윤리경영연구 2024년 12월, 제24권 제2호, pp. 099-125
http://dx.doi.org/10.34273/kjbe.2024.24.2.004

인공지능 윤리와 기업의 책임있는 AI 활동

김지혜[1]·조봉순[2]

AI 기술의 급속한 발달에 따라 기업들은 AI를 활용하여 운영 효율성을 향상시키고 새로운 비즈니스 모델을 창출하고 있다. 하지만 AI 활용이 증가하면서 예상치 못한 문제들도 발생하고 있는데 알고리즘 편향, 개인정보 침해, 저작권 문제 등과 같은 윤리적 문제들이 나타나고 있다. 그 결과 AI 기술과 서비스를 제공하는 기업에 대해서는 단순히 기술적 성능의 향상뿐만 아니라 AI를 책임감 있게 사용하며 사회적, 윤리적 책임을 다해야 한다는 사회적 요구가 커지고 있다. 기업들은 AI 기술을 책임감 있게 개발하고 배포하기 위해 '책임있는 AI(Responsible AI)'라는 이름으로 AI의 윤리적 측면에 대한 기업의 대응활동 방향을 마련하고 있다. 본 연구는 먼저 AI와 관련된 윤리적 원칙으로 중시되는 내용들을 살펴본다. 그리고 AI 윤리적 원칙을 실천하기 위한 기업의 '책임있는 AI' 활동을 AI 라이프사이클에 따라 전략 및 계획수립 단계, 생태계 구축 단계, 개발 및 배포 단계, 모니터링 및 보완의 4단계로 구분하여 단계별 주요 활동 내용을 소개한다.

핵심어: 인공지능, AI 윤리, 책임있는 AI, 윤리 원칙

* 논문투고일: 2024년 11월 25일 논문수정완료일: 2024년 12월 14일 논문게재확정일: 2024년 12월 14일
1) ㈜이노크루 교육사업본부장, jihyekim@innocrew.co.kr, 제1저자
2) 서강대학교 교수, bcho@sogang.ac.kr, 교신저자

<사례 1>

스타트업 스캐터랩은, 20대 여대생 컨셉의 챗봇 이루다 1.0을 2020년 12월 23일 출시하여 3주만에 80만명의 이용자를 확보하면서 MZ세대의 많은 관심을 끌었다. 그러나 이용자들의 이루다 챗봇에 대한 성희롱 문제, 동성애, 장애인, 흑인 차별 등 사회적 혐오를 조장하는 내용을 그대로 학습하여 전달하는 이슈, 데이터 사용 과정에서 개인정보 유출 의혹 등 예상하지 못한 다양한 논란이 불거졌고, 결국 출시 3주만에 스캐터랩은 서비스 중단을 결정하였다. 스캐터랩은 이후 서비스 DB와 딥러닝 대화모델 폐기를 발표하는 등 어려움을 겪었고, 새로운 버전은 1년 이상 지난 후에야 출시될 수 있었다.

<사례 2>

아마존은 2014년 더 효과적인 엔지니어 채용을 위해 스코틀랜드의 수도 에든버러에 엔지니어 팀을 결성해 인공지능 시스템을 개발했다. 해당 시스템은 머신러닝을 기반으로 지원자의 이력서에서 약 5만 개의 키워드를 추출하여 자사에 적합한 인재를 선별하는 방식으로 채용의 효율성을 높이려는 시도였다. 이 시스템은 지난 10년간 회사에 제출된 이력서 패턴을 학습하여 지원자를 평가하였는데, 업계에 남성 엔지니어가 많았기 때문에 AI는 남성이 업무에 더 적합하다고 학습하였다. 그 결과 이력서에서 여성과 관련된 단어가 많이 발견되면 점수를 낮게 주는 경향이 나타났으며, 이러한 여성 차별적 편향은 결국 2017년 AI 채용 프로그램을 폐기하는 결정적 요인이 되었다.

Ⅰ. 서 론

인공지능(Artificial intelligence: AI)은 우리 시대의 가장 혁신적인 기술 중 하나로 자리잡으며 다양한 업무와 산업 분야에서 그 존재감을 확대해가고 있다. 기술 발전의 속도가 빨라지면서 기업들은 AI를 활용하여 운영 효율성을 향상시키고, 새로운 비즈니스 모델을 창출하고 있다. 2024년에 발간된 McKinsey의 Global Survey on AI에 따르면 AI를 사용하고 있는 기업의 비율이 2023년까지는 50%대에 머물렀으나 2024년에는 72%로 급속히 늘어났으며, 특히 생성형 AI의 경우 2023년 33%였던 것이 2024년에는 65%로 크게 증가하였다. 산업별로는 에너지 및 자원 분야와 전문서비스 분야에서 AI활용이 가장 급속히 증가한 것으로 나타났으며, 활용 영역별로는 마케팅 및 영업 분야가 가장 많았고 다음으로는 제품 및 서비스 개발 분야인 것으로 나타났다(McKinsey, 2024).

그러나 위의 두 사례에서 보듯이 AI 활용이 증가하면서 예상치 못한 문제들도 함께 발생하고 있는데. 예를 들면 알고리즘 편향, 개인정보 침해, 저작권 문제 등과 같은 윤리적 문제들이 동시에 부각되고 있다. AI Incident Database에 따르면 자율주행차량의 보행자

사고나 안면인식 오류로 인한 부정확한 범죄자 체포 등 AI 관련 비윤리적 사고가 해마다 증가하고 있으며, 2023년에는 총 123건으로 전년 대비 32.3%가 증가하였다고 한다(AI Incident Database, 2023). 이에 따라 AI의 윤리적 문제에 대한 일반인의 위기의식도 높아지고 있다. 경기연구원이 2022년 발간한 <인공지능의 윤리적 쟁점에 관한 탐색적 연구> 보고서에서 시민들을 대상으로 진행한 인공지능 윤리 인식조사에 따르면, 응답자의 54%가 AI 윤리 문제가 심각하다고 인식하였으며, 미래에는 더욱 심각해질 것이라고 응답한 비율은 58.2%로 나타났다(배영임·김유나, 2022).

AI가 인간을 대신하는 역할을 수행하는 정도가 커질수록 AI의 윤리적 측면에 대한 우려도 커지고 있다. 이에 따라 철학, 인지과학, 기술 공학 등 여러 분야의 전문가들은, 윤리적 기준을 준수하는 AI 개발을 위해 윤리 원칙을 제시해 왔으며(예: 김명주, 2017; 이우권, 2020; 최경석, 2020; Floridi, 2013), AI 개발을 담당하는 엔지니어, 프로그래머, 그리고 기업체의 윤리적 태도를 고양하기 위한 다양한 노력이 이루어지고 있다. 특히 AI 기술과 서비스를 제공하는 기업들에 대해서는 단순히 기술적 성능의 향상뿐만 아니라 AI를 책임감 있게 사용하며 사회적, 윤리적 책임을 다해야 한다는 사회적 요구가 커지고 있으며, 이는 기업이 AI를 통한 생산성 향상 외에 AI의 윤리성을 높이기 위한 방법에 대한 심도 있는 고민을 필요로 하고 있다. AI의 활용은 기업의 입장에서는 생산성 증대와 새로운 시장의 개척이라는 긍정적 효과를 기대할 수 있지만, 잘못될 경우 기업 이미지 손상과 법적 문제까지 초래할 수 있는 리스크가 존재한다. 기업은 AI 알고리즘이 해를 끼칠 경우 법적 책임에 대한 우려를 해결해야 하고, 이러한 문제들은 데이터 프라이버시, 알고리즘 투명성, 결정 과정의 공정성, 그리고 AI의 사용이 인간의 일자리에 미치는 영향 등을 포함한다(Zhou, Chen, Berry, Reed, Zhang, & Savage, 2020).

이러한 측면을 고려하여 기업들은 AI 기술을 책임감 있게 개발하고 배포하는 전략을 수립해야 할 필요가 있다. 최근에는 기업 차원의 노력이 '책임 있는 AI(Responsible AI)' 라는 이름으로 AI의 윤리적 측면에 대한 기업의 대응활동 방향을 마련하고 있다. 본 연구는 AI 윤리에서 중시되는 원칙을 정리하고, 이를 실천하기 위한 기업의 활동을 책임있는 AI 프로세스를 중심으로 살펴보고자 한다.

Ⅱ. AI와 AI 윤리

AI는 기계나 컴퓨터 시스템이 인간과 유사한 지능을 발휘하는 기술 분야로서, 환경을 인식하고, 학습 및 지능을 활용하여 정의된 목표를 달성하는 데 필요한 행동을 취할 수 있는 소프트웨어와 방법을 개발하고 연구하는 컴퓨터 과학의 한 분야이다. AI는 "인간의 학습, 추론, 자가 교정 능력 등을 모방할 수 있는 컴퓨터 시스템 또는 기계의 능력"으로 정

의된다(Russell & Norvig, 2020). 캘리포니아 공과대학의 발표 자료에 의하면 “AI의 발전은 원래 인간처럼 사고할 수 있는 기계를 만들고자 하는 아이디어에서 시작되었으며, 이는 기술 혁신을 촉진하고 새로운 AI 응용 프로그램의 개발로 이어졌다”고 한다(Caltech Science Exchange, 2022). AI는 주로 기계 학습(machine learning)과 딥 러닝(deep learning) 기술을 활용하여 데이터 패턴을 파악하고 예측 모델을 만들고 이를 통해 음성 인식, 이미지 처리, 자연어 처리 등의 영역으로 확장된다. 이러한 기술은 금융 서비스, 자동차 산업, 고객 서비스 등 다양한 산업에서 광범위하게 활용되고 있다.

AI 기술의 발전은 인간과 사회에 많은 기회와 혜택을 제공하지만 동시에 문제들을 야기하기도 한다. AI는 방대한 데이터에서 패턴을 학습하는 통계적 방법을 사용하는데 예를 들면 백인 남성에 비해 상대적으로 자료의 원천으로는 소수인 여성, 장애인, 유색인종의 데이터는 수집하기 어려워 이로 인한 AI 알고리즘에 편향이 생길 수 있다. 기업의 인사부서가 구직자의 이력서를 AI 시스템으로 1차 필터링할 경우, 이러한 AI 시스템의 편향 때문에 일부 소수 집단의 구직자들이 차별을 받을 수 있는 것이다. 프로퍼블리카의 연구에서 범죄자의 재범 가능성을 예측하기 위해 AI 알고리즘을 사용하였더니 AI는 흑인의 재범 가능성을 백인보다 훨씬 높게 예측했다(배영임·김유나, 2022에서 인용). 이러한 결과는 AI 기술이 차별적 판단을 할 수 있음을 의미하며 가치중립적이라고 생각되는 AI 기술이 사회적, 윤리적 문제를 발생시키고 있음을 보여준다.

AI 시스템의 기술적인 문제는 시스템의 개발 과정에서 개발자가 의도적으로 가치판단을 내린 것이 아니라 시스템 구현상의 활동이 가치중립적이지 않은 결과를 만들어 내는 경우이다. AI가 윤리적 문제를 발생시키는 또 다른 경우는 AI 개발과 관련하여 인간이 가치판단의 기준을 적용하는 경우이다. 예를 들면, 트롤리 딜레마의 경우처럼 자율주행차가 위험 상황에서 누구를 우선적으로 보호할 것인가는 개인마다, 상황마다 우선순위에 대한 판단이 다를 수 있다. 운전자를 먼저 보호할 것인가, 아니면 동승자를, 아니면 보행자를 우선시할 것인가? 동승자나 보행자가 어린이, 임산부, 고령자일 경우 우선순위가 달라지는가 등의 이슈는 자율주행차를 개발하는 개발자가 무엇을 우선시하는 것으로 프로그램을 개발하는 가에 따라 의사결정의 결과가 달라지고, 위기 상황에서 누가 생존 또는 사망할 것인지가 결정된다. 따라서 AI 개발자의 윤리적 기준이 어떻게 적용되는지가 중요한 의사결정 기준이 된다. AI가 기술적으로 고도화되고 AI의 의사결정 자율성의 범위와 정도가 확대될수록 AI의 판단으로 인한 윤리적 위험성이 증대될 것으로 예상된다.

이러한 윤리적 의사결정에 대한 사회적 합의를 만들기 위해 정부나 민간이 주체가 되어 다양한 시도가 이루어지고 있다. EU, 미국, 영국, 독일 등 주요국들은 국가 차원에서 AI 윤리 가이드라인을 마련하여 AI 기본원칙을 제시하고 이를 추진하기 위한 범국가적 차원의 AI 거버넌스를 도입하고 있다(한국지능정보사회진흥원, 2022). 우리나라도 과학기술정보통신부는 2020년 「인공지능 윤리기준」, 2021년 「신뢰할 수 있는 인공지능 실현전략」을

발표하였다. 개인정보보호위원회, 방송통신위원회, 금융위원회는 2021년 관련 자율점검표·가이드 라인을 제정하였으며, 국가인권위원회는 2022년 「인공지능 개발과 활용에 관한 인권 가이드라인」을 마련하였다(박소영, 2023). EU의 경우 AI 생성물 표기를 의무화하는 내용을 포함한 AI 개발과 사용을 규제하는 인공지능법(AI Act)을 2024년 8월 공식 발효하였다. 이는 AI 윤리가 단순히 가이드라인으로서 AI 생태계의 방향성 설정의 기준이 되는 상징적 의미를 넘어서 실제 생태계 내 여러 이해관계자의 활동을 제약하는 규제의 지침이 되었다는 것을 뜻한다. 향후 국가 차원에서의 규제가 증가된다는 것은 기업에게는 도전적인 과제가 추가된다는 의미가 될 것이다.

이에 따라 정부뿐만 아니라 민간 기업들도 AI 윤리의 제정 및 실행을 위한 활동을 확대하고 있으며 AI 기술개발, 법·제도 정비, 윤리 교육 강화, 평가 체계 구축 등 다양한 전략을 추진하고 있다. 기업은 AI 윤리 원칙을 제정하는 것만으로는 충분하지 않으며 실제 AI 개발, 운영, 서비스 제공 등 사업활동 전반에 걸쳐 AI 윤리 원칙을 어떻게 실천할 것인지 구체화하는 것이 필요하다.

AI 기술의 급속한 발전과 함께 여러 윤리적 문제들이 대두됨에 따라 이를 해결하고자 하는 다양한 접근 방식과 관점이 등장하고 있는데, ‘AI 윤리(AI ethics)’, ‘윤리적 AI(ethical AI)’, ‘책임있는 AI(responsible AI)’ 등 개념들이 혼재되어 사용되고 있다. ‘AI 윤리’는 AI 기술의 개발과 적용 과정에서 발생하는 윤리적 문제들을 다루는 것으로 AI 기술의 개발과 사용에 관련된 도덕적 원칙과 규범을 의미하는데, 기술의 혁신과 발전이 인류 사회에 미치는 영향에 대한 논의를 포함하며 AI 시스템의 설계, 구현, 배포 및 사용에 있어서 윤리적 고려 사항을 강조한다. ‘윤리적 AI’는 AI가 기본적인 윤리적 가이드라인을 준수하는 경우 이를 윤리적 AI라고 한다. ‘윤리적 AI’는 AI 기술적 관점의 설계, 개발, 사용이 윤리적 원칙과 가이드라인을 준수하여 AI가 인간의 삶을 향상시키고, 사회적 공동선에 기여할 수 있도록 하는 것을 목표로 한다(Jobin, Lenca, & Vayena, 2019). 윤리적 AI에서는 AI 시스템 자체가 윤리적인 행동과 결정을 할 수 있도록 설계되어야 하는 것을 강조하고 있으며, 이는 AI 시스템이 인간의 윤리적 기준을 준수하며 작동하도록 설계된다는 개념이다. 즉, AI의 의사결정 과정에서 윤리적 판단을 내릴 수 있도록 하는 기술적 구현에 논의의 초점이 있다. ‘책임있는 AI’는 윤리적 원칙들을 통합하면서 AI가 사회적, 법적, 도덕적 책임을 갖고 개발 및 배포되어야 한다는 활동에 초점을 맞추어 설명한다. 즉 AI와 관련한 행위의 주체가 AI 윤리 원칙에 부합하는 정책, 절차 및 기준을 수립하기 위한 구조나 역할을 설정하는 노력을 할 때 책임있는 AI가 구현된다고 할 수 있다. 정리하면 AI 윤리는 “AI가 가져야 할 윤리적 기준이 무엇인지”를, 윤리적 AI는 “설계되고 배포된 AI가 윤리적 기준에 부합하는지”를, 그리고 책임있는 AI는 “AI 윤리를 적용하기 위해서는 AI 관련 주체가 어떤 활동을 해야 하는지”를 서술할 때 사용하는 개념이다.

Ⅲ. AI 윤리 원칙

먼저 AI 윤리로는 어떤 원칙이 제시되는지 몇 가지 예를 살펴보자. 하버드대학교 Berkman Klein 센터는 2016년부터 2019년 사이 발표된 36개의 주요 AI 윤리 원칙 문서를 종합 분석하여 8개의 핵심 테마를 도출하고, 각 테마별로 자주 언급되는 원칙들을 정리하였다(Fjeld, Achten, Hilligoss, Nagy, & Srikumar, 2020)(<표 1>). 개인정보 보호는 AI 시스템이 사용자들의 개인정보를 존중하고 보호해야 함을 의미하며, 책임성은 AI 시스템과 그 개발자들이 그들의 행동과 결정에 대해 책임을 질 수 있는 메커니즘을 구축하는 것을 뜻한다. 안전 및 보안은 AI 시스템이 사용자에게 안전하고 높은 보안성을 제공해야 함을 강조하고, 투명성 및 설명 가능성은 AI 시스템이 이해할 수 있는 방식으로 의사 결정 과정을 투명하게 설명할 수 있게 하는 것을 의미한다. 공정성과 비차별은 AI 시스템이 편향과 차별을 방지하며 모든 사용자에게 공정하게 작동하도록 하는 것을 목표로 한다. 인간의 기술 통제는 AI 시스템에 대한 인간의 감독과 통제를 유지하는 것을 의미하고, 전문적 책임은 개발자들이 높은 전문성과 윤리적 표준을 준수하도록 권장하는 것이다. 마지막으로, 인간적 가치의 증진은 AI가 인간의 복지와 윤리적 가치를 지원하도록 보장하는 것을 의미한다. 이러한 테마들은 AI 기술의 발전과 윤리적 고려 사항을 균형 있게 다루어, 사회에 긍정적인 영향을 미치면서 부정적인 영향을 최소화하는 AI 시스템을 개발하는 것이 중요함을 강조하고 있다.

EU 집행위원회의 과학 및 신기술 윤리 그룹(Group on Ethics in Science and New Technologies)은 2018년 과학과 신기술에 대한 윤리그룹의 윤리적 원칙을 9가지로 발표했다(European Commission, 2018). 이는 새로운 기술을 선보이는 과정에서 고려해야 하는 다양한 원칙들을 담고 있다. 이 원칙들은 인간의 존엄성, 정의와 공정성, 보안과 안전, 자율성, 민주주의, 데이터 보호와 프라이버시, 책임, 법치와 책무성, 지속가능성 등의 광범위하고 다양한 가치를 포함한다. 예를 들어, 헬스케어 AI는 환자의 개인 정보를 보호하고, 채용 AI는 공정하게 작동해야 하며, 자율주행자동차 AI는 안전하게 운영되어야 한다. 이러한 원칙들은 AI 시스템이 인간의 기본 권리와 자유를 침해하지 않도록 하고, 모든 사용자에게 공정하게 작동하며, 데이터를 보호하고, 법적 책임을 질 수 있도록 설계되어야 함을 강조한다. 이는 AI 기술이 지속 가능한 발전에 기여하고, 사회적 책임을 다하는 방향으로 나아가도록 돕는 중요한 가이드라인을 제공한다.

영국의 데이터 사이언스 및 인공지능을 위한 국립연구소인 Alan Turing Institute는 AI 기술 설계와 사용의 원칙으로 FAST를 제시하였다(Leslie, 2019). FAST는 공정성(Fairness), 책임성(Accountability), 지속가능성(Sustainability), 그리고 투명성(Transparency)을 의미한다. 공정성은 누구도 차별이나 편견으로 인해 피해받지 않도록 AI를 설계하고 실행해야 함을 의미하고, 책임성은 AI 설계부터 실행까지 모든 과정을 인간이 설명할 수 있어야 하

는 원칙이다. 지속가능성은 AI의 혁신이 안전하고 그 결과가 윤리적이어야 함을 강조하며, 투명성은 AI의 모든 절차가 정당화될 수 있고 알고리즘의 결과가 모두에게 이해될 수 있어야 함을 뜻한다. 이 4가지 원칙 중 투명성과 책임성은 AI 전 과정에 적용되어야 하는 절차 관련 메커니즘으로 특히 중요하며, 공정성과 지속가능성은 시스템의 설계, 실행 및 결과물이 가져야 할 특성으로 설명된다.

〈표 1〉 하버드대학교 Berkman Klein 센터: AI 윤리 8대 테마

테마	설명	하위 원칙
개인정보 보호 (Privacy)	AI 시스템이 사용자의 개인정보를 존중하고 보호하도록 함	Consent/ Ability to Restrict Processing/ Right to Erasure/ Recommends Data Protection Laws/ Control over the Use of Data/ Right to Rectification/ Privacy by Design/ Privacy (Other/General)
책임성 (Accountability)	AI 시스템과 개발자가 자신의 행동과 결정에 책임을 지는 메커니즘을 창출	Verifiability and Replicability/ Impact Assessments/ Environmental Responsibility/ Creation of a Monitoring Body/ Remedy for Automated Decision/ Evaluation and Auditing Requirement/ Ability to Appeal/ Liability and Legal Responsibility/ Recommends Adoption of New Regulations/ Accountability Per Se
안전 및 보안 (Safety and Security)	AI 시스템의 안전성과 사용 안전성의 필요성 강조	Safety/ Security by Design/ Security/ Predictability
투명성 및 설명 가능성 (Transparency and Explainability)	AI 시스템을 이해할 수 있고 의사결정 프로세스가 투명하도록 함	Transparency/ Open Source Data and Algorithms/ Right to Information/ Notification when Interacting with AI/ Explainability/ Open Government Procurement/ Notification When AI Makes a Decision about an Individual/ Regular Reporting
공정성과 비차별 (Fairness and Non-Discrimination)	편견과 차별을 피하기 위한 AI 시스템의 공정성을 촉진	Non-discrimination and the Prevention of Bias/ Fairness/ Inclusiveness in Impact/ Representative and High Quality Data/ Equality/ Inclusiveness in Design
인간의 기술통제 (Human Control of Technology)	AI 시스템에 대한 인간의 감독과 통제를 유지	Human Review of Automated Decision/ Ability to Opt out of Automated Decisions/ Human Control of Technology
전문적 책임 (Professional Responsibility)	개발자가 높은 전문 표준과 윤리적 관행을 준수하도록 장려함	Accuracy/ Consideration of Long Term Effects/ Responsible Design/ Multi-stakeholder Collaboration/ Scientific Integrity
인간적 가치 증진 (Promotion of Human Values)	AI가 인간의 행복과 윤리적 가치를 지원하도록 보장	Human Values and Human Flourishing/ Leveraged to Benefit Society/ Access to Technology

한편, 국내 연구로 박소영(2023)은 주요 국가와 국내외 기업들이 중시하는 AI 윤리 4가지를 FATE(Fairness, Accountability, Transparency, Ethics)로 정리하여 제시하였으며, 그 내용은 아래 <표 2>와 같다.

〈표 2〉 AI 4대 공통윤리: FATE

원칙	설 명
공정성 (Fairness)	특정 집단에게 기회나 자원을 불공정하게 할당하거나, 특정 집단에 대한 고정관념을 갖게 하여 사회적 약자를 계속 취약한 위치에 처하게 하는 불평등 재생산 금지
책임성 (Accountability)	인공지능을 설계·개발·배포·운용하는 자는 인공지능 시스템이 적절하게 기능하도록 설계하고, 지속적으로 관리, 감독하는 체계를 구축하여 자신의 인공지능으로 하여금 부작용이 발생하는 것을 방지해야 함
투명성 (Transparency)	인공지능 판단이 중요한 영향을 미치는 영역에서는 인공지능이 사용된다는 사실뿐만 아니라 인공지능이 사용하는 데이터, 변수, 알고리즘 작동 방식에 대한 기본 정보를 제공
윤리의식 (Ethics)	인권을 존중하고 민주적 가치를 보장하며, 인류의 공동이익을 위하는 방향으로 인공지능을 활용

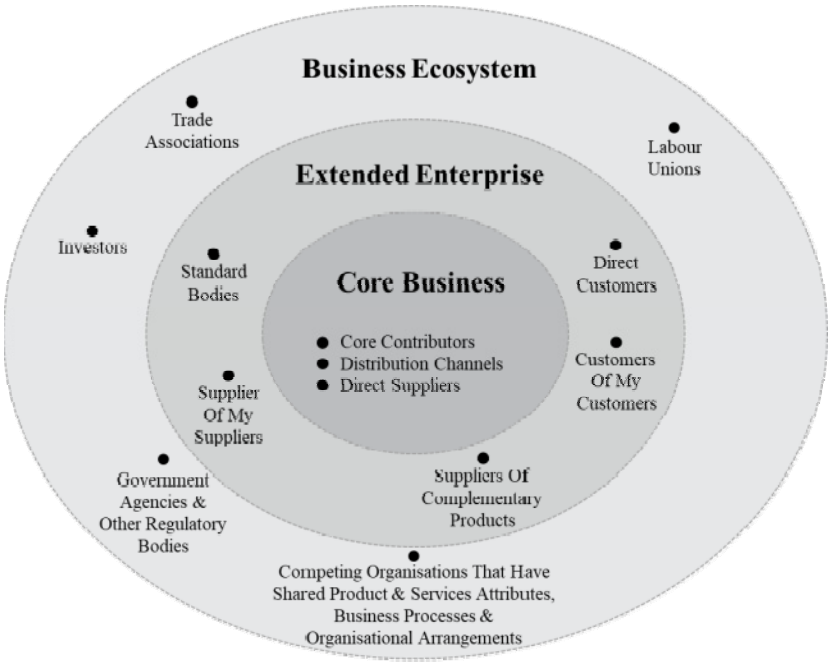
Ⅳ. 책임있는 AI와 AI 생태계

앞서 설명한 모든 윤리적 원칙들을 통합하고 AI가 사회적, 법적, 도덕적 책임을 갖고 개발 및 배포되도록 하기 위해 AI와 관련된 주체들이 실천해야 하는 활동은 책임있는 AI라는 명칭으로 언급된다. 책임있는 AI란 AI 시스템을 개발, 운영, 배포하는 과정이 윤리적이고 투명하며 사용자 개인 정보와 사회적 가치를 존중하는 방식으로 이뤄지는 것을 의미한다(삼일PwC경영연구원, 2023). 이는 AI 기술의 투명성, 신뢰성, 안전성, 프라이버시 보호 및 공정성 등 AI 윤리로 강조되는 원칙들을 보장하기 위한 체계적인 접근을 포함한다.

책임있는 AI가 구현되기 위해서는 AI와 직간접적으로 관련된 생태계 내의 모든 이해관계자들이 AI 윤리의 원칙을 이해하고 이를 준수하려는 의지가 필요하다. AI 생태계는 AI를 기반으로 상호작용하는 다양한 이해관계자 집단이라고 정의할 수 있을 것이다(심진보, 2021). 생태계 내에는 다양한 이해관계자가 존재하는데 이를 생태계의 핵심 부분과 주변 부분으로 나눌 수 있다. Moore(1993)의 분류에 따르면 핵심사업(core business)과 확장 사업(expanded enterprise)에 해당하는 공급자, 유통업자가 핵심 부분이며, 에코시스템(business ecosystem)에 해당하는 소비자, 규제기관, 정부, 투자자 등은 주변 부분이다. AI 생태계에서는 핵심 부문은 AI 모델을 만드는 개발사업자, AI 모델을 이용하여 서비스를 제공하는 이용사업자, AI 서비스 사용자, 그리고 데이터를 저장하는 플랫폼을 제공하는 배포자이며,

법이나 제도를 통해 직접 또는 간접적으로 영향을 주는 정부나 시민단체 등은 생태계의 주변 부분에 해당한다고 할 수 있다(유순덕, 2020). 최근 급속도로 확산 사용되고 있는 생성형 AI산업에서는 AI연산모델에 사용되는 반도체 제조사도 중요한 이해관계자로 포함된다(유재홍·안성원·안미소·노재원, 2023).

〈그림 1〉 AI 에코시스템



출처: 유순덕(2020)

EU의 AI법은 이해관계자를 제공자, 배포자, 공인대리인, 수입업자, 유통업자, 제품 제조업자, 공급자 등 7개 유형으로 세분화하고, 각 유형별로 필요한 의무와 책임을 규정하고 있다(<표 3> 참조). 예를 들어 제공자(provider)와 배포자(deployer)에 대해서는 투명성 의무를 별도로 규정하고 있는데, 제공자는 산출물이 기계적으로 판독이 가능하도록 하여야 하고 인공적으로 생성되었음을 검출할 수 있어야 한다. 또한, AI시스템 배포자는 시스템이 작동 중임을 알리고 개인정보 처리를 포함한 투명성 의무조항을 준수해야 한다고 명시되어 있다.

〈표 3〉 EU AI법의 AI 이해관계자 분류

AI 이해관계자	정 의
제공자 (provider)	AI시스템 또는 범용 AI모델을 개발하거나 자신의 이름 또는 상표 하에 유료 또는 무료로 AI시스템 또는 범용 AI모델을 개발하여 출시하거나 AI시스템을 서비스 개시하는 자연인이나 법인, 공공기관, 관청 또는 기타 기구
배포자 (deployer)	AI시스템이 자체 권한에 따라 AI 시스템을 사용하는 자연인이나 법인, 공공기관, 관청 또는 기타 기구 개인적, 비전문가적 활동 과정에서 사용되는 경우는 제외함
공인대리인 (authorized representative)	AI시스템 또는 범용 AI모델 제공자로부터 의무와 절차를 대신 이행하고 준수할 권한을 서면으로 위임받고 이를 수락한 자로서 유럽연합 내에 소재하거나 설립된 자연인이나 법인
수입업자 (importer)	유럽연합에 소재하거나 설립된 자연인이나 법인으로서 제3국에서 설립된 자연인이나 법인의 이름 또는 상표를 지닌 AI시스템을 출시하는 자
유통업자 (distributer)	공급망에 속하는 자연인이나 법인 중 제공자 또는 수입업자가 아닌 자로서 유럽연합 시장에 AI시스템을 제공하는 자
제품 제조업자 (product manufacturer)	AI를 안전요소로 포함시키는 경우를 포함하여 최종 제품 생산에 책임있는 자
공급자 (supplier)	고위험성 AI시스템 제공자에게 AI시스템, 도구, 서비스 등을 제공하는 자

Ⅴ. 책임있는 AI를 위한 단계별 기업의 활동

기업은 AI 생태계의 각 부분에서 중요한 역할을 맡고 있다. 생성형 AI의 예로 살펴보면, AI 반도체 생산 기업에는 삼성전자, SK하이닉스, NVIDIA, Intel, Google 등이 있으며, AI 데이터 수집, 저장 및 학습을 위한 클라우드 기업으로는 네이버, KT 클라우드, MS Azure, Amazon AWS, Google GCP 등이 있다. AI 플랫폼으로는 Naver HyperClova, Open AI GPT-4, DeepMind Gopher 등이, 그리고 AI 서비스 기업으로는 건강 AI 챗봇 서비스를 하는 굿닥, 자기소개서 생성을 지원하는 잡브레인, 이미지 생성 서비스를 제공하는 미드저니 등이 있다(유재홍 등, 2023).

이처럼 AI 생태계에서 기업은 중심과 주변에서 직·간접적으로 AI의 윤리적 문제와 관련이 되어있어 보다 책임있는 AI 활동을 적극적으로 시도하고 있다. 특히 AI 라이프사이클의 각 단계에서 발생할 수 있는 윤리적 문제들을 사전에 이해하고 해결하는 것이 책임있는 AI 활동에 기여할 수 있다. 이를 위한 세부 활동을 AI 라이프사이클 단계에 따라 살펴보면 다음과 같다(Huang, Zhang, Mao, & Yao, 2023).

〈표 4〉 AI 라이프사이클의 각 단계에 따른 윤리적 고려사항(Huang et al., 2023)

단계	해당 단계의 윤리적 고려 사항
사업 분석 (Business Analysis)	• 투명성, 공정성: 설계된 AI 제품의 아키텍처에 어떤 변수, 기능, 프로세스가 포함되어 있으며, 이것이 합리적이고 도덕적으로 문제가 되지 않는가? • 책임성, 민주주의 & 시민권, 지속가능성
데이터 엔지니어링 (Data Engineering)	• 프라이버시: 데이터 세트에 포함된 민감하고 개인적인 정보의 데이터 보안을 어떻게 보장할 것인가? • 투명성: 데이터 수집 절차를 소비자에게 어떻게 투명하게 할 것인가? • 공정성: 데이터가 대표적이고, 관련성 있으며, 정확하고, 일반화 가능한가? • 민주주의와 시민권: 최종 사용자가 자신의 데이터 사용을 어떻게 통제할 수 있게 할 것인가?
기계 학습 모델링 (ML Modeling)	• 투명성: 모델의 결정 또는 추론 과정을 어떻게 이해할 수 있게 할 것인가? • 안전성: 모델의 정확성, 신뢰성, 보안성, 견고성은 어떠한가? • 공정성: 모델 출력이 다른 그룹의 사람들에게 불균등한 결과를 보여주지는 않는가?
모델 배포 (Model Deployment)	• 프라이버시: 배포된 모델을 통해 개인 정보가 다시 식별되지 않도록 보장하는 방법은? • 안전성: 배포된 모델의 안전성을 어떻게 확보할 것인가, 예를 들어 악의적인 수정이나 공격으로부터?
운영 및 모니터링 (Operation and Monitoring)	• 프라이버시: 운영 및 모니터링 과정에서 개인정보 보호를 어떻게 보장할 것인가? • 공정성: AI 제품이 영향을 미치는 사람들에게 차별적 또는 불평등한 영향을 미치지 않는가? • 민주주의 & 시민권: 사용자의 시민권을 침해하지 않는가?

Huang 등의 모형(2023)과 유사하게 PwC는 기업 내부에서 책임있는 AI에 대한 공감대와 가이드라인을 지속적이고 일관성 있게 가져가기 위해서는 AI 모델 개발 및 전 라이프사이클에 걸친 전사적인 거버넌스를 갖추어야 함을 강조하면서, AI 개발 프로세스 별로 발생하는 리스크와 제어 방법에 초점을 맞춘 End-to-end 거버넌스 모델을 6단계로 제시하였다(<그림 2>). 이 모델에 따르면 첫 번째 단계는 기업의 장기적 목표와 비전에 맞춘 AI 전략 수립, 산업 표준 및 규정 준수, 내부 정책 및 규범 설정을 포함하는 전략 수립이다. 두 번째 단계인 계획은 AI 프로그램 감독, 구체적인 계획 수행 방안 마련, 포트폴리오 관리로 구성된다. 세 번째 단계는 생태계 구축으로, 기술 로드맵 수립, 벤더 평가 및 소싱, 변화 관리를 포함한다. 네 번째 단계는 개발로서 비즈니스 요구와 데이터 이해, 솔루션 설계 및 데이터 추출, 데이터 전처리 및 AI 모델 구축을 포함한다. 다섯 번째 단계는 배포를 통한 모델 성능 평가 및 체크인, 운영 환경에서 성능 모니터링, 전환 실행 및 시스템 통합을 포함한다. 마지막 단계는 모니터링 & 보고로 지속적인 모니터링 및 성능 유지, 정기적인 컴플라이언스 보고를 포함한다.

기존 제품 또는 서비스의 프로세스보다 더욱 복잡해진 AI 서비스의 라이프사이클을 효과적으로 관리하려면 더욱 상세하고 체계적인 절차가 필요한데 PwC의 6단계 모델은 각 업무 단계 플로우에서 명확한 역할과 책임, 추적 가능성과 지속적인 평가의 필요성을 강조한다. 이를 통해서 기업들은 책임있는 AI 모델에 대한 전반의 과정을 빠짐없이 검토하여 AI 시스템의 신뢰성과 투명성을 유지할 수 있게 된다. 이 모델에 따르면 AI 서비스 기업은 전략적 이해를 바탕으로 AI의 적용 범위를 설정하고, AI 모델의 배포 타당성을 단계별로 검토해야 한다. 또한, 적용 범위가 확정되면, 배포 후 사용자의 피드백을 바탕으로 모델을 지속적으로 개선하거나 필요에 따라 중단해야 한다. 단순히 기능을 배포하는 것을 넘어, 모델의 상태를 지속적으로 모니터링하고 관리하여 AI 모델의 라이프사이클 전반을 관리하여 품질을 높여야 한다는 것이 책임있는 AI를 위한 프로세스의 핵심 목적이라고 할 수 있다.

〈그림 2〉 PwC의 책임있는 AI 거버넌스 모델

출처: 삼일PWC경영연구원(2023)

이하에서는 AI 서비스 라이프사이클에 따라 기업이 윤리적 도전과 이슈를 이겨내기 위한 책임있는 AI 활동을 어떻게 하고 있는지 구체적인 사례를 살펴보기로 하자. 아래에서는 각 기업 사례를 IEEE 모형과 PwC 모델의 세부 사항을 사업 단계에 맞추어 △전략 및 계획수립, △생태계, △개발 및 배포, △모니터링 및 보완의 4단계로 구분하여 단계별 주요 활동 내용 그리고 실제 기업의 사례를 살펴본다. 사례로 소개되는 기업은 신문기사나 문헌

에 소개된 내용, 그리고 AI윤리 국제 심포지엄(2024 World AI Cannes Festival: WAICF)에서 발표된 자료를 참조하여 선정하였다.

1. 전략 및 계획 수립: 윤리 원칙 제정 및 윤리 거버넌스 구조 구축

첫번째 단계인 전략 및 계획 수립 단계는 기업의 AI 전략을 수립하고, 이를 실행하기 위한 구체적인 계획을 마련하는 단계이다. 이는 IEEE의 사업계획단계에 해당하고, PwC 모델에서는 1. 전략, 2. 계획에 해당한다. 이 단계에서는 시스템에 대한 이해를 바탕으로 AI 윤리 기준을 설정하여 이를 조직 내외부에 공개하고, 이후 개발, 배포, 그리고 모니터링 과정에 대한 책임을 가지는 거버넌스 구조를 수립하는 것이 필요하다.

Microsoft, IBM, Google 등과 같은 선도 기업들은 AI 윤리 가이드라인을 자체적으로 개발하여 이를 홈페이지 등에 공개하고 있다. 대부분의 기업들은 앞서 말한 AI 윤리의 원칙들을 전부 또는 일부 포함하고 있다. 아래의 표는 국내외 기업들의 AI 윤리 원칙을 비교한 것이다. 윤리 원칙 중 무엇이 포함되고 무엇이 보다 중시되어야 하는 지는 업종의 특성에 따라 달라지게 되는데 예를 들면 어느 회사는 투명성을, 어떤 회사는 안전을 보다 강조한다.

〈표 5〉 주요기업 및 기관의 인공지능 윤리 원칙

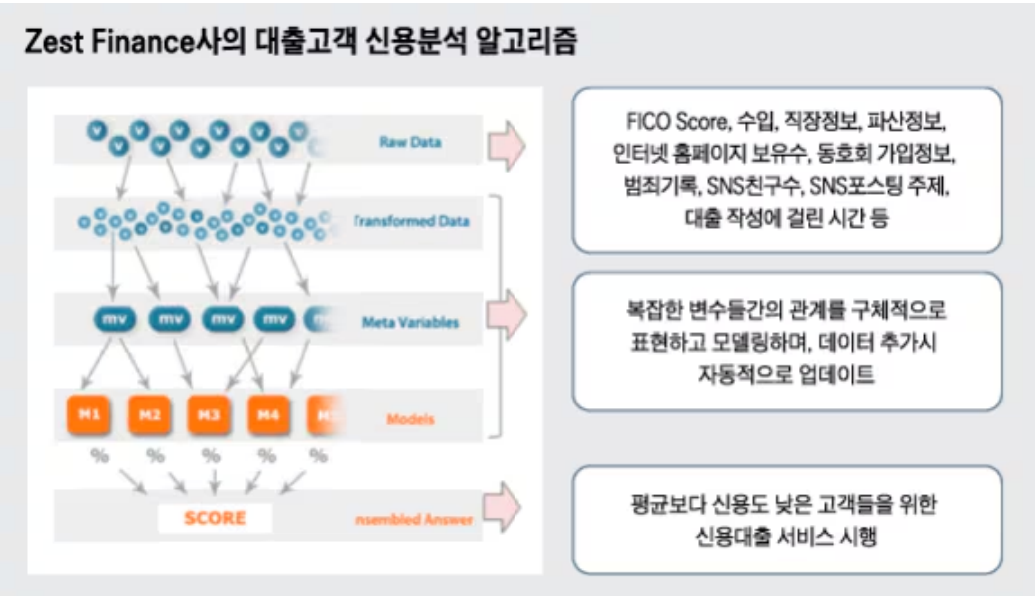
	Humanity Human-centered	Responsibility Accountability	Privacy Security	Safety Reliability	Transparency Explanability
Google	O	O		O	
Microsoft		O	O	O	O
IBM	O		O		O
Adobe		O			O
OpenAI	O			O	
LG AI Research	O	O		O	O
Kakao	O		O		O
Naver	O		O	O	O
Samsung	O		O		O
SK Telecom	O		O	O	O
Upstage	O	O	O	O	O

출처: 박찬준·이원성·김윤기·김지후·이활석(2023)

ZestFinance는 신용분석 알고리즘을 개발하여 핀테크 회사에 신용 평가 정보를 제공하는 소프트웨어 개발회사이다. Google의 CIO였던 Douglas Merrill이 창업한 이 회사는 설명

가능한 머신러닝 모델을 통해 1만여 개의 변수 데이터를 사용해 개인의 신용도를 분석하며, 신용 결정 과정에서의 투명성과 공정성을 강조한다. 일반적으로 금융기관은 대출 결정을 위해 10~20여 개의 변수만을 사용하고 정확한 신용평가 기준을 공개하지 않지만, Zest-Finance는 최종 신용점수를 계산하는 머신러닝 모델의 결정 과정을 상세히 문서화하고, 각 변수가 모델의 예측에 미치는 영향을 공개한다. 이를 통해 AI 시스템의 투명성을 높이고 사용자가 AI의 결정을 신뢰할 수 있도록 돕는다.

〈그림 3〉 ZestFinance의 대출고객 신용분석 알고리즘



이미지 출처: https://biz.chosun.com/site/data/html_dir/2017/03/21/2017032101045.html

전략 및 계획수립 단계에서 기업이 책임있는 AI 윤리를 실천하기 위해서는 AI에 의한 결정이 사회적 기준과 법적 요구사항을 준수하도록 책임 수용성을 중심에 둔 의사결정 체계를 강화해야 한다. 이를 위해 기업 내에 윤리 위원회를 구성하거나 윤리 감사 시스템을 도입하는 등의 활동이 이루어지고 있다. 예를 들어, Google은 AI 윤리를 심사하는 고급 기술 외부자문 위원회를 설립하여 윤리적 의사결정을 강화했다. 이 위원회는 AI 프로젝트의 윤리적 측면을 평가하고, 그 결과에 대한 책임을 지도록 설계되었다. 이는 Google이 AI 기술을 개발하고 배포할 때 사회적 기준과 법적 요구사항을 준수하도록 보장한다. 위원회의 존재는 내부 결정 과정에 투명성을 부여하며, 기업의 책임 있는 리더십을 강화하는 데 기여한다. Facebook 역시 AI 의사결정 과정에서의 책임 수용성을 높이기 위해 AI 연구윤리팀을 설립하였다. 이 팀은 AI 기술을 통해 이루어지는 모든 의사결정이 회사의 윤리 기준

을 준수하는지 감시한다. 이를 통해 의사결정 과정에 투명성을 제공하고, 사용자 데이터를 처리하는 방식에 대한 신뢰를 증진시키려 노력한다. 또한, 이 팀은 정기적인 리뷰와 감사를 통해 AI 알고리즘의 공정성을 평가하고, 필요한 경우 조정하는 과정을 관리한다. 네이버는 AI 윤리자문 프로세스(CHEC)를 구축하여 서비스 출시전에 네이버 AI 윤리 준칙을 기반으로 외부관점에서 발생가능한 이슈를 도출하여 이를 검토하여 개선방안을 제시하는 시스템을 구축하였다(장진철·노재원·유재홍·조지연, 2024).

2. 생태계 구축: 이해관계자 간의 선순환 비즈니스 모델

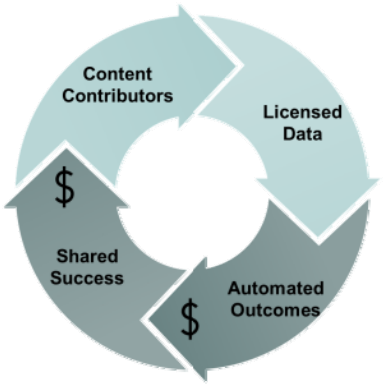
AI 서비스 라이프사이클의 2단계 생태계 구축 단계는 PwC 모델에서 3단계 생태계에 해당한다. 이 단계는 AI 프로젝트 수행에 필요한 외부 자원과 이해관계자 등과 협력 관계를 구축하는 단계로서 외부 자원과 이해관계자 간의 선순환 비즈니스 모델을 만들어 나가는 것이 필요하다. 리 카이푸(2019)는 지속 가능한 혁신을 촉진하기 위해서는 효과적인 AI 생태계 구축이 필요한데 이는 기술적 역량과 인적 자원의 결합을 통해 가능하다고 주장했다. 효과적인 AI 생태계 구축을 통한 비즈니스 성공의 사례로 셔터스톡을 살펴보자.

셔터스톡의 데이터 사이언스 및 AI 수석 디렉터인 Alessandra Sala는 MIT Sloan과의 인터뷰에서, 셔터스톡 AI가 AI 기술을 사용하여 생성된 이미지에 대한 저작권료를 원본 이미지 작가들에게 제공함으로써 새로운 비즈니스 모델을 도입했다고 밝혔다(MIT Sloan Management Review, 2023). 셔터스톡은 AI 생성 콘텐츠에 사용된 원본 이미지의 작가들에게 적절한 저작권료를 지급하여 저작권을 존중하고 법적 분쟁의 위험을 줄이는 한편, 시장에서 신뢰를 구축하고 투명한 거래 조건을 제공함으로써 사용자와 창작자 모두의 권리가 보호받고 있음을 보장한다. 이러한 접근 방식은 지속 가능한 비즈니스 모델을 제시하며, 기술 혁신과 창작자 권리 보호 간의 균형을 잘 이루어 AI 기술의 책임 있는 사용을 촉진하고, 장기적으로 기업의 경쟁력을 강화하는 방법을 제시한다.

게티이미지(Getty Images)가 “생성형 AI가 만든 컨텐츠는 저작권을 침해할 소지”가 있기 때문에 AI를 사용하여 만든 콘텐츠를 허용하지 않겠다는 입장과 셔터스톡의 사업전략은 정반대인 것이다. 셔터스톡의 사업 모델은 생성형 AI 콘텐츠의 윤리적 사용과 기존 예술가들의 지속 가능한 창작 활동 지원에 중점을 두고 있다. 셔터스톡은 생성형 AI를 통해 생성된 이미지, 비디오, 3D 오브젝트 등의 콘텐츠에 대해, 해당 콘텐츠에 사용된 원 콘텐츠 제작 예술가들에게 로열티를 지급하는 새로운 비즈니스 모델을 도입했다. 이는 콘텐츠가 AI에 의해 생성되어 판매될 때마다 원작자에게 수익이 돌아가는 순환 경제의 개념을 실현한 것이다. 이로써 셔터스톡은 예술가들이 자신의 작품이 AI 기술에 의해 재창조되어도 정당한 대가를 받을 수 있는 환경을 조성한다. 이러한 셔터스톡의 비즈니스 모델은 원소스의 데이터 수집 퀄리티에 따라 모델의 가치가 높아지는 AI 서비스의 긍정적인 영향을

미치고, 더 많은 원소스를 제공한 작가에게 더 많은 로열티가 제공되는 선순환 모델을 형성하는 것을 목표로 한다. 셔터스톡 AI 서비스의 향후 방향은 기술적 혁신과 윤리적 기준의 균형을 지속적으로 모색하는 데 중점을 두고 있다. 생성형 AI 기술의 발전에 따라 셔터스톡은 예술가들의 창작물이 새로운 형태로 재해석되는 과정에서 그 가치가 존중받을 수 있도록 지속적인 노력을 기울일 예정이다. 또한, AI로 생성된 콘텐츠의 진위를 판별하고, 이를 명확히 표시하는 방법에 대한 연구와 개발을 강화할 계획이라고 한다. 셔터스톡은 이를 통해 AI 기술이 가져올 변화 속에서도 사용자와 예술가 모두에게 투명성과 신뢰를 제공하고자 한다. 셔터스톡의 사례는 글로벌 기업이 AI 서비스를 도입하며 겪게 되는 도전과 기회를 잘 보여주고 있으며, AI 기술의 발전은 무한한 가능성을 열어주지만, 이와 동시에 여러 이해관계자간의 윤리적, 법적 문제에 대한 신중한 고민과 접근이 필요함을 강조한다.

〈그림 4〉 셔터스톡AI - 선순환 비즈니스 모델



3. 개발 및 배포: 데이터 보호와 투명성 강화

개발 및 배포 단계는 IEEE 모형의 데이터 엔지니어링, 기계학습 모델링, 모델 배포 단계를 포함하며, 이는 PwC 모형에서는 4. 개발과 5. 배포에 해당한다. 이 단계는 AI 솔루션을 실제로 개발하고 이를 운영 환경에 배포하는 과정으로, 운영 환경에서 발생하는 문제에 대한 선행 관리가 필요하다. Russell & Norvig(2020)에 따르면 AI 솔루션의 성공적인 개발 및 배포는 철저한 데이터 관리와 운영 환경에 맞춘 최적화 과정을 요구한다고 주장한다. AI 시스템이 신뢰할 수 있고 일관된 성능을 발휘하려면 데이터의 품질과 일관성을 보장해야 하며, 운영 환경에서 발생할 수 있는 다양한 변수와 요구사항을 고려하여 솔루션을 최적화해야 한다는 것을 의미한다. 이에 따라 기업은 AI 기술을 사용하면서 수반되는 다양한 위험, 특히 개인의 프라이버시 침해, 데이터 보안 위반, 그리고 AI 결정으로 인한 부

정적인 사회적 영향을 최소화할 방안을 고려해야 한다.

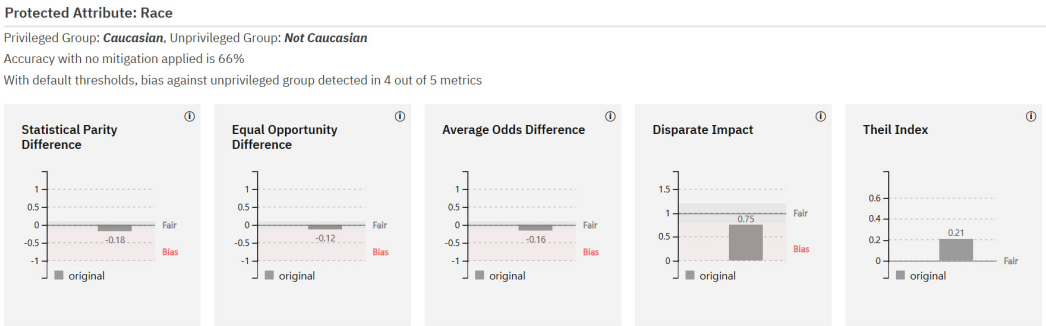
Amazon은 AI 기술을 사용하여 고객 데이터 보호를 강화하고자 AWS에서 다양한 보안 프로토콜과 서비스를 제공한다. 이러한 조치는 고객 데이터를 보호하고 데이터 보안 위반을 방지하기 위한 것으로 데이터 암호화, 정기적 보안 감사 및 위험 평가를 통해 데이터 프라이버시를 보장하며, 클라우드 서비스를 통한 보안 인프라를 강화하고 있다. 이를 통해 고객 데이터의 무결성을 유지하고, 기업이 데이터 관련 규제를 준수할 수 있도록 한다.

Meta는 Stable Signature 기술을 통해 콘텐츠의 출처와 수정 이력을 투명하게 관리하는 데 중점을 두고 있다. 이 기술은 디지털 아트, 미디어 콘텐츠, 그리고 다양한 형태의 창작물에 대한 메타데이터를 생성하여 콘텐츠의 진위와 출처를 추적할 수 있도록 설계되었다. 이는 사용자들이 콘텐츠를 신뢰할 수 있는 기반을 마련하고, AI를 활용한 창작물을 더욱 발전시키는 동시에 위험성을 줄이는 것에 기여한다.

IBM은 대형 IT 기업으로서 AI 모델의 공정성을 보장하고 편향성을 완화하기 위한 오픈소스 툴킷인 AI Fairness 360(AIF360)을 무상으로 제공하고 있다. 이 툴킷은 알고리즘의 편향을 검사하고 이해하기 쉬운 형태로 결과를 제공하여 개발자와 이해관계자들이 AI 시스템을 보다 잘 이해하고 필요한 조정을 할 수 있도록 지원한다. AI Fairness 360은 공정성을 측정하고 편향성을 완화하기 위해 70개 이상의 공정성 지표와 11개의 편향 완화 알고리즘을 제공한다. 지표는 일반적인 성능 지표 외에 그룹 공정성, 개인 공정성 지표가 포함되어 있다. 이 툴킷은 편향 정도를 체크하는 데 사용될 뿐 아니라 편향을 완화하기 위한 알고리즘 제공 기능을 중시하여 AI 모델을 개발하기 위한 데이터 전처리 단계, 모델링 단계, 모델링 사후 단계별로 선택하여 사용 가능하다(이재호·조남용, 2023).

아래 그림은 IBM 홈페이지에서 테스트할 수 있는 데모용 데이터 분석 결과로, 백인에 비해 비백인 범죄 기소자가 재범할 확률을 계산하는 AI 모델이 편향된 결과를 보이는 정도를 나타낸다. 이 분석에서는 공정성 지표 5가지 중 4가지에서 편향이 드러나고 있음을 확인할 수 있다.

〈그림 5〉 IBM - AI Fairness 360 작동 원리



IBM과 유사하게 Google은 2019년 Fairness Indicator를 통해 AI 모델의 성능과 함께 공정성 지표를 제공한다. Microsoft는 2020년 Fairlearn이라는 도구를 통해 공정성을 검증하며 또한 공정성을 높일 수 있는 알고리즘을 같이 제공하고 있다. 이러한 기술들의 주요 목표는 서비스 과정에서의 공정성과 투명성을 증가시키고 사용자들이 해당 콘텐츠의 신뢰성을 판단할 수 있도록 돕는 것이다. 이러한 사례들은 AI 기술 혁신과 윤리적 책임 사이의 균형을 추구하려는 기업들의 노력을 잘 보여주고 있다.

4. 모니터링 및 보완: 지속적 모니터링 및 즉각적 대응 체계

모니터링 및 보완 단계는 AI 라이프사이클의 마지막 단계로서 IEEE 모형에서는 운영 및 모니터링에 해당하고, PwC 모델에서는 6. 모니터링 & 보고에 해당한다. 마지막 모니터링 및 보완 단계는 AI 시스템의 운영 상태를 지속적으로 모니터링하고 컴플라이언스 보고를 수행하여 즉각적 대응 체계를 만들어가는 단계이다. 지속적인 모니터링은 AI 시스템의 지속적인 개선을 위해 필요하다(Sajid, 2024). 이를 통해 기업은 문제 발생 상황에도 더욱 빠르게 대응할 수 있으며, 문제발생으로 인한 소비자의 신뢰를 회복하는 데 중요한 역할을 한다. 스캐터랩의 예를 살펴보자.

스캐터랩은 2020년 12월 22일에 인공지능 기반의 챗봇 '이루다 1.0'을 공식 출시하였다. 이 서비스는 사용자와의 대화를 통해 학습하는 인공지능 기술을 기반으로 하여, 자연스러운 대화를 지향했다. 그러나 서비스는 출시 초기부터 개인정보 수집에 대한 동의 절차가 충분히 이루어지지 않았으며, 차별적 표현을 사용하는 등의 문제로 사회적 논란이 일었다. 이로 인해 출시 후 약 3주 만에 서비스가 일시 중단되었다. 이 사건은 스캐터랩에 대한 대중의 신뢰 저하를 초래했으나, 회사는 이를 계기로 내부적인 변화와 서비스 개선에 박차를 가하기로 결정했다. 2021년 한 해 동안, 스캐터랩은 개인정보 보호조치를 강화하고, 사용자의 데이터를 취급하는 프로세스를 전면 재검토하였다. 또한, 인공지능 모델이 부적절한 표현을 사용하지 않도록 개선하는 어뷰징 모델을 개발하여 '이루다 2.0'에 적용했다.

2022년 1월, 스캐터랩은 '이루다 2.0'의 클로즈 베타 서비스를 시작하면서, 이전의 문제점을 개선한 새로운 버전을 소비자에게 선보였다. 이 새로운 버전에서는 개인정보 보호와 윤리적 기준을 철저히 준수하는 것을 강조하였다. 회사는 이 과정에서 통해 투명성을 제고하기 위해 홈페이지에 기업 내 변화 과정을 상세히 고지하였으며, 이러한 공개적인 소통 활동은 소비자와의 신뢰를 점진적으로 회복하는 데 기여했다. 스캐터랩의 노력은 기술적인 개선뿐만 아니라, 기업 문화와 고객 서비스에 대한 접근 방식에서도 변화를 가져왔다. 회사는 지속적으로 소비자의 피드백을 수집하고, 이를 서비스 개선에 반영하며, 사용자 경험을 중심으로 한 개발을 지속적으로 추구하고 있다. 이 과정은 기술 기업으로서의 책임과 윤리적인 운영을 강조하는 산업 전반의 트렌드와도 부합한다.

〈그림 6〉 스캐터랩 홈페이지 AI 윤리 항목

2) 이루다 2.0 오픈 베타 테스트에 이르기까지 스캐터랩의 노력과 조치들

- **5회 '업의 본질' 주제의 타운홀:** 전 직원과 함께 AI 챗봇의 가치와 의미에 대한 공유와 토론 진행
- **1년 8개월의 시간:** 데이터베이스 구축, 가명 처리, 어뷰징 모델 마련, 페널티 조치 마련, 베타 테스트
- **5가지 준칙:** 스캐터랩 AI 챗봇 윤리 원칙을 회사가 지향하는 '친밀한 관계' 기준에 맞춰 5가지로 정리
- **5단계 사전 점검:** 알파 테스트, 전문가 테스트, 클로즈 베타테스트, 제한적 오픈 베타테스트, 오픈 베타테스트
- **99% 안전하게 발화한 비율:** 루다가 안전하게 발화한 비율이 99% 이상인지 지속적으로 확인

2. 스캐터랩 AI 챗봇 윤리 점검표 세부 내용

> 스캐터랩의 AI 챗봇 윤리 준칙의 가치를 인공지능 윤리기준 10대 핵심 요건별로 재구성하여 총 21개의 점검 항목으로 제시했습니다. 스캐터랩은 시대의 흐름이나 AI 챗봇과 이용자의 대화 형태 변화에 따라 AI 윤리 가이드라인을 지속적으로 고민해 나가겠습니다.

- **목적**: 스캐터랩의 경험과 사례를 바탕으로 AI 챗봇에 적용할 수 있는 점검문항을 구체적으로 명시해 지속적으로 윤리 기준을 실천합니다. 또한, 기업의 구체적인 사례를 바탕으로 정부와 각계 전문가의 의견을 조합한 최종 결과물을 공개함으로써 인공지능을 개발 및 운영하는 기업과 산업에 도움이 되고자 합니다.
- **구성**: 본 자율점검표는 인공지능 윤리기준의 10대 핵심요건을 기준으로 총 21개의 점검 항목을 제시하였습니다.

핵심 요건	인권보장	프라이버시 보호	다양성 존중	침해금지	공공성	연대성	데이터 관리	책임성	안정성	투명성
문항 수	2	2	4	2	2	1	2	2	2	2

이러한 사례들은 기업이 전반적인 AI 서비스의 라이프사이클 내에서 책임 있는 AI 운영을 위한 전략을 수립하고 실천하고 있음을 보여준다. 이러한 노력은 기업들이 기술적 혁신과 사회적 책임 사이의 균형을 이루는 데 필수적이며, 특히 윤리적 문제와 법적 요구사항에 대응할 수 있는 기반이 된다. 각 단계는 상호 연결되어 있으며, AI 서비스의 전략 및 계획 수립, 생태계 구축, 개발 및 배포, 모니터링 및 보완에 이르기까지 종합적으로 실행될 때 AI의 지속적 개선과 신뢰 구축에 중요한 역할을 한다. 이러한 통합적 접근은 기업이 AI 시스템의 사회적 영향을 고려하여 투명성과 책임을 유지할 수 있게 하며, 장기적으로 소비자 신뢰와 브랜드 이미지 제고에 기여할 수 있을 것이다.

이상의 사례 내용을 정리하면 아래 <표 6>와 같다.

〈표 6〉 기업의 책임있는 AI 활동의 단계별 사례

AI 서비스 단계	주요 윤리 원칙	책임있는 AI 활동(예)
전략 및 계획 수립	책임성 인간의 기술 통제 인간 가치의 증진	• AI윤리원칙의 제정 및 공개 • AI윤리위원회 등 의사결정체계 구축
생태계 구축	지속가능성 인간적 가치 포용성	• 이해관계자와의 공전을 위한 비즈니스 모델 개발 • 이해관계자간 법적, 윤리적 문제가능성 사전 검토
개발 및 배포	개인정보보호 안전 및 보안 공정성	• 기술개발 과정의 공개 • 개인 프라이버스 보호을 위한 데이터 암호화 • 보안 인프라 강화 • 데이터 학습시 공정성 확보를 위한 자가진단 툴킷 제공
모니터링 및 보완	투명성 개인정보 보호	• 운영 상태의 지속적 모니터링을 통해 윤리원칙 준수 여부를 확인 • AI윤리 감사위원회 운영 • AI윤리 프라이버시 관련 보고서 발행

Ⅵ. 결 론

본 연구는 AI 서비스 도입 과정에서 기업이 직면하는 윤리적 문제를 탐구하고, AI 기술의 급속한 발전이 비즈니스 환경에서 갖는 의미를 중심으로 윤리적 이슈와 기업의 대응 전략을 소개하였다. AI 기술은 데이터 분석, 고객 서비스, 제품 개발 등 다양한 분야에서 기업의 생산성과 운영 효율성을 높여주었지만, 개인정보 보호, 데이터 투명성, 알고리즘 편향과 같은 윤리적 문제도 발생시켰다. 이에 본 연구에서는 기업들이 이러한 윤리적 도전에 어떻게 대응하는지를 책임 있는 AI 개념을 적용하여, AI 서비스 라이프사이클별로 구체적인 사례를 통해 살펴보았다. 특히 AI 기술을 선도하는 기업들은 기술적 측면뿐만 아니라 윤리적, 사회적 책임까지 고려해야 한다는 필요성을 인식하고 이에 대해 전략적 대응을 하고 있었다. AI를 윤리적으로 사용하는 것은 기업의 지속 가능성과 신뢰성 강화에 중요한 요소로 작용하여, 고객과의 관계를 강화하고 장기적인 성공을 도모하는 데 중요한 역할을 한다는 인식이 기반에 깔려 있다.

현재 AI 기술과 이를 활용한 서비스 모델이 급속히 확대됨에 따라, 이전에 경험하지 못한 윤리적 문제가 새롭게 대두되고 있다. 이에 정부, 민간 단체, 기업이 함께 AI 윤리 체계를 확립하고자 개별적, 공동적으로 노력하고 있다. 이러한 노력은 AI 윤리가 무엇을 담아야 하는지에 대한 고민에서 출발하며, 이는 윤리 원칙의 제정과 연관된다. 현재까지의 결과를 보면 투명성, 책임성, 개인정보보호, 안전과 보안 등이 주요 원칙으로 공통 제시되

고 있다. 윤리 원칙의 설정은 정부주도, 민간 기관 또는 개별 기업이 담당하고 있으나, 이해관계자를 참여시키는 것도 중요하다.

최근 연구에 따르면 2020년까지 발표된 47개의 인공지능 윤리 원칙 및 가이드라인 중 38%인 18개만이 이해관계자를 윤리 원칙 및 가이드라인 제정에 이해관계자가 관여한 것으로 나타났다(Bélisle-Pipon, Monteferrante, Roy, & Couture, 2022). 윤리 원칙 및 가이드라인 제정에 주로 참여하는 이해관계자로는 산업계 대표자, 학계, 정부부처, NGO, 협회 등의 순으로 나타났는데, 해당 연구 결과는 이해관계자의 관여 정도가 높을수록 윤리 원칙 및 가이드라인의 구체성과 유용성이 높은 것을 보여주고 있다. 위 연구 결과의 특징으로는 민간 기업이 이해관계자의 참여 정도가 가장 낮다는 점이며, 이러한 연구 결과는 윤리 원칙이 보다 포괄적이고 실효성 있는 방향으로 나아가기 위해서는 다양한 이해관계자의 적극적인 참여가 필요함을 시사한다.

AI 기술의 급속한 발전과 보편화는 전세계적으로 다양한 규제환경을 형성하도록 촉진하였다. 앞서 설명한 바와 같이 EU의 경우 가장 적극적으로 AI법안을 제정하여 실효성있게 규제를 진행하고 있어서 세계적인 벤치마크가 되고 있다. 미국은 AI 시장 장악을 위한 정부의 지원이 특징적인데 동시에 정부 차원의 규제도 진행되고 있다. 2023년 7월 바이든 정부는 선도적인 서비스 기술 7개사(Amazon, Anthropic, Google, Inflection, Meta, Microsoft, OpenAI)로부터 안전, 보안, 신뢰의 3항목을 중심으로 하는 Voluntary AI Commitment를 약속받았다. 미국의 접근 방식은 혁신을 촉진하면서도, 소비자 보호와 안전을 보장하는 데 중점을 두고 있으며, 시민의 자유를 보호하는 과정이 민주적 가치와 일치하는 AI 기술의 개발을 강조한다. 중국의 경우 네트워크 안전법, 데이터 안전법, 개인정보보호법이라는 기존의 데이터 3법에 이어 2023년 8월 '생성형 AI 서비스 관리방법'이라는 규제안을 시행했다. 국가 차원에서의 AI 거버넌스 노력뿐 아니라 국제간 협력도 구체적으로 이루어지고 있는데 OECD AI 원칙과 같은 다자간 노력은 AI 사용에 대한 글로벌 기준을 기준을 마련하였다. 국가 간, 그리고 국제 기관 간 협력은 AI 기술의 책임 있는 발전과 윤리적 사용을 촉진하는 포괄적인 글로벌 프레임워크 구축을 촉진하고 있다. 주요 국가간 AI 국제협력 현황에 대한 자료는 한국지능정보사회진흥원(2023)의 연구자료 '인공지능 국제협력 현황 및 특징 분석'에 구체적으로 기술되어 있다.

기업의 AI 윤리에 대한 접근은 크게 두 가지로 나뉜다. 첫째는 규제를 준수하며 최소한의 문제를 해결하자는 규제준수적 접근이고, 둘째는 AI 기술 고도화를 통해 윤리적 문제를 해결하겠다는 기술주도 접근이다. 규제준수적 접근에서는 조직내 컴플라이언스 팀을 포함한 규제관련 대응 팀을 구성하여 법적 문제를 사전에 방지하기 위해 AI 시스템을 관리하고, 국내외 규범을 분석하며 대응 방안을 개발한다. 또한 AI 시스템에 대한 감사 절차를 도입하여 규제사항을 지속적으로 모니터링하고 보고하는 시스템을 갖추는 데 노력을 기울인다. 기술주도 접근은 AI 기술의 효율성을 유지하며 윤리관련 규범을 준수하기 위한 기

술적 혁신을 중시한다. 예를 들어 앞서 설명한 IBM의 AI Fairness 360, Google의 Fairness Indicator, Microsoft의 Fairlearn과 같이 기업은 AI 시스템의 편향을 줄이고 설명 가능성을 높이기 위한 기술을 개발하여 AI 기술을 적용함으로써 발생할 수 있는 윤리적 문제를 사전에 파악하여 공정성을 높이는 방향으로 보완하려고 노력한다. 또한 ZestFinance의 신용분석 시스템, Meta의 Stable Signature 기술은 AI 모델의 의사결정을 기준을 명확하게 설명할 수 있는 기능을 구현하여, 사용자가 AI 결과를 신뢰할 수 있도록 돕는 동시에 규범이 요구하는 투명성과 신뢰성을 충족시키는 시도로 평가된다.

본 연구는 다양한 기업 사례를 통해 AI 도입 과정에서 발생하는 윤리적 문제들을 살펴보았다. 그러나 연구의 범위가 AI 기술개발과 서비스 사업을 선도하는 특정 대기업과 선진국 중심으로 한정되어 중소기업이나 개발도상국에서 겪고 있는 윤리적 문제에 대한 분석은 상대적으로 부족하다. 이러한 한계는 연구 결과의 일반화 가능성을 제한하며 다양한 경제 조건과 문화적 배경을 가진 기업들의 AI 도입 경험을 충분히 반영하지 못했음을 의미한다. 현재 AI 기술의 개발은 글로벌 빅테크 기업들이 경쟁적으로 기술개발을 선도하고 있기 때문에 책임있는 AI와 관련된 활동도 아무래도 이들 기업을 중심으로 살펴볼 수 밖에 없으며 이들의 활동이 기업의 반응을 대표한다고 볼 수 있다. 하지만 향후 AI 기술이 보급되면서 어플리케이션 사용이 확대되면 기술침해나 탈취로부터 취약한 중소 스타트업에 대한 보호를 통한 AI 생태계 확장에 대한 정책적 지원, 예를 들면 AI참여 주체간의 정보공유나 공동 기술개발을 위한 정부의 지원(유재흥 등, 2023)이 필요할 것으로 생각된다.

또한 AI 기술의 빠른 발전과 확산으로 인해 본 연구는 최신 기술 동향과 새롭게 등장하는 윤리적 문제들을 포괄적으로 다루지 못한 것이 한계로 지적될 수 있다. 향후 연구에서는 AI에 대한 윤리적 문제는 산업별, 지역별로 다르게 나타날 수 있음을 고려하여 다양한 산업과 지역에 걸쳐 연구를 확대할 필요가 있다. 특히 국가별로 법적 규제와 문화적 특성에 따른 윤리적 이슈 민감도가 다르므로, 국가별 비교 연구가 필요하며, 이를 통해 보다 효과적인 AI 원칙 제정과 AI 윤리에 관한 글로벌 스탠다드를 마련하는 데도 기여할 수 있을 것이다.

본 연구에서는 AI 윤리의 철학적 기반, 그리고 AI 윤리의 핵심 내용이 무엇인지에 대한 심도있는 이론적 고찰이 이루어지지 않았다. 예를 들면 AI 윤리에서 강조하는 다양한 원칙이 기존 윤리학에서 강조하는 공리주의(utililtarianism)나 덕 윤리(virtue ethics)와 어떤 관련성을 가지고 있는지를 살펴볼 필요가 있을 것이다. 하지만 본 연구는 국내외의 다양한 AI 윤리기준을 소개하고 이를 실천하기 위해 기업이 어떠한 활동을 하는지에 관한 구체적인 사례를 제시하는 것을 연구의 목표로 하였기 때문에, AI 윤리가 과연 무엇이 되어야 하는지, 그리고 이들이 어떤 윤리적이고 철학적인 기준에 따라 도출되는지에 관한 이론적인 부분에 대해서는 충분히 설명되지 않았다. 본 연구에서 다루어지지 않았다고 해서 이러한 연구 주제가 가지는 의미가 중요하지 않다는 것은 아니며 최근 발간되고 있는 철학 전

문 서적을 참고할 수 있을 것이다(이중원, 2019).

본 연구는 기업이 AI를 책임감 있게 도입하고 운영하는 방법에 대한 사례를 제공하여, AI 윤리 연구자와 실무자들에게 유용한 자료를 제공하고자 하였다. AI 기술의 사회적 영향을 폭넓게 이해하고, 기업과 사회가 이를 책임감 있게 활용할 수 있는 방안을 모색하는 데 본 연구가 기여할 수 있기를 바란다.

[참고 문헌]

김명주. (2017). 인공지능 윤리의 필요성과 국내외 동향. 「한국통신학회지」, 34(10), 45-54.

리 카이푸. (2019). 「AI 슈퍼파워」 이콘.

박소영. (2023). 인공지능의 FATE(공정성·책임성·투명성·윤리의식)를 위한 입법 논의 동향과 시사점. 「이슈와 논점」, 2111.

박찬준·이원성·김윤기·김지후·이활석. (2023). 초거대 언어모델 연구 동향. 「정보과학회지」, 41(11), 8-24.

배영임·김유나. (2022). 「인공지능의 윤리적 쟁점에 관한 탐색적 연구」. 경기연구원

삼일 PWC경영연구원. (2023). 「책임있는 AI」. https://www.pwc.com/kr/ko/insights/issue-brief/responsible-ai.html

심진보. (2021). 「글로벌 주요국들의 인공지능 생태계 현황 분석」, 한국콘텐츠학회 2021 국내종합학술대회

유순덕. (2020). 인공지능 비즈니스 생태계 연구. 「한국인터넷방송통신학회 논문지」, 20(2), 21-27.

유재홍·안성원·안미소·노재원. (2023). 「생성 AI산업 생태계 현황과 과제」, 소프트웨어 정책연구소.

이우권. (2020). 「인공지능의 윤리적 사용을 위한 정책대안에 관한 연구」. 한국정책논집, 20(1), 1-15.

이재호·조남용. (2023). 「사람과 공존하는 AI의 필요조건, AI 공정성」, 삼성SDS 인사이트 리포트, https://www.samsungsds.com/kr/insights/ai-fairness.html

이중원 편. (2019) 「인공지능의 윤리학」, 한울아카데미.

장진철·노재원·유재홍·조지연. (2024). 「책임있는 AI를 위한 기업의 노력과 시사점」, 소프트웨어정책연구소.

최경석. (2020). 인공지능이 인간 같은 행위자가 될 수 있나? 「생명윤리」, 21(1), 71-85.

한국지능정보사회진흥원. (2022). 「주요국 인공지능 거버넌스 분석(상)」. 한국지능정보사회진흥원.

한국지능정보사회진흥원. (2023). 「인공지능(AI) 국제협력 현황 및 특징 분석」, 10호(2023. 12. 31.)

AI Incident Database. (2023). Retrieved from https://incidentdatabase.ai

Bélisle-Pipon, J. C., Monteferrante, E., Roy M. C., & Couture, V. (2022). Artificial intelligence ethics has a black box problem. *Ai & Society*. 38(4), 1507-1522.

Caltech Science Exchange. (2022). *Artificial Intelligence: The Good, the Bad, and the Ugly* [Video]. Yaser Abu-Mostafa. Caltech Science Exchange. Retrieved from https://scienceexchange.caltech.edu/topics/artificial-intelligence-research/artificial-intelligence-experts/yaser-abu-mostafa-artificial-intelligence

European Commission. (2018). *Directorate-General for Research and Innovation and European Group on Ethics in Science and New Technologies, Statement on artificial intelligence, robotics and 'autonomous' systems,* Publications Office, 2018, https://data.europa.eu/doi/10.2777/531856.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI,* The Berkman Klein Center for Internet & Society at Harvard University.

Floridi, L. (2013). Distributed morality in an information society. *Science and Engineering Ethics,* 19(3), 727-743.

Huang, C., Zhang, Z., Mao, B., & Yao, X. (2023). An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence,* 4(4), 807. https://doi.org/10.1109/TAI.2022.3194503

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence.* 1, 389-399. 2019.

Leslie, D. (2019). *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector.* The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529

McKinsey & Company. (2024). *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value.*

MIT Sloan Management Review. (2023). *Ethically Sourced Creativity: Shutterstock's Alessandra Sala* [Audio interview]. Retrieved from https://sloanreview.mit.edu/audio/ethically-sourced-creativity-shutterstocks-alessandra-sala/

Moore, J. F. (1993). Predators and prey: the new ecology of competition. *Harvard Business Review,* 71(3), 75-83.

Russell, S., & Norvig, P. (2020), *Artificial Intelligence: A Modern Approach (Pearson Series in Artifical Intelligence),* 4th Edition, Pearson.

Sajid, H. (2024). Why continuous monitoring is essential for maintaining AI integrity. WiseCube. Retrieved from https://www.wisecube.ai/blog/why-continuous-monitoring-is-essential-for-maintaining-ai-integrity/

Zhou, J., Chen, F., Berry, A., Reed, M., Zhang, S., & Savage, S. (2020). A survey on ethical principles of AI and implementations. *2020 IEEE Symposium Series on Computational Intelligence(SSCI),* Canberra, ACT, Australia, 3010-3017.

AI Ethics and Firms'Responsible AI Activities

Jihye Kim* ·Bongsoon Cho**

With the rapid development of artificial intelligence (AI) technology, companies are using AI to improve operational efficiency and create new business models. However, with the increased use of AI comes unexpected ethical challenges, such as algorithmic bias, privacy infringement, and copyright issues. As a result, there is a growing social demand for companies that provide AI technologies and services not only to improve their technical performance, but also to use AI responsibly and fulfill their social and ethical responsibilities. In order to develop and deploy AI technologies responsibly, companies are developing a direction for their responses to the ethical aspects of AI under the name of "Responsible AI."

This study first reviews the ethical principles related to AI. In addition, companies' responsible AI activities for practicing AI ethical principles are introduced at four stages according to the AI lifecycle: strategy and planning stage, ecosystem constructio stage, development and deployment stage, and monitoring and supplementation stage.

Keywords: artificial intelligence, AI ethics, responsible AI, ethical principles

* Innocrew Co. Manager, First Author
** Sogang University Professor, Corresponding Author

■ 저자 소개

김지혜(Jihye Kim) 서강대학교에서 경영학석사를 취득하였으며, 현재 주식회사 이노크루의 교육사업본부 본부장으로 재직 중이다. 윤리경영, 청렴 및 반부패와 관련한 교육 사업을 운영하며, 기업의 지속가능한 발전을 위한 교육 프로그램 개발과 관련한 연구를 하고 있다.

조봉순(Bongsoon Cho) State University of New York at Buffalo에서 조직행동으로 박사학위를 취득하였고 현재 서강대학교 경영대학 교수로 재직 중이다. 리더십, 조직문화, 갈등관리, 인적자원관리 전반적 이슈에 대한 연구를 하고 있다. 그동안 「윤리경영연구」, 「인사조직연구」, 「경영학연구」, 「Journal of Organizational Behavior」, 「Human Resource Management」, 「International Journal of Human Resource Management」 등의 학술지에 여러 편의 논문을 게재해왔다.