

부산대 AI 윤리헌장(AI Ethics Charter)

부산대학교 AI 윤리헌장

부산대학교는 인공지능이 미래 지식사회와 대학의 혁신을 견인하는 핵심 기술임을 인식하며, AI의 책임 있는 개발과 활용이 학문 공동체의 신뢰를 형성하는 핵심 가치임을 선언한다. 우리는 인간 중심성, 책임성, 공정성, 투명성, 안전성, 그리고 지속가능성을 대학의 보편 원칙으로 삼아, 다음의 윤리 기준을 준수할 것을 밝힌다.

I. 기본 원칙

1. 인간 중심성 (Human-Centric AI)

AI는 인간의 존엄성과 권리를 최우선으로 보호하며, 인간의 판단을 보조하는 방향으로 설계·활용한다.

2. 책임성과 설명 가능성 (Accountability & Explainability)

AI의 개발자·운영자·사용자는 결과에 대한 책임을 가진다. 주요 기능, 데이터 처리, 판단 근거는 설명 가능해야 하며, 중요한 의사결정은 그 근거와 과정을 명확히 설명할 수 있어야 한다.

3. 공정성과 비차별 (Fairness & Non-Discrimination)

AI는 어떠한 형태의 편향·차별을 유발해서는 안 되며, 정기적인 점검과 개선을 통해 공정성을 확보한다.

4. 투명성과 개방성 (Transparency & Openness)

AI 시스템의 목적·한계·사용 여부는 구성원에게 명확히 공개한다. 대학은 공익적 지식 확산을 위하여 가능한 범위에서 연구 결과를 개방한다.

5. 안전성과 보안 (Safety & Security)

AI 시스템은 기술적·데이터적·물리적 안전을 확보해야 하며, AI기본법에 따른 고영향 AI는 별도의 강화된 검증·관리 체계를 갖춘다.

6. [신설] 지속가능성 (Sustainability & Green AI)

AI 개발 및 활용 시 에너지 소비와 탄소 배출을 최소화하는 지속가능한 원칙을 준수하며, 환경적 책임을 다한다.

II. 대학 구성원 행동 규범

1. 학생

- AI를 학습·연구에 보조적으로 활용하되, 학업 성실성을 훼손하지 않는다.
- 생성형 AI 사용 시 출처, 활용 범위를 명시한다.
- 딥페이크 등 AI 생성 콘텐츠를 타인에게 피해를 주는 방식으로 사용하지 않는다.

2. 교원

- AI 기반 교육·평가·연구 수행 시 윤리 기준을 준수하며 학생을 지도한다.
- AI 결과물의 오류·편향·환각(Hallucination) 발생 가능성을 검토하고 보완한다.
- AI 활용 기준을 수업 초기에 명확히 공지하고, 학생의 학습 과정을 평가한다.

3. 직원

- AI 기반 행정 시스템을 투명하게 운영하고 데이터 보호 책임을 준수한다.
- 이상 징후 발견 시 즉시 보고하고 개선을 요청한다.
- 구성원의 개인정보 처리 시 최소 수집 원칙을 준수한다.

4. [신설]연구자 및 개발자

- 개발하는 AI 모델의 사회적 영향을 사전에 평가하고, 편향·혐오 발언 등 부정적 요소를 제거한다.
- 데이터 수집 시 저작권과 개인정보 보호 법규를 준수하며, 불법 수집 데이터 사용을 금지한다.
- AI 훈련 데이터의 출처를 투명하게 관리하고, 생성 결과물의 저작권 침해 여부를 검토한다.
- 연구 설계 단계에서 환경 영향을 고려하고, 지속 가능한 컴퓨팅 방법론을 우선 적용한다.

5. 대학

- AI 윤리·안전 관리체계를 구축하고 정기적 점검·보완을 실시한다.
- 법령·국제 기준 변화를 반영해 기준을 지속적으로 고도화한다.
- 구성원 대상 AI 윤리 교육 프로그램을 운영한다.

Ⅲ. 시행 및 관리

부산대학교는 “AI 윤리·안전 위원회”를 중심으로 윤리 기준의 운영·점검·교육 체계를 구축하며, AI를 신뢰할 수 있는 방향으로 활용하는 책임 대학(Responsible AI University)으로서의 역할을 수행한다.

※ 위 내용은 AI 기본법 및 개인정보보호법에 기반하여 작성되었음

📌 영역별 AI 윤리 가이드라인 (AI Ethics Guide)

📌 교육(Education) 영역

🔹 학습자 보호

- AI 기반 학습지원 도구는 학습자의 사고력과 창의성을 확장하는 데 사용되어야 하며, 비판적 사고를 저해하거나 학습 과정을 생략하게 만드는 방식의 활용(단순 대필 등)은 금지한다.
- 평가 자동화 도구 사용 시 교원의 최종 확인을 필수로 하며, 기계적 판단만으로 성적을 부여하지 않는다.
- 학생의 개인정보, 성적, 상담 기록 등 민감 정보는 학습 분석이나 AI 훈련 목적으로 활용될 때 엄격한 비식별화 조치와 동의 절차를 거쳐야 한다.

🔹 교육 윤리

- 교육용 데이터는 정확성·다양성·대표성을 확보해야 하며, 부정확하거나 편향된 자료 사용을 금지한다.
- AI 활용 여부에 따라 시험을 구분하여 실시하되, AI 활용이 금지된 시험에 대해서는 엄격한 감독을 실시한다.
- 학생이 AI를 도구로 사용하여 창의적으로 수정·재구성한 결과물에 대한 저작권은 학생에게 있음을 인정한다.

🔹 투명한 활용

- 수업에서 AI 도구 활용 여부, 허용 범위, 평가 기준을 학기 초에 명확히 공지한다.
- AI 생성 결과물은 인간 창작물과 혼동되지 않도록 명확히 표시한다.

📌 연구(Research) 영역

🔹 연구윤리 준수

- AI는 연구자의 사고·검증을 대체할 수 없으며, 생성형 AI 사용시 기여도·출처를 공개한다.
- 연구 부정행위(표절·위조·변조)를 유발할 수 있는 AI 활용을 금지한다.

🔹 연구데이터 보호

- 개인정보·민감정보는 암호화·비식별화 등 최고 수준의 보호조치를 적용한다.
- 연구 참여자 보호 기준(IRB, GDPR 등)을 준수한다.

🔹 지속가능한 연구 수행

- AI 모델 개발 시 전력 소모와 탄소 배출을 최소화하도록 환경 영향을 고려한 연구 설계를 수행한다.
- 사전 학습 모델의 미세 조정·경량화 알고리즘 등 지속 가능한 컴퓨팅 방법론을 우선 적용한다.

🔹 고영향 연구 관리(인공지능기본법 참조)

- 의료·생명·국방 등 고영향 분야 연구는 별도 안전성 검증 절차와 사전 평가를 거친다.
- 잠재적 오용 가능성에 대비한 예방 조치를 마련한다.

📌 행정(Administration) 영역

🔹 공공적 운영


- AI 기반 의사결정시스템은 구성원 권익을 최우선으로 고려하며, 정책 결정은 인간이 최종 승인한다.
- 주요 행정 과정에서 AI 사용 여부를 투명하게 공개한다.

▶ **데이터 거버넌스**

- 대학은 데이터 정확성·보안·접근권한을 관리하는 체계를 유지한다.
- 행정 데이터 처리 및 AI 활용 시 관련 법령·국제 기준을 준수한다.

▶ **신뢰성 검증**

- AI 시스템의 정확성·편향·보안성은 주기적으로 점검하며, 외부 인증을 통해 객관성을 확보한다.
- 오류·위험 발생 시 즉각 중단하고 개선 프로세스를 가동한다.

 **AI 윤리헌장 및 윤리 가이드라인 활용 방안**

- ▶ AI 윤리헌장 및 윤리가이드라인은 매년 AI 윤리위원회의 심의를 통해 갱신되며, 총장이 최종 결재한다.
 - ▶ AI 윤리헌장 및 윤리가이드라인은 홈페이지 및 LMS에 게시하여 모든 구성원이 언제든지 접근하여 확인할 수 있도록 조치한다.
 - ▶ 모든 구성원은 AI 윤리헌장 및 윤리가이드라인에 대한 교육을 연1회 이상 실시한다.
 - 부산대학교 AI 윤리 온라인 필수교육 개발 및 전 구성원 수료 확인
 - 학기 초 모든 교육과정 시작 단계에서 AI 윤리교육 시행
-