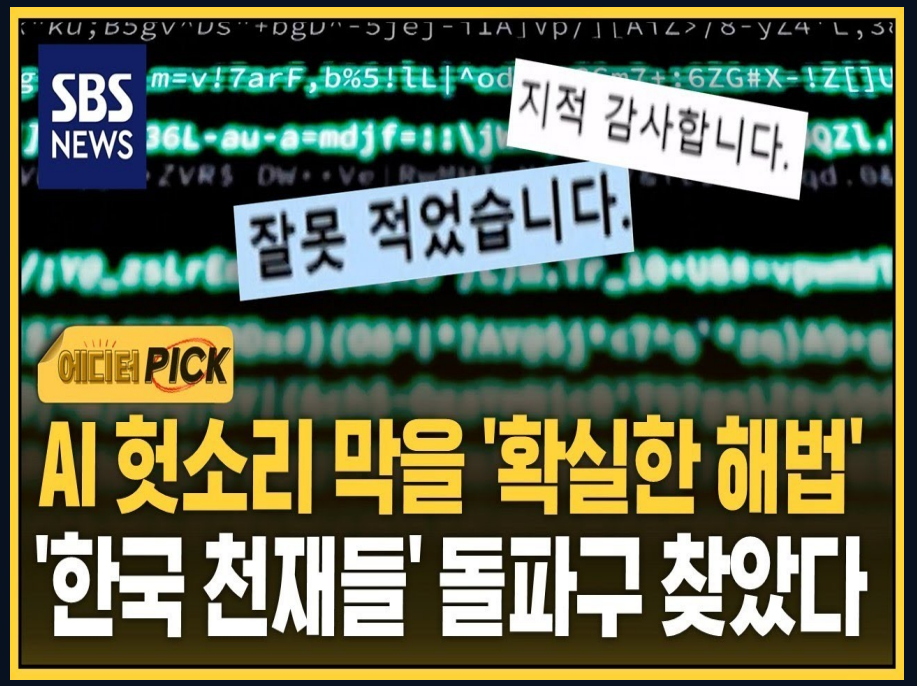
DREAMLABS COLLECTOR

AI 헛소리 막을 ' 확실한 해법 '..' 한국 천재들 ' 돌파구 찾았다 (에디터픽) / SBS



SBS 뉴스 보도에 따르면 , 국내 연구팀이 AI 의 ' 할루시네이션 (환각)' 현상 원인을 규명하고 이를 방지할 수 있는 돌파구를 찾았습니다 . 연구팀은 AI 가 학습 전 ' 아무것도 모르는 상태 ' 를 먼저 학습하도록 하는 예열 단계를 도입하여 , AI 가 모르는 것에 대해 과신하지 않고 ' 모른다 ' 고 답하게 만들었습니다 . 이 방법은 기존 대규모 학습 AI 에는 적용하기 어렵지만 , 새로 개발되는 대화형 AI 나 자율주행 AI 등에 도입 가능성이 제시되었습니다 . 해당 연구 결과는 국제 학술지 ' 네이처 머신 인텔리전스 ' 에 게재되어 그 중요성을 인정받았습니다 . 이 기술은 AI 의 신뢰성을 높여 다양한 분야에서의 활용도를 증대시킬 잠재력을 가집니다 .

CHANNEL

SBS 뉴스

VIDEO ID

dcsvMlyhhAI

source: https://www.youtube.com/watch?v=dcsvMlyhhAI | video_id: dcsvMlyhhAI | generated: 2026-06-05T16:09:45Z

Executive Summary

영상 시청 전 빠른 정보 습득을 위한 요약

SUMMARY

SBS 뉴스 보도에 따르면, 국내 연구팀이 AI의 '할루시네이션(환각)' 현상 원인을 규명하고 이를 방지할 수 있는 돌파구를 찾았습니다. 연구팀은 AI가 학습 전 '아무것도 모르는 상태'를 먼저 학습하도록 하는 예열 단계를 도입하여, AI가 모르는 것에 대해 과신하지 않고 '모른다'고 답하게 만들었습니다. 이 방법은 기존 대규모 학습 AI에는 적용하기 어렵지만, 새로 개발되는 대화형 AI나 자율주행 AI 등에 도입 가능성이 제시되었습니다. 해당 연구 결과는 국제 학술지 '네이처 머신 인텔리전스'에 게재되어 그 중요성을 인정받았습니다. 이 기술은 AI의 신뢰성을 높여 다양한 분야에서의 활용도를 증대시킬 잠재력을 가집니다.

source: https://www.youtube.com/watch?v=dcsvMlyhhAI | video_id: dcsvMlyhhAI | generated: 2026-06-05T16:09:45Z

Video Structure

영상 구성과 논리 흐름

01

AI 할루시네이션 문제 제기 및 사례 (이정후 홈런, 세종대왕 맥북)

02

일반인 인터뷰를 통한 문제점 인식 (정확성, 불명확한 인용)

03

국내 연구팀의 할루시네이션 원인 규명 (과신 문제)

04

연구팀의 해결책 제시 (예열 단계, '아무것도 모른다' 학습)

05

해결책의 효과 및 적용 가능 범위 설명 (과신 제거, 신규 AI 적용)

06

연구 결과의 학술지 게재 및 의의

source: https://www.youtube.com/watch?v=dcsvMlyhhAI | video_id: dcsvMlyhhAI | generated: 2026-06-05T16:09:45Z

Key Ideas

정보게시물로 전환할 핵심 아이디어

01

AI 할루시네이션 : AI 가 허위 정보를 사실처럼 확신하여 답변하는 현상 .

02

과신 (Overconfidence): AI 가 근거 없이 특정 답변에 높은 확률을 부여하여 잘못된 정보를 확신하는 경향 .

03

예열 단계 (Warm-up Phase): AI 가 본격적인 학습에 앞서 무의미한 데이터를 학습하여 ' 아무것도 모르는 상태 ' 를 인지하게 하는 과정 .

04

' 모른다 ' 고 답하는 능력 : 예열 단계를 통해 AI 가 불확실한 정보에 대해 솔직하게 모른다고 인정하는 능력 .

05

신규 AI 적용 가능성 : 기존 대규모 학습 AI 보다 새로 개발되는 대화형 AI 및 자율주행 AI 에 효과적인 적용이 기대됨 .

06

인간 두뇌 발달 착안 : 인간의 인지 발달 과정에서 영감을 얻어 AI 학습 방법을 개선 .

source: https://www.youtube.com/watch?v=dcsvMlyhhAI | video_id: dcsvMlyhhAI | generated: 2026-06-05T16:09:45Z

DreamLabs Application

DreamLabs 내부 적용 관점

01

DreamLabs 의 신규 AI 모델 개발 시, 초기 학습 단계에 '예열 단계'를 도입하여 할루시네이션 현상을 선제적으로 방지하는 방안 검토.

02

자율주행 AI 또는 고신뢰성이 요구되는 AI 시스템 개발 시, 본 연구의 '모른다'고 답하는 능력을 통합하여 안전성 및 신뢰도 향상.

03

내부 AI 기반 리서치 및 분석 도구의 답변 신뢰도를 높이기 위한 후처리 또는 재학습 방안 연구.

04

AI 모델의 '과신' 정도를 측정하고 제어하는 내부 평가 지표 및 방법론 개발에 본 연구의 원리 활용.

05

AI 윤리 및 책임성 (AI Ethics & Accountability) 프레임워크에 '불확실성 인지 및 표현' 기능을 포함하는 연구.

06

KAIST 등 국내 연구기관과의 협력을 통해 AI 할루시네이션 방지 기술의 추가적인 발전 및 DreamLabs 제품 적용 가능성 탐색.

source: https://www.youtube.com/watch?v=dcsvMlyhhAI | video_id: dcsvMlyhhAI | generated: 2026-06-05T16:09:45Z

Verification Required

모델 추론 /metadata 한계 / 원본 확인 필요

01

연구팀의 정확한 구성원 및 소속 기관 (KAIST 백세범 교수 외 다른 연구자 확인 필요).

02

네이처 머신 인텔리전스에 게재된 논문의 정확한 제목, 저자 목록, 초록 및 전문 검토를 통한 상세 방법론 및 결과 확인.

03

예열 단계에 사용된 '무의미한 데이터'의 구체적인 정의 및 학습 방식에 대한 기술적 세부 사항.

04

해당 기술이 '새로 개발되는 대화형 AI 나 자율주행 AI' 에 적용될 때의 구체적인 성능 향상 지표 및 한계점.

05

기존 대규모 학습 AI 에 적용하기 어려운 이유에 대한 기술적 설명 및 우회 적용 가능성 여부.

06

이정후 홈런 개수, 세종대왕 맥북 등 할루시네이션 사례에 대한 연구팀의 실제 실험 결과 및 개선 효과 데이터.

source: https://www.youtube.com/watch?v=dcsvMlyhhAI | video_id: dcsvMlyhhAI | generated: 2026-06-05T16:09:45Z

Source & Download Metadata

게시물과 문서 산출물 추적 정보

METADATA

Title: AI 헛소리 막을 ' 확실한 해법 '..' 한국 천재들 ' 돌파구 찾았다 (에디터픽) / SBS
Channel: SBS 뉴스
Video ID: dcsvMlyhhAI
Source URL: https://www.youtube.com/watch?v=dcsvMlyhhAI
Playlist ID: PLHwM6idVO2zyqi2IZeDAiP5QBqRXd2Zyh
Generated at: 2026-06-05T16:09:45Z
Source basis: metadata_and_model_inference