Hanbit eBook

Realtime 15

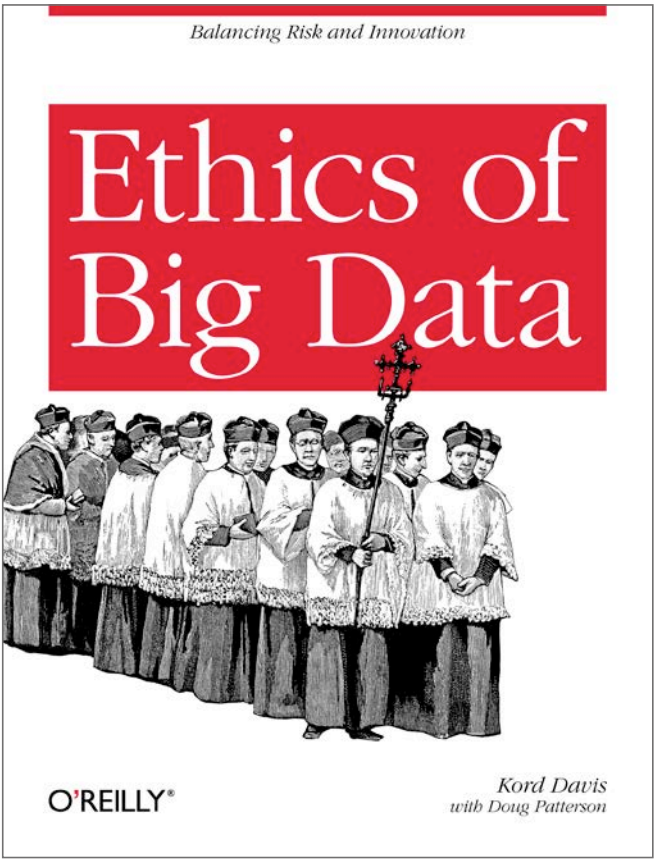
윤리적인 빅데이터 사용을 위한 정책 가이드

빅데이터 윤리

Ethics of Big Data

코드 데이비스 지음 / 강석주 옮김

O'REILLY® 한빛미디어 Hanbit Media, Inc.



이 도서는 O'REILLY의
Ethics of Big Data의
번역서입니다.

윤리적인 빅데이터 사용을 위한 정책 가이드

빅데이터 윤리

윤리적인 빅데이터 사용을 위한 정책 가이드 **빅데이터 윤리**

초판발행 2013년 3월 11일

지은이 코드 데이비스 / **옮긴이** 강석주 / **펴낸이** 김태헌
펴낸곳 한빛미디어(주) / **주소** 서울시 마포구 양화로 7길 83 한빛미디어(주) IT출판부
전화 02-325-5544 / **팩스** 02-336-7124
등록 1999년 6월 24일 제10-1779호
ISBN 978-89-7914-369-0 15560 / **정가** 9,900원

책임편집 배용석 / **기획** 김창수 / **편집** 이중민
디자인 표지 여동일, 내지 스튜디오 [밈], 조판 김현미
마케팅 박상용, 박주훈, 정민하

이 책에 대한 의견이나 오탈자 및 잘못된 내용에 대한 수정 정보는 한빛미디어(주)의 홈페이지나 아래 이메일로 알려주십시오.
한빛미디어 홈페이지 www.hanb.co.kr / **이메일** ask@hanb.co.kr

Published by HANBIT Media, Inc. Printed in Korea
Copyright ⓒ 2013 HANBIT Media, Inc.
Authorized Korean translation of the English edition of *Ethics of Big Data, ISBN 9781449311797*
ⓒ 2012 Kord Davis and Doug Patterson. This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.
이 책의 저작권은 오라일리 사와 한빛미디어(주)에 있습니다.
저작권법에 의해 보호를 받는 저작물이므로 무단 복제 및 무단 전재를 금합니다.

지금 하지 않으면 할 수 없는 일이 있습니다.
책으로 펴내고 싶은 아이디어나 원고를 메일(ebookwriter@hanb.co.kr)로 보내주세요.
한빛미디어(주)는 여러분의 소중한 경험과 지식을 기다리고 있습니다.지금 하지 않으면 할 수 없는 일이 있습니다.

지은이_ **코드 데이비스**Kord Davis

코드 데이비스는 캡 제미니Cap Gemini의 수석 컨설턴트였으며, 거의 20년 동안 오토 태스크, 마이크로소프트, 인텔, 자비의 성모 수녀회 의료원Sisters of Mercy Healthcare, 나이키, 본빌 전력국Bonnevill Power Administraion, BPA, 북서부 에너지 연합Northwest Energy Alliance, NEEA, 빌과 멜린다 게이츠 재단, 웨스턴 디지털, 플루크, 메릭스, 로드웨이 익스프레스, 가든버거 등 100개가 넘는 다양한 규모의 회사에 비즈니스 전략, 분석, 기술을 컨설팅했다. 정보통신 업계의 경력과 전공한 철학 지식을 바탕으로 열정적인 호기심, 정력적인 분석과 업무를 더 빨리, 보다 쉽고 능숙하게 할 수 있는 방법을 연구하고 있다. 전문적인 교육을 받은 작업 그룹 조력자로, 리드 칼리지Reed College에서 철학을 전공했다. 또한 통신, 시스템 모델링, 기업 변신에 관한 자격증을 보유하고 있다.

옮긴이_ **강석주**

초등학교 때 부모님께서 사주신 Apple II가 아직도 재산 목록 1호다. 날씨가 좋은 날이면 어김없이 한강을 따라 인라인스케이트를 타는 마니아지만, 해가 떨어지면 틈틈이 아두이노와 프로세싱을 이용해 밤새 무언가 뚝딱뚝딱 만들기도 한다. 2년 전부터 아이폰에 사로잡혀 cocos2d for iPhone을 이용한 게임 개발에 심취해있다.

저자 서문

철학과 비즈니스가 언제나 서로 잘 맞는 것은 아니다. 일반적으로 철학은 밝혀낸 진실이 실제로 미치는 영향에 대해 별로 관심이 없는 반면에, 비즈니스는 운용 결과에 대해 많은 관심을 두고 있다.

그리고 윤리는 매우 무거운 단어다. 윤리는 학문으로 자리 잡은 분야지만, 윤리에 대한 선입견 때문에 사람들은 이 주제를 이야기하기 꺼린다. 완전히 이해하지 못한 내용을 이야기하기는 매우 어렵고, 때로는 윤리라는 단어 자체가 '이것은 하고 저것은 하면 안 되고'와 같은 명령이나 판단을 의미하기 때문이다.

이 책에서 이런 어려움은 가능한 한 피하려고 했다. 윤리에 대한 질문이 어렵기 때문이 아니라, 윤리 문제가 철학적 사고와 질문에서 도움을 받으려는 회사에 장벽을 만들 수 있기 때문이다. 윤리적인 질문이 줄 수 있는 도움은 많다. 예를 들어 필자의 접근 방식이 가진 큰 특징은 비즈니스 환경, 시장, 회사, 문화, 지리적인 특징, 조직 규모와 성숙도가 모두 다루기 어려운 복잡하고 다양한 환경에 영향을 준다는 점을 인식하게 하는 것이다. 여러분 환경에 대해서는 여러분이 훨씬 더 익숙하다. 따라서 여러분의 회사에 윤리적인 질문을 던지는 가장 좋은 방법은 여러분이 더 잘 알 것이다.

사람들은 필자에게 "어떻게 철학과를 졸업하고 컨설팅을 할 수 있느냐?"고 질문하곤 한다. 이 질문에 대한 답은 필자가 어떻게 대답해야 할지 알게 될 때마다 계속 바뀌어왔다. 이 답은 일반적으로 철학과 비즈니스의 관계, 구체적으로는 윤리와 빅데이터의 관계와 관련 있다.

할머니가 전자 프로젝트 키트(고급형 세트는 아직도 이베이에서 판매하고 있다!)를 선물해주신 5학년 때부터 필자는 기술에 관심이 있었다. 이때 전선, 배터리, 콘덴

서, 저항을 조합하면 아주 새로운 회로를 만들어 재미있는 기능을 실행할 수 있음을 알게 되었다. 또한 고등학교에서 프로그래밍 수업을 듣고, 거의 5년간 미국 해안 경비대에서 무선 기사로 복무하면서 기술이 가진 매력을 경험하였다. 이때는 전파를 보내면 디지털 정보를 전달할 수 있으며, 트랜지스터와 반도체를 결합하여 강력하고 새로운 정보 처리 및 저장 장치를 만들 수 있다는 사실을 알았다. 그리고 소프트웨어를 이용하여 장치들이 놀라운 일을 할 수 있게 만들 수도 있다는 사실도 알았다.

이쯤 되면 아마 필자가 물리학이나 컴퓨터 과학을 공부하는 편이 좋았을 것으로 생각할 수도 있다. 그러나 두 분야의 장점을 비교했을 때, 필자의 관심을 끈 것은 철학이었다. 철학은 열정적으로 분석하고, 모든 사물의 본질을 파악하며, 과학과 기술이 어떻게 작동하는지 이해할 수 있는 열린 접근법을 제공하기 때문이다. 이 책을 쓰는 동기는 철학을, 구체적으로는 기술 발전을 위한 도구에 윤리적인 질문을 적용하여 우리의 삶을 개선하고 더욱 편하게 만드는 것이다.

빅데이터 시대가 도래했다. 대량의 정보를 생성, 연결, 연관, 결합, 저장, 처리, 보고하는 것이 빅데이터의 본질이라는 점에서 빅데이터는 독특한 기술이다. 그러나 많은 정보(특히 사람과 사람 사이의 행동에 대한 정보)를 수집하는 일은 필자가 빅데이터의 '강제성'이라고 부르는 문제를 일으킨다. 좋든 싫든 빅데이터를 사용해 사회적, 정치적, 법적 권한을 침해하지 않는지, 혹은 심각한 윤리 문제는 없을지 생각하게 한다.

이 문제는 길고 복잡하며 매우 중요한 이야기지만, 아직 사회적으로 많은 관심을 받고 있지 않다. 그리고 비즈니스 환경에 윤리를 적용하는 일에 익숙지 않기 때문

에 관심을 끌기도 어렵다. 그만큼 윤리라는 단어는 매우 무거운 단어다. 따라서 이 책이 여러분과 여러분의 회사가 새로운 관점에서 윤리적인 질문을 명시적으로 도입할 수 있는 능력을 발전시키는 데 도움이 되길 바란다. 이 책에서 설명하는 방법, 개념, 용어는 여러분의 고유한 환경에서 회사를 개선시킬 윤리적인 질문을 적용하는 가장 좋은 방법을 판단하는 데 도움을 줄 것이다.

문제와 답을 이해하는 관점에서 보면 이 주제는 아직 초기 단계다. 우리는 계속 문제를 해결해야 한다(현재로서는 다른 대안이 없다). 역사상 그 누구도 빅데이터가 제공하는 뜻하지 않은 결과물 때문에 위험을 겪거나 해결할 기회를 갖지 못했다. 필자는 이 책이 토론의 일부가 되고 빅데이터가 발전함에 따라 책의 내용도 함께 발전하길 바란다.

한빛 eBook 리얼타임

한빛 eBook 리얼타임은 IT 개발자를 위한 eBook 입니다.

요즘 IT 업계에는 하루가 멀다 하고 수많은 기술이 나타나고 사라져 갑니다. 인터넷을 아무리 뒤져도 조금이나마 정리된 정보를 찾는 것도 쉽지 않습니다. 또한 잘 정리되어 책으로 나오기까지는 오랜 시간이 걸립니다. 어떻게 하면 조금이라도 더 유용한 정보를 빠르게 얻을 수 있을까요? 어떻게 하면 남보다 조금 더 빨리 경험하고 습득한 지식을 공유하고 발전시켜 나갈 수 있을까요? 세상에는 수많은 종이책이 있습니다. 그리고 그 종이책을 그대로 옮긴 전자책도 많습니다. 전자책에는 전자책에 적합한 콘텐츠와 전자책의 특성을 살린 형식이 있다고 생각합니다.

한빛이 지금 생각하고 추구하는, 개발자를 위한 리얼타임 전자책은 이렇습니다.

1. eBook Only - 빠르게 변화하는 IT 기술에 대해 핵심적인 정보를 신속하게 제공합니다.

500페이지 가까운 분량의 잘 정리된 도서(종이책)가 아니라, 핵심적인 내용을 빠르게 전달하기 위해 조금은 거칠지만 100페이지 내외의 전자책 전용으로 개발한 서비스입니다. 독자에게는 새로운 정보를 빨리 얻을 수 있는 기회가 되고, 자신이 먼저 경험한 지식과 정보를 책으로 펴내고 싶지만 너무 바빠서 엄두를 못 내시는 선배, 전문가, 고수분에게는 보다 쉽게 집필하실 기회가 되리라 생각합니다. 또한 새로운 정보와 지식을 빠르게 전달하기 위해 O'Reilly의 전자책 번역 서비스도 준비 중이며, 조만간 선보일 예정입니다.

2. 무료로 업데이트되는, 전자책 전용 서비스입니다.

종이책으로는 기술의 변화 속도를 따라잡기가 쉽지 않습니다. 책이 일정한 분량 이상으로 집필되고 정리되어 나오는 동안 기술은 이미 변해 있습니다. 전자책으로 출간된 이후에도 버전 업을 통해 중요한 기술적 변화가 있거나, 저자(역자)와 독자가 소통하면서 보완되고 발전된 노하우가 정리되면 구매하신 분께 무료로 업데이트해 드립니다.

3. 독자의 편의를 위하여, DRM-Free로 제공합니다.

구매한 전자책을 다양한 IT기기에서 자유롭게 활용하실 수 있도록 DRM-Free PDF 포맷으로 제공합니다. 이는 독자 여러분과 한빛이 생각하고 추구하는 전자책을 만들어 나가기 위해, 독자 여러분이 언제 어디서 어떤 기기를 사용하시더라도 편리하게 전자책을 보실 수 있도록 하기 위함입니다.

4. 전자책 환경을 고려한 최적의 형태와 디자인에 담고자 노력했습니다.

종이책을 그대로 옮겨 놓아 가독성이 떨어지고 읽기 힘든 전자책이 아니라, 전자책의 환경에 가능한 최적화하여 쾌적한 경험을 드리고자 합니다. 링크 등의 기능을 적극적으로 이용할 수 있음은 물론이고 글자 크기나 행간, 여백 등을 전자책에 가장 최적화된 형태로 새롭게 디자인하였습니다.

앞으로도 독자 여러분의 충고에 귀 기울이며 지속해서 발전시켜 나가도록 하겠습니다.

지금 보시는 전자책에 소유권한을 표시한 문구가 없거나 타인의 소유권한을 표시한 문구가 있다면 위법하게 사용하고 계실 가능성이 높습니다. 이 경우 저작권법에 의해 불이익을 받으실 수 있습니다.

다양한 기기에 사용할 수 있습니다. 또한 한빛미디어 사이트에서 구입하신 후에는 횟수에 관계없이 다운받으실 수 있습니다.

한빛미디어 전자책은 인쇄, 검색, 복사하여 붙이기가 가능합니다.

전자책은 오탈자 교정이나 내용의 수정보완이 이뤄지면 업데이트 관련 공지를 이메일로 알려드리며, 구매하신 전자책의 수정본은 무료로 내려받으실 수 있습니다.

이런 특별한 권한은 한빛미디어 사이트에서 구입하신 독자에게만 제공되며, 다른 사람에게 양도나 이전되지 않습니다.

차례

01 **빅데이터, 큰 영향** 1

1.1 빅데이터는 무엇인가? 5
1.2 빅데이터 시대가 우리에게 요구하는 것 8
1.3 빅데이터는 윤리적으로 중립이다 12
1.4 무엇을 해야 하는지 말하는 것이 아니다 14
1.5 중요 개념과 용어 15

02 **가치와 행동** 18

2.1 가치의 표명 19
2.2 가치를 행동으로 바꾸기 21
2.2.1 빅데이터의 4가지 항목: 신원, 프라이버시, 소유권, 평판 22
2.2.2 윤리 질문의 장점 27
2.2.3 가치의 역할 29
2.3 윤리 결정점 31
2.3.1 이 모든 것의 의미는? 35

03 **현재 관행** 38

3.1 기업 정책 요약 39
3.1.1 구입과 판매 40
3.1.2 옵트인과 옵트아웃 42
3.1.3 통합을 통한 연관 43
3.1.4 데이터 소유권 47
3.1.5 가치의 표명 49

3.1.6 윤리 불일치 49
3.1.7 다른 이름의 정책 50
3.1.8 문화적 가치 53
3.1.9 그래서 무엇을 배웠나? 54

04 **가치와 행동의 일치** 56

4.1 방법과 도구 56
4.2 일치 방법론 구조 60
4.2.1 질문 60
4.2.2 분석 64
4.2.3 표명 72
4.2.4 실행 73
4.2.5 가치 페르소나 74
4.2.6 GDM: 사례 연구 77
4.2.7 일치의 장점 81

1 | 빅데이터, 큰 영향

원칙보다 권리를 중시하는 사람은 원칙과 권리 둘 다 잃는다.

– 드와이트 아이젠아워Dwight D. Eisenhower

나의 가치를 알리기 위해 작품을 쓰기로 결심했다.

– 시드니 포이티어Sidney Poitier

타깃[01]도 알고, 애플도 안다. 그리고 엄청난 경제적, 사회적, 정치적 가치를 생산하고 있는 페이스북, 트위터, 익스피디아, 링크드인, 넷플릭스, 전국 각지의 정치 단체 등도 빅데이터 시대가 도래했으며 앞으로도 계속될 것임을 안다. 기업들은 점점 더 빅데이터 기술에 의존하고 있으며, 이런 회사는 전례 없이 늘어나고 있다.

2012년 2월 16일 뉴욕 타임즈에 대형 유통 회사인 타깃이 고객의 임신 시기를 알 수 있다는 기사가 나왔다. 타깃은 이 이야기에 아무런 응답을 하지 않았지만, 어쨌든 이 기사는 보도되었다. 이후 엄청나게 많은 댓글과 후속 뉴스는 타깃이 대응한 방법이 적절한가에서부터 프라이버시 공개에 대한 폭넓은 대중적 관심에 이르기까지 수많은 질문을 쏟아부었다.

2011년 4월 20일, 두 명의 연구자는 아이폰이 각 장비의 위치를 숨겨진 파일에 주기적으로 기록함을 밝혔다. 애플은 그 주장을 인정하지 않았지만, 결과적으로 프라이버시 침해에 대한 걱정을 일으켰던 이유가 파일을 생성하고 저장하는 방법에 원인이 있었음이 밝혀졌다. 또한 이 방법을 사용했다는 사실은 실제 윤리적인 문제에 직접적인 영향을 끼쳤다.

---

01 의류와 일용품을 할인 판매하는 대형 유통 회사

누가 이런 결정에 참여했을까? 엔지니어 혼자 연구실 구석에 앉아 기능을 설계하고 구현했을까? 이 기능이 필요한지 경영진과 폭넓게 이야기는 했을까? 얼마나 상세히 보안 및 다양한 위험에 관해 이야기했을까?

2011년 8월 페이스북 모바일 앱에서 '연락처' 기능을 사용하는 모바일 디바이스의 주소록에 있는 이름과 전화번호를 공개한다는 루머가 퍼지자, 페이스북은 많은 비난을 받았다. 페이스북은 이에 대해 이 기능이 어떻게 작동하는지, 페이스북 계정에서 정보를 제공하려면 어떻게 해야 하는지 설명했다. 그런데 왜 애초에 기능을 처음 사용할 때 자세한 정보와 방법을 공개하지 않았던 것일까?

2011년, 고대디GoDaddy의 CEO가 아프리카코끼리 사냥과 논란의 여지가 있는 SOPAStop Online Privacy Act에 대한 지지 의사를 공개적으로 표현했을 때, 이에 반감이 있던 고객 수만 명이 도메인을 다른 회사로 옮겼다. 2011년 봄에 카이로에서 발생한 정치 폭동을 자사의 제품을 홍보하는 데 이용한다고 생각되었을 때, 케네쓰 콜Kenneth Cole 브랜드는 손해를 입었다. 후에 사과했지만 브랜드 이미지는 계속 손상되었다. 2010년, 월마트는 협력 광고 회사 중 하나가 가짜 온라인 커뮤니티를 만들어 새로 연 지점을 지원했다는 의혹을 받았다(당시에는 이런 방법이 잘 알려지지 않았었다). 이러한 사례는 비슷한 상황에서 여러분의 회사가 어떻게 대응할지 고민을 남길 것이다.

앞의 사례는 모두 빅데이터를 이용했다는 공통점이 있음에 주목하자. 사례에서 알 수 있듯이, 빅데이터는 기존 기술보다 더 빠른 시간에 훨씬 큰 영향을 미치며 회사의 브랜드 이미지, 고객 관계, 매출에 직접적으로 반영된다. 빅데이터 덕분에 더 많은 사람이 참여하고, 이야기하고, 반응할 수 있게 되었기 때문이다. 이제 빅데이터는 우리의 가치와 행동에 새로운 고민을 던져주고 있다. 빅데이터가 널리 보급됨에 따라 기업이 만들어내는 정보는 빠르게 변화하고 증가하게 되었다. 결국 빅데이터 현상은 점차 윤리적인 문제를 제기하게 된 것이다.

브래드 피터스Brad Peters가 포브스에 기고한 대로, 빅데이터는 말 그대로 사회적 계약을 바꾸어 놓았다[02]. 이 변화는 실로 복잡하다. 이 책을 쓰는 주요 목적 하나는 개인이나 기업에 이 문제를 더 명백하고 투명하게 논의해야 한다고 제안하는 것이다. 그리고 이 논의는 본질적으로 윤리적인 요소를 포함한다.

비록 윤리 문제는 개개인에 대한 문제지만, 그 영향은 다양한 곳까지 미친다. 또한 빅데이터가 프라이버시 보호에 대한 걱정을 일으키는 것과 마찬가지로, 다음과 같은 개인의 신원에 대한 새로운 의문도 일으킨다. 누가 개인 데이터를 소유하는가? 현재 사용자의 온라인 정보가 더 많이 노출되고 더 많은 데이터에 접근할 수 있게 되면, 우리 회사에는 어떤 영향을 미칠 것인가?

실제로 개인이나 기업이 빅데이터 윤리에 대한 문제를 고려할 때는, 다음과 같은 네 가지 요소가 많은 영향을 미친다.

- **신원**

  오프라인 신원Identity과 온라인 신원은 어떤 관계에 있나?

- **프라이버시**

  누가 데이터에 대한 접근을 통제해야 하나?

- **소유권**

  누가 데이터를 소유하는가? 데이터에 대한 권리를 이전할 수 있는가? 데이터를 생성하고 사용하는 사람들은 어떤 규제를 받는가?

- **평판**

  어떤 데이터를 믿을지 어떻게 판단할 수 있을까? 자신에 대한 정보든 외부에 대한 정보든, 빅데이터는 정보량과 이 정보를 활용하는 방법을 기하급수적으로 늘린다. 이런 현상 때문에 다른 사람이 우리를 어떻게 인지하고 판단할지를 관리하는 일이 더욱 복잡해졌다.

---

02 http://www.forbes.com/sites/bradpeters/2012/07/12/the-age-of-big-data

개인과 기업이 데이터를 처리하는 방법에 관심을 두는 것은 당연하다. 여러분이 기업에서 어떤 역할을 담당하든, 관련 기술 분야에서 일하든, 오늘날 거의 모든 사람의 삶은 빅데이터 기술에 영향을 받는다. 이러한 점 때문에 빅데이터가 주는 장점뿐만 아니라 정말 엄청나게 많은 사람에게 뜻하지 않은 결과를 줄 수 있는 잠재적 위험성이 동시에 존재함을 느끼게 된다.

일례로 뉴욕 법원에서 판사로 일하는 개리 브라운Gary Brown은 최근 저작권 침해자를 찾아내기에 IP 주소는 충분한 단서가 되지 않는다는 견해를 밝혔다[03]. 이는 저작권 문제에 대한 판결이었지만, 빅데이터 윤리의 주요 4가지 문제에도 상당한 영향을 미칠 것이다. 당연한 말이지만 사람이 IP 주소가 아니라면(그리고 사람과 IP 주소가 같다고 생각할 사람은 없을 것이다), 특정 IP 주소로 생성된 데이터가 합법적으로 특정한 개인과 연관될 수 있을까?

온라인 마케터는 이 문제를 오랜 시간 고민했다. 그러나 빅데이터의 뜻하지 않은 위험은 타깃 마케팅 문제보다 훨씬 범위가 넓다. 예를 들어 여러분이 집에서 다른 사람과 같은 시간대에 컴퓨터를 사용한다면 구글은 여러분이 선호하는 내용을 어떻게 알아낼 수 있을까? 집에서 필자가 관심을 갖는 바닷가 휴양지 정보를 검색할 때, 인터넷 회선을 공유한다는 이유로 바다를 무서워하고 수영도 못하는 다른 사람과 연관 짓는 일은 어불성설이다. 즉, 타깃 마케팅에서 연관성을 높이는 일은 상당히 어려운 일이며, 빅데이터로 인한 위험은 단지 온라인 광고에서 끝나지 않는다.

구체적인 사례를 보면 개인이나 회사가 직면한 어려움을 잘 알 수 있다. 예를 들어 노인의 혈당치와 심장을 감시하는 장치가 컴퓨터와 IP 주소를 공유한다고 가정해보자. 일반적으로 의료 장치가 생성한 모든 데이터는 의료 기관에 저장되는데 병원에 데이터 유출 사고가 발생해서, 노인들의 건강 상태 정보와 여러분의 데이터가 뒤섞였다고 생각해보자. 결국 당신과 같은 주소에 사는 성(性)별이 같고 성(姓)도

---

03 http://torrentfreak.com/judge-an-ip-address-doesnt-identify-a-person-120503/

같은 노인의 정보가 뒤섞인다.

물론 데이터가 뒤죽박죽되어 엉뚱한 사람에게 치료 서비스 청구서가 발송될 수 있지만, 이는 단순히 경제적 문제인 것은 아니다. 이미 IP 주소 기반으로 사용자를 구분하는 기술이 널리 사용됨을 고려하면 법적으로 IP 주소와 특정 개인을 연결하지 못하게 하는 것도 위험할 수 있다.

또한 인터넷보다 더 오래 존재해온 스푸핑Spoofing[04]도 염두에 둬야 한다. 이 사례에서는 특정 IP 주소에서 생성된 데이터는 단 한 사람의 데이터라고 가정했기 때문에, 가족과 여러분의 데이터를 혼동하여 여러분이 응급실에 실려가는 재앙이 발생할 수도 있다.

결국 브라운 판사의 견해는 프라이버시, 신원, 평판, 데이터 소유권이라는 미묘한 문제를 연구해야 할 필요성을 일깨워준다. 또한 빅데이터 기술과 사용자 관계의 복잡성과 미묘함을 이해하는 것은 매우 중요한 일임을 보여준다.

1.1 빅데이터는 무엇인가?

지금쯤 여러분은 "다른 데이터는 안 그런가?", "모든 기업은 지금까지 데이터를 관리하기 위해 노력해오지 않았던가?"라는 생각을 할 수도 있다.

하지만 빅데이터는 지금까지 행해져 온 데이터 관리의 개념을 뒤바꾼다. 우선 새로운 데이터가 생성되는 양은 엄청나다. 우리는 전 세계적으로 생성되는 데이터의 양이 엑사바이트(10^{18})나 제타바이트(10^{21}) 단위로 계산해야 하는 시대에 살고 있다. 또한 2025년에는 인터넷 데이터양이 지구에 사는 모든 인류의 두뇌 용량을 초과할 것으로 예상한다. 이러한 현상에 발맞춰 생성되는 데이터의 출처와 데이터 종류는 새로운 기술이 등장하는 속도만큼 빠르게 다양화되고 있다.

---

04 인터넷에서 다른 사람인 것처럼 가장하는 행동이다.

실제로 차량 모니터의 성능 계측기, 생산량 측정 장치, 의료 기기, 스마트 그리드 에너지 장치, 모두가 데이터를 생성한다. 게다가 모두 빠른 속도로 데이터를 만들어낸다. 그 결과 시간이 지남에 따라 데이터 출처의 수와 형태가 다양해져서 데이터 생성, 획득, 처리, 출력 속도는 기하급수적으로 증가하는 중이다. 이를 뒷받침하는, 전 세계 데이터의 90%는 최근 2년에 생성된 데이터라는 보고서도 널리 알려져 있다.[05] 즉, 빅데이터 혁명은 복잡한 이벤트 처리, 온라인 트랜잭션 포착, 모바일 컴퓨팅 제품과 서비스 개발, 대량 이벤트 데이터 실시간 처리 등에 많은 변화를 가져왔다.

현재 빅데이터는 너무 방대해서 SQL과 같은 전통적인 데이터베이스 시스템으로 처리하고 분석할 수 없는 데이터라고 정의한다. 기술적으로는 기존 데이터베이스 기술의 능력과는 차별화된 하둡Hadoop 클러스터, 블룸Bloom 필터, R과 같은 데이터 분석 도구, 새로운 데이터 스토리지, 처리 기법들에 기반을 둔다는 정의를 내리고 있다. 그래서 빅데이터라는 용어는 시간이 지나면 바뀔 수도 있다. 지금은 크다고 생각하더라도, 조만간 처리 능력에 비해서 작은 데이터가 될 수 있기 때문이다. 이런 측면에서 보면 빅데이터에서 '크기'는 한 가지 측면일 뿐이다. 좀 더 넓은 관점에서 보면 사용할 수 있는 정보의 양, 다양성, 속도가 증가하면 회사가 관리해야 하는 정보의 복잡도를 기하급수적으로 증가시키며, 지금까지 비즈니스를 해오면서 접하지 못했던 문제를 일으킨다. 즉, 빅데이터 관련 요소에서도 위험과 윤리 문제는 발생할 수 있다는 의미다.

지난 몇 년간 빅데이터 기술을 사용하는 여러 회사와 일하면서, 엄청나게 다양한 빅데이터 환경에서 윤리 문제와 제약에 관한 의견이 합의점에 도달하지 못하는 상황을 자주 봤다. 그리고 비즈니스 환경에서 윤리 문제를 논의하는 공식적이고 명확한 구조가 없으므로 사람들은 각자의 도덕성에 의존하곤 했다.

---

05 http://www.economist.com/node/21537967

물론 이런 상태는 좋은 출발점이긴 하지만, 곧 아무런 결론도 나지 않는 “그건 무시 무시해…”, “아니야, 그렇지 않아”라는 논쟁으로 빠져버릴 수 있다. 다시 말해 이런 상황에서는 토론이 무의미해지고, 회의가 끝나도 아무런 답도 얻지 못한다는 것이다. 그리고 이렇게 뜻하지 않은 결과 때문에 일어나는 피해는 빅데이터 혁명이 제공하려고 했던 가치를 줄인다.

그래서 비즈니스 혁신을 꿈꾸는 사람은 빅데이터 기술에 기반을 둔 다양한 제품과 서비스를 새로 설계하고 구현함으로써 만들 수 있는 잠재적인 가치에 열광한다. 하지만 사용할 수 있는 정보의 크기, 다양성, 속도가 엄청나다는 점에서 새로운 의문이 생겨난다. 그리고 이런 의문 중 일부는 사람들의 성격, 행동, 선호, 관계, 위치에 대한 대량의 데이터를 획득, 저장, 사용하는 일이 어떤 결과를 가져올 것인가에 대해서다.

기본적으로 이런 의문은 윤리 문제이며, 여러분의 가치관이 무엇이고 그 가치관을 제품과 서비스를 만드는데 어떻게 적용할 것이냐는 문제와 관련 있다. 그리고 혁신이 가져다줄 유용함과 해로움을 어떻게 조절할 것인가의 핵심은 여러분의 가치관이다. 의식하지 못할 수도 있지만, 여러분의 '가치관'은 대량의 데이터에서 얻는 정보에 의존하는 제품과 서비스를 설계하고 실행하는 방법을 알려주며, 뜻하지 않게 발생하는 위험과 약속된 혜택을 비교해서 올바른 판단을 내리는 데 결정적인 역할을 한다.

이 말은 빅데이터가 제공할 수 있는 혁신의 혜택과 위험 사이에서 적절한 균형을 유지해야 함을 의미한다. 즉, 기업이 빅데이터로 인한 혁신의 혜택과 위험 사이에 균형을 유지하도록 도와주는 윤리 토론의 명백한 구조를 만들 수 있게 하는 것이 이 책을 쓰는 목적 중 하나다.

1.2 빅데이터 시대가 우리에게 요구하는 것

겨우 한두 세대 전만 해도 쉽게 '공공기관의 눈을 피해' 미국에서 사라질 수 있었다. 하지만 오늘날에는 누구도 데이터라는 흔적을 남기지 않고 할 수 있는 일이란 거의 없다. 따라서 어느 정도 지식과 장비를 갖춘 수사 기관은 누구든 추적할 수 있다.[06] 그런데 우리 사회나 정부, 법 제도는 아직도 투명성, 연관성, 결합을 요구하는 다가올 빅데이터 시대에 적응해 있지 않으며, 그럼에도 전 세계 정부는 빠른 속도로 새로운 법을 입안하고, 논의하고, 승인하고 있다.

빅데이터는 영구적이다. 예를 들어 미국 국회 도서관은 2006년 이후 기록된 모든 트윗[07]을 저장한다. 이는 이 데이터를 영구적으로 보관하기 위함이다. 페이스북은 계정을 삭제해도 계정에 관련된 데이터를 모두 삭제하지 않음을 묵시적으로 시인했다.[08] 그런데 이런 영구성은 비즈니스나 사회가 여태 겪어보지 못했다.

에릭 프리먼Eric Freeman과 데이비드 겔런터David Galernter는 이러한 현상을 설명하기 위해 라이프스트림Lifestream이라는 용어를 만들었다.[09]

> "… 온라인 활동의 일기 역할을 하는, 시간순으로 정렬된 일련의 문서. 여러분이 만든 모든 문서와 다른 사람이 여러분에게 보낸 문서는 모두 여러분의 라이프스트림에 저장된다. 이 스트림의 꼬리에는 온라인 탄생 증명서를 시작으로 과거의 문서가 있으며, 꼬리에서부터 현재로 이동하면서 차례로 좀 더 최신의 문서(작성 중인 문서나 새로운 이메일 등)를 담는다. 다른 문서(사진, 연락, 영수증, 영화, 음성 메일, 소프트웨어)는 그 사이에 저장되어 있다. 현재에서 미래로 이동하면 앞으로 필요하게 될 문서(일정, 달력, 할 일 목록)를 담을 것이다."

---

06 http://www.wired.com/vanish/2009/11/ff_vanish2/

07 트위터에 남기는 글이나 사진 등을 의미한다.

08 http://arstechnica.com/gadgets/2012/05/on-facebook-deleting-an-app-doesnt-delete-your-data-from-their-system/

09 http://cs-www.cs.yale.edu/homes/freeman/lifestreams.html

프리먼과 겔런터는 개인의 온라인 정보를 관리하는 소프트웨어 아키텍처와 프레임워크를 정의하기 위해 라이프스트림을 사용하려 했지만, 이 개념은 빅데이터의 영속성이 개인의 삶에 있어 중요한 부분에 어떤 영향을 미치는지 이해하는 데 더 유용하다. 빅데이터는 종종 '메타데이터'를 포함하는데, 메타데이터는 개개인에 대한 하나 이상의 정보 계층을 추가할 수 있다. 예를 들면 빅데이터 아키텍처와 기술을 사용하면 각 트위터 메시지를 생성할 때 실제 물리적 위치를 연결할 수 있다.

이렇게 추가된 정보 계층이 가진 의도는 명백하다. 금융이나 사회적인 트랜잭션의 본질과는 거의 상관없는 다양한 부가 정보가 담겠다는 뜻이다. 이런 정보와 빅데이터 기술을 이용하면 지구 상 어디든 여러분의 여행 경로를 재구성할 수 있으며, 서로 관련 없는 개인 정보를 의도적으로 결합할 수 있다. 그리고 이를 완전히 막는 것은 불가능하다. 봄 휴가 때 찍은 사진이 취업할 때 참고 자료로 쓰일 것이라고 생각하는 사람은 없지 않겠는가? 빅데이터 때문에 이런 상황에서 우리의 평판을 관리하는 방법이 크게 변했다.

이런 데이터 흔적을 활용하는 것은 인간의 행동과 활동을 더 넓고 깊게 관찰할 수 있게 해주는 빅데이터 기술의 한 예일 뿐이다. 빅데이터와 관련 있는 모든 형태의 혁신은 이런 관찰을 통해 가치 있는 제품과 서비스를 새로 만든다. 그리고 풍요로운 데이터는 마케팅, 관리, 교육, 연구 개발, 의료, 정부, 서비스와 우리 삶의 다양한 측면을 개선할 수 있음을 약속하고 있다. 또한 빅데이터는 분석을 통해 효과적인 교육 정책을 만들거나 특정 지역에서의 위험한 기후 조건을 예측하는 능력을 향상하는 데 이미 사용되고 있기도 하다.

빅데이터 기술은 개인과 회사에 대한 프라이버시, 평판, 소유권, 신원과 같은 중요한 개념의 의미에 영향을 미친다. 원래 정보의 소유자뿐만 아니라 원래 정보를 이용해 혁신적인 제품과 서비스를 구현하려는 사람도 수집하고 결합할 수 있기 때문에, 정보가 우리 손을 떠나면 어떻게 사용되는지 통제하지 않으며 통제할 수도 없

다. 이 때문에 빅데이터가 만드는 '강요 기능forcing function'[10]은 데이터를 다루는 방법에 새롭고 긴급한 문제를 제기한다. 이 문제는 흔히 한 회사가 통제할 수 있는 관리 범위를 벗어나기 때문에 곤란하다.

그리고 빅데이터는 어떤 국가 인구의 몇 배나 되는 사용자를 온라인 커뮤니티 하나에 모일 수 있게도 한다. 페이스북은 잘 알려진 대표적인 예고, 이외에도 인터넷에는 수천 개의 온라인 커뮤니티가 있으며 각 커뮤니티는 각 사용자에 대한 정보 중 고유한 일부 측면을 담고 있다. 그런데 우리는 이제서야 우리 신원에 미치는 이런 현상의 영향, 소유권 개념, 사람과 사람 관계, 신뢰, 평판 등을 보는 방법, 우리 삶의 전통적인 자기 관리 측면을 이해하기 시작했을 뿐이다.

데이터는 사람의 개성과 행동을 그대로 담기도 하므로, 획득한 데이터를 올바르게 사용하거나 남용하는 방법은 매우 많을 것이다. 데이터를 올바르게 사용해서 얻는 직접적인 혜택은 지금 실현되고 있다. 그런데 개인 데이터를 얻고, 결합하고, 판매하고, 찾아내고, 재판매하고, 다른 데이터와 결합하는 데는 데이터 남용에 대한 우려가 현실화되고 있다.

그리고 이런 위험은 개인에 국한된 것이 아니며 기업에도 마찬가지로 적용된다. 기업은 고객에게 해를 끼치는 일을 목적으로 삼지 않으며, 병원은 환자의 비밀을 누설하는 일을 목적으로 삼지 않는다. 또한 비영리 연구 기관은 실험 주체의 신원을 알 수 있는 정보를 공유하는 일을 목적으로 삼지 않는다.

그러나 빅데이터 기술을 사용하는 일상적인 작업을 통해 다양한 위험은 점점 더 많이 증가한다. 그리고 위험의 형태, 규모, 영향을 미리 결정하기 어렵다. 사회 전체적으로 보면 빅데이터 시대의 의미를 이해하기 시작했을 뿐이기 때문이다.

다음 사실을 생각해보자.

---

10 어쩔 수 없이 시급하게 처리하거나 고민해야 하는 기능 또는 그러한 상황, 이슈를 말한다.

- 웹 브라우저나 과거 위치(예를 들어 가슴 통증에 대한 정보를 가진 웹사이트 방문이나 자동차의 GPS 상세 기록) 정보에 기반을 두고 보험료율을 설정할 때의 사회적 경제적 영향[11]

> **NOTE_** OnStar는 프라이버시 침해 문제로 이 결정 사항을 재빠르게 백지화했다. 자세한 정보는 http://www.computerworld.com/s/article/9220337/OnStar_reverses_course_on_controversial_GPS_tracking_plans를 참조한다.

- 고용 여부를 결정하는 데 유전자 정보를 사용.
- 위치 정보, 소셜 네트워크 기록, 웹 접속 기록을 이용해 범죄 행위 '예측하기'. 영화 '마이너리티 리포트'와 같은 '예측 단속'이 이미 일부 주요 도시 지역에서 적용되고 있다.[12]
- 개인의 범죄 기록, 웹 접속 기록, 데이트 평가 사이트에 접속하는 스마트폰 '데이트 앱'으로 찍은 사진에 기반을 두고 개인 메타데이터를 판독.

이러한 사실 때문에 빅데이터 혁신으로 누리는 혜택 자체가 위험에 처해있다. 실제로 2011년 후반과 2012년 초 의회에 제출된 SOPA(Stop Online Privacy Act)는 다양한 업계, 기업, 개인에게 강력한 반발을 샀다. 제안한 법안 항목이 빅데이터와 같은 기술을 사용해 만드는 미래 혁신을 심각히 제한한다는 우려가 주요 원인이다.[13]

법안을 지지하는 국회 의원이 이해 관계자로부터 이런 기술의 작동 방법과 어떻게 이런 혁신이 가능했는지에 대해 잘못된 정보를 들었거나, 아니면 인터넷과 빅데이터 기술이 운영되는 현실을 아예 모른다는 주장도 있다. 어찌 되었든 SOPA는 우리 사회의 주요 경제적 사회적 부분이 어떻게 움직이는지에 대한 투명하고 명백한 논

---

11 http://www.wired.com/threatlevel/2011/09/onstar-tracks-you/

12 http://www.cbsnews.com/8301-18563_162-57412725/lapd-computer-program-prevents-crime-by-predicting-it/

13 http://en.wikipedia.org/wiki/Stop_Online_Piracy_Act

의가 부족할 때, 이러한 기술로부터 혜택을 누릴 수 있는 사회 전체적인 능력이 심각하게 제한될 수 있음을 보여주는 전형적인 예다.

결국 빅데이터가 가진 '강요 기능'은 기업과 개인이 쌓는 데이터에 더 깊숙이 파고들 수 있으므로, 빅데이터 기술이 해를 끼치기보다 혜택을 주게 하려면 위험과 혁신을 적절히 조절하는 일이 시급하다.

1.3 빅데이터는 윤리적으로 중립이다

빅데이터 기술은 이윤을 창출하거나 더 큰 사회적 혜택을 누리게 하는 혁신적인 제품과 서비스를 만들 수 있게 정보를 연결하지만, 다른 기술과 마찬가지로 윤리 측면에서는 중립적이다. 다시 말하면 빅데이터 자체나 빅데이터 기술을 사용하는 데 있어서 선악에 관한 판단 능력은 존재하지 않는다. 빅데이터 기술은 가치 체계를 가지고 있지 않다. 그러나 개인이나 기업에 윤리 질문의 답을 요구해야만 빅데이터가 개인이나 기업의 가치에 일치하도록 사용할 수 있다.

이를 위해서는 가치를 명백히 탐구하고 윤리 관점을 바로 세운 뒤 토의를 거쳐야 한다. 쉬운 일은 아니다. 윤리는 상당히 개인적인 주제이며, 선, 악, 정의, 부정과 같은 양극화된 단어를 많이 사용할 수밖에 없다. 또한 사람은 각각 개인적인 도덕성을 가지므로, 윤리에 관한 기준도 사람마다 다를 수밖에 없다. 즉, 우리가 개인적으로 믿는 것과 회사의 임직원으로서 빅데이터로 수행하려는 일의 관계를 윤리적으로 표현하는 데는 공통된 용어가 부족하며, 이는 생산적인 토의 과정을 제한하고 합의를 이끌어 내는 데 장애가 될 수 있다.

이 책은 운영 정책, 실제 상황, 성문화된 법에 영향을 주려는 책은 아니다. 방금 소개한 주제는 회사 경영진, 관리자, 판사, 선출직 공무원들이 다뤄야 한다. 또한 이 책은 (전통적인 관점에서의) 비즈니스 윤리에 관한 책도 아니다. 비즈니스는 주로 이익과 혁신에 관심을 두기 때문에, 이윤을 창출하는 사업이나 시급히 바뀌는 시장

요구를 만족하는 제품과 서비스를 지속해서 개발하는 데 영향을 줄 때만 공식적으로 윤리 질문에 관심을 두게 된다.

그러나 비즈니스의 본질에는 사회적 요소가 있으며, 사실 빅데이터와 소셜미디어 때문에 이런 진실에 과장이 섞여 주목받았을 뿐이다. 모든 상거래, 어떤 가치를 가진 재화와 서비스를 교환하는 단순한 행동도 본질적으로 사람과 관련 있는 활동이다. 그리고 모든 사람은 나름의 가치관을 가진다. 이 책의 목적은 그런 가치관을 표현하는 장소인 비즈니스 환경에서 윤리 토론을 쉽게 할 수 있는 구조를 만들고 여러 단체가 가치관에 맞는 행동을 하도록 돕는 것이다.

빅데이터가 가진 '강요 기능'은 비즈니스 기능과 개인의 가치를 더욱 긴밀히 연결되게 만든다. 즉, 생성되는 데이터의 엄청난 양, 다양성, 속도를 통해 기업의 행동이 개인의 삶에 훨씬 더 깊이 파고들게 만든다. 또한 빅데이터 제품의 설계, 개발, 판매, 관리 업무가 프라이버시, 평판, 소유권, 신원과 같은 용어가 가진 일반적인 의미를 바꾸면서, 개인의 삶에 미치는 영향과 충격도 커지고 있다.

빅데이터가 엄청나게 크고 우리 생활 어디에나 존재하기 때문에 신원, 프라이버시의 진화, 데이터 소유의 의미, 온라인 데이터의 흔적이 생활 전반의 평판에 미치는 영향에 대해 새로운 질문을 하지 않을 수 없다. 비즈니스에서 교육에 이르기까지, 단순한 연구기관에서 제조업과 전문 서비스 기업에 이르기까지 모든 조직은 자신의 고객과 업무, 그리고 측정할 수 있는 거의 모든 사항에 대한 엄청난 양의 정보를 갖고 있다. 지난 5년 동안 빅데이터 기술이 급속하게 발달하기 전에는 조직 프로세스나 정책의 변화는 고객의 삶에 영향을 주지 않거나 주더라도 즉각적인 것은 아니었다. 또한 고객의 개인 데이터에 접근할 수 있는지는 전형적으로 고객 데이터에 접근할 수 있는 개인이나 기업의 숫자 문제였다.

이제는 빅데이터가 엄청난 규모와 속도로 운영되기 때문에 정책이나 실행이 변화하면 많은 사람에게 더 많이, 더 빠르게 영향을 준다. 따라서 비즈니스 기능의 변화

는 개인의 삶에 훨씬 더 큰 영향을 미치며, 전통적인 운영 방법의 확장은 관리는 고사하고 일일이 다 기억할 수도 없을 정도로 여러 가지 면에서 우리의 일상적인 삶에 영향을 미친다. 지금 말할 수 있는 것은, 빅데이터가 성장함에 따라 일반적인 상호 작용에 대한 법이나 사회적 규범, 경제, 심지어 일반적인 기대조차 어떻게 바뀔지 알 수 없다는 점이다.

윤리 토론을 장려하는 것은 앞에서 말했던 바와 같이 어떻게 바뀔지 모르기 때문이다. 빅데이터가 주는 혁신의 위험과 혜택 사이의 균형을 맞추기 위해 가치와 실행을 일치하게 하는 공개적이고 명백한 대화는, 균형을 맞춰야 하는 상황에서 여러분의 상황을 유리하게 이끌 수 있는 방법의 하나다. 즉, 어떤 윤리 결정점Decision Point으로 바뀔 때나 윤리적 판단을 실행으로 옮겨야 할 때를 식별하는 일은 투명한 관점과 정책을 개발할 때 더 많은 형식적인 부분과 상황에 맞는 부분을 함께 토론할 수 있는 능력을 갖추는 첫 번째 단계다.

1.4 무엇을 해야 하는지 말하는 것이 아니다

이 책은 윤리적 기준으로 데이터를 다루기 위한 확고한 규칙 목록을 나열하지 않는다. 사실 이런 이슈는 해당 비즈니스 모델, 분야, 업계에 따라 세세하게 달라질 수밖에 없다. 따라서 이 책은 이런 이슈를 직접 언급할 때의 장점을 보여주고, 윤리 질문에 대한 일관적인 접근법을 개발하는 데 중요한 요소들을 이야기하고, 토론 구조를 만들고 장려하는 데 필요한 사항을 설명한다. 토론은 회의실, 임원 회의, 법정뿐만 아니라 실무자 회의, 복도, 구내식당에서도 명백하고, 협력적이고, 투명하게 할 수 있다. 그리고 명백하고 투명한 토론을 하는 목적은 고객과 사업 협력자의 관계에 대한 위험을 더 잘 이해하고, 위험을 줄이고, 빅데이터 혁신의 혜택을 더 잘 표현하기 위한 것이다.

부정적인 인식과 언론보도가 매출에 영향을 준다는 사실을 생각해야 한다. 데이터를 비윤리적으로 취급한다는 인식만으로도 내부적으로는 비즈니스 목표에 대한 지

원이 줄어들고, 외부적으로는 고객과의 관계가 나빠지는 부정적인 결과를 가져올 수 있다. 이는 방금 소개한 투명성이나 올바른 관리의 문제가 아니다. 빅데이터가 진화하고 사람의 삶에 더 많이 파고들어 영향을 줌에 따라 행동과 가치를 계속 일치하는 것에 대한 더 폭넓은 윤리 질문이다.

간단히 말해 이 책은 여러분의 데이터를 어떻게 사용해야 하는지 설명하지 않는다. 오늘날 빅데이터를 중요하게 생각하는 기업에 의해 발생하는 윤리 문제 토론에 생산적으로 참여하고 기업 안에서 행동과 가치를 일치시키는 방법을 소개한다. 이러한 구조는 모든 기업이 고객, 협력사, 그 외 관련자들과 서로 신뢰하고 분명한 가치관에 따라 한결같이 협업할 수 있는 기업이 될 수 있도록 돕는 일련의 도구를 제공한다.

1.5 중요 개념과 용어

윤리 결정점을 찾아내면 개인 데이터와 연관되는 비즈니스 운영, 제품, 서비스의 가치를 주는 정책과 관점을 개발하는 데 도움이 된다. 그러려면 여러분이 어떤 가치를 가지며 어디에서 가치와 운영이 일치되지 않는지 알아야 한다. 그리고 이는 전문화된 용어를 사용하는 복잡한 일이 될 수 있다. 그럼 정책을 개발하는 데 있어 유용한 용어를 살펴보자.

● 권리와 관심

사람들은 흔히 프라이버시를 권리라고 말하는데, 권리에 대해 말할 때는 이런 권리가 절대적이라고 생각하기 때문에 당면한 문제를 예단하기 마련이다. 이런 예단을 피하려면 프라이버시는 어떤 종류의 관심이고, 권리는 가장 강력하며 절대적인 관심 일부라는 것을 명백히 이해해야 한다.

예를 들어 여러분의 의료 데이터의 사용에 관한 절대적인 프라이버시는 이 데이터와 관련해 아무런 프라이버시 위험이 없음을 규정하는 권리를 포함한다. 그러나 여러분이 의식을 잃고

응급실에 실려와 치료를 받았다고 가정해보자. 그동안 데이터는 계속 생성되며, 이 데이터는 여러분이나 이런 상황에 있는 다른 사람을 잘 치료하기 위해 유용하게 사용할 수 있다. 이때 여러분의 동의 없이 이 데이터를 사용할 수 없거나 공유하는 일을 금지해야 한다고 생각하는가? 바로 옆의 간호사나 의사에게도? 물론 그렇게 할 수 있는 권리도 있다고 주장하고 싶겠지만, 권리가 있다고 생각하는 것은 우리가 지지할 수 있는 토론 결과가 되어야 하지 전제가 되어서는 안 된다.

그럼에도 권리를 가진다는 것은 윤리적인 견해이기 때문에 문제는 훨씬 더 복잡하다. 절대적인 정의를 지지하는 일은 윤리적 입장을 포함하며, 이 입장이 가치에 맞는지 객관적으로 판단할 수 있는 능력을 감소시킨다. 이때 프라이버시 관심(권리가 아닌)의 관점에서 생각해보면 윤리적 주장의 강점을 자유롭고 객관적으로 판단할 수 있게 된다.

- **개인 데이터**

개인 식별 정보와 그 외 정보에 대해 일반적으로 생각하는 차이점은 대체로 극복할 수 있는 기술 제약 때문에 인위적으로 만든 것이다. 따라서 토론을 진행하려면 개인이 프라이버시에 대해 걱정할 때 문제가 되는 데이터에 대한 매우 폭넓은 용어가 필요하다. 참고로 여기에서 언급하는 개인 데이터는 개인에 의해 생성되는 모든 데이터를 의미한다.

- **책임 기업**

올바른 일을 실행하는 것과 여러 사람이 올바르다고 생각하는 일을 실행하는 것의 차이는 이 주제에서 상당히 중요하다. 책임 기업Responsible Organization이란 데이터를 처리함에 있어 자신의 가치관에 부합하는 것과 다른 이들이 기대하는 방법으로 처리하는 것 모두에 관심을 두는 기업을 말한다. 책임 기업은 서로 다른 이 두 가지 걱정을 조율하는 법을 찾아야 한다.

빅데이터는 크고, 빠르고, 다양한 정보를 담을 수 있다. 또한 실제로 존재하며, 경제적 이익, 사회적 혜택, 문화적 발전을 약속한다. 그리고 이전에는 윤리 문제에 답할 필요가 없었던 장소와 환경에서 윤리 문제에 대한 답을 요구한다. 개인과 기업은 어떻게 반응해야 할 것인가? 이 책은 명시적이고, 투명하고, 생산적인 윤리 질문에 참여하는 방법을 지지한다.

2장은 이런 윤리 질문으로 여러분의 가치와 행동을 일치시켜 혁신을 증진함과 동시에 위험을 감소시킬 수 있게 하는 방법을 설명한다. 먼저 성공적이고 잘 운영되는 회사에서도 존재하는 윤리 불일치 사례를 설명하고, 가치와 행동을 더욱 잘 일치시키기 위한 윤리 질문에 참여하는 데 필요한 용어와 구조를 설명한다.