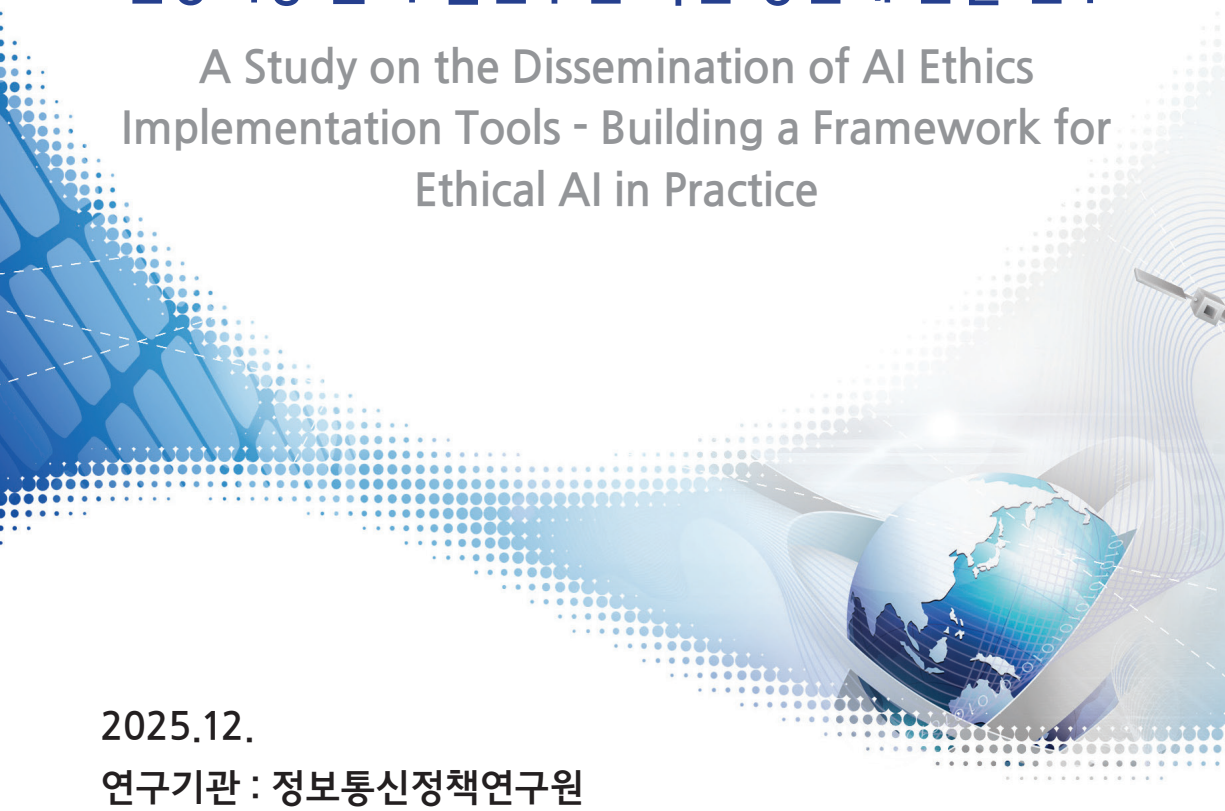
방송통신정책연구 RS-2025-16063586

정책연구 25-45

인공지능 윤리 실천수단 확산 방안에 관한 연구

A Study on the Dissemination of AI Ethics Implementation Tools - Building a Framework for Ethical AI in Practice



2025.12.

연구기관 : 정보통신정책연구원

과학기술정보통신부
Ministry of Science and ICT

IITP 정보통신기획평가원
Institute of Information & Communications Technology Planning & Evaluation

방송통신정책연구 RS-2025-16063586
정책연구 25-45

인공지능 윤리 실천수단 확산 방안에 관한 연구

A Study on the Dissemination of AI Ethics Implementation Tools - Building a Framework for Ethical AI in Practice

이현경/김휘홍/문정욱/조성은/연소라/이시직/김지혜/양기문/강하연/김소담/권지혜

주관연구기관 : **정보통신정책연구원**

이 보고서는 2025년도 과학기술정보통신부 방송통신발전기금 방송통신정책연구사업의 연구결과로서 보고서 내용은 연구자의 견해이며, 과학기술정보통신부의 공식입장과 다를 수 있습니다.

제 출 문

과학기술정보통신부 장관 귀하

본 보고서를 『**인공지능 윤리 실천수단 확산 방안에 관한 연구**』의 연구결과보고서로 제출합니다.

2025년 12월

연 구 기 관: 정보통신정책연구원
총괄책임자: 이 현 경 연구위원
참여연구원: 김 휘 홍 부연구위원
문 정 욱 실 장
조 성 은 연 구 위 원
연 소 라 부연구위원
이 시 직 부연구위원
김 지 혜 전문연구원
양 기 문 전문연구원
강 하 연 연 구 원
김 소 담 연 구 원
권 지 혜 연 구 원

목 차

표 목 차

그 림 목 차

요 약 문

1. 제 목

인공지능 윤리 실천수단 확산 방안에 관한 연구

2. 연구 목적 및 필요성

「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하 「인공지능기본법」)의 제정·시행은 국내 인공지능 윤리 거버넌스를 제도적으로 정착시키는 중요한 전환점이 되고 있다. 「인공지능기본법」은 민간 주도의 자율적 윤리 실천 확산과 민·관·학 협력 네트워크 구축을 통해 안전하고 신뢰할 수 있는 인공지능 생태계를 조성하는 것을 주요 방향으로 설정하고 있다. 이에 따라 민간자율위원회의 설치·운영을 체계적으로 지원하기 위한 표준 지침을 마련하고 보급하는 것이 윤리 거버넌스의 실효성을 제고하는 핵심 과제로 부상하고 있다.

그러나 현행 제도 환경에서는 민간의 인공지능윤리위원회 구성·운영을 위한 구체적인 지침이 충분히 마련되어 있지 않아 실제 현장에서의 적용과 운영에 어려움이 존재하는 상황이다. 이로 인해 민간 조직 내에서 윤리 거버넌스를 제도화하고 지속적으로 운영하는 데 한계가 발생하고 있다. 이러한 문제를 해소하기 위해서는 법적 강제에 의존하기보다는 민간의 자율적 윤리 준수를 유도할 수 있는 정책 도구의 개발이 필요하다. 이와 더불어 이해관계자의 폭넓은 참여를 기반으로 인공지능 윤리정책 포럼을 운영하고, 이를 통해 정책 논의와 현장 경험을 연계함으로써 윤리적 쟁점에 대한 사회적 공감대를 형성하는 것이 어느 때보다 중요한 시점이다. 궁극적으로는 이러한 노력을 통해 인공지능 생태계 전반의 윤리성과 신뢰성을 확보할 수 있는 기반을 조성하는 것이 중요하다.

3. 연구의 구성 및 범위

본 연구는 기업의 자율적인 인공지능 윤리원칙 준수를 실질적으로 지원하기 위한 방안을 체계적으로 분석하는 것을 주요 내용으로 한다. 이를 위해 인공지능 분야뿐만 아니라 유사 기술 영역에서 민간 주도로 운영되고 있는 자율적 거버넌스 사례를 폭넓게 검토하고, 각 사례의 운영 구조와 성과, 한계를 비교·분석한다. 이러한 분석 결과를 토대로 기업 환경에 적용 가능한 시사점을 도출하고, 국내 실정에 부합하는 자율적 윤리 실천 지원 방안을 모색한다.

아울러 전문가 조사와 연구반 운영을 통해 민간자율인공지능윤리위원회의 구성 및 운영을 위한 표준 지침을 개발한다. 이 과정에서는 학계, 산업계, 연구기관 등 다양한 분야의 전문가 의견을 수렴하여 지침의 실효성과 현장 적용 가능성을 제고한다. 개발된 표준 지침은 국내외 우수 사례와 함께 체계적으로 정리하여 관련 기관과 기업에 배포함으로써, 민간 부문의 자율적 윤리 거버넌스 확산을 지원할 계획이다.

한편, 인공지능 윤리와 신뢰성에 관한 사회적 논의를 활성화하기 위해 산·학·연이 참여하는 공론의 장을 마련한다. 인공지능 윤리정책 포럼을 운영하여 인공지능 윤리와 관련한 국내외 주요 정책 및 기술 동향을 공유·논의하고, 이를 바탕으로 범부처 차원의 정책 과제를 발굴한다. 특히 다양한 이해관계자의 관점을 포괄적으로 반영할 수 있는 거버넌스 체계를 구축함으로써, 정책 논의의 대표성과 정당성을 강화한다.

이와 함께 정책 과제의 구체화와 실행 가능성을 높이기 위해 산·학·연 네트워크를 구성·운영한다. 해당 네트워크는 연구 결과와 현장 경험을 연계하는 협력 플랫폼으로 기능하며, 인공지능 윤리 정책의 실질적 발전과 지속 가능한 협력 구조 형성에 기여할 것으로 기대된다.

4. 연구 내용 및 결과

본 연구는 「인공지능기본법」의 시행과 인공지능 기술 확산에 따라 제기되는 윤리적 쟁점에 대한 정책 논의 기반을 강화하고, 민간 영역에서 인공지능 윤리를 자율적으로 점검·운영할 수 있는 실천 수단을 마련하는 것을 목적으로 수행되었다. 인공지능 기술

의 사회적 영향력이 확대되는 상황에서 정부 주도의 일방향 규제보다는 민간의 자율적 책임과 다수의 이해관계자 협력이 중요해짐에 따라 실효성 있는 인공지능 윤리 거버넌스 구축 방안을 다각도로 모색하였다. 이를 위해 민간자율인공지능윤리위원회 표준지침(안)을 마련하고 인공지능 윤리정책 포럼을 운영하여 주요 윤리 쟁점을 논의하였다.

주요 연구 내용으로서 먼저 국내외 인공지능 윤리 거버넌스 동향을 분석하였다. 미국, 영국, 유럽연합(이하 EU) 등 주요국의 정책 기조와 국내 주요 테크 기업들의 자율 거버넌스 운영 실태를 검토하고 전략적 집행형, 기술-정책 연계형, 윤리 자문형, 선언적 원칙형의 네 가지 유형으로 분류하여 각 유형별 작동 기제를 진단하였다. 이를 통해 민간 기업이 인공지능 윤리를 조직 내부에서 체계적으로 내재화하기 위한 실행 프로세스의 중요성을 도출하였다.

민간자율인공지능윤리위원회 표준지침(안)은 「인공지능기본법」 제28조를 실천적으로 뒷받침하기 위해 개발되었다. 국내외 기업 및 기관의 인공지능 윤리위원회 운영 사례와 기관생명윤리위원회, 동물실험윤리위원회 등 유사 제도 분석을 바탕으로, 법학·기술·윤리 전문가로 구성된 연구반의 지속적 논의를 통해 위원회의 설치, 구성, 운영 전반을 포괄하는 13개 조문의 가이드라인을 설계함으로써 민간 조직이 자율적으로 윤리 거버넌스를 설계·운영할 수 있는 실무적 기준을 제시하였다. 특히 위원회 구성 시 성별의 다양성, 외부 전문가 참여, 전문성 확보를 최소 요건으로 제시하여 자율성을 존중하면서도 공정성과 중립성을 유지할 수 있는 제도적 균형점을 마련하였다.

아울러 민·관·학 협력의 중추적 공론장으로서 제4기 인공지능 윤리정책 포럼을 구성하고 운영하였다. 포럼은 학계·산업계·법조계·시민사회·국제기구·공공 등 다양한 분야의 전문가로 구성되었으며, 「인공지능기본법」 시대의 아젠다 전환, 소외계층의 인공지능 리터러시 강화, 데이터 거버넌스와 저작권 문제 등 급변하는 환경 속의 핵심 현안들을 입체적으로 논의하였다. 이를 통해 윤리정책과 관련된 주요 쟁점과 시사점을 정리하고 사회적 신뢰 형성을 위한 정책 자문 기능과 향후 정책 연구에 활용 가능한 논의 결과를 축적하였다.

본 연구는 인공지능 윤리가 선언적 구호를 넘어 실질적 규범으로 작동할 수 있는 실천수단을 마련하고 그 확산 방안을 도출하였다. 민간자율인공지능윤리위원회 표준지침(안)을 통해 민간 조직이 자율적으로 윤리적 의사결정 체계를 구축·운영할 수 있는 실

무적 기준을 제시하였으며, 인공지능 윤리정책 포럼을 통해 다양한 이해관계자의 의견을 정책 논의에 연결하는 소통 창구로서 거버넌스의 수용성을 제고하였다.

5. 정책적 활용 내용

해당 연구는 인공지능 기술의 확산에 따라 심화되는 윤리적 위험과 사회적 신뢰 문제에 대응하기 위해, 선언적 윤리 원칙을 넘어 실질적으로 작동 가능한 인공지능 윤리 거버넌스 체계를 구축하는 것을 목표로 추진되었다. 이러한 연구 방향과 성과는 2024년 12월 발표된 「대한민국 AI 행동계획」의 과제 82, '모두의 AI를 위한 AI 윤리 확산·고도화'에 구체적으로 반영되며 정책적 활용 성과로 이어졌다.

과제 82의 검토 배경에서는 생성형 인공지능의 고도화와 활용 확산으로 인해 기존 2020년 「인공지능 윤리기준」만으로는 새로운 기술 환경과 위험 요인을 충분히 포괄하기 어렵다는 점을 명확히 지적하고 있다. 이는 본 연구가 지속적으로 문제의식을 제기해 온 지점과 정확히 맞닿아 있다. 본 연구는 기존 윤리기준의 한계를 전제로, 국제적 정합성을 갖춘 윤리기준의 재정비 필요성과 함께, 개발·이용 전 과정에서 사업자가 스스로 점검할 수 있는 실행 가능한 윤리 체계의 필요성을 강조해 왔다. 이러한 문제 인식은 행동계획에서 윤리기준 고도화와 자율 점검 체계 마련이라는 정책 권고로 구체화되었다.

또한 과제 82에서 명시된 '사회 각계의 의견 수렴을 위한 AI 윤리 포럼 운영' 계획은, 그간 본 연구를 통해 운영·축적되어 온 인공지능 윤리정책 포럼의 성과와 경험을 제도화 단계로 확장한 것으로 볼 수 있다. 본 연구는 학계, 산업계, 법조계, 시민사회, 공공부문 등 다양한 이해관계자가 참여하는 윤리정책 포럼을 통해 인공지능 윤리를 둘러싼 쟁점을 공론화하고, 정책 설계 과정에서 사회적 합의를 도출하는 메커니즘을 실험해 왔다. 이러한 포럼 운영 경험은 행동계획에서 2027년 1분기부터 인공지능 윤리 포럼을 공식적으로 운영하고, 그 결과를 차년도 국가 인공지능 정책에 반영하겠다는 정책 권고로 이어졌다.

이는 인공지능 윤리정책 포럼이 단순한 자문이나 일회성 논의 기구가 아니라, 국가 인공지능 거버넌스 체계 내에서 정책 환류 기능을 수행하는 지속적 플랫폼으로 자리매김할 가능성을 제도적으로 인정받았다는 점에서 중요한 의미를 가진다. 다시 말해, 본

연구는 윤리 정책을 '정부 주도의 일방적 기준 설정'이 아닌, '사회적 논의를 통한 공동설계(co-design)'의 영역으로 확장하는 데 기여하였고, 이러한 접근이 국가 인공지능 행동계획에 반영되었다고 평가할 수 있다. 본 연구는 「대한민국 AI 행동계획」 과제 82가 제시하는 세 가지 핵심 방향, 즉 ① 인공지능 윤리기준의 고도화와 국제 정합성 확보, ② 개발·이용 전 과정에서 활용 가능한 자율 점검 체계 구축, ③ 사회적 합의를 기반으로 한 윤리 확산 메커니즘 마련이라는 정책 목표의 형성과 구체화에 실질적으로 기여하였다.

6. 기대효과

민간자율인공지능윤리위원회 표준지침의 보급과 확산은 자율규제를 중심으로 한 인공지능 안전성과 신뢰성 확보의 핵심 수단으로 기능할 것으로 기대된다. 해당 지침은 중소기업을 비롯한 교육·연구 기관 등이 인공지능 윤리위원회를 설치·운영하는 과정에서 실무적 가이드로 활용될 수 있으며, 조직 규모나 역량에 따른 격차를 완화하는 데 기여할 수 있다. 또한 이러한 표준지침의 확산은 인공지능 분야를 넘어 타 기술·정책 영역에서도 윤리 거버넌스 체계가 확산될 수 있는 기반을 마련하는 효과를 지닌다.

아울러 전문가 네트워크의 구성과 포럼 운영을 통해 인공지능 윤리와 신뢰성에 관한 최신 글로벌 논의를 지속적으로 공유·검토하고, 주요 윤리 쟁점에 대한 폭넓은 의견 수렴을 추진한다. 포럼을 통해 축적된 논의 결과는 향후 정책 수립 및 제도 개선 과정에서 자문 자료 및 논의 기반 자료로 활용함으로써, 정책 설계의 전문성과 현실 적합성을 제고하는 데 기여할 수 있다.

이와 같은 노력을 통해 민간 중심의 윤리 실천 체계를 안정적으로 정착시키고, 인공지능의 안전하고 책임 있는 활용을 촉진하는 것이 궁극적인 목표이다. 특히 법적 근거에 기반한 실효성 있는 실행 수단을 마련함으로써 자율적 윤리 실천이 선언적 수준에 그치지 않고 실제 운영과 의사결정 과정에 반영되도록 유도한다.

나아가 이해관계자의 참여와 공개적 논의를 전제로 한 정책 추진은 정책 실행 과정의 투명성과 정당성을 확보하는 데 중요한 역할을 한다. 이는 인공지능 거버넌스 전반에 대한 사회적 신뢰를 제고하고, 지속 가능한 정책 이행을 가능하게 하는 기반으로 작용할 것이다.

SUMMARY

1. Title

A Study on the Dissemination of AI Ethics Implementation Tools - Building a Framework for Ethical AI in Practice

2. Objective and Importance of Research

The enactment and forthcoming implementation of the Framework Act on the Advancement of Artificial Intelligence and the Establishment of a Trust-Based Foundation (hereinafter, the "AI Framework Act") mark a significant turning point in Korea's AI governance landscape. Beyond a regulatory approach centered solely on laws and institutions, there is a growing need to expand practical instruments for AI ethics through the development of voluntary ethical governance in the private sector and strengthened cooperation among the public sector, industry, and academia. In particular, the AI Framework Act includes provisions on the development and dissemination of standard guidelines for the establishment and operation of private-sector autonomous AI ethics committees (Article 28), as well as the creation of a safe and trustworthy AI usage environment. Accordingly, it has become urgent to identify and expand concrete mechanisms that support voluntary ethical practices by private actors. This is essential not only to respond to global expectations regarding the implementation of AI ethics, but also to enable domestic organizations to align with international standards for responsible AI use.

As AI technologies increasingly permeate society, the need for transparent and fair responses to their potential risks has also intensified. To this end, it is necessary to establish public forums and governance structures in which academia, industry, and

civil society can jointly participate, deliberate on AI ethics issues, and reflect diverse stakeholder perspectives in policy development. The AI Ethics Policy Cooperation Network, formed by the Korea Information Society Development Institute (KISDI) in 2022, provides an important foundation for such collaborative discussions. Building upon this network, the continued operation of a dedicated forum is required. In particular, the AI Ethics Policy Forum can function as a core platform by providing policy advice, leading public discourse on key AI ethics issues, and coordinating differing stakeholder opinions.

Furthermore, rapid socio-economic and technological changes continue to generate new ethical challenges across various sectors and application domains. Addressing these challenges effectively requires moving beyond government-led norm-setting toward strengthening the capacity of private organizations to autonomously identify, deliberate, and operationalize ethical considerations in their AI use. The establishment and practical operation of voluntary ethical governance within private entities has thus emerged as a critical component of responsible AI utilization.

Against this backdrop, establishing an AI ethics implementation framework that links private-sector autonomy with government policy objectives has become a key policy task. It is increasingly important to support private actors in taking the lead in ethical responsibility while ensuring coherence with national policy goals. The enactment and enforcement of the AI Framework Act provide an institutional basis for embedding AI ethics governance in Korea, emphasizing the expansion of private-led voluntary ethical practices and the construction of cooperative networks among government, industry, and academia.

3. Contents and Scope of the Research

In this context, this study conducts an in-depth analysis of measures to support corporate compliance with voluntary AI ethics principles. It reviews cases of private-

sector self-governance not only in the AI domain but also in related technological fields, examining their operational structures, achievements, and limitations. Based on this comparative analysis, the study derives actionable implications and proposes support measures tailored to the domestic context.

In parallel, expert surveys and research group activities are undertaken to develop standard guidelines for the establishment and operation of private autonomous AI ethics committees. By incorporating perspectives from academia, industry, and research institutions, the guidelines are designed to enhance practical applicability and effectiveness. The finalized standard guidelines, together with identified best practices, are systematically compiled and disseminated to relevant organizations, thereby supporting the broader diffusion of voluntary ethical governance in the private sector.

To further stimulate societal dialogue on AI ethics and trustworthiness, this initiative establishes a public deliberation platform involving industry, academia, and research institutions. Through the operation of the AI Ethics and Trustworthiness Forum, major domestic and international trends in AI ethics are discussed, and cross-ministerial policy agendas are identified. The forum is structured to ensure that diverse stakeholder perspectives are meaningfully reflected, thereby strengthening the representativeness and legitimacy of policy discussions. In addition, an industry-academia-research network is organized and operated to support the concretization and implementation of policy agendas, functioning as a collaborative platform that links research findings with field-level experience.

4. Research Results

This study was conducted with the objective of strengthening the policy discussion framework for ethical issues arising from the implementation of the AI Basic Act and the rapid diffusion of AI technologies, while also developing practical mechanisms that enable the private sector to autonomously assess and operationalize AI ethics. As the

societal impact of AI continues to expand, the study recognizes that effective AI ethics governance increasingly depends not on unilateral, government-led regulation, but on voluntary responsibility within the private sector and collaboration among multiple stakeholders. Accordingly, the research explored multidimensional approaches to building an effective and practicable AI ethics governance framework. To this end, a draft set of Standard Guidelines for Voluntary AI Ethics Committees was developed, and an AI Ethics Policy Forum was organized to deliberate on key ethical issues.

As a core component of the research, trends in domestic and international AI ethics governance were first analyzed. The study examined policy approaches in major jurisdictions such as the United States, the United Kingdom, and the European Union, alongside the operational practices of voluntary governance mechanisms among leading domestic technology companies. These cases were categorized into four types-strategic execution-oriented, technology-policy integrated, ethics advisory-oriented, and declarative principle-oriented-and the operational dynamics of each type were assessed. Through this analysis, the research identified the critical importance of institutionalized execution processes for systematically embedding AI ethics within private organizations.

The draft Standard Guidelines for Voluntary AI Ethics Committees were developed to provide practical support for the implementation of Article 28 of the AI Basic Act. Drawing on case studies of AI ethics committees operated by domestic and international companies and institutions, as well as analyses of analogous bodies such as Institutional Review Boards and Animal Experiment Ethics Committees, the guidelines were formulated through sustained deliberation by a research group composed of legal, technical, and ethics experts. The resulting framework consists of 13 guideline provisions covering the establishment, composition, and operation of voluntary AI ethics committees, offering practical standards that enable private organizations to independently design and operate ethical governance structures. In particular, the guidelines specify minimum requirements-such as gender diversity, participation of external experts, and professional expertise in committee composition-thereby establishing an institutional

balance that respects organizational autonomy while ensuring fairness and neutrality.

In addition, the Fourth AI Ethics Policy Forum was convened and operated as a central deliberative platform for public-private-academic collaboration. The forum brought together experts from academia, industry, the legal community, civil society, international organizations, and the public sector to engage in multidimensional discussions on pressing issues in a rapidly evolving environment, including agenda shifts in the era of the AI Basic Act, strengthening AI literacy for marginalized groups, and challenges related to data governance and copyright. Through these discussions, the forum identified key ethical issues and policy implications, accumulated deliberative outcomes that can inform future policy research, and fulfilled an advisory role in support of trust-building in AI governance.

Overall, this study advanced practical means through which AI ethics can function as an operative normative framework rather than remaining at the level of declarative principles, and identified pathways for their broader dissemination. By providing operational standards through the draft Standard Guidelines for Voluntary AI Ethics Committees, the research enabled private organizations to autonomously establish and manage ethical decision-making structures. Furthermore, through the AI Ethics Policy Forum, it enhanced the acceptability and inclusiveness of AI governance by serving as a communication channel that connects diverse stakeholder perspectives to policy deliberation.

5. Policy Suggestions for Practical Use

This research was conducted with the objective of establishing a practically operable AI ethics governance framework that goes beyond declarative principles and responds to the ethical risks and trust challenges emerging from the widespread deployment of artificial intelligence. The direction and outcomes of this research were substantively reflected in Task 82, “Advancing and Disseminating AI Ethics for AI for All,” of the

Republic of Korea AI Action Plan announced in December 2024, resulting in tangible policy uptake.

The background section of Task 82 explicitly recognizes that, with the rapid advancement and diffusion of generative AI, the 2020 AI Ethics Guidelines alone are no longer sufficient to address newly emerging risks and evolving technological environments. This assessment closely aligns with the core problem awareness that has underpinned this research. From the outset, the study emphasized the limitations of the existing ethics framework and highlighted the need to both reassess international alignment and develop actionable ethical mechanisms that can be practically applied by developers and deployers throughout the AI lifecycle. These concerns were translated into concrete policy recommendations in the Action Plan, namely the upgrading of AI ethics guidelines and the establishment of self-assessment mechanisms.

Moreover, the Action Plan's proposal to institutionalize an AI Ethics Forum for broad stakeholder engagement builds directly upon the accumulated experience and outcomes of the AI Ethics Policy Forum developed through this research. Over multiple iterations, the forum served as a deliberative platform bringing together academia, industry, the legal community, civil society, and public institutions to surface ethical issues, mediate perspectives, and inform policy design through social dialogue. The Action Plan's commitment to operating such a forum from the first quarter of 2027 onward-and to feeding its outcomes into subsequent national AI policies-represents a transition from experimental consultation to institutionalized policy feedback.

This development signifies official recognition of the AI Ethics Policy Forum not merely as an advisory or ad hoc discussion body, but as a sustained policy platform embedded within the national AI governance framework, capable of generating structured policy input. In this sense, the research contributed to reframing AI ethics policymaking from a top-down, government-centric exercise into a process of social co-design grounded in deliberation and consensus-building, an approach that is now explicitly reflected in the national Action Plan.

6. Expectations

The dissemination and expansion of standard guidelines for private autonomous AI ethics committees are expected to play a central role in securing AI safety and trustworthiness through self-regulation. These guidelines can serve as practical manuals for small and medium-sized enterprises, as well as educational and research institutions, when establishing and operating AI ethics committees, thereby reducing disparities in organizational capacity. Moreover, the diffusion of such standards can lay the groundwork for the expansion of ethical governance frameworks beyond the AI domain into other technological and policy areas.

Through the formation of expert networks and the continuous operation of forums, the initiative facilitates engagement with the latest global discussions on AI ethics and trustworthiness and systematically gathers opinions on emerging ethical issues. The outcomes of these forum discussions can be provided as advisory and reference materials for policy formulation, enhancing the expertise and contextual relevance of policy design.

Ultimately, these efforts aim to institutionalize a private-sector-centered ethical practice framework and promote the safe and responsible use of AI. By establishing effective implementation mechanisms grounded in legal authority, voluntary ethical practices are encouraged to move beyond declaratory commitments and become embedded in actual organizational operations and decision-making processes. At the same time, stakeholder participation and open deliberation contribute to ensuring transparency and legitimacy in policy implementation, thereby strengthening societal trust in AI governance and enabling sustainable policy execution.

CONTENTS

제1장 서 론

제1절 연구의 필요성 및 목적

1. 연구의 필요성 및 목적

「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하 「인공지능기본법」)이 2026년 1월 22일 시행될 예정인 가운데, 국내 인공지능 거버넌스 환경은 법·제도 중심의 규율을 넘어 민간의 자율적 윤리 실천과 이를 뒷받침하는 협력적 거버넌스 체계의 구축이 중요한 과제로 부상하고 있다. 특히 「인공지능기본법」은 민간자율인공지능윤리위원회의 구성·운영을 위한 표준지침의 마련·보급(제28조)과 안전하고 신뢰할 수 있는 인공지능 이용환경 조성(제27조)을 주요 내용으로 포함하고 있어, 민간이 자율적으로 인공지능의 신뢰성을 구축할 수 있도록 지원하는 실천 수단의 확산이 시급한 상황이다. 이는 인공지능 윤리 실천에 대한 국제적 이행 요구에 대응하고, 국내 기업과 기관이 글로벌 기준에 부합하는 책임 있는 인공지능 활용 체계를 갖추는 데 필수적인 기반이 된다.

아울러 인공지능 기술이 사회 전반에 미치는 영향이 확대됨에 따라, 기술의 잠재적 위험에 대해 투명하고 공정하게 대응하기 위한 사회적 논의 구조의 필요성도 커지고 있다. 이를 위해 학계, 산업계, 시민사회가 함께 참여하여 인공지능 윤리 이슈를 논의하고, 다양한 이해관계자의 관점을 정책과 제도에 반영할 수 있는 공론의 장과 거버넌스 체계를 지속적으로 운영할 필요가 있다. 정보통신정책연구원이 2022년부터 운영해 온 「인공지능 윤리정책 협력 네트워크」는 이러한 협력적 논의를 위한 기반으로서, 이를 토대로 한 포럼의 지속적 운영이 요구된다. 특히 '인공지능 윤리정책 포럼'은 관련 정책에 대한 자문 기능을 수행하고, 주요 윤리 이슈에 대한 사회적 토론을 주도하며, 이해관계자 간 의견을 조정하는 역할을 수행할 수 있는 핵심적 플랫폼으로 기능할 수 있다.

더 나아가 사회·경제적 변화와 기술 발전이 가속화됨에 따라, 산업 및 서비스 영역별로 새로운 윤리적 쟁점이 지속적으로 등장하고 있다. 이러한 변화에 효과적으로 대응하기 위해서는 정부 주도의 규범 제시를 넘어, 민간 기업 내부에서 인공지능 활용 과정에

서 발생하는 윤리적 문제를 자율적으로 논의하고 구체화할 수 있는 역량이 중요해지고 있다. 책임 있는 인공지능 활용을 실현하기 위해서는 민간 주체가 자율적 윤리 거버넌스를 구축하고 이를 실질적으로 운영하려는 노력이 점차 핵심 요소로 자리 잡고 있다.

가. 법적 근거

2025년 1월 21일 제정된 「인공지능기본법」은 인공지능 기술의 발전과 신뢰 확보를 동시에 달성하기 위한 국가 차원의 제도적 기반을 마련하였다는 점에서 중요한 의미를 지닌다. 특히 제4장 제27조 및 제28조는 각 기관의 자율성을 존중하는 윤리 체계의 구축과 함께, 사회 각계의 의견을 수렴하고 정책 간 정합성을 조율하는 국가 차원의 인공지능 윤리 정책 추진 체계 마련의 필요성을 명확히 하고 있다.

「인공지능기본법」 제27조는 인공지능 윤리의 제도화와 자율적 실천이 병행되어야 함을 명시적으로 요구하고 있으며, 이를 통해 인공지능 기술의 건전한 활용을 위한 민관 협력 기반의 윤리 정책 거버넌스 구축이 필수적임을 강조하고 있다. 특히 과학기술정보통신부 장관이 사회 각계의 의견을 수렴하여 인공지능 윤리원칙의 실천 방안을 수립하도록 권고하고 있다는 점은, 중앙정부가 조정자이자 촉진자로서 역할을 수행해야 함을 제도적으로 뒷받침하는 규정으로 평가된다.

〈표 1-1〉 본 연구의 추진 근거 - 「인공지능기본법」 제27조

〈제27조(인공지능 윤리원칙 등) 내용〉

① 정부는 인공지능윤리의 확산을 위하여 다음 각 호의 사항을 포함하는 인공지능 윤리원칙(이하 “윤리원칙”이라 한다)을 대통령령으로 정하는 바에 따라 제정·공표할 수 있다.
1. 인공지능의 개발 · 활용 등의 과정에서 사람의 생명과 신체, 정신적 건강 등에 해가 되지 아니하도록 하는 안전성과 신뢰성에 관한 사항
2. 인공지능기술이 적용된 제품·서비스 등을 모든 사람이 자유롭고 편리하게 이용할 수 있는 접근성에 관한 사항
3. 사람의 삶과 번영에의 공헌을 위한 인공지능의 개발·활용 등에 관한 사항

② 과학기술정보통신부장관은 사회 각계의 의견을 수렴하여 윤리원칙이 인공지능의 개발·활용 등에 관여하는 모든 사람에 의하여 실현될 수 있도록 실천방안을 수립하고 이를 공개 및 홍보·교육하여야 한다.

③ 중앙행정기관 또는 지방자치단체의 장이 인공지능윤리기준(그 명칭 및 형태를 불문하고 인공지능윤리에 관한 법령, 기준, 지침, 가이드라인 등을 말한다)을 제정하거나 개정하는 경우 과학기술정보통신부장관은 윤리원칙 및 제2항에 따른 실천방안과의 연계성·정합성 등에 관한 권고 또는 의견의 표명을 할 수 있다.

자료: 「인공지능기본법」 제4장 인공지능윤리 및 신뢰성 확보, 제27조

또한 제28조는 인공지능 관련 사업자 및 단체가 ‘민간자율인공지능윤리위원회’를 설치할 수 있도록 규정하고, 인공지능 기술을 윤리적으로 연구·개발·활용하기 위한 다양한 업무를 자율적으로 수행할 수 있는 근거를 제시하고 있다. 다만, 이러한 위원회를 아직 설치·운영하지 않고 있는 민간 중소기업이나 연구기관의 경우, 인공지능 윤리와 관련한 전문 지식과 운영 경험이 충분하지 않을 가능성이 있으며, 이로 인해 제도의 취지가 현장에서 실질적으로 구현되는 데에는 한계가 발생할 수 있다.

〈표 1-2〉 본 연구의 추진 근거 - 「인공지능기본법」 제28조

〈제28조(민간자율인공지능윤리위원회의 설치 등) 내용〉

① 다음 각 호의 기관 또는 단체는 원칙을 준수하기 위하여 민간자율인공지능윤리위원회(이하 "민간자율위원회"라 한다)를 둘 수 있다.
1. 인공지능기술 연구 및 개발을 수행하는 사람이 소속된 교육기관 · 연구기관
2. 인공지능사업자
3. 그 밖에 대통령령으로 정하는 인공지능기술 관련 기관

② 민간자율위원회는 다음 각 호의 업무를 자율적으로 수행한다.
1. 인공지능기술 연구·개발·활용에 있어서 윤리원칙의 준수 여부 확인
2. 인공지능기술 연구·개발·활용의 안전 및 인권침해 등에 관한 조사·연구
3. 인공지능기술 연구·개발·활용의 절차 및 결과에 관한 조사·감독
4. 해당 기관 또는 단체의 연구자 및 종사자에 대한 윤리원칙 교육
5. 인공지능기술 연구·개발·활용에 적합한 분야별 인공지능윤리 지침 마련
6. 그 밖에 윤리원칙 구현에 필요한 업무

③ 민간자율위원회의 구성·운영 등에 필요한 사항은 해당 기관 또는 단체 등에서 자율적으로 정한다. 다만, 그 구성을 특정한 성(性)으로만 할 수 없으며, 사회적·윤리적 타당성을 평가할 수 있는 경험과 지식을 갖춘 사람 및 그 기관 또는 단체에 종사하지 아니하는 사람을 각각 포함하여야 한다.

④ 과학기술정보통신부장관은 민간자율위원회의 공정하고 중립적인 구성·운영을 위하여 표준 지침 등을 마련하여 보급할 수 있다.

자료: 「인공지능기본법」 제4장 인공지능윤리 및 신뢰성 확보, 제28조

이와 더불어 인공지능 윤리 정책 추진의 법적 근거로는 「지능정보화 기본법」과 「정보통신산업진흥법」도 중요한 역할을 한다. 지능정보화 기본법은 2020년 전부 개정을 통해 데이터와 인공지능 등 핵심 기술 기반과 산업 생태계를 강화하는 한편, 제4차 산업혁명 과정에서 발생할 수 있는 부작용에 대응하기 위한 사회적 안전망을 마련하는 데 목적을 두고 있다. 이를 통해 국가 경쟁력을 제고하고 국민의 삶의 질 향상에 기여하고자 하였다.

특히 동 법률에는 지능정보서비스 등의 사회적 영향 평가를 포함하여 지능정보사회에서 발생할 수 있는 역기능을 해소하고 예방하기 위한 규정이 신설되었다. 제56조는 지능정보서비스 등의 사회적 영향 평가를 규정하고 있으며, 제62조는 지능정보사회 윤리에 관한 조항을 신설하여 인공지능과 지능정보기술의 활용이 사회적 가치와 윤리 원칙에 부합하도록 유도하고 있다. 이러한 법적 체계는 「인공지능기본법」과 함께 우리나라 인공지능 윤리 정책을 뒷받침하는 중층적 제도 기반으로 기능하고 있다.

나. 연구의 목표

본 과제는 법적 근거에 기반하여 인공지능 윤리 정책의 실효적 이행을 지원하는 것을 목적으로 한다. 구체적으로는 인공지능 윤리정책 포럼을 구성하고 이를 체계적으로 운영함으로써 사회 각계의 의견을 폭넓게 수렴하고, 민간자율인공지능윤리위원회의 설치 및 운영을 지원하기 위한 표준 지침을 마련하는 데 중점을 둔다. 이를 통해 선언적 수준에 머무르지 않고 현장에서 실제로 작동 가능한 인공지능 윤리 실천 체계의 구축에 기여하고자 한다.

이러한 접근은 정부가 추진하는 인공지능 윤리 정책의 이행력을 제고하는 동시에, 민간 주체의 자율성과 책임성을 강화하는 효과를 기대할 수 있다. 궁극적으로는 공공과 민간이 협력하는 윤리 거버넌스 기반을 통해 신뢰받는 인공지능 생태계를 조성하는 데 핵심적인 역할을 수행할 것으로 평가된다.

다. 연구의 내용

본 과제는 민간자율인공지능윤리위원회(이하 민간자율위원회)의 설치와 운영을 지원하기 위한 표준지침 마련과 관련 기반 연구를 주요 내용으로 한다. 민간자율위원회는 「인공지능산업 발전과 신뢰 기반 조성 등에 관한 기본법(약칭:인공지능기본법)」 제28조에 근거하여, 단일의 기관 또는 단체가 내부에 임의적으로 설치할 수 있는 위원회를 의미한다. 인공지능 기술의 연구·개발을 수행하는 교육 및 연구기관과 인공지능 사업자 등은 인공지능 기술의 연구·개발·활용 과정에서 윤리 원칙의 준수 여부를 점검하고 윤리적 운영을 도모하기 위해 민간자율위원회를 설립·운영할 수 있으며, 이를 위해 관련 정보의 공유와 실무에 활용 가능한 가이드라인 마련이 요구된다.

이러한 표준지침 마련의 취지는 인공지능 윤리의 확보를 법적으로 강제하기보다는, 인공지능 기술을 개발·연구·활용하는 기업, 연구기관, 교육기관, 공공기관 등이 자체적인 거버넌스 조직을 활용하여 인공지능 윤리 원칙을 자율적으로 준수하도록 유도하는 데 있다. 즉, 민간의 자율성과 책임을 전제로 한 윤리 거버넌스의 확산을 통해 현장 중심의 인공지능 윤리 실천을 촉진하고자 하는 것이다.

아울러 본 과제는 인공지능 윤리정책 포럼의 구성과 운영을 중요한 추진 과제로 포함한다. 인공지능 확산에 따라 발생하는 국내외 주요 현안과 사회적 영향에 대해 민·관·

학의 주요 이해관계자가 함께 참여하여 논의하는 공론장의 지속적인 운영은 국내 인공지능 산업의 지속 가능한 성장과 사회적 신뢰 확보를 위해 필수적이다. 이에 따라 그간 국내에서 운영되어 온 인공지능 윤리정책 포럼의 성과와 경험을 계승·발전시켜 새로운 포럼을 구성하고 운영함으로써, 민간자율위원회 표준지침 마련을 포함한 정부의 주요 정책 현안에 대해 폭넓은 의견을 수렴할 수 있는 공식적인 소통 창구로 활용하고자 한다.

2. 연구 추진체계와 전략

가. 민간자율위원회 표준지침 마련 연구

본 과제는 민간자율인공지능윤리위원회 표준지침 마련을 위해 국내외 민간 자율 인공지능 거버넌스 사례를 체계적으로 조사·분석하고, 이를 토대로 실효성 있는 표준지침을 도출하는 것을 주요 내용으로 한다.

먼저 국내외 민간 자율 인공지능 거버넌스 사례 연구를 수행한다. 「인공지능기본법」 제28조에 따라 민간자율위원회의 공정하고 중립적인 구성과 운영을 지원하기 위한 표준지침을 마련하기 위해, 유사 윤리위원회의 유형과 운영 현황을 조사한다. 민간자율위원회의 설치 여부는 원칙적으로 각 기관 또는 단체의 자율에 속하나, 동조 제3항에서는 성별, 전문성, 투명성을 고려한 위원 구성 요건을 규정하고 있으며, 제2항에서는 위원회의 업무 범위까지 제시하고 있어 자율성과 실효성 측면에서 논의의 여지가 존재한다. 또한 동조 제4항에서 규정한 정부의 표준지침 마련 및 보급 권한은, 각 기관이나 단체가 내부에 임의로 설치하는 위원회를 대상으로 한다는 점에서 타 법률에서는 유사 사례를 찾아보기 어려운 이례적인 규정으로 평가된다.

헌법상 보장된 영업의 자유와 결사의 자유에 따라 기업을 포함한 기관이나 단체는 내부 조직 구성 방식에 있어 기본적으로 자율성을 갖지만, 생명윤리 및 안전에 관한 법률상의 기관생명윤리위원회, 호스피스·완화의료 및 임종과정에 있는 환자의 연명의료결정에 관한 법률상의 의료기관윤리위원회, 동물보호법상의 동물실험윤리위원회와 같이 특정 윤리위원회의 설치와 운영을 법률로 의무화한 사례도 존재한다. 이러한 선행 사례를 비교·분석함으로써 민간자율위원회에 대한 표준지침 마련의 법적·제도적 맥락을 검토한다.

아울러 거대 글로벌 빅테크 기업을 포함한 인공지능 사업자 및 인공지능 기술 교육·연구기관 내에서 운영되고 있는 내부 인공지능 윤리 거버넌스 사례를 분석한다. Google, Microsoft, IBM, Meta, 네이버, LG, 카카오, KB금융그룹, SK텔레콤, 엔씨소프트, 크래프톤, NHN 등의 사례를 대상으로, 기업 내 자율위원회 또는 거버넌스 구조의 변화와 운영 실태를 검토한다. 이를 통해 위원회의 성격, 구성 방식, 의사결정의 구속력 등 국내 기업에 공통적으로 적용 가능한 요소를 도출하고, 전문가 조사 및 유형화 작업에 활용할 수 있도록 정리한다.

이와 함께 표준지침 마련을 위한 연구반을 운영한다. 인공지능 관련 전문가와 인공지능 윤리정책 포럼 위원 등을 대상으로 전문가 조사를 실시하여, 민간자율위원회의 설립과 운영에 필요한 표준지침 구성 요소에 대한 의견을 수렴한다. 해당 의견은 연구반에 제공되어 표준지침 초안에 실질적으로 반영된다. 「인공지능기본법」은 민간자율위원회 구성 요건으로 특정 성별로만 구성하지 않을 것, 사회적·윤리적 타당성을 평가할 수 있는 경험과 지식을 갖춘 인사를 포함할 것, 그리고 외부 인사를 포함할 것을 최소한의 의무 요건으로 제시하고 있으므로, 이러한 법적 요건을 전제로 현장 적용 가능성을 높이는 방향으로 지침 내용을 구체화한다.

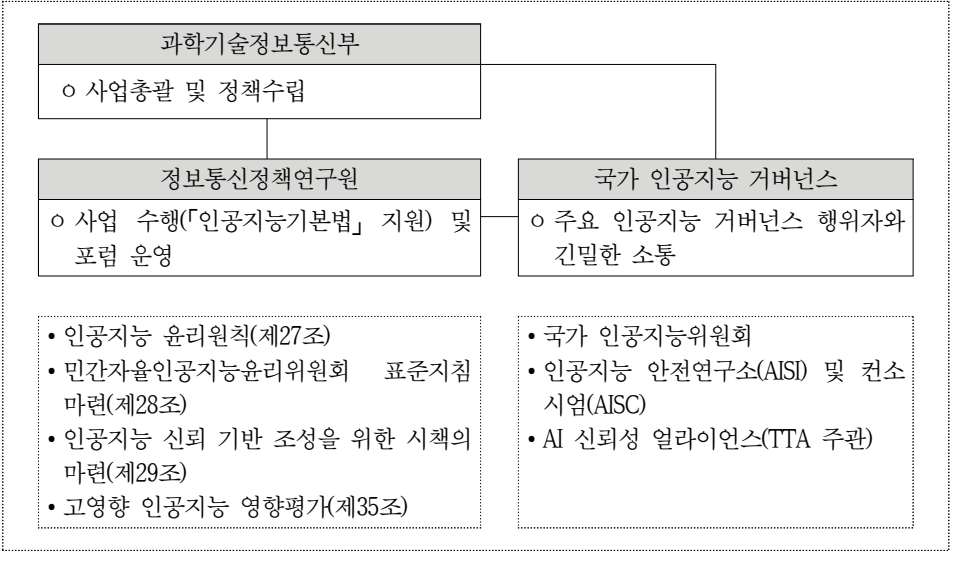
연구반 운영을 통해 도출되는 표준지침은 적용 대상의 특성을 고려하여 두 가지 유형으로 구분하여 마련한다. 첫째는 인공지능 기술의 연구 및 개발을 수행하는 교육·연구기관을 위한 지침이며, 둘째는 인공지능 사업자를 대상으로 한 지침이다. 「인공지능기본법」이 민간자율위원회 설립 대상으로 제시한 이 두 유형의 기관은 조직 구성과 역할, 강조해야 할 수행 내용에 차이가 있을 수 있으므로, 전문가 조사 결과를 바탕으로 법학 전문가로 구성된 연구반에서 세부 내용을 정리한다.

마지막으로 마련된 표준지침과 함께 우수 사례를 정리하여 배포한다. 민간자율위원회 구성과 운영에 관한 표준 지침과 국내외 민간 자율적 인공지능 거버넌스의 우수 사례를 민간 현장에서 쉽게 활용할 수 있도록 체계적으로 정리하여 배포함으로써, 민간 부문의 자율적 인공지능 윤리 거버넌스 확산을 지원하고자 한다.

나. 인공지능 윤리정책 포럼의 구성 및 운영

본 과제는 인공지능 윤리 및 신뢰성 정책의 실효성을 제고하기 위해 인공지능 윤리정책 포럼을 구성·운영하고, 이를 통해 사회적 논의와 정책 자문 기능을 체계화하는 것을 목적으로 한다.

〔그림 1-1〕 인공지능 윤리체계 확산 기반 조성을 위한 부처간 협의 계획



자료: 연구진 작성

우선 포럼 구성에 앞서 정책 수요 조사를 실시한다. 과학기술정보통신부의 정책 수요를 확인하고, 기존의 국가 인공지능 거버넌스 체계에서 충분히 다루기 어려운 윤리적 쟁점이나 윤리의식 확산과 관련된 사안을 포괄적으로 검토한다. 이를 위해 포럼 위원 구성 방향과 주요 논의 안건 선정에 대해 주무 부처와 긴밀한 협의를 진행함으로써, 포럼이 실질적인 정책 지원 기구로 기능할 수 있도록 한다.

이러한 협의 결과를 바탕으로 제4기 인공지능 윤리정책 포럼을 구성한다. 포럼 위원은 과학기술정보통신부와의 소통을 전제로 약 20명 내외로 구성하며, 학계, 산업계, 법조계, 공공 부문, 시민사회, 국제기구 등 다양한 분야의 전문가를 포함한다. 특히 여타 국가 인공지능 거버넌스 체계에서는 시민단체의 참여가 제한적인 경우가 많다는 점을

고려하여, 인공지능 윤리정책 포럼을 통해 시민의 목소리를 정책 논의에 반영할 수 있는 기회를 마련하는 데 중점을 둔다. 이는 국가 주도의 윤리 원칙이 형식적으로는 진보적 가치를 표방하면서도, 실제로는 구조적 불평등이나 차별을 은폐할 수 있다는 비판을 반영한 것으로, 윤리 거버넌스의 대표성과 포용성을 강화하기 위한 시도로 평가된다.

포럼 운영의 목표는 인공지능의 범용적 활용이 확대되는 환경에서 발생하는 다양한 윤리적 쟁점에 대해 심층적인 논의를 진행하고, 이에 대한 정책적 대응 방안을 모색하는 데 있다. 포럼을 통해 정부 부처가 추진 중인 주요 인공지능 윤리 정책 과제에 대해 자문과 논의를 제공함으로써, 정책 추진 과정에서 절차적 정당성을 확보할 수 있을 것으로 기대된다. 특히 국책 사업 추진 과정에서 발생할 수 있는 사회적 갈등을 예방하고 사회적 합의를 도출하기 위해서는, 다양한 이해관계자가 정책 과정에서 배제되지 않고 담론 형성에 참여할 수 있는 공론장의 마련이 중요하다.

포럼은 학계, 산업계, 법조계, 공공 부문, 시민사회, 국제기구 등 각 부문을 대표하는 전문가 약 20명으로 구성되며, 인공지능 윤리에 관한 다양한 관점을 포괄하는 동시에 집중된 논의를 통해 심도 있는 해결책을 도출할 수 있도록 설계한다. 운영 방식은 출범식을 포함하여 격월로 포럼을 개최하는 것을 원칙으로 하며, 총 3회 내외의 회의를 진행할 예정이다. 다만 구체적인 운영 시기와 방식은 과학기술정보통신부와의 협의를 거쳐 최종 확정한다.

포럼의 주요 논의 내용으로는 인공지능 윤리 정책의 기본 방향에 대한 검토, 국제사회의 관련 논의 동향에 대한 대응 방안, 그리고 윤리 정책 실행 수단의 개선 및 활용도 제고 방안 등이 포함된다. 구체적으로는 인공지능 기본권 영향평가(안), 인공지능 윤리 기준 자율점검표(안), 민간자율인공지능윤리위원회 운영 방안, 인공지능 윤리 교육의 방향성 등이 주요 논의 대상이 된다. 또한 과학기술정보통신부의 정책 수요를 적극 반영하여, 생성형 인공지능 기술의 확산과 함께 새롭게 부각되는 거짓 정보, 편향, 혐오 표현, 허위·조작 정보 등 윤리적 이슈에 대한 정책적 접근을 유도하는 논의를 중점적으로 진행할 계획이다.

다. 추진 전략 및 방법

과제를 수행하기 위해 문헌 연구와 함께 인공지능 산업계 및 법학자를 대상으로 한 전문가 조사를 실시하고, 표준지침 마련을 위한 연구반을 운영함으로써 민간 인공지능 자율 거버넌스를 뒷받침할 수 있는 실질적인 기반을 조성한다.

아울러 본 과제는 민간의 자율성과 정부 정책 간의 일관성 사이에서 균형을 추구한다. 민간 주체의 자율적 거버넌스를 장려하되, 정부 정책과의 정합성을 유지할 수 있도록 지침과 운영 체계를 체계적으로 설계함으로써 정책 전반에 대한 신뢰도를 제고하고자 한다. 이를 위해 정보통신정책연구원의 인공지능 윤리 연구진이 축적해 온 전문성을 바탕으로, 기존에 개발된 인공지능 윤리 콘텐츠와 연계하고, 인공지능 기본권 영향평가와 자율점검표 등 관련 정책 도구에서 요구되는 내용을 반영하여 정합성 있는 표준지침을 개발한다.

또한 다층적 거버넌스 연계와 협력 기반 구축을 중요한 추진 방향으로 설정한다. 기존 제1·2·3기 인공지능 윤리정책 포럼의 성과를 계승하고, 학계와 산업계로 구성된 전문가 자문단 네트워크를 적극 활용함으로써 정책적 완성도와 실효성을 갖춘 성과물을 도출하고자 한다. 이와 함께 국가인공지능위원회, AI 신뢰성 얼라이언스, 인공지능 안전연구소 등 주요 거버넌스 기구와의 정책 연계를 고려하여, 포럼 논의 결과가 실질적인 정책 제안으로 이어질 수 있도록 연계 기반을 마련한다.

마지막으로 본 과제는 정책 수요에 기반한 안건 발굴과 대응적 논의를 중시한다. 과학기술정보통신부 등 주무 부처와의 긴밀한 협의를 통해 현안 중심의 논의 주제를 설정하고, 국제사회의 인공지능 윤리 이슈와 국내에서 민감하게 대두되는 사안에 대해 선제적으로 대응할 수 있는 논의 체계를 구축함으로써 정책 환경 변화에 능동적으로 대응하고자 한다.

〈표 1-3〉 주요 연구 구성 요약

구분	내용
민간자율위원회 표준지침 마련 연구	▸「인공지능기본법」에 근거해 자율성과 참여 기반의 윤리체계 마련 -(민간의 자율적 인공지능 윤리위원회 현황 및 주요 기업 사례 연구) -(전문가 조사 및 연구반 운영을 통한 표준지침 개발) 민간자율위원회를 활용하고자 하는 주체가 목적, 특징, 특성에 맞추어 기관 내 자율 거버넌스를 확립할 수 있도록 표준지침 제공
인공지능 윤리정책 포럼 구성 및 운영	▸다양한 이해관계자의 의견을 반영하는 인공지능 윤리 거버넌스 체계 구축 ▸정책 수요 기반 안건 발굴 및 대응적 논의 -(정부 정책과 민간의 자율성 간 균형 확보) 산업·기술 변화에 대응하는 책임 있는 인공지능 활용 확산

자료: 연구진 작성

라. 연구의 활용방안 및 기대효과

본 연구개발 성과는 민간자율인공지능윤리위원회의 설립과 운영을 촉진하기 위한 실무적 가이드로 활용될 수 있다. 중소기업, 스타트업, 연구기관 등 다양한 민간 주체가 인공지능 윤리위원회를 자율적으로 설치·운영할 수 있도록 구체적인 절차와 구성 기준을 제시함으로써, 현장에서 즉시 적용 가능한 실용적 지침으로 기능할 것으로 기대된다. 특히 「인공지능기본법」의 취지에 부합하는 민간자율인공지능윤리위원회 표준지침을 제공함으로써, 민간 분야에서 인공지능 거버넌스 사례를 공유하고 각 기관이 자체적인 표준지침을 마련하는 데 기초자료로 활용될 수 있다.

아울러 본 연구 결과는 정부의 인공지능 윤리 정책 집행 과정에서 자문 및 근거 자료로 활용 가능하다. 인공지능 기본권 영향평가(안), 윤리 자율점검표, 윤리 교육 체계 설계 등 주요 윤리 정책 과제를 추진하는 과정에서 정책 설계와 실행을 뒷받침하는 참고자료로 기능할 수 있다. 이를 통해 정책 추진의 일관성과 실효성을 제고하는 데 기여할 것으로 기대된다.

또한 민간자율 윤리위원회 모델은 향후 보건의료, 교육, 국방 등 다른 정책·산업 분야로 확장될 수 있는 범용 윤리 거버넌스 모델로 발전할 가능성을 지닌다. 개발된 표준지침과 민간의 우수 사례를 국내외 학술대회, 국제 거버넌스 네트워크, 정부 보고서 등을 통해 공유함으로써, 우리나라 인공지능 윤리 모델에 대한 국제적 인지도와 신뢰도를 제

고하는 데에도 활용할 수 있다.

이러한 연구를 통해 산업 성장과 기술 혁신을 저해하지 않으면서도 인공지능의 위험과 부작용을 예방할 수 있는 자율규제 기반의 인공지능 윤리 체계를 정립하고 확산하는 데 기여할 것으로 기대된다. 현재 우리나라는 인공지능 산업의 성장과 기술 혁신을 유지하는 동시에, 급변하는 기술 환경에 대응하여 부작용을 최소화할 수 있는 정책적 대응이 요구되는 시점에 있다. 이 과정에서 강제력과 구속력을 갖는 법 규제로는 다루기 어려운 영역에 대해 윤리적·연성적 접근 방식의 체계를 구축하는 것이 특히 중요하다. 전 세계적으로 인공지능 혁신이 빠르게 전개되는 상황에서, 이른바 '글로벌 AI 경쟁'에서 선도적 역할을 수행하기 위해서는 인공지능 윤리 역량을 갖춘 시민과 기업의 역할이 핵심적인 요소로 작용한다.

또한 다양한 이해관계자의 참여를 전제로 한 인공지능 윤리정책 포럼 운영을 통해 윤리 정책 추진 과정의 투명성과 사회적 수용성을 제고할 수 있다. 이는 정부 정책의 정당성과 실행력을 강화하고, 국민적 신뢰를 기반으로 한 지속 가능한 정책 추진을 가능하게 하는 효과로 이어질 것으로 기대된다.

나아가 본 연구는 유럽연합 인공지능법(EU AI Act), 경제협력개발기구 인공지능 원칙(OECD AI Principles), 유네스코(The United Nations Educational, Scientific and Cultural Organization, UNESCO) 인공지능 윤리 권고안(Recommendation on the Ethics of Artificial Intelligence) 등 국제사회의 주요 윤리기준과 정합성을 갖춘 한국형 인공지능 윤리 이행 모델을 제시함으로써, 국제 협력과 국제 거버넌스 논의에서 우리나라의 위상과 리더십을 강화하는 데 기여할 수 있다.

마지막으로 인공지능 윤리·신뢰성 정책의 추진 과정과 성과를 국민에게 투명하게 공개하고, 의견을 수렴하는 체계를 마련함으로써 국민의 참여권을 보장하고 공공 신뢰를 제고할 수 있다. 이를 통해 인공지능의 윤리적 활용을 기반으로 국민의 안전과 권리, 포용성이 보장되는 디지털 사회로의 이행을 촉진하고, 인공지능에 대한 사회적 신뢰를 지속적으로 강화하는 효과를 기대할 수 있다.

제2장 국내외 윤리 거버넌스 동향

제1절 공공 인공지능 윤리 거버넌스

1. 해외 현황

가. 미국

미국에서는 인공지능의 안전성과 신뢰성을 제고하기 위한 다양한 정책·제도적 장치가 단계적으로 마련되어 왔다. 먼저 인공지능 안전연구소(AI Safety Institute)는 고도화된 인공지능 모델의 평가 역량을 강화하고, 인공지능 안전과 관련된 규칙을 제정·시행하려는 규제 당국이 활용할 수 있는 기술적 가이드를 제공하는 역할을 수행할 것으로 기대된다. 이는 인공지능 안전, 보안, 신뢰에 대한 연방 정부의 권한과 책임을 강화하려는 조치의 일환으로, 인간이 생성한 콘텐츠의 진위 확인, 인공지능 생성 콘텐츠의 워터마킹, 해로운 알고리즘 차별의 식별 및 완화, 투명성 확보, 프라이버시 보호형 인공지능의 도입 등 개발자의 역할이 중요한 분야를 포괄하고 있다.

이와 함께 미국 국립표준기술연구소(National Institute of Standards and Technology, NIST)가 2023년에 발표한 「인공지능 위험관리 프레임워크(AI Risk Management Framework)」는 산업별 위험 수준에 따라 기업이 자율적으로 적용할 수 있는 관리 체계를 제시하고 있다. 해당 프레임워크는 법적 강제력을 갖추고 있지는 않으나, 기업이 위험 기반 접근을 토대로 인공지능 거버넌스를 구축하도록 유도하는 사전적 윤리·기술 표준으로 기능한다는 점에서 정책적 의미를 지닌다.

또한 2022년에 발표된 「인공지능 권리장전(AI Bill of Rights)」은 인공지능 기술 확산 과정에서 미국 시민의 권리를 보호하기 위한 자율 규제 가이드로서, 인공지능이 공공에 부정적인 영향을 미치지 않도록 하기 위한 다섯 가지 기본 원칙을 제시하였다. 해당 원칙은 안전하고 효과적인 시스템의 보장, 알고리즘 기반 차별로부터의 보호, 데이터 프라이버시 보장, 통지 및 설명의 제공, 인간에 대한 보호를 핵심 내용으로 하고 있다.

이러한 정책 흐름 속에서 미국 전임 바이든 대통령은 2023년 10월 30일 인공지능의

안전한 활용과 국가 안보 차원의 대응을 강화하기 위해 「안전하고 신뢰할 수 있는 인공지능에 관한 행정명령(EO 14110)」에 서명하였다. 해당 행정명령은 안전 및 보안, 혁신과 경쟁 촉진, 근로자 지원, 형평성과 시민권 보호, 소비자 보호, 개인정보 보호, 연방정부의 인공지능 활용, 국제 관계에서의 리더십 확보 등 여덟 가지 원칙을 중심으로 연방정부 차원의 인공지능 정책 방향을 제시하였다.

그러나 이후 출범한 트럼프 대통령 행정부는 EO 14110을 포함한 기존 인공지능 정책을 철회하거나 수정하는 방향으로 정책 기조를 전환하였다. 2025년 1월 23일 서명된 「미국 인공지능 리더십 장벽 제거 행정명령(EO 14179)」은 산업 중심의 인공지능 개발 환경 조성을 강조하며, 미국의 경제적 경쟁력과 국방력 강화를 핵심 목표로 설정하고 있다. 해당 행정명령은 미국의 인공지능 리더십 강화, 이념적 편향 제거, 기존 정책의 철회, 새로운 인공지능 행동 계획 수립 등을 주요 내용으로 포함하고 있으며, 인공지능 거버넌스에 대한 접근 방식이 규범·안전 중심에서 산업·경쟁력 중심으로 이동하고 있음을 보여준다.

나. 영국

영국은 인공지능 정책에 있어 윤리와 혁신의 조화를 핵심 원칙으로 삼고, 정부가 기술 기업, 학계, 규제 기관 간의 대화를 촉진하는 메커니즘을 마련함으로써 인공지능 개발이 윤리적 원칙에 부합하는 동시에 혁신을 저해하지 않도록 하는 접근 방식을 취하고 있다. 이러한 거버넌스 구조 하에서 영국의 기술 기업들은 정책 형성과 이행 과정에 적극적으로 참여하고 있으며, 다수의 기업이 정부의 요구 사항을 충족하는 수준을 넘어 이를 보완하거나 상회하는 자체적인 윤리 프레임워크를 개발·운영하고 있는 것으로 나타난다.

영국에서 정의하는 '인공지능 안전'은 고의적이거나 우발적으로 인공지능으로 인해 발생할 수 있는 피해에 대한 이해, 예방 및 완화를 포괄하는 개념이다. 여기에는 개인, 집단, 조직, 국가 또는 전 세계적 차원에서 발생할 수 있는 신체적·정신적·사회적·경제적 피해 등 다양한 유형의 위해가 포함된다. 이러한 폭넓은 정의는 인공지능의 위험을 단순한 기술적 오류의 문제를 넘어 사회 전반의 영향으로 인식하고 있음을 보여준다.

영국 정부는 인공지능의 위험을 규율함에 있어 증거 기반의 비례적 접근을 중시하고

있으며, 이를 위해 거버넌스 체제(Governance Regime)에 실질적인 통찰을 제공할 수 있는 기초연구를 적극적으로 지원하고 있다. 특히 새로운 유형의 인공지능을 검사·평가·테스트함으로써 안전성에 대한 지식을 축적하고, 예측하기 어려운 기술 발전으로부터 사람들을 보호하는 데 있어 학자와 개발자의 역할을 중요하게 인식하고 있다. 이는 규제와 연구를 병행하는 영국식 인공지능 안전 접근의 특징으로 평가된다.

이와 더불어 영국은 인공지능 관련 단일 규제 기관을 설립하여 전반적인 권한을 부여하고, 상황별로 맞춤형 접근 방식을 채택하는 방안을 추진하고 있다. 2023년 말에는 국제 인공지능 안전 정상회의(Global AI Safety Summit)를 개최하여 국제적 논의를 주도하였으며, 국제 인공지능 감시기구를 런던에 유치하려는 계획을 통해 국제 인공지능 거버넌스에서의 주도적 역할을 강화하고자 하고 있다.

영국은 인공지능의 사회적 영향 평가 역시 중요한 정책 과제로 다루고 있다. 사회적 영향 평가는 심리적 영향, 프라이버시 침해, 조작과 설득 가능성, 편향된 생성물과 추론, 민주주의에 대한 영향, 시스템화된 차별 등 다양한 요소를 포함한다. 이러한 평가는 인공지능 시스템 평가의 우선순위 중 두 번째 단계로 설정되어 있으며, 개발 기업이 사후 평가(post-deployment)를 수행할 수 있도록 요구되는 영역이기 때문에, 사전에 이를 고려한 설계와 운영이 필요하다는 점이 강조되고 있다.

또한 영국은 인간 감독자가 인공지능 시스템의 행동을 효과적으로 통제하기 어려운 상황에서 발생할 수 있는 '통제력 상실'의 위험을 주요 이슈로 상정하고 있다. 이에 따라 인공지능 시스템이 인간 감독자를 속일 수 있는 능력, 자동적으로 복제할 수 있는 능력, 인간의 개입 시도를 조정하거나 무력화할 수 있는 능력, 더 강력한 인공지능 시스템을 스스로 창조할 수 있는 능력 등과 같은 핵심 기능을 중심으로 위험 평가를 수행하고 있으며, 이는 고도화된 인공지능에 대한 사전적 안전 관리의 중요한 기준으로 활용되고 있다.

다. 캐나다

캐나다는 정책 네트워크를 기반으로 한 인공지능 윤리 거버넌스 모델을 구축하고 있으며, 다양한 이해관계자가 참여하는 분산형 인공지능 정책 생태계를 특징으로 한다. 이러한 구조 하에서 정부, 연구기관, 민간 부문 등 각 주체는 비교적 자율적인 방식으로

정책 형성 과정에 관여하고 있으며, 중앙집권적 규제보다는 협력과 조정을 중심으로 한 거버넌스 접근을 취하고 있다.

범캐나다 인공지능 전략은 이러한 정책 기조를 구체화한 대표적 사례로, 정부는 캐나다 경제와 사회 전반에 걸친 인공지능 도입을 촉진하기 위해 세계적 수준의 인재와 연구 역량을 상용화·지원하는 프로그램과 연계된 정책을 추진하고 있다. 이를 통해 캐나다에서 창출된 지식과 아이디어가 해외로 유출되지 않고 자국 내에서 활용·상용화될 수 있도록 지원하는 데 중점을 두고 있다. 이와 함께 인공지능 자문위원회의 개편, 인공지능에 관한 국제협력 참여, 인공지능 시스템 관리자를 위한 가이드 발간, 자발적 인공지능 행동 강령에 대한 신규 서명국 발표 등 다양한 정책적 노력이 병행되고 있다.

다만 캐나다의 인공지능 거버넌스 구조는 정부와 연구기관 중심으로 운영되는 경향이 강해, 시민사회나 산업계의 실질적 영향력이 상대적으로 제한된다는 한계도 지적되고 있다. 이에 따라 정책 기관 간 연계성을 강화하고 시민사회의 참여를 확대할 필요성이 지속적으로 제기되고 있으며, 규제 설계 과정에서 대표성과 포용성을 제고하기 위한 거버넌스 구조 개편의 필요성도 함께 논의되고 있다. 이러한 과제는 캐나다형 분산 거버넌스 모델의 지속가능성을 확보하기 위한 핵심 쟁점으로 평가된다.

라. 호주

호주는 연방제 국가의 특성을 반영하여 다층적인 인공지능 윤리 거버넌스 구조를 구축하고 있으며, 주정부와 연방정부 간의 상호작용을 통해 정책의 실험과 확산을 병행하는 접근 방식을 취하고 있다. 대표적으로 뉴사우스웨일스(New South Wales, NSW)주는 호주 최초로 인공지능 윤리 정책을 수립하고, 이를 연방 차원으로 확산하는 모델을 제시하였다. 이 과정에서 주정부 차원의 실험적 정책 모델과 연방 표준 간의 상호 피드백이 이루어지며, 하향식(top-down)과 상향식(bottom-up)이 결합된 융합적 정책 체계가 형성되고 있다.

이와 함께 호주는 공공부문을 중심으로 인공지능 보증 프레임워크(AI Assurance Framework)를 도입하여, 공공조달 과정에서 인공지능 시스템이 윤리 기준과 성능 요구사항을 충족하는지를 사전에 평가하도록 하는 체계를 설계하였다. 해당 보증 제도는 법적 강제력을 전제로 하기보다는, 공공기관과 시민 간의 신뢰를 확보하기 위한 예방적

수단으로 기능한다는 점에서 특징을 지닌다. 이를 통해 공공 영역에서 활용되는 인공지능 시스템의 책임성과 신뢰성을 사전에 점검하고, 위험을 최소화하는 것을 목표로 하고 있다.

연방정부 차원에서는 이른바 '퓨처 테크(Future Tech)' 정책을 통해 인공지능, 양자기술, 로봇공학을 국가 핵심 기술로 선정하고, 인공지능 로드맵을 발표하였다. 해당 로드맵은 인공지능을 통한 경제 성장 촉진과 일자리 창출을 지원하는 동시에, 인공지능 윤리 프레임워크 개발 등 필요한 규제 수단을 함께 마련하는 것을 주요 내용으로 하고 있다. 이는 기술 혁신과 윤리·규범 간의 균형을 추구하는 호주 정부의 정책 기조를 반영한다.

아울러 2024년에는 공공서비스 전반에서 인공지능의 활용이 확대됨에 따라, 호주 정부는 공공서비스를 위한 인공지능의 책임 있는 사용을 안내하기 위해 '자발적 인공지능 안전 표준(Voluntary AI Safety Standard)' 발표하였다. 해당 정책은 거버넌스, 보증, 투명성에 관한 기본 요건을 설정함으로써, 공공서비스 기관과 공무원 전반에 걸쳐 인공지능 활용에 대한 일관된 접근 방식을 확립하는 데 목적을 두고 있다. 관련된 후속 조치로 '인공지능 도입을 위한 가이드라인(Guidance for AI Adoption)'을 2025년 10월에 발표하는 등 지속적인 노력을 기울이고 있다. 이를 통해 호주는 공공 부문에서의 인공지능 활용이 책임성과 신뢰성을 기반으로 이루어질 수 있도록 제도적 기반을 강화하고 있다.

마. 싱가포르

싱가포르는 책임감 있는 인공지능 활용을 제도적으로 뒷받침하기 위해 검증 중심의 거버넌스 체계를 구축하고 있으며, 이를 통해 기술 혁신과 윤리적 책임의 균형을 도모하고 있다. 이러한 접근의 일환으로 2023년 6월, 정보통신미디어개발청(Infocomm Media Development Authority, IMDA) 산하에 인공지능 검증 재단(AI Verify Foundation)이 출범하였다. 해당 재단은 책임 있는 인공지능 사용을 위한 검증 도구의 개발을 목표로 하며, 2024년에는 인공지능 거버넌스 프레임워크를 발표하고, 이를 강화하기 위한 핵심 요소와 세부 지침을 제시하였다. 이는 기업과 공공기관이 인공지능 시스템의 신뢰성과 책임성을 객관적으로 점검할 수 있도록 지원하는 기반으로 기능하고 있다.

싱가포르의 인공지능 거버넌스 프레임워크는 생성형 인공지능이 야기할 수 있는 윤리

적·기술적 리스크를 체계적으로 관리하는 동시에, 혁신을 저해하지 않는 정책 환경을 조성하는 것을 목표로 한다. 이를 위해 해당 프레임워크는 총 아홉 개의 핵심 차원으로 구성되어 있다. 구체적으로는 책임성, 데이터 관리, 신뢰할 수 있는 개발 및 배포, 사고 보고 체계, 테스트 및 보증, 보안, 콘텐츠 출처 관리, 안전과 연구개발(R&D) 연계 강화, 공공 이익을 위한 인공지능 활용을 포함한다. 이러한 다차원적 구조는 인공지능의 전 주기적 관리와 사회적 영향에 대한 포괄적 대응을 가능하게 한다.

싱가포르는 이와 같은 프레임워크를 기반으로 국제적 인공지능 거버넌스 협력을 선도하고자 하며, 검증 가능한 기준과 도구를 중심으로 글로벌 논의에 기여하는 전략을 취하고 있다. 이는 규범 제시를 넘어 실질적인 실행 가능성과 측정 가능성을 중시하는 싱가포르식 인공지능 거버넌스 모델의 특징을 보여준다.

바. 중국

중국은 인공지능 거버넌스에 있어 중앙집권적 규제 방식을 채택하고 있으며, 국가 차원의 정책 방향과 규범을 통해 인공지능 개발·활용 전반을 관리하는 특징을 보인다. 중국 정부는 인공지능 개발 기업에 대해 보건의료, 금융, 고용 등 사회적 영향이 큰 핵심 분야에서 알고리즘이 편향되거나 불공정한 결과를 초래하지 않도록 할 책임을 명확히 부여하고 있다. 특히 민감 정보를 다루는 인공지능 시스템에 대해서는 엄격한 데이터 보안 기준을 함께 규정함으로써, 개인정보 보호와 국가 안보 차원의 위험을 동시에 관리하고자 하고 있다.

이러한 제도적 환경 속에서 중국의 주요 기술 기업들은 국가 기준에 부합하는 기업 차원의 인공지능 거버넌스 프레임워크를 구축하고 있으며, 다수의 기업이 내부 인공지능 윤리위원회를 설립하여 정책 이행과 기술 개발을 연계하고 있다. 이는 국제 인공지능 기술 경쟁력에서 우위를 유지하려는 중국 정부의 전략적 의지를 반영하는 동시에, 기업이 국가 정책과 정합적인 방향으로 기술 혁신을 추진하도록 유도하는 구조로 평가된다. 대표적인 기업으로는 알리바바, 바이두, 화웨이, 센스타임, 아이플라이텍, 메그비, 텐센트 등이 있다.

중국의 주요 기술 기업들은 정부와의 협력을 통해 인공지능 기술의 사회적 적용을 확대하고 있다. 화웨이는 정부와의 공동 연구 및 프로젝트를 통해 스마트시티 및 공공 안

전 시스템 개발에 참여하며, 인공지능 기술의 공공 영역 활용을 적극적으로 추진하고 있다. 센스타임은 정부 기관과의 공동 프로젝트를 통해 인공지능 기술의 윤리적 적용을 모색하는 한편, 인공지능 윤리 및 거버넌스 위원회를 설립하고 상하이교통대학교와 협력하여 인공지능 윤리 연구센터를 공동 설립하는 등 윤리적 인공지능 개발을 위한 제도적 기반을 구축하고 있다. 아이플라이텍은 정부 및 학계와의 협력을 통해 인공지능 기술의 사회적 가치를 실현하는 프로젝트를 수행하고 있으며, 칭화대학교와 공동으로 의료용 인공지능 로봇 '샤오이(Xiaoyi)'를 개발·활용하는 사례가 대표적이다. 메그비 역시 정부 기관과 협력하여 인공지능 기술의 윤리적 적용을 추진하고, 인공지능 윤리 행동 강령 발표와 함께 인공지능 윤리위원회 및 연구소를 설립하였다. 텐센트는 정부 주도의 인공지능 정책 및 윤리 기준 수립 과정에 협력하며, 국가 차원의 인공지능 거버넌스 체계 구축에 참여하고 있다.

한편 중국에서는 인공지능 윤리 기준과 기술 표준화를 위한 학술적·제도적 연구도 활발히 이루어지고 있다. 칭화대학교, 베이징대학교, 중국과학원(CAS), 중국사회과학원(CASS) 등 주요 대학 및 연구기관은 인공지능 윤리 기준과 기술 표준화 연구, 생성형 인공지능의 사회적 영향 분석, 위험 시나리오 평가 등을 수행하고 있으며, 이를 통해 정부의 정책 수립과 실행을 이론적으로 뒷받침하는 자문기관 역할을 수행하고 있다. 이들 기관은 국가표준화관리위원회(SAC) 등과 협력하여 공정성, 투명성, 책임성에 관한 국가 차원의 기준 마련에도 기여하고 있다.

이와 같은 중앙정부 주도의 정책 방향과 민관 협력 구조를 바탕으로, 중국은 국가 차원의 인공지능 개발 및 혁신 전략과 디지털 대기업의 기술 축적을 결합하여 투자와 연구개발을 지속적으로 확대하고 있다. 그 결과 다수의 인공지능 유니콘 기업이 등장하였으며, 벤처 기업부터 산업 분야 서비스 기업, 대학 및 연구기관에 이르기까지 폭넓은 주체가 시장에 진출하고 있다. 이러한 환경은 교육, 금융, 의료 등 다양한 분야에서 인공지능 응용을 가속화하는 동력으로 작용하고 있다.

사. 일본

일본은 인공지능 정책과 윤리 거버넌스를 국가 비전과 사회 전반의 미래상에 연계하여 설계해 온 국가로 평가된다. 일본 정부는 2019년 3월, '소사이어티 5.0' 실현을 목표

로 「인간 중심의 인공지능 사회원칙」을 발표하고, 인공지능 발전이 인간의 존엄과 사회적 가치에 부합하도록 유도하는 기본 방향을 제시하였다. 해당 원칙은 인간 중심 원칙, 교육·리터러시 원칙, 프라이버시 보호 원칙, 안전 확보 원칙, 공정 경쟁 원칙, 공정성·설명 책임·투명성 원칙, 혁신 원칙의 일곱 가지로 구성되어 있으며, 기술 발전과 사회적 책임 간의 균형을 중시하는 일본의 정책 기조를 반영하고 있다.

일본은 이러한 원칙을 구체화하는 과정에서 윤리적·법적·사회적 함의(Ethical, Legal, and Social Implications, 이하 ELSI) 접근방식을 적극적으로 활용하고 있다. ELSI 프레임워크를 기반으로 정부의 미래지향적 기술 활용 구상에 윤리적 우려를 체계적으로 반영하고자 하며, 인공지능 기술의 사회적 영향과 위험을 사전에 검토하는 정책적 시도를 지속하고 있다. 이는 기술 수용 과정에서 발생할 수 있는 사회적 갈등을 완화하고, 공공의 신뢰를 확보하기 위한 일본 특유의 점진적 접근으로 해석된다.

일본의 인공지능 윤리 거버넌스는 정부 기관과 산하 연구기관을 중심으로 민간 기업, 학계, 시민사회단체 등이 참여하는 다층적 이해관계자 구조를 통해 운영되고 있다. 다만 이러한 구조 속에서도 정책 결정 과정에서 이해관계자 간 실질적인 영향력의 비대칭성이 존재한다는 점이 지속적인 논의 과제로 제기되고 있다. 즉, 참여의 폭은 넓으나 실제 정책 형성 과정에서의 발언권과 영향력은 주체별로 차이가 있다는 점에서 거버넌스의 대표성과 균형성에 대한 고민이 병존하고 있다.

한편 일본은 인공지능 정책을 국가 전략 차원에서 체계적으로 정비해 왔다. 2019년 11월에는 연구개발(R&D), 산업, 거버넌스를 통합하는 「인공지능 전략 2019」를 수립하고, 인간 존엄성, 다양성, 지속가능성을 핵심 가치로 설정한 가운데 네 가지 전략 목표를 제시하였다. 이후 이를 개정·보완하여 대규모 재해 대응, 사회 실증 확대 등 새로운 정책 과제를 반영한 「인공지능 전략 2022」를 2022년에 발표함으로써 전략 범위를 확장하였다. 이는 인공지능을 단순한 산업 기술이 아니라 사회 문제 해결과 국가 회복력 강화의 수단으로 인식하고 있음을 보여준다.

국제 협력 측면에서도 일본은 적극적인 행보를 보이고 있다. 2023년 5월 히로시마에서 개최된 G7 정상회의를 계기로 '히로시마 인공지능 프로세스' 논의를 주도하며, 생성형 인공지능 개발자를 대상으로 한 이행 원칙과 행동 규범에 대한 국제적 합의를 도출하였다. 이어진 G7 디지털 장관 회의를 통해서는 세계 최초로 인공지능에 특화된 포괄

적 국제 규범에 대한 최종 합의안을 마련하는 데 기여하였다. 이러한 움직임은 일본이 인공지능 거버넌스 분야에서 글로벌 규범 형성과 리더십 확보를 주요 정책 목표로 설정하고 있음을 보여준다.

2. 국내 현황

가. 국가 인공지능 전략위원회

국가인공지능위원회는 「인공지능기본법」 제7조에 따라 인공지능 발전과 신뢰 기반 조성을 위한 주요 정책을 심의·의결하기 위해 대통령 소속으로 설치된 범국가적 최고 거버넌스 기구이다.[1] 위원회는 인공지능 정책을 단일 부처 차원이 아닌 국가 전략 차원에서 종합적으로 조정·의결하는 역할을 수행하며, 우리나라 인공지능 거버넌스 체계의 최상위 의사결정 기구로 기능하고 있다.

위원회는 주로 인공지능 정책 전반에 대한 심의·의결을 하는 기능을 수행하고 있다. 구체적으로는 국가 인공지능 기본계획의 수립·변경 및 이행 점검, 인공지능 관련 정책·연구개발·투자 전략 수립, 인공지능 산업 발전과 경쟁력을 저해하는 규제의 발굴 및 개선, 인공지능 데이터센터 등 인프라 확충, 산업 및 공공부문에서의 인공지능 활용 촉진, 국제 규범 마련을 포함한 국제협력, 고영향 인공지능의 규율 및 이에 따른 사회적 변화와 정책 대응 등이 핵심 심의 사항으로 규정되어 있다.

아울러 위원회는 국가기관 및 인공지능 사업자를 대상으로 인공지능의 올바른 사용, 인공지능 윤리의 실천, 인공지능 기술의 안전성과 신뢰성 확보를 위한 권고 또는 의견을 표명할 수 있는 권한을 가져 인공지능 윤리정책의 확산에 있어서도 중요한 역할을 한다. 위원회의 권고나 의견 표명이 이루어진 경우, 해당 국가기관의 장은 법령·제도 개선이나 실천 방안을 마련하도록 함으로써, 위원회의 논의가 실질적인 정책 조치로 연계될 수 있도록 제도적 장치를 두고 있다. 향후 인공지능 윤리정책과 신뢰성 정책이 개별 사업이나 부처 단위가 아닌 범정부적 전략 속에서 추진될 수 있는 제도적 기반을 제공한다는 점에서 중요한 의미를 지닌다.

---

1) 정식 명칭은 대통령 직속 '국가인공지능전략위원회'로 변경되었다.

나. 인공지능 윤리정책 포럼

우리나라는 인공지능 기술의 확산에 따른 잠재적 위험에 투명하고 공정하게 대응하기 위해, 민·관·학이 함께 참여하는 윤리정책 협력 네트워크를 단계적으로 구축해 왔다. 2022년부터 학계, 산업계, 시민사회가 공동으로 참여하여 인공지능 윤리 이슈를 논의할 수 있는 공론의 장과 소통 창구를 마련함으로써, 다양한 이해관계자의 의견을 정책에 반영하기 위한 기반을 조성해 왔다. 이는 학계, 기업, 시민사회, 정책입안자 등 폭넓은 이해관계자로부터 지속적으로 의견을 수렴하고 피드백을 제공하는 EU의 '유럽 인공지능 연합(European AI Alliance)' 운영 사례와 유사한 방향성을 지닌다.

이러한 협력 구조의 핵심 축으로서, 우리나라에서는 국가 인공지능 윤리정책에 대한 자문과 이해관계자 의견 조정을 목적으로 인공지능 윤리정책 포럼이 단계적으로 구성·운영되어 왔다. 2022년 제1기 인공지능 윤리정책 포럼을 시작으로, 2023년 제2기 인공지능 윤리정책 포럼, 2024년에는 제3기 인공지능 윤리·신뢰성 포럼이 구성되어 활동하였다. 이들 포럼은 산·관·학·연 전문가들이 참여하여 인공지능 윤리와 관련된 주요 정책 방향을 논의하고, 사회적 신뢰 형성을 위한 자문 기능을 수행해 왔다.

아울러 2023년 5월 11일에는 '인공지능 윤리·신뢰성 강화를 위한 현장 간담회'를 개최하여, 인공지능 윤리·신뢰성 관련 정책 및 사업 추진 현황을 공유하고 기업의 자율적 노력 사례를 논의하는 한편, 윤리와 신뢰성 확산을 위한 실질적인 방안을 모색하였다. 해당 간담회에는 과학기술정보통신부(제2차관, 인공지능기반정책관 등)를 비롯하여 네이버, LG AI연구원, 제네시스랩 등 산업계, 학계, 그리고 정보통신정책연구원(KISDI), 한국정보통신기술협회(TTA) 등 유관기관이 참여하였다.

제1기와 제2기 인공지능 윤리정책 포럼은 관련 정책에 대한 자문 제공, 인공지능 윤리 이슈에 대한 사회적 토론 주도, 이해관계자 간 의견 조정 등의 기능을 수행하며 국가 인공지능 윤리정책의 기본 방향을 제시하는 데 기여하였다. 특히 인공지능 윤리의 세부 쟁점에 보다 체계적이고 효율적으로 대응하기 위해 윤리분과, 기술분과, 교육분과 등 세 개의 전문 분과위원회를 운영하였으며, 약 30명 내외의 위원을 중심으로 학계, 교육계, 산업계, 시민단체, 법조계, 공공 부문 등 다양한 분야의 전문가를 초빙하여 논의를 진행하였다. 이를 통해 인공지능 윤리 확보를 위한 주요 정책 과제에 대해 각계의 의견을 공식적으로 수렴하고, 정책 추진 과정의 투명성을 제고하는 창구로 기능하였다.

〔그림 2-1〕 제1기·2기 인공지능 윤리정책 포럼

	
제1기 인공지능 윤리정책 포럼	제2기 인공지능 윤리정책 포럼

한편 2024년에 운영된 제3기 인공지능 윤리·신뢰성 포럼은 인공지능에 대한 신뢰 기반을 형성하기 위한 사회적 논의의 구심점 역할을 수행하였다. 이 포럼은 인공지능 윤리·신뢰성 분야의 범부처 과제 발굴과 정책 과제 구체화를 지원하는 데 중점을 두었으며, 인공지능 윤리에 관한 다양한 관점을 포괄하고 심도 있는 논의를 통해 실질적인 해결책을 모색하기 위해 별도의 분과를 두지 않고 통합 운영 방식으로 구성되었다. 또한 생성형 인공지능 기술 발전과 국제 인공지능 논의 동향에 대응하여, 새로운 윤리 쟁점에 대한 의견 수렴과 정책적 대응 방안을 중심으로 논의를 전개하였다.

〔그림 2-2〕 제3기 인공지능 윤리·신뢰성 포럼

	
제3기 인공지능 윤리·신뢰성 포럼	인공지능 윤리·신뢰성 현장간담회

이와 함께 2024년 9월에는 대통령을 위원장으로 하는 '국가인공지능위원회'가 발족되었다. 해당 위원회는 인공지능 분야 전문가로 구성된 30명의 민간위원과 주요 부처 장관급 정부위원 10명, 대통령실 과학기술수석 등으로 구성되어, 범국가적 차원의 인공지능 정책을 심의·의결하는 민·관 협력 기구로 기능하고 있다. 이는 우리나라 인공지능 거버넌스가 개별 정책 논의를 넘어 국가 전략 차원의 통합적 의사결정 체계로 확장되고 있음을 보여주는 중요한 제도적 진전으로 평가된다.

다. 인공지능 안전연구소 및 'AI 안전 컨소시엄'

2024년 인공지능 안전연구소 설립을 계기로 'AI 안전 컨소시엄'이 구성되어, 우리나라 인공지능 안전 분야의 산업계·학계·연구기관 역량을 결집하는 협력 체계가 마련되었다. 해당 컨소시엄은 총 24개의 전문 기업 및 기관이 참여하여,[2] 체계적인 공동 연구와 협력을 통해 국가 차원의 인공지능 안전 체계를 정립하고 글로벌 대응 역량을 강화하는 것을 목표로 하고 있다.

'AI 안전 컨소시엄'은 국내 기업과 기관을 대상으로 연중 상시 참여 접수를 진행하며, 참여 규모를 지속적으로 확대해 나갈 예정이다. 또한 과학기술정보통신부와 국가인공지능위원회를 비롯한 국내 협의체와 긴밀한 협력 체계를 유지하는 한편, 국제 인공지능 안전연구소 네트워크(AISI Network)와의 연계를 통해 글로벌 논의와 협력에도 적극적으로 참여하는 기능을 수행할 것으로 기대된다. 이러한 움직임은 우리나라 인공지능 거버넌스가 자율규제와 안전 중심의 국제 흐름에 대응하여 점진적으로 고도화되고 있음을 보여준다.

라. AI 신뢰성 얼라이언스

우리나라에서는 민간 주도의 인공지능 신뢰성 확보를 지원하기 위한 협력 체계가 본격적으로 구축되고 있다. 우선 과학기술정보통신부와 한국정보통신기술협회(TTA)는 2025년

2) 2025년 10월 기준, LG AI연구원, 카카오, 네이버, 삼성전자, SK텔레콤, KT, 코난테크놀로지, 이스트소프트, 포티투마루, 뤼튼테크놀로지스, 트웰브랩스, 라이너, 크라우드웍스, 고려대 정보보호대학원, 서울대, 성균관대 인공지능융합원, 숭실대, 연세대, KAIST, 소프트웨어정책연구소(SPRi), 정보통신기획평가원(IITP), 정보통신정책연구원(KISDI), 한국지능정보사회진흥원(NIA), 한국정보통신기술협회(TTA) 등 기업 13개, 대학 6개, 연구기관 5개가 참여하고 있다.

발족한 'AI 신뢰성 얼라이언스'를 통해 국내 기업과 전문가들이 참여하는 민간 협의체를 구성하고, 인공지능 신뢰성 인증 제도의 마련과 확산을 위한 의견 수렴과 논의를 추진한다. 해당 얼라이언스는 정부가 추진하는 인공지능 자율규제 도입의 일환으로, 알고리즘 투명성 확보, 환각(Hallucination) 예방, 개인정보 및 지식재산권 보호 등 인공지능 활용 과정에서 발생할 수 있는 부작용에 대응하기 위한 구체적인 기술적·정책적 기준을 민간 주도로 도출하는 역할을 수행한다.

이와 관련하여 TTA는 이미 2023년 말 'AI 신뢰성 인증(CAT)' 제도를 시행하여, 포티투마루 등 일부 스타트업을 대상으로 인증을 부여한 바 있다. 이러한 선행 사례는 향후 AI 신뢰성 얼라이언스를 중심으로 보다 폭넓은 기업 참여와 제도 확산이 이루어질 수 있는 기반으로 작용할 것으로 기대된다.

제 2 절 민간 인공지능 윤리 거버넌스[3)]

1. 민간의 자율적 인공지능 윤리 거버넌스

인공지능 기술의 확산과 함께 사회·기술 환경이 빠르게 변화함에 따라, 새로운 윤리적 이슈가 지속적으로 제기되고 있다. 이러한 변화에 대응하여 민간 기업 내부에서 윤리적 쟁점을 자율적으로 논의하고, 이를 구체적인 실천으로 발전시킬 수 있는 책임감 있는 인공지능 활용을 위한 민간의 자율적 노력이 점차 중요해지고 있다. 특히 테크 기업을 중심으로 신뢰할 수 있는 인공지능을 실현하기 위해 다양한 형태의 거버넌스 체계를 자율적으로 운영하고 있으며, 인공지능의 윤리적 활용을 뒷받침하는 기업 문화의 정착에 중점을 두는 경향이 나타나고 있다.

기업들은 조직 내 리더십을 중심으로 윤리의 내재화를 도모하고 있으며, 인공지능 민간 자율거버넌스를 운영하거나 자체적인 인공지능 윤리 원칙을 수립함으로써 윤리적 판단의 전문성과 타당성을 확보하고 있다. 이러한 노력의 일환으로 IBM의 'AI 윤리 위원

3) 본 절의 내용은 연구 과정에서 별도로 작성한 KISDI Premium Report [특집 3호] "포용적 AI 기본사회 구현을 위한 정책 방향: 보편적 접근, 안전한 서비스, 책임 있는 활용"의 <Chapter 3. 안전한 AI 서비스>의 내용을 바탕으로 재작성되었다.

회', 마이크로소프트의 'AETHER 위원회', LG의 'AI 윤리 위원회', 카카오의 '그룹기술윤리 소위원회', KB금융그룹의 '윤리위원회', SK텔레콤의 'AI 추구 가치 자문단', 엔씨소프트의 'ESG 경영위원회', 크래프톤의 'AI 윤리 위원회' 등 다양한 형태의 위원회와 자문기구가 운영되고 있다. 더 나아가 일부 기업은 윤리위원회와 별도로 관련 분야 연구 및 이슈 분석을 위한 워킹그룹이나, 윤리정책의 실행을 담당하는 전담조직을 함께 구성함으로써 다층적인 거버넌스 구조를 구축하고 있다.

본 절에서는 이러한 흐름을 바탕으로 주요 글로벌 및 국내 테크 기업들의 인공지능 윤리 거버넌스 구조를 분석하고, 자율적 인공지능 거버넌스가 실제로 작동하는 수준에 따라 유형화를 시도하였다. 유형 분류는 조직 내 권한 구조, 의사결정 체계, 운영의 지속성, 기술과 윤리 간의 연계 수준, 외부에 대한 투명성 등을 종합적으로 고려하여 이루어졌다.

이러한 분석은 포용적 인공지능 기본사회 구현을 위한 정책 방향, 즉 보편적 접근, 안전한 서비스 제공, 책임 있는 활용이라는 정책 목표와 맞닿아 있다. 본 절의 내용은 2025년을 기준으로 기업의 공개 자료와 연구기관의 조사 자료를 토대로 작성되었으며, 공개되지 않은 기업 내부 거버넌스 정보는 포함하지 않았다. 또한 기업별 실제 운영 방식의 변화 가능성을 고려할 때, 향후 추가적인 검토와 보완이 필요하다.

분석 대상은 공개된 인터넷 기반 자료를 중심으로 실질적인 인공지능 윤리 관련 활동이 확인된 총 14개 기업으로, 마이크로소프트, IBM, 메타, 앤트로픽, 알리바바, 바이두, 네이버, 카카오, LG, KB금융그룹, SK텔레콤, KT, 엔씨소프트, 크래프톤을 포함한다. 이들 기업은 자율적 인공지능 윤리 거버넌스의 작동 양상에 따라 A) 전략적 집행형, B) 기술-정책 연계형, C) 윤리 자문형, D) 선언적 원칙형의 네 가지 유형으로 분류하고, 각 유형에 해당하는 기업을 매칭하였다.[4)]

4) 본 절은 2025년 기준 기업 공개자료 및 연구기관 조사자료를 기반으로 작성하였으며, 공개되지 않은 기업 내부 거버넌스 정보는 포함하지 않았으며, 기업별 실제 운영 변화는 추후 추가 검토가 필요하다.

〈표 2-1〉 민간 인공지능 윤리 자율거버넌스 유형 정의

구분	거버넌스 정의	주요 특징
A. 전략적 집행형	기업 내 최고 의사결정기구 (이사회, CEO 직속 조직 등)가 실질적으로 운영 및 결정하는 구조	전사 전략 반영, 규정 준수 도구 및 실행력 확보
B. 기술-정책 연계형	기술조직과 윤리 정책 조직이 협력적으로 운영되며 일부 의사결정 참여는 있으나, 전략적 권한은 제한적	기술 실행력에 강점, 정책 반영 수준은 케이스별 상이
C. 윤리 자문형	주로 자문 성격의 위원회가 존재하며, 실질 권한은 없으나 정책 방향 제시	최고경영층과 연결성은 있으나 실행 구조는 미미, 교육적 접근 방식으로 보완
D. 선언적 원칙형	윤리 원칙은 존재하나, 이를 실현할 조직·체계·모니터링 수단은 부재	선언적 윤리원칙, 실질 운영 구조는 불명확

자료: 연구진 작성(2025)

가. 전략적 집행형 거버넌스

먼저, '전략적 집행형(A)'은 기업 내 최고 의사결정기구(이사회 또는 CEO 직속 조직 등)가 인공지능 윤리 거버넌스를 실질적으로 운영하고 주요 사항을 결정하는 구조로 정의된다. 이 유형에 해당하는 기업으로는 마이크로소프트, IBM, 앤트로픽, 네이버, SK텔레콤, KT, KB금융그룹 등이 있으며, 이들 기업은 인공지능의 윤리적 활용을 전사적 전략에 반영함으로써 비교적 높은 정책 실행력을 확보하고 있다.

마이크로소프트는 'AETHER' 위원회, 신뢰할 수 있는 인공지능 사무국(Office of Responsible AI, ORA), 공학 분야 책임 AI 전략기구(RAiSE) 등으로 구성된 다층적 거버넌스 구조를 통해 AI 윤리 원칙을 전사 전략에 내재화하고 있다(<표 2-2 참고>). AETHER 위원회는 윤리 이슈 분석을 담당하고, ORA는 규정 준수 기능을 수행하며, RAiSE는 기술 구현을 담당하는 등 각 조직이 유기적으로 연계되어 실행력, 독립성, 지속성을 동시에 확보하고 있다는 점에서 전략적 집행형의 대표적 사례로 볼 수 있다.

〈표 2-2〉 마이크로소프트 사내 인공지능 윤리 거버넌스 구성

위원회	목적 및 구성	역할 및 기능
Aether(AI, Ethics, and Effects in Engineering and Research) Committee	- 책임있는 인공지능(Responsible AI, RAI) 관련 주요분야 전문가, 개발 대표, 주요 부서의 부서장이 추천한 대표 등이 포함된 조직으로 2017년도에 설립 이후 책임 있는 AI(RAI: Responsible AI) 관련 문제, 기술, 프로세스 및 모범사례 등을 제시할 것을 목적으로 함 ※ 상설 워킹 그룹 외 광범위한 커뮤니티와 회사 외부 전문가 활용	- 공정성과 포용성, 인간-AI의 상호작용 및 협업, 투명성, 신뢰성 및 안전, 보안과 개인정보 분야의 이슈 분석, 개발에 중점
Office of Responsible AI(ORA)	- 책임 있는 AI 구현을 위한 전사적 규칙을 설정하고, 여기에 참여하는 팀의 역할과 책임을 정의	- 시스템의 개발·배포 단계에서 민감한 사례를 검토하여 MS의 AI 원칙 준용이 가능하도록 지원하고, 새로운 규범 및 표준 형성을 위해 노력
Responsible AI Strategy in Engineering (RAISE)	- 엔지니어링 그룹 전체에 MS의 책임 있는 AI 규칙 및 프로세스를 구현	- RAI에 대한 MS의 전사적 규칙을 구현할 수 있는 Azure ML에 적용된 도구 및 시스템 집합(1ES)을 개발 - Aether, ORA와 공동으로 협력하여 RAI 요구사항을 업무에 통합할 수 있는 엔지니어링 관행을 식별하고, 개발 및 구현 - RAI 규칙 및 요구사항을 모니터링하고 시행할 수 있도록 규정 준수 도구를 구현하고, 피드백 매커니즘을 운영

자료: 연구진 작성(2025)

마이크로소프트는 2022년 6월「책임 있는 인공지능(Responsible AI, RAI) 표준지침 v2」를 발표하고, 인공지능 기술의 개발과 배포 전 과정에서 윤리적 원칙을 실질적으로 구현하기 위한 내부 표준을 체계화하였다. 해당 표준지침은 선언적 원칙에 머무르지 않고, 제품 개발과 운영 단계에서 직접 적용 가능한 실행 기준을 제시한다는 점에서 기업 주도의 자율적 인공지능 윤리 거버넌스의 대표적인 사례로 평가된다.

본 지침은 연구개발자, 변호사, 디자이너, 엔지니어, 주제별 전문가 등 30인 이상의 핵심 실무 그룹이 공동으로 참여하여 수립되었다. 이를 통해 인공지능 시스템이 내포할 수 있는 고유한 위험을 다각도로 인식하고, 사회적 요구를 반영한 책임 있는 인공지능 제품 개발 요건을 정의하였다. 이러한 다학제적 구성은 기술적 관점에 치우치기 쉬운 인공지능 개발 과정에 법·윤리·디자인 관점을 통합하려는 구조적 시도로 해석된다.

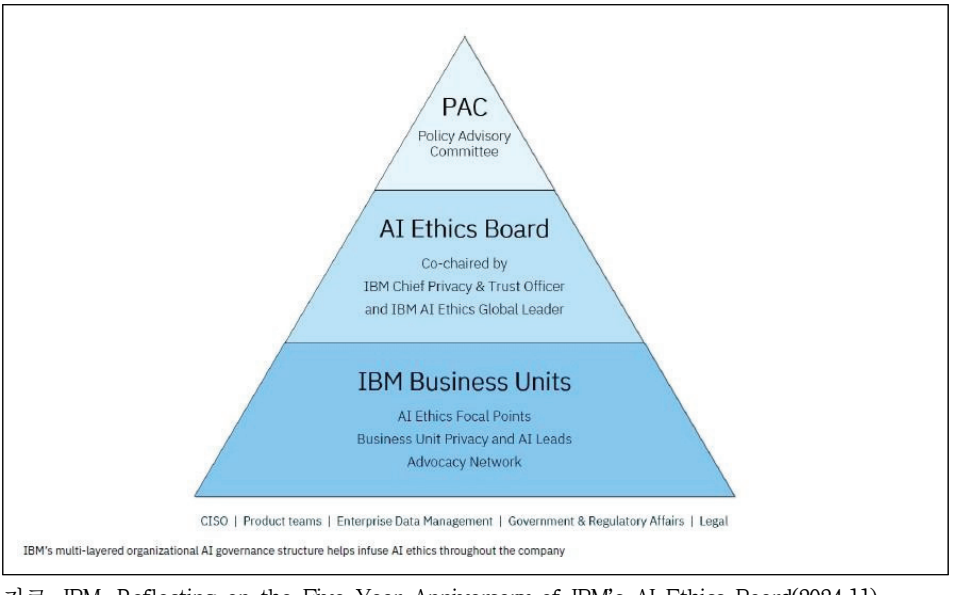
〈표 2-3〉 마이크로소프트 '책임 있는 인공지능(RAI) 지침 V2'('22.6)

구성	내용
책임 목표	(A1) MS의 인공지능 시스템은 영향 평가를 사용 (A2) 사람, 조직, 사회에 중대한 악영향을 미칠 수 있는 시스템을 식별하여 추가 감독 및 요구 사항을 적용 (A3) 문제에 대한 유효한 솔루션을 제공함으로써 목적에 부합 (A4) 적절한 데이터 거버넌스 및 관리 (A5) 사람의 감독 및 제어 지원 기능이 포함
2. 투명성 목표	(T1) 시스템 작동의 의사결정 및 이해에 대한 이해관계자 요구를 지원 (T2) 이해관계자가 정보에 입각한 선택을 할 수 있도록 인공지능 시스템의 성능과 한계에 대한 정보를 제공 (T3) 인공지능 시스템과의 상호작용을 공개하도록 설계
3. 공정성 목표	(F1) 소외계층을 포함하여 식별된 인구통계집단에 유사한 품질의 서비스 제공 (F2) 필수 도메인에서 리소스 또는 기회를 할당하도록 설계됨으로써 집단 간 결과 격차를 최소화하도록 설계 (F3) 사람, 문화 또는 사회를 묘사/표현할 때 고정관념화, 비화, 삭제할 수 있는 가능성을 최소화하도록 설계
4. 신뢰성과 안전 목표	(RS1) MS는 인공지능 시스템이 안정적이고 안전하게 작동할 것으로 예상되는 운영 요소와 범위를 평가·문제 해결·관련 정보를 제공 (RS2) 예상 가능하거나 이미 알려진 장애 해결에 소요되는 시간을 최소화 (RS3) 지속적인 모니터링, 피드백 및 평가
5. 개인정보 보호와 보안 목표	(PS1) MS 개인정보보호표준에 따라 개인정보를 보호하도록 설계 (PS2) MS 보안 정책에 따라 보안을 유지하도록 설계
6. 포용성	(I1) MS 접근성 표준에 따라 포용성을 갖추도록 설계

자료: 연구진 작성(2025)

IBM의 인공지능 윤리위원회(AI Ethics Board) 역시 부사장급 인사가 공동 위원장을 맡아 최고경영진이 직접 관여하는 구조로 운영되며, 윤리기준의 개발뿐 아니라 교육, 기술 검토 등 실행적 기능까지 포괄하고 있다. IBM의 인공지능 윤리 접근방식은 혁신과 책임의 균형을 이루어, 신뢰할 수 있는 인공지능을 대규모로 도입할 수 있도록 지원하는 것을 목표로 신뢰와 투명성의 원칙에 따라 인공지능 윤리위원회를 중심으로 인공지능 신뢰 구축 노력을 다방면으로 진행 중이다. 구체적으로는 내부 인공지능 거버넌스 프로세스를 감독하고, 책임감 있고 안전한 방식으로 기술을 도입할 수 있도록 합리적인 가드레일(기술 사용 사례 검토 및 모범 사례 홍보, 내부 교육 실시, 이해관계자 그룹의 참여 주도 등)을 만드는 역할을 하고 있는 것으로 보인다.

〔그림 2-3〕 IBM 인공지능 거버넌스 구조



자료: IBM, Reflecting on the Five-Year Anniversary of IBM's AI Ethics Board(2024.11)

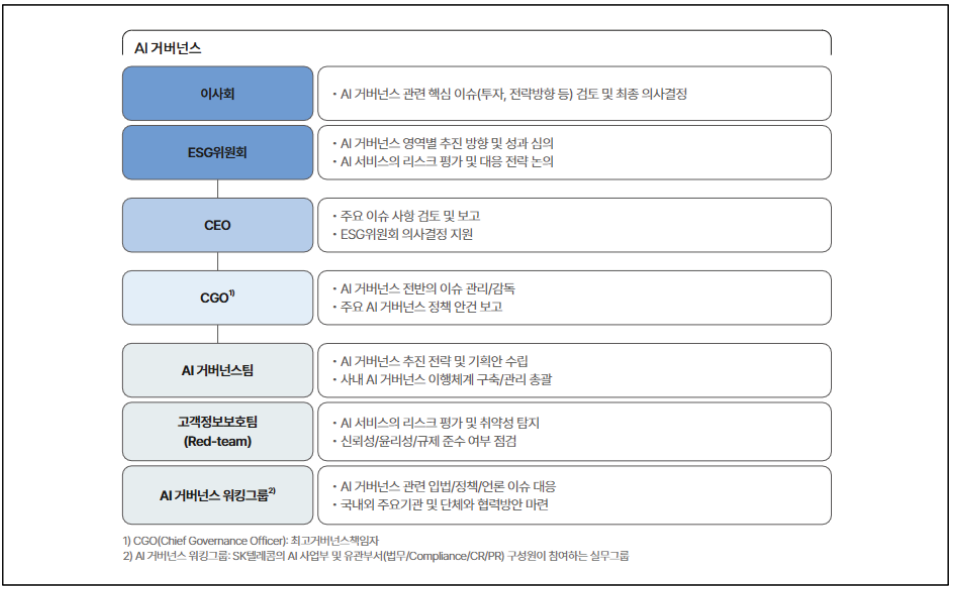
앤트로픽은 다른 기술 기업과 차별화된 법인 구조를 채택하고 있는 사례로, 법적으로 공익 목적을 내재화한 공익기업(Public Benefit Corporation, PBC) 형태로 설립되었으며, 장기혜택신탁(Long-Term Benefit Trust, LTBT)을 운영하고 있다. 델라웨어 주 기업법상 공익법인은 이사가 주주의 재정적 이익과 정관에 명시된 공익 목적, 그리고 기업 활동

으로 실질적 영향을 받는 이해관계자의 이익을 균형 있게 고려할 수 있도록 명시적으로 허용하고 있다. 앤트로픽의 장기혜택신탁은 기업의 재정적 이해관계로부터 독립적으로 설계[5]되었으며, 인공지능 안전, 국가 안보, 공공정책, 사회적 기업 분야의 전문성을 갖춘 5인의 독립적인 수탁자로 구성되어 있다는 점에서 높은 제도적 독립성을 갖는다.

국내 사례로는 네이버가 2024년 최고경영자 직속 국제 인공지능 연구·정책 조직인 '퓨처 AI 센터'를 신설하고, 위험관리위원회 및 위험관리 워킹그룹을 함께 운영함으로써 사내 실무 단계부터 최고 의사결정 단계까지 윤리적 리스크와 안전 이슈를 포괄적으로 다룰 수 있는 체계를 구축하고 있다. SK텔레콤은 'AI 거버넌스 워킹그룹', '고객정보보호팀(Red Team)', 'AI 거버넌스팀'으로 이어지는 체계를 정립하고, 주요 이슈를 최고 의사결정기구인 ESG위원회에 보고하는 구조를 통해 경영진과 실무진 간의 유기적 연결을 강화하고 있다. 특히 최고거버넌스책임자(Chief Governance Officer, CGO)가 인공지능 거버넌스 전반을 관리·감독하며, 주요 정책 안건을 이사회 및 CEO에 보고하도록 하고 있다는 점이 특징이다. KB금융그룹 역시 'AI 윤리위원회'를 인공지능 거버넌스 원칙을 뒷받침하는 최고 의사결정기구로 두고, 고위험 서비스 승인, 윤리기준 및 위험관리 정책 수립 등 인공지능의 윤리적 활용과 관련한 핵심 전략과 의사결정을 수행하고 있다.

5) 신탁은 새로운 주식(Class T)을 신설하였는데, Class T 주식은 신탁에 주요 마일스톤에 따라 점진적으로 앤트로픽의 이사진을 선출하고 사임하게 할 수 있는 권한을 부여하고 있다. 이 주식은 신탁이 회사의 구조나 사업을 중대하게 변경할 수 있는 특정 행위에 대해 통지를 받아야 하는 "보호 조항"도 포함되며, 신탁은 4년 이내로 이사진의 과반수를 새롭게 선출할 수 있다.

〔그림 2-4〕 SKT 인공지능 거버넌스 전담 조직



자료: SK Telecom Annual Report 2024(2025.8)

나. 기술-정책 연계형 거버넌스

‘기술-정책 연계형(B)’은 기술 조직과 정책 조직 간의 협력 및 내부 환류 체계를 갖춘 구조로, 최고 수준의 전략적 권한은 제한적이나 기술 실행력 측면에서 강점을 지니는 유형이다. 이 유형에는 카카오, LG, KT, 크래프톤, 알리바바, 바이두 등이 포함되며, 정책 반영 수준은 기업별로 상이하게 나타난다. 카카오는 ‘그룹기술윤리 소위원회’를 CA협의체 ESG위원회 산하에 두어 기술윤리 정책 수립과 컨트롤 타워 기능을 수행하고 있으며, 실무와 실행력 중심의 거버넌스를 갖추고 있다. 다만 10개 계열사의 기술윤리 총괄 및 실무 책임자로 구성되어 있음에도 불구하고,[6] 실질적인 결정 권한이나 강제력이 낮고 이사회 또는 CEO 직속의 전략 기구와 연계되지 않아 기술-정책 연계형으로 분류하였다. 카카오는 국내 기업으로 최초로 인공지능 ‘알고리즘 윤리 헌장’을 발표(’18)하고, ‘카카오 그룹의 책임 있는 AI를 위한 가이드라인(’23.3)’을 발표하였다.

6) 카카오, 카카오게임즈, 카카오 모빌리티, 카카오뱅크, 카카오스타일, 카카오엔터테인먼트, 카카오엔터프라이즈, 카카오페이, 카카오헬스케어, 디케이테크인 10개 계열사로 구성

〔그림 2-5〕 카카오 알고리즘 윤리 헌장 및 그룹기술윤리 소위원회

그룹기술윤리 소위원회 조직도

카카오 그룹기술윤리 활동 내역

자료: 2024 카카오 그룹 기술윤리 보고서(2024.12)

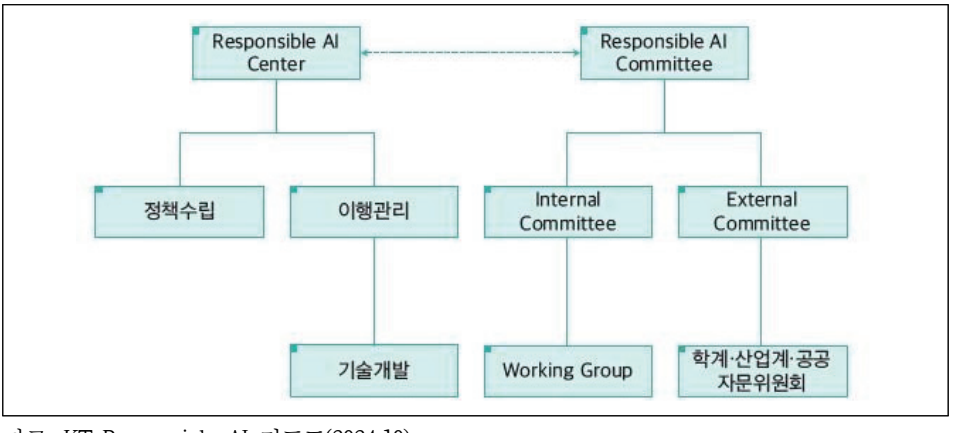
LG는 'AI 윤리위원회', 'AI 윤리 사무국', 'AI 윤리 워킹그룹'을 중심으로 체계적인 조직을 구축·운영하며, 인공지능 기술의 혁신성과 윤리적 책임 간의 균형을 추구하고 있다. 'AI 윤리위원회'는 LG 그룹 'AI 윤리 원칙'을 제정하고 'AI 윤리 원칙'의 이행 현황을 검토하거나 윤리 관련 이슈가 발생하였을 때 의사결정 자문 등을 진행하는 역할을 하는 반면, 'AI 윤리 사무국'은 인공지능 윤리점검 프로세스를 설계하고 사내 구성원에 대한 인공지능 윤리 교육을 진행하는 역할을 담당하는 것으로 알려져 있다. 특히 2024년부터는 인공지능 서비스 기획, 설계, 테스트 전 과정에 걸쳐 개인정보 보호와 인공지능 윤리 관점에서 전문 조직이 단계별 점검을 수행하는 체계를 운영하고 있다.

〔그림 2-6〕 LG 인공지능 거버넌스 조직구성

자료: 2024 LG AI 윤리 책무성 보고서(2025.2)

KT는 2024년 ‘책임감 있는 인공지능 센터(RAIC)’를 신설하여 임직원 인식 제고와 확산에 주력하고 있으며, 학계·산업계·공공 부문이 참여하는 외부 자문위원회를 운영하고 있다. RAIC는 인공지능 윤리원칙을 고도화하고, 책임감 있는 인공지능 거버넌스를 수립하고, 책임감 있는 인공지능 평가 체계의 구축을 담당하고 있다. 다만 주요 안건에 대한 전략적 의사결정 권한, 예컨대 이사회 보고 및 의결 구조가 명확하지 않아 기술-정책 연계형으로 분류하였다. 이 외에도 한국적 문화와 가치를 반영한 인공지능 솔루션 개발을 위해 인공지능 프레임워크 고도화 목적으로 ’24년에 자문위원회를 출범하였으며, 인공지능 모델을 객관적으로 검증하고 국내외 규제를 반영하는 역할을 수행하는 것으로 알려져 있다.

〔그림 2-7〕 KT Responsible AI 추진 거버넌스



자료: KT Responsivle AI 리포트(2024.10)

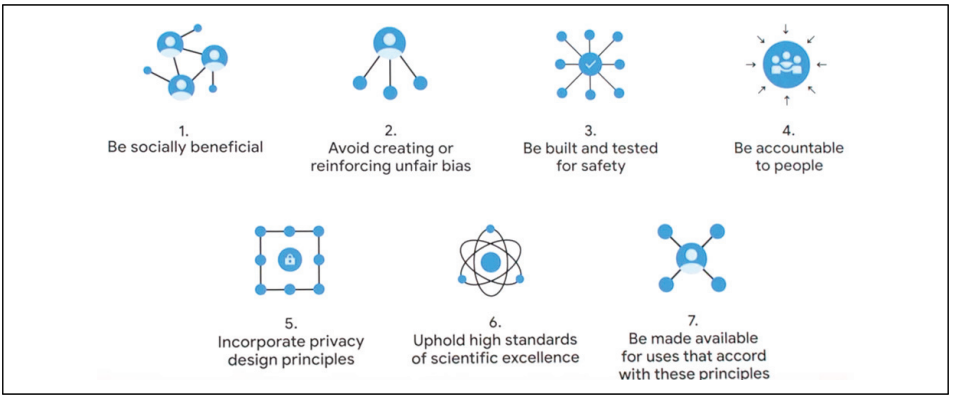
크래프톤은 2023년 ‘AI 윤리위원회’를 출범시켜 지속적인 숙의와 토론을 통해 인공지능 윤리 이슈를 사전에 개선하기 위한 정책 수립 기능을 수행하고 있으나, 법무·데이터·프라이버시 부서 구성원의 자발적 참여에 기반하고 있어 고위 전략 의사결정 구조와의 연계성은 상대적으로 미흡한 것으로 평가된다. 알리바바와 바이두 역시 CTO 산하 기술위원회 또는 기술윤리위원회를 통해 인공지능 윤리를 검토하고 윤리 교육과 안전 메커니즘을 운영하고 있으나, 최고 의사결정기구와의 연결은 명확하지 않은 실무 중심의 운영 구조를 보인다.

다. 윤리 자문형 및 선언적 원칙형 거버넌스

‘윤리 자문형(C)’은 자문 성격의 윤리위원회를 운영하되 실질적인 의사결정 권한이 제한적이며, 주로 정책 방향 제시에 역할이 한정된 기업 유형을 의미한다. 이 유형은 실행 체계가 상대적으로 취약하거나, 과거에는 윤리적 리더십이 두드러졌으나 현재는 조직적 지속가능성이 낮아진 경우에 해당한다. ‘선언적 원칙형(D)’은 인공지능 윤리 원칙과 같은 문서상 기준은 존재하나, 이를 실질적으로 이행할 조직, 체계, 모니터링 수단이 부재한 경우로 구분하였다. 또는 선언적 원칙 및 이행 조직이 한때 자체적으로 굉장히 활발하게 있었다 하더라도 정치적 환경 변화로 인해 기존의 조직이 전면 해체되거나 실질적인 업무를 수행하기 어려울 정도로 조직이 약화되었을 시 ‘선언적 원칙형’으로 분류하였다.

대표적으로 구글과 메타는 인공지능 윤리 원칙을 비교적 선도적으로 수립한 기업으로 평가되나, 이후 주요 윤리 조직이 반복적으로 축소·해산되면서 구조적 실행 체계와 정책 연계성이 약화되었다는 평가가 존재한다. 구글은 2018년 인공지능 윤리 원칙을 수립하고 RESIN과 같은 조직을 운영하며 한때 전략적 집행형으로 분류될 수 있을 정도의 선도적 위치에 있었으나, 외부 자문기구인 ATEAC이 2019년 출범 일주일 만에 해체되고, 주요 윤리 조직이 2024년 이후 축소 또는 해산되는 등 지속가능성 측면에서 한계를 드러냈다. 현재는 ‘책임 있는 인공지능과 인간중심 기술팀(RAI-HCT)’을 중심으로 연구 및 도구 개발에 집중하고 있다. 메타 역시 기술자와 윤리학자로 구성된 ‘책임 있는 인공지능팀(RAI)’을 운영했으나 2023년 해산되었으며, 제3자 팩트체크 폐지와 맞물려 지속적인 인공지능 윤리 거버넌스 구조가 공백 상태에 놓여 있다는 평가를 받고 있다.

〔그림 2-8〕 구글 인공지능 원칙의 7대 목표



자료: Medium(2023.5.13.), "How effective is Google's Bold and Responsible Approach to AI?"

2. 민간부문 주요 인공지능 윤리원칙 현황

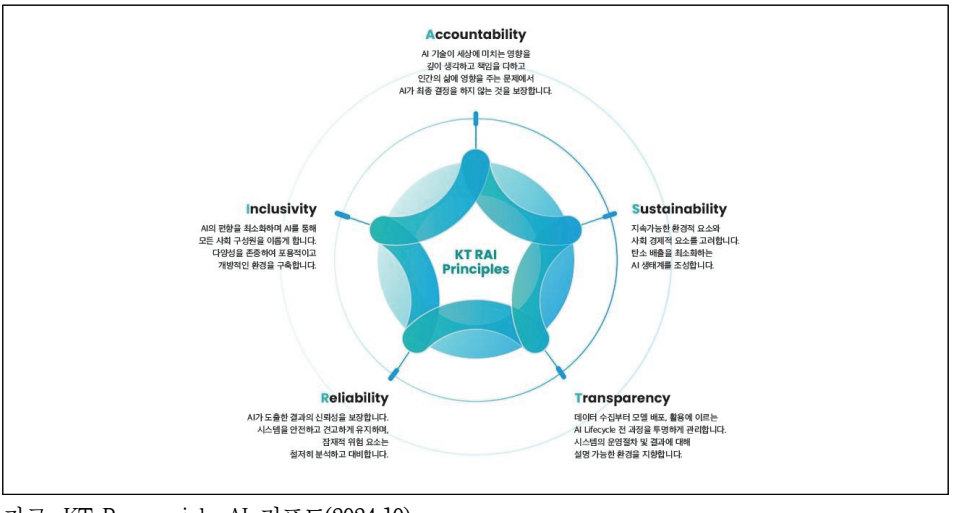
구글, 마이크로소프트, IBM, 알리바바 등 주요 글로벌 기업들은 비교적 이른 시점에 인공지능 윤리원칙을 선언하고, 이를 기술적 프레임워크 구축, 개발자 교육 체계 운영 등 구체적인 실행 도구와 병행하여 추진해 왔다. 이러한 기업들은 인공지능 윤리를 단순한 선언적 가치에 그치지 않고, 실제 개발·운영 과정에 반영하기 위한 다양한 내부 절차와 지원 체계를 마련해 왔다는 점에서 특징을 지닌다. 다만, 기업이 활동하는 국가 및 지역의 정치적·제도적 환경 변화에 따라 민간 자율 인공지능 거버넌스 구조가 비교적 빈번하게 조정되는 경향이 있으며, 이로 인해 윤리 원칙의 일관된 이행과 실행 수단의 지속성에는 일정한 한계가 존재할 수 있다.

한편, 국내 정보통신기술 기업들은 글로벌 기업에 비해 상대적으로 최근에 인공지능 윤리원칙을 수립한 경향을 보이며, 윤리 원칙과 함께 '기업 준칙' 또는 '실무 지침'의 형태로 구체적인 내부 규범을 병행하여 보유하는 경우가 많다. 이러한 특징은 K-디지털 규범, 개인정보보호법, 「인공지능기본법」 등 국내 법·제도와의 정합성을 확보하는 데 중점을 둔 결과로 해석할 수 있다. 즉, 국내 기업의 인공지능 윤리 체계는 글로벌 논의와의 정합성뿐만 아니라 국내 제도 환경에 대한 대응을 주요 고려 요소로 삼고 있다.

〔그림 2-9〕 네이버 인공지능 윤리준칙 5대 원칙

자료: 네이버TV(2022.07.28.), 네이버 AI 윤리 준칙의 방향성(최종 검색일: 2025.04.22.)

〔그림 2-10〕 KT Responsible AI 윤리원칙



자료: KT Responsivle AI 리포트(2024.10)

2020년에 발표된 국가 인공지능 윤리기준을 기준으로 민간 부문의 인공지능 윤리원칙을 살펴보면, 대부분의 기업 윤리원칙은 국가 인공지능 윤리기준이 제시한 핵심 가치를 전반적으로 포괄하고 있는 것으로 나타난다. 특히 다양성 존중과 투명성 원칙의 경우, 주요 기업들이 동일하거나 유사한 용어를 사용하여 공통적으로 강조하고 있으며, 이는 민간 부문 전반에서 해당 가치에 대한 공감대가 형성되어 있음을 시사한다. 또한 안전

성 원칙 역시 대부분의 기업이 중요하게 인식하고 있는 핵심 가치로 확인되며, 인공지능 기술의 책임 있는 활용을 위한 기본 전제로 자리 잡고 있다.

〈표 2-4〉 민간 인공지능 윤리 원칙 및 표준지침(국외 기업)

기업명	윤리원칙 및 표준지침	주요 구성 내용	실행력 확보 수단
Google	7대 인공지능 윤리 원칙(2018)	사회적 유익, 편향 최소화, 안전성, 책임성, 프라이버시, 과학적 우수성, 원칙 부합성	미미(실행 구조 해체)
Microsoft	RAI 표준지침 v2(2022)	책임성, 공정성, 안전성, 포용성, 투명성, 프라이버시	인공지능 개발자 대상 가이드라인, 교육 플랫폼
IBM	신뢰와 투명성 원칙(2018)	설명가능성, 공정성, 견고성, 투명성, 프라이버시	개발자 교육, 윤리 5대 고려항목 제시
Meta	책임감 있는 인공지능의 5대 원칙	개인정보 보호와 보안, 공정성과 포용성, 견고성과 안전성, 투명성과 조절, 책임과 거버넌스	미미(실행 구조 해체)
Alibaba	6대 인공지능 기술 윤리 원칙	인간 중심성, 포용성, 안전성 및 신뢰성, 프라이버시 보호, 책무성, 개방성 및 협력	기술윤리위원회 운영, 각 사업부 지정 책임자, 설명가능한 인공지능 및 프라이버시 보호 기술 투자
Baidu	인공지능 윤리 원칙(2023) 확인	안전성, 통제 가능성, 평등한 성장, 인류의 자유와 가능성 창출	기술윤리위원회 운영, 교육, 안전 메커니즘 구축

자료: 연구진 작성(2025)

〈표 2-5〉 민간 인공지능 윤리 원칙 및 표준지침(국내 기업)

기업명	윤리원칙 및 표준지침	주요 구성 내용	실행력 확보 수단
네이버	AI 윤리 준칙(2021)	전문 + 5개 조항	ASF 관리체계 구축 중
카카오	책임 있는 AI 가이드라인(2023 개정)	10개 조항(구 알고리즘 윤리 헌장 포함)	전사 온라인 교육('21)
LG	LG AI 윤리원칙(2022)	5대 핵심 가치 기반	각 단계별 윤리점검 프로세스, 인공지능 윤리 인식조사 및 교육
KB 금융	AI 윤리기준(2022)	설계-운영-관리 단계별 7대 가치	사내 의견수렴 기반 수립
SKT	AI 추구 가치(2021), T.H.E. AI 원칙(2024)	Telco, Humanity, Ethics 중심 가치	외부 자문단 및 교육 자료
KT	Responsible AI 윤리원칙(2024)	기술·윤리 융합된 5대 원칙	RAIC 중심 실행, 자문위 운영
엔씨소프트	NC AI 윤리 프레임워크	기술 개발·운영 반영형 프레임워크	ESG 위원회, 정책 내 반영 노력
크래프톤	AI 윤리원칙(2024.7)	-	윤리위원회 주관, 논의 지속
NHN	AI 윤리원칙	-	선언 수준 원칙, 세부지침 미공개

자료: 연구진 작성(2025)

3. 소결

본 절에서는 민간 기업의 인공지능 자율 거버넌스를 '전략적 집행형(A)', '기술-정책 연계형(B)', '윤리 자문형(C)', '선언적 원칙형(D)'의 네 가지 유형으로 분류하고, 각 유형 간 차이를 비교·분석하였다. 분석 결과, 조직의 권한 수준과 실행 범위의 깊이가 민간 자율 거버넌스의 실질적 차이를 만들어내는 핵심 요인으로 확인되었다. 단순히 조직 구조를 갖추는 것만으로는 충분하지 않으며, 거버넌스가 실제로 작동하기 위해서는 실행 프로세스의 제도화가 병행되어야 한다. 구체적으로는 내부 체크리스트의 활용, 위험평가 기준의 마련, 자체 교육 시스템 구축, 이슈 발생 시 대응 프로토콜 등의 요소가 갖추어질 때 비로소 자율 거버넌스가 실효성을 가질 수 있다.

한편, 민간 기업별 인공지능 윤리 원칙은 대부분 기업이 자발적으로 개발하고 있으며, 각 기업의 경영 철학과 핵심 가치에 부합하는 요소를 중심으로 선택·구성하는 경향이 나타난다. 그러나 인공지능 윤리 원칙이 사내 조직 운영, 내부 규범, 서비스 운영 방식, 기술 설계 과정 등과 유기적으로 연결되지 않을 경우, 윤리 원칙은 선언적 수준에 머무를 가능성이 크다.

〔그림 2-11〕 민간자율윤리위원회 수립 및 윤리원칙 운영 기업

자료: 연구진 작성('25.3 기준)

기업의 사업 구조와 기술 특성에 따라 윤리 원칙의 내재화 방식에도 차이가 나타난다. 기술 개발 중심의 기업인 마이크로소프트, IBM, LG, 바이두 등은 인공지능 윤리 원칙을 모델 개발 과정과 직접 통합하려는 경향을 보이며, 기술 설계 단계에서부터 윤리적 고려를 반영하는 데 중점을 두고 있다. 반면 구글, 메타, 네이버, 카카오와 같은 플랫폼 기반 기업은 알고리즘 책임성, 이용자 보호, 콘텐츠 관리 등 서비스 운영 과정에서 발생하는 윤리적 이슈에 상대적으로 더 큰 비중을 두는 방식을 선호한다. 금융 및 통신 기반 기업인 KB금융그룹, SK텔레콤, KT 등은 규제 환경과의 연계를 고려하여 위험 기반 관리 체계를 구조화하는 방향으로 인공지능 윤리 원칙을 내재화하는 경향이 뚜렷하게 나타난다.

인공지능 기업이 안전하고 신뢰할 수 있는 인공지능의 개발과 활용을 실현하기 위해서는 윤리 원칙을 선언하는 수준을 넘어, 이를 조직 운영과 기술 개발 과정에 제도적으

로 정착시킬 수 있는 실행 프로세스의 구축이 필요하다. 현재 민간 부문에서 제도화가 가능한 실행 활동은 여러 영역에 걸쳐 제시되고 있으나, 기업의 역량과 환경을 고려하여 단계별 우선순위를 정립할 필요성이 제기된다.

우선 거버넌스 기반 조성 측면에서는 사내 인공지능 윤리 담당 위원회의 설치 및 운영이 핵심적인 출발점이 된다. 이를 토대로 내부 인공지능 윤리 원칙과 세부 가이드라인을 수립하고, 인공지능 개발·활용 과정에 참여하는 협력 업체에 대해서도 윤리 준수를 요구하는 체계를 마련할 필요가 있다. 아울러 인공지능 시스템의 설계와 운영 전반에 다양성과 포용성을 반영한 운영 원칙을 수립함으로써, 편향과 차별의 위험을 사전에 완화할 수 있는 기반을 구축해야 한다.

다음으로 위험관리 및 책임성 확보를 위해서는 기술적·관리적 통제가 병행되어야 한다. 인공지능 모델의 안전성을 사전에 검증할 수 있는 체계를 구축하고, 인공지능 의사결정 과정에 대한 검증 시스템을 마련함으로써 오류나 위험을 체계적으로 관리할 필요가 있다. 또한 인공지능 제품과 서비스가 사회 및 이용자에게 미치는 영향을 평가하는 절차를 도입하고, 정기적인 자율 점검과 내부 감사 시스템을 운영함으로써 지속적인 개선을 도모해야 한다. 이와 함께 인공지능으로 인해 발생할 수 있는 이용자 피해에 대응하기 위한 구제 절차와 고객 지원 체계를 갖추는 것이 책임성 확보의 중요한 요소로 작용한다.

마지막으로 투명성 확보와 조직 및 사회 전반의 인식 제고를 위해서는 정보 공개와 교육 활동이 병행되어야 한다. 기업은 인공지능 관련 투명성 보고서를 발간·공개함으로써 개발 및 운영 현황에 대한 사회적 신뢰를 제고할 수 있으며, 사내 직원을 대상으로 한 인공지능 윤리 교육을 통해 윤리 원칙의 실질적 내재화를 촉진할 수 있다. 더 나아가 시민이 참여하는 인공지능 윤리 모니터링 기구를 운영함으로써, 기업의 자율적 윤리 실천을 외부의 시선에서 점검하고 사회적 소통을 강화하는 방안도 고려할 수 있다.

〈표 2-6〉 국가 인공지능 윤리기준과 민간 주요 윤리원칙 간 비교

구분	원칙	10대 핵심요건										기타
		인권 보장	프라이버시 보호	다양성 존중	침해 금지	공공성	연대성	데이터 관리	책임성	안전성	투명성	
대한민국	**인공지능 윤리기준**	**인권 보장**	**프라이버시 보호**	**다양성 존중**	**침해 금지**	**공공성**	**연대성**	**데이터 관리**	**책임성**	**안전성**	**투명성**	
IBM	Foundational Pillars of Trustworthy AI		프라이버시	공정성						견고성	투명성, 설명 가능성	
Google		인간 중심성	프라이버시	공정성					책임성	신뢰성, 안전성	투명성, 설명 가능성	견고성
Microsoft	Responsible AI Principles		프라이버시 및 보안	포용성, 공정성					책무성	신뢰성·안전성	투명성	
Meta			개인정보 보호와 보안	공정성, 포용성					책임성	안전성	투명성	견고성, 통제성
LG	LG AI 윤리원칙	인간 존중		공정성					책임성	안전성	투명성	
카카오	알고리즘 윤리헌장	사회 윤리 준수 및 인류 편익 행복 추구	프라이버시 보호	차별 경계		아동·청소년 보호	기술 포용성	학습 데이터 운영			알고리즘에 대한 설명	알고리즘 독립성
네이버	AI 윤리 준칙	사람을 위한 AI 개발	프라이버시 보호와 정보 보안	다양성의 존중				프라이버시 보호와 정보 보안		안전을 고려한 서비스 설계	합리적인 설명과 편리성의 조화	
KB금융	AI 윤리기준	공정과 포용	데이터 관리	공정과 포용		디지털 역량	참여와 협력	데이터 관리	안전과 책임	안전과 책임, 통제 가능성	투명한 활용	통제 가능성
NH농협은행	AI 윤리원칙			다양한 가치를 존중하는 AI				개인정보 보호와 보안을 중시하는 AI	책임지는 안전한 AI	책임지는 안전한 AI	투명하고 신뢰할 수 있는 AI	농업인과 고객의 행복을 위한 AI

구분	원칙	10대 핵심요건										기타
		인권 보장	프라이버시 보호	다양성 존중	침해 금지	공공성	연대성	데이터 관리	책임성	안전성	투명성	
SK텔레콤	AI 추구가치		사생활 보호	다양한 의견 포용	무해성	사회적 가치				기술 안정성	투명성	지속 혁신
KT	Responsible AI 윤리원칙			포용성					책임성	신뢰성	투명성	지속 가능성
NC 소프트	AI 윤리 핵심가치		데이터 보호 중시	비편향성				데이터 보호 중시			투명성 추구	
크래프톤	AI 윤리 원칙	사람을 위한 AI	개인정보 보호	다양성에 대한 존중				개인정보 보호			신뢰를 위한 투명성 확보	

자료: 연구진 작성('25.3 기준)

제 3 장 민간자율인공지능윤리위원회 표준지침(안)

제 1 절 표준지침 마련의 근거

1. 인공지능 윤리원칙의 확산 및 실천 지원

「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하, 「인공지능기본법」) 제28조에 따른 민간자율인공지능윤리위원회(이하, 민간자율위원회)의 설치 목적은 국가가 일방적으로 주도하는 하향식 규제의 한계를 극복하고, 민간이 주체적으로 참여하는 상향식 윤리 거버넌스를 구축하는 데 있다. 동조 제1항은 위원회 설치를 강제하지 않고 임의규정으로 두어, 각 기관이나 단체가 처한 상황과 필요에 따라 자율적으로 설치 여부를 결정할 수 있게 했다. 이는 획일적인 규제 적용 대신 각 기업의 특성에 부합하는 맞춤형 윤리 실천을 가능하게 하여 자발적인 참여를 이끌어내기 위함이다.

이러한 자율적 접근 방식은 대기업뿐만 아니라 중소기업, 스타트업 등 다양한 민간 영역의 주체들이 부담 없이 민간자율위원회를 도입할 수 있는 환경을 조성한다. 이는 결과적으로 특정 선도 기업에 국한되지 않고 산업 생태계 전반으로 인공지능 윤리의식이 확산되는 효과를 가져오며, 정부의 정책적 노력만으로는 달성하기 어려운 사회 전반의 자생적인 윤리 문화를 조성하는 중요한 기제가 된다.

동조 제1항이 “윤리원칙을 준수하기 위하여”라고 명시한 것은 민간자율위원회의 목적이 단순히 추상적인 윤리 담론을 논의하는 데 있지 않음을 시사한다. 민간자율위원회는 인공지능의 연구·개발·활용 각 단계에서 윤리원칙이 지켜지고 있는지 확인하고 점검하는 실천 중심의 역할을 한다. 즉, 민간자율위원회는 선언적인 윤리원칙(법 제27조)이 실제 기술 개발 과정과 서비스 운영 절차에 체계적으로 반영되고 내재화될 수 있도록 하는 실질적이고 제도적인 토대로서 기능한다.

한편, 「인공지능기본법」 제27조 제1항 제2호는 “인공지능기술 연구 및 개발을 수행하는 사람이 소속된 교육기관·연구기관”도 민간자율위원회를 둘 수 있는 주체로 규정하고 있다. 이때 ‘교육기관·연구기관’이란 인공지능기술 연구 및 개발을 전문적인 업(業)으로

수행하는 자가 해당 업무를 수행하기 위해 소속된 기관을 의미한다. 이러한 기관에는 「고등교육법」 상의 교육기관을 비롯하여 「정부출연연구기관 등의 설립·운영 및 육성에 관한 법률」에 따라 설립된 연구기관, 「특정연구기관 육성법」 상의 연구기관이 포함된다.

2. 민간자율위원회의 구성·운영상 자율성

「인공지능기본법」 제28조 제3항 본문은 민간자율위원회의 구성과 운영에 필요한 사항을 해당 기관이나 단체(동조 제1항 각 호)가 자율적으로 정하도록 규정하고 있다. 이는 국가가 획일적인 기준을 강요하는 대신, 기업이나 연구기관이 각자의 규모, 업무 특성, 조직 문화에 최적화된 형태로 위원회 모델을 설계할 수 있도록 광범위한 재량권을 부여한 것이다. 이에 따라 각 주체는 위원회의 구체적인 구성 방식, 회의 개최 주기, 의사결정 절차 등을 스스로 정하여 내부 규정으로 수립할 수 있다. 이러한 자율성은 경직된 규제의 한계를 극복하고, 각 현장에 가장 적합한 실효적인 윤리 실천 체계를 구축하도록 유도하는 강력한 동기로 작용할 수 있다.

그러나 이러한 자율성이 무제한 허용되는 것은 아니며, 위원회가 사회적 책임을 다할 수 있도록 최소한의 공익적 요구사항이 법적으로 강제된다. 「인공지능기본법」 제28조 제3항 단서는 민간자율위원회의 구성에 있어 특정 성(性)으로만 구성하지 않도록 규정하여 성별 다양성을 확보하고, 사회적·윤리적 타당성을 평가할 수 있는 경험과 지식을 갖춘 외부 전문가와 해당 기관에 종사하지 않는 사람을 각각 위원으로 포함하도록 하여 조직 내부의 논리에 치우치지 않은 객관성을 확보하고 있다. 이는 민간의 자율적인 운영을 최대한 보장하되, 전문성과 독립성이라는 공익적 가치를 훼손하지 않기 위한 조화로운 법적 장치로 이해할 수 있다.

3. 전 생애주기에 걸친 인공지능 윤리원칙의 내재화

민간자율위원회는 기업이 대외적으로 표방하는 윤리원칙이 선언적 구호에 그치지 않고, 조직 내부에 체계적으로 스며들어 내재화될 수 있도록 하는 구체적인 수단을 제공한다. 민간자율위원회는 「인공지능기본법」 제28조 제2항 각 호에 규정된 업무를 자율적으로 수행하는데, 여기에는 실제 사례에 대한 검토(제2호), 연구자 및 종사자에 대한 윤

리 교육(제4호), 기관이나 단체의 특성에 맞는 맞춤형 인공지능 윤리 지침의 수립(제5호) 등이 포함된다. 이러한 활동들은 추상적인 윤리 개념을 현장 실무자들이 적용할 수 있는 형태의 구체적인 행동 규범으로 전환하는 역할을 한다.

민간자율위원회의 업무 범위는 인공지능 기술의 연구-개발-활용에 이르는 전 생애주기를 포괄한다. 법 제28조 제2항은 윤리원칙 준수 여부의 확인뿐만 아니라, 안전 및 인권침해 가능성에 대한 조사·연구(제2호), 절차와 결과에 대한 감독(제3호)까지 위원회의 업무에 포함하고 있다. 이는 법률이 인공지능 개발 단계에서의 기술적 검증뿐만 아니라, 서비스가 출시된 이후 사회적으로 미칠 영향에도 관심을 기울이고 있음을 보여준다. 위원회는 인공지능 시스템의 성능 지표를 넘어, 해당 기술이 이용자(법 제2조 제8호) 또는 영향받는 자(법 제2조 제9호)의 기본권을 침해하거나 차별을 유발할 가능성은 없는지 등 사회적 영향과 인권 보호의 관점에서 기술을 평가한다. 이러한 포괄적인 접근법은 기업이 이윤 추구 과정에서 놓칠 수 있는 사회적 가치를 재확인하고, 안전하고 신뢰할 수 있는 인공지능 서비스를 제공하도록 유도하는 핵심적인 기제에 해당한다.

4. 법 제28조 제4항에 따른 표준지침 수립 및 보급

민간자율위원회가 실질적인 기능을 수행하기 위해서는 이를 뒷받침할 제도적 기준이 필수적이다. 이에 「인공지능기본법」 제28조 제4항은 “과학기술정보통신부장관은 민간자율위원회의 공정하고 중립적인 구성·운영을 위하여 표준지침 등을 마련하여 보급할 수 있다”라고 명시하여 표준지침 마련의 직접적인 법적 근거를 두고 있다. 이는 정부가 민간의 자율적인 윤리 활동을 존중하되, 그 운영이 방임되거나 형식적인 수준에 그치지 않도록 표준 모델을 제시할 수 있는 권한을 법률로써 부여한 것이다.

다른 한편, 표준지침의 마련은 민간 기관이 윤리위원회를 처음 도입할 때 겪을 수 있는 시행착오를 줄여주는 지원책으로서의 성격도 가진다. 법률은 위원회의 운영에 필요한 사항을 원칙적으로 해당 기관이 자율적으로 정하도록 하고 있으나, 구체적인 운영 노하우가 부족한 중소기업이나 스타트업에게는 이러한 완전한 자율이 오히려 부담으로 작용할 수 있다. 따라서 정부가 마련하는 표준지침은 회의 운영 방식, 안건 심의 절차, 이해상충 방지 규정 등 모범적인 운영 가이드라인을 제시함으로써, 민간 기업이 더 쉽

고 효율적으로 윤리 거버넌스를 구축하고 실효성 있는 윤리 통제 활동을 수행할 수 있도록 돕는 법적·제도적 기반이 된다.

제 2 절 민간자율위원회 표준지침(안) 도출 과정

1. 유사 윤리위원회 사례

가. 기관생명윤리위원회

기관생명윤리위원회(IRB)는 연구의 윤리적 수행과 안전을 확보하기 위해 「생명윤리 및 안전에 관한 법률」(이하, 「생명윤리법」)에 근거하여 설치되며, 엄격한 구성 및 운영 체계가 특징이다. 물론 기관생명윤리위원회와 「인공지능기본법」 제28조에 따른 민간자율위원회는 법적 성격이 매우 다르므로 참고에 주의가 필요하다. 그럼에도 기관생명윤리위원회의 조직 및 운영에 관한 사항은 민간자율위원회가 실질적인 공정성과 독립성을 확보하기 위해 참고할 수 있는 표준 모델이라는 점은 부인하기 어렵다.

기관생명윤리위원회는 기관장 직속으로 설치될 수 있으나, 그 운영과 심의에 있어서는 「생명윤리법」 제11조 제5항에 따라 독립성을 보장받아야 한다(보건복지부, 2022). 위원회는 위원장 1인을 포함하여 5인 이상의 위원으로 구성하되, 특정 성(性)으로만 구성할 수 없다(「생명윤리법」 제11조 제1항). 위원회의 공정성과 다양성을 확보하기 위해, 사회적·윤리적 타당성을 평가할 수 있는 비과학계 위원 1인 이상과 해당 기관에 종사하지 않으며 이해관계가 없는 외부 위원 1인 이상을 반드시 포함하여 구성하여야 한다(「생명윤리법」 제11조 제1항; 보건복지부, 2022).

심의의 공정성을 담보하기 위하여 위원회는 위원, 연구자, 심의 안건 간의 이해상충을 관리하는 절차를 명문화하여 운영하여야 하며, 모든 위원은 임기 내 심의하는 모든 안건에 대하여 이해상충 여부를 사전에 공개할 의무를 진다(보건복지부, 2022, 18면 이하). 구체적으로 심의 대상 연구에 참여하거나 관여하는 위원은 해당 안건의 심의 및 의결에서 법적으로 배제되는 제척대상이 된다(「생명윤리법」 제11조 제3항). 또한, 연구자는 공정한 심의를 기대하기 어려운 위원에 대해 기피를 신청할 수 있으며, 위원 스스로 심의의 공정성을 해칠 우려가 있다고 판단할 경우 회피할 수 있는 절차를 두어 심의 과정의

투명성을 확보하고 있다.

기관생명윤리위원회는 연구 계획의 승인에 그치지 않고 연구의 시작부터 종료까지 전 과정을 관리·감독하여야 한다. 위원회는 해당 기관에서 수행 중인 연구의 진행 과정 및 결과에 대해 조사하고 감독할 권한을 가지며, 필요한 경우 무작위 표본조사 등을 실시할 수 있다(보건복지부, 2022, 5면). 특히 연구 수행 중 생명윤리 또는 안전에 중대한 위해가 발생하거나 발생할 우려가 있는 경우, 기관장은 지체 없이 위원회를 소집하여 심의하고 그 결과를 보건복지부장관에게 보고하여야 한다(「생명윤리법」 제11조 제4항). 연구계획서, 심의 결과, 서면동의서 등 연구 관련 기록은 연구가 종료된 시점부터 3년간 보관하여야 하며, 보관 기간이 경과한 개인정보 관련 문서는 파기하는 것을 원칙으로 한다(「생명윤리법」 제19조; 보건복지부, 2022, 70면).

나. 동물실험윤리위원회

동물실험윤리위원회 역시 민간자율위원회의 구성 및 운영에 관한 표준지침을 구상함에 있어 참고할 수 있는 모범 사례로서, 위원회의 독립성 확보와 사후 관리 체계 구축에 관한 중요한 시사점을 제공한다.

동물실험윤리위원회는 조직 내부의 이익 대변을 방지하고 윤리적 심의의 객관성을 확보하기 위하여 법령 차원에서 구성의 다양성을 엄격히 강제하고 있다. 「동물보호법」 제53조 제1항에 따르면, 위원회는 위원장 1인을 포함하여 3인 이상 15인 이하의 위원으로 구성하되, 해당 동물실험시행기관과 이해관계가 없는 사람이 전체 위원 수의 3분의 1 이상 포함되어야 한다(동조 제4항). 특히 주목할 점은 위원의 자격 요건을 세분화하여 전문성과 윤리성의 균형을 맞춘 점이다. 위원회는 수의사로서의 전문성을 가진 사람뿐만 아니라, 민간단체가 추천하는 사람 또는 변호사나 법학·동물보호복지학 전공자 등 인문·사회적 소양을 갖춘 외부 전문가를 각각 1인 이상 반드시 포함하여야 한다(「동물보호법」 제53조 제3항; 농림축산검역본부, 2023, 3면).

위원회의 심의가 형식적인 통과 의례로 전락하는 것을 방지하기 위해 법령은 회의 성립과 의결에 있어 필수 위원의 참석을 의무화하고 있다. 「동물보호법 시행령」 제21조 제2항은 위원회의 회의를 재적위원 과반수의 출석으로 개의하고, 출석위원 과반수의 찬성으로 의결한다고 규정하면서, 단순히 정족수만 채우는 것이 아니라 수의사 위원과 민

간단체 추천 위원이 반드시 포함되어야 회의를 개의할 수 있다는 절차적 요건을 두고 있다(「동물보호법 시행령」 제21조 제3항).

동물실험윤리위원회 제도의 가장 큰 특징은 사전 심의에 그치지 않고, 승인 후 점검을 제도화했다는 점이다. 농림축산검역본부의 가이드라인(2023, 49면 이하)에 따르면 위원회는 승인된 계획서대로 실험이 수행되고 있는지, 실험동물의 복지 상태가 적정한지를 지속적으로 감독해야 한다. 「동물보호법」 제55조 제1항 및 「동물보호법 시행령」 제22조 제1항은 위원회가 반기별로 1회 이상 해당 동물실험시설의 운영 실태를 확인하고 평가하도록 의무를 부과하고 있다. 만약 심의받은 내용과 다르게 실험을 진행하거나 윤리적 규정을 위반한 경우, 위원회는 「동물보호법」 제55조 제2항에 따라 해당 실험의 중지를 요구할 수 있으며, 기관장은 특별한 사유가 없으면 이에 따라야 한다.

다. 소결

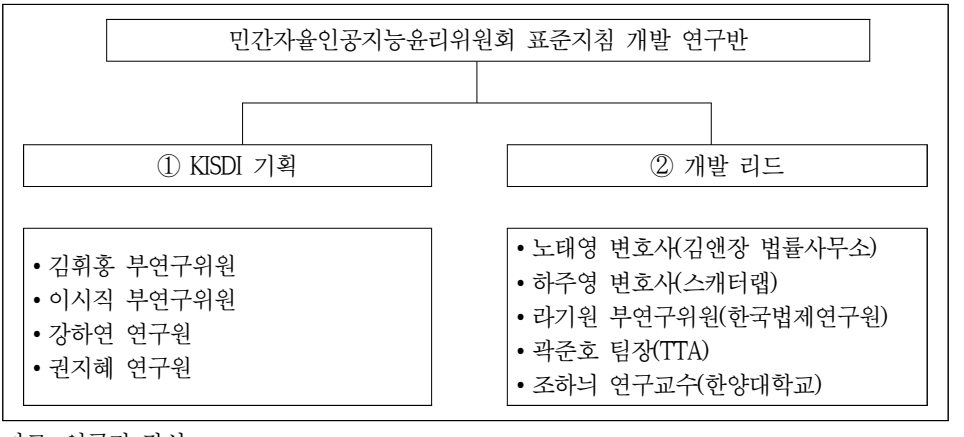
「생명윤리법」에 따른 기관생명윤리위원회와 「동물보호법」에 따른 동물실험윤리위원회는 「인공지능기본법」에 따른 민간자율위원회가 지향해야 할 투명하고 공정한 거버넌스의 지향점을 제시한다. 그러나 이를 참고할 때는 법적 성격의 근본적인 차이를 반드시 고려해야 한다. 검토한 두 위원회는 인간 또는 동물의 신체와 생명을 다루는 연구에 대해 법적으로 강제되는 의무로서의 성격이 강하게 나타난다. 따라서 위원회의 권한이 매우 강력하고 절차가 엄격하며, 정부 차원의 평가 및 인증 제도까지 연계되어 있는 경성 규범(Hard Law)의 체계 내에 있다. 반면, 「인공지능기본법」의 민간자율위원회는 기업이 자율적으로 설치 여부를 결정하는 임의 기구이며, 그 운영 또한 민간의 자율성을 존중하는 연성 규범(Soft Law)의 성격을 띤다. 따라서 기관생명윤리위원회와 동물실험윤리위원회의 규정을 기계적으로 도입하여 지나치게 엄격한 제척사유나 복잡한 심의 절차를 모든 인공지능 사업자에게 강요할 경우, 오히려 기업의 혁신을 저해하거나 위원회 설치 자체를 기피하게 만드는 부작용을 낳을 수 있다. 그러므로 표준지침을 마련할 때는 독립성, 공정성, 사후 관리에 관한 양 위원회의 취지는 계승하되, 절차는 간소화할 필요가 있다. 예컨대, 중소기업이나 스타트업의 경우, 외부 위원 선임의 부담을 줄여주거나, 운영 방식을 인공지능 개발 환경에 맞춰 폭넓게 인정하는 등 유연한 적용이 필수적이다. 즉, 자율적 책임을 강화하는 방향으로 이 모델들을 참고하여야 한다.

2. 표준지침 연구반 운영

가. 연구반 구성

민간자율위원회 표준지침(안)을 마련하기 위해 학문·연구, 산업, 법조 등 분야별 전문가 총 9인으로 연구반을 구성하고, 정보통신정책연구원 김휘홍 부연구위원이 연구반 운영을 총괄하였다. 연구반의 구성원과 역할은 다음과 같다.

〔그림 3-1〕 「민간자율인공지능윤리위원회 표준지침(안)」 개발 연구반 구성



자료: 연구진 작성

나. 연구반 활동

연구반은 2025년 7월부터 11월까지 약 5개월간 운영하였으며, 7월 10일 킥오프(Kick-off) 회의를 시작으로 총 5차례의 온오프라인 회의를 개최하였다. 킥오프 회의에서는 목표 및 표준지침 구성안을 공유하고 추진 일정 조율 및 유사 위원회를 검토하였으며, 위원회 설치 주체, 구성 및 운영에 대한 자율성, 공정하고 중립적인 구성 방안 등 핵심 내용을 도출하였다.

〔그림 3-2〕 「민간자율인공지능윤리위원회 표준지침(안)」 개발 연구반 사진

킥오프 회의 이후에는 표준지침의 실효성 및 확산 가능성 제고 방안을 논의하고, 각 연구진의 담당 영역을 배분하였으며, 연구 위원별로 각자 맡은 영역에서 자료를 조사하고 원고를 작성하였다. 온라인 및 오프라인 회의를 통해 진행 상황을 공유하고 작업 내용에 대해 논의하였으며, 추후 민간자율위원회 표준지침(안)의 실효성과 완성도를 높이기 위해 학계·산업계·법조계 등 인공지능 분야 외부 전문가들의 의견을 다각도로 수렴하였다.

〈표 3-1〉 「민간자율인공지능윤리위원회 표준지침(안)」 개발 연구반 활동

차수	일자	방식	주요 내용
1	'25년 7월 10일	오프라인	- 킥오프(Kick-off) 회의
2	7월 25일	오프라인	- 표준지침 실효성 및 확산 가능성 제고 방안 논의
3	8월 18일	온라인	- 목차별 지침 내용 검토
4	9월 3일	온/오프라인	- 표준지침 설계 방안 논의
5	11월 3일	오프라인	- 표준지침 초안 최종 점검

자료: 연구진 작성

다. 표준지침 개발 원칙 및 구성

본 민간자율위원회 표준지침(안)은 3가지 기본 원칙을 바탕으로 개발되었다. 첫째, 인공지능 윤리원칙의 확산 및 실천 지원 목적을 기반으로 개발되었다. 「인공지능기본법」 제28조 제1항에 따라 본 표준지침은 획일적 규제보다는 맞춤형 윤리 실천을 가능하게 하는 자율적 윤리 거버넌스를 지향하며, 윤리원칙이 추상적 논의에 그치지 않도록 실제 준수와 실천에 초점을 둔다. 둘째, 「인공지능기본법」 제28조 제3항에 따라 민간자율위원회의 설치·구성·운영상 자율성이 보장되며, 위원회가 사회적 책임을 다할 수 있도록 법률이 정하고 있는 최소한의 공익성을 요구하는 방향으로 개발하였다. 셋째, 「인공지능기본법」 제28조 제2항 각 호의 업무에 대해 업무별 사례, 예시 등을 마련하여 기관 또는 단체가 특성에 맞는 인공지능 윤리 지침의 수립, 윤리교육, 사례 검토 등을 통해 추상적 윤리원칙을 실무 차원에서 구현하도록 개발하였다.

표준지침(안)은 총 13개의 조문으로 구성하였다. 내용은 크게 '목적', '구성' 및 '운영'으로 구성하였으며, 제1조는 목적에 대한 내용이다. 제2조~제6조 구성에 관한 내용은 '위원회 수행 업무 범위', '업무 수행 방법', '위원 자격', '위원 수', '위원의 해촉', '위원의 임기', '위원장의 직무', '간사' 등으로 구성되어 있다. 제7조~제13조 운영에 관한 내용은 '회의 소집'과 '의견 청취', '의사결정', '위원의 제척, 기피, 회피', '회의록', '후속조치', '인공지능 윤리원칙 준수 보고서' 등으로 구성되어 있다.

3. 표준지침(안) 의견수렴

표준지침(안)의 완성도를 제고하고자 연구반 초안 작성 과정에서 도출된 쟁점과 이견사항을 중심으로 서면 자문을 실시하였다. 자문은 2025년 10월 27일부터 11월 3일까지 중소·스타트업 관계자 6인을 대상으로 진행되었으며, 주요 검토 의견은 다음의 표와 같다.

〈표 3-2〉 「민간자율인공지능윤리위원회 표준지침(안)」 개발 방향에 대한 서면 자문 예시

구분	주요 내용
민간자율위원회 취지	인공지능사업자의 판단에 따른 내부 자율규제 장치로 기업 현황과 산업 특성에 맞게 철저히 자율적으로 운영되어야 함
위원 구성	기업 규모와 인공지능 활용 수준 등 사업 방향을 고려해 필요 최소한의 인원으로 구성하는 것이 현실적임
주요 전문성	법률·시장·기술 전문가의 협력을 통한 실질적 리스크 관리
외부 전문가 확보	정부·협회 차원에서 공동 전문가 풀을 운영하여 기업의 부담을 줄이고, 등록 전문가 활용 시 법정 요건을 충족한 것으로 인정하는 유연한 제도적 장치 필요
결과보고서 공개	결과보고서의 공개는 원칙적으로 기업 자율에 맡겨야 하며, 필요 시 협회 차원의 익명 집계 등의 수준으로 공개하는 방안 가능

자료: 연구진 작성

표준지침의 실효성을 제고하고 산업계의 수용성을 높이기 위해, 본 지침이 새로운 규제 부담으로 작용하거나 기업 활동을 위축시키지 않아야 한다는 의견이 지배적이었다. 이에 따라 위원 수의 최소화 구성 및 정부·협회 차원의 전문가 풀(Pool) 운영 필요성 등 현장의 여건을 반영한 실무적 제언들이 강조되었다. 이러한 중소·스타트업의 의견을 적극 수렴하여 표준지침 초안을 도출하였다.

이어 12월 5일부터 12월 12일까지 산업계뿐만 아니라 학계·법조계 등 외부 전문가 14인을 대상으로 제2차 전문가 의견조사를 실시하여 다양한 의견을 수렴하였다. 초안에 대해 전문가들은 표준지침 초안의 '소규모 기업 적용 가능성'과 '최소 기준 설정'에 전반적으로 동의하였으나, 실무 예시 및 세부 기준의 보완이 필요하다고 지적하였다. 예를 들어, 위원장의 권한 규정이 다소 추상적이어서 실질적인 권한 행사 시 부담이 될 수 있다는 점과, 인공지능 산업의 특수성을 반영한 위원 해촉 사유의 명확화가 필요하다는 의견이 제시되었다. 투자 관계, 용역·컨설팅 제공, 지분 보유, 특정 프로젝트를 매개로 직·간접적인 이해관계를 맺는 경우가 빈번하다는 것이 이유였다. 그밖에 다양한 의견을 바탕으로 조문 간 상충 가능성을 검토하고 세부 수정을 거쳐 표준지침 수정안을 도출하였으며, 이 과정에서 조문의 조정은 최소 기준 제시라는 본래의 취지를 훼손하지 않는 범위 내에서 신중하게 수행하였다.

제 3 절 민간자율위원회 표준지침(안) 소개

1. 민간자율위원회의 설치 및 구성

가. 민간자율위원회의 구성

1) 위원 수

위원회의 구성 인원은 「인공지능기본법」 제28조 제3항 본문에 따라 해당 기관 또는 단체의 자율적인 판단에 맡겨져 있다. 다만 기관의 규모와 인공지능 활용 범위를 고려하여 권장 인원을 제시할 수 있으며, 일반적으로 소규모 기관의 경우 3~5인, 대규모 기관의 경우 7인 이상의 위원으로 구성하는 것이 바람직하다. 아울러 위원회가 실질적인 의사결정 기구로 기능하기 위해서는 다수결 원리가 원활히 작동할 수 있도록 위원장 1인을 포함하여 최소 3인 이상의 홀수 인원으로 구성하는 것이 필요하다.

2) 위원 구성

개별 위원의 구성에 관한 사항도 원칙적으로 해당 기관 또는 단체의 자율적인 판단에 따라 결정할 수 있으며, 내부 임직원을 위원회의 위원으로 임명하는 방식도 가능하다. 다만 내부 임직원을 위원으로 위촉하는 경우에는 위원회가 해당 기관 또는 단체에 대해 실질적인 영향력을 행사할 수 있도록 집행력을 갖춘 인사를 포함하는 것이 바람직하며, 나아가 책임 있는 인공지능 윤리원칙의 실현을 도모하기 위해 기관장 또는 관련 부문의 임원급 인사를 최소 1인 이상 포함하는 것이 권장된다. 또한 2인 이상의 내부 임직원이 위원으로 참여하는 경우에는 다양한 관점에서의 논의가 이루어질 수 있도록 각 위원의 전문 분야를 고려하여 균형 있게 구성하는 것이 바람직하다.

3) 위원 구성 시 고려 사항

위원회 구성 시 「인공지능기본법」 제28조 제3항 단서에 규정된 다음 세 가지 요건이 준수되어야 하며, 이는 위원회의 전문성, 독립성, 다양성을 보장하기 위하여 법률이 설정한 최소한의 기준이다.

① 다양성

위원회는 특정 성별로만 구성될 수 없으며, 양 성별을 각각 최소 1인 이상 포함하여야 한다. 이는 성별에 따른 다양한 관점을 위원회 논의 과정에 반영함으로써 의사결정의 균형성과 공정성을 확보하기 위한 것이다.

② 전문성

위원회에는 사회적·윤리적 타당성을 평가할 수 있는 경험과 지식을 갖춘 사람을 최소 1인 이상 포함하여야 한다. 여기에서 말하는 '전문성'이란 인공지능 윤리를 비롯하여 이에 인접한 공학, 윤리학, 법학, 사회학, 심리학 등의 분야에서 관련 학위나 자격을 취득하였거나, 해당 전문 분야에서의 실무 경험과 경력을 통해 축적된 전문성을 의미한다. 구체적인 전문성의 인정 여부는 사회 일반의 통념과 기준에 비추어 합리적인 범위 내에서 개별 기관 또는 단체가 자율적으로 판단할 수 있다.

③ 독립성

위원회 위원 구성 시 해당 기관 또는 단체에 종사하지 않는 외부 인사를 1인 이상 위촉해야 한다. 해당 요건은 위원회의 객관성과 중립성 확보에 목적을 두고 있으며, 외부 위원이 해당 기관 또는 단체에 종속되어 있지 않았을 때 비로소 윤리적 판단을 내릴 수 있다는 고려가 반영된 것이다.

독립성은 단순히 '해당 기관 또는 단체 소속이 아니다'라는 소극적 조건만으로는 충분하지 않고, 해당 기관 또는 단체와 직·간접적인 경제적 이해관계에 있다면, 독립성의 부재를 이유로 외부 위원으로 위촉할 수 없다. 또한 이해관계의 유무는 금전적·경제적 대가가 제공되었는지, 거래관계 등 계약적 이해관계가 존재하는지 등을 종합적으로 고려하여 판단할 수 있다.

예를 들어 외부 위원의 자격을 심사할 때, 위촉일로부터 2년 이내 ① 임직원 관계(해당 기관의 현직 임직원, 겸직자, 파견 근무자), ② 용역·거래 관계(해당 기관과 계속적·지속적인 계약 관계(컨설팅, 자문, 납품, 용역 수행 등), ③ 주식 보유·투자 관계(해당 기관의 지분을 일정 수준 이상 보유한 주주, 또는 경영 의사결정에 실질적으로 영향력을 미칠 수 있는 투자자), ④ 경제적 수익 의존 관계: 고문료·자문료 등 일정 규모 이상의

경제적 보상을 해당 기관으로부터 받아온 경우라면 외부 위원으로 선임하는 것은 적절하지 않다.

나아가 위원을 위촉할 때, 위촉 후보자는 위의 예시를 참고하여 '이해관계 현황 서약서'를 작성·제출하고, 해당 기관 또는 단체의 장은 이를 검토하여 위촉 여부를 판단해야 한다. 이러한 절차를 문서로 기록·보관한다면, 위원회의 구성에 관한 공정성과 투명성을 확보하는 중요한 근거가 될 수 있을 것이다.

또한 동일인이 위원 구성 요건 중 둘 이상의 요건을 동시에 충족하는 경우, 예를 들어 외부 위원이면서 동시에 사회적·윤리적 전문성을 갖춘 경우에는 독립성과 전문성 요건을 모두 충족한 것으로 볼 수 있다. 이는 「인공지능기본법」 제28조 제3항 단서에서 요건별로 서로 다른 인물을 반드시 둘 것을 요구하고 있지 않다는 점에 근거한 내용으로, 전문성과 독립성, 전문성과 다양성, 독립성과 다양성 등 다양한 조합이 가능하며 이를 통해 위원 수를 최소화하면서도 법에서 요구하는 구성 요건을 충족하는 위원회 구성이 가능하다.

나. 위원의 위촉과 임기

1) 위촉 권한과 절차

위원의 위촉은 '해당 기관 또는 단체의 장'이 수행하는 것이 일반적이며, 기업의 경우에는 이사회 등 내부 의사결정기구가 위촉의 주체가 될 수도 있다. 위촉 과정에서는 앞서 설명한 위원 구성 시 고려 사항에 따라 「인공지능기본법」 제28조 제3항 단서에서 정한 필수 요건이 충족되고 있는지를 반드시 확인하여야 하며, 위원회의 목적과 기능에 부합하는 인재를 선발하는 것이 중요하다. 또한 위원을 위촉할 때에는 위원의 역할과 책임, 임기, 활동 범위 등을 사전에 명확히 안내하여야 하며, 임기의 시작일과 종료일을 회의록 및 위원 명부에 체계적으로 기록·관리하는 것이 바람직하다.

2) 위원의 임기

위원의 임기는 해당 기관 또는 단체가 상황을 고려하여 자유롭게 정할 수 있지만, 2년 임기와 연임을 허용하는 사례가 가장 일반적이다. 먼저, '2년의 임기'는 위원이 위원회 업무에 충분히 숙달되어 전문성을 발휘할 수 있으면서 동시에 새로운 관점의 주기적인

유입 가능성을 담보하는 기간이고, '연임'을 허용한다면 기존 위원의 전문성을 계속해서 활용할 수 있고 업무의 연속성을 바탕으로 위원회 운영상 안정성을 확보할 수 있다는 장점이 있다.

다만, 위원의 연임 여부는 기관 또는 단체의 장이 해당 위원의 활동 성과와 민간자율위원회의 운영상 필요에 따라 판단할 수 있으며, 연임 여부를 결정할 때에는 출석률, 기여도, 이해 상충 관리 준수 여부 등을 종합적으로 고려할 필요가 있다.

나아가 위원회 구성의 다양성과 독립성을 확보하기 위해서는 ① 동일 위원이 장기간 재임하여 위원회 운영의 활력이 저하되지 않도록, 연속 2회(최대 4년)까지만 연임을 허용하는 것이 권장된다. ② 다만, 전문성이 특별히 요구되는 경우(예: 법학, 윤리학, 기술 분야 핵심 전문가), 기관 사정에 따라 추가 연임이 가능하며, ③ 연임 시에는 임기 종료 시점마다 활동 평가 및 이해상충 여부를 검토하여 재위촉 여부를 결정하는 등의 권고 기준을 고려할 수 있다.

3) 후임자의 위촉과 임기

위원회 위원 수가 3인 미만이 될 경우, 위원회의 기능이 사실상 정지될 수 있으므로 가능한 한 신속하게 후임 위원을 위촉하여 위원회 활동의 공백을 최소화할 필요가 있다. 이는 다수결 원리가 작동할 수 있는 최소한의 정족수를 유지하기 위한 것으로, 필요하다면 최소 30일 이내에 결원을 보충하도록 하는 권장 기준을 둘 수 있다.

한편 위원의 결원에도 불구하고 위원 수가 3인 이상으로 유지되는 경우에는 결원 보충 여부를 기관 또는 단체의 재량에 맡길 수 있으나, 위원회의 안정적인 운영을 위해 가능한 한 후임 위원을 위촉하거나 임명하는 것이 바람직하다.

아울러 위원의 해촉이나 사퇴 등으로 새롭게 위촉된 위원의 임기는 전임 위원의 잔여 임기로 하는 것이 적절하다. 이는 위원회 운영의 연속성을 확보하고, 전체 위원의 임기를 조화롭게 관리하기 위한 조치이다.

다. 위원장

1) 위원장의 선출

위원장의 리더십과 전문성은 위원회의 실효성과 영향력에 직접적인 영향을 미치므로, 위원장 선출 과정에서는 전문성, 소통 능력, 중립성 등 다양한 요소를 종합적으로 고려할 필요가 있다. 유사 위원회의 사례를 살펴보면, 위원장은 위원 중에서 호선으로 선출하는 방식이 일반적이다. 호선 방식은 위원들 간의 상호 신뢰와 합의를 기반으로 하여 위원회의 독립성과 민주적 운영을 보장하는 가장 합리적인 방법이라 할 수 있다.

2) 위원장의 역할

위원장은 위원회를 대표하여 회의를 소집·주재하는 등 위원회 업무 전반을 총괄하는 핵심적인 역할을 수행하고, 회의 진행 시 위원 전원이 충분히 토론할 수 있도록 발언 기회를 균등하게 보장하고, 의결 과정이 절차적으로 정당하게 이루어지도록 관리할 책임이 있다. 또한 위원장은 모든 위원이 의안의 내용을 충분히 이해할 수 있도록 필요한 설명을 제공하고, 위원들이 충분한 토의와 숙고를 거쳐 의결할 수 있도록 회의를 운영하여야 한다.

3) 위원장의 대행

위원장이 부득이한 사유로 직무를 수행할 수 없는 경우에는 위원장이 지명한 위원이 그 직무를 대행한다. 이 경우 위원장이 임기 개시와 동시에 직무대행자를 미리 지명해 두는 것이 바람직하다. 이러한 위원장 직무대행 체제는 위원회 운영의 연속성을 확보하고, 긴급한 상황에서도 신속하고 안정적인 의사결정을 가능하게 한다.

라. 위원의 해촉

위촉권자(일반적으로 해당 기관 또는 단체의 장)는 위원이 질병이나 기타 개인적인 사정 등 위원 개인의 의사와 무관한 불가항력적인 사유로 직무를 수행할 수 없게 된 경우 해당 위원을 해촉할 수 있다. 또한 위원이 직무와 관련하여 이해충돌 행위, 기밀 누설, 부정행위 등 비위 사실이 있는 경우에도 해촉할 수 있다.

아울러 위원이 직무를 성실히 수행하지 아니하거나 품위 손상, 그 밖의 사유로 위원회의 위원으로서 적합하지 않다고 인정되는 경우, 예를 들어 지속적으로 회의에 불참하거나 위원회 활동에 대한 현저히 소극적인 태도, 위원으로서의 품위를 손상시키는 행위 등을 한 경우 해촉이 가능하다.

한편, 이해충돌의 우려 등으로 위원 스스로 객관적이고 공정한 직무 수행이 곤란하다고 판단하여 사임 의사를 명확히 밝힌 경우에도 위원의 자발적 사퇴로 처리할 수 있다.

마. 간사

1) 간사의 임명

위원장은 위원회의 원활한 운영을 위하여 위원이 아닌 자 중에서 간사 1인을 행정 지원 인력으로 임명할 수 있다. 임명된 간사는 회의 운영의 연속성과 회의록 등 기록 관리의 전문성을 확보하기 위한 역할을 수행하며, 기관 또는 단체 내 전담 인력을 활용하는 것이 바람직하다. 이 경우, 위원장은 선출 직후 간사 임명 여부와 간사의 구체적인 업무 범위를 문서로 명확히 정함으로써, 위원회의 효율적인 운영은 물론 간사의 업무 부담을 줄일 수 있다. 한편, 간사가 부재한 경우에는 위원장이 해당 기관 또는 단체 소속 인원 중에서 적합한 자를 지정하여 간사의 업무를 대행하게 할 수 있다.

2) 간사의 업무 범위

간사는 위원회의 원활한 운영을 위하여 행정 지원, 교육 지원, 안건에 대한 사전검토 지원 및 그 밖에 위원장이 지시하는 사항을 수행한다. 구체적으로 간사의 업무에는 회의 준비, 회의록 작성, 안건 관리, 관련 자료의 배포 등 위원회의 조직적이고 체계적인 운영을 위한 행정적 지원이 포함된다. 간사가 작성한 회의록은 위원회 회의 종료 후 일정 기간 내에 위원 전원에게 회람되어야 하며, 위원장의 승인을 거쳐 최종 확정·보관한다. 다만, 회의 준비 등 행정 지원 과정에서 간사의 역할은 위원회의 효율적인 운영을 돕기 위한 보조적 기능에 한정되며, 위원회의 심의·의결 등 의사결정을 대체할 수 없다.

다음으로는 위원회 신임 위원 대상 오리엔테이션이나 정기적 인공지능 윤리교육 운영 보조 등 교육 지원 업무를 수행하며, 회의 상정 안건에 대한 기본 사실관계 정리, 관련 자료 수집 및 정리 등 사전검토 지원 업무도 수행한다. 그 밖에 필요에 따라 현장 조사 보조, 외부 전문가 섭외 등 위원장이 지시하는 업무를 수행할 수도 있다.

한편, 간사는 업무 수행 과정에서 알게 된 사항에 대하여 비밀을 유지하여야 하며, 특히 회의록 및 관련 자료의 작성·보관·관리와 관련하여 필요한 보안 조치를 준수하여야 한다.

2. 민간자율위원회의 운영

가. 회의 종류와 소집

1) 회의 종류와 진행 방식

위원회 회의는 기본적으로 회의의 정례화(정기회의)로 예측가능성을 확보하는 동시에 탄력적 소집(임시회의)을 가능하게 함으로써 신속성도 확보할 수 있도록 운영하는 것이 바람직하다. 구체적으로 정기회의는 매월, 분기별 또는 연 2회 등 일정한 주기에 따라 소집하여 기본적인 현안 점검과 정기 보고를 실시하고, 「인공지능기본법」 제28조 제2항 각 호에 따른 반복적·예정된 업무 등을 처리함으로써 위원회 운영의 지속성과 안정성을 도모할 수 있다. 한편 임시회의는 예상치 못한 정책 변경이나 사고 발생 등 긴급한 의사결정이 필요한 사유가 발생한 경우에 회의를 소집함으로써, 주요 현안에 대해 신속하고 유연하게 대응할 수 있도록 하는 장점이 있다. 이러한 운영 방식은 위원회가 정례적인 업무 수행과 함께 돌발적인 상황에도 효과적으로 대응할 수 있는 기반을 마련하는 데 기여할 수 있다.

다음으로 회의 진행 방식에 있어서는 '대면 중시 원칙'을 기본으로 하되, 비대면 회의에 대한 현실적 요구를 반영한 절충적이고 현대적인 운영 형태를 채택함으로써 원활한 소통과 신뢰 형성은 물론, 상황에 따른 탄력적 대응 가능성을 확보할 필요가 있다. 대면 회의는 비언어적 소통이 가능하고 실시간 상호작용이 활발하며, 위원들의 책임감을 제고하는 데 효과적이라는 점에서 심도 있는 논의와 신뢰 기반의 의사결정이 이루어질 수 있다. 다만, 현안에 대해 긴급하고 즉각적인 의사결정이 필요하나 시간적 제약으로 인해 대면 회의 소집이 곤란한 경우이거나 감염병 확산이나 자연재해, 위원의 해외 체류, 교통마비 등과 같은 객관적 장애로 인해 대면 회의가 불가능한 경우에는 예외적으로 비대면 회의를 진행할 수 있다. 이와 같은 예외 판단에 대한 권한은 위원장에게 부여할 수 있으나, 자의적인 결정이 이루어지지 않도록 일반적인 편의나 개인적 선호만을 이유로 비대면 회의를 선택하는 것을 제한하는 의미로 "특별히" 필요한 상황으로 한정함으로써,

대면 회의 원칙을 유지하면서도 불가피한 상황에 대한 합리적인 대응이 가능하도록 할 수 있다.

2) 정기회의 소집 빈도

회의의 정례화를 위한 주기는 원칙적으로 해당 기관 또는 단체가 자신의 활동 현황과 업무 수행의 필요성 등을 종합적으로 고려하여 내부 규정에 따라 자율적으로 정할 수 있다. 기관이나 단체의 규모, 업무의 성격, 위원회에 부여된 역할과 책임 등에 따라 적절한 주기는 달라질 수 있으며, 이러한 특성을 반영한 유연한 기준 설정이 바람직하다.

정기회의의 소집 주기는 매월, 분기별, 연 2회 등 다양한 방식이 고려될 수 있으나, 연 1~2회 실시하는 것이 일반적이다. 이는 「인공지능기본법」 제28조 제2항 각 호에 따른 업무의 목표를 달성하고, 위원회 업무 수행 절차를 표준화하는 데 필요한 적정한 소집 빈도로 평가되기 때문이다. 이러한 주기 설정은 위원회의 안정적인 운영과 함께 효율적인 업무 추진을 가능하게 하는 기반이 될 수 있다.

3) 임시회의 소집 요건

위원장은 예상치 못한 상황에 대한 조사와 대응이 필요한 경우 비정기적으로 임시회의를 소집할 수 있다. 임시회의의 소집 요건은 해당 기관 또는 단체가 자율적으로 결정하여 내부 규정에 반영할 수 있으며, 일반적으로 위원장의 소집권, 일정 수 이상 위원의 요구권, 상위 기관의 요청 등을 소집 요건으로 규정하여, 신속성과 절차적 정당성 간의 균형을 도모할 수 있다.

구체적으로 상위 기관이 긴급 현안 발생, 관련 정책 변경 등으로 위원회의 즉시 대응이 필요할 때 임시회 소집이 발동될 수 있으며, 이는 위원회의 자율성을 존중하면서도 기관 또는 단체 업무와의 연계성을 확보하는 장치가 될 수 있다. 또한 위원장의 직무 또는 위원회 업무 수행에 대하여 중대한 이견이 있거나 외부 환경 변화로 기존 방침의 재검토가 필요할 경우, 정기회의에서 미처 다루지 못한 현안 등이 있는 경우 위원의 집단 발의를 통해 임시회를 소집할 수 있다. 이 때 위원의 3분의 1 이상의 소집 요구 요건은 과반수보다는 낮아 소수 보호 기능을 하면서 동시에 무분별한 회의 소집 요구를 방지하여 합리적 다수의 의사를 존중하기 위한 균형점에 해당한다고 볼 수 있다.

한편, 임시회의를 소집하는 때에도 사전 안건 상정 원칙은 준수되어야 하며, 심의·의결의 대상도 사전에 고지된 안건으로 제한하여 회의 범위를 명확히 할 필요가 있다. 물론 사전 고지 원칙의 경직성이 때로는 위원회 운영의 실용성을 가로막는 장애물로 작용할 수 있으나, 이를 완화하기 위하여 예를 들어 '출석위원 3분의 2 이상'이라는 특별 다수결 요건을 두어 안건 추가 가능성을 마련할 수도 있다.

나. 회의 소집 절차

위원회 회의 소집 통지는 기본적으로 위원회의 운영에 있어 절차적 투명성을 확보하고, 위원들이 상정된 안건에 대해 충분히 숙고할 수 있는 시간을 보장하기 위한 중요한 절차이다. 원칙적으로 회의 소집 통지의 내용과 통지 기간은 해당 기관 또는 단체가 내부 규정을 통해 자율적으로 정할 수 있으며, "개최일 7일 전" 통지는 다수의 유사 위원회에서 채택하고 있는 일반적인 기준으로, 위원들이 복잡한 안건을 검토하고 사전에 의견을 정리하기에 합리적인 기간으로 평가된다. 또한 회의의 일시와 장소를 사전에 통지함으로써 위원들이 업무 일정을 미리 조정하고 회의 참석을 준비할 시간을 확보하여 위원의 실질적인 참여를 보장하는 데 기여할 수 있다.

또는 회의 소집 통지는 서면, 이메일, 전자문서 등 전자적 방식으로 할 수 있으며, 전체 재적 위원에 대하여 개별적으로 이루어져야 하며, 대표자를 통한 간접 통지나 일부 위원을 제외한 통지는 절차상 하자로 평가될 수 있으므로 유의할 필요가 있다. 한편, 행정 지원 업무를 담당하는 간사를 임명한 경우에는 간사가 위원장의 지시에 따라 회의 소집 통지와 관련된 절차를 수행할 수 있다.

다만 긴급한 현안 처리를 위해 임시회의를 소집하는 경우에도 절차적 정당성과 투명성이 유지될 수 있도록 소집 사유와 통지 기간 등을 위원회의 운영 원칙이 훼손되지 않도록 일정한 절차적 기준을 갖출 필요가 있다.

임시회의를 소집할 때에는 긴급한 현안의 처리 필요성이나 업무 수행과 관련한 중대한 이견의 존재 등 소집 사유를 구체적으로 특정하여 재적 위원 전원에게 전달하고, 임시회의의 통지 기간은 해당 기관 또는 단체가 위원회의 구성 및 운영에 관한 내부 규정을 통해 자율적으로 정할 수 있다. 다만 위원들의 회의 참석과 안건에 대해 실질적으로 검토할 시간을 보장하기 위하여 적어도 회의 개최 48시간 전까지 회의의 일시와 장소,

회의 안건 및 심의자료를 전달하는 것이 바람직하다.

나아가 불가피한 사정이 있는 경우에는 회의 당일에 회의 안건과 심의자료를 배포하는 방안도 고려할 수 있으며, 더 나아가 재적 위원 전원의 동의가 있는 경우에는 예외적으로 회의 안건만을 통지한 상태에서 임시회의를 소집하는 것도 허용될 수 있다. 임시회의 소집에 관한 통지 방식은 정기회의와 마찬가지로 서면 또는 이메일, 전자문서 등 전자적 방식에 따르며, 모든 재적 위원에게 개별적으로 이루어져야 하며, 이를 통해 임시회의의 신속성과 함께 절차적 투명성도 함께 확보할 수 있다.

다. 전문가 의견 청취

위원회는 실질적이고 전문적인 심의·의결을 위해 필요에 따라 기관 또는 단체 내부의 담당자나 관계 전문가를 회의에 출석하게 하여 의견을 청취하거나, 사전에 서면으로 의견을 요청할 수 있다. 관계 전문가는 해당 안건과 관련된 분야에서 전문성을 갖춘 자로서, 해당 분야의 교수나 연구원, 실무 경험이 풍부한 기술자, 법률 전문가, 시민단체 활동가 등이 포함될 수 있다. 다만 외부 전문가는 회의 안건과 직접적인 이해관계가 없을 것이 요구되며 이를 확인하기 위해 경제적·인적 이해관계 여부를 사전에 확인할 필요가 있는데, 이는 심의·의결의 공정성과 신뢰성을 확보하기 위한 중요한 절차이다.

위원회가 대면 논의를 통해 안건을 심의하고 의결하는 성격을 갖는 점을 고려할 때, 의견 청취 역시 원칙적으로 회의에서 직접 대면 방식으로 의견을 청취하고 질의·응답을 진행하는 것이 적절하다. 다만 일정이나 물리적 제약 등으로 이러한 절차가 어려운 경우에는 예외적으로 서면을 통한 의견 청취도 가능하다.

의견 청취는 회의의 원활하고 효과적인 진행을 위한 조치이므로, 위원장의 제안에 따라 진행할 수 있고, 위원장뿐만 아니라 각 위원 역시 의견 청취가 필요하다고 판단하는 경우 위원장에게 이를 제안할 수 있다. 다만 이러한 의견 청취 절차가 회의 안건의 심의 및 의결을 방해하거나, 기타 부당하다고 판단되는 경우에는 위원이 이에 대해 이의를 제기할 수 있고, 위원장은 위원들의 의견을 종합적으로 고려하여 의견 청취를 계속 진행할지 여부를 결정할 수 있다. 또한 의견 청취를 위해 대상자가 회의에 출석하는 경우, 의견 청취의 대상 안건과 관련 자료를 사전에 제공하고, 대면으로 진행하는 회의 소집과 진행에 관한 일반적인 원칙을 준용하여 진행한다.

교육기관 및 연구기관의 학문의 자유(헌법 제22조)를 실질적으로 보장하기 위해서는 민간자율위원회가 특정 연구를 심의할 때 해당 연구책임자나 참여자에게 충분한 의견 진술의 기회를 제공하는 것이 권장된다. 위원회의 '윤리원칙 준수 여부 확인'이나 '조사·감독권' 행사는 연구자의 연구 지속 여부와 대외적 평판에 상당한 영향을 미칠 수 있다. 특히 인공지능 연구는 기술적 복잡성이 매우 높고, 적용 맥락에 따라 윤리적 판단 기준이 가변적이어서, 단순한 서면 심사만으로는 연구의 진정한 의도나 복잡한 기술 구조를 명확히 파악하는 데 한계가 있을 수 있다. 따라서 연구자의 학문의 자유를 부당하게 침해하지 않으면서 위원회가 합리적인 결정을 내리기 위해서는, 연구자의 직접적인 설명과 위원 간의 질의응답 및 심도 있는 논의를 통한 상호 이해 과정이 반드시 선행되어야 한다.

라. 회의 성립과 의사결정

1) 회의 성립

위원회 회의가 적법하게 개최되기 위해서는 최소한의 정족수(의사정족수)를 사전에 정해 둘 필요가 있다. 핵심은 정족수를 정함에 있어서 관련 법령에서 정한 기준을 준수하되, 해당 기관 또는 단체의 규모와 운영 여건을 반영하여 위원회가 탄력적으로 운영될 수 있도록 하는 것이 중요하다. 의사정족수는 재적위원 전원의 출석을 요구하는 것은 아니지만, 의사결정의 대표성과 민주적 정당성을 확보할 수 있는 적정한 수준으로 정하여야 한다.

일반적으로 많은 회의체에서는 재적위원의 과반수 출석을 의사정족수로 정하고 있으므로, 위원회 역시 이에 준하여 운영하는 방안을 고려할 수 있다. 다만 의사정족수는 회의의 절차적 정당성을 담보하기 위한 취지를 고려할 때 이를 임의로 변경하거나 조정하는 것은 원칙적으로 허용되지 않는다. 그럼에도 불구하고 본 위원회가 자율기구로서의 성격을 고려할 때 좀 더 유연한 운영을 허용할 필요성이 있고, 기업들이 인력 및 자원의 부족으로 제한된 인원으로 위원회 및 회의를 운영해야 하는 상황이 발생할 수 있다는 점을 고려할 때 회의 개최의 긴급한 필요성 및 정족수 충족이 어려운 불가피성이 인정된다면 의결로 이를 조정하는 것은 예외적으로 허용될 수 있다.

한편 「인공지능기본법」 제28조 제3항 단서는 다양성, 전문성, 외부성을 고려하도록 세

가지 기준을 규정하고 있으므로, 각 기준에 해당하는 위원이 최소 1인 이상 출석하도록 하는 것이 적절하다. 이러한 장치는 법률의 요구사항을 온전히 준수하여 제도적 정당성을 확보하고 각 관점의 균형적 반영을 위한 실질적 다양성을 구현하는 데 도움이 될 수 있다. 예를 들어 3인으로 구성된 위원회의 경우, 출석한 2인이 종합적으로 성별 다양성, 전문성, 외부성 요건을 모두 충족할 때 회의를 개의할 수 있다. 참고로 회의 개시 전에는 위원장 또는 간사가 출석 현황을 확인하여 의사정족수 충족 여부와 함께 이해 상충이 있는 위원의 제외 여부를 확인하고 심의 진행을 개시한다.

2) 의사결정

위원회의 업무는 안건 심의에 이어지는 회의체 의사결정, 즉 의결을 기초로 수행될 수 있으며, 이에 요구되는 정족수인 의결정족수를 사전에 정해 둘 필요가 있다. 일반적으로 회의체 기구에서는 출석위원 과반수의 찬성으로 의결하는 방식을 채택하고 있으므로, 위원회 의결정족수도 이에 준하여 정할 수 있다. 다만 안건과 관련하여 이해 상충이 있는 위원은 “해당 안건에 관하여” 의사정족수에 포함되지 않으며, 의결정족수도 그를 제외한 출석 인원을 기준으로 판단한다.

한편, 실효성 중심의 관점에서 볼 때, 위원회의 의결 내용은 해당 기관이나 단체에서 실제로 이행될 수 있도록 구체적으로 제시되어야 한다. 지나치게 추상적이거나 일반적인 의결 내용은 후속 조치를 불분명하게 하여 구체적인 이행을 어렵게 만들고, 나아가 전체 의결 사항에 대한 이행 가능성이나 실효성을 저하시킬 수 있으므로, 심의 및 의결 과정에서 이행가능성과 구체적인 실행 방안에 대해 충분히 논의하는 것이 중요하다.

또한 의사결정 단계에 이르기 이전부터 안건 자체를 가급적 구체적이고 명확하게 구성하는 방안도 고려할 수 있다. 예를 들어 ‘인공지능 대출심사 알고리즘 차별 문제 개선 방안 검토’와 같은 포괄적인 안건보다는 ‘인공지능 대출심사 알고리즘의 성별 편향성 개선을 위한 성별 변수 제거 및 대체 지표 도입(안)’과 같이 실행 내용이 명확한 안건이 보다 실질적이고 이행 가능성이 높은 논의를 가능하게 만들 수 있다.

마지막으로 위원회에서 의결된 사항은 원칙적으로 관련 법령과 해당 기관 또는 단체의 내규 및 규정에 위반되지 않아야 한다. 만약 위원장이 의결 내용이 법령이나 내부 규정과 충돌할 가능성을 인지한 경우에는 사전에 해당 기관 또는 단체 담당자를 통해

관련 내용을 파악하고 그 내용을 위원들에게 안내하는 것이 바람직하다.

3) 타 연구윤리 규정 및 거버넌스 체계와의 연계

교육기관 및 연구기관은 「학술진흥법」 상 연구윤리지침이나 「생명윤리법」에 따른 기관생명윤리위원회 등 기존의 연구윤리 규정 및 거버넌스 체계를 이미 갖추고 있는 경우가 많다. 따라서 민간자율위원회의 운영이나 지침이 이러한 기존 체계와 충돌하지 않고 유기적으로 조화 및 연계될 수 있도록 각별히 주의해야 한다. 구체적으로는 기존 연구윤리 체계와의 심의·검증 절차 중복을 방지하고, 인공지능 연구·개발 활동에 대한 윤리적 검토의 일관성을 확보해야 한다. 또한 각 위원회 간 심의 결과와 정보를 상호 공유하는 협력 체계를 구축함으로써 연구자와 개발자의 과도한 행정적 부담을 방지하는 방향으로 운영되어야 한다. 예를 들어, 인공지능 기술 연구가 인간대상연구에도 해당하여 「생명윤리법」 상 기관생명윤리위원회의 심의를 동시에 받아야 하는 경우, 중복되는 범위 내에서 두 위원회 간의 협력 혹은 통합 심의를 가능하게 하는 방안을 적극적으로 도모할 수 있다.

마. 위원의 제척, 기피 및 회피

위원의 제척은 위원회의 공정성과 중립성을 확보하기 위한 장치이며, 위원회에서 객관적 심의를 위해 특별한 이해관계가 있는 위원을 배제하는 것을 주요 내용으로 한다. 여기서 “특별한 이해관계가 있는 위원”이란, 안건과 관련된 업무에 해당 기관 또는 단체의 내·외부에서 관여하고 있거나, 이에 개인적인 이해관계를 가지고 있는 경우 등을 의미한다. 이 때, 직접 업무에 관여하고 있는 경우에는 명확히 판단할 수 있지만, ‘기타 이해관계가 있는 경우’에는 스스로의 판단 또는 다른 위원들의 의견을 기초로 위원장이 이해 상충 여부를 판단할 수 있다. 그리고 “출석하지 못한다”의 의미는 위원 본인의 의사와 무관한 강제적 배제이며, 다른 위원들의 동의가 있어도 심의 및 의결에 참여할 수 없다. 다만 이해 상충이 있는 위원의 출석 자체가 심의 및 의결을 방해하거나 장애가 되는 경우가 아니라면 출석 자체는 허용할 수 있다.

또한 이해 상충이 있는 위원은 의사정족수 및 의결정족수 계산 시 재적 위원 수에서도 제외되는데 예를 들어 재적 인원 10인 중 이해 상충 위원 1인이 있는 경우, 9인을

기준으로 하여 재적 위원 과반수는 5인, 출석 위원의 과반수는 3인이 된다. 다만, 이해 상충이 있는 위원으로 인하여 정족수의 충족이 불가능한 경우는 예외로 할 수 있고, 이해 상충 있는 위원을 제외하면 남성 위원만 남는 경우, 예외적으로 남성 위원만으로 의사정족수는 충족된 것으로 볼 수 있다. 나아가 제척 이외에 제3자에 의한 기피신청, 위원 스스로의 회피 신청을 통해서도 심의, 의결에서 제외될 수 있는데, 만약 이해 상충이 있는 위원으로 인하여 정족수의 충족이 불가능한 경우 위원장은 위원을 보충 지명 또는 위촉하여 정족수 요건을 충족하고 정상적인 심의 및 의결이 이루어질 수 있도록 조치할 수 있다.

바. 회의록 작성 및 보관

1) 회의록 작성

위원회의 회의록은 체계적이고 신뢰할 수 있는 회의 기록을 통해 의사결정 과정의 투명성과 추후 검증 가능성을 보장하는 장치이다. 위원회의 회의록은 전문성과 일관성을 담보하기 위하여 원칙적으로 회의에 참석한 간사가 작성하며, 회의 진행에 관한 일체를 기록한다. 회의록에는 회의에서 채택된 안건, 안건별 및 일반 토의 사항, 최종 의결 사항을 포함하여 관련 위원 발언이 포함될 수 있으며, 회의의 공정성과 투명성을 확보하기 위하여 이해 상충이 있는 위원은 해당 안건의 심의·의결에서 배제되었다는 점을 회의록에 기재할 수 있다.

위원회 간사는 부득이한 경우를 제외하고 회의 종료 후 일정 기간 내에 회의록을 작성하며, 이 경우 회의록은 자유 양식으로 기록하며, 수기 혹은 전자적 형태를 포함하여 보관 및 열람이 가능한 형태로 작성되어야 한다. 구체적으로 회의록 작성 기한은 해당 기관 또는 단체의 내부 규정으로 정할 사항이지만, 타 위원회 사례들을 참고할 때 “회의 종료 후 15일” 내에 회의록을 작성하는 것이 일반적이다. 참고로 “부득이한 경우”라고 함은 예를 들어 질병, 사고, 천재지변 등 객관적 사유는 물론 녹음 장비 오작동, 자료 손실 등 기술적 장애를 포함한다고 볼 수 있다.

2) 회의록 확인과 채택

간사는 작성된 회의록을 특정 안건에서 제척된 위원을 제외하고 '회의에 참석한 모든 위원'이 회의록을 회람할 수 있는 상태로 제공하고, 확인, 수정 요청 및 이의제기 신청을 받는다. 여기서 회람 기간, 수정 요구, 이의제기 등에 관한 방식은 해당 기관 또는 단체가 내부 규정으로 자유롭게 정할 수 있다.

위원장은 회람을 마친 위원들의 수정 요구나 이의를 검토하고, 필요한 경우 회의록을 수정하며, 승인한 최종안을 정식 회의록으로 채택할 수 있다. 나아가 의사결정의 연속성을 보장하고, 이전 결의 사항의 실행 여부를 확인하기 위하여 차기 회의에서 직전 회차의 회의록 내용을 보고하도록 규정하는 것도 고려해 볼 수 있다.

3) 회의록의 공개 여부

회의록의 전부 또는 일부의 공개 여부는 기본적으로 해당 기관 또는 단체가 내부 규정으로 자유롭게 결정할 수 있다. 위원들의 솔직한 의견 개진을 촉진해 심의 품질을 확보하고, 인공지능 기술, 개인정보, 영업비밀 등 민감한 내용의 유출을 방지하기 위해 원칙적 비공개로 규정할 수 있고, 비공개를 원칙적으로 하더라도 예외적으로 위원회는 회의록의 공개 여부, 공개의 인적·내용적 범위, 공개 방식 등을 의결로 정할 수 있다. 이 경우, 위원회 의결 등에 따라 회의록을 공개하더라도 이름, 주민번호, 연락처 등 개인 식별 정보는 제외되어야 하며, 핵심 기술, 전략 정보 등 영업비밀은 해당 기관 또는 단체의 상황을 고려하여 제외할 수 있다. 또한 회의록의 공개 방식과 관련하여 주요 이해당사자 간 이메일, 해당 기관 또는 단체의 인트라넷 게시판, 해당 기관 또는 단체의 홈페이지 내 경영 정보란 등을 활용할 수 있고 이는 공개의 인적·내용적 범위에 따라 달리 설정할 수 있다.

사. 후속 조치

위원회의 의결 사항이 충분히 이행되지 않을 경우, 위원회 운영의 실효성 자체가 크게 영향을 받을 수 있다. 이에 따라 의결 사항이 합리적이고 충실하게 이행될 수 있도록 위원회 스스로는 물론 해당 기관 또는 단체의 구성원 및 이해관계자들의 노력도 필요하다.

위원회의 심의·의결을 거친 사항은 해당 기관 또는 단체의 인공지능 윤리원칙 준수에 관하여 중요한 내용이고, 특히 심의·의결 사항의 구체적인 이행은 결국 해당 기관 또는 단체의 대표자 또는 실무 담당자를 통하여 이루어질 수밖에 없는 점을 고려할 때, 심의·의결된 내용과 그 취지를 충분히 이해할 수 있도록 사전에 관련 정보를 제공할 필요가 있다.

다만, 이 경우 회의록 전체를 공개하기보다는 위원회의 의결에 이르게 된 배경과 취지, 주요 고려 사항 등을 정리하여 전달함으로써, 관계자들이 해당 의결의 의미를 충분히 인식하고 공감하여 적극적으로 이행에 참여할 수 있도록 유도하는 것이 바람직하다.

아. 보고서 작성 및 공개

1) 윤리원칙 준수 보고서의 작성

해당 기관 또는 단체는 위원회의 다양한 성과를 '윤리원칙 준수 보고서'에 담아 주기적으로 발간할 수 있다. 해당 윤리원칙 준수 보고서는 위원회의 활동 내용을 이해관계자들에게 공유하고, 해당 기관 또는 단체의 윤리원칙 준수 노력을 알리는 동시에 기관이나 단체가 수행한 업무와 의사결정 과정을 이해관계자에게 투명하게 공유하고, 필요한 피드백을 수용하여 업무 수행 방식의 개선에 기여할 수 있다는 점에서 중요한 의미를 가진다. 또한 보고서 작성 과정에서 기관이나 단체는 그간의 활동을 스스로 점검하고 평가하면서 그간의 이슈와 해결책을 진단하고 대응할 역량을 제고하는 데 이바지할 수 있다.

다만, 회의록에 포함된 내용이 업무상 비밀이나 핵심 기술과 관련되어 외부에 공개하기 어려운 경우, 위원회가 심의·의결한 사안을 공개하는 것만으로는 그 과정에서 고려된 요소와 맥락을 충분히 설명하기에 부족할 수 있다. 이 경우, 보고서에는 회의록을 공개하지 않더라도, 안건을 심의·의결한 배경, 필요성 및 중요성을 가능한 한 상세히 기록하여 이해관계자들의 이해와 공감을 끌어내는 데 도움을 줄 수 있다.

한편, 보고서의 내용과 구성은 기관이나 단체가 자율적으로 결정할 수 있으며, 위원회의 업무 수행을 중심으로 작성하되 보고서의 구성도 「인공지능기본법」 제28조 제2항 각 호의 업무를 참고하여 작성하는 방안을 고려해 볼 수 있다. 보고서 내용은 단순히 모든 항목을 나열하기보다는, 해당 시점에서 기관이나 단체, 이해관계자들의 요구가 높고 중

요하다고 판단되는 내용 위주로 선별하여 정리하는 것이 효과적이다. 이 경우 '중요도'의 평가는 이해관계자의 관심사, 해당 기관 또는 단체와 유사한 분야 및 영역의 당시 주된 이슈나 과제, 법령이나 규제의 변화, 기관 또는 단체에 발생할 수 있는 위험이나 영향 및 그 지속성, 긴급성, 구체성 등을 종합적으로 고려하여 결정할 수 있다.

또한 보고서에는 위원회가 심의·의결한 내용뿐만 아니라, 국내외 유사 사례와 관련 정보 등 다양한 참고 자료를 함께 기술함으로써 이해관계자들이 보고서 내용을 보다 깊이 이해할 수 있도록 하는 것이 바람직하다. 이를 통해 보고서는 기관이나 단체의 윤리원칙 준수 활동을 투명하게 보여주는 동시에, 내부적으로는 업무 수행 역량을 강화하고 외부적으로는 이해관계자의 신뢰와 공감을 얻는 중요한 수단이 될 수 있다.

2) 윤리원칙 준수 보고서의 공개

위원회는 인공지능 윤리원칙 준수 보고서를 공개함으로써 해당 기관 또는 단체의 인공지능 연구, 개발 및 활용 전반에 대한 신뢰성을 제고하는 데 기여할 수 있다. 보고서는 가급적 해당 기관 또는 단체의 공식 웹사이트를 통하여 공개함으로써, 이해관계자를 포함한 다양한 주체들에게 인공지능 윤리원칙 준수를 위한 노력과 관련 활동 내용을 효과적으로 안내하고 전달하는 것이 바람직하다.

아울러 인공지능의 연구, 개발 및 활용에 직접적인 영향을 받는 이해관계자에 대해서는 보고서 작성 및 발표 후에 설명회를 개최하거나, 관련 문의에 답변할 수 있는 전담 소통 창구를 마련하는 방안도 함께 고려될 수 있다.

제 4 장 인공지능 윤리정책 포럼

제 1 절 인공지능 윤리정책 포럼 구성

제4기 포럼의 위원 구성은 정책 자문 기능, 현장성, 국제 정합성, 공공성을 동시에 고려해서 구성하였다. 제4기 인공지능 윤리정책 포럼은 인공지능 윤리와 신뢰성에 관한 정책 논의를 폭넓고 입체적으로 수행하기 위해 다분야·다주체 참여 원칙에 따라 구성되었다. 위원 구성은 학계, 산업계, 법조계, 시민사회, 국제기구, 공공부문 등 인공지능 생태계의 주요 이해관계자를 균형 있게 포함하는 것을 기본 방향으로 삼고 있다.

학계 위원은 행정학, 철학, 윤리교육, 과학기술 정책 등 인공지능 윤리의 이론적·규범적 기반을 제공할 수 있는 분야의 전문가들로 구성되었다. 이를 통해 인공지능 윤리 논의가 기술적 논점에 국한되지 않고, 공공성, 가치, 교육, 사회 제도 전반으로 확장될 수 있도록 하였다.

산업계 위원은 대규모 플랫폼 기업, 글로벌 기술 기업, 국내 인공지능 전문기업, 스타트업 등 다양한 산업 스펙트럼을 대표하는 인사들로 구성되었다. 이는 실제 인공지능 개발·활용 현장에서 발생하는 윤리적 쟁점과 기업의 자율적 대응 경험을 정책 논의에 반영하기 위한 취지로 해석된다.

법조계 위원은 개인정보 보호, 책임성, 규제 설계 등 법·제도적 쟁점을 다루기 위해 다수의 실무 경험을 가진 전문가들로 구성되었으며, 시민단체 위원은 소비자 보호와 이용자 권익의 관점에서 인공지능의 사회적 영향을 점검하는 역할을 담당하도록 배치되었다. 또한 국제기구 소속 위원을 포함함으로써 유네스코 등 국제 윤리 규범 및 글로벌 논의 동향과의 연계성을 확보하고자 하였다.

공공부문 위원은 인공지능 정책 연구, 표준화, 안전, 교육 등 정책 집행 및 지원 기능을 수행하는 기관 소속 전문가들로 구성되어, 포럼 논의가 실제 정책으로 연계될 수 있는 제도적 연결고리를 강화하는 역할을 하도록 구성하였다.

〈표 4-1〉 제4기 인공지능 윤리정책 포럼위원 명단

구분	성명	소속
위원장(1인)	문명재	연세대학교 행정학과 교수
학계(3인)	박성필	한국과학기술원 문술미래전략대학원장
	변순용	서울교육대학교 윤리교육과 교수
	이상욱	한양대학교 철학과 및 인공지능학과 교수
산업계(8인)	김동환	포티투마루 대표
	김민성	한국IBM 상무
	김유철	LG AI연구원 전략부문장
	문병순	카카오모빌리티 미래플랫폼 경제연구소장
	박선민	구글코리아 상무
	송대섭	네이버 이사
	윤희식	한국마이크로소프트 이사
	이영복	제네시스랩 대표
법조계(3인)	김도엽	김앤장 법률사무소 변호사
	손지원	법무법인 혁신 변호사
	이동익	법무법인 선운 변호사
시민단체(2인)	이정수	한국소비자단체협의회 사무총장
	정지연	한국소비자연맹 사무총장
국제기구(1인)	김은영	유네스코한국위원회 유네스코의제정책센터장
공공(5인)	김명주	인공지능안전연구소(AISI) 소장
	문정욱	정보통신정책연구원 디지털사회전략연구실장
	신준호	한국정보통신기술협회 AIX혁신단장
	안성원	소프트웨어정책연구소 AI정책연구실장
	이현숙	한국과학창의재단 SW·AI 팀장

자료: 연구진 작성

다만 향후 포럼이 인공지능의 사회적 영향 전반을 다루는 공론장으로서 기능하기 위해서는 시민사회 참여의 확대, 취약계층 관점의 구조적 반영, 이해 상충 관리 장치 강화 등을 통해 포용성과 신뢰성을 한 단계 더 제고하는 방향으로의 발전이 필요할 것으로 판단된다. 인공지능 윤리정책 포럼의 지속적 발전을 위해서는 다음의 한계점을 유의하여 향후 포럼을 설계할 필요가 있다.

첫째, 시민사회 참여의 비중은 여전히 제한적이라는 한계가 존재한다. 시민단체 위원이 포함되기는 하였으나, 전체 구성에서 차지하는 비중은 상대적으로 낮아 산업계·공공부문 중심의 논의 구조를 완전히 보완하기에는 한계가 있다. 향후에는 이용자 단체, 디지털 권리 단체, 청년·취약계층을 대변하는 시민사회 주체의 참여 확대를 검토할 필요가 있다.

둘째, 노동·지역·취약계층 관점의 직접적 대표성은 부족하다. 인공지능 윤리가 노동시장 변화, 지역 격차, 사회적 약자에 미치는 영향을 주요 의제로 다루고 있음에도 불구하고, 해당 집단을 직접적으로 대표하는 주체의 참여는 제한적이다. 이는 논의의 포괄성과 현장성을 제약하는 요소로 작용할 수 있다.

셋째, 산업계 비중이 상대적으로 높은 구조는 정책 자문 과정에서 이해 상충 가능성에 대한 지속적인 관리 필요성을 제기한다. 이는 부정적 요소라기보다는, 포럼 운영 과정에서 이해 상충 관리 원칙, 투명한 논의 구조, 공개성 강화 등을 통해 보완되어야 할 운영상의 과제로 볼 수 있다.

제2절 제4기 인공지능 윤리정책 포럼 운영

본 절에서는 제1·2·3차 포럼의 개최 결과를 정리하고자 한다. 출범식과 함께 이루어진 제1차 포럼에서는 정보통신정책연구원의 인공지능 윤리 사업 추진 경과와 2025년 인공지능 윤리정책 포럼의 운영 방안을 발표하였다. 또한, 「인공지능기본법」 시대의 인공지능 윤리 거버넌스에 대한 기조 강연이 진행된 이후 인공지능 격차 완화와 인공지능 윤리 기술 발전 및 기업 지원 등에 대한 토론을 진행하였다. 제2차 포럼은 유네스코의 인공지능 윤리 이행 동향과 시사점, 최신 법적 사례를 중심으로 인공지능의 윤리적 활용을 고찰하는 발표를 진행하고, 소외계층의 인공지능 리터러시 및 교육, 저작권 보호 및 데이터 거버넌스 등에 대해 논의했다. 제3차 포럼은 '2025 인공지능 윤리 공개 세미나'의 종합토론으로 갈음되었다. 해당 토론은 포럼 위원장과 위원들이 참여하여 「인공지능기본법」 시행과 인공지능 윤리정책 아젠다 전환 및 사회적 과제에 대해 심도 있는 의견을 개진하였다.

〈표 4-2〉 제4기 인공지능 윤리정책 포럼 개최 현황

회차	개최일	주요 논의사항
제1차	7월 31일	- 인공지능 윤리 사업 추진 경과 및 제4기 포럼 운영 방안 발표 - 「인공지능기본법」 시대의 인공지능 윤리 거버넌스 발표 - '모두의 AI' 실현을 위한 인공지능 격차 완화, 인공지능 윤리 기술 발전 및 기업 지원 등에 관한 토론
제2차	9월 26일	- 유네스코 인공지능 윤리 이행 동향과 시사점 - AI for Everyone, AI on Trial - 소외계층 관련 인공지능 리터러시와 교육, 저작권 보호 및 데이터 거버넌스 등에 관한 토론
제3차	11월 27일	- 포스트 「인공지능기본법」 시대, 인공지능 윤리정책의 아젠다 전환에 관한 토론 - 기업의 인공지능 윤리원칙 현장 적용 애로사항과 제도적 해소 방안 논의

자료: 연구진 작성

1. 제1차 포럼 개최 결과

인공지능 기술 주권 확보와 책임 있는 인공지능 윤리 기반 마련을 위해 2025년 7월 31일 '제4기 인공지능 윤리정책 포럼'이 공식 출범되었다. 이날 행사에서 연세대 문명재 교수가 포럼위원장으로 위촉되었으며 학계, 산업계, 법조계 등 각 분야 전문가 23명이 참여하여 「인공지능기본법」과 정부의 소버린 인공지능 전략과 연계하여 인공지능 윤리 실천 방안 수립, 국민 체감형 정책 설계, 인공지능 3대 강국 도약을 위한 정책 과제 발굴 등을 집중적으로 논의할 예정이다.

이번 포럼에서는 정보통신정책연구원 문정욱 실장이 연구원에서 수행한 인공지능 윤리정책 과제의 추진 경과와 주요 내용을 발표하며 2024년 연구 성과를 공개하였다. 이어 정보통신정책연구원 이현경 연구위원이 제4기 인공지능 윤리정책 포럼의 추진 배경, 구성, 논의 주제 등을 포함한 운영 계획을 공유하고, 대내외 환경 변화를 분석하며 포럼에서 논의가 필요한 사항들을 제시하였다. 이후 문명재 포럼위원장이 「인공지능기본법」 시대의 인공지능 윤리 거버넌스를 주제로 기조 강연을 진행하며, 인공지능 발전에 따른 인공지능 윤리 중요성 확대와 다양한 배경을 가진 참여자들의 균형 잡힌 논의를 통한 포용적 거버넌스의 필요성을 강조했다. 마지막으로 인공지능 격차 완화와 인공지능 윤

리 기술 발전 및 기업 지원 등과 관련하여 포럼 위원 전원이 참여하는 자유 토론이 진행되었다.

〈표 4-3〉 제1차 인공지능 윤리정책 포럼 개최 개요

- 일시: 2025.7.31.(목), 오후 2시
- 장소: 포스트타워 스카이홀
- 참석: 인공지능 윤리정책 포럼위원장 및 위원, 과기정통부 관계자 등 20인 내외
- 내용: 포럼 위원 위촉, 인공지능 윤리 사업 추진 경과 및 제4기 포럼 운영 방안 발표, 「인공지능기본법」 시대의 인공지능 윤리 거버넌스 발표 및 '모두의 AI' 실현을 위한 인공지능 격차 완화, 인공지능 윤리 기술 발전 및 기업 지원 등에 관한 종합토론

가. 주요 발표내용

1) '인공지능 윤리 사업 추진 경과 및 포럼 운영방안' 발표내용

정보통신정책연구원 디지털사회전략연구실 문정욱 실장과 이현경 연구위원은 정보통신정책연구원에서 실시하는 인공지능 윤리정책 관련 주요 과제의 추진 경과 및 포럼 운영 방안을 주제로 발표를 진행하였다. 주요 발표 내용은 다음과 같다.

〔그림 4-1〕 '인공지능 윤리 사업 소개' 발표

자료: 정보통신정책연구원, 인공지능 윤리 사업 소개, 2025.7.31.

정보통신정책연구원은 2020년 12월 ‘사람이 중심이 되는 「인공지능(AI) 윤리기준」’을 마련한 이후, 인공지능 윤리 실천과 확산을 위한 정책적 노력을 지속적으로 추진해 왔다. 2025년 1월 「인공지능기본법」이 제정됨에 따라, 기존의 인공지능 윤리 실천 중심 정책 기조를 계승하는 동시에 이를 더욱 체계적으로 정비한 인공지능 윤리정책을 본격적으로 추진할 계획이다. 「인공지능기본법」이 안전한 인공지능의 개발·활용을 위한 최소한의 법적 기준을 제시한다면, 인공지능 윤리는 그보다 넓은 범위에서 사회적 책임과 가치 판단을 요구하는 역할을 수행한다. 이는 법적 규제가 미치기 어려운 영역에서 인공지능의 신뢰성과 책임성을 제고하고 규제 공백을 보완하는 실질적인 수단으로 기능한다. 이러한 점에서 「인공지능기본법」 시행 이후에도 인공지능 윤리의 중요성은 더욱 강조될 전망이며, 정보통신정책연구원은 법에 명시된 윤리 원칙과 실천 조항을 토대로 제도적 지원체계를 강화하고, 기업이 책임감을 가지고 인공지능을 개발·활용할 수 있는 환경 조성을 위해 지속적으로 노력해 나갈 예정이다.

정보통신정책연구원은 인공지능 윤리의 실질적 이행을 위해 다양한 정책 수단을 추진하고 있다. 우선 인공지능 서비스의 잠재적 긍·부정 영향력을 사전에 진단하고 이를 체계적으로 관리하기 위해 인공지능 윤리 영향평가를 시행 중이다. 2023년 프레임워크의 구축을 시작으로 2024년에는 딥페이크 등 영상합성 서비스를 대상으로 시범 평가를 실시하였으며, 2025년에는 인공지능 채용 서비스를 대상으로 공정성과 투명성을 중점적으로 검증할 계획이다. 또한 기업이 인공지능 윤리기준 준수 여부를 스스로 점검할 수 있도록 산업 현장의 특수성을 반영한 분야별 자율점검표를 지속적으로 개발·보급하고 있다. 2022년 챗봇·작문·영상 분야를 시작으로 2023년 채용, 2024년 영상합성 서비스 분야를 대상으로 자율점검표의 현장 적용을 지원하며 분야별 점검표를 개발했다. 2025년에는 헬스케어 분야로 영역을 확대한다. 이와 함께 국민의 인공지능 윤리 리터러시 제고를 위한 교육을 강화하고 있다. 초·중·고교생과 일반인, 전문 개발자에 이르기까지 대상별 수준과 역할에 맞춘 교육 콘텐츠를 개발, 배포하고 있다. 아울러 2020년부터 축적된 연구 성과와 국내외 인공지능 정책 동향 등을 체계적으로 관리하는 통합 플랫폼인 인공지능 윤리 소통채널을 운영 중이다. 이날 포럼에서는 그간의 연구 결과를 담은 보고서, ‘AI 산업계 실무자를 위한 AI 윤리’ 교육 콘텐츠의 인쇄본과 사업 관련 자료를 QR 코드 형태로 포럼 위원들에게 공개하였다.

〈표 4-4〉 인공지능 윤리사업 주요 내용

구분	세부 활동 내용
인공지능 윤리 영향평가 시행	- 인공지능 기반 서비스의 윤리적 영향력 식별 및 합리적 관리를 돕기 위한 인공지능 윤리영향평가 시행('24~)
인공지능 윤리기준 자율점검표	- 인공지능 윤리 규범 실천을 위한 구체적 방안으로서 인공지능 윤리기준 준수 여부를 스스로 점검할 수 있는 공통 및 분야별 자율점검표 개발('22~)
인공지능 윤리교육 콘텐츠	- 사회 경제, 기술적 변화와 함께 새롭게 제기되는 윤리적 이슈를 논의하고 사회 구성원의 인공지능 리터러시 함양을 위해 인공지능 윤리교육 콘텐츠 개발('21~)
인공지능 산업계 실무자 대상 인공지능 윤리교육 콘텐츠	- 인공지능 산업계 실무자의 역량 발전을 위해 인공지능 생애주기(AI Life Cycle)에 따른 인공지능 윤리교육 콘텐츠 개발
인공지능 윤리 소통채널	- 인공지능 윤리의 사회 정착과 신뢰성 향상을 지원하고, 사회 구성원과 공유 및 소통을 위한 채널 구축 및 운영('23~)
인공지능 윤리정책 포럼	- 기술 중심의 인공지능 발전 흐름 속에서 윤리와 신뢰의 원칙을 기반으로 한 정책 거버넌스 구축 및 지속 운영('22~)

자료: '정보통신정책연구원, 인공지능 윤리 사업 소개, 2025.7.31.' 내용을 토대로 연구진 재구성

〔그림 4-2〕 '제4기 인공지능 윤리정책 포럼' 발표

자료: 정보통신정책연구원, 제4기 인공지능 윤리정책 포럼, 2025.7.31.

다음으로 '제4기 인공지능 윤리정책 포럼'의 간사를 맡은 정보통신정책연구원 이현경 연구위원이 제4기 포럼의 운영계획(안)과 대내외 환경 및 주요 논의에 대한 발표를 진행하며, 포럼이 나아가야 할 구체적인 비전을 제시하였다. 자세한 내용은 아래와 같다.

먼저 제4기 인공지능 윤리정책 포럼의 추진 배경과 목적은 현 정부의 ‘소버린 AI’ 구축 및 ‘AI 3대 강국 도약’이라는 국가 전략과 연계하여 공공의 신뢰를 확보하고 윤리적 활용 기반을 마련하는 데 있다. 특히 「인공지능기본법」의 제정은 윤리 논의를 축소하는 계기가 아니라, 오히려 인공지능 윤리를 사회 전반에 확산시키는 법적·정책적 기반이 되었다고 분석한다. 이에 따라 포럼은 학계, 산업계, 법조계, 시민사회, 국제기구, 공공 등 다양한 이해관계자들이 참여하는 사회적 논의의 구심점으로서, 인공지능 기술 발전 흐름 속에서 윤리적 기반과 공공 신뢰를 동시에 확보할 수 있는 방안을 모색할 예정임을 밝혔다.

이어지는 국내외 현황 발표에서는 최근 급변하는 정책 환경에 주목하였다. 국내의 경우 국가인공지능위원회의 설치 및 운영 규정의 개정을 통해 범부처 차원의 전략 조율과 심의·의결 기능이 대폭 강화되었다. 국제적 환경 역시 역동적으로 변화하고 있다. G7은 인공지능 네트워크 구축을 제안하며 국제 공조를 강화하고 있으며, 유네스코는 이행 준비도 평가 방법론과 글로벌 감독 기관 네트워크를 제시하며 시민사회와 학계의 참여를 독려하고 있다. EU는 투명성과 저작권, 안전을 핵심으로 하는 ‘인공지능 실행 규범(Code of Practice)’ 등을 발표하여 실무적 기준을 선제적으로 제시했다. 이에 대응하여 미국은 자국 중심주의와 혁신 가속화를 골자로 한 ‘미국 인공지능 행동계획(America’s AI Action Plan)’을 통해 전방위적 규제 철폐 의지를 드러냈으며, 중국 또한 주도권 경쟁에 적극적으로 참여하고 있다.

마지막으로 향후 포럼의 역할에 대해서는 국제 기술 패권 경쟁과 규제 및 혁신 사이의 균형점을 찾고 한국의 인공지능 윤리 정책 방향을 제시하는 데 집중하겠다고 밝혔다. 아울러 학계, 산업계, 시민사회의 의견을 통합하여 실무적이고 수용성 높은 윤리 가이드라인을 도출할 계획임을 알렸다. 제4기 인공지능 윤리정책 포럼은 기술 패권 경쟁과 국제 기준 정립이 교차하는 시점에서 위원들의 전문성을 바탕으로 한국이 국제 인공지능 윤리 거버넌스에서 선도적인 역할을 수행할 수 있도록 다각적인 정책 대안을 모색하고자 하며, 이에 대한 포럼 위원들의 적극적인 참여를 당부하였다.

2) 「AI 기본법」 시대의 인공지능 윤리 거버넌스' 발표내용

문명재 연세대학교 교수는 제4기 인공지능 윤리정책 포럼 위원장으로서 'AI 기본법 시대의 인공지능 윤리 거버넌스'라는 제목으로 기조 강연을 진행하며, 인공지능 기술의 발전이 가져올 기회와 위기를 동시에 조망하였다. 본 강연에서는 인공지능의 혜택과 위험 간의 균형을 핵심으로 국제 규제 동향, 인공지능 윤리 트렌드 변화와 향후 과제를 종합적으로 제시하였다. 주요 내용은 다음과 같다.

〔그림 4-3〕 'AI 기본법 시대의 AI 윤리 거버넌스' 기조 강연

자료: 문명재, AI 기본법 시대의 AI 윤리 거버넌스, 2025.7.31.

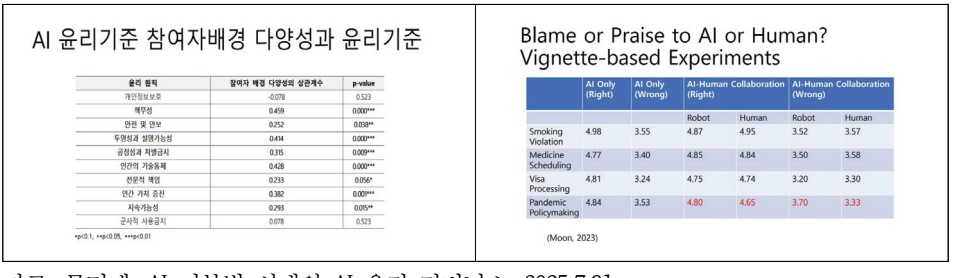
인공지능 기술은 인류에게 막대한 혜택을 제공할 잠재력을 지닌 동시에, 가장 큰 위험을 초래할 수 있는 기술로 평가되고 있다. '스탠퍼드 인간 중심 인공지능 연구소(HAI)'의 보고서에 의하면 딥페이크를 포함한 인공지능 관련 사고는 매년 급격히 증가하고 있으며, 이는 기술 발전의 속도를 사회적 대응이 따라가지 못하고 있음을 보여준다. 실제로 중국의 대입 시험 기간 중 인공지능 채팅 서비스 차단, 학술 논문 심사 과정에서의 인공지능 조작 사건, 생성형 인공지능의 가짜 정보 생성 및 환각 문제 등은 과거에 상상하기 힘들었던 부작용이 이미 현실화 되고 있음을 시사한다.

이러한 위험 속에서 국제 인공지능 정책의 패러다임은 과거 선언적 담론 중심의 윤리 논의를 넘어 실질적 집행력을 갖춘 규제 법제화 단계로 진입하고 있다. 미국, EU, 중국 등 강대국들이 자국의 경쟁력과 직결된 핵심 자산으로 인식하고 있으며, 자국의 이익 보호와 국제적 우위 선점을 위해 인공지능 정책을 적극 활용하고 있다. 우리나라 또한 'AI 3대 강국 도약'을 목표로 다양한 정책을 추진 중이다. 이러한 국제 정세에 따라 인

공지능 윤리 정책의 흐름 또한 변화하고 있다. 2018년 '아실로마 원칙'을 기점으로 정점에 달했던 자율적 가이드라인 중심의 접근은 점차 둔화되는 반면, EU의 인공지능법이나 한국의 「인공지능기본법」처럼 구체적인 법제화와 과징금 규범 등 강제성을 띤 정책은 증가 추세를 보인다.

국내외 다양한 인공지능 윤리 원칙을 종합적으로 분석한 결과, 공통적으로 강조되는 핵심 가치로 인간 존엄성, 사회 공공성, 기술의 합목적성을 확인할 수 있었다. 또한, 이를 뒷받침하는 핵심 키워드로는 투명성, 설명가능성, 책무성, 공정성이 가장 빈번하게 언급되었다. 아울러 현재의 인공지능 거버넌스 구조는 정부와 산업계가 주도하는 형태이지만, 시민사회와 윤리 전문가의 참여가 확대될수록 책임성과 안전성과 같은 주요 가치가 더욱 균형 있게 강조되는 경향이 나타나는 것으로 분석되었다.

〔그림 4-4〕 'AI 기본법 시대의 AI 윤리 거버넌스' 기조 강연



자료: 문명재, AI 기본법 시대의 AI 윤리 거버넌스, 2025.7.31.

한편, 인공지능이 공공 서비스나 정책 결정 과정에 활용될수록 인간과 인공지능의 협업 방식과 그에 따른 책임 귀속 문제가 중요한 쟁점으로 부상하고 있다. 관련 연구에 따르면 업무 성과가 긍정적으로 나타날 경우에는 그 공이 인공지능 시스템에 귀속되는 경향이 강한 반면, 오류나 문제가 발생하면 이를 활용한 담당자에게 책임이 집중되는 책임의 비대칭성이 확인된다. 이러한 현상은 향후 공공부문을 비롯한 다양한 영역에서 인공지능을 도입할 때 책임 주체와 책임 범위를 어떻게 설정할 것인지에 대한 중요한 정책적 과제를 제기한다.

향후에는 포괄적인 인공지능 윤리원칙을 넘어 각 산업 분야의 특수성을 고려한 맞춤형 윤리기준과 구체적 실행 가이드라인, 역량 평가 체계를 구축하는 것이 필요하다. 더

나아가 산업계의 혁신 요구와 시민사회의 불안과 우려 사이에서 적절한 균형을 이루는 포용적 인공지능 거버넌스의 확립이 필수적이다. 궁극적으로 인공지능과 사회의 균형을 위해서는 급변하는 기술 환경 속에서 고정된 해답을 찾기보다는 지속적인 성찰과 조정을 통해 윤리적 균형을 끊임없이 모색해 나가는 노력이 요구된다.

나. 주요 토론내용

이어진 순서에서는 '민간에서의 인공지능 윤리 거버넌스'를 비롯한 인공지능 윤리 정책에 대한 위원 간 자유토론이 진행되었으며, '모두의 AI 실현을 위한 인공지능 격차 완화', '인공지능 윤리 기술 발전 및 기업 지원' 등 다양한 의견이 개진되었다.

〔그림 4-5〕 제1차 포럼 종합토론

먼저 '모두의 AI'를 실현하기 위해 디지털 격차의 사회적 해결 방안 마련이 필요하며, 디지털 포용에 대한 논의를 확대하고, 인공지능 산업 진흥과 더불어 소외자 대상 정책이 병행되어야 한다고 의견을 모았다. 구체적으로 공정거래위원회와 한국소비자원에서 실시한 소비자 인식조사 결과 중, 인공지능에 대한 높은 기대치와 긍정적 인식과 동시에 딥페이크, 프라이버시 보호, 차별, 의인화 등에 대한 높은 우려를 공유하며, 인공지능 소비자 간의 격차가 존재하고 인공지능 기술의 상용화에 따른 소비자 불안감 또한 고조됨을 언급하며 소비자 차원에서의 의견 수용 및 소비자 행동 교육과 인식 전환에 대한 고려가 필요하다는 제언이 나왔다. 또한 인공지능으로 인한 일자리 전환에 대응하여 인공지능 시대에 적합한 직무 재설계, 직업 교육 및 재교육 등을 통한 격차 완화의 필요성도 언급되었다.

인공지능 윤리 기술 발전 및 기업 지원에 대한 구체적인 논의 또한 활발하게 이루어졌다. 인공지능 분야는 기존의 규제기관 처분 및 과징금과 같은 접근보다 기업의 자발적인 프레임워크를 유도하는 것이 관건임을 지적하며 정부에서 윤리 거버넌스 및 프레임워크 예시를 제시하고 관련된 인센티브를 제공하는 방식이 제안되었다. 특히 인공지능 분야는 법적인 규제보다는 이슈가 되는 사항을 중심으로 윤리적 접근을 통해 우선 대응하는 방안이 적절하며, 이를 위해 실제 우수사례 및 기술·서비스 개발·이용자 경험에 대한 공유가 필요하다고 강조하였다. 인공지능의 역기능 방지 및 예방을 위해 가장 빠르게 조치할 수 있는 수단이 인공지능 윤리라는 점에 다시 한번 의견을 모았으며, 향후 'Agentic AI', 'Physical AI'로의 전환에 대응하여 관련된 역기능에 대한 선제적 논의를 진행하고 이를 통해 참고 사항을 제시해야 한다는 제언이 나왔다. 그 외에도 초신뢰·초지능 인공지능(Super safe & Intelligent AI)이 각광받는 상황에서 이를 위한 대규모 언어모델(Large Language Model, 이하 LLM)의 편향성 사전 완화 및 통제성 강화 등의 인공지능 윤리 기술 연구가 활발함에 따라 관련된 논의의 필요성 또한 제기되었다.

이외에도 유네스코가 주재하는 감독기관 네트워크 등 관련 국제기구 혹은 국가적 논의 및 활동에 적극적인 참여 및 주도 노력의 필요성과 기술 발전으로 인해 변화한 현 상황과 미래를 고려하여 윤리기준 및 원칙 등의 지속적인 개선 등이 제안되었다. 또한, 「인공지능기본법」에서 '고영향 인공지능' 개념에 대한 해석의 여지가 다양하여 아직은 법적 안정성이 부족한 측면이 존재함에 따라, 기업의 자율성을 보장하는 동시에 이를 구체화하는 방안 등 기존의 윤리기준에 국한하지 않는 입법 정책 논의로의 포럼 역할 확장이 제안되었다.

2. 제2차 포럼 개최 결과

'제4기 인공지능 윤리정책 포럼' 2차 회의가 2025년 9월 26일(금) 명동 포스트타워 스카이홀에서 개최되었다. 이번 회의에서는 인공지능 리터러시 교육의 방향성을 논의하며, 기존의 소외계층 대상 디지털 교육을 넘어, 시민이 공정하게 사회에 참여할 수 있는 실질적 역량을 함양하는 정책 대안으로 설계할 필요성을 검토하였다.

회의에서는 유네스코의 '인공지능 윤리 권고(2021)' 이행 동향과 시사점, 최신 법적 분

쟁 사례를 통해 본 인공지능의 윤리적 활용에 대한 주제 발표가 각각 진행되었다. 이어진 자유토론에서는 참석 위원 전원이 참여하여 소외계층의 인공지능 리터러시 및 교육, 저작권 보호 및 데이터 거버넌스 등의 정책 현안에 대한 심도 있는 의견을 교환하였다.

〈표 4-5〉 제2차 인공지능 윤리정책 포럼 개최 개요

- 일시: 2025.9.26.(금), 오후 2시
- 장소: 포스트타워 스카이홀
- 참석: 인공지능 윤리정책 포럼위원장 및 위원, 과기정통부 관계자 등 20인 내외
- 내용: 유네스코 인공지능 윤리 권고 이행 동향과 시사점 발표, 법적 분쟁 사례를 통한 인공지능의 윤리적 활용 방안 발표 및 소외계층 관련 인공지능 리터러시와 교육, 저작권 보호 및 데이터 거버넌스 등에 관한 종합토론

가. 주요 발표내용

1) '유네스코 AI 윤리 이행 동향과 시사점' 발표내용

유네스코 한국위원회 의제정책센터 임시연 선임전문관은 '유네스코 AI 윤리 이행 동향과 시사점'을 주제로 발표를 진행하였다. 해당 발표는 유네스코의 인공지능 윤리 권고와 2025 글로벌 포럼의 성과를 중심으로 '인간 중심의 인공지능' 확산을 위한 유네스코 및 국제 거버넌스 체계를 분석하고, 향후 우리나라가 견지해야 할 전략적 역할과 구체적인 대응 방향을 제안하였다. 주요 내용은 아래와 같다.

〔그림 4-6〕 ‘유네스코 AI 윤리 이행 동향과 시사점’ 발표

자료: 임시연, 유네스코 AI 윤리 이행 동향과 시사점, 2025.9.26.

제41차 유네스코 총회는 인공지능 알고리즘 편향성, 프라이버시 침해, 정보 왜곡 등 윤리적 위험에 대응하기 위해 193개국 만장일치로 ‘인공지능 윤리 권고(Recommendation on the Ethics of AI)’를 채택하였다. 이는 인권과 존엄성, 비차별을 핵심 원칙으로 하는 최초의 국제 규범으로, 유네스코는 해당 권고가 선언에 그치지 않고 각국 정책에 실질적으로 내재화될 수 있도록 준비도 평가 방법론(Readiness Assessment Methodology, 이하 RAM)과 윤리적 영향 평가(Ethical Impact Assessment, 이하 EIA) 등의 구체적인 실행 체계를 운용하고 있다. 먼저 RAM은 국가별 법·제도·문화적 인프라를 종합 진단하여 정책 역량 강화를 지원하는 도구로 2025년 기준 70개국 이상이 도입하였다. 특히 아세안 국가들은 평가 결과를 국가 디지털 전략에 반영하는 등 실효성 있는 정책 수단으로 활용되고 있다. EIA는 인공지능 시스템 도입 시 발생할 수 있는 집단 배제, 환경적 부담 등 잠재적 위험을 사전에 평가하는 제도적 장치이다. 이는 단순한 기술 성능 평가를 넘어 인권 보장과 사회적 책임을 제도화하는 핵심 수단으로 평가되며 많은 국가에서 법제화되거나 주요 정책 절차에 통합되고 있다.

이러한 흐름 속에서 2025년에 개최된 제3차 인공지능 윤리 국제포럼은 인공지능 윤리를 국제적 공공재(Global Public Goods)로 확립하는 계기가 되었다. 해당 포럼을 통해 각국 규제기관 간 지식 공유 플랫폼인 ‘인공지능 감독기구 글로벌 네트워크(Global Network of AI Supervisory Authorities, GNAIS)’가 출범하였으며, 전문가 그룹인 ‘국경 없는 전문가 네트워크(AI Ethics Experts Without Borders, AIEB)’를 100명 규모로 확대하고 시민사회와 학계의 참여를 제도화함으로써 다자간 거버넌스의 기틀을 다졌다. 또한 교

육, 언론, 성평등, 취약계층, 환경, 사법 등 다양한 분야별 아젠다에 대한 합의를 도출하며 규범적 영향력을 넓혔다. 유네스코의 인공지능 권고 및 인공지능 윤리 관련 협력 현황은 아래의 <표 4-6>에서 확인할 수 있다.

〈표 4-6〉 유네스코 인공지능 권고 및 AI 윤리 관련 협력 현황

구분	세부 활동 내용
유네스코 인공지능 권고	- 제41차 유네스코 총회에서 193개국 만장일치 채택('21) 인공지능 윤리 분야 최초의 글로벌 규범 문서로서 인공지능의 잠재적 위험 예방과 인류 공동 가치 추구를 위한 포괄적인 가이드라인
준비도 평가, RAM	- 인공지능 준비 방법론(Readiness Assessment Methodology, RAM)은 국가별 법·정책·인프라·교육·문화 등을 종합적으로 진단하며 '25년 기준 70개국 이상에서 사용 중으로 감사보다 접근성과 참여성이 높은 도구라고 평가됨
윤리 영향평가, EIA	- 유네스코의 윤리영향평가(Ethical Impact Assessment)는 인공지능 시스템 등의 사회·윤리적 영향을 사전에 분석하여 불평등 심화, 집단 배제, 환경 부담 등 부정적 결과 예방 등 사회적 책임과 인권 보장을 제도화하는 핵심 수단으로 평가되며 여러 국가에서 법제화 혹은 정책 절차에 포함 중임
분야별 윤리 아젠다	- 교육, 언어 다양성, 성평등, 환경, 과학, 청년·아동, 장애인 권리, 언론·저널리즘 등의 분야를 중점적으로 다루며 차세대 기술인 신경기술과 양자기술 분야로 인공지능 윤리 거버넌스를 확장 중임
인공지능 윤리 국제포럼	- 제3차 유네스코 인공지능 국제포럼('25)은 인공지능 윤리를 전세계가 공유해야 할 공공재로 공식화하고 다양한 지역의 균형적 대표성을 강화하였으며, 인공지능 감독기구 글로벌 네트워크(GNAIS)와 시민사회·학계 글로벌 네트워크 출범 등의 성과를 거둠 * GNAIS: Global Network of AI Supervisory Authorities
주요 이니셔티브	- 인공지능 윤리와 리터러시를 중심으로 훈련·지침·연구·참여가 유기적으로 연결된 복합적 생태계 조성을 위해 다양한 이니셔티브를 추진 중이며 대표적으로 ▲국경 없는 전문가 네트워크(Experts Without Borders, AIEB), ▲공무원 인공지능 리터러시 훈련, ▲인공지능과 법의 지배(AI & Rule of Law), ▲글로벌 인공지능 윤리·거버넌스 관측소(Global AI Ethics and Governance Observatory, GAEGO) 등이 있으며, 이를 통해 지침 개발·교육 확산·시민 참여, 민간 협력 등이 진행 중임
국제협력	- 권고의 국제 규범화 및 표준·정책·훈련을 아우르는 실행력 제공을 위해 UN, G20, G7 등과 다자 거버넌스를 연계하고 지역 전략과 국제 담론 확산을 위해 노력하며, 국제 규범의 연결자와 현장 적용 촉진자 역할을 수행하고자 함

자료: '임시연, 유네스코 AI 윤리 이행 동향과 시사점, 2025.9.26.' 내용을 토대로 연구진 재구성

이처럼 유네스코를 중심으로 국제 인공지능 거버넌스가 실행 중심으로 전환됨에 따라, 한국은 국내 인공지능 윤리의 국제 정합성을 제고하고 국제사회에서의 인공지능 윤리 리더십을 강화할 필요가 있다. 이를 실현하기 위한 네 가지 구체적 방안은 다음과 같다. 첫째, 공공부문의 인공지능 리터러시를 강화하여야 한다. 둘째, 유네스코 등의 국제 지표를 국내 실정에 맞게 최적화함으로써 국제 규범과 국내 지표 간의 연계성을 공고히 해야 한다. 셋째, 한국의 인공지능 정책 현황과 구체적인 실증 사례를 국제사회에 적극적으로 공유해야 한다. 마지막으로 아시아·태평양 지역의 인공지능 윤리 허브 구축을 통해 지역 내 협력을 주도하고 전략적 리더십을 확대해야 한다.

결론적으로 인공지능 혁신과 규제는 상충하는 가치가 아니라 상호 보완적인 관계이며, 윤리 거버넌스는 인공지능 기술의 신뢰성과 공공성을 제고하기 위한 필수 조건이다. 한국은 이러한 국제 흐름에 발맞춰 기술적 우위를 넘어 규범적 가치를 선도함으로써 인간 중심의 인공지능 생태계를 조성하는 데 주도적인 역할을 수행해야 할 것임을 재차 강조하였다.

2) ‘AI for Everyone, AI on Trial?’ 발표내용

다음으로 뉴욕주 변호사이자 성균관대 겸임교수를 역임하고 있는 홍정순 변호사가 ‘AI for Everyone, AI on Trial? - 노동자, 예술가, 취약계층의 AI 소송 이야기’라는 제목으로 발표를 진행하였다. 발표는 먼저 사례를 바탕으로 인공지능 관련 주요 소송 유형과 현황을 살펴보고, 인공지능 시스템에 적용 가능한 규제 법률과 향후 규제 적용 시나리오를 제시하는 구성으로 이루어졌다. 주요 내용은 다음과 같다.

〔그림 4-7〕 'AI for Everyone, AI on Trial?' 발표

자료: 홍정순, AI for Everyone, AI on Trial?, 2025.9.26.

전 세계적으로 인공지능 시스템의 신뢰성과 책임에 대한 논의가 활발해지는 가운데, 노동, 심리, 저작권이라는 세 가지 핵심 분야를 중심으로 주요 소송 사례를 검토하였다. 첫째, 노동 분야에서는 '딜리버루(Deliveroo)'와 '우버(Uber)'의 사례처럼 인공지능 알고리즘의 편향적이고 불공정한 결정으로 인한 노동자 차별, 통제 및 감시 가능성, 권리 침해를 초래할 수 있는지에 대한 공방이 이어졌다. 이에 대해 법원은 인공지능의 결정이 인간의 기본권을 침해한다면 위법이며, 기업이 해당 알고리즘의 정당성을 입증해야 한다고 판단하는 추세이다. 둘째, 심리적 측면에서는 '캐릭터.AI(Character.AI)' 소송 사례와 같이 생성형 인공지능 챗봇에 대한 과의존, 폭력 행위 조장, 아동 대상 성폭력 및 성착취, 유해 콘텐츠 노출 문제가 주로 제기되고 있다. 현재 관련 소송이 진행 중이나, 안전 조치가 미흡한 경우 제조사에 책임이 있으며, 아동과 같은 취약계층을 위한 엄격한 보호 체계 마련이 필수적이라는 사회적 목소리가 강화되고 있다. 마지막으로 저작권 분야에서는 '뉴욕타임스(The New York Times)'와 '디즈니(Disney)' 등이 인공지능 학습 데이터 및 그 결과물의 저작권 침해 여부를 두고 소송을 진행하고 있다. 이에 대해 법조계는 적법하게 확보된 데이터의 학습은 공정이용으로 볼 여지가 있으나, 인공지능의 결과물이 원작자의 시장 가치를 훼손하거나 대체한다면 저작권 침해로 간주될 가능성이 있다고 분석하고 있다.

다음으로 인공지능 시스템에 적용 가능한 규제 법률을 지식재산권 보호, 데이터 보호, 설명 가능하고 비차별적이며 책임성 있는 인공지능 시스템 구축을 위한 규제로 나누어 제시하였다. 먼저 인공지능 시스템 및 생성물에 대한 지식재산권 보호 측면에서는 특허

법, 저작권법, 영업비밀보호 및 공정거래법이 주요하게 작동한다. 이 영역에서는 인공지능이 생성한 결과물이 원작자의 권리를 정당하게 보호하는지, 그리고 궁극적으로 무엇을 위해 인공지능 저작권을 보호해야 하는지에 대한 근본적인 질문을 던질 필요가 있다. 다음으로 인공지능 데이터 이용에 관한 데이터 보호 규제는 각국의 데이터 관련법과 EU의 일반 개인정보 보호법(General Data Protection Regulation, 이하 GDPR)을 중심으로 이루어질 수 있다. 데이터의 불법 수집 및 이용은 개인의 인권 침해뿐만이 아닌 데이터 독점과 악용 문제로 이어질 수 있다. 따라서 이용자의 데이터를 활용하는 전 과정에서 명확한 법적 근거와 투명성이 확보되었는지 검토하는 것이 필수적이다. 마지막으로 인공지능 편향성으로 인한 사회적 차별이나 예측 불가능한 결과에 대한 책임 회피 문제 해결을 위해 EU '인공지능법(AI Act)', 미국 '알고리즘 책임법(Algorithmic Accountability Act)' 같은 수단이 마련되고 있다. 이러한 규제들이 실제 현장에서 어떻게 적용될 수 있는지 구체적인 사례를 기반으로 구성한 시나리오를 제시하였으며, 그 내용은 아래의 <표 4-7>과 같다.

〈표 4-7〉 AI 규제 법률 적용 시나리오 예시

사례 구분	주요 내용 및 법적 쟁점	관련 법률 및 규제 결과
청소년 범죄 예측	- 청소년의 환경(가족, 친구 등)을 분석해 범죄 가능성을 예측·모니터링하는 시스템	- EU AI Act: '수용 불가한 위험'으로 간주되어 금지(Art. 5) - 한국 「인공지능기본법」: '고영향 AI'로서 투명성 및 안전성 의무 부과 대상
'Clearview AI' 안면 인식	- 웹사이트에서 500억 개 이상의 얼굴 이미지를 무단 수집하여 수사기관 등에 제공	- GDPR: 정보주체 동의 없는 생체정보 수집으로 거액의 과징금 부과 - EU AI Act: 무단 수집 데이터 기반 안면 인식 DB 확장은 금지
독일의 신용 점수 분석	- 과거 기록 및 사회적 행동 기반 프로파일링을 통해 상환 가능성을 예측하는 시스템	- EU AI Act: 개인 특성에 기반한 자동화된 예측 시스템으로 금지 대상 - GDPR: 프로파일링에 대한 정보주체 거부권 및 설명요구권 인정
TikTok (플랫폼 규제)	- 어린이 개인정보 과다 수집 및 광고 타게팅의 불투명성, 기만적 광고 방치	- GDPR: 미성년자 데이터보호 위반 - DSA(디지털 서비스법): 광고 투명성 요건 및 플랫폼 악영향 최소화 의무 위반으로 제재

사례 구분	주요 내용 및 법적 쟁점	관련 법률 및 규제 결과
인공지능 탑재 드론	- 감시, 데이터 수집, 자율 주행 기능을 갖춘 드론의 운용	- EU AI Act: 고위험 AI로 분류되며, 데이터 품질 관리, 투명성 확보, 인간 감독(Human-in-the-loop) 및 엄격한 데이터 거버넌스 준수 의무 존재

자료: '홍정순, AI for Everyone, AI on Trial?, 2025.9.26.' 내용을 토대로 연구진 재구성

마지막으로 인공지능이 인류의 가능성을 확장하고 난제를 해결하는 '모두를 위한 AI'를 실현하기 위해서는 기술적 혁신을 넘어선 다각적인 노력이 선행되어야 함을 다시 한번 강조하였다. 우선 인공지능은 단순한 기계적 도구가 아니라 우리 사회의 제도, 문화, 가치관과 끊임없이 상호작용하며 변화를 만들어내는 사회적 기술(Socio-Technology)임을 이해하여야 한다. 따라서 인공지능 기술의 사회적 영향력을 면밀히 분석하여 실효성 있는 윤리 가이드라인을 정립하고, 이러한 논의가 추상적인 선언에 그치지 않도록 기술의 오남용을 방지하고 책임 있는 혁신을 유도할 수 있는 최소한의 법적 안전장치를 체계적으로 구축해야 한다. 더불어 인공지능을 수동적으로 소비하는 이용자의 역할에서 벗어나, 인간의 가치와 기술이 조화롭게 어우러지는 미래를 설계하는 사회 설계자가 되어야 한다. 인간의 존엄성을 최우선 가치로 두고, 인공지능 시스템이 진정으로 인류의 안녕을 위해 작동하고 있는지 끊임없이 성능을 평가하며 사회적 기대치를 설정하는 과정을 지속해야 한다. 결국 이와 같은 인간 중심의 가치를 바탕으로 기술과 사회가 능동적으로 상호작용하며 공동 진화(Co-evolution)를 이룰 때, 인공지능은 비로소 인간을 소외시키지 않고 우리 삶의 질을 높이는 진정한 파트너로 자리 잡을 수 있을 것이며, 결국 인공지능의 미래는 기술 그 자체에 있는 것이 아니라 기술을 어떻게 이해하고 어떤 제도적·윤리적 틀 안에서 활용하느냐에 달려 있음을 다시 한번 역설하였다.

나. 주요 토론내용

이어진 순서에서는 인공지능 리터러시를 비롯한 인공지능 윤리 정책에 대한 위원 간 자유토론이 진행되었으며, '소외계층 관련 인공지능 리터러시 및 교육', '저작권 보호 및 데이터 거버넌스' 등에 관한 다양한 의견이 개진되었다.

〔그림 4-8〕 제2차 포럼 종합토론

자료: ‘제4기 인공지능 윤리정책 포럼’ 제2차 회의, 2025.9.26.

먼저 디지털 및 인공지능 소외계층 대상의 교육 및 인공지능 리터러시 제고 방안에 대한 토론이 진행되었다. 인공지능 활용 교육과는 달리, 인공지능으로 인한 부작용 및 주의 사항에 대한 교육과 특히 소외계층과 아동·청소년 맞춤형 접근이 부족하다는 지적이 제기되었으며, 해당 영역에 대한 고민과 정책 노력의 필요성이 강조되었다. 더 나아가 포용적 인공지능과 관련하여 노년층에 대한 정책적 고민 또한 제안되었다. 최근 1인 노년 가구가 급증하는 추세이나, 핵심적으로 조망되지 않는 부분이므로 더욱 심도 있는 논의가 필요하다는 제언이 나왔다. 또한, 기본적으로 사람 중심의 인공지능 확립이 가장 중요하며, 인권침해 등 인공지능 윤리에 대한 고민이 앞으로도 지속되어야 한다는 것에 의견을 모았다.

저작권 보호 및 데이터 거버넌스에 대한 논의 또한 주요하게 진행되었다. 국내 인공지능 기술 도약 및 기업 진흥을 위해 옵트아웃(Opt-out)[7] 방식 도입, 국가의 데이터 지원 등 저작권 문제 완화 노력이 필요하나, 국제법과의 연계성과 해외 공룡 기업의 데이터 독식 등 부작용에 대한 고려 또한 필요함이 강조되었다. 또한, 공정이용 및 저작권의 최종 목적이 저작권자 보호에서 창작물의 공익적 활용으로 확대되는 추세에 따라 법조계에서 활발하게 진행되는 인공지능 창작의 공익성 및 필터링, 라이선스 등 데이터 거버넌스 차원의 주요 논의 내용들을 공유하였다. 아울러, 인공지능 학습 데이터를 불법으

7) 정보주체의 동의를 받지 않고 개인정보를 수집·이용한 후, 당사자가 거부 의사를 밝히면 개인정보 활용을 중지하는 방식

로 획득한 데이터와 변형 및 재생산을 통해 생성한 데이터로 분리하여 대응하는 국제적 추세를 소개하며, 전자의 경우 저작권 침해에 해당되나 후자의 경우에는 데이터 변형 등의 인공지능 기술 발전 및 기업 혁신을 저해하지 않도록 별도의 논의가 필요하다는 의견이 제시되었다.

이외에도 현재 인공지능은 완성된 기술이 아니므로 부정확성, 안정성 등의 문제가 존재하므로 기술 발전의 지원과 동시에 기업의 내부적 시스템 및 체계 정립 지원의 필요성이 언급되었다. 또한 인공지능 알고리즘을 이용한 자동화 결정에 대한 투명성·설명가능성·책임성이 특히 중요하며, 미래 혁신 기술로 손꼽히는 'Agentic AI'[8] 또한 인적 개입의 최소화를 지향한다는 점을 고려한다면 인공지능에 대한 윤리적 접근의 중요성은 더욱 높아진다. 따라서 이에 대한 새로운 고민이 필요하다는 의견이 제기되었다. 아울러 유네스코를 비롯하여 아세안, 남미 등의 많은 국가가 한국의 사례에 관심이 높으므로 인공지능 보안 및 안전, 감독기구, 글로벌 네트워크 등의 국제 협력과 이를 통한 국내 산업 경쟁력 확대를 위한 논의가 지속되어야 한다고 제언했다.

3. 2025 인공지능 윤리 공개 세미나

국내 인공지능 윤리 정책을 실제 이행 중심의 제도적 체계로 전환하기 위한 논의의 장으로, 11월 27일 서울 양재 엘타워에서 '2025 인공지능 윤리 공개 세미나'가 개최되었다. 이번 세미나는 기조 강연을 시작으로 인공지능 채용 서비스 윤리 영향평가 중간 결과 발표, 민간 자율 인공지능 윤리위원회 표준지침(안) 소개, 종합토론 순으로 진행되었다. 특히 제4기 인공지능 윤리정책 포럼 위원을 포함한 산·관·학·연 전문가는 물론 일반 시민들까지 폭넓게 참여하여 국내 인공지능 윤리 정책의 향방에 대한 높은 관심을 드러냈다. 이번 세미나는 온·오프라인 병행 개최를 통해 누구나 자유롭게 의견을 개진할 수 있는 열린 논의의 장이 되었으며, 온라인으로 실시간 중계되었다.

8) Agent AI란 이용자가 정의한 목표를 달성하기 위해 인공지능 모델을 활용하여 환경과 상호작용하는 시스템으로 추론(reasoning), 계획(planning), 외부 도구를 사용한 행동의 실행(execution)을 결합하여 작업을 수행

〔그림 4-9〕 2025 인공지능 윤리 공개 세미나

자료: ‘2025 인공지능 윤리 공개 세미나’, 2025.11.27.

문명재 제4기 인공지능 윤리정책 포럼위원장의 주재로 진행된 이번 종합 토론은 ‘「인공지능기본법」 시행과 인공지능 윤리 정책 아젠다 전환 및 사회적 과제’를 주제로 다루었다. 특히, 2026년 「인공지능기본법」 시행이라는 전환기적 시점을 맞아 향후 정책 방향과 현장의 실천 과제에 대한 심도 있는 논의가 이루어졌다. 토론자로는 제4기 인공지능 윤리정책 포럼 위원인 김도엽 변호사(김앤장 법률사무소), 김동환 대표(포티투마루), 김명주 소장(인공지능 안전연구소), 김유철 부문장(LG AI연구원), 박성필 원장(KAIST 문술미래전략대학원), 이상욱 교수(한양대학교)와 간사인 이현경 연구위원(정보통신정책연구원)이 참여하였다. 토론자들은 전문성과 포럼 활동 경험을 토대로 인공지능 윤리정책의 아젠다 전환 및 기업의 애로사항 해소를 위한 제도적 방안을 중심으로 열띤 토론을 펼쳤으며, 주요 토론 내용은 다음과 같다.

〔그림 4-10〕 2025 인공지능 윤리 공개 세미나

자료: '2025 인공지능 윤리 공개 세미나', 2025.11.27.

가. 포스트 「인공지능기본법」 시대, 인공지능 윤리정책의 아젠다 전환

포스트 「인공지능기본법」 시대에 향후 인공지능 윤리가 과연 어떤 방향으로 진행되어야 하는가에 대한 질문에 대해서 포티투마루 김동환 대표는 우선 「인공지능기본법」이 인공지능 개발부터 서비스 제공, 사후 관리까지 전 생애주기를 포괄적으로 규율하는 최초의 근거법으로서 큰 의미를 갖는다고 평가하였다. 또한 사후 규제보다는 사전 예방에 중점을 두고 있으며, 기술 진흥을 넘어 윤리·안전성·신뢰성을 국가 차원의 성장 기준으로 제시하며, 기업과 정부의 역할과 책임을 보다 구체화한다는 점에서 의의가 있다고 언급하였다. 「인공지능기본법」 시행을 통해 향후에는 선언적 윤리에서 벗어나 실질적 이행 중심으로 규율이 전환될 것으로 예상되며, 이를 위해 기업 현장에서 작동 가능한 안전성 기준과 책임 추적 체계, 특히 중소·스타트업을 고려한 지원 인프라 구축이 중요하다고 주장하였고, 궁극적으로 「인공지능기본법」은 규제 자체보다 조직 내 인공지능 윤리 거버넌스의 정착과 실효적 이행을 촉진하는 계기가 되어야 함을 다시 한번 강조하였다.

LG AI연구원 김유철 부문장 역시 규제보다는 기업이 스스로 신뢰성을 증명하고 책임을 지는 '민간 자율규제'로의 이행을 강조하며, 정부는 기업이 이를 실천하기 쉽도록 가이드라인을 제공하는 데 집중해야 한다고 제언하였다. 또한 인공지능 윤리정책은 윤리 원칙과 거버넌스 구축의 선언 단계, 프로세스 수립과 실행을 통한 이행 단계, 성과를 외부에 공유하는 확산 단계로 구분될 수 있으며, 정보통신정책연구원에서 개발 중인 민간

자율인공지능윤리위원회 표준지침은 이러한 전 과정을 포괄하고 있다는 점에서 중요한 역할을 수행할 것이라고 평가하였다. 아울러 향후에는 다수의 기업이 인공지능 윤리정책을 보다 쉽게 이행할 수 있도록 실무 중심의 가이드 제공이 필요하다고 주장하였다.

한양대학교 이상욱 교수는 국제 정세가 이미 윤리를 넘어 적극적인 규제 준수를 요구하는 단계로 바뀌었음을 지적하며, 이를 선제적으로 내재화하여 역량으로 전환해야 한다고 제언하였다. 구체적으로 인공지능 윤리자율점검표의 수행이나 관련 인증을 통해 기업은 윤리적 고려 사항을 선제적으로 확인하고 문제를 예방할 수 있으며, 이를 통해 내부 윤리 역량과 인식을 강화할 수 있다. 기업 입장에서 새로운 제도 도입은 비용 부담을 발생시켜 거부감을 초래할 수 있으나, 국제적 현실을 고려할 때 인공지능 윤리를 적극적으로 내재화하고 기업 역량으로 전환하는 것이 바람직하다. 이를 위해 정부 정책은 단순 규제 집행에 그치지 않고, 기업이 규제 대비 역량을 강화할 수 있도록 컨설팅과 지원 중심으로 설계되어야 함을 강조하였다.

인공지능 안전연구소 김명주 소장 또한 인공지능 윤리가 기업 경쟁력의 핵심으로 부상할 것이라는 견해에 동의하며, 「인공지능기본법」 시행 이후 시민들이 인공지능 윤리에 대한 관심이 더욱 높아질 것으로 분석하였다. 이에 따라 기업은 기술력뿐만 아니라 부작용 대응과 윤리적 책임 또한 기업 가치 평가의 중요한 요소로 고려해야 하며, 정부는 투명성, 공정성 등의 윤리 원칙을 평가하고 인증할 수 있는 기준과 제도를 마련하고 제공할 필요성이 커지고 있음을 지적하였다. 특히 「인공지능기본법」은 EU의 인공지능법 대비 의무 규제가 적기 때문에 기업은 단순한 법 준수보다는 소비자의 기대와 기준에 맞춰 윤리적 활동을 강화할 필요가 있다고 강조하였다.

정보통신정책연구원 이현경 연구위원은 「인공지능기본법」 시행으로 인공지능 윤리의 역할과 과제가 더욱 확대될 것으로 전망하였다. 실제 현장에서는 「인공지능기본법」에 대한 인식이 아직 부족하며 분야별·주제별 인공지능 윤리교육과 확산이 필요하다고 진단하였다. 아울러 시민 권리 보호와 구제 정책, 특히 설명 요구권이나 인공지능 오남용 피해구제와 관련된 국제적 논의가 아직 활발하지 않으므로 우리나라에서 선도적으로 논의를 시작할 필요가 있다고 언급하였다. 또한 미성년자 및 취약계층 보호를 위해서 인공지능 윤리의식을 확산하고 다양한 교육 방법을 선제적으로 개발·보급해야 한다고 제언하였다.

김앤장 법률사무소 김도엽 변호사는 「인공지능기본법」으로 인해 연성 규범이었던 인공지능 윤리원칙의 일부가 법적 규정으로 전환되었으며, 특히 투명성을 포함한 다양한 윤리 원칙의 구현에 기업들의 관심이 높아지고 있다고 언급하였다. 이러한 상황에서 국제적으로 권장되는 영향평가 제도가 사전적으로 평가 가이드를 제공한다는 점에서 의미가 있으며, 정보통신정책연구원에서 시행 중인 인공지능 윤리영향평가가 해당 역할을 충실히 수행할 수 있을 것이라고 평가하였다. 또한, 기업이 인공지능 거버넌스를 구축할 때 윤리 원칙을 실제 내부 규정과 절차 안에서 어떻게 적용할지가 핵심이므로, 향후 기업 내 인공지능 윤리 원칙의 구체적 실행 방법과 전체 체계 구축 방향을 제시하는 연구가 필요하다고 제언했다.

한국과학기술원 문술미래전략대학원 박성필 원장은 인공지능은 변화 속도가 매우 빠른 기술 특성을 가지고 있으므로, 이에 대응하기 위해 예견적 거버넌스의 중요성을 강조하였다. 「인공지능기본법」은 이러한 필요성에 따라 신속히 제정되었으나, 빠른 입법 과정에서는 시민의 충분한 이해와 사회적 공감대를 형성하는 데 한계가 있을 수 있다. 따라서 법 시행 과정에서 지속적인 피드백과 검증, 시행령 보완을 통해 법의 적절성, 인권 보호, 실현 가능성 등을 점검하고 개선하는 노력이 필요하다고 주장하였다. 아울러 준비 단계에서의 테스트와 시행 과정에서의 반복적인 점검을 통해 법 적용상의 문제를 발견하고 수정하여, 사회적 공감대를 지속적으로 반영하여 제도의 완성도를 높여야 한다고 강조하였다.

〔그림 4-11〕 '2025 인공지능 윤리 공개 세미나' 종합 토론

자료: '2025 인공지능 윤리 공개 세미나', 2025.11.27.

나. 기업의 애로사항과 제도적 해소 방안

토론의 두 번째 세션에서는 인공지능 윤리원칙을 기업 현장에 적용할 때 발생하는 구체적 애로사항과 이를 해소하기 위한 제도적 대응 방안이 논의되었다. 기업 패널들은 인공지능 윤리 준수의 실무적 한계와 규제 불확실성을 주요 고충으로 꼽았다. 김유철 부문장은 규제 설계에 있어 신중한 접근을 강조하며, 기업 혁신 동력을 저해하지 않는 산업 육성 중심의 정책 환경 조성을 주문했다. 기업이 인공지능 도입 시 가장 중요하게 고려하는 요소로 정확성, 보안, 규제 정합성을 제시하고, 특히 정확성과 보안은 단순한 모델 성능을 넘어 투명성·공정성·안전성·강건성 등의 윤리적 가치가 충족될 때 확보될 수 있다고 분석했다. 기업들은 이미 규제 이전에 시장 요구에 부응하는 방식으로 인공지능 윤리를 자연스럽게 이행하고 있으며, 단편적이고 과도한 규제는 산업 생태계 전반을 위축시키고 혁신의 역동성을 저해할 수 있으므로 규제 설계에 각별한 주의가 필요하다고 지적했다. 아울러 부처별 상이한 가이드라인이 기업 현장에 정책적 혼선을 야기하고 있음을 언급하며, 범부처 차원의 일관된 기준 마련과 세밀한 정책 조율이 필요하다고 제언했다.

김동환 대표는 기업이 인공지능 윤리를 이행하는 과정에서 비용·시간·인력의 부담이 필연적으로 수반된다고 지적했다. 특히 인공지능의 안전성 및 신뢰성 확보를 위한 검증 도구 도입과 인증·평가 과정에서 발생하는 제반 비용이 기업에 상당한 부담이 된다고 언급하였다. 인력 측면에서도 인공지능 윤리 전문 인력 확보와 사내 거버넌스 구축은 개별 기업에게 부담으로 작용하므로 실무 중심 인력 양성과 사내 거버넌스 수립을 위한 정책적 지원이 뒷받침되어야 한다고 강조했다. 또한 기업의 부담을 완화하고 자발적인 인공지능 윤리 경영을 유도하기 위해 공공 차원의 검증 도구 개발 및 배포, 인증 비용 지원, 세제 혜택 등 실질적인 인센티브 제도 마련을 촉구하였다. 아울러 국제 시장 진출 시 중복 인증에 따른 자원 낭비를 방지하기 위해 국가 간 인증 상호 인정 체계 구축의 필요성도 함께 역설했다. 또한, 결국 기업이 가장 어려워하는 요소는 규제 자체가 아닌 불확실성이라고 지적하며, 이를 해소하기 위해 「인공지능기본법」에 대한 명확한 해석 기준과 가이드라인, 컨설팅 지원 등이 중요하다고 강조하였다.

토론자들은 이러한 기업의 고충을 해소하기 위해 규제자가 아닌 조력자로서의 정부 역할을 강조하며 실효성 있는 제도적 대안을 제시했다. 이상욱 교수는 대다수 기업이

기술 개발과 서비스 제공 과정에서 인공지능 윤리 지침을 준수하려는 의지가 있음에도 불구하고, 추상적인 윤리 원칙을 실제 기술 현장에 구체적으로 적용하는 실무적 번역 단계에서 어려움을 겪고 있다고 분석했다. 이를 해결하기 위해 단순한 원칙의 나열이 아니라, 추상적 규범을 실제 기업 사례에 맞게 구체화한 산업별 모범 사례를 발굴하여 보급해야 한다고 제언하였다. 아울러 산업별·기업별 특성에 맞춘 사례들을 수집, 공유, 학습할 수 있는 체계적 거버넌스 마련의 필요성도 강조하였다. 이를 위해 현장 밀착형 컨설팅 지원을 강화하고, 기업이 스스로 역량을 키울 수 있도록 지침 유포와 교육 지원을 병행해야 한다고 덧붙였다.

김도엽 변호사는 법률 실무적 관점에서 「인공지능기본법」 시행을 앞둔 기업의 현실적인 우려를 전달하며 구체적인 지원책을 제안했다. 「인공지능기본법」상 고영향 인공지능 등의 기준이 모호해 기업이 규제 대상 여부 판단에 혼란을 겪고, 증빙 자료 제출 시 영업비밀 유출에 대한 우려가 크다는 점을 지적했다. 이에 따라 초기 단계에서는 시나리오별 검토 지침을 마련하여 기업의 예측 가능성을 높여야 한다고 강조했다. 또한 자체 거버넌스 구축이 어려운 중소기업·스타트업을 위해 재정적 지원과 안내 제도를 강화하고 과기정통부·정보통신정책연구원 차원의 원스톱 상담창구 운영이 필요하다고 언급했다. 마지막으로 기업의 해외 진출을 돕기 위해 국내법이 국제적 기준과 보조를 맞춰야 함을 역설했다. 급변하는 해외 규제 동향을 기업이 개별적으로 파악하기 어려우므로 정부 차원에서 국제적 규제 정보를 통합 안내하고 국내외 법제 정합성을 확보해야 한다고 강조했다.

기술적 지원과 법적 정비에 대한 심도 있는 제언 또한 이어졌다. 김명주 소장은 「인공지능기본법」 시행에 따른 기업 혼란을 방지하기 위한 정부 차원의 기업 지원과 국제적 수준의 안전성 검증 체계 구축을 제안했다. 먼저, 기업이 「인공지능기본법」의 법적 요건을 원활히 이행할 수 있도록 돕는 실무적인 지원 창구가 필수적이라고 언급했다. 또한, 인공지능의 신뢰성을 담보하기 위한 안전성 테스트와 평가의 중요성을 강조했다. 인공지능 안전연구소는 해외 주요국 연구소와 협력하여 국제 기준에 맞춘 테스트를 수행하며 일부 국내 기업도 이를 활용한 경험이 있다고 설명했다. 이러한 테스트는 법적 의무가 아니라 국제적 경쟁력 확보 및 강화 수단으로 적극 활용할 것을 제안했다. 나아가, 국제 시장의 사실상 표준을 선점하기 위해서는 기업들이 안전 테스트에 적극 참여

하고, 정부가 적시에 법적·기술적 서비스를 제공하여 기업이 체감할 수 있는 실질적인 도움을 주는 체계가 구축되어야 한다고 제언했다.

박성필 원장은 기업의 혼란을 줄이기 위해 규제의 불확실성을 해소하고 법제도를 통합·단순화해야 한다고 제언했다. 현재 기업들은 「인공지능기본법」뿐만 아니라 개인정보보호법, 데이터 관련 법령 등 여러 규제가 산재해 있어 중복 대응에 어려움을 겪고 있으며, 특히 자체적인 대응 역량이 부족한 중소기업과 스타트업의 경우 법규 준수 부담이 큰 상황이라고 진단했다. 이러한 문제를 해결하기 위해 EU의 디지털 옴니버스 패키지(Digital Omnibus Package), GDPR과 같이 데이터 규제를 포괄적으로 단순화·통일화한 사례를 참고하여 기업이 직관적으로 이해할 수 있는 통합적 규제 체계를 마련하는 정책적 노력이 필요하다고 강조했다. 이러한 작업은 단순히 여러 법을 물리적으로 합치는 것이 아니라, 기업 현장의 실제 상황을 반영하고 현장에서 축적된 산업별 모범사례를 기반으로 실제 기업 현실에 맞게 규제를 재정립해 나가는 과정이어야 한다고 제언했다. 이를 위해 기업과 시민사회의 의견을 적극 수렴하고 실무적인 모범사례를 폭넓게 축적할 필요가 있다고 덧붙였다.

이현경 연구위원은 「인공지능기본법」과 같은 법적 규제만으로는 통제하기 어려운 영역이 존재함을 지적하며 기업 현장의 실질적인 윤리 실천 역량과 실행력 확보의 중요성을 역설했다. 단순히 인공지능 윤리원칙을 선언하는 단계에 그치지 않고 이를 실제 기업 경영과 조직 운영 체계에 내재화하는 것이 핵심적인 도전 과제라고 평가했다. 특히 전 세계적 거대 기술 기업들조차 윤리원칙 선언 이후 경영 환경 변화에 따라 윤리 전담 조직을 해체하는 등의 사례가 빈번하다는 점을 예시로 들며, 인공지능 윤리의 지속적 이행이 기업 내 결정권자의 의지에 따라 좌우되는 현실을 지적했다. 이에 따라 기업의 인공지능 윤리가 기업 내부 상황에 영향을 받지 않고 지속적으로 유지 및 강화될 수 있도록 학계와 소비자, 시민사회가 지속적으로 관심을 두고 건설적인 피드백을 제공해야 한다고 제언하였다. 결국 법적 규제를 넘어 실행력을 갖춘 민간 중심의 자율적 인공지능 윤리 거버넌스를 공고히 구축하는 것이 가장 큰 과제임을 강조하였다.

〔그림 4-12〕 '2025 인공지능 윤리 공개 세미나' 종합 토론

자료: '2025 인공지능 윤리 공개 세미나', 2025.11.27.

다. 마무리 발언

토론을 마무리하며 토론자들은 인공지능 시대의 핵심 과제로 법·제도의 예측 가능성과 투명성 확보, 기업과 이용자의 윤리의식 제고, 사회적 신뢰와 수용성을 높이기 위한 지속적인 공론화와 교육의 필요성을 재차 강조했다. 먼저 김도엽 변호사는 인공지능이 선택이 아닌 일상의 필수 기술로 자리 잡는 상황에서 법적 예측 가능성과 투명성 확보의 중요성을 강조하며, 인공지능 윤리 포럼이 사회적 발전과 보호를 조화시키는 가이드라인을 제시하기를 희망한다고 밝혔다. 김동환 대표는 우리나라가 인공지능 윤리 분야에서 제도적 준비를 선도해 왔다고 평가하면서도, 법제화와 시스템 구축을 넘어 기업과 일반 국민의 윤리의식을 실질적으로 함양할 수 있는 인공지능 윤리교육의 중요성을 강조했다. 김명주 소장은 완벽하지 않은 인공지능을 사용하는 시대에 이용자의 윤리가 중요하며, 이용자의 윤리기준이 높아질수록 기업의 책임 수준도 함께 향상될 수 있다고 설명했다. 박성필 원장은 사회적 수용성을 높이기 위해 기술과 규제에 대한 사회적 이해, 예측 가능성, 정부와 기업에 대한 신뢰 구축이 필수적이며, 이를 위한 지속적인 공론화 노력이 필요하다고 밝혔다. 김유철 부문장은 일자리 감소나 범죄 활용과 같은 문제에 대해 막연한 공포를 경계하며 객관적인 통계와 효과 분석을 바탕으로 인공지능 정책을 세워야 하며, 포럼이 사회적 합의를 이끄는 중추적인 역할을 수행해야 한다고 제안했다. 이현경 연구위원은 법 제도의 정비와 함께 인공지능 윤리교육이 제도의 공백을 보완할 핵심 수단임을 강조하며, 교육 대상의 특성과 상황을 고려한 맞춤형 교육을 더

욱 확대해 나갈 계획임을 밝혔다. 마지막으로 문명재 위원장은 “The Show Must Go On”이라는 말로 토론을 매듭지으며 인공지능 윤리가 선언적 구호에 그치지 않고 우리 사회에서 실질적으로 이행될 수 있도록 앞으로도 지속적인 논의와 실천을 멈추지 않고 이어가야 한다고 강조하며 토론을 마무리했다.

제5장 결론 및 시사점

제1절 결과 요약

인공지능 윤리 실천수단 확산 방안에 관한 연구를 마무리하는 시점에 '국가인공지능전략위원회'에서 「대한민국 인공지능 행동계획(초안)」(이하, 인공지능 행동계획 초안)(2025.12)을 발표하였다. '인공지능 행동계획 초안'은 인공지능 혁신 생태계 조성, 범국가 인공지능 기반 대전환, 국제 인공지능 기본사회에 기여라는 세 개의 정책 축을 중심으로 총 98개의 과제로 구성되어 있다.[9] 그 중에서도 82번 과제로 제시된 〈모두의 AI를 위한 AI 윤리 확산·고도화〉 과제는 해당 연구와 긴밀한 연관성을 갖는다고 평가할 수 있다. 이러한 환경 변화 속에서 2020년에 수립된 「인공지능 윤리기준」의 한계를 점검하고, 국제적 정합성과 현장 적용성을 동시에 확보하는 방향으로 윤리 정책을 재설계해야 할 필요성이 분명해지고 있다.

이러한 맥락에서 그간 운영되어 온 인공지능 윤리정책 포럼의 활동은 매우 시의적절한 정책적 실천으로 평가할 수 있다. 포럼은 인공지능 윤리를 기술 규범이나 선언적 원칙의 문제로 한정하지 않고, 실제 정책 설계와 이행 과정에 다양한 이해관계자의 목소리를 반영할 수 있는 공론장으로 기능해 왔다. 이는 「대한민국 인공지능 행동계획」이 강조하고 있는 바와 같이, 사회적 합의를 기반으로 윤리 확산 메커니즘을 구축해야 한다는 정책 방향과 직접적으로 부합한다. 특히 생성형 인공지능 확산에 따른 새로운 윤리 쟁점을 선제적으로 논의하고, 민간의 자율적 실천과 정부 정책 간의 연결을 모색해 왔다는 점에서 포럼의 논의는 정책적 예측성과 준비성을 제고하는 역할을 수행해 왔다.

또한 인공지능 윤리정책 포럼은 윤리기준 고도화와 자율 점검 체계 마련이라는 향후 정책 과제의 실험적·준비 단계로서 중요한 기능을 해왔다. 행동계획에서 제시된 바와 같이, 정부는 2026년까지 인공지능 윤리기준을 고도화하고, 개발·이용 전 생애주기에 걸

9) 해당 문서는 최종본은 아니며, 20일간('25.12.16~'26.1.4) 공개 의견 수렴 및 향후 각계 의견을 거쳐 수정·보완될 예정이다.

친 윤리 자율 점검표를 정비할 계획을 밝히고 있다. 이러한 정책 과제가 실효성을 갖기 위해서는, 법·제도적 설계 이전에 현장의 요구와 우려, 적용 가능성에 대한 충분한 논의와 검증이 선행되어야 한다. 포럼은 바로 이러한 역할을 수행할 수 있는 정책 실험 공간으로서, 윤리기준의 추상성을 낮추고 정책 도구의 현장 적합성을 높이는 데 기여해 왔다.

나아가 행동계획에서 2027년 1분기부터 인공지능 윤리 포럼을 본격 운영하고, 그 결과를 차년도 국가 인공지능 정책에 반영하겠다는 계획을 명시한 점은, 그간의 포럼 운영 경험이 일회성 논의에 그치지 않고 제도화 단계로 이행하고 있음을 의미한다. 이는 인공지능 윤리정책 포럼이 단순한 자문 기구를 넘어, 국가 인공지능 거버넌스 체계 내에서 지속적·구조적으로 기능하는 정책 플랫폼으로 자리매김할 가능성을 보여준다. 특히 「인공지능기본법」에 따라 확산의 근거가 명확해졌으며, 법을 기반으로 한 거버넌스의 운영이 가능해졌다고 볼 수 있다.

본 연구는 「인공지능기본법」 제28조에 근거한 민간자율위원회의 설치 목적과 법적 성격을 규명하고, 이를 실효적으로 운영하기 위한 표준지침(안)을 제안하는 것을 목적으로 수행되었다. 이를 위하여 먼저 법률의 내용을 분석하여 국가 주도의 하향식 규제가 갖는 한계를 극복하기 위해 민간이 주체적으로 참여하는 상향식 윤리 거버넌스의 필요성과 법률이 위원회 설치를 임의 규정으로 둔 취지를 고려하여 기업 규모와 특성에 맞는 맞춤형 윤리 실천의 유도라는 방향성을 도출하였다. 또한, 위원회 운영에 있어 민간의 폭넓은 자율성을 보장하되, 자율이 방임으로 흐르지 않도록 공정성과 중립성이라는 공익적 가치를 조화시키는 방안을 법 제28조 제3항(구성의 다양성 및 외부 전문가 참여)과 제4항(표준지침 마련 근거)을 통해 확인하였다.

본 연구의 최종 결과물로서 제안된 '민간자율인공지능윤리위원회 표준지침(안)'은 「인공지능기본법」 제28조가 지향하는 자율규제와 사회적 책무를 현장에서 조화롭게 구현하기 위한 구체적인 실천 수단이다. 이 지침(안)을 통해 단순히 위원회의 설치를 권고하는 수준을 넘어, 위원회가 실질적인 윤리 통제 기구로서 기능하기 위해 갖추어야 할 거버넌스 구조, 운영 절차, 투명성 확보 방안을 체계적으로 제시하였다.

첫째, 위원회 구성에 있어 절차적 독립성과 다양성을 제도화하였다. 표준지침(안)은 기업 내부에 설치되는 위원회가 경영진의 간섭이나 이익 논리에 포섭되는 것을 방지하

기 위해, 구성 단계에서부터 견제 장치를 마련하였다. 법 제28조 제3항 단서를 구체화하여, 위원회 구성 시 특정 성(性)이 위원 전원을 차지하지 못하도록 성별 균형을 의무화하였으며, 기술적 판단에 치우치지 않도록 사회적·윤리적 타당성을 평가할 수 있는 외부 전문가와 비종사자의 참여를 필수 요건으로 규정하였다. 이는 위원회가 다양한 이해관계자의 관점을 반영하고, 객관적인 시각에서 인공지능 시스템의 위험성을 진단할 수 있도록 하는 핵심 기제에 해당한다.

둘째, 기업 맞춤형 유연한 운영 모델과 윤리적 통제의 균형점을 제시하였다. 연구 결과물은 기업의 규모나 업종에 따라 위원회 운영 방식이 달라야 함을 인정하고, 회의 주기나 의사결정 방식 등 세부 운영 사항에 대해서는 기업의 자율권을 폭넓게 인정하였다. 이를 통해 대기업부터 스타트업까지 각자의 여건에 맞는 맞춤형 윤리 거버넌스를 구축하되, 윤리적 안전망이라는 본질적인 기능은 놓치지 않도록 설계하였다.

셋째, 투명성과 자율적 환류(Feedback) 체계를 강화하였다. 표준지침(안)의 결과물 중 하나는 위원회의 활동내역과 윤리적 의사결정 과정을 담은 연차 보고서의 발간을 권고한 점이다. 이는 기업이 수행한 윤리적 노력과 성과를 이해관계자들에게 투명하게 공개하여 사회적 신뢰를 확보하는 수단이 된다. 또한, 보고서 작성 과정 자체가 기업 스스로 지난 활동을 회고하고 윤리적 역량을 진단하여 부족한 점을 개선하는 자율적 환류 체계로 작동하도록 하여, 외부의 강제 없이도 지속 가능한 윤리 경영이 가능하도록 유도하였다.

제2절 정책적 제언

향후 인공지능 윤리정책 포럼은 몇 가지 방향에서 전략적으로 발전할 필요가 있다. 첫째, 윤리기준 고도화와 자율 점검표 및 기타 정책 결과물 개발 과정에서 실질적인 정책 공동설계(co-design) 기구로서의 역할을 강화할 필요가 있다. 단순한 의견 수렴을 넘어, 정책 초안 단계부터 다양한 이해관계자가 참여하는 구조를 통해 정책의 정합성과 수용성을 높여야 한다. 둘째, 생성형 인공지능을 포함한 고위험·고영향 인공지능에 대한 논의를 심화하여, 기술 발전 속도를 따라잡는 '대응적 윤리'에서 나아가 '예측적 윤리 거버넌스'로의 전환을 모색해야 한다. 셋째, 국제 윤리기준과의 정합성을 지속적으로 점

검하고, 국내 논의 결과를 국제 협력과 규범 형성 논의로 환류시키는 연결 창구로서의 기능도 강화할 필요가 있다.

인공지능 윤리정책 포럼의 그간 활동은 「대한민국 인공지능 행동계획」이 제시한 윤리 확산·고도화 방향을 선제적으로 준비하고 실험해 온 과정으로 평가할 수 있다. 이는 우리나라가 인공지능 윤리를 규제의 사후적 보완 수단이 아닌, 기술 발전과 사회적 신뢰를 함께 견인하는 핵심 정책 축으로 인식하고 있음을 보여준다. 향후 포럼이 보다 제도화되고 정책 연계성이 강화될 경우, 우리나라는 국제사회가 요구하는 책임 있는 인공지능 거버넌스를 실질적으로 구현하는 선도 국가로 자리매김할 수 있을 것이다.

또한 본 연구의 결과물로 민간 주도의 자율적 인공지능 윤리 거버넌스를 확립하기 위한 제도적 토대로서 '민간자율인공지능윤리위원회 표준지침(안)'을 개발하여 제시하였다. 그러나 이 지침이 실효적으로 기업 현장에서 작동하고, 건강한 인공지능 생태계 조성에 기여하기 위해서는 다음과 같은 정책적 후속 조치가 뒤따라야 한다.

첫째, 인공지능 사업자 대상 설명회 개최를 통한 표준지침(안)의 효용성 전파 및 인식 제고이다. 표준지침(안)의 확산을 가로막는 가장 큰 장벽은 기업들의 규제 피로감과 오해일 것이다. 많은 인공지능 사업자는 민간자율위원회를 또 하나의 비용이자 행정적 부담으로 인식할 가능성이 크다. 따라서 정부는 온오프라인 설명회를 적극적으로 개최하여 이러한 인식을 전환해야 한다. 설명회에서는 본 표준지침(안)이 획일적인 강제 규범이 아니라, 기업의 규모와 특성에 맞춰 유연하게 변형 가능한 가이드라인임을 강조해야 한다. 특히, 표준지침(안)에 따른 민간자율위원회 설치가 향후 발생할 수 있는 리스크를 예방하고, 국제 시장에서 경쟁력을 확보하는 효과적인 자율적 리스크 관리 수단임을 인공지능 사업자들에게 구체적으로 설명하고 설득하는 과정이 필수적이다.

둘째, 모범 사례(Best Practice) 발굴 및 확산을 위한 실무자 심층 인터뷰 추진이다. 제도 도입 초기에는 참고할 만한 선례가 부족하여, 도입 의지가 있는 기업조차 민간자율위원회 구성과 운영에 막막함을 느낄 수 있다. 따라서 설명회 이후, 표준지침(안)을 선도적으로 도입하여 운영 중인 기업의 실무자들을 대상으로 심층 인터뷰를 추진하여 기업 규모별로 외부 위원을 섭외한 경험, 개발 속도를 저해하지 않으면서 윤리성을 확보한 사례, 사후 모니터링의 구체적 기법 등 현장의 생생한 경험을 발굴해야 한다. 이렇게 수집된 성공 사례와 장애 요인 극복 경험은 실무 운영 매뉴얼의 형태로 가공되어 후발

주자들이 겪을 시행착오를 줄여주는 실질적인 안내서가 될 것이다.

셋째, 도입 기업의 의견 수렴을 통한 표준지침(안)의 지속적 고도화이다. 인공지능 기술은 급변하고 있으며, 산업 현장에서 마주하는 윤리적 쟁점 또한 계속 변화한다. 따라서 이번 연구를 통해 마련된 표준지침(안)을 고정불변의 원칙으로 두어서는 안 된다. 설명회와 시범 운영 과정에서 제기된 인공지능 사업자들의 목소리를 지속적으로 청취하여 지침을 현행화해야 한다. 이를 위하여 실제 운영 과정에서 불필요하게 경직된 절차는 없는지, 혹은 새롭게 등장한 인공지능 이슈 등을 다루기 위해 추가되어야 할 기준은 무엇인지에 대한 피드백을 확인하는 과정이 필수적이다.

참고문헌

국내 문헌

과학기술정보통신부(2023).『지능정보화 기본법』.

과학기술정보통신부(2024).『정보통신산업 진흥법』.

과학기술정보통신부(2024).『인공지능 발전과 신뢰 기반 조성 등에 관한 기본법』.

네이버(2021).『AI윤리준칙』.

네이버(2024.08.07.).『NAVER AI Safety Framework』. (https://www.clova.ai/tech-blog/ko-naver-ai-safety-framework-asf)

농림축산검역본부(2023).『동물실험윤리위원회(IACUC) 표준운영가이드라인』.

대한민국 정책브리핑(2024.11.27.).『'한국 AI안전연구소' 출범…AI 위협에 체계적·전문적 대응』. (https://www.korea.kr/news/policyNewsView.do?newsId=148936794)

문정욱 외(2023).『인공지능 윤리 의식 확산을 위한 정책연구』. 정보통신정책연구원 정책연구 23-24-01.

문정욱 외(2024).『AI 윤리·신뢰성 확보를 위한 실천 방안 및 정책연구』. 정보통신정책연구원 정책연구 24-18.

보건복지부(2022).『생명윤리법 관련 기관 운영지침. 기관생명윤리위원회 관리 안내』.

이나라(2025.11.26).『안전하고 책임있는 AI 도입을 위한 거버넌스 체계 구축』. 저작권동향, 2025-18호. 한국저작권위원회.

이재숭(2022).『중국의 인공지능(AI) 윤리 정책에 관한 연구』. 중국학 제80호, pp.69~87. 대한중국학회.

전자신문(2025.09.30.).『AI 신뢰성 검·인증 체계 확산한다…산·학·연 얼라이언스 출범』. (https://www.etnews.com/20250930000345)

카카오(2018).『AI 윤리』.

카카오(2023).『2023 카카오 공동체 기술윤리 보고서』.

카카오(2024).『2024 카카오 그룹 기술윤리보고서』.

황성기·황승흠(2003).『인터넷은 자유공간인가?: 사이버 공간의 규제와 표현의 자유』. 커뮤니케이션북스.

AP 신문(2023.05.15.).『농협은행, AI 윤리원칙 수립과 운영체계 구축 완료』. (https://www.apnews.kr/news/articleView.html?idxno=3010023)

KB 금융그룹(2022).『AI 윤리기준』.

KRAFTON AI.『Ethical AI』. (https://www.krafton.ai/ko/ai-ethics/principle/)

KT(2024).『KT ESG Report 2024』.

KT Responsible AI Center(2024).『KT Responsible AI Report』.

LG AI연구원(2025).『LG AI 윤리 책무성 보고서 2024』.

LG AI연구원(2026).『LG AI 윤리 책무성 보고서 2025』.

NC소프트(2022.9.26.).『AI Ethics Framework | Data Privacy』.

NC소프트(2022.9.27.).『AI Ethics Framework | Unbiased』.

NC소프트(2022.9.30.).『AI Ethics Framework | Transparency』.

SK(2025).『SK Telecom Annual Report 2024』.

해외 문헌

Alibaba Cloud Community(2022.9.23.). “Alibaba’s CTO On Everything You Wanted To Know About AI Ethics.”

Anecdotes(2025.9.28.). “AI Regulations in 2025: US, EU, UK, Japan, China and More.”

AnJie Broad Law firm(2025.1.20.). “AI Ethics: Overview (China).” China Law Vision.

Anthropic(2023.9.19.). “The Long-Term Benefit Trust.”

Australian Government. “Technology.” Department of Industry, Science and Resources. (https://www.industry.gov.au/science-technology-and-innovation/technology)

Australian Government(2024).『National Framework for the Assurance of Artificial Intelligence in Government』. Department of Finance.

Australian Government(2025). “Artificial Intelligence.” Department of Industry, Science and Resources. (https://www.industry.gov.au/science-technology-and-innovation/technology/

artificial-intelligence)

Baidu. 『Baidu AI Ethics Measures』.

Baidu. 『Technology Ethics』.

Cabinet Office of Japan(2022.4.22.). “AI Strategy 2022 - Realizing a Human-Centric Society (Society 5.0).”

Cabinet Office, Government of Japan(2019.6.11). 『AI Strategy 2019: AI for Everyone: People, Industries, Regions and Governments』. Integrated Innovation Strategy Promotion Council.

CONNECTCX(2025.4.17.). “Jack Ma Urges Ethical AI to Empower Humanity Amid China’s Tech Surge.”

Creemers, R. & Webster, G.(2021.8.20.). “Translation: Personal Information Protection Law of the People’s Republic of China - Effective Nov. 1, 2021.” Stanford University, DigiChina.

CSIS (Center for Strategic and International Studies). Tagui Ichikawa(2025.6.17). “Norms in New Technological Domains: Japan’s AI Governance Strategy.”

Cyberspace Administration of China(2023.7.10.). “Interim Measures for the Management of Generative Artificial Intelligence Services.”

Department of Industry, Science and Resources(2024). 『Voluntary AI Safety Standard』. Australian Government.

European Union(2025.8.11). “Artificial Intelligence Act.(EU Regulation 2024/1689).”

Google(2025). “Responsible AI Progress Report.”

Government of Canada. “Guide on the use of generative artificial intelligence.”

Government of Canada(2024). “Eight organizations to join Canada’s voluntary AI code of conduct.”

Government of Canada(2024). “Even more organizations adopting Canada’s voluntary code of conduct on artificial intelligence development.”

Government of Canada(2025). “AI Strategy for the Federal Public Service 2025-2027: Overview.”

Government of Japan(2024.2.9). “The Hiroshima AI Process: Leading the Global Challenge to Shape Inclusive Governance for Generative AI.”

GovTech Singapore(2025.1.21.). “AI in Education: Transforming Singapore’s education system with student learning space.”

IBM. “Reflecting on the Five-Year Anniversary of IBM’s AI Ethics Board.”

IBM. Kelly Churchill. “Foundations of trustworthy AI: governed data and AI, AI ethics and an open diverse ecosystem.”

Infocomm Media Development Authority(IMDA). “Artificial Intelligence in Singapore.”

Infocomm Media Development Authority(IMDA)(2023). “AI Verify: Testing Toolkit for Responsible AI.”

Lexology(2025.1.8). “Japan: Regulations Relating to Children’s Personal Information.”

Medium(2023.5.13.). “How effective is Google’s Bold and Responsible Approach to AI?”

Meta(2024.9.25.). “Connect 2024: The Responsible Approach We’re Taking to Generative AI.”

Microsoft. “Responsible AI at Microsoft.”

Microsoft(2024). “2024 Responsible AI Transparency Report.”

Microsoft(2025). “2025 Responsible AI Transparency Report.”

National Artificial Intelligence Centre(2025). 『Guidance for AI Adoption』. Australian Government, Department of Industry, Science and Resources.

National Institute of Standards and Technology(2023.1.26). 『Artificial Intelligence Risk Management Framework (AI RMF 1.0)』. U.S. Department of Commerce.

New South Wales Government, Australia(2020). 『Artificial Intelligence Ethics Policy』. Digital NSW.

OECD. “OECD AI Policy Observatory – National AI Policies Database.”

People’s Republic of China, Ministry of Foreign Affairs(2025.7.26.). “Global AI Governance Action Plan.”

Responsible AI UK(RAI UK). https://rai.ac.uk/

SECNewgate. Matthew Ford(2025.2.6.). “Google’s ethical crossroads: Lifting the ban on

AI for weapons and surveillance."

The Verge. Davis, W.(2023.11.19.). "Meta disbanded its Responsible AI team."

The White House. Executive Orders 14110. "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." Federal Register.

The White House. Executive Orders 14179. "Removing Barriers to American Leadership in Artificial Intelligence." Federal Register.

The White House Office of Science and Technology Policy(OSTP)(2022). "Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People."

UK Government(2025.1.13.). "AI Opportunities Action Plan." ISBN 978-1-5286-5362-6.

UK Government, Office for Artificial Intelligence(2021.9.22.). "National AI Strategy."

UK Parliament POST (Parliamentary Office of Science and Technology). "Artificial intelligence: ethics, governance and regulation." DOI: https://doi.org/10.58248/HS51

UNESCO(2021). "Recommendation on the Ethics of Artificial Intelligence."

UNESCO(2023). "Readiness Assessment Methodology – A Tool of the Recommendation on the Ethics of Artificial Intelligence."

White & Case LLP(2025.5.29.). "AI Watch: Global regulatory tracker – China."

부 록

부록 1: 민간자율인공지능윤리위원회 구성 및 운영에 관한 표준지침(안)

민간자율인공지능윤리위원회 구성 및 운영에 관한 표준지침(안)

제1조(목적) 이 규정은 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」 제28조에 따라 안전하고 신뢰할 수 있는 인공지능기술 연구·개발·활용을 위하여 기관 또는 단체가 자율적으로 설치하는 민간자율인공지능윤리위원회(이하 “민간자율위원회”라 한다)의 구성 및 운영 등에 필요한 사항을 규정함을 목적으로 한다.

제2조(민간자율위원회의 구성) ① 민간자율위원회는 위원장 1인을 포함하여 3인 이상의 위원으로 구성한다. 위원 수는 홀수로 구성한다.

② 민간자율위원회는 하나의 성(性)으로만 구성할 수 없으며, 위원은 다음 각 호의 어느 하나에 해당하는 자를 각 1명 이상 포함하여야 한다.

1. 사회적·윤리적 타당성을 평가할 수 있는 경험과 지식을 갖춘 사람
2. 해당 기관 또는 단체에 소속되지 않고, 이해관계가 없는 외부인

③ 제2항 각 호의 요건은 동일인이 2개 이상의 요건을 동시에 충족하는 경우에도 그 충족으로 본다.

④ 제2항의 위원은 해당 기관 또는 단체의 장이 임명 또는 위촉하며, 위원장은 위원 중에서 호선한다.

제3조(위원의 해촉) 해당 기관 또는 단체의 장은 위원이 다음 각 호의 어느 하나에 해당하는 사유가 있을 때에는 해임 또는 해촉할 수 있다.

1. 질병 등으로 인하여 직무를 수행할 수 없게 된 경우
2. 직무와 관련된 비위 사실이 있는 경우
3. 직무태만이나 품위손상 등의 사유로 위원에 적합하지 아니하다고 인정되는 경우
4. 이해상충 우려 등 위원 스스로 직무 수행이 곤란하여 사임 의사를 밝히는 경우

제4조(위원의 임기 등) ① 위원의 임기는 2년으로 하며, 연임할 수 있다.

② 위원 중 결원이 발생한 때에는 위원을 새로 위촉 또는 임명할 수 있다. 단, 결원으로 인해 위원 수가 3인 미만일 때에는 후임 위원을 반드시 위촉 또는 임명하여야 한다.

③ 제2항에 따라 위촉 또는 임명된 위원의 임기는 전임자의 잔여 임기로 한다.

제5조(위원장의 직무 등) ① 위원장은 민간자율위원회를 대표하며, 회의를 소집하고 주재하는 등 위원회의 직무를 총괄한다.

② 위원장이 부득이한 사유로 직무를 수행할 수 없을 때에는 위원장이 사전에 지명한 위원이 그 직무를 대행한다.

제6조(간사) ① 위원장은 위원회의 원활한 운영을 위하여 위원이 아닌 간사 1인을 임명할 수 있다.

② 간사는 다음 각 호의 업무를 수행한다.

1. 위원회 회의 준비, 회의록 작성, 안건 관리 등 행정 지원 업무
2. 위원회 구성원 대상의 윤리교육, 오리엔테이션 진행
3. 위원회의 심의 대상에 대한 사전검토 및 관련 자료 정리
4. 위원장이 부여한 기타 업무

③ 간사가 부재중일 때에는 위원장이 지정한 자가 업무를 대행한다.

제7조(회의 소집) ① 민간자율위원회 회의는 정기회의와 임시회의로 구분한다.

② 위원장은 법 제28조 제2항 각 호에 따른 업무 수행을 위하여 정기적으로 회의를 소집한다.

③ 위원장은 회의를 소집하고자 할 경우 회의의 일시·장소, 회의 안건과 심의자료를 회의 개최일 7일 전까지 각 위원에게 서면 또는 전자적 방식으로 통지하여야 한다.

④ 위원장은 다음 각 호의 어느 하나에 해당하는 경우 임시회의를 소집할 수 있다.

1. 법 제28조 제1항의 기관 또는 단체의 장이 회의 소집을 요구하는 경우
2. 위원의 3분의 1 이상으로부터 소집 요구가 있는 경우
3. 법 제28조 제2항 각 호에 따른 업무수행을 위하여 위원장이 필요하다고 인정하는 경우

⑤ 제4항에 따라 임시회의를 소집할 때에는 위원장이 회의 안건과 심의자료를 회의 개최 48시간 전까지 회의에 참석하는 위원들에게 송부하여야 한다. 단, 긴급하거나 부득이한 사유가 있는 경우 위원 전원의 동의로 사전 통지 없이 회의를 개최할 수 있다.

⑥ 임시회의는 사전 고지한 안건에 한하여 심의·의결한다 단, 출석위원 3분의 2 이상이 동의하는 때는 안건을 추가할 수 있다.

⑦ 회의는 대면회의를 원칙으로 한다. 다만, 위원장이 긴급하거나 부득이한 사유로 특별히 필요하다고 인정하는 경우에는 비대면회의로 이를 대체할 수 있다.

제8조(의견 청취) 회의에 상정된 안건의 심의 및 의결을 위하여 필요하다고 인정하는 경우 민간자율위원회는 관계부서의 장, 임직원, 또는 외부전문가 등을 출석하게 하거나 서면으로 의견을 들을 수 있다. 단, 관계 전문가의 회의 출석에 대해서는 해당 전문가의 출석 및 의견 진술에 필요한 범위에서 제7조 제3항 및 제5항의 규정을 준용한다.

제9조(의사결정) ① 회의는 재적위원 과반수의 출석으로 개의한다. 그러나 민간자율위원회는 회의 개최 및 심의, 의결이 긴급하게 필요하나 본 정족수 충족이 어려운 불가피한 사정이 있는 경우 법 제28조 제3항의 요건을 준수하는 범위에서, 그 결의로 구성 및 인원 등을 고려하여 회의 개최를 위한 의사정족수를 달리 정할 수 있다.

② 민간자율위원회는 해당 안건에 관하여 출석위원 과반수의 동의로 의결한다. 그러나 민간자율위원회는 그 결의로 구성 및 인원 등을 고려하여 의결정족수를 달리 정할 수 있다.

③ 민간자율위원회는 결의된 내용이 기관 또는 단체에서 효과적으로 이행 및 실행될 수 있도록 안건을 가급적 구체적이고 명확하게 구성하고 결의하여야 한다.

제10조(위원의 제척, 기피, 회피) ① 회의 안건에 관하여 특별한 이해관계가 있는 위원은 회의에서 제척된다.

② 해당 회의를 통하여 심의 및 의결된 결과로 인하여 그 업무 수행에 중대한 영향을 받는 자는 회의 위원에게 공정한 심의, 의결을 기대하기 어려운 특별한 사정이 있는 경우 위원회에 기피 신청서를 제출하여 해당 위원에 대한 기피 신청을 할 수 있다. 이 경우 위원장은 기피신청 사유가 정당한지 검토하여 기피신청 인정 여부를 결정한다.

③ 위원은 위 제1항 및 제2항의 사유에 해당하는 경우 위원회에 회피신청서를 제출하여 스스로 해당 안건의 심의, 의결을 회피할 수 있다.

④ 위원의 제척, 기피 또는 회피로 인하여 회의가 의사정족수 또는 의결정족수를 충족하지 못하는 경우 위원장은 위원을 보충 지명 또는 위촉하여야 한다.

⑤ 본 조의 위원의 제척, 기피, 회피에 관한 규정은 제8조에 따른 관계전문가 등에 대해서도 준용된다.

제11조(회의록) ① 간사는 회의록을 작성하여 관리하여야 한다.

② 간사는 부득이한 경우를 제외하고 회의 종료 후 15일 이내에 회의록을 작성하여야 하며, 위원장이 회의록을 확정하기 전까지 회의에 참석한 모든 위원의 확인을 받아야 한다.

③ 위원장은 위원들의 수정 및 이의제기를 검토하여 필요한 경우 회의록을 수정하고, 최종안에 대하여 승인함으로써 정식 회의록으로 채택한다.

④ 간사는 차기 회의에서 직전 회차의 회의록 내용을 민간자율위원회에 보고하여야 한다.

⑤ 위원회가 공개의 필요성을 인정하여 의결한 경우에는 회의록의 전부 또는 일부를 공개할 수 있다. 이 경우에도 개인정보보호법에 따른 민감정보 및 고유식별정보, 영업비밀 및 경영상 중요한 정보는 제외하고 공개하여야 한다.

제12조(후속 조치) ① 민간자율위원회는 결의된 안건을 해당 기관 또는 단체의 구성원 및 이해관계자들에게 공지할 수 있다. 이 경우 공지 내용은 안건 내용에 추가로 해당 안건이 결의된 배경, 취지 및 목적, 이행방안 등의 내용을 합리적으로 포함하여 구성원 및 관계자들이 이를 이해하고 이행할 수 있도록 조치할 수 있다.

② 기관 또는 단체는 제1항에 따라 공지된 내용을 그 직후 개최되는 민간자율위원회에서 보고하거나 별도로 위원들에게 개별 통지한다.

제13조(인공지능 윤리원칙 준수 보고서) ① 민간자율위원회는 해당 기관 또는 단체의 인공지능 윤리원칙 준수를 위한 활동 및 성과를 정리한 보고서(이하 “윤리원칙 준수 보고서”라 한다)를 정기적 또는 부정기적으로 발행 및 공개할 수 있다.

② 제1항의 보고서에는 법 제28조 제2항 각 호와 관련된 기관 또는 단체의 활동을 포함할 수 있다. 단 민간자율위원회는 보고서 작성의 취지와 목적, 해당 기관 또는 단체의 상황을 고려하여 그 범위 및 내용을 자율적으로 정할 수 있다.

③ 민간자율위원회는 제1항의 보고서를 해당 기관 또는 단체의 홈페이지에 게시하거나 그에 준하는 방법으로 외부 또는 이해관계자들에게 공개할 수 있다. 〈끝〉

부록 2: 민간자율인공지능윤리위원회에 관한 1차 전문가 조사지

ID					

민간자율인공지능윤리위원회에 관한 전문가 조사

KISDI 정보통신정책연구원

안녕하십니까?

정보통신정책연구원의 디지털사회전략연구실의 연구책임자입니다.

과학기술정보통신부와 정보통신정책연구원에서는 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하, 「인공지능기본법」) 제28조 제4항에 따라 '민간자율인공지능윤리위원회'(이하 '민간자율위원회')의 구성·운영에 관한 표준지침을 개발하고 있으며, 본 전문가 조사를 통해 민간자율위원회의 **1) 인적 구성 방식**, **2) 사회적·윤리적 전문성의 종류**, **3) 외부 전문가 확보의 현실성**, **4) 위원회의 실효성 확보 방안**과 관련하여 귀하의 고견을 청하고자 합니다.

귀하의 의견은 표준지침 개발의 귀중한 기초자료로 활용됩니다. 전문가의 소속기관에 대한 정보는 직접적으로 보고서에 활용되지 않습니다. **소속 기관을 밝히기 어려운 분께서는 1) 학계, 2) 산업계(민간), 3) 공공, 4) 시민단체 중 택1**하여 적어주십시오.

본 조사 참여에 대해 소정의 전문가 자문비가 지급될 예정입니다. 조사 완료 후 자문비 지급을 위해 이메일을 보내드릴 예정이오니 링크에 접속하셔서 필요한 정보를 채워주시기를 바랍니다. 바쁘시더라도 성심껏 응답해 주시기를 바랍니다.

※ 본 조사와 관련된 문의나 의견이 있으시면 아래로 연락해 주시기를 바랍니다.

2025.10.27.

조사 발주기관		**조사 문의처**	**이시직 부연구위원** (043-531-4357, potential47@kisdi.re.kr)
조사 주관기관	**정보통신정책연구원**	**조사 책임자**	**김휘홍 부연구위원** (043-531-4290, hwihong@kisdi.re.kr)

<전문가 인적 사항>

소 속	
직 위	
전 공	
성 명	

- 1 -

[민간자율위원회 취지의 이해]

「인공지능기본법」 제28조 제1항은 인공지능 윤리원칙의 준수를 위해 특정 기관 또는 단체가 '민간자율인공지능윤리위원회'(이하 '민간자율위원회')를 자율적으로 설치·운영할 수 있다고 규정하고 있습니다. 이는 기업이 윤리원칙의 준수 여부를 스스로 평가하고 점검하도록 유도하는 자율규제 방식에 해당합니다. 이어서 동조 제2항은 민간자율위원회가 자율적으로 수행할 수 있는 업무를, 제3항은 구성·운영의 원칙적 자율성을 규정하고 있습니다. 이처럼 법률은 민간자율위원회와 관련하여 기업의 자율성을 최대한 허용함으로써 산업 전반에의 인공지능 윤리원칙 확산을 도모하고 있습니다.

A1 전문가께서는 「인공지능기본법」 제28조의 민간자율위원회의 취지에 관하여 어떠한 의견을 가지고 계시는지, 그리고 민간자율위원회가 산업 전반으로 확산하기 위해 고려해야 할 점은 무엇이라고 보시는지 자유롭게 기술해 주십시오.

답변

[민간자율위원회의 위원 구성]

「인공지능기본법」 제28조 제3항 본문은 민간자율위원회의 자율적인 구성·운영을 규정하고 있을 뿐 위원 수에 대해서는 명시적으로 규정하고 있지 않습니다. 이에 따라 민간자율위원회의 위원 수는 인공지능사업자들이 자유롭게 결정할 수 있습니다.

A2 - 1 1) 전문가께서는 위원 수를 정할 때 고려해야 할 요소(예: 의사결정의 효율성, 기업 또는 단체의 규모와 역량 등)는 무엇이라고 생각하시는지 그리고 실제로 민간자율위원회의 기능을 효과적으로 수행하기 위해 위원 수를 어느 정도로 구성하는 것이 바람직하다고 보시는지 의견을 부탁드립니다.

답변

[민간자율위원회의 구성 방식]

A2 - 2 2) 민간자율위원회의 유연한 구성을 허용한 법률의 취지를 고려하면 '적정 위원 수'를 제시하는 방식 외에 '전문성 확보에 기반한 위원회 구성'을 생각해 볼 수 있습니다. 예를 들어,

- 대기업: 전문영역별(기술, 윤리, 법률, 경영, 고객 담당 등)로 전담 위원 배치
- 중소기업/스타트업: 한 명이 여러 전문영역을 겸임(기술+경영 또는 고객 담당+법률 등)

민간자율위원회 구성의 유연성을 유지하되, 윤리원칙의 준수라는 기능에 부합하도록 위원회의 인적 구성을 제시한다면, 전문가께서는 어떤 방식으로 제도를 설계하는 것이 좋을지 다양한 관점에서 의견을 주시기를 바랍니다.

답변

[위원의 사회적·윤리적 전문성]

「인공지능기본법」 제28조 제3항 단서는 민간자율위원회의 구성에 관하여 "사회적·윤리적 타당성을 평가할 수 있는 경험과 지식을 갖춘 사람 포함"할 것을 규정하고 있습니다. 이는 인공지능/IT 기술 분야의 전문성과는 구별되는 인문·사회과학적 전문성을 가리키는 것으로 보입니다.

A3 - 1 **1) 전문가께서는 민간자율위원회의 구성 시 어떤 전문영역이 포함되어야 한다고 생각하십니까? 예시된 전문영역 또는 전문가께서 생각하시는 전문영역의 중요도를 선택해 주시기를 바랍니다.**

전문영역	전혀 중요하지 않다	중요하지 않다	보통이다	중요하다	매우 중요하다
법률/규제 전문가	①	②	③	④	⑤
윤리/철학 전문가	①	②	③	④	⑤
시장/산업(비즈니스) 전문가	①	②	③	④	⑤
인권·사회적 책임 전문가	①	②	③	④	⑤
해당 업종(예: 의료, 금융 등) 전문가	①	②	③	④	⑤
기타(응답자 직접 기재) _____	①	②	③	④	⑤

A3 - 2 **2) 귀하께서 어떤 전문영역을 "매우 중요" 또는 "중요하다"고 생각하신 이유와 추가로 필요하다고 생각하신 영역이 있다면, 그 이유를 자유롭게 적어주시기를 바랍니다.**

답변

[외부 전문가 확보]

인공지능사업자들은 「인공지능기본법」 제28조 제1항에 따라 민간자율위원회를 자율적으로 설치·운영할 수 있으나, 동조 제3항 단서는 민간자율위원회를 설치할 경우, "그 기관 또는 단체에 종사하지 아니하는 사람"을 포함하도록 규정하고 있습니다. 또한 동조 제2항의 위원회 업무를 수행하여야 하므로 결국 해당 분야의 '전문성을 겸비한 외부인'이 필요하다고 할 수 있습니다.

A4 - 1 1)「인공지능기본법」 제28조 제3항 단서의 "그 기관 또는 단체에 종사하지 아니하는 사람"은 해당 기관과 이해관계가 없는 사람, 즉 '계약적·경제적 관계에 있지 아니한 사람'으로 이해할 수 있습니다. 이때 외부 전문가 확보에 현실적으로 어떤 어려움이 있을 것으로 예상하십니까? 법률의 취지를 존중하면서도 현장성을 확보할 수 있는 대안은 무엇이라고 보십니까?

답변

[외부 전문가와 정보제공]

A4 - 2 2)「인공지능기본법」 제28조 제2항 각호는 민간자율위원회가 자율적으로 수행할 수 있는 업무를 예시하고 있습니다(제1호: "인공지능기술 연구·개발·활용에 있어서 윤리원칙의 준수 여부 확인"). 이러한 업무를 수행하려면 외부 위원에게도 인공지능기술 연구·개발·활용에 관한 정보가 제공되어야 할 것입니다. 이때 해당 기관 또는 단체의 핵심 기술 또는 업무상 비밀이 유출될 수 있다는 우려가 제기될 수 있습니다.

전문가께서는 이러한 우려를 완화하고 동시에 민간자율위원회의 원활한 운영을 확보하기 위해 정보제공의 범위와 절차를 어떻게 설계할 수 있을지 다양한 관점에서 의견을 주시기를 바랍니다.

답변

[민간자율위원회의 실효성]

「인공지능기본법」 제28조 제1항은 인공지능사업자 등이 "윤리원칙을 준수하기 위하여" 민간자율위원회를 둘 수 있다고 규정하고 있으며, 운영 등에 필요한 사항은 공정성과 중립성을 존중하는 범위에서 자율적으로 정할 수 있습니다(제3항 본문 및 제4항).

A5-1 1) 민간자율위원회 제도의 확산을 위하여 자문기능만을 허용한다면, 귀하는 이러한 위원회가 인공지능 윤리원칙 준수에 얼마나 이바지할 것이라고 보십니까? 윤리원칙에 기반한 자문이 기관 또는 단체의 내규나 내부 정책에 내재화할 수 있으려면 추가로 어떤 조치를 고민할 필요가 있다고 보십니까?

답변

[결과보고서 작성 및 공개]

A5 - 2 2) 민간자율위원회 운영을 통한 윤리원칙 준수 노력과 성과를 대중에 알릴 수 있는 수단으로 연 1회 결과보고서(예: 투명성 보고서의 일부) 공개를 고려해 볼 수 있습니다. 물론 간략한 결과보고에 불과하더라도 중소 스타트업에는 부담으로 작용할 수 있습니다. 귀하께서는 이러한 결과보고서 공개에 관하여 어떠한 의견이신지 자유롭게 기술해 주시기를 바랍니다. 그리고 이를 활성화하기 위해 고려되어야 하는 점에 대해서도 의견을 부탁드립니다.

답변

A6 [표준지침 개발에 관한 조언]

민간자율위원회 표준지침의 개발에 관하여 앞의 문항에서 미처 기술하지 못한 조언이 있으시다면, 자유롭게 작성해 주십시오.

답변

- 응답해 주셔서 대단히 감사합니다 -

부록 3: 민간자율인공지능윤리위원회에 관한 2차 전문가 조사지

ID					

「민간자율AI윤리위원회 표준지침」 전문가 의견조사

KISDI 정보통신정책연구원

안녕하십니까,

여러 업무로 바쁘신 중에도 의견조사에 흔쾌히 응해주셔서 진심으로 감사드립니다.

정보통신정책연구원(KISDI) 연구진은 「인공지능기본법」 제28조에 따라 AI 사업자 등이 기관 또는 단체 내에 민간자율AI윤리위원회(이하, '위원회')를 구성·운영할 때 참고할 수 있도록 '표준지침(안)'을 개발하고 있습니다.

표준지침(안)은 크게 위원회 '구성'과 '운영'에 관한 13개의 조문으로 구성되어 있습니다. 위원회 구성과 관련하여, 법적 제약을 반영한 위원 수와 자격, 임기 및 운영진에 관한 조항을 두었으며, 운영 측면에서는 회의 전 과정을 절차화하여 공정성을 확보하는 한편, 윤리원칙 준수 보고서 등 후속 조치를 제안하였습니다. 「인공지능기본법」 제28조는-인적 구성에 관한 법적 제약을 제외하면-위원회의 구성과 운영을 AI 사업자 등의 자율에 맡기고 있습니다. 따라서 **본 표준지침(안)은 참고 자료일 뿐, 정답을 제시하지 않습니다**. 다만, 기업 현장의 의견을 반영할 때 비로소 '중소기업·스타트업의 지원'이라는 표준지침(안)의 목표를 달성할 수 있다고 판단하여 여러 전문가분의 고견을 청하게 되었습니다.

귀하의 의견은 오직 표준지침(안) 개발을 위한 기초자료로 활용됩니다. 또한 아래 인적 사항은 보고서 등에 활용되지 않습니다. 바쁘시더라도 소중한 시간을 할애하시어 성심껏 응답해 주시면, 최선을 다하여 건전한 AI 생태계 조성에 이바지할 수 있는 표준지침(안)을 개발하겠습니다. 조사와 관련하여 문의나 의견이 있으시면 아래로 연락해 주시기를 바랍니다.

감사합니다.

연구기관: 정보통신정책연구원 디지털사회전략연구실

문 의 처: 김휘홍 부연구위원
(043-531-4290, hwihong@kisdi.re.kr)

* 본 **'표준지침(안)'**은 의견 수렴을 위한 자료이며, **최종 결과물이 아닙니다**. 전문가분들의 의견을 반영하여 수정·보완할 예정이므로, 내부적으로만 활용해 주시고 **외부에 유출되지 않도록 주의**를 부탁드립니다.

<전문가 인적사항>	
성 명	
소 속	
직 위	
전 공	

※ 조사지와 함께 첨부한 '민간자율인공지능윤리위원회 구성 및 운영에 관한 표준지침'을 참고하셔서 다음 질문에 대한 의견을 작성해 주십시오.

의견 작성

1. 귀하께서는 표준지침(안) 각 조항에 관하여 어떻게 생각하십니까?
삭제 · 추가 · 보완이 필요한 조항과 그 이유를 작성해 주십시오.

o 개별 조항에 대한 의견

조문 번호	의견
예) 제5조 제1항	내용

* 칸이 부족하시다면, 추가하여 작성해 주십시오.

2. 귀하께서는 현재의 표준지침(안)의 내용에 관하여 전체적으로 어떻게 생각하십니까?
실효적인 표준지침(안)의 개발을 위해 연구진이 고려해야 할 점을 자유롭게 작성해 주십시오.

o 의견 작성

- 감사합니다 -

● 저 자 소 개 ●

이 현 경

- The George Washington University 정책학 박사
- 현 정보통신정책연구원 연구위원

김 휘 홍

- Ludwig Maximilian University of Munich 공법학 박사
- 현 정보통신정책연구원 부연구위원

문 정 욱

- 고려대학교 행정학 박사
- 현 정보통신정책연구원 디지털사회전략연구실 실장

조 성 은

- Rutgers University 커뮤니케이션학 박사
- 현 정보통신정책연구원 연구위원

연 소 라

- Texas A&M University 경제학 박사
- 현 정보통신정책연구원 부연구위원

이 시 직

- 연세대학교 법학 박사수료
- 현 정보통신정책연구원 부연구위원

김 지 혜

- 서강대학교 국제학 석사
- 현 정보통신정책연구원 전문연구원

양 기 문

- 연세대학교 정보시스템학 석사
- 현 정보통신정책연구원 전문연구원

강 하 연

- 충남대학교 정책학 석사수료
- 현 정보통신정책연구원 연구원

김 소 담

- 서울대학교 지능정보융합학 석사
- 현 정보통신정책연구원 연구원

권 지 혜

- 호서대학교 데이터사이언스학 석사
- 현 정보통신정책연구원 연구원

방송통신정책연구 RS-2025-16063586
정책연구 25-45

인공지능 윤리 실천수단 확산 방안에 관한 연구

2025년 12월 일 인쇄
2025년 12월 일 발행

발행인 과학기술정보통신부 장관
발행처 과학기술정보통신부
세종 가름로 194 세종파이낸스센터
Homepage: www.msit.go.kr
인 쇄 (사)아름다운사람들