

판례평석

인공지능채팅 서비스의 개인정보보호법 위반으로 인한 제재처분 사례 분석

- 비정형 데이터 수집·이용에 관한 법적쟁점 -

Case Study of the AI Chatting Service's Violation of the PIPA

- Legal Issues concerning the Collection and Use of Unstructured data -

박 규 홍(Kyuhong Park)*

목차

- I. 사안의 배경
- II. 개보위의 인공지능채팅 서비스에 대한 제재처분의 요지
- III. 주요 법적 쟁점
- IV. 비정형 데이터 활용시 규제리스크 해소를 위한 방안
- V. 결론

I. 사안의 배경

개인정보보호위원회(이하 '개보위'라 한다)는 2021. 4. 28. 인공지능(이하 'AI') 챗봇 '이루다'의 개발사인 (주)스캐터랩(이하 '피심인')의 카카오톡 대화 등 비정형 데이터 처리에 대해 과징금과 과태료를 합산하여 총 1억 330만 원의 제재금을 부과하였다. 개보위는 (i) 이루다의 개발운영과 관련하여 카카오톡 대화 수집 및 이용 목적을 구체적으로 특정하지 아니하고 단순히 '신규 서비스 개발'로만 기재하여 동의를 받은 것은 적법한 동의로 보기 어렵고 또한 (ii) '이루다'의 서비스 운영 시 사용되는 응답 DB 내 카카오톡 대화를 가명 처리를 하지 않고 공유한 것도 개인정보보호법 위반으로 판단하였다.

위 사건에서 문제가 된 카카오톡 대화내용은 정

형화되지 않은 데이터, 이른바 '비정형 데이터'로 분류할 수 있다. 여기서 비정형 데이터란 일정한 규격이나 형태를 지닌 숫자데이터와 달리 형태와 구조가 규격화되지 않은 음성·영상·사진·문자 등의 데이터를 의미 하는데, 정보처리 용량 및 속도의 향상과 인공지능(이하 'AI') 기술의 발달로 이러한 비정형 데이터의 활용도는 급격히 증가하고 있다. 개보위의 인공지능채팅 서비스 관련 제재 처분은, AI 알고리즘 개발을 위해 방대한 학습 데이터가 필요한 기업에 적지 않은 부담으로 작용할 것으로 보인다.

아래에서는 위 사건과 관련한 법적 쟁점을 살펴보고, 그 해결책을 모색하고자 한다.

II. 개보위의 인공지능채팅 서비스에 대한 제재처분의 요지

개보위 조사 결과, 피심인은 ① '이루다' 인공지능(AI) 모델의 개발을 위한 알고리즘 학습 과정에서, 카카오톡 대화에 포함된 이름, 휴대전화번호, 주소 등의 개인정보를 삭제하거나 암호화하는 등의 조치를 전혀 하지 않고, 약 60만 명에 달하는 이용자의 카카오톡 대화문장 94억건을 이용하였다. ②

* 법무법인(유한) 세종 파트너변호사.

‘이루다’ 서비스 운영 과정에서는, 20대 여성의 카카오톡 대화문장 약 1억 건을 응답 DB로 구축하고, ‘이루다’가 응답 DB에 저장된 대화문장 중 한 문장을 선택하도록 운영하였다. 그 결과 피심인은 자사의 앱 서비스인 ‘텍스트앳’과 ‘연애의 과학’에서 수집한 카카오톡 대화를 2020. 2.부터 2021. 1.까지 페이스북 이용자 대상의 챗봇 서비스의 AI 개발 및 운영에 이용한 것으로 확인되었다.

개보위의 조사에 따르면, 피심인은 선행 서비스인 ‘텍스트앳’과 ‘연애의 과학’ 개인정보처리방침에 ‘신규 서비스 개발’이 포함시키고, 이용자가 로그인함으로써 동의한 것으로 간주하는 방식으로 이용자가 ‘이루다’ 서비스 개발 및 운영에 이용자의 카카오톡 대화를 이용하도록 개인정보수집 동의를 한 것으로 처리하여 왔다.

개보위는 이러한 개인정보수집 동의에 대하여, 이용자가 ‘이루다’와 같이 자신의 대화내용이 타인에게 드러날 수 있는 내용의 ‘신규 서비스 개발’ 목적의 이용에 동의하였다고 보기 어렵고, ‘신규 서비스 개발’이라는 기재만으로 이용자가 ‘이루다’ 개발과 운영에 카카오톡 대화가 이용될 것에 대해 예상하기도 어려우며, 이용자의 개인정보 자기결정권이 제한되는 등 이용자가 예측할 수 없는 손해를 입을 우려가 있다고 보았다.

한편, 피심인은 개발자 코드 공유 및 협업 사이트인 Github에 2019. 10.부터 2021. 1.까지 이름 22건(성은 미포함)과 지명정보(구·동 단위) 34건, 성별, 대화 상대방과의 관계(친구 또는 연인) 등이 포함된 카카오톡 대화문장 1,431건과 함께 인공지능(AI) 모델을 공개하였다. 이에 대하여, 개보위는 피심인이 가명정보를 불특정 다수에게 제공하면서 ‘특정 개인을 알아보기 위하여 사용될 수 있는 정보’를 포함하였으므로 「개인정보 보호법」 제28조의 2제2항 위반에 해당한다고 판단하였다.

개보위는 조사 과정에서 정보주체가 명확히 인지할 수 있도록 알리고 동의를 받지 않은 사실 등 추가 위반사실을 확인하였으며, ‘이루다’와 관련된 사항을 포함, 총 8가지 「개인정보 보호법」 위반행

위에 대하여 피심인에 총 1억 330만원의 과징금과 과태료를 부과하고 시정조치를 명령하였다.

Ⅲ. 주요 법적 쟁점

현재 비정형 데이터 자체를 직접 규율하는 법은 마련되어 있지 않다. 따라서, 비정형 데이터에 개인정보에 해당하는 부분은 개인정보보호법에, 지식재산권(저작권 등)에 해당하는 부분은 저작권법 등 지식재산권 관련법에 따른 규제 또는 제한을 받는다.

1. 개인정보 보호법 관련

먼저 비정형 데이터에 개인정보가 포함되어 있는 경우, 해당 데이터를 활용하고자 하는 기업은 원칙적으로 정보주체로부터 개인정보의 수집·이용에 관한 동의를 받아야 한다. 비정형 데이터의 활용을 검토하는 단계에서는 정보주체로부터 다시 개인정보의 수집·이용에 대한 동의를 받기 어렵다. 결국 이러한 문제를 해소하기 위하여, 개인정보 수집시 정보주체로부터 동意的 대상과 범위를 추상적이거나 불명확하게 기재하여 동의를 얻게 된다.

이에 따라 정보주체가 수집목적에 기재된 내용으로부터 장래에 개발될 비정형 데이터를 활용한 서비스에서 자신의 개인정보가 활용되는 양태를 구체적으로 예상하기 어려운 경우, 해당 정보주체의 동意的는 개인정보보호법상 적법한 동意的로 인정받지 못할 가능성이 있고, 이는 수집 목적 외의 이용행위로 판단되어 이용이 금지되는 결과로 이어진다.

데이터 3법 개정으로 도입된 가명 처리를 통한 해결을 모색할 수 있다. 비정형 데이터 중 개인정보에 해당하는 부분은 가급적 가명·익명 처리를 통해 개인식별 가능성을 제거하여 개인정보보호법에 따른 규제 리스크를 최소화하는 시도가 그것이다. 데이터3법 개정으로 도입된 개인정보보호법상 가명처리의 특례는 신규·서비스 개발 목적의 가명 처리를 허용한다.

그러나, 데이터 활용시 가명처리·익명처리로 해결하는 방안의 문제는 비정형 데이터에 포함된 개인정보를 완전하게 찾아내어 제거하는 것이 사실 상 매우 어렵다는데 그 문제가 있다. 현행법 해석으로는 고의·중과실이 없이 식별가능성을 제거하지 못한 행위를 처벌하지 않는 방식으로 규제리스크를 해소해 주어야 할 것이나, 사업자의 입장에서 어느 수준으로 비정형 데이터의 가명화를 하여야 할 것인지 여부에 대한 판단이 매우 어려울 수 있다.

2. 지식재산권 및 부정경쟁방지 관련

비정형 데이터가 저작물에 해당하거나 보호가치 있는 타인의 성과물인 경우에는 지식재산권 침해가 문제될 수 있다.

AI 학습 및 빅데이터 분석 등에 활용되는 데이터가 「저작권법」의 보호를 받는 저작물에 해당할 경우 복제권 및 전송권 침해 등이 문제될 수 있다. 데이터 분석을 통해 새로운 의미를 도출하는 ‘데이터 마이닝’을 위한 저작물 이용은 「저작권법」 제35조의3에 따른 공정이용에 해당한다는 견해도 있으나 현행 「저작권법」에서는 이에 관한 명문 규정을 두고 있지 않아 다툼의 여지가 있고, 개별 사례에서의 구체적인 저작물 이용 방식과 내용에 따라 저작권 침해 여부에 대한 판단이 달라질 수 있다.

특정 데이터가 ‘타인의 상당한 투자나 노력으로 만들어진 성과’에 해당하는 경우, 이를 무단으로 사용하는 행위는 「부정경쟁방지 및 영업비밀보호에 관한 법률」(이하 ‘부정경쟁방지법’) 제2조 제1호 카목에 따른 부정경쟁행위¹⁾에 해당할 가능성이 있다. 일본의 경우 2018년 부정경쟁방지법 개정을 통해 ‘한정제공데이터²⁾’의 부정취득, 사용, 공개를 영업비밀의 부정취득 등에 준한다고 하여 비정형 데이터에 관한 보호를 명시적으로 규정하나, 우리나라에

서는 아직 이러한 입법적 배려는 논의만 존재한다.

3. 시사점

빅데이터의 주축으로 주목받고 있는 비정형 데이터의 활용은 알고리즘 고도화 및 서비스 개선 수요에 따라 향후에도 지속적으로 늘어날 것이다. 비정형 데이터의 활용이 늘어남에 따라 개인정보 보호 등 관련 이슈가 지속적으로 발생하면서, 이에 대한 규제기관의 관심도 높아질 것이 자명한 가운데, 최근 ‘이루다’ 관련 제재 사례는 비정형 데이터 활용이 규제리스크를 내포하고 있음을 확인한 사례이다.

IV. 비정형 데이터 활용시 규제리스크 해소를 위한 방안

1. 개관

규제샌드박스는 새로운 제품·서비스를 출시할 때 일정 기간 동안 기존의 규제를 면제 또는 유예시켜 주는 제도로, 2019년 「정보통신 진흥 및 융합 활성화 등에 관한 특별법」, 「산업융합촉진법」, 「금융혁신지원 특별법」, 「지역특화발전특구에 대한 규제특례법」 등 4개법에 도입되었다(최근 「스마트도시 조성 및 산업진흥 등에 관한 법률」과 「연구개발 특구의 육성에 관한 특별법」에도 추가로 도입되었다). 규제샌드박스에서 사업 시행을 위하여 부여하는 특례의 대상이 되는 ‘규제’가 공법적 금지규정만을 의미하는지, 아니면 일반 사인의 권리에 의한 금지까지도 제한적 범위의 특례 부여를 통하여 허용되도록 할 수 있는지에 관해서는 견해가 나뉠 수 있다.

2. 규제샌드박스를 이용한 규제리스크 해소가능성

개인정보보호에 관한 규제는 어디까지나 공법(公法)적인 금지규정이므로, 그에 대한 특례를 설

1) 타인의 상당한 투자나 노력으로 만들어진 성과 등을 공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 자신의 영업을 위하여 무단으로 사용함으로써 타인의 경제적 이익을 침해하는 행위
2) 업으로서 특정한 자에게 제공한 정보로 전자적 방법에 의하여 상당량이 축적, 관리되고 있는 기술상 또는 영업상의 정보

정하는 것은 허용된다고 봄이 상당하고, 이미 공공 기관 전자고지서비스 사례에서 주민등록번호의 CI(Connected Information) 변경시 정보주체의 개별 동의를 얻지 않도록 하는 특례가 인정된 바 있다.

그러나, 저작권 등 지식재산권의 경우에는 이와 달리 봄이 상당하다. 저작권법의 목적은 ① 저작자의 권리로서의 저작권 보호와 ② 저작물의 공정한 이용을 도모함으로써 문화 및 산업 향상 발전에 이바지하는 것이다. 저작권법상 전자의 목적은 저작권의 보호로, 후자는 저작권의 제한이라는 방식으로 규정되는데, 저작권자의 권리를 보호함과 더불어 저작물의 이용 또는 일반 대중의 저작물 향유가 위축되지 않고 인류의 문화적 총량을 증가시키기 위해 필요한 만큼으로 저작권의 내용과 범위를 합리적으로 제한해야 한다는 것이 저작권 제도의 근본적 긴장관계를 이루어야 한다. 따라서 저작권법을 해석함에 있어 개인의 사권으로서의 저작권 보호만을 지나치게 강조한 나머지 이용자들의 정당한 권익이나 공공의 이익이 저해되지 않도록 할 필요가 있다.

저작권법의 입법 및 해석에 공익적 관점이 고려되고 이에 따라 저작권자의 권리가 제한 된다고 하더라도, 저작권의 경우에도 저작권 제도의 또 다른 목적인 ‘문화 및 산업 발전’을 위해 일정한 범위 내로 권리가 제한되는 것일 뿐이므로, 저작물의 이용 관계를 공법상 관계로 보기는 어렵고, 여기에 규제 샌드박스를 통한 특례를 부여하기는 쉽지 않은 면이 있다.

V. 결론

결국 앞서 살펴본 바를 종합하면, 비정형 데이터를 이용하기 위해서는 개인정보보호법과 저작권법, 부정경쟁방지법 등의 특례가 인정되어야 할 것으로 보인다.

현행 규제샌드박스는 기존 공법적 규제를 풀어 주는 방식에 맞게 설정된 제도라는 점을 고려할 때,

규제샌드박스를 통한 규제특례를 부여하는 방안은 개인정보에서의 정보주체의 개인정보자기결정권 보호, 저작권자의 저작권 보호와 같은 권리보호 관점에서 그 성격이 맞지 않는 측면이 있다.

따라서 데이터의 공정이용행위를 구체화하고, 이를 허용하는 내용의 입법적 논의가 진행될 필요가 있으며, 이러한 논의는 데이터 경제의 도래라는 시대적 상황을 고려하여 개별 법령³⁾이 파편적으로 개정되기 보다는 신중하게 검토될 필요가 있다고 생각한다.

3) 참고로, 데이터 마이닝 과정에서의 저작물 이용 관련, ‘자동화된 분석 기술을 통해 다수의 저작물을 포함한 대량의 정보를 분석하는 경우’ 필요한 한도에서 저작물의 복제전송을 허용하는 내용의 「저작권법 전부개정안」(도종환 의원 대표발의, 의안번호 2107440)이 발의되어 현재 국회 문화체육관광위원회에 계류 중이다.