

2023.11.14.(화) 조간용

이 보도자료는 2023년 11월 14일 오전 06:00부터 보도할 수 있습니다.



보도자료

담당 부서: 서울디지털재단
AI·빅데이터팀

사진 없음 ☐ 사진 있음 ☒ 쪽수: 5쪽

경영전략실장		주성환	570-4612
AI·빅데이터팀장(겸직)		주성환	570-4612
담당자	사업운영	이정환	570-4641
	언론홍보	유채란	570-4625
관련 누리집 (메뉴)		http://sdf.seoul.kr	

서울디지털재단, 시민의 안전한 AI 일상화를 위해 지자체 최초로 생성형 AI 윤리 가이드라인* 발표..

- 서울시민 AI 일상화 3대 프로젝트 중 하나인 윤리 가이드라인 발표
- 3대 기본원칙 기반으로 생성형 AI 위험 유형에 따라 9대 핵심요건 도출
- 재단 “시민교육과 윤리문화 확산을 통해 AI 건전 활용에 앞장설 것”

※ AI확산 이면의 AI위험(가짜뉴스·개인정보 침해 등)에 대한 시민 안전을 위해 선제적 활용

- ☐ 서울디지털재단(이사장 강요식)은 지난 10월 26일 제시한 서울시민 인공지능(AI) 일상화 3대 프로젝트 중 하나인 서울시 생성형 AI 윤리 가이드라인을 발표했다.
 - 서울시 생성형 AI 윤리 가이드라인은 정부의 디지털 권리장전 제정과 분야별 윤리 가이드라인을 마련한다는 정부 정책의 기초를 이어받기 위한 노력의 일환으로 마련되었다.
- ☐ 이번 윤리 가이드라인은 생성형 AI 특성과 서울시 핵심가치를 기반으로 연구한 것이 특징이며, 이용자를 중심으로 운영자·개발자가 지켜야 할 윤리적 준수사항을 함께 포함하였다.

- 본 윤리 가이드라인은 민간기업에 대한 규제적 접근이 아닌 이용자·운영자·개발자가 생성형 AI를 안전하게 활용할 수 있도록 자율적으로 준수할 수 있는 규범을 제시한다는 취지이다.
- 산·학·연 여러 전문가의 의견을 수차례 수렴하여 윤리 가이드라인에 대한 추진방향 및 3대 기본원칙을 도출하였고, 이를 바탕으로 9대 핵심요건을 제시하였다.
- 윤리 가이드라인은 ‘이용자 중심, 위험예방, 지속가능성’을 3대 기본 원칙으로 선정하였다.
 - ‘기본원칙’은 생성형 AI의 특성(이용자의도에 따라 능동적인 결과 생성)과 서울시 핵심가치(삶의질, ‘도시경쟁력’)를 토대로 도출하였다.
 - ‘이용자 중심’은 생성형 AI에 대한 이용자의 참여 정도에 따라 새로운 위험요인이 나타나기 때문에 ‘이용자를 중심’으로 한 윤리 가이드라인 마련이 필요하다는 내용이다.
 - ‘위험예방’은 삶의 질을 향상시키기 위해서는 AI의 활용만큼이나 위험을 최소화할 필요가 있기 때문에 생성형 AI의 위험을 기반으로 핵심요건을 도출해야 한다는 내용이다.
 - ‘지속가능성’은 도시경쟁력을 약화시키는 범죄, 환경오염 등의 도시 문제를 생성형 AI를 통해 해소할 필요가 있으며, 이를 위해 윤리 가이드라인에 지속가능성의 개념이 필요하다는 내용이다.
- 윤리 가이드라인은 3대 기본원칙을 바탕으로 ‘이용자의 검증 역량, 책임성, 데이터 권리와 책임, 사용가능성 확보, AI 학습데이터 안전성, AI 윤리 소양, 공공성, 생태적 지속가능성, 보안성’을 9대 핵심요건으로 제시하였다.

- 핵심요건 도출을 위해 국내외 생성형 AI의 위험을 제시한 문헌과 발표 자료를 조사·분석하고 위험의 유사성에 따라 유형을 구분하였다.
 - 또한 유형별 실제 사례를 분석해 이해관계자별로 착안할 수 있는 쟁점을 도출하고 이를 기존 윤리원칙이 어떻게 해석하는지 분석하였으며, 분석 결과를 기반으로 이해관계자가 준수해야 할 핵심요건을 도출하였다.
- 또한, 이해관계자(이용자·운영자·개발자)가 윤리 가이드라인의 핵심 요건을 준수하기 위해 알아야 할 핵심요건별 위험유형과 실제사례와 함께 배경과 주의사항이 ‘가이드라인’에 제시된다.
- 윤리 가이드라인이 적극 확산되고 활용되기 위해서는 이해관계자가 처한 위험 유형을 정확하게 인지하고, 이에 맞는 준수사항을 인지할 필요가 있다.
- 강요식 이사장은 “생성형 AI가 확산되는 가운데, 가짜뉴스, 개인정보 침해 등 AI 위험 우려도 커지는 상황이다. 이에 재단은 시민의 안전을 위해 선제적으로 지자체 최초의 생성형 AI 윤리 가이드라인을 발표하게 되었다”며, “이를 통해 시민들이 안전하고 편리하게 AI를 활용하기를 기대하며, 우리 재단도 시민교육, 윤리문화 확산 활동을 통해 건전한 AI 활용문화 조성을 위해 노력할 것”이라고 밝혔다.

붙임 : 1. 서울시 생성형 AI 윤리 가이드라인 기본원칙

2. 서울시 생성형 AI 윤리 가이드라인 비전과 핵심요건



비전

신뢰 기반의 AI 도시, 서울

추진방향과
기본원칙

- 생성형 AI와 전통적 AI의 차이 → (원칙1) **이용자 중심**의 윤리 가이드라인
- 市 핵심가치 ‘삶의 질’과 AI → (원칙2) **위험예방**을 위한 윤리 가이드라인
- 市 핵심가치 ‘도시경쟁력’과 AI → (원칙3) **지속가능성**을 중시하는 윤리 가이드라인

9대
핵심요건

1. 이용자의 검증 역량 (허위 정보 답변) ▶ 이용자 는 활용 상황에 따라 생성형 AI의 도출 정보를 검증하고 활용하는 역량을 확보 ▶ 이용자 는 생성형AI 결과물을 객관성이 요구되는 상황에 활용할 경우 편향·차별 등의 내용이 포함되어 있는지 점검	4. 사용가능성 확보 (결과물의 저작권 침해) ▶ 이용자 는 생성형AI 결과물을 활용할 때 사용가능성을 확보하기 위해 노력 ▶ 이용자 는 생성형AI 결과물을 활용할 때 타인의 권리침해 여부를 법·제도적으로 고려	7. 공공성 (사회적 문제 야기) ▶ 이용자 는 생성형 AI를 보조적 수단으로 인지·활용해야 하며 공공성을 침해하지 않도록 노력 ▶ 운영자 는 생성형AI 도입 시 해당 서비스 관련 일자리에 미치는 영향을 고려
2. 책임성 (악의적 활용) ▶ 이용자 는 생성형AI를 개발 목적에 부합하도록 활용 ▶ 이용자 는 생성형AI 결과물이 사회 전반에 부정적인 영향을 미치지 않도록 노력 ▶ 이용자 는 자신이 생성을 유도한 생성형 AI의 결과물에 대한 책임을 인지	5. AI 학습데이터의 안전성 (저작권 침해와 오용·편향 데이터 학습) ▶ 개발자 는 타인의 저작권을 침해하지 않은 범위에서 데이터를 학습 ▶ 개발자 는 오용·편향데이터를 학습에 활용하지 않아야 함	8. 생태적 지속가능성 (탄소 多배출) ▶ 이용자·운영자 는 생태적으로 설계된 서비스를 도입·활용하기 위해 노력 ▶ 개발자 는 데이터 구축 단계에서 최소한의 탄소배출을 위해 노력
3. 데이터 권리와 책임 (개인정보·중요정보 유출) ▶ 이용자 는 개인·중요정보가 생성형 AI에 노출되지 않도록 유의 ▶ 이용자 는 생성형AI를 활용함과 동시에 서비스가 이용자 정보를 수집한다고 인식 ▶ 이용자 는 개인정보에 대한 책임과 권한이 자신에게 있음을 인식	6. AI 윤리 소양 (이용자 이해도 증진) ▶ 이용자 는 생성형 AI 특성을 이해하고 사용상의 문제점을 인식하려고 노력해야 하며, 운영자 는 이용자의 AI 윤리 소양을 가질 수 있도록 노력	9. 보안성 (생성형 AI 자체의 보안) ▶ 이용자·운영자 는 생성형 AI가 가진 보안 위험에 유념 ▶ 개발자 는 생성형AI에 대한 보안사항을 적시에 설명·통보 ▶ 이용자 는 생성형AI로 보안상 피해를 입은 경우 책임과 보호조치를 요구해야 하며, 운영자·개발자 는 보안사고 발생 시 명확한 책임 주체를 설정하고 통보