AI 추천 목록에 추천 이유를 공개해야 할까?

똑똑 · 에디터 나루 · 발행 2026.09.30

추천 이유를 보여 주면 이용자가 왜 목록이 나왔는지 묻고 선택을 수정할 수 있습니다.

추천 이유를 보여 주면 이용자가 왜 목록이 나왔는지 묻고 선택을 수정할 수 있습니다. 그럴듯한 설명이 잘못된 추천을 정당화할 수 있습니다.

쉬운 설명

추천 이유를 보여 주면 이용자가 왜 목록이 나왔는지 묻고 선택을 수정할 수 있습니다. “비슷한 책을 읽어서” 같은 짧은 설명과 실제 추천 모형의 작동을 완전히 공개하는 것은 다릅니다.

핵심

- 추천 이유를 보여 주면 이용자가 왜 목록이 나왔는지 묻고 선택을 수정할 수 있습니다.
- “비슷한 책을 읽어서” 같은 짧은 설명과 실제 추천 모형의 작동을 완전히 공개하는 것은 다릅니다.
- 그럴듯한 설명이 잘못된 추천을 정당화할 수 있습니다.

핵심 질문 · 대상과 범위부터 읽기

추천 이유를 보여 주면 이용자가 왜 목록이 나왔는지 묻고 선택을 수정할 수 있습니다. 설명 문구의 정확성도 검증해야 합니다.

추천 설명에는 “이 책을 지난주에 클릭했다”처럼 실제 사용한 신호를 적을 수 있지만 시스템이 왜 점수를 높였는지 모두 보여 주기는 어렵습니다. 그래서 설명 문장이 실제 입력과 얼마나 일치하는지 검사해야 합니다. 듣기 좋은 문구는 검증 가능한 이유와 다릅니다.

개념 구분 · 같아 보이는 말을 나누기

“비슷한 책을 읽어서” 같은 짧은 설명과 실제 추천 모형의 작동을 완전히 공개하는 것은 다릅니다. 개인정보와 악용 가능성도 있습니다.

추천을 잘못 이해한 학생이 관심 없는 책을 계속 누르는 경우, 이유 공개와 함께 “이 주제 줄이기” 같은 수정 기능이 있어야 선택권이 생깁니다. 반대로 친구의 독서 기록을 근거로 삼았음을 구체적으로 밝히면 그 친구의 사생활을 드러낼 수 있습니다.

적용 사례 · 가상의 장면에서 판단하기

가상 독서 앱이 특정 주제만 계속 추천합니다. “지난주에 클릭한 역사 책 때문에”라는 설명이 있으면 학생이 관심 범위를 넓히는 선택을 할 수 있습니다.

토론은 내부 코드를 공개할지의 단일 선택으로 끝나지 않습니다. 추천에 쓰인 정보의 큰 범주, 이용자가 바꿀 수 있는 설정, 오류 신고 절차를 먼저 공개하고 위험이 큰 결정에는 더 자세한 설명을 요구하는 단계적 방식을 검토할 수 있습니다.

비교표 · 이 글의 세 가지 판단 기준

확인 항목	본문에서 확인한 것	읽을 때 주의할 점
핵심 범위	추천 이유를 보여 주면 이용자가 왜 목록이 나왔는지 묻고 선택을 수정할 수 있습니다.	“비슷한 책을 읽어서” 같은 짧은 설명과 실제 추천 모형의 작동을 완전히 공개하는 것은 다릅니다.
가상 사례	가상 독서 앱이 특정 주제만 계속 추천합니다.	그럴듯한 설명이 잘못된 추천을 정당화할 수 있습니다.
후속 확인	설명과 실제 추천 입력의 일치율, 이의 제기, 추천 끄기, 민감 정보 노출을 평가합니다.	설명만 바뀌고 목록은 그대로라면 사용자는 실제 선택 기준을 이해하기 어렵습니다.

본문 설명을 비교하기 위해 직접 만든 표입니다. 기관 원문의 도표나 실제 조사 결과는 아닙니다.

찬성 · 이유 공개를 기본으로 한다

주장

이유 공개를 기본으로 한다

설명

이용자의 선택권과 오류 발견 가능성을 높입니다.

근거

UNESCO AI 윤리 권고는 맥락에 맞는 투명성과 설명 가능성을 강조합니다.

근거의 한계

그럴듯한 설명이 잘못된 추천을 정당화할 수 있습니다. 반대로 모든 내부 수치를 공개하면 이용자가 읽고 판단하기 어려울 수 있습니다.

예시

특정 주제만 추천될 때 입력 이유를 보고 설정을 바꾸는 경우입니다.

반론

설명이 과장되거나 개인 정보를 노출할 수 있습니다.

성립 조건

설명 정확도·개인정보 보호를 함께 검증할 때입니다.

신중·반대 · 설명의 범위를 제한한다

주장

설명의 범위를 제한한다

설명

복잡한 내부 작동을 한 줄로 오해하게 만들거나 민감 정보를 드러낼 위험을 줄입니다.

근거

UNESCO 권고도 투명성과 개인정보·안전 사이의 긴장을 인정합니다.

근거의 한계

그럴듯한 설명이 잘못된 추천을 정당화할 수 있습니다. 반대로 모든 내부 수치를 공개하면 이용자가 읽고 판단하기 어려울 수 있습니다.

예시

친구의 행동을 특정할 수 있는 이유가 노출되는 경우입니다.

반론

불투명해져 오류를 바로잡기 어려울 수 있습니다.

성립 조건

추천 기준의 큰 범주와 이의 제기 통로는 제공할 때입니다.

한 단계 더 · 조건을 바꾸면 답도 달라집니다

추천 이유의 정확성을 시험할 때는 같은 입력을 바꾸었을 때 설명도 그에 맞게 바뀌는지 봅니다. 설명만 바뀌고 목록은 그대로라면 사용자는 실제 선택 기준을 이해하기 어렵습니다.

그럴듯한 설명이 잘못된 추천을 정당화할 수 있습니다. 반대로 모든 내부 수치를 공개하면 이용자가 읽고 판단하기 어려울 수 있습니다.

직접 확인 · 원문과 사례에 적용하기

설명과 실제 추천 입력의 일치율, 이의 제기, 추천 끄기, 민감 정보 노출을 평가합니다.

어떤 친구 집단인지, 내 행동 정보가 쓰였는지, 목록에서 제외된 자료는 무엇인지 묻습니다. 개인을 드러내지 않는 범위의 검증 가능한 설명이 필요합니다.

직접 생각해 보기

추천 이유에 “친구들이 좋아해서”만 적었다면 추가할 질문은?

확인 설명

어떤 친구 집단인지, 내 행동 정보가 쓰였는지, 목록에서 제외된 자료는 무엇인지 묻습니다. 개인을 드러내지 않는 범위의 검증 가능한 설명이 필요합니다.

AI 추천 목록에 추천 이유를 공개해야 할까?

추천 이유를 보여 주면 이용자가 왜 목록이 나왔는지 묻고 선택을 수정할 수 있습니다. 설명 문구의 정확성도 검증해야 합니다. “비슷한 책을 읽어서” 같은 짧은 설명과 실제 추천 모형의 작동을 완전히 공개하는 것은 다릅니다. 개인정보와 악용 가능성도 있습니다.

추천 이유에 “친구들이 좋아해서”만 적었다면 추가할 질문은?

어떤 친구 집단인지, 내 행동 정보가 쓰였는지, 목록에서 제외된 자료는 무엇인지 묻습니다. 개인을 드러내지 않는 범위의 검증 가능한 설명이 필요합니다.

추천 이유 한 줄이면 알고리즘을 충분히 이해할 수 있나요?

그럴듯한 설명이 잘못된 추천을 정당화할 수 있습니다. 반대로 모든 내부 수치를 공개하면 이용자가 읽고 판단하기 어려울 수 있습니다. 설명과 실제 추천 입력의 일치율, 이의 제기, 추천 끄기, 민감 정보 노출을 평가합니다.

원문과 출처

- UNESCO · AI 윤리 권고 — https://www.unesco.org/en/artificial-intelligence/recommendation-ethics: 설명 가능성, 투명성, 인간의 책임; 본문의 가상 수치나 학교 사례를 제공한 자료는 아닙니다.
- UNESCO · 2023 교육 기술 보고서 — https://gem-report-2023.unesco.org/technology-in-education/: 교육 기술의 접근성·근거·부작용을 함께 보는 관점; 본문의 가상 수치나 학교 사례를 제공한 자료는 아닙니다.

제작 방식과 정정 안내: https://www.dokdok.co/about#editorial

온라인 원문 · 수정된 내용은 아래 페이지에서 확인하세요.

https://www.dokdok.co/reading/explain-ai-recommendations