빅데이터와 빅퀘스천

- 빅데이터 활용에 대한 인문학적 비판과 질문 -

유강하*

〈차 례〉

1. 머리말
2. 개인정보 유출과 프라이버시의 침해
3. 오독과 조작의 문제
 1) 데이터 오독의 가능성
 2) 데이터 조작의 위험성
4. 인간의 생명과 존엄성에 관한 문제
 1) 범죄·치안 영역에서의 빅데이터 사용
 2) 의료 영역에서의 빅데이터 사용
5. 맺음말

【국문초록】

빅데이터, 사물인터넷(IoT), 인공지능은 현대 사회의 중요한 화두이다. 이 가운데서도 빅데이터는 지금까지 공공의 영역과 기업의 이윤 추구에 생산적인 결과를 도출해 왔다. 현재까지의 데이터를 바탕으로 미래를 예측할 수 있다는 빅데이터에 대한 전망은 빅데이터의 적극적 활용으로 이어지게 되었다. 그러나 빅데이터 사

* 江原大學校 人文科學硏究所 HK助敎授

용에도 우려할 만한 지점이 존재한다. 빅데이터는 인간과 삶, 사회에 유용하게 활용될 것이라는 믿음에서 만들어졌지만, 지금까지 빅데이터는 주로 기업의 이윤 창출과 효용성의 극대화에 적극적으로 활용되었고, 이러한 과정 속에서 '인간'이라는 기본적인 철학이 붕괴되고 있는 듯이 보인다. 빅데이터 활용이 가속화되는 이 시점에, 인문학은 도덕적, 윤리적 감수성을 제공한다는 점에서 진지하게 다루어질 필요가 있다. 특히 다음과 같은 세 영역에서 인문학적 성찰과 질문이 필요해 보인다. (1) 개인정보 유출과 프라이버시의 침해 문제, (2) 오독과 조작의 문제, (3) 인간의 생명과 존엄성에 관한 문제가 그것이다. 빅데이터가 야기하는 문제는 인간, 삶, 개별성, 존엄성이라는 근원적인 문제라는 점에서 소홀히 다룰 수 없다.

현재 우리 사회에는 빅데이터가 인류를 더 나은 미래로 이끌어 줄 것이라는 전망이 우세하지만, 빅데이터를 통해 분명히 알 수 있는 것은 과거와 현재일 뿐, 그것이 곧 미래에 대한 정확한 예측을 의미하지는 않는다. 빅데이터가 분명한 방향성을 제시할 수는 있지만, 그것이 인류의 건강한 미래로 직결된다고 단언하기는 어렵다. 빅데이터가 인간과 삶, 인류의 미래를 위해 사용되는 것이라면, 다수(big)의 흐름에 앞서 '인간 삶'이라는 기본 전제 위에서 의심하고 비판하며, 건강한 방향성을 설정하려는 인문학적 비판이 필요하다. 빅데이터가 창출한 가치에 대한 환호에 앞서, 그것이 과연 무엇을 위해 존재하는지에 대해 인문학이 던지는 근원적이고 큰 질문, 즉 빅퀘스천이 필요하다.

주제어: 빅데이터, 빅퀘스천, 알고리즘, 오독, 조작, 프라이버시, 개인정보, 인간의 존엄성

1. 머리말

빅데이터, 사물인터넷, 인공지능이 중요한 화두로 제시되고 있다. 이 가운데서도 빅데이터는 우리의 현실과 연동되어 유용하게 활용되고 있다. 빅데이터 분석가는 떠오르는 유망 직업군 가운데 하나가 되었고, 이러한 시대적 수요는 관련 대학과 대학원의 설립과 산업수요에 걸맞은 인재 양성으로 이어지고 있다.[1] '빅데이터'라는 용어가 학술대회는 물론이고, 대중서와 대중강연에 빈번하게 등장하는 것은 빅데이터가 오늘날 우리의 현실에게 차지하고 있는 위상을 설명한다.

빅데이터는 문자 그대로 '커다란, 많은'(big) 정보(자료)라는 의미를 갖고 있다. 빅데이터는 용량(volume), 빠른 속도(velocity), 다양성(variety)으로 대표되는 3V에 가치(value)를 더해 '4V'라는 간결한 용어로 설명되기도 한다. 실제로 지난 수년 간 빅데이터는 그 유용성을 입증해 왔다. 지난 2008년, 구글(Goggle)은 감기와 관련된 급격한 검색어의 급증을 통해 미국 내에서 독감이 유행할 것임을 예측하였다. 이는 미국질병통제센터보다 약 2주 빠른 예측으로서, 이 예측이 적중하면서 빅데이터는 미래를 바꿀 새롭고도 중요한 기술로 각광받기 시작하였다. 이후, 빅데이터 기술은 공공 영역까지 확대되어 그 가능성을 확인시켜 주었다.[2] 앞으로 빅데이터의 사용범위는 확대되고 유용성은 증가하며, 보다 큰 파급력을 가

1) 김기봉 외, 『빅데이터와 사물인터넷 시대 인문학의 재구성 방안 연구』, 경제·인문사회연구회, 2015, pp.73-77.

2) 빅데이터는 공공의 영역과 기업의 이윤을 최대화 하는 데 활용되어 왔다. "대표적으로 넷플릭스사의 유료인터넷 드라마 '하우스 오브 카드'(House of Cards)의 성공, 구글의 감기지도(Flu-Map)로 확인된 질병의 예측과 예방, 서울시 시내버스의 효율적인 노선변경, 스페인의 스마트 가로등과 쓰레기통 등이 빅데이터 활용의 성공적인 사례로 꼽힌다." 유강하, 「빅데이터와 사물인터넷 시대의 비판적 해석과 인문학적 상상력」, 『시민인문학』 30, 2016, p.97.

져올 것으로 전망된다. 빅데이터가 미래를 예측할 수 있다는 기대와 환상은 빅데이터와 관련된 기사의 제목과 내용을 통해서도 충분히 읽어낼 수 있다.

"데이터 자체는 황금이 아니라 흙덩어리다. 하지만 그 안에는 모래알 같은 황금 알갱이들이 박혀있다."[3]

"모든 기업이 빅데이터에서 금맥을 찾는다. (……) 데이터를 '4차 산업혁명 시대 원유'라 부르는 이유다."[4]

"무궁무진한 공공데이터 활용해 '금맥' 캐는 시대 열렸다."[5]

"일자리 금맥 빅데이터시장 급성장…2019년 490억 달러 예상"[6]

빅데이터에 대한 다양한 비유, 즉 황금, 금광(金鑛), 금맥, 원유, 블루오션 등의 용어는 빅데이터가 우리 사회에 어떻게 받아들여지고 있는지 잘 설명해준다. 이러한 기사는 빅데이터의 사용이 자본의 증식, 부의 창출이라는 연장선상에서 이해되고 있음을 보여준다.

빅데이터가 각광받게 된 것은 비용 절감, 가치 창출을 가능케 하고, 무엇보다 기존방식으로는 얻기 어려운 해답을 인류에게 제공해줄 것이라는 기대 때문이다. 공공의료, 치안 등 공공 영역에서 빅데이터를 활용하는 이유일 것이다. 인문학 분야에서조차 빅데이터를 통해 연구 방식 등을 논의하는 것이 현실이다.[7]

---

3) 에릭 시겔, 고한석 옮김, 『빅데이터의 다음 단계는 예측분석이다』, 이지스퍼블리싱, 2016, p.27.

4) 「[글로벌기업, 그들은 한국에 무엇인가] 4차 산업혁명 원유 '데이터' 쓸어 담는 그들」'(『전자신문』 2017/12/26)http://www.etnews.com/20171226000244

5) (『동아일보』(2017/08/07)) http://news.donga.com/3/all/20170807/85707589/1

6) (『매일경제』(2017/10/10))http://news.mk.co.kr/newsRead.php?year=2017&no=668918

7) 박정식 · 손정훈, 「빅데이터를 활용한 대중의 인문학 인식 연구」, 『글로벌

그러나 빅데이터가 우리에게 장밋빛 미래만을 보여주는 것은 아니다. "빅데이터는 큰 것부터 작은 것까지 전모를 볼 수 있게 하는 망원경인 동시에 현미경"의 역할을 하기도 하지만, 판옵티콘에 갇힌 죄수를 감시하는 빅브라더의 역할도 담당한다.[8] 빅데이터는 만능열쇠가 아닐 뿐만 아니라, 빅데이터가 초래하는 문제의 해결 방법도 간단하지 않다. 빅데이터의 활용 영역이 '인간의 삶'이라는 영역이라는 점도 간과되어서는 안 된다. 일부 영역에서 증명된 빅데이터의 효용성을 '모든' 영역으로 확대하는 것에 대한 문제에 대해서는 비판과 의심이 필요하다.

빅데이터는 인간과 삶, 사회에 유용하게 활용될 것이라는 믿음에서 만들어졌지만, 지금까지 빅데이터는 주로 기업의 이윤 창출과 효용성의 극대화에 적극적으로 활용되었고, 이러한 과정 속에서 '인간'이라는 기본적인 철학이 붕괴되고 있는 듯이 보인다. 빅데이터 활용이 가속화되는 이 시점에, 인문학은 도덕적, 윤리적 감수성을 제공한다는 점에서 진지하게 다루어질 필요가 있다. 이 글에서는 인문학이 물을 수 있고, 물어야 하는 질문 또는 의심을 크게 세 가지로 나누어 살펴보고자 한다.

문화콘텐츠』 25, 2016, pp.65-80; 문상호, 「엔그램 뷰어를 이용한 인문학의 빅데이타 사례 연구」, 『예술인문사회융합멀티미디어논문지』 5-6, 2015, pp.57-65 등.

8) 김기봉 외, 같은 책, p.16.

2. 개인정보 유출과 프라이버시의 침해[9)]

빅데이터는 작은 데이터들이 모인 집합적 데이터를 말한다. 많은 사람들은 빅데이터의 유용성에 대해서 인정하면서도, 프라이버시와 개인정보의 유출에 대해서는 민감하게 반응하기 마련이다. 빅데이터의 문제점과 관련하여 가장 많이 제기되었고 연구된 분야 또한 개인정보유출과 프라이버시에 관한 문제이다.[10)]

현대 사회에서 개인은 수많은 데이터를 생성하는데, 때로 이 데이터들은 무차별적으로 수집되고 가공되며, 재사용되고 있다.[11)] 이 과정에서 개인정보유출, 프라이버시 침해 등의 문제가 발생할 가능성이 높고, 실제로 이러한 개인정보나 개인의 프라이버시가 기업의 비즈니스에 활용되고 있는 상황이다.[12)] 이는 개인정보와

---

9) 개인정보와 프라이버시는 구별되는 개념이다. 차상육은 우리나라 "개인정보보호법 제2조에서는 '개인정보'란 살아있는 개인에 관한 정보로서, 당해 정보에 포함된 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보를 말한다고 정의하고 있다. (……) 한편으로, 프라이버시에 관해서는 개인정보관련 법령상에 그 정의는 없다."고 말하면서, 개인정보와 프라이버시는 중복되는 부분도 있지만, 양자가 반드시 일치하는 것은 아니라고 설명하고 있다. 차상육, 「빅데이터 환경과 프라이버시의 보호」, 『IT와 法연구』 8, 2014, p.202.

10) Fred H. Cate 손금주·이영조 옮김, 「빅데이터 활용에 있어서의 프라이버시 보호」, 『경제규제와 법』 8-1, 2015; 이대희, 「빅데이터와 개인정보보호: 통지와 동의의 원칙을 중심으로」, 『정보법학』 19-2, 2015 등. 국내에서 지금가지 이루어진 빅데이터 관련 연구는 프라이버시와 개인정보보호에 집중되어 있다고 말할 수 있다. 연구 초기에는 프라이버시와 개인정보보호의 중요성에 대한 연구가 주로 이루어져 왔으나, 개인정보보호와 정보이용의 균형을 주장하는 연구도 꾸준히 증가하고 있다.

11) "대량의 개인정보는 우리가 직접적인 거래를 하지 않거나 매우 제한적인 거래를 하고 있는 제3자들에 의해 수집된다. 뉴욕타임즈에 따르면, 액시엄(Acxiom)만해도 50조개의 데이터 거래에 관여하고 있으며 거의 모든 거래는 개인으로부터 직접 수집하는 형태의 거래가 아니라고 한다." Fred H. Cate, 손금주·이영조 옮김, 위의 논문, p.22.

12) 미국에서 ok큐피드라는 웹사이트를 운영하여 상당한 양의 데이터를 축적

프라이버시 보호와 기업의 이윤 창출 가운데 어디에 우선순위를 두어야 할지의 문제로 이어진다.

빅데이터 전문가인 데이븐포트는 빅테이터에서 비롯된 일상의 편의성과 필요성을 다음과 같이 강조한다. "인터넷 쇼핑몰에서 내가 둘러본 상품과 구매 내용 등을 분석해 취향에 맞는 제품을 자동으로 추천해주고, 구글이 내 이메일과 검색 자료에 바탕해 내가 관심이 있을 법한 광고를 보여"주며, 유저들이 "일일이 개입하지 않아도 컴퓨터가 알아서 적절한 판단을 내"[13]려준다. 그러나 그가 언급한 "쿠키 정보를 활용한 온라인 맞춤광고 사례, 구매정보를 활용한 고객 신분·성향분석 사례, SNS 및 위치정보를 활용한 맞춤형 광고 사례, 의료데이터를 활용한 질병 예측 사례, 범죄 DB를 이용한 범죄감시 시스템구축"[14] 등은 개인정보 유출과 프라이버시 침해 문제로 이어지는 대표적 사례이다.

데이븐포트는 빅데이터 사용으로 발생할 수 있는 부작용에 대해서는 설명하지 않고, "미국인들의 경우 구글과 페이스북 같은 서비스를 무료로 이용하면서 자신들의 데이터가 비즈니스에 활용되고, 이로 인한 개인정보 유출의 위험이 있는 것을 '공짜의 대가'로 인식한다"고 말한다. 개인정보 유출의 위험은 조건 없이 공짜로

할 수 있었던 크리스천 러더는 데이터 수집에 대해 다음과 같이 밝힌 바 있다. "조금 창의적인 단어선택과 몇 차례의 시도, 약간의 잔머리만 있으면 사람들이 마음에 품은 독백을 수많은 청중 앞에 드러낼 수 있다. 사람들은 터무니없는 생각은 물론 남의 마음에 상처를 주는 생각까지 검색창에 털어놓는다."(크리스천 러더, 이가영 옮김, 『빅데이터 인간을 해석하다』, 다른, 2015, p.175.) 그는 이렇게 내밀하게 토로된 데이터가 사람들의 속내를 알아내는 유용한 정보가 되고, 자신조차 그것을 사용하고 있다고 말한다. 아무에게도 말할 수 없는, 내밀한 개인의 비밀들이 데이터로 저장되어 비즈니스에 활용되고 있다는 사실을 알 수 있다.

13) 「한국은 빅데이터 金鑛(금광·goldmine)인데… 제대로 캐질 못해」, 『조선비즈』(2015/10/14) http://srchdb1.chosun.com/pdf/i_service/pdf_ReadBody.jsp?Y=2015&M=10&D=14&ID=2015101400058

14) 차상육, 위의 논문, p.199.

구글과 페이스북을 사용하기 위해 감수해야 하는 리스크라는 논리다. 또한 그는 한국은 그 어떤 나라보다 질 높은 빅데이터를 보유하고 있지만, 이를 활용하는 데 뒤쳐져 있다며, 빅데이터 사용에 있어서 환경의 변화가 필요하다고 언급하기도 하였다.[15] 그러나 우리사회는 개인정보 유출로 인한 개인적·사회적 문제를 경험한 바 있고,[16] 유사한 문제들은 여전히 진행 중이다.

데이븐포트가 지적한 것처럼 '개인 정보 유출의 위험'이 사용대가와 등가의 가치를 가지는지에 대해서는 고민이 필요하다. 지금까지 빅데이터는 기업의 이윤 추구를 목표로 하는 비즈니스 영역에서 성과를 도출했고,[17] 공공분야에도 긍정적으로 기여한 측면이 있는 것이 사실이다. 기업의 영리적 사용은 경제적 발전이라는 논리로, 범죄·치안 등 공공의 영역에서는 공익 추구라는 근거로 사용되었다. 그런데 샘플조사가 아닌 전수조사를 위해서는 정보 공여를 원하지 않는 개인의 정보까지도 포함하게 된다. 빅데이터를 활용하는 측에서는 개인 식별을 불가능하게 하는 장치를 사용한다고 하지만, 개인정보 유출이라는 근본적 문제는 해결하기 어렵다. 빅데이터가 경제발전과 공익이라는 목적 추구를 위해 사용되지만, 그 과정에서 인간이 소외되고 개인정보가 도구화되는 현실로 이어지는 것이다. 개인의 프라이버시가 침해되고, 그에 대한 암묵적 간

15) "한국에서도 이러한 부분에 대한 사회적 고민과 공감대가 있어야 할 것 같다." 『조선비즈』(2015/10/14), 같은 글.

16) 빅데이터의 비즈니스에 따른 정보권력의 출현 가능성과 이로 인해 벌어지는 문제들, 특히 신용카드 회사에서 누출된 개인정보가 제2차 피해로 이어지는 현실적 문제들이 바로 그것이다. 2014년 1월에 발표된 금융기관의 개인정보유출 사건은 개인정보의 중요성과 심각성을 일깨우는 사건이 되었다. 차상육, 위의 논문, pp.194-195.

17) "정부보다 기업들은 이미 빅데이터의 산업화를 보다 더 발빠르게 구축하고 있다. 기업들은 나이, 성별, 구매이력 등 고객의 개인정보에서 관심사나 구매 가능성을 찾아내 마케팅에 활용하는 빅데이터는 업종과 분야를 불문하고 기업에는 꼭 필요한 것이라고 외친다." 차상육, 위의 논문, p.195.

과를 강요받고 있는 현실에서, 과연 빅데이터가 무엇을 위해 사용되어야 하고, 또 어떠한 방향으로 나아가야 하는지에 대해 비판과 질문이 필요하다. 개인정보보호에 대한 요청, 프라이버시 침해에 대한 비판, 빅데이터의 존재 의미를 묻는 근원적 질문이 개인·사회의 발전을 가로막는 것은 아니다. 오히려 이러한 인문학적 비판과 질문은 빅데이터가 개인이나 기업의 이익을 대변하는 것이 아니라, 인간의 삶에 복무하기 위한 도구라는 점을 상기시킨다는 점에서 의미가 있다.

빅데이터 사용에는 '사회적 기여'라는 수식어가 동반되곤 한다. 그러나 맥락을 살펴보면 그 '사회'는 공공의 이익보다 기업의 이익을 지칭하는 경우가 많다. 앞서 언급한 것처럼, '황금', '금맥', '원유' 등으로 지칭되는 빅데이터는 공공영역보다는 사적인 이익이나 기업의 이익 증대, 비용 절감의 목적으로 활용되어 왔다.[18] 기업의 이익과, 성공적인 비즈니스 모델의 수립이 개인과 사회의 이익과 항상 일치하는 결과로 이어지지는 않는다. 빅데이터 전문가 양성에 앞서, 빅데이터의 활용의 목적과 방향성을 설정하는 것이 중요하다. 개인정보와 프라이버시 보호에 대한 문제는 사회적 합의가 필요한 문제로서, 빅데이터 전문가뿐만 아니라 법률・사회・인

18) 이에 해당하는 사례를 찾는 것은 결코 어렵지 않다. 미국의 체이스은행에서는 부동산 담보대출 리스크를 줄이기 위해 빅데이터 분석을 실시하였다. 이용자들의 데이터를 활용한 부동산담보대출 예측 시스템은 실제로 사용되어 고객 이탈을 막고, 미국 최대의 은행으로 성장하는 데 크게 기여하였다.(에릭 시겔, 고한석 옮김, 위의 책, p.220; p,239.) 많은 데이터를 토대로 만들어진 의사결정 나무의 개발과 실행은 결국 기업의 이익에 복무하는 유용한 도구로 활용되었다.(에릭 시겔, 고한석 옮김, 위의 책, p.27.) 기업의 성공적인 인수를 설명하는 이 사건의 외피를 한 겹 들추어 보면, '서브프라임 모기지' 사태로 고통을 받은 사람들의 이야기가 숨어 있다. 이 은행 이용자들의 개인적 정보는 낱낱이 분석되어 기업의 키우는 황금 알갱이의 역할을 하였다. 그러나 수많은 이용자들은 데이터 제공자로 이용되었을 뿐, 아무런 이익을 얻지 못했고 이후 경제적인 고통에 시달려야 했다.

문학 영역의 전문가들이 함께 하는 진일보한 학문적 논의와 실천이 필요하다.

3. 오독과 조작의 문제

많은 사람들은 빅데이터를 통해 미래를 정확하게 '예측'할 수 있다고 생각한다. 그러나 분석과 해석을 거치지 않은 데이터는 무의미한 데이터 더미에 불과하다. 빅데이터의 유용성은 쓸모 있는 데이터 발굴과 해석에 달려 있는데, 문제는 데이터를 분석하고 해석하는 과정에 인간의 선입관이 작용하기 쉽다는 것이다. 결과적으로 빅데이터에 대한 기대는 의도적이고 작위적인 데이터 해석으로 이어질 수 있다. 원하는 결과 도출을 위한 데이터의 취사선택도 가능한데, 이것은 데이터 오독과 조작의 문제로 이어질 가능성을 배제할 수 없다. 사람들의 기대와 달리 빅데이터는 오용될 가능성도 적지 않은 것이다. 여기에서는 오독과 조작의 가능성과 그 위험성을 따로 나누어 살펴보도록 할 것이다.

1) 오독의 위험성

데이터로부터 유의미하고 가치 있는 정보를 찾기 위해서는 디지털 마이닝(mining)이 필요하다. 그런데 빅데이터의 핵심을 차지하는 SNS 데이터는 비정형 데이터로서, 그것이 담고 있는 의미가 매우 다양하다.[19] 이러한 데이터를 기계적으로 해석할 경우, 오독을 면하지 못하리라는 염려는 이미 오래전부터 지속되어 왔다. 기계는 데이터를 모을 수 있지만, 자기과시를 위한 글, 위로받기 위

19) 윤상오, 위의 논문, p.108.

한 과잉감점의 분출, 거짓말, 디지털 잔해,[20] 행간에 숨겨진 의미[21] 등 인간의 내밀한 정서와 감정을 정확하게 읽어내기는 쉽지 않다. "특히 문맥을 고려해야 하는 다의적(多義的) 표현이자 유머, 풍자, 아이러니와 같은 복잡한 표현을 분석할 때 오류가 발생하기 쉽다."[22] 따라서 데이터를 기계적이나 과학적으로 분석하거나, 다수(big)가 가리키는 방향성을 따르다 보면 때로 잘못된 해석이나 결론에 이를 수 있다.

기계가 사람의 언어나 감정 등을 이해하는 것은 쉬운 일이 아니다. 효율성, 효용성을 근거로 기계에 모든 해석을 미루어두는 것은 오히려 오독을 심화시키는 것일 수 있다. 빅데이터를 통해 "예측을 위한 모델링을 하려는 모든 시도들은 일반적 원칙을 수립해야 하고 확보한 데이터에서 잡음을 걸러내야 하는 핵심적 도전과제에 직면"[23]해야 하지만, 어떤 것이 잡음이고, 잡음을 어떻게 걸러내

---

20) 빅토르 마이어 쇤버거 · 케네스 쿠키어, 이지연 옮김, 『빅데이터가 만드는 세상』, 21세기 북스, 2015. p.209.

21) "트위터와 시(詩)의 닮은 점을 살펴봤지만, 결국 둘은 같다고 말할 수 없다. 시에는 행간이라는 것이 있다. 우리가 어떤 시를 처음 봤을 때는 행을 읽지만, 몇 번을 읽은 뒤부터는 행간을 봐야 하고, 그 다음에는 자신의 마음을 봐야 한다. 좋은 시일수록 행간은 늪이 된다. 그 행간에 점점 빠져들수록 반대로 독자의 마음속에는 무언가가 떠오른다. 시는 독자의 마음에 돌기를 심어 넘어지게끔 하는 것이고, 말뚝을 박아 그 사유와 감성의 늪에서 빠져 나오지 못하게 하는 것이다." 편석준, 『구글이 달로 가는 길』, 레드우드, 2015, p.186.

22) 이오준 · 박승보 · 정다울 · 유은순, 「소셜 빅데이터를 이용한 영화 흥행 요인 분석」, 『한국콘텐츠학회논문지』 14-10, 2014, p.530. 빅데이터 분석을 통해, 영화흥행 요소를 분석했던 위의 연구자들은 같은 논문에서 다음과 같이 밝혔다. "지금까지 진행된 국내외 선행연구들을 살펴보면 영화에 대한 관객의 선호도와 흥행의 추이를 예측하기 위해 SNS의 비정형 데이터를 분석하여 감성의 극성을 분류하거나 데이터의 사계열적 변화를 추적하는 정량적 관점이 주를 이루고 있다. 하지만 언어 표현의 풍부함과 복잡한 내재적 속성 때문에 텍스트가 함축하고 있는 감성을 정확하게 식별하고 극성 값을 부여해야 하는 감성 분석은 매우 어려운 작업이다." 같은 논문, p.530.

야 하는가는 여전히 요원한 문제로 남는다.

빅데이터는 기존의 샘플조사의 한계를 뛰어 넘은 전수조사이기 때문에 더욱 과학적이고 실수가 없는 것처럼 보이기도 한다. 그러나 빅데이터 분석은 질문(질의어)을 던지는 과정에서 전수조사의 의미를 상실하고, 분석가의 의도에 따른 결과를 도출하는 도구로 악용될 수 있다. 또한 전문가들이 지적하는 것처럼 기계가 인간의 언어를 이해하고 해석하는 것은 여전히 쉽지 않은 문제이다. 그것은 여전히 기계적인 분석을 벗어나지 못하기 때문이다. 인간의 언어는 복잡하고 다층적이다. 인간의 언어를 분석하고 의미를 정확하게 해석하는 데 있어 인문학적 접근과 해석은 불가결한 요소로 보인다. 빅데이터 사용에 있어 해석과 분석이 중요한 문제라고 한다면, 인문학의 참여와 인문학적 접근은 보다 신뢰할 만한 해석과 분석으로 나아가는 데 유의미한 역할을 하게 될 것이다.

2) 조작의 위험성

> "세상에는 세 가지 종류의 거짓말이 있다. 거짓말, 새빨간 거짓말, 그리고 통계이다."
>
> 마크 트웨인

마크 트웨인이 비판한 통계는 샘플조사를 기반으로 한 기존의 통계방식이지만, 빅데이터 역시 이러한 염려에서 벗어나기는 어렵다. 무엇보다 우려되는 것은 빅데이터를 사용자에 대한 윤리의식을 강제할 수 있는 방법이 쉽지 않아 보인다는 점이다. 빅데이터 사용자에게 윤리와 도적적 감수성을 요구할 수 있지만, 데이터 조작은 발견하기 쉬운 문제가 아니다. 데이터 조작, 또는 데이터 분석 조작은 현재 데이터분석가가 적기 때문이라는 현실적 문제도

23) 에릭 시겔, 고한석 옮김, 위의 책, p.227.

있지만, 빅데이터 분석의 조작을 사전에 감지하거나 사후에 검증하는 것이 어렵기 때문이기도 하다.[24] 데이터조작 혹은 데이터분석조작은 개인이나 기업의 소기 목적달성이나 이익추구와 밀접하게 연관되는 경우가 많다.[25] 이는 개인의 피해뿐만 아니라 사회적 혼란과 갈등의 유발로 이어질 수 있다. 결국 공동체 내에서의 신뢰붕괴라는 위험을 초래할 수 있다는 점에서 위협요소로 꼽힌다.

데이터 조작 위험성에 대한 우려를 쉽게 떨쳐낼 수 없는 또 다른 이유는, 빅데이터 사용에 인간의 욕망이 개입되어 있기 때문이다. 빅데이터의 유용성 입증은 기업의 이윤과 밀접하게 맞물려 있다. 이것은 기업의 이윤뿐만 아니라, 인간(단체나 개인)의 이익 추구와 욕망 실현을 위해 비윤리적으로 사용될 수 있다는 것을 의미하고, 데이터의 조작 등의 문제로도 이어질 수 있게 된다. "데이터가 무책임한 사람들의 손에 들어갈 경우, '현실 마이닝'의 힘은 끔찍한 결과를 초래할 수 있다"[26]는 염려는 여기에서 비롯된다.

빅데이터는 일방적인 주장이 아니라 우리가 직면해야 할 현실이고, 그것이야말로 우리가 받아들여야 할 사실이라는 주장도 있다.[27] 그러나 많은 사람들의 생각이 언제나 옳은 것은 아니다. 빅데이터는 말 그대로 '많은'(big) 데이터이다. 우리가 빅데이터에 보내는 무한한 신뢰와 기대는 그것이 엄정한 과학에 기댄 정확한 데이터라고 믿는 데서 출발한다. 많은(big) 데이터, 많은(big) 사람들이 원하는 것이 옳다는 주장은 받아들이기 어렵다. 때로 빅데이터

24) 윤상오, 위의 논문, p.109.

25) "데이터 선별이나 조작 등 일련의 데이터 마사지는 많은 경우, 연구자의 현실적 이익과 깊이 연동되어 있다." 김호연, 「디지털 기술과 인문학의 융합 - 데이터(data) 활용을 중심으로」, 『시민인문학』 30, 2016, p.79. 이어 김호연은 이러한 기만행위가 역사 속에서 부단히 이루어져 왔고, 갈릴레이와 뉴턴의 경우도 여기에 해당한다고 설명하였다. 김호연, 같은 논문, pp.79-80.

26) 빅토르 마이어 쇤버거 · 케네스 쿠키어, 이지연 옮김, 위의 책, p.169.

27) 크리스천 러더, 이가영 옮김, 위의 책, p.180.

의 핵심을 이루고 있는 '많은'(big) 데이터 흐름은 자본에 의해 의도적으로 형성될 수도 있기 때문이다.[28] 특정한 의도와 목적으로 데이터를 발생시키고, 데이터의 흐름을 조작한 사례는 우리의 현실에서도 찾아볼 수 있다. 우리 사회에서 특정 기관의 정책이나 선거와 관련된 댓글 공작, 기업에서의 댓글 조작 등은 유관한 데이터를 증식·노출시킴으로써, 이익을 얻으려는 시도와 연관되어 있다. 고의적인 온라인 댓글 작업을 통해 소기의 목적을 달성하려는 노력은 일부 정치인이나 기업에 국한되지 않는다.

이처럼 국내외의 다양한 사례는 빅데이터에서 도출된 결과가 상대적으로 더 중요하고 유용한 것이라는 공식이 언제나 성립하지 않는다는 것을 보여준다. 자본과 의도가 충분하다면, 빅데이터 조작은 고의적이고 악의적으로 이루어질 수 있다. 인문학은 이에 대해 윤리적 문제를 보다 촘촘하게 제기하고, 법률 등 다른 분야의 전문가들과 함께 이러한 문제를 해소할 수 있는 방법을 모색하며, 빅데이터의 사용이 보다 바람직한 흐름으로 이어지도록 건강한 방향성을 제시해야 한다.

4. 인간의 생명과 존엄성에 관한 문제

지금까지 빅데이터는 기존의 샘플 조사로 어려운 문제들을 예측

28) 미국에서 자신의 세를 과시하기 위해 팔로워를 구입했던 사실이 그 증거가 될 수 있다. "공화당 공천 후보가 결정되기 전에 뉴트 깅리치(Newt Gingrich)는 '제 트위터 팔로워 수가 다른 후보들의 팔로워를 합친 것보다 여섯 배가 많습니다.'라고 거들먹거렸다. 사실 깅리치는 이 가운데 90퍼센트를 돈으로 매수했다. 미트 롬니도 팔로워를 샀다. 7월 어느 날, 몇 분 만에 롬니의 팔로워가 2만 명 더 늘었다. 이는 그 시각 전까지 그가 얻은 팔로워보다 20배가 많은 숫자다."(크리스천 러더, 위의 책, pp.284-285) 이런 고의적 조작에 의해 형성된 갑작스러운 팔로워 증가를 21세기 워터게이트로 볼 수 있다는 의견은 과장이 아니다.

하고 성과를 보여 왔다. 빅데이터의 사용을 주장하는 사람들은 "[빅데이터는] 사람들의 소비 활동을 예측한다. 사랑하고, 일하고, 출산하고, 이혼할 확률을 예측한다. 투표 성향, 판단, 의사결정을 예측한다. 이직하고, 그만두고, 갈아탈 확률을 예측한다. 누가 사고를 낼 것인지 예측한다. 누가 병들고, 누가 죽을지 예측한다. 누가 거짓말하고, 속이고, 훔치고, 살인할지 예측한다."[29]고 말한다. 어떤 사람이 어떤 성향을 가졌고, 이직할 것이며, 물건을 구매할 것인지에 대한 예측은 가능해보인다. 그러나 누가 사고를 낼 것이고, 누가 죽을 것이며, 누가 훔치고 살인할 것인지에 대한 예측은 그 신뢰성과 효율성에도 불구하고 위험성을 내포하게 된다.

1) 범죄·치안 영역에서 빅데이터 사용의 위험성

인간의 생명과 존엄성과 관련된 영역에서의 빅데이터 활용은 양날의 검이라고 할 수 있다. 개인정보 유출 또는 프라이버시 침해 문제에 대해서는 비판적인 입장이 우세하지만, 범죄와 치안, 질병관리, 공중보건 등 공공 영역에서 빅데이터 활용이 공적 가치를 증진시키는 데 기여할 것이라고 판단하는 경우가 적지 않기 때문이다. 그러나 이들 중에서도 범죄예방, 질병관리는 인간의 기본권, 존엄권, 생명권을 직접적으로 다룬다는 점에서, 신중하게 다루어질 필요가 있다. 현재, 범죄·치안 관련된 영역에서의 빅데이터 활용은 긍정적인 측면이 강조되어 나타나는 경향이 있다. 흔히 접할 수 있는 기사의 구체적인 내용은 다음과 같다.

> "싱가포르와 미국의 경우 테러를 막고 범죄자를 색출하기 위해 빅데이터를 활용한다."[30]

---

29) 에릭 시겔, 고한석 옮김, 위의 책, pp.29-36.

30) 「[잇츠IT] 규제 완화 vs 강화, 빅데이터 논란」(『머니S』(2017/09/14)

"[프레드폴(Predpol)은] 범죄가 일어날 가능성이 있는 곳을 알려주는 것이다. 프레드폴을 사용하고 있는 영국의 켄트경찰서는 '어느 지역을 순찰해야 하는지를 알려주는 것은 아주 중요하다'며 '프레드폴 도입 이후 14만 건에 이르렀던 범죄가 10만 건으로 줄었다'고 설명했다."[31]

"2011년 7월 미국 캘리포니아주 샌타크루즈 섬의 주차장에서 '예비 자동차 절도범' 두 명이 체포됐다. 자동차 안을 엿보던 모습이 '때마침' 현장에 있던 경찰관들에게 적발됐다. 이들을 수색하던 경찰관이 주머니에서 마약도 발견했다. 이들은 곧바로 구금됐다. 당시는 샌타크루즈 경찰이 범죄 예측 프로그램인 '프레드폴'(PredPol·예측 치안이라는 의미의 Predictive Policing의 준말)을 도입한 직후였다. 경찰은 '그 장소에서 차량 절도가 발생할 것으로 예상된다'는 프레드폴의 예측으로 현장에 미리 대기하고 있었다. 이 사건을 보도한 뉴욕타임스(NYT)는 '경찰이 예지력을 갖는 것은 이제 평범한 일이 됐다'고 평했다."[32]

빅데이터를 통해 인간이 '예지력'을 가지게 되었다는 건조한 기사는, 빅데이터라는 '믿을 만한' 첨단 과학을 통해 미래를 통제할 수 있게 되었다는 기대로 읽힌다. 이어지는 구체적 사례와, 명확한 숫자를 동반한 효과는 신뢰할 만한 근거로 제시되는데, 이는 마치 영화 〈마이너리티 리포트〉(2002)의 '프리 크라임'(pre-crime) 시스템이 현실화 된 것처럼 보이기도 한다.

범죄 예측 시스템은 '프레드폴' 이전에도 존재해 왔다. 범죄 예측 시스템을 도입한 캘리포니아 주 산타크루스 시는 이 시스템을 이용해 강도 사건 가운데 25%를 정확하게 예측하였다고 발표하였고, 『타임』지는 이 시스템을 2011년을 빛낼 50대 발명 가운데 하

---

http://moneys.mt.co.kr/news/mwView.php?no=2017082413008076163&outlink=1

31) 「프리크라임(Pre-Crime)…현실화 된다」(『아시아경제』(2015/3/12) http://www.asiae.co.kr/news/view.htm?idxno=2015031210072221504

32) 「미국 프레드폴, 예비 범죄자 잡아내 … 1년 새 강도 27% 줄어」(『중앙일보』(2017/12/9)) http://news.joins.com/article/22189106

나로 선정하였다.[33] 미국의 오리건 주에서는 이 시스템이 범죄인에게 형량을 정하는 데 있어 참고대상이 되고 있다.[34] 예비 살인자를 예측하는 것은 영화 〈마이너리티 리포트〉에 묘사된 '프리 크라임' 시스템의 재현이라고 할 수 있다. 한 사람에 대한 분석이 아니라, 빅데이터 분석에 의해 판결에 참고하거나 예비범죄자를 걸러내는 데 사용하는 것은 '가능성'이 아니라, 이미 현실에서 작동되거나 현실화가 가능하도록 연구·고안되고 있는 것이다.[35]

그러나 빅데이터를 통한 범죄 통계분석은 효과적일 수는 있어도 모든 범죄를 예측할 수는 없고, 오히려 여기에서 또 다른 문제들이 발생하게 된다.[36] 범죄를 짓지 않은 사람들이 예비 범죄자 군

33) 에릭 시겔, 고한석 옮김, 위의 책, pp,107-108.

34) "오리건 주에서는 배심원단으로부터 유죄판결을 받은 중범죄인에게 판사가 형량을 선고할 때 도움을 받을 수 있는 범죄 예측 시스템을 도입하였다. (……) 오리건 주의 재범 위험성 예측 모델은 매우 훌륭한 성과를 거두고 있다. 기계 학습 시스템은 지난 5년간 오리건 주의 범죄자 55,000명에 대한 기록을 분석하여 이 모델을 구축하였다. 이 모델은 지난 30년간의 350,000명의 범죄자 기록에 대한 백테스트를 통해서 검증되었다." 에릭 시겔, 고한석 옮김, 위의 책, pp.124-125.

35) 이러한 연구는 우리나라에서도 급격히 증가하고 있는 추세이다. 김성준, 「한국에서 빅데이터를 활용한 범죄예방시스템 구축을 위한 연구」, 『한국인터넷방송통신학회 논문지』 17-5, 2017, pp.217-221; 최민제·노규성, 「빅데이터 융합 기반 범죄예방에 관한 탐색적 연구 - 성남시 사례 분석을 통해」, 『디지털융복합연구』 14-11, 2016, pp.125-133; 권양섭, 「범죄예방과 수사에 있어서 빅데이터 활용과 한계에 관한 연구」, 『법학연구』 65, 2017, pp.179-198; 김영진, 「빅데이터 활용과 인권 보호에 관한 시론적 연구」, 『법과정책』 23-1, 2017, pp.33-59. 등. 최근 범죄 분야에서 빅데이터를 활용하자는 연구는, 많은 경우 방법과 기술, 효용성 입증에 기울어진 측면이 있다.

36) "영국의 더럼대학교 데이비드 월 범죄학 교수는 '범죄 통계분석 기술은 효과적일 수 있는데 모든 범죄를 예측해낼 수는 없다'고 한계를 지적했다. 범죄예측시스템에 대한 전 세계적 관심이 확대되면서 동시에 개인 정보 침해는 물론 그 부작용에 대한 우려도 높아지고 있다." 「프리크라임(Pre-Crime)…현실화 된다」, 『아시아경제』(2015/3/12) http://www.asiae.co.kr/news/view.htm?idxno=2015031210072221504

으로 분류된다는 점에서, 인간의 기본권 침해의 문제를 피할 수 없게 된다. 반체제 인사나 양심수 등에 대한 기본권 침해는 이미 심각하게 이루어지고 있는 현실이기도 하다.[37]

더욱 큰 문제는 빅데이터 해석에 있어 인간의 편견과 의도가 개입할 수 있다는 점이다. 캐시 오닐은 빅데이터와 알고리즘, 이를 기반으로 한 프레드폴의 위험성에 대해 지적한다. 프레드폴은 범죄 유형, 발생 장소와 시점이라는 핵심데이터를 기반으로 하는데, "프레드폴 시스템을 적용할 때, 경찰들에게 선택권이 주어"지고, 경찰은 1군 범죄 뿐만 아니라, 경미한 2군 범죄까지 범죄의 범위에 넣을 수 있다. 캐시 오닐에 따르면 가난한 동네에서 종종 일어나는 2군 범죄는 반사회적 행동(antisocial behavior)이라고 부를 만한 경미한 것이므로, 경찰의 자의적 선택에 따라 선택되는 2군 범죄는 오히려 분석을 왜곡시킬 위험이 있다. 이렇게 적발된 정보는 프레드폴에 입력되어 데이터화 되고, 다시 경찰력의 증대로 이어진다. 캐시 오닐은 이것을 유해한 피드백 루프가 활성화되는 과정이라고 말한다. 또한 이러한 데이터는 피해자 없는 범죄자들, 즉 대부분 가난한 지역에 사는 흑인이나 히스패닉계를 교도소에 보내는 불합리한 결과로 이어진다고 지적한다.[38]

과학적이고 체계적인 빅데이터와 알고리즘조차 인간의 편견이나 선입관을 전혀 배제할 수 없고, 그럴 경우 인권침해라는 심각한 부작용을 피할 수 없다는 결론에 이르게 된다. 빅데이터는 그것의 사용이 인간의 삶을 개선하고 고양시킬 수 있다는 점에서 긍정적으로 받아들여지고 있으나, 범죄를 사전에 차단한다는 이유로 인

37) "국제인권단체 휴먼라이츠워치(HRW)가 지난달 낸 보고서에 따르면 중국 정부는 반체제 인사나 소수민족과 관련된 의료 기록, 피임 방법, 수퍼마켓 배달 기록 등 거의 모든 개인정보를 수집해 행동 방식을 예측하고 있다." 「미국 프레드폴, 예비 범죄자 잡아내 … 1년 새 강도 27% 줄어」, 『중앙일보』(2017/12/9). http://news.joins.com/article/22189106

38) 캐시 오닐, 김정혜 옮김, 『대량살상 수학무기』, 흐름출판, 2017, pp.151-152.

권 훼손이 공공연하게 일어나는 현실에 대해서는 진지한 논의가 필요하다. 인문학적 비판이 질문이 제기되지 않는다면, 빅데이터에 근거한 '예비' 범죄자에 대한 단죄는 새로운 우생학이 될 수도 있을 것이다.[39]

2) 의료 영역에서 빅데이터 사용의 위험성

빅데이터 사용이 신중하게 고려되어야 하는 또 다른 영역은 의료 분야이다. 다른 위험 요소와 비교해 보았을 때, 의료 분야에서의 빅데이터 활용은 긍정적으로 검토되어 왔던 것이 사실이다. 기대수명이 연장되고 고령화 되는 사회에서, 빅데이터는 의료 영역에서의 비용절감이나, 전염병 확산에 대한 적절한 대응, 신약 개발 등과 연결되며 긍정적으로 받아들여져 왔다.

> "공공부문 중에서도 의료분야의 빅데이터 활용은 비용절감이라는 이슈를 함께 제기하는데, 이는 대부분의 국가에서 의료보장개혁이 막대한 국가예산을 필요로 하는 정책과제 중 하나이기 때문이다. (……) 빅데이터 활용을 통해 전국적으로 환자 개인 및 환자치료에 대한 데

39) 캐시 오닐은 데이터에 기반하여 인간을 평가하는 것을 '디지털 골상학'이라는 용어로 표현하였다. 캐시 오닐, 김정혜 옮김, 위의 책, p.181.
범죄데이터 사용과는 다소 차이가 있지만, 생체정보를 기반으로 한 데이터는 의료와 치안 영역에서 활용되거나, 사용에 대한 기대를 불러일으키고 있다. "전 세계에서 가장 큰 영국의 범죄자 DNA 데이터뱅크는 유럽사법재판소에 의해 규모가 축소됐지만, DNA 정보가 포함된 국가보건서비스(NHS)의 전자건강기록은 더 큰 보건의료 사안들에 초점을 맞춘 법률의 일부로 포함되면서 아무도 모르는 새 상품화되었다. (……) 열성적 지지자들은 뇌파계와 기능성자기공명영상을 써서 살아 있는 뇌의 활동을 들여다봄으로써 잠재적 사이코패스, 범죄자, 테러리스트가 범죄를 저지르기 전에 찾아낼 수 있다고 주장한다." 힐러리 로즈·스티븐 로즈, 김명진·김동광 옮김, 위의 책, pp.33-34. 범죄데이터가 생체정보를 기반으로 한 생물학적데이터와 결합된다면, 이는 생물학적 근거로 인간을 단죄하고, 차별하는 우생학의 재현에 다름 아닐 것이다.

이터베이스를 구축함으로써 의료종사자 및 정책입안자들로 하여금 전염병과 같이 빠르게 전파되는 병원균에 대한 보다 신속하고 종합적인 대응을 할 수 있게 한다. 또한 광범위한 의료데이터의 집적으로 유전자 연구, 제약연구 등이 더욱 활발하게 이루어진다. 특히 개인의 전자의료 기록은 각 개인의 진료기록, 투약 및 부작용 정보, 예방접종기록, 방사선 관련 의료기록, 가족력, 결제내역 등을 포함하고 있어 의료기관 간 정보를 효율적으로 공유할 수 있는 체계를 제공해준다."[40]

우리나라에서도 공공의료분야에서 빅데이터를 적극 활용하고 있고, 국민건강관리공단에서는 연구자의 신청을 받아, 심의의 과정을 거친 후 연구자에게 데이터를 제공하고 있다.[41] 건강에 관한 개인의 민감정보가 비식별화 또는 익명화처리 후에 사용되기 때문에 문제가 되지 않는다는 주장도 있지만, 비식별화된 정보라 하더라도 다른 정보와의 매칭이 이루어진다면 재식별이 가능하다는 주장이 엇갈리고 있는 현실이다.[42] 이렇게 유출된 민감정보는 개인의 사회생활에 직접적 영향을 줄 수 있는 가능성이 있다는 점에서 문제로 제기되고 있다.[43] 또한 건강 관련 빅데이터가 웨어러블 기기와 결합하여 상용화되었을 때 나타나는 문제도 적지 않다. 사적 기업이 개인의 건강관련정보를 수집·저장할 수 있고, 해킹 피해가 발생할 가능성도 무시할 수 없다.[44] 건강·생명에 관한 문제인

40) 김기환, 「공공부문 빅데이터의 활용성과 위험성」, 『정책분석평가학회보』 23-2, 2013, pp.8-9.

41) https://nhiss.nhis.or.kr/bd/ab/bdaba001cv.do

42) http://www.doctorsnews.co.kr/news/articleView.html?idxno=120109

43) "데이터베이스에 대한 관점은 젠더, 나이, 건강 상태 등에 따라 달라지는 경향이 있다. …… 그들[데이터 제공에 반대하는 사람들]은 유전 정보가 고용주에게 팔릴 경우, 자칫 그 데이터 때문에 취업이 어려워질 수 있다고 생각했다. 어떤 사람들은 보험에 가입하는 데 미칠 수 있는 영향에 대해 우려했다." 힐러리 로즈·스티븐 로즈, 김명진·김동광 옮김, 『급진과학으로 본 유전자 세포 뇌』, 바다출판사, 2015, p.235.

44) 최규진·김병수, 「건강관련정보(data concerning health)의 활용과 문제점」,

만큼 발생하게 될 문제도 심각하고, 폐해도 적지 않을 것으로 예상된다.

이 외에도 의료 진료를 위한 의사결정 지원도구 프로그램인 APACHE-3에 의해 제시되는 생존확률 수치는 인간의 생명을 포기하게 만드는 근거가 되기도 한다.[45] 빅데이터를 기반으로 한 인공지능 의사 왓슨(Watson)의 사용도 같은 맥락에서 주의를 요한다. 빅데이터에 기반한 왓슨에 의한 오진 및 오류는 치명적 결과로 이어질 수 있는데, 더 큰 문제는 이때 발생하는 문제의 법적 책임소재를 분명히 할 수 없다는 점에 있다.[46] 빅데이터의 제시에 따라 인간 생명에 대한 판단이 내려지는 것에는 주의가 필요하다. 지금까지 인류가 질병, 고통에 대항하며 지탱시켜 온 가능성, 기적, 희망 등은 무용한 것으로 전락될 수 있다는 우려 또한 지울 수 없다.

빅데이터가 의학의 진보를 가져올 것이라는 기대도 있지만, 상

『생명윤리』 18-1, 2017, p.11.

45) "APACHE-3는 의료진료를 위한 의사결정 지원도구 프로그램으로 자칫하면 죽을 수도 있는 중환자들의 위험을 예측하기 위해 만들어졌다. APACHE-1은 1981년 조지 워싱턴 대학병원의 의사들에 의해 만들어졌는데, 중환자에 대한 치료 경험을 전산화한 것에서부터 비롯됐다. 이후, APACHE-3까지 개발됐는데, 여기에는 60만 명 이상의 중환자에 관한 DB가 들어 있다. APACHE-3는 실제 치료와 치료 절차를 개선해 병의 회복과 비용 절감에 많은 도움을 줄 수 있다. 하지만 의사들에게 APACHE-3의 말은 권고로 시작했다가 나중에 지시와 명령이 될 수도 있다." 편석준 · 이정용 · 고광석 · 김준섭 지음, 『사물인터넷: 실천과 상상력』, 미래의 창, 2015 pp.60-61.

46) "인공지능의 인격을 인정한다면 형사책임에 있어서 양벌규정을 통하여 자연인 외에도 책임을 추궁할 가능성도 존재할 것이다. (……) 이러한 인공지능에 대하여 독자적인 인격으로 대우할 필요성이 적으며, 우리나라 현행법상 이를 인정하는 명문의 규정도 존재하지 않으므로, 이를 인정하는 것은 현시점에서는 어렵다고 보인다." 장연화 · 백경희, 「왓슨의 진단 조력에 대한 현행법상 형사책임에 관한 소고」, 『형사법의 신동향』 55, 2017, p.329.

대적으로 가치를 인정받기 어려운 마이너리티 데이터에 대한 적은 관심은 오히려 의학의 진보를 더디게 하는 요소가 될 수도 있을 것이다. 의학의 발전을 추동한 요소 가운데 하나는 '빅'의 흐름이 보여주는 규칙이 아니라, 예외들이었다.[47] 빅데이터의 활용은 인간의 해석과 직관을 축소시킬 수 있고, 이것이 오히려 의학의 발전을 저해하는 요소가 될 수도 있을 것이다.

의료영역에서의 빅데이터 활용은 의료비용의 손실 감소, 사회적 이익의 증대라는 목적과는 부합해 보이지만, 이러한 흐름 속에서 개인정보나 프라이버시, 존엄성 등의 문제는 상대적으로 가볍게 다루어져 온 것이 사실이다. 개인의 생체정보(민감정보) 활용에 대한 절차를 투명하게 하여, 향후 발생할 문제를 사전에 차단하는 적절한 방식도 진지하게 고려될 필요가 있다. 인문학은 새로운 의료윤리의 제시를 제시하고, 각 분야의 전문가와 공조하여 옵트아웃(opt-out) 제도의 도입과 시행 등 개인정보 보호와 의료데이터의 활용이 적절히 균형을 잡을 수 있도록 비판과 의문을 제기하는 역할을 담당해야 할 것이다.

빅데이터는 공적 영역에서의 활용을 넘어 거대화, 상업화 되고 있다. 비용을 절감하는 시스템과, 인간의 존엄과 가치를 보호하고 향상시키는 시스템은 같은 층차에서 논의되기 어렵다. 빅데이터를 이용하려는 목적은 그것이 이롭기 때문이다. 하지만, 과연 무엇을 그리고 누구를 위한 이로움인가? 인간의 삶과 존엄성이라는 맥락을 거세당한 빅데이터의 이용은 자본의 증식과 경제적 이익이라는 귀결로 수렴될 것이다.

---

47) "'정상적인 것'들은 규칙을 가르쳐준다. 하지만 법칙을 가르쳐주는 것은 '예외들'이다." 싯다르타 무케르지, 『의학의 법칙』, 문학동네, 2017, p.59.

5. 맺음말

많은 데이터의 수집, 마이닝, 해석과 분석을 거쳐 비즈니스 모델링으로 이어지거나 공공부문에 활용되어온 빅데이터는 우리의 삶을 바꾸어놓을 획기적인 과학으로 다가오고 있다. 빅데이터에 대한 수요와 전망은 우리 사회가 빅데이터에 거는 기대를 읽게 한다. 빅데이터는 많은 양의 데이터에 수학을 적용해 확률을 추론하는 것이고,[48] 이 효용성은 이미 다양한 분야에 적용되어 증명된 바 있다. 빅데이터는 개인과 기업, 공공분야의 미래를 예측하여, 더 나은 방향으로 나아가게 하는 데 분명 유용한 해답을 줄 수 있다. 그러나 이 지점에서 빅데이터에 대한 깊은 믿음을 되돌아볼 필요가 있다. '많은'(big) 것이 확실하고 의미 있는 것인가? 또 '많은' 것에서 도출한 해석이 가장 질적으로 정확하고 가치 있는 해석인가에 대한 질문이 필요하다.

크리스천 러더는 "한 사람의 이야기와 우리 모두의 이야기는 다르다. 집합 데이터가 필요한 이유는 바로 개인의 일화로는 절대 승자를 가릴 수 없는 논쟁의 답을 준다는 데 있다. 데이터는 주장이 아니라 우리가 직면해야 할 사실을 보여준다."[49]고 말하면서, 한 사람의 개별적 사례가 아니라, 다수가 가리키는 흐름이야말로 우리가 직면해야 할 현실이라고 강조하였다. 그러나 빅데이터가 커다란(big) 흐름과 방향을 보여준다고 말할 수 있지만, 빅(big)이 항상 '옳음'이나 '높은 가치'와 등가적 가치를 가진다고 말하기는 곤란하다. 설령 정확하다 하더라도, 그것이 인류와 세계, 생태를 위한 것인지에 대한 되물음은 더욱 필요하다. 인간적 가치가 배제된 것이라면 재고되어야 한다.

---

48) 빅토르 마이어 쇤버거 · 케네스 쿠키어, 이지연 옮김, 위의 책, p.27.
49) 크리스천 러더, 이가영 옮김, 위의 책, p.180.

데이터의 수집보다 중요한 것은 사용이고, 이때 필요한 것이 비판적 해석과 인문학적 상상력, 끊임없는 큰(big) 되물음이다. 빅데이터에 대한 맹신에 제동을 걸고, 그것의 의미를 되살피는 것도 필요하다. 빅데이터는 인간처럼 생각하거나 성찰하지 않기 때문이다. 빅데이터와 알고리즘을 맹신하는 흐름에 "데이터 처리 과정은 과거를 코드화할 뿐, 미래를 창조하지 않는다. 미래를 창조하려면 도덕적 상상력이 필요하다. 그런 능력은 오직 인간만이 가지고 있다. 우리는 더 나은 가치를 알고리즘에 명백히 포함시키고, 우리의 윤리적 지표를 따르는 빅데이터 모형을 창조해야 한다. 그렇게 하려면 가끔은 이익보다 공정성을 우선시해야 한다"[50]는 캐시 오닐의 비판은 무겁고 진지하게 다가온다.

빅데이터가 인간과 삶, 인류의 미래를 위해 사용되는 것이라면, 다수(big)의 흐름에 앞서 '인간 삶'이라는 기본 전제 위에서 의심하고 비판하며, 건강한 방향성을 설정하려는 인문학적 성찰이 누락되어서는 곤란하다. 빅데이터를 활용해야 한다는 목소리는 점점 높아질 것이다. 빅데이터가 창출해온 가치, 기대되는 가치에 대한 환호에 앞서, 그것이 과연 무엇을 위해 존재하는지에 대해 인문학이 던지는 근원적이고 큰 질문, 즉 빅퀘스천이 필요하다.

참고문헌

1. 기본자료

「프리크라임(Pre-Crime)… 현실화 된다」, 『아시아경제』(2015/3/12)http://www.asiae.co. kr/news/view.htm?idxno=2015031210072221504

「무궁무진한 공공데이터 활용해 '금맥' 캐는 시대 열렸다」, 『동아일보』(2017/08/07)) http://news.donga.com/3/all/20170807/85707589/1

50) 캐시 오닐, 김정혜 옮김, 위의 책, p.337.

「보건의료 빅데이터, 안심해도 되나」, 『의협신문』(2017/11/27)) http://www.doctorsnews.co.kr/news/articleView.html?idxno=120109
「일자리 금맥 빅데이터시장 급성장…2019년 490억 달러 예상」, 『매일경제』(2017/10/10))http://news.mk.co.kr/newsRead.php?year=2017&no=668918
「한국은 빅데이터 金鑛(금광 · goldmine)인데… 제대로 캐질 못해」, 『조선비즈』(2015/10/14) http://srchdb1.chosun.com/pdf/i_service/pdf_ReadBody.jsp?Y=2015&M=10&D=14&ID=2015101400058
「[글로벌기업, 그들은 한국에 무엇인가] 4차 산업혁명 원유 '데이터' 쓸어 담는 그들」, 『전자신문』(2017/12/26) http://www.etnews.com/20171226000244
「[잇츠IT] 규제 완화 vs 강화, 빅데이터 논란」, 『머니S』(2017/09/14) http://moneys.mt.co.kr/news/mwView.php?no=2017082413008076163&outlink=1
「미국 프레드폴, 예비 범죄자 잡아내 … 1년 새 강도 27% 줄어」, 『중앙일보』(2017/12/9)http://news.joins.com/article/22189106
「빅데이터 공개에 따른 개인정보 유출 우려…'비식별화 기술이 관건'」, 『스페셜경제』(2016/04/04) http://www.speconomy.com/news/articleView.html?idxno=66813
http://www.doctorsnews.co.kr/news/articleView.html?idxno=120109
https://nhiss.nhis.or.kr/bd/ab/bdaba001cv.do

2. 논문

Fred H. Cate, 손금주 · 이영조 옮김, 「빅데이터 활용에 있어서의 프라이버시 보호」, 『 경제규제와 법』 8-1, 2015, pp.21-33.
권양섭, 「범죄예방과 수사에 있어서 빅데이터 활용과 한계에 관한 연구」, 『법학연구』 65, 2017, pp.179-198.
김기환, 「공공부문 빅데이터의 활용성과 위험성」, 『정책분석평가학회보』 23-2, 2013, pp.1-27.
김미영 · 이정훈, 「보안강화를 통한 개인정보보호법상의 옵트아웃제 도입 방안 제언-산업보안과 빅데이터활용산업 발전의 교집합을 위하여」, 『한국사회안전학지』(11-2), 2016, pp.113-139.
김성준, 「한국에서 빅데이터를 활용한 범죄예방시스템 구축을 위한 연구」, 『한국인터넷방송통신학회 논문지』(17-5), 2017.

김영진, 「빅데이터 활용과 인권 보호에 관한 시론적 연구」, 『법과정책』 23-1, 2017, pp.33-59.
김호연, 「디지털 기술과 인문학의 융합 - 데이터(data)활용을 중심으로」, 『시민인문학』 30, 2016, pp.70-95.
문상호, 「엔그램 뷰어를 이용한 인문학의 빅데이타 사례 연구」, 『예술인문사회융합멀티미디어논문지』 5-6, 2015, pp.57-65.
박정식 · 손정훈, 「빅데이터를 활용한 대중의 인문학 인식 연구」, 『글로벌문화콘텐츠』 25, 2016, pp.65-80.
소원근 · 김하균, 「사물인터넷 시대의 개인정보과잉이 정보프라이버시 보호반응에 미치는 영향」, 『경영과 정보연구』 36-1, pp.81-94.
유강하, 「빅데이터와 사물인터넷 시대의 비판적 해석과 인문학적 상상력-영화 〈마이너리티 리포트〉를 중심으로」, 『시민인문학』 30, 2016, pp.96-116.
윤상오, 「빅데이터의 위험유형 분류에 관한 연구」, 『한국지역정보화학회지』 16-2, 2013, pp.93-122.
이대희, 「빅데이터와 개인정보 보호: 통지와 동의의 원칙을 중심으로」, 『정보법학』 19-2, 2015, pp.127-158.
이미경 · 최인호 · 정세훈, 「기업 페이스북에서 긍정 · 부정 댓글이 소비자들의 태도, 규범 및 구매행동에 미치는 효과-자기감시성의 조절효과」, 『韓國言論學報』 57-4, 2013, pp.51-71.
이오준 · 박승보 · 정다울 · 유은순, 「소셜 빅데이터를 이용한 영화 흥행 요인 분석」, 『한국콘텐츠학회논문지』 14-10, 2014, pp.527-538.
장연화 · 백경희, 「왓슨의 진단 조력에 대한 현행법상 형사책임에 관한 소고」, 『형사법의 신동향』 55, 2017, pp.316-3346.
차상육, 「빅데이터 환경과 프라이버시의 보호」, 『IT와 法연구』 8, 2014, pp.193-159.
최규진 · 김병수, 「건강관련정보(data concerning health)의 활용과 문제점」, 『생명윤리』 18-1, 2017, pp.1-13.
최민제· 노규성, 「빅데이터 융합 기반 범죄예방에 관한 탐색적 연구 - 성남시 사례 분석을 통해」, 『디지털융복합연구』 14-11, 2016, pp.125-133.

3. 단행본

김기봉 외, 『빅데이터와 사물인터넷 시대 인문학의 재구성 방안 연구』, 경제 · 인문사회연구회, 2015.
빅토르 마이어 쇤버거 · 케네스 쿠이어, 이지연 옮김, 『빅데이터가 만드는 세상』, 21세기 북스, 2015.

싯다르타 무케르지, 강병철 옮김, 『의학의 법칙들』, 문학동네, 2017.
에릭 시겔, 고한석 옮김, 『빅데이터의 다음 단계는 예측분석이다』, 이지스퍼블리싱, 2016.
캐시 오닐, 김정혜 옮김, 『대량살상 수학무기』, 흐름출판, 2017.
크리스천 러더, 이가영 옮김, 『빅데이터 인간을 해석하다』, 다른, 2015.
패트릭 터커, 이은경 옮김, 『네이키드 퓨처 – 당신의 모든 움직임을 예측하는 사물인터넷의 기회와 위협』, 와이즈베리, 2014.
편석준, 『구글이 달로 가는 길』, 레드우드, 2015.
편석준 · 이정용 · 고광석 · 김준섭 지음, 『사물인터넷: 실천과 상상력』, 미래의창, 2015.

(**투고일**: 2018. 2. 19. **심사완료일**: 2018. 3. 15. **게재확정일**: 2018. 3. 21.)

유강하
소속: 강원대학교 인문과학연구소
주소: (24341) 강원도 춘천시 강원대학길1 인문대학 2-309
전자우편: jiangxia430@gmail.com

[Abstract]

Big Data and Big Question
:Humanistic Criticism and Question in Using Big Data

Yu, Kang-Ha(Kangwon National University)

Big data, Internet of Things, and Artificial Intelligence are important topics in modern society. So far, big data has produced useful results in modern society. The prospect that big data can predict the future leads to active utilization of big data. However, there are problems in using big data. Big data has been actively utilized mainly in profit-making or profit maximization. In this process, the basic philosophy for 'human being' is collapsing. Humanities are important because they provide moral sensitivity. It seems that humanities reflection and questions are needed to solve the following three problems. (1) personal information leakage and privacy invasion (2) misreading and rigging of data (3) human life and dignity. Problems arising from the use of big data should be treated seriously in terms of issues of human, life, human individuality, and dignity.

People believe that big data will lead humans to the perfect future, but big data only tells the past and present. Although big data can present the direction of the future, but But it is not synonymous with human-being's bright future. If big data is used for human-being, life and future, humanistic questions should be asked. Before expecting the value created by big data, we have to ask humanistic questions about why big data exists.

Key words : big data, big question, algorithm, misreading, rigging, personal information, privacy, human dignity