**빅데이터의 이면**

**- 개인정보의 활용에 관하여 -**

20152689

산업보안학과 황성은

* **서론**

최근 4차 산업혁명 시대에 진입하여 국가 및 기업들이 이전과는 다르게 급진적인 변화에 대응하고 있다. 정보통신기술, 즉 ICT(Information & Communication Technology)와 관련하여 수많은 이슈들이 생성되고 있으나, 그 중에서도 핵심이 되는 요소는 데이터이다. 이는 빅데이터 분야 뿐만 아니라 클라우드 서비스, 인공지능(AI), 블록체인 등에 활용되며, 데이터의 질과 양이 결과의 퀄리티에 영향을 미친다.

이러한 데이터들을 제대로만 활용할 수 있다면 기업 입장에서는 혁신 기술을 선도하는 열쇠이자 기존 시스템보다 효율적인 의사결정을 도출하는 무기가 될 수 있다. 맥킨지(McKinsey)에 따르면 교육, 의료, 마케팅 등 다양한 분야에서 빅데이터의 잠재적 경제 가치는 최소 3조 달러에서 5조 달러 이상으로 평가하고 있다. 그러나, 여기서 간과할 수 없는 점은 비즈니스에 핵심적인 데이터 즉, 활용할 수 있는 데이터의 상당수가 개인정보와 연관된다는 사실이다.

개인정보에는 이름, 주소, 직업 등 개인의 구체적 신상이 포함되며, 국가기관 및 기업의 데이터베이스에 저장되고 있는 상황이다. 이는 혁신 기술을 접하는 사용자로서 편리함과 만족감을 얻는 동시에, 개인정보 주체로서 정보의 유출가능성에 대한 불안감을 가지고 있다는 것을 의미한다. 개인정보보호로 인해 비즈니스에 관한 핵심 데이터를 활용하지 못한다면 4차 산업의 선구자가 될 수 없다. 그러므로 데이터의 활용과 개인정보보호 사이에서 합리적인 위치를 찾는 것이 본 레포트의 주요 목적이다.

* **본론**

먼저, ‘빅데이터’란 기존 데이터베이스 관리도구의 능력을 넘어서는 대량의 정형 혹은 데이터 베이스 형태가 아닌 비정형 데이터를 말한다. 이러한 데이터를 기반으로 가치를 생산하고 결과를 분석하는 기술 또는 가공되지 않은 정보를 분석하여 의사결정에 도움이 되는 정보를 도출하는 과정 또한 빅데이터의 영역에 속한다.

실제로 정보와 데이터는 인간의 활동과 분리될 수 없는 직접적인 관련이 있으며, 산업의 발전을 위해서는 국가 및 기업차원에서 국민의 개인정보를 활용하는 것이 불가피하다. 그렇다면, 현재 어떠한 방식으로 빅데이터의 기반 데이터가 수집되는지에 대한 과정을 알아보고 빅데이터의 사용이 개인정보 침해에 얼마나 영향을 미치는지를 개인정보 주체인 이용자의 관점에서 살펴보도록 하겠다.

예컨대, 휴대전화를 개통할 경우 계약서의 필수적인 동의가 요구되는 항목은 성명부터 시작하여 예금주 정보, 서비스이용기록, 개인위치정보에 이르기까지 다양하다. 다시 말해서 계약자 자신에 대한 거의 모든 개인정보를 제공해야만이 휴대폰의 개통이라는 서비스를 얻을 수 있다는 것이다. 그 뿐만이 아니다. 인터넷에 가입을 할 경우에도 병원에서 진료를 받고 약을 받을 때도 우리의 개인정보는 각 기관의 데이터베이스에 고스란히 쌓인다.

이러한 활동들이 결국 데이터의 기반이 되는 흔적들로 남겨지게 되고 기관 및 기업은 남겨진 흔적들을 이용해 개개인의 정보를 추적하고 수집, 분석을 한다. 이 과정을 통해 기업은 이용자에게 적합한 마케팅을 시도하는 것이다.

그렇다면, 이용자 입장에서는 일련의 개인정보를 제공함으로써 “자신에게 적합한 광고나 정보를 가져와준다면 시간-비용을 고려할 때 더 효율적이지 않나?”라고 생각할 수 있다. 즉, 편의를 위해 개인정보를 매개로 기업과 일종의 거래를 하는 것이다. 그러나 빅데이터는 단순하지 않다. 기업은 다양한 경로로 얻은 데이터를 저장하고 분석하여 개인의 소비 생활을 예측하고 더 나아가 인종, 성별, 나이와 같은 수많은 기준에 따라 차별하는 일도 가능하다.

이는 정부 및 공공기관에도 그대로 적용될 수 있는 중요한 사안이다. 정부도 개인의 정보를 수집하고 이를 활용한다. 휴대전화를 개통한 순간부터 GPS 정보의 열람이 가능해 개인이 어느 시기에 어디에 있는지, 연락의 빈도가 높은 사람은 누구인지 등 정말 사소한 정보까지도 파악할 수 있다. 이 뿐만이 아니다. 의료보험과 같은 국가의 공공 데이터 영역도 데이터베이스와 밀접한 관계를 갖는 데, 정부는 법 개정을 통해 공공 데이터까지도 민간에 개방, 공유를 한다고 한다.

이미 정부는 빅데이터 산업의 활성화를 위해 동의없이 정보주체의 개인정보를 얻을 수 있는 정책을 추진해왔다. ‘가명정보’의 도입과 과기부의 ‘빅테이터 플랫폼 및 센터 구축 사업’이 그러한 예이다. 그렇다면, “데이터베이스 상의 개인정보가 안전한 영역에서 활용되고 있는가?” 전혀 그렇지 않다. 주요한 문제는 빅데이터로 사용에 대한 사회적인 합의도 이루어지지 않은 상황에서 이를 활용함으로써 야기되는 개인정보 침해 위험에 대한 적절한 대응책 및 가이드라인이 충분치 않다는 것이다.

일례로, ‘개인정보 비식별 조치 가이드라인’에 의해서 공공기관이 가지고 있던 3억 건의 개인정보가 기업에 제공되었으며, 2014년 국내 카드 3사의 1억 건이상의 고객들의 정보가 내부 근무자에 의해 불법 유출된 사건이 있었다. 가장 민감한 건강 및 의료관련 개인정보도 이를 피할 수 없었다. 한국약학정보원은 약 8년간 수집한 의료정보를 해외 의료관련 빅데이터 기업에게 정보를 매각하였고 건강보험심사평가원이 국민의 표본 데이터를 제작하여 민간보험사에 팔아 넘겼다.

물론, 이 데이터들은 합법적인 규정을 통해 ‘비식별화’ 되었거나 ‘가명정보’이므로 특정인에 대한 정보까지는 알 수 없다. 그러나, 문제는 다량의 데이터를 비교분석한 ‘재식별화’의 기능을 이용해서 일정한 연관성을 가진 개인에게 가입 차별 등의 불이익을 가할 수 있다는 것이다.

이처럼 민간과 공공영역을 불문하고 개인정보의 침해사고는 끊이지 않고 있다. 빅데이터의 등장 및 개인정보 관련 법의 개정으로 인해 개인정보의 활용 빈도가 늘어났고 이에 비례하여 피해사례의 발생이 증가한 반면 개인정보 침해사고에 대한 법적 처벌은 현재까지도 미미한 수준이다.

일례로, 1억 건 이상의 개인정보를 유출시킨 카드 3사는 고작 수백만원의 과태료만을 받았으며, 과거 기업 개인정보 유출사태의 집단 소송에 참여했던 변호사는 법원의 소극적 판결로 인해 패소하였다. 이러한 개인정보의 불감증이 존재하는 사회에서 개인정보보호의 유출이 걷잡을 수 없어질 경우, 과연 어떠한 부정적인 영향이 발생할 지에 대해 심히 우려스럽다.

다행인 점은 법적으로 미국, 유럽 등의 주요 국가들이 빅데이터의 활용으로 발생하는 문제점을 인지하고 법을 개정하였으며, 한국도 이에 따라 개인정보의 비식별 데이터와 관련하여 정보 주체가 큰 영향을 받지 않도록 유연하게 법 규정을 추진하고 있다는 것이다.

기술적인 측면에서도 이를 방지할 수 있는 몇 가지의 수단이 존재한다. 첫번째는 ‘익명정보’의 사용이다. 익명 정보는 가명 정보나 비식별 정보와 비슷하나, 원본의 개인 정보를 다른 정보와 분석 및 결합하더라도 개인을 특정할 수 없도록 하는 조치이다. 두번째는 빅데이터의 위험 분석 기술이다. 즉, 해당 기업의 데이터베이스가 보유한 빅데이터가 어떠한 방식으로 사용되면 위험 가능성이 증가하는 지에 대한 분석 자료를 제공하는 기술이다. 이러한 보호 수단을 통해 개인 및 기업은 개인정보에 대한 침해가능성을 감소시킬 수 있을 것이다.

* **결론**

정리하면, 4차 산업에서 파생된 빅데이터라는 기술은 국가 및 기업에 필수불가결하다. 그러나, 기술의 활용이 현재까지 개인정보의 유출에 상당한 취약성을 가지고 있으며, 이로 인해 개인정보 침해 및 프라이버시 보호에 관한 문제를 유발하고 있다.

이러한 문제를 해결하기 위해 법, 기술, 정책분야에서 다양한 수단 및 방법들이 도출되고 있다. 그럼에도 불구하고 국가 및 기업과 개인정보 주체 간의 충돌로 인해 구체적인 합의점이 정해지지 않았고 이러한 상황속에서 허술한 법망을 통해 기업과 기업, 기업과 정부 간의 개인정보 유출 사례가 점점 증가함에 따라 개인정보 주체의 불안감도 커지고 있다.

현재 빅데이터에 대한 비식별화가 개인의 정보보호를 위한 조치로 시행되고 있으나, 이것은 완전한 해결방안이 아니다. SKT와 한화보험은 각 기업이 보유한 비식별화된 데이터를 결합하는 경우, 특정한 개인의 데이터가 도출된다고 분석했다. 즉, 비식별화를 거친 개인정보라고 하더라도 다른 데이터를 활용하여 새로운 정보유출로 이어질 수 있다. 또한, 현행 가이드라인에 의하면 전체적인 상황 및 심각성을 고려할 때, 개인정보 유출에 대한 처벌은 너무나도 미약한 수준이다.

그러므로 정부와 기업 그리고 개인정보 주체들이 서로 Win-Win 할 수 있도록 구체적인 합의를 통해 표준화된 가이드라인이 만들어진다면 현재 이슈가 되고 있는 문제들은 얼마든지 해결이 가능하다고 생각한다. 이를 위해서 정보 주체에 대한 기술적 및 관리적 보호조치가 필요하며, 기업도 빅데이터 기술을 활용함에 있어서 심각한 제약을 받지 않도록 패러다임이 바뀌어야 한다.

국가 및 기업의 관점에서 빅데이터의 규제를 완화하는 대신 개인정보의 비식별화 및 익명화를 필수로 하여, 빅데이터를 자유롭게 활용할 수 있도록 해야 한다. 반면, 개인정보 주체에게는 개인정보의 자기통제권을 더욱 높여야 하며, 빅데이터의 편의성 이면에 부정적인 영향이 있다는 것을 정확히 인식하고 정보를 합리적으로 관리 및 사용해야한다. 정보 주체들의 적극적인 자세와 기업의 노력이 지속된다면 안전하고 합리적인 스마트 사회가 구축될 수 있다고 생각한다.

* **Reference**

<https://www.boannews.com/media/view.asp?idx=44851&page=1&kind=2>

“빅데이터, 수집 시 부터 철저히 비식별화 조치해야”

<http://biz.khan.co.kr/khan_art_view.html?artid=201711050932001&code=920100>

““빅데이터 규제 풀어달라” 아우성, 해외는 강화 추세”

<https://ko.wikipedia.org/wiki/> “한국 위키피디아”

<https://www.mckinsey.com/mgi/overview> “맥킨지 글로벌 연구소”

<https://www.boannews.com/media/view.asp?idx=48312&page=1&kind=2>

“정부통합전산센터, 빅데이터 기반 행정지원 서비스 제공”