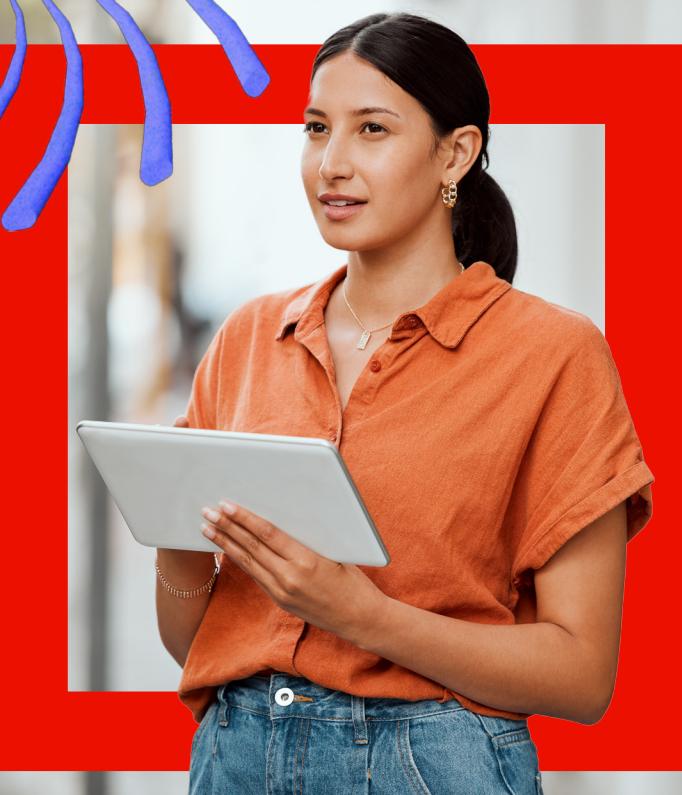




데이터시트

Adobe Firefly 윤리 및 편향

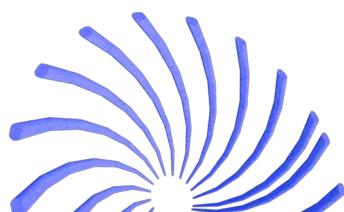


올바른 방식으로 구현된 생성형 AI는 기존 워크플로에 새로운 가능성과 역량을 더하고 창의력을 증폭시킵니다. 그러나 대부분의 신기술과 마찬가지로, 생성형 AI 또한 저작권 침해부터 유해한 영향과 편향에 이르기까지 해결해야 할 여러 과제와 위험이 있습니다. Adobe는 클라우드 기술 전반에서 10년 이상 쌓아온 AI 혁신의 유산을 바탕으로 신뢰, 책임감, 투명성이라는 [AI 윤리 원칙](#)을 준수하는 Firefly 생성형 AI를 개발 및 배포했습니다.

Adobe는
크리에이티브 생성형
AI를 신중하게
개발하므로, 사용자는
자신이 생성하는
콘텐츠를 안심하고
활용할 수 있습니다.

상업적으로 안전한 설계

Firefly의 독점적인 생성형 AI 모델은 상업적으로 안전하게 사용할 수 있도록 설계되었습니다. 이는 Adobe가 라이선스 및 저작권 문제를 책임감 있게 해결했다는 의미이며, 사용자는 결과물을 안심하고 게시할 수 있습니다. Adobe는 Adobe Stock과 같은 라이선스 콘텐츠와 저작권이 만료된 공용 도메인 콘텐츠의 수많은 이미지, 오디오, 영상을 기반으로 Firefly 생성형 AI 모델을 학습시켰습니다.



IP 면책

일부 고객은 Firefly 결과물과 관련된 청구 발생 시 계약상의 IP 면책이 포함된 이용 권한을 구매할 수 있습니다. 즉, Adobe는 Firefly 결과물이 제3자의 특허, 저작권, 상표, 퍼블리시티권 또는 프라이버시권을 직접적으로 침해하거나 위반한다고 주장하는 청구가 제기될 경우, 법적 방어와 금전적 손해 배상에 대한 책임을 집니다. Firefly 결과물과 관련된 청구에 대한 Adobe의 책임은 고객이 Adobe와 체결한 기준 계약상의 기술 기반 IP 청구에 적용되는 제한과 동일하게 적용됩니다. Adobe 면책 조항은 Firefly 결과물의 수정, Firefly 결과물이 사용된 컨텍스트, 맞춤형 학습을 위해 제공된 콘텐츠 또는 고객 계약을 위반한 Firefly 결과물의 사용을 포함하여 고객의 행위나 콘텐츠에 기반하거나 그로 인해 발생한 청구에는 적용되지 않습니다.

[Firefly의 상업적 안전 및 면책에 대해 자세히 알아보기](#)



비즈니스에 안전한 설계

AI 기능을 출시하기 전에 Adobe 테스트 팀은 유해하고 부정적인 결과를 완화하고, 분류 및 필터링이 필요한 언어의 우선순위를 정하고, 즉각적인 차단 및 거부 목록을 작성하고, 업무에 적합하지 않은 콘텐츠(NSFW)의 노출을 줄이기 위한 세부적인 프로세스를 따릅니다.

[Firefly의 콘텐츠 조정에 대해 자세히 알아보기](#)

신뢰

Adobe는 AI 지원 툴의 결과에 책임을 가지며, AI 윤리 원칙을 실천하기 위해 다음과 같은 표준 프로세스를 운영하고 있습니다.



수립: 거버넌스 절차를 수립하여 학습 데이터 및 AI 모델을 추적합니다.



교육: 엔지니어 및 제품 관리자에게 AI 윤리 문제를 교육합니다.



설립: AI 개발 보급을 감독하고 AI 윤리 문제에 대한 공론의장을 제공하기 위한 AI 윤리 심사 위원회를 설립합니다.



의무화: 서비스 개발 프로세스에 AI 영향 평가를 의무화하여 새로운 AI 기술을 배포하기 전에 AI 윤리 검토가 이루어지도록 합니다.



개발: 배포 후 부정적인 AI 영향에 대한 피드백 및 개선을 보장하기 위한 프로세스를 개발합니다.

Adobe는 내부 및 외부로부터 AI에 대한 우려 사항을 접수하고 대응하는 전담 프로세스를 통해 적절한 시정 조치를 취합니다. 내부 AI 윤리 심사 위원회 외에도, Firefly에는 피드백 매커니즘이 내장되어 있습니다. 따라서 특정 기능이 편향적이거나 부정확하다고 인식되는 결과를 생성하는 경우 사용자가 이를 보고할 수 있습니다. 이러한 사용자 커뮤니티와의 피드백 루프는 Adobe 팀이 피해를 최소화하고 신뢰와 책임감을 유지하는 중요한 방법입니다.

[Adobe의 신뢰 노력에 대해 자세히 알아보기](#)



책임감

Adobe는 우리가 세상에 제공하는 혁신과 기업 활동이 미치는 영향에 대한 책임을 중요하게 생각합니다. 우리는 AI 개발과 사용에 사려 깊은 보호 장치를 마련하면 잠재적으로 해로운 영향을 완화하면서 사회에 혜택을 제공할 수 있는 AI의 잠재력을 최대한 실현할 수 있다고 믿습니다. 책임 있는 AI 개발을 위한 Adobe의 접근 방식은 다음과 같습니다.



설계: AI 시스템을 신중하게 설계합니다.



평가: 최종 사용자와 상호 작용하는 방식을 평가합니다.



수행: 실사를 실시하여 원치 않는 유해한 편향을 완화합니다.



평가: AI 기술이 인간에게 미치는 영향을 평가합니다.



피해 및 편향 완화

AI는 학습에 사용된 데이터에 따라 결과가 달라지며, 모든 데이터에는 편향이 있습니다. 책임 있는 AI 개발을 위해서는 잠재적인 편향과 그로 인한 피해를 예상하고, 이를 완화하기 위한 선제적 조치를 취하며, 피해 확산을 측정 및 문서화하고, 예상치 못한 유해한 결과를 모니터링하고 대응할 수 있는 시스템을 마련해야 합니다. Adobe는 포용성을 염두에 두고 설계하지만, AI 시스템이 인간의 속성(예: 인종, 성별, 소득, 종교, 장애 등)과 관련된 유해한 선입견이나 고정관념을 확산시킬 수 있는 잠재적으로 불공정하거나 차별적이거나 부정확한 결과를 초래할 가능성에 대해 지속적으로 평가합니다. 또한 책임과 윤리에 대한 약속을 이행하기 위해 다음과 같은 전담 리소스와 표준 프로세스를 구축하여 Firefly 생성형 AI의 영향을 지속적으로 평가하고 있습니다.

AI 윤리 위원회 - 개발팀과 협력하여 개발의 모든 단계에서 AI 윤리 원칙을 준수하고, 잠재적으로 유해한 영향이나 편향을 사전에 완화하기 위해 노력합니다.

AI 윤리 심사 위원회 - 모든 Adobe AI 기반 기능 및 제품을 공개 출시 전에 엄격하게 검토합니다.

투명성

Adobe는 투명성이 신뢰를 형성한다고 믿으며 고객, 파트너, 업계 동료들이 Adobe 제품에 갖는 신뢰를 중요하게 생각합니다. Adobe Firefly에서 생성된 모든 이미지에 Content Credentials를 적용하여, 이러한 투명성을 AI로 생성된 결과물로까지 확대하고 있습니다. Adobe가 창립 멤버로 참여한 CAI(Content Authenticity Initiative)를 통해 제공되는 기술 에코시스템의 일환인 Content Credentials는 온라인에서 잘못된 정보의 확산을 방지하고 신뢰성과 공신력을 쌓는 데 기여하는 새로운 유형의 변조 방지 메타데이터입니다. 이는 크리에이터가 자신의 작품에 대한 인정을 받을 수 있도록 할 뿐만 아니라, 이미지에 생성형 AI가 사용되었음을 표시하여 다른 사용자나 소비자에게 이미지가 어떻게 만들어졌는지 알릴 수 있습니다.

Adobe의 투명성 원칙에 대해 자세히 알아보세요.

담당자에게 문의하여 Adobe가 지속적으로 혁신하고, 학습하고, 성장하면서 책임감 있는 AI 기술 개발을 위해 어떻게 노력하는지 알아보세요.

