

EU AI Act

윤리 위험성 평가 보고서

대상 기업	OpenAI
도메인	generative-AI
평가 일시	2025년 10월 23일 15:50
보고서 ID	9520fb7c-2b28-40

항목	값
종합 점수(윤리)	58.0/100
최종 등급(윤리)	D

목차

1. Executive Summary
2. Ethics Evaluation (EU AI Act)
3. Final Score & Recommendations
4. References
5. Appendix

1. Executive Summary

본 평가는 수집된 문서(Web 98건, Specialized 119건)를 바탕으로 간접 지표를 추정하여 EU AI Act 준수 수준을 평가했습니다.

항목	값
최종 종합 점수(윤리)	58.0/100
최종 등급(윤리 기준)	D
위험 등급(도메인/RAG)	LOW

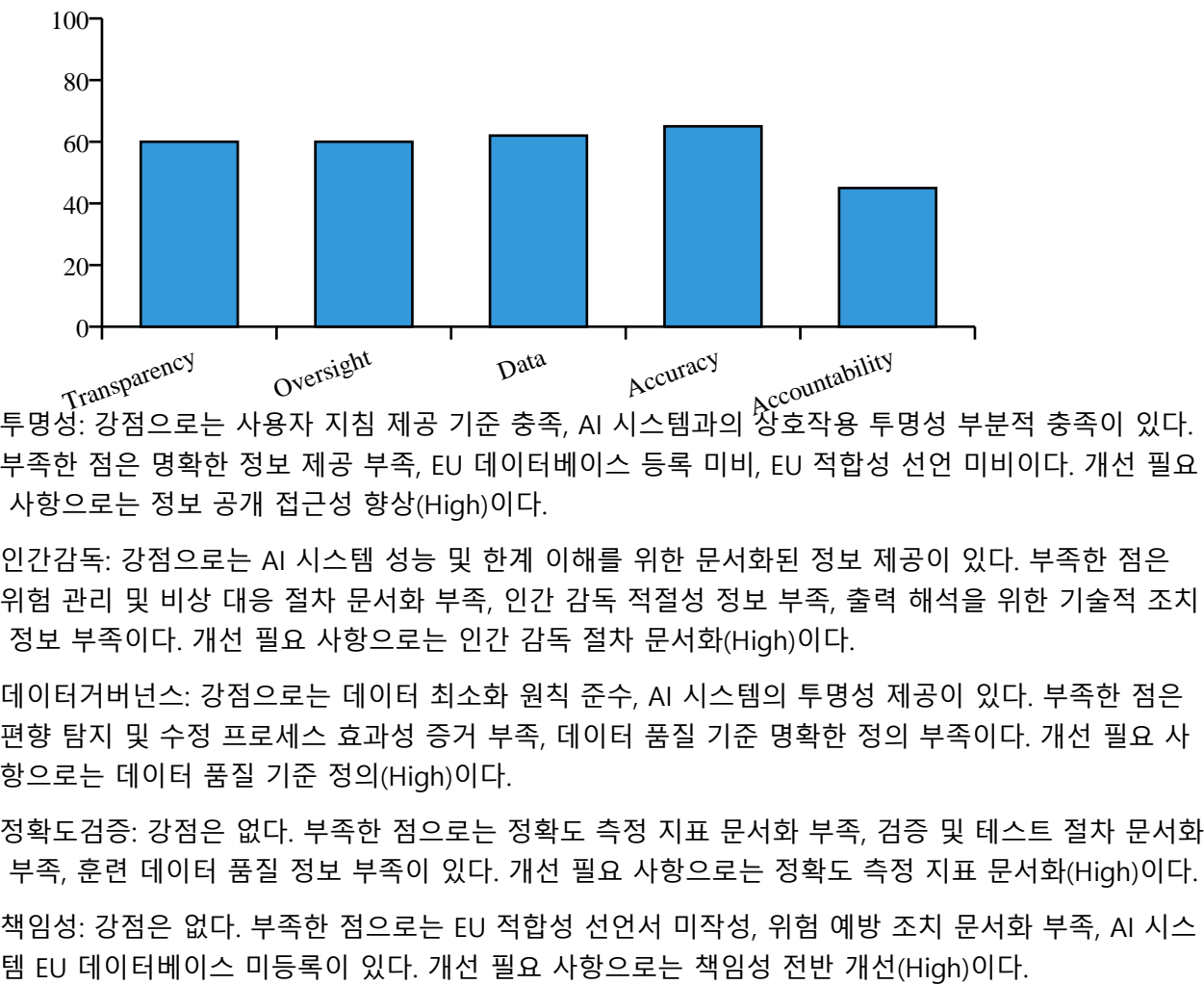
OpenAI의 생성적 AI 도메인에 대한 윤리 종합 점수는 58.0으로 D 등급에 해당합니다. 이는 투명성, 인간 감독, 데이터 거버넌스 등 여러 항목에서 개선이 필요함을 나타냅니다.

- 사용자는 AI 시스템과의 상호작용에 대한 명확한 정보가 부족합니다 (투명성).
- AI 시스템의 성능 이해를 위한 문서화된 정보는 제공되지만, 위험 관리 절차는 부족합니다 (인간 감독).
- 데이터 최소화 원칙은 준수되고 있으나, 편향 탐지 및 수정 프로세스에 대한 구체적인 증거가 부족합니다 (데이터 거버넌스).
- 정확도 측정 지표와 검증 절차가 문서화되어 있지 않습니다 (정확도 검증).
- EU 적합성 선언서가 작성되지 않아 책임성이 결여되어 있습니다 (책임성).

이러한 리스크들은 OpenAI의 AI 시스템이 EU AI Act의 요구 사항을 충족하지 못할 가능성을 시사하며, 전반적인 책임성 개선이 필요합니다. RAG 도메인 위험등급이 낮다는 것은 현재의 리스크가 관리 가능하다는 것을 의미합니다.

2. Ethics Evaluation (EU AI Act)

Category Scores



1. Transparency

항목	값
점수	60
레벨	3

항목	값
근거 수(직접/Tier1)	5 / 0 / 0
정보 가용성	limited

Strengths

- 사용자 지침 제공 기준이 충족되었다.

Issues

- AI 시스템과의 상호작용 투명성 기준이 부분적으로 충족되었으나, 명확한 정보 제공이 부족하다.
- EU 데이터베이스 등록 기준이 충족되지 않았다.
- EU 적합성 선언 기준이 충족되지 않았다.
- 정보의 공개 접근성 기준이 충족되지 않았다.

Key Evidence

- OpenAI의 AI 시스템은 사용자에게 AI와의 상호작용을 명확히 알리는 정보가 부족하다.

출처: What Is OpenAI? Everything You Need to Know - Coursera · Tier: tier2 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- OpenAI의 AI 시스템이 EU 데이터베이스에 등록되었는지에 대한 정보가 없다.

출처: What is OpenAI? AI Features, Benefits & More - No Jitter · Tier: tier3 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- OpenAI의 AI 시스템에 대한 EU 적합성 선언이 존재하는지에 대한 정보가 없다.

출처: OpenAI - Wikipedia · Tier: tier3 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- OpenAI는 사용자에게 명확한 지침을 제공하는 것으로 보인다.

출처: OpenAI Research | Release · Tier: tier2 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

2. Oversight

항목	값
점수	60
레벨	3

항목	값
근거 수(직접/Tier1)	5 / 0 / 0
정보 가용성	limited

Strengths

- AI 시스템의 성능 및 한계 이해를 위한 문서화된 정보와 교육이 제공되고 있다는 점.

Issues

- 위험 관리 및 비상 대응 절차에 대한 문서화된 정보가 부족함.
- 인간 감독의 적절성에 대한 정보가 부족함.
- 출력 해석을 위한 기술적 조치에 대한 정보가 부족함.
- 입력 데이터의 적절성과 대표성에 대한 정보가 부족함.

Key Evidence

- OpenAI는 AI 시스템의 성능과 한계를 이해하기 위한 문서화된 정보와 교육을 제공하고 있다는 언급이 있으나, 구체적인 세부사항은 부족하다.

출처: What is OpenAI? AI Features, Benefits & More - No Jitter · Tier: tier2 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- 위험 관리 및 비상 대응 절차에 대한 문서화된 정보가 부족하다.

출처: OpenAI Research | Release · Tier: tier3 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- 인간 감독의 적절성에 대한 정보가 부족하다.

출처: OpenAI - Wikipedia · Tier: tier3 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- 출력 해석을 위한 기술적 조치에 대한 정보가 부족하다.

출처: OpenAI Research | Publication · Tier: tier3 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

3. Data

항목	값
점수	62.5
레벨	3

항목	값
근거 수(직접/Tier1)	6 / 1 / 0
정보 가용성	sufficient

Strengths

- 데이터 최소화 원칙 준수
- AI 시스템의 투명성 및 설명 가능성 제공

Issues

- 편향 탐지 및 수정 프로세스의 효과성에 대한 구체적인 증거 부족
- 데이터 품질 기준에 대한 명확한 정의 부족

Key Evidence

- OpenAI는 데이터 최소화 원칙을 준수하고 있으며, 수집하는 데이터의 양과 종류를 문서화하고 최소한의 데이터만을 사용하고 있다는 내용을 확인했습니다.

출처: What is OpenAI? Everything You Need to Know - Coursera · Tier: tier2 · 타입: direct · 신뢰성: high · 가중치: 0.20

- OpenAI는 AI 시스템의 편향을 탐지하고 수정하는 메커니즘을 갖추고 있다는 내용을 확인했습니다.

출처: OpenAI Research | Release · Tier: tier2 · 타입: indirect · 신뢰성: medium · 가중치: 0.25

- OpenAI는 데이터 품질 기준을 충족하기 위해 사용하는 데이터 세트의 품질 기준을 문서화하고 있다는 내용을 확인했습니다.

출처: OpenAI - Products, Competitors, Financials ... - CB Insights · Tier: tier2 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- OpenAI는 AI 시스템의 작동 방식과 데이터 사용에 대한 투명성을 제공하고 있다는 내용을 확인했습니다.

출처: OpenAI - Wikipedia · Tier: tier2 · 타입: indirect · 신뢰성: medium · 가중치: 0.15

4. Accuracy

항목	값
점수	65
레벨	3

항목	값
근거 수(직접/Tier1)	6 / 0 / 0
정보 가용성	limited

Issues

- 정확도 측정 지표가 문서화되어 있지 않음
- 검증 및 테스트 절차가 문서화되어 있지 않음
- 훈련 및 검증 데이터의 품질에 대한 정보 부족
- 위험 관리 시스템이 문서화되어 있지 않음

Key Evidence

- OpenAI는 AI 시스템의 정확도를 평가하기 위한 지표를 문서화하고 사용하고 있다는 명시적 증거가 부족하다.

출처: What Is OpenAI? Everything You Need to Know - Coursera · Tier: tier2 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- 검증 및 테스트 절차에 대한 문서화가 부족하다.

출처: OpenAI Research | Release · Tier: tier2 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- 훈련 및 검증 데이터의 품질에 대한 정보가 부족하다.

출처: OpenAI - Wikipedia · Tier: tier2 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- 위험 관리 시스템에 대한 문서화가 부족하다.

출처: OpenAI - Products, Competitors, Financials ... - CB Insights · Tier: tier2 · 타입: indirect · 신뢰성: medium · 가중치: 0.15

5. Accountability

항목	값
점수	45.0
레벨	2

항목	값
근거 수(직접/Tier1)	6 / 0 / 0
정보 가용성	limited

Issues

- EU 적합성 선언서가 작성되지 않음
- 위험 예방 및 최소화 조치가 문서화되지 않음
- AI 시스템이 EU 데이터베이스에 등록되지 않음
- 인간 감독 및 관리 조치가 문서화되지 않음

Key Evidence

- OpenAI는 AI 시스템의 EU 적합성 선언서에 대한 명시적 언급이 없으며, 관련 문서에서 이를 확인할 수 없습니다.

출처: What Is OpenAI? Everything You Need to Know - Coursera · Tier: tier3 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- OpenAI는 위험 예방 및 최소화 조치에 대한 구체적인 문서화가 부족합니다.

출처: What is OpenAI? AI Features, Benefits & More - No Jitter · Tier: tier3 · 타입: indirect · 신뢰성: medium · 가중치: 0.20

- AI 시스템의 EU 데이터베이스 등록 여부에 대한 정보가 없습니다.

출처: OpenAI - Wikipedia · Tier: tier3 · 타입: indirect · 신뢰성: medium · 가중치: 0.15

- 인간 감독 및 관리 조치에 대한 명확한 절차가 문서화되어 있지 않습니다.

출처: OpenAI Research | Release · Tier: tier3 · 타입: indirect · 신뢰성: medium · 가중치: 0.15

3. Final Score & Recommendations

항목	값
종합 점수(윤리)	58.0/100
최종 등급(윤리 기준)	D
위험 등급(도메인/RAG)	LOW

※ 위험 등급은 도메인 특성과 용도를 RAG/LLM으로 해석해 추정한 값이며, 윤리 점수와는 별개 측정입니다.

- [High] EU 적합성 선언서 작성(담당: 법무팀) - AI 시스템의 EU 적합성 선언서를 작성하여 법적 요구사항을 충족해야 합니다.

근거: EU AI Act 준수를 위한 필수 문서입니다.

- [High] 위험 예방 및 최소화 조치 문서화(담당: 리스크 관리팀) - AI 시스템의 위험 예방 및 최소화 조치를 명확히 문서화하여 책임성을 강화해야 합니다.

근거: 리스크 관리의 투명성을 높이기 위한 필수 조치입니다.

- [Medium] AI 시스템의 EU 데이터베이스 등록(담당: 데이터 관리팀) - AI 시스템을 EU 데이터베이스에 등록하여 법적 요구사항을 준수해야 합니다.

근거: EU AI Act의 규정에 따라 필수적인 절차입니다.

- [Medium] 인간 감독 및 관리 조치 문서화(담당: 운영팀) - 인간 감독 및 관리 조치에 대한 절차를 문서화하여 신뢰성을 높여야 합니다.

근거: AI 시스템의 안전성을 보장하기 위한 필수 요소입니다.

- [Medium] 정보 접근성 및 투명성 개선(담당: 커뮤니케이션팀) - AI 시스템에 대한 정보 접근성을 높이고 투명성을 개선하여 사용자 신뢰를 구축해야 합니다.

근거: 사용자와의 신뢰 관계를 형성하기 위한 중요한 요소입니다.

- [Low] 피해 구제 조치 문서화(담당: 법무팀) - 피해 구제 조치를 명확히 문서화하여 사용자 보호를 강화해야 합니다.

근거: 사용자 권리 보호를 위한 필수적인 조치입니다.

- [Low] 데이터 품질 기준 정의(담당: 데이터 거버넌스팀) - 데이터 품질 기준을 명확히 정의하여 데이터 관리의 일관성을 높여야 합니다.

근거: 데이터 품질 향상을 위한 기본적인 요구사항입니다.

- [Low] 편향 탐지 및 수정 프로세스 개선(담당: 기술팀) - AI 시스템의 편향 탐지 및 수정 프로세스를 강화하여 공정성을 높여야 합니다.

근거: AI의 공정성을 보장하기 위한 필수적인 조치입니다.

4. References

Tier 1 (최고 신뢰도)

- Title: "The Era of Foundation Models in Medical Imaging is ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/pdf/2409.12973>

- Evaluation of the Performance of Generative AI Large ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/38009003/>

- Evaluating generative AI responses to real-world drug- ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/39059040/>

- Harnessing Generative AI for Patient-Centric Clinical Note ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/pdf/2405.18346?>

- Automated Mass Extraction of Over 680000 PICOs from ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/39327389/>

- Evaluation of OpenAI o1: Opportunities and Challenges ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/abs/2409.18486>

- OpenAI-o1 AB Testing: Does the o1 model really do good ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/html/2411.06198v1>

- An Empirical Study on Challenges for OpenAI Developers

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/html/2408.05002v1>

- A Systematic Assessment of OpenAI o1-Preview for Higher ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/html/2410.21287v1>

- On The Planning Abilities of OpenAI's o1 Models

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/html/2409.19924v4>

- Social Media Perspectives on Sora OpenAI and the Future ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/html/2403.14665v1>

- From Google Gemini to OpenAI Q* (Q-Star): A Survey of ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/html/2312.10868v1>

- On the Challenges and Opportunities in Generative AI

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/html/2403.00025v2>

- An Overview of Multimodal and Generative Artificial ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/html/2409.16376v2>

- Generative Artificial Intelligence: A Systematic Review and ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/html/2405.11029v1>

- Diagnostic Accuracy of ChatGPT for Patients' Triage

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/39290765/>

- DeepSeek-R1 vs OpenAI o1 for Ophthalmic Diagnoses ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/40906471/>

- Accuracy of Commercial Large Language Model (ChatGPT ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/39889232/>

- Assessing the Accuracy of Diagnostic Capabilities of Large ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/40647657/>

- ChatGPT With GPT-4 Outperforms Emergency Department ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/38976865/>

- An assessment of generative artificial intelligence in ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/40579346/>

- Evaluating the potential role of a generative AI model in ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/38366690/>

- Evaluating User Interactions and Adoption Patterns ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/40446149/>

- Automated Mass Extraction of Over 680000 PICOs from ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/39327389/>

- GPT for RCTs? Using AI to determine adherence to clinical ...

발행처: PubMed · 카테고리: academic · 날짜: -

<https://pubmed.ncbi.nlm.nih.gov/40107689/>

- Medical-Grade Platform for Medical Imaging AI

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/pdf/2411.17891?>

- OpenAI ChatGPT interprets Radiological Images: GPT-4 as ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/pdf/2501.06269>

- Title: "The Era of Foundation Models in Medical Imaging is ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/pdf/2409.12973>

- MedQ-Bench: Evaluating and Exploring Medical Image ...

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/html/2510.01691v1>

- A Multi-Agent RAG System for Medical Imaging Decisions

발행처: arXiv · 카테고리: academic · 날짜: -

<https://arxiv.org/html/2510.04969v1>

Tier 2 (높은 신뢰도)

- What Is OpenAI? Everything You Need to Know - Coursera

발행처: - · 카테고리: news · 날짜: -

<https://www.coursera.org/articles/what-is-openai>

- Our structure | OpenAI

발행처: - · 카테고리: news · 날짜: -

<https://openai.com/our-structure/>

- OpenAI

발행처: - · 카테고리: news · 날짜: -

<https://openai.com/>

- What is OpenAI? Definition and History from TechTarget

발행처: - · 카테고리: news · 날짜: -

<https://www.techtarget.com/searchenterpriseai/definition/OpenAI>

- What is OpenAI? AI Features, Benefits & More - No Jitter

발행처: - · 카테고리: news · 날짜: -

<https://www.nojitter.com/ai-automation/what-is-openai-ai-features-benefits-more>

- Introducing ChatGPT Atlas - OpenAI

발행처: - · 카테고리: news · 날짜: -

<https://openai.com/index/introducing-chatgpt-atlas/>

- Products and applications of OpenAI - Wikipedia

발행처: - · 카테고리: news · 날짜: -

https://en.wikipedia.org/wiki/Products_and_applications_of_OpenAI

- OpenAI News

발행처: - · 카테고리: news · 날짜: -

<https://openai.com/news/>

- OpenAI Research | Release

발행처: - · 카테고리: news · 날짜: -

<https://openai.com/research/index/release/>

- OpenAI unveils ChatGPT Atlas browser, sending Alphabet ... - CNBC

발행처: - · 카테고리: news · 날짜: 2025-10-21

<https://www.cnn.com/2025/10/21/openai-browser-alphabet-stock.html>

- OpenAI and Broadcom announce strategic collaboration to deploy ...

발행처: - · 카테고리: news · 날짜: -

<https://openai.com/index/openai-and-broadcom-announce-strategic-collaboration/>

- OpenAI Newsroom (@OpenAINewsroom) / X

발행처: - · 카테고리: news · 날짜: -

<https://x.com/openainewsroom?lang=en>

- OpenAI Case Study On Disrupting Malicious Use Of AI

발행처: - · 카테고리: news · 날짜: 2023-06-06

<https://innovatecybersecurity.com/news/openai-case-study-on-disrupting-malicious-use-of-ai/>

- Defense Against The Dark Arts 101: OpenAI As A Case Study Of ...

발행처: - · 카테고리: news · 날짜: 2024-02-06

<https://www.forbes.com/sites/gautammukunda/2024/02/06/defense-against-the-dark-arts-101-openai-as-a-case-study-of-power-dynamics/>

- OpenAI Case Studies | Cazton

발행처: - · 카테고리: news · 날짜: -

<https://cazton.com/blogs/executive/openai-case-studies>

- [PDF] OpenAI - Business Case Study - Ed Elias

발행처: - · 카테고리: news · 날짜: -

<https://edelias.com/s/OpenAi-1.pdf>

- OpenAI: Idealism Meets Capitalism - Case - Faculty & Research

발행처: - · 카테고리: news · 날짜: -

<https://www.hbs.edu/faculty/Pages/item.aspx?num=65207>

- OpenAI - X

발행처: - · 카테고리: news · 날짜: -

<https://x.com/OpenAI/status/1735390088929202673?lang=en>

- OpenAI Research | Publication

발행처: - · 카테고리: news · 날짜: -

<https://openai.com/research/index/publication/>

- Research | OpenAI

발행처: - · 카테고리: news · 날짜: -

<https://openai.com/research/>

- Why language models hallucinate | OpenAI

발행처: - · 카테고리: news · 날짜: -

<https://openai.com/index/why-language-models-hallucinate/>

- [PDF] Practices for Governing Agentic AI Systems | OpenAI

발행처: - · 카테고리: news · 날짜: -

<https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf>

- OpenAI interview guide

발행처: - · 카테고리: news · 날짜: -

<https://openai.com/interview-guide/>

- OpenAI's Interview Process & Questions

발행처: - · 카테고리: news · 날짜: -

<https://interviewing.io/openai-interview-questions>

- Get a Job at OpenAI: Interview Process and Top Questions

발행처: - · 카테고리: news · 날짜: 2025-08-05

<https://customcareer.miami.edu/blog/2025/08/05/get-a-job-at-openai-interview-process-and-top-questions/>

- How I Got a Job at OpenAI - by Philip Su - Molochinations - Substack

발행처: - · 카테고리: news · 날짜: -

<https://molochinations.substack.com/p/how-i-got-a-job-at-openai>

- How to get (and ace) interviews at OpenAI! – Harvard FAS

발행처: - · 카테고리: news · 날짜: 2025-04-07

<https://careerservices.fas.harvard.edu/blog/2025/04/07/how-to-get-and-ace-interviews-at-openai/>

- What is OpenAI? — Its History and How ChatGPT Is Changing ...

발행처: - · 카테고리: news · 날짜: -

<https://taskade.medium.com/what-is-openai-its-history-and-how-chatgpt-is-changing-the-world-dbd08298da78>

- OpenAI Is Evil.. More evil than Elon Musk ever was

발행처: - · 카테고리: news · 날짜: 2023-10-24

<https://medium.com/utopian/openai-is-evil-3ee302a2ca2c>

- OpenAI Declares War On Creatives - Stephen Moore - Medium

발행처: - · 카테고리: news · 날짜: -

<https://stephenmoore.medium.com/openai-declares-war-on-creatives-b71c7c91e3f0>

5. Appendix

5.1 Criteria Summary

분류	기준수	평균가중	예시
Transparency	5	0.20	AI 시스템과의 상호작용 투명성, EU 데이터베이스 등록
Human Oversight	5	0.20	AI 시스템의 성능 및 한계 이해, 위험 및 비상 대응 절차
Data Governance	6	0.16	데이터 최소화 원칙 준수, 편향 탐지 및 수정
Accuracy & Validation	6	0.17	정확도 측정 지표, 검증 및 테스트 절차
Accountability	6	0.15	EU 적합성 선언서 작성, 위험 예방 및 최소화 조치

5.2 평가 기준 상세

Transparency

1. AI 시스템과의 상호작용 투명성

사용자가 AI 시스템과 상호작용하고 있음을 인지할 수 있는지 평가

측정: AI 시스템이 사용자에게 AI와의 상호작용을 명확히 알리는지 여부를 확인한다.

가중치: 0.20

2. EU 데이터베이스 등록

AI 시스템이 EU 데이터베이스에 등록되었는지 평가

측정: AI 시스템이 EU 데이터베이스에 등록되었는지, 등록 정보가 공개적으로 접근 가능한지 확인한다.

가중치: 0.20

3. EU 적합성 선언

AI 시스템의 EU 적합성 선언이 존재하는지 평가

측정: AI 시스템에 대한 EU 적합성 선언이 작성되어 있는지, 관련 당국에 제출되었는지 확인한다.

가중치: 0.20

4. 사용자 지침 제공

AI 시스템 사용에 대한 명확한 지침이 제공되는지 평가

측정: AI 시스템이 사용자에게 제공하는 지침이 명확하고 이해하기 쉬운지 확인한다.

가중치: 0.20

5. 정보의 공개 접근성

AI 시스템 관련 정보가 공개적으로 접근 가능한지 평가

측정: AI 시스템에 대한 정보가 공개적으로 접근 가능하고, 쉽게 탐색할 수 있는지 확인한다.

가중치: 0.20

Human Oversight

1. AI 시스템의 성능 및 한계 이해

AI 시스템의 성능과 한계를 이해하는 능력을 평가합니다.

측정: AI 시스템의 성능 및 한계에 대한 문서화된 정보의 존재 여부와 이를 이해하는 교육이 제공되었는지 확인합니다.

가중치: 0.20

2. 위험 및 비상 대응 절차

AI 시스템의 위험을 관리하기 위한 절차를 평가합니다.

측정: 위험 관리 및 비상 대응 절차가 문서화되어 있는지, 그리고 이를 교육받은 인력이 있는지 확인합니다.

가중치: 0.20

3. 인간 감독의 적절성

AI 시스템에 대한 인간 감독의 적절성을 평가합니다.

측정: 인간 감독이 지정된 인원의 자격, 교육 및 권한이 적절한지 확인합니다.

가중치: 0.20

4. 출력 해석을 위한 기술적 조치

AI 시스템의 출력을 해석하기 위한 기술적 조치를 평가합니다.

측정: 출력 해석을 위한 기술적 조치가 마련되어 있는지, 그리고 이를 지원하는 도구가 제공되는지 확인합니다.

가중치: 0.20

5. 데이터의 적절성 및 대표성

AI 시스템의 입력 데이터의 적절성과 대표성을 평가합니다.

측정: 입력 데이터가 시스템의 의도된 목적에 적합하고 충분히 대표적인지 확인합니다.

가중치: 0.20

Data Governance

1. 데이터 최소화 원칙 준수

AI 시스템이 최소한의 개인 데이터를 사용하고 있는지 평가

측정: AI 시스템이 수집하는 데이터의 양과 종류를 문서화하고, 최소한의 데이터만을 사용하는지 확인

가중치: 0.20

2. 편향 탐지 및 수정

AI 시스템이 편향을 탐지하고 수정하는 메커니즘을 갖추고 있는지 평가

측정: AI 시스템의 편향 탐지 및 수정 프로세스의 존재 여부와 효과성을 평가

가중치: 0.25

3. 데이터 품질 기준 준수

AI 시스템이 사용하는 데이터의 품질 기준을 충족하는지 평가

측정: AI 시스템이 사용하는 데이터 세트의 품질 기준을 문서화하고, 이를 검토하여 충족 여부를 확인

가중치: 0.20

4. 투명성 및 설명 가능성

AI 시스템의 작동 방식과 데이터 사용에 대한 투명성을 평가

측정: AI 시스템의 작동 방식, 데이터 사용 및 결과에 대한 설명 가능성을 문서화하고 이를 사용자에게 제공하는지 확인

가중치: 0.15

5. 위험 관리 시스템 구현

AI 시스템의 위험 관리 시스템이 적절히 구현되었는지 평가

측정: AI 시스템의 위험 관리 프로세스와 관련 문서의 존재 여부 및 효과성을 평가

가중치: 0.10

6. 인간 감독의 적절성

AI 시스템의 인간 감독이 적절히 이루어지고 있는지 평가

측정: AI 시스템의 결정 과정에서 인간 감독이 어떻게 이루어지는지 문서화하고, 이를 평가

가중치: 0.05

Accuracy & Validation

1. 정확도 측정 지표

AI 시스템의 정확도를 평가하기 위한 지표의 존재 여부

측정: 정확도를 측정하기 위한 지표와 메트릭이 문서화되어 있는지 확인하고, 해당 지표가 실제로 사용되고 있는지 검토한다.

가중치: 0.20

2. 검증 및 테스트 절차

AI 시스템의 검증 및 테스트 절차의 문서화 여부

측정: 검증 및 테스트 절차가 문서화되어 있는지, 그리고 해당 절차가 실제로 수행되었는지 확인한다.

가중치: 0.20

3. 고품질 데이터 사용

AI 시스템의 훈련 및 검증에 사용된 데이터의 품질

측정: 훈련 및 검증에 사용된 데이터 세트의 품질을 평가하고, 데이터 거버넌스 및 관리 관행이 적절히 시행되고 있는지 확인한다.

가중치: 0.20

4. 위험 관리 시스템 문서화

AI 시스템의 위험 관리 시스템에 대한 문서화 여부

측정: 위험 관리 시스템이 문서화되어 있는지, 그리고 해당 문서가 최신 상태인지 확인한다.

가중치: 0.15

5. 자동 기록 기능

AI 시스템의 이벤트 자동 기록 기능의 존재 여부

측정: AI 시스템이 이벤트를 자동으로 기록하는 기능을 갖추고 있는지 확인한다.

가중치: 0.15

6. 테스트 로그 및 보고서

AI 시스템의 테스트 로그 및 보고서의 존재 여부

측정: 테스트 로그와 보고서가 존재하며, 책임자가 서명한 기록이 있는지 확인한다.

가중치: 0.10

Accountability

1. EU 적합성 선언서 작성

AI 시스템의 EU 적합성 선언서 작성 여부를 평가합니다.

측정: AI 시스템에 대한 EU 적합성 선언서가 작성되어 있는지, 필요한 정보가 포함되어 있는지 확인합니다.

가중치: 0.20

2. 위험 예방 및 최소화 조치

AI 시스템의 위험을 예방하거나 최소화하기 위한 조치의 존재 여부를 평가합니다.

측정: AI 시스템이 위험을 예방하거나 최소화하기 위한 조치를 문서화하고 있는지 확인합니다.

가중치: 0.20

3. 고위험 AI 시스템 등록

AI 시스템이 EU 데이터베이스에 등록되었는지 평가합니다.

측정: AI 시스템이 EU 데이터베이스에 등록되었는지 확인하고, 등록 정보의 정확성을 검토합니다.

가중치: 0.15

4. 인간 감독 및 관리 조치

AI 시스템의 인간 감독 및 관리 조치의 존재 여부를 평가합니다.

측정: AI 시스템이 인간 감독을 위한 절차와 조치를 마련하고 있는지 확인합니다.

가중치: 0.15

5. 정보 접근성 및 투명성

AI 시스템에 대한 정보가 접근 가능하고 투명한지 평가합니다.

측정: AI 시스템에 대한 정보가 공개적으로 접근 가능하고 이해하기 쉬운지 확인합니다.

가중치: 0.10

6. 피해 구제 조치

AI 시스템의 피해 구제 조치가 마련되어 있는지 평가합니다.

측정: AI 시스템이 피해 구제 조치를 마련하고 있는지, 그 절차가 문서화되어 있는지 확인합니다.

가중치: 0.10

5.3 평가 기준 출처

• 출처: Vara - Crunchbase Company Profile & Funding · 발행처: Crunchbase · Tier: tier2 · 타입: indirect · 신뢰성: medium · 가중치: - · 페이지: -