JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 21, No. 5, pp. 969-978, May 2020

Check for updates

지능형 소프트웨어의 안전성 확보를 위한 정책적 개선 방안

박 태 형[1] · 강 상 욱[2*]
[1]소트프웨어정책연구소 책임연구원
[2]상명대학교 컴퓨터과학과 교수

Policy Proposal for Assuring the Safety of Intelligent Software

Tae-Hyoung Park[1] · Sang-ug Kang[2*]

[1]Principal Researcher, SW Technology Research Team, Software Policy & Research Institute, Gyeonggi-do 13488, Korea

[2]Professor, Department of Computer Science, Sangmyung University, Seoul 03016, Korea

[요 약]

본 논문에서는 지능형 소프트웨어의 안전성에 대해 약 인공지능에 기반하여 개괄적으로 연구하였다. 기존 연구에서는 전통적 소프트웨어의 안전성과 보안성을 기획, 개발, 테스트, 운영 및 유지보수 등의 단계별로 나누어서 제시하고 있다. 하지만, 본 연구에서는 다양한 딥러닝 알고리즘과 여러 안전성과 관련된 보고서들을 바탕으로 4차 산업혁명 시대에 걸맞은 지능형 소프트웨어의 안전성에 대한 이슈들을 도출하였다. 도출된 이슈별로 미국, 일본, EU, 중국, 우리나라에서의 지능형 소프트웨어 발전 및 보안에 관한 정책을 조사 분석하고 적정 정책들이 제시되고 있는지의 여부를 점검하였다. 이를 바탕으로 우리나라에서의 미진한 정책들을 중심으로 개선된 정책 방향을 제시하고 안전한 소프트웨어 사용을 위한 10가지의 권고안도 제시하였다.

[Abstract]

This paper studied the safety of intelligent software focused on the weak artificial intelligence and proposed considerations to build up government based policies. The existing related papers and reports explained the software safety by dividing the software development process into four steps - design, development, test and operation. However, we extracted various aspects of safety issues on intelligent softwares from deep learning algorithms and related reports. Then we also analyzed existing worldwide policies by issues and scrutinized the appropriateness of them. Based on the analysis, some useful political proposals are suggested and some missing policies are pointed out. In addition, ten recommendations for the use and development of safe intelligent software are proposed in order to summarize the whole content of the paper.

색인어 : 지능형 소프트웨어, 딥러닝, 소프트웨어 안전성, 소프트웨어 안전성 정책, 인공지능

Key word : Intelligent software, Deep learning, Software safety, Software satety policy, Artificial intelligence

http://dx.doi.org/10.9728/dcs.2020.21.5.969

Received 16 March 2020; **Revised** 15 May 2020
Accepted 25 May 2020

***Corresponding Author; Sang-ug Kang**

Tel: +82-2-781-7588
E-mail: sukang@smu.ac.kr

Ⅰ. 서 론

지능형 소프트웨어는 기계학습을 시작으로 최근의 딥러닝에 이르기까지 괄목할만한 발전이 있었다. 4차 산업혁명의 주요 동력이 될 지능형 소프트웨어의 기술력은 삶의 질을 높이고, 긍정적인 사회적 효과가 있지만 동시에 다양한 위험성을 내재하고 있다. 따라서, 우리사회가 이러한 위험성에 대처할 수 있는지 여부를 살펴보는 것은 향후 안정적이고 지속적인 지능형 소프트웨어 기술 발전에 필수적이다. 이에 우리나라뿐만 아니라 미국, 일본, EU, 중국 등에서는 지능형 소프트웨어의 발전과 안전성 확보에 대한 국가적 정책을 제시하고 있다. 하지만, 이러한 정책이 다양한 딥러닝 알고리즘에서 기인한 문제점 및 안전성에 관한 관련 보고서에서 도출된 이슈에 잘 대처할 수 있는지 여부는 검증이 필요하다. 만약 미진한 부분이 있다면, 이를 보완할 수 있는 정책적 제언과 안전한 인공지능의 개발과 활용에 필요한 권고안도 필요한 시점이다.

본 연구에서는 우선 기존의 소프트웨어 안전성과 보안성 확보 프레임워크를 살펴보았다. 그리고 다양한 딥러닝 알고리즘과 관련 보고서를 조사 분석한 후에, 기존 프레임워크로 해결하기 어려운 지능형 소프트웨어에 특정적인 보안성 및 안전성 이슈를 도출하였다. 또한, 도출된 이슈별로 미국, 일본, 중국, EU, 우리나라의 정책 등을 조사 분석하여 4차 산업혁명 시대에 필수적인 역할을 수행할 지능형 소프트웨어의 잠재적인 위험성에 정책적으로 잘 대응하는 지 여부를 알아보았다. 마지막으로 이 결과를 바탕으로 국내의 실정에 맞는 개선된 정책 방향을 이슈별로 제안하였고, 안전한 지능형 소프트웨어의 개발 및 사용을 위한 10가지의 권고안도 제안하였다.

Ⅱ. 전통적 소프트웨어 안전성 확보 프레임워크

1998년 국제기구인 IEC (International Electrotechnical Commission)는 IEC 61508 [1]을 제정하여 전기/전자/프로그램 가능한 전자 시스템에 대한 기능안전을 명시하였고, 2010년 개정하였다. IEC 61508은 기능안전의 대표적인 표준으로서 전자기기, 원자력, 의료기기, 프로세스 산업, 자동차 등 다양한 분야의 기능안전에 관한 주요한 표준이다. 안전 생명주기를 통하여 시스템의 개념단계에서 구현, 양산, 관리, 폐기에 이르기까지의 전체 단계에 대한 위험 분석 및 평가, 안전 무결성 수준 (Safety Integrity Level)을 설정한다. 여기서 안전 무결성 수준이란 시스템의 허용 고장 수준을 말하며 네 단계로 분류되어 있다 [2]. IEC 61508은 목표로 하는 안전 달성을 위해 어떤 안전 기능을 추가할 것인가를 결정하는 안전기능 요구사항과, 안전기능의 달성 가능한 성능 정도에 따른 안전 무결성 수준을 결정하는 절차를 제시한다. 즉, 시스템 전체 안전 기능 요구사항은 안전 무결성 수준과 함께 결합하여 하위 시스템에 요구사항으로 할당함으로써 하위 시스템의 구조 또는 각 시스템의 기능들을 구현하는 기술 및 측정법을 결정한다. 안전기능 요구사항은 시스템을 구성하는 하위시스템 또는 컴포넌트의 생성에 영향을 미치고 안전 무결성 요구사항은 시스템 구조에 영향을 미친다.

ISO/IEC Guide 51 [3]은 제품 표준에 안전에 관한 측면을 포함시키는 데 필요한 요구사항과 권장사항을 제공하는 가이드라인이며 [그림2-1]과 같이 제품의 라이프사이클에 걸쳐 안전에 관한 사항을 규율한다. 이 가이드라인을 준수하여 IEC 61508 (기능안전표준), IEC 61511 (프로세스 산업안전표준) 등의 하위 규격이 생성되었다. 또한 IEC 61508을 근간으로 소프트웨어 안전이 필요한 각 산업군별 특성을 고려한 산업군별 안전규격을 제정하고 있다. [그림 1]의 소프트웨어 안전성 확보 프로세스는 정보통신산업진흥원에서 2016년 발표한 SW 안전성 공통 개발가이드 [4]의 IEC 61508의 안전수명주기를 기반으로 분석, 개발, 테스팅, 운영으로 재구성하였다.

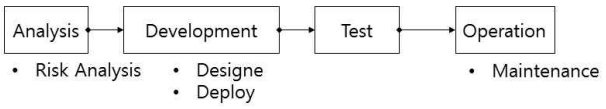

그림 1. 소프트웨어 안전성 확보 프로세스
Fig. 1. The process for assuring software safetyness

국내에서도 주요 산업 도메인별로 국제 표준을 도입하여 적용하려는 노력이 진행 중이지만 선진국에 비해서는 초기단계에 머물러있다 [5]. 안전성 자체에 대한 인증은 아니지만, 국내 공공기관에 납품하는 소프트웨어는 소프트웨어산업진흥법에 의한 품질인증 제도인 GS (Good Software) 인증을 받아야 한다. 2001년부터 시행되었으며, ISO/IEC 9126, 25051 등 국제 표준에 근거하여 품질 특성별 결함을 점검한 후에 인증한다. 인증기관은 한국정보통신기술협회 (2000년)와 한국산업기술시험원 (2008년) 이다. 인증 과정에서 품질 평가모델을 기반으로 제품의 기능성, 사용성, 신뢰성, 효율성, 상호 운용성, 표준 적합성, 성능 등을 평가하고 발견된 결함에 대해서는 수정 후 재평가가 가능하다.

Ⅲ. 전통적 소프트웨어 보안성 확보 프레임워크

소프트웨어 개발보안은 보안의 주요 목적인 기밀성, 무결성, 가용성의 달성을 방해하는 소프트웨어 취약점을 소프트웨어 개발 단계에서부터 사전에 제거하는 것을 목표로 한다. 즉, 소프트웨어 개발 단계별로 수행하는 일련의 보안활동을 통하여 안전한 소프트웨어를 구현하고 운영하기 위한 소프트웨어 개발체계이다. 안전한 소프트웨어는 보안이 위협받는 상황에서도 시스템을 신뢰할 수 있는 상태로 유지할 수 있도록 해야 한다. 미국 국립표준기술연구소 (National Institute of Standards and Technology, NIST)의 연구결과에 따르면, <표 2-2>에서 볼 수 있듯이 설계과정에서 발생한 결함이 제품 출시 후에 발견 및 조치될 경우에는 설계단계에서 결함을 수정하는 비용 대비 30

배의 비용이 발생한다고 발표하였다 [6]. 소프트웨어 개발보안의 종류에는 CLASP (Comprehensive, Lightweight Application Security Process) [7], Seven Touchpoint [8], TSP-Secure [9] 등이 있으나 대표적인 것은 MS-SDL (Microsoft Secure Development Lifecycle) [10] 이다. MS-SDL이 적용된 소프트웨어는 이전에 비해 50% 이상 취약점이 감소하는 효과가 있었다. 비록 다양한 소프트웨어 개발보안 방법론이 존재하지만, 공통적으로 적용되는 단계를 살펴보면 "분석 → 설계 → 구현 → 검증 → 유지보수" 라고 할 수 있다.

유지보수 단계에서 수행할 수 있는 추가적인 보안 활동에는 소프트웨어 보안성 인증이 있다. 소프트웨어 인증은 국제적으로 사용되는 CCRA (Common Criteria Recognition Arrangement) [11]와 국내에서 사용되는 ITSCC (IT Security Certification Center) [12]가 있다. CCRA는 정보보호제품의 안전성을 회원국 간 상호 인정하여 활용을 증진시키는 국제협약이다. CCRA에서 사용하는 평가 기준은 CC (Common Criteria)로 정보 보호 제품의 평가 기준을 규정한 국제 표준(ISO 15408)이다. CC의 주 초점은 객관적인 평가이나, 보안 요구 사항을 개발하는 사람들에게 중요한 표준을 제공하기도 한다. CC는 미국, 캐나다, 프랑스, 독일, 네덜란드, 영국 등 6개국의 노력의 합작품으로, 유럽의 ITSEC (Information Technology Security Evaluation Criteria) [13], 미국의 TCSEC (Trusted Computer System Evaluation Criteria) [14], 캐나다의 CTCPEC (Canadian Trusted Computer Product Evaluation Criteria) [15]와 같은 표준을 바탕으로 만들어졌다.

표 1. 소프트웨어 개발 단계에 따른 결함 수정 비용 (단위, 배)
Table 1. The cost for fault correction divided by the software development steps (unit, times)

type	design stage	coding stage	integration stage	beta stage	comercial stage
design fault	1	5	10	15	30
coding fault		1	10	20	30
integration fault			1	10	20

Ⅳ. 지능형 소프트웨어의 안전성 및 보안성 이슈

제4차 산업혁명을 대비하는 사회 대부분의 영역에서 지능형 소프트웨어의 도입이 추진됨에 따라, 기존의 법제도 규제에 적용되지 않는 규제의 문제에서부터 지능형 소프트웨어 사용에 따른 일자리 감소와 같은 사회경제적 문제에 이르기까지 다양한 이슈가 지능형 소프트웨어와 함께 부상하고 있다. 본 절에서는 지능형 소프트웨어의 다양한 이슈 가운데 안전성과 관련하여 기 논의된 주요 이슈를 살펴보고, 이와 함께 지능형 소프트웨어의 발전과 함께 제기되고 있는 새로운 관점의 기술적 이슈를 논의한다.

4-1 규제 및 거버넌스의 사각 지대

최근 부상하고 있는 지능형 소프트웨어에 대한 법제도 및 규제는 기술의 발전에 비해 많이 뒤처진 상태이다 [16]. 현재 사회 전반에서 추진되고 있는 인공지능의 도입을 활성화하고, 관련하여 제기되는 다양한 이슈를 다루기 위한 규제 및 거버넌스 측면의 논의가 시작되어 활발하게 진행 중이지만, 관련 법제도는 아직까지 미성숙한 상태라 할 수 있다. 그리고 이러한 지능형 소프트웨어에 대한 법제도적 논의는 활성화에 초점을 두고 추진이 우선시되고 있기에, 지능형 소프트웨어의 안전성을 다루는 규제 및 거버넌스 역시 미흡한 상황이다.

지능형 소프트웨어의 안전성과 관련된 대부분의 이슈들은 기존의 법령·규범·거버넌스 체계 하에서 적용되지 않는 새로운 문제로 부상할 것이 예상되고 있어, 안전성 이슈가 규제 및 거버넌스의 사각지대에서 사회적인 위험으로 자리 잡게 될 가능성이 높다. 또한 지능형 소프트웨어에 의해 사용되는 데이터에서의 개인정보나 지적재산권의 보호에 대해 현재의 개인정보보호법이나 관련 법률은 이를 보호하기 위한 적절한 제도적 장치를 갖추고 있지 못하고 있다. 따라서 현재의 규제 및 거버넌스가 다루고 있지 못하는 위험 요소를 최소한으로 해소하고 보장하기 위해, 지능형 소프트웨어의 안전, 보안 등을 보장할 수 있는 적절한 법제도 체계의 구축과 정비가 요구된다. 이는 단순히 법제도적 정책의 문제가 아니기에, 우선적으로 지능형 소프트웨어의 안전성이 기술의 구현과 발전에 있어, 기술 아키텍처상의 코드로서 규제될 수 있도록 하는 논의가 시작되어야 하며, 이와 함께 현실의 법제도적 규제를 통해 지능형 소프트웨어의 안전성을 법적 근거 하에 보장할 필요가 있다.

4-2 의사결정에서의 불명확성과 책임성

현재까지의 지능형 소프트웨어 기술은 내부적으로 의사결정과 판단이 어떻게 이뤄지는지 알 수 없는 블랙박스 모형으로, 지능형 소프트웨어 기술로 인한 문제가 발생한 경우 어떠한 원인이 작용했는지 알기 어렵다. 또한 학습에 사용되는 데이터와 관련하여, DB의 품질에 있어 일부 오류가 존재할 수밖에 없는 점을 고려하면, 지능형 소프트웨어의 의사결정이 잘못된 결론을 내릴 가능성을 배제할 수 없다. 이러한 문제로 인해, 지능형 소프트웨어 개발과정에 의사결정에 대한 책임소재를 설정하고자 하지만, 사회적으로 합의되거나 결정된 기준이 없어 어려움이 있다. 특히 지능형 소프트웨어가 어느 범주까지 사람을 대신하여 판단할 수 있는가와 관련하여, 기술적[1]이나 법적 판단[2]은 어느 정도 효율적으로 가능하다고 인식되고 있는 반면, 윤리

1) 기술적 판단은 한정된 변수와 조건 하에서의 일반적 수준의 공학적 또는 의학적 판단을 의미하며, 이러한 판단은 일정한 정보에 따라 일률적인 판단이 이뤄지므로 실수, 비일관성, 부정 등의 인적오류 요인을 배제할 수 있는 지능형 소프트웨어 활용.

2) 법적 판단은 성문화된 법률과 규정에 따른 판단을 의미하며, 새로운 지능형 사회에서 발생할 수 있는 상황들이 법 규정이 명확하게 성문화되어 법 적용이 용이한 경우 일관된 판단 가능.

적 성격의 문제[3]에 있어서는 지능형 소프트웨어의 불명확한 의사결정으로 인해, 판단이 부적절하거나 편향된 의사결정이 발생 가능하다는 여러 부작용이 인식되고 있다. 지능형 소프트웨어의 윤리적 판단은 곧 개발자나 제작자의 윤리적 입장을 반영할 가능성이 존재하며, 의사결정을 위해 학습하는 데이터가 편향된 경우, 왜곡된 판단을 할 가능성이 있다.

한편 지능형 소프트웨어의 설계 시 고려하지 못한 상황 또는 조건에 대한 판단이 인명 피해나 재산 손실을 초래할 경우, 그 책임을 누구에게 귀속시킬 수 있는가에 대해서는 아직 명확하게 정립되지 않은 상태이며, 일부 경우에 대해 제조물책임(Product Liability)과 같은 기존 규제의 적용을 논의하고 있지만 제한적으로 적용될 수밖에 없는 한계를 가진다. 따라서 지능형 소프트웨어의 판단과 의사결정 과정을 사람이 이해하고 분석할 수 있도록 하는 투명성이 요구되며[4], 이와 함께 판단이나 의사결정이 안전함을 신뢰할 수 있도록 하는 책임성을 어떻게 설정할 것인가에 대한 사회적 논의와 기준의 정립이 요구된다.

특히, 안전의 측면에서는 지능형 소프트웨어의 판단과 그에 따른 책임의 문제가 이상적으로 해결될 수 없기에, 아직까지는 의료 등과 같이 생명과 관련된 특정 분야의 경우, 최종적으로 사람에 의해 지능형 소프트웨어의 판단을 보정 또는 검토·결정할 수 있는 방향으로 개발되도록 추진할 필요가 있다.

4-3 학습데이터의 품질과 편향성

지능형 소프트웨어는 기본적으로 데이터에 기반 한 학습에 의존하기에 데이터의 확보는 지능형 소프트웨어의 선행요소로서 항상 언급된다. 그리고 양질의 빅데이터 보유가 인공지능의 수준을 결정하여 기업 간 격차 기술 격차를 넘어 국가 간 격차를 심화할 것으로 예측된다. 특히 인간의 역할을 대신한다는 관점에서, 지능형 소프트웨어는 정확한 의사결정을 지원하는 것을 넘어 중립적이고 공정성을 갖출 필요가 있는데, 고품질의 편향되지 않은 데이터는 의사결정에 대한 신뢰성을 보장할 뿐 아니라, 안전성 측면에서도 잘못된 데이터로 인한 오작동을 하지 않도록 하는 중요한 요소라 할 수 있다.

하지만 현재 추진되고 있는 대부분의 관련 기술은 민간에서 개발되고 있고, 데이터 역시 수많은 정보원천에서 수집되고 이들이 결합되어 빅데이터로 통합되고 있어, 이들 데이터가 어떠한 의도를 가지고 편향된, 잘못된 데이터인지 확인한다는 것은 불가능한 것이 현실이다. 또한 최근에는 이러한 데이터의 이슈는 단순히 잘못된 편향적인 데이터의 사용의 문제를 넘어, 지능형 소프트웨어 자체의 기능에 대한 공격의 이슈로도 등장하고 있어, 안전성에 있어 데이터의 문제는 더욱 심화될 가능성이 높은 상황이다. 구체적으로 지능형 소프트웨어 모델의 데이터 학습을 통해 오판을 유도하는 데이터를 반복 주입하는 공격 기법 등이 등장하고 [17], 이러한 기법은 인공지능 모델의 판단 오류를 유발한다는 관점에서 안전성을 크게 저해 가능하나, 특별한 대책이 부재한 상황이다.

4-4 새로운 안전성 및 보안성 확보

지능형 소프트웨어는 그 자체의 안전성 및 보안성이 강화될 필요가 있다. 지능형 소프트웨어의 내부 의사결정 과정에 의도하지 않은 외부의 개입이 반영되지 않도록 개발에서부터 학습과 활용의 전 단계에서 구현된 시스템을 안전하게 유지해야 하고, 구현된 소프트웨어가 오작동하지 않고, 그 결과가 신뢰할 수 있는지 검증해야 한다. 지능형 소프트웨어 기술 자체의 복잡성이 높고, 다양한 신기술과 결합됨에 따라 오작동 유발의 요인을 식별하는 것이 어렵고, 파급력 역시 확대될 수 있기에 개발 단계에서부터 안전한 지능형 소프트웨어를 구현하기 위한 노력이 필요하다. 또한 스웨덴의 철학자인 닉 보스트롬 (Nick Bostrom)이 주장한 바와 같이 [18], 지능형 소프트웨어 자체의 안전성이 내재적으로 보장되도록 연구자와 안전 관계자들 간의 협력 관계가 구축되어야 하며, 기업들은 안전을 검토하는 노력을 지능형 소프트웨어 자체의 연구개발과 동등하거나 더 높은 수준에서 병행해야 할 필요가 있다.

4-5 프라이버시 침해

인공지능은 다양한 분야에서 인간을 대체하는 기술로서 도입되어 개인에 대한 데이터를 수집·분석·활용하는 경우가 많기 때문에, 개인정보나 프라이버시의 관점에서 아직까지는 잠재적이지만, 새로운 침해 요소로 부상할 가능성이 높다. 특히, 치안 등의 목적으로 감시 및 통제 기능을 위해 사용되는 인공지능은 개인에 대한 많은 정보를 수집·분석하여 범죄를 예방하고 사회의 위험도를 낮출 수 있지만, 이러한 과정에서 어떠한 프라이버시 침해가 발생할지 알 수 없기 때문에, 이에 대한 충분한 고려와 논의가 필요하다. 특히 최근에는 기존의 텍스트 위주의 개인정보를 넘어, 영상정보와 같은 엄청난 양의 비정형 개인정보가 가정과 같은 사적 영역에서도 실시간으로 수집·분석되고 있고, 이들 데이터가 인공지능의 학습에 사용되는 과정이 클라우드 환경에서 이뤄짐에 따라 개인정보와 프라이버시 침해의 가능성이 점차 높아지고 있다. 예를 들어, 소프트뱅크가 개발한 인공지능 로봇인 페퍼(Pepper)는 학습한 데이터를 클라우드를 통해 공유한다. 또한 지능형 소프트웨어의 개발이 경쟁적으로 추진됨에 따라 데이터의 확보도 경쟁적으로 이뤄지고 있고, 이와 함께 데이터 개방과 공유 관점이 부각되고 있어 개인정보와 사생활 침해의 가능성을 높이고 있다. 따라서 지능형 소프트웨어의 개발 단계부터 개인정보의 이슈를 어떻게 해결할 지에 대한 윤리적이고 법률적인 검토가 동반되어 잠재적인 개인정보

3) 윤리적 판단은 행위의 옳고 그름에 대한 규범적 판단으로, 판단의 기준 자체가 사회 및 문화 요소에 따라 다양하여 고려해야할 변수가 많고 일관된 판단을 내리기 어려워 지능형 소프트웨어를 이용한 윤리적 판단 자체가 가능한지에서부터 판단의 효율성까지 여러 관점에서 부작용의 우려제기.

4) 미 방위고등연구계획국(DARPA)는 이러한 지능형 소프트웨어 내부의 결정과정을 사람이 이해할 수 있도록 하는 Explainable Artificial Intelligence (XAI) 연구를 수행중임.

및 프라이버시 침해의 위험성을 낮추어야 한다.

4-6 기술의 오남용

현대 사회에서 신기술은 종종 범죄나 전쟁의 수단으로 활용되며, 지능형 소프트웨어 역시 그렇게 될 가능성이 존재한다. 먼저, 범죄적 목적으로 오용되는 경우, 그 파급효과는 현대 정보사회의 안전성에 대한 심각한 위해로 작용할 수 있으며, 사회 경제적으로도 치명적이고 심각한 손실·피해를 초래할 수도 있다. 특히 전쟁의 수단으로서 인공지능이 사용되는 이른바 킬러로봇, 자율살상무기가 부상하고 있는데, 이는 기존 무인화된 자동기기의 성능을 넘어 지능형 소프트웨어에 의해 공격 대상을 자율적으로 판단하여 공격할 수 있다. 따라서 전통적인 윤리적 기준이 무시되어 무차별적인 인명 살상 등을 초래할 수 있어 전쟁의 가능성을 높이고 군비경쟁을 더 격화시킬 수 있는 위험성이 있다. 이미 영국의 Taranis UCAV (Unmanned Combat Air Vehicle), 미 해군의 자율운항무인함정 (Autonomous Unmanned Surface Vehicle) Sea Hunter, 보잉의 무인잠수정 Echo Voyager, 러시아의 무인탱크 URAN-9 등이 자율살상무기로 등장하고 있다. 이에 따라 2014년 4월 최초로 제네바에서 자율살상무기에 관한 UN 특정재래식무기금지협약 회의가 개최되었고 [19], 2017년 11월 제네바에서 열린 UN 특정재래식무기금지협약 회의에서 자율살상무기 정부전문가그룹 회의가 진행되어 지능형 소프트웨어로 움직이는 킬러로봇에 대한 규제 논의가 진행 중에 있다.

민간 분야에서도 Campaign to Stop Killer Robots과 같은 NGO를 중심으로 킬러로봇의 규제에 대한 논의가 진행되고 있다. 2015년 7월 부에노스아이레스에서 열린 International Joint Conference on Artificial Intelligence 에서는 인공지능 기술이 부가된 자율로봇의 무기화 자제를 촉구하는 공개서한이 발표되었고, 이에 일론 머스크, 스티븐호킹, 노암 촘스키, 데미스 허사비스 등 개발 관련자 2,587명을 비롯한 1만 7,972명의 주요 인사가 동의하여 각국 정부에게 무기 개발 자제를 촉구하였다 [20]. 따라서, 지능형 소프트웨어를 이용한 범죄나 오용을 규제할 법제도적 논의가 시작되어야 하며, 특히 안전과 연관된 지능형 소프트웨어 기술의 개발과 활용에 대한 검증과 규제에 관한 사회적 합의를 도출하여야 한다.

Ⅴ. 국내 인공지능 정책동향

지능형 소프트웨어의 활용과 확산은 국가 전반적 구조 변화를 불러오고 있고 오작동·오남용 등에 따른 위험과 피해의 문제를 수반할 수밖에 없으며, 이에 대한 정책적 대응은 어느 때보다 중요한 시점이 되었다 [21]. 인공지능 발전단계의 초기에 있는 국가일수록 인공지능 산업·기술의 육성과 확산에 초점을 맞추기 때문에, 앞서 언급한 위험이나 피해를 사전적으로 대응하기 위해서는 기술구현과 활용 가치를 극대화하는 동시에 부작용과 리스크를 최소화 할 수 있는 정책 마련이 필요하다. 정부는 제4차 산업혁명의 주도권을 선점하고 지능정보사회를 앞당기기 위한 전략을 2016년부터 추진하고 있다. 정부가 정의하는 지능정보사회란 고도화된 정보통신기술 인프라를 통해 생성·수집·축적된 데이터와 지능형 소프트웨어가 결합한 지능정보기술이 경제·사회·삶 모든 분야에 보편적으로 활용됨으로써 새로운 가치가 창출되고 발전하는 사회이다. 또한 데이터와 지식이 노동, 자본 등의 기존 생산요소 보다 중요해지고 다양한 제품과 서비스의 융합으로 이종 산업간 경계가 붕괴되며, 지능화된 기계를 통한 자동화가 지적노동 영역까지 확장되는 등 경제·사회 전반에 혁신적인 변화가 발생한다. 2016년 5월에 'K-ICT 전략 2016' 10대 전략산업에 지능정보산업이 포함되었고 [22], 2016년 12월에는 '지능정보 사회 중장기 종합대책'을 수립하였다 [23]. 인간 중심 지능정보사회 실현이라는 비전과 함께 아래의 전략을 제시하였다.

- 기업·국민- 정부·학계의 파트너십 구축
- 기술·산업·사회를 포괄한 균형 있는 정책 추진
- 전략적 지원을 통한 기술 및 산업 경쟁력 조속 확보
- 사회적 합의를 통한 정책 개편 및 역기능 대응체계 구축

이를 위해 기술·산업·사회 분야별 정책방향을 설정하고 이를 달성하기 위한 전략과제를 선정하였다. 인공지능의 안전성에 대한 내용은 주로 사회 분야의 12개 과제에 담겨있다. 그 중 안전 관련해서는 사이버 위협, 인공지능 오작동 등 역기능 대응을 12번째 과제로 다루고 있다. 지능형 소프트웨어로 발생할 수 있는 역기능을 효과적으로 차단하기 위한 주요 세부 과제는 다음과 같다.

- 사이버 위협에 대응한 지능형 자율 방어체계 실현
- 인간과 사물을 포함한 지능형 통합인증체계 구축
- 지능정보 SW의 안전성 평가체계 마련
- 인공지능 보안인력 양성 및 글로벌 공조체계 강화

정부는 2017년 11월 4차산업혁명위원회와 논의를 거쳐 4차 산업혁명 대응계획을 발표하였다 [24]. 여기서는 총론 위주의 접근을 넘어 국민이 체감하는 성과와 새로운 변화를 창출하기 위한 구체적인 청사진을 제시하였다. 모두가 참여하고 모두가 누리는 산업혁명 구현을 목표로 정부는 민간의 혁신역량이 극대화될 수 있도록 시장 환경을 개선하는 조력자 역할과 공공분야 선제 도입으로 민간의 마중물 역할을 한다는 계획이다.

- 다양한 신산업 창출 및 튼튼한 주력산업 육성
- 고질적 사회문제를 해결하여 국민 삶의질 제고
- 양질의 일자리 창출 및 미래 일자리 변화에 대응
- 누구나 이용 가능한 지능화 기술·데이터·네트워크 확보

이 중에 안전과 관련된 세 번째 항목은 미래사회 변화 대응에서의 사이버 안전망 확립에 대해 다루고 있다. 이는 앞서 발표한 지능정보사회 중장기 종합대책의 사이버 위협 및 인공지능 오작동 등 역기능 대응의 연장선 상에 있다.

Ⅵ. 국내외 인공지능 이슈별 정책 분석

다양한 국내외 인공지능 정책에서 안전성과 관련된 부분을 '규제 및 거버넌스 사각지대', '의사결정 불명확성과 책임성', '데이터 품질과 편향성', '안전성 및 보안성', '기술 오남용' 등 앞서 제시한 5가지 이슈로 정리·분석해 보았다. 분석한 정책자료는 미국의 Preparing for the Future of Artificial Intelligence [25], The National Artificial Intelligence Research and Development Strategic Plan [26], Federal Automated Vehicles Policy [27] 이다. EU는 RoboLaw [28] 프로젝트, Civil Law Rules on Robotics [29], SPARC (The Partnership for Robotics in Europe) [30]이다. 일본은 New Robot Strategy [31], 일본재흥전략[5] 이며, 중국은 "인터넷+" AI 3년 행동실시 방안[6]이다.

규제 및 거버넌스 사각지대 분야 국내외 인공지능 관련 정책

미국	Preparing for the Future of AI · (중점 정책 분야) 대중을 보호와 혁신 추구를 위한 정책, 법률, 규제환경 조성 · (권고안 9) AI와 규제 · (권고안 22) 전세계적 고려사항과 보안 Federal Automated Vehicles Policy · 표준 주 정책: 자율주행차 규제 국가프레임워크, 현행 규제수단, 현대적 규제수단
중국	인터넷+" AI 3년 행동실시방안 · (중점추진과제) 행동실시방안에 대한 국가차원 전략 지원
EU	로봇법 프로젝트 · 로봇규제가이드라인 - 로봇시장 발전을 위한 규제 필요 - 해당 도메인의 이해관계자 참여 - 로봇에 대한 기술적 규제 모색 - 기본권 보호를 위한 원칙의 제시 - 잠재적 위험에 대한 법적 책임 원칙 논의 로봇시민법 · Civil Law Rules on Robotics - 사회적 유대 해체 회피 - 로봇 연구에 대한 평등한 접근 · 인공지능 로봇 가이드라인 - 전자인간으로 법적 지위 부여 - 로봇담당국 신설을 통한 법적, 윤리적 문제 관리 SPARC - 투자와 혁신 증진 및 위험 감소를 위한 민관 협력 촉진 - 법적, 윤리적, 사회적 이슈를 해결하기 위한 정책 개발
국내	지능정보사회 중장기종합대책 · (추진전략) 사회적 합의를 통한 정책 개편 및 역기능 대응체계 구축 · (전략과제) 사이버 위협 및 인공지능 오작동 등 역기능 대응 - 인공지능 보안인력 양성 및 글로벌 공조체계 강화 4차산업혁명 대응계획 · (기본방향) 누구나 이용할 수 있는 세계 최고 수준의 지능화기술·데이터·네트워크 확보

안전성 및 보안성 분야 국내외 인공지능 관련 정책

미국	Preparing for the Future of AI · (중점 정책) 테스트베드 구축 및 파일럿 프로젝트 지원 · (권고안 8) AI와 규제 · (권고안 18,19) 공공성, 안전, 거버넌스 · (권고안 22) 전세계적 고려사항과 보안	The National AI R&D Strategic Plan · (전략) Ethical, Legal, and Societal Implications · (전략) Safety and Security Federal Automated Vehicles Policy · 자율주행차 성능지침
일본	로봇 신전략 · 인공지능과 인간사회에 관한 간담회 - (이슈) 인공지능에 의한 사고, 오작동 등의 책임 소재	인공지능 개발 가이드라인, 공적인증제도 운영계획 · 인공지능 시스템의 위험억제에 관한 원칙 - 안전의 원칙, 보안의 원칙
EU	로봇법 프로젝트 · 로봇규제 가이드라인 : 각 도메인별 활용 사례중심의 접근 로봇시민법 (Civil Law Rules on Robotics) - 로봇에 의해 야기되는 피해에서 인간 보호, 로봇에 의한 돌봄 거절, 로봇을 마주한 인간의 자유 보호 · 인공지능 로봇 가이드라인 - 비상 상황시 즉각 대응을 위한 킬 스위치 의무적 장착 - 로봇 사고 발생시 시스템 코드에 접근할 수 있는 권한 제공	
국내	지능정보사회 중장기종합대책 · (전략 과제) 사이버 위협 및 인공지능 오작동 등 역기능 대응 - 강화된 사이버위협에 대응한 지능형 자율방어체계 실현 - 인간과 사물을 포함한 지능형 통합인증체계 구축 - 지능정보SW의 안전성 평가체계 마련 4차산업혁명 대응계획 · (전략 과제) 미래 일자리 변화에 대응한 인재 성장지원과 일자리 안전망을 강화하고 인간 중심 윤리체계를 확립 - 데이터 수집 및 인공지능 알고리즘 개발시 오작동·남용 등 역기능을 예방을 위해 인간 중심 윤리를 정립	

의사결정 불명확성과 책임성 분야 국내외 인공지능 관련 정책

미국	Preparing for the AI · (권고안 16) 공공성, 안전, 거버넌스 · (권고안 22) 전세계적 고려사항과 보안	The National AI R&D Plan · (전략) Human-AI Collaboration · (전략) Ethical, Legal, and Societal Implications · (전략) Safety and Security
일본	로봇 신전략 · 인공지능과 인간사회에 관한 간담회 - (이슈) 인공지능에 의한 사고, 오작동 등의 책임 소재	인공지능 개발 가이드라인과 공적인증제도 운영계획 · 인공지능의 위험억제 원칙 - 투명성의 원칙 - 제어가능성 원칙 · 이용자 수용성 향상 원칙 -이용자 지원의 원칙 - 책임의 원칙
중국	"인터넷+" AI 3년 행동실시 방안 · (중점추진과제) AI 신흥 산업 육성 및 발전	
EU	로봇시민법 (Civil Law Rules on Robotics) - 로봇이 처리한 개인정보 관리, 로봇에 의한 조작 위험에 맞서는 인간 보호	

5) 日本経済再生本部, "日本再興戦略 改訂 2015". Available on https://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/dai1jp.pdf

6) 国务院, "'互联网+'人工智能三年行动实施方案". Aailable on http://www.miit.gov.cn/n1146290/n1146392/c4808445/part/4808453.pdf

데이터 품질/편향성 분야 국내외 인공지능 관련 정책	
미국	Preparing for the AI · (중점 정책) 대중이 사용 가능한 데이터세트 구축 지원 · (권고안 7) AI와 규제 The National AI R&D Plan · (전략) Standards and benchmarks · (전략) Datasets and Environments
일본	로봇 신전략 ·인공지능과 인간사회에 관한 간담회 - (이슈) 인공지능에 의한 사고, 오작동 등의 책임 소재
중국	"인터넷+" AI 3년 행동실시방안 · (중점추진과제) AI 신흥산업 육성 및 발전
국내	4차산업혁명 대응계획 · (기본방향) 누구나 이용 가능한 세계적 지능화기술 및 데이터·네트워크 확보 · (전략 과제) 미래 일자리 변화에 대응한 인재 성장지원과 일자리 안전망을 강화하고 인간 중심 윤리체계를 확립 - 데이터 수집 및 인공지능 알고리즘 개발시 오작동·남용 등 역기능을 예방을 위해 인간 중심 윤리를 정립

기술 오남용 분야 국내외 인공지능 관련 정책	
미국	The National AI R&D Strategic Plan · (전략) Ethical, Legal, and Societal Implications
일본	인공지능 개발 가이드라인과 공적인증제도 운영계획 · 인공지능 시스템의 위험억제에 관한 원칙 - 윤리의 원칙
EU	로봇시민법 (Civil Law Rules on Robotics) - 인간 향상 기술에 대한 접근 제한, 로봇세 부과
국내	지능정보사회 중장기종합대책 · (추진전략) 사회적 합의를 통한 정책 개편 및 역기능 대응체계 구축 · (전략 과제) 사이버 위협 및 인공지능 오작동 등 역기능 대응 -사이버 위협에 대응한 지능형 자율방어체계 실현, 인간과 사물 통합형 지능형 인증체계 구축 4차산업혁명 대응계획 · (전략 과제) 미래 일자리 변화에 대응한 인재 성장지원과 일자리 안전망을 강화하고 인간 중심 윤리체계를 확립 - 데이터 수집 및 인공지능 알고리즘 개발시 오작동·남용 등 역기능을 예방을 위해 인간 중심 윤리를 정립

국내외 정책은 지능형 소프트웨어의 안전성 이슈를 일부 인식하고 이를 해결하기 위한 방향을 논의하는 단계에 위치하고 있음을 확인할 수 있다. 먼저, 정책적 논의가 가장 활발하게 진행되고 있는 미국의 경우 5가지 안전성 이슈를 대부분 인식하고 있고, 관련된 규제나 연구개발의 방향과 안전성에 대한 기술적 지침, 윤리적 고려사항을 논의하고 있다. EU의 경우, 미국과 비슷한 양상으로써 법적, 기술적, 윤리적 원칙과 규제 방향을 수립함으로써 다양한 환경변화에 대응하기 위한 거시적 정책들을 추진하고 있다. 특히, EU의 정책이 수립되면 그를 위반하지 않는 범위 내에서 각 회원국의 상세 정책이 수립되는 구조로 인해 인공지능에 대한 원칙과 방향의 수립이라는 거시적 접근이 이루어질 수밖에 없는 특성이 반영되었다. 로봇과 같은 응용분야를 중심으로 지능형 소프트웨어를 추진하고 있는 일본은 인공지능의 안전성을 보장하기 위한 규제 보다 개발에 있어 안전성의 위험을 억제하기 위한 고려해야 할 원칙을 정립하여 지침을 제공하고자 한다. 지능형 소프트웨어를 핵심 산업기술로서 인식하는 중국은 산업 육성과 발전을 위한 정책에서 인공지능 관련 규제 정비나 데이터베이스 구축, 인간-기계 인터페이스 등 핵심기술 개발을 추진과제로 다루면서 일부 안전성을 보장할 수 있는 사항의 추진을 명시하고 있는 수준이다. 하지만 인공지능의 안전성 이슈를 아직까지는 심각하게 인식하지 않고 있다. 우리나라의 경우, 인공지능 관련 정책은 인공지능의 역기능 해결을 위한 대응의 부분에서 정책의 기본방향 및 추진전략을 설정하고 안전성과 보안성, 윤리 정립을 위한 세부 과제를 제시하는 수준으로 안전성 이슈를 다루고 있어, '의사결정'의 안전성 이슈와 같이 인공지능의 안전성 이슈 전반에 보다 많은 관심이 요구된다고 볼 수 있다. 전반적으로 다른 국가들과 비슷한 수준으로 보완해야 할 바가 많은 상황이다.

Ⅵ. 인공지능 안전성 확보를 위한 개선 방안

6-1 안전관리 관점의 거버넌스 구축을 통한 사각지대 해소

국내의 소프트웨어 안전 활동은 자동차, 항공, 철도, 의료 등의 분야에서 국제표준에 근거한 활동이 주를 이루고 있다. 정보보호의 안전관리 체계인 예방-탐지-대응-복구 프로세스에 기반 하여 각 단계에서 필요한 소프트웨어 안전 관련 법제도적 사항, 소프트웨어 안전 활동, 국제 표준 등 여러 고려 사항들에 대한 논의를 추진하고 있다 [5]. 그럼에도 불구하고, 소프트웨어 안전을 확보하기 위한 거버넌스가 구축되기 어려운 것은 소프트웨어의 편재성에 기인하는데, 소프트웨어가 어느 부처의 소관인지 어느 분야에만 국한되는지 특정하기 어렵다는 것이다. 이점은 지능형 소프트웨어 관련 거버넌스에도 공히 적용된다고 할 수 있다. 인공지능을 통해 생산되는 다양한 결과나 현상들을 거시적인 측면에서 접근하는 것은 인공지능을 구성하는 소프트웨어와 관련된 문제들을 어떻게 해결할 것인가의 문제로 좁힐 수 있기 때문에, 더욱 소프트웨어 안전 확보 체계가 시사 하는 바가 큰 것이다. 따라서 인공지능의 안전성 확보를 위한 거버넌스 구축은 결국, 예방-탐지-대응-복구라는 안전관리의 관점에서 접근하는 것이 위험의 발생으로 인한 피해를 최소화 하는 효과적인 방안이다. 인공지능이 활용될 수 있는 분야의 정부, 관계 공공기관, 각 도메인의 전문가 및 안전 전문가, 안전공학 전문가, 시민사회 등이 선제적으로 거버넌스를 구축하고, 큰 틀 안에서 각 안전 단계별로 구체적인 법제를 도출함으로써 제도적 사각지대를 해소할 수 있다.

6-2 인공지능 수준별 책임성 확보

인공지능의 적용으로 인해 발생되는 위험은 내부적으로 어떻게 의사결정이 이루어지고 행동하게 되는지 알 수 없기 때문

에 예측하기도 어려울 뿐만 아니라 그 양태가 매우 다양하다. 이러한 상황은 새로운 환경과 새로운 데이터를 통해 스스로 학습하기 때문에 더욱 심화될 것이다. 지능형 소프트웨어의 의사결정에서의 불명확성과 이로 인해 발생한 문제의 책임에 대해 국제사회에서 다양한 논의가 진행되고 있고, 그 개발과정 및 제조물 판매에 대한 기업의 책임소재에 대한 사회적 합의를 이끌어 내기 위한 시도들이 있지만, 여전히 결정하기 매우 어려운 문제이다. 이와 관련하여 자율주행 차의 자율주행 수준에 따른 지침을 마련한 미국의 Federal Automated Vehicles Policy에서 좋은 시사점을 얻을 수 있다. 미국은 미국 자동차기술학회에서 정의 (SAE J3016)하고 있는 자율주행 수준 등급에 따라 제작사 및 관련 회사가 고려해야 하는 자율주행 차 성능지침 및 안전평가서를 제시하여 이를 준수토록 한다. 관계 기관의 규정이나 지침 또는 고시 등을 통해 개발자와 제조사로 하여금 그것이 유발할 수 있는 피해를 개발단계에서 상세히 하고, 인공지능 관련 제품의 판매단계에서 이를 고지하도록 하는 제도적 장치가 필요하다. 법제도적 장치를 통해 인공지능의 안전성을 평가할 수 있는 지침을 마련하고, 이것이 검증된 경우에만 실제 적용할 수 있도록 해야 한다. 또한 지능형 소프트웨어 알고리즘은 기업의 입장에서 영업비밀에 해당되기 때문에 온전히 공개할 수 없다. 따라서 지능형 소프트웨어에 의해 도출되는 결과에 대한 설명의 범위를 규정할 필요가 있다. 특히 공개 소프트웨어를 활용한 경우, 설명의 범위를 확대하고, 이를 의무화 하는 방안도 고려해야 한다. 나아가 지능형 소프트웨어가 구현 가능 범위 내에서 현행 법제도를 위반하지 않거나 또는 준수하고 있다는 사실을 소명하도록 책임을 부과하는 방안도 고려해야 한다.

6-3 학습데이터의 품질 제고를 위한 정보 공개 · 공유 활성화

양질의 데이터 확보는 지능형 소프트웨어의 성패를 결정짓는 필수요건이다. 그러나 양질의 데이터는 인공지능 알고리즘이 영업비밀인 것과 마찬가지로 기업의 중요한 자산이기 때문에 개인 및 중소기업은 이를 확보하기 매우 어렵다. 그렇기 때문에 정부가 다양한 산업 및 연구 수요를 파악하여 양질의 데이터세트를 구축하고 공유하는 정책이 필요하다. 특히 데이터세트의 무결성 및 가용성은 결과의 신뢰성을 보장하는 데 필수적인 요소이다. 하지만 사이버공간에 존재하는 무작위의 데이터의 무결성이나 신뢰성을 확인하기는 어렵다. 따라서 데이터의 품질 향상을 위한 체계를 마련하고 데이터세트를 검증하는 공개적인 장을 마련하여야 한다. 미국의 인공지능 R&D 전략 계획에서 예로 제시하고 있는 국토안보부의 IMPACT 프로그램은 좋은 사례가 될 수 있다. 이 프로그램은 사이버 위험 및 신뢰성에 대한 분석을 위한 일종의 정보 마켓으로서 국제 사이버 보안 R&D 커뮤니티, 핵심 인프라 제공 업체 및 정부 지원자 간의 경험적 데이터 공유를 지원하고 있다. 인공지능 데이터세트의 구축, 시험, 검증 등에도 마찬가지로 적용할 수 있는 모델인 것이다. 또한 자발적 데이터세트의 공개와 공유를 촉진하기 위한 정책 및 재정적 인센티브를 제공하는 방안도 고려해 볼만 하다. 한편, 지능형 소프트웨어의 오류나 오작동에 의해 발생한 사고 사례의 공유와 전파 또한 중요한 문제인데, 다양한 현실 데이터를 통해 새로이 학습한 지능형 소프트웨어가 유발하는 오류나 오작동은 예측이 불가능하기 때문이다. 이는 안전관리 관점에서 탐지나 대응의 단계에서 얼마나 신속하게 대처하느냐의 문제이며 그 신속성에 따라 피해를 최소화 할 수 있다. 이러한 측면에서 정보공유분석센터는 정보공유의 체계화에 좋은 모델이다.

6-4 인공지능 기술 오남용 억제를 통한 안전성 강화

세계 각국은 인공지능 분야를 선점하기 위해 치열하게 경쟁하고 있다. 로봇 강국인 일본뿐만 아니라 지능형 소프트웨어 분야에서 후발주자 임에도 불구하고 괄목상대하고 있는 중국도 국가적 차원의 노력을 하고 있다 [20]. 이러한 상황은 인공지능, IoT, 빅데이터 등 신기술의 도래로 주도권을 잡은 국가가 없는 경우 발생하는 일반적 현상이다. 각 국은 인공지능 산업의 육성 및 진흥에 초점을 두고 기술을 선점하고 각국 및 기업의 산업경쟁력을 극대화 하고자 한다. 그러나 일반적으로 안전의 확보는 규제적 관점에서 접근하기 때문에 인공지능 산업의 육성과 진흥을 위해서는 우선적으로 고려되기 어렵다. 정보보호를 너무 강조하면 기업이나 개인의 불편과 산업 효율성이 어느 정도 감소하는 것에서 유추해 볼 수 있는데, 안전 규제를 강화하면 산업의 육성이나 진흥의 속도는 감소할 수밖에 없다. 그럼에도 불구하고, 인공지능의 안전성을 확보해야 하는 이유는 모든 사물이 연결되고, 사이버 공간과 물리적 공간이 유기적으로 의사소통하는 초연결 지능정보사회에서 지능형 소프트웨어에 의해 유발되는 사고의 피해는 막대할 수 있기 때문이다. 그런 의미에서 소프트웨어의 안전성 평가체계를 마련을 위해 기존의 인증체계 내에 소프트웨어 안전성 심사를 추가하는 방안과 설계 시부터 보안성이 확보될 수 있도록 하는 평가체계를 개발하여 이를 국제 표준 인증평가로 확대하는 정책 계획을 담고 있는 우리나라의 '지능정보사회 중장기 종합대책'은 의미가 크다. 인공지능 기술 수준이 초기단계에 막 접어든 상황이며 인공지능 산업도 기반이 마련되어 있지 않은 단계라서 산업의 육성에 치중할 수밖에 없음에도 불구하고 이러한 안전성 확보를 위한 정책들을 병행하는 것은 상당한 의미가 있다. 그러나 신기술 분야의 초기단계에서는 정부의 정책 방향이나 R&D 투자 계획 등이 민간시장에 매우 큰 영향을 미치기 때문에, 자칫 지능형 소프트웨어 산업부문의 개발 추진력을 상실케 할 수도 있다. 따라서 안전성을 확보하도록 하는 제도적 장치와 함께 이를 구현하는 경우, 인센티브를 제공하는 방안을 반드시 고려해야 한다. 기업과 연구자가 개발하는 지능형 소프트웨어가 현행 법규를 준수하고 있으며 안전성 확보를 위한 제도적인 노력을 거쳤음을 인증받도록 하고, 그럴 경우 이에 상응하는 인센티브를 제공하는 것이 지능형 소프트웨어의 육성과 안전 확보를 동시에 달성하는 가장 효과적인 방안이 될 것이다.

Ⅶ. 인공지능 도입을 위한 안전성 권고안

사회의 다양한 분야에 지능형 소프트웨어가 활용되고 국가적 정책이나 세계적 추세를 고려할 때 그 활용 폭과 증가 속도가 증가할 것으로 예상된다. 지능형 소프트웨어는 일부분에서 분명히 기존의 판단 및 분류 알고리즘보다 뛰어나며, 종종 사람의 능력보다 앞선다는 것을 보여준다. 하지만 여전히 오류가 있으며 안전성 문제가 발생할 수 있다. 인공지능을 개발, 도입, 활용하고자 하는 자에게는 본 연구 결과를 집대성하여 마련한 권고안을 제안한다. 이 권고안은 약 인공지능을 사용할 경우에 해당하며, SF 영화나 소설에서 사람처럼 행동하는 일반 인공지능에 대한 내용은 아님을 밝혀둔다.

① 인공지능도 기존의 소프트웨어처럼 완벽하게 안전하지는 않다. 하지만 많은 사람들이 위험에도 불구하고 비행기를 타듯이 이를 사용함에 따른 허용 가능한 위험의 정도를 먼저 정하라.

② 만약 인공지능이 안전을 고려할 필요가 없는 곳에 사용된다고 가정하면 안전성을 고려하지 않아도 된다. 인공지능이 얼마나 위험하고 중요한 곳에 사용되는지를 먼저 판단하고, 위험도 증가에 비례하여 안전성도 더욱 엄격하게 적용해야 한다.

③ 인공지능은 학습 데이터에 의존성이 있는 결과를 도출한다. 학습 데이터가 가능한 모든 위험 상황을 반영하도록 구축되어 인공지능이 발생 가능한 위험에 적절하게 대처할 수 있도록 의료, 교육, 교통 등의 분야별로 데이터를 협력하여 구축하라.

④ 인공지능은 지속적으로 학습하며 성능을 향상시키려고 한다. 이러한 변화에 따른 안전수준이 일정 수준 이상으로 유지되는지 안전 모니터링 활동을 해야 한다.

⑤ 인공지능은 주된 기능은 판단이다. 사람의 경우, 같은 상황에 대해 사람마다 다른 판단을 하고 그 판단은 불완전 할 수 있다. 인공지능은 복잡한 상황에서 과연 편견이나 소수의견을 미반영하는 등의 문제없이 사람보다 훨씬 뛰어난 판단을 할 수 있는지에 대한 연구를 추진할 필요가 있다.

⑥ 가짜 뉴스나 편견으로 인공지능의 활용에 대한 정책을 제시하면, 이 또한 사회적인 안전성을 해치는 요소가 될 수 있다. 따라서 정부 공무원, 회사 임원 등 정책적 책임이 있는 사람들은 인공지능에 대한 정확한 진실을 알아야 한다.

⑦ 시장경제 논리는 항상 인공지능 선점에 따른 이익과 안전성 미준수에 따른 불이익을 저울질 한다. 정부는 인공지능의 진흥보다는 안전성이 우선이 되는 정책을 운영하고, 인공지능을 개발할 때는 일정한 투명성을 확보하고 협업 체계를 구축해야 한다.

⑧ 인공지능의 안전성에 대해 만족할 만한 수준의 사회적 합의가 있기 전까지는 인공지능은 보조적이고 자문적인 수준으로 활용되어야 한다.

⑨ 인공지능과 관련된 안전성 및 보안성 확보 프로세스를 마련하기 위한 R&D를 활발하게 추진하고 국제적인 연구 공조체계를 구축해야 한다.

⑩ 정보화로 인해 정보격차가 존재하였듯이, 인공지능으로 인한 지능격차가 발생할 가능성이 높다. 인공지능 기술 발달의 초기부터 지능 격차 해소를 위한 정치한 정책이 연구되고 마련되어야 한다.

Ⅷ. 결 론

기존에도 안전성과 보안성 측면에서 소프트웨어의 기획, 개발, 테스트, 운영 및 유지보수에 대해 단계별 다양한 기술개발과 표준을 제정해왔다. 본 연구에서는 전통적인 소프트웨어 안전과 보안정책이 지능형 소프트웨어에도 그대로 통용될 수 있는지에 대해 연구 분석하여 정책적 대안을 제시하였다. 지능형 소프트웨어의 안전에 대한 개선 정책 방향으로는;

① 안전관리 관점의 거버넌스 구축을 통한 사각지대 해소
② 인공지능 수준별 책임성 확보
③ 학습데이터의 품질 제고를 위한 정보 공개·공유 활성화
④ 인공지능 기술 오남용 억제를 위한 안전성 강화

등을 제안하였으며, 안전한 지능형 소프트웨어 사용을 위한 권고안도 제안하였다. 본 연구의 결과는 지능형 소프트웨어 안전관련 법·제도 및 표준 강화에도 활용 할 수 있으며, 지능형 소프트웨어의 위험에 대해 신속하게 대처하고 조치할 수 있는 프레임워크 및 전략 마련으로 막대한 손실을 막을 수 있는 경쟁력 상승에도 기여할 수 있을 것으로 기대한다.

감사의 글

본 논문은 한국정보화진흥원의 지원으로 수행된 연구보고서 [32]의 내용을 포함하고 있음. 또한 이 논문의 일부는 2018년도 정부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2018R1A2B6006754).

참고문헌

[1] IEC TR 61508: Functional safety of electrical electronic programmable electronic safety-related systems
[2] Sooyeon Lee, “H/W, S/W safety standard of IT convergence industry IEC 61508’, Korea Testing Laboratory, 2014.
[3] ISO/IEC Guide 51:1999 Safety aspects — Guidelines for their inclusion in standards
[4] National IT Industry Promotion Agency, “SW Safety Common Development Guide”, NIPA, Seoul, Nov. 2016.
[5] T. H. Park et. al, “A Study on the System of Ensuring SW Safety for a SW Oriented Society-Focus on Testing, Evaluation, Certification”, Software Policy & Research Institute, Gyeonggi-do, Jan. 2016

[6] Ministry of the Interior and Safety & KISA, “Software Development Security Guide”, Korea Internet & Security Agency, Seoul, 2017
[7] Viega, J. “The CLASP Application Security Process. Secure Software”, 2005. Available on https:llwww.ida.liu.se/ -TDDC90/literatureIpapers/ clasp -external. pdf
[8] McGraw, G., Seven Touchpoints for Software Security, 2006 Available on http://www.swsec.com/resources/touchpoints/
[9] CERT. "TSP-Secure", 2010 Available on http://www.cert.org/ secure-coding/secure.html>
[10] Microsoft Security Develoepement Iifecycle (SDL). Available on http://www.microsoft.com/france/securite/sdl/
[11] Common Criteria Part 1, CCMB-2005-08-001, Version 2.3, August 2005.
[12] ITSCC (Information Technology Security Certificate Center), November, 2007.
[13] ITSEC: Information Technology Security Evaluation Criteria (Provisional Harmonised Criteria, Version 1.2, 28 June 1991)
[14] DOD 5200.28-STD "Department of Defense Trusted Computer System Evaluation Criteria", 1985
[15] E. Mate Bacic, "The Canadian trusted computer product evaluation criteria," [1990] Proceedings of the Sixth Annual Computer Security Applications Conference, Tucson, AZ, USA, 1990, pp. 188-196.
[16] S. W. Son and Y. M. KIM, “A Study on International Discussion and Legal Policy on Artificial Intelligence Technology”, Korea Legislation Research Institute, Seoul, Oct. 2016
[17] D. S. Choi, “Artificial Intelligent and Security”, *The Journal of The Korean Institute of Communication Sciences*, Vol. 34, No. 10, pp. 31-37, Oct. 2017.
[18] Nick Bostrom, “What happens when our computers get smarter than we are?,” TED 2015, March 2015. Available on https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are
[19] Center for the Study of the Drone, “Debating “Killer Robots” at the United Nations,” April 13, 2015. Available on http://dronecenter.bard.edu/debating-killer-robots-at-the-united-nations
[20] Kangbong Lee, “AI weapon is not that dangerous yet” Science Times, 2005. Available on http://www.sciencetimes.co.kr/ ?p=139237&cat=36&post_type=news&paged=213
[21] T. H. Park et. al, “Software Safety Industry Trends Study”, Software Policy & Research Institute, Gyeonggi-do, April 2017.
[22] Ministry of Science and ICT, “K-ICT Strategy”, Korea, May. 2016
[23] Joint Ministries, “Comprehensive countermeasures for mid- to long-term intelligent information society”, Korea, Dec. 2016
[24] Joint Ministries and 4th Industrial Revolution Committee, “Person-centered 4th Industrial Revolution Response Plan for Innovation Growth”, Korea, Nov. 2017
[25] Executive Office of the President National Science and Technology Council Committee on Technology, “Preparing for the Future of Artificial Intelligence”, US, October, 2016
[26] National Science and Technology Council, “The National Artificial Intelligence Research and Development Strategic Plan”, US, October, 2016
[27] NHTSA, “ Federal Automated Vehicles Policy”, US, September, 2016
[28] EU, “D6.2 Guidelines on Regulating Robotics”, Brussels, September, 2014
[29] European Parliament, “European Civil Law Rules in Robotics”, Brussels, October, 2016
[30] The Partnership for Robotics in Europe (SPARC), https://www.eu-robotics.net/sparc/index.html
[31] The Headquarters for Japan’s Economic Revitalization, “New Robot Strategy”, Japan, October, 2015
[32] National Information Society Agency, “A study on software safety management in the 4th industrial revolution", 2017.

박태형(Tae-Hyoung Park)

2004년 : 고려대학교 일반대학원 (행정학석사)
2011년 : 고려대학교 정보보호대학원 (공학박사-정보보호정책)

2004년～2008년: 한국행정연구원 연구원
2011년～2014년: 고려대학교 정보보호대학원 연구교수
2014년～현 재: 소프트웨어정책연구소 SW기술연구팀 책임연구원(SW안전)
※관심분야 : 정보보호(Personal Information), SW안전(Software Safety), 디지털전환(Digital Transformation)

강상욱(Sang-ug Kang)

1996년 : University of Southern California (공학석사)
2011년 : 고려대학교 (공학박사-멀티미디어 보안)

1993년～1994년: 한국 IBM 주식회사
1996년～2002년: 삼성전자 중앙연구소
2002년～2012년: 한국정보화진흥원
2012년～현 재: 상명대학교 컴퓨터과학과 교수
※관심분야 : 인공지능, 디지털 저작권, 멀티미디어 보안 등