요청할 시, 안전신뢰문서를 열람할 수 있도록 제공할 수 있는가?	☐	☐
	[4]	관계기간, 이용자 등이 열람을 요청한 문서에 영업비밀 등이 포함되었는지 확인하였는가?	☐	☐
	[5]	영업비밀 등이 포함된 경우, 관련 법령에 따라 해당 내용을 제외한 부분만 선별하여 제공 가능한지 법률 검토를 받았는가?	☐	☐
고영향 AI 영향 평가	[1]	고영향 AI을 이용한 제품 또는 서비스를 제공하는 경우, 사전에 해당 제품 또는 서비스가 사람의 기본권에 미치는 영향을 평가하였는가?	☐	☐
	[2]	고영향 AI 영향 평가를 통해 사전 수립한 위험 시나리오를 법률 전문가의 조력을 받아 검토하였는가?	☐	☐
AI 투명성 확보 의무	[1]	제품 · 서비스가 고영향 AI에 기반하여 운용된다는 사실을 이용자에게 사전에 고지하였는가? 과태료 * AI 투명성 확보 의무 관련 상세 내용은 AI 기본법 시행령안 제 22조 (인공지능 투명성 확보 의무) 참고바랍니다.	☐	☐
사실 조사	[1]	AI 기본법 위반 사항 또는 혐의에 대한 사실 조사 필요 시, 충분한 자료나 증거를 제공할 수 있는가?	☐	☐
	[2]	사실조사 결과에 따라 위반행위의 중지나 시정 명령을 받을 경우 대응할 수 있는 조직과 인력, 절차 등이 마련되어 있는가?	☐	☐

2. 추가 고려사항

2.1. GenAI

2.1.1 GenAI 개발

점검항목		요구사항 및 체크리스트	Y	N
GenAI 학습 데이터 관리	[1]	GenAI 모델 구축을 위한 학습 과정에서 저작권법상 보호되는 저작물이 학습 데이터 내 포함되어 있는지 확인하였는가?	☐	☐
	[2]	타인의 저작물을 이용할 경우, 사전에 저작권자와 이용 허락 계약 등의 방법으로 적법한 이용 권한을 확보하였는가?	☐	☐
	[3]	GenAI 학습에 이용되는 저작물의 권리자가 누구인지 명확하지 않거나 알 수 없는 경우, 저작권법상 법정허락 제도 등을 활용함으로써 적법한 이용 권한을 확보하는 방안을 고려하였는가?	☐	☐
	[4]	AI 사업자는 저작권자와 계약 체결 시 저작물의 이용 목적 범위, 기간 등에 대해 구체적으로 명시하였는가?	☐	☐
GenAI 법적 보호	[1]	범용 AI를 개발하거나 이를 이용해 서비스를 제공하는 사업자는 이용계약 체결 시 결과물의 저작권 및 법적 책임 귀속을 명확히 규정하였는가?	☐	☐
	[2]	GenAI 사업자는 저작권, 명예훼손, 영업비밀 유출 등 법적 리스크의 범위를 충분히 이해하고, 이를 예방하기 위한 내부 절차 및 법적 보호 조치를 수립하였는가?	☐	☐
	[3]	GenAI 서비스가 저작권 침해가 우려되는 요청을 인식하고, 거부 응답을 제공하거나 제한하도록 설계하였는가?	☐	☐
	[4]	GenAI 사업자는 이용약관 또는 고지를 통해 생성된 결과물의 저작권 및 책임 범위를 이용자에게 명확히 안내하였는가?	☐	☐
	[5]	GenAI 사업자는 이용자가 기존 저작물과 동일하거나 유사한 콘텐츠를 생성하지 않도록 주의사항을 명확히 안내하였는가?	☐	☐

2.1.2 GenAI 활용

점검항목		요구사항 및 체크리스트	Y	N
GenAI 저작권 등록	[1]	GenAI 산출물이 저작권법 보호를 받기 위한 저작권 등록 시, 인간의 창작적 기여 부분이 있는지 확인하였는가? * 인간의 창작적 기여 여부에 대한 상세 내용은 한국저작권위원회가 발간한 '생성형 인공지능 활용 저작물의 저작권 등록 안내서'를 참고바랍니다.	☐	☐
	[2]	GenAI 저작권 등록 및 향후 분쟁에 중요한 자료로 활용하기 위해 산출물 생성 및 창작 과정을 영상 등으로 기록하였는가?	☐	☐
	[3]	GenAI 저작권 등록 시, GenAI 산출물 부분과 인간이 창작한 부분을 구분하여, 저작권 보호 범위를 명확히 하였는가?	☐	☐
GenAI 저작물 영리 목적 활용	[1]	GenAI 결과물이 AI 학습 데이터에 무단으로 포함되는 것을 방지하고, 저작권 분쟁을 예방하기 위해 다양한 기술적 수단(예. 크롤러 접근 제한, API 접근 권한 관리 등)을 고려하였는가?	☐	☐
	[2]	GenAI 결과물이 GenAI 학습에 이용되는 것을 원하지 않는 경우, 그에 반대하는 의사를 적절한 방식으로 명시(예. 공식 홈페이지 상 AI 학습 이용 금지 선언 등) 하였는가?	☐	☐
	[3]	저작권 보호를 강화하기 위해 저작물에 저작권자 정보, 이용 허락 범위, 출처 등 저작권 정보를 나타내는 권리정보를 삽입하였는가?	☐	☐
	[4]	GenAI 결과물을 영리 목적으로 이용할 경우 제3자의 권리 침해 여부를 사전에 확인하였는가?	☐	☐
	[5]	GenAI 서비스 약관 상 명시되어 있는 결과물에 대한 저작권 귀속, 상업적 이용 가능 여부 등 제반 규정을 확인하였는가?	☐	☐
	[6]	글이나 이미지, 영상 등을 GenAI로 생성하여 이용할 경우, 해당 사실을 적절한 방식으로 표시하였는가?	☐	☐

2.2. Agentic AI

2.2.1 Agentic AI 기획

점검항목		요구사항 및 체크리스트	Y	N
Agentic AI 업무 정의	[1]	Agentic AI의 업무 범위와 의사결정 권한 수준을 관련 부서 및 이해관계자 간 협의를 통해 명확히 합의하였는가?	☐	☐
	[2]	Agentic AI가 업무 수행에 필요한 시스템과 데이터에만 접근할 수 있도록, 작업 단위별 접근 권한을 세분화하여 관리하고 있는가?	☐	☐
	[3]	Agentic AI가 활용하는 도구 목록, 호출 권한, 사용 조건 및 승인 절차를 정의하고 이를 문서화하여 관리하고 있는가?	☐	☐
Agentic AI 작업 방식 설계	[1]	각 Agent의 업무 범위, 데이터 접근 범위, 의사결정 권한을 명확히 구분하여, 서로의 역할이 겹치거나 충돌하지 않도록 체계적으로 관리하고 있는가?	☐	☐
	[2]	각 Agent별 공급 업체가 다른 경우, AI 시스템 연계 및 Agent 간 호환 여부 등을 검토하였는가?	☐	☐
	[3]	하나의 Agent가 수행하는 업무 혹은 의사결정이 다른 Agent의 판단이나 결과에 영향을 미치지 않도록 설계하였는가?	☐	☐
	[4]	여러 Agent가 함께 작업할 때를 대비하여 우선순위, 승인 절차, 충돌 처리 기준 등을 사전에 정의하고 이를 문서화하였는가?	☐	☐
Agentic AI 의인화 제한 정책	[1]	Agentic AI 설계 시 업무 수행 목적과 무관한 불필요한 인간적 특성 (예. 성격, 감정적 표현 등)을 과도하게 부여하지 않도록 설계하였는가?	☐	☐
	[2]	이용자가 Agentic AI를 사람처럼 인식하거나 감정적 애착을 형성하지 않도록, Agentic AI의 감정적 의인화를 제한하는 정책을 마련하였는가?	☐	☐

2.2.2 Agentic AI 운영 및 활용

점검항목		요구사항 및 체크리스트	Y	N
Agentic AI 위험관리	[1]	동일 작업을 여러 Agent가 수행할 경우, 중복 실행이나 과도한 리소스 사용을 방지하는 통제 기준을 마련하였는가?	☐	☐
	[2]	모든 Agent의 로그를 통합적으로 수집 · 관리하여, Agent 간 상호작용, 의사결정 과정, 출력 결과를 지속적으로 모니터링하고 있는가?	☐	☐
	[3]	Agentic AI 행동 범위를 제한하는 정책(예. 승인 없는 API 호출 금지, 개인정보 처리 제한 등)을 수립하였는가?	☐	☐
	[4]	Agentic AI가 스스로 처리할 수 없는 요청이나 예기치 못한 상황이 생겼을 때, 담당자가 대신 업무를 이어받아 처리할 수 있도록 업무 전환 절차와 대응 프로세스를 수립하였는가?	☐	☐
	[5]	Agentic AI가 요청을 처리할 수 없을 경우, 이용자에게 사유와 후속 조치(예. 대체 시스템 연결 등)을 제공하는 절차가 마련되어 있는가?	☐	☐
Agentic AI 변화관리	[1]	Agentic AI 도입 시 법적 · 윤리적 기준(예. 노동법, 저작권법 등) 상 문제 소지가 없는지 법률 검토를 받았는가?	☐	☐
	[2]	Agentic AI 도입에 따른 임직원의 우려와 피드백을 수집 · 반영할 수 있는 공식 소통 채널을 마련하였는가?	☐	☐
	[3]	Agentic AI와 사람이 협업하는 방식에 대해 임직원에게 충분한 설명과 안내를 제공하였는가?	☐	☐
	[4]	Agentic AI 도입으로 업무 축소 및 전환 대상자 대상 재교육 또는 직무 전환 프로그램을 제공하고 있는가?	☐	☐

2.3. Physical AI

2.3.1 Physical AI 기획

점검항목		요구사항 및 체크리스트	Y	N
Physical AI 업무 · 환경 정의	[1]	사람이 직접 수행하기 어려운 작업(예. 고위험 작업 등)에 우선 적용할 수 있도록 Physical AI 도입이 필요한 업무 영역을 식별 및 정의하였는가?	☐	☐
	[2]	Physical AI 시스템이 적용되는 산업 현장의 특수성(예. 자재 특성, 작업 절차 등)을 충분히 이해하고 설계하였는가?	☐	☐
	[3]	Physical AI의 모델 학습 및 동작 시나리오 설계 시 산업별 안전 기준, 설비 구조, 환경 조건을 고려하였는가?	☐	☐
	[4]	Physical AI 시스템이 업무를 수행하는 환경(예. 공장 구조, 온도 등)을 명확하게 정의하여 문서화하였는가?	☐	☐
	[5]	해양산업 전문가와 실무진들이 Physical AI의 기획 · 개발 단계에 참여하여 기획 및 설계 적정성을 검증하였는가?	☐	☐
Physical AI 데이터 관리	[1]	Physical AI 시스템이 주변 환경을 실시간으로 정확하게 인지할 수 있도록 양적 · 질적으로 충분한 데이터를 수집 및 활용하고 있는가?	☐	☐
	[2]	영상 · 음성 · 위치 등 실시간 데이터 수집 시 작업자에게 사전 고지 및 동의를 받는 절차가 마련되어 있는가?	☐	☐
	[3]	개인영상정보, 국가핵심기술 등 민감정보에 대해 암호화 · 비식별화 · 접근제한 등의 보안 조치를 적용하고 있는가?	☐	☐

2.3.2 Physical AI 운영 및 활용

점검항목		요구사항 및 체크리스트	Y	N
Physical AI 운영 및 활용	[1]	Physical AI를 실제 현장에 적용하기 전에, 가상 시뮬레이션이나 테스트베드 환경에서 주요 동작을 미리 검증하고 최적화하였는가?	☐	☐
	[2]	Physical AI 시스템의 물리적 작업공간과 안전구역을 구분하고, 운영 범위를 시각적으로 명확히 표시(예. 작동 구역 라인 표기, 접근 제한 구역 설정 등)하였는가?	☐	☐
	[3]	Physical AI 시스템이 사람 또는 다른 기계와 상호작용할 때, 충돌 · 낙상 등 물리적 사고를 예방할 수 있는 안전기능(예. 비상 정지, 충돌 회피 등)이 구현되어 있는가?	☐	☐
	[4]	데이터 누락 및 오인식 발생 시, Physical AI 시스템이 기본 기능을 유지할 수 있도록 이중 안전장치가 마련되어 있는가?	☐	☐
	[5]	Physical AI 시스템이 사람이나 시설물과 직접 접촉할 때 발생 가능한 리스크(예. 부상, 재산 피해, 운영 중단 등)에 대비하여 엄격한 안전 규정과 대응 메뉴얼을 마련하였는가?	☐	☐
Physical AI 법적 보호	[1]	Physical AI 관련 사고 발생 시, 책임 주체(예. 운영사 · 제조사 · 개발사 등) 를 명확히 구분할 수 있는 관리체계를 마련하였는가?	☐	☐
	[2]	Physical AI 모델 · 센서 · 하드웨어 등 공급망(예. 협력사, 외주 개발사 등)에 대한 책임 · 품질 · 보안 기준을 계약에 명시하고 있는가?	☐	☐
	[3]	사고 원인 분석 및 재발 방지를 위해, Physical AI 시스템의 작동 로그, 오류 이력, 유지보수 내역 등을 체계적으로 기록 · 보존하고 있는가?	☐	☐
	[4]	Physical AI 시스템 도입 · 변경 · 업데이트 시, 법적 · 안전 리스크에 대한 사전 검토 및 내부 승인 절차를 수립 및 이행하고 있는가?	☐	☐
	[5]	Physical AI 시스템 사고 발생에 대비해 보험, 면책, 법적 대응 프로세스를 사전에 마련하고 정기적으로 검토하고 있는가?	☐	☐

Chapter 03. 서비스 도입 체크리스트

점검항목		요구사항 및 체크리스트	Y	N
Physical AI 변화관리	[1]	Physical AI 도입 이후 사람과 AI 간 역할 분담을 명확히 정의하고, 임직원에게 충분한 안내와 설명을 제공하였는가?	☐	☐
	[2]	Physical AI 도입 과정에서 현장 직원의 의견을 수렴하고 피드백을 반영하는 절차를 운영하고 있는가?	☐	☐
	[3]	Physical AI 도입으로 업무 축소 및 전환 대상자 대상 재교육 또는 직무 전환 프로그램을 제공하고 있는가?	☐	☐
	[4]	Physical AI 도입 이후 기존 직원의 업무 피로도, 안전성, 만족도 등 인적 영향 지표를 정기 점검하기 위한 체계를 마련하였는가?	☐	☐

안전한 AI 도입 및 활용을 위한
해양산업 인공지능(AI) 가이드라인

부록.
용어집

한국해양진흥공사
KOREA OCEAN BUSINESS CORPORATION

| 용어집 |

용어명	정의
가상 시뮬레이션	모델에 의해 의도되는 명령들을 정확하게 수행하고 있는가를 검증하는 작업
고영향 AI	사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능시스템
네트워크 (Network)	지리적으로 떨어져 있는 다른 위치에 있는 장치(전화기, 팩스, 컴퓨터, 단말기 등) 간에 정보를 교환할 수 있도록 이들 장치를 상호 접속하기 위하여 사용되는 전기 통신 기기와 장치, 전송로의 결합
대규모 언어 모델 (LLM, Large Language Model)	대규모의 텍스트 데이터를 활용하여 학습한 파라미터의 수가 매우 큰 모델로, 자연어 이해와 생성 작업에 탁월한 성능을 보이는 심층 신경망(deep neural network) 모델
대리 변수 (Proxy Variable)	어떤 특정한 변수에 대해 직접적으로 획득이 곤란하거나 사용이 어려운 경우, 혹은 반영이 제대로 이루어지지 않는 경우에 원래 변수 대신하여 사용되는 변수
데이터 생애주기	데이터 입력에서 데이터 폐기에 이르기까지의 일련의 과정
데이터 암호화	데이터베이스 보안을 위한 하나의 기법으로서, 데이터의 실제 내용을 허가받지 않은 사람이 볼 수 없도록 은폐하기 위해 데이터를 암호로 바꾸는 것
딥러닝, 심층 기계 학습 (DL, Deep Learning)	일반적인 기계 학습 모델보다 더 깊은 신경망 계층 구조를 이용하는 기계 학습으로, 여러 개의 은닉층(hidden layer)으로 구성된 인공 신경망을 활용
라벨링 (labeling)	기계학습을 위해 태그(레이블)가 지정된 데이터 세트를 준비하는 과정
망분리	중앙행정기관과 공기업 등의 내부 정보 유출을 막고, 컴퓨팅 악성 코드 따위가 내부망에 침투하지 못하게 물리적으로 차단하려는 목적으로 인터넷망과 완전히 분리된 환경에서 업무를 보도록 인터넷망과 업무망을 망(네트워크)분리 하는 것
머신러닝, 기계 학습 (ML, Machine Learning)	인공지능의 한 분야로 데이터의 집합을 컴퓨터에 학습시킨 후, 새로운 데이터에 대한 질문에 컴퓨터가 대답할 수 있게 만드는 작업
메타데이터	다른 데이터를 정의하고 기술하는 데이터 또는 다양한 형식의 다른 데이터의 내용 또는 구조를 설명하는 데이터로, ISO/IEC 11179(Information Technology – Metadata registries (MDR)) 표준에 따르면 메타데이터는 데이터 그 자체는 아니지만, “다른 데이터를 정의하고 기술하는 데이터(data that defines and describes other data)”라고 정의
모델링	물리적 환경으로 수집된 데이터를 바탕으로 입력값과 출력값 간의 상호 인과관계에 대한 컴퓨터 연산 프로세스 혹은 함수 형태로 표현한 모형
민감정보	개인정보처리자는 사상 · 신념, 노동조합 · 정당의 가입 · 탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 정보주체의 사생활을 현저히 침해할 우려가 있는 개인정보로서 대통령령으로 정하는 정보

| 용어집 |

용어명	정의
비식별	개인정보의 일부 또는 전체를 삭제 · 변환 · 대체 등의 방법으로 처리하여, 특정 개인을 직접 또는 간접적으로 식별할 수 없도록 만드는 절차
사용자 경험 (UX, User Experience)	사용자가 어떤 시스템, 제품 혹은 서비스를 직 · 간접적으로 이용하면서 느끼고 생각하게 되는 총체적 경험으로 단순히 기능이나 절차상의 만족뿐 아니라 전반적으로 지각 가능한 모든 면에서 사용자가 참여, 사용, 관찰하고 상호 교감을 통해서 알 수 있는 가치 있는 경험
사용자 오남용	사용자가 의도적으로 또는 부주의로 AI 시스템의 설계된 목적이나 윤리적 지침에 반하는 방식으로 AI 시스템을 사용하여 해를 끼치거나 부당한 이득을 얻는 행위
사용자 인터페이스 (UI, User Interface)	사용자와 컴퓨터 시스템의 하드웨어나 소프트웨어 구성요소 사이에 정보교환을 가능하게 하는 인터페이스
사후학습된 모델	파운데이션 모델에 추가적인 데이터를 대량 학습시켜 성능을 고도화한 모델
생성형 AI (GenAI, Generative AI)	사용자의 명령에 대응하여 텍스트, 이미지, 동영상, 기타 미디어를 생성할 수 있는 인공지능
설명 가능성 (Explainability)	AI 시스템이 왜 이런 결정을 내리게 되었는지 설명할 수 있는 능력
소프트웨어 (S/W, Software)	컴퓨터를 동작시키고 컴퓨터에 어떤 일을 처리할 순서와 방법을 지시하는 명령어의 집합인 프로그램과 프로그램의 수행에 필요한 절차, 규칙, 관련 문서 등의 총칭
심층 신경망 (DNN, Deep Neural Network)	입력층(input layer)과 출력층(output layer) 사이에 다중의 은닉층(hidden layer)을 포함하는 인공 신경망(ANN: Artificial Neural Network)으로 다양한 비선형적 관계 학습 가능
안전모드	수행할 어셈블리를 컴파일 시 지정한 관련 어셈블리의 해당 버전과만 연동되도록 제한하는 버전 정책의 일종
알고리즘 (Algorithm)	유한한 단계 내에 주어진 문제의 답을 얻기 위해 잘 정의한 명령어의 유한한 집합 또는 특정한 일을 수행하기 위해 연산의 순서를 정의한 규칙의 유한한 집합
어플리케이션 (Application)	어떤 종류의 작업 수행을 돕기 위해 설계된 컴퓨터 프로그램
엣지 케이스 테스팅 (Edge case testing)	극단적이고 비정상적인 입력값, 조건, 상황에 AI가 어떻게 반응하는지 테스트하는 절차
오픈소스 (Open Source)	소프트웨어의 설계도에 해당하는 소스 코드를 인터넷 등을 통하여 무상으로 공개하여 누구나 그 소프트웨어를 개량하고, 이것을 재배포할 수 있도록 하는 소프트웨어
워터마크	어떤 파일에 관한 저작권 정보 등을 식별할 수 있도록 디지털 이미지나 오디오 및 비디오 파일에 삽입한 비트 패턴

| 용어집 |

용어명	정의
원시 데이터 (source data)	자료가 발생하거나 생성하는 것이 개인이나 조직에 의해 만들어지는 데이터
이상값 (outlier)	전체 데이터 분포에서 현저하게 벗어난 값
인공지능 (AI, Artificial Intelligence)	하나 이상의 주어진 작업을 수행하기 위해 모델 형태로 보유한 지식을 획득, 처리, 생성 및 적용하는 능력
인공지능 모델 (AI Model, Artificial Intelligence Model)	AI 시스템을 구성하는 필수 요소로 AI 시스템 자체를 구성하지는 않으며, 대규모 언어 모델(LLM)과 또는 알고리즘 같은 모델
인공지능 서비스 (AI Service, Artificial Intelligence Service)	인공지능을 적용한 컴퓨터 시스템에 의해 인간의 지능을 모사하여 이를 이용하고자 하는 고객이나 사용자의 요구에 의해 제공되는 모든 서비스
인공지능 시스템 (AI System, Artificial Intelligence System)	자율성을 가지고 작동할 수 있도록 설계된 시스템으로, 적응력을 발휘할 수 있으며, 입력값을 받아 명시적 또는 암묵적 추론을 통해 환경에 영향을 미칠 수 있는 콘텐츠 · 예측 · 추천 · 결정과 같은 산출물을 생산하는 소프트웨어
입력 변수	시스템 내의 다른 변수들과는 무관하면서 외부로부터 시스템에 작용하는 변수
자율운항선박	자율화시스템을 통해 의사결정을 지원하고 선박의 제어 및 관리의 전체 또는 일부를 자율화시스템이 대신할 수 있는 선박으로 유인, 무인 또는 원격으로 운항
재현성	동일한 코드, 데이터, 구성 및 매개변수를 사용하여 동일한 조건에서 시스템의 동작이나 출력을 일관되게 재현할 수 있는 능력
저작물	인간의 사상 또는 감정을 표현한 창작물(소설 · 시 · 논문 · 강연 등 어문 저작물, 음악 저작물, 연극 및 무용 · 무언극 등을 포함하는 연극 저작물, 회화 · 서예 · 도안 · 조각 · 공예 · 응용미술 작품과 그 밖의 미술 저작물, 건축물 · 건축을 위한 모형 및 설계도서를 포함하는 건축 저작물, 영상 저작물, 지도 · 도표 · 설계도 · 약도 · 모형 및 그 밖의 도형 저작물, 컴퓨터 프로그램 저작물 등)과 원저작물을 번역 · 편곡 · 각색 · 영화 제작 및 그 밖의 방법으로 작성한 '2차적 저작물' 및 편집물로서 그 소재의 선택 또는 배열에 창작성이 있는 편집 저작물
적대적 공격	인공지능 시스템이 의도하지 않은 잘못된 판단이나 동작을 수행하도록 입력 데이터, 학습 데이터 또는 모델 자체를 의도적으로 조작하는 공격 기법
전처리 작업	원자료(raw data)를 데이터 분석이나 기계 학습에 적합하도록 가공하는 과정
정제	원자료(raw data)의 오류, 중복, 이상치, 불일치, 결측값 등을 식별하고 제거하거나 보정하여 분석 및 학습에 적합한 정확하고 신뢰할 수 있는 형태로 만드는 품질 확보 기술

| 용어집 |

용어명	정의
증강	기계학습의 성능을 개선하기 위해 기존 데이터에서 새로운 데이터를 생성하는 작업
차별적 데이터	차별을 금지하는 성별 · 종교 · 학력 · 장애 · 사회적 신분 · 국적 등의 데이터
추적가능성	사용자가 예측 및 프로세스를 추적할 수 있는지 여부를 나타내는 AI의 속성
파라미터, 매개변수 (parameter)	프로그램을 위해서 기술한 고정값과 의미 있게 주어진 변수로, 함수와 같은 수학적 개체의 특성이나 출력에 영향을 주는 변수
파운데이션 모델	광범위한 정보를 포함하는 대규모 데이터가 미리 학습되어 다양한 분야에 기본적인 틀처럼 적용할 수 있는 대규모 인공지능 모델
파인튜닝된 모델	특정 작업이나 도메인에 특화된 데이터셋을 추가로 훈련시켜 미세 조정한 모델
편향 (bias)	가용한 데이터가 모집단이나 연구 현상을 적절히 표현하지 못하여 데이터셋의 특정 요소가 과장되거나 축소되어 표현될 때 발생하는 오류
하드웨어 (H/W, Hardware)	컴퓨터를 비롯한 시스템의 물리적 구성품으로, 컴퓨터 프로그램, 절차, 규칙, 관련문서 등의 소프트웨어에 대응하는 용어 (컴퓨터 하드웨어는 중앙 처리 장치(CPU: Central Processing Unit), 기억 장치(memory device), 입력 장치(input device), 출력 장치(output device)로 구분)
학습 데이터	인공지능 모델을 학습시키는 데 사용되는 데이터(세트)를 지칭 (인공지능 알고리즘과 모델이 패턴을 인식하고, 예측하거나, 특정 작업을 수행하도록 가르치는 데 기초로 작용)
환각 (hallucination)	인공지능 학습 모델이 실재하지 않거나 확인되지 않은 내용을 사실인 것처럼 제시하는 현상
AI 거버넌스	조직의 AI 시스템이 전략 · 윤리 · 법적 기준에 맞게 운영되도록 관리하는 전사적 운영체계
AI 리스크	인공지능 시스템의 설계, 개발, 배포 및 사용 과정에서 AI가 개인과 기업에 미칠 잠재적인 부정적 결과 또는 손해 발생 가능성으로 정의되며, 위험 발생 형태에 따라 수행 리스크, 보안 리스크, 통제 리스크, 경제적 리스크, 사회적 리스크, 윤리적 리스크 등으로 분류
AI 리터러시	개인과 조직이 AI 시스템과 도구를 비판적으로 이해하고 평가하며, 이를 도구로써 안전하고 윤리적으로 활용할 수 있는 주체적 능력
AI 생명주기	인공지능 시스템의 기획 · 개발부터 운영 · 폐기까지 전 과정에 걸쳐 수행되는 단계별 활동과 관리 절차
AI 영향평가	인공지능 시스템의 개발 및 배포 전반에 걸쳐 해당 시스템이 개인, 사회, 경제, 문화, 윤리, 법률 등에 미칠 수 있는 잠재적 영향과 위험을 사전에 식별, 분석 및 완화하기 위한 체계적인 평가 절차

| 용어집 |

용어명	정의
AI 윤리원칙	인공지능의 개발과 활용 전 과정에서 인간의 존엄성과 권리를 보호하고, 공정성, 책임성, 투명성 등 윤리적 가치를 실현하기 위한 원칙과 기준
Agentic AI	명시적 지시 없이도 스스로 목표를 설정하고, 계획을 수립해 다양한 도구와 시스템을 활용하여 실제 작업을 수행할 수 있는 고도의 자율적 인공지능 기술로, 단일 명령에 따라 정해진 작업만 수행하는 기존 인공지능과 달리, 스스로 문제를 인식하고 목표를 설정한 뒤, 계획을 수립하고 적절한 도구나 다른 AI Agent와 연동하여 작업을 수행하는 고도의 자율성과 능동적 구조를 갖춘 인공지능 기술
AI Agent	사용자의 의도를 이해하고, 자율적으로 계획을 수립한 뒤, 외부 도구 및 시스템과 상호작용하여 실제 작업을 수행하는 실행 중심의 인공지능 시스템으로, 단순히 질문에 답하거나 정보를 제공하는 챗봇 형태의 인공지능 시스템을 넘어서 실제 컴퓨터 시스템에서 활용 가능한 도구와 인터페이스를 능동적으로 호출하여 행동을 수행할 수 있도록 설계된 고도화된 인공지능 시스템
AIOps (Artificial Intelligence for IT Operations)	머신러닝 및 자연어 처리(NLP)와 같은 인공지능 기술을 사용하여 IT 시스템 관리 방식을 자동화하고 개선하는 기술 플랫폼 및 접근 방식
AIS (Automatic Identification System)	선박 자동 식별 시스템 (선박의 위치, 침로, 속력 등 항해 정보를 실시간으로 제공하는 첨단 장치)
AMR (Autonomous Mobile Robots)	자율 이동 로봇 (자율적으로 작동하며 고정된 경로나 트랙 없이도 통제되지 않은 환경에서 탐색할 수 있는 로봇)
API (Application Programming Interface)	애플리케이션 계층과 플랫폼 시스템 계층 사이에 존재하며 플랫폼상에서 실행하는 애플리케이션 개발을 용이하게 하는 소프트웨어 라이브러리 집합
ETA (Estimated Time of Arrival)	도착예정시간
KPI (Key Performance Indicator)	핵심 성과 지표 (기업이 목표를 달성하기 위해서 핵심적으로 관리해야 하는 요소들에 대한 성과지표)
LLMOps (Large Language Model Operations)	전체 수명 주기 동안 AI 모델의 개발, 배포 및 관리를 가속화하는 전문적 사례 및 워크플로
MLOps (Machine Learning Operations)	머신 러닝 모델을 구축하고 실행하기 위해 어셈블리 라인을 만들도록 설계된 일련의 과정

| 용어집 |

용어명	정의
Physical AI	센서와 액추에이터 등 물리 장치와 결합되어, 실제 환경에서 자율적으로 인식, 판단, 행동을 수행하는 인공지능 기술로 텍스트, 이미지, 음성 등 디지털 정보를 다루는 기존 인공지능과 달리, 물리적 실체를 갖고 환경과 능동적으로 상호작용하는 지능체 현실 세계의 물리 법칙과 공간 구조, 환경을 학습에 반영해 실제 환경에서 발생하는 변화를 이해하고, 다음 상황을 추론하거나 상호작용을 수행할 수 있는 인공지능
PoC, Proof of Concept	제품, 기술, 정보 시스템 등이 조직의 특수 문제 해결을 실현할 수 있다는 증명 과정으로, 아직 시장에 나오지 않은 신제품 및 신규 서비스 등에 대한 사전 검증을 하는 작업
RAG 기반 모델	기업 내부 데이터베이스 등 외부 지식 기반과 연결하여 성능을 최적화시킨 모델
SaaS, Software as a Service	인터넷환경에서 사용자가 원하는 소프트웨어를 서비스 형태로 제공하는 서비스로, 공급 업체가 하나의 플랫폼을 이용해 다수의 고객에게 소프트웨어 서비스를 제공하고, 사용자는 이용한 만큼 돈을 지불하는 방식으로 유통

부록.

참고문헌

kobc 한국해양진흥공사
KOREA OCEAN BUSINESS CORPORATION

| 참고문헌 |

공통

- 금융위원회, “ 금융분야 AI 운영 가이드라인 ”, 2021.07
- 금융위원회, “ 금융분야 AI 개발 · 활용 안내서 ”, 2022. 08
- 금과학기술정보통신부 · 한국정보통신기술협회, “ 2024 신뢰할 수 있는 인공지능 개발 안내서 ”, 2024.02
- 해양금융센터, “ 글로벌 해운업 동향파악 및 국내 해운업 현황보고 ”, 2024.09
- 디지털플랫폼정부위원회, “ 공공부문 초거대 AI 도입 · 활용 가이드라인 2.0 ”, 2025.04
- 한국인터넷진흥원, “ 인터넷·정보보호 법제동향 ”, 2025.06
- 과학기술정보통신부, “ AI 기본법 하위법령집 ” , 2025.09
- Gartner, “ Hype Cycle for AI and Cybersecurity ”, 2025.08
  Available: Hype Cycle for AI and Cybersecurity, 2025

AI 윤리

- 관계부처 합동, “ 사람이 중심이 되는 AI 윤리기준 ”, 2020.12
- 과학기술정보통신부 · 정보통신정책연구원, “ 2025 인공지능 윤리기준 실천을 위한 자율점검표(안) ”, 2025.02
- 네이버, “ Naver Integrated Report 2023 ”, 2024.07
  Available: Sustainability Reports I NAVER Corp.
- LG AI 연구원, “ 2024 LG AI 윤리 책무성 보고서 ”, 2025.02
  Available: LG AI Research About
- 웨이크업, “ AI 리터러시, 인공지능 시대의 필수 생존 역량 ”, 2025.08
  Available:https://www.wakeupnews.co.kr/news/articleView.
- 국제인공지능윤리협회, “ 감정교류 AI 윤리 가이드라인 ”, 2025.09
- 행정안전부, “ 공공 AI 어떻게 사용해야 할까? '공공부문 인공지능 윤리원칙' 마련 추진 ”, 2025.11

| 참고문헌 |

AI 보안 / 리스크

- 금융위원회, “ 금융분야 AI 보안 가이드라인 ”, 2023.04
- 한국지능정보사회진흥원, “ 美 NIST「AI 위험관리 프레임워크(AI RMF) 1.0」분석 및 시사점 ”, 2023.07
  Available: 국회도서관 국가전략정보포털
- 방송통신위원회, “ 방통위, ‘인공지능(AI) 서비스 이용자 피해 신고창구’ 개설 ”, 2024.12
- 법무부, “ 해외규제 모니터링 제5호 ”, 2024.10
  Available: https://www.moj.go.kr/bbs/moj/177/589371/artclView.do
- 방송통신위원회 · 정보통신정책연구원, “ 생성형 인공지능 서비스 이용자 보호 가이드라인 ”, 2025.02
- 국가전략정보포털, “ 딥시크, 한국 이용자 개인정보 · 입력어 해외 무단 이전 ”, 2025.04
- 카카오뱅크, “ 대화형 AI 서비스 이용약관 ”, 2025.05
  Available:https://www.kakaobank.com/Help/Documents/ProductDeclaration/History/view/16644
- 정보통신정책연구원, “ 미국과 한국의 AI 채용 분야 정책 현황 ”, 2025.05
- 국가정보원 · 국가보안기술연구소, “ 국가 망 보안체계 보안 가이드라인 ”, 2025.09
- ISO/IEC 23894,“Artificial Intelligence – Risk Management", 2023. 02
- ABA Banking Journal , “ Mass. AG reaches settlement with student loan firm for $2.5M over AI lending bias ”, 2025.08
  Available: Mass. AG reaches settlement with student loan firm for $2.5M over AI lending bias | ABA Banking Journal

| 참고문헌 |

데이터

- 개인정보보호위원회, “ 인공지능(AI) 개발·서비스를 위한 공개된 개인정보 처리 안내 ”, 2024.07
- 과학기술정보통신부 · 한국지능정보사회진흥원, “ 2024 빅데이터 플랫폼 & 센터 ”, 2024.11
- 과학기술정보통신부 · 한국지능정보사회진흥원 · 한국정보통신기술협회, “ AI 데이터 품질관리 가이드 ”, 2025.02
- 한국저작권위원회, “ 미국 생성형 AI를 활용한 이미지의 저작권 등록 사례 ”, 2025.02
- 해양수산과학기술진흥원, “ 해양수산과학기술정책 · 기술동향 ”, 2025.05
  Available: https://www.kimst.re.kr/u/data/magazine_01/board.do
- 한국저작권위원회, “ 생성형 인공지능 활용 저작물의 저작권 등록 안내서 ”, 2025.06
- 한국저작권위원회, “ 생성형 인공지능 결과물에 의한 저작권 분쟁 예방 안내서 ”, 2025.06
- 법무법인 세종, “ AI 개발을 위한 저작물 학습과 공정이용에 관한 최근 미국 판례 동향 ”, 2025.07
  Available: https://www.shinkim.com/kor/media/newsletter/2894
- 개인정보보호위원회, “ 생성형 인공지능(AI) 개발·활용을 위한 개인정보 처리 안내서 ”, 2025.08
- 과학기술정보통신부 · 한국지능정보사회진흥원, “국가데이터 통합 연계를 위한 데이터 카탈로그 표준 가이드 v1.0 ”, 2025.10
- 한국지능정보사회진흥원, “ 한국형 데이터 스페이스(K-Data Space) – AX 시대, 데이터 공유 · 활용 패러다임 전환 전략 ”, 2025.10

| 참고문헌 |

AI 모델 / 시스템

- 과학기술정보통신부·정보통신산업진흥원, “소프트웨어사업 요구사항 분석·적용 가이드 ”, 2021.01
- 과학기술정보통신부· 정보통신산업진흥원, “ 기업 공개소프트웨어 거버넌스 가이드 ”, 2021. 11.
- 한국지능정보사회진흥원 · 업스테이지, “ Open-ko LLM 리더보드 ”, 2024.08
- 중앙일보, “ 젠슨 황 ‘다음 먹거리는 피지컬 AI’ 삼성·LG·SK ‘우리도 준비’[CES 2025] ”, 2025.01
  Available: https://www.joongang.co.kr/article/25306533
- 싱가포르 해양항만청, “ 300 Maritime Workers to Build Digital and Technical Capabilities under Enhanced Career Conversion Programme by WSG and MPA ”, 2025.03
  Available: https://www.mpa.gov.sg/media-centre/details/joint-media-release
- 소프트웨어정책연구소, “ 피지컬 AI의 현황과 시사점 ”, 2025.05
- 삼성SDS, “ Agentic AI란 무엇인가? – 뛰는 AI 에이전트, 나는 Agentic AI의 시대 ”, 2025.06
  Available: https://www.samsungsds.com/kr/insights
- TheGuardian, “ ChatGPT offered bomb recipes and hacking tips during safety tests ”, 2025.08
  Available: https://www.theguardian.com/technology
- NVIDIA, “ 자율주행 자동차를 위한 DRIVE 인프라 ”
  Available: https://www.nvidia.com/ko-kr/self-driving-cars
- Korea Business Review, “ 쿠팡 해지 절차 논란, 법 · 심리 · UX로 본 플랫폼 이탈 장벽의 구조 ”, 2025.12
  Available: https://www.koreabizreview.com/detail

안전한 AI 도입 및 활용을 위한
해양산업 인공지능(AI) 가이드라인

kobc 한국해양진흥공사
KOREA OCEAN BUSINESS CORPORATION
부산광역시 해운대구 마린시티2로 38, C1동 7층 | Tel. 051-795-1500 | www.kobc.or.kr