데이터의 품질과 인공 지능 시스템의 신뢰성

곽준호 TTA AI시험검증팀 책임연구원

1. 머리말

인공 지능 시스템은 다양한 구성 요소로 이루어져 있다. 인공 지능 시스템의 신뢰성을 고려할 때 이러한 요소를 모두 염두에 두어야 할 것이다. 먼저 인공 지능 모델이 필요할 것이며, 모델을 기반으로 기능 및 동작을 구현하는 소프트웨어 혹은 하드웨어가 있다. 또한 사용자와 상호작용이 필요한 경우, 입출력을 위한 인터페이스 역시 존재한다. 이러한 시스템을 직접적으로 구성하지는 않으나, 인공 지능 시스템 구현에 필수 불가결한 요소가 있다. 데이터이다. 그렇다면, 인공 지능 시스템의 신뢰성을 고려할 때, 데이터는 어떤 부분을 고려해야 하는가? 지난 199호 저널에서는 지도학습을 위한 데이터의 품질 관리 요구사항 동향을 소개한 바 있는데, 나아가 데이터 품질과 인공 지능 시스템의 신뢰성은 어떤 관계가 있는가?
본고에서는 이에 대한 물음을 해결할 수 있는 접근을 위하여 인공 지능 시스템의 품질을 매개체로 인공 지능 시스템의 신뢰성과 데이터의 품질을 분석해보고자 한다. 이를 위하여, 먼저 인공 지능 시스템의 신뢰성과 품질의 관계, 그리고 인공 지능 시스템의 품질과 데이터의 품질의 관계를 차례대로 논하고자 한다.

2. 인공 지능 시스템의 신뢰성과 품질

2.1 인공 지능 시스템의 품질

인공 지능 시스템은 기본적으로 소프트웨어와 함께 구현된다. 따라서 인공 지능 시스템의 품질은 전반적으로 소프트웨어 시스템의 품질 체계를 따르게 된다. 국제표준에서도 인공 지능 시스템의 품질에 대한 해석이 소프트웨어 시스템의 품질 체계의 관점에서 이루어지고 있다. 소프트웨어 시스템 품질은 ISO/IEC 25010에서 '시스템의 이해관계자가 명시적/암묵적으로 필요로 하는 수요(needs)를 만족시키는 정도'라고 정의되어 있으며, 이는 구체화된 품질 속성(Characteristics)을 통해 상세히 다루어지고 있다. 인공 지능 시스템의 품질은 ISO/IEC 25010에서 다루어지고 있는 품질 속성에 인공 지능 관점의 하위 속성(Sub-characteristics)을 추가함으로써 그 정의를 확장하고 있다. 이는 ISO/IEC 25059에서 현재 표준화 진행 중이며, 아직 논의 중이기는 하나, 인공 지능 시스템의 품질의 해석이 소프트웨어 시스템의 품질 기반 위에서 이루어지고 있는 것은 자명한 사실이다.
[그림 1]은 소프트웨어 시스템의 품질 모델을 나타낸 것이며, 품질을 정의하는 속성(녹색 음영) 및 하위 속성(음영 부분 하단)이 명시되어 있다. 이 중, 인공 지능 시스템 관점에서는 속성 중

기능 적합성, 사용성, 신뢰성(Reliability), 보안성에 대해서 추가 및 수정이 이루어지고 있다.

[그림 1] 소프트웨어 시스템 품질 모델[1]

2.2 인공 지능 시스템의 신뢰성

인공 지능 시스템의 신뢰성(Trustworthiness)은 어떻게 정의되고 있을까? 우선 국제표준을 살펴보면, ISO/IEC 24028에서는 인공 지능 시스템의 신뢰성을 '검증가능한 방식(Verifiable way)으로 이해관계자의 기대치(Expectation)를 충족시킬 수 있는 능력'이라 정의하고 있다. 이 때 고려해야 할 것은 '신뢰성'이라는 용어가 다양하게 해석될 수 있다는 것이다. 소프트웨어시스템에서는 'Reliability', 'Dependability'와 같은 용어를 모두 신뢰성이라 지칭한다. 아주 엄격하게 구분되는 개념은 아니나, 주로 소프트웨어 시스템이 내포하고 있는 결함과 문제점을 해결하고 대응하는 관점에서 정의되고 있다. 그러나 인공 지능에서는 'Trustworthiness'를 신뢰성으로 새로이 개념 정의하고 있다. 이는 인공 지능 시스템 특성상, 시스템 자체가 내포하고 있는 결함과 문제점보다는 사용되는 맥락과 사람과의 상호작용에서 기인하는 이슈가 더욱 부각되기 때문으로 해석된다.

인공 지능 시스템의 신뢰성 정의와 개념은 인공 지능의 윤리가 부각되면서 영향을 받은 것으로 보인다. 인공 지능의 윤리는 2010년대 중후반부터 본격적으로 국제사회에서 논의되기 시작했으며, 다양한 원칙과 가이드라인이 마련된 바 있다. 우리나라 역시 '인공 지능 윤리기준'이 2020년 말에 마련되어 인권보장, 프라이버시 보호, 다양성 존중, 침해금지, 공공성, 연대성, 데이터 관리, 책임성, 안전성, 투명성으로 구성된 10대 요건을 선정한 바 있다. 이러한 윤리 요건에는 연대성이나 인권 보장과 같이 기술적으로는 검증 불가능한 영역이 존재한다. 이에 대응하여 국제표준에서는 실제로 '검증가능한' 영역을 '신뢰성'으로 정의한 것이다. 또한 인공 지능 시스템의 신뢰성의 개념 정의는 아직 넘어야 할 산이 많다. 비록 ISO/IEC 24028을 통해서 하위 속성과 키워드가 정의되었으나, 아직도 국제사회뿐 아니라 국내에서도 그 개념과 정의에 대해 공감대가 충분히 이루어지지 않은 상황이다. 인공 지능 기술의 발달과 혁신 추세를 보았을 때, 구체적인 기술적 정의와 개념은 지속적으로 논의가 진행되고 있으며, 향후 변동성이 있을 것으로 예측된다.

2.3 품질과 신뢰성의 비교

그렇다면 인공 지능 시스템의 신뢰성과 품질은 무엇이 다른 것일까? 앞서 언급한 ISO/IEC 표

준에서 제시한 정의에서 그 힌트를 얻을 수 있다. ISO/IEC는 품질을 이해관계자의 '수요'라고 정의하였고, 신뢰성을 이해관계자의 '기대치'로 정의하였다. 수요는 시스템을 사용하는 자가 가치 창출을 위해 필요로 하는 특정 목적과 기능을 의미하는 데 반해, 기대치는 이러한 목적 및 기능 외에도, 사용자가 생각하기에 올바르게 동작하거나 마땅한 결과를 내놓는 것을 의미한다. 따라서 신뢰성은 품질보다 더 넓은 범위이며 더 많은 요소를 충족해야 하는 특성인 것이다.

3. 인공 지능 시스템의 품질과 데이터의 품질

3.1 데이터가 인공 지능 시스템의 품질에 미치는 영향

인공 지능 시스템의 품질과 데이터의 품질의 관계를 살펴보기 위해서는 인공 지능 시스템 구현 과정에서 데이터가 활용되는 과정과 그 영향을 살펴보아야 한다. 인공 지능 시스템을 구현하는 과정에서 인공 지능 모델은 학습이 필요하다. 학습을 하기 위한 데이터를 준비하여야 하며, 다단계에 걸쳐 데이터가 필요하다. 모델 훈련을 위한 데이터, 초기에 구현된 모델의 성능 확인을 위한 검증 데이터 및 테스팅 데이터, 마지막으로 학습이 완료된 인공 지능 모델이 실제 상황에서 판단 및 추론을 하기 위해 입력되는 데이터이다. 이들은 모두 모델이 전산적 기법을 통해 처리할 수 있게 가공된 데이터일 것이며, 모델이 수행하고자 하는 목적에 맞게 특정 현상의 특성을 적절하게 반영하고 있는 데이터일 것이다. 이러한 요소들은 결국 모델의 성능에 크게 영향을 미치게 된다.

예를 들어, 모델 훈련을 위하여 수집된 데이터가 모델 훈련에 맞게 포맷이 가공되어 있지 않거나, 데이터의 내용이 안 맞아서 현상을 왜곡할 수도 있다. 이 경우 모델은 적절한 판단 및 추론이 어렵거나, 심지어 훈련 자체가 어려울 수 있다.

이를 통해 데이터가 얼마나 잘 만들어졌는지가 인공 지능 시스템의 품질에서 언급하는 '이해관계자 수요의 만족'에 지대한 영향을 끼치는 것을 알 수 있다.

3.2 데이터의 품질

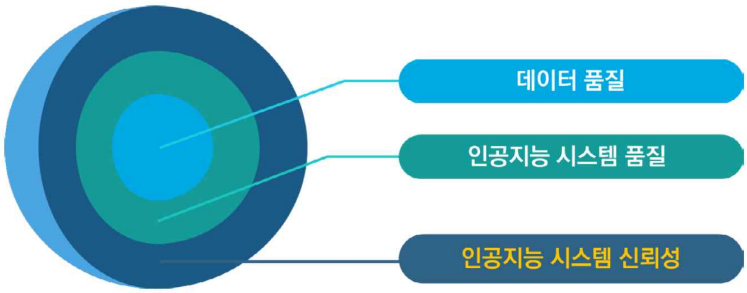
이러한 데이터의 중요성은 결국 데이터의 품질로 귀결된다. 데이터의 품질은 ISO/IEC 25012에서 '이해관계자가 제시한 요구사항을 만족하는 정도'로 정의되어 있다. 인공 지능 시스템의 품질 관점에서 시스템 이해관계자가 필요로 하는 수요에 포함될 것이며, 이 중 핵심인 인공지능모델의 정확하고 적절한 판단 및 추론을 위하여 데이터의 품질은 반드시 확보되어야 하는 대상인 것이다.

물론, 인공 지능 시스템에서 사용된 데이터는 기존 소프트웨어 시스템에서 사용되는 데이터와는 차별점이 있으며, 국내외 표준에서도 이러한 영역을 식별하여 데이터 품질을 새로이 정의하려 노력하고 있다. 지난 199호 저널에서 소개한 바있는 '지도학습을 위한 데이터 품질관리 요구사항(TTAK.KO-10.1339)'에서는 인공 지능 시스템에서 사용되는 데이터 중, 지도학습 계열의 인공 지능 모델을 활용하는 경우의 데이터 품질을 정의하고 데이터 구축 과정별로 요구사항을 제시하고 있다. 또한 ISO/IEC JTC1/SC42 인공 지능 위원회에서는 'ISO/IEC 5259: Data quality for analytics and Machine Learning' 시리즈를 통해 인공 지능 시스템에서 사용되는 데이터 품질에 대한 용어와 정의, 품질 측정 방법, 요구사항과 가이드라인, 품질 프로세스 프

레임워크와 거버넌스 표준화를 진행하는 중이다.

4. 맺음말

본고에서는 인공 지능 시스템의 신뢰성과 데이터의 품질의 관계에 대하여 알아보았다. 요약하자면, [그림 2]와 같이 데이터 품질은 인공 지능시스템의 품질로 연결되며, 인공 지능 시스템의 품질이 확보되고 나서야 인공 지능 시스템 신뢰성 확보를 논할 수 있다. 그만큼 데이터 품질 확보는 인공 지능 시스템 신뢰성 확보에 있어 기본적인 출발점이라 볼 수 있다.



[그림 2] 데이터 품질, 인공 지능 시스템 품질과 신뢰성

데이터 품질, 인공 지능 시스템 품질, 그리고 인공 지능 시스템 신뢰성에 이르기까지 필요성이 제기되고 활발하게 체계화 및 표준화가 이루어지고 있으나, 아직까지도 많은 부분들이 정립되지 않은 상태이다. 또한 국내외적으로도 이들의 관계와 명확한 범위에 대해서 혼동을 하는 경우가 적지 않다.

이러한 기술적 상황에도 불구하고 실생활에서는 다양한 인공 지능 시스템을 기반으로 제공되는 서비스와 제품이 확산되고, 기술 혁신이 일어나면서 산업 현장 적용 역시 활발하게 이루어지고 있다. 이에 따라 인공 지능 시스템을 구현하는 산업계 및 연구계는 품질 및 신뢰성 확보를 위해 많은 고민을 하고 있다. 앞으로도 인공 지능기술로 인한 사회적 이슈와 문제점은 우리가 그동안 인식하지 못했던 형태로도 발생할 것이며, 이는 데이터의 품질을 비롯하여 인공 지능 시스템의 품질 및 신뢰성에 대한 지속적인 개념 정의와 표준화가 필요함을 의미한다.

※ 본 연구는 '인공 지능 학습용 데이터 구축'과제와 '인공 지능 신뢰성 기반조성' 과제의 일환으로 수행되었다.

[참고문헌]

[1] ISO/IEC 25010: 2011 Systems and software engineering-Systems and software Quality Requirements and
Evaluation(SQuaRE) - System and software quality models
[2] TTAK.KO-10.1339. 지도학습을 위한 데이터 품질관리 요구사항

※ 출처: TTA 저널 제201호