

# 기업의 AI 및 알고리즘 이용에 관한 지침 : 미국 FTC의 ‘Using Artificial Intelligence and Algorithms’을 중심으로

양기문\*

## 1. 개요

지난 14일, 구글은 인종 차별적 판단을 내린 당사의 AI 서비스에 대해 사과의 입장을 밝혔다. 구글의 이미지 식별 AI 서비스가 피부색에 따라 다른 결과를 도출했기 때문이다. 독일의 IT전문매체인 알고리즘워치의 이미지 식별 실험에서, 구글의 서비스는 밝은 피부를 가진 사람이 체온계를 들고 있는 사진에는 체온계로, 어두운 피부를 가진 사람이 체온계를 들고 있는 사진에는 총으로 분류했다. 이에 대해 구글은 시스템적 편향이라 인정할 수는 없지만, 차별적 결과로 인해 발생할 수 있는 피해에 대해 사과하고 결과와 인종차별의 관련성을 인식하는 것이 중요하다고 밝혔다. 이러한 결과가 나온 것은 해당 AI 서비스가 충분한 학습 과정을 거치지 않았기 때문일 것이다. 이미 편향된 데이터 잘못 학습했을 가능성도 존재한다.

앞에서 소개한 구글 사례뿐만 아니라 아마존 채용시스템에 의한 특정 직군 여성 지원자 배제, 범죄예측 프로그램의 과도한 흑인 재범률 예측 사례 등에서 알 수 있듯 오랜 시간 동안 형성된 인간의 편향이 사회에 만연하고 있으며, 그런 사회를 반영한 데이터를 활용해

\* ICT전략연구실 연구원, 043-531-4296, gimun@kisdi.re.kr

AI 서비스를 제공할 경우 개발자가 의도하지 않을지라도 차별을 조장하고 영속화할 가능성이 있다. 『대량살상수학무기』(2017)의 저자 캐시 오닐은 “알고리즘이 공정할 것이라는 생각을 버리고, 알고리즘을 민주적으로 통제해야 한다”고 주장하며, “우리 사회에 차별이 이미 내재해 있기 때문에 알고리즘에도 차별이 발생할 수 밖에 없다”고 말하기도 했다.<sup>1)</sup>

이러한 배경에서 미국 연방거래위원회(Federal Trade Commission, 이하 FTC)<sup>2)</sup>는 위원회 웹페이지 블로그를 통해 AI 및 알고리즘 사용에 대한 지침을 발표하였다. 본 지침은 기업이 AI 및 알고리즘을 사용하는 과정에서 소비자에게 발생할 수 있는 위험을 어떻게 관리할지에 대한 방향을 제시하기 위해 만들어졌다. 이에 본고에서는 미국 FTC에서 발표한 ‘Using Artificial Intelligence and Algorithms(AI 및 알고리즘 사용, 2020.04)’ 지침의 주요 내용을 소개하도록 하겠다.

## 2. AI 및 알고리즘 사용 지침의 주요 내용

FTC는 이번 지침을 통해, 기업이 AI 및 알고리즘을 사용할 때 ① 투명성, ② 설명 가능성, ③ 공정성, ④ 견고성과 실증적 타당성, ⑤ 책임성을 갖출 것을 당부했다. 항목별 주요 내용은 다음과 같다.

### (1) AI 및 알고리즘 활용의 투명성 제고(Be transparent)

먼저, 기업은 AI 기반의 자동화된 도구를 사용함에 있어 소비자를 기만하면 안 된다. 특히, 챗봇(Chatbot)과 같이 소비자와 직접 상호작용을 하는 AI 도구를 사용할 때 소비자를 오도하지 않도록 주의해야 한다. 허위 계정을 생성해 잠재고객의 데이트 서비스 가입을 유도한 애슐리 메디슨(Ashley Madison), 가짜 팔로워를 생성해 판매하여 조희수 등을

- 
- 1) 상세한 내용은 한겨레(2018.11.1) “알고리즘이 불평등 강화할 수도...민주적 통제 나서야” 참조
  - 2) 미국 대통령 직속 독립행정 기관으로 독과점을 규제하고 불공정거래를 규제하기 위해 1914년에 설립되었으며, 우리나라의 공정거래위원회와 유사한 업무를 수행하고 있음

조작한 데부미(Devumi) 사례와 같이 기업이 가짜 프로필, 허위 팔로워 및 가입자, 가짜 ‘좋아요’, 답페이지 등 부적절하게 AI·알고리즘을 활용하여 소비자를 기만하는 행위는 법에 따라 사기로 간주할 수 있으며, FTC의 시행조치를 받을 수 있다.

다음으로, 민감한 데이터를 수집할 때에는 투명성을 담보로 해야 한다. 알고리즘 학습자료로 활용하기 위해 이용자의 음성, 시각적 데이터 및 민감한 데이터를 비밀리에 수집하는 경우 FTC의 조치가 내려질 수 있다. 작년 FTC는 '18년 4월에 업데이트 된 페이스북의 얼굴인식 관련 정책이 사용자를 기만한다고 지적했는데, 기본적으로 얼굴인식 기능이 활성화되도록 설정하여 실질적인 사용자의 동의 없이 얼굴 정보를 수집했다는 이유에서이다.

한편, 공정신용보고법(Fair Credit Reporting Act, 이하 FCRA)에 의하면 신용, 고용, 보험, 주택 또는 유사한 혜택과 거래의 적격성에 관한 의사결정을 자동화하기 위해 소비자의 정보를 조합하는 기업은 신용평가사(Consumer reporting agency)라 판단할 수 있으며, 그에 따른 의무가 부과된다. FCRA에 의해 부과된 의무에 따라 신용평가사는 제3자 공급업체가 제공한 정보를 바탕으로 자동화된 결정을 내리는 경우, 그 과정에서 발생한 불리한 조치를 투명하게 소비자에게 알려야 하며, 이를 통해 소비자는 자신의 정보를 확인하고 부정확한 정보를 수정할 수 있는 권리를 행사할 수 있다.

〈표 1〉 주요 내용 : AI 및 알고리즘 활용의 투명성 제고

주요 내용		세부 내용
1	소비자 기만 금지	• AI 기반 서비스를 제공하는 경우 소비자를 기만하지 않도록 주의 : 허위 계정, 가짜 프로필, 답페이지 등 AI 툴의 부적절한 활용 경계
2	민감한 데이터 수집 시 투명성 담보	• 알고리즘 학습을 위해 생체정보 등 이용자의 민감 데이터를 수집할 경우 투명하게 고지 : 민감 데이터를 비밀리에 수집할 경우 FTC 조사 대상이 될 수 있음
3	불리한 조치에 대한 통지	• 자동화된 의사결정으로 발생한 불리한 조치를 투명하게 통지 : 정보 접근 권한 및 부정확한 정보를 수정할 권리를 소비자에게 부여

자료: FTC(2020.04)를 참고하여 재작성

## (2) AI 및 알고리즘을 활용한 의사결정에 대한 설명 (Explain your decision to the consumer)

기업은 알고리즘에 의해 소비자에게 불리한 결정을 내려야 할 경우 그 이유를 구체적으로 설명해야 한다. 예를 들어 알고리즘 기반 신용평가에서 개인의 신용이 낮게 나온 이유에 대해 단순히 “신용점수가 낮다”, “기준에 부합하지 않는다”라고 말하는 것이 아닌 “신용 체납이 있다” 또는 “신용 참조 건수가 충분하지 않다” 같은 구체적인 이유를 밝혀야 한다. 이는 기업이 판단 과정에서 어떤 데이터를 활용하였고, 그 데이터가 특정 결정을 내리는데 어떻게 활용되었는지를 파악하고 이용자에게 설명할 수 있어야 함을 의미한다.

또한, 기업은 구체적인 설명과 함께 주요 요인을 공개해야 한다. 기업이 신용점수를 근거로 소비자의 대출 신청을 거부하거나 신용 거래에 불리한 조건을 제시하는 경우 신용점수에 관한 설명과 함께 신용점수에 부정적인 영향을 미친 상위 주요 요인을 최소 4가지 공개해야 한다.

AI·알고리즘에 의해 거래조건이 변경될 때에도 소비자에게 반드시 알려야 한다. 10여 년 전, FTC는 신용카드사 컴퓨크레딧(CompuCredit)이 고객의 이용 패턴을 신용 한도에 반영하였고, 이 사실을 고객에게 알리지 않았으므로 FTC의 규정을 위반했다고 지적한 바 있다. FTC는 컴퓨크레딧이 고객들이 신용카드로 현금 서비스를 이용하거나, 술집, 나이트클럽 등 특정 장소에서 자주 결제할 경우 신용한도를 줄일 수 있었음에도 이 사실을 가입 고객에게 알리지 않았고, 신용한도를 낮출 때도 통지하지 않았기 때문에 규정 위반이라 판단하였다.

〈표 2〉 주요 내용 : AI 및 알고리즘을 활용한 의사결정에 대한 설명

주요 내용		세부 내용
1	결정에 대한 구체적 이유 설명	• 소비자에게 불리한 결정을 내릴 때, 그 이유를 구체적으로 설명 : 활용 데이터 및 데이터 활용방법을 구체적으로 설명할 수 있어야 함
2	영향을 미친 상위 주요 요인 공개	• 부정적 결과에 대한 설명과 함께 주요 요인 공개 : 결과에 영향을 준 상위 주요 요인을 최소 4가지 공개할 수 있어야 함
3	거래조건 변경 시 통지	• AI·알고리즘에 의해 거래조건이 변경될 때 소비자에게 반드시 통지 : 이용자에게 고지하지 않을 경우 FTC 규정 위반이라 판단할 수 있음

자료: FTC(2020.04)를 참고하여 재작성

### (3) 결과의 공정성에 대해 보장(Ensure that your decisions are fair)

먼저, 특정 집단이나 계층에 대한 차별이 발생하지 않도록 주의해야 한다. 무신경한 AI의 사용은 취약계층에 대한 차별을 일으킬 수 있으며, 이는 신용업무 취급 차별금지법(Equal Credit Opportunity Act, 이하 ECOA)과 1964년 민권법(제7편) 등 연방 평등 기회 법과도 관련된 문제이다. FTC는 ECOA에 따라 AI 사용에 있어 인종, 종교, 국적, 성별 등에 따른 차별적 영향을 금지하고 있다. 예컨대 기업이 소비자의 우편번호에 근거하여 신용 결정을 내림으로써 특정 인종이 거주하는 지역에 불리한 영향을 준다면 FTC는 ECOA를 근거로 이러한 관행에 이의를 제기할 수 있다. 따라서 기업은 알고리즘이 취약계층에 차별적 영향을 주지 않도록 사용 전과 후 엄격히 테스트하고 관리해야 한다.

또한, 기업은 입력 데이터의 차별적 요소뿐만 아니라 별개로 도출되는 결과에도 차별이 발생하지 않았는지 통제·관리해야 한다. 차별이 발생할 것으로 예상되는 데이터를 사용하지 않았음에도 취약계층에게 차별이 발생한다면 이에 대한 추가 검토를 해야한다는 것이다.

마지막으로, 의사결정에 사용되는 정보에 대한 접근 권한과 수정 기회를 소비자에게 제공해야 한다. FCRA에 의하면 소비자는 취업, 신용, 보험 등 자신과 관련한 의사결정에 활용되는 정보가 무엇인지 알 권리가 있으며, 정보가 부정확하다고 생각될 경우 해당 정보에 대해 이의를 제기할 수 있다. 기업은 이러한 권리를 소비자에게 안내해 그들의 정보 접근 권한 및 수정 기회를 보장해야 한다.

〈표 3〉 주요 내용 : 결과의 공정성에 대해 보장

주요 내용		세부 내용
1	특정 집단·계층에 대한 차별 금지	• AI 사용 시 인종, 종교, 국적, 성별 등 특정 집단·계층에 대한 차별 금지 : 알고리즘이 차별을 유발하지 않도록 사용 전후 엄격한 관리가 필요
2	결과의 공정성 보장	• 입력값의 차별적 요소와 별개로 도출되는 결과에도 차별이 발생하지 않도록 통제·관리 : 검토를 통해 결과의 공정성도 보장해야 함
3	정보 접근 권한 및 수정 기회를 소비자에게 제공	• 의사결정에 대한 판단근거로 사용되는 정보에 대한 접근 권한과 수정 기회를 소비자에게 제공 : 기업은 이용자의 권리를 소비자에게 안내해 권리와 기회를 보장해야 함

자료: FTC(2020.04)를 참고하여 재작성

#### (4) 데이터·모델의 견고성 및 실증적 타당성 보장

(Ensure that your data and models are robust and empirically sound)

기업이 신용, 고용, 보험 또는 이와 유사한 거래에 대한 소비자 정보를 제3자에게 제공하는 경우, 해당 기업은 신용평가사(Consumer reporting agency)로 분류될 수 있으며, 정보의 정확성과 최신성을 유지하는 등 FCRA를 준수해야 한다. 무엇보다도, 기업은 개인 신용정보의 정확성을 가능한 최대로 보장하고 오류를 수정할 수 있는 기능과 함께 소비자들이 자신의 정보에 접근할 수 있도록 합리적인 절차를 마련할 의무가 있다.

신용평가사가 아닌 기업의 경우에도 고객의 데이터를 자동화된 의사결정 시스템 활용하려는 제3자에게 제공한다면 데이터의 정확성을 확인할 의무가 있다. 신용평가사에 정보를 제공하는 기업들(Furnishers)이 여기에 해당되며, 이들은 제공 데이터가 정확성·무결성을 갖도록 명문화된 정책과 절차를 마련해야 한다. FTC는 이러한 정책 및 절차를 유지하지 못한 기업에 조치를 취하고 벌금을 부과하고 있다.

또한, 기업은 AI 모형이 설계 의도와 맞게 작동하고, 불법적 차별을 일으키지 않도록 AI 모형의 유효성을 검사하고 재확인해야 한다. 명확하고 실증적인 통계 원칙과 방법론을 사용하여 주기적으로 AI 모형의 유효성을 재검증해야 하며, AI 모형은 예측력 유지를 위해 필요에 따라 조정되어야 한다.

〈표 4〉 주요 내용 : 데이터·모델의 견고성 및 실증적 타당성 보장

주요 내용		세부 내용
1	정보의 정확성과 최신성 유지	<ul style="list-style-type: none"> <li>• 소비자 정보를 제3자에게 제공하는 경우 정보의 정확성과 최신성 유지 : 정확성을 최대한 보장하고, 정보 접근 및 오류 수정을 위한 합리적인 절차를 마련할 의무가 있음</li> </ul>
2	명문화된 정책과 절차 마련	<ul style="list-style-type: none"> <li>• 제공 데이터가 정확성·무결성을 갖도록 명문화된 정책과 절차 마련 : 정책 및 절차를 유지 못할 시 FTC가 제재할 수 있음</li> </ul>
3	AI 모형의 유효성 검사	<ul style="list-style-type: none"> <li>• 기업은 AI 모형이 설계 의도와 맞게 작동하고 불법적 차별을 일으키지 않도록 AI 모형의 유효성을 검사 : 명확하고 실증적인 통계 원칙과 방법론을 사용하여 모형의 유효성을 재검증하고 예측력 유지를 위해 모형을 조정해야 함</li> </ul>

자료: FTC(2020.04)를 참고하여 재작성

### (5) 법령 준수, 윤리, 공정성 및 비차별성에 대한 책임 견지

(Hold yourself accountable for compliance, ethics, fairness, and nondiscrimination)

빅데이터의 분석 결과가 소비자에게 편견이나 피해를 유발하는 것을 방지하기 위해 알고리즘 운영자는 알고리즘을 사용하기 전 자가 질문을 통해 다음의 사항을 점검해야 한다 : △ 데이터가 얼마나 대표성을 지니고 있는가? △ 데이터 모델의 편향성을 보정할 수 있는가? △ 빅데이터 기반 예측이 얼마나 정확한가? △ 빅데이터에 의존할 경우 윤리적 또는 공정성 문제가 발생할 우려는 없는가?

또한, 개발된 AI 툴이 악용될 가능성이 없는지, 접근 제어 및 기타 기술을 이용해 AI 툴의 남용을 막을 수 있는지를 검토해야 한다. 한 예로, FTC의 음성복제 기술 워크숍에 참여한 한 기업은 당사의 기술이 사기 등 범죄에 악용될 경우를 대비해 사용자를 검토하고, 해당 기술을 자체 서버에서 운영하는 등 기술의 남용 방지에 심혈을 기울이고 있다.

마지막으로, 개발한 AI에 대해 책임을 다하는 방법에 대한 고려가 필요하다. 이를 위해 독립적인 표준을 마련하고 외부 전문가를 활용하는 등 객관적인 시각에서 당사의 AI를 평가하는 체계를 구축하는 것을 고려해볼 수 있다.

〈표 5〉 주요 내용 : 법령 준수, 윤리, 공정성 및 비차별성에 대한 책임

주요 내용		세부 내용
1	자가 점검을 통한 편견·피해 방지	• 알고리즘으로 인한 피해 방지를 위해 사용 전 자가 점검 실시 : 대표성, 편향성 보정 가능성, 정확성, 문제 발생 가능성 등 점검
2	무단 사용에 대한 알고리즘 보호	• 개발된 AI 툴의 악용 및 무단이용 가능성 검토 : 범죄에 악용될 가능성을 검토하고 대비책을 마련해야 함
3	책임 메커니즘 구축방안 고려	• 개발한 AI에 대해 책임을 다하는 방안 고려 : 객관적인 시각에서 당사가 개발한 AI를 평가할 수 있는 체계를 구축하는 것을 고려해야 함

자료: FTC(2020.04)를 참고하여 재작성

### 3. 결어

의도하지 않았다고 할지라도, AI 및 알고리즘이 우리 사회에 만연한 차별을 반영하고 가시화할 수 있음이 다양한 사례를 통해 드러나고 있다. 심지어는 특정 집단과 계층에 대한 차별을 영속화하고 증폭시킬 수 있다는 우려도 계속 제기되고 있다. 불행히도, AI 알고리즘의 불합리한 결과를 인지하고 누군가가 이의를 제기할 때에는, 이미 많은 사람이 모르는 사이 해당 AI·알고리즘의 편향된 영향을 받은 후가 된다. 그렇기 때문에 향후 발생할 수 있는 예기치 못한 기술적·사회적 위험을 최소화하기 위한 선제적 대응이 필요하며, 이러한 맥락에서 FTC가 제시한 AI 및 알고리즘 사용에 관한 지침의 문제해결 접근법은 시의적절한 것으로 보인다. 특히 기존 법규만으로는 설명할 수 없었던 구체적 행동 모델과 사례를 기업에 소개, 실용적인 대책을 제시하였다는 점에서 그 의의가 크다고 할 수 있다.

우리나라에서도 AI 기반 서비스가 발전하고 상용화되어 일상속에서의 활용이 늘어남에 따라 AI·알고리즘에 대한 알 권리, 설명요구권 등에 대한 이용자 권리의식도 점차 높아지고 있다. 2019년 방송통신위원회와 정보통신정책연구원이 진행한 “지능정보사회 이용자 패널조사”에 따르면 ‘이용자는 AI 기반의 서비스를 사용하다 발생한 사고로 인한 피해나 불이익에 대해 설명을 요구할 권리가 있다’라는 의견에 49.3%가 찬성했다. 특히, ‘이용자는 알고리즘이 선별한 뉴스의 선별기준을 알 권리가 있다’라는 의견에는 전년 대비 11.7%p가 늘어난 49.2%가 동의하는 것으로 나타났다. AI 서비스 이용자의 권리의식은 높아졌지만 ‘잘 모르겠다’나 ‘중립’으로 응답한 의견도 상당수를 차지하였으며, 이는 여전히 적지 않은 수의 이용자가 스스로의 권리에 대해 제대로 인지하지 못하고 있음을 시사한다.

이러한 상황을 종합적으로 고려해보았을 때, FTC의 “Using Artificial Intelligence and Algorithms”은 인공지능이 본격화되는 미래 사회에 발생할 수 있는 위험을 방지하기 위한 실질적인 지침이자 상황에 따라 우리나라 이용자 권리 보장 및 권익 향상을 위한 좋은 참고자료가 될 수 있다. 물론, 이를 위해서는 한국의 기술적·사회적 환경에 맞게 현실화하는 과정이 선행되어야 할 것이다.



### 〈참고문헌〉

방통위(2020.05), “2019년 지능정보사회 이용자 패널조사 결과 발표” 보도자료  
AlgorithmWatch(2020.04.08), “Google apologizes after its Vision AI produced  
racist results”, <https://algorithmwatch.org/en/story/google-vision-racism/>  
FTC(2020.04), “Using Artificial Intelligence and Algorithms”,  
<https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms>