

추론개인정보의 법적 취급

I. 들어가며

II. 추론개인정보

1. 추론개인정보 기술 현황
2. 추론개인정보에 대한 규제 필요성

III. 추론개인정보에 대한 개인정보 보호법의 적용

1. 추론된 정보는 개인정보인가?
2. 추론개인정보에 대한 규제체계
3. 민감정보인 추론개인정보의 처리
4. 추론개인정보에 대한 정보주체의 권리

IV. 결론



조상현
변호사

I. 들어가며

“의식은 없지만 지능이 매우 높은 알고리즘이
우리보다 우리 자신을 더 잘 알게 되면
사회, 정치, 일상에 어떤 일이 일어날까?”

— 유발 하라리, 『호모 데우스』 마지막 문장

알고리즘이 우리보다 우리 자신을 더 잘 알게 되는 메커니즘은 무엇인가? 통상의 개인정보는 개인에 관한 관찰 가능한 정보이다. 가령 이름, 생년월일, 주소, 물품구매내역, 생김새 등이 있다. 반면 기호, 심리상태, 성격, 정치적 성향, 성적 지향과 같이 개인 내면에 관한 개인정보는 관찰 불가능할 뿐 아니라 개인들도 자기자신에 대해 정확히 알지 못하는 경우도 많다. 알고리즘이 우리보다 우리 자신을 더 잘 알게 된다는 말의 의미는 관찰 가능한 정보로부터 일련의 추론과정을 거쳐 내장된(embedded) 관찰 불가능한 정보를 추출할 수 있게 됨을 의미한다. 디지털 데이터의 축적, 데이터 연산 및 저장 능력의 지수적 증가, 그리고 고도화된 인공지능 알고리즘의 등장으로 기업은 이미 개인에 관하여 관찰 불가능한 정보를 추론할 수 있게 되었고, 이를 상업적 목적으로도 이용하고 있다. 본고에서는 추론의 근거가 되는 개인정보를 “근거정보”, 추론 결과 도출된 개인정보를 “추론 개인정보”라 정의한다.

추론개인정보가 이미 널리 활용되고 있음에도 각국의 프라이버시법은 관찰 불가능한 정보의 추론에 대해서는 별다른 규제를 하지 않으면서 관찰 가능한 개인정보의 수집 단계에 규제를 집중하는 경향이 있다. 개인의 온라인 행태정보와 사회적 관계에 관한 정보를 이용하면 관찰된 정보와는 전혀 다른 새로운 개인정보를 추론할 수 있음에도 수집된 정보로부터 추론된 정보의 취급은 법적 회색 지대에 놓여 있다. 개인에 대해 10의 정보를 가지고 100의 정보를 추론할 수 있다면 그렇게 추론된 정보는 현행 개인정보 보호법 하에서 어떠한 보호를 받을 수 있는가? 어떤 보호를 받아야 하는가?

본고에서는 현재기술로 추론할 수 있는 추론개인정보에 대해 살펴본다. 또한, 추론개인정보가 개인정보 보호법상 개인정보에 해당함에도 법률의 미비로 이에 관한 적절한 규제가 이루어지고 있지 않음을 보이고, 민감정보 처리제한 규정을 통한 사전통제, 정보주체의 개인정보 열람권 등을 통한 사후통제가 가능함을 제시한다. 하지만 현행법의 해석만으로는 추론분석(inferential analytics)의 프라이버시 위협에 충분히 대항할 수 없다. 장기적으로 추론개인정보에 대한 체계적인 규제를 확립하기 위한 노력이 필요하다.

II. 추론개인정보

1. 추론개인정보 기술 현황

빅데이터를 활용한 개인적 속성의 추론은 맞춤형 온라인 광고 등 상업적 활용이 매우 활발한 영역이다. 기술기업을 중심으로 개인정보 추론력이 비약적으로 향상되고 있는 점에 비추어 본격적으로 그 프라이버시 침해 위험을 검토할 필요가 있다.

언론 등을 통해 공개된 자료에 따르면 페이스북은 이용자가 일절 공개하지 않은 성적 지향¹⁾, 인종²⁾, 정치적 견해³⁾, 심리 상태⁴⁾ 등을 페이스북 이용내역을 근거로 추론할 수 있으며 이를 기반으로 맞춤형 광고를 집행했다는 의혹을 받은 바 있다. 나아가 페이스북 이용자의 친구 네트워크, “좋아요” 표시 등 소셜네트워크 상에서 이루어진 행태정보를 활용하면 페이스북이 아닌 제3자 또한 이와 유사한 추론을 정확하게 할 수 있다.⁵⁾ (Kosinski, 2013)에 따르면 페이스북 “좋아요”를 통한 분석만으로 [그림1]에 이분법적으로 표현된 변수들에 대해 상당히 정확한 수준의 추론력이 확보되었다. 더욱 놀라운 점은 단 58,000명의 이용자가 어떤 게시물에 “좋아요” 표시를 했는지 만을 가지고 이러한 수준의 개인정보 추론력이 확보되었다는 점이다.

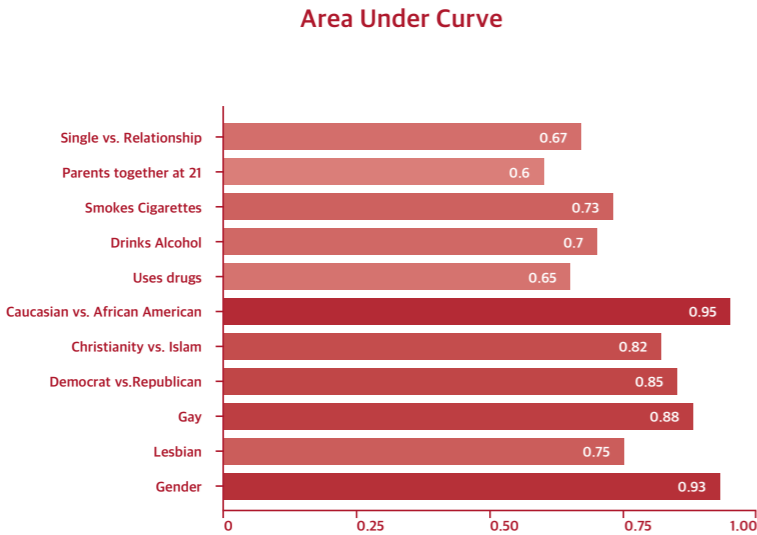


그림 1 [(Kosinski, 2013) 개인적 속성의 추론 정확성]

페이스북 “좋아요”는 이용자와 특정 항목 간에 양의 연관(positive association)이 있음을 나타내는 디지털 기록에 불과하다. 이용자의 검색기록, 콘텐츠

1) José González Cabañas/Ángel Cuevas & Rubén Cuevas, Facebook Use of Sensitive Data for Advertising in Europe, 2018. 2. 14., <https://arxiv.org/abs/1802.05030>

2) Annalee Newitz, Facebook’s Ad Platform Now Guesses at Your Race Based on Your Behavior, ARS TECHNICA, 2016. 3. 18., <https://arstechnica.com/information-technology/2016/03/facebooks-ad-platform-now-guesses-at-your-race-based-on-your-behavior/>

3) Jeremy B. Merrill, Liberal, Moderate or Conservative? See How Facebook Labels You, New York Times, 2016. 8. 23., <https://www.nytimes.com/2016/08/24/us/politics/facebook-ads-politics.html>

4) Sam Machkovech, Report: Facebook helped advertisers target teens who feel “worthless”, ARS TECHNICA, 2017. 5. 1. <https://arstechnica.com/information-technology/2017/05/facebook-helped-advertisers-target-teens-who-feel-worthless/>

5) Michael Kosinski et al., “Private traits and attributes are predictable from digital records of human behavior”, Proceedings of the National Academy of Sciences Apr 2013, 110 (15) 5802-5805

6) Michael Kosinski, “Facial recognition technology can expose political orientation from naturalistic facial images”, Sci Rep 11, 100 (2021)

7) Yilun Wang/Michael Kosinski, “Deep neural networks are more accurate than humans at detecting sexual orientation from facial images”, Journal of Personality and Social Psychology, 114(2), 246-257 (2018)

8) AP News, Chinese ‘gait recognition’ tech IDs people by how they walk, 2018. 11. 6., <https://apnews.com/article/bf75d-d1c26c947b7826d270a16e2658a>

9) WIRED, This Robot Can Guess How You’re Feeling by the Way You Walk, 2020. 5. 18., <https://www.wired.com/story/proxemo-robot-guesses-emotion-from-walking/>

10) Forbes, How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did, 2012. 2. 16., <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=66b08d4a6668>

11) Privacy International, “Examples of Data Points Used in Profiling”, 2017.

12) 이를 협업 필터링(collaborative filtering)이라고 한다.

츠 소비기록, 물품구매기록 역시 특정한 항목과 이용자 사이의 양의 연관을 나타내는 데이터이다. 이러한 데이터를 보유한 개인정보처리자는 (Kosinski, 2013)와 같은 간단한 분석만으로도 개인에 관한 정보를 추론 할 수 있다. 중국은 사회 신용점수시스템(Social Credit Scoring System)을 통해 국가 단위에서 이러한 개인정보 추론 기술을 도입해 활용하고 있다. 한편 온라인 콘텐츠 추천 및 광고를 위해 디지털 행태정보와 사회관계망 데이터를 활용하는 것에서 나아가 추론에 활용되는 근거정보의 범위가 확장되는 현상도 발견된다. 얼굴 사진만으로 정치 성향⁶⁾과 성적 지향⁷⁾을 추론할 수 있다는 연구가 발표되었고, 사람의 걷는 모습 만을 가지고 그 사람을 특정하고,⁸⁾ 나아가 심리상태를 추론할 수 있는 알고리즘⁹⁾도 존재한다.

그 외 대표적인 개인정보 추론 실증 사례로는 2012년 미국의 대형 유통업체 타겟(Target)이 이용자의 물품구매내역 데이터를 이용해 여성 청소년의 임신여부를 추론해 개인화된 할인 쿠폰을 발행한 사례가 있다.¹⁰⁾ 임신 여부와 같은 민감정보를 해당 이용자가 타겟에서 산 물품의 내역으로부터 추론한 것이다. 구체적으로, 임신 3개월에 도달한 여성은 무항 로션을 자주 구매하며 임신 첫 20주 기간에는 칼슘, 마그네슘, 아연 보충제 등을 구매한다는 사실에 기초해 이러한 구매패턴을 보이는 여성을 임신여성으로 추론한 것이다. 소비자의 물품구매기록에 임신사실이라는 민감한 개인정보가 내장되어 있었다고 할 수 있다. Privacy International은 2017년 Examples of Data Points Used in Profiling 보고서를 통해 컴퓨터과학 문헌에 나타난 개인정보 추론 사례들을 집대성하기도 하였는데 대부분 제한된 공개 데이터로 한 분석 임에도 한 종류의 개인정보로부터 전혀 다른 종류의 개인정보를 높은 정확도로 추론하였다.¹¹⁾

2. 추론개인정보에 대한 규제 필요성

이상에서 살펴본 바와 같이 적법하게 보유한 근거정보로부터 근거정보와는 전혀 다른 별개의 추론개인정보를 추론할 수 있는 기술은 이미 존재한다. 이러한 알고리즘의 가장 기초적인 작동 방식은 이용자A와 이용자B가 여러 항목에서 유사하다는 관찰결과로부터 이미 알려진 이용자A의 속성을 이용자B도 가지고 있다고 추론하는 것이다.¹²⁾ 추론은 개인정보처리자 내부에서 행해지므로 대상이 되는 정보주체는 통상 자신에 대해 어떤 추론이 이루어지고 있는지 알 길이 없으며(투명성 문제) 왜 그러한 추론이 이루어졌는지 이유를 알 수 없는 경우가 많다(설명 가능성 문제).

나아가 추론개인정보를 통해 규제받지 않는 형태로 민감정보가 처리될 가능성도 존재한다. 개인정보 보호법 제23조 제1항은 정보주체의 별도의 동의 또는 법령에 근거하지 않고서는 민감정보를 처리할 수 없다고 규정하나, 각 개인적 속성들 간의 상관관계를 이용하면 민감정보를 쉽게 추론할 수 있다(대리변수의 문제). 가령 앞서 살핀 타겟(Target) 사례에서와 같이 구매내역정보를 통해 임신 여부를 쉽게 추론할 수 있다면 구매내역정보 수집 및 이용에 대한 일반적인 규

제만으로는 민감정보가 보호된다고 할 수 없다. 민감정보가 아닌 근거정보로부터 민감정보인 추론개인정보를 생성하는 행위에 대한 규제가 있어야만 민감정보 처리를 제한하는 개인정보 보호법 제23조의 입법목적은 달성할 수 있다.

III. 추론개인정보에 대한 개인정보 보호법의 적용

1. 추론된 정보는 개인정보인가?

개인정보 보호법, 캘리포니아 소비자 프라이버시법(California Consumer Privacy Act; “CCPA”), EU 일반정보보호규정(General Data Protection Regulation; “GDPR”)은 모두 개인정보(personal information 또는 personal data)를 보호대상으로 하고 있다. 개인에 관하여 추론된 정보가 개인정보에 해당하지 않는다면 이들 법이 적용될 여지가 없으므로 추론된 정보의 개인정보성을 먼저 검토한다.

가. 해외 입법례

CCPA는 소비자의 기호, 특성, 심리상태, 선호 등을 반영한 소비자 프로파일링을 만들기 위해 개인정보로부터 추론한 정보를 명시적으로 개인정보(personal information)의 일종으로 규정하였다.^{13) 14)} 이때 추론 행위는 “사실, 증거 기타 정보 또는 데이터로부터 정보, 데이터, 가정, 결론을 얻는 행위”로 넓게 정의된다.¹⁵⁾ 새롭게 도입된 조항이라 달리 해석될 여지는 있겠으나 CCPA 하에서 정보주체는 추론개인정보에 대해 알 권리(right to know), 접근권(right to access), 삭제권(right to deletion), 거부권(right to opt out) 등을 행사할 수 있다.

CCPA와 달리 GDPR은 개인정보(personal data)의 정의에 개인에 관하여 추론된 정보를 직접 포섭하고 있지는 않다. 하지만 GDPR 제2조에서 개인정보는 식별된 또는 식별가능한 개인에 관한 일체의 정보(any information relating to an identified or identifiable natural person)로 폭넓게 정의되며, 유럽연합 제29조 작업반(Article 29 Working Party; “WP29”)에 따르면 정보의 내용이 객관적일 필요 없이 주관적인 정보, 의견, 평가(subjective information, opinions, assessments)도 개인정보에 포함된다.¹⁶⁾ 한편, GDPR이 추론개인정보를 과소보호하고 있다는 비판도 존재한다.¹⁷⁾ (Wachter, 2019)는 GDPR은 정보주체에게 자신의 개인정보를 이용해 자신에 대한 추론을 하는 것에 대해 제한된 지배권만을 허용하며, 민감정보 추론 및 추론 결과에 대한 이의제기권 등 보호가 부족하다고 비판하였다. 나아가 정보주체에게 “합리적인 추론 요구권(right to reasonable inference)”을 부여하여 고위험 추론에 대한 책임소재를 명확히 하여야 한다고 주장하였다.

GDPR 법문상 추론개인정보가 개인정보에 해당하는지 여부가 불분명한 가운데 WP29는 추론개인정보도 개인정보로 취급되어야 한다는 입장인 것으로 보인다. WP29 자동화된 의사결정 및 프로파일링에 관한 가이드라인은 프로파

13) California Consumer Privacy Act, 1798.140 (o) (1) “Personal information” means information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household. Personal information includes, but is not limited to, the following if it identifies, relates to, describes, is reasonably capable of being associated with, or could be reasonably linked, directly or indirectly, with a particular consumer or household: [...] (K) Inferences drawn from any of the information identified in this subdivision to create a profile about a consumer reflecting the consumer’s preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes.

14) 한편, 2020년 11월 California Privacy Rights Act(“CPRA”)가 제정되면서 민감정보 개념과 개인정보 정정권이 도입되는 등 CCPA를 보다 강화하였다. CPRA는 2023년 1월 시행예정이다.

15) California Consumer Protection Act, 1798.140 (m).

16) WP29, Opinion 4/2007 on the concept of personal data, 2007. 6. 20., 6면.

17) Sandra Wachter/Brent Mittelstadt, “A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI”, Columbia Business Law Review, 2019(2).

18) WP29, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, 8면.

19) WP29, 앞의 가이드라인, 17면.

20) WP29, Guidelines on the right to data portability, 9면.

21) 개인정보보호위원회, 개인정보 보호 법령 및 지침·고시 해설서, 2020. 12., 81면.

22) 박광배/채성희/김현진, “빅데이터 시대 생성 정보의 처리 체계 - 추론된 정보의 처리에 관한 우리 개인정보보호법의 규율과 개선 방안에 관한 고찰”, 정보법학 제21권 제2호(2017), 175면; 최경진, 온주 개인정보보호법(2016. 7. 6.) 제15조.

일링의 대상이 되는 개인정보를 개인이 제공한 정보(data provided directly by the individuals), 관찰된 정보(data observed about the individuals), 추론된 정보(inferred data such as profile of the individual that has already been created)로 구분하고 있다.¹⁸⁾ 이 가이드라인은 프로파일링의 결과로 생성된 추론 정보가 개인에 의해 직접 제공되지 않은 “새로운” 개인정보임을 명확히 한다.¹⁹⁾ WP29 데이터 이동권에 대한 가이드라인 역시 GDPR 제20조 제1항에서 정보주체가 제공한 정보(data provided by the individuals)만을 데이터 이동권의 대상으로 삼은 취지를 설명하면서 비록 추론된 정보(inferred data)가 개인정보이기는 하나 컨트롤러(controller)에 의해 생성된 정보인 만큼 정보주체의 데이터 이동권 범위에서 제외된다고 명시하였다.²⁰⁾

나. 국내법

국내법이 추론개인정보를 개인정보로 취급하는지는 역시 불분명하다. 개인정보 보호법 제2조 제1호는 개인정보를 살아 있는 개인에 관한 정보로서 (1) 개인을 알아볼 수 있는 식별정보, (2) 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 식별가능정보, (3) 가명정보로 정의한다. 개인정보 보호법은 추론된 정보에 대해 별도의 규정을 두고 있지 않으므로 추론된 정보가 개인정보에 해당하는지 여부는 개인정보처리자를 기준으로 한 식별가능성에 달려있다. 현재 맞춤형 콘텐츠 및 광고 제공의 목적으로 활용되고 있는 추론개인정보는 대부분의 경우 별다른 무리없이 이러한 식별가능성 요건을 만족할 것이다.

개인정보보호위원회 유권해석은 추론개인정보의 개인정보성을 전제로 한 입장인 것으로 보인다. 개인정보보호위원회 해설서에 따르면 “개인정보 ‘수집’이란 정보주체로부터 직접 이름, 주소, 전화번호 등의 정보를 제공받는 것뿐만 아니라 정보주체에 관한 모든 형태의 개인정보를 취득하는 것을 말한다 ... 개인정보처리자가 직접 수집하지 않아도 업무처리 과정에서 개인정보가 생산되거나 생성된 경우도 적지 않다.”²¹⁾ 추론개인정보는 개인정보처리자가 업무처리 과정에서 생산·생성한 개인정보에 해당하며 이 경우 추론이라는 개인정보 취득 과정은 개인정보 보호법상 “수집”에 해당한다는 취지이다. 학설도 대체로 개인정보보호위원회의 입장과 같다.²²⁾

한편, 추론개인정보는 알고리즘이 확률론적 추론에 기반해 생성한 정보이기 때문에 “정보성”을 인정할 수 없다는 반론도 가능하다. 가령 특정 이용자가 남성인지 여성인지를 확정적으로 표시하는 통상의 개인정보와 달리 추론개인정보는 ‘95% 확률로 남성이다’와 같은 확률론적 정보라는 것이다. 하지만 이러한 견해는 정보의 개념을 ‘불확실성(uncertainty)의 제거’로 이해하는 정보 이론(Information Theory)에 반할 뿐만 아니라, 만일 정보의 정확성을 정보성의 요건으로 이해한다면 오래되어 변경내역을 반영하지 못했거나 각종 오류로 인해 부정확한 정보는 모두 정보성이 부정되어야 한다. 즉, 정확한 정보만이 정보라고 해

석한다면 통상의 방법으로 수집한 각종 개인정보에도 오류의 가능성이 있다는 문제를 해결할 수 없다. 개인정보 보호법이 개인정보가 부정확할 가능성을 전제로 정보주체의 정정권(제36조)을 규정하고 있는 점을 보아도 이러한 해석은 체계정합적이지 않다. 정보의 정확성은 정보성의 구성요소로 볼 수 없으므로 확률적 추론도 정보성을 인정해야 할 것이다.²³⁾ 개인정보보호위원회도 정보가 반드시 '사실'이거나 '증명된 것'이 아닌 부정확한 정보 또는 허위의 정보일지라도 특정한 개인에 관한 정보이면 개인정보가 될 수 있다는 견해이다.²⁴⁾

추론개인정보를 개인정보로 인정할지 여부는 깊은 고민을 필요로 한다. 이하에서 살펴보듯이 추론개인정보가 온전히 개인정보로 인정될 경우 그 생성과 이용에 관하여 개인정보 보호법이 전면적으로 적용되어야 하기 때문이다. 개인정보 개념을 차등화하여 각기 다른 법률효과를 부여하자는 주장도 제기된 바, 이를 진지하게 검토할 필요가 있다.²⁵⁾

2. 추론개인정보에 대한 규제체계

앞서 살핀 바와 같이 개인정보위원회 해설서는 개인정보 생성(추론) 행위를 포함해 개인정보처리자가 개인정보를 취득하는 모든 행위를 개인정보 수집으로 규정한다.²⁶⁾ 학설 중에는 추론개인정보의 생성이 개인정보 수집이라는 전제 하에 근거정보를 이용해 추론개인정보를 생성하는 행위의 적법성은 제1단계로서 근거정보를 새로운 추론개인정보의 생성 목적으로 이용하는 것이 허용되는지(근거정보의 이용)에 대한 검토와 더불어 제2단계로서 새로운 정보의 생성(수집) 행위 자체가 허용되는지(추론개인정보의 수집)를 검토하여야 한다는 견해가 있다.²⁷⁾ 이 견해에 따르면 개인정보처리자는 추론개인정보 생성시에 정보주체의 동의 등 개인정보 보호법 제15조 제1항 각호에 따른 요건을 충족하여야 한다.

하지만 추론개인정보의 취득을 개인정보 수집으로 관념하는 것은 법체계상 적절하지 않으며 개인정보 이용 현실에도 부합하지 않는다고 생각된다. 첫째, "수집하다"의 사전적 정의는 "(취미나 연구를 위하여) 여러 가지 물건이나 재료를 찾아 모으다" 이므로,²⁸⁾ 개인정보의 수집이란 기존에 존재하는 개인정보를 취득하는 것을 의미한다고 해석하는 것이 문리해석에 부합한다. 이에 반해 추론개인정보는 근거정보로부터 새롭게 생성된 개인정보이다. 둘째, 개인정보 보호법 제2조 제2호는 개인정보 "처리"를 "수집, 생성, ... 그 밖에 이와 유사한 행위"로 구분한다. 즉, 개인정보의 수집은 생성과 법률상 구분되는 개념이다. 추론개인정보를 취득하는 과정은 수집보다 생성에 해당한다. 셋째, 개인정보의 추론을 수집이라 인식한다면 개인정보처리자가 근거정보를 이용해 추론개인정보를 생성할 때 마다 개인정보 보호법 제15조에 따른 고지와 정보주체의 동의를 득할 의무를 지게 된다. 이러한 결과는 추론개인정보 활용 현실에 부합하지 않으며 개인정보처리자에게 이를 요구하는 것은 추론개인정보의 활용가능성을 극도로 저하시킨다. 실제 맞춤형 콘텐츠 추천 및 광고 등을 위해 추론개인정보를 이용하고 있는

23) 물론 확률론적 정보의 정보가치(informational value)가 0인 경우도 있다. 가령 일반적인 서비스의 이용자에 대해 "이용자가 남성일 확률은 50%이다"와 같은 정보는 정보 가치가 없으며 아무런 근거정보 없이도 생성할 수 있다. 이러한 경우 정보성을 부정할 여지가 있다고 생각한다.

24) 개인정보보호위원회, 앞의 해설서, 11면.

25) 이성엽, "개인정보의 개념의 차등화와 개인정보 이동권의 대상에 관한 연구", 경제규제와 법, 12(2).

26) 개인정보보호위원회, 앞의 해설서, 81면.

27) 박광배/채성희/김현진, 앞의 글(주22) 179면.

28) 표준국어대사전

것으로 예상되는 기업들의 개인정보처리방침에는 추론개인정보의 "수집"을 전제로 한 조항은 발견되지 않는다.

현재 추론개인정보를 생성하기 위해 근거정보를 이용하는 목적은 대부분 추론개인정보를 통해 맞춤형 콘텐츠 추천 및 광고를 위해서일 것이다. 맞춤형 콘텐츠 및 광고 제공 과정에서 개인정보 보호법의 규제대상으로 포착되는 개인정보의 처리행위는 맞춤형 콘텐츠 및 광고 제공을 위한 근거정보 이용에 국한된다([그림2] 참조). 주요 기술기업의 개인정보처리방침이 추론개인정보를 언급하지 않는 점에 비추어 추론개인정보의 생성은 별도의 독립된 개인정보 처리로 인식되지 않고, 맞춤형 콘텐츠 및 광고 제공을 위한 근거정보 이용의 부산물 정도로 인식되는 것으로 보인다. 이들 기업이 추론개인정보를 처리(생성)하는 법적 근거는 근거정보 수집 시 고지된 이용목적에 포괄적으로 해석하거나 개인정보 보호법 제15조 제3항에 따라 당초 수집 목적과 합리적으로 관련된 범위 내 개인정보 이용으로 구성할 것으로 예상된다. 다시 말해 추론개인정보의 생성에 대한 법적 통제는 근거정보의 이용목적에 의존하며 근거정보가 이 이용목적내에서 이용된다면 어떠한 추론개인정보가 생성되는지는 개인정보 보호법이 규제하지 못한다.

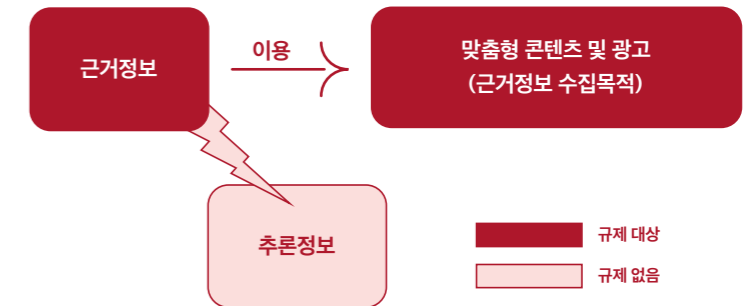


그림 2 현행 개인정보처리과정의 이해

하지만 맞춤형 콘텐츠 및 광고 제공을 위한 개인정보의 처리의 실질을 살펴보면 [그림2]가 아니라 [그림3]과 같이 인식해야 한다. 개인정보처리자는 먼저 근거정보를 기반으로 이용자에 관한 추론개인정보를 생성하며(이용자 프로파일) 이 추론개인정보와 근거정보를 동시에 이용하여 맞춤형 콘텐츠와 광고를 제공한다. 한편, 생성된 추론개인정보는 즉시 폐기되지 않고 다음 서비스 제공을 위

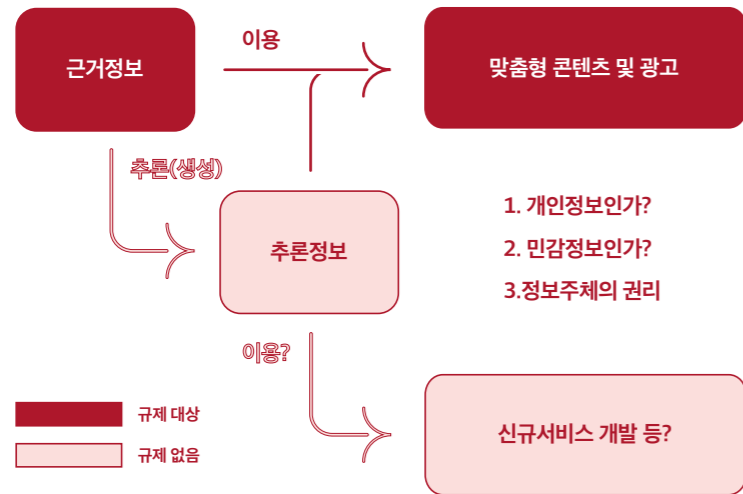


그림 3 개인정보처리과정의 새로운 이해

해 존속할 수 있으며 이용자의 행태정보 등 추가되는 근거정보를 기반으로 지속적으로 업데이트 된다. 많은 경우 이렇게 저장된 추론개인정보는 신규서비스 개발에도 이용될 것으로 추정된다.

이와 같이 추론개인정보의 생성을 근거정보의 이용과는 별개의 과정으로 이해해야 하는 가장 중요한 이유는 추론개인정보가 수집된 근거정보보다 민감한 정보인 경우가 많기 때문이다. 가령 타겟(Target)의 사례에서와 같이 구매 내역 정보는 그 자체로 민감정보가 아니지만 이로부터 임신사실을 추론할 수 있다면 추론된 정보는 민감정보에 해당한다. 근거정보를 맞춤형 콘텐츠 및 광고 제공이라는 수집목적 범위 내에서 이용하는 행위만을 규제대상으로 인식한다면 그 과정에서 개인정보처리자가 정보주체의 임신사실이라는 민감정보를 취득했다는 사실은 주목 받지 못한다. WP29 역시 수집된 근거정보가 아닌 추론된 개인정보가 개인에 관한 보다 민감한 정보를 담고 있을 가능성이 높음을 우려한 바 있다.²⁹⁾

추론개인정보를 생성하고 이용하는 과정을 별개의 개인정보 처리로 이해한다면 이에 대한 규제를 신설할 필요가 있다는 결론에 이르게 된다. 이때 추론개인정보에 대한 규제는 추론개인정보를 생성하는 과정을 개인정보의 수집으로 보아 정보주체의 동의를 득할 것을 요구하는 방식일 수 없다. 이미 개인정보처리자의 지배영역에 들어온 근거정보의 이용을 규제하는 것이기 때문에 추가적인 동의 과정을 요구하는 것은 자연스럽지 않다. 또한 근거정보 수집 시에 정보주체에게

29) WP29, Opinion 03/2013 on Purpose Limitation, 47면.

제공할 이용목적 고지 내용에 추론으로 생성할 개인정보 내역을 추가하는 방식도 적절하지 않다. 근거정보로부터 추론가능한 정보는 데이터 축적과 기술발전 에 따라 얼마든지 확장될 수 있기 때문이다. 구체적인 내용은 본고의 범위를 벗어나지만 기본적으로 고위험 추론행위 금지 등의 사전통제, 그리고 추론개인정보에 대한 정보주체의 권리보장을 중심으로 규제를 설계해야 한다고 생각한다.

3. 민감정보인 추론개인정보의 처리

고위험 추론행위에 대한 사전통제 방안으로 추론개인정보가 민감정보인 경우 민감정보 처리제한에 관한 조항(개인정보 보호법 제23조)을 적용하는 방안을 검토할 수 있다. 개인정보 제23조 제1항은 개인정보처리자가 정보주체에게 별도의 동의를 받거나 법령에서 민감정보의 처리를 허용하는 경우 외에는 민감정보를 처리하지 못하도록 규정한다. 개인정보 보호법의 “처리”는 “생성”을 포함하는 개념이므로 민감정보를 생성하는 경우 개인정보 보호법 제23조가 적용되는 것으로 해석될 수 있다. 즉, 달리 법률이 허용하지 않는 경우 개인정보처리자가 민감정보인 추론개인정보를 생성하기 위해서는 정보주체의 별도 동의를 받아야 한다. 이상의 해석을 그대로 적용한다면 추론개인정보 생성에 대한 강력한 사전통제로 작동할 수 있으나 현실에서는 추론개인정보의 생성 시에 개인정보 보호법 제23조를 적용해 별도의 동의를 구하는 실무는 존재하지 않는다.

한편, 방송통신위원회가 2017. 2. 7. 발간한 온라인 맞춤형 광고 개인정보 보호 가이드라인은 이용자의 동의 없이 행태정보를 이용·분석하여 구 정보통신망법 제23조 제1항에 따른 민감정보를 수집·생성·활용하여서는 아니된다고 규정하였다.³⁰⁾ 본 가이드라인을 해석함에 있어 개인정보 보호법에 흡수되기 전의 구 정보통신망법 제23조 제1항은 개인정보 보호법 제23조와는 달리 민감정보의 “처리”가 아니라 “수집”을 제한하고 있었다는 점에 주목하여야 한다. 즉, 방송통신위원회가 온라인 맞춤형 광고 개인정보보호 가이드라인을 통해 법률이 예정하지 않은 민감정보의 생성을 규율하고자 시도한 것으로 평가할 수 있다. 이 가이드라인은 개인정보보호 업무 소관이 개인정보보호위원회로 이전된 현재에도 여전히 효력이 유지되고 있는 것으로 보인다.

민감정보인 추론개인정보의 처리 제한에 관해서는 EU의 접근을 참고할 필요가 있다. GDPR 제9조는 개인정보 보호법과 마찬가지로 민감정보(special categories of data)의 처리를 제한하고 있는데 그 뉘앙스는 개인정보 보호법과 미묘하게 다르다. 즉, GDPR 제9조는 인종 또는 민족, 정치적 견해, 종교적 또는 철학적 신념, 노동조합의 가입여부 등을 “드러내는(revealing)” 정보의 처리를 제한한다. “개인정보를 드러내는(personal data revealing)”이라는 표현은 인종 등 보호 변수를 직접 또는 간접적으로 알게 하는 정보의 처리를 제한하려는 의도로 해석된다.³¹⁾ 다시 말해 그 자체로는 민감정보가 아닌 근거정보를 통해 민감정보를 추론할 수 있는 경우라면 그 근거정보 역시 민감정보로 취급한다고 할 수 있다.^{32) 33)} European Data Protection Board 역시 2019. 3. 13. 성명에서 개인의

30) 방송통신위원회, 온라인 맞춤형 광고 개인정보 보호 가이드라인

31) WP29, Advice Paper on Special Categories of Data, 6면, 10면

32) Sandra Wachter/Brent Mittelstadt, 앞의 글(주17), 72면

33) Tal Zarsky, "Incompatible: The GDPR in the Age of Big Data", Seton Hall Law Rev. 47, (2017)

성격, 심리상태, 정치적 성향 등 민감정보를 예측하는 도구가 정치적 목적을 위해 이용되는 것을 우려하며 정치적 성향을 드러내는 개인정보는 GDPR상 민감정보에 해당하기 때문에 이에 관한 규제를 적용 받는다고 지적하였다.³⁴⁾ 이를 종합하면 GDPR 하에서는 민감정보를 추론할 수 있게 하는 근거정보 역시 민감정보로 취급될 가능성이 있음을 의미한다.

그러나 근거정보로부터 추론 가능한 추론개인정보의 내용과 범위는 사전에 확정할 수 없다. 알고리즘의 고도화, 학습데이터의 축적, 새로운 근거정보의 추가 등을 통해 미래에 그 범위가 확장될 수 있기 때문에 근거정보에도 민감정보 처리제한에 대한 규제를 전면적으로 적용한다면 잠재적으로 모든 개인정보를 민감정보로 취급해야 하는 문제가 발생할 수 있다. 이에 민감정보의 범위가 무제한 확장되는 것을 제한하기 위해 정보처리의 목적과 추론의 신뢰성을 요건으로 하여 이들 요건을 만족하는 경우에만 민감정보 추론의 근거가 되는 근거정보를 민감정보로 볼 수 있다는 논의가 이루어진 바 있다.³⁵⁾ 유럽사법재판소(Court of Justice of the European Union)는 개인이 특정 유럽의회 의원 사무실 직원으로 근무하였다는 사실 만으로는 해당 개인의 정치적 성향을 신뢰성 있게 추론할 수 있는 근거정보가 될 수 없으므로 해당 근무사실 정보는 민감정보가 아니라고 판시한 바 있다.³⁶⁾

EU의 논의와 같이 민감정보를 신뢰성 있게 추론 가능한 경우 그 기초가 되는 근거정보 자체를 민감정보로 보는 접근법이든, 개인정보 보호법 법문 그대로 민감정보의 생성(추론)을 법률이 제한한 민감정보 처리의 일환으로 보는 접근법이든 추론을 통한 민감정보의 처리에는 일정한 제약을 가할 필요가 있다. 다만 민감정보 처리제한에 관한 규정을 모든 추론개인정보에 그대로 적용하는 것은 추론개인정보 활용 현실에 부합하지 않는다. 한가지 검토할 수 있는 입법방안으로는 2021. 4. 21. 제안된 EU Artificial Intelligence Act가 인공지능 시스템의 사용목적(intended purpose)을 기준으로 규제대상 고위험 인공지능 시스템을 정의한 것과 같이³⁷⁾ 추론개인정보의 사용목적에 따라 고위험 추론개인정보를 정의하여 통제장치를 마련하는 방안을 검토할 수 있다.

4. 추론개인정보에 대한 정보주체의 권리

가. 개인정보 전송 요구권

추론개인정보의 존재를 직접적으로 인식하고 이에 대한 법률효과를 부여하는 법조항으로는 개인정보 전송 요구권이 있다. 개인정보보호위원회가 2021. 1. 6. 입법예고한 개인정보 보호법 일부개정법률(안)은 일반적인 개인정보 전송 요구권(제35조의 2) 도입을 예고하였다. 이 조항은 정보주체가 개인정보처리자에 대하여 자신의 개인정보를 본인, 또는 다른 개인정보처리자 등에게 전자적으로 처리 가능한 구조화된 형식으로 전송할 것을 요구할 수 있는 권리로 GDPR 제 20조(Right to data portability), 신용정보법 제33조의2(개인신용정보의 전송

34)

European Data Protection Board, Statement 2/2019 on the use of personal data in the course of political campaigns, 2019. 3. 13.

35)

Sandra Wachter/Brent Mittelstadt, 앞의 글(주17), 73면이하

36)

Case T-190/10, Kathleen Egan & Margaret Hackett v. European Parliament, 2012 E.C.R. I-165.

37)

Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Article 6, Annex 3

38)

Paul De Hert et al., "The right to data portability in the GDPR: Towards user-centric interoperability of digital services", Computer Law & Security Review Volume 34, Issue 2, 2018. 4., p. 199

39)

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation); Document 52012PC0011: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52012PC0011>

40)

WP29, Guidelines on the right to data portability(주20)는 이를 명백히 한다. 3면 ("Since it allows the direct transmission of personal data from one data controller to another, the right to data portability is also an important tool that will support the free flow of personal data in the EU and foster competition between controllers. It will facilitate switching between different service providers and will therefore foster the development of new services in the context of the digital single market strategy.")

요구)와 대체로 유사한 내용으로 입법될 것으로 예상된다. GDPR은 전송대상 개인정보를 정보주체가 제공한 정보로 한정하여 추론개인정보를 전송대상에서 제외하고 있으며, 신용정보법 역시 신용정보제공·이용자등이 개인신용정보를 기초로 별도로 생성하거나 가공한 신용정보를 전송대상 개인신용정보의 범위에서 제외하고 있다.

GDPR 이 개인정보 이동권의 대상에서 추론개인정보를 제외한 이유는 컨트롤러의 지적재산권 등 경제적 이익을 보호하기 위함인 것으로 이해된다. 특히 온라인사업자가 복잡한 알고리즘을 통해 정보주체(소비자)에 대해 추론한 정보를 경쟁 사업자가 무상으로 취득하는 것을 방지하고자 하는 목적이 가장 크다고 할 수 있다.³⁸⁾ 이는 GDPR 원인³⁹⁾ 이 개인정보 이동권의 대상을 "처리되는 정보(data undergoing processing)"로 규정하였으나 최종안에서는 그 범위를 "정보주체가 컨트롤러에게 제공한 개인정보(personal data concerning him or her which he or she has provided to the data controller)"로 축소한 점에서 알 수 있다. 추론개인정보의 생성에는 개인정보처리자의 노력이 투입되는 만큼 이러한 정보를 경쟁사업자에게 제공하지 않을 정당한 이익을 인정할 수 있으므로 개인정보 전송 요구권의 범위에서 추론개인정보를 제외하는 입법은 타당하다고 생각된다. 개인정보 보호법에 개인정보 전송 요구권을 도입함에 있어서도 이러한 취지를 반영할 필요가 있다.

나. 개인정보 열람권 등

개인정보 보호법은 개인정보주체의 개인정보 열람권(제35조), 정정·삭제권(제36조), 처리정지요구권(제37조)을 규정하고 있는데, 추론개인정보에 대해서도 이러한 정보주체의 권리를 인정할 수 있을지 문제된다. 우선, 현행 개인정보 보호법은 제35조 제4항 각호의 사유가 없는 한 정보주체로 하여금 개인정보를 열람할 수 있도록 하여야 한다고 규정하므로 추론개인정보가 개인정보에 해당하는 이상 개인정보처리자가 정보주체의 추론개인정보 열람요구를 적법하게 거절할 방법은 없다고 할 것이다. 입법론적으로도 개인정보 열람권의 경우 개인정보 전송 요구권과는 달리 개인정보처리자의 지적재산권 보호 및 경쟁 사업자의 무임승차(free-riding) 방지를 고려할 필요가 없으므로 개인정보 열람권의 대상을 제한해야 할 필요가 없다. 개인정보 열람권은 정보주체와 개인정보처리자 양자간의 문제인 반면 개인정보 전송 요구권은 (대부분의 경우) 개인정보를 수령하는 제3의 개인정보처리자가 등장하는 삼자관계라는 차이가 있기 때문이다. 나아가 개인정보 열람권은 순수하게 개인정보자기결정권에서 파생되는 권리라고 볼 수 있는 반면 개인정보 전송 요구권은 개인정보자기결정권과 더불어 개인정보가 특정 사업자에게 lock-in 되는 것을 방지해 사업자간 경쟁을 촉진하기 위한 정책목표⁴⁰⁾도 반영한 입법이다. 따라서 헌법상 기본권인 개인정보자기결정권의 발현인 개인정보 열람권 등의 범위를 제한함에 있어서는 개인정보 전송 요구권의 범위를 제한하는 경우보다 신중할 필요가 있다.

개인정보주체가 추론개인정보에 대한 열람권 등을 행사하면 개인정보처리자에게는 상당한 부담으로 작용할 소지가 있다. 특히 추론개인정보의 특성상 개인정보처리자가 개인정보를 “저장”하고 있지 않은 경우가 많을 수 있고, 저장된 형태의 데이터가 존재하는 경우에도 행태정보 등 근거정보를 반영해 수시로 업데이트 된다는 특성이 있으며, 추론개인정보를 배경지식이 없는 일반인이 이해할 수 있는 형태로 변환하기 위해서는 상당한 노력이 투입되어야 할 것이다. 나아가 개인정보처리자가 보유한 추론개인정보에는 민감한 내용이 포함되어 있을 가능성이 높는데, 앞서 살핀 대로 적절한 규제의 부재로 인해 이러한 민감정보의 처리가 가능했던 측면이 있다. 개인정보처리자가 자신에 대해 민감한 정보를 추론하고 있다는 사실을 인지한 소비자들의 반발도 예상된다.

그럼에도 추론개인정보에 대한 개인정보 열람권 행사는 적극 장려되어야 한다. 인공지능 기술이 발달할수록 개인정보처리자는 훨씬 정확하게 민감한 개인정보를 추론할 수 있게 될 것인 반면 현행법상 추론개인정보의 처리를 사후적으로라도 통제할 수 있는 수단은 개인정보 열람권 뿐이기 때문이다. 열람권은 일종의 사후적 투명성 요건으로 부적절한 개인정보 처리(추론)에 대한 민간차원의 압력으로 작용할 수 있을 것이다. 시민단체 등의 임팩트 소송(impact litigation)을 통해 추론개인정보의 법적 성격과 이에 대한 정보주체의 권리를 명확히 하는 방안도 검토할 수 있다. 장기적으로는 개인정보 보호법 하에 추론개인정보의 생성, 이용에 관한 규제와 더불어 합리적 추론요구권(right to reasonable inference)⁴¹⁾ 등 정보주체의 권리를 신설할 필요가 있다.

IV. 결론

추론개인정보 개념을 사용하지는 않았지만 동일한 문제의식에 기초한 규제가 다수 제안된 바 있다. 개인정보보호위원회가 2021. 1. 6. 입법예고한 개인정보 보호법 개정안 제37조의2는 자동화된 개인정보 처리에만 의존하여 개인에게 법적 효력 또는 중대한 영향을 미치는 의사결정을 정보주체가 거부, 이의제기, 설명요구할 수 있는 권리를 신설하였다. 개정안 제37조의2는 추론개인정보 이용 단계에서 이루어지는 중대한 의사결정에 한하여 정보주체의 배제권 등 권리를 신설한 것으로 볼 수 있다. 공정거래위원회가 2021. 3. 5. 입법예고한 전자상거래법 개정안 제18조는 온라인판매사업자에게 맞춤형 광고 고지의무를 부여하고 소비자가 맞춤형 광고 대신 일반광고를 선택할 수 있도록 하여야 한다고 규정한다. 맞춤형 광고는 대표적인 추론개인정보를 이용하는 행위인데, 본 규정은 광고와 소비자보호라는 특수한 맥락에서 정보주체가 자신의 추론개인정보가 이용되는 것을 인지하고 opt-out할 수 있는 권리를 신설한 것으로 볼 수 있다.

개인정보 보호와 소비자 보호라는 별개의 영역에서 제안된 것처럼 보이는 이들 규제는 개인에 대한 추론분석(inferential analytics)에 대항할 수 있는 정보주체·소비자의 권리를 신설 했다는 공통점이 있다. 이러한 권리를 “추론개인

정보에 대한 개인정보자기결정권”으로 일반화 및 확장할 필요가 있다. 방대한 데이터의 축적, 연산능력의 지수적 발전, 데이터 경제가 약속하는 부가가치로 인해 미래의 프라이버시 위협은 추론분석에서 비롯될 가능성이 매우 높기 때문이다. 현재 법적 회색지대에 놓여져 있는 개인정보 추론행위를 규제영역 내로 포섭해야만 추론분석에 의한 프라이버시 위협에 대항할 수 있다.

41)

Sandra Wachter/Brent Mittelstadt, 앞의 글(주17)