발 간 등 록 번 호
12-1071400-000016-01

4차산업혁명시대 산업별 인공지능 윤리의 이슈 분석 및 정책적 대응방안 연구

-최종보고서-

수행기관 : 정보통신정책연구원

대통령직속
4차 산업혁명위원회

발 간 등 록 번 호
12-1071400-000016-01

4차산업혁명시대 산업별 인공지능 윤리의 이슈 분석 및 정책적 대응방안 연구

-최종보고서-

수행기관 : 정보통신정책연구원

대통령직속
4차 산업혁명위원회

제 출 문

4차산업혁명위원회 귀하

이 보고서를 『**4차산업혁명시대 산업별 인공지능 윤리의 이슈 분석 및 정책적 대응방안 연구**』 과제의 최종 보고서로 제출합니다.

2018년 10월

주관연구기관명: 정 보 통 신 정 책 연 구 원

주관연구책임자: 이원태 디지털사회정책그룹장

참 여 연 구 원: 김정언 ICT전략연구실장

선지원 부 연 구 위 원

이시직 연 구 원

과 제 자 문 위 원: 박혜경 교 수(충남대학교)

안수현 교 수(한국외대)

정채연 교 수(포항공대)

최은창 펠 로 우

(프리인터넷 프로젝트)

한희원 교 수(동국대)

발주기관 검수: 4차산업혁명위원회 사회혁신팀

태병민 팀장, 지병석 사무관

목 차

제 1 장 서 론 ····· 9
제 1 절 연구의 필요성 및 목적 ····· 9
제 2 절 연구의 내용과 범위 ····· 12
제 2 장 인공지능 산업동향 및 산업 · 사회 변화상 ····· 15
제 1 절 인공지능 관련 기술 및 산업동향 ····· 15
1. 전 세계 인공지능 시장 규모 ····· 15
2. 국내 인공지능 시장 규모 ····· 16
제 2 절 인공지능 적용산업 현황 및 활용사례 분석 ····· 19
1. 인공지능 활용 산업 분야 분류 및 적용산업 현황 ····· 19
2. 국내외 인공지능 적용산업 포지셔닝 ····· 24
3. 시사점 ····· 26
제 3 절 4차 산업혁명의 산업(경제) · 사회변화상 분석 ····· 28
1. 산업 구조의 변화 ····· 28
2. 고용 구조의 변화 ····· 32
3. 직업 구조의 변화 ····· 37
4. 데이터 기반 사회의 도래 ····· 39
5. 의사결정 시스템의 보편화에 따른 삶의 질 향상과 사회적 역기능 ····· 41
6. 인간－로봇 공존을 위한 규범(윤리 · 법) 환경의 변화 ····· 43
제 3 장 국내 · 외 인공지능 윤리관련 정책 및 규범 동향 분석 ····· 48
제 1 절 논의의 배경 ····· 48
제 2 절 학계 및 민간연구기관차원에서의 인공지능 윤리 논의 동향 ····· 50
1. 하버드 버크만센터－MIT 미디어랩 ····· 50
2. 옥스퍼드대 미래의 삶 연구소(Future of Life Institute) ····· 52
3. 프린스톤 인공지능과 윤리를 위한 대화(Princeton Dialogues on AI and Ethics) ···· 55

4. 캠브리지대 실존적 위기 센터(The Centre for the Study of Existential Risk) ····· 56
제 3 절 국가별 정부 및 의회 차원에서의 인공지능 윤리관련 논의 동향 ····· 58
1. 미국 ····· 60
2. 유럽연합 ····· 63
3. 영국 ····· 73
4. 프랑스 ····· 77
5. 중국 ····· 80
제 4 절 기업차원에서의 인공지능 윤리 논의 동향 ····· 86
1. 인공지능 파트너십(Partnership on AI) ····· 86
2. 마이크로소프트 에테르(Aether) ····· 87
3. 딥마인드 윤리 사회(DeepMind Ethics & Society) ····· 89
4. 구글(Google) ····· 90
5. 카카오의 알고리즘 윤리헌장 ····· 93
제 5 절 국제기구(글로벌)차원에서의 인공지능 윤리 논의 동향 ····· 94
1. 경제협력개발기구(OECD) ····· 94
2. 유엔 살상로봇 금지 캠페인(Campaign to Stop Killer Robots) ····· 96
3. 인터넷 소사이어티(ISOC) ····· 98
4. 국제전기전자학회(IEEE) 글로벌 이니셔티브 ····· 100
5. 전미 컴퓨터학회(ACM) ····· 102
제 6 절 소결 및 정책적 시사점 ····· 106
1. 책임 있는 인공지능을 위한 인공지능 거버넌스 ····· 106
2. 윤리・규제정책의 필요성: 주인－대리인’(principal-agent) 문제 ····· 106
3. 알고리즘 문해력의 필요성 ····· 108
4. 개인정보 보호와 양립가능한 데이터 활용 ····· 109
5. 연성규범 제시 필요성 ····· 110
6. 인공지능 윤리 가이드라인의 제정 주체 ····· 111

제 4 장 산업 분야별 인공지능 윤리 이슈 분석 ····· 112
제 1 절 인공지능 윤리 이슈의 분석 원리 및 검토대상 분야의 선정 ····· 112

1. 기술의 생애단계와 행위자연결망(Actor-Network) ····· 112
2. 인공지능 윤리이슈 검토대상 분야의 선정: 제조, 의료, 금융, 국방 ····· 114
제 2 절 제조 분야 인공지능 윤리이슈: 자율주행자동차를 중심으로 ····· 114
1. 논의의 배경 ····· 114
2. 자율주행자동차의 윤리 이슈 및 쟁점 ····· 119
제 3 절 금융 분야: 로보어드바이저를 중심으로 ····· 136
1. 논의의 배경 ····· 136
2. 본 연구에 적용하는 '금융윤리'의 개념과 특징 ····· 137
3. 로보어드바이저 산업과 국제적 규율 현황 ····· 140
4. 로보어드바이저 산업·시장 참여자 및 각 참여자별 윤리 이슈 ····· 153
제 4 절 의료 분야 인공지능 윤리 이슈 ····· 155
1. 논의의 배경 ····· 155
2. 의료 분야 인공지능 윤리 이슈 및 쟁점 ····· 158
제 5 절 국방 분야: 치명적 자율무기체계를 중심으로 ····· 168
1. 논의의 배경 ····· 168
2. 국방 분야 인공지능 윤리 이슈 및 쟁점 ····· 175

제 5 장 인공지능윤리 이슈 정책적 대응방안: AI 윤리 대응 프레임워크 ····· 185
제 1 절 자율 윤리 대응 프레임워크 구축의 필요성 ····· 185
제 2 절 자율 윤리 대응 프레임워크 구축 시의 고려 사항 ····· 187
1. 개관 ····· 187
2. 법적 규제와의 경계 설정 ····· 187
3. 민관학 협력 모델 ····· 192
4. 서비스의 이용자에 대한 고려 ····· 194
5. 인공지능 기술 발전 속도 고려 ····· 196
6. 사전 예방과 사후 대응의 조화 ····· 197
7. 인공지능 기술의 생애주기에 따른 윤리 이슈 고려 ····· 197
8. 프레임워크의 범용성 ····· 198
제 3 절 종합적 자율 윤리 대응 프레임워크의 구축 ····· 199

1. 원칙 ····· 199
2. 행위자별 자율 윤리 대응의 구조 ····· 200
3. 쟁점별 자율 윤리 대응의 구조 ····· 202
제 4 절 개별 산업 영역에의 적용 ····· 203
1. 개관 ····· 203
2. 제조 분야 ····· 204
3. 금융 분야 ····· 208
4. 의료 분야 ····· 216
5. 군사 분야 ····· 225
제 6 장 결론 및 정책 제언 ····· 227
제 1 절 본 연구의 결론적 논의 ····· 227
제 2 절 정책 제언 ····· 228
참고문헌 ····· 231

표 목 차

〈표 2－1〉 전 세계 인공지능 시장 규모 예측 ····· 15
〈표 2－2〉 세부 인공지능 기술 수준 및 격차(미국=100) ····· 18
〈표 2－3〉 인공지능 활용 산업 분야 분류체계 ····· 20
〈표 2－4〉 도출된 17개 인공지능 적용산업 분야 분류 및 적용예시 ····· 21
〈표 2－5〉 해외 인공지능 제품/서비스별 적용 산업 집중도 ····· 24
〈표 2－6〉 국내 인공지능 제품/서비스별 적용 산업 집중도 ····· 25
〈표 2－7〉 집중도 Mattix 레벨 설정 기준 ····· 26
〈표 2－8〉 스마트 기술에 의한 제조·서비스 분야 고용 대체가능성 ····· 35
〈표 2－9〉 지능정보기술로 인한 일자리 창출 효과 추정 ····· 36
〈표 2－10〉 산업 분야별 요구 직무역량 변화 전망(2015~2020) ····· 38
〈표 3－1〉 인공지능 윤리와 거버넌스의 세부 프로젝트 ····· 51
〈표 3－2〉 ‘유럽을 위한 인공지능’에 포함된 윤리적, 법적 체계의 요약 ····· 70
〈표 3－3〉 영국 상원 보고서의 인공지능 규범(AI Code) ····· 74
〈표 3－4〉 카카오 알고리즘 윤리 헌장 ····· 94
〈표 4－1〉 윤리적 딜레마 상황에서의 가치판단 기준 ····· 128
〈표 4－2〉 저티스와 손튼의 3법칙과 로하스의 4법칙 ····· 134
〈표 4－3〉 각 단계 행위자별 윤리이슈 ····· 135
〈표 4－4〉 인공지능의 시기별 분류 ····· 139
〈표 4－5〉 로보어드바이저의 유형 ····· 140
〈표 4－6〉 각 단계 행위자별 윤리이슈 ····· 153
〈표 4－7〉 각 단계 행위자별 윤리이슈 ····· 168
〈표 4－8〉 각 단계 행위자별 윤리이슈 ····· 184
〈표 5－1〉 최근 발간된 인공지능 관련 윤리 가이드라인 ····· 191

그 림 목 차

〔그림 1－1〕 인공지능 윤리담론의 변화과정 ····· 11
〔그림 1－2〕 인공지능 알고리즘의 윤리적 개입 구조 ····· 12
〔그림 1－3〕 연구목표 및 주요내용 ····· 14
〔그림 2－1〕 세계 인공지능 시장 규모 전망 ····· 16
〔그림 2－2〕 국내 인공지능 시장 규모 ····· 17
〔그림 2－3〕 인공지능 기술 수준 ····· 18
〔그림 2－4〕 인공지능 기술 격차 ····· 18
〔그림 2－5〕 지능정보사회 중장기 종합대책 ····· 19
〔그림 2－6〕 인공지능 R&D 전략 ····· 19
〔그림 2－7〕 적용산업별 AI 잠재적 경제효과 및 영향 ····· 22
〔그림 2－8〕 산업별 인공지능 유스케이스 분석 프로세스 ····· 23
〔그림 2－9〕 국내·외 인공지능 제품/서비스 비중 ····· 23
〔그림 2－10〕 해외 인공지능 제품·서비스 기술 적용도 Matrix 분석 결과 ····· 27
〔그림 2－11〕 국내 인공지능 제품·서비스 기술 적용도 Matrix 분석 결과 ····· 27
〔그림 2－12〕 제4차 산업혁명의 주요 변화 동인 ····· 29
〔그림 2－13〕 2006년과 2016년 시가총액 10대 기업 비교 ····· 29
〔그림 2－14〕 미국 내 직종의 대체확률별 고용인원 ····· 33
〔그림 2－15〕 미래 산업 분야에서 요구하는 직무역량 ····· 39
〔그림 3－1〕 미 백악관 주재 인공지능 워크숍의 주요 내용 ····· 61
〔그림 3－2〕 뉴욕시 회의의 알고리즘 소스코드 공개 의무화 법안 ····· 62
〔그림 3－3〕 ‘유럽을 위한 인공지능’ 정책의 3대 분야. ····· 66
〔그림 4－1〕 로보어드바이저 투자자문 및 자산관리서비스 절차 ····· 143
〔그림 5－1〕 윤리와 법의 경계 ····· 189
〔그림 5－2〕 윤리 규범과 법 규범의 대상 관계도 ····· 190

〔그림 5－3〕 국내 자율규제기구의 사례 ····· 192
〔그림 5－4〕 독일 공공사업 계획 및 추진단계에서의 주민 참여 모델 ····· 195
〔그림 5－5〕 독일 Industrie 4.0 표현도 ····· 196
〔그림 5－6〕 인공지능 기술 수용 과정도 ····· 198
〔그림 5－7〕 편익－리스크 비교 그래프(예시) ····· 201
〔그림 5－8〕 행위자별 자율 윤리 대응 프레임워크 ····· 202
〔그림 5－9〕 쟁점별 자율 윤리 대응 프레임워크 ····· 203
〔그림 5－10〕 제조 분야 윤리 대응 프레임워크 ····· 208
〔그림 5－11〕 금융 분야 윤리 대응 프레임워크 ····· 215
〔그림 5－12〕 의료 분야 윤리 대응 프레임워크 ····· 225
〔그림 5－13〕 국방 분야 윤리 대응 프레임워크 ····· 226

제1장 서 론

제1절 연구의 필요성 및 목적

2016년 '알파고 충격' 이후 4차 산업혁명은 우리 시대의 화두가 되었고, 4차산업혁명을 추동하는 핵심기술로서의 인공지능은 사회경제적 변화는 물론, 오랫동안 우리의 의식과 행동을 규정해온 규범 체계마저 근본적으로 재검토할 것을 요구하고 있다. 4차산업혁명이 가져온 변화에 대한 정책 대응의 양상 또한 초기에는 기술개발, 산업진흥에 초점을 둔 투자 및 연구개발이 강조되었다가, 최근에 와서는 점차 사회문화적 역기능이나 윤리적 규범 이슈에 대한 정책 대응이 중요하게 부각되었다. 물론 4차산업혁명의 정책 대응과정에서 윤리적, 규범적 대응의 중요성이 강조된 데에는 인공지능 기술이 다양한 산업 분야 및 서비스 영역에서 적용되는 과정에서 발생가능한 윤리적 이슈들이 학계는 물론, 언론, 시민사회에서 끊임없이 제기되었기 때문이다.[1] 그리고 최근 인공지능 스피커의 목소리가 인간과 구별되지 않아서 발생가능한 윤리적 문제, 인공지능에 의한 면접 채용의 공정성 논란에서 보듯이, 점차 인공지능 기술의 서비스 채택 범위가 넓어질수록 인공지능 기반의 의사결정이나 선택의 공정성, 합리성에 대해 지속적으로 문제 제기하는 일이 많아지고 이에 대한 윤리적 규제 논의도 수반될 것이다.

그런데 문제는 지금까지 인공지능의 윤리적 이슈를 둘러싼 논의가 몇가지 이유로 인해 제대로 방향을 잡지 못하고 있다는 점이다. 그것은 첫째, 인공지능의 산업적·사회적 활용이 증가함에 따라 산업 활성화를 위한 입법 논의에 치중하면서도, 인공지능과 관련한 사고발생 및 오작동의 위험성에 대비한 법적인 책임의 배분이나 권리의 범위 설정에 초점을 둔 법제도적 논의에 비해 인공지능의 윤리적 설계를 포함한 윤리적 논의는 상대적으로 소홀하다는 점이다. 일부 과학자 등 전문가 중심으로 인공지능의 윤리적 개발 원칙에 관한 논의가 제기

1) 지능정보사회 규범 또는 인공지능의 윤리적 기준을 다룬 최근 학계의 주요 연구로는 이원태(2016), 이원태 외(2017), 심우민(2016; 2017), 권헌영 외(2017), 한국포스트휴먼연구소(2017), 토비 월시(2018) 등이 대표적이다.

되고는 있지만, 인공지능 관련 기업들의 자발적 참여 보다는 정부, 의회 중심의 다양한 법제화 시도가 더 많다는 것이다.[2] 더구나 시민사회 및 시민 개개인의 차원에서 윤리적 논의는 이제 막 시작단계에 머물러 있어 인공지능이 인간과 사회에 미치는 파급효과에 대한 윤리적 평가와 개발 원칙에 대한 논의는 좀더 활성화될 필요가 있다.

둘째, 인공지능의 기술적 발전 속도가 워낙 빠르고 그 발전 방향 또한 예측할 수 없기 때문에 법이든, 윤리든 규범적 프레임워크를 섣불리 설정하기가 쉽지 않은 데다가 입법적 대응과 윤리적 대응 사이의 우선순위가 나라마다, 기술이 적용되는 산업 분야에 따라 상대적일 수밖에 없다는 것이다. 더구나 인공지능 기술의 특성이 방대한 데이터, 복잡한 알고리즘으로 인해 통제와 예측이 쉽지 않을 뿐만 아니라 알고리즘의 의도하지 않은 차별 및 편견 개입에 따른 피해 원인을 특정하기도 어려운 것이 현실이다. 따라서 다양한 영역에서 인공지능에 관한 윤리적 논의를 보다 풍부하게 제기해 신기술 혁신과 규범 정립 사이의 간극에 따른 갈등을 최소화하도록 일종의 윤리적 완충 지대가 필요할 것으로 보인다. 다시 말해 인공지능 알고리즘 기술의 일반화에 대한 사회적 저항 및 혼란을 최소화하기 위해서라도 인공지능 윤리 기준에 대한 공론화가 필요하다는 것이다. 일사불란한 법체계 수립 보다는 기술, 산업, 사회의 복합적 변화에 대비한 다양한 윤리적 논의가 더 필요하다는 것이다. 이처럼 4차 산업혁명 시대의 본격화에 대비한 인공지능 윤리 규범의 소환은 기술의 자율성이 기술 및 사물에 대한 인간의 자율성을 훨씬 능가하는 상황에서 인간의 정체성 위기와 무기력을 극복하고 최소한의 윤리적 존재 가치를 회복하려는 움직임이라고 할 수 있다(이원태, 2018).

셋째, 인공지능 윤리 이슈에 대한 선행적・사전적 논의의 필요성에도 불구하고 윤리 논의의 과잉 자체가 심리적 규제효과로 작용하여 인공지능 개발 및 산업 분야의 이해당사자들이 인공지능 윤리 문제에 자발적, 적극적인 관심과 참여를 위축시킨다는 점이다. 이 때문에 정부, 의회가 인공지능 윤리 규제 담론을 주도하는 상황을 초래하게 되고 인공지능 윤리 문제의 제기가 잠재적인 규제 갈등 이슈로 전환될 가능성도 내포한다는 것이다.

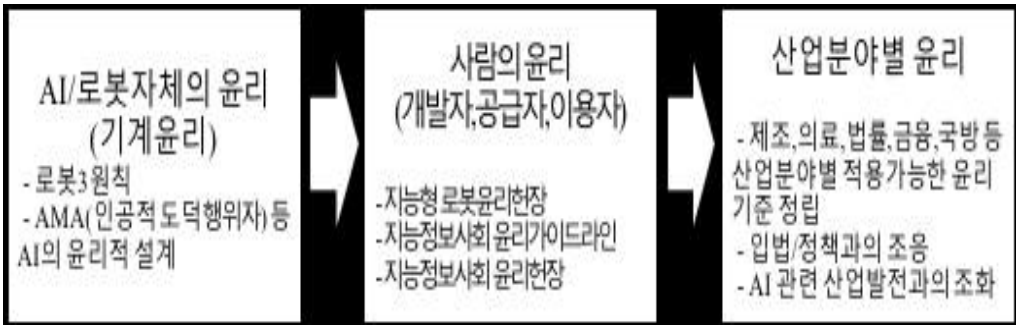
넷째, 인공지능의 잠재적, 현재적 리스크에 대한 막연한 두려움과 예측에 기반해 인공지능의 윤리적 이슈가 인간 및 사회에 대한 일반적, 총론적 규범 논의에 그치거나 행위자 중심의 도덕 윤리론 중심으로 논의된다는 점이다. 기존의 윤리학적 담론 틀에 갇혀 논의되거나 지

2) 법제도적 접근에는 기본법 등 일반적 입법대응, 관련법 개정 등 개별적 입법대응, '가이드라인' 등 법령 개정 없는 부처 소관권한 대응으로 구분 가능하다(김태오, 2017).

나치게 먼 미래의 시점에서 추상적, 거시적 차원에서만 논의되는 경우가 많다는 것이다. 이 때문에 국내외적으로 인공지능의 윤리적 이슈에 대한 논의가 다양하게 진행되고 있으나, 인공지능 기술이 활용되고 있는 구체적 산업 분야 및 개별 적용 서비스에서 발생되는 윤리 이슈 검토 및 체계적 대응 논의는 상대적으로 미흡한 실정이다.

이상의 최근 인공지능 윤리 이슈를 둘러싼 논의 및 대응방식의 한계점을 고려할 때 인공지능 윤리 논의의 방향 전환이 요구된다. 물론 인공지능 기술 및 서비스의 일반화에 따른 사회적 저항 및 혼란을 최소화하기 위한 공통의 윤리 기준 정립도 필요하겠지만, 이제는 구체적인 적용 분야별의 특성을 고려한 윤리적 행위 원칙에 관한 논의 틀 마련도 고민하기 위해 다양한 이해당사자들의 참여와 소통이 요구된다는 것이다.

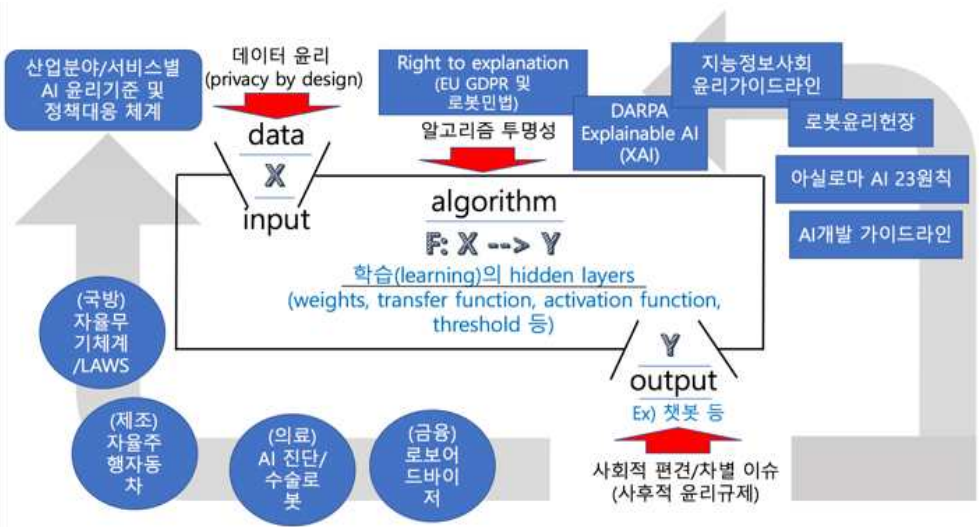
〔그림 1-1〕 인공지능 윤리담론의 변화과정



이에 본 연구는 인공지능의 기술 및 서비스, 산업별로 윤리적 이슈를 조사 정리하여, 인공지능 윤리 이슈에 대해 체계적·구체적으로 분석하며, 더 나아가 현재 인공지능 기술 개발 및 상용화에 따른 주요 산업 분야의 구체적이고 개별적인 윤리 이슈 분석을 통해 적용 가능하고 실제적인 정책 대응방안을 아울러 제시하고자 한다. 그래서 본 연구는 인공지능이 폭넓게 활용되고 윤리 이슈가 쟁점화되는 산업 분야를 제조, 의료, 금융, 국방 등 4개 영역으로 설정하고 각각 자율주행자동차, AI의료서비스, 로보어드바이저, 치명적 자율무기체계(LAWS: Lethal Autonomous Weapons System)의 윤리이슈 분석 및 정책적 대응방안을 도출하고자 한다. 더 나아가, 본 연구는 오늘날 인공지능의 윤리 이슈는 데이터 윤리 문제에서부터 점차 알고리즘 개발 단계에서의 윤리적 규제 이슈, 그리고 산업 분야별 서비스 적용 및 활용 차원의 윤리적 규제 이슈로 진화, 복잡하게 발전하는 것이기 때문에 인공지능 윤리이슈의 정책

적 대응 프레임워크를 모델링하고 이의 적용 방안도 모색하고자 한다.[3)]

〔그림 1－2〕 인공지능 알고리즘의 윤리적 개입 구조



제 2 절 연구의 내용과 범위

본 연구는 서론(1장)을 비롯해서 인공지능 기술·산업 현황 및 사회변화상 분석(2장), 국내외의 인공지능 윤리 연구 및 정책동향 분석(3장), 주요 산업 분야별 인공지능 윤리이슈 분석(4장), 인공지능 윤리이슈에 대한 정책대응 방안 도출(5장), 결론 및 정책제언(6장) 등 모두 6개의 장으로 구성되어 있다.

먼저 제2장은 인공지능 기술·산업동향 및 사회경제적 변화상을 다각적으로 검토하였다. 글로벌 시장조사기관 보고서 등을 통해 국내·외 인공지능 시장규모를 살펴보고, 인공지능 적용산업 현황 및 활용 사례를 분석하기 위해 정책적 중요도에 따라 17개 산업 분야로 세분화하여 국내외 기업(264개)의 345개 인공지능 제품 및 서비스의 유스케이스 분석을 통해 산업집중도를 도출하였다. 더 나아가 지능정보기술 확산에 따른 수확체증 가시화, 제조업의 서비스와, 지식 중심사회, 고용 및 직업 변화, 사회적 역기능 초래 등 미래 사회·경제 구조

3) 본 연구에서는 선정된 산업 분야를 중심으로 설계－개발－시험/임상－상용화－이용 등 인공지능 기술의 구체적인 적용 및 활용단계에서 발생하는 행위주체별 윤리이슈를 중심으로 분석하고 더 나아가 이에 기반하여 윤리적 정책 대응 프레임워크도 구축하고자 하였다.

변화상을 구체적으로 제시하였다.

3장에서는 국내외의 다양한 영역에서 논의되는 인공지능 윤리 관련 연구 및 주요국의 정책 대응 현황을 국가별 정부 및 의회, 학계 및 민간연구기관, 국내외 IT기업, 국제기구 등으로 세분화해서 분석하였다. 특히 개발자, 과학자 등 전문가들이 인공지능 윤리 이슈에 대해 집단적 문제 제기를 하는 것과는 달리, 기업이나 사업자들이 뒤늦게 인공지능 윤리 개발원칙을 제시하는 등 소극적으로 대응하는 등 영역별 대응의 특성을 비교 검토하였다.

4장에서는 제조, 금융, 의료, 국방 등 2장에서 도출한 주요 산업 분야별로 인공지능 인공지능 윤리이슈를 조사, 분석하였다. 선정된 4개 산업 분야를 중심으로 설계-개발-시험/임상-상용화-이용 등 인공지능 기술의 구체적인 적용 및 활용 단계에서 발생하는 행위주체별 윤리이슈를 도출하고, 이의 내용을 분석하였다.

5장에서는 제조업, 금융, 의료산업, 국방 등 산업 분야별 인공지능 윤리 관련 이슈의 해결을 모색하기 위해 전문가 간담회, 세미나 등의 의견수렴 과정을 통해 인공지능 윤리이슈 대응을 위한 정책 설계방향 및 대응방안을 제시하였다. 앞에서 살펴본 산업 분야별 인공지능 윤리 이슈에 적절히 대응하고 각종 인공지능 윤리 규범들을 산업영역에서 관철시키기 위해 그와 관련한 정책 대응의 구조 또는 프레임워크를 마련한다는 것이다. 본 연구에서는 윤리적 정책 대응 프레임워크를 구축하기 위해 일반적 윤리 거버넌스의 특징, 인공지능 기술의 특성, 인공지능 기술의 사회적 수용성 및 그에 따른 사회현안 등을 염두에 두었으며, 특히 인공지능 기술의 생애주기 특성을 반영해 발전단계별로 각각의 행위주체별 또는 윤리 관련 쟁점별로 대응구조를 종합적으로 제시하였다.

6장에서는 본 연구의 전체적인 결론과 함께 정책적 제언을 제시하였다. 여기서는 4차산업혁명시대의 신기술 윤리 관련 정책 연구대상 및 범위를 인공지능 외에도 빅데이터, 클라우드 컴퓨팅, 사이버 보안 등 여타의 기술 영역으로 확대하고, 지속적인 윤리 대응을 위해 4차산업혁명위원회 내에 ‘신기술 윤리분과’를 상설화하는 방안, 본 연구에서 마련한 인공지능 윤리이슈 대응정책 프레임워크 및 적용 기준을 각 산업 분야에 적용하기 위한 공론화 과정(예시: ‘AI윤리 해커톤’)을 4차산업혁명위원회가 주도하기 위한 방안 등을 제시하였다.

〔그림 1-3〕 연구목표 및 주요내용

제 2 장 인공지능 산업동향 및 산업·사회 변화상

제 1 절 인공지능 관련 기술 및 산업동향

1. 전 세계 인공지능 시장 규모

전 세계 인공지능 시장규모 전망은 각 조사기관마다 적용하는 인공지능의 범위 및 분류 기준이 크게 달라 관련 시장의 절대적인 성장 잠재력을 예측한 결과에도 큰 편차를 보이고 있다. IDC(2018. 3)에 따르면, 2018년 전 세계 인식·인공 지능 시장은 전년대비 약 54.2% 증가한 191억 달러규모에 이르고, 2016년부터 연평균 46.2%씩 성장하여 2021년에는 522억 달러에 기록할 것으로 전망하였다. Statista(2018. 6)도 올해 세계 인공지능 시장 규모가 2017년(48억2000만 달러) 대비 52.2% 성장한 73억5000만 달러 규모에 달하고, 연평균 약 51%씩 성장하여 2025년에는 898억5000만 달러 규모로 성장할 것으로 예측하였다. 또한 Gartner(2018. 4)는 올해 글로벌 인공지능 비즈니스 가치가 전년(6,920억 달러)대비 70% 성장한 1조2000억 달러 규모를 달성하고, 매년 두 자릿수 성장률을 기록하여 2020년에는 올해보다 3배 증가한 약 3조9000억 달러 규모에 이를 것으로 예상하고, 전 세계 인공지능 시장은 2017년 125억 달러 수준에서 2022년 1,132억 달러에 달할 것으로 전망하였다.

〈표 2－1〉 전 세계 인공지능 시장 규모 예측

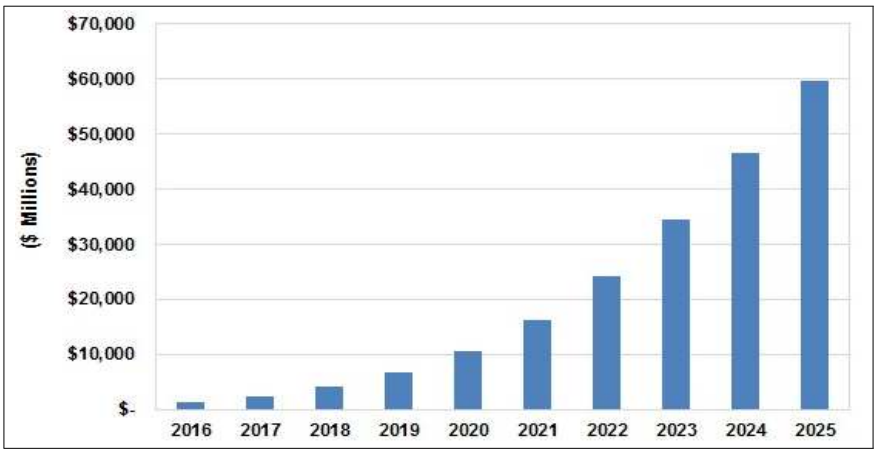
	2017	2018	2019	2020	2021	2022
비즈니스 가치	692	1,175	1,901	2,649	3,346	3,923
성장률(%)	-	70	62	39	26	17

자료: Gartner(2018. 4).

이 밖에도 Tractica(2017)에 따르면, 전 세계 인공지능 시장은 14억 달러('16. 기준) 규모에서 연평균 52% 성장하여 2025년에 598억 달러 규모를 기록할 것으로 예측하였다.

〔그림 2－1〕 세계 인공지능 시장 규모 전망

(단위: 백만 달러)



자료: Tractica(2017).

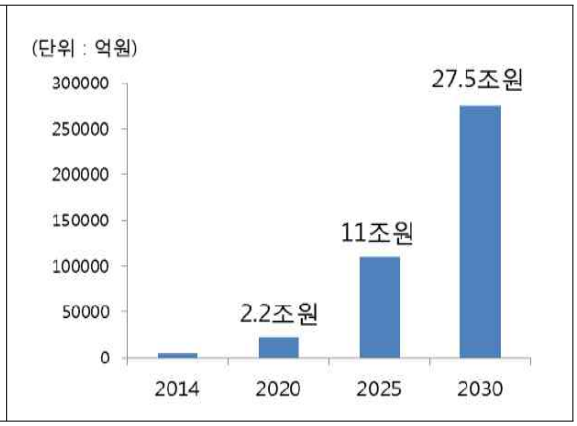
한편, 주요 조사기관마다 인공지능 시장규모 전망이 상이하지만 대부분 연평균 50%를 상회하는 높은 성장률을 예측하고 있어, 도입 단계의 인공지능 기술이 향후 상용화 단계로 발전하면서 폭발적인 성장세와 타산업과의 지능형 융합을 견인할 것으로 예상된다.

2. 국내 인공지능 시장 규모

국내에서는 인공지능과 관련한 정확한 기준이나 데이터가 부재하여 관련 시장 규모를 산출하기 쉽지 않다. 다만, 과학기술정보통신부(2016)는 SW 및 알고리즘, 영상처리 및 영상인식, 음성인식 및 통번역 등 3개 부문으로 분류한 국내 인공지능 산업이 2016년 5조 4,000억 원에서 2020년에는 11조 1,000억 원으로 약 2배 이상 성장할 것으로 예측한 바 있다. 또한 KT경제경영연구소(2016)도 국내 인공지능 시장을 로봇 산업 수치에 기초하여 향후 다양한 산업군으로의 적용을 가정하여 산출한 결과, 2020년에는 2조 2,000억 원, 2025년에는 11조 원, 2030년에는 약 27～30조 원의 규모로 확대될 것으로 전망하였다.

〔그림 2-2〕 국내 인공지능 시장 규모

자료: 과학기술정보통신부(2016).



자료: KT경제경영연구소(2016).

미국, 일본, 중국 등 4차 산업혁명 선도국가들은 인공지능의 높은 잠재력에 주목하고, 각각 '국가 AI R&D 전략계획('16. 10)', 'AI 산업화 로드맵('17. 4)', '차세대 AI 발전규획('17. 7)'을 발표하는 등 인공지능 기술력을 선점하기 위한 국가 차원의 대책을 추진하고 있다.

또한 글로벌 기업들도 IT 기업을 중심으로 인공지능 원천 기술 확보를 위한 기술 투자를 확대하고, 인공지능 스타트업의 인수합병 및 제휴가 활발해지며 인공지능 플랫폼을 선제적으로 구축하기 위한 경쟁이 치열해지고 있다. 그러나 국내 AI 산업은 아직까지 초기 단계에 머물러 있으며 주요 선진국 대비 기술 정도나 시장 규모가 미흡한 수준이다(한국무역협회, 2018).

정보통신기술진흥센터(2017)에 따르면, 한국의 인공지능 분야 기술 수준은 미국(100 기준)의 73.9로, 미국과의 기술격차는 유럽 1.1년, 일본 1.5년, 한국 2.2년, 중국 2.3년으로 나타났다. 특히, 중국의 인공지능 기술은 미국의 71.8로 2015년(66.1%) 대비 0.5년의 기술격차를 단축하면서 빠르게 한국을 추격하고 있다.

〔그림 2－3〕 인공지능 기술 수준

주: 2016년 기준(미국=100).
자료: IITP(2017).

〔그림 2－4〕 인공지능 기술 격차

주: 미국과의 격차
자료: IITP(2017).

세부적으로는 기초연구(73.6), 응용 개발(74.4), 사업화(73.4) 중 사업화 부문에서 미국과 가장 큰 기술격차가 있었으며, 일본은 전체 점수가 81.9로 기초연구, 응용개발, 사업화 등 각 부문에서 모두 한국을 앞섰다.

〈표 2－2〉 세부 인공지능 기술 수준 및 격차(미국=100)

	유럽	일본	한국	중국
기초연구	86.5	82.0	73.6	71.3
응용 및 개발	86.5	82.1	74.4	72.2
사업화	85.7	81.5	73.4	71.7
상대 기술수준	86.3	81.9	73.9	71.8
상대 기술격차	1.1	1.5	2.2	2.3

자료: IITP(2017).

이처럼 우리나라는 주요 선도국에 비해 인공지능 기술력 및 격차가 점차 심화되고 있으며, 장기적인 연구·개발 체계 구축과 전문 인력 양성 등에서도 여전히 낮은 수준에 머물고 있는 상황이다.

이에 최근 우리 정부도 역동적인 기술혁신 생태계를 구축하기 위해 지능정보기술의 발전에 주목하고 시급한 대비가 필요한 지능정보기술 및 산업발전 대책을 담은 ‘지능정보산업 발전방안(’16. 3)’, 지능정보사회에 대비하여 기술·산업·사회 분야 정책을 아우르는 ‘지능정보사회 중장기 종합대책(’16. 12)’, ‘4차 산업혁명 대응계획(’17. 11)’, 세계적 수준의 인공

지능 기술력 및 R&D 생태계 확보를 위한 '인공지능(AI) R&D 전략('18. 5)' 등 범국가적 대응전략을 마련하여 역량을 집중하고 있다.

〔그림 2-5〕 지능정보사회 중장기 종합대책

자료: 관계부처합동(2017).

〔그림 2-6〕 인공지능 R&D 전략

자료: 과학기술정보통신부(2018. 5).

제 2 절 인공지능 적용산업 현황 및 활용사례 분석

1. 인공지능 활용 산업 분야 분류 및 적용산업 현황

본 절에서는 인공지능 적용산업 현황을 파악하기 위해서 우선 국내에 소개된 각종 문헌조사를 통해 인공지능 활용 산업 분야를 분류하였다. 먼저, 과학기술정보통신부(舊미래창조과학부, 2016)는 "지능정보사회 중장기 종합대책"을 통해 12개의 지능정보화 촉진 대상 산업을 기반으로 인공지능 적용산업을 분류하였고, 맥킨지(2016)는 "16년 지능정보기술을 통한 사회경제 혁신방안 연구"에서 5개 산업 분야와 4개 라이프분야 총9개를 제시하였다. 또한 NIPA(2017)는 기존 과학기술정보통신부(2016)에서의 지능정보화 촉진 대상 산업을 기반으로 지능정보기술이 적용되는 산업 분야별 경제효과 분석을 위해 15개 산업 분야로 분류하

는 한편, 과학기술정보통신부(2018)는 미래사회 파급효과가 높은 산업 분야의 집중 육성을 위해 마련된 “인공지능(AI) R&D전략”을 통해 10개 산업 분야를 발표하였다.

이에 본 연구에서는 위 4가지 문헌에서의 인공지능 활용 산업 분야 분류체계와 정책적 중요도에 따라 17개 산업 분야로 세분화하여 현재 서비스 중인 지능정보기술 적용 유스케이스를 다각적으로 분석하였다.

〈표 2－3〉 인공지능 활용 산업 분야 분류체계

분류 \ 연번	1		2		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
맥킨지 (2016)	의료		제조			금융	유통	교통	도시	주거	웰니스	기타							
舊미래부 (2016)	의료	바이오	제조	에너지	로봇	금융	유통	교통	－		복지	－	국방	행정			－	－	보안
NIPA (2017)	의료		제조		에너지	금융	유통	교통	도시	생활	환경	농업	국방	행정사무			교육	문화관광	안전
과기정통부 (2018)	의료		제조		에너지	금융	－	이동체	도시	－	복지	농수산업	국방	－			－	－	안전
본 연구의 분류체계	의료		제조		에너지	금융	유통	교통	도시	홈	웰니스	농업	국방	법률	사무관리	공공분야	교육	문화관광	안전

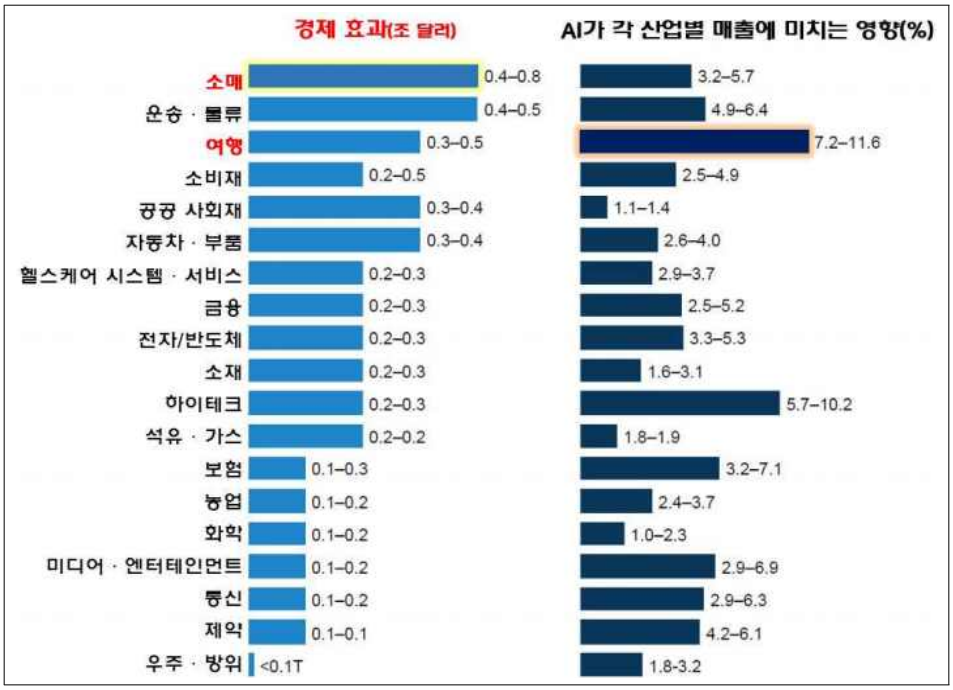
자료: 맥킨지(2016), “16년 지능정보기술을 통한 사회경제 혁신방안 연구”, 과학기술정보통신부(2016), “지능정보사회 중장기 종합대책”, NIPA(2017), “지능정보사회 구현을 위한 심층통계 경제효과 및 세부 이행전략 분석연구”, 과학기술정보통신부(2018), “인공지능(AI) R&D전략”.

〈표 2－4〉 도출된 17개 인공지능 적용산업 분야 분류 및 적용예시

순번	산업 분야	산업별 인공적용 적용 예시
1	의료	진단의 정확도 증가, 약물의 오남용방지 등 맞춤형 정밀의료 서비스 제공
2	제조	3D 프린팅 등으로 생산기간 단축 및 수요 맞춤형 다품종 소량생산
3	에너지	에너지의 소비 패턴을 통한 비용절감과 필요한 에너지는 자가 생산 및 충족하고, 잉여 에너지는 판매하는 'E(Energy)－프로슈머' 등장
4	금융	사람의 개입 없는 로보어드바이저 기반 개인자산 분석, 금융 사기 탐지 및 대응
5	유통	어디서든 사용자가 원하는 물건을 구입하는 무노력 쇼핑(Zero-Effort Shopping) 실현
6	교통	사고 없이 안전하게 운행하는 무인버스, 무인택시의 상용화
7	도시	교통 정체구간 분석을 통하여 실시간 교통관제로 운전시간 및 연료, 이산화탄소 배출비용 감소
8	홈	가사 자동화 시스템으로 인한 가사노동 시간 단축과 스마트 홈 시스템을 기반으로 한 에너지 관리 효율화 및 보안 강화
9	웰니스	생활운동 유인 및 건강상태 모니터링을 통한 건강증진 효과 및 대기·수질오염 측정 모니터링 시스템을 통한 실시간 대응으로 관련 질병 예방과 사망자수 감소
10	농업	자율주행 장비, 온습도 센서, 양액 재배, 관리시스템 등 농업의 생산성 향상 및 노동력 저감을 위한 지능형 시스템 개발
11	국방	지능형CCTV, 빅데이터를 활용한 국제범죄, 테러 등의 범죄를 사전에 예측하여 신속하게 대응
12	법률	법률, 세금 등 다양한 법 관련 분야에 대한 실시간 대화형 통합민원상담 서비스 제공
13	사무관리	사무실 내 업무환경 조성 및 문서 관리, 인사·조직 관리 등 ERP 서비스를 통해 기업의 전반적인 관리 서비스 제공
14	공공분야	생애주기별 실업, 주거, 돌봄, 질병 등 사회적 위험에 대응하는 통합형 복지서비스 제공
15	교육	학습패턴, 공부방식 분석 등을 통한 개인 맞춤형 교육 서비스 제공
16	문화·관광	VR·AR을 통해 방문자가 직·간접적으로 현장 체험할 수 있는 서비스 제공
17	안전	최근 발생하는 싱크홀 등 지하구조물 관리 및 태풍, 홍수 등 재해재난 예방 서비스 제공

한편, 맥킨지(2018. 4)에서는 산업 전반에 AI를 응용한 미국 내의 400여개의 사례를 분석하여 이들이 창출하는 잠재적 경제 효과를 전망한 결과를 발표하였다. 소매·여행·금융·농업 등 미국의 19개 전통산업과 9개 비즈니스 영역에 AI기술을 접목할 경우 연간 3.5조~5.8조 달러의 경제적 효과가 창출될 것으로 추산하였다. 또한, 소매부분이 연간 최대 8,000억 달러로 가장 높은 효과를 창출할 것으로 전망한 가운데 운송·물류, 여행, 소비재, 공공 사회재, 자동차 부품 부문이 두각을 보였다.

〔그림 2－7〕 적용산업별 AI 잠재적 경제효과 및 영향



자료: 맥킨지(2018. 4).

본 연구에서는 맥킨지(2018. 4)보고서에서 분류한 AI기술 접목에 따른 19대 전통분야 분류를 통해 국내 유스케이스에 적용하면 제조 분야를 세분화할 수 있지만 지식정보서비스 산업 등에 대한 분류가 미비하여 국・내외 연구결과를 기초를 한 17개 AI 적용산업 분류를 통해 지능정보기술 적용현황을 분석하였다.

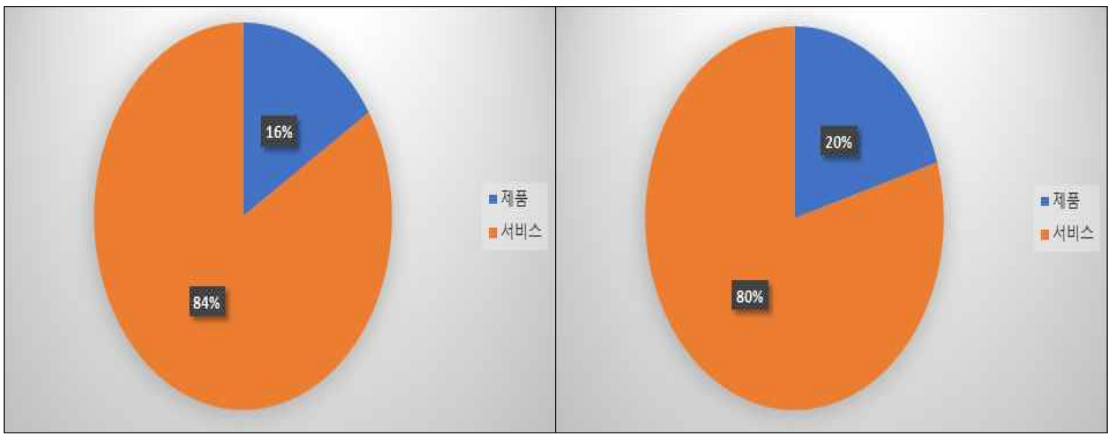
따라서 본 연구에서의 유스케이스 분석 대상은 해외의 경우, CB insights에서 제시한 인공지능 스타트업 100개 기업 및 주요 인공지능 기업(Google, IBM, 아마존, 애플 등)중심으로, 국내의 경우 지능정보산업협회로부터 제공받은 인공지능 기업을 대상으로 하였다. 구체적으로 분석대상 기업은 총264개(국내 기업은 155개, 국외 기업은 109개)이며, 분석대상이 된 제품/서비스는 총 345개(국내 제품 216개, 국외 제품/서비스 129개)를 이용하여 인공지능 유스케이스를 분석하였다. 조사 방법은 문헌 및 홈페이지 조사를 바탕으로 시행하였으며, 유스케이스 등이 제시된 문헌 등을 참고하였다.

〔그림 2-8〕 산업별 인공지능 유스케이스 분석 프로세스

단 계	연구내용	수행방법
국내외 주요 제품 조사	• 국외 출시(예정) 제품 조사 • 국내 출시(예정) 제품 조사	• 문헌조사 • 온라인 조사
제품의 유형화	• (산업) 17개 산업유형	• 산업별 Grouping
국내외 비교분석	• 국외/국내 제품의 산업/기술분야별 집중도 분석	• Matrix 분석
국내 Gap 도출	• 국내외 비교 분석을 통한 중점영역 및 공백분야 도출	• Gap 분석

분석대상인 총 345개의 인공지능 제품/서비스는 국내·외 모두 인공지능 제품보다는 SW·플랫폼·솔루션 등 서비스 위주로 개발 중인 것으로 조사되었다. 이는 대부분의 국내외 기업들이 인공지능과 결합되어야 디바이스에 초점을 맞추기보다는 기존 IT 다바이스(PC, 스마트폰, 음성인식 스피커 등)를 활용한 서비스 등을 개발하는데 초점을 맞추는데 기인하고 있다. 또한, 인공지능 HW는 개발비용 및 인프라 등의 진입장벽이 높으며 오랜 시간 검증이 필요한 분야로 스타트업 위주로 구성되고 있는 인공지능 산업 생태계에 비추어볼 때, 아래와 같은 제품/서비스 비중은 지속적으로 이어질 전망이다.

〔그림 2-9〕 국내·외 인공지능 제품/서비스 비중



국내 인공지능 제품/서비스 비중 국외 인공지능 제품/서비스 비중

2. 국내외 인공지능 적용산업 포지셔닝

가. 해외 인공지능 적용산업 포지셔닝

CB insights의 인공지능 100개 스타트업을 포함하여 주요 인공지능 관련 기업의 인공지능 제품・서비스 129개에 대해서 산업 집중도를 조사하였다. 조사방식은 각 기업의 제품 및 서비스를 도출하고, 현재 산업 분야에서 적용되고 있는 유스케이스를 추적하여 해당되는 산업군에 포지셔닝하였다. 대부분의 제품 및 서비스의 경우 단일 산업군보다는 다양한 산업에서 활용되고 있음에 따라 복수로 포지셔닝하였으며, 그 결과 17개 산업 분야 중 웰니스 분야에 대한 산업 집중도가 가장 높았으며 에너지 분야에 대한 제품/ 서비스 개발이 저조한 것으로 조사되었다.

〈표 2－5〉 해외 인공지능 제품/서비스별 적용 산업 집중도

구분	의료	국방	교육	농업	법률	웰니스	제조	홈	금융	문화관광	유통	사무관리	도시	교통	에너지	안전	공공
시각지능	10	7	2	3	1	9	5	7	5	2	2	2	6	9		3	5
언어지능	6	5	6		2	11	3	15	9	4	4	1	3	5			4
자연어처리	4	2	4		2	6	2	16	7	2	4			3			2
음성지능	4	6	7	1	2	13	3	16	7	4	3	2	5	7			1
공간지능		4		1		4		2	1	1	1		5	8			
기계학습	20	8	10	2	2	21	11	15	14	7	10	5	6	8	1	3	7
지식표현	10	2	2		2	4	3	12	7	3	4		1	3			6
DB관리	19	12	4	3	2	26	18	9	27	6	6	14	7	9	1	8	12
합계	73	46	35	10	13	94	45	92	77	29	34	24	33	52	2	14	37
순위	4	6	9	16	15	1	7	2	3	12	10	13	11	5	17	14	8

17개 산업 분야 중 웰니스(94개)로 가장 집중되어 있어 해외 인공지능 기술이 개인의 신체적, 정신적 건강 등 개인 맞춤형 기술개발에 집중하고 있는 것으로 분석된다. 2번째로 집중

도가 높은 홈(92개)의 경우, 음성비서 등의 AI 홈 디바이스 등이 급속히 보급되고 있으며 이에 따른 다양한 서비스들이 개발되고 있는 것으로 분석된다. 최근 글로벌 가전업체는 자체 인공지능 플랫폼 등을 개발하여 가전의 기본 기능 외에 인공지능 기술 등을 결합하여 소비자 지향형 서비스를 부가하고 있다. 3번째로 집중도가 높은 의료(73개)의 경우, 인공지능 정밀진단 분야에 적극적으로 인공지능 기술이 도입되고 있으며 신약개발에 있어서도 적극적으로 활용되고 있다.

가장 낮은 산업 집중도를 나타내는 에너지 분야는 국내외적으로 에너지를 자가 생산하고 잉여 에너지를 판매하는 'E(Energy)－프로슈머'에 대한 기술이 미흡한 것으로 분석되지만, 향후 인공지능 제품/서비스 시장의 틈새시장으로 향후 적용 확대 가능성이 있다.

나. 국내 인공지능 적용산업 포지셔닝

지능정보산업협회에서 제공받은 인공지능 전문가들을 통해 조사된 155개의 국내 인공지능 제품/서비스 216개에 대한 산업 집중도를 분석하였다. 조사방식은 해외 인공지능 적용산업 집중도 방식과 동일하다. 17개 산업 분야에 대한 제품/서비스 중복 선택을 통해 국내 산업 집중도를 분석한 결과 웰니스 분야에 가장 집중되고 있으며, 국내에서 인공지능 기술도 기술 진입장벽이 낮은 개인 맞춤형 산업으로 집중되고 있는 것으로 분석된다. 해외와 비교할 때 제조에 산업 집중도가 높은 것을 확인할 수 있는데, 최근 스마트공장화 맞물어 제조 자동화 기술로 인공지능이 적극 활용되고 있는 것으로 판단된다. 또한, 금융권을 중심으로 챗봇 및 로보어드바이저 등을 형태로 인공지능 활용 서비스가 확산됨에 따라 관련 서비스들이 다수 분포하고 있다.

〈표 2－6〉 국내 인공지능 제품/서비스별 적용 산업 집중도

구분	의료	국방	교육	농업	법률	웰니스	제조	홈	금융	문화관광	유통	사무관리	도시	교통	에너지	안전	공공
시각지능	6	4	5			18	14	8	3	1	5	3	9	13		5	3
언어지능	6	1	3	1	1	21	13	17	10	12	18	4	7	3	1	1	8
자연어처리	2	1	1	1	1	4	1	10	4	4	8	1	1	1	1	1	

구분	의료	국방	교육	농업	법률	웰니스	제조	홈	금융	문화관광	유통	사무관리	도시	교통	에너지	안전	공공
음성지능	6	1	7	1	1	22	5	18	3	8	8	1	4	4	1	1	4
공간지능		3	1	1		3	8	8	1	1	1		4	6			
기계학습	17	2	6	5	3	52	22	15	35	16	24	9	16	7	2	9	14
지식표현	8	3	6	1	1	25	19	14	14	9	15	3	7	4	4	3	7
DB관리	10	3	6	4	1	38	33	14	41	5	23	13	10	8	6	9	22
합계	55	18	35	14	8	183	115	104	111	56	102	34	58	46	15	29	58
순위	9	14	11	16	17	1	2	4	3	8	5	12	6	10	15	13	6

3. 시사점

현재 인공지능 제품·서비스의 개발은 서비스 개발에 집중되고 있어 기술력 확보를 통해 인공지능 로봇, 자율주행 분야 등 제품 영역으로 확장 개발이 필요하다. 특히, 국내에서는 인공지능 원천기술을 포함한 알고리즘 개발보다는 어플리케이션 개발을 중심으로 한 서비스 개발을 바탕으로 한 인공지능 SW 분야에 많은 기업들이 치중되어 있다. 국내외 인공지능 제품 및 서비스에 대한 Matrix 분석을 통해 인공지능 적용이 미흡한 산업을 분석하여 요구된다. 아래 그림과 같이 국내외 Matix 분석 결과에 따라 상대평가를 통해 집중도의 등급을 1~4단계로 구분하여 각 산업별 집중도를 도식화하였다. 집중도의 등급은 해외/국내 Matrix 분석 결과에 대해 집중도 수치를 4단계로 구분하여 적용 집중도 레벨을 설정하였다.

〈표 2－7〉 집중도 Mattix 레벨 설정 기준

레벨 / 구분	LV.1	LV.2	LV.3	LV.4
해외 빈도	1~7	8~14	15~21	22~28
국내 빈도	1~13	14~26	27~39	40~52

〔그림 2－10〕 해외 인공지능 제품·서비스 기술 적용도 Matrix 분석 결과

구분	의료	국방	교육	농업	법률	웰니스	제조	홈	금융	문화 관광	유통	사무 관리	도시	교통	에너지	안전	공공
시각	LV. 2	LV. 1	LV. 1	LV. 1	LV. 1	LV. 2	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 2		LV. 1	LV. 1
언어	LV. 1	LV. 1	LV. 1		LV. 1	LV. 2	LV. 1	LV. 3	LV. 2	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1			LV. 1
자연어 처리	LV. 1	LV. 1	LV. 1		LV. 1	LV. 1	LV. 1	LV. 3	LV. 1	LV. 1	LV. 1			LV. 1			LV. 1
음성	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 2	LV. 1	LV. 3	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1			LV. 1
공간		LV. 1		LV. 1		LV. 1		LV. 1	LV. 1	LV. 1	LV. 1		LV. 1	LV. 2			
기계	LV. 3	LV. 2	LV. 2	LV. 1	LV. 1	LV. 3	LV. 2	LV. 3	LV. 2	LV. 1	LV. 2	LV. 1	LV. 1	LV. 2	LV. 1	LV. 1	LV. 1
지식	LV. 2	LV. 1	LV. 1		LV. 1	LV. 1	LV. 1	LV. 2	LV. 1	LV. 1	LV. 1		LV. 1	LV. 1			LV. 1
DB	LV. 3	LV. 2	LV. 1	LV. 1	LV. 1	LV. 4	LV. 3	LV. 2	LV. 4	LV. 1	LV. 1	LV. 2	LV. 1	LV. 2	LV. 1	LV. 2	LV. 2

〔그림 2－11〕 국내 인공지능 제품·서비스 기술 적용도 Matrix 분석 결과

구분	의료	국방	교육	농업	법률	웰니스	제조	홈	금융	문화 관광	유통	사무 관리	도시	교통	에너지	안전	공공
시각	LV. 1	LV. 1	LV. 1			LV. 2	LV. 2	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1		LV. 1	LV. 1
언어	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 2	LV. 1	LV. 2	LV. 1	LV. 1	LV. 2	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1
자연어 처리	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	
음성	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 2	LV. 1	LV. 2	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1
공간		LV. 1	LV. 1	LV. 1		LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1		LV. 1	LV. 1			
기계	LV. 2	LV. 1	LV. 1	LV. 1	LV. 1	LV. 4	LV. 2	LV. 2	LV. 3	LV. 2	LV. 2	LV. 1	LV. 2	LV. 1	LV. 1	LV. 1	LV. 2
지식	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 2	LV. 2	LV. 2	LV. 2	LV. 1	LV. 2	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1
DB	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 3	LV. 3	LV. 2	LV. 4	LV. 1	LV. 2	LV. 1	LV. 1	LV. 1	LV. 1	LV. 1	LV. 2

국내외 인공지능 제품·서비스는 산업 분야에서는 공통적으로 홈, 웰니스 분야에 집중되어 있는 것으로 분석되었다. 국내 인공지능 적용 17대 산업에 인공지능의 기술들이 다양하게 적용되고 있으나, 기술의 집중도가 부족하여 인공지능 기술의 고도화가 필요하다. 특히 국내 인공지능 제품 및 서비스는 해외 대비 웰니스와 기계학습, DB분야에 집중되어 있으며 타 산업 및 기술의 적용이 부족하다.

언어지능, 자연어처리, 음성지능에 대해서 해외에서는 홈서비스에 집중적으로 활용하고 있으나 국내에서는 해외 인공지능 플랫폼 등을 활용한 제한적인 기술 개발을 통해 서비스가 되고 있어 상대적으로 집중도가 높지 않다. 국내에서도 응용서비스 개발뿐만 아니라. 텐서플로우, 파이토치 등 인공지능 플랫폼 개발 및 공유 등을 통해 인공지능 기술 기반을 마련해야 홈 서비스 등 다양한 산업군에 적용이 확대될 수 있을 것으로 판단된다.

제 3 절 4차 산업혁명의 산업(경제)·사회변화상 분석[4)]

1. 산업 구조의 변화

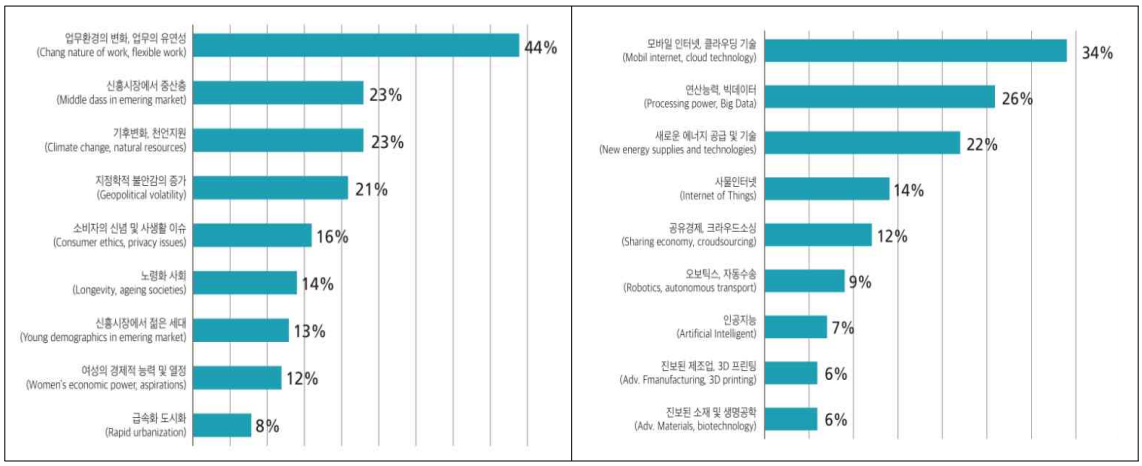
인공지능, IoT, 빅데이터, 클라우드 등 ICT 기반 기술들이 제4차 산업혁명의 핵심동인으로 작용하고 있다. ICT 핵심기술들은 범용기술로 작용하면서 생산성 개선을 통한 산업 고도화에 기여하고 있을 뿐만 아니라, 비즈니스 전 과정의 '디지털화'를 촉진시킴에 따라 기존의 산업구조를 빠르게 변화시키고 있다. 특히, ICT 기업들이 다양한 산업 영역으로 진출하면서 가치사슬 전반의 혁신을 야기하고, 기존의 수직적 산업 영역 구분이 붕괴되고 있다. 구글과 애플 등 글로벌 ICT기업들은 모바일 OS 경쟁력을 기반으로 자동차, 의료 등 타산업으로 영향력을 확대시키고 있으며, 아마존은 오프라인 진출을 본격화하며 오프라인 업체들을 위협하고 있다. 이처럼 ICT 기업들의 타산업으로의 진출이 활발해지면서 ICT 생태계 경쟁 양상이 전산업으로 확대되고 있다.

---

4) 김정언 외(2017), 이호영 외(2016)을 중심으로 보완·재정리 하였음

〔그림 2－12〕 제4차 산업혁명의 주요 변화 동인

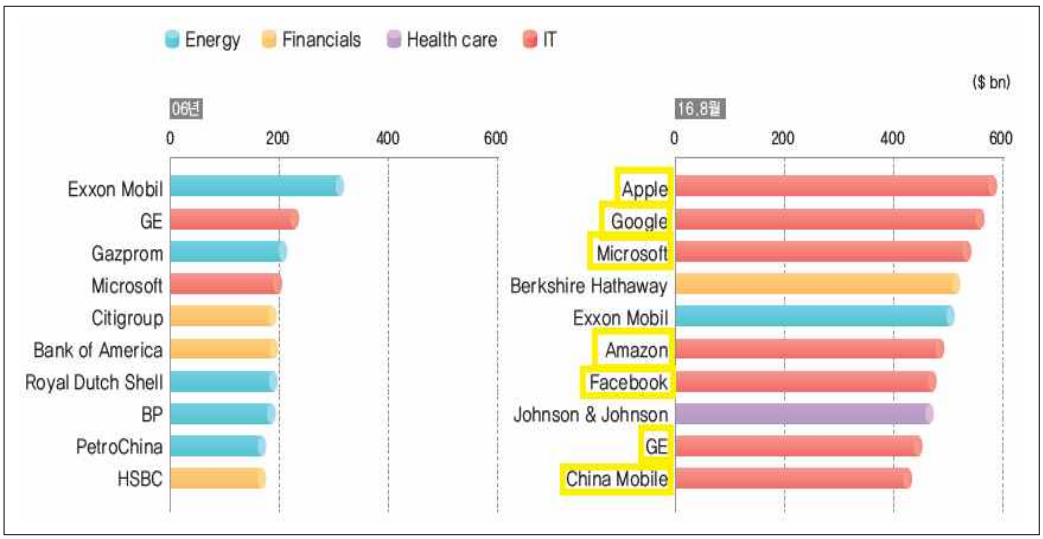
〈사회－경제학적 주요 변화 동인〉 〈기술적 주요 변화 동인〉



자료: WEF(2016b)

제4차 산업혁명 시대에 들어서면서 데이터와 지식이 노동, 자본 등 기존의 생산요소를 압도하는 새로운 경쟁원천으로 부상하고 있다. 대규모 설비투자와 인건비 절감 여부 보다는 기술혁신 여부가 기업의 성과에 결정적인 요인으로 작용하면서 데이터와 핵심기술이 결합된 지식자원의 중요성이 확대되고 있다. 데이터와 핵심기술들이 산업의 경쟁 원천으로 부상하면서 이들을 보유한 혁신적인 기업들이 산업을 주도하고 있다. 도서 및 유통 분야의 아마

〔그림 2－13〕 2006년과 2016년 시가총액 10대 기업 비교



자료: 관계부처 합동(2017. 1)

존, 개인운송 분야의 우버, 사진 분야의 인스타그램 등이 대표적인 사례들이다. 데이터를 확보할 수 있는 생태계를 구축하고, 이를 활용할 수 있는 핵심기술을 보유한 기업이 시장을 주도하고 많은 이윤을 창출하고 있다. 〔그림 2－13〕을 보면, '16년 8월 기준으로 전 세계 시가총액 10대 기업 대부분이 데이터와 기술 기반의 혁신적인 ICT 기업임을 확인할 수 있다.

ICT 융합의 확산, ICT 기업들의 타산업 진출 등에 따라 산업의 경쟁 구도 역시 변화하고 있다. 최근 몇 년간 ICT 부문에서 뚜렷하게 나타나고 있는 플랫폼 기반의 생태계 경쟁이 확산되고 있다. 인공지능, 사물인터넷, 빅데이터, 클라우드 등 ICT 기술의 활용 증가는 보다 많은 사용자가 플랫폼 생태계에 참여할 수 있는 환경을 조성하고 있다. 플랫폼 기반 생태계의 핵심은 많은 사용자가 플랫폼에 직접 참여하여 데이터를 지속적으로 생성하는 동시에 활용하는 구조를 형성한다는 점이다. 이러한 플랫폼을 통해 관련 제품과 서비스들이 유기적으로 연결되어 통합 서비스가 제공됨으로써 기존의 제품・서비스(Stand alone 형태)를 압도하게 된다.

제4차 산업혁명의 핵심기술인 지능정보기술은 학습을 통해 성능이 지속적으로 발전되는 특징을 갖고 있기 때문에 우수제품을 시장에 먼저 출시하여 생태계를 구축한 플랫폼 기업들이 시장 지배력을 확대하며 빠르게 성장할 전망이며, 이에 따라 글로벌 시장에서 플랫폼 생태계 확보를 위한 경쟁은 더욱 치열해 질 것으로 보인다. 애플, 구글 등과 같이 글로벌 생태계를 주도하는 인공지능 기반의 새로운 플랫폼 기업의 등장도 예상해 볼 수 있다. 스타트업과 중소벤처기업들에게는 응용 서비스 분야의 경우 이미 형성된 글로벌 플랫폼을 통해 글로벌 시장 진출이 가능해 지면서 새로운 성장의 기회가 발생할 수 있을 것으로 판단된다.

한편, 제4차 산업혁명의 진전에 따른 산업구조 변화에 있어서 우리가 간과해서는 안될 한 가지는 제조업의 서비스화가 빠르게 진행되고 있다는 점이다. 신기술 및 플랫폼 비즈니스와 같은 혁신적인 패러다임과 결합되면서 맞춤형 소량생산, 스마트팩토리 등 제조 ・ 공정 측면의 혁신이 활발하게 일어나고 있다. 기존 저비용 기반의 대량생산과 유통에서 벗어나 인터넷을 통한 맞춤형 생산・유통・소비가 가능한 디지털 제조 시대가 시작되었다. 누구든지 아이디어를 디지털화하고 시제품 공유를 통한 피드백을 통해 제품의 완성도를 높일 수 있는 기회가 주어지기 때문에 개인이나 벤처기업 등도 소규모 자본으로 생산이 가능해졌다(이은민, 2016). 또한, 인터넷 기반의 온디맨드 경제가 확산되어 다품종 맞춤형 소량생산에 대한 수요가 증가하였고, 더불어 스마트 기기들의 발전과 빅데이터, 클라우드가 결합되어 디지털

제조공정이 가능해짐에 따라 스마트팩토리가 확산되고 있다. 가까운 미래에는 IoT를 통해 누적된 빅데이터를 클라우드로 공유하고, 빅데이터를 분석해 생산시뮬레이션을 가동하는 스마트팩토리가 일반화될 것으로 예상된다.[5] 스마트팩토리의 도입으로 에너지·인건비 등의 비용 절감 및 부가가치 증대로 인한 생산성 증가를 기대하며 제조업의 질적 고도화를 통한 경제 성장의 기반이 마련될 것으로 예상된다(조윤정, 2015). 이미 GE 등 해외 선도 기업은 스마트팩토리의 도입으로 납기 단축, 라인 정지시간 최소화, 불량 축소, 에너지 절감 등의 성과를 보이고 있으며, 국내에서도 '14년 스마트팩토리 시범사업으로 그 효과를 입증하였다(조윤정, 2015).[6]

이처럼 디지털 제조공정이 가능해짐에 따라 학습이 가능한 인공지능과 로봇이 점차적으로 노동을 대체하고 있으며, 적은 비용으로 더 높은 생산성을 얻게 되었다. 이러한 변화로 인해 비용절감 등을 목적으로 해외에 제조 공장을 이전했던 기업들이 해외 인건비 상승 등 각종 비용 상승으로 인해 오프쇼어링 매력을 상실하고, 고객 니즈에 대한 보다 신속한 대응 필요성 증대와 기술 혁신으로 인한 비용 감소 등이 겹쳐 다시 자국으로 돌아오는 리쇼어링으로 방향을 전환하고 있다.[7] 여기에 사물인터넷과 3D 프린터, 로봇 등 신기술을 적극 접목한다면 자국에서 제조하더라도 충분히 원가 경쟁력이 있고, 오히려 시장 수요에 맞춰 민감하게 대응할 수 있다는 판단이 리쇼어링을 가속화 시키고 있다.

최근 많은 글로벌 기업들은 제조업에 ICT 서비스를 결합하는 비즈니스를 추진하고 있다. 과거 제조업 기업들은 제품의 생산과 판매에 주된 가치를 두었지만, 점차 소비자 데이터를 기반으로 한 서비스(마케팅, 애프터 서비스 등)에 더 많은 가치를 두고 있다.[8] 쉽게 말해 제

---

5) 스마트 공장의 시장규모를 살펴보면, 글로벌 시장 기준으로 '12년 1,552억 달러를 기록하였으며, '18년에는 2,460억 달러 규모로 성장할 것으로 전망된다. 국내 스마트 공장 시장규모는 '12년 24억 달러로 세계시장의 1.5% 수준이었으나, '18년 44억 달러로 연평균 11.2%의 성장을 기대하고 있다(홍승호 외, 2015. 12. 11)

6) '14년 277개 중소·중견기업을 대상으로 스마트공장 시범사업을 진행하였는데 그 결과, 생산성 22.7%, 품질 33%, 비용 23%, 매출 17%의 개선효과가 나타났다. 향후 확대 적용으로 '17년까지 제조업 부가가치율과 생산성 향상을 기대하고 있다.

7) 대표적으로 GE는 해외에 만들었던 공장을 미국으로 이전시키거나 신설 공장을 미국에 건립하는 형태로 리쇼어링을 강화하고 있다. '12년 6월, 1억 7천만 달러 규모의 산업용 배터리 공장을 미국에 지었고 같은 해 중국에 있던 온탕기 공장과 멕시코에 있던 가전 공장을 미국으로 옮겨 왔다(송재용, 2016. 4. 4).

8) 이와 같은 변화를 제조업의 서비스화(Product Servitization)라고 일컫는다(장병열, 2014).

조와 서비스의 융합으로 제품을 단순히 만들어 유통하는 것에 그치지 않고 서비스 모델을 융합함으로써 그 가치를 확장하는 것이다(문형철, 2016). 애플이 하드웨어 및 소프트웨어와 서비스를 연계하는 플랫폼인 아이튠즈(I-tunes) 와 앱스토어(App Store)를 구축하여 소비자와 콘텐츠 제공자를 서로 연결시키는 소비 생태계를 구축한 것이 대표적인 예이다. 애플이 스마트폰 등의 기기와 플랫폼을 결합하여 사용자에게 고부가가치 서비스를 제공한 것이 애플을 글로벌 시가총액 1위의 기업(2015년 기준)으로 성장시킨 결정적인 원동력이었다(이은민, 2016). 국내의 대표적인 예로 웅진코웨이의 정수기 비즈니스를 들 수 있는데 단순히 정수기를 제조하여 판매하는 것이 아니라 렌탈과 AS라는 서비스 모델을 융합하여 새로운 비즈니스를 창출하였다. 웅진코웨이는 렌탈 서비스 사업으로 ’00년 대비 ’09년 500%의 매출성장률을 보였으며 AS 시스템인 ‘코디와 코닥’ 도입으로 12,000여개의 신규 일자리도 창출하였다(김성덕, 2000).

기존의 제조업은 기술의 표준화와 저비용 생산으로 더 이상 성능과 가격으로 제품의 차별화가 어려운 상황이었다. 이러한 상황에서 제조업의 서비스화는 제품 차별화를 위한 대안이 되고 있으며, 이러한 추세는 향후에도 지속될 것으로 판단된다.

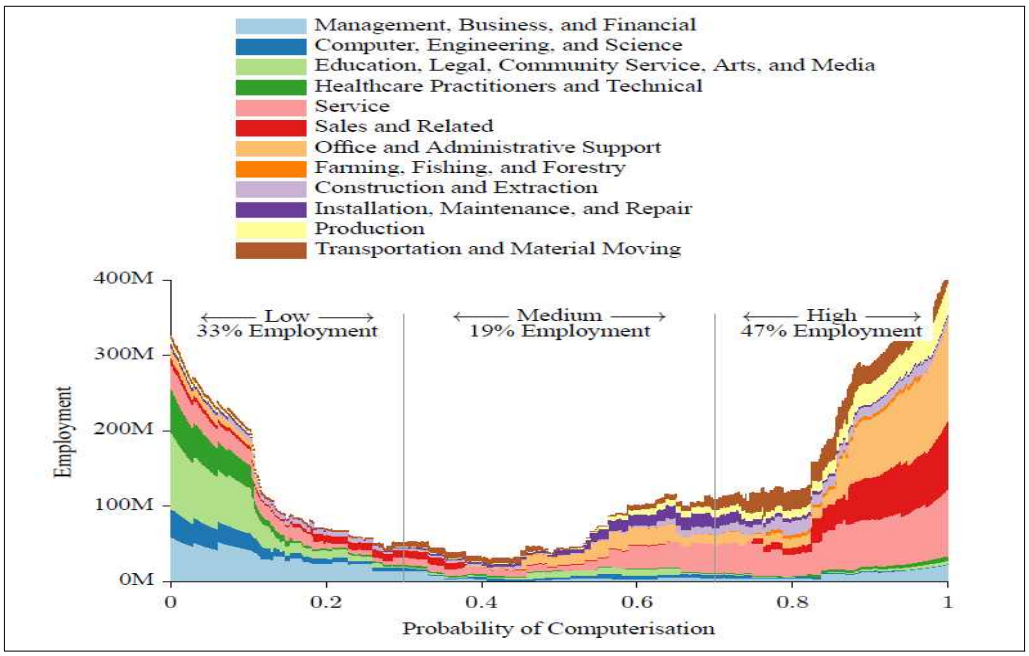
2. 고용 구조의 변화

’16년 세계경제포럼(World Economic Forum)에서는 제4차 산업혁명을 화두로 던지면서 유전학, 인공지능(AI), 로봇기술(Robotics), 나노기술 및 바이오기술 등 지능정보기술의 발전이 비즈니스 모델을 빠르게 변화시키고 있음을 보였다. 이런 비즈니스 모델의 대대적인 변화는 향후 노동시장의 고용구조에 지대한 영향을 미칠 것이며, 현재 글로벌 산업의 변화에 영향을 미치는 많은 주요 요인들은 일자리의 소멸과 생성 및 직업 구조에 광범위한 영향을 미칠 것으로 보인다(WEF, 2016).

기술의 발달이 인간의 노동력을 대체할 것인가에 대해서는 지속적인 논쟁이 있어왔다. 기술 혁신은 전통적인 산업의 생산성을 향상시켜 시장 규모를 증가시키거나 혹은 기존 산업과 융합하여 새로운 시장을 창출함으로써 노동 수요의 증가와 새로운 직업군을 만드는 등 인적 자본과 상호 보완적인 역할을 수행한다. 이와는 반대로 기술 혁신으로 인하여 사람이 하던 일을 기계가 대신할 수 있게 되고, 이는 노동 수요를 감소시키거나 혹은 관련 직업군의 소멸로 이어지기도 한다. 지능정보시대 이전의 기술 발달이 단순반복적 작업에 한해서 인간의

노동력을 대체해 온데 반하여, 인공지능 기술의 발달은 보다 복잡한 사고 능력을 요하는 일자리마저 기술이 인간을 대체할 수 있음을 보여주고 있다.[9] 이러한 현상으로 인하여 사회 전반에 걸쳐 기술이 인간 노동력을 대체할 것인가에 대한 논란이 급증하고 있다.[10]

〔그림 2－14〕 미국 내 직종의 대체확률별 고용인원



자료: Frey and Osborne(2017) Figure III, p.37.

하지만, 그간 기술의 발전이 노동시장에 미치는 영향을 살펴봤을 때, 혁신적 기술이 개발되고 발전해도 경제적 효용가치에 대한 검증, 관련 법제도의 정비, 사회적 합의가 전제되어야 사회에 온전히 적용될 수 있기 때문에 기술이 완전하게 사람을 대체하는 데는 상당한 시간이 소요됨을 알 수 있다. 따라서 혁신적 기술이 도입되기까지의 시간 동안 노동시장과 근로자들이 새롭게 바뀌어가는 환경에 적응하므로 일자리의 대체 폭은 제한적일 수 있고, 따라서 지능정보기술의 발달로 인해 감소할 수 있는 일자리 수에 대한 현재의 예측치는 과대

9) ’16년 초, 게임의 전개가 다양해 오랫동안 컴퓨터가 정복하지 못했던 바둑에서 구글의 인공지능 알파고가 인간을 대표한 이세돌에게 승리를 거둬 복잡한 사고를 요하는 분야에서마저 기계의 인간 대체에 대한 우려를 자아냈다.

10) ’15년 테크프로 리서치(Tech Pro Research)는 인공지능에 관한 인식 수준을 알아보고자 설문을 진행하였는데, 응답자의 63%는 인공지능이 비즈니스에 도움이 될 것이라고 기대하고 있으나, 34%는 인공지능 관련 기술로 인해 일자리를 잃게 될 가능성에 두려움을 나타냈다.

계상된 것일 수 있다.

현재 진행 중인 제4차 산업혁명은 그 어느 때 보다 기술혁신의 파급력이 크다는 점에서 기술과 인적자본 간 대체효과와 보완효과에 대한 불확실성을 내포하고 있다. 이러한 불확실성 속에서 제4차 산업혁명 시대의 고용에 관한 많은 연구와 토론이 진행되고 있으며, 아직까지 기술의 일자리 대체 전망에 대한 비관론과 낙관론이 대립하여 공존하는 가운데 뚜렷한 결과에 도달하지 못하고 있다.

Frey and Osborne(2017)은 미국의 노동시장을 대상으로 기술의 일자리 대체 가능성을 분석한 결과, 향후 20년 이내에 미국 내 일자리의 47%가 대체확률 0.7 이상인 고위험군 직종에 속하는 것을 도출하였다(〔그림 2－14〕 참조). 독일 제조업 분야에서는 기계가 인간의 업무를 대체함에 따라 생산부문 120,000개(부문 내 4%), 품질관리부문 20,000개(부문 내 8%) 및 유지부문 10,000개(부문 내 7%)의 일자리가 감소하고 생산계획부문의 반복형 인지업무(Routine cognitive work)도 20,000개 이상의 일자리가 사라질 것으로 예측되고 있다(Boston Consulting Group, 2015). 이러한 현상은 ’25년 이후 더욱 가속화 될 것으로 전망되고 있다. 호주는 노동시장의 39.6%(약 5만 명의 노동인력)가 수십 년 내 컴퓨터에 의해 대체 될 것으로 예상하고 있고, 그 중 18.4%는 업무에서의 역할이 완전히 사라질 가능성이 높을 것으로 보고 있다(CEDA, 2015).

기술의 일자리 대체 가능성에 대한 부정적인 연구결과들에 이어 최근 ’16년 다보스 포럼에서는 ’20년까지 인공지능과 로봇의 영향으로 전 세계적으로 총 710만개의 일자리가 소멸되고, 새롭게 200만개의 일자리가 창출되어 총 510만개의 일자리 감소를 예상했다(WEF, 2016b). 정지형 외(2014)는 직업별 전산화 위험도 측정 방식을 도입하여 스마트 기술 발전에 따른 우리나라 산업별 고용 대체가능성을 분석하기 위해 전문가조사를 이용하였다. 그 결과 제조업보다 오히려 의료, 운수, 교육, 상담 등 서비스 분야 고용이 많이 대체될 수 있다는 결과를 제시하였다.

Frey and Osborne(2017)의 연구방법을 참조하여 우리나라 노동시장에서 기술진보에 따른 고용대체 고위험군 일자리의 비중이 얼마나 되는지를 분석한 결과 향후 20년 이내 약 55～57%의 일자리가 소멸할 것으로 나타났다(김세움 2015, 최창옥 2015). 이런 연구결과들은 연구대상 및 방법론에 차이가 있으나 거의 모든 직종에서 향후 10～20년간 자동화, 인공지능 및 로봇에 의한 고용대체가 이뤄진다는 것을 말하고 있다.

〈표 2-8〉 스마트 기술에 의한 제조·서비스 분야 고용 대체가능성

대분류 (기술群)	소분류 (세부기술)	스마트기술에 의한 고용 대체가능성								
		제조		서비스						합계
		조립 운반	공정 관리	의료	법률	운수	금융	교육	상담	
스마트 컴퓨팅	**빅데이터 분석**	1	3	3	3	2	3	3	3	21
	딥러닝	0	0	3	3	2	2	2	3	15
	감성 컴퓨팅	1	0	2	0	2	1	3	3	12
	NLQA	1	1	3	3	2	3	3	3	19
	자동통역	1	1	3	1	3	1	3	2	15
스마트 머신	웨어러블디바이스	2	1	3	0	0	0	1	3	10
	개인비서로봇	1	1	2	2	0	0	2	3	11
	자율주행자동차	0	0	0	0	3	0	0	0	3
	학습적응형 로봇	3	1	0	0	0	0	0	0	4
	물류자동화 로봇	3	1	0	0	1	0	0	0	5
	무인 택배머신	0	0	0	0	3	0	0	0	3
스마트 인프라	**클라우드컴퓨팅**	1	3	2	3	2	3	3	3	20
	사물인터넷(IoT)	1	3	1	1	2	1	3	1	13
	5세대 이동통신	1	2	2	1	3	3	3	1	16
합계		16	17	24	17	25	17	26	25	167

자료: 정지형 외(2014).

그러나 제4차 산업혁명이 고용 구조에 미치는 영향에 관하여 부정적인 전망만 있는 것은 아니다. 기술 발전의 노동시장 파급력에 관한 그간의 역사에 비춰볼 때, 제4차 산업혁명의 혁신을 단순히 일반적인 자동화의 발전으로 간주할 경우 기술혁신이 꼭 수많은 직업의 소멸, 고용 감소로 이어질 이유는 없다. 제4차 산업혁명으로 인하여 고숙련(High- skilled) 노동자에 대한 수요가 증가할 뿐만 아니라, 관련한 기술 직군 및 산업 분야에서 새로운 일자리가 등장할 것이라는 예측도 존재한다. 특히 산업계에서는 인공지능, 3D 프린팅, 빅데이터 및 산업로봇 등 제4차 산업혁명의 주요 변화 동인과 관련성이 높은 기술 분야에서 200만개의 새로운 일자리가 창출되고, 그 중 65%는 신생직업이 될 것이라는 전망도 있다(GE, 2016). 또한 독일 제조업 분야 내 노동력의 수요는 대부분 IT와 S/W 개발 분야에서 경쟁력을 가진 노동자를 대상으로 나타날 것이고, 특히 IT 및 데이터 통합 분야의 일자리 수는 110,000개(약

96%)가 증가하고, 인공지능과 로봇 배치의 일반화로 인해 로봇 코디네이터 등 관련 분야 일자리가 40,000개 증가 할 것으로 전망되고 있다(Boston Consulting Group, 2015).

이러한 낙관적인 견해는 지능정보기술의 노동 보완효과가 클 경우를 상정한 것이다. 이 경우 생산량이 증대하면서 노동 수요가 증가 하고 이에 따라 공급에 변화가 일어나면서 임금 간 상호작용으로 순고용의 큰 변화 없이 노동시장이 작동하게 된다. 이 과정에서 점진적으로 직업과 직무의 변화가 일어날 수 있다. Autor(2015)는 기술의 인간 일자리 대체에 관한 비관론이 기술혁신의 노동 보완 측면을 간과하고 있음을 비판하고, 오히려 기술 진보로 인해 시장이 확대되고, 이에 따르는 노동수요 증대를 통해 순고용은 큰 변화가 없을 것이라 전망했다.

〈표 2－9〉 지능정보기술로 인한 일자리 창출 효과 추정

구분	설명	고용창출 효과
해외 AI 관련 산업 종사자 예측치 기반	미국, 캐나다, 호주 등 선진국 AI 관련 업종의 미래 고용 예측치를 기반으로 한국의 고용 증가분 추정	10~45만명
외부기관 예측자료 기반	한국고용정보원 등의 산업 및 직종별 인력 수요 예측치 중 AI, 빅데이터 등 신산업 직종 수치 합산	60~80만명

자료: 관계부처합동(2016. 12. 27)

미국, 캐나다, 호주 등 해외 선진국 AI 관련 업종의 미래 고용 예측치를 기반으로 우리나라의 고용 증가분을 추정한 결과, 지능정보기술의 발달로 인한 고용창출 효과는 10～45만명에 달할 것으로 분석됐다. 또한, 한국고용정보원 등의 산업 및 직종별 인력 수요 예측치 중 AI, 빅데이터 등 신산업 직종 수치를 합산한 결과 해외 예측치 기반의 추정보다 더 큰 60～80만 명의 신규 고용이 창출될 것이라는 결과를 얻었다(관계부처합동, 2017).

물론 기존 일자리 기반의 연구는 지능정보기술의 발전에 따른 새로운 고용수요 창출을 온전히 예측하지 못하기 때문에 미래 고용 변화를 예측하는 것은 쉽지 않다. 새로운 산업과 직업 출현은 예상하기 어려우나, 기존 일자리 대체는 상대적으로 예측 가능하여 이후 지능정보기술에 따른 고용영향분석은 부정적인 것이 일반적이다. 하지만 지능정보기술 발전에 따라 SW 엔지니어 등 기존 일자리 창출 뿐 아니라 신규 직종 창출, 직종 다양성 확대 등의 고용 환경의 구조적 변화가 생길 것은 분명하다. 특히 지능정보기술은 인간의 활동과 지능을

보완하거나 대체할 것으로 예상되므로 전혀 예상하지 못한 직종의 출현 또한 가능하다.

3. 직업 구조의 변화

제4차 산업혁명에 따른 일자리 지형의 변화는 단순히 고용 구조에만 영향을 미치는 것에 그치지 않고, 요구되는 노동의 종류를 변화시키거나 고용 인력의 "직무역량(Skills & Abilities) 변화"에 영향을 미치고 있다.

Muro and Andes(2015)는 로봇 활용도가 고용에 미치는 영향을 파악하기 위해 17개국 14개 산업의 로봇 활용도를 분석하였는데, 그 결과 고용과 로봇 활용 간 상관관계는 발견할 수 없는 반면 로봇의 이용으로 인하여 요구되는 노동의 종류가 변화되는 것을 보였다. 즉, 자동화의 대표적 사례인 로봇이 생산성을 향상시키나 직업은 감소시키지 않고, 대신 비숙련 노동의 수요를 감소시킨다는 것이다. 지능정보기술의 발달로 인한 자동화는 직업을 대체하기 보다는 직무를 변화시키고 시장의 확대를 이끌 수 있다. Remus and Levy(2015)는 특정 변호사 직종 내 직무별 자동화의 노동대체 가능성 및 피고용인의 직무별 시간 분배를 모두 고려한 분석에서 인공지능으로 인한 법률서비스 자동화는 변호사와 보조인의 대체효과(약 13% 투입시간 감소)도 있으나 법률서비스 시장 확대를 고려하면 고용에 큰 변화가 없을 것이라고 전망했다.[11)]

Bessen(2016)은 '80년부터 '13년까지의 데이터를 이용하여 컴퓨터의 활용이 고용에 미치는 영향을 분석하였는데, 이 기간 동안 평균 이상으로 컴퓨터를 활용한 산업의 고용은 증가한 반면, 그렇지 않은 산업의 고용은 상대적으로 부진하였음을 보였다. 이는 문제가 노동의 대체가 아닌 노동의 종류에서의 차이 즉 'skill gap'에 있음을 나타낸다.

---

11) Remus and Levy(2015)에서는 법률회사 피고용인의 직무를 6개 카테고리로 구분하고, 이 가운데 현재 자동화 효과가 큰 직무(문서 검토 등)에 투입되는 시간은 불과 4.1%, 어느 정도 대체 효과가 예상되는 직무의 시간은 39.7%, 대체 효과가 거의 없을 것으로 예상되는 직무(사례 심층 분석 등)의 시간 비중은 56.0%로 나타났다.

〈표 2－10〉 산업 분야별 요구 직무역량 변화 전망(2015~2020)

구분(%)	기초/인프라		소비자		에너지		금융서비스		보건		정보통신기술		미디어		이동수단		전문서비스		평균	
	현재	2020	현재	2020	현재	2020	현재	2020	현재	2020	현재	2020	현재	2020	현재	2020	현재	2020	현재	2020
복합문제 해결능력	42	33	28	31	49	38	35	39	35	36	36	46	－	－	32	24	35	38	36	36
사회적 능력	17	17	26	27	27	28	32	23	30	28	20	19	27	32	22	20	26	24	20	19
공정 능력	10	19	21	22	24	29	36	34	25	36	26	25	27	31	18	22	37	29	18	18
체계적 능력	22	26	28	25	24	18	23	22	－	－	26	24	－	－	16	23	16	16	16	17
지원관리 능력	21	15	38	35	29	24	20	20	－	－	16	19	28	32	26	28	24	29	14	13
기술적 능력	25	20	20	18	29	22	5	16	－	－	22	20	－	－	26	21	19	18	14	12
인지 역량	10	19	13	25	－	－	15	23	35	36	20	23	－	－	11	27	19	22	11	15
콘텐츠 능력	6	13	－	－	－	－	22	24	－	－	19	18	－	－	22	28	11	15	10	10
신체적 역량	－	－	－	－	－	－	－	－	－	－	－	－	－	－	－	－	－	－	5	4

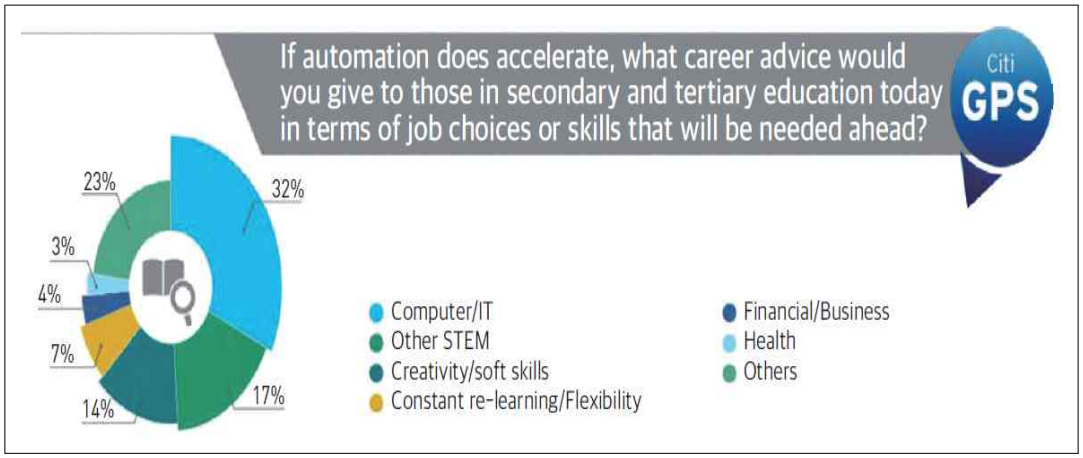
자료: WEF(2016b), 김진하(2016) 재인용

WEF 보고서에 따르면 제4차 산업혁명은 고용 인력의 직무역량 안정성(Skills Stability)에도 영향을 미치고, 산업 분야가 요구하는 주요 능력 및 역량에도 변화가 생겨 ‘복합문제 해결능력(Complex Problem Solving Skills)’ 및 ‘인지능력’ 등에 대한 요구가 높아질 것으로 전망되고 있다(WEF, 2016).

다수의 전망 보고서에서도 ‘컴퓨터/IT’ 및 ‘STEM(Science, Technology, Engineering, Mathematics)’ 분야의 지식이 효율적인 업무수행을 위해 필요함을 강조하고 있다(Oxford Univ., 2016). 특히 미국 제조업계에서는 ’18년까지 전체 일자리의 63%가 STEM 분야의 교육 이수를 요구하고, 첨단제조 분야의 15% 이상이 STEM 관련 고급학위(석사 이상)를 필요로 할 것으로 전망하고 있다(GE, 2016). 또한 미래사회에는 새로운 역할과 환경에 적응할 수 있는 유연성을 갖추고, 다양한 하드스킬(Hard Skills)을 활용할 수 있는 인력이 요구된다고 말하고 있다. 이와 함

께 지속적인 학제간 학습(Interdisciplinary Learning)이 필요할 것이라고 전망하고 있다. 로봇이나 기계를 다루는 전문적인 직업 노하우를 정보통신기술(ICT)과 접목할 수 있는 역량과 더불어 다양한 지식의 활용을 기반으로 소프트스킬(Soft Skills)이 미래사회에서 더욱 중요한 역량이 될 것으로 보고 있다(Boston Consulting Group, 2015).

〔그림 2-15〕 미래 산업 분야에서 요구하는 직무역량



자료: Technology at Work v2.0(Osford Univ., 2016), 김진하(2016) 재인용

직무역량과 더불어 자동화 또는 인공지능 등 기술 및 기계의 발전으로 노동력이 대체되더라도 창의성 및 혁신성 등과 같은 인간만의 주요 능력 및 영역은 자동화되지 않을 것으로 전망되고 있다. Mckinsey는 미국 내 800개 직업을 대상으로 업무활동의 자동화 가능성을 분석한 결과, 800개 중 5% 만이 자동화 기술로 대체되고 2,000개 업무 활동 중 45%만이 자동화될 것으로 분석하고 있다. 그리고 인간이 수행하는 업무 중 창의력을 요구하는 업무(전체 업무의 4%)와 감정을 인지하는 업무(전체업무의 29%)는 자동화되기 어려울 것으로 보고 있다(Mckinsey, 2015).

4. 데이터 기반 사회의 도래

기업의 의사 결정과 마케팅, 고객 분석과 대응, 새로운 제품과 서비스 개발에서 데이터의 중요성은 이제 더 강조할 필요가 없다. 인공지능을 기반으로 하는 지능정보산업이나 제4차 산업혁명이 국가 전략 과제의 중심이 되면서, 이제 데이터는 지능의 중심에 놓이는 가장 핵

심 자원으로 인정받고 있다.

알리바바의 마윈이 '세상은 IT 시대에서 DT(Data Technology: 데이터 기술) 시대로 가고 있다'는 인터뷰를 하고, 데이터 자본을 논하는 시대에 와 있다(한상기 2015. 3. 10).[12] 또한 데이터는 글로벌 경제의 각 분야에 격변을 가져오고 있다. 기업들은 자신들의 고객, 거래처, 조업에 관한 엄청난 양의 데이터들을 포착하여 거래 데이터로 쌓아놓고 있다. 수 억, 수 조개가 될 네트워크 센서들은 모바일폰, 스마트기기, 자동차, 그리고 산업용 기계나 군사 무기에 부착되어 사물인터넷 시대를 열어가고 있다. 사실상 기업은 물론이고 정부나 개인에 이르기까지 모든 활동들은 부산물로 데이터를 생산하고 있다. 예를 들어 지난 날 개인 간의 대화에 불과했던 사회적 교류의 산물들은 소셜네트워크사이트(Social Network Sites, SNS)에 고스란히 축적되고 그 결과물들은 페이스북, 인스타그램, 링크드인 등 플랫폼을 소유한 기업의 중요한 비즈니스 자원이 되어주고 있다. 그리고 점점 더 많아지는 멀티미디어 콘텐츠는 빅데이터의 기하급수적인 양적 성장에 중요한 역할을 하고 있다.

맥킨지에서는 빅데이터를 전형적인 데이터베이스 소프트웨어 툴이 포착하고 저장하고 관리하고 분석할 수 있는 역량을 넘어서는 데이터세트를 의미한다고 정의한다. 물론 이러한 정의는 대단히 주관적인 것인데 맥킨지는 이것이 의도적인 것이라고 덧붙인다. 이것은 데이터세트의 크기 자체가 기술 발전에 따라 커지기 때문에 빅데이터 개념 자체도 상대적이어야 한다는 것을 의미한다. 또한 일반적으로 사용하는 툴이 무엇인가에 따라 부문 별로, 또 산업별로 빅데이터의 정의는 달라질 수 있다. 벅 우디(Buck Woody)는 비슷한 맥락에서 "빅데이터는 동시에 작업할 경우 컴퓨터 연산 비용이 많이 드는 대규모 데이터세트"라고, 즉 당신이 가진 기술력으로 빠르게 처리하고 이용할 수 없는 데이터라고 정의한다.[13] 하지만 빅데이터는 무엇보다도 양적으로 크다는 개념을 넘어서며 속도와 다양성의 도전 앞에서 이에 응답할 수 있는 데이터를 의미한다.

중요한 것은 빅데이터라고 지칭되는 현상이 컴퓨터 과학의 발전으로 인해 가능해졌고 스

---

12) 알리바바 그룹의 마윈 회장은 2015년 베이징에서 개최된 빅데이터산업 설명회에서 "IT시대는 자기제어와 자기관리를 위주로 한다면 DT시대는 대중 서비스와 생산력 촉발을 위주로 하게 된다"며 "양자 사이에 특정한 기술적 차이가 있는 것처럼 보이지만 사실은 사고관념 측면의 차이"라고 갈파한 바 있다(연합뉴스, 2015. 6. 9).

13) http://sqlblog.com/blogs/buck_woody/archive/2011/10/18/big-data-and-the-cloud-more- hype-or-a-real-workload.aspx (검색일: 2016. 11. 20)

마트 사물(smart objects)과 연결된 지능(connected intelligence)을 통해 구현되는 빅데이터 기반의 네트워크 사회를 이전의 인터넷 기반의 정보사회와 구별 짓고 있다는 것이다. 이는 단순히 빅데이터가 글로벌한 현상이 되었음을 의미하는 데 그치지 않고 기업과 정부, 시민과 소비자의 삶과 의사 결정을 근본적으로 변화시키는 중요한 터닝포인트가 도래했음을 일깨워주고 있다. 인터넷이 검색이나 전자상거래 등 새로운 산업을 창출하면서 기업은 실시간으로 생성되는 데이터뿐만 아니라 비구조화된 데이터에 관심을 기울이기 시작하였다. 그 대표적인 기업이 구글인데 구글은 데이터 관리 및 활용에 있어서 패러다임 시프트를 가져온 것으로 유명하다(Forbes, 2016. 12. 18). 하지만 이러한 터닝포인트는 또 다른 측면에서는 우리가 알고 있는 프라이버시의 종말을, 플랫폼과 데이터를 소유한 기업으로의 권력 집중을 예고하고 있다.

5. 의사결정 시스템의 보편화에 따른 삶의 질 향상과 사회적 역기능

가. 삶의 모습·환경 변화

’16년 인공지능(알파고)과 인간(이세돌 9단)의 바둑 대결에서 지난 20년 동안 컴퓨팅 환경과 기술의 발전, 엄청난 양의 데이터를 통한 학습, 새로운 알고리즘 개발 등의 연구와 각 기업 및 정부의 끊임없는 투자와 노력으로 인공지능 기술의 성능이 급성장한 것을 확인할 수 있었다. 최근 구글, 페이스북, 아마존 등 글로벌 기업들을 중심으로 개인비서서비스, 자산관리를 위한 로보어드바이저, 암 진단 서비스, 법률 서비스 등 다양한 영역에서 인공지능 기술 기반의 서비스가 출시되고 있다. 나아가 이러한 인공지능에 대한 의존성이 확대됨에 따라 개인의 삶에 중요한 영향을 미치는 다수의 의사결정들까지도 이미 인공지능 알고리즘에 의해 이루어지고 있으며, 앞으로 이러한 현상은 더욱 확대될 것으로 예상된다.

우선, 인공지능 기술을 사용한 시스템 설계 및 시뮬레이션 활성화로 각종 서비스의 비용 감소와 품질향상으로 삶의 편의성 및 소비자 후생을 증대시킬 것이다.[14] 예를 들어 네덜란드 암스테르담에 세워진 ‘디엣지(The Edge)’는 친환경 스마트빌딩으로 건물 전 구역에 설치된 약 3만 개의 센서를 통해 사무실 직원 수, 실내외 온도, 냉난방 상황, 조명의 밝기 등의 정보

14) 관계부처 합동(2016. 12. 27)의 ‘지능정보사회 중장기 종합대책’에 따르면, 전 산업의 지능정보화로 인하여 국내의 경우, 신규매출 85조 원, 비용절감 200조 원, 소비자후생 175조 원 등 약 460조 원(2030년 기준)에 달하는 총 경제효과를 창출할 것으로 예상된다.

를 중앙서버에서 실시간으로 수집, 분석하여 건물 곳곳의 조명과 냉난방 스위치를 자동으로 조절하고 있다(매일경제, 2016. 10. 10). 또한 IBM의 인공지능 왓슨(Watson for Oncology)을 통한 암 진단 및 치료의 정확도가 90% 이상으로 향상됨에 따라 환자입장에서는 의료비용을 절감하고 고품질의 의료 서비스를 제공받을 수 있는 환경으로 변화되고 있다.

또한 지능정보기술의 활용을 통해 노동시간의 단축이나 재택근무, 원격근무, 스마트워크 등의 활성화로 여가시간이 증대하고 단순 반복적 업무로부터 해방되어 자발적 학습과 취미 활동 활성화로 일과 노동의 균형을 회복하는데 도움을 줄 수 있다. 노인・장애인・아동・저숙련 노동자 등 취약계층 및 빈곤계층에 대한 복지사각지대 예측을 강화하여 복지행정의 내실화도 가능하게 될 것이다.

나. 인공지능과 사회 윤리의 충돌

인공지능 중심 기술이 지식과 규칙 기반에서 딥러닝을 활용한 새로운 학습의 시대로 넘어가면서 최근에는 학습 데이터의 품질과 공정성이 인공지능 시스템이 가질 수 있는 편견과 왜곡 또는 차별의 문제를 야기할 수 있다는 논의를 불러왔다(Byrbes, 2016. 3. 28).

시스템 내부의 의사 결정 과정에 대한 로직은 과거에는 저장된 지식과 추론 규칙을 통해 파악이 가능했다. 그런데 소프트웨어가 전 산업 분야의 기반이 되고, 자동화가 진행되면서 수많은 데이터를 학습한 인공지능 시스템은 이제 그 결정 과정에 대한 면밀한 조사와 분석이 어렵게 되었다. 때문에 학습 과정에 사용한 데이터와 학습 알고리즘에 대한 사전 검토와 분석이 선행되어야 한다.

자동화된 시스템이 어떤 의사 결정을 내리는 경우, 지금까지는 인간이 의사 결정 과정에 존재하거나, 최종 의사 결정자였다. 그러나 자동화의 수준이 고도화되면서 암묵적으로, 또는 인간이 의도적으로 배제된 환경에서 기계에 의한 판단이 과연 인간 사회의 가치와 일치하는 것인가를 확인해야 할 필요가 생겼다.[15] 예를 들어 다음의 경우와 같은 문제가 발생했을 때 곧바로 대두되는 윤리적 문제, 즉 로봇의 판단이 사회에 반하는가 하는 문제를 생각해보자. 범인이 숨어 있는 집에 다른 시민이 있음에도 불구하고 드론이 사격할 수 있는가? 자율주행 자동차는 보행자를 피하기 위해 다른 차와 충돌하거나 승객을 위험에 빠지게 할 수 있는가?

15) 이코노미스트의 로봇 윤리 특별 섹션에서는 2012년 이런 문제를 다음과 같은 상황으로 예시하고 있다(Economist, 2012. 6. 2).

재난 구조 로봇은 사람들이 패닉에 빠질 수도 있음에도 현재 상황을 정확하게 알려줘야 하는가? 이러한 논의는 자연스럽게 인공지능 시스템의 윤리적 판단이 이루어질 수 있는지 또는 기계는 도덕적인 인공행위자(Artificial Moral Agent, AMA)가 될 수 있는가에 대한 토론을 불러왔다.

특히 자율주행자동차의 실용화가 가까운 미래에 이루어질 수 있다는 상황은 기계의 판단이 사회의 윤리적 판단이나 가치 시스템과 일치할 수 있는가에 대한 학술적, 사회적, 법률적 논의의 시발점이 되고 있다. 결국 이러한 기술 구현이 어떻게 이루어지고 누가 구현된 결과를 검증할 것인가 라는 문제는 엔지니어의 몫으로만 남을 수 없음이 명백해졌다.

6. 인간-로봇 공존을 위한 규범(윤리·법) 환경의 변화

가. 지능정보기술의 윤리 및 안정성 확보를 위한 논의 가속화

제4차 산업혁명 시대의 새로운 성장 동력으로서 인공지능, 로봇 기술의 발전 및 관련 산업 진흥 못지않게 중요한 것이 스스로 판단하고 의사결정을 내리는 로봇에 대한 사회적 수용과정에서의 윤리적 논의이다. MIT Technology Review(2015)는 영국 철학자 필리파 푸트(Philippa Foot)가 제시한 '트롤리 딜레마(Trolley Dilemma)' 상황을 자율주행자동차에 적용 가능하도록 3가지 경우[16]로 나누어 설문조사를 실시하였다. 그 결과, 응답자 대부분은 희생자를 최소화하도록 자율주행자동차를 만들어야 한다고 답하면서도 탑승자의 안전이 우선적으로 보장되지 않는 자율주행자동차에 탑승하기를 거부했다. 이는 인공지능, 로봇기술의 상용화에 앞서, 로봇이 직면한 돌발 상황에서 로봇이 어떠한 선택이든 내려야 할 경우 로봇이 어떠한 선택을 내려야하는지, 로봇이 내린 결정이 과연 옳은 선택이었는지, 옳은 선택이었다는 기준은 무엇인지, 로봇의 의사결정에 따라 희생자가 발생한 경우 그 책임을 누가 부담할 것인지 등의 다양한 윤리적 이슈에 대한 충분한 사회적 합의가 필요하다는 것이다.

이러한 인공지능 또는 로봇의 윤리담론에서 대표적으로 언급되는 사례가 아이작 아시모프(Isaac Asimov)의 '로봇의 3원칙'이다. 이 후에도 '로봇이 초지능을 바탕으로 스스로 추론

16) (1) 직진을 하는 경우에는 여러 명의 보행자를 치게 되고, 방향을 꺾으면 한 명을 치게 되는 상황, (2) 직진을 하는 경우에는 한 명의 보행자를 치게 되고, 방향을 꺾으면 자동차 탑승자가 위험한 상황, (3) 직진을 하는 경우에는 여러 명의 보행자를 치게 되고, 방향을 꺾으면 자동차 탑승자가 위험한 상황

하고 의사결정을 내리는 행위의 주체가 될 수 있는가?'에 관하여 '트롤리 문제(The Trolley Problem)'[17], '차이니즈 룸 논쟁(Chinese room Argument)'[18], '도덕적 튜링 테스트(Moral Turing Test)'[19] 등의 다양한 논의가 전개되었다. 최근에는 본격적으로 로봇공학자의 윤리, 규칙기반의 인간 친화적 인공지능 시스템 구축, 나아가 인공지능, 로봇자체의 윤리적 판단과 책임 등 로봇에 대한 윤리적 접근방식이 다양해지고 있다.[20] 또한 윌러치・알렌(2014)은 '인공적 도덕 행위자(AMA: Artificial Moral Agent)'의 개념을 통해 의도하지 않아도 도덕적 의사결정을 내려야하는 기계나 소프트웨어가 도덕적 판단을 할 수 있도록 알고리즘 설계단계에서 도덕과 윤리기준을 프로그래밍 하는 방법론(윌러치・알렌(노태복 역), 2014; 한상기, 2015)을 제시하고 있다. 한편, 미국의 인공지능 전문가 엘리저 유드코프스키(Eliezer Yudkowsky)는 로봇이 인간에게 우호적으로 행동하도록 설계되어야 한다는 '우호이론(friendliness theory)'을 제시하였다. '17년 1월 인류미래연구소(FLI: Future of Life Institute)는 인간에게 유용하고 혜택을 주는 착한 인공지능(beneficient AI)을 개발해야 한다는 이른바 '아실로마 AI원칙(ASILOMAR AI PRINCIPLES)'을 도출하기도 하였다.[21]

나아가 인공지능, 로봇기술이 군사용, 해킹 등 이른바 '치명적 자율무기 시스템(LAW:

---

17) 트롤리 딜레마(Trolley Dilemma)란 영국의 철학자 필리파 풋(Philippa Foot)과 미국의 철학자 주디스 자비스 톰슨(Judith Jarvis Thomson)이 고안한 윤리학 분야의 사고실험으로, 다섯 사람을 구하기 위해 한 사람을 죽이는 것이 도덕적으로 허용 가능한지에 대한 질문이다(두산백과).

18) 차이니즈 룸 논쟁(Chinese room Argument)이란 존 설(John Searle)이 튜링 테스트로 기계의 인공지능 여부를 판정할 수 없다는 것을 논증하기 위해 고안한 사고실험으로, 인간 또한 외부에서 접한 자극의 반응으로 습득된 기억속에서 대응방식을 도출하여 표현하게 되는데 이러한 골자를 그대로 따라한 중국어 방의 핵심인 '영어만 할 줄 아는 사람' 이 '인간의 지능' 과 같다고 볼 수 없다는 것이 중국어 방 논변이 갖는 논리적 바탕이다(위키백과).

19) 도덕적 튜링 테스트(Moral Turing Test)란 로봇에 인간의 도덕성을 주입하여 도덕적 기준을 갖추기 위한 실험으로 도덕적 튜링 테스트 설계에 있어 하향식 방법과 상향식 방법 그리고 최상의 장점을 결합한 하이브리드 방식이 있다(한상기, 2016).

20) 이와 관련하여 고인석(2014)은 로봇윤리에 대한 논의가 ① 로봇을 설계, 제작, 관리하는 공학자의 관점과 ② 로봇이 실현하는 행위의 도덕적 지위와 함의를 분석하는 윤리학자 등 두 가지 관점에서 접근하고 있으며, 한상기(2016)는 닉 보스트롬과 엘리저 유드코프스키가 '인공 지능의 윤리' 에세이에서 제시한 ① 로봇 공학자의 전문가적 윤리, ② 로봇 안에 프로그램된 '모럴 코드', ③ 로봇에 의해 윤리적 추론이 이루어질 수 있는 자기 인식 능력을 의미하는 로봇 윤리에 추가로 ④ 사용자 윤리까지 고려해야 한다고 말한다.

21) '아실로마 AI원칙(ASILOMAR AI PRINCIPLES)'의 구체적인 내용은 '진보넷 홈페이지(http://act.jinbo.net/wp/29625/)'에서 확인 가능하다.

Lethal Autonomous Weapon system)'에 활용되면서, 인간의 생명·신체에 해를 가하거나 인간 고유의 기본권 침해 가능성에 대한 우려가 지속되고 있다. 이에 따라 윤리적 차원의 논의를 넘어 법·제도적 대응의 필요성도 점차 높아지고 있다. 뿐만 아니라 자율주행자동차가 상용화 될 경우, 운전자 및 운행자를 기준으로 규정된 현행 자동차 관련 법령, 사고발생 시 책임주체 그리고 보험제도 등 일상생활 영역에서도 다양한 법적 환경의 변화도 함께 요청되고 있다.

특히, 유럽연합(EU)의 경우 로봇, 인공지능 기술의 발전에 따라 야기되는 다양한 윤리적·법적 이슈에 대한 규범적 차원에서의 '법제화' 움직임이 가장 활발하다. 우선, 로봇기술연구단체인 유럽로봇연구네트워크(EURON: European Robotics Research Network)가 '07년 4월 이탈리아 로마에서 개최된 국제로봇자동화학회(ICRA: IEEE International Conference on Robotics and Automation)에서 발표한 'EURON 로봇윤리 로드맵(The EURON roboethics roadmap)'을 시작으로, EU 집행위원회는 '12년부터 2년 동안 진행한 '로봇법 프로젝트(RoboLaw Project)'의 결과물로서 '14년 9월에 '로봇규제에 관한 가이드라인'(Guidelines on Regulating Robotics)을 제정하였다. 최근 '16년 5월에는 유럽의회 법사위원회(Committee on Legal Affairs)가 '로봇법 규칙 초안을 위한 보고서(Draft Report)'을 위원회에 제출하였고, '16년 10월에는 영국하원 과학기술 상임위원회는 로봇과 인공지능의 윤리적·법적 이슈를 다루는 '로봇공학과 인공지능(Robotics and artificial intelligence)' 보고서를 발표하는 등 유럽연합 차원에서 로봇, 인공지능 기술이 야기할 윤리적·법적 영향을 고려한 로봇규제의 근거를 마련하기 위한 노력이 한창 진행 중이다. 특히, 최근에 발표된 '로봇법 규칙 초안을 위한 보고서(Draft Report)'는 유럽연합 차원의 로봇법 법제화 방향을 담고 있을 뿐만 아니라, 유럽연합 회원국을 포함하여 인공지능, 로봇 등에 관한 법제화를 시도하는 각국에게 지침으로 작용할 수 있다(Committee on Legal Affairs, 2016). 전자적 인격(electronic person) 부여, 로봇 등록제 도입, 로봇기술헌장 마련, 로봇기술 규제기구 창설, 지적재산권 및 개인정보의 보호, 로봇기술 표준화, 법적 책임 등 로봇법 관련 법제화 과정에서 고려해야 할 원칙과 주요내용도 구체적으로 제시하고 있다는 점에서 중요한 의미를 갖는다.

이 밖에도 미국의 백악관은 '16년 5월부터 7월까지 5회에 걸친 워크숍을 통해 '인공지능의 미래를 위한 준비(Preparing for the Future of Artificial Intelligence)' 보고서를 10월에 발표한데 이어 12월에는 '인공지능과 자동화가 경제에 미치는 영향(Artificial Intelligence, Automation

and the Economy)' 보고서를 발표하였다. 또한 스탠포드 대학은 '16년 9월에 발표한 '인공지능과 2030년의 삶(ARTTFICIAL INTELLIGENCE AND LIFE IN 2030)'보고서를 통해 교통, 가정/서비스 로봇, 의료, 교육, 빈곤지역, 공공안전・보안, 고용・직장, 엔터테인먼트 등 8대 분야에 미칠 인공지능의 사회적 영향을 분석하였다. 이를 통해 규제와 정책을 통한 인공지능의 역기능에 대비할 필요가 있음을 강조하였다. '16년 12월 국제전기전자기술자협회(IEEE: Institute of Electrical and Electronics Engineers)에서도 인권, 책임, 투명성, 교육 등 4대 쟁점을 다룬 인공지능 윤리 지침서인 '윤리적인 디자인(Ethically Aligned Design)'을 발표하였다. 아울러 우리나라도 '07년에 로봇기술의 윤리적 발전방향과 로봇의 개발・제조・사용 시 지켜야할 윤리적 가치 및 행동지침을 담은 '로봇윤리 헌장' 초안을 발표한 바 있지만 더 이상 논의가 지속되지는 못하였다.

나. 인간중심 법체계 환경의 변화

최근 인공지능 기술의 비약적인 발전은 강한 인공지능(strong AI)의 출현을 굳이 상정하지 않더라도 현행 인간 중심의 법체계에 큰 혼란을 야기할 가능성이 높다. 물론 우리 「헌법」상 기본권의 주체와 「민법」상 권리와 의무의 주체, 「형법」상 형사책임의 주체 등은 모두 자연인(自然人)인 '사람'으로 한정하고 있으므로, 인간 중심의 현행법 체계상으로는 인공지능, 로봇에게 권리주체성을 인정하기는 어렵다. 다만, 기계 또는 알고리즘이 인간의 개입 없이 스스로 법률행위를 하거나 새로운 콘텐츠를 생성하고, 심지어 특정 영역에서는 인간을 넘어서는 행위결과를 만들어내는 사례가 증가하고 있다. 예를 들어 인공지능은 로봇저널리즘(기사작성), 투자자의 자산 운용(로보어드바이저), 영화, 음악, 미술 등의 다양한 영역에서 창의성 높은 결과물을 만들어내고 있다. 특히, 저작권 영역에서는 인간의 창작물과 인공지능이 만들어낸 결과물 사이의 경계도 모호해지는 등 창작행위가 더 이상 인간만의 독점 영역이 아닐 수 있다는 불안감이 커지고 있다. 따라서 인간과 유사하게 자율성을 바탕으로 생각하고 행동을 하는 인공지능에게 법적지위를 부여할 것인가, 인공지능이 권리와 법적책임의 주체가 될 수 있는지에 대한 고민이 필요한 시점이다.

물론 현재까지 인공지능에 의해 만들어진 대부분의 결과물들은 인간의 의도 내지 개입 하에 인공지능이 도구적으로 활용되어 왔다. 그러나 지속적으로 학습하고 지능화하는 인공지능 기술의 특성 상 인간이 관여되지 않는 상황에서 인공지능이 타인의 저작물을 이용하거나 스스로 만든 결과물에 대한 저작권을 어디까지 인정할 것인가에 대한 법률적인 검토가 필요

할 수 있다. 또한 인공지능이 만든 음악, 그림, 기사 등 창작물이 타인의 저작물을 침해하는 경우, 그 책임을 누가 부담할 것인지 등에 대한 사항은 현 상황에서도 중요한 쟁점이 된다. 예를 들어 로봇저널리즘이 타(他) 언론 기사의 내용을 무단으로 재사용하여 기사를 작성하거나 로봇저널리즘에 의해 작성된 기사가 오보일 경우, '정정보도' 청구의 대상 및 책임주체는 누가 될 것인지에 대한 보다 심도 깊은 논의가 필요하다. 마찬가지로 인간의 개입 없이 인공지능이 스스로 계약을 체결하는 경우, 자율성을 가진 인공지능에게 대리인으로서의 법인격을 부여할 것인지 그리고 만약 인공지능이 행한(법률)행위의 효과가 이용자 입장에서 전혀 의도하지 않거나 반할 경우 그 책임을 누구에게 귀속시킬 것인지 등에 대한 민사법적 쟁점도 제기될 수 있다.

이처럼 로봇, 인공지능 등 지능정보기술의 급속히 발전이 인간의 생활, 삶 전반에 큰 변화를 야기하고 있지만, 여전히 인간과 유사하거나 인간을 초월하는 수준으로의 기술 발전은 금명간 이루어지지 못할 것이라는 것이 중론이다. 확실치 않은 기술적 발전, 소위 '강한 인공지능(Strong AI)'을 예상하면서까지 현행 인간중심의 법체계를 바꾸는 것은 신중할 필요가 있다. 다만, 최근 유럽연합(EU) 의회에서 인공지능을 탑재한 로봇에게 '전자인간(electronic personhood)'이라는 특수한 형태의 법적지위를 부여하는 결의안을 채택한 바와 같이,[22] 적어도 인간과 동등한 법적지위는 아니더라도 인공지능을 가진 로봇이 인간에게 도움을 주는 존재로서 일정한 권리를 갖고 행동할 수 있는 특수한 형태의 법적지위를 부여하는 방안은 검토될 필요가 있다.

22) https://www.theguardian.com/technology/2017/jan/12/give-robots-personhood-status-eu-committee-argues

제3장 국내·외 인공지능 윤리관련 정책 및 규범 동향 분석

제1절 논의의 배경

빅 데이터와 인터넷 사물(IoT)은 인공지능 응용 프로그램 및 서비스가 성장할 수 있는 환경을 조성하고 있다. 금융 투자, 사이버 안보, 헬스케어, 운송, 스마트시티, 인재 선발 등 많은 분야에서 적용되기 시작한 인공지능은 인간에 의한 전통적 판단 방식을 빠르게 자동화하고 있으며 전례 없는 경제적 효율성을 안겨주고 있다. 한편으로는 그로 인해 초래되는 사회·윤리적 문제에 대한 관심도 커지는 추세이다. 이러한 배경에서 인공지능 산업육성 전략과 함께 윤리 가이드라인의 마련이 중요한 정책 이슈로 부상하였다.[23)]

2017년 이후 학계, 사적영역, 공공영역의 행위주체들(actors)은 인공지능 윤리 원칙과 정책 보고서를 발표하고 있다. 주요 국가들의 정부(의회), 대학 연구소, IT 기업 연합체, 싱크탱크, 민간 이니셔티브, ITU 및 OECD 등 국제기구, World Wide Web Consortium, IEEE 등 국제 표준화기구가 윤리·규범 정책 논의에 참여하고 있는 것이다. 인공지능 윤리를 둘러싼 논의는 2015년도에는 기계윤리 담론, 초지능(superintelligence)의 출현에 의한 인류의 절멸, 자동화로 인한 대량 실업 등이 주된 소재였으나 2017년 후반기부터 인공지능의 산업적 활용 과정에서 발생할 수 있는 실질적 문제로 초점이 옮겨가고 있다. 각 국가가 발표한 인공지능 전략 보고서는 인공지능 산업 진흥을 도모하고, 인공지능 생태계를 구축을 위한 전략에 초점을 두고 있다. 인공지능 윤리·규범 정책은 이러한 거시적 국가 목표를 달성하기 위한 제도적 정비 작업으로 여겨지고 있다.

해외 인공지능 윤리 관련 보고서들이 포함하는 초점들은 책무성(accountability), 편향적 판단(bias)의 교정을 위한 조사(auditability), 프라이버시 보호와 보안, 공정성(fairness), 자동화된 판단에 대한 설명(explanation), 위험의 관리(Managing AI risk), 특정한 집단만이 인공지능을 활용되어서는 안된다는 포괄성(inclusiveness) 등이다. 인공지능 윤리를 다룬 보고서들의 초

---

23) Darrell West, John Allen (2018) “How Artificial Intelligence is Transforming the World” Brookings Institution; Yale Law School Information Society Project (2017), “Governing Machine Learning,”

점들은 대동소이하지만 그 발표 주체가 누구인가에 따라서 인공지능 윤리을 접근하는 관점과 강조점이 달라지는 양상을 보이고 있다. 인공지능의 고도 발전에 따른 대중의 불안감과 우려가 과대 포장되었음을 지적하는 견해도 존재한다.[24] 그러나 궁극적으로 인공지능 산업이 국가경쟁력의 수준을 결정하고, 사회적 경제적 격변을 가져오게 될 것이므로 윤리적 논쟁점에는 대응이 필요하다는 점에는 세계적 공감대가 형성되고 있다.

한편 인공지능 연구개발(R&D) 분야에서 치열한 각축전을 벌이고 있는 글로벌 IT 기업들이 'Partnership on AI'을 구성하여 자발적으로 인공지능 관련 윤리 원칙을 밝히는 흐름도 주목할만 한다.[25] 구글, 딥마드, 마이크로소프트는 인공지능의 공익적 활용 사례들, 사회적 역할을 강조하고 있다. 글로벌 IT기업들이 이처럼 자발적으로 인공지능 윤리 원칙을 표명하는 이유는 무엇일까? 이는 산업계가 인공지능과 관련한 문제점들에 선제적으로 대응하여 '자율규제'(self-regulation)로 충분하다는 공감대를 얻기 위한 전략으로 보인다.

인공지능 윤리・규범 정책을 둘러싼 최근의 국제적 논의에서는 세 가지의 경향이 발견된다. 첫째, 2017년 이후 각국의 정부 및 의회는 국가 차원의 인공지능 산업 육성을 위한 기초적 정비하면서 인공지능 윤리・규범을 동시에 주목하고 있다. 이를 토대로 국가 경쟁력을 높이고, 장기적으로 지속가능한 인공지능 생태계를 구축한다는 전략이다. 만일 인공지능 윤리 논의가 국제적 공감대 및 합의 과정을 거쳐 규범적 효력을 지닌 '거버넌스' 준칙이나 기술 표준으로 자리잡게 된다면 국가들에게는 인공지능 전략에 긍정적・부정적으로 작용하게 될 것이다. 이처럼 인공지능 윤리・규범 논의는 인공지능 국가 전략과 인공지능 산업에 직접적 영향을 줄 수 있으므로 각 국은 글로벌 인공지능 윤리・규범 논의의 전개에 긴밀한 관심을 유지하고 있다.

둘째, 2018년 현재 다양한 글로벌 행위주체들이 발표한 인공지능 윤리 가이드라인들과 원칙들은 어떻게 '책임 있는 인공지능;(Responsible AI)을 구현할 것인가?"로 수렴된다고 할 수 있다. 그러나 각국 정부의 인공지능 전략은 세계적 각축전에서 우위를 차지하기 위한 움직이라는 의미가 크다. 예컨대 유럽 집행위원회, 프랑스 정부, 일본 정부의 인공지능 국가 전

24) Erik Brynjolfsson, Andrew McAfee (2014), The Second Machine Age. 참조

25) Partnership on AI https://www.partnershiponai.org 2016년 9월 페이스북, 구글, 마이크로소프트(MS), 아마존, IBM은 인류와 사회에 도움을 주기 위한 인공지능 파트너십(Partnership on Artificial Intelligence to Benefit People and Society)을 맺고 비영리로 인공지능 관련 사회 윤리적 문제를 연구하기로 하였다.

략은 인공지능 산업의 역동적 발전을 도모하여 투자를 증대시켜 국가 경쟁력을 강화한다는 지향점을 분명히 하고 있다.

셋째, 인공지능 윤리정책은 알고리즘에 공급하는 데이터에 대한 적절한 거버넌스를 위한 규제와 개인정보 보호와 밀접한 관계에 놓여있다. 2018년 5월 시행에 들어간 유럽연합의 일반개인정보보호규제(GDPR)는 각 국가별로 추진하는 인공지능 윤리 및 규범정책의 수립과 집행과정에 앞으로도 상당한 영향을 미칠 것으로 예상된다.[26)27)] 이러한 관련성을 주목하고 데이터 거버넌스에 초점을 두는 전략을 선택한 국가는 영국이다.[28)]

제 2 절 학계 및 민간연구기관차원에서의 인공지능 윤리 논의 동향

1. 하버드 버크만센터－MIT 미디어랩

가. 주요활동

하버드대 버크만 센터(Berkman Klein Center)와 MIT대 미디어랩(Media Lab)은 민간 및 공공 영역의 의사 결정자에게 지침을 제공하는 근거 중심의 인공지능 윤리 및 거버넌스 연구 프로젝트를 시작했다. 공공재로서 인공지능 사용을 강화하기 위한 인공지능 윤리 및 거버넌스에 초점을 맞춘 학제 간 연구자들의 네트워크를 구축하고 산업계, 학계, 시민 사회 및 정부 간의 관계 강화하고 있다.[29)] 협력 그룹은 네트워크화 된 협력을 연구하는 학자들의 학제 간 공동체로서 온라인상의 일련의 경험적 연구 프로젝트를 포함한다.

---

26) General Data Protection Regulation (EU) 2016/679

27) 유럽연합의 2차 법원(Secondary Source) 가운데 규제(Regulation)는 법적으로 그 구속력이 해당사항에 관하여 모든 회원국들에서 일반적으로 적용(general application)되기 때문에 규칙의 일부 조항에 대하여 효력발생의 유보를 둔다든가 예외를 둘 수 없다. qksais에 유럽연합의 지침(Directive)은 회원국의 입법조치로 인하여 “달성될 수 있는 결과에 대해서만(as to the result to be achieved)” 구속력이 있다. 김상만 (2018) “유럽연합(EU)의 입법영향 분석제도와 모범사례 연구” 국회입법조사처 (2013. 11. 3) 18면 참조

28) Information Commissioner UK,“Big data, artificial intelligence, machine learning and data protection” (Sept 2017) 참조

29) https://cyber.harvard.edu/topics/ethics-and-governance-ai

나. 인공지능 윤리와 거버넌스(Ethics and Governance of AI)

하버드대와 MIT가 진행하는 인공지능 윤리와 거버넌스(Ethics and Governance of AI)는 현재 다음의 세 가지 영역에 초점을 두고 프로젝트를 진행하고 있다.

- 자율성과 국가(Autonomy and the State): 정부는 인공지능을 통해 자동화된 기술의 소비자이자 규제자로서 역할을 수행한다. 인공지능은 인간의 자율성, 적법 절차 및 정의에 대한 기존의 생각을 어떻게 바꾸고 있는가?
- 자율성 및 플랫폼(Autonomy and the Platform): 어떻게 소셜 미디어 플랫폼에 적용된 인공지능 자동화 기술은 시민 생활에 영향을 미치고 민주적 규범을 재구성하는가?
- 대학의 역할(The University－A Countervailing Force): 점차 더 자동화되고 있는 세계에서 인간의 자율성을 유지하기 위해 대학이 할 수 있는 공동체 구축 및 교육 노력은 무엇인가?

위와 같은 인공지능 윤리 연구의 기본방향에 입각해서 두 연구소는 인공지능 윤리와 거버넌스를 구체화하기 위한 세부 프로젝트를 아래와 같이 추진하고 있다.

〈표 3－1〉 인공지능 윤리와 거버넌스의 세부 프로젝트[30)]

주요 프로젝트	주요 내용
AGTech 포럼	변호사, 학자 및 이해 관계자, 혁신 기업들이 인공지능에 대하여 협력하는 기회를 제공
인공지능 알고리즘과 정의	법원이 알고리즘을 사용하면서 인공지능의 편향성 공정성에 대한 질문이 제기되었으므로 이에 대한 탐구를 진행
인공지능 글로벌 거버넌스 및 포괄성	인공지능 경제가 성장하고 있지만 그 혜택에 불평등이 제기되므로 정책 입안자는 인공지능 시스템을 다룰 때 어려운 문제에 직면하게 되므로 이에 대한 연구를 진행
인공지능 미디어 및 정보 품질	자율 시스템이 온라인 컨텐츠를 선택하는 역할이 증가함에 따라 인공 지능이 인간의 판단, 의견 및 인식에 미치는 영향을 탐구할 필요가 있음
인공지능의 투명성 및 설명 가능성	인공지능 시스템에 책임을 지우는 방법은 무엇인가? 인간이 이해하고 알아볼 수 있는 정보를 인공지능이 제공할 있는가에 중점을 두고 있음.

30) https://cyber.harvard.edu/projects-tools

2. 옥스퍼드대 미래의 삶 연구소(Future of Life Institute)

가. 주요활동

오늘날 사용되는 인공지능은 특정한 역할을 수행하도록 설계되어 있기에 “제한된 AI” (narrow AI)에 해당한다. 인공지능 시스템이 우리의 자동차, 비행기, 심장 박동기, 자동거래 시스템 그리고 전원장치를 제어한다면 인공지능의 안전한 작동은 중요하다. 또한 치명적인 자동화 무기에 대한 개발 경쟁을 막는 것도 인공지능의 안전에 중요하다. 장기적으로는 인지 작업에서 인간보다 우월한 “포괄적 AI” (general AI)가 출현하여 인간 존재를 위협하는 경우의 위험에 대비하는 것이다.

나. 인공지능의 위험 시나리오

미래의 삶 연구소는 인공지능 위험 시나리오는 구체적으로 밝히고 있다. 인공지능은 그 어떤 인간보다 더 똑똑해질 가능성이 있으므로 인간의 지능으로는 인공지능이 어떻게 행동할지 예측할 수 없다. 인공지능이 위험 요소가 될 수 있는 시나리오 가운데 가장 가능성이 높은 두 가지는 다음과 같다.[31)]

인공지능이 치명적 작업을 하도록 프로그램되어 있는 경우: 예를 들어 인공지능 시스템이 장착된 군사용 자율무기들은 인명을 살상하도록 프로그램 되어 있으므로 대량의 사상자를 쉽게 유발할 수 있다. 더욱이 인공지능 무기 경쟁이 가속화된다면 전쟁이 발발했을 때 인간은 이미 배치된 자율적인 무기들을 통제할 수 없게 될 수 있다. 이러한 위험은 인공지능의 자율성 수준이 증가함에 따라서 커지게 된다.

인공지능이 유익한 작업 목표 달성을 위해 파괴적인 방법을 사용하는 경우: 인공지능이 작업하는 목표를 인간의 가치기준과 완벽하게 일치(align)하지 못하였을 때 발생한다. 예를 들어 자율 자동차에게 공항으로 가능한 가장 빨리 데려다줄 것을 요구한다면 과속과 교통법규 위반이 정당화되는가? 인공지능이 다른 고려요인들을 스스로 고려하여 판단하는가? 만약 초지능 인공지능이 지구 환경보존 프로젝트에 투입된다면 생태계를 파괴하는 부작용을 일으키는 인간을 목표 달성에 대한 위협으로 받아들일 수 있다.[32)]

---

31) Benefits & Risks of Artificial Intelligence - Future of Life Institute https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/

32) Ibid

다. 아실로마 인공지능 원칙 (Asilomar AI Principles) 공개[33)]

미래의 삶 연구소(FLI)가 개최한 '유익한 인공지능 콘퍼런스'(Beneficial AI Conference)는 학계와 산업계의 인공지능 연구자 그룹을 모으고 유익한 인공지능에 전념하기 위해 5 일 동안 경제, 법률, 윤리 및 철학 분야의 지도자들과 토론을 진행한다.[34)] 미국 캘리포니아주 아실로마에서 2017년에 합의된 인공지능에 대한 아실로마 원칙들은 인공지능의 힘이 미래에 어떻게 모든 사람의 삶을 개선하는 데 사용될 수 있을 것인가를 대중들이 이해하고 생각해 볼 수 있는 계기를 제공한다.

아실로마 원칙은 IT 기업들이나 국제기구 등이 유사한 인공지능 윤리 원칙을 고찰하도록 유도하는 기폭제의 역할을 하였다고 평가된다. 아실로마 인공지능 23가지 원칙 가운데 윤리와 가치(Ethics and Values)와 직결된 항목들은 16가지이며 최근에 산업계에서 실제로 제기되는 인공지능 윤리 이슈들은 대부분 이하의 항목에 긴밀하게 관련되어 있다.[35)]

- 안전(Safety): 인공지능 시스템은 작동 수명 전반에 걸쳐 안전하고 또 안전해야 하며, 적용 가능하고 실현 가능할 경우 그 안전을 검증할 수 있어야 한다.
- 작동 실패의 투명성(Failure Transparency): 인공지능 시스템이 실패, 오작동 일으킬 경우에는 그 이유를 확인할 수 있어야 한다.
- 사법적 투명성(Judicial Transparency): 사법제도에 인공지능 시스템이 사용된다면, 제3자의 검증과정이 필요하며 자동화된 판단에 대하여 설명을 제공할 수 있어야 한다.
- 책임(Responsibility): 인공지능 시스템의 디자이너와 설계자는 인공지능의 사용, 오용 및 행동의 도덕적 영향에 관한 이해관계자이며, 책임과 기회를 가져야 한다.
- 가치 정렬(Value Alignment): 고도로 자율적인 인공지능 시스템은 작동하는 동안 그의 목표와 행동이 인간의 가치와 일치하도록 설계되어야 한다.
- 인간의 가치(Human Values): 인공지능 시스템은 인간 존엄성, 권리, 자유, 문화적 다양성에 적합하도록 설계되어 운용되어야 한다.
- 개인정보 보호(Personal Privacy): 개인은 자기 데이터를 이용, 관리, 통제할 권리를 가져

33) Asilomar AI Principles - Future of Life Institute https://futureoflife.org/ai-principles/
34) Future of Life Institute, Annual Report 2017 https://futureoflife.org/wp-content/uploads/2018/02/FLI-2017-Annual-Report.pdf?x93895
35) Asilomar AI Principles, Future of Life Institute https://futureoflife.org/ai-principles/

야 한다

- 공동이익(Shared Benefit): 인공지능 기술은 최대한 많은 사람에게 혜택을 주어야 한다.
- 공동번영(Shared Prosperity): 인공지능에 의한 경제적 번영은 널리 공유되어야 한다.
- 인간에 의한 통제 (Human Control): 인간의 의사결정에 따라서 인공지능 시스템을 통제한다
- 인공지능 무기 경쟁(AI Arms Race): 치명적인 자동화 무기의 군비 경쟁은 방지되어야 한다.

라. 자동화 무기 개발 방지를 촉구하는 공개서한[36)]

미래의 삶 연구소(FLI)에 소속된 토비 월시(Toby Walsh) 스튜어트 러셀(Stuart Russell) 등이 주도한 자동화 무기 개발 방지를 촉구하는 공개서한은 2015년에 열린 IJCAI (Joint Conference on Artificial Intelligence)에서 공개되었다. 이하는 서한의 전문에서 주요 내용을 발췌한 것이다.

"자율 무기들(Autonomous Weapons)은 인간의 개입이 없이도 표적을 선택하고 교전을 시작합니다. 자율 무기의 범주에는 사전에 입력된 판단기준(pre-defined criteria)을 충족하는 대상을 검색하여 제거하는 무장 쿼드코퍼(armed quadcopters)가 포함되지만 인간이 결정을 내리는 순항 미사일 또는 원격 무인 비행기(drone)는 포함하지 않습니다. 인공지능을 자율 무기 시스템에 사용하는 것은 합법적이지 않지만 몇 년 안에 실현 될 수있는 지점에 도달했습니다. 자율 무기는 화약, 핵무기에 이어서 세 번째 혁명으로 묘사되기도 합니다.

예를 들어, 인간 병사를 기계로 대체하면 사상자를 줄일 수 있지먼 전투가 시작되는 한계점을 낮추게 된다는 문제가 있습니다. 오늘날 인류가 관심을 가져야 하는 핵심 질문은 세계적 인공지능 군비 경쟁을 시작하지 못하게 막는 것입니다. 어떤 주요 군사력이 인공지능 무기 개발을 추진한다면 세계 무기 경쟁은 사실상 피할 수 없으며이 기술 궤적의 종말점은 명백합니다. 자율 무기는 카라시니코프(AK-47) 소총처럼 보편적인 무기가 될 수 있습니다. 핵무기처럼 비싸지 않거나 구하기 힘든 원자재도 없기 때문에 모든 중요한 군대에 자율 무기들이 사용될 것입니다. 자율 무기가 암시장과 테러리스트들의

36) Open Letter on Autonomous Weapons, Future of Life Institute https://futureoflife.org/open-letter-autonomous-weapons/

손에 들어가는 것은 시간 문제 일뿐입니다. 자율 무기는 암살, 국가 불안정화, 특정 종족 집단 살해와 같은 일에 이상적으로 사용될 수 있습이다. 따라서 군사용 인공지능 무기 경쟁은 인류에게 유익하지 않을 것이라고 생각합니다. 인공지능으로 사람을 죽이기 위한 새로운 도구를 만들지 않고도 전장을 인간, 특히 민간인에게 더 안전하게 만들 수있는 다양한 방법이 있습니다.

대부분의 화학자와 생물 학자들은 화학 무기나 생물학 무기에 관심이 없듯이 대부분의 인공지능 연구자들도 인공지능 무기의 제작에 관심이 없습니다. 다른 연구자들이 인공지능 무기를 만들게 된다면 인공지능에 대한 대중의 반발을 불러일으킬 수 있습니다. 요약하자면, 우리는 인공지능이 여러 측면에서 인류에게 유익하게 사용될 큰 잠재력을 가지고 있다고 믿습니다. 군사용 인공지능 무기 경쟁은 나쁜 생각입니다. 인간의 통제를 벗어나는 공격적인 자율 무기는 금지되어야 하고 예방되어야 합니다.”

3. 프린스톤 인공지능과 윤리를 위한 대화(Princeton Dialogues on AI and Ethics)

가. 주요활동

프린스톤대학의 인간가치연구소(UCHV)와 정보기술 정책센터(CITP)는 공동으로 인공지능 시스템의 개발이 사회에 사용되는 경우에 발생할 수 있는 윤리적 딜레마에 대한 연구를 진행하고 있다.[37] 이 프로젝트의 목표는 윤리적 프레임 워크를 위한 지적 추리 도구를 개발하는 것이다.

나. 인공지능 윤리논쟁 사례연구

인공지능과 윤리의 교차점에서 존재하는 논쟁점들을 토론하기 위한 가상 사례연구를 개발한다.[38] 인공지능과 관련한 윤리적 사례 연구는 2017～18년에 진행된 프린스턴 대학의 학제 간 워크샵 시리즈에서 개발되었다. 알고리즘을 사용과 관련하여 충분히 발생할 수 있는 난제적 상황들을 설정하고 윤리적 질문들을 제기함으로써 인공지능 알고리즘의 윤리, 데이터 윤리에 다양한 공론화를 유도하고 있다.

- 사례연구 1. 자동화된 건강 관리 프로그램(Automated Healthcare App)

37) Princeton Dialogues on AI and Ethics https://aiethics.princeton.edu/

38) https://aiethics.princeton.edu/case-studies/

- 사례연구 2. 동적 사운드 식별 문제 (Dynamic Sound Identification)
- 사례연구 3. 학교 문제 해결을 위한 위한 개인정보의 사용 (Optimizing Schools)
- 사례연구 4. 법 집행 챗봇(Law Enforcement Chatbots)

4. 캠브리지대 실존적 위기 센터(The Centre for the Study of Existential Risk)

가. 주요활동

영국 캠브리지대의 실존적 위기센터(CSER)는 2012년에 설립되었으며 인류의 존재를 위협할 수 있는 합성생물학, 극단적 기후변화, 분자 나노기술 등을 연구하고 있다. 최근에는 인공지능의 오용이 초래하는 위험을 주목한다.[39] 인공지능이 인류를 어떻게 위협하고 불확실한 미래로 이끄는가에 대해서 철학자, 과학자 및 소프트웨어 엔지니어가 연구를 진행한다.

나. ‘해악적 인공지능 보고서’(Malicious AI Report)의 발행

캠브리지대 실존적 위기 센터는 미래의 삶 연구소(FLI), OpenAI, 뉴아메리카 안보센터, 전자 프런티어 재단(EFF)과 공동으로 ‘인공지능의 해악적 사용: 예측, 예방, 완화’ 보고서를 작성하여 2018년 2월 공개하였다.[40] 이 보고서는 고의적으로 인공지능을 악용하는데서 발생하는 보안 위협을 주목하고 있다. 인공지능을 이용한 공격 등 위협의 특성, 위협의 변화 양상을 진단하고 위협을 예방・완화하기 위한 방안을 제시하고 있다.

1) 인공지능으로 인한 위협(landscape of threats)의 변화

- 현존하는 위협의 확장: 효율적 인공지능 시스템을 사용하게 되면서 공격 행위자, 공격 속도, 공격 대상이 확대되었다. 인공지능의 익명성은 주저함을 사라지게 만들었고, 환경적・시간적 제약이 소멸되었으므로 비용편익에 대한 고민 없이 공격이 가능하다.[41]
- 새로운 위협의 등장: 인공지능을 사용하거나, 인공지능 시스템의 숨은 취약점을 찾아

---

39) The Centre for the Study of Existential Risk https://www.cser.ac.uk/

40) Miles Brundage et al., “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation” (Feb. 2018) Executive Summary pp.2~6; 인공지능 악용에 따른 위협과 대응 방안" Special Report 2018-12 한국정보화진흥원 참조 (2018. 8. 3.)

41) Ibid

내 공격하는 방식은 종래에는 존재하지 않던 위협들이다. 인공지능이 적용된 사회 인프라의 취약점을 공격하는 형태가 등장하고 있다. (ex: 가짜 뉴스(fake news)의 생산과 전파, 알고리즘에 오류를 일으켜 자율주행차 사고유발, 인공지능 알고리즘에 의한 음성의 모방과 허위 음성의 합성)

- 전형적인 위협 특성들의 변화: 인공지능을 이용한 공격은 익명성으로 인하여 행위자를 찾아서 책임을 귀속(attribute)시키는 작업은 극히 어려워지고 있다. 이같은 환경의 변화에 기인하여 공격의 빈도와 규모가 커졌으며 정밀한 표적화 공격이 가능해졌다.

2) 인공지능에 의한 보안 위협 시나리오

- 디지털 보안 (Digital Security): 알고리즘 코드의 취약점을 자동으로 발견하여 악용 코드를 생성, 인공지능 시스템의 블랙박스 모델 추출, 공격대상 선택 및 우선순위 선정 방법의 향상, 기계학습을 통하여 사이버 공격 대상의 우선순위 지정 등이 가능하다.
- 물리적 보안 (Physical Security): 인공지능이 적용된 물리적 대상을 공격하는 수단은 방어용 기술보다 빠르게 발전하여 비대칭이 심화되고 있으므로 정책적 개입이 필요한 분야로 여겨진다. (ex: 악용을 위해 인공지능 시스템에 부여된 목적을 변경, 로봇을 해킹하여 원격 조정, 자동화된 스웜(swarm)공격 수행)
- 정치적 보안 (Political Security): 소셜 미디어 플랫폼에서 봇(bot)을 이용하여 정치적 프로파간다의 실행이 증가하는 경향을 보이고 있다. 인공지능을 사용하여 콘텐츠를 자동적으로 생성하고 게시하는 방식이 확대되고 있다. (ex: 인공지능을 이용하여 발언을 조작하는 가짜 비디오(Deepfakes)가 확산되지만 진위를 가리기 어려움

3) 인공지능 위협에 대응하기(intervention)

인공지능의 악용을 차단하고 대응하려면 다양한 이해관계자 간의 논의(multi-stakeholder dialogues)를 통한 공개 토론을 활성화할 필요가 있다. 또한 인공지능을 사용한 위협들에 대응하기 위하여 다음의 권고안들이 제안되었다.

- 사이버 보안 공동체에서 배우기: 인공지능을 이용한 공격과 사이버 보안의 교차점에 위치한 문제를 해결하려면 적색팀(red team)을 구성하여 실제 어떤 공격을 받을 수 있는지 검증을 진행하고, 인공지능의 취약점이 발견되면 책임감 있게 공개할 필요가 있다.
- 개방형 모델에 대한 조사: 인공지능과 머신러닝 기술의 이중적 용도가 분명해졌으므

로 인공지능 알고리즘의 개방성(openness)이 과연 적합한지, 어떤 방법으로 공개하는 편이 적절한지에 대한 규범과 제도를 재고할 필요가 있다.

- 책임 문화의 제고: 인공지능 연구개발 담당자들에게 인공지능이 악용될 수 있는 위험을 주의시키고, 윤리적·사회적 책임감을 일깨우는 교육이 중요한 의미를 가지게 되었다.
- 기술적, 정책적 해결책의 모색: 인공지능 악용을 예방하고 완화하기 위한 제도적, 정책적 방안으로는 개인정보 보호를 위한 기술적 조치, 공공 안전을 위해 방어적 목적으로 인공지능 사용하기, 악의적 공격을 규제하는 법제도 마련 등을 고려해야 한다.
- 규제를 통한 대응: 인공지능에 의한 공격은 그 피해가 국경을 넘어 확산될 수 있으므로 국제적으로 통용되는 범국가적 법규정을 마련할 필요가 있다. 그러나 아직 국제적 기준이나 규범은 마련되지 않고 있다. 한편, 입안자가 기술적 이해가 부족하다면 인공지능의 악의적 사용을 규제하기 위한 입법이나 대응책이 잘못 설계될 수도 있다.

제 3 절 국가별 정부 및 의회 차원에서의 인공지능 윤리관련 논의 동향

주요 선진국들은 국가역량을 결정하고 세계 경제의 패권을 좌우하게 될 인공지능의 영향력을 인지하고 인공지능 국가전략(National AI Strategy)을 준비하고 있다. 국가별 인공지능 정책은 인공지능 산업 진흥을 위한 투자를 늘이고, 관련 제도를 정비하여, 인공지능 생태계를 구축하려는 전략이 반영되어 있다. 아울러 인공지능 활용 과정의 부작용, 자동화가 초래하는 문제에 대응하기 위한 윤리 원칙 및 가이드라인도 포함하고 있다.[42)]

캐나다는 2017년 전 세계에서 처음으로 인공지능 국가전략 Pan-Canadian Artificial Intelligence Strategy을 발표하였다.[43)] 캐나다의 전략은 인공지능의 발전으로 말미암은 경제적, 윤리적, 정책적, 법적 함의를 고찰하고, 캐나다가 인공지능 논의를 선도할 수 있는 글로벌 리더십을 갖추어야 한다는 점을 언급하였다. 캐나다는 주로 연구 및 인재 중심의 전략을 강조하였다. 아랍 에미리트 연방은 인공지능부(Ministry of Art Intelligence)를 처음으로 만들었고

42) Matt Chessen “The AI Policy Landscape” Medium.com (Mar 31, 2017)

43) Canadian Institute for Advanced Research, ‘Pan-Canadian Artificial Intelligence Strategy’ https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy

정부의 성과를 높이기 위한 인공지능 활용에 중점을 두고 있다. 영국은 2018년 4월 '인공지능 분야 협력(AI Sector Deal)'에 이어 인공지능 산업을 육성하여 데이터 기반 혁신의 중심에 서겠다는 비젼을 담은 '웅대한 도전(Grand Challenges)'을 발표하였다.[44] 한편, 2017년 발표된 중국 정부의 '차세대 인공지능 발전계획(新一代人工智能發展規划)'은 미국 등 서구의 경쟁 국가들에게 상당한 도전을 던져주고 있다.[45] 중국의 목표는 2030 년까지 인공지능 기술을 통해 세계를 선도한다는 계획이다.[46]

미국, 이스라엘, 러시아는 아직 공식적 인공지능 국가 정책이나 전략을 내놓지 않고 있다.[47] 미국 국가과학기술위원회(NSTC)의 네트워킹·정보기술 연구개발 위원회(NITRD)가 2016년에 '국가 인공지능 연구개발 전략계획'(National Artificial Intelligence R&D Strategic Plan)을 발표한 바 있다.[48] 또한 오바마 대통령 집권기의 백악관 국가과학기술 자문위원회(NSTC)는 '인공지능의 미래 준비하기'(Preparing for the Future of Artificial Intelligence) 보고서를 내놓았으나[49] 이를 미국 정부의 공식 인공지능 국가 전략으로 보기는 어렵다.[50] 한편 유럽연합 회원국들이 2018년 4월에 발표한 '인공지능에 관한 EU 협력 선언'은 유럽이 미국과 중국의 인공지능 기업들과 경쟁하기 위해 공공 연구재원을 마련하고, 공공 분야 데이터의 활용을 위해 협력해야 한다는 점을 강조하였다.[51]

ABB와 이코노미스트지(Economist)는 자동화준비지표(Automation Readiness Index)를 발표

---

44) United Kingdom Govenment Policy paper 'AI Sector Deal' (26 April 2018); Policy paper 'The Grand Challenges' (21 May 2018)

45) 國務院關于印發, 新一代人工智能發展規划的通, 國發〔2017〕35号

46) China's Artificial Intelligence Strategy Poses a Credible Threat to U.S. Tech Leadership Council on Foreign Relations (4 December 2017)

47) Ganesh Bell "Why countries need to work together on AI" World Economic Forum (16 Sep 2018)

48) National Science and Technology Council Networking & Information Technology Research and Development Subcommittee www.nitrd.gov/news/national_ai_rd_strategic_plan.aspx

49) Executive Office of the President National Science and Technology Council Committee on Technology "Preparing for the Future of Artificial Intelligence" Obama White House (Oct 2016) https://obamawhitehouse.archives.gov/sites/.../preparing_for_the_future_of_ai.pdf

50) 이러한 관점에 대하여는 다음 기사를 참조 바람 John Dealaney "France, China, and the EU All Have an AI Strategy. Shouldn't the US?" The Wired (May 20, 2018)

51) 2018 년 4 월 10일 발표된 EU Declaration on Cooperation on Artificial Intelligence에는 유럽연합 소속 24개 국가들이 서명했다. 뒤이어 2018년 5월에는 4개국이 추가로 참여하여 28개국이 서명했다. https://ec.europa.eu/digital-single-market/en/events/digital-day-2018

했다. 이 지표는 인공지능과 로봇공학 등 고도 지능기술의 확산을 준비하는 상황을 국가별로 수치화한 것이다. 2018년도의 종합 순위에서 한국이 1위(91.3), 독일이 2위(89.6), 싱가포르가 3위(87.3), 일본이 4위(82.6), 캐나다가 5위(81.8)을 차지했다.[52] 우리나라가 아직 인공지능 국가 전략을 공식화하지 않았음에도 이처럼 높은 순위로 평가된 이유는 직업교육과 훈련을 위한 정부 보조금 지출 등 인공지능 경제에 대비하기 위한 공공영역 투자가 높게 평가받았기 때문이다.(GDP 대비 연구개발비 4.23% 투입).[53][54]

한편, 전체적으로는 자동화준비지표에서 상위에 위치한 국가들조차 인공지능 정책의 수립에 있어서는 여전히 초기 단계에 머물고 있다.[55] 이하에서는 인공지능 윤리・규범정책 논의를 선도하고 있는 유럽 집행위원회와 유럽 국가들을 위주로 검토하겠다.

1. 미국

가. 빅데이터 및 인공지능 알고리즘의 윤리적 이슈 논의

미국은 정부 차원에서 대대적으로 인공지능의 사회적 영향에 관한 연구를 진행하지만, 규범 수립은 주로 기업, 학계 등 민간 영역에서 이루어지고 있다. 다만, 오바마 행정부 당시 백악관 국가과학기술위원회를 중심으로 빅데이터 알고리즘 윤리 규범을 포함한 미래전략을 수립하는 가운데 인공지능 윤리 논의의 단초를 제공한 바 있다. 즉 2016년 3월 미국 백악관은 이른바 '빅데이터 알고리즘의 윤리적 가이드라인'으로도 알려진 ''Big Data: A Report on Algorithmic System, Opportunity and Civil Right' 라는 보고서를 통해 빅데이터 분석, 머신러닝 및 알고리즘 자체의 내부 작용에 의한 차별, 왜곡, 편견 제공 등의 위험성을 경고하였던 것이다.

또한 2016년 10월 백악관은 '인공지능 국가개발 연구전략', '인공지능의 미래를 위한 준비', 그라고 2016년 12월 '인공지능과 자동화가 경제에 미치는 영향' 등 일련의 3부작 보고

---

52) The Economist (2018) "The Automation Readiness Index: Who Is Ready for the Coming Wave of Innovation?" p.5

53) Ibid. p. 14

54) 세부지표에는 혁신환경(Innovation Environment)이 있으며 '윤리와 안전', '인프라', '연구와 혁신'이 평가항목이다. '윤리와 안전'은 '기술윤리와 안전제도'(Technology ethics and safety institutions) '사이버 보안 전략'(Cybersecurity strategy)이 구비되어 있는가를 측정하여 평가한다.

55) The Economist (2018) p.5

서를 통해 인공지능의 위험성 등에 대한 윤리적 대응을 강조하였다. 특히 미국 백악관 과학기술정책실은 학계, 비영리기구, 국가경제회의 등과 함께 인공지능 규범에 관한 공론장을 워크숍 형태로 운영하는 등 인공지능 윤리 이슈에 대한 사회적 논의를 활성화하고자 노력한 것으로 보인다.

〔그림 3－1〕 미 백악관 주재 인공지능 워크숍의 주요 내용

순서	주제	핵심 이슈 및 시사점	기술·연구·정책 사례
제1차 (2016.03.24.기술정책랩, 워싱턴 대학교 법과대학)	·인공지능의 법과 거버넌스 시사점 (Legal and Governance Implications of Artificial Intelligence)	·고용과 보험 등 여러 분야에 미칠 위험과 영향 고려 필요 ·자동차나 의료사고로 인한 사망 대폭 감소 가능성 높음 ·인공지능 제어 문제는 법률 전체에 문제가 될 수 있음 ·인공지능 알고리즘이 규범 문제를 포함하며, 동시에 규제정립에 유연성이 필요함	·워싱턴대학교 다학제적 기관인 '기술 정책 랩(Tech Policy)' 창설 ·구글 포토의 AI가 흑인을 고릴라로 본 것은 백인만이 인간이라는 가정이 개입된 오류
제2차 (2016.06.07. Computing Community Consortium)	·사회 이익을 위한 인공지능 (Artificial Intelligence for Social Good)	·도시 컴퓨팅(Urban Computing), 생태계 지속가능성(Environmental Sustainability), 건강(Health), 공공복지(Public Welfare) 논의 ·데이터를 통한 생태계 지속가능성 정책 도출 가능 ·건강과 공공복지와 관련 있는 의학에서 AI에 의존 가능성 있음 ·데이터 과학과 AI가 과도한 폭력(excessive force)을 예측한다는 부정적 의견	·오바마의 정밀의료계획(Precision Medicine Initiative)과 문샷 프로젝트(Cancer Moonshot)
제3차 (2016.06.28. 카네기멜론 대학)	·인공지능 안전과 제어 (Safety and Control for Artificial Intelligence)	·러닝 알고리즘 분석과 예측 모델을 바탕으로 한 '본질 안전(intrinsic approaches)' AI 알고리즘 ·수학 기반의 안전 모델과 추론 ·AI와 러닝 기능성을 통합한 시스템의 안전 보장 ·오작동 대응 알고리즘 개발 및 독립 정보를 주는 센서 배치	·원전 센서들로 즉각적인 위험요소 발견 가능
제4차 (2016.07.07. Artificial Intelligence Now)	·단기적 인공지능 기술의 사회적, 경제적 함의 (The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term)	·사회적 불평등, 노동, 금융시장, 헬스케어, 윤리 등 다섯 가지 주제의 향후 5-10년 내 변화 초점 ·인공지능 시스템이 인프라가 될 때 적법 절차와 공정성의 가장 높은 수준을 고려해야 함 ·특히 급변할 것으로 예상되는 노동시장에 대한 정책 필요	·범죄위험평가(criminal risk assessment) : 알고리즘으로 재범비율 산정하면 흑인은 더 높은 예측범죄율 도출 ·인공지능 연구 위한 컴퓨터 과학자 그룹 FAL MT

나. 자율주행 안전설계를 위한 12가지 가이드라인

미국은 자율주행자동차 운행과 관련해 각 주마다 시험운행 요건 차이가 있으나 일부 선제적 입법화를 추진하는 가운데 윤리적 가이드라인을 모색하고 있다. 예컨대 2017년 9월 미국 정부는 'A Vision for Safety2.0: Automated Drving Systems'를 통해 자율주행시스템의 안전, 운영설계 범위, 위기 상황 최소화 및 데이터 기록, 인증방법, 사이버보안, 소비자교육 및 훈련 등의 원칙 및 기준을 제시하였다.

다. 뉴욕시 회의의 알고리즘 코드 공개 의무화 법안 및 캘리포니아 주의 봇규제(Anti-Bots)

최근 미국에서는 뉴욕과 캘리포니아에서 윤리적 규범 보다는 인공지능과 로봇에 대한 법적 규제를 모색하는 움직임이 주목받았다. 뉴욕시 의회의 경우 지난 2017년 8월 인공지능 알고리즘의 법적 규제 일환으로 알고리즘 또는 자동화된 결정 시스템의 소스 코드 공개를 의무화하는 법안을 추진한 바 있다.

〔그림 3－2〕 뉴욕시 회의의 알고리즘 소스코드 공개 의무화 법안

또한 2018년 9월 28일 미국 캘리포니아 주는 봇(bot)을 통한 허위정보 유포 등을 최소화하기 위해 '봇'의 존재를 알리지 않고 온라인 상의 의사소통 및 상호작용을 위해 로봇을 사용하는 것을 금지하는 법안을 통과시켰다. 미국 최초의 로봇금지법으로 '블레이드 러너법'이라고보 불린다. 이 법은 2019년 7월부터 시행될 예정인데, 동 법률은 캘리포니아 주민을 상대로 봇을 사용한 의사소통 또는 상호작용임을 밝히지 않고 의도적으로 인공적 정체성을 오인시키는 경우를 불법으로 다룬다.[56] 동 법률의 규제 대상은 "상업적 거래에서 상품 또는 서비스의 구매 또는 판매를 촉진(incentivize)하거나 선거에서 투표에 영향(influence)을 주기 위해 의사소통의 내용을 의도적으로 속이는 목적"을 가진 행위이다. 봇을 이용하여 온라인 상

56) Senate Bill No. 1001 CHAPTER 892. An act to add Chapter 6 (commencing with Section 17940) to Part 3 of Division 7 of the Business and Professions Code, relating to bots.

거래를 위한 광고 등에서 영업주체의 혼동을 일으키는 사칭통용 행위(passing-off) 및 봇을 이용한 악의적 여론 조작을 규제하려는 시도라고 볼 수 있다. 봇(bots)은 이용자의 질문에 대답하고 채팅을 할 수 있고 질문에 대한 지능적인 대답도 가능하다.[57] 그러나 소셜미디어 서비스 트위터에 자동으로 게시물을 올리는 '트위터 봇(bot)' 계정은 스팸 게시물을 게시하거나 퍼뜨리는 용도로 사용되어 선거전을 혼탁하게 만들기도 하였다. 이러한 배경에서 2018년 2월 트위터 사는 "스팸 게시물을 활용한 가장 흔한 규정 위반 행위 중 하나는 특정 트윗을 인위적으로 늘리거나 많아 보이게 하기 위해 다수의 계정을 쓰거나 개발자 플랫폼을 악의적으로 이용"하는 행위를 금지한 바 있다.

2. 유럽연합

가. 인공지능에 관한 EU 협력 선언

1) 개요

유럽 집행위원회(European Commission)의 마리야 가브리엘(Mariya Gabriel) 집행임원은 2018 년 4월 10일 디지털 데이 (Digital Day)에 때를 맞추어서 '인공지능에 관한 EU 협력 선언'(EU Declaration on Cooperation on Artificial Intelligence)을 발표하였다. 이 선언은 인공지능 분야에 대한 유럽연합 소속 국가들의 기본적 합의를 담고 있다. 24개 국가들이 참여하였고 그 후 5월에 루마니아, 그리스, 사이프러스, 크로아티아도 서명하였고 노르웨이도 합류하기로 하였다.[58]

이미 다수의 유럽연합 회원국들이 인공지능에 관련한 국가 계획을 발표했지만, 유럽 연합 차원에서 공동적인 협력과 역량을 결집하기 위한 접근방향을 밝힌 것이다. 인공지능에 관한 EU 협력 선언은 회원국들이 인공지능의 기회를 확보하고 도전 과제들을 공동으로 대응하도록 하는 유럽 차원의 합의 형성이 핵심이라고 할 수 있다. 또한 인공지능 분야에서 유럽이 거둔 성과와 디지털 싱글 마켓 (Digital Single Market)의 진전을 토대로 삼고 있다.

---

57) Jim Martin "What are Bots and Chatbots?" Tech Advisor (26 Sep 2017)

58) EU Declaration on Cooperation on Artificial Intelligence의 내용 전문은 다음의 링크를 참조. https://docs.google.com/viewerng/viewer?url=https://ec.europa.eu/jrc/communities/sites/jrccties/files/2018aideclarationatdigitaldaydocxpdf.pdf

2) 주요 내용

‘인공지능에 관한 EU 협력 선언’의 주요 내용은 인공지능이 제공하는 기회를 최대한 활용하려면 유럽 국가들은 최고 수준의 데이터 보호를 포함하여 함께 협력해야 한다는 것이다. 유럽연합 회원국들은 다음의 사항들에 협력하기로 하였다. 첫째, 유럽 차원의 인공지능 분야 기술과 산업계 역량의 강화 및 활용을 진흥시키기 위해서는 공공 영역 데이터(public sector data)에 대한 더 나은 접근이 요구된다. 공공 영역의 데이터는 인공지능 분야의 연구개발에 영향을 미치고 혁신적인 비즈니스 모델을 촉진하고 경제 성장과 새로운 직업군을 창출하기 위한 필수적인 조건이기 때문이다. 둘째, 인공지능이 제기하는 사회 경제적 문제의 해결. 이는 인공지능의 활용으로 인한 노동 시장의 변화와 유럽의 교육 및 훈련 시스템의 현대와 등을 포함한다. 셋째, 프라이버시와 개인 정보 보호, 투명성, 책무성을 포함하는 인공지능 윤리 원칙이다. 인공지능의 시대가 도래한다고 해도 유럽연합의 기본적인 권리와 가치에 기초한 적절한 합법적이고 윤리적인 프레임워크가 보장되어야 한다는 취지이다.

3) 비개인적 데이터(non-personal data)의 활용

이러한 배경에서 유럽 집행위원회는 비개인적 데이터(non-personal data)를 자유로운 흐름을 촉진하고 e-health에 사용되는 의료용 데이터를 공유하기 위한 방법을 모색하고 있다.[59] 디지털 싱글 마켓을 담당하는 유럽 집행위원회의 안드루스 안쉽(Andrus Ansip)은 “디지털 싱글 마켓은 빠르게 형태를 갖추어 가고 있지만 데이터가 없다면 인공지능, 고성능 컴퓨팅 및 기타 기술들을 최대한 활용하기 불가능하다”고 밝히며 데이터의 현명한 사용을 강조했다. 그가 주장하는 발전 방향은 유럽 내에서 공공 영역의 데이터(public sector data)가 자유롭게 사용되어야 하며, 데이터 재사용(re-use)에는 상업적 목적도 포함된다는 것이다. 이를 통해서 유럽연합을 ‘공유 데이터 영역’(common data space)으로 만든다면 데이터 이용 비용이 크게 낮아져 경제 성장이 촉진될 것이라는 논리이다.[60]

유의할 점은 어디까지나 비개인적 데이터는 유럽연합 일반개인정보보호규제(General Data Protection Regulation, GDPR)의 준수가 요구된다. 인공지능 산업의 발전을 위하여 데이터 접

---

59) Data in the EU: Commission steps up efforts to increase availability and boost healthcare data sharing Brussels, (25 April 2018)
http://europa.eu/rapid/press-release_IP-18-3364_en.htm

60) “Communication: Towards a common European data space” (April 25, 2018)
https://ec.europa.eu/digital-single-market/en/news/communication-towards-common-european-data-space

근성 및 재사용 수준 개선을 위한 정책의 대상이 되는 것은 상업적 영역의 데이터 및 비상업적 목적을 위한 공공영역 기관(법률, 교통, 기상 및 금융 등)이 생성한 데이터이다. 이러한 유형의 데이터를 재사용하려면 'PSI 지침'으로 불리는 '공공영역 정보의 재사용에 관한 지침'(Directive 2003/98/EC)의 적용을 받는다.[61] PSI 지침은 시민이 어떤 공공영역의 정보에 접근하는 것보다 공공영역의 정보를 '재사용' 할 때의 경제적 측면에 중점을 두고 있다. PSI 지침에 따라서 유럽연합 회원국들은 재사용 할 수있는 많은 최대한의 정보를 제공하도록 장려된다. 공공 영역 데이터를 적극적으로 활용하도록 권장하고 '공유 데이터 영역'을 만들려는 유럽연합의 정책은 우리나라에도 시사점을 준다고 하겠다.[62]

나. 유럽을 위한 인공지능 (Artificial Intelligence for Europe) 정책

1) 개요

유럽 집행위원회는 유럽연합 국가들이 합의한 '인공지능에 관한 EU 협력선언'의 취지를 구체화하여 '유럽을 위한 인공지능'(Artificial Intelligence for Europe) 정책을 2018년 4월 25일 발표하였다.[63] 이는 인공지능이 제공하는 새로운 도전에 대처하는 동시에 그 기회를 최대한 활용하기 위한 본격적인 실행방안을 담고 있다.[64] 이러한 정책이 제시된 배경에는 지식노동의 자동화, 로봇, 자율 주행차량 등으로 인한 경제적 영향력이 2025년까지 해다마 6조~13조 유로(€)에 달할 것이라는 예측이 존재한다. 한편, 유럽은 북미 지역과 아시아 지역 국가들과 비교하여 2016년 기준으로 사적 영역의 투자가 2조 4천 억~3조 2천 억 유로(€) 뒤쳐져 있다.

유럽 집행위원회는 최근 유럽 회원국들의 '선언'을 이끌어내고 유럽을 위한 인공지능의 실행계획을 적극적으로 발표하였다. 무엇보다도 유럽지역에서 공공영역의 데이터의 이동과 재사용을 늘리고, 디지털 싱글 마켓을 진전을 강조하고 있다. 이러한 적극적 행보에는 미국

61) Directive 2003/98/EC on the re-use of public sector information

62) 월드와이드웹(WorldWideWeb) 재단이 2018년 발표한 공개 데이터 평가(Open Data Barometer)에서 영국과 캐나다가 공동 1위, 호주가 3위, 우리나라와 프랑스가 공동 4위에 랭크되었다. The Open Data Barometer (ODB) https://opendatabarometer.org/barometer

63) EU Commission (25 April 2018) "Artificial Intelligence for Europe" https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

64) A European approach on Artificial Intelligence www.europa.eu/rapid/press-release_MEMO-18-3363_en.htm

과 중국에 비해서 유럽의 인공지능 산업의 발전 속도가 밀리고 있다는 현실적인 위기감도 상당히 반영되어 있다.[65)]

〔그림 3－3〕 '유럽을 위한 인공지능' 정책의 3대 분야.[66)]

2) 주요내용

'유럽을 위한 인공지능'은 인공지능에 대한 접근방식을 세 가지로 나누었다. 첫째, 공공 영역의 기술 및 민간 영역의 인공지능 기술 개발과 활용의 촉진, 둘째, 인공지능이 초래한 사회 경제적 변화에 대한 대응, 셋째, 적절한 윤리적 및 법적 체계의 마련과 보장이 그것이다. '유럽을 위한 인공지능'도 유럽 연합 내에서 국경을 넘나드는 자유로운 데이터 흐름(free flow of data)를 포함한 디지털 싱글 마켓의 생성이 인공지능의 발전에 핵심적이라는 점을 강조하였다.

① 인공지능 연구개발 재정 지원 확대 및 공공 및 민간 영역의 활용 증진

유럽 연합의 공공 및 민간 영역은 'Horizon 2020 연구 및 혁신 프로그램'에 따라서 2020 년까지 인공지능 연구 및 혁신을 위한 투자를 200 억 유로 증가시킬 계획이고 2018 년에만 15 억 유로의 투자를 하고 있다.[67)] 이러한 투자는 인공지능을 활용한 운송에서 건강에 이르는 분야들의 개발을 지원할 예정이다. 유럽 전역에 산재한 인공지능 연구 센터들을 연결하고 관련 인공지능 리소스에 대한 접근을 제공하는 "인공지능 주문형 플랫폼"(AI-on-demand platform) 개발을 지원할 예정이다.[68)] 전략 투자를 위한 유럽 기금 (European Fund for Strategic

65) Factsheet: Artificial Intelligence for Europe (25 April 2018) https://ec.europa.eu/digital-single-market/en/news/factsheet-artificial-intelligence-europe

66) Ibid

67) Horizon 2020 https://ec.europa.eu/programmes/horizon2020/en/

68) AI4EU initiative https://ai4eu.org;

Investments)은 2020년까지 총 5억 유로 이상의 투자를 주요 분야에 걸쳐서 집행할 예정이며 신생 기업들도 지원할 예정이다. 또한 인공지능 분야의 투자를 촉진하는 환경을 지속적으로 조성할 예정이다. 데이터는 인공지능의 재료가 되므로 유럽 위원회는 재사용 가능한 더 많은 공개 데이터-공공 유틸리티, 환경 데이터, 연구 및 보건 데이터를 포함-를 공개(opne)하는 법안 및 쉽게 데이터를 이용(access)하기 위한 방법을 제안하고 있다.

공공영역에 활용되는 인공지능은 공공 행정 기관들이 시민들에게 맞춤형 대응을 제공하게 하고, 교통 사고를 감소시키고, 기후 변화에 대처하고, 사이버 보안 위협을 예측하는 데 도움을 줄 수 있다. 중소기업, 비기술 회사 및 공공 행정 분야 등 잠재적 사용자들도 인공지능 주문형 플랫폼을 통해서 최신 알고리즘 및 전문 지식에 대한 조언을 얻고 쉽게 접근 할 수 있다. 디지털 혁신 허브(Digital Innovation Hub)는 주변의 중소기업들이 디지털 기회를 활용할 수 있도록 지원하는 지역 생태계로서 테스트, 기술, 비즈니스 모델, 금융, 시장 인텔리전스 및 네트워킹에 대한 전문 지식을 제공한다. 예를 들어, 자동차 금속 부품을 생산하는 소규모 회사는 지역 허브와 상의하고 인공지능을 사용하여 제조 프로세스를 개선하는 방법에 대한 조언을 구할 수 있다.[69]

② 인공지능에 의한 사회적, 경제적 변화의 준비

인공지능에 의해 많은 일자리가 생겨날 것이지만 수많은 일자리들이 사라지거나 불가피하게 그 형태가 달라질 것이다. 이에 대비하고자 유럽 집행위원회는 회원국들에게 유럽 교육 기금을 투입하여 교육 훈련 시스템을 현대화하고 노동시장 전환을 지원하도록 장려하고 있다. 유럽 지역에 더 많은 인공지능 분야의 인재들을 유지하기 위해 비즈니스 교육 파트너십을 지원하고 유럽 사회 기금의 재정 지원을 받아 교육 계획을 수립할 예정이다. 여기에는 디지털 기술, 과학, 기술, 공학 및 수학 (STEM)의 역량, 기업가 지원, 인공지능 분야 전문 지식 등 첨단 디지털 기술 교육의 지원도 포함된다. 2018년부터 2020년까지 해마다 6천 명의 졸업생을 대상으로 제공되는 첨단 디지털 기술 훈련과정(Digital Opportunity Traineeship)은 실제로 디지털 경험을 얻을 수있는 기회를 제공하고 있다. 이러한 프로그램은 기술 격차를 줄이고 인공지능 분야 등 정보통신 기술을 강화하는 방법으로 평가되고 있다.[70]

---

69) Digital Innovation Hubs www.s3platform.jrc.ec.europa.eu/digital-innovation-hubs

70) Digital Opportunity traineeships: boosting digital skills on the job https://ec.europa.eu/digital-single-market/en/digital-opportunity-traineeships-boosting-digital-skills-job

③ 적절한 윤리적 및 규범적 프레임워크의 마련

인공지능 알고리즘은 편향적 판단 등으로 인하여 전례 없는 윤리적 및 법적 문제를 제기할 수 있다. 이러한 문제들에 대응하기 위하여 '유럽 인공지능 협의체'(European AI Alliance)가 2018년 6월에 출범하였다. 이 협의체는 인공지능 윤리 가이드라인을 작성하고 인공지능 관련 문제들에 대한 중장기적 권고안을 마련하기 위해서 설립되었다. 이 협의체가 2018년 말까지 발표할 예정으로 있는 '인공지능 윤리 가이드라인'에는 알고리즘 투명성, 책무성 및 공정성 문제의 해결방안이 담길 것으로 보인다.[71)]

i) 알고리즘 투명성

인공지능 알고리즘은 대학 입학, 금융권 융자의 대출, 주식 시장의 고빈도 거래, 디지털 정보의 필터링 같은 많은 결정을 내리고 있다. 알고리즘 투명성은 인공지능의 책임감과 공정성과 관련한 중요한 요소로 부상했다. 알고리즘의 투명성은 소스 코드의 공개와는 무관하며, 유럽연합 내에서 현재 알고리즘의 투명성을 담보하기 위한 방법으로 일반 개인정보보호 규제(GDPR), 그리고 주식 시장에서의 알고리즘을 이용한 초빈도 거래에 대한 지침(Markets in Financial Instruments Directive, MiFID Ⅱ) 등이 존재한다MiFID II/ MiFIR은 2018년 1월부터 적용되기 시작한 새로운 규제로서 파생 상품 시장을 포함한 금융 시장의 투명성과 감독을 개선하고, 파생 상품 시장의 단점을 해결하여 투자자들을 위한 보호를 강화할 것으로 전망된다.[72)]

ii) 알고리즘 인식 제고(Algorithmic Awareness Building)

알고리즘은 공공 정책의 관련성 결정에 점차 중요한 역할을 하므로 유럽 집행 위원회는 그에 대한 인식을 제고하고 알고리즘 결정이 제기하는 도전과 기회에 대한 증거를 구축하고자 알고리즘 투명성에 대한 심층 분석을 수행하기로 하였다. 이 연구는 알고리즘이 수행하는 역할에 대한 심층적 정책 관련 분석, 특히 알고리즘이 중개하는 정보 흐름, 필터링, 개인화를 분석한다.[73)] 알고리즘 인식 제고(AlgoAware) 프로젝트는 세 가지 목적을 가지고 있다. 첫째, 온라인 플랫폼에서 알고리즘의 역할에 대한 폭넓고 공유된 이해를 증진하고 동시에

---

71) European AI Alliance https://ec.europa.eu/digital-single-market/en/european-ai-alliance

72) Markets in Financial Instruments (MiFID II)- Directive 2014/65/EU

73) 알고리즘 인식제고, 개방적이고 참여적 접근을 위해 AlgoAware https://www.algoaware.eu 사이트가 구축되었다. 이 사이트는 과학적 통찰력과 연구에 기초한 인공지능 정책 설계를 가능하게 할 것으로 전망된다.

대중의 인식을 높이고 토론을 촉진. 둘째, 알고리즘의 사용으로 제기되는 문제를 파악하고 과학적 근거를 수집. 셋째, 알고리즘이 제기하는 문제점들에 대한 해결책 제시 (정책 선택방안, 기술적 해결책, 사적 주체와 시민 사회의 행동을 포함).

iii) 제조물 책임 지침 (Product Liability Directive)의 새로운 해석

유럽 집행위원회는 결함이있는 인공지능 제품으로 발생하는 불법행위와 그 손해배상 책임 등에 대한 법적 명확성을 보장하기 위해 2019 년 중반까지 제조물 책임 지침의 해석에 관한 가이드라인을 발행할 예정이다. 1985 년에 제정된 유럽연합 제조물 책임 지침은 결함있는 제품에 대한 책임 규정으로서 소비자를 보호하고 기업이 혁신적인 제품을 시장에 내놓도록 장려하는 기능을 한다.[74] 인공지능이 다양한 전자 제품에 통합되고 그 이용과정에서 사람에게 중대한 피해를 입힌다면 그 제품에 결함이 입증되는 경우에 생산자는 보상을 지불해야 한다. 파손이나 사고로 이어지는 사건이 발생하면 그 원인이 밝혀져야만 책임을 귀속시킬 수 있는데 인공지능이 적용된 제품의 경우는 그 원인을 밝히는 것이 쉽지 않으므로 '신기술 디지털 기술에 대한 책임'의 관점에서 제조물 책임 지침의 개념을 명확히하는 가이드라인이 필요하게 되었다.

iv) 일반개인정보보호규제(GDPR)이 인공지능 연구개발에 미치는 영향

GDPR은 유럽 지역의 사용자 데이터 수집 및 처리에 대한 전례없는 제한을 부과하고 준수하지 않는 회사에 대한 엄중한 처벌을 가하는 근거가 된다.[75] 인공지능 기반 시스템을 통한 프로파일링을 포함한 자동화된 의사결정의 대상이 되는 데이터 주체는 그 결정과 관련된 의미있는 정보를 제공받을 권리가 있다. 또한 개인들은 자동화된 온라인 신용 신청을 거부하고, 전자적 모집 관행의 대상이 되지 않을 권리가 있다. GDPR은 프로파일링에 반대할 권리(제21조)와 자동화 처리의 대상이 되지 않을 권리(제22조)를 규정하고 있다.

v) 유럽 인공지능 협의체(European AI Alliance)의 인공지능 윤리 가이드라인

기업, 학계, 소비자 단체, 노동조합, 정책 입안자, 시민사회 대표자 등 이해관계자들(multi-

---

74) Directive 85/374/EEC (25 July 1985); Directive 1999/34/EC (10 May 1999)

75) 우리나라를 포함한 유럽연합 이외의 국가들도 'GDPR 적정성평가'를 받기 위해 개인정보보호 규제를 재정비하고 있다. (GDPR 제45조 '적정성 결정에 근거한 데이터의 이전' 참조) 유럽연합과 동등한 개인정보보호 수준이 인정하는 적정성 평가를 통과하면 유럽 지역 시민들의 개인정보를 이전 및 처리할 수 있기 때문이다. 박미사 외 "EU 개인정보보호 적정성평가의 주요 내용과 시사점" 정보통신기술진흥센터 주간기술동향 1802호(2017. 6. 28)

stakeholders)은 유럽 인공지능 협의체를 통해서 인공지능의 영향과 관련한 광범위한 공개 토론에 공식적으로 참여할 수 있다. 인공지능 협의체는 52명의 인공지능 분야 전문가들로 고위그룹(AI HLG: High-Level Expert Group on Artificial Intelligence)을 구성하여 워크샵 등을 통해 의견을 수렴하고 있다.[76] 그 검토 범위에는 인공지능의 공정성, 안전성, 사회적 영향, 알고리즘 투명성 등이 포함되며, 개인정보 보호(프라이버시), 인간 존엄성, 민주주의, 소비자 보호 및 차별 금지를 비롯한 유럽연합 기본권 헌장에 미치는 영향도 고려할 예정이다. 동 협의체는 유럽 집행위원회의 자문 기관 '과학 및 신기술 윤리 그룹'(EGE)이 2018년 3월에 발표한 '인공지능, 로봇 및 자율 시스템에 관한 성명'을 토대로 인공지능 윤리 가이드라인을 작성할 예정이다.[77]

〈표 3－2〉 '유럽을 위한 인공지능'에 포함된 윤리적, 법적 체계의 요약[78]

• 유럽 인공지능 협의체(European AI Alliance)에게 과학 및 신기술 윤리그룹(EGE)과 협력하여 인공지능 윤리 가이드라인, 안전기준, 기술 호환성(technical intero perability) 기준을 개발하도록 요청 (2018년 말까지) • 기술의 발전을 반영하여 유럽연합 제조물 책임 지침(Product Liability Dire ctive)의 해석 가이드라인을 제공하기로 결정 (2019년 중반) • 인공지능·사물인터넷·로봇공학에서 발생하는 법적 책임 및 안전 체계에 대한 잠재적 격차를 조사한 보고서를 발간하기로 결정 (2019년 중반) • 알고리즘에 대한 인식 제고(Algorithmic Awareness Building) 시범사업을 실시하여 인공지능이 초래하는 도전에 대한 정책 대응을 설계하고 관련 자료를 축적하기로 함. • 국가별 및 유럽 수준의 소비자 조직과 데이터 보호 감시기구를 지원함으 로써 인공지능이 적용된 제품과 서비스에 대한 이해를 구축. 이 과정은 유럽 소비자 자문그룹과 유럽 개인정보 보호 이사회(Data Protection Board)와 협력하여 진행.

76) European Commission, High-Level Expert Group on Artificial Intelligence https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

77) European Commission, European Group on Ethics in Science and New Technologies (2018) "Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems"

78) Wolfgang Maschek et al.,"New EU Strategy on Artificial Intelligence European Public Policy" Squire Patton Boggs (April 2018)

다. 분석

1) 윤리적 가이드라인

유럽 집행위원회는 인공지능 산업의 육성을 최우선적 목표로 설정하고 2018년 연말까지 윤리적 가이드라인을 준비하도록 유럽 인공지능 협의체에 위탁하였다. 유럽에서의 뚜렷한 특징은 GDPR을 통한 개인정보 보호와 공공 데이터 이용가능성의 확대를 디지털 싱글 마켓과 4차 산업혁명의 핵심조건으로 인정하고 있다는 점이다. 윤리적 가이드라인은 아직 마련 중에 있으므로 현재 단계에서 분석하기는 어렵지만 '유럽을 위한 인공지능'에 포함된 윤리적, 법적 체계는 참조할 만한 정책으로 볼 수 있다. 유럽 집행위원회가 윤리 원칙을 발표한다면 실효성 있는 사실상의 규범으로서 유럽권역 이외의 국가들에게도 파급효과를 미칠 것으로 예견된다. 한편 유럽연합 인공지능 협의체는 다중 이해관계자들이 의견을 표명하는 포럼(multi-stakeholder forum)의 방식으로 인공지능 윤리에 대한 다양한 논의들을 수렴하고 있다는 점은 주목할만 하다.

2) 인공지능 개발을 견제하는 GDPR 준수의무

유럽연합의 GDPR은 높은 수준의 개인정보 보호를 요구하지만 유럽연합은 인공지능 개발에서 미국과 중국 등에 뒤처지고 있다. 유럽연합이 GDPR 준수의무(Compliance)를 통해 인공지능 머신러닝에 공급하는 재료가 되는 개인 데이터를 통제하고, 윤리적 기준(ethics standards)을 중요시하는 이유는 세계적 인공지능 연구개발 경쟁을 견제하기 수단으로 여겨지기도 한다.[79]

전문가들은 대체적으로 일반개인정보보호규제(GDPR)가 인공지능 개발에 부정적 영향을 주거나 데이터를 입수하는 과정에 시간과 비용을 증가시킬 것으로 평가하고 있다.[80] 유럽 시민들의 데이터를 인공지능 분야에서 활동하는 기업들은 높은 법적 위험을 부담하게 되므로 GDPR은 인공지능의 개발과 사용에 부정적 영향을 주게 된다. GDPR의 준수의무로 인하여 데이터 수집이 어려워지고, 인공지능 시스템 및 응용 프로그램의 제작과 배포에 비용과 시간이 더 들게 되므로 인공지능 개발 속도도 늦어지게 된다.[81] 예를 들어, 인공지능 개인비서,

79) European Commission, European Group on Ethics in Science and New Technologies (2018) "Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems"

80) Ben Dickson (2018) "How GDPR Will Impact the AI Industry" PC Review (May 17, 2018); Nick Wallace and Daniel Castro (2018) "The Impact of the EU's New Data Protection Regulation on AI" Center for Data Innovation (March 27, 2018); Daniel Mikkelsen et al., (2017) "Tackling GDPR compliance before time runs out" McKinsey & Company (August 2017)

81) Alessandro Mantelero(2018) "AI and Big Data: A blueprint for a human rights, social and ethical impact

로보 어드바이저, 영화 스트리밍 추천, 신용 애플리케이션, 전자적 채용 또는 작업장 모니터링 등 개인들을 직접적 대상으로 하는 자동화된 데이터 처리는 GDPR에 의해 제한을 받게 된다. GDPR 제22조에 따르면, 이 경우 사업자는 모든 개인들에게 명시적 동의를 얻고 기록하는 과정을 거쳐야 한다.[82)]

3) 디지털 싱글 마켓과 유럽 인공지능 정책의 연계성

유럽 집행위원회의 인공지능 정책과 비젼은 디지털 싱글 마켓이라는 거시적 목표와 연결되어 진행되고 있다. 2016년 4월에 본격적으로 시작된 유럽의 디지털 싱글 마켓 정책은'모두를 위한 연결된 디지털 시장'을 구축하여 유럽의 경제성장과 고용을 촉진하고 디지털 혁신의 기회를 살리자는 지향점을 가지고 있다.[83)] 이 정책은 유럽연합 소속 회원국들의 규제와 정책이 디지털 경제의 특성을 제대로 반영하지 못하고 파편화되어 있다는 문제의식을 기초로 하고 있다. 불과 5%에 머무르고 있는 유럽 내 회원국들 간 디지털 거래를 활성화해야 할 필요가 있었다. 이러한 배경에서 디지털화를 통한 혁신의 기회를 잡으려면 회원국들이 협조하고, 각 국가마다 다른 제도적 장벽을 허물어서 디지털 시장을 유연하게 통합해야 한다는 정책적 공감대가 형성되었다.[84)]

디지털 싱글 마켓 정책의 추진 과정은 4차 산업혁명을 가속화하고 인공지능 산업의 진흥을 원하는 유럽연합의 목표를 보여준다고 할 수 있다. 디지털 싱글 마켓 정책의 범위에는 전자상거래, 콘텐츠는 물론, 데이터화와 연계 가능한 온라인시장을 통합하고 시장규범은 물론, 산업・인프라 정책의 통합까지 포괄한다.[85)] 디지털 싱글 마켓과 인공지능 정책의 관련성을 보여주는 자료는 2017년 5월 유럽 집행위원회가 펴낸 '디지털 싱글 마켓 전략 실행 중기검토' 보고서이다.[86)] 동 보고서의 부제목 '인공지능 역량구축'(Building artificial intelligence

---

assessment" Computer Law & Security Review Vol 34, Issue 4 (Aug 2018) 754-772

82) David Roe "Understanding GDPR and Its Impact on the Development Of AI" CMS WiRE (Apr 26, 2018)

83) Digitising European Industry-Reaping full benefits of a Digital Single Market (19 April 2016) https://ec.europa.eu/digital-single-market/en/news/communication-digitising-european-industry-reaping-full-benefits-digital-single-market

84) Digital single market https://ec.europa.eu/commission/priorities/digital-single-market_en

85) "EU 디지털 단일시장 전략의 평가와 시사점" Global Strategy Report 17-010, KOTRA (2017) 3면-8면 참조

86) Digital Single Market Mid-term Review (10 May 2017) https://ec.europa.eu/digital-single-market/en/news/digital-single-market-mid-term-review

capacities)은 디지털 싱글 마켓이 지향하는 중요 목표가 무엇인가를 말해준다고 하겠다.[87] 동 보고서는 유럽연합이 재사용이 가능한 많은 공유 데이터(common data)를 제공하고 공유 데이터를 이용(access)하는 비용을 낮추어야 한다고 지적하고 있다.

3. 영국

가. 영국의 인공지능: 준비, 의향, 역량? (AI in the UK: ready, willing and able?)

1) 개요

영국 정부가 민간 분야와 체결한 '인공지능 분야의 협력(AI Sector Deal)'이 2018년 4월 발표되었다.[88] 뒤이어 인공지능 산업을 육성하여 데이터 기반 혁신의 중심에 서겠다는 비젼을 담은 '웅대한 도전(Grand Challenges)'이 공개되었다.[89] 동 보고서는 인공지능과 머신러닝을 세계 경제를 변화시키큰 범용기술로 정의하고 영국을 인공지능 및 데이터 혁명의 최전선에 위치시킨다는 목표를 설정하였다.

영국의 인공지능 윤리 보고서 '영국의 인공지능: 준비, 의향, 역량?'은 상원에 설치된 인공지능 특별위원회(Board of Lords AI Select Committee)가 작성하였다. 클레멘트 존스(Clement-Jones)는 200 명 이상의 업계 전문가들을 접촉하여 작성한 보고서를 발표하면서 영국은 국민의 이익을 위해 인공지능을 활용하고, 인공지능의 윤리적 발전을 통해 국제 사회를 이끌 수 있는 기회를 가지고 있음을 강조하였다. 보고서의 취지는 영국은 인공지능이 주는 기회를 최대한 활용해야 하지만 인공지능의 개발과 사용과정에서 발생하는 위험 요소와 오용의 문제는 윤리 원칙을 채택하여야 완화될 수 있다는 것이다. 나아가 2018년 9월에는 영국 정부는 세계경제포럼(World Economic Forum)과 시민을 위한 공공분야의 인공지능 사용 가이드라린을 공동 설계하기로 하였다.[90] 많은 공공기관들이 인공지능 편향에 대한 우려, 개인정보

---

87) European Commission, Brussels, 10.5.2017 COM(2017) 228 final, The Mid-Term Review on the implementation of the Digital Single Market Strategy 참조 (10 May 2017) https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52017DC0228&from=EN

88) United Kingdom Govenment Policy paper 'AI Sector Deal' (26 April 2018) www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal

89) United Kingdom Govenment, 'The Grand Challenges' (21 May 2018); United Kingdom Government "Industrial Strategy: Building a Britain fit for the future" (Nov. 2017) 198~199 pages

90) "United Kingdom Partners with World Economic Forum to Develop First Artificial Intelligence Procurement Policy",World Economic Forum Press Release(20 Sep 2018)

보호, 책임성, 투명성 및 복잡성 때문에 인공지능 기술의 활용에 조심스러운 입장이지만 영국 정부는 세계경제포럼 센터와 함께 인공지능 정부조달에 대한 정책과 가이드라인을 개발하기로 했다.

〈표 3－3〉 영국 상원 보고서의 인공지능 규범(AI Code)[91]

‘영국 인공지능: 준비, 의향, 역량’ 보고서는 권장 사항으로서 국가 간, 국제적으로 채택 될 수 있는 인공지능 규범(code)을 설정하였다. 상원 위원회가 제시한 5가지 인공지능 원칙은 다음과 같다. － 인공지능은 인류의 공동 이익과 이익을 위해 개발되어야 한다. － 인공지능은 투명성과 공정성의 원칙에 따라 작동해야 한다. － 인공지능이 개인, 가족, 지역 사회의 데이터 권리 또는 개인정보를 감소시켜서는 안된다. － 모든 시민은 인공지능을 통해서 정신적, 정서적, 경제적 번영을 누리도록 교육받을 권리를 가져야 한다. － 인간을 해치거나 파괴하거나 속이는 자율적 힘을 인공지능에 절대로 부여해서는 안된다.

2) 주요내용

‘영국의 인공지능: 준비, 의향, 역량?’ 보고서는 인공지능 윤리와 관련하여 어떻게 데이터를 관리하고 보호할 것인가에 상당한 비중을 두고 있다. 데이터와 관련하여 다음의 권고를 담고 있다.[92]

- 개인은 자신의 데이터에 대한 개인의 통제력과 개인의 사용 방식을 잘 이해할 수 있어야 한다. 데이터를 수집하고 이용하는 현재의 방식을 변경해야만 모두가 데이터에 공정하고 합당한 접근을 할 수 있고 시민과 소비자의 개인 정보를 보호 할 수 있다. 이를 위해서 오픈 데이터(open data), 윤리자문 위원회(ethics advisory boards)의 설치, 데이터 보호법의 마련, ‘데이터 이동성’(data portability) 및 데이터 신탁(data trusts) 등 새로운 프레임워크를 개발하여야 한다.
- 기술 기업들에 의한 데이터 독점은 피해야하며 이를 위해서는 경쟁이 필요한다. 그러므로 경쟁 당국과 정부는 영국에서 운영되는 대형 기술 기업들의 데이터 사용이 적정한

91) UK can lead the way on ethical AI, says Lords Committee (16 April 2018) www.parliament.uk/business/committees/committees-a-z/lords-select/ai-committee/news-parliament-2017/ai-report-published/

92) The House of Lords Select Committee on Artificial Intelligence Report of Session 2017-19 “AI in the UK: ready, willing and able?” (16 April 2018)

지를 검토해야 한다.

- 과거 데이터에 포함된 편향이 무의식적으로 자동화된 시스템에 내장되어서는 안된다. 그러므로 정부는 인공지능에서 사용되는 데이터 세트를 감사하기 위한 새로운 접근법의 개발을 장려하고 인공지능 전문가의 훈련 및 모집에서 다양성을 장려해야 한다.
- 인공지능의 투명성이 요구되므로 인공지능 위원회를 설치하여 산업계가 제작한 인공지능이 중요하거나 민감한 결정을 내릴 때 소비자에게 알리도록 자발적 메커니즘을 수립해야 한다. 다섯째, 기초 교육 단계에서 아이들은 인공지능을 사용할 수 있도록 교육과정이 준비되어야한다. 인공지능의 윤리적 설계와 사용은 교육 커리큘럼의 필수적인 부분이 되어야 한다.
- 정부는 인공지능의 개발과 확산을 촉진하기 위해 인공지능 조달(procurement)을 위한 목표치를 설정하여 추진해야 한다. 이러한 정부 조달은 투자를 발생시키게 되므로 인공지능을 활용한 공공정책 과제의 해결을 장려할 수 있다. 영국의 국민건강 서비스(NHS)가 구글 딥마인드 인공지능을 이용해 혈액 검사 결과를 분석하고 급성 질환이나 기타 질병 징후를 분석하는 시스템을 구축한 사례는 인공지능을 활용한 헬스케어(AI for healthcare)로서 인상적 진보로 평가된다.[93] (2016년 딥마인드는 국민건강서비스(NHS)와 160만 명의 환자 데이터에 접근할 수 있는 협약을 체결했다. 국민건강서비스(NHS)는 구글 딥마인드의 인공지능을 이용하여 혈액 검사 결과를 분석하여 급성 신부전 위험등 질병징후를 분석하는 시스템을 도입했다. 개인정보 침해 논란이 일어나자 환자들의 정보 보호를 위해 데이터 익명화 작업이 진행되었다.)[94]
- 인공지능 시스템이 오작동을 일으키거나 사용자에게 해를 입힐 때 책임을 묻기 위한 법률조항이 충분한지 여부가 불명확하다. 이 점을 명확하게 하기 위한 검토가 필요한다. 의회의 법률위원회가 이 문제를 조사하기를 권고한다.

---

93) Macy Bayern “DeepMind, NHS use anonymized patient data in AI to avoid regulatory hurdles” Tech Republic (July 2, 2018)

94) 공공기관이 보유한 환자정보를 인공지능 기업에 제공한 영국의 사례는 개인의 동의를 받지 않은 처방전 개정보가 합법인가를 둘러싸고 벌어진 국내 약학정보원 사건 논쟁과 대비된다. 약학정보원 사건의 민사소송 제1심 판결은 개인정보보호법 제18조 제2항 제4호 해석과 관련하여 비식별화가 제대로 이루어지지 아니한 채 IMS에게 제공된 개인정보에 대하여는 구체적 법익이 침해되지 않았다고 보아 손해배상 책임을 인정하지는 않았다. (서울중앙지방법원 2015고합665 사건)

정부는 국가 정책 프레임워크를 입안할 때 인공지능 정책(AI policy in the UK)이 산업 진흥 전략들과 정확하게 보조를 맞출 수 있도록 고려해야만 한다.

나. 영국 데이터 윤리 및 혁신 컨설팅 센터

2018년 6월 출범한 영국 미디털 문화 미디어 스포츠 장관 산하 '데이터 윤리 및 혁신 컨설팅 센터'(Centre for Data Ethics and Innovation Consultation)의 활동은 인공지능에 공급되는 데이터의 사용과 관련성이 깊다.[95] 데이터 사용 방식의 발전은 때때로 낯선 윤리적 문제를 야기하는데 그럼에도 불구하고 데이터 거버넌스를 유지해야만 대중의 신뢰를 유지할 수 있다는 것이 센터(CDEI)의 설립목적이다.

- 기업, 시민, 공공영역에서 데이터를 안전하게 사용하고 인공지능 혁신을 가능하는 명확한 규칙이 필요하다. 센터(CDEI)는 데이터 및 인공지능 관리와 관련하여 데이터 사용과 새로운 데이터 중심 기술의 윤리적 혁신을 장려하는 방법을 정부에게 권고한다.
- 윤리적이고 혁신적인 데이터 사용을 유도하기 위한 모범 사례를 검토하고 명확한 정책 또는 규제조치의 필요성을 정부에 조언한다. 인공지능 등 급속히 발전하는 기술로 인하여 등장하는 새로운 윤리적 문제를 다루고 데이터 사용의 모범 사례를 모아서 데이터 중심의 혁신을 지원한다.
- 데이터와 인공지능이 윤리적으로 사용되도록 보장하여 책임있는 혁신의 성장을 돕고 영국을 신뢰할 수있는 데이터 중심 비즈니스의 허브로 만든다
- 데이터 사용 및 인공지능 관련 문제가 복잡하고 광범위하다는 것을 인지하고 컨설팅을 제공할 것이며 대중의 견해를 이해하고 국제적인 논쟁을 주도한다.
- 데이터 사용 및 인공지능이 관리되고 규제되는 방식에 대한 이해관계가 있는 모든 사람들과 광범위하게 소통한다. (기업, 규제 기관, 지방 정부, 학계 및 시민 사회를 포함)

다. 분석

상원 위원회가 발표한 '영국의 인공지능: 준비, 의향, 역량?' 보고서는 각자의 입장에 따른 데이터 관리에 대하여 구체적으로 권고하고 있다. 기업들에게는 오픈 데이터에 공정하고 합

95) Department for Digital, Culture, Media & Sport "Centre for Data Ethics and Innovation Consultation" (13 June 2018) www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation

당한 접근을 할 수 있도록 보장하지만 특정한 거대 기업이 방대한 양의 데이터를 제어할 수 있는 강력한 힘을 가지지 못하도록 정부에게 영국 내의 대규모 기술 회사들의 데이터 독점 여부를 조사할 것을 촉구하였다. 또한 시민과 소비자들의 개인 프라이버시는 데이터 보호 법률을 통해서 보호되어야 한다는 점을 강조하고 있다.[96]

영국은 인공지능의 발전이 데이터 사용과 관련한 복잡하고 광범위한 논란을 불러일으킨다는 사실을 인지하고 2018년 6월 '데이터 윤리 및 혁신 컨설팅 센터'(CDEI)를 설립하여 자문을 제공하기 시작하였다. CDEI는 데이터 사용과 새로운 데이터 중심 기술의 윤리적 혁신을 장려하는 방법에 대해 정부에 조언하므로 데이터 사용과 관련한 실무에서 모호함과 불명확성을 상당부분 해소하는 긍정적 효과를 가져올 것이다. 인공지능 산업계는 어떤 데이터를, 어떻게 사용하는 것이 합법적인가가 불명확하여 상당한 비용과 경제적 손실을 입고 법적 분쟁으로 이어지는 사례들이 많았다. 그러므로 적절한 모범사례와 조언을 제공하고 개선 방안을 권고하는 정부 소속 자문기관의 활동은 크게 유용할 것으로 보인다.

4. 프랑스

가. '의미있는 인공지능을 위하여'(For a Meaningful Artificial Intelligence) 보고서

1) 개요

2018년 3월에 발표된 프랑스 의회 의원 세드릭 빌라니(Cédric Villani)가 작성한 "의미있는 인공지능을 위하여" 보고서는 인간을 위한 인공지능(AI for humanity)을 전면에 내세우면서도 비즈니스·경제 분야의 혁신을 위하여 필수불가결한 인공지능의 활용 방식을 권고하고 있다. 프랑스 인공지능 국가 전략은 인공지능 경제의 새로운 핵심자원으로 부상한 데이터와 기술인재 확보에 비중을 두고 있다.[97]

---

96) 영국은 현재 유럽연합 회원국 자격을 탈퇴하는 수순을 밟고 있지만 GDPR 규제를 준수하겠다고 밝히고 있다. United Kingdom Government "The exchange and protection of personal data" https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/639853/The_exchange_and_protection_of_personal_data.pdf

97) Cédric Villani, "For a Meaningful Artificial Intelligence:Towards a French and European strategy" (Mar 8, 2018) www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf"

2) 주요내용[98)]

세드릭 빌라니의 보고서가 담고 있는 권고 내용은 크게 세가지로 나눌 수 있으며 특히 네 번째 항목에 '인공지능 윤리'에 대한 내용이 존재한다.

- 인공지능 기술경쟁을 위한 데이터의 중요성: 국가주도의 데이터 확보 전략의 수립이 필요하다. 오픈 데이터의 확보화 동시에 데이터 보호방안을 마련해야 한다. 또한 인공지능 생태계를 구축하고 강화해야한 한다.
- 인공지능 연구에서 프랑스의 잠재력 증대: 학제 간 인공지능 연구를 위한 네트워크를 구축하고 공공연구 인프라를 강화할 필요가 있다.
- 인공지능의 확대로 인하여 노동시장 변화 대응: 직무가 자동화되는 범위에 대하여 예측하고 대응할 필요가 있다. 인공지능 기술개발에 대비한 교육의 변화가 필요하므로 대책을 마련해야 한다
- 인공지능으로 인한 사회변화 및 대응: 친환경적 인공지능 기술을 개발하여 생태계와 융합을 추구할 필요가 있다. 또한 인공지능의 역기능을 연구하고 파악할 수 있는 방안을 마련해야 한다. 인공지능이 야기하는 윤리 문제를 교육하고 법제도를 마련해야 한다.

3) 인공지능 윤리를 위한 권고

"의미있는 인공지능을 위하여" 보고서 5장은 "인공지능 윤리란 무엇인가?"라는 제목 하에 다음 내용을 담고 있다. 블랙박스 해부하기(Opening the 'Black Box'), 설계 단계에서 고려하는 인공지능 윤리(Considering Ethics from the Design Stage), 데이터를 집합적 권리로 접근하기기(Considering Collective Rights to Data) 어떻게 우리는 계속해서 인공지능을 통제할 수 있나?(How Do We Stay in Control?) 인공지능 윤리를 위한 특수한 거버넌스방안 (Specific Governance of Ethics in Artificial Intelligence).

데이터 오류나 편향이 사회적 불평등을 재생산할 수 있는 문제에 대하여는 데이터 알고리즘을 연구・감시하거나 인공지능 기술 전문가 그룹을 구성하여 데이터 알고리즘 분석을 진행하는 방식을 제안하였다. 사용자 친화적 인터페이스 구축으로 사용자가 데이터 알고리즘

98) "프랑스 AI 권고안 리뷰와 ICT 정책 검토", 해외 ICT R&D 정책동향 2018-1호 21면 참조, 정보통신기술진흥센터

을 잘 이해할 수 있도록 하는 방안이 주장되었다. 프랑스에 윤리자문위원회를 구성하여 인공지능으로 인해 발생하는 윤리적 문제에 대한 공개 토론을 실시하고 교육 및 법적 가이드라인를 마련하는 방안을 권고하였다. 인공지능의 설계와 연구개발을 담당하는 기술자 및 연구원들을 대상으로 윤리 교육을 실시하여 인공지능 기술의 사회적 영향력에 대한 인식을 제고하는 방안이 주장되었다. 또한 국방안보 분야와 같은 공적 영역에서 인공지능 기술을 활용할 경우, 결과 책임에 대한 법적 가이드라인과 감시가 필요하다는 점도 지적되었다.

나. 프랑스 개인정보감독기구(CNIL)의 인공지능 윤리 문제 보고서

1) 개요

프랑스 개인정보감독기구(CNIL)는 인공지능 윤리문제 보고서를 2018년 8월에 공개하였다. 이 보고서의 목적은 인공지능과 관련한 윤리적 문제의 개요를 제시하고 시민사회에 인공지능 윤리 문제에 대한 집단적 인식을 높이는 것이다. 이 보고서는 2017년에 진행된 대규모 공개 토론의 결과로서 약 3천명이 60개 연구소, 공공기관, 노동조합, 싱크탱크, 회사 등 제기 한 45 가지 논점에 참여했다. 이처럼 수개월 동안 진행된 공개적 토론은 집단적 사고 과정을위한 토대를 마련하는 데 도움이된다.[99)]

2) 인공지능 윤리의 6 가지 주요 관심사

이 보고서는 공개적 토론이 진행되기 이전에 알고리즘과 인공지능의 사용과 관련된 6 가지 윤리적 문제를 다음과 같이 정리하였다.

- 자율적인 기계: 자유의지(Free Will)를 가지므로 위협인가?
- 편향, 차별 및 배제
- 알고리즘 프로파일링
- 인공지능을 강화하기 위한 대규모 파일링: 새로운 균형을 추구하려면?
- 데이터의 품질, 수량, 관련성: 인공지능에 의해 큐레이션 된 데이터
- 인공지능에 의해서 도전받는 인간의 정체성

3) 인공지능 윤리 정책을 위한 권고사항

공공영역, 정부, 시민사회 (기업, 시민 등) 모두를 위한 6가지 정책 권고가 제시되었다.

---

99) Algorithms and artificial intelligence: CNIL's report on the ethical issues (25 May 2018) www.cnil.fr/en/algorithms-and-artificial-intelligence-cnils-report-ethical-issues

- “알고리즘 체인”에 연결된 인공지능 설계자, 전문가, 시민 등 모든 관련자를 교육
- 알고리즘 시스템을 이해할 수 있도록 만들기 (인공지능과 사용자가 문제를 해결하도록 중재하기).
- 인간의 자유와 이해관계를 위한 알고리즘 시스템의 설계 개선
- 알고리즘 조사(auditing algorithms)를 위한 국가 플랫폼 구축
- 윤리적 인공지능 연구에 대한 인센티브 확대 및 일반 관심 연구 프로젝트 참여
- 인공지능 비즈니스 분야에 윤리를 강화하기.

다. 분석

세드릭 빌라니가 작성한 “의미있는 인공지능을 위하여” 보고서는 프랑스를 인공지능 분야의 선도국가로 부상시키려는 계획을 담고 있으나 인공지능 기술개발을 위한 교육 시스템 구축에 비중을 크게 두고 있고 인공지능 윤리에 대하여는 체계적 대응책을 담고 있지 못하다. 한편, 프랑스 개인정보감독기구(CNIL)는 인공지능 윤리 문제 보고서는 공개적 토론을 통해 얻어진 결과물이라는 점에서 유의미하며, 인공지능 윤리를 둘러싼 논쟁점들에 대한 분명한 초점과 권고를 일목요연하게 제시하고 있다. 프랑스 개인정보감독기구의 인공지능 윤리 문제 보고서는 데이터와 인공지능 윤리와의 관련성을 충분히 파악하고 있다고 평가할 수 있다.[100] 따라서 우리나라도 인공지능 윤리 가이드라인을 도출하기 위해 다수의 참여자들에게 숙의적 공론화 과정을 거치거나 공개적 토론이 가능한 포럼을 설치하는 방안을 고려해 볼만한다.

5. 중국

가. ‘차세대 인공지능 발전계획’(新一代人工智能發展規划的通)

1) 개요

2017년 8월 중국 국무원은 중국 인공지능 발전의 우위 구축을 위한 ‘차세대 인공지능 발전

100) 프랑스 개인정보보호 감독기구(CNIL)는 유럽연합의 일반개인정보보호규제(GDPR) 이행을 위해 2가지 이슈에 대한 공개검토를 2017년에 진행한 바 있다. 그것은 데이터 투명성(transparency)과 국제 데이터 이전(international data transfer)이다. 그 목적은 GDPR과 관련된 구체적인 질문과 GDPR 해석의 어려움, 그리고 모범사례(best practices)를 수집하는데 있었다. Privacy Impact assessment (PIA) https://www.cnil.fr/en/privacy-impact-assessment-pia

계획(新一代人工智能發展規划)'을 발표하였다.[101] 인공지능을 통해 국가경쟁력을 높이겠다는 중국 지도부의 강력한 의지와 비젼이 담긴 발전계획은 3단계 전략목표와 6대 중점임무로 구성된다. 2025년까지 인공지능 기술측면에서는 세계적 수준에 도달하고 인공지능을 스마트 제조, 스마트 의료, 스마트 시티, 스마트 농업에 광범위하게 적용하여 차세대 인공지능 핵심 산업을 육성한다는 전략이다. 산업자원부는 2017년 12월에 차세대 인공지능 산업의 발전을 촉진하기 위한 3 개년 계획을 발표하여 구체적 행동에 들어갔다.[102] 2018년 9월 중국 상해에서 개최된 '세계인공지능컨퍼런스'에서 시진핑 주석은 인공지능이 제기하는 복잡한 윤리적, 법적 문제를 해결하기 위해 전 세계가 함께 노력할 것을 촉구하는 등 인공지능 분야에서 중국의 글로벌 리더십을 다기지 위한 준비를 계속하고 있다.[103]

2) 주요내용

중국 차세대 인공지능 발전계획이 담고 있는 6대 중점 임무는 ① 개방형・협동형 인공지능 기술혁신 체계 구축 (기초 이론, 통용 기술, 혁신 플랫폼, 고급인력 등의 측면을 강화) ② 고효율 스마트 경제의 육성 (스마트 경제를 육성하여 인공지능 신흥 산업을 발전시키고, 스마트 산업의 업그레이드를 추진하며, 인공지능 혁신기지를 설립) ③ 안전하고 편리한 스마트 사회 건설 (고효율 스마트 서비스를 발전시키며, 사회 관리의 스마트화 수준을 향상) ④ 인공지능 영역의 군민융합(軍民融合) 강화 (인공지능 영역의 군민융합을 강화하고, 군민 쌍방간의 인공지능 기술 응용을 촉진시키며, 군민 혁신자원을 공동 건설 및 공유) ⑤ 안전하고 효율성 높은 스마트 인프라 체계 구축 (인터넷, 빅데이터, 슈퍼컴퓨터의 고효율화를 통해 고효율의 스마트 인프라 설비 체계를 구축) ⑥ 중대 과학기술 프로젝트 계획이다.[104]

중국의 차세대 인공지능 발전계획 5조는 '보호조치'(五. 保障措施)라는 소제목 아래 ① 인공지능의 발전을 위한 법률, 규정 및 윤리 규범의 제정촉진, ② 인공지능 개발을 지원하기 위한 주요 정책의 개선, ③ 인공지능 기술 표준 및 지적 재산권 시스템 수립, ④ 인공지능 안

---

101) 國務院關于印發, 新一代人工智能發展規划的通, 國發〔2017〕35号
http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

102) 차세대 인공지능 산업 발전 촉진을 위한 3 개년 계획 실행 및 배포에 관한 산업 자원부 고시 促進新一代人工智能産業發展三年行動計划 (2018-2020年)
www.miit.gov.cn/n1146295/n1652858/n1652930/n3757016/c5960820/content.html

103) "China calls for global collaboraton on AI ethics and law" The Global Legal Post (17 Sept. 2018)

104) "중국의 인공지능(AI) 전략". 과학기술 & ICT 정책기술동향 No. 101, 한국정보화진흥원 (2017. 9. 29)

전 감독 및 평가 시스템 구축, ⑤ 인공지능 노동 훈련의 적극적 강화를 순차적으로 설명하고 있다. 이하에서는 ‘ 인공지능의 발전을 위한 법률, 규정 및 윤리 규범의 제정촉진’, ‘인공지능 안전 감독 및 평가 시스템 구축’을 구체적으로 살펴보겠다.

3) 인공지능 발전을 위한 법률, 규정 및 윤리 규범의 제정촉진

- 중국의 차세대 인공지능 발전계획 5조 1항은 인공지능 발전을 위해 법률, 규정, 윤리 규범 제정을 촉진한다는 내용이다.(五、保障措施（一）制定促進人工智能發展的法律法規和倫理規范)
- 인공지능과 관련된 법적, 윤리적 및 사회적 문제에 대한 연구를 강화하고 인공지능의 건전한 발전을 보장하기 위해 법률, 규정 및 윤리 체계를 확립한다.
- 인공지능의 민사 및 형사책임과 관련한 법적 문제는 개인정보 보호, 재산권 보호, 정보보안, 추적가능성 및 귀속, 인공지능에 대한 권리 ・의무・ 책임의 명확한 법적 귀속 등이다.
- 자율주행차량, 서비스 로봇 등 인공지능 기술이 적용되는 하위 분야에 초점을 두고 안전관리 규정을 연구개발하여 신기술의 신속한 적용을 위한 법적 기반을 마련한다.
- 인공지능과 관련한 행동과학 및 윤리 문제 연구를 수행하고, 인간-컴퓨터 협업을 위한 윤리적 및 다단계 판단구조의 윤리 체계를 수립한다.
- 인공지능 설계자에게 적용되는 윤리 및 행동 규범을 수립한다. 인공지능이 초래하는 잠재적 위험과 이점에 대한 평가를 강화한다. 복잡한 인공지능의 적용과정에서 응급상황이 발생하였을 때 해결책을 구축한다.
- 인공지능의 글로벌 거버넌스에 적극적으로 참여하고, 로봇의 안전 감독 등 인공지능에 대한 주요 국제 공통 논제에 대한 연구를 강화한다. 인공지능 법규 및 국제규칙에 대한 국제 협력을 강화하고, 전 지구적 도전 과제를 공동으로 해결한다.

4) 인공지능 개발을 지원하는 주요 정책의 개선

- 중국의 차세대 인공지능 발전계획 5조 2항은 인공지능 개발을 지원하는 주요 정책의 개선을 촉진한다는 계획을 담고 있다.(五、保障措施 (二）完善支持人工智能發展的重点政策。)
- 인공지능 중소기업 및 신생 기업에게 재정 및 세금 인센티브를 지원하고, 첨단기술 기업에게는 세금 인센티브 및 연구개발 경비의 공제를 통해 인공지능 기업의 발전을

지원한다.

- 데이터 개방・보호정책의 이행을 개선하고 공공 데이터를 개방하여 공공영역과 기업이 공공 데이터의 상업적 가치를 충분히 활용하고 인공지능 분야의 혁신을 촉진하도록 지원한다
- 인공지능에 적응하기 위하여 교육, 의료, 보험, 사회 지원 및 기타 정책 시스템을 개선하고 인공지능으로 인한 사회적 문제에 효과적으로 대응한다.
- 중소기업 인공지능을 지원하고 신생기업에게는 세금 인센티브를 부여하고 하이테크 인공지능 기업에게는 세금 인센티브와 연구개발 비용 공제 정책을 통해 사업개발을 지원한다

5) 인공지능 안전 감독 및 평가 체계의 구축

- 중국의 차세대 인공지능 발전계획 5조 4항은 인공지능 분야에 대한 안전 감독 및 평가체계를 구축한다는 구체적 계획이 담겨있다.(五、保障措施 (四) 建立人工智能安全監管和評估体系。
- 인공지능이 국가안보 및 기밀성에 미치는 영향에 대한 연구 및 평가를 강화하고 인력, 기술, 자재 및 관리를 위한 보안 시스템을 개선한다. 인공지능 보안을 위한 모니터링 및 조기경보 메커니즘을 구축한다. 인공지능 기술 개발의 연구성과들을 추적하여 기술과 산업의 발전 추세를 파악한다. 인공지능에 의해 초래되는 위험의 인식, 위험 평가 및 예방의 중요성을 강조하고, 미래 지향적 예방 지침을 강화한다
- 단기적으로는 인공지능이 고용에 미치는 영향에 초점을 맞춘다. 장기적으로는 인공지능이 사회 윤리에 미치는 영향에 초점을 맞추고 제어가능한 범위 내에서 인공지능 개발이 안전하게 규제되도록 보장한다.
- 개방적이고 투명한 인공지능 감독 시스템을 수립하고 개선한다. 설계책임 및 응용 프로그램의 감독으로 구성된 '2계층'의 감독 구조를 구현한다. 인공지능 알고리즘 설계, 제품 개발 및 결과로 이어지는 전체 프로세스에 대한 감독을 실현한다.
- 인공지능 산업 및 기업을 장려하고, 경영을 강화하며, 데이터 남용, 개인 사생활 침해 및 도덕 및 윤리 위반에 대한 처벌을 강화한다. 인공지능 네트워크 보안 기술의 개발을 강화하고 인공지능 제품・시스템의 보안을 강화한다. 인공지능 응용 프로그램에 대한 동적 평가 메커니즘을 구축하고 인공지능의 설계, 제품・시스템의 복잡성, 위

험, 불확실성, 해석 가능성, 경제적 영향에 대한 검사 방법 및 지표체계를 개발한다. 인공지능 보안 인증을 홍보하고 인공지능 제품·시스템의 성능을 평가할 수있는 교차 검사 플랫폼을 구축한다.

나. 분석

중국의 차세대 인공지능 발전계획의 요체는 혁신적 국가 건설을 가속화하기 위해 인공지능의 발전의 중요한 전략적 기회를 잡아야 한다는 것이다. 이러한 전제에서 인공지능의 발전을 위한 토대로서 관련된 법률과 윤리 규범을 정비해야 하고, '인공지능 안전 감독 및 평가 시스템을 구축해야 한다는 구상을 담고 있는 것이다. 동 발전계획이 담고 있는 인공지능의 편향 및 부작용에 대한 대응, 개인정보 보호 필요성 등은 유럽연합 국가의 인공지능 윤리·규범 정책과 비교하여 차이를 발견하기 어렵다.

인공지능 산업 분야의 기술수준과 인공지능 글로벌 거버넌스에서 세계를 선도하겠다는 중국의 계획은 서구의 경쟁국가들을 압박하고 있다.[105)106)] 중국의 인공지능굴기(人工智能 的崛起)를 가능하게 만든 경쟁력의 원천은 3만명에 이르는 엔지니어와 과학자들, 인공지능 시스템 훈련에 필요한 풍부한 데이터이다. 중국에는 데이터 수집과 사용을 막는 장애물이 없기 때문에 인공지능 산업에 필요한 거대한 데이터베이스를 축적할 수 있었다.[107)] 반면, 중국 대도시들은 CCTV로 수집한 데이터를 머신러닝으로 분석하는 고성능 안면 인식 시스템(AI-powered surveillance)을 시민 감시와 경찰 활동에 널리 사용하는 중이다.[108)] 또한 중국이 밝힌 '인공지능 분야의 민군융합 강화'는 인공지능을 탑재한 자동화 살상무기의 등장을 막으려는 국제사회의 움직임과 정면으로 충돌한다.[109)] 이는 인공지능 윤리와 안전을 국가정책

---

105) "China's Artificial Intelligence Strategy Poses a Credible Threat to U.S. Tech Leadership" Council on Foreign Relations (4 Dec. 2017); Darrell West, John R. Allen "How artificial intelligence is transforming the world" (April 24, 2018)

106) 1,500억 달러를 인공지능 산업육성에 투자하고 2030 년까지 중국을 선두 주자로 올려놓겠다는 중국 국무원의 계획은 미국의 정책 입안자, 기업가들에게 충격을 안겨주었다. "America v China: The Battle for Digital Supremacy" The Economist (March 15, 2018); Beijing Wants A.I. to Be Made in China by 2030 , New York Times. (July 20, 2017)

107) "Intelligent Machines China's AI Awakening "MIT Technologu Review (Oct. 10, 2017)

108) Paul Mozur, John Markoff, "Is China Outsmarting American Artificial Intelligence?" New York Times (May 28, 2017); Paul Mozur "Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras" New York Times (July 8, 2018)

에 실제로 고려하지 않는다는 의구심을 던져준다. 그렇지만 옥스포드대 미래의 삶 연구소(FHI)가 발간한 '인공지능 중국몽 해독하기'(Deciphering China's AI Dream)에 따르면 중국은 인공지능 윤리와 안전에 대한 심도 깊은 고찰을 진행 중이다.[110)]

중국 산업자원부가 2017년 12월 발표한 '차세대 인공지능 산업발전 촉진을 위한 3개년 계획'(促進新一代人工智能産業發展三年行動計划, 2018-2020年)은 차세대 인공지능 발전계획의 구체적인 행동방안에 해당한다.[111)] 3개년 계획은 인공지능 관련 정책 및 법률 및 규제의 연구를 수행하고, 산업계가 합리적으로 데이터를 공개하도록 촉진하며, 인공지능과 관련한 이슈에 관하여 양자 간 및 다자간 국제협력을 장려한다는 내용을 담고 있다.[112)] 그러나 인공지능 윤리 원칙이나 규범제정을 위한 공개 토론, 의견수렴 절차는 전혀 없으며, 핵심 지도부의 비전을 국가기관들이 하향식(Top-down)으로 실행하는 방법론을 택하고 있다.

한편 중국 국가 표준화 관리위원회(國家標准化管理委員會)는 개인정보 보호를 강화를 위하여 '정보보안기술 개인정보 보안규범'(信息安全技術 个人信息安全規范)을 시행하기 시작했다.[113)] '개인정보 보안규범'은 국가표준으로 개인정보의 수집, 저장, 공유 방법뿐 아니라 사용자 동의 절차 등 세부 사항들을 담고 있다. 이 보안규범은 중국 내 개인정보가 미흡하다는 비판을 의식하여 마련된 것으로 시행 시점(2018년 5월)도 유럽연합 GDPR와 동일하다. 비록 규정 자체는 상세하지만 개인정보 보호 수준을 향상시킬 수 있을지 의문이다. 중국에서 국가표준은 높은 수준의 법률과 조치를 구현하는 도구로서 기능하지만 중국이 시행하고 있는 수백 가지 사이버 보안표준들에 '개인정보 보안'을 추가하는 방식은 부적절하다는 비판도 있다.[114)]

---

109) Campaign to Stop Killer Robots https://www.stopkillerrobots.org

110) "Deciphering China's AI Dream" Future of Humanity Institute, (Mar 1, 2018) p.3

111) 인공지능 관련 분야의 제조혁신 센터 건립 촉진, 핵심 연구소 설립, 합리적 데이터 공개, 주요 산업 지원, 핵심적인 인공지능 응용분야 확대 지원, 인공지능 산업의 혁신촉진, 인공지능의 핵심 역량의 획기적 향상, 인공지능 산업 지원 시스템의 구축이 핵심적 내용이다.

112) 《促進新一代人工智能産業發展三年行動計划 (2018-2020年)》解讀 참조

113) 2017년 12월에 공개하고 2018년 5월 1일부터 시행에 들어갔다.

114) Samm Sacks "New China Data Privacy Standard Looks More Far-Reaching than GDPR" Center for Strategic and International Studies (January 29, 2018)

제 4 절 기업차원에서의 인공지능 윤리 논의 동향

글로벌 IT기업들이 인공지능 윤리 원칙을 제시하는 이유는 인공지능 활용 과정에서 발생하는 위험이나 부작용, 기술 문해력이 없는 대중들의 반감이나 사회적 불신을 방지하려는 전략으로 볼 수 있다. 인공지능의 연구개발 주체가 인공지능으로 제기되는 윤리적 우려와 잠재적 부작용을 사전에 인식하고 대응하고 있으므로, 인공지능 거버넌스는 '자율적인 규제'를 통해서 가능하다는 공감대를 확산하려는 의도로 풀이된다.

1. 인공지능 파트너십(Partnership on AI)

가. 출범 계기

2016년에 출범한 인공지능 파트너십은 글로벌 기업들의 개방적 협력 사례로서 구글, 마이크로소프트(MS), 아마존, IBM, 페이스북이 주축으로 만들어 졌다. 2017년에는 애플(Apple)도 참여하고 있다. '사람과 사회에 이익을 주는 인공지능 파트너십'이라는 뜻의 '파트너십온AI'는 인공지능 기술의 활용단계에서 발생하는 법적, 윤리적 도전과제를 해결하고 인공지능이 사람과 사회에 미치는 영향에 대한 토론이 가능한 열린 플랫폼을 제공한다.[115)]

나. 주요 활동

인공지능 파트너십은 다중 이해당사자(multistakeholder) 조직으로 IT 기업들뿐만 아니라 인공지능 학계, 정책・윤리 전문가들도 참여하고 있다. 산업계 리더, 학자, 비영리 단체 및 전문가와 함께 최선의 인공지능 관행을 공동으로 개발한다. 인공지능에 대한 대중의 인식을 향상 시키며 인공지능 기술에 대한 모범 사례를 연구하고 공식화하는 역할을 한다.[116)]

- 인공지능 기술의 연구, 개발, 테스트 분야에서 최상의 실행 방법과 접근 방식을 개발, 공유
- 핵심 기술, 잠재적 이익 및 비용 등 인공지능에 대한 대중의 이해를 증진
- 인공지능 미래에 대한 토론과 참여를위한 개방적이고 포괄적 플랫폼을 제공하고 주요 이해 관계자가 이러한 중요한 대화에 완전히 참여할 수있는 지식, 자원 및 전반적 역량을 갖도록 보장

---

115) The Partnership on AI https://www.partnershiponai.org/

116) https://www.partnershiponai.org/faq/

• 사회 친화적인 인공지능의 활용을 모색

다. 분석

인공지능이 활용 과정에서 차별, 오작동, 편향, 인체에 대한 손상 등 부작용이 빈번히 발생한다면 인공지능 거버넌스를 위한 규범적 수단이 요구될 것이다. 그러나 인공지능 산업계는 강력한 법적 구속력을 가진 경성규범(hard law)이 등장하기를 원하지 않는 입장이다. 인공지능 생태계의 확대와 투자가 위축될 수 있고, 인공지능 응용 제품 및 서비스는 출시전에 공식적 사전 테스트를 거치거나 사후적 검증 절차 등이 필요해 진다면 부담이 되기 때문이다. 테스트와 검증 과정에서 인공지능 관련한 영업비밀과 중요한 알고리즘 설계기술이 그대로 노출될 위험도 크다고 할 수 있다. 인공지능 산업계는 데이터와 알고리즘 작동방식에 대한 문해력(algorithm literacy)이 부족한 규제정책의 입안자들이 인공지능 윤리·규범 정책을 제시하고 규제하려 든다면 기술 현실을 도외시할 수 있다는 우려를 제기한다.[117] 반면에 인공지능이 사회, 경제 전반에 미치는 영향이 막대한데도 불구하고 아무런 감시도 없이 기업들에게 과도한 인공지능 통제권을 허용하는 방식은 곤란하다는 목소리도 높다.[118]

2. 마이크로소프트 에테르(Aether)

가. 주요활동

마이크로소프트는 2017년 인공지능 분야의 난제 해결에 집중할 새로운 연구 및 인큐베이션 허브 '마이크로소프트연구소AI(Microsoft Research AI)'를 신설하기로 했다. 인공지능 윤리를 전담하는 조직 에테르(AETHER: AI and Ethics in Engineering and Research)가 2018년 3월에 출범하였다. 에테르는 인공지능 윤리를 연구하는 조직이자 사내 위원회 형태이며 인공지능 플랫폼과 경험을 기업의 핵심 가치와 원칙에 뿌리 내리고 사회에 이익이되도록 활동하고 있다. 한편 인공지능 엔지니어와 개발자를 대상으로 개설된 AI 전문 교육과정(Microsoft Professional Program in AI)에는 인공지능 윤리(Ethics for AI)가 기본과정으로 포함되어 있다.

117) Responsible AI relies on Data Literacy (June 08, 2018) https://2018.semantics.cc/responsible-ai-relies-data-literacy

118) Lee Rainie, Janna Anderson "Code-Dependent: Pros and Cons of the Algorithm Age" Pew Research Center (Feb 8, 2017)

나. 인공지능 윤리 원칙의 마련

마이크로소프트는 네 가지 인공지능 윤리 원칙을 구현하겠다고 밝혔다.[119)]

- 공평 (Fairness): 인공지능은 존엄성을 파괴하지 않으면서 효율성을 극대화하여야 한다. 인공지능의 자동화된 판단에 존재할 수 있는 편향성에 대비해야 한다.
- 책무성 (Accountability): 인공지능 알고리즘은 책무성을 가져야 한다.
- 투명성 (Transparency): 인공지능은 투명해야 한다.
- 윤리 (Ethics): 인공 지능은 인류를 돕고 지적 개인 정보 보호를 위해 설계되어야 한다.

다. 인공지능 안면 인식기술에 대한 규제[120)]

2018년 7월 마이크로소프트는 “안면 인식기술: 공공 규제와 기업의 책임에 대한 필요성”을 발표하면서 정부 규제를 촉구하고 나섰다. 안면 인식기술은 사생활과 표현의 자유와 기본적 인권 보호에 핵심적 문제를 불러일으키므로 기술 회사의 책임이 커진다는 점을 인정하면서 정부 규제와 기업들에게 수용이 가능한 규범의 개발을 요구하였다. 그 취지는 안면 인식기술과 같이 공공안전과 민주적 자유 사이에 적절한 균형을 요구하는 문제에 대하여는 기업의 자율적 판단이 아니라 시민들에 의해 선출된 연방 정부와 의회가 그 역할을 담당해야 한다는 것이다. 기술 사용을 관리하는 효과적인 방법은 정부가 적극적으로 사용 자체를 관리하는 것이고 기술이 사회 전반에 걸쳐 더 광범위하게 사용될 우려가 있는 경우, 정부의 역할이 필요하다는 점을 강조하였다. IT기업이 이처럼 공개적으로 규제의 필요성을 강조한 사례는 찾아보기 어렵다. 공개적 규제를 촉구했음에도 불구하고 정부 역시 국가안보 등을 이유로 안면인식 기술 사용을 확대하고자 하는 입장이어서 규제가 도입될 수 있을지는 의문이다.

---

119) Microsoft AI principles https://www.microsoft.com/en-us/ai/our-approach-to-ai

120) Brad Smith “Facial recognition technology: The need for public regulation and corporate responsibility” (Jul 13, 2018) https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/

3. 딥마인드 윤리 사회(DeepMind Ethics & Society)

가. 출범 배경

1) 조직

딥마인드(Deepmind)는 2014년 구글에 인수됐으며 알파벳(Alphabet) 그룹의 일부다. 딥마인드는 2017년 10월 인공지능이 제기하는 윤리적 이슈와 사회적 영향의 분석을 담당하는 DMES(DeepMind Ethics & Society)를 신설하였다.[121] DMES는 "모든 인공지능 제품은 사람의 합리적인 통제 아래 있을 때 최대 가치를 갖는다"는 전제 하에 미래의 인공지능 연구개발에 반영하기 위해서 인공지능 윤리를 연구하고 있다.

2) 영국 국민건강서비스(NHS) 데이터 이용과 불법 논쟁

딥마인드는 영국 국민건강서비스(NHS)과 환자 의료정보 공유 협약을 맺고 신장질환을 모니터링하는 스마트폰 헬스케어 앱 스트림스(Streams) 개발을 진행해 왔다. 그러나 영국 개인정보위원회(ICO)는 2017년 7월 딥마인드의 연구와 국민건강서비스(NHS)의 개인정보 처리 방식이 영국 개인정보보호법 규정을 위반한다는 결정을 내리고 시정을 권고하였다. "환자들은 딥마인드가 자신들의 진료기록·약물 과다복용 여부 등 보호받아야 할 정보까지 열할 수 있단 사실을 공지 받지 못했다"고 결정이유를 밝혔다.[122] 그러나 이 사건은 국민건강서비스(NHS)의 부실한 정보 관리와 환자들과의 소통 방식에서 비롯됐다는 지적이 많았다. 국민건강서비스(NHS)가 영국 내 다른 병원들과 체결한 환자 정보 공유 계약의 관리 상태에 문제가 있었던 것이다. 이 사건 이후 NHS는 정보 보호 업체를 통해서 환자들의 데이터와 정보를 비식별화하는 처리 시스템을 도입했다.[123] 딥마인드는 인공지능의 과학분야 활용 단계에서 불거지는 윤리 논쟁에 대비하고자 DMES을 신설하였다.

---

121) DeepMind Ethics & Society https://deepmind.com/applied/deepmind-ethics-society ; DMES(DeepMind Ethics & Society)가 공개한 인공지능 윤리연구를 수행하면서 준수하는 원칙들은 다음과 같다. ① 글로벌 사회와 환경에 도움이 되는 방식으로 기술을 개발한다 ② 엄격하고 철저히 증거에 기반을 둬 연구한다. ③ 자금 조달 방식 방식 등 모든 과정은 투명하고 개방적이어야 한다 ④ 다양한 목소리를 담고 학구적이어야 한다 ⑤ 포괄적으로 협업한다.

122) Information Commissioner UK "Royal Free NHS Foundation Trust-Google DeepMind trial failed to comply with data protection law" (July 2017)

123) NHS England, De-Identification Data Items https://www.england.nhs.uk/publication/de-identification-data-items/

나. 인공지능 윤리 연구분야

DMES는 인공지능 윤리와 관련된 연구분야를 다음과 같이 정하고 연구를 진행하고 있다.[124)]

- 개인정보 보호, 투명성, 공정성(Privacy, Transparency and Fairness)
- 경제적 영향: 포괄성과 공평한 이용 (Economic impact: Inclusion and Equality)
- 인공지능 거버넌스와 책무성 (Governance and Accountability)
- 인공지능 위험의 관리: 오남용 의도하지 않은 결과(Managing AI risk: Misuse and Unintended Consequences)
- 인공지능의 도덕성과 가치(AI Morality and Values)

다. 분석

DMES는 인공지능 산업계가 운영하는 윤리 전담조직 가운데 가장 체계적으로 인공지능의 윤리적 이슈와 사회적 영향을 다루고 있다. DMES는 단지 딥마인드 내부의 기술자들에게 윤리를 실천할 수 있도록 돕는 역할에 그치지 않고, 인공지능이 글로벌 사회에 미치는 영향을 예견하고 방향을 제시하여 모두의 이익을 위해 효과를 발휘할 수 있도록 도움을 준다는 목표로 활동하고 있다. DMES는 인공지능이 단지 기술적 효용성에만 치중하지 않고 사회의 우선순위와 관심사에 따라서 형성되고 사용되어야 한다는 지향점을 가지고 있다. 인공지능을 폐쇄적인 연구로 진행하지 않고 인공지능 기술의 향후 영향에 대한 집단적 반응을 면밀히 조사하여 설계에 반영하는 역할을 담당한다.

4. 구글(Google)

가. 주요활동

1) 군사용 드론 연구 논란

구글은 미국 국방부와의 공동 연구 프로젝트 ‘메이븐(Maven)’를 진행하면서 인공지능 플랫폼(텐서플로 프로그래밍 키트)을 제공하였다.[125)] 제공된 기술은 객체 인식을 지원하는 오

124) https://deepmind.com/applied/deepmind-ethics-society/research/

125) 텐서플로(TensorFlowh)는 구글 독자 머신러닝 엔진으로 알파고를 비롯해 검색, 음성인식, 번역 등 구글 서비스에 적용된다. https://www.tensorflow.org/tutorials/ 참조

픈소스 텐서플로 API로서 특정인의 이미지를 구분하기 위한 것이다. 구글은 이러한 인공지능 기술이 공격용으로는 사용되지 않는다고 설명하였지만 메이븐 프로젝트가 인공지능을 드론 등 첨단무기에 도입하기 위한 기술제공이라며 반대해 왔다. 그 실체는 무인 항공기가 수집한 영상 정보를 머신러닝으로 자동 분석해 타격 목표의 정밀도를 높이는 시스템 이었다. 4000명 이상의 구글 직원들이 메이븐 프로젝트를 폐지하라는 청원에 참여하였다.[126)]

2) 인공지능의 공익적 활용

구글이 드론을 이용한 인명 살상 등 군사적 목적으로 인공지능 기술을 국방부에 지원한다는 윤리적 논란이 크게 불거지자 구글은 메이븐 프로젝트를 중단하고 2018년 6월 인공지능 활용에 대한 구글의 원칙을 공개적으로 발표하였다.[127)] 이와 함께 인공지능 기술을 책임있게 개발하고, 인공지능 기술을 활용하지 않는 특정한 분야를 확립하겠다는 약속을 제시하였다. 또한 구글은 인공지능이 중요한 문제를 해결하고 일상 생활에서 사람들을 돕는 기술을 개발하는데 사용된다는 점을 적극적으로 홍보하기 시작했다. 그 핵심은 인공지능 기술이 혁신을 촉진하고 전 세계 정보를 체계화하여 유용한 정보를 보편적으로 접근가능하고 활용할 수 있다는 것이다.

나. 구글의 인공지능 활용 목표

1) 사회적 유익성

인공지능은 의료, 보안, 에너지, 운송, 제조 및 엔터테인먼트 등 분야에서 혁신적 영향을 미칠 것이므로 광범위한 사회적, 경제적 요인을 고려하여 전반적 이익이 예상되는 위험과 단점을 상회한다고 판단될 때 활용할 것이다. 국가의 문화, 사회규범, 법규범을 존중하면서 고품질의 정보를 쉽게 이용할 수 있도록 노력한다.

2) 불공정한 편향의 생성·강화의 금지

인공지능 알고리즘과 데이터 세트는 불공정한 편견을 반영, 강화 또는 감소시킬 수 있다. 공평하고 공정한 판단은 구단순하지 않고 문화와 사회에 따라 다르므로 구글은 인종, 성별, 국적, 소득, 성적 취향, 능력 및 정치적·종교적 신념과 같은 민감한 특성에 대한 부당한 영향력을 주지 않기 위해 노력할 예정이다.

---

126) Scott Shane, Cade Metz and Daisuke Wakabayashi “How a Pentagon Contract Became an Identity Crisis for Google” New York Times May 30, 2018)

127) Artificial Intelligence at Google: Our Principles https://ai.google/principles/

3) 강력한 안전 및 보안의 실행

인공지능이 유해한 위험을 결과를 야기하지 않도록 강력한 안전 및 보안 방법을 개발하여 실행할 예정이다. 인공지능 시스템을 적절하게 신중하게 설계하고 인공지능 안전 연구의 모범 사례에 따라 개발할 것이다. 제약 조건이있는 환경에서 인공지능 기술을 테스트하고, 배포 후에도 작업을 모니터링할 예정이다.

4) 책임을 지는 인공지능 시스템

인공지능에 대한 피드백, 관련 설명 및 불만 제기를 위한 적절한 기회를 제공하는 시스템을 설계할 예정이다. 인공지능 기술은 적절한 인간의 지시와 통제를 받도록 할 것이다.

5) 개인정보 보호 원칙을 인공지능 설계에 통합하기

개인정보 보호 원칙을 인공지능 기술 개발 및 사용에 반영 할 예정이다. 사용자에게 통지하여 동의 기회를 제공하고, 프라이버시 보호 장치가 있는 아키텍처를 장려하며, 데이터 사용에 대한 적절한 투명성과 제어를 제공한다.

다. 인공지능 기술 활용이 금지되는 분야들

- 전반적으로 해악을 주거나 일으킬 가능성이있는 기술. (중대한 위해의 위험이 있다면 이익이 중대한 위험보다 우위 있다고 믿는 분야에서만 진행하고 이 경우에도 적절한 안전 제약을 부과)
- 인명 피해를 야기하거나 직접적인 손상를 가하는 목적의 무기 또는 기타 기술.
- 국제 규범을 위반하는 감시활동(surveillance)을 통해서 얻어진 정보를 사용하는 기술
- 국제법 및 인권 원칙에 위배되는 기술.

라. 분석

구글이 밝힌 구글의 인공지능 활용 원칙은 윤리와 사회적 가치를 고려하여 인공지능의 활용 영역을 신중하게 판단하겠다는 공개적 약속이다.[128] 즉, 인공지능 기술을 어디에든 유용하다는 이유로 사용하지 않고 사회적 가치를 존중하여 공공의 이익이 되는 방향으로 사용될 수 있도록 내부 기준을 확립하겠다는 의지로 이해할 수 있다. 주목이 필요한 부분은 인공지능 기술 활용이 금지되는 분야들이다. 이는 메이븐 프로젝트에 대한 비판을 회피하기 위하여 만들어진 원칙으로 판단된다. 구체적으로 살펴보자면 구글은 인체에 직접적 손상을 가하

128) Ibid

는 영역을 위하여 인공지능을 설계하거나 배포하지 않겠다는 점을 밝혔다.

그러나 앞으로 구글이 군사 분야에 인공지능의 활용을 절대적으로 금지하거나 미국 국방부와의 협력을 전면적으로 금지한다는 입장으로 보기는 어렵다. 위 조항들에 따르면 인공지능 활용으로 인한 이익과 위험의 비교형량 및 판단을 구글 스스로 진행하기 때문에 객관적 준거를 결여하고 있다고 볼 수 있다. 이는 자율적 규제의 한계라고 할 수 있다. 만일 구글이 밝힌 윤리적 원칙을 스스로 위반하기로 했을 때, 고용인에 불과한 개발자 혹은 기술자들이 거부하는 경우에 법적인 보호를 받을 수 없고 내부 고발자로서 공개하는 방식을 취할 수 있을 따름이다. 그러므로 인공지능 윤리 원칙은 반드시 강력한 처벌이 수반되는 것은 아니더라도 객관적으로 공식화된 준거로 정립할 필요가 있다.

5. 카카오의 알고리즘 윤리헌장

국내에서는 지난 1월말 카카오가 국내 기업으로는 처음으로 홈페이지를 통해 인공지능 알고리즘 윤리헌장을 공표해 크게 주목받았다. 이는 알고리즘의 개발 및 운영 방식이 우리 사회가 요구하는 윤리 기준에 부합하도록 노력한다는 선언으로, 사회적 차별 금지, 데이터 윤리, 독립성, 설명의무 등 최근 활발하게 논의되고 있는 인공지능 윤리의 핵심 원칙들이 포함되어 있다. 알고리즘 기술이 인간의 도덕적 가치와 윤리 원칙의 관점에서 인간과 조화를 이루어야 하고 궁극적으로는 인간에게 도움이 되어야 한다는 기업의 사회적 책임 윤리를 천명한 것이라 할 수 있다. 이는 인터넷 기업이 알고리즘에 영향을 받는 이용자들과의 지속가능한 신뢰 관계를 형성하는 데 매우 중요한 원칙 표명이라고 할 수 있다. 물론 여기서 이용자는 일반 사용자 뿐만 아니라 알고리즘과 관련된 서비스 및 비즈니스에 직간접적 이해관계를 갖는 다양한 파트너들을 포함하는 것이어서, 향후 알고리즘 규제체계 정립의 방향을 설정하는 데 중요한 전환점이 될 것으로 보인다. 지난 2017년 '지능정보사회 윤리 가이드라인' 등과 같이 정부 및 국회 차원의 규범 형성 노력은 있었지만, 민간 기업 차원에서 인공지능 윤리기준을 공표한 것은 처음이라 향후 인공지능 시대의 중요한 현안이 급부상할 '알고리즘 규제'의 주체, 범위 및 성격에 대해 성찰할 수 있는 계기를 제공했다고 볼 수 있다.

〈표 3－4〉 카카오 알고리즘 윤리 헌장

카카오 알고리즘 윤리 헌장
1. 카카오 알고리즘의 기본 원칙 카카오는 알고리즘과 관련된 모든 노력을 우리 사회 윤리 안에서 다하며, 이를 통해 인류의 편익과 행복을 추구한다.
2. 차별에 대한 경계 알고리즘 결과에서 의도적인 사회적 차별이 일어나지 않도록 경계한다
3. 학습 데이터 운영 알고리즘에 입력되는 학습 데이터를 사회 윤리에 근거하여 수집, 분석, 활용한다
4. 알고리즘의 독립성 알고리즘이 누군가에 의해 자의적으로 훼손되고나 영향받는 일이 없도록 엄정하게 관리한다
5. 알고리즘에 대한 설명 이용자와의 신뢰 관계를 위해 기업 경쟁력을 훼손하지 않는 범위 내에서 알고리즘에 대해 성실하게 설명한다.

제 5 절 국제기구(글로벌)차원에서의 인공지능 윤리 논의 동향

1. 경제협력개발기구(OECD)

가. 인공지능 정책의 분석과 개발

경제협력개발기구(OECD)는 인공지능 정책 분석, 대화 및 참여, 모범 운영 사례 확인을 통해 인공지능 기술과 응용 프로그램의 경제적 사회적 영향과 정책적 함의를 파악하는 작업에 상당한 노력을 기울이고 있다.[129] OECD의 활동에는 인공지능의 영향에 대한 측정방법 개선, 노동시장 변화의 분석과 대안 개발, 디지털 시대의 개인정보 보호정책, 인공지능 기반의 의사 결정에 대한 책임, 보안 및 안전 등 정책 문제들이 포함된다. OECD 디지털 경제위원회(Committee on Digital Economy Policy)는 2016 년부터 인공지능 원칙에 대한 권고안을 제시할 필요성을 논의하기 시작했다.

나. OECD의 분석 작업의 초점

2017년 10월에 열린 인공지능에 관한 OECD 컨퍼런스 〈지능형 기계 - 스마트 정책〉에는 기

129) 2017년 12월 인터넷 거버넌스 포럼(IGF)에서는 OECD AI 세션을 진행하였다.

술자, 고위 정책 입안자 및 시민 사회, 노동 및 비즈니스 대표들이 인공지능 정책에 대한 논의를 진행하였다. 그 주된 초점은 다음과 같다. i) 자율적인 인공지능 시스템과 머신러닝이 일상 생활의 일부가 되는 상황에서 어떤 종류의 정책 및 제도적 프레임 워크가 필요한가? ii) 인공지능 시스템이 경제와 사회에 널리 활용되고 있는바 인공지능은 사회 전반에 어떻게 이익을 보장 할 수 있을까? OECD가 진행하여 온 인공지능 정책 분석의 초점은 다음과 같은 질문으로 요약될 수 있다.[130)]

- 혁신 생태계와 규제 체계가 인공지능이 개발을 촉진 할 수 있는 방법은?
- 인공지능이 제공하는 기회를 통해 더 나은 삶을 어떻게 지원할 수 있는가?
- 중소기업을 포함한 기업이 어떻게 인공지능으로의 전환에서 기회를 얻을 수 있나?
- 인공지능이 불평등을 악화시키지 않도록 보장 할 수있는 방법은 무엇인가?
- 인공지능은 경쟁력, 혁신 및 지속가능한 성장을 어떻게 개선 할 수 있나?
- 정부와 기업은 시민들에게 더 나은 서비스를 제공하기 위해 AI를 어떻게 활용할 수 있나?
- 전환의 부정적인 영향을 최소화하면서 어떻게 시민, 교육자 및 기업이 미래의 일자리를 준비하는가?
- 인공지능을 사용할 때 편견을 완화시켜 모든 사람을 위해 봉사 할 수있는 방법은 무엇인니까?
- 인공지능 시스템은 어떻게 안전하고, 안전하며, 투명하게 책임을 질 수 있는가?

다. OECD 인공지능 관련 권고안의 준비

OECD의 디지털 경제 정책위원회는 인공지능의 신뢰와 채택을 촉진하기 위해 인공지능 발전을 위한 권고안을 준비하기로 합의했으며 이 작업을 위해서 전문가 그룹(AIGO)을 구성하였다.[131)] OECD 권고안은 법적 구속력은 없지만 회원국들의 정책 수립에 있어서 표준적 권고로 기능하므로 이행력이 확보되는 사례가 다수 존재한다.[132)] 디지털경제정책위원회(CDEP)

130) OECD initiatives on AI www.oecd.org/going-digital/ai/oecd-initiatives-on-ai.htm 참조

131) OECD creates expert group to foster trust in artificial intelligence (Sept. 13. 2018) www.oecd.org/going-digital/ai/oecd-creates-expert-group-to-foster-trust-in-artificial-intelligence.htm

132) OECD, “Ethical artificial intelligence 10 essential ingredients” www.oecd-forum.org/users/75928-dr-christina-j-colclough/posts/29527-10-principles-for-ethical-artificial-intelligence

는 2019년에 발표하는 것을 목표로 OECD 인공지능 원칙에 대한 전문가 그룹을 설립하기로 결정했다.디지털경제정책위원회(CDEP)는 인공지능의 개발과 윤리적 원칙을 개발하기 위해 다중 이해관계자들(multi-stakeholders)이 참여하는 토론을 추진할 계획이다. 또한 2019 년에 시작될 예정으로 있는 OECD AI 정책 전망대Observatory)는 인공지능 규제정책 및 공공정책에 통찰력을 제공할 것으로 보인다. 정책 전망대는 OECD뿐만 아니라 다양한 이해 관계자들과 함께 위원회를 구성하여 유망한 인공지능 응용 프로그램을 찾고 경제적, 사회적 영향을 분석하며 그 정보를 최대한 공유하는 역할을 할 예정이다.[133)]

2. 유엔 살상로봇 금지 캠페인(Campaign to Stop Killer Robots)

가. 주요활동

킬러 로봇은 자율살상 무기 LAWS(lethal autonomous weapons systems)로 불리며 인간의 조정이 없이도 목표를 식별하고 공격할 수 있는 시스템을 의미한다. 아직 인공지능은 군사 분야에서 방어용 감시 기능을 수행하거나, 인간 조작자의 판단과 행동을 돕는 역할에 그치고 있으나 완전한 자율무기 시스템이 능동적인 공격에 사용될 수 있는 가능성도 커지고 있다.[134)] 유엔군축연구소(UNIDIR)도 자율적 무기시스템의 증가 추세를 분석하여 보고서로 내놓았다. 유엔 살상로봇 금지 캠페인은 다양한 센서와 인공지능이 발전하면서 인간의 통제 없이 목표를 찾아 공격하는 방식에 우려를 표하고 캠페인을 시작했다. 인간의 생명을 빼앗는 결정은 절대로 기계에 위임해서는 안된다는 것이다.

나. 자율살상무기의 반대

자동화 무기체계는 이미 중국, 이스라엘, 한국, 러시아, 영국 및 미국을 비롯한 다수의 국가들에서 개발되거나 배치되어 있다. 유엔 살상로봇 금지 캠페인은 국제사회에 문제의 심각성을 알리고 금지규범의 제정을 촉구하는 활동을 진행하고 있다.[135)] 주로 유럽연합의 개별 국가들의 의회를 대상으로 자동화된 무기체계의 개발을 금지하는 법안이나 결의안을 촉구

---

133) http://www.oecd.org/going-digital/ai/oecd-initiatives-on-ai.htm

134) UNIDIR “The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence” (Sep 4, 2017)

135) Campaign to Stop Killer Robots www.stopkillerrobots.org

하는 활동을 벌이고 있다. 도살자로봇(Slaughterbots)'이라는 제목의 동영상 홍보물을 통해서 자율 살상무기의 치명성을 홍보하기도 하였다.[136] 그러나 유엔 특정재래식무기금지협약(CCW)의 범주에 자율살상무기 시스템을 포함시키는 활동에서는 성과를 거두지 못하고 있다.

다. 유엔 특정재래식무기금지협약(CCW)과의 관계

국제조약법에 따른 특정 무기의 금지 또는 제한은 유엔 특정재래식무기금지협약(CCW)에 의해서 규율되고 있다.[137] 그러나 킬러로봇 등 자율살상 무기 시스템을 유엔 특정재래식무기금지협약이 금지하는 목록에 추가적으로 올리기 위한 국제적 합의는 이루어지지 못하고 있다. 유엔 군축위원회(UNDC)는 자율살상 무기(LAWS)와 관련하여 각국 정부의 전문가그룹(GGE)을 통해서 논의를 진행 중이다.[138] 국가별 입장은 브라질, 이라크, 우간다 등 세계 19개 국가들이 자율살상무기 시스템의 금지에 찬성하지만 호주 등은 반대하는 입장이다.[139]

위험한 자율살상무기의 형태가 무엇인지, 그 심각성에 대한 공감대가 형성되고 난 이후에 국제적 규제가 논의되어야 한다는 입장이다. 자율무기 금지에 반대하는 논리는 인간의 판단에 비해서 자율무기 시스템이 더 빠르고 정확하게 타격하므로 무고한 민간인을 살상할 수 있는 위험을 낮출 수 있다는 점 등이다.

인간 조작자의 관여가 전혀 없는 완전한 자율 무기(fully autonomous weapons)가 개발된다면 기계적 판단에 따라서 대량학살 등이 발생할 수 있고, 군사적 필요성의 원칙 등 국제 인도법의 기준들을 충족하지 못할 것으로 보인다.[140] 그러나 자동화된 무기 시스템에 대한 정확한 정의는 각 국가마다 약간씩 다르기 때문에 금지 규제가 도입되려면 '자동화의 수준'

---

136) Slaughterbots https://www.youtube.com/watch?v=ytfc_qcEIKw

137) CCW 특정 재래식무기 금지협약의 정식 명칭은「과도한 상해나 무차별한 영향을 초래하는 특정 재래식무기의 사용 금지 또는 제한에 관한 협약(Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects)」

138) 2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS) https://www.unog.ch/80256EE600585943/(httpPages)/7C335E71DFCB29D1C1258243003E8724

139) Mary Wareham "Support Grows for Killer Robots Ban US and Russia Among States Who Object to Prohibition" Human Rights Watch (Sept. 5, 2018) ; "Country Statements on Killer Robots" Campaign to Stop Killer Robots (March 2014) www.stopkillerrobots.org/wp-content/uploads/2013/03/KRC_CountryStatus_14Mar2014.pdf

140) Campaign to Stop Killer Robots "Retaining human control of weapons systems" www.stopkillerrobots.org/wp-content/uploads/2018/03/KRC_Briefing_CCWApr2018.pdf

‘인공지능 시스템에 의한 자율적 판단’이 무엇인가에 대한 명확한 법적 개념정의와 ‘자동화된 살상무기군’의 상세한 분류부터 진행되어야 할 것으로 보인다. 자동화된 무기 시스템을 전장에서 사용하는 경우 자국민 피해가 감소하는 등 군사적 이익이 크기 때문에 국제조약 등을 통해 금지되기는 쉽지 않을 것으로 전망된다.[141)]

3. 인터넷 소사이어티(ISOC)

가. 주요활동

인터넷 소사이어티(Internet Society)는 인터넷의 발전을 위해 설립된 국제 비영리 기구로서 모든 인터넷 관련기구들 가운데 최상위에 위치한다. 인터넷 기술개발이나 운용 관리상의 제반 문제들을 총괄한다. 인터넷 소사이어티는 2017년 ‘인공지능 및 머신러닝에 대한 정책 보고서’를 발간하였다.[142)] 보고서의 내용 가운데 ‘인공지능 설계・사용의 윤리적 고려’와 ‘책임 있는 사용’은 인공지능 윤리・규범 정책과 관련성을 가지므로 살펴볼 필요가 있다. 인터넷 소사이어티가 밝힌 이하의 원칙과 권고사항은 국제전기전자학회(IEEE) 글로벌 이니셔티브가 2017년에 공개한 윤리적 설계(Version. 2)와 연결하여 이해할 필요가 있다.

나. 인공지능의 설계・사용에서 윤리적 고려(Ethical Considerations in Deployment and Design)

1) 원칙: 인공지능 시스템 설계자는 기술에 ‘이용자 중심의 접근’(user-centric approach)을 적용할 필요가 있다. 설계자는 인공지능이 시스템이 인터넷과 인터넷 이용자들이 보안 위험에 노출되지 않도록 ‘집합적 책임’(collective responsibility)[143)]을 부담한다.

2) 권고사항

- 윤리적 표준의 채택: 윤리적 고려를 위한 원칙 및 표준의 준수가 필요하며, 연구자들 및 산업계가 나아갈 발전 방향이 제시되어야 한다.[144)]

---

141) “How can international law regulate autonomous weapons?” Brookings Institute (April 10, 2018)

142) Internet Society “Artificial Intelligence and Machine Learning: Policy Paper” (April 27 2017.pdf)

143) ISOC, “Collaborative Security: An approach to tackling Internet Security issues” (2015), www.internetsociety.org/collaborativesecurit

144) 윤리적 표준의 사례는 IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

- 혁신 정책에서 윤리적 고려의 촉진 자금 조달의 선결적인 조건으로 윤리적 표준을 준수할 것을 혁신 정책에 포함시켜야 한다.

다. 책임 있는 사용(Responsible Deployment)

1) 원칙

인공지능 행위자(AI agent)는 자율적으로 행동하고 인간의 지시 없이도 시간이 흐르면서 자신의 행동을 조정(adapt)할 수 있는 능력이 요구된다. 인공지능이 배치되기 이전에 이러한 역량이 안전한지 여부를 점검하고 지속적인 모니터링을 해야 한다.

2) 권고사항

- 인간의 통제: 자율 시스템은 인간이 작동을 중단하거나 시스템을 종료할 수 있어야 한다. (“작동 종료 스위치”). 특히 인간의 생명과 안전에 위험을 미치는 인공지능 시스템 설계에 새로운 자율적 의사 결정 방식을 도입하는 때에는 인간의 통제를 통합해야 할 필요가 있다.
- 안전 최우선: 모든 자율 시스템은 사용에 앞서서 광범위한 테스트를 받아야 한다. 인공지능 행위자가 디지털 또는 물리적 환경과 안전하게 상호작용하고 의도한대로 작동되는가를 확실히 해두어야 한다. 자율 시스템은 작동 중에 모니터링되고 필요하면 업데이트 또는 수정되어야 한다.
- 개인정보의 보호: 인공지능 시스템은 데이터 책임성이 요구된다(data responsible). 인공지능 시스템은 필요한 데이터만 사용하고 데이터가 더 이상 필요하지 않다면 삭제해야 한다 (“데이터 최소화"). 인공지능 시스템은 데이터의 전송 중 및 유휴 중에 데이터를 암호화해야 하며, 인가된 사람에게만이 접근이 가능하도록 제한되어야 한다 (“데이터 접근 통제”). 인공지능 시스템은 개인정보법 및 모범 사례에 따라서 데이터를 수집, 사용, 공유, 저장해야 한다.
- 데이터 공급의 신중성: 인공지능 시스템에 제공하는 지침과 데이터에 대한 신중한 고려가 필요한다. 인공지능 시스템은 편향되거나, 부정확하거나, 불완전하거나, 오해의 소지가있는 데이터로 훈련되어서는 안된다.
- 인공지능 시스템의 안전성: 인터넷에 연결된 인공지능 시스템은 안전하게 유지되어야 한다. 인공지능 시스템과 인터넷 모두를 보호하기 위해 안전은 중요하다. 맬웨어

에 감염된 인공지능 시스템은 차세대 봇넷이 될 수 있다. 인공지능 시스템의 기기, 시스템 및 네트워크 보안에는 상향화된 표준이 적용되어야 한다.

- 책임있는 공개: 선의로 행동하는 보안 연구자는 기소 또는 법적 조치에 대한 우려 없이 책임감 있게 인공지능 시스템의 보안을 검사 할 수 있어야 한다. 보안 취약점이나 기타 설계 결함을 발견하면 그 문제를 해결가능한 최상의 위치에 있는 사람에게 책임감 있게 공개해야 한다.

4. 국제전기전자학회(IEEE) 글로벌 이니셔티브

가. 주요활동

1884년에 설립된 국제전기전자학회(IEEE)은 전기전자공학 전문가들의 국제 조직이자 사실상 표준화 기구(de facto standard-setting organization)이다. 주요 역할은 전기 전자기술 분야의 전문가 협회, 컴퓨터 통신, 전력, 에너지 기술 등의 표준을 정하고 공표한다. 최근의 흐름은 공식적 표준화기구에 비해서 사실상 표준화기구의 위상이 강화되고 있다는 점이다. 사실상 표준은 일반적 또는 공식적으로 국제적으로 정한 표준은 아니지만 실질적 대중성을 지니고 있어 시장원리에 의한 지배력을 가지게 된다.[145)]

나. 자율적 지능 시스템에 반영되어야 할 윤리적 고려

인공지능이 인간의 삶과 비즈니스에 어떤 영향을 주는가는 중요한 논의 주제가 되었다. 국제전기전자학회(IEEE)는 인공지능 기술개발에 있어서 인간의 복리(well-being)를 우선하기 위한 3가지 판단 도구를 표준화하는 작업을 진행 중이다. 일상에 사용되는 인공지능 및 자동화 시스템이 윤리적인가를 판단하는 표준을 제정하여 인간의 복리와 안전을 보장하겠다는 의도이다. 다시 말해서 점차 자동화 및 인텔리전트 시스템이 일상 생활에서 더 중요한 역할을 수행하게되므로 자동화된 시스템에 적용된 기술들은 윤리적 고려사항이 반영되어야 하고 이를 측정할 표준도 요구된다.

이러한 배경에서 국제전기전자학회(IEEE) 산하의 자율적 지능 시스템의 윤리적 설계를 위한 글로벌 이니셔티브는 지능형 제품, 서비스 개발을 위한 가이드라인을 개발하였다.[146)] 2017

---

145) “사실표준화기구” 2005년 정보통신표준화백서(2006. 2) 한국정보통신기술협회 www.tta.or.kr/data/androReport/whiteBook/2-3.pdf

년에 발표된 '윤리적으로 조정된 설계'(Ethically Aligned Design, EAD. Ver. 2)은 글로벌 이니셔티브 산하 13 개 위원회 의견을 반영하여 작성되었다.[147] '윤리적으로 조정된 설계'(Ver.2) 문서에 포함되어 있는 이슈, 권고안들은 자율적 시스템이 인간 사회의 안전, 복지, 이익을 증진시킬 수 있도록 윤리적 고려사항을 구체화하여 인공지능 기술개발의 방향을 유도하는 기준으로 기능한다고 할 수 있다.[148] 동 가이드라인이 표명하는 일반 원칙은 다음과 같다.

- 인권(Human Rights): 자율적 지능 시스템의 윤리적 설계, 개발, 실행은 국제적으로 인정되는 인권을 인정해야 한다
- 복리(Well-being): 자율적 지능 시스템의 윤리적 설계, 개발, 실행은 인간 복리의 측정을 우선시해야 한다.
- 책무성(Accountability): 자율적 지능 시스템의 윤리적 설계와 조작자는 책임과 책무성이 요구된다.
- 투명성(Transparency): 자율적 지능 시스템의 윤리적 설계와, 개발, 실행은 투명해야만 한다.
- 오남용에 대한 인식(Awareness of misuse): 자율적 지능 시스템의 윤리적 설계, 개발, 실행은 오남용을 최소화해야 한다.

다. '윤리적으로 조율된 설계(EAD)'와 법적 규제 권고

자율적 지능 시스템의 개발, 활용과 관련하여 많은 논의가 오가고 있지만 정작 법적으로 어떻게 다루어야 하는지는 아직 결정되지 않았다. 자율적 지능 시스템의 법적 규제와 관련하여 '윤리적으로 조율된 설계(EAD)'는 다음과 같은 4가지 영역게 걸쳐서 초점과 함께 권고안을 제시하고 있다.[149] 이하의 질문은 공개적 토론을 유도하기 위한 것이며 법률가 커뮤니

146) http://ethicsinaction.ieee.org/

147) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (Dec. 2017). "윤리적으로 조정된 설계: 인간의 복리을 우선하기 위한 인공지능 및 자동화 시스템 비전"(Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous systems) Ver. 2. https://standards.ieee.org/industry-connections/ec/autonomous-systems.html;IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2016) 윤리적으로 조정된 설계 초판은 https://standards.ieee.org/develop/indconn/ec/ead_v1.pdf

148) Alison DeNisco Rayome, "IEEE announces 3 AI standards to protect human well-being in the robot revolution" Tech Republic (Nov. 11. 2017)

149) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (Dec. 2017). "윤리적으로 조정된 설계: 인간의 복리을 우선하기 위한 인공지능 및 자동화 시스템 비전"(Ethically Aligned Design:

티의 참여 등 공개적 의견 수렴을 거쳐서 점차 구체화될 것으로 전망된다.

- 자율적 지능 시스템의 법적지위
- 정부가 자율적 지능 시스템을 이용하는 경우: 투명성과 개인의 권리
- 자율적 지능 시스템에 의해 초래된 해악의 법적 책임
- 자율적 지능 시스템의 투명성, 책무성, 확증가능성

또한, 최근 2018년 7월 IEEE와 IEEE 표준협회(IEEE Standards Association)는 자율 및 지능형 시스템의 윤리를 위해 공개 커뮤니티(OCEANIS, Open Community for Ethics in Autonomous and Intelligent Systems)를 9개 표준화 기구들(SDO)과 함께 시작하였다. 이는 정보통신 기술 분야에서 개발되는 자율 및 지능형시스템이 윤리적으로 조화될 수 있도록 표준을 개발하기 위한 협력을 촉진하는 글로벌 포럼이다. OECIANIS는 인공지능 기술과 관련한 비즈니스 및 정책결정을 지원하는 표준을 개발하기 위한 모든 관련 국내 및 국제 조직 간의 광범위한 협력을 도모한다. OECIANIS는 자율 및 지능형 시스템의 개발 및 보급과정에서 나타나는 복잡한 윤리적 문제는 비공개 협약을 통해서가 아니라 공동의 협의를 통해서만 해결될 수 있다고 전제하고 있다. 이는 IEEE의 자율 밑 지능형 시스템 윤리 글로벌 이니셔티브 (Global Initiative on Ethics of Autonomous and Intelligent Systems)의 방향과도 일치한다.[150]

5. 전미 컴퓨터학회(ACM)

가. 주요활동

전미 컴퓨터학회(USACM)는 컴퓨터 소프트웨어 연구 단체로서 2017년에 '알고리즘 투명성과 책무성에 대한 성명'을 발표했다.[151] 이 성명에는 인공지능 시스템의 소유자, 설계자가

Ver. p. 141-161

150) IEEE가 추진하는 OCEANIS가 처음 제기된 것은 2017년 봄이었고 기밀로 다루어던 사항이었는데 이번에 공개로 전환된 것이다. OCEANIS는 EAD 산하의 각 표준화 워킹 그룹들이 만든 표준들을 실제 표준으로 만들려는 움직임으로 볼 수 있다. 만약 IEEE가 EAD를 표준화할 경우 인공지능 분야의 각국 특허권 취득, 등록 및 산업 지배력에도 상당한 영향을 줄 것으로 예측된다. OCEANIS에 대한 자세한 소개는 https://standards.ieee.org/news/2018/oceanis.html 를 참조할 것

151) Association for Computing Machinery US Public Policy Council (USACM) "Statement on Algorithmic Transparency and Accountability" (Jan. 12, 2017) www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

고려해야할 사항들이 담겨있다. 2018년에는 컴퓨터 기술이 사회 전반에 영향력을 발휘하고 있으므로 기술자들의 책임이 변화되었다고 밝히면서 강화된 행동강령 및 전문가 행동(ACM Code of Ethics and Professional Conduct)을 공개하였다.[152] 성명과 행동강령은 자율적 가이드라인이어서 법적 구속력은 없지만, 인공지능의 윤리와 긴밀한 관계에 있는 알고리즘 투명성과 책무성을 위해 역할을 할 것으로 기대된다.

나. '알고리즘 투명성과 책무성에 대한 성명'

1) 인식(Awareness): (인공지능 알고리즘이 적용된) 분석 시스템의 소유자, 설계자, 사용자, 및 기타 이해관계자들(stakeholders)은 설계, 실행, 이용과정에서 발생할 수 있는 편향(bias)의 가능성들을 인식하여야 한다. 편향은 개인들과 사회에게 해악을 초래할 수 있다.
2) 이용과 수정(Access and redress): 규제자들(Regulators)은 알고리즘에 기반한 결정에 의해 악영향을받는 개인 및 그룹이 의문을 제기하고 바로 잡을 수 있는 메커니즘을 채택하도록 장려해야 한다.
3) 책무성(Accountability): 기관들은 자신들이 사용하는 알고리즘에 의한 결정에 대해 책임을 져야합다. 알고리즘이 어떻게 결과를 산출하는지 자세히 설명 할 수 없는 경우에도 마찬가지이다.
4) 설명가능성(Explanation): 알고리즘에 의한 결정을 사용하는 시스템 및 기관들은 알고리즘이 따르는 절차(procedures) 및 수행된 특정한 결정들(specific decisions)에 대한 설명을 작성하는 것이 좋습니다. 이러한 설명은 공공정책의 맥락에서 특히 중요하다.
5) 데이터의 출처파악(Data Provenance)[153]: (인공지능 알고리즘에 공급된) 훈련용 데이터가 수집된 방법에 대한 서술(description)은 알고리즘 작성자에 의해서 유지되어야하며 인간 또는 알고리즘에 의한 데이터 수집 절차(data-gathering process)에 의해 유도될 수

152) ACM Code of Ethics and Professional Conduc (July 2018) www.acm.org/code-of-ethics.

153) 데이터 프로비넌스(Data Provenance)는 데이터의 출처 및 근원정보를 의미하며, 데이터의 버전관리, 기관인증, 출처 추적 등을 의미한다. 데이터의 신뢰성(Trust) 문제를 해결하기 위한 메타데이터에 속한다. 백두권 "빅데이터 시대에서 메타데이터의 활용법은?" ICT Standard Weekly 한국정보통신기술협회 2015-01호 (2015. 1. 5); Boris Glavic "Big Data Provenance: Challenges and Implications for Benchmarking" Revised Selected Papers of the First Workshop on Specifying Big Data Benchmarks - Volume 8163 Pages 72-80, 2012

있는 잠재적인 편향들을 조사하여야 한다. 공개적으로 데이터를 조사한다면 수정할 수 있는 기회가 최대치가 될 수 있다. 그러나 개인정보 및 영업비밀 보호에 대한 우려, 분석방식의 노출(revelation of analytics) 우려 또는 악의적 행위자에 의하여 시스템 조작될 수 있는 우려가 있으므로 자격을 갖춘 허가된 개인들에게만 조사를 위한 접근을 제한 할 수 있다.

6) 감사가능성(Auditability): (인공지능) 모형, 알고리즘, 데이터 및 결정은 위험이 의심되는 경우 감사를 받을 수 있도록 기록되어야 한다.

7) 확증과 검사(Validation and Testing): 연구개발 주체들은 (인공지능) 모형을 테스트하기 위하여 엄격한 방법을 사용해야 하고 검사 방법과 결과를 문서화해야 한다. 특히, 모형이 차별적 해악(discriminatory harm)을 입히는지 여부를 평가하고 결정하기 위한 검사를 정기적으로 수행해야 한다. 검사 결과는 공개하도록 권장된다.

다. 전미 컴퓨터학회 윤리강령Code of Ethics)

전미 컴퓨터학회(USACM)가 2018년에 개정한 윤리강령(Code of Ethics)은 자율적인 행태준칙이며 강행성은 없다. 1조는 기본 윤리 원칙, 2조는 전문가의 책임에 대한 설명을 담고 있다. 3조는 리더십 역할을 수행하는 전문가를 위한 내용이다. 이 윤리강령은 컴퓨터 전문가들을 대상으로 발표되었고, 실무자들이 연구개발 과정에서 스스로 윤리적 결정을 내리도록 일정한 행동준칙을 제시하고 있다. 이하는 새로운 윤리강령 1조~2조 가운데서 인공지능 윤리와 관계 깊은 조항만을 선별하여 축약한 것이다.

1) 피해의 회피

1.1 컴퓨터 전문가들은 컴퓨터 기술로 인하여 발생하는 피해를 최소화 해야만 한다. 피해는 건강, 안전, 신체의 안전, 개인정보(프라이버시) 침해에 관련된 위협을 말한다.

1.2 ‘피해’는 정당하지 않은 물리적・정신적 손상, 정당하지 않은 정보의 파괴, 정보의 공개, 재산, 명성에 대한 정당하지 않은 손상 등이다. 실수 또는 의도하지 않았지만 피해가 발생한 경우, 그 행위의 책임자는 피해를 최대한 복구시키기 위해 노력해야만 한다. 피해로 이어질 수 있는 시스템 위험의 징후가 발견되면 즉시 알려야 한다. 만일 결정권자가 보고를 받고서도 아무런 조치를 취하지 않는다면 고발해야 한다.

2) 투명성과 공정성

1.3 컴퓨터 전문가들은 시스템 역량과 한계, 발생 가능한 문제들을 투명하게 완전히 공개

해야 한다. 이에 대하여 고의적으로 허위 주장을 하거나 오해를 불러일으키거나, 데이터를 조작하거나 위조해서는 안된다.

3) 차별에 대한 주의

1.4 새로운 기술의 도입과 활용은 기존에 없던 차별을 야기할 수 있다. 그러므로 새로운 사항을 결정하고 시도한다면 사후에도 관심을 유지하고 현상을 관찰해야 한다.

4) 개인정보의 존중

1.6 각종 기술의 발전으로 인하여 개인정보(privacy)는 가장 위협받고 있다. 최신 기술은 개인의 정보를 빠르게 수집하고, 모니터링하고, 교환한다. 이러한 데이터의 흐름 속에서 당사자도 모르게 개인정보의 침해가 일어날 수 있다. 그러므로 개인정보가 어떻게 수집되고 활용되는가를 이해하고, 정당한 권리를 가진 자들에게 그 과정을 설명해 주어야 한다. 개인정보의 사용은 법규제의 테두리를 벗어나서는 안 된다. 수집하고 활용하는 개인정보가 식별 가능한 정보로 변할 수 있는지 여부, 유통되는 과정 중에 유출될 수 있는지 여부를 주의 깊게 살펴야 한다. 데이터의 변질이나 불법적 접근, 실수로 인한 유출의 가능성도 최대한 줄여야 한다.

5) 대중 인식

2.7 대중들이 컴퓨터 관련 기술들을 보다 더 잘 이해하고 인지할 수 있도록 돕고, 알리고, 권장해야 한다. 특히 컴퓨터 시스템의 영향력과 한계, 취약점, 기회를 쉽고 친절하고 알기 쉽게 알려줄 수 있어야 한다.

6) 안전

2.9 컴퓨터 시스템을 설계하거나 도입할 때 원래 의도한 대로 기능을 발휘하도록 하고 사고나 악의에 의한 남용, 조작, 마비 상황에 미리 대비하고 최대한 차단해야 한다. 새로운 시스템의 도입으로 인하여 취약점들이 생겨나고 새로운 공격 가능성도 늘어난다. 그러므로 이에 대비하기 위한 장치 및 정책을 마련하는 것도 컴퓨터 전문가들의 책임이다. 여기에는 취약점 모니터링, 패치, 취약점에 대한 보고 등이 포함된다.

제 6 절 소결 및 정책적 시사점

1. 책임 있는 인공지능을 위한 인공지능 거버넌스

인공지능 윤리와 관련 해외 보고서와 국가별 인공지능 전략들을 종합하면 인공지능 분야 의육성과 더불어 책임 있는 인공지능(Responsible AI)을 위한 논의가 역동적으로 진행되고 있음을 파악할 수 있었다. 인공지능의 활용 과정에서 편익을 확대하고 신기술이 초래할 수 있는 해악과 부작용을 줄이기 위해 적절한 정책적 개입(policy interventions)이 필요하다는 점에는 점차 공감대가 형성되고 있다. 인공지능 분야에서 연구개발의 자유를 최대한 보장하는 자유방임(Laissez-faire)을 인공지능 윤리・규범 정책의 주류적 흐름으로 보기는 어렵다. 물론 인공지능 산업계의 자율규제 활동도 있지만 주요 국가들의 국가 전략은 인공지능 육성을 주안점으로 두면서도 인공지능을 기업들의 전적인 판단에 맡겨두거나 영리적 이해에 의해 좌우되지 않도록 윤리적 거버넌스 구축이 필요함을 전제로 하고 있기 때문이다.

산업계는 막연한 우려에 근거한 규제나 불명확한 규제는 비효율적이고, 연구개발의 자유를 막고, 4차 산업혁명 분야의 투자를 저해시킬 것이라고 우려하기도 한다. 그러나 인공지능 통제권을 기업에게만 맡기기에는 인고지능이 사회적, 정치적, 경제적 힘이 막대하다. 그러므로 인공지능 국가전략에는 인공지능 윤리 문제에 적절히 대응하기 위한 공적 거버넌스의 틀을 포함시키는 것이 타당할 것이다.

다만, 책임 있는 인공지능을 구현하기 위한 구체적 방법에 대하여는 여전히 논의가 필요하다고 할 수 있다. 한편, 인공지능 산업의 글로벌 각축전에 참여하는 국가들이 모두 개인정보 보호, 공정성, 살상용 자동화 무기(LAWS) 금지에 동의하지는 않고 있으며 각자의 접근은 상당한 차이를 보이고 있다.

2. 윤리・규제정책의 필요성: 주인－대리인'(principal-agent) 문제

아직 가시화되지 않은 위험을 지나치게 우려하거나 인공지능 알고리즘을 막연히 불신하는 시각도 문제일 수 있지만 인공지능 알고리즘을 설계하는 개발자만이 이해하는 정보 비대칭 상황은 '주인－대리인'(principal-agent) 문제를 구조화하므로 도덕적 해이가 발생할 수 있

다.[154] 양자는 상충되는 이해관계를 가지게 되므로[155] 경제학은 주인－대리인 관계를 불완전 정보로 인한 도덕적 해이의 문제로 다룬다. 대리인은 재량 범위 안에서 자기의 역량과 지식을 활용하여 업무를 처리하게 된다. 대리인은 정보비대칭을 이용하여 이익을 극대화하려는 인센티브가 있으므로 주인에게는 손실이 발생하게 된다. 이와 동일한 시장실패(market failure)가 인공지능 연구개발에서 나타날 수 있다.[156]

IT 기업들이 알고리즘이 기회 배분, 정치 과정, 공평성과 차별, 배제 등에 미치는 영향을 고려하기보다 플랫폼 경쟁과 이익확보에만 주력하게 된다면 윤리적 고려는 우선시되지 않을 것이다. 인공지능 설계와 매개변수 정보를 보유한 '대리인'(인공지능 개발 기업)과 모니터링이 가능한 기술 문해력을 갖추지 못한 '주인'(개인들 또는 소비자)은 정보비대칭 상황에 놓이게 된다. 인공지능 개발 기업들이 안전 의무와 사전적 검사를 이행하는지 여부를 알 수 없다면 도덕적 해이가 발생할 수 있다.[157] 그럼에도 불구하고 인공지능이 사회에 미치는 해악 가능성이나 부정적 파급효과를 검증하고 분석하기 위한 투자는 극히 미미한 실정이다.[158] 위험한 작업에 투입되는 고차원적 인공지능 또는 자동화된 기계적 판단이 윤리적으로 수용 가능한지를 제3자가 확인하고 사전에 검증하는 절차가 제도적으로 정착될 필요가 있다. 그러나 인공지능 개발사들이 알고리즘 자체를 영업비밀로 다루고 어떤 경우에도 비공개 한다면 이러한 절차는 시행하기가 곤란할 것이다.[159]

---

154) Sappington, David E. M. (1991). "Incentives in Principal–Agent Relationships". Journal of Economic Perspectives. 5 (2): 45-66

155) 김영세, 게임이론: 전략과 정보의 경제학 제8판, 박영사 (2018) 477~451면

156) 인간은 정서적 의사소통이 중요한 존재인데 긴밀한 의사소통이 불가능한 복잡한 인공지능은 필연적으로 인간에게 해를 입힐 것이라는 우려에 대하여는 다음의 논문을 참조 Matthias Scheutz (2012) "The Affect Dilemma for Artificial Agents: Should We Develop Affective Artificial Agents?"IEEE Transactions on Affective Computing, Vol 3, Issue4, 2012

157) Anderson, Michael; Anderson, Susan Leigh (2007). "Machine Ethics: Creating an Ethical Intelligent Agent". AI Magazine. American Association for Artificial Intelligence. 28 (4): 15-26 (Winter 2007)

158) "Building AI is like launching a rocket: Meet the man fighting to stop artificial intelligence destroying humanity" ZDNet, (March 5, 2015); 인공지능이 사회에 어떤 사회적, 제도적 영향을 미치는가를 측정하는 지표는 아직 개발되지 않았다.

159) Burrell, Jenna(2016). "How the machine 'thinks': Understanding opacity in machine learning algorithms". Big Data & Society, 3(1),2016.

3. 알고리즘 문해력의 필요성

불투명하고 잠재적으로 편향된 수학적 모델이 알고리즘을 통해 수많은 개인들의 삶에 영향을 줄 수 있는데도 개발자가 진지한 관심을 기울이거나 오류에 대한 책임을 지지 않고 정부도 관심이 없다면 부정적 효과는 걷잡을 수 없이 커지게 될 것이다. 예컨대 머신러닝도 인간 판단처럼 도덕적 위험과 오류를 재현하거나 심지어 더 확대 할 수도 있다.[160] 예컨대 데이터를 통해서 차별, 부정적 평가, 암묵적 편견도 학습될 수 있다.[161] 알고리즘 내부에 숨어있는 편견은 채용이나 기회의 배분에서 가난한 지역이나 사회적 약자를 차별할 수 있다. 그러나 알고리즘에 의한 자동화된 판단과정에서 편향, 오류, 무작위성이 발견된다면 일반인으로서는 기술적 문해력의 부족으로 인하여 그것이 과연 의도된 것인지 설계 상의 통제 불가능한 변수들 때문인지 알기 어렵다.[162] 이러한 입증책임의 문제는 데이터 문해력에 기인하는 것이다. 따라서 종래의 제조물책입법을 인공지능 관련 제품이나 서비스에 그대로 적용하기 어렵게 만드는 요인이기도 하다.[163]

크리스티안 샌드빅(Christian Sandvig), 캐시 오닐(Cathy O'Neil)은 인터넷 플랫폼과 다양한 인공지능 솔루션에 적용된 알고리즘의 검증과 조사를 통해 책무성(accountability by auditing)을 확보할 필요가 있다고 주장한다. 책임있는 인공지능의 구축은 기술 문해력에 의존하므로 입법자, 규제기관, 또는 알고리즘조사(audit)을 실시하는 주체는 알고리즘 문해력을 갖출 필요가 있다.[164] 이러한 의미에서 데이터 문해력은 책임있는 인공지능을 구현하고 알고리즘 책

---

160) Sendhil Mullainathan, Ziad Obermeyer "Does Machine Learning Automate Moral Hazard and Error" American Economic Review: Papers & Proceedings 2017, 107(5): 476-480

161) Paul Christiano et al., "Deep reinforcement learning from human preferences" (12 Jun 2017) arXiv:1706.03741 [stat.ML]

162) Will Knight "Forget Killer Robots－Bias Is the Real AI Danger" MIT Technologu Review (October 3, 2017); UNIRIR (2018) "Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies"

163) 이러한 점을 감안한 개정된 제조물책임법이 2017. 4. 18. 공포되었다. 제조업자 및 공급업자의 책임을 강화하고, 피해자의 입증책임을 완화하는 내용을 담고 있으며 2018. 4. 19.부터 시행 중이다. 피해자의 입증책임 완화규정(제3조의2)은 피해자가 3가지 간접사실을 입증하면, 제조물에 결함이 있었고(결함의 존재), 그 제조물의 결함으로 인하여 손해의 발생(결함과 손해 사이의 인과관계)을 추정하도록 하였다. 간접사실 3가지는 ① 해당 제조물이 정상적으로 사용되는 상태에서 피해자의 손해가 발생하였다는 사실, ② 그 손해가 제조업자의 실질적 지배영역에 속한 원인으로부터 초래되었다는 사실, ③ 그 손해가 해당 제조물의 결함 없이는 통상적으로 발생하지 아니한다는 사실이다.

164) Christian Sandvig et al, (2014) "Auditing Algorithms: Research Methods for Detecting Discrimination on

무성(algorithmic accountability)을 담보하기 위한 수단이라고 할 수 있다. 알고리즘 조사에 대한 규제가 마련되면 정부 또는 공익단체가 일정한 경우 알고리즘을 검증하여 알고리즘의 정확하고 공정한 작동을 압박할 수 있다.[165]

4. 개인정보 보호와 양립가능한 데이터 활용

인공지능 경제에서 데이터의 공급은 절대적으로 중요한 요소인데 캠브리지 애널래티카에 의한 페이스북(Facebook) 개인정보 유출, GDPR의 시행 등 규제 틀의 변화로 인하여 최근 데이터 프라이버시에 대한 인식이 증가하고 있다.[166] 많은 국가들이 인공지능 산업발전의 기회를 잡기 위해 데이터의 필요성을 인지하고 있다. 이러한 상황에서 개인정보의 활용을 위한 거버넌스 개선은 인공지능 산업의 발전을 좌우하는 요소로 부상하고 있다. 그러나 어떤 데이터가 '개인정보'로 보호되고, 어떤 데이터가 '공유 가능한가'의 구분이 명확하지 않다면 산업계에는 활용가능한 데이터를 공급받을 수 없고, 개인들은 개인정보가 유출될 것이라는 두려움을 가지게 된다. 이러한 대립적 관점은 거듭해서 지적되어 왔다.[167]

머신 러닝 알고리즘을 이용하여 데이터가 자동으로 수집하고 분석되지만 인공지능으로 인해 개인식별 가능한 데이터가 대량으로 수집되기 때문에 데이터 마스킹(Data Masking)에 주의를 기울일 필요가 있다.[168] 개인정보라고 해도 비식별화(de-identification) 조치를 거치면 사용이 가능하다고 할 수 있다.[169] 다만, 산업계가 요구하는 대량 데이터의 입수와 활용은 민

Internet Platforms" Conference proceeding at 'Data and Discrimination: Converting Critical Concerns into Productive Inquiry' of the 64th Annual Meeting of the International Communication Association. (May 22, 2014); Cathy O'Neil (2018) "Audit the algorithms that are ruling our lives" Financial Times (Jul 30, 2018)

165) Christian Sandvig et al, Ibid p. 17

166) 영국 정보위원회(Information Commissioner's Office)는 이 사건에 대한 조사 끝에 데이터 분석 결과를 공유한 업체들을 형사고발하고 벌금을 부과했다. Information Commissioner' Office. "Investigation into the use of data analytics in political campaigns" (11 July 2018)

167) 아주대 산학협력단 (2016) "인공지능 및 로보틱스 분야의 개인정보 보호에 관한 연구" 개인정보보호위원회.

168) 데이터 마스킹에는 데이터 난독화(Data Obfuscation), 암호화(Encryption)를 포함한다. Alyssa Provazza "Artificial intelligence data privacy issues on the rise" TechTarget(26 May 2017).

169) '개인정보 비식별화 조치 가이드라인'이 개인 정보의 보호에 우선하여 산업적 활용에 치중하고 있다는 비판이 제기되었다. 심우민, "개인정보 비식별조치에 관한 입법정책적 대응과제" 국회입법조

감한 개인정보가 안전하게 처리된다는 대중의 신뢰가 보장되는 때에만 가능할 것이다. 자동적인 의사결정 과정에 공급되는 개인정보는 원칙적으로는 개별적으로 동의를 얻어 적법하게 수집되어야 한다. 그러나 이 과정에 시간과 비용이 소요되므로 가명・익명처리된 개인정보가 재식별되지 않는다면 허용할 필요가 있다. 그러 비식별화를 위한 기술적 조치 방법이 있다고 해도 그 실행에 필요한 사회적 신뢰 수준을 확보할 필요가 있다. 우리나라는 유럽연합 소속은 아니지만 GDPR 적정성 결정(adequacy decision)을 받기 위한 준비단계에 있으므로 간접적으로 GDPR의 영향 하에 있다고 할 수 있다.[170]

5. 연성규범 제시 필요성

인공지능 시대가 본격화되면서 입법자로서는 인공지능이 현실적으로 야기할 수 있는 윤리적 논란들을 정확하게 인지하고, 이를 사전에 방지하거나・사후에 조정할 수 있는 가이드라인이나 규제를 도입할 필요가 있다. 기술적 설계 과정을 감안하여 인공지능과 관련한 윤리적 가이드라인 등을 연성규범(soft law) 형태로 제시한다면,[171] 규제의 불투명성・불확실성에서 비롯되는 부정적 외부효과(externalities)는 감소될 수 있을 것이다.[172] 다만, 인공지능 연구개발과 활용에 대한 정부의 직접 규제나 경성규범(hard law)의 집행은 사회전체적으로 비용이 과다하게 소요되므로 연성규범의 형태가 적절할 것으로 보인다. 인공지능 연구개발자를 대상으로 삼은 규제가 마련된다고 하더라도 효과적인 판단과 집행을 위한 기술 문해력을 규제기관이 충분히 갖출 수 있는지도 불확실하기 때문이다. 연성규범은 정식적인 법률제정 절차를 거치지 않고 발표되기에 법적 구속력이 없으나 아무 효력이 없지는 않다.[173] 즉 간접적으로 인공지능 산업계의 설계자 또는 개발자의 행위와 적용방식에 실질적 영향력을 미치

사처 현안보고서 305호-20170524) (2017. 5. 24); 한국인터넷진흥원, “개인정보의 안전성 확보조치 기준(2016-35호) 해설서(개정)” (2017. 1)

170) GDPR 제45조 ‘적정성 결정에 근거한 데이터의 이전’ 참조

171) Melvin Eisenberg, “The Architecture of American Corporate Law: Facilitation And Regulation”, 2 Berkeley Bus. L. J. 167, 182 (2005).

172) Ronald Coase (1960), “The Problem of Social Cost” Journal of Law and Economics, Vol. 3 (Oct., 1960). pp.17-19

173) 최난설현 (2013) “연성규범(Soft Law)의 기능과 법적 효력” 인하대학교 법학연구 제16집, 2호 (2013. 7) 87~123면 참조

는 행위규범(rule of conduct)이 될 수 있다. 유럽연합에서도 인공지능 윤리 가이드라인이 성립되면 회원국들에게 연성규범으로 제시될 것으로 예상된다.

6. 인공지능 윤리 가이드라인의 제정 주체

2019년을 전후하여 인공지능 분야의 선도 국가들은 각자의 인공지능 국가전략을 발표할 것으로 예상된다. 인공지능 산업 분야에 대한 국가별 접근방향과 청사진을 공개하는 것이 더 많은 투자를 유인하고 시장 참여자들이 행동을 결정하는데 필요한 신호를 줄 수 있기 때문이다. 인공지능 국가전략의 수립에 있어서 정책입안자는 인공지능 분야의 투자와 육성을 위한 전략뿐만 아니라 윤리적 논쟁점와 대중의 불안감·우려를 충분히 고려해야 한다.

우리의 경우 '유럽 인공지능 협의체'(The European AI Alliance)처럼 인공지능 산업진흥을 위한 전략수립과 별개로 인공지능 윤리 가이드라인 제정을 위한 광범위한 의견 수렴과 검토를 진행하도록 숙의형 포럼을 운영하는 방식도 고려할 필요가 있다. 이와 관련하여 인공지능 정책의 수립에는 다중이해관계자 모형에 입각해야 한다는 주장이 최근 주목받고 있다. 인공지능 윤리 가이드라인 작성을 위한 공론화 과정에는 인공지능 산업계, 정책전문가, 시민사회, 윤리학자 등 다중 이해관계자들(multi-stakeholders) 참여가 필수적으로 여겨지는 흐름이 뚜렸하다.[174] 요컨대 톱다운 방식(Top-down)이 아닌 숙의형 공론화 과정이나 다중이해관계자 참여형 포럼이 권장되고 있다. OECD 인공지능 이니셔티브와 인터넷 소사이어티(ISOC)도 다중 이해관계자에 의한 거버넌스(Multistakeholder Governance) 정립을 권장하고 있다.[175] 유럽연합 인공지능 협의체도 광범위한 다중 이해관계자들이 의견을 표명하고 토론하는 포럼(multi-stakeholder forum)으로 역할을 수행하고 있음을 참조할 필요가 있다.[176]

---

174) 대통령 직속 4차산업혁명위원회가 주도하는 규제· 제도혁신 해커톤은 이러한 맥락에서 순기능을 하고 있다. 해커톤, 대통령직속 4차산업혁명위원회 https://www.4th-ir.go.kr/hackathon/list

175) OECD initiatives on AI www.oecd.org/going-digital/ai/oecd-initiatives; Internet Soceity (ISOC) "Artificial Intelligence and Machine Learning Policy Paper" (April, 27 2017) p.27

176) The European AI Alliance의 참여자들은 포럼 방식(forum-style setting)으로 AI HLEG의 전문가들과 토론할 수 있다. AI HLEG가 작성한 초안 문서뿐만 아니라 비공개 및 공개질문에 대한 피드백도 제공된다. https://ec.europa.eu/digital-single-market/en/european-ai-alliance 참조

제 4 장 산업 분야별 인공지능 윤리 이슈 분석

제 1 절 인공지능 윤리 이슈의 분석 원리 및 검토대상 분야의 선정

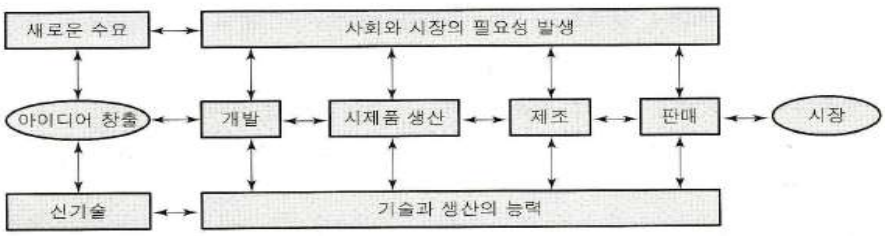
1. 기술의 생애단계와 행위자연결망(Actor-Network)

신기술의 생애단계와 관련하여서는 기술혁신론(시장견인모델, 기술주도모델, 커플링모델), 기술성장론, 기술수용주기모델, 제품수명주기모델 등이 연관되어 있다. 기술혁신론은 기술의 발명－마케팅－판매에 이르는 일련의 상업화 과정으로 요약되며 기술성장론은 신기술의 특허(발명)을 기준으로 '태동－성장－성숙－쇠퇴－회복'이라는 5단계 성장론을 제시한다. 또, 첨단기술제품이 초기 시장을 거쳐 대중화되는 과정을 설명하는 기술수용주기모델과 신기술의 도입－성장－성숙－쇠퇴를 순환구조로 설명하는 제품수명주기모델도 있다(이공래, 2001: 77-113; 김정홍, 2003: 12-14). 이러한 이론 및 모델들은 신기술과 관련하여 '탄생－상업화－시장 정착'이라는 일련의 생애주기를 전제하고 있다는 공통점이 있다.[177)]

본 장에서도 인공지능이 첨단 신기술이라는 점에서 이러한 이론과 모델들의 공통 전제를 수용하여 기술의 생애단계를 '기술의 생산～소비' 구도로 단순화한 분석 틀을 설정하였다. 그런데 기술의 생산～소비에 이르는 생애 과정과 연관을 맺고 있는 다양한 행위자 집단들은 기술의 각 단계마다 서로 다른 업무 특성, 역할(role), 이해 관심사(interest)를 가지고 있기 때문에 그에 따른 상호 논쟁과 충돌이 불가피하다. 예를 들어 인공지능기술의 의료적 활용은

---

177) 이 가운데 시장견인모델과 기술주도모델을 통합한 Rothwell & Zegveld의 커플링모델(coupling model)은 아래와 같은 도식(커플링모델에 의한 기술혁신의 피드백)을 제시하고 있다(김정홍, 2003: 13-14.에서 재인용).



AI닥터 논란을 부른 의료 인공지능 프로그램에서 디지털헬스케어, 건강보조기기 등에 이르기까지 매우 광범위하다. 이렇게 볼 때, 의료 인공지능과 상호작용하게 되는 행위자 그룹은 비단 의사, 간호사와 같은 의료집단에 국한되지 않음을 알 수 있다. 특히 보건의료기술이 고안－개발－출시(상용화)－수용(안착) 등 기술의 생산～소비에 이르는 일련의 생애주기를 따라 확산된다는 점을 고려할 때, 인공지능 의료기술과 연관되는 행위자 집단(group)은 상당히 많을 것으로 예측된다. 따라서 인공지능 윤리와 관련한 보다 광범위한 이슈들을 파악하기 위해서는 '신기술의 생애단계'에 어떠한 행위자 집단이 얽혀 있는지, 또 그들이 어떠한 방식으로 연결되어 있는지 살펴보아야 한다.

특히, 본 장에서는 기술의 생애단계별 행위자 역할을 분석하기 위해 행위자연결망이론(ANT: Actor-Network Theory)을 적용하였다. ANT에 따르면 사회적 네트워크는 다양한 행위자들이 연결(네트워크)되어 있는 형태인데, 여기서 행위자란 인간행위자 뿐만 아니라 비(非)인간행위자들을 포함하는 개념으로 정의되어 있다. 비인간행위자로는 기계, 기술, 부속품, 문서, 텍스트, 지식, 조직체계 등 사회의 구성원인 인간과 상호작용하는 '모든' 주체들을 말한다. 이들은 '행위자－연결망'의 단위체에서 각각 행위소로 기능한다. 그리고 이 모든 행위자들은 다양한 강도(느슨하거나 긴밀한 정도)의 연결망으로 네트워크를 형성하고 있으며, 행위자와 연결망은 서로가 서로를 구성한다. 즉, 행위자는 연결망 없이 행위를 할 수 없고, 연결망은 행위자들로 이루어져 있다는 것이다(김환석, 2006: 73; 브루노 라투르 외・홍성욱, 2010; Bloor, 1976(1991); Latour & Woolgar, 1979(1986); Latour, 1987).[178)]

그렇다면 산업 분야별 인공지능을 중심으로 하는 행위자연결망은 어떻게 짜여 져 있을까? 우선 신기술의 제안(발명)과 관련하여 과학기술자, 발명가, 기술혁신가 등이 주요 행위자로 자리 잡고 있을 것이며 기술의 인・허가 및 정책적 관리를 위해 정책・행정・법조계 관료들이 연결되어 있을 것이다. 또, 최첨단의 신기술이 등장하면서 몰고 올 사회적 변화와 윤리적 이슈의 파장을 이해・예측하기 위해 윤리학, 철학, 심리학, 사회학 등 유관 학계의 학자들도 관여되었을 것으로 판단된다. 여기에 비인간행위자로 인공지능기술 자체, 인공지능의 기술적 특성들(뉴럴 네트워크 시스템, 딥 러닝 등), 인공지능의 입력물에 해당하는 빅 데이터, 윤

178) ANT는 1980년대 중반을 전후해 과학기술사회학(STS) 분야의 이론으로 등장하였으나, 역동적 이론 구조와 특이성에 힘입어 2000년대 이후 유럽 등을 중심으로 금융, 지리, 경영, 생태 등의 분야에서까지 적극적으로 활용되고 있다.

리적 가치와 규범 등도 연관 될 수 있다.

그런데 ANT에서 또 하나 주목해야 할 이론적 전제는, 하나의 행위자－연결망 단위체에서 특정 행위소의 역할 취약성, 행위소 간의 연결고리의 단절이 발생할 경우 행위자－연결망 단위체의 안정화를 이루기 어려워진다는 내용이다. 이 전제를 적용하여 볼 때, 인공지능의 사회적 활용이 안정화에 이를지 여부를 예측하기 위해서는 각 기술 생애단계별로 어떠한 행위자들이 관여되어 있으며 그들의 역할 특성, 행위자 간 상호 연결고리 등을 파악해야 한다. 그리고 각 행위자들의 역할과 행위자 간 연결망의 취약정도를 면밀히 분석함으로써, 앞으로 발생할 가능성이 있는 이슈에 효과적인 대응책을 제안할 수 있을 것으로 판단된다.

2. 인공지능 윤리이슈 검토대상 분야의 선정: 제조, 의료, 금융, 국방

본 장에서 검토할 인공지능 윤리이슈의 대상은 제조, 의료, 금융, 국방 등 4개 산업 분야로 선정하였다. 구체적으로 제2장에서 도출한 산업별 인공지능 유스케이스 분석결과를 토대로 국내외 인공지능 제품/서비스별 산업 집중도가 높은 것으로 나타난 제조, 의료, 금융 분야를 선정하고, 최근 국제적・국가안보 차원의 법적・윤리적 쟁점을 야기하는 치명적자율무기체계를 대표하는 국방 분야 등 4개 산업 분야를 인공지능 윤리이슈의 검토대상으로 선정하였다.

제 2 절 제조 분야 인공지능 윤리이슈: 자율주행자동차를 중심으로[179)]

1. 논의의 배경

가. 인공지능의 상용화

상당수의 자동차 기업들은 2020년 전후를 자율주행자동차의 출시 목표 시점으로 삼아 기술 개발에 매진하고 있다(김대식, 2015). 이에 따라 지난 2015년 국토교통부는 자율주행자동차 상용화를 위한 지원방안을 발표하면서, 2016년 2월부터 시험운행허가제도를 마련하여

179) 본 절은 이상돈・정채연, “자율주행자동차의 윤리화의 과제와 전망”, 「IT와 법연구」 제15집, 경북대학교 IT와 법연구소, 2017, 281-325면; 이상돈・정채연, 『인공지능과 자율주행 자동차, 그리고 법』(7인 공저), 세창출판사, 2017, 55-93면에서의 논의를 해당 연구의 목적에 맞게 재구성하고 수정・보완한 것임을 밝혀둔다.

2018년 시범운행을 거쳐 2020년에 일부 상용화를 추진할 계획임을 밝힌 바 있다(국토교통부・산업통상자원부・미래창조과학부, 2015). 이러한 자율주행자동차의 상용화는 공리적 관점에서 정당화되고 있기도 하다. 예컨대 자율주행자동차의 도입으로 인한 운행시간 단축 및 연비 절감 등 교통 효율성 제고, 교통약자의 이동권 보장 등이 고려되기도 한다. 특히 자율주행자동차의 도입 정당성은 교통사고의 극적인 감소와 밀접하게 연결되어 있다. 현재 1년 동안 전세계적으로 발생하는 교통사고 사망자는 120만 명에 달하는 것으로 확인되고 있으며(한국포스트휴먼연구소・한국포스트휴먼학회, 2016), 기존 교통사고의 90% 이상은 음주운전, 졸음운전, 운전미숙 등을 포함한 인간의 과실 및 부주의에 의해 발생한다고 한다(이종영・김정임, 2015).

나. 자동차 사고의 일상성 및 정상성

자율주행자동차로 인해 상당 정도의 교통사고가 감소할 것으로 예측된다고 할지라도, 다양한 변수들이 상호작용하는 주행환경의 성격상 자동차사고는 완전히 제거될 수는 없는 것이다. 즉, 자율주행자동차 역시 일정 정도의 사고를 동반하게 될 것이다. 실제로 2016년 2월, 구글 카(Google Car)는 전방의 모래주머니를 피하기 위해 차선을 변경하여 지나가려 하였으나 옆 차선에서 주행 중이던 버스와 접촉사고를 일으킨 바 있으며, 이는 구글 카가 실질적으로 자신의 과실에 기인한 사고임을 인정한 최초의 사건이라고 할 수 있다.[180] 또한 2016년 5월, 테슬라(Tesla)의 모델 S는 오토파일럿(autopilot) 모드로 주행 중 대형트럭과 충돌하여 운전자가 사망하는 사고가 발생하였고, 테슬라는 밝은 낮이라 흰색의 트럭 측면을 구별하지 못하여 발생한 사고일 수 있다는 분석을 내놓은 바 있다.[181]

다. 인간사회에서 지능화된 의사결정의 영향력

이러한 일련의 자동차 사고는 자율주행자동차와 같은 인공지능의 지각, 판단, 행동이 우리의 생활세계에 분명히 영향을 미치게 된다는 점을 시사한다. 이 지점에서 인공지능의 법적 쟁점뿐만 아니라 윤리적 쟁점이 논의되기 시작한다. 현재 기술 수준을 고려할 때 인공지능의 상용화가 급속히 전개될 것으로 논의되는 분야는 자율주행자동차뿐만 아니라 소위 킬러로봇(killer robot)이라 지칭되기도 하는 군사용 로봇과 의료진단 및 수술에서 활용되는 의료

180) https://www.theguardian.com/technology/2016/feb/29/google-self-driving-car-accident-california

181) https://electrek.co/2016/05/26/tesla-model-s-crash-autopilot-video

로봇 등이며, 이들 모두 인간의 생명 및 신체에 대한 일련의 의사결정과정을 수반한다는 점에서 윤리적 요청이 강하게 제기되는 것이다.

라. 자율적 로봇으로서 자율주행자동차의 윤리적 지위

1) 로봇으로서 자율주행자동차

로봇은 인간이 특정한 목적을 위해 제작한 유용한 도구로서 인공물(artifact)이면서도(고인석, 2012a), 다른 인공물과 구별되는 독자적인 속성을 갖는다. 즉, 지능을 가진 인공물(intelligent artefacts)로서 지능로봇은 인간의 능동적인 정신을 모사하면서, 그 자체로 행동의 능동성을 지닌 존재라는 점에서 다른 인공물들과 구별되는 특성을 갖는다고 한다(고인석, 2014). 이렇듯 여타 인공물과 구별되는 지점에서, 지능로봇을 법윤리적으로 특별하게 취급해야 하는 필요성이 등장하게 된다.

지능로봇은 "외부 환경의 변화를 인식하고, 스스로 상황을 판단하여, 자율적으로 동작하거나 인간과 상호작용을 하는 로봇"으로 정의될 수 있다(김대식, 2015). 이러한 정의에서 알 수 있듯이 인지 센서를 통해 내・외부 환경을 인식하는 지각(sensing), 지능을 바탕으로 행동을 결정하는 판단 및 처리(processing), 이러한 결정을 실행하는 일련의 행동(acting)은 로봇기술의 기본적인 구성요소로 제시되어 왔다(김대식, 2015). 한편 자율주행자동차의 통상적인 정의(definition)로 "목표지점이 설정되면 인위적인 추가 조작 없이 스스로 주행환경을 인식하면서 목표지점까지 자율적으로 운행하는 자동차(이중기・황창근, 2016)", "일정한 조건하에서 스스로 주행환경을 인식하면서 자율적으로 운행하는 자동차(이중기, 2016a)" 등이 제시되어 왔다. 이러한 개념 정의에 따를 때,[182] 자율주행자동차는 지각(sensing)－정보처리(computing)－운동 출력(motion)이라는 로봇의 정의적 요소를 전형적으로 구현하고 있다고 하겠다(고인석, 2014). 특히 자율주행자동차의 경우, 도로교통체계 인프라와 총체적으로 결합된 형태로서 이해될 수 있다는 점에서 일종의 '교통통제 로봇시스템'과 결합되어 사회적 기반시설로 작동하는 '사회기반형 로봇'으로 분류될 수도 있다.[183] 인공지능이 체화된 지능

182) 이중기(2016b)는 "목표지점이 설정되면 인위적인 추가 조작 없이 스스로 주행환경을 인식하면서 목표지점까지 자율적으로 운행하는 자동차"로 정의내린 바 있으며, 자동차관리법 제2조 제1호의3에서는 자율주행자동차를 "운전자 또는 승객의 조작 없이 자동차 스스로 운행이 가능한 자동차"로 정의하고 있다.

183) 사회기반형 로봇의 예로서 교통통제 로봇시스템을 제시하는 고인석(2012b: 191) 참조할 것.

형 자동차 역시 로봇의 한 유형으로 분류될 수 있다면, 자율주행자동차를 물건으로서 자동차와 자율적인 지능을 갖춘 로봇이라는 이중적인 지위로 인식하는 것이 가능해지게 된다(이중기, 2016b). 즉, 자율주행자동차의 지능정보화는 “운전자가 운행하는 행위객체로서의 물건”이라는 지위와 “스스로 운전하는 행위주체”라는 지위를 동시적으로 인정할 가능성을 연다(이중기, 2016a). 결론적으로 로봇이라는 지위를 갖는 자율주행자동차 역시 로봇윤리의 범주 안에서 이론적・실천적으로 다루어질 수 있게 된다.

2) 자율주행자동차의 자율성에 대한 확장적 재해석

이렇듯 지능정보를 갖춘 자율주행자동차는 인간 운전자의 개입 없이 스스로 도로상황을 포함한 주행환경을 인식하고 의사결정하며 운행한다는 점에서, 적어도 현상적인 차원에서는 자율적인 지능로봇으로 볼 수 있다(이중기, 2016c). 자율주행자동차에 대한 법윤리적 쟁점을 다루는데 있어서 특히 자율주행자동차의 ‘자율성’(autonomy)은 중요한 고려요소라고 할 수 있다.

실제 자율주행자동차는 ‘자동화된 자동차’(automated vehicle)의 연장선에서 이해될 수 있다(이중기・오병두, 2016). 2013년 5월, 전미 고속도로 교통안전청(National Highway Traffic Safety Administration; NHTSA)은 자율주행자동차의 자동화 수준을 5단계로 분류하여 제시한 바 있다.[184] NHTSA의 자동화 수준을 기준으로 볼 때, 현재 시점에서 자동차의 자동화 기술은 2레벨과 3레벨 사이에 위치하는 것으로 평가되고 있으며(이중기, 2016a), 현재 진행 중인 기술 개발은 3레벨에서 4레벨에 해당하는 자율주행자동차의 상용화 단계를 목표로 추진 중

184) NHTSA의 자동차 자동화 수준을 정리하면 아래와 같다(http://www.nhtsa.gov).

0레벨	비－자동화 (No automation)	운전자가 자동차 조향, 제동 등에 대한 완전한 단독의 통제권한을 갖고 있음 운전자가 도로상황 모니터링 등에 대해 항시 전적인 책임을 지며, 자동차는 주행 관련 사항에 대한 경보 정도의 부수적・편의상 역할을 수행함
1레벨	특정 기능 자동화 (Function-specific automation)	자동차가 한 가지 이상의 통제 기능을 독립적으로 수행할 수 있는 자동화 단계
2레벨	통합 기능 자동화 (Combined-function automation)	두 가지 이상의 주요 통제 기능이 자동화되고 이들이 함께 결합되어 작동하는 단계(예: 크루즈 컨트롤과 차선 유지 장치의 결합) 운전자는 여전히 도로 및 주행상황에 대한 모니터링에 있어 책임을 지며, 항시 그리고 갑작스러운 통보에도 운행을 통제할 수 있도록 준비되어야 함

에 있다. 실제로 법윤리적 쟁점이 가시화되는 수준은 3-4레벨의 스펙트럼 사이에 위치하고 있는 자율주행자동차라고 할 수 있겠다.[185]

크루즈 컨트롤, 차선유지 장치 등이 장착되어 있는 기존의 '일반' 자동차 역시 인식, 판단, 주행에 있어서 상당 정도의 지능정보화 및 자동화가 이루어져 있다고 할 수 있지만,[186] 앞으로 도입될 자율주행자동차와의 실질적인 구별 기준은 운행자의 개입 비중 차이에 있다고 할 것이다. 지능로봇에 대한 윤리규범을 설정하는데 있어서 기본적인 두 축은 인공지능 알고리즘의 자율성과 인간 주체의 통제권이라고 할 수 있다(이원태, 2015). 기계의 자율성 역량과 인간의 통제 권한은 반비례의 관계에 놓여있다고 할 수 있다. 자율주행자동차의 개념을 이해하는데 있어서 운행자 없는(driverless) 자율적인 자기주행(self-driving) 로봇은 중요한 기준점을 제공해준다. 물론 자율주행자동차가 인간 운전자의 실질적인 개입 없이 운행을 지배할지라도, '지능'과 '자율성'은 상이한 개념으로 이해되어야 하며, 따라서 '자율'주행자동차라는 표현은 부적절하다는(misnomer) 의견 역시 존재한다(O. Etzioni, 2016). 사실상 자율주행자동차의 자율적인 판단이란 알고리즘에 따라 도출된 자동화된 판단이라는 점에서 진정한 의미의 자율성을 갖고 있다고 볼 수 없으며, 이에 따라 현재의 다수 의견은 로봇으로서의 자율주행자동차가 도덕 행위자로 인정될 수는 없다고 한다(윤지영・임석순, 2011). 다시 말해, 현재 기술수준에서 구현될 수 있는 자율주행자동차의 자율성의 정도를 인간의 자의식 및 자유의지와 동일하게 이해하여 윤리적 판단의 주체성을 적극적으로 인정하는 것은 이론적・실천적으로 무리가 따른다고 할 수 있다(이중기・오병두, 2016).

다만, 인간의 자율성과 동일한 수준으로 인정할 수 없을지라도 독립적으로 인식 및 행동

3레벨	제한된 자율주행 자동화 (Limited self-driving automation)	운전자가 주행 중 도로상황에 대해 항상 모니터링할 것이 요청되지 않으며, 다만 간헐적인 통제(occasional control)가 가능해야 함 특정 상황에서는 운전자가 자동차에게 통제권한을 양도할 수 있음 운전자에게 통제권한이 다시 전환될 때에는 운전자가 이를 충분히 확보하도록 적절한 시간이 주어짐
4레벨	완전 자율주행 자동화 (Full self-driving automation)	자동차는 운전자의 개입에 대한 기대 없이 전체 운행을 안전하게 지배함

185) 따라서 앞으로의 논의 역시 NHTSA 기준 3~4레벨의 자율주행자동차를 전제로 하여 전개하도록 한다.
186) 자동거리 조절기능 및 사고방지 감지장치 등을 예로 들 수 있다.

하는 인공지능의 자율적인 기능을 인정할 여지는 남아 있다. 자율성의 개념이 기존에 획득된 지식에 대한 분석을 바탕으로 독립적으로 판단하고 행동하는 것을 본질적인 내용으로 삼는다면(양종모, 2016), 자율주행자동차 역시 모종의 자율성을 획득하고 있다고 볼 수 있는 것이다. 다시 말해, 인간을 대신해 자율적 기능을 수행하는 (자의식을 전제로 하지는 않는) '의식 없는 지능'을 근거로 하여, 그것의 행위주체성을 인정하려는 이론적 시도가 이루어지고 있다는 점에 주목해볼 필요가 있다(한국포스트휴먼연구소・한국포스트휴먼학회, 2016).

자율주행자동차의 자율성은 인식－판단－행동이라는 일련의 과정에서 작동하는 알고리즘의 맥락에서 이해할 수 있다. 알고리즘이란 "컴퓨터가 소프트웨어를 통해 데이터를 처리하는 일련의 구조"를 가리킨다(한국포스트휴먼연구소・한국포스트휴먼학회, 2016). 프로그래머 등 인간의 작업을 수반하여 설계된 알고리즘에 따라 이루어지는 판단은 진정한 의미의 자율적인 판단이라고 보는데 분명한 한계가 있지만, 알고리즘 작동에 따른 행위의 결과를 프로그래머가 사전에 판단할 수 없는 지평이 존재한다는 점에서 알고리즘적 판단의 자율성이 인정될 여지가 발견되는 것이다. 즉, 알고리즘의 작동은 비가시성을 기본적인 속성으로 가지고(한국포스트휴먼연구소・한국포스트휴먼학회, 2016),[187] 인공지능의 알고리즘에 기반한 판단 및 행동을 모든 상황에 대해 예측할 수는 없으며(양종모, 2016), 이에 따라 알고리즘 작동방식에 대한 통제 권한을 온전히 확보하기는 어렵다는 점에서 자율적 판단에 대한 윤리적 우려가 시작되는 것이다.

2. 자율주행자동차의 윤리 이슈 및 쟁점

가. 사후적・사전적 차원의 윤리적 접근

1) 사후적 차원의 윤리적 접근: 자율주행자동차 관련 책임 귀속의 곤란성

먼저 자율주행자동차의 알고리즘에 따라 이루어진 '자동화된 결정'(automated decision)에 대한 책임 귀속 문제(responsibility ascription problem), 책임의 공백 내지 진공 상태가 발생하

187) 이로 인해 파생되는 일반 대중에 대한 비공개성 내지 비접근성의 문제는 설계자 및 제작자의 윤리적 책임을 좀 더 강화시키는 정책 방향성을 이끌기도 한다. 곧, 인공지능 알고리즘의 기술적 구현과정에서 공개성, 투명성, 책무성 등의 원리가 중요한 윤리적 요청으로 강조되고 있으며(이원태, 2015), 법정책적으로 기관윤리위원회(Institutional Review Board; IRB)의 역할 수행이 로봇윤리의 맥락에서 논의되고 있다.

게 된다. 위에서 언급한 바와 같이, ‘운행지배’에 있어서 인간 운전자와 자율주행자동차의 개입 정도는 대항적 관계에 놓여 있으며, 이때 자율주행자동차의 자동화 내지 자율성 수준은 운행자의 주의의무 위반 여부를 판단하거나 제조물책임의 법리를 적용하는데 있어서도 큰 영향을 미친다고 하겠다. 예컨대 실질적인 의미에서 운행지배를 하지 않거나 극히 제한된 범위 내에서만 통제권한을 가진 인간 운전자에게 운행자의 주의의무 및 운행자 책임을 온전히 인정하기 곤란한 지평이 존재한다. 특히 실제 충돌이 불가피한 상황에서 최대한 안전성을 확보하기 위해 내리는 의사결정은 사전계획 없이 매우 급작스럽게 고도의 스트레스를 동반하여 이루어진다는 점을 고려할 때, 더욱 그러하다고 한다(N. J. Goodall, 2014). NHTSA 기준 3레벨 이상의 자동화가 이루어진 경우 운전자는 주의를 집중한(attentive) 상태를 계속적으로 유지할 필요는 없으며 사전경보를 받고 일정 시간이 지난 후에나 자동차에 대한 통제권한을 되찾게 되는데, 이때 수초에 이르는 짧은 시간 안에 인간 운전자로 하여금 상황에 대한 전적인 통제권한 및 이에 상응하는 책임을 요청하는 것은 비현실적이기 때문이다(N. J. Goodall, 2014).

이렇듯 자율주행자동차의 경우, 인간 운전자의 운행에 대한 통제권한이 현저히 축소됨에 따라 운행자 책임에서 제조자 책임으로 전환되는 등 책임구조의 변화가 예측되고 있다는 점에서(이중기・황창근, 2016) 제조자의 윤리적 책무가 특히 중요하게 고려되고 있다고 하겠다. 즉, 자율주행자동차와 관련된 일련의 법적 책임이 운행자로부터 제조자에게로 전환되어야 함이 점차 설득력을 얻어 가고 있는 것이다(J.-F. Bonnefon & A. Shariff & I. Rahwan, 2015). 다만, 자율주행자동차의 프로그래머 및 알고리즘 개발자의 책임 귀속 역시 온전히 인정되기 어렵다는 점이 지적되고 있다. 일단 자율주행자동차의 알고리즘 소프트웨어가 기존 제조물책임법리상 제조물인지 여부에 대한 논란이 여전히 존재하고 있고(이중기・황창근, 2016), 위에서 언급했듯이 알고리즘 작동에 따른 판단 및 행동을 알고리즘 개발자 역시 완벽하게 예측할 수 없다는 점에서 개발자에 대한 책임 귀속이 불분명할 수밖에 없다(김대식, 2015). 이로 인해 결국 손해 및 피해는 발생하지만 “누구도 책임지지 않는 사태”가 발생할 가능성(이중기・황창근, 2016)이 존재하게 되는 것이다. 이에 대해 책임보험제도의 역할을 강조하는 등 인공지능의 행위 결과에 대한 책임 귀속의 문제를 해결하는 방식이 적극적으로 논의되고 있다. 이러한 논의 자체가 일련의 윤리적 고려를 수반하고 있다고 볼 수 있다. 즉, 손해의 공평한 분배 문제는 기본적으로 도덕적 영역 안에 속하는 것이며(J.-F. Bonnefon &

A. Shariff & I. Rahwan, 2015), 책임의 형평적 분배와 관련해서는 공리주의적 관점이 특히 많이 반영될 수 있다.

2) 사전적 차원의 윤리적 접근: 자동차 사고의 일상성 내지 정상성

자율주행자동차의 책임 귀속 문제가 사후적 차원에서의 윤리적 접근을 요청한다면, 자동차 사고의 일상성 및 정상성이라는 맥락에서 당위적으로 요청되는 자율주행자동차의 윤리적 프로그래밍은 사전적 차원의 윤리적 접근이라고 볼 수 있다. 로봇에 대한 윤리적 고려는 바로 그것의 판단 및 행동이 현실세계에서 일련의 작용과 반응을 가져온다는 점에서 요청되며, 자율주행자동차의 경우 외부세계에 대한 영향은 충돌 등을 동반한 사고에서 전형적으로 나타난다. 물론 자율주행자동차의 도입으로 인해 교통사고가 획기적으로 감소할 것으로 예측되고 있기도 하다. 이와 관련하여 보다 급진적인 견해에 따르면, 자율주행자동차는 절대 충돌하지 않을 것이므로 자동차의 자동화로 인한 교통사고는 거의 존재하지 않을 것이고, 따라서 자율주행자동차의 의사결정체계와 관련해 특별한 윤리적 고려가 요청되지 않는다고 한다.[188] 그러나 자동차의 기계적 장비 및 자율주행 소프트웨어 프로그램의 결함 내지 오작동으로 인한 교통사고의 발생 가능성은 분명히 잔재하고 있다(신동현, 2016). 나아가 자동차의 주행환경은 상당히 다양하고 가변적인 변수들의 영향을 받게 된다는 점을 고려할 때, 자율주행자동차의 모든 센서를 비롯해 알고리즘 기능이 완벽하게 작동할 때에도 일정 비율의 자동차 사고는 발생하게 될 것이다(N. J. Goodall, 2014). 즉, 완전무결한 자율주행자동차의 개발이 가능하다고 할지라도 자동차 사고는 일상적으로 발생하게 되며, 근본적으로 자동차 사고는 피할 수 없는(unavoidable) 것이다(P. Lin, 2016). 이러한 자동차 사고의 일상성 내지 정상성은 자율주행자동차의 사고에 대한 책임분배의 문제와 더불어, 자율주행자동차가 주

188) 자율주행자동차의 의사결정체계에서 윤리적 고려를 요청하지 않는다는 반론 역시 다수 존재하고 있다. 구달은 이러한 반론을 다음과 같이 9가지 주장으로 압축하여 정리한 바 있다: 1) 자율주행자동차는 절대 (혹은 거의) 충돌하지 않을 것이다; 2) 복잡한 윤리적 결정을 요청하는 충돌은 극도로 있음직하지 않다; 3) 자율주행자동차는 충돌에 대해 절대 (혹은 거의) 책임지게 되지 않을 것이다; 4) 자율주행자동차는 여타 자율주행차와 절대 충돌하지 않을 것이다; 5) 레벨2와 3에서는 인간이 통제권한을 항상 가질 수 있으므로, 인간 운행자가 윤리적 의사결정에 대한 책임을 지게 될 것이다; 6) 인간은 주행 중 혹은 충돌 시 윤리적 판단을 거의 내리지 않는다; 7) 자율주행자동차는 법을 준수하도록 프로그래밍될 수 있으며, 이를 통해 윤리적 상황에서의 문제는 해결된다; 8) 자율주행자동차는 단순히 손해를 항상 최소화하도록 시도하며, 이는 공리주의적 윤리 체계를 의미한다; 9) 자율주행자동차로 인해 창출되는 전반적인 이익이 비윤리적 자동차로 인한 위험보다 크다(N. J. Goodall, 2014).

행 중 예측되는 충돌상황 등에 대해 어떠한 의사결정을 내려야 하는지와 같은 윤리적 쟁점, 곧 자율주행자동차의 윤리적 의사결정체계에 대한 문제를 본격적으로 제기하게 된다. 단순히 미연에 자동차 사고를 예방하기 위한 충돌회피 알고리즘만으로는 자율주행자동차의 상용화가 가져오게 될 다양한 사회문제들을 해결할 수 없으며, 보다 구체적인 윤리적 고려가 요청되는 것이다.

나. 자율주행자동차의 행위주체성과 윤리적 책무성

로봇으로서 자율적인 인식 및 판단을 수행하는 자율주행자동차에 대한 적절한 윤리적 고려의 필요성은 그것의 자율성에 기반한 법윤리적 책임을 법리적・법정책적으로 가능하도록 하기 위한 적극적인 논의의 지평을 열었다.

먼저 로봇윤리 담론에서는 강 인공지능(strong AI)과 같은 수준의 자의식 및 자유의지를 상정하지 않고서도 자율적인 주체성을 논의하기 위한 다양한 이론적 탐구가 시도되고 있다. 다시 말해, 모종의 자율적인 행위능력을 갖춘 자율기술시스템의 독립적인 주체적 지위를 인정하는데 있어서, 반드시 인간 수준에 도달한 인공지능의 출현이 전제되는 것은 아니라는 입장이 설득력 있게 제안되고 있는 것이다(신상규, 2016). 고인석은 로봇이 철학적 의미의 진정한 주체성을 갖지 못한다고 할지라도, 로봇이 주체적인 역할을 수행하는 '현상적 차원'에서의 행위가 일상화될 것이며, 이로 인해 인간은 인공적 시스템에 생활세계의 판단 및 선택의 일부를 위임하거나 의존하게 될 것으로 예측한다(고인석, 2011). 자율주행자동차 역시 일종의 자동화된 인공적인 시스템이라고 볼 수 있으며, 이때 자율주행차는 가치 판단 및 이에 기반한 행위 실행을 부분적으로 위임받게 되는 것이다(고인석, 2011). 로봇의 행위에 대한 법윤리적 평가에 있어서 그것의 자율성 보유 여부 자체에 대한 논의로 깊이 들어가기보다는, 이렇듯 현상적인 측면이 강조되고 있다. 로봇의 '현상적 자율성', 즉 "로봇이 현상적 차원에서 자율적 주체인 것처럼 행동할 수 있다"는 데에 방점을 두고 로봇에 대한 윤리적 검토가 이루어지고 있는 것이다(고인석, 2011). 이러한 맥락에서 준－자율적(semi-autonomous) 내지 유사자율적(quasi-autonomous)인 존재로서 로봇의 지위가 주장되기도 한다(변순용・송선영, 2012). 자율주행자동차 역시 주변환경을 인식하고 판단능력을 보유하며 운행행위를 수행한다는 점에서 적어도 '현상적 차원'에서의 자율적인 주체성을 인정할 가능성이 열린다(이중기, 2016b).

다시 말해 여타 인공물과 지능로봇을 구별하는 기준으로서 현상적 차원에서 인간과 유사하게 수행하는 주체적 기능 및 행위의 능동성이 주목받고 있으며(윤지영・임석순, 2011), 이를 바탕으로 자의식과 자유의지가 인간과 동등한 수준은 아닐지라도 자율적인 인공 행위자(artificial agent)로서의 지위를 인정할 가능성이 제시되고 있다(신상규, 2016). 이러한 맥락에서 주목할 수 있는 개념은 바로 윤리적 '행위 주체성'(agency)이다. 행위 주체성[189]은 "더 이상 인간의 직접적 조작에 의해 작동하거나 지속적인 개입을 필요로 하는 수동적 존재가 아니며, 일종의 직권 위임에 의해 스스로의 자율적 판단을 통하여 작동하는 능동적 행위자"(신상규, 2016)를 의미한다. 즉, 로봇은 "문제의 그 고유한 결정과 실행이 이루어지는 동안 어떤 인간 주체에 의해서도 직접적인 통제를 받지 않는다(고인석, 2012a)." 인공지능 담론에서 행위 주체성 개념은 역량(capability)을 전제로 하는 개념이라고 할 수 있다. 히마(Himma)는 행위 주체성이 행위(act or action)로서 여겨질 수 있는 무엇인가를 할 수 있는 상태(being capable of doing something)에 대한 관념이라고 할 수 있으며, 개념적으로 해당 행위를 수행할 수 있는 역량을 갖추었을 때 행위자로서 인정될 수 있다고 말한다.[190] 이렇듯 행위자 내지 작인(agent)으로서 로봇의 지위는(고인석, 2012a) 소위 강한 인공지능 혹은 약한 인공지능의 구별기준에 대한 논쟁으로부터도 자유롭다고 할 수 있다.[191]

나아가 자율적인 도덕적 행위자에 대한 유연한 개념 틀을 바탕으로 하는 도덕적・윤리적 책무(accountability) 개념 역시 주목받고 있다. 이와 관련하여 최근의 지능정보사회에 대한 현실적인 적합성을 갖고 있는 이론적 틀로서 정보윤리가 조명 받고 있다. 정보윤리의 관점에서는 도덕적이라고 평가될 수 있는 행동을 수행할 수 있는 모든 자율적 존재를 도덕적 행위자로 간주하므로, 인간뿐만 아니라 동물, 나아가 인공적 행위자까지도 도덕적 행위자라는 개념 안에 포섭된다(신상규, 2016). 이러한 도덕적 행위자 개념은 도덕적 책임을 질 수 있는 존재를 이성적 판단능력을 갖춘 인간에게만 한정 짓고 있는 기존의 책임(responsibility) 개념

189) 행위자성(agency) 개념에 대해 보다 자세히는, D. J. Gunkel(2012) 18면 이하 참고.

190) K. E. Himma, "Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent Have to Be a Moral Agent?", Ethics and Information Technology, Vol. 11, Iss. 1, 2009, pp.19～20.

191) 자율주행자동차에 대해 권리능력이 아닌 (한정적) 행위능력은 인정될 수 있는 여지가 있다는 점에서 법적으로 로마시대의 노예와 유사하게 취급할 수 있다는 제안이 이루어지기도 한다(이중기, 2016c: 9).

을 넘어서고자 하는 맥락에서 제안된 것이라고 한다. 또한 근대적인 책임이론은 발생한 결과에 대한 인과성을 전제로 하지만, 지능정보사회에서 새로이 도래하는 다양한 양상들은 기존 인과성 중심의 책임 개념만으로는 책임 주체를 확정지을 수 없다는 한계를 내재적으로 갖고 있다(신상규, 2016). 이러한 인과성 개념 적용의 곤란성은 인공지능 알고리즘 작동에서 현저히 나타남을 앞서 살핀 바 있다. 플로리디(Floridi)는 도덕적 '책임'의 기준이 인간 중심적으로만 사유하는 도덕의 내재적 한계를 넘어설 수 없다고 주장하면서(신상규, 2016), 전통적인 의미의 책임을 귀속시킬 수 없는 도덕적 행위자들의 선택 및 행동이 급속히 증대되고 있는 현대의 지능정보사회에 적합한 새로운 개념이 제안될 필요성을 역설한 바 있다(신상규, 2016). 도덕적 행위에 대한 규범적 평가에는 기존의 책임 개념을 넘어서는 맥락이 포함되어 있으며, 도덕적 행위자가 책임은 질 수 없는 주체라고 할지라도 도덕적인 행동을 하도록 해야 한다는 당위적 요청이 이론적으로 가능하다는 것이다(신상규, 2016). 이러한 맥락에서 플로리디는 책임을 넘어서는 책무성 내지 책망(censure) 개념을 제안하고 있다(신상규, 2016). 즉, 도덕적 책임 개념과 도덕적 책무 개념은 구별될 수 있다는 전제에서, 한편으로 행위자가 자신의 의도 및 의지에 따라 자발적으로 행위하였다면 도덕적 책임을 지는 것이고(이상형, 2016), 다른 한편으로 행위자가 도덕적인 행위를 하였다면 해당 행위는 도덕적이라고 평가되어야 하며, 이때 도덕적 책무가 발생한 것으로 이해할 수 있다는 것이다(이상형, 2016). 결국 도덕적 책무 개념은 정신 및 감정이 없는 인공 행위자들에게도 부여될 수 있게 된다(이상형, 2016). 이렇듯 인공지능 기술이 상용화되면서 인간의 규범 및 가치판단, 그리고 행위기대가 기계적 알고리즘으로 대체되는 경향이 증대되고 있으며, 이에 따라 인간 중심의 규범체계를 넘어서는 새로운 규범체계가 '알고리즘 책무성'(algorithm accountability)의 관점에서 정립될 필요성이 제기되고 있다.[192]

행위 주체성이나 책무성 등 새로운 개념들의 창설이 적극적으로 검토되는 까닭은 지능로봇과 같은 비인간적 주체들의 행위 유형 및 범주가 확대될 것으로 예측되는 오늘날의 시점에서 새로운 개념적 프레임을 도입하는 것이 갖는 합리성에 근거하는 것으로 해석된다.[193] 이러

192) 알고리즘 책무성은 안전, 보안, 신뢰의 관점에서 인공지능 알고리즘에 대한 적절한 통제를 가능케 하는 제도화 방안을 포괄한다(이원태, 2015: 22).

193) 기존의 개념적 프레임이 자율주행자동차의 책임 등에 대한 문제에 근본적으로 접근하는데 한계가 있음을 지적하는 신상규(2016), 278면 참고.

한 이해에 바탕을 둘 때, 지능형 자율주행자동차에 대해서 역시 행위 주체성과 도덕적 책무를 인정할 수 있게 되며, 이는 자율주행자동차에 대한 윤리적 규범화 요청으로 연결된다.

다. 자율주행자동차의 윤리적 프로그래밍

1) 사전적 단계로서 윤리적 프로그래밍

이러한 윤리적 규범화의 요청은 기술적으로는 자율주행자동차에 대한 윤리적 프로그래밍으로 현실화된다. 자율주행자동차의 윤리적 프로그래밍은 사후적 차원에서 논의되는 책임귀속의 법리상 난제로 인해, 사전적 규율의 차원에서 자율주행자동차를 도덕적・윤리적으로 책임적인 주체로 만들어야 한다는 규범적 당위성에 대한 요청이 강조됨에 따라 활발하게 논의되고 있다. 이에 대해 사후적인 처벌과 보상의 관점에서 (윤리적 행위자의 보편적인 책무성에 입각한) 예방적 점검의 윤리학으로 나아가는 전환이 이루어지고 있다고 평가된다(김형준, 2017). 로봇의 자율적인 판단 및 행동은 알고리즘과 그에 기초한 일련의 학습과정을 통해 이루어진다는 점에서 인간의 그것과 구별될 수 있다(고인석, 2011). 따라서 자율주행자동차의 주의의무 및 윤리판단이 이루어져야 할 주행상황의 설정은 제조 당시 사전적으로(ex ante) 미리 결정되어 프로그래밍을 통해 반영되어야 한다(이중기, 2016a). 특히 기존의 법적・윤리적 규범체계를 프로그래밍하는 것은 로봇 사용에 따른 피해를 최소화하는 방안으로써 중요하게 고려되고 있다(이원태, 2015). 나아가 이러한 사전적 프로그래밍은 제조자의 관점에서 자율주행자동차 제조를 위한 일종의 인증 기준으로 작동하게 될 것으로 예측되고 있다(이중기, 2016a).

2) 윤리적 인공지능의 구현을 위한 접근방법

구달(Goodall)은 자율주행자동차의 충돌상황이 갖는 특성과 이에 따른 고려사항을 아래와 같이 정리한 바 있다: i) 자율주행자동차는 거의 반드시 충돌상황을 수반하게 될 것이다; ii) 해당 충돌상황에 선행하는 자율주행자동차의 의사결정에는 도덕적・윤리적 판단요소가 담겨져 있다; iii) 인간의 복잡한 도덕체계를 인공지능 소프트웨어에 효과적으로 구현(encode)하기 위한 명백한 방법은 존재하지 않는다(N. J. Goodall, 2014). 자율주행자동차의 의사결정이 외부 세계에 미치는 영향은 일련의 윤리적 고려를 필수적으로 요청하며, 따라서 자율주행자동차의 윤리적 판단이 기술적으로 가능하기 위한 프로그래밍 방법론들이 제안되고 있다. 윤리적 프로그래밍을 통해 도덕적・윤리적 로봇을 만들어 가는 과정은 인간의 규범적 가치기준을 인공지능에 전이(transfer)하여 로봇에 구현하는 것을 의미한다(김윤명, 2016). 윤

리적으로 자율적인 기계를 만들기 위해 컴퓨터에 인간의 윤리체계를 알고리즘으로 짜 넣는 작업은 인공지능 연구 중에서도 가장 난해한 도전으로 악명이 높으나(P. Lin, 2016), 다양한 방식으로 지속적으로 시도되고 있다.

라. 자율주행자동차와 트롤리 문제에서 윤리적 딜레마

자율주행자동차의 윤리적 프로그래밍에 대한 요청이 가장 적극적으로 논의되고 있는 쟁점은 자율주행자동차의 운행 중 발생할 수 있는 윤리적 판단의 딜레마 상황, 바로 트롤리 문제(Trolley Problem)일 것이다. 트롤리 문제란 1967년 영국의 윤리학자이자 철학자인 필리파 풋(Philippa Foot)이 제시한 윤리적 딜레마에 대한 사고 실험(thought-experiments)을 지칭하는 개념이다. 먼저 트롤리 문제는 브레이크가 고장난 전차가 다가오는 철로 위에 5명의 사람이 있고 다른 철로 위에는 1명의 사람이 있을 경우, 철로를 변경할 수 있는 조종대 앞에 서 있는 자의 딜레마 상황을 묘사하고 있다. 조종대 앞의 사람은 5명의 사람이 죽도록 내버려 둘 것인가 아니면 철로를 변경하여 5명을 살리고 1명을 희생시킬 것인가. 이러한 트롤리 문제는 윤리판단에서의 맥락에 대한 민감성을 시사한다는 점에서도 의의를 가진다. 예컨대, 5명이 있는 철로에 한 사람을 밀어 떨어뜨리는 것은 철로를 변경하는 것과 동일한 결과, 곧 5명의 사람을 살리는 대신 1명의 사람을 희생시키는 결과를 가져오지만 양자에 대한 도덕적·윤리적 반응은 상당히 다를 것이다(N. J. Goodall, 2014).

이러한 트롤리 문제는 자율주행자동차의 맥락에서 오늘날 다시 주목받고 있다. 극단적인 차악(the lesser of two evils)을 선택하도록 하는 트롤리 문제는 상당히 희소한 상황적 설정을 전제로 하지만 이는 특정한 변수를 기준으로 하는 사고실험이기 때문이지, 트롤리 문제에서와 같이 극단적이고 희소한 상황에서만 윤리적 판단이 요청된다는 것을 의미하지는 않는다. 즉, 윤리적 판단은 위험적 요소가 존재할 때마다 요청되는 것이다(N. J. Goodall, 2014). 자율주행자동차의 주행 시 요청되는 윤리적 판단은 비단 충돌 상황에만 국한된 것은 아니다. 예컨대 바로 옆을 달리고 있는 자전거 운전자를 배려하여 안전공간을 확보하기 위해 속도를 조절하거나 차선을 조금 비켜 달리는 것과 같이, 위험적 요소에 대해 일상적인 차원에서 이루어지고 있는 일련의 '고려들'이 모두 윤리적 판단을 수반하고 있는 것이다(N. J. Goodall, 2014).

1) 자율주행자동차의 윤리적 딜레마와 판단기준 및 우선순위 문제

트롤리 문제가 자율주행자동차의 맥락으로 현실화된, 피할 수 없는 교통사고에서의 도덕적 딜레마 상황은 생명, 신체, 재산이 상호 충돌하는 법익으로서 주로 제시된다(이중기・오병두, 2016). 이러한 윤리적 딜레마 상황 설정은 보행자들 사이에서 선택이 강제되는 상황과 자율주행자동차의 탑승자와 보행자 사이에서 선택이 강제되는 상황으로 대표된다. 전자의 상황에서는—예컨대 한 명의 보행자와 다수의 보행자의 설정과 같이—사망자 수(death toll) 등 인명피해가 판단의 기준으로 고려될 수 있다. 이와 더불어 어린이와 노인의 설정과 같이 보행자의 특성을 윤리적 판단의 변수로 삼아야 할지 역시 논의의 대상이 되고 있다. 이들 모두는 기본적으로 윤리적 판단에서 공리주의적 차원의 고려가 개입되어야 하는지에 대한 사고 실험이라고 할 수 있다. 후자와 같이 자율주행자동차의 탑승자와 외부 보행자 사이의 선택을 강제하는 상황 설정에서의 윤리적 판단은 탑승자와 보행자 사이의 우선순위와 관계된다.

이러한 윤리적 판단의 스펙트럼에서 고려될 수 있는 다양한 변수들과 이들 사이의 우선순위를 다루는 사고실험 도구로 '모럴 머신'(Moral Machine)을 예로 들 수 있다. MIT 미디어랩(Media Lab)은 자율주행자동차의 윤리적 딜레마 상황에서 이루어지는 가치판단의 기준에 대한 집단지성을 분석할 수 있는 일종의 도구로서 모럴 머신이라는 사고 실험 플랫폼을 구축하였다.[194)]

특히 미디어랩의 사고 실험 플랫폼은 자율주행자동차의 인지 센서가 해당 특성을 정확하게 식별할 수 있는 기술수준을 전제로 하여, 탑승자와 보행자의 다양한 특성을 시나리오 상황설정에 활용하고 있다.[195)] 여기에서 제시되고 있는 자율주행자동차의 탑승자와 보행자의 유형 및 개별적 특성은 다음과 같다: 남성/여성, 임산부, 노인, 어린이(children), 아기(baby), 운동선수(athletes), 비만인(large person), 의사, 경영진(executive), 노숙자(homeless person), 범죄인, 강아지/고양이.

이러한 유형 및 특성을 바탕으로 미디어랩에서 설정하고 있는 윤리적 딜레마 상황에서의 가치판단 기준들을 정리하면 다음과 같다.

---

194) 사용자들은 기존의 자율주행자동차 딜레마 상황 시나리오에서 외부 관찰자로서 일련의 선택을 하면서 축적된 결과 데이터를 바탕으로 자신의 판단기준 가중치를 비교・대조할 수 있으며, 직접 시나리오를 작성하여 공유할 수도 있다(http://moralmachine.mit.edu/).

195) 자율주행자동차 관련 윤리적 쟁점을 내포하는 대표적인 사례들에 대한 소개로, P. Lin(2016) p.73 이하 참고.

〈표 4－1〉 윤리적 딜레마 상황에서의 가치판단 기준

	가치판단 기준
1	사회적 가치 선호도 (Social Value Preference)
2	희생자 수의 중요도 (Saving More Lives)
3	종에 대한 선호도 (Species Preference)
4	연령 선호도 (Age Preference)
5	성별 선호도 (Gender Preference)
6	(보행자에 대한) 탑승객 보호 선호도 (Protecting Passengers (vs. pedestrian))
7	신체 건강 선호도 (Fitness Preference)
8	개입에 대한 회피 선호도 (Avoiding Intervention)
9	법규 준수 선호도 (Upholding the Law)

위 설정에서 ‘사회적 가치 선호도’란 의사 및 경영진 對 노숙자 및 범죄인의 비교・대조 구도를 바탕으로, 사회에서 보다 인정받는 지위를 충돌 우선순위에서 고려하는 것을 의미한다. ‘희생자 수의 중요도’는 가장 대표적인 공리주의적 고려로서 자동차 사고로 인한 사망자 수 등 인명피해와 관련된다. ‘종에 대한 선호도’는 강아지, 고양이와 같은 애완동물과 인간 간의 우선순위 문제를 다룬다. ‘연령 선호도’는 노인, 평균연령의 일반인, 어린이, 아기 사이, 그리고 ‘성별 선호도’는 남성과 여성 사이의 보호 선호도이다. ‘탑승객 보호 선호도’는 탑승객과 보행자 사이에 어느 대상을 우선적으로 보호할지의 문제이다. ‘신체 건강 선호도’는 운동선수 對 비만인의 구도를 바탕으로 한다. 또한 ‘개입에 대한 회피 선호도’는 자율주행자동차가 충돌 상황을 예측한 직후, 별도의 추가적인 의사판단 없이 주행하던 차선을 그대로 유지할지, 독자적인 의사판단을 바탕으로 차선에서 이탈하는(swerve) 등 적극적으로 개입할지 여부에 대한 것이다. 나아가 ‘법규 준수 선호도’는 특히 보행자가 횡단보도를 건널 때 교통신호를 준수하는지 여부에 따라, 예컨대 탑승자와 보행자 사이의 보호 우선순위를 정할 때 보호 의무의 정도에 있어서 차등을 둘지의 문제라고 할 수 있다.

윤리적 딜레마 상황에서의 판단기준 및 우선순위는 기본적으로 충돌 최적화(crash-optimization) 알고리즘과 밀접한 상관성을 갖고 있다. 그런데 이러한 충돌 최적화는 궁극적으로는 표적화(targeting)를 의미하게 된다(P. Lin, 2016). 즉, 어떤 선택을 따르던 누군가의 희생이 불가피한 딜레마적 상황에서, 보호할 대상을 '선택'한다는 것은 자동적으로 희생될 대상을 '표적'으로 삼는 행위를 수반하게 된다. 사망자 수의 많고 적음과 같이 보다 객관적인 지표의 경우에도 표적화는 이루어지지만 공리주의적 맥락에서 사회적 수용도가 좀 더 높을 가능성이 있다. 하지만 이러한 표적화가 특히 사회적 지위, 연령, 성별 등의 기준으로 이루어진다면 주관적이고 상대적인 성격의 가치평가를 바탕으로 할 수밖에 없다는 점에서 논쟁의 여지가 많다고 하겠다. 다시 말해, 충돌 최적화 알고리즘이 의도적이고 체계적인 차별을 수반할 수 있는 위험성을 내재하고 있음이 지적되고 있다(P. Lin, 2016). 예컨대 IEEE의 윤리헌장에서는 인종, 종교, 젠더, 장애 여부, 나이, 국적, 성적 지향성 등에 따라 차별하지 않고 모든 인간을 공평하게 대우해야 한다는 내용이 명시되어 있으며, 이때 표적화는 분명 윤리적으로 부적합한 것이 된다(P. Lin, 2016). 물론 표적화는 평등원칙 위반 등 헌법상의 위헌성으로부터도 자유로울 수 없으며, 이에 따라 민주적 헌정질서에서 명백하게 불법적인 선택 혹은 표적을 설정하도록 설계하는 제조사를 허용할 수는 없을 것이다(P. Lin, 2016).

또한 표적화는 실천적 지평에서도 모순점을 갖고 있다. 대표적으로 표적화의 대상이 위험에 대한 과도한 부담을 부당하게 짊어지게 될 수 있다는 점이 지적되고 있다. 충돌 최적화의 관점에서 판단해보면 상대적으로 피해가 적을 것으로 예측되는 대상을 표적의 대상으로 삼게 될 가능성이 높다. 즉, 일반 승용차에 비해 안전도가 높은 SUV(sports utility vehicle)를, 헬멧 미착용자에 비해 착용자를 표적의 대상으로 삼게 되는 것이다(P. Lin, 2016). 그러나 헬멧을 착용한 오토바이 운전자를 표적화하다면 안전을 고려해 헬멧을 착용하는 신중한 판단을 했다는 이유로 차별적으로 페널티(penalty)를 받게 되는 사뭇 비합리적인 결과를 초래하고 말게 된다(N. J. Goodall, 2014). 또한 오히려 각자의 안전을 확보하기 위해 헬멧을 착용하지 않도록 하는 전략적 행동으로 유도할 수도 있다는 점에서 법정책적 부작용이 예측되기도 한다(P. Lin, 2016).

다음으로 탑승자와 보행자 사이의 우선순위 문제는 i) 어떠한 전제에서도 탑승자의 보호를 우선시하는 '자기보존'(self-preservation)과 ii) 탑승자보다 보행자의 보호를 우선시하는 '자기희생'(self-sacrifice)이라는 양 극단의 입장 사이에 존재하는 긴장관계와 관계한다. 자기

보존 내지 자기보호(self-protection)의 입장은—자기희생의 관념이 자율주행자동차에 대한 구매 욕구를 저하시킬 것이고, 반대로 탑승자를 최우선적으로 고려하는 윤리적 프로그래밍이 구매자의 지지를 받게 될 것이라는(N. J. Goodall, 2014)—자율주행자동차의 제조자 및 구매자의 관점을 반영한다는 점에서, 현실적인 윤리적 대안으로 논의되고 있다(N. J. Goodall, 2014). 이렇게 볼 때, 자율주행자동차의 사고 시 선호도, 가중치, 우선순위 체계가 운전자, 보행자, 산업, 시장 등 일반 사회에 미치게 될 영향 역시 윤리의 실질적 내용을 형성하는데 있어서 중요하게 고려되고 있다고 하겠다(이중기・오병두, 2016). 다만 자율주행자동차와 탑승자를 중심으로만 충돌 최적화를 판단하는 것을 윤리적이라고 평가하기에는 한계가 있다. 반대로 공리주의적 관점에서 요청될 경우, 자기희생을 감내하도록 강요하는 것 역시 반드시 윤리적이라고만 평가할 수는 없다. 이는 자기보존과 반대로 자율주행자동차의 제조자와 소비자의 지지를 받을 수 없다는 현실적인 한계 역시 가지고 있다.

나아가 자율주행자동차의 개입 여부 자체에 대한 논쟁이 있을 수 있다. '개입에 대한 회피 선호도'는 가치판단의 최소화 기준이라고 풀이할 수 있다. 자율주행자동차가 충돌 상황에 직면하였을 때, 기존의 주행 상태를 그대로 유지하는 것과 일련의 선택을 통해 차선을 변경하는 등 보다 적극적인 개입이 이루어지는 것의 법윤리적 평가가 상이할 수 있다. 이는 결과주의자(consequentialist)와 비결과주의자(non-consequentialist) 사이의 입장 차이로 설명되기도 한다. 결과주의자의 경우 공리주의적 관점에서 산출된 평가에 따라 더 나은 결과를 가져오는 선택을 정당화할 수 있지만, 비결과주의자는 수학적 결과만을 고려하지 않으며, 특히 인위적인 개입이 추가적으로 이루어지는 것 자체에 대한 문제의식을 강하게 제기한다. 즉, 트롤리 문제에 자율주행자동차 상황을 대입해본다면, 자율주행자동차가 차선을 그대로 유지하는 것은 5명의 사람이 죽게 되는 결과를 초래하기는 하지만 단지 죽도록 내버려두는 행위(let die)에 해당하게 되며, 이와 비교해 차로를 변경하는 행위는 1명의 사람이 희생되지만 해당 사람을 죽이려는 의도가 적극적으로 개입하게 된다는 점에서 양자에 대한 가치평가는 상이할 수 있다는 것이다(P. Lin, 2016). 이는 작위와 부작위에 대한 평가, 즉 죽이는 행위와 죽도록 내버려 두는 행위에 대한 도덕적 구별에 바탕을 두고 있다(P. Lin, 2016). 이러한 맥락에서 자율주행자동차의 윤리적 판단에 대한 일률적 프로그래밍이 가진 한계를 지적하면서, 이러한 기준이 정부 차원에서의 개입에 의해 이루어질 경우 후견주의적인 속성을 내재할 수밖에 없다는 견해(오도빈・유지원・양규원・구보민, 2017) 역시 존재한다.

마지막으로 교통법규 등 법적 규율의 실질적 내용이 자율주행자동차의 로봇윤리를 구성하는 체계 내적인 요소가 될 것으로 예측된다(이중기・오병두, 2016). 교통법규의 준수여부를 고려하게 될 경우, 예컨대 보행자가 신호체계를 준수하여 횡단보도를 건너는지 여부가 보행자에 대한 보호 우선순위에 영향을 미치게 될 수 있다. 이러한 법규 준수 여부는 표적화에 대한 문제제기와 같이 자율주행자동차의 주관적인 윤리판단이 갖는 위험성이 지적되면서, 최소한의 객관적인 규범적 테두리로서 기능할 수 있음이 강조되고 있다.

2) 자율주행자동차의 윤리적 프로그래밍을 위한 이론적 시도

자율주행자동차의 윤리적 판단기준은 자율자동차의 윤리적 프로그램을 설계하는 단계에서 실질적인 내용을 형성하게 된다. 인간 운전자의 경우 충돌 상황에서 실시간으로 의사판단이 이루어지지만, 자율주행자동차의 경우 충돌 상황에서 어떻게 판단할지에 대해 프로그래머에 의해 사전에 규정되어야 한다는 것이 양자의 차이라고 할 수 있겠다(N. J. Goodall, 2014). 예측되는 충돌 상황에 대해 발생 및 결과의 가능성, 안전성 등을 종합적으로 고려하여 가장 최선의 충돌 방식을 선택하는 역량에 있어서, 자율주행자동차가 인간보다—특히 신속성 및 확실성을 고려할 때—우월할 것으로 평가되고 있으며(N. J. Goodall, 2014), 트롤리 문제와 같은 윤리적 딜레마 상황에서도 로봇 시스템에게 맡기는 것이 오히려 바람직할 수 있다는 주장이 펼쳐지기도 한다(고인석, 2012a). 자율주행자동차는 도로주행이라는 특성상 다양한 양상으로 나타날 수 있는 충돌 궤적(trajectory)의 대안들을 예측할 수 있으며, 이 중에서 가장 낮은 피해를 가져오는 대안을 선택할 수 있는 역량, 즉 불가피한 충돌 상황에서 '최적의 충돌 전략'(optimal crashing strategies)을 제시할 수 있는 역량을 갖추고 있다는 것이다(N. J. Goodall, 2014). 다만, 자율주행자동차의 윤리적 프로그래밍은 시뮬레이션 상황 시나리오에서 이미 알려진 결과에 대한 단일한 선택지들을 바탕으로 이루어지지만, 실제 도로주행의 현실상황에서는 결과발생 여부가 불확실하며 해당 시점에는 알려지지 않았던 다른 선택지들과 연동될 수 있는 상황적 맥락의 복잡성 때문에 한계가 있다는 점이 동시에 지적되기도 한다(N. J. Goodall, 2014).

현재 자율주행자동차의 윤리적 프로그래밍에서 하향식 접근으로 가장 많이 활용되고 있는 윤리이론은 공리주의적 관점에서 해악 최소화 알고리즘과 의무론적 관점에서 아시모프의 로봇 3법칙이라고 한다(이중기・오병두, 2016). 이들은 각기 장단점을 가지므로 자율주행자동차의 윤리체계를 구성하는데 있어서 상호보완적인 관계에 놓인 것으로 평가받는다(N.

J. Goodall, 2014).

먼저 자율주행자동차의 윤리적 딜레마 상황에서는 공리주의적 접근이 가장 일반적으로 검토되고 있다. 이러한 공리주의적 접근방식은 i) 자기보존 알고리즘(self-preservation algorithm), ii) 자기희생 알고리즘(self-sacrifice algorithm), iii) 해악 최소화 알고리즘(harm-minimizing algorithm)으로 정리될 수 있다고 한다(이중기・오병두, 2016). 위에서 살펴보았듯이, 자기보존의 경우 자율주행자동차와 탑승자의 안전을 극단적으로 보호함으로써 보행자 등의 피해를 과도하게 초래할 수 있어 시민사회에서 실제 수용되기 어려운 난점을 갖고 있고, 자기희생이라는 극단의 이타주의 역시 자율주행자동차의 상용화를 고려할 때 현실화되기 어렵다는 한계를 가진다(이중기・오병두, 2016). 해악 최소화 알고리즘은 자기보존과 자기희생이라는 양 극단의 관점을 지양하고, 해당 상황에 대한 해악(harm)을 구체적으로 산정함으로써 윤리적 딜레마를 해결하고자 한다(이중기・오병두, 2016). 다만 효용 계산의 주관성 문제와 마찬가지의 맥락에서, 해악 역시 그 정의에 따라 산출방식이 각기 상이할 수밖에 없다는 점은 내재적 한계라고 하겠다(이중기・오병두, 2016). 즉, 해악 최소화 알고리즘은 동일한 법익의 충돌 상황에서 효용 및 해악에 대한 계산을 통해 보다 객관적인 판단기준을 제시한다는 장점을 가지지만, 결국 그 정의 및 기준이 주관적이거나 불명확할 수 있다는 한계를 내재하고 있는 것이다(이중기・오병두, 2016). 예컨대 충돌의 있고 없음, 중대한 피해와 경미한 피해와 같이 선택이 명확한 경우가 아닌 한, 결정의 정당성에 대한 결과주의적 판단이 달라질 수밖에 없다. 나아가 해악의 정의 및 기준이 분명하게 마련된다고 하더라도 계산에 따른 양적 기준에 따라 손해를 산출하는 것이 바람직하지 않은 결과를 초래할 수도 있다(N. J. Goodall, 2014). 인간과 달리 컴퓨터는 공리주의적 계산을 즉각적으로 냉정하게(dispassionately) 수행할 수 있으나(J.-F. Bonnefon & A. Shariff & I. Rahwan, 2015), 이로 인해 인간의 일반적 상식(common sense)과 불일치할 수 있는 지점이 발생할 수 있으며, 윤리적 판단이 수학적 계산으로 환원되는 것이 갖는 궁극적인 위험성 역시 지적될 수 있겠다. 예컨대 특정 개인 혹은 집단의 안전이 향상되는 측면은 분명 누군가의 희생이라는 대가를 치렀을 때 가능한 것이라는 점에서, 전반적인 사망자(fatalities)의 감소가 반드시 윤리적이라고만 볼 수는 없는 것이다(P. Lin, 2013). 이러한 한계에도 불구하고 공리주의적 관점은 자율주행자동차의 윤리를 설계하는데 있어서 가장 기본적인 이론틀을 제공해주고 있다. 이는 사회적 수용도에 대한 고려 역시 반영한다. 경험적으로 분석된 실험 윤리(experimental ethics)는 시민들의 도덕적 기대를

반영하고 있다는 점에서 자율주행자동차의 사회적 수용을 위해 중요하게 고려되고 있는데, 한 실증연구에 따르면 일반인들이 사망자 수를 최소화하도록 설계된 공리주의적 자율주행자동차에 대해 상대적으로 높은 수용도를 보였다고 한다.[196)]

의무론의 관점에서는 아시모프의 로봇공학 3법칙을 적극적으로 변용하려는 시도가 지속적으로 이루어지고 있다. 위 법칙은 윤리적 난제 상황에 봉착했을 때 상충되는 이익들 중 하나를 선택해야 되는 로봇의 상황에서 해결 방안을 고안하고자 했다(이중기・오병두, 2016). 특히 위계서열적 구조의 단순명료함으로 인해 알고리즘화 작업이 비교적 용이하다는 점에서, 로봇공학자들로부터 그 실효성이 높게 평가받고 있다(이중기・오병두, 2016). 하지만 로봇공학 3법칙의 추상성은 지속적으로 비판받아 왔으며(N. J. Goodall, 2014), 동일한 위계서열 내에서의 우선순위에 대한 고려가 이루어질 수 없다는 난점이 있다(이중기・오병두, 2016). 예컨대 기존 로봇공학 3법칙의 제1법칙은 트롤리 문제와 같은 딜레마적 상황에서의 도덕적 판단에 대해 그 어떤 지침을 제공해주지 못한다. 즉, 위계서열구조를 바탕으로 상충적인 이익들 간의 조정이 이루어질 수 있으나, 동일서열에서의 갈등상황은 해결할 수 없다는 한계를 가지며(이중기・오병두, 2016), 이는 자율주행자동차의 윤리적 딜레마 문제를 해결하기 위한 윤리적 프로그래밍의 이론적 기초로 삼기에는 실천적 한계가 있음을 의미한다. 이에 따라 윤리학적 관점에서 갖는 한계를 수정 및 보완하여, 자율주행자동차의 상황에 적절한 형태로 변용하려는 시도가 이루어지고 있다. 이중기, 오병두는 이러한 연구의 대표적인 예로 저디스(Gerdes)와 손튼(Thornton)의 3법칙(J. C. Gerdes & S. M. Thornton, 2016과 로하스(Rojas)의 4법칙[197)]을 제시한 바 있다.

---

196) 저자들은 세 가지 설문조사에 대한 분석을 바탕으로 이러한 결과를 발표하였는데, 다만 자율주행자동차 자신과 탑승객보다 타인을 우선시하는 자기희생적 선택에 대해서는 좀 더 신중한 태도를 보였다고 한다(J.-F. Bonnefon & A. Shariff & I. Rahwan, “Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?”).

197) R. Rojas, “I, Car: The Four Laws of Robotic Cars”(http://www.inf.fu-berlin.de/inst/ag-ki/rojas_home/documents/tutorials/I-Car-Laws.pdf).

〈표 4－2〉 저티스와 손튼의 3법칙과 로하스의 4법칙

저디스와 손튼의 3법칙	로하스의 4법칙
제1법칙 자동화된 자동차는 보행자 또는 자전거운전자와 충돌해서는 안 된다. 제2법칙 자동화된 자동차는 충돌을 회피하는 것이 제1법칙과 상충되는 경우를 제외하고는, 다른 자동차와 충돌해서는 안 된다. 제3법칙 자동화된 자동차는 충돌을 회피하는 것이 제1법칙 또는 제2법칙과 상충되는 경우를 제외하고는, 주변의 그 어떤 대상과도 충돌해서는 안 된다.	제1법칙 자동차는 인간을 해치거나 혹은 부작위로 인해 인간에게 위험을 초래해서는 안 된다. 제2법칙 자동차는 제1법칙과 상충되는 경우를 제외하고는, 교통법규를 준수해야 한다. 제3법칙 자동차는 인간이 내린 명령이 제1법칙 또는 제2법칙과 상충되는 경우를 제외하고는, 인간의 명령에 복종해야 한다. 제4법칙 자동차는 제1법칙, 제2법칙 또는 제3법칙과 상충되지 않는 한, 자신의 존재를 보호해야 한다.

저디스와 손튼의 3법칙은 충돌대상을 유형화하고 보행자/자전거운전자, 다른 자동차, 기타의 순으로 위계서열에 따른 보호의 우선순위를 설정하여 이를 윤리적 지침의 기본적인 표본으로 삼았다(이중기・오병두, 2016). 로하스의 4법칙은 객관적인 판단기준으로 교통법규를 제2법칙으로 설정하고 있는 것이 흥미롭다.

해악 최소와 알고리즘과 아시모프의 로봇공학 3법칙의 현대적 변용은 모두 하향식 접근방식에 해당한다고 볼 수 있다. 이와 더불어 상향식 접근방식 역시 자율주행자동차의 윤리적 프로그래밍에 있어서 고려해볼 수 있다고 하겠다. 다만 알고리즘 판단의 궁극적인 추적 불가능성 및 조작 위험성을 고려할 때, 알고리즘의 투명성(transparency)을 높이고 명백히 비윤리적인 행위를 차단하는 사전 작업이 뒷받침되지 않는다면 상향식 접근법에만 의존하여 자율주행자동차의 윤리적 알고리즘을 완성하는 것이 사회적으로 수용되기는 어려울 것이다(N. J. Goodall, 2014). 특히 상향식 접근법에 기초한 윤리학습을 통해 자율주행자동차가 운전자의 편견 및 잘못된 운전습관을 학습할 가능성까지 열려 있게 된다는 부작용에 대한 지적 역시 존재한다(이중기・오병두, 2016).

이에 따라 자율주행자동차의 윤리적 프로그래밍에서도 하향식과 상향식 접근법의 적절한 조화를 꾀하는 혼합식 접근법이 제안된 바 있다. 이러한 혼합형 접근법의 한 예로 구달은 i) 일반적・보편적으로 승인된 원칙들에 근거하여 충돌의 여파를 최소화하는 행위를 하도록 하향식 접근에 기반한 도덕체계를 구축하고(1단계로서 하향식 접근법에 따른 이성적 윤리); ii) 기계학습에 기반하여 현실세계의 충돌 시뮬레이션 시나리오에서 인간의 의사결정과정을 학습하여 인간과 유사한 가치판단이 가능할 수 있도록 하며(2단계로서 상향식 접근법에 따

른 인공지능 접근법과 결합한 혼합식 접근법); iii) 마지막으로 자율자동차로 하여금 자신의 의사결정을 자연어(natural language)로 표현하도록 하여, 고도로 복잡한 논리체계에 근거하는 의사결정을 인간이 이해할 수 있는 형태로 전환하여 가독성을 높이고, 이를 통해 사후 교정이 가능하도록 하는 일련의 단계들을 주장한 바 있다(N. J. Goodall, 2014).

또 다른 혼합형 접근법의 에로서 이중기, 오병두는 수정・변용된 아시모프의 로봇공학 3법칙과 같은 의무론적 요소와 해악 최소화 알고리즘과 같은 공리주의적 요소가 결합된 하향식 접근법을 기본적인 접근방식으로 하되, 해당 지역의 운행상 관행 등 문화적 규범 등을 기계학습에 의해 반영할 수 있도록 상향식 접근법으로 보완하는 혼합형 접근법을 제안한 바 있다(이중기・오병두, 2016).

〈표 4－3〉 각 단계 행위자별 윤리이슈

단계	행위주체	윤리 이슈
개발단계	개발자	• 인간에 대한 안전보호 의무 및 사회윤리적 공리를 반영하는 알고리즘 개발 － 하향식의 윤리적 프로그래밍 강조
	제조사	• 알고리즘 개발 과정의 투명성 제고
	공적 주체 (중앙 및 지방 행정정부)	• 자율주행자동차의 개발 및 상용화를 위한 파지티브(positive) 규제 방식 채택 － 금지조항의 최소화 및 디폴트적 허용
	소비자 (운행자 포함)	• 윤리적 기준의 자율적 선택권 보장에 대한 요청
	시민단체 (보행자 포함)	• 사회적 수용성 판단을 위한 공론의 의견 형성 및 반영 요청
인허가단계	개발자 및 제조사	• 시범운행 단계에서의 윤리 기준 충족 및 운행결과에 대한 정보공개의 책무성
	공적 주체	• 인허가를 위한 윤리적 검증 기준의 정립 및 적용
	소비자 및 시민단체	• 인허가 과정에서의 윤리문제에 대한 모니터링 및 의견개진 절차 요청
상용단계	개발자	• 알고리즘에서 사회윤리 및 사회적 수용성의 지속적인 변화 추이 반영 － 상향식의 윤리적 프로그래밍 강조
	제조사	• 자율주행자동차의 제조물책임보험을 통한 사후적 차원에서의 윤리 문제 고려
		• 법인 내 기관윤리위원회 조직 및 운영을 통한 윤리문제의 절차주의적 해결

단계	행위주체	윤리 이슈
상용단계	공적 주체	• 국가·사회적 차원에서 공적 위험분배를 위한 사회보험 등 공적 기금 확보
		• 인허가단계의 윤리적 기준 충족 여부에 대한 사후적 검증 도구 마련
		• 스마트 시티 및 지능형 도로교통체계와의 연계에 따른 보안 문제 해소 방안
	소비자	• 윤리적 프로그래밍에 대한 선택권 및 자기결정권 보장의 지속적 유지 – 프로그래밍 변경가능성 포함
	시민단체	• 사회윤리 및 사회적 수용성 관련 지속적인 의견 형성을 위한 공론에서의 촉매 기능
		도로교통체계의 지능화로 인한 프라이버시 및 개인정보 침해 예방 및 모니터링

제 3 절 금융 분야: 로보어드바이저를 중심으로

1. 논의의 배경

금융 분야에서의 인공지능의 발전 및 이용은 현재 빠른 속도로 진화하고 있으며, 단순한 온라인 신용분석부문에서부터 투자자문을 제공하는 로보어드바이저 그리고 더 나아가 자동으로 개인자산을 운용하고 재투자하는 투자일임·펀드매니저 영역까지 점차 확대되고 있다. 최근에는 인공지능을 활용하여 거래소와 감독기관이 규제와 감독에도 이용하는 경향을 보이고 있다. 이러한 기술발전은 다양한 산업에 영향을 미치는데 이중 특히 비구조화된 데이터가 풍부한 금융 분야는 그 변화가 빠를 것으로 전망된다.[198)]

특히 로보어드바이저는 금융 분야만이 아니라 이후 의료자 문, 소비자 시장에서도 상품선택에 활용될 수 있다. 금융 분야는 종래 규제가 많고 한 나라의 경제에 있어 주요 기능을 하는 금융시장이 관련되어 있다는 점에서 이 분야에서의 로보어드바이저 관련 윤리 이슈에의 선제적인 대응은 다른 산업에도 유익한 시사점을 줄 것으로 기대된다.

이러한 이유에서 이하에서는 금융자문서비스를 제공하는 로보어드바이저 서비스에 포커스를 두어 금융 분야에서의 인공지능과 관련한 윤리이슈를 중심으로 문제점을 추출하고 대응방안을 정책적·사회적으로 모색해 보기로 한다.

198) 금융 분야는 이미 데이터가 집적되어 있을 뿐 아니라 ICT기술의 발달로 종전에 수치로만 사람의 반응을 예측할 수 있다면 이제는 SNS 등을 통해 사람의 말, 음성, 표정 등을 통해 행동을 보다 정확하게 예측할 수 있게 되었기 때문이다.

2. 본 연구에 적용하는 ‘금융윤리’의 개념과 특징

가. 직무윤리와의 차이

통상 윤리라고 할 경우 인간으로서 지켜야 할 당연한 선을 말한다. 반면에 지금까지 금융분야에서 빈번히 이야기되는 금융윤리는 직무윤리를 말하는 경우가 많다.

금융산업은 서비스 산업으로서 신용에 바탕을 두고 있으며 신용의 정도가 그 기업의 가장 중요한 자산이며 금융산업이 지탱하는 축이라 할 수 있다. 때문에 금융산업에 대한 믿음, 신뢰, 신용이라는 무형의 가치에 포커스를 두고 비록 법적으로 규제대상으로 하지 않더라도 윤리차원에서 자율규제기관의 규정 내지 회사의 자율적인 통제를 통해 규율하도록 권고해 왔다. 왜냐하면 이러한 직무윤리는 법규(law and regulation)와 공통된 부분이 많은 동시에 법규의 결함을 보충하는 기능을 갖고 있기 때문이다. 많은 법규가 마치 그물망처럼 촘촘하게 만들어져 있지만 거기에는 늘 결함이 있기 마련이다. 직무윤리는 이 같은 법규의 사각지대를 메우는 기능을 갖는다. 또한 법규는 원래 타율적인 것이지만 윤리는 이를 준수하는 자의 자율적인 선택에 의한 것이기 때문에 타율적 준수에서 오는 문제점과 한계를 극복할 수 있다는 장점이 있다. 금융관련법규의 궁극적인 목적은 금융산업의 건전성을 확보하고 금융소비자를 보호하는 것인데, 이러한 법의 취지는 법만으로는 부족하고, 윤리의 보충적 기능에 의하여 달성될 수 있다. 즉, 윤리 규범이 존재하지 않고 제대로 실행되지 않는 경우에 점차 엄격한 법에 의해서만 구속되어 금융기관들은 경제적 활동에 제약을 받게 되어 급변하는 환경에 신속하게 대처하는데 많은 제약을 받을 수밖에 없다. 또한 공정하고 자유로운 경쟁의 전제조건이 직무윤리라고 할 수 있으며 최근에는 윤리 자체가 경쟁력이며, 경제적으로도 이득이 된다고 보고 있다. 윤리적 행동이 초래하는 부정적 효과는 오히려 이를 만회하기 위한 비용과 시간이 소요되며, 이를 규제하기 위한 법적 규제와 같은 타율적 규제의 증가로 더 큰 사회적 비용을 초래할 수 있다. 아울러 직무윤리의 자발적 준수는 금융투자업 종사자 자신을 보호하는 장치가 된다. 특히 금융시장과 산업에서 금융상품을 판매・대리・중개・자문하는 경우 직무종사자로서 관련법상 금융서비스이용자(‘금융소비자’)의 이익과 충돌하는 경우 금융소비자의 이익을 우선시해야 하는 의무가 부과되어 있고, 이는 금융소비자의 재산을 위탁받아 관리 운영하고 있는 이상 고도의 주의의무가 부과되고 있다. 이러한 의무는 법적 의무이며, 위반 시 제재가 부과된다. 그러나 금융시장, 금융회사의 독특한 성격상 관련법규가 다수인바, 우려되는 모든 경우를 법적으로 규율하는 것은 현실적으로 불가능하다. 더욱

이 금융상품의 분석, 판매자인 금융기관과 금융소비자 간의 정보비대칭문제도 심화되어 있어 사후적인 감독에는 한계가 있다. 이 때문에 도덕성과 신뢰는 금융상품거래 및 시장발전에 필수적이라고 보고 있다. 이러한 점을 고려하여 금융시장과 금융산업에서는 지금까지 윤리가운데 특히 금융산업 종사자로서 사회규범화하여 자율적으로 강제할 필요성이 큰 경우에 직무윤리로 정립하는 경향이 있다.

나. 직무윤리 이상의 윤리에 초점을 둘 필요성

본 연구에서 초점을 두고 접근하는 윤리는 금융종사자의 직무윤리보다는 더 넓은 포괄적인 금융윤리이다. 왜냐하면 현재 금융윤리에 속하는 직무윤리는 근본적으로 금융회사가 제공하는 금융상품의 질을 금융소비자가 정확하게 판단이 어렵다고 보고 이를 위하여 공적 차원에서 금융법제도에 근거하여 마련된 진입규제, 상품규제, 행위규제 등을 보완하는 차원에서 윤리를 접근하고 있기 때문이다. 특히 금융상품을 제조하고 이를 금융소비자에게 판매하는 금융업자에 대해서는 직접 금융법제도에 마련된 진입규제, 상품규제, 행위규제가 적용된다. 직무윤리는 바로 이러한 법제도를 통해 요구되는 행동규칙을 보완하는 역할을 한다.

인공지능이 제공하는 로보어드바이저 서비스는 기존의 금융회사와 금융소비자간의 종래의 문제 즉 정보비대칭문제를 더 심화시킨다는 점에서 기존의 규율체계와 아주 다르지는 않을 것으로 보인다. 그러나 제도적인 측면은 물론, 윤리적인 면에서 보다 세심한 행동규율이 요구된다. 그 이유는 우선, 금융상품・서비스와 관련한 직무윤리는 통상적으로 금융산업만을 포괄하는 반면에서 인공지능을 탑재한 경우에는 비금융산업, 즉 인공지능을 설계・개발한 비금융회사가 넓게 금융산업권에 포섭되기 때문이다. 이들은 현행 법률에 의해 금융관련 법상 규율대상이 되지 않은 반면에 인공지능에 대한 원천적인 정보를 갖고 있는 자로서 전문성과 정직, 그리고 엄격한 윤리가 요구되기 때문이다.

무엇보다 이러한 새로운 기술의 활용은 종래 금융시장과 산업에서 요구되는 윤리를 뛰어넘는 새로운 문제를 유발할 가능성이 크다는 점도 고려하지 않을 수 없다. 전통적으로 인공지능이 존재하지 않았던 때 인간의 행위만으로도 공정성이 요구되고 금융거래상대방 보호가 강하게 요구되는 금융산업에서 인공지능이 이용되는 경우에는 새로운 불공정행위유형이 등장할 수 있고 그로인한 피해는 대규일 것이기 때문이다. 이로 인해 금융거래에 참여하는 일반인의 취약성은 더욱 심화될 가능성이 크다.

한편, 현재 로보어드바이저 분야의 기술발전의 수준은 매우 빠른 속도로 진화하고 있어 어느 정도의 기준으로 금융어드바이저의 금융윤리를 요구하고 확립할 것인가는 단언하기는 매우 어렵다. 지금은 인공지능이 사람을 대체할 정도로 일반적이지는 않고 그 시기에 대해서도 예측이 불가능하지만, 적어도 사람의 지시나 개입 없이 자율적으로 판단할 수 있는 고도의 인공지능이 예상보다 빨리 도래할 수도 있음을 간과할 수 없다. 사람의 개입없이 이루어지는 인공지능에 의한 지시와 인공지능 스스로 의사결정을 내리게 되면 그로 인하여 야기되는 다양한 위험들이 있음은 쉽게 예측된다는 점은 제도적 측면외에 윤리적 측면에서도 대응이 필요한 대목이다. 다만 현재 로보어드바이저와 관련한 금융윤리를 논함에 있어서 국제적 동향은 인간을 행위자로 보고 있는 현재 상황을 전제로 금융윤리를 논의하고 로보어드바이저 자체를 '전자人'으로 보고 논의를 하지는 않는다.

다. 로보어드바이저 분야에서 금융윤리의 범위와 대상

앞서 기술한 바와 같이 현재의 수준에 비추어 볼 때 로보어드바이저의 등장에도 불구하고 규제 내지 윤리를 논함에 있어 사람의 역할이 주가 될 수밖에 없다. 물론 인공지능의 자율성의 수준에 따라 '전자人'을 전제로 논의를 하는 것이 필요하지만, 현재 그러한 단계에서는 인간의 행동을 통해 통제가 가능하다는 전제하에서 금융윤리를 접근하는 것이 지금까지의 국제적 동향이다.

〈표 4-4〉 인공지능의 시기별 분류

시기	1950년대~	1980년대~	2010년대~	2045년~
종류	검색프로그램	if~then형식의 rule-base[199]	기계학습	범용적 인공지능
특징	초보적 프로그램	전문가수준의 문제해결능력을 가진 인공지능	자연언어처리 음성영상인식 강화학습[200] 심층학습	기술적 특이성[201]

출처: 佐藤 廣大, 人工知能・ビッグデータを活用した資産運用への期待と課題, 野村資本市場クォータリー 2017年春号, 野村資本市場研究所, 66頁의 표 수정

199) if~then이라는 형식의 규칙을 쌓아서 문제해결을 하는 시스템을 말한다.

200) 보상이 최대화되는 행동을 하도록 학습된 인공지능을 말한다.

201) 기술적 특이점에 대해서는 미래학자인 Ray Kurzweil이 2045년에 기술적 특이점에 도달하여 컴퓨터에 의한 인공지능의 능력이 전 인류의 지능을 상회한다고 예언한 바 있다.

그러한 동향은 한동안 계속될 것으로 보인다. 국제적으로 로보어드바이저로 인한 새로운 유형의 문제, 그중에서도 금융시장의 존속과 안정성을 위협할 가능성에 대해서는 이러한 위험을 예방하고 탐지하는 기술에 대한 투자를 지원하고 뒷받침하는 등 감독기법(대표적으로 RegTech, SuperTech)을 이용하여 통제하려는 움직임도 있다. 따라서 결국 궁극적으로 인간을 궁극적인 통제자로 보고 있음을 감안할 필요가 있다.

3. 로보어드바이저 산업과 국제적 규율 현황

가. 개념

금융 분야에서 로보어드바이저(robo-advisor)는 2002년 미국 언론에서 사용된 용어로 아직까지 이에 대한 합의된 정의는 없는 상황이다. ‘robo’는 자동화(automated)라는 의미로 사용되었으며, 비대면으로 자동화된 투자자문 또는 자산관리서비스를 제공하는 자를 일컫는 말로 사용되고 있다. “자동화”란 미리 짜여진 알고리즘을 이용하여 사람과 유사하게 투자자문 또는 자산관리서비스를 제공하는 것을 말한다. “비대면”이란 온라인 또는 모바일 환경에서 비대면으로 투자자문 또는 자산관리서비스를 제공하는 것을 말한다.

나. 유형

로보어드바이저는 독립적인 투자자문 서비스 제공 또는 플랫폼 직접 개발 여부에 따라 통상 4가지 유형으로 구분된다.

〈표 4－5〉 로보어드바이저의 유형

순수 로보어드바이저	금융당국에 등록한 투자자문업자 또는 투자일임업자가 독립적으로 온라인 투자자문 또는 자산관리서비스를 제공할 목적으로 플랫폼을 직접 개발한 경우
하이브리드 로보어드바이저	기존 투자자자문업자 또는 투자일임업자가 오프라인 뿐 아니라 온라인으로도 투자자문 또는 자산관리 서비스를 제공할 목적으로 로보어드바이저 플랫폼을 개발한 경우
로보어드바이저 어댑터	제3자가 제공하는 로보어드바이저 플랫폼을 이용하여 온라인 투자자문 또는 자산관리서비스를 제공하는 경우
로보어드바이저 개발자	제3자에게 제공할 목적으로 로보어드바이저 플랫폼을 전문적으로 개발하는 IT전문회사

현재 국내에서는 순수 로보어드바이저 출현은 지연되고 있는 상태이고 하이브리드 로보어드바이저와 로보어드바이저 어댑터가 시장을 선정하고 있는 상태이다. 국제적으로도 대체로 하이브리드 로보어드바이저가 대세이다.

다. 로보어드바이저의 특징과 현황

로보어드바이저는 서비스 한계비용이 매우 낮고 무한 복제가 가능하며 언제 어디서나 이용이 가능하다는 점에서 효과가 있다. 그러나 완전한 개인맞춤형 서비스를 제공하거나 자유로운 의사소통과 상호작용을 하기에는 한계가 있다, 다만 향후 기술의 발전으로 완전한 자율 AI수준이 될 경우 딥러닝(Deep learning)과 완전한 사람 수준의 서비스를 제공할 것으로 기대되고 있다. 그러나 현재 수준은 고객과의 상호작용에 문제가 있는 로보어드바이저의 단점을 보완하기 위해 실제 사람이 보조하는 추세이다.

현재 로보어드바이저는 자동화된 최적화로 투자자가 쉽게 투자결정할 수 있도록 돕고 고객의 투자성향과 목적에 적합하게 개인 맞춤형 자산과리서비스를 설계하고 제공할 것으로 기대되고 있다. 아울러 자산관리시장에서 자문보수와 서비스 질에 대한 경쟁을 촉진하며, 그간 부유층의 향유물처럼 인식되어온 투자자문 및 자산관리서비스의 대중화를 이끌 것으로 기대되고 있다.

국내에서 로보어드바이저 시장은 아직 공식적으로 통계가 나오고 있지 않지만 2020년 말 운용자산은 1조원을 넘어설 것으로 기대되고 있다(한국과학기술정보연구원, 2017). 국제적으로 보면 미국 IT 전문시장조사회사인 International Data Corporation사가 2016년에 발표한 바에 의하면 2020년까지 470억 달러정도 늘어날 것으로 예상하고 있다.[202] 이중 단기적으로 투자 확대가 예상되는 분야로 금융, 증권・투자, 제조 분야를 들고 있다.

라. 로보어드바이저 알고리즘 체계

금융윤리를 논하기에 앞서 금융윤리가 필요한 영역을 알기 위해서는 우선 로보어드바이저의 알고리즘 체계와 단계를 파악할 필요가 있다. 로보어드바이저에 내재한 위험과 잘못 이용될 경우 이용자는 물론 금융시장에 미치는 피해를 예방할 수 있기 때문에 무엇보다 사전에 제대로 설계될 것(well-design)이 필요하고 필수적이다.

다만 제대로 설계된 알고리즘이라고 하여 모든 단계에서 요구되는 것이 모두 법제도화될

202) http://www.idc.com/getdoc.jsp?containerId=prUS41878616

필요가 있는지는 별개의 문제이다. 즉 윤리의 영역과 구분하여 법적 규율을 제도화할 수도 있지만, 이를 자율규제영역에서만 다룰 수도 있기 때문이다. 국가별로 규제의 수위와 방법이 다른 것도 이러한 이유에서이다.

로보어드바이저 알고리즘이 잘 설계되어 있다고 하는 경우란 적어도 다음의 사항이 충족된 경우가 해당된다. 즉 일반적으로 로보어드바이저 서비스절차에 따라 알고리즘은 크게 (1) 고객진단 (2) 자산배분 (3) 리밸런싱 모듈로 구성되며 각 단계별 다음의 요소가 고려된다.

1) 고객진단단계

"고객진단단계"란 고객의 투자성향과 투자목적으로 파악하고 고객유형을 분류하기 위한 알고리즘을 말한다. 어느 나라나 대체로 금융투자회사의 경우 오프라인 또는 온라인 환경과 상관 없이 투자권유를 하기 전에 고객의 적합성 평가를 실시하도록 하고 있다. 때문에 로보어드바이저의 고객 진단 알고리즘도 법규에서 정한 적합성원칙을 동일하게 준수하여야 한다. 이 때의 고객유형 분류는 이후 자산배분과 리밸런싱에도 영향을 미치기 때문에 고개진단 알고리즘은 매우 중요하다. 고객유형 분류가 잘못된 경우 로보어드바이저의 자산관리역량과 무관하게 투자결과는 고객에게 불리하게 나오기 때문이다.

따라서 고객진단알고리즘이 법과 고객의 이익과 정합성을 갖기 위해서는 다양한 검증장치가 설계 시 마련되어야 한다. 예컨대 (i) 법규에서 요구하는 수준에 미달하는 소수의 온라인 질문으로 고객의 정합성을 평가할 경우 고객유형 분류에 오류발생이 가능하므로 이를 예방하도록 설계되어야 하고, (ii) 온라인 질문으로 추출한 고객정보를 일관되고 체계적인 방법으로 이용하지 않을 경우 고객 유형 분류에 오류가 발생할 가능성이 있으므로 이 역시 예방되도록 설계되어야 한다. 아울러 (iii) 고객의 비일관적이거나 모순된 답변을 해소할 수 없는 경우 고객 유형 분류에 오류가 발생할 수 있음을 인식하고 이를 예방해야 한다.

2) 자산배분단계

"자산배분"은 고객유형에 적합한 투자자산을 추천하거나 투자자산을 배분하기 위한 알고리즘을 말한다.

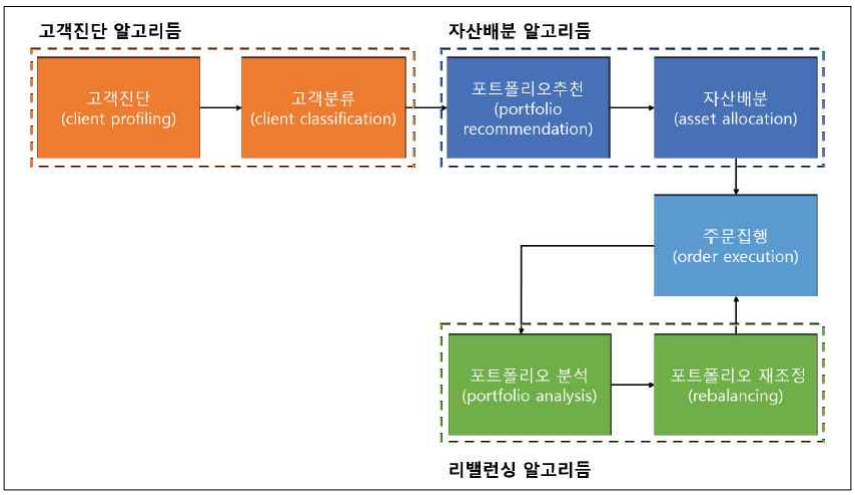
이 경우 이용되는 알고리즘은 (i) 타당성(자산배분알고리즘에서 미리 설정된 가정은 타당해야 하며 자산배분알고리즘은 주지된 기초모형에서 요구되는 요건을 충족해야 한다), (ii) 체계성(자산배분알고리즘은 미리 설정된 가정과 절차에 따라 일관적이고 체계적인 방법으로 효율적인 자산배분을 도출할 수 있어야 한다), (iii) 효율성(자산배분 알고리즘이 기초모형

을 따르지 않을 경우 자산배분이 더 효율적임을 증명할 수 있어야 한다), (iv) 적합성(자산배분 알고리즘이 추천한 자산배분은 고객에게 적합하여야 하고, 적합성 충족 여부는 자산배분 알고리즘이 추천한 자산배분의 손실위험과 기대수익률이 고객의 투자성향과 투자목적에 비례하여야 한다. (v) 보완성(자산배분 알고리즘이 사용한 이론모형의 한계가 무엇인지를 인식하고 이를 보완할 수 있는 방법론 또는 기타 다른 이론적 기반의 사용을 고려할 수 있어야 한다) 등의 요건을 충족하여야 한다.

3) 리밸런싱단계

"리밸런싱단계"란 고객의 투자목적에 맞게 포트폴리오의 성과를 분석하고 재조정하는 알고리즘을 말한다. 즉 리밸런싱 알고리즘은 고객의 이익을 최우선으로 하는지에 대하여 주기적으로 검토하여야 한다. 리밸런싱전략 또는 전술이 시장상황에 따라 적합하지 않을 경우 불필요한 거래비용 뿐만 아니라 고객의 자산배분의 손실위험을 높일 수 있다. 아울러 리밸런싱 알고리즘이 채택한 리밸런싱 기준이 시장상황에 적합한지 또는 고객 최선대우 원칙에 위배되는지 여부를 주기적으로 검토하여야 한다.

〔그림 4-1〕 로보어드바이저 투자자문 및 자산관리서비스 절차



마. 인공지능의 발달에 따른 로보어드바이저의 추가적 위험 요소

급성장하고 있는 인공지능기술이 매우 새로운 것이기 때문에 그 부작용이나 피해는 종전의 유형과 다르게 나타날 수 있다는 점이 가장 위험한 요소로 빈번히 지적되고 있다. 대표적으로 시스템에러로 인한 시장의 혼란이 가장 우려되고 있다. 특히 딥러닝을 통해 스스로 학

습을 하는 자율형 인공지능이 이용될수록 시스템 에러의 발생가능성이 더 커질 수 있다.

시스템 에러가 발생하는 원인은 매우 다양한데 대표적으로 다음과 같은 요인이 꼽히고 있다.[203] 우선, ① 과잉적합(overfiting)의 가능성이다. 이는 과도한 학습이라고 칭할 수 있는데, 사용할 수 있는 척도가 너무 많거나 너무 복잡해서 과거의 단순한 정보만으로 적합하다고 판단될 위험이 없지 않다는 점이다. 과거의 정보에 기초할 경우 테스트하여 좋다고 나오더라도 실제 적용시에는 나쁜 결과가 나오는 것과 같은 문제이다. ② 둘째, 영향을 미치는 요인들이 너무 많다는 점이다. 시장의 움직임에 영향을 미치는 요소들은 매우 많은데, 기계학습을 한다고 해도 모든 요인을 완전하게 분석하기란 쉽지 않다. ③ 셋째, 심층학습을 하는 인공지능은 학습한 데이터로 알고리즘에 의해 자동적으로 결과를 생성해 내는데, 이것을 자산운용에 응용할 경우 사후적으로는 매매의 성과형태라는 결과로만 판단할 수밖에 없다는 점도 우려를 낳고 있다. 즉 왜 그러한 결과가 되었는지 전혀 그 과정을 알 수 없다는 점이다. 실무에서는 이를 '블랙박스'의 문제라고 부르기도 한다. 블랙박스는 어떻게 작동하는지를 자세히 알지 못하더라도 투입에 비해 산출이 기대 이하이면 하자 또는 결함이 있다고 추정이 가능하다. 이러한 점에서 투입과 산출에 대한 투명성을 강조한다고 하더라도 어떻게 작동하여 그러한 결과가 나오는지는 현재의 기술로는 설명할 수 없는 문제가 있다. ④ 마지막으로 인공지능에 의한 자산운용이 고도로 발달하여 광범하게 보급될 경우 여러 알고리즘이 서로 통신하여 가치급변을 야기할 수도 있다는 우려도 제기되고 있다.

바. 로보어드바이저에 대한 글로벌 규제동향

1) 법제도적 동향

최근 1~2년간 각국에서 로보어드바이저에 대한 규제를 정비하거나 마련하려는 움직임이 나타난다. 그러나 각 국가마다 규제의 정도와 내용에 차이를 보이고 있다. 다만 건전한 영업행위 확립과 투자자보호라는 목적과 방향 면에서는 대체로 일치하고 있다. 그러나 이러한 규제는 최소한의 법적 규제라는 점에서 윤리적인 차원의 고도의 행동규칙을 정한 수준에는 이르지 못한다.

특히 글로벌 규제에서 초점을 두는 부분은 다음의 부문으로 이중 일부는 법적 규제수준으로 준수를 요구하는 국가도 있는 반면에 그렇지 않은 국가들은 자율규제 차원에서 준수를

203) 자세한 내용은 佐藤 廣大, 人工知能・ビッグデータを活用した資産運用への期待と課題, 野村資本市場クォータリー 2017年春号, 野村資本市場研究所, 77~78면 참조.

권고하는 수준에 그치고 있다. 그러나 로보어드바이저의 진화 수준에 따라 후자의 영역은 법제화로 규제 수준이 상향될 것으로 예상된다.

법적 규제의 관심과 초점이 맞추어지는 영역은 크게 다음과 같이 다섯 가지로 정리할 수 있다.

첫째, 고객확인 및 적합성원칙 준수(know your customer and suitability)이다. 이는 현재 오프라인 투자자문/일임업자 등 증권회사, 자산운용회사, 자문회사등과 동일하게 로보어드바이저에게도 적합성원칙을 준수하도록 요구하는 것을 말한다. 이 경우 국가에 따라서는 자문서비스를 제공할 금융상품에 적합한 고객을 선정할 수 있는 절차(filtering process) 뿐 아니라 인터페이스 설계(interface design)에서도 자문서비스를 제공할 고객의 특성에 맞게 기술중립적으로 설계하도록 요구하고 있다(영국 FCA).

둘째, 알고리즘 디자인 및 감독이다. 로보어드바이저에 사용된 알고리즘은 적어도 다음 요건을 갖추어야 한다. 충분한 전문성을 갖춘 투자전문가가 알고리즘을 개발하고 상시 감시하는 과정에 밀접하게 참여해야 하고,알고리즘 가정(assumption)은 일반적으로 인정된 투자이론에 기반해야 하며, 알고리즘 가정에 대한 설명서를 알기 쉬운 언어로 작성하여 고객이 언제라도 볼 수 있도록 공시하는 한편, 제3자 알고리즘은 엄격히 심사(due diligence)한 후 자기책임 하에 이용해야 한다. 그 외에 호주 증권감독기관인 ASIC는 2016년 3월에 알고리즘을 제공하는 IT업자에 대하여 다음의 사항을 요구하는 규제를 발표한 바 있다. 로보어드바이저 제공업자는 자문업자에게 요구되는 최소교육 및 역량기준을 충족하는 책임자를 적어도 1명 이상 설치해야 한다. 또한 로보어드바이저 제공업자는 로보어드바이저 알고리즘이 적정하게 설계되고, 모니터되고, 시험되는지를 다음과 같은 방법과 절차에 따라 확인해야 한다.

Ⓐ 알고리즘의 목적, 범위, 설계에 관하여 명확하게 기술한 시스템디자인문서를 작성해야 하며, 이 안에는 알고리즘의 의사결정순서도(decision tree)와 규칙이 포함할 것

Ⓑ 알고리즘 테스트의 범위를 설명한 테스트전략 문서 작성. 이 안에는 테스트계획, 사례, 결과, 결점해결안 및 최종결과를 포함하여야 함. 알고리즘의 건전성 테스트는 서비스 개시 전에 반드시 설시. 이후에는 정기적으로 실시

Ⓒ 알고리즘 수정을 관리하기 위한 절차를 마련해야 함. 이 안에는 알고리즘으로의 승인되지 않은 접근을 감시하고 방지하기 위한 보안절차 포함

Ⓓ 알고리즘의 변화를 통제, 감시, 재구성

Ⓔ 시장 변화 및 법제도 변화로 영향받는 요인이 있을 때마다 알고리즘을 검토·갱신

Ⓕ 알고리즘안에 오류가 감지되면 자문서비스를 중단할 수 있는 권한과 절차 마련

Ⓖ 로보어드바이저에 의해 제공된 투자자문에 대한 시의적절한 검토를 통해 알고리즘의 성능을 감시하고 감독할 수 있는 인적 및 기술적 자원을 적절하게 보유

Ⓗ 아울러 자격을 갖춘 투자자문사가 IT업체의 로보어드바이저 품질을 감시하고 시험할 수 있도록 절차를 마련

셋째, 공시 기준과 비용 등에 대한 투명성 강화가 이루어져야 한다. 서비스범위, 수수료 체계, 로보어드바이저 작동체계, 투자위험 등에 대해 투자자가 알기 쉽게 공시하거나 설명하여야 한다. 대표적으로 미국에서는 로보어드바이저와 고객간의 상호관계에 한계가 있다는 점을 감안하여 비즈니스 모델에 대하여 다음과 같이 고객에게 상세하게 제공하도록 규제하고 있다(SEC, 2017년 2월 Guidance Update, No. 2017-02).

- 고객에게 알고리즘이 고객의 계좌를 관리하고, 포트폴리오를 추천하고 자동으로 리밸런싱하는데 사용된다는 점을 설명
- 알고리즘에 사용된 가정과 알고리즘의 한계를 설명
- 알고리즘 이용으로 발생할 수 있는 위험에 대해 설명(예를 들어 시장상황과 특별한 관계가 없거나 고객의 기대와 달리 자주 리밸런싱이 일어날 수 있는 위험 또는 시장여건의 오랜 변화를 적시에 반영하지 못할 위험)
- 시장이 위기상황에 직면한 경우 알고리즘의 거래가 중지되거나 다른 방어적인 조치가 취해질 수 있다는 점을 설명
- 제3자가 알고리즘의 개발, 관리, 소유와 관련된 경우 그 사실과 함께 이해상충의 소지를 설명
- 로보어드바이저에 의해 직접 고객에게 부과되는 수수료, 고객이 직간접적으로 부담할 수 있는 비용에 대해 설명
- 상시 감시 및 고객계좌관리에 사람이 얼마나 관여하는지 설명
- 로보어드바이저가 포트폴리오를 추천할 때 고객이 제공한 정보를 얼마나 이용하는지 그 한계에 대해 설명

- 고객이 언제 어떻게 로보어드바이저에게 제공한 정보를 업데이트 해야 하는지 설명
- 로보어드바이저가 제공하는 투자자문서비스의 범위에 대해 고객에게 오해의 소지가 없게 분명하게 알려야 함
- 로보어드바이저는 핵심설명사항, 투자위험, 면책사항 등 중요정보를 고객에게 온라인으로 제공할 때 다음 사항을 고려
  → 핵심설명사항은 고객이 합리적인 투자의사를 결정할 수 있도록 계약을 체결하기 전에 제공
  → 핵심설명사항은 팝업상자와 같은 웹기술을 이용해 특별히 강조
  → 고객의 추가설명이 필요한 정보를 제공할 때에는 고객과 상호작용이 가능한 웹기술(tooltips등) 활용
  → 중요정보를 공시할 경우 운영체계, 웹 브라우저, 기기 등에 영향받지 않도록 기술중립적으로 설계
- 로보어드바이저가 신의성실의무에 따라 고객의 이익을 최우선으로 대우하고 적합한 투자자문을 제공하기 위해 온라인으로 질문할 경우 다음의 사항을 고려할 것
  → 온라인 질문이 로보어드바이저가 고객의 금융상황과 투자목적에 근거하여 고객에 적합하고 적정한 투자자문을 제공하기에 충분한 정보를 추출할 수 있어야 함
  → 온라인 질문이 직관적으로 이해할 수 있게 쉬운 언어로 작성되어야 하며 추가적인 설명이 필요한 경우에 부연이나 예시를 제시할 수 있어야 함
  → 고객이 온라인 질문에 비일관적으로 응답할 경우 고객에게 이를 알려 고객이 제공하는 정보를 수정하게 하거나 시스템에서 이를 자동적으로 적시해 재검토를 요구할 수 있게 함
- 로보어드바이저는 고객이 자신의 투자성향과 투자목적에 맞지 않게 자산배분을 조정할 경우 이에 대하여 추가적인 자문을 제공할 수 있어야 함

넷째, 거래 시 고객의 최선 집행 원칙을 고려해야 한다. 합리적으로 설계된 절차와 방식으로 포트폴리오 주문을 관리할 수 있어야 하고, 이를 통해 고객에게 최선집행(best execution)을 제공할 수 있어야 한다.

마지막으로 데이터 보호와 사이버안전에 유의하여야 한다. 이를 위해 최고 단계의 데이터 암호화, 제3자 데이터 리스크관리 위탁, 사이버 보안 보험가입, 보안사고 대비책 마련 등을

충족시켜야 한다. 또한 상시 감시의무, 기술지원 인력, 개인정보 수집 제한, 적합성 자가진단 금지 등도 규제대상으로 하고 있다. 무엇보다 로보어드바이저가 알고리즘에 의해 작동된다는 점, 고객과의 상호작용에 한계가 있는 점, 인터넷으로 투자자문서비스를 제공한다는 점을 고려하여 다음의 준법감시프로그램을 운용하도록 하고 있다.

- 알고리즘 코드의 개발, 시험운용, 검증시험과 알고리즘 성능에 대한 사후 모니터링
- 온라인 질문이 신의성실의무에 따라 적합한 투자자문을 제공하기에 충분한 정보를 고객으로부터 추출하는지에 대한 검증
- 알고리즘 코드의 변화를 고객에게 알린 경우 고객의 자산배분에 어떤 영향을 미치는지에 대한 검토
- 로보어드바이저에 사용되는 알고리즘 코드 또는 소프트웨어를 개발, 소유, 관리하는 제3자에 대한 적정한 감시
- 사이버보안 위협의 방지, 감시, 대응정책 및 절차
- 로보어드바이저의 투자자문서비스를 마케팅하기 위해 사용하는 소셜 및 다른 형태의 미디어 사용의 적정성
- 고객 계좌와 핵심 투자자문서비스 시스템의 보호

FINRA(미국증권업협회)도 다음의 사항을 금융회사가 갖추어야 한다고 제시하고 있다(2016).

- 알고리즘에 사용된 방법론과 가정이 로보어드바이저가 제공하고자 하는 서비스에 잘 맞는지를 평가
- 알고리즘에 사용될 데이터 입력값 이해
- 알고리즘 결과가 회사의 기대와 상응하는지를 평가하기 위한 시험
- 알고리즘에 사용된 모델이 시장 또는 다른 여건의 변화에 따라 적정하게 작동하는 지를 평가
- 로보어드바이저가 의도 또는 계획한 대로 작동하는지를 확인하기 위해 일반적인 상황에서 로보어드바이저의 결과를 시험
- 로보어드바이저 감독에 책임있는 자를 명확하게 지정
- 회사는 로보어드바이저를 오프라인에서 투자자문도구로 활용할 경우 임직원에게 로보

어드바이저는 허가된 사람만 가능한 점, 로보어드바이저에 사용된 가정과 한계, 로보어드바이저 사용에 부적합한 고객 유형을 교육할 것

2) 법제도적 규제의 시사점

앞서 살펴본 현재 각국의 법제도적 규제 동향에서는 공통적으로 고객진단 알고리즘, 자산배분 알고리즘, 리밸런싱 알고리즘 등 단계별로 문제되는 사항을 해소할 수 있는 개별 조치들을 마련하고 있고, 특히 알고리즘 관리체계에 대해서도 규율하는 경향을 보이고 있다. 즉 알고리즘 기술서 작성 및 관리－알고리즘 기술서 작성절차－알고리즘 수정절차－알고리즘 상시감시－알고리즘 성능 평가－알고리즘 오류대응－알고리즘 보안－제3자 위탁관리－알고리즘 관리지원 등 프로세스 전반에 대하여 규율을 하고 있음을 알 수 있다.

다만 이러한 규율을 어느 정도로 상세하게 하느냐, 그리고 공적 규율을 통해서 할 것인지 아니면 자율규제기관을 통해서 규율할 것인지는 국가별로 차이를 보이고 있다. 공적 규율대상으로 하지 않을 경우 자율규제기관에서 규정을 통해 하거나 직무윤리의 영역으로 요구할 수 있을 것이다.

다만, 현재의 규율은 사람의 통제가 가능함을 전제로 로보어드바이저 서비스에 대해 법과 제도를 통해 규율하고 있다는 점에서 한계를 찾을 수 있다. 따라서 향후 인공지능의 발전정도에 따라서는 사람의 통제없이 인공지능 스스로 로보어드바이저 서비스를 제공하는 것에 대응하는 방안도 함께 마련되어야 할 것이다. 이 경우에도 법과 제도등의 공적・자율규제기관의 통제를 통해서 할 것인지는 결정되어야 할 것이지만 넓게 윤리적 관점에서 위험에 대한 통제요소들을 사전에 추출할 필요는 있다. 이하에서 이에 대해 분설하기로 한다.

첫째로, 자율형 AI의 행위에 대한 사법적 취급을 명확히 해야 한다.

사람의 통제 없이 인공지능 스스로 로보어드바이저 서비스를 제공하는 경우 무엇보다 권리와 의무의 주체가 될 수 있는지가 사전에 명확히 할 필요가 있다. 인공지능이 의사표시를 하여 매매계약을 성립시켰다고 하더라도 현행법상 인(자연인 내지 법인)만이 권리의무의 주체가 된다. 따라서 인공지능이 한 계약은 현행법적으로는 인공지능 이용자(또는 운용자)와 거래상대방 간에 성립하게 된다. 그런데 이는 어디까지나 인공지능의 판단에 따라 매매한다는 이용자의 의사가 표시되어 있다고 전제할 때 가능하다. 그러나 만약 인공지능이 그 이용자가 전혀 상정하지 않은 계약을 체결하는 경우(예컨대 계약상대방을 선택하고 계약조건을 협상하고 계약을 체결한 경우 등) 내심의 의사와 인공지능이 한 구체적인 표시의 차이를 착

오로 주장하여 계약을 취소할 수 있는지가 문제된다. 민법 제109조는 법률행위 내용의 중요부분에 착오가 있는 경우 의사표시를 취소할 수 있다고 규정하고 있다. 그러나 그 착오가 중대한 과실로 인한 때에는 취소하지 못한다고 하고 있다(동조 단서). 표시에 관한 중요한 착오의 경우 취소가 되겠지만, 인공지능이 스스로 상정한 범위 내에서 행위한 것을 표시의 착오로 보기는 어렵고 동기의 착오에 해당하기 때문에 취소가 인정되기는 현행법상 어려울 것으로 보인다.

이처럼 인공지능이 자체 학습을 통해 자율적으로 판단하고 행동하는 경우 현행법 하에서는 법적 취급이 마련되어 있지 않아 불확실하다. 반면에 외국의 경우를 보면 고도의 인공지능을 이용자의 대리인(agent)으로 보는 입법례가 있다. 대표적으로 통일상법전 제2B편 초안이 해당하는데(U.C.C. §2B-102(22)), 이 초안을 이어받아 통일전자거래법에서도 전자적 대리인의 개념(Section 2(6))과 전자적 대리인에 의한 자동거래(Section 2(2))를 규정하고 있다(최경진, 2015: 209).

한편, 인공지능의 행위가 채무불이행이 되거나 불법행위에 이르는 경우도 쉽게 상정할 수 있다. 이 경우 누가 책임을 질 것인지가 문제된다. 불법행위에 기해 책임을 물을 수 있기 위해서는 현행법상 인공지능의 소유자에게 고의 내지 과실이 있을 것으로 요구하고 있다(민법 제750조). 민법에서 말하는 과실이란 구체적인 결과의 발생을 예상할 수 있음에도 불구하고 그 결과의 발생을 회피하기 위하여 필요한 조치를 하지 않았을 것이 요구된다. 그렇다면 기계학습 내지 심층학습으로 자기가 스스로 학습하고 자율적으로 판단한 인공지능을 상정할 때 그러한 결과의 발생을 예상할 수 있는지 내지 예상할 수 있는 상황이 무엇인가가 문제될 수 있다. 이 경우 피해자구제는 곤란해질 수밖에 없다. 왜냐하면 이러한 과실요건을 인공지능의 소유자가 충족하지 못할 경우 피해는 발생하였지만 그럼에도 불구하고 아무도 책임을 지지 않는 결과를 가져오기 때문이다. 더구나 과실의 판단에 있어 결과발생을 예상할 수 있었는지 내지 예상되는 결과를 회피하기 위해 필요한 장치를 갖추지 못하였는지는 통상 일반적으로 불법행위자가 속하는 평균적인 그룹을 전제로 판단하는데 이 경우 인공지능의 소유자가 기성소프트웨어로 제작된 인공지능을 구입·사용한 경우에는 예견가능성이 인정될 여지가 줄어들게 된다. 결국 고도로 발달된 인공지능의 행위로 인해 그와 거래한 상대방이 일반소비자인 경우 일반 불법행위책임에 기해서는 보호받지 못할 가능성이 매우 크다. 이러한 문제점을 인식하여 한편에서 동물 점유자의 책임을 유추하여

불법행위책임을 묻거나[204] 인공지능은 동물보다 더 위험하다는 이유를 들어 엄격하게 위험책임을 지우자는 의견도 없지 않다.

소프트웨어 개발자에게 책임을 물을 수 있는지의 여부도 문제될 수 있다. 피해발생을 예상할 수 있었다면 책임을 묻는 것이 어렵지 않지만 개발된 소프트웨어를 판매한 이후 기계가 스스로 학습하여 그 결과로 피해가 발생한 경우, 이러한 경우까지 소프트웨어 개발자에게 책임을 물을 수 있는지는 의문이기 때문이다. 소프트 개발자가 결과발생에 대하여 구체적으로 예상할 수 없는 경우까지 책임을 지게 한다면 너무 가혹할 뿐 아니라 인공지능산업의 발전에도 저해가 되기 때문이다(福岡眞之介・桑田寛史・料屋 惠美, 2017: 193).[205] 그 외 발생한 손해와 관련하여 책임범위를 제한할 필요가 있는 것은 아닌지에 대해서도 논의할 필요가 있다. 즉 민법상의 원칙인 상당인과관계의 범위로 할지 아니면 인공지능의 특수성을 고려하여 책임범위를 한정해 두어야 할지 등에 대해서도 의견이 분분할 것으로 예상된다.[206]

둘째로, 인공지능의의 행위에 대한 형사법적 취급을 명확히 해야 한다. 형사책임을 지우는 것과 관련해서는 더 곤란한 문제에 직면할 것으로 예상된다. 고도의 지능형 인공지능이 자신의 판단으로 피해를 가하여도 처벌할 수 없기 때문이다. 양심과 형벌감수성이 결여된 인공지능에게 인격이나 책임능력, 형벌능력을 인정할 수 없기 때문이다(손영화, 2016: 135). 이와 달리 인공지능에게 인격을 인정할 경우 형벌을 부과할 수 있다는 접근방법도 불가능하지는 않지만 인공지능에게 있어 처벌의 의미가 있는지 근본적으로 의구심이 없지 않다. 한편, 가해행위의 책임자로서 그 이용자에게 책임을 지우려고 하더라도 위법성이 미흡하다고 판

---

204) 그러나 기계를 동물에 준하여 취급할 수 있는지, 즉 유추를 정당화하는 비교가능성이 있는지는 의문이라고 한다. 왜냐하면 지능형 로봇은 이용자가 활성화와 비활성화중 선택할 수 있지만, 동물에서는 이와 유사한 상황이 전제되지 않기 때문이다. 또한 동물의 위험성은 동물의 본능적 행동에서 비롯되지만 지능형 로봇은 본능이라는 것을 알지 못한다. 더욱이 동물점유자의 책임은 위험책임의 원리에 의하여 강화된 과실책임 내지 위험책임과 비슷하게 운용되는 중간책임에 해당하는데, 그와 같은 엄격한 책임은 유추와 친하지 않다(김진우, 2017: 33).

205) 개발자는 소프트웨어의 용도와 동작조건의 안전성을 담보할 수 있도록 일정한 범위에서만 인공지능이 학습하고 동작하도록 제한을 두고 어떠한 학습을 하더라도 상정한 것 외에는 동작할 수 없도록 대응할 가능성이 크다. 그러나 이 역시 너무 보수적인 접근방법일 수 있고 인공지능산업의 발전을 더디게 할 우려가 없지 않다.

206) 이에 대하여 책임범위에 별도 제한을 두지 않더라도 손해배상은 손해와 가해원인 사이에 상당인과관계가 있는 범위로 제한되는 점(민법 제393조)과 이 분야에 대해 의무보험 내지 보상기금의 도입을 전제로 책임범위를 제한하지 않는 것이 바람직하다는 주장도 있다(김진우, 2017: 40).

단될 가능성이 높아 여전히 한계가 예상된다. 이처럼 앞서 한 논의에 의하면 민사배상이건 형사처벌이건 귀책에 따른 현행법상의 책임추궁장치들은 고도의 지능형 인공지능에게는 그다지 효과가 없을 수 있다.

셋째로, 인공지능의 행위에 대한 제재 방법을 명확히 해야 한다. 현행법상 마련된 형사벌과 행정제재인 과징금・과태료・임직원 지위 박탈 등의 제재수단은 사실상 인공지능에는 유용하지 않게 된다. 이처럼 제재 방법은 인공지능 관련 행위책에 대한 대응방안을 고려할 때 매우 어려운 과제를 노출시킨다. 현행법상 사법상의 규율은 물론, 산업규제 및 시장질서 차원에서 마련해 둔 규제적 성격의 규범을 인공지능의 자체 판단과 외부환경과의 상호작용에 의해 규정위반사항이 발생하더라도 이를 위해 마련한 제재규정들의 효과 및 이에 기한 억지효과는 더 이상 기대하기 어렵기 때문이다. 때문에 향후 제재효과를 기대할 수 있도록 현행 규정을 보완하거나 대체하는 제도를 마련하여야 할 것이다.

넷째로 인공지능이 투자자에게 부담하는 의무를 확대해야 할 것이다. 인공지능을 이용하는 단계에서는 무엇보다 기존의 금융관련법제에서 활발하게 논의가 되지 않았던 금융회사의 선관주의의무에 관하여 보다 주의를 기울이고 정비해야 할 것으로 보인다. 예컨대 투자자의 이해를 돕기 위하여 추가적인 설명이나 예시를 충분히 활용하거나 일관성 없는 답변이 되지 않도록 추가적인 질문항목이나 조치를 충분히 반영하여 투자자성향 및 이해도를 파악하는 방안들을 이용할 필요가 있다.[207] 아울러 분쟁의 단초가 되는 복잡한 알고리즘에 대한 접근성 담보차원에서 분쟁시 투자자가 요구할 경우 알고리즘에 대한 정보를 제공하는 권리를 명시하는 방안에 대해서도 마련할 필요가 있다.

다섯째, 알고리즘 관련 금융회사 내부통제・임직원 보수연수를 강화할 필요가 있다. 미국 규제기관은 대체로 위험관리를 강화하고 알고리즘 개발 및 사전에 설정한 대로 작동되는지 테스트하는 한편, 사후 모니터링할 의무를 부과하면서, 한편으로는 알고리즘소스코드를 감독기관에 제공하는 것으로 알고리즘이 야기하는 위험에 접근하는 방법을 취하고 있다. 또한 알고리즘거래전략(algorithmic trading strategies)을 설정・개발・수정을 주로 담당하는 자와 이들 업무를 일상적으로 감독하는 자 내지 지시자의 경우 자격시험에 합격한 자만 가능하게 하고 있으며 이들에게 보수연수를 의무화하고 있다. 이러한 접근방법은 지능형 인공지능을

207) 미국의 경우 투자자성향을 파악하기 위한 지침에서도 이를 요구하고 있다. 자세한 것은 SEC, IM (Division of Investment Management) Guidance Update ROBO-ADVISERS, 2017.2 P.6

범용화시키는 단계에서도 유용할 것으로 보인다.

여섯째, 지능형 인공지능 알고리즘에 윤리사항 입력을 의무화할 필요가 있다. 알고리즘 관련 임직원의 연수 강화 외에 알고리즘에 윤리사항을 입력하는 것을 의무화하는 방안도 생각해 볼 수 있다. 사전에 설정한 범위 내에서만 판단하고 투자결정을 하는 것으로 예상하여 프로그램을 만들 수 있지만 지능형 인공지능의 경우 외부와의 상호작용에 의해 자율적으로 판단함으로 인해 야기되는 위험을 해소하기 위해서이다. 임직원에 대해 보수연수를 의무화하는 것과 마찬가지로 지능형 인공지능 알고리즘도 윤리사항을 준수하도록 하는 방안이 그것이다.

마지막으로 국제적 감독 공조 강화를 고려해 볼 수 있다. 로보어드바이저에 관한 국내 감독기관들은 국내 차원에서의 규율에만 관심을 갖고 있는 것으로 보인다. 그러나 로보어드바이저의 이용확대 속도와 정도를 고려할 때, 그 규모가 확대될 경우 금융시장에 미치는 영향은 매우 폭발적일 수 있다. 예컨대 동일한 알고리즘의 확대 및 동일한 판단이 이루어지는 경우 가히 금융시장이 경험할 혼돈은 상상하기가 어렵지 않다. 그러한 점에서 한 국가의 감독에만 초점을 맞추어 대응하는 것은 적절하지 않다.

4. 로보어드바이저 산업·시장 참여자 및 각 참여자별 윤리 이슈

가. 금융산업·시장 참여자

이상의 기술한 바에 따라 로보어드바이저 산업에 참여한 각 참여자별 대응이 필요한 윤리 이슈를 정리하면 다음과 같다.

〈표 4-6〉 각 단계 행위자별 윤리이슈

개발단계 (디자인, 프로그래밍)	개발자	• 알고리즘 개발단계에서의 윤리 특수성, 금융규제의 목적 투입 • 투자전문가가 투입(인정된 투자이론에 기초하여 설계)·상시감독
	자율규제기관	AI 개발 촉진정책과 규제의 균형
	공적규제기관	AI 개발 촉진정책과 규제의 균형
	금융종사자	AI 이용과 관련한 윤리교육, 금융규제 준수 교육
허가단계 (판매·광고· 이용)	개발자	허가를 위한 금융윤리적 자가검증도구 개발 및 적용
	공적규제기관	허가과정에서 금융윤리적 검증도구 개발 및 역량 강화 (알고리즘과 데이터 평가능력, 이용자에게 제시 및 의사결정하게 제시되는 선택구조, 정보기술인프라, 인공지능이용이 대규모화될 경우 시장에 미치는 영향 등)

이용단계 (판매・광고・이용)	금융종사자	금융자문인	• AI 로보어드바이저의 특수성 및 위험 교육 • 이용자 보호 프로세스 준수 • 이해상충 방지 프로세스 준수 • AI위험을 쉽게 알 수 있는 방법으로 설명・공시기법 선택 • 개인정보 보호
		〔금융회사〕 은행・증권회사・보험회사・자산운용회사・선물회사	• 이용자 보호를 위한 최적화된 알고리즘 이용 및 검증 담보 • 이해상충 방지 프로세스 검증 • AI위험을 쉽게 알 수 있는 방법으로 설명・공시기법 선택 • 알고리즘 투명성 담보 • 개인정보 보호 • 해킹방지등 침해방지를 위한 인적・물적 시설 적정수준 담보 • 금융시장 안정성 담보 • 이용자 피해 발생시 책임범위
	자율규제기관 (거래소, 금융협회 등)	• 금융회사 이용자 보호와 규제의 균형 • 금융시장 보호(금융시장의 공정한 가격발전기능과 안정성 확보) • 금융윤리의 규범화 수준과 범위 • AI기술에 관한 정보 및 대응 공유 • AI 기술의 수혜 격차 문제	
	공적 규제기관	• 금융회사 이용자 보호와 규제의 균형 • 금융시장 보호와 규제의 균형 • 금융윤리의 규범화 수준과 범위 • AI기술에 관한 정보 및 대응 공유(감독기관간, 전문기관간, 전문가간) • AI기술의 수혜 격차 문제	
	이용자	• 결정과 선택권 문제 • 윤리적 투자 • 금융정보의 활용과 프라이버시 보호 • 해킹에의 안전	
	기타(기관투자자)	• 결정과 선택권 문제 • 윤리적 투자	

나. 고려해야 할 세부 지표

위의 윤리이슈에 대하여 법적 규율과 윤리적 대응 중 어떤 방법을 선택할 것인지의 문제는 다음의 세부 지표를 고려해야 판단해야 할 것이다.

– AI 발전(딥러닝과 기계학습 포함)・이용 규모 및 정도 (자율형 AI의 범용화, AI간의 통신 등 담합가능성 등)

- AI 설계 검증기술의 발전 정도
- AI에 내재된 위험정도(오류, 해킹, 잘못된 설계, Blackbox 대응, 금융시장에 미치는 영향 등)
- 감독기관의 검사·감독 역량
- 금융회사 내부통제·금융소비자보호 수준
- AI 로보어드바이저에 대한 금융소비자의 이용정도(집단행동의 정도)·이해수준·교육수준
- 금융소비자의 로보어드바이저에 내재한 위험 인식과 선호도 정도
- 로보어드바이저 관련 법제도 정비수준(헌법·사법적 전자인간 인정, 형사법제·행정법제 정비정도, AI 관련 상품다양화, 상품간 규제차별 해소, AI 이용시 규제준수비용, 경쟁의 담보 등)
- AI의 불공정행위유형과 피해 구제방법과 수준
- 국가 간 공조와 정보 공유수준

제 4 절 의료 분야 인공지능 윤리 이슈

1. 논의의 배경

인공지능은 딥 러닝(deep learning) 기능을 핵심으로 보유하고 있는데 이는 '주어진 데이터의 특징을 스스로 학습하여 매우 정확한 답을 찾아낼 수 있는' 능력을 의미하는 것으로, 최근 들어 이러한 능력을 활용한 첨단 응용기술이 사회의 여러 분야에서 적용되고 있으며 그 대표적인 분야가 바로 보건의료 분야이다.

의료 인공지능은 맞춤의료, 정밀의료, 검사 남용 및 과잉진단 해소 등은 물론 오진 및 부작용 감소에도 기여할 것으로 기대되고 있다. 또, 질병의 사전 예측을 통해 예방 및 치료 효과를 제고함으로써 고령화 시대의 건강과 장수 욕구에 크게 부응할 전망이다. 의료계에서는 인간의사의 과도한 노동시간을 덜어주고 환자-의사 사이의 진료 환경을 개선해줄 수 있을 것으로 예상하고 있다. 반면, 의료 인공지능의 등장으로 의사의 권한 축소, 개인정보 유출, 의료 수혜 격차 등과 같은 문제들도 등장할 전망이어서 다양한 논쟁이 줄을 잇고 있다. 의료 인공지능의 등장과 관련하여 이러한 상반된 변화들이 예상되는 이유는, 바로 인공지능의 기

술적 특성 때문이다. 특히, 인간의 생체(몸)를 직접적으로 다루는 의료 분야와 연관되면서 이러한 기대와 반감은 갈수록 고조되는 양상이다.

인공지능(人工知能 ・AI: artificial intelligence)은 인간의 인식, 판단, 추론, 문제 해결, 그 결과로서의 언어나 행동 지령, 더 나아가서는 학습 기능과 같은 인간의 두뇌 작용과 기능을 기계(컴퓨터)를 통해 구현하는 기술로 정의되어 있다(한국사전연구사 편집부, 1994). 그리고 인간의 뇌 구조를 모방한 '뉴럴 네트워크(neural network)'시스템을 통해 '빅 데이터(big data)'를 딥 러닝(deep learning: 심층학습)하는 것이 인공지능의 핵심 메커니즘이다(뉴턴하이라이트121, 2018; 마쓰오 유타카・박기원, 2015; 미야케 요이치로・모리카와 유키히토・김완섭, 2017; EY어드바이저리・부윤아, 2016; 최윤식, 2016).

이러한 인공지능의 특징과 작동 메커니즘은 현재 다양한 분야에 응용되고 있는데, 전(全)세계적으로 고령화 추세가 심화되면서 특히 보건의료 분야에서의 활용이 활발하게 진행되고 있다. 인공지능의 의료적 활용은 이른바 'AI의사의 출현'과 같은 표현으로 화제가 되면서 사회적으로 큰 우려와 윤리적 논쟁을 불러왔는데, 그 이유는 인공지능이 인간만이 가지고 있다고 여겨져 왔던 판단, 추론은 물론 창조성, 영감의 능력에까지 도전하는 형태로 발전할 것이라는 전망에 따른 것이다.

그렇다면 의료 인공지능, 혹은 인공지능 의료기술로 언급할 수 있는 보건의료 분야의 인공지능 활용이 과연 '인간의사를 추월한 AI의사의 등장'으로 귀결될 것인가? 이러한 질문에 대해 현재로선 명확하게 '예, 아니오'로 대답할 수는 없다. 인공지능은 인간의 능력을 초월하는 기능을 갖추는 단계로 발전해 나갈 수 있기 때문에 이러한 능력이 의료 분야에서 활용될 경우 결국에는 인간의사의 의료능력을 초월하는 AI의사의 출현이 불가피할 수 있다. 그러면 신기술의 종국적 전망이 AI의사의 출현, 이에 따른 인간의사의 헤게모니 상실 등으로 이어질 가능성을 어떻게 제어할 것인가? 이러한 윤리적 논쟁과 우려를 해소하기 위해서는 인공지능 의료기술에 대한 존재론적 개념 규정이 우선되어야 한다. 그리고 신기술에 대한 존재론적 규정은 ELSI(Ethical・Legal・Social Implication: 윤리적・법적・사회적 영향) 개념들에 기반 하여 정리하는 것이 바람직하다.

인공지능은 종국적으로 인간의 능력을 초월하는 기능을 지향하여 발전할 수 있다. 그러나 인공지능 기술의 수준을 어디까지 발전시킬 것이며, 어디까지 기술의 활용을 허용할 것인가

라는 문제는 기술적인 영역이 아니라 사회적인 수준에서 합의해야만 한다. 왜냐하면 기술을 제안하고 구성하고 그것을 활용해나가며 경우에 따라서 만들어낸 기술을 폐기처분하는 것도 모두 인간과 기술의 관계 맥락에서 비롯되는 '사회문제(social problems)'의 영역을 벗어날 수 없기 때문이다. 앞서도 언급한 바와 같이 인공지능은 앞으로 인간보다 더 뛰어난 능력을 구사할 수 있다, 그러나 향후 인공지능을 어느 수준으로까지 발전시킬 것인가는 결국 인간의 의지와 선택에 달려 있다. 하지만 이러한 인간의 의지와 선택을 과학기술자, 발명가, 기술혁신가 그리고 정책가, 법률가 등의 양심과 이해관계 맡겨둘 수는 없다.

현재의 인공지능은 레벨1(단순한 제어프로그램 수준), 레벨2(단순한 제어프로그램의 조합을 통해 다양한 작동패턴을 구사하는 수준. 고전적 인공지능, 청소로봇 등), 레벨3(데이터를 바탕으로 규칙과 지식을 스스로 학습하는 기계학습(machine learning)이 가능한 수준. AI 체스프로그램), 레벨4(딥 러닝을 적용한 인공지능)에 이르는 수준으로 다양하게 구현되어 왔는데, 현재 레벨4의 단계가 적극적으로 구현되고 있다(뉴턴 하이라이트121, 2018). 여기서 레벨3까지의 인공지능을 '약한 인공지능(Artificial Narrow Intelligence)'이라고 부르며, 레벨4 단계의 인공지능을 '강한 인공지능(Artificial General Intelligence)' 또는 일반 인공지능, 범용 인공지능 등으로 부르기도 한다. 강한 인공지능은 약한 인공지능과 달리 스스로 사고하며, 종류를 가리지 않고 다양한 문제를 인간 정도의, 혹은 인간을 능가하는 수준으로 해결할 수 있다.[208)]

그러나 고도의 정밀성을 제공하는 인공지능의 이러한 특징과 능력들은 인간의 판단과 추론 능력을 뛰어넘을 수 있다는 점에서 기대와 동시에 다양한 논쟁을 낳고 있다. 인공지능의 활용으로 급속하게 변화하고 있는 보건의료 환경은 우리 사회에 어떠한 변화를 가져올 것인가? 과연 인간에게 건강과 장수의 꿈을 실현해줌으로써 유토피아 사회를 열어줄 것인가? 특히, 인공지능의 손에 인간의 몸(생체)을 전적으로 맡기게 된다는 것을 어떠한 의미로 받아들여야 할 것인가?

---

208) 인공지능 관련 학자, 미래학자들의 경우 강한 인공지능은 '순환적 자기개선(recursive self-improvement)'을 통해 더 나은 프로그램을 반복적으로 프로그래밍함으로써 지능 대폭발(intelligence explosion) 과정을 거쳐 초인공지능(Artificial Super Intelligence)으로 발전할 것이라고 예측하기도 한다(최윤섭, 2018).

2. 의료 분야 인공지능 윤리 이슈 및 쟁점

가. 일반적 윤리이슈

1) 인간의 몸에 대한 양화.수치화.등급화와 차별 가능성

① 후향적 연구의 허점

인공지능이 의료적으로 활용되는 과정에서 가장 광범위하게 나타나는 현상은 인간의 몸에 대한 양화 및 수치화일 것이다. 현재 인공지능을 활용한 의료적 상황은 데이터(data)에 기반한 후향적 연구(retrospective study)로 진행되고 있다. 즉, 축적된 데이터로 진료 시점을 기준하여 환자의 몸 상태를 규정한다. 진료 시점을 기준으로 산출되는 몸에 대한 수치화는 특정 신체에 대한 정확하고 정밀한 진단이라는 측면과 동시에 변화무쌍한 몸의 가소 가능성이라는 예외를 배제하는 결과를 낳는다. 개인의 의지, 노력, 의학적으로 규명하기 어려운 드라마틱한 신체 상황의 반전을 용인하지 않는 허점이 있다.

② 몸의 수치 환원적 규정 강제

몸에 대한 수치화 작업은 정신의학 분야나 특발성(비정형성) 질환과 같이, 데이터에 기반하여 규명하기 어려운 질환에 대해 과도한 수치 환원적 규정을 강제할 가능성이 높다. 또, 이 분야에서 중요시되는 임상경험, 직관력 등과 같은 경험칙의 판단 가치를 저평가하고 의료 인력의 탈숙련화를 부추길 수 있다. 이처럼 인공지능에 의존하는 진료 경험이 과도해지면, 환자들이 원로의사나 권위 있는 의사보다 인공지능의 판단을 더 신뢰할 수 있다. 이러한 점에서 인공지능의 의료적 활용 과정에서 의료인의 전인적 역할을 어떠한 수준에서 남겨놓을 것인가에 대한 논쟁이 예상된다.

③ 몸의 등급화와 차별

데이터에 기반 한 몸의 수치화는 무엇보다 몸의 품질에 대한 등급화, 나아가 사회적 차별로 이어질 수 있는 근거로 악용될 소지가 있다. ‘건강한 몸-불건강한 몸-병든 몸’ 등으로 몸의 질서가 위계화 되면, 개인들에게 부여된 신체등급의 부정적 라벨(label)은 학업, 취업, 고용 과정에서 부당한 차별 근거로 활용되거나 보험료 할증과 같은 형태로 경제적 부담을 더 가중시킬 가능성이 높다.[209)]

---

209) 실제로 미국에서는 1980년대 후반부터 비만자로 분류된 사람들이 정상체중인 사람들에 비해 소득이 적다는 보고가 제시되어 왔는데, 그 원인을 탐색한 결과 ‘비만자들이 기업에게 상대적으로 높은 건강보험 할증을 부담시키며 장기 결근 등으로 낮은 생산성을 보인다’며 고용주들이 비만자들을 고

2) 기술 수혜의 격차(건강형평성) 증가 논란

인공지능의 의료적 활용은 높은 정확성・정밀도에 힘입어 건강수명을 연장하려는 사람들로부터 크게 선호될 전망이다. 그러나 이러한 의료－기술적 수혜가 보편화되기는 어렵다. 정확한 진단의 성취로 불필요한 검사와 과잉 진료에 따른 환자의 경제적 부담이 감소하는 장점도 있지만, 일반적으로 고가의 의료비를 지불할 능력이 있는 사람과 그렇지 못한 사람 간에는 뚜렷한 수혜의 격차가 발생할 것이다(박혜경, 2017).

3) 유사・불법・자가 의료행위의 확산 우려

의료 수혜의 격차는 이른바 무허가, 비인가, 유사 의료기기의 활용이나 가짜 생체 데이터의 생산, 불법 생체 데이터 가공 등으로 이어질 수 있다. 그 결과 유사 의료행위, 불법 의료행위, 자가 의료행위 등이 광범위하게 확산될 가능성이 증가할 수 있다. 여기에, 전 세계적으로 의료 전문가와 병・의원에 전적으로 의존하는 전통적 방식의 진단-치료 시스템(생의학적 모형)을 비판하는 흐름이 이러한 현상을 자극할 수 있다(박혜경, 2017).

4) 기술적 실직(technological unemployment) 갈등, 탈숙련화 출현

인공지능은 거대한 생체, 유전적 데이터를 수치적, 시각적으로 분석하고 특징을 잡아 정밀한 진단 및 해결 방안을 제시하는 메커니즘을 가지고 있다. 의료 분야 중 CT, MRI 판독에서의 인공지능 도입은 영상판독, 마취과 등 특정 분야 의료 일자리 상실 및 감소, 인력 감축 등으로 이어질 수 있으며 일부 진료과목에서는 의사의 역할 축소에 따른 대량 감원이나 인공지능에 의한 대체, 탈숙련화 사태도 나타날 수 있다.

현재 인공지능의 딥 러닝 컨볼루션 신경망은 영상 인지・판독의 정확도가 매우 뛰어나기 때문에 영상의학과, 피부과, 안과의 안저 사진 판독, 병리과, 치과, 소아청소년과의 엑스레이 골연령 판독, 소화기내과의 (캡슐)내시경 분야 등에 광범위한 공급이 예상되고 있다. 실제로 2018년 5월 국내의 뷰노메이드 본에이지가 딥러닝 기반의 인공지능으로는 최초로 식약처의 의료기기 허가를 받은 바 있다(최윤섭, 2018). 또, 의료 인공지능의 높은 정밀성은 의료비 경감에도 기여하고 있다. 존슨앤존슨의 일명 자동 마취로봇인 수면마취 유도장치(세데시스)의 도입은 수면내시경 의료비를 90% 경감(최윤섭, 2018)해 주고 있어 의료계의 도입 전망이 매우 긍정적이다. 따라서 관련 분야 전문 의료 인력의 역할 및 기능 변화가 불가피할 것으로

용할 때 꺼리는 경향이 있었기 때문이었다. 서구 여러 나라에서도 건강과 소득이 비례 관계에 있다는 많은 연구 보고들이 있다(박혜경, 2017).

예상된다.[210)]

나. 신기술의 생산단계 윤리이슈와 대응방안

1) 의료 '로봇－사피엔스'의 출현과 법적 인격 부여 논란

인공지능은 다양한 장점과 유용성이 있으며 향후 그 능력이 계속 진화할 것으로 예상된다. 따라서 인간의사의 능력을 초월하는 로봇의사(AI의사)의 출현 가능성과 이에 따른 2차적 파급 현상으로 인간의사와 AI의사를 둘러싼 의료 주도권 다툼이 일어날 수 있다. 나아가 AI의사가 인간의사를 대체할 가능성도 배제하기 어렵다.

이러한 예상의 배경에는 의료 인공지능의 급속한 발전과 활용이 자리 잡고 있다. 미국에서는 2017년에 안과의 질병 진단을 인공지능에 맡겨 정확도를 이미 검증 받았으며(최인준, 2018), 싱가포르의 경우도 인공지능을 활용한 안저 진단이 유효한 방법임을 확인받았다고 발표했다. 일본의 경우는 안저 사진을 분석한 인공지능의 진단 시스템을 의료현장에 도입해 의사의 검진 부담을 줄여나간다는 방침이다. 또 안과 의사가 없는 자치단체에 이러한 시스템을 도입하여 어디서나 안과 검진을 받을 수 있도록 하여 질병을 조기 발견하는 것도 목표로 삼고 있다. 나아가 의사나 설비가 부족한 해외 지역에 이 시스템을 보내는 것도 검토하고 있다. 현재 미얀마의 의사 및 의료기관과 제휴해 그러한 준비가 진행되고 있다(뉴턴하이라이트121, 2018). 이미 "인공지능이 80%의 의사를 대체한다"라는 주장이 제기되기도 하였다(최윤섭, 2018).

앞으로 의료 인공지능의 능력이 더욱더 향상되고, 인공지능을 더 신뢰하는 사례가 반복적으로 나타나게 된다면 인공지능이 독립적인 진단과 치료, 처방의 권한을 갖게 된다고 볼 수 있다. 이렇게 될 경우, 의료 인공지능에게 법적 인격과 자격(의사면허)을 부여할 것인가 말 것인가의 문제가 대두될 수 있다. 의사의 권한 위임을 허용하지 않는다하더라도, 의료 인공지능에게 어떠한 수준의 자격을 부여할 것인가의 문제는 계속 제기될 수 있다.

---

210) 미국 암센터 토론회에서 메릴랜드 대학교의 엘리엇 시겔박사는 "영상의학과 전문의를 (완전히) 대체하기 위해서는 수천 가지 영상의학과 문제에 대해서 모두 개별적인 인공지능이 개발되어야 한다. 이를 위해서는 그 수천가지 모든 문제에 대한 개별적인 학습 데이터, 판독 결과, 그리고 개발비가 필요하다. 이런 인공지능이 모두 개발되는 데 얼마나 걸릴까?"라며 상당한 시간이 소요될 것임을 언급한 바 있다(최윤섭, 2018).

2) 의료 인공지능의 의료행위・의료기기 여부 논란

의료 인공지능이 개인들의 생체정보와 유전정보를 직접적으로 대량 수집하는 활동은 곧바로 의료 인공지능의 시스템 운용이 의료행위인지 아닌지의 논란으로 이어지고 있다. 의료 인공지능을 의료기기로 분류해야 하며, 그 활동도 의료행위로 보아야 한다고 주장하는 근거로는 의료 인공지능이 수집・제공하는 빅 데이터가 정보제공자인 개인의 건강에 큰 영향을 줄 수 있으며 따라서 체계적인 의료적 관리의 필요성이 있다는 점이 제기되고 있다. 특히 의료기관 내부에서 지속적인 의학적, 임상적 목적과 용도로 활용되는 의료 인공지능의 활동은 '의료행위'로 구분하여 규정할 필요성이 있다는 주장이다.

의료 인공지능의 독립적인 진단－처방 기능의 등장도 이 논란에 불씨를 제공하고 있다. 현재 의료 인공지능(왓슨 포 온콜로지의 경우)은, 식약처와 FDA의 규정에 따라 비(非)의료기기로 분류되어 있다. 분류 기준은 의료 인공지능이 의사의 독립적인 판단(진단－치료)을 보장하는가 여부이다. 그런데 2018년 4월 인간의사의 관여 없이 인공지능이 '진단'까지 내리도록 인허가 받은 사례가 최초로 등장하였다. 미국 FDA가 승인한 안저 사진 판독을 통한 당뇨성 망막병증을 진단하는 IDx-DR가 바로 그 사례이다(최윤섭, 2018). 앞으로 이러한 사례들이 더 많이 등장할 것으로 예상되기 때문에 의료 인공지능의 행위를 의료 보조행위로 제한하기는 어려울 것 같다.

현재도 중환자실이나 응급실에서 환자의 활력 징후에 따라서 인공지능이 환자의 산소 공급, 약물 주입 등을 결정하는 시스템이 있다면 의료진의 판단과 독립되어 있다고 볼 수도 있다, 특히, 병원 밖의 일상생활에서 일반 사용자의 웨어러블에 내장된 인공지능이 부정맥을 진단하고, 저혈당 쇼크를 예측하는 경우도 의사의 독립적인 판단을 벗어난 경우에 해당한다(최윤섭, 2018). 후자와 같은 경우는, 의료기기가 아님에도 불구하고 광범위하게 개인들의 생체정보, 유전정보 등을 '대량으로', 그리고 '직접적으로' 수집할 뿐만 아니라 수집한 데이터를 의료진에게 진료 및 치료 근거로 제시한다는 점에서, 오히려 의료 인공지능이 독립적인 진단 과정에 더 많이 관여한다고 볼 수 있다. 실제로 인공지능 소프트웨어 및 인공지능기술이 적용된 기기가 의료기기에 해당하는지 여부, 의료기기와 건강관리제품 간 구분의 모호(식약처, 〈의료기기와 개인용 건강관리(웰니스)제품 판단기준(지침) 마련〉. 2015. 7월), 인공지능이 적용된 의료기기의 신고대상 및 허가대상 구분 모호, 인공지능이 적용된 의료기기의 등급 및 품목 모호(2016. 12. 식약처. 〈빅데이터 및 인공지능(AI) 기술이 적용된 의료기기의

허가.심사 가이드라인(안)〉 마련 및 의견수렴 절차 진행 중) 등도 논란을 가중시키고 있다(윤혜선, 2017).

의료행위에 대해 현행 실정법상에 명시적인 정의 규정이 부재한 상태라는 점도 논란의 근거가 될 수 있다. 현행법상으로는, 국가가 법률로써 소수에게 면허를 주어 의료행위를 허용하고 있으며 따라서 무면허 의료행위에 대해서는 일률적・전면적으로 금지하고, 위반 시 그 치료결과와 관계없이 형사적 처벌을 하도록 규정하고 있다. 그러나 2001년, 2007년의 대법원 판례는 의료행위에 대해 포괄적으로 확대해석이 가능하도록 판결을 내린 바, “의학적 전문지식을 가진 의료인이 진료. 검안. 처방, 투약과 외과적 시술을 시행하여 질병을 예방하거나 치료하는 것, 의료인이 시행하지 않으면 사람의 생명, 신체나 공중위생에 위해를 발생시킬 우려가 있는 행위”를 의료행위로 보았다(윤혜선, 2017). 이는 의료 인공지능의 활동에 대해 의료행위 개념 혹은 범위에 대한 논란이 이어질 충분한 근거가 될 수 있다.

3) 의료 인공지능의 정확성・활용효용성・사용편의성 논란

① 의료 인공지능의 정확성・정밀성 논란

약한 인공지능이 적용된 의료기술의 경우 기계학습에 따른 한계가 분명하다. 기계학습을 하는 약한 인공지능의 경우 한 가지 문제만 풀도록 고안되어 있으며, 자의식이 없고, 스스로 무엇을 할지 판단하지 못한다. 또한 기계학습의 경우 문제를 푸는 것을 배우기 위해서는 반드시 학습데이터가 필요하다. 즉, 기존에 학습한 데이터가 없고 문제가 정의되어 있지 않은 완전히 새로운 영역에 대해서는 인공지능이 근본적으로 해결할 수 없다. 인공지능은 임상적 소견이 축적된 의료 환경에서는 뛰어난 능력을 발휘할 수 있지만 아직 원리나 치료법이 밝혀지지 않은 질병, 정량적으로 나타내기 어렵거나, 환자의 수가 많지 않고, 질병의 유형이 지나치게 세분화되어 있어서 충분히 데이터가 쌓이기 어려운 질병 등에서는 인공지능의 접근이 특히 어렵다(최윤섭, 2018; 김진구, 2018).

실제로 국내 의료현장에서도 현재 사용 중인 의료 인공지능의 데이터 수집 및 제공 수준이 임상적 활용가치가 높을 정도로 의미 있게 축적되어 있지 않은 상태이기 때문에 의료기기로 분류하여 활용하는 것은 적합하지 않다는 의견을 제시하고 있다. 의료계와 보건복지부 등이 왓슨을 의료기기로 분류하지 않는 이유로, 데이터의 업데이트가 너무 잦다는 데 있으며 의료는 축적 데이터가 더 중요하고, 따라서 현재로서는 왓슨을 의료기기로 관리하기가 어렵고 따라서 당분간은 범용화에 이르기는 쉽지 않은 상태로 보인다는 주장이다(장

동경, 2017).[211]

② 의료 인공지능의 활용효용성 논란

동일한 인공지능 의료기술이라 하더라도 누가, 어떤 환자를 대상으로, 어떻게 활용할지에 따라서도 이런 의학적 효용은 달라지며, 의사나 환자, 병원에 미치는 영향도 달라질 수 있다(최윤섭, 2018). 현재의 의료 인공지능은 설명력과 해석력을 확인하기가 어렵다. 즉, 인공지능이 어떠한 결과를 내놓았을 때, 왜 그런 결론을 내리게 되었는지 그 과정과 이유를 투명하게 설명할 수 있는 (기술적 알고리즘이) 마련되어 있지 않기 때문에 기본적으로 블랙박스와 다름이 없다.[212] 이러한 인공지능의 특징은 딥 러닝 능력에서 비롯되는 것인데, 문제는 딥 러닝의 계산 과정이 일종의 블랙박스와 같아서 투명하지 않기 때문에, 그 인공지능을 만든 사람도 세부적인 작동 원리를 파악하기가 쉽지 않다는 데 있다.[213] 결국 이러한 인공지능의 특징 때문에 의료 현장에서 인공지능을 활용할 때에 의료진이 의료 인공지능의 효용성을 얼마나 신뢰하고 수용하게 될 것인지 기대하기 어렵다. 특히, 오랜 임상 경험에 기반 하여 자신의 의학적 판단을 절대적으로 신뢰하는 숙련의가 의료 인공지능의 효용성을 조건 없이 신뢰하고 활용하기란 쉽지 않을 것이다.

한편, 현재 사용하고 있는 의료의 많은 부분(약물기전, 의료기기 예, 아스피린)이 이미 블랙박스이므로, 인공지능이 블랙박스라는 점이 별 문제가 되지 않는다는 의견도 적지 않다. 그 이유는 의료 인공지능이 블랙박스를 모두 해독하지 않더라도, 임상시험을 통한 정확성과

---

211) 우리나라 식품의약품안전처 산하 식품의약품안전평가원 의료기기심사부의 〈빅데이터 및 인공지능 기술이 적용된 의료기기의 허가・심사 가이드라인(안)〉(2016. 12)에 따르면, 환자의 질병 진단・예방 등의 목적으로 의료정보를 분석, 진단, 또는 예측하기 위해 제조된 소프트웨어는 의료기기에 해당될 수 있지만, 의료 문헌을 분석하여 해당 문헌에서의 환자의 질병 진단법 또는 치료법을 검색, 요약하거나, 환자의 처방전 및 의약품 목록을 제시하는 소프트웨어(의료정보검색)는 의료기기에 해당하지 않는다.

212) 왓슨 포 온콜로지는 판단 과정이 상대적으로 투명하며, 결과에 대한 설명력도 가진다, 내부적으로 어떠한 과정을 거치는지 알기 매우 어려운 딥러닝과는 달리, 왓슨 포 온 콜로지는 판단 결과에 대한 '근거'를 제시하기 때문이다. 각 권고안이 달린 근거 버튼을 누르면, 해당 권고안에 대한 진료 가이드라인, 논문. 임상시험 결과 등을 보여주기 때문에 판단 과정에 대한 힌트라도 의료진이 얻을 수 있다. 만약 왓슨이 제시한 근거가 적절하지 않다고 생각하면, 의료진은 왓슨의 권고를 받아들이지 않을 수 있다(최윤섭, 2018).

213) 예를 들어, 알파고의 경우 몇천년 동안 인간이 쌓아온 바둑에 대한 실력을 단기간에 능가했지만, 특정한 한 수가 구체적으로 어떠한 계산 과정을 거쳐서 도출된 것인지를 완전히 파악하기는 매우 어렵다.

안전성의 검증을 통해서 인허가 여부가 결정될 것이다.

③ 의료 인공지능의 사용편의성 논란

인공지능 의료기술의 진료현장 도입 후 효과적인 활용 가능성, 즉 과연 정확도, 정밀성이 좋다고 해도 의료 현장에서 사용할 때 사용편의성이 있을지에 대한 논란을 예상할 수 있다. 의료 인공지능의 활용에 있어서 의료진의 진료습관, 기존 업무 방식에 대한 강한 의존, 실제 활용 과정에서 비용부담과 의미 있는 결과 도출 간의 뚜렷한 상관성 확보가 어려운 상태이다. 이렇게 볼 때 아직은 의사의 판단에 더 의존할 확률이 높다. 의사가 단독으로 판단하기 어려울 경우, 복수의 의료진이 협의를 하는데, 그렇게 하여도 판단이 어려운 경우에는 인공지능을 활용할 것으로 예상된다.

그러나 의료 인공지능이 의료진에게 제시하는 경고시스템을 의사들이 무시할 수도 있다. 그 이유는 너무 사소한 경고가 반복됨으로써 경고피로현상이 있을 수 있고, 의사들이 시스템 사용이 불편하여 경고를 무시하거나, 자신의 판단을 더 신뢰함으로써 이러한 현상이 발생할 수 있기 때문이다(최윤섭, 2018). 이러한 현상은 의사들의 진료'문화'와 깊은 관련이 있는 부분이며, 따라서 인공지능 의료기술의 현장 사용 편의성에 대한 논란은 쉽게 개선되기 어려울 가능성도 높다.

4) 의료 인공지능의 오작동 우려

의료 인공지능의 가장 큰 기술적 위험요인은 오작동이다. 인공지능의 오작동 원인은 process의 오류로 인한 기기 오작동, 기술적 결함, 하드웨어의 불량 등을 우선 꼽을 수 있다. 또 사용자의 과실 및 고의에 의한 오작동, 잘못된 정보제공 등에 의해서도 발생할 수 있다(박혜경, 2017). 전자의 오작동은 기술의 인허가 과정과 개발자 및 제조업자의 품질 관리 및 사후 A/S 등을 통해 예방이 가능할 것으로 예상된다. 그러나 후자의 요인들에 의한 오작동은 예방이 쉽지 않을 수 있다. 사용자가 범죄적 목적을 가지고 고의로 프로그램을 교란하거나, 환자가 신체 정보를 왜곡 또는 가공하여 제공할 경우 인공지능이 산출하는 잘못된 결과물로 인해 의료진의 판독 오류와 오진(誤診)이 발생할 수 있다.

그러나 이보다 더 심각한 상황은 인공지능 플랫폼 통합과정에서 발생하는 오작동, 인공지능 시스템의 복잡도가 커짐에 따라 인공지능의 오작동 유발 요인을 사용자가 인식하거나 파악할 수 없는 경우이다(이예하, 2017; 장동경, 2017). 이른바, 인공지능의 블랙박스로서의 특성이 이러한 문제를 일으키는 핵심 원인이다.

다. 신기술의 소비단계 윤리이슈와 대응방안

1) 의료 인공지능의 신뢰도 제고에 따른 의료분쟁 시 책임 논란

의료 인공지능의 능력과 유용성의 향상은 환자가 의사보다 인공지능을 더 선호하는 현상으로 이어질 수 있다. 현재 의료 인공지능은 비(非)의료기기로 분류되어 있기 때문에 인간의사의 독립적인 진단 판정과 권한이 보장되어 있다(김진구, 2017). 그리고 인공지능이 제안한 검진 결과와 치료법을 환자에게 적용하기 위해서는 환자 본인이나 보호자의 동의를 반드시 받아야한다. 그런데 이 과정에서 만약 환자가 의사보다 인공지능의 판단을 더 신뢰한다면 의사의 진료 판정 과정에 새로운 논쟁거리가 등장할 수밖에 없다. 실제로 2017년 1월 한 국내 일간지가 보도한 바에 따르면 길병원에서 2016년 11월부터 진료한 100여 명의 환자 중에서 의사와 왓슨의 판단이 다를 경우에 환자들은 모두 의사보다 왓슨을 더 신뢰했다(이언, 2018). 이러한 경우 의료사고나 과실에 의한 분쟁 발생 시 의사, 인공지능(제조사), 환자 간의 책임 소재가 논란이 될 수 있다.

2) 데이터(의료정보) 확보와 관리의 투명성・합리성 문제

인공지능을 통해 수집되는 데이터는 개인의 생체정보, 유전정보 등 고유하고 민감한 의료정보에 해당한다. 더욱이 의료 인공지능이 처리하는 정보는 그 규모 또한 어마어마한 빅 데이터라는 특징이 있다.[214] 그런데 이러한 대규모의 민감성 의료정보에 대한 수집, 처리, 관리 과정에 대한 투명하고 합리적인 체계가 확보되어 있지 못한 상태여서, 다양한 문제가 출현할 것으로 예상된다.

먼저, 데이터를 확보하는 과정의 문제를 들 수 있다. 데이터 제공자에 대한 보상, 의료계 입장에서 볼 때 데이터 취득 비용이 모호하다는 문제가 있다. 이는 데이터의 소유관계에 대한 모호성에서 비롯된다(윤혜선, 2017). 방대한 데이터를 수집하는 과정에서 데이터를 어떠한 방법으로 수집・확보해야 의미 있는 데이터를 생산할 수 있을까? 바로 질 좋은 데이터를 어떻게 모을 것인가의 문제가 대두한다(김형회, 2017).

다음으로, 수집・확보한 데이터를 어떻게 보관・관리할 것인가의 문제도 나타날 수 있다. 인공지능의 등장으로 데이터의 수집량은 어마어마하게 늘어날 것이며 시간이 흐를수록 더 큰 용량이 축적될 것이다. 빅데이터를 보관・관리하는 과정에서 데이터의 손실, 훼손, 변형

214) IBM에 따르면, 사람 1명이 일생에 걸쳐 생산하는 data는 1,100Tb에 이른다고 한다(이예하, 2018: 9).

등이 발생하지 않도록 안전하게 보관・관리하는 장치가 요구된다. 수집한 빅데이터에서 의미 있는 정보를 신속하고 정확하게 추출하여 사용하기 위해 미리부터 체계적으로 분류하고 보관하는 기법도 마련할 필요가 있다.

3) 데이터(의료정보)의 상업적・범죄적 유출 및 악용, 오・남용 피해

의료정보는 정치적 이해관계 달성, 금전적 이익 등을 목적으로 한 의료집단 및 관계자, 그리고 이들과 연결된 외부인, 독자적 외부침입자들을 통해 의도적 유출・거래, 부정한 열람, 복제 등의 위험에 처할 수 있다(윤혜선, 2017; 박혜경, 2017). 의료정보에는 생체정보, (희귀) 유전정보, 질병이력, 복용하는 약물, 이용하는 병의원, 식사 및 운동 습관 등 개인들의 건강과 관련한 제반 사생활 정보가 대량으로 담겨있다는 점에서 유출 시에는 개인의 신체에 대한 낙인과 차별, 이로 인한 취업・고용 차별, 보험 할증 및 보험 가입 거절 피해부터 범죄적 악용(취약정보를 악의적으로 편집・가공・유포하는 행위 등)으로 이어지는 2차 피해로 확대될 수 있다.[215)]

고인(故人)의 생전 데이터가 유통되거나, 개인의 건강상태가 과거의 부정적인 데이터로 남아 있게 될 경우 낡은 데이터의 유통으로 인해 사자와 생존자 및 그 가계에 불명예와 실질적 피해가 초래될 수 있다. 개인에 대한 생체, 유전 정보 등 민감 정보의 대량 수집, 빅데이터의 축적 및 확대재생산, 가짜 정보의 생산 및 유통 등이 끊임없이 반복될 수 있기 때문이다.

의료 인공지능의 판단 결과에 대한 의사와 환자의 과도한 신뢰 및 의존과 함께 기기제조업계, 기기유통업계, 병원 경영자, 제약회사, 메디컬언론 등이 의료 인공지능의 판단 결과에 대해 과도한 신뢰를 조장함으로써 의료데이터의 오・남용이 발생할 수 있다는 점도 우려되는 문제이다. 특히, 전문 의료진의 해석과 판단 범위를 벗어난 데이터의 오・남용은 개인의 왜곡된 생체정보를 양산함으로써 위해를 초래할 수 있다는 점에서 엄중히 관리되어야 할 부분이다.

데이터의 보안 및 투명한 관리는 국가적 차원에서도 주의를 기울여야 한다. 인공지능의 의료적 활용 과정에서 나타날 수 있는 위험성은 사회적 재난으로 이어질 수 있다. 최근 세계적 역병의 규명 과정에서 특정한 지역이나 국가들이 질병 발생지로 지목되는 오해를 산 바가 있었다. 빅데이터는 단순한 숫자가 아니라 명징한 결과를 생산할 뿐 만 아니라, 그 생성 과정이 블랙박스화되어 있기 때문에 지속적・체계적으로 철저히 관리해야만 한다. 국제사

215) 현재 국내 의료계에서는 희귀유전질환 데이터를 수집중이다. 이러한 민감성 정보는 매우 잘 관리하지 않으면 안 된다(김형회, 2017: 90).

회에서 의료적 비난의 화살을 맞게 될 경우 세계적 오명을 입음과 동시에 국가적 위신의 추락을 겪을 수 있다.

라. 신기술 재생산 과정의 윤리이슈와 대응방안

1) 의료정보의 자산화와 공익적 활용 과정의 문제

의료 인공지능의 활용이 보편화되기에 이르면 대량의 의료정보가 축적되고 의미 있고 활용가치가 높은 정보의 재생산이 가능해질 것이다. 이렇게 되면 중요한 의료 데이터를 공유하고 이전하는 문제가 발생할 것이다.

의료계에 인공지능의 활용이 보편화되면, 의료 데이터의 활용가치는 갈수록 높아질 것이다. 때문에 그러한 상황에서 유용한 의료데이터를 원활하게 활용하는 문제와 프라이버시 문제를 기술적, 법적으로 어떻게 지킬 수 있는가는 상호 충돌이 불가피(이중원, 2017)한 측면이 있어 해결하기가 매우 어려운 과제이다. 특히, 활용가치가 큰 국가적 수준의 데이터는 지적재산의 의미와 가치를 갖게 될 수 있다. 이렇게 될 경우 프라이버시권의 지나친 보호는 특수한 의료정보의 공익적 활용을 제한한다는 점에서 문제가 발생할 수 있다. 또, 개인 생체, 유전정보를 기반으로 축적된 의료데이터는 공공 자산화 과정에서 소유권 다툼으로 이어질 수 있다(윤혜선, 2017).

2) 의료 인공지능의 사회화 알고리즘 구성의 필요

장기적으로 의료 인공지능으로 하여금 상식, 윤리관, 도덕관념, 감정, 권리를 갖게 해야 할 것인가? 이미 미국에서는 자폐증 환자는 도울 목적으로 한 감정인지로봇을 수년 전부터 개발해왔다. 이렇게 될 경우 의료 인공지능이 싱귤레리티(singularity: 기술적 특이점)를 지나 자율성을 가진 로봇으로 진화할 가능성은 더욱 높아진다(뉴턴하이라이트121, 2018).

그러나 상식, 감정은 무엇인가에 대한 개념 규정, 또 시대나 문화권에 따라 변할 수 있다는 점에서 어떤 감정을 얼마나 부여할 것인가는 필요적, 기술적으로도 난제이다. 하지만 전 세계적으로 의료의 패러다임은 환자중심, 의사－환자간 커뮤니케이션의 확대 등을 강조하고 있다. ‘의사의 공감 능력이 치료 결과를 좌우한다’는 많은 연구 결과도 있다(최윤섭, 2018). 그러나 현실은 그렇지 않다. 커뮤니케이션 훈련을 받은 종양내과의사는 전체의 5%에 지나지 않는다는 결과도 있다(최윤섭, 2018). 결국 현재의 프로그램 설계 시스템을 어떠한 수준에서 선택할 것인가의 문제가 남게 된다.

〈표 4－7〉 각 단계 행위자별 윤리이슈

단계	행위주체	윤리 이슈
개발단계	개발자	• 알고리즘 개발 단계에서 국가별/지역별 윤리의 특수성 고려
	공적 주체	• AI 개발을 위한 촉진 정책과 규제의 적절한 균형
	의사	• AI 이용과 관련한 윤리 교육(의과대학에서의 교육 포함)
인허가단계	개발자 및 제조사	• 허가를 위한 윤리적 자가검증 도구 개발 및 적용
	공적 주체	• 허가 과정에서의 윤리적 검증 도구 개발
	소비자 및 시민단체	• 인허가 과정에서의 윤리문제에 대한 모니터링 및 의견개진 절차 요청
상용단계	의사	• 수가 책정의 문제 • 의료사고 시 책임 면제 범위의 문제
	병원 경영자	• 병원의 수익 창출에 최적화된 알고리즘 적용의 문제
	공적 주체	• 의료사고 시의 손해배상 체계와 사전 행정적 규제의 검토 • 윤리적 내용의 규범화 범위 • 의료 기술의 수혜 격차 문제
	소비자	• 데이터 활용과 프라이버시 보호의 문제 • 소비자의 선택권 문제 • 자가진료 내지 무허가 의료행위의 문제

제 5 절 국방 분야: 치명적 자율무기체계를 중심으로

1. 논의의 배경

인공지능(AI)의 국방 무기체계에 대한 문제인 치명적 자율무기(LAWs)는 이미 전 세계적으로 공개된 비밀이다. 현재도 적지 않은 전장에서 LAWs가 배치되어 활동하고 있고 그 추세는 점점 더 확대되어 갈 것으로 사료된다. 그런데 한편으로 치명적 자율무기 체계는 미래 국가안보의 핵심으로 각국이 최고도의 보안을 유지하면서 진전 상황이나 적용영역을 초비밀로 베일의 장막 뒤에서 진행하고 있는 과학기술의 무한 경쟁이다(한희원, 2018a; 한희원, 2018b; Daniel S. Hoadley & Nathan J. Lucas, 2018).[216)]

216) 남호주대학 산하 방위시스템연구소 소장인 안토니 핀 교수는 “자율 살상무기는 ‘파이어 앤 포겟(Fire & Forget)’ 무기로서 재래식 무기보다 민간인 피해를 절대적으로 줄이고 전쟁을 조기에 끝낼 수 있을 것”이라고 주장한다. 꼭 필요한 명령만 내려 로봇무기가 무고한 살상을 방지할 수 있는 프로그램을 개발하면 예컨대 로봇은 여성들을 성폭행하지 않는 것처럼 불필요한 피해를 최소화 할 수 있는 장점도 있다고 주장한다. 또한, Thomas Simpson과 Vincent Muller는 오히려 누가 명령을 내렸는지 기록하기가 더 쉬워 책임자도 명확하게 가려낼 수 있다고 주장한다(Simpson, Thomas W;

물론 제4차 산업의 다른 많은 기술도 국부(國富)를 형성할 것으로 고도의 보안이 필요하지만 인류의 생명권・자유권・재산권이라는 천부인권에 직접적인 타격을 가할 수 있는 LAWs의 전쟁 무기적 특성으로 인해서 그 어떤 과학기술보다 국가안보 차원의 법적・윤리 쟁점을 야기할 것으로 사료된다.

일례로 2018년 4월 5일 세계 인공지능(AI) 및 로봇 연구 분야 학자 50여 명은 한국과학기술원(KAIST)이 한화시스템과 공동으로 추진하는 인공지능 무기 연구에 항의해 카이스트와의 관계를 끊겠다고 선언했다. 토비 월시 오스트레일리아 뉴사우스웨일스대 교수 등 세계 30개 국가 인공지능 및 로봇 공학 연구자 50여 명은 “카이스트처럼 명망이 있는 기관이 인공지능 무기개발을 통해 군비경쟁을 가속하는 것처럼 보이는 것은 유감이라면서 카이스트의 어떤 부분과도 협력을 보이콧하겠다”는 내용의 공개편지를 발표했다.

그들은 앞으로 카이스트를 방문하지 않고, 카이스트의 방문자를 받지도 않을 것이며, 카이스트와 관련한 연구에 참여하지 않겠다고 압박했다. 이에 대해 카이스트는 곧바로 “국방인공지능융합연구센터는 대량살상무기, 공격무기 등 인간 윤리에 위배되는 연구와 통제력이 결여된 자율무기를 포함한 인간 존엄성에 어긋나는 연구 활동을 수행하지 않을 것”이라는 내용의 성명을 발표하고 항의 서명한 학자들에게 신성철 카이스트 총장 명의의 서신을 보냈다.

하지만 항의자들의 보이콧 선언은 철회되지 않았다. 적극적으로 해명에 나선 카이스트는 센터의 설립 목적을 방위산업 물류 시스템, 무인 항법, 지능형 항공훈련 시스템 등에 대한 알고리즘 개발로 “살상용 무기나 공격용 무기개발이 아니다”라고 밝혔다. 카이스트 홍보실은 “한화시스템은 토털 아이티 솔루션 소프트웨어 업체로, 연구센터의 주요 목적은 공군 조종사들이 모의훈련 할 때 사용할 인공지능 기반 프로그램을 개발하는 것이다. 우리나라 무기개발 시스템에서는 대학이 무기개발 과정에 참여할 수 없다”고도 해명했다.[217] 대한민국은 이미 인공지능(AI) 기반 무기연구 개발 착수단계에서부터 난항을 겪었던 것이다.

Muller, Vincent C., 2016).

217) 카이스트는 2018년 2월 20일 한화시스템과 ‘국방인공지능융합연구센터’를 공동 설립하고 개소식을 열었었다(한겨레, 2018. 4. 5).

가. 자율무기체계의 현재와 치명적자율무기(LAWs)의 개념

1) 자율무기가 위험을 초래한 역사적 사례

미래 인공지능(AI)은 과연 전쟁에는 어떠한 영향을 줄까? 일각에서는 가공할 핵무기의 억지력(deterrence)처럼 역설적으로 치명적 자율무기가 무력충돌을 자제시켜 줄 것으로 긍정적으로 예측하며, 차라리 모든 전쟁을 영원히 종식시키기 위하여 모든 나라가 더욱 가공할 AI 기반 무기를 갖는 것을 허용하자고 제안한다(한희원, 2018b). 그러나 만약 자동무기 시스템에 오류가 나거나 혼동의 소지가 있거나 예상대로 작동하지 않는다면 과연 어떤 일이 벌어질까. 먼저 2개의 역사적 사례를 본다(맥스 테그마크, 2017).

① 1988년 미국 미사일 순양함 빈센스 함 오발 사례

미국의 이지스급 순양함에 장착된 근접방어 무기체계 팔랑스는 대함미사일과 전투기 같은 위협 요인을 스스로 발견하고 추적해 공격한다. 이에 로보크루저라는 별명을 갖는다. 1988년 7월 3일 이란-이라크 전쟁에서 이란과 교전 중에 빈센스 함의 레이더 자동 시스템이 항공기 접근을 경고했다. 당시 윌리엄 로저스 3세 함장은 빈센스 함이 급강하하는 이란의 F-14 전투기의 공격을 받게 되었다고 추론하고 이지스 시스템 발사를 승인했다. 그러나 격추된 것은 이란 항공 655편이었고 결국 민간인 290명 전원이 사망했다. 조사결과 빈센스 함의 이용자 자동 인터페이스의 자율적인 판단이 잘못된 것이었다. 자율적인 기계는 레이더에 나타나는 어느 점이 민간항공기인지, 어느 점이 공격을 위한 전투기인지 구분을 하지 못했던 것이다. 자율무기체계 판단 시스템은 숫자로 판단했기 때문이었다. 사례에서 물론 최종 결정권자는 로저스 3세 함장이었지만 그는 제한된 시간의 압박 속에서 자율무기 체계의 정보를 신뢰할 수밖에 없었던 것이었다.

② 제3차 세계대전의 위험성

쿠바 미사일 위기 때인 1962년 10월 27일, 미국 구축함 11척과 항공모함 랜돌프와 소련잠수함 B-59가 대치중이었다. 당시 미군이 알지 못한 사실은 소련잠수함의 배터리가 거의 방전되어 에어컨이 멈췄고 함정의 실내 온도가 섭씨 45도 넘게 올라가는 상황이었다. 다수의 승무원이 이산화탄소 중독으로 기절을 했다. 함내 사정을 모르는 미군은 소련잠수함을 수면 위로 부상시키기 위해 작은 기뢰를 투하하기 시작했다. 그런데 미군도 알지 못한 비밀이었는데 소련잠수함 B-59는 소련 본국의 재가가 필요 없이 가동할 수 있는 자동 핵무기 발사 시스템을 갖추고 있었다. 함장 발렌틴 그리고리에비치 사비츠키는 핵 어뢰 발사를 결심하고

“우리는 죽을 것이지만 미군도 다 침몰시킬 것이다. 우리는 우리 소련 해군의 명예를 실추시키지 않을 것!” 이라며 자동시스템 가동을 결심했다고 한다. 그런데 핵 어뢰를 발사하기 위해서는 함장 이외에 2명이 더 동의해야 했는데, 한 사람은 핵 발사에 동의한 소련 공산당 정치위원이었고 다른 한 사람은 부함장 바실리 아르키포프였다. 그런데 부함장 아르키포프는 핵 어뢰 투하는 3차 세계대전 유발이라고 반대하였고 결국 시스템은 가동하지 않았다. 이건은 치명적 자율무기에서 의미 있는 인간통제가 얼마나 중요한지는 알려주는 사례로 회자된다. 전문가들은 만약 소련의 B-59가 인공지능(AI)이 통제하는 완전 자율적인 무기체계 시스템이어서 사람이 의사결정에 개입하지 않게 되어있었다면, 세계는 제3차 세계대전을 맞았을 것으로 예상한다.

2) 치명절자율무기의 개념

① 개념

치명적 자율무기(Lethal Autonomous Weapons: LAWs), 일명 살인 로봇은 인간의 개입 없이 시스템 스스로가 목표물을 선택하고 공격하도록 설계된 자율무기를 의미한다. LAWs는 LAWS(lethal autonomous weapon systems, 치명적 자율무기 시스템), LAR(lethal autonomous robots, 치명적 자율로봇), 로봇 무기(robotic weapons), 단순하게 살상용 로봇(killer robots)이라고도 불린다(Krishnan, Armin, 2009; Anderson & Waxman, 2012).

그러나 사실 현재까지도 LAWs가 무엇인지 국제적으로 합의된 개념 자체가 없는 상태이다. UN 특별보고관 크리스토프 헤인즈(Christof Heyns)는 LAWs를 일단 활성화되면 더는 인간의 개입 없이 목표를 스스로 선택하고 완수하는 무기 시스템이라고 개념 정의한다(C. Heyns, 2014). 미국 국방부 지침(The US Department of Defense Directive)은 일단 활성화되면 더는 인간의 개입 없이 스스로의 자율적인 판단으로 목표를 선택하고 수행할 수 있는 무기체계라고 정의하고 있다(Department of Defense Directive, 2012). 국제적십자위원회(International Committee of the Red Cross: ICRC)는 ‘핵심기능(critical functions)’에서 ‘자율권(autonomy)’을 갖는 무기로 인간의 개입 없이 기계 스스로가 목표물을 검색, 식별, 추적하여 목표에 대해 제압, 무력 사용, 무력화, 살상과 파괴 등 각종 공격을 할 수 있는 무기라고 개념 정의한다(C. Heyns, 2014).

LAWs에 대한 개념정의와 관련하여 ‘의미 있는 인간통제(meaningful human control)’ 개념이 중요하다. 의미 있는 인간통제(MHC)란 공격할 대상의 선택과 효과에 대한 인간의 직접

적인 통제를 의미한다(한희원, 2018b). 인류의 신형병기인 LAWs는 어떤 방향으로든지 지구의 미래를 새롭게 형성할 것으로 보인다. 그런데 부상하는 LAWs는 냉전 시대처럼 경쟁하는 두 나라 사이의 경쟁이 아니라 천천히 변하는 법규범과 급속한 과학기술 발전 사이에 놓여 있는 법적・윤리적 쟁점이다.

② 유형[218)]

#1. 인공지능(AI) 기반 무기 개관

킬러로봇, 인공지능 레이더, 자율주행 탱크, 로봇 보초병, 인공지능 드론, 인공지능 폭격기, 인공지능 유도탄, 인공지능 기관총, 살상용 로봇

#2. 1991년 개봉 영화 '터미네이터 2'

○ 1997년 미국 정부는 인공지능(AI) '스카이넷'에 모든 군 지휘권과 무기통제권 일임. 스카이넷은 자신을 위협하는 인간을 모두 '적'으로 간주: 인류 말살에 나섬. 스카이넷은 러시아에 핵미사일을 쏘아 미・러 간 핵전쟁을 유발

#3. 2005년 개봉 영화 '스텔스'

○ 인공지능 무인 스텔스기인 '에디'의 인공지능 회로에 문제 발생: 통제 불능

○ 피아 식별 불가 & 민간인 공격 & 러시아를 적대국으로 인식하고 폭격 시도

#4. 2015년 1월 러시아 스푸트니크 보도, 시리아 배치 인공지능 자율무기

○ 무인로봇 기갑차량 '플랫폼－M'－자율무기 개발을 위한 전술 실험장 활용

○ 신형 AI 기반 무인 전투차량 우란－9

우란－9: 본격적인 공격용 전투차량(기관포와 대전차 로켓 장착)

이슬람국가(ISIS)를 공격－무인기계가 인간을 공격한 대표적 사례.

#5. 러시아 100% 인공지능 레이더망－기계가 100% 목표물 자율 선택

#6. 미국 신형 인공지능 자율추적 미사일

#7. 인공지능 자율 무인기－알카에다와 IS 소탕 작전에 이미 실전 투입

---

218) Scharre and Horowitz, An Introduction to Autonomy in Weapon Systems, p. 13; Noel Sharkey, "Killing Made Easy: From Joysticks to Politics," in Robot Ethics, p. 116; Schmitt and Thurnher, "Out of the Loop," p. 238; Johnson, Meyers, Richards, et al., Unmanned Effects (UFX): Taking the Human Out of the Loop, p. 5; 북핵보다 치명적인 인공지능 무기, 중앙일보, 2018. 2. 17, available at, https://news.joins.com/article/22374485.

MQ-9 리퍼(계량형 프레데터)

양날개 길이: 20m 작전 시간: 14시간, GPS 유도 합동정밀직격탄(JDAM), 레이저 유도탄(GBU-12), 헬파이어 미사일 등

MQ-9 리퍼(계량형 프레데터)

양날개 길이: 20m 작전 시간: 14시간 무장: GPS 유도 합동정밀직격탄(JDAM), 레이저 유도탄(GBU-12), 헬파이어 미사일 등

– 2015년 6월 '알카에다 아라비아반도지부(AQAP)'를 이끌어 온 나세르 알 우하이시 저격, 저비용 무인기 군집기술(LOCUST)'을 활용해 드론 떼를 상륙전의 선봉에 배치.

#8. 2017년 11월 미 해군의 인공지능 자율무기

– 무인 함정 '씨 헌터(Sea Hunter)' 배치 공식 선언

길이 40m, 최대 시속 50㎞, 최장 3개월 원거리 해상에서 적 잠수함 탐지

– 무인 잠수정 '에코 보이저(Echo Voyager)'

최대 1개월간 자율 운항하면서 적 잠수함 정보수집

#9. 미래 인공지능(AI) 로봇 군인

– 24시간, 365일 전투 수행 가능, 피로, 수면, 휴식시간 불필요

– 작전도 스스로 전개

③ 미국 펜타곤의 전략과 각국의 현재까지의 동향 개관

미국 국방부는 현재 국제사회의 논란을 의식하면서 국방용 로봇을 100% 인공지능 로봇이 아닌 '지능 확장(IA · Intelligence Augmentation)형 로봇' 개념으로 개발한다고 말한다. 기본적으로 미국은 '자율무기에서도 인간 통제력'을 절대 놓아선 안 된다면서 마지막 공격 스위치는 인간의 몫으로 남겨두어야 한다는 태도이다(Department of Defense Directive, 2012). 사실 미국은 사람을 살상하거나 인간의 개입 없이 공격 목표를 설정하고 자율적으로 공격하는 자체적으로 설정한 소위 적절한 수준(appropriate levels)과 기준을 충족하는 자율무기 시스템을 이미 인증했다(Department of Defense Directive, 2012; Gubrud, Mark, 2015).[219] 2015년 미국 해군은 미래에는 무인무기시스템이 '새로운 정상(new normal)'이 되어야 한다고 공표했고, 로버트 워크(Robert Work) 국방부 부장관은 2016년 "국방부는 치명적 권한(lethal authority)을

219) 더 나아가 자율적으로 목표물을 식별하고 공격하는 반 자율적 헌터 킬러는 인증조차 필요로 하지 않는다(한희원, 2018b).

기계에 위임하지 않을 것이지만 어떤 “독재 정권”이 그렇게 할 수 있기 때문에 이것을 다시 생각할 필요가 있다.”고 말했다(한희원, 2018b 재인용).

중국과 러시아는 미국 국방부가 2017년 회계연도 인공지능(AI) 관련 프로젝트 예산으로 120억～150억 달러를 의회에 요청하면서 로버트 워크 국방부 차관이 “검은 커튼 뒤에 무엇이 있을지 경쟁자들이 궁금해하기 바란다.”라고 하는 발언에 주목하고 있다고 한다. 한편 멕시코, 칠레, 쿠바, 이집트, 교황청, 파키스탄 등은 자율무기에 대한 선제적이고 포괄적인 금지를 요구했다. 반면에 일본, 스리랑카, 스위스, 프랑스, 아일랜드, 대한민국은 LAWs를 반대하지 않으면서 ‘의미 있는 인간통제(MHC)' 개념 토론에 참여할 용의를 밝혔다.[220)]

나. 치명적 자율무기체계와 국가안보

치명적 자율무기 시장이 결코 미국을 비롯한 소수의 강대국에게만 기회를 제공하는 독점적 무대는 아니다. 인공지능(AI) 시장의 국제 경쟁자들은 미국이 혁신적인 국방용 AI 애플리케이션을 놓고 연구에 연구를 거듭한다는 사실을 잘 알고, 그들 또한 무한 경쟁해야 한다는 압력을 받고 있을 것임은 자명하다. 대표적으로 이웃 나라 중국은 이 점에서 선두적인 경쟁국으로서 2030년까지 AI 개발에서 세계적인 선두를 차지하겠다는 계획을 2017년에 내놓았다. 현재 중국은 AI를 이용하여 빠르고 더 많은 정보에 입각한 결정을 내리고 다양한 유형의 군사 무기를 개발하는 데 주력하고 있다. 2017년 7월 20일, 중국 정부는 2030년까지 AI의 선두를 차지하려는 계획을 상세히 기술한 전략을 발표했고, 그러자 두 달도 채 되지 않아 블라디미르 푸틴은 이 분야의 선도자가 될 것이라고 공식적으로 발표했다(Daniel S. Hoadley & Nathan J. Lucas, 2018).[221)]

현실적으로도 인공지능 자율무기는 또한 그 누구도 예후를 정확히 예측할 수 없고, 독특한 형태의 조작에 취약할 것이며, 전장이라는 돌발적인 극한상황에서 인간-기계 상호작용에 문제를 제기할 수도 있을 것이다. 많은 전문가들은 AI가 혁명적인 영향은 아니더라도 최소

---

220) Campaign to Stop Killer Robots, Country Views on Killer Robots, 14 April, 2016(http://www.stopkillerrobots.org/wp-content/uploads/2016/04/CountryViews_14Apr2016.pdf).

221) 이에 d OpenAI의 설립자와 치명적 자율무기 기술분야 114명의 국제 지도자들이 공동 서명한 서한을 유엔에 제출했다. 그들은 자율적인 무기는 “무기가 그 어느 때보다도 더 빨리 싸울 수 있다”고 경고했다. “An Open Letter to the United Nations Convention on Certain Conventional Weapons,” August 20, 2017(https://www.dropbox.com/s/g4ijcaqq6ivq19d/2017%20Open%20Letter%20to%20the%20United%20Nations%20Convention%20on%20 Certain%20Conventional%20Weapons.pdf?dl=0).

한 미래 전쟁에 진화적인 영향을 미칠 것이라고 믿는다. 분명히 AI는 군사적인 맥락에서 많은 이점을 제공할 수 있는 잠재력을 가지고 있지만 뚜렷한 도전도 불러올 수 있다. 인공지능 기술은 자율작동을 촉진함으로써 인간을 배제하는 경향을 보일 것이고, 군사적인 의사결정에서 인간이 관여하는 대화와 외교가 아니라 실행 전략 같은 과학기술 위주의 판단을 강조할 수 있을 것으로, 그 결과 군사행동의 속도와 규모를 증가시킬 수 있을 것으로 판단된다. 법적・윤리적 논의 없이는 자율무기 체계는 국제질서를 새롭게 재편할 약육강식의 정글의 법칙이 적용될 영역임을 보여준다. LAWs는 이미 국가안보의 핵심 쟁점이 된 것이다.

2. 국방 분야 인공지능 윤리 이슈 및 쟁점

치명적 자율무기는 단적으로 상업적 목적이 농후한 인공지능 의사 왓슨과는 기술의 생산단계, 기술의 소비단계에서 전혀 다른 경로를 보인다. LAWs는 각국이 극도의 보안 아래에 연구와 생산 그리고 시행착오를 반복하여 더욱 정교화하는 관계로 그 누구도 생산과 활용에 따른 관계자별 윤리 이슈를 망라하여 예상하기는 어렵다. 단적으로 오늘날 미국, 중국, 러시아, 영국, 프랑스 등은 원격조정 살상 드론이나 전투차량을 국제기준의 법적・윤리적 논의가 없는 상태에서 이미 운용하는 것으로 알려져 있다. 모두 자국의 국가이익과 국가안보를 내세운 결과이다.

자료를 보면 미국 육군은 원격조종 전투차량(Remotely Operated Vehicles: ROVs)의 운용을 공식화했다. 2007년 10월 포스트－밀러(Foster Miller)사는 3대의 원격조종 기관총 탑재로봇을 이라크에 파견했다. 동 로봇에는 특수무장정찰원격행동 시스템(SWORDS)이 설치돼 있다고 한다(웬델 월러치・콜린 알렌, 2014). 아이로봇 사는 팩봇(Packbot)을 이라크에 집중적으로 보낸 회사로 2008년 후반기에 이미 워리어 X700(Warrior X700)을 선보인 것으로 알려졌고 성능은 지속적으로 개선되는 것으로 알려져 있다.[222] 이미 실전 배치 중인 로봇 제품은 원격조종 차량과 기관총에 그치지 않는다. 크루즈 미사일과 같은 반자동 로봇 시스템은 이미 폭탄을 실어 나른다. 미국의 경우에 2005년 뉴욕 타임스(New York Times)는 이미 펜타곤은 인간 군인을 자동화 로봇으로 대체할 목표를 세워놓았다고 보도했었다(웬델 월러치・콜린 알렌, 2014).

---

222) Army Guide, Warrior x700, http://www.army-guide.com/eng/product4196.htm.

이처럼 인간 살상이 궁극적인 목표인 치명적 자율무기(LAWs)는 인간의 존엄성을 유지하면서 사람의 생명을 연장하는 것이 궁극적인 목표인 인공지능(AI) 기반 의료로봇이나 인간 파트너 로봇 등의 이슈와는 전혀 다르게, 개발되어 현장 배치되면 그 이후의 진행은 자율무기의 독자적인 판단으로 작동된다. 결국 프로그래밍부터 현실적 작동까지 윤리 이슈가 통합적으로 작동된다는, 즉 GIGO가 가장 강력히 작동되는 인공지능(AI)영역이라고 할 수 있다(한희원, 2018b). 그러므로 다른 산업부문에서의 인공지능(AI) 로봇과는 제기하는 윤리 이슈와 대응방안이 매우 색다르지 않을 수 없다. 그래도 치명적 자율무기체계의 본격적인 도래가 불가피하다면 통일적인 이해를 위해 다음의 4가지 영역에 대해 가능한 사고를 해봄으로써 이해의 폭을 넓힘과 동시에 그나마 현재의 국내법과 국제법 체계 그리고 윤리적・법적・도덕적 체계에 부합되는 LAWs를 염두에 둔 고민을 해보기로 한다.

가. 치명적자율무기체계의 일반적 윤리이슈

① 윤리적・도덕적 해이의 문제

스튜어트 러셀(Stuart Russell)은 치명적 자율무기가 반윤리적이고 비인간적이라면서 인공지능(AI) 기반 치명적 자율무기시스템의 가장 큰 문제점은 전투원과 비전투원을 구별하는 것이 어렵다는 점과 전쟁을 쉽게 생각하도록 하여 더 많은 살상을 초래할 수 있다는 점을 지적했다.[223] 완전자율무기 체계는 전장에서 사람을 배제하게 됨으로써 군인과 시민, 정치인과 정책결정권자가 전쟁의 참상을 직접 경험할 기회를 배제하고 추상적이고 일반적인 희망에 기초하여 정책 결정을 하게 함으로써 전쟁의 문턱을 낮추고 전쟁으로 가는 자동화된 도로를 열어 줄 수 있을 것이라는 것이다. 또한, LAWs가 일단 실전에 배치되면 전쟁에 대한 인간 통제를 어렵게 만들 것이고 살인 책임자의 경계를 모호하게 할 수 있다는 점도 지적했다(한희원, 2018b).

② 생명권 경시의 문제

치명적 자율무기 쟁점의 본질은 윤리적인 문제로 그 실질은 치명적 자율무기가 설령 합법적인 군사목표만을 대상으로 제작되었다고 하는 경우에도 불가피하게 생명권 침해 쟁점을 야기한다는 사실이다(한희원, 2018b). 물론 살인은 전쟁에서는 매우 당연한 정당행위라는 사

223) Russell, Stuart(27 May 2015). “Take a stand on AI weapons”. International weekly journal of science. 521; Crootof, Rebecca (2015). “The Killer Robots Are Here: Legal and Policy Implications”. 36 Cardozo L. Rev: 1837, p. 388.

실에 비추어 보면, 즉 현행 전쟁법의 법리에 입각하는 경우에도 전쟁에 개입되어 있는 전투원은 언제든지 어떤 방법이나 수단에 의해서도 사망할 위험성에 개입되어 있는 존재라는 사실에 비추어보면 전투원에게 삶에 대한 권리, 즉 생명권이 과연 주어지는 것인가에 대한 의문이 제기되는 것도 사실이다(한희원, 2018b). 그러나 명백한 것은 인간은 그 누구나 설령 무력충돌 현장에서도 자의적으로는 살해당하지 않는다는 관점에서 본다면 모든 인간은 비인간적으로 살해되지 않을 권리가 있다. 불가피하게 LAWs는 생명권 경시 흐름을 초래할 위험성을 피할 수 없다(한희원, 2018b).

③ 인간성의 피폐화와 인간의 수단화 문제

철학자 임마누엘 칸트는 인간 공동체 사회의 핵심적 도덕법칙으로 2가지 정언명령을 제시했다. 제2의 정언명령이 인간 존엄성 목적대우의 원칙이다(한희원, 2011). 치명적 자율무기는 생명권 경시, 인간 존엄성 와해와 같은 법적 쟁점을 일으키지만, 그 근저에는 출발부터 사람을 목적으로서가 아니라 단순한 공격대상, 즉 수단으로 취급한다는 사실이다. 과학기술적으로 킬러로봇은 사람들을 사람으로 인식하지 않고 대상으로 인식하게 프로그래밍된다. 어떤 인간 목표는 공격대상으로 코딩되고, 다른 인간 목표는 죽이지 않는 대상으로 코딩되는 것이다. 현재 수준에서 인공지능 로봇이 이해하는 것은 약속기호인 2진법의 1과 0 형태의 숫자 비트 정보이다(한희원, 2018b).

이러한 연유 등으로 치명적 자율무기는 정의론 관점에서 본다면 불가피하게 인간성 피폐화를 일으켜 인간을 동물이나 무감각한 로봇으로 하향 수준화시킬 위험성이 있다는 것이 킬러로 로봇의 윤리문제라고 할 것이다.[224] 이러한 이유 등으로 어쨌든 무력충돌에서의 살상에도 도덕적 배려가 반드시 필요하고 위와 같은 사려 깊고 도덕적인 결정을 내릴 책임은 프로그래머들의 어깨에 뿐만이 아니라 군사행동을 승인하는 정치 지도자들에게도 당연히 부과되어야 할 것이다(Simpson, Thomas W & Muller, Vincent C, 2016).

나. 신기술 출현에 따른 윤리 이슈

오늘날 기술적으로 급성장하고 있는 인공지능(AI)은 잠재적으로 국가안보에 직접적인 영향을 미치는 과학기술개발 분야이다. 이에 펜타곤은 정보수집 및 정보분석, 병참, 사이버 공간 운영, 지휘 및 통제, 다양한 군용 자율차량 분야에 집중했고, 이라크와 시리아 내 저항세

224) 인간성 수단화에 대한 각종 사례 분석은 한희원(2011) 참조.

력의 목표물 파악을 위해 AI 알고리즘을 사용하는 프로젝트 마븐(Project Maven)이라는 선봉책을 통해 AI 시스템을 전투에 통합했다.[225)]

그런데 이러한 일련의 사건들을 2017년 미국 의회가 청문회에서 다루면서 윤리 이슈가 개방적으로 논의되었다. 의회에서 제기된 몇 가지 의문들은 다음과 같다.어떤 종류의 군사 AI 어플리케이션이 가능하고, 만약 있다면, 어떤 제한을 두어야 하는가? 방어를 위해 AI를 채택할 때 얻을 수 있는 고유한 장점과 취약점은 무엇인가? AI가 미래 전쟁을 어떻게 변화시킬 것이며, 미국 경쟁국들과의 군사적 균형에 어떤 영향을 미칠 것인가? 쟁점을 정리하면 다음과 같다(맥스 테그마크, 2017).

① AI 개발을 위한 민간 및 정부 자금의 적절한 균형은 무엇인가?
② 국민의 대의기구인 의회는 군사 AI 적응을 용이하게 하는 국방획득 개혁사업에 어떤 영향을 미칠 수 있는가?
③ (막연한 반대가 아니라)의회나 국방부가 인공지능(AI) 무기개발을 효과적으로 감독하려면 어떤 변화가 필요한가?
④ 군사 목적 인공지능(AI) 애플리케이션의 개발 보급을 위해서 필요한 규제변경에는 어떠한 것이 있는가?
⑤ 국제경쟁사로부터 미국 인공지능(AI) 기술을 보호하고 지속적으로 미국의 우위를 지키기 위해서는 의회는 어떤 조치를 취해야 하는가?

의회는 이러한 질문에 대한 답변과 AI 기술의 미래 궤적에 영향을 미치는 데 사용할 수 있는 많은 예산권과 입법권이라는 통제장치를 보유하고 있다. 논의되는 쟁점을 살펴보면 알 수 있듯이, 치명적 자율무기 분야에서도 가장 선도적인 위치에 있는 것으로 알려진 미국의 경우에도 LAWs에 대한 의회 차원의 논의의 핵심도 찬성/ 반대, 즉 금지/허용 차원을 넘어서서, 이미 실전 배치된 현신을 반영하면서 과연 어느 수준까지 의회가 관여할 것인가에 대한 철저한 국익 중심적 관점인 것을 알 수 있다. 그것은 비단 미래 인공지능 무기가 국방 분야에서 국방안보 확보를 위한 방편으로서 뿐만이 아니라, 팩스 아메리카나라는 거대함 담론에서 경쟁국가와는 차원을 달리하는 총체적인 국가경쟁력 우월성의 시각에서 논의되는 쟁점

225) Adam Frisk, What is Project Maven? The Pentagon AI project Google employees want out of, *available at* https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/.

으로 이해된다. 이것은 인터넷을 비롯하여 인류의 삶에 불가결한 과학기술 제품이 사실은 국방 분야에서 먼저 개발되어 활용되었던 점에서도 알 수 있다.

다. 신기술 개발/생산단계에서의 윤리 이슈

LAWs의 국가사회적 윤리 쟁점을 더욱 복잡하게 하고 논란을 촉발시키는 것은 인류의 행복증진을 목적으로 하는 상용분야에서의 인공지능(AI) 기술이, 약간의 변형만 거치면 치명적 자율무기로의 구동이 가능하다는 사실 때문이다. 이 때문에 민간 상용분야에서의 AI 기술을 군사적으로 전용하는 것에 대해 펜타곤은 AI와 같은 시스템을 군사 목적으로 전용하고 획득하기 위해서는 그 필요성 판단과 정당성 등의 확보를 위해 국방 무기 획득과정(DAP: Defense Acquisition Process)을 채택한다.[226] 이 경우에도 상용 목적의 AI 기술이 군사 목적으로 전용됨으로 인한 위험성으로 민간부문의 AI 연구가 위축될 가능성도 배제할 수 없을 것이다.

한편 AI는 국가안보 분야로 진입함에 따라 고려해야 할 여러 가지 특성을 갖고 있다. 첫째, 오늘날 사물인터넷이 등장한 것처럼 AI는 사실상 모든 것에 통합될 가능성이 있기 때문에 전지전능한 기술로 1세기 전에 전기(電氣)가 했던 것처럼 불활성 물체를 활기차게 할 것이다. 두 번째로 인공지능(AI) 기반 LAWs는 구동 알고리즘의 교환만으로 성능이 개선되고 변경되는 것으로 본체 그 자체는 상대적으로 투명하기 때문에 무기 제품으로의 통합이 즉시 인식되지 않는다는 사실이다. 그러나 실질에 있어서 무기전용이 가능한 상용 인공지능(AI) 애플리케이션과의 통합은 매우 엄청난 결과를 초래할 것이다. 이러한 이유로 LAWs 논의와 관련된 국가・사회적 정책과제는 어느 수준의 인공지능 기술과 역량을 LAWs와의 연계산업으로 판단할 것인가의 문제가 있다.

국가와 사회의 민주화와 개방화에 따라서 특히 치명적 자율무기라는 신기술 개발단계에서는 분명히 다양한 국가적・국제적 쟁점이 제기될 것은 현재까지의 논의 자체로도 명백해 보인다. 특히 우리의 경우에는 사드(THAAD) 무기 배치에서 이미 경험한 것처럼 LAWs에 대해 시민사회단체에서의 논의가 전개되는 경우에는 국가안보와 국가이익 수호를 위한 쟁점에 대해서도 추상적 인권 같은 또 다른 가치 기준으로 논의가 전개될 위험성도 보인다. 그

226) Adam Frisk, What is Project Maven? The Pentagon AI project Google employees want out of, *available at*, https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/.

뿐만이 아니다.

국제사회에서 독자적 의사결정 권한을 갖지 못하는 우리의 경우에 상대 국가들의 견제는 물론이고 UN 같은 공적 국제기구의 방침, 휴먼라이츠워치나 앰네스티 인터내셔널 같은 국제인권기구나 국제환경기구의 문제 제기를 또한 무시할 수 없는 처지다. 특히 국제사회에서 독자적인 의사결정권을 갖는 소수의 강대국을 제외하고는 아무리 국방안보나 국가이익과 관련된 신기술 개발이라고 하더라도 기계가 인간을 살해대상으로 자율적으로 정한다는 그 파괴적 파급효과 그러므로 인간 본성(humanity)에 대한 본질적인 침해위반 가능성으로 인하여 LAWs 개발이 공표되는 순간 국제 연대적 저항에 부딪칠 가능성은 거의 명백해 보인다고 사료된다.

그러므로 인류 역사에서 현재까지는 가장 파괴적인 무기로 알려진 핵무기 개발이 극비의 국책사업으로 진행된 것처럼, 현실적으로 LAWs의 개발단계에서의 윤리 이슈는 할 것인가 아니면 말 것인가? 라는 선택의 문제와 개발한다고 하는 경우에는 초극도의 보안을 유지하는 절차와 방법의 문제로 집중될 것이다.

라. 신기술의 실험단계에서의 윤리 이슈

새로운 무기의 실험단계에서의 주된 쟁점은 안전성과 효용성일 것이다. 이미 정책 결정으로 개발과 생산이 판정되었다고 한다면, 역설적으로 치명적 자율무기는 목표물을 제대로 타격하고 효율적으로 제압, 즉 살상할 수 있는 그야말로 치명적인 무기로 개발되어야 할 것이기 때문이다. 이러한 점 때문에 LAWs는 인류의 건강을 증진하고 생명을 연장하거나 살려내려고 고안되는 인공지능 왓슨과 같은 인공지능 의료로봇과는 차원이 다른 윤리 쟁점을 야기한다.

그러나 한 가지 명백한 것은 치명적 자율무기의 경우에 실전 단계에서도 그렇지만 실험단계에서 가장 중요한 정책적 윤리적 쟁점은 국제전쟁법이 요구하는 무기로서의 적합성으로 그 가운데서 핵심은 '의미 있는 인간통제(meaningful human control)'의 구현일 것이다. 사실 LAWs에 대한 보편적인 개념 정의를 개발하는 것과 관련하여 중요한 것이 '의미 있는 인간통제(MHC)' 개념이다. 의미 있는 인간통제(MHC)란 공격할 대상의 선택과 효과에 대한 인간의 직접적인 통제를 의미한다. 인간이 무기를 효과적으로 통제할 수 있을 때 완전히 자율적인 무기의 배치로 인해 발생하는 윤리적 그리고 법적 문제가 발생하지 않을 것이기 때문

에 MHC가 LAWs의 필수요건이 되어야 한다는 것이다(Huffingtonpost, 2013. 7. 15). 그러나 '인간통제'의 정확한 수준이라든가 '의미성'의 필수적 매개 변수가 무엇이어야 하는지 여전히 모호하다. 여하튼 '의미 있는 인간통제(MHC)'를 달성하려면 인간 통제자가 공격목표물과 동원된 LAWs 시스템의 내용과 LAWs가 취하려는 행동의 적법성에 대한 제반 정보를 정확히 제공받을 수 있는 위상이 주어져야 하는 것이 필수적인 요청일 것이다.

한편 위 카이스트 사태에서 파이낸셜 타임스가 "한화시스템이 무고한 민간인까지 무차별 살상한다는 비난을 받아온 집속탄(대형 폭탄이 수백 개의 소형 폭탄으로 분리돼 다수 인명을 살상하는 무기) 생산에 관여하고 있다"고 보도한 데 대해 한화시스템 관계자는 "집속탄은 (주)한화가 만드는 것으로, 한화시스템은 2015년 한화그룹에 흡수 합병된 삼성탈레스가 전신인 회사"라고 밝히기도 했다. 세계는 이미 각국 무기에 대한 정보를 아는 것이고, 조금이라도 연관된 정보로 신무기 개발과 배치에 대한 주권국가의 행보에 대해서도 제약을 가하는 것이 국제질서의 현실이다.

더 나아가 일부 분석가들은 LAWs를 통제하려는 노력이 비록 근본적으로 그렇게 하도록 설계되지 않았더라도 치명적인 시스템에 사용될 수 있는 응용 프로그램에 대한 엄격한 통제 때문에 다른 유용한 군사 AI 기술의 개발이 억제될 수도 있을 것이라는 우려를 제기했다. 최근 비공개 증언에서 어느 전문가는 "만약 기술을 규제한다면 전쟁을 더 인간적으로 만들 수 있는 길을 포기하는 것이 될 것"이라면서 군사용 AI 애플리케이션 규제의 타당성 판단 기준은 "인간이 전시 의사결정에서 어떤 역할을 하기를 원하는가?" 가 될 것이라고 라고 알기 쉽게 설명한다(한희원, 2018b).

마. 신기술의 실전배치 단계에서의 윤리 이슈

LAWs의 실전배치와 관련된 윤리적 이슈는 두말할 것도 없이 제대로 성능을 발휘하여 적을 적으로 인식하고 아군을 아군으로 인식함에 차질이 없을 것인가의 문제이다. 그런데 인류 무기 개발역사에서 예컨대 폭탄의 경우에 그 폭발력 증강에만 초점이 맞추어졌었지, 대량 살상성의 윤리적 쟁점에 대한 논의는 큰 고민거리가 아니었다. 그러므로 치명적 살상무기의 살상능력을 윤리 쟁점으로 논의하는 것은 부적합해 보인다. 그러나 인공지능(AI) 기반 치명적 자율무기는 다른 군사 무기와 달리 국가와 사회에 제기하는 법적・윤리적 쟁점이 한 가지 있을 수 있다. 바로 프라이버시 쟁점이다. 한편 인공지능(AI) 기반 무기는 사이버 공간

에서도 중요한 역할을 할 것으로 기대된다. 이에 미국 사이버 사령관 마이클 로저스 소장은 2016년 상원 군사위원회 증언에서 사이버 공간에서 인간 정보에만 의존하는 것은 '잃어버리는 전략(a losing strategy)'이라고 말했다.[227] 예컨대 기존의 인간이 작동하는 사이버 방어 도구는 이전의 악성 코드와 일치 사례를 찾아 해커를 차단한다. 그러므로 해커들은 그 코드의 일부만 수정하면 인간 방어 수단을 피할 수 있다. 그러나 AI 사이버 방어 도구는 네트워크의 행동 패턴 변화 예후를 자동적으로 인식하고 이상 징후를 탐지하도록 훈련되어 이전에 예상치 못한 공격 방법이나 미래 사이버 공격에 대한 보다 포괄적인 장벽을 제공할 수 있을 것으로 예상된다(WIRED, 2017. 9. 27).

바. 국제관계적 윤리 이슈

치명적 자율무기체계와 관련된 국제관계적 쟁점은 사드배치의 경험처럼 직접적으로는 무기개발과 현장배치와 관련된 인접국가, 즉 우리의 경우에는 북한은 물론이고 중국이나 일본과의 외교 안보적 관계도 있을 것이지만, 동 쟁점은 생략하고 아래에서는 국제기구와 국제단체와의 관계에 대한 문제를 살펴본다.

2013년 출범한 살인로봇금지 국제캠페인(Campaign to Stop Killer Robots)[228]은 '자율무기 개발・생산・사용에 대한 포괄적이고 선제적 금지'를 구호로 내세우면서 국제법의 제정과 각국의 국내법 제정을 포함한 관련 조치를 국제사회에 호소하고 있다. 휴먼 라이츠 워치는 "LAWs 개발, 생산 및 사용에 대한 선제적 금지"를 요구하면서 LAWs가 무력충돌에서의 구별의 원칙과 비례의 원칙 그리고 군사적 필요성의 원칙을 포함하여 국제인도법의 표준 요구조건을 충족할 수 있는지에 대해 의문을 제기했다.[229] 2014년 국제여성평화・자유연맹(Women's International League for Peace and Freedom: WILPF)은 LAWs가 의미 있는 인간통제 없이 작동할 수 있는 가능성에 심각한 우려를 제기하고 자율살상무기를 프로그래밍하는 자체가 이성과 법에 대한 근본적인 도전이라고 LAWs 그 자체를 비판했다(Women's International League for Peace and Freedom, 2014). 국제 적십자위원회(ICRC)는 자율무기체계의 개발중단이나 생산 및 사용금지 같은 국제운동에는 참여하지 않기로 선택하면서도, 각국

227) https://www.defense.gov/News/Article/Article/938941/national-security-experts-examine-intelligence-challenges-at-summit/.

228) https://www.stopkillerrobots.org/.

229) http://www.hrw.org/topic/arms/killer-robots.

은 LAWs의 진전된 개발이나 무력 현장에의 배치에 앞서서 “자율무기의 핵심적 기능(critical functions)에 있어서 무기 자체의 자율성이 제기하는 근본적인 법적 및 윤리적 문제를 고려할 것”을 촉구했다. ICRC는 또한 자율무기의 핵심적 기능을 포함하여 각국이 개발하거나 구입하는 신무기와 전쟁의 신기술에 대한 완전한 법적 재검토를 할 것을 촉구했다(한희원, 2018b).

순수 민간단체의 동향으로는 2015년 7월 1,000명이 넘는 인공지능 전문가들이 인공지능자율무기 금지를 촉구하는 결의안에 서명했다. 결의안은 부에노스아이레스에서 개최된 제24회 인공지능 국제합동회의(IJCAI-15)[230]에 제출되어 대표적으로 스티븐 호킹(Stephen Hawking), 엘론 머스크(Elon Musk), 스티브 워즈니악(Steve Wozniak), 노암 촘스키(Noam Chomsky), 스카이페(Skype)의 공동 창립자인 쟌 탈린(Jaan Tallinn), 구글 딥마인드(Google DeepMind) 공동 발명자인 데미스 허사비스(Demis Hassabis) 등이 추가 서명했다. 서한은 “치명적인 자동화 무기는 일단 개발되면 더 큰 규모로 무장충돌을 일으킬 것으로 전쟁의 3차 혁명이 될 수 있다며 인류의 피해가 엄청난 수준에 이를 것이고 킬러로봇이 전투의 문턱을 낮춰 인간의 삶을 크게 훼손할 것”이라고 주장했다(한희원, 2018b).

G7, 경제개발기구(OECD), 아시아태평양경제협력체(APEC)등 많은 공적 국제기구들도 인공지능(AI) 기반 자율무기에 대한 쟁점을 논의했다. 그 중심에 국제연합(UN)이 있다. UN은 치명적 자율무기(LAWs)를 ‘특정재래식 무기금지협약(convention on certain conventional weapons: CCW)’과의 상관관계에서 집중적으로 논의하고 있다. 상세는 후술하겠지만 CCW는 “전투원에게 불필요하거나 부당한 고통을 초래하거나 민간인에게 영향을 미칠 수 있는 특정 유형의 무기의 사용을 금지 또는 제한”하는 국제조약이다. UN 군축포럼(disarmament forums)은 LAWs가 제기하는 문제점을 해결할 수 있는 매우 적절한 통로이다. UN 군축포럼에서의 집약된 논의는 UN 총회에서 무기통제나 군비경쟁제한 등에 대한 결의안이나 권고로 발령될 수 있다. 이러한 전개의 일환으로 UN 인권이사사회(Human Rights Council)는 크리스토프 헤인즈(Christof Heyns) UN 특별보고관(Special Rapporteur)에게 킬러로봇이 어떻게 국제전쟁규범의 요건을 충족시킬 수 있을 것인지에 대해서 조사할 것을 의뢰했고 제출받은 보고서를 2014년 4월 1일 UN 총회에 제출했었다.

---

230) International Joint Conference on Artificial Intelligence (IJCAI-15).

〈표 4-8〉 각 단계 행위자별 윤리이슈

단계	행위주체	윤리 이슈
정책결정단계	군	• 치명적자율무기 개발을 필요성 및 허용여부 • 방어목적의 AI를 채택할 때 얻을 수 있는 고유한 장점과 취약점은 무엇인가 • AI가 미래 전쟁을 어떻게 변화시킬 것인가에 대한 연구 • LAWs는 경쟁국들과의 군사적 균형에 어떠 영향을 미칠 것인가 • 국가안보 최고수호자로서의 윤리의식 • LAWs의 사고책임에 대한 법적·윤리적 쟁점에 대한 신속한 대응책이 강구되었는가 • 쟁점들에 대한 국제법 준수 및 대응 입법은 준비되었는가
	최고통수권자	
	윤리/법률가	
개발단계	개발자	• 알고리즘에 어떤 윤리기준을 주입시킬 것인가 • 전쟁법에서 요구하는 의미 있는 인간통제를 어떻게 구현할 것인가
	윤리/법률가	• 법적/윤리적 기준을 제시할 수 있는가
	운용책임자	• 운용자가 필요로 하는 별도의 기준은 없는가
실험단계	개발자 및 제조사	• LAWs의 안정성은 담보되었는가 • 지향하는LAWs는 극대의 살상이라는 효용성은 보장되는가 • 테러리스트 등에 대한 보안수준
	공적 주체	
	소비자 및 시민단체	
국제협력단계	국방/외교	• LAWs에 대한 국제 보편의 기능적, 윤리적 기준은 무엇인가
	윤리/법학자	• 기존 국제법질서와의 관계는 어떻게 되는가
	운용책임자	• 보편과 특수의 조화를 어떻게 도모할 것인가
실전배치단계	부대책임자	• LAWs의 살상력은 충분하가 • 아군과 적군, 민간인과 무장군인을 정확히 구분할 수 있는 여부
	정책담당자	• 상대세력의 LAWs에 필적하는지 여부 • 정책부서와 정보주서의 유기적인 협조가 이루어지는가 • 사이버공간에서도 구동이 가능한 수준의 LAWs인지 여부 • 프라이버시 유출에 대한 대비가 적절히 이루어졌는지 여부 • 현장 배치된 LAWs의 프로그램의 해킹에 대한 대비가 충분한지 여부
	윤리/법률가	• 어떤 유형의 피해를 인정할 것인지 여부

제 5 장 인공지능윤리 이슈 정책적 대응방안: AI 윤리 대응 프레임워크

제 1 절 자율 윤리 대응 프레임워크 구축의 필요성

제4장까지의 내용을 통해 살펴본 각국의 인공지능 관련 윤리 규범 형성 노력과 산업 분야별 윤리 이슈의 도출 작업은 산업에 인공지능을 실제로 적용할 때의 리스크를 대비할 수 있는 의제를 제시한다는 측면에서 그 자체로도 의미 있는 일이다. 그러나 이러한 의제에 구체적인 실행 방안이 뒤따르지 않는다면, 윤리 규범은 명목에 지나지 않게 될 것이다. 따라서 윤리 규범과 윤리 이슈에 대한 논의를 더욱 실효성 있게 만들기 위해서는, 이들을 바탕으로 한 정책적 대응 방안을 제시할 필요가 있다.

국가가 구속력을 가지고 일반적 행동 기준을 제시하는 법적인 규제와는 달리, 윤리 규범과 관련해서는 도덕률에 대한 개별 행위 주체들이 이를 준수함에 있어서 자율성이라는 요소가 존재한다(Kreß, 2012: 29; 권헌영 외, 2017: 34). 즉, 윤리적 대응의 영역에서는 공적 주체의 간섭을 통해서가 아니라, 윤리 규범에 대한 사회적 합의와 그에 따른 수범자의 자율적인 행위가 중요성을 가지는 것이다. 과학기술기본법 제4조 제4항은 “과학기술인은 자율을 바탕으로 과학기술 활동을 수행하되 과학기술이 미치는 사회적·윤리적 영향을 고려하여 진실성 있게 수행하여야 하며, 경제와 사회의 발전을 위하여 과학기술의 역할이 매우 크다는 점을 인식하고 자신의 능력과 창의력을 발휘하여 이 법의 기본이념을 구현하고 과학기술의 발전에 이바지하여야 한다”고 규정하고 있다. 동조의 규율이 지극히 포괄적이고 선언적인 성격을 가지기는 하지만, 실체적인 활동을 하는 기업(개발자 내지 사업자)가 윤리 준수의 책임을 지고 있다는 것이 과학기술 분야에서의 입법자의 의도라는 점을 알 수 있다.[231] 특히 시장

231) 인공지능의 사용과 직접적으로 관련된 규율은 아니지만, 정보통신망법 제44조의 4 역시 “정보통신서비스 제공자단체는 이용자를 보호하고 안전하며 신뢰할 수 있는 정보통신서비스를 제공하기 위하여 정보통신서비스 제공자 행동강령을 정하여 시행할 수 있다”고 규정하여, 행위자들이 스스로 구축하는 자율 윤리통제의 중요성을 강조하고 있다.

행위자의 인공지능 윤리 통제는 공권력의 법집행에 의한 강제가 아니라 시장 행위자의 자율에 따라 이루어지는 것이 중요하다는 점을 인식할 필요가 있다. 이를 위해 윤리 규범의 준수가 규제로 작용하는 것이 아니라 장기적인 관점에서 시장 행위자의 이윤 창출에도 도움이 될 것이라는 점에 대한 이해와 설득이 필요하다. 예컨대 자율 대응을 하지 않았을 경우 사업자의 명성의 훼손이 나타날 수 있다는 점, 사전 윤리 대응을 통해 향후 나타날 수 있는 소비자 피해에 따른 리스크를 방지할 수 있다는 점, 윤리적 대응의 내용과 범위를 명확히 할 경우 그 틀 안에서 오히려 기술 개발자의 자유가 확대될 수 있다는 점 등을 고려할 수 있다.

한편 국가정보화기본법 제40조는 국가와 지방자치단체에게 “건전한 정보통신윤리를 확립하기 위하여 미풍양속을 해치는 불건전한 정보의 유통을 방지하고 건강한 국민정서를 함양하며, 불건전한 정보로부터 국민을 보호하기” 위한 시책을 마련할 의무를 부여하고 있다. 동조의 규정에 비추어 볼 때, 윤리 문제 대응의 책임을 전적으로 사인(私人)인 시장 행위자에게만 맡기고 있지는 않다는 점을 고려해야 할 것이다. 입법상의 근거는 찾아보기 어렵지만, 개발자 내지 공급자와 공적 주체 외에 인공지능 기술을 이용한 제품 내지 서비스의 이용자 역시 윤리 대응 프레임워크에 있어 중요한 한 주체라고 할 수 있다. 한 편으로는 이용자가 인공지능 기술 사용의 수혜자로서 윤리 규범에 따른 책임을 져야 할 필요가 있으며, 다른 한 편으로는 비윤리적 인공지능 기술 사용의 피해자로서의 지위를 가질 수도 있기 때문이다.

따라서 공적 주체 및 시장 행위자가 수범해야 하는 윤리와 더불어 이용자가 가지는 행위자와 피보호자로서의 양 측면을 반영한 윤리를 포괄할 수 있는 프레임워크의 구성이 필요하다고 말할 수 있다. 또한 그러한 프레임워크는 이미 문제가 제기되어 있는 개별 사안에 대한 해법을 제시하는 데에 그치지 않고, 인공지능 기술 적용의 전 과정에서 유기적으로 연결되어, 대부분의 잠재적 윤리 쟁점에 대해 기능적이고 효과적으로 적용될 수 있는 것이어야 할 것이다. 즉, 향후 제기될 수 있는 인공지능 윤리 문제에 대해 인공지능 사용과 관련되어 있는 모든 관계자가 스스로 효과적이고 합리적으로 대응할 수 있는 자율적 구조를 마련하는 것이 본 프레임워크의 궁극적 지향점이다. 제4장에서 살펴본 바 있는 인공지능 산업별 윤리 쟁점들의 특징을 반영한 윤리 규범을 실제로 관철하기 위해서는, 자율 윤리 대응 프레임워크 구축을 위해 고려해야 하는 원칙들이 무엇인지 고찰해야 할 것이다. 다음 단계로 이와 같은 원칙에 입각하여 산업 영역별 사업 수행의 과정에 따라 각각의 행위자별로 또는 각각의 쟁점별로 행동 규범 내지 준칙을 살펴볼 필요가 있다. 마지막으로 이러한 사안별 행동규범

들을 체계적으로 연계하여 종합적 프레임워크와 분야별 특성을 반영한 프레임워크를 도출해 낼 수 있을 것이다. 이러한 프레임워크는 단순히 윤리 규범을 제시하는 데에 그치지 않고, 그러한 윤리 규범의 실천 방향을 제시한다는 점에서, 최근 각 분야에서 연구되고 있는 인공지능 윤리 강령 내지 가이드라인들과 차별성을 가진다고 할 수 있을 것이다.

이하에서는 위 내용들을 단계별로 살펴보면서, 인공지능 윤리 이슈들을 포괄할 수 있는 정책적 수단으로서 자율 윤리 대응 프레임워크를 제안하고자 한다.

제 2 절 자율 윤리 대응 프레임워크 구축 시의 고려 사항

1. 개관

적절한 자율 윤리 대응 프레임워크를 구축하기 위해서 일반적인 윤리 거버넌스의 특징, 인공지능 기술 적용의 특수성 및 현재 우리 사회에서의 인공지능 기술 수용성 및 그에 따른 사회 현안의 해결 등을 염두에 두어야 한다. 즉, 윤리적 대응의 개념적 특성(법적 규제와의 구별), 프레임워크 구축 과정에서의 절차적 문제(협력 모델의 고려), 인공지능 기술 자체가 촉발시킬 수 있는 쟁점(이용자 보호 및 인공지능 기술 개발 속도 고려), 프레임워크의 실질적 적용 가능성(사전 배려와 사후 통제의 조화, 생애주기에 따른 단계별 이슈, 프레임워크의 범용성) 등을 고려해야만, 윤리적 대응의 프레임워크를 효과적으로 작동할 수 있도록 구축할 수 있을 것이다. 이하에서 각각의 범주에서 고려할 사항들을 살펴보는 한편, 각각의 고려 사항들을 구체적으로 실천하고 있는 각 분야의 거버넌스 실례를 찾아보기로 한다.

2. 법적 규제와의 경계 설정

먼저 윤리 대응의 구조와 공적 주체가 행하는 구속력을 가지는 규제와의 경계를 설정함으로써, 윤리 문제에의 대응이 결과적으로 산업 성장을 위축시키는 규제가 되지 않도록 하는 노력이 필요하다. 이를 위해 먼저 규율의 대상, 행위의 성질 및 규율의 속성 측면에서 각각의 문제를 고려한 후, 윤리적 책임과 법적 책임의 경계를 구성해야 할 것이다.

그러나 윤리와 법은 사회 구성원들에게 일정한 의무를 요구하는 사회적 합의라는 측면에서 동일한 스펙트럼 안에 있으며, 그 본질적인 특징에 대한 구분이 쉽지 않다. 즉, 사회적 합

의에 따른 규율이라는 본질적인 개념 면에서는 구별이 되지 않지만, 입법자가 법의 영역으로 설정한 부분을 법이라고 지칭한다는 점에서 외관에 따른 형식적 경계만이 가능한 것이다. 인공지능 기술 적용 영역에서의 윤리와 관련해서는 아직 사회적 합의가 명확히 구축된 것이 아니라, 기술의 발전에 따라 형성되어 가는 중 혹은 형성될 것이 예상되는 것이라고 평가할 수 있다. 또한 인공지능에 특유한 법적 규율 역시 아직 정립되어 있지 않다. 인공지능 윤리의 관철을 위한 프레임 워크 구축 수 법적 규율과 윤리적 규율의 경계는 반드시 필요한 것이므로, 양 규범의 형성 과정을 중심으로 여러 요소들에 대한 고찰이 필요하다.

첫 번째로 행위의 양태에 대한 고려가 필요하다. 행위가 자발적인 것이었는지, 비자발적인 것이었는지, 그리고 행위의 결과에 대해 행위자가 인식을 하였는지, 하지 못하였는지의 여부는 법적 책임과 윤리적 책임의 존부를 판단하는 요소 중 하나가 될 수 있다. 예컨대 행위자가 그 결과에 대해 충분히 인식을 하고 자발적으로 행위했을 경우에는 일반적으로 법적이든 윤리적이든 높은 수준의 책임을 지게 될 것이라는 점을 예상할 수 있다.

다음으로 윤리 규범의 특징상 사회 환경적인 측면을 고려해야 한다. 윤리는 사회적 합의의 산물이므로, 구체적인 윤리 규범의 형성 이전에 사회적 규범에 대한 고려가 필요한 것이다. 즉, 윤리 규범이라는 직접적인 규율로 넘어가기 이전 단계에서 일정한 가치에 대한 사회적 합의가 선행되어야 하는 것이다. 사회적 규범에 대한 합의 절차는 다양한 방식으로 이루어진다. 사회 구성원들의 자연스러운 합의에 의해 이루어질 수도 있으나, 사회적 가치를 구성하려는 정책 참여자들의 노력을 통해 구현되기도 한다. 특히, 각 산업 분야에의 인공지능의 적용과 같은 기존 사회에 존재하지 않았던 현상에 대한 사회적 가치를 형성할 때에는, 각종의 논의 플랫폼 등을 통해 사회적 규범을 확인하려는 노력을 선행할 필요가 있을 것이다.[232] 민관학의 협력 플랫폼을 통해 사회적으로 합의된 규범을 담은 가이드라인을 제정하는 것도 대안의 하나가 될 수 있다. 사회적 규범의 존재 유무는 윤리적 비난 가능성으로 이어진다. 즉, 행위자의 행위를 비난할 사회적 합의가 존재하지 않는다면 윤리 책임 역시 존재하기 어렵게 될 것이다.

마지막으로 보다 직접적으로는 규범 구성의 주체와 관련하여 입법자가 직접 또는 행정청에 대한 수권을 통하여 법규를 제정했는지의 여부가 법과 윤리의 경계를 가르는 기준이 된

232) 예컨대 지능정보 기반의 산업을 노동 환경에 수용하는 정책의 방향성을 제시한 독일의 「노동 4.0 (Arbeiten 4.0)」은 의제 발굴을 위해 사회적 가치의 확인 과정을 거친 바 있다(BMAS, 2016).

다. 대외적 구속력을 갖는 법규는 입법자(법률) 또는 입법자의 명시적인 수권을 받은 행정청(법규명령)만이 제정할 수 있다. 법적 규율이 이루어지지 않은 윤리 규범에 대해서는 강제력을 부여할 공권력의 개입이 존재하지 않는 것이 원칙이다. 즉, 자율성이라는 요소가 윤리 규범을 강하게 특징짓는 것이다. 그러나 법적 구속력을 가지는 규범이 아니더라도 국가가 주도하여 제정한 윤리 규범은 시장 행위자들에 대하여 실질적인 규범력을 가질 위험성이 있다. 따라서 인공지능 기술에 대한 자율 윤리대응 프레임워크를 구축함에 있어서도 윤리 규범의 기본적인 특징인 자율성이 사라지지 않도록 유의하여야 할 것이다.

이상의 기준에 따라 하나의 스펙트럼 안에서 윤리 책임과 법적 책임이 경계를 이루는 모습을 표현한다면 다음과 같은 구조를 떠올릴 수 있을 것이다.

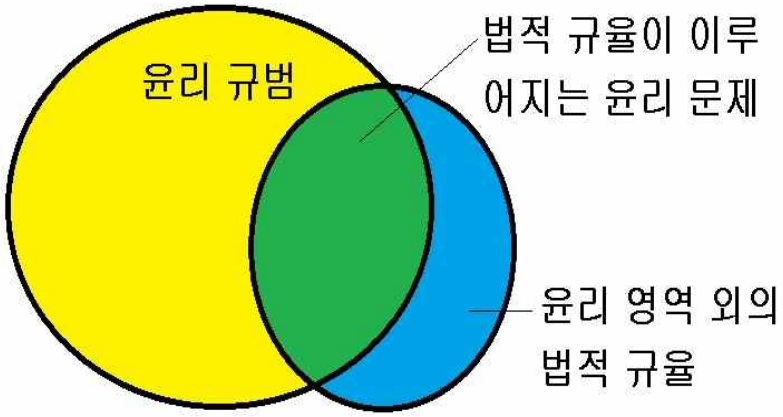
〔그림 5-1〕 윤리와 법의 경계

자료: Lyu, 2013: 408

이에 더하여 대응의 대상이 되는 행위 내지 쟁점에 대한 고찰 역시 필요하다. 즉, 이미 살펴본 대로 법규범은 구속성을 가지는 규범으로서, 법규범이 통제하는 대상은 입법자가 결정하는 사안에 제한된다. 반면, 윤리 규범은 개인에게 넓은 판단의 여지가 부여되는 사안들을 대상으로 한다. 예컨대 인공지능 기술의 오용으로 인해 인적·물적 피해가 발생했을 경우 입법자는 민사 책임법상의 규율을 통해 손해배상 책임과 관련된 법률관계를 규정한다. 또는 인공지능 기술이 범죄에 활용되었을 경우에는 형법 및 관계 형사법률이 행위자에 대한 처벌

규정을 두게 된다. 그밖에도 인공지능 기술이 공공의 이익에 위해를 가할 수 있다고 입법자가 판단하는 때에는 그 기술의 활용에 대하여 행정상의 제재를 유보할 수 있다(김동희, 2016: 35; 김유환, 2016: 17; 박균성, 2017: 15; 정하중, 2016: 29; 홍준형, 2018: 21; Maurer/Waldhoff, 2017: § 6, Rn. 11; Wolff/Bachof/Stober/Kluth, 2017: § 18, Rn. 14). 법적 규제가 이루어지는 범위는 최대한 좁게 해석되어야 하며, 국민의 대표인 의회(입법자)가 정한 제한적인 영역에 국한되어야 할 것이다. 윤리 문제는 이와 같은 위법성을 가지는 사안들을 포괄하기는 하지만, 입법자가 법률로 규율하겠다는 의지를 표명하지 않은 영역까지를 포함한 넓은 범주를 대상으로 한다. 따라서 윤리 문제와 법적 문제를 구분하고 양 규율 사이의 경계 설정을 고려할 때, 윤리 규범은 법 규범에 비해 훨씬 넓은 영역을 대상으로 한다는 점을 염두에 두어야 할 것이다. 물론 윤리적 문제에 속하지 않음에도 불구하고 입법자의 의사에 따라 법적 규율의 대상이 되는 영역 역시 존재한다. 윤리와 법의 관계를 그림으로 표현하면 다음과 같을 것이다.

〔그림 5-2〕 윤리 규범과 법 규범의 대상 관계도



개별적 행위에 대해 법적 규제를 가하는 것은 가능하지도 않거니와, 바람직하지도 않다.[233] 물론 법적 규율이 필요한 영역도 존재하지만, 이는 입법자의 결단에 따라 엄격하고도 명확하며, 예측 가능하게 규정되어야 할 것이다. 특히 아직 사회적으로 완전히 수용되지 않은 인

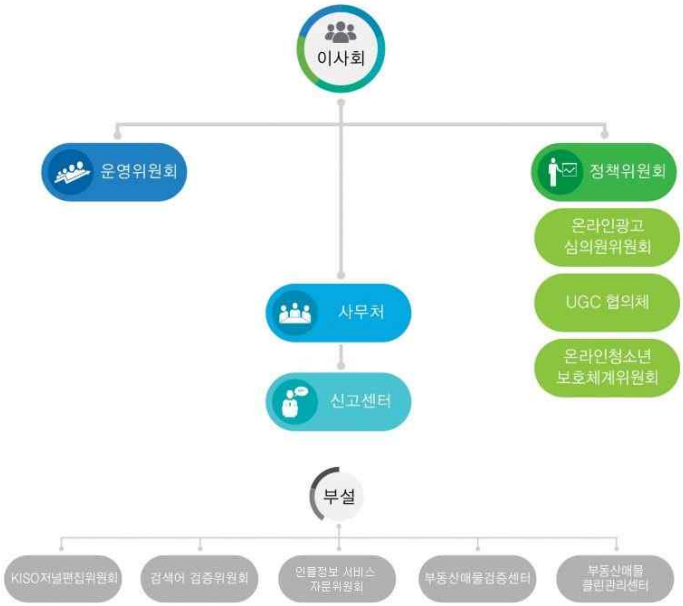
233) 특히 형법학의 관점에서 보더라도 "일반적인 윤리적 문제에 형법을 수단으로 투입하여 해결하는 것"은 국가의 임무가 아니다(이석배, 2008: 155).

공지능 기술의 적용 시 발생하는 문제에 대해 지나친 법적 규율을 가할 경우, 자칫 기술 개발을 위축시킬 수 있을 것이다. 따라서 윤리 대응의 프레임워크는 일차적으로 자율적으로 윤리 규범을 자율적으로 정립할 수 있도록 해야 하며, 이차적으로 그렇게 생성한 규범을 준수하고 관철하는 일 역시 자율적으로 행할 수 있도록 구성해야 할 것이다. 또한 법규범을 통한 통제는 최소화하되, 일정한 법적 규율이 필요한 경우에도 행정의 자의적 규제가 발생하지 않도록 법률을 통한 구체적이고도 명확한 수권을 전제로 삼아야 할 것이다.

〈표 5－1〉 최근 발간된 인공지능 관련 윤리 가이드라인

국가	가이드라인명	발간 주체	발간 시기
우리 나라	지능정보사회 윤리 가이드라인	정보문화 포럼 한국정보화진흥원	2018. 9.
	카카오 알고리즘 윤리 헌장	카카오	2018. 1.
독일	자동화 및 네트워크화된 자동차에 대한 윤리 규범 (EthischeRegeln fur den automatisierten und vernetzten Fahrzeugverkehr)	연방 교통 및 디지털인프라스트럭처부 윤리위원회(BMVI)	2017. 6.
미국	아실로마 인공지능 원칙 (Asilomar AI Principles)	Future of Life Institute	2017. 1.
	인공지능과 인권(Artificial Intelligence &Human Rights)	Berkman Klein Center(Havard Univ.)	2018. 9.
일본	전문가를 위한 인공지능 가이드라인	일본 인공지능 협회(JSAI)	2017. 5.
유럽연합	로봇 영역의 민사적 규율에 대한 집행위원회에 대한 권고	유럽 의회	2017. 2.
OECD	인공지능 윤리의 난제－기계에의 도덕성 구성을 위한 3개 가이드라인 (The Hard Problem of AI Ethics – Three Guidelines for Building Morality into Machines)	VyacheslavPolonski박사	2018. 2.

〔그림 5－3〕 국내 자율규제기구의 사례

국내에 존재하는 자율규제기구로서 인터넷상의 표현물에 대한 자율 규제를 표방하며 인터넷 플랫폼 사업자들이 조직한 “한국인터넷자율정책기구(KISO)”를 들 수 있다. KISO는 인터넷 사업자들이 이용자들의 표현의 자유를 신장하는 동시에 이용자들의 책임을 제고해 인터넷이 신뢰받는 정보 소통의 장이 될 수 있도록 하고, 인터넷 사업자들이 이용자 보호에 최선의 노력을 기울이는 등 사회적 책무를 다하는 한편, 공정하고 투명한 정책을 통해 인터넷 공간의 질서와 사회적 합의를 도출하기 위해 노력하는 것을 목표로 하고 있다. 이와 같은 목적을 위해 기구 강령 및 가이드라인을 수립하고, 회원사 등으로부터 요청 받은 인터넷 게시물 등의 여러 정책에 관한 사항을 다루는 한편, 국제 자율규제기구와의 교류협력 및 국제기구 활동에 참여하고, 정기적으로 자율규제백서를 발간하고 있다.
〈한국인터넷자율정책기구의 조직도〉


자료: 한국인터넷자율정책기구 국문 브로셔(2017)

3. 민관학 협력 모델

위에서 살펴본 대로 윤리대응 프레임워크는 행위자의 자율에 기반하지만, 그 내용의 형성 과정에서 각 영역의 시장 참여자와 공적 주체가 대등한 관계로서 협력 모델을 구축하는 것 역시 장점을 지닌다. 따라서 프레임워크 구축 시 이러한 민관협력 모델의 원용도 고려해 볼 수 있을 것이다.

민관협력의 모델은 현대 행정에서 행정 주체와 시민의 수직적인 관계를 수평적인 관계로 바꾸고 공적 과제 수행에 민간 주체가 참여하는 형태에서 비롯한다. 일반적으로 본래 공적 주체의 임무였던 과업을 민간 부문과 공공 부문이 공통의 목표로 설정하고, 특정 프로젝트

를 수행하기 위하여 계약 등의 방식을 통해 책임과 위험을 공유하며 장기적 관점에서 협력하는 제도를 민관협력이라 지칭한다(류현숙, 2017: 308; Seon, 2018: 29; Ziekow, 2006: 627). 즉, 공적 주체가 주도하는 영역에 대하여 사인이 참여하여 협력 관계를 이룸으로써, 참여와 협력의 민주주의 원리를 실현하는 한편, 사인의 창의적이고 효율적인 능력을 공적 과제에 활용하는 것이다. 윤리 대응 프레임워크에서는 반대로 민간 주도의 원칙 아래 공적 주체가 일정 부분을 기여하는 방식을 생각해 볼 수 있다. 즉, 보편적인 공적 과제 수행에서의 민관협력 모델을 반전하여 원용할 수 있다.

단, 전통적인 민관협력의 모델에서 사인이 보유하고 있는 능력을 통하여 과업의 수행을 더욱 효율적으로 만들 것을 요구하는 것처럼(Seon, 2018: 42), 윤리 대응의 프레임워크에서도 공적 주체의 참여와 그에 따른 협력을 통한 장점이 존재해야 할 것이다. 인공지능 관련 윤리 대응 문제는 사인이 주도권을 쥐는 것이 바람직한 영역이기는 하지만, 공적 주체가 일정 역할을 수행할 경우 여러 가지 장점을 취할 수 있다. 가령 시장에서의 행위자들은 다양한 이해관계의 대립 사이에 놓여 있어, 자율적 윤리 규범 형성을 위한 행위자들 사이의 합의를 유연하게 도출하는 데에 어려움이 있을 수 있으나, 공적 주체는 그러한 이해관계의 충돌에서 어느 정도 자유로운 지위에 있다는 점을 생각해 볼 수 있다. 이러한 이해관계 중립성은 공적 주체가 윤리 대응 프레임워크에의 참여를 통해 대립 관계에 있는 이해관계를 조정하는 것을 가능케 한다. 또한 일부 영역의 관련 정보(예컨대 기술 관련 정보, 윤리 규범의 국제동향 등)에의 접근에 있어서도 민간 부문에 비해 공적 주체가 유리한 입장에 있을 수 있다. 이러한 장점은 효율적이고 합리적인 윤리 대응 프레임워크 구축을 유리하게 할 수 있을 것이다.

따라서 윤리 대응의 프레임워크에서도 이를 원용하여 규범의 형성 단계 및 규범의 실효성 확보 단계에서 민간 행위자와 공적 주체가 협력하는 모델을 고려해 볼 수 있을 것이다. 구체적으로는 윤리 규범 구축을 위한 사회적 합의 단계에서 공사 협력 형태의 협의체를 구성하는 방안, 기술 개발 단계에서 윤리적 표준에 맞는 기술 개발을 위한 지원체를 구성하는 방안 및 인공지능 기술 사용 단계에서 윤리 규범 준수에 대한 공사 협력 형태의 감시체 구축 방안 등을 생각해 볼 수 있다.[234] 또한 나아가 학계의 연구자까지 포함하는 민관학 협력 형태의 플

234) 인공지능 윤리 분야에서 이러한 방식의 민관협력 플랫폼으로는 독일 정부가 발족을 예정하고 있는 "인공지능 전략을 위한 온라인 컨설팅"의 예를 참조할 수 있다. 2018년 12월 정식 발족 예정인 이 플랫폼에는 연방 정부의 부처 중 노동사회부, 에너지경제부 및 교육연구부가 참여하였으며, 각 기업, 연구소 및 시민단체들이 참여하여 인공지능 전략에 관한 공동의 대응을 해 나갈 예정이다. 상세

랫폼 역시 고려해 볼 수 있다.

4. 서비스의 이용자에 대한 고려

인공지능 기술이 분야별로 사회적으로 수용되어 적용된 후의 이용자 보호와 관련한 쟁점에 대해서는 최근 많은 논의가 진전되어 왔다(이병준, 2017: 119; 이정구/민대환/권헌영, 2017: 81; 한문승/이주헌, 2014: 337). 특히, 과거의 이용자 정책이 공정경쟁 측면에서의 시장 규제를 통한 이용자 보호에 치중되어 왔다면, 서비스 기반 기술이 고도화되고 복잡해진 현재에는 지능정보산업 영역에서 독자적인 이용자 보호가 필요하다는 점이 지적되고 있다(최경진, 2017: 15). 인공지능 기술이 수용되고 실제 사회에 적용될 경우에도, 기존의 재래 산업에서의 이용자 보호와는 다르게 독자적이면서도, 특히 자율적인 형태를 갖추고 있는 이용자 보호의 수단을 마련해야 할 것이다. 윤리 대응의 측면에서도 인공지능 기술을 통해 구현되는 서비스를 이용하는 이용자 보호의 관점을 충분히 고려해야 할 것이다.

서비스의 객체라는 관점에서 이용자를 보호의 대상으로 바라볼 수도 있지만, 인공지능 기술이 구현된 영역에서 적극적으로 인공지능을 활용하는 주체로도 생각할 수 있다. 따라서 단순히 이용자에 대한 보호만을 강구하는 프레임워크가 아니라, 이용자가 윤리 수범자의 한 축으로서 적극적으로 참여할 수 있는 모델 역시 고려해야 할 것이다. 이는 일종의 디지털 시민권(digitale Bürgerschaft)의 개념과 비교해서도 설명할 수 있을 것이다. 디지털 시민권은 온라인을 바탕으로 한 기술 발전 사회에서 온라인 기술을 기반으로 능동적으로 참여하는 시민을 의미하므로(김만권, 2014: 76; 최윤국, 2003: 301), 인공지능 기술이 적용되는 사회에서도 이러한 개념을 적용할 수 있다. 즉, 인공지능 기술을 이용하는 시민이 윤리 문제에 있어서도 적극적인 수범자가 되는 것이다.

두 가지 측면의 이용자 보호 및 참여가 보장되기 위해서는 윤리 규범의 수립, 인공지능 기술의 개발과 기술의 사회적 영향에 대한 평가 및 실제 기술 적용 과정에서 이용자의 의견을 반영할 수 있는 절차를 마련해야 할 것이다. 이용자의 참여 모델과 관련해서는 독일 대형 공공사업 추진 시 사업 계획에서 구체적 수행 단계에 이르기까지 시설의 이용자에 해당하는

한 내용은 다음 인터넷 사이트를 참조하라. 〈https://www.konsultation-ki.de/KI/Redaktion/DE/Artikel/fortschritt-durch-kuenstliche-intelligenz.html〉 (2018. 10. 26. 최종 방문)

주민들이 적극적인 이니셔티브를 갖고 참여하고 있는 모델을 참고해 볼 수 있을 것이다(선지원, 2018: 142, 144). 인공지능 기술의 수용 과정에서도 윤리 규범의 수립, 인공지능 기술의 개발과 기술의 사회적 영향에 대한 평가 및 실제 기술 적용 과정에서 이용자의 의견을 적극 반영할 수 있는 절차를 강구해 볼 수 있다.

〔그림 5-4〕 독일 공공사업 계획 및 추진단계에서의 주민 참여 모델

<table>
<tr><td rowspan="8">계획 단계</td><td rowspan="3">수요 계획
(Bedarfsplanung)</td><td>연방교통망계획(Bundesverkehrswegeplan) 준비 및 프로젝트 신청</td><td></td></tr>
<tr><td>연방교통망계획의 수립</td><td>– 환경적합성평가에 수반한 정식 참여 절차 강제</td></tr>
<tr><td>수요 계획 및 건설될 노선의 수립</td><td></td></tr>
<tr><td colspan="3">↓</td></tr>
<tr><td rowspan="2">공간이용계획
(Raumordnung)</td><td>사전조사 및 공간이용 기초자료의 작성</td><td></td></tr>
<tr><td>공간이용계획절차</td><td>– 대부분의 경우 공간이용계획행정청 주도의 정식 참여 절차 시행</td></tr>
<tr><td colspan="3">↓</td></tr>
<tr><td colspan="2">노선 결정(Linienbestimmung)</td><td>– 경우에 따라 환경적합성평가에 수반한 정식 참여 절차 시행</td></tr>
<tr><td colspan="3">↓</td></tr>
<tr><td rowspan="7">사업자 선정 단계</td><td colspan="2">경쟁참여자 모집 공고
(Teilnahmewettbewerb)</td><td>☜ 예비타당성조사</td></tr>
<tr><td colspan="3">↓</td></tr>
<tr><td colspan="3">조건 제시 단계(Angebotsphase)</td></tr>
<tr><td colspan="3">↓</td></tr>
<tr><td colspan="3">협상 단계(Verhandlungsphase)</td></tr>
<tr><td colspan="3">↓</td></tr>
<tr><td colspan="2">실시협약 체결(Vertragsabschluss)</td><td>☜ 타당성 조사</td></tr>
<tr><td colspan="4">↓</td></tr>
<tr><td colspan="2" rowspan="2">계획 확정 및 허가
(Zulassung)</td><td>초안 작성, 허가 계획, 계획확정절차 기초자료의 작성</td><td>경우에 따라 조기 공공참여</td></tr>
<tr><td>계획확정절차
(Planfeststellungsverfahren)</td><td>– 청문행정청이 주도하는 정식 참여 절차 강제</td></tr>
</table>

자료: 선지원, 2018: 142.

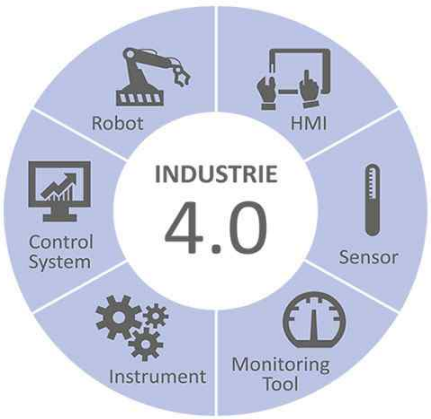
5. 인공지능 기술 발전 속도 고려

인공지능 기술의 적용과 관련한 윤리는 인공지능 기술의 발전 속도에 발맞추어 정립되는 것이 바람직할 것이다. 인공지능 기술이 미처 실현되지 않은 상황에서의 윤리 규범 통제 논의는 선부를 뿐만 아니라, 기술 개발을 위축시킬 수 있는 위험성을 지닌다.

따라서 프레임워크의 구축 시에도 인공지능 기술의 개발 정도를 고려할 필요가 있다. 이를 위해 먼저 윤리 규범의 구축 단계에 참여하는 주체들이 인공지능 기술의 발전 정도를 이해하는 일이 선행되어야 한다. 또한 개발과 실험 단계에서 예측하는 리스크의 범위를 기술의 개발 범위 내로 한정시키는 것도 기술 개발 정도를 넘어서는 지나친 윤리 규율을 방지할 수 있는 방법일 것이다.

프레임워크 구축 시 기술 발전 속도를 고려한 예로 독일 정부가 ICT 기술의 발전에 상응하도록 제조업의 구조를 개선하기 위한 목적으로 발표한 이른바 "산업 4.0(Industrie 4.0)"[235] 을 들 수 있다. 산업 4.0 정책결정을 위한 논의 구조를 비롯하여 정책 결정 시 나타나는 문제들을 해결하기 위하여 부수적으로 구축한 플랫폼들을 참조할 수 있을 것이다. 더구나 산업 4.0은 경제 구조의 디지털화를 통한 사회 구조 변화를 정면으로 다루고 있는 사례이므로, 윤리 대응의 문제를 넘어서서 인공지능 기술의 사회적 수용이라는 거시적인 구조의 거버넌스를 마련할 때에도 참고할 수 있는 정책 사례일 것이다.

〔그림 5－5〕 독일 Industrie 4.0 표현도



자료: http://www.piovan.com/en/winfactory4_0

235) 산업 4.0의 주요 내용에 대해서는 다음 웹페이지를 참조하라. Plattform Industrie 4.0 – Industrie 4.0 〈https://www.plattform-i40.de/I40/Navigation/EN/Industrie40/industrie40.html〉 (2018. 9. 30. 최종 접속)

6. 사전 예방과 사후 대응의 조화

자율 윤리 대응은 기본적으로 시장 행위자가 선제적으로 발생할 수 있는 윤리 쟁점을 파악하여 그에 따른 문제를 방지함으로써 사업 시행 과정에서의 리스크를 방지한다는 의미를 가진다. 이러한 측면에서 자율 윤리 대응 프레임워크는 사전 예방의 성격을 강하게 띤다고 할 수 있다.

사전 예방이란 위험이 실제로 나타나지 않은 경우에도 "위험의 잠재적 결과가 심각하고 돌이킬 수 없는 경우"에는 위험 방지를 위한 조치를 미리 취해야 한다는 관념이다(류현숙, 2017: 40). 1970년대 이후 각종 환경, 기후, 식품 등의 문제를 다루는 국제 조약에서 사전 예방의 원칙을 원용하고 있다(류현숙, 2017: 41). 이는 미래의 위험을 예측하고 모니터링하는 한편, 위험의 회피를 위하여 각종 시스템을 구축하거나 사전에 위험을 야기하는 작용을 중지시키는 등의 조치로 나타날 수 있다(김은성, 2010; 류현숙, 2017).

인공지능과 관련해서는 기술의 상용화 이전 단계에서 모니터링을 통해 리스크를 평가하여 그러한 리스크가 수인 가능한 것인지 판단하는 방식으로 사전 예방의 원칙을 관철할 수 있을 것이다(이른바 "Responsible Risk and Innovation"). 또한 개발자 외에 일반적인 상식을 가진 비전문가가 모니터링에 참여함으로써 불확실성과 파급 효과에 대한 선제적 대응을 하는 것이 가능할 것이다. 이미 언급한 디지털 시민권의 관점에서 이용자 역시 인공지능 기술의 이용 윤리에 대해 사전 준비하는 일이 필요하다.

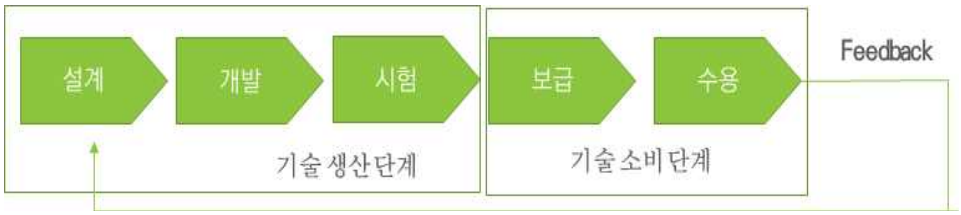
그러나 프레임워크를 실제로 실행시키는 과정에서 프레임워크 내에서도 예측하지 못했던 위해가 발생할 가능성도 존재한다. 따라서 이런 경우에 대비하여 사후적으로 문제를 시정할 수 있는 구조를 마련해 두는 일도 필요하다. 기술의 발전 정도 등을 고려하여, 사전 예방이 중요한 분야와 사후 대응이 중요한 분야를 분리하여 대응하는 방안도 고려해 볼 수 있을 것이다.

7. 인공지능 기술의 생애주기에 따른 윤리 이슈 고려

특정 기술의 적용에 따른 리스크를 효율적으로 관리하는 방법의 하나로, 해당 기술을 생애주기별로 절차화하여 각 단계별로 대응을 해 나가는 방식을 생각해 볼 수 있다. 인공지능 영역에서도 이와 같은 절차별 대응 방법론은 의미를 가질 수 있다. 일반적으로 인공지능 기

술의 수용은 기술의 생산과 소비의 두 단계로 나눌 수 있을 것이다. 윤리는 이 두 단계에서 모두 의미를 가진다. 산업 영역에 따라 차이가 있을 수 있지만, 먼저 기술 생산 단계는 설계, 개발, 시험의 단계로 세분할 수 있을 것이며, 기술 소비 단계는 보급과 수용의 두 단계로 다시 나눌 수 있을 것이다. 인공지능 기술의 수용 과정을 그림으로 표현하자면 다음과 같다.

〔그림 5-6〕 인공지능 기술 수용 과정도



이러한 각 단계별 윤리 이슈를 인식하고, 해당 단계에 적합한 대응 방안을 고려한다면 효율적이고 합리적인 대응이 가능할 것이다.

8. 프레임워크의 범용성

윤리 대응을 위해 구축하는 프레임워크는 다양한 영역의 인공지능 관련 쟁점에 대응할 수 있도록 보편성 및 포괄성을 가지는 한편, 유연성 또한 갖추어야 한다.

특정한 윤리 이론만으로 전체를 아우르는 윤리규범이나 대원칙을 설명하는 것은 복잡한 고도의 발전 사회인 현대사회, 특히 지능정보화 사회에서 사안에 따른 대처를 어렵게 할 위험이 있다. 즉, 인공지능 윤리와 관련해서 사안별로 쟁점에 대한 논의를 가능하게 하는 합리적이고 범용적인 구조를 만드는 것이 필요하다. 이를 위해 기존의 법규범과는 다른 포스트모더니즘적 거버넌스를 구축할 필요가 있는 것이다. 이런 의미에서 윤리 대응의 프레임워크는 규범의 내용 자체보다 규범을 관철할 수 있도록 만드는 자연스러운 논의의 장을 마련하는 데에 역점을 두어야 할 것이다. 물론 이러한 보편적 프레임워크를 실제 사례에 적용할 때에는 각 산업 영역별 특수성을 고려할 수 있어야 한다.

프레임워크의 범용성은 산업 분야별로 의미를 가질 뿐만 아니라, 윤리 규범을 수용하는 행위자에 대해서도 적용되어야 할 것이다. 즉, 관련 정책을 수립하고 집행하는 공적 주체, 인공지능 기술을 개발하고 보급하는 개발자 및 사업자를 비롯하여, 인공지능 기술을 직접

이용하는 일반 시민에 이르기까지, 인공 지능 기술의 사회적 수용 단계에 개입하는 모든 행위자를 포괄할 수 있는 프레임워크를 구성해야 할 것이다. 이들 모두 윤리 규범을 구성할 때 사회적 가치를 제공하는 참여자이자, 형성된 규범의 수범자이기 때문이다.

프레임워크의 범용성을 추구하기 위해 먼저 사전 단계에서 보편적 가치를 담은 윤리 규범을 구축하도록 노력하는 일이 필요하다. 또한 예기치 못한 윤리적 쟁점이 발생했을 때, 논의 플랫폼이 구축되도록 하는 구조를 만드는 것이 중요하다. 그러한 논의 플랫폼은 국가 주도의 위원회나 협의체가 아닌, 지능정보사회 구성원 모두가 협력적 방식으로 참여하는 자연스러운 형태가 되는 것이 바람직할 것이다.

그러나 필요한 최소한의 경우에는 그러한 자율 협의체가 작동될 수 있도록 지원하는 법적 장치를 마련하는 것도 고려할 수 있다.

제 3 절 종합적 자율 윤리 대응 프레임워크의 구축

1. 원칙

범용적인 자율 윤리 대응의 프레임워크는 종과 횡의 두 가지 방향으로 구축해 볼 수 있다. 먼저 종으로는 인공지능 기술이 준비되고, 개발되며, 실험을 거쳐 현실에 적용되기까지의 과정을 생각해 볼 수 있다. 횡으로는 두 가지 차원의 구조를 만들어 낼 수 있을 것이다. 첫 번째로 인공지능 기술의 설계, 개발 및 활용에 관계하는 당사자－사업자 내지 개발자, 공적 주체 및 일반 이용자－별로 종으로 만들어진 각각의 과정에 대응시킬 수 있을 것이다. 두 번째로는 행위자가 아닌 인공지능 기술의 개발과 적용 과정에 나타나는 쟁점들을 횡축에 대응시킬 수 있을 것이다.

이 때 위에서 살펴본 고려 요소들을 통해 도출된 프레임워크 구축의 지침들을 적용할 수 있을 것이다. 즉, 법 규범과 윤리 규범의 구별이라는 관점에서 각각의 레이어별로 법적 규제를 행해야 하는 경우(입법이 필요한 경우), 일정한 가이드라인이 필요한 경우 및 자율 통제만으로 충분한 경우를 나누어 볼 수 있다. 이 때, 법적 규제가 필요한 영역은 입법자의 결단에 따른 엄격한 수권을 요하며, 가이드라인의 경우에도 실질적인 강제규범화가 이루어지지 않도록 하는 배려가 필요하다. 또한 각각의 레이어에서 협의체의 구축이 필요한 경우을 제

시한다. 인공지능으로 인한 위해가 발생하는 경우를 대비하여, 윤리 차원을 넘어선 피해보상 체계가 필요한 부분과 사회적 리스크 분배 체계가 필요한 부분을 나누어 고려한다. 이 때 사회적 리스크 분배라 함은 인공지능 기술이 완전히 사회적으로 수용되어, 그 기술의 향유자인 사회 공동체가 그로 인한 리스크 역시 부담한다는 의미이다. 인공지능의 오작동 등에 대하여 기술적 지원을 제공할 주변 기구를 비롯하여, 보편적 프레임워크로부터 쟁점의 특수성 고려를 가능케 할 논의 플랫폼 역시 적절한 위치에 마련해야 할 것이다.

이하에서는 이러한 구조를 바탕으로 각각의 인공지능 적용 절차와 영역별 행위자 및 쟁점별로 레이어를 구축한 후, 프레임워크상 횡축을 두 가지 차원으로 나누어 논의를 진행하고자 한다.

2. 행위자별 자율 윤리 대응의 구조

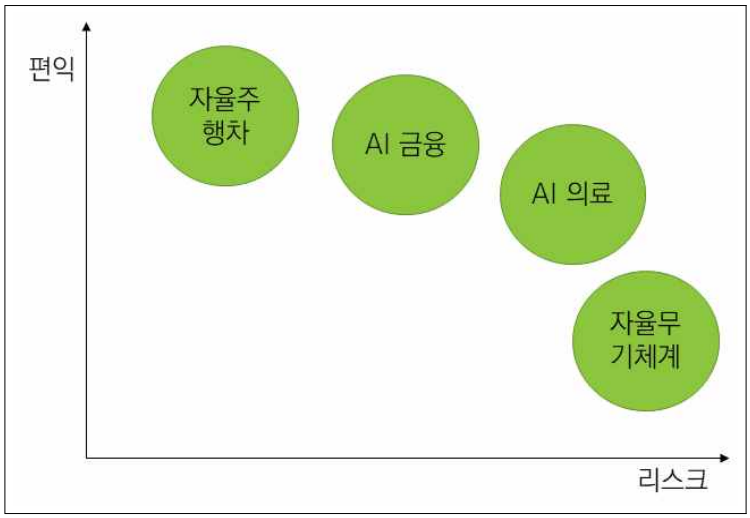
먼저 행위자별로 인공지능 기술 적용 과정에서의 자율 윤리 대응의 역할을 나누어 그 프레임워크를 구축해 보기로 한다.

먼저 인공지능 기술을 개발하여 사회적으로 수용하고 실제로 사용하기까지의 절차를 도식화할 필요가 있을 것이다. 도식화된 절차에 따라 사업자/개발자, 공적 주체 및 이용자에 각각의 역할을 부여한다. 개발 이전의 단계에서는 사회 구성원들의 합의에 기반한 사회적 가치를 정립해야 하고, 이를 바탕으로 한 인공지능 기술 관련 자율적 윤리 규범을 정립할 필요가 있다. 이 과정은 제조자/공급자, 공적 주체 및 이용자가 모두 참여하는 공론화 방식을 통해 이루어져야 할 것이다. 또한 이러한 윤리 규범은 향후의 인공지능 개발 및 적용 단계에서 지속적으로 준수될 수 있는 내용이어야 할 것이다(Responsible Risk and Innovation). 법적 규율이 필요한 내용에 대해서는 입법자의 명확하고 구체적인 규정이 이루어져야 한다. 설계 단계에서는 제조자/공급자의 자율적인 영향 평가와 기술에 대한 윤리적 리스크 진단이 필요하다. 이 때 공적 주체는 영향 평가의 요령 등을 기술적 지침 등의 형태로 제공할 수 있을 것이다. 또한 영향평가의 과정에서 일반 이용자의 의견을 수렴하는 절차를 둘 수 있을 것이다. 임상단계가 필요한 경우에도 개발자는 윤리 규범을 준수해야 하며, 공적 주체와 이용자는 임상과정에 의견을 제시할 수 있을 것이다. 인공지능 기술의 실제 사용 단계에서는 경우에 따라 법적으로 규율된 논의 플랫폼이 구축되어 있어야 한다. 이러한 논의 플랫폼에는 비전문가 역

시 참여하여 일반적으로 통용되는 사회적 가치가 논의에 반영될 수 있도록 해야 할 것이다. 제조자 내지 공급자는 윤리 규범을 준수하되, 윤리 규범을 넘어서는 물적인 위해에 대해서는 피해 구제를 행해야 한다. 공적 주체는 명백한 위법 행위에 대해서는 법적·행정적 규제를 행하되, 필요한 경우 기술 표준, 위험 인증 등을 제공하는 등 인공지능 기술에 대한 지원을 행할 수 있다. 다만, 일반적인 윤리 기준에 대해서는 간섭을 자제해야 할 것이다.

인공지능 기술의 적용이 높은 사회경제적 후생 효과를 가지는 경우, 그에 따른 리스크 역시 사회적으로 수용하는 것을 합의할 수 있다. 특히 이용자들의 합의가 이루어진 경우, 인공지능 사용에 따른 리스크는 사회 공동체가 모두 부담하게 되며, 이러한 상태가 인공지능 기술이 완벽하게 사회적으로 수용된 경우일 것이다. 이는 인공지능 기술의 수용으로 인한 사회적 편익과 리스크의 관계를 고려함으로써 판단할 수 있다.

〔그림 5-7〕 편익-리스크 비교 그래프(예시)



리스크에 비해 사회적 편익이 높은 것으로 사회적 합의가 이루어진다면, 해당 기술은 사회적으로 수용된 것이라고 볼 수 있을 것이다. 그러한 기술의 경우 발생하는 리스크를 사회 공동체가 흡수하려는 노력이 필요하다.

지금까지 살펴본 행위자별 자율 윤리 대응의 프레임워크를 그림으로 정리하면 다음과 같다.

〔그림 5－8〕 행위자별 자율 윤리 대응 프레임워크

<table>
<tr><th></th><th></th><th>개발자/공급자</th><th>공적 주체</th><th>이용자</th></tr>
<tr><td>개발
이전 단계</td><td>☞</td><td colspan="3">• 윤리 규범 구성의 기반이 될 사회적 가치에 대한 합의 / 사회적 협의체
→ 윤리 / 리스크 커뮤니케이션
• 사회적 규범을 바탕으로 한 자율 윤리 규범 정리 / 법적 규율 경계 고려
• 연구윤리(RRI) 고려 / 윤리 교육 실행</td></tr>
<tr><td>설계 단계</td><td>☞</td><td>• 영향평가
• 자가리스크 진단</td><td>• 영향평가의 요령 제공</td><td>• 영향평가 단계에서 의견 제시</td></tr>
<tr><td>개발 단계</td><td>☞</td><td>• 제조 시 자율 윤리 규범 준수</td><td colspan="2"></td></tr>
<tr><td>시험 단계</td><td>☞</td><td>• 윤리 규범 준수</td><td>• 안전기준 구축
• 리스크 평가 / 인증</td><td>• 의견 제시</td></tr>
<tr><td rowspan="3">소비 단계</td><td rowspan="3">☞</td><td colspan="3">• 논의 플랫폼의 구축 / 민관협력의 사회적 협의체</td></tr>
<tr><td rowspan="2">• 피해 구제</td><td rowspan="2">• 위법 행위에 대한 제재
• 오작동에 대비한 기술적 지원 체계</td><td>• 신기술 리스크의 사회적 수용 → 기술의 완전한 수용</td></tr>
<tr><td>사회적 협의체 필요</td></tr>
<tr><td>피드백 단계</td><td>☞</td><td colspan="3">• 소비 단계에서 형성 및 축적된 경험을 신규 기술의 도입 시 반영</td></tr>
</table>

3. 쟁점별 자율 윤리 대응의 구조

자율 윤리 대응의 구조는 행위자별로 나눌 수도 있지만, 인공지능 기술의 적용 시에 제기될 수 있는 주요 쟁점별로 구분할 수 있다. 인공지능 기술에 따른 윤리적 쟁점은 여러 가지를 설정할 수 있으나, 본 연구에서는 제4장까지의 내용에서 주로 다루어진 알고리즘 책임성, 설명 가능성 및 공정거래의 쟁점을 채택하여 프레임워크를 예시하고자 한다. 인공지능 기술의 개발자 내지 사업자는 각각의 쟁점별로 자율 윤리 대응의 프레임워크 안에서 행위하게 된다. 행위자별 자율 윤리 대응 구조와 마찬가지로 먼저 인공지능 기술의 적용 단계를 구축한 후, 각각의 쟁점별로 그에 상응하는 윤리 대응 프레임워크를 제시하고자 한다.

개발 이전의 단계에서는 행위자별 윤리 대응 프레임워크와 마찬가지로, 구성원 간의 합의가 이루어진 사회적 가치에 기반하여 각 윤리 쟁점을 포괄하는 윤리 규범을 정립하게 된다. 이 때 기본권 침해를 포함하여 명백한 권리 침해가 이루어지는 경우는 법적 규율의 대상이 된다. 설계에서 임상에 이르는 단계에서는 개발자 및 공급자의 자율적인 영향 평가 내지 리스크 진단이 이루어져야 하며, 이 때 자율적으로 정립한 윤리 규범이 적용된다. 사용 단계에서는 행위자별 프레임워크와 마찬가지로 논의 플랫폼의 구축이 필요하며, 기술 리스크를 사

회적으로 수용할 것인지의 여부도 사회적 합의에 기반하여 결정하게 된다. 그밖의 사후 윤리 문제에 대해서는 개발자 및 공급자가 자율적인 피해 구제 노력을 기울이는 한편, 법적 피해에 대해서는 법적 책임을 진다.

이와 같은 내용을 고려하여 쟁점별 윤리 대응 프레임워크를 표현하면, 다음 그림과 같다.

〔그림 5-9〕 쟁점별 자율 윤리 대응 프레임워크

<table>
<tr><th></th><th></th><th colspan="2">알고리즘 책임성</th><th colspan="2">설명 가능성</th><th colspan="2">공정거래 쟁점</th></tr>
<tr><td rowspan="2">개발 이전 단계</td><td rowspan="2">☞</td><td rowspan="2">알고리즘 규제 알고리즘 윤리 구축</td><td rowspan="2">규제 필요성이 인정되면 법적 규제</td><td rowspan="4">설명 가능한 알고리즘 설계</td><td rowspan="4">윤리 강령 필요</td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td>설계 단계</td><td>☞</td><td rowspan="2">알고리즘 윤리의 준수</td><td rowspan="2">자율 대응</td><td></td><td></td></tr>
<tr><td>개발 단계</td><td>☞</td><td>공정한 보조금</td><td>법적 규제</td></tr>
<tr><td>시험 단계</td><td>☞</td><td>영향 평가</td><td>영향 평가 기준 및 기술 표준</td><td>알고리즘의 설명 가능성 검사</td><td>자율 대응</td><td></td><td></td></tr>
<tr><td>소비 단계</td><td>☞</td><td colspan="2"></td><td>이용자에 대한 설명</td><td>자율 대응 이용자보호 가이드라인</td><td>이용자 간 차별 시정</td><td>자율 대응 경우에 따라 법적 규제</td></tr>
<tr><td rowspan="2">피드백 단계</td><td rowspan="2">☞</td><td colspan="6">논의 플랫폼을 통해 형성 및 축적된 경험을 신규 기술 도입 시 반영</td></tr>
<tr><td colspan="6">민관학 협의체의 구성 및 기능</td></tr>
</table>

제 4 절 개별 산업 영역에의 적용

1. 개관

범용적인 프레임워크는 개별 사례에서는 특히 논의 플랫폼을 통하여 각 인공지능 활용 영역의 기술적 특성 및 해당 산업 분야의 특성 고려하는 방향으로 발전할 수 있을 것이다. 분야별, 특성별로 강조점이 달라지게 하는 구조는 위에서 살펴본 프레임워크상의 논의 플랫폼을 통해 가능해질 것이다.

논의 플랫폼은 개별 산업의 특성을 고려하는 역할을 수행하는 한편, 반대로 개별 인공지능 적용 분야에서 발생한 쟁점이 보편적인 것이라고 판단하는 경우에는 프레임워크에 대한

수정을 가능케 하는 기능을 수행할 수도 있다. 즉, 프레임워크는 사전적인 배려를 기초로 하지만, 사후의 논의 플랫폼을 통해 사전 통제의 구조에 일정한 피드백을 제공할 수도 있는 것이다. 또한 이러한 산업 분야별 프레임워크는 인공지능 기술의 개발 및 상용화 단계별, 행위자별, 쟁점별 체크리스트의 기능도 수행할 수 있을 것이다.

이하에서는 제4장의에서 구분한 인공지능 기술 적용의 분야별로 고려해야 할 프레임워크상의 특수성에 대해 살펴보고자 한다.

2. 제조 분야

가. 논의의 출발점

제조 분야에서 강조될 수 있는 윤리 쟁점은 알고리즘 설계 단계에서의 윤리 기준 준수 및 사후 인공지능 오류로 인한 피해 발생 시의 책임 문제일 것이다.

전자에 대해서는 개발자 스스로 윤리 규범 정립에 참여함으로써 윤리 규범에서 벗어나지 않는 알고리즘을 구축하려는 노력을 하는 것이 중요하다. 보편적 윤리 대응의 프레임워크에 이와 같은 알고리즘 개발자의 참여를 강조하는 일이 필요할 것이다. 후자의 경우 일반적으로는 기존의 민사법적 관점에서의 접근이 필요할 것이나, 책임 문제와 관련하여 윤리적 성격이 더 짙을 때는 민사법적 해결이 적절치 않을 수 있다. 이와 같은 문제를 개발자가 책임질 것인지 리스크를 사회적으로 수용할 것인지, 혹은 법적 규율의 테두리 안에 넣을 것인지 자율 통제의 영역에 남겨둘 것인지에 대해 논의 플랫폼상의 협의 절차를 거칠 수 있을 것이다.

이하에서 제조 분야에 인공지능을 적용할 때의 윤리 문제에 대한 정책 대응 프레임워크를 구성함에 있어 특히 고려해야 할 사항에 대해 살펴본 후, 최종적으로 프레임워크의 예시를 제시하고자 한다.

나. 윤리의 규범적 속성과 절차주의적 거버넌스 및 기관윤리위원회의 역할(정채연, 2017: 37-40)

먼저 자율주행자동차 관련 윤리적 쟁점들과 관련된 실질적인 내용을 (윤리헌장 등의 형태를 가지는) 가이드라인을 통해 제시하고자 하는 시도의 적합성을 고찰해 볼 필요가 있다. 해당 가이드라인이 최소한의 윤리적 기준이며, 법적 책임을 판단하는데 있어서 유일한 근거가

될 수는 없다는 점을 명시한다고 할지라도, 가장 핵심적인 윤리적 지침으로서 위치할 가능성이 높다.

지능정보사회와 관련된 많은 법제정 시도에서 특히 중앙행정기관 내의 위원회가 윤리와 관련하여 주도적인 역할을 수행하도록 설계되어 있다. 기존의 지능형 로봇윤리헌장 역시 정부가 지능형 로봇 개발자・제조자 및 사용자가 지켜야 할 윤리로 제정하여 공표하는 것이다(지능형 로봇 개발 및 촉진법 제18조 제1항).

이는 윤리의 규범적 속성을 고려할 때 원칙론상으로는 적합하지 않다고 판단된다. 자율주행차를 비롯한 제조 분야에서 보다 구체적인 윤리적 실질을 채우는 역할은 기관윤리위원회에서 수행하는 것이 바람직할 것으로 보인다. 이러한 맥락에서 로봇윤리와 관련해 정부주도의 위원회가 갖는 권한이 커지게 되면, 자칫 다원주의적 절차주의 및 통합적인 거버넌스 구상이 와해될 수도 있다는 점을 상기해야 할 것이며, 지능정보기술에 대한 과도한 규제를 우려하는 공론을 고려할 때에도 기관윤리위원회의 실질적인 역할에 대한 적극적인 관심이 필요할 것이다.

자율주행자동차의 윤리적 쟁점을 다루는 데 있어서 전반적으로 탈중앙화된 거버넌스로의 방향성을 가져야 할 것이다. 탈중앙화된 거버넌스에서는 중앙정부가 직접적으로 통제하는 역할을 수행하기 보다는 관련 이해당사자들의 주도적인 참여에 대한 동기를 부여하고 모니터링하는 간접적인 조정 역할을 담당하게 된다.

절차주의의 맥락에서 기관윤리위원회의 의의 및 역할이 중요하다는 점을 고려할 때, 기관윤리위원회 설치를 실제 보편화할 수 있도록 하는 제도적 뒷받침이 필요할 것으로 보인다. 로봇윤리의 실질적인 지침을 마련하기 위해서는, 실제 지능정보기술의 연구・개발・활용이 이루어지는 현장 및 실무적 차원에서 윤리적인 가치판단의 결과들을 축적하는 것이 매우 중요하다. 위에서 언급한 윤리의 규범적 속성을 고려할 때, 로봇공학 분야 및 쟁점을 규율하는 데 있어서 다원적이고 탈중앙화된 절차주의가 기본적인 정책 방향성으로 고려될 수밖에 없으며, 이때 실질적인 윤리지침을 마련하는 역할을 하는 기관윤리위원회는 이렇듯 절차주의적 맥락에서 정당성을 갖게 된다.

기관윤리위원회의 설치가 해당 기관의 연구 및 개발 활동에 대한 강력한 규제정책의 실시를 의미하는 것은 아닐 수 있다. 절차주의적 방향성은 기본적으로 위원회의 운영 및 구성 등 절차와 관련된 제반사항을 제시하고, 윤리적 실질은 해당 위원회의 자율적인 판단역량에 맡

겨 놓으며 그 결정에 법적 정당성을 부여하는 것이다.

특히 로봇윤리와 생명윤리가 갖는 공통적인 속성을 고려하여 생명윤리법상의 기관생명윤리위원회 관련 규정과 비교・검토하는 것이 유의미할 수 있을 것이다. 즉, 인간의 존엄성과 과학기술개발 및 연구의 자율성이 상충되는 경우에 이를 조화롭게 해결하기 위한 규제 방향성을 제시한다는 점에서 생명윤리법과 상관성을 갖는다고 할 수 있다.

생명윤리 및 안전에 관한 법률 제10조는 생명윤리 및 안전을 확보하기 위하여 인간대상연구를 수행하는 자가 소속된 교육 연구 기관 또는 병원 등은 기관생명윤리위원회를 설치하여야 한다고 규정하고 있다. 또한 생명윤리법은 기관생명윤리위원회의 구성 및 운영과 관련된 절차상의 원칙에 대해서도 규정하고 있다(생명윤리법 제11조).

따라서 특히 절차주의의 맥락에서 기관윤리위원회 내적 구성의 이질성 및 다양성을 담보하는 일이 요청된다.

나. 사회적 수용성 관련 의사소통 절차에서 시민의 자발적 참여 보장

자율주행자동차의 윤리적 프로그램이 상용화의 사전적 단계에서 이미 요청되며 이때 시민사회의 사회적 수용성이 기본전제로서 확보되어야 함은 물론이다. 이때의 사회적 수용성이란 당시의 객관적 판단 준거를 바탕으로 보편적 합의를 추구하되, 다만 그 합의의 내용이 지속적으로 변화 가능성에 열려 있는 것이다. 특히 시민의 자발적 선택권이 충분히 확보되고 보장되었는지 여부가 사회적 수용성을 검토하는 데 있어서 핵심적인 기초가 되어야 할 것이다.

이 때 윤리적 프로그래밍의 실질적인 내용을 다원적 행위자들의 의견을 수렴할 수 있는 절차를 통해 규정하는 것이 절차주의적 거버넌스라고 할 수 있다. 이를 위해서는 운행자, 보행자, 소비자, 제조사, 개발자, 도로교통 관련 공적 기구 등 다양한 주체들이 참여하는 의사소통 절차가 보장되어야 한다. 이는 기관윤리위원회 뿐만 아니라 다양한 민주적 채널을 통해 이루어질 수 있도록 구상할 수 있을 것이다.

다. 국가 및 사회적 차원의 공동 위험 관리 및 공적 책임 분배를 위한 사회적 자본의 창출

자율주행자동차의 기술이 아무리 고도로 발전한다고 하더라도 자동차 사고의 일상성 및 정상성이 단숨에 해소되기는 어렵다. 또한 기존의 책임 법리에 따라 손해의 형평성 있는 분배가 완벽하게 이루어지는 데에도 한계가 있다. 특히 윤리적 딜레마를 동반한 사례에서 특

정인에게 과도하게 책임을 넘어서는 책무를 부여하는 것 역시 법윤리적으로 바람직하다 보기 어려울 것이다. 따라서 자율주행자동차의 상용화에 대비하여 일종의 공동 위험 관리 및 공적 책임 분배를 위한 국가 및 사회적 차원에서의 공적 자본을 창출하는 것이 필요할 것이다. 이는 책임보험을 통해서도 부분적으로 해결할 수 있겠지만, 사적 차원을 넘어 공적 자본을 통해 공동의 문제로 해결하려는 규율 방식에 대한 검토가 좀더 필요한 것으로 보인다.

라. 스마트 시티, 사물인터넷(IoT) 및 블록체인 등 4차 산업혁명 핵심기술들과의 연관관계와 규율방식

자율주행자동차에 대한 규율이 이루어지는데 있어서 초연결사회(ultra connected world)의 기본 전제, 즉 스마트 시티(smart city), 사물인터넷(Internet of Things; IoT), 빅 데이터(big data), 그리고 블록체인(blockchain) 기술이 자율주행자동차의 상용화와 총체적으로 논의되고 있다는 점을 주의 깊게 살필 필요가 있다.

먼저 자율주행자동차는 특히 스마트 시티 기획과 깊이 관련되어 규율될 수밖에 없다. 따라서 스마트 시티에 내재한 윤리적 쟁점 역시 동시적이고 유기적으로 고려하여야 할 것이다. 자율주행자동차와 도로교통체계의 디지털화의 궁극적인 목표는 도시 내의 정보와 자원이 초연결되는 스마트 도시의 구축이라고 할 수 있다(배영임・최준규・신혜리, 2018).

스마트 시티와 관련하여 가장 주된 윤리적 쟁점이 되는 것은 바로 보안(security) 문제이다. 도로, 터널, 에너지 분배 등과 같은 도시의 핵심적인 인프라에 embedded 되어 있는 다양한 장치들이 연결되게 됨에 따라, 도시 자산과 인명에 중대한 손해를 끼치는 사이버 공격의 위험을 경시할 수 없게 되었다. 이에 따라 스마트 시티의 보안 문제를 해결하기 위한 기술적 대안으로서 블록체인 기술의 적용 여부가 검토되고 있다(Qi・Feng・Liu/Mrad,.2017).

특히 사물인터넷(IoT)과 결합된 자율주행자동차 및 도로 인프라에서 생성 및 수집되는 개인정보에 대한 보호의 맥락에서 블록체인 기술의 활용 역시 고려해 볼 수 있다.

마. 인허가 과정에서 인공지능 기술의 특수성 및 규제의 유연성 고려

지능형 자율주행자동차의 윤리적 프로그래밍이 개발되고 그 성능을 검증하기 위해서는 임상적 단계에서의 성능검증뿐만 아니라 허가된 이후의 지속적인 모니터링이 요청될 수 있다. 사실상 자율주행자동차의 윤리적 프로그래밍은 하향식과 상향식이 혼종하는 혼합식 프로그래밍을 통해 지속적인 학습이 이루어지게 될 것이며, 이에 따라 개별 사안에서의 윤리

적 판단이 지속적으로 변천하게 될 것이므로, 인허가를 위한 1차적인 검증에 그쳐서는 안 된다. 특히 윤리 규범의 특수성을 고려할 때, 해당 윤리적 판단에 대한 사회적 수용성이 사례의 세분화 및 구체화 정도에 따라 변화가능하다는 점이 규제 방향성에서 충분히 고려되어야 할 것이다.

바. 제조 분야 윤리 대응의 프레임워크 예시

이와 같은 내용을 종합하여 제조 분야에서 윤리 문제를 대응할 프레임워크를 다음과 같이 고안할 수 있을 것이다.

〔그림 5－10〕 제조 분야 윤리 대응 프레임워크

<table>
<tr><th></th><th></th><th>특수 쟁점</th><th>개발자/공급자</th><th>공적 주체</th><th>이용자</th></tr>
<tr><td>개발 이전 단계</td><td>☞</td><td></td><td colspan="3">윤리 / 리스크 커뮤니케이션을 통한
〈자율주행차 윤리 가이드라인 등〉 구축</td></tr>
<tr><td>설계 단계</td><td>☞</td><td>알고리즘 구축</td><td rowspan="2">영향평가
자가리스크 진단</td><td>영향평가의 요령 제공</td><td rowspan="2">영향평가 단계에서 의견 제시</td></tr>
<tr><td>개발 단계</td><td>☞</td><td>타 기술과의 연관성 고려</td><td>스마트시티 구축 등의 지원</td></tr>
<tr><td>시험 단계</td><td>☞</td><td>인허가 단계의 윤리 검증</td><td>윤리 규범 준수</td><td>안전 및 윤리 기준 구축</td><td>의견 제시</td></tr>
<tr><td rowspan="2">소비 단계</td><td rowspan="2">☞</td><td rowspan="2">책임 주체로서의 자율주행차
책임보험제
소비자 자기결정권</td><td colspan="3">논의 플랫폼의 구축 / 민관협력의 사회적 협의체</td></tr>
<tr><td>피해 구제</td><td></td><td>리스크에 대한 수용</td></tr>
<tr><td>피드백 단계</td><td>☞</td><td>자동차 사고의 일상성 고려</td><td colspan="3">공동체 차원의 리스크 관리</td></tr>
</table>

3. 금융 분야

가. 논의의 출발점 및 원칙

금융 분야에서 인공지능이 적용될 수 있는 내용과 그에 따른 윤리 쟁점은 인공지능을 이용한 금융 기법 개발과 그에 선행하는 알고리즘에 대한 문제를 비롯하여 해당 기술을 이용하여 서비스를 수행하는 주체들의 행위윤리 문제를 중심으로 제기한 바 있다.

전통적으로 법과 제도는 금융시장에 참여한 금융소비자를 보호하기 위하여 다양한 보호

장치를 마련하고 있다. 예컨대 금융상품의 구매로 인한 위해를 사전에 예방하고, 금융업자의 파산에 대비하게 하며, 금융상품과 업자에 대한 정보비대칭으로 인해 약취의 대상이 되는 것을 예방하고 있다. 아울러 불평등한 취급에 대해서는 공정한 취급을 금지하고 취약계층을 배려하는 조치를 마련하기도 한다. 그 외 독점규제, 자금세탁방지, 금융회사의 부당한 고객 대우 문제도 예방하고자 하고 있다.

이하의 윤리적 이슈는 이처럼 금융회사와 금융소비자 간의 관계에 국한하지 않고 로보어드바이저산업에 직・간접적으로 참여하는 모든 관련자에게 요구되는 윤리적 이슈에 대하여 대응방안을 모색하고자 한다. 그 이유는 무엇보다 제4장에서 기술한 바와 같이 인공지능의 발전정도에 따라 차이가 있지만, 기본적으로 black box의 문제로 로보어드바이저라는 투자자문알고리즘이 잘 설계될 것(well design)이 무엇보다 중요하기 때문이다. 잘 설계된 알고리즘은 로보어드바이저가 자문 시 금융소비자에게 적합한 상품을 선택・자문할 수 있게 하고 시장의 상황이나 법제도 변화에 상응하도록 조정하기 때문이다. 따라서 잘 설계되었는지 알 수 있도록 설계문서의 작성 및 기록・보관・관리도 요구되어야 하고 이후에 수정이 없도록 할 필요가 있다.

아울러 판매단계에서는 로보어드바이저에 내재한 위험과 특징을 금융회사는 금융소비자에게 잘 설명하고 이해시켜야 할 것이다. 이 경우 현행 법제도하에서 요구되는 선관의무에 기초하여 잘 이해할 수 있는 방법과 기법을 사용하고 이와 관련한 기록을 보관・관리하여야 할 것이다. 아울러 이러한 판매가 금융소비자의 이익을 위하여 행해지도록 내부통제와 준법감시프로그램이 효과적으로 작동될 필요가 있다. 이를 위하여 인적・물적 시설이 적정하게 갖추어져야 할 것이다. 따라서 감독기관은 이러한 점을 감독하여야 할 것이다. 감독기관 역시 로보어드바이저의 파급효과에 대하여 적극적으로 주의를 기울여야 할 것이다. 예컨대 로보어드바이저의 불공정유형과 피해유형을 조기에 알 수 있는 기법과 국제적인 움직임을 지체없이 파악할 수 있는 인적・물적 시설과 역량을 증대시켜야 한다.

로보어드바이저에 대한 수요는 일반 이용자 뿐 아니라 기본적으로 비용과 수익에 민감한 기관투자자의 관심에 촉발되었고 그 발전을 선도하고 있음을 고려할 때 기관투자자에 대한 기술발전에 상응한 윤리교육이 필요할 것이다. 왜냐하면 로보어드바이저의 기술에 투자 및 이용하려는 자는 대부분 기관투자자이기 때문이다.

나. 참여자 유형별 정책 대응 시 고려 사항

1) 비금융회사(IT회사): well design 담보

로보어드바이저는 금융소비자의 이익을 추구하고 최선의 이익을 낼 수 있도록 처음부터 제대로 개발되는 것(well designed)이 중요하다. 먼저 “well design”의 정의를 고찰할 필요가 있다. 왜냐하면 빅데이터 수집과 분석이 가능한 AI를 전제한다면 슈퍼사람보다 더 나아야할 것으로 볼 수 있기 때문이다. 그러나 이 때의 기준은 적어도 최소한 사람이 하는 경우에 기대되는 수준의 행동(즉 금융소비자의 이익을 위하여 금융규제를 충실히 준수하는 자)을 하도록 설계되어 있어야 한다는 것이다. 물론 이보다 더 잘 되어 있다면 이는 로보어드바이저를 이용하는데 따른 비용보다 효과가 더 크다고 할 수 있다.

따라서 적어도 최소한 사람이 하는 경우에 기대되는 수준으로 행동할 수 있게 디자인된 것으로 보기 위해서는 설계단계에서 요구되는 다음 요소들을 충실히 담보할 필요가 있다.

- 충분한 전문성을 갖춘 투자전문가가 알고리즘을 개발하고 상시 감시하는 과정에 밀접하게 참여
- 알고리즘 가정(assumption)은 일반적으로 인정된 투자이론에 기반
- 알고리즘이 추구하는 결과
- 알고리즘이 디자인한대로 이행하고 있음을 보여주는 증명방식
- 로보어드바이저가 고려하거나 거절한 대안 및 그 설명방식
- 알고리즘 가정에 대한 설명서를 알기 쉬운 언어로 작성하여 고객이 언제라도 볼 수 있도록 공시
- 교육 및 역량기준을 충족하는 책임자 최소 1명이상 설치
- 알고리즘이 적정하게 설계되고, 모니터되고, 시험되는지를 다음과 같은 방법과 절차에 따라 확인
  - ☑ 알고리즘의 목적, 범위, 설계에 관하여 명확하게 기술한 시스템디자인문서를 작성, 이 안에 알고리즘의 의사결정순서도(decision tree)와 규칙 포함
  - ☑ 알고리즘 테스트의 범위를 설명한 테스트전략 문서 작성. 이 안에 테스트계획, 사례, 결과, 결점해결안 및 최종결과를 포함. 알고리즘의 건전성 테스트는 서비스 개시 전에 반드시 설시. 이후에는 정기적으로 실시
  - ☑ 알고리즘 수정을 관리하기 위한 절차 마련. 이 안에 알고리즘으로의 승인되지 않은 접근을 감시하고 방지하기 위한 보안절차 포함
  - ☑ 알고리즘의 변화 통제, 감시, 재구성
  - ☑ 시장 변화 및 법제도 변화로 영향받는 요인이 있을 때마다 알고리즘 검토·갱신
  - ☑ 알고리즘안에 오류 감지시 자문서비스를 중단할 수 있는 권한과 절차 마련
  - ☑ 로보어드바이저에 의해 제공된 투자자문에 대한 시의적절한 검토를 통해 알고리즘의 성능을 감시하고 감독할 수 있는 인적 및 기술적 자원을 적절하게 보유
  - ☑ 투자자문사가 IT업체의 로보어드바이저 품질을 감시하고 시험할 수 있는 절차 마련

2) 금융회사(금융상품・서비스판매)

현재 AI 발전수준에 대응하여 궁극적으로 인간의 통제가 가능하다는 전제 하에 요구되는 윤리적 요소는 다음과 같이 정리할 수 있다. 이중 무엇보다 이용자의 의사결정에 영향을 미치는 선택구조(decision architecture)에 주목할 필요가 있다. 금융회사가 설명・공시를 함에 있어 이용자들은 최근의 행동경제학의 연구결과에 의하면 프레임효과에 노출되어 왜곡된 의사결정을 할 가능성이 크기 때문이다. 때문에 비대면채널로 제공되는 로보어드바이저의 경우 무엇보다 의사결정을 하게 되는 선택구조를 디자인하고 전달하는데 편향된 의사결정이 이루어지지 않도록 할 필요가 있다. 아울러 감독기관도 이러한 부분에 주목할 필요가 있다.

- 고객확인 및 적합성원칙 준수
- 자문서비스를 제공할 금융상품에 적합한 고객을 선정할 수 있는 절차(filtering process), 인터페이스 설계(interface design)에서도 고객의 특성에 맞게 기술중립적으로 설계
- 공시 기준과 비용 등 투명성 강화
  - ☑ 서비스범위, 수수료 체계, 고객이 직간접적으로 부담할 수 있는 비용, 로보어드바이저 작동체계, 투자위험 등에 대해 투자자가 알기 쉽게 공시・설명
  - ☑ 고객에게 알고리즘이 고객의 계좌를 관리하고, 포트폴리오를 추천하고 자동으로 리밸런싱하는데 사용된다는 점을 설명
  - ☑ 알고리즘에 사용된 가정과 알고리즘의 한계를 설명
  - ☑ 알고리즘 이용으로 발생할 수 있는 위험에 대해 설명(예를 들어 시장상황과 특별한 관계가 없거나 고객의 기대와 달리 자주 리밸런싱이 일어날 수 있는 위험 또는 시장여건의 오랜 변화를 적시에 반영하지 못할 위험)
  - ☑ 시장이 위기상황에 직면한 경우 알고리즘의 거래가 중지되거나 다른 방어적인 조치가 취해질 수 있다는 점을 설명
  - ☑ 제3자가 알고리즘의 개발, 관리, 소유와 관련된 경우 그 사실과 함께 이해상충의 소지를 설명
  - ☑ 상시 감시 및 고객계좌관리에 사람이 얼마나 관여하는지 설명
  - ☑ 로보어드바이저가 포트폴리오를 추천할 때 고객이 제공한 정보를 얼마나 이용하는지 그 한계에 대해 설명
  - ☑ 고객이 언제 어떻게 로보어드바이저에게 제공한 정보를 업데이트 해야 하는지 설명
  - ☑ 로보어드바이저가 제공하는 투자자문서비스의 범위에 대해 고객에게 오해의 소지가 없게 분명하게 명시
  - ☑ 로보어드바이저는 핵심설명사항, 투자위험, 면책사항 등 중요정보를 고객에게 온라인으로 제공할 때 다음 사항을 고려할 것
    - → 핵심설명사항은 고객이 합리적인 투자의사를 결정할 수 있도록 계약을 체결하기 전에 제공
    - → 핵심설명사항은 팝업상자와 같은 웹기술을 이용해 특별히 강조
    - → 고객의 추가설명이 필요한 정보를 제공할 때에는 고객과 상호작용이 가능한 웹기술(tooltips등) 활용
    - → 중요정보를 공시할 경우 운영체계, 웹 브라우저, 기기 등에 영향받지 않도록 기술중립적으로 설계

- 로보어드바이저가 신의성실의무에 따라 고객의 이익을 최우선으로 대우
- 온라인으로 질문할 경우 다음의 사항을 고려
  → 온라인 질문이 로보어드바이저가 고객의 금융상황과 투자목적에 근거하여 고객에 적합하고 적정한 투자자문을 제공하기에 충분한 정보를 추출
  → 온라인 질문이 직관적으로 이해할 수 있게 쉬운 언어로 작성되어야 하며 추가적인 설명이 필요한 경우에 부연이나 예시를 제시
  → 고객이 온라인 질문에 비일관적으로 응답할 경우 고객에게 이를 알려 고객이 제공하는 정보를 수정하게 하거나 시스템에서 이를 자동적으로 적시해 재검토를 요구
- 고객이 자신의 투자성향과 투자목적에 맞지 않게 자산배분을 조정할 경우 이에 대하여 추가적인 자문을 제공
- 거래시 고객의 최선집행
  ☑ 합리적으로 설계된 절차와 방식으로 포트폴리오 주문을 관리, 이를 통해 고객에게 최선집행(best execution)을 제공
- 데이터 보호와 사이버안전
  ☑ 최고단계의 데이터 암호화, 제3자 데이터 리스크관리 위탁, 사이버 보안 보험가입, 보안사고 대비책 마련
- 상시 감시 및 준법감시 프로그램 구비
  ☑ 알고리즘에 사용된 방법론과 가정이 로보어드바이저가 제공하고자 하는 서비스에 잘 맞는지를 평가
  ☑ 알고리즘에 사용될 데이터 입력값 이해
  ☑ 알고리즘 결과가 회사의 기대와 상응하는지를 평가하기 위한 시험
  ☑ 알고리즘에 사용된 모델이 시장 또는 다른 여건의 변화에 따라 적정하게 작동하는 지를 평가
  ☑ 알고리즘 코드의 개발, 시험운용, 검증시험과 알고리즘 성능에 대한 사후 모니터링
  ☑ 로보어드바이저를 오프라인에서 투자자문도구로 활용할 경우 임직원에게 로보어드바이저는 허가된 사람만 가능한 점, 로보어드바이저에 사용된 가정과 한계, 로보어드바이저 사용에 부적합한 고객 유형 교육
- 로보어드바이저의 투자자문서비스를 마케팅하기 위해 사용하는 소셜 및 다른 형태의 미디어 사용의 적정성
- 로보어드바이저 감독에 책임있는 자를 명확하게 지정

다만, 인공지능의 발전단계가 자율형 인공지능이 범용화 될 경우 사람이 아닌 인공지능에 직접 금융규제를 적용하여 금융종사자에게 요구되는 지식과 규제준수 역량을 요구할 것인지는 좀더 검토가 필요하다. 전통적으로 금융서비스의 질을 거래상대방이 평가할 수 없기 때문에 통상 이러한 서비스를 제공하는 자에 대하여 진입규제, 행위규제, 상품규제를 하는 것이 보통이다. 특히 이들 상품과 서비스를 팔 수 있는 자격요건으로 전문성을 갖추도록 하는 진입규제와 교육을 고려하여야 한다. 지금까지 이러한 행동규율은 인공지능을 사용하는 금융기관에게 초점을 맞추어 요구되어 왔지만, 자율형 인공지능의 출현 시 이들에게 직접 진입규제와 교육을 요구하는 것이 필요하고 타당한지는 단언하기 어렵다.

3) 금융소비자

금융소비자는 정보 수집과 분석 등 정보비대칭의 취약점은 물론, 로보어드바이저의 규모가 확대될 경우 그로 인한 시장변동에 의해 집단적 피해에 노출될 가능성이 크다. 때문에 금융소비자 개인이 로보어드바이저에 내재된 위험을 알고 행동하는 데에는 한계가 있다. 무엇보다 로보어드바이저의 이용비중이 급격히 확대될 경우 모든 로보어드바이저가 시장에 일방향의 투자・매도지시가 내려질 경우 내지 로보어드바이저 사이에 담합이 이루어질 경우 투자자들은 공포상황에서 합리적인 판단의 일환으로 집단적 행동을 할 수밖에 없다. 그리고 그 결과는 오히려 비합리적인 결과를 낳는다. 즉 대량매도는 더욱 시장의 가격을 하락시켜 시장공포를 낳을 수 있기 때문이다.

아울러 로보어드바이저에 대한 지식과 정보를 금융회사로부터만 제한적으로 취득하는 것으로는 정보의 품질을 비교・신뢰하기 어렵다. 따라서 다양한 공신력 있는 정보채널을 통해 수집・제공받을 수 있게 할 필요가 있다. 이러한 점에서 금융교육의 기회를 확대할 필요가 있다.

4) 감독기관 등 공적 주체

첫째로, 금융 분야에서는 공적 주체의 전문성과 감독역량 강화가 매우 중요하다. 금융시장의 존속과 안정성을 위협할 가능성에 대해서는 이러한 위험을 예방하고 탐지하는 기술에 대한 투자를 지원하고 뒷받침하는 등 감독기법(대표적으로 RegTech, SuperTech)을 마련하고 지원할 필요가 있다. 특히 로보어드바이저의 이용 규모가 늘어날수록 감독기관은 더 적극적으로 로보어드바이저의 품질을 평가하고 점검하여야 한다. 왜냐하면 동일 방향의 투자지시는 금융시장에 폭발적인 충격을 가할 수 있기 때문이다. 현재 수준의 금융규제는 집단적 확대를 전제로 하지 않고 있지만, 향후 시장규모가 상당한 비중에 달할 경우에는 기존의 감독방식은 한계를 가질 위험성이 있다.

둘째로, 공적 주체가 금융 관련 기술 발전에 상응한 윤리 생태계를 조성할 필요가 있다. 감독기관은 정부의 예산에 기초하여 감독할 수밖에 없어 인적・물적 전문역량을 갖추는데 한계가 있다. 때문에 감독기관의 감독을 지원할 수 있는 금융시장 참여자들을 양산・확대하고 이들의 도움을 받는 환경을 만들 필요가 있다.

마지막으로 로보어드바이저 산업과 감독 관련 연구 지원을 고려해야 한다. 현재 로보어드바이저를 어떻게 규율할 것인지에 관한 연구와 분석에 관한 연구가 매우 드문 상황이므로

이에 대한 지원이 필요하다. 아울러 현재의 감독수준과 역량을 파악하는 것도 중요한 일이다. 이를 위해 현재 수준을 조사하고 그에 따른 감독수준이 충분한지 정기적으로 검토하는 일이 필요하다. 이를 위해서는 감독보고서나 조사결과를 정기적으로 홈페이지에 게시하는 것이 바람직할 것이다. 이러한 조사는 현재 금융회사들이 로보어드바이저 서비스 구입 내지 개발 현황을 정기적으로 파악함으로써 가능할 것이다.

다. 대응방식과 관련한 고려 사항

1) 정책적 지원 방식

정책적 지원과 관련하여 금융소비자에 대한 공신력 있는 정보채널 제공, 양질의 금융교육 기회 확대 및 감독기관의 역량 강화를 위한 정책 수행을 고려해 볼 수 있다.

또한 금융 분야에서 인공지능을 적용하고, 그러한 인공지능을 윤리적으로 사용할 환경을 마련하기 위한 정부 지원 방안 역시 고려해야 할 것이다. 감독기관의 감독을 지원할 수 있는 금융시장 참여자들을 양산・확대하고 이들의 도움을 받는 환경을 조성해야 할 것이며, 로보어드바이저 산업과 감독 관련 연구 지원이 필요할 것이다. 인공지능 전문금융인력 양성과 금융윤리 교육 기법과 교육 내용을 전파하는 기관에 대한 효과적인 지원체계를 마련함으로써, 관련 교육의 수행을 촉진할 수 있을 것이다. 또한 기술의 안전성을 담보하고 검증하는 정보기술인프라의 확보도 필수적이다.

기관투자자에 대한 고려도 필요하다. 투자와 수익에 민감한 속성상 윤리적 이슈를 간과하거나 배제하는 경우가 적지 않으므로 과학기술의 발전과 그에 상응하는 윤리적 이슈를 고려하여 투자 및 지원하도록 환경과 인센티브를 제공할 필요가 있다.

금융 업무 수행과 관련하여 정부가 모범 기준을 포함한 윤리강령(Code of Ethic)과 행위기준(Standards of Conduct) 등의 가이드라인을 제시하는 것도 고려해 볼 수 있다.

필요한 경우 입법을 통한 환경 개선도 생각해 볼 수 있다. 로보어드바이저의 발전 내지 이용 규모에 따라 기존에 법제화된 것 외에 앞서 제시한 요소들을 모두 법제화하거나 자율규제기관과 이원적으로 규율하는 방안을 생각해 볼 수 있다.

2) 사회적 대응

로보어드바이저에 대해서는 기술의 진화에 따라 이용자가 피해 노출 내지 불이익을 볼 가능성이 없지 않으나 이에 선제적으로 피해를 예방하고 기술의 진화에 따른 혜택이 비용을 초과할 수 있도록 생태계를 조성할 필요가 있다. 이에 가장 기여하는 것이 바로 그에 맞는

윤리를 마련하고 모든 구성원에게 인식을 강화시키는 것이라고 할 것이다. 문제는 로보어드바이저의 발전 단계를 누구도 예측할 수 없다는 점이다. 때문에 사회적으로 필요한 것은 무엇보다 이러한 발전 정도를 모두가 쉽게 알 수 있고, 비교할 수 있고, 문의할 수 있는 네트워크가 넓게 구축하는 일이다. 이는 반드시 감독기관만의 역할은 아니라고 할 것이다. 소비자단체, 연구기관, 학계, 금융회사, 금융회사 관련기관, 감독기관, 정부 모든 부처에 인공지능 관련한 네트워크를 마련하여, 관련 정보를 쉽게 수집하고 제공받을 수 있는 시스템을 구축해야 할 것이다. 단기적으로는 반드시 단일한 하나의 시스템을 요하지 않고, 링크와 같은 방식을 고려할 수도 있을 것이다. 아울러 이에 대한 인식과 주의를 환기시키는 다양한 방법을 마련할 필요가 있다.

라. 소결

위의 내용을 고려하여 금융 분야에서의 윤리 대응 프레임워크를 다음과 같이 제시할 수 있다.

〔그림 5－11〕 금융 분야 윤리 대응 프레임워크

<table>
<tr><th></th><th></th><th>특수 쟁점</th><th>개발자/공급자</th><th>공적 주체</th><th>이용자</th></tr>
<tr><td rowspan="2">개발 이전 단계</td><td rowspan="2">☞</td><td rowspan="2">AI 개발 촉진과 규제 균형
AI 이용 관련 금융 종사자에 대한 교육</td><td colspan="3">윤리/리스크 커뮤니케이션을 통한 사회적 합의</td></tr>
<tr><td>종사자 교육 시행</td><td>규제 개선
교육 프로그램 제공</td><td>AI 금융 시스템에 대한 이해</td></tr>
<tr><td>설계 단계</td><td>☞</td><td rowspan="3">알고리즘 윤리적 구축(Flash Crash 방지)
윤리적 자가 검증</td><td rowspan="3">영향평가
자가리스크 진단
투자 전문가의 참여</td><td rowspan="3">상시적 금융 감독
리스크 진단 도구 제공</td><td rowspan="3">의견 제시</td></tr>
<tr><td>개발 단계</td><td>☞</td></tr>
<tr><td>시험 단계</td><td>☞</td></tr>
<tr><td rowspan="2">소비 단계</td><td rowspan="2">☞</td><td rowspan="2">투자자 보호
윤리적 투자</td><td colspan="3">윤리적 투자에 대한 사회적 합의</td></tr>
<tr><td>투자자 피해 구제</td><td></td><td>투자 리스크에 대한 수용</td></tr>
<tr><td>피드백 단계</td><td>☞</td><td>기관 간 정보 공유
해킹 대비</td><td colspan="3">데이터 공유에 대한 사회적 합의 필요</td></tr>
</table>

4. 의료 분야

가. 논의의 출발점

의료 분야의 기술 개발에서는 특히 임상 단계의 중요성을 강조할 수 있다. 한편 의료과실 시의 책임소재 문제 역시 법적인 논의 뿐만 아니라 윤리적인 영역에서도 문제될 수 있다. 특히, 인공지능 기술의 활용에 따른 의료사고의 경우, 인공지능 자체에만 문제가 있었고 의료인이 주의의무 등을 위반하지 않았을 때, 의료인의 과실이 조각됨으로써, 피해는 존재하지만 책임은 존재하지 않는 상황이 발생할 가능성이 있다. 이와 같이 민사법적 책임이 존재하지 않는 경우, 윤리적인 해결 방법을 모색할 수 있을 것이다. 이 과정에서 특히 논의 플랫폼을 통하여 인공지능 기술의 사회적 수용과 리스크의 사회공동체 부담이 인정된다면, 사회보험 등 사회법적 수단을 통하여 피해자가 피해를 보전받을 가능성이 열릴 수 있을 것이다.

나. 신 기술의 출현과 관련한 고려 사항

의료 분야에 적용되는 새로운 인공지능 기술의 출현으로 인해 야기되는 문제들을 윤리적 측면에서 접근하여 고려할 필요가 있다. 특히 의료 기술은 개인의 신체적 자유와 밀접한 연관을 가지므로, 정책 대응 수립 시에도 이를 유의할 필요가 있다.

먼저 의료 인공지능을 통해 개인의 생체정보, 유전정보 등 각종 민감성정보가 광범위하게 수집・활용되는 과정에서 인간의 몸에 대한 등급화 현상이 사회 내에서 잘못된 문화로 자리잡을 수 있다는 점을 인식해야 할 것이다. 혐오, 차별, 불평등 등 일체의 편견과 왜곡된 가치에 대한 사회적 의식 개선 작업이 필요하다. 또한 미디어, 학교, 가정 등 주요 사회화 기관의 역할 재규정 및 강화가 요구된다. 의료계는 과도한 생의학 패러다임에서 벗어나 생물사회의학 패러다임을 수용하는 노력에 적극적으로 임해야 할 것이다.

의료 인공지능을 향후 어느 수준까지 발전시킬 것인가의 문제를 관련 행위자들의 양심과 이해관계에만 맡겨둘 수는 없는 일이다. 오히려 기술의 발전단계와 연동하여 지속적인 사회적 합의를 이어나가야 하며, 이를 위해 상설 거버넌스 협의체 구성 등을 모색할 필요가 있다.

정책적 대응 방안은 기술 제안, 구현, 시험 등 기술개발단계에서부터 원천기술 차원에서 통제 장치를 확보하는 방식으로 설정해야 한다. 이런 측면에서 의료정보 데이터의 익명성 처리, 의료정보 데이터의 식별불가능성 보장, 의료정보 데이터의 연결불가능성 보장, 의료정

보 데이터에 대한 접근 제어 기능 등을 고려하여 정책 대응 방안을 강구해야 할 것이다.

다. 형평성과 관련한 고려 사항

의료 인공지능의 수혜 격차는 건강에 대한 불평등을 가속할 소지가 있다. 고령화에 뒤 이어 건강에 대한 관심이 고조되면서 건강형평성에 대한 요구는 계속 증가할 것이므로 이와 관련한 의료계, 보험업계, 정책 당국의 균형 있는 협의가 요구된다. 의료 수혜의 격차를 복지적 차원에서 줄여나가고자 한다면 국가의 재정적 부담, 수혜 대상자에 대한 선별 문제 등 복지정책 전반의 논의들이 불가피할 것이다. 의료수가나 비급여 인정 여부, 공공의료 범위 설정 등과 관련한 사회적 합의 과정이 필요할 것이다(박혜경, 2017).

따라서 의료 인공지능 활용과 관련한 의료수가 인정 내지 비급여 인정 문제(윤혜선, 2017), 국민건강보험제도 및 민간의료보험제도의 수정, 취약 계층에 대한 공공의료 차원의 첨단 의료서비스 지원 방안 등을 고려해야 할 것이다.

라. 불법 및 자가 의료행위와 관련한 고려 사항

의료 수혜 격차를 등에 업고, 의료 인공지능 소프트업계가 프로그램의 이용자 저변 확대 차원에서 무료 혹은 저가 소프트웨어 이용 서비스를 제공할 수 있다. 2017년 10월 지브라 메디컬 비전은 다양한 종류의 영상의료 데이터를 1달러에 판독해주는 서비스(AI1)를 구글 클라우드를 기반으로 전세계를 대상으로 출시하였다. 루닛 역시 2017년 북미영상의학회(RSNA)에서 클라우드 기반으로 무료로 흉부 엑스레이에서 폐암 결절, 결핵, 기흉 등 주요 폐질환을 실시간 진단할 수 있는 루닛 인사이트(Lunit Insight)를 발표한 바 있다(최윤섭, 2018). 향후 이러한 사례가 국내외에 대거 출현할 수 있으므로 관찰감시시스템을 구축하고 상시 가동할 필요가 있다.

의료 인공지능 소프트웨어 프로그램의 무료, 저가 보급 과정에서 인・허가를 받지 않은 업체들이 생겨나고 불법, 무허가, 불량 소프트웨어가 대량으로 유통될 수 있다. 가짜 프로그램 이용에 따른 소비자들의 신체적 위해, 폭리 피해 등 의료적・비의료적 폐해가 발생하지 않도록 지속적인 법제도적 감시와 단속을 준비해야 한다.

의료 수혜 불평등 개선, 유사・불법・자가 의료 행위로 인한 폐해 예방을 위해 의료업계에 인공지능 도입이 활성활 될 수 있도록 정책적 지원이 필요하다. 이러한 차원에서 불법, 가짜 의료 인공지능 소프트웨어업체에 대한 법적 제재 규준을 마련할 필요가 있고, 유사, 저

급 의료에 대해 인공지능 소프트웨어를 활용한 불공정 거래를 감시 내지 단속할 방안을 강구해야 하며, 소비자 피해보상 및 구제 규정 역시 고려해야 할 것이다.

마. 기술적 실직과 관련한 고려 사항

의료 분야에서 인공지능이 신속하고도 적극적으로 도입될 것으로 예상되는 부분은 자동화로 편리성이 제고될 부분, 비용 절감과 효율성이 높은 영역, 난이도 및 위해도가 적지만 의사들이 수행하기 싫어하는 영역, 의료행위에 보험 수가가 낮은 영역 등이다. 의료 인공지능은 병원단위, 학과(학교 단위) 차원에서 특히 대량의 영상 의료 데이터를 집중적으로 판독해야 하는 건강검진센터나 관련 전문병원 등에서 판독정확성, 판독효율, 비용 절감 등을 도모하기 위해 도입을 서두를 수 있다(최윤섭, 2018). 대규모 인원 감축이 예상되는 분야에 대한 전문 인력 조정 등을 의료계와 정부가 함께 검토해야 할 것으로 판단된다.

의료 분야에서 인공지능의 도입 과정에서 병원경영진(경영이익), 인공지능 공급기업(영업이익), 직접적인 관련 진료 분야(일자리 감소) 간의 갈등이 예상된다. 실제로 세데시스 도입 과정에서 존슨앤존슨과 미국 마취과의사협회와의 갈등이 불거진 바 있다.

의료 인공지능의 도입으로 관련 진료과목 전공자들은 탈숙련화를 겪을 수 있지만, 인공지능을 활용할 수 있는 교육과정 수행 및 새로운 기능 습득을 통해 더 효과적으로 업무에 임할 수도 있다(최윤섭, 2018). 따라서 직종 전체를 대체한다기보다 특정 작업을 대체하거나 직종 내 인력 구성 및 역할 변화가 발생할 것으로 예상된다(장동경, 2018). 의과대학과 의료 관련 연구소 등에서는 이러한 패러다임 변화에 대응할 수 있는 협업 구조 논의 등 전략 구상이 요구된다.

정책적으로는 의료보험수가가 낮지만 인력 감원에 따른 의료적 피해가 큰 영역에 대한 수가 조정 작업이 필요하고, 병원경영기업, 경영대리인, 의료 인공지능 공급 사업자, 진료 분야 전문가집단 간 갈등조정을 위한 갈등관리위원회 등 정책기구의 구성 및 대응 역시 고려해야 한다. 또한 의료교육 패러다임을 조정(영상의학과의 인공지능 교육 및 연구 참여 활성화)하고, 영상의학 파트, 마취의학 파트 의료인력과 관련한 대학정원은 하향 조정할 필요가 있다. 의과대학의 교육과정(커리큘럼) 및 의사 수련 과정 역시 변화할 필요가 있다.

바. 생산 단계에서의 고려 사항

인공지능의 능력이 인간의 역량을 뛰어넘는 부분이 발생할 경우, 인간의사의 개입 없이

독립적인 진단-처방 등의 업무수행을 행사하는 경우에 대비하여 기술적 발전 수준을 어디까지 허용할 것인가에 대한 과학기술자, 시민, 정부(국가), 시장(market) 간의 사회적 합의가 지속적으로 이어져야 한다. 인공지능 기술의 발전상을 고려하여 자격지위의 재조정 등을 검토해야 할 것이다.

이를 위해 구체적으로는 의료 인공지능(소프트웨어) 등 공식 용어 사용을 공지하고 공유화하는 한편.[236] 'AI의사'라는 용어 사용 시 불공정거래관련 규율을 적용(과장광고)하는 방안을 생각해 볼 수 있다.

병·의원 밖에서 개인들이 일상의 건강관리를 목적으로 웨어러블기기를 통해 수집하는 데이터를 인공지능 프로그램을 통해 활용하는 경우는 현재로서는 의료행위로 분류하기 어렵다. 그러나 향후 인공지능의 활용은 의료기관 외부에서 의료기관과 외부를 연결하는 기기들(U헬스케어, 원격진료, 웨어러블기기 등)을 통해 광범위하게 확산될 것으로 예상된다. 예를 들어, 당뇨나 고혈압, 치매 등 만성질환이나 관리형 질환의 경우 앞으로는 의료기관 외부에서 수집한 생체정보를 기반으로 의료적 진단과 처방이 내려지는 경우가 더 늘어날 것이다. 이렇게 될 경우 의료행위의 경계는 매우 모호해질 수 있다. 의료 인공지능의 발전 추이를 살펴가면서 '의료행위'에 대한 법적 개념 및 범위도 유연하게 재검토·조정해나갈 필요가 있다.

의료 인공지능이 비의료기기로 분류되어 있다는 점으로 인해 영리목적의 상업자본이 개입하여 부당하게 사용되는 방향으로 이어질 우려가 있다. 현재 복지부는 왓슨이 보다 발전된 의학교과서의 개념으로, 평소 의사들이 진단과 처방을 내림에 있어서 관련 서적과 논문 등을 참고하는 것과 같은 성격으로 봐야 한다고 결론을 내린 상태이다. 즉 의료기기나 장비가 아닌 만큼 허가를 필요로 하지 않고, 의료신기술 등으로도 분류되지 않는 만큼 왓슨을 사용하는 것은 의료법상에 저촉될 아무런 문제가 없다는 판단이다(이언, 2017: 35). 하지만 의료기기로 분류되어 있지 않기 때문에, 의료기관이 아닌 곳에서 전문 의료진이 아닌 사람들에 의해서 의료 인공지능이 활용될 수 있다는 점은 의료적 관리 망을 벗어나게 됨으로써 여

236) 보건의료 분야에 활용하는 인공지능 기술의 존재론적 지위를 인간의 통제 범위 안에서 제한하는 수준으로 규정하는 것은 인간의사와 AI의사 간의 헤게모니 쟁탈이라는 과장된 우려와 윤리적 논쟁을 벗어날 수 있는 길을 열어 준다. 이처럼 기술에 대한 법적 규제는 기술의 발전 단계와 사회윤리적 합의 수준을 통합적으로 고려하여 유연하게 조정해야 한다.

러 가지 문제를 초래할 수 있다. 개인의 생체정보와 유전정보를 직접적으로 수집하고, 필요에 따라서는 이를 상업적, 범죄적 용도로 활용할 가능성을 배제하기 어렵다.

의료 인공지능이 비의료기기로 분류되어 있다는 점은 공식적인 의료적 관리체계에서 벗어나 있다는 것을 의미한다. 의료 인공지능이 의료기기로 분류된다면 임상연구를 통해서 정확성과 안전성을 충분히 검증하고 인허가를 받은 이후에 시장에 출시할 수 있다. 이 경우 인공지능을 시장에 출시하기 전에 어느 정도로 검증할 것인지는 제조사의 결정에 달려있다(최윤섭, 2018: 393). 그러나 의료기기가 아닐 경우 '의료적으로' 충분히 검증되지 않은 상태로 출시될 수밖에 없다. 그런데 의료 인공지능을 활용하는 과정은 의료적 용도와 상당히 흡사하며, 필요에 따라서 의료적 근거 자료로 제공되기도 하므로 사용 과정에서의 문제가 반드시 비의료적인 영역에만 머무른다고 보기 어렵다. 무면허 의료행위, 유사 의료행위의 출현으로 이어질 수도 있으며, 제조사, 사용자, 의료진 간의 임상적 유용성 및 안전성 평가·검증 논란도 발생할 수 있다.

이와 같은 쟁점에 대한 정책 대응방안을 구축하기 위해서는 구체적으로 다음 사항들을 고려해야 할 것이다.

- 신속한 신(新)의료기술 및 신(新)의료기기 인허가제도 도입
- 인공지능기술이 적용된 기기나 앱을 규율하는 법규정이 없는 경우 임시허가, 규제 샌드박스, 비조치의견서 등 유연한 규제수단 활용 모색
- 인공지능의 안정성, 안전성 확보를 위한 제도적 방안으로 인공지능 기기의 품질인증제도 및 안전 관리감독제도의 도입 강구
- 제조물 책임법의 정비, 보안 및 관리 강화를 위한 민형사상 책임제도 강화, 행정형벌제도의 정비 등도 고려할 수 있음
- 비의료기기로서의 부당한 상업적 사용 우려(박혜경, 2017)

사. 인공지능 활용 편의성 관련 고려 사항

의료 인공지능은 다양한 장점과 활용가치를 제공해줄 것으로 기대되는 만큼 전문 의료인력이 편리하고 효율적으로 활용할 수 있도록 사용편의성을 제고할 필요성이 있다. 만약 이러한 활용효용성 및 사용편의성 등이 확보되지 못할 경우, 전통적인 전문 의료진의 권위에 의존하는 진료 관행이 개선되기 어렵다. 따라서 의료업계의 진료'문화' 개선을 위한 패러다

임 전환 노력이 지속적으로 요구된다.

이를 위해 먼저 의료 인공지능의 정확성, 안정성, 효용에 대한 검증시스템 확보 차원에서 개발자에게 아실로마 선언의 투명성 원칙의 준수를 요구할 필요가 있다.

개발자, 디자이너 등은 의료 인공지능의 기술적 수준을 향상하는 노력 이외에 기기와 프로그램의 구체적인 사용법, 사용자 인터페이스, 사용자 경험 등에 주목하여 의료진의 진료 프로세스에의 수용 정도와 활용도를 높일 수 있도록 시스템을 구현해야 할 것이다. (약의 복수 처방을 경고하는 전자의무기록의 전자처방시스템(CDSS)와 같은 수준의 기술이 아니더라도) 효용성과 정확도 높은 결과를 도출하기 위한 활용 과정의 핵심 정보 안내지침(인공지능 활용 가이드라인 등)을 마련하여 제시할 필요성이 있다. 의료계가 기존 진료시스템을 바꾸지 않고도 접목해서 활용할 수 있도록 현재의 상황에 맞는 '의료계친화시스템'으로 인공지능을 개발해야 할 필요성이 있으므로,[237] 이에 상응하는 고려가 이루어져야 할 것이다.

아. 신기술 소비와 관련한 고려 사항

의료사고나 과실, 분쟁의 수준이 아니더라도, 인공지능과 의사의 판단이 불일치할 경우 어떻게 대처할 것인가에 대한 정책적 원칙이 필요하다.

현행법상 왓슨의 경우는 비(非)의료기기로 분류되어 법적 인격이 인정되지 않고 있기 때문에, 왓슨이 의사를 조력하는 위치라고 볼 때 의료행위에 대한 지휘·감독권은 의사에게 있다(장연화, 2017). 따라서 의사는 왓슨과의 판단 차이가 발생했을 경우, 인공지능의 진단 및 처방결과를 선택하더라도 그것이 불러올 수 있는 부작용 등에 대해 환자에게 충분히 고지해야할 의무를 준수해야 한다. 그럼에도 불구하고 인공지능에 의존한 결정으로 중대한 의료분쟁이 발생했을 경우에도 의사에게 책임을 물을 수 있는가라는 문제에 관련해서는 논란이 예상된다. 인공지능이 의료 분야에서 상용화되고, 의사의 의료 인공지능에 대한 의존도 및 권한 부여가 증가할 경우, 만일의 의료분쟁과 관련해 의사에게 주어지는 면책권이 이전보다 훨씬 증가할 수 있기 때문이다. 의사의 독점적 권한과 면책권 간의 논쟁에 대한 세밀한 법적 보완이 필요하다.

의료 인공지능의 활용 과정에서 의료분쟁 시 의사의 의료 독점권과 면책권 문제는, 인공

237) 구글의 인공지능 병리 현미경의 경우 증강현실 기능을 채용함으로써 의료 전문인력이 기존의 업무 시스템을 유지하는 가운데 업무 효율과 편의를 획기적으로 향상시킬 수 있도록 기여하고 있다(최윤섭, 2018: 376).

지능(제조사)과의 진료 독립권 및 면책권과도 연동되어 있다. 따라서 두 행위자의 양적・질적 진료 참여 과정, 진료 내용, 분업 정도 등을 통합적으로 고려하여 권한과 책임의 배분를 조정하는 작업이 필요할 것으로 판단된다.

의료사고 및 의료과실(malpractice)을 둘러싼 의사와 인공지능(제조사) 간의 책임 배분의 핵심에는 의료사고의 정확한 원인 규명이 자리 잡고 있다. 의료 인공지능이 초래할 사고 원인으로는 인공지능의 데이터 활용 및 진단 오류, 인간의사의 의료 인공지능에 대한 과도한 신뢰관습에 따른 탈숙련화와 오진, 의료 인공지능의 정확성의 한계와 검증 부재, 인공지능이 제공한 데이터에 대한 과도한 신뢰와 의존, 오진 판독결과의 도출, 인공지능 오작동에 따른 사고 등 매우 다양한 요인들이 있을 수 있다. 원인 파악을 위한 다각도의 확인 작업을 관리할 주체가 필요하다.

인공지능에 대한 신뢰도 증가는 환자의 치료법 선택권한을 증대하는 결과를 불러올 수 있다. 여기에는 전 세계적으로 치료법에 대한 환자의 자기결정권 강화 패러다임의 증가 분위기도 가세할 것이다. 환자의 의사결정권을 존중하는 의료계의 '공유된 의사결정(shared decision making)' 및 시민권에 대한 관대한 의료문화 형성도 필요해 보인다.

이를 위해 다음의 대응 방안 역시 고려해 볼 수 있다.

- 의사와 인공지능제조사에 의료 인공지능 사용과 관련한 의료사고책임보험 가입의 의무화(정채연, 2017)
- 의료 인공지능의 상용화에 대비하여 의사의 환자에 대한 인공지능 활용과 관련한 설명(고지) 및 사전 동의(informed consent) 의무화
- 기술적인 문제와 의료행위 관련 문제를 동시에 검토하여 의료사고의 원인을 판정할 주체(개인 혹은 협의체) 필요
- 인공지능제조사와 인・허가권자, 품질인증관리업계 사이의 책임 배분과 관련한 대응 준비

자. 데이터 관리와 관련한 고려 사항

의료 인공지능이 수집, 가공, 생산하는 의료정보는 개인정보라는 고유성 외에도 특정 사회, 나아가 인류 전체의 건강과 보건의료의 수준을 제고하는 데 의미 있게 활용될 전망이다. 따라서 데이터의 생산~유통~소비에 이르는 제반 과정을 통합적인 관점에서 체계적으로

통제 및 관리할 수 있는 시스템을 만들 필요가 있다.

이를 위해 다음과 같은 사항들을 구체적으로 고려할 필요가 있다.

- 투명하고 합리적인 데이터의 확보처리체계 마련 시급
- 표준화된 데이터, 질 좋은 데이터의 수집 기법 확충 필요(김형회, 2017)
- 의료데이터 표준화를 위한 단일 의학용어체계 마련, 의료데이터 항목의 공통화 추진 필요(김형회, 2017)
- 의료데이터의 검증 및 품질관리 문제
- 데이터의 수집・처리・이용의 투명성 강화
- 건강정보 개념 확립 및 유형화(의료정보, 유전자정보, 건강진단정보, 개인생체 정보 등)(윤혜선, 2017)
- 의료데이터 디지털화 및 통합관리(윤혜선, 2017)
- 의료데이터(특히 유전자데이터) 저장장치 확보 문제
- 기계학습용 데이터 관리 및 낡은 데이터의 업데이트 제도화 필요
- 데이터 손실. 변형(가공) 방지 등 품질관리 문제(윤혜선, 2017)
- 의료정보관리서비스학, 의료정보관리서비스업의 육성 지원 필요

인공지능의 진화와 관련한 기술적 범주의 문제들은 기술적 성취라는 측면에서 그 해결점을 찾는 것에 한계가 있을 수 있다. 기술은 그 자체로 완벽에 도달하였다고 평가되기도 어렵지만 특정 기술이 완벽하다는 평가를 받았다 해도 그러한 기술적 성취 수준이 늘 '절대성'을 갖는다고 보기는 어렵다. 왜냐하면, 기술은 '사회'라는 역동적인 시・공간에 놓여 살아 있는 유기체와 같은 특성을 가지고 있기 때문이다. 실제로 많은 기술은 사회를 구성하는 다양한 측면(요소)들, 즉 정치・경제・사회・문화・이데올로기는 물론 기술 생산자이자 사용자인 사회구성원들과 상호작용하는 과정에서 그 실재가 예측불허의 존재로 변형될 수 있다. 이렇게 되면 기술은 오・작동과 같은 기술적 결함 외에 사회와의 상호작용 과정에서 만들어지는 '사회－기술적 오・작동'을 출현시킬 수 있다. A. 기든스(Giddens), U. 벡(Beck) 등의 거시 사회이론가들은 이러한 과학기술의 불확실성이 초래한 문제를 쟈거노트((juggernaut: 절대적인 힘의 파장), 위험(danger) 등으로 표현한 바 있는데, 이들의 우려는 고도로 발달한 최첨단의 과학기술이 '구조화'되었음을 시사한다.[238] 인공지능의 의료적 활용 과정에서 나타날 수 있

는 이러한 위험성은 '사회적 재난'으로 이어질 수 있으므로, 이를 사회적 재난관리시스템에 포함하여 통제할 필요가 있다(박혜경, 2017). 이와 관련하여 프레임워크 수립 시에 특히 고려해야 할 사항은 다음과 같다.

- 데이터 폐기 절차와 관련한 문제(개인정보의 신속한 업데이트 시스템 구축, 낡은 데이터로 인한 개인 평가 오류 발생 차단)
- 개인정보보호(프라이버시권 강화), 개인정보이용동의의 강화
- 가짜 정보 유통 검색 기법(기술) 마련과 제재 대책
- 의료 빅데이터 활용 시 개인프라이버시 보호와 공익적 활용 간의 충돌 조정(이중원, 2017)
- 자산가치가 큰 국가적 수준의 데이터 유출 우려(장동경, 2017)
- 생체정보의 상업적, 범죄적 악용에 대한 제재
- 데이터 보안, 암호화, 비식별화 수단 개발 및 활용을 위한 제도적, 정책적 지원 필요(윤혜선, 2017)
- 해킹 테러 등 데이터 안보를 위한 기술적, 제도적 장치 마련 필요(손해배생책임의 문제, 인과관계 입증의 문제, 국민건강보험제도 및 민간보험제도 개선 등 고려 필요)

차. 신기술 재생산과 공익적 활용 관련 고려 사항

데이터 공유는 환자의 불필요한 검사 생략, 의료기관 간 의료정보 공동이용에 의한 경제성 및 효율성 증대, 통계, 임상시험, 역학 연구 용이 등 다방면에서 대단히 유용하다(윤혜선, 2017: 61). 공익목적의 이러한 의료정보의 공유는 국가 내에서 이루어질 수도 있지만, 국제적 차원에서도 요구될 수 있기 때문에 공익적 활용과 개인정보보호 문제 간의 충돌이 발생할 수 있다. 균형점을 찾아 해결할 필요가 있다.

개인 의료정보는 공공 자산화하거나 국제적 이전 과정에서 소유권을 누가 가질 것인가의 문제로 쟁점화할 수 있다. 원천적으로 개인 소유로 남겨둘 것인지, 진료과정에서 제공된 이

238) 앤서니 기든스, 울리히 벡 등은 이처럼 과학기술이 사회 속의 다른 여러 요소들과 뒤엉켜 발생하는 통제할 수 없는 거대한 위험을'쟈거노트(Juggernaut)'라고 명명하고 강력하게 경고한 바 있다.「저거노트(Juggernaut)」는 힌두교의 신인 크리슈나의 이름으로, '행복과 파멸을 동시에 의미하는 저항할 수 없는 힘'을 의미하는 말로 인용되고 있다.

후 시점부터는 병의원에 소유권을 넘겨줄 것인지, 국가 간 정보 이전 과정에서는 국가가 소유권을 갖고 누리도록 할 것인지, 다양한 문제들이 예상되므로 체계적인 논의와 권리 정돈이 필요하다.

카. 소결

이상의 내용을 종합하여 의료 분야에서의 윤리 대응 프레임워크를 구성하면 다음과 같다.

〔그림 5－12〕 의료 분야 윤리 대응 프레임워크

<table>
<tr><th></th><th></th><th>특수 쟁점</th><th>개발자/공급자</th><th>공적 주체</th><th>이용자</th></tr>
<tr><td>개발 이전 단계</td><td>☞</td><td>인간 중심주의 강조
의료 정보 데이터 취급에 유의</td><td>의료 종사자에 대한 교육
(의과대학 포함)</td><td>데이터 법제 정비
교육 내용 제공</td><td></td></tr>
<tr><td>설계 단계</td><td>☞</td><td>알고리즘 구축</td><td></td><td>영향평가의 요령 제공</td><td rowspan="2"></td></tr>
<tr><td>개발 단계</td><td>☞</td><td></td><td>기술영향평가</td><td>의료기기 인허가 기준</td></tr>
<tr><td>임상 단계</td><td>☞</td><td>임상 윤리 준수</td><td>윤리적 자가 검증
임상 표준 고려</td><td>표준
의료품질 제시</td><td></td></tr>
<tr><td rowspan="2">소비 단계</td><td rowspan="2">☞</td><td rowspan="2">유사, 불법 및 자가 의료행위
기술수혜 격차 유의
경영자와 의료인 충돌 고려</td><td colspan="3">논의 플랫폼의 구축 / 민관협력의 사회적 협의체</td></tr>
<tr><td>의료사고 피해 구제</td><td>불법의료행위 규율
과대광고 규율</td><td>의료인의 책임 인식
리스크에 대한 수용</td></tr>
<tr><td>피드백 단계</td><td>☞</td><td>의료 정보 데이터의 공익적 재활용</td><td colspan="3">데이터 활용에 대한 사회적 합의</td></tr>
</table>

5. 군사 분야

군사 분야의 인공지능 기술은 다른 분야에 비해 인권 측면 및 안보 측면에서 민감성을 지닌다. 때문에 군사 분야에서의 윤리 문제를 다룰 때에도 그와 같은 측면을 고려해야 할 것이다. 특히 국제 인권협약, 국제규범 등을 준수할 필요성이 있는 반면, 안보의 관점의 특수성도 고려해야 할 것이다. 요컨대 군사 분야에서의 쟁점은 윤리 문제와 안보 필요성 사이의 균형을 어떤 지점에서 찾을 것인지의 문제이다.

〔그림 5－13〕 국방 분야 윤리 대응 프레임워크

<table>
<tr><th></th><th></th><th>특수 쟁점</th><th>개발자/공급자</th><th>공적 주체</th><th>이용자</th></tr>
<tr><td rowspan="2">개발 이전 단계</td><td rowspan="2">☞</td><td rowspan="2">인권 존중
국제인권협약 준수
LAWS 개발의 국방상 필요성 검토</td><td colspan="3">인권 수준에 대한 사회적 합의와 리스크 커뮤니케이션
전력상 LAWS의 필요성에 대한 사회적 합의</td></tr>
<tr><td colspan="2"></td><td rowspan="4"></td></tr>
<tr><td>설계 단계</td><td>☞</td><td rowspan="2">알고리즘 구축</td><td rowspan="2">알고리즘
구축에서의
전쟁법 준수</td><td rowspan="4">안보
수호자로서의
책임 인식</td></tr>
<tr><td>개발 단계</td><td>☞</td></tr>
<tr><td>시험 단계</td><td>☞</td><td>안전성 보장
효용성 고려
테러리스트에 대한 보안</td><td>윤리 규범 준수</td></tr>
<tr><td rowspan="2">소비 단계</td><td rowspan="2">☞</td><td rowspan="2">상대 세력의 전력 고려</td><td colspan="3">논의 플랫폼의 구축 / 민관협력의 사회적 협의체</td></tr>
<tr><td>오폭 등에 대한
피해 구제</td><td>사고 책임에 대한
법적 대응
절차 마련</td><td>운용
책임자로서의
윤리적 기준 준수</td></tr>
<tr><td>피드백 단계</td><td>☞</td><td></td><td colspan="3">공동체 차원의 리스크 관리
피해 유형의 검토</td></tr>
</table>

제 6 장 결론 및 정책 제언

제 1 절 본 연구의 결론적 논의

4차산업혁명이 본격화되는 가운데 인공지능 관련 윤리·법·규제 등 일련의 규범담론들이 형성, 확산되고 있지만, 규범적 대응방식이 입법 논의를 서둘러 제기하는 등 단기적 대응에 치중하는 현상이 여전히 극복되지 못하고 있다. 그래서 본 연구는 인공지능이 사람들의 사회규범에 미치는 영향에 대한 윤리적 고려 및 검토가 더 필요하다고 본다. 윤리적 논의가 실제 기술을 적용/활용하는 기업 및 시민사회 차원에서 좀더 활발하게 이루어져야 한다는 것이다. 이왕이면 추상적·일반론적 윤리담론에서 벗어나 산업 분야별 윤리논의로 좀더 구체적인 방향으로 전환하고, 가급적이면 향후 각 분야에서 인공지능 윤리 이슈가 발생할 것에 대비해 윤리적 대응 프레임워크를 모델링하고 이의 활용기준을 구체적으로 마련하는 방향으로 전개되어야 한다는 것이다.

본 연구에서 인공지능 윤리 이슈가 빈번하게 발생할 것에 대비한 정책 대응 프레임워크를 모델링하는 이유는 법적 규율을 성급하게 확정하는 것보다 윤리적 이슈에 대해 각계의 이해당사자들이 자율적으로 대응할 수 있는 사회경제적 역량을 형성, 축적하는 일이 더 중요하다는 판단에서이다. 그래서 인공지능 기술이 적용되는 분야의 특성을 고려한 단계별, 관련 행위자별로 구분해 대응방안을 구체적으로 논의할 수 있었다.

또한 본 연구는 인공지능 윤리 대응 프레임워크 구축시에 반드시 고려해야 할 사항을 밝히고 있는데 사실 이들은 인공지능 윤리의 기본 원칙이라고 해도 과언이 아닐 것이다. 즉 법적 규제와의 경계를 설정하는데 있어서 윤리적 대응이 산업성장을 위축시키지 말아야 하고, 최근 규범의 성격이 윤리와 법의 혼융·절충 추세에 있지만 규율의 대상 및 속성을 고려해 단계적으로 접근할 필요가 있다는 것이다.

결국 문제는 막연한 두려움과 추상적 예측에 입각한 일반적·총론적 접근이 아니라 각 분야에서 발생하는 인공지능 윤리 이슈에 대응할 수 있는 틀을 구성하고 이를 구체적으로 적용한 기준을 마련하는 것이라 하겠다. 즉 설계－개발－시험/임상－상용화－이용 등 인공지

능 기술의 구체적 적용 및 활용 단계에서 발생하는 다양한 인공지능 윤리 이슈들에 대해 인공지능 기술의 발전속도, 분야별 특성, 행위자 인식(규범준수) 수준 등을 고려한 윤리적 대응 프레임워크를 구축하고, 이를 각급 단위에서 활용가능한 '윤리적 체크리스트(ethical checklist)'를 만들어 보는 것이라 하겠다.

다시 말해 기술개발에서 서비스 상용화 등에 이르기까지 인공지능 생애주기의 전 과정에서 소위 '윤리적 레이어(ethical layer)'을 삽입시켜 윤리적·사회적 영향평가(impact assessment)를 체계화, 내면화하는 일이 중요하다는 것이다. 물론 그 과정에는 개발자, 일반 시민을 포함한 다양한 이해관계자의 자발적 참여와 소통, 즉 멀티스테이크홀더리즘(multi-stakeholdersim)의 원리가 작용해야 할 것이다.

제 2 절 정책 제언

앞에서도 논의했듯이, 본 연구의 가장 중요한 정책적 결론은 인공지능 윤리 이슈에 대한 공통의 정책 대응 프레임워크를 모델링하고, 발전단계, 행위주체, 쟁점별 특성을 고려해서 대응방안을 세분화했으며, 더 나아가 개별 산업 분야에서 적용할 수 있도록 대응방안을 구체화하였다는 점이다. 문제는 이와 같이 마련된 윤리대응 프레임워크 및 행위 기준을 다양한 이해당사자들과 함께 어떻게 실천적으로 적용할 것인가이며, 특히 이 과정에서 4차산업혁명위원회의 역할은 무엇인가 하는 점이다. 그래서 본 연구는 인공지능 윤리이슈에 대한 정책 대응방안을 보다 효과적으로 적용하고 이해당사자들의 자발적 참여와 소통을 유인하기 위해 다음과 같은 정책 방안들을 제언하고자 한다.

첫째, 4차산업혁명에 대한 윤리적 정책 대응은 인공지능 기술에만 국한되지 않는다는 것이다. 지능정보사회 또는 지능정보기술에 대비해 윤리 가이드라인 논의가 있듯이, 4차산업혁명시대의 신기술 윤리 관련 정책 연구대상 및 범위를 인공지능 외에도 사물인터넷, 빅데이터, 클라우드 컴퓨팅, 로봇, 스마트시티, 사이버 보안 등 여타의 기술 영역으로 확대해야 한다는 것이다. 이점과 관련해 본 연구는 4차산업혁명위원회 내에 '신기술 윤리분과'를 상설화하는 방안을 제안하고자 한다. 물론 4차산업혁명위원회가 모든 신기술의 윤리적 이슈에 대응하기는 쉽지 않을 것이다. 그래서 각계 전문가 및 이해관계자들의 의견 수렴 절차를 거쳐 매년 윤리적 이슈의 중요도 및 시급성에 비추어 1~2개의 주요한 신기술 및 서비스를 선

정하고 윤리적 대응기준을 마련하는 것이 바람직하다고 본다.

둘째, 본 연구에서 마련한 인공지능 윤리이슈 대응정책 프레임워크 및 적용 기준을 다양한 산업 분야 및 개별 서비스에 적용하기 위해서는 개발자, 사업자, 이용자단체 등 각 분야의 이해당사자들이 함께 모여 숙의하는 절차와 과정이 충분히 확보되어야 한다는 것이다. 이를 위해 본 연구는 1기 4차산업혁명위원회가 추진한 바 있는 '데이터 해커톤'의 경험을 살려 소위 '인공지능 윤리 해커톤'이라는 공론화 절차를 제안하고자 하며, 이를 관련 정부부처 및 전문가들과의 협업 하에 4차산업혁명위원회가 주도해야 한다고 본다. 인공지능 윤리 이슈 대응 방안에는 한가지의 정답만 있는 것은 아닐 것이다. 현재로서는 개발자, 전문가 차원의 다양한 윤리 논의도 좀더 필요한 상황이며, 앞으로 4차산업혁명위원회가 이같은 기업, 개발자, 전문가, 시민사회, 공적 주체가 참여하는 인공지능 윤리이슈 대응 거버넌스의 중심적 역할을 수행할 필요가 있다.

셋째, 인공지능의 성능이 고도화될수록 예상하지 못하거나 의도하지 않은 행위 위험성이 증가하는 등 더욱 복잡해지는 윤리적 이슈에 대응한 안전한 인공지능 개발・활용 기준의 구체적 준거를 마련해야 한다는 것이다. 다시 말하면, 인공지능 윤리 이슈를 안전한 인공지능 개발 기준의 정립과 결부시켜야 한다는 것이다. 이를 위해 인공지능 R&D 관련 정책 및 사업에 윤리적 협의 절차를 적극 포함시킬 필요가 있으며, IEEE와 같이 인공지능 윤리기준의 표준화를 구체화하는 국제기구와의 글로벌 협력도 추진할 필요가 있다. 예를 들면, 인공지능 기술의 제작과정 관리・감독, 윤리적 기준을 갖춰 출시되었는지의 점검, 필수 제한규정[239] 등과 같이 인공지능 설계자, 제품의 제공・공급자 및 이용자들이 안전하고 윤리적으로 인공지능을 이용할 수 있도록 산업별, 서비스별 인공지능 윤리 규범준수 기준을 구체적으로 제시할 필요가 있다는 것이다. 따라서 인공지능 판매 및 제공시 사전 설명과 동일한 기능・성능을 제공하는지에 대한 판단, 최적사용・권장사항・경고사항에 대한 판단, 권고 또는 제시된 표준 준수 여부, 알고리즘 추적성 판단 등 인공지능의 윤리적 건전성과 안전성을 평가하기 위한 표준을 관련 이해관계자들과 함께 제정하는 작업에 4차산업혁명위원회가 적극 관여할 필요가 있다.

넷째, 최근 인공지능 기술의 발전 속도가 가속화되는 추세이며 현 시점에서 인공지능의

239) '필수제한 규정'이란 금융당국이 대출신청 결정과정을 고객에게 설명하지 못하면 사업인가나 재정지원에 제한 두는 것처럼 정책입안자는 필요한 제한 규정을 두는 것을 말한다.

적용 및 자율적 판단이 점차 증가하는 상황이므로 보다 중장기적 관점에서의 윤리적 대응 방안을 마련하기 위한 학제 간 연구협력 체계를 강화해야 한다는 것이다. 인공지능의 사회적 파급효과 및 발전경로를 쉽게 예측하기 힘들므로 피해 발생 이후의 법적 처리를 상정하는 사후규제의 틀을 넘어 알고리즘 자체를 규제해야 한다는 견해가 제기되는 바, 인공지능 윤리 이슈를 알고리즘의 윤리적 규제원칙에 관한 논의로 구체화할 필요가 있다.[240] 다만, 알고리즘의 윤리적 규제 방향이 인공지능 관련 산업 발전 및 비즈니스 생태계에 지나친 규제 효과를 발생시키지 않도록 신중하게 논의할 필요가 있다.

다섯째, 인공지능 윤리가 4차산업혁명 시대의 디지털 시민역량의 핵심 요소라는 관점에서 윤리적 인공지능 개발 및 이용 역량 강화를 위한 다학제적 교육과정 및 학습 지원의 근거로도 활용 가능하도록, 디지털 교육과정에 인공지능 윤리 관련 사항을 적극 반영하는 권고안을 4차산업혁명위원회가 제시할 필요가 있다. 최근 4차산업혁명 대응정책의 일환으로 소프트웨어교육, 코딩교육 등 ICT역량에 초점을 둔 교육혁신 및 디지털 리터러시 교육이 강조되고 있으나, 인공지능의 윤리적 이슈에 대한 이해, 알고리즘의 윤리적 개발 기준 등에 관한 교육 내용을 중심으로 미디어 리터러시 또는 디지털 리터러시를 고도화할 필요가 있다.

240) 프로파일링을 포함한 자동화된 개인 의사결정(automated individual decision-making, including profiling)에 대해 언제든지 설명을 요구할 권리를 법제화한 EU의 GDPR 규정은 대표적인 예이며, 이는 사전예방의 원칙(precautionary principle) 하에서 제기된 '알고리즘 투명성(algorithmic transparecy)', '알고리즘 책무성(algorithmic accountability)' 등의 다양한 윤리담론을 부각시키는 계기를 마련하였다.

참 고 문 헌

〔국내문헌〕

고인석(2011), “아시모프의 로봇 3법칙 다시 보기: 윤리적인 로봇 만들기”,《철학연구》, 제93집, 철학연구회.

_____(2012a), “로봇이 책임과 권한의 주체일 수 있는가”,《철학논총》, 제67집, 새한철학회.

_____(2012b), “체계적인 로봇윤리의 정립을 위한 로봇 존재론, 특히 로봇의 분류에 관하여”,《철학논총》, 제70집, 새한철학회.

_____(2014), “로봇윤리의 기본 원칙: 로봇 존재론으로부터”,《범한철학》, 제75집, 범한철학회.

과학기술정보통신부(2018. 5),『인공지능(AI) R&D 전략』.

관계부처 합동(2016. 12),『지능정보사회 중장기 종합대책』.

권헌영・백승조・장규현・주문호・김법연・윤상필・김미리(2017),『지능정보사회에서의 인터넷윤리 기준 연구』, 방송통신위원회.

국토교통부・산업통상자원부・미래창조과학부(2015),『자율주행차 지원 방안』.

김대식(2015), “지능형 로봇기술의 발전 현황과 전망”,『법과학을 적용한 형사사법의 선진화 방안(VI)』, 한국형사정책연구원.

김동희(2016),『행정법 Ⅰ』, 박영사.

김만권(2014), “디지털 시민권과 시민정치의 활성화”, 《문화와 정치》 제1권 제2호, 한양대학교 평화연구소.

김유환(2016),『현대 행정법강의』, 법문사,

김윤명(2016),『인공지능과 리걸 프레임, 10가지 이슈』, 커뮤니케이션북스.

김은성(2010),『사전예방원칙의 정책타당성 분석 및 제도화 방안: 유전자변형생물체, 광우병, 테러를 중심으로』, 한국행정연구원

김정언・최계영・조유리・강준모・이학기・김민식・이은민・이시직・정원준・손가녕・양수연・최주한(2017),『제4차 산업혁명 선도를 위한 과학기술・ICT 기반 국가정책방안 연구』, 정보통신정책연구원.

김진구(2017), “식약처, 인공지능 ‘왓슨’ 의료기기 아니다”, 《조선일보》, (2017. 11. 23).

김형준(2017), “「지능정보사회 기본법(안)」 전문가 세미나 토론문”, 《지능정보사회 기본법 입법세미나》, 연세대학교 법학연구원 공공거버넌스와 법 센터.

김형회(2017), “인공지능의 의료적 활용과 생명윤리－의학적 관점에서 토론”, 《제1회 국가생명윤리포럼－인공지능(AI)의 의료적 활용과 생명윤리》, 국가생명윤리심의위원회・국가생명윤리정책연구원.

김환석(2006), 『과학사회학의 쟁점들』, 문학과지성사.

NIPA(2017), “지능정보사회 구현을 위한 심층통계 경제효과 및 세부이행전략 분석연구”.

뉴턴하이라이트121(2018), 『인공지능－기초부터 미래의 전망까지』, (주)아이뉴턴.

류현숙(2017), 『인공지능 기술 확산에 따른 위험 거버넌스 연구』, 한국행정연구원.

마쓰오 유타카(2015), 『인공지능과 딥러닝－인공지능이 불러올 산업구조의 혁신과 변화』, 박기원(역), 동아엠앤비.

맥스 테그마크(2017), 『Life3.0』, 백우진(역), 동아시아.

맥킨지(2016), “16년 지능정보기술을 통한 사회경제 혁신방안 연구”.

미래창조과학부(2016), 『ICT R&D 중장기 기술로드맵 2020』.

미야케 요이치로・모리카와 유키히토(2017), 『인공지능70－재미있게 알아보는 AI 키워드』. 김완섭(역), 제이펍.

박균성(2017), 『행정법강의』, 박영사.

박혜경(2017), “인공지능(AI)의 의료적 활용과 생명윤리의 몇 가지 쟁점－시민사회 관점의 논의”, 《제1회 국가생명윤리포럼－인공지능(AI)의 의료적 활용과 생명윤리》, 국가생명윤리심의위원회・국가생명윤리정책연구원.

배영임・최준규・신혜리(2018), “블록체인 기반 공공 플랫폼 구축을 위한 제언”, 《이슈&진단》, 경기연구원.

변순용・송선영(2012), “로봇윤리의 이론적 기초를 위한 근본 과제 연구”, 《윤리연구》, 제88호, 한국윤리학회.

브루노 라투르(2010), 『인간・사물・동맹』, 홍성욱(역), 이음.

선지원(2018), “대형 공공건설사업 추진 시 주민 참여에 대한 소고” 《경제규제와 법》 제11권 제1호, 서울대학교 법학연구소.

신동현(2016), “자율주행자동차 운행의 법적 문제에 관한 시론”,《과학기술법연구》, 제22집 제3호, 한남대학교 과학기술법연구원.

신상규(2016), “자율기술과 플로리디의 정보 윤리”,《철학논집》, 제45집, 서강대학교 철학연구소.

심우민(2017),“개인정보 비식별조치에 관한 입법정책적 대응과제” 국회입법조사처 현안보고서 305호-20170524) (2017. 5. 24), 국회입법조사처.

아주대 산학협력단(2016),『인공지능 및 로보틱스 분야의 개인정보 보호에 관한 연구』, 개인정보보호위원회 위탁연구과제 보고서.

양종모(2016), “인공지능의 위험의 특성과 법적 규제방안,”《홍익법학》, 제17권 제4호, 홍익대학교 법학연구소.

오도빈・유지원・양규원・구보민(2017), “탈후견주의적 시각에서 바라본 자율주행 자동차의 일률적 프로그래밍의 한계”,《고려법학》, 제84호, 고려대학교 법학연구원.

O. Etzioni(2016), “The Future of AI”,《2016 국제법률 심포지움: 4차 산업혁명의 도전과 응전: 사법의 미래》, 대법원.

월시, 토비(2018),『AI의 미래 생각하는 기계: 인공지능 시대 축복인가?』, 이기동(역), 도서출판 프리뷰.

웬델 월러치・콜린 알렌(2014),『왜 로봇의 도덕인가?』, 노태복(역), 메디치미디어.

윤지영・임석순(2011), “지능형 로봇시대를 대비한 국가별 전략 및 제도적 정비 현황”,『로봇윤리 공론화 전략 연구』, 한국로봇산업진흥원.

윤혜선(2017), “인공지능의 의료적 활용에 관한 법적・정책적 쟁점”,《제1회 국가생명윤리포럼－인공지능(AI)의 의료적 활용과 생명윤리》, 국가생명윤리심의위원회・국가생명윤리정책연구원.

이병준(2017), “자율규제기관인 GUCC와 자율규제로서의「이용자보호방안 가이드라인」의 의의와 한계”,《외법논집》 제41권 제4호, 한국외국어대학교 법학연구소.

이상돈・정채연(2017), “자율주행자동차의 윤리화의 과제와 전망”,《IT와 법연구》, 제15집, 경북대학교 IT와 법연구소.

_____(2017),『인공지능과 자율주행 자동차, 그리고 법』(7인 공저), 세창출판사.

이상형(2016), “윤리적 인공지능은 가능한가?”,《법과 정책연구》 제16권 제4호, 한국법정책학회.

이석배(2008), “생명윤리에 대한 형법적 보호의 범위와 한계”,《동아법학》 제42호, 동아대학교 법학연구소.

이언(2017), “인공지능의 의료적 활용 사례(왓슨)”,《제1회 국가생명윤리포럼－인공지능(AI)의 의료적 활용과 생명윤리》, 국가생명윤리심의위원회・국가생명윤리정책연구원.

이예하(2017), “인공지능의 의료적 활용 전망 및 한계”,《제1회 국가생명윤리포럼－인공지능(AI)의 의료적 활용과 생명윤리》, 국가생명윤리심의위원회・국가생명윤리정책연구원.

EY어드바이저리(2016),『2020년, 인공지능이 내 곁으로 다가왔다』, 부윤아(역), 매일경제신문사.

이원태(2015), “인공지능의 규범이슈와 정책적 시사점”,《KISDI Premium Report》, 15-7, 정보통신정책연구원.

_____(2018), “ ‘카카오 알고리즘 윤리헌장’의 의미: 개발자 자율규제 의지 천명... ‘이용자 윤리’ 보완 필요”, 월간 신문과 방송, 2018년 5월호.

이정구・민대환・권헌영(2017), “클라우드컴퓨팅 발전 및 이용자 보호에 관한 법률의 쟁점 및 개선방안”, JITAM, 사단법인 한국정보기술응용학회.

이종영・김정임(2015), “자율주행자동차 운행의 법적 문제”,《중앙법학》, 제17집 제2호, 중앙법학회.

이중기(2016a), “자율주행차의 발전단계로 본 운전자와 인공지능의 주의의무의 변화와 규범적 판단능력의 사전 프로그래밍 필요성”,《홍익법학》, 제17권 제4호, 홍익대학교 법학연구소.

_____(2016b), “자율주행자동차: 로봇으로서의 윤리와 법적 문제”,《국토》, 통권 제416호, 국토연구원.

_____(2016c), “인공지능을 가진 로봇의 법적 취급: 자율주행자동차 사고의 법적 인식과 책임을 중심으로”,《홍익법학》, 제17권 제3호, 홍익대학교 법학연구소.

이중기・오병두(2016), “자율주행자동차와 로봇윤리: 그 법적 시사점”,《홍익법학》, 제17권 제2호, 홍익대학교 법학연구소.

이중기・황창근(2016), “자율주행자동차 운행에 대비한 책임법제와 책임보험제도의 정비필요성: 소프트웨어의 흠결, 설계상 흠결 문제를 중심으로”,《금융법연구》, 제13권 제1호, 한국금융법학회.

이중원(2017), "'인공지능으로 인한 의료 패러다임 변화의 윤리적 측면'에 대한 토론", 《제1회 국가생명윤리포럼－인공지능(AI)의 의료적 활용과 생명윤리》, 국가생명윤리심의위원회・국가생명윤리정책연구원.

이호영・손상영・이원태・조성은・김희연・문정욱・이시직・양수연(2016), 『ICT 기반 사회현안 해결방안 연구』, 정보통신정책연구원.

장동경(2017), "'인공지능의 의료적 활용과 생명윤리'에 대한 토론－의학적 관점에서", 《제1회 국가생명윤리포럼－인공지능(AI)의 의료적 활용과 생명윤리》, 국가생명윤리심의위원회・국가생명윤리정책연구원.

장연화(2017), "왓슨의 진단 조력에 대한 현행법상 형사책임에 관한 소고", 《형사법의 신동향》, 제55권, 대검찰청.

정보문화포럼(2018), 지능정보사회 윤리가이드라인, 한국정보화진흥원.

정보통신기술진흥센터(2017), 『2016년 ICT 기술수준조사 보고서』.

정채연(2017), "'몇 가지 쟁점들과 법정책적 방향성'에 대한 토론", 《제1회 국가생명윤리포럼－인공지능(AI)의 의료적 활용과 생명윤리》, 국가생명윤리심의위원회・국가생명윤리정책연구원.

정하중(2016), 『행정법개론』, 법문사.

최경진(2017), "지능정보사회와 이용자 보호 이슈의 특징", 지능정보사회에서의 이용자 보호 이슈와 정책방향 학술세미나 자료집, KISDI－사이버커뮤니케이션학회 공동세미나.

최난설현(2013), "연성규범(Soft Law)의 기능과 법적 효력" 《인하대학교 법학연구》, 제16집, 2호.

최윤국(2003), "EU의 디지털 격차와 시민권－디지털경제로의 이행을 중심으로", 《유럽연구》 제18권, 한국유럽학회.

최윤섭(2018), 『의료인공지능』. 클라우드나인.

한국로봇산업진흥원(2018), '행위주체별 입장에서 본 로봇윤리', 로봇윤리포럼 자료집(2018. 9. 20).

한국무역협회(2018), "우리 기업의 인공지능(AI)을 활용한 비즈니스 모델", 《TRADE FOCUS》.

한국사전연구사 편집부(1994), 『컴퓨터정보용어대사전』, 한국사전연구사.

한국인공지능법학회・충남대학교 법학연구소(2018), '인공지능과 법: 기술, 법, 정책' 한국인공지능법학회 학술대회 자료집(2018. 5. 4).

한국인터넷윤리학회(2018), ‘인공지능시대 인터넷 윤리의 실제와 적용’, 2018년 한국인터넷윤리학회 춘계학술대회 자료집(2018. 5. 25).

한국인터넷진흥원(2017),『개인정보의 안전성 확보조치 기준(2016-35호) 해설서 (개정)』.

한국정보화진흥원(2017), “중국의 인공지능(AI) 전략”《과학기술 & ICT 정책기술동향》, No. 101.

_____(2018), “인공지능 악용에 따른 위협과 대응 방안”《Special Report》2018-12.

한국포스트휴먼연구소・한국포스트휴먼학회(2016),『포스트휴먼 시대의 휴먼』, 아카넷.

_____(2018),『제4차산업혁명과 새로운 사회윤리』, 아카넷.

한문승・이주헌(2014), “이용자보호 관점의 통신서비스 웹사이트 평가모형 연구”, 한국경영정보학회 학술대회

한희원(2011),『정의로운 산책: 정의론』, 삼영사.

_____(2018a),『인공지능(AI) 법과 공존윤리』, 박영사.

_____(2018b), “인공지능(AI) 치명적자율무기(LAWs)의 법적・윤리적 쟁점에 대한 기초연구”,《중앙법학》, 제20집 제1호.

_____(2018c), “인공지능(AI)의 법인격 주체 가능성의 이론적 기틀에 대한 기초 연구”,《중앙법학》, 제20집 제3호.

홍준형(2018),『시민을 위한 행정법 입문』, 박영사.

〔해외문헌〕

中華人民共和國, 工業和信息化部, 促進新一代人工智能産業發展三年行動計划 (2018-2020年), 2017年12月14日.

中華人民共和國, 國務院, 國務院關于印發, 新一代人工智能發展規划的通, 國發〔2017〕35号, 2017年7月8日.

Alessandro Mantelero(2018). “AI and Big Data: A blueprint for a human rights, social and ethical impact assessment” Computer Law & Security Review Vol 34, Issue 4 (Aug 2018)

Alison DeNisco Rayome(2017). “IEEE announces 3 AI standards to protect human well-being in the robot revolution” Tech Republic.

Alyssa Provazza(2017). “Artificial intelligence data privacy issues on the rise” TechTarget.

Anderson & Waxman(2012). “Law and Ethics for Robot Soldiers,”.

Association for Computing Machinery US Public Policy Council (USACM) “Statement on Algorithmic Transparency and Accountability” (Jan. 12, 2017).

Ben Dickson(2018). “How GDPR Will Impact the AI Industry” PC Review.

Bloor, D. 1976(1991). Knowledge and Social Imagery. University of Chicago Press.

Boris Glavic(2012). “Big Data Provenance: Challenges and Implications for Benchmarking” Revised Selected Papers of the First Workshop on Specifying Big Data Benchmarks – Vol. 8163. 72-80.

Brad Smith(2018). “Facial recognition technology: The need for public regulation and corporate responsibility”.

Bundesministerium für Arbeit und Soziales(2016), 『Wertewelten Arbeiten 4.0』

Cawley, J.(2003), “What Explains Race and Gender Differences in the Relationship between Obesity and Wages?“, Gender Issues, Summer/2003.

_____(2014), ‘The Obesity Wage Penalty’, The New York Times.

C. Heyns(2014). “Report of the UN Special Rapporteur on extrajudicial, summary or arbitrary executions”, A/HRC/23/47.

“California Enacts Anti-Bot and IoT Laws” The National Law Review (October 4, 2018)

California State Senate Bill No. 1001 CHAPTER 892. An act to add Chapter 6 (commencing with Section 17940) to Part 3 of Division 7 of the Business and Professions Code, relating to bots.

Cathy O’Neil(2018). “Audit the algorithms that are ruling our lives” Financial Times.

Christian Sandvig et al,(2014). “Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms” Conference proceeding at ‘Data and Discrimination: Converting Critical Concerns into Productive Inquiry’ of the 64th Annual Meeting of the International Communication Association.

Commission Nationale de l’informatique et des Libertés “Algorithms and artificial intelligence: CNIL’s report on the ethical issues” (25 May 2018)

Council on Foreign Relations “China’s Artificial Intelligence Strategy Poses a Credible Threat to U.S. Tech Leadership” (4 Dec. 2017)

D. J. Gunkel(2012). The Machine Question: Critical Perspectives on AI, Robots, and Ethics, The MIT Press.

Daniel Mikkelsen et al.,(2017). “Tackling GDPR compliance before time runs out” McKinsey & Company.

Daniel S. Hoadley & Nathan J. Lucas(2018). “Artificial Intelligence and National Security”, CRS Report.

Darrell M. West, John R. Allen(2018). “How Artificial Intelligence is Transforming the World” Brookings Institution.

David Roe(2018). “Understanding GDPR and Its Impact on the Development Of AI” CMS WiRE.

Department of Defense Directive(2012). “Autonomy in Weapon Systems”.

Erik Brynjolfsson, Andrew McAfee(2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies, W. W. Norton & Company.

European Commission(2016. 4. 19). “Digitising European Industry-Reaping full benefits of a Digital Single Market”.

_____(2018. 4. 10). “EU Declaration on Cooperation on Artificial Intelligence”.

_____(2018. 4. 25). “Artificial Intelligence for Europe”

_____(2018. 4. 25). “Data in the EU: Commission steps up efforts to increase availability and boost healthcare data sharing”.

European Commission(2018. 4. 25). “Factsheet: Artificial Intelligence for Europe”.

_____, “The Mid-Term Review on the implementation of the Digital Single Market Strategy“ Brussels, 10.5.2017 COM(2017) 228 final, (10 May 2017)

_____, European Group on Ethics in Science and New Technologies(2018). “Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems”.

Future of Life Institute(2018). 『Annual Report 2017』.

Gartner(2018. 4). “Forecast: The Business Value of Artificial Intelligence, Worldwide, 2017-2025”.

Government of United Kingdom(2018. 5. 21). ‘The Grand Challenges’.

_____, Department for Digital, Culture, Media & Sport “Centre for Data Ethics and Innovation Consultation” (13 June 2018).

Gubrud, Mark(2015). “Semi-autonomous and on their own: Killer robots in Plato’s Cave”. Bulletin of the Atomic Scientists.

Hans J. Wolff/Otto Bachof/Rolf Stober/Winfried Kluth(2017), 『Verwaltungsrecht Ⅰ』, C. H. Beck.

Hartmut Kreß(2012), “Grenzlinien zwischen Recht und Ethik”, ZRP, 28.

Hartmut Maurer/Christian Waldhoff(2017), 『Allgemeines Verwaltungsrecht』, C. H. Beck.

Hyeon-Suk Lyu(2013), “Internet Policy in Korea: A Preliminary Framework for Assigning Moral and Legal Responsibility to Agents in Internet Activities”, Government Information Quarterl, Volume 29, Issue 3, 394.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous systems) Version 2.

IEEE Standards Association, “Open Community for Ethics in Autonomous and Intelligent Systems (OCEANIS)” https://standards.ieee.org/news/2018/oceanis.html

IEEE, Open Community for Ethics in Autonomous and Intelligent Systems (OCEANIS). https://ethicsstandards.org

Information Commissioner’s Office(2017). “Big data, artificial intelligence, machine learning and data protection”.

_____(2017). “Royal Free NHS Foundation Trust-Google DeepMind trial failed to comply with data protection law”.

Internet Society(2017). “Artificial Intelligence and Machine Learning: Policy Paper” (April 27, 2017)

Jan Ziekow(2006), “Public Private Partnership - auf dem Weg zur Formierung einer intermediaeren Innovationsebene?”, Verwaltungsarchiv.

J. C. Gerdes & S. M. Thornton(2016). “Implementable Ethics for Autonomous Vehicles,” Autonomous Driving: Technical, Legal and Social Aspects(M. Maurer & J. C. Gerdes & B. Lenz & H. Winner(eds.)), Springer.

J.-F. Bonnefon & A. Shariff & I. Rahwan(2015). “Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?”. Retrived from https://pdfs.semanticscholar.org/13d4/56d4c53d7b03b90ba59845a8f61b23b9f6e8.pdf.

Jenna Burrell(2016). “How the machine ‘thinks’: Understanding opacity in machine learning algorithms”. Big Data &Society, 3(1).

JF Zabriskie(2009). “Bots, Scrapers, and Other Unwanted Visitors to Your Web Site: Can You Keep Them Out?” The Computer & Internet Lawyer, Vol 26, No. 7.

Jiweon Seon(2018), 『Die Kontrolle von Public Private Partnerships im Beschaffungswesen im Rahmen des europäischen Beihilfen- und Vergaberechts』, Peter Lang.

Krishnan, Armin(2009). Killer robots: Legality and ethicality of autonomous weapons (Aldershot: Ashgate).

Latour, B(1987), “Science in Action: How to Follow Scientists and Engineers through Society”, Havard University Press.

_____(1993), “We Have Never Been Modern”, Havard University Press.

Latour, B. & Woolgar, S. 1979(1986), “Laboratory Life: The Construction of Scientific Facts”, Princeton University Press.

Lee Rainie & Janna Anderson(2017). “Code-Dependent: Pros and Cons of the Algorithm Age” Pew Research Center.

Macy Bayern(2018). “DeepMind, NHS use anonymized patient data in AI to avoid regulatory hurdles” Tech Republic.

Mary Wareham(2018). “Support Grows for Killer Robots Ban US and Russia Among States Who Object to Prohibition” Human Rights Watch.

Matt Chessen(2017). “The AI Policy Landscape” Medium.com.

Matthias Scheutz(2012). “The Affect Dilemma for Artificial Agents: Should We Develop Affective Artificial Agents?”IEEE Transactions on Affective Computing, Vol 3, Issue4.

McLean, R. & Moon, M.(1980), ‘Health, Obesity, and Earnings’, PUBLIC HEALTH BRIEFS, September 1980, Vol. 70(No. 9).

Melvin Eisenberg(2005). “The Architecture of American Corporate Law: Facilitation And Regulation”, 2 Berkeley Bus. L. J. 167, 182.

Michael Anderson, Susan Leigh Anderson(2007). “Machine Ethics: Creating an Ethical Intelligent Agent”. AI Magazine. American Association for Artificial Intelligence. 28 (4): 15－26.

Miles Brundage et al.,(2018). “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation”.

N. J. Goodall(2014). “Ethical Decision Making During Automated Vehicle Crashes”, Transportation Research Record: Journal of the Transportation Research Board, No. 2424, Transportation Research Board of the National Academies.

Nick Wallace and Daniel Castro(2018). “The Impact of the EU’s New Data Protection Regulation on AI” Center for Data Innovation.

P. Lin(2013). “The Ethics of Saving Lives with Autonomous Cars Are Far Murkier than You Think,” Wired Opinion. Retrived from https://www.wired.com/2013/07/the-surprising-ethics-of-robot-cars.

_____(2016). “Why Ethics Matters for Autonomous Cars”, Autonomous Driving(M. Maurer & J. C. Gerdes & B. Lenz & H. Winner(eds.)), Springer.

Paul Mozur(2018). “Inside China’s Dystopian Dreams: A.I., Shame and Lots of Cameras” New York Times.

Paul Mozur, John Markoff(2018). “Is China Outsmarting American Artificial Intelligence?”.

Register, C. & Williams, D.(1990), ‘Wage Effects of Obesity among Young Workers’, SOCIAL SCIENCE QUARTERLY, Vol. 71(No.1).

Samm Sacks(2018). “New China Data Privacy Standard Looks More Far-Reaching than GDPR” Center for Strategic and International Studies.

Scott Shane, Cade Metz and Daisuke Wakabayashi(2018). “How a Pentagon Contract Became an Identity Crisis for Google” New York Times.

Sendhil Mullainathan, Ziad Obermeyer(2017). “Does Machine Learning Automate Moral Hazard and Error” American Economic Review: Papers & Proceedings 2017, 107(5).

Simpson, Thomas W & Muller, Vincent C(2016). “Just war and robots’ killings”. Philosophical Quarterly.

Tractica(2017). “Enterprise AI Revenue by Region, World Markets: 2016-2025”.

UNIDIR(2017). “The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence”.

UNIRIR(2018). “Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies”

United Kingdom Govenment Policy paper ‘AI Sector Deal’ (26 April 2018); Policy paper ‘The

Grand Challenges' (21 May 2018)

United Kingdom, The House of Lords Select Committee on Artificial Intelligence Report of Session 2017–19 "AI in the UK: ready, willing and able?" (16 April 2018)

Will Knight(2017). "Forget Killer Robots-Bias Is the Real AI Danger" MIT Technologu Review.

_____(2017). "China's AI Awakening 中國 人工智能 的崛起 " MIT Technologu Review.

Women's International League for Peace and Freedom(2014). "Statement to the CCW Meeting of Experts on Lethal Autonomous Weapon Systems".

World Economic Forum(2018. 9. 20). "United Kingdom Partners with World Economic Forum to Develop First Artificial Intelligence Procurement Policy".

Yale Law School, Information Society Project(2017). "Governing Machine Learning:Exploring the Intersection Between Machine Learning, Law, and Regulation".

Zi Chu et al.(2012). "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?" IEEE Transactions on Dependable and Secure Computing 2012 Vol. 9 Issue No. 06–Nov.-Dec.

〔기타〕

《조선일보》(2018. 4. 13), "1호 '인공지능의사' 탄생…진단서도 척척".

《중앙일보》(2018. 2. 17), "북핵보다 치명적인 인공지능 무기", Retrived from https://news.joins.com/article/22374485.

《한겨레》(2018. 4. 5), "카이스트 보이콧" 선언한 세계 학자들, AI무기 개발 비판, Retrived from http://www.hani.co.kr/arti/science/science_general/839279.htm.

한국인터넷자율정책기구 국문 브로셔(2017).

《Huffingtonpost》(2013. 7. 15). "Killer Robots and the Rule of Law", Retrived from https://www.huffingtonpost.com/A-View-from-the-United-Nations-/killer-robots-and-the-rul_b_3599657.html.

《WIRED》(2017. 9. 27), "Firewalls Don't Stop Hackers, AI Might", Retrived from https://www.wired.com/story/firewalls-dont-stop-hackers-ai-might/.

http://moralmachine.mit.edu/)

https://www.theguardian.com/technology/2016/feb/29/google-self-driving-car-accident-california

https://electrek.co/2016/05/26/tesla-model-s-crash-autopilot-video

http://www.army-guide.com/eng/product4196.htm.

https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/.

https://www.dropbox.com/s/g4ijcaqq6ivq19d/2017%20Open%20Letter%20to%20the%20 United%20Nations%20Convention%20on%20Certain%20Conventional%20Weapons.pdf?dl=

http://www.stopkillerrobots.org/wp-content/uploads/2016/04/CountryViews_14Apr2016.pdf.

https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/.

https://www.defense.gov/News/Article/Article/938941/national-security-experts-examine-intelligence - challenges-at-summit/.

https://www.konsultation-ki.de/KI/Redaktion/DE/Artikel/fortschritt-durch-kuenstliche-intelligenz.html

http://www.piovan.com/en/winfactory4_0

https://www.plattform-i40.de/I40/Navigation/EN/Industrie40/industrie40.html

https://www.stopkillerrobots.org/

● 저 자 소 개 ●

이 원 태
- 서강대 정치외교학과 졸업
- 서강대 정치학 박사
- 현 정보통신정책연구원 연구위원 및 디지털사회정책 그룹장

선 지 원
- 한양대 법학과 졸업
- Universität Regensburg 법학 박사
- 현 정보통신정책연구원 부연구위원

박 혜 경
- 현 충남대 사회학과 교수

정 채 연
- 현 포항공대 인문사회학부 교수

한 희 원
- 현 동국대 법학과 교수

김 정 언
- 고려대 경제학과 졸업
- Iowa State Univ. 경제학 박사
- 현 정보통신정책연구원 ICT전략연구실장

이 시 직
- 연세대 법학과 졸업
- 연세대 법학 석사(박사과정)
- 현 정보통신정책연구원 연구원

안 수 현
- 현 한국외대 법학전문대학원 교수

최 은 창
- 현 프리인터넷 프로젝트 펠로우

4차산업혁명시대 산업별 인공지능 윤리의
이슈 분석 및 정책적 대응방안 연구

2018년 10월 일 인쇄
2018년 10월 일 발행

발행인 김 대 희
발행처 정 보 통 신 정 책 연 구 원
충청북도 진천군 덕산면 정통로 18
TEL: 043-531-4114 FAX: 043-535-4695～6
인 쇄 인 성 문 화
ISBN 12-1071400-000016-01