

AI 산업 발전을 위한 데이터 법제의 주요 쟁점과 개선방향

— 자율주행 정밀지도 및 학습데이터 활용을 중심으로

I. AI 산업 발전을 위한 데이터 법제도의

중요성과 우리나라의 규제 현황

1. AI 기술과 데이터의 중요성
2. 데이터 활용에 관한 우리나라의 규제 현황

II. 자율주행을 위한 정밀지도 관련 규제 개선 필요성

III. AI 학습데이터 활용에 관한 규제 개선 필요성

1. 가명정보와 AI 학습데이터 활용
2. 가명처리를 통한 학습데이터 활용의 문제점
3. 가명처리 없는 AI 학습데이터 활용의 필요성
4. 가명처리 없는 학습데이터 활용에 있어 고려사항

IV. 결론



고환경
법무법인(유) 광장
변호사



이일신
법무법인(유) 광장
변호사



이상후
법무법인(유) 광장
변호사

I. AI 산업 발전을 위한 데이터 법제도의 중요성과 우리나라의 규제 현황

1. AI 기술과 데이터의 중요성

4차 산업혁명 시대는 사물인터넷(Internet of Things; IoT), 클라우드컴퓨팅(Cloud Computing), 빅데이터(Big Data), 인공지능(Artificial Intelligence; AI) 기술 등과 같은 파괴적 혁신 기술이 선도하고 있다. 그 중에서도 AI 기술은 2000년대 이후 컴퓨팅 파워의 성장, 알고리즘의 등장, 스마트폰의 보급 및 네트워크 발전에 따른 데이터 축적으로 급격히 진보하여 인간의 지적 기능을 일부 대체할 수 있는 수준으로 발전할 것으로 예상되고 있다.¹⁾ 이와 같은 AI 기술 발전으로 견인되는 4차 산업혁명의 문명사적 변화는 단순한 기술적 차원을 넘어 인문사회 등 국가와 사회 전반에 혁명적인 영향을 미칠 것으로 예상되며, 최근 코로나19로 인해 산업 환경이 비대면 환경으로 변화되고 있으므로 그 속도는 더욱 빨라질 것이다.²⁾

4차 산업혁명, 특히 AI 기술 발전에 대해 논의할 때, 데이터는 원유 또는 핵심 자원으로 이야기된다. AI 기술은 인간의 지적 기능을 일부 대행하거나 보완하는 것이 기술적 목표라는 점에서 인문학적 이해와 판단이 가능하도록 하기 위한 '양질의 학습 데이터'의 확보가 중요하다. AI가 학습하는 데이터가 제한적이거나 부정확한 정보인 경우 이러한 데이터의 편향성으로 인해 AI가 적용되는 기술의 안전성에 문제가 발생하거나 신뢰하기 어려운 결과가 초래될 확률이 훨씬 높아지게 된다.

최근 발표된 OpenAI³⁾의 GPT-3 알고리즘의 경우, 1,750억개의 매개변수를 갖고 약 1조개에 가까운 단어로 이루어진 데이터셋과 책에서 추출한 말뭉치, 영문 위키피디아 등을 통해 학습되었다.⁴⁾ 이러한 양질의 학습데이터를 대량 확보함으로써 OpenAI는 인간이 작성한 글과 구별하기 어려운 수준의 글을 작성하는 것으로 평가받는 GPT-3 알고리즘을 완성할 수 있었다. 그러나 한편으로는, 연구자들 스스로 학습된 데이터에 편향(bias)이 존재함으로써 인해 고정관념(stereotype) 또는 편견(prejudice)이 있는 콘텐츠를 생성하게 될 것을 우려한 바 있다.⁵⁾

2. 데이터 활용에 관한 우리나라의 규제 현황

AI 기술 발전으로 인류는 많은 혜택을 누리게 될 것으로 예상되는 반면, 광범위한 데이터의 활용으로 인해 개인정보와 프라이버시 침해가능성 또한 커지고 있는 것이 사실이다. 특히 개인 스마트폰 보급과 사물인터넷으로 인한 초연결 서비스의 등장 등으로 인해 연간 생산되는 데이터 중 개인데이터의 비중이 약 75%에 이르는 것으로 추정되고 있다.⁶⁾

이러한 이유로 세계 각국은 4차 산업혁명 시대로 전환하는 과정에서 개인의 프라이버시를 충분히 보호하면서도 데이터 활용을 촉진하는 법제 마련을 위해 최선의 노력을 다하고 있다.⁷⁾

1) 관계부처 합동, 인공지능 국가전략, 2019. 12. 1면 이하

2) 경제협력개발기구(OECD)도 4차산업혁명 시대 디지털 변혁(Digital Transformation)의 사회·경제적 영향을 분석하고, 관련 공공 정책적 방향을 제시하는 "Going Digital Project"를 추진하고 있으며, 2019~2020년에는 인공지능 및 블록체인 등을 중심으로 위 프로젝트의 두 번째 단계를 추진 중에 있다. 2019년 5월에는 프랑스 파리에서 열린 각료사회에서 '고잉 디지털(Going Digital)' 보고서를 채택하였다(<http://www.oecd.org/going-digital> 참조).

3) 테슬라사의 창립자 일론 머스크를 비롯한 투자자들이 설립한 인공지능 연구기관(<http://openai.com/> 참조).

4) Tom B. Brown 외 30, "Language Models are Few-Shot Learners", OpenAI, 2020. 7. 22, 5-8면. <https://arxiv.org/pdf/2005.14165.pdf> (2020. 9. 28. 최종 확인)

5) 예를 들어, GPT-3 알고리즘은 여성에 대해 묘사할 때 "beautiful", "gorgeous"와 같은 단어를 남성을 설명할 때에 비해 더 많이 사용한다. 남성 묘사와 더 연관된 단어는 "Large", "Mostly", "Lazy", "Fantastic", "Eccentric", "Protect" 등이고, 여성 묘사와 더 연관된 단어는 "Optimistic", "Bubbly", "Naughty", "Easy-going", "Petite", "Tight" 등이다. Tom B. Brown 외 30, 앞의 논문, 34-39면.

6) 한국데이터진흥원, 국내 개인데이터 이용환경 분석 연구보고서, 2017. 11. 1면(출처 IDC, IDC Digital Universe Study, 2011. 1.)

그러나 우리나라의 경우 기술 패러다임 변화에 대한 전반적인 대응이 늦어짐으로 인해 데이터 활용 관련 평가지표들이 주요 비교 국가 대비 낮은 수준에 머물고 있는 것으로 확인되고 있다. 스위스 국제경영개발원(IMD)⁸⁾은 우리나라의 빅데이터 활용과 분석 수준이 2017년을 기준으로 전 세계 63개국 중 56위, 2018년을 기준으로 31위에 불과한 것으로 보고하고 있다.⁹⁾ 또한 데이터 가치사슬(구축, 유통, 활용) 내 시장형성이 미국의 1/400 수준으로 미진하고, 미국의 인공지능 기술수준을 100으로 하였을 때 우리나라의 인공지능 기술수준은 78에 불과하며, 국내 기업의 빅데이터 이용률은 7.5%에 불과한 수준이다.¹⁰⁾

한편, 연간 생산되는 데이터의 대부분이 개인과 관련한 데이터인 상황에서, 우리나라 개인정보보호법에는 OECD 주요 국가 대비 가장 규제가 높은 수준으로 사전 동의 및 행사처벌 위주의 형식적인 규제가 중심이 되고 있어 데이터를 활용하고자 하는 기업들의 법적 불확실성과 위험이 매우 크다고 평가되고 있다.¹¹⁾ 실제로 정보주체에게 미치는 위험이 없거나 미미한 경우, 특히 새로운 서비스 개발이나 기능 개발을 위해 개인정보를 활용하여 정보주체에게 혜택을 제공하고자 하는 경우 조차 개인정보를 최소화 수집할 때 정보주체에게 고지한 목적과 다르다는 이유로 개인정보보호법령 위반이 문제될 수 있는 상황으로, 인공지능 기술의 발전을 위한 산업적 활용은 물론 원천기술 개발 등 세계 각국과의 인공지능 기술 패권경쟁에서도 뒤처질 우려가 있는 상황이다.¹²⁾

다행히 우리나라 정부는 이러한 문제점을 뒤늦게나마 인식하고 2018년부터 데이터 및 AI 산업 활성화 정책을 본격적으로 추진하는 한편,¹³⁾ 4차 산업혁명위원회 해커톤 등 논의를 거쳐¹⁴⁾ 2018. 11. 15. 개인정보 보호법 개정안 등¹⁵⁾을 국회에 발의, 2020. 1. 9. 데이터 3법이 국회 본회의를 통과하여 2020. 8. 5.자로 시행되기에 이르렀다.

특히 최근 정부는 디지털 뉴딜을 추진하면서, 4대 분야 중 하나로 “Data-Network-AI (D.N.A) 생태계 강화”를 선정하고, 그 내용으로 “데이터 수집, 개방, 활용에서부터 데이터 유통 및 인공지능 활용에 이르기까지 데이터 전주기 생태계를 강화하고, 민관 합동 데이터 컨트롤 타워 마련을 통해 데이터 경제 전환 가속화를 추진” 하겠다는 의지를 밝힌 바 있다. 구체적으로 “민간 데이터 산업 활성화 및 데이터 기반 서비스·정책 개발 지원을 위해 정밀도로지도, 안전.취약 시설물 관리 정보 등 공공데이터의 개방과 연계를 확대하고, ‘21년까지 개방 가능한 14.2만개 공공데이터 전체를 신속히 개방하고, 공동 빅데이터 분석시스템을 구축하여 각 기관이 개별 시스템을 구축하지 않고도 타 기관 데이터와 연계.분석할 수 있도록 지원”하고, “수요가 많고 기술적으로 구현 가능한(언어 말뭉치, 자율주행 영상데이터 등) 인공지능 학습용 데이터를 ‘25년까지 1,300종 추가 구축하고, 중소.스타트업의 인공지능 기술 개발.적용을 위해 인공지능 학습용 데이터 가공.바우처를 기업에 제공하여 일자리 창출과 인공지능 고도화를 지원할 계획”을 발표하였다.¹⁶⁾

이하 필자들은 본 원고에서 AI 산업 발전에 필수적인 데이터 처리와 관련한

7)

유럽의회는 2018년 12월 AI 윤리 가이드라인(AI Ethics Guideline)의 초안을 발표하고 500개가 넘는 의견을 수렴하여 2019년 4월 8일 “신뢰할 수 있는 AI를 위한 윤리 가이드라인”(Ethics Guidelines for Trustworthy AI)을 발간하였다(<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> 참조). 일본 내각부는 2018년 5월부터 인간중심의 AI 사회원칙 검토회의를 총 8회에 걸쳐 개최하여 “인간 중심의 AI 사회 7원칙(안)”을 수립하였는데, 위 원칙은 개인정보 확보에 있어 개인의 자유·존엄·평등이 침해되지 않도록 하고 기업이 수집하는 개인 데이터의 정확성, 정당성 및 이용자 참여 가능성 등을 확보하여 AI 기술의 편익 극대화가 필요하다는 내용 등을 담고 있다(한국정보보호진흥원, “인공지능 윤리 가이드라인 - 일본과 EU 사례를 중심으로”, 지능정보사회 법제 이슈 2019. 1. 20면). 미국의 “연방 빅데이터 연구 및 개발 전략 계획(The Federal BIG DATA Research and Development Strategic Plan)”(2016), 중국의 “차세대 AI 발전규획”(2017), 독일의 “AI made in Germany Plan”(2018)은 모두 이러한 노력의 일환으로 볼 수 있다(관계부처 합동, -혁신성장 전략투자- 데이터·AI경제 활성화 계획[19~23년], 2019. 1. 16, 1면).

8)

IMD, “IMD WORLD DIGITAL COMPETITIVENESS RANKING 2017”, “IMD WORLD DIGITAL COMPETITIVENESS RANKING 2018” 참조. 스위스 국제경영개발원은 2017년부터 “기업이 의사결정에 빅데이터의 활용과 분석을 활용하는 정도”를 기준으로 각국의 빅데이터 활용과 분석 수준을 평가하고 있다.

9)

관계부처 합동, -혁신성장 전략투자- 데이터·AI경제 활성화 계획[19~23년], 2019. 1. 16, 4면

10)

관계부처 합동, 데이터를 가장 안전하게 잘 쓰는 나라를 만들겠습니다, 2018. 8. 31, 3면

11)

4차 산업혁명위원회, 4차 산업혁명 대정부 권고안, 71면

12)

미국 백악관은 2020년 5월 중국 당국이 "자유롭고 개방적인 규칙에 기반한 질서"를 약용하여 "국제 시스템을 자국에 유리하게 재정립하려고 한다"고 비난하는 전략 문건을 공개하고, AI, 양자컴퓨팅 등 중국 기업이 첨단 기술을 획득하는 것을 방지하는 보다 강력한 절차를 요구하고 있다. 이에 따라 미국 정부는 중국 전자 대기업인 화웨이와 칩 제조에 미국 기술이나 기계를 사용하는 것을 차단하는 등 국제 조치를 취한 바 있다. 한편, 중국은 검색엔진 결과의 검열, 중국 당국에 사용자 데이터, 소스 코드 등을 넘기는 것을 거부하는 국제 기업 들을 봉쇄하는 조치를 취하고, "메이드 인 차이나 2025" 같은 프로그램을 통해 "대량 혁신"을 육성하고 AI 등 전략 부문에 대한 보조금 증액 노력도 대폭 강화하였다(MIT Technology Review, "The virus that split the world", 2020. 9. 12면). 이처럼 최근 첨단 기술 영역에서도 미국과 중국을 중심으로 한 기술인족주의(technonationalism)에 기초한 패권주의가 점차 강화되고 있는 상황이다.

13)

과학기술정보통신부는 2018년 5월과 6월 각각 "데이터 산업 활성화 전략"과 "Korea 4.0 실현을 위한 인공지능(AI) R&D 전략"을 발표하였고, 2018년 8월에는 대통령이 직접 데이터 경제로의 전환을 천명하기도 하였다(관계부처 합동, "혁신성장 전략투자- 데이터·시경제 활성화 계획" [19~23년], 2019. 1. 16, 4면). 또한 2019. 12. 관계부처 합동으로 "인공지능 국가전략"을 발표하였는데, "AI 인프라 확충"과 관련하여 양질의 데이터 자원 확충을 위한 공공데이터 전면 개방(~21), 데이터 생산, 유통, 활용을 지원할 공공/민간 데이터 지도의 연계(21) 등을 추진하는 한편 "법제도 정비"와 관련하여 선하용-후규제의 기본방향 하에서 AI분야 포괄적 네거티브 규제 로드맵 수립(20), AI 시대 기본이념과 원칙, 역기능 방지 사책 등 기본법제 마련(20), (가칭) 미래사회 법제정비단 발족(20)을 통해 분야별 법제 정비 주도 전략을 추진하겠다는 내용을 담고 있다.

14)

2018년 5월 25일 GDPR 시행에 앞서 2018년 2월과 4월 두 차례에 걸쳐 4차산업혁명위원회 '규제·제도혁신 해커톤'을 개최하고, '개인정보의 보호와 활용의 균형 방안 마련'과 관련하여 정부, 공공기관, 법조계, 산업계, 시민단체까지 참여하는 대토론회를 통해 GDPR을 참조하여 개인정보 개념의 명확화, 가명정보 개념의 도입 및 활용 등에 관한 큰 틀의 합의를 이끌어 내었으며, 국회의 4차산업혁명특별위원회도 2018. 5. "개인정보 보호와 활용을 위한 특별 권고안"을 제시한 바 있다.

15)

2018. 11. 15. 국회에 발의된 법률안은 개인정보 보호법 개정안, 정보통신망 이용촉진 및 정보보호에 관한 법률("정보통신망법") 개정안, 신용정보의 이용 및 보호에 관한 법률("신용정보법") 개정안, 위치정보의 보호 및 이용 등에 관한 법률("위치정보법") 개정안 등 4개 법률로써 '데이터 4법'이었으나, 위치정보법 개정안을 제외한 나머지 법률 개정안들이 국회 본회의를 통과하였다.

여, (i) 정부가 데이터 산업 활성화 분야 중 하나로 언급한 자율주행을 위한 정밀지도에 관한 규제 개선 필요성을 검토하고, (ii) AI 기술 발전을 위한 학습데이터 활용에 관한 규제 개선 필요성에 대해서 살펴보도록 한다.

II. 자율주행을 위한 정밀지도 관련 규제 개선 필요성

자율주행자동차 기술을 개발하는데 있어 필수적인 정보 중 하나가 지도 정보이다. 특히 사람이 읽을 수 없는 형태의 지도가 아닌, 기계가 읽을 수 있는(machine-readable) 데이터 형태의 지도의 작성·활용이 필수적인데, 현행 『공간정보의 구축 및 관리 등에 관한 법률』("공간정보관리법")은 이러한 데이터 형태의 지도를 모두 "지도"로 포괄하여 규정¹⁷⁾ 하고(제2조 제10호), 기계가 읽을 수 있는 데이터 형태의 지도에 관한 별도의 규정을 두고 있지 않아 사람이 읽고 작성하는 지도에 대한 규제가 기계용 지도에도 그대로 적용되는 실정이다. 또한, 자율주행 자동차는 기존의 지도보다 상세하고 정밀한 지도 데이터를 수집하여 작성하여야 하며, 나아가 주행중인 자동차로부터 실시간으로 도로의 정보를 수집하여 지도를 업데이트 하는 등 지도 정보를 기존의 공간정보관리법 제정 당시 예상하지 못했던 형태로 활용할 필요성이 있다.

이와 관련하여 국토교통부는 2016년 자율주행차 상용지원을 위해 정밀도로지도 구축 계획을 발표한 바 있다.¹⁸⁾ 위 계획에 따라 구축되는 정밀도로지도는 자율주행에 필요한 차선, 도로시설, 표지시설 정보를 정확도 25cm로 제작한 전자지도로, 국토교통부, 국토지리정보원, 한국도로공사 및 14개 민간사가 참여하여 민간-공공 협력 하에 제작되고 있다. 그럼에도, 현행 규제를 살펴보면 국토교통부 국가공간정보 보안관리규정 별표1에 따라 해상도가 90m보다 정밀하고 3차원 좌표가 포함된 공간정보를 공개제한 정보로 규정하고 있고, 이 중 도로지역에 관한 정보는 보안성 검토를 거쳐 국가안보상 위해 요인이 없는 경우 공개하도록 되어 있어 정밀도로지도가 구축되더라도 민간의 원활한 활용이 불가능하고, 민간에서 직접 구축한 경우에도 유사한 규제가 적용될 것이라는 지적이 있는 실정이다.¹⁹⁾

또한 현행 공간정보관리법은 기본측량성과, 기본측량기록 또는 국토교통부장관이 간행한 기본측량성과 및 기본측량기록을 사용하여 지도나 그 밖에 필요한 간행물을 활용한 지도등을 간행하여 판매하거나 배포하려는 자는 국토교통부장관의 심사를 받도록 하고 있다(제15조 제3항). 공공측량성과에 대해서도 마찬가지이다(제18조 제3항). 측량성과심사는 국토교통부장관으로부터 국토지리정보원이 위탁받아 수행하되, 실무적으로는 국토지리정보원이 지정한 측량성과 심사수탁기관인 공간정보품질관리원이 심사하고 있다(제105조 제2항). 또한, 항공촬영, 지도제작 등의 업무를 하기 위해서는 등록기준을 갖추어 국토교통부장관 또는 시도지사에게 등록하여야 한다(제44조 제2항). 이러한 심사 절차는 사

람이 읽고 작성하는 지도를 전제로 하는 것으로, 위와 같은 기계용 지도 작성·배포에 그대로 적용되어야 할 필요가 있는지 의문이다.

또한, 공간정보관리법 시행규칙은 간행심사 절차를 규정하면서 간행심사를 받아 지도등을 간행한 자는 그 지도등을 수정하여 간행한 경우 수정간행한 지도 등의 사본을 측량성과 심사수탁기관에 제출하도록 규정하고 있다(제17조 제2항). 자율주행자동차는 수많은 센서를 통해 주행 과정에서 도로 정보 등을 실시간으로 수집하고 이러한 정보가 지도 데이터에 실시간으로 반영되어야 하는데, 자율주행자동차의 정확한 주행을 위해 사용되는 정밀도로지도는 도로 등의 사소한 변경이 즉각적으로 업데이트되어 시스템에 반영되지 않으면 자율주행 과정에서 오작동 등 사고 위험이 높아질 우려가 있다. 따라서 정밀도로지도는 신속한 업데이트가 필수적이다. 그러나 공간정보관리법 시행규칙에 따른 간행심사 절차 규정은 자율주행자동차의 지도 데이터의 신속한 업데이트에 걸림돌이 될 수 있고, '수정'의 범위나 의미, 제출 기한을 명시하고 있지 않아 실무적으로 혼란을 야기할 우려가 있다.

20대 국회에서 이러한 문제점을 시정하기 위한 공간정보관리법 일부개정법률안(강병원 의원 대표발의)²⁰⁾이 발의된 바 있다. 법안의 주요 내용은 (1) 기계용 지도를 정의하고 (2) 기계용 지도의 간행에 관하여 간행심사 시 심사사항을 간소화하고 기간의 제한을 두는 간이심사 절차를 신설하고, (3) 지도등의 수정간행의 범위를 명시하며 (4) 수정간행시 사본제출의무를 사후조치 의무로 규정하는 한편 (5) 기계용 지도의 수정간행시 사본제출 의무를 분기별로 국토교통부장관에게 보고하도록 하여 사본제출의무를 경감하는 내용을 담고 있다.

그러나 아쉽게도 해당 법안은 20대 국회 임기만료와 함께 폐기되었다. 21대 국회에서 공간정보관리법 개정안이 재발의되어 차후 AI를 이용한 자율주행자동차 기술에 있어 고속도로와 같은 산업적 근간으로 볼 수 있는 기계용 정밀지도 관련 산업도 함께 발전될 수 있는 토대가 마련될 필요가 있다.

III. AI 학습데이터 활용에 관한 규제 개선 필요성

1. 가명정보와 AI 학습데이터 활용

개정 데이터 3법은 '가명정보'의 개념을 도입하여 일정 범위, 특히 과학적 연구 목적으로 정보주체의 동의 없이 가명정보를 이용 및 제공하는 것을 허용하고 있다. 개정 전 개인정보 보호법에 의하면, 비식별 조치를 통해 더 이상 개인정보에 해당하지 않는 이른바 '익명정보'에 대해서만 동의의 규제가 배제되었으나, 데이터 분석 기술이 급속도로 발전하면서 완전한 비식별 조치가 사실상 불가능하게 되었고, 개인정보의 재식별 우려로 인한 법적 리스크로 인해 사업자들이 비식별 조치가 기반하여 데이터를 정보주체의 동의 없이 활용하는 것에 상당한 부담이 있었다.

16) 과학기술정보통신부, 디지털 뉴딜, 코로나 이후 디지털 대전환을 선도합니다, 2020. 7. 15.

17) 공간정보관리법 제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

10. "지도"란 측량 결과에 따라 공간상의 위치와 지형 및 지명 등 여러 공간정보를 일정한 축척에 따라 기호나 문자 등으로 표시한 것을 말하며, 정보처리시스템을 이용하여 분석, 편집 및 입력·출력할 수 있도록 제작된 수치지형도(항공기나 인공위성 등을 통하여 얻은 영상정보를 이용하여 제작하는 정사영상지도(正射影像地圖)를 포함한다)와 이를 이용하여 특정한 주제에 관하여 제작된 지하시설물도·토지이용현황도 등 대통령령으로 정하는 수치주제도(數値主題圖)를 포함한다.

18) 국토교통부, '정밀도로지도' 자율주행차 상용화 앞당긴다., 2016. 9. 28.

19) 윤서연, "자율주행용 정밀도로지도 구축을 위한 민간-공공 협력에 대한 정책제언," 국토정책 Brief 750호, 2020, 1-6면.

20) 의안번호 2022853, 제안일자 2019. 10. 8.

개정 개인정보 보호법은 “가명정보”를 “개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없는 특정 개인을 알아볼 수 없도록 처리”하는 “가명처리”를 통해 “원래의 상태로 복원하기 위한 추가 정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보”를 의미하는 것으로 정의하고 있다(제2조 제1호 다목, 제1의2호). 또한 개인정보처리자는 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 처리하는 것이 가능하다고 규정하고 있다(제28조의2 제1항).

개정 신용정보법에서도 가명정보를 별도로 정의하고 활용에 관한 근거 규정을 두고 있다. 개정 신용정보법은 “가명처리”를 “추가정보를 사용하지 아니하고는 특정 개인인 신용정보주체를 알아볼 수 없도록 개인신용정보를 처리하는 것”으로 규정하고, “가명정보”를 “가명처리한 개인신용정보”로 정의하고 있다(제2조 제15호, 제16호). 개정 신용정보법에 의하면 통계작성, 연구, 공익적 기록보존 등을 위하여 가명정보를 이용, 제공하는 경우 신용정보주체의 동의가 요구되지 않는 것으로 규정하고 있다(제32조 제6항 제9의2호, 제33조 제1항 제4호).

여기서의 ‘통계작성, (과학적) 연구’의 의미에 관하여, 개정 신용정보법은 “통계작성에는 시장조사 등 상업적 목적의 통계작성을 포함하며, 연구에는 산업적 연구가 포함된다”는 명문의 규정을 두고 있으나(제32조 제6항 제9의2호), 개정 개인정보 보호법은 “과학적 연구”란 “기술의 개발과 실증, 기초연구, 응용연구 및 민간 투자 연구 등 과학적 방법을 적용하는 연구를 말한다”고 규정할 뿐(제2조 제8호), ‘과학적 연구’에 산업적 목적의 연구가 포함된다는 명문의 규정을 두고 있지 않다. 그러나 개정 개인정보 보호법은 그 제안이유에서 “데이터를 기반으로 하는 새로운 기술·제품·서비스의 개발 등 산업적 목적을 포함하는 과학적 연구, 통계작성, 공익적 기록보존 등의 목적으로도 가명정보를 이용할 수 있”다고 설명하여 가명정보를 이용할 수 있는 과학적 연구에 산업적 목적이 포함된다는 점을 밝히고 있으며, 최근 개인정보보호위원회가 발간한 ‘가명정보 처리 가이드라인’에서는 “과학적 연구”에 대해 “기술의 개발과 실증, 기초연구, 응용연구 뿐만 아니라 새로운 기술, 제품, 서비스 개발 등 산업적 목적을 위해서도 수행이 가능하며, 민간 투자 연구, 기업 등이 수행하는 연구도 가능”하고, “통계작성”에 대해 “시장조사와 같은 상업적 목적의 통계 처리도 포함”된다고 설명하여, 민간의 산업적 연구 목적, 상업적 통계작성 목적으로의 활용이 가능하다는 점을 명확히 밝히고 있다.²¹⁾

따라서 개정 개인정보 보호법 및 신용정보법에 의하면 AI 기술 및 알고리즘 관련 연구를 위해서 정보주체의 동의 없이 개인(신용)정보를 가명처리한 후 해당 가명정보를 AI 학습데이터로 활용하는 것이 가능하게 되었다.²²⁾²³⁾

21) 개인정보보호위원회, 가명정보 처리 가이드라인, 2020. 9. 5면

22) 한편, 위치정보법은 데이터 3법과 함께 개정되지 않아 가명정보에 관한 규정을 두고 있지 않다. 개인위치정보와 관련하여 위치정보법은 일반적으로 개인정보 보호법의 특별법적 지위에 있다고 볼 수 있으므로, 개인정보 보호법의 가명정보에 관한 규정은 개인위치정보에 대해서는 적용되지 않는다는 견해가 있을 수 있다. 그러나 일반법과 특별법의 관계는 반드시 어떤 법령의 전체를 대상으로 하는 것이 아니라 개별 규정마다 달리 판단될 수 있다(대법원 2012. 5. 24. 선고 2010두16714 판결 등 참조). 즉, 위치정보법의 개별 규정이 개인정보 보호법의 개별 규정과 모순, 저촉되는 경우에만 위치정보법이 개인정보 보호법의 특별법으로 우선 적용된다고 보아야 한다. 위치정보법 제21조 제2호는 “통계작성, 학술연구 또는 시장조사를 위하여 특정 개인을 알아볼 수 없는 형태로 가공하여 제공하는 경우”에는 예외적으로 개인위치정보를 이용, 제3자에게 제공할 수 있다고 규정하고 있는데, “특정 개인을 알아볼 수 없는 형태로 가공하여”라는 문구를 가명처리를 규정한 것으로 보는 한편 학술연구 역시 과학적 연구와 같은 개념으로 보는 것이 그동안의 규제기관 입장이었다. 따라서 위치정보법 제21조 제2호가 가명정보 처리에 관한 개정 개인정보 보호법 제28조의2와 모순, 저촉된다고 보기 어렵다고 할 것이다. 결국 현행 위치정보법을 개정하지 않고도 위치정보법상의 개인위치정보를 가명처리하여 개정 개인정보 보호법상의 가명정보의 활용에 관한 규정에 따라 활용할 수 있다고 판단된다

23) 참고로 최근 보건복지부는 개인정보 보호법 개정에 따른 생명윤리법의 해석에 관하여, 개정 개인정보 보호법의 ‘가명처리’는 생명윤리법의 ‘익명화’에 포함되는 것으로 해석되므로, 의료기관에서 진료목적으로 수집된 의료데이터 등을 개인정보 보호법 상의 가명처리를 통해 연구목적 등으로 이용하려는 경우, “연구대상자 등에 대한 기존의 자료나 문서를 이용하는 연구”로 간주하고 기관 차원에서 가명처리가 확인된 경우 IRB(기관생명윤리위원회) 심의 및 동의를 면제할 수 있다고 해석함으로써(생명윤리정책과-2605, 2020. 8. 4.), 생명윤리법에서 규율하는 유전자 정보 등 의료데이터에 대해서도 개인정보 보호법상 가명처리 제도가 적용될 수 있다는 입장을 취하고 있다.

2. 가명처리를 통한 학습데이터 활용의 문제점

데이터 3법 개정으로 산업적 연구 목적으로 정보주체의 동의 없이 가명정보를 이용하는 것이 허용되었으나, AI산업 활성화의 필수조건인 AI 학습데이터 확보가 충분히 이루어질 것으로 기대하기는 어렵다.²⁴⁾

우선 AI 학습데이터 확보와 관련하여 반드시 가명처리가 전제되는 경우, 다른 정보와 쉽게 결합하더라도 개인을 식별할 수 없는 수준으로 대규모의 디지털 데이터에 대한 가명처리가 이루어져야 하는데 그에 대한 검증, 적정성 평가 등의 과정에서 상당한 시간과 막대한 비용이 소요될 수 밖에 없다. 특히 개정 개인정보 보호법은 가명정보의 처리 중 재식별 행위와 재식별된 가명정보의 활용에 대해 형사처벌, 전체 매출액의 3퍼센트 이하의 과징금 등의 엄격한 법적 제재를 예정하고 있다. AI와 빅데이터 처리 기술이 발달할수록 가명정보의 처리 과정 중 특정 개인의 재식별 가능성은 함께 높아지게 되므로 관련한 법적 리스크는 더욱 커질 수 밖에 없다는 점도 AI 학습데이터 활용에 큰 걸림돌이 될 것으로 보인다. 특히 개정 개인정보 보호법은 “개인정보처리자는 가명정보를 처리하는 과정에서 특정 개인을 알아볼 수 있는 정보가 생성된 경우에는 즉시 해당 정보의 처리를 중지하고, 지체 없이 회수·파기하여야 한다”라고 규정하고 있어(제28조의5 제2항²⁵⁾) AI를 통한 대규모의 디지털 데이터를 학습데이터로 분석할 경우 그 과정에서 특정 개인을 식별할 수 있는 정보에 대한 즉각적인 처리 중지·파기 기능을 함께 구현해야 하는데 이를 기술적으로 구현하는 것이 가능한지도 의문인 상황이다.

3. 가명처리 없는 AI 학습데이터 활용의 필요성

따라서 4차 산업혁명 시대에 있어, 일상화되고 있는 대규모 디지털 데이터 처리와 관련한 인공지능 학습에 대해서는 (개별 정보주체에 대한 프로파일링 내지 중요한 권리의무에 영향을 미치는 자동화된 의사결정과 관련된 것이 아닌) 단순 학습행위와 관련된 경우에는 접근통제, 암호화 등 일정한 기술적, 관리적 보호조치를 전제로 하여 가명처리 없이 개인정보를 활용할 수 있도록 하는 방안을 적극적으로 고려할 필요가 있다.

실제 저작권법과 관련하여 AI 학습을 위한 디지털 저작물 활용에 대해서는 이러한 관점에서 법 개정 작업이 이루어지고 있다. 우리나라는 2006년 이후 14년 만에 저작권법 전부개정을 추진하고 있으며, 그 주요 내용 중 하나로 인공지능 학습 및 빅데이터 분석 등을 위하여 저작물의 이용이 필요하다고 인정되는 범위 내에서 저작물의 이용허락을 받지 않아도 되도록(침해 면책) 허용하는 조항의 신설이 논의되고 있다. 영국의 경우 저작권, 디자인 및 특허법(Copyright, Designs and Patents Act)에 따라 비상업적 연구 목적의 텍스트 및 데이터 분석을 위한 복제를 허용(제29A조)하고 있으며, EU는 ‘디지털 단일시장 저작권지침’에서 과학적 연구 목적의 텍스트/데이터 마이닝 등을 허용하고 있다.²⁶⁾ 일본은 2018. 5. 저작권법 개정을 통해, 기술개발을 위한 시험의 경우 등 당해 저작물에 표현된 사상

24)

이와 관련하여, 개인정보를 보호하면서 데이터를 분석할 수 있는 방법으로 암호화된 상태에서 연산하는 경우에도 연산 결과가 보존되는 ‘동형 암호’ 등의 방법도 제안되고 있으나, 현재로서는 평문에 비해 분석에 필요한 컴퓨터의 용량·연산량이 상당히 증가하는 등의 단점도 함께 지적되고 있다.

25)

이를 위한 경우 3천만원 이하의 과태료를 처해 질 수 있다(개인정보보호법 제75조 제2항 제7의2호).

26)

문화체육관광부, 창작자 권리보호는 강하게, 저작물 이용은 손쉽게, 2020. 7. 1, 7면

27)

신현철, “제4차 산업혁명시대의 일본 저작권법 개정과 시사점,” 법학논총 26권 3호(2019. 12), 조선대학교 법학연구원, 12~16면

또는 감정을 스스로 향수하거나 타인이 누리게 하도록 하는 것을 목적으로 하지 않는 경우에는 그 필요가 인정되는 한도에서 저작물을 이용할 수 있도록 하고 있으며(제30조의4), 전자계산기를 사용하여 정보를 검색하거나 정보 분석을 하여 그 결과를 제공하는 자는 공중예의 제공 또는 제시가 이루어진 저작물에 관하여 그 행위의 목적상 필요하다고 인정되는 한도에서 당해 행위에 수반하여 경미한 이용 등을 할 수 있도록 하고 있다(제47조의5).²⁷⁾

개인(신용) 정보의 AI 학습데이터 활용에 관하여서도 이러한 관련 법 개정논의를 적극적으로 참고할 필요가 있다.

4. 가명처리 없는 학습데이터 활용에 있어 고려사항

가명처리 없는 학습데이터 활용을 일정 범위에서 허용하더라도, 학습데이터 활용을 통한 과학적 연구 과정에서의 윤리적 통제는 준수되어야 할 것이다.

과학적 연구에 관하여 EU에서는 그에 대한 윤리적 통제 기준 마련(ethics by design)이 이미 중요한 이슈로 논의되고 있다. European Data Protection Supervisor(“EDPS”)가 발간한 “Preliminary Opinion on Data Protection and Scientific Research”에서는, 연구자가 AI 시스템을 운용하는 경우 전문가적 윤리적 기준으로서 safeguard 및 감독이 요구되고, IT, 엔지니어링 리서치에 대한 윤리적 감독이 종종 결여되는 경우를 고려하여 European Commission은 향후 모든 EU 지원 AI 프로젝트에 ‘ethics by design’이 적용될 수 있도록 노력하기로 하겠다고 설명하고 있다. 특히 EDPS는 위 의견에서 윤리적 통제에 관하여 리스크 기반 접근법(risk-based approach)이 고려되어야 한다고 설명하면서, 리스크 판단의 척도로 “데이터의 민감성 또는 고도의 개인적 성격, 정보주체의 취약성, 대규모의 처리 활동, 모니터링의 조직적 성격, 과학기술적 솔루션의 혁신적인 이용 또는 적용, 개인의 평가, 데이터셋의 결합, 자동화된 의사 결정에 미치는 법적 또는 그와 유사한 중요한 영향”이 포함되어야 한다는 의견을 제시하고 있다.²⁸⁾

우리나라도 AI를 활용한 산업적 연구가 활발히 이루어질 것을 대비하여, 비윤리적 연구 내지 개인정보의 오남용으로 이어지지 않도록 하는 명확한 가이드라인 마련에 관한 전문가들의 논의가 본격적으로 필요할 것으로 생각된다. 특히 AI 관련 산업적 연구 과정에서 민감정보, 특히 건강정보에 기초한 의료 연구(medical research)가 본격화될 수 있으므로 AI 관련 연구의 윤리적 통제 기준(ethics by design) 마련에 관한 논의와 준비를 본격적으로 시작할 필요가 있다.

28)

EDPS, “Preliminary Opinion on Data Protection and Scientific Research”, 2020. 1. 6, 24면

IV. 결론

금번 데이터 3법의 개정으로 데이터의 보다 자유로운 활용에 관한 상당한 진전이 있었다고 평가되지만 여전히 데이터 3법은 형식적인 사전 동의 규제를 원칙으로 하는 등 기본적으로 4차 산업혁명 이전 오프라인을 통한 개인정보 처리를 전제로 하는 규제 프레임워크를 그대로 유지하고 있다.

개인정보자기결정권이라는 근대적 자유권에만 기초한 기존의 접근 방식은 개인정보의 이용 형태, 범위 등이 날로 복잡해지는 현 상황에서 정보주체에게 모든 의사결정을 스스로 하도록 하는 부담으로 작용할 뿐 아니라 오히려 개인에게 책임이 전가되는 역설적 결과로 이어지고 있다. 따라서 개인데이터의 디지털 전환(digital transformation), 빅데이터, 인공지능 등과 같은 데이터 처리 기술 등의 본격적인 확산 등 4차 산업혁명을 전제로 한 개인정보보호법제의 근본적인 패러다임 전환이 필요하다. 4차 산업혁명 시대의 개인정보보호법제는 형식적인 사전 동의 규제를 합리적으로 개선하고 데이터 활용에 대한 기업의 투명성을 제고하는 방향으로 개선할 필요가 있다. 또한 개인정보보호법 위반에 대해서는 과태료, 형사처벌 위주였던 관련 규정을 과징금 규제 위주로 변경하는 한편 개인정보처리자에게 개인정보 처리에 있어 서비스 내지 제품 설계 단계부터 프라이버시 보호 의무(privacy by design and default)를 부담하도록 그 내용을 전반적으로 개선하는 것이 중요할 것으로 생각된다.

정밀지도, AI 학습데이터의 활용에 관한 법제도도 위와 같은 4차 산업혁명 시대의 새로운 패러다임에 맞추어 개정하는 것이 반드시 필요할 것으로 생각된다. 세계 각국과의 AI 기술 패권경쟁에서 뒤처지지 않기 위해서는 금번 데이터 3법의 개정만 만족하여서는 안 되며, 기존의 낡은 패러다임을 벗어나 4차 산업혁명 시대 새로운 데이터 처리 기술을 반영한 법제도의 전면적인 개정을 서둘러야 할 것이다.

