

비만대사연구학회 춘계학술대회: 간행위원회/연구윤리위원회 세미나

# 인공지능 활용과 연구윤리의 적용

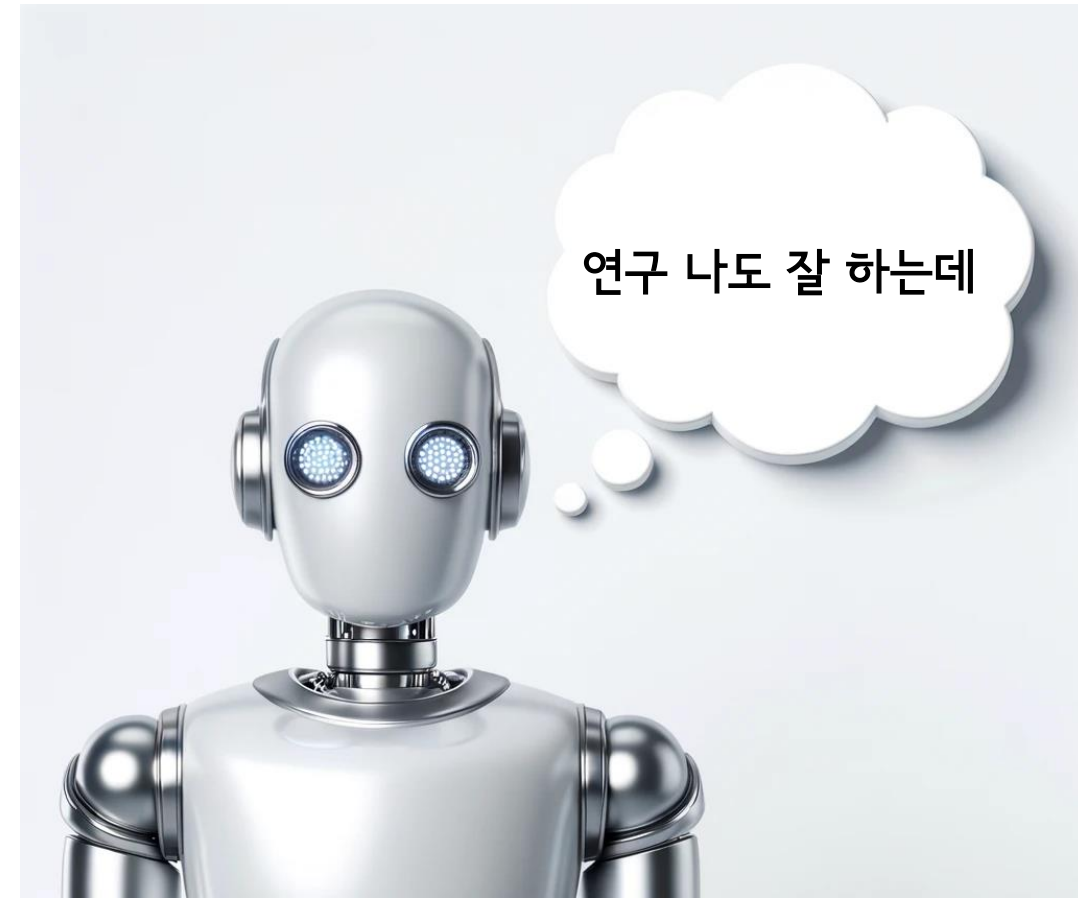
2025-03-30

인제의대 약리학교실 안상진

sangzinahn@inje.ac.kr

# 발표 내용

- 요즘 연구용 AI의 수준
- 연구용 AI의 한계점과 윤리적 고민 사항
- 앞으로 어떻게 해야 하는가?



# 요즘 연구용 AI의 수준

AI Scientist

Data Analyst

Deep Research

# AI로 생성한 논문이 peer review 통과

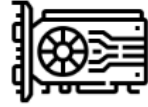
- ICLR 2025 workshop에 3개의 논문 투고
- 1개의 논문이 3명의 리뷰어 평가 결과 6.3점을 받아 acceptance (상위 45%)



# AI Scientist from Sakana AI



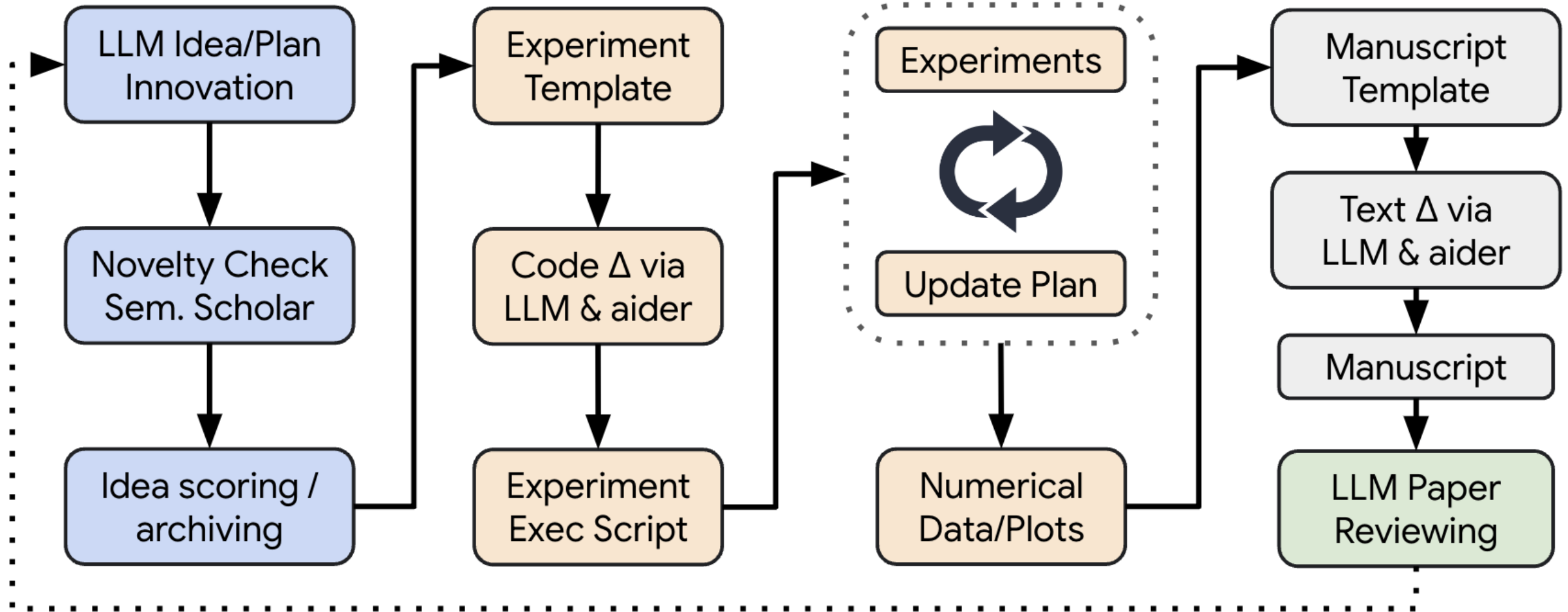
## Idea Generation




## Experiment Iteration



## Paper Write-Up



# Data Analyst from OpenAI



**Data Analyst**  
By ChatGPT

Drop in any files and I can help analyze and visualize your data.

**Other**  
Category


**Capabilities**

- Code Interpreter & Data Analysis

**Ratings**

Not enough ratings yet

**More by ChatGPT**



**Web Browser**  
I can browse the web to help you gather information or conduct research  
By ChatGPT

Start Chat

I want to write a paper about the factors that affect stable weekly warfarin dose. Do some scientific analysis using machine learning models. Visualize the results.

Analyzed ▾

Great! Your dataset has 3,345 observations and 17 variables. Here's a quick breakdown of the data:

- Target variable:** Warfarin (weekly dose in mg).
- Predictors include:**
  - Demographics:** Gender, Age, Race, Height, Weight
  - Clinical factors:** Diabetes, CHF, ValveReplacement, Aspirin, Tylenol, Simvastatin, Amiodarone, Vitamin, Smoking
  - Genetic factors:** CYP2C9, VKORC1

These features are already encoded numerically.

**Here's the plan for the scientific analysis:**

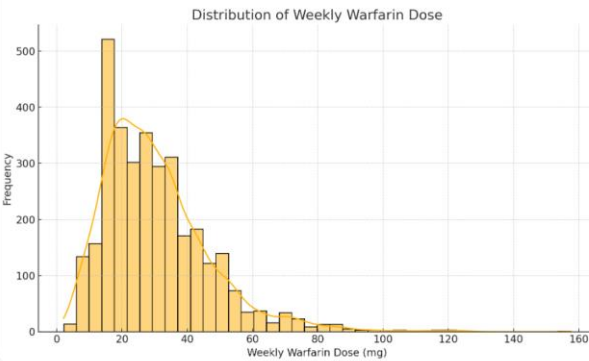
- Exploratory Data Analysis (EDA):** Understand the distributions and relationships.
- Correlation Analysis:** Check correlations with Warfarin dose.
- Feature Importance** using:
  - Linear Regression
  - Random Forest
  - Gradient Boosting
- Model Performance Comparison**
- Visualization:** Feature importance, predictions vs actual, etc.
- Interpretation** for inclusion in your paper.

Shall we start with EDA and visualization to get a feel for the data? [-]

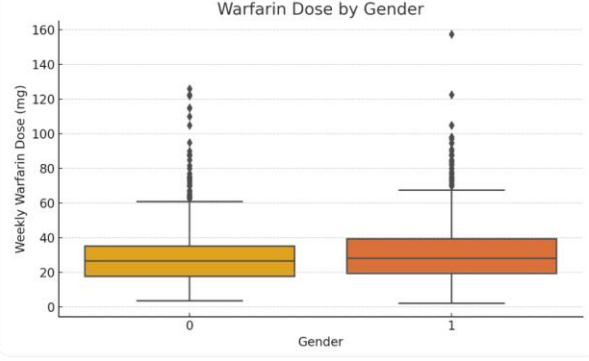
Continue

Analyzed ▾


**Distribution Of Weekly Warfarin Dose**



**Warfarin Dose By Gender**



**Warfarin Dose By Race**



# Data Analyst from OpenAI

Skip to modelling

## Model Performance Comparison

	Model	MSE	RMSE	R2 Score
1	Linear Regression	151.7349677447367	12.318074839224542	0.4831254347846
2	Random Forest	175.7988854966517	13.258917206795271	0.4011533804131
3	Gradient Boosting	152.45619590655298	12.34731533194779	0.4806686214434

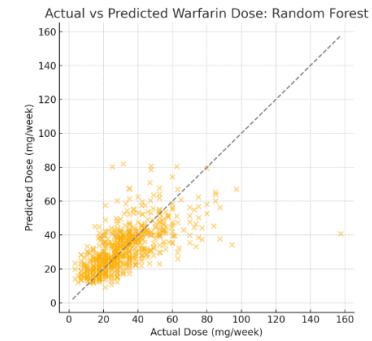
Here's a comparison of the three models (Linear Regression, Random Forest, Gradient Boosting) based on Mean Squared Error, Root Mean Squared Error, and R<sup>2</sup> Score.

Next, we can:

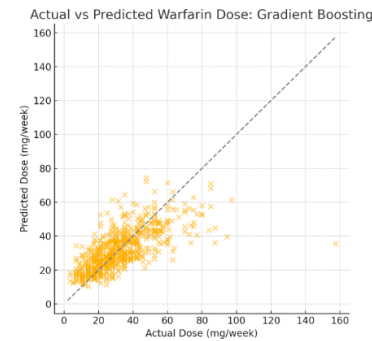
- Visualize actual vs predicted doses
- Analyze feature importance (especially for tree-based models)
- Summarize key findings for your paper

What would you like to do next? [?]

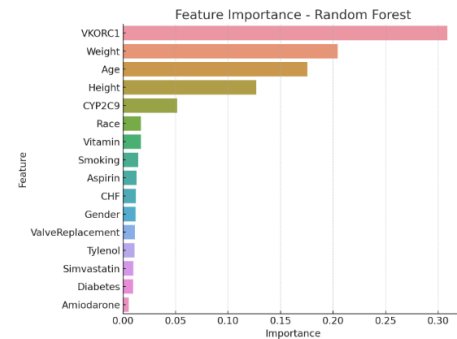
Actual Vs Predicted Warfarin Dose: Random Forest



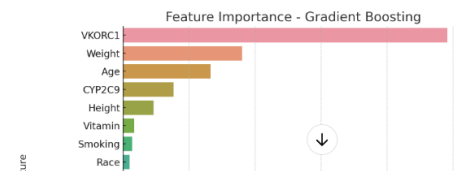
Actual Vs Predicted Warfarin Dose: Gradient Boosting




Feature Importance - Random Forest



Feature Importance - Gradient Boosting



# Deep Research from OpenAI

ChatGPT o3-mini ▾ 

**What can I help with?**

GLP1-RA 약물이 근손실을 일으킨다고 들었습니다. 최근에 나온 약물에서는 괜찮은지 궁금합니다.

Deep Research 기능 활성화

GLP-1 수용체 작용제(GLP-1 RA) 약물이 근손실을 유발할 수 있다는 점에 대해 관심을 갖고 계시군요. 최근 출시된 약물에 대해 근손실과의 관련성을 조사해 드릴 수 있습니다.

조사를 위해 아래 사항을 확인해 주세요:

- 특정 약물(예: 세마글루타이드, 티르제파타이드 등)에 대한 정보가 필요하신가요, 아니면 전체적으로 최근 GLP-1 RA 약물에 대한 검토를 원하시나요?
- 근손실과 관련된 특정 연구, 임상시험 결과, 전문가 의견 등을 포함한 자료가 필요하신가요?
- 특정 환자군(예: 비만 환자, 당뇨병 환자, 고령층 등)에 대한 정보가 필요하신가요?

더 자세한 정보를 알려주시면 보다 정확한 연구 결과를 제공해 드릴 수 있습니다.

조사 방향성 확인 질문에 대해 응답

1. 전체적으로 최근 GLP-1 RA 약물에 대한 검토를 원합니다.
2. 임상시험 결과가 포함되면 좋겠습니다.
3. 비만과 당뇨 쪽에 관심이 있습니다.



# Deep Research from OpenAI

ChatGPT o3-mini

Share



GLP-1 RA 약물에 대한 근육 손실 가능성:

- 근손실과 관련된 특정 연구, 임상시험 결과, 전문가 의견 등을 포함한 자료가 필요하신가요?
- 특정 환자군(예: 비만 환자, 당뇨병 환자, 고령층 등)에 대한 정보가 필요하신가요?

더 자세한 정보를 알려주시면 보다 정확한 연구 결과를 제공해 드릴 수 있습니다.

1. 전체적으로 최근 GLP-1 RA 약물에 대한 검토를 원합니다.
2. 임상시험 결과가 포함되면 좋겠습니다.
3. 비만과 당뇨 쪽에 관심이 있습니다.

알겠습니다! 최근 GLP-1 수용체 작용제(GLP-1 RA) 약물에 대한 근손실 가능성을 검토하고, 특히 비만 및 당뇨 환자에서의 영향을 조사하겠습니다. 최신 임상시험 결과를 포함하여 연구 및 전문가 의견을 참고해 정보를 제공하겠습니다.

조사가 완료되면 알려드리겠습니다.

Evaluating GLP-1 medications

11 sources

Message ChatGPT



Search

Deep research

ChatGPT can make mistakes. Check important info.

Activity

11 Sources



I'm analyzing how GLP-1 RAs might link to muscle loss in obesity and diabetes, considering clinical trial data, muscle mass changes, and expert opinions.

Read [dom-pubs.onlinelibrary.wiley.com](https://dom-pubs.onlinelibrary.wiley.com)

I'm considering the latest research on GLP-1-based therapies and mitigation strategies for changes in lean body mass. This likely includes updates from GLP-1 and GIP RA trials, possibly featuring tirzepatide data.

Read more from [dom-pubs.onlinelibrary.wiley.com](https://dom-pubs.onlinelibrary.wiley.com)

I'm noting that lean mass loss varies significantly across studies, with some showing 40-60% loss and others 15% or less. GLP-1 RAs and SGLT2 inhibitors often result in 20-50% lean mass loss.

Read more from [dom-pubs.onlinelibrary.wiley.com](https://dom-pubs.onlinelibrary.wiley.com)

I'm gathering data on how semaglutide and liraglutide impact lean mass during weight loss, noting significant decreases compared to total weight loss.

온라인 자료 검색을 통해  
스스로 자료조사 진행

# Deep Research from OpenAI

알겠습니다! 최근 GLP-1 수용체 작용제(GLP-1 RA) 약물에 대한 근손실 가능성을 검토하고, 특히 비만 및 당뇨 환자에서의 영향을 조사하겠습니다. 최신 임상시험 결과를 포함하여 연구 및 전문가 의견을 참고해 정보를 제공하겠습니다.

조사가 완료되면 알려드리겠습니다.

**4분 동안 조사/작성**  
**자료 출처가 표기된 보고서 형식**

Research completed in 4m · 17 sources

## GLP-1 수용체 작용제와 근손실: 최신 연구 동향

### GLP-1 RA의 체중 감량 효과와 근손실 우려

GLP-1 수용체 작용제(GLP-1 RA)는 비만 및 제2형 당뇨병 환자에서 체중 감량과 혈당 조절을 크게 개선하는 약물입니다. 세마글루타이드(semaglutide)와 티르제파타이드(tirzepatide) 등 최신 GLP-1 RA는 식욕 억제와 포만감 증가를 통해 음식 섭취를 줄여 체중을 줄이며, 그 효과가 일부 비만 수술에 필적할 정도로 크다는 보고도 있습니다. 그러나 급격하고 많은 체중 감소가 이루어지면서 근육량(제지방량) 감소에 대한 우려가 제기되었습니다. 일반적으로 체중 감량을 동반하는 어떤 방법에서도 일정 부분 근손실이 발생하며, GLP-1 RA로 인한 체중 감소에서도 예외가 아니라는 점이 전문가들에 의해 강조되었습니다. 중요한 것은 이러한 근손실이 정상적인 범주의 적응 현상인지, 아니면 근기능에 악영향을 줄 정도로 과도한 현상인지를 파악하는 것입니다.

### 비만 및 당뇨 환자에서 근육량 감소 여부와 기전

비만 환자는 정상체중인에 비해 절대 근육량이 더 많지만, 체중 감소 시 근육량도 함께 감소하는 경향이 있습니다. 식이 조절, 약물 요법, 수술 등 어떤 방법으로 체중을 감량해도 일정 비율로 근육과 같은 제지방 조직이 줄어들게 됩니다. 이는 체중 부하 감소로 근육에 가해지는 자극이 줄고, 섭취 칼로리 감소로 인한 근단백 합성 저하 등이 원인입니다. 실제로 칼로리를 약 30% 제한하면 식후 근육단백 합성률이 떨어진다는 보고도 있습니다. GLP-1 RA는 주로 섭취 열량을 감소시키기 때문에 이러한 기전으로 인한 근손실이 나타날 수 있습니다. 특히 단백질 섭취가 부족하거나 저항성 운동을 병행하지 않을 경우 근육 분해가 촉진되어 근육량 감소가 두드러질 수 있습니다.

다행히 GLP-1 RA 자체가 근육에 독성으로 작용한다는 근거는 없으며, 오히려 대사 개선 효과로 근육의 질을 향상시킬 가능성이 언급됩니다. GLP-1 RA 치료로 인슐린 감수성이 개선되고 근육 내 지방 침착이 감소하여, 비록 근육 양은 줄더라도 근육 기능은 유지되거나 향상될 수 있다는 분석이 있습니다. 한 연구에 따르면 제지방 감소율이 근육량 감소율보다 커서 신체 기능이 개선되는 경향도 관찰되었습니다. 요약하면, 체중 감량으로 인한 근육량 감소는 대부분 생리적인 범위이며, 이를 최소화하기 위해 충분한 단백질 섭취와 근력운동이 중요합니다.

특히 제2형 당뇨병 환자의 경우, 초기 근육 상태와 대사 상황이 달라 근손실 양상이 다를 수 있습니다. 당뇨병 환자는 인슐린 저항성과 염증 등으로 근육 내 지방이 끼어있는 경우가 많고, GLP-1 RA 치료로 당 조절이 개선되면 근육 분해를 촉진하는 스트레스 호르몬 수준 감소 등 긍정적 효과가 나타날 수 있습니다. 실제 임상에서 당뇨병 환자를 대상으로 한 GLP-1 RA 연구에서는 근육량 감소가 체중 감소에 비례하여 나타났으며, 대조군과 비교해 과도하지 않았습니다. 결국 비만 및 당뇨 환자에서 GLP-1 RA로 인한 근육량 변화는 체중 감량 정도에 좌우되며, 적절한 관리한다면 근감소로 인한 기능 저하 위험은 크지 않을 것으로 보입니다.

### 최신 임상시험 및 연구 결과 요약

### 최신 임상시험 및 연구 결과 요약

최근 몇 년간 발표된 임상시험과 연구들은 GLP-1 RA 치료 시 \*\*체중 감소 구성성분(지방 vs 근육)\*\*에 대한 다양한 데이터를 제공했습니다. 핵심 연구 결과를 정리하면 다음과 같습니다:

- 세마글루타이드 (비만 대상, STEP1 시험)** - 68주 동안 비만 성인을 주 1회 세마글루타이드 2.4mg으로 치료한 결과, 평균 체중이 약 15.3kg 감소하였고(총 체중의 약 -15%), 이 중 근육량(제지방량)은 약 6.9kg 감소했습니다. 전체 감량 체중의 약 45%가 근육 등 제지방 조직 손실에 해당하며, 나머지 55%는 지방 감량으로 이루어졌습니다. 절대적인 근육량은 줄었지만 체지방이 더 크게 감소하여 신체 조성상 제지방 비율은 오히려 증가한 것으로 나타났습니다. 즉, 세마글루타이드로 체중 감량 시 상당한 근손실이 동반될 수 있으나, 감량된 체중 중 대부분은 지방이라는 점이 확인되었습니다.
- 티르제파타이드 (비만 대상, SURMOUNT-1 시험)** - 주 1회 티르제파타이드 고용량 투여로 평균 22.1kg의 체중 감소(-21% 정도)가 보고되었고, 그 중 근육량은 약 5.7kg 감소하여 전체 감량의 25% 정도만 제지방 손실로 나타났습니다. 세마글루타이드에 비해 체중 감량 폭이 크지만 근손실 비율은 오히려 낮아(근손실 비율 약 25%) 체지방 감량 위주의 효과를 보였습니다. 이는 티르제파타이드의 동시 GIP 작용 효과나 전반적인 체중 감소 폭의 차이 등으로 설명될 수 있으며, 근육 보존 측면에서는 비교적 양호한 결과로 해석됩니다.
- 세마글루타이드 (당뇨병 환자, SUSTAIN-8 연구 부분적)** - 제2형 당뇨병 환자에서 세마글루타이드 1.0mg을 투여한 연구의 체성분 분석에 따르면, 약 5~6%의 체중 감소가 일어나는 동안 근육량은 약 4~5% 감소하였습니다. 체중 감소 대비 근손실 비율은 40%대로 STEP1 연구와 유사했으나, 전체 체구성에서 제지방 비율은 1% 가량 상승하여 체지방이 더 많이 줄었음을 보여주었습니다. 즉, 당뇨병 동반 비만 환자에서도 체중이 줄면 그에 비례한 근손실이 발생하지만, 체지방 감량 효과가 지배적임을 알 수 있습니다.
- 리라글루타이드 + 생활습관개선** - 한 연구에서는 리라글루타이드(liraglutide)를 투여하면서 식이 및 운동을 병행한 경우 약 6.6%의 체중 감소에 단지 2.5%의 근육량 감소만 나타나 감량 체중 중 근손실 비율이 15% 내외에 그쳤습니다. 또 다른 소규모 연구에서는 리라글루타이드로 체중을 약 0.7% 줄이는 동안 근육량에는 변화가 없었다는 보고도 있습니다. 이러한 결과들은 적극적인 운동 및 영양 관리가 동반될 경우 근육량 보존이 훨씬 용이함을 시사합니다.
- 기타 연구 및 메타분석** - 여러 임상시험의 메타분석에 따르면, GLP-1 RA로 인한 체중 감소 중 평균 20~30% 정도가 제지방(근육) 감량으로 구성되는 것으로 추산됩니다. 다만 연구마다 편차가 커서, 어떤 보고에서는 감량 체중 중 근손실 비율이 5~15%에 불과한 반면 다른 연구에서는 40~60%까지 높게 보고된 경우도 있습니다. 이러한 이질성은 연구 대상의 차이(비만 정도, 당뇨 유무, 연령), 체중 감소 속도와 기간, 측정 방법(DXA vs MRI) 및 생활습관 개입 정도에 따라 달라지는 것으로 분석됩니다. 중요하게, GLP-1 RA에 의한 근손실 비율은 식이요법이나 위절제술 등 다른 체중감량 방법에서 관찰되는 범위와 대체로 유사하며, 특별히 GLP-1 RA가 다른 방법보다 근육을 더 많이 손실시킨다는 증거는 없다는 점입니다. 예를 들어 한 연구에서는 세마글루타이드 1.0mg이나 티르제파타이드로 체중을 감량한 당뇨병 환자군 모두 전체 감량 중 15% 미만이 근손실로 나타나 대조군과 큰 차이가 없었으며, GLP-1 RA로 근육이 "과도하게" 빠진다는 근거는 발견되지 않았습니다. 이는 GLP-1 RA로 인한 체중감량의 "질(質)"이 비교적 양호하며, 감량된 체중 중 상당 부분이 지방 조직에서 나오기 때문으로 해석됩니다.

### 전문가 의견 및 권장 사항

전문가들은 현재까지의 증거를 바탕으로 GLP-1 RA로 인한 근손실 위험이 낮게 우려할 수준은 아니며, 체중 감량의 이점이 훨씬 크다고 조언합니다. 2024년 JAMA에 발표된 한 전문가 견해 논문은 \*\*\*GLP-1 기반 비만 치료로 유도된 현저한 체중 감소가 신체 허약(frailty)이나 근감소증을 일으

# 연구용 AI의 한계점과 윤리적 고민 사항

Inaccuracies (hallucination, bias)

Privacy and security

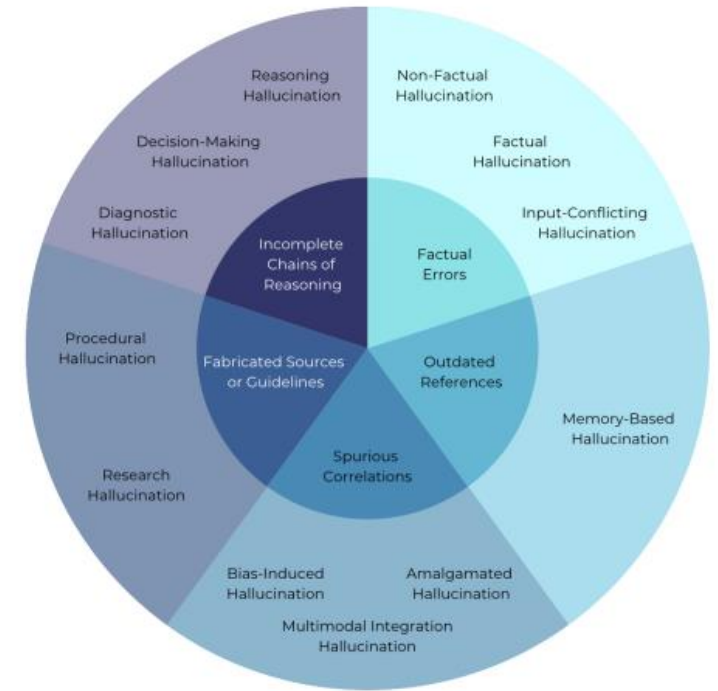
Authorship and transparency

Plagiarism

Peer review

# Inaccuracies

- **Hallucination:** LLM이 존재하지 않거나 거짓인 정보를 사실인 것처럼 만들어내는 현상
- **Bias:** 온라인 문서에 포함된 편향이 LLM 출력 결과에 포함될 수 있음
- Open access journal의 정보가 deep research 도구를 활용할 때 과대표 되는 위험
- 잘못된 과거 정보가 덮어써지지 않고 반복 재생산될 수 있는 위험
- 연구자의 AI 출력 내용 검토가 중요



**Fig. 2: A visual taxonomy of medical hallucinations in LLMs, organized into five main clusters.** (a) **Factual Errors:** Hallucinations arising from incorrect or conflicting factual information, encompassing Non-Factual Hallucination, Factual Hallucination, and Input-Conflicting Hallucination. (b) **Outdated References:** Errors stemming from reliance on obsolete guidelines or data, illustrated by Memory-Based Hallucination. (c) **Spurious Correlations:** Hallucinations that merge or misinterpret data in ways that produce unfounded conclusions, including Bias-Induced Hallucination, Amalgamated Hallucination, and Multimodal Integration Hallucination. (d) **Fabricated Sources or Guidelines:** Inventions or misrepresentations of medical procedures and research, covering Procedural Hallucination and Research Hallucination. (e) **Incomplete Chains of Reasoning:** Flawed or partial logical processes, such as Reasoning Hallucination, Decision-Making Hallucination, and Diagnostic Hallucination.

is moderate drinking beneficial for health?

Pro Filter Share

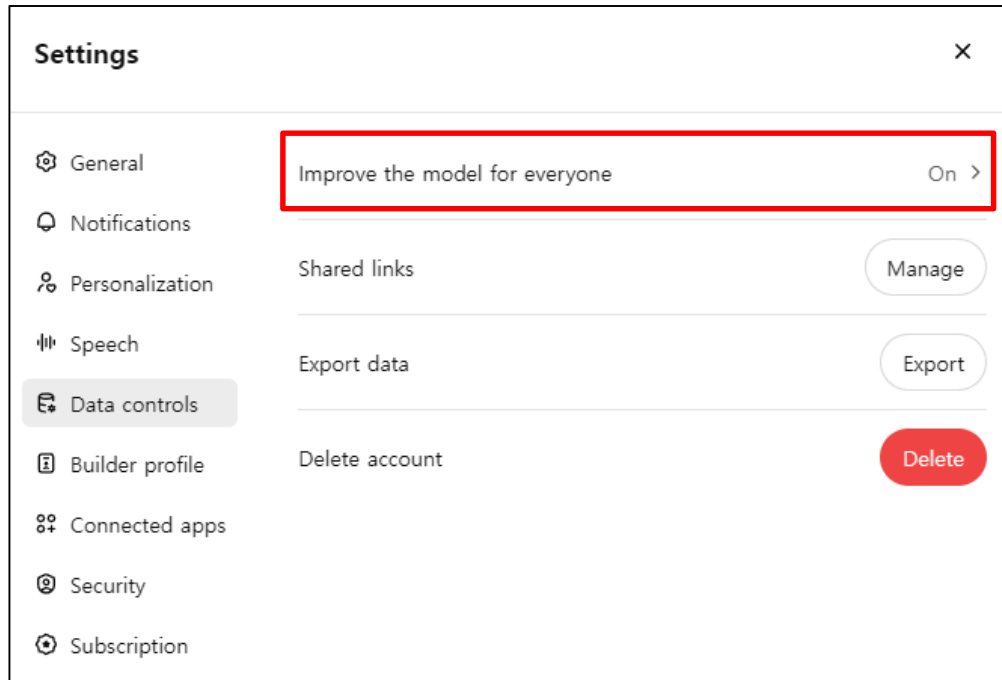
**Consensus Meter** Beta  
Results from 13 relevant papers

Yes 54% Possibly 15% Mixed 23% No 8%

Moderate drinking, defined as up to one drink per day for women and up to two for men, has been associated with certain health benefits, particularly concerning cardiovascular health. However, the evidence is mixed and should be interpreted with caution.

# Privacy and security

- 환자 정보, 미공개 연구결과 노출
- AI 서비스의 privacy 관련 설정과 규정 확인해보기
- Synthetic data 이용, Local model 활용



ChatGPT 설정

We will not use your Inputs or Outputs to train our models, unless: (1) your conversations are flagged for Trust & Safety review (in which case we may use or analyze them to improve our ability to detect and enforce our [Usage Policy](#), including training models for use by our Trust and Safety team, consistent with Anthropic's safety mission), or (2) you've explicitly reported the materials to us (for example via our feedback mechanisms), or (3) you've otherwise explicitly opted in to the use of your Inputs and Outputs for training purposes.

Claude 설정



# Publisher AI policy

- Report use of text generation tools
- Authors are responsible for contents of the paper
- LLM tools cannot take responsibility, thus cannot be an author

## Nature

First, no LLM tool will be accepted as a credited author on a research paper. That is because any attribution of authorship carries with it accountability for the work, and AI tools cannot take such responsibility.

Second, researchers using LLM tools should document this use in the methods or acknowledgements sections. If a paper does not include these sections, the introduction or another appropriate section can be used to document the use of the LLM.

## arXiv

1. continue to require authors to report in their work any significant use of sophisticated tools, such as instruments and software; we now include in particular text-to-text generative AI among those that should be reported consistent with subject standards for methodology.
2. remind all colleagues that by signing their name as an author of a paper, they each individually take full responsibility for all its contents, irrespective of how the contents were generated. If generative AI language tools generate inappropriate language, plagiarized content, errors, mistakes, incorrect references, or misleading content, and that output is included in scientific works, it is the responsibility of the author(s).
3. generative AI language tools should not be listed as an author; instead authors should refer to (1).

## Science

**Artificial intelligence (AI).** Text generated from AI, machine learning, or similar algorithmic tools cannot be used in papers published in *Science* journals, nor can the accompanying figures, images, or graphics be the products of such tools, without explicit permission from the editors. In addition, an AI program cannot be an author of a *Science* journal paper. A violation of this policy constitutes scientific misconduct.

Editorial

Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?

Siobhan O'Connor<sup>a1</sup> **ChatGPT<sup>b</sup>**

Nurse Education in Practice  
Volume 66, January 2023, 103537

Corrigendum

Corrigendum to “Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?” [Nurse Educ. Pract. 66 (2023) 103537]

Siobhan O'Connor

# Plagiarism while using AI

- 저작권: 이용자가 입출력 내용 모두에 대해 이용 권한을 가진다.
- AI의 출력을 내가 쓴 글처럼 복/붙하면?
  - 초안을 꼭 스스로 작성하고 수정하는 방식으로 접근(originality)
- AI의 도움을 받아 논문 내용을 참고한 경우, 인용을 철저히 표기하면 문제가 없다.
  - AI의 아이디어 출처를 알기 어렵다
  - AI와 토론하는 과정에서 생성된 아이디어가 기존 문헌에 있는지 확인 필요
  - LLM에게 학습데이터를 기여한 사람들에게 적절한 보상이 주어졌는가?

## Content

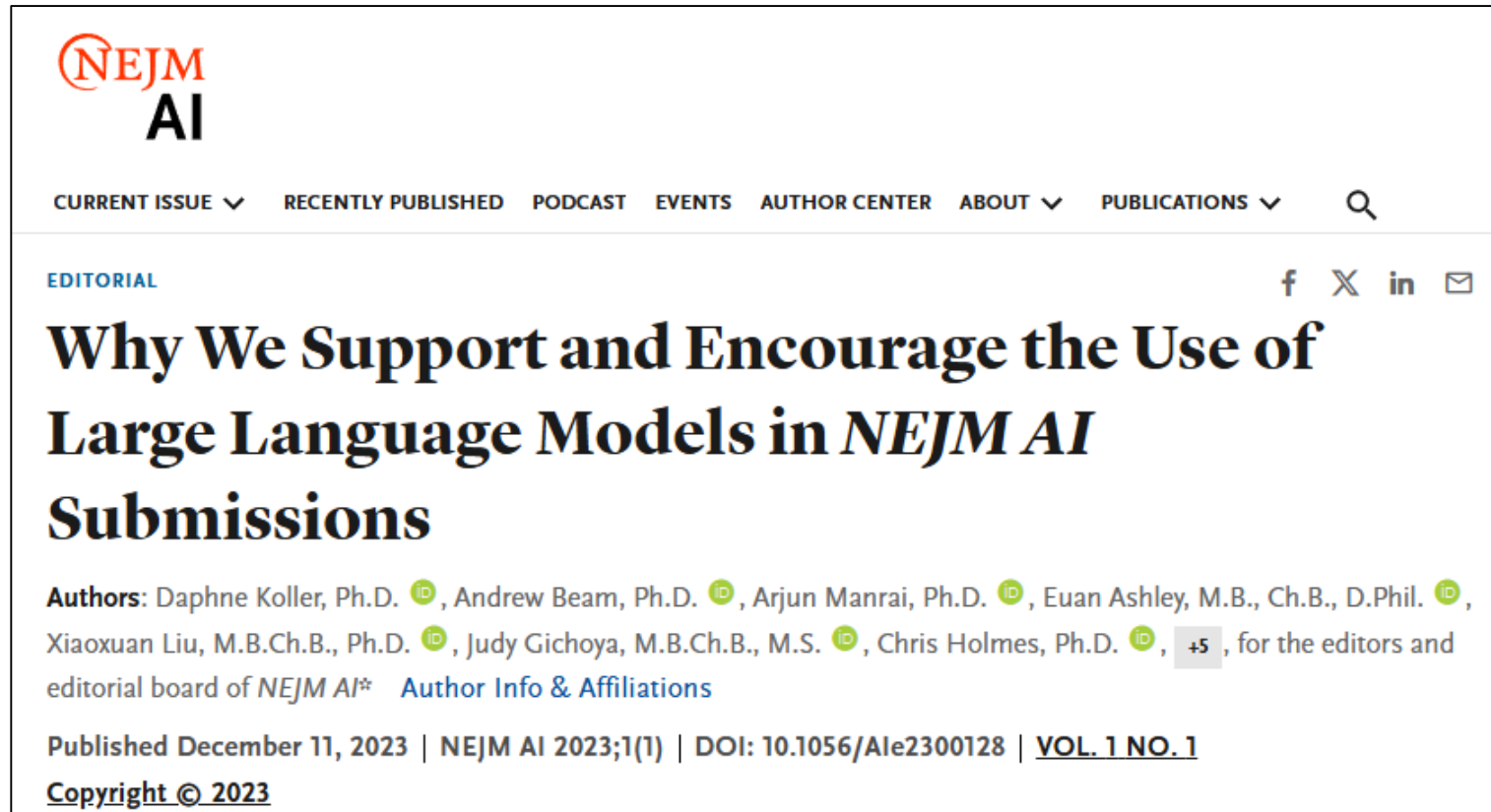
## ChatGPT terms of use

**Your content.** You may provide input to the Services (“Input”), and receive output from the Services based on the Input (“Output”). Input and Output are collectively “Content.” You are responsible for Content, including ensuring that it does not violate any applicable law or these Terms. You represent and warrant that you have all rights, licenses, and permissions needed to provide Input to our Services.

**Ownership of content.** As between you and OpenAI, and to the extent permitted by applicable law, you (a) retain your ownership rights in Input and (b) own the Output. We hereby assign to you all our right, title, and interest, if any, in and to Output.

# NEJM AI에서 LLM 활용을 장려하는 이유

- 언어장벽(영어가 모국어가 아님, 언어장애)을 줄여준다
- 연구와 논문의 질을 올려준다
- 금지하더라도 정확하게 잡아낼 수 없다
- 금지하면 규정을 지키는 사람이 역으로 손해를 본다











**NEJM AI**

CURRENT ISSUE ▾ RECENTLY PUBLISHED PODCAST EVENTS AUTHOR CENTER ABOUT ▾ PUBLICATIONS ▾ 🔍

EDITORIAL f X in ✉

## Why We Support and Encourage the Use of Large Language Models in *NEJM AI* Submissions

**Authors:** Daphne Koller, Ph.D. , Andrew Beam, Ph.D. , Arjun Manrai, Ph.D. , Euan Ashley, M.B., Ch.B., D.Phil. , Xiaoxuan Liu, M.B.Ch.B., Ph.D. , Judy Gichoya, M.B.Ch.B., M.S. , Chris Holmes, Ph.D. , , for the editors and editorial board of *NEJM AI*\* [Author Info & Affiliations](#)

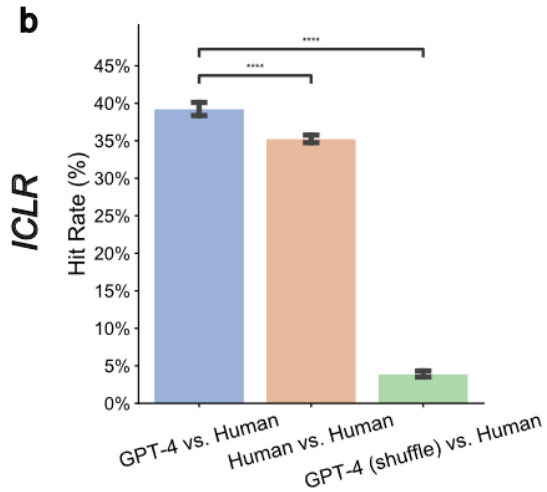
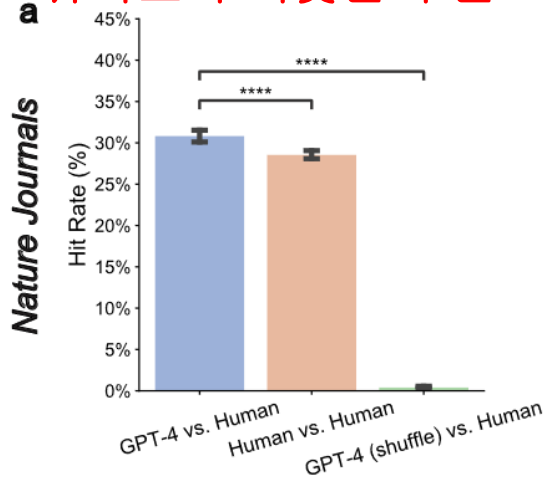
Published December 11, 2023 | *NEJM AI* 2023;1(1) | DOI: 10.1056/AIe2300128 | [VOL. 1 NO. 1](#)

Copyright © 2023



# Peer review에의 활용

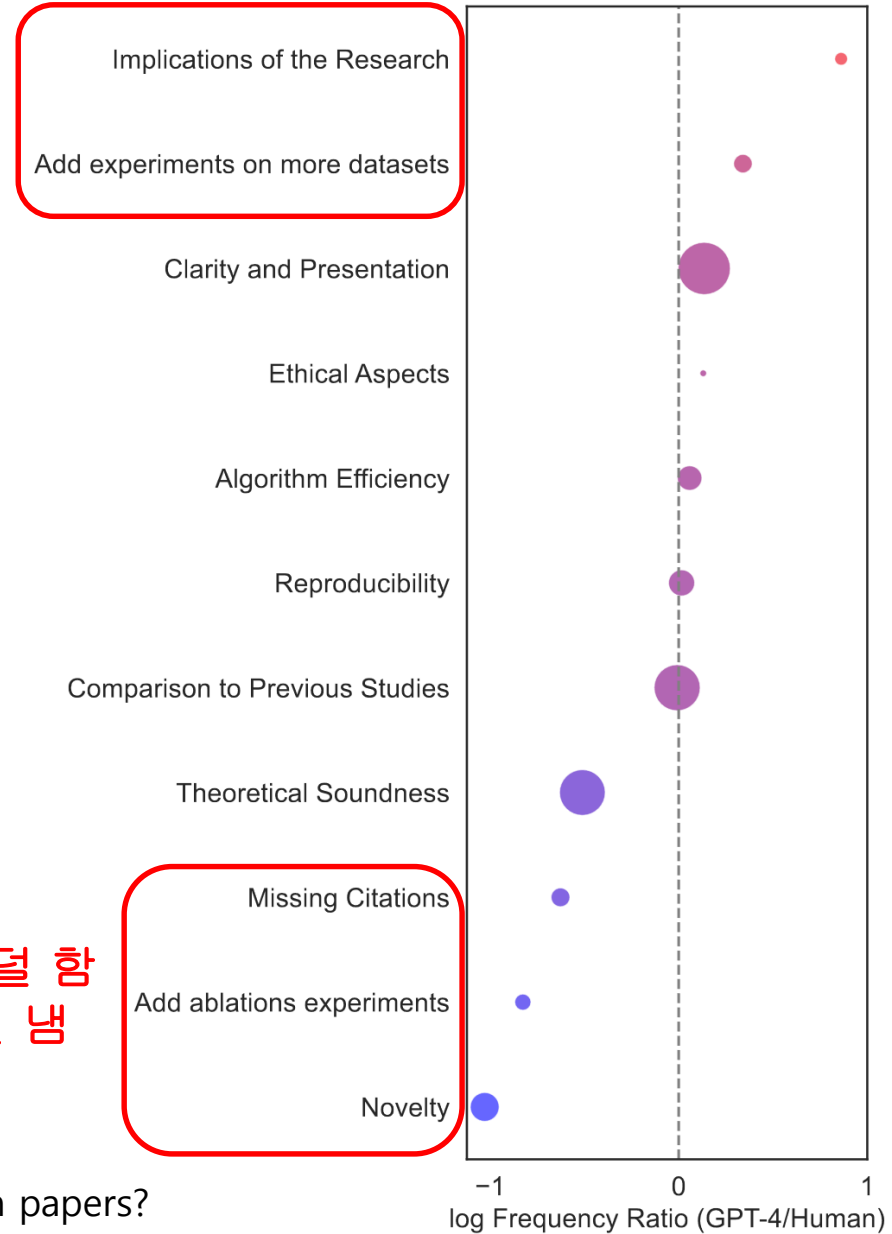
심사자간 유사도와 GPT-인간 유사도가 비슷한 수준



연구 결과의 파급효과 샘플수 추가에 대한 내용이 많음

연구의 novelty에 대한 강조를 덜 함  
ablation 실험에 대한 의견을 덜 냄  
Citation 추가 의견 안 냄

←인간이 많이 언급      LLM이 많이 언급→



# Peer review에의 활용

- (1) 15.8%의 리뷰가 AI 도움을 받음
- (2) AI리뷰가 인간리뷰보다 점수 높을 확률 53.8%
- (3) 5-6점 논문 중 AI리뷰가 포함되면 4.9%p accept 확률 높아짐

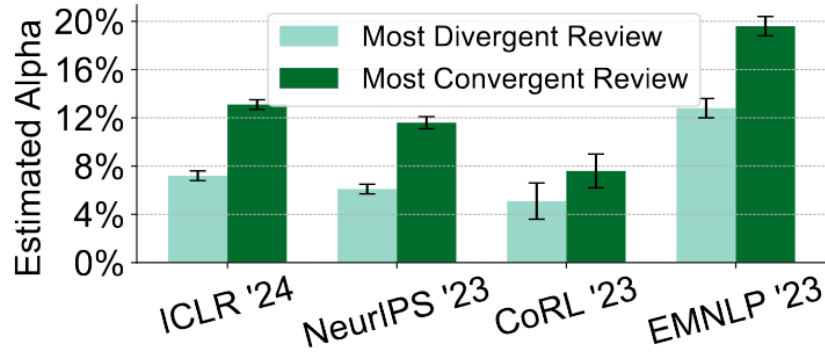


Figure 10: **The homogenization effect.** “Convergent” reviews (those most similar to other reviews of the same paper in the embedding space) tend to have a higher estimated  $\alpha$  as compared to “divergent” reviews (those most dissimilar to other reviews). See Supp. Table 26 for full results.

embedding이 중심에 가까운 리뷰의 AI 점수가 더 높음

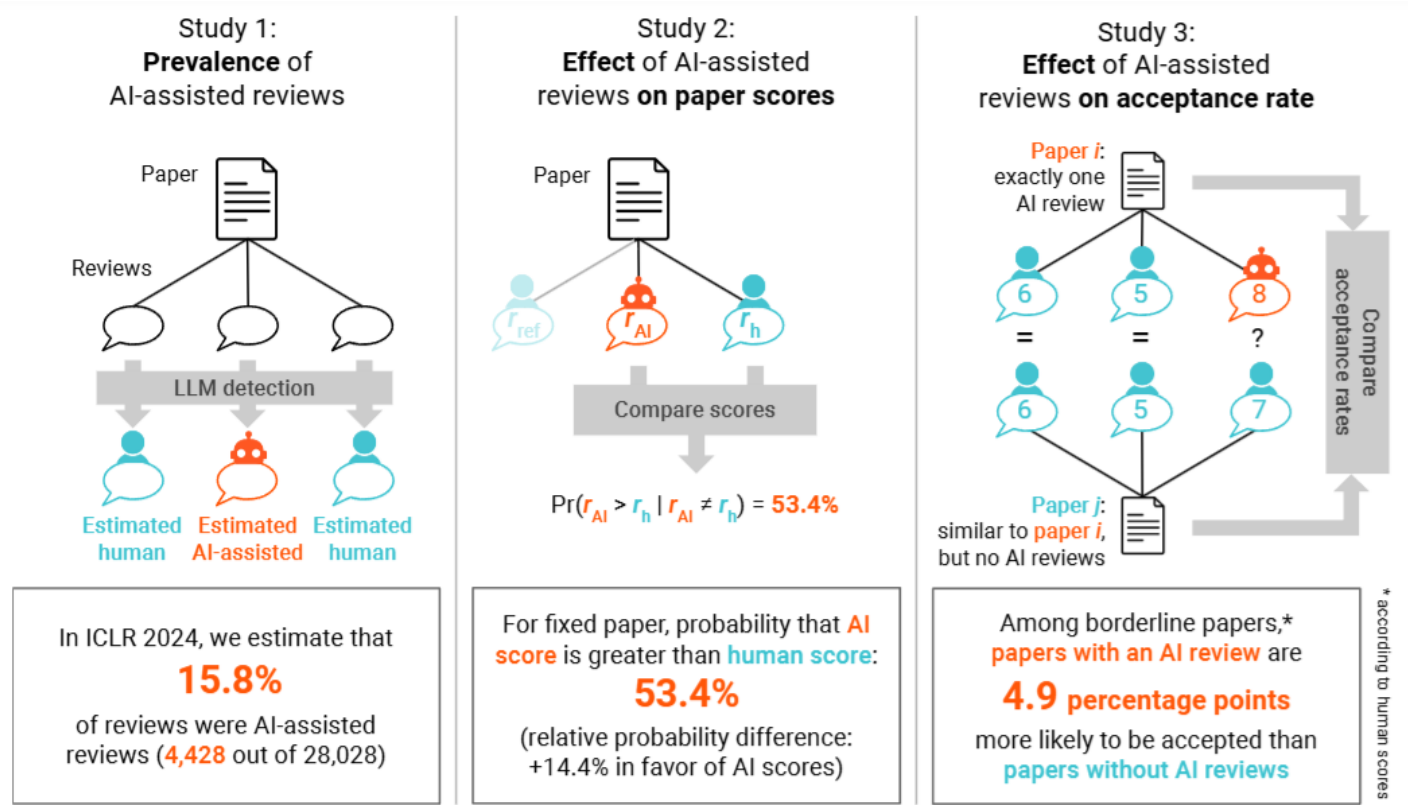


FIG. 1. Overview of our quasi-experimental approach to estimate the prevalence and causal effects of AI-assisted reviews. **Study 1: Estimating the prevalence of AI-assisted reviews** by classifying each review as human or AI-assisted using an out-of-the-box LLM-detection model. **Study 2: Estimating the effect of AI-assisted reviews on paper scores** by comparing the scores of human and AI-assisted reviews assigned to the same paper (thus controlling for properties of the reviewed paper). **Study 3: Estimating the effect of AI-assisted reviews on acceptance rate:** we match papers into pairs  $(i, j)$  such that (1)  $i$  and  $j$  are similar in content, (2)  $i$  and  $j$  received the same number  $m$  of reviews, (3)  $i$  received exactly one AI-assisted review, and  $j$  none, (4)  $i$ 's  $m - 1$  human scores are identical to  $m - 1$  of  $j$ 's  $m$  human scores. We then estimate the causal effect of AI-assisted reviews on paper acceptance as the difference in acceptance rates between  $i$  and  $j$  in matched pairs.

# 앞으로 어떻게 해야 하는가?

LLM 관련 정책 고려사항  
AI agent의 시대가 오고 있다

# LLM 관련 정책 고려사항

- 정확한 AI 검출 불가
  - False positive, false negative
- 모호한 영역이 많음
  - 대필과 침삭의 경계
- 시간의 흐름에 따른 변화
  - 지금은 불완전하지만, 갈수록 정교한 결과
  - 오타 자동 수정처럼, 문장/문단수준 자동수정 버튼(자연스럽게 사용)
- 올바른 사용법에 대한 교육
  - AI 사용을 금지하면 올바른 사용법을 교육할 수 없음
  - 논문 작성시 인간의 입력을 충분히 넣고, 검토의 중요성 강조
  - 본의 아니게 학습데이터로 활용되지 않도록 주의

# AI agent의 시대가 오고 있다

- 대부분의 연구를 AI가 하는 시대에 인간의 역할과 윤리적 고민은 무엇일까?

