

# AI기술 활용 통계의 등장 배경과 품질 기준의 필요성

The Background of AI Technology Utilization Statistics  
and Necessity of Quality Standards



김정민 선임연구원  
소프트웨어정책연구소  
디지털통계센터  
jungmink26@spri.kr

전이슬  
소프트웨어정책연구소  
디지털통계센터  
jys@spri.kr

## Executive Summary

국내외를 막론한 조사환경 악화와 데이터 확보 유형의 증가는 기존 통계자료의 대안을 고려하는 계기로 작용하고 있다. 특히 민간 기업들은 자체 보유한 정보를 토대로 시의성과 상세성을 갖춘 통계를 제공하고 있으며 이는 데이터의 대표성 및 신뢰도 검증이 어려운 구조적 문제를 안고 있음에도 통계 수요자로 하여금 각광받는 추세이다. 각국의 통계기관은 이를 국가통계의 위기로 바라보고 국가통계 혁신을 위한 활로 모색에 한창이다.

한국은 2021년 말 유럽에서 시행 중인 제도를 참고해 실험적 통계 제도를 신설하였고, 이를 통해 국가통계 작성 방법에 다양한 통계 생산 방식을 수용하고자 하는 의지를 표명하였다. 실험적 통계는 국가가 유예기간을 두고 새로운 방식을 적용한 통계의 품질을 관리 후 국가통계 승인 여부를 결정하도록 한 제도로서, 유예기간 동안 AI기술을 활용해 작성한 통계를 민간에 공식적으로 공개 가능하다는 장점을 지닌다. 더불어 2022년 통계청은 빅데이터 활용했거나 AI기술을 작성 기법으로 활용하는 통계를 데이터과학 활용 통계라 정의함으로써 관련 통계의 등장 가능성을 공식적으로 언급하였다. 전통적 통계 기법 외 작성기법에 대해 배타적인 입장을 견지해왔던 통계 분야로서 최근 행보는 괄목할만한 변화로 볼 수 있다.

이처럼 통계작성에 AI기술 활용의 기회가 개방되었음에도 현실적인 장애물이 존재하는 것으로 나타났다. 품질 문제가 핵심으로, 글로벌 관점에서도 논의 초기다 보니 통계로서 갖추어야 할 품질 요소가 아직 불명확하다는 지적이다.

본 보고서는 이러한 문제 인식에 기반하여 AI기술 활용 통계에 고려되어야 할 품질 지표(안)를 제시하고자 하였다. 이를 위해 첫 번째로 해외 국가 또는 기관의 관련 문헌들을 검토하였으며, 두 번째로는 과학기술정보통신부와 한국정보통신기술협회가 발간한 「2022 신뢰할 수 있는 인공지능 개발 안내서(안)」를 기준하여 AI기술 활용 통계에서 공통적으로 고려할만한 품질 요소를 선별하였다. 분석 결과 총 10가지 유형의 품질 지표를 도출하였으며 세부 내용은 다음과 같이 제시되었다.

시기술 활용 통계의 품질 지표(안)

진단 지표		세부 내용
방법론	설명 가능성	· (정의) 결과로 도출되는 산출물뿐만 아니라 분석에 활용되는 알고리즘의 기초가 되는 논리를 설명할 수 있는 것 · (측정 방향) 통계 이용자에게 시모델의 세부적인 명세와 작동 원리 등을 상세히 제공하고 이해도를 높이기 위한 노력을 평가
	정확성	· (정의) 측정하도록 설계된 현상을 정확하게 설명하는 수준 · (측정 방향) 정확도(Accuracy), 정밀도(Precision), 재현율(Recall) 등 알려진 성능 척도 활용 · (제도적 고려) 불확실성의 정량화 - 시모델의 성능 지표로 활용되는 수치 각각의 구간별 의미를 부여해야만 다수 모델의 평가를 객관적인 기준상에서 수행할 수 있음 - 성능 관리 측면에서의 인적 책임 명확화
	재현성	· (정의) 동일 실험의 재현을 통해 결과를 복제할 수 있는 것 · (측정 방향) 품질 합의사항에 따른 기준으로 평가가 가능할 것임 · (제도적 고려) 시방법론 자체의 비결정적 특성을 전제로 Method reproducibility를 만족하기 위한 최적화 기준을 정할 필요
	편중성 <sup>1)</sup>	· (정의) 시모델 생성 과정의 학습용 데이터 참조의 편중 정도 · (측정 방향) 시모델 편향성 측정에 활용되는 알려진 방법론 활용 · (제도적 고려) 불확실성의 정량화 - 시모델의 편향성 측정에 활용되는 수치 각각의 구간별 의미를 부여해야만 통계적 활용에 적합한 시모델을 판단 가능
	지속 가능성	· (정의) 시성능 유지를 위한 신규 학습용 데이터 추가에 기반한 재학습 · (측정 방향) 품질 합의사항에 따른 기준으로 평가가 가능할 것임 · (제도적 고려) - 재학습 주기 기준 마련(성능 척도와 연계) - 신규 학습 데이터 세트 구축 소요 기간에 따른 통계 공표 일시 중단 주기 고려 - 시모델 성능 평가를 위한 학습용 데이터 기준 명확화
수집 데이터	편향성	· (정의) 데이터의 내재적·외재적 속성이 편향을 지니는 특성 · (측정 방향) 시신뢰성 국제 기준상의 보호 변수 포함 여부, 데이터 편향성과 관련된 통계 분야의 합의 기준 등을 활용 · (제도적 고려) 통계 작성 목적의 시학습용 데이터 편향 기준 마련
	외적 타당성	· (정의) 관측된 효과 및 관계를 규모가 더 크거나 상이한 인구, 환경, 상황 등으로 이전할 수 있는 것 · (측정 방향) 데이터 출처, 특성 등을 정성적으로 분석해 모집단 추정이 가능할 것인지 판단 · (제도적 고려) 통계의 목적과 특성을 고려해 외적 타당성을 선택적 요건화 하는 방안도 검토 주제일 것임
통계 운영 환경	시스템 안정성	· (정의) 통계 데이터 수집, 가공, 공표가 이루어지는 환경의 변화 또는 장애발생에 의해 통계 공표에 악영향을 끼치는 수준 · (측정 방향) SW시스템의 네트워크 사양 및 유사시 대처 방안 등 · (관련 고려사항) 수집 데이터가 휘발성을 지닐 시, SW시스템의 일시적인 문제가 통계의 시계열 안정성을 위협할 가능성
경제성	적시성	· (정의) 개념회에서 알고리즘 구현, 처리, 통계 생산에 이르는 결과를 도출하는데 소요되는 시간 · (측정 방향) 공표 주기에 따라 상이하나 통상적 기준에 맞춰 통계작성이 가능한지 검토
	비용 효과	· (정의) 자본 투입 대비 산출의 효과성 · (측정 방향) 고정 비용(fixed Cost)과 지속적 비용(ongoing Cost)의 효율 평가

1) 시알고리즘의 최적화를 위해 특정 데이터의 반영 비중을 높게 가져가는 일련의 동작 과정은 높은 성능을 보장하기 위한 고의적 여과(filtering)로도 볼 수 있어 편향이 아닌 편중으로 정의

### Executive Summary

The deterioration of the survey environment and the increase in the type of data securing are acting as an opportunity to consider alternatives to existing statistical data at domestic and abroad. In particular, private companies provide timely and detailed statistics based on their own information, which is a trend that statistical consumers are in the spotlight even though they have structural problems that are difficult to verify the representativeness and reliability of data. Statistical institutions in each country view this as a crisis in national statistics and are in the midst of seeking a way to innovate national statistics.

At the end of 2021, Korea established an experimental statistical system by referring to the system being implemented in Europe, and through this, it expressed its willingness to accept various statistical production methods in the method of preparing national statistics. Experimental statistics are a system in which the state manages the quality of statistics applied with a grace period and then decides whether to approve national statistics, and has the advantage of being able to officially publish statistics prepared using AI technology to the private sector during the management period. Subsequently, in 2022, the National Statistical Office officially mentioned the possibility of the emergence of related statistics by defining statistics that utilize big data or use AI technology as writing techniques as data science utilization statistics. In addition to traditional statistical techniques, it is a statistical field that has maintained an exclusive position on writing techniques, and its recent move can be seen as a remarkable change.

As such, it was found that there are realistic obstacles even though the opportunity to use AI technology was opened for statistical composition. It is pointed out that the quality factor that needs to be equipped as a statistic is still unclear as the quality issue is at the core and the discussion is in the early stages from a global perspective.

Based on this problem recognition, this report attempted to present quality indicators (Drafts) to be considered in AI technology utilization statistics. To this end, related documents from overseas countries or institutions were reviewed first, and the quality factors that AI technology utilization statistics could consider in common were selected based on the 2022 Reliable Artificial Intelligence Development Guide published by the Ministry of Science and ICT and the Korea Information and Communication Technology Association. As a result of the analysis, a total of 10 types of quality indicators were derived, and details were presented as follows.

## Quality indicators of AI technology utilization statistics (Draft)

Diagnostic indicators		Details
Methodology	Explainability	<ul style="list-style-type: none"> <li>· (Definitions) capable of explaining the logic underlying the algorithm used for analysis as well as the output derived from the results</li> <li>· (Measurement Direction) To provide statistical users with detailed specifications of the AI model, operating principles, etc. and to evaluate efforts to improve their understanding</li> </ul>
	Accuracy	<ul style="list-style-type: none"> <li>· (Definitions) Level that accurately describes the phenomenon designed to be measured</li> <li>· Utilize known performance scales such as accuracy, precision, and reproducibility (direction of measurement)</li> <li>· Quantification of uncertainty (institutional considerations) <ul style="list-style-type: none"> <li>- The evaluation of many models can be carried out on an objective basis only when the meaning of each section of the numerical values used as performance indicators of AI models is given</li> <li>- Clarify human responsibility in terms of performance management</li> </ul> </li> </ul>
	Reproducibility	<ul style="list-style-type: none"> <li>· (Definition) reproduction of the same experiment to replicate the results</li> <li>· (Measurement direction) Evaluation will be possible based on the quality agreement</li> <li>· (Institutional consideration) Based on the non-deterministic nature of the AI methodology itself, it is necessary to establish optimization criteria to satisfy method reproducibility</li> </ul>
	Partiality <sup>1)</sup>	<ul style="list-style-type: none"> <li>· (Definitions) Degree of bias in reference to learning data in AI model creation process</li> <li>· (Measurement Direction) Utilize known methodologies used to measure bias in AI models</li> <li>· Quantification of uncertainty (institutional considerations) <ul style="list-style-type: none"> <li>- It is possible to determine the AI model suitable for statistical use only by giving the meaning of each section of the numerical value used to measure the bias of the AI model</li> </ul> </li> </ul>
	Sustainability	<ul style="list-style-type: none"> <li>· (Definition) Re-learning based on the addition of new learning data to maintain AI performance</li> <li>· (Measurement direction) Evaluation will be possible based on the quality agreement</li> <li>· (Institutional Consideration) <ul style="list-style-type: none"> <li>- Establishing criteria for re-learning cycles (linked to performance scales)</li> <li>- Consider how often statistics are suspended according to how long it takes to build a new learning dataset</li> <li>- Clarification of learning data criteria for AI model performance evaluation</li> </ul> </li> </ul>
Collected data	Deflection	<ul style="list-style-type: none"> <li>· (Definitions) Characteristics of internal and external attributes of data being biased</li> <li>· (Measurement direction) Using the criteria for the inclusion of protection variables in international standards of AI reliability and the criteria for consensus in statistical fields related to data bias, etc</li> <li>· (institutional consideration) Establishment of data bias criteria for AI learning for statistical purposes</li> </ul>
	External validity	<ul style="list-style-type: none"> <li>· (Definitions) The ability to transfer observed effects and relationships to larger or different populations, environments, situations, etc</li> <li>· (Measurement direction) Qualitative analysis of data sources, characteristics, etc. to determine whether population estimation is possible</li> <li>· (institutional consideration) A plan to selectively require external validity in consideration of the purpose and characteristics of statistics will also be a subject of review</li> </ul>
Statistical operational environment	System stability	<ul style="list-style-type: none"> <li>· (Definitions) Levels that adversely affect statistical publication due to changes in the environment in which statistical data are collected, processed, or published or to the occurrence of a failure</li> <li>· (Measurement direction) Network specification of SW system and countermeasures in case of emergency, etc</li> <li>· (Related Considerations) The possibility that temporary problems in SW systems threaten the time series stability of statistics when collected data is volatile</li> </ul>
Affordability	Timeliness	<ul style="list-style-type: none"> <li>· (Definitions) Time taken to produce results from conceptualization to algorithm implementation, processing, and statistical production</li> <li>· (Measurement direction) Review whether statistics can be prepared according to normal standards, although they differ according to the publication cycle</li> </ul>
	Cost effectiveness	<ul style="list-style-type: none"> <li>· (Definitions) Effectiveness of calculation relative to capital input</li> <li>· (Measurement Direction) Evaluation of efficiency of fixed cost and longing cost</li> </ul>

1) A series of operational processes that take a high proportion of specific data to optimize AI algorithms can also be seen as intentional filtering to ensure high performance, which is defined as biased rather than biased.

## 서론

### ▣ 국내외를 막론한 조사환경 악화와 데이터 확보 유형의 증가는 기존 통계자료의 대안을 고려하는 계기로 작용

○ 수년간에 걸쳐 1차 데이터 수집 시(면접조사) 응답률이 하락하고 민간 기업의 통계 생산이 용이해짐에 따라 국가통계기관의 역할을 우려하는 시선 증가<sup>2)</sup>

-면접조사는 원천적 접촉 불가능, 조사 거부, 접촉 곤란, 응답부담 호소 등으로 인해 가구는 물론 사업체 대상의 조사도 어려움을 겪는 상황<sup>3)</sup>

-민간 기업의 통계는 데이터의 대표성·신뢰도에 관한 투명한 검증이 어려운 한계\*가 있음에도, 시의성과 상세성을 갖춰 통계 수요자에게 각광받는 추세

\* 통상적으로 민간 기업은 통계 산출을 위해 자사의 내부 데이터를 활용하므로 외부에서 통계 생산 체계를 파악하기 어려움

○ 각국의 통계기관은 기존의 통계를 보완하거나 대체하는 데 활용 가능한 새로운 유형의 데이터 및 방법론 도입을 추진하며 활로를 모색

-새로운 데이터 유형으로서 행정자료\*와 민간 빅데이터\*\*가 주목받는 중

\* 국내 통계청은 국가통계 작성환경의 변화에 대응하여 개정된 「행정자료 이용에 관한 법률」(2015)에 입각해 꾸준히 행정자료 기반 통계 개발을 확대

\*\* 민간 기업에서 생산하는 정형데이터(모바일 사용량, 기업 공시 등) 위주로 활용 검토가 확대되는 추세이나 현재까지는 국가승인통계 외 수량적 데이터로서 관리

-통계 생산 프로세스의 현대화, 신규통계 생산 등을 목적으로 한 시기술 도입 또한 활발히 연구되고 있는 상황<sup>4)</sup>

\* 무응답 대체(Imputation), 이상치 탐색(Outlier detection) 등 통계 신뢰성 증가를 위해 필요한 과업에 시기술을 도입 하는 시도는 이미 세계 여러 국가에서 채택하였으며, 한국 또한 통계청에서 관련 사업이 추진 중인 것으로 확인

2) UNECE(2018), "The use of machine learning in official statistics"

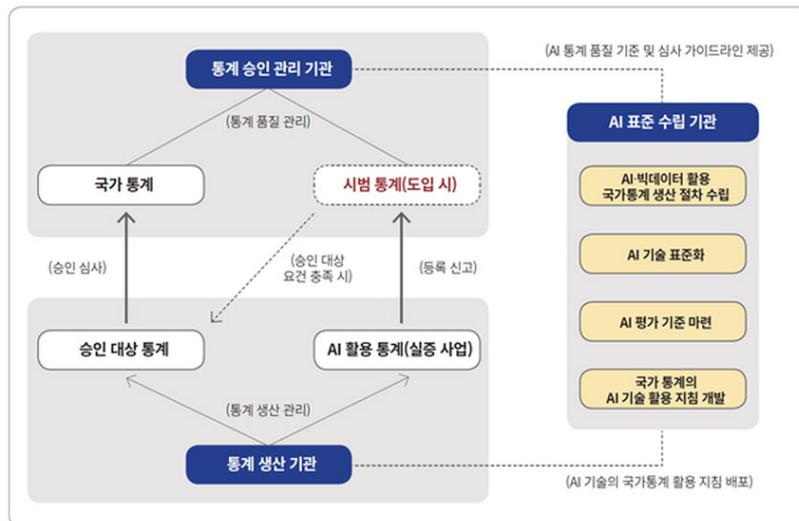
3) 통계청(2019), "빅데이터 활용통계의 국가통계 승인관리방안 연구"

4) 소프트웨어정책연구소(2021), "국가통계 시도입 활성화 방안 연구"

■ 한국은 2021년 말 유럽의 실험적통계(experimental statistics) 제도를 참고한 신규 통계 관리 제도를 신설함으로써 향후 국가통계 작성 방법에 다양한 통계 생산 방식을 수용하고자 하는 의지를 표명

- 실험적통계는 국가가 유예기간을 두고 새로운 방식을 적용한 통계의 품질을 검증 후 국가통계 승인 여부를 결정하도록 하는 샌드박스(Sandbox) 제도임
    - 국내의 경우 실험적통계를 「통계법 시행령 제1장 제2조의 2」<sup>5)</sup>에 의거, 법 적용 대상이 아닌 수량적 정보의 일종으로 해석
    - 실험적통계로 등록 시 국가통계로 승인되지 않더라도 민간에 공개 가능
- \* 통계법 상 국가통계로 지정되지 않은 통계는 이용자에게 공표할 수 없는 점을 감안할 때, 실험적통계의 공개 기준은 통계 작성자에게 충분한 유인책이 될 것으로 예상

[그림 1] 국가 통계 승인 관리 체계 구상안(실험적통계<시범 통계> 도입 시)



\* 출처 : 소프트웨어정책연구소(2021)

5) 통계작성기관이 통계를 원활하게 작성하기 위한 사전 준비 또는 사후 확인과정에서 통계작성대상이나 절차 또는 방법 등의 적합성 및 타당성, 오차의 발생여부 등을 확인·점검하기 위하여 시험적으로 작성하는 수량적 정보

- '22년 9월 기준 9종의 실험적통계가 공식 등록된 상황으로, 현재는 민간 빅데이터의 수치 정보를 집계하는 유형이 주류<sup>6)</sup>

[표 1] 실험적통계 등록 사례('22.09 기준)

연번	작성주체	통계명	작성목적
2021-001	통계청, SKT	통신모바일 인구이동량 통계	- 통신 모바일 데이터를 이용하여 국내 인구 이동 패턴 변화 등을 조기에 파악
2021-002	통계청	사회적현물이전을 반영한 소득 통계	- 국제지침 상 소득으로 간주하지 않는 사회적현물 이전을 소득의 일종으로 반영하여 통계 산출
2021-003	국립공원공단	모바일 기반 국립공원 이용량 통계	- 모바일 빅데이터 및 표본조사를 혼용해 국립공원 이용량 통계 작성 및 이용 활성화
2021-004	서울시, KT	서울 생활인구 현황	- 공공빅데이터와 통신데이터를 이용하여 서울의 특정지역, 특정 시점에 존재하는 모든 인구 파악
2022-001	한국은행	뉴스심리지수 <sup>7)</sup>	- 인터넷 포털사이트의 경제분야 뉴스기사에 나타난 경제심리를 지수화
2022-002 2022-003 2022-004 2022-005	통계청	사업체 매출 및 영업일, 구인, 노란우산공제, 가계지출 등	- 신용카드 이용, 구인·구직, 통신사 등 민간 빅데이터를 활용해 가계, 사업체, 일자리, 공중보건 분야의 속보성 통계 제공

\* 출처 : 통계청(2022)

- ▣ 본 보고서는 국내 실험적통계 제도 시행에 맞춰 대두된 데이터과학 활용 통계의 정의와 포괄범위를 짚어보고, 최근 통계분야에서 SI기술에 관한 품질 이슈가 거론되는 원인과 주요 현안을 분석해 보고자 함

<sup>6)</sup> 수집 데이터 출처의 확장(설문조사→민간 빅데이터)이 주로 시도

<sup>7)</sup> 감성 분석을 위해 딥 러닝을 활용

## 신규 제도와 시기술 활용 통계

### 1. 실험적통계 제도

#### ▣ (정의) 새로운 유형의 데이터를 활용하거나 새로운 방식을 적용하여 실험적으로 작성하는 통계<sup>8)</sup>

- 작성 이후 신뢰성 및 타당성의 확인·점검이 필요한 통계로서 통계법을 적용 받지 않고 외부 공개가 가능한 장점이 있는 반면, 관리 기간<sup>9)</sup> 이후 국가통계로 인정되지 않는 경우 폐기될 가능성 또한 존재
- 데이터의 안정적 확보를 전제로 활용 가능한 데이터 유형을 특정하지 않았기 때문에 정형 데이터뿐만 아니라 비정형 데이터 활용도 가능

#### ▣ (등록 기준) 통계청이 제공한 체크리스트에 기반한 검토 절차를 수행한 후 등록 여부를 결정

- 체크리스트는 국가통계로서 갖춰야 할 신뢰성과 중복성, 데이터 대표성 등과 더불어 기존 통계 작성 방법과의 다른 유무<sup>10)</sup>를 중점적으로 확인

[표 2] 실험적통계 체크리스트 중 핵심 확인요소

체크리스트 주요 항목	진단 요소
· 기존 승인통계 명칭 중 유사 또는 중복되는 통계가 없는가?	통계 명칭의 중복성
· (유사 통계가 존재할 경우) 작성주기를 단축·기존 통계를 보완하는 내용이 있거나 다른 데이터나 다른 방법론을 사용하는가?	실험적통계 활용 시 기존통계 대비 이점이 존재하는지 여부
· 공표 시 이용자 유의사항을 적절하게 제공하는가?	이용자의 실험적통계 오용 방지
· 기초자료의 커버리지에 대한 정보가 제공되는가?	대표성, 모집단 추정 가능여부
· 통계 작성을 위해 사용하는 데이터가 기존과 다른 새로운 유형인가?	데이터의 유형
· 통계작성에 적용한 방법론이 기존과 다른 새로운 방식인가?	방법론의 유형
↳ 가공을 위해 사용된 기초자료에 대한 상세한 설명이 있는가?	활용 데이터 명세의 상세성
↳ 추정 및 가공 방법에 대한 상세한 설명이 있는가?	활용 방법론 명세의 상세성
· 원천데이터를 제공하거나 데이터를 수집하는 기관은 적절한가?	데이터 수집처의 적절성

\* 출처 : 통계청(2021) 자료 참고 후 내용 발췌

8) 계정 정책정보(2022)

9) 실험적통계 등록 후 관리기간은 제도 시행 초기인 점을 감안하여 유동적일 것으로 예상되나 약 3년으로 추정

10) 새로운 데이터·방법론을 활용했는지 여부와 설명자료 유무를 판단하는 것이므로 활용한 기술 또는 방법론의 신뢰성을 검증하는 것은 아님

## 2. 실험적통계 제도의 의의

### ▣ 새로운 작성기법으로 도출된 수량적 정보를 공개하고 승인 통계로 인정받을 기회를 부여한다는 점에서 가치가 큼

- 공식적인 제도의 취지는 정책환경 변화(조사환경 악화, 빅데이터 등장)에 대응하기 위한 시의성 높은 통계·데이터 생산 촉진
- 과거부터 현재까지의 통계청의 관련 연구를 검토해볼 때 실험적통계 제도 시행은 아래와 같은 순차적인 통계 발전 가능성을 염두에 둔 것으로 파악됨
  - ① 통계 분야의 빅데이터 활용: 원천 데이터 수집 방식을 면접 조사 위주에서 행정자료, 공공 빅데이터까지 확장
  - ② 민간 빅데이터의 활용과 통계 가공: 민간 빅데이터의 활용을 허용하고 이에 기반 한 가공통계 생산을 장려
  - ③ 데이터과학을 통계 가공에 활용: 데이터의 가공을 위한 방법론으로 데이터과학(AI알고리즘 포함)의 수용을 검토
  - ④ AI기술 활용 통계(진행 중): AI기술 고유의 특성을 파악해 통계 분야 도입 전 개선해야 할 제도개선 사항 도출

### ▣ 현시점에서 등록된 실험적통계는 상기 논의의 발전 단계 중 ①~②에 집중되어 있으나, 제도 시행 초기인 만큼 ③~④에 해당하는 통계 사례도 빠른 시일 내 증가할 것으로 예상

- 데이터과학과 AI기술은 전문가의 견해에 따라 분류 범위가 불명확한 지점이 존재<sup>11)</sup>해, 두 방법론은 순차적 달성 단계가 아닌 유관 기술 분야로 보는 것이 적절
- 데이터과학과 AI기술의 기능적 달성 목표는 크게 분류(Classification)와 군집화(Clustering)로 볼 수 있기 때문에 유사 특성을 지니는 데이터들을 자동 분류하여 유용한 수치 정보를 제시하거나 예측하는 형태로 활용될 여지가 큼

<sup>11)</sup> 대표적인 사례로 규칙 기반 학습(Rule based learning) 알고리즘은 오랜시간 동안 AI기술의 핵심 아이디어로 간주되었으나, 최근 신경망 이론에 기반한 딥러닝(Deep Learning)이 각광받음에 따라 규칙 기반 학습을 데이터과학의 일종으로 분류하는 경우도 존재

### 3. 데이터과학 활용 통계의 포괄범위와 AI

#### ▣ 데이터과학은 데이터를 정제, 집계, 조작하는 데이터 준비과정에서부터 고급 데이터 분석을 수행하는 작업까지 모든 과정을 총칭하는 용어<sup>12)</sup>로, 관점에 따라 포함되는 기술의 범위가 상이

- 넓은 의미에서 데이터과학은 빅데이터 처리, 데이터에 기반한 통계적 추정, 기계학습에 기반한 AI알고리즘 등을 모두 포괄하는 개념으로 해석
- 좁은 의미에서 데이터과학은 독자적인 영역을 구축한 빅데이터 처리와 AI학습 알고리즘의 범주를 제외한 빅데이터 계량·가공 방법으로 간주

#### ▣ 국내 통계청은 2022년 보고서<sup>13)</sup>를 통해 데이터과학 활용 통계를 빅데이터를 활용하거나 AI기술을 적용하여 생산하는 통계로 조작적 정의

- 세부적으로는 빅데이터·데이터과학·AI 기술을 별도의 특징으로 분류하되, 이를 모두 총칭할 수 있는 용어로서 광의의 데이터과학 개념을 차용
- 해외 문헌 검토 결과 차후 구체화 된 적용 기술별 개선과제 도출 시 빅데이터의 활용과 AI기법의 적용을 분리 해서 볼 가능성<sup>14)</sup> 존재

[표 3] 데이터과학의 업무 응용 분야와 세부 기술 유형

응용 분야	알고리즘	비고
분류	의사결정나무, 신경망, 베이지안 모델, 유도 규칙, K-최근접 이웃 등	일반적으로 머신러닝, 딥 러닝 활용
회귀 분석	선형회귀, 로지스틱 회귀	일반 통계 모델, 딥 러닝 모두 활용
이상 탐지	거리기반, 밀도기반, 지역 특이값 요소(LOF)	'분류'와 본질적으로 동일한 목표
시계열 분석	지수평활, 자기회귀 누적 이동평균, 회귀분석	일반 통계 기법을 주로 활용
군집화	k-평균, 밀도기반 군집화	AI·일반통계기법 외 데이터과학 분야
연관성 분석	반발패턴성장 알고리즘, 선형적 알고리즘 등	AI·일반통계기법 외 데이터과학 분야

\* 출처 : 통계개발원(2022.04) 자료 인용 후 수정

12) oracle.com의 데이터 과학의 정의 일부 인용

13) 통계개발원(2022.04), "데이터과학 활용통계 승인 개선방안 연구"

14) 해외 연구에서는 통계의 빅데이터 활용과 AI기술 활용을 명확히 분리하여 바라보는 접근이 많음

#### 4. 통계의 작성에 AI기술 활용, 현실적인 문제는 없는가

▣ (시기적 적절성) 국내 차원의 실험적 통계 제도 신설, 통계 분야에 AI기술 활용을 고려하고 있는 현 상황을 감안하면 향후 공공조직이 AI기술 활용 통계의 작성을 시도해 볼 충분한 여건이 마련되어 있다고 판단됨

○ 해외 국가들의 AI기술 활용 통계 실증 사례가 축적되고 있으므로 활용 부문 및 방법론을 사전 검토하는데 점차 용이해지는 추세

\* 범국가 프로젝트로 UN유럽경제위원회(UNECE)의 HLG-MOS Machine Learning Project: WP1(파일럿 연구 부문) 가 운영 중이며 매년 국가별 실험 성과가 공유

\* 독일 연방 통계국이 발간한 「Machine Learning in Official Statistics」의 조사 결과도 참고하는데 적절한 문헌

○ 과거 연구에서 국가통계의 AI도입 유형을 두 가지로 구분<sup>15)</sup> 한 바 있으며, 통계 생산 프로세스의 현대화 부문의 경우 이미 국내 차원의 실증 사업을 수행 중

\* 과학기술정보통신부 2022년 디지털 공공서비스 혁신 프로젝트의 일환으로 선정된 "인공지능 기반 통계분류 자동화" 사업이 대표적으로, 국가통계 생산과정에 AI기술을 활용해 산업분류 코드 할당의 정확도 향상, 내검 자동화를 추진

[그림 2] 국가통계 AI도입 유형의 두 가지 관점과 기대효과



\* 출처 : 국가통계방법론 심포지움(2021.9), "AI·빅데이터 기술 활용 국가통계 생산과 당면과제" 발표자료

15) 소프트웨어정책연구소의 2020년 보고서를 통해 국가통계 AI도입 유형을 크게 「통계 생산프로세스의 현대화」와 「신규 통계 생산」으로 구분한 바 있음. 본 고에서 다루는 AI기술 활용 통계는 「신규 통계 생산」유형에 해당

### ▣ (국가통계 승인 가능성) 현시점에서 통계작성에 활용 가능한 SI기술에 대한 명세 및 제한이 없는 것은 이점으로 작용할 수 있으나, 통계작성 조직이 달성해야 할 품질 목표가 불확실해 국가통계 승인 가능성을 예측하기 어려운 상황

- 통계작성자는 국가통계 승인을 목표로 실험적통계 작성을 추진할 것이므로 작성 과정에서 어떤 사항을 준수 하는 것이 승인에 유리한지가 관심사일 것임
- 현행 실험적통계의 등록 심사는 통계작성을 위해 활용한 방법론 자체의 적절성을 판단하지 않기 때문에, 관리 기간 이후 통계 활용이 불가한 것으로 결론 시 개선의 여지 자체가 제한될 가능성
  - \* 공공조직의 통계작성은 계약상의 비용(Cost)을 포함한 공적 자원이 소비되므로 개념 증명(Proof of Concept) 형태의 투자로 접근하는 데 한계가 존재

### ▣ (통계 제도 보완 동향) SI기술 활용 통계의 다양한 유형을 아우를 수 있는 품질관리 방안과 진단체계 변화 필요성을 인지하는 단계

- 실험적통계의 국가통계 승인 여부 판단을 위해서는 SI기술을 활용해 작성된 통계의 신뢰성을 기존 품질진단 체계에 맞춰 평가할 수 있는지 검토 필요
  - \* 통계개발원 연구 결과<sup>16)</sup>에 따르면 국가통계 승인제도 개편의 일환으로 데이터과학 활용 통계 주요 심사내용 및 심사 기준, 통계작성 자료원 및 작성 방법을 체계적으로 범주화할 수 있는 대안 검토가 필요함을 시사
- 실험적통계 품질관리를 위해서는 표준화된 품질 가이드 및 지침도 필요할 것이므로 진단체계 개정 논의와 더불어 이슈화될 것으로 짐작해볼 수 있음

### ▣ (소결) SI기술 활용 통계를 작성을 추진할 정책적 유인은 충분하나, 통계에 활용하는 SI분야 방법론의 품질 기준이 불명확해 활성화 장애물로 작용

- 3장을 통해 해외 연구 사례, 통계 품질에 대한 정확한 문제를 짚어보고 자체적인 분석 결과를 제시해보고자 함

<sup>16)</sup> 통계개발원(2021.04), "데이터과학(Data Science) 기초 연구"

## III 시기술 활용 통계의 품질 이슈 분석

### 1. 시기술 활용 통계의 품질, 무엇을 고려해야 하는가

▣ 시기술 활용 통계의 품질 이슈를 명확화하기 위해 품질 검토 부문<sup>17)</sup>을 크게 두 가지 차원으로 구분하여 접근



- (수집 데이터) 통계 산출의 근간이 되는 자료원의 형태로서 면접 조사, 행정 자료, 민간 빅데이터 등 현재 고려되는 모든 종류의 정보
- (방법론) 데이터를 가공 또는 분석하기 위한 작성 기법으로 기계학습, 심화학습, 강화학습 등으로 분류되는 모든 시알고리즘

<sup>17)</sup> 기존 통계 대비 시기술 활용 통계의 품질 진단 관점의 차이를 유도하기 위함으로 공통적으로 요구되는 정책 거버넌스 관련 품질 특성(통계 중복성, 분류체계 활용, 자료 수집체계 등)는 검토범위에서 제외

## ▣ 수집 데이터로 빅데이터를 활용할 시 고려해야 할 품질 요소는 관련 연구 사례가 많아 비교적 발굴에 수월한 편

[표 4] 빅데이터 품질 진단 주요 참고 사례

국가	참고자료명	비고
한국	한국지능정보사회진흥원(2019), “데이터 품질관리 가이드”	· 총 9개 차원의 품질관리 지표 제시
	한국데이터산업진흥원(2016), “데이터 품질 인증 제도(Data Quality Certification, DQC)”	· 데이터 품질의 인증을 크게 데이터, 데이터관리, 데이터보안으로 구분하고 각각에 대한 진단지표 제시
UN	UNECE(2014), “Framework for the Quality of Big Data”	· 빅데이터의 품질 검증 단계를 투입(input), 처리(throughput), 산출(output)로 구분 후 각각의 진단지표 제시

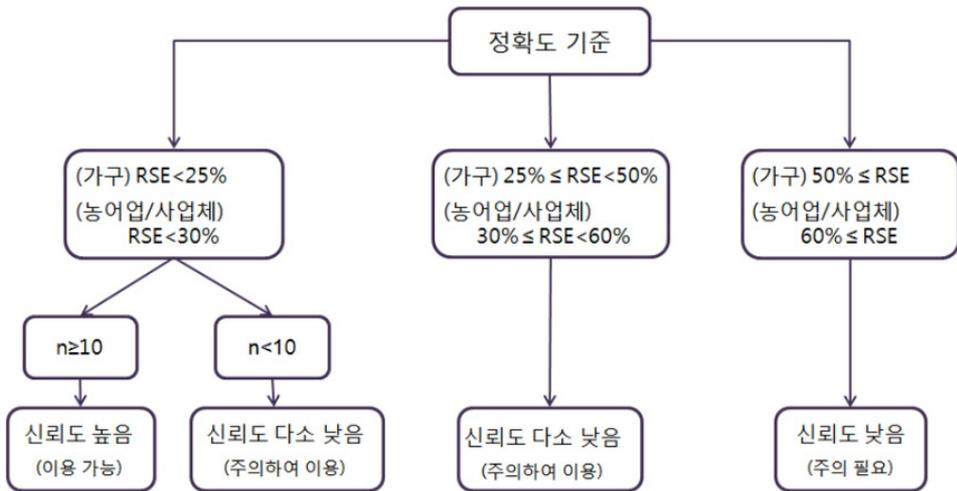
## ▣ 반면 AI알고리즘의 활용은 도입 적절성이 충분히 증명되지 않았고, 해당 방법론을 통계에 활용하는데 요구되는 품질 합의사항<sup>18)</sup>이 검토되지 않은 상황

- 기존 통계작성에 활용하는 방법론의 경우 통계 분야에서 오랜 기간 정립된 이론이므로 별도의 검증 과정이 불필요
- 실제 국내 품질관리체계 또한 방법론의 신뢰성이 아닌 투명성을 진단<sup>19)</sup>하고 있는데, 이는 잘 알려진(Well-known) 방법론의 활용이 전제되기 때문
  - \* 기존 통계 방법론의 경우 품질 합의사항이 필요한 사안에 대해 국가법령정보의 [별표] 형태로 기준이 명확히 규정되어 있음[그림 3 참고]

18) 통계청예규 제254호 [별표]의「유사중복통계 판정기준」,「표본조사 정확도 평가 기준」등이 이에 해당

19) 현행 품질관리체계는 활용한 방법론과 가공과정, 변동 사항 등에 관한 기록을 중요시함

[그림 3] 표본조사 정확도 평가기준



\* 출처 : 국가법령정보센터(2017.12)

- 즉 SI기술을 통계작성에 활용하기 위해서는 SI의 특성을 통계 품질관리체계에 내재화하고 주관적 판단 요소를 규정화하는 작업이 관건이 될 것임
  - (내재화) SI기술 고유의 특성으로 인해 발생하는 통계 작성환경의 변화를 진단하여 기존 품질관리 체계가 고려하지 못한 진단 영역을 고안
  - (규정화) 연구자의 주관에 따라 결과해석이 달라질 여지가 있는 사항을 식별해 표준화된 결과해석 기준을 정의

## 2. 국내외 관련 연구 사례 검토

※ SI기술 활용 통계의 품질을 주제로 논의 중인 국가 또는 기관은 소수로, 대다수 국가가 현시점에도 도입에 초점을 맞추어 가능한 통계 유형을 발굴하는 데 집중

※ 그러므로 본 장에서는 정확히 해당 주제와 일치하지 않더라도 참조 가능한 문헌 위주로 정리하였음을 참고

### 통계 알고리즘을 위한 품질 프레임워크(UN유럽경제위원회)<sup>20)</sup>

- (문헌 개요) UNECE, HLG-MOS 기계학습 프로젝트의 품질부문 실무 그룹에서 개발한 통계 품질진단 프레임워크
- (문헌 특징) '알고리즘'이란 용어를 시모델에 한정하지 않고 추정이나 예측을 하기 위한 절차(Process) 또는 일련의 규칙으로 규정하여 전통적 통계 및 기계학습(ML) 방법론 모두를 하나의 체계로 포괄
- (진단 체계) 5가지 차원으로 품질진단 체계를 제안

[표 5] QF4SA의 통계 알고리즘 품질진단 체계

진단 지표	주요 참고 사항
설명가능성(Explainability)	· 결과로 도출되는 산출물뿐만 아니라 분석에 활용되는 알고리즘의 기초가 되는 논리를 설명할 수 있는 것
정확성(Accuracy)	· 측정하도록 설계 된 현상을 정확하게 설명하는 수준으로 정의 · 예: 모델 불확실성에 대한 추정값이 정의될 필요
재현성(Reproducibility)	· Method reproducibility: 원래 사용한 것과 동일한 데이터 및 알고리즘을 활용하여 결과를 복제할 수 있는 것 · Inferential reproducibility: 다른 연구 설계와 분석 방법론을 선택해도 유사한 결과를 산출할 수 있는 것
적시성(Timeliness)	· 개념화에서 알고리즘 구현, 처리, 통계 생산에 이르는 결과를 도출하는데 소요되는 시간으로 정의
비용효과(Cost effectiveness)	· 해당 비용과 비교하여 그 결과가 얼마나 효과적인지로 정의 · 인프라 구축, 직원 교육 등의 고정 비용(Fixed cost)과 통계 생산 비용과 같은 지속적 비용(Ongoing cost) 등이 여기에 해당

### 공인 테스터 시테스팅 강의 계획서(국제 소프트웨어 테스팅 자격 위원회)<sup>21)</sup>

- (문헌 개요) 시분야 테스터 자격 발급을 위해 한국·독일·중국 등이 공동 제작한 커리큘럼 가이드라인으로써 시기반 시스템의 품질진단 절차 및 측정 기법 등을 정의하고 있음
- (문헌 특징) 시기반 소프트웨어 시스템의 품질 테스팅을 위한 가이드로 통계 분야 시모델에는 적용하기 어려운 척도도 다수 존재
- (진단 체계) 8가지 차원으로 시모델의 품질 특성을 다룸

20) A Quality Framework for Statistical Algorithms(QF4SA)

21) (2021) International Software Testing Qualification Board(ISTQB), AI Testing Syllabus

[표 6] 시 기반 시스템 품질 특성

품질 특성	주요 참고 사항	참고 가치
유연성 및 적응성 (Flexibility and Adaptability)	· 운영 환경의 변화 등 새로운 상황에서 해당 시스템을 운용 가능한지에 관한 척도	○
자율성 (Autonomy)	· 인간의 감독과 제어(개입)에 완전히 독립적인 상태를 의미 · 협의로서는 인간의 개입 없이 성능을 보장할 수 있는 기간	×
진화 (Evolution)	· 외적 제약에 대한 대응으로서 자체 시스템을 향상시키는 시스템의 능력을 의미(예시: 자가 학습기능)	×
편향 (Bias)	· 시스템이 제공하는 산출물과 '공정한 결과'로 간주되는 산출물 간의 거리에 대한 통계적 측정값	○
윤리 (Ethics)	· 시스템이 윤리적 방식으로 사용되는지에 대한 정성적 판단	△
부작용 및 보상해킹 (Side effects and Reward hacking)	· 특정 과업을 수행하는데 초점을 맞춰 다른 변수를 무시해 부정적 결과를 초래하거나, 모델 설계의 방향을 왜곡하는 '교묘한' 또는 '쉬운' 방법을 시모델이 채택하는 경우	×
투명성, 해석가능성 및 설명가능성 (Transparency, Interpretability and Explainability)	· 시모델을 생성하는 데 활용한 알고리즘 및 학습 데이터 확인의 수월함(투명성), 다양한 이해관계자가 시모델을 이해할 수 있는 것(해석력), 시 기반 시스템에서 특정 결과가 어떻게 산출되는지 쉽게 판단할 수 있는 것(설명력)	○
안전과 인공지능 (Safety and AI)	· 시 기반 시스템이 사람, 재산 또는 환경에 해를 끼치지 않을 것이라는 기대로 정의	△

▣ ML기반 생활비 및 식료품 소비 데이터 편집 사례연구(영국 통계청)<sup>22)</sup>

- (문헌 개요) 생활비 및 식료품 소비조사의 개인 소득 데이터 중 의심스러운 기록을 식별하는 데 ML의 활용 여지를 검토한 실험
- (문헌 특징) 해당 실험은 통계 작성 외 목적으로 시를 활용한 사례이나, 시기술의 실제 도입을 위해 검토해야 할 과제를 제시하고 있어 참고
- (주요 시사점) 연구를 통해 시기술 실제 도입을 위한 제도적 개선과제 시사

22) (2020) Office for National Statistics(ONS), "Editing of LCF(Living Cost and Food) Survey Income data with Machine Learning

[표 7] 시기술의 통계 분야 실제 도입을 위한 검토 과제

제안 사항	제안 사유
시모델 평가를 위한 기준 데이터 수집 계획의 수립	· 시모델 성능평가를 위한 수집 데이터 기준이 불명확
정확도 관련 척도에 대한 임계치 설정	· 실험에서 활용한 성능척도는 실제 도입 적절성을 판단하기 위한 임계치 기준이 없음
효과적인 모델 모니터링 및 유지관리 규약	· 현재 ML기반 통계 도입은 초기 모델 배포만 고려함 · 모델의 동작 모니터링 및 주기적 재학습은 예측결과 성능을 지속적으로 유지하는데 중요
모델 재학습에 따른 책임 지정	· 모델 성능 검토 주기뿐만 아니라, 모델 성능 임계치 및 관련 절차 구현에 대한 인적 책임이 명확히 지정되어야 함
중단 및 운영을 위한 규약	· 주기적 시모델 재학습을 위해 신규 학습 데이터 세트를 구축할 시 많은 시간 소요 예상 · ML솔루션을 활용하기 위해서는 이를 고려해 모델 중단에 대한 규약도 마련될 필요

#### ▣ ML을 활용한 에너지 균형 통계 추정 사례연구(벨기에 통계청)<sup>23)</sup>

- (문헌 개요) 에너지 생산 및 소비에 관한 국가통계를 대상으로 선정하고 결측치 처리를 시모델의 결과 값으로 자동 대체
- (주요 시사점) 랜덤 포레스트, 신경망 등 시알고리즘의 실험 결과를 집계하기 위하여 모델 생성의 랜덤 시드(seed)를 고정하는 임의의 조치를 시행<sup>24)</sup>

#### ▣ 인공지능 영역에서의 통계의 역할(Sarah Friedrich 외 10인)<sup>25)</sup>

- (문헌 개요) 시분야의 통계학적인 해석을 위해 방법론, 연구계획 및 설계, 데이터 품질 등 다방면을 다룬 논문
- (주요 시사점) 시분야 품질 평가를 위한 통계학 관점의 7가지 고려사항을 언급

23) (2020) Vlaamse Instelling voor Technologisch Onderzoek(VITO), "Early estimates of energy balance statistics using machine learning"

24) 시알고리즘의 확률적 모델 생성을 제한하는 것은 결과로 제시한 모델의 최적화를 포기했음을 의미하므로, 시모델의 품질 기준 부재를 역설적으로 보여주는 사례

25) Sarah Friedrich 외 10인(2021), "Is there a role for statistics in artificial intelligence?"

[표 8] AI기술 품질 평가를 위한 통계적 고려사항

구분	세부 내용	참고 가치
유효성 (Validity)	· AI분야의 내적·외적 타당성 고려 필요 ↳ 내적 타당성: 체계적 편향(Systemic bias) 상태를 회피하는 것 ↳ 외적 타당성: 관측된 효과 및 관계를 규모가 더 크거나 상이한 인구, 환경, 상황 등으로 이전할 수 있는 것	○
대표성 (Representativity)	· 충분히 큰 데이터가 곧 대표성을 의미하지 않으므로 관련 판단은 통계학적 접근이 필요	○
편향 (Bias)	· 신중하지 않은 데이터 수집은 결론 왜곡을 야기	○
모델 안정성 및 재현성 (Model stability and Reproducibility)	· 실험 결과의 재현성은 과학적 방법론의 본질적인 부분으로서 타자의 연구를 복제할 수 없다는 것은 심각한 결과 초래	△
데이터 품질 평가 (Assessment of Data quality)	· 모든 분석과정에서 데이터 품질 보장이 우선시되어야 함	○
인과성 및 연관성 구분 (Distinguishing between Causality and Association)	· AI는 연관성에 기반한 대량의 데이터에서 구조와 관계를 탐색할 수 있으나 이는 인과 관계를 기정사실화 하는 것과 다름을 인지하여야 함 · 연관성을 인과성으로 대체하려는 노력 필요	△
불확실성 및 해석력 평가 (Evaluating uncertainty and Interpretability)	· 불확실성의 정량화는 AI분야에서 간과되는 경우가 많으나, 신뢰성 평가를 위해서는 중요한 척도임	○

### 3. AI솔루션 신뢰성 검증 항목 기반 비교 분석

▣ (추진 방향) AI솔루션의 신뢰성 검증 항목을 기반으로 통계 도메인에 공통적으로 활용될 수 있는 AI품질 검증 사안을 선별하고 차이점을 분석해보고자 함

▣ (대상 문헌) 신뢰할 수 있는 인공지능 개발 안내서<sup>26)</sup>

- 문헌 개요: 민간 기업의 AI솔루션 개발 신뢰성 제고 차원에서 작성된 가이드라인
- 문헌 특징: AI소프트웨어 및 서비스 개발 기업을 대상으로 작성된 실무지침에 근접한 문서로서 AI기술 관점의 요구사항이 다수 포함

<sup>26)</sup> 과학기술정보통신부, 한국정보통신기술협회(2022), "2022 신뢰할 수 있는 인공지능 개발 안내서(안)"

\* 시기술 고유의 특성 및 품질 합의를 위한 사항을 식별하기 위해서는 기존 통계 기법의 활용을 전제로 개발된 진단체계 검토만으로는 한계가 존재한다고 판단

[표 9] 시활용 통계의 품질 요소 분석 결과

시솔루션 신뢰성 검증 항목		항목별 주요 내용	통계 작성 관점 검토 의견
1	인공지능 시스템에 대한 위험관리 계획 및 수행	· 시스템 구동 중 발생가능한 모델 오인식, 기능 오동작, 보안 및 개인정보 이슈와 같은 위험 요소 사전에 인식	· 공표주기가 실시간인 시기술 활용 통계를 허용할 것인가?(★) · 허용 시 시기술 활용 통계를 SW 시스템 관점에서 볼 필요(★)
2	데이터의 활용을 위한 상세 정보 제공	· 수집 데이터의 올바른 활용과 문제 발생 시 원인 추적에 위해 상세 정보 제공	· 서로 다른 국가통계가 동일 출처의 수집 데이터를 활용하는 사례가 발생할 여지가 존재하므로 수집 데이터 명세 양식을 표준화할 필요
2-2	데이터 출처 기록, 관리	· 오픈소스(Open source, OSS) 데이터셋 활용으로 인한 데이터 버전 관리	· 데이터셋 버전 정보 관리를 품질관리 체계에 포함 여부
3	데이터 강건성 확보를 위한 이상 데이터 제거	· 인공지능 모델 학습에 활용된 데이터는 이상 값, 중독 및 회피 등에 영향을 받지 않아야 함 · 점검 및 방어 기법 적용을 통한 강건성 확보	· 데이터 이상값(Outlier) 점검은 정형 및 비정형 학습용 데이터(Text, 이미지 등)를 활용하는 모든 통계 작성 과정에서 중요 → 이상 값에 대한 검토 기준은 현재도 존재
3-1	데이터 이상값 식별 및 정상·오류 여부 점검	· 데이터 노이즈 · 학습 데이터 내 편향(Bias) · 라벨링 문제(오류, 누락) 등	· 학습 데이터 내 편향성에 대한 점검 척도 별도 고려(★)
4	수집 및 가공된 학습 데이터의 편향 제거	· 데이터 수집 시, 학습을 위한 특성 선택, 데이터 라벨링 및 샘플링 시에도 편향 발생이 가능하므로 제거방안 마련 필요	· 통계작성에 활용 되는 데이터 수집 과정에서의 편향성 검증을 진단 요소로 추가 가능할 것임
4-1	데이터 수집 시, 인적·물리적 요인에 의한 편향 완화 방안을 마련	· 인적 요인으로 인한 편향 · 물리적 요인에 의한 편향	· 체크리스트 수준에서 진단 가능할 것으로 보임
4-2	학습에 사용되는 특성 (Feature)를 분석하고 선정기준을 마련했는지 여부	· 보호변수는 데이터 학습 시 반영되지 않아야 하는 특성	· 통계 분야는 데이터 비식별 여부에 매우 엄격하므로 문제 발생 여지는 적을 것으로 판단 · 보호변수를 별도 여과할 필요가 있는지 검토 필요
4-3	데이터 라벨링 편향 방지	· 지도학습은 학습 데이터 라벨링이 요구되므로 편향이 발생할 여지 · 라벨링 결과의 평가 및 작업 기준 교육 등을 통한 방지 노력	· 실효성 여부 검토 수준에서 참고 가능
4-4	데이터 분포 검증을 통한 데이터 샘플링 수행	· 증화추출법 등 편향 방지를 위한 샘플링 기법 활용	· 통계 분야에 이미 널리 적용

AI솔루션 신뢰성 검증 항목		항목별 주요 내용	통계 작성 관점 검토 의견
5	OSS 라이브러리 보안성 및 호환성 확보	<ul style="list-style-type: none"> <li>· AI모델 설계 및 개발을 위해 활용하는 OSS의 버전 관리 및 보안 위험 요소 점검</li> <li>· 주요 이슈: OSS라이선스 문제, 호환성, 보안취약점 등</li> </ul>	<ul style="list-style-type: none"> <li>· (SW시스템의 관점에서 볼 시) 통계 작성에 OSS를 사용할 가능성이 커 고려 필요</li> </ul>
6	인공지능 모델의 편향 제거 기법 적용	<ul style="list-style-type: none"> <li>· 편향완화 기법(Pre-processing)</li> <li>· 학습 중 적용(In-processing)</li> <li>· 학습 후 적용(Post-processing)</li> </ul>	<ul style="list-style-type: none"> <li>· 통계 분야 또한 활용 필요</li> </ul>
7	인공지능 모델 공격에 대한 방어 대책 수립	<ul style="list-style-type: none"> <li>· 적대적 의도를 가진 사용자에 의한 공격을 방지해야 함</li> </ul>	<ul style="list-style-type: none"> <li>· AI모델의 명세를 투명하게 공개함을 원칙으로 한다면 고려할 필요가 없는 사안</li> </ul>
8	인공지능 모델 명세 및 출력 결과에 대한 설명 제공	<ul style="list-style-type: none"> <li>· 인공지능 모델의 예측 결과에 대한 사용자 신뢰 확보를 위하여 모델 정보 및 결과도출 과정 설명이 제공되어야 함</li> </ul>	<ul style="list-style-type: none"> <li>· 설명가능성은 AI기술 활용 통계의 품질을 구성하는 주요 척도(★)</li> </ul>
9	인공지능 모델 출력에 대한 신뢰도 (Confidence value) 제공	<ul style="list-style-type: none"> <li>· 정확도(Accuracy)</li> <li>· 정밀도(Precision)</li> <li>· 재현율(Recall)</li> </ul>	<ul style="list-style-type: none"> <li>· 결과의 불확실성을 나타내는 지표로써, 통계조사의 오차와 유사</li> </ul>
9-2	신뢰도가 낮을 경우 조치방안	<ul style="list-style-type: none"> <li>· 신뢰 구간(Confidence interval)에 따른 의미 정의</li> <li>· 신뢰도가 기대에 못 미치는 경우 적절한 조치방안 및 대안 필요</li> </ul>	<ul style="list-style-type: none"> <li>· AI모델의 신뢰도 측정 결과 구간 별 의미를 정의하는 것이 품질 판단 차원에서 매우 중요(★)</li> <li>→ 품질 합의사항에 해당</li> </ul>
10	인공지능 시스템 구현 시 발생 가능한 편향 제거	<ul style="list-style-type: none"> <li>· 시스템 구현 단계에서 편향을 고려하지 않는다면 설계자 또는 개발자 배경지식에 의해 인공지능 시스템이 편향될 수 있음</li> </ul>	<ul style="list-style-type: none"> <li>· 정성적 기준을 마련하는 방식을 통해 연구자의 주관적 개입을 최소화하거나 평가 가능</li> <li>· 조사통계도 마찬가지로 연구자의 주관에 의해 설문지가 편향될 가능성은 항상 존재</li> </ul>
11	인공지능 시스템의 안전 모드 구현	<ul style="list-style-type: none"> <li>· 인공지능 시스템을 통해 생성되는 결과나 의사결정이 사회에 부정적인 영향을 미칠 수 있으므로 안전 모드를 구현할 필요</li> </ul>	<ul style="list-style-type: none"> <li>· 통계 공표 주기가 짧고 데이터가 휘발성을 가진 경우는 데이터 수집에 문제가 발생 시, 통계 안정성에 중대한 문제 발생 가능(★)</li> </ul>
11-1	공격, 성능 저하 및 사회적 이슈 등의 문제 발생 시 안전모드 적용 여부	<ul style="list-style-type: none"> <li>· 고장 안전(Fail-safe) 보장</li> <li>· 상황에 대한 안내 필요</li> </ul>	<ul style="list-style-type: none"> <li>· (SW시스템의 관점에서) 휘발성 데이터 수집과 관련한 네트워크 오류 등을 대처하기 위한 대책이 요구</li> <li>· 상황 발생 후 후속 조치는 국가통계에서 허용되지 않는 상태(State)임(★)</li> </ul>

시솔루션 신뢰성 검증 항목		항목별 주요 내용	통계 작성 관점 검토 의견
12	인공지능 시스템의 설명에 대한 사용자의 이해도 제고	· 서비스 제공자는 사용자에게 결과가 이해가능한지(Understandability), 해석가능한지(Interpretability), 설명가능한지(Explainability) 평가	· 8번 항목과 동일
13	인공지능 시스템의 추적가능성 확보	· 문제 원인 추적을 위한 시스템 로그 · 데이터 모니터링 · 인공지능 모델과 사람 간 의사결정 기여도 추적 방안	· 실시간성을 가진 통계의 경우 좌측 내용의 전체 고려가 필요 · 일반적인 시활용 통계는 수집 데이터 모니터링 수준으로 분야를 한정할 수 있음
13-2	학습 데이터 변경 이력을 주기적으로 관리	· 인공지능 모델은 사용 데이터에 따라 학습모델도 달라짐 · 학습 데이터가 변경될 경우, 학습 데이터 버전관리 및 변경 발생 원인 추적이 필요함	· 데이터 출처의 유형별 예외상황에 대한 대처 방안이 수립되어 있는지 검토 필요
13-3	학습 데이터의 업데이트 이력을 주기적으로 관리	· 인공지능 학습모델은 신규 데이터 사용 시 성능 하락의 가능성이 있어 신규 데이터를 기존과 더하여 추가 학습이 필요할 수 있음 · 이 경우 학습 데이터에 포함된 신규 데이터 비율에 따른 모델 성능 변화 추적이 가능하도록 관리가 필요	· 통계 분야에서는 사전학습 된 시 모델을 재학습 없이 활용하거나 재학습 주기를 길게 상정할 것으로 예상됨(★) · 이는 통계가 시계열 안정성을 중요시하기에 발생 가능한 문제로서 신규 데이터 추가를 통한 시모델 재학습의 위험관리기준 마련이 필요할 것임(★)

#### 4. 시기술 활용 통계의 품질 지표(안)

※ 아래 제시된 품질 지표(안)은 보고서 상의 관련 연구 문헌 검토 및 유관 지표 기반 비교분석 결과를 종합한 결과임

※ 해당 지표는 시기술 활용 통계 작성자의 관점에서 작성 전 참고할 수 있는 품질 요소로서 의의가 있으므로 실제 국가통계 관리 체계상의 품질 진단사항과는 관련이 없음을 참고

[표 10] 시기술 활용 통계의 품질 지표(안)

진단 지표		세부 내용
방법론	설명 가능성	<ul style="list-style-type: none"> <li>· (정의) 결과로 도출되는 산출물뿐만 아니라 분석에 활용되는 알고리즘의 기초가 되는 논리를 설명할 수 있는 것</li> <li>· (측정 방향) 통계 이용자에게 시모델의 세부적인 명세와 작동 원리 등을 상세히 제공하고 이해도를 높이기 위한 노력을 평가</li> </ul>
	정확성	<ul style="list-style-type: none"> <li>· (정의) 측정하도록 설계된 현상을 정확하게 설명하는 수준</li> <li>· (측정 방향) 정확도(Accuracy), 정밀도(Precision), 재현율(Recall) 등 알려진 성능 척도 활용</li> <li>· (제도적 고려) 불확실성의 정량화                             <ul style="list-style-type: none"> <li>- 시모델의 성능 지표로 활용되는 수치 각각의 구간별 의미를 부여해야만 다수 모델의 평가를 객관적인 기준상에서 수행할 수 있음</li> <li>- 성능 관리 측면에서의 인적 책임 명확화</li> </ul> </li> </ul>
	재현성	<ul style="list-style-type: none"> <li>· (정의) 동일 실험의 재현을 통해 결과를 복제할 수 있는 것</li> <li>· (측정 방향) 품질 합의사항에 따른 기준으로 평가가 가능할 것임</li> <li>· (제도적 고려) 시방법론 자체의 비결정적 특성을 전제로 Method reproducibility를 만족하기 위한 최적화 기준을 정할 필요</li> </ul>
	편중성 <sup>27)</sup>	<ul style="list-style-type: none"> <li>· (정의) 시모델 생성 과정의 학습용 데이터 참조의 편중 정도</li> <li>· (측정 방향) 시모델 편향성 측정에 활용되는 알려진 방법론 활용</li> <li>· (제도적 고려) 불확실성의 정량화                             <ul style="list-style-type: none"> <li>- 시모델의 편향성 측정에 활용되는 수치 각각의 구간별 의미를 부여해야만 통계적 활용에 적합한 시모델을 판단 가능</li> </ul> </li> </ul>
	지속 가능성	<ul style="list-style-type: none"> <li>· (정의) 시성능 유지를 위한 신규 학습용 데이터 추가에 기반한 재학습</li> <li>· (측정 방향) 품질 합의사항에 따른 기준으로 평가가 가능할 것임</li> <li>· (제도적 고려)                             <ul style="list-style-type: none"> <li>- 재학습 주기 기준 마련(성능 척도와 연계)</li> <li>- 신규 학습 데이터 세트 구축 소요 기간에 따른 통계 공표 일시 중단 주기 고려</li> <li>- 시모델 성능 평가를 위한 학습용 데이터 기준 명확화</li> </ul> </li> </ul>

27) 시알고리즘의 최적화를 위해 특정 데이터의 반영 비중을 높게 가져가는 일련의 동작 과정은 높은 성능을 보장하기 위한 고의적 여과(filtering)로도 볼 수 있어 편향이 아닌 편중으로 정의

진단 지표		세부 내용
수집 데이터	편향성	<ul style="list-style-type: none"> <li>· (정의) 데이터의 내재적·외재적 속성이 편향을 지니는 특성</li> <li>· (측정 방향) AI신뢰성 국제 기준상의 보호 변수 포함 여부, 데이터 편향성과 관련된 통계 분야의 합의 기준 등을 활용</li> <li>· (제도적 고려) 통계 작성 목적의 AI학습용 데이터 편향 기준 마련</li> </ul>
	외적 타당성	<ul style="list-style-type: none"> <li>· (정의) 관측된 효과 및 관계를 규모가 더 크거나 상이한 인구, 환경, 상황 등으로 이전할 수 있는 것</li> <li>· (측정 방향) 데이터 출처, 특성 등을 정성적으로 분석해 모집단 추정이 가능할 것인지 판단</li> <li>· (제도적 고려) 통계의 목적과 특성을 고려해 외적 타당성을 선택적 요건화 하는 방안도 검토 주제일 것임</li> </ul>
통계 운용 환경	시스템 안정성	<ul style="list-style-type: none"> <li>· (정의) 통계 데이터 수집, 가공, 공표가 이루어지는 환경의 변화 또는 장애발생에 의해 통계 공표에 악영향을 끼치는 수준</li> <li>· (측정 방향) SW시스템의 네트워크 사양 및 유사시 대처 방안 등</li> <li>· (관련 고려사항) 수집 데이터가 휘발성을 지닐 시, SW시스템의 일시적인 문제가 통계의 시계열 안정성을 위협할 가능성</li> </ul>
경제성	적시성	<ul style="list-style-type: none"> <li>· (정의) 개념화에서 알고리즘 구현, 처리, 통계 생산에 이르는 결과를 도출하는데 소요되는 시간</li> <li>· (측정 방향) 공표 주기에 따라 상이하나 통상적 기준에 맞춰 통계작성이 가능한지 검토</li> </ul>
	비용효과	<ul style="list-style-type: none"> <li>· (정의) 자본 투입 대비 산출의 효과성</li> <li>· (측정 방향) 고정 비용(fixed Cost)과 지속적 비용(ongoing Cost)의 효율 평가</li> </ul>

## IV 결과 요약

### ▣ (연구 배경) 2021년 하반기부터 본격 시행된 실험적통계 제도는 빅데이터 또는 SI기술을 활용해 작성한 통계를 대상으로 국가통계 승인 기회를 제공한다는 점에서 많은 관심을 받고 있음

- 해외 국가들을 중심으로 SI기술 활용 통계의 실증 사례가 축적되고 있고 통계 분류 자동화 분야의 경우 국내 차원에서도 실증 사업이 진행 중인 상황이므로 SI기술 활용 통계의 작성을 시도해 볼 충분한 여건이 마련된 상황
- 반면 SI기술을 통계작성에 활용하고자 할 때 요구되는 특화 된 품질 기준이 부재한 상황으로 통계작성자·통계기관 모두 SI기술의 품질 측정 기준 마련에 관한 수요가 커지는 추세

### ▣ (SI기술 활용 통계의 품질 고려사항) 일반적인 조사통계와는 달리 SI기술 활용 통계는 수집 데이터의 품질뿐만 아니라, 방법론의 신뢰성을 고려해야 하는 문제

- SI의 특성을 통계 품질 관리체계에 「내재화」 할 필요
  - SI기술 고유의 특성으로 인해 발생하는 통계작성 환경의 변화를 진단하여 기존 품질관리 체계가 고려하지 못한 진단 영역을 식별
- SI기술 활용 과정에서 발생하는 주관적 판단 요소를 「규정화」할 필요
  - 연구자의 주관에 따라 결과해석이 달라질 여지가 있는 사항을 식별해 표준화된 결과 해석 기준을 정의해야 함

### ▣ (SI기술 활용 통계 품질 지표 분석) 해외 연구 사례 검토, SI솔루션 신뢰성 검증항목 기반 비교 분석을 통해 총 10가지 유형의 품질 지표를 발굴 및 제안



## 참고 문헌

### 1. 국내문헌

- 통계청(2019), “빅데이터 활용통계의 국가통계 승인관리방안 연구”
- 소프트웨어정책연구소(2021), “국가통계 시도입 활성화 방안 연구”
- 통계개발원(2022.04), “데이터과학 활용통계 승인 개선방안 연구”
- 통계개발원(2021.04), “데이터과학(Data Science) 기초 연구”
- 과학기술정보통신부, 한국정보통신기술협회(2022), “2022 신뢰할 수 있는 인공지능 개발 안내서(안)”

### 2. 국외문헌

- UNECE(2018), “The use of machine learning in official statistics”
- UNECE(2020), “A Quality Framework for Statistical Algorithms(QF4SA)”
- AI Testing Syllabus(2021), “International Software Testing Qualification Board(ISTQB)”
- Office for National Statistics(2020), “Editing of LCF(Living Cost and Food) Survey Income data with Machine Learning”
- Vlaamse Instelling voor Technologisch Onderzoek(2020), “Early estimates of energy balance statistics using machine learning”
- Sarah Friedrich 외 10인(2021), “Is there a role for statistics in artificial intelligence?”