

언어능력, 쓰기교수, 그리고 교수학습의 철학

2024년 12월 6일 금요일

제10차 국어교육학회 세미나

국어교육의 이론6: 다중언어시대의 언어교육과 인공지능

김성우 (서울대학교 | 캣츠랩)

발표 순서

1. 인공지능의 언어 능력 그리고 인간과 인공지능의 언어능력의 근본적 차이
2. 쓰기 교수에서 인공지능의 활용: 피드백을 중심으로
3. 생성형 인공지능의 부상과 리터러시 생태계의 변동
4. 생성형 인공지능과 교수학습의 철학: AI for UDL (보편적 학습설계를 위한 인공지능)
5. 결론: “생성 없는 생성을 경계하며”

인공지능의 언어 능력 그리고 인간과 인공지능의 언어능력의 근본적 차이

언어학자도 구별하지 못 하는 AI 생성 텍스트

- 연구 결과에 따르면, 여러 가지 근거를 사용하여 텍스트를 평가했음에도 불구하고, 리뷰어들은 AI와 인간의 글을 구별하는 데 크게 성공하지 못했으며, 전체적으로 정확한 판별률은 38.9%에 불과했습니다.
- 영어의 경우라는 점을 고려해야 하는 결과라고 판단됨.
- 하지만 한국어의 경우에도 점차 이런 현상이 나타날 것으로 예상됨.
- Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, 2(3), 100068. <https://doi.org/10.1016/j.rmal.2023.100068>

인공지능의 언어능력

"LLM은 형식적 역량(formal competence)에서 놀라울 정도로 뛰어난 성과를 보이지만, 기능적 역량(functional competence) 과제에서는 여전히 고르지 못한 성과를 보이며, 종종 전문적인 미세 조정(fine-tuning)이나 외부 모듈과의 결합이 필요합니다.

우리는 인간처럼 언어를 사용하는 모델이 되기 위해서는 이 두 가지 역량을 모두 숙달해야 하며, 이를 위해 형식적 언어 역량과 기능적 언어 역량에 특화된 별도의 메커니즘이 나타날 필요가 있을 것이라고 가정합니다."

arXiv:2301.06627v2 [cs.CL] 4 Nov 2023

DISSOCIATING LANGUAGE AND THOUGHT IN LARGE LANGUAGE MODELS

A PREPRINT

Kyle Mahowald*

The University of Texas at Austin
mahowald@utexas.edu

Anna A. Ivanova*

Massachusetts Institute of Technology
annaiv@mit.edu

Idan A. Blank

University of California Los Angeles
iblack@psych.ucla.edu

Nancy Kanwisher

Massachusetts Institute of Technology
ngk@mit.edu

Joshua B. Tenenbaum

Massachusetts Institute of Technology
jbt@mit.edu

Evelina Fedorenko

Massachusetts Institute of Technology
evelina9@mit.edu

November 7, 2023

ABSTRACT

Large language models (LLMs) have come closest among all models to date to mastering human language, yet opinions about their linguistic and cognitive capabilities remain split. Here, we evaluate LLMs using a distinction between formal linguistic competence—knowledge of linguistic rules and patterns—and functional linguistic competence—understanding and using language in the world. We ground this distinction in human neuroscience, showing that formal and functional competence rely on different neural mechanisms. Although LLMs are surprisingly good at formal competence, their performance on functional competence tasks remains spotty and often requires specialized fine-tuning and/or coupling with external modules. In short, LLMs are good models of language but incomplete models of human thought.

*The two lead authors contributed equally to this work.

	SKILLS REQUIRED FOR SUCCESSFUL LANGUAGE USE	EXAMPLE OF A FAILURE
FORMAL COMPETENCE	linguistic knowledge phonology, morphology, syntax, lexical/compositional semantics...	The keys to the cabinet is on the table.
FUNCTIONAL COMPETENCE	formal reasoning logic, math, planning...	Fourteen birds were sitting on a tree. Three left, one joined. There are now eleven birds.
	world knowledge facts, concepts, common sense...	The trophy did not fit into the suitcase because the trophy was too small.
	situation modeling discourse coherence, narrative structure...	Sally doesn't own a dog. The dog is black.
	social reasoning pragmatics, common ground, theory of mind...	Lu put the toy in the box and left. Bo secretly moved it to the closet. Lu now thinks the toy is in the closet .

Figure 1: Successful use of language relies on multiple cognitive skills, some of which (required for formal competence) are language-specific and some (required for functional competence) are not. A failure to acquire a particular skill would result in a specific type of language use deficit. Determining whether a particular failure stems from a gap in formal competence or functional competence is key to evaluating and improving language models.

But! 할루시네이션

- “학습 데이터에 의해 정당화되지 않는 것처럼 보이는 AI의 자신감 있는 응답”을 가리키는 일명 ‘할루시네이션(hallucination)’의 문제는 여러 각도에서 논의되어 왔다. 실세계로부터 도출된 데이터에 기반하지 않은 진술임에도 ‘아무런 거리낌 없이’ 응답으로 제시하는 경우는 텍스트 요약, 대화, 표로 된 데이터 등에 대한 설명, 번역 등 실로 다양한 맥락에서 관찰된다.
- 학술적 글쓰기와 관련하여 챗GPT를 사용하면서 흥미롭게 관찰한 것은 알고리즘이 생성하는 가짜 인용(fake citation)인데, 쉽게 말해 없는 논문을 지어내 근거를 제시하는 것이다.

Hallucination is Inevitable: An Innate Limitation of Large Language Models

Ziwei Xu Sanjay Jain Mohan Kankanhalli
School of Computing, National University of Singapore
ziwei.xu@u.nus.edu {sanjay,mohan}@comp.nus.edu.sg

Abstract

Hallucination has been widely recognized to be a significant drawback for large language models (LLMs). There have been many works that attempt to reduce the extent of hallucination. These efforts have mostly been empirical so far, which cannot answer the fundamental question whether it can be completely eliminated. In this paper, we formalize the problem and show that it is impossible to eliminate hallucination in LLMs. Specifically, we define a formal world where hallucination is defined as inconsistencies between a computable LLM and a computable ground truth function. By employing results from learning theory, we show that LLMs cannot learn all of the computable functions and will therefore always hallucinate. Since the formal world is a part of the real world which is much more complicated, hallucinations are also inevitable for real world LLMs. Furthermore, for real world LLMs constrained by provable time complexity, we describe the hallucination-prone tasks and empirically validate our claims. Finally, using the formal world framework, we discuss the possible mechanisms and efficacies of existing hallucination mitigators as well as the practical implications on the safe deployment of LLMs.

"학습 이론의 결과를 활용하여, 우리는 대규모 언어 모델(LLM)이 모든 계산 가능한 함수를 학습할 수 없으며 따라서 항상 환각을 일으킬 수밖에 없음을 보여줍니다. 형식적 세계는 훨씬 더 복잡한 실제 세계의 일부이기 때문에, 실제 세계를 다루는 대규모 언어 모델에서도 환각은 불가피합니다."

인공지능의 화용 능력



A feasibility study for the application of AI-generated conversations in pragmatic analysis

Xi Chen ^{a,*}, Jun Li ^b, Yuting Ye ^{c,**}

^a University of Central Lancashire, Preston, UK

^b Southern University of Science and Technology, Shenzhen, PR China

^c Southern University of Science and Technology, Business School 335, SUSTech, Shenzhen, PR China

ARTICLE INFO

Article history:

Received 11 August 2023

Received in revised form 9 January 2024

Accepted 19 January 2024

Available online 21 February 2024

Keywords:

ChatGPT

Speech act

Pragmalinguistic

Sociopragmatic

Pragmatic competence

ABSTRACT

This study explores the potential of including AI-generated language in pragmatic analysis – a field that has primarily been conducted on human language use. With the rapid growth of large language models, AI-generated texts and AI-human interactions constitute a growing field where pragmatics research is expanding to. Language data that humans used to hold a full authorship may also involve modifications made by AI. The foremost concern is thus the pragmatic qualities of AI-generated language, such as whether and to which extent AI data mirror the pragmatic patterns we have found in human speech behaviours. In this study, we compare 148 ChatGPT-generated conversations with 82 human-written ones and 354 human evaluations of these conversations. The data are analysed using various methods, including traditional speech strategy coding, four computational methods developed in NLP, and four statistical tests. The findings reveal that ChatGPT performs equally well as human participants in four out of the five tested pragmalinguistic features and five out of six sociopragmatic features. Additionally, the conversations generated by ChatGPT exhibit higher syntactic diversity and a greater sense of formality compared to those written by humans. As a result, our participants are unable to distinguish ChatGPT-generated conversations from human-written ones.

© 2024 Elsevier B.V. All rights reserved.

- 우리의 설계된 프롬프트로 생성된 ChatGPT 대화는 화용론적 분석(pragmatic analysis)에서 인간과 유사한 데이터로 간주될 수 있습니다.
- 이 대화는 화용언어적 특징(pragmalinguistic features) 다섯 가지 중 네 가지, 그리고 사회화용론적 특징(sociopragmatic features) 여섯 가지 중 다섯 가지에서 인간 대화와 동등한 성과를 보였습니다.
- ChatGPT는 구문적 다양성(syntactic diversity)과 적절한 격식 수준(formality levels)의 선택에서 인간 참여자들을 능가했습니다.
- 우리의 참여자들은 ChatGPT가 생성한 대화를 인간이 작성한 대화와 구별하지 못했습니다.

Chen, Xi, Li, Jun & Ye, Yuting (2024). A feasibility study for the application of AI-generated conversations in pragmatic analysis. Journal of pragmatics, 223. 14–30.

하지만 인공지능과 인간의 언어는 다르다!

- 인간의 언어 습득: 인간은 일상 환경에서 돌보는 이와 의사소통 상호작용을 통해 언어를 배우며, 이러한 상호작용은 언어의 형태, 의미, 의사소통 기능에 기반한 언어적 지식 습득에 필수적이다.
- 기계 학습의 한계: 현재의 대규모 언어 모델(LLMs)은 주로 텍스트 데이터에 의존하며, 인간과 같은 맥락적이고 의사소통적인 학습이 부족하여 데이터 의존성, 논리적/실용적 추론의 제한, 편향에 취약하다.
- 대안적 접근법 제안: 기계가 맥락적 의사소통 상호작용에 참여하여 언어 구조를 학습하도록 설계하는 접근법이 제안되며, 이는 보다 인간적인 언어 처리로 나아갈 수 있는 잠재력을 제공한다.

인간과 LLM이 언어를 배우는 방법의 차이

1. 학습의 기초

- 인간: 의미 있는 의도적이고 상황적으로 맥락화된 상호작용(situated communicative interactions)을 통해 언어를 습득한다. 이 과정은 환경적 단서와 사회적 교류에 의존한다.
- 기계: 텍스트 데이터(textual data)를 통해 언어를 학습하며, 단어 간 공출현(co-occurrence) 패턴에 집중하고 상황적 또는 의도적 맥락을 포함하지 않는다.

2. 학습 메커니즘

- 인간: 두 가지 상호보완적인 인지 기술을 활용한다: 의도 읽기(intention reading): 상황적 맥락(제스처나 환경적 단서 포함)을 바탕으로 의사소통 의도를 추론. / 패턴 찾기(pattern finding): 상호작용 간 언어적 패턴을 일반화하여 생산적 도식을 형성.
- 기계: 분포적 언어학(distributional linguistics) 원칙을 사용하며, 단어 공기 패턴을 분석하여 텍스트 코퍼스(corpus)에서 통계적 관계를 캡처하는 임베딩(embedding)을 생성.

인간과 LLM이 언어를 배우는 방법의 차이

3. 입력 규모와 효율성

- 인간: 제한적이지만 풍부한 다중양식 상호작용(multimodal interactions)을 통해 언어를 습득. 5세까지 약 6천만 개 토큰(token), 20세까지 약 2억 개 토큰에 노출.
- 기계: 언어 모델 훈련에는 수조(trillion) 단위의 텍스트 데이터가 필요. 예컨대, GPT-3 모델은 3천억 개 토큰으로 학습, Chinchilla 모델은 1.4조 개 토큰으로 학습됨.

4. 인식론적 확실성

- 인간: 특히 초기 단계에서는 언어적 가설에 불확실성이 수반되지만, 성공적인 의사소통을 통해 반복적으로 정교화.
- 기계: 출력의 인식론적 상태(epistemological status)가 균일해, 사실과 허구가 혼합된 환각(hallucination)이 발생. 출력물이 진실(truth)이나 의도(intent)에 기반하지 않음.

인간과 LLM이 언어를 배우는 방법의 차이

5. 적응성과 진화

- 인간: 상호작용을 통해 언어 체계를 동적으로 발전시키며, 성공적인 의사소통을 기반으로 구조(construction)를 강화하거나 약화.
- 기계: 의사소통 성공과 연계된 적응 메커니즘이 없으며, 사전 정의된 텍스트 데이터의 통계적 패턴에 의존.

6. 편향과 맥락 인식

- 인간: 동적이고 사회적으로 맥락화된 언어 해석을 통해 편향을 극복 가능.
- 기계: 훈련 데이터에 내재된 편향(bias)을 반영하거나 증폭. 나이, 인종, 성별, 종교, 성적 지향 등과 관련된 편향이 포함될 가능성.

인간과 LLM이 언어를 배우는 방법의 차이

7. 생성 능력

- 인간: 맥락적으로 적합한 프레임워크 내에서 새로운 발화를 생성하고 이해.
- 기계: 통계적으로 가능성이 높은 텍스트를 생성하며, 맥락적으로 그럴듯하지만 사실적이거나 의미적으로 정확하지 않을 수 있음.

8. 의사소통 범위

- 인간: 초기에는 맥락에 의존하지만 점차 추상적이고 분리된 주제에 대해 논의할 수 있도록 도식이 발달.
- 기계: 추상적이고 분리된 담론을 생성할 수 있으나, 경험적 기반이 없기 때문에 추상적 개념의 내재적 이해는 부족.

인간과 LLM이 언어를 배우는 방법의 차이

9. 추론과 추리(inference)

- 인간: 귀납적 추론(inductive reasoning)을 통해 패턴을 발견하고, 가설적 추론(abductive reasoning)을 통해 의도 읽기를 수행하여 논리적이고 실용적인 추론을 가능하게 함.
- 기계: 의도적 의사소통 맥락(intentional communicative contexts)의 부재로 인해 논리적, 실용적 추론에 어려움을 겪음. 의도 재구성이 필요한 실용적 추론(pragmatic inference)은 특히 어려움.

10. 지식 표현의 본질

- 인간: 언어적 지식을 현실 세계 경험(real-world experiences)과 상호작용을 기반으로 구축. 추상적 개념조차 감각-운동(sensory-motor) 또는 사회적 경험에 간접적으로 연결.
- 기계: 단어의 공기 행동을 캡처하는 고차원 벡터 임베딩(vector embedding)으로 언어적 지식을 표현. 현실 세계 경험에 기반하지 않음.

쓰기 교수에서 인공지능의 활용: 피드백을 중심으로

학생들의 작문 능력 개발을 촉진하는 중재자로서의 AI

양적 분석에 따르면, AI 지원 교육을 받은 학생들이 대조군에 비해 **작문 실력과 동기 부여에서 모두 유의미한 향상**을 보인 것으로 나타났다.

질적 결과에서는 AI의 혁신적인 교육적 역할과 작문 실력 및 동기 부여에 대한 긍정적인 영향을 인정하는 의견에서부터 **맥락적 정확성과 지나친 의존성에 대한 우려**에 이르기까지 다양한 관점이 제시되었다.

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*, 1260843.



OPEN ACCESS

EDITED BY
Valentin Riemer,
University of Ulm, Germany

REVIEWED BY
Bin Zou,
Xi'an Jiaotong-Liverpool University, China
Musa Nushi,
Shahid Beheshti University, Iran

*CORRESPONDENCE
Cuiping Song
✉ songcp2023@163.com

RECEIVED 18 July 2023
ACCEPTED 28 November 2023
PUBLISHED 15 December 2023

CITATION
Song C and Song Y (2023) Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Front. Psychol.* 14:1260843. doi: 10.3389/fpsyg.2023.1260843

COPYRIGHT
© 2023 Song and Song. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students

Cuiping Song^{1*} and Yanping Song²

¹School of Foreign Studies, North Minzu University, Yinchuan, Ningxia, China, ²School of Public Administration, Central South University, Changsha, Hunan, China

Introduction: This mixed-methods study evaluates the impact of AI-assisted language learning on Chinese English as a Foreign Language (EFL) students' writing skills and writing motivation. As artificial intelligence (AI) becomes more prevalent in educational settings, understanding its effects on language learning outcomes is crucial.

Methods: The study employs a comprehensive approach, combining quantitative and qualitative methods. The quantitative phase utilizes a pre-test and post-test design to assess writing skills. Fifty EFL students, matched for proficiency, are randomly assigned to experimental (AI-assisted instruction via ChatGPT) or control (traditional instruction) groups. Writing samples are evaluated using established scoring rubrics. Concurrently, semi-structured interviews are conducted with a subset of participants to explore writing motivation and experiences with AI-assisted learning.

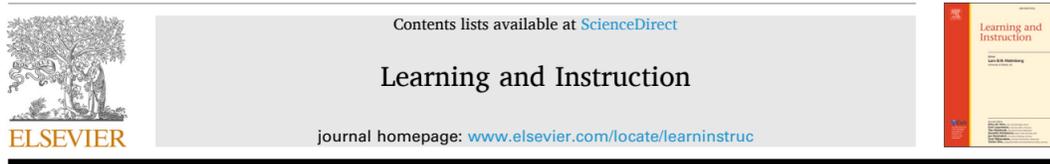
Results: Quantitative analysis reveals significant improvements in both writing skills and motivation among students who received AI-assisted instruction compared to the control group. The experimental group demonstrates enhanced proficiency in various aspects of writing, including organization, coherence, grammar, and vocabulary. Qualitative findings showcase diverse perspectives, ranging from recognition of AI's innovative instructional role and its positive influence on writing skills and motivation to concerns about contextual accuracy and over-reliance. Participants also reflect on the long-term impact and sustainability of AI-assisted instruction, emphasizing the need for ongoing development and adaptation of AI tools.

Discussion: The nuanced findings offer a comprehensive understanding of AI's transformative potential in education. These insights have practical implications for practitioners and researchers, emphasizing the benefits, challenges, and the evolving nature of AI's role in language instruction.

연구 결과 해석에서 주의할 점

- 참가자들의 숙련도 수준: 중급에서 상급 중급
- 작문 과제의 장르: IELTS(논증적 작문) -- LLM은 다른 장르에 비해 이 장르를 평가하는 방법에 대해 상대적으로 많은 지식을 축적했을 가능성이 높음.
- “Ahn과 Chung(2020), Chung과 Ahn(2022)의 연구는 기계 번역이 작문에 미치는 영향이 장르에 따라 어떻게 달라지는지를 분석했다. **두 연구 모두 공통적으로 학습자가 논증적 작문에서 기계 번역 결과물을 서사적 작문보다 더 높은 수용률로 받아들였음을 보여주었다.**”
- Chung과 Ahn(2022)은 논증적 작문이 표준화된 구조와 특정 어휘를 사용하는 반면, 서사적 작문은 창의적이고 추상적인 개념을 표현하기 위해 높은 수준의 어휘 다양성을 요구하기 때문에, 논증적 작문이 기계 번역에 더 적합하다고 설명

피드백 제공자로서의 AI (1)



Comparing the quality of human and ChatGPT feedback of students' writing

Jacob Steiss^{a,*}, Tamara Tate^a, Steve Graham^b, Jazmin Cruz^c, Michael Hebert^a, Jiali Wang^a, Youngsun Moon^a, Waverly Tseng^a, Mark Warschauer^a, Carol Booth Olson^a

^a School of Education University of California, Irvine, USA
^b Mary Lou Fulton Teachers College Arizona State University, USA
^c WestEd, USA

ARTICLE INFO

Keywords:
Automated writing evaluation
Formative feedback
Writing instruction

STRUCTURED ABSTRACT

Background: Offering students formative feedback on their writing is an effective way to facilitate writing development. Recent advances in AI (i.e., ChatGPT) may function as an automated writing evaluation tool, increasing the amount of feedback students receive and diminishing the burden on teachers to provide frequent feedback to large classes.

Aims: We examined the ability of generative AI (ChatGPT) to provide formative feedback. We compared the quality of human and AI feedback by scoring the feedback each provided on secondary student essays. We scored the degree to which feedback (a) was criteria-based, (b) provided clear directions for improvement, (c) was accurate, (d) prioritized essential features, and (e) used a supportive tone.

Sample: 200 pieces of human-generated formative feedback and 200 pieces of AI-generated formative feedback for the same essays.

Methods: We examined whether ChatGPT and human feedback differed in quality for the whole sample, for compositions that differed in overall quality, and for native English speakers and English learners by comparing descriptive statistics and effect sizes.

Results: Human raters were better at providing high-quality feedback to students in all categories other than criteria-based. AI and humans showed differences in feedback quality based on essay quality. Feedback did not vary by language status for humans or AI.

Conclusion: Well-trained evaluators provided higher quality feedback than ChatGPT. Considering the ease of generating feedback through ChatGPT and its overall quality, generative AI may still be useful in some contexts, particularly in formative early drafts or instances where a well-trained educator is unavailable.

"인간 평가자는 기준 기반(category-based)을 제외한 모든 범주에서 학생들에게 **더 높은 품질의 피드백을 제공**하는데 뛰어났다"

교사의 전문성: 모든 평가자는 학생들의 작문 교육과 피드백 제공 경험을 보유하고 있었으며, **16명 중 12명은 작문 교육에서 15년 이상의 경력**을 가지고 있었다.

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., ... & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 101894.

AI의 작문 피드백에 관한 학습자의 평가

1. 다양한 기능을 갖춘 **글쓰기 도우미** (Multi-tasking writing assistant):

- 검색 엔진 (Search engine) (93%)
- 사고 유도자 (Thought Provoker) (86%): 아이디어 생성과 파트너 역할, 촉진자로서의 역할
- 초급 수준 글쓰기 지원 (Entry-level writer) (78%): 기술적 또는 내용적 글쓰기를 지원
- 교정자 (Proofreader) (72%)

2. 1분 이내에 응답하는 개인 **가상 튜터** (Private virtual tutor on-demand within a minute or less) (87%)

3. **디지털 동료** (Digital peer) (72%)

Education and Information Technologies
<https://doi.org/10.1007/s10639-024-12878-7>



Exploring students' perspectives on Generative AI-assisted academic writing

Jinhee Kim¹ · Seongryeong Yu² · Rita Detrick¹ · Na Li^{3,4}

Received: 6 February 2024 / Accepted: 21 June 2024
© The Author(s) 2024

Abstract

The rapid development of generative artificial intelligence (GenAI), including large language models (LLM), has merged to support students in their academic writing process. Keeping pace with the technical and educational landscape requires careful consideration of the opportunities and challenges that GenAI-assisted systems create within education. This serves as a useful and necessary starting point for fully leveraging its potential for learning and teaching. Hence, it is crucial to gather insights from diverse perspectives and use cases from actual users, particularly the unique voices and needs of student-users. Therefore, this study explored and examined students' perceptions and experiences about GenAI-assisted academic writing by conducting in-depth interviews with 20 Chinese students in higher education after completing academic writing tasks using a ChatGPT4-embedded writing system developed by the research team. The study found that students expected AI to serve multiple roles, including multi-tasking writing assistant, virtual tutor, and digital peer to support multifaceted writing processes and performance. Students perceived that GenAI-assisted writing could benefit them in three areas including the writing process, performance, and their affective domain. Meanwhile, they also identified AI-related, student-related, and task-related challenges that were experienced during the GenAI-assisted writing activity. These findings contribute to a more nuanced understanding of GenAI's impact on academic writing that is inclusive of student perspectives, offering implications for educational AI design and instructional design.

GenAI 기반 학문적 글쓰기의 문제점에 대한 학생들의 평가

1. AI 관련 걸림돌 (AI-related):

- 환각 (Hallucination) (92%)
- 맥락적 이해 부족 (Lack of contextual understanding) (92%)
- 고차원적 사고 부족 (Lack of higher-order thinking) (87%)
- 인간 인식 부족 (Lack of human awareness) (76%)
- 언어적 문화 인식 부족 (Lack of cultural awareness in language) (64%)
- 관계 기술 부족 (Lack of relationship skills) (62%)
- 교육적 기술 부족 (Lack of pedagogical skills) (62%)
- 상호 운용성 부족 (Lack of interoperability) (62%)
- 설명 가능성 결여 (Absence of explainability) (58%)

GenAI 기반 학문적 글쓰기의 문제점에 대한 학생들의 평가

2. 학생 관련 걸림돌 (Students-related):

- AI 문해력 부족 (Lack of AI literacy) (98%)
- AI에 대한 부정적 태도 (Negative attitude toward AI) (84%)
- 고차원적 사고 부족 (Lack of higher-order thinking) (80%): 비판적 사고, 논리 구성 등
- 과제 주제 지식 부족 (Lack of task topic knowledge) (80%)
- 글쓰기 기술 부족 (Lack of writing skills) (76%)

3. 과제 관련 걸림돌 (Task-related):

- 시간 제약 (Time constraints) (89%)

AI 리터러시 프레임워크 (Warschauer, et al.)

AI 리터러시 프레임워크: 작문 교육에 AI 도구를 통합하는 다섯 단계

- "첫째, 학생들은 AI 작문 도구의 기능, 강점, 약점, 그리고 편향성에 대한 기본적인 이해를 가져야 한다. (Ng et al., 2021, "Know and Understand").
- 둘째, 학생들은 논문 작성, 이메일 작성, 또는 배경 정보 수집과 같은 특정 커뮤니케이션 과제를 수행하기 위해 AI 작문 도구에 접근하고 이를 탐색할 수 있어야 한다. (Ng et al., 2021, "Use and Apply").

AI 리터러시 프레임워크 (Warschauer, et al.)

- 셋째, 학생들은 자신의 작문 목표에 적합하고 유용한 콘텐츠를 생성하도록 AI를 전문적으로 프롬프트할 수 있어야 한다.
- 넷째, 학생들은 작문 및 수정 과정 전반에서 AI가 생성한 콘텐츠의 정확성을 확인해야 한다.
- 마지막으로, 학생들은 AI가 생성한 텍스트를 자신의 글에 윤리적이고 효과적으로 통합하는 방법을 배우고, 작문 과정에서 AI를 사용했음을 명시하고 인용해야 한다.

생성형 AI 작문 피드백 연구가 놓치고 있는 것

- 챗GPT의 작문 피드백이 학생들의 개인차를 충분히 반영하지 못 하지만 교사가 의도한 피드백 영역을 어느 정도 반영한다는 것(Evmenova et al., 2024)은 분명 의미있는 결과이다. 이를 고려한다면 교사는 교육적 목표에 정교하게 정렬된 (aligned) 피드백 생성을 위해 더 노력함과 동시에 학생 개개인에게 꼭 필요한 피드백을 주는 일은 기계에 온전히 맡길 수 없다는 점을 명심해야 한다.
- 하지만 여기에는 결정적인 변수가 빠져 있다. **대개 학생들은 피드백을 열심히 읽지 않는다.** 피드백을 읽어야만 하는 장치가 있거나, 교사의 피드백을 높은 수준으로 신뢰하거나, 교사와 학생 사이의 관계가 강한 유대로 엮여 있지 않다면 대부분 휘리릭 읽고 마는 것이다.
- 이런 종류의 연구에서 언급되지 않는 **또 다른 요인은 개개인의 집중력**이다. 피드백도 일종의 '읽기 자료'이고, 이것을 세심하게 읽어내기 위해서는 학생의 집중력이 요구된다. 하지만 많은 연구들은 특정 맥락에 처한 학생들이 어느 정도로 집중해서 피드백을 읽어낼 수 있는지 좀처럼 언급하지 않는다.

생성형 AI 작문 피드백 연구가 놓치고 있는 것

- 근본적으로 **AI를 활용한 작문교육이 이제 어떤 모습이 되어야 하는지에 관한 명확한 관점이 결여되어 있는 연구가 대부분**이다.
- 현재의 연구들은 AI 피드백의 정확성이나 효율성에만 주목할 뿐, 학생들이 실제로 경험하는 심리적 역동성과 그들의 학습 맥락이 지닌 복잡성을 충분히 포착하지 못하고 있다.
- 이 같은 진단은 단순히 연구의 미비함을 지적하는 것이 아니라, 우리가 AI와 인간의 협력적 관계를 어떻게 교육적으로 재구성할 것인가에 대한 본질적 질문을 제기한다. 피드백의 기술적 구현이나 효과성 측정을 넘어, 학생들이 자신의 글쓰기를 통해 세계를 이해하고 표현하는 과정에서 AI가 어떤 교육적 동반자가 될 수 있는지에 대한 깊이 있는 탐구가 필요한 시점이다.

이후 논의는 아래 세 문헌을 참고하여 작성되었습니다.

1. 김성우. (2024). *인공지능은 나의 읽기-쓰기를 어떻게 바꿀까. 지금 준비해야 할 문해력의 미래*. 유유.
2. 김성우. (2024). 보편적 학습 설계를 위한 인공지능 생성형 인공지능과 리터러시 교육의 상상력. *안과밖*, 56, 123-154.
3. 김성우. (2023). 생성형 AI 의 부상과 리터러시 생태계의 변동: 변화의 지형과 비판적 메타-리터러시의 가능성을 중심으로. *문화과학*, 114, 167-187.

생성형 인공지능의 부상과 리터러시 생태계의 변동

생성 인공지능과 리터러시 생태계의 변동

1. 리터러시 관행의 변화: 인공지능 매개 읽기 쓰기의 부상
2. 리터러시 실천에 있어서의 읽기와 쓰기의 전도 및 실시간 통합
3. 속도의 생태계의 변화, 그리고 생산성 담론의 부상
4. 저자성과 윤리의 새로운 논점들

1. 리터러시 관행의 변화: 인공지능 매개 읽기 쓰기의 부상

인공지능 매개 읽기의 지속적 증가

인공지능을 매개로 한 읽기의 지속적 증가가 예상된다. 거대언어모델의 자연어 처리 능력은 광범위하다. 텍스트 요약, 단순화, 상세 설명, 바꿔 쓰기 등이 대표적인 예이다. 텍스트가 담고 있는 정서를 감지하거나 키워드를 추출하는 작업도 가능하다. 텍스트 내에서의 다양한 변형, 텍스트 간의 비교, 대조, 통합, 추출, 텍스트를 구성하는 언어를 넘나드는 횡단적 읽기가 가능해지고 있는 것이다.

이는 읽기에 있어 양날의 검으로 작용할 것이다. 반면에 전통적인 텍스트 읽기의 방식 중 정독(精讀)이나 숙독(熟讀) 혹은 미독(味讀) 등 텍스트와 맥락, 자신의 생각과 감정을 세심히 살피는 독서는 그 감소세가 더욱 뚜렷해질 것이다.

요약의 효용과 한계를 인식하기

- “책을 시시하게 만드는 것이 ‘요약’이다. 요약이 필요 없다는 말이 아니다. 요약은 읽기와 공부에서 아주 중요하다. 내가 읽으며 알고자 하는 것의 포인트를 잡아주는 것이 요약이다. 요약은 일종의 이정표 역할을 한다. 요약 정리된 것을 포스트로 삼아 책 전체의 흐름을 파악하고 그 사이사이에 생략된 것을 떠올리며 재구성할 수 있기 때문이다. 따라서 어느 지점에서 읽은 것을 요약하는지에 따라 책은 전혀 다른 것으로 재구성된다. 책을 읽을 때 마다 밑줄 치고 요약한 내용이 달라지는 것이 이런 이유에서다.”
- “이런 요약은 읽기를 전제한다. 문제는 읽지 않고 요약하고, 요약된 것만을 보는 경우다. 읽기를 생략한 요약이 무엇보다 치명적인 이유는 요약된 것은 ‘핵심’이기 때문에 매리언 울프의 ‘다시, 책으로’에 따르면 “진실을 찾는 고된 훈련에 나서기도 전에 이미 진실을 안다고 생각”하게 만든다.” (엄기호)

2. 리터러시 실천에 있어서의 읽기와 쓰기의 전도 및 실시간 통합

전통적인 글쓰기 시나리오

- 전통적인 학술적 글쓰기에서 읽기와 쓰기는 유기적으로 연결된다. 교육의 맥락에서도 대부분의 쓰기 과제는 읽기를 기반으로 수행된다. 특정한 글을 읽고 비판적 코멘터리를 작성하거나, 여러 논문을 엮어 문헌 고찰을 제출하기도 한다.
- 생성 인공지능으로 글을 쓸 경우 이 순서는 전도되고 변형된다. 학부 과정 과제로 종종 부과되는 특정 주제에 대한 보고서 작성 시나리오를 상상해 보자.
- 종래의 글쓰기를 구성하는 ‘읽기->쓰기’의 방향성은 생성 인공지능 기반 글쓰기에서 ‘쓰기->읽기’로 변화한다. 전자에서 읽기의 대상은 주로 타인이 쓴 텍스트이며 쓰는 주체는 필자 자신이다. 하지만 후자에서는 ‘쓰기->읽기’의 방향성이 우세하며, 쓰기는 ‘프롬프팅+생성’으로, 읽기의 대상은 생성된 텍스트와 요약된 문헌이 주를 이룰 공산이 크다.

3. 속도의 생태계의 변화, 그리고 생산성 담론의 부상

- 리터러시 기반 과업의 속도가 빨라지고 생산성이 증가될 것이라는 예측 속에서 한 가지 반드시 짚고 넘어가야 할 것이 있다. 바로 인공지능의 리터러시 과업 수행과 인간의 리터러시 행위 속도 간의 비대칭이다.
- 그러나 이러한 가속화의 흐름 속에서도 여전히 '상수'로 남는 것은 인간이 글을 읽는 속도이다. 온전히 프롬프팅을 기반으로 글을 생산한다고 해도, 좋은 글을 쓰고자 하는 필자라면 각 단계에서 생산되는 응답을 읽는 일을 건너뛸 수는 없다.
- 측정 방법에 따라 또 개인의 읽기 능숙도에 따라 편차는 있으나, 한국어의 경우 분당 평균 200여 단어, 영어의 경우 논픽션 장르는 238 단어, 픽션 장르는 260단어 정도를 읽는다.
- (몸을 경유하지 않은) 생산은 엄청나게 빨라지고 있지만, 그것을 이해하는 몸의 속도는 변하지 않았다!

- 신중한 독자는 다른 세계를 상상하고, 저자가 인도해 준 여정을 따르면서 이탈하고, 행간을 자신의 경험과 의견, 감정과 예측으로 채우며, 필자의 주장에 반론을 펼치기도 한다.
- 글 읽기가 촉발하는 과정은 물성을 가진 텍스트로 외화되지 않으나 몸 안에서 엄연히 또 시끄럽게 진행된다.
- “인상적인 것을 중요한 것이라고 오해한다” - E. M. 포스터
- 아울러 외부로 만들어내는 것(output)만 생성적이라는 사고는 지극히 반-생성적이다. 읽기를 하는 동안에도 우리의 마음 속에는 온갖 심상과 상상, 복기와 예측, 공감과 갈등이 생성된다.

4. 저자성과 윤리의 새로운 논점

AI는 공저자가 될 수 있는가?

- 이러한 변화는 학술적 글쓰기의 관점에서 흥미롭고도 까다로운 문제로 이어진다. 챗GPT의 언어생산능력은 인간만이 글쓰기의 주체가 아닐 수 있음을 물리적으로 증명해 내고 있으며, 인간과 챗GPT가 함께하는 대화적 문제 해결은 공저(co-authoring) 활동의 핵심 요소이다.
- 이 두 가지 측면을 종합하면 결국 “챗GPT는 (공)저자가 될 수 있는가?”라는 질문, 즉, 저자성의 구성요건은 무엇인가라는 질문으로 응축된다.

국제의학학술지 편집인위원회의 저자 요건

1. 저작물의 고안 또는 설계, 또는 저작물을 위한 데이터의 획득, 분석 또는 해석에 대한 실질적인 기여
2. 저작물의 초안을 작성하거나 중요한 지적 내용에 대해 비판적으로 수정함
3. 출판될 버전에 대한 최종 승인
4. 저작물의 정확성 또는 무결성과 관련된 문제가 적절히 조사되고 해결될 수 있도록 저작물의 모든 측면에 대한 책임을 진다는 데 동의함.

- 기여의 영역에서 인공지능의 역할을 어느 정도 인정한다고 해도 두 번째 ‘윤리적인 책무’ 영역에서 보다 근본적인 문제에 봉착한다.
- 인공지능은 특정한 학술적 결과물에 대한 과학적, 사회적 책임을 질 수 있는가? 데이터와 프로토콜, 해석과 설명 등에 문제가 발견될 경우 이를 수정하고 필요한 경우 법적 책임을 질 수 있는가? 현재의 윤리적, 학술적, 제도적 테두리 안에서는 가능하지 않다.
- 결론적으로, 현재의 저자 인정 요건을 생성 인공지능에 적용했을 때 발생하는 여러 논점 중에서 저자의 윤리적 책무와 이로부터 파생되는 사회적이고 법적인 책임의 문제는 해소되지 않는다.

인공지능과 교수학습의 철학: AI for UDL (보편적 학습설계를 위한 인공지능)

보편적 학습설계(UDL)

- “**보편적 학습설계(UDL, Universal Design for Learning)**는 학습자들이 교실을 포함한 모든 환경을 인식하고 상호작용하는 방식에 있어서 개인 간에 엄청난 다양성이 존재한다는 전제 하에 학습자들이 학습목표를 성취할 수 있도록 교수자료와 활동을 다양하게 설계하는 것을 의미한다. 보편적 학습설계(UDL)는 **보편적 설계(UD)에서 지향하는 유연성의 개념을 물리적인 시설 뿐 아니라 교육과정 및 교수-학습 방법으로 확대 적용하고자 하는 접근**으로 교육과정을 설계 하면서 모든 학생들이 배움의 기회를 가질 수 있도록 형평성을 구현하는 것을 의미한다.
- 즉 교수학습의 목표, 방법, 평가 등을 설계할 때 학습자들이 학습자료를 이해하는 데 방해가 되는 모든 장애물을 제거함으로써 모든 학습자들의 학습을 촉진하여 개별화 학습이 가능하고 학습자들에게 최상의 학습환경을 제공하는 것을 기본개념으로 한다(Bray & McClaskey, 2017).”
(조윤정 등, 2019)

보편적 학습설계(UDL)

- UDL을 기반으로 실제 수업을 디자인하기 위해 참조해야 할 기준으로 가장 널리 활용되는 것은 **CAST의 보편적 학습설계 가이드라인(Ui Guideline)**이다. 이에 따르면 UDL은 다양한 방법으로 학습자들의 **참여**를 가능하게 해야 하고, 여러 방식으로 지식과 경험을 **전달**해야 하며, 학생들이 복수의 방식으로 **행동하고 표현**할 수 있는 토대를 제공해야 한다.
- 이는 학습자에게 무제한의 자유를 부여하거나 수업 상황을 방관하는 것과는 전혀 다르다. UDL을 성공적으로 실행하기 위해서는 교수자가 학습 내용뿐 아니라 학습 환경 및 상호작용에 대한 전문성을 갖추어야 하며, 무엇보다 학습자에 대한 풍부한 이해를 위한 체계적 접근법을 체화해야 한다.

Provide multiple means of
Engagement



Affective Networks
The "WHY" of Learning

Provide multiple means of
Representation



Recognition Networks
The "WHAT" of Learning

Provide multiple means of
Action & Expression



Strategic Networks
The "HOW" of Learning

Access	Provide options for Recruiting Interest	Provide options for Perception	Provide options for Physical Action
Build	Provide options for Sustaining Effort & Persistence	Provide options for Language & Symbols	Provide options for Expression & Communication
Internalize	Provide options for Self Regulation	Provide options for Comprehension	Provide options for Executive Functions
Goal	Expert learners who are...		
	Purposeful & Motivated	Resourceful & Knowledgable	Strategic & Goal-Directed

UDL 가이드라인에 따른 시활용 활동 예시

UDL 가이드라인	디자인 활동	내용
Provide options for Perception	다양한 미디어 제공 및 미디어 간 변환	텍스트, 이미지, 비디오 간 변환 및 이들의 병치를 통하여 학습자료 인지 수준 제고
Provide options for Language & Symbols	기존 가독성 지수를 활용한 텍스트 수준 조정	문장과 문단의 길이, 어휘 수준, 문법 복잡도 등의 조정
Provide options for recruiting interest	장르 변환을 활용한 BTS+Shakespeare 콜라보 수업	학생의 생활세계에서 시작하여 문학교육의 목표로 나아가기
Provide options for expression and communication	다른 시각으로 텍스트 재구성하기	대상 텍스트를 다른 관점에서 재구성하여 비판적으로 이해하기
<p>-> 비판적 메타 리터러시(critical meta-literacy)의 함양을 통해 시의 외부를 상상하고 내부의 '빈틈'을 간파하기</p>		

예시: 이미지에서 텍스트 산출 (ChatGPT)



Can you explain this image in detail?

Certainly!

The image depicts a picturesque landscape, likely during late afternoon or early evening, given the warm hues and lighting. Here are some prominent features:

- 1. Sky:** The sky occupies a significant portion of the image. It is filled with scattered cumulus clouds that have a tinge of gold, indicating either a setting or rising sun. The clouds create a dramatic effect against the blue backdrop of the sky.
- 2. Water:** A calm body of water is seen in the foreground. Its surface reflects the sky above, along with the tree and other elements from its surroundings.
- 3. Lone Tree:** Dominating the water is a solitary, lush green tree that seems to be growing on a small piece of land or an island. The tree is dense, tall, and appears healthy.

가독성 지수를 활용한 텍스트 난이도 조정

- **Lexile700:** "Schools need to change how they teach because things are different now. We have more technology, fewer young people, and more older people. Old ways of teaching just gave facts, but we need new ways that help students think better, work well with others, and stay active."
- **Lexile 1900:** "Given the confluence of digital evolution, demographic shifts characterized by low fertility and an aging populace, there's an exigent need for an educational renaissance. The antiquated paradigm, which predominantly focused on a unidirectional transfer of knowledge, proves inadequate. Contemporary pedagogical approaches should holistically cultivate cognitive proficiencies—ranging from ... (이하 생략)"

BTS+Shakespeare 콜라보 수업 (장르변환)

As in the looking glass my gaze doth meet,
Thy heart shall melt in twain, so true and clear,
For I possess a glow, a star most sweet,
And in my presence, thou shalt boogie near.

Shakespearean Sonnet (1st)	Original Lyric (1st)
In smoothness like butter, I, a thief concealed,	Smooth like butter, like a criminal undercover
Doth make thy heart to break, as trouble's kin.	Gon' pop like trouble breaking into your heart like that, ooh

다른 시각에서 텍스트 재구성하기

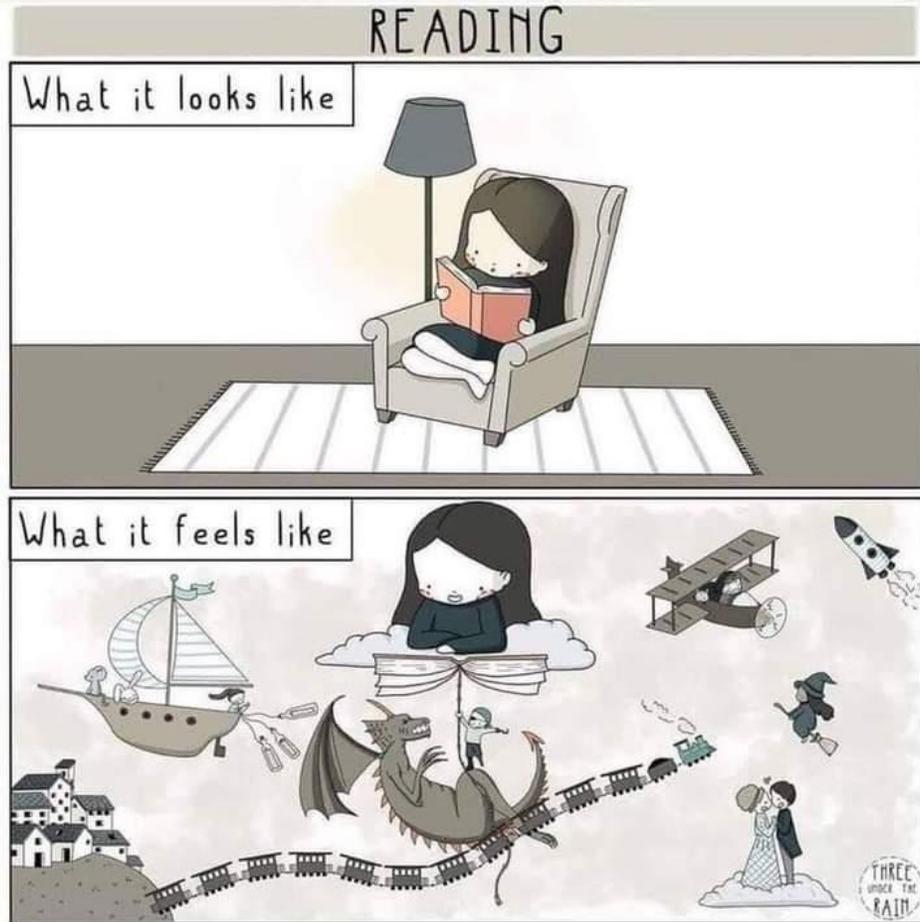
- Declaration of Independence의 첫부분을 주고 “Can you rewrite the following paragraph from the critical standpoint of black slaves at the time? You may want to incorporate their tragic experiences into the new writing.”라는 프롬프트를 제시.
- 생명, 자유, 행복 추구라는 양도할 수 없는 권리에 대한 선언은 인간의 가장 기본적인 존엄성마저 부정당한 사람들의 귀에 공허하게 울립니다. 국민의 상당수가 족쇄에 묶여 침묵하고 있는데 어떻게 정부가 피지배자의 동의로부터 권력을 얻었다고 주장할 수 있을까요? 억압적인 정부가 우리에게 가하는 바로 그 억압을 외면하면서 '국민이 정부를 변경하거나 폐지할 권리'에 대해 말하는 대담함은 씩씩한 아이러니입니다. 노예들은 '가법고 일시적인 원인'이 아니라 잔인함과 비인간화의 끊임없는 공세를 견뎌왔습니다. (번역한 것. 이하 생략)

**결론: 생성(becoming) 없는 생성
(generation)을 경계하기**

생성의 속도에 속지 말아야 할 이유

- 리터러시 기반 과업의 속도가 빨라지고 생산성이 증가될 것이라는 예측 속에서 한 가지 반드시 짚고 넘어가야 할 것이 있다. **바로 인공지능의 리터러시 과업 수행과 인간의 리터러시 행위 속도 간의 비대칭이다.** 다양한 인공지능 알고리즘은 대량의 텍스트를 단시간 내에 생산하는 역량을 지니고 있고, 다양한 텍스트를 변형하거나 요약할 수 있다.
- 그러나 이러한 가속화의 흐름 속에서도 **여전히 '상수'로 남는 것은 인간이 글을 읽는 속도이다.** 온전히 프롬프팅을 기반으로 글을 생산한다고 해도, 좋은 글을 쓰고자 하는 필자라면 각 단계에서 생산되는 응답을 읽는 일을 건너뛸 수는 없다. 측정 방법에 따라 또 개인의 읽기 능숙도에 따라 편차는 있으나, 한국어의 경우 분당 평균 200여 단어, 영어의 경우 논픽션 장르는 238 단어, 픽션 장르는 260단어 정도를 읽는다.

읽기도 생성적이라는 것을 기억하기



“이러한 의미에서 쓰기만이 아니라 읽기도 생성적입니다. 책을 읽으면서 생각하고 고민하고 상상합니다. 웃고 울며 감동하고 절망합니다. 그렇게 저자가 펼쳐놓은 세계와 교감하며 역동적인 경험을 하기에 자신도 모르게 변화합니다. 이렇듯 읽기의 경험을 통과하며 우리는 이전과는 다른 지식과 감정, 열망과 기억을 머금은 사람이 됩니다. 나아가 읽기를 통한 생성은 쓰기를 통한 생성의 탄탄한 기반이 됩니다. **텍스트를 비롯한 미디어를 세상에 내놓는 것만이 생성적이라고 생각하는 것만큼 반생성적인 사고도 없습니다.**”

‘생성 없는 생성’을 경계하기

생성형 AI (Generative AI)의 생성(Generation)

- 데이터를 바탕으로 새로운 콘텐츠를 만들어내는 능력. ‘만들기’
- 텍스트, 이미지, 음악, 영상 등을 빠르게 생성. 기존 AI와 달리 창작물을 직접 생산. (예시: GPT 모델의 텍스트 생성, DALL-E의 이미지 생성 등)

철학적 의미의 생성(Becoming)

- 새로운 존재가 되는 과정, 내적 변화의 경험. ‘되기’
- 글쓰기 과정에서 자연스럽게 발생하는 변화, 주제 고민, 자료 탐색, 문장 구성 과정에서 일어나는 사고의 확장

‘생성 없는 생성’을 경계하기

“역사상 읽히지 않는 글이, 청취되지 않는 오디오가, 시청되지 않는 비디오가 가장 많이 생산되는 지금, 우리가 더 나은 존재로 변화하고 있는지 엄중히 물어야 합니다. 대규모의 생태계 파괴와 저임금 노동 착취 위에 지어진 생성형 인공지능은 지금 이 순간에도 디지털 쓰레기를 양산하고 있습니다. 가치 없는 미디어의 대량생산은 행성을 파괴하는 재앙입니다.”

-> ‘인공지능의 시대’를 선언하고 인공지능을 중심으로 리터러시 생태계를 변화시키려하기 보다, **깊게 읽고 정성껏 쓰고 마음을 다해 소통하는 일의** 가치와 아름다움을 다시 되돌아보아야 할 이유가 여기 있습니다.

감사합니다.

