Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.33, NO.6, Dec. 2023 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2023.33.6.975

안전한 AI 서비스를 위한 국내 정책 및 가이드라인 개선방안 연구*

김 지 연,[1†] 석 병 진,[2] 김 역,[2] 이 창 훈[3‡]
[1,2,3]서울과학기술대학교 (대학원생, 연구원, 교수)

A Study on the Improvement of Domestic Policies and Guidelines for Secure AI Services*

Jiyoun Kim,[1†] Byougjin Seok,[2] Yeog Kim,[2] Changhoon Lee[3‡]
[1,2,3]Seoul National University of Science and Technology
(Graduate student, Researcher, Professor)

요 약

인공지능 기술의 발전이 가속화되며 다양한 산업 분야에서 데이터를 활용한 자동화 및 지능화를 가능하게 하는 AI 서비스의 제공이 증가하며, AI 활용으로 발생할 수 있는 AI 보안 위험에 대한 우려가 높아지고 있다. 이에 국외에서는 AI 규제의 필요성과 중요성을 인지하고 관련 정책 및 규제 마련에 주력하고 있다. 국내에서도 이러한 움직임을 보이고 있으나, AI 규제에 대한 구체화가 이루어지지 않은 실정이다. 따라서, 기존의 정책안이나 가이드라인을 비교 분석하여 공통 요소 도출 및 보완점 파악, 국내 AI 규제 방향에 대해 논의할 필요성이 있다. 본 논문에서는 AI 라이프 사이클에서 발생할 수 있는 AI 보안 위험에 대해 조사하고, 각 위험에 대한 분석을 통해 국내 AI 규제 수립에 고려되어야 할 사항 6가지를 도출한다. 이를 토대로 국내 AI 정책안 및 가이드라인을 분석하고, 보완사항을 확인한다. 또한, 미국, EU의 AI 법률의 주요 내용 검토와 본 논문의 분석 결과를 기반으로 국내 AI 정책안과 가이드라인에 대한 개선방안을 제시한다.

ABSTRACT

With the advancement of Artificial Intelligence (AI) technologies, the provision of data-driven AI services that enable automation and intelligence is increasing across industries, raising concerns about the AI security risks that may arise from the use of AI. Accordingly, Foreign countries recognize the need and importance of AI regulation and are focusing on developing related policies and regulations. This movement is also happening in Korea, and AI regulations have not been specified, so it is necessary to compare and analyze existing policy proposals or guidelines to derive common factors and identify complementary points, and discuss the direction of domestic AI regulation. In this paper, we investigate AI security risks that may arise in the AI life cycle and derive six points to be considered in establishing domestic AI regulations through analysis of each risk. Based on this, we analyze AI policy proposals and recommendations in Korea and validate additional issues. In addition, based on a review of the main content of AI laws in the US and EU and the analysis of this paper, we propose measures to improve domestic guidelines and policies in the field of AI.

Keywords: AI policy analysis, guidelines, domestic polices

Received(09. 12. 2023), Modified(10. 27. 2023), Accepted(10. 27. 2023)

* 본 논문은 2023년도 한국정보보호학회 하계학술대회에 표한 우수논문을 개선 및 확장한 것임.

* 이 연구는 2022년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임('20018637').

† 주저자, jiyeon97@seoultech.ac.kr

‡ 교신저자, chlee@seoultech.ac.kr(Corresponding author)

I. 서 론

코로나 19 이후 촉발된 디지털 대전환 시대에서 인공지능(AI, Artificial Intelligence)은 자동화와 지능화를 가능하게 하는 핵심적인 ICT Information and Communications Technology) 기술로 국·내외에서 주목받고 있다. AI 시스템은 클라우드, 빅데이터, 블록체인, 메타버스 등 대부분의 ICT 기술과의 융합이 가능하다. 이에 융합 ICT 기술이 빠르게 발전하고 있으며, AI 기술 주도의 ICT 패러다임 변화가 나타나고 있다.

AI 중심사회가 도래하며 AI를 활용한 서비스 도입이 산업 전 분야에 적용되며 우리의 삶의 방식에도 영향을 미치고 있다. AI 기술은 다양한 분야에서 사회적, 경제적 혜택을 제공하고 있는 반면에 AI의 확산은 많은 양의 데이터 수집과 알고리즘의 활용으로 발생하는 사회적, 윤리적 문제 및 AI의 역기능에 대한 우려도 존재한다. 더 나아가, 생성형 AI (Generative AI) 및 초거대 AI 등과 같은 새로운 AI 형태가 대두되면서 신뢰성과 안전성 문제, 데이터 유출 및 프라이버시 침해에 대한 우려, AI 모델 공격 등의 보안 위험에 대한 우려도 더욱 증가하였다.

다양한 AI 보안 위험은 AI 시스템을 설계, 개발, 배포 및 운용하는 각 단계의 과정 속에서 발생할 수 있다. 때문에, 안전한 AI 생태계의 발전과 활용을 위한 정책 방향 수립뿐만 아니라 프라이버시 정책도 함께 고려되어야 한다. 국내·외에서는 AI 중심사회에서 다양하게 발생하는 보안 위험을 고려하고 안전한 AI 활용을 위한 기술적 보안과 법·제도 등 정책적 대응방안을 마련의 필요성을 인지하였다. 국외에서는 인공지능에 대한 프라이버시 규칙 설정 및 AI 법안 등과 같은 규제와 대응책 마련을 추진하고 있다[1].

미국의 경우, 2023년 3월 22일 AI 기술을 선도하고 발전시키면서 재앙적인 피해를 막기 위해 새로운 감독체계에 대한 초안을 작성하여 회람했다고 발표하였다. 유럽은 AI가 초래할 수 있는 사회적 위험을 예방하고 신뢰할 수 있는 AI 기술 개발을 위해 2021년 4월 21일 유럽 인공지능법(EU AI Act)을 제안하였고, 2023년 6월 EU 의회를 통과하여 법안 제정이 추진되고 있다. 중국의 경우, 생성형 AI 서비스 관리 방안 초안을 발표하고, 기업이 AI 서비스를 출시하기에 앞서 제품의 안전성을 평가해 당국에 제출하도록 하였다[2].

국내에서는 2019년부터 '인공지능 국가전략'을 발표하여 AI 경쟁력 혁신, AI 활용 전면화, 사람 중심의 AI 구현을 목표로 지원 정책을 추진하고 있다. 2021년 7월 1일에는 '인공지능 육성 및 신뢰 기반 조성 등에 관한 법률안[3]'이 제안되어 2023년 2월 14일 국회 과학기술방송정보통신위원회 법안2소위원회를 통과하였다. 개인정보보호위원회에서는 2023년 8월 3일 AI에 대한 전체적인 프라이버시 침해 위험을 최소화하고 데이터를 안전하게 활용하기 위한 'AI 시대 안전한 개인정보 활용 정책방향[4]'을 발표하였다.

AI 산업은 전 세계를 아우르며 성장하고 있기 때문에 AI 국제 규범 마련에 대한 노력도 중요하다. 국내·외에서는 AI 글로벌 규범 또는 표준을 개발하고자 국제 기구와 산업 단체들의 국제협력 움직임도 지속적으로 행해지고 있다. 우리나라도 국제협력 및 국제적 공조 체계 강화를 위해 글로벌 AI 사업자와 국내 AI 사업자와의 소통 활성화에 대한 의지를 보였다[4].

우리나라에서는 안전한 AI 서비스 활용을 위한 조치와 정책 마련, 규제 개선에 대한 논의와 노력이 진행되고 있지만, 구체적인 요구사항이나 기준, 표준화에 대한 정립은 미비한 실정이다. 또한, AI에 대한 규제나 정책에 대해 살펴보고자 할 때, 참조할 수 있을 만한 대표적인 자료나 가이드가 제시되어 있지 않은 측면이 있다. 뿐만 아니라 제도적 수립에 대한 움직임의 실행 수준 역시 미미한 편이다. 이와 같은 문제 외에도 향후 국제협력 활성화를 위해서라도 국내 AI 규제 마련에 대한 논의와 실행은 촉진되어야 할 필요성이 있다.

본 논문은 다음과 같이 구성된다. 2장에서는 AI 라이프 사이클에서 발생할 수 있는 AI 보안 위험을 조사하고 이를 통해 국내 AI 규제 수립을 위한 고려사항을 도출한다. 3장에서는 현재 AI 법제가 체계화된 미국과 EU의 AI 법안을 검토한다. 4장에서는 2장에서 도출한 고려사항을 기반으로 국내 AI 정책 및 가이드라인을 분석한다. 이러한 분석 결과와 3장의 검토 내용을 토대로 국내 AI 정책과 규제 방향에 대한 개선점 및 보완사항을 제시한다.

II. AI 보안 위험

본 절에서는 AI 라이프 사이클에서 발생 가능한 공격 유형을 분석하고, 국내 AI 규제 수립에 고려되어야 할 고려사항 6가지를 도출한다.

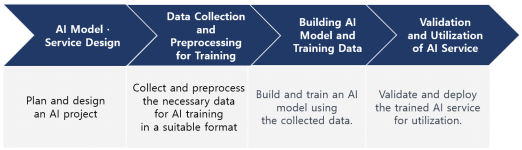


Fig. 1. AI Life Cycle

AI 라이프 사이클은 AI 모델·서비스 설계, 학습 데이터 수집 및 전처리, AI 모델 구축 및 데이터 학습, AI 서비스 검증 및 활용의 각 단계를 가진다. 우리나라의 경우 각 단계에 정확한 정의가 없어, [4]의 'AI 단계별 데이터 처리기준과 보호조치'와 [5]의 'AI 모델 개발주기 단계', [8]의 'AI Life Cycle Stage'를 참조하여 [그림 1]과 같이 재구성하였다.

2.1 AI 공격 유형

[표 1]은 AI 모델을 공격할 수 있는 대표적인 공격 유형으로, 각 공격들은 상기 AI 라이프 사이클 상에서 다음과 같은 보안 위협을 발생시킬 수 있다.

AI 모델·서비스 설계 단계에서는 모델의 설계 결함으로 성능 저하나 취약점이 존재할 수 있으며 모델 개발 시 악성 데이터의 사용으로 보안 위험들이 발생할 수 있다. ①은 데이터에 잘못된 데이터를 주입하여 공격하는 유형으로 학습 데이터 수집 및 전처리 단계에서 발생할 수 있다.

AI 모델 구축 및 데이터 학습 단계에는 모델이 특정 입력에 취약하게 만들거나 강화되게 만드는 공격인 ②, ③이 해당된다. ④, ⑤, ⑥, ⑦은 모델을 사용하거나 분석하는 과정에서 나타날 수 있는 공격으로, AI 서비스 검증 및 활용단계에서 발생할 수 있다. [표 1]의 공격은 AI 라이프 사이클의 한 단계에서만 발생하는 것이 아니라 여러 단계에 걸쳐 나타날 수 있다.

2.2 고려사항 도출

안전하게 AI 서비스를 이용하기 위해서는 AI 규

Table 1. AI Attack Type

	Attack Type	Description
①	Poisoning attack	• Attack that injects contaminated or erroneous data into the training dataset to lower the model's accuracy
②	Backdoor attack	• Training the model by including specific patterns known as "triggers" in the training data • Aiming to induce incorrect classifications when similar inputs with the triggers are encountered • A type of data poisoning attack[5]
③	Adversarial attack	• Security risks that can arise in adversarial environments due to vulnerabilities inherent in machine learning algorithms[7] • An attack that involves minimal modification of input images to cause misclassification by the artificial intelligence model[6]
④	Membership inference	• An attack that analyzes the response results after querying the AI model in the deployment phase to determine if specific data was included in the AI model's training[6]
⑤	Evasion attack	• An attack that involves minimal manipulation of input data in the deployment phase of a trained AI model to deceive machine learning, leading the AI model to make incorrect judgments[5, 6, 7] • The manipulated data is referred to as an adversarial example[6]
⑥	Inversion attack	• An attack that involves throwing numerous queries at a machine learning model, then analyzing the resulting outputs to extract the data used for model training[7]
⑦	Model extraction attack	• Attack where a similar model is extracted or cloned from the original AI model[5] • Like the Inversion attack, this is a type of attack that continuously queries the machine learning model and analyzes the results[7]

제 설정에 관한 요구사항들과 기준이 수립되어야 할 필요성이 있다. 이에 〔표 1〕의 공격 유형을 기반으로 국내 AI 규제 수립에 고려되어야 할 6가지의 고려사항을 도출하여 〔표 2〕로 나타내었다.

고려사항 1(신뢰성)과 3(정확성)은 ③(적대적공격)으로부터 도출하였다. 적대적 공격은 딥러닝 모델에서 이미지 분류 등의 작업에서 모델의 성능을 저하하거나 모델이 오인식하도록 데이터 값을 조작하는 공격으로, AI 시스템의 신뢰성 및 정확성을 저해시킬 수 있다. 적대적 공격을 통해 AI 모델이 잘못된 결과를 제공하면, 이를 활용한 상황의 왜곡과 사용자의 피해가 초래될 수 있다. 따라서, 고려사항 1(신뢰성)과 3(정확성)은 사용자들이 AI 시스템을 사용할 때 고려되어야 하는 중요한 요소이다.

고려사항 2(안전성)는 주로 학습 데이터에 악의적인 데이터 또는 패턴을 삽입하여 모델의 오분류를 유도하는 ①(오염공격)과 ②(백도어공격)에서 도출하였다. 이러한 공격이 금융 데이터에서 발생하게 되면 고객의 금융 정보가 비정상적으로 해석되거나 노출될 수 있다. 따라서, AI 시스템의 데이터 활용과정에서는 데이터 프라이버시와 관련된 안전성이 고려되어야 한다.

고려사항 4(윤리성)는 앞서 살펴본 공격 유형 ①(오염공격), ②(백도어공격)와 ④(멤버십추론)을 기반으로 도출하였다. 멤버십추론 공격은 AI 모델에 질의하고, 그 응답 결과를 분석하여 입력 데이터가 AI 모델의 학습 데이터에 포함되어 있는지 확인한다. 이는 사용자의 개인정보를 유추하거나 민감 정보 노출과 같은 프라이버시 침해를 야기할 수 있다. 따라서, AI 시스템에서 데이터를 활용할 때 개인정보 보호 및 윤리적 문제에 대한 관리가 고려되어야 한다.

투명성은 윤리성 문제와 상호 보완적인 관계에 있다. 고려사항 6(투명성)은 고려사항 4(윤리성), ⑥(전도 공격), ⑦(모델 추출 공격)을 기반으로 도출하였다. 전도 공격은 모델의 작동 방식 및 학습에 사용된 데이터를 복원하는〔9〕 공격 유형으로, 모델의 투명성이 저하되어 AI 시스템 사용자가 모델의 동작을 이해하기 어려워진다. 모델 추출 공격의 경우, 공격자가 모델을 복제하거나 악용하여 모델의 투명성에 대한 우려를 발생시킨다.

고려사항 5(위험관리)는 악성 데이터를 식별하지 못하고 오류를 발생하도록 하는 공격인 ⑤(회피공격)에서 도출하였다. 회피공격으로 인해 변조된 데이터는 적대적 예제라고 부르며, 이는 AI 시스템을 사용하는 실제 환경에 사고나 위협을 초래할 수 있다. 따라서, 위험 분석을 통해 가능한 공격 유형과 영향을 파악하는 위험관리 대응책을 마련할 필요성이 있다.

〔표 1〕의 공격 유형은 AI 라이프 사이클의 전 단계에서 발생할 수 있다. 도출된 고려사항은 상호 연관성을 가지고 있어 이를 보장하기 위해서는 종합적인 접근이 필요하다. 6가지 고려사항들이 AI 시스템의 설계, 개발, 운용 단계에서 공통으로 고려된다면 국내 AI 규제 수립과 관련된 중요한 원칙과 가이드라인으로 활용될 수 있다.

III. 국외 AI 정책

본 절에서는 미국과 EU에서 추진되고 있는 AI 정책에 대해 설명한다.

3.1 미국 - NIST AI RMF 1.0

NIST AI 100-1 AI RMF 1.0〔8〕은 2023년 1월 미국의 NIST(National Institute of Standards and Technology)에서 발표한 인공지능 위험관리 프레임워크이다. AI RMF(Artificial

Table 2. Consideration

	Consideration	
1	Reliability	Are the sources of the collected data clear and reliable?
2	Safety	Is it possible to verify that data privacy protection is well done in the process of data utilization?
3	Accuracy	Can you confirm the accuracy of the results derived from the use of AI services?
4	Ethicality	Do AI services comply with prevailing ethical norms?
5	Risk management	Is the risk management system for the AI service model established?
6	Yransparency	Can you explain how the AI service model works and decisions?

Intelligence Risk Management Framework)의 목표는 AI의 많은 리스크를 관리하고, 신뢰할 수 있고 책임 있는 AI 시스템의 개발과 사용을 촉진하기 위해 AI 시스템을 설계, 개발, 배치 또는 사용하는 조직에 자원을 제공하는 것이다.

해당 프레임워크는 시민의 자유와 권리에 대한 위협과 같은 AI 시스템의 예상되는 부정적 영향을 최소화하는 동시에 긍정적인 영향을 극대화할 수 있는 기회를 식별하는 접근 방식을 제공한다. 또한, AI 시스템과 관련된 사람, 조직, 생태계에 대한 잠재적 피해와 같은 새로운 리스크가 대두될 때 이를 해결하기 위해 설계되었다.

AI RMF는 AI 라이프 사이클 및 차원 전반에 걸쳐 AI act(행위자)가 사용하기 위한 것으로, 각 단계에 따라 AI 행위자들은 서로 다른 위험 관점을 가질 수 있다고 강조하였다. AI RMF 내에서 모든 AI 행위자는 리스크를 관리하고 신뢰할 수 있고 책임 있는 AI의 목표를 달성하기 위해 협력한다. 경제협력개발기구(OECD)는 AI 행위자를 “AI를 배치하거나 운영하는 조직 및 개인을 포함하여 AI 시스템 라이프 사이클에서 적극적인 역할을 수행하는 자”로 정의하고 있다. 여기서, AI 라이프 사이클이란 ① 계획 및 설계(plan and design), ② 데이터 수집 및 전처리(collect and process data), ③ 모델 구축 및 활용(build and use model), ④ 확인 및 검증(verify and validate), ⑤ 배포 및 활용(deploy and use), ⑥ 운영 및 모니터링(operate and monitor), ⑦ 사용 및 영향(use or impacted by)의 단계를 가지는 일련의 과정을 말한다.

3.1.1 신뢰할 수 있는 AI 시스템의 특성

AI 신뢰성(Trustworthiness)을 향상시키기 위한 접근법은 부정적인 AI 리스크를 줄일 수 있다. AI RMF는 신뢰할 수 있는 AI의 특성을 설명하고 이를 해결하기 위한 지침을 제공한다. 신뢰할 수 있는 AI를 만들기 위해서는 AI 시스템의 사용 상황에 따라 각 특성의 균형을 맞추어야 한다. 신뢰할 수 있는 AI 시스템의 7가지 특성으로는 ① 유효성 및 신뢰성(valid and reliable), ② 안전성(safe), ③ 보안 및 복원성(secure and resilient), ④ 책임과 투명성(accountable and transparent), ⑤ 설명 및 해석 가능성(explainable and interpretable), ⑥ 개인정보 보호 강화(privacy-enhanced), ⑦ 공정성(fair-with harmful bias managed)이 있다. 이중 ① 특성은 신뢰성의 필수 조건이며, 다른 신뢰성 특성의 기초로 표시된다. ③ 특성은 다른 모든 특성과 관련되어 있다.

AI 리스크를 관리할 때, 조직은 이러한 특성의 균형을 맞추는 데 있어 어려운 결정에 직면할 수 있다. 신뢰성의 특성을 이해하고 처리하는 것은 AI 라이프 사이클 내에서 AI 행위자의 특정 역할에 달려 있다. 어떤 주어진 AI 시스템에 대해서도 AI 설계자나 개발자는 배포자와는 다른 특성에 대한 인식을 가질 수 있다.

3.1.2 AI RMF Core

AI RMF Core는 AI 리스크를 관리하고 신뢰할 수 있는 AI 시스템을 책임감 있게 개발하기 위한 다이얼로그, 이해, 활동을 가능하게 하는 결과와 행동을 제공한다. Core에는 거버넌스(Govern), 매핑(Map), 측정(Measure), 관리(Manage)의 4가지 기능이 해당되며, AI 리스크 관리 활동을 구성한다. 리스크 관리는 AI 시스템 라이프 사이클 차원 전반에 걸쳐 지속적이고 적시에 수행되어야 한다.

AI RMF Core 기능은 조직 외부의 AI 행위자들의 견해를 포함하고 다학제적인 관점을 반영하는 방식으로 수행되어야 한다. 각 기능은 카테고리와 하위 카테고리로 분류되어 있어, 프레임워크 사용자는 그들의 자원과 능력을 기반으로 AI 리스크 관리를 위해 필요에 맞는 가장 적합한 기능들을 적용할 수 있다.

3.1.2.1 거버넌스(Govern)

Govern은 AI 리스크 관리 전반에 주입되어 프로세스의 다른 기능을 가능하게 하는 cross-cutting 기능으로 리스크 관리 문화가 조성되고 존재한다. 특히, 규정 준수 또는 평가와 관련해서 다른 각 기능에 통합되어야 하는데, 강력한 Govern은 조직의 리스크 문화를 촉진하기 위해 내부 관행과 규범을 추진하고 강화할 수 있다. 거버넌스에 대한 관심은 AI 시스템의 수명과 조직의 계층에 걸쳐 효과적인 AI 위험 관리를 위한 지속적이고 본질적인 요구사항이다.

3.1.2.2 매핑(Map)

Map은 AI 시스템과 관련된 리스크를 프레임하기 위한 상황을 설정하고 리스크를 식별한다. Map 기능이 완료된 후 프레임워크 사용자는 AI 시스템 영향에 대한 충분한 상황적 지식을 가지고 AI 시스템 설계, 개발 또는 배치 여부에 대한 초기 go/no-go 결정을 알려주어야 한다. 진행이 결정되면 조직은 AI 리스크 관리 노력을 지원하기 위해 Govern 기능에 적용된 정책 및 절차와 함께 Measure과 Manage 기능을 활용해야 한다. 시간이 지남에 따라 상황, 기능, 위험, 이점 및 잠재적 영향이 진화해도 AI 시스템에 Map 기능을 계속 적용하는 것이 프레임워크 사용자의 의무이다.

3.1.2.3 측정(Measure)

Measure는 AI 리스크 및 관련 영향을 분석, 평가, 벤치마킹 및 모니터링하기 위해 정량적, 정성적 또는 혼합 방법 도구, 기술 및 방법론을 사용한다. Map 기능에서 확인된 AI 리스크와 관련된 지식을 활용하여 Manage 기능을 알려준다. Measure 기능을 완료한 후에는 측정 기준, 방법 및 방법론을 포함한 객관적이고 반복·확장이 가능한 TEVV(Test, Evaluation, Verification and Validation) 프로세스가 마련되고 준수되며 문서화된다. 지식, 방법론, 리스크 및 영향은 시간이 지남에 따라 진화하기 때문에 Measure 기능을 AI 시스템에 계속 적용하는 것이 프레임워크 사용자의 의무이다.

3.1.2.4 관리(Manage)

Manage는 리스크의 우선순위를 지정하고 예상되는 영향에 따라 조치하는 기능을 가진다. 이는 Govern 기능에 의해 정의된 대로 정기적으로 매핑되고 측정된 리스크에 리스크 자원을 할당하는 것을 수반한다. Manage 기능을 완료한 후에는 리스크 우선순위를 정하고 정기적인 모니터링 및 개선 계획을 수립한다. 방법, 상황, 리스크 및 관련 AI 행위자의 필요나 기대가 시간이 지남에 따라 진화하기에 배포된 AI 시스템에 Manage 기능을 계속 적용하는 것은 프레임워크 사용자의 의무이다.

3.2 EU - AI Act

AI Act[10]은 2021년 4월 유럽 위원회에서 발표된 인공지능에 대한 EU 규제 프레임워크에 관한 제안서이다. 2021년 12월 EU 회원국의 대표로 이루어진 협의회는 AI Act에 관한 일반적인 입장에 동의했다. 해당 법안은 2023년 6월 EU 의회를 통과하며 입법 절차 진행 중에 있다.

AI Act에서 AI 기술은 다양한 분야에서 경제적, 사회적 혜택을 가져올 것이라는 기대와 동시에 AI 기술이 적용된 제품과 기술로 인해 사용자의 안전 위험과 기본권 위협에 대한 우려를 보였다. 해당 법안에서는 AI 시스템을 '(특정) 기술과 접근법에 의해 개발되었으며 인간이 정의한 목표 집합에 대한 내용, 예측, 권장 사항 또는 상호작용하는 환경에 영향을 미치는 결정과 같은 출력을 생성할 수 있는 소프트웨어'로 정의하였다.

AI Act 초안은 다음과 같은 특정 목적을 달성하고자 한다.

(ⅰ) EU 시장에 출시된 AI 시스템이 안전하고 기존 EU 법률을 준수하는지 확인한다,
(ⅱ) AI에 대한 투자와 혁신을 촉진하기 위한 법적 확실성을 보장한다.
(ⅲ) AI 시스템에 적용되는 기본권 및 안전 요구사항에 대한 EU 법률의 거버넌스 및 효과적인 집행을 강화한다.
(ⅳ) 합법적이고 안전하며, 신뢰할 수 있는 AI 애플리케이션을 위한 단일 시장 개발을 촉진하고 시장 분절을 방지한다.

제안된 AI 프레임워크는 AI 시스템의 기술 중립적 정의를 명시하고 위험기반 접근(risk-based approach)을 채택하였으며, EU에서의 AI 시스템 개발, 시장 출시, 사용을 위한 요구사항과 의무를 규정하고 있다. 또한 EU 위원회는 NLF(New Legislative Framework)의 논리를 따를 것을 제안하였다. NLF란 적합성 평가와 CE marking 사용을 통해 다양한 제품들이 EU 시장에 출시될 때, 해당 법률을 준수하도록 보장하는 EU 접근법이다. 이는 EU 내 설립된 AI 시스템 제공자들 또는 EU 시장에서 AI 시스템을 출시하거나 EU에 서비스를 제공하는 제3국에 적용된다. 뿐만 아니라 EU에 위치하고 있는 AI 시스템 사용자들, EU에서 해당 시

스템에 의해 생성된 산출물이 사용되는 제3국에 위치한 AI 시스템 제공자와 사용자들에게도 적용된다.

3.2.1 위험기반 접근법(Risked-based approach)

AI Act는 법적 개입이 구체적인 위험 수준에 맞게 조정되는 위험기반 접근법을 따른다. 이는 AI가 가진 불투명성, 복잡성, 데이터 의존성, 자율적 행동과 같은 특성으로 인해 사용자들의 기본권과 안전에 부정적인 영향을 미칠 수 있다는 우려를 해결하기 위해 채택되었다. 위험기반 접근법은 ① 허용할 수 없는 위험(Unacceptable risk: Prohibited AI practices) ② 고위험(High risk: Regulated high risk AI systems) ③ 제한된 위험(Limited risk: Transparency) ④ 낮고 최소한의 위험(Low and minimal risk: No obligations)으로 분류되며 피라미드 모형으로 나타낸다.

3.2.1.1 허용할 수 없는 위험(Unacceptable risk)

허용할 수 없는 위험(Unacceptable risk: Prohibited AI practices)은 사람들의 안전, 생계, 권리에 명백한 위협으로 간주되는 유해한 AI 관행들을 명백히 금지한다. 이는 다음의 4가지 경우에 해당되는 경우 금지된다.

(ⅰ) 유해한 조작적 '잠재의식 기술'을 사용하는 AI 시스템
(ⅱ) 특정 취약계층(신체적 또는 정신적 장애)을 이용하는 AI 시스템
(ⅲ) 공공기관 또는 이들의 행동을 사회적 점수화하는 목적으로 사용하는 AI 시스템
(ⅳ) 제한된 경우의 수를 제외한 법 집행 목적으로 일반인이 접근할 수 있는 공간에 있는 '실시간' 원격 생체 인식 시스템

3.2.1.2 고위험(High risk)

고위험(High risk: Regulated high risk AI systems)은 사람들의 안전 또는 기본권에 부정적인 영향을 미치는 경우 AI 시스템을 규제한다. 이는 '제품의 안전 구성요소 또는 EU 보건 및 안전 조화 법률에 해당하는 시스템'과 '8가지 특정 영역에 배치된 시스템'이라는 두 가지 범주로 구분된다. 8가지 특정 영역으로는 '자연인의 생체 인식 및 분류', '중요 인프라의 관리와 운영', '교육과 직업 훈련', '고용, 근로자 관리와 자영업에 대한 접근', '필수 민간 서비스와 공공 서비스 및 혜택에 대한 접근과 향유', '법 집행', '이주, 망명과 국경 통제 관리', '정의와 민주적 절차의 행정'이 있다.

이 고위험 AI 시스템은 사전 적합성 평가 요건과 기타 요구사항(위험관리, 테스트, 기술적 견고성, 데이터 훈련 및 데이터 거버넌스, 투명성, 인적 감독, 사이버 보안 등) 규칙이 적용된다. 얼굴 인식 기술(FRT, Facial Recognition Technology)의 경우 고위험 또는 저위험 사용 특성에 따라 차별화된다.

3.2.1.3 제한된 위험(Limited risk)

제한된 위험(Limited risk: Transparency)은 인간과 상호작용하는 시스템(챗봇 등), 감정 인식 시스템, 생체 인식 분류 시스템, 이미지, 오디오 또는 비디오 컨텐츠를 생성·제작하는 AI 시스템(딥페이크 등)으로 제한적으로 투명성 의무가 적용된다.

3.2.1.4 낮고 최소한의 위험(Low and minimal risk)

낮고 최소한의 위험(Low and minimal risk: No obligations)은 위험이 낮거나 최소한인 다른 모든 AI 시스템으로 추가적인 법적 의무를 준수하지 않고도 EU에서 개발과 사용이 가능하다.

IV. 국내 AI 정책

본 절에서는 국내 AI 제도 및 정책, 가이드라인에 대해 살펴보고, 2장에서 도출한 고려사항을 토대로 이를 분석한다.

4.1 AI 법률안 및 정책안

2023년 2월 14일 '인공지능 육성 및 신뢰 기반 조성 등에 관한 법률안[3]'은 국회 과학기술방송정보통신위원회 법안2소위원회를 통과하였다. 이는 인공지능산업의 육성을 도모하면서 인간이 인공지능의 개발·제공 및 이용에 있어서 지켜야 할 윤리적 원칙 등을 규정하여 인공지능을 신뢰할 수 있는 기반을 마련

함으로써 인공지능이 산업과 사회 그리고 인간을 위하는 것이 되도록 이바지하는 것을 목적으로 제안되었다. 해당 법률안은 인공지능의 기본원칙(제3조) 5가지와 인공지능 사업자의 윤리(제7조) 6가지, 이용자의 윤리(제8조) 4가지 등을 제안함으로써, 윤리적인 측면에 대해 강조하고 있다. 제9조(기본계획의 수립·시행)에서는 인공지능 이용자 보호, 인공지능의 공정성·투명성·책임성 확보 및 불법적 사용 등 부작용 해소 방안 등에 대한 사항이 포함되어야 하는 기본계획을 명시하고 있다. 해당 조항을 통해 [3]에서 안전성과 투명성에 대해 고려하고 있음을 확인할 수 있다. 또한, 제12조(인공지능사회를 위한 정책 및 인공지능기술 개발) ①의 3에서 인공지능 및 인공지능 기술 개발·활용으로 인한 위험요인의 평가 및 대응방안 연구에 대한 사업을 추진할 수 있다는 근거 제시를 통해 위험관리를 고려하고 있음을 확인하였으나, 보안 사고 발생 대응 및 조사에 대해서는 제시하고 있지 않았다.

2023년 3월 과학기술정보통신부는 지난해 9월 논의되었던 디지털 정책 구상(뉴욕구상)에서 제시된 디지털 신질서 정립 방안을 단계별로 마련할 계획이라고 밝혔다. 또한, 디지털 권리장전(가칭)을 23년 하반기까지 마련하여 부처별 신질서 정립을 본격화할 예정이라고 밝혔다[11].

2023년 4월 개최된 "디지털플랫폼정부 실현계획 보고회"에서는 인공지능·데이터 시대의 전환기적 도전에 대응하기 위한 정부혁신 전략인 디지털플랫폼정부의 청사진과 4대 핵심과제를 제시하였다. 과학기술정보통신부에서는 '초거대 AI 경쟁력 강화 방안'을 제시하였다. 해당 방안에는 초거대 AI 규제 개선과 제도 정립 추진 등 범국가 AI 혁신 제도·문화 정착을 위한 계획을 포함하고 있다. 개인정보보호위원회는 AI로 인한 국민의 프라이버시 침해 우려를 최소화할 수 있는 방안 마련과 공공부문에 대한 관리·감독 체계 전면 정비 계획을 밝혔다[12]. 세부 계획으로 AI 데이터의 안전한 활용 및 처리를 위한 신뢰성 구축에 대해 제시하며 신뢰성 및 안전성에 대해 고려하였다. 또한, 포괄적인 계획의 제안을 통해, 정확성, 윤리성, 위험관리, 투명성을 고려하고 있다.

2023년 8월 개인정보보호위원회는 "인공지능 시대 안전한 개인정보 활용 정책 방향[4]"을 발표하였다. 주요 추진과제로는 (ⅰ) 불확실성 해소를 위한 원칙 기반 규율체계 정립, (ⅱ) AI 개발·서비스 단계별 개인정보 처리기준 구체화, (ⅲ) 민·관 협력을 통한 분야별 가이드라인 마련, (ⅳ) AI 글로벌 협력 체계 공고화에 대해 제시하였다. 또한, AI 환경에서

Table 3. List of domestic AI polices and guidelines

Date	Document	Publisher	Ref.
'21.05	AI Personal Information Protection Self-inspection Table	Personal Information Protection Commission	[14]
'21.07	AI Guidelines for Financial Sector	Financial Supervisory Service	[15]
'22.08	Guide to AI Development and Utilization in the Financial Sector	Financial Services Commission and 9 institutions	[16]
'23.02	Proposed Act on the Development of Artificial Intelligence and the Establishment of Trust Base, etc	Pilmo Jung and 23 others	[3]
'23.02	Data Quality Guidelines and Implementation Guide for Artificial Intelligence Learning v.3.0	National Information Society Agency	[17]
'23.02	2023 Self-inspection table for the implementation of artificial intelligence ethics standards [proposal] (revision)	Ministry of Science and ICT · Korea Information Society Development Institute	[18]
'23.04	AI Security Guidelines for Financial Sector (revision)	Financial Security Institute	[5]
'23.04	Digital Platform Government Realization Plan	Personal Information Protection Commission · Cooperation of related ministries	[12]
'23.08	Policy Direction for Safe Use of Personal Information in the Age of Artificial Intelligence	Personal Information Protection Commission	[4]

개인정보 보호 원칙 시 고려사항과 AI 단계별 리스크 분석 및 대응 수립 계획을 언급함으로써 신뢰성, 안전성, 위험관리 고려사항을 만족하였다. [4]에서는 AI 서비스 단계에서 개인정보 수집·처리의 투명성을 다루고 있으나, 2.2절의 고려사항 6(투명성)은 AI 모델을 중심으로 다루고 있어 두 내용 간에 차이가 있다. 해당 발표문에서는 현행 「개인정보 보호법」 체계 하에서 그간의 해석례·의결례·판례 등을 종합하여 AI 개발·서비스 기획, 데이터 수집, AI 학습, 서비스 제공 등 단계별로 개인정보를 어떠한 원칙과 기준에 입각하여 처리할 수 있는지에 대해 최대한 구체화하였다[13].

[표 3]은 앞서 살펴본 국내 AI 제도 및 정책과 국내에서 발행된 AI 가이드라인을 정리한 목록이다.

4.2 AI 가이드라인 및 지침서

[5]는 보안관점에서 AI 서비스를 운영계와 개발계로 나누어 구성요소와 각 기능에 대해 안내하였다. 이전과 달리 AI 공격유형별 예방책을 제시해 줌으로써 각 단계별 보안 고려사항의 이해를 돕고 있으나 윤리적인 측면에 대한 언급은 없는 것이 특징이다.

[17]은 인공지능 학습용 데이터 생애주기에 따른 다양한 점검항목과 AI 모델 평가지표, 단계별 보안 고려사항을 제시하였다. 또한, AI 학습용 데이터의 품질관리 프레임워크와 품질관리 원칙, 지표에 대해 구체적이고 체계적으로 제시하여 신뢰성, 정확성, 투명성 고려사항을 만족시켰다. [17]은 품질관리에서 프라이버시 보호 검증에 대한 내용을 포함하며 안전성 고려사항을 만족하였다.

[5, 17]은 AI 모델 평가지표에 관한 내용을 다루고 있다는 점에서 공통점을 지닌다. 이를 통해 데이터와 AI 모델의 밀접한 연관성을 확인할 수 있다.

[14]는 인공지능 관련 개인정보보호 6대 원칙(적법성, 안전성, 투명성, 참여성, 책임성, 공정성)을 도출하였다. 이후 발간된 AI 개인정보보호 안내서는 이 윤리원칙을 포함하는 추세를 보인다. 해당 문서는 단계별(또는 상시)로 법령상 준수해야 할 의무 또는 권장 내용에 대한 점검 및 확인사항을 제시하고 있다.

[15]는 금융 분야에서의 AI 시스템 운영에 대한 가이드라인으로 기획 및 설계, 개발, 평가 및 검증, 도입, 운영 및 모니터링 단계별 지침을 제시하였다.

[16]은 [15]에서 제시된 항목들을 구체화한 세부 안내서로, 신뢰성 확보, AI 시스템 활용 결과에 대한 정확성 여부 확인, AI 윤리원칙 부합 여부, 학습 데이터 출처에 대한 조치 수행에 관한 체크리스트를 포함하고 있다. 이외에도 각 항목별로 점검항목, 필요성, 체크리스트 확인을 통해 AI 시스템 활용에 도움이 될 수 있다.

[18]은 2020년 발표된 '인공지능(AI) 윤리기준'에서 제시하는 3대 원칙(인간 존엄성 원칙, 사회의 공공선 원칙, 기술의 합목적성 원칙)과 10대 핵심요건(인권보장, 프라이버시 보호, 다양성 존중, 침해금지, 공공성, 연대성, 데이터 관리, 책임성, 안전성, 투명성)을 기반으로 한다. 이 중 데이터 관리, 투명성 부문에서 각각 신뢰성, 투명성에 대한 고려사항 만족 여부를 확인할 수 있었다. [18]은 국내 인공지능 윤리 가이드와 인공지능 역기능 사례별 윤리적 고려사항을 제공하고 있어 AI 기술의 적절한 활용을 지원하고 관련 종사자에게 유용한 자료로 활용될 수 있다.

[5, 14, 16, 18]은 AI 서비스의 위험 관리에 대한 언급과 위험 평가 체크리스트 항목 등을 제시하였으나, 표준 체계가 아닌 AI 서비스를 제공하는 각

Table 4. Analysis for AI plans and policies and guidelines (Each type of AI attack can occur in all stages of the AI life cycle.)

	Consideration	Basis for derivation (through AI attack type)		[3]	[4]	[5]	[12]	[14]	[16]	[17]	[18]
1	Reliability	③	→	O	O	O	O	O	O	O	O
2	Safety	①, ②		O	O	△	O	△	△	O	△
3	Accuracy	③		X	△	O	△	X	O	O	X
4	Ethicality	①, ②, ④		O	X	X	△	O	O	△	O
5	Risk management	⑤		△	O	△	△	△	△	△	△
6	Transparency	⑥, ⑦		O	△	O	△	O	O	O	O

기업의 위험 관리 기준을 따른다. 또한, 프라이버시 보호와 「개인정보 보호법」 준수를 위한 노력에 관한 내용을 포함하고 있으나 확인에 관한 내용은 다루지 않았다. AI 서비스의 위험 관리 체계 정립 및 프라이버시 보호 검증 과정에 관한 향후 연구가 필요하다.

4.3 AI 정책안 및 가이드라인 분석 결과 및 개선방안

〔표 3〕을 대상으로 본 논문에서 설정한 고려사항의 만족 여부를 분석하여 〔표 4〕로 나타내었다. 행에는 2절에서 설정한 6가지 고려사항을, 열에는 분석대상을 표시하였다. 각 분석 결과는 고려사항에 직접적인 안내 여부에 따라 O·X로 표시하였고, 해당 내용을 언급만 한 경우에는 △로 나타내었다.

〔표 4〕를 통해 1(신뢰성)과 6(투명성) 측면에서는 대부분의 문서가 관련 내용을 잘 제시하고 있음을 확인하였다. 고려사항 2(안전성), 3(정확성), 4(윤리성), 5(위험관리) 측면에서는 상대적으로 기준 체계가 미비한 것으로 분석되었으나, 최근 발행된 문서들에서는 이에 대한 중요성에 대해 인식하고, 관련 내용을 다루고 있는 것을 확인하였다. 분석대상을 기준으로 AI 시스템 활용에 공통적으로 고려되는 부분은 고려사항 1(신뢰성), 2(안전성), 5(위험관리), 6(투명성)이라는 점을 확인하였다. 분석된 문서들은 2.2절에서 제시한 AI 규제 수립 시 고려되어야 할 고려사항을 전체적으로 다루기보다는 특정 부분에 집중된 내용을 다루는 모습을 보인다. 이러한 분석 결과를 통해 국내 가이드라인 및 정책안의 공통점과 보완점을 확인하였고, 향후 본 논문에서 제시한 6가지 고려사항의 상호보완 연구를 기대한다.

〔표 5〕는 미국, EU, 우리나라의 정책적 특성을 요약한 표이다. 미국과 EU의 AI 법안은 위험 수준에 따라 AI 규제를 다루는 점과 안전성, 투명성에 대해 다루고 있다는 점에서 공통점을 보이고 있다. 그러나 미국은 각 AI 라이프 사이클 단계에서 발생할 수 있는 AI 행위자 관점에서의 리스크 관리를 위한 리스크 관리 프레임워크를 제시하고 있고, EU는 AI 시스템을 통해 발생할 수 있는 위험 수준을 기준으로 법적 수준이 조정되는 위험기반 접근법을 제시하는 데에 차이가 있다. 구체적인 법률안 확정을 목전에 두고 있는 EU 외의 다른 국가들은 AI 규제의 필요성을 알고 노력을 기울이고 있다. 미국의 경우 2028년까지 AI RMF 1.0의 계속적인 검토를 진행하며 현실과 기술의 발전에 맞게 내용을 수정해나갈 계획을 밝혔다〔8〕.

우리나라의 경우 AI를 신뢰할 수 있는 기반 마련과 AI 서비스 활용을 위해 여러 정책안 및 가이드라인을 제시하고 있다. 하지만 AI를 활용하기 위해 필수로 요구되는 기준이나 원칙이 정립되지 않았다. 또한, AI 서비스로 인해 발생할 수 있는 위험 관리에 대한 체계 수립의 미비한 점을 확인할 수 있었다. 최근 〔4〕에서는 AI 단계별 리스크 분석 및 대응체계 수립에 대한 계획을 발표하였으나, AI 활용 속도가 빠르게 증가하고 있는 만큼 위험관리 체계 마련도 시급해 보인다.

우리나라는 AI에 대한 규제나 정책에 대해 살펴보고자 할 때, 참조할 수 있을 만한 대표적인 자료나 기준이 모호한 측면도 있다. 뿐만 아니라 제도적 수

Table 5. Policy characteristics of th USA, EU, Korea

	USA	EU	Korea
Document	NIST RMF 1.0 〔8〕	AI Act 〔10〕	〔3〕
Main point	Risk Management	Risked-based approach	Building a trustworthy foundation for AI
Feature	• Trustworthiness(valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable privacy-enhanced, fair-with harmful bias managed) • AI RMF Core(Govern, Map, Measure, Manage)	• Unacceptable risk: Prohibited AI practices • High risk: Regulated high risk AI systems • Limited risk: Transparency • Low and minimal risk: No obligations	• Proposed AI basic principles, AI operator ethics, and user ethics • Presents the necessity of establishing a foundation for AI social safety

립에 대한 움직임의 실행 수준 역시 미미한 편이다. 이는 우리나라가 높은 인공지능 기술력을 보유하고 있으나, 상대적으로 표준화 정립에 대한 수준이 낮음을 보여준다. 이러한 상황은 AI 기술의 빠른 발전으로 인해 기존 법규가 이에 맞춰가지 못하고 다양한 AI 적용 분야와 잠재적 위험 요소 때문에 정책 수립이 복잡하여 여러 이해관계자 간의 의견 충돌이 발생하기 때문으로 보인다. 정부는 이를 극복하기 위해 AI 윤리와 안전에 대한 가이드라인, 개인정보 보호와 데이터 규제, AI 산업 지원 등 다양한 측면에서 노력하고 있다. 그러나 국외에 비해 이러한 노력이 미미한 측면도 있어 AI 관련 정책 및 규제에 대한 더 많은 논의와 연구가 필요하다.

따라서, 현재 정부에서 제시한 로드맵 및 정책 수립 방향안에 대한 지속적인 검토와 실행을 적극적으로 추진하고, 법안 수립 진행에 속도도 높여야 할 것으로 사료된다. 본 논문에서 제시한 고려사항과 개선방안은 향후 AI 기술 및 신기술 관련 제도가 수립될 때, 신뢰성, 안전성, 위험관리, 투명성 측면이 강화되어야 함을 강조하였다. 특히 잠재적 위험을 최소화하기 위한 위험관리 기반 마련 노력의 필요성 강조를 통해 AI 기술 활용과 기술발전에 도움이 될 것으로 기대한다.

V. 결 론

본 논문에서는 AI 라이프 사이클의 각 단계에서 발생할 수 있는 AI 기술/서비스에 대한 보안 위험요소를 식별하였다. 이를 기반으로 국내 AI 규제 수립 및 가이드 개발을 위한 고려사항 6가지를 도출하였다. 또한, 미국과 EU에서 발표된 인공지능 정책의 주요 내용을 검토하였다. 검토 내용과 도출한 고려사항을 토대로 과학기술정보통신부, 개인정보보호위원회, 금융감독원 등 여러 국가 기관에서 발행한 국내 정책안 및 가이드라인 9권을 분석하였다. 분석 결과를 통해 국내 AI 정책안 및 가이드라인의 보완사항을 확인하고 개선방안을 제시하였다.

본 논문은 AI 규제 수립 및 가이드 개발 시 고려되어야 할 고려사항을 분석하여 이에 대한 중요성을 강조하였다. 또한, 미국과 EU의 AI 규제안 검토와 분석 결과는 국내 정책 수립을 위한 자료로 참고할 수 있다. 향후, 제안 내용을 활용하여 국내 AI 규범 및 정책 수립, 가이드라인이 보완되길 기대한다. 또한, 국제협력을 위한 AI 국제 규범 마련 추진에도 도움이 되길 기대한다. 향후 연구로는 국가정보원에서 발간된 "챗GPT 등 생성형 AI 활용 보안가이드라인"을 분석하고자 한다.

References

[1] Dowon Kim and 6 others, "ChatGPT (ChatGPT) Security Threats and Implications", KISA Insight 2023 Vol.3, pp.8-16, May. 2023.

[2] KISA, "The rise of GhatGPT and the trend of personal information protection regulations in major countries", 2023 Privacy Report VOL 4, pp. 8-9, May. 2023.

[3] Pilmo Jung and 23 others, "Proposed Act on the Development of Artificial Intelligence and the Establishment of Trust Base, etc", pp. 1-26, Jul. 2021.

[4] Personal Information Protection Commission, "Policy Direction for Safe Use of Personal Information in the Age of Artificial Intelligence", pp. 1-20, Aug. 2023.

[5] Financial Security Institute, "AI Security Guidelines for Financial Sector", AGR-VII-2023-②-405, Apr. 2023.

[6] Jinho Yoo and 3 others, "Analysis of Security Issues and the Emergence of AI-Centered Society" KISA Insight 2022 Vol.3, pp. 1-29, Apr. 2022.

[7] LG CNS, "AI attack types", https://www.lgcns.com/blog/cns–tech/ai-d ata/9616/ (accessed 2023-05-19).

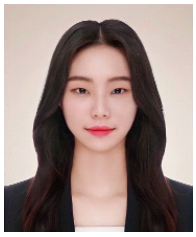
[8] National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)", Jan. 2023.

[9] Choi Daeseon, "Security and privacy issues in the AI era", TTA Journal Vol.199 no.1079, pp. 1-7, Jan. 2022.

[10] European Parliament, "Artificial

Intelligence Act", EPRS, Jun. 2023.
〔11〕 Ministry of Science and ICT, "Launch of Digital New Order Establishment Consultative Body", Korea Policy Briefing, Mar. 2023.
〔12〕 Personal Information Protection Commission·Cooperation of related ministries, "Digital Platform Government Realization Plan", Presidential Committee on the Digital Platform Government, Apr. 2023.
〔13〕 KOIT, "Personal Information Protection Commission AI Policy direction", https://www.koit.co.kr/news/articleView.html?idxno=115531, (accessed 2023-09-01).
〔14〕 Personal Information Protection Commission, "AI Personal Information Protection Self-inspection Table", May. 2021.
〔15〕 Financial Supervisory Service, "AI Guidelines for Financial Sector", Jul. 2021.
〔16〕 Financial Services Commission and 9 institutions, "Guide to AI Development and Utilization in the Financial Sector", Aug. 2022.
〔17〕 National Information Society Agency, "Data Quality Guidelines and Implementation Guide for Artificial Intelligence Learning v.3.0", Feb. 2023.
〔18〕 Ministry of Science and ICT·Korea Information Society Development Institute, "2023 Self-inspection table for the implementation of artificial intelligence ethics standards〔proposal〕", Feb. 2023.

〈저자소개〉



김 지 연 (Jiyoun Kim) 학생회원
2020년 2월: 서울과학기술대학교 컴퓨터공학과 졸업
2020년 11월~2021년 2월: 서울과학기술대학교 전기정보기술연구소 연구원
2021년 4월~2022년 5월: 서울과학기술대학교 전기정보기술연구소 연구원
2023년 3월~현재: 서울과학기술대학교 일반대학원 컴퓨터공학과 석사과정
〈관심분야〉 정보보호, 인증, 정책, 암호모듈평가 등

석 병 진 (Byoungjin Seok) 종신회원
2017년 8월: 서울과학기술대학교 컴퓨터공학과 졸업
2019년 2월: 서울과학기술대학교 컴퓨터공학과 석사
2022년 2월: 서울과학기술대학교 컴퓨터공학과 박사
2022년 3월~현재: 서울과학기술대학교 전기정보기술연구소 연구원
〈관심분야〉 정보보호, 암호학, 암호분석, 디지털포렌식 등

김 역 (Yeog Kim) 종신회원
1992년: 성신여자대학교 전산학과 이학사
2003년: 고려대학교 정보보호대학원 공학석사
2010년: 고려대학교 정보경영전문대학원 공학박사
2014년 3월~현재: 세종사이버대학교 시간강사
2017년~현재: 서울과학기술대학교 전기정보기술연구소 연구원
〈관심분야〉 정보보호, 디지털포렌식, 암호모듈평가 등

이 창 훈 (Changhoon Lee) 종신회원
2001년: 한양대학교 자연과학부 수학전공 학사
2003년: 고려대학교 정보보호대학원 석사
2008년: 고려대학교 정보경영전문대학원 정보보호전공 박사
2008년 4월~2008년 12월: 고려대학교 정보보호연구원 연구교수
2009년 3월~2012년 2월: 한신대학교 컴퓨터공학부 조교수
2012년 3월~2015년 3월: 서울과학기술대학교 컴퓨터공학과 조교수
2015년 4월~2020년 3월: 서울과학기술대학교 컴퓨터공학과 부교수
2020년 4월~현재: 서울과학기술대학교 컴퓨터공학과 교수
〈관심분야〉 정보보호, 암호학, 지능형 사이버보안 위협, 블록체인, 디지털포렌식 등