공공부문 인공지능 윤리기준

2026. 10.

행 정 안 전 부
(공공인공지능혁신과)

목 차

Ⅰ. 추진배경 및 경과 ····· 1

1 추진 배경 ····· 1

2 추진 경과 ····· 2

Ⅱ. 윤리기준 개요 ····· 3

1 개요 ····· 3

2 기본 체계 ····· 3

3 기준 구성 ····· 4

4 이행 절차 ····· 4

Ⅲ. 세부 내용 ····· 5

1 윤리 기준 ····· 5

2 점검 항목 ····· 10

《붙임》 인공지능 윤리기준 세부 자율 점검사항 ····· 12

공공부문 인공지능 윤리기준

Ⅰ. 추진배경 및 경과

1 추진 배경

○ **공공부문 AI 도입**을 통한 **대국민 서비스**와 **업무방식**의 **혁신**과 함께 **신뢰기반 구축** 필요

- 정부 효율성과 생산성 향상을 위한 **공공부문**의 **AI 전환**과 함께 **국민의 기본권 침해** 및 **다양한 위험요소**에 대한 사전 **대비체계** 마련

> ■ **【이재명 정부** 123대 **국정과제】** 세계 최고 AI 민주정부 실현-**공공 AI 신뢰기반**-공무원의 **AI 윤리 내재화**

○ **세계적**으로 **윤리 문제**에 **적극 대응**하고 있고, **일본 총무성(행안부**와 유사)의 경우 **AI 활용원칙**('18.7월) 등을 통해 **윤리적 과제** 검토

※【국외 사례】 EU: AI 윤리 가이드라인('19.4.), OECD: AI 원칙('19.5.), 영국: 공공부분 인공지능 활용 가이드라인('19.6.), AI Playbook('25.2.), UNESCO: AI 윤리 권고('21.11.), 미국: AI 행정명령('23.10.)

- 우리나라는 인공지능 정책, 개인정보 보호 등 AI 관련 소관 부처를 중심으로 **업무 특성**에 맞는 **윤리원칙, 가이드라인** 등을 운영하고 있음

○ **인공지능**의 **급속한 발전**으로 법이 따라가지 못하는 문제가 발생할 수 있으므로 사후 규제보다 **예방적·자율적 윤리원칙** 중요

- **공공부문**의 **AI 활용**은 **국민**에게 **직접적인 영향**을 미치므로 민간과 차별화된 **윤리기준 제시**로 **공공**의 **가치**를 **지향**

- **인공지능기본법**의 **제정**('26.1.22. 시행) 및 **인공지능데이터행정법 개정**('26.8.28. 시행)에 따라 **공공부문 인공지능의 윤리적 활용을 위한 자율적 기준 제정 필요**

> **※ 인공지능 및 데이터 기반 행정활성화에 관한 법률**
> 제31조(**인공지능 활용 윤리기준**) ① **행정안전부장관은 공공기관이 안전하고 효율적으로 인공지능을 도입 및 활용할 수 있도록 공정성·투명성·책임성·안전성 등의 신뢰 기반 조성에 필요한 사항**과 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」 제27조에 따른 인공지능 윤리원칙을 반영하여 **공공기관에 적용되는 윤리기준을 제정하고 이를 공표**하여야 한다.

2 추진 경과

◇ AI 윤리 분야의 **전문가 중심**으로 **워킹그룹·자문회의**를 통해 **윤리기준** 초안 **작성**, 관계기관 의견수렴 및 전문가 검토를 통한 **최종안 확정**

○ (**적용대상**) 중앙부처 및 지자체, 공공기관으로 설정
○ (**접근방법**) 공공부문의 특성에 맞는 **핵심가치** 반영
○ **(주요관점)** 국민, 행정, 기술 관점에서 접근
○ (**실행방안**) 실무 중심의 **점검항목**(단계별 이행)으로 구성
○ (**국제표준**) ISO/IEC 42001 핵심요소 중 **국민과 관련있는 사항** 검토
○ (**구성체계**) 행정혁신·국민신뢰를 목표로 **윤리기준**의 용어를 **구체화**
○ (**사례반영**) 정부, 민간, 해외의 **윤리원칙 사례**를 참고

〈 **AI 윤리기준 주요 제정사례**(중앙부처) 〉

제정시기		부처	주요내용
'20년	12월	과기부	• **사람**이 중심이 되는 인공지능 윤리기준 * 3대 원칙(인간 존엄성, 사회 공공선, 기술 합목적성), 10대 요건(인권보장 등)
'21년	5월	개보위	• AI 관련 **개인정보보호** 6대 원칙 * 6대 원칙(적법성, 안전성, 투명성, 참여성, 책임성, 공정성)
	6월	방통위	• 인공지능 기반 **미디어** 추천 서비스 이용자 보호 기본원칙 * 3개 핵심원칙(투명성, 공정성, 책무성), 5개 실행원칙(이용자를 위한 정보공개 등)
	7월	금융위	• **금융분야** 인공지능 가이드라인 * 4개 핵심가치(금융산업의 책임성, AI 학습용 데이터의 정확성·안전성 확보 등)
'22년	5월	인권위	• 인공지능 개발과 활용에 관한 **인권** 가이드라인 * 6개 주요내용(인간의 존엄성 및 개인의 자율성과 다양성 보장 등)
	8월	교육부	• **교육분야** 인공지능 윤리원칙 * 대원칙(사람의 성장을 지원하는 인공지능), 10개 세부원칙(인간성장의 잠재성 등)
'23년	6월	산자부	• 인공지능 윤리 **국가표준**(KS) * 10개 고려항목(투명성, 공정성, 무해성, 책임성, 사생활 보장성, 편익성 등)
'25년	2월	과기부	• 2025 인공지능 윤리기준 실천을 위한 자율점검표(안)
'26년	7월	개보위	• 공공AX 프라이버시 보호 안내서
	8월	과기부	• 국가 인공지능 윤리원칙(전면 개정) *** 3대 가치(인간 존엄성, 사회 공공선, 인류의 지속가능성), 7대 원칙(인간중심성 등)**

Ⅱ. 윤리기준 개요

1 개요

○ (**정의**) 공공 AI 도입·활용에 적용하기 위한 **윤리적 가치**와 **행동 지침**

○ (**적용기관**) **중앙행정기관, 지방자치단체, 공공기관** 및 **지방공사·공단**

※ 인공지능 및 데이터 기반 행정 활성화에 관한 법률 제2조제5호

○ (**적용주체**) **인공지능 기술**을 **도입**하여 **활용**하고 **서비스**를 **제공**하는 **공공부문**의 **종사자**

2 기본 체계



○ **(1단계) 적용: AI 기초 활용**

- 인공지능을 검색·추천, 문서 작성, 분석결과 제공 등 행정 보조도구로 활용
- 도서 추천, 회의록 자동 요약, 다국어 번역 서비스 등

○ **(2단계) 응용: AI 의사결정 지원**

- 인공지능이 분석·예측·시뮬레이션을 통해 의사결정의 근거를 제공
- 수요예측 기반 의사결정 지원(위기 가구 예측), 신호체계 개선안 제시 등

○ **(3단계) 융합: AI 자율결정 기반**

- 사람의 관리·통제하에 인공지능이 실시간으로 판단·결정·행동까지 수행
- 자율주행 차량, 스마트 교통제어, 홍수 예·경보 시스템 등

3 기준 구성

목표	**공공부문 인공지능**을 통한 **행정혁신*** 촉진과 **국민신뢰**** 구축 ***혁신** → 신기술 도입, ****신뢰** → 윤리기준의 목적

윤리 기준	**국민**	**공공성**	공공 AI를 공공의 이익과 국민 삶의 질 향상을 위해 제공한다.
		형평성	공공 AI를 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 AI가 국민에게 피해를 주지 않도록 안전한 방식으로 운영한다.
		프라이버시 보호	공공 AI가 개인정보와 사생활 등을 침해하지 않도록 보호장치를 마련한다.

4 이행 절차 및 주체

준비	점검	조정	환류
AI 도입·활용 현황 분석	AI 윤리기준 점검항목 단계별 확인	AI 윤리기준 점검 결과 조정	AI 윤리기준 반영 사항 모니터링 및 점검 체계 개선
적용기관	적용기관	적용기관	적용기관·행정안전부

Ⅲ. 세부 내용

1 윤리기준

① 공공성 (Public Interest)

공공 AI를 **공공의 이익**과 **국민 삶의 질 향상**을 위해 제공한다.

○ AI의 도입 및 활용이 민주주의를 강화하고 사회적 가치를 창출할 수 있도록 정책을 수립하여 집행한다.

○ AI 서비스는 국민 삶의 질을 높이는 방향으로 지속적인 발전을 추구한다.

○ AI 서비스는 국민과 정부 간 소통을 강화하고 행정에 대한 신뢰를 높이도록 한다.

※ 참고 사항

■ **(취지)** 인공지능의 발전은 인류 전체의 안전과 삶의 질 향상에 기여하는 방향으로 이루어져야 하며, 이를 위해 국민의 참여와 국내외적 협력을 통한 공감대가 형성되어야 한다. 특히, 공공부문에서 활용되는 인공지능은 더욱더 특정 개인이나 집단의 이익이 아닌 사회 전체적인 공익 증진에 기여해야 함

■ **(관련 규정) 전자정부법 제3조** : 공무원 등은 담당업무를 전자적으로 처리할 때 국민의 편익을 우선적으로 고려해야 함 / **행정절차법 제52조제1항** : 행정청은 행정과정에서 국민의 의견을 적극적으로 청취하고 이를 반영하도록 노력해야 함

■ **(적용 예시)** 00부는 AI 기반 도시계획 수립 시스템을 도입하면서 정책 결정 과정에 국민의 의견이 충분히 반영될 수 있도록 온라인 참여 플랫폼을 운영하고, AI는 다양한 제안을 분석하여 정책 초안을 마련하여 최종 결정은 공청회를 통해 확정함

② 형평성 (Equity)

공공 AI를 **모든 국민**에게 **공정하고 차별 없이 제공**한다.

○ AI 서비스는 사회적 약자와 소외계층을 포함한 모두에게 혜택을 제공한다.

○ AI 서비스는 사회 구성원의 다양성을 보장하고, 사회적 차별이 발생하지 않도록 한다.

○ AI 서비스는 사회적 약자와 디지털 소외계층이 불편 없이 이용할 수 있도록 접근성과 사용 편의성을 보장한다.

※ 참고 사항

■ **(취지)** 공공부문은 모든 국민의 권리를 보호하고 평등하게 서비스를 제공해야 하므로 인공지능이 편향된 결정을 하지 않도록 관리하고, 사회적 약자와 소외계층을 보호하는 방향으로 운영되어야 함

■ **(관련 규정) 지능정보화기본법 제3조제4항** : 국가 등은 지능정보기술을 활용하거나 지능정보서비스를 이용할 때 사회의 모든 구성원에게 공정한 기회가 주어지도록 노력해야 함 / **전자정부법 제19조** : 행정기관 등은 경제적·지역적·신체적 또는 사회적 여건으로 국민이 전자정부서비스에 접근하거나 이를 활용하는데 어려움이 발생하지 않도록 해야 함

■ **(적용 예시)** 00시청은 AI를 이용해 청년 일자리 지원 대상을 선정하는 과정에서 특정 지역이나 학력 수준에 따라 차별이 발생할 수 있다는 문제를 발견하여 학습 데이터를 재점검하고 다양한 배경의 청년들이 공정하게 평가받을 수 있도록 기준을 개선함

③ 투명성 (Transparency)

공공 AI의 도입과 활용에 대한 **과정을 투명하게 공개**한다.

○ AI 도입 여부와 활용 과정을 모든 국민에게 투명하게 공개한다.

○ AI 서비스의 학습에 이용되는 데이터의 개요와 결과물이 작동하는 방식 등을 알기 쉽게 공개한다.

○ AI 서비스 결과에 대한 국민의 의견을 반영하여 개선한다.

※ 참고 사항

■ **(취지)** 공공부문의 의사결정은 그 과정이 투명하고 결과에 대해 설명할 수 있어야 하고, 인공지능이 의사결정에 영향을 주거나 대신 결정하는 경우 국민의 알권리 보장을 위해 필요한 정보를 제공할 수 있어야 함

■ **(관련 규정) 공공기관의 정보공개에 관한 법률 제3조, 제6조제1항** : 공공기관이 보유·관리하는 정보는 적극 공개해야 하고, 정보를 투명하게 공개하는 조직문화 형성에 노력해야 함 / **행정절차법 제5조제1항** : 행정청이 행하는 행정작용은 그 내용이 구체적이고 명확하여야 함

■ **(적용 예시)** 00군청은 AI를 이용해 복지 수급 대상자를 선정하는 시스템을 도입하면서 주민들에게 결과에 대한 이유를 알 수 있도록 설명 요청 기능을 함께 제공하여 판단 기준과 주요 데이터를 안내하고 이의신청 절차도 고지함

④ 책임성 (Accountability)

공공 AI에 대한 **책임 주체를 명확히** 정하고, **제도적 기준**에 따라 활용한다.

○ AI 서비스의 기획부터 제공까지 전 과정의 단계별 책임 주체를 명확히 지정한다.

○ AI 서비스 결과에 대해 인간의 검토와 개입이 가능하도록 인적 감독 체계와 사후관리 체계를 마련한다.

○ AI 서비스 결과가 관련 법령과 규정을 준수하도록 관리한다.

○ AI 서비스 결과에 대한 사회적 책임을 다한다.

※ 참고 사항

■ **(취지)** 인공지능 기술이 발전에 따라 자동화된 업무 처리와 서비스 제공이 예상되며 이에 따라 발생가능한 피해를 예방하고 피해 구제의 실효성을 확보하기 위해 책임 소재를 분명히 해야 함

■ **(관련 규정) 지능정보화기본법 제4조제3항** : 국가기관 등은 지능정보서비스를 제공·이용할 때 안전성·신뢰성 및 공정성 확보를 위하여 노력해야 함 / **국가배상법 제2조제1항** : 국가 등은 공무원이 고의 또는 과실로 법령을 위반하여 타인에게 손해를 입힌 경우 배상해야 함

■ **(적용 예시)** 00시청은 AI를 활용한 인허가 업무 과정에서 오류가 발생하자 책임 소재를 명확히 하기 위해 최종 결정은 담당 공무원이 검토하도록 절차를 강화하고, 오류 발생시 신속한 재심사와 피해 구제를 할 수 있는 제도를 마련함

⑤ 안전성 (Safety)

공공 AI가 **국민에게 피해를 주지 않도록 안전한 방식**으로 운영한다.

○ AI 시스템이 기술적 측면과 조직적 차원에서 안전성을 확보하도록 운영한다.

○ AI 시스템이 해킹 및 데이터 유출 등 외부 공격에 대응하도록 방안을 마련한다.

○ AI 기술의 오류 및 오작동에 대한 예방 조치를 마련하고, 재발 방지를 위해 안전대책을 적용한다.

※ 참고 사항

■ **(취지)** 공공 서비스는 잠시라도 중단될 경우, 국민의 일상 생활에 미치는 영향이 크므로 끊김이 없이 안정적으로 운영되어야 하고, 장애가 발생하더라도 신속히 복구될 수 있어야 함

■ **(관련 규정) 지능정보화기본법 제4조제3항** : 국가기관 등은 지능정보서비스를 제공·이용할 때 안전성·신뢰성 및 공정성 확보를 위하여 노력해야 함

■ **(적용 예시)** 00시청은 AI 기반 재난 예보 시스템을 도입했지만, 태풍 예보 중 시스템에 장애가 발생하여 경보발령이 지연되어, 백업 알고리즘과 수동 전환 체계를 마련하고 정기적인 시스템 점검 체계를 구축함

⑥ 프라이버시 보호 (Privacy)

공공 AI가 **개인정보와 사생활 등을 침해하지 않도록 보호장치를 마련**한다.

○ AI 시스템은 개인정보를 안전하게 저장·처리하고, 무단 접근과 유출 방지를 위해 조치를 마련한다.

○ AI 서비스는 설계 및 학습 과정에서 개인의 사생활이 보호될 수 있도록 관리적·기술적 조치를 적용한다.

○ AI 서비스에 대해 사용자 본인의 개인정보에 대한 열람, 정정 등의 권리를 행사할 수 있도록 한다.

※ 참고 사항

■ **(취지)** 인공지능은 각종 데이터를 학습하여 발전하지만, 개인정보 등 프라이버시 보호를 위해 데이터가 안전하게 관리되고 목적에 맞게 활용되어야 함

■ **(관련 규정) 개인정보보호법 제5조제1항** : 국가 등은 사생활 보호를 위한 시책을 강구해야 하고, 개인정보 보호지침을 정하여 준수하도록 함

■ **(적용 예시)** 00시청은 AI를 활용해 교통 흐름을 분석하기 위해 개인 식별 정보가 포함된 영상 데이터를 처리하면서 자동 익명화 기술을 적용하고, 분석 결과를 통계 형태로만 제공하도록 함

2 자율점검 항목 (총 68개)

* 항목별 세부 자율점검 사항 : 붙임

대항목	중항목	소항목
공공성 (8개)	공익 및 사회적 가치	공익 증진, 공공 서비스 기여, 공공 가치 추구, 삶의 질 향상
	국민 소통과 참여	소통 체계 구축, 국민 참여
	신뢰 구축	지속적 평가, 정책 반영
형평성 (12개)	공정한 혜택과 차별 금지	공정한 혜택, 사회적 격차 감소, 편향성 점검
	다양성과 포용성 보장	포용성 보장, 서비스 디자인, 지속적인 모니터링, 다양성 반영
	사회적 약자와 디지털 소외계층 보호	접근성, 편의성, 디지털 격차 해소
투명성 (12개)	정보공개	도입 목적 공개, AI 적용 여부 공개, 평가 결과의 공개, 개선 이행 과정 공개, 데이터 출처 공개
	설명 가능성 및 국민 이해 제고	설명 가능성, 서비스 한계 및 오류 가능성, 용어 일관성
	운영 개선 및 환류 체계 구축	외부 전문가 참여 구조, 이의제기 처리, 최신 정보 반영, 관련 근거 제시
책임성 (12개)	책임의 주체 설정과 분배	책임 주체 명확화, 책임 소재 추적, 법적 기준 준수, AI 결과물 합법성 검토, 데이터 합법적 수집·이용
	감사·감독 및 사후관리	외부 감사, 자동화 영향 검토, 인적 감독 체계, 사후 책임, 윤리 검토 체계
	사회적 책임 및 피해 대응	피해 구제, 피해 재발 방지
안전성 (15개)	보안 및 접근 통제	보안 체계, 데이터 접근 관리, 접근 권한 관리, AI 공격 대응 계획
	예방·점검·모니터링	시스템 안정성, 정기적 기술 점검 및 관리, 시험 및 검증, 잠재적인 위험 분석 및 대응, 주의 사항 안내
	위기·장애 대응 및 복구	위기 대응 및 복구 체계, 오남용 및 악용 방지 체계, 장애 감지 시스템, 사후 대책 마련, 오작동 대응, 비상 연락 체계 구축
프라이버시 보호 (8개)	개인정보 처리 적법성	개인정보 수집 제한, 사생활 침해 방지 평가, 목적 외 사용 제한
	관리적·기술적 조치	개인정보 안전 관리, 불필요한 개인 식별·추론 방지, 개인정보 파기
	정보주체 권리 보호	신고 및 구제 절차, 정보 주체 통제권 보장

붙임 **인공지능 윤리기준 세부 자율점검 사항**

◈ 분류기준 : 1단계-적용(AI 기초 활용), 2단계-응용(AI 의사결정 지원), 3단계-융합(AI 자율결정 기반)

대항목	중항목	소항목	점검사항	분류기준		
				1 (27개)	2 (57개)	3 (68개)
공공성 (8)	공익 및 사회적 가치	공익 증진	1. AI 서비스를 공익을 증진하는 방향으로 운영하고 있는가?	●	●	●
		공공 서비스 기여	2. AI 서비스가 공공 서비스의 효과성을 높이는 데 기여하고 있는가?	●	●	●
		공공 가치 추구	3. AI 서비스가 사회적 가치 창출에 기여하고 있는가?			●
		삶의 질 향상	4. AI 서비스가 국민 삶의 질을 향상시키는 데 기여하고 있는가?			●
	국민 소통과 참여	소통체계 구축	5. AI 서비스에 대하여 다양한 사용자 계층의 의견 수렴과 국민 신뢰 제고를 위한 소통 체계가 마련되어 있는가?	●	●	●
		국민 참여	6. AI 서비스의 도입에 따른 효과에 대해 일반 국민을 비롯한 주요 이해관계자들이 참여하고 논의할 수 있는 공식적이고 공개된 창구가 마련되어 있는가?			●
	신뢰 구축	지속적 평가	7. AI 서비스가 사회에 미치는 영향과 공익 기여도에 대해 정기적으로 모니터링하고 평가하고 있는가?		●	●
		정책 반영	8. AI 서비스의 운영 방향이 국가 정책 및 사회적 요구를 적절히 반영하고 있는가?		●	●
형평성 (12)	공정한 혜택과 차별 금지	공정한 혜택	9. AI 서비스가 모든 사용자에게 공정한 혜택을 제공하도록 설계되어 있는가?	●	●	●
		사회적 격차 감소	10. AI 서비스를 통해 사회적 격차를 감소시키기 위한 노력이 이루어지고 있는가?			●
		편향성 점검	11. AI 알고리즘과 데이터의 편향성을 정기적으로 점검하고 점검 결과에 따라 수정·보완하는 절차가 마련되어 있는가?	●	●	●
	다양성과 포용성 보장	포용성 보장	12. AI 서비스 개발·운영 과정에서 사회적 약자와 소외계층의 요구를 반영하고 있는가?		●	●
		서비스 디자인	13. AI 서비스에 문화적, 언어적 다양성을 고려한 디자인을 적용하고 있는가?	●	●	●
		지속적인 모니터링	14. AI 서비스에서 소외계층이 배제되지 않도록 지속적인 모니터링을 실시하고 있는가?		●	●
		다양성 반영	15. AI 서비스에 대한 다양한 계층과 지역의 의견을 균형 있게 반영하고 있는가?		●	●

대항목	중항목	소항목	점검사항	분류기준		
				1 (27개)	2 (57개)	3 (68개)
	사회적 약자와 디지털 소외 계층 보호	접근성	16. AI 서비스에 고령자, 장애인, 이주민 등 사회적 약자가 편리하게 접근할 수 있는가?	●	●	●
		편의성	17. AI 서비스 제공 시 연령과 학력 등에 관계없이 국민 누구나 이해할 수 있는 쉬운 용어를 사용하고 있는가?	●	●	●
			18. AI 서비스를 모든 국민이 어떤 환경에서든 편리하게 이용할 수 있도록 다양한 채널을 제공하고 있는가?	●	●	●
			19. AI 서비스의 사용자 환경이 직관적이고 단순하게 설계되어 모든 사용자가 복잡한 조작 없이 사용할 수 있는가?	●	●	●
		디지털 격차 해소	20. AI 서비스의 접근성 및 디지털 취약계층의 활용 능력 향상을 위하여 교육, 기기 지원, 정보 제공 등을 하고 있는가?	●	●	●
투명성 (12)	정보공개	도입 목적 공개	21. AI 시스템 도입 목적과 필요성, 담당 기관 등에 대해 설명(공개) 하고 있는가?		●	●
		AI 적용 여부 공개	22. AI가 적용되는 행정 서비스의 영역과 범위를 이용자에게 공지하고 있는가?	●	●	●
		평가 결과의 공개	23. AI 서비스에 대한 평가 결과를 정기적으로 공개하고 있는가?		●	●
		개선 이행 과정 공개	24. AI 서비스 평가 결과 및 이용자 의견에 따른 개선 이행 과정을 공개하고 있는가?		●	●
		데이터 출처 공개	25. AI 학습에 활용된 데이터의 출처와 수집 방법을 공개하고 있는가?	●	●	●
	설명 가능성 및 국민 이해 제고	설명 가능성	26. AI가 내린 결정이나 결과에 대해 기술적으로 설명할 수 있는가?		●	●
			27. AI 서비스의 의사결정 기준이나 조건 등을 사용자가 이해하기 쉽게 설명하고 있는가?	●	●	●
		서비스 한계 및 오류 가능성	28. AI 서비스의 기능, 적용 범위의 한계, 오류 가능성을 안내하고 있는가?	●	●	●
		용어 일관성	29. AI 서비스 내 관련 용어나 개념이 표준화되어 일관성 있게 사용되고 있는가?	●	●	●
	운영 개선 및 환류 체계 구축	외부 전문가 참여 구조	30. AI 서비스의 설계 과정에 외부 전문가 또는 이해관계자들이 공식적으로 참여하는 절차가 마련되어 있는가?		●	●
		이의제기 처리	31. AI 서비스 결과에 대한 사용자의 이의제기를 체계적으로 접수하고 처리하는 절차가 운영되고 있는가?		●	●
		최신 정보 반영	32. AI 서비스 결과의 정확성과 신뢰성을 높이기 위해 최신 정보를 반영할 수 있는 체계가 마련되어 있는가?		●	●
		관련 근거 제시	33. AI 서비스가 제공하는 결과에 대해 관련 법령 등의 근거를 함께 제시하고 있는가?		●	●

대항목	중항목	소항목	점검사항	분류기준		
				1 (27개)	2 (57개)	3 (68개)
책임성 (12)	책임의 주체 설정과 분배	책임 주체 명확화	34. AI 서비스의 기획, 개발, 도입, 운영, 위기 대응, 폐기 등 단계별 책임자와 담당 조직이 명확히 지정되어 있는가?	●	●	●
		책임 소재 추적	35. AI 시스템이 의사결정을 하거나 오류가 발생했을 때 책임 소재를 추적할 수 있는 기록 체계가 구축되어 있는가?		●	●
		법적 기준 준수	36. AI 서비스 운영에 관련된 법적 요구사항과 책임 범위를 명확히 정의하고 있는가?	●	●	●
		AI 결과물 합법성 검토	37. AI 서비스 결과물이 관련 법령과 규정에 부합되는지 정기적으로 검토하고 있는가?		●	●
		데이터 합법적 수집·이용	38. AI에서 학습·활용하는 데이터가 적절한 경로를 통해 수집·이용 되고 있는가?	●	●	●
	감사·감독 및 사후관리	외부 감사	39. AI 시스템에 대한 외부 감사를 받을 수 있는 체계를 갖추고 있는가?			●
		자동화 영향 검토	40. AI 서비스의 자동화된 결정이 개인에게 중대한 영향을 미칠 수 있는 경우, 그로 인한 피해 및 간접 영향을 사전에 검토하는 절차가 마련되어 있는가?			●
		인적 감독 체계	41. AI 시스템의 결정에 대한 인간의 검토와 개입이 가능한 감독 체계가 마련되어 있는가?			●
		사후 책임	42. AI 서비스에 대한 사후 책임 소재를 밝히기 위한 기술적·관리적 장치가 마련되어 있는가?			●
		윤리 검토 체계	43. AI 서비스에 대한 윤리적 이슈 발생 시 검토할 수 있는 내부 기구나 위원회가 마련되어 있는가?		●	●
	사회적 책임 및 피해 대응	피해 구제	44. AI 서비스의 문제 발생 시 대상자에게 신속하게 알리고 적절한 후속 조치를 이행하는 절차가 마련되어 있는가?			●
		피해 재발 방지	45. AI 서비스의 문제점이나 한계가 발견될 경우 이를 신속하게 개선하고, 재발을 방지하는 환류 체계가 마련되어 있는가?			●
안전성 (15)	보안 및 접근 통제	보안 체계	46. AI 시스템에 대한 보안 위협을 탐지하고 대응할 체계가 구축되어 있는가?		●	●
		데이터 접근 관리	47. AI 시스템의 운영에 필요한 개인정보를 포함한 데이터의 접근 통제 및 권한 관리에 대한 정책이 수립되어 있는가?		●	●
		접근 권한 관리	48. AI 시스템에 대한 권한별 접근 통제와 인증 체계가 구현되어 있는가?		●	●
		AI 공격 대응 계획	49. AI 시스템의 모델 및 데이터에 대한 공격(프롬프트 해킹 등)에 대해 대응 계획을 주기적으로 갱신·관리하고 있는가?		●	●
	예방·점검·모니터링	시스템 안정성	50. AI 시스템이 예상하지 못한 상황에서 안정적으로 작동하도록 설계되어 있는가?		●	●

대항목	중항목	소항목	점검사항	분류기준		
				1 (27개)	2 (57개)	3 (68개)
		정기적 기술 점검 및 관리	51. AI 시스템 장애 예방을 위한 정기적으로 기술 점검과 관리를 실시하고 있는가?		●	●
		시험 및 검증	52. AI 서비스 개시 전 다양한 시나리오로 시험·검증하는 체계를 갖추고 있는가?		●	●
		잠재적인 위험 분석 및 대응	53. AI 서비스 운영 중 나타날 수 있는 잠재적 위험을 지속적으로 발굴하고 대응책을 마련하고 있는가?			●
		주의 사항 안내	54. AI 서비스의 안전한 이용을 위한 주의사항, 사용자 오류 시 대응 방안 등을 명확히 안내하고 있는가?	●	●	●
	위기·장애 대응 및 복구	위기 대응 및 복구 체계	55. AI 시스템 장애, 재난, 데이터 손실 등 위기 상황에 대비한 백업 및 복구를 위한 매뉴얼을 마련하고 있는가?		●	●
		오남용 및 악용 방지 체계	56. AI 서비스가 범죄, 허위 정보 유포 등 공공의 이익을 해치는 방향으로 악용되지 않도록 장치 및 체계가 마련되어 있는가?	●	●	●
		장애 감지 시스템	57. AI 시스템의 비정상적인 행동 패턴이나 오작동을 감지하고 대응하는 체계가 마련되어 있는가?		●	●
		사후 대책 마련	58. AI 시스템의 결함이 발생할 경우 보완 및 재발 방지 대책이 마련되어 있는가?		●	●
		오작동 대응	59. AI 서비스 오류, 오작동 등으로 정상적인 서비스 제공이 불가능할 경우 긴급 제동 또는 대체 서비스 제공 등의 대응체계가 마련되어 있는가?		●	●
		비상 연락 체계 구축	60. AI 서비스 관련 보안·안전 문제가 발생했을 경우 사용자가 신속하게 연락 및 신고할 수 있는 체계가 구축되어 있는가?	●	●	●
프라이버시 보호 (8)	개인정보 처리 적법성	개인정보 수집 제한	61. AI 시스템이 목적 달성에 필요한 최소한의 개인정보만을 수집하고 있는가?	●	●	●
		사생활 침해방지 평가	62. AI 시스템이 사생활의 비밀과 자유를 침해할 수 있는 요소를 사전에 식별·평가하고 조치하는 체계가 마련되어 있는가?		●	●
		목적 외 사용 제한	63. AI 서비스를 위해 수집한 정보의 활용 범위를 명확하게 명시하고, 명시된 목적에 따라 사용되도록 통제 체계가 구축되어 있는가?	●	●	●
	관리적·기술적 조치	개인정보 안전 관리	64. AI 시스템 개발·활용 과정에서 처리되는 개인정보의 안전한 관리를 위한 적절한 관리적·기술적 보호조치가 적용되어 있는가?	●	●	●
		불필요한 개인 식별·추론 방지	65. AI 서비스에서 불필요한 개인 식별, 개인정보 추론 가능성을 최소화하는 조치를 수행하고 있는가?		●	●
		개인정보 파기	66. AI 시스템에서 수집된 개인정보를 활용 이후 안전하게 파기하고 있는가?	●	●	●
	정보주체 권리 보호	신고 및 구제 절차	67. AI 서비스로 인해 개인정보 및 사생활 침해나 권리 제한이 발생한 경우 신고·이의제기·구제 절차가 마련되어 있는가?		●	●
		정보 주체 통제권 보장	68. AI 서비스에 대해 사용자가 자신의 데이터를 열람, 정정, 삭제할 수 있는 권한이 보장되어 있는가?	●	●	●