

## 갈릴레오의 LLM 환각 지수 평가에서 GPT-4가 가장 우수












### KEY Contents

- 주요 LLM의 환각 현상을 평가한 'LLM 환각 지수'에 따르면 GPT-4는 작업 유형과 관계없이 가장 우수한 성능을 보였으며 GPT-3.5도 거의 동등한 성능을 발휘
- 오픈소스 모델 중에서는 메타의 라마2가 RAG 없는 질문과 답변 및 긴 형식의 텍스트 생성에서 가장 우수한 성능을 발휘

### ● 주요 LLM 중 GPT-4가 가장 환각 현상 적고 GPT-3.5 터보도 비슷한 성능 기록

- 머신러닝 데이터 관리 기업 갈릴레오(Galileo)가 2023년 11월 15일 주요 LLM의 환각 현상을 평가한 'LLM 환각 지수(LLM Hallucination Index)'를 발표
  - 생성 AI의 환각 현상은 AI 시스템이 잘못된 정보를 생성하거나, 현실과 다른 부정확한 결과를 내놓는 현상으로, 기업의 AI 도입을 가로막는 주요 장애물이며, 환각 지수는 신뢰할 수 있는 생성 AI 구축을 위해 환각을 평가하고 측정하는 구조화된 접근방식을 제공
  - 환각 지수는 △검색 증강 생성(Retrieval-Augmented Generation, RAG)\*을 포함한 질문과 답변 △RAG 없는 질문과 답변 △긴 형식의 텍스트(보고서나 기사, 에세이) 생성의 3개 작업 유형에 대하여 환각을 기준으로 LLM의 순위를 평가
    - \* 기존에 학습된 데이터가 아닌 외부 소스(데이터셋, 데이터베이스, 문서 등)에서 가져온 정보를 검색해 활용하는 기술
- 3개의 작업 유형 평가 전체에서 오픈AI의 GPT-4가 최고의 성능을 기록했으며, GPT-3.5 터보도 GPT-4와 거의 동등한 성능을 발휘
  - 메타의 라마2(Llama-2-70b)는 RAG 없는 질문과 답변 유형에서 오픈소스 모델 가운데 가장 우수했고 긴 형식의 텍스트 생성에서도 GPT-4에 준하는 성능을 기록했으나, RAG 포함 질문과 답변에서는 허깅 페이스의 제퍼(Zephyr-7b)가 라마2를 능가

#### 〈갈릴레오의 LLM 환각 지수(RAG 포함 질문과 답변 기준)〉

Developer	Model	Context Adherence Score
	gpt-4-0613	0.76
	gpt-3.5-turbo-0613	0.75
	gpt-3.5-turbo-1106	0.74
	zephyr-7b-beta	0.71
	gpt-3.5-turbo-instruct	0.68
	llama-2-70b-chat	0.68
	llama-2-13b-chat	0.68
	mistral-7b-instruct-v0.1	0.67
	llama-2-7b-chat	0.65
	falcon-40b-instruct	0.60
	mpt-7b-instruct	0.58

출처: Galileo, LLM Hallucination Index, 2023.11.15.

## 영국 옥스퍼드 인터넷 연구소, AI 기술자의 임금이 평균 21% 높아

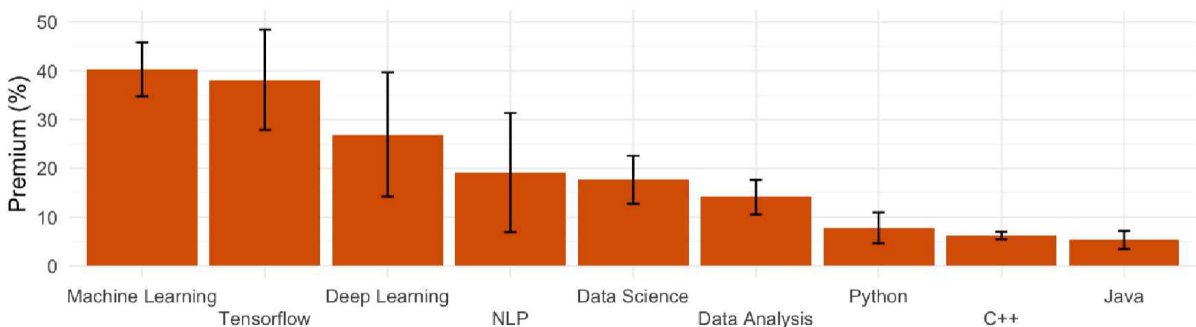
### KEY Contents

- 옥스퍼드 인터넷 연구소의 연구에 따르면 특정 기술의 경제적 가치는 다른 기술과 결합 가능성이 높을수록 높게 평가됨
- AI의 확산은 기술의 경제적 가치에 크게 영향을 미치며, AI 기술을 가진 근로자는 평균 21%, 최대 40% 높은 임금을 받을 수 있음

### ● AI 기술 중 머신러닝, 텐서플로우, 딥러닝의 임금 프리미엄이 높게 평가




- 옥스퍼드 인터넷 연구소(Oxford Internet Institute)가 2023년 10월 24일 962개 기술과 2만 5천 명을 대상으로 한 연구에서 AI를 포함한 주요 기술의 경제적 가치를 분석한 결과를 발표
  - 연구에 따르면 한 기술의 경제적 가치는 근로자의 여타 역량과 얼마나 잘 결합하는지를 보여주는 '상보성(complementarity)'에 따라 결정됨
  - 특정 기술은 다른 기술과 결합 가능성이 높을수록 경제적 가치가 높아지며, 일례로 데이터 분석과 같은 기술은 여타 고부가가치 기술과 결합할 수 있어 가치가 높지만, 사진 리터칭 같은 기술은 특정 기술과만 결합할 수 있어 가치가 낮게 평가됨
  - 대부분 직업은 여러 기술의 조합이 필요하며, 근로자의 재교육에서 경제적 효율성을 높이려면 기존 기술과 신기술 간 상보성을 극대화할 필요
- AI의 확산은 기술의 경제적 가치에 크게 영향을 미치는 요소로, AI 기술을 가진 근로자는 평균적으로 21% 높은 임금을 획득 가능
  - AI 기술 중 근로자에 대한 경제적 가치(시간당 임금 증가율 기준) 측면에서 상위 5개 기술은 머신러닝(+40%), 텐서플로우(+38%), 딥러닝(+27%), 자연어처리(+19%), 데이터 과학(+17%) 순

〈AI 기술 유형 평균 기술 대비 갖는 임금 프리미엄〉



출처: Oxford Internet Institute, AI comes out on top: Oxford Study identifies the economic value of specific skills, 2023.10.24.

## II. 주요 행사 일정

행사명	행사 주요 개요		
CES 2024		<ul style="list-style-type: none"> <li>- 미국 소비자기술 협회(CTA)가 주관하는 세계 최대 가전·IT·소비재 전시회로 5G, AR&amp;VR, 디지털헬스, 교통·모빌리티 등 주요 카테고리 중심으로 기업들이 최신의 기술 제품군을 전시</li> <li>- CTA 사피로 회장은 가장 주목받는 섹터로 AI를 조명하였으며, 모든 산업을 포괄한다는 의미에서 ‘올 온(All on)’을 주제로 한 이번 전시에는 500곳 이상의 한국기업 참가 예정</li> </ul>	
	기간	장소	홈페이지
	2024.1.9~12	미국, 라스베이가스	<a href="https://www.ces.tech/">https://www.ces.tech/</a>
AIMLA 2024		<ul style="list-style-type: none"> <li>- 머신러닝 및 응용에 관한 국제 컨퍼런스(AIMLA 2024)는 인공지능 및 머신러닝의 이론, 방법론 및 실용적 접근에 관한 지식과 최신 연구 결과 공유</li> <li>- 이론 및 실무 측면에서 인공지능, 기계학습의 주요 분야를 논의하고, 학계, 산업계의 연구자와 실무자들에게 해당 분야의 최첨단 개발 소식 공유</li> </ul>	
	기간	장소	홈페이지
	2024.1.27~28	덴마크, 코펜하겐	<a href="https://ccnet2024.org/aimla/index">https://ccnet2024.org/aimla/index</a>
AAAI Conference on Artificial Intelligence		<ul style="list-style-type: none"> <li>- AI 발전 협회 컨퍼런스(AAAI)는 AI 연구를 촉진하고, AI 분야 연구원, 실무자, 과학자, 학생 및 공학자 간 교류의 기회 제공</li> <li>- 컨퍼런스에서 AI 관련 기술 발표, 특별 트랙, 초청 연사, 워크숍, 튜토리얼, 포스터 세션, 주제 발표, 대회, 전시 프로그램 등 진행</li> </ul>	
	기간	장소	홈페이지
	2024.2.20~27	캐나다, 밴쿠버	<a href="https://aaai.org/aaai-conference/">https://aaai.org/aaai-conference/</a>



홈페이지 : <https://spri.kr/>

보고서와 관련된 문의는 AI정책연구실(jayoo@spri.kr, 031-739-7352)으로 연락주시기 바랍니다.

경기도 성남시 분당구 대왕판교로 712번길 22 글로벌 R&D 연구동(A) 4층  
22, Daewangpangyo-ro 712beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea, 13488