

개방적/인간친화적 인공지능 체계 기술 표준화 동향¹

한국전자통신연구원 전종홍 · 이승윤

1. 서론

최근 빅데이터/클라우드, 스마트기기, GPU 기술 등의 급격한 발전 속에서 신경회로망과 딥러닝 기술이 보편화되면서 다양한 산업 분야에서 인공지능 기술을 결합시키는 다양한 사례들이 나오고 있다. 제조, 교통, 의료, 금융, 농업 등 수많은 산업 분야에 인공지능이 적용되면서 그동안 경험하지 못했던 새로운 변화들을 발생시키고 있고, 이로 인해 인공지능 적용과 보편화는 새로운 4차 산업혁명을 촉발시키는 핵심 기술 요소로까지 평가 받고 있다.

인공지능은 미래 ICT 기술융합 트렌드를 주도하는 핵심 기술로 전망되기에 아마존, 구글, IBM, Microsoft, Nvidia 등 글로벌 IT 기업들은 인공지능 관련 스타트업들을 인수하고, 사업 조직을 신설하는 등 인공지능을 활용한 다양한 비즈니스 기회를 발굴하고 이를 선점하기 위한 많은 노력들을 하고 있다. 시장에서는 로봇, 자율주행차, 드론, 챗봇 등 인공지능 기술을 결합시킨 다양한 제품들이 빠르게 등장하고 있으며, 빠른 속도의 발전을 거듭하고 있다.

인공지능 기술의 발전 속도도 빨라져 인공지능이 인간 꺾는데 퀴즈 7년, 체스 30년, 바둑은 47년 정도만에 인간의 능력을 넘어서는 수준까지 도달하였다. 이러한 인간과 기계의 전반적인 수준차(Human-machine parity)가 2050년이면 사람의 50% 수준, 2075년이면 사람의 90% 수준까지 따라올 것이라 예상되고 있기까지 하다 [1].

인공지능의 빠른 발달은 경제적·사회적 효과에 대한 기대뿐 아니라 자동화로 인한 일자리 대체, 통제 불능 문제 등 부정적 영향에 대한 우려감도 높이고 있다. 세계경제포럼(WEF)은 2017년 위험보고서에서 인공지능의 위험성을 주요 위험요소로 꼽았고, 엘런 머스크, 스티븐 호킹, 빌 게이츠와 같은 여러 전문가들도 인공지능의 위험성과 인류의 미래에 대해 불안감을 경고한 바 있다 [2, 10, 21].

이처럼 인공지능의 빠른 진화 속에서 100년 이내에 인공지능(人工自我)과 자율성(自律性), 그리고 자기학습 능력까지 갖는 초지능체(Superintelligence)로 발전할 가능성은 점점 높아지고 있다. 이에 인간의 통제를 벗어난 자율적 초인공지능체 등장으로 인한 여러 위험들을 사전에 예방하고 인류의 안전성을 보장하도록 해야 한다는 목소리도 높아지고 있다.

그동안 인간이 발명하였던 세탁기, 자동차 및 비행기가 인간을 위해 안전해야 하는 것과 똑같은 이유인 인공지능 기반의 다양한 기술들도 인간을 위한 친화성과 안전성이 매우 높은 기준하에서 설계되어야 한다. 본 고에서는 이러한 개방적/인간친화적 인공지능 체계를 위한 국내외 동향을 살피고, 이를 통해 향후 검토 및 대응이 필요한 주요 이슈들을 살펴보고자 하였다.

¹ ※ 본 연구는 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행하였음 (No. R0127-16-1042, 라이프케어 위한 스마트 웨어러블 표준 개발).

2. 인공지능의 진화에 따른 이슈들

AI(Artificial Intelligence)라는 용어 자체는 1956년에 최초로 만들어졌지만 인공지능의 역사는 1940년대로 거슬러 올라간다. 인공지능 아이디어는 Alan Turing이 1950년 발행한 논문인 "Computing Machinery and Intelligence"에서 본격적으로 시작되었다. Turing은 논문을 통해 "과연 기계는 생각을 할 수 있을까?"라는 질문을 제기했고, 기계가 어린아이의 학습 능력 수준을 갖출 수 있다는 가능성을 제시했다. 이후 인공지능 연구는 수십 년간 우여곡절을 거쳐오다 1990년 말 영상인식과 의학 진단 등 실용적인 응용분야에 연구가 집중되면서 관련 연구들이 활발해지기 시작했다 [3].

버클리대 설(Searle, 1980) 교수는 인공지능을 인간의 일을 얼마나 수행할 수 있는지에 따라 弱인공지능과 强인공지능으로 분류하였다. 전자는 어떤 문제를 실제로 사고하거나 해결할 수 없는 컴퓨터 기반의 인공적인 지능을 만들어 내는 것으로 세탁기 퍼지기능, 로봇청소기, 전력관리 등과 같이 자율성이 없는 인간의 다양한 능력 가운데 일부만 구현한다. 후자는 어떤 문제를 실제로 사고하고 해결할 수 있는 컴퓨터 기반의 인공적인 지능을 만들어 내는 것으로 인간의 지능이 가지는 학습·추리·적응·논증 기능을 갖춘 자율성이 있다 [4].

인공지능의 진화방향을 弱인공지능에서 스스로 사고·판단·예측, 스스로 학습·진화, 두뇌를 모사하는 인지컴퓨팅 등 强인공지능 기술로 진화할 것으로 전망되고 있다. 전략 게임, 번역, 자율주행, 영상인식 분야에 적용되는 약한 인공지능은 그동안 놀라운 발전을 거뒀다. 약한 인공 지능은 여행 계획, 소비자 추천 시스템, 광고 타겟팅과 같은 많은 상업 서비스들을 가능케 했고, 최근에는 의학 진단, 교육, 과학 연구 분야에서 더욱 중요한 응용 기회를 모색하면서 다양한 변화들을 촉발시키고 있다 [3, 4, 10].

인공지능 분야의 기술 스펙트럼이 넓고 다양하지만 최근 전반적인 인공지능 기술 진화 과정에서 등장하고 있는 이슈들을 요약해보면 크게 4가지 범주로 나눌 수 있다.

- 1) 개방성 : 인공지능 기술은 기존의 전통적 ICT 산업의 테두리를 넘어 전산업 및 사회의 모든 영역에 걸쳐 응용 가능하기 때문에 고도의 융합적 속성을 지니고 있다. 이러한 융합적 속성을 갖는 인공지능 기술을 다수의 협력적 모델 속에서 보다 빠르게 혁신하며 개선하기 위해서는 좀더 개방적인 구조로 만들어야 된다는 공감대가 높아지고 있다. 과거 전통적인 연구방식이 소수 연구그룹 중심의 폐쇄적인 연구를 진행하였다면, 최근 구글, 페이스북, MS, Open AI 등에서는 인공지능 플랫폼과 학습에 필요한 데이터들을 오픈소스와 오픈데이터로 개방하고, 이를 통해 취약성과 위험성을 공동으로 발굴하고 공동의 진화를 이루자는 시도들을 추진하고 있다. 특히 인공지능 지능처리 능력의 핵심적인 요소는 다양한 데이터의 효과적인 활용이 필요하기에 데이터 개방과 활용에 대한 이슈 또한 중요한 이슈라 할 수 있다. 개방적 인공지능 체계에 대한 관심이 높아지는 이유는 개방성을 높이는 것만이 인공지능의 활용시 발생할 수 있는 취약성과 위험성을 줄이고, 빠른 혁신이 가능하도록 하는 좋은 방법이기 때문이다 [5, 19].
- 2) 신뢰성 : 딥러닝 등의 인공지능 관련 기술들은 대부분 설명가능성(explainability)와 투명성(transparency)의 문제를 갖고 있다. 딥러닝 기술은 블랙박스과 같아서 학습시킨 데이터를 기반으로 통계적으로 답을 찾아내고, 사후 수학적 모델링을 통해 최종 결과를 다시 검증하지만, 그럼에도 왜 그런 결과가 나왔는지, 각 노드별로 어떤 역할을 했는지는 완벽하게 설명하기는 어려운 상황이다. 특히 향후 인공지능에 대한 의존성이 확대됨에 따라 개인의 삶에 중요한 영향을 미치는 많은 의사결정들이 인공지능을 통해 처리될 경우에 신뢰성 확보는 더더욱 중요한 이슈가 된다. 그러나 사용자 수준에서 인공지능 관련 제품의 신뢰성을 확인하기는 불가능에 가깝기 때문에 인공지능 기술이 적용된 제품의 지능처리 과정이 정확하고 안전하게 작동

하는지 신뢰성을 객관적으로 시험할 수 있는 기준과 방법, 체계가 필요하다. 특히 다양한 사업자와 기업의 기술을 활용하는 인공지능 생태계에서 인공지능 기술의 신뢰성 평가를 위해서는 더욱더 객관성을 갖는 공통의 표준과 평가체계가 필요하다 [3, 5, 10, 19].

- 3) 자율성 : 일반적으로 자율성(Autonomy)이란 어떠한 행위를 인간의 통제력 없이 오직 시스템으로만 제어하는 것을 의미하지만, 기술적인 측면에서보면 다양한 조건과 주변 환경을 관측(Observe)하고, 판단(Orient)해서, 결정(Decide)한 후 행동(Act)하는 OODA 의사 결정 루프(Loop)를 거쳐 처리할 수 있는 능력을 의미한다. 예를 들어, 완전 자율주행차량은 인간의 간섭이나 개입없이 자율적인 통제를 거쳐 목적지에 도착한다. 이러한 자율성은 차량, 항공기에 적용될 수 있지만, 자동 금융 트레이딩과 자동 콘텐츠 큐레이션 시스템같은, 더 큰 시나리오에 적용될 수도 있다. 또한 자율성은 운영 중 자가진단을 거쳐 결함시 자가 조치를 취하거나, 보안 취약점을 발견해 자가 조치를 하는 것을 가능하도록 한다. 인간 개개인이 감당하기 힘들만큼 엄청난 양의 정보를 인간보다 훨씬 빠른 속도로 처리할 수 있는 능력 덕분에 인공지능의 의사결정 과정에 인간이 개입할 여지가 줄어들고 있고, 이는 자율성의 증가로 이어지고 있다. 자율성의 증가는 판단, 결정, 행동의 과정과 결과에 대한 책임성으로 이어지고, 자율성의 수준과 그 결과에 따른 다양한 윤리적/규범적/법적 문제 등과 같은 실제적 이슈들을 발생시키고 있다 [3, 6, 7, 8, 9, 10, 11]
- 4) 협력성 : 자동화(Automation)는 스스로 움직이게 만드는 능력이다. 인공지능 기술을 통해 자동화 능력은 갖는 다양한 기계들이 등장하면서 인간을 보조하는 수준을 넘어 인간의 일자리를 대체는 단계에까지 발전하고 있다. 많은 전문가들은 당분간 범용 인공지능의 실현은 힘들 것이라고 판단하지만, 반면 어떤 이들은 '지능 폭발'과 '기술적 특이점(Singularity)'이 일어나 기계가 인간의 지능 수준을 넘어 빠르게 진화할 것이라고 예상한다. 긍정적으로 미래를 보는 연구자들은 인간의 협력자, 어시스턴트, 팀메이트, 트레이너가 되는 방식으로 안전하고 도덕적으로 인공지능이 사용될 것이라고 생각하기도 한다. 로봇 기술과 인공지능 기술의 발달과정에서 인간과 기계가 경쟁할 것인가? 아니면 인간과 기계가 서로의 단점을 보완하여 협동하면 효율성을 높이는 방향으로 협력할 것인가? 만약 협력한다면 어떤 모델로 상호협력하는 모델을 만들 것인가? 인간과 인공지능의 지위와 역할은 어떻게 규정되며 어떻게 책임과 보상을 나누고 협력해야 하는가와 같은 다양한 이슈들이 등장하고 있다 [3, 6, 7, 10, 22].
- 5) 안전성 : 인공지능이 생활에 밀접하게 적용될수록 중요하게 확인해야 할 사항이 안전성이다. 자율적 의사결정 기능을 부여한 인공지능이 설계 시 고려하지 못했던 조건에 대해 오작동을 일으키거나 최초에 의도했던 대로 작동하지 않으면 인명피해나 재산손실을 유발할 가능성이 크다. 그러므로 우리가 일상에서 사용하는 세탁기, 자동차 및 비행기가 안전해야 하는 것과 똑같이 인공지능 기술 요소들도 안전성이 매우 높은 표준으로 설계되어야 하고 검증 가능해야 한다. 이를 위해서는 주요 항목별로 위험요소를 도출한 후 이에 필요한 안전 요소와 요구사항을 정의하고, 이에 기반한 국제 표준 수준의 공통 안전기준과 이를 시험/검증하고 인증할 수 있는 체계를 만드는 것이 필요하다 [10, 12, 13, 14, 15, 19].

이처럼 인공지능 진화과정에서 발생하고 있는 다양한 이슈들을 해결하고 인간에게 보다 편리하고 안전하며 친근한 인공지능 기술이 될 수 있도록 하기 위한 노력들을 통칭하여 개방형/인간친화적 인공지능이라 한다 [16].

3. 국제 동향

3.1 인공지능의 윤리/규범/법 제도 동향

자율주행자동차, 의료분야 등에서 이미 인공지능, 로봇 기술이 상용화단계에 진입하고 있으며, 머지않은 미래에 인공지능 기술은 단지 인간의 생활을 조력하는 단계를 넘어 인간의 삶의 일부로 편입되는 이른바 포스트휴먼(post-human)의 시대가 도래 할 것으로 전망되고 있다. 단순히 집적된 데이터의 활용을 넘어서서, 스스로 데이터를 분석 및 학습하고 이에 기반해 특정 목표를 실행에 옮길 것으로 예견되는 인공지능의 확산은 규범 환경 또는 구조적 측면에서 <표1>과 같이 과거와는 다른 새로운 차원의 문제를 제기하면서 새로운 규범체계 정립의 필요성을 높이고 있다. 특히 킬러로봇의 군사적 이용, 드론 해킹 사고, 자율주행자동차 법 적용 논란 등 최근 인공지능과 로봇 확산에 따른 인간권리 침해 가능성이 높아지면서 윤리적 디자인의 차원을 넘어 법제도적 규제체계를 마련해야 한다는 논의도 증가하고 있다 [7].

구분	세부 이슈	주요 내용
인공지능의 자율성 범위	로봇/인공지능의 법적 책임 문제	-로봇도 도덕주체(moral agent)가 될 수 있는가, 즉 인공지능 로봇이 범죄를 저지를 경우 인간과 마찬가지로 로봇에게도 사법처리 가능한가 -로봇/인공지능 관련 기기의 법적 권리 여부를 검토하고 '제조물 책임법' 등 기존 법체계의 한계점 개선 방안 마련
	로봇/인공지능의 인격성 부여문제	-의인화 기술의 발전에 의해 친밀감, 동질감을 제공하는 인공지능 로봇의 인격성을 어디까지 인정해야 하는가
	인공지능 알고리즘의 신뢰성 문제	-인공지능 알고리즘에 기반한 자율적 의사결정(판단, 계약, 거래 등)을 얼마나 신뢰할 수 있는가
	인공지능의 안전성 문제	-인공지능 관련 소프트웨어, 프로그램의 오류나 오작동의 문제를 기술적으로 최소화하고 이를 규제하기 위한 방안 마련
	인공지능의 신뢰할만한 평가체계 구축문제	-인공지능의 사회문화적 영향을 체계적으로 평가하고 활용할 수 있는 가이드라인 마련 -인공지능, 로봇 등의 자율성 범위를 제한하거나 규정하기 위한 사회적 합의 틀 마련
인간의 권리침해 및 통제권	지능정보사회의 인권 개념 재정립 문제	-인공지능 확산에 따른 인간의 존엄성과 인권에 대한 가치 혼란이 발생하므로 지능정보사회를 살아가는 인간의 기본권 개념을 재정립할 필요
	사생활 침해 및 프라이버시 보호 문제	-인공지능 관련 데이터의 연결 및 공유가 확산됨에 따른 프라이버시 침해를 최소화하고 피해를 보상하기 위한 기제 마련
	인공지능 알고리즘의 책임성 강화 문제	-인공지능 관련 기술 및 제품 설계-생산과정에서 알고리즘 자체의 도덕적 책임성을 적극 부과하는 원칙이나 기준 마련
	인공지능 관련 갈등조정 거버넌스 구축 문제	-인공지능 규제와 관련한 이해당사자들 간의 이익갈등 및 권리상충 문제를 효율적으로 조정, 해결하기 위한 거버넌스(법제도 등) 구축
	인공지능 확산에 따른 정보격차 해소문제	-인공지능 활용에 따른 사회적 양극화 및 정보 격차를 해소하기 위해 '코드 리터러시(code literacy) 함양 등 지능정보사회 교육체계 구축
	창의성 등 인간고유의 역량 강화 문제	-일상적 의사결정을 인공지능 기계에 위임할 경우 인간의 비판적 사고가 약화될 우려가 커지므로 창의성, 통찰력 등 인간고유의 역량을 증대시키기 위한 정책 추진

<표1: 인공 지능의 규범적 주요 이슈들 [7]>

주요 국가별로도 인공지능의 확산에 따른 사회경제적 변화를 연구하고 대응하기 위한 다양한 노력들을 추진하고 있다.

1) 미국

미국 백악관은 2016년 10월 'AI의 미래를 위한 준비(Preparing for the Future of Artificial Intelligence)'라는 제목의 인공지능(AI) 특별보고서를 발표하는데 이어 2016년 12월말에는 인공지능이 향후 수십 년 동안 미국 경제에 끼칠 영향을 예측한 "인공지능과 자동화가 경제에 미치는 영향(Artificial Intelligence, Automation and the Economy)" 보고서를 발표했다. 뒤를 이어 국가과학기술위원회(NSTC) 산하 네트워킹·IT R&D 소위원회(NITRD)는 인공지능 R&D 전략계획을 수립하여 발표하였다. 그리고 인공지능 분야 연방연구 R&D의 체계적 지원을 위해 다음과 같은 7대 전략을 제시하였다 [3, 6].

- ① 인공지능 R&D에 대한 장기적 투자
- ② 인간-인공 지능 협업을 위한 효과적 방안 개발
- ③ 인공지능의 윤리적·법적·사회적 함의 이해 및 대응
- ④ 인공지능 시스템의 안전 보안 확보
- ⑤ 인공지능 훈련·시험용 공공데이터 공유 환경 조성
- ⑥ 표준 및 벤치마킹을 통한 인공지능 기술 측정 및 평가
- ⑦ 국가적 인공지능 R&D 인력수요 파악

2) 유럽연합

유럽연합에서는 로봇 관련 법규 및 규제대응을 위한 정책연구 프로젝트(2012.3-2014.3)로 '로봇법(RoboLaw)' 프로젝트를 진행하고 이를 통해 '로봇규제 가이드라인(Guidelines on Regulating Robotics)'를 도출하였다. EU의 로봇법 프로젝트는 로봇기술의 법률적, 윤리적 이슈 검토를 통해 새로운 규범체계를 정립하고자 하는 연구목표 하에 이탈리아, 네덜란드, 영국, 독일 등 4개국 4개 연구소가 참여했고 특히 자율주행차, 수술로봇, 로봇인공기관, 돌봄 로봇 등 4가지 연구대상의 윤리적, 법률적 분석을 통해 규제정책의 근거를 마련하고자 하였다. 유럽연합의 로봇규제 가이드라인은 로봇의 잠재적 위험에 대한 법적 책임 원칙도 적극 논의할 것을 강조하고 있다. 로봇 기술의 복잡성 및 자율성 증가로 인해 인간의 통제력 감소뿐만 아니라 책임 소재도 불분명해지고 있기 때문에, 개발자(설계자), 제조사(생산자), 소유자, 사용자 등 주체 간 법적 책임 배분을 어렵게 만들고 있다는 점 때문이다 [9].

3) 일본

일본 총무성 정보통신정책연구소는 2016년 인공지능(AI)을 사용한 네트워크 사회의 영향과 과제를 검토하기 위한 "AI 네트워크화 검토회의"를 운영한 후 2016년 6월에 20가지 항목의 구체적인 리스크를 보고서로 정리하였다. 2016년 4월 일본에서 개최된 주요 7개국(G7) 정보통신장관회의에서는 인간이 AI를 안심하고 안전하게 사용 할 수 있도록 AI 개발 원칙 방안에 대한 논의를 안건으로 제안하였다. 2016년 10월부터는 "AI 네트워크화 검토회의"를 "AI 네트워크 사회 추진 회의"로 확대하여 사회 전체의 AI 네트워크화 추진을 위한 사회 경제적·윤리적·법적 문제를 종합적으로 검토하는 작업을 추진 중에 있는 등 AI 안전이용 분야에서 주도권을 잡기 위한 여러가지 시도들을 추진 중에 있다 [4].

일본 총무성은 또한 기업체가 개발하는 AI 안전성과 보안성을 평가하는 공적 인증제도를 운용할 계획을 수립 중에 있다. 일본 정부는 인증제도 도입으로 AI 사용자와 개발자 책임범위를 명확히 하고, 안전을 담보해 AI 개발과 보급을 촉진시키며, 인증 평가는 정부가 아닌 제3의 기관에서 담당하도록 할 계획이다. 현재 AI 연구개발과 관련한 특별한 규제나 지침은 없지만 일본 총무성에 설치할 전문가 회의에서 상반기 중 연구개발지침을 개발하고, 2018년 법개정과 함께 공적인증 제도를 개시한다는 목표로 있다 [17].

4) 기타

IEEE(국제전기전자기술자협회)에서는 인공지능의 개발은 한쪽으로 편향되지 않으며 궁극적으로 인류에 혜택을 주는 "윤리적 디자인"을 추구하여야 하며, 이러한 활동을 장려하기 위해 학계, 정부, 기업 등에서 인공지능, 법, 윤리, 철학, 정

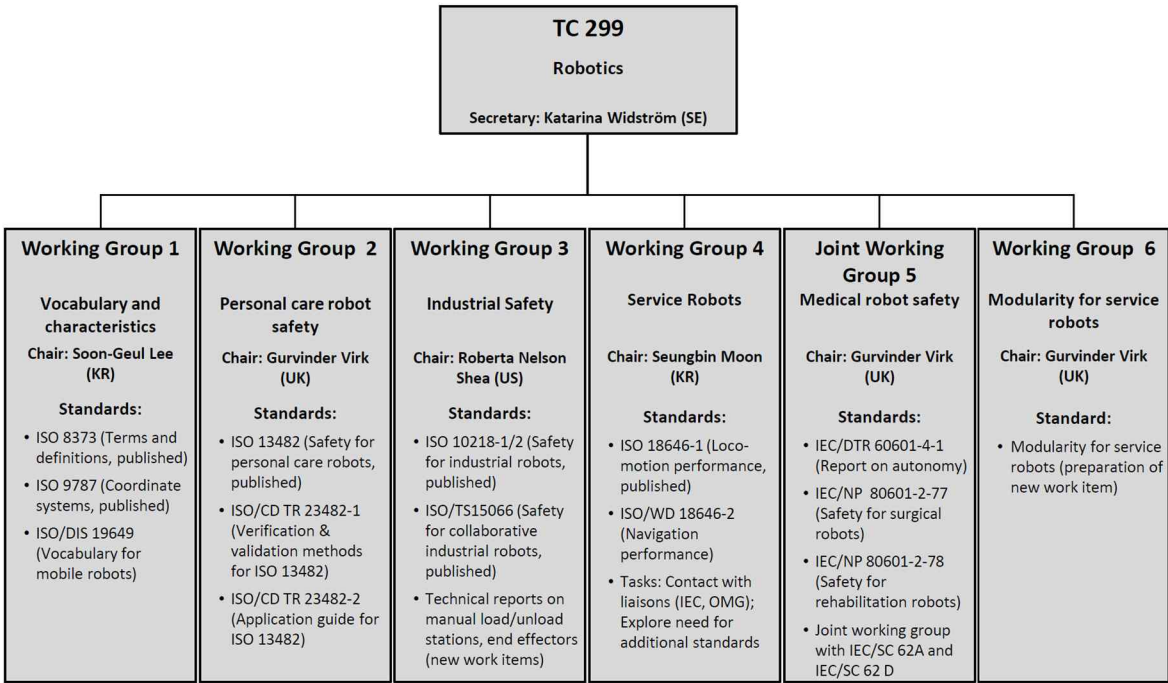
책 등의 분야에 종사하는 100인 이상의 전문가가 참여하여 작업한 "Ethically Aligned Design" 문서를 공개하였다. IEEE에서는 이러한 윤리적 설계/개발 지침을 통해 인공지능 기술과 제품들로 인한 사용자 피해를 예방할 수 있도록 하며, 추후 다양한 규격과 표준들을 통해 인공지능의 개발과 생산 과정에서의 윤리적 사항들을 감독할 수 있을 것으로 기대하고 있다 [18].

3.2 인공지능의 신뢰성/안전성 국제 표준화 동향

유럽연합은 최근 '인공지능을 탑재한 로봇'을 '전자인격'을 갖는 '전자시민'으로 정의하였다. 이처럼 로봇은 물리적으로 실제화된 인공지능 형태중 하나라할 수 있다. 제조, 물류, 공공장소, 가정 등 다양한 환경에서 인간과 가까이 공존하며 상호작용하는 지능로봇의 안전과 성능에 기반한 신뢰성 확보는 성장하는 로봇 시장에 중요한 이슈이다. 일반적으로 가정과 공공 환경에서 개인지원로봇이 활용될 때, 로봇은 위험 상황에 즉각적이고 적절하게 대처 할 수 없는 어린이나 노약자를 포함하여 로보틱스 기술에 익숙하지 않은 다양한 사람과 접촉하게 된다. 그러므로 안전은 더욱 중요한 이슈라 할 수 있다. 이에 미국, 유럽 등 로봇 선진국에서는 다수의 기업들이 표준화 활동에 적극적으로 참여하고 있으며, 성능평가 및 안전, 신뢰성, 규제대응 등을 염두에 둔 제품개발 및 개선을 통해 기술장벽을 적극적으로 준비·구축해 나가고 있다 [13, 14, 15].

그동안은 기계류의 안전 표준인 ISO 12100과 ISO 13849-1을 참조하여 안전 확보를 위한 위험 평가를 수행하였으나, 자율 이동 및 사람과 근접에서 상호작용 하는 서비스 로봇의 특성을 고려할 때 어려움이 있어, 이를 해결하기 위해 개인 지원 로봇의 안전 요구조건을 규정하는 ISO 13482를 제정하게 되었다. 이와 같이 개인지원로봇의 안전 요구사항에 대한 국제 표준이 제정됨에 따라 유럽, 중국, 일본, 중동 등 해외 시장은 국제 표준에 기반한 서비스 로봇의 인증을 요구하고 있다 [14].

최근 승격되어 신설된 ISO TC299은 <그림1>과 같은 구조로 WG들 - WG2(개인지원로봇 안전), WG3(산업용 로봇 안전성), WG4(서비스로봇 성능), JWG5(의료용 로봇 안전) - 을 구성하고 로봇의 성능과 안전에 대한 국제 표준들을 개발 중에 있다.



<그림1: ISO TC299 구조 및 개발 표준 현황 [12]>

ISO 및 IEC 등에서 개발된 로봇의 안전성과 관련된 표준들은 다음과 같다

- ISO 12100 –Safety of machinery –General principles for design –Risk assessment and risk reduction
- ISO 13849-1 –Safety of machinery –Safety-related parts of control systems –Part 1: General principles for design
- IEC 62061 –Safety of machinery –Functional safety of safety-related electrical, electronic and programmable electronic control systems
- ISO 10218-1 –Robots and robotic devices –Safety requirements for industrial robots –Part 1: Robots Requirements for the design of manipulators for industrial environments
- ISO 10218-2 –Robots and robotic devices –Safety requirements for industrial robots –Part 2: Robot systems and integration
- ISO/TS 15066 –Robots and robotic devices –Collaborative robots
- ISO 13482 – Robots and robotic devices – Safety requirements for personal care robots
- IEC 62061 and IEC 61508-3 –Functional safety of electrical/electronic/ programmable electronic safety-related systems –Part 3: Software requirements

예를 들어 서비스로봇의 경우 ISO 13482는 서비스 로봇의 기계적, 전기 전자적 결함과 움직임에 의해 발생할 수 있는 위험원(hazard)를 정의하고 로봇의 안전 확보를 위한 본질적 안전 설계와 보호 조치, 사용상의 주의 정보, 이를 위해 필요한 안전 요구사항 등에 대해 규정하고 있다. 또한 위의 위험 요소를 감소시키기 위한 방어 수단으로 제어 시스템을 구현한 경우 ISO 13849-1 또는 IEC 62061에 따라 위험 평가를 통해 제어 시스템의 기능에 대한 성능 수준(Performance Level)을 결정하고 검증(Verification and Validation)하도록 하고 있다 [14].

앞서 살펴본 로봇의 사례들과 같이 인공지능 기술 분야에서도 용어에 대한 표준화를 비롯해, 응용별, 분야별 특성에 따른 분류를 하고, 이에 기반한 위험요소와 안전 요구사항, 사용상 주의 정보를 정의하고, 위험 평가에 따른 성능 평가와 검증이 가능하도록 하는 국제 표준화 노력들이 요구된다.

3.3 개방형 인공지능 기술 동향

최근 인공지능이 빠르게 확산된 주요 배경 중 하나에는 오픈 인공지능 플랫폼과 오픈소스의 확산이라는 개방형 노력들이 있다. 과거의 인공지능 연구는 대부분 특정 연구 그룹을 통해서만 연구가 진행되고, 그 결과는 논문 등을 통해 일부 공개되는 수준이었다면, 최근에는 SW 소스코드와 데이터 등을 오픈하고 협력하는 모델을 통해 빠른 오픈 이노베이션을 빠르게 도모하는 형태를 추구하고 있다. 이는 오픈소스의 확산, 클라우드의 확산과 같은 ICT 생태계의 기본적인 변화 흐름에 기인한 면도 있지만, 구글/MS/IBM 등과 같은 주요 기업들의 인공지능 클라우드 사업과 연계된 측면도 있다고 할 수 있다.

인공 지능 연구의 기업 쏠림 현상도 가속화 되고 있다. 구글, 페이스북, 마이크로소프트 그리고 중국 바이두가 2015년 한 해 동안 인공지능 관련 기업을 매입하는데 쓴 돈이 85억 달러(약 10조 원)다. 페이스북은 2015년과 2016년 각각 파리와 베를린에 인공지능연구소 문을 열고, 구글은 2015년 독일 국책 인공지능연구소를 통째로 인수하여 유럽 인재들을 유혹하고 있다. 인공지능에 대한 연구 중심도 대학교 또는 국책 연구소에서 기업으로 빠르게 옮겨가고 있다 [20].

구글은 인공지능 연구조직 내에서 머신러닝과 신경망연구를 위해 개발하였던 머신러닝 프레임워크인 TensorFlow를 2015년 11월 오픈 소스로 공개하였다. 페이스북의 인공지능 개발조직인 FAIR(Facebook Artificial Intelligence Research)도 머신러닝 프레임워크인 Torch기반으로 제작된 인공지능 모듈들을 오픈소스로 공개하였다. MS는 얼굴을 인식하는 Face API, 음성인식을 수행하는 Speech API, 이미지를 인식하는 Vision API, 자연어 인식 기능을 수행하는 LUIS 등으로 이루어

지는 프로젝트 옥스퍼드(Project Oxford)의 API 키를 2015년 10월에 공개하였다. 엘론 머스크 등이 10억달러를 공동 출자하여 2015년 12월 창설한 비영리 인공지능연구기관인 OpenAI에서도 인공지능 플랫폼인 OpenAI Gym을 2016년 4월에 공개한 바 있다. 이 밖에도 Open ML(Machine Learning)을 비롯해 다양한 오픈소스 기반의 인공지능 도구들이 개발되고 있다 [5].

이러한 인공지능 플랫폼의 공개에도 불구하고 기업들이 보유하고 있는 데이터 공개나 개방형 데이터 규격 등에 대한 움직임은 아직 상대적으로 많지 않은 상태이다. 인공지능을 위한 효과적인 알고리즘 개발을 위해서는 대량의 학습용 데이터와 시험용 데이터들이 필요하게 된다. 데이터의 확보에는 많은 시간과 노력이 소요되기에, 초기 연구자나 규모가 작은 스타트업의 입장에서 이러한 데이터의 확보가 큰 난관 중에 하나라 할 수 있다. 이러한 데이터 축적에 따른 네트워크 효과는 중장기적으로 실제적인 연구성과에서도 많은 격차를 발생시키게 된다. 주요 기업들이 인공지능 플랫폼을 개방하는 배경에는 시스템을 개방하고 사용자들이 사용하도록 함으로써, 여러가지 활용 사례들과 이에 사용되는 데이터 세트들을 획득하여 축적/활용하려는 의도도 있다고 할 수 있다. 이처럼 개방형 인공지능 플랫폼과 개방형 데이터, 오픈소스 기술들은 인공지능 기술 선도를 위한 중요한 방법으로 활용되고 있다 [5, 19, 20].

4. 결론

지금까지 인공지능 진화과정에서 등장하고 있는 다양한 이슈들을 살펴보고, 이중에서도 특히 부각되고 있는 윤리적/규범적/법적 이슈들, 그리고 신뢰성/안전성 이슈와 동향에 대해 살펴보았다.

인공지능 기술 발전의 초기단계인 현재, 인공지능에 관한 제도적/정책적 이슈들 중 다수는 빅데이터, 사물인터넷 등이 제기하는 이슈들과 많은 부분이 중첩되고, 인공지능에 고유한 이슈들은 개입하기에는 덜 성숙하여 이슈화 자체 또는 이슈에 대한 대응방안을 고려함에 있어서 신중한 접근이 요구된다는 특징이 있다.

이러한 이유로 EU는 수년간 다양한 분야의 전문가 컨소시엄을 통해 인공지능의 혜택과 위험성을 평가하고 관련 이슈를 정리하여 그에 대한 대응방안을 연구해 왔으며, 그 결과를 기초로 제도적 방안을 모색하고 있으며, 미국의 경우에도 인공지능 일반에 대한 선부른 제도적 정책적 접근 대신 로봇, 자율주행자동차, 드론, 의료기기 등 개별 응용영역별로 현실적으로 제기되는 이슈들에 대하여 관련 부처 간의 협력을 통해 증거에 기반을 두어 정책적, 제도적 대응방안 모색을 시도하고 있다 [19].

인공지능 기술은 弱인공지능에서 强인공지능으로 발전하면서 초지능화 될 가능성이 높으며, 단지 인간의 생활을 조력하는 단계를 넘어 앞으로 50년 이내에는 인간의 삶의 일부로 편입되며 포스트휴먼(post-human)의 시대가 도래 할 것으로 전망되고 있다. 로봇에 대한 윤리적/법적 지위가 논의되고, 로봇과 인공지능의 위험을 막기 위한 킬스위치가 검토되며, 1942년 아이작 아시모프에 의해 만들어졌던 로봇3원칙이 그 기반으로 논의되는 지금, 인류에게는 그동안 경험하지 못했던 다양하고 새로운 판단기준들을 만들어야 하는 시대가 도래하고 있는 것이다. 100년 이내에 인공지능이 인간을 능가할 것으로 전망되는 상황에서 우리는 인류가 만든 인공지능이 최대한 인간친화적이며 인간과 동일한 목표를 추구하도록 만전을 기해야 한다.

Yudkowsky[21]가 인공지능의 부정적 영향으로 언급했던 것처럼, 인공지능의 가장 큰 위험은 사람들이 자기가 인공지능을 잘 이해한다고 선불리 결론을 내린다는 사실이다. 불완전하며 절대선(絕對善)이지 못한 인간이 인공지능을 다룬다는 것이 위험한 것처럼, 도래하고 있는 인공지능 시대에 대응하는 가장 중요한 방법은 인간 스스로의 부족함을 인정하고 최대한 윤리적/절차적/제도적 방법 속에서 오류를 최소화시키고 위험을 최소화시킬 수 있도록 하는 방법뿐이다. 전자 인격(electronic personhood)이 현실화되고 있는 시기에, 개방형 구조를 바탕으로 인간친화적이면서도 안전한 인공지능 체계로 설계되고 운영될 수 있도록 하는 다양한 법제도와 더불어, 인공지능의 신뢰성/안전성을 보장할 수 있도록 하는 기술 표준/규격을 만드는 노력이 빠르게 진행되어야 할 것이다.

참고문헌

- [1] Nick Bostrom, "Superintelligence: Paths, Dangers, Strategies", Oxford University Press, 2014
- [2] World Economy Forum, "Global Risks Report 2017", <http://reports.weforum.org/global-risks-2017/>
- [3] 카카오 정책지원팀 번역, 美 백악관 '인공지능의 미래를 위한 준비' 보고서, <https://brunch.co.kr/@kakao-it/42>
- [4] 김병운, "인공지능 동향분석과 국가차원 정책제언", 한국정보화진흥원, 정보화정책, Vol.23 No.1, 2016
- [5] 남충현, "오픈소스 AI : 인공지능 생태계와 오픈 이노베이션", KISDI Premium Report 16-10, 2016.12.21
- [6] 한국과학기술기획평가원, "과학기술&정책, 기술 동향", No. 81, 2016.11.4
- [7] 이원태, "인공지능의 규범 이슈와 정책적 시사점", KISDI Premium Report 15-07, 2015.12.7
- [8] 최은창, "인공지능 시대의 법적·윤리적 쟁점", Future Horizon, Spring 2016 제28호, 2016.5, 18-21
- [9] 이원태, "유럽연합(EU)의 로봇법(RoboLaw) 프로젝트", KISO Journal, 제23호, <http://journal.kiso.or.kr/?p=7496>
- [10] 한국과학기술기획평가원, "2015 기술영향 평가, 제2권 인공지능", 연구보고 2016-048, 2016.1
- [11] 한상기, "왜 지금 인공지능 윤리를 논의해야 하는가?", <http://slownews.kr/54060>
- [12] Theo Jacobs, "Safety requirements and standardisation for robots: software do's and don'ts", ROS-Industrial Conference 2016
- [13] 한국로봇산업표준원, "로봇표준 – Part 4: 국내·외 표준화 현황", Vol. 1, 2016.8
- [14] 정영숙, 이경호, 서범수, "서비스 로봇의 안전 및 성능 평가 기술", 한국통신학회지, 33(8), 36-43, 2016
- [15] 박현섭, 전진우, "국·내외 지능형 로봇 분야 표준화 동향 및 시사점", KEIT PD Issue Report, Vol 16-02, 2016.02
- [16] Eliezer Yudkowsky, "Creating Friendly AI 1.0: The Analysis and Design of Benevolent Goal Architectures", 2001
- [17] IP nomics, "日, AI에 공식 인증 부여한다", <http://www.ipnomics.co.kr/?p=59139>
- [18] The IEEE Standards Association, "Ethically Aligned Design", 2016
- [19] 김윤정, 윤희선, "인공지능 기술의 활용과 발전을 위한 제도 및 정책 이슈", KISTEP, ISSUE PAPER 2016-07
- [20] 강정수, "산업으로서의 인공지능과 기업·국가별 격차 심화", <http://slownews.kr/56114>
- [21] Eliezer Yudkowsky, "Artificial Intelligence as a Positive and Negative Factor in Global Risk", Global Catastrophic Risks, Oxford: Oxford University Press, 2008
- [22] 김윤명, "人工知能(로봇)의 법적 쟁점에 대한 試論的 考察", 정보법학회 학회지, 제20권 제1호, 2016.5.10