인공지능 규제: 공식화와 거버넌스를 위한 메타프레임워크

요약

이 논문에서는 공식화에서 지속 가능한 거버넌스에 이르기까지 국제 공공정책 수립의 모든 단계를 아우르는 인공지능(AI) 규제의 메타 프레임워크를 제시한다. 2009~2019년 출간된 인공지능 규제(AIR) 관련 문헌에 대한 방대한 체계적 검토를 바탕으로, 연구와 실천을 위한 복잡한 과학적 시나리오를 만들어낸 15개의 독특한 프레임워크와 여러 이론이 담긴 '프레임워크'라는 레이블 아래 구성된 지식의 분산체가 확인됐다. '애자일과 윤리'만큼 다양한 이론과 원칙이 발견됐다. 따라서 이러한 대량의 지식을 응집력 있고 합성적이며 일반적인 이론적 도구로 통합하기 위한 구조화된 분석 방법이 뒤따라 개발되었다. 그 결과로 만들어진 '에어 프레임워크'는 인공지능의 사용과 적용이 무엇을, 언제, 어떻게 규제해야 하는지와 관련하여 사회가 집단적으로 사고하고 정보에 입각한 정책 결정을 내릴 수 있는 신뢰할 만한 관점을 제공한다. 게다가, 새로운 프레임워크는 이 지역의 최신 발전을 미래의 연구가 프레임화되어 출판된 문헌에 추가될 수 있는 형식으로 정리하고 있다. 인공지능이 사회에 미치는 (잠재적) 영향은 막대하며, 따라서 이 기술의 담론, 사회적 협상 및 적용은 용어, 거버넌스, 사회적 가치 측면에서 공통적인 근거에 의해 제시되어야 한다.

1. 소개

인공 지능이 우리의 일상적인 행동과 눈에 띄지 않는 방식으로 널리 보급됨에 따라, 예외적인 개념과 시나리오에서 전례없는 법적 문제를 도입했다.

같은 관점에서 기계 학습 솔루션에서 데이터 처리의 불투명성은 법적으로 이례적인 문제를 일으킬 가능성을 증가시킨다.

연구에서는 이러한 반향을 바탕으로 인공지능 규정과 관련이 있는 문헌을 방대하게 탐색하고, 개인 및 법인과 시스템 간의 운영과 관계를 내장된 인공지능으로 규제하기 위해 성찰과 행동이 가능한 통합적인 이론적 틀로 처리, 그룹화하고자 했다.

2. 인공지능을 규제하는 이유

매일 제공되는 여러 '지능형 솔루션'중 의료용 로봇, 드론, 자율주행차 등의 각기 다른 시스템과 관련된 책임, 보안, 지적 재산권, 프라이버시에 대한 의문이 제기되고 있다. 위험 관련 의사 결정 수준을 보여주면서, 개발자가 알고리즘에 대한 전략적 방어를 가르치기 위해 게임 이론을 사용하는 경우 기계 학습과 게임 이론을 결합하여 사용하고 있다. 두 알고리즘 간의 게임은 자원이 절대적으로 부족할 때만 다른 알고리즘의 프로세스를 끝낼 것이라고 예측했다. 그러나, 더 지능적인 알고리즘이 도입되었을 때, 그것은 즉시 약한 알고리즘을 중지시켰다. 이 사례는 특정 규칙을 따라야 할 뿐만 아니라 복잡한 윤리적 결정을 내려야 하는 상황에서 자율적 체계를 발견할 수밖에 없다는 생각을 강화한다.

이러한 모든 위험을 고려할 때 인공지능이 창출한 기회를 활용하기 위해서는 새로운 도덕적 책임 귀속 모델을 위임하고 정의하기 위한 모범 사례를 수립하는 것이 중요하다. 위험 평가 모델은 빅데이터 및 인공지능 응용 프로그램에 대한 지원과 유연성을 제공할 수 있다. 비록 실천하는 것이 어렵지만, 이 초지능에 대한 도덕성을 높이는 것이 우선순위가 되어야 한다. 윤리적이고 도덕적인 관점에서, 결정은 윤리적인 틀의 원칙에 위배되지 않는 한 받아들일 수 있다고 간주된다.

규제 이유에는 제조업체가 안정적으로 운영할 수 있는 법적 체계를 이해할 필요성, 소비자와 사회가 그들에게 해를 끼치거나 악영향을 미칠 수 있는 장치로부터 보호할 필요성 및 비즈니스 기회의 필요성이 포함된다.

아직 규제가 부족한 산업에서는 혁신은 자유롭게 허용하되 특정 유형의 피해가 발생할 경우 담당자가 부담해야 한다는 것이 일반적인 접근이다.

3. 규제할 수 있는 최선의 방법을 찾는 것

행동 패턴을 표준화하려는 시도를 나타내기 위해 사용되었을 때, '규제'라는 용어는 법을 의미하는 것으로 가정된다. 현재는 인공지능이 지원하는 제품과 서비스가 초래하는 피해를 사법적으로 해결하기 위해 몇 안 되는 기존 법에 의존하고 있다. 한편으로는 사례가 늘어나고 있고 다른 한편으로는 입법부는 기술적 진보에 비해 느린 속도로 움직이고 있는 것으로 보인다.

세계적으로 생산되고 있고 상용화된 기술과 로봇이 만든 발명품을 다루는 법의 범위에 대한 문제가 아직 해결되지 않았다. 그러한 문제는 기술 산업에 구축된 복잡한 네트워크를 고려하면 훨씬 더 넓은 차원에 도달할 수 있기 때문에 전 세계에 분산된 데이터로부터 제품이 학습의 대상이 될 수 있다.

대규모 데이터 분석은 인공지능 규제 딜레마와 관련된 핵심 과제가 적절하게 생산되고 배치된다는 것을 입증하고 있음을 밝혔다. 가장 옹호되는 전략 중 하나는 지능형 시스템을 훈련시킬 때 활용되는 투명성, 전체 생산 과정, 특히 의사 결정 규칙, 방법 및 기반을 개방하는 것이다. 특정 상황에서는 알고리즘을 통해 규제가 시행되어야 할 것이다. 그래서 자율 시스템에는 매개 변수가 미리 정의된 표준 내에 있는지 확인하는 보호자 알고리즘이 있다. 개방형 데이터와 유사한 전략은 세계적인 이해를 지향하는 코딩 모델을 만들기 위한 '설명 가능한 인공지능'(Intelignable Artificial Intelligence, XAI) 표준이다.

인공지능을 규제하기 위해 제안된 이론 중 일부는 계약적·외계약적 책임 또는 엄격한 책임에 기반을 두고 있으며, 인공지능의 경우 설계자, 규제기관, 사용자 사이에 도덕적 책임이 분산되어 있기 때문에 흠잡을 데 없는 책임 모델을 채택하고 있다. 로봇의 행동에 책임을 묻려는 시도로 인해 일부 국가가 각 유닛에 법적 정체성을 부여할 수 있는 가능성을 고려하게 되었다. 계약 관계에 있는 당사자들이 다른 독립체에 의해 법적으로 대표될 수 있다면, 시스템 또한 마찬가지라고 주장할 수 있다. 그에 대한 반론으로는 로봇에 대한 손해배상 청구가 불가능한 상황에서 '로봇책임'이라는 용어를 '로봇에 대한 간접책임'으로 대체해야 하므로 형사책임을 물을 수 없다. 따라서, 그러한 제품들이 사회에 미치는 영향 또한 책임이 되어야 한다.

또한 인공지능 규제에 동기를 부여하는 우려 중에는 일자리 감소에 대항하는 것을 목표로 연구 모델의 혼란을 최소화하는 것을 목표로 하는 접근법이 있다.

규제 대상 영역에 관심이 집중되면서 규제 대상에 대해 제대로 알지 못한 채 디지털 기술을 법제화하려는 시도에 대한 비판이 이루어지고 있다. 이러한 위험을 최소화하기 위해서는 점진적인 규제 전략을 사용할 수 있다. 위험을 완화하기 위해 규제 기관은 윤리를 기반으로 한 테스트를 통해 안전성과 효능이 입증될 때까지 특정 알고리즘을 시장에 도입하는 것을 금지할 수 있다.

차별에 저항하고 스마트 알고리즘을 기반으로 한 의사 결정에 대한 설명권을 강화하는 일반 데이터 보호 규정(GDPR)이 완성되었다. 2017년 유럽의회 법무위원회는 유럽 로봇 및 인공지능 기관의 창설을 권고하는 보고서를 발표하고 규제 모델의 진화와 관련된 복잡성을 감안할 때 엄격하고 부드러운 법률들을 조합하는 것을 제안한다. 유럽 입법에서 또 다른 중요한 점은

규제 체계를 만들 필요성을 강조하는 하원이 발표한 보고서였다. 미국에서는 인공지능의 개발과 시행에 관한 연방자문위원회 설립을 통해 인공지능을 활용하고 상용화하기 위한 조건을 정의하고자 하는 HR4625가 제시되었다. 또한, 몇몇 국가들은 인공지능의 개발과 사용을 규제하기 위한 정책과 법률을 만들겠다는 의사를 보여 왔다.

4. 방법

문헌에서 발견된 인공지능 규제에 대한 국제 논의를 조사하는 것을 목표로 출판물 통계 분석을 진행하고, 논문을 체계적으로 검색하고 목록화하여 해당 논쟁의 심화 과정이 향후 규제 노력의 근거가 되고 있음을 입증하기 위한 질적 분석을 수행했다.

이 연구를 위해 ScienceDirect, JSTOR, SpringerLink, PROQUEST, IEEE, Scopus, DOAJ 및 Google Scholar 데이터베이스에서 제목 및 주제별로 검색하여 2009년에서 2019년 6월 사이에 발표된 자료를 수집하였다. 영어로 발표된 상호 검토된 연구 논문만 수집하였고, 중복은 허용하지 않았다. 표본에서 근위 토론 사례를 제거하는 것을 목표로 모든 주제의 논문을 읽고, 규제가 논의중이며 주요 주제가 아닌 주제를 파악하여 선정하였다.

이 샘플은 인구 통계를 구조화 할 때 출판 연도, 저널, 저자, 저자 기관, 저자의 연구 분야, 국가, 키워드와 같은 특정 매개 변수에 따라 분류되었다. 우리는 또한 각 기사에 대한 개념, 발견, 기여도, 의제, 접근법, 방법 및 연구 주제를 적었다. 주제를 분류할 때 '위험', '윤리', '규제 방법', '기존 규제', '프레임워크'등의 용어를 고려하였다. 초록을 분석한 후, 추가적인 읽기와 토론을 위해 51개의 기사로 구성된 샘플을 선택하였다.

5. 결과

타임라인 측면에서는 2015년 이후 논문의 94%가 게재되었으며, 이후 매년 생산량이 증가하고 있음을 강조할 필요가 있다. 47개의 논문에서 배타적인 질적 접근법이 지배적으로 나타난 반면, 오직 4개의 논문에서만 혼합된 접근 방식이 발견되었다. 이것은 그러한 초기 주제를 다룰 때 예측되는 점이다. 초기 논쟁은 이 경우에 탐구적이며, 이는 질적연구의 상당수를 설명한다.

이 샘플은 인공지능 규정에 관심을 갖는 연구 분야의 진보를 반영한다. 연구 대상인 인공지능은 전통적으로 컴퓨터과학(IT)과 공학에 속하지만 법학·경영학·철학 등 다른 분야에서 규정에 대한 관심이 높아지고 있다. 전체 표본 중 법학 분야 연구자가 53%를 차지하고 있고, 정보기술(IT)이 43%로 그 뒤를 이었다. 경우에 따라 같은 논문을 다른 분야의 연구자가 공동으로 작성하는 경우도 있다.

비 배타적으로 '위험'(41%), '윤리'(16%), '규정 방법'(65%), '기존 규정'(8%), '프레임워크'(26%)로 나눠 표본의 주요 대상 분석에 각별한 주의를 기울였다. 주목할 점은 인공지능에 적용되는 위험과 윤리에 대한 우려가 해를 거듭할수록 지속되고 있다는 점이다. 그러나, 어떻게 규제할지에 대한 논의는 2016년에서야 중요해졌다. 인공지능 규제 프레임워크 논의와 관련해서는 2017년 이후 최대 건수의 논의가 발생했으며 표본에서 발견된 최대 15건의 제안을 추가해 다음에 발표·분석할 예정이다.

5.1. 실험 기술의 윤리적 문제를 위한 모델

로봇이 실험적 기술이라는 것을 전제로, 이 모델은 자율 시스템이 내린 결정과 관련된 윤리적 딜레마를 최소화하고자 한다. 이 제안은 로봇이 사람 및 환경과 상호 작용함에 따라 잠재

적인 윤리적 문제를 예상하기 위해 구축된 실험 기술을 배치하기 위한 16가지 조건을 기반으로 의사결정 프로세스를 지원한다. 조건은 3개 그룹으로 나뉘어 피해예방(남성이 아닌 조건), 선행(선의 조건), 자율성과 정의 존중 등을 목표로 하고 있다. 위험에 대한 우려는 "빨간 버튼" 조건에 대한 예측으로 확대된다. 또한 점진적인 상호 작용 전략의 일환으로이 모델을 구현하는 것을 권장한다.

5.2. 대화형 규제 거버넌스 모델

이 제안은 이해 관계자의 귀속이 강조되는 기술 개발 및 법률 제정 프로세스를위한 대화 형 거버넌스 모델을 기반으로한다. '규제혁신'과 '일시적 실험입법'이라는 표현을 사용하고, 혁신의 수명주기 성숙 단계에서 에이전트 간의 적절한 행동 순서를 고려하는 등 지속적인 학습과 법적 틀의 점진적 진화가 필요하다는 점은 주목할 만하다. 제안된 모델은 다음과 같은 구성 요소를 포함한다.

- 기존 법률에 따라 로봇에 대한 새로운 개념 모델을 만들도록 안내하는 R2T(Regulation-to-Technology) 거시 프로세스.
- 기술적 진화에 따른 필요에 따라 법을 조정하는 T2R(Technology-To-Regulation) 거시적 프로세스.
- R2T와 T2R이 공유하는 데이터 저장소.

이 하이브리드 인공지능 거버넌스 모델의 주요 이점 중 점진적 전략으로 톱다운 방식과 상향식 규제 행동의 통합을 강조하여 새롭고 끊임없이 변화하는 대상을 규제함으로써 발생하는 위험을 최소화할 필요가 있다.

5.3. 인공지능 개발과 배치를 위한 윤리 모델

철학적 원칙과 인권과 복지 유지 차원에 기초한 인공지능 개발과 배치를 위해 제안된 윤리적 프레임워크는 인공지능의 핵심 기능을 위한 윤리적 관점, 권리(의무론적 윤리), 손해배상과 상품(목적론적 윤리), 덕(이태 윤리), 공동체(공동체 윤리), 대화(소통 윤리), 번영(번창성 윤리)으로 나뉜다. 이러한 핵심 기능은 윤리적 문제의 식별, 인간 의식의 발달, 협력 참여, 책임 및 인공 지능의 무결성 등과 같은 것으로 간주된다.

5.4. 역량기반 인공지능 규제모델

각 국가권력이 가진 역량과 장단점을 고려하여 책임 배분을 하는 것을 바탕으로 한 '인공지능 규제모델'을 제시하였다. 이 모델은 규제 과정의 대리인으로서 행정부, 입법부, 사법부의 규제 기관을 인정한다.

제안된 모델에서 법제처는 사용자와 사회안전 측면에서 인공지능을 사용하는 제품과 서비스를 인증하는 권한을 규제기관에 부여하는 법령을 마련하기로 했다. 연구자 그룹의 지원을 받아, 규제 기관은 기술 발전을 모니터링하고, 지능형 학습 과정과 인공지능 활용에서의 위험을 파악하고, 기술 추천을 하고, 기술이 공표된 목적에 맞게 적용되고 있는지 여부를 확인하는 데 더 민첩하고 유능하게 반응할 수 있을 것이다. 기업의 제품이나 서비스에 손해가 발생하면 인증을 받은 기업은 더 관대한 규칙에 따라 판단을 받는 반면, 인증을 받지 않은 기업은 더 엄격한 규칙의 적용을 받는다. 법원은 그 기관들이 인증을 진행중인 상황을 고려하여 기업들이 입은 손실과 손해에 대해 판단할 것이다.

5.5. 사회에 의해 유지되는 규제 모델

사회 계약 이론으로부터 영감을 받은, 사회에 의해 지속되는 규제 모델은 '맨 인 루프'모델을 '소사이어티 인 루프' 모델로 바로잡는다.

대화형 학습 머신(맨 인 루프)은 사용자의 피드백으로부터 민첩성과 효율성을 배우므로, 이를 통해 생성된 지식들이 풍부해진다. 그것이 제품과 서비스에 인공 지능을 사용함으로써 발생하는 문제를 학습하는 데 사용되었다면, 소사이어티 인 루프는 사회가 이러한 요소를 통제하고 능동적으로 식별할 수 있도록 하는 거버넌스 도구가 될 것이다. 안전, 개인 정보 보호 및 정의와 관련된 개념 간의 갈등은 이 모델을 통해 이익을 볼 수 있을 것이다. 이 관계는 다음과 같이 요약될 수 있다. 소사이어티 인 루프 = 맨 인 루프 + 사회적 계약.

5.6. 로봇공학의 원리

로봇과 관련된 모든 에이전트의 책임을 강조하면서 로봇 설계자, 제조업체 및 사용자에 대한 5가지 원칙을 수립했다. 이 원칙의 주요 목적은, 로봇이 도구인 반면 인간은 실제 책임이 있는 주체라는 점을 강조하는 것이다.

규제기관이 수행하는 감사에 이 원칙을 사용할 기회를 파악할 수 있고, 이것은 입법 과정에서의 수정이나 제작 과정에 반드시 반영되어야만 한다.

5.7. 애자일 인공지능 거버넌스

이는 형식적인 규제 조치와 딥러닝과 머신러닝 기반 제품 및 서비스의 생산 및 상용화 간의 시간적 불일치 문제에 대한 대안이 될 것이다. 이 제안의 성공은 시장과 학자, 정부, 보험사, 조직화된 시민사회가 얼마나 노력하느냐에 달려 있다. 이 모델은 거버넌스 조정 위원회와 글로벌 거버넌스 조정 위원회가 수행하는 행동을 예측한다. 국제적 접근도 현재 상황에 따라 이들이 더욱 취약해지고 있다는 점을 고려하여 아직 인공지능 규제 역학에 참여하고 있지 않은 몇몇 국가에 균형을 제공하기 위한 수단으로 거론되고 있다. 이 모델은 법안이 작성되는 동안 위험을 완화하는 권고지침(soft law)이다. 소프트 거버넌스 부분에는 업계 표준, 소셜 코드, 연구실, 인증 관행, 절차 및 프로그램이 포함된다. 하드 거버넌스 부분에서는 법, 규정, 규제 그룹을 집중적으로 다룬다.

제안된 모델은 입법이 성숙해지는 동안 실제 표준 수립을 강화하는 방식으로 인공지능을 다루기 위해 관계 네트워크를 고려한다.

5.8. 지속 가능한 인공지능 개발

AI기반 솔루션의 전체 라이프 사이클에 대한 우려는 지속 가능한 AI개발(SAID) 프레임워크를 고안할 때 고려하는 주요 토대였다. 거버넌스 구조의 렌즈 아래에서 분석된 SAID는 기술(데이터, 아키텍처 및 알고리즘 설계), 소셜(사회 영역에서 인공지능 활용의 잠재적 결과 분석), 거버넌스(알고리즘이 국가 및 국제 의사 결정에 영향을 미치는 방식)와 같이 계층화된다.

5.9. 로봇 자동화를 위한 윤리적 프레임워크

인공지능을 이용한 자동화에 여러 이해관계자를 통합하는 것과 관련하여, 이 프레임워크는 인공지능 사용에 대한 윤리적 근거를 찾기 위해 이해관계자 이론(Stakeholders Theory)과 사회 계약 이론(Social Contract Theory)을 통합한다. 이 제안에서는 이해 관계자를 근로자, 시장, 정부, 경제 및 사회 전반으로 간주한다. 그리고 고용 시장에 미치는 영향과 이해 관계자

간의 새로운 행동과 관계가 크게 강조된다. 이 프레임워크는 이해당사자 파악에서 사회계약 분석, 영향평가, 마지막으로 근로계약 종료나 위반 위험을 줄이기 위한 조치에 이르는 일련의 단계를 따른다.

제품과 서비스에 인공지능을 도입함에 따라 직업을 변경하게 되는 노동자들을 이해당사자로서 고려하는 유일한 제안이라는 점에 주목할 필요가 있다.

5.10. 학습 알고리즘을 규제하기 위한 지능형 모델

편향을 포함하는 지능형 서비스에 대항하는 전략에 초점을 맞춘 이 모델은 알고리즘이 머신러닝 프로세스의 기본 요소(데이터, 테스트 알고리즘, 의사 결정 모델)를 평가해야 한다고 제안한다. 이 제안은 편견 없는 솔루션을 보장하기에는 코드의 투명성이 부족하다는 명제에 근거하며, 방대한 양의 데이터로 학습을 시킬 때에도 여전히 편향이 존재할 수 있음을 인정한다. 또한 알고리즘이 복잡성이 커짐에 따라 이러한 문제를 자동으로 식별하는 데 어려움이 있다는 것을 인식하고 있다. 또한 다음 연구는 학습 과정에서 편향 관련 문제를 감지하는 것으로 분류할 수 있는 알고리즘의 특성을 분석한다.

5.11. 틀로서의 보편적 인권 선언

이 모델은 각 특정 윤리 영역과 관련된 여러 체계가 국제적 규모로 민간이나 정부 내에서 인공지능을 규제하기에는 미흡하다는 주장에 따른 것이다. 그 격차로 인해 세계인권선언(the Universal Declaration of Human Rights)은 사회에 미치는 영향에 따른 인공지능의 효과적인 규제에 필요한 접근으로 여겨졌다.

5.12. 로봇의 윤리적 평가를 위한 소프트웨어 요구 모델

이 제안은 프로젝트 수립 중 로봇을 평가하기 위한 시스템에서 고려해야 할 일련의 일반적인 사양을 제시한다. 사용자의 감정 상태와 같은 제안된 사양에는 다른 요소가 고려된다. 그 제안은 업계와 규제기관 모두에 의해 활용될 수 있는 것으로 보인다. 두 경우 모두 로봇 프로젝트에서 빨간 버튼이 필요하다는 신호를 보내는 첫 번째 빨간 깃발이 될 수 있다.

5.13. 윤리적 결정에 대한 윤리적 판단 모델

윤리적 결정을 해결하는 것이 피하는 것보다 낫다는 점을 고려하여 저자는 그 결정을 내리고 설명할 수 있는 능력을 모두 갖춘 윤리적 딜레마에 직면한 에이전트에서 구현할 수 있는 형식적인 논리 모델을 제안한다.

모델의 기능을 구축할 때 '결정', '사건', '효과'의 개념을 고려하였고, 윤리적 기본 원칙 또한 반영하였다. 결과론적 윤리, 의무론적 윤리 및 이중 효과 원칙은 불수용(┬), 불수용(┴) 또는 미확정(?)이라는 세 가지 가능한 결과를 반환하는 판단 함수에서 공식으로서 사용되었다.

5.14. 아실로마 인공지능 원리

아실로마 컨퍼런스(Asilomar Conference)에서 제안한 거버넌스 모델은 수천명의 전문가가 서명한 23개의 인공지능 원칙이다. ‘연구 문제’, ‘윤리와 가치’, ‘장단기 문제’로 분류된 이 원칙은 동기 및 자금 지원에서부터 그 영향에 대한 이익 및 판단 기준의 평가에 이르기까지 AI가 내장된 제품 또는 서비스의 생애 주기에 모두 적용된다.

5.15. 신뢰할 수 있는 인공지능을 위한 유럽 윤리 지침

유럽연합 집행위원회는 새로운 인공지능 거버넌스를 지향하기 위한 가이드라인을 만드는 것을 목표로 전문가 그룹을 통해 매우 포괄적인 구조를 바탕으로 3단계로 나눠 '신뢰할 수 있는 인공지능 윤리지침'을 작성했다. 가장 높은 계층은 기본적인 인권에 기초한 네 가지 윤리적 원칙을 다룬다. 두 번째 계층에는 라이프 사이클 전반에 걸쳐 AI 기반 시스템 또는 서비스에 대한 애플리케이션 및 지속적인 평가에 필요한 일곱 가지 핵심 요구 사항이 포함되어 있다. 기본 계층을 위해서는, 각 특정 시스템에 대한 상위 계층의 주요 요구 사항을 충족하도록 한 권장 사항 목록이 작성되었다.

6. 프레임워크 접근법

샘플에서 제안된 15가지 프레임워크를 통해 각각 채택한 접근법을 분석하여 표 1과 같은 결과를 도출했다. 윤리 지침이 존재한다는 사실만으로는 소프트웨어 개발 산업에 어떠한 영향을 미치기에 충분하지 않다. 따라서 윤리적 원칙에 기반을 둔 모델은 그러한 권장 사항을 충족시킬 수 있도록 법적 메커니즘을 필요로 한다.

표 1 - 저자가 작성한 프레임워크에서 탐구된 접근법의 비교 표

접근법	30	14	32	33	34	36	37	38	23	18	39	41	43	44	46
제도권 역량		■		■			■								■
국제적 접근							■	■			■				■
애자일 사고							■								
하이브리드		■		■			■							■	
연속 상호작용		■												■	
규제기관		■		■			■								■
규정 연표	■	■		■			■	■						■	■
점진적 개선	■	■		■	■		■							■	■
윤리원칙	■		■			■			■		■	■	■	■	■
사회계약					■				■						
고용시장									■						
이해관계자 영향		■	■			■	■		■		■			■	■
지배구조		■	■	■			■	■	■		■				■
프로세스 정의		■							■	■					■
규제기관으로서의 기술										■		■	■		

사회계약을 고려하는 프레임워크는 정부와의 공동 제작에 사회가 참여하는데 가장 개방적인 프레임워크 중 하나이다. 그 모델들은 시민들을 뛰어난 이해관계자로 본다. 고용시장에 미치는 영향에 대한 우려도 이해관계자에게 미치는 영향을 평가하는 방법이다. 규제를 점진적으로 도입하는 것은 리스크 완화 전략이지만 법제처와 규제기관 간의 연속적인 상호작용과도 결합할 수 있어 입법 과정에서 지속적인 개선이 가능하다.

대화형 규제 거버넌스 모델, 역량 기반 규제 모델, 애자일 거버넌스, 아실로마 원칙 및 신뢰할 수 있는 AI 제안을 위한 유럽 윤리 지침은 더 많은 주제를 포괄한다. 유럽의 제안은 신뢰할 수 있는 AI가 합법적이고 윤리적이며 강력해야 한다고 강조한다. 그 외에 규제 과정에 관련된 모든 당사자 간의 관계와 더 경직되고 유연한 메커니즘 사이에서 균형을 찾기 위해 연구하기도 한다. 애자일 거버넌스 제안은 모두 입법부가 참여하는 '대화형 규제 거버넌스모델'과 '역량기반 규제모델'이라는 정식 규제에 대한 기존의 조치를 배제하지 않는다는 점에 주목할

필요가 있다. 따라서 이는 합의된 기준이 합의되어 시행되고, 법적 장치가 공식화되지 않은 상태에서 위험이 완화되는 과도기적 상황을 구성하고 있는데, 이는 피드백이 규제 수단의 성숙의 근거가 되는 동적 규제의 개념과 매우 유사하다.

AI4PEOPEL에서 제시한 제안과 프레임워크 간의 관계는 간과해서는 안된다. 인공지능을 가장 잘 사용하는 방법에 대한 기준을 설정해야 한다는 주장들을 분석할 때, 기술 개혁을 둘러싼 경쟁을 발전시킬 방법을 확인할 수 있었다.

이러한 조치의 범위는 이러한 우려를 제기한 샘플의 모델보다 더욱 이해관계자를 포괄적으로 포함한다. 글로벌 인공지능 규제 중심의 행동 간 시너지 발굴의 필요성을 고민하고, 인공지능 애자일 거버넌스에서는 국가별 거버넌스위원회와 국제적 맥락에서의 글로벌 거버넌스위원회 신설이 제안됐는데, 이는 샘플에 포함된 2건의 문건에서도 언급되었다.

기존 소프트웨어 기반 규제 모델의 수가 적음에도 불구하고 인공지능 솔루션의 복잡성이 높아질수록 시스템 규칙이 많아져 결국 결합된 시스템에서 해당 규칙 간의 충돌 가능성이 높아지기 때문에 유사한 모델이 발생할 가능성이 높다. 그러므로, 그것은 인간이 따라갈 수 있는 능력을 뛰어넘는 문제이다.

7. 인공지능 규제 메타프레임워크

일부 프레임워크에서의 보충성은 인공 지능과 로봇 공학의 영향으로 설계, 법률 및 교육의 결합이 필요할 것이라는 인식을 확인시켜 준다. 정치적 및 사회적 맥락에 포함된 다학제적 주제를 다루기에는 프레임워크가 불충분하다는 주장이 있었기 때문에, 검토된 샘플에서 각 모델의 주요 기여도를 포함하는 인공지능 규제 메타 프레임워크가 구축되었다(그림 1).

정부의 독점적인 역량은 입법부, 행정부, 사법부 전반에 걸쳐 분산될 것이다.

법을 만드는 것과는 별개로 입법 기관을 공개하여 해당 법안을 사회, 학계(B)와 논의해 지속적인 피드백(F)을 받을 수 있도록 유지하는 것이 중요하다. 입법부가 법령(J)을 통해 만든 규제기관으로 강하게 대표되는 행정부는 입법부가 입법에 미치는 영향과 그로 인한 진보를 조사하는 지속적인 과정의 일환으로 입법부와의 관계를 구축할 것이다.

규제기관과 마찬가지로 규제기관(T2R)에서 얻은 지식은 입법기관(R2T)이 논의하고 승인한 법률을 토대로 내부 업무 프로세스를 구조화한다. T2R과 R2T 프로세스 간의 이러한 동시성의 품질과 효율성은 입법 및 규제기관이 공유하는 데이터베이스를 통해 강화된다.

규제기관의 역량 중에는 인공지능 시스템의 개발과 학습 과정을 평가하는 모델을 만들고 적용하는 것이 눈에 띈다. 입법부와 마찬가지로 규제 기관의 개방적인 관행 또한 사회와 학계로부터 피드백을 받는 것이 바람직하다(F). 성공적인 평가를 거쳐 규제기관에 제품을 제출한 기업은 활동분야(운송·헬스케어·연예·교육·군 등) 내에서 인증서(C)를 받게 된다. 평가 과정의 엄격함과 성격은 각 분야마다 다를 수 있다. 인증서 발급을 통해 자신이 소비하는 제품과 서비스의 안전 수준과 위험성을 이미 사회에 투명하게 알리고 있기 때문에 법이 통과되기 전에 적용해야 할 전략이 될 수 있다. 정부와 인증 기업의 광고 캠페인도 그 전략을 강화할 것이다. 법원에 의한 법 집행은 시행중인 법률뿐만 아니라 새로운 법률에 따른 해석과 관련하여 지속적인 학습 과정을 거치게 될 것이다. 법령에 인증이 편입된 국가에서는 미인증 기업과 관련된 사건에 대한 결정을 인증 기업과는 다르게 처리한다. 따라서 법원은 각 회사의 인증된 제품 및 서비스에 대한 최신 정보가 필요하다. 지속적인 학습 과정을 고려하면 규제 기관은 인공지능 시스템과 관련된 의사 결정 결과를 받아 입법부와 공유하는 데이터베이스에 저장한다.

업계와 서비스 제공자는 신속한 과정을 통해 가능한 한 명확하게 명시된 규제기관의 인증규정(I)을 받는 한편, 규제기관이 요구하는 개발과정이 앞으로 나아가지 못하게 하는 조건에 대해서는 피드백(F)을 제공해야 할 것이다.

규제기관이 실시하는 감사는 3차원으로 이뤄질 것이다. 먼저, 각국의 세계인권선언과 관련 법률을 포함하는 윤리 원칙에 대한 감사(M. 윤리적 문제나 딜레마가 수반되지 않더라도 이해관계자(P)에게 미치는 영향을 평가하기 위한 감사를 통해 두 번째 차원이 발생한다. 이 분석을 통해 사회가 새롭게 마련한 제도에서 발생하는 향후 문제점을 파악할 수 있었다. 신뢰할 수 있는 인공 지능 시스템의 기본 요소의 실패도 확인될 것이다(투명성, 프라이버시, 인간 복지, 책임성 등). 그리고 마지막으로 인공지능 시스템을 구축할 때 기술적 절차나 대상이 된(L) 학습 과정에 대한 감사가 뒤따랐다.

규제 기관이 기대하는 효율성과 지식은 이를 해결하도록 설계된 윤리적 딜레마에 대한 공식 표현 모델(H), 편향된 학습 프로세스를 식별하는 시스템(Q), 로봇 행동에서 윤리를 평가하는 시스템(V) 및 개발 프로세스 평가 모델(D)에 달려 있다. 법정에서 세계인권선언은 아직 법에 의해 규제되지 않았거나 법적 차원에서 대우받을 필요가 없는 다양한 상황을 해석하는 토대가 될 것이다.

국가 차원에서는 정부기관과 업계 대표, 서비스 제공자, 학자(G) 등이 모이는 인공지능 거버넌스위원회를 통해 우선조치와 산업표준 인정을 촉진하기 위한 논의를 할 수 있다. 합의된 표준(N)은 입법부가 입법에 대한 조정을 논의하는 동안 일부 기술적 차원에서 앞으로 나아갈 수 있도록 한다. 이러한 기준의 사용과 관련된 리스크 관리 기준(O)은 국가 위원회와 업계 간에 협의될 것이다.

인공지능 서비스의 많은 구성 요소 및 글로벌 도달 제품은 각국 위원회(A)의 대표로 구성된 국제 거버넌스 위원회에서 합의할 조치를 부과한다. 많은 경우, 생산 과정의 투명성은 국제 협약을 통해서만 실현 가능하다.

8. 결론

브라질에서, 그리고 전 세계적으로 인공지능을 규제해야 할 필요성과 시급성에 대해서는 반론의 여지가 없어 보인다. 기술의 발달된 특성 때문이든, 그 영향이 구조적으로 사회적 기준에 영향을 미치기 때문이든, 주제의 복잡성 또한 분명하다.

2009~2019년 발표된 51개 논문으로 구성된 표본을 통한 문헌 연구에서 인공지능과 관련된 위험성과 윤리적 딜레마를 파악하고 확대하고, 정부가 모니터링하고 있는 다양한 양식을 통해 인공지능을 규제하는 모델을 모색하려는 상당한 노력이 드러났다.

우리는 지난 58년에 파괴적인 혁신이 어떻게 일어났는지와 같이 법에 대한 인식의 재구성이 일어나고 있다는 사실을 깨달았다. 토론에 참여하는 전문가들의 특성이 전부 다르다는 것은 주제가 연구되고 있는 복잡성과 성숙도를 분명히 보여준다. 이 같은 심도 있는 접근은 한편으로는 일정하게 연구를 지연시켰을 수도 있지만, 다른 한편으로는 부적절한 규제 해법이 공식화되는 것을 막은 측면도 있다.

논의된 프레임워크는 보완적인 접근법을 기반으로 하므로 독립적으로 분석할 경우엔 불충분하다. 메타 프레임워크(그림 1)로서 제안된 통합 접근법은 다양한 지식 영역을 연결하는 여러 에이전트의 존재와 주제의 후기성을 고려할 때 인공지능 거버넌스의 배치에 가장 적절한 전략으로 보인다. 제시된 AIR 프레임워크의 시야를 확장하면 관련 에이전트가 자신의 역할을 식별하는 동시에 점진적이고 중단없는 배치를 위한 로드맵을 수립할 수 있다.

이와 더불어 이 새로운 현실의 균형을 맞추고 있는 그대로의 세상을 고려하는 새로운 보상과 처벌 모델을 만드는 데 기여할 것이다.

AIR 프레임워크의 각 구성 요소를 더 가까이 모으는 것을 넘어 지속 가능한 규제를 위해 참여하는 에이전트 간의 동기화가 필요하다. 그 여정을 위해서는 규제 거시과정에 정부의 3대 주체(집행부, 입법부, 사법부)와 학자들이 연대하는 것이 결정적 역할을 할 것이다.

논의를 주도하는 국가들은 아마도 포괄적이고 효과적인 거버넌스를 위해 필요한 기관들 간의 파트너십과 합의를 마련하고 규제 절차를 시작할 준비가 되어 있을 것이다. 그럼에도 불구하고 첨단 규제 모델을 보유한 국가에서 임베디드 인공지능이 탑재된 제품을 출시하는 것 자체가 이런 점에서 아직 준비되지 않은 국가들에게 동일한 안전 수준을 보장하는 것은 아니다.

제안된 메타 프레임 워크의 개선뿐만 아니라 검토된 샘플에 제시된 프레임워크의 발전에 대한 경험적 분석과 연구가 가능하도록 실제 시나리오에서 프레임워크를 사용한 솔루션을 공식화하는 데에는 아직 지속된 일이 없다. 따라서 성숙 수준이 확립 될 인공지능 거버넌스의 참조 모델을 만들게 되었으며, 이는 공동 노력으로 국제기구에 의해 모니터링 될 수 있다. 우리와 미래 세대가 우리의 삶을 살아가는 방식은 그 협력에 달려 있다.