

Article

Introduction of the First AI Impact Assessment and Future Tasks: South Korea Discussion

Jonggu Jeong 

Law Research Institute, Seoul National University, 1 Gwanak-ro, Seoul 08826, Gwanak-gu, Korea; lawecon@snu.ac.kr

Abstract: South Korea introduced the artificial intelligence impact assessment and was the first case of introducing the artificial intelligence impact assessment as national-level legislation. Artificial intelligence impact assessments will be helpful in deciding whether to introduce artificial intelligence by comparing costs and benefits. However, South Korea's approach had limitations. First, an impact assessment was introduced only in the public sector. Second, artificial intelligence impact assessments were voluntary. Third, the subject of artificial intelligence impact assessments was limited to society. Fourth, it is necessary to establish a relationship with other impact assessments. Fifth, specific details were incomplete.

Keywords: AI; AI impact assessment; intelligent information service; South Korea



Citation: Jeong, Jonggu. 2022. Introduction of the First AI Impact Assessment and Future Tasks: South Korea Discussion. *Laws* 11: 73. <https://doi.org/10.3390/laws11050073>

Academic Editor: Esther Salmerón-Manzano

Received: 27 June 2022

Accepted: 23 September 2022

Published: 29 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

This study aims to examine the artificial intelligence impact assessment (2020) introduced in South Korea. Artificial intelligence has brought enormous benefits to mankind by increasing productivity (OECD 2022; Tu et al. 2022), but it can lead to costs in terms of human rights infringement, social problems, and environmental destruction (Weidinger et al. 2021; Stanford HAI 2021). Artificial intelligence impact assessments will be helpful in deciding whether to introduce artificial intelligence by comparing costs and benefits. This is a prerequisite for achieving sustainable development in a society in which artificial intelligence brings great change. Sustainable development will be possible only when the benefits from the introduction of artificial intelligence outweigh the costs.

This study assumes that the benefits that can be obtained through artificial intelligence are larger than the cost. This is also the reason why many countries invest huge amounts of money in establishing national strategies for artificial intelligence and leading the industry (European Commission 2019) (Stanford HAI 2021). However, it is not yet clear what the possible costs will be. There are only sporadic discussions going on. First, we introduce what artificial intelligence is. Second, we systematically analyze the cost aspect of artificial intelligence through a literature review of previous papers. Third, we present the status of the discussion of artificial intelligence impact assessments in chronological order through a case study. Fourth, a case study will be conducted on the artificial intelligence impact assessment (2020) introduced in South Korea as a legal amendment in 2020. Fifth, we suggest limitations and future challenges of the South Korean artificial intelligence impact assessment legislation.

2. The Emergence and Development of Artificial Intelligence Technology (Benefit)

2.1. How Artificial Intelligence Works

Artificial intelligence is a problem-solving process that derives results through data processing. It can be understood as a kind of function. When data are input as an input variable, it encompasses a series of processes of mapping the output variable by applying an algorithm (Russell and Norvig 2022). In natural language communication, the input

variable is a natural language as a query, and the output variable is a natural language as an answer. In computer vision, the input variable is an image, and the output variable is a description of the image. In the recommendation system, the input variable is the user's behavior data, and the output variable is the suggestion according to the user's characteristics.

Artificial intelligence is sometimes divided into strong artificial intelligence and weak artificial intelligence based on which problems can be solved by artificial intelligence. Weak AI is artificial intelligence that can only perform tasks well for a specific purpose. On the other hand, a strong artificial intelligence is an artificial intelligence that can solve a variety of problems. Unlike weak artificial intelligence, strong artificial intelligence is often referred to as artificial general intelligence (AGI) in that it has a general purpose for solving problems. Artificial intelligence developed and used today has not yet reached strong artificial intelligence and is evaluated to be at the level of weak artificial intelligence (Surden 2019).

2.2. Rule-Based Method (Expert System)

Until the advent of machine learning, artificial intelligence was mainly implemented in a rule-based way. The rule-based method is a method in which a person inputs rules for a given situation in advance by coding a computer program and provides answers to the presented questions based on the input rules. At this time, many of the rules were inputted by experts in various fields, including law, medical care, and finance, through an apprenticeship curriculum.

Expert systems were studied from the 1980s to the 1990s, focusing on fields such as law, medical care, and finance. However, in general, these expert systems have been evaluated as not very successful. Cases other than the entered situation could not be handled. As the number of rules increased, management became more difficult. In the end, due to limited scalability, difficult management, and performance limitations, the expert system did not perform as expected by society, which was a decisive factor in the stagnation of artificial intelligence research for a while (Taulli 2021).

2.3. Learning-Based Method (Introduction of Machine Learning)

To overcome the limitations of the expert system introduced above, a machine learning technique was introduced. Machines are often not very good at tasks that humans can easily perform. This is called Moravec's paradox (Moravec 1990), which is the paradox that what is easy for humans is difficult for computers, and conversely, what is difficult for humans is easy for computers. This comes from the fact that it is difficult to code a machine to understand what a human could easily perform using intuition. Machine learning is a series of problem-solving methods that derive functions by learning on their own based on sample data. Even without a human expert inputting rules one by one, a function is inferred from the sample data, which is called learning. As a result, the machine was able to learn even things that were difficult for humans to code by regularizing, and it was possible to achieve good results in problems that were difficult to solve with the existing rule-based method (Goodfellow 2016).

2.4. The Use of Artificial Neural Networks and the Emergence of Deep Learning

An artificial neural network is an engineering implementation that simulates the information transmission system of a human brain neural network. An artificial neural network is a network structure in which artificially generated neurons are connected in several layers, consisting of an input layer, one or more hidden layers, and an output layer. Each layer is composed of several artificial neurons, and each artificial neuron works by applying an activation function to an input value to derive a result value and then transmitting it to the next neuron. Each artificial neuron multiplies each input value by a weight, adds them, then inputs the value to the activation function and transmits the result to the next neuron. Such an artificial neural network seeks the weight value between

connected artificial neurons so that the error between the value of the output variable derived from the activation function and the value of the true output variable of the given data is minimized, such as in machine learning. Deep learning is a technique for learning a deep neural network based on an artificial neural network (Taulli 2021).

Deep learning is widely used to recognize complex patterns in text, voice, image, and video in that it can process unstructured data that were previously difficult to handle. In the ImageNet Challenge, which is a famous task in the field of computer vision, it was confirmed that deep learning can significantly lower the error rate, and deep learning came into the spotlight. In 2011, before deep learning was applied, the error rate, which was equivalent to 25%, decreased significantly with the application of deep learning and then dropped to 3.57% in 2015. The processing of unstructured data has been recognized as a difficult task to be dealt with by simple machine learning. This seems to be easily recognized by going through several layers. After deep learning became very successful in the field of image recognition, it was widely used in various fields, such as natural language processing and autonomous vehicles (Goodfellow 2016).

2.5. The Emergence and Development of Large-Scale Language Models

The natural language processing technique used in the mainstream language model today is transformer, which was announced by Google in 2017. The recurrent neural network (RNN) model, which is a previously used language model, performed calculations sequentially, so even when analyzing a single sentence, calculations had to be repeated several times. As a result, parallel processing techniques using multiple computing units cannot be used, and there is a limit to efficiently learning artificial neural networks. On the other hand, the Transformer model uses an attention mechanism (Vaswani et al. 2017), which gives different weights considering which word to pay attention to even when learning one sentence.

According to the attention mechanism, it is not necessary to perform calculations sequentially. Thus, parallel processing using multiple arithmetic units becomes easy. As a result, in training the artificial neural network, fewer resources are required, and the artificial neural network can be trained at high speed. Moreover, the artificial neural network model using the attention mechanism performs better than the recurrent neural network model. This is because the existing recurrent neural network did not learn the relationship between words that are separated from each other in a sentence due to long-term dependencies, whereas the attention mechanism does not process the input sequentially but should focus on a certain word. This is because the problem was overcome by assigning different weights according to their importance (Joshi 2019).

Transformers became the basis of the mainstream language models that appeared later. Google's BERT (Devlin et al. 2018) model (Devlin et al. 2018) extended the encoder part of the transformer, and OpenAI's GPT (Radford et al. 2018, 2019) model (Radford et al. 2018, 2019; Brown et al. 2020) extended the decoder part. BERT, announced by Google in 2018, is a pre-learning model that was trained using a large amount of training data by expanding the encoder part of the transformer to increase the size of the artificial neural network. Since the introduction of BERT, it has been widely recognized as showing the highest level of performance (SOTA) in various natural language understanding (NLU) tasks. On the other hand, GPT, which OpenAI announced sequentially from 2018 to 2020, is a pre-learning model built by expanding the decoder part of the transformer, increasing the size of the artificial neural network, and inputting a large amount of learning data.

While BERT focuses on natural language understanding (NLU), GPT focuses more on natural language generation (NLG). GPT-3 significantly increased its scale by building a super-large language model that is 1000 times larger than GPT-1 and 100 times larger than GPT-2. Through this, we succeeded in implementing a very small data learning technique (i.e., few-shot learning) that can train an existing language model according to the task by presenting only a few examples. As a result, a high-performance and versatile natural language processing model that can be used for general purposes without fine-tuning

has emerged. The language model of GPT-3 is known at the level of writing novels and coding programs. In the second half of 2021, the deep learning model in which transfer learning occurs by being pre-trained with a large amount of data was called a large-scale AI language model (foundation model). This seems to be a paradigm shift following machine learning and deep learning.

3. Problems Arising from the Development of Artificial Intelligence Technology (Cost)

3.1. Human Rights Risks

First, the issue of fairness was raised in terms of non-discrimination. In the case of incorrect labeling in supervised learning (e.g., the ImageNet Roulette case) (Crawford and Paglen 2019; Ruiz 2019) and in unsupervised learning, the case where the training dataset is biased (e.g., the Amazon Recruitment Algorithm case and the COMPAS case) are typical (Lauret 2019; Angwin et al. 2016). Famous biases of learning data, such as misrepresentation, underrepresentation, and overrepresentation, are externally expressed and can lead to biased expression, hate speech, demeaning speech, and performance differences.

Second, the issue of transparency was raised in terms of the information subject's right to be explained (Kaminski 2019b; Aslam et al. 2022). This is because artificial intelligence can perform well but can have poor explanatory power. Large-scale AI language models are more problematic than existing models. This is because a large-scale AI language model may provide a plausible explanation and not be able to provide a true explanation. Artificial intelligence is a kind of data processing technology. Even if the performance is excellent, if the decision-making process cannot be explained, it is difficult to adequately address the violation of the rights of the affected person (Mittelstadt et al. 2018).

The question of accountability was also raised (Kaminski 2019a). If you inflict harm on people, you will be held liable. In the case of indirect discrimination based on sensitive attributes, responsibility for indirect discrimination is limited only in cases specified in individual laws. However, in the process of using artificial intelligence, although it is difficult to prove the exact causal relationship and damage, problematic situations may arise. To prevent this case, AI ethics imposes accountability on developers and operators. Responsibility is a concept that includes ex ante control as opposed to legal responsibility or moral responsibility, which frequently emphasizes the ex post aspect.

3.2. Social and Environmental Risks

Artificial intelligence technology will develop and become more widely accepted by society. However, artificial intelligence could increase the unemployment rate and reduce wages. As such, human labor can be partially replaced by machines, as in the example of cobot (Lee 2018). Automating these tasks can have a negative impact on employment (Webb 2019). For example, the number of customer service staff will decrease by 2029 as the number of automated devices increases (US Bureau of Labor Statistics 2021). A different perspective is also presented. Jobs are likely to be replaced by artificial intelligence, but this is only for the long run (Lambert and Cone 2019). A bigger risk is that advances in artificial intelligence technology could create a relatively large wage gap between new high-paying jobs and low-wage jobs that are threatened by replacement (Salomons 2019). This aspect can be detrimental to groups with limited access to technology (Sambasivan and Holbrook 2018).

Artificial intelligence technology can produce content at a low cost in a short amount of time, which used to be produced in a small amount while investing a lot of time. As a result, the profitability of innovative, creative activities may decrease. This feature is particularly noticeable in large-scale AI models. A famous example is DALL·E 2, announced by OpenAI in 2022. Existing large-scale artificial intelligence language models just wrote naturally. However, DALL·E 2 draws a picture when you input a natural language. Rather than selecting a photo of an actual object, DALL·E 2 draws a picture even if the object is an

imaginary existence. Creativity, which had been perceived as a solo stage for humans, was threatened.

Artificial intelligence uses a significant amount of energy in its learning and operation processes. The model must be trained by processing large amounts of various data. Large-scale AI models, which have been widely developed since 2021, will consume more energy than the existing model. This can result in significant environmental costs (Bender et al. 2021). Two main reasons are suggested for this. First, a significant amount of energy is required to train and operate the model. The model emits a lot of carbon (Patterson et al. 2021). Second, in order to sufficiently cool the data center in the computational operation, huge amounts of cooling water must be supplied (Mytton 2021).

4. Preceding Discussion for AI Impact Assessment

4.1. *The Emergence and Development of Cost-Benefit Analysis*

As we have seen so far, advances in AI technology bring not only benefits but also costs. These costs occur not only in terms of human rights but also in terms of society/environment. Artificial intelligence technology can become a useful tool only when the benefits outweigh the costs. This approach is called a cost-benefit analysis, and a typical example is an environmental impact assessment (EIA). Once the environment is destroyed, it cannot usually be restored to its original state. Even if it is possible to restore it, remedies can be expensive. Therefore, the environmental impact assessment was first introduced in the United States in 1970 to establish an environmentally sound business plan while considering both the benefits and costs that can be obtained through the actions that affect the environment (Canter 1982).

In addition, cost-benefit analysis is being implemented in various ways in many fields. Gender impact assessment (GIA) is the process of comparing and assessing, according to gender-relevant criteria, the current situation and trend with the expected development resulting from the introduction of the proposed policy (EIGE 2016). Privacy impact assessment (PIA) is a systematic assessment of a project that identifies the impact that the project might have on the privacy of individuals and sets out recommendations for managing, minimizing, or eliminating that impact (OAIC 2021). There are also other safety impact assessments.

4.2. *Introduction of Risk-Based Approach in the Field of Artificial Intelligence Technology*

Cost-benefit analysis is also being introduced in the field of artificial intelligence. The benefit at this time is the improvement of welfare obtained by artificial intelligence technology. Additionally, the cost at this time is the human rights infringement or the risk in the social/environmental field caused by using artificial intelligence technology. This has been discussed in the field of artificial intelligence ethics for the past five years. As such, the discussion of risk-based approaches by introducing cost-benefit analysis to the field of artificial intelligence technology has a short history.

This was discussed in four aspects. The first is the personal information impact assessment. The European Union's Data Protection Impact Assessment (2018) is a famous example. The second is an artificial intelligence risk assessment tool. The Canadian Government's Canadian Algorithmic Impact Assessment Tool (2019) and the European Union's High-Level Expert Group on AI's Assessment List on Trustworthy Artificial Intelligence (2020) are notable examples. The third is artificial intelligence human rights impact assessment. The European Union's Human Rights, Democracy and Rule of Law Impact Assessment of AI system (2021) is a famous example. The fourth is the AI Impact Assessment Act (draft). The Algorithmic Accountability Act (2019) in the United States is a notable example.

4.3. *Data Protection Impact Assessment (DPIA) (2018)*

In the General Data Protection Regulation of the European Union, which came into force in 2018, key provisions related to automated decision-making systems are included.

A famous provision is Article 35 (Data Protection Impact Assessment), which states that the personal information protection impact assessment would be mandatory for high-risk personal information processing. General Data Protection Regulation (GDPR) seeks to implement meaningful oversight of artificial intelligence and other automated decision-making systems through this provision. However, the limitations of this regulation were pointed out in several aspects.

The main content of DPIA is for automated decision making and other AI-based systems in general. However, it is not an independent mechanism targeting AI itself (Nahmias and Perel 2021). Therefore, it cannot properly deal with problems other than the protection of personal information. For example, although the principle of fairness was mentioned in Article 5 (1)(a) of the GDPR and was expected to be considered in the process of carrying out DPIA, in practice, the principle of fairness was hardly applied in the process of carrying out DPIA (Kasirzadeh and Clifford 2021; Nahmias and Perel 2021; Kaminski and Malgieri 2019).

4.4. Risk Assessment Tool

The Government of Canada has been using the Canadian Algorithmic Impact Assessment Tool since 2019. This is a risk assessment tool that public institutions must implement. In other words, it is an impact assessment that Canadian government agencies must complete before using AI. The answer to the question will determine what action must be taken to mitigate the risk.

The European Union published the *Assessment List on Trustworthy Artificial Intelligence* in July 2020. The European Commission commissioned the High-Level Expert Group on AI in June 2018 and has been studying how to regulate AI. As a result, ethics guidelines for trustworthy AI were published in April 2019. Based on this, the *Assessment List* was presented.

4.5. Human Right Impact Assessment

The Commission on Human Rights of the Council of Europe issued a recommendation for *The Black Box of Artificial Intelligence: 10 Steps to Protect Human Rights* (2019). This was to suggest ways to prevent and mitigate the negative impact of artificial intelligence on human rights. The areas that the recommendation focused on included human rights impact assessment. Subsequently, the Council of Europe adopted the recommendation CM/Rec (2020) of the Ministerial Committee on Human Rights Impacts of Algorithmic Systems in 2020 (CAHAI-PDG 2021).

In the Appendix, there were *Guidelines for Response to Human Rights Impacts of Algorithmic Systems*, which aims to protect human rights and individual freedoms stipulated in the European Convention on Human Rights from technological development by providing guidelines to state and private actors in relation to the design and development of algorithmic systems. This led to the *Human Rights, Democracy and Rule of Law Impact Assessment of AI Systems* (2021) published by the AD HOC Committee on Artificial intelligence (CAHAI).

4.6. AI Impact Assessment Act

The Algorithmic Accountability Act was introduced in the US Senate in 2019 but failed to be enacted. However, as in the EU's GDPR, this is a significant example of an attempt to use the benefits of impact assessment to oversee artificial intelligence and other automated decision-making systems. Any company that uses an automated decision-making system should submit an impact assessment on fairness, bias, discrimination, and personal information protection and security. Since there are various types of automated decision-making systems, the single regulatory framework proposed by the bill is not effective enough to adequately regulate multiple AI systems. To ensure effective policy implementation, it was necessary to legislate a sectoral approach regarding supervisory regulations (Chae 2020).

It is also necessary to refer to the conformity assessment of the European Union AI Act of 2021 (draft). The European Union has previously divided artificial intelligence into four types (that is, unacceptable risk/high risk/limited risk/minimal risk). Among them, to use high-risk AI, a conformity assessment was required in advance. This is different from evaluating the impact of its use as it only targets the technology itself, such as safety certification. However, given that it is a mandatory evaluation to use artificial intelligence technology, it is necessary to consider the conformity assessment in preparing the impact assessment bill in the future.

4.7. Summary

Artificial intelligence impact assessment is the implementation of artificial intelligence ethics in the form of an impact assessment. As the history of artificial intelligence technology is short, the history of AI impact assessment is also not so long. However, the above legislative examples had the following limitations. First, the European Union's Data Protection Impact Assessment (2018) is a personal data protection scheme and is not aimed at artificial intelligence itself. Thus, the social and environmental risks, which can be induced by artificial intelligence, cannot be adequately dealt with.

Second, Canada's *Directive on Automated Decision-Making: Algorithmic Impact Assessment (AIA)* (2019) and the EU's *Assessment List on Trustworthy Artificial Intelligence* (2020) are difficult to understand as typical impact assessment models. Of course, this can also be viewed as an AI impact evaluation model in a broad sense. In general, the impact assessment model refers to the National Environmental Policy Act (NEPA) (1969) style. NEPA presupposes public participation through transparency and a comment framework and is mainly focused on the public sector model (Selbst 2021).

Third, the European Union's Human Rights, Democracy and Rule of Law Impact Assessment of AI Systems (2021) has a narrow evaluation target. This is because the problems arising from the development of artificial intelligence technology encompass not only human rights infringement but also social and environmental problems. To pursue sustainable development, an impact assessment model that can additionally cover society and the environment should have been introduced.

Fourth, the US Algorithmic Accountability Act (2019) was only a bill and thus had no binding power. In addition, the individual and specific risks of each AI service were not sufficiently considered, and a single regulation was applied. Later, the Algorithmic Accountability Act of 2022 was introduced, which also had the same limitations.

To overcome the limitations of these existing methods, Korea introduced AI impact assessments by amending the Framework Act on Intelligent Informatization in 2020. The contents are reviewed below, and their significance and limitations are evaluated in detail.

5. Introduction of AI Impact Assessment in South Korea (2020)

5.1. Introduction of Social Impact Assessment on Intelligent Informatization Services

In the process of providing services to users, artificial intelligence systems may have negative impacts on the environment or society owing to unintended consequences. Even though a bypass design is possible with the same performance, excessive energy consumption can cause excessive environmental costs by increasing carbon emissions. In the process of using artificial intelligence systems in society, they can replace a lot of work performed by people, which can lead to social problems such as unemployment and poverty.

To achieve sustainable development by maximizing the benefits of artificial intelligence and minimizing the costs, South Korea introduced a social impact assessment on intelligent informatization services. In the Framework Act on Intelligent Informatization in 2020, Article 56 (Social Impact Assessments of Intelligent Information Services) is an impact assessment that only targets artificial intelligence (South Korea 2020). This is the world's first case of impact assessment legislation implemented based on the NEPA style.

5.2. Subject and Target of the Social Impact Assessment on Intelligent Informatization Services

The state and local governments may survey and assess the following matters with respect to how the utilization and spread of intelligent information services have far-reaching effects on citizens' lives that affect society, economy, culture, and citizens' daily lives. The specific contents include: (1) safety and reliability of intelligent information services; (2) impacts on the information culture, such as closing the digital divide, protection of privacy and ethics for the intelligent information society; (3) impacts on the society and the economy, such as employment, labor, fair trade, industrial structure, rights and interest of users; (4) impacts on information protection; and (5) other impacts of intelligent information services on the society, economy, culture and citizens' daily lives.

The subjects of a social impact assessment of intelligent information services are the national and local governments. They can implement the assessment arbitrarily. It is not the duty of the state and local governments. Korea has a technology impact assessment system, which is distinguished in that technology impact assessment is mandatory annually by the government. The target of the social impact assessment of intelligent information services is that the use and spread of intelligent information services, which have a huge effect on people's lives, have an impact on society, economy, and culture.

The target of the social impact assessment of intelligent information services is the impact of the use and spread of intelligent information services, which have a huge effect on people's lives, society, economy, and culture. Therefore, it is distinguished from technology impact assessment, which is aimed at predicting the future due to technological development. Technology itself is the only target of the technology impact assessment and cannot be the target of the social impact assessment.

5.3. Evaluation Items of the Social Impact Assessments of Intelligent Information Services

The most important factor of social impact assessment on intelligent information services is the safety and reliability of intelligent informatization services (Article 56.1.1.). In fact, safety and reliability are the most controversial items while forming the basis of intelligent informatization services, including artificial intelligence. Safety can refer to technical, administrative, and physical safeguards. Considering the relationship with the impact on information security (Article 56.1.4.), this should be understood from the perspective of industrial safety, not information security. Reliability can be interpreted as the governance system itself that encompasses the entire AI ethical discourse. It is designed to address the risk of human rights infringement. It may include a human rights impact assessment.

Social Impact Assessments of Intelligent Information Services address the impact of artificial intelligence on society. This is because, among the evaluation items, there are effects on information culture (Article 56.1.2.) and social and economic effects (Article 56.1.3.). For the impact on information culture (Article 56.1.2.), this assessment refers to the information gap bridging, privacy, and intelligent information society ethics. For the impact on information culture (Article 56.1.3.), this assessment refers to the information gap bridging, privacy, and intelligent information society ethics. It also deals with the impact of artificial intelligence on the environment. This is because, among the evaluation items, there is an impact of the intelligent information service on the society, economy, culture, and daily life of the people (Article 56 (1) 5). They are evaluated to deal with social and environmental risks. It was a point that any type of AI impact assessment that has been proposed so far has not been able to be adequately dealt with.

5.4. Evaluation Procedure of the Social Impact Assessments of Intelligent Information Services

After the national or local government decides to investigate and evaluate the social impact of artificial intelligence, the Minister of Science and Technology has an obligation to disclose the results of the social impact assessment. After that, the Minister can recommend necessary measures, such as improving the safety and reliability of the intelligent information service, to national agencies and business operators. It is distinguished from

the existing environmental impact assessment in terms of effectiveness. This is because the publicly announced results of the social impact assessment are not reflected in policy but are only recommended.

Further detailed procedures were not stipulated in the law. However, since the social impact assessment for intelligent information services is a risk management system, the evaluation-communication-management process is expected to be applied when materialized later. At this time, the evaluation is to judge the risk of the artificial intelligence service based on the social impact assessment. As it is a government-led impact assessment, it may lead to overregulation, so it is necessary to secure the independence, objectivity, and expertise of the assessment agency. Additionally, communication means collecting the opinions of stakeholders, experts, or general citizens by disclosing the results of the social impact assessment. Furthermore, management means recommending measures necessary to improve reliability to suppress risks induced by artificial intelligence services. This can be understood as governance that can systematically manage risks beyond simply reflecting them into policies such as environmental impact assessments.

6. Conclusions: Significance and Limitations of the Social Impact Assessments of Intelligent Information Services

The social impact assessment on intelligent information services in South Korea includes all risk-based approaches that have been tried so far. Data protection impact assessments are included by considering not only privacy (Article 56.1.2.) but also the impact on information security (Article 56.1.4.). Human rights impact assessments are also included by considering the safety and reliability of intelligent informatization services (Article 56.1.1.). As the social impact assessment addressed the impact on society and the environment that the existing impact assessment system did not have, a more extensive evaluation became possible. However, it was not specified in detail at the same level as the Risk Assessment Tool of the European Union or Canada. This is because the social impact assessment of South Korea's intelligent information service is in its initial stage. The sub-legislation has not yet been enacted. There are five limitations, including this, which will be described below.

The AI impact assessment in South Korea has the following limitations. First, an impact assessment was introduced only in the public sector. This is because the impact assessment was limited to cases where the national and local governments used intelligent information services. There is a legislative vacuum in the private sector. It is necessary to expand and apply the impact assessment to the private sector as well. Second, the impact assessment is voluntary. As a result, there is a limitation that the impact assessment, in this case, depends on the discretion of the state or local government. However, it is only meaningful in that the impact assessment can be used for mid- to long-term policy establishment and technology development and introduction. The AI impact assessment needs to be made compulsory, at least in the public sector. In the private sector, if this is completed, it is necessary to think about how to provide incentives. Note that the personal information impact assessment of the Personal Information Protection Act divides the public and private sectors. Public institutions are required to undergo a privacy impact assessment when there is a concern about personal information infringement; however, private organizations may arbitrarily undergo a privacy impact assessment or certify an information protection (and personal information protection) management system (ISMS-P). If it is acquired by the private sector, additional points are given in public bidding.

Third, the subject of impact assessment is limited to society. The enormous impact of artificial intelligence technology on human rights and the environment has been overlooked. Points on human rights and the environment as subjects of the impact assessment should also be added. To this end, the Framework Act on Intelligent Informatization should be amended. Fourth, it is necessary to establish a relationship with other impact assessments. Privacy is mentioned as the subject of the social impact assessment. However, privacy is also subject to the privacy impact assessment of the personal information protection act. Since

the subject of impact assessment is overlapping, it is necessary to think about which one to perform first, which outcome to prioritize, and whether to adjust the evaluation target. Fifth, the social impact assessments of Intelligent Information Services lack details. Since the assessment only contains the basic contents of the country's intelligent informatization policy, it lacks specific content, form, and procedure. Therefore, it is necessary to develop an AI impact assessment system and standards. In this case, it is necessary to take a different approach to individual artificial intelligence. Artificial intelligence technology can be divided into natural language processing, computer vision, robotics, and rule-based systems. This is because applying a single evaluation criterion to a different object can lead to asymmetric results.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

- AD HOC Committee on Artificial Intelligence (CAHAI) Policy Development Group. 2021. *Human Rights, Democracy and Rule of Law Impact Assessment of AI systems, Council of Europe*. Strasbourg: Council of Europe (Europe).
- Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. *ProPublica*. Available online: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (accessed on 1 June 2022).
- Aslam, Nida, Irfan Ullah Khan, Samiha Mirza, Alanoud AlOwayed, Fatima M. Anis, Reef M. Aljuaid, and Reham Baageel. 2022. Interpretable Machine Learning Models for Malicious Domains Detection Using Explainable Artificial Intelligence (XAI). *Sustainability* 14: 7375. [CrossRef]
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Paper present at the 2021 ACM Conference on Fairness, Accountability, Virtual Event, March 3–10.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, and et al. 2020. Language Models are Few-Shot Learners. *arXiv* arXiv:2005.14165.
- Canter, Larry W. 1982. Environment Impact Assessment. *Impact Assessment* 1: 6–40. [CrossRef]
- Chae, Yoon. 2020. US AI regulation guide: Legislative overview and practical considerations. *The Journal of Robotics, Artificial Intelligence & Law Volume* 3: 1.
- Crawford, Kate, and Trevor Paglen. 2019. Excavating AI: The politics of images in machine learning training sets. *AI & Society* 36: 1105–16. [CrossRef]
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* arXiv:1810.04805.
- European Commission. 2019. Policy and Investment Recommendations for Trustworthy Artificial Intelligence. Available online: <https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence> (accessed on 1 June 2022).
- European Institute for Gender Equality (EIGE). 2016. *Gender Impact Assessment*. Helsinki: EU Publication Office (Europe), p. 8.
- Goodfellow, Ian. 2016. *Deep Learning*. Cambridge: MIT Press.
- Joshi, Prateek. 2019. How do Transformers Work in NLP? A Guide to the Latest State-of-the-Art Models. Available online: <https://www.analyticsvidhya.com/blog/2019/06/understanding-transformers-nlp-state-of-the-art-models/> (accessed on 1 June 2022).
- Kaminski, Margot E. 2019a. Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability. *University of Colorado Law Legal Studies Research Paper*. Available online: <https://scholar.law.colorado.edu/articles/1265/> (accessed on 1 June 2022).
- Kaminski, Margot E. 2019b. The Right to Explanation, Explained. *Berkeley Technology Law Journal* 34. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3196985 (accessed on 1 June 2022).
- Kaminski, Margot E., and Gianclaudio Malgieri. 2019. Algorithmic impact assessments under the GDPR: Producing multi-layered explanations. *International Data Privacy Law*. Available online: <https://academic.oup.com/idpl/article/11/2/125/6024963> (accessed on 1 June 2022).
- Kasirzadeh, Atoosa, and Damian Clifford. 2021. Fairness and data protection impact assessments. Paper present at the 2021 AAAI/ACM Conference on AI, Ethics, and Society, Virtual Event, May 19–21.
- Lambert, James, and Edward Cone. 2019. How Robots Change the World—What Automation Really Means for Jobs, Productivity. Technical Report, Oxford Economics. Available online: <https://www.oxfordeconomics.com/resource/how-robots-change-the-world/> (accessed on 1 June 2022).
- Lauret, Julien. 2019. Amazon's Sexist AI Recruiting Tool: How Did It Go So Wrong? *Medium*. Available online: <https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e> (accessed on 1 June 2022).
- Lee, Kai-Fu. 2018. *AI Superpowers: China, Silicon Valley, and the New World Order*. Boston: Houghton Mifflin Harcourt.

- Mittelstadt, Brent, Chris Russell, and Sandra Wachter. 2018. Explaining Explanations in AI. In Proceedings of FAT*19: Conference on Fairness, Accountability, and Transparency. *arXiv* arXiv:1811.01439.
- Moravec, Hans. 1990. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge: Harvard University Press.
- Mytton, David. 2021. Data centre water consumption. *NPJ Clean Water* 4: 1–6. Available online: <https://www.nature.com/articles/s41545-021-00101-w> (accessed on 1 June 2022). [CrossRef]
- Nahmias, Yifat, and Maayan Perel. 2021. The oversight of content moderation by AI: Impact assessments and their limitations. *Harvard Journal on Legislation* 58: 145. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3565025 (accessed on 1 June 2022).
- OECD. 2022. 2nd International Conference on AI in Work, Innovation, Productivity and Skills. Available online: <https://oecd.ai/en/work-innovation-productivity-skills> (accessed on 1 June 2022).
- Office of the Australian Information Commissioner (OAIC). 2021. *Guide to Undertaking Privacy Impact Assessments*; Melbourne: Australian Government (Australia).
- Patterson, David, Joseph Gonzalez, Quoc Le, Chen Liang, Lluís-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. 2021. Carbon Emissions and Large Neural Network Training. *arXiv* arXiv:2104.10350.
- Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners. Available online: https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (accessed on 1 June 2022).
- Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving Language Understanding by Generative Pre-Training. Available online: https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf (accessed on 1 June 2022).
- Ruiz, Cristina. 2019. Leading Online Database to Remove 600,000 Images after Art Project Reveals Its Racist Bias. *The Art Newspaper*. Available online: <https://www.theartnewspaper.com/2019/09/23/leading-online-database-to-remove-600000-images-after-art-project-reveals-its-racist-bias> (accessed on 1 June 2022).
- Russell, Stuart, and Peter Norvig. 2022. *Artificial Intelligence: A Modern Approach*, 4th ed. London: Pearson.
- Sambasivan, Nithya, and Jess Holbrook. 2018. Toward responsible AI for the next billion users. *Interactions* 26: 68–71. [CrossRef]
- Salomons, Anna. 2019. New Frontiers: The Evolving Content and Geography of New Work in the 20th, May 2019. Working Paper. Available online: <https://app.scholarsite.io/david-autor/articles/new-frontiers-the-evolving-content-and-geography-of-new-work-in-the-20th-century> (accessed on 1 June 2022).
- Selbst, Andrew D. 2021. An institutional view of algorithmic impact assessments. *Harvard Journal of Law & Technology* 35: 117–91.
- Stanford HAI. 2021. AI Policy and National Strategies. Artificial Intelligence Index Report 2021. Available online: https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf (accessed on 1 June 2022).
- South Korea. 2020. Framework Act on Intelligent Informatization. Available online: <https://www.law.go.kr/LSW//lsInfoP.do?lsiSeq=218737&chrClsCd=010203&urlMode=engLsInfoR&viewCls=engLsInfoR#EJ56:0> (accessed on 1 June 2022).
- Surden, Harry. 2019. Artificial Intelligence and Law: An Overview. *Georgia State University Law Review*. 35. Available online: <https://scholar.law.colorado.edu/cgi/viewcontent.cgi?article=2340&context=articles> (accessed on 1 June 2022).
- Taulli, Tom. 2021. *Artificial Intelligence Basics: A Non-Technical Introduction*. New York: Apress.
- Tu, Menger, Sandy Dall’erba, and Mingque Ye. 2022. Spatial and Temporal Evolution of the Chinese Artificial Intelligence Innovation Network. *Sustainability* 14: 5448. [CrossRef]
- US Bureau of Labor Statistics. 2021. Interpreters and Translators: Occupational Outlook Handbook. Technical Report, U.S. Available online: <https://www.bls.gov/ooh/media-and-communication/interpreters-and-translators.htm> (accessed on 1 June 2022).
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. Available online: <https://arxiv.org/abs/1706.03762> (accessed on 1 June 2022).
- Webb, Michael. 2019. The Impact of Artificial Intelligence on the Labor Market. Available online: <https://ssrn.com/abstract=3482150> (accessed on 1 June 2022).
- Weidinger, Laura, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, and et al. 2021. Ethical and social risks of harm from Language Models. *arXiv* arXiv:2112.04359.