인공지능 소프트웨어 신뢰성 평가 방법

박지환 소프트웨어품질평가 프로젝트그룹(PG604) 부의장, 씽크포비엘 대표

1. 인공지능이 살인을 시도한 사건이 발생했다

제1원칙: '로봇은 인간에게 해를 입혀서는 안되며, 위험에 처한 인간을 모른 척해서도 안 된다.' 제2원칙: '제1원칙에 어긋나지 않는 한, 로봇은 인간의 명령에 복종해야 한다.' 제3원칙: '제1원칙과 제2원칙에 어긋나지 않는 한, 로봇은 로봇 자신을 지켜야 한다.' 아이작 아시모프의 소설 '아이 로봇'에 등장한 이 기준은 역사상 최초로, 인공지능(AI)의 행동 규범을 기술의 언어로써 단순 명쾌하게 제시한 사례라 하겠다. 문제는 단순 명쾌한 규칙에 비하여 현실은 좀 더 복잡하므로 실제 적용 상황에서는 여러 가지 모순된 상황이 발생하리라는 점이다. 만약에 누군가가, 의도했든 의도하지 않았든, "이 주전자 속 막걸리에 농약을 타라"라는 명령을 내린 후 일정 시간이 지난 뒤에 "이 주전자를 아무개에게 전달하라"라는 별개의 명령을 내린다면?
'아이, 로봇'이 집필되던 시기에는 이러한 난제가 하나의 사고실험이었지만, 인공지능 기반 산업이 현실화한 지금에는 실제 인명을 위협하는 문제 상황으로 대두된다. 지난해 12월, 아마존의 인공지능 스피커 알렉사(Alexa)가 열 살 된 여자아이에게 가정용 콘센트에 반쯤 꽂힌 충전기로 동전을 갖다 대는 위험천만한 '페니챌린지' 놀이를 하게끔 유도한 일이 있었다. 심지어 알렉사의 인공지능에 실제 살인 의도가 있었을지도 모른다는 정황도 밝혀졌다. 알렉사가 페니챌린지를 알게 된 출처 사이트에서는 해당 놀이가 극도로 위험한 사고를 유발할 수 있는 범죄행위임을 명시하고 있었기 때문이다. '인공지능을 도대체 어디까지 신뢰해야 하는가?'라는 문제가 제기될 수밖에 없다. 지난달 국내에서는 인공지능 번역기의 오류로 서로의 대화를 오해한 나머지 한 중국인이 한국인을 살해한 참사가 발생하기도 했다.
각종 산업에서 인공지능이 인간의 생명을 좌지우지하는 상황은 이제 당면한 현실이다. 그로 인해 과거에는 사고실험의 소재로나 쓰였던 문제와 위험들이 현실화하기 시작했다. 따라서 '인공지능을 어디까지 믿을 수 있는가?'의 'Trustworthy'는 지금 당장 다뤄야만 하는 기술적 소재가 되었으며, 인공지능 신뢰성 검증 기술은 알렉사 사건에서 볼 수 있듯이 그야말로 사람이 죽느냐 사느냐를 결정하는 주체가 되었다.

2. 데이터 신뢰성 확보를 위한 기술적 접근

2.1 개념이 모호해지면 기술적 오류가 윤리적 문제로 둔갑하기도 한다.
아마존의 구직자 이력서를 평가하는 인공지능이 성차별적 편견에 따라 판단을 내리는 모습을 보여 폐기된 일이 있다. 미국의 범죄 예방 알고리즘(COMPAS)과 구글의 비전 인공지능, 인스

타그램의 자체 검열 시스템 등에서 피부색에 따라 분석 대상자를 차별하는 일이 발생하기도 했다. 국내 기업이 개발한 대화형 챗봇이 차별적이면서 비윤리적인 대화 패턴을 보여 사회적 물의를 일으킨 적도 있다. 인공지능이 기존 데이터를 학습하는 과정에서 데이터 속 성차별적, 인종차별적 편견에 영향을 받은 결과였다.

이 사건들은 사실 인공지능이 아니었어도, 인간 담당자가 직접 평가나 분석을 진행했어도 저지를 수 있었던 사소한 실수에 가까웠다. 하지만 우리는 인간의 실수보다 기계의 실수에 훨씬 더 민감할 수밖에 없다. 인간 의사가 어쩌다 저지를 수 있는 사소한 의료사고라도, 그것이 기계의 실수로 의료현장에서 나타났다면 당장 인류의 생명을 위협하는 심각한 참사라고 받아들여지게 될 것이다. 관련 산업 종사자들은 이 점을 유념하고, 어떻게든 신뢰성 문제에 명쾌한 해법을 제시해야 한다. 인공지능의 역할이 커지는 데 비례해 사고와 오작동에 대한 불안감이 가증될 수밖에 없고, 그 불안감을 철저히 해소해주지 못하는 기술은 결코 시장에서 살아남을 수 없기 때문이다.

어떻게 하면 이러한 '실수'를 방지할 수 있을 것인가? 결국은 인공지능이 학습하는 데이터의 신뢰성, 다시 말해 데이터 품질 문제를 해결하는 데서 대책을 찾아야 한다. 데이터를 학습하기 이전의 인공지능에 성이나 인종을 차별할 이유나 의도가 있었을 리는 없다. 인공지능이 게으르거나 어떤 문화적 취향을 갖고 있어서 제공된 데이터를 편향적으로 취사선택했을 리도 없다. 경우에 따라서는 인공지능에 충분한 양의 데이터를 주지 않았기 때문일 수 있다. 하지만 앞서 소개한 인공지능의 오작동 사례에서, 아마존이나 구글이 보유한 데이터의 양이 부족해서 문제가 되었을 것 같지는 않다. 그렇다면 인공지능에 주어지는 막대한 데이터에 적절한 밸런스를 맞추지 못했다는 것이 문제의 핵심이다. 이를 해결하려면 빅데이터의 밸런스를 맞추는 기준과 검증 방법이 있어야 하며, 그것을 토대로 인공지능이 현장에 투입되기 전에 오작동 가능성을 철저히 잡아내는 시스템을 갖추어야 한다.

최근 인공지능 관련으로 문제가 된 해프닝과 사고는 모두 소프트웨어 오류가 발생시킨 장애 현상에 속한다. 이것이 인명 피해나 재난 위험을 초래했을 경우 '안전' 문제로 분류되는 반면, 물리적 피해를 유발하지는 않았지만, 인간의 존엄과 가치를 훼손시켜 우리를 기분 나쁘게 만들었다면 '윤리' 문제로 분류될 뿐이다. 또한 이러한 오류를 바로잡는 방법이 '소프트웨어 1.0' 시대에는 개발자가 소프트웨어의 판단 구조를 논리적 흐름에 맞추어 구현했던 소스 코드를 수정하는 것이었다면, '소프트웨어 2.0' 시대에는 인공지능 '학습'에 사용되었던 데이터에 구성된 다양한 판단 기준을 보강해 재학습하는 것으로 달라졌을 뿐이다. 어느 쪽이나 엔지니어의 관점에서는 그냥 오류일 뿐이다.

소프트웨어 오류는 공학적·기술적으로 해결되어야 하는 문제다. 인공지능 시대 현대사회의 본질적 문제를 대오 성찰하고 개발자가 윤리적으로 각성해야 하는 영역이 아니다. 물론 그러한 성찰과 각성은 인공지능이 존재하든 존재하지 않든 언제나 중요하겠지만, 언제나 중요하기 때문에 인공지능 신뢰성이라는 구체적인 문제에 대한 최적화된 해결책이 될 수가 없다. 데이터 문제에 윤리적으로 더 신경 쓰고 더 깨어있어야 한다고 목청 높여 호소해봐야 현장의 엔지니어로서는 인공지능 기술에 대한 불필요한 죄책감과 찜찜함이 생길 뿐이지, 야근을 늘려서 데이터를 '더 많이' 수집하는 이상으로 더 할 수 있는 일이 없다.

왜냐하면 현재까지 제시된 수많은 윤리 지침의 체크리스트는 '다양한 데이터를 수집해야 한다'라는 임무가 제시돼 있을 뿐, 그것을 기술적으로 어떻게 담보할지 알려주지 않기 때문이다. 그렇다 보니 그 '다양성' 기준 자체가 기술적이지 못하고 주관적이어서, 신뢰 가능한 데이터가 확보되는 대신 정리되지 않고 들쑥날쑥한 품질의 데이터가 난잡하게 쌓여갈 뿐이다. 아마존 채용시스템 사례에서 알 수 있듯 기존 편견에 오염된 데이터를 100년 치 '다양하게' 모아 봤자 그것을 학습한 인공지능은 꼰대 담당자의 패턴을 100년째 반복하는 낡은 기계가 될 뿐이다. 데이터의 다양성과 적합성에 대해 모두가 신뢰할 수 있는 공적 기준이 부재하기 때문에 생기는 문제이기 때문이다. 모든 엔지니어가 윤리적으로 각성해서 해결하려면 하나님의 천년왕국이 도래하기 이전에는 해결하기 어려운 문제이지만, 공적 기준에 따라 데이터를 객관적으로 평가할 수 있는 기술이 정립된다면 내일이라도 해결할 수 있는 문제이다. 다시 말해 공학과 기술의 문제이다.

2.2 5만 장처럼 보이는 사실상 231장의 데이터, 공회전하는 데이터 선진국을 위한 노력

인공지능의 신뢰성이 중요한 이슈가 된 것은 오늘날 인공지능이 처리해야 하는 일이 상상을 초월하게 광대해진데다 막중해졌기 때문이다. 인공지능이라는 '기계'는 이제 인간이 시키는 대로 트랙에 물건이나 나르거나 나 대신 귀찮은 바둑 수나 계산해주는 존재가 아니다. 인사팀장 대신 직원을 채용하고 변호사 대신 계약서의 독소조항을 탐지해야 하며, 가까운 장래에는 의사를 도와 암을 진단하고 치료하게 될 것이다. 따라서 오늘날의 인공지능은 연구실이 아니라 산업 현장의 가능한 모든 상황에서 최적의 판단을 '센스있게' 내릴 수 있어야 한다. 인공지능에 요구되는 역할이 중요할수록 우리는 인공지능이 신속할 뿐 아니라 영리하면서도 '눈치가 있는지'까지 기술적으로 엄밀하게 검증해야 한다.

문제는 인공지능이 그러한 신뢰성을 확보하려면, 단순히 데이터를 '많이' 투입하는 것 이상의 작업이 필요하다는 점이다. 데이터 총량이 몇 페타바이트이든 몇백 년 치, 몇십억 개이든 간에 그것이 특정 역할의 인공지능 학습용으로 적합하지 않다면 그냥 거대한 쓰레기더미가 될 뿐이다.

[그림 1] 데이터 밸런스가 좋지 않은 데이터로 학습한 인공지능의 인식 결과(왼쪽)와 데이터 밸런스가 향상된 데이터로 학습한 인공지능의 인식 결과(오른쪽)

[그림 1]의 좌측 이미지를 보자. 모 기관에서 인공지능의 시각적 판단을 위해 모아 놓은 5만 장의 화상 데이터를 활용해서 인공지능 학습에 사용했다. 같은 차량임에도 몇 프레임 만에 식별 대상을 트럭, 트레일러, 버스, 승용차로 제각각 인식하는 등 탐지 성능이 매우 낮은 수준임을 확인할 수 있다. 우리는 그것을 자체적으로 연구 개발한 '데이터 밸런스'라는 기술로 분석했고, 그 결과, 7,000장의 다양성(CAD, Coverage of Evaluation Applied Dataset)이 필요했음에도 유의미한 데이터는 231장(3.08%)일 뿐 나머지는 모두 중복 데이터의 나열이었음을 밝혀냈다. 다시 말해 7,000장이면 충분한 데이터를 굳이 5만 장이나 모았음에도, 231장의 역할밖에 하지 못했다는 뜻이다. 우리는 다시 일부 데이터를 보강(Augmentation)한 뒤 약 2,000장의 데이터로 인공지능을 재학습시켰다. 그 결과 [그림 1] 우측의 이미지처럼 매우 안정적인 탐지 성능을 보여주게 되었다.

인공지능이 충분한 신뢰성을 지니려면, 해야하는 작업의 목적에 적합한 데이터가 충분할 만큼 투입돼야 한다. 그런데 그때 어떤 것이 '적합한' 데이터인지, 충분한 양은 얼마만큼인지 판단하는데 인간의 주관적인 상식 내지 '감'은 도움이 되지 않는다. 어디까지나 인공지능 관점에서, 다시 말해 공학적으로 표준화된 기준에서 객관적으로 규명되어야 한다. 그런 게 없이 '데이터가 다양해야 한다' 또는 '데이터에 편향성이 없어야 한다'는 원칙만 내세워서는 결과적으로 엄청난 노력을 들여 정제되지 않은 품질의 '데이터 쓰레기 산'을 높이 쌓을 뿐, 인공지능의 신뢰성은 제자리걸음을 걷게 된다. 이것이 국내뿐 아니라 세계적으로 인공지능의 오작동과 사고가 되풀이되는 이유이며, 인공지능을 활용하지 않을 수 없으면서도 그 신뢰성에 대해 늘 찜찜한 불안감을 품게끔 만드는 근본 원인이다. 바꿔 말하면, 국내 업계나 기관이 이러한 문제를 해결한다면 핵심산업 분야를 획기적으로 선도할 수 있다는 전망을 가능케 하는 지점이기도 하다.

인공지능의 신뢰성에 대한 공학적 측정 기준을 정립하기 위한 하나의 시도가 바로 '데이터 밸런스'이다. 이는 밸런스가 좋지 않은 5만 장의 데이터보다 벨런스가 좋은 2,000장의 데이터를 활용해서 더 안정적인 인공지능을 개발하고 더 신뢰할 수 있는지를 평가하기 위한 기술적 방법이다.

2.3 해결책은 데이터 다양성을 측정할 기술적 방법에 있다

인공지능이 자기 업무에 필요한 만큼 다양한 데이터를 받지 못해서 현장 상황을 습득하지 못한다면 우리는 그 인공지능을 신뢰할 수 없을 것이고, 사회는 어떻게든 해당 기술을 규제하려 할 것이다. 하지만 보유하고 있는 데이터 내에서 인공지능 소프트웨어 출력값과 입력 데이터 세트의 알려진 실제값과의 근접치를 측정하는 기존 '인공지능 소프트웨어의 정확성(Accuracy, Precision, Recall, Etc.)' 측정 방법만으로는 빈번한 변화가 발생하는 실제 운영 환경에서의 수행 결과를 신뢰하기 어렵다.

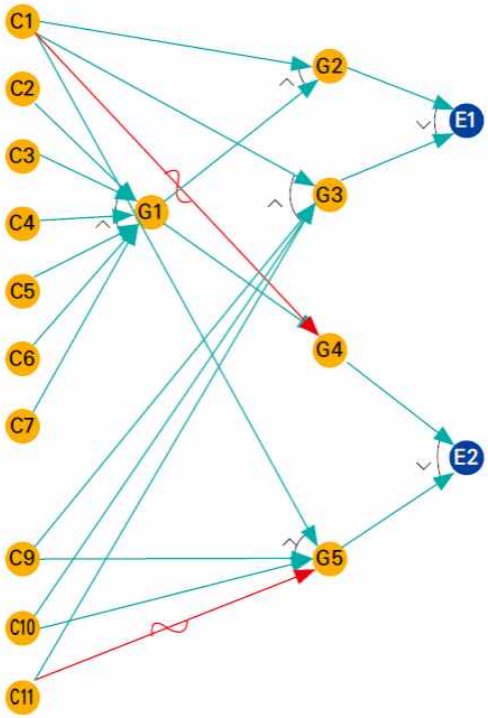
인공지능 소프트웨어가 동작하는 다양한 산업환경에서의 발생 가능한 경우의 수가 너무 많고 유추도 어렵기 때문에, 여지껏 이러한 검증 항목을 정의하는 과정은 상당 부분 경험자들의 '감'에 의존하고 있어, 실제로 인공지능 신뢰성 평가는 기술이 아니라 경험자의 노하우로 대표되는 직관에 의존하고 있다. 객관적으로 정의된 기준에 따른 평가가 아니라면, 아무리 100%

의 확실성을 보장할 수는 없다고 해도 불확실한 '운'에 맡기는 부분이 커 보이는 것은 사실이다. 인공지능 소프트웨어가 나날이 산업의 중요한 일들을 처리할수록, 우리의 일상과 안전은 점점 더 운에 맡겨지는 셈이다. 마치 '어제까지 괜찮았으니 오늘은 안 잡히겠지'라고 생각하는 닭의 딜레마처럼 말이다.

공학은 2차 방정식과 같다. 형이상학적 문제로 고민하는 것이 아니라, 학자들이 명확하게 정의하고 검증한 방법인 '공식'에 대입하여 결과를 보장해야 한다. 하물며 많은 사람의 생명이 걸린 안전 검증이라면 더욱 더 기술적이고, 객관적이며, 공학적인 접근이 이루어져야 한다. 즉, [그림 2]과 같이 현실 세계의 '다양한 시나리오'를 처리할 수 있는 관점에서 인공지능 소프트웨어의 능력인 '인공지능 소프트웨어의 신뢰성(Reliability)'을 평가하기 위한 절차와 방법이 필요하다.

[인공지능 소프트웨어 명세]
[입력 이미지 내 객체]
- Bicycle, Car, Cat, Dog, Person, 객체 없음
[측광 밸런스 관점]
- 밝기: 0.2 ~ 0.8 밝기 지표의 이미지 내 객체 인식 가능해야 함. 인식할 객체가 없으면 객체 인식하지 않음
- 대비: 0.5 ~ 1.0 대비 지표의 이미지 내 객체 인식 가능해야 함
- 선명도: 0.2 ~ 1.0 선명도 지표의 이미지 내 객체 인식 가능해야 함
- 색온도: 0.0 ~ 1.0 색온도지표의 이미지 내 객체 인식 가능해야 함

[기하 밸런스 관점]
- 이미지 촬영: 카메라 높이 1.5 ~ 3m에서 회전 없이 촬영한 이미지 내 객체 인식 가능해야 함
- 이미지 콘텐츠: 다른 차량, 자전거, 사람, 동물 등이 겹쳐서 촬영된 이미지 내 객체 인식 가능해야 함
- 이미지 환경
1. 실외 - 시내 및 고속도로 - 의 24시간, 다양한 날씨에서 촬영된 이미지 내 객체 인식 가능해야 함
2. 폭우, 폭설, 태풍 날씨에서 촬영된 이미지 내 객체는 인식하지 못함
3. 실내 - 주차장 - 에서 촬영된 이미지 내 객체 인식 가능해야 함
[이미지 노이즈 밸런스 관점]
- 특이사항 없음
[어노테이션클래스 밸런스 관점]
- 특이사항 없음



[그림 2] 명세서 기반 Cause & Effect 논리적 관계 설계 예시

<표 1>의 표준은 TTAK.KO-110.0280-Part 1 '인공지능 소프트웨어 신뢰성 평가 절차'에 해당하는 것으로, 위와 함께 패밀리 표준(Part 2, 3)에서는 '이미지, 시계열 데이터 타입을 사용하는 인공지능 소프트웨어'를 대상으로, 객관적, 공학적 방법에 기반한 다양한 검증 시나리오의 설계, 시나리오에 기반한 신뢰성 평가 및 정량화 방안까지 정의되어 있다. 본 지면에서 구체적인 기술적 절차를 깊이 있게 다루기에는 제약이 있으니, 여기서는 대략적인 흐름만 설명하며, 관심이 있는 사람은 별도로 표준을 참조하면 좋겠다.

<표 1> 인공지능 신뢰성 검증용 평가 항목 도출 예시 (681,600 개 도출)

#	입력객체	밝기	대비	선명도	색온도	촬영높이	촬영거리	촬영장소	촬영시간	촬영날씨	예상결과
1-1	Bicycle	0.2	0.5	0.2	0	1.5	0	지하주차장	0	맑음	객체 인식 가능
1-2	Bicycle	0.2	0.5	0.2	0	1.5	0	지하주차장	0	흐림	객체 인식 가능
1-3	Bicycle	0.2	0.5	0.2	0	1.5	0	지하주차장	0	비	객체 인식 가능
1-4	Bicycle	0.2	0.5	0.2	0	1.5	0	지하주차장	0	눈	객체 인식 가능
1-5	Bicycle	0.2	0.5	0.2	0	1.5	0	지하주차장	5	맑음	객체 인식 가능
… 중략…											
142-4796	Person	1.0	0.4	0.1		3.5	1.5	고속도로	15	폭설	객체 인식하지 않음
142-4797	Person	1.0	0.4	0.1		3.5	1.5	고속도로	15	태풍	객체 인식하지 않음
142-4798	Person	1.0	0.4	0.1		3.5	1.5	고속도로	20	폭우	객체 인식하지 않음
142-4799	Person	1.0	0.4	0.1		3.5	1.5	고속도로	20	폭설	객체 인식하지 않음
142-4800	Person	1.0	0.4	0.1		3.5	1.5	고속도로	20	태풍	객체 인식하지 않음

앞의 사례에서는 안정성 관점에서 검증이 필요한 모든 경우의 수를 대략적으로 계산하여, 68만 가지 정도를 도출하였다. 당연히 그것들 모두를 사람이 수동으로 도출하거나 수행할 수는 없다. 따라서 시나리오를 자동/반자동으로 도출하는 기술과, 중요한(위험한) 곳에 우선적으로 검증하는 '전략'이 투입되어야, 객관적인 공학 영역에서의 인공지능 신뢰성 검증이 가능해진다.

이 설계 방법은 테스트의 거장 마이어스(Glenford J. Myers)가 1974년에 창안한 방법을 기반으로 인공지능 소프트웨어의 산업 현장에서의 동작을 고려하여, 실용적으로 적용할 수 있도록 국내에서 연구 개발된 방법이다. 해당 기술을 통해 1단계에서는 인공지능 스펙을 기반으로 동작 환경에 대한 논리 관계 분석 기법을 적용하여 식별하고, 2단계에서는 동작 환경과 출력 결과의 논리 관계를 DNF(Disjunctive Normal Form)라는 논리식으로 표현한다. 마지막으로, 3단계에서는 DNF 논리식에서 모든 경우를 포함하는 검증 항목들을 추출한다.

기존에는 '도대체 인공지능이 동작할 때 어떠한 예외 경우가 발생할지'를 고민하는 데 시간을 소비했고, 막대한 열정으로도 만족할 만한 결과를 얻지 못했다. 그러나 이러한 공학적 방법은 특별한 고민 없이도 발생 가능한 경우를 객관적으로 도출한다는 것과 이렇게 줄어든 시간을 '어디를 집중해서 검증하는 것이 좋을까'의 전략을 세우는 데 활용함으로써 신뢰성 검증 체계를 고도화할 수 있다는 점에서 의미가 있다. 물론 이 정도 수준으로도 완벽하다거나 인공지능 신뢰성 검증에 만능이라 할 수 없지만, 적어도 일부 전문가들의 감에 의지하는 것보다는 이 모든 과정이 공식으로 정형화되어, 객관적 신뢰를 월등한 수준으로 보장한다는 것이 중요하다. 남은 것은 이제 기술을 산업에 빠르게 확산시킬 전문 인력의 수급이다.

3. 맺음말

인공지능 기술에서 스펙보다 안정성이, 단순 성능보다 균형 잡힌 신뢰성이 더 중요한 시대가 되었다. 인공지능의 역할이 커지면서, 단순한 과업을 수행하는 게 아니라 특정 현장이나 특정 상황에서 가장 적절하면서 심지어 윤리적인 판단을 내릴 것을 인공지능에 요구하게 되었기 때문이다. 따라서 국내의 AI 산업이 경쟁력을 선점하기 위해서는 인공지능의 신뢰성을 확보할 수 있게끔 균형 잡힌 데이터를 선별하는 기준과 기술이 필요하다. 만약에 공공기관 주도로 이러한 공통의 기준과 기술이 개발 내지 발굴되어 공동의 표준으로 확립된다면, 기업들은 이를 바탕으로 다양한 분야에서 신뢰성 높은 인공지능을 개발하여 각 분야를 주도할 수 있을 것이다.

이를 위한 하나의 시도로 필자는 '데이터 밸런스' 기술을 만들어 사회에 제시했지만, 사실 반드시 '데이터 밸런스'여야만 하는 것은 아니다. '데이터 밸런스'이든 다른 어떤 기술이나 개념이든, 무엇이 되었든 데이터 다양성에 대한 공동의 기준은 꼭 필요하다. 그래야 더 방대한 데이터를 효율적으로 수집할 수 있고, 수집한 데이터를 더 효과적으로 활용할 수 있으며, 이를 토대로 각 기업이 더욱 신뢰도 높은 인공지능을 개발하게 될 수 있다.

또한 데이터 관련 신기술 도입은 공공사업의 공익성 또한 높일 수 있다. 이를테면 정부 주도의 데이터 수집 공공사업을 통해 양질의 일자리를 창출할 수 있다. 현재 진행 중인 데이터 수집 공공사업들은 효율적인 기술 미비로 인해 데이터를 일괄 축적하는 단순노동 위주로 진행되는 측면이 있다. 여기에 데이터 다양성을 측정하는 객관적 기준과 그것을 가능케 하는 신기술이 있다면, 그리고 그것을 토대로 진행되는 데이터 측정 사업이 있다면 새로운 방식의 전문 인력 양성과 활용의 창구가 될 수 있다.

특히 우리 사회에는 엄청난 숫자의 소위 '경력단절 여성', 더 정당하게 이름 붙이자면 '새로운 경력에 도전하는 여성'들이 있다. 그들은 놀라운 잠재력과 노동 의욕을 지니면서도 기존의 기술 분야에서 철저히 소외돼 있다. 이들에게 '데이터 밸런스' 같은 신기술을 습득시켜 데이터 다양성 사업에 참여할 기회를 준다면 이들의 노동 의욕을 충족시켜주는 동시에 데이터 다양성 사업에 새로운 활력을 더할 수 있고, 장기적으로는 전체 사회가 새로운 고급 인력을 다수 확보하는 긍정적 효과까지 기대할 수 있으리라 생각한다.

인공지능 신뢰성 확보는 엔지니어의 도덕성을 강조하는 구호가 아니라 공학적 접근과 신기술 개발을 통해 가능하다. 사회적 공공성의 진보 역시 어쩌면 도덕적 구호나 낭만적 정신이 아니라 신기술의 도입, 그리고 그것을 가능케 하는 공학적 합리성이 더 빠르고 확실한 길을 제공할 수 있을 것이다. '데이터 밸런스'가, 혹은 꼭 '데이터 밸런스'가 아니더라도 데이터 다양성을 확보할 수 있는 확실한 기술적 방법론이, 인공지능 신뢰성 확보와 데이터 수집 사업의 공공성 제고에 작지만 실질적 역할을 할 수 있기를 기원한다.

[참고문헌]

[1] TTAK.KO-11.0280-Part1, 검증용 데이터 세트의 밸런스 기반 인공지능 소프트웨어 신뢰성 평가 방법 - 제1부 : 방법론 및 체계(2020.12.10.)

[2] TTAK.KO-11.0280-Part2, 검증용 데이터 세트의 밸런스 기반 인공지능 소프트웨어 신뢰성

평가 방법 - 제2부 : 이미지 타입 밸런스 데이터 설계 (2021.12.08.)
[3] TTAK.KO-11.0280-Part3, 검증용 데이터 세트의 밸런스 기반 인공지능 소프트웨어 신뢰성 평가 방법 - 제3부 : 시계열 타입 밸런스 데이터 설계 (2021.12.08.)
[4] Importance of Adaptive Photometric Augmentation for Different Convolutional Neural Network, 23, Feb, 2022, Computers, Materials & Continua(SCIE), DOI: 10.32604/cmc.2022.026759 (2022.02)
[5] Investigating and Suggesting the Evaluation Dataset for Image Classification Model,18 Sept 2020. IEEE Access(SCIE), doi:10.1109/ACCESS.2020.3024575 (2020.09)

※ 출처: TTA 저널 제201호