[Introduction]

Course introduction - AI 원칙 적용

AI의 발전과 영향

- 일상 속 **AI:** 우리는 이미 교통 예측, 날씨, 콘텐츠 추천 등 다양한 AI 기술과 상호작용하고 있습니다.
- 기술 발전 속도**:** 스탠퍼드 대학의 2019년 보고서에 따르면, 2012년 이후 AI 컴퓨팅 성능은 약 3.5개월마다 두 배로 증가하며 기하급수적으로 발전하고 있습니다.
- 정확도 향상**:** ImageNet 이미지 분류 챌린지에서 AI의 오류율은 2011년 26%에서 2020년 2%까지 감소했습니다. 이는 인간의 오류율(5%)보다 낮은 수치입니다.
- 진입 장벽 감소**:** 과거에는 소수의 전문 엔지니어만 AI를 개발할 수 있었지만, 이제는 AI 전문 지식이 없는 사람들도 AI를 구축할 수 있게 되었습니다.

책임 있는 **AI**의 필요성

- **AI**의 한계**:** AI는 결코 완벽하지 않으며, 책임 있는 개발을 위해서는 잠재적 문제, 한계, 의도치 않은 결과를 이해해야 합니다.
- 사회적 편향 증폭**:** 기술은 사회의 모습을 반영하므로, 올바른 관행이 없으면 AI는 기존의 편견이나 문제를 복제하고 증폭시킬 수 있습니다.
- 고유한 원칙**:** '책임 있는 AI'에 대한 보편적인 정의나 간단한 공식은 없으며, 각 조직은 자신들의 가치와 사명을 반영한 고유한 AI 원칙을 개발해야 합니다.

Google의 책임 있는 **AI** 접근 방식

- **4**가지 핵심 아이디어**:** 책임 있는 AI의 공통 주제는 투명성**,** 공정성**,** 책임**,** 그리고 개인정보 보호입니다.
- **Google**의 가치**:** Google은 모두를 위한 **AI,** 책임 있고 안전한 **AI,** 개인정보를 존중하는 **AI,** 과학적 우수성에 기반한 **AI**를 목표로 합니다.
- 프레임워크 활용**:** Google은 AI 원칙을 책임 있는 의사결정을 위한 프레임워크로 활용합니다. 이 작업은 계속 진행 중이며, Google은 그 과정을 공유하며 커뮤니티와 협력하고자 합니다.

AI 개발의 핵심은 **'**사람**'**

- 인간의 역할**:** AI 개발에서 기계가 중심적인 의사결정 역할을 한다는 오해가 있지만, 실제로 데이터를 만들고, 모델을 훈련시키고, 배포 방식을 결정하는 것은 사람입니다.
- 가치관의 개입**:** 모든 결정 지점에서 사람들은 자신의 가치관에 따라 선택을 내립니다. 따라서 개발의 모든 단계에서 책임감 있는 선택이 이루어지도록 고려와 평가가 필수적입니다.
- 핵심 목표**:** AI에 대한 복잡한 정의에 얽매이기보다, 기술을 책임감 있게 개발하는 것이 가장 중요한 목표입니다.

본 과정의 목표는 Google의 책임 있는 AI 여정을 공유함으로써, 여러분의 조직이 자체적인 AI 전략을 수립하는 데 도움을 주는 것입니다.

Google and responsible AI - 책임 있는 **AI**의 중요성

- **AI**의 광범위한 영향**:** AI 기술은 삶을 편리하게 만드는 혁신적인 기회를 제공하지만, 동시에 기술 제공자에게 깊은 책임을 요구함.
- 윤리적 문제**:** AI 혁신은 의도치 않은 결과를 초래할 수 있으며, 여기에는 머신러닝의 공정성 문제, 편향의 확산, AI로 인한 실업, AI 의사결정에 대한 책임 소재 등이 포함됨.
- 윤리적 설계의 필요성**:** 단순히 논란의 여지가 있는 용도뿐만 아니라, 모든 AI 활용 사례에서 윤리적 문제나 의도치 않은 결과를 방지하고, 더 나은 이점을 얻기 위해 책임 있는 AI 관행이 필요함.

An introduction to Google's AI principles - Google의 접근법**:** 책임 있는 **AI =** 성공적인 **AI**

- 신뢰 구축**:** 책임감을 가지고 AI를 개발하는 것은 더 나은 모델을 만들고, 고객 및 이해관계자와의 신뢰를 구축하는 길임.
- 프로세스의 중요성**:** 신뢰가 깨지면 AI 프로젝트가 중단되거나 실패하고, 최악의 경우 피해를 줄 수 있음. 따라서 Google은 일련의 평가와 검토를 통해 일관된 접근 방식을 유지하며, 모든 프로젝트가 AI 원칙에 부합하도록 함.
- 문화의 역할**:** 책임 있는 AI 개발을 위해서는 건강한 토론을 수용하는 집단적 가치 시스템에 기반한 문화가 필수적임.

모두를 위한 책임 있는 **AI**

- 반복적인 실천**:** 책임 있는 AI는 한 번에 완성되는 것이 아니라, 헌신, 규율, 그리고 지속적인 학습과 조정을 필요로 하는 반복적인 과정임.
- 작은 시작의 중요성**:** 조직의 규모나 리소스와 관계없이, 책임 있는 AI는 작은 단계부터 시작하는 것이 중요함. 회사 가치와 제품이 만들고자 하는 영향에 대해 정기적으로 고민하는 것만으로도 큰 도움이 될 수 있음.
- 커뮤니티의 힘**:** Google은 AI 개발 커뮤니티의 한 목소리일 뿐이며, 책임 있는 AI의 진정한 핵심은 커뮤니티와의 협력에 있음.
- 신뢰할 수 있는 프로세스 구축**:** 모든 결정에 모두가 동의할 수는 없으므로, 사람들이 그 과정을 신뢰할 수 있는 강력한 프로세스를 구축하는 것이 중요함.

[The Business Case for Responsible AI]

The Economist Intelligence Unit report - 책임 있는 **AI**의 비즈니스적 가치

- 경제적 잠재력**:** PricewaterhouseCoopers의 예측에 따르면, AI는 2030년까지 전 세계 GDP를 14% 증가시켜 최대 15.7조 달러의 가치를 창출할 것으로 기대됨.

- 성공의 핵심**:** Google은 AI의 책임 있고, 포용적이며, 공정한 배포가 이러한 성장을 실현하는 핵심 요소라고 믿음. 책임 있는 AI는 장기적인 신뢰를 바탕으로 성공적인 AI와 동의어임.
- 전략적 우위**:** 책임 있는 AI 프로그램과 실천은 비즈니스 리더에게 전략적이고 경쟁적인 이점을 제공함.

'Staying Ahead of the Curve: The Business Case for Responsible AI' 보고서

- 보고서 개요**:** Google이 후원하고 The Economist Intelligence Unit(EIU)이 개발한 보고서로, AI 중심의 세계에서 책임 있는 AI 관행의 가치를 제시함.
- 데이터 기반 연구**:** 보고서는 광범위한 데이터 기반 연구, 업계 전문가 인터뷰, 경영진 설문조사를 통해 개발되었으며, 개발자, 업계 리더, 최종 사용자의 의견을 반영함.
- 주요 비즈니스 영향**:** 보고서는 책임 있는 AI가 조직의 핵심 비즈니스 고려 사항에 미치는 영향을 7가지 영역으로 분류함.
    1. 제품 품질 향상**:** 제품의 품질을 높이는 효과.
    2. 인재 확보**,** 유지**,** 참여 증진**:** 인재를 끌어들이고, 유지하며, 이들의 참여도를 높이는 데 기여.
    3. 데이터 관리**,** 보안 및 개인정보 보호 개선**:** 더 나은 데이터 관리, 보안, 개인정보 보호에 이바지.
    4. 규제 준수**:** 현재 및 미래의 AI 규제에 대한 준비 태세 확보.
    5. 수익 증대**:** 매출과 순이익 성장에 기여.
    6. 이해관계자 및 투자자 관계 강화**:** 이해관계자 및 투자자와의 관계를 돈독히 하는 데 도움.
    7. 강력한 신뢰와 브랜드 이미지 유지**:** 신뢰와 브랜드 이미지를 강력하게 유지하는 효과.

The business case for responsible innovation - 책임 있는 **AI**의 비즈니스적 가치 **(7**가지 하이라이트**)**

1. 제품 품질 향상

- 윤리적 검토**:** EIU 설문 응답자의 **97%**가 윤리적 AI 검토가 제품 혁신에 중요하다고 동의함. 이 검토는 데이터 세트, 하위 그룹별 모델 성능, 의도된 결과 및 의도치 않은 결과의 영향을 면밀히 조사함.
- 위험 감소**:** 책임 있는 AI 관행을 조기에 도입하면, 윤리적 위반으로 인한 제품 출시 지연, 개발 중단, 심지어 시장 철수와 같은 위험을 줄여 개발 비용을 절감하는 효과.
- 신뢰 구축**:** 책임 있는 AI는 편향으로 인한 해를 줄이고, 투명성을 개선하며, 보안을 강화함으로써 제품을 향상시킴. 이는 이해관계자와의 신뢰를 구축하고, 제품 가치와 경쟁 우위를 높이는 핵심 요소.

2. 인재 확보 및 유지

- 최고 인재 유치**:** 오늘날의 최고 인재들은 역동적인 업무와 좋은 급여 이상의 것을 추구함. 특히 윤리적 문제와 같은 자신들의 가치와 부합하는 이슈를 다루는 고용주에 대한 충성도가 더 높음.
- 생산성 향상**:** 최고 인재는 평균보다 생산성이 400% 더 높으며, 소프트웨어 개발과 같은 복잡한 직업에서는 800%까지 높음.
- 비용 절감**:** 우수한 인재를 교체하는 데는 막대한 비용이 소요되므로, 책임 있는 AI 관행을 통해 직원 신뢰와 참여를 구축하는 것은 인재를 유지하는 효과적인 방법.

3. 데이터 약속 보호

- **AI** 채택의 장애물**:** EIU 설문조사에 따르면, 사이버 보안 및 데이터 프라이버시 우려는 AI 채택의 가장 큰 장애물임.
- 소비자 신뢰**:** 90% 이상의 소비자가 자신의 데이터 사용 방식에 대한 우려가 있을 경우 해당 기업의 제품을 구매하지 않겠다고 밝힘.
- 데이터 침해의 막대한 비용**:** 데이터 침해는 평균 392만 달러의 비용을 초래하며, 비즈니스 손실이 가장 큰 재정적 피해를 줌.
- 투자 수익**:** 데이터 프라이버시 강화에 1달러를 투자하면 평균 2.70달러의 수익을 얻을 수 있다는 연구 결과.

4. AI 규제에 대한 선제적 대비

- 규제 도입**:** 전 세계적으로 AI 규제에 대한 요구가 높아지고 있으며, 유럽연합 등 여러 정부가 AI 규제 법안을 추진 중임.
- 경쟁 우위**:** 책임 있는 AI를 개발하는 조직은 새로운 규제가 발효될 때 규제 미준수 위험을 줄이고, 심지어 규제 논의에 생산적으로 기여함으로써 상당한 이점을 얻을 수 있음.
- **GDPR** 사례**:** GDPR(개인정보보호법) 사례에서 규제 미준수 비용은 준수 비용보다 2.71배 높았음. 이는 AI 규제에 대한 사전 대비의 중요성을 시사함.

5. 수익 성장 증진

- 시장 확대**:** AI 공급업체에게 책임 있는 AI는 더 큰 목표 시장, 경쟁 우위, 그리고 기존 고객과의 관계 개선으로 이어짐.
- 기업 윤리**:** 조사에 따르면, 91%의 경영진이 윤리적 고려 사항을 제안 요청(RFP) 프로세스에 포함하고 있으며, 66%는 윤리적 우려 때문에 특정 AI 공급업체와 협력을 거부한 경험이 있음.
- 윤리적 행동과 재무 성과**:** ESG(환경, 사회, 지배구조)에 투자하는 기업이 주식 시장에서 더 좋은 성과를 보였으며, '세계에서 가장 윤리적인 기업'은 5년간 대형주 지수를 14.4% 초과 달성함.
- 소비자 행동**:** 닐슨 설문조사에 따르면, 66%의 소비자가 지속 가능하고 윤리적으로 설계된 제품과 서비스에 대해 기꺼이 더 많은 비용을 지불할 의향이 있음.

6. 파트너십 강화

- 가치 기반 투자**:** 투자자들은 개인적 가치에 따라 포트폴리오를 조정하려는 경향이 증가하고 있으며, 지속 가능한 투자에 대한 관심이 높아짐.
- **ESG**와 책임 있는 **AI:** ESG 평가 기준에 아직 책임 있는 AI가 포함되지 않았지만, 사회적 책임 기업에 대한 투자 경향은 책임 있는 AI를 우선시하는 기업에 자금이 재배치될 것임을 시사함.

- 자금 조달 증가**:** 책임 있는 AI 스타트업에 대한 투자는 2013년 800만 달러에서 2020년 3억 3500만 달러로 급증함.

7. 강력한 신뢰 및 브랜드 유지

- 평판 리스크 감소**:** 책임 있는 AI 관행이 없는 기업은 대중의 부정적 여론, 브랜드 가치 훼손, 부정적 언론 보도와 같은 위험에 노출됨.
- 브랜드 가치 향상**:** 책임 있는 AI 관행은 고객 신뢰와 충성도를 약화시키는 것을 막고, 관련 조직 및 브랜드 가치를 높이는 잠재력을 가짐.
- 미래를 위한 결정**:** 책임 있는 AI는 기업에 도덕적 의무와 더불어, 명백한 비즈니스적 가치를 제공함. 오늘날의 책임 있는 결정은 미래에 발생할 수 있는 부정적 결과를 예방하는 기회가 됨.

[AI's Technical Considerations and Ethical Concerns]

AI's technical considerations and ethical corners

윤리적 딜레마와 **AI**의 윤리

- 윤리적 딜레마의 정의**:** 두 가지 이상의 행동 방침 중 하나를 선택해야 하지만, 각 선택이 모두 도덕적 원칙을 위반하게 되는 어려운 상황. 결정을 내리지 않는 것 또한 하나의 결정임.
- 도덕적 유혹과의 구분**:** 윤리적 딜레마는 올바른 것과 그른 것 중 하나를 선택하는 '도덕적 유혹'과는 다름. 도덕적 유혹은 잘못된 행동이 자신에게 이익이 될 때 발생함.
- **AI**와 윤리**:** AI는 사회에 미치는 영향이 크기 때문에, AI 개발 과정에서 많은 윤리적 딜레마에 직면할 수 있음. 따라서 AI 커뮤니티에서는 윤리에 대한 논의가 최우선 과제여야 함.

윤리의 개념과 역할

- 윤리의 정의**:** 가치를 명확히 하고, 그 가치에 기반하여 결정을 내리고 정당화하는 지속적인 과정. 궁극적으로 사회 구성원 모두가 함께 번영할 수 있게 함.
- 주관성 및 다양성**:** 윤리에는 주관성과 문화적 상대성이 존재하며, 다양한 관점과 경험을 바탕으로 한 윤리적 논의가 중요함.
- 규칙이나 체크리스트의 한계**:** 윤리는 법이나 정책과 달리 규칙이나 체크리스트로 해결하기 어려운 새로운 도덕적 문제에 직면했을 때 독창적인 해결책을 필요로 함.
- 법과 윤리의 차이**:** 윤리는 법률이나 공식적인 시스템에 의해 강제되지 않는, 서로에게 기대하는 가치와 행동을 반영함. 비윤리적이지만 합법적인 행위가 있는 반면, 영웅적인 시민 불복종처럼 윤리적이지만 불법적인 행위도 존재함.

책임 있는 **AI**의 부상

- 기술의 위험성**:** 21세기 기술의 사회, 정치, 환경적 영향이 빠르게 확대되면서, AI 기술은 의도치 않은 해악을 놀라운 속도와 규모로 복제할 수 있는 잠재력을 지니고 있음.
- 경영진의 인식 변화**:** Capgemini의 2020년 설문조사에 따르면, AI 관련 문제점을 인식하는 경영진의 수가 2019년에 비해 두 배 증가함.
- 윤리적 헌장 채택**:** AI 개발에 대한 가이드라인을 제공하는 '윤리적 헌장'을 정의한 조직의 비율이 같은 기간 동안 5%에서 45%로 크게 증가함.

신뢰 구축의 중요성

- 윤리와 신뢰의 관계**:** 조직의 윤리 정의는 사용자와 팀, 그리고 사회 전반과의 신뢰 관계를 형성하는 데 필수적임.
- 관계의 기반**:** 이러한 신뢰 없이는 강력한 고객 관계가 존재할 수 없음

Concerns about artificial intelligence - AI의 주요 윤리적 문제 및 과제

- 투명성 부족**:** AI 시스템이 복잡해짐에 따라 의사결정 과정을 이해하기 어려워지며, 이는 사용자에게 정보에 입각한 선택을 방해하고, 개발자가 시스템의 실패나 의도치 않은 해를 예측하기 어렵게 만듦.
- 불공정한 편향**:** AI는 사회에 존재하는 편향을 드러내고 증폭시킬 수 있음. 특히 학습 데이터의 특정 집단(인종, 성별 등)에 대한 과소 또는 과대 대표성은 심각한 문제로, 감시 시스템이 소수 집단을 범죄자로 오인할 가능성과 같은 해를 초래할 수 있음.
- 보안 취약점**:** AI 시스템은 사이버 공격에 악용될 수 있는 잠재적 취약점을 가짐. AI는 딥페이크와 같은 새로운 형태의 조작 기술을 가능하게 하며, 이는 사회의 안전과 보안에 심각한 영향을 미칠 수 있음.
- 개인정보 침해**:** AI는 방대한 양의 데이터를 신속하게 수집, 분석, 결합할 수 있는 능력을 바탕으로 데이터 악용, 원치 않는 추적, 안면 인식 및 프로파일링과 같은 사생활 침해 위험을 높임.
- **AI** 유사과학**:** 과학적 근거가 부족한 AI 시스템이 마치 신뢰할 수 있는 것처럼 포장되어 해를 끼칠 수 있음. 예를 들어, 얼굴 특징만으로 범죄 성향이나 신뢰도를 판단하는 알고리즘은 개인과 커뮤니티에 해를 가하고, AI의 유익한 사용 사례에 대한 신뢰를 훼손할 수 있음.
- 책임성 부재**:** AI 시스템은 모든 사람의 요구와 목표를 충족하도록 설계되어야 하지만, 이를 위한 명확한 목표, 투명성, 인간의 개입 가능성 등이 부족할 수 있음.
- **AI**로 인한 실업 및 탈숙련화**:** AI가 일상적인 작업을 자동화하면서 실업을 초래하고 인간의 능력을 저하시킬 수 있다는 우려가 존재함. 하지만 과거 기술 혁신처럼 새로운 직업 창출의 기회도 함께 가져올 것.

생성형 **AI**의 고유한 문제

- 환각 **(Hallucinations):** AI 모델이 현실과 동떨어지거나 완전히 조작된 콘텐츠를 생성하는 현상.
- 사실성 **(Factuality):** 생성형 AI 모델이 생성한 정보의 정확성과 진실성과 관련된 문제.
- 의인화 **(Anthropomorphization):** AI 모델에 인간과 같은 특성이나 행동을 부여하여 사용자가 AI를 실제 인간처럼 인식하게 만드는 문제.

윤리적 문제의 원인과 해결책

- 주요 원인**:** Capgemini 설문조사에 따르면, 윤리적 문제가 발생하는 주요 원인으로는 윤리적 **AI** 시스템에 할당된 리소스 부족**,** 다양성이 부족한 개발팀**,** 윤리적 **AI** 행동 강령 부재 등이 있음. 또한 AI를 서둘러 도입해야 한다는 압박감도 중요한 원인으로 지적됨.
- 사회적 이점**:** 윤리적 문제는 비난의 대상이 아니라, AI가 재료 개발, 의학적 발견, 예측 시스템, 비용 효율적인 상품 및 서비스 제공 등 사회적으로 유익하게 사용될 기회를 찾는 데 초점을 맞춰야 함.
- 거버넌스의 역할**:** 책임 있는 AI 거버넌스 및 프로세스를 구축하는 것은 윤리적 문제를 해결하고, 고객과 사회 전반에 해를 끼치지 않도록 하는 핵심적인 방법임.

[Creating AI Principles]

How Google’s AI principles were developed - Google의 AI 원칙 수립 과정

- 수립 배경**:** Google의 오랜 미션과 가치를 바탕으로, AI의 책임 있는 개발과 활용을 위해 구체적인 지침이 필요했음. 이는 연구, 제품 개발, 비즈니스 결정에 적용될 수 있는 **AI** 원칙으로 구체화됨.
- 업계 동향**:** Berkman Klein Center 보고서와 Capgemini 연구에 따르면, 책임 있는 AI 사용에 대한 가이드라인을 개발한 조직의 수가 2019년에서 2020년 사이에 40% 증가하는 등 AI 원칙 수립이 활발해지는 추세.
- 수립 과정**:**
    1. 시작**:** 2017년 여름, Sundar Pichai CEO가 Google을 'AI 우선 회사'로 선언하면서 시작됨.
    2. 전문가 그룹 구성**:** AI 기술 전문성뿐만 아니라 사용자 연구, 법률, 공공 정책, 개인정보 보호, 온라인 안전, 지속 가능성 등 다양한 배경과 인구 통계를 가진 교차 기능 전문가 그룹을 구성함.
    3. 광범위한 의견 수렴**:** 다양한 국가, 성별, 인종, 민족, 연령대 사람들의 의견을 반영하고, 핵심 그룹 외의 팀과 외부 전문가로부터도 피드백을 받아 폭넓은 관점을 통합함.
    4. 연구 기반 원칙 초안 작성**:** AI에 대한 사람들의 우려, 학술 연구, 미디어, 심지어 대중문화(SF 소설, TV 쇼)까지 조사하여 비책임적인 **AI**의 기준을 파악하고, 이를 바탕으로 원칙 초안을 작성함.
    5. 반복적인 검증 및 세분화**:** 초안을 외부 전문가들에게 검토받고, 피드백을 통해 원칙 목록을 통합하고 세분화하는 과정을 거침.

AI 원칙의 결과와 의의

- 원칙의 내용**:** Google의 AI 원칙에는 AI에 대한 7가지 '목표'와 함께, 명확한 가이드라인을 제공하기 위해 추구하지 않을 **4**가지 **AI** 애플리케이션 목록이 포함됨. 만들지 않을 것을 명확히 하는 것이 무엇을 만들지 정의하는 것만큼 중요함을 강조.

- 실제 적용**:** 2018년 6월 원칙을 발표한 이후, 이는 일상 대화와 제품 개발 프로세스에 통합되어 모든 Google 직원이 의사결정을 내릴 때 공유하는 윤리적 약속으로 기능함.
- 지속적인 발전**:** 책임 있는 AI 분야는 계속 진화하고 있으며, Google은 이 분야의 선구적인 연구와 옹호자 커뮤니티 덕분에 많은 발전을 이루었음. Google 역시 끊임없이 배우고 방법을 개선해 나갈 것.
- 조직별 맞춤화**:** 회사의 미션, 가치, 지리적 위치, 조직 목표에 따라 원칙이 달라질 수 있음. 고객 지원 챗봇 회사와 광범위한 컨설팅 회사의 AI 원칙이 다를 수 있는 것처럼, 각 조직에 맞는 원칙을 만드는 것이 중요함.
- 궁극적 목표**:** 조직의 AI 원칙은 그 조직의 정신을 전달하고, AI 거버넌스의 기반을 제공하는 역할을 함.

Ethical issue spotting - AI 거버넌스와 윤리적 문제 파악

- 문제 파악 **(Issue Spotting)**의 중요성**:** AI 거버넌스의 핵심 요소로, AI 프로젝트에서 발생할 수 있는 잠재적인 윤리적 문제를 인식하는 과정. 조직의 **AI** 원칙은 이러한 문제를 식별하는 가이드 역할을 함.
- 체크리스트의 한계**:** 윤리적 문제 파악을 체크리스트로 효율화하려는 시도는 실패함. 새로운 기술과 함께 등장하는 예측 불가능한 위험과 고유한 상황(사용 사례, 고객, 사회적 맥락) 때문에 경직된 체크리스트는 적합하지 않음.
- 적응형 프로세스의 필요성**:** 모든 경우에 맞는 단순한 체크리스트를 만드는 것은 불가능하며, 각 사안의 사실관계를 신중하게 검토하는 것이 중요함. 새로운 기술의 빠른 발전 속도에는 고정된 답이 아닌 적응형 프로세스가 요구됨.

문제 파악을 위한 '렌즈' 활용

- 조류 관찰 비유**:** 윤리적 문제를 조류에 비유하여, 훈련된 '조류 관찰자'처럼 연습을 통해 문제를 더 쉽고 정확하게 식별할 수 있음을 강조함.
- 다양한 관점의 중요성**:** 한 개인이 모든 것을 볼 수 없으므로, 여러 사람이 함께 검토하는 것이 효과적.
- 윤리적 렌즈**:** 도덕 철학자들이 수천 년 동안 개발해 온 윤리적 렌즈는 윤리적 문제를 다양한 각도와 관점에서 구조적으로 검토하는 데 도움을 줌. 모든 상황에 하나의 접근 방식을 고집하기보다는, 여러 렌즈를 활용하여 의사결정의 결과를 평가하고, 인권 및 의무에 미치는 영향을 파악하며, 올바른 인격과 일치하는지를 검토하는 방식이 효과적.
- 추가 학습**:** 윤리적 렌즈에 대해 더 자세히 알아보려면 Markkula Center for Applied Ethics의 자료를 참고할 수 있음.

The ethical aims of Google’s AI principles

Google의 **7**가지 **AI** 원칙과 윤리적 목표

1. 사회에 이롭도록

- 목표**:** 건전한 사회 시스템과 제도를 지원하고, 고용, 주택, 교육 등 필수 서비스에 대한 불공정한 거부를 막는 것을 목표로 함. 취약 계층에 대한 의도치 않은 해악과 위험을 줄이는 데 중점을 둠.

2. 불공정한 편견을 만들거나 강화하지 않도록

- 목표**:** 사람과 그룹에 대한 공정하고, 정의로우며, 공평한 대우를 촉진하는 AI를 개발하는 것. 학습 데이터에 존재할 수 있는 역사적 편견의 영향을 최소화하고, 모든 사용자가 제품을 유용하게 사용할 수 있도록 함.
- 불공정성의 발생**:** 불공정성은 문제 정의, 데이터 수집, 모델 훈련, 사용 방식 등 머신러닝 생애 주기의 모든 단계에서 발생할 수 있음.
- 해결을 위한 질문**:** 모델의 목적, 대상 사용자, 데이터의 수집 및 라벨링 방식, 모델의 테스트 및 검증 방식 등 각 단계에서 어려운 질문을 던지는 것이 핵심.

3. 안전하게 구축하고 테스트하도록

- 목표**:** 개인과 공동체의 안전(신체적 무결성 및 건강)과 시스템, 인프라의 보안을 보장하는 AI를 만드는 것. 안전에 중요한 애플리케이션에 대한 효과적인 감독과 테스트, AI 시스템 행동에 대한 제어, 기계 지능에 대한 의존도 제한을 목표로 함.

4. 사람에게 책임을 다하도록

- 목표**:** 사람의 권리와 독립성을 존중하고, 사용자가 AI 상호작용에서 거부할 수 없는 상황을 제한하는 것. 사용자의 동의를 얻고, 오용, 부당한 사용 또는 오작동에 대한 보고 및 시정 조치를 위한 경로를 제공하는 데 중점을 둠.

5. 개인정보 보호 원칙을 포함하도록

- 목표**:** 개인과 그룹의 개인정보 및 안전을 보호하는 것. 개인 식별 정보와 민감한 데이터를 특별히 취급하고, 사용자가 데이터 사용에 대해 명확하게 인지하고 동의할 수 있도록 보장함.

6. 과학적 우수성의 높은 기준을 지키도록

- 목표**:** AI 분야의 지식을 발전시키기 위해 과학적으로 엄격한 접근 방식을 따르는 것. 개방적 탐구, 지적 엄격성, 진실성, 협력을 통해 과학적 주장의 신뢰도를 확보함. 교육 자료와 연구 결과를 공유하여 AI 유사과학을 피하고 유익한 AI 애플리케이션 개발을 돕는 것을 포함.

7. 이러한 원칙에 부합하는 용도로만 사용 가능하도록

- 목표**:** Google이 사회에 미치는 고유한 영향에 대한 책임을 다하는 것. 잠재적으로 유해하거나 악용될 수 있는 애플리케이션을 제한하고, 기술이 유해한 용도로 적용될 가능성을 고려함. Google은 자체 기술뿐만 아니라 고객과 파트너에게 제공하는 기술도 AI 원칙을 준수하도록 보장함.

AI 애플리케이션을 추구하지 않을 **4**가지 영역

- 전반적인 해악을 초래할 가능성이 있는 **AI** 애플리케이션**.**
- 주요 목적이 사람에게 상해를 입히는 무기 또는 기술**.**
- 국제적으로 인정된 규범을 위반하는 감시 기술**.**
- 국제법 및 인권을 위반하는 목적을 가진 기술**.**

원칙의 실천

- 거버넌스 구축**:** AI 원칙을 실행하기 위해 새로운 프로젝트, 제품, 거래에서 발생하는 다면적인 윤리적 문제를 평가하기 위한 공식적인 검토 프로세스와 거버넌스 구조를 확립.
- 고려와 절충**:** AI 원칙은 어떤 경우에는 상충할 수 있으므로, 원칙을 적용하고 위험을 완화하기 위한 신중한 고려와 절충이 필요함.
- 지속적인 실천**:** 원칙 수립은 첫 단계일 뿐이며, 이를 실제에 적용하기 위한 프로세스와 프로그램이 중요함.

[Operationalizing AI Principles: Setting Up and Running Reviews]

Google’s AI Governance

- **AI** 거버넌스 및 검토 프로세스

AI 원칙 실행의 중요성

- 지침으로서의 원칙**:** AI 원칙은 모든 윤리적 문제에 대한 즉각적인 답을 제공하지 않지만, 기술 개발에서 평가해야 할 가치와 기준을 정립하는 출발점 역할을 함.
- 지속적인 노력**:** 원칙을 실제로 적용하려면 지속적이고 꾸준한 노력이 필요하며, 명확한 책임 목표가 있을 때 기술 도구의 효용성이 극대화됨.
- 책임 있는 문화 조성**:** 전통적인 제품 개발 주기에서 부족하기 쉬운 책임 있는 AI 문화를 조성하는 데 전념하는 프로세스가 필요함.

일반적인 오해와 극복 방안

- 오해 **1:** 윤리적인 사람을 고용하면 윤리적인 **AI**가 보장된다**.**
    - 현실**:** 윤리적인 사람이라도 경험과 배경에 따라 같은 문제에 대해 다른 결론을 내릴 수 있음. '윤리적 맹점'에 빠질 수 있으므로, 윤리적 의사결정을 위한 실천과 논의의 장을 마련하는 것이 중요함.
- 오해 **2:** 책임 있는 **AI**를 위한 체크리스트를 만들 수 있다**.**
    - 현실**:** AI 기술은 빠르게 발전하고, 각 제품의 기술적 세부사항과 사용 맥락이 고유하므로, 단순한 체크리스트는 효과적이지 않음. 체크리스트는 비판적 사고를 제한하고 윤리적 맹점을 유발할 수 있음.

    - 대안**:** 윤리적 문제 해결을 위한 '도덕적 상상력'을 발휘하고, 문제 파악 능력을 기를 수 있도록 지원하는 프로그램과 실천 방안이 필요함.

Google의 **AI** 거버넌스 위원회 구조

- 구조적 접근**:** Google은 AI 원칙에 부합하도록 새로운 프로젝트, 제품, 거래를 평가하기 위한 공식적인 검토 위원회 구조를 구축함.
- **1.** 중앙 **'**책임 있는 혁신 팀**'**:
    - 역할**:** 다양한 제품 영역에 걸쳐 AI 원칙에 대한 공통된 해석을 확립하고, 일상적인 운영과 초기 평가를 담당함.
    - 구성**:** 사용자 연구원, 사회 과학자, 윤리학자, 인권 전문가, 정책 및 개인정보 보호 고문 등 다양한 전문가들로 구성됨.
- **2.** 고위 전문가 그룹**:**
    - 역할**:** 기술적, 기능적, 애플리케이션 전문 지식을 제공하고, 신흥 기술에 대한 전략과 가이드라인을 수립함. 필요한 경우 검토에 자문을 제공함.
- **3.** 고위 경영진 협의회**:**
    - 역할**:** 여러 제품과 기술에 영향을 미치는 가장 복잡하고 어려운 문제를 다루는 최상위 의사결정 기구. 선례를 만드는 결정을 내리고, 최고 수준의 책임을 부여함.
- **4.** 임베디드 검토 위원회**:**
    - 역할**:** 특정 제품 영역 내부에 맞춤형으로 존재하며, 해당 제품의 기술, 사용 사례, 학습 데이터, 사회적 맥락 등 고유한 상황을 고려하여 '책임 있는 혁신 팀'과 긴밀하게 협력함.
- 핵심 원칙**:** 모든 검토 과정에서 다양한 사람들의 참여를 보장하고, 심리적 안전이 확보된 환경에서 토론과 토의가 이루어져야 신뢰성 있고 강력한 결과물을 얻을 수 있음.

Google Cloud’s review process

- **Google Cloud**의 검토 프로세스

개요

- 목적**:** Google Cloud는 AI 플랫폼(Vertex AI), MLOps, API 등을 통해 기업의 AI 구축을 돕는 다양한 기술을 제공하며, 이러한 기술의 윤리적 영향을 평가하는 맞춤형 AI 원칙 검토 프로세스를 운영함.
- 두 가지 검토 체계**:** 책임 있는 AI 개발을 위해 고객 **AI** 거래 검토와 클라우드 **AI** 제품 개발 검토라는 두 개의 연결되면서도 명확히 구분된 검토 체계가 존재함.
- 핵심 질문**:** 제안된 사용 사례가 AI 원칙에 부합하는지, 그리고 의도된 이점을 실현하고 위험을 완화하기 위해 솔루션을 어떻게 설계하고 통합해야 하는지에 대한 질문에 답하는 것을 목표로 함.

1. 고객 **AI** 거래 검토

- 목표**:** 거래가 진행되기 전에 AI 원칙과 충돌할 위험이 있는 사용 사례를 식별하는 것.
- 단계별 프로세스**:**
    1. 영업 거래 제출**:** 영업 담당자가 고객 기회를 제출하거나, 자동화된 프로세스가 거래를 검토 대상으로 지정함.

2. 예비 검토**:** Cloud AI 원칙 팀이 제출된 거래를 검토하고, 심층 검토가 필요한 거래의 우선순위를 정함. 이 단계에서는 관련 선례를 적용하고, 잠재적 위험을 논의하며, 추가 정보를 요청함.
3. 검토**,** 논의 및 결정**:** 제품, 정책, 영업, AI 윤리, 법률 등 여러 부서의 리더들로 구성된 위원회가 모여 특정 거래와 사용 사례에 AI 원칙이 어떻게 적용되는지 논의하고 최종 결정을 내림. 결정은 '진행', '진행 불가', '특정 조건 충족 시 진행', '상위 위원회로 이관' 등으로 이루어짐.

2. 클라우드 **AI** 제품 개발 검토

- 목표**:** 제품이 출시되기 전에 고급 기술을 사용하는 제품을 어떻게 평가, 범위 설정, 구축하고 관리할지 결정하는 것.
- 단계별 프로세스**:**
    1. 파이프라인 개발**:** Cloud AI 원칙 팀이 제품 개발 초기 단계부터 검토가 이루어지도록 제품 파이프라인을 추적함. 이는 '설계부터 윤리'를 적용하는 데 중요함.
    2. 예비 검토**:** 출시 일정 또는 위험도에 따라 검토 우선순위를 정함.
    3. 검토 브리프 준비**:** 검토 회의 전에 Cloud AI 원칙 팀이 제품 관리자, 엔지니어, 윤리 전문가 등과 협력하여 제품의 목표, 데이터, 잠재적 위험 등을 심층적으로 분석한 검토 브리프를 작성함. 이 브리프는 영향을 받는 모든 이해관계자 그룹을 고려하고, 윤리적 딜레마를 논의하며, 잠재적 해악을 해결하기 위한 '정렬 계획(alignment plan)'을 포함함.
    4. 논의 및 정렬**:** 위원회 구성원들이 검토 브리프를 미리 숙지한 후, 회의에서 라이브로 제품을 검토함. 이 과정을 통해 추가적인 윤리적 문제를 발견하고, 책임 있는 AI를 제품 설계에 통합하는 결정을 내림.
    5. 승인**:** 최종 정렬 계획이 위원회와 제품 리더의 승인을 받은 후, 제품 개발 로드맵에 통합되고, AI 원칙 팀이 그 실행을 추적함.

검토 프로세스의 결과 및 특징

- 정렬 계획 **(Alignment Plan):** 각 제품 또는 솔루션에 고유하며, 기술적 해결책뿐만 아니라 기술의 목적을 축소하거나**,** 특정 고객에게만 제공하거나**,** 책임 있는 사용을 위한 교육 자료를 제공하는 등의 비기술적 조치를 포함할 수 있음.
- 지속적인 개선**:** 비슷한 문제에 대한 반복적인 검토를 통해 특정 정책이 수립되고, 이는 향후 검토 프로세스를 간소화하는 선례가 됨.
- 성장과 진화**:** 이 프로세스는 시간이 지남에 따라 성장하고 진화하고 있으며, Google은 이 프레임워크가 다른 조직에도 유용하게 적용될 수 있기를 바람.

Celebrity Recognition Case Study

- **Google Cloud**의 얼굴 인식 기술 사례 연구

사례 연구 개요

- 결과**:** 2019년, Google Cloud는 미디어 및 엔터테인먼트 고객을 대상으로 전문가가 촬영한 라이선스 콘텐츠에 등장하는 유명인을 식별하는 **Celebrity Recognition(**유명인 인식**)** API를 출시함. 이는 Google Cloud의 첫 번째 엔터프라이즈용 얼굴 인식 제품으로, 맞춤화가 불가능한 사전 훈련된 AI 모델임.

- 배경**:** 얼굴 인식 기술은 불공정한 편견을 유발할 수 있는 주요 문제로 인식되었으며, Google Cloud는 고객의 요청에도 불구하고 2016년 초기부터 Cloud Vision API에 이 기능을 포함하지 않기로 결정함.

AI 원칙 검토 프로세스

- 윤리적 우려**:** Google은 얼굴 인식 기술의 오용에 대한 광범위한 우려를 공유했으며, 특히 다음 세 가지를 핵심적으로 고려함.
    1. 공정성**:** 기존 편견을 강화하거나 증폭시키지 않아야 하며, 특히 소수 집단에 영향을 미치지 않도록 함.
    2. 감시**:** 국제적으로 용인되는 규범을 위반하는 감시에 사용되지 않아야 함.
    3. 개인정보 보호**:** 적절한 투명성과 제어를 제공하여 사람들의 개인정보를 보호해야 함.
- 제한적 범위 설정**:** 이러한 우려를 완화하고 기술을 AI 원칙에 부합하는 기업용 사례에 활용하기 위해, Google은 유명인 인식이라는 매우 제한된 범위의 애플리케이션을 개발하기로 결정함.
- 외부 전문가 협력**:** 제품 출시 준비를 위해 외부 전문가 및 시민권 단체(Business for Social Responsibility, BSR)와 협력하여 심층적인 인권 영향 평가를 수행함. 이 평가는 일반적인 얼굴 인식 API를 제공하지 않기로 한 초기 결정을 재확인시켜 줌.

기술적 및 윤리적 조치

- 안전장치 구현**:** BSR의 권장사항을 바탕으로 다음과 같은 안전장치를 구현함.
    - 자격을 갖춘 고객에게만 API를 허용 목록**(allow list)** 방식으로 제공.
    - 유명인 데이터베이스를 신중하게 정의하고 사전 정의된 목록으로 제한.
    - 유명인이 목록에서 제외를 요청할 수 있는 옵트아웃**(opt-out)** 정책 도입.
    - API에 대해 확장된 서비스 약관 적용.
- 공정성 분석**:** 여러 차례의 공정성 테스트를 통해 API 성능을 피부 톤과 성별 그룹별로 평가함.
    - 오류 원인 발견**:** 피부 톤 라벨의 부정확성과, 소수의 배우들이 오인식의 상당 부분을 차지하고 있음을 발견함. 특히 어두운 피부 톤을 가진 남성 배우들에게 오류율이 높았음.
    - 문제 해결**:** Joy Buolamwini와 Timnit Gebru의 **'Gender Shades'** 연구에서 사용된 Fitzpatrick 스킨 톤 척도에 따라 라벨을 재조정함. 또한, 훈련 데이터 세트에 유명인들의 다양한 연령대 이미지를 추가하여 모델이 나이가 든 모습을 인식하지 못하는 문제를 해결함.
- 결론**:** 미디어 내의 재현**(representation)** 문제를 인식하고, 솔루션의 범위를 엄격하게 제한하며, 공정성을 위해 API를 철저히 테스트하고 개선한 후에야 비로소 API를 출시할 수 있었음.

결과 및 의의

- 성공적인 통합**:** 이 사례는 책임 있는 **AI** 개발이 곧 성공적인 **AI** 통합으로 이어진다는 것을 보여줌.
- 지속적 개선**:** 2020년 다른 기술 기업들이 얼굴 인식 사업을 제한하거나 중단한 상황에서도, Google은 AI 거버넌스 프로세스를 통해 AI 원칙에 부합하는 제품을 개발할 수 있었음.

- 새로운 도구 개발**:** 이 과정에서 얻은 교훈을 바탕으로, Google은 이미지 내 재현성을 더 잘 이해하는 데 도움이 되는 개선된 피부 톤 척도인 **Monk Skin Tone(MST)** 척도를 발표함.

[Operationalizing AI Principles: Issue Spotting and Lessons Learned]

Issue spotting process

- **AI** 거버넌스**:** 윤리적 문제 파악을 위한 질문 중심 접근법

문제 파악 **(Issue Spotting)**의 목적

- 개요**:** AI 거버넌스와 검토 과정의 핵심으로, AI 사용 사례의 잠재적인 윤리적 문제를 식별하는 과정.
- 접근 방식**:** Google은 단순한 체크리스트가 아닌, 개발자들이 기술에 대해 비판적으로 사고하도록 유도하는 질문을 중심으로 문제 파악을 진행함.
- 질문의 근원**:** 이러한 질문들은 윤리적 의사결정 프레임워크에 뿌리를 두고 있으며, 추가 정보 탐색과 최악 및 최선의 시나리오 고려를 강조하여 간과될 수 있는 윤리적 문제를 드러냄.

문제 파악을 위한 질문의 범위

- 제품 정의**:** 해결하려는 문제, 의도된 사용자, 사용되는 데이터, 모델 훈련 및 테스트 방식 등을 포함.
- 맥락**:** 사용 사례의 목적과 중요성, 사회적 이점, 오용 가능성 등을 포함.
- 기본 가정**:** 모든 AI 사용 사례에는 개선할 점이 있다는 가정하에, 사회적으로 명백히 이로운 경우에도 문제를 제기하며 비판적으로 접근함.
- 심층 검토**:** 문제 파악 과정에서 AI 원칙과 충돌할 수 있는 문제가 발견되면, 심층적인 검토가 이루어짐. 특히, 감시**(Surveillance)** 또는 **합성 미디어(Synthetic Media)**와 같은 복잡한 영역은 더 면밀한 검토가 필요함.

가상의 사례**: ASD Carebot**에 대한 문제 파악

- 가정**:** 미취학 아동을 위한 저렴한 클라우드 기반 AI 챗봇(음성, 제스처, 감정 분석, 맞춤형 학습 모듈 포함)을 개발한다고 가정.
- 원칙별 질문**:**
    - 사회적 이점**:** 이 기술이 최선의 치료법인지, 그리고 자폐 스펙트럼 장애(ASD)가 이러한 형태의 개입을 필요로 하는 문제인지.
    - 불공정한 편견**:** 개발팀의 구성, 공정성이 고려되어야 할 부분, 직접 영향을 받는 사람들의 의견 반영 여부, 훈련 데이터의 대표성 등.
    - 안전**:** 모델이 예상대로 작동하지 않거나 시간이 지남에 따라 성능이 저하될 경우 발생할 수 있는 안전 문제.

    - 개인정보 보호**:** Carebot이 수집할 데이터의 종류, 개인정보 보호 위험이 있는 데이터 세트, 그리고 민감한 사용 사례에 적합한 보호 설계 원칙.
    - 책임성**:** 시스템에 대한 인간의 감독을 어떻게 보장할지, 어떤 종류의 동의가 필요한지, 그리고 Carebot이 자신을 '친구'로 소개하는 것이 적절한지.
    - 과학적 우수성**:** 제품 개발에 필요한 전문성을 갖추었는지, 외부 전문가와 협력해야 할 필요가 있는지, 그리고 성능을 보장하기 위한 테스트 및 검토 기준.
    - 원칙에 부합하는 용도**:** 솔루션이 저렴하고 접근성이 좋아 광범위하게 사용될 수 있는지.

결론

- 가치**:** 문제 파악 질문을 통해 팀은 사용 사례의 잠재적 이점과 해악을 비판적으로 평가할 수 있음.
- 책임 있는 접근법 형성**:** 이러한 철저한 검토 과정을 거쳐야만 새로운 AI 애플리케이션에 대한 책임 있는 접근 방식을 수립할 수 있음.

What we've learned from operationalizing AI Principles: Challenges

- 책임 있는 **AI** 개발 과정에서의 도전 과제

1. 책임 있는 **AI**의 효과 측정 난이도

- 문제점**:** 책임 있는 AI의 효과를 측정하는 것은 기술적 성과를 평가하는 것보다 어려움. 특히, 잠재적 해악을 예방하는 완화 조치의 효과를 수치화하기 어려움.
- 해결책**:** 전통적인 비즈니스 지표와 다른 지표들을 사용함.
    - 추적**:** 문제 및 완화 조치, 그리고 제품에 적용된 방식 추적.
    - 측정**:** 고객 신뢰 구축 및 거래 성공 가속화에 대한 AI 거버넌스의 영향 측정.
    - 피드백 수집**:** 설문조사 및 고객 피드백을 통해 최종 사용자의 경험과 인식을 수집하여 영향력과 추세를 파악함.

2. 윤리적 딜레마와 가치 충돌

- 문제점**:** AI 원칙을 적용할 때 옳고 그름이 명확한 결정보다는 윤리적 딜레마가 자주 발생함. 위원회 구성원 각자의 가치관, 경험, 전문성이 충돌하여 많은 논쟁이 일어날 수 있음.
- 해결책**:** 이러한 딜레마와 숙고 과정 자체를 AI 원칙 검토의 핵심 목표로 받아들임. 솔직하고 열린 대화를 통해 선택에 따른 **상충 관계(trade-offs)**를 파악하는 것이 중요함.

3. 주관성 및 문화적 상대성

- 문제점**:** AI 원칙 적용이 주관적이거나 문화적으로 상대적으로 보일 수 있음.

- 해결책**:**
    - 명확한 프로세스**:** 잘 정의된 검토 및 의사결정 프로세스를 통해 신뢰를 구축함.
    - 현실 기반**:** 기술적, 연구적, 비즈니스 현실에 기반하여 완화 조치를 현실적인 문제에 연결함.
    - 문서화**:** 의사결정 과정을 문서화하여 투명성과 책임성을 확보함.
    - 선례 기록**:** 일관성을 위해 이전 사례에 대한 포괄적인 기록을 유지하고, 현재 사례가 이전 사례와 관련하여 어떻게 다른지 평가함.

4. 외부 전문가 및 이해관계자 참여의 어려움

- 문제점**:** 외부 전문가나 영향을 받는 그룹으로부터 직접적인 의견을 듣는 것이 매우 중요하지만, 현실적으로 어려움. 한 사람이 특정 그룹의 모든 관점을 대변할 수 없기 때문임.
- 해결책**:** 제품이 모든 사람을 위해 만들어지도록 최대한 광범위한 목소리를 듣기 위해 노력함.

결론

- 지속적인 도전**:** 책임 있는 AI 개발은 항상 문제와 도전 과제에 직면하는 여정임.
- 핵심**:** 이러한 문제를 인식하고, 최소화하며, 완화하기 위해 노력하는 것이 시작점임.

What we've learned from operationalizing AI Principles: Best practices

- **Google**의 책임 있는 **AI** 개발 모범 사례

1. 다양성을 갖춘 검토 위원회 구성

- 중요성**:** 문화적 배경, 전문 지식, 직급 등 다양한 관점을 가진 위원회를 구성하는 것이 중요함.
- 효과**:** 사용자 기반을 폭넓게 대변하는 위원회는 AI 원칙을 더 정확하게 해석하고, 실행 가능하며 실용적인 솔루션을 도출함.

2. 상향식 및 하향식 지원 확보

- 상향식 **(Top-down):** 고위 리더십의 AI 원칙 채택에 대한 공식적인 지지가 필요함.
- 하향식 **(Bottom-up):** 팀원들의 자발적인 참여와 의견이 책임 있는 AI 문화를 내재화하는 데 핵심적인 역할을 함.

3. 팀 교육을 통한 책임 있는 **AI** 문화 구축

- 실천**:** 제품 및 기술 팀에게 기술 윤리에 대한 교육을 제공하고, 비기술 직군에게는 AI가 사회와 비즈니스에 미치는 영향에 대한 이해를 높이도록 독려함.
- 결과**:** 윤리가 기술 개발 및 제품 우수성과 직접적으로 연결되는 기업 문화를 조성함.

4. 비즈니스 목표와 윤리적 목표의 일치

- 인식**:** 책임 있는 AI가 곧 성공적인 AI라는 점을 인식해야 함.
- 장기적 관점**:** 윤리적 문제 해결을 위해 때로는 속도를 늦추는 과정이 필요하지만, 모두에게 잘 작동하는 제품을 출시하는 것이 장기적으로 비즈니스와 사회에 이로움.

5. 거버넌스 프로세스의 투명성 추구

- 신뢰 구축**:** 프로세스와 참여 인력에 대한 투명성을 통해 신뢰를 구축함.
- 투명성과 기밀성의 균형**:** 개별 검토의 세부사항은 기밀을 유지해야 할 때도 있지만, 거버넌스 프로세스 자체에 대한 투명성은 신뢰성과 신뢰 구축에 기여함.

6. 결정 사항 및 선례 기록 관리

- 목적**:** AI 원칙 적용에 대한 결정, 완화 조치, 선례 등을 기록하는 시스템을 구축함.
- 효과**:** 미래의 작업과 검토를 위한 일관된 기준을 제공하고, 투명성을 확보하며, 책임 있는 AI 이니셔티브를 확장하는 데 도움이 됨.

7. 겸손한 접근 태도 유지

- 자세**:** AI 기술과 세계는 끊임없이 변화하므로, 항상 배우고 개선할 수 있다는 겸손한 태도를 가져야 함.
- 유연성**:** 일관성을 유지하면서도 새로운 연구와 의견에 개방적이고 반응하는 섬세한 균형을 유지해야 함.

8. 심리적 안전에 대한 투자

- 필요성**:** 팀원들이 위험을 감수하고 취약성을 드러낼 수 있는 심리적 안전을 확보하는 것이 중요함.
- 위험 탐색**:** '만약에(what if)'라는 질문을 자유롭게 탐색하고 오용 가능성을 논의해야 잠재적 문제를 발견할 수 있음.
- 실용성**:** '분석 마비(analysis paralysis)'를 피하기 위해 기술적, 비즈니스적, 사회적 현실에 기반하여 문제 파악을 진행해야 함.

9. 효율성보다 신중한 검토 우선

- 균형**:** 제품 개발 목표와 포괄적인 AI 검토에 필요한 시간 사이의 균형이 필요함.
- 결과**:** 효율성에만 집중하면 고객에게 해를 끼칠 수 있는 잠재적 문제를 놓칠 수 있음.
- 건전한 논쟁**:** 건전한 논쟁과 숙고를 통해 위험과 완화 방안을 탐색하는 것이 중요함.

10. 모든 **AI** 애플리케이션에 대한 관심

- 전제**:** 모든 AI 애플리케이션은 잠재적인 윤리적 문제를 가지고 있다는 가정에서 출발해야 함.
- 사고 확장**:** 명백히 이롭거나 무해해 보이는 사용 사례라도 '만약에'라는 질문을 통해 모든 시나리오를 탐색함으로써 포괄적인 완화 조치를 개발할 수 있음.
- 가이드 역할**:** AI 원칙 검토는 이러한 대화를 이끌어가는 프레임워크 역할을 함.

[Continuing the Journey Towards Responsible AI]

Continuing the journey towards responsible AI

- 책임 있는 **AI**의 중요성 및 실천

- 신념**:** Google은 책임 있는 AI를 구축하기 위한 엄격한 평가가 성공적인 AI를 만드는 데 필수적인 요소라고 믿음. 제품과 기술은 모든 사람에게 유용해야 함.
- **AI** 원칙의 역할**:** Google의 AI 원칙은 공동의 목표를 달성하도록 동기를 부여하고, 전 세계 사람들에게 최선의 이익이 되는 방향으로 기술을 사용하도록 이끌며, Google의 미션과 핵심 가치에 부합하는 결정을 내리도록 도움.

지식의 활용과 도전

- 모두의 역할**:** 책임 있는 AI 적용에는 모두가 역할을 해야 함. 이 과정을 통해 Google이 AI 원칙을 개발하고 조직 내에서 실행한 방법에 대한 이해를 얻기를 바람.
- 행동 촉구**:** 배운 교훈과 모범 사례를 바탕으로 팀과 협력하여 자신만의 **AI** 원칙과 검토 프로세스를 개발하는 도전 과제를 제시함.
- 토론의 가치**:** AI 여정의 어느 단계에 있든, 책임 있는 AI가 비즈니스 맥락에서 무엇을 의미하는지에 대해 팀과 토론하는 것은 AI 원칙을 수립하는 데 매우 유용한 과정.

지속적인 발전과 협력

- 완벽하지 않은 시스템**:** 인간이든 AI 기반이든 완벽한 시스템은 없으므로, 이를 개선하는 작업은 결코 끝나지 않음.
- 학습과 공유**:** Google은 책임 있는 AI와 관련하여 지속적으로 배우고 있는 내용을 Google 및 Google Cloud Responsible AI 페이지를 통해 업데이트하고 공유할 것.
- 협력 방안**:** Google과의 다음 프로젝트나 비즈니스 목표를 위해 Google Cloud 담당자나 ML 전문 파트너에게 문의하거나, 책임 있는 AI 관련 질문은 Google Cloud 책임 있는 AI 팀에 직접 문의 가능.
- 마무리**:** 책임 있는 AI에 대한 Google의 변함없는 헌신을 다시 한번 강조하며, 이 여정에 함께해 준 것에 대한 감사를 표함.