ACT ON CREATING FOUNDATION FOR TRUST AND FOSTERING ARTIFICIAL INTELLIGENCE

COMMENTS FROM BSA | THE SOFTWARE ALLIANCE

AUGUST 2, 2021

BSA | The Software Alliance (**BSA**) welcomes this opportunity to provide our comments to Lawmaker Jung Pil-Mo's Office and the Science, ICT, Broadcasting and Communications Committee (**SIBCC**) regarding the bill to establish an "Act on Creating Foundation for Trust and Fostering Artificial Intelligence" (**Bill**).

BSA is the leading advocate for the global software industry before governments and in the international marketplace. Our members[6] are at the forefront of software-enabled innovation that is fueling global economic growth, including cloud computing, data analytics, and artificial intelligence (**AI**) products and services. As leaders in the development of cutting-edge technology, BSA's members have unique insights into both the tremendous potential of these new technologies and the government policies that can best support their responsible use and ensure continued innovation of such technologies.

BSA works closely with the governments around the world to promote the development of policies that encourage the responsible development and use of AI.[7] To that end, BSA has identified five key pillars for Responsible Artificial Intelligence. These pillars reflect how both industry and government have important roles to play in promoting the benefits and mitigating the potential risks involved in the development, deployment, and use of AI:

1. **Building Confidence and Trust in AI Systems:** Highlighting industry efforts to ensure AI systems are developed in ways that maximize fairness, accuracy, data provenance, explainability, and responsibility.

2. **Sound Data Innovation Policy:** Promoting data policies that are conducive to the development of AI and other new data-driven technologies including reliable legal mechanisms that facilitate cross-border data transfers, legal certainty for value-added

[6] BSA's members include: Adobe, Altium, Amazon Web Services, Atlassian, Autodesk, Aveva, Bentley Systems, Box, Cisco, CNC/Mastercam, Dassault, DocuSign, IBM, Informatica, Intel, MathWorks, Microsoft, Nikon, Okta, Oracle, PTC, Rockwell, Salesforce, ServiceNow, Siemens Industry Software Inc., Slack, Splunk, Synopsys, Trend Micro, Trimble Solutions Corporation, Twilio, Workday, and Zoom.

[7] BSA AI Policy Overview, accessible at https://ai.bsa.org/

300 Beach Road
#30-06 The Concourse
Singapore 199555

P: +65 6292 2072
F: +65 6292 6369
W: bsa.org

Regional Representative Office
UEN: S97RF0005K

services (e.g., text and data mining, machine learning), and enhanced access to non-sensitive government data.

3. **Cybersecurity and Privacy Protection:** Advocating for policies that strengthen enhanced security measures and respect informed consumer choices while ensuring the ability to deliver valuable tailored products and services.

4. **Research and Development:** Supporting investment in efforts that foster confidence and trust in AI systems, promote coordination and collaboration between industry and government, and help grow the AI workforce pipeline.

5. **Workforce Development:** Identifying opportunities for government and industry to collaborate on initiatives to prepare the workforce for the jobs of the future.

BSA acknowledges both the importance of AI and the potential risk associated with certain uses of the technology. In response to the risk of bias, BSA published a report titled "**Confronting Bias: BSA's Framework to Build Trust in AI"**[8] to provide a guide that organizations can use to perform impact assessments to identify and mitigate risks of bias that may emerge throughout an AI system's lifecycle.

GENERAL OBSERVATION

BSA is encouraged that Lawmaker Jung-Pil-Mo's Office drafted the Bill in an effort to both promote the benefits and enhance public trust in AI by establishing safeguards around high risk uses of the technology. BSA shares these objectives and offers below some recommendations that are aimed at ensuring the Bill is clear to stakeholders who will be subject to its requirements.

BSA recommends the National Assembly work closely and transparently with a range of experts, including representatives from affected companies such as BSA members, and closely consider similar international efforts regarding AI and ethics. An assessment of the international perspective and approach on how to realize trustworthy AI would ensure that Korean AI developers remain competitive in global markets while contributing to strengthen trust in new technologies.

RECOMMENDATIONS

Clearly Defining Important Concepts

"AI Users" and "AI Business Operators"

The Bill appropriately recognizes that responsibility for AI oversight will necessarily be shared by multiple stakeholders. To that end, the Bill seeks to distinguish between "AI Business Operators" and "AI Users." The term "AI Business Operators" seems intended to cover organizations that are engaged in the development, manufacture, or distribution of AI technology or services to other organizations that ultimately deploy them (i.e., AI Users). While this seems to be the intention, the definition of "AI Business Operators" (Article 2(6)) is far broader than the intention and would encompass any organization that is "engaged in economic activities related to the AI Industry". Given that organizations that *use* AI technologies or services are also engaged in "economic activities related to the AI industry," the definition of AI Business Operators could be interpreted as encompassing AI Users. Similarly, the definition of "User" (Article 2(7)) simply means "a person who uses a product or AI service(s) using AI technology", there is a risk that the concept of AI User could refer also to some AI Business Operators. The consequence of this definitional overlap will create

[8] Confronting Bias: BSA's Framework to Build Trust in AI, accessible at https://ai.bsa.org/

tremendous uncertainty for both AI Business Operators and AI Users about which organizations should be responsible for complying with the Bill's substantive requirements.

We recommend that the definitions be revised as follows:

> *6. The term "AI business operator" means a person ~~engaged in economic activities related to the AI industry~~* ***or organization engaged in the development, manufacture, or distribution of an AI product, technology, or service****.*
>
> *7. The term "user" means a person* ***or organization*** *who uses a product or AI service(s) ~~using~~* ***incorporating*** *AI technology* ***as the final user of the product or AI service(s) and not as a person or organization engaged in providing a product or AI service(s) for use by another person or organization****.*

"AI Used in Special Areas"

By focusing on "Artificial Intelligence Used in Special Areas", the Bill seeks to establish important guardrails around uses of AI that pose significant risks to the public. The definition of "Special Areas" therefore plays a critical role in distinguishing between the applications of AI that will be subject to the Bill's substantive requirements. The definition identifies sectors where the use of AI could implicate grave risks to the public in the event of a system malfunction. However, a purely sector-based approach to identifying risk would be overly broad. Even in "high risk" sectors — such as the medical field — there are low risk uses of AI. For instance, the mere use by a hospital of an AI system to streamline its payroll processes should not render the underlying software as a "Special Area" subject to regulation.

In another example of how this overbroad sector-based application of risk, Article 20(2) requires that "a person who carries out tasks with AI for special use as specified in Items E through G of Subparagraph 2 of Article 2 shall not make final evaluation or make decision only relying on AI for special use." This appears to impose an absolute prohibition against purely automated decision making in all applications of AI in the sectors covered by Items E through G. However, this would mean that even low- or no-risk applications of AI in these sectors would be prohibited, such as the use of AI by information and communication service providers (under Item G) to determine the optimal routing of Internet traffic.

With this in mind, we recommend adjusting the definition of "artificial intelligence used in special areas" (Article 2(2)) to be clear that it "means AI that **both** falls under any of the following items, **and** which may endanger human life or body, or harm human dignity…".

It will also be important, when implementing this law, for the Government to work with relevant stakeholders to provide clear guidance on how industry and the Government will assess what may "endanger human life or body, or harm human dignity" to reduce the subjectivity and uncertainty of these important concepts.

Transforming Ethical Principles into Legal Standards will Require Care

Since 2019, the Government of Korea has made important strides towards the recognition of key ethical principles for the development and use of responsible AI. For example, the Korea Communications Commission (**KCC**) and the Korea Information Society Development Institute (**KISDI**) have developed and announced the "Principles for Protection of Users in the Intelligent Information Society" which state core ethical principles for AI, including "People-centered Service", "Transparency", and "Prohibition of Discrimination". The KCC/KISDI Principles were developed in consultation with stakeholder experts from the private sector, including BSA, and were intentionally designed as voluntary measures to assist enterprises and users in advancing the ethical development

and use of AI. Moreover, in June of this year, KCC and KISDI announced the "Basic Principles to Protecting the Users of AI-Based Recommender System". This too provides a voluntary guideline for the private sector to reflect and educate themselves of the core principles to better protect individual AI users.

Unlike the prior efforts by the KCC and KISDI, this Bill appears to be aimed at transforming high-level ethical principles into legally enforceable standards. We caution, however, that although the Basic Principles set forth in Article 3 and the Ethical Principles identified in Chapter 2 of the Bill reflect universal values that all organizations should strive to adhere to in their business practices, they lack the objectivity that is needed for enforceable legal requirements. To the contrary, many of the principles identified in the Bill are inherently subjective. For instance, there is no uniform standard for determining when an AI infringes "on the good of the whole community," or "basic rights of human beings". Given the inherently subjective nature of many of the principles outlined in the Bill, there is a great risk that transforming them into enforceable legal standards will create tremendous uncertainty and unreasonable liability for organizations that may be subject to these requirements. Moreover, considering the aspirational nature of the proposed Ethical Principles, we recommend the Bill advocate establishing governance-based safeguards requiring AI Users and AI Business Operators to implement reasonable risk management processes related to the underlying principles.

Reporting Requirement for "AI used in Special Areas"

The Bill also obligates AI Business Operators to "file a report on its technical and administrative measures to protect the life and body of users and respect human dignity" when they intend to develop, manufacture, and distribute AI Used in Special Areas (Article 21). This effectively establishes a regime where the Ministry of Science and ICT (**MSIT**) must approve (or reject) uses of AI. We have several concerns with this. First, given the expansive definition of "AI Used in Special Areas" and the vast number of AI systems that may fall within its scope, a requirement for MSIT to review the technical underpinnings, including those related to development and manufacturing, of all such systems before they are made available is likely untenable. Second, advance reporting by AI Business Operators on their technical and administrative measures is unlikely to provide MSIT with meaningful insight into the risk profile of AI deployment scenarios. In many instances, AI Business Operators provide AI Users with general purpose AI tools and resources that must be customized and re-trained for the particular end-use application. Because AI Business Operators may be unable to monitor or control how AI Users deploy the underlying AI, they will often lack information about the "technical and administrative measures" that are ultimately needed to control for potential risks. Reporting obligations should therefore be contextual, risk based and, should be done during the course of deployment by the AI User and as the risks and impacts of such deployment are determined. Third, requiring all "AI Used in Special Areas" to undergo approval risks imposing a huge and disproportionate burden on the industry, especially small- and medium-sized enterprises, which would have a chilling effect on innovation in AI technology and products in Korea.

Private Autonomous AI Ethics Committees Should Operate Without Direct Government Interventions

BSA welcomes the Bill's proposal encouraging AI Business Operators to establish AI Ethics Committees (Article 16). Effective AI risk management should be underpinned by a governance framework that establishes the policies, processes, and personnel that will be used to identify, mitigate, and document risks throughout the AI system's lifecycle. The purpose of such a governance framework is to promote understanding across organizational units — including product development, compliance, marketing, sales, and senior management — about each entity's role and responsibilities for promoting effective risk management during the design, development, and deployment of AI systems. While we agree that organizations should be encouraged to establish Autonomous AI Ethics Committees, we are concerned that the MSIT certification process contemplated by Article 17 may

stifle the flexibility that is necessary for organizations to structure their policies and processes in a manner that is tailored to their unique profile. We therefore recommend eliminating the certification clauses (Article 17).

Explainability Requirements Should Be Risk-Based and Context-Specific

Article 20(3) indicates that a future Presidential Decree will set forth requirements for "a person who performs tasks by using AI for special use" to provide explanations regarding the "decision-making principle of the AI" upon request of the "other party." As AI is integrated into high-stakes decision-making processes that can have consequential impacts on the public, "explainability" has emerged as a foundational element for promoting trust. However, while there is a consensus that "explainability" can play an important role in promoting trust in AI, there is as-yet no universal understanding of what it means for a system to be explainable or the specific contexts in which explainability should be required.

For instance, the European Union's High-Level Expert Group on AI (**HLEG**) recently acknowledged that output-level explanations may not always be possible as a technical matter, and noted that "in those circumstances, other explainability measures (e.g. traceability, auditability and transparent communication on the AI system's capabilities)" can help to achieve the goal of promoting trust.[9] The HLEG also acknowledged that "the degree to which explainability is needed depends on the context and the severity of the consequences of erroneous or otherwise inaccurate output to human life."[10] Consistent with the insights of the HLEG, we urge the National Assembly to revise Article 20(3) to ensure that explainability requirements will be applied in a context-dependent fashion taking into consideration the relevant roles and responsibilities of the organization, different mechanisms by which explainability can be accomplished, and the circumstances that warrant the implementation of this obligation.

CONCLUSION

BSA appreciates the opportunity to comment on the Bill to establish an 'Act on Creating Foundation for Trust and Fostering Artificial Intelligence'. We hope this submission is useful to the consultation process. Please let us know if you have any questions or would like to discuss comments in more details. You may contact Mr. Geun Kim, Korea Country Manager at guenk@bsa.org.

[9] https://ec.europa.eu/futurium/en/ai-alliance-consultation

[10] https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

인공지능 육성 및 신뢰 기반 조성 등에 관한 법률안(11261)에 대한 BSA | THE SOFTWARE ALLIANCE 의견서

2021 년 8 월 2 일

BSA | The Software Alliance (BSA)는 인공지능 육성 및 신뢰기반 조성 등에 관한 법률안(의안번호 11261)에 대한 의견을 전달할 수 있도록 기회를 주신 정필모 의원실과 국회 과학기술정보방송통신위원회에 감사의 말씀을 드립니다.

BSA는 정부와 국제시장 보다 앞서 글로벌 소프트웨어 산업을 주도하고 있는 소프트웨어 연합입니다. BSA를 구성하는 회원들은[1] 현재 클라우드 컴퓨팅, 데이터 분석, 인공지능 제품 및 서비스 등 전 세계 경제의 성장을 가속화하고 있는 소프트웨어 기반의 혁신을 이끌어 나가고 있습니다. 이렇게 선구자 역할을 하는 BSA는 그 동안의 경험들을 바탕으로 신기술들이 가지고 있는 엄청난 성장 가능성과 이를 책임 있게 활용할 수 있도록 도우면서 지속적인 혁신도 보장할 수 있도록 진흥하는 정부의 정책들에 대해 특별한 통찰력을 가지고 있습니다.

BSA는 전세계 여러 정부와 긴밀히 협력하며 책임 있는 인공지능의 사용과 개발이 이루어질 수 있도록 지원하는 정책 개발에 힘써 왔습니다[2]. 이러한 노력의 결과로, BSA는 책임 있는 인공지능 개발에 필수적인 5개 원칙을 정립한 바 있습니다. BSA 원칙들은 산업계와 정부가 인공지능의 개발, 배치 및 사용에 수반되는 잠재적 위험을 완화하고 편익을 촉진하는 데 얼마나 중요한 역할을 하는지 적시 되어 있습니다.

1. **인공지능 시스템에 대한 믿음과 신뢰성 제고:** 공정하고, 정확하고, 정보의 출처가 분명하며, 설명 가능하고, 책임 있는 인공지능 시스템을 개발하기 위한 산업계의 노력을 부각

2. **데이터 혁신 정책:** 국경간 데이터 전송을 촉진하는 신뢰성 있는 법적 메커니즘, 부가가치 서비스에 대한 법적 확실성(예: 텍스트 및 데이터 마이닝, 기계 학습), 민감도가 낮은 정부 데이터에 대한 접근성 강화 등 인공지능 및 기타 새로운 데이터 기반 기술의 개발에 도움이 되는 데이터 정책 추진

---

[1] BSA 회원사: Adobe, Altium, Amazon Web Services, Atlassian, Autodesk, Aveva, Bentley Systems, Box, Cisco, CNC/Mastercam, Dassault, DocuSign, IBM, Informatica, Intel, MathWorks, Microsoft, Nikon, Okta, Oracle, PTC, Rockwell, Salesforce, ServiceNow, Siemens Industry Software Inc., Slack, Splunk, Synopsys, Trend Micro, Trimble Solutions Corporation, Twilio, Workday, and Zoom.

[2] BSA AI Policy Overview, accessible at https://ai.bsa.org/

300 Beach Road
#30-06 The Concourse
Singapore 199555

P: +65 6292 2072
F: +65 6292 6369
W: bsa.org

Regional Representative Office
UEN: S97RF0005K

3. **사이버 보안 및 사생활 보호:** 향상된 보안 조치를 강화하고 정보에 입각한 소비자의 선택을 존중하는 동시에 가치 있는 맞춤형 제품과 서비스를 제공할 수 있는 정책 추진

4. **연구 개발:** 인공지능 시스템에 대한 신뢰를 높이고, 산관 간 조정과 협력을 촉진하며, 인공지능 인력 파이프라인 육성에 도움이 되는 시도들에 대한 투자를 지원

5. **인력 개발:** 정부와 업계가 협력하여 미래의 일자리를 준비할 수 있는 기회 파악

또한, BSA 는 인공지능의 중요성과 인공지능의 편향성으로 발생할 수 있는 위험에 대해서도 잘 인지하고 있습니다. 편향성에 대응하기 위해, BSA 는 최근 **'편향성에 맞서다: AI 의 신뢰 구축을 위한 BSA 프레임워크'** [3] 리포트를 게재하여 인공지능을 다루는 이해관계자들이 인공지능의 라이프사이클에 걸쳐 편향성 여부에 대해 점검하고 이에 따른 위험을 선별 및 완화시키는 방법을 소개하였습니다.

개요

BSA는 정필모 의원실에서 위험성이 있는 인공지능에 대한 안전장치를 만들어 궁극적으로 인공지능에 대한 이익을 증진하는 동시에 국민 신뢰 제고하기 위한 목적으로 법안을 작성하셨다고 생각하고 있으며, 이를 것을 매우 고무적으로 생각하고 있습니다. BSA는 법안의 목적에 동감하고 있으며, 현재 법안에 적시된 요구사항을 직접 적용 받는 이해관계자들이 법안을 보다 명확히 이해하기 위해, 아래와 같은 권고사항들을 제언 드립니다.

BSA는 국회가 인공지능과 그 윤리와 관련하여 BSA 회원사들을 포함한 다양한 분야의 전문가와 투명하게 협력하고 타 선진국에서 선제적으로 진행된 사례들을 적극 고려하실 것을 권고 드립니다. 신뢰할 수 있는 인공지능을 어떻게 구현하고 평가할 것인가에 대한 국제적 시각과 시도는 글로벌 시장에서 국내 인공지능 개발자들이 경쟁력을 유지하는 동시에 기술에 대한 신뢰를 강화하는데 기여할 것으로 사료됩니다.

권고사항

법안 내 중요 개념에 대한 명확한 정의가 내려져야 합니다.

"인공지능사업자"와 "이용자"

법안은 "인공지능사업자"와 "이용자"를 별도로 정의하고 있으며, 이는 인공지능의 관리 감독에 대한 책임을 여러 이해관계자들이 분담해야 한다는 점을 적절히 반영하고 있다고 생각됩니다. 하지만 법안은 '인공지능사업자'를 인공지능 기술이나 서비스를 개발, 제조 또는 보급하는 단체(즉, 이용자)를 모두 포괄하는 의미로 정의되어 있습니다. 이에 대한 결과로 '인공지능사업자'(제2조 제6항)의 정의는 그 의도보다 훨씬 광범위하게 적용될 수 있으며, '인공지능 산업 관련 경제활동에 종사하는' 모든 단체들을 포괄하는 의미를 가지게 됩니다. 인공지능 기술이나 서비스를 '이용'하는 단체도 '인공지능 산업 관련 경제활동'을 하고 있다는 점에서, '인공지능사업자'의 정의가 '이용자'의 정의를 아우르는 정의로 해석될 여지가 있습니다. 이와 유사하게, 법안 제2조 7항에서 정의하는 "이용자"는 "인공지능기술을 활용한 제품 또는 인공지능서비스를 이용하는 자를 말한다"고 정의하고

[3] Confronting Bias: BSA's Framework to Build Trust in AI, accessible at https://ai.bsa.org/

있는데, 이는 인공지능사업자가 이용자의 정의 안에 해석될 수 있는 위험도 존재합니다.

이렇게 명확하지 않고 중복된 정의는 인공지능사업자 및 이용자 모두에게 법안의 실질적 요건을 준수할 책임이 있는 단체가 어떤 단체인지에 대해 큰 불확실성을 야기할 것입니다.

따라서 BSA 는 아래와 같이 원문을 수정하시기를 제언 드립니다:

> *6. "인공지능사업자" 란 ~~인공지능산업과 관련된 경제활동을 영위하는 자를~~ **인공지능 서비스를 개발, 제조하거나 인공지능 상품, 기술, 서비스를 유통하는 단체 및 개인을 말한다.***
>
> *7. "이용자" 란 인공지능기술이 포함된 제품 또는 서비스를 최종 이용하는 ~~자~~ **단체 및 개인을 말하며, 인공지능기술이 포함된 제품 또는 서비스의 중간단계에서 관여하는 단체 및 개인은 제외한다.***

"특수한 영역에서 활용되는 인공지능"

법안은 '특수한 영역에서 활용되는 인공지능'을 중심으로 국민에게 중대한 위험이 있는 인공지능 활용에 대한 안전장치를 구축하고 있습니다. 따라서 "특수한 영역"에 대한 정의가 이 법안의 실질적 요구사항이 적용될 인공지능을 구분하는 중요한 역할을 하게 될 것이라 사료됩니다. 현재 법안은 "특수한 영역"에 대한 정의에 따라 시스템 오작동이 발생할 경우 국민에게 중대한 위험을 초래할 수 있는 인공지능의 여러 분야를 직접 규정하고 있습니다. 그러나 위험을 식별하기 위해 임의로 영역을 규정하여 접근하는 방법은 지나치게 광범위할 규제로 적용될 것입니다. 의료 분야와 같은 '고위험' 분야에서도 낮은 위험도의 인공지능이 활용되고 있습니다. 예를 들어, 병원에서 인공지능 시스템을 활용해 급여절차를 간소화하기도 하는데, 이러한 소프트웨어가 "특수한 영역에서 활용되는 인공지능"으로 규정되어서는 안 될 것입니다.

이러한 광범위한 분야 별 위험에 또 다른 예로서, 제20조 2항은 " 특수활용 인공지능 중 제2조 2항 '마'목부터 '사'목까지의 인공지능을 사용하여 업무를 수행하는 자는 특수활용 인공지능에만 의존하여 최종적인 평가 또는 의사결정 업무를 수행하여서는 아니 된다"고 규정하고 있습니다. 이는 마목부터 사목에 해당하는 모든 인공지능 분야에서 순수하게 자동화된 의사결정 자체를 금지하는 것으로 규정 보여집니다. 이는 인터넷 트래픽의 최적 라우팅을 결정하기 위해 정보통신사업자('사' 목)가 인공지능을 이용하는 등 위험도가 낮거나 아예 없는 분야에 대한 인공지능의 적용도 금지하게 되는 조치입니다.

보다 합리적인 인공지능 활용을 위하여, "특수한 영역에서 활용되는 인공지능"(2조 2항)의 정의를 수정하여 "다음 항목 중 **모두**에 해당하며, 인간의 생명이나 신체를 위태롭게 하거나 인간의 존엄성을 해칠 수 있는 인공지능을 의미한다"는 점을 명확히 할 것을 권고 드립니다.

또한 정부는 이 법의 시행에 맞추어 관련 이해관계자들과 협력하여 중요한 개념인 제2조 2항 "사람의 생명·신체에 위험을 줄 수 있거나 부당한 차별 및 편견의 확산 등 인간의 존엄성을 해칠 위험이 있는 인공지능"의 주관성과 불확실성을 줄이기 위해 이를 어떻게 평가할 것인지 명확한 지침을 제공하는 것이 중요할 것입니다.

인공지능 윤리 원칙의 법률화는 보다 신중하게 접근해야 합니다.

정부는 2019년부터 신뢰가능한 인공지능의 개발과 활용을 위한 핵심 윤리원칙의 인식을 위해 중요한 진전을 이뤄왔습니다. 예를 들어, 방송통신위원회와 정보통신정책연구원은 '사람중심 서비스', '투명성', '차별 금지' 등 인공지능의 핵심 윤리원칙을 담은 '이용자 중심의 지능정보사회를 위한

원칙'을 개발하고 발표했습니다. 방송통신위원회와 정보통신정책연구원이 발표한 원칙은 BSA 등 민간 이해관계자 전문가와 협의해 개발됐으며 기업과 이용자가 인공지능의 윤리적 발전과 이용을 증진할 수 있도록 지원하기 위한 자발적 방안으로 개발되었습니다. 더욱이 두 기관은 올해 6월 'AI 기반 추천 서비스 이용자 보호 기본원칙'을 발표했다. 이 원칙도 마찬가지로, 인공지능 이용자를 보호하기 위해 민간이 핵심 원칙을 스스로 반영하고 교육할 수 있는 가이드라인을 작용하고 있습니다.

하지만 방송통신위원회와 정보통신정책연구원의 이전 노력과는 달리, 이 법안은 높은 수준의 윤리 원칙을 법적으로 집행 가능한 기준으로 변경하려는 목적이 보여집니다. 제 3 조에 명시된 기본 원칙과 법안의 제 2 장에 명시된 윤리 원칙은 모든 단체가 사업 관행에서 준수해야 하는 보편적 가치를 반영하고 있지만, 집행 가능한 법적 요건에 필요한 객관성이 결여되어 있음을 주의해야 합니다. 반대로, 법안에 추가된 원칙들은 본질적으로 주관적인 사항들이 많습니다. 예를 들어, 인공지능이 "전체 공동체"의 "이익"을 언제 침해하는 것인지에 판단할 수 있는 공통된 기준이 없습니다. 이처럼 본질적으로 주관적인 성격을 띄고 있는 여러 원칙들이 집행 가능한 법적 표준으로 전환되게 되면, 이로 인해 법적 요구사항의 대상이 될 수 있는 여러 단체에게 엄청난 불확실성과 불합리한 책임이 발생할 위험이 큽니다. 따라서, 제안된 윤리원칙대신 인공지능사업자 및 이용자가 합리적인 위험관리를 이행하도록 하는 등 거버넌스 기반의 안전장치 구축에 중점을 둘 것을 권고 드립니다.

'특수한 영역에서 활용되는 인공지능'의 신고 의무

법안은 '특수한 영역에서 활용되는 인공지능'을 개발, 제조, 유통하고자 하는 인공지능사업자에게 '이용자의 생명과 신체를 보호하고 인간의 존엄성을 존중하기 위한 기술·관리적 조치'(제 21 조)를 의무적으로 마련하도록 하고 있습니다. 이는 과학기술정보통신부가 인공지능의 사용을 승인(또는 거부)해야 하는 체제가 구축된 것입니다. BSA 는 이 조항과 관련하여 아래와 같은 우려사항을 전달합니다. 첫째, '특수한 영역에서 활용되는 인공지능'이라는 광범위한 정의와 그 범위에 포함될 수 있는 개발 및 유통에서 사용되는 인공지능을 포함한 방대한 양의 인공지능 시스템들을 감안할 때, 과학기술정보통신부가 모든 인공지능 시스템의 기술적 토대를 사전에 검토해야 한다는 요구는 실행하기 어려울 것으로 보여집니다. 둘째, 인공지능사업자가 신고를 통해 선제적으로 제공하는 기술적·관리적 조치는 인공지능 배치 시나리오에 따른 위험 프로필에 대한 의미 있는 통찰력을 과학기술정보통신부에 제공하기 어렵습니다. 대부분의 경우, 인공지능사업자는 이용자가 특정 최종 사용 애플리케이션에 맞게 맞춤화하고 재배치할 수 있도록 하는 범용 인공지능 개발 도구와 자원을 제공합니다. 인공지능사업자는 이용자들이 인공지능 개발 도구들을 어떻게 배치하는지를 모니터링하거나 통제하지 못할 수 있기 때문에, 잠재적 리스크를 완벽히 통제하는 데 필요한 '기술적·행정적 조치'에 대한 정보가 부족한 경우가 많습니다. 따라서 신고의 의무는 상황에 따라 위험에 기반해야 하며, 이용자가 인공지능을 유통하는 과정과 해당 유통 과정에서 위험과 영향이 결정되는 시기에 수행되어야 합니다. 또한, 모든 '특수한 영역에서 활용되는 인공지능'에 이러한 승인을 받도록 하는 것은 산업, 특히 중소기업에 크고 불균형적인 부담을 줄 것이며, 이는 국내 인공지능 기술 및 제품 혁신에 심각한 악영향을 가져올 것입니다.

민간자율인공지능윤리위원회는 정부의 개입 없이 자발적으로 운영되어야 합니다.

BSA 는 법안이 인공지능사업자에게 자체적으로 민간자율인공지능윤리위원회를 설치(제 16 조)하도록 독려하는 것을 지지합니다. 효과적인 인공지능의 위험 관리는 인공지능

시스템의 라이프사이클 전체에 걸쳐 위험을 식별, 완화 및 문서화하는 데 사용될 정책, 프로세스 및 인력을 확립하는 관리 체계가 뒷받침되어야 합니다.

이러한 관리 체계의 목적은 제품 개발, 컴플라이언스, 마케팅, 영업 및 고위 경영진을 포함한 조직 단위 전반에 걸쳐 인공지능 시스템의 설계, 개발 및 배치 단계에서 효과적인 위험 관리를 촉진하는 각 기업의 역할과 책임에 대한 이해를 증진하는 것입니다. 결론적으로, 민간자율인공지능윤리위원회 설립을 장려해야 한다는 법안에 내용에는 공감하지만, 제 17 조에 따라 과학기술정보통신부가 이를 인증할 수 있도록 하는 절차는 단체가 고유한 프로파일에 맞게 정책과 프로세스를 구조화하는데 필요한 유연성을 억압할 수 있다는 우려가 있습니다. 따라서, 민간자율인공지능윤리위원회의 인증 조항(제 17 조)을 삭제하실 것을 권고 드립니다.

설명의 의무는 위험에 기반해야 하며, 상황에 따라 구체적으로 규정되어야 합니다.

법안의 제 20 조 3 항은 '특수활용 인공지능을 사용하여 업무를 수행하는 자'가 '상대방'의 요청이 있는 경우에는 해당 '인공지능이 갖는 의사결정 원리 및 최종결과 등 대통령령이 정하는 사항'을 설명하도록 규정하고 있습니다. 인공지능이 사람들에게 큰 영향을 미칠 수 있는 고도의 의사결정 과정에 포함되면서, '설명가능성'이 인공지능의 신뢰를 높이는 기본 요소로 떠오르고 있습니다. 이처럼 인공지능에 대한 신뢰를 높이는 데 '설명가능성'이 중요한 역할을 할 수 있다는 공감대가 형성 있지만, 시스템이 무엇을 의미하는지 또는 설명가능성이 요구돼야 하는 구체적인 상황에 대해서는 보편적으로 공감대가 이루어지지 않고 있습니다. 예를 들어, 유럽연합의 인공지능 고위 전문가 그룹(European Union's High-Level Expert Group on AI, HLEG)은 인공지능의 신뢰성을 높이기 위해[4], 최근 결과물 수준의 설명이 기술적 문제로 항상 가능하지 않을 수 있음을 인정하고 "그런 상황에서는 다른 설명 조치(예: 인공지능 시스템의 역량에 대한 추적성, 감사성, 투명한 커뮤니케이션)가 도움이 될 수 있다"고 언급한 바 있습니다. 또한 "설명가능성이 필요한 정도는 상황에 따라 다르며, 인간의 생활에 대한 부정확한 결과의 심각성에 따라 달라진다"[5] 고 인정했습니다. HLEG 의 주장과 같이, 법안 제 20 조 3 항의 설명가능성은 관련 단체의 역할과 책임을 규정하고, 설명가능성을 달성할 수 있는 다른 방법들도 함께 제시하며, 상황에 따라 책임이 달라지는 방식으로 법제화되어야 합니다.

결론

마지막으로, 회원사들을 대표하여 ' 인공지능 육성 및 신뢰 기반 조성 등에 관한 법률안'에 대한 의견을 개진할 기회를 주신점에 대해 감사의 말씀을 드립니다. BSA 는 이 의견서가 법안 논의 과정에서 유용한 참고자료가 되기를 희망하고 있습니다. 의견서 관련 질문이 있으시거나, 더 자세한 내용에 대해 논의가 필요하신 경우, 아래 연락처로 편하게 회신주시기 바랍니다. 감사합니다.

김근 BSA 코리아 대표 guenk@bsa.org

[4] https://ec.europa.eu/futurium/en/ai-alliance-consultation

[5] https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment