

부패범죄에 AI 범죄예측 시스템 적용의 문제점과 개선방안*

Problems and Improvements of AI Crime Prediction System for Corruption Crime

김 병 수(Kim, Byung Soo)**

ABSTRACT

Corruption crime is not limited to the deviance of one actor, but is a crime with destructive power that can destroy the social structure. Although the state and society are making great efforts to reduce corruption crime, it is not easily detected or completely eradicated due to the confidentiality and customary nature of corruption crime. Therefore, it is necessary to explore whether there is an effective way to reduce corruption by introducing an AI crime prediction system. This is because many countries are already seeing a positive effect of reducing the crime rate by introducing AI crime prediction systems. However, it is necessary to examine whether the AI crime prediction system can be applied to corruption crimes and what problems there are if it is applied. Furthermore, we intend to explore the improvement measures for applying the AI crime prediction system to corruption crimes.

AI crime prediction systems that handle personal information can infringe on individual rights. And the algorithm of AI crime prediction system has problems of opacity, inaccuracy, and bias. AI crime prediction systems are becoming a new means of control for citizens, such as 'Big Brother'.

As a solution to bias and inaccuracy, the AI crime prediction system should be descriptive, accurate, and audited by an objective agency. In order to limit the individual control function of the AI crime prediction system, the AI crime prediction system must also apply the principle of proportionality to minimize the infringement of the subject's rights (the principle of minimal infringement). And AI crime prediction systems should be used as a last resort (principle of complementarity). It should be used individually according to the subject's risk of recidivism (principle of individuality). The AI crime prediction system should not be implemented by administrative disposition, but should be imposed by judicial review of the court (principle of judicial control). If the analysis of the AI crime prediction system is applied according to the principle of imposition of such security measures, the individual control function of the AI crime prediction system can be limited as much as possible and the efficiency as a control method for corruption crimes can be increased.

Key words: corruption crime, crime prediction system, artificial intelligence, causes of corruption crimes, algorithm, crime prevention

* 이 논문은 2021년 한국부패학회·동아대학교 법학연구소가 공동개최한 동계학술대회의 발표논문을 수정·보완한 것임.

** 동의과학대학교 경찰경호행정과 교수, 법학박사.

I. 들어가며

사회의 많은 분야에서 머신러닝 및 딥러닝 기술을 바탕으로 한 인공지능(AI)이 빅데이터 속에서 숨겨진 패턴이나 상관관계를 찾아내고 있다. 이를 바탕으로 지금까지 알지 못했던 새로운 지식을 발굴하여 앞으로 어떠한 일이 발생할 것인지를 전망하고 있다. 이러한 인공지능의 예측은 통계적으로도 유의미한 결과를 내놓고 있어 금융, 마케팅, 운송, 의료, 공공 서비스 등 사회전반에 활용되고 있다. 이러한 AI의 높은 예측력은 범죄예방분야에서도 활용되고 있는데, 그것이 바로 AI 범죄예측 시스템이다. AI 범죄예측 시스템은 범죄와 관련된 대규모의 정보 즉 빅데이터(Big Data)를 수집 및 분석하여 범죄의 의심, 위험 그리고 명백한 사례 등을 탐지하고 예측하는 시스템이다. 이미 AI 범죄예측 시스템을 도입한 선진국에서는 범죄율이 감소하는 등 가시적인 성과를 나타나고 있다.¹⁾ 따라서 이러한 AI 범죄예측 시스템을 활용하여 부패범죄를 줄일 수 있는 방안을 모색할 필요가 있다. 왜냐하면 부패범죄는 한 행위자의 일탈에 그치는 것이 아니라 사회구조를 무너뜨릴 수 있는 파괴력을 지니는 범죄로서 한 국가의 문제만이 아니라 전 세계적으로도 심각한 문제이기 때문이다. 또한 부패범죄가 가지고 있는 은밀성과 관행성 때문에 쉽게 적발되지도 않을 뿐만 아니라 완전히 근절되기도 어렵기 때문에 AI 범죄예측 시스템을 도입하여 부패범죄를 줄일 수 있는 효과적인 방안이 없을지 모색할 필요가 있다.

따라서 본 논문에서는 이러한 AI 범죄예측 시스템을 부패범죄에 적용할 수 있을지(II) 그리고 적용한다면 어떠한 문제점이 있는지를 검토할 필요가 있다(III). 나아가 부패범죄에 AI 범죄예측 시스템을 적용하기 위한 개선방안이 무엇인지 모색하고자 한다(IV).

II. 부패범죄에 AI 범죄예측 시스템 적용의 필요성

1. 국가경쟁력을 저해하는 부패범죄

2020년 국가경쟁력 평가에서 우리나라는 평가대상국 63개 중에서 R & D 투자액 1위, 과학인프라 3위, 기술인프라 13위, 디지털 경쟁력 10위 등 경제분야에서는 높은 경쟁력을 보

1) Zach Friend, Predictive Policing: Using Technology to Reduce Crime, FBI Law Enforcement Bulletin(Feb. 8, 2018).; Jeffrey S. Paul and Thomas M. Joiner, Integration of Centralized Intelligence with Geographic Information Systems, Geography and Public Safety, Oct. 2011, at 7.; Jennifer Bachner, Predictive Policing: Preventing Crime with Data and Analytics, Washington, DC: IBM Center for The Business of Government, 2013.; C. Francescani, "NYPD Expands Surveillance net to Fight Crime as Well as Terrorism," Reuters, June 21, 2013.

이고 있지만, 한 국가의 부패수준을 나타내는 국제투명성기구(ITI)의 국가별 부패인식지수(CPI)에서는 2020년 전체 180개국 중에서 31위에 불과하다.²⁾ 즉 우리나라가 경제적으로는 선진국의 반열에 올랐지만 사회에 만연하는 부패로 인하여 국제적인 국가경쟁력에서는 떨어진다. 이처럼 부패는 사회의 투명성을 판단하는 바로미터이자 국가의 국제적인 경쟁력 지표이다. 부패범죄를 척결하여 국가경쟁력에 걸맞는 우리 사회의 투명성 확보가 절실하다. 하지만 그 동안 청탁금지법(부정청탁 및 금품등 수수의 금지에 관한 법률), 이해충돌방지법(공직자의 이해충돌 방지법), 공수처법(고위공직자범죄수사처 설치 및 운영에 관한 법률) 등 부패범죄를 감소시키는데 국가와 사회가 많은 노력을 기울이고 있지만, 부패범죄는 끊임없이 발생하고 있다. 이것은 부패범죄가 가지고 있는 은밀성과 관행성으로 인하여 쉽게 적발되지도 않을 뿐만 아니라 완전히 근절되기도 어렵기 때문이다.

2. 부패범죄 수사의 문제점

부패범죄는 보통 내부자 간에 은밀하게 이루어지기 때문에 부패범죄의 인지 자체가 어려워 수사가 개시되지도 못하는 경우가 많다. 그리고 부패범죄의 피해자는 전체 국민이거나 국가이기 때문에 피해자가 특정되지 않아 피해자의 고소나 고발이 제대로 이루어지지 않는다. 은밀하게 진행된 부패범죄가 세상에 드러난 경우에도 공익단체나 국가기관이 고소 또는 고발을 하지 않는 이상 형사절차로 진행되기 어렵다. 또한 부패범죄는 직접적인 피해자가 없어 증거를 수집하거나 확보하는데도 어려움이 있다. 이 때문에 고소 또는 고발이 이루어진다고 해도 증거부족으로 유죄를 입증하기가 쉽지 않다.

공직자에 의한 부패범죄는 대표적인 화이트칼라 범죄로서 수사기관이 ‘축소수사’, ‘파주기식 수사’ 등 차별적인 수사를 하는 경향이 있었다. 과거 우리나라에서는 고위공직자의 부패범죄에 대한 수사가 정치적인 성향에 따라 좌우되거나 정치권력, 경제권력 등의 외압에 의해 공정하게 진행되지 못한 선례가 많았다. 특히 공직자의 부정부패 사건에 대하여 정치적인 외압이나 고려에 의하여 검찰의 기소재량권이 남용되었다.³⁾ 수사기관은 정치권력이나 경제권력이 원하는 대로 ‘표적수사’ ‘먼지털이식 수사’ 등 편파수사를 진행하였다는 의혹을 받아왔다.⁴⁾

2) 국민권익위원회, 2020년 부패인식 조사결과, 2020.

3) 정신교, “뇌물죄의 양형합리화 방안”, 치안정책연구 제23호, 치안정책연구소, 2009, 136면.

4) 김병수, “검찰개혁의 필요성과 추진방향”, 한국부패학회보 제25권 2호, 2020, 109면; 이주희, “고위공직자 부패 방지에 관한 형법적 고찰”, 한양법학 제24권 제3집, 한양법학, 2013, 287면.

3. 부패범죄에 AI 범죄예측 시스템 적용 현황

범죄통계에 따르면 범죄를 신고하는 전화의 절반이 도시 전체 주소 중에서 3.5%에 해당하는 특정 지역에서 걸려 왔고,⁵⁾ 범죄가 많이 발생하는 장소는 전체 도시의 도로 중에서 4.5%에 해당하는 특정 도로에서 발생하였고,⁶⁾ 전체 범죄자 중의 10%에 해당하는 범죄자가 전체 강력범죄의 50%를 저지른다고 한다.⁷⁾ 또한 연쇄살인범의 상당수는 어린시절 동물학대와 방화와 같은 일탈행위를 한 경험이 있으며,⁸⁾ 부모나 주위 사람으로부터 학대를 당한 경험이 있고⁹⁾ 그에게 학대를 행한 사람은 대부분 그의 부모였다.¹⁰⁾ 이처럼 범죄자와 범죄의 시간과 장소 등에는 일정한 경향이 발견되면서 범죄예측의 연구가 시작되었다. 오늘날에는 인공지능이 빅데이터에서 일정한 패턴을 찾아내고 이를 근거로 군집·최적화·예측·분류 등을 위한 판단규칙을 스스로 학습하는 머신러닝을 사용하여 범죄를 예측하고 있다.¹¹⁾

이러한 AI 범죄예측 시스템은 크게 인적 기반 AI 범죄예측 시스템과 장소 기반 AI 범죄예측 시스템으로 나눌 수 있다.¹²⁾ 인적 기반 AI 범죄예측 시스템은 범죄자 및 피해자의 정보를 수집하여 장래의 범죄발생을 예측하는 방법으로, 범죄 피해의 대상이 될 가능성이 높은 집단이나 개인을 전망하는 ‘범죄 피해자 예측’과 범죄를 범할 가능성이 높은 개인을 예측하는 ‘범죄자 예측’ 등이 여기에 해당한다. 재범의 위험성을 측정하는 연구¹³⁾가 이러한 인

- 5) Sherman, L.W., Gartin, P.R., & Buerger, M.E., “Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place,” *Criminology*, 27, 1989, pp.27-55.
- 6) 시애틀에서 14년간의 범죄 데이터를 분석한 Harthorne은 범죄의 절반이 도시 전체 도로의 4.5%에서 발생하는 것을 밝혀냈다. Harthorne M., City : A Few streets Responsible for Half of seattle's Crime; komonews.com/archive/city-a-few-streets-responsible-for-half-of-seattles-crime.
- 7) Nath, S.V., Crime Pattern Detection Using Data Mining, Oracle Corporation, 2006.
- 8) Wright, J., & Hensley, C. (2003). et. al.
- 9) Holmes, S. T., Tewksbury, R., & Holmes, R. M. (1999). Fractured Identity Syndrome. *Journal of Contemporary Criminal Justice*, 13(3), p.265 ; Whitman, T. A., & Akutagawa, D. (2004). Riddles in serial murder: A synthesis. *Aggression and Violent Behavior*, 9, p. 699.
- 10) Singer, S. D., & Hensley, C. (2004). Applying Social Learning Theory to Childhood and Adolescent Firesetting: Can It Lead to Serial Murder?. *International journal of offender therapy and comparative criminology: official organ of the Association for Psychiatric Treatment of Offenders (APTO)*, 48(4), p.465.; Wright, J., & Hensley, C. (2003). From Animal Cruelty to Serial Murder: Applying the Graduation Hypothesis. *International Journal of Offender Therapy and Comparative Criminology: official organ of the Association for Psychiatric Treatment of Offenders(APTO)*, 47(1), p.88.
- 11) 이태익, “[이태익의 과학에세이] 스물데이터 AI와 뉴턴의 사과”, IT조선, 2018. 7. 7., http://it.chosun.com/site/data/html_dir/2018/07/06/2018070602627.html
- 12) 윤상현/신상화, “범죄 예측의 현재와 미래- 한국형 범죄예측 시스템 개선·도입에 대한 제언 -”, 한국공안행정학회보, 제71호, 2018, 251-253면.
- 13) 김태완, “데이터 마이닝 기법을 활용한 청소년 재범 관련요인 연구”, 교정복지연구(제35호), 한국교정복지학회(2014. 12); 이수정, “성인 판결전 조사 활용, 재범 위험성 평가도구 개발”, 동향/연구보고서, 법무부(2010).

적 기반 범죄예측 시스템을 활용하였다. 그리고 장소 기반 AI 범죄예측 시스템은 장소, 계절 등과 같은 사회환경에 관한 정보를 활용하여 범죄 발생의 가능성이 높은 시간이나 장소를 전망하는 ‘범죄 예측’이 여기에 해당한다. 아래에서는 인적 기반 AI 범죄예측 시스템과 장소 기반 AI 범죄예측 시스템을 활용하고 있는 사례들과 부패범죄에 AI 범죄예측 시스템을 사용한 사례를 살펴보겠다.

1) 인적 기반 AI 범죄예측 시스템

영국 런던경찰청의 범죄예측 시스템인 OASys(Offender Assessment System)는 대표적인 인적 기반 AI 범죄예측 시스템이다. OASys는 특정인의 개인정보 특히 조직범죄 데이터베이스에 있는 우범자들의 범죄기록뿐만 아니라 주거, 소득, 교육, 친구관계, 생활습관, SNS 기록 등을 바탕으로 행동의 패턴을 알아내고 재범 가능성을 예측하고 있다.¹⁴⁾ 그리고 NDAS(National Data Analytics Solution)은 빅데이터를 검토하여 범죄예측에 유의미한 지표를 구축하였는데, 영국 경찰은 이러한 지표들을 활용하여 범죄 발생 가능성이 크다고 예측되는 사람들을 상담하고 범죄예방을 위한 지원활동을 하고 있다. 이것은 특정한 사회 관계망(예를 들어 조직범죄단체, 폭력 서클, 경기장 홀리건 등)에 속한 사람들은 앞으로 범죄를 저지를 가능성이 높다는 전제를 바탕으로 하고 있다.¹⁵⁾ NDAS와 OASys의 범죄예측 시스템은 범죄발생 위험성이 큰 시간과 장소를 지정하는 것보다는 특정인을 주목하여 장차 범죄자가 될 가능성이 높은 사람을 주요 대상으로 한다는데 특징이 있다.¹⁶⁾

2) 장소 기반 AI 범죄예측 시스템

장소 기반 AI 범죄예측 시스템에는 미국 LA 경찰청(LAPD)이 예방적 경찰활동(Predictive Policing)으로 개발한 프레드폴(PredPol, PREDictive POLicing)이 대표적이다. 프레드폴은 과거의 범죄에 대한 자료를 머신러닝 알고리즘을 이용하여 분석하여 언제(시간) 어디서(장소) 어떤 유형의 범죄가 발생할지를 예측한다. 여기서 나온 결과에 기초하여 범죄 발생의 위험성이 높은 지역과 시간에 경찰력을 우선적으로 투입하고 경찰의 순찰경로도 그 우선순위에 따라 조정하는 등의 기능을 제공한다.¹⁷⁾ 프레드폴은 다양한 범죄유발 요인들을

14) 유병후, “빅데이터를 활용한 범죄예방서비스 창출방안”, 한국범죄심리연구 제16권 제3호, 2020. 71면; 박민지/서정범, “사물인터넷을 활용한 경찰활동에 관한 연구- 그 현황과 허용성에 관한 논의를 중심으로 -”, 경찰학연구 제19권 제3호, 경찰대학, 2019. 9, 101면.

15) 행정안전부/한국정보화진흥원, 월간 디지털 정부 최신 해외 정책·기술 동향 제40호, 2018, 40면.

16) 윤영석, “범죄징후 예측시스템에 대한 비판적 검토-2019년 법무부의 발표내용을 중심으로-”, 과학기술법연구 제26집 1호, 2020, 142-143면.

17) PredPol available from <http://www.predpol.com/>

효과적으로 시공간적으로 반영하기 위하여 위험영역모델(Risk Terrain Model)과 인근 반복 모델(Near Repeat Model)과 같은 시공간 분석(Spatial-temporal Analysis) 통계기법을 주로 사용한다.¹⁸⁾ 이러한 시스템들을 사용하는 프레드폴은 지역의 다양한 환경에 관한 정보를 수집 및 분석하여 범죄 예측의 정확성을 향상시키고 도시 및 광역 지역의 범죄에 유연하게 대처할 수 있다는 점이 특징이다.¹⁹⁾

우리나라도 범죄유형, 범죄발생 공간정보, 인구통계 등에 관한 범죄통계자료와 같은 정형 데이터 뿐만 아니라 주민신고 등과 같은 비정형데이터를 분석하여 범죄유형과 장소 및 시간을 예측하는 지오프로스(GeoPros)나 CLUE(Crime Layout Understanding Engine) 등이 개발되었다. 이를 통해서 범죄경향분석, 핫스팟(hot spot), 범죄 위험지수 등의 정보를 제공하고 이러한 정보를 우선적으로 경찰력을 투입할 장소를 결정하거나 경찰의 순찰경로 선정하는데 활용하고 있다.²⁰⁾

3) 부패범죄에 AI 범죄예측 시스템 적용 사례

미국 보건의료분야에서 부당하고 불법적인 재정지출이 증가하자 보건복지부소속 감사부서(OIG)와 법무부(DOJ) 등이 협업하여 보건의료분야의 부패예방 및 적발을 위해 공공데이터를 개방하고 공유하였다. 이를 위해 데이터 Portal 등을 활용하여 데이터의 투명성과 개방성을 한층 확대하였고 그와 함께 머신러닝을 이용한 알고리즘을 개발하여 보건의료 분야에서 재정지출의 불법적이고 부당한 패턴을 탐지하였다. 이를 통해서 부당청구 등으로 이미 발생한 재정손실을 회수하는 것뿐만 아니라 보다 더 효율적이고 효과적으로 사전에 부패행위를 탐지할 수 있어 보건의료분야의 부패방지과 예방에 도움을 주고 있다.²¹⁾ 또한 금융기관 및 금융정보기관 그리고 금융관련 서비스업체들이 미국은행(US Bank)과 협업하여 인공지능을 이용한 이상거래 탐지시스템(FDS)의 일종인 자금세탁방지시스템(AML System)을 운영하여 은행계좌 간 이체내역, 신용카드 사용내역 등을 모니터링하여 정상적인 거래패턴에서 벗어난 이상 징후를 탐지하여 수사기관에 신고하고 있다.²²⁾

중국은 2012년부터 중국공산당 내부의 감시기관과 중국과학원이 공동개발한 인공지능시스템(Zero Trust)으로 중앙 및 지방정부의 150개 데이터베이스에 접속하여 관련 정보들을 이용하여 정부자금 유용, 공무원의 횡령, 권력남용, 친족고용 등 범죄를 적발하고 있다. 이

18) 윤상현/신상화, 앞의 글, 251-253면.

19) 윤상현, “범죄예측 기술 어디까지 왔나?”, 치안정책리뷰 제 60호, 치안정책연구소, 2018, 5-8면.

20) 윤상현, 앞의 글, 5-8면.; 한국전자통신연구원, “위험 상황 초기 인지를 위한 ICT 기반의 범죄 위험도 예측 및 대응 기술 개발,” 한국전자통신연구원 사업계획서, 2018.

21) 한세익, “AI기반 청렴성 증강모델”, 한국부패학회보 제26권 제4호, 2021, 172-173면; 김민정, “공공데이터를 활용한 부패방지 사례분석 - 주요국 보건의료분야를 중심으로-”, 감사원 감사연구원, 2019.

22) 금융보안원, “머신러닝 기반의 이상거래 탐지시스템 동향”, 2017년 8월 25일 배포한 보도자료 참조.

시스템은 주택철거를 비롯해 부동산 이전이나 토지 획득, 사회기반시설 건설 등에서 수상한 거래를 잡아내는 데 유용하다. 또한 은행예금이 비정상적으로 급증하는 경우도 감지가 가능하며, 공무원 본인은 물론 가족 그리고 지인의 이름으로 새 자동차를 구입하거나 부동산을 매입하거나 정부계약에 응찰했는지도 확인이 가능하다. 이 같은 신호가 발생하고 이상한 거래가 탐지되거나 일정 기준을 넘어설 경우, 당국은 일정 단계에서 감시대상자에게 상급자를 보내 부패행동을 중단하도록 경고하고 있다.²³⁾

멕시코는 일련의 개혁조치를 통해 지난 10년 동안 경제가 성장하고 부패문제가 해소되는 가시적인 성과를 이루었다. 멕시코 국세청(Tax Administration Services)이 인공지능 알고리즘을 이용하여 부패와 관련한 불규칙적인 활동을 분석하고 식별할 수 있는 분석도구를 개발하였다. 이를 활용하여 납세기업 간 부정영업을 적발하는 프로젝트를 시범적으로 운영하여 6개월간 3,500건의 부정거래를 탐지하고 1,200개의 부정거래 회사를 적발하였다. AI의 활용이 없었다면 부패와 관련한 부정행위를 식별하고 분석하는데 18개월 이상의 많은 기간이 소요되었을 것이다. 그러나 AI의 활용함으로써 시간과 자원이 절약할 수 있었던 것이다. 그 결과, 멕시코는 경제의 많은 분야에서 가시적인 성과가 나타났다. 예를 들어 독점적 지배영역이었던 통신사업이 이러한 AI를 활용함으로써 경쟁체제로 전환되어 연결비용 등이 대폭 절감되었다.²⁴⁾

4. 부패범죄에 AI 범죄예측 시스템 적용의 필요성

부패범죄가 가지고 있는 은밀성과 관행성 때문에 쉽게 적발되지도 않을 뿐만 아니라 정치, 경제, 사회의 구조적인 문제로 인하여 완전히 근절되기도 어렵다. 부패범죄는 은밀하게 이뤄지기 때문에 부패와 관련한 불규칙한 활동은 식별하고 분석하는데 많은 시간과 자원이 소요된다. AI 범죄예측 시스템을 활용하면 데이터의 수집, 분석, 동적 시각화, 모델링에 있어 시간과 자원을 절약할 수 있다. 이러한 결과는 미국은행(US Bank)과 멕시코 국세청(Tax Administration Services) 등의 사례에서도 확인할 수 있다. 그리고 AI 범죄예측 시스템을 활용하여 탐지한 결과물의 객관성도 높일 수 있다.²⁵⁾ 특히 공직자에 의한 부패범죄는 대표적인 화이트칼라 범죄로서 수사기관이 정치권력이나 경제권력에 의해 휘둘려 공정하게 수사를 진행하지 못하고 그 수사결과에 대해서도 편파수사라는 비난을 받았던 선례가

23) 동아일보, “中, 부패방지 AI로 공무원 8721명 횡령 등 적발”, 2019년 2월4일자, 출처 <https://www.donga.com/news/Inter/article/all/20190204/93985541/1>

24) 한세익, “AI기반 청렴성 증강모델”, 한국부패학회보 제26권 제4호, 2021, 174-175면; P. Aarvik, Artificial Intelligence-a promising anti-corruption tool in development setting? U4 Report, 2019-1, p. 4.

25) A. Petheram · I. N. Asare, From Open Data to Artificial Intelligence: The next Frontier in Anti-Corruption, Oxford Insights. Oxford Insights. 2018.

많았다. 그러나 수사기관과 달리 외압으로부터 영향을 받지 않는 AI 범죄예측 시스템은 스스로 부패범죄를 식별할 수 있어 수사의 공정성을 확보할 수 있다. 이와 같은 장점이 있는 AI 범죄예측 시스템을 활용하여 부패범죄를 줄일 수 있는 효과적인 방안을 모색할 필요가 있다.

Ⅲ. 부패범죄에 AI 범죄예측 시스템 적용의 문제점

1. AI 범죄예측 시스템의 대상 범죄로서 형평성

AI 범죄예측 시스템은 그 대상이 될 범죄를 선정하는 것이 우선 과제인데, AI 범죄예측 시스템의 대상 범죄를 어느 범위까지로 한정할 것인가는 개별 범죄들간의 형평성의 문제를 야기할 수 있다. 즉 ‘테러범죄’ 또는 ‘살인범죄’ 등과 같이 중대범죄에 한정할 것인가 절도나 폭행과 같은 경미한 범죄까지 포함할 것인가는 형평성의 문제뿐만 아니라 국가자원의 배분권 문제에도 관련이 있기 때문이다.²⁶⁾ 예를 들어 부패범죄를 범죄예측 시스템의 대상 범죄로 한다면 테러범죄나 살인범죄 등과 같은 중대범죄만큼이나 사회적 위험성이 있는지와 한정된 국가자원을 중대범죄인 테러범죄나 살인범죄보다 우선적으로 배분해야 하는지도 검토되어야 할 것이다.

2. 헌법과 법률상의 권리침해

개인정보를 개인식별정보이든 비(非)개인식별정보이든 민감정보이든 비(非)민감정보이든 불문하고 AI 범죄예측 시스템으로 분석하게 되면 필연적으로 헌법상 보장된 국민의 기본권인 사생활의 비밀과 자유 및 개인정보자기결정권 등이 침해될 수밖에 없다. 왜냐하면 아무리 익명화되고 비식별조치가 된 정보라도 다수의 정보가 모여지게 되면 특정 개인을 추정할 수 있게 되는데 수많은 정보가 모여지는 빅데이터의 특성상 개인의 프라이버시와 개인정보자기결정권이 침해될 수 밖에 없다. 실제로 2018년 서울시 민생사법경찰단은 불법대부, 다단계 등 민생범죄 수사에 AI 범죄예측 시스템을 이용한 ‘인공지능 수사관’을 활용해 수사를 진행하는 것이 개인정보보호법을 위반했다는 개인정보보호위원회의 결정이 나오면서 수사에 제동이 걸렸다.²⁷⁾

26) 윤해성 등, 앞의 글, 141-146면.

27) 서울시 민생사법경찰단(민사단)의 ‘인공지능을 활용한 민생범죄 수사지원 분석사업’(인공지능 수사관)이 개인정보보호법을 위반했다는 개인정보보호위원회(개보위)의 결정이 나왔다. 빅데이터·인공지능(AI) 등 새로운 기술을 범죄수사 등에 사용할 경우 정보인권을 고려한 제도적 환경이 조성돼야

헌법상 평등권에 대한 침해도 우려된다. 즉, AI 범죄예측 시스템으로 부패범죄의 조사대상이 된 특정인과 그렇지 않은 일반인 사이에는 형평성의 문제가 발생한다. 정보의 주체인 국민에 대해서 기본권 보호가 균등하게 되어야 하는데 부패조사 대상자는 그렇지 않은 일반인과 달리 기본권이 더 제한된다는 것이다. 또한 AI 범죄예측 시스템이 과거의 부패범죄와 연관된 사건이나 전과 등의 데이터를 바탕으로 앞으로 발생할 범죄를 예측하기 때문에 부패범죄경력이 과거에 있었던 특정인을 일반인보다 더 집중적으로 감시 및 통제하게 된다. 이러한 감시 및 통제는 대상자를 잠재적인 범죄자로 낙인을 찍는 효과를 유발하게 되는데 결과적으로 대상자를 일반인과 차별하게 되는 문제가 발생한다. 또한 조사 대상자의 범위를 어디까지로 설정할지도 문제가 된다. 만약 대상자를 부패범죄의 피의자, 피고인, 수형자 등으로 한정할 경우 형의집행및수용자처우에관한법률과 개인정보보호법 등에 규정이 있는지도 문제가 될 것이다.²⁸⁾

과도하게 AI 범죄예측 시스템을 사용하여 부패범죄의 예방을 시도하는 것은 일반인의 행복추구권과 행동 자유권을 침해하게 될 것이다. 따라서 AI 범죄예측 시스템의 사용에 의한 부작용도 고려하여야 한다. 예를 들어 AI 범죄예측 시스템이 특정인 갑이 12시간 이내에 부패범죄를 저지를 것이라는 예측결과를 도출했다고 하자. 그래서 담당자가 적절한 감시와 통제에 나섰고 그 결과 부패범죄가 방지되었다면, 그것이 그 담당자의 감시와 통제로 인하여 부패범죄가 예방된 것인지 아니면 애초부터 AI 범죄예측 시스템의 예측 자체가 틀린 것인지 누구도 알 수가 없다. 실제로 오류가 발생한 때도 오류가 발생하였다는 사실을 알 수 없으므로 오류를 바로잡으려는 기회조차 가질 수 없다는데 더 문제가 있다. 오류를 검증할 수 있는 유일한 방법은 위와 같은 사안에서 담당자가 감시와 통제에 나서지 않는 것이다. 만약 12시간 이내에 갑이 정말로 부패범죄를 저지른다면 AI 범죄예측 시스템이 올바르게 예측한 것이며, 그렇지 않다면 적어도 당해 예측에 대해서는 AI 범죄예측 시스템이 틀린 것이다. 그러나 이와 같은 무작위성에 의존하여 범집행을 할 수는 없다. AI 범죄예측 시스템이 부패범죄 발생을 예측하였다면 담당자는 반드시 감시와 통제에 나서야 한다. 그런데 AI 범죄예측 시스템의 오류율이 1%라고 하더라도 100회의 감시와 통제 시도 중에 1회는 아무런 잘못이 없는 사람에게 불필요한 감시와 통제를 한 결과가 될 것이다. 그런데 그것이 옳지 않은 감시와 통제였다는 것은 알 수 없으므로 억울하게 부당한 대우를 받은 사람은 사후에 구제를 받을 가능성도 희박하다.²⁹⁾ 범죄예측의 정확성이 보장되지 않으면 부정확한 범죄예측을 근거한 예방책은 유해하거나 무익하다. 아직 벌어지지 않은 사건을 다루는 AI 범죄예측 시스템은 고도의 정확성이 보장되어야 한다.³⁰⁾ 부정확한 범죄예측으로 개인을 감시하고

한다는 지적이 나온다. 한겨레, “AI 수사관, 개인정보 수집은 위법”...SNS 불심검문 제동 걸려”, 2019. 6.11. <http://www.hani.co.kr/arti/economy/it/897375.html>

28) 윤해성 등, 앞의 글, 141-146면.

29) 윤영석, 앞의 글, 169면.

30) Mark C. Niles, “Preempting Justice : “Precrime” in fiction and in fact”, Seattle journal for social

통제한다면 개인의 자율성과 인권을 크게 침해하게 될 것이다.³¹⁾

3. 알고리즘의 불투명성, 편향성 및 부정확성

1) 알고리즘의 불투명성

인공지능에는 주어진 정보(data)를 자동적으로 분석하고 처리하는 과정이 있는데 이것을 인공지능의 알고리즘이라고 한다.³²⁾ 그런데 이러한 알고리즘이 정보를 입력받아 결론을 도출되는 과정이 불투명하여 충분히 공개되거나 설명되지 않는다는 문제점이 알고리즘의 불투명성이다.³³⁾ 이러한 알고리즘의 불투명성은 AI 범죄예측 시스템의 대표적인 문제점 중의 하나이다. 왜냐하면 AI 범죄예측 시스템으로 인하여 부정확한 범죄예측이 발생하게 되면 왜 잘못된 범죄예측이 발생했는지 원인을 규명하여야 하는데 인공지능의 처리과정 자체가 공개되지 않거나 설명되지 않아서 원인의 규명은 물론 피해나 그 책임의 소재를 찾는 데 어려움을 준다는 것이다.³⁴⁾ 또한 알고리즘의 불투명성은 피고인과 변호인이 알고리즘의 오류를 검토하고 주장하기 어렵게 만들어 결과적으로는 피고인의 방어권을 침해할 수 있다.

알고리즘이 불투명한 원인은 소스코드와 같은 알고리즘의 세부 정보가 공개되더라도 전문적 지식이 없는 일반인은 이를 읽고 이해하는 것은 매우 어렵다는 점이다. 또한 알고리즘을 개발하거나 설계한 전문가들조차도 인공지능의 작동원리나 내부구조를 정확하게 파악할 수 없다는 알고리즘의 특성을 ‘블랙박스(black box)’라고 하는데 이것은 알고리즘의 불투명성의 원인으로 설명하고 있다.³⁵⁾ 그리고 알고리즘 자체가 특정 개인이나 기업의 소유물로서 영업비밀 등으로 보호되고 있다. 영업비밀과 같이 개인이나 기업이 보유한 기술상 또는 경영상 정보는 기업의 경쟁력을 좌우하는 중요한 요소이기 때문에 명확한 공개를 요구하기가 쉽지 않다.³⁶⁾

2) 알고리즘의 편향성

범죄예측의 결과가 특정 계층이나 소수집단에 불리하도록 편향되게 산출되는데 이 때문

justice Vol. 9 Iss. 1, 2010. 1., p. 303.

31) Jennifer C. Daskal, "Pre-crime restraints: the explosion of targeted, noncustodial prevention", Cornell law review Vol. 99 Iss. 2, 2014. 1, p. 332.

32) 크리스토퍼 스타이너/박지유(역), 『알고리즘으로 세상을 지배하라』, 에이콘(2016). 14면.

33) Andrea Roth, Trial by Machine, 104 Geo. L. J. 1245 (2016), 제1269쪽.

34) 양종모, "인공지능 알고리즘의 편향성, 불투명성이 법적 의사결정에 미치는 영향 및 규율 방안", 법조 제723권, 법조협회, 2017. 80면.

35) 양종모, 앞의 글, 81면, 93면.

36) State v. Loomis, 881 N. W. 2d 749 (Wis. 2016).

에 범죄예측이 공정하지 못하다는 문제가 알고리즘의 편향성이다. 예를 들어 대표적인 화이트 칼라범죄인 고위 공직자의 부패범죄는 잘 드러나지 않지만 비교적 하위 공직자들의 부패범죄는 고위 공직자들의 부패범죄에 비해 잘 적발된다면 이러한 데이터가 축적이 되고 AI 범죄예측 시스템이 이를 학습하여 적용하게 되면 하위 공직자들이 편중된 감시와 수사가 집중될 것이고 더 많은 하위 공직자의 부패범죄가 적발되어 처벌되게 된다. 결국 알고리즘의 편향성의 문제는 사법절차에서의 형평성의 문제를 야기한다. 이러한 알고리즘의 편향성은 편향된 데이터의 문제일 수도 있지만, 알고리즘의 개발자가 의도적으로 알고리즘을 편향적으로 설계할 때도 발생할 수 있다. 또한 개발자가 지닌 편향된 지식과 경험이 무의식중에 알고리즘에 그대로 투영되어 발생할 수도 있다.³⁷⁾ 그러나 알고리즘의 편향성의 가장 주요한 원인은 편향된 데이터이다. 알고리즘이 학습하는 데이터들 중에는 인종, 성별 등 차별적인 요소들이 많은데 그 결과 편향적인 판단을 하게 되는 것이다. 예를 들어 과거 유죄판결과 같이 범죄사건을 토대로 한 데이터가 많은데 이러한 데이터를 기초로 미래의 범죄를 예측하다 보니 과거 범죄사건과 관련된 지역이나 사람은 그렇지 않은 지역이나 사람보다 더 집중적인 감시와 통제를 받아 우범지역이나 범죄자로 낙인찍히게 된다는 것이다. 이러한 편향성의 대표적인 사례로 Northpoint가 개발한 교정 범죄자 프로파일링 관리시스템인 컴파스(COMPAS, Correctional Offender Management Profiling for Alternative Sanctions)는 범죄자 재범률을 계산하여 재범의 위험성을 예측한 결과 흑인 피고인을 고위험군으로 잘못 분류하는 편향이 2배나 되는 오류가 나타났다. 즉 흑인이 강력범죄를 저지를 위험성이 백인보다 77% 더 높게 측정하여 인종차별적이란 비판을 받았다.³⁸⁾

3) 알고리즘의 부정확성

AI 범죄예측 시스템은 축적된 데이터를 바탕으로 장래 범죄의 발생을 예측하게 되는데, 이러한 범죄예측은 범죄 발생 가능성이 높다는 확률을 근거로 하는 것이다. 따라서 100% 정확하게 예측하는 것은 현재로서는 불가능하다. 결과적으로 부정확한 예측이 발생할 수 있다. 이러한 알고리즘의 부정확성 또한 문제가 된다. 예를 들어 부정확한 예측으로 인하여 부패범죄를 저지를 것으로 예측하고 예방조치를 취했지만 실제로는 부패범죄를 저지르지 않은 잘못된 긍정 오류(false positive)의 경우 무고한 특정인을 ‘부패’혐의자로 만들게 되며, 부당하게 부패 혐의를 쓴 특정인은 사회적 비난과 감시 등으로 인권이 침해되는 대가를 치르게 된다. 반대로 더 이상 부패범죄를 저지르지 않을 것으로 예측된 특정인에게 예방조치 등을 하지 않았지만 그 특정인이 실제로 부패범죄를 저지르게 된 잘못된 부정 오류(false

37) 양종모, 앞의 논문, 74-77면.

38) Amy B. Cyphert, Reprogramming Recidivism: The First Step Act and Algorithmic Prediction of Rick, 51 Seton HALL L. REV. 331(2020), p.339.

negative)의 경우 부패범죄를 방치하는 결과가 되는데, 이 역시 사회방위에 위협이 되고 범죄로 인한 피해로 인해 사회전체의 비용이 증가하게 된다.

결국 인공지능 알고리즘의 불투명성, 편향성 및 부정확성이 해결되지 아니하면 예측의 결과 및 판단의 신뢰성을 담보하기는 어렵다. 이는 최종적으로 AI 범죄예측 시스템에 대한 불신을 넘어 범집행을 포함한 사법절차 전체에 불복을 초래한다.³⁹⁾

4. 새로운 통제수단으로서 AI 범죄예측 시스템

중국은 전국에 설치된 CCTV로 수집된 영상 데이터를 분석하고 AI 안면인식기술을 활용하여 범죄위험을 예측하는 스카이넷(SkyNet) 시스템을 사용하고 있다. 추적 대상자의 신원과 위치뿐만 아니라 탑승했던 차량, 최근 만난 사람, 이동한 동선 등을 자동으로 추적하고 분석하는 기능도 가지고 있다. 이에 대하여 개인의 사생활까지 감시하는 것은 과도한 통제에 의한 인권침해가 아니냐는 비판을 받고 있다.⁴⁰⁾ 이처럼 AI 범죄예측 시스템이 개인을 감시하고 통제하는 사회통제 시스템으로 둔갑하면 얼마든지 ‘빅브라더’의 출현은 얼마든지 가능하게 된다. 공산당과 같은 소수의 권력층이 정보를 독점한 사회에서는 AI 범죄예측 시스템은 고위험군으로 분류된 사람들에 대한 감시와 감독 등을 정당화하는 새로운 통제수단으로 등극할 수 있다.

AI 범죄예측 시스템은 발생하지도 않은 사건에 대해서도 범죄예방이라는 명목으로 구속과 처벌을 가능하게 할 수 있다. 독재자나 권력자들은 AI 범죄예측 시스템의 설계자나 개발자를 조종하여 빅데이터를 가지고 편향된 알고리즘을 만들 수 있다. 독재자나 권력자들은 이것을 가지고 자신들의 반대세력이나 정적들을 고위험군의 범죄 용의자 그룹으로 설정해 놓고 감시와 통제를 손쉽게 하려고 할 것이다. 그러나 감시나 통제대상이 된 일반인들은 블랙박스과 같은 알고리즘의 불투명성 때문에 그러한 사실을 알지도 못할 것이다.⁴¹⁾

39) 윤해성 등, 앞의 글, 254, 255쪽.

40) Horace White, “Is China’s Skynet detrimental to the entire human race?”, Modern Tokyo Times, 2019년 1월 16일 수정, 2019년 12월 30일 접속,
<http://moderntokyotimes.com/is-chinas-skynet-detrimental-to-the-entire-human-race/> 참조

41) 유강하, “빅데이터와 사물인터넷 시대의 비판적 해석과 인문학적 상상력-영화<마이너리티 리포트>를 중심으로”, 시민인문학 제30호, 2016, 100면 이하.

Ⅳ. 부패범죄에 대한 효율적인 통제방안으로서 개선방안

1. AI 범죄예측 시스템의 대상으로서 부패범죄와 특징인

어느 범죄에 AI 범죄예측 시스템을 우선적으로 사용하느냐는 국가자원의 효율적인 배분의 차원에서 형평성의 문제를 야기한다. 살인범죄나 테러범죄 만큼 부패범죄도 AI 범죄예측 시스템을 사용하여 시급히 척결해야 할 범죄이므로 부패범죄에 AI 범죄예측 시스템을 사용하더라도 형평성의 문제는 발생하지 않는다. 왜냐하면 부패범죄는 한 행위자의 일탈에 그치는 것이 아니라 사회구조를 무너뜨릴 수 있는 파괴력을 지니는 범죄로서 한 국가의 문제만이 아니라 전 세계적으로도 심각한 문제이기 때문이다. 또한 부패범죄가 가지고 있는 은밀성과 관행성 때문에 쉽게 적발되지도 않을 뿐만 아니라 완전히 근절되기도 어렵기 때문에 AI 범죄예측 시스템을 적용하여 부패범죄를 줄일 수 있는 효과적인 방안을 모색할 필요가 있다.

부패범죄에는 수뢰 및 사전수뢰(형법 129조), 제3자 뇌물제공(형법 130조), 수뢰후부정처사 및 사후수뢰(형법 131조), 알선수뢰(형법 132조), 뇌물공여(형법 133조), 특정범죄가중처벌등에관한법률상 뇌물, 알선수재(동법 2조, 3조, 4조), 국제상거래에 있어서 외국공무원에 대한 뇌물방지법위반죄, 배임수재(형법 357조), 배임중재(형법 357조), 금융기관 임직원의 수재 등(특정경제범죄가중처벌등에관한법률 5조, 6조, 7조, 8조, 9조), 사법기관에 대한 청탁 명목 금품 수수행위(변호사법 110조), 공무원에 대한 청탁 명목 금품 수수행위(변호사법 111조) 등이 있다.⁴²⁾

AI 범죄예측 시스템의 대상이 될 특징인은 부패범죄를 저지른 경력이 있는 범죄자나 우범자가 해당될 것이다. 이들에 대한 범죄기록, 재판기록, 판결문, 수사기록 등을 AI로 분석하여 범죄 발생을 예측하게 될 것인데, 보통은 부패범죄를 저지른 경력이 있는 범죄자는 부패범죄의 유혹이 많은 직책에서 배제될 가능성이 높기 때문에 이들에 대한 재범예측은 의미가 없을 수도 있다. 하지만 부패범죄와 관련된 부서(공무원 또는 중재인의 근무하는 부서 등)이나 직무(또는 지위) 등이 예측될 수가 있다는 점에서 범죄예측의 의미가 부여될 수도 있다.

2. 개인의 privacy와 개인정보보호 방안

개인정보를 다루는 AI 범죄예측 시스템은 개인의 권리를 침해할 수 있다. 그렇다고 대상

42) 부패범죄 전담재판부 설치 등에 관한 예규(재형 2003-14) 개정 2009. 8. 28. [재판예규 제1281호, 시행 2009. 9. 1.]

이 되는 개인정보의 수집에 대하여 일일이 개인의 동의를 받거나 법원의 영장을 받는 것은 현실적으로 불가능하다. 이를 개선하기 위해서 2013년 프라이버시에 관한 최대회의인 ‘제35회 데이터보호 프라이버시 커미셔너 국제회의’에서 채택한 ‘프로파일링에 대한 결의’를 검토할 필요가 있다.⁴³⁾ 여기서는 특정인의 특징을 분석하여 미래의 행동을 예측하는 프로파일링(profiling)의 실시조건을 6가지를 열거하고 있다. 즉 첫째, 프로파일링을 실시하기 전에 프로파일링의 필요성과 실용성을 점검해야 한다. 둘째, 프로파일링에 의한 정보의 수집은 합법적 목적을 위해 필요한 한도 내에서 행해져야 한다. 셋째, 프로파일링에 의해 발생한 결과가 긍정적이든 부정적이든 해당 정보주체가 이를 개선하거나 회피할 수 있는 적당한 조치(예를 들어 해당정보를 삭제 또는 수정을 요청할 장치 등)를 마련하여야 한다. 넷째, 정보주체에게 프로파일링이 실시된 것에 대한 정보를 최대한 제공하여야 한다. 다섯째, 프로파일링이 특히 개인에 대한 중대한 법적 효과를 가지거나 이익이나 지위에 영향을 미치는 결과를 야기하는 경우, 개인 정보에 대해서는 정보주체에게 액세스권 및 정정권을 보장하여야 한다. 여섯째, 모든 프로파일링은 적절한 감독을 받아야 한다.⁴⁴⁾ 이러한 6가지 조건들에 부합하도록 개인정보보호법이 개정되었다. 개인정보보호법 제16조 1항에 의하여 “개인정보를 수집하는 경우에는 그 목적에 필요한 최소한의 개인정보를 수집하여야 한다”고 규정하여 개인정보의 수집을 필요 최소한으로 제한하고 있다. 개인정보를 수집 및 이용하는 경우에는 정보주체의 사전 동의를 받도록 하고(제15조 제1항 1호) 다만, 정보주체 또는 그 법정대리인이 의사표시를 할 수 없는 상태에 있거나 주소불명 등으로 사전 동의를 받을 수 없는 경우로서 명백히 정보주체 또는 제3자의 급박한 생명, 신체, 재산의 이익을 위하여 필요하다고 인정되는 등의 경우에는 정보주체의 사전 동의를 받지 않는 예외를 인정하고 있다(동법 제15조 1항 5호).⁴⁵⁾ 정보주체는 개인정보처리자에 대하여 자신의 개인정보 처리의 정지를 요구할 수 있고(동법 제37조 1항), 자신의 개인정보를 열람한 정보주체는 개인정보처

43) 35th International Conference of Data Protection and Privacy Commissioners, Resolution on Profiling, adopted on 24 Sep. 2013.

(출처; https://privacyconference2013.org/Resolutions_and_Declarations).

44) 차상욱, 빅 데이터(Big Data) 환경과 프라이버시 보호, IT와 법 연구 제8집, 2014.

45) 제15조(개인정보의 수집·이용) ① 개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 개인정보를 수집할 수 있으며 그 수집 목적의 범위에서 이용할 수 있다.

1. 정보주체의 동의를 받은 경우
2. 법률에 특별한 규정이 있거나 법령상 의무를 준수하기 위하여 불가피한 경우
3. 공공기관이 법령 등에서 정하는 소관 업무의 수행을 위하여 불가피한 경우
4. 정보주체와의 계약의 체결 및 이행을 위하여 불가피하게 필요한 경우
5. 정보주체 또는 그 법정대리인이 의사표시를 할 수 없는 상태에 있거나 주소불명 등으로 사전 동의를 받을 수 없는 경우로서 명백히 정보주체 또는 제3자의 급박한 생명, 신체, 재산의 이익을 위하여 필요하다고 인정되는 경우
6. 개인정보처리자의 정당한 이익을 달성하기 위하여 필요한 경우로서 명백하게 정보주체의 권리보다 우선하는 경우. 이 경우 개인정보처리자의 정당한 이익과 상당한 관련이 있고 합리적인 범위를 초과하지 아니하는 경우에 한한다.

리자에게 그 개인정보의 정정 또는 삭제를 요구할 수 있고(동법 제36조 1항), 정보주체는 개인정보처리자가 개인정보보호법을 위반한 행위로 손해를 입으면 개인정보처리자에게 손해배상을 청구할 수 있도록 함으로써(동법 제39조 1항) 프로파일링에 의해 발생한 결과를 정보주체가 이를 개선하거나 회피할 수 있는 적당한 조치를 마련하였다.

3. 알고리즘의 문제 해결방안

1) 외부 감사기관의 설치와 설명 가능한 알고리즘

알고리즘의 처리과정을 개발자나 설계자도 정확하게 알지 못하도록 블랙박스처럼 되어 있다는 알고리즘의 불투명성을 해결하는 방법으로는 설명 가능한 알고리즘을 개발하는 방법이 있다. 2017년 발표된 아실로마 23원칙⁴⁶⁾이나 유럽연합(EU)의 인공지능윤리 가이드라인 그리고 OECD의 AI 권고안(OECD Recommendation of the Council on Artificial intelligence, 이하 AI원칙)⁴⁷⁾에서도 알고리즘의 설계자나 알고리즘을 사용하는 기업은 그 알고리즘 내부 논리에 대해 감사기관 등에게 설명할 수 있어야 한다는 설명가능성을 명시하고 있다. 이러한 설명가능성을 확보하기 위해서는 사전적·사후적 단계에서 공적인 기관에 의하여 알고리즘에 대한 감사가 필요하다.⁴⁸⁾ 따라서 알고리즘의 주체나 피해자를 대신하여 알고리즘의 적절성을 감시하고 심사하는 객관적인 기관이나 조직을 설치하여야 할 것이다.⁴⁹⁾

2) Daubert 기준

알고리즘의 부정확성 해결방안으로 미국의 과학수사기법의 신뢰성에 관한 Daubert 기준을 사용한 사례가 있다. AI 머신러닝을 사용한 시스템이 오류가 발생하자 그 오류를 수정하는 과정을 반복함으로써 안면인식 시스템의 정확성을 높이고 신뢰성이라는 Daubert 기준을 통과하였다.⁵⁰⁾ Daubert 기준처럼 인공지능의 오류를 반복적으로 수정하는 과정을 통해서 알고리즘의 부정확성의 문제를 해결할 수가 있을 것이다. AI 범죄예측시스템의 오류를 반

46) 한국인공지능법학회, 인공지능과 법, 박영사, 2020. 66면

47) OECD. Recommendation of the Council on Artificial Intelligence. C/MIN(2019)3/FINAL

48) Michael Veale, Max V. Kleek, and Reuben Binns, Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making, CHI 2018 Paper, Montréal, QC, Canada, 2018. p.1.

49) Tal Z. Zarsky, Transparent Predictions, 2013 U. Ill. L. Rev. 1503, 1515 (2013). at 1558-1562, 1563-1564, 1566.

50) Nawara, John., Machine Learning: Face Recognition Technology Evidence in Criminal Trials, 49 U. Louisville L. Rev. 601(2011). 604, 2011.

복적인 수정과정을 통해서 범죄예측의 정확성을 높여 범죄예방의 효과성이 입증된다면 알고리즘의 부정확성의 문제를 해결할 수가 있을 것이다. 이를 위해서는 오류가 발생하지 않도록 내부적인 통제절차 및 정확성에 대한 인증절차를 마련하는 방안이 필요하다.⁵¹⁾

3) 알고리즘의 책무성과 투명성

알고리즘의 편향성의 문제를 해결하기 위해서는 어떤 데이터가 편향된 데이터인지 식별하는 방법과 이러한 편향된 데이터를 학습하여도 편향된 범죄예측을 하지 않는 알고리즘을 개발하는 방법이 필요하다. 하지만 현재로서는 편향된 데이터를 배제하는 방법이나 데이터의 편향성을 극복할 수 있는 알고리즘의 개발은 진행 중이고 다만 알고리즘의 편향성이 얼마나 위험한지를 평가하는 방법이 개발되어 활용되고 있다. 즉 알고리즘에 설계자나 개발자가 의도적으로 또는 무의식적으로 편향성을 알고리즘에 투영하였는지를 알고리즘의 작동방식을 통해서 잠재적인 편향성을 밝히는 방법이다. 이러한 방식은 코넬대학교 연구진과 산업체 연구진이 공동으로 개발하였는데, 먼저 문제된 알고리즘을 모방해서 모델을 만들고 그 모델에 초기 데이터를 투입하여 그 모델의 편향성이 위험한지를 점수로 표시한다. 그리고 나서 초기 데이터를 투입하여 학습한 이후의 모델을 만들어 그 모델의 편향성이 얼마나 위험한지 점수를 내고 이를 이전 모델과 비교한다. 이를 통해서 기초 데이터 중 어떤 것이 최종 점수에서 영향을 주며 어떤 변수가 중요한지를 추정해 볼 수 있다. 이를 통하여 편향성에 영향을 주는 요인들을 발견하고 그에 대한 조치를 취할 수 있게 된다.⁵²⁾

알고리즘의 편향성을 극복하기 위해서는 궁극적으로는 알고리즘의 설계자나 개발자가 알고리즘을 편향되게 설계하지 않도록 하여야 한다(알고리즘 책무성; algorithmic accountability). 이를 위해 설계과정이나 개발과정에서 알고리즘의 요소 및 매트릭스(metrics)를 투명하게 공개하여야 한다(알고리즘의 투명성; algorithmic transparency). 물론 공개가 필요한 알고리즘의 요소 및 매트릭스가 무엇인지는 합의를 통하여 공개할 대상에 대한 표준화가 필요할 것이다. 또한 알고리즘의 투명성이 확보되지 위해서는 신뢰할 수 있는 외부 기관에 의한 감사가 이루어져야 할 것이다.⁵³⁾

51) 황경조/조재웅/서창호, “상호의존정보 기반 공정성을 보장하는 기계학습”, 2019년도 대한전자공학회 추계학술대회 논문집, 2019.11, 667-671면.

52) Tan, S., Caruana, R., Hooker, C., & Lou, Y., T., Detecting bias in black-box models using transparent model distillation arXiv preprint arXiv: 1710.0616, 2017.

53) 주현경/정재연, “범죄예측 및 형사사법절차에서 알고리즘 편향성 문제와 인공지능의 활용을 위한 규범 설계”, 법학논총 제27권 제1호, 2020, 134면; Nicholas Diakopoulos, Algorithmic Accountability Reporting: On the Investigation of Black Boxes, Two Center for Digital Journalism - A Tow/Knight Brief, Columbia Journalism School, 2013. p.28.

4. 부패범죄 유발인자의 발견과 부패범죄 유발인자의 수치화·계량화

AI 범죄예측 시스템을 활용하면 수많은 형사사법정보에서 부패범죄의 성립에 영향을 주는 유의미한 인자를 발견할 수 있다. 이를 통해 부패범죄의 발생에 상관성이 있는 ‘부패범죄 유발인자’를 발견하게 된다.⁵⁴⁾ 그리고 어떤 부패범죄 유발인자가 부패범죄의 발생에 상관성이 있는지를 수치화한다면 부패범죄 발생에 개별 부패범죄 유발인자간의 우열이 발생할 것이고 이는 결국 부패범죄 유발인자의 계량화·수치화로 발전할 수 있을 것이다. 부패범죄 유발인자들 사이의 우열과 이를 계량화 및 수치화함으로써 부패범죄의 재범 위험성을 예측할 수 있다. 각 부패범죄 유발인자가 부패범죄의 발생 가능성에 어떠한 영향을 미치는지를 통계적으로 산출하여 예측의 정확성을 보다 더 높일 수 있다. 이를 통하여 발견한 부패범죄의 경향성 및 부패범죄 유발인자들은 부패범죄예방과 부패범죄 대응전략의 근거자료로 활용할 수 있을 것이다. 그리고 예방적 개입이 필요한 부분에도 부패관련 인자들의 계량화·수치화가 판단기준을 제시하는데 도움이 될 것이다.

5. 범죄예측 시스템의 개인통제기능의 제한

AI 범죄예측 시스템은 감시 및 통제장치로서 악용될 우려가 있다. 이러한 AI 범죄예측 시스템의 개인통제기능을 제한하기 위해서는 AI 범죄예측 시스템도 대상자의 권리침해를 최소화하기 위하여 비례성의 원칙을 적용하여야 한다(최소침해의 원칙). 그리고 AI 범죄예측 시스템은 최후수단으로서 사용되어야 한다(보충성의 원칙). 대상자의 재범위험성에 따라 개별적으로 사용하여야 한다(개별성의 원칙). AI 범죄예측시스템은 행정처분에 의하여 행해져서는 안 되고 법원의 사법심사에 의하여 부과되어야 한다(사법통제의 원칙). 이와 같은 보안처분의 부과원칙에 따라 AI 범죄예측 시스템의 분석을 적용한다면 AI 범죄예측 시스템 개인통제기능을 최대한 제한할 수 있고 부패범죄의 통제방안으로서 효율성을 높일 수 있게 될 것이다.

한편 AI 범죄예측 시스템의 이용을 비례성의 원칙에 의하여 제한하더라도 그 대상이 되는 사람들은 감시와 통제를 받는다는 사회적 거부감과 공식적인 부패조사에 대한 반감이 필연적으로 존재할 수밖에 없다. AI 범죄예측 시스템의 이용을 위한 부패조사를 거부감 없이 수용하기 위해서는 성숙한 시민의식의 제고와 함께 부패범죄가 없는 세상을 위한 사회적 공감대가 형성되는 것이 필요할 것이다.

54) 이 때 ‘상관성’(correlation)이란 “두 데이터 값 사이의 통계적 관련성을 수량화 한 것”을 말한다. 빅토르 마이어 쾰베르거·케네스 쿠키어, 이지연(옮김), 빅데이터가 만드는 세상, 21세기북스, 2013, 102면.

V. 나가며

부패범죄는 한 행위자의 일탈에 그치는 것이 아니라 사회구조를 무너뜨릴 수 있는 파괴력을 지니는 범죄이기에 한 국가의 문제만이 아니라 전 세계적으로도 심각한 문제로 여겨지고 있다. 이러한 부패범죄를 감소시키는데 국가와 사회가 많은 노력을 기울이고 있지만, 부패범죄가 가지고 있는 은밀성과 관행성 때문에 쉽게 적발되지도 않을 뿐만 아니라 완전히 근절되기도 어렵다. 이에 AI 범죄예측 시스템을 도입하여 부패범죄를 줄일 수 있는 효과적인 방안이 없을지 모색할 필요가 있다. 이미 많은 나라에서 AI 범죄예측 시스템을 도입하여 범죄율이 감소하는 긍정적인 효과를 보고 있기 때문이다. 그런데 AI 범죄예측 시스템을 부패범죄에 적용할 수 있을지 그리고 적용한다면 어떠한 문제점이 있는지를 검토할 필요가 있다. 나아가 부패범죄에 AI 범죄예측 시스템을 적용하기 위한 개선방안이 무엇인지 모색하고자 한다.

부패범죄가 가지고 있는 국가와 사회에 대한 심각성과 은밀성과 관행성 때문에 AI 범죄예측 시스템을 적용하여 부패범죄를 줄일 수 있는 효과적인 방안을 모색할 필요가 있다. AI 범죄예측 시스템은 필연적으로 개인의 권리를 침해할 수 있으므로 개인정보보호법에 의하여 개인정보를 보호하여야 한다. 알고리즘의 불투명성을 해결하기 위해서는 알고리즘의 설계자나 알고리즘을 사용하는 기업은 그 알고리즘 내부 논리에 대해 감사기관 등에게 설명할 수 있어야 한다. 알고리즘의 부정확성 해결방안으로 미국의 과학수사기법의 신뢰성에 관한 Daubert 기준을 사용할 필요가 있다. 알고리즘의 편향성을 극복하기 위해서는 궁극적으로는 알고리즘의 설계자나 개발자가 알고리즘을 편향되게 설계하지 않도록 하여야 한다. 이를 위해 설계과정이나 개발과정에서 알고리즘의 요소 및 매트릭스(metrics)를 투명하게 공개하여야 한다. AI 범죄예측 시스템의 개인통제기능을 제한하기 위해서는 AI 범죄예측 시스템도 대상자의 권리침해를 최소화하기 위하여 비례성의 원칙을 적용하여야 한다. AI 범죄예측 시스템의 이용을 비례성의 원칙에 의하여 제한하더라도 그 대상이 되는 사람들은 감시와 통제를 받는다는 사회적 거부감과 공식적인 부패조사에 대한 반감이 필연적으로 존재할 수밖에 없다. AI 범죄예측 시스템의 이용을 위한 부패조사를 거부감 없이 수용하기 위해서는 성숙한 시민의식의 제고와 함께 부패범죄가 없는 세상을 위한 사회적 공감대가 형성되는 것이 필요할 것이다.

참고문헌

- 김병수, “검찰개혁의 필요성과 추진방향”, 한국부패학회보 제25권 2호, 2020.
- 김상락·강만모, “빅데이터 분석 기술의 오늘과 미래”, 정보과학지 제32권 1호, 2014. 1.
- 김태완, “데이터 마이닝 기법을 활용한 청소년 재범 관련요인 연구”, 교정복지연구(제35호), 한국교정복지학회(2014. 12).
- 박민지/서정범, “사물인터넷을 활용한 경찰활동에 관한 연구- 그 현황과 허용성에 관한 논의를 중심으로 -”, 경찰학연구 제19권 제3호, 경찰대학, 2019. 9.
- 박민지/서정범, “전자감독제도의 예방적 경찰활동으로서의 합의-범죄징후 예측시스템의 허용성을 중심으로-, 고려법학 제99호, 2020.
- 빅토르 마이어 쾨베르거·케네스 쿠키어, 이지연(옮김), 빅데이터가 만드는 세상, 21세기북스, 2013.
- 양주아/양문승, “경찰의 과학치안 활동 현황 및 전망”, 한국경찰연구 제16권 제4호, 2017. 12.
- 양종모, “인공지능 알고리즘의 편향성, 불투명성이 법적 의사결정에 미치는 영향 및 규율 방안”, 법조 제723권, 법조협회, 2017.
- 유병후, “빅데이터를 활용한 범죄예방서비스 창출방안”, 한국범죄심리연구 제16권 제3호, 2020.
- 유강하, “빅데이터와 사물인터넷 시대의 비판적 해석과 인문학적 상상력-영화<마이너리티 리포트>를 중심으로”, 시민인문학 제30호, 2016.
- 윤상현/신상화, “범죄 예측의 현재와 미래- 한국형 범죄예측 시스템 개선·도입에 대한 제언 -”, 한국공안행정학회보, 제71호, 2018.
- 윤상연, “범죄예측 기술 어디까지 왔나?,” 치안정책리뷰 제60호, 치안정책연구소, 2018. 6.
- 윤영식, “범죄징후 예측시스템에 대한 비판적 검토-2019년 법무부의 발표내용을 중심으로-”, 과학기술법연구 제26집 1호, 2020.
- 윤해성 등, 범죄 빅데이터를 활용한 범죄예방시스템 구축을 위한 예비 연구(I), 형사정책연구원, 2014.
- 임승욱 외, “범죄 발생환경, 행동패턴 및 심리정보 등 융합정보 적용형 엔트로피 필터링 예측분석기반의 실시간 범죄 예측·예방 시스템 개발,” 정보통신·방송 연구 개발사업 최종보고서, 전자부품연구원, 2016. 10.
- 이수정, “성인 판결전 조사 활용, 재범 위험성 평가도구 개발”, 동향/연구보고서, 법무부(2010).
- 이주희, “고위공직자 부패 방지에 관한 형법적 고찰”, 한양법학 제24권 제3집, 한양법학, 2013.
- 주현경/정채연, “범죄예측 및 형사사법절차에서 알고리즘 편향성 문제와 인공지능의 활용을 위한 규범 설계”, 법학논총 제27권 제1호, 2020.
- 정신교, “뇌물죄의 양형합리화 방안”, 치안정책연구 제23호, 치안정책연구소, 2009.
- 정용찬, 빅데이터, 커뮤니케이션북스, 2013.
- 정재준, “사물인터넷(IoT)의 치안대책 활용 방안”, 경찰학연구 제18권 제1호, 경찰대학, 2018.
- 차상욱, 빅 데이터(Big Data) 환경과 프라이버시 보호, IT와 법 연구 제8집, 2014.
- 한세익, “AI기반 청렴성 증강모델”, 한국부패학회보 제26권 제4호, 2021.
- 한국인공지능법학회, 인공지능과 법, 박영사, 2020.
- 한국전자통신연구원, “위험 상황 초기 인지를 위한 ICT 기반의 범죄 위험도 예측 및 대응 기술 개발,” 한국전자통신연구원 사업계획서, 2018.

한국전자통신연구원, “지능형 치안서비스 기술 동향”, 전자통신동향분석 제34권 제1호, 2019. 2.
 행정안전부/한국정보화진흥원, 월간 디지털 정부 최신 해외 정책·기술 동향 제40호, 2018.
 황경조/조재웅/서창호, “상호의존정보 기반 공정성을 보장하는 기계학습”, 2019년도 대한전자공학회
 추계학술대회 논문집, 2019.11.

- Andrea Roth, Trial by Machine, 104 Geo. L. J. 1245 (2016).
- Amy B. Cyphert, Reprogramming Recivism: The First Step Act and Algorithmic Prediction of Rick, 51 Seton HALL L. REV. 331(2020).
- C. Francescani, “NYPD Expands Surveillance net to Fight Crime as Well as Terrorism,” Reuters, June 21, 2013.
- Harthorne M., City : A Few streets Responsible for Half of seattle’s Crime; komonews.com/archive/city-a-few-streets-responsible-for-half-of-seattles-crime.
- Holmes, S. T., Tewksbury, R., & Holmes, R. M. (1999). Fractured Identity Syndrome. Journal of Contemporary Criminal Justice, 13(3).
- Jeffrey S. Paul and Thomas M. Joiner, Integration of Centralized Intelligence with Geographic Information Systems, Geography and Public Safety, Oct. 2011.
- Jennifer Bachner, Predictive Policing: Preventing Crime with Data and Analytics, Washington, DC: IBM Center for The Business of Government, 2013.
- Nath, S.V., Crime Pattern Detection Using Data Mining, Oracle Corporation, 2006.
- Nawara, John., Machine Learning: Face Recognition Technology Evidence in Criminal Trials, 49 U. Louisville L. Rev. 601(2011). 604, 2011.
- Nicholas Diakopoulos, Algorithmic Accountability Reporting: On the Investigation of Black Boxes, Two Center for Digital Journalism – A Tow/Knight Brief, Columbia Journalism School, 2013.
- Mark C. Niles, “Preempting Justice : “Precrime” in fiction and in fact”, Seattle journal for social justice Vol. 9 Iss. 1, 2010. 1.
- P. Aarvik, Artificial Intelligence—a promising anti-corruption tool in development setting? U4 Report, 2019-1.
- Perry, W. L. (2013), Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations, RAND Corporation.
- Singer, S. D., & Hensley, C. (2004). Applying Social Learning Theory to Childhood and Adolescent Firesetting: Can It Lead to Serial Murder?. International journal of offender therapy and comparative criminology: official organ of the Association for Psychiatric Treatment of Offenders (APTO), 48(4).
- Sherman, L.W., Gartin, P.R., & Buerger, M.E., “Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place,” Criminology, 27, 1989.
- Tan, S., Caruana, R., Hooker, C., & Lou, Y., T., Detecting bias in black-box models using transparent model distillation arXiv preprint arXiv: 1710.0616, 2017.
- Tal Z. Zarsky, Transparent Predictions, 2013 U. Ill. L. Rev. 1503, 1515 (2013).

- Whitman, T. A., & Akutagawa, D. (2004). Riddles in serial murder: A synthesis. *Aggression and Violent Behavior*, 9.
- Wright, J., & Hensley, C. (2003). From Animal Cruelty to Serial Murder: Applying the Graduation Hypothesis. *International Journal of Offender Therapy and Comparative Criminology*: official organ of the Association for Psychiatric Treatment of Offenders (APTO), 47(1).
- Zach Friend, Predictive Policing: Using Technology to Reduce Crime, *FBI Law Enforcement Bulletin*(Feb. 8, 2018).

투고일자 : 2022. 06. 07

수정일자 : 2022. 06. 26

게재일자 : 2022. 06. 30

<국문초록>

부패범죄에 AI 범죄예측 시스템 적용의 문제점과 개선방안

김 병 수

부패범죄는 한 행위자의 일탈에 그치는 것이 아니라 사회구조를 무너뜨릴 수 있는 파괴력을 지니는 범죄이기에 한 국가의 문제만이 아니라 전 세계적으로도 심각한 문제로 여겨지고 있다. 이러한 부패범죄를 감소시키는데 국가와 사회가 많은 노력을 기울이고 있지만, 부패범죄가 가지고 있는 은밀성과 관행성 때문에 쉽게 적발되지도 않을 뿐만 아니라 완전히 근절되기도 어렵다. 이에 AI 범죄예측 시스템을 도입하여 부패범죄를 줄일 수 있는 효과적인 방안이 없을지 모색할 필요가 있다. 이미 많은 나라에서 AI 범죄예측 시스템을 도입하여 범죄율이 감소하는 긍정적인 효과를 보고 있기 때문이다. 그런데 AI 범죄예측 시스템을 부패범죄에 적용할 수 있을지 그리고 적용한다면 어떠한 문제점이 있는지를 검토할 필요가 있다. 나아가 부패범죄에 AI 범죄예측 시스템을 적용하기 위한 개선방안이 무엇인지 모색하고자 한다.

부패범죄가 가지고 있는 국가와 사회에 대한 심각성과 은밀성과 관행성 때문에 AI 범죄예측 시스템을 적용하여 부패범죄를 줄일 수 있는 효과적인 방안을 모색할 필요가 있다. AI 범죄예측 시스템은 필연적으로 개인의 권리를 침해할 수 있으므로 개인정보보호법에 의하여 개인정보를 보호하여야 한다. 알고리즘의 불투명성을 해결하기 위해서는 알고리즘의 설계자나 알고리즘을 사용하는 기업은 그 알고리즘 내부 논리에 대해 감사기관 등에게 설명할 수 있어야 한다. 알고리즘의 부정확성 해결방안으로 미국의 과학수사기법의 신뢰성에 관한 Daubert 기준을 사용할 필요가 있다. 알고리즘의 편향성을 극복하기 위해서는 궁극적으로는 알고리즘의 설계자나 개발자가 알고리즘을 편향되게 설계하지 않도록 하여야 한다. 이를 위해 설계과정이나 개발과정에서 알고리즘의 요소 및 매트릭스(metrics)를 투명하게 공개하여야 한다. AI 범죄예측 시스템의 개인통제기능을 제한하기 위해서는 AI 범죄예측 시스템도 대상자의 권리침해를 최소화하기 위하여 비례성의 원칙을 적용하여야 한다. AI 범죄예측 시스템의 이용을 비례성의 원칙에 의하여 제한하더라도 그 대상이 되는 사람들은 감시와 통제를 받는다는 사회적 거부감과 공식적인 부패조사에 대한 반감이 필연적으로 존재할 수밖에 없다. AI 범죄예측 시스템의 이용을 위한 부패조사를 거부감 없이 수용하기 위

해서는 성숙한 시민의식의 제고와 함께 부패범죄가 없는 세상을 위한 사회적 공감대가 형성되는 것이 필요할 것이다.

주제어: 부패범죄, 범죄예측 시스템, 인공지능, 부패범죄 유발인자, 알고리즘, 범죄예방

