

과학기술&ICT 정책·기술 동향

Science, ICT Policy and Technology Trends



CONTENTS

I 이슈 분석	1	2. ICT	35
폰 데이터 라이선스 2기 EU의 기후변화 대응 및 과학기술 경쟁력 강화 논의: 청정산업계획 및 방위산업 경쟁력을 중심으로		능동형 인공지능 시대를 여는 열쇠, AI 에이전트	35
		AI 환각(Hallucination), 산업계의 신뢰성 위기와 대응 전략	39
		랜섬웨어 공격 증가와 중소기업의 사이버 위협 증대	42
		중국, 알리바바 AI 플랫폼 진단모델 FDA 혁신 기기 지정	45
		일본, 차량용 반도체 공동개발 '기술격차 극복'	48
II 주요 동향	14	III 단신 동향	50
1. 과학기술	14	1. 해외	50
미국 백악관, 생물안보 및 제약 제조 기반 강화를 위한 행정명령 발표	14	2. 국내	59
CSIS, 데이터센터 전력 확보를 위한 4개국의 전략 비교	16		
일본 특허청, AI 기술과 관련한 발명 보호의 주요 쟁점 검토	18	IV 주요 통계	64
중국 국가에너지국, '수소에너지 발전 보고서 2025' 발표	20		
EU 집행위원회, '호라이즌 유럽' 중간 성과 평가 보고서 발간	23		
OECD, 기업의 AI 도입을 유도하기 위한 정책 방향 제시	25		
IEA, 2025 글로벌 전기차 전망 보고서 발간	29		
OECD, 시민참여 촉진을 위한 신흥 디지털 기술의 활용 가능성 검토	32		



- 과학기술&ICT 정책 · 기술 동향 보고서는 한국과학기술기획평가원 기관고유사업의 일환으로 추진되고 있으며, 과학기술정보통신부의 지원 및 정보통신기획평가원(IITP)의 협조를 통해 발간되고 있습니다.
- 관련 자료는 <https://www.kistep.re.kr/gps/>를 통해서도 서비스를 이용할 수 있으며, 보고서 내용에 대한 문의는 아래와 같이 주시기 바랍니다.

과학기술
동향

 **KISTEP** 한국과학기술기획평가원
Korea Institute of S&T Evaluation and Planning
TEL: 043-750-2481
E-mail: wona@kistep.re.kr

ICT 동향

 **IITP** 정보통신기획평가원
Institute of Information & Communications
Technology Planning & Evaluation
TEL: 042-612-8240
E-mail: itzme@iitp.kr



2 AI 환각(Hallucination), 산업계의 신뢰성 위기와 대응 전략

→ 오픈AI 신규 모델의 환각 문제와 산업계 파장

- 최신 추론형 AI 모델의 환각 현상 심화, o4 미니 모델 환각률(Hallucination Rate) 48%로 이전 모델 대비 급증
 - 2025년 4월 오픈AI(Open AI)가 출시한 추론형 AI 모델 ‘o3’와 ‘o4 미니’는 코딩 및 시각적 추론 능력 향상에도 불구하고 환각률 대폭 증가
 - 오픈AI는 “모델이 이전보다 더 많은 사용자 요청에 응답하게 되면서, 정확한 결과뿐 아니라 잘못된 결과도 함께 증가”한 것이라고 분석
- 성능 향상과 환각률 증가의 역설, 고도화된 추론 능력이 오히려 신뢰성 문제 야기
 - o3와 o4 미니는 화이트보드 그림, 도표, 그래프 등 시각 정보를 분석하고 회전된 이미지도 처리하는 등 향상된 능력을 구현하여 사용자 경험 측면에서 획기적 발전
 - AI 전문가들은 향상된 시각 정보 처리와 추론 능력이 역설적으로 모델의 ‘과신(overconfidence)’ 문제를 악화시켰을 가능성 제기
- 산업계 영향 : 추론형 AI 모델에 대한 신뢰성 위기 및 법률, 회계 등 고정확도 요구 분야 도입 위기
 - 기업들은 AI 활용 전략을 재검토하고 검증 절차를 강화하고 있으나, 환각 문제의 근본적 해결에는 한계가 있다는 지적
 - 소비자 인식 변화에 따라 향후 AI 서비스의 평가는 정확성이 핵심 기준으로 부상할 것으로 예상됨

→ AI 환각 현상의 발생 요인 및 배경

- (기술적 배경) 대규모 언어 모델(LLM)의 근본적 작동 원리와 추론 능력 향상의 딜레마
 - 대규모 언어 모델은 ‘다음에 올 가능성이 높은 단어 예측’이라는 확률적 텍스트 생성 방식에 기반하여 작동하며, 문맥상 자연스러우나 사실과 다른 정보가 생성될 가능성 존재
 - o3, o4 미니와 같은 시각 정보 통합 모델은 다양한 입력 형식을 처리하며 더 복잡한 추론이 가능하나 이 과정에서 오류 전파 및 증폭 가능성 증가
- (데이터 기반 배경) 학습 데이터의 한계와 지식 표현 문제

- 대규모 언어 모델이 학습하는 인터넷 데이터에는 오류, 허위 정보, 모순된 내용이 다수 포함되어 있어 모델도 이를 학습하게 되는 문제 발생 가능성 상존
- AI 모델은 '사실'과 '허구'를 직접 구분하는 능력 없이 단어 간 통계적 연관성만을 학습하여, 사실이 아닌 정보를 '환각'할 수 있음
- (산업적 배경) 기업 간 경쟁 심화와 성능 위주 개발 관행
 - 오픈AI, 앤트로픽(Anthropic), 구글(Google) 등 주요 AI 기업 간 경쟁 격화로 신속한 모델 출시 압박이 커지고 충분한 안전성 및 정확성 검증 시간 부족 문제 발생 가능
 - 대규모 언어 모델 개발에 소요되는 막대한 비용(수천만~수억 달러)으로 인해 환각 문제 해결보다 기능 확장에 투자 집중
- (사회적 배경) 사용자 기대와 AI 한계 간의 간극 확대
 - ChatGPT 등 대화형 AI의 광범위한 보급으로 사용자들의 AI 능력에 대한 과도한 신뢰와 기대가 형성되어 환각 현상에 대한 인식과 경계심 부족
 - 디지털 리터러시 격차로 인한 사용자들의 AI 출력 결과의 무비판적 수용 경향과 전통 미디어 쇠퇴 및 AI 기반 콘텐츠 생성 증가에 따른 잘못된 정보의 대규모 확산 위험의 고조

→ AI 환각 대응 기술 개발 및 도입 사례

- 주요 AI 기업의 환각 대응 전략 및 기술 구현 사례
 - 오픈AI는 'o3'와 'o4 미니' 모델에 검색 기반 지식 증강(RAG)과 다중 검증 시스템을 적용함으로써 환각 문제 대응
 - 앤트로픽은 클로드(Claude) 모델에 자체 검증 메커니즘을 도입하여 사실 기반 응답의 강화 및 전반적인 신뢰성 향상을 추구하는 전략을 전개
 - 구글은 제미니(Gemini) 모델에 실시간 검색 연동 기능과 함께 불확실성 표현 기능을 통합하여, 환각 발생 가능성을 사전에 제어하고 정보 신뢰도를 제고하는 방식의 대응 추진
- 산업별 환각 대응 전략 및 적용 사례
 - (의료 분야) AI 모델이 제공하는 정보에 대한 의료 전문가의 검증 프로세스를 도입하고, 의학 지식 데이터베이스와 연동하여 환각으로 인한 위험을 최소화
 - (법률 분야) AI가 생성한 법적 정보에 대해 판례 및 법령 데이터베이스와의 자동 교차 검증 시스템을 구축하여 존재하지 않는 법률이나 판례 인용 문제 방지
 - (교육 분야) 학생들에게 AI 환각 탐지 능력을 가르치는 디지털 리터러시 프로그램을 도입하고, AI 학습 도구에 사실 검증 메커니즘을 추가하는 방식으로 대응



➔ AI 환각 관련 한계점 및 발전 과제

- 환각 탐지 기준의 부재와 도메인별 성능 편차, 불확실성 처리에 따른 모델 활용성 저하 우려
 - AI 환각을 완벽하게 탐지하기 위한 ‘최적 표준(gold standard)’ 부재, 특히 새로운 정보나 최신 사건에 대한 판단이 어려운 상황에서 환각과 정확한 추론을 구분하는 경계 모호
- AI 환각 대응을 위한 전문 인력 부족과 교육 체계 미비
 - 도메인 지식(의학, 법률 등)과 AI 기술 이해를 동시에 갖춘 ‘하이브리드 전문가’ 수요 급증에도 불구하고 인력 양성 체계 미비로 인한 시장 불균형 심화
- 사용자 인식 제고의 한계와 접근성 확보 과제
 - 환각 위험과 대응 방법에 대한 대중 교육 프로그램 및 사용자 친화적 도구 부족, AI 리터러시 격차로 인한 디지털 불평등 심화 우려

출처 : ZDNet Korea 외 (2025.4.)

<https://arxiv.org/pdf/2401.06796>

<https://cdn.openai.com/gpt-4-5-system-card-2272025.pdf>

https://documentation.suse.com/suse-ai/1.0/pdf/AI-preventing-hallucinations_en.pdf

<https://techcrunch.com/2025/04/18/openais-new-reasoning-ai-models-hallucinate-more/>

<https://futurism.com/openai-admits-gpt45-hallucinates>

<https://research.aimultiple.com/ai-hallucination/>

<https://au.finance.yahoo.com/news/openais-reasoning-ai-models-hallucinate-210944177.html>

<https://www.ynetnews.com/business/article/rjqvyk7jlg>

<https://www.insidehook.com/internet/openais-new-models-have-hallucination-problem>

<https://zdnet.co.kr/view/?no=20250420122338>