The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)
Vol. 24, No. 2, pp.15-21, Apr. 30, 2024. pISSN 2289-0238, eISSN 2289-0246

https://doi.org/10.7236/JIIBC.2024.24.2.15
JIIBC 2024-2-3

인공지능 서비스 거버넌스 연구

Governance research for Artificial intelligence service

유순덕*

Soonduck Yoo*

요 약 본 연구의 목적은 일반적인 서비스 뿐만 아니라 공공정책 등에 인공지능 서비스를 도입할 때, 어떤 단계를 통해 진행하고 점검할 수 있는지 방안에 대해 제안을 하고 있다. 이를 위해 인공지능 서비스 관리와 거버넌스 툴킷에 대해 제시하고 공공정책에 인공지능 서비스를 제공할 때, 어떻게 해야 하는지에 대한 내용을 연구하였다. 첫째, 인공지능 서비스의 개발 방향과 개발하지 말아야 할 내용에 대한 지침을 제공하고 있다. 둘째, 개발을 하는 경우 인공지능 거버넌스 툴킷에서 제공하고 있는 설계, 개발, 배포단계별로 검토해야 하는 체크 리스트를 통해 내용을 점검 후 진행하는 것을 권장하고 있다. 셋째, 인공지능 서비스를 운영할 시 1) 기획설계, 수명주기, 3) 모델 구축 및 검증, 4) 배포 및 모니터링, 5) 책임에 대한 각각 원칙과 관련 내용을 명확히 제시하고 이에 충족하고 있는지에 대해 점검을 해야 한다. 인공지능 서비스의 거버넌스 측면은 궁극적으로 제공되는 서비스에 대한 위험 측면을 완화하려는 노력으의 일환이므로 등장할 수 있는 위험관리 측면에서도 연구가 이루어져야 한다. 우리는 인공지능에 제공하는 장점을 수용하면서 한계 및 위험요소에 대한 적극적인 대응 방안으로 마련해야 한다. 인공지능 기술을 적극적 활용하여 효율적으로 정책 수립하여 고부가 가치를 생성하고 사회에 의미 있는 영향을 제공할 수 있도록 노력해야 한다.

Abstract The purpose of this study is to propose a framework for the introduction and evaluation of artificial intelligence (AI) services not only in general applications but also in public policies. To achieve this, the study explores AI service management and governance toolkits, providing insights into how to introduce AI services in public policies. Firstly, it offers guidelines on the direction of AI service development and what aspects to avoid. Secondly, in the development phase, it recommends using the AI governance toolkit to review content through checklists at each stage of design, development, and deployment. Thirdly, when operating AI services, it emphasizes the importance of adhering to principles related to 1) planning and design, 2) the lifecycle, 3) model construction and validation, 4) deployment and monitoring, and 5) accountability. The governance perspective of AI services is crucial for mitigating risks associated with service provision, and research in risk management aspects should be conducted. While embracing the advantages of AI, proactive measures should be taken to address limitations and risks. Efforts should be made to efficiently formulate policies using AI technology to create high value and provide meaningful societal impacts.

Key Words : Artificial intelligence service, Service gervenance, Framwork, Process and principles

*정회원, 한세대학교 경영학과
접수일자 2023년 12월 30일, 수정완료 2024년 3월 9일
게재확정일자 2024년 4월 5일

Received: 30 December, 2023 / Revised: 9 March, 2024 / Accepted: 5 April, 2024
*Corresponding Author: koreasally@gmail.com
Dept. of International Business, Hansei University, Korea

I. 서 론

인공지능은 주어진 목표에 대해 실제 또는 가상 환경에 영향을 미치는 예측, 권장 사항 또는 결정과 같은 출력을 생성할 수 있는 작업을 수행할 수 있는 모든 시스템이다[1][2].

편견의 원인은 다양하다. 과거의 편견이나 바람직하지 않은 상태를 포함하여 일부 편견은 데이터에 내재되어 있다[3]. 즉, 모델에서 복제해서는 안 되는 기존 패턴이다. 표현 편향은 속성 누락, 표본 디자인, 하위 모집단의 데이터 전체 또는 부분 부재로 인해 정보가 불완전할 때 발생한. 측정 편향은 모델에 포함되어야 하는(포함되지 않는) 변수의 생략(포함)으로 인해 발생한다[4]. 방법론적 오류로 인해 다른 편향이 나타난다. 예를 들어 검증 프로세스, 측정항목 정의 및 결과 평가(평가 편향)의 오류로 인해 훈련 중에 편향이 발생하고, 대상 모집단에 대한 잘못된 가정으로 인해 편향이 발생한다. 모델 정의에 영향을 미친다. 결과에 대한 부적절한 해석이나 현실 세계 또는 데이터 캡처 방법의 패턴의 일시적인 변화로 인해 모델의 오용 및 모니터링으로 인해 편향이 발생한다. 이 툴킷의 여러 섹션을 통해 이러한 편향의 주요 이유가 제시되고 이를 완화하기 위한 다양한 조치가 제안된다.

강력하고 책임감 있는 의사결정 또는 의사결정 지원 AI 시스템을 구축하려면 가능한 모든 편견 원인을 신중하게 고려해야 한다. 결함 조사 및 가정 문서화; 시스템이 충족해야 하는 알고리즘 공정성 목표 및 기준에 대한 명확한 정의로 해야하며 시스템의 특정 상황에서 한계와 허용 가능한 오류를 이해해야 한다. 이 과정은 의사결정에서 바람직하지 않거나 편향된 결과를 피하기 위한 모니터링 조치의 구현 하는 것이다[5][6].

AI 시스템 수명 주기 동안 기계학습 방법을 구축하고 적용할 때 발생하는 일반적인 업무와 이와 관련된 프로세스와 점검항목에 대한 연구가 필요하다.

정책 입안자와 기술팀은 수명주기의 각 단계에서 AI 시스템이 올바르게 작동하도록 책임을 져야 한다.

이와 관련하여 툴킷의 한 장은 공공정책을 위해 AI를 사용할 때 책임 관련 문제를 탐색하고 이를 해결하기 위한 실제 메커니즘에 대한 이해가 이루어져야 한다.

본 연구의 목적은 일반적인 서비스 뿐만 아니라 공공정책 등에 인공지능 서비스를 도입할 시 어떤 단계를 통해 진행하고 점검할 수 있는 방안에 대한 제안을 하고 있다.

II. 인공지능 서비스 관리

1. 머신러닝 및 의사결정 지원 시스템

기계학습의 대표적인 방법인 머신러닝 방법은 AI 시스템이 사용할 수 있는 유일한 유형의 알고리즘은 아니지만 최근 몇 년간 가장 많이 성장한 알고리즘이다. 이러한 방법은 인간이 입력한 명시적 또는 상징적 명령 대신 시스템이 패턴과 추론을 통해 자동화된 방식으로 행동을 학습할 수 있도록 하는 일련의 기술로 구성되어 있다[7].

머신러닝은 인공지능(AI)의 하위 집합이며 비즈니스부터 공공정책 및 서비스 제공에 이르기까지 다양한 맥락에서 조치나 개입을 알리기 위해 의사 결정자가 기계학습 방법을 점검하기 위해 더 많이 사용하고 있다. 실제로 이러한 방법은 다양한 수준의 성공을 거두었으며 이 방법이 사회에 미치는 긍정적 또는 부정적인 성과와 영향을 이해하는 방법에 대한 관심이 커지고 있다[8]. 따라서 의사결정 또는 의사결정 지원을 위해 머신러닝 기술을 사용할 때 발생하는 가장 일반적인 몇 가지 과제가 있다.

이와 관련하여 구현오류와 편견을 감지 및 완화하고 회사, 공공부문 기관 또는 사회에 바람직하지 않은 결과가 발생할 가능성을 평가하는 것을 포함하고 있다.

경제협력개발기구(OECD)는 AI 시스템을 주어진 목표에 대한 결과(예측, 권장 사항 또는 결정)를 생성하여 환경에 영향을 미칠 수 있는 기계 기반 시스템으로 설명한다.

첫째, 실제 및 또는 가상 환경을 인식하기 위해 기계 및 또는 인간기반 데이터와 입력을 사용한다.

둘째, 자동화된 방식(예: 기계학습) 또는 수동 분석을 통해 이러한 인식을 모델로 추상화한다.

셋째, 모델 추론을 사용하여 결과에 대한 옵션을 공식화했다. AI 시스템은 다양한 수준의 자율성으로 작동하도록 설계되었다[7][9]

여러 방법으로 점검된 결과를 기반으로 의사결정자는 인공지능 기술로 등장한 결과물에 대해 신뢰를 제공하고 있다.

2. 인공지능 서비스 관리

인공지능 서비스를 적용 시 윤리적 AI를 지원하는 원칙은 다양하게 개발되고 있다. 그러나 현재 AI 서비스를 어떻게 개발해야 하는지, 개발하지 말아야 하는지에 대한 높은 수준의 지침과 원칙에 대한 연구가 진행되고 있다.

기본적으로 인공지능 서비스 관리 원칙은 첫째, AI 서비스를 어떻게 개발해야 하는지와 둘째, 개발하지 말아야 하는지에 대한 높은 수준의 지침을 제공하는 것을 목적으로 한다.

AI 서비스를 개발하기 위해 다음과 같은 단계를 수행을 할 수 있다.

표 1. AI 서비스 개발 단계별 수행 내용
Table 1. AI service development step-by-step details

구분	내용
문제 정의 및 목표 설정	어떤 문제를 해결할 것인지 명확히 정의하고 목표를 설정
데이터 수집 및 전처리	품질 좋은 데이터는 효과적인 모델 훈련의 핵심으로 필요한 데이터를 수집하고 정제
모델 선택 및 설계	사용할 모델을 선택하고 설계하는 것으로 특정 문제에 맞는 모델을 선택하는 것이 중요하며 미리 훈련된 모델을 사용하거나, 필요에 따라 새로운 모델을 설계할 수 있음
모델 훈련	선택한 데이터로 모델을 훈련시키며. 이 과정에서 하이퍼 파라미터를 조정하고 성능을 평가
평가 및 성능 향상	훈련된 모델을 평가하고 성능을 향상시키기 위해 모델을 조정하거나 추가 데이터를 수집 할 수 있음
배포	훈련이 완료된 모델을 서비스에 통합하고 배포함
모니터링 및 유지보수	서비스가 운영되는 동안 모델의 성능을 지속적으로 모니터링 하고 필요에 따라 업데이트 또는 재훈련을 수행
보안 및 개인 정보 보호	AI 서비스에는 보안과 개인 정보 보호에 대한 고려 사항이 있으며 적절한 보안 프로토콜을 적용하고 사용자 데이터를 안전하게 처리해야함

AI 서비스를 개발하지 말아야 하는지에 대한 지침으로 다음과 같은 내용을 검토할 수 있다.

표 2. AI 서비스를 개발하지 말아야 하는지에 대한 지침
Table 2. Guidance on whether or not to develop AI services

구분	내용
비투명성과 해석 불가능성	AI 모델이 작동하는 방식을 이해하기 어려운 경우, 특히 결정의 근거를 설명할 수 없는 경우에는 신중해야 함
사용자 프라이버시 보호	사용자의 개인정보를 수집하고 처리할 때는 규정과 법규를 준수해야 하며. 개인 정보 보호에 대한 적절한 보안 및 처리 방법을 확보해야 함
공정성과 편향	모델이 특정 집단이나 속성에 대해 편향되거나 불공정한 예측을 하는 경우, 이에 대한 조치를 취해야 하며. 특히, 모델이 편견을 갖고 있는 경우, 이를 완화하거나 교정하는 방법을 찾아야 함
데이터의 질과 다양성	모델을 훈련 시킬 때 사용하는 데이터가 충분히 다양하고 대표성을 가져야 하며 부족하거나 특정한 편향이 있는 데이터로 모델을 훈련시키면 예측의 정확도와 일반화 능력이 저하될 수 있음
사용자 동의와 투명성	AI 서비스를 사용하는 사용자에게 목적, 데이터 수집 및 활용에 대한 명확하고 이해하기 쉬운 정보를 제공하고, 그들의 동의를 받아야 함
법적 규정과 준수	해당 국가 또는 지역의 법적 규정과 준수 사항을 철저히 따라야 하며 데이터 보호, 저작권, 소비자 보호 등 다양한 법적 요구 사항을 고려해야 함
고객 서비스 및 응답	AI 서비스에 대한 피드백과 불만 사항에 신속하게 대응하고, 필요에 따라 모델이나 서비스를 조정하거나 개선해야 함
사회적 영향과 윤리	개발한 AI 서비스가 사회적 영향을 미칠 수 있는 경우, 윤리적 고려와 사회적 책임을 고려해야 하며 서비스가 부정적인 영향을 미칠 수 있는 가능성에 대비하고 적절한 조치를 취해야 함

III. 인공지능 서비스 거버넌스

1. 인공지능 서비스 거버넌스 툴킷

가. 사용 대상자

인공지능 거버넌스 툴킷은 다양한 분야에서 사용 할 수 있으며 우선적으로 공공정책을 수립하는 기술팀에서 가이드라인을 제공하는 것으로 하고 있다. 거버넌스 툴킷은 공공정책을 위한 기계학습 알고리즘을 적용하는 기술팀을 위한 것이다. 그러나 이 기술은 다른 응용 분야에서 흔히 발생하는 문제도 다루고 있다.

인공지능 서비스를 위한 모범 사례에 대한 명확성[10]을 확보하기 위해 사용되는 거번넌스의 목적은 위험 영역을 식별하고 의사결정자의 목표에 반하는 결과를 방지하기 위한 완화 조치를 권장하는 것이다. 이러한 결과에는 바람직하지 않은 결과, 부적절한 타겟팅으로 인한 자원 낭비 또는 의사 결정자가 달성하고자 하는 목표를 약화시키는 등의 결과를 포함하고 있다.

나. 거버넌스 프로세스

인공지능 서비스를 효과적으로 운영하기 위해서는 적절한 거버너스 환경을 확보해야 한다[11][12]. 이를 위해 우리는 인공지능을 활용한 서비스를 개발 시 다음과 같은 단계를 거치도록 권장하고 있다. 서비스를 위해 해결해야 할 문제를 인식하고 해당 서비스를 개발하고 마지막으로 배포하는 단계를 가지고 있다.

표 3. 서비스 생성 절차도
Table 3. Service creation process diagram

구분	내용
설계	문제 인식
개발	데이터 수집
	생성 및 테스트 알고리즘
	업데이트 알고리즘
배포	배포

다. 설계 단계

임무 과제에 대해 팀의 이해를 공유하려면 먼저 AI 기술이 해결할 주요 문제를 식별해야 한다. 설계 단계에서 문제 인식을 하기 위해 다음과 같은 질문이 필요하다.

표 4. 설계 단계의 문제 인식 주요 질문
Table 4. Problem recognition key questions during the design phase

구분	주요 내용
설계 단계	이 알고리즘은 어떤 임무 목적을 충족합니까?
	알고리즘 프로세스의 예상되는 임무 결과는 무엇입니까?
	이 데이터를 수집/보유할 권한이 있습니까?
	사용하려고 생각하는 데이터의 출처는 무엇입니까? 직접 수집하셨나요? 다른 사람이 귀하를 대신하여 수집했습니까?

정보를 공유하려면 인공지능에 맡겨진 임무에 대한 이해가 필요하다. 이 단계의 경우, 결과가 일반언어로 설명되어야 하며 AI 기술 보안, 개인정보 보호 관련으로 문제 사항 없는지 검토가 이루어져야 한다.

라. 개발 단계

인공지능 서비스 개발 단계에서 다음 사항에 대한 검토가 이루어져야 한다.

첫째, 설계에 기초한 데이터를 수집을 해야 한다. 이 단계에서 충분하고 신뢰가 있는 데이터 수집이 이루어져야 한다. 그리고 데이터 수집 시 다음과 같은 사항을 검토해야 한다. 이는 데이터를 적절한 방식으로 수집했는지에서부터 적절한 데이터가 이용되었는지, 데이터의 보관 방법에 대한 충분한 검토가 이루어졌는지를 확인을 하는 과정이다.

표 5. 개발 단계에서 데이터 수집시 주요 질문
Table 5. Key questions when collecting data during the development phase

구분	주요 내용
개발 단계의 데이터 수집	귀하에게는 정보 수집 및/또는 사용할 법적 권한이 있습니까? 데이터 수집에 관한 공개 공지가 있나요?
	전송 중 및 저장 중인 데이터는 어떻게 보호됩니까?
	제안된 데이터 세트에 개인정보가 포함이 되어있는 경우 공지는 어떻게 되나요? 제공되고 동의가 수집되었는가?
	데이터가 대중으로부터 직접 수집되었나요?
	데이터의 소유자는 누구입니까? 액세스는 어떻게 관리되나요?
	데이터가 얼마나 시의적절합니까? 문제 진술과 얼마나 오랫동안 관련될 것인가?
	데이터가 충분히 대표성이 있나요? 잠재적인 편견의 원인이 있습니까? 데이터는 얼마나 대표성/편향성이 있나요? 관련성/편향을 어떻게 조사했나요? 편견이 법적/윤리적/공공 신뢰에 미치는 영향은 무엇입니까? 허용 가능한 데이터 품질/편향은 누가 결정합니까?
	테스트를 준비할 때 데이터에 무엇을 했나요?
	데이터는 어디에/얼마나 오랫동안 보관되나요?

둘째, 테스트 알고리즘의 생성이다. 이를 위해 (1) 모델 훈련 및 선택 프로세스는 대화 진행, (2) 반복 학습으로 훈련하여 모델 언어 습득, (3) 원하는 데이터를 얻을 수 있도록 효율적인 모델 언어 습득체계를 진행해야 한다.

표 6. 배포 단계의 주요 질문
Table 6. Key questions during the deployment phase

구분	주요 내용
개발 단계의 알고리즘 생성 및 테스트	정확도 임계값은 무엇입니까? 즉, "정확한" 것으로 인정되려면 출력이 얼마나 정확해야 합니까(위양성률 대 위음성률)?
	잠재적으로 편향된 출력을 결정하는 메커니즘은 무엇입니까? 잠재적으로 편향된 출력을 어떻게 결정합니까? (데이터 품질 및 무결성)
	실험을 어떻게 재현할 수 있나요? 재현성에 대한 장벽이 있습니까? (감사/회 계)
	훈련 데이터 세트가 대표적인지 어떻게 확인합니까? 알고리즘이 정확한지 어떻게 확인합니까? 데이터가 완전하고 알고리즘이 정확한지 확인하기 위해 어떤 방법이 마련되어 있나요?
	예상되는 결과는 무엇입니까?
	재현성을 위해 알고리즘 버전과 테스트 프로토콜을 어떻게 문서화하고 있나요? 알고리즘은 무엇을 하도록 설계되었나요? 다른 기능도 있나요?
	알고리즘을 수정할 수 있나요? 원하는 출력/결과를 얻지 못한 경우 알고리즘을 변경할 수 있는 유연성이 있습니까? 그렇다면 프로세스는 무엇이며 이로 인해 발생할 수 있는 위험은 무엇입니까?
	초기결과가 예상치 못한 경우, 작업하려면 얼마나 많은 작업이 필요합니까? 테스트를 재현하시겠습니까? 결과가 예상치 못한 경우 개인정보가 필요합니까?
	편향/강화 편향 및/또는 이질적 처리의 가능성을 어떻게 평가합니까? 추가 허용 여부기능 및/또는 신뢰도/신뢰도를 높이기 위해 정확도/ 신뢰도 임계값에 대한 표준 및/또는 기대치는 무엇입니까?
	더 적은 PII 데이터로 유사하고 효과적인 결과를 얻을 수 있습니까? 데이터가 양식에서 나오고 알고리즘이 양식을 읽고 출력을 제공하면 어떻게 될까요? 그런 다음 사용자는 양식 정보를 변경한다. 사용자와 개발자 커뮤니티는 양식/알고리즘/출력의 변경사항에 대해 어떻게 소통합니까? 중요한 변경 사항이 있는 경우 거버넌스 팀은 잠재적으로 "데이터 수집" 단계로 돌아가야 한다.

셋째, 사용하고 있는 알고리즘의 업데이트가 이루어져

야 한다. 이를 위해 (1) 이미 습득이 된 모델을 통하여 실적이 좋은 것으로 보이는 평가 지표, 새 데이터에 대한 모델 테스트를 진행하고, (2) 학습모델을 일반화하여 비즈니스에 부합할 수 있도록 한다.

마. 배포 단계

배포 하기 전에 추적된 결과물의 잠재적 편향/강화 편향 및 이질성을 평가해야 한다. 또한, 사용자의 매커니즘에서 논의하고 살펴보는 것이 필요하다.

표 7. 배포 단계의 주요 질문
Table 7. Key questions during the deployment phase

구분	주요 내용
배포단계	단계적 배포인가요? 그렇다면 왜 그렇습니 까? 시간이 지남에 따라 AI의 목적이 변경되었는지를 어떻게 확인합니까?
	출력에 PII 또는 기타 CUI가 포함됩니까? 그렇다면 출력은 어떻게 보호되고 액세스는 제한됩니까?
	개인이 어떻게 알 수 있나요? 그들은 알고리즘 프로세스를 거치고 있습니까? 관련 SORN/ PIA 또는 기타 공지가 업데이트, 검토 및 게시되었습니까?
	개인이 시스템에 있는 자신의 데이터에 접근/ 수정할 수 있나요? 그렇다면 어떻게?
	상당한 시간이 지났거나 조건이 변경된 경우 동의를 업데이트할 수 있는 방법이 있습니 까?

2. 공공정책을 위한 AI 툴킷

가. 공공정책 수명주기의 의사결정

AI는 공공정책 수립을 대체하지 않으며. 그 기능은 의사결정을 위한 정보를 제공하여 공공정책 개발 주기를 지원하는 것이다. AI 지원 공공정책 주기는 다음 단계로 구성된다.

그림 1. AI 지원 공공정책 주기
Fig. 1. AI-enabled public policy cycle

첫째, 문제 식별단계에서 모든 AI 프로젝트는 공공 정책이 해결하려고 하는 문제와 해당 문제의 가능한 원인 및 결과를 올바르게 식별하는 것부터 시작해야 한다.

둘째, 정책 수립단계에서 특정 사람, 단위 또는 프로세스에 적용되는 것으로 간주 되는 개입 또는 정책이 수립된다. 우리는 일반적으로 그러한 정책이 목표 집단에 적용될 때 이점이 있다는 증거가 있다고 가정한다.

셋째, 의사결정/의사결정 지원 시스템 단계에서 개입이 정의되면 AI 주기는 의사결정/의사결정 지원 시스템의 설계 및 개발로 시작되며, 그 결과는 선택한 개입에 초점을 맞추거나 안내하는 데 사용된다.

넷째, 정책 시행단계에서 공공정책은 시범 프로젝트나 일반적으로 대규모 프로젝트로 시행한다.

다섯째, 정책 평가단계에서 정책 조치의 효과성, 신뢰성, 비용, 예상 및 의도하지 않은 결과, 기타 관련 특성을 평가한다. 결과가 긍정적이면 개입이 계속되거나 규모가 확대된다.

공공정책 결정 주기와 병행하여 AI 시스템 개발에는 다음 단계를 포함하는 자체 수명주기로서[7][13] (i) 계획 및 설계, (ii) 데이터 수집 및 처리, (iii) 모델 구축 검증, (v) 배포 및 모니터링 등이 있다.

나. 공공정책 수립을 위한 AI 툴킷

공공 부문에서 AI는 생산성 향상과 공공 서비스 품질 향상을 약속한다. 예를 들어, 정책 입안자는 소셜 네트워크 활동을 실시간으로 분석함으로써 AI 시스템을 활용하여 가장 시급한 사회문제와 요구 사항에 대해 보다 정확하고 증거 기반의 평가를 얻을 수 있다. AI 시스템의 결과와 예측은 정책 수립, 구현 및 평가에 영향을 미칠 수 있다.

이러한 배경에서 전 세계 정부는 공공정책 개발을 지원하기 위해 AI의 힘을 활용할 수 있는 관련 기술을 갖추고 있다. 그러나 AI 기반 공공 정책이 사람들의 삶과 복지에 큰 영향을 미칠 수 있다는 점을 고려할 때, AI 사용으로 인한 기회를 포착하고 AI 사용으로 인해 발생하는 문제를 해결하기 위한 적절한 보호 장치가 마련되어 있는지 확인하기 위한 체계적인 접근방식이 필요하다.

AI 시스템에 대한 기술 표준 및 규범의 개발은 여전히 AI 커뮤니티의 미해결 과제이지만, 이 툴킷에서는 AI 수명 주기 동안 편견을 방지하고 완화하기 위한 주요 기술적 측면과 조치를 설명하고 있다.

3. 공공정책을 위한 AI 시스템 프레임워크

AI 시스템 수명주기를 분석을 위한 지침 프레임워크로 사용하는 이 툴킷은 AI 기술을 사용하여 의사결정 프로세스와 결과를 개선하려는 공공정책팀에 기술 지침을 제공한다.

AI 시스템 수명주기의 각 단계(계획 및 설계, 데이터 수집 및 처리, 모델 구축 및 검증, 배포 및 모니터링)에 대해 이 툴킷은 공공 정책 맥락에서 AI를 사용할 때 발생하

는 가장 일반적인 과제를 식별하고 탐지 및 완화를 위한 실제 메커니즘을 간략하게 설명한다.

유럽 연합의 GDPR(General Data Protection Regulation, 일반 데이터 보호 규정)과 같은 규정에서는 책임을 조직이 적절한 기술 및 조직적 조치를 마련하고 요청 시 수행한 작업과 그 효과를 입증할 수 있어야 하는 요구 사항으로 정의하고 있다.

정책 입안자와 기술팀은 수명주기의 각 단계에서 AI 시스템이 올바르게 작동하도록 책임을 져야 한다. 이와 관련하여 툴킷은 공공정책을 위해 AI를 사용할 때 책임 관련 문제를 탐색하고 이를 해결하기 위한 실제 메커니즘을 간략하게 설명하는데 전념하고 있다.

공공정책을 위한 AI의 책임감 있는 사용을 장려한다는 목표에 충실하여 툴킷의 각 영역에는 실제 구현을 안내하는데 도움이 되는 체크리스트가 포함되어 있다. 데이터 문제를 평가하고 AI 시스템의 특성, 가정, 수명주기 전반에 걸쳐 구현된 위험 완화 조치를 문서화 하는데 도움이 되는 "데이터 프로필" 도구와 "모델 카드"도 제공하고 있다.

표 8. 공공정책을 위한 AI 툴킷 운영 방향
Table 8. AI Toolkit Operation Direction for Public Policy

구분	내용
기획 및 설계	문제의 올바른 정의와 공공정책 대응
	AI 원칙
데이터 수집 및 처리	데이터 품질 및 이용 가능한 데이터의 관련성
	대상 모집단에 대한 데이터 검증 및 완전성
모델 구축 및 검증	검증 샘플의 부재 또는 부적절한 사용
	데이터 유출
	분류 모델
	계량화되지 않은 오류와 인간의 평가
	공정성과 차별화된 성과
배포 및 모니터링	성능 저하
	모델 효율성을 평가하기 위한 실험
책임	예측의 해석 및 설명
	추적성

IV. 결 론

우리는 인공지능 서비스가 제공하고 있는 업무의 효율성과 편리성을 통해 기술과 시장이 동시에 성장하고 있다. 기술에 발달은 긍정적인 요소와 부정적인 요소를 제공하고 있어 어떻게 부정적인 요인을 완화하느냐가 중요한 과제로 등장하고 있다.

본 연구는 인공지능 서비스 관리와 거버넌스 툴킷에 대해 제시하고 공공정책에 인공지능 서비스를 제공시 어떻게 해야 하는지에 대한 방법을 제시하고 있다.

현재 인공지능 서비스가 해결해야 하는 것은 머신러닝을 이용하여 개발되는 인공지능 서비스에서 등장 할수 있는 구현오류와 편견을 감지 및 완화하고 회사, 공공부문 기관 또는 사회에 바람직하지 않은 결과가 발생할 가능성을 평가하는 것을 해야 한다.

기본적으로 인공지능 서비스 관리 원칙은 첫째, AI 서비스를 어떻게 개발해야 하는지와 둘째, 개발하지 말아야 하는지에 대한 높은 수준의 지침을 제공하는 것을 목적으로 한다.

인공지능 서비스에 사용되는 거버넌스의 목적은 위험 영역을 식별하고 의사결정자의 목표에 반하는 결과를 방지하기 위한 완화 조치를 권장하는 것이다. 이러한 결과에는 바람직하지 않은 결과, 부적절한 타겟팅으로 인한 자원 낭비 또는 의사 결정자가 달성하고자 하는 목표를 약화시키는 등의 결과를 포함하고 있다.

인공지능 서비스 거버넌스를 위해 사례, 개발, 배포의 단계를 통해 각각 점검해야 할 체크 리스트를 제공하고 있다.

특히, 인공지능 서비스를 공공정책 개발에 적용하기 위해 공공정책 의사결정 지원으로 1) 문제 식별, 2) 정책 수립, 3) 의사결정 지원, 4) 정책 시행, 5) 정책 평가의 단계를 제시하고 있다. 이 과정에서 AI 기반 공공 정책이 사람들의 삶과 복지에 큰 영향을 미칠 수 있다는 점을 고려할 때, AI 사용으로 인한 기회를 포착하고 AI 사용으로 인해 발생하는 문제를 해결하기 위한 적절한 보호 장치가 마련되어 있는지 확인하기 위한 체계적인 접근방식이 필요하다.

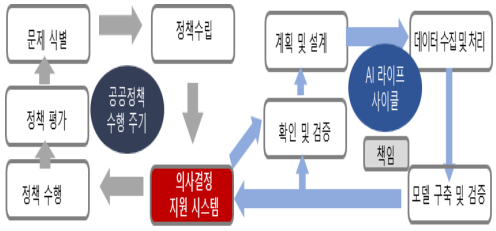

그림 2. 의사결정 지원 시스템과 공공정책 생애주기
Fig. 2. Decision support systems and public policy life cycle

공공정책을 위한 AI의 책임감 있는 사용을 장려한다는 목표에 충실하여 툴킷의 각 영역에는 실제 구현을 안내하는데 도움이 되는 체크 리스트가 포함되어 있다. 데이터

문제를 평가하고 AI 시스템의 특성, 가정, 수명주기 전반에 걸쳐 구현된 위험 완화 조치를 문서화하는데 도움이 되는 데이터 프로필 도구와 모델 카드도 제공된다.

이 연구의 시사점은 인공지능 서비스를 정책에 도입시 활용할 수 있는 거버넌스 툴킷에 대한 조명을 통해 현업에서 적용할 수 있는 방안에 대한 가이드 라인을 제공하고 있다.

첫째, 인공지능 서비스의 개발 방향과 개발하지 말아야 할 내용에 대한 지침을 제공하고 있다.

둘째, 개발을 하는 경우, 인공지능 거버넌스 툴킷에서 제공하고 있는 설계, 개발, 배포단계별로 검토 해야할 체크 리스트를 통해 내용을 점검 후 진행하는 것을 권장하고 있다.

셋째, 인공지능 서비스를 운영 시 1) 기획설계, 수명주가, 3) 모델 구축 및 검증, 4) 배포 및 모니터링, 5) 책임에 대한 각각 원칙과 관련 내용을 명확히 제시하고 이에 충족하고 있는지에 대해 점검을 해야 한다.

본 연구는 거버넌스 측면 중심으로 논의하고 있다. 인공지능 서비스의 거버넌스 측면은 궁극적으로 제공되는 서비스에 대한 위험 측면을 완화하려는 노력으의 일환이므로 등장할 수 있는 위험관리 측면에서도 연구가 이루어져야 한다.

우리는 인공지능에 제공하는 장점을 수용하면서 한계 및 위험 요소에 대한 적극적인 대응 방안으로 마련해야 한다. 인공지능 기술을 적극적 활용하여 효율적으로 정책 수립하여 고부가가치를 생성하고 사회에 의미 있는 영향을 제공할 수 있도록 노력해야 한다.

References

[1] African Development Bank, Organisation for Economic Co-operation and Development, United Nations, and World Bank. A Toolkit of Policy Options to Support Inclusive Green Growth. World Bank, 2012.

[2] Artificial Intelligence Risk Management Framework, NIST, National Institute of Standards and Technology

[3] Soonduck Yoo, "Research on the evaluation model for the impact of AI services." International Journal of Internet, Broadcasting and Communication 15, no. 3, 191-202, 2023

[4] Suresh, Harini, and John V. Guttag. "A framework for understanding unintended consequences of machine learning." arXiv preprint arXiv:1901.10002 2, no. 8, 2019.

[5] Ávalos, Roberto Sánchez, Felipe González, and Teresa Ortiz. "Responsible use of AI for public policy: Data science toolkit.", 2021.

[6] Andre Wirjo, Sylwyn Calizo Jr., Glacer Niño Vasquez, and Emmanuel A. San Andres, Artificial Intelligence in Economic Policymaking

[7] “AI Policy Observatory”, OECD 2019c

[8] Barocas, Solon, and Andrew D. Selbst. "Big data's disparate impact." California law review, 671-732 2016.

[9] Han-gyu Lim, “Study on the revision of copyright law on artificial intelligence”, Soongsil University Graduate School master’s thesis, 2022.

[10] Jobin, Anna, Marcello Ienca, and Effy Vayena. "The global landscape of AI ethics guidelines." Nature machine intelligence 1, no. 9 (2019): 389-399.

[11] Artificial Intelligence Governance Toolkit, 2022

[12] AI Governance Toolkit, 2022 linklaters

[13] Artificial Intelligence in Society, published by Korea Intelligence and Information Society Agency, November 2019.

저 자 소 개

유 순 덕(정회원)

- 1991년 2월 : 국민대학교 수학과 (학사)
- 1994년 2월 : 연세대학원 수학과 (이학석사)
- 1995년 12월 : 영국 뉴카슬 대학 응용수학 (석사)
- 2010년 3월 ~ 2013년 2월 : 한세대학교 IT융합박사
- 2013년 9월 ~ 현재 : 한세대학교 조교수
- 관심분야 : 전자금융, 창업 및 벤처, 빅데이터, 정부정책, 개인정보 및 보안