전승재 변호사(법무법인 바른, '해커 출신 변호사가 해부한 해킹판결' 저자)

인공지능 학습은 개인정보 처리인가? 통계 처리인가?

법무부가 출입국 심사 중 확보한 1억7000만건의 안면 이미지를 인공지능(AI) 개발업체에 맡겨 논란이 되고 있다. 얼굴은 그 자체로 개인을 알아볼 수 있는 정보로서 대표적인 개인정보이다. 가령 엑스레이나 컴퓨터단층촬영(CT) 사진은 환자 이름표를 떼면 이것을 촬영한 의료기관 이외에는 어느 환자의 것인지 알아보기 어려우니 개인정보가 아닐 수도 있겠지만, 얼굴 사진은 이름표를 떼도 여전히 개인정보라는 점에서 특히 주의를 요한다. 이번 사건에서 개인정보 보호법상 쟁점은 두가지이다.

첫째 법무부가 AI 개발업체라는 제3자에게 개인정보를 제공한 것인지, 혹은 개인정보의 처리를 위탁한 것인지가 문제 된다. 전자인 '3자 제공'은 법무부와 AI 개발업체가 이른바 '제휴' 관계로서 개인정보 처리에 관하여 양쪽 모두 독자적 권리와 책임을 갖는다는 뜻이고, 이럴 때 적법요건은 정보주체의 '동의'이다. 후자 '처리위탁'은 법무부가 AI 개발업체에 일종의 '하도급'을 준 것으로서 모든 대외적 책임은 법무부가 지고 업체는 법무부가 시킨 대로만 개인정보를 처리한다는 뜻이며, 이럴 때는 정보주체의 동의를 받을 필요는 없고 위탁업무의 내용과 수탁자를 '공개'하기만 하면 된다. 법무부는 "처리위탁에 해당하므로 동의를 받지 않아도 적법하다."고 해명한 바 있는데, 이 부분에 대해서는 개인정보 보호위원회의 조사가 진행 중이다.

둘째 AI 학습을 통해 자동화된 안면인식 시스템을 개발하는 것이 개인정보의 목적 외 이용에 해당하는지 문제 된다. 시민단체 측은 법무부가 출입국관리법에 근거하여 생체정보를 처리할 권한은 있지만 그 데이터를 민간기업

이 개입한 AI 식별 시스템 개발 용도로 쓸 권한은 없으므로 개인정보 보호법 위반이라고 주장한다.

한편 법무부는 AI 학습 결과물에는 개인정보가 포함되어 있지 않으므로 개인정보 보호법 적용 대상이 아니라고 맞선다. 이 부분을 좀 더 자세히 살펴보자.

법무부가 민간업체로 하여금 개발하도록 한 것은 '이 얼굴(현재 카메라에 찍힌 것)과 저 얼굴(기존 데이터베이스에 저장된 것)이 같은 사람의 것이냐'는 질문에 대답하는 AI다. 학습 데이터로 제공된 1억7000만 건의 안면 이미지는 동일인에 대한 사진끼리 묶여 있다. 각 사진은 안면 특징점별 좌푯(벡터)값으로 변환되는데, 동일인의 사진이라도 촬영 각도나 표정 등에 따라서 좌푯값이 모두 다르다. AI는 동일인을 찍은 사진끼리 갖는 공통점과 타인을 찍은 사진과의 차이점을 관찰하면서 '좌푯값이 이렇게 들어오면 동일인일 가능성이 몇 퍼센트이고, 저렇게 들어오면 타인일 가능성이 몇퍼센트이다'라는 확률값을 일일이 기록해 둔다. 이 과정을 거치면 안면 특징점 좌푯값 패턴별로 동일인일 확률을 각각 기록한 거대한 행렬(matrix)이 만들어지며, 학습 데이터가 많을수록 이 확률값은 정밀해진다. 그러한 확률값 시퀀스가 바로 AI 학습 결과물이다. 이후 AI를 실제 운영하는 단계에서는 임의의 두 사진(학습 데이터에 포함되지 않았던 새 사진도 무방하다)을 위 행렬에 입력으로 넣어서 두 사진상의 인물이 동일인일 확률을 구하도록 하고, 그 확률이 일정한 기준이 넘으면 동일인으로 판정하게 된다.

중요한 것은 확률값 시퀀스가 기록된 행렬에는 학습 데이터인 안면 특징점 좌푯값이 그대로 남아 있지 않다는 점이다. 만약 그 원본값이 남아 있다면 AI 학습 과정에서 개인정보가 복사된 셈이다. 하지만 AI를 그렇게 단순하게 구현하지 않는다. 원본값을 관찰한 세세한 패턴별로 나누어 복잡하게 파편화시킨 뒤 확률값 시퀀스를 연산하는 방식으로 AI 학습 알고리즘을 구현한다. 이렇게 하면 학습 결과물에는 원본값이 남지 않으며, 원본값을 역산해내는 것도 현실적으로 불가능할 것으로 보인다. 법무부로서는 이 점을 보다 명확하게 해명할 필요가 있다.

요컨대 AI 학습은 '통계처리'의 일종이다. 예컨대 개개인의 소득정보는 당연히 개인정보이지만, 이것을 가지고 평균이나 표준편차를 내면 개인별 특성이 제거되며 통계값이 특정 개인에게 연결되지도 않는다. 이에 '통계처리 목적의 동의'를 정보주체에 따로 받지 않는다. AI 학습도 빅데이터의 패턴을 관찰하여 확률값을 연산하는 과정에서 개별 데이터의 특성을 남겨놓지 않는다는 점에서 마찬가지로 별도 동의를 받지 않아도 될 것이다. 나아가 개인별 소득정보를 평균 낸 '국민소득'

같은 통계값을 정보주체의 동의 없이 3자에게 제공할 수 있듯이, 학습 결과물로서 만들어진 확률값 시퀀스 및 AI 모델 등도 법무부가 민간업체와 공유할 수 있을 것이다. 학습 데이터는 개인정보이므로 당연히 반출이 안 되지만, 그 데이터를 통계처리하여 얻어 낸 노하우는 따로 활용할 수 있어야 한다.

덧붙이면 통계처리로서의 AI '학습'과 그 후 실제 서비스 운영 시 '개인정보 처리'는 구분해야 한다. 법무부가 안면 이미지 1억7000만 건을 학습시켜 AI 모델을 만드는 단계까지는 통계처리일 수 있다. 반면에, 개발된 AI 모델에 실제 우리나라 국민의 여권 사진을 등록시켜 출·입국 심사에 쓰면 이것은 개인정보 처리에 해당하므로 개인정보 보호법에 따른 엄격한 통제를 받는다.

한편 학습 데이터와 뗀 안면인식 AI 모델만을 외국에 수출하여 외국 공항에서 안면인식 용도로 쓴다면 이것은 별도의 개인정보 처리로서 그 국가 법률의 적용을 받는다.

이처럼 AI 학습과 서비스의 '운영'을 구분하여 법적 규율을 달리할 수 있다는 점에 대해서 보다 심도 있는 학술적·정책적 논의가 필요하다. 쉽게 말하면 기술을 개발하는 단계부터 규제할 것이냐, 그 기술을 쓰는 단계에서 비로소 법적 통제를 할 것이냐의 문제이다.

(저작권자/세계일보)