1. AI 윤리원칙

삼성전자의 AI 윤리원칙 약속

삼성전자는 인류에 공헌하는 AI 를 위해 AI 윤리원칙을 준수합니다. 삼성전자는 AI 기술을 통해 언제 어디서나, 안전하고, 도움을 주며, 지속적으로 학습하는 인간 중심 기기를 만들고자 합니다.

삼성전자의 모든 구성원은 AI 개발과 이용에 있어 '공정성', '투명성', '책임성' 이라는 AI 윤리 원칙을 준수하겠습니다.

✓ 공정성

- AI 의 모든 단계에서 인권을 존중하며, 공정성 및 다양성을 추구합니다.
- AI 로 인해 불공정한 편견이 조장되거나 강화되지 않도록 노력합니다.
- 누구나 쉽게 접근할 수 있는 AI 를 만들도록 노력합니다.

인권 존중

AI 의 모든 단계에서 인권을 존중하겠습니다. 삼성전자는 국제인권장전, UN 기업과 인권 이행 원칙(UNGPs) 등 인권과 관련된 국제 기준을 존중하고, 인권을 존중하겠다는 삼성전자 인권 기본원칙을 발표하였습니다. 이러한 원칙을 바탕으로 AI 의 모든 단계에서 인간의 존엄성, 자율성, 인간 생명과 안전에 관한 기본권, 표현·정보의 자유 등 인간 중심의 가치를 존중하고 지키겠습니다.

공정성 및 다양성 존중

AI 의 모든 단계에서 공정성 및 다양성을 존중합니다. 특히, 합리적인 이유 없이 성별, 민족, 국적, 소득, 성적 취향, 정치·종교적 신념 등에 따라 특정 개인이나 집단이 차별 받지 않도록 노력합니다.

불공정한 편향 방지

AI 로 인해 불공정한 편견이 조장되거나 강화되지 않도록 노력합니다. 데이터 수집 단계부터 성별, 인종 등 데이터의 편향성을 최소화하고, 편향성을 극복할 수 있는 검증 툴 등의 기술 연구를 지속하겠습니다.

포용성 및 접근성 보장

인류 전체가 AI 기술의 혜택을 받을 수 있도록, 누구나 쉽게 접근할 수 있는 AI 를 추구합니다. 성별이나 신체적 능력 등과 상관 없이 모든 사용자가 AI 제품 및 서비스를 편리하게 사용할 수 있도록 접근성 기능 개발과 연구를 지속하겠습니다.

✓ 투명성

- 사용자가 자신이 AI 와 상호 작용한다는 것을 인지할 수 있도록 합니다.
- 기술적으로 가능한 범위 내에서 설명 가능한 AI 를 위해 노력합니다.
- AI 서비스 사용자의 프라이버시가 보호될 수 있도록 노력합니다.

투명성

사용자가 자신이 AI 와 상호작용한다는 것을 알 수 있도록 합니다. AI 는 사용자에게 스스로를 인간으로 오해될 수 있도록 해서는 안되며, AI 를 활용한 제품 및 서비스라는 점을 알리겠습니다.

설명 가능한 AI

기술적으로 가능한 범위 내에서 설명 가능한 AI 를 위해 노력합니다. 사용자가 이해할 수 있는 언어로 AI 의 목적과 기능, 한계 등을 설명할 수 있도록 노력합니다.

개인정보보호

삼성전자는 AI 를 운영함에 있어 개인정보를 목적 내 최소한으로, 투명하게 수집·사용하고 안전하게 이용하며, 사용자의 선택권을 존중합니다. 또한, 개인정보보호를 위해 한발 앞서 위험 요소들에 대비하여 강력한 보안 기술을 활용합니다.

✓ 책임성

- AI 의 사회적 책임을 다할 수 있도록 노력합니다.
- AI 가 안전하고, 보안이 유지되도록 노력합니다.
- AI 를 통한 사회적 기여가 기업의 문화가 될 수 있도록 노력합니다.

사회적 책임

AI 의 사회적 책임을 다할 수 있도록 노력합니다. AI 가 사회적으로 미칠 수 있는 영향을 식별하고, 이에 따른 AI 윤리 이슈를 자율 점검할 수 있도록 관련 임직원 가이드 및 교육을 운영합니다.

안전성 및 보안성

AI 의 모든 단계에서 안전하고, 정보 보안이 유지되도록 노력합니다. AI 가 가져올 수 있는 위험 수준에 비례하여 적절한 위험 예방을 할 수 있도록 노력합니다. 해킹과 같은 공격에 취약점이 없도록 정보 보안 프로세스를 운영합니다.

AI 를 통한 사회 기여

AI 를 통한 사회 기여가 기업의 문화가 될 수 있도록 노력합니다. AI 윤리 교육 등을 통해 AI 윤리 중요성에 대한 인식을 제고하고, 개발·연구하는 AI 과제를 통해 사회 기여할 수 있는 부분이 있는지 지속적으로 검토합니다.

2. AI 윤리원칙 운영

삼성전자는 공정성, 투명성, 책임성의 AI 윤리 원칙을 실천할 수 있도록 구체적이고 실행 가능한 가이드를 마련하고 AI 거버넌스 협의회를 구성하여 자체 점검 프로세스를 운영하고 있습니다.

AI 윤리원칙 준수를 위한 삼성전자의 노력

공정성(Fariness)	투명성(Transparency)	책임성(Accountability)
- 접근성 사무국 운영을 통한 삼성전자 접근성 정책 준수 - 자가진단 체크리스트 제공으로 AI 가 활용하는 데이터의 공정성 및 편향성 평가 - AI 서비스 제공에 있어 편향성 등의 공정성 위반이 감지되는 경우 모니터링/경고 시스템	- 삼성전자 개인정보보호 정책 준수 - 사용자가 AI 와의 상호 작용을 인지 가능하도록 설계 - 모델 카드 및 데이터 카드 작성 가이드	- AI 윤리 이행 가이드 제공 및 임직원 대상 필수 교육 운영 - AI 거버넌스 협의회 운영 - AI Safety 점검 및 Risk 완화를 위한 AI Safety 점검 프로세스 운영 - AI Summit 참여 등 업계, 학계, 시민 사회 및 정부와 지속적인 협력

✓ 공정성(Fairness)

- 삼성전자 접근성 정책 준수 확인

접근성 사무국은 삼성전자 제품 및 서비스를 사용하는 모든 고객의 접근성 경험 향상을 위해 노력하고 있습니다. 당사는 모든 이해관계자가 동등하고 편리하게 제품과 서비스를 이용할 수 있도록 포용성 있는 AI 를 개발하고 있습니다.

- AI 데이터에 대한 편향성 평가

자가진단 체크리스트를 통해 AI 가 활용하고 생성하는 데이터가 인종, 성별, 나이 등 편향된 내용을 담고 있지 않는지 확인하고 편향된 내용을 담은 데이터는 폐기됩니다. 이후 동일한 편향 데이터를 사용하거나 생성하지 않도록 자체적으로 피드백 프로세스를 운영하고 있습니다.

- 공정성 위반이 감지되는 경우 모니터링/경고 시스템

삼성전자는 공정성 위반에 대한 모니터링 및 경고 시스템을 마련하였습니다. 특히, 삼성전자로부터 제공되는 AI 서비스의 편향성 발생 등과 같은 공정성 위반 위험이 감지되는 경우, 이에 대한 모니터링/경고 시스템이 작동하여 중대한 위반의 현실화를 방지하고자 노력합니다.

✓ 투명성(Transparency)

- 삼성전자 개인정보보호 정책 준수 확인

삼성전자는 개인정보를 목적 내 최소한으로, 투명하게 수집·사용하고 안전하게 이용하며, 사용자의 선택권을 존중하는 것을 원칙으로 합니다.

AI 관련 데이터의 수집, 처리, 활용의 전 과정에서 개인정보 침해 여부에 대한 확인 프로세스를 수립하여 운영 중이며, 강력한 보안기술을 적용하여 데이터를 안전하게 보호하고 있습니다.

- 사용자의 AI 에 대한 인지 및 이해 향상 노력

삼성전자는 AI 설계 단계부터 사용자가 AI 와 상호 작용하거나, AI 를 사용하고 있다는 사실을 인지할 수 있고, 이러한 사실을 투명하게 공개하도록 개발자 등 임직원에게 가이드 하고 있습니다.

또한 다양한 이해관계자가 이해할 수 있는 명확한 언어로 AI 의 목적과 기능, 특별한 주의가 필요한 기능 및 한계 등을 설명하고 있습니다.

- 모델 및 데이터 카드 작성 가이드

AI 윤리원칙에 입각한 AI 모델 및 서비스 개발·평가 과정과 데이터 관련 정보 기입 양식 및 가이드를 제공하여, 개발 과정의 투명성 확보와 책임 있는 프로세스 구축을 위해 노력하고 있습니다.

✓ 책임성(Accountability)

- AI 윤리 이행 가이드 제공 및 임직원 교육

• AI 윤리 이행 가이드 제공

AI 기술이 활용되는 제품과 서비스를 설계·개발·배포·구현·운용하는 과정에서 임직원이 AI 윤리원칙을 어떻게 이행할 수 있을지 내부 가이드라인 및 체크리스트를 제공하여 AI 윤리원칙 이행의 방법을 가이드 하고 있습니다

• AI 윤리 의무 교육 진행

임직원들의 AI 윤리에 대한 이해도를 높이고 중요성에 대한 인식 제고를 위해 글로벌 AI 윤리 및 규제 동향, 임직원을 위한 AI 윤리 가이드 교육을 진행하고 있습니다. 2024 년도부터는 사내 AI 개발자·전문가 대상의 AI 윤리 교육 과정을 신설하고, AI 윤리 교육을 전임직원 대상 의무 교육으로 진행하고 있습니다.

- AI 거버넌스 협의회 운영

삼성전자 AI 윤리 원칙에 따른 개발 프로세스와 개발자용 도구, 윤리적인 AI 개발을 위한 가이드를 지원·교육·감독 할 수 있도록 AI 전략팀, 삼성리서치, Compliance 팀 등 유관 부서들이 함께 AI 거버넌스 협의회를 운영하고 있습니다. AI 거버넌스 협의회의 중요한 사안은 대표이사 주관의 AI 전략협의체에 보고하며, AI 윤리 개발 문화를 정착시키고 발전시키기 위해 지속 노력하고 있습니다.

- AI Safety 점검 프로세스 운영

AI Safety 를 위한 점검 프로세스를 수립하여, 생성형 AI 관련 주요 리스크 (모델 기능 오류 유도, 생성 결과물 악용, 개인정보 유출, 유해 컨텐츠 생성, 편향성, 위험정보 생성 등)에 대한 Safety 점검과 AI Red Team 활동을 포함한 Risk 완화 활동을 수행하고 있습니다.

- 국제 표준 준수

삼성전자는 소비자들이 신뢰하며 사용할 수 있는 AI 제품과 서비스를 제공하기 위해 국제 표준에 맞춘 체계적인 관리체계를 운영합니다.

2023 년 삼성전자 생활가전 사업부는 국내 최초로 국제 표준인 '인공지능 경영시스템(ISO/IEC 42001)' 인증을 받았습니다. 이 인증은 AI 제품이나 서비스를 만들고 제공할 때, 회사가 AI 윤리를 지키고 신뢰성에 관한 위험을 책임감 있게 관리하는지를 중점적으로 평가합니다. 한국표준협회(KSA)는 AI 가전 제품·서비스의 기획, 개발, 양산, 폐기까지 전 생애주기에 걸쳐 보안, 공정성, 투명성, 그리고 데이터와 시스템 품질을 적절하게 관리하고 있는지 확인하여 이 인증을 부여합니다.

- 글로벌 이니셔티브와 파트너십 참여

2023 년 11 월 삼성전자는 영국 블레츨리 파크에서 열린 세계 최초 AI Safety Summit 에 참석하여 글로벌 AI 안전 거버넌스 구축에 동참했습니다.

이 회의에서는 인간 중심의 안전한 AI 개발과 AI 를 사용하며 발생할 수 있는 위험을 식별하고 문제를 해결하는 것의 시급성을 확인했습니다.

또한 AI Safety 관련 국제 협력 강화를 위한 논의와 연구를 지속하기로 약속하는

'블레츨리 선언'이 채택되었습니다. 2024 년 5 월 삼성전자는 서울에서 개최된 AI Seoul Summit 에 참여하여, 'Seoul AI 기업 서약(Seoul AI Business Pledge)'에 동참했습니다. 2025 년 2 월에는 파리에서 개최된 AI Action Summit 에 참여하고, AI Safety Framework 을 공지하는 등 'Seoul AI 기업 서약'을 이행했습니다.

삼성전자는 AI 의 사회적 영향에 대한 이해를 높이고, AI 기술을 책임 있는 방식으로 활용하기 위해 국내외 다양한 이해관계자들과 협력합니다.

국제표준화기구(ISO/IEC)의 인공지능 기술위원회(JTC1 / SC42) 에서 진행되는 인공지능 국제 표준화 논의에 참여하며 글로벌 기준에 부합하는 AI Safety 와 신뢰성 확보를 위해 노력합니다. 또한 한국 정부 주관 "산업 인공지능 표준화 포럼"에 참여해 학계, 연구기관 과 산업계의 전문가들과 올바른 정책 수립을 위해 소통합니다. 포럼은 AI 신뢰성에 대한 평가기준과 윤리 가이드라인 수립, 양질의 데이터 축적, AI 적용 산업별 상호운용성 확보 등을 위한 표준화를 추진합니다.