

<https://doi.org/10.7236/JIIBC.2024.24.6.1>
JIIBC 2024-6-1

생성형 인공지능(AI) 정보시스템의 데이터 품질 감리 점검 방안

Audit Inspection Method for Data Quality of Generative AI Information System

송춘광*, 장문석**, 박유리***, 전삼현****

Chun-Kwang Song*, Moon-Seok Jang**, Yu-Ri Park***, Sam-Hyun Chun****

요약 생성형 인공지능(AI)이 다양한 산업 분야에서 활용됨에 따라, 생성형 AI의 성능과 신뢰성의 핵심 요소인 양질의 데이터 확보가 중요하다. AI 모델을 효과적으로 학습시키기 위해서는 대규모 학습데이터의 구축과 데이터의 품질 확보가 필수적이며, 데이터 품질에 대한 감리(Audit)를 통해서 데이터 구축 절차별로 수행하는 작업에 대해서 객관적인 품질 점검을 통한 품질개선이 중요하다. 본 연구에서는 생성형 AI 기반 정보시스템을 위한 AI 학습데이터 품질 확보를 위해서, 데이터 구축 단계와 단계별 활동에 대해서, 절차, 산출물 및 성과의 관점으로 감리 점검항목과 세부 검토항목을 제시한다.

Abstract As generative artificial intelligence (AI) is used in various industries, it is important to secure high-quality data, which is a key factor in the performance and reliability of generative AI. In order to effectively train AI models, it is essential to build large-scale training data and secure data quality, and it is important to improve quality through objective quality checks for each data construction procedure through data quality audits. In this study, we present guidelines for audit and detailed review items from the perspective of procedures, outputs, and outcomes for each step of data construction and each activity to secure AI training data quality for generative AI-based information systems.

Key Words : Audit, Data Quality Management, Generative AI, Large-scale AI, Training Data

1. 서론

최근 인공지능, 데이터 중심의 디지털 심화 시대의 도래에 따라 디지털플랫폼정부위원회는 '디지털플랫폼정부 실현계획'을 통해서, 공개된 정부 문서를 AI가 읽을 수

있는 형태로 데이터화하고 학습함으로써 국민이 공공서비스를 쉽게 이용할 수 있도록 서비스 품질을 제고하고, 민간 초거대 AI를 적극 활용하여 복지, 재난, 청년 등 사회 현안을 해결하고자 초거대 AI 추진 방향을 발표하였다^[6]. 또한 각 기관은 정부의 추진 방향에 맞춰 공공부문

*송실대학교 IT정책경영학과

**송실대학교 IT정책경영학과

***송실대학교 IT정책경영학과

****송실대학교 IT정책경영학과

접수일자 2024년 10월 22일, 수정완료 2024년 11월 22일

게재확정일자 2024년 12월 6일

Received: 22 October, 2024 / Revised: 22 November, 2024 /

Accepted: 6 December, 2024

*Corresponding Author: shchun@ssu.ac.kr

Graduate School of IT Policy and Management, Soongsil University, Korea

에 생성형 AI 서비스 개발을 위해서 준비하고 있으며^[9], 공공부문 실무에 생성형 AI 기술을 적용하기 위한 정보 시스템 구축 사업이 점차 확대되고 있다.

생성형 AI 학습용 데이터 구축 사업은 생성형 AI 모델, 특히 GPT(Generative Pre-trained Transformer)와 같은 대규모 언어 모델을 성공적으로 학습시키기 위해서 필수적인 요소이며, 생성형 AI가 점점 다양한 산업 분야에서 활용됨에 따라, AI 모델을 효과적으로 학습시키기 위한 대규모 데이터 구축이 중요해졌다. 과학기술 정보통신부와 한국지능정보사회진흥원에서는 생성형 AI 모델을 위한 대규모 학습용 데이터 구축 및 관련 연구 지원을 위해서 초거대 AI 데이터 구축사업을 2023년부터 추진하고 있다^[6].

공공 정보시스템 구축 사업에서 생성형 AI가 핵심 기술로 주목받으면서 데이터의 정확성, 편향성 최소화, 일관성 및 신뢰성을 보장하는 것이 공공 서비스의 신뢰성 제고에 중요한 고려사항이며 AI 데이터의 품질 확보는 AI 시스템의 성능과 신뢰성을 높이고, 공공 서비스를 효과적으로 제공하는 데 중요한 역할을 한다. AI 데이터 품질은 AI 시스템의 품질과 신뢰성으로 연결되므로 데이터 품질 확보는 AI 시스템 신뢰성 확보에 있어 기본적인 출발점이라 볼 수 있다^[3].

그러나, AI 데이터 품질에 대한 객관적이고 체계적인 점검을 위해서, 감리 수행 시에 활용할 수 있는 데이터 품질에 대한 감리 점검 단계 및 감리 점검항목에 대한 제시가 미흡한 상태이다.

본 연구에서는 생성형 AI 정보시스템의 성능과 신뢰성에 중요한 요소인 생성형 AI 학습데이터에 대한 객관적인 품질 점검을 위해서, 생성형 AI 학습데이터 구축계획의 수립과 구축 절차별 활동에 대해서 절차와 산출물, 그리고 성과의 3가지 감리 관점과 기준을 제시하고, 감리 점검 수행시 활용할 수 있는 감리 점검항목과 세부 검토항목을 제시한다.

II. 관련 연구

초거대 AI 기술은 일반적인 인공지능과 비교하여 규모, 학습능력, 문제 해결 능력 등에서 발전을 이룬 AI 시스템으로써 대량의 데이터와 처리능력을 기반으로 더욱 복잡하고 광범위한 작업을 수행할 수 있는 AI의 진화된 형태이다^[7]. 생성형 AI는 초거대 AI 모델을 기반으로 텍스트나 이미지, 음성 등 새로운 정보, 콘텐츠 생성에 활

용된다. 초거대 AI는 대규모 학습용 데이터와 초거대 매개변수를 통해 사전학습(Pretrained)된 기반 모델(Foundation Model)을 다양한 방법으로 적응학습(Adaptation Learning)을 수행하며, 학습을 위한 이미지 캡션, 음성 등 다양한 유형의 데이터를 포함한다^[7].

공공 및 민간 분야에서 생성형 AI의 도입이 가속화되면서 데이터 품질의 중요성이 부각 되고 있으며 초거대 AI 모델을 개발하고 운영하는 과정에서, AI의 성능과 신뢰성은 데이터 품질에 따라 크게 좌우된다. 학습데이터가 AI 모델의 성능을 결정짓는 핵심 요소이기 때문에, 학습용 데이터의 구축 과정에서 발생할 수 있는 오류나 편향성, 일관성 결여 문제는 정보시스템의 성능과 공정성을 저해할 수 있다. 성공적인 AI 학습을 위해서는 양질의 데이터 확보가 중요하며, 이를 위한 데이터 품질 연구는 매우 중요하다^[5].

초거대 AI 데이터 관점의 품질관리 역량 확보를 위해서는 품질관리 방법 및 절차의 체계적인 접근이 필요하며 초거대 AI 데이터는 기존의 지도학습용 데이터와 다른 고유의 특징을 보유하고 있다^[7]. 초거대 AI 데이터는 매우 거대한 데이터가 필요하며, 초거대 AI 학습용 말뭉치 데이터는 자기 지도학습으로 사전학습을 수행하여 기존의 지도학습 데이터와 달리 라벨링이 없거나 부분적으로 수행한다. 또한, 공정 단계별 구축되는 지도학습용 데이터와 달리 초거대 AI 데이터는 사전 학습된 기반모델(Foundation Model)을 다양한 방법으로 반복하여 조정하는 구축 절차를 수행하며, 학습(Training), 검증(Validation), 테스트(Test) 데이터의 유형도 다양한 양상을 보인다^[7].

생성형 AI 기반 정보시스템 구축 사업에서 학습데이터의 품질은 구축되는 정보시스템의 성능과 요구되는 업무 기능에 대한 신뢰성에 영향을 주는 부분^{[1][2]}으로, AI 학습데이터 구축 시, 데이터 품질 확보를 위한 체계적인 데이터 품질관리가 중요하다.

AI 데이터 품질에 대해서 발주기관 및 구축 사업자로부터 독립된 자가, 제3자의 관점에서 객관적인 데이터 품질 점검을 위한 감리를 수행함에 있어서, 구축 대상 사업의 준비 계획, 구축, 학습 및 운용 등 절차별 수행 활동에 대한 적정성을 점검하고, 구축 활동 과정에서 작성하는 산출물에 대한 점검 그리고, 사업 목표 대비 성과의 점검 등 감리 관점과 기준을 수립하고, 체계적인 감리 점검방안에 따른 객관적인 감리의 수행이 중요하다.

한국지능정보사회진흥원의 지능정보기술 감리 실무 가이드('23년 2월)^[8]에서는 빅데이터 정보화 사업과 클라

우드 정보화 사업에 대한 감리를 수행함에 있어서 필요한 사업 단계별 감리 점검 항목을 제시하였으나, 인공지능 정보화 사업에 대해서는 감리 점검 항목이 제시되지 않았고, 인공지능 학습용 데이터품질관리 가이드라인('24년 1월)^[7]가 배포되어 AI 학습용 데이터 구축에 대한 데이터 품질관리 체계를 제시하였으나, 감리 수행을 위한 감리 관점과 감리 점검항목은 제시되지 않아서, 인공지능 관련 사업에 대한 감리 수행을 위한 실무 지침이 미흡한 상태이다.

인공지능 시스템 신뢰성 감리 프레임워크 논문에서는 인공지능 시스템의 신뢰성 관점의 감리 프레임워크에 기반하여 감리 점검항목을 제시하였고^[3], 인공지능 기술을 적용한 정보시스템의 감리 점검방안 논문에서는 인공지능 적용시스템의 구축 절차별 감리 점검항목을 제시하였으나^[4], 인공지능 시스템을 위한 데이터 품질 관점의 감리 점검방안은 제시되지 않았다.

III. 생성형 AI 데이터 품질에 대한 감리관점 및 감리 점검항목

한국지능정보사회진흥원에서 발간된 '인공지능 학습용 데이터 품질관리 가이드라인(v3.1)'^[8]에서는 인공지능 학습용 데이터 구축시 구축계획 수립 단계부터 데이터 획득·수집, 정제, 가공, 학습 및 운영·활용 단계까지 절차, 산출물 및 품질관리 활동을 제시하고 있으며, 초거대 AI 학습용 데이터 구축을 위한 기초자료로도 활용할 수 있도록 관련 품질관리 활동을 제시하고 있다.

인공지능 학습용 데이터 품질관리 가이드라인에서 제시한 데이터 품질관리 활동에 대해서, 구축 절차와 절차별로 생성되는 산출물 그리고, 성과의 3가지 감리 관점 및 기준을 제시하고, 구축 사업의 단계별 활동에 대해서 감리 점검을 위한 감리 점검항목과 세부 검토항목을 제시한다.

1. 감리 관점 및 기준

감리 수행의 관점과 기준으로, 절차, 산출물 및 성과의 3가지 관점을 제시한다. 절차는 사업에 대한 구축 운영 계획 및 절차 수립의 적정성을 검토하며, 산출물은 구축 절차에 따라 생산되는 문서 등에 대한 적정성의 검토, 그리고, 성과는 사업의 성과목표 및 기대효과의 달성 가능성 및 달성 여부에 대한 검토이다.

표 1. 감리관점 및 점검기준 코드표

Table 1. Code table of Audit viewpoint and Inspection Criteria

항목	코드	명칭
절차	1-1	계획 적정성
	1-2	절차 적정성
	1-3	준수성
산출물	2-1	기능성
	2-2	무결성
	2-3	편의성
	2-4	안정성
	2-5	보안성
	2-6	효율성
	2-7	준거성
	2-8	일관성
성과	3-1	실현성
	3-2	충족성

2. 준비 계획단계 절차에 대한 감리 점검항목

가. 구축계획수립 절차의 감리 점검항목

구축계획수립 절차는 인공지능 데이터 구축에 필요한 목적, 범위 및 절차 등에 대한 사업수행계획을 수립하는 단계이며, 데이터 구축 시 목적에 대한 명확한 정의 여부, 데이터 구축 절차 및 조직 구성의 적정성 여부, 임무 정의에 대한 적정성 여부, 법·제도 준수, 품질목표 및 점검기준에 대한 적정성을 중점 점검한다.

표 2. 구축계획수립 절차 점검항목

Table 2. Checklist for the construction plan establishment

점검 항목	검토항목	감리 관점
1. 사업수행 및 구축계획을 적정하게 수립하였는지 여부	1-1.사업수행계획서에 학습 데이터 구축목적 및 범위를 명확하게 정의하였고, 수행계획을 적정하게 수립하였는가?	1-1 2-1 3-2
	1-2.사업수행계획서 및 품질관리계획서에 품질 관리방안을 적정하게 수립하였는가?	1-1 2-1 3-2
	1-3.사업수행계획서 및 품질관리계획서에 품질 목표 수준을 적정하게 수립하였고, 품질관리 기준을 명확하게 정의하였는가?	1-1 2-1 3-2
2. 데이터 구축절차 및 조직 구성이 적정한지 여부	2-1.데이터 명세 정보 및 데이터 구성을 명확하게 정의하였는가?	1-1 2-2,2-8 3-2
	2-2.데이터 구축을 위한 절차, 조직, 도구 등 역할과 책임 수립에 대해서 적정하게 검토하였는가?	1-1 2-1 3-2
	2-3.데이터셋 개요, 구성내역, 설계기준 및 분포 현황, 구조, 구축 과정, 검수 및 품질활동 내역을 검토하였는가?	1-1 2-1 3-2

3. 임무정의에 대한 적절성 검토 여부	3-1. 구축목적의 일차성과 일관성 확보를 위한 계획을 충분히 검토하였는가?	1-1 2-8 3-2
4. 품질목표 및 점검기준 수립 여부	4-1. 사업수행기관이 제출한 사업수행계획서 등 검증관련 자료를 토대로 품질지표기준을 수립하고, 품질검증 지표 및 목표에 대해 합의하였는가?	1-1 2-1 3-2

3. 구축단계 절차에 대한 감리 점검항목

가. 데이터 획득/수집 절차의 감리 점검항목

데이터 획득/수집 절차는 인공지능 학습에 필요한 데이터를 획득/수집 또는 생성하는 단계이며, 데이터 획득/수집 방법 및 기준의 현행화 여부, 데이터 법적 근거의 검토여부, 수집 도구 및 저장환경의 검토여부, 원시데이터 품질검사 수행 여부 등, 범·제도 준수, 데이터 다양성 확보, 데이터 편향성 방지 및 윤리 준수의 적정성을 중점 점검한다.

표 3. 데이터 획득/수집 절차 점검항목
Table 3. Checklist for data acquisition/collection procedure

점검 항목	검토항목	감리 관점
1. 획득/수집방법 및 기준의 현행화 여부	1-1. 데이터 획득 및 확보 방법을 현실적 가능성에 맞게 계획을 수립하고, 절차와 법 제도적 적절성을 검토하였는가?	1-3 2-7 3-2
	1-2. 획득항목 및 구축량 획득/수집단계 품질기준 및 조직, 절차를 수립하였고, 도구를 확보하였는가?	1-3 2-7 3-2
	1-3. 데이터 획득 방법 및 기준, 교육, 검사에 대한 체계를 수립하였는가?	1-2 2-7 3-2
	1-4. 데이터 획득 기준에 대한 변경 관리를 수행하였는가?	1-3 2-2,2-7 3-2
2. 데이터 법적 근거의 검토 여부	2-1. 법 제도 준수(개인정보, 초상권, 저작권) 근거자료를 확보하였는가?	1-3 2-5,2-7 3-2
	2-2. 데이터 획득 및 수집의 편향성 방지를 위한 방안을 수립하였는가?	1-2 2-2,2-4,2-7 3-2
3. 수집 도구 및 저장환경의 검토 여부	3-1. 획득/수집에 사용될 도구를 확보하고 사용 방법을 검토하였는가?	2-1,2-3 3-2
	3-2. 획득/수집 데이터 저장환경을 적절하게 구축하였는가?	2-4 3-1
4. 원시데이터 품질검사 수행 여부	4-1. 원시데이터가 인공지능을 활용하여 문제를 해결하는데 요구되는 다양성, 신뢰성, 충분성, 균일성, 사실성, 공정성 등 기준에 적합하게 구성되어 있는지 확인하였는가?	1-2 2-1,2-8 3-1
	4-2. 파일포맷, 해상도, 프레임 레이트 등 기술 적합성을 확인하였는가?	1-2 2-1,2-2,2-7 3-1

4-3. 데이터에 대한 클래스 분포, 인스턴스 분포, 분석 가능 항목의 통계적 분포를 확인하였는가?	1-2 2-1 3-1
---	----------------

나. 데이터 정제 절차의 감리 점검항목

데이터 정제 절차는 획득한 원시데이터로부터 불필요한 중복 제거 및 개인정보 비식별화 등 일련의 전처리를 수행하는 단계이며, 정제방법 및 기준 현행화 여부, 개인정보 비식별화 여부, 원천데이터 품질검사 수행 여부 등 정제기준의 명확성, 중복방지, 정제작업방식의 적정성을 중점 점검한다.

표 4. 데이터 정제 절차 점검항목
Table 4. Checklist for data cleansing procedure

점검 항목	검토항목	감리 관점
1. 정제방법 및 기준 현행화	1-1. 구축목적에 맞게 적절한 데이터 선별을 위한 정제기준을 명확하게 수립하였는가?	1-1 2-1,2-7,2-8 3-1
	1-2. 정제단계별 품질 기준, 조직, 절차를 적절하게 구성하였고, 도구를 확보하였는가?	1-2 2-1 3-1
	1-3. 데이터 정제 방법 및 정제 결과에 대한 교육 및 훈련을 적절하게 수행하였는가?	1-2 2-1 3-1
	1-4. 데이터 정제 결과에 대한 검사 절차와 기준을 수립하였는가?	1-2 2-1,2-7 3-2
	1-5. 데이터 정제 기준에 대한 변경 관리를 수행하였는가?	1-3 2-8 3-2
2. 개인정보/민감정보 비식별화 등 법적 준거 확보 여부	2-1. 개인정보보호 등 비식별화 기준 및 방안 등 법적 준거를 적절하게 확보하였는가?	1-3 2-5,2-8 3-2
3. 정제 도구 및 저장환경의 검토 여부	3-1. 정제단계에 사용될 저장도구를 확보하였고 사용방법을 숙지하였는가?	1-2 2-1 3-1
	3-2. 정제 데이터 저장환경을 적절하게 구축하였는가?	1-2 2-4,2-5 3-1
4. 원천데이터 품질검사 수행 여부	4-1. 사용하기 적합한 데이터 여부에 대한 품질검사를 수행하였는가?	1-3 2-2,2-8 3-1
	4-2. 원천데이터 품질검사를 위한 검사 도구를 확보하였고 검사결과를 충분히 검토하였는가?	1-3 2-1,2-7,2-8 3-1
	4-3. 품질검사결과에 따른 보완조치를 적절하게 수행하였는가?	1-3 2-1,2-7,2-8 3-2

다. 데이터 가공 절차의 감리 점검항목

데이터 가공 절차는 AI모델의 기능과 목적에 맞게 가

공 처리하는 절차로, 초거대 AI 데이터는 적응학습 (Adaptation Learning)¹⁷⁾을 위한 데이터셋을 구축하는 단계이며, 데이터 가공방법 및 기준의 현행화 여부, Adaptation Learning 데이터 품질검사 수행여부 등 가공 처리 활동별 적정성을 중점 점검한다.

표 5. 데이터 가공 절차 점검항목
 Table 5. Checklist for data processing procedure

점검 항목	검토항목	감리 관점
1. 데이터 가공 방법 및 기준의 현행화 여부	1-1. Adaptation Learning 데이터구축방법 및 기준을 명확하게 정의하였는가?	1-1 2-1,2-2,2-8 3-2
	1-2. Adaptation Learning에 필요한 데이터가 적정하게 구축되었는가?	1-2 2-1,2-2,2-8 3-2
	1-3. 데이터 형식 및 입력값의 범위에 대한 기준을 명확하게 마련하였는가?	1-1 2-2,2-7 3-2
	1-4. 데이터 가공 방법 및 기준, 교육, 검사체계를 적정하게 수립하였는가?	1-1 2-7,2-8 3-2
	1-5. 데이터 가공 기준에 대한 변경 관리를 수행하였는가?	1-3 2-2,2-8 3-1
2. 데이터 가공 도구 및 저장환경의 검토 여부	2-1. 가공에 사용될 저작도구를 확보하였고, 사용 방법을 충분히 검토하였는가?	1-2 2-1 3-1
	2-2. 데이터 가공 기준에 대한 변경 관리를 수행하였는가?	1-2 2-2,2-8 3-1
	2-3. Adaptation Learning 용 데이터 저장환경을 구축하였는가?	1-2 2-4 3-1
3. Adaptation Learning 데이터 품질검사 수행 여부	3-1. 정확성(구문, 의미)을 중점으로 검사를 수행하였는가?	1-2 2-2,2-7,2-8 3-1
	3-2. Adaptation Learning 용 데이터 품질검사를 위한검사 도구를 확보하였고, 검사 결과를 충분히 검토하였는가?	1-2 2-2,2-8 3-1
	3-3. 품질검사 결과에 따른 보완 조치를 적정하게 수행하였는가?	1-2 2-2,2-8 3-1

라. 데이터 학습 절차의 감리 점검항목

데이터 학습 절차는 AI 모델을 학습시키고 학습된 AI 모델의 성능을 향상시키거나 보정하는 활동을 수행하는 단계이며, 구축목적 AI모델 합치성 확인, 학습결과 확인 및 최적화, 품질검증결과 보완조치 여부 등, 데이터 학습의 유효성, AI모델 성능, 품질오류에 대한 분석 활동의 적정성을 중점 점검한다.

표 6. 데이터 학습 절차 점검항목
 Table 6. Checklist for data learning procedure

점검 항목	검토항목	감리 관점
1. 구축목적-AI모델 합치성 확인	1-1. 데이터를 인공지능 모델로 학습시켰을 때 영역별 품질검사 기준을 확인하였는가?	1-3 2-1,2-2,2-8 3-1
	1-2. 구축 목적에 부합하는 AI 모델의 선정 여부를 확인하였는가?	1-3 2-2,2-7,2-8 3-1
	1-3. AI 모델 구축/운영 및 테스트를 위한 환경 구축 정보를 검토하였는가?	1-3 2-4,2-8 3-1
	1-4. AI 모델 정보/라이선스 가이드를 충분히 검토하였는가?	1-3 2-7,2-8 3-1
2. 학습결과 확인 및 최적화	2-1. 학습용 데이터가 AI 모델로 학습시켰을 때 목표로 했던 수준의 성능 달성 가능성이 충분히 검사하였는가?	1-3 2-1,2-4 3-1
	2-2. AI 모델(알고리즘)의 성능 검사를 계획에 따라 적정하게 수행하였는가?	1-3 2-4,2-7 3-1
	2-3. 모델의 단계별 유효성을 적정하게 검사하였는가?	1-3 2-2,2-6,2-7 3-1
	2-4. 학습모델, 활용 서비스 개발에 가능한 학습데이터셋 품질오류 및 이슈 내용을 분석하여 보완조치를 반영하였는가?	1-3 2-4,2-7 3-1
3. 품질검증결과 보완조치	3-1. 최종데이터 제3자 품질검증결과가 미달성인 경우, 보완조치를 수행하여 결과를 적정하게 보고하였는가?	1-3 2-7 3-1

4. 운영·활용 절차에 대한 감리 점검항목

가. 데이터 운영/활용 절차의 감리 점검항목

데이터 운영/활용 절차는 데이터 하자 및 유지보수, 품질개선 활동을 수행하는 단계이며, 데이터 하자 및 유지보수 수행 여부, 사용자 품질개선 의견의 반영여부 등, 유지보수 계획 및 품질개선 활동의 적정성에 대해서 중점 점검한다.

표 7. 데이터 운영/활용 절차 점검항목
 Table 7. Checklist for data operation/utilization procedure

점검 항목	검토항목	감리 관점
1. 데이터 하자 및 유지보수 수행 여부	1-1. 구축 데이터에 대한 하자보수 및 유지보수를 적정하게 수행하였는가?	1-3 2-4,2-7 3-1
2. 사용자 품질개선요건의 반영 여부	2-1. 데이터 품질에 대한 하자 및 유지보수 계획을 수립하였는가?	1-1 2-7,2-8 3-2
	2-2. 사용자들의 품질에 대한 개선 의견이 있는 경우 반영 등 유지보수를 적정하게 수행하였는가?	1-2 2-3,2-6 3-1

IV. 결 론

본 연구는 생성형 AI를 적용한 정보시스템의 성능과 신뢰성에 중요한 영향을 미치는 핵심 요소인 데이터 품질을 확보하기 위해서, AI 학습데이터 구축 단계별 세부 활동들에 대해서 절차, 산출물 및 성과의 3가지 감리 관점을 제시하였고, 감리원의 감리 업무 수행시 참고할 수 있는 감리 점검항목과 세부 검토항목을 제시하였다.

생성형 AI 데이터에 대한 구축 단계별 감리 수행을 통해서 준비 계획, 구축, 학습 및 운용 단계 활동의 적정성을 점검하고 객관적인 데이터 품질 점검과 미흡한 사항에 대한 개선 권고 활동을 통해서, 데이터 품질을 확보하고 생성형 AI 정보시스템의 성능과 신뢰성 제고에 도움이 될 것으로 기대된다.

References

- [1] Hyun-mok Oh, Seo-youn Lee, Young-hoon Chang, "A Study on the Artificial Intelligence (AI) Training Data Quality : Fuzzy-set Qualitative Comparative Analysis (fsQCA) Approach", Information Systems Review, Vol. 26, No. 1, pp. 19-56, 2024.
<https://www.earticle.net/Article/A443444>
- [2] Soo-Sang Kim, "A Study on the Factors Affecting Continuous Usage Intention of Data Driven Artificial Intelligence Product Service System", Doctoral Dissertation, 2019.
<http://oasis.dcollection.net/common/orgView/200000214845>
- [3] Chan-Woo Jung, Hang-Bae Jang, Young-Ji Ju, Jung-Duk Kim, "A Study of Artificial Intelligence System Reliability Audit Framework", Korean Journal of Industry Security, Vol. 13, No. 1, pp. 29-54, 2023.
<https://www.earticle.net/Article/A429816>
- [4] Jong-Bae Kim, Yun-Jin Kim, Hyun-seok Jeong, Jin-ho Park, "Guidelines for Audit of Information Systems using Artificial Intelligence Technology", Journal of The Korea Society of Information Technology Policy & Management, Vol. 11, No. 4, pp. 1329-1337, 2019.
<https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artilId=ART002493304>
- [5] Hyun-jung Lee, Jae-ho Lee, Eui-hyun Jung, Su-wook Ha, "Data Quality Metrics of Bias for Artificial Intelligence", Conference of IEIE, 2021.
<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE10591578>
- [6] Digital Platform Government Committee, National Information Society Agency, "Guidelines for the Introduction and Utilization of AI in the Public Sector", 2024.
<https://www.dpg.go.kr/DPG/contents/DPG03020000.do?schM=view&id=20240423103225685873&schBcid=reference>
- [7] National Information Society Agency, "Data Quality Management Guidelines for AI Training(v3.1)", 2024.
<https://aihub.or.kr/aihubnews/qlityguidance/view.do?currMenu=135&topMenu=103&nttSn=10269>
- [8] Eun-joo Hong, "Public Digital Innovation, Preparing for Large scale AI and Cloud", Insight Report, 2024.
<https://www.samsungds.com/kr/insights/digital-governent-innovation.html>
- [9] Yong-Man Ha, Jong-Wook Jang, "Effect Analysis of a Artificial Intelligence Attention Redirection Compensation Strategy System on the Data Labeling Work Attention Concentration of Individuals with Developmental Disabilities", The Journal of the Institute of Internet, Broadcasting and Communication, Vol. 24, No. 2, pp. 119-125, 2024.
<https://www.earticle.net/Article/A448264>
- [10] Sang-Uk Kim, Ji-Hwan Ha, "Development of LncRNA-Disease Associations Prediction Model using CNN-RNN Feature Extraction based on LncRNA Sequence Data", The Journal of KIIT, Vol. 22, No. 10, pp. 1-11, 2024.
DOI: <https://doi.org/10.14801/jkiit.2024.22.10.1>
- [11] Nak-Hun Choi, Jong-Seok Oh, Jong-Rok Ahn, Key-Sun Kim, "A Development of Defeat Prediction Model Using Machine Learning in Polyurethane Foaming Process for Automotive Seat", Journal of the Korea Academia-Industrial cooperation Society, Vol. 22, No. 6, pp. 36-42, 2021.
<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE10572371>

저 자 소 개

송 춘 광(정회원)



- 송실대학교 일반대학원 IT정책경영학과 박사과정
- 서강대학교 정보통신대학원 석사 졸업
- 관심분야 : 방송미디어, 인공지능, 빅데이터 분석

장 문 석(정회원)



- 송실대학교 일반대학원 IT정책경영학과 박사과정
- 성균관대학교 정보통신대학원 석사 졸업
- 관심분야 : 정보보안, ISMS, 개인정보, 인공지능, 빅데이터 분석

박 유 리(정회원)



- 송실대학교 일반대학원 IT정책경영학과 박사과정
- 송실대학교 일반대학원 석사 졸업
- 관심분야 : 정보보안, 개인정보보호, 정보보호교육, IT정책 등

전 삼 현(정회원)



- 송실대학교 일반대학원 IT정책경영학과 교수
- 독일 Frankfurt University 법학박사
- 관심분야 : IT정책경영, 자본시장법, 회사법