생성형 AI
활용 가이드라인

2024년 5월(ver 1.0)

연세대학교
연구처 연구윤리센터

생성형 AI 활용 가이드라인

생성형 AI는 예술, 음악, 소프트웨어 코드, 글쓰기 등 새롭고 독창적인 콘텐츠를 만들 수 있는 인공지능의 한 분야입니다. 사용자가 프롬프트를 입력하면 인터넷의 기존 사례에서 학습한 내용을 기반으로 인공지능 기술의 한 종류로서 이미지, 비디오, 오디오, 텍스트 등을 포함한 대량의 데이터를 학습하여 사람과 유사한 방식으로 문맥과 의미를 이해하고 새로운 데이터를 자동으로 생성해 주는 기술을 의미합니다.

최근 주요 생성형 AI 서비스로는 텍스트 생성형 AI(ChatGPT, Bard, Bing 등), 이미지 생성형 AI(DALL-E, Stable Diffusion, Midjourney 등), 음성 생성형 AI(클로바더빙, VoiceBox, Typecast 등), 영상 생성형 AI(VCAT, Runway, Invideo 등), 글쓰기 AI(Magic Write, Notion AI 등) 등 다양하게 활용되고 있습니다.

생성형 AI는 빠른 데이터 생성, 다양한 응용 가능성, 사용자 맞춤형 콘텐츠 제공 등을 제공하면서 그 기술은 끊임없이 발전하고 있으며, 대학 교육에 생성형 AI 활용은 교육 및 연구에 변화를 가져오고 있습니다.

첫째, 질문에 대한 답변을 생성하거나 텍스트 기반의 대화를 통해 교육 및 연구자료를 빠르게 수집하고, 다양한 콘텐츠를 생성할 수 있습니다.

둘째, 문서 요약, 문서 번역, 음성 합성 등을 다양한 언어 활용을 할 수 있습니다.

셋째, 시, 소설, 음악, 이미지 등 작품 및 컴퓨터 활용과 프로그램 코드를 생성하여 창조적인 일을 수행할 수 있습니다.

넷째, 창의적인 작업을 시작하는 시점에서 어떻게 시작할 수 있을지, 관련된 기존의 자료들은 무엇이 있는지 등을 제공함으로써 창의적 아이디어를 도출할 수 있습니다.

하지만, 생성형 AI 활용의 단점으로는 부정확하고 편향된 정보 제공, 저작권 침해, 개인정보 유출, 과도한 의존에 의한 사고력 및 역량 저하 및 표절이나 부정행위와 같은 학문적 진실성 위반 문제 등을 발생시킬 수 있습니다.

특히, 생성형 AI 활용 시 거짓 정보를 그럴듯하게 유포하는 '환각(hallucination)' 및 학습한 시점에서의 제한된 정보만 활용하고 있다는 점을 명심하십시오.

생성형 AI 활용 가이드라인

따라서, 생성형 AI 활용 시에는 다음의 사항을 고려해야 합니다.

생성형 AI에 대한 활용 방안

1. **비판적 활용:** AI가 학습한 데이터에 편향성이나 오류가 존재할 경우 잘못된 답변을 내놓을 수 있기 때문에 그 결과물을 맹목적으로 신뢰할 수 없고, 비판적으로 사실확인을 한다.
2. **상호합의적 활용:** 생성형 AI 활용 원칙에 대해 교수자와 학습자 간의 합의 후 생성형 AI를 사용해야 한다. 교수자가 생성형 AI 활용을 허용한다면, 교수자는 학습자가 최소한의 규칙을 잘 따를 수 있도록 지도하며, 학습자는 생성형 AI 활용 시 교수자가 제시한 지침을 숙지하고, 교수자의 가이드를 따른다.
3. **창의적 활용:** 생성형 AI는 학습 대체자가 아닌 학습을 도와주는 도구임을 인식하며, 생성형 AI 답변을 그대로 받아들이기 보다는 이를 재해석하고 변형하여 새로운 지식을 창조하여 더 나은 결론을 만들 수 있어야 한다.
4. **윤리적 활용:** 생성형 AI 정보를 그대로 활용하는 것은 표절행위일 수 있으므로 출처를 정확히 표기하고 부정행위를 피한다. 또한, 생성형 AI를 사용하는 과정에서 사이버 공격이나 인종이나 지역에 대한 편견 등 사회적 문제가 발생할 수 있으므로 사회의 이로운 방향으로 사용한다.
5. **보안적 활용:** 생성형 AI는 개인정보 및 기밀정보 등의 보안이 취약함을 인지하고, 이와 관련된 정보를 입력하지 않는다.
6. **개방적 활용:** 생성형 AI는 자연어처리를 자동화하거나 글쓰기를 지원하고, 새로운 아니디어를 제시하는 등 사용자의 업무에서 효율성과 생산성을 높일 수 있다는 장점이 있으므로 새로운 학습 도구임을 받아들이고, 빠르고 적절하게 대응해야 하는 대상임을 인식한다.

생성형 AI 활용 시 이용자가 스스로 판단과 검토를 통해 책임있는 윤리의식을 갖고, 비판적인 사고로 올바르게 사용한다면 부정적 기능은 최소화하면서, 긍정적이고 생산적으로 활용할 수 있을 것입니다.

이에 대학 내 교육 및 연구 수행 시 생성형 AI를 윤리적이고 책임있게 활용하기 위한 보편적인 가이드라인을 제시하고자 합니다.

※ 본 가이드라인은 연구재단 등 정부기관 및 학술단체 등에서 연구자를 위한 가이드라인이 마련되거나, 생성형 AI 발전 동향에 따라 활용 방안이 변경될 수 있으므로 지속적으로 업데이트 예정입니다.

Ⅱ. 학습자를 위한 생성형 AI 활용 방안

1. 수업 정책을 확인하고 준수하십시오.

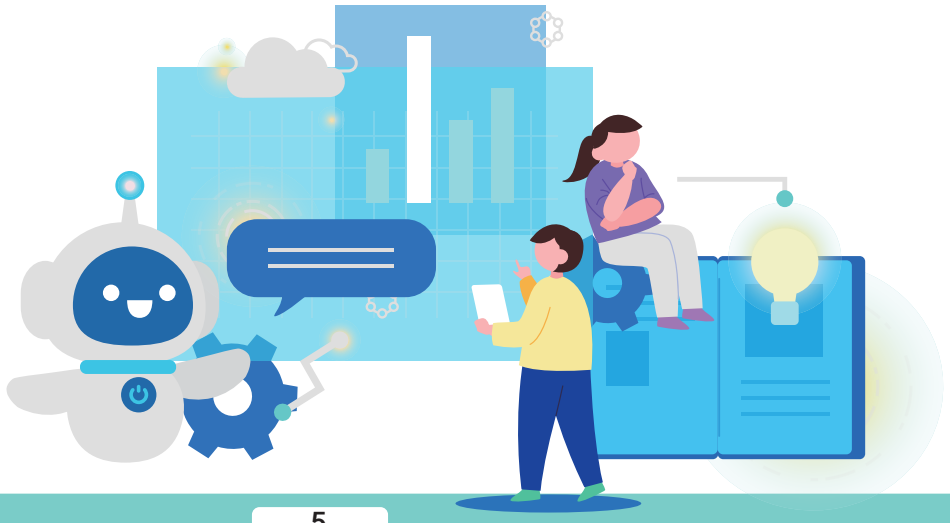
- 수업에 따라 생성형 AI 활용이 허용되거나 금지될 수 있습니다. <강의계획서> 및 수업 정책에 따라 활용여부를 확인하십시오. 생성형 AI를 활용하는 것이 수업에 따라 부정행위로 처리될 수 있습니다.
- 생성형 AI 활용 시 교수자가 제시한 지침을 숙지하고, 가이드를 따르십시오.
- 성적 평가를 위한 과제로 제출하여 발생하는 모든 문제에 대한 책임은 학생 본인에게 있음을 숙지하여 주십시오.

2. 생성형 AI를 활용한다면 사실 여부를 확인하십시오.

- 생성형 AI는 그 특성상 답변이 거짓 정보를 생성할 수 있으므로 항상 '검증과 확인'이 필요합니다.
- 생성형 AI의 단점과 한계를 알고, 결과물에 의심하고 비판적 사고를 가지십시오.
- AI를 활용할 때 학습 목표를 항상 염두에 두고 자율적인 학습과 탐구를 중요시하며, AI에 지나치게 의존하지 않도록 주의하십시오.

3. 생성형 AI를 활용한다면 윤리성, 편향성에 주의하십시오.

- 생성형 AI는 비 윤리적인 질문에 답변을 거부하도록 훈련되었으나, 우회적 질문으로 비윤리적으로 활용할 수 있는 결과물 제시가 가능하므로 주의하십시오.
- 편향적이거나, 차별적 데이터 학습 시 편향적 답변이 도출되어 연구에 반영될 수 있으므로 주의하십시오.



Ⅱ. 학습자를 위한 생성형 AI 활용 방안

4. 생성형 AI를 활용한다면 자료 및 출처를 한번 더 확인하십시오.

- 생성형 AI를 활용할 경우, 다른 검색을 통해 한번 더 확인해 보십시오.

5. 생성형 AI를 활용한다면 출처 표기를 하십시오.

- 생성형 AI를 학습에 활용한다면 출처를 정확하게 표시하고, 산출물을 가공하지 않고 그대로 붙여 활용하는 것은 표절 등 부정행위로 간주될 수 있으므로 주의하십시오.
- 생성형 AI 출처표시 방법은 생성 AI 모델, 프롬프트 내용, 생성 AI 플랫폼 등을 작성하는 방법이 있으며, 과제 제출 시 교수자의 가이드를 따라 주십시오.

예시 1	ChatGPT3.5(2024.5.1.). "프롬프트 내용". https://chat.openai.com
예시 2	**(APA 스타일)** OpenAI. (Year). ChatGPT (Month Day version) [Large language model]. https://chat.openai.com
예시 3	**(MLA 스타일)** "Exact prompt you used" prompt. ChatGPT, Day Month version, OpenAI, Day Month Year, chat.openai.com.

<출처: Guide for referencing & acknowledging the use of artificial intelligence tools>

6. 생성형 AI 자료 활용시 다음의 사항을 확인하여 주십시오.

- 생성형 AI가 저작권자의 사용허가 없이 인터넷 기사, 웹사이트 게시글 등을 학습용 데이터로 이용하는 경우 저작권 문제가 발생할 수 있습니다.
- 생성형 AI가 학습하는 정보에 개인정보(인터넷상 공개된 개인정보)가 포함되어 개인정보 보호 규정에 문제가 발생할 수 있습니다.
- 생성형 AI가 생성한 코드는 에러나 보안상 취약점이 있을 수 있으므로 주의하십시오.

<첨부자료>

생성형 AI를 현명하게 활용하기 위한 체크리스트

1 저작권
생성형 AI의 결과물을 활용할 때 생성형 AI를 활용해서 얻은 결과물이라고 출처를 표기했나요? 네 ☐ 아니오 ☐

2 권리침해
생성형 AI를 활용할 때 타인의 권리가 침해될 수 있는 텍스트, 오디오, 이미지 등을 사용하지 않았나요? 네 ☐ 아니오 ☐

3 명예회손
생성형 AI에 질문이나 정보를 입력할 때 특정인의 명예를 훼손하거나, 차별하는 내용이 포함되어 있지는 않나요? 네 ☐ 아니오 ☐

4 혐오표현
생성형 AI가 제시한 정보에 개인, 기관 등 특정 대상을 비난하거나, 가치관이나 주장을 일방적으로 혐오하는 내용이 포함되어 있지 않나요? 네 ☐ 아니오 ☐

5 정보유출
생성형 AI로 정보를 얻거나 콘텐츠를 제작하기 위해 개인정보, 기업기밀 등 민감한 정보를 제공하지는 않았나요? 네 ☐ 아니오 ☐

6 허위조작정보
생성형 AI로 가짜뉴스, 스팸 등을 만들기 위해 사실이 아닌 부정확한 정보나 조작된 내용을 일부러 입력하지는 않았나요? 네 ☐ 아니오 ☐

7 정보편향
생성형 AI가 결과로 제시한 정보에 한쪽으로 치우친 편향적인 내용이 없는지 확인하였나요? 네 ☐ 아니오 ☐

8 환각현상
생성형 AI가 제공한 정보가 모두 정답은 아니라는 생각을 하며 잘못된 정보가 있는지 사실 확인을 위해 교차검증을 했나요? 네 ☐ 아니오 ☐

9 오남용
생성형 AI가 주는 편리함에만 의존하지 않고 먼저 충분히 생각하고 고민한 후에 생성형 AI는 보조적 수단으로 활용하였나요? 네 ☐ 아니오 ☐

10 창의성
생성형 AI가 제시한 결과를 그대로 사용하지 않고, 재해석하거나 자신의 생각과 아이디어를 덧붙여 생산적으로 활용하였나요? 네 ☐ 아니오 ☐

<출처: 생성형 AI 윤리 가이드북. 한국정보사회진흥원. 2023>

<첨부자료>

챗GPT 등 생성형 AI 활용 보안 가이드라인

윤리적 사용	· 서비스 사용 시 윤리적 가치 고려 및 상호 존중
데이터 처리방침 준수	· 개인정보 및 민감 정보 관련 법률 및 규정 준수 · 적절한 데이터 보호 기술 사용
인공지능 생성 콘텐츠 위험 인지	· 사실여부 확인 및 객관적 분석을 통한 비판적 콘텐츠 수용 · 감정적 자극 유도 및 소셜 미디어 콘텐츠에 대한 비판적 자세 견지
데이터 제어 설정	· 데이터 제어 설정을 통한 채팅 기록 및 모델 학습 비활성화
개인정보 입력 금지	· 개인정보 침해 및 악용 방지를 위한 민감한 개인정보(건강, 종교, 정치적 성향 등) 입력 금지 · 금융적 손실 및 신용도 피해 방지를 위한 금융 관련 정보(신용카드 번호, 은행계좌 정보, 비밀번호 등) 입력 금지
기관 내 민감 정보 입력 금지	· 기관 내부자료 등 민감정보 입력 금지로 기관 정보 유출 차단 · 가명화 및 익명화를 통한 실제 개인정보와 기관과의 관계 유추 가능성 차단
보안 질문 회피	· AI 모델과 대화 중 보안 관련 질문 발생 시 답변 자제 · 필요시 고객 지원 센터 문의를 통한 올바른 절차 준수
인증 정보 입력 금지	· 계정 침해 위협 방지를 위한 인증 정보 (사용자 이름, 비밀번호, 인증코드 등) 입력 금지

챗GPT 등 생성형 AI 활용 보안 수칙

01 **민감한 정보**(非공개 정보, 개인정보 등) **입력 금지**
* 설정에서 「대화 이력 & 학습」 기능 非활성화

02 생성물에 대한 **정확성·윤리성·적합성** 등 再검증

03 가짜뉴스 유포·불법물 제작·해킹 등 **범죄에 악용금지**

04 생성물 활용 시 지적 재산권·저작권 등 **법률 침해·위반 여부 확인**

05 악의적으로 거짓 정보를 입력·학습 유도하는 등 **非윤리적 활용 금지**

06 연계·확장프로그램 사용 시 보안 취약여부 등 **안전성 확인**

07 로그인 계정에 대한 **보안설정 강화 및 보안관리 철저**
*'다중 인증(Multi-Factor Authentication)' 설정 등

<출처: 챗GPT 등 생성형 AI 활용 보안 가이드라인, 국가정보원>

<첨부자료>

생성형 AI 활용에 대한 국제 학술단체의 입장(예시)

학술단체	생성형 AI 자료 사용
국제학술지 출판윤리위원회(COPE)	· 연구 출판에서 ChatGPT 또는 거대언어모델(LLM)과 같은 AI tool의 사용이 급속히 늘어나고 있다. · 논문의 원고 작성, 이미지 또는 그래픽 요소 제작, 데이터 수집 및 분석 등에서 AI 도구를 사용한 저자는 논문의 재료 및 방법론(또는 유사한 섹션)에서 어떤 AI 도구가 어떻게 사용되었는지를 명확하게 공개해야 한다. · 저자들은 AI tool에 의해 산출된 부분을 포함하여 그들의 논문 원고 내용에 대해 전적으로 책임을 져야 한다. 따라서 저자들은 어떠한 출판윤리 위반에 대해서도 책임을 져야 한다.
국제의학학술지 편집인위원회(ICMJE)	· 논문 투고 단계에서 학술지는 저자들이 제출한 연구결과의 산출에 AI 지원 기술(예시:LLMs, 챗봇, 이미지 생성기)을 활용했는지 여부를 공개하도록 요구해야 한다. AI기술을 활용한 저자들은 그 기술을 어떻게 사용하였는지를 표지서신(Cover letter)과 제출원고(submitted work)에 기술해야 한다. · AI는 부정확하고(incorrect), 불완전하고(incomplete), 편향적이면서(biased)도 권위가 있는 것처럼 보이는 결과물을 생성할 수 있으므로 저자들은 결과물을 주의 깊게 검토하고 편집해야 한다. · 전체 인용(full citations)을 포함하여 모든 인용된 자료에 대해 적절한 저작자표시(appropriate attribution)가 있었는지를 확인하는 것은 인간이 담당해야 한다.
Nature誌	· LLM tool을 사용하는 연구자는 연구방법(methods) 또는 사사(acknowledgements)를 표기하는 섹션에 해당 사실을 적시해야 한다.
Science誌	· AI, 기계학습 또는 유사한 알고리즘 도구를 활용하여 생성된 텍스트는 편집자(editor)의 명시적인 허락이 없는 상태에서 Science 저널의 논문에서 사용할 수 없다. 그와 같은 도구를 사용하여 만든 그림, 이미지, 그래픽도 마찬가지이다.

<출처: 연구개발에서 인공지능 도구의 활용과 관련된 연구윤리 이슈 분석. 한국연구재단>

(출처자료 발췌 정리)

<참고자료>

1. 생성형 AI 윤리 가이드북. 한국정보사회진흥원. 2023
2. 챗GPT 활용방법 및 주의사항 안내. 행정안전부. 2023
3. 챗GPT 등 생성형 AI 활용 보안 가이드라인. 국가정보원. 2023
4. 연구개발에서 인공지능 도구의 활용과 관련된 연구윤리 이슈 분석. 한국연구재단. 2023
5. 주요 국가의 인공지능(AI) 관련 연구윤리 정책 동향 조사. 한국연구재단. 2020
6. 생성형 AI 활용 가이드. UNIST. 2023
7. Guide for referencing & acknowledging the use of artificial intelligence tools. University of New South Wales. 2023