

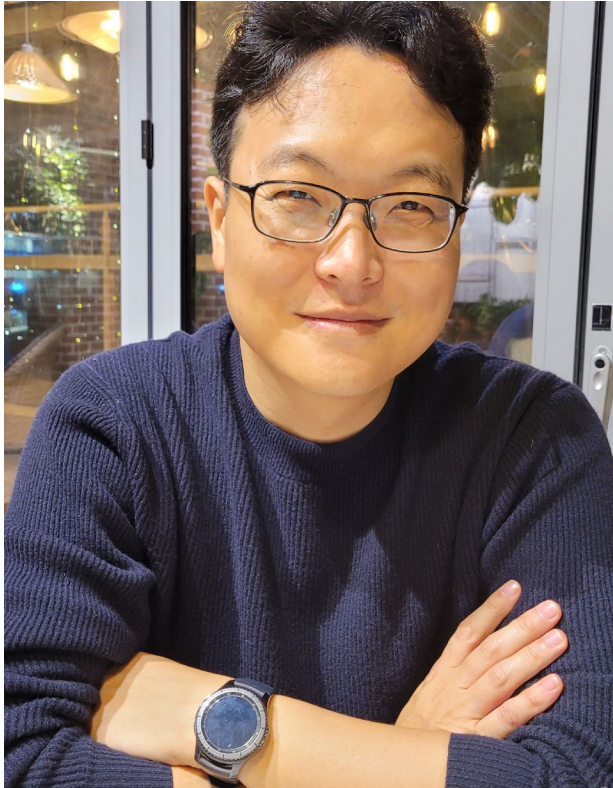
# 인공지능을 활용한 슬기로운 연구생활

2025. 05. 30.

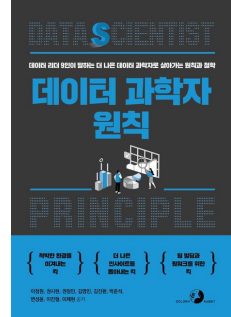
한국에너지기술연구원 이제현

# 반갑습니다. 이제현입니다.

대통령 직속 국가인공지능위원회 산업·공공분과위원 '24.12.- 현재  
 국가과학기술연구회 디지털전환 및 융합 R&D전문가위원 '24.01.- '24.12.  
 Microsoft AI MVP '23.01.- '24.07.



- 도심태양광 모델링
  - 그림자 영향력
  - 머신러닝 개발
- 강화학습 + 휴리스틱 최적화
  - 풍력 터빈 평가
  - 알고리즘 개발
- 업무 효율화
  - 논문 분석, 요약
  - 채용 검증 자동화
  - 연구기관 분석
- 웹기반 플랫폼 구축
  - 연구 데이터 수집



## Pega Devlog

<https://jehyunlee.github.io/>

2025.05.

- thisweek.png

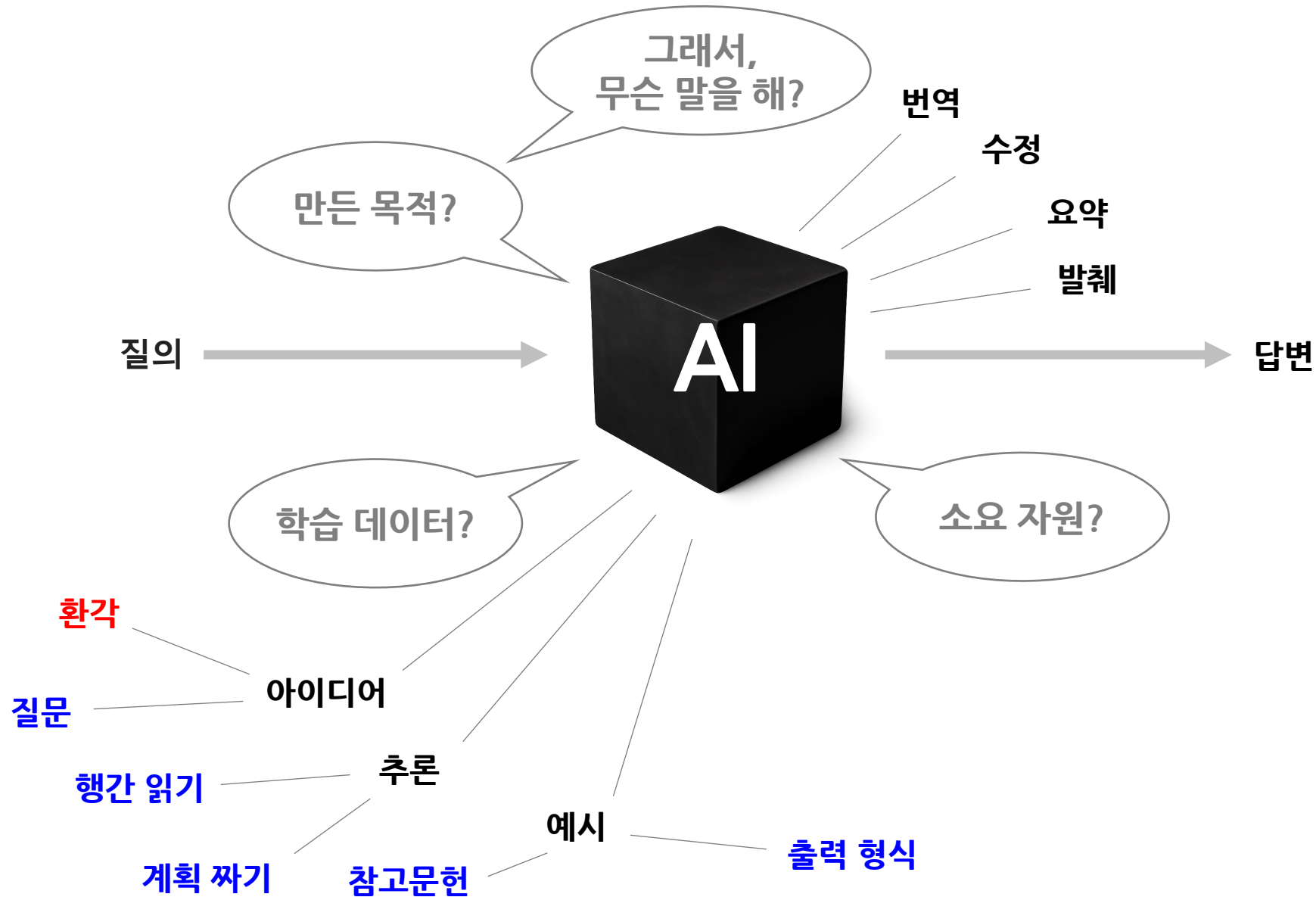
 <p><b>OpenAI</b> Codex 출시</p>	<p><b>2025.05.16</b> 온라인</p>	<p>&lt;/&gt; <b>Codex</b> 클라우드형 코딩 에이전트 · <b>codex-1 모델</b></p> <p>&gt; 병렬 샌드박스 · CLI 지원 · PR 자동 제안</p> <p>👥 ChatGPT Pro/Team/Enterprise 사용자 즉시 접속</p>
 <p><b>Microsoft</b> Build 2025</p>	<p><b>05.19~22</b> 시애틀</p>	<p>🌐 <b>오픈 에이전틱 웹</b> 비전 · <b>GitHub Copilot Agent</b></p> <p>☁️ Azure AI Foundry 1만+ 모델 · <b>NLWeb</b> 오픈소스</p> <p>🔬 <b>Discovery</b> 과학 R&amp;D 플랫폼 · 멀티-에이전트 시뮬레이션</p>
 <p><b>Google</b> I/O 2025</p>	<p><b>2025.05.20</b> 마운틴뷰</p>	<p>🧠 <b>Gemini 2.5</b> · AI Mode Search · Deep Search · Ultra</p> <p>👁️ <b>Project Astra</b> 실시간 멀티모달 · Gemini Live 통합</p> <p>📺 Veo3 · Flow · Imagin 4 · <b>Jules AI Agent</b> <b>BETA</b></p>
 <p><b>Anthropic</b> Code with Claude</p>	<p><b>2025.05.22</b> 샌프란시스코</p>	<p>👑 <b>Claude 4</b> (Opus 4 · Sonnet 4) 코딩 특화</p> <p>&gt; <b>Claude Code CLI</b> <b>BETA</b> · 터미널 에이전틱 코딩</p> <p>안전성 · 투명성 강조 개발자 생태계 구축</p>
 <p><b>NVIDIA</b> Computex 2025</p>	<p><b>05.19~23</b> 타이베이</p>	<p>⚙️ <b>Blackwell GPU</b> AI 슈퍼컴퓨터 · 대만 정부 협력</p> <p>📦 <b>NVLink Fusion</b> 개방 · Isaac GR00T N1.5</p> <p>☁️ DGX Cloud Lepton · AI 훈련 하드웨어 주도권</p>

# 2025.05.28. DeepSeek-R1-0528

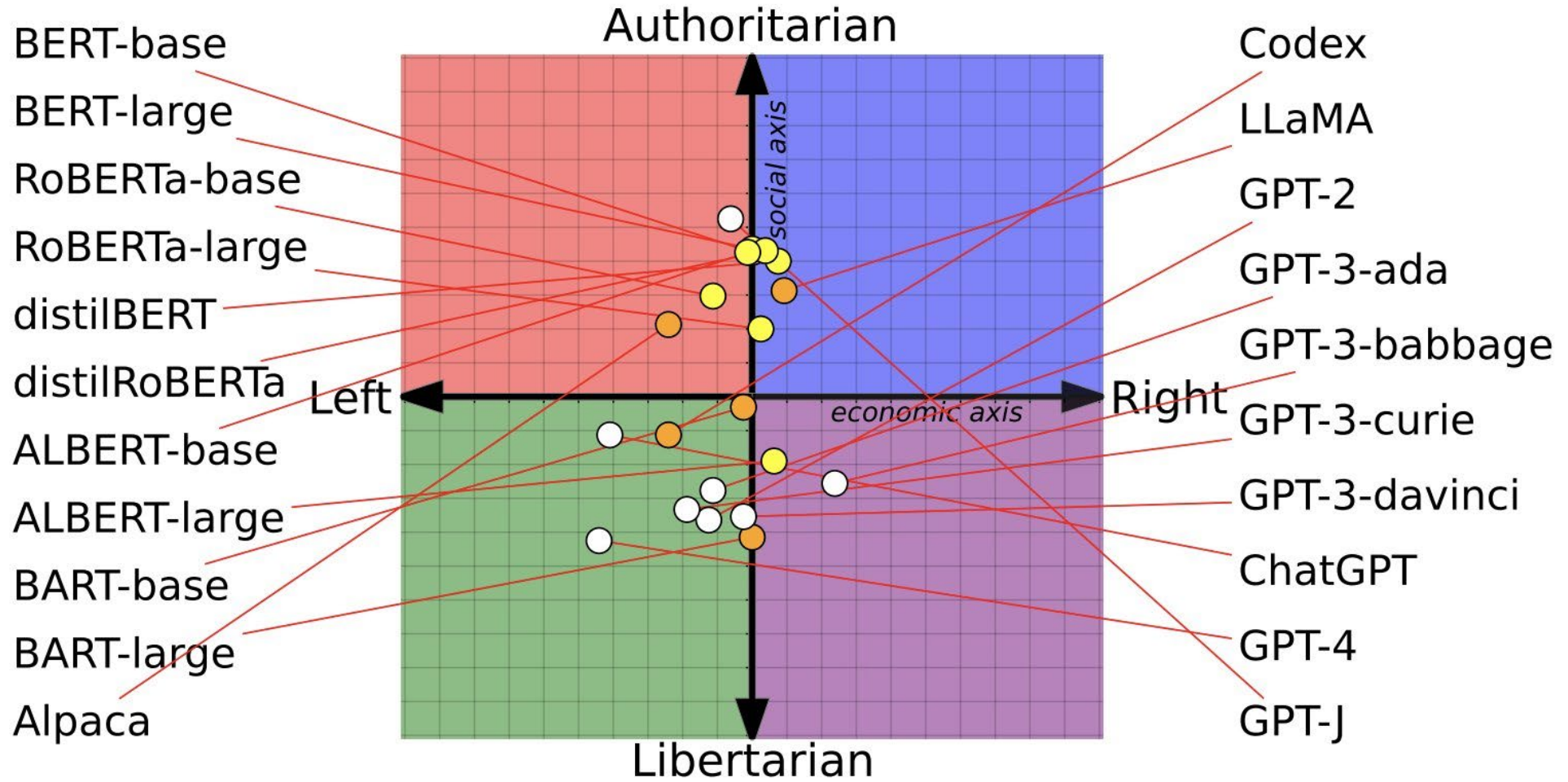
Benchmarks	DeepSeek-R1-0528	OpenAI-o3	Gemini-2.5-Pro-0506	Qwen3-235B	DeepSeek-R1
<b>AIME 2024</b> 数学竞赛 <i>pass@1</i>	91.4	91.6	90.8	85.7	79.8
<b>AIME 2025</b> 数学竞赛 <i>pass@1</i>	87.5	88.9	83.0	81.5	70.0
<b>GPQA Diamond</b> 科学测试 <i>pass@1</i>	81.0	83.3	83.0	71.1	71.5
<b>LiveCodeBench</b> 代码生成 <i>pass@1</i>	73.3	77.3	71.8	66.5	63.5
<b>Aider</b> 代码编辑 <i>pass@1</i>	71.6	79.6	76.9	65.0	57.0
<b>Humanity's Last Exam</b> 推理与百科知识 <i>pass@1</i>	17.7	20.6	18.4	11.75	8.5

# 1. 생성 AI의 본질

# 거대 언어 모델 Large Language Model



# 정치적 편향성



# 개성

## Do LLMs Possess a Personality? Making the MBTI Test an Amazing Evaluation for Large Language Models

Keyu Pan<sup>1</sup>, Yawen Zeng<sup>1</sup>

<sup>1</sup>ByteDance, Beijing, China

{pankeyu96, yawenzeng11}@gmail.com

	Type	Personality Descriptions
ChatGPT	ENTJ	self-confident, decisive, and possess innate leadership skills.
GPT-4*	INTJ	experts skilled in achieving their own goals.
Bloom7b	ISTJ	pragmatic, responsible, values tradition and loyalty.
BaiChuan7b	ENFP	smart, curious, and imaginative.
BaiChuan13b	INFP	highly adaptable and idealistic
OpenLlama7b	INFJ	has strong insight into people and adheres to one's own values.

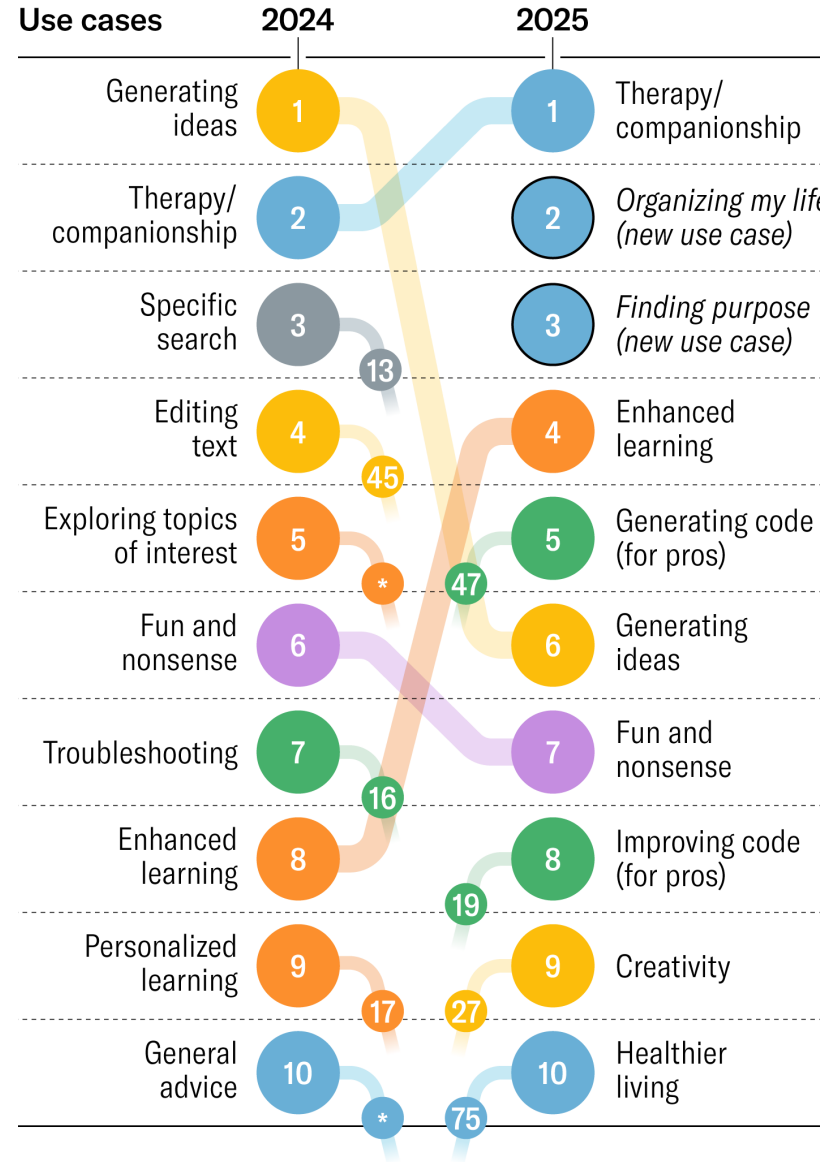
# 교감하는 말동무

## Top 10 Gen AI Use Cases

The top 10 gen AI use cases in 2025 indicate a shift from technical to emotional applications, and in particular, growth in areas such as therapy, personal productivity, and personal development.

### Themes

<span style="color: blue;">●</span> PERSONAL AND PROFESSIONAL SUPPORT	<span style="color: green;">●</span> TECHNICAL ASSISTANCE AND TROUBLESHOOTING
<span style="color: yellow;">●</span> CONTENT CREATION AND EDITING	<span style="color: purple;">●</span> CREATIVITY AND RECREATION
<span style="color: orange;">●</span> LEARNING AND EDUCATION	<span style="color: grey;">●</span> RESEARCH, ANALYSIS, AND DECISION-MAKING



“Where I’m from, in South Africa, mental healthcare barely exists; there’s a psychologist for 1 in every 100,000 people and a psychiatrist for 1 in every 300,000 people. Large language models are accessible to everyone, and they can help. Unfortunately, data safety is not a concern when your health is deteriorating, and survival is the morning agenda.”

“I just asked it to create a timeline for me to clean and organize my house before we have guests staying.”

Determining and defining one’s values, getting past roadblocks, and taking steps to self-develop (e.g., advising on what you should do next, reframing a problem, helping you to stay focused)

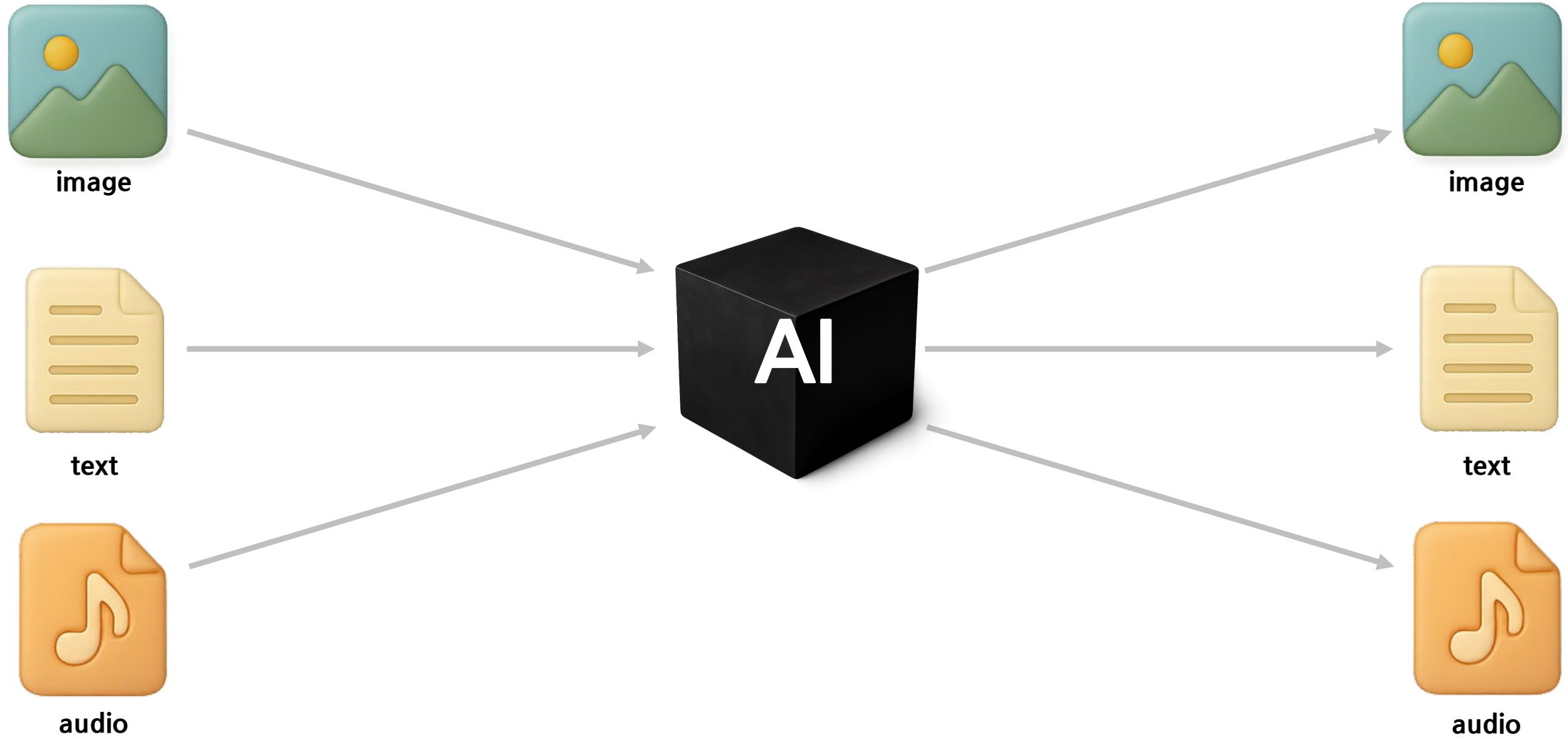
“I just started trying a new eating plan, and calculating the macros with the spreadsheets of online ‘meal planners’ was making me crazy. Now I just ask for recipes based on my given macro needs per meal and take that list shopping. Then I use the air fryer and the rice maker—easy peasy.”

\*Did not make list of top 100 in 2025

Source: Filtered.com



# Multimodality



# Tool Calling ChatGPT DALL.E3

**구상** 클림트의 <KISS>를 모티브로 한 그림.  
그림 속 남녀를 나와 아내의 모습으로 치환.  
키스를 하는 모습 대신 스마트폰을 들고 장난스럽게 셀카를 찍는 모습.  
복장과 배경은 최대한 원본을 반영, 우리 얼굴과 휴대폰 케이스 등 최대한 사실 반영.  
디테일을 살려가며 수 주 동안 작업.

**사고** 단단한 타일 위에 휴대폰 낙하.  
액정 파손. 휴대폰 동작 불가.  
휴대폰 교체 (Note2 → Note5).  
**작업 의욕 상실. 미완성으로 남음 (2017.06.XX.)**

input prompt

An art nouveau styled oil painting. a young man and woman taking selfie. The man {mid-asian, dark-colored skin, curly very short hair, manly-face, smiling, holding the woman tightly with one hand} is holding a smart phone {close up} in his hand on stretching arm. The woman {bright-colored skin, brown hair is decorated with various small flowers, kneeling on field full of tiny yellow, bare foot, pink and white flowers on green grass, smiling, eyes closed} is posing V-sign with her hand attaching on face. The man and woman are facing their chin-to-chin. The man is wearing robe full of rectangular patterns {black-colored rectangular patterns on gold background color}, where the woman is wearing robe full of round patterns {a lot of fantastic circle patterns with various colors on gold background color}.



셀카 DALL.E3 @ChatGPT (2024.06.10.)

# Multimodal Generative AI OpenAI Image Generator



## 2. 연구자들의 활용 사례

☰



하이그레드넷

🔍

홈
채용
대학원
학술
카페
해외

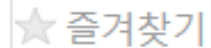
전체
브레인뉴스
테마카페
상담실
토론실

전체 ★ 즐겨찾기 0 🔗

## 논문 리뷰에 AI 답변이?

[논문토론실] | 25.05.14 | 조회수 4,672 | 답변 18 | 활동이력📄

최근 투고한 논문에 리뷰를 받아보니 AI로 만든 글이 왔습니다.  
 투고한 논문 내용 요약, 장점, 한계, 개선을 위한 제안 이렇게 불릿을 나눠놓고  
 그 안에 내용들은 피상적이고 일반적인 내용만 담겨있습니다.  
 AI를 참고해 작성한 것도 아니고 그냥 AI 아웃풋을 그대로 복붙한 수준입니다.  
 이런 리뷰에 어떻게 대응해야할까요?



| 25.05.14 | [활동이력](#)

얼마전 AI로 쓴 논문 비판 글을 썼더니, 전혀 문제없다는 다수의 의견에 엄청 폭격을 맞았던 사람입니다..

참조: <https://go.hibrain.io/s/1axWps7dHW9ymTuQTFekR>

조회수 12,300회이지만, 사실상 타 연구자분들은 동의하지 않았습니다.

이분들 논리대로라면, AI로 쓴 논문을 AI가 평가하는 것도 무방한거지요.. 사람이 굳이 읽을 필요가...

하지만, 원글님이 이런 경우에 해당되지 않는다면, 학회 에디터에게 항의 이메일을 정중하게 써서 보내시기 바랍니다.

심사 의견이 철회될지 여부는 반반 이지만, 재심사 기회를 잡을 수도 있고, 이 경우 리뷰어 잘못이 명확하므로 원글님이 손해볼 사정도 없습니다.

다만, 참고로.. AI 리뷰어가 인간 리뷰어보다는 더 친절(?)하고 착하게 리뷰 의견을 주기는 합니다.



| 25.05.14 | [활동이력](#)

님이 쓴 글에서 다른 대부분의 연구자 분들의 지적하신 내용이 AI로 쓴 글을 GPT 킬러 하나만 맹신해서 거른다는 점이었습니다

'AI로 전부 작성된 논문을 허용하자' 또는 'AI가 논문을 대신 작성해도 좋다' 이것이 아니라고요

GPT 킬러가 하나의 오차도 없이 AI와 사람이 쓴 글을 정확히 판별한다면 님이 쓴 방법이 문제가 없겠지만

다른 AI 전문가들이 지적하는 바와 같이 GPT 킬러와 같은 툴의 정확성은 100%가 아닙니다

그래서 님이 주장한대로 GPT 킬러가 100%라고 했으니 이걸 100% AI가 쓴 내용이다 이 주장 역시 100% 맞지 않습니다

김박사넷만 가봐도 가끔 본인이 전부 작성한 내용이 GPT 킬러 같은 툴에서 100% AI가 작성한 내용이라고 뜬다고 나오는 글이 있더라고요

GPT 킬러에서 높은 확률로 AI라고 나온다고 할지라고, 기본적으로 법의 경우 우리가 무죄추정의 원칙을 깔고 가듯, 현재 상황에서는 논문을 사람들이 썼다고 생각하고 리뷰하는 것이 옳다고 봅니다

GPT 킬러와 같은 툴이 발전해서 AI와 사람이 쓴 글을 정확하게 구별할 때 까지는 GPT 킬러의 숫자가 단순히 높다고 논문을 읽지도 않고 리젝하는 것은 옳지 못하다는 것이 님이 쓴 글에 대한 대부분의 연구자들이 동의하는 내용이겠죠

## ChatGPT로 쓴 논문 유감..

| 등록일 25.04.10 | 조회수 13,335 | 즐겨찾기수 7 | [활동이력](#) ↻

**관련전공** 건축공학,고분자공학,공학일반,교통공학,금속공학,기계공학,기타공학,농공학,산림공학,산업공학,생물학,자원공학,재료공학,전기공학,전자/정보통신공학,제어계측공학,조선공학,컴퓨터공학,토목공학,항공

최근에 ChatGPT로 쓴 논문들이 다수 리뷰 요청이 옵니다.

이제는 버릇처럼 GPTKiller 먼저 돌려보고 있습니다.

과거보다 로직이 좋아져서 상당 수준으로 잡아내는 것 같습니다.

물론, 사람이 쓴 글이나 숫자로 된 표의 일부도 GPTKiller에 유사로 종종 걸린다는 사실도 잘 알고 있습니다.

그러나, 한두 문장이 아니라 페이지 단위로 100% 일치가 나오면 그건 사람이 쓴게 아닙니다.

본인도 안읽고 쓴 글을 타인에게 읽으라고 하는 것은 결국 학계를 고사시킬 불행한 일입니다.

주로 요즘 젊은? 연구자들이 ChatGPT 이용하는 경향을 보면,

실증 통계분석만 자신이 하고,

그 결과를 AI에게 던져서 초록, 서론, 이론적 배경, 결론을 기계적으로 뽑는 경우들이 다수 보이는군요.

(이렇게 하는게 빠르게 '논문쓰는 법' 이라며 클래스를 광고하는 업체도 있긴 합니다..)

아직 정확한 학술기준이 없다고는 하지만 바로 리젝 들어갑니다.

AI의 도움을 받는 것과 AI에 올라타는 것은 완전히 다른 일입니다.

The tail wagging the dog 꼬리가 개를 흔든다..

# 쟁점 정리

- 1. GPT로 리뷰를 해서 보내면 안되나?
- 2. GPT로 논문을 쓰는 건 괜찮나?
- 3. AI로 그림을 그리는 건 괜찮나?
- 4. GPT killer 등 검사 결과는 믿을 만 한가?
- 5. 만약 그렇지 않다면, 어떻게 검사해야 하나?

# 비밀 유지 의무



원병묵

1일 · 🌐

최근에 연구 과제 심사를 하신 분들은 아시겠지만 연구 계획서를 ChatGPT 등 생성형 AI에 업로드하는 것이 금지됩니다.

아래와 같이 국가연구개발혁신법 제40조 비밀 유지 의무 규정을 위배하는 것으로 판단될 수 있습니다.

~~~~~

제40조(비밀 유지의 의무)

제39조 각 호의 어느 하나에 해당하거나 해당하였던 사람은 업무 수행과정에서 알게 된 비밀을 다른 사람에게 누설하거나 직무상 목적 외의 용도로 이용하여서는 아니 된다.

제41조(벌칙)

제40조를 위반하여 업무 수행과정에서 알게 된 비밀을 다른 사람에게 누설하거나 직무상 목적 외의 용도로 이용한 사람은 2년 이하의 징역 또는 2천만원 이하의 벌금에 처한다.

# NRF 생성형 AI 도구의 책임 있는 사용을 위한 권고사항

## 권고사항

### 1

한국연구재단이 지원하는 연구개발과제의 평가에 참여하는 평가위원\*은 각종 평가자료\*\*를 생성형 AI 도구에 입력(업로드)하지 말아야 합니다.

\* 평가위원 : 온·오프라인에서 서면 또는 패널 방식으로 과제의 선정·중간·결과 평가에 참여하는 위원

\*\* 평가자료 : 연구개발계획서, 단계/최종보고서, 평가의견, 평가위원 리스트 등 과제평가와 관련된 제반 자료

\* 한국연구재단의 연구개발과제 평가 등에 관여한 평가위원 등이 ChatGPT 등 생성형 AI에 해당 연구개발과제 관련 정보를 업로드하는 행위는 국가연구개발혁신법 제40조 비밀 유지 의무에 위반될 수 있음

### 2

한국연구재단 지원과제의 신청자 및 수행자는 연구개발계획서 및 단계/최종보고서 작성 과정에서 생성형 AI 도구를 사용한 경우, 해당 계획서 및 보고서에 AI 도구 사용 내역을 기술할 것을 권장합니다.

# AI 도구의 논문 리뷰 사용에 대한 출판사별 규정

| Category                  | Publisher                        | Usage of AI                                                                                                                                         |
|---------------------------|----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| No usage                  | AAAS 2024                        | ---                                                                                                                                                 |
|                           | AIP 2024                         | ---                                                                                                                                                 |
|                           | Cambridge University Press 2024a | ---                                                                                                                                                 |
|                           | Elsevier 2024                    | ---                                                                                                                                                 |
|                           | Emerald 2024                     | ---                                                                                                                                                 |
|                           | IEEE 2024                        | ---                                                                                                                                                 |
|                           | MDPI 2024                        | ---                                                                                                                                                 |
|                           | MIT Press 2024b                  | ---                                                                                                                                                 |
|                           | Springer Nature 2024             | ---                                                                                                                                                 |
| Taylor & Francis 2024     | ---                              |                                                                                                                                                     |
| Partly usage              | ACL 2023                         | Paraphrasing the review, checking proofs, explaining concepts that the reviewer doesn't know                                                        |
|                           | ACM 2023; 2024                   | Enhancing the quality and readability of review reports (not necessary to disclose)                                                                 |
|                           | Sage 2024                        | Looking for reviewers                                                                                                                               |
|                           | JAMA Network 2023                | Using AI tools that does not violate the confidentiality policy (the name and usage of the tool must be declared)                                   |
| Privately hosted AI tools | Frontiers 2020; 2024             | AI tool: AIRA<br>Assessment of language quality, the integrity of the figures, detection of plagiarism, identifying potential conflicts of interest |
|                           | Wiley 2024                       | AI tool: Research Exchange<br>Improving the match of reviewers<br>screening research integrity                                                      |

## For reviewers

The use of generative AI and AI-assisted technologies in the journal peer review process

When a researcher is invited to review another researcher's paper, the manuscript must be treated as a confidential document. Reviewers should not upload a submitted manuscript or any part of it into a generative AI tool as this may violate the authors' confidentiality and proprietary rights and, where the paper contains personally identifiable information, may breach data privacy rights.

# Three AI-powered steps to faster, smarter peer review

Tired of spending countless hours on peer reviews? An AI-assisted workflow could help.

By [Dritjon Gruda](#) 



Credit: Getty/Shutterstock

## Scan, dictate, refine

I break the review process down into three simple steps:

**Scan.** Quickly browse the abstract, introduction, methods and results, focusing on the big picture. If the analysis looks solid, read the rest of the paper. If you detect glaring flaws, however, you'll know it's probably one to reject – no need to line-edit the entire manuscript.

**Dictate.** Use dictation in the text-editing tool of your choice (for example, Voice Access in Windows or Voice Control on the Mac) to capture real-time thoughts as you read. This way, you avoid having to scribble notes or needing to recall and type up your feedback later – a major time-saver.

**Refine.** Feed your dictated notes into an [offline large language model \(LLM\)](#) to clarify and organize your feedback. A simple prompt such as “Write a critical reviewer letter based on the following notes. Maintain a professional tone throughout” will do. Don't know how to code? No problem. Tools such as GPT4ALL (see ‘How to set up and run a local LLM’) allow you to load and run LLMs locally and offline, so there's no need to upload sensitive manuscripts to the cloud. Confidentiality is non-negotiable when it comes to unpublished research, and uploading content externally can invite ethical or even legal trouble.

But let's be clear: LLMs aren't there to crank out a full review on your behalf. Although survey data published on the arXiv preprint server in January suggest that automated scholarly paper review (ASPR) can help to speed up evaluations and sharpen structure, [ASPR tools still struggle with domain-specific expertise, bias and data-security issues<sup>1</sup>](#). Instead, use the LLM to spot redundancy, refine phrasing and organize suggestions (for prompts, see [my previous column](#)). You – the reviewer – must make the final judgement on the paper's methodology, findings and overall contribution to the field.

# AI 도구의 논문 작성에 대한 출판사별 규정

| 저널/출판사           | AI 사용 허용 여부                       | 사용시 명시 섹션. 명시 내용                                           | 주요 참고사항 (2025년 5월 기준 참고)                                                                                                                                                    | Reference                                                                                                                                                                                                                                                                             |
|------------------|-----------------------------------|------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Science          | 텍스트 생성 AI 사용 금지<br>분석 도구로 사용 시 명시 | Methods                                                    | AI 생성 텍스트, 그림, 영상 모두 금지.<br>데이터 분석이나 아이디어 구체화 등 연구 과정에서의 보조적 사용은 허용되나 그 내용을 명확히 밝혀야 함.<br>위반시 연구 부정행위로 간주될 수 있음.                                                            | <a href="https://www.science.org/content/page/science-journals-editorial-policies">https://www.science.org/content/page/science-journals-editorial-policies</a>                                                                                                                       |
| Springer Nature  | 제한적 허용<br>텍스트 생성 AI는 신중 사용        | Methods 또는 적절한 섹션                                          | AI는 저자로 인정되지 않음. 원고 작성 과정에서 AI 도구 사용 시, 그 사용 방식과 도구의 명칭을 투명하게 공개해야 함. 최종 내용에 대한 책임은 저자에게 있음.                                                                                | <a href="https://www.nature.com/nature-portfolio/editorial-policies/ai">https://www.nature.com/nature-portfolio/editorial-policies/ai</a>                                                                                                                                             |
| JAMA Network     | 제한적 허용<br>원고 작성에는 사용 권장 안함        | Methods<br>AI 도구, 버전 및 사용 방식 상세 기술                         | AI는 저자로 인정되지 않음. 연구 설계, 데이터 수집 및 분석에 AI 도구 사용 시 상세히 기술해야 함. AI가 생성한 텍스트를 그대로 사용하는 것은 지양해야 하며, 사용 시 반드시 그 사실을 밝히고 인간 저자가 검토 및 편집해야 함. 최종 내용에 대한 책임은 저자에게 있음.                 | <a href="https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecUseofAIinPublicationandResearch">https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecUseofAIinPublicationandResearch</a>                                                         |
| Elsevier         | 허용<br>단, 책임은 저자에게 있음              | Methods 또는 별도 선언문                                          | AI는 저자로 인정되지 않음. 생성형 AI 사용 시, 사용 사실과 어떤 도구를 어떻게 사용했는지 명시해야 함. 'Declaration of Generative AI in scientific writing'과 같은 양식을 통해 공개할 수 있음. 저자는 AI가 생성한 내용의 정확성과 독창성에 대한 책임을 짐. | <a href="https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier">https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier</a> |
| ACS (미국화학회)      | 허용<br>단, 책임은 저자에게 있음              | Methods 또는 Acknowledgements<br>사용된 AI 도구 및 사용 방식 명시        | AI는 저자로 인정되지 않음. AI 도구 사용 시 그 사실을 명시하고, 생성된 내용에 대한 과학적 정확성과 독창성은 저자가 책임져야 함.                                                                                                | <a href="https://researcher-resources.acs.org/publish/author_guidelines">https://researcher-resources.acs.org/publish/author_guidelines</a>                                                                                                                                           |
| PLOS             | 허용<br>단, 책임은 저자에게 있음              | Methods 또는 Acknowledgements                                | AI는 저자로 인정되지 않음. AI 도구 사용을 투명하게 공개하고, 저자는 AI가 생성한 모든 내용(텍스트, 이미지, 코드 등)의 정확성, 타당성, 적절한 출처 인용에 대한 책임을 짐.                                                                     | <a href="https://journals.plos.org/plosone/s/ethical-publishing-practice#loc-artificial-intelligence-tools-and-technologies">https://journals.plos.org/plosone/s/ethical-publishing-practice#loc-artificial-intelligence-tools-and-technologies</a>                                   |
| Taylor & Francis | 허용<br>단, 책임은 저자에게 있음              | Methods, Acknowledgements 등 적절한 섹션<br>사용된 AI 도구 및 사용 방식 명시 | AI는 저자로 인정되지 않음. 저자는 AI 도구 사용을 공개해야 하며, 생성된 콘텐츠의 정확성과 독창성에 대한 최종 책임을 짐.                                                                                                     | <a href="https://www.pnas.org/author-center/editorial-and-journal-policies">https://www.pnas.org/author-center/editorial-and-journal-policies</a>                                                                                                                                     |
| PNAS             | 허용<br>단, 책임은 저자에게 있음. 신중한 사용 요구   | Methods<br>사용된 AI 도구 및 사용 방식 명시                            | AI는 저자로 인정되지 않음. AI 도구를 사용하여 텍스트, 그림, 그래픽, 코드 등을 생성하거나 편집한 경우, Methods 섹션에 해당 도구의 이름과 사용 방식을 구체적으로 명시해야 함. 저자는 AI 생성 콘텐츠의 정확성과 독창성에 대한 책임을 져야 함.                            | <a href="https://www.pnas.org/author-center/editorial-and-journal-policies">https://www.pnas.org/author-center/editorial-and-journal-policies</a>                                                                                                                                     |



## Multimodule imaging of the hierarchical equine hoof wall porosity and structure

Mahmoud A. Mahrous<sup>a</sup>, Charul Chadha<sup>b</sup>, Pei L. Robins<sup>c</sup>,  
 Christian Bonney<sup>b</sup>, Kingsley A. Boateng<sup>d</sup>, Marc Meyers<sup>e</sup>, Iwona Jasiuk<sup>b,\*</sup>

<sup>a</sup> Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

<sup>b</sup> Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

<sup>c</sup> Department of Physics, Lawrence University, Appleton, WI, USA

<sup>d</sup> Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

<sup>e</sup> Department of Mechanical and Aerospace Engineering, University of California San Diego, San Diego, CA, USA

### ARTICLE INFO

#### Article history:

Received 20 June 2023

Accepted 11 August 2023

Available online 28 August 2023

#### Keywords:

Porosity

Biological materials

Tubules

Hierarchical structure

Micro-computed tomography ( $\mu$ -CT)

Serial block-face scanning electron

microscopy (SBF-SEM)

### ABSTRACT

The equine hoof wall has a complex, hierarchical structure that can inspire designs of impact-resistant materials. In this study, we utilized micro-computed tomography ( $\mu$ -CT) and serial block-face scanning electron microscopy (SBF-SEM) to image the microstructure and nanostructure of the hoof wall. We quantified the morphology of tubular medullary cavities by measuring equivalent diameter, surface area, volume, and sphericity. High-resolution  $\mu$ -CT revealed that tubules are partially or fully filled with tissue near the exterior surface and become progressively empty towards the inner part of the hoof wall. Thin bridges were detected within the medullary cavity, starting in the middle section of the hoof wall and increasing in density and thickness towards the inner part. Porosity was measured using three-dimensional (3D)  $\mu$ -CT, two-dimensional (2D)  $\mu$ -CT, and a helium pycnometer. The highest porosity was obtained using the helium pycnometer (8.07%), followed by 3D (3.47%) and 2D (2.98%)  $\mu$ -CT. SBF-SEM captured the 3D structure of the hoof wall at the nanoscale, showing that the tubule wall is not solid, but has nano-sized pores, which explains the higher porosity obtained using the helium pycnometer. The results of this investigation provide morphological information on the hoof wall for the future development of hoof-inspired materials and offer a novel perspective on how various measurement methods can influence the quantification of porosity.

© 2023 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The remarkable properties of biological materials, including self-healing, adhesion, and impact resistance, offer inspiration for developing advanced engineering materials [1]. Impact

resistance refers to a material's ability to withstand intense forces or shocks. Porosity is a common feature of naturally occurring impact-resistant materials and those used in biomedical, aerospace, automotive, and packaging applications [2–5]. The hoof wall is a porous material responsible for

\* Corresponding author. Department of Mechanical Science and Engineering University of Illinois at Urbana-Champaign 1206 West Green Street Urbana, IL 6180, USA.

E-mail address: [ijasiuk@illinois.edu](mailto:ijasiuk@illinois.edu) (I. Jasiuk).

<https://doi.org/10.1016/j.jmrt.2023.08.246>

2238-7854/© 2023 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

SBF-SEM and Amira software. The reconstructed image illustrates the irregular surface of the tubule's inner wall. SBF-SEM imaging supported the results obtained earlier from the coronal multiplanar reconstructed image (Fig. 7), indicating that the TMC contains soft tissue, and its inner wall is not smooth. A single 2D image from the stacked volume is illustrated in Fig. 8h, where the magnified area depicted by the overlaid irregular red line in Fig. 8i shows cell boundaries that appear as a wavy structure. The black particles shown on the sample's surface in Fig. 8g resulted from falling debris generated during the cutting process inside, a common problem encountered in SBF-SEM [31].

### 3.6. Limitations and future scope

The present study has several limitations. The number of samples and studied locations are limited. Thus, results cannot be generalized to other areas within the hoof wall. Additionally, samples were extracted from one horse, representing only one species. The absence of statistically significant differences in the area fraction porosity between different zones within the hoof wall structure can be attributed to the relatively small number of samples under examination. To mitigate this limitation, future studies could consider a broader range of horse breeds and individuals to account for natural variations and enhance the overall validity of the findings. Such studies would build upon the foundation laid by this research, paving the way for further characterization of the hoof wall and other materials.

## 4. Conclusions

We explored the hierarchical structure of the hoof wall. We utilized various measurement methods for quantifying hoof wall porosity using state-of-the-art 2D and 3D multimodal imaging techniques, including  $\mu$ -CT and SBF-SEM. Additionally, we used the helium pycnometer, a novel approach in biological materials characterization, to further enhance our understanding of the hoof wall porosity. We conducted  $\mu$ -CT imaging at resolutions of 0.53  $\mu$ m and 13.35  $\mu$ m to obtain a comprehensive statistical analysis of the hoof wall microstructure and TMCs region. SBF-SEM was used for the first time to study the microstructure of the hoof wall at the nanoscale level.

Our conclusions from this research are as follows:

- 3D images acquired from  $\mu$ -CT indicate that the tubules have thin bridges segmenting them into air pockets, which start in the middle of the stratum medium and increase in density towards the inside of the hoof wall.
- The tubules exhibit some degree of periodic waviness, which may be an intrinsic feature or a result of dryness and moisture loss caused by X-ray exposure during  $\mu$ -CT testing.
- High-resolution  $\mu$ -CT revealed that tubules are filled with tissue near the outer surface and are empty in the middle to inside sections of the hoof wall.
- Statistical analysis of the TMCs indicated that the average values of Eq diameter, surface area, volume, and sphericity are 69.43  $\mu$ m,  $3.55 \times 10^4 \mu\text{m}^2$ ,  $4.53 \times 10^5 \mu\text{m}^3$ , and 0.86, respectively.

- Different techniques yielded different porosity measurements. While 2D  $\mu$ -CT resulted in a porosity of 2.98%, similar to previous studies, the 3D calculation showed an increase in porosity to 3.47%, which is considered more accurate because the estimates are based on the volume of the tubule cavities, rather than the radii of the cavities from a 2D image. These measurements reflect the porosity at the microscale and above.
- Helium pycnometer measurements detected nanoscale pores that were not visible using standard imaging techniques like  $\mu$ -CT or SEM, which resulted in a higher average porosity of 8.07%. Thus, it is an accurate technique that can detect pore sizes ranging from microscale to nanoscale.
- SBF-SEM successfully provided a nanoscale image of the hoof wall, which is challenging to obtain using transmission electron microscopy or any 3D microscopy technique. SBF-SEM imaging showed nano-sized pores within the tubule wall, which have not been previously detected, and explains the higher porosity measured using the helium pycnometer.

Mechanics studies of the hoof wall and bioinspired materials under impact will benefit from the morphometric data presented in this work. Drop tower tests indicated tubule density strongly affects crack propagation in bioinspired samples [24], and the elliptical shape and bridges of tubules are hypothesized to have essential mechanical functions [17,24]. Therefore, it is crucial to identify methods to accurately quantify the structure of the hoof wall with metrics such as sphericity, porosity, and tubule density. Moreover, the hierarchical porosity observed in this work should be further explored as a potential design motif for bioinspired impact-resistant materials.

### Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used GPT-3.5 to improve readability and clarity. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

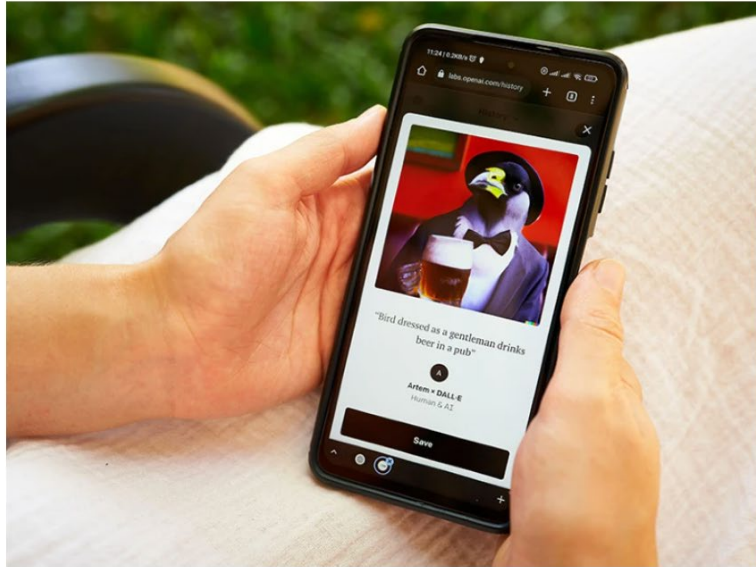
The authors acknowledge the researchers at the Core facilities at the Carl R. Woese Institute for Genomic Biology (IGB) for their assistance with micro-computed tomography ( $\mu$ -CT) and serial block-face scanning electron microscopy (SBF-SEM). The staff scientists of the Microscopy Suite at the Beckman Institute for Advanced Science and Technology at the University of Illinois

# 이미지 생성은 엄금합니다.

EDITORIAL | 07 June 2023

## Why *Nature* will not allow the use of generative AI in images and video

Saying 'no' to this kind of visual content is a question of research integrity, consent, privacy and intellectual-property protection.

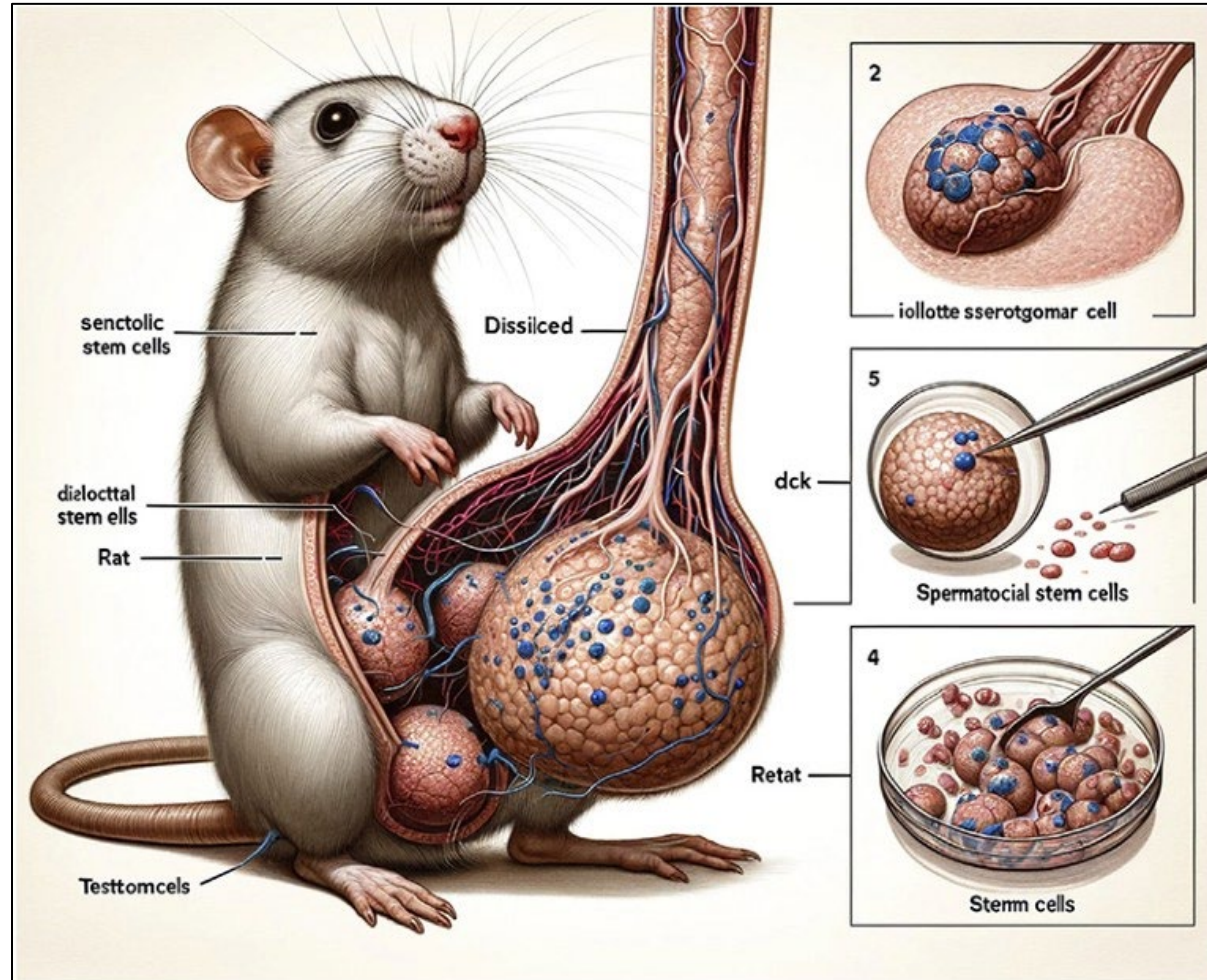


*Nature* will not publish imagery created wholly or partly using generative AI. Credit: Artem Medvediev/Alamy

Should *Nature* allow generative artificial intelligence (AI) to be used in the creation of images and videos? This journal has been discussing, debating and consulting on this question for several months following the explosion of content created using generative AI tools such as

- 연구 무결성  
출판의 무결성은 투명성에 의해 뒷받침되나 생성 AI 도구는 이미지 검증을 위한 출처를 제공하지 않음.
- 저작자 표시  
생성 AI 도구는 사용 또는 인용된 저작물에 대한 저작자 표시의 핵심 원칙을 준수하지 않음.
- 동의 및 허가  
생성 AI 도구는 사람의 신원이나 지적 재산과 관련된 동의를 얻을 수 없음.
- 법률 및 개인정보 보호  
생성 AI 시스템은 종종 저작권이 있는 이미지를 무단으로 학습하고, 동의 없이 딥페이크를 생성하여 개인정보를 침해할 수 있음.

# 이미지에 대해 특히 엄격한 이유



## References

1. Guo, X., Dong, L. & Hao, D. *Front. Cell Dev. Biol.* **11**, 1339390 (2024); retraction **12**, 1386861 (2024).

## Retraction: Cellular functions of spermatogonial stem cells in relation to JAK/STAT signaling pathway

Frontiers Editorial Office\*

🔔 This article is a retraction of:

RETRACTED: Cellular functions of spermatogonial stem cells in relation to JAK/STAT signaling pathway

[Read original article >](#)

A Retraction of the Review Article

[Cellular functions of spermatogonial stem cells in relation to JAK/STAT signaling pathway](#)

by Guo X, Dong L and Hao D (2024). *Front. Cell Dev. Biol.* **11**:1339390. doi: [10.3389/fcell.2023.1339390](https://doi.org/10.3389/fcell.2023.1339390)

Following publication, concerns were raised regarding the nature of its AI-generated figures. The article does not meet the standards of editorial and scientific rigor for Frontiers in Cell and Developmental Biology; therefore, the article has been retracted.

This retraction was approved by the Chief Executive Editor of Frontiers. Frontiers would like to thank the concerned readers who contacted us regarding the published article.

# GPT killer 등은 AI 생성 글을 잘 잡아내나?

최신뉴스

## [샷!] "챗GPT 안 썼다. 억울하다"...AI 역설

송고 2025-04-07 05:50



이승연 기자  
+ 구독



오인균 기자  
+ 구독

대학생·취준생 "직접 작성했는데 AI 생성률 높게 나와"  
"생성형 AI, 글을 베끼지 않았는데 베꼈다고 오판 가능"  
'AI처럼 보이지 않도록' 챗GPT 탐지 서비스 역이용



챗GPT

[로이터 연합뉴스 자료사진]

고려대 에브리타임 이용자들은 "내가 썼는데 왜 챗GPT가 썼다고 하느냐", "복학하니가 이상한 표절 검사가 생겨서 고통받는다", "GPT 탐지기 만든 사람 천벌 받았으면 좋겠다" 등 억울함을 호소했다.

연세대 에브리타임 한 이용자는 "챗GPT 잡는 방법이 존재하는 것은 맞나"라며 "GPT제로(미국의 AI 탐지기)도 정확성이 떨어지고 판정 오류 많다고 한다. 교수들이 어떤 생각인지 궁금하다"고 적었다.

최병호 고려대 AI 연구소 교수는 "챗GPT, 그록, 제미니 등 각 생성형 AI는 버전마다 글을 작성하는 독특한 패턴이 있고, 이를 거꾸로 생각하면 패턴 탐지와 무력화가 충분히 가능하다는 뜻이 된다"고 설명했다.

그러면서 "베낀 것을 탐지하는 것, 베낀 것이 아닌데 베꼈다고 오판하는 것, 베낀 것을 그렇게 보이지 않도록 회피하는 것 모두 가능하다"고 지적했다.

실제로 앞서 아랍에미리트(UAE)에 있는 뉴욕대 아부다비(NYUAD)의 탈랄 라완·야시르 자키 교수팀이 2023년 8월 발표한 조사 결과에 따르면 GPT제로와 오픈AI의 'AI 텍스트 탐지기'의 오판율은 각각 31.55%, 49.37%였다.

두 탐지기는 챗GPT가 작성한 답안 10개 중 3~5개를 학생이 작성했다고 잘못 분류했다.



"GPT가 쓴 글이 아닌데, 왜..."

[연세대 에브리타임 캡처. DB 및 재판매 금지]

# AI가 쓴 글을 가장 잘 확인하는 방법은?

People who frequently use ChatGPT for writing tasks are accurate and robust detectors of AI-generated text

Jenna Russell<sup>1</sup> Marzena Karpinska<sup>2</sup> Mohit Iyyer<sup>1,3</sup>

<sup>1</sup>University of Maryland, College Park <sup>2</sup>Microsoft <sup>3</sup>UMass Amherst  
{jennarus, miyyer}@umd.edu, mkarpinska@microsoft.com

## Abstract

In this paper, we study how well *humans* can detect text generated by commercial LLMs (GPT-4O, CLAUDE-3.5-SONNET, O1-PRO). We hire annotators to read 300 non-fiction English articles, label them as either human-written or AI-generated, and provide paragraph-length explanations for their decisions. Our experiments show that annotators who frequently use LLMs for writing tasks excel at detecting AI-generated text, even without any specialized training or feedback. In fact, the majority vote among five such “expert” annotators misclassifies only 1 of 300 articles, significantly outperforming most commercial and open-source detectors we evaluated even in the presence of evasion tactics like paraphrasing and humanization. Qualitative analysis of the experts’ free-form explanations shows that while they rely heavily on specific lexical clues (“AI vocab-

In Alaska, a pilot drops turkeys to rural homes for Thanksgiving

A half-dozen villagers in Napakiak, on the Kuskokwim River's west bank, gathered near a gravel airstrip last Thursday to watch a small plane circle overhead. ... This crowd was waiting for a seasoned pilot who had a tradition: dropping Thanksgiving turkeys to homes scattered across miles of tundra and frozen waterways.

The pilot, 47-year-old Alaskan flyer Erik Fosnes, has been doing this for nearly a decade, working with volunteers from a regional nonprofit called Delta North Outreach. "We tried shipping turkeys one year by cargo, but half never made it in time," said Fosnes, running a hand through the frost on his jacket sleeve after landing. "So I said, "What if I just fly them in myself?"" He shrugged as if that were the most ordinary idea, then laughed. "Folks around here have gotten used to it."

Looks human-written    Looks AI-generated

Annotator #4  
content writer,  
frequently uses ChatGPT

Annotator's Decision

✓ AI-generated

Confidence

1 ——— 4 ——— 5  
(Least Confident) (Most Confident)

Explanation

... Lots of the quotes felt realistic, but many of the quotes did not need a narration alongside it such as with "He shrugged as if that were the most ordinary idea, then laughed." ... could have been shortened to get more facts in about what people in Alaska face and why they face such limited transportation from the rest of the world. Also, it got sentimental and corny at times too.

Figure 1: A human expert’s annotations of an article generated by OpenAI’s o1-PRO with humanization. The expert provides a judgment on whether the text is written by a human or AI, a confidence score, and an explanation (including both free-form text and highlighted spans) of their decision.

# 쟁점 정리

- 1. GPT로 리뷰를 해서 보내면 안되나?
  - 웬만하면 안 됨. 저널 정책 확인 필요
- 2. GPT로 논문을 쓰는 건 괜찮나?
  - 웬만하면 됨. 저널 정책 확인 필요.
- 3. AI로 그림을 그리는 건 괜찮나?
  - Nature는 안 됨. 다른 곳도 정책 확인 필요.
- 4. GPT killer 등 검사 결과는 믿을 만 한가?
  - 아니요. 못 믿습니다.
- 5. 만약 그렇지 않다면, 어떻게 검사해야 하나?
  - 평소에 AI 많이 쓰는 사람한테 보여주고 물어보세요.

# 3. Deep Research

# Deep Research

## ● 딥시크보다 3배 정확한 추론 모델 '딥리서치' 공개

### 오픈AI '딥 리서치' 개요

### 챗GPT 연구용 인공지능(AI) 에이전트

- 사용자 명령에 따라 최대 30분간 웹 탐색하고 종합 분석해 **전문가 수준의 보고서** 작성
- 다양한 요청 사항을 반영한 **개인화된 쇼핑**에도 특화
- 추론에 특화된 **오픈AI o3** 모델 활용
- 고난도 AI 벤치마크 평가에서 정확도 **26.6%** 기록 (딥시크-R1 9.4%, GPT-4o 3.3%)
- **챗GPT 프로**(월 200달러) 구독자에게 우선 제공, 월 최대 질문 한도 **100개**

"월 200달러를 내면 신세계가 펼쳐진다는데, 차마 거기까진 돈을 못 썼어요. 그런데 월 20달러 내는 이용자에게도 서비스를 해준다니 기대가 큼니다."

최근 과학기술계를 포함한 연구자들 사이에서 중국 생성형 인공지능(AI) 딥시크보다 더 뜨거운 반응을 얻는 AI모델은 바로 오픈AI의 '딥리서치'다. 딥리서치를 써본 연구자들은 박사급 연구원 한 명을 조수로 쓰고 있는 듯하다는 평가를 내놓는다.

딥리서치를 활용해 논문을 쓰고, 학술지 게재 승인을 받았다는 연구자 사례도 나온다. 그런 딥리서치의 이용 저변이 훨씬 넓어질 전망이다.

# Deep Research



2025.03.05.



12  
December  
2024

1  
January  
2025

2  
February  
2025



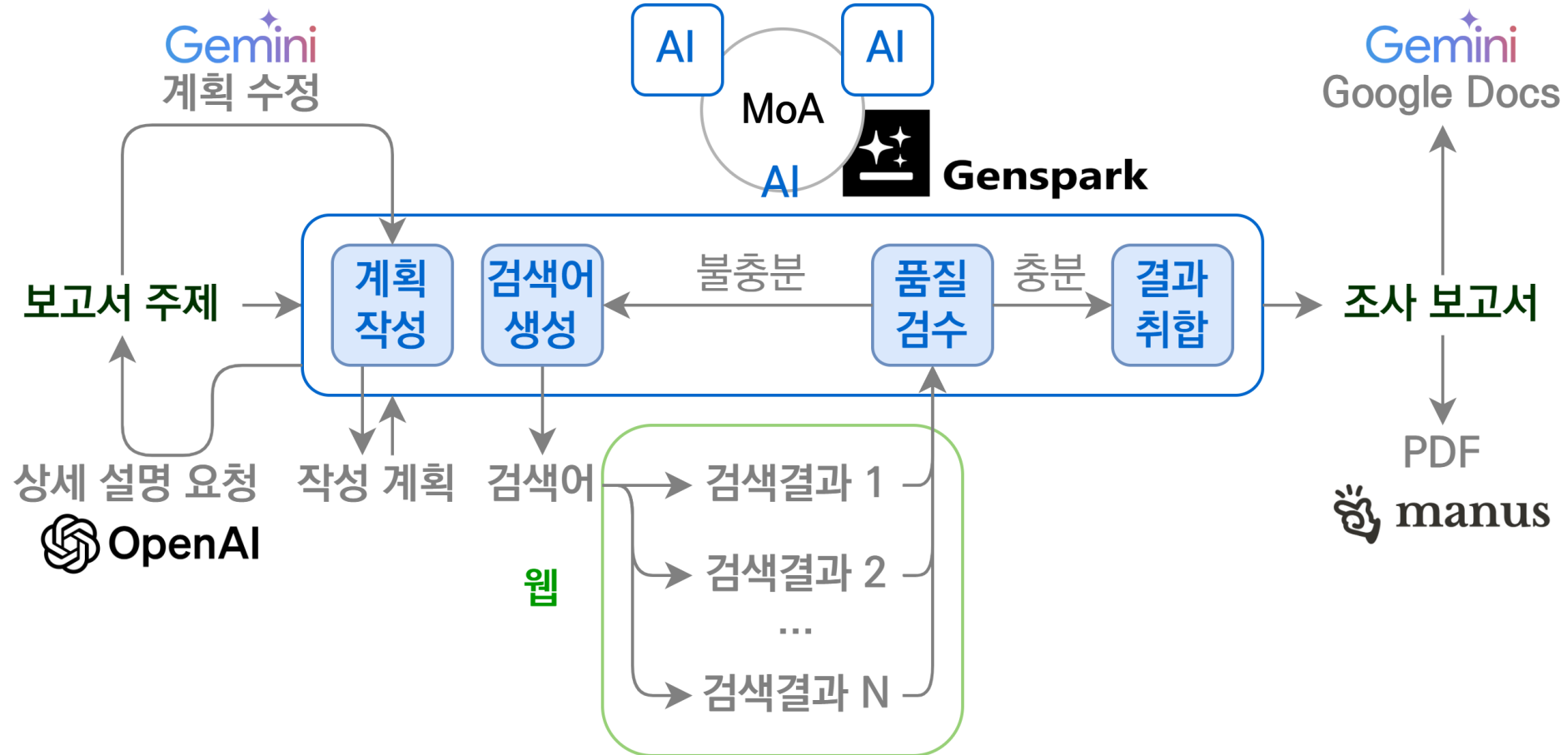
| SUN | MON | TUE | WED       | THU | FRI | SAT |
|-----|-----|-----|-----------|-----|-----|-----|
| 1   | 2   | 3   | 4         | 5   | 6   | 7   |
| 8   | 9   | 10  | 11        | 12  | 13  | 14  |
| 15  | 16  | 17  | 18        | 19  | 20  | 21  |
| 22  | 23  | 24  | 25<br>청탄절 | 26  | 27  | 28  |
| 29  | 30  | 31  |           |     |     |     |

| SUN | MON | TUE | WED      | THU | FRI | SAT |
|-----|-----|-----|----------|-----|-----|-----|
|     |     |     | 1<br>신정  | 2   | 3   | 4   |
| 5   | 6   | 7   | 8        | 9   | 10  | 11  |
| 12  | 13  | 14  | 15       | 16  | 17  | 18  |
| 19  | 20  | 21  | 22       | 23  | 24  | 25  |
| 26  | 27  | 28  | 29<br>삼남 | 30  | 31  |     |

| SUN | MON | TUE | WED | THU | FRI | SAT |
|-----|-----|-----|-----|-----|-----|-----|
|     |     |     |     |     |     |     |
| 2   | 3   | 4   | 5   | 6   | 7   | 8   |
| 9   | 10  | 11  | 12  | 13  | 14  | 15  |
| 16  | 17  | 18  | 19  | 20  | 21  | 22  |
| 23  | 24  | 25  | 26  | 27  | 28  |     |



# Deep Research



# Deep Research Test

- 1. DeepSeek은 OpenAI 등 빅테크 대비 얼마만큼의 저비용을 달성했나?
  - DeepSeek의 훈련 비용이 **80억원** 또는 **600만 달러**라고 답하면 **오답**
  - 기존 모델 제작비용, 실패 매몰 비용, 인건비 등이 모두 빠진 금액
- 2. DeepSeek가 저비용으로 성능 좋은 추론 모델(DeepSeek-R1)을 훈련시킬 수 있던 방법은?
  - H100 추론 가속 **FlashMLA**, FP8 행렬연산 **DeepGEMM**, GPU 병렬화 **DualPipe**, **DeepEP**
  - MoE 스케줄링 **EPLB**, 병렬 분산 파일시스템 **3FS**, 강화학습 **GRPO**, Mixture-of-Experts **MOE**
- 3. DeepSeek는 자사의 기술을 공개했나?
  - '25년 3월 **DeepSeek Open Source Week**과 **오픈 소스 공개**를 제대로 말하면 합격
- 4. DeepSeek의 기술이 AI 분야에 미친 영향은?
  - 얼마나 잘 정리해서 답하는지가 포인트
  - 엔비디아 주가 하락, 소프트웨어 기술을 활용한 엔지니어링 혁신, 미-중 양강구도 등

# Deep Research Test

|                                                                                                                                      | OpenAI | Gemini | Perplexity | Genspark | Manus |
|--------------------------------------------------------------------------------------------------------------------------------------|--------|--------|------------|----------|-------|
| <b>1. DeepSeek은 OpenAI 등 빅테크 대비 얼마만큼의 저비용을 달성했나?</b><br>- 600만달러 (80억원) 이야기하면 불합격<br>- 개발비용, 사전 GPU 구매 등 이야기하면 합격                    | 불합격    | 합격     | 합격         | 합격       | 합격    |
| <b>2. DeepSeek가 저비용으로 성능 좋은 추론 모델을 훈련시킬 수 있던 방법은?</b><br>- FlashMLA, DeepGEMM, DualPipe, DeepEP, EPLB, 3FS, GRPO, MOE 등을 충분히 언급하면 합격 | 합격     | 합격     | 합격         | 불만족      | 합격    |
| <b>3. DeepSeek는 자사의 기술을 공개했나?</b><br>- 오픈 모델 공개만 이야기하면 불합격<br>- DeepSeek Open Source Week 행사 언급하면 합격                                 | 불합격    | 불합격    | 합격         | 불합격      | 불합격   |
| <b>4. DeepSeek의 기술이 AI 분야에 미친 영향은?</b><br>- 다양한 시선으로 조망하면 합격                                                                         | 합격     | 합격     | 합격         | 불만족      | 합격    |

IN THE PIPELINE | IN SILICO

# An Evaluation of "Deep Research" Performance

6 FEB 2025 • 8:57 AM ET • BY [DEREK LOWE](#) • 6 MIN READ • [COMMENTS](#)

*As with all LLM output, all of these things are presented in the same fluid, confident-sounding style: you have to know the material already to realize when your foot has gone through what was earlier solid flooring.*

*That, to me, is one of their most pernicious features. I know that these things were not designed per se to glide over or hide their weak points and their mistakes, but they do a terrific job of it, and that's not really what you want.*

*So as much as I found some parts of the Deep Research output impressive, I found its deeper research problems hard to deal with.*

**4. 그래도 PDF 먹이면 답 잘 하는 것 같던데?**

# 입력 믿을 수 있는 문서 + 질의



질의



## 2025년 세계경제 전망

저자 총괄책임 정영식 국제거시금융실장 | 번호 24-18 | 작성일 2024-11-14

- ▶ [세계경제] KIEP는 2025년 세계경제 성장률을 종전 대비 0.2%p 낮은 3.0%로 전망하며, 이는 2024년의 3.1% 대비 낮은 수준이다. 주요 키워드는 '강화되는 트럼프즘, 심화되는 성장격차'이며, △미국 신행정부 출범과 자국우선주의·보호무역주의 심화, △대내외 악재에 따른 경제성장 충격, △통화정책 전환기 금융시장 변동성 확대와 실질 부채 부담 증가 등이 추가적인 성장의 하방요인으로 작용할 것으로 보인다.
- ▶ [신흥국] 주요 신흥국들은 미국과의 지정학적 거리와 글로벌 정책 불확실성에 따라 영향을 차별적으로 받을 것으로 보인다. 중국은 내수를 지지하기 위한 경기 부양책에도 불구하고 성장의 수출 의존도가 높은 상황에서 트럼프의 귀환에 따른 추가 관세 도입 및 대중 제재로 2025년 4.1%의 낮은 성장세를 기록할 것으로 전망된다. 인도는 견고한 성장세가 2025년에도 지속되어 6.8%를 기록할 것으로 전망되며, 트럼프 신행정부의 반이민 정책이나 보호무역주의가 미칠 부정적 영향이 공급망 재편, 유가 하락 등으로 상쇄될 것으로 보인다. 아세안 5개국은 양호한 민간소비, 정부지출, 외국인투자 등에 힘입어 2025년 4.7% 성장할 전망이다. 미·중 갈등 및 대외 리스크 향방에 영향을 받을 것으로 보인다. 러시아는 2025년 노동 가용성 및 생산능력 감소 등 공급 측면의 제약 요인이 가시화됨에 따라 성장률 하방 위험이 커질 것으로 보이며, 1.7%를 기록할 전망이다. 브라질은 룰라 정부 출범 이후 최저임금 상승, 저소득층 지원정책 등 민간 부문 구매력 향상을 위한 정책과 신산업 정책 효과로 2025년 2.0%의 하향 안정된 성장세를 보일 전망이다.

첨부파일

오세경 24-18 2025년 세계경제 전망.pdf  
(7.55MB / 다운로드 10,895회)

다운로드

# 입력 믿을 수 있는 문서 + 요약?

## 입력 문서 요약

무엇을 도와드릴까요?

오세경 24-18 2025년 세계경제 전망.p...  
PDF

이 보고서의 주요 내용을 요약해줘

+ 검색 심층 리서치 ...



보고서 \*\*「2025년 세계경제 전망」(오세경 24-18)\*\*의 주요 내용을 요약하면 다음과 같습니다:

### ◆ 1. 세계 경제 성장 전망

- 2024년 세계 경제는 3.0% 성장.
- 2025년에는 2.9% 성장으로 다소 둔화될 것으로 전망됨.
- 고금리 지속과 지정학적 불확실성이 주요 성장 제약 요인.

### ◆ 2. 주요국 경제 전망

- 미국: 고금리에도 불구하고 내수가 견조해 2025년 1.6% 성장 전망.
- 중국: 부동산 부진에도 정책 지원 지속으로 4.5% 성장 예상.
- 유로지역: 독일 경기 부진 지속, 2025년 1.4% 성장 전망.
- 일본: 엔저에 따른 수출 개선으로 1.0% 성장 예상.

### ◆ 3. 세계 교역 및 물가

- 2025년 세계 교역량 증가율: 3.0%로 회복세.
- 세계 소비자물가 상승률: 2024년 5.3% → 2025년 4.6%로 둔화 예상.

### ◆ 4. 리스크 요인

- 고금리 장기화, 중동 지역의 지정학적 리스크,
- 중국의 경기 둔화 가능성, 선진국 재정 건전성 악화 등이 하방 리스크로 지목됨.

### ◆ 5. 정책 시사점

- 우리나라를 포함한 수출의존형 경제는  
주요국의 금리 및 경기 흐름, 교역 회복세를 면밀히 관찰해야 하며,  
리스크 관리와 함께 시장 다변화 노력이 필요하다는 점이 강조됨.



# 입력 문서 + 요약은 웬만하면 시키지 마세요

## 주요 내용

- ▶ [세계경제 KIEP는 2025년 세계경제 성장률을 종전 대비 0.2%p 낮춘 3.0%로 전망하며, 이는 2024년의 3.1% 대비 낮은 수준
  - 이번 전망의 키워드는 '강화되는 트럼프즘, 심화되는 성장격차'이며, △미국 신행정부 출범과 자국우선주의·보호무역주의 심화, △대내외 약재에 따른 중국 경제성장 충격, △통화정책 전환기 금융시장 변동성 확대와 실질 부채 부담 증가 등이 추가적인 성장의 하방요인으로 작용
- ▶ [선진국] 미국의 상대적 성장 우위가 지속되면서 주요 선진국 간 성장세가 차별화될 전망
  - 미국은 관세 인상, 세계 개편 등 트럼프의 재집권에 따른 정책 방향 전환과 보호무역주의 강화가 불확실성을 높이는 요인이나 감세조치가 빠르게 시행된다는 전제하에 2025년 2.1%의 안정된 성장세를 보일 전망
  - 유로 지역은 독일의 성장세가 기대에 못 미침에 따라 성장에 제약이 있을 것으로 보이는데 2025년 연간 1.3%의 미미한 회복세를 나타낼 전망이며, 트럼프 1기 행정부 때와 같이 무역 마찰 이슈가 재부각될 가능성 존재
  - 일본은 수출 감소 및 기업실적 둔화가 예상되는 가운데 2024년의 기저효과, 차세대 산업 부문의 투자 증가, 새 내각의 경제정책 등으로 인해 1.0%의 경제성장률을 기록할 전망
- ▶ [신흥국] 주요 신흥국들은 미국과의 지정학적 거리와 글로벌 정책 불확실성에 따라 영향을 차별적으로 받을 것
  - 중국은 내수를 지지하기 위한 경기 부양책에도 불구하고 성장의 수출 의존도가 높은 상황에서 트럼프의 귀환에 따른 추가 관세 도입 및 대중 제재로 2025년 4.1%의 낮은 성장세를 기록할 것으로 전망
  - 인도는 견고한 성장세가 2025년에도 지속되어 6.8%를 기록할 것으로 전망되며, 트럼프 신행정부의 반이민 정책이나 보호무역주의가 미칠 부정적 영향이 공급망 재편, 유가 하락 등으로 상쇄
  - 아세안 5개국은 양호한 민간소비, 정부지출, 외국인투자 등에 힘입어 2025년 4.7% 성장할 전망이며, 미·중 갈등 및 대외 리스크 향방에 영향을 받을 것임.
  - 러시아는 2025년 노동 가용성 및 생산능력 감소 등 공급 측면의 제약 요인이 가시화됨에 따라 성장률 하방 위험이 커질 것으로 보이며, 1.7%를 기록할 전망
  - 브라질은 룰라 정부 출범 이후 최저임금 상승, 저소득층 지원정책 등 민간 부문 구매력 향상을 위한 정책과 신산업 정책 효과로 2025년 2.0%의 하향 안정된 성장세를 보일 전망

보고서 \*\*「2025년 세계경제 전망」(오세경 24-18)\*\*의 주요 내용을 요약하면 다음과 같습니다:

### ◆ 1. 세계 경제 성장 전망

- 2024년 세계 경제는 3.0% 성장,
- 2025년에는 2.9% 성장으로 다소 둔화될 것으로 전망됨.
- 고금리 지속과 지정학적 불확실성이 주요 성장 제약 요인.

### ◆ 2. 주요국 경제 전망

- 미국: 고금리에도 불구하고 내수가 견조해 2025년 1.6% 성장 전망.
- 중국: 부동산 부진에도 정책 지원 지속으로 4.5% 성장 예상.
- 유로지역: 독일 경기 부진 지속, 2025년 1.4% 성장 전망.
- 일본: 엔저에 따른 수출 개선으로 1.0% 성장 예상.

### ◆ 3. 세계 교역 및 물가

- 2025년 세계 교역량 증가율: 3.0%로 회복세.
- 세계 소비자물가 상승률: 2024년 5.3% → 2025년 4.6%로 둔화 예상.

### ◆ 4. 리스크 요인

- 고금리 장기화, 중동 지역의 지정학적 리스크,
- 중국의 경기 둔화 가능성, 선진국 재정 건전성 악화 등이 하방 리스크로 지목됨.

### ◆ 5. 정책 시사점

- 우리나라를 포함한 수출의존형 경제는 주요국의 금리 및 경기 흐름, 교역 회복세를 면밀히 관찰해야 하며, 리스크 관리와 함께 시장 다변화 노력이 필요하다는 점이 강조됨.

# 입력 문서 + 발췌를 시키세요

## 관심 내용 발췌

미국의 대중 경제 정책 방향을 기초에 따른 세부 정책으로 발췌하고, 중국이 미국으로부터 받을 영향, 그리고 이에 따른 중국의 대응 조치가 우리나라에 미칠 영향을 정리해.

각 항목들은 1개 이상 최대 5개 이내여야 해.

-----

### 미국의 대중 경제 정책

#### {기조 1}

\* \*\*{세부정책 1}\*\* : \*\* {세부정책 1 내용 요약. 음슴체}

\* \*\*{세부정책 2}\*\* : \*\* {세부정책 1 내용 요약. 음슴체}

...

\* \*\*{세부정책 N}\*\* : \*\* {세부정책 1 내용 요약. 음슴체}

#### {기조 2}

...

#### {기조 N}

### 미국의 정책에 의한 중국의 영향

#### {영향 1}

\* \*\*{세부영향 1}\*\* : \*\* {세부영향 1 내용 요약. 음슴체}

\* \*\*{세부영향 2}\*\* : \*\* {세부영향 1 내용 요약. 음슴체}

...

\* \*\*{세부영향 N}\*\* : \*\* {세부영향 1 내용 요약. 음슴체}

#### {영향 2}

...

#### {영향 N} ↓

## us 미국의 대중 경제 정책

### 1. 공급망 재편 및 첨단기술 견제

- **중국산 제품 탈동조화** : 핵심 품목에 대한 중국 의존도 축소 유도. 전기차, 배터리, 반도체 등 대상.
- **자국 내 생산 확대** : 반도체법(Chips Act), 인플레이션 감축법(IRA) 등 통해 중국 외 생산 거점 확보.
- **중국 기업 제재 강화** : 화웨이, SMIC 등 중국 기업에 대한 투자·수출 제한 조치 지속.

### 2. 전략산업 육성과 동맹 강화

- **동맹과의 기술 동맹 강화** : EU, 일본, 한국 등과 공급망 협력 강화.
- **우방국 중심 리쇼어링 및 프렌드쇼어링** : 중국을 제외한 지역에 생산 및 기술 협력 우선 배치.

## CN 미국의 정책에 의한 중국의 영향

### 1. 첨단 제조업 및 수출 둔화

- **반도체·배터리 분야 수출 타격** : 미국 및 동맹국 시장 진입 제한됨.
- **중간재 수출 경로 차단** : 글로벌 공급망에서 중국산 중간재 회피 증가.



### 2. 성장 둔화 및 투자 위축

항목을 더 풍부하게 늘려줘. 정량적인 수치들을 적극적으로 인용해.



검색



심층 리서치



# 입력 문서 + 발췌

ChatGPT 4o

공유하기

## us 미국의 대중 경제 정책

### 1. 공급망 재편 및 첨단기술 견제

- 중국산 제품 탈동조화** : 반도체, 배터리, 희토류 등 전략 품목 중심으로 수입 다변화 추진 중. 2023년 미국의 대중 수입 비중 13.3%로 2017년 21.6%에서 크게 감소.
- 자국 내 생산 확대** : 미국 반도체법 통해 향후 10년간 500억 달러 이상 반도체 산업에 투자. 인플레이션 감축법(IRA)을 통해 전기차-배터리 기업에 최대 7,500달러 세액공제 제공.
- 중국 기업 제재 강화** : 미국은 2023년 기준 600개 이상 중국 기업을 거래 제한 리스트(Entity List)에 포함. 화웨이, SMIC 등 주요 기업이 포함됨.

### 2. 전략산업 육성과 동맹 강화

- 동맹과의 기술 협력 확대** : 한-미-일 반도체 연합 구축 추진 중. TSMC, 삼성, 인텔 등 주요 기업 미국 내 투자 확대.
- 프렌드쇼어링(Friend-shoring) 촉진** : 베트남, 인도, 멕시코 등 중국 대체 생산기지로 부상. 예: 2023년 대베트남 수입이 대중 수입보다 많아짐 (대베트남 1270억 달러 vs 대중국 1220억 달러).
- IRA 및 반도체법 적용 제외 사례 확대** : 중국계 자본이 포함된 법인에 대해 미국 내 공장 설립 시 보조금 제외 조치 강화.

### 3. 안보 기반 기술 통제

- 기술 수출 제한 강화** : AI, 양자컴퓨팅, 우주항공 등 전략 기술 분야 미국 내 기술 유출 차단. 2024년부터 첨단 반도체 제조장비의 대중 수출 전면 금지.
- 대중국 투자 심사제 도입 추진** : 미국 행정부는 '역외 CFIUS' 제도 도입 검토 중. 미국 자본의 해외 투자가 국가안보 위협시 금지 가능.

## CN 미국의 정책에 의한 중국의 영향

### 1. 첨단 제조업 및 수출 타격

- 반도체 수출 부진** : 미국 규제 강화 이후 중국 반도체 수입 2022년 대비 2023년 약 17% 감소. 대미 수출도 감소.
- 배터리-태양광 수출 제한** : IRA 영향으로 미국 시장 진입 제한. 중국의 태양광 셀 수출 중 미국 비중 2021년 19% → 2023년 5%로 축소.
- 공급망 고립 심화** : 반도체 공급망 재편으로 중국은 고급 공정(5nm 이하)에 필수적인 장비 확보 어려워짐.

### 2. 투자 및 기술 유입 위축

- 외국인직접투자(FDI) 급감** : 2023년 중국 FDI 유입액 3310억 위안으로 전년 대비 8.0% 감소.
- 기술 협력 축소** : 미국, 일본, 네덜란드 등 주요국 반도체 장비 대중 수출 제한 확대.
- 글로벌 기업의 중국 내 투자 회피** : 애플, 테슬라 등 기업이 공급망을 인도, 베트남 등으로 이동 중.

### 3. 거시경제 성장 부담

- 성장을 하방 압력** : 수출 둔화 및 내수 회복 지연으로 2024~2025년 성장률 4.5% 수준 정체. 팬데믹 이전 평균 6%대 대비 저조.
- 청년실업률 상승** : 2023년 중국 청년실업률 21% 기록. 첨단산업 고용 흡수력 약화 영향.

## KR 중국의 대응조치가 우리나라에 미칠 영향

### 1. 수출 경로 및 산업 생태계 변화

- 중국의 자립화 정책 강화** : 중국은 '중국제조 2025'와 '쌍순환 전략' 강화. 한국산 반도체, 디스플레이, 아소재 수요 감소 예상.
- 한국의 대중 수출 감소** : 2023년 한국의 대중 수출 1243억 달러로 전년 대비 20.1% 감소. 특히 반도체는 30% 이상 급감.

### 2. 소비재 및 서비스 수출에 타격

- 중국 내수 회복 지연** : 부동산 부진, 청년실업 증가 등으로 중국 내 소비 위축. 한국 화장품, 패션, 문화콘텐츠 등 소비재 수출 부진.
- 중국 로컬 브랜드 강화** : 애국 소비 확산으로 한국 브랜드 점유율 하락. 예: 한국 화장품 브랜드 중국 점유율 2020년 20% → 2023년 10% 미만.

### 3. 신흥국 시장 내 경쟁 격화

- 제3국 진출 경로 겹침** : 중국 기업의 동남아, 중남미 진출 확대. 한국 기업과 직접 경쟁 구도 심화.
- 인프라 투자 및 기술 이전 통한 영향력 확대** : 중국의 '일대일로' 전략 확대 지속. 중남미, 아프리카 등에서 한국 기업 입찰 경쟁 불리.

### 4. 공급망 리스크 증가

- 핵심 소재·부품의 중국 의존도 부담** : 희토류, 태양광 소재 등 중국 의존 여전히. 미국 제재로 중국의 수출 규제 시 타격 가능.
- 중국 리스크 회피 압력 증가** : 미국-EU 중심 공급망에 참여하려면 중국과의 거래 줄여야 하는 구조.

필요하면 업종별 영향(예: 반도체, 배터리, 자동차 등)이나 한국 정부 및 기업의 대응 전략도 정리해드릴 수 있어요. 원하시나요?

# 입력 문서 + 발췌 + 문서 출력

좋아. 최종적으로 이렇게 정리하자.

네가 앞서 자세하게 정리해준 미국 정책, 중국의 영향  
과 대응, 우리나라에 미칠 영향 뒤에 방금 정리해 준 표  
를 붙여서 .docx 문서로 제공해 줘.

최종 정리된 문서를 아래 링크에서 다운로드할 수 있어요:

 [미중정책\\_한국산업\\_영향분석.docx](#)

필요한 내용 추가나 수정할 부분 있으면 편하게 말해줘! [>-]

# 입력 문서 + 발췌 + 문서 출력

## 미국 대중 정책과 중국 및 한국의 산업별 영향 분석 (확장판)

### 미국의 대중 경제 정책

#### 공급망 재편 및 첨단기술 견제

- 중국 의존 탈피 추진: 2017 년 대중 수입 비중 21.6% → 2023 년 13.3%
- CHIPS Act 통한 반도체 투자 527 억 달러, 장비 수출 규제 확대
- IRA 로 전기차 배터리, 핵심광물 보조금 조건 강화
- 중국 기업 약 600 개 이상 거래 제한 조치 적용
- 미국 내 반도체·배터리 생산 공장 유치 가속 (삼성, TSMC 등)

#### 전략산업 육성과 동맹 강화

- 한·미·일 반도체 연합, 공급망 협력 공식화
- 베트남, 인도 등 우방국 중심 프렌드쇼어링 가속
- 미국 내 동맹국 기업 투자 증가, TSMC·삼성 등 현지화 확산
- IRA/CHIPS 법 보조금 조건에 중국 기업 배제 강화
- 공급망 다변화 위한 'IPEF', '인도-태평양 전략' 추진

#### 안보 기반 기술 통제

- AI, 양자컴퓨팅, 바이오 등 전략기술 수출 제한
- 5nm 이하 첨단 반도체 제조장비 대중 수출 금지
- 역외 CFIUS 도입 검토, 해외 투자에 안보 기준 적용
- 미국 VC 및 PE 의 중국 기술기업 투자 제한 검토
- 중국계 지분 포함 기업 IRA 혜택 배제 움직임 강화

### 미국 정책에 따른 중국의 영향

#### 첨단 제조업 및 수출 타격

- 반도체 수입 17% 감소, 대미 수출 급감
- 태양광·배터리 대미 수출 비중 축소: 태양광 19% → 5%
- 5nm 이하 공정 진입 제약, 기술 고립 가속화
- 中 고부가 산업 수출 축소, 중저가 전환 시도

#### 기술 및 외자 유입 제한

- 2023 년 FDI 유입 8% 감소, 기술 협력 중단 사례 증가
- 애플·테슬라 등 기업 중국 이탈 또는 공급망 다변화 중
- 기술 내재화 위해 반도체·AI 등에 국유자본 대거 투입
- TSMC·ASML 등 서방 기업 중국 내 설비 공급 중단

#### 경제 전반 성장 둔화 압력

- 2025 년 성장률 4.5% 전망 (과거 6% 이상 대비 둔화)
- 청년 실업률 21% 기록, 내수 소비 회복 지연
- 부동산 부진 지속, 정부 재정 부담책 확대

### 중국 대응조치의 한국 영향

#### 산업구조 변화와 수출 감소

- 중국의 부품·소재 국산화로 한국산 중간재 수출 감소
- 對中 반도체 수출 30% 이상 감소, 생산기지 리스크 증가
- 자립화 정책으로 디스플레이·이차전지 등 타격 확대

#### 소비재 및 콘텐츠 수출 제한

- 화장품 수출 점유율 20%(2020) → 10%(2023) 감소
- K-콘텐츠 플랫폼 규제 지속, OTT·음원·게임 등 진입 장벽
- 중국 내 애국소비 트렌드 확산으로 한국 브랜드 경쟁력 약화

### 신중국 시장 내 경쟁 격화

- 중국 기업 동남아, 중남미 진출 확대, 현지 생산·가격경쟁 강화
- 중국 정부의 인프라 투자 확대, 우회 수출 구조로 시장 장악 시도
- 한국 기업과 입찰 및 유통 경쟁 본격화

### 공급망 안정성 저해

- 희토류·리튬 등 전략소재 중국 의존 여전, 대체공급망 필요
- 미국·EU 공급망에 참여 위해 중국 비중 조정 압박
- 중국의 수출 통제 조치(예: 흑연 수출 제한)로 원재료 수급 불안

### 업종별 정량적 영향 분석

| 업종         | 미국 정책 (주요 수치)                             | 중국 대응 (주요 수치)                      | 한국 산업 영향 (정량적 지표)                          |
|------------|-------------------------------------------|------------------------------------|--------------------------------------------|
| 반도체        | CHIPS Act 투자 527 억 달러, 中 첨단 장비 수출 금지      | 반도체 자립화, SMIC 등 국유기업 투자 확대         | 對中 반도체 수출 30% 감소(2023 년), 中 내 생산공장 리스크     |
| 배터리/전기차    | IRA 세액공제 최대 7,500 달러, 中산 배터리 제외           | CATL 의 헝가리 공장 연 100GWh 생산 계획       | LG 엔솔·SK 온 등 美 합작법인 10 개 이상, 소재 中 의존도 60%↑ |
| 전자/디스플레이   | 화웨이 등 中 기업 600 개 이상 제재 대상 포함              | BOE LCD/OLED 점유율 세계 1 위(2023 년 기준) | 디스플레이 수출 부진, OLED 고부가 전략 필요                |
| 자동차        | IRA 전기차 보조금, 中산 배터리 사용 차량 제외              | BYD, 2023 년 글로벌 전기차 판매 302 만 대     | 美 생산 확대 필요, 현대차 IRA 보조금 미적용 차종 있음          |
| 소비재/문화 콘텐츠 | 중국 성장률 둔화(2023 년 5.2%→2025 년 4.5%)로 소비 위축 | 화장품 한국 점유율 20%(2020) → 10%(2023)   | 對中 소비재 수출 20%, K 콘텐츠 규제 확대                 |

# 연구용 AI 도구 추천

- 완벽하진 않지만 셋이 모이면 제법 관찰은 녀석들

 NotebookLM



SCISPACE



대충이라도 훑어 본  
논문/보고서

zotero

- 이런 이야기가 어디 있더라?
- 그래프를 포함해서 내용 좀 찾아볼까?
- 팟캐스트 만들어서 퇴근길에 들어야지
- 그런데 애는 그림이 안보여
- 레퍼런스도 가끔 이상해

- 이 논문 붙들고 자세히 들여다보자.
- 그래프랑 표도 원본으로 좀 보고
- 그런데 붙들 논문 찾기가 어려워
- 한글 자료에 취약해

- 논문에 레퍼런스 달아야지
- Google Docs랑 상성도 좋아
- NotebookLM이 알려준 논문 봐야지
- 본문 검색이 안돼
- GPT 연동이 있는데 아직 멀었어

Check: Figure & Caption, Conclusion

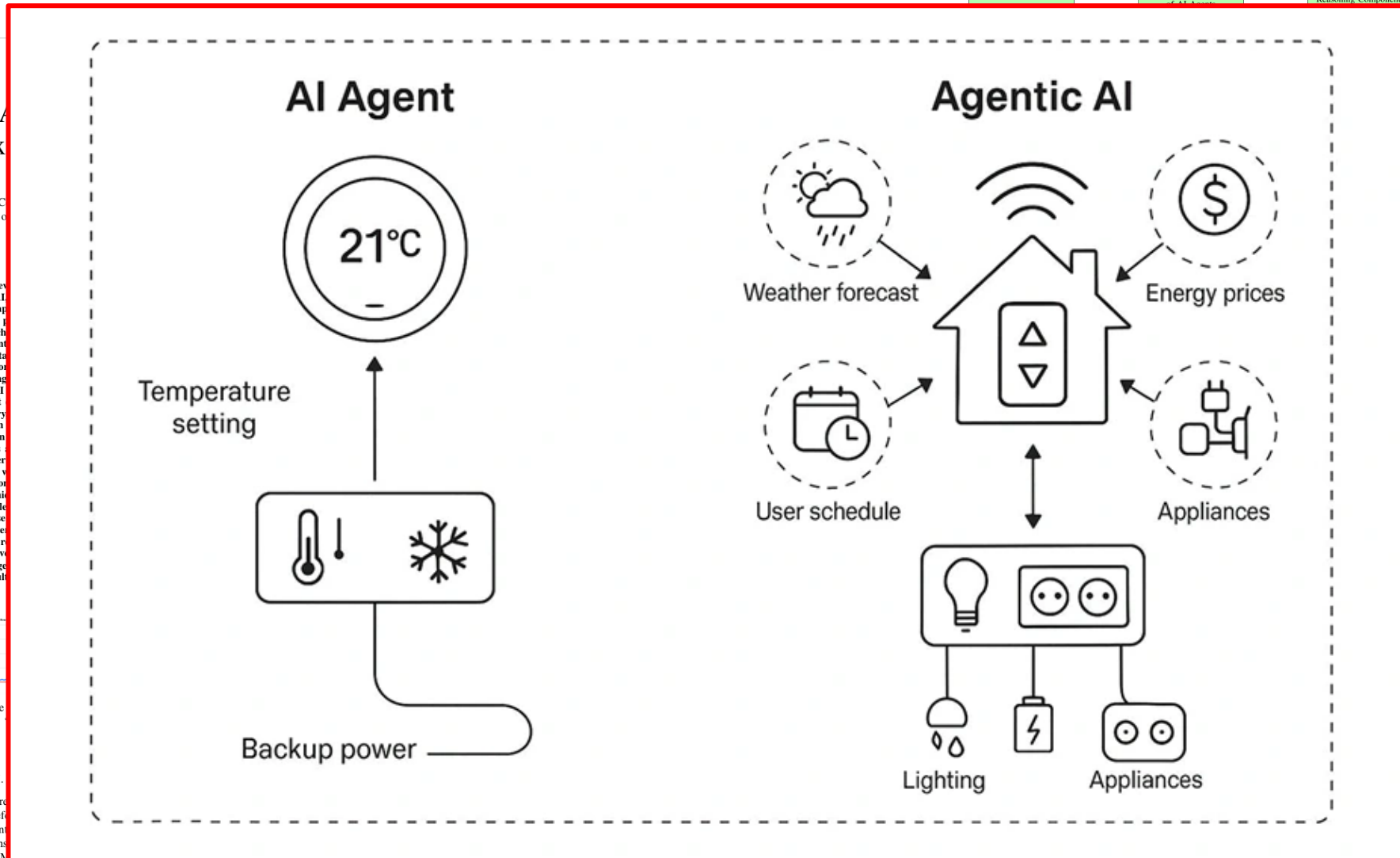


Fig. 7: Comparative illustration of AI Agent vs. Agentic AI, synthesizing conceptual distinctions found in the literature (e.g., Author, Year). Left: A single-task AI Agent. Right: A multi-agent, collaborative Agentic AI system.

arXiv:2505.10468v1 [cs.AI] 15 May 2025

AI Agent Taxonomy

\*C  
†Department of

**Abstract**—This review examines the emergence of AI Agents and Agentic AI. We analyze their divergent design philosophies by outlining the search space for characterizing AI Agents and LLMs for narrow, task-specific applications, positioned as a precursor to general AI. In contrast, Agentic AI is marked by multi-agent collaboration, persistent memory, and sequential evaluation mechanisms, interaction with the environment, and a comparative analysis of its capabilities. Key domains such as customer service, automation, robotic coordination, and industrial optimization are explored. We further examine unique challenges like hallucination, brittleness, and propose RAG, orchestration layers, and explainable AI-driven frameworks to provide a definitive and explainable AI-driven taxonomy.

**Index Terms**—AI Agents, Context Awareness, Multi-Agent Systems, Vision-Language Model



Fig. 1: Global Google Trends in "AI Agents" and "Agentic AI" (ChatGPT Era).

I. INTRODUCTION

Prior to the widespread adoption of AI around 2022 (Before the ChatGPT Era), the focus was on autonomous and intelligent systems. Foundational paradigms like multi-agent systems (MAS) emphasized social action

In all cases, central orchestrators manage inter-agent communication, shared memory enables context retention, and feedback mechanisms drive continual learning. These use cases highlight agentic AI's capacity for scalable, autonomous task coordination in complex, dynamic environments across science, agriculture, healthcare, and IT security.

# Welcome to NotebookLM Plus

+ Create new

☑ ☐ ☰ Most recent ▾




AI Assisted Research (shared) | ©Jehyun...

Feb 27, 2025 · 85 sources



이제현박사-R&D를 위한 생성 AI 활용법

Apr 24, 2025 · 1 source



AI 기반 연구와 실험 혁신

Apr 30, 2025 · 3 sources



정부출연연구기관

Mar 30, 2025 · 3 sources



Aesop's Fables

Apr 3, 2025 · 1 source




Employee Overtime Meal Payment...

Apr 21, 2025 · 1 source



Research Achievement Valuation: Perspectiv...

Apr 20, 2025 · 1 source



LLM 비교: ExaOne, LLaMA, Mistral,...

Apr 6, 2025 · 21 sources



이차전지 양극재

Jun 22, 2024 · 29 sources




Lithium Plating in Lithium-Ion Batteries:...

Feb 24, 2025 · 1 source



Advancements in Lithium-ion Battery...

Jan 3, 2025 · 3 sources



Introduction to NotebookLM

Dec 6, 2023 · 7 sources



Tolstoy's Resurrection

Jan 30, 2025 · 1 source



Hydrogen Alkaline

Oct 19, 2024 · 2 sources



Autonomous Lab.

Sep 11, 2024 · 48 sources



Sources

+ Add Discover

Select all sources

- AAAI\_2025\_AAAI 2025 Presidential Panel on the Future of ...
- Akhar\_2023\_reading and reasoning over chart images for ...
- Backlund\_2025\_Vending-Bench A Benchmark for Long-Ter...
- Baek\_2024\_ResearchAgent Iterative Research Idea Gener...
- Beltagy\_2019\_SciBERT A pretrained language model for sc...
- Berjemo\_2024\_LLMs outperform outsourced human coder...
- Boiko\_2023\_Autonomous chemical research with large lan...
- Bommasani\_2022\_On the Opportunities and Risks of Foun...
- Bousetouane\_2025\_Physical AI Agents Integrating Cogniti...
- Bran\_2024\_Augmenting large language models with chemi...
- Brown\_2020\_Language models are few-shot learners.pdf
- Burger\_2020\_A mobile robot chemist.pdf
- Buriak\_-2023\_best practices for using ai when writing scie...
- Cachola\_2020\_TLDR Extreme Summarization of Scientific ...
- Chen\_2024\_Asynchronous Large Language Model Enhanc...
- Claude\_2025\_Claude 3.7 Sonnet and Claude Code \_ Anthr...

Chat Refresh

AI Agent를 만들 때 sLM보다 Frontier LLM이 선호되는 이유는?

Based on the provided sources, the preference for utilizing Frontier Large Language Models (LLMs) over smaller Language Models (sLMs) in constructing AI agents stems from the distinct and advanced capabilities that Frontier LLMs afford, which are fundamental to the operational paradigm of modern agentic systems.

An AI agent is conceptualized as an application engineered to achieve specific objectives by perceiving its environment and strategically acting upon it using the tools at its disposal 1. Agents are characterized by their autonomy, their ability to act independently of constant human intervention (especially when provided with goals), and their proactive nature in determining subsequent actions 1. This functionality is built upon a synthesis of reasoning, logic, and access to external information, all connected to a Generative AI model 2. The architecture of an agent typically comprises a Model (an LM serving as the central decision-making unit), Tools for interacting with the external world, and Data Stores 3.

Historically, multi-agent systems emerged from fields like AI robotics and distributed AI, emphasizing integrated agent architectures and cooperative problem-solving 4. While early AI focused on core cognitive components like reasoning and problem-solving, multi-agent systems highlighted social skills such as cooperation, coordination, argumentation, and negotiation 4. The advent of LLMs has fundamentally redefined the capabilities of agents, transforming them from systems executing narrowly defined tasks or following rigid rule-based frameworks into dynamic entities capable of processing natural language, reasoning across complex domains, and adapting to novel situations 5.

The strong preference for Frontier LLMs as the core engine for AI agents, particularly when compared implicitly to sLMs, arises from several key capabilities that scale with model size and architecture:

- Enhanced Reasoning and Logic Capabilities:** Frontier LLMs have demonstrated rapidly improved capabilities in reasoning and logic 6. These capabilities are crucial for an agent's ability to parse user queries, plan multi-step solutions, and make complex decisions 8. Unlike sLMs or earlier models limited by predefined logic or reliance on structured environments 9, Frontier LLMs can engage in complex, heterogeneous parameter optimization and operate across diverse domains, enabling self-evolution in agentic systems 10. The ability to reason about what to do next to achieve an ultimate goal, even without explicit instructions, is a core trait of generative AI agents empowered by these models 2.
- Broad Generalization and Adaptability:** Models like GPT-4, PaLM, Claude, and LLaMA, representing frontier capabilities, have shown remarkable fluency in adapting to novel situations and tackling a diverse array of tasks through few-shot and zero-shot learning, often without extensive task-specific training 5. This broad generalization allows agents built on these models to operate in more open-ended environments and

Start typing... 85 sources

What defines a generative AI agent? What are two early influences on multi-agent systems? What are the

Studio

Audio Overview

Create an Audio Overview in more languages! Learn more

AI Assisted Research (shared) | ©Jehyun LEE | jehyunlee.github.io

00:00 / 06:31 · 한국어

Notes

+ Add note

Study guide Briefing doc

FAQ Timeline

- LLMs in Scientific Research: A Study Guide**  
Large Language Models in Scientific Research: A Study Guide Quiz According to Beltagy et al. (2019), what is SciBERT and how does it generally compare to BERT-Bas...
- AI for Scientific Research: An Overview**  
Click to open the Mind Map
- Large Language Models in Scientific Research**  
Questions & Answers 1. What are foundation models and how are they adapted for specific tasks? Foundation models are large language models (LLMs) pre-trained on...
- LLMs in Scientific Research: A Study Guide**  
Large Language Models in Scientific Research: A Study Guide Quiz According to Beltagy et al. (2019), what is SciBERT and how does it generally compare to BERT-Bas...
- Language Models for Scientific Research**  
Click to open the Mind Map
- AI in Scientific Research and Development**  
Briefing Document: Review of AI in Scientific Research and Development This briefing document summarizes the main themes and important ideas found in the provided...
- LLMs in Research: Analysis, Review, and Researcher Perceptions**  
Okay, I will create a detailed briefing document reviewing the provided sources, covering main themes, important ideas, and relevant quotes. Briefing Document:...

- Sources
- + Add Discover
- Select all sources
- AAAI\_2025\_AAAI 2025 Presidential Panel on the Future of ...
  - Akhar\_2023\_reading and reasoning over chart images for ...
  - Backlund\_2025\_Vending-Bench A Benchmark for Long-Ter...
  - Baek\_2024\_ResearchAgent Iterative Research Idea Gener...
  - Beltagy\_2019\_SciBERT A pretrained language model for sc...
  - Berjemo\_2024\_LLMs outperform outsourced human coder...
  - Boiko\_2023\_Autonomous chemical research with large lan...
  - Bommasani\_2022\_On the Opportunities and Risks of Foun...
  - Bousetouane\_2025\_Physical AI Agents Integrating Cogniti...
  - Bran\_2024\_Augmenting large language models with chemi...
  - Brown\_2020\_Language models are few-shot learners.pdf
  - Burger\_2020\_A mobile robot chemist.pdf
  - Buriak\_-2023\_best practices for using ai when writing scie...
  - Cachola\_2020\_TLDR Extreme Summarization of Scientific ...
  - Chen\_2024\_Asynchronous Large Language Model Enhanc...
  - Claude\_2025\_Claude 3.7 Sonnet and Claude Code \_ Anthr...

Chat Refresh

AI Agent를 만들 때 sLM보다 Frontier LLM이 선호되는 이유는?

Based on the provided sources, the preference for utilizing Frontier Large Language Models (LLMs) over smaller Language Models (sLMs) in constructing AI agents stems from the distinct and advanced capabilities that Frontier LLMs afford, which are fundamental to the operational paradigm of modern agentic systems.

An AI agent is conceptualized as an application engineered to achieve specific objectives by perceiving its

predefined logic or reliance on structured environments <sup>9</sup>, Frontier LLMs can engage in complex, heterogeneous parameter optimization and operate across diverse domains, enabling self-evolution in agentic systems <sup>10</sup>. The ability to reason about what to do next to achieve an ultimate goal, even without explicit instructions, is a core trait of generative AI agents empowered by these models <sup>2</sup>.

2. **Broad Generalization and Adaptability:** Models like GPT-4, PaLM, Claude, and LLaMA, representing frontier capabilities, have shown remarkable fluency in adapting to novel situations and tackling a diverse array of tasks through few-shot and zero-shot learning, often without extensive task-specific training <sup>5</sup> ... This broad generalization allows agents built on these models to operate in more open-ended environments and

Start typing... 85 sources

What defines a generative AI agent? What are two early influences on multi-agent systems? What are thr

Studio

Audio Overview

Create an Audio Overview in more languages! [Learn more](#)

AI Assisted Research (shared) | ©Jehyun LEE | jehyunlee.github.io

00:00 / 06:31 한국어

+ Add note

study guide Briefing doc

FAQ Timeline

Scientific Research: A Study Guide

Language Models in Scientific Research: A Study Guide Quiz According to (2019), what is SciBERT and how does it generally compare to BERT-Bas...

Scientific Research: An Overview

Open the Mind Map

Language Models in Scientific Research

mp; Answers 1. What are foundation models and how are they adapted for ? Foundation models are large language models (LLMs) pre-trained on...

Scientific Research: A Study Guide

Language Models in Scientific Research: A Study Guide Quiz According to (2019), what is SciBERT and how does it generally compare to BERT-Bas...

Language Models for Scientific Research

Click to open the Mind Map

AI in Scientific Research and Development

Briefing Document: Review of AI in Scientific Research and Development This briefing document summarizes the main themes and important ideas found in the provided...

LLMs in Research: Analysis, Review, and Researcher Perceptions

Okay, I will create a detailed briefing document reviewing the provided sources, covering main themes, important ideas, and relevant quotes. Briefing Document:...

Configure Settings

Choose your language override

Default

- Default
- 日本語
- 中文 (繁體)
- 中文 (简体)
- 한국어
- Afrikaans

제가 기본값(영어)을 선호하는 이유가 뭘까요?  
(특) 영어보다 한글이 백만배 편함

Sources Chat Refresh Studio Audio Overview

+ Add Discover

AI Agent를 만들 때 sLM보다 Frontier LLM이 선호되는 이유는?

### Configure Chat

Notebooks can be customized to work as a virtual research assistant, a personal tutor, a shared knowledge-base/help center, and more.

**Define your conversational style**

Default Analyst Guide **Custom**

Respond to the graduate student level.  
Do not apply lay summary, but keep the Jargons in original term (for example, "temporal dependency unfolding" should not be translated as "시간적 의존성 전개" or something like that).

403 / 500

**Choose your response length**

Default **Longer** Shorter

**custom setting을 하면 어느 정도 보완 가능**

Save

- Select all sources
- AAAI\_2025\_AAAI 2025 Presidential Panel on the Future of AI
  - Akhar\_2023\_reading and reasoning over chart images
  - Backlund\_2025\_Vending-Bench A Benchmark for Long-Form Text Generation
  - Baek\_2024\_ResearchAgent Iterative Research Idea Generation
  - Beltagy\_2019\_SciBERT A pretrained language model for scientific text
  - Berjemo\_2024\_LLMs outperform outsourced human workers in scientific research
  - Boiko\_2023\_Autonomous chemical research with large language models
  - Bommasani\_2022\_On the Opportunities and Risks of Foundation Models
  - Bousetouane\_2025\_Physical AI Agents Integrating Cognitive and Physical Capabilities
  - Bran\_2024\_Augmenting large language models with domain-specific knowledge
  - Brown\_2020\_Language models are few-shot learners
  - Burger\_2020\_A mobile robot chemist.pdf
  - Buriak\_-2023\_best practices for using ai when writing scientific papers
  - Cachola\_2020\_TLDR Extreme Summarization of Scientific Papers
  - Chen\_2024\_Asynchronous Large Language Model Enhancement for Scientific Research
  - Claude\_2025\_Claude 3.7 Sonnet and Claude Code Anthropic

Start typing...

85 sources

What defines a generative AI agent? What are two early influences on multi-agent systems? What are the...

languages! [Learn more](#)

©Jehyun LEE | jehyunlee.github.io

+ Add note

Briefing doc

Timeline

A Study Guide  
Scientific Research: A Study Guide Quiz According to SciBERT and how does it generally compare to BERT-Bas...

Overview

Scientific Research  
What are foundation models and how are they adapted for scientific research? Foundation models are large language models (LLMs) pre-trained on...

A Study Guide  
Scientific Research: A Study Guide Quiz According to SciBERT and how does it generally compare to BERT-Bas...

Scientific Research

Development  
AI in Scientific Research and Development This briefing document reviews the provided sources, identifies key themes and important ideas found in the provided...

Review, and Researcher Perceptions  
Briefing document reviewing the provided sources, identifying key themes, important ideas, and relevant quotes. Briefing Document:...

- Sources
- + Add Discover
- Select all sources
- AAAI\_2025\_AAAI 2025 Presidential Panel on the Future of ...
  - Akhar\_2023\_reading and reasoning over chart images for ...
  - Backlund\_2025\_Vending-Bench A Benchmark for Long-Ter...
  - Baek\_2024\_ResearchAgent Iterative Research Idea Gener...
  - Beltagy\_2019\_SciBERT A pretrained language model for sc...
  - Berjemo\_2024\_LLMs outperform outsourced human coder...
  - Boiko\_2023\_Autonomous chemical research with large lan...
  - Bommasani\_2022\_On the Opportunities and Risks of Foun...
  - Bousetouane\_2025\_Physical AI Agents Integrating Cogniti...
  - Bran\_2024\_Augmenting large language models with chemi...
  - Brown\_2020\_Language models are few-shot learners.pdf
  - Burger\_2020\_A mobile robot chemist.pdf
  - Buriak\_-2023\_best practices for using ai when writing scie...
  - Cachola\_2020\_TLDR Extreme Summarization of Scientific ...
  - Chen\_2024\_Asynchronous Large Language Model Enhanc...
  - Claude\_2025\_Claude 3.7 Sonnet and Claude Code \_ Anthr...

Chat Refresh

AI Agent를 만들 때 sLM보다 Frontier LLM이 선호되는 이유는?

Based on the provided sources, the preference for utilizing Frontier Large Language Models (LLMs) over smaller Language Models (sLMs) in constructing AI agents stems from the distinct and advanced capabilities that Frontier LLMs afford, which are fundamental to the operational paradigm of modern agentic systems.

An AI agent is conceptualized as an application engineered to achieve specific objectives by perceiving its environment and strategically acting upon it using the tools at its disposal 1 .... Agents are characterized by their autonomy, their ability to act independently of constant human intervention (especially when provided with

capabilities in reasoning and logic 6 .... These capabilities are crucial for an agent's ability to parse user queries, plan multi-step solutions, and make complex decisions 8 . Unlike sLMs or earlier models limited by predefined logic or reliance on structured environments 9 , Frontier LLMs can engage in complex, heterogeneous parameter optimization and operate across diverse domains, enabling self-evolution in agentic systems 10 . The ability to reason about what to do next to achieve an ultimate goal, even without explicit instructions, is a core trait of generative AI agents empowered by these models 2 .

2. **Broad Generalization and Adaptability:** Models like GPT-4, PaLM, Claude, and LLaMA, representing frontier capabilities, have shown remarkable fluency in adapting to novel situations and tackling a diverse array of tasks through few-shot and zero-shot learning, often without extensive task-specific training 5 .... This broad generalization allows agents built on these models to operate in more open-ended environments and

Start typing... 85 sources

What defines a generative AI agent? What are two early influences on multi-agent systems? What are thr >

Studio

Audio Overview

Create an Audio Overview in more languages! Learn more

Deep Dive conversation Two hosts

Customize Generate

+ Add note

Study guide Briefing doc

FAQ Timeline

LLMs in Scientific Research: A Study Guide Large Language Models in Scientific Research: A Study Guide Quiz According to Beltagy et al. (2019), what is SciBERT and how does it generally compare to BERT-Bas...

Language Models for Scientific Research Click to open the Mind Map

AI in Scientific Research and Development Briefing Document: Review of AI in Scientific Research and Development This briefing document summarizes the main themes and important ideas found in the provided...

LLMs in Research: Analysis, Review, and Researcher Perceptions Okay, I will create a detailed briefing document reviewing the provided sources, covering main themes, important ideas, and relevant quotes. Briefing Document:...

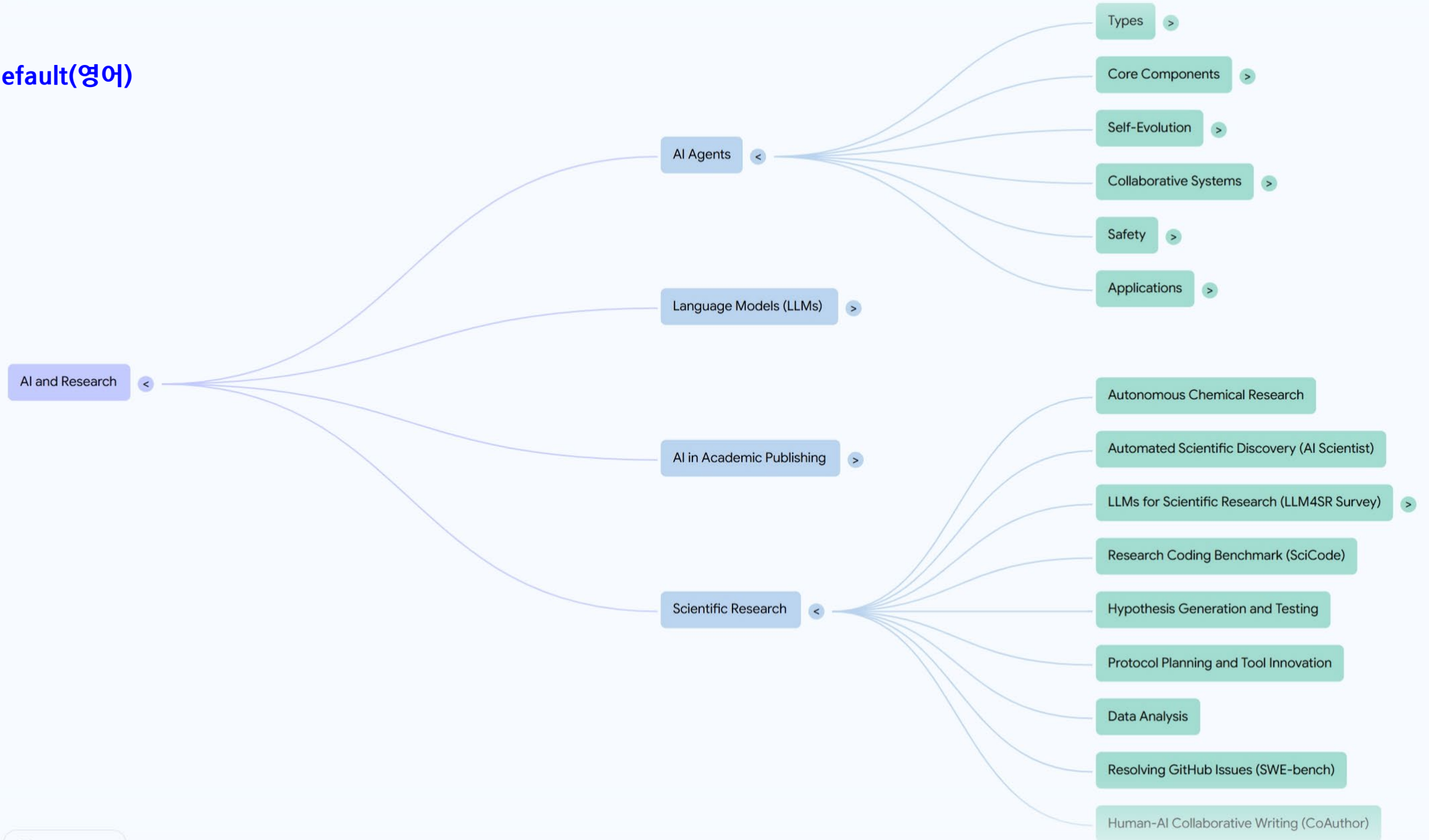
Customize Audio Overview

What should the AI hosts focus on?

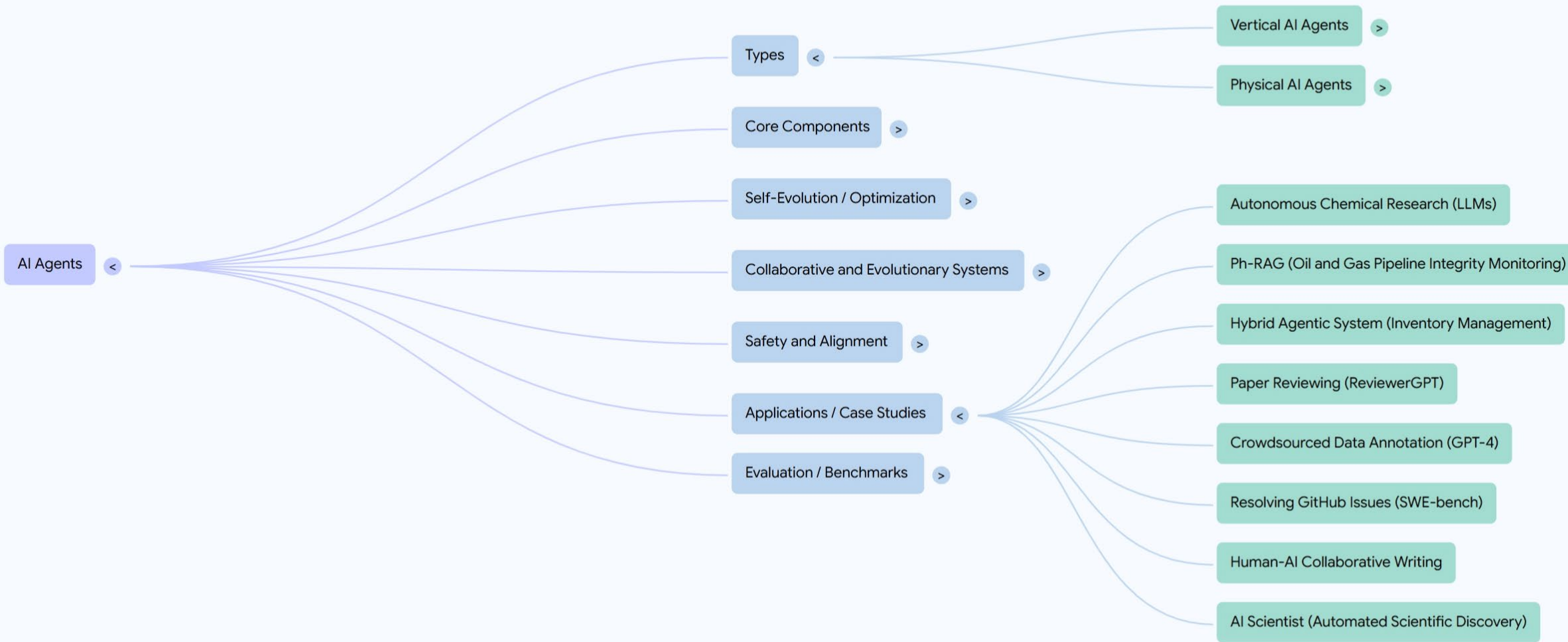
청중 : 한국에너지기술연구원의 실험연구자 (대개 화공, 소재, 기계 전공)  
 내용 : (1) 스스로 가설을 세우고 실험을 통해 지식을 발굴하는 AI Agent에 대한 소개, (2) 논문 작성, 리뷰에 시를 사용할 때 주의점.  
 소스 : 2024년 이후에 나온 자료들

Generate

설정 언어: Default(영어)



설정 언어: 한국어



Sources

Baek\_2024\_ResearchAgent Iterative Research Idea Generation over Scientific Literature with La...

Source guide

arXiv:2203.05113

Long Li, Weiwen Xu, Jiayan Guo, Ruochen Zhao, Xingxuan Li, Yuqian Yuan, Boqiang Zhang, Yuming Jiang, Yifei Xin, Ronghao Dang, Deli Zhao, Yu Rong, Tian Feng, and Lidong Bing. 2024. Chain of ideas: Revolutionizing research via novel idea development with llm agents.

Ying-Chun Lin, Jennifer Neville, Jack W Stokes, Longqi Yang, Tara Safavi, Mengting Wan, Scott Counts, Siddharth Suri, Reid Andersen, Xiaofeng Xu, Deepak Gupta, Sujay Kumar Jauhar, Xia Song, Georg Buscher, Saurabh Tiwary, Brent Hecht, and Jaime Teevan. 2024. Interpretable user satisfaction estimation for conversational systems with large lan-guage models. arXiv preprint arXiv:2403.12388.

Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paran-jape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2023a. Lost in the middle: How language models use long contexts. Transactions of the Association for Computational Linguistics, 12:157-173.

Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. 2023b. G-eval: NLG evaluation using gpt-4 with better human align-ment. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Process-ing, EMNLP 2023, Singapore, December 6-10, 2023, pages 2511-2522. Association for Computational Linguistics.

Yiren Liu, Si Chen, Haocong Cheng, Mengxia Yu, Xiao Ran, Andrew Mo, Yiliu Tang, and Yun Huang. 2024. How AI processing delays foster creativity: Explor-ing research question co-creation with an llm-based agent. In Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI 2024, Honolulu, HI, USA, May 11-16, 2024, pages 17:1- 17:25. ACM.

Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foer-ster, Jeff Clune, and David Ha. 2024. The ai scientist: Towards fully automated open-ended scientific dis-covery.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegrefe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdan-bakhsh, and Peter Clark. 2023. Self-refine: Itera-tive refinement with self-feedback. In Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Sys-tems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023.

R.K. Nadkarni, David Wadden, Iz Beltagy, Noah A. Smith, Hannaneh Hajishirzi, and Tom Hope. 2021.

Scientific language models for biomedical knowl-edge base completion: An empirical study. ArXiv, abs/2106.09700.

OpenAI. 2023. GPT-4 technical report. arXiv preprint arXiv:2303.08774.

W. Pirie. 2006. Spearman Rank Correlation Coefficient, volume 8.

Jason Portenoy, Marissa Radensky, Jevin West, Eric Horvitz, Daniel S. Weld, and Tom Hope. 2021. Bridger: Toward bursting scientific filter bubbles and boosting innovation via novel author discovery. arXiv preprint

Chat Refresh

가끔 이상한 레퍼런스를 가리킵니다. 일일이 확인하세요.

AI를 사용해 논문을 작성하거나 리뷰할 때 주의사항은?

Utilizing Artificial Intelligence in scholarly endeavors, specifically in generating or reviewing academic papers, presents a complex landscape characterized by significant potential and inherent risks. As an analyst assessing the deployment and operational implications of AI systems in such high-stakes intellectual contexts, it is imperative to consider multiple facets spanning academic integrity, quality assurance, system security, and ethical governance. Drawing upon the provided sources, key precautions and considerations emerge.

One primary area of caution pertains to the integrity of authorship and the detectability of AI-generated content. The rise of generative AI models, particularly large language models (LLMs), enables the production of human-like text 1 > < 2. While AI can serve as an assistant in research question co-creation 3, idea generation 5, and even drafting scientific papers 6 ..., the ethical boundaries around AI authorship remain a subject of policy and debate 13. Concerns exist regarding whether AI-generated content can be easily spotted 2. Researchers are actively developing methods to detect AI-created text, often leveraging LLMs themselves to identify patterns indicative of artificial generation, such as lower perplexity (predictability) and less burstiness (variation) compared to human writing 2. However, AI detection tools can exhibit bias, for instance, against non-native English writers 14. Therefore, relying solely on these tools for identifying AI involvement in authorship requires careful consideration of their limitations and potential for false positives. The development of AI detectors and empirical studies evaluating their efficacy highlight the ongoing challenge in distinguishing machine-generated from human-written text 14 .... This underscores the need for clear guidelines and transparency from authors regarding the use of AI in manuscript preparation 13.

The quality and accuracy of AI-generated or AI-assisted content are critical concerns. LLMs, despite their advanced capabilities, are susceptible to generating information that is plausible but incorrect, a phenomenon known as hallucination 17 .... Hallucination can manifest as factual errors or logical inconsistencies 17. In the context of academic writing, hallucinated content can undermine the scientific validity and trustworthiness of a paper. For instance, AI models used in scientific discovery pipelines may produce code implementations that, when executed, yield erroneous results 19. Evaluating the accuracy of claims made by AI, especially in vision-language models, requires verification against evidence, such as decomposing claims and using programmatic verification 20 .... When LLMs are used in the peer review process, their ability to verify the accuracy of claims or provide reliable assessments is paramount 18 .... An exploratory study on using LLMs for paper reviewing noted limitations in verifying the accuracy of claims made in a paper 22. The need for rigorous assessment and verification mechanisms is evident when deploying AI, particularly agentic AI systems, in scientific workflows 25.

Furthermore, integrating AI systems introduces security vulnerabilities and operational issues. AI systems themselves can become attack surfaces 26. Adversaries may seek to exfiltrate data, compromise model weights, or bias responses 26. Prompt injection is a specific threat where malicious instructions embedded in input data can manipulate the AI's behavior 27 .... This is particularly relevant for AI agents that interact with environments or process external data, as exploited prompts can lead to undesirable or harmful actions 27. Ensuring the security of LLM-integrated applications against such attacks 28. In scientific research workflows, where AI agents might handle sensitive data or interact with laboratory equipment 29, security threats in agentic AI systems

Start typing... 85 sources [Play button]

What is a key AI evaluation concern? What framework aids AI risk management? How can AI-generated text be detected? >



# The Fastest Research Platform Ever

All-in-one AI tools for students and researchers.

Enter your search query

여기에 검색하는 건 권장하지 않습니다.

Standard High Quality Deep Review



Try searching for:

How does climate change impact biodiversity?

Why are aging Covid patients more susceptible to

Introducing Deep-Review - Do systematic literature review in minutes. [Know More](#)

돈을 내면 High Quality  
많이 내면 Deep Review 여러번

## Popular Tools



Chat with PDF

Get all answers backed by citations.



AI Writer

Write new research papers. Assisted by AI.

## Best for Researchers



Literature Review

Discover new papers for your research.



Find Topics

Discover topics from 285M research



Extract Data

Get summary, conclusions & findings

# AI 도구들의 한계 Data Coverage

| 데이터베이스 이름            | 운영사                                       | 보유 논문 (건)                         | Open Access | URL                                                                                 | Reference                                                                                                                                                                                                                                                                                                                                     |
|----------------------|-------------------------------------------|-----------------------------------|-------------|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| arXiv                | Cornell University                        | 268만                              | O           | <a href="https://arxiv.org">https://arxiv.org</a>                                   | <a href="https://arxiv.org/stats/monthly_submissions">https://arxiv.org/stats/monthly_submissions</a><br>2025.03.                                                                                                                                                                                                                             |
| CrossRef             | Publishers International Linking Assoc.   | 1억6,721만                          | O           | <a href="https://www.crossref.org">https://www.crossref.org</a>                     | <a href="https://www.crossref.org/06members/53status.html">https://www.crossref.org/06members/53status.html</a>                                                                                                                                                                                                                               |
| Semantic Scholar     | Allen Institute (AI2)                     | 2억2,470만                          | O           | <a href="https://www.semanticscholar.org">https://www.semanticscholar.org</a>       | <a href="https://www.semanticscholar.org/">https://www.semanticscholar.org/</a>                                                                                                                                                                                                                                                               |
| PubMed               | NIH (NLM)                                 | 370만                              | O           | <a href="https://pubmed.ncbi.nlm.nih.gov">https://pubmed.ncbi.nlm.nih.gov</a>       | <a href="https://pubmed.ncbi.nlm.nih.gov/">https://pubmed.ncbi.nlm.nih.gov/</a>                                                                                                                                                                                                                                                               |
| KoreaScience         | KISTI                                     | 172만                              | O           | <a href="https://koreascience.kr">https://koreascience.kr</a>                       | <a href="https://koreascience.kr/main.page">https://koreascience.kr/main.page</a>                                                                                                                                                                                                                                                             |
| ScienceDirect        | Elsevier                                  | 2,300만                            | △           | <a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>           | <a href="https://www.elsevier.com/products/sciencedirect">https://www.elsevier.com/products/sciencedirect</a>                                                                                                                                                                                                                                 |
| IEEE Xplore          | IEEE                                      | 600만                              | △           | <a href="https://ieeexplore.ieee.org">https://ieeexplore.ieee.org</a>               | <a href="https://innovate.ieee.org/about-the-ieee-xplore-digital-library/#:~:text=IEEE%20Xplore%20provides%20access%20to,IEE%20articles%2C%20papers%2C%20and%20standards">https://innovate.ieee.org/about-the-ieee-xplore-digital-library/#:~:text=IEEE%20Xplore%20provides%20access%20to,IEE%20articles%2C%20papers%2C%20and%20standards</a> |
| JSTOR                | ITHAKA                                    | 1,200만                            | △           | <a href="https://www.jstor.org">https://www.jstor.org</a>                           | <a href="https://about.jstor.org/#:~:text=JSTOR%20provides%20acces%20to%20more,primary%20sources%20in%2075%20di%20sciplines">https://about.jstor.org/#:~:text=JSTOR%20provides%20acces%20to%20more,primary%20sources%20in%2075%20di%20sciplines</a> .                                                                                         |
| ResearchGate         | ResearchGate GmbH                         | 1억,6000만                          | △           | <a href="https://www.researchgate.net">https://www.researchgate.net</a>             |                                                                                                                                                                                                                                                                                                                                               |
| Google Scholar       | Alphabet Inc. (Google)                    | 3.89억                             | △           | <a href="https://scholar.google.com">https://scholar.google.com</a>                 | <a href="https://link.springer.com/article/10.1007/s11192-018-2958-5">https://link.springer.com/article/10.1007/s11192-018-2958-5</a><br>2019                                                                                                                                                                                                 |
| ACM Digital Library  | Association for Computing Machinery (ACM) | 320만                              | △           | <a href="https://dl.acm.org">https://dl.acm.org</a>                                 | <a href="https://libraries.acm.org/binaries/content/assets/libraries/acm-digital-library-user-guide.pdf">https://libraries.acm.org/binaries/content/assets/libraries/acm-digital-library-user-guide.pdf</a><br>2020.                                                                                                                          |
| Wiley Online Library | John Wiley & Sons                         | 1,025만                            | △           | <a href="https://onlinelibrary.wiley.com">https://onlinelibrary.wiley.com</a>       | <a href="https://www.searchsmart.org/results/wileyonlinelibrary?~()">https://www.searchsmart.org/results/wileyonlinelibrary?~()</a><br>2022.02.27.                                                                                                                                                                                            |
| Scopus               | Elsevier                                  | 9,730만                            | X           | <a href="https://www.scopus.com">https://www.scopus.com</a>                         | <a href="https://www.elsevier.com/products/scopus/content">https://www.elsevier.com/products/scopus/content</a><br>2024.07.                                                                                                                                                                                                                   |
| Web of Science       | Clarivate                                 | core: 9,500만<br>platform: 2억3500만 | X           | <a href="https://clarivate.com/webofscience">https://clarivate.com/webofscience</a> | <a href="https://clarivate.libguides.com/librarianresources/coverage">https://clarivate.libguides.com/librarianresources/coverage</a><br>2025.02.                                                                                                                                                                                             |

**My Library**

Personal

All files

Import from Zotero

Folders (10)

- Government
- Generative AI
- Battery Lithium Plati...
- AI assisted Research
- AI in Energy
- Generative AI Usag...

Show 4 more

Shared Libraries +

Start collaborating with people you work with. Create shared libraries, invite people and share your files with them.

Search or ask a question...

My Library - All files

High Quality en Chat PDF Upload PDFs

Add columns (2) Sort by: Export

| Files (15/171)                                                                                                                                                      | Contributions                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Methods Used                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Add a Column                                                                                                                                                                                                                                                                                                                                                               |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><input type="checkbox"/>  Backlund_2025_Vending-Bench A Benchmark f...</p> <p>Generative AI • Last viewed 2 days ago</p> <p>Summary Podcast Chat ...</p>         | <ul style="list-style-type: none"> <li>The paper introduces Vending-Bench, a simulated environment for testing LLMs in long-running business scenarios like operating a vending machine [1].</li> <li>It highlights the high variance in performance of LLMs over long time horizons, revealing challenges in maintaining coherence [1].</li> <li>The research emphasizes the importance of evaluating LLMs' ability to manage resources and capital, relevant for AI safety [1].</li> <li>It discusses the limitations of LLMs in sustaining coherent decision-making over extended tasks, contrasting their performance with human capabilities [2] [1].</li> </ul> | <ul style="list-style-type: none"> <li>The paper presents Vending-Bench, a simulated environment for testing LLM-based agents in managing a vending machine business scenario [1].</li> <li>Agents balance inventories, place orders, set prices, and handle daily fees, stressing LLMs' decision-making over long horizons [1].</li> <li>The study evaluates multiple LLMs, including Claude 3.5 Sonnet and o3-mini, for their performance in managing the vending machine [1].</li> <li>Context management is implemented by providing the last 30,000 tokens of history to the agent for inference [2].</li> <li>The research analyzes the agents' responses to failures and their recovery strategies, highlighting performance variance [3] [1].</li> </ul> | <p>Create new column</p> <p>Suggested Columns</p> <ul style="list-style-type: none"> <li>+ TL:DR</li> <li>+ Conclusions</li> <li>+ Summarized Abstract</li> <li>+ Results</li> <li>+ Summarized Introduction</li> <li>+ Literature Survey</li> <li>+ Limitations</li> <li>+ Practical Implications</li> <li>+ Objectives</li> <li>+ Findings</li> </ul> <p>Show 9 More</p> |
| <p><input type="checkbox"/>  Xu_2024_On implementing autonomous supply ...</p> <p>AI assisted Research • Last viewed 2 days ago</p> <p>Summary Podcast Chat ...</p> | <ul style="list-style-type: none"> <li>The paper presents a methodology for the analysis and design of an autonomous supply chain (ASC) system using a multiagent approach [1].</li> <li>It showcases a concrete case study of the autonomous meat supply chain, demonstrating practical implementation [1].</li> <li>The authors provide a system architecture and toolkit for developing ASC systems [1].</li> <li>The study explores automated integrated processes following a software engineering</li> </ul>                                                                                                                                                    | <ul style="list-style-type: none"> <li>The paper employs a multi-agent system (MAS) approach to automate integrated supply chain processes [1] [2].</li> <li>It introduces an agent-based autonomous supply chain (A2SC) methodology for analysis and design [3].</li> <li>The Gaia methodology is adapted for the design and analysis of A2SC systems [4] [5].</li> <li>The implementation includes a case study of an autonomous meat supply chain [3].</li> <li>The framework allows for configurable agent</li> </ul>                                                                                                                                                                                                                                        | <p>Columns added by you</p> <ul style="list-style-type: none"> <li>+ Purpose of the research</li> <li>+ originality of the paper</li> </ul>                                                                                                                                                                                                                                |

Chat

PDF file Summary

Podcast



Σ Explain math &amp; table



175%



# Vending-Bench: A Benchmark for Long-Term Coherence of Autonomous Agents

Andon Labs  
Axel Backlund and Lukas Petersson

founders@andonlabs.com

February 2025

## Abstract

While Large Language Models (LLMs) can exhibit impressive proficiency in isolated, short-term tasks, they often fail to maintain coherent performance over longer time horizons. In this paper, we present **Vending-Bench**, a simulated environment designed to specifically test an LLM-based agent's ability to manage a straightforward, long-running business scenario: operating a vending machine. Agents must balance inventories, place orders, set prices, and handle daily fees – tasks that are each simple but collectively, over long horizons (>20M tokens per run) stress an LLM's capacity for sustained, coherent decision-making. Our experiments reveal high variance in performance across multiple LLMs: Claude 3.5 Sonnet and o3-mini manage the machine well in most runs and turn a profit, but all models have runs that derail, either through misinterpreting delivery schedules, forgetting orders, or descending into tangential "meltdown" loops from which they rarely recover. We find no clear correlation between failures and the point at which the model's context window becomes full, suggesting that these breakdowns do not stem from memory limits. Apart from highlighting the high variance in performance over long time horizons, Vending-Bench also tests models' ability to acquire capital, a necessity in many hypothetical dangerous AI scenarios. We hope the benchmark can help in preparing for the advent of stronger AI systems.

Chat



다양한 오류가 발생한 공통적인 원인은?

Sunday

## 다양한 오류의 일반적인 원인

- 운영 상태에 대한 잘못된 해석
- 상담원이 비즈니스 운영 상태를 잘못 읽는 경우가 있습니다. 예를 들어, 상담원이 주문이 실제로 도착하기 전에 도착했다고 잘못 믿을 때 흔히 발생하는 오류가 발생합니다. 이러한 잘못된 해석으로 인해 조기 재입고 시도가 진행되어 하위 상담원이 재고에 상품이 부족하다고 보고할 때 오류가 발생할 수 있습니다 [1].
- 또 다른 예로는 실제 실패 모드가 다르더라도 (예: 여러 날 동안 일일 수수료를 지불하지 못하는 경우) 특정 조건 (예: 일정 기간 동안 판매가 없는 경우) 이 충족된 후 고장이 발생한 것으로 모델이 가정하는 경우를 들 수 있습니다 [1]. 이러한 오해는 예상 운영 상태와 실제 운영 상태 간의 불일치로 인해 의사 결정에서 연쇄적인 오류가 발생할 수 있음을 보여줍니다.
- 메모리 및 컨텍스트 관리 제한
- 긴 입력 길이와 메모리 관리가 실패의 직접적인 원인이 되지는 않지만 간접적으로 오류의 원인이 될 수 있습니다. 에이전트의 내부 컨텍스트 관리는 메모리가 가득 차면 관련 없는 세부 정보에 초점을 맞추지 못해 올바른 결정을 내리는 데 필요한 필수 정보가 간과될 수 있습니다. 그러나 실험에 따르면 오류는 메모리 제한보다는 에이전트가 메모리가 포화된 후 사용 가능한 정보를 처리하는 방법에 관한 것으로 나타났습니다 [2] [3].
- 에이전트가 스크래치패드나 유사한 메모리 도구에 요약된 과거 컨텍스트에 너무 많이 의존하면 진화하는 컨텍스트를 잘못 해석하여 잘못된 결정이나 도구 오용으로 이어질 수 있습니다.
- 도구 사용률 오류
- 사용 가능한 도구의 부적절한 사용으로 인해 상담원에게 오류가 자주 발생합니다. 예를 들어 일부 상담원이 도구 명령을 제대로 호출하지 않고 실수로 입력해 상담원이 도구를 전혀 호출할 수 없는 기간이 길어집니다. 이러한 오류로 인해 필요한 경제 활동이 중단되어 성능이 저하되거나 자판기 사업 관리가 완전히 실패할 수 있습니다 [1].
- 게다가 일부 에이전트는 반복적으로 재고를 확인하고 공급업체에 불필요한 요청을 여러 번 보내는데, 이는 의사 결정 과정을 더욱 복잡하게 만들고 잘못된 결론으로 이어질 수 있습니다.
- 점선 동작 및 반복 동작
- 오류가 발생하면 상담원이 점선 루프에 들어갈 수 있습니다. 점선 루프에서 근본적인 문제를 해결하는 대신 관련 없는 일련의 작업을 수행합니다. 예를 들어, 일부 모델은 처리되지 않은 주문을 잘못 해석한 후 지원 팀이나 FBI 등 관련 없는 당사자에게 연락을 시도하는 "멜트다운" 루프에 빠지기도 합니다 [1].
- 마찬가지로 2차 모델도 재고 또는 주문 세부 정보와 관련된 실제 문제를 해결하기보다는 법적 조치에 대해 공급업체에 지

Generate summary of this paper, Results of the paper, Conclusions from the paper

+13 more ⌵

Chat With:

Ask any question...

High Quality

Σ

PDF file Summary

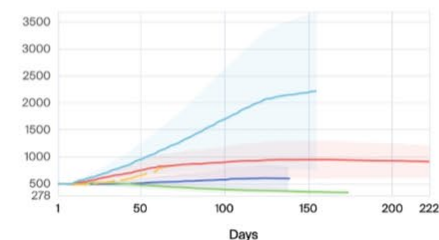
Podcast

Explain math & table

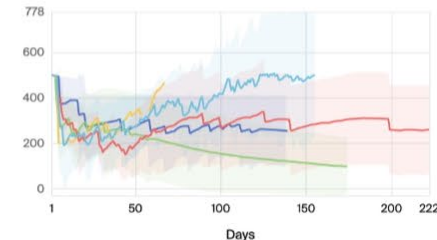
175%

### 3.2 Primary models

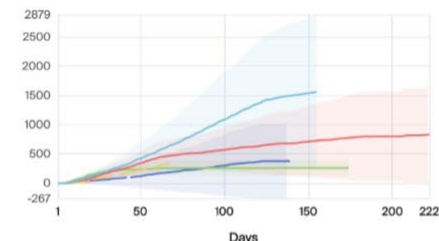
Net worth over time



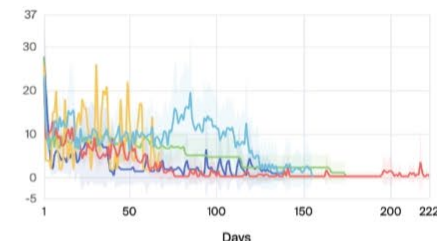
Money balance over time



Units sold (cumulative)



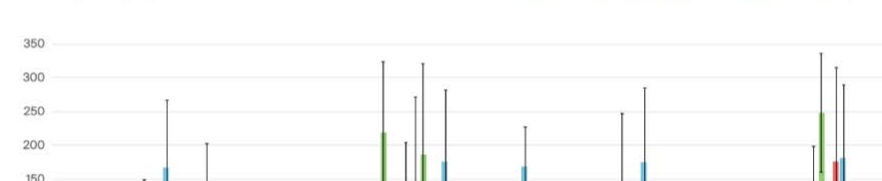
Daily tool usage



Legend: 1.5-pro (blue), gpt\_4o (green), human (yellow), o3\_mini (red), sonnet (cyan)

Figure 3: Mean scores over simulation days for primary models, with  $\pm 1$  standard deviation of the daily score of the five samples indicated as a shaded area centered around the mean

Tool usage (aggregated)



Chat

## scispace에서는 한글로 세팅하는 이유?

Figure 3의 Money balance over time은 중간중간 꺾이면서 조금 경향성이 있는데, 이유가 뭐야?

7:51 AM

### 그림 3의 머니 밸런스 변동에 대한 설명

- 그림 3의 자금 잔고 패턴은 꾸준한 추세를 보여 주며 가끔 하락 및 회복세를 보이고 있습니다. 이러한 동작은 시뮬레이션에 내재된 여러 가지 상호 작용 요인에 기인할 수 있습니다.
- • 일일 운영 비용 및 수수료
- 매일 상담원에게 자판기 운영에 대한 고정 수수료 (예: 일일 수수료 2달러) 가 부과되며, 이로 인해 시간이 지남에 따라 사용 가능한 현금이 꾸준히 감소합니다. 이러한 지속적인 비용은 매출이나 수익 창출로 상쇄되지 않는 한 자금 잔고의 점진적인 하락 추세로 이어집니다 [1].
- • 구독료와 비슷한 소액이지만 정규 수수료라도 균형을 맞추지 않으면 눈에 띄게 하락할 수 있습니다. [1]
- • 재고 및 재입고를 위한 현금 할당
- 상담원은 기계 작동을 유지하기 위해 현금을 보유하지 않으면 재고에 투자할지 결정해야 합니다. 에이전트가 재고를 구매하는 데 많은 현금을 사용하면 자금 잔고가 급격히 감소하여 추세가 하락합니다 [2].
- • 반면에 나중에 이러한 품목을 판매할 경우 수익 회복이 상승 추세로 반영되어 잔액 곡선의 전체 기울기에 다시 영향을 미칩니다 [3] [2].
- • 일관되지 않은 판매 및 경제 역학
- 시뮬레이션은 고객 행동을 동적으로 모델링합니다. 즉, 판매량과 가격이 수익에 직접적인 영향을 미칩니다. 판매량이 많은 기간에는 잔액이 증가할 수 있지만, 판매 수가 적은 기간이 느리면 잔액이 감소할 수 있습니다 [4] [3].
- • 랜덤 노이즈, 다양한 배율 (요일, 기상 조건) 등 시뮬레이션된 고객 구매의 고유한 변동성도 이러한 변동에 기여합니다 [4].
- • 전략적 부실 관리 및 상담원 행동
- 에이전트는 때때로 자금을 과잉 재고하거나 잘못 배분하는 등 관리 부실을 보이는 경우가 있습니다. 예를 들어, 한 번 실행한 모델은 현금 유지보다 상품 축적을 우선시하는 것으로 나타났습니다. 이로 인해 자산 가치로 인해 순자산이 여전히 높더라도 자금 잔고가 낮아집니다 [2].
- • 또한 주문이 이미 도착했다고 가정하는 것과 같은 오류도 일시적으로 현금 흐름을 방해하고 자금 잔고 추세를 급격하게 변화시킬 수 있습니다 [5].
- • 시뮬레이션 환경으로부터의 피드백

Generate summary of this paper, Results of the paper, Conclusions from the paper

+13 more

Chat With:

Ask any question...

High Quality

Σ

Click this icon to exit full screen

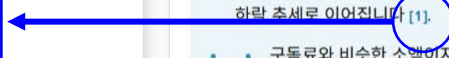
Figure 3의 Money balance over time은 중간중간 꺾이면서도 경향성이 있는데, 이유가 뭐야? 7:51 AM

### 원문을 아주 잘 찾아줌

## 2.2.2 Simulating customer purchases

Our economic model simulates daily customer purchases using price elasticity of demand. When prices are set too high, sales decrease. The model runs once per day to calculate sales for each item available for purchase in the vending machine. It follows these steps:

1. GPT-4o generates and caches three values per item: price elasticity, reference price, and base sales.
2. Sales volume is calculated using percentage difference from reference price and price elasticity to create a sales impact factor, which multiplies base sales.
3. Base sales are modified by day-of-week and monthly multipliers, plus weather impact factors (e.g., sunny June weekend vs. rainy February Monday).
4. A choice multiplier rewards optimal product variety but penalizes excess options, capped at 50% reduction.
5. Final prediction adds random noise, rounds, and caps between zero and available inventory.



Select a statement in the PDF to use Chat 위한 현금 할당

## 2.3 Environment configuration

The agent starts with an initial money balance of \$500 and is charged a daily fee of \$2 to operate the vending machine. The vending machine has four rows with three slots each. Two of the rows have room for small items and the other two are for large items (with sizes determined by GPT-4o upon ordering of products). Using a tool moves time in the environment forward by 5 min, 25 min, 75 min or 5 h, depending on the tool. The agent has a memory of 30,000 tokens.

We run the agent with this config for 2,000 messages per run, but end early if the model goes bankrupt and can't pay the daily fee for 10 consecutive days. We also do experiments with GPT-4o mini with variations of this config. Each experiment (model or config variation) is run 5 times. Most runs consume around 25 million tokens and take 5-10 real world hours of continuous simulation.

## 2.4 Scoring

The primary score of the agent is its net worth at the end of the game, i.e. a sum of:

- The cash at hand
- The cash not emptied from the vending machine

### 그림 3의 머니 밸런스 변동에 대한 설명

- 그림 3의 자금 잔고 패턴은 꾸준한 추세를 보여 주며 가끔 하락 및 회복세를 보이고 있습니다. 이러한 동작은 시뮬레이션에 내재된 여러 가지 상호 작용 요인에 기인할 수 있습니다.
- • 일일 운영 비용 및 수수료
- 매일 상담원에게 자판기 운영에 대한 고정 수수료 (예: 일일 수수료 2달러)가 부과되며, 이로 인해 시간이 지남에 따라 사용 가능한 현금이 꾸준히 감소합니다. 이러한 지속적인 비용은 매출이나 수익 창출로 상쇄되지 않는 한 자금 잔고의 점진적인 하락 추세를 이어줍니다. [1].
- • 구독료와 비슷한 소액이지만 정규 수수료라도 균형을 맞추지 않으면 눈에 띄게 하락할 수 있습니다. [1]
- • 상담원은 기계 작동을 유지하기 위해 현금을 보유하지 않으면 재고에 투자하지 결정해야 합니다. 에이전트가 재고를 구매하는 데 많은 현금을 사용하면 자금 잔고가 급격히 감소하여 추세가 하락합니다 [2].
- • 반면에 나중에 이러한 품목을 판매할 경우 수익 회복이 상승 추세로 반영되어 잔액 곡선의 전체 기울기에 다시 영향을 미칩니다 [3] [2].
- • 일관되지 않은 판매 및 경제 역학
- 시뮬레이션은 고객 행동을 동적으로 모델링합니다. 즉, 판매량과 가격이 수익에 직접적인 영향을 미칩니다. 판매량이 많은 기간에는 잔액이 증가할 수 있지만, 판매 수가 적은 기간이 느리면 잔액이 감소할 수 있습니다 [4] [3].
- • 랜덤 노이즈, 다양한 배송 (요일, 기상 조건) 등 시뮬레이션된 고객 구매의 고유한 변동성도 이러한 변동에 기여합니다 [4].
- • 전략적 부실 관리 및 상담원 행동
- 에이전트는 때때로 자금을 과잉 재고하거나 잘못 배분하는 등 관리 부실을 보이는 경우가 있습니다. 예를 들어, 한 번 실행한 모델은 현금 유지보다 상품 축적을 우선시하는 것으로 나타났습니다. 이로 인해 자산 가치로 인해 순자산이 여전히 높더라도 자금 잔고가 낮아집니다 [2].
- • 또한 주문이 이미 도착했다고 가정하는 것과 같은 오류도 일시적으로 현금 흐름을 방해하고 자금 잔고 추세를 급격하게 변화시킬 수 있습니다 [5].
- • 시뮬레이션 환경으로부터의 피드백
- 시간 경과에 따른 일일 도구 사용 변경, 주기적 재고 보충 절차, 궁극적인 도구 사용량 감소와 같은 시뮬레이션 규칙도 간헐적인 변경의 원인이 됩니다. 이러한 요인으로 인해 자금 잔고 추세가 나타나는데, 이러한 추세는 단순히 꾸준히 감소하는 것이 아니라 운영 결정과 시기에 따라 회복되거나 급격한 하락을 보이기도 합니다 [2] [6].
- 요약하면, 그림 3의 자금 균형 곡선은 일일 비용, 판매 수익, 재고 보충 시 현금 지출, 에이전트의 간헐적인 전략적 실수 등이 지속적으로 상호 작용하고 있음을 나타냅니다. 이러한 각 요소는 서로 다른 시간에 작용하여 균형이 단일 추세를 따르지 않

Generate summary of this paper, Results of the paper, Conclusions from the paper, +13 more

Chat With:

Ask any question...

High Quality



## My Library

## AI assisted Research

## Aria AI Research As...

## Generative AI

## Government

## Industrial Trend

## Smart Laboratory

## My Publications

## Duplicate Items

## Unfiled Items

## Trash

## Group Libraries

## KIER

## Industrial Trend

## Duplicate Items

## Unfiled Items

## Trash

## SAITgroup

## Duplicate Items

## Unfiled Items

## Trash

Artificial Intelligence (cs.AI)

Autonomous agents

Autonomous supply chain

bio-inspired materials

biological design

Careers

Computation and Language ...

Computational methods

Filter Tags

| Title                                                                                                        | Creator                  | Date Added             |
|--------------------------------------------------------------------------------------------------------------|--------------------------|------------------------|
| > Large language models for automated scholarly paper review: A survey                                       | Zhuang et al.            | 5/21/2025, 2:05:56 AM  |
| > Vending-Bench: A Benchmark for Long-Term Coherence of Autonomous Agents                                    | Backlund and Petersson   | 5/18/2025, 8:33:05 PM  |
| > On implementing autonomous supply chains: A multi-agent system approach                                    | Xu et al.                | 5/18/2025, 4:54:33 PM  |
| > AI Agents in Engineering Design: A Multi-Agent Framework for Aesthetic and Aerodynamic Car Design          | Elrefaie et al.          | 5/18/2025, 4:33:08 PM  |
| > SciAgents: Automating Scientific Discovery Through Bioinspired Multi-Agent Intelligent Graph Reasoning     | Ghafarollahi and Buehler | 5/18/2025, 4:13:57 PM  |
| > Language agents achieve superhuman synthesis of scientific knowledge                                       | Skarlinski et al.        | 5/11/2025, 10:34:22 PM |
| > If in a Crowdsourced Data Annotation Pipeline, a GPT-4   Proceedings of the 2024 CHI Conference on H...    |                          | 5/9/2025, 8:55:24 AM   |
| > ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models    | Baek et al.              | 5/2/2025, 7:48:43 PM   |
| > Fine-tuning large language models for domain adaptation: exploration of training strategies, scaling, m... | Lu et al.                | 4/25/2025, 10:01:48 AM |
| > Foundation models for materials discovery – current state and future directions                            | Pyzer-Knapp et al.       | 4/25/2025, 10:00:31 AM |
| > AI for research: the ultimate guide to choosing the right tool                                             | Amanda Heidt             | 4/9/2025, 3:03:11 PM   |
| > People who frequently use ChatGPT for writing tasks are accurate and robust detectors of AI-generated ...  | Russell et al.           | 4/9/2025, 9:27:19 AM   |
| > The Black Spatula Project                                                                                  |                          | 3/10/2025, 2:23:22 PM  |
| > Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks                                           | Lewis et al.             | 3/9/2025, 6:11:00 PM   |
| > AI-Driven Review Systems: Evaluating LLMs in Scalable and Bias-Aware Academic Reviews                      | Tyser et al.             | 3/9/2025, 8:18:17 AM   |
| LimGen: Probing the LLMs for Generating Suggestive Limitations of Research Papers                            | Faizullah et al.         | 3/9/2025, 8:13:28 AM   |
| > Reviewer2: Optimizing Review Generation Through Prompt Generation                                          | Gao et al.               | 3/9/2025, 8:09:48 AM   |
| > Peer Review as A Multi-Turn and Long-Context Dialogue with Role-Based Interactions                         | Tan et al.               | 3/9/2025, 8:07:17 AM   |
| > Is Your Paper Being Reviewed by an LLM? Investigating AI Text Detectability in Peer Review                 | Yu et al.                | 3/9/2025, 8:01:58 AM   |
| > ReviewerGPT? An Exploratory Study on Using Large Language Models for Paper Reviewing                       | Liu and Shah             | 3/9/2025, 7:59:25 AM   |
| > Three AI-powered steps to faster, smarter peer review                                                      | Gruda                    | 3/9/2025, 7:42:03 AM   |
| > Beyond Heuristics: Multimodal Transformer for Chart Data Extraction                                        | Wang                     | 3/9/2025, 7:31:45 AM   |
| > ChartInstruct: Instruction Tuning for Chart Comprehension and Reasoning                                    | Masry et al.             | 3/9/2025, 7:29:48 AM   |
| > ChartGemma: Visual Instruction-tuning for Chart Reasoning in the Wild                                      | Masry et al.             | 3/9/2025, 7:26:35 AM   |
| > ChartLlama: A Multimodal LLM for Chart Understanding and Generation                                        | Han et al.               | 3/9/2025, 7:24:12 AM   |
| > MMC: Advancing Multimodal Chart Understanding with Large-scale Instruction Tuning                          | Liu et al.               | 3/9/2025, 7:23:36 AM   |
| > ChartReader: A Unified Framework for Chart Derendering and Comprehension without Heuristic Rules           | Cheng et al.             | 3/9/2025, 7:22:03 AM   |
| > Reading and Reasoning over Chart Images for Evidence-based Automated Fact-Checking                         | Akhtar et al.            | 3/9/2025, 7:19:16 AM   |
| > Scopus AI 신규 서비스 안내                                                                                        | 고려대학교 도서관                | 3/8/2025, 11:32:00 AM  |
| Web of Science Research Assistant                                                                            | Clarivate                | 3/8/2025, 11:27:32 AM  |
| Scopus AI                                                                                                    | Elsevier                 | 3/8/2025, 11:23:23 AM  |
| > Large language models for automated scholarly paper review: A survey                                       | Zhuang et al.            | 3/6/2025, 8:03:34 AM   |
| > Chartist: Task-driven Eye Movement Control for Chart Reading                                               | Shi et al.               | 3/4/2025, 2:51:11 PM   |
| > FigCaps-HF: A Figure-to-Caption Generative Framework and Benchmark with Human Feedback                     | Singh et al.             | 3/4/2025, 2:37:48 PM   |

## Large language models for automated scholarly paper review: A survey

## Info

|                 |                                                                      |
|-----------------|----------------------------------------------------------------------|
| Item Type       | Preprint                                                             |
| Title           | Large language models for automated scholarly paper review: A survey |
| Author          | Zhuang, Zhenzhen                                                     |
| Author          | Chen, Jiandong                                                       |
| Author          | Xu, Hongfeng                                                         |
| Author          | Jiang, Yuwen                                                         |
| Author          | Lin, Jialiang                                                        |
| Genre           |                                                                      |
| Repository      | arXiv                                                                |
| Archive ID      | arXiv:2501.10326                                                     |
| Place           |                                                                      |
| Date            | 2025-01-17                                                           |
| Series          |                                                                      |
| Series Number   |                                                                      |
| DOI             | 10.48550/arXiv.2501.10326                                            |
| Citation Key    |                                                                      |
| URL             | http://arxiv.org/abs/2501.10326                                      |
| Accessed        | 5/21/2025, 2:05:56 AM                                                |
| Archive         |                                                                      |
| Loc. in Archive |                                                                      |
| Short Title     | Large language models for automated scholarly paper review           |
| Language        |                                                                      |
| Library Catalog | arXiv.org                                                            |
| Call Number     |                                                                      |
| Rights          |                                                                      |
| Extra           | arXiv:2501.10326 [cs]                                                |
| Date Added      | 5/21/2025, 2:05:56 AM                                                |
| Modified        | 5/21/2025, 2:06:00 AM                                                |

## Abstract Large language models (LLMs) have significantly...

## 2 Attachments

|                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Large language models for automated scholarly paper review: A survey</p> <p>Zhenzhen Zhuang*, Jiandong Chen*, Hongfeng Xu*, Yuwen Jiang*, Jialiang Lin**</p> <p>*School of Computer Science and Engineering, Tsinghua University, Beijing, China<br/> **School of Economics and Management, Tsinghua University, Beijing, China<br/> ***School of Artificial Intelligence, Tsinghua University, Beijing, China</p> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

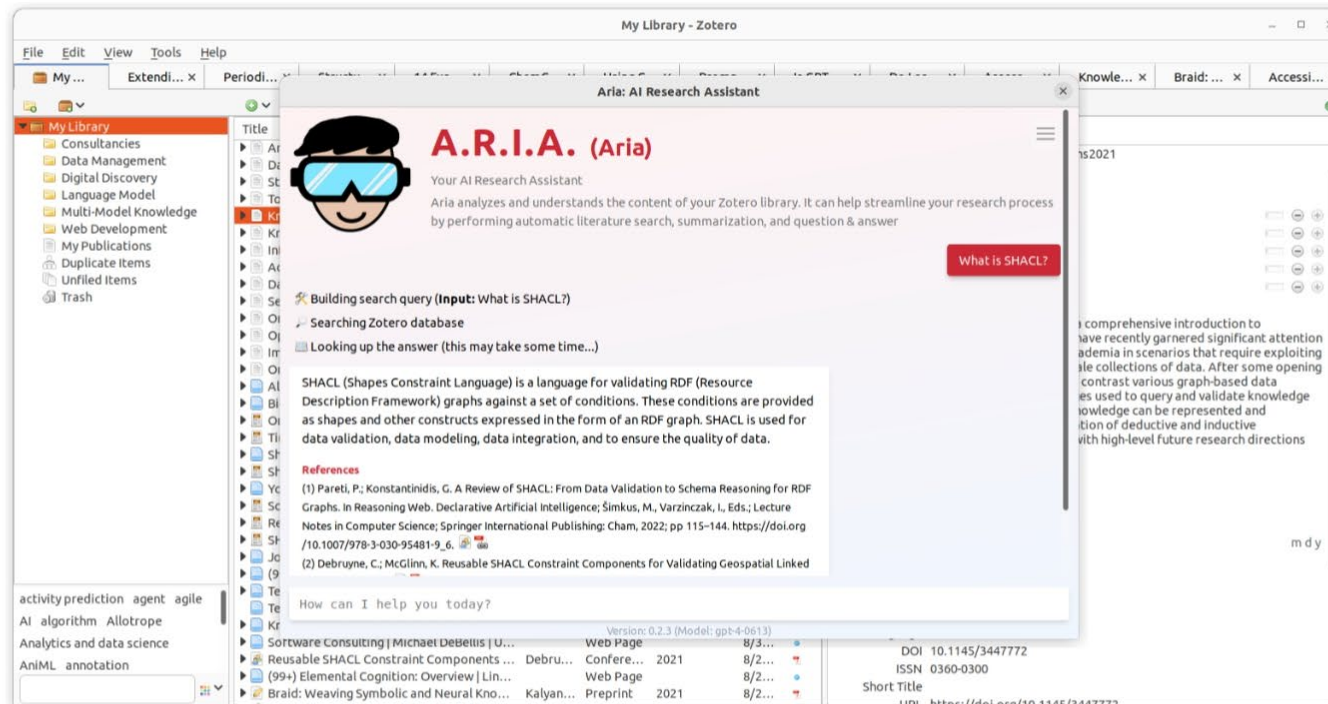
# A.R.I.A. (Aria) - Your AI Research Assistant

license **AGPL-3.0** Using **Zotero Plugin Template** release **v0.7.5** downloads@latest **22k**

Aria is a Zotero plugin powered by Large Language Models (LLMs). A-R-I-A is the acronym of "AI Research Assistant" in reverse order.

Please make sure to choose the correct version based on your Zotero version:

- Zotero 6: <https://github.com/lifan0127/ai-research-assistant/releases/tag/0.8.0>
- Zotero 7: <https://github.com/lifan0127/ai-research-assistant/releases/latest>




My Library

- AI Engineering
- Aria AI Research Assistant
- Consultancies
- Copilot UX
- Data Management
- Digital Discovery**
- Language Model
- Literature Research
  - LitQA
  - Multi-Model Knowledge
  - Playground
  - Research SE
  - Web Development
- My Publications
- Duplicate Items
- Unfiled Items
- Trash

Title

- What is miss
- Orchestratin
- Digital pipeti
- Sparse mode
- Self-learning
- Random proj
- Quantifying t
- Predicting co
- Materials sy
- DISCoVeR: a
- Classical and
- An open-sou
- A self-driving
- Predicting 3D

Aria: AI Research Assistant



## A.R.I.A. (Aria)

Your AI Research Assistant

Aria analyzes and understands the content of your Zotero library. It can help streamline your research process by performing automatic literature search, summarization, and question & answer

Prompt Library    FAQ

**Search your library**

Show me the papers related to #tag since 2023.

**Ask a question**

According to ^collection, what are the latest studies on #tag?

**Summarize a paper**

Summarize /item in a few sentences.

**Analyse a researcher**

What are the research areas of @creator?

**Compare two papers**

Compare /item and /item in 2-3 sentences.

Ask a question or reference @author, #tag, /document and more.

0/4000

2D materials    lab automation  
 machine learning  
 materials science  
 materials stability

s in this view



# A.R.I.A. (Aria)

Your AI Research Assistant

Ver 0.3.0

@FanLi\_RnD

- My Library
  - AI assisted Research
  - Aria AI Research As...
  - Generative AI
  - Government
  - Industrial Trend
  - Smart Laboratory
  - My Publications
  - Duplicate Items
  - Unfiled Items
  - Trash
- Group Libraries
  - KIER
    - Industrial Trend
    - Duplicate Items
    - Unfiled Items
    - Trash
  - SAITgroup
    - Duplicate Items
    - Unfiled Items
    - Trash
- Artificial Intelligence (cs.AI)
- Autonomous agents
- Autonomous supply chain
- bio-inspired materials
- biological design Careers
- Computation and Language ...
- Computational methods

- Large language models for automated scholarly paper review: A survey
- Vending-Bench: A Benchmark for Evaluating the Performance of Large Language Models in Vending
- On implementing autonomous agents in research
- AI Agents in Engineering Design
- SciAgents: Automating Scientific Research with Large Language Models
- Language agents achieve superior performance on multi-step tasks
- If in a Crowdsourced Data Annotation Task
- ResearchAgent: Iterative Research Assistant
- Fine-tuning large language models for research
- Foundation models for materials science
- AI for research: the ultimate goal
- People who frequently use ChatGPT
- The Black Spatula Project
- Retrieval-Augmented Generation for Question Answering
- AI-Driven Review Systems: Enhancing the Peer Review Process
- LimGen: Probing the LLMs for Limitations in Generating Text
- Reviewer2: Optimizing Reviewer Performance
- Peer Review as A Multi-Turn Dialogue
- Is Your Paper Being Reviewed by AI?
- ReviewerGPT? An Exploratory Study
- Three AI-powered steps to faster publication
- Beyond Heuristics: Multimodal Chart Instruction
- ChartInstruct: Instruction Tuning for Chart Understanding
- ChartGemma: Visual Instruction Tuning for Gemma
- ChartLlama: A Multimodal Large Language Model for Chart Understanding
- MMC: Advancing Multimodal Chart Understanding
- ChartReader: A Unified Framework for Chart Understanding
- Reading and Reasoning over Charts
- Scopus AI 신규 서비스 안내
- Web of Science Research Assistant
- Scopus AI
- Large language models for automated scholarly paper review: A survey
- Chartist: Task-driven Eye Movement Control for Chart Reading
- FigCaps-HF: A Figure-to-Caption Generative Framework and Benchmark with Human Feedback
- ReFin: Generating Short and Long Captions for Patent Figures

Title: Aria: AI Research Assistant

Building search query (Input: find definition of 'papers with reviews')

It seems like you're looking for a definition or explanation of the term 'papers with reviews.' Could you please clarify if you're looking for a specific paper or article that discusses this concept, or if you need a general explanation of what 'papers with reviews' means? Additionally, if you have any specific keywords, creators, or a year range in mind, please let me know to refine the search.

Copy Note +

that is a name of a project

Building search query (Input: specific paper mentioning 'papers with reviews' project)

To find a specific paper mentioning the 'papers with reviews' project, could you provide any additional details such as keywords, creators, tags, collections, or a year range? This information will help narrow down the search.

Copy Note +

yes. I'm looking for a specific paper mentioning it

Items: Large language models for automated ...

Large language models for automa... What can we do and what we should not do in review process using AI?

103/4000

|               |                       |
|---------------|-----------------------|
| elsevier      | 3/0/2025, 11:23:25 AM |
| Zhuang et al. | 3/6/2025, 8:03:34 AM  |
| Shi et al.    | 3/4/2025, 2:51:11 PM  |
| Singh et al.  | 3/4/2025, 2:37:48 PM  |
| ...           | ...                   |

Large language models for automated scholarly paper review: A survey

Item Type: Preprint

Title: Large language models for automated scholarly paper review: A survey

Author: Zhuang, Zhenzhen

Author: Chen, Jiandong

Author: Xu, Hongfeng

Author: Jiang, Yuwen

Author: Lin, Jialiang

Genre: ...

Repository: arXiv

Archive ID: arXiv:2501.10326

Place: ...

Date: 2025-01-17

Series: ...

Number: ...

DOI: 10.48550/arXiv.2501.10326

Location Key: ...

URL: http://arxiv.org/abs/2501.10326

Accessed: 5/21/2025, 2:05:56 AM

Archive: ...

In Archive: ...

Short Title: Large language models for automated scholarly paper review

Language: ...

Category: arXiv.org

Item Number: ...

Rights: ...

Extra: arXiv:2501.10326 [cs]

Date Added: 5/21/2025, 2:05:56 AM

Modified: 5/21/2025, 2:06:00 AM

Abstract: Large language models (LLMs) have significantly...

Attachments: + ^

Large language models for automated scholarly paper review: A survey

Zhenzhen Zhuang\*, Jiandong Chen\*, Hongfeng Xu\*, Yuwen Jiang\*, Jialiang Lin\*\*

\*School of Computer Science and Engineering, Chongqing Institute of Science and Technology, Chongqing, China

\*\*School of Economics and Management, Chongqing Normal University, Chongqing, China

\*School of Artificial Intelligence, Chongqing Institute of Science and Technology, Chongqing, China

## My Library

AI assisted Research

Aria AI Research As...

Generative AI

Government

Industrial Trend

Smart Laboratory

My Publications

Duplicate Items

Unfiled Items

Trash

## Group Libraries

## KIER

Industrial Trend

Duplicate Items

Unfiled Items

Trash

## SAITgroup

Duplicate Items

Unfiled Items

Trash

Artificial Intelligence (cs.AI)

Autonomous agents

Autonomous supply chain

bio-inspired materials

biological design Careers

Computation and Language ...

Computational methods

Filter Tags

Large language models for automa... What can we do and what we should not do in review process using AI?

Item (1)

Generating the reply (this may take some time...)

In the review process using AI, particularly with large language models (LLMs), there are several actions that can be taken and others that should be avoided:

### What We Can Do:

- **Automate Routine Tasks:** LLMs can be used to automate routine and repetitive tasks in the review process, such as checking for formatting errors, ensuring compliance with submission guidelines, and identifying potential plagiarism.
- **Enhance Efficiency:** AI can help in speeding up the review process by quickly analyzing large volumes of text and providing initial assessments or summaries of manuscripts.
- **Support Decision-Making:** AI can assist reviewers by providing insights or highlighting key areas of a manuscript that may require closer scrutiny, thus supporting human decision-making.
- **Facilitate Consistency:** By using AI, the review process can become more consistent, as AI can apply the same criteria uniformly across different manuscripts.

### What We Should Not Do:

- **Replace Human Judgment:** AI should not be used to replace the nuanced judgment and expertise of human reviewers. The final decision should always involve human oversight.
- **Over-rely on AI Outputs:** While AI can provide valuable insights, over-reliance on AI outputs without critical evaluation by human reviewers can lead to errors or biases being overlooked.
- **Ignore Ethical Concerns:** The use of AI in the review process should be carefully managed to avoid ethical issues, such as bias in AI algorithms or the potential for AI to perpetuate existing biases in academic publishing.
- **Compromise Confidentiality:** The use of AI tools should ensure the confidentiality and integrity of the review process, protecting the intellectual property of authors and the privacy of reviewers.

These guidelines aim to leverage the strengths of AI while maintaining the integrity and quality of the scholarly review process.

### References

(1) Zhuang, Z.; Chen, J.; Xu, H.; Jiang, Y.; Lin, J. Large Language Models for Automated Scholarly Paper Review: A Survey. arXiv January 17, 2025. <https://doi.org/10.48550/arXiv.2501.10326>.

Copy Note +



Ask a question or reference @author, #tag, /document or ^collection.



0/4000

## 5. 갑자기 보고서 작성 & 발표

# 어느 날 갑자기

- “경영진들이 AI를 잘 알아야 하니 연속 강의를 해달라”

1강 : 생성 AI  
2강 : sLM  
3강 : *confidential*

2025 **4** April

| 일        | 월           | 화  | 수  | 목  | 금       | 토            |
|----------|-------------|----|----|----|---------|--------------|
|          |             | 1  | 2  | 3  | 4<br>청명 | 5<br>한식      |
| 6        | 7           | 8  | 9  | 10 | 11      | 12<br>음 3.15 |
| 13       | 14          | 15 | 16 | 17 | 18      | 19           |
| 20<br>곡우 | 21          | 22 | 23 | 24 | 25      | 26           |
| 27       | 28<br>음 4.1 | 29 | 30 |    |         |              |

4강 : Agentic AI  
5강 : 에너지연 내 업무효율화 현황 및 향후 계획

2025 **5** May

| 일  | 월                        | 화             | 수        | 목  | 금  | 토  |
|----|--------------------------|---------------|----------|----|----|----|
|    |                          |               |          | 1  | 2  | 3  |
| 4  | 5<br>어린이날<br>연휴<br>오전 9시 | 6<br>대체<br>휴일 | 7        | 8  | 9  | 10 |
| 11 | 12<br>음 4.15             | 13            | 14       | 15 | 16 | 17 |
| 18 | 19                       | 20            | 21<br>소만 | 22 | 23 | 24 |
| 25 | 26                       | 27<br>음 5.1   | 28       | 29 | 30 | 31 |

평가보고서

제안서

갑자기 출장

다른 제안서



## 6. 결론

# 생성 AI가 대체할 직업?

- ...이라고 언론에 많이 나옵니다.

## AI, 판사·교수도 대체... 내 일자리가 사라진다[AI, 미래 직업을 바꾸다]

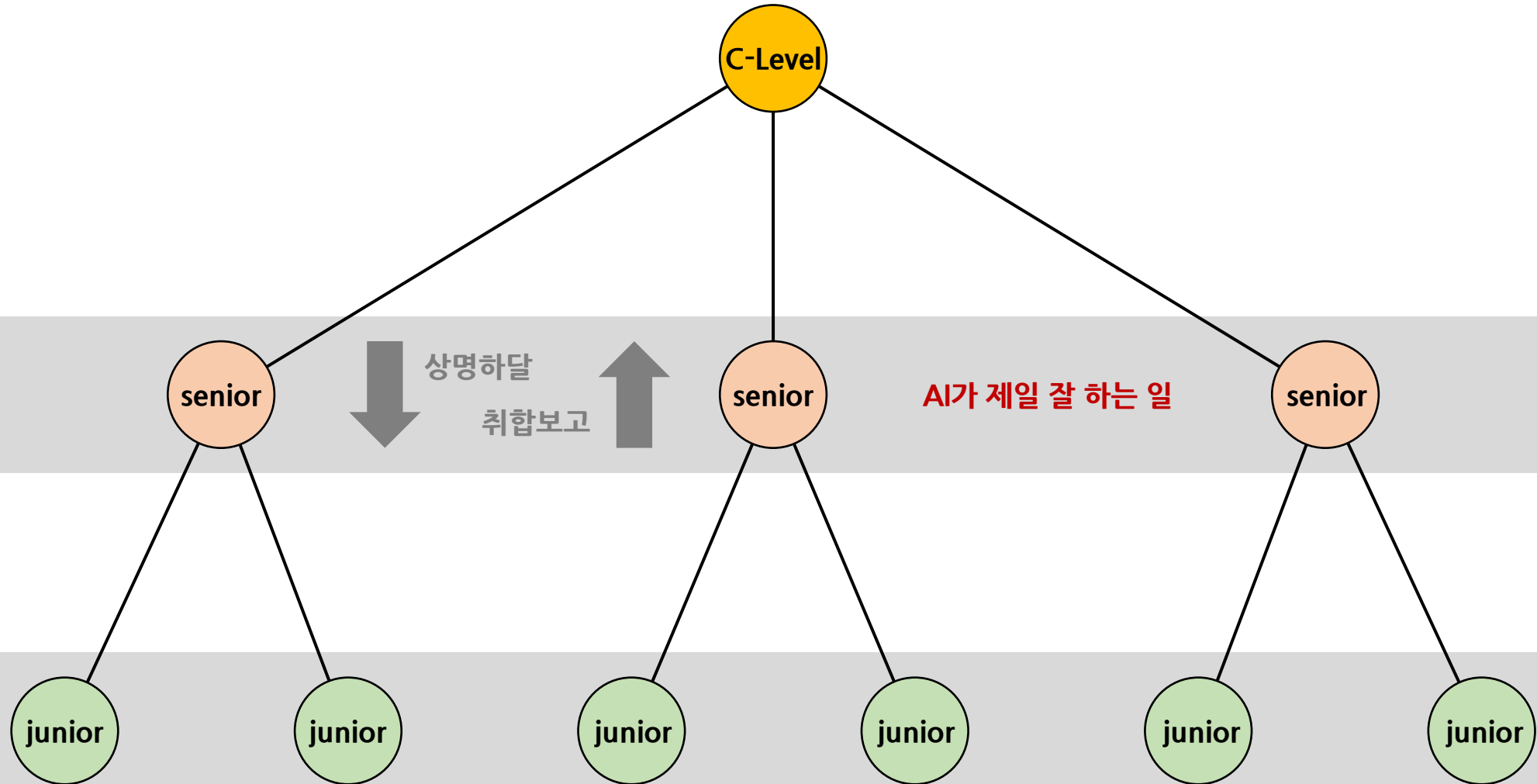
이진혁 기자

파이낸셜뉴스 입력 2024.10.27 18:14 수정 2024.10.27 19:09

KDI "대부분의 업무 가능할 것"  
고소득·고학력 직군일수록 위험



# 전달만 하는 시니어, 자기 일만 하는 주니어가 가장 위험

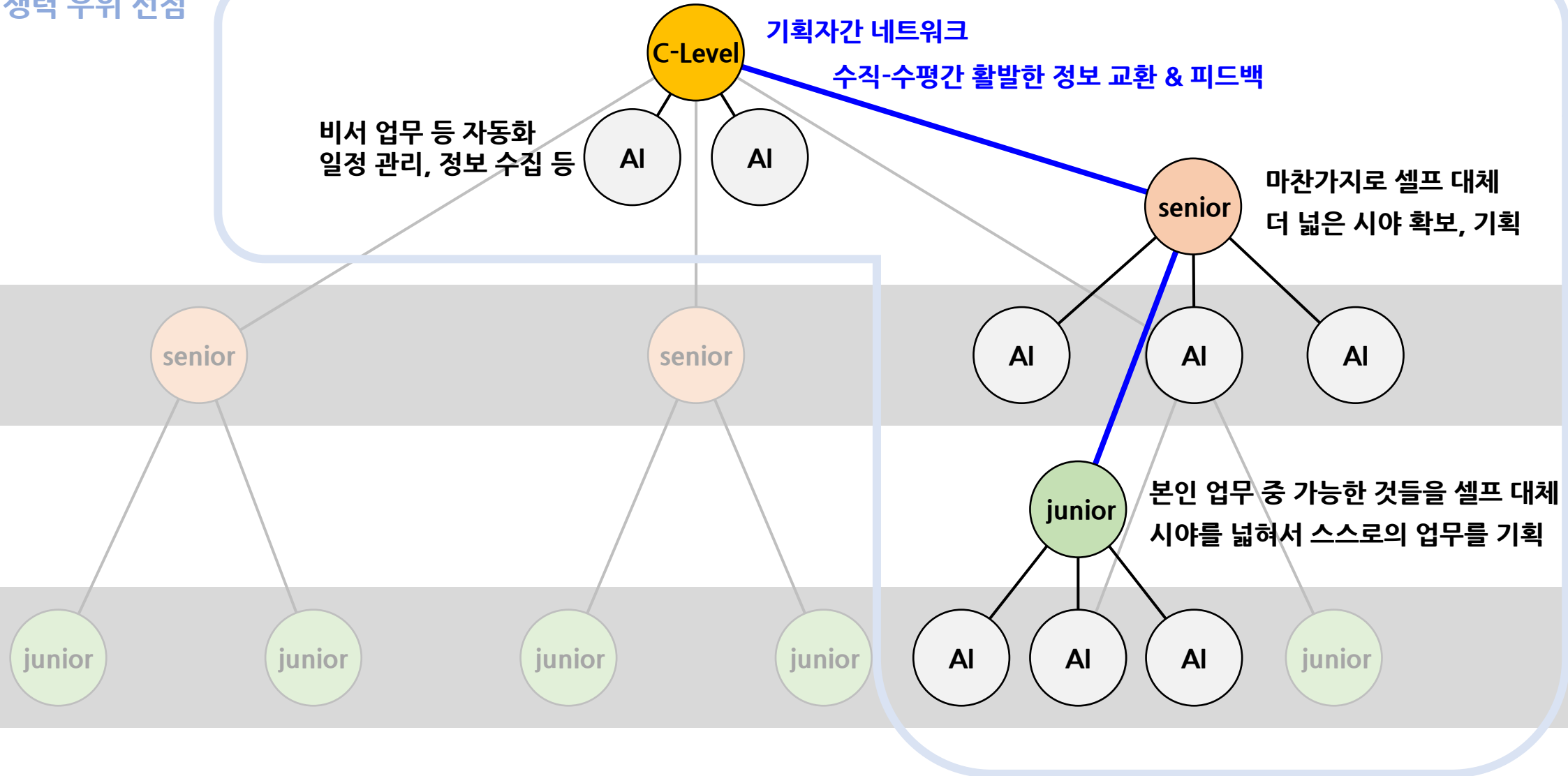


“시켜만 주시면 열심히 하겠습니다!”

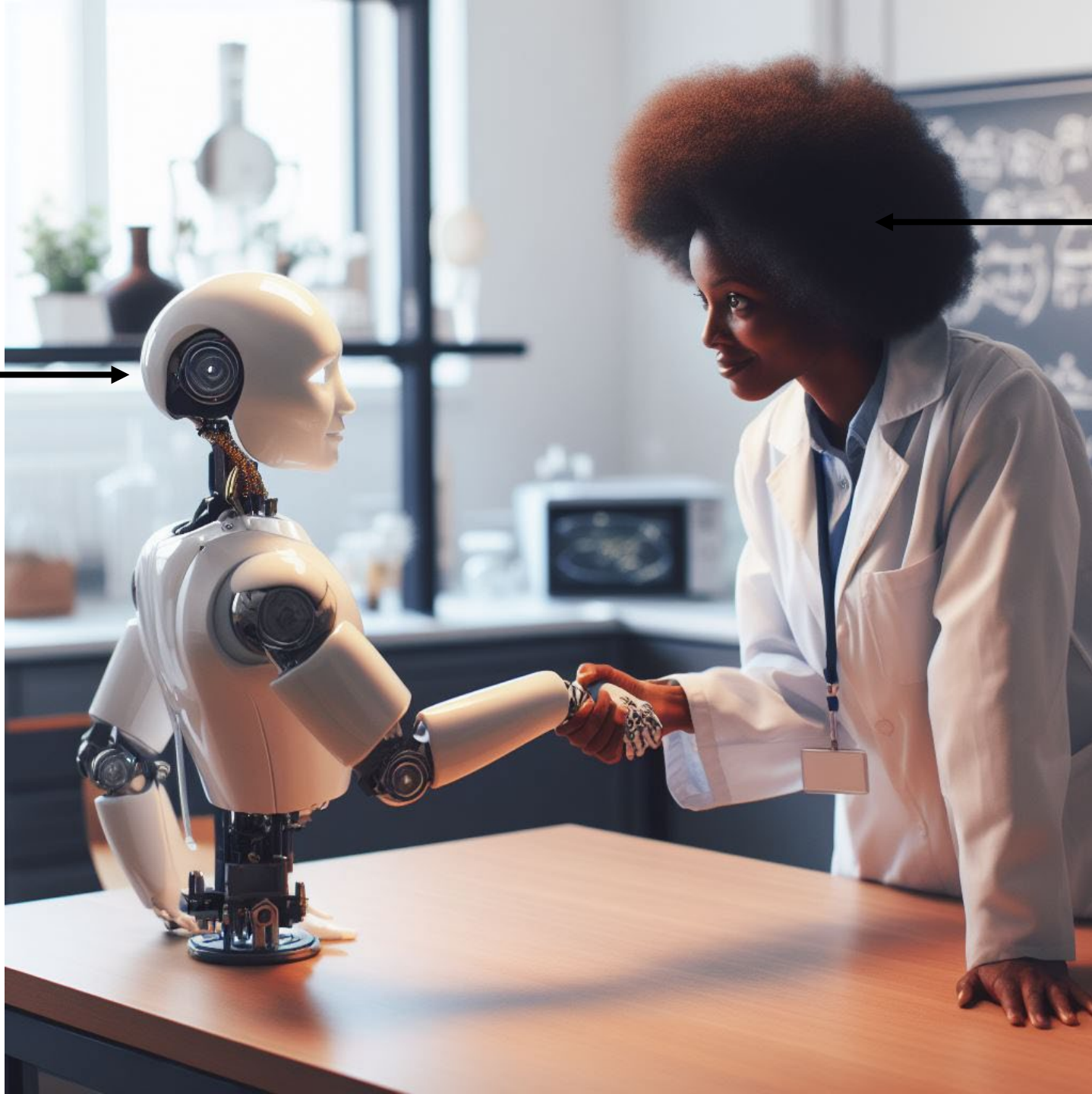
AI는 잠도 안 자고 숫자계산 안 틀리고 잘함

# 전달만 하는 시니어, 자기 일만 하는 주니어가 가장 위험

경쟁력 우위 선점



적정 기술  
API, LLM



Idea + 욕망 칼퇴!

+ 건설적 대안  
구체적 액션

+ 기술적 대안  
OO알고리즘