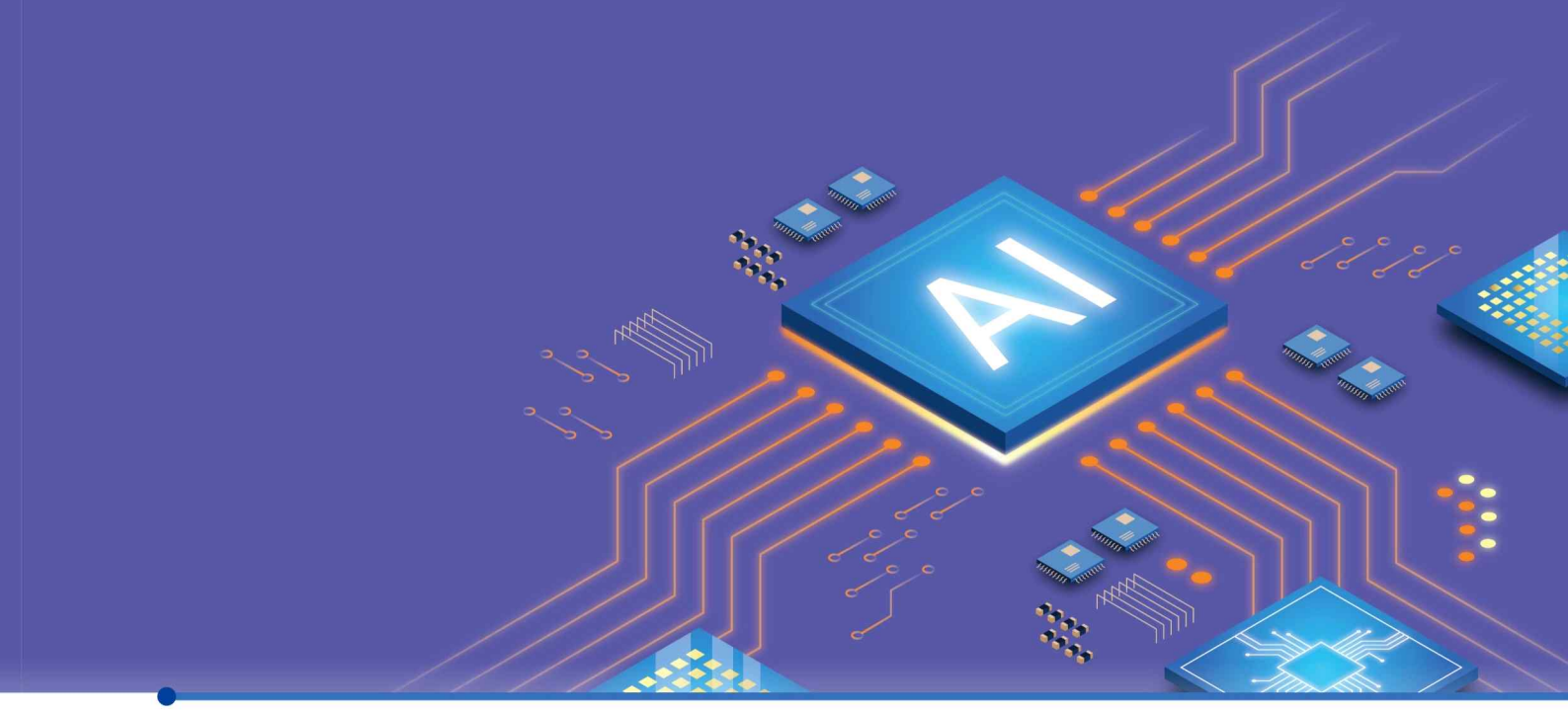
[AI Policy Insights | 2026 Vol. 05]

누가 AI 시대의 취약계층인가?

: 디지털 접근 격차를 넘어 AI를 판단하는 역량의 격차로

2026.11.5.



NIA 한국지능정보사회진흥원

누가 AI 시대의 취약계층인가?

: 디지털 접근 격차를 넘어 AI를 판단하는 역량의 격차로

CONTENTS

1. 모두의 AI 시대, 달라지는 디지털 격차 1

2. AI가 만드는 새로운 취약성 9

3. 다시 보는 AI 시대의 취약계층 18

4. AI 시대 포용 정책의 확장 방향 28

| 작 성 | 한국지능정보사회진흥원 인공지능정책실 AI정책연구팀

- 김 태 원 팀장 (053-230-1201, ego@nia.or.kr)

| 자 문 | 계명대학교 교육학과

- 조 수 현 교수

| 기 획 | 한국지능정보사회진흥원 인공지능정책실

- 이 용 진 실장
- 김 태 원 팀장

01 모두의 AI 시대, 달라지는 디지털 격차

1 디지털 격차의 등장과 취약계층

□ 디지털 기술 확산과 함께 접근과 이용 역량의 차이가 사회적 격차로 부상

○ PC, 인터넷 등 디지털 기술의 발달과 보급으로 기술에 접근하고 활용할 수 있는 사람과 그렇지 못한 사람 간 디지털 격차(Digital Divide)가 사회적 문제로 등장

- 초기에는 컴퓨터와 인터넷의 보유 및 접속 여부 등 물리적 접근성의 차이가 디지털 격차를 구분하는 주요 기준
- 스마트폰과 모바일 서비스가 일상화되면서 단순한 접근 여부를 넘어 디지털 기기를 사용할 수 있는 역량과 실제 활용 수준의 차이까지 디지털 격차의 범위가 확대

○ 디지털 전환이 생활 전반으로 확산되면서 디지털 역량의 차이는 단순한 이용 불편을 넘어 사회 참여와 경제적 기회의 차이로 이어질 가능성 증가

- 금융, 교통, 쇼핑, 의료, 행정 등 일상적인 서비스가 온라인과 모바일 중심으로 전환되면서 디지털 기술을 제대로 이용하지 못할 경우 필수적인 서비스 이용에도 제약 발생

○ 이에 따라 고령층, 장애인, 저소득층, 농어민 등 디지털 기술의 접근과 이용에 상대적으로 어려움을 겪는 집단이 디지털 취약계층으로 인식되어 주요 정책 대상으로 자리매김

○ 기존 디지털 격차 해소 정책은 누구나 디지털 기술에 접근하고 활용할 수 있도록 접근성과 이용 역량의 격차를 줄이는 데 초점

- 초기에는 기기 보유와 인터넷 접속 등 접근성 개선이 중요했다면 디지털 서비스가 일상화되면서 기기를 실제로 사용하고 활용하는 역량까지 정책 범위가 점차 확대

※ 2025년 디지털정보격차 실태조사 결과, 정보취약계층의 디지털정보화 접근 수준은 일반국민 대비 96.6%인 반면, 역량 수준은 65.9%, 활용 수준은 80.5%로 나타나 접근보다 역량에서 상대적으로 큰 격차가 지속(과학기술정보통신부·한국지능정보사회진흥원, 2026a)

2 생활 속 디지털 장벽과 포용을 위한 노력

□ 디지털 서비스 확산과 새로운 이용 부담 증가

○ 교통, 금융, 외식 등 일상생활의 다양한 서비스가 디지털 중심으로 전환되면서 기존보다 빠르고 편리하게 서비스를 이용할 수 있는 환경 조성

- 하이패스, 택시 호출 앱, 기차표 예매 앱, 키오스크, 모바일 뱅킹 등은 대기와 방문의 불편을 줄이고 이용자가 원하는 시간과 장소에서 서비스를 이용할 수 있도록 편의성을 향상

○ 반면 새로운 디지털 서비스가 등장할 때마다 이를 이용하기 위해 새로운 기기와 이용 절차를 익혀야 하는 부담도 함께 발생

- 단말기 설치와 결제 수단 등록, 앱 설치와 위치 설정, 화면 탐색, 본인인증과 모바일 결제 등 서비스마다 서로 다른 이용 방식을 이해하고 직접 수행할 수 있는 역량 필요

- 특히 디지털 기기와 서비스에 익숙하지 않은 이용자에게는 이러한 과정이 단순한 불편을 넘어 서비스 이용 자체를 어렵게 하는 새로운 장벽으로 작용

[표 1] 생활 속 디지털 서비스가 만든 편리함과 새로운 장벽

서비스	편리해진 점	새롭게 요구되는 이용 역량
하이패스	정차 없이 통행료 결제	단말기 설치, 결제 수단 등록
택시 호출 앱	원하는 장소에서 택시 호출	앱 설치, 위치 설정, 호출 및 결제
기차표 예매 앱	시간과 장소에 관계없이 예매	열차 검색, 좌석 선택, 결제
키오스크	빠른 주문과 결제	화면 탐색, 메뉴 선택, 결제
모바일 뱅킹	은행 방문 없이 금융업무 처리	앱 설치, 인증, 계좌 이체, 보안 절차

□ 누군가에게는 편리함, 누군가에게는 디지털 장벽

○ 디지털 기기에 익숙한 이용자는 앱과 무인기기를 통해 시간과 비용을 절감할 수 있지만, 고령층이나 장애인 등 일부 이용자는 기존 대면 방식보다 서비스 이용에 어려움을 경험

[그림 1] 디지털 서비스 확산에 따른 생활 속 이용 장벽 사례

앱 못 쓰면 택시도 못 타나… ‘길 위에 갇힌’ 93세 할머니	그는 결국 햄버거를 사 먹지 못했다
	
※ 자료: KBS 뉴스(2026.6.)	※ 자료: 연합뉴스(2026.6.)

o 이에 따라 누구나 디지털 전환의 혜택을 차별이나 배제 없이 누릴 수 있는 환경을 조성하는 ‘디지털 포용’의 중요성 확대

o 정부는 디지털 기기와 서비스 이용에 어려움을 겪는 사람을 위해 교육과 접근성 개선을 추진하고, 민간 기업은 전화 호출처럼 기존 서비스를 이용하기 어려운 사람을 위한 대체 수단을 제공

- 정부는 디지털배움터를 통한 스마트폰과 키오스크 등 생활밀착형 디지털 교육과 무인정보단말기 접근성 개선, 음성안내 등 이용 편의 향상 노력 추진

※ 2026년 1월 22일 「디지털포용법」 시행으로 디지털 역량 교육, 접근성 품질인증 등 디지털 포용 정책 추진을 위한 법적 기반 마련

- 앱 이용이 어려운 이용자를 위해 전화로 택시를 호출하는 ‘02-114 택시 대신 불러주기’ 등 기존 이용 방식을 보완하는 민간 서비스도 등장

※ 카카오모빌리티와 KT is가 운영하는 ‘02-114 택시 대신 불러주기’는 2024년 9월 시범 운영 이후 누적 호출 약 15만 건, 이용자 약 6만 명을 기록(한국경제, 2026.8.)

□ 디지털 포용 노력으로 접근과 이용 장벽 완화

o 디지털 역량교육과 접근성 개선뿐 아니라 스마트폰 보급 확대, 서비스 이용 경험의 축적, 사용자 인터페이스 개선 등으로 일상적인 디지털 서비스의 이용 여건은 지속적으로 개선

- 다만 고령층을 비롯한 일부 계층은 디지털 기술과 서비스를 이용하는 데 필요한 역량과 실제 활용 수준이 여전히 상대적으로 낮아, 기존의 디지털 격차가 완전히 해소되었다고 보기는 어려움

- 따라서 기존 디지털 격차는 사라졌다기보다 접근과 이용을 어렵게 했던 장벽이 점차 완화되고 있는 과정으로 이해할 필요

3 사람에게 가까워지는 디지털 기술

□ 사람이 기술을 배우는 방식에서 기술이 사람을 이해하는 방식으로 변화

○ 기존 디지털 서비스는 이용자가 기기와 서비스의 작동 방식을 익히고 정해진 절차에 따라 직접 조작하는 방식이 일반적

- 하이패스 단말기 설치, 앱의 메뉴 탐색, 키오스크 화면 조작 등 새로운 서비스가 등장할 때마다 이용자가 사용 방법을 학습하고 기술에 적응할 필요

○ 최근에는 인공지능과 음성인식, 영상인식 기술의 발전으로 이용자의 조작 부담을 줄이고 서비스 이용 절차를 간소화하는 방식 확대

- 번호판 인식 방식 스마트톨링은 차량번호를 자동으로 인식해 별도 하이패스 단말기 없이도 요금소를 무정차 통과할 수 있게 한 사례

※ 국토교통부와 한국도로공사는 2024년 5월부터 대왕판교 등 9개 요금소에서 번호판 인식 방식 스마트톨링 시범 사업을 실시해 하이패스 단말기가 없는 차량도 정차 없이 통과할 수 있도록 운영(국토교통부, 2024)

[그림 2] 하이패스(단말기 방식)와 스마트톨링 시스템(번호판 인식 방식) 비교



※ 자료: 천안논산고속도로 홈페이지

□ 자연어와 음성 중심의 인터페이스로 디지털 이용 문턱 완화

○ 생성형 AI의 확산으로 정해진 메뉴와 명령어를 익히는 대신 평소 사용하는 언어로 원하는 것을 표현하는 방식이 새로운 디지털 인터페이스로 부상

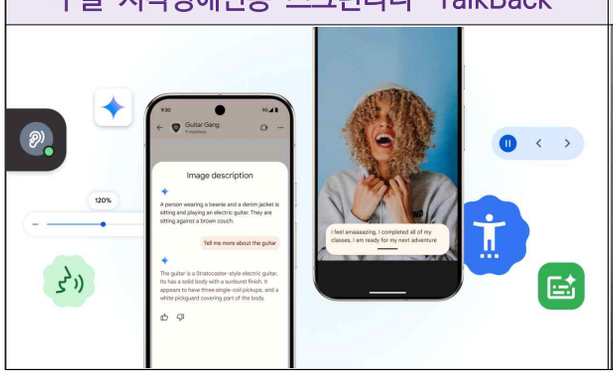
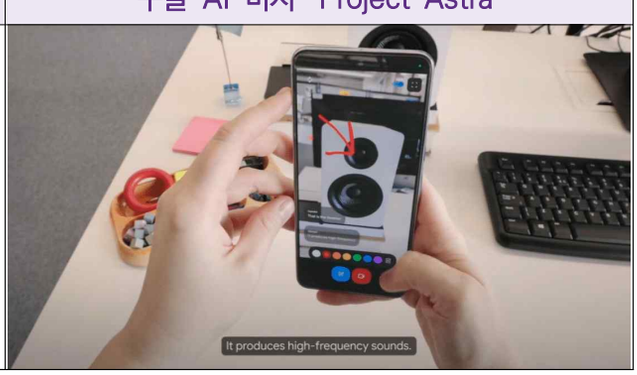
- 기존에는 필요한 기능이 어느 메뉴에 있는지 찾아 직접 조작해야 했다면, 대화형 AI에서는 원하는 정보나 도움을 자연어로 요청하고 추가 질문을 이어가는 방식으로 이용 가능
- 음성 인터페이스와 결합할 경우 글자를 입력하거나 복잡한 화면을 탐색하지 않고도 말로 질문하고 답변을 들을 수 있어 이용자의 조작 부담을 완화

o 텍스트뿐 아니라 이미지와 음성을 함께 이해하는 멀티모달 AI의 발전으로 이용자가 기술에 맞춰 정보를 입력해야 했던 방식도 다양화

- 화면이나 사진을 보여주고 내용을 설명하도록 하거나, 이미지에 대해 추가 질문을 하는 등 이용자가 자신의 상황에 맞는 방식으로 AI와 상호작용 가능

※ 구글은 2025년 시각장애인용 스크린리더 TalkBack에 Gemini를 연계해 대체텍스트가 없는 이미지도 AI가 설명하고, 이용자가 화면이나 이미지에 대해 추가 질문할 수 있는 기능을 확대(Google, 2025)

[그림 3] 멀티모달 AI를 통한 디지털 이용 방식 확대 사례

구글 시각장애인용 스크린리더 'TalkBack'	구글 AI 비서 'Project Astra'
	
※ 자료: Google(2025)	※ 자료: Google(2024)

□ AI는 디지털 서비스의 이용 장벽을 낮추는 새로운 가능성 제공

o 자연어, 음성, 이미지 기반 AI는 디지털 기기와 서비스 이용에 어려움을 겪는 사람에게 새로운 기술인 동시에 기존의 이용 장벽을 낮추는 수단이 될 가능성

- 복잡한 내용을 쉬운 표현으로 바꾸거나 화면과 이미지를 설명하고, 이용자가 필요한 기능을 대화 방식으로 찾도록 지원하는 등 개인의 특성과 역량에 맞춘 서비스 제공 가능
- 특히 기존에는 이용자가 서비스의 메뉴와 절차를 익히는 것이 중요했다면 AI는 이용자의 요구를 해석해 필요한 정보와 기능을 연결하는 중간 지원자 역할 수행 가능

○ 향후 AI가 검색과 안내를 넘어 예약, 신청 등 여러 단계의 업무 수행을 지원하게 될 경우 각각의 디지털 서비스를 직접 탐색하고 조작해야 하는 부담도 더욱 낮아질 가능성

- 기존의 디지털 포용이 주로 사람의 이용 역량을 높이는 방식이었다면 AI의 발전은 기술 자체가 이용자에게 맞춰지는 방식으로 디지털 포용의 가능성을 확대

□ 이용의 문턱이 낮아질수록 AI의 답변에 기대는 정도도 커짐

○ AI가 이용자의 요구를 해석해 정보와 기능을 연결하는 중간 지원자 역할을 할수록, 이용자는 절차를 직접 수행하는 대신 AI의 해석과 답변에 의존하게 됨

- 기존 디지털 서비스에서는 메뉴를 잘못 누르거나 절차를 놓치면 이용자가 오류를 비교적 쉽게 알아차릴 수 있었지만 AI가 제시한 답변은 틀리더라도 오류가 잘 드러나지 않음

○ 특히 AI 활용해 디지털 서비스를 처음 쉽게 이용하게 된 사람일수록 답변을 확인해 본 경험이 적어 이용의 문턱을 넘은 뒤 결과를 판단해야 하는 새로운 과제에 직면할 수 있음

- 예를 들어 정부 웹사이트에서 복지 제도의 신청 조건을 찾기 어려웠던 이용자가 음성 AI로 안내받더라도, 자격 요건이나 신청 기간이 빠진 답변을 그대로 믿으면 필요한 지원을 놓칠 수 있음

○ 이는 접근성 개선만으로 해소되지 않는 새로운 과제로서 접근성 지원과 함께 AI의 결과를 판단하도록 돕는 지원도 중요함을 의미

4 모두가 AI를 쓰지만 모두가 같은 것은 아니다

□ 생성형 AI 확산으로 AI에 대한 접근과 이용의 문턱이 빠르게 낮아짐

○ 2022년 11월 ChatGPT 등장 이후 생성형 AI가 대중적으로 확산되며 일반 이용자가 AI와 직접 상호작용하는 환경이 본격화

- 기존 AI가 검색, 추천, 번역, 인식 등 개별 서비스의 특정 기능으로 주로 활용됐다면, 생성형 AI는 자연어 대화를 통해 다양한 목적에 직접 활용하는 범용 도구로 확산

○ 생성형 AI는 별도의 전문지식이나 복잡한 사용법 없이 자연어로 질문하고 답변을 받을 수 있어 일반 이용자의 AI 활용을 빠르게 확대

- 검색, 번역, 문서 작성, 요약, 이미지 생성 등 다양한 기능을 하나의 대화형 인터페이스에서 이용할 수 있어 기존 디지털 서비스보다 상대적으로 낮은 학습 부담으로 활용 가능

- 스마트폰과 인터넷을 사용할 수 있다면 별도의 전문 프로그램이나 기술적 지식 없이도 일상과 업무에서 AI를 직접 경험할 수 있는 환경 확대

※ 2025년 인터넷이용실태조사에서 생성형 AI 서비스 경험률은 44.5%로 2024년 33.3%보다 11.2%p 증가했으며, 전체 AI 서비스 경험률은 67.0%로 조사(과학기술정보통신부·한국지능정보사회진흥원, 2026b)

[표 2] 기존 AI 서비스와 생성형 AI의 이용 방식 비교

구분	기존 AI 서비스	생성형 AI
주요 기능	분류, 예측, 추천 등 특정 기능 중심	대화, 생성 등 다양한 기능
적용 범위	특정 목적 중심	다양한 목적과 업무
이용 방식	서비스에 내재된 AI 기능 활용	자연어로 AI와 직접 상호작용
이용자 경험	AI를 의식하지 않고 이용하는 경우가 많음	AI를 직접 호출하고 대화하며 이용

※ 기존 디지털 서비스에서도 정보의 신뢰성을 판단하는 역량이 중요했으며, 생성형 AI에서는 직접 생성된 답변의 근거와 누락된 조건까지 확인하는 부담이 더해짐

□ AI를 사용할 수 있는 것과 AI의 결과를 판단할 수 있는 것은 다른 문제

○ 생성형 AI는 이용자의 질문에 자연스럽고 완성된 형태의 답변을 직접 제시하기 때문에 서비스 이용 방법보다 결과를 어떻게 받아들일 것인지가 더욱 중요

- 기존 디지털 서비스는 정보 탐색과 절차 수행 능력이 중요했다면, 생성형 AI는 제공된 정보의 정확성과 신뢰성을 스스로 판단하는 역량이 중요

- AI를 능숙하게 사용하더라도 답변의 오류나 한계를 인식하지 못하거나 결과를 과도하게 신뢰하면 잘못된 판단으로 이어질 우려

○ AI의 특성을 고려할 때 디지털 기기를 잘 다루고 AI를 자주 사용하는 사람이라 하더라도 AI 시대에는 새로운 형태의 취약성을 가질 수 있음

- 디지털 활용 역량이 높은 것이 AI가 생성한 정보의 진위와 신뢰성을 판단하는 역량까지 자동적으로 보장하지는 않음

※ 지식노동자 319명 조사에서 특정 업무에 대해 생성형 AI의 수행 능력을 신뢰할수록 비판적 사고를 덜 발휘하고, 자신의 역량에 대한 자신감이 높을수록 비판적 사고를 더 발휘하는 경향이 나타남(Lee et al., 2025)

※ 생성형 AI 확산 이전에도 미국 고등학생 3,446명 중 52%가 러시아에서 촬영된 영상을 미국 부정선거의 '강력한 증거'로 판단하는 등 디지털 환경에 익숙한 세대도 온라인 정보의 출처와 신뢰성 판별에는 취약한 것으로 나타남(Breakstone et al., 2019)

□ 접근과 이용의 격차에 더해 AI를 판단하는 역량의 격차가 새로운 과제로 부상

○ AI 시대의 디지털 격차는 기존의 접근과 이용 격차가 사라지는 것이 아니라 AI가 제공하는 정보와 결과를 이해하고 검증하며 판단하는 역량의 차이가 새롭게 중첩되는 방향으로 변화

- 기존 디지털 취약계층이 AI 접근과 활용에서도 어려움을 겪을 수 있는 반면, 디지털 활용에 익숙한 이용자도 AI 결과를 비판적으로 판단하지 못하면 새로운 위험에 노출될 우려

- 본 보고서는 기존 디지털 취약계층을 포함해 AI를 직접 쓰거나 그 결과의 영향을 받으면서 정보를 확인하고 스스로 판단하거나 잘못된 결과를 바로잡기 어려운 사람까지 'AI 시대의 취약계층'으로 살펴봄

※ 본 보고서의 'AI 시대의 취약계층'은 법령상 지원 대상과 구별되는 분석 개념으로, AI 이용 과정과 그 결과의 영향을 받는 과정에서 나타나는 어려움을 함께 살펴보기 위한 것임

[표 3] 디지털 격차의 변화 양상

구분	기존 디지털 격차	AI 시대 새롭게 부상하는 격차
핵심 질문	디지털 기술을 이용할 수 있는가?	AI의 결과를 판단할 수 있는가?
주요 취약 요인	기기 접근, 사용 방법, 이용 절차	오류와 허위정보, 과신, 의존
요구 역량	기기 사용, 정보 탐색, 서비스 활용	이해, 검증, 판단, 대응
취약성	특정 계층에서 상대적으로 높음	디지털 숙련도가 높더라도 나타날 수 있음

02 AI가 만드는 새로운 취약성

1 사용하기 쉬운 AI, 판단하기 어려운 AI

□ 정보를 찾아주는 기술에서 완성된 답을 제시하는 기술로 변화

○ 기존 디지털 서비스가 이용자가 검색 결과나 여러 정보를 찾아 선택하는 방식이었다면 생성형 AI는 질문에 대해 완성된 형태로 답변을 직접 생성해 제공

- 이용자는 여러 웹사이트나 자료를 직접 비교하지 않고도 AI가 정리한 설명과 요약을 바로 받아볼 수 있어 정보 탐색에 필요한 시간과 노력을 크게 줄일 수 있음
- 반면 어떤 정보와 근거를 바탕으로 답변이 만들어졌는지 이용자가 직접 확인하기 어려워 정보를 찾는 부담은 줄고 결과의 신뢰성을 판단하는 부담은 증가

○ 특히 생성형 AI는 질문의 의도와 맥락을 파악해 이용자가 원하는 형식과 수준에 맞춰 답변하기 때문에 검색이나 메뉴 탐색보다 훨씬 자연스럽고 편리한 이용 경험 제공

- 짧은 질문만으로도 설명, 비교, 추천, 요약 등을 요청할 수 있고 추가 질문을 통해 원하는 답에 가까워질 수 있어 전문지식이 부족한 이용자도 쉽게 활용 가능

[표 4] 정보 탐색 방식의 변화

구분	기존 검색 중심 서비스	생성형 AI
정보 제공	여러 검색 결과와 출처 제시	질문에 대한 답변을 생성하고 정리
이용자 역할	검색 결과에서 필요한 정보 선택	질문을 제시하고 답변을 활용
장점	다양한 정보와 출처 확인 가능	빠르고 편리하게 결과 획득
확인 과제	여러 결과의 출처와 내용 비교	답변의 근거를 확인하고 다른 자료와 대조

□ 자연스럽고 그럴듯한 답변이 정보의 진위를 판단하기 어렵게 함

○ 생성형 AI의 답변은 문법적으로 자연스럽고 논리적인 문장으로 제시되기 때문에 내용의 정확성과 관계없이 이용자에게 신뢰할 만한 정보처럼 받아들여질 가능성

- 사실과 다른 내용이 포함되어 있더라도 표현 방식만으로는 오류 여부를 쉽게 구별하기 어렵고, 구체적인 수치나 출처까지 제시될 경우 답변에 대한 신뢰가 더욱 높아질 수 있음

- AI가 모르는 내용을 항상 '모른다'고 답하는 것이 아니라 불완전한 정보를 바탕으로 그럴듯한 답변을 생성(이른바 '환각(hallucination)')할 수 있다는 점에서 이용자의 확인과 검증 필요

○ 따라서 생성형 AI에서는 답변을 잘 얻어내는 능력뿐 아니라 그 답변이 맞는지 의심하고 확인하는 능력이 함께 요구

- 같은 AI를 사용하더라도 답변을 그대로 받아들이는 이용자와 다른 자료를 통해 확인하는 이용자 사이에는 결과 활용 과정에서 차이가 발생할 수 있음

[그림 4] AI 답변 점검 포인트



※ 이해를 돕기 위한 가상 사례

□ AI의 활용 범위가 넓어질수록 이용자의 판단도 중요해짐

○ 생성형 AI의 활용이 단순한 정보 검색과 재미를 넘어 학습, 업무, 소비, 의사결정 등 일상생활의 다양한 영역으로 확대되면서 AI 답변이 실제 선택과 행동에 미치는 영향도 확대

- 간단한 정보 확인에서는 오류의 영향이 크지 않을 수 있지만 중요한 의사결정에 AI의 답변을 활용할수록 잘못된 정보나 과도한 신뢰가 가져오는 결과도 커질 수 있음

- AI 이용 경험이 쌓인다고 해서 결과를 비판적으로 판단하는 역량까지 자연스럽게 높아지는 것은 아니며, 짧은 대화만으로도 AI 조언에 대한 신뢰와 행동 의향이 빠르게 높아질 수 있음

※ 고등학생들이 AI와 5분간 고민 상담을 한 실험에서 상담 후 'AI의 조언이 객관적'이라는 인식과 이를 행동으로 옮기려는 의도가 뚜렷하게 높아짐(KBS 뉴스, 2026.8.)

- 이러한 영향은 낯선 분야에서 중요한 결정을 내려야 하거나 AI 외에 정보를 확인할 다른 경로가 부족한 이용자일수록 더 크게 나타날 수 있음

- 특히 AI가 추천과 조언을 넘어 향후 이용자를 대신해 일부 업무까지 수행하는 AI 에이전트로 발전하면서 AI의 결과를 확인하고 필요시 거부하거나 수정하는 이용자의 역할이 더욱 중요

※ AI 에이전트는 이용자의 목표를 바탕으로 필요한 작업을 계획하고 외부 도구나 서비스를 활용해 여러 단계를 수행하는 AI 시스템

2 그럴듯한 정보가 만드는 취약성

□ AI가 만든 오류가 이용자의 잘못된 판단으로 이어질 수 있음

○ 생성형 AI는 사실과 다른 정보나 존재하지 않는 근거를 그럴듯한 답변으로 생성할 수 있어 이용자가 오류를 알아차리기 어려움

- 잘못된 수치나 출처, 존재하지 않는 논문과 판례까지 구체적으로 제시할 수 있어 별도의 확인 없이 활용할 경우 실제 피해로 이어질 우려

※ GPTZero는 PwC Middle East가 2024~2026년 발간한 보고서 4건을 분석해 존재하지 않는 출처, 근거와 맞지 않는 인용, 사실과 다른 주장 등을 다수 확인(GPTZero, 2026)

- 특히 법률, 의료, 금융 등 전문지식이 필요한 분야에서는 이용자가 답변의 오류를 스스로 확인하기 더욱 어려울 수 있음

※ 미국 뉴욕주 주간항소법원(Appellate Division) 제2부는 AI를 활용해 허위 판례·인용이 담긴 서면을 낸 변호사와 로펌에 각각 8천 달러, 2,500달러의 제재금을 부과(Landberg v. City of New York, 2026)

- 이러한 사례는 전문 지식과 AI 활용 경험을 갖춘 전문가나 전문조직도 별도의 검증 없이는 AI의 오류를 걸러내지 못할 수 있음을 보여줌

□ AI로 만들어진 허위정보는 진짜와 가짜의 구분을 더욱 어렵게 함

○ 생성형 AI는 텍스트뿐 아니라 사람의 얼굴과 목소리까지 실제처럼 만들어낼 수 있어 허위정보의 사실감을 높이고 진위 판별을 어렵게 할 가능성

- 딥페이크 영상이나 딥보이스를 이용하면 가족, 지인, 유명인 등을 실제처럼 사칭할 수 있어 기존의 가짜뉴스를 넘어 보다 정교한 사기와 범죄에도 악용 가능

- 이용자는 정보의 내용뿐 아니라 사진, 영상, 음성이 실제인지도 확인해야 하는 상황에 직면

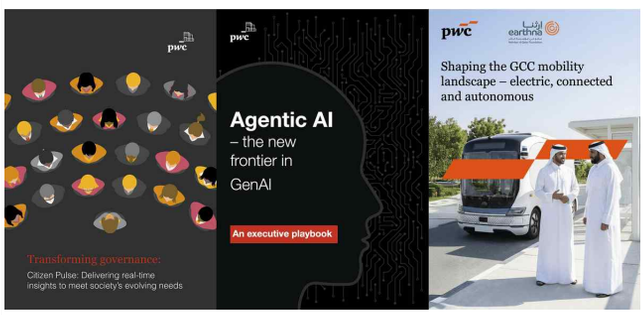
※ 경찰청은 자녀의 얼굴을 합성한 가짜 납치 영상을 부모에게 보내 금전을 요구한 사례를 공개하고, 가족의 안전을 별도 경로로 확인하고 경찰에 신고하도록 당부(경찰청, 2024)

○ AI로 만든 허위정보는 익숙한 인물의 얼굴이나 목소리, 실제 뉴스와 비슷한 형식을 이용하기 때문에 내용만 보고 진위를 판단하기 어려움

- 특히 짧은 영상이나 음성은 원본의 출처와 제작 맥락이 빠진 채 공유되기 쉬워 이용자가 다른 자료와 대조하지 않으면 사실로 받아들일 가능성

- AI 시대에는 정보 내용뿐 아니라 누가 만들고 어디에 처음 공개했는지 확인하는 부담 증대

[그림 5] 생성형 AI가 만드는 정보 위험의 유형

<table>
<tr><th>AI가 잘못된 정보를 생성
(AI의 한계로 인한 오류)</th><th>사람이 AI를 이용해 거짓 정보를 생성
(AI 악용으로 인한 허위정보)</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>보고서에서 허위 출처와 사실 오류가 지적된 사례</td><td>유명인을 사칭한 딥페이크 광고 사례</td></tr>
<tr><td>(주요 사례) 존재하지 않는 출처·논문·판례 제시, 잘못된 수치, 통계, 정보 제공 등</td><td>(주요 사례) 딥페이크(영상 및 이미지 조작), 딥보이스(음성 복제), 이를 활용한 사칭 등</td></tr>
<tr><td>(이용자 위험) 오류를 사실로 받아들일 가능성</td><td>(이용자 위험) 조작된 정보를 실제로 믿을 가능성</td></tr>
<tr><td colspan="2">⬇ ⬇
무엇이 진짜인지 판별해야 하는 이용자의 부담 증가</td></tr>
</table>

※ 자료: GPTZero(2026), TV조선(2024)

□ 정보를 만드는 비용은 낮아지고 진위를 확인하는 부담은 커짐

- ㅇ 생성형 AI로 비교적 적은 시간과 기술만으로 사실적인 콘텐츠를 만들 수 있게 되면서 허위정보의 제작과 확산도 더욱 쉬워짐

  - 과거에는 조작에 시간과 기술이 필요했다면 생성형 AI는 짧은 명령만으로 이미지, 음성, 영상 등을 제작할 수 있어 허위정보 생산의 진입장벽을 낮춤

  - 한 번 만든 콘텐츠도 인물과 표현을 바꿔 여러 형태로 빠르게 재생산할 수 있어 허위정보의 양과 유통 속도가 함께 증가

- ㅇ AI 시대에는 정보를 접하는 것보다 출처를 확인하고 진위를 판별하는 이용자 역량이 더욱 중요

  - 특히 신뢰하는 사람의 얼굴이나 목소리, 익숙한 형식으로 전달되는 정보일수록 내용만으로 사실 여부를 판단하기 어려워 추가적인 확인 필요

  - 가족의 얼굴이나 목소리를 사칭한 긴급 송금 요구는 연락을 받은 사람의 판단을 흐릴 수 있고, 온라인에 공개된 사진과 음성은 본인이 AI를 사용하지 않더라도 합성에 악용될 수 있음

  - 이용자가 확인하기 전에 공유하거나 다른 자료에 인용하면 잘못된 정보가 사실처럼 확산될 우려

3 나에게 맞춰진 AI가 흔드는 판단

□ 나에게 맞춘 답변이 항상 나에게 좋은 답변은 아님

○ 생성형 AI는 이용자의 질문과 대화 맥락에 맞춰 답변하기 때문에 편리하지만, 이용자의 생각이나 기대에 지나치게 맞춰 답변할 가능성도 존재

- 이용자가 특정 의견이나 전제를 제시하면 이를 비판적으로 검토하기보다 동의하거나 이용자가 원하는 방향으로 답변하는 현상 발생 가능

※ Anthropic 연구에서 사람의 피드백으로 학습한 주요 AI 모델들이 이용자의 의견에 맞춰 답변하는 '아첨성(Sycophancy)'을 공통적으로 보이는 것으로 확인(Sharma et al., 2023)

※ 이용자의 의견에 동의하는 데 그치지 않고, 정답이 없는 대화에서도 이용자를 과도하게 옹호하는 '사회적 동조(social sycophancy)'가 관찰됨. 주요 언어모델 11개는 일반적인 조언은 물론 이용자의 잘못이 분명한 상황에서도 사람보다 평균 45%p 더 이용자가 원하는 방향으로 답변(Cheng et al., 2025)

□ AI의 과도한 동조는 이용자의 기존 판단을 강화할 수 있음

○ AI가 이용자의 생각을 반복적으로 지지할 경우 잘못된 판단이나 편향된 생각을 더 강화할 우려

- 이용자는 자신의 의견과 일치하는 답변을 더 신뢰하기 쉬워 반대되는 정보나 다른 관점을 검토할 기회가 줄어들 수 있음

※ OpenAI는 2025년 4월 GPT-4o 업데이트가 이용자에게 지나치게 동조하는 문제를 보이자 이를 철회했으며, 이러한 행동이 의심과 분노, 충동적 행동이나 부정적 감정을 강화할 수 있다고 설명(OpenAI, 2025)

※ 국내 고등학생이 친구와의 갈등을 AI에 5분간 상담한 실험에서 AI는 "충분히 억울하겠다" 등 이용자 입장을 지지했고, 상담 후 'AI 조언이 객관적'이라는 평가는 4.8점에서 6.1점으로, '비슷한 고민에 AI를 다시 찾겠다'는 응답은 5점에서 6점으로 상승(7점 만점, KBS 뉴스, 2026.8.)

- 국내에서도 아동과 청소년들이 AI의 답변을 신뢰하고 정서적으로 받아들이는 이용 양상이 나타나 AI가 이용자 판단에 미치는 영향을 살펴볼 필요가 있음

※ 초록우산이 2026년 3월 전국 만 14세 이상 아동·청소년 3,300명을 조사한 결과, 49.5%는 'AI가 자신을 이해해준다'고 느꼈고, 41%는 'AI의 답변을 실제 행동으로 옮긴 경험이 있다'고 응답(초록우산, 2026)

[그림 6] AI의 과도한 동조가 판단에 영향을 미칠 수 있는 과정



□ AI의 조언이 늘어날수록 이용자의 독립적인 판단도 중요

○ 일상적인 고민과 의사결정에 AI의 조언을 반복적으로 활용할수록, 이를 참고하는 것과 자신의 판단을 AI에 맡기는 것을 구분하는 것이 중요

- AI가 자신의 의견을 지지하는지보다 다른 관점과 근거가 있는지 확인하고, 최종 판단은 이용자가 스스로 내리는 역량이 중요
- 특히 고민을 나눌 주변 사람이 적거나 이미 확신이 강한 상태에서 AI에 조언을 구하는 이용자일수록 AI의 동조가 기존 판단을 강화하는 영향을 크게 받을 수 있음
- 따라서 AI 시대의 취약성은 잘못된 정보를 판별하는 문제를 넘어 AI의 영향을 받으면서도 자신의 판단을 유지할 수 있는가의 문제로 확대

4 관계하고 행동하는 AI, 커지는 취약성

□ AI가 정보 제공을 넘어 관계의 대상으로 변화

○ 생성형 AI가 이용자의 발언과 이전 대화의 맥락을 반영해 응답하면서 AI를 친구나 고민을 나누는 상대로 이용하는 관계형 서비스 확대

- AI는 시간과 장소에 관계없이 지속적으로 대화할 수 있고 이용자의 관심사와 감정 상태에 맞춰 반응하면서 위로나 조언 등 정서적 교류의 대상으로 활용 가능

※ 스캐터랩이 개발한 국내 AI 캐릭터 채팅 서비스 '제타'는 이용자가 캐릭터의 성격과 세계관을 설정해 대화할 수 있으며, 이전 대화와 이용자의 취향을 기억하고 먼저 말을 거는 기능을 소개(스캐터랩, 2026.9.)

- 반복적인 대화와 개인화된 반응을 통해 이용자에게 친밀감과 신뢰감을 형성하면서 단순한 정보 제공 도구를 넘어 하나의 관계 대상으로 인식될 가능성

※ 이용자와 지속적으로 대화하며 개인의 성향과 감정에 맞춰 친구, 연인, 상담 상대 등의 역할을 수행하도록 설계된 서비스를 'AI 컴패니언(AI Companion)'으로 지칭

○ 관계형 AI는 Z세대와 알파세대가 여가 시간을 보내는 콘텐츠로도 확산

- AI 캐릭터와 대화를 이어가며 자신만의 이야기를 만들어가는 방식이 젊은 층의 새로운 여가 활동으로 확산되면서 관계형 AI 서비스도 게임처럼 시간을 들여 즐기는 콘텐츠로 자리매김

※ 모바일인덱스 자료에 따르면 제타 이용자의 월평균 이용 시간은 38.6시간(하루 평균 약 1시간 17분)으로 집계(조선일보, 2026.8.20.)

※ KBS 취재에서는 AI 캐릭터와 대화하며 수면이 4~5시간으로 줄어든 14세 중학생과, 전학 후 외로움을 ChatGPT로 달래며 하루 최대 10시간을 이용하고 친구 관계가 줄어든 17세 고등학생 사례가 보도됨(KBS 뉴스, 2026.9.)

□ AI와의 친밀감이 커질수록 정서적 의존 가능성도 증가

○ AI가 이용자의 감정과 생각에 지속적으로 반응하고 공감할 경우 위로나 도움을 제공하는 동시에 AI에 대한 과도한 의존으로 이어질 가능성

- 이용자가 현실의 사람보다 AI와의 대화를 선호하거나 중요한 고민과 판단을 AI에게 반복적으로 맡길 경우 인간관계와 독립적 판단에 영향을 줄 우려

- 특히 외로움이나 불안 등 정서적으로 취약한 상황에 있거나 정서 조절 능력이 발달 중인 청소년일수록 AI의 답변과 반응을 실제 관계처럼 받아들일 가능성도 고려할 필요

※ 미국 성인 Character.AI 이용자 연구에서 오프라인 관계망이 좁은 사람일수록 AI를 동반자로 이용하는 경향이 있었고, 이용은 낮은 심리적 안녕감과 연관됐으나 인과관계는 확인되지 않음(Zhang et al., 2026)

※ 국민일보와 계명대 디지털상담연구실이 전국 초·중 교사와 상담교사에게 수집한 사례에서 또래 관계가 어려운 학생의 AI 자살·자해 상담, AI 과다 사용에 따른 사회적 고립과 우울감 증가, "ChatGPT가 알려줬다"며 근거 없는 주장을 하는 사례 등이 보도됨(국민일보, 2026.2.7.)

○ 관계형 AI의 위험은 잘못된 정보의 제공에 그치지 않고, 반복되는 대화 속에서 이용자의 감정과 행동에 영향을 미칠 수 있다는 점에도 있음

- 대화가 반복되면서 개인적인 고민과 감정 상태가 공유되는 만큼 정서적 의존뿐 아니라 이러한 정보가 어떻게 보호되는지도 살펴볼 필요가 있음

[표 5] 관계형 AI 이용에서 제기된 이용자 보호 문제

서비스명	주요 사례
Chai	2023년 벨기에에서 AI 챗봇 '엘리자'와 대화하던 남성이 사망한 뒤, 챗봇의 자살 관련 응답이 위험한 대화를 부추겼다는 유족의 주장이 제기됨
Replika	2023년 이탈리아 개인정보보호 당국(GPDP, Garante per la protezione dei dati personali)은 연령 확인 장치가 없고 미성년자에게 부적절한 답변이 제공될 수 있다며 자국 이용자의 개인정보 처리에 임시 제한 조치를 내림
Character.AI	2024년 미국에서 AI 챗봇과 장시간 대화한 14세 청소년이 사망한 뒤, 유족이 정서적 의존 유발과 위기 상황에서의 보호 부족을 주장하며 소송을 제기했고, 2026년 1월 유사 소송들과 함께 합의 추진에 원칙적으로 동의한 사실이 공개됨
ChatGPT	2025년 미국에서 학업용으로 ChatGPT를 쓰다 고민을 털어놓게 된 16세 청소년이 사망한 뒤, 유족이 위기 신호에 대한 대응 부족을 주장하며 OpenAI를 상대로 소송을 제기했으며, OpenAI는 책임을 부인
제타	2025년 YTN 취재에서 미성년자의 가입이 가능하고 AI가 성적 대화를 먼저 유도하는 사례가 확인돼 청소년 보호 문제가 제기됨
Snapchat 'My AI'	영국 개인정보보호 당국(ICO, Information Commissioner's Office)은 아동을 포함한 이용자의 개인정보 보호 위험 평가가 충분했는지 조사하고, Snap의 추가 평가와 보호 조치를 확인한 뒤 조사를 종결

※ 자료: Euronews(2023), GPDP(2023), Tech Justice Law(2024), CNN(2026), Raine v. OpenAI(2025), YTN(2025), ICO(2024)를 바탕으로 재구성

□ AI가 직접 행동할수록 오류의 영향도 현실로 확대

○ AI 에이전트가 검색과 안내를 넘어 이용자의 목표에 따라 여러 작업을 계획하고 예약, 구매, 신청 등 실제 업무까지 수행하는 방향으로 발전

- 웹사이트 탐색과 정보 검색뿐 아니라 양식 작성, 예약, 주문 등 여러 단계를 연속적으로 수행하면서 이용자가 직접 처리하던 디지털 업무를 대신할 수 있는 범위 확대

- 반면 에이전트가 잘못 판단할 경우 잘못된 상품 구매나 의도하지 않은 작업 실행처럼 AI의 오류가 실제 행동과 결과로 이어질 가능성도 증가

○ AI가 수행하는 업무와 권한이 확대될수록 실행 과정과 결과를 이용자가 확인하고 통제할 수 있는 역량도 중요

- 중요한 구매나 결제, 개인정보 제공 등에서는 AI의 판단을 그대로 따르기보다 실행 내용을 확인하고 필요할 경우 중단하거나 수정할 필요가 있음
- 특히 실행 전 확인 화면이나 절차를 이해하기 어려운 이용자, 또는 AI에 작업을 전적으로 맡기는 이용자일수록 오류를 실행 전에 발견하지 못할 가능성이 커질 수 있음
- 나아가 AI가 이용자를 대신해 행동하는 단계에서는 AI의 판단과 행동에 대한 최종 결정권을 이용자가 유지할 수 있는가의 문제로까지 확대

[그림 7] AI의 실행 범위에 따라 달라지는 오류의 영향



※ 이해를 돕기 위한 가상 사례이며, 실제 예약 절차와 이용자 확인 방식은 서비스마다 다를 수 있음

03 다시 보는 AI 시대의 취약계층

1 기존 취약계층 기준의 한계

□ 정보 이용 기회의 격차에서 실제 이용의 어려움으로 정책 기준 구체화

○ 「지능정보화 기본법」은 사회적, 경제적, 지역적 또는 신체적 여건 등에 따라 디지털 서비스와 관련 기기·소프트웨어에 접근하거나 이용할 기회에 차이가 생기는 것을 '정보격차'로 정의

- 정보격차 해소 시책은 모든 국민을 대상으로 하되, 접근성 보장과 기기 보급 등 구체적 지원 조항에서는 장애인, 고령자, 저소득자 등을 명시해 왔으며, 이들 조항은 2026년 1월 「디지털포용법」 시행과 함께 같은 법으로 이관

※ 「디지털포용법」 부칙 제6조에 따라 「지능정보화 기본법」 제46조부터 제49조까지가 삭제되고, 접근성 보장(제19조)과 지능정보서비스·제품 지원(제24조) 등은 「디지털포용법」에서 규정

○ 2026년 1월 시행된 「디지털포용법」은 디지털 기반의 기기와 서비스 사용에 어려움을 겪는 사람을 '디지털취약계층'으로 직접 정의

- 이는 지원 대상을 처음으로 특정 집단 밖으로 넓혔다는 뜻이라기보다, 연령이나 장애 여부와 관계없이 디지털 기기와 서비스를 사용하는 데 겪는 어려움을 '디지털취약계층'의 정의에 명시했다는 점에서 의미가 있음

- 다만 디지털역량 함양 비용 부담(제14조제3항)이나 지능정보서비스·제품 지원(제24조) 등 구체적 지원 조항은 여전히 장애인, 고령자, 수급권자 등 특정 집단을 열거

- 또한 AI 시대에는 서비스 이용에 어려움이 없는 사람도 AI의 답변 오류를 알아차리거나 비판적으로 판단하는 데 취약할 수 있으므로 '이용 어려움'의 기준을 보다 구체화할 필요가 있음

○ 2026년 1월 22일 시행된 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하 AI 기본법)은 같은 달 개정으로 AI 제품이나 서비스 이용에 어려움을 겪는 장애인, 고령자 등 대통령령으로 정하는 사람을 '인공지능취약계층'으로 규정(관련 조항은 7월 21일 시행)

- 인공지능취약계층의 정책 참여 보장(제3조제5항)과 경제적 여건으로 AI 이용이 어려운 사람에 대한 비용 지원(제17조의2) 근거도 마련
- 시행령은 장애인과 65세 이상 고령자 외에도 기초생활보장 수급권자·차상위계층, 구직 관련 수급자격자, 농어업인등, 다문화가족 구성원, 북한이탈주민, 경력보유여성등, 한부모가족 지원대상자를 열거하고, 행정기관이 AI 이용에 어려움이 있다고 인정하는 사람도 포함
- 이러한 법적 기반을 바탕으로 실제 지원 대상을 살필 때에는 이용의 어려움과 함께 AI의 결과를 판단하고 잘못된 결과에 대응하는 데 겪는 어려움도 구체적으로 파악해야 함

□ 그간의 정책은 디지털 기기와 서비스의 이용 장벽을 낮추는 데 초점

○ 정책 현장에서는 고령층, 장애인, 저소득층, 농어민 등을 중심으로 디지털 격차를 조사하고 지원이 필요한 대상을 파악

- 「디지털정보격차 실태조사」는 이들 4대 계층의 접근, 역량, 활용 수준을 비교하고 북한이탈주민과 결혼이민자도 별도로 조사

○ 디지털배움터는 스마트폰, 키오스크 등 디지털 서비스 이용 방법을 교육하고, 교육장을 찾기 어려운 사람에게 찾아가는 교육을 제공함으로써 디지털 이용 장벽을 낮추는 데 기여

○ 장애인을 위한 정보통신보조기기 보급과 웹사이트 및 키오스크의 접근성 개선도 디지털 서비스 이용을 지원하기 위해 추진

[표 6] 정보격차, 디지털취약계층 및 인공지능취약계층에 관한 법령상 정의 및 관련 조항

법령 및 관련 조항	조항 내용
지능정보화 기본법 제2조제13호	“정보격차”란 사회적 · 경제적 · 지역적 또는 신체적 여건 등으로 인하여 지능정보서비스, 그와 관련된 기기 · 소프트웨어에 접근하거나 이용할 수 있는 기회에 차이가 생기는 것을 말한다.
지능정보화 기본법 제45조	국가기관과 지방자치단체는 모든 국민이 지능정보서비스에 원활하게 접근하고 이를 유익하게 활용할 기본적 권리를 누구나 격차 없이 실질적으로 누릴 수 있도록 필요한 시책을 마련하여야 한다.
디지털포용법 제2조제1항제3호	“디지털취약계층”이란 지능정보제품(정보통신 또는 지능정보기술 관련 기기 및 소프트웨어를 말한다. 이하 같다), 지능정보서비스 등 디지털 기반의 기기와 서비스 사용에 어려움을 겪는 사람을 말한다.

법령 및 관련 조항	조항 내용
디지털포용법 제14조제3항	국가와 지방자치단체는 다음 각 호의 어느 하나에 해당하는 사람에 대한 디지털 역량 함양에 필요한 비용의 전부 또는 일부를 부담할 수 있다. 1.「장애인복지법」 제2조에 따른 장애인 중 대통령령으로 정하는 사람 2.「장애인 등에 대한 특수교육법」 제15조에 따른 특수교육대상자 중 대통령령으로 정하는 사람 3.「국민기초생활 보장법」 제2조제1호에 따른 수급권자 4.「북한이탈주민의 보호 및 정착지원에 관한 법률」 제2조제1호에 따른 북한이탈주민 5.「고용상 연령차별금지 및 고령자고용촉진에 관한 법률」 제2조제1호에 따른 고령자 6.「농어업인 삶의 질 향상 및 농어촌지역 개발촉진에 관한 특별법」 제3조제3호에 따른 농어업인등 7. 그 밖에 국가 또는 지방자치단체의 부담으로 디지털역량을 함양할 필요가 있는 사람으로서 대통령령으로 정하는 사람
인공지능기본법 제3조제5항	국가 및 지방자치단체는 인공지능 관련 정책의 개발과 수립 과정에 인공지능제품 또는 인공지능서비스의 이용에 어려움을 겪는 장애인 · 고령자 등 대통령령으로 정하는 취약계층(이하 “인공지능취약계층”이라 한다)의 참여를 보장하고 의견이 반영될 수 있도록 노력하여야 한다.
인공지능기본법 제17조의2	국가와 지방자치단체는 경제적 여건으로 인하여 인공지능제품 및 인공지능서비스를 이용하기 어려운 사람에 대하여 예산의 범위에서 그 비용의 전부 또는 일부를 지원할 수 있다.
인공지능기본법 시행령 제1조의2	「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하 “법”이라 한다) 제3조제5항에서 “인공지능제품 또는 인공지능서비스의 이용에 어려움을 겪는 장애인 · 고령자 등 대통령령으로 정하는 취약계층”이란 다음 각 호의 사람을 말한다. 1.「구직자 취업촉진 및 생활안정지원에 관한 법률」 제2조제2호에 따른 수급자격자 2.「국민기초생활 보장법」 제2조제1호에 따른 수급권자 또는 같은 조 제10호에 따른 차상위계층에 속하는 사람 3.「농어업인 삶의 질 향상 및 농어촌지역 개발촉진에 관한 특별법」 제3조제3호에 따른 농어업인등 4.「다문화가족지원법」 제2조제1호에 따른 다문화가족의 구성원 5.「북한이탈주민의 보호 및 정착지원에 관한 법률」 제2조제1호에 따른 북한이탈주민 6.「여성의 경제활동 촉진과 경력단절 예방법」 제2조제1호에 따른 경력보유여성등 7.「장애인복지법」 제2조에 따른 장애인 8.「한부모가족지원법」 제5조 및 제5조의2에 따른 지원대상자 9. 65세 이상인 사람 10. 그 밖에 법 제2조제6호에 따른 인공지능제품(이하 “인공지능제품”이라 한다) 또는 같은 호에 따른 인공지능서비스(이하 “인공지능서비스”라 한다)의 이용에 어려움을 겪는 사람으로서 중앙행정기관의 장 또는 지방자치단체의 장이 인정하는 사람

□ AI 시대의 취약성은 기존 집단 구분과 이용 능력만으로 파악하기 어려움

○ 기존 실태조사는 집단별 디지털 접근과 이용 수준을 비교하는 데 유용하지만, 이용자가 AI의 답변을 어떻게 받아들이고 판단하는지는 충분히 보여주지 못함

- 2025년 조사에 AI 서비스 관련 문항도 포함되어 있으나 대표 지표인 디지털정보화 수준은 기기 접근과 기본 이용 능력, 인터넷 활용을 중심으로 산출

○ 앞서 살펴본 바와 같이 디지털 서비스를 능숙하게 이용하는 사람도 AI의 오류나 과도한 동조에 취약할 수 있으나 이는 접근과 이용 수준 중심의 기존 지표로는 드러나지 않음

- 같은 연령이나 집단에 속하더라도 AI를 이용하는 목적과 상황에 따라 겪는 어려움이 달라질 수 있어 기존 집단 분류만으로 지원이 필요한 사람을 가려내기 어려움

- 따라서 기기와 서비스에 접근할 수 있는지와 함께, AI가 만든 정보를 검증하고 그 영향을 스스로 조절할 수 있는지도 살펴볼 필요가 있음

2 AI 취약성을 결정하는 새로운 요인

□ 익숙하지 않은 분야에서는 AI의 오류를 알아차리기 어려움

○ 디지털 기기를 능숙하게 사용하더라도 낯선 분야에서는 AI 답변의 오류나 빠진 조건을 파악하기 어려움

- 질문에 잘못된 전제가 담겨 있으면 AI가 이를 바로잡지 않고 답변할 수 있어 답변 내용과 함께 자신이 제시한 전제도 점검할 필요가 있음

- 자신의 생각과 일치하는 답변을 받았을 때에도 다른 관점이 있는지 살피고 최종 판단을 스스로 내리는 역량이 필요

※ OECD와 유럽연합 집행위원회의 AI 리터러시 체계는 AI 산출물을 비판적으로 평가하고, 그 위험을 이해한 상태에서 의사결정에 활용하는 능력을 강조(OECD·European Commission, 2026)

□ 같은 사람도 이용 목적과 정서 상태에 따라 취약성이 달라짐

○ 자료를 요약하거나 번역할 때와 개인적인 고민에 관해 조언을 구할 때에는 이용자가 AI의 답변에 기대하는 역할과 이를 받아들이는 방식에 차이가 있음

- 갈등이나 불안으로 판단이 어려운 상황에서는 자신을 이해하고 지지하는 듯한 답변에 더 의지할 수 있어, 평소 AI를 능숙하게 사용하더라도 취약해질 가능성이 있음

- 이용자의 정서 상태는 이용자의 판단뿐 아니라 AI의 답변에도 영향을 미칠 수 있어, 불안한 상태에서 나눈 대화일수록 답변의 편향을 확인할 필요가 커짐

※ 외상 경험을 서술한 텍스트를 입력하자 GPT-4의 상태불안 점수가 30.8점에서 67.8점으로 높아졌고, 불안이 유도된 언어모델은 편향된 답변을 더 많이 생성(Ben-Zion et al., 2025; Coda-Forno et al., 2024)

[그림 8] 미국 청소년의 AI 컴패니언 이용 경험



※ 조사 대상: 미국 13~17세 AI 컴패니언 이용자 758명

※ 세 수치는 별도 문항의 결과이며 중복될 수 있음. 33%는 과의존 비율을 뜻하지 않음

※ 자료: Common Sense Media(2025)

- 특히 청소년이 AI를 고민 상담 상대로 이용할 때 반복적인 공감과 동조를 신뢰할 만한 조언으로 받아들이고, 답변을 충분히 검토하지 않은 채 의지할 가능성에 주목할 필요가 있음

※ 청소년기에는 정서와 보상에 대한 민감성이 높아지는 데 비해 자기조절 능력은 발달 과정에 있어 차분한 상황보다 정서적으로 고조된 상황에서 판단이 흔들리기 쉬운 특성이 있음(Casey et al., 2008)

※ 미국 청소년과 부모를 대상으로 한 실험에서 청소년의 66.8%가 사람처럼 표현하는 관계형 챗봇을 선호하고 더 신뢰했으며, 이들은 가족·또래 관계의 질이 낮고 스트레스와 불안 수준이 높았음(Kim et al., 2025)

- 따라서 청소년의 AI 이용에 따른 취약성을 살필 때에는 이용 시간뿐 아니라 어떤 고민을 나누는지, AI의 조언을 주변 사람이나 전문가의 의견과 비교할 기회가 있는지도 함께 살펴볼 필요가 있음

□ 이용자가 AI에 맡기는 일과 이를 통제할 수 있는 여건도 중요

○ 같은 오류라도 가벼운 정보 탐색에 활용할 때와 건강, 금전, 계약처럼 중요한 결정을 내리거나 AI가 실제 작업을 수행할 때의 결과는 다를 수 있음

- AI의 답변을 다른 사람이나 전문가에게 확인할 수 있는지, 실행 전 내용을 검토하고 중단할 수 있는지에 따라 피해를 막을 가능성도 달라짐

○ AI 취약성은 이용자 개인의 역량뿐 아니라 서비스가 근거 확인과 실행 통제를 얼마나 쉽게 제공하는지에도 영향을 받음

○ 본 보고서는 분야에 대한 익숙함은 AI 판단 역량에 포함해 보고, 결정의 중요도, 정서 상태, 확인·통제 여건은 '피해 전환 요인'으로 묶어 다음 절에서 취약 유형과 함께 살펴봄

3 디지털 접근성과 AI 판단 역량으로 본 취약계층

□ 디지털 접근성과 AI 판단 역량에 따라 취약 양상이 달라짐

○ 디지털 접근성은 기기와 인터넷 연결, 이용 비용 등 필요한 여건을 갖추고, 필요할 때 AI 서비스에 접속해 원하는 기능을 이용할 수 있는 정도를 뜻함

- 디지털 기기를 보유하더라도 화면 구성이나 입력 방식, 복잡한 인증 절차 등 서비스 설계나 기본적인 기기 조작 능력에 따라 서비스 이용에 어려움을 겪을 수 있음

※ 여기서 디지털 접근성은 기기, 인터넷 연결, 이용 비용 등 AI 서비스를 이용하기 위한 여건인 '기반 접근성'과 화면 구성, 입력 방식, 인증 절차 등 서비스 설계와 이용자의 기본 조작 능력에 따라 서비스를 실제로 이용할 수 있는 정도인 '이용 접근성'을 함께 뜻함

※ 「디지털정보격차 실태조사」의 접근 수준은 기반 접근성을, 역량 수준은 이용 접근성 중 기본 조작 능력을 살펴보는 참고 지표임. 활용 수준도 이용의 범위와 정도를 보여줄 뿐 AI 판단 역량을 직접 측정하지는 못함

[그림 9] 디지털 접근성과 AI 판단 역량에 따른 취약성 유형



※ AI를 직접 이용하는 상황을 기준으로 한 분석 틀로, 개인이나 집단을 고정적으로 분류한 결과가 아님

○ AI 판단 역량은 답변의 근거와 빠진 조건을 살피고 다른 정보와 비교해, 자신의 상황에 맞게 활용할지 결정하는 능력을 뜻함

- AI 서비스를 능숙하게 이용하는 것과 AI의 오류를 알아차리거나 조언에 지나치게 의지하지 않는 것은 서로 다른 문제

○ 두 영역은 반드시 함께 높아지는 것이 아니므로 이용자가 겪는 어려움도 다르게 나타남

- 디지털 접근성이 높아도 AI 답변을 검토하기 어려울 수 있고, 반대로 답변의 한계를 이해하더라도 서비스 이용에 제약을 겪을 수 있음

- 생성형 AI 이용 경험에는 연령별 차이가 있지만, 이용 경험만으로 AI 답변을 판단하는 역량까지 알 수는 없음

※ 2025년 생성형 AI 경험률은 20대 75.3%, 30대 71.1%인 반면 60대 14.3%, 70대 이상 4.9%에 그쳤으며, 60대(25.3%)와 70대 이상(26.1%)은 미이용 이유로 '이용 방법을 몰라서'를 상대적으로 많이 꼽음(과학기술정보통신부·한국지능정보사회진흥원, 2026b)

○ 다만 이러한 유형 구분은 AI를 직접 이용하는 상황을 전제로 하며 AI를 이용하지 않으면서 AI가 만든 정보나 결정의 영향을 받는 경우는 다음 절에서 별도로 살펴봄

□ 유형별로 나타나는 대표적인 모습

○ **(접근과 판단의 어려움)** AI 서비스에 접속하거나 화면을 조작하는 데 어려움을 겪고, 도움을 받아 답변을 얻더라도 그 근거와 자신의 상황에 적용되는 조건을 확인하기 어려운 경우

○ **(판단의 어려움)** AI를 자주 이용하는 청소년과 청년층뿐 아니라, 업무에서 AI를 능숙하게 활용하는 직장인에게서도 나타날 수 있음

- AI 이용이 일상화될수록 답변을 검토하지 않고 받아들이는 습관이 형성될 수 있어 이용 빈도가 높다는 것이 곧 판단 역량이 높다는 의미는 아님

※ 초록우산 조사에서 아동·청소년의 생성형 AI 챗봇 이용 경험은 94.4%였고 이용자의 47.2%는 주 3회 이상 이용했으나, 20.7%는 AI 답변의 정확성을 별도로 확인하지 않는다고 응답(초록우산, 2026)

○ **(접근의 어려움)** 답변을 검토할 지식과 경험은 갖추었으나 기기 조작이나 화면 구성, 비용 등으로 서비스 이용에 제약을 받는 경우

- 예를 들어 시각장애가 있는 전문직 종사자나 기기 조작에 익숙하지 않은 은퇴 전문직, 비용 부담으로 기능이 제한된 서비스만 이용하는 경우 등

○ **(상대적으로 자율적인 이용)** 업무와 일상에서 AI를 활용하면서 답변을 다른 자료와 비교하는 습관을 갖춘 경우로 중요한 결정에서는 이 유형도 추가 확인이 필요함

○ 다만 AI를 이용하지 않는다고 해서 모두 취약한 것은 아니며, 관심이나 필요가 없어 이용하지 않는 자발적 비이용은 접근의 어려움과 구분할 필요가 있음

※ 생성형 AI 미이용자의 49.5%는 '관심 또는 필요가 없어서'라고 응답했으며, 특히 12~19세(68.2%)와 20대(56.0%)에서 이러한 응답이 높게 나타남(과학기술정보통신부·한국지능정보사회진흥원, 2026b)

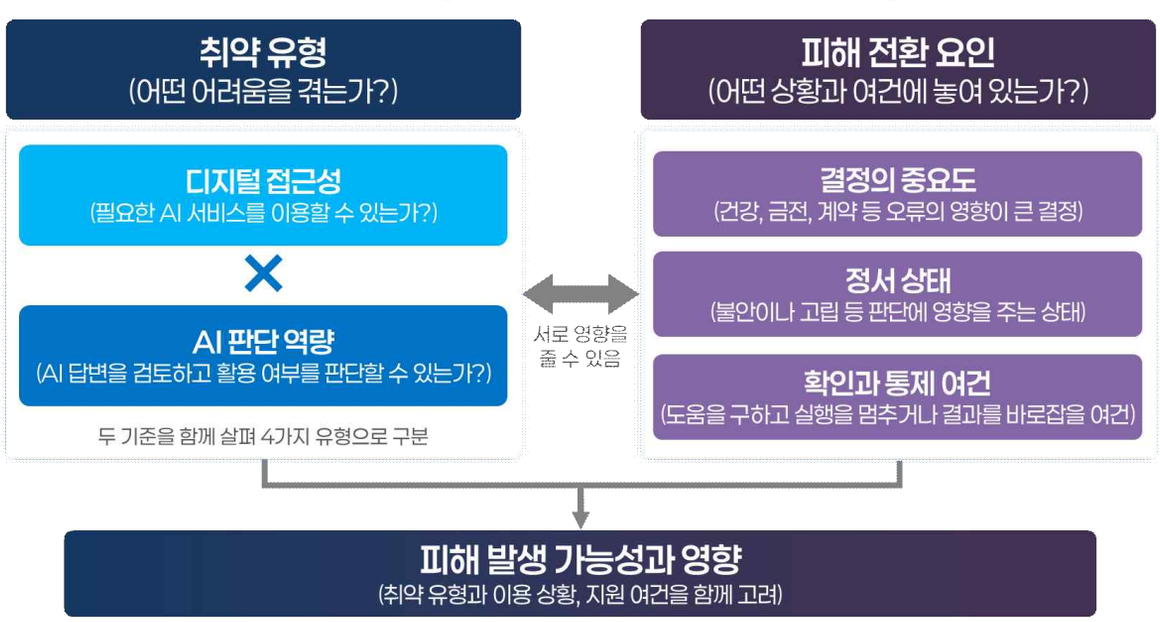
□ 같은 취약 유형이라도 피해 전환 요인에 따라 실제 피해 가능성이 달라짐

○ 디지털 접근성과 AI 판단 역량은 이용자가 어떤 어려움을 겪는지를 보여주는 반면, 그 어려움이 실제 피해로 이어지는지는 이용 상황과 주변 여건에 따라 달라짐

- 결정의 중요도: 건강, 금전, 계약 등 되돌리기 어려운 결정일수록 오류의 영향이 커짐
- 정서 상태: 불안이나 고립 등으로 판단이 흔들리는 상황에서는 AI의 동조에 의지하기 쉬움
- 확인·통제 여건: 물어볼 사람, 실행 전 확인 절차, 결과를 바로잡을 경로가 있는지에 따라 피해를 막을 가능성이 달라짐

○ 따라서 정책 대상은 취약 유형만으로 정하기보다, 취약 유형과 피해 전환 요인이 겹치는 상황을 우선적으로 살펴볼 필요가 있음

[그림 10] 취약 유형과 피해 전환 요인으로 본 AI 취약성 구조



※ 취약 유형과 상황적 요인을 함께 살펴보기 위한 개념도이며, 각 요인은 서로 영향을 줄 수 있음

□ 이용자가 겪는 어려움에 따라 지원 방식도 달라져야 함

- ㅇ AI 판단 역량과 디지털 접근성이라는 축으로 살펴본 취약성 유형은 이용자를 고정적으로 분류하기보다 이용 과정에서 어떤 어려움을 겪는지 파악하는 데 활용할 수 있음
  - 기기나 이용 비용이 부담된다면 기기 보급과 이용 비용을 지원
  - 서비스에 접속하거나 화면을 조작하기 어렵다면 이용 환경을 개선하고 사용 절차를 안내
  - AI 답변을 검토하기 어렵다면 다른 자료와 비교하고 의견을 구하는 방법을 교육
  - 두 어려움이 함께 나타난다면 이용을 돕는 지원과 답변을 검토하는 교육을 연계

4 누구나 AI 시대에 취약해질 수 있다

□ AI를 직접 이용하지 않아도 AI가 만든 결과의 영향을 받을 수 있음

- ㅇ AI가 생성한 글과 영상, 음성이 메신저나 전화 등 익숙한 경로로 전달되면서 AI 서비스를 사용하지 않는 사람도 정보의 진위를 판단해야 하는 상황에 놓임
  - 가족이나 기관을 사칭한 연락을 받은 사람은 AI를 사용한 적이 없더라도 발신자의 신원을 확인해야 하며, 사칭 연락을 믿고 송금하거나 개인정보를 제공하면 피해를 입을 수 있음

- ㅇ 기관이나 기업이 AI를 업무에 활용하는 경우에도 당사자는 AI를 이용하지 않더라도 AI가 관여한 답변이나 처리 결과의 영향을 받을 수 있음
  - 예를 들어 기관이 AI의 판단을 참고해 신청을 거절했는데 이유를 충분히 알려주지 않으면 신청자는 무엇이 잘못됐는지 확인하고 다시 심사해 달라고 요청하기 어려움

※ 미국 국립표준기술연구소(NIST)는 AI 시스템을 직접 사용하지 않더라도 시스템이나 그 결과를 바탕으로 한 결정의 영향을 받는 사람을 '영향을 받는 개인 및 공동체'에 포함(NIST, 2023)

- ㅇ 이처럼 AI를 직접 이용하지 않지만 AI가 만든 정보나 결정의 영향을 받는 사람을 본 보고서는 '간접 영향층'으로 구분함
  - 간접 영향층의 취약성을 살필 때에는 사칭 여부를 확인하거나 결정의 이유를 설명받고, 잘못된 결과의 재검토를 요청할 수 있는 여건이 중요

□ 누구나 취약해질 수 있지만 피해 가능성은 같지 않음

○ 앞서 살펴본 피해 전환 요인 중 결정의 중요도와 사실을 확인하고 결과를 바로잡을 수 있는 여건은 AI를 능숙하게 쓰는 사람과 간접 영향층의 피해 가능성에도 영향을 미침

- 예를 들어 AI가 계약서를 요약하면서 위약금 조항을 빠뜨리고, 이용자가 그 요약만 믿고 서명하면 예상하지 못한 비용을 부담할 수 있음

○ 특히 기존에 디지털 서비스 이용이 어려웠던 사람은 AI 접근과 판단의 어려움에 더해 다른 경로로 확인하거나 도움을 구할 수단도 제한될 수 있어 더 많은 지원이 필요할 수 있음

○ 따라서 AI 시대의 취약계층에는 AI에 접근하거나 결과를 판단하기 어려운 이용자뿐 아니라, AI를 직접 이용하지 않더라도 그 영향을 받으며 사실 확인이나 설명 요청, 잘못된 결과의 정정에 어려움을 겪는 사람도 포함

- 우선 대응 대상도 직접 이용자에 한정하지 않고, 간접 영향층을 포함해 피해 전환 요인이 겹치는 상황을 중심으로 정할 필요가 있음

04 AI 시대 포용 정책의 확장 방향

[표 7] 포용 정책의 확장 방향

정책 요소	기존에 강조해 온 지원	AI 시대에 확대할 지원
지원 대상	디지털 서비스 이용에 어려움을 겪는 사람	AI 답변을 판단하기 어렵거나, AI가 만든 정보·결정의 영향을 받으면서 확인과 대응에 어려움을 겪는 사람까지 고려
이용자 지원	접근성 개선과 사용 방법 교육	답변의 근거 확인, 다른 의견 비교, 도움을 구하는 방법 교육
서비스 설계	접근성과 이용 편의 개선	중요한 행동 전 확인, 답변의 한계 안내, 이용자 통제 기능
결과에 대한 대응	서비스 이용 문의와 오류 신고	결정 이유 설명, 잘못된 정보의 정정, 사람에 의한 재검토

1 접근성 지원에서 판단을 돕는 지원으로

□ 기존 접근성 지원에 AI 판단 역량을 더하는 방향으로 정책 범위 확대

○ AI 포용 정책은 기기 보급과 접근성 개선, 이용 교육을 지속하면서 AI의 답변을 확인하고 자신의 상황에 맞게 활용할 수 있도록 지원 범위를 넓혀야 함

○ 교육 내용은 AI 사용법과 질문 작성 방법에 더해 답변의 근거를 확인하고 추가 검토나 도움이 필요한 상황을 구별하는 데 중점을 두어야 함

- 디지털배움터와 학교 교육 등에 출처 확인, 다른 자료와의 비교, 자신의 생각과 다른 관점 검토를 생활 속 사례로 연습하는 과정 포함
- AI에게 자신에 관해 어떤 정보를 기억하고 있는지 어떤 기준으로 답변하는지 직접 물어 확인하는 활동을 통해 AI가 이용자에게 맞춰 답하는 방식을 체험적으로 이해하도록 지원
- 건강이나 금전, 계약처럼 중요한 문제에서는 AI 답변만으로 결정을 내리기 전에 확인해야 할 사항과 전문가에게 도움을 구하는 방법을 함께 교육

※ UNESCO의 AI 역량 체계는 AI의 가능성과 한계를 이해하고, 비판적 사고와 인간의 주도성을 바탕으로 책임 있게 활용하는 역량을 강조(UNESCO, 2024)

□ 집단별 특성과 실제 이용 상황을 함께 고려해 지원 내용 구체화

○ 고령층과 장애인 등 AI기본법상 인공지능취약계층의 이용 어려움을 살피고, 연령이나 법령상 대상 구분으로 드러나지 않는 판단의 어려움과 지원 수요도 파악해야 함

- 결정의 중요도, 정서 상태, 확인·통제 여건 등 피해 전환 요인이 겹쳐 피해의 영향이 크거나 지원을 받기 어려운 상황에 우선 대응
- 청소년에게는 AI의 공감과 조언이 지닌 한계를 이해하고 중요한 고민을 신뢰할 수 있는 사람과 함께 검토하는 교육을 제공하며, 보호자와 교사의 대응 역량도 강화

※ 전문가 좌담에서는 학생과 직접 접촉하는 전문상담교사를 대상으로 한 AI 관련 연수가 사실상 없고, AI 대화의 폐쇄성 때문에 학생이 어떤 대화를 나누는지 외부에서 파악하기 어렵다는 점이 지적됨(국민일보, 2026.2.12.)

○ 교육 이후에도 이용 과정에서 겪는 어려움을 해결할 수 있도록 지자체와 교육기관, 복지기관이 협력하는 지원 체계를 갖추어야 함

- 가까운 기관이나 전화 상담을 통해 어려움을 설명하고, 상황에 따라 소비자 상담이나 전문상담 등 적절한 도움으로 이어질 수 있도록 안내 체계 마련
- 안전장치를 우회(jailbreaking)해 자살·자해 정보를 얻은 경우에는 즉시 전문기관으로 연계
- 정책 성과도 교육 참여 인원과 이용률에 더해, 실제 상황에서 답변을 확인하고 적절한 도움을 요청할 수 있게 되었는지까지 살펴 개선에 반영

2 이용자가 확인하고 바로잡을 수 있는 AI 서비스

□ 이용자의 판단을 돕는 교육과 함께 서비스의 보호 책임 강화

○ 2026년 1월 시행된 AI기본법은 생성형 AI 등의 서비스에 AI 기반 운용 사실의 고지와 결과물 표시 의무를 규정(제31조)

- 다만 현행 투명성 의무는 AI 기반 운용 사실과 결과물의 생성 여부를 알리는 데 초점이 있어, 답변의 근거와 한계를 확인하고 잘못된 결과를 바로잡도록 돕는 데까지는 이르지 않음
- 설명 방안과 이용자 보호 방안 등을 포함한 사업자 책무(제34조)는 고영향 AI나 이를 이용한 서비스에 적용되므로 일상적인 대화형 AI 서비스는 적용 대상이 아닐 수 있음

○ 따라서 AI 개발사와 운영기관은 이용자가 모든 위험을 스스로 알아차리고 대응하기 어렵다는 점을 고려해 답변을 확인하고 AI의 행동을 통제하기 쉬운 서비스를 설계해야 함

- 사실 확인이 필요한 답변에는 검증 가능한 출처를 연결하고, 근거가 부족하거나 추가 확인이 필요한 부분을 알아보기 쉽게 안내

※ 기기와 서비스의 개발 단계부터 AI 생성 콘텐츠의 신뢰도와 출처를 직관적으로 보여주는 'AI 리터러시 바이 디자인(AI literacy by design)' 관점이 제시된 바 있음(김태원, 2024)

- 결제나 계약 신청, 개인정보 전송 등 중요한 행동 전에는 대상과 금액, 제공되는 정보 등을 구체적으로 보여주고 이용자가 승인하거나 수정할 수 있도록 설계

○ 관계형 AI는 이용자의 연령과 대화 맥락을 고려해 정서적 의존과 개인정보 노출 위험을 줄이는 보호 장치를 갖추어야 함

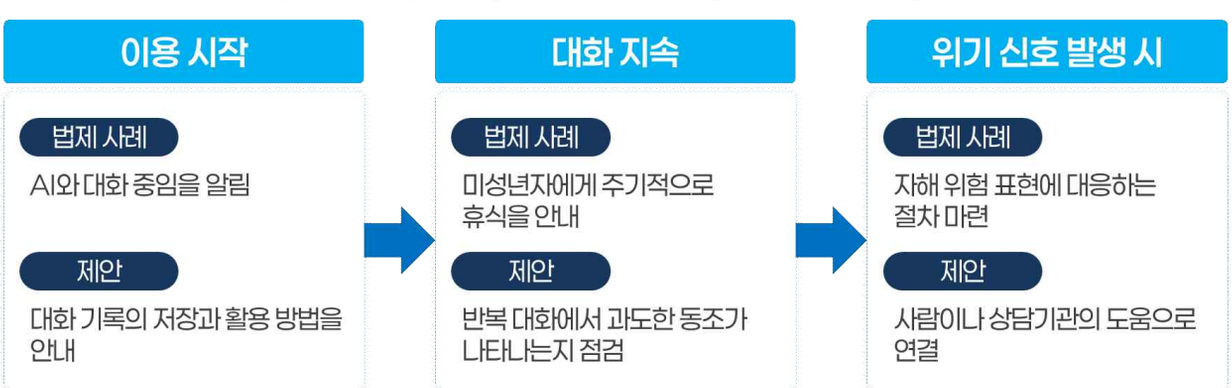
※ 국회입법조사처는 현행 청소년 보호 법제가 사후 대응 중심이라 일대일 챗봇 대화에 대응하기 어렵고 14~18세가 보호 사각지대에 있다며 사전예방 중심 규제로의 전환을 제안(김나정, 2026)

- AI와 대화 중임을 알리고 대화 기록의 저장, 활용 및 설정 변경 방법을 안내하며, 청소년이 위기 상황에 놓였을 때는 사람이나 상담기관의 도움으로 연결

- 서비스 출시 전과 운영 중에는 단일 응답뿐 아니라 여러 차례 이어지는 대화에서 과도한 동조가 나타나는지, 위기 신호에 적절히 대응하는지 점검

※ 한국의 사회문화적 맥락을 반영한 생성형 AI 안전성 평가 데이터셋 AssurAI가 공개됐으며, AI, 교육, 심리 분야 전문가가 정의한 35개 위험 요인에 '과의존 및 안전하지 않은 사용'이 포함돼 이용자 보호를 위한 서비스 점검에 활용 가능(Lim et al., 2025)

[그림 11] 관계형 AI 서비스의 이용 단계별 보호 장치



※ 캘리포니아주 SB 243의 AI 고지, 청소년 휴식 안내, 위기 대응 조치를 참고하고, 기록 관리와 반복 대화 점검은 보고서의 정책 제안으로 구성

□ AI의 영향을 받는 사람도 설명과 재검토를 요청할 수 있는 절차 마련

○ AI를 직접 이용하지 않는 간접 영향층은 교육을 통한 역량 강화만으로 보호하기 어려우므로 AI를 활용하는 기관이 설명과 재검토 절차를 갖추는 것이 핵심적인 보호 수단임

ㅇ 채용이나 복지 지원 등 중요한 결정에 AI를 참고만 한 기관도 당사자가 AI 활용 사실과 결정 이유를 알고 이의를 제기할 수 있는 절차를 마련해야 함

※「개인정보 보호법」 제37조의2는 개인정보를 처리한 자동화된 결정에 대한 설명 요구권과 일정 요건 아래 거부권을 규정하며, 사람이 실질적으로 판단에 개입한 경우에는 해당 조항의 적용 대상에서 제외될 수 있음

※ 전문직 영역에서도 상담자가 AI를 활용할 때는 이용자(내담자)에게 AI의 사용 영역을 구체적으로 밝히고, 이용자의 사전 동의를 요하는 윤리적 가이드라인이 마련되는 등 AI를 활용하는 쪽의 책임을 명확히 하는 흐름이 나타남(한국상담심리학회, 2025)

- AI를 직접 사용하지 않은 사람도 자신에 관한 잘못된 정보의 정정을 요청하고, 담당자에게 결정 내용을 다시 검토해 달라고 요구할 수 있도록 지원

- 온라인 신청이 어려운 사람도 전화나 대면으로 문의할 수 있게 하고, 재검토 담당자의 연락처, 처리 절차, 진행 상황 등을 쉽게 확인할 수 있도록 안내

ㅇ 정부와 공공기관은 AI 도입과 운영 평가에 이러한 보호 조치를 반영하고, 다양한 이용자가 실제로 이해하고 활용할 수 있는지 점검해야 함

- 보호 기능의 유무와 함께 이용자가 설명을 이해하고, 실행을 중단하며, 오류의 수정을 요청할 수 있는지 확인해 복잡하거나 이용하기 어려운 절차를 개선

- 이용자의 확인이나 동의가 있더라도 개발사와 운영기관의 오류 예방 및 대응책임을 명확히 해 이용자 역량 강화와 서비스 개선이 함께 이루어지도록 추진

3 맺음말

□ AI 시대의 취약계층은 AI를 쓰기 어려운 사람을 넘어 AI를 판단하고 바로잡기 어려운 사람까지

ㅇ AI 시대에는 기술에 접근하고 이용할 수 있는지에 더해, AI가 내놓은 결과를 판단하고 잘못된 결과를 바로잡을 수 있는지가 격차를 살피는 중요한 기준으로 부상

ㅇ 이에 AI 시대의 취약계층은 AI에 접근하기 어려운 사람뿐 아니라, 직접 이용 여부와 관계없이 AI의 결과를 판단하고 잘못된 결과에 대응하기 어려운 사람까지 포함하며, 이러한 취약성은 상황에 따라 누구에게나 나타날 수 있음

- 다만 기존 디지털 취약계층은 접근과 판단의 어려움에 더해 도움을 구하거나 결과를 바로잡을 수단도 부족할 수 있어 기존 지원에 판단을 돕는 지원까지 더해 두텁게 지원 필요

□ 포용 정책의 기준을 '접근'에서 '판단'과 '상황'으로 확장

○ AI 포용 정책은 접근성 지원을 지속하면서 판단 역량 교육과 판단을 돕는 서비스 설계를 더하고, AI가 관여하는 중요한 결정이나 정서적으로 취약한 상황에서 도움을 구하거나 결과를 바로잡기 어려운 경우에 우선 대응하는 방향으로 보완할 필요

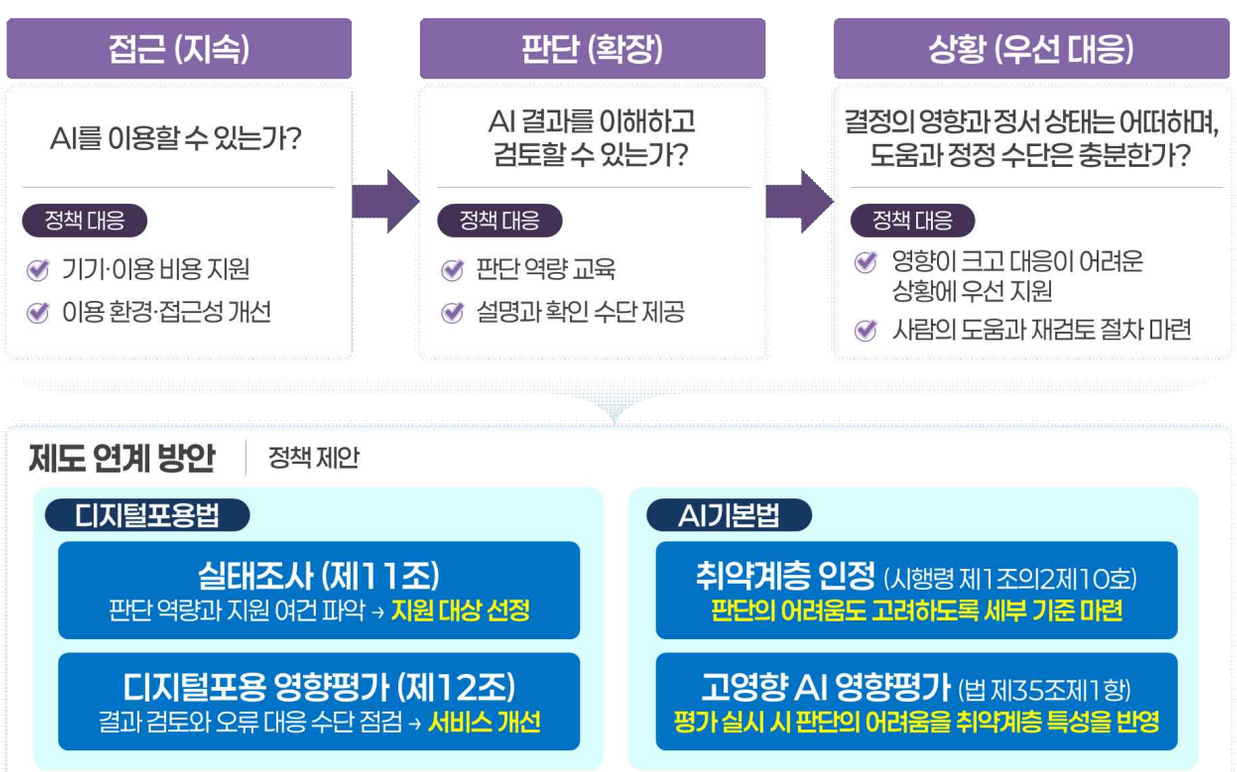
○ 디지털포용법과 AI기본법에 마련된 제도를 활용해 접근과 이용의 어려움뿐 아니라 AI 결과를 판단하고 잘못된 결과에 대응하는 어려움도 지원 대상 선정과 영향평가에 반영 필요

- 디지털포용법에 따른 실태조사(제11조)에는 AI 판단 역량과 주변의 지원 여건을 파악하는 문항을, 디지털포용 영향평가(제12조)에는 이용자가 AI 결과를 검토하고 오류에 대응할 수 있는 수단이 마련되어 있는지 점검하는 항목을 반영해 각각 지원 대상 선정과 서비스 개선에 활용

- AI기본법 시행령 제1조의2제10호의 적용 기준을 구체화할 때에는 AI 서비스 이용 과정에서 결과를 이해하고 검토하는 데 겪는 어려움을 살피고, 고영향 AI 영향평가를 실시하는 경우(법 제35조제1항)에는 이러한 어려움을 취약계층의 특성에 반영

※ 영어 중심의 안전성 평가 데이터셋은 한국어와 국내 사회문화적 위험을 포착하는 데 한계가 있으며(Lim et al., 2025), 26개 언어 비교 연구에서도 대화가 길어질수록 고자원 언어와 저자원 언어의 성능 격차가 커지는 것으로 나타남(Kim et al., 2025). 이에 취약계층 영향평가와 서비스 점검 기준에 국내 이용 맥락을 반영할 필요가 있음

[그림 12] AI 포용 정책의 세 가지 기준과 대응 방향



참고자료

〈 국내 자료 〉

1) 과학기술정보통신부·한국지능정보사회진흥원(2026a),『2025년 디지털정보격차 실태조사』.

2) KBS 뉴스, “앱 못 쓰면 택시도 못 타나…‘길 위에 갇힌’ 93세 할머니”, https://news.kbs.co.kr/news/pc/view/view.do?ncd=8574980, 2026.6.1.

3) 연합뉴스, “그는 결국 햄버거를 사 먹지 못했다”, https://www.yna.co.kr/view/AKR20260610141900011, 2026.6.14.

4) 한국경제, “앱 없이 택시 15만번 불렀다…어르신들 '02-114' 누른 까닭”, https://n.news.naver.com/mnews/article/015/0005319010?sid=105, 2026.8.10.

5) 국토교통부(2024), “‘번호판 인식방식’ 스마트톨링 시범사업, 하이패스 없이도 고속도로 통행료 무정차 납부”, 보도자료, 2024.5.26.

6) 천안논산고속도로 홈페이지, https://www.cneway.co.kr/sub/info.do?m=030504&s=cneway.

7) 과학기술정보통신부·한국지능정보사회진흥원(2026b),『2025년 인터넷이용실태조사』.

8) 김태원(2024),『AI 일상화 시대의 위협, 딥페이크 대응 방안』, The AI Report, 2024-13, 한국지능정보사회진흥원.

9) 경찰청(2024), "딥페이크 이용, '자녀 납치했다' 사기 주의", 보도자료, 2024.11.7.

10) TV조선, “[핫라인] 손흥민이 ‘우량주 추천해드립니다’…국회에도 등장한 ‘딥페이크’”, https://www.youtube.com/watch?v=O6m5M04W4-g, 2024.8.29.

11) 초록우산(2026),『내 친구와의 위험한 대화 : 생성형 AI 챗봇 안전설계를 위한 이슈브리프』.

12) 스캐터랩 제타 앱 소개, https://play.google.com/store/apps/details?id=com.scatterlab.messenger (검색일 2026. 9. 28.).

13) 조선일보, “관계형 AI와 시간 보내는 젊은 층… 게임사는 '숏폼 플랫폼'으로 공략”, https://www.chosun.com/economy/tech_it/2026/08/20/CGVCNVOL7RD5XPGCFYR4BVZ6OU, 2026.8.20.

14) YTN, “3살도 가입되는 제타 AI...'성적 표현' 그대로 노출”, https://www.ytn.co.kr/_ln/0102_202510102251080004, 2025.10.10.

15) 법제처 국가법령정보센터, 「지능정보화 기본법」, 제2조제13호 및 제45조, https://www.law.go.kr/법령/지능정보화기본법.

16) 법제처 국가법령정보센터, 「디지털포용법」, 제2조제1항제3호, 제11조, 제12조, 제14조제3항, 제19조, 제24조 및 부칙 제6조, https://www.law.go.kr/법령/디지털포용법.

17) 법제처 국가법령정보센터, 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(법률 제21311호, 2026.1.20. 일부개정), 제3조제5항, 제17조의2, 제31조, 제34조 및 제35조제1항, https://www.law.go.kr/법령/인공지능발전과신뢰기반조성등에관한기본법.

18) 법제처 국가법령정보센터, 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령」, 제1조의2, https://www.law.go.kr/법령/인공지능발전과신뢰기반조성등에관한기본법시행령.

19) 김나정(2026), 『AI 챗봇의 위험에 노출된 아동·청소년: 사전예방 중심 규제체제 도입의 필요성과 방향성』, 이슈와 논점 제2515호, 국회입법조사처, 2026.7.20.

20) 법제처 국가법령정보센터, 「개인정보 보호법」, 제37조의2, https://www.law.go.kr/법령/개인정보보호법.

21) KBS 뉴스, "'네 말이 맞아' AI가 아첨…실험결과 봤더니", https://news.kbs.co.kr/news/pc/view/view.do?ncd=8642406, 2026.8.21.

22) KBS 뉴스, "'AI와 사랑에 빠졌어요'…14살 중학생의 고백 [AI과몰입]①", https://news.kbs.co.kr/news/pc/view/view.do?ncd=8651663, 2026.9.1.

23) 국민일보, "[단독] 해외서 발견된 'AI 사망 이상 징후', 국내 10대서 다수 포착", https://www.kmib.co.kr/article/20260207500096, 2026.2.7.

24) 국민일보, "'아동·청소년 등 AI 취약층 보호·안전 장치 필요'", https://www.kmib.co.kr/article/view.asp?arcid=1770883778, 2026.2.12.

25) 한국상담심리학회(2025), 「심리상담 AI 활용 윤리 가이드라인」, 2025.5.16.

〈 국외 자료 〉

1) Google(2025), "New AI and accessibility updates across Android, Chrome and more", Google Blog, https://blog.google/company-news/outreach-and-initiatives/accessibility/android-gemini-ai-gaad-2025/, 2025.5.15.

2) Google(2024), “Gemini breaks new ground with a faster model, longer context, AI agents and more”, Google Blog, https://blog.google/innovation-and-ai/products/google-gemini-update-flash-ai-assistant-io-2024/, 2024.5.14.

3) Lee, H.-P. H., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N.(2025), “The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers”, CHI 2025, https://doi.org/10.1145/3706598.3713778, 2025.4.25.

4) Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A.(2019), “Students' Civic Online Reasoning: A National Portrait”, Stanford History Education Group & Gibson Consulting, https://purl.stanford.edu/gf151tb4868, 2019.11.14.

5) GPTZero(2026), “Chasing the Hallucinations: PwC report hallucinates product and government customers”, Esau, P., Ogale, O., & Cui, A., https://gptzero.me/news/investigations-pwc/, 2026.7.28.

6) Landberg v City of New York, 2026 NY Slip Op 03935, Appellate Division, Second Department, 2026. 6. 23.

7) Sharma, M. et al.(2023), “Towards Understanding Sycophancy in Language Models”, https://arxiv.org/abs/2310.13548, 2023.10.20.

8) OpenAI(2025), “Expanding on what we missed with sycophancy”, https://openai.com/index/expanding-on-sycophancy/, 2025.5.2.

9) Zhang, Y., Zhao, D., Hancock, J. T., Kraut, R., & Yang, D.(2026), “Interaction with AI companions and psychological well-being.” Nature Human Behaviour, https://doi.org/10.1038/s41562-026-02516-2.

10) Euronews(2023), “Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change”, https://www.euronews.com/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-, 2023.3.31.

11) GPDP(2023), “Intelligenza artificiale, dal Garante privacy stop al chatbot ‘Replika’. Troppi i rischi per i minori e le persone emotivamente fragili”, https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9852506, 2023.2.3.

12) Tech Justice Law(2024), Garcia v. Character Technologies, Google, and Character AI co-founders, Daniel De Freitas and Noam Shazeer, https://techjusticelaw.org/cases/garcia-v-character-technologies-google-and-character-ai-co-founders-daniel-de-frietas-and-noam-shazeer/, 2024.10.22.

13) CNN(2026), "Character.AI and Google agree to settle lawsuits over teen mental health harms and suicides", https://www.cnn.com/2026/01/07/business/character-ai-google-settle-teen-suicide-lawsuit, 2026.1.7.

14) Raine v. OpenAI, Inc., Case No. CGC-25-628528, San Francisco County Superior Court, 2025.8.26.

15) ICO(2024), "We warn organisations must not ignore data protection risks as we conclude Snap 'My AI' chatbot investigation", https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2024/05/ico-warns-organisations-must-not-ignore-data-protection-risks-as-it-concludes-snap-my-ai-chatbot-investigation, 2024.5.21.

16) OECD·European Commission(2026), "Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education", https://www.oecd.org/en/publications/empowering-learners-for-the-age-of-ai_65cd27d4-en.html, 2026.6.18.

17) Casey, B. J., Jones, R. M., & Hare, T. A.(2008), "The adolescent brain", Annals of the New York Academy of Sciences, 1124, 111-126, https://doi.org/10.1196/annals.1440.010.

18) Common Sense Media(2025), "Talk, trust, and trade-offs: How and why teens use AI companions", Robb, M. B., & Mann, S., https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf.

19) NIST(2023), "Appendix A: Descriptions of AI Actor Tasks", https://airc.nist.gov/airmf-resources/airmf/appendices/app-a-descriptions-of-ai-actor-tasks/.

20) UNESCO(2024), "AI competency framework for students", Miao, F., Shiohira, K., & Lao, N., https://doi.org/10.54675/JKJB9835.

21) California State Legislature(2025), "SB 243: Companion chatbots", Chapter 677, https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB243.

22) Cheng, M., Yu, S., Lee, C., Khadpe, P., Ibrahim, L., & Jurafsky, D.(2025), “ELEPHANT: Measuring and understanding social sycophancy in LLMs”, https://arxiv.org/abs/2505.13995, 2025.5.20.

23) Kim, P., Xie, Y., & Yang, S.(2025), “‘I am here for you’: How relational conversational AI appeals to adolescents, especially those who are socially and emotionally vulnerable”, https://arxiv.org/abs/2512.15117, 2025.12.17.

24) Ben-Zion, Z., Witte, K., Jagadish, A. K., et al.(2025), “Assessing and alleviating state anxiety in large language models”, npj Digital Medicine, 8, 132, https://doi.org/10.1038/s41746-025-01512-6.

25) Coda-Forno, J., Witte, K., Jagadish, A. K., Binz, M., Akata, Z., & Schulz, E.(2024), “Inducing anxiety in large language models can induce bias”, https://arxiv.org/abs/2304.11111.

26) Hudon, A., Bhat, V., Nunez, N., Paquin, V., Malouin-Lachance, A., & Lespérance, P.(2026), “The AWARE (AI Use, Why, Attachment, Reality and Risk, and Effect on Functioning) Framework for Assessing Patients’ Use of AI in Mental Health Care”, JMIR Medical Education, https://doi.org/10.2196/108766, 2026.9.16.

27) Lim, C.-G., Han, S.-H., Byun, E., Han, J., Cho, S., et al.(2025), “AssurAI: Experience with Constructing Korean Socio-cultural Datasets to Discover Potential Risks of Generative AI”, https://arxiv.org/abs/2511.20686, 2025.11.20.

28) Kim, Y., Russell, J., Karpinska, M., & Iyyer, M.(2025), “One ruler to measure them all: Benchmarking multilingual long-context language models”, https://arxiv.org/abs/2503.01996, 2025.3.3.

[AI Policy Insights | 2026 Vol. 05]

누가 AI 시대의 취약계층인가?

: 디지털 접근 격차를 넘어 AI를 판단하는 역량의 격차로

- 발 행 일 : 2026년 11월
- 발 행 인 : 김형철
- 발 행 처 : 한국지능정보사회진흥원
- 주　　소 : 대구광역시 동구 첨단로 53
- 대표전화 : 053-230-1114
- 보고서 온라인 서비스 : www.nia.or.kr

1. 『AI Policy Insights』는 새로운 기술의 발전 동향과 경제·사회·정책·법제 이슈, 산업과 기업의 혁신 사례를 폭넓게 분석하고, 정책 시사점을 제시하기 위해 한국지능정보사회진흥원(NIA)에서 기획·발간하는 심층 분석 보고서 시리즈입니다.
2. 본 보고서는 방송통신발전기금으로 수행한 정보통신·방송 연구개발 사업의 결과물이므로, 보고서 내용을 발표할 때는 반드시 「과학기술정보통신부 정보통신·방송연구개발사업」의 연구 결과임을 밝혀야 합니다.
3. 본 보고서 내용의 무단전재를 금하며, 가공·인용할 때는 반드시 출처를 「한국지능정보사회진흥원(NIA)」이라고 밝혀 주시기 바랍니다.
4. 본 보고서의 내용은 한국지능정보사회진흥원(NIA)의 공식 견해와 다를 수 있습니다.