

클로드 아버지의 충격 고백 ” 우리도 모른다 “전세계 AI 천재들이 경악한 현대 AI 의 근본 문제 I 딥토크 EP 14

이 영상은 워싱턴대 이수인 교수님과 딥토크를 통해 현대 AI 의 근본적인 문제인 '블랙박스' 현상과 AI 해석 불가능성에 대해 다룹니다. 이수인 교수님은 '설명 가능한 AI(XAI)' 분야의 선구자로서, AI 의 의사결정 과정을 투명하게 이해하기 위한 핵심 기술인 'SHAP' 을 개발했습니다. 영상은 AI 개발자조차 내부 작동 원리를 완전히 알지 못하는 현실을 지적하며, 앤트로픽 CEO 의 경고와 챗 GPT 설명의 함정을 예시로 들어 AI 신뢰성 문제를 심층적으로 탐구합니다. 궁극적으로 AI 를 100% 신뢰하고 활용할 수 있는 미래를 위한 XAI 의 중요성과 그 적용 사례를 제시합니다. 또한, AI 를 활용한 치매 정복이라는 야심찬 포부와 AI 시대의 필수 마인드에 대해서도 논의합니다.

CHANNEL

SEIZE 씨즈

VIDEO ID

kV9H9rZvLK0



Executive Summary

영상 시청 전 빠른 정보 습득을 위한 요약

SUMMARY

이 영상은 워싱턴대 이수인 교수님과 딥토크를 통해 현대 AI의 근본적인 문제인 '블랙박스' 현상과 AI 해석 불가능성에 대해 다룹니다. 이수인 교수는 '설명 가능한 AI(XAI)' 분야의 선구자로서, AI의 의사결정 과정을 투명하게 이해하기 위한 핵심 기술인 'SHAP'을 개발했습니다. 영상은 AI 개발자조차 내부 작동 원리를 완전히 알지 못하는 현실을 지적하며, 앤트로픽 CEO의 경고와 챗 GPT 설명의 함정을 예시로 들어 AI 신뢰성 문제를 심층적으로 탐구합니다. 궁극적으로 AI를 100% 신뢰하고 활용할 수 있는 미래를 위한 XAI의 중요성과 그 적용 사례를 제시합니다. 또한, AI를 활용한 치매 정복이라는 야심찬 포부와 AI 시대의 필수 마인드에 대해서도 논의합니다.

Video Structure

영상 구성과 논리 흐름

01

AI 블랙박스 : 개발자조차 AI 내부 작동 원리를 모르는 이유 (02:15)

02

AI 파라미터와 인류가 AI 를 이해하지 못할 때 생기는 일 (03:25, 05:26)

03

'설명 가능한 AI(XAI)'의 선구자 , 이수인 교수 소개 및 삼성호암상 수상 후기 (06:17, 07:03)

04

설명 가능한 AI(XAI)의 개념과 엔트로픽 CEO의 경고 , 챗 GPT 설명의 함정 (09:54, 13:12, 16:00)

05

이수인 교수가 개발한 AI 해석 기법 'SHAP'과 새롭리 별류의 원리 및 적용 사례 (17:38, 19:59, 22:56)

06

일반 사용자의 XAI 기법 활용 가능성 및 생성형 AI 100% 신뢰 시험 논의 (25:51, 26:33)

Key Ideas

정보게시물로 전환할 핵심 아이디어

01

****AI 블랙박스 문제****: 현대 AI, 특히 딥러닝 모델은 복잡한 내부 구조로 인해 의사결정 과정을 추적하기 어려운 '블랙박스' 특성을 가집니다. (메타데이터 기반 추론)

02

****설명 가능한 AI (XAI)****: AI의 예측이나 결정이 왜 그렇게 나왔는지 인간이 이해할 수 있도록 설명하는 기술 및 연구 분야입니다. (메타데이터 기반 추론)

03

****SHAP (SHapley Additive exPlanations)****: 이수인 교수가 개발한 AI 해석 기법으로, 각 입력 특성이 모델의 예측에 얼마나 기여했는지 정량적으로 평가합니다. (메타데이터 기반 추론)

04

****새롭리 벨류****: 게임 이론에서 각 플레이어의 기여도를 공정하게 분배하는 개념으로, SHAP의 수학적 기반이 됩니다. (메타데이터 기반 추론)

05

****AI 신뢰성 문제****: AI의 불투명성은 의료, 금융 등 고위험 분야에서의 적용에 대한 신뢰성 문제를 야기하며, 엔트ropic CEO도 이 점을 경고했습니다. (메타데이터 기반 추론)

06

****생성형 AI의 한계****: 챗 GPT와 같은 생성형 AI가 제공하는 설명은 때때로 피상적이거나 잘못된 정보를 포함할 수 있어 주의가 필요합니다. (메타데이터 기반 추론)

DreamLabs Application

DreamLabs 내부 적용 관점

01

****AI 모델 투명성 확보****: DreamLabs 에서 개발하거나 활용하는 AI 모델 (예 : 추천 시스템 , 데이터 분석 모델) 의 의사결정 과정을 SHAP 과 같은 XAI 기법으로 분석하여 투명성과 신뢰성을 높일 수 있습니다 .

02

****연구 결과 해석 및 검증****: AI 기반 연구 결과의 신뢰도를 높이기 위해 , 모델이 특정 결론에 도달한 핵심 요인을 XAI 로 설명하고 검증하는 데 활용할 수 있습니다 .

03

****윤리적 AI 개발 가이드라인****: AI 의 ' 블랙박스 ' 문제를 인지하고 XAI 를 적극적으로 도입함으로써 , DreamLabs 의 AI 개발 프로세스에 윤리적이고 책임감 있는 AI (Responsible AI) 원칙을 강화할 수 있습니다 .

04

****사용자 신뢰도 향상****: AI 기반 서비스나 제품을 개발할 때 , 사용자에게 AI 의 작동 원리나 결정 이유를 설명할 수 있는 기능을 통합하여 사용자 신뢰도와 수용도를 높일 수 있습니다 .

05

****AI 교육 및 역량 강화****: 내부 연구원 및 개발자를 대상으로 XAI 및 SHAP 기법에 대한 교육을 진행하여 AI 해석 역량을 강화하고 , 더 깊이 있는 연구를 수행할 수 있도록 지원합니다 .

Verification Required

모델 추론 /metadata 한계 / 원본 확인 필요

01

*** 클로드 아버지의 증격 고백 '우리도 모른다' ***의 정확한 맥락 **: 영상 제목에 언급된 이 발언이 엔트로픽 CEO 의 발언인지 , 어떤 상황에서 어떤 뉘앙스로 언급되었는지 영상 시청을 통한 정확한 확인이 필요합니다 .

02

**엔트로픽 CEO 의 경고 내용 구체화 **: 엔트로픽 CEO 가 AI 해석 불가능성에 대해 구체적으로 어떤 내용을 경고했는지 영상 시청을 통해 확인해야 합니다 .

03

** 챗 GPT 설명의 '합정' 구체적 사례 **: 챗 GPT 의 설명이 어떤 점에서 '합정' 인지 , 구체적인 예시와 함께 제시되는지 영상 시청을 통해 확인이 필요합니다 .

04

**SHAP 의 금융 AI 대출 심사 거절 설명 사례 상세 내용 **: 실제 사례의 구체적인 데이터와 설명 방식에 대한 상세한 내용 확인이 필요합니다 .

05

** 이수인 교수의 'AI 로 치매 정복' 포부의 구체적인 계획 **: 이 야심찬 목표에 대한 구체적인 연구 방향이나 계획이 영상에서 제시되는지 확인이 필요합니다 .

Source & Download Metadata

게시물과 문서 산출물 추적 정보

METADATA

Title: 클로드 아버지의 충격 고백 ” 우리도 모른다 “ 전세기 AI 천재들이 경악한 현대 AI 의 근본 문제 | 딥토크 EP 14
Channel: SEIZE 씨즈
Video ID: kV9H9rZvLK0
Source URL: <https://www.youtube.com/watch?v=kV9H9rZvLK0>
Playlist ID: PLHwM6idVO2zyqi2IZeDAiP5QBqRXd2Zyh
Generated at: 2026-06-27T15:31:49Z
Source basis: metadata_and_model_inference