

국가연구개발 AI 연구윤리 가이드

2026. 9.



과학기술정보통신부



목 차



I . 개요.....	1
II . AI 활용의 윤리적 위험과 원칙.....	2
III . AI 활용 권고사항.....	4
IV . AI 활용 서식(예시).....	7

I. 개요

□ 목적

- 국가R&D 수행의 모든 단계에서 연구자의 **윤리적인 AI 활용**을 지원하는 가이드를 개발하여 국가연구개발의 연구 진실성 제고
 - 본 가이드는 AI 활용을 규제하기 위한 것이 아니며, 연구개발 과정에서 연구자의 윤리적인 AI 활용을 지원하기 위해 작성

□ 배경

- 연구자의 AI 활용 범위가 지속적으로 확대되고 있으며, 그 영향은 개별 연구뿐만 아니라 국가연구개발 전반에 미치고 있음

와일리(Wiley) 조사 (2025)	엘스비어 조사 (2025)
· 연구자의 생성형 AI 사용경험* 증가(전년도 57%→84%) * 업무의 어떤 면으로든 활용하는 비율 · 연구출판과 직접 연관된 일에 활용(전년도 45%→62%) ※ 연구 특화 모델보다는 주로 범용적 챗봇(ChatGPT 등)을 활용(80%)	· 업무에 AI를 활용하는 연구자 (전년도 37% → 58%) ※ 연구 결과의 요약(61%), 문헌 검토(51%), 과제 제안서 초안(41%), 연구 데이터 분석(38%), 연구 논문(보고서) 초안 작성(38%)에 AI를 활용

- 새로운 유형의 연구윤리 우려가 증가함에 따라 **국가연구개발의 연구진실성 확보를 위한 제도적 대응 필요**

※ 연구자의 은닉 프롬프트를 활용한 학술 리뷰 조작 등 AI를 활용한 신종 연구부정 행위에 대응하기 위해, 국가 차원의 가이드라인 정립이 필요

□ 권고 적용 대상 및 모델

- (적용 대상) 국가연구개발에서 **AI를 활용하는 모든 연구자를 주 대상**으로 하며, 연구자 지원을 위한 **연구개발기관의 역할도 일부 제시**
- (적용 모델) 권고사항은 **생성형 AI(LLM*) 활용을 기준으로** 하되 자체 개발·학습시킨 AI 모델이나 에이전틱 AI 등도 적용 가능한 범위 내에서 포함
 - * LLM(Large Language Model, 대규모 언어 모델)은 방대한 데이터를 학습해 사람처럼 글을 읽고 쓰고 대화할 수 있는 초대형 인공지능 모델

Ⅱ. AI 활용의 윤리적 위험과 원칙

□ AI 활용의 윤리적 위험

① 연구 진실성 훼손

- AI는 검증되지 않은 가설이나 허위 정보를 포함해 부정확하거나 잘못된 내용을 생성할 위험이 있음
- AI를 활용한 아이디어 탐색 및 자료 정리·분석 과정에서 기존 연구 동향 왜곡, 데이터 누락·오류·중복이 발생할 우려가 있음

② 편향 재생산

- AI는 기존 자료와 학습 데이터에 기반해 과거의 한계와 왜곡을 그대로 답습하고 증폭시키는 경향이 있음
- 연구 자료 수집·분석 시 객관성과 타당성을 저해할 수 있고, 특정 연구 주제나 방향성만을 강조하여 연구 다양성을 훼손할 수 있음

③ 보안사항 유출

- 미출원 기술이나 기술이전 협상 중인 기술 등 비공개·보안 사항을 외부 AI에 입력할 경우, 연구 기밀이 유출될 위험이 있음
- 외부 AI에 입력된 데이터는 시스템 학습 등에 활용되어 기관의 지식재산권과 국가 주요 기술의 보안성을 심각하게 훼손할 수 있음

④ 개인정보 유출

- 개인정보, 의료 데이터 등 민감한 정보를 외부 AI에 입력하는 경우 해당 정보가 외부 서버에 저장되거나 유출될 가능성이 높음
- ※ 글로벌 AI 서비스의 경우, 업로드된 자료가 해외 데이터센터나 해외 법인으로 이전되는 등 개인정보의 국외 이전 위험이 높음

⑤ 부적절한 연구행위

- AI 생성물(이미지, 도표 등)은 학습 데이터를 기반으로 하므로, 정확한 출처를 명시하지 않을 경우 표절 및 부적절한 인용 등의 문제 소지가 있음
 - AI가 생성한 코드는 기존 개발자들이 짜놓은 코드를 복제한 것일 가능성이 있으므로, 검증 없이 활용 시 라이선스 분쟁의 소지가 있음

□ AI 활용 기본원칙

<국가연구개발 AI 활용 기본원칙>

국가연구개발에서 AI의 역할은 도구로 제한되며, 최종 결과물에 대한 책임과 권리는 AI를 활용한 연구자에게 귀속됨

- (보조 도구) AI는 문헌 조사 요약, 번역, 문법 교정, 기초 코딩, 데이터 정리, 시뮬레이션 등 연구의 효율성을 높이는 보조적 도구 역할에 국한되어야 함
- (저자성 인정불가) AI는 법적·도덕적 의무를 질 수 없으므로 저자 및 발명자로서의 지위·권리를 가질 수 없음을 원칙으로 함

□ 연구자와 연구개발기관의 책임

- 연구자와 연구개발기관은 국가연구개발 전 과정에서 AI를 윤리적으로 활용하기 위해 각자의 역할과 책임을 다해야 함
 - (연구자) AI 활용의 기술적·윤리적 한계를 명확히 인식하고, 주체적 판단 역량을 함양하여 AI를 책임 있게 활용
 - (연구개발기관) 연구자의 윤리적 AI 활용을 위한 제도적 지원체계 (지침·규정·매뉴얼) 구축 및 윤리적 위험에 대한 예방·대응 체계 마련

Ⅲ. AI 활용 권고사항

AI는 연구 기회를 확대하는 동시에 윤리적 위험 요인을 내포하고 있으므로, 위험 요인에 대한 주의와 기본원칙에 기반한 권고사항 준수를 통해 책임 있는 AI 활용을 도모

□ 국가연구개발 AI 활용 권고사항

- ① (사실관계 검증) 연구자는 AI 생성물의 오류, 핵심 자료의 누락, 출처의 존재 및 진위 등을 검증해야 함
 - AI의 환각(hallucination)에 기인한 연구부정 및 부적절한 연구행위를 방지하고 연구 진실성을 확보하고자 노력해야 함
- ② (주체적인 판단) 연구자는 AI가 생성한 자료를 활용 시 논리적 타당성과 전문 지식에 따라 검토하고, 자신의 관점을 반영하여 재구성해야 함
 - 편향으로 인한 연구 결과의 왜곡을 예방하며 연구개발 결과물에 최종 책임을 지는 연구자의 필수 의무에 해당
- ③ (투명성 확보) 연구자는 연구 과정 전반에서 AI를 활용한 경우 모델·버전, 목적·범위* 등을 최종 성과물 제출(또는 발표) 시 투명하게 공개해야 함
 - * 정보 수집, 데이터를 통한 추론, 연구데이터 정리, 문법·오타 점검, 번역 등
 - AI 활용 여부를 공개하는 것은 연구 결과의 신뢰성을 높이고, 연구자가 최종 책임자임을 분명히 하는 데 목적이 있음
- ④ (결과 신뢰성 관리) 연구자는 AI를 활용하더라도 연구개발 성과의 구현 과정을 논리적으로 설명할 수 있어야 함
 - 국가연구개발에서 결과에 대한 신뢰는 원활한 후속 연구 진행, 기술이전 등 성과 활용, 연구부정행위 예방 등 연구 진실성 확보를 위한 필수 요소임

- ⑤ (정책·규정의 준수) 연구자는 AI 활용 시 관련 법령뿐만 아니라 연구개발기관의 규정이나 지침, 가이드 등을 확인하고 준수해야 함
- 논문 출판의 경우 학술지(출판사)별로 AI 활용 정책이 다르므로, 투고 전 관련 정책을 점검하고 준수해야 함
- ⑥ (보안 확보) 연구자는 보안과제 등 보안이 필요한 연구에 AI를 활용할 경우, 연구개발기관이 허가한 보안 환경 내의 도구를 사용해야 함
- 연구개발기관은 연구자가 AI 활용 시 보안 사항이 외부로 유출되지 않도록 제도적·기술적 보안 환경을 마련해야 함
- ⑦ (개인정보 보호) 연구자는 연구개발 전 과정에서 개인정보가 포함된 연구데이터를 AI 서비스에 입력하지 않도록 주의해야 함
- 연구개발 과정에서 부득이하게 개인정보를 활용해야 하는 경우 비식별화 처리 등의 보안 조치를 사전에 이행해야 함

☞ 연구자는 연구 수행 전 AI 활용 권고사항의 내용을 사전에 확인하고, 최종 결과물을 제출(또는 발표)하기 전, AI 활용 권고사항의 준수 여부를 점검할 수 있도록 '<연구자 자가진단표>'를 활용할 것을 권장

□ 평가 과정에서 AI의 활용

- ① (정보 유출) 평가자는 연구계획서, 결과보고서, 논문 등 평가 대상 자료를 외부 AI에 입력하지 않도록 해야 함
- 평가 대상 자료(보고서, 논문)를 외부 AI에 입력할 경우 보안사항, 개인정보 등이 외부로 유출될 위험이 있음
- ② (평가 취지 왜곡) 국가연구개발 과제 평가 및 논문 심사 시, 평가자는 자신의 전문적 지식과 주체적인 판단에 기반하여 평가를 수행해야 함
- AI를 보조적으로 활용하더라도 평가자의 검토나 판단 없이 수용하는 것은 전문가 평가의 취지와 공정성을 훼손함

- ③ (부적절 행위) 피평가자는 AI를 활용하여 **정당한 심사기준을 회피하거나 평가를 왜곡하려는 시도**(은닉 프롬프트 등)를 해서는 안 됨
- 해당 행위는 평가의 신뢰성을 심각하게 저해하는 연구부정행위에 해당하며 관련 법령에 따라 엄격한 제재처분의 대상이 될 수 있음

<평가에서 AI 활용 관련 동향>

- ▲ 동료평가(Peer Review) 시 학술지 또는 출판사가 AI 활용을 제한하거나 금지하는 경우*가 많아, 사전에 AI 활용 정책을 확인하고 준수해야 함
 - * 해외 학술지 및 연구관리기관(Science, Nature Portfolio, NSF, NIH 등) 다수에서는 평가에서 AI 활용 제한
- ▲ 국가연구개발사업(과제) 평가 참여시 중앙행정기관(혹은 전문기관)이 AI 활용을 제한하는 경우*가 존재하므로, 기준을 확인하고 준수해야 함
 - * 한국연구재단 등의 전문기관에서는 평가 시 평가자가 자료를 AI에 업로드하는 것을 금지

IV. AI 활용 서식(예시)

□ (서식) AI 활용 연구자 자가진단표(예시)

연구 수행 전 각 항목의 내용을 사전에 확인하고 최종 결과물 제출 전 실제 준수 여부를 점검하시기 바랍니다.

<연구자 자가진단표>			
		사전 확인	사후 점검
주체적인 검증과 판단	1. 허위 참고문헌, 변형된 데이터 및 왜곡된 이미지 등이 포함되지 않도록 검토하고 출처 표기 및 원본 자료를 확인하였는가	☐	☐
	2. 연구개발 과정에서 핵심 분석과 판단을 연구자가 직접 수행하여 맥락과 내용, 독창성 및 차별성 등을 검토하였는가	☐	☐
부적절한 연구행위 예방	3. AI 도구의 기술적 한계와 윤리적 위험 요소를 사전에 파악하고, 이를 경감하기 위한 방안을 미리 고려했는가	☐	☐
	4. AI를 동원한 부적절한 행위(은닉 프롬프트 삽입 등) 및 저작권 침해 문제 등이 발생할 가능성을 인식하고 유의하였는가	☐	☐
투명성 확보	5. AI 도구의 사용 여부, 모델 이름·버전, 활용 기간 및 활용 방법 등을 관련 규정에 따라 투명하게 공개하였는가	☐	☐
결과 신뢰성 관리	6. AI를 활용한 연구 결과가 주어진 조건에서 재현될 수 있는지 유의하였는가	☐	☐
규정 준수	7. 소속 기관·학술지·전문기관의 AI 관련 규정 및 지침이 지속적으로 개정·보완됨을 인식하고, 주기적으로 확인하여 최신 기준을 준수하였는가	☐	☐
연구보안 확보	8. 보안이 요구되는 자료(예: 미발표 연구 데이터, 원 자료(raw data), 특허 출원 예정 기술 등)를 상용 AI 서비스에 업로드하지 않도록 유의하였는가	☐	☐
	9. 보안 및 민감과제의 연구에서 AI를 활용하는 경우 관련 기관의 절차와 규정을 준수하였는가	☐	☐
개인정보 보호	10. AI에 데이터 입력 전 개인정보 및 민감정보에 비식별화 및 보안 조치를 수행하였는가	☐	☐

□ (서식) AI 활용 여부 공개(예시)

※ 국가연구개발 수행 과정에서 다양한 용도 및 범위로 AI를 활용한 경우 보고서 등의 결과물에 활용 여부를 공개하는 데에 활용(양식은 필요에 따라 수정하여 활용 가능)

기본 정보

	내용
연구과제명(논문)	신뢰할 수 있는 데이터 생성 및 검증을 위한 생성형 인공지능 기술 개발
연구책임자(저자)	김○○
연구수행일자	20yy.mm.dd - 20yy.mm.dd.

활용 도구 정보 및 내용

	내용
도구명 (모델)*	ChatGPT, Claude(Opus 5)
활용 내용	○ 선행 연구 탐색 - 문헌 고찰을 위한 기존 연구의 검색 및 정리 과정에서 Claude를 활용하여 참고문헌 후보를 선별하고, 주요 개념과 키워드 및 핵심 주장을 정리하였음. ○ 도식 - 방법론의 도식(Fig o.o) 구성 및 생성을 위해 Claude를 활용하였으며, 해당 도식은 연구 과정과 결과를 적절하게 설명하고 있음 ○ 문장 교정 - 영문 초록의 표현 및 문법 교정에 ChatGPT를 사용하였음 - 결론의 비문 수정과 적절한 표현 추천을 위해 ChatGPT를 사용하였음
활용 기간	20yy.mm. ~ 20yy.mm.

* 특정한 분석 및 연구를 위한 AI 도구의 경우, 통용되는 모델명 혹은 기능(목표) 등을 명시

최종 책임 확인

본 연구의 핵심 아이디어, 자료와 분석, 결론 및 최종 원고에 대한 책임은 저자(연구자)에게 있으며, 저자는 생성형 AI의 활용 과정을 투명하고 성실하게 관리하였음을 확인합니다.

20 . . .

성 명: (서명)