이루다 서비스와 AI 모델 학습에서의 개인정보의 처리

강태욱
법무법인(유) 태평양
변호사, 법학박사

Ⅰ. 들어가며

머신러닝 기반의 AI 학습 모델에 있어서 대량의 데이터의 사용은 필수적이다. Open AI로 일컬어지는 GPT-3 알고리즘의 경우 약 1조개에 가까운 데이터셋과 말뭉치, 위키피디아 등 공개된 텍스트 데이터를 통하여 학습이 이루어졌다.

인공지능 챗봇 서비스로 지칭할 수 있는 이루다 서비스는 페이스북 메신저 내에서 AI 챗봇이 이루다를 친구로 하여 이용자와 대화를 할 수 있도록 하는 방식의 NLP(Natural Language Processing) 서비스이다. 이루다 서비스는 Retrieval 방식의 마이크로소프트의 XiaoIce 기반 프레임워크에서 출발하여 독자적으로 수집한 대화 데이터를 이용하여 학습함으로써 개발되었다.[1)]

이 서비스는 2020. 12. 22. 출시되었고 약 3주간의 서비스를 진행하던 중 그 서비스를 중단하였다. 이루다 서비스는 대화 내용 중의 편향된 발화 내용으로 인하여 언론에서 기사화되었고, 이후 이루다 서비스를 통하여 개인정보가 발화되는 과정에서 개인정보 보호법 위반 이슈가 있는 것은 아닌지가 이슈화되었으며, 개인정보보호위원회는 이루다 서비스의 개인정보 보호법 위반 이슈를 점검하여 지난 4. 28. 총 8가지 항목에 대하여 개인정보 보호법을 위반하였다고 판단하여 과징금 5,550만원, 과태료 4,780만원 및 시정명령을 발하는 처분을 하였다.[2)]

이 사건에서 여러 가지 쟁점이 제기되었지만 개인정보 보호법의 해석과 관련하여 가장 논란이 된 것은 크게 다음의 3가지 이슈로 볼 수 있다.[3)] 첫째, 회사는 정보주체로부터 카카오톡 대화를 수집하였는데, 대화를 수집하는 과정에서 일방 당사자만의 동의를 얻고 상대방 당사자의 동의는 없이 카카오톡 대화를 수집, 이용한 것이 개인정보 보호법 위반이 아닌지 여부, 둘째, 회사가 카카오톡 대화를 포함한 정보주체의 개인정보를 수집하는 과정에서 "신규 서비스 개발"이라는 목적을 고지하고 동의를 받았는데, 이와 같은 목적에 비추어 이루다 서비스의 개발에 있어서 목적 내 이용이라고 볼 수 있는지 여부, 셋째, 이루다 서비스의 개발 과정에서 발화용 DB에 있어서 메타 데이터 중 식별자는 모두 삭제하고 대화 데이터만 남겼으며 그 중 숫자 등 식별성이 있는 정보에 대하여 모델링 방식에 의하여 비식별 처리를 하였는바 이러한 처리 행위의 결과물을 익명 정보 내지 가명정보로 볼 수 있는지 여부 등이다.[4)]

이하에서는 위 세 가지 쟁점에 대한 개인정보보호위원회의 의견을 일별하고 그에 대한 필자의 의견을 덧붙이기로 한다.

Ⅱ. 일방 당사자의 동의에 의한 개인정보의 수집 여부

쌍방 당사자에 의한 대화가 이루어진 경우, 이러한 대화가 '개인정보'를 포함하고 있어 개인정보성을 가지는 것을 전제로 하는 한, 이러한 대화 내용은 쌍방 당사자 모두의 개인정보에 해당하는 것을 보아야 할 가능성이 높다. 즉, 대화자 A와 대화자 B가 대화한 내용의 경우 그 양 당사자가 모두 해당 대화에 대하여

1) 김종윤, 오픈도메인 챗봇 루다 육아일기 : 탄생부터 클로즈베타까지의 기록, Naver Deview 컨퍼런스 발표자료 참조.

2) 개인정보보호위원회, 이루다 개발사 ㈜ 스캐터랩에 과징금, 과태료 등 제재처분, 2021. 3. 29 보도자료, 1면

3) 그 이외에도 여러 쟁점이 있고 처분청인 개인정보보호위원회와 이루다 서비스 사 측의 입장이 서로 상이하나 본 고에서는 그 중 일부 쟁점만을 다루기로 한다.

4) 챗봇 서비스의 법적 이슈를 포괄적으로 다룬 논문으로는, 양종모, 인공지능 챗봇 알고리즘에 대한 몇 가지 법적 고찰, 홍익법학, 21권 1호, 2020. 참조.

개인정보의 주체로서 인정될 여지가 있다. 이 경우 일방 당사자가 그 대화 전체에 대하여 처분권한이 있는지 여부가 논란이 될 수 있다.

이와 관련하여, 개인정보보호위원회는 일방 당사자가 대화 내용을 회사에 제공한 행위(즉, 회사의 수집에 동의한 행위)는 일방 당사자의 개인정보로써 수집이 이루어진 것이므로 허용되는 행위로 판단한 것으로 보인다.[5] 더 구체적인 논거는 다음과 같이 생각해 볼 수 있다. 첫째, 개인정보 보호법은 개인정보파일을 운용하는 것을 기준으로 개인정보처리자 해당 여부를 정하고 있는데, 대화를 전송한 자는 개인정보 보호법 상의 '이용자'에 해당하고 이 이용자를 기준으로 개인정보 파일을 관리하고 있으나 대화 상대방에 대하여는 별도로 해당 상대방을 기준으로 개인정보파일을 운용하고 있는 것이 아니므로 대화 상대방에 대하여는 개인정보처리자의 지위에 있지 아니하다는 점이다.[6] 둘째, 당사자 사이의 대화 내용을 일방 당사자가 녹음하더라도 통신비밀보호법 상 '감청'에 해당하지 않는다고 판단한 바와 같이(대법원 2008. 10. 23. 선고 2008도1237 판결), 일방 당사자는 일정한 범위 내에서 해당 대화 내용에 대한 처분 권한을 가지며, 회사의 수집에 동의하여 준 행위는 이러한 처분 권한 내의 행위로 볼 수 있는 것이다.[7]

메신저 내에서의 대화 내용 자체의 개인정보성과 관련하여서는 두 가지로 나누어서 살펴볼 필요가 있다. 먼저 대화 내용 자체에 아무런 개인정보를 포함하지 아니하였더라도 개인이 식별되는 Meta data와 완결된 상태에서 대화 내용 자체를 개인정보로 볼 수 있을 것인가의 문제이다. 즉, 대화 내용 자체로 개인을 식별할 수 없는 경우에도 개인정보성을 인정할 것인가의 하는 점이다. 다음으로는 대화 내용에 개인정보가 포함된 경우에 이를 개인에 관한 정보로 보아야 할 것인가의 문제이다. 첫번째 점에 대하여는 대체로 개인정보성을 인정할 필요가 없다는 결론으로 논의가 모이는 것으로 보이나, 두번째 점에 대하여는 많은 논란이 있는 것으로 보인다(첫번째 쟁점과 두번째 쟁점을 과연 구별할 수 있을 것인가의 의문도 들 수 있다). 그 중 하나는 전체 컨텍스트 측면에서 보았을 때 카카오톡 대화는 컨텐츠의 영역에 속하는 것이므로 이를 개인정보 보호법 상 보호되는 개인정보와는 다른 관점에서 접근하여야 한다는 시각이다.[8] 또다른 관점은 그 형식이 어떠하든 그 안에 개인정보라고 볼 수 있는 정보가 포함되어 있다면 개인정보 보호법 상 보호되어야 한다는 시각이다.

이러한 문제는 실무상으로는 open text, open speech와 관련하여 자주 발생하는 문제이다. 개인정보처리자는 특정한 내용의 개인정보를 수집하려는 의도가 전혀 없었음에도 정보주체 내지 이용자가 특정한 개인정보를 임의로 입력하여 처리자가 개인정보를 수집하게 되는 경우이다. 예컨대, 개인정보처리자가 입사지원자로부터 온라인 내지 웹사이트를 통하여 자기소개서를 제출받으면서 주민등록번호나 민감정보를 입력할 것을 전혀 예정하지 않았는데 정보주체가 임의로 이러한 정보들을 입력하는 경우를 생각해 볼 수 있다. 이러한 경우에 개인정보처리자가 주민등록번호 법정주의를 위반하였다거나 민감정보에 대한 별도 동의 원칙을 위반하여 개인정보를 수집하였다고 볼 수 있을 것인지는 논란이 될 수 있다.[9]

5) 개인정보보호위원회, 앞의 보도자료, 6면. 한편, 법원도 서울고등법원 2015. 2. 9. 선고 2014노2820 사건에서 빅데이터 업체가 API를 이용하여 트위터 정보를 수집한 사안에 대하여 적법한 수집으로 본 바 있다(대법원 확정).

6) 서울서부지방법원 2019. 2. 14. 선고 2018노556 판결은 B 라디오 작가인 피고인 A가 청취자 D가 피고인에 대한 항의글을 게시하자 D에 대한 고소사건에서 피고인 A를 대리하던 변호사에게 D의 주소 및 연락처를 교부하여 개인정보를 목적외 이용하였다는 취지로 고소되었는데, 이에 대하여 법원은 피고인 A가 B 라디오의 개인정보 DB에 접근 권한이 있으므로 개인정보처리자에 해당한다는 검사의 주장을 배척하고 A는 개인정보파일을 운영하는 개인정보처리자에 해당한다고 볼 수 없다고 판단하였다(위 사건은 대법원 2019. 7. 15. 선고 2019도3215 판결로 확정되었다). 이러한 판단은 '개인정보처리자'의 개념을 좁게 해석함으로써 개인정보 보호법 위반에 따른 형사처벌의 위험성이 확대되는 것을 방지하려는 정책적 판단으로 생각되며, 기존에 개인정보처리자의 개념을 좁게 해석한 기존의 대법원 판결례들과도 일응 궤를 같이한다고 평가할 수 있다.

7) 한편, 법원은 카카오톡 단체 대화방 내에 일방 당사자의 대화에 대하여 압수수색을 허용한 영장의 범위에 해당 대화방 내의 다른 이용자의 대화도 포함된다고 보았다. 미디어 오늘, 2019. 10. 5.자 기사 참조.

8) 대화내용 자체는 법률에서 정하는 개인정보의 범주에 해당하지 않는다는 견해로 News1, 2014. 10. 12.자 기사("카카오톡 대화내용, 법률상 '개인정보' 여부 놓고 공방") 참조.

9) 이러한 경우에도 개인정보처리자가 적법한 동의를 받지 아니할 경우 그 책임을 져야 한다는 견해는, 소병수/김형진, 소셜미디어상의 개인정보 활용과 보호, 법학연구 24집 1호, 인하대학교 법학연구소, 196면.

메신저 대화 내용 역시 마찬가지의 관점에서 생각해 볼 수 있을 것이다. 대화 내용 중에 개인정보를 입력하는 것이 비록 금지되어 있지는 않지만 메신저를 운영하는 사업자로서는 대화 내용 내에서 개인정보를 수집하려는 의사는 전혀 없었다는 점에서 이러한 대화 파일 자체를 개인정보 보호법 상의 개인정보 파일로 볼 것인지의 문제가 발생하게 되는 것이다.

좀더 심도깊은 논의가 필요한 부분이나, 개인정보보호위원회는 이러한 메신저의 대화 내용에 대하여 '개인정보'를 포함하고 있다고 하더라도, 적어도 일방 당사자가 그 수집에 동의하는 행위는 양 당사자간에서는 일방에게 허용된 범위의 행위로 보았다는 점에서 오랫 동안 논의되어 왔던 쟁점에 대한 개인정보보호위원회의 입장을 명확히 하였다는데 그 의의가 있다 할 것이다.

Ⅲ. '신규서비스' 개발과 목적외 이용

이루다 서비스의 개발에 활용된 대화 정보는 회사가 운영하고 있던 '텍스트앳'이라는 서비스와 '연애의 과학'이라는 서비스로부터 각 수집한 정보들이다. 이러한 정보들은 그 수집 과정에서 텍스트앳 및 연애의 과학 서비스 이용자로부터 수집 과정에서 '신규서비스 개발'이라는 목적을 고지하고 동의를 받았다.[10]

이에 대하여 개인정보보호위원회는 위 각 서비스에서 '분석의 대상이 되는 메시지'를 '신규 서비스 개발' 목적으로 이용한다는 점을 명시하고 동의를 받았다는 점만으로는, 카카오톡 대화가 이루다 서비스의 학습, 운영에 이용될 것이라는 점을 이용자가 예상하기 어렵다라고 판단하여 목적 범위 내의 이용에 해당하지 않는다고 판단하였다.[11]

각 항목들에 대하여 살펴보면 위원회의 위와 같은 결론은 쉽사리 동의하기 어렵다.

'텍스트앳' 서비스는 대화 내용에 기반하여 대화 상대방의 감정 분석을 의뢰하는 서비스로, 이용자가 텍스트앳 서비스 내에서 감정 분석을 의뢰하면 회사는 알고리즘에 기반하여 대화 내용을 분석한 후 이를 이용자에게 알려주는 방식의 서비스이다. '연애의 과학' 서비스는 좀더 연인간의 대화에 집중하여 일방 당사자가 요청하는 경우 알고리즘에 기반하여 대화 내용을 분석한 후 그에 따라 이용자가 요청하는 연애에 대한 자문을 해 주는 내용의 서비스이다.

두 서비스는 기본적으로 AI 학습 모델을 기반으로 한 통계학적인 분석 서비스로서 공통적으로 메신저 대화 내용을 분석하여 상대방의 감정을 예측하고 그에 대한 반응을 하여 주는 Sentiment Analysis(감성 분석)[12] 서비스의 일종에 해당한다.

위와 같은 요소들, 즉 대화 내용에 기반한 통계학적 분석 서비스라는 점, AI 감성 분석 서비스의 일종에 해당하는 점은 이루다 서비스에서도 마찬가지의 요소에 해당한다. 즉, 이루다 서비스는 대화 분석 서비스가 아니라 챗봇 서비스이기는 하나 기존 대화의 분석을 기반으로 한다는 점에서, 텍스트앳 서비스와 같은 대

10) 본건에서는 동의를 받는 방법이 적정하였는지의 이슈도 제기되었으나 논의의 편의상 본 글에서는 이 부분의 논의는 제외하기로 한다.

11) 개인정보보호위원회, 앞의 보도자료, 3면.

12) 감성 분석은 자연어인 텍스트에 들어있는 의견이나 감성, 평가, 태도 등의 주관적인 정보를 컴퓨터를 통해서 분석하는 연구방식을 지칭하며, 데이터 사이언스의 한 분야이다.

화 분석 서비스에 있어서 대화 대상이 되는 챗봇 서비스의 개발은 충분히 사업자라면 생각할 수 있는 '신규 서비스의 개발'의 범위 내에 포함된다고 봄이 타당하다.

이와 관련하여, 대법원은 소위 '부활충전' 사건에서 목적외 이용 여부는 정보주체가 선불폰 개통 당시 예상한 서비스 범주 내의 행위였는가를 기준으로 보아야 한다고 하면서, 사업자가 동의를 받은 '서비스 제공'은 이동전화 서비스를 이용자 동의 하에 공급하는 것이지 해당 서비스를 사업자가 일방적으로 공급하는 것을 의미하는 것은 아니다라고 판단하였고(대법원 2018. 7. 12. 선고 2016두551178 판결), 이러한 판시 취지에 비추어 목적외 이용이 좁게 해석되어야 한다는 견해도 보인다. 그러나, 위 사안은 '서비스 제공'이라는 해당 동의의 필수적인 사용 범위의 해석에 대한 것으로 이를 '고객 유지만을 목적으로 하고 사업자에게는 마케팅 목적이 인정되는 경우'까지 확대될 수는 없다는 취지의 판단인바, 해당 서비스에 대한 것임이 아님이 명백한 '신규' 서비스의 의미에 대하여까지 좁게 해석할 것으로 단정적으로 판단하여야 한다는 의미로 새길 수는 없다. 오히려 위 판결의 판시의 취지에 비추어 보면, 이루다 서비스는 기존의 텍스트앳이나 연애의 과학 서비스의 대화 내용 분석과 그 서비스의 내용이 유사하고, 기존 서비스의 이용자의 입장에서는 '신규 서비스'로서 예상할 수 있는 범위에 속하며, 서비스 개발을 위하여 가명처리 내지 익명처리가 이루어진 점 등을 고려하여 이용자들이 불측의 손해를 입을 가능성이 낮은 점 등을 종합하여 보면 오히려 '수집 목적 범위 내'라고 봄이 타당하다.

나아가, 기반 플랫폼이나 이용 대상이 다소 상이하다고 하더라도 이러한 사정만으로 목적외 이용에 해당한다고 단정할 것은 아니다. 앱 서비스에 있어서 신규 서비스는 기본적으로 새로운 앱을 통한 서비스를 전제로 하는 것이므로 기존의 서비스 내에서 새로운 기능이 추가되는 것으로 좁게 해석하여야 할 것은 아니며, 이용 대상의 경우에도 텍스트앳이나 연애의 과학 역시 대학생이나 젊은 직장인을 대상으로 하고 있음이 명백한 서비스로서 20대의 여성 대학생을 persona로 상정한 이루다 서비스와 비교하여 그 이용 대상 층이 전혀 다른 것으로 볼 수도 없다.

'신규 서비스 개발'이라는 문언 자체가 사전적으로 그 범위를 명확히 규정하기는 어렵다는 점에서 무한정 확장할 경우에는 정보주체의 권리 보장이 될 수 없다는 반론도 가능할 것으로 보인다. 그러나, 그러한 이유로 서비스 내의 새로운 기능 추가 정도로 목적 범위를 한정하는 것은 사업자의 개인정보의 이용 범위를 과도하게 제한하는 결과가 될 수 있다.[13] 이와 관련하여 "가입 후 회원의 카카오톡 상의 개인적 대화 내용, 신상정보, 소셜미디어 상에 포스팅한 데이터가 수집되며, 이러한 정보는 추후 개발될 예정인 AI 채팅로봇 서비스에 답변의 일환으로 활용될 수 있다."라는 식의 표현을 할 필요가 있다는 견해도 있으나,[14] 합리적 관련성 원칙에 따른 동의 없는 처리의 기준으로서 위와 같은 예시가 제시될 수 있는 경우라는 몰라도, 사전적으로 위와 같은 내용을 모두 고지하고 동의를 받는다는 것은, 개발 과정에서 그 개발 목적이 변화할 수 있는 현실에 부합하는 대안이라고 보기는 어려워 보인다. 이와 관련하여서는 감독기관의 의견이 조속히 제시될 필요가 있다.

13) GDPR 하에서 AI의 학습과 빅데이터의 활용은 목적 제한 원칙이 충돌할 수 있다는 점에 대해서는, The Norwegian Data Protection Authority, Artificial intelligence and privacy, 2018, 1., pp. 16-19.

14) 소병수/김형진, 앞의 논문, 190면 등.

Ⅳ. 가명처리 및 익명처리에 있어서 비식별화의 수준

자연어 처리와 관련한 AI 모델은 발화하는 데이터를 임의로 생성하는 Generative 모델과 일정한 대화 내용을 기반으로 그대로 발화하는 Retrieve 모델로 나눌 수 있다.[15] 이루다 서비스 모델인 Retrieve 모델의 경우 발화 DB에 개인에 관한 정보가 포함된 경우에는 해당 정보가 그대로 발화가 이루어지게 되므로 발화 DB의 내용에 대한 비식별 조치가 전제되어야 한다. 이에 회사 역시 발화 DB에 대하여 개인정보 비식별 처리를 수행하였으나 그럼에도 불구하고 다수의 개인에 관한 정보라고 볼 수 있는 데이터들이 발화 과정에서 생성되었다. 이 부분의 쟁점은 발화 DB의 생성에 있어서 개인정보의 비식별 처리가 어느 정도로 완벽하여야 하는가의 정책적 판단의 문제이다.[16]

개인정보 보호법은 원칙적으로 정보주체 개개인을 보호 대상으로 삼고 있으므로, 정보주체 1인의 개인정보에 대한 법 위반 행위가 발생하더라도 이러한 행위를 법 위반 행위로 보고 있다. 휴대전화번호 4자리 사건(대전지방법원 논산지원 2013. 8. 9. 선고 2013고단17 판결)이나, 비상연락망 사건(대법원 2017. 4. 13. 선고 2014도7598 판결) 모두 단 한 건의 개인정보의 처리가 문제가 된 사안이다.

그런데, 비식별 처리에 있어서도 이처럼 동일한 기준을 적용하여야 할 것인지는 의문이다. 왜냐하면, 처리하여야 할 데이터의 양이 너무 많아서 적법한 비식별처리가 이루어졌다고 하더라도 해당 데이터베이스 내에 개인정보가 100% 제거되었다는 점을 단정하기 어렵고, AI 학습이라 처리 목적의 성격 상 외부에 공개되는 방식의 처리에 해당하지 아니하며, 데이터의 발화가 이루어진다고 하더라도 그로 인하여 개인이 식별될 가능성은 매우 낮기 때문이다.[17]

참고로 이루다건에 대한 개인정보보호위원회의 현장조사서에 의하면, 회사는 응답후보 DB의 생성을 위해서 NER(Named Entity Recognition) 모델[18]을 통한 사람 이름으로 분류되는 문장의 삭제, 선정적 키워드의 삭제 필터링, PNR(Personal Name Recognition)을 통하여 NER 모델이 검출하지 못한 실명 필터링, 숫자 및 영문 포함 문장 삭제, 사적 표현의 문장 필터링, 선정적 대화 내용의 필터링, 테스트 과정에서 확인되는 문자 및 해당 문장과 유사한 벡터값을 가진 문장의 삭제 등의 과정을 거쳤으며, 그 이후에도 지속적으로 테스트 과정에서 확인되는 개인정보성이 인정될 수 있는 정보에 대하여는 삭제하는 과정을 수행하였다. 역시 현장조사서의 내용에 따르면 실제 발화된 문장 총 720만건 중 숫자는 0건, 행정지역 중 동/호수가 포함된 문장은 1건(동까지 표시된 것은 232건), 문자로 표시된 핸드폰 번호로 추정되는 정보는 3건, 계좌번호 및 카드번호는 0건의 발화가 이루어진 것으로 조사된 바 있다.

15) 김병필, 이루다 사례를 통해서 본 자연어처리 인공지능의 법정 쟁점, 정보법학회 사례연구회, 2021. 3., 36면.

16) GDPR 하에서의 AI 기술의 활용을 위한 기술적 조치와 관련하여서는, UK Information Commissioner's Office (ICO), Guidance on Big data, Artificial Intelligence, Machine-Learning and Data Protection, 2017, p.38. 등 참조.

17) Big Data 내 context의 완전한 비식별화의 어려움에 대해서는 Yordanka Ivanova, Re-using Personal Data for Statistical and Research Purposes in the Context of Big Data and Artificial Intelligence, APF 2019, p.119.

18) 자연어 처리 과정에서 미리 정의된 인명, 지명 등을 인식하여 추출하는 기법을 지칭한다.

빅데이터를 학습하지만 실제로 parameter만이 엔진에 저장되고 발화 DB 없이 문장을 새롭게 생성해는 Generative 방식의 AI 모델의 경우에도 해당 모델로부터 학습된 개인정보가 추출될 수 있는 위험이 존재하며,[19] 앞서 설명한 여러 형태의 단어 인식 모델을 통하여 특정 정보를 인식하여 이를 삭제하더라도, 개인정보가 식별될 어느 정도의 가능성이 존재할 수밖에 없다.[20] 식별될 가능성이 완벽하게 없도록 하기 위해서는 모든 학습 데이터를 전수조사하여 개인정보를 삭제하여야 하는데, 이루다건의 응답 후보군 DB만 하더라도 약 1억건의 대화문이 존재하고, 실제로 학습에 활용된 DB는 약 90억건에 이르는 대화라는 점을 고려할 때 이러한 대화문을 전수 조사하여 개인정보성을 가진 부분을 100% 삭제하여야 한다는 것은 기술적으로 기대가능성이 있는 것인지는 매우 회의적이며, 적어도 이러한 엄격한 기준의 고수는 빅데이터의 학습을 전제로 하는 AI 산업에 있어서 치명적인 장벽이 될 가능성이 매우 높다.

Ⅴ. 이루다 사건 이후

이루다 처분은 4차산업혁명의 총아라고 불리는 AI 및 빅데이터 산업에 있어서 개인정보의 활용과 관련하여 어느 정도의 요건을 충족하여야 할 것인지에 대한 일응의 기준이 제시된 우리나라 최초의 사례라고 할 수 있다. 나아가, 개인정보를 활용하려는 기업이 개인정보 보호법 관련 이슈를 충분히 인식하고 법 준수를 위하여 노력하여야 한다는 점을 생각하게 한 사례라고도 할 것이다. 앞으로 개인정보보호위원회를 비롯하여 학계와 업계의 풍부한 논의를 통하여 개인정보 보호법이 지향하는 보호와 활용의 균형점을 적절히 찾을 수 있기를 기대해 본다.

19) Carlini et al, Extracting Training Data from Large Language Models, 2020. 참조. 위 논문은 GPT-2 모델에 대한 privacy attack을 통하여 개인에 관한 정보를 추출해 낸 사례이다. 물론, 이 정보들이 모두 '곧바로 개인을 특정하여 식별할 수 있는 정보'라고 볼 것인지, 또는 그 공개성에 비추어 모델 자체의 폐기를 요구할 정도의 심각한 개인정보에 대한 침해가 발생하는 것인지는 전혀 다른 쟁점이다.

20) 진료기록에서 개인식별정보가 삭제되지 못할 확률이 3.75%, 언어모델로부터 개인정보가 추출될 확률은 0.13%, 현재 연구수준 상 언어모델로부터 개인정보가 재생성될 확률은 0.0005%라는 연구 결과 및 공개된 정보를 기준으로 하더라도 zero-risk가 달성될 수 없다는 설명은, 김병필, 앞의 발표자료, 70면.