인공지능에 기반한 형사법상 의사결정 연구
- 설명요구권과 영업비밀보호 간 균형모색을 중심으로 -

김혜인* / 정종구**

목 차

Ⅰ. 서론

대한민국 정부는 2017년 11월 4차 산업혁명위원회를 출범시키고, 2019년 10월에는 대통령 'AI 기본구상'을 발표하는 등 다양한 방면에서 인공지능을 활용하려는 의지를 보였다.[1] 또한 최고의 디지털 정부를 구현하기 위해 주요 전자정부 시스템을 진단하고 개선하여 인공지능 기반 디지털 정부로 이행하고, 맞춤형, 지능형 공공서비스를 제공하기 위해 공공부문이 선도적으로 인공지능을 도입하겠다고 선언하였다.[2] 여기에는 맞춤형 문화복지, 특허정보 제공, 환경오염 대응, 고정업무 효율화, 국민생활 안전(범죄 발생 예측 및 대응), 노인복지 강화, SoC 안전 확보 등이 구체적인 목표로 제시되었다. 이와 같은 동향에서 볼 때, 인공지능은 공공의 영역으로 편입될 가능성이 크다. 따라서 인공지능이 공공 의사결정을 담당하였을 때 생길 수 있는 문제와 이에 대한 제도적 준비를 고민해볼 필요가 있다.

해외의 경우 인공지능에 의한 의사결정의 도입이 여러 법적 쟁점 논의로 비화되었다. 미

* 주저자, 서울대학교 법학전문대학원 박사과정
** 공동저자, 서울대학교 법학전문대학원 교육지원실장 (변호사)

1) 관계부처 합동, '인공지능 국가전략', 2019.12., 9쪽.
2) 관계부처 합동, 앞의 책, 36쪽.

국의 경우 켄터키, 위스콘신, 캘리포니아, 펜실베니아 등에서 형사법 영역에 인공지능을 도입했다가 편향성의 문제, 근거조항의 부재에 대해 논란이 생겼다.[3][4] 이처럼 공공서비스에서 인공지능을 도입하려다가 한계에 봉착하여 좌절된 사례는 빈번하게 목도된다.[5] 특히 인공지능을 통한 데이터 주도의 자동화된 의사결정(automated decision-making)이 공적인 영역으로 도입되었을 경우 지니게 될 영향력을 고려할 때, 소위 알고리즘 권력(algorithmic power)을 어떻게 견제하며 조화를 이루어낼 것인가에 대한 논의가 시급하다.[6]

본고에서는 우선 ① 인공지능에 기반한 공공분야의 의사결정이 어떻게 이루어지고 있으며 여기서 제기되는 법적 쟁점이 무엇인지 살펴본다. 다음으로 ② 공공분야 중에서도 실질적으로 인신구속과 직결되어 특히 중요한 형사법에서 범죄예측 및 재범가능성에 대한 의사결정을 토대로 양형에 영향력을 행사했던 인공지능의 예시와 관련판례를 위주로 형사법에서 불거질 수 있는 알고리즘의 편향성 문제를 짚어보고 이러한 편향성에 대해 인공지능을 둘러싼 이해관계자들 어떤 책무성을 부담하는지를 검토한다. 마지막으로 ③ 공적 영역에서 인공지능의 활용을 둘러싼 영업비밀(trade secret) 논의와 의사결정에 의해 영향을 받는 주체의 설명을 요구할 권리(right to explanation)에 대한 논의를 다루고 인공지능의 발전을 저해하지 않으면서도 설명을 요구할 권리가 보장될 수 있는지 살펴본다. 마지막으로 ④ 설명가능한 인공지능(explainable artificial intelligence)을 활용하여 양측의 균형을 도모하기 위한 입법의 가능성을 모색한다.

II. 인공지능에 기반한 공적인 의사결정의 현황 및 쟁점

1. 인공지능을 활용한 공적인 의사결정의 예시 및 현황

알고리즘(Algorithm)이란 어떤 문제를 해결하기 위한 절차(과정)을 기술해놓은 것이

3) Bavitz, Christophre *et al.*,"Assessing the Assessments: Lessons From Early State Experiences in the Procurement and Implementation of Risk Assessment Tools", Berkman Klein Center for Internet & Society Research Publication, 2018, pp. 13-14.

4) Houston Federation of Teachers, Local 2145 v. Houston Independent School District (HISD) 사건처럼 공교육 부문에 사기업의 인공지능인 EVAAS를 도입하여 교사들을 평가하려다가 적법절차(due process) 문제가 제기될 수 있음을 근거로 더 이상 쓰이지 않게 된 사례도 있었다(Houston Federation of Teachers, Local 2415, et al., v. Houston Independent School District, Amended Summary Judgment Opinion, May 4, 2017, https://www.govinfo.gov/content/pkg/USCOURTS-txsd-4_14-cv-01189/pdf/USCOURTS-txsd-4_14-cv-01189-0.pdf; Courthouse News Service, 'Houston Schools Must Face Teacher Evaluation Lawsuit', May 8, 2017, available at https://www.courthousenews.com/houston -schools-must-face-teacher-evaluation-lawsuit/.)

5) Michele Caianiello, Criminal Process faced with the Challenges of Scientific and Technological Development, *European Journal of Crime, Criminal Law and Criminal Justice,* Vol. 27 (2019), pp. 275-276.

6) 주현경/정채연, '범죄예측 및 형사사법절차에서 알고리즘 편향성 문제와 인공지능의 활용을 위한 규범설계', 조선대학교 법학논총, 제27집 제1호, 2020, 117쪽.

며, '입력된 자료를 처리하여 일정한 출력값을 유도하는 규칙, 절차, 과정 등의 모든 내용의 총체' 내지 '과업을 수행하기 위해 확정된 과정 또는 단계별로 형식화된 묶음'이다.[7] 이러한 알고리즘을 컴퓨터가 인식할 수 있는 형태로 표현한 것이 프로그램이며, 인공지능 작동의 기본원리는 기존의 데이터를 알고리즘에 입력하고, 이후 어떠한 과정을 거쳐 결론을 도출하는 것에서 비롯된다.[8]

최근 우리 정부는 향후 인공지능을 행정에 활용하는 방안을 내놓은 바 있어 귀추가 주목된다. 2017년 과학기술정보통신부의 '지능정보사회 중장기종합대책(2017)'은 전장 전력 극대화, 지능형 범죄대응 시스템 구축을 위한 범죄예방 및 검거역량 강화, 지능정보기술을 활용한 행정 및 복지서비스의 구현, 지능정보기술을 활용한 미래형 교통, 유통, 도시 인프라 등을 언급하였다. 행정안전부 역시 '지능형 전자정부 기본계획(2017)'에서 민원빅데이터 분석이 가능한 AI신문고, 기존 정책데이터 분석을 통해 숨은 정책수요를 발굴하고, 선제적으로 대응하는 WISE 정부 등을 구축하겠다는 계획을 세운 바 있다. 알고리즘은 잠재적으로 비용을 절감하고, 자원제약을 극복하며, 인간을 자유롭게 하며, 예측의 정확성을 향상시키고, 많은 작업을 한꺼번에 처리할 수 있다는 장점을 가지기에 공공분야에 점차 적용되고 있는 것으로 보인다.

하지만 활성화되는 적용방법과 향후 활용방안에도 불구하고, 인공지능의 공공분야 도입에 대한 우려 또한 만만치 않다. '법의 지배(rule of law)' 대신 알고리즘이 법을 지지, 수정하거나 무효화하는 '알고리즘의 지배(Rule of Algorithm)'가 올 수도 있다는 염려가 대표적이다.[9] 현존하는 법 규정이 사회적 개념인데 반해, 알고리즘은 인간에 의해 목적과 가치가 입력되어 사회적인 면모는 갖추었지만, 그 작동방식에서 인간이 파악하기 어려운 블랙박스(black box)를 지니는 기술적 개념이다.[10] 또한 알고리즘은 법치국가 원리에 입각한 제3자에 의한 객관적인 심사가 가능한 절차의 산물이 아니기에 민주적 법치국가 원리의 지배에 입각하지는 않는다는 비판을 받게 된다.[11] 더 나아가 알고리즘에 의한 직간접적 행위조종은 동의를 이용하여 법적 보호를 형해화하거나, 민주적 정당성을 저해할 수 있다는 의견도 있다.[12] 즉 인공지능을 행정에 적용하기 위한 조건이나 합법성 조건에 대해서 심도있는 논의를 전개하고 미리 대비해야 한다.

특히 공공분야 중에서도 형사법 영역은 직접적으로 인신을 구속하는 등 인간의 기본권과 직결된 부분과 맞닿아 있기에 우선적으로 고민해볼 주제이다. 간략하게 형사분야의 인공지능 적용례를 살펴보면, 이미 다양한 분야에서 변주가 일어나고 있는 것을 알 수 있

7) 선지원, '인공지능 알고리즘 규율에 대한 소고: 독일의 경험을 중심으로', 경제규제와 법, 제12권 제1호, 통권 제23호, 2019, 27쪽.
8) 양자는 구별되지 않고, 알고리즘으로 불리고 있다. 김중권, "인공지능시대에 알고리즘에 의한 행위조종과 가상적 행정행위에 관한 소고", 공법연구, 제48집 제3호, 2020, 290쪽.
9) 김중권, 앞의 논문, 289쪽.
10) 김중권, 앞의 논문, 295-296쪽.
11) 김중권, 앞의 논문, 298쪽.
12) 김중권, 앞의 논문, 299쪽.

다. 우선 범죄예방과 관련해서는 수사기관이 가진 정보와 외부 공개 정보가 빅데이터가 되어 인공지능의 알고리즘이 막대한 양의 데이터를 분석한 뒤 범죄유형, 범죄시간 및 장소를 도출하고, 머신러닝이나 딥러닝을 통해 범죄예방 및 수사방법을 제시하여 범죄에 대응할 수 있으리라는 기대를 받고 있다.[13] 실제로 서울 서초구는 범죄통계 정보를 이용해 범죄발생 가능성을 예측하는 인공지능 CCTV 기술을 2020년 하반기에 도입할 것으로 발표하였다.[14]

위와 같은 예시에서 유추해보건대 형사법에서 인공지능의 적용은 이미 현실이다. 다만 이러한 새로운 경향이 기존의 체제가 추구하는 여러 가치들과 정합적으로 공존할 수 있는지에 대해서는 좀 더 면밀히 살펴보아야 한다. 이를 위해 우선 인공지능의 적용이 공적인 의사결정에 어떠한 법적쟁점을 유발하는지 살펴본 뒤, 실제 판례가 선고된 범죄예측 및 재범가능성 예측 인공지능을 위주로 분석한다.

2. 인공지능을 활용한 공적 의사결정에 대한 법적 쟁점

공공분야의 의사결정에 대한 인공지능의 개입가능성이 점차 높아짐에 따라, 학계에서는 의사결정 알고리즘에 대한 규범적인 대응방안을 다방면으로 연구하고 있다. 우선 알고리즘이 인간의 행위에 영향을 주는 규제적 역할을 하기 시작했다는 점에 주목하고 있는데, 알고리즘에 의한 행위규제는 공식적인 의회입법 절차나 공적영역에서의 담론을 통해 형성되는 것이 아니라는 점에서 투명성이나 관리가능성 측면에서 불확실성을 내포한다는 점이 지적된다.[15] 또한 알고리즘은 자동화 속성을 통해 인간과 상호작용하는데, 이 과정에서 과거 데이터를 학습하게 되고 그에 따른 편견이나 선입견이 주입되어 편향성(bias)이 생겨날 수 있다.[16] 따라서 적절한 안전장치가 마련되지 않으면, 확증되고 공고화된 편향 (인종, 소득, 직업 등에서 비롯한 차별적 대우 등)이 데이터에 의한 판단으로서 객관적이라고 포장된 채 공적인 의사결정을 좌우할 위험이 발생한다.

하지만 위와 같은 문제가 현재 구현되어 있는 법제에 반영되어 있는지는 회의적이다. 인공지능 관련 법률로 「전자정부법」, 「뇌연구촉진법」, 「지능형 로봇 개발 및 보급 촉진법」, 「소프트웨어산업 진흥법」, 「정보통신 진흥 및 융합활성화에 관한 특별법」 등이 있으나 직접 공적인 의사결정을 함에 있어 인공지능을 활용할 때 논란이 될 수 있는 인공지능의 특수성이나 위험을 미연에 방지할 수 있는 법률은 없는 것으로 보인다.[17] 즉 공적 영역에서 의사결정을 함에 있어 적용되는 알고리즘은 공익과 적법한 절차에 대한 고민이 구

13) 최정일, '빅데이터 분석을 기반으로 하는 첨단과학기법의 현황과 한계: 범죄예방과 수사의 측면에서', 한국법학회 법학연구, 제20권 제1호, 통권 77호, 2020, 65쪽.
14) 고현실, AI로 범죄 예측하는 CCTV 기술, 하반기 서초구에 적용, 연합뉴스, 2020. 1. 2.
15) 박상돈, '헌법상 자동의사결정 알고리즘 설명요구권에 관한 개괄적 고찰', 헌법학연구, 제23권 제3호, 2017, 200-201쪽.
16) 박상돈, 앞의 논문, 201쪽.
17) 김도승, '인공지능 기반 자동행정과 법치주의', 미국헌법연구, 제30권 제1호, 2019, 113쪽.

체화되어야 하는 분야이며, 이를 위해서는 기술적 적법절차(technological due process)에 대한 논의가 선행되어야 한다.[18] 다만 행정과정이 자동화되는 경우는 실정법적인 근거가 부재하므로 만일 공적 의사결정 영역에서 인공지능이 활용되는 경우 기본권제한의 측면이 존재한다면 법률유보 원칙상 법적인 근거마련이 시급하다.[19]

이에 대해 2017. 12. 28.에 제안된 데이터 기반 행정 활성화에 관한 법률안에서는 데이터 기반 행정을 언급하고 있는데, 데이터를 수집·저장·가공·분석 등의 방법으로 정책 수립 및 의사결정에 활용함으로써 객관적으로 과학적으로 수행하는 행정(안 제2조 제2호)을 정의하고 있는 점이 주목된다. 하지만 여전히 정책의사결정 단계에서 데이터 기반 행정의 적용 가능성, 한계, 특수성의 내용은 다루지 못하고 있는 점은 한계로 보인다. 동 법안은 데이터기반행정 활성화 위원회 등의 추진체계, 데이터의 등록 및 제공 절차, 데이터 기반행정 표준화, 데이터 통합관리 플랫폼, 데이터분석센터, 데이터기반행정 전문기관, 데이터기반행정 실태 점검 및 평가를 명시하지만 기존의 전자행정 법제의 일반적 체계를 답습하는 것으로 평가되며, 인공지능 기술의 행정분야 적용에 대한 공법적 정당화 요소를 결여하고 있기 때문이다.[20] 공적인 의사결정 중에서 특히 문제되는 지점은 상술한 바와 같이 형사법의 영역이다. 편파적인 인공지능을 이용한 공적인 의사결정이 무고하거나 가벼운 범죄를 저지른 이에게 가혹한 형벌을 과하게 된다면 국민의 기본적인 권리를 위법하고 부당하게 침해할 수 있기 때문이다. 이는 국제적으로 보편적인 가치인 인권, 그리고 국내법적으로 과잉금지원칙을 위배하여 신체의 자유를 침해할 우려가 있다.

위와 같은 쟁점들과 더불어 가장 첨예하게 논의되고 있으며 실제로 사건화되어 본 연구에서 논하고자 하는 문제는 인공지능의 편향성과 그에 따른 책무성의 이슈이다. 인공지능에게 법적 책임을 물어야 할 경우, 알고리즘에 따른 처리과정이 인간 개입이 부재하거나 어려웠던 상태였다는 점에서 블랙박스(black box) 문제가 제기된다. 특히 법적인 판단을 할 때 결과예측의 정확성이나 효율성을 논외로 하더라도 결과예측에 이르게 된 절차적 정당성 역시 확보해야 한다는 요청이 있는데, 과연 이를 진지하게 수용하고 있는지 확인해볼 필요가 있다. 이와 같은 문제의식을 바탕으로 다음 장에서는 형사법에 있어서 짚고 넘어가야 할 편향성과 책무성의 문제를 다룬 뒤, 설명할 의무를 기본권 차원에서 강제할 수 있는지를 모색한다.

III. 형사법 절차에서 인공지능의 편향성과 책무성

형사법의 절차는 사실인정과 법률적용 및 양형절차로 구성된다. 이때 인공지능은 현재

18) Anne L. Washington, 'How to Argue with an Algorithm: Lessons from the COMPAS-PROPUBLICA Debate', *Colorado Technology Law Journal*, Vol. 17, 2018, p. 136.
19) 김중권, 앞의 논문, 303쪽.
20) 김도승, 앞의 논문, 114쪽.

양형절차에서 주로 그 활용이 고려되고 있다.[21] 가령 사법기관에서 의사결정 지원도구로서 고안된 통계적 예측모델에 기반하여 재범 위험성 예측 알고리즘을 활용함으로써 기존 판결에서 언급된 인자[22]를 고려함과 동시에 모든 관련 정보를 집약해 처리하게 되며, 이로써 형사사법 시스템의 공정성이 높아지고 형사사법 자원을 고위험 범죄자에게 집중할 수 있으리라고 기대된다.[23]

양형의 경우 이론보다는 경험에 의존하는 면이 많고 언어화가 어려운 분야로 뽑히고 있다. 만일 대법원의 양형위원회를 통해 축적된 양형 데이터베이스를 인공지능과 결부시킨다면 양형의 맥락에서 인공지능을 활용할 수 있다.[24] 즉 이 부분이 형사법 절차 중에서 인공지능을 적용할 개연성이 가장 큰 부분일 수 있다. 현재 형사소송법 제70조 제2항은 구속사유를 심사함에 있어 범죄의 중대성, 재범의 위험성, 피해자 및 중요 참고인 등에 대한 위해 우려 등을 고려하도록 하고 있고, 형법 제51조는 형을 정함에 있어 1. 범인의 연령, 성행, 지능과 환경, 2. 피해자에 대한 관계, 3. 범행의 동기, 수단과 결과, 4. 범행 후의 정황 등을 고려하도록 하고 있다.

만약 법원이 여러 공적 데이터, 통계, 기타 피고인에 대한 데이터를 기반으로 범죄예측 모델을 운영하고 여기에 인공지능을 접목시키게 된다면, 현재 운영되고 있는 양형위원회의 역할을 넘어 개인별로 맞춤형의 특화된 분석과 평가를 수행할 수 있게 된다. 다만 이러한 시나리오가 현실화되기 위해서는 인공지능의 편향성과 그에 따른 책무성의 문제가 해소되어야 한다.[25] 우리 헌법은 법치국가원리 뿐만 아니라 평등권과 그에 따른 평등원칙 및 자유로운 의사결정에 따른 자기책임원칙을 규정하고 있기 때문이다(헌법 제10조, 제11조 등 참조).

1. 형사법 절차에서의 알고리즘 편향성의 문제

알고리즘 편향성(algorithm bias)은 근본적인 알고리즘의 본래 속성에서 기인한다. 인공지능은 자동화된 추론을 수행하는 과정에서 통계를 기반으로 작동하기 때문에 인간의 규범적인 개입이나 통계 없이 특정 정보의 추가나 배제를 결정할 수 있다.[26] 이러한 내

21) 정용기/송기복, '인공지능(AI)의 발전과 형사사법의 주요논점, 한국경찰연구 제18권 제2호, 12-16면; 양천수, '인공지능과 법체계의 변화: 형사사법을 예로 하여', 법철학연구, 제20권 제2호 (2017), 57-58쪽.

22) 대법원 2016.8.24. 선고 2016두34929 판결 (종전에 범한 범죄의 종류와 성격, 범정, 형 집행 기간의 행태, 형 집행 이후의 사회적 활동 및 태도, 생활환경, 성행 등), 대법원 2015. 2. 26. 선고 2014도17294판결 (성폭력 범죄의 재범 위험성 유무 판단을 직업과 환경, 당해 범행 이전의 행적, 그 범행의 동기, 수단, 범행 후의 정황 등을 고려하여 한다.)

23) 양종모, '인공지능 알고리즘의 편향성, 불투명성이 법적 의사결정에 미치는 영향 및 규율 방안', 法曹, Vol. 723, 2017, 68-69쪽.

24) 정용기/송기복, 앞의 논문, 17-18쪽.

25) 최정일, 앞의 논문, 71쪽.

26) 주현경/정채연, '범죄예측 및 형사사법절차에서 알고리즘 편향성 문제와 인공지능의 활용을 위한 규범설계', 조선대학교 법학논집, 제27집 제1호, 2020, 120쪽.

재적 위험성 외에도 우선화, 분류, 연관, 필터링 등의 과정을 거치는 인공지능의 알고리즘은 그 과정에서 개발자가 기준설정, 훈련데이터 분류 및 선정 등을 수행하기에 어느 정도 인간의 사고(세계관, 역사관, 윤리관)가 사전에 개입되게 된다.[27]

설령 알고리즘이 매우 중립적일 지라도 트레이닝을 위해 학습하는 데이터 자체에 오류, 왜곡, 편향이 존재할 수 있으므로, 이 경우 데이터 집합체(data set)에서 기인하는 편향이 발생하고 증폭될 수 있다.[28] 또한 딥러닝(deep learning) 기술은 데이터를 스스로 학습하고 패턴을 찾아내기에 프로그래머 역시 인공지능 알고리즘의 판단이 이루어지게 된 추론과정 및 작동방식을 모두 규명하는데는 한계를 가진다. 따라서 인공지능 편향성은 사전에 예방하거나 사후에 원인규명 및 책임귀속을 행하기 어렵다는 특징이 있다.[29] 나아가 인공지능의 편향성은 인간의 편견을 그대로 반영함을 넘어 영속화될 가능성도 있다.[30]

형사법 절차 중에서는 보석(bail), 가석방(parole), 그리고 양형(sentencing) 등에 필요한 재범률 예측에 인공지능이 활용된 적이 있었는데 이때 편향의 문제가 가장 잘 드러났다. 미국의 형사절차에서는 형량을 결정할 때 재범예측 알고리즘(Recidivism Algorithm)을 이용하기도 하는데, 미국의 COMPAS(Correctional Offender Management Profiling for Alternative Sanctions)가 대표적이다. COMPAS는 머신러닝[31]을 이용하여 정부의 데이터베이스에 직접 연결되어 137개의 지표를 분석하고, 이를 통해 범죄자의 재범가능성을 판단한다. 이러한 판단에 따른 결과는 재판 전 구금 및 가석방 심사, 피고인에 대한 법관의 양형에 적용되고 있다.[32] 그러나 COMPAS의 공식에 인종적 편향이 내재되어 있고 심사의 설계 자체에서 흑인 피고인이 백인 피고인보다 높은 재범 가능성을 보이도록 설계되었다는 비판이 제기되었다. 이에 제조사인 Northpointe는 COMPAS가 인종적으로 중립적이며, 모든 집단에게 동일한 비율로 적용되어 정확한 심사를 구현하고 있다고 하였다. 이와 같은 공방 과정에서 알고리즘의 공정성(fairness) 문제가 부각되기도 하였다.[33]

위와 같은 문제는 State v. Loomis 사건에서 본격적으로 점화되었는데, 본 사건에서는 위스콘신 주 대법원(Supreme Court of Wisconsin)이 성범죄 전력을 가지고 총격 현장에서 도주하다가 체포된 에릭 루미스(Loomis)[34]가 COMPAS 프로그램을 통해 재범 가능

27) 김성용/정관영, '인공지능의 개인정보 자동화 처리가 야기하는 차별 문제에 관한 연구', 서울대학교 법학, 제60권 제2호, 2019, 326쪽.
28) 김성용/정관영, 앞의 논문, 326쪽.
29) 주현경/정채연, 앞의 논문, 122쪽.
30) 양종모, '인공지능 알고리즘의 편향성, 불투명성이 법적 의사결정에 미치는 영향 및 규율 방안', 法曹, Vol. 723, 2017, 75쪽.
31) '기계학습'으로 번역되기도 하며, '기계가 일일이 코드로 명시하지 않은 동작을 데이터로부터 학습하여 실행할 수 있도록 하는 알고리즘을 개발하는 연구 분야'로 정의된다. 장민선, '인공지능(AI) 시대의 법적 쟁점에 관한 연구', 한국법제연구원 연구보고 18-10, 2018, 47쪽.
32) Michael E. Donohue, 'A Replacement for Justitia's Scales?: Machine Learning's Role in Sentencing", *Harvard Journal of Law & Technology*, Vol. 32, No. 2, 2019, p. 661.
33) 주현경/정채연, 앞의 논문, 127쪽.
34) 2013년에 미국 위스콘신 주에서 발생한 차량 이용 총격사건에 사용된 차량을 운전하여 체포되었고, 주

성이 높다는 의견이 담긴 선고전 조사보고서(pre-sentencing investigation report)를 토대로 징역 6년을 선고한 판결(Wisconsin v. Loomis)이 미국 헌법 제14조의 적법절차(due process) 원칙에 위반되는지를 다루었다.[35] Loomis는 항소하면서, 첫째, 양형에 쓰인 집단 통계 데이터가 본인이 저지른 죄와 관련성이 낮아서 불공평하다는 점을 주장했고, 두 번째, 피고인에게 인공지능 알고리즘에 대한 확인 혹은 이의제기의 기회가 주어지지 않는 점에서 절차가 투명성을 결여하고 있으며, 본인의 적법절차에 따른 권리를 침해당했다고 항변하였다.[36] 더 나아가 Loomis는 영업비밀(trade secret)이라는 이유로 보호되고 있는 사기업이 제작한 알고리즘의 소스코드(source code), 즉 판단과정 및 작동방식이 공개되어야 한다고 주장했다.

이에 2017년 위스콘신 주 대법원은 양형절차에서는 집단 데이터를 사용하는 COMPAS가 제시한 결과가 반영되었지만 법관이 개별적인 판단을 하였고, 종국적인 결정은 보조기구인 COMPAS의 결과를 부수적으로 참고한 법관의 판단에 의한 것이었으므로, 적법절차의 원칙의 위반이 아니라고 보았다.[37] 요컨대 본 사건에서 위스콘신 주 대법원은 COMPAS의 합헌성을 판단하기보다는 COMPAS의 사용에 여러 제한을 두었다고 평가된다. 재판관이 형사사법절차에서 COMPAS를 단독으로 이용하여 피고인의 감금 여부 자체를 결정하는데 쓰지 않아야 하고, 양형 시 알고리즘의 분석과 독립적인 이유를 들어야 하며, 선고 전 보고서에도 알고리즘의 한계에 대한 경고 표시를 하도록 여러 조건을 부여했기 때문이다.[38] 즉 COMPAS가 편향적이라거나 헌법과 합치하지 않는다고 판결한 것은 아니지만, COMPAS의 위험성을 인지하고 일단 유보적인 태도를 취한 것으로 보인다.

또한 2016년 NGO인 ProPublica는 COMPAS의 예상재범률 계산 알고리즘이 흑인에 대해 편파적이고 편향적인 판단을 하고 있고 흑인에게 좀 더 가혹한 선고를 내리는 데 기여할 수 있다는 결론을 담은 연구자료를 발표하였다.[39] 특히 COMPAS는 범죄관련 관계, 생활방식, 성격 및 태도, 가족과 사회로부터의 배제 등의 영역을 두고 변수를 평가하는데, 영업비밀을 근거로 연방정부의 감독을 받지 않기에 보다 공적인 영역인 형사절차에 요구되는 투명성이 결여되었다는 논란에 휩싸였다.[40]

정부로부터 기소되었다. 루미스는 5개의 기소 사유 중 '경찰로부터 도주 시도', '자동차 소유자 동의 없이 자동차 운전'한 점에 대해서만 유죄를 인정했고, 나머지 3가지에 대해서는 항변했다. 이에 순회법원(Circuit court)은 판결전 조서(pre-sentence investigation) 명령을 내렸고, 판결 전 조사 보고서에는 COMPAS의 위험평가 결과가 첨부되었다. 남중권, '머신러닝 알고리즘의 데이터 처리에 대한 법적 제한의 한계: 개인정보보호와 차별금지의 측면에서', 충북대학교 과학기술과 법, 제10권 제1호, 2019, 70쪽.

35) *State v. Loomis*, 881 N.W.2d 749, 767, Wis. 2016.

36) *Ibid.*, 757.

37) *Ibid.*, 749, 767.

38) *Ibid.*, 767.

39) Jeff Larson *et al.*, How We Analyzed the COMPAS Recidivism Algorithm, ProPublica, May 23, 2016, available at https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm [https://perma.cc/3V5S-W874].

40) 다만 이와는 반대로 알고리즘이 실제 재범확률을 인간보다 더 정확하게 예측한다는 정반대되는 주장

이러한 불확실성과 혼란에도 불구하고 Loomis의 대법원에 대한 항고는 받아들여 지지 않았고[41], 이후 또 다시 COMPAS에 비슷한 논점을 제기한 Henderson v. Stensburg 사건도 기각되었다.[42] 이처럼 Loomis사건에서 점화된 인공지능의 편향성 문제는 아직 논란이 많은 분야이며, 과연 사기업의 관할 하에 있는 혹은 사기업이 개발한 인공지능 알고리즘이 투명성이나 책무성에 있어 공적 영역에서 요구되는 수준에 도달하지 못한 것으로 보이는데도 불구하고 사용되어야 하는지에 대한 공방은 아직 진행 중이다.[43] 위와 같이 편향성 문제가 종식되지 않은 상태에서, 형사법 절차상 인공지능의 책무성은 편향성에 대응할 수 있는 방법으로서 어떻게 다루어져야 할지 고민이 필요하다.

2. 형사법 절차상 알고리즘의 책무성

책무성(accountability)이란 주로 자기 자신의 행동을 설명할 수 있는 능력에 기반한 것이다. 인공지능 맥락에서는 의사결정 과정을 설명하고 오류 또는 예기치 않은 결과를 식별하는 능력을 의미한다.[44] 알고리즘 책무성(accountability)은 주로 알고리즘의 편향성에 대응하기 위해 제안된 것이며, 이를 위해 알고리즘의 투명성(transparency), 설명가능성(explainability), 감사가능성(auditability), 공정성(fairness)를 요구하는 것으로 이해되기도 한다.[45]

여기서 투명성이란 정보처리가 공개적이고 이해할 수 있어야 한다는 원칙을 의미하며, 편향성과 위해의 인지, 접근가능과 시정, 설명가능성, 데이터 출처의 공개, 감사가능성, 검증과 테스팅을 포괄한다. 그리고 알고리즘 투명성은 해당 기업의 영업비밀에 해당하는 알고리즘의 공개범위를 어떻게 획정하고, 어떤 방식으로 누가 관리하고 통제하는가의 사전규제 문제와도 연결된다.[46]

그리고 설명가능성은 인공지능의 입력부터 출력까지 기술적 차원에서 어떻게 작동하는지에 관한 전반적인 세부사항을 요구하기 보다는, 어떤 요인(혹은 요소)들이 특정 상황에

을 가지는 연구도 진행되고 있어 이 부분에 대한 판단을 상당히 유보적일 수 밖에 없으며 주의를 요한다 : Zhiyuan Jerry Lin et al., 'The Limits of Human Predictions of Recidivism', Science Advances, 2020, pp. 1-8.

41) Loomis v. Wisconsin, 137 S. Ct. 2290 (2017). Both the state of Wisconsin and the United States Solicitor General filed briefs defending COMPAS; *See* Brief for the United States as Amicus Curiae, Loomis v. Wisconsin, 137 S. Ct. 2290, 2017, No. 16-6387, 2017 WL 2333897; *See also* Brief in Opposition, Loomis v. Wisconsin, 137 S. Ct. 2290, 2017, No. 16-6387.

42) Henderson v. Stenburg, 2020 U.S. Dist. LEXIS 28386.

43) 주현경/정채연, 앞의 논문, 131쪽.

44) 책무는 자의식이나 자유의 문제로부터 자유롭고, 궁극적으로 행위를 수행한 주체인 행위자에게 초점을 두고 있어 책임(responsibility)보다 적용하기 적합하다. 이중원, '인공지능에게 책임을 부과할 수 있는가?: 책무성 중심의 인공지능 윤리 모색', 과학철학, 22권 2호, 2019, 91면.

45) 이원태, '알고리즘 규제의 두가지 차원과 정책적 함의', 사회과학연구, 제32집 2호, 2020, 205쪽; 주현경/정채연, 앞의 논문, 134쪽.

46) 이원태, 앞의 논문, 205쪽.

서 어떻게 특정한 결과 산출에 작용하는지를 면밀하게 보는 것을 의미한다.[47] 또한 설명 가능성은 입력 정보와 출력 결과 사이 연관관계 뿐만 아니라, 최종적인 출력 결과 및 관련 행위가 함축하는 사회적, 도덕적, 법적 함의들을 밝혀내고, 이 함의들이 기존의 규범틀 안에서 어떠한 문제를 일으키는지 드러내기 까지를 함축한다.[48]

앞서 편향의 문제가 데이터 차원에서, 그리고 알고리즘 차원에서 모두 일어날 수 있음을 확인하였다. 이에 위에서 언급된 책무의 여러 요소를 데이터 레벨의 편향, 그리고 알고리즘 레벨의 편향을 살핀다.

1) 데이터 편향과 책무성

형사법 절차에서 사용되는 인공지능은 수집된 데이터를 이용하게 되는데, (1) 데이터 수집 주체, (2) 수집되는 데이터의 범위, (3) 상기 데이터를 활용한 인공지능 알고리즘 이용 및 적용 범위 등이 문제된다. 데이터 수집과 활용의 주체가 누구인지, 수집한 범죄기록 등의 공적인 데이터를 어디까지 공개할 것인지, 또한 만약 공개의 범위가 협소하다면 어느 제3자가 알고리즘을 중립적으로 검토할 것인가에 대한 문제도 제기된다. 그리고 수집되는 데이터의 외연이 어디까지 인지도 정립되어야 한다.

재범가능성을 판단하는 알고리즘이 문제된 경우 알고리즘 그 자체보다 데이터에 먼저 문제를 제기해야 한다는 의견도 있다. 데이터 중에는 잘못 선택된 데이터, 불완전하거나 부정확한 또는 시기에 맞지 않는 데이터, 편중된 표본, 역사적 편향성을 지닌 데이터 등이 문제시될 수도 있다.[49] 알고리즘은 입력한 데이터로 훈련되기에 현실의 데이터가 이미 편향적이라면 컴퓨터가 산출하는 알고리즘의 결과물도 편향적일 수밖에는 없다. 다만 알고리즘이 자체 문제를 시정할 수 있는 장치(feedback loop)를 탑재하도록 권유해 볼 수는 있을 것이다.

또한 개별 데이터 뿐만 아니라, 집단 데이터의 이용에 관해서도 편향이 생겨날 가능성이 있다. COMPAS는 회귀분석을 하는 공식을 이용하는데, 여기에 적용하는 데이터는 주로 집단의 평균 수치다.[50] 하지만 어떤 집단에 기반한 편견이 통계적인 증거로 뒷받침된다고 해도, 그것 자체가 차별의 근거가 될 수 있는지는 미지수다. 또한 데이터 세트를 구성하는 실제 인구 내 샘플의 구성도 결과의 해석에 중요한 요소인데, COMPAS의 경우에는 2004-2005년 사이 다양한 교정시설, 보호관찰 기관 등에서 수집한 30,000건의 평가를 기반으로 하고 있다.[51]

47) 이중원, 앞의 논문, 95쪽.
48) 이중원, 앞의 논문, 95쪽.
49) 박상돈, 앞의 논문, 201쪽.
50) John Lightbourne, 'Damned Lies & Criminal Sentencing Using Evidence-Based Tools', *Duke Law & Technology Review*, Vol. 15, 2015, p. 329.
51) Northpointe, Practioner's Guide to COMPAS Core 27, 2015, available at http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf.

COMPAS는 2011년 통계를 2004-2005년도 자료와 비교하여 7,381명 (남자 5,681명, 여자 1,700명)을 포괄하는 표본집단(Core Norm Group)을 구성하였고, 형사법 절차상 관여된 성인의 실제 비율도 고려하여 설계되었다.[52] 이러한 통계학적인 엄밀성에 대한 부연설명에도 불구하고, 이와 같은 집단 데이터에서 도출한 판단이 개개인에게 적법하게 적용될 수 있는지는 달리 생각해볼 여지가 있는 쟁점이다. Loomis의 집단 데이터 세트와 개인에 대한 평가 간 간극이 있을 수 있다는 주장에 대해, Ann Walsh Bradley 재판관은 COMPAS의 평가는 집단 데이터에 근거한 재범률임을 인정하고, 양형에 있어 개인을 고려하는 것이 중요하다는 점을 상기했다. 하지만 COMPAS의 평가가 단독적인 평가 기준이 아니고, 재판관의 판단이 개입되어 활용되는 것에 불과하기 때문에 필요할 경우 법정이 평가에 동의하지 않거나 재량을 발휘하는 등 개인화할 여지가 충분히 있다고 판시했다.[53]

2) 알고리즘 자체의 편향성과 책무성

State v. Loomis 사건의 재판관들이 COMPAS와 같은 알고리즘이 정확한 데이터를 넣어도 틀린 예측을 할 수 있는 가능성을 고려하지 않았다는 비판이 제기되기도 한다.[54] 이에 데이터 상의 문제 외에 알고리즘 자체의 문제도 따로 생각해 보아야 한다.

Loomis 측은 선고 전 조서내용을 공개하지 않아 적법절차에 위반되었다고 판시한 Gardiner v. Florida 사건을 들며, COMPAS가 영업비밀로 보호받는 사기업의 알고리즘이라 소스코드와 가중치를 공개하지 않은 것은 적법절차를 확보하지 못한 부분이라 지적하였다.[55] 그러나 위스콘신 주 대법원은 양쪽에서 COMPAS의 위험평가보고서(risk assessment report)를 보유하고 있기에 COMPAS의 제조사 측에서 주요한 정보를 숨겼다고 보기 어렵다고 판단하였다.[56] 그러나 Loomis 측은 위험평가 보고서 만으로는 선고 전 조서를 반박, 설명 혹은 보충할 수 없을 만큼 중요한 정보가 빠져있다는 점을 항변하였다.[57] Loomis측 변호인단은 피고인이 COMPAS가 그의 범죄전력과 선고 전 기입한 질문서상 개인정보를 이용한다는 점은 인지했지만, 이 정보가 그의 재범률 산정에 어떻게 이용되는지 예상할 수는 없었다고 주장했다.[58] 이에 따라 Loomis는 알고리즘의 합헌성에 문제를 제기한 것이었고 과연 소스코드가 공개되어야 하는지가 쟁점이 되었다.

하지만 알고리즘의 편향성을 판단함에 있어 소스코드를 보는 것은 그다지 효과적이지 않다는 반론도 있었다. 프로그래머가 위헌적인 코드를 사용할 가능성도 희박하지만, 코드 자체를 살펴보는 데서 명시적으로 확연한 차별을 찾아내기도 힘들기 때문이다.[59] 따라서

52) *Ibid.*
53) State v. Loomis, 881 N.W. 2d, 764-765.
54) Anne L. Washington, 'How to Argue with an Algorithm: Lessons from the COMPAS-PROPUBLICA Debate', *Colorado Technology Law Journal,* Vol. 17 (2018), p. 134.
55) 430 U.S. 349, 351, 1977.
56) *State v. Loomis*, 881 N.W. 2d, 761.
57) *Ibid.,* 760.
58) Rebecca Wexler, When a Computer Program Keeps You In Jail, N. Y. Times, Jun. 13, 2017.

해당 사안에서 판시한 내용만으로는 책무성의 근거를 찾기가 어렵다고 평가할 수도 있다.

그렇다면 직접 COMPAS의 제조사가 제공한 알고리즘 설명서를 참조하여 책무성을 판단해볼 수도 있을 것이다. 현재 Equivant사가 제공하는 실무자 가이드를 참조하면 COMPAS는 범죄 관여도, 불이행 전력, 폭행 전력, 현 범죄 폭력성, 범죄 동료 및 구성원 친밀도, 약물 남용, 재정상황, 교육 및 직업 숙련도, 범의, 가족력, 사회환경적 요소, 레저, 거주 안정도, 사회 적응력 등등의 다양한 요소를 참조한다.[60]

그러나 알고리즘 내 요소 간 '가중치'는 공개되고 있지 않고 있다.[61] 요소의 대략적인 위험척도(risk scale)는 제공하지만, 중요한 계수(coefficient)는 비밀로 하고 있다. 이와 같은 부분은 편향에 취약성을 노출할 수 있고, 어느 정도 책무성을 요하는 부분이 될 수 있다. 예를 들자면 COMPAS는 'Violent Recidivism Risk Score'를 다음과 같이 정의한다.[62]

Violent Recidivism Risk Score =
(age＊－w)+(age－at－first－arrest＊－w)+(history of violence＊w) + (vocation education ＊ w) + (history of noncompliance ＊ w)[63]

여기서 w는 가중치(weight)를 구성하는데, 각 요소의 w수치는 Equivant의 연구 데이터에 따라 산정되고 공개되지 않고 있다. 그러나 알고리즘을 훈련시키는 데이터와 입력한 요소마다 주어진 가중치 등은 편향과 책무성의 유무를 판단할 때 특히 중요하다.[64] 예를 들면 미국의 경우에는 우편번호(Zip code)에 가중치를 부여하게 되면, 이는 곧 인종별로 주거지가 특정되는 지역이 있는 경우, 우편번호가 인종 요소의 proxy가 되어 결국은 인종에 가중치를 두는 것과 비슷한 효과를 낼 수 있다.[65]

또한 알고리즘의 타당성(validity)과 형법의 의무 간에도 문제적인 요소가 내포되어 있다. COMPAS는 상기 식에서 도출한 수치를 기준이 되는 집단(norm group)과 비교하여 10분위 순위(decile rank)로 제공한다. Equivant사는 형사사법, 심리학, 의학 등에서 널리 쓰이는 타당성 평가 척도(Receiver Operating Characteristic (ROC) Curve)를 이용하는데, 통상 0.70이 산업계가 인정하는 기준이고, COMPAS는 이를 충족한다고 주장한다.[66] 하지만 ROC의 수치가 0.7이라는 것은 무작위로 고른 고위험도 개인이 무작위로 고른 저

59) Ellora Israni, 'Algorimatic Due Process: Mistaken Accountability and Attribution in State v. Loomis', Jolt Digest.
60) Equivant, Practitioner's Guide to COMPAS Core, 2019, available at https://www.equivant.com/wp-content/uploads /Practitioners-Guide-to-COMPAS-Core-040419.pdf.
61) Ligtbourne, *supra* note 50, pp. 330-331.
62) Equivant, *supra* note 60, p. 33.
63) 설명서에서는 각 요소(age, age at first arrest 등등)에 각기 다른 가중치를 부여했다고 설명하였다. 따라서 상기 식도 엄밀하게는 w1, w2, w3... 등으로 구별하는 것이 이해에 더 도움이 될 것으로 사료된다.
64) Israni, *supra* note 59.
65) Katherine Noyes, Will big data help end discrimination—or make it worse?, Fortune, Jan. 15, 2015.
66) Equivant, *supra* note 60, p. 12.

위험도 개인보다 고위험으로 분류될 확률이 70%라는 뜻이며, 반대로 말하면, 저위험도 개인이 실제로 고위험도 개인보다 고위험군으로 분류될 확률이 30%나 된다는 뜻이기도 하다.[67] 이는 의심스러울 때는 피고인의 이익을 고려하고, 무죄를 추정하는 형사법의 대원칙과 정합성이 의심되는 대목이다.[68] 나아가 추가적으로 비판을 받는 부분은 성별 구분에 있어서 데이터를 알고리즘에 반영할 때인데, COMPAS는 남자와 여자 norm group을 미리 따로 구별하고 있어서, 남녀 간 재범률 차이가 반영되지 않고 있다고 비판된다.[69]

비록 알고리즘의 편향에 대한 책무의 문제가 State v. Loomis 사건에서는 그다지 비중있게 다뤄지지 않았지만, Equivant사의 설명서 상의 알고리즘 자체는 설명가능성이나 투명성을 결여하고 있어 책무성을 부담해야 하는 대상으로 보인다. 가중치 w에 대한 소스코드 없이는 피고인이 효과적으로 약 30%에 육박하는 잘못 분류될 가능성을 방어하기 어려워 보이기 때문이다. 하지만 무작정 이에 대한 반대급부로 알고리즘의 소스코드를 모두 공개하라는 명령이나 입법을 추진한다면, 인공지능 알고리즘의 개발의 유인을 저해하고, 공공분야의 혁신 역시 도태될 수 있다. 이에 다음 장에서는 인공지능 알고리즘의 적용에서 가장 경합하는 가치인 인공지능 알고리즘 개발 기업의 영업비밀의 보호와 개인의 알고리즘에 대한 설명을 요구할 권리를 각각 정의하고, 무엇이 우선시되고 입법을 통해 보호받아야 하는지 고민해 본다.

Ⅳ. 자동화된 의사결정에 대한 설명을 요구할 권리와 영업비밀보호 간 균형의 모색

알고리즘을 규제하는 규제 거버넌스를 어떻게 설계할지에 대해서는 행위주체마다 각기 다른 형태를 주장하고 있다. 윤리 가이드라인을 만들거나, 기업이 알고리즘 내 스스로 규제 메커니즘을 확보하거나 하는 방법도 있지만, 본고에서는 국가가 직접적으로 개인에게 (사전적인) 권리를 보장할 수 있는 방법을 착안해보는 것을 목표로 한다.

1. 자동화 의사결정에 대한 설명을 요구할 권리

기존의 알고리즘 규제 거버넌스의 논의는 정부, 기업, 개발자 등 행위주체와 몇 가지 규제원칙이 주로 논하여져 왔지만, 이제는 알고리즘의 수혜자이자 잠재적 피해자도 될 수 있는 이용자의 보호를 위한 논의가 이루어야 한다는 지적이 있다.[70] 알고리즘 시대의 이용자 보호는 인간주체로서 향유해야 할 기본권의 보호뿐만 아니라, 합리적 의사결정 주체로서의 소비자/이용자 선택권의 권익문제와도 결부되고, 알고리즘 책무성의 여러 원칙들

67) Lightbourne, *supra* note 50, p. 336.
68) Frederik J. Zuiderveen Borgesius, "Strengthening Legal Protection Against Discrimination by Algorithms and Artificial Intelligence", *The International Journal of Human Rights* (2020), , p. 14.
69) Lightbourne, *supra* note 50, p. 332.
70) 이원태, 앞의 논문, 209쪽.

과도 긴밀하게 연결된다.[71] 특히 형사법 절차에서 성공적으로 인공지능을 활용한다고 하더라도, 적법절차의 보장을 위해 피고인의 방어권에 포섭될 수 있는 당사자의 이의제기 가능성의 문제가 남아있다. 형사법적 결정은 개인의 신체의 자유라는 기본권을 침해하기에 알고리즘에 대한 정보권, 즉 자동화 의사결정에 대해 설명을 요구할 권리의 문제가 남아있다.[72]

EU GDPR(General Data Protection Regulation) 제22조는 정보주체의 자동화된 의사결정 거부권을 보장하고 있는데, 이를 통해 인공지능 알고리즘에 투명성과 책무성을 요구하고 있다. 동 조항은 정보주체는 프로파일링 등, 본인에 관한 법적 효력을 초래하거나 이와 유사하게 본인에게 중대한 영향을 미치는 자동화된 처리에만 의존하는 결정의 적용을 받지 않을 권리를 가진다고 명시한다. 그러나 GDPR이 컨트롤러에게 설명의무를 부과한 것인지, 아니면 정보주체에게 설명을 요구할 권리를 인정한 것인지에 대해 아직 논의가 이어지고 있으며, 컨트롤러의 설명의무의 내용이 무엇인지에 대해서도 확정된 사항은 없다.[73] 그리고 동 조문 제2조 제2항에서는 범죄의 예방, 수사, 탐지, 기소 및 형벌 집행의 목적으로 관계당국에 의해 이루어지는 처리에는 적용되지 않는다고 명시한다. GDPR 자체의 설명을 요구할 권리에도 아직 모호한 부분이 남아있는데 더해 형사적인 부분에는 더욱 조심스러운 예외를 설정한 것이다.

그럼에도 불구하고 형사법 절차에는 여전히 설명을 요구할 권리를 주장해볼 가능성이 남아있다. EU는 2016/680 Directive (27 April 2016)[74] 에서 데이터 처리가 적법, 정당, 투명해야 하며, 유럽 인권 및 기본적 자유 보호협약(European Convention for the Protection of Human Rights and Fundamental Freedoms)의 제6조와 유럽연합 헌장 제47조의 공정한 재판을 받을 권리(right to fair trial)에서 이를 도출하한다.[75] 또한 자연인은 본인의 개인정보 처리와 관련한 위험, 규칙, 보호장치(safeguard), 권리 등을 인지해야 할 수 있어야 한다고 명시한다.[76] 또한 유럽평의회(Council of Europe)의 개인정보 자동처리에 관한 개인보호협약 협의위원회는 범죄행위 예방, 수사, 기소, 형집행을 '경찰분야'로 포섭하였고, Draft practical guide on the use of personal data in the police sector (15 February 2018)에서는 'data subject's rights' 부분에 접근권(right of access)을 규정하고 있다.[77] 그리고 조사나 다른 경찰 분야 절차를 위해 피고인의 데이

71) 이원태, 앞의 논문, 209-210쪽.
72) 주현경/정채연, 앞의 논문, 149쪽.
73) 이선구, '알고리즘의 투명성과 설명가능성: GDPR을 중심으로', 서울대학교 인공지능정책 이니셔티브 이슈페이퍼 2019-2 '미디어 알고리즘과 민주주의', 2019, 3-4쪽.
74) European Union, 'Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, L119/89.
75) *Ibid.*, para. 26.
76) *Ibid.*

터가 수집되었다면, 상황이 허락하는 즉시 정보제공자가 데이터에 접근을 할 수 있도록 해야 한다고 적시하였다.

그러나 접근권은 정정권(right of rectification) 이나 삭제권(right of erasure)로 더 나아갈 토대를 마련하지만, 만약 경찰의 임무, 국가이익(공공안보, 국가안보 등) 등의 타인의 자유와 권리의 수호, 그리고 조사를 위해서는 제한되거나 배제될 수도 있다고 하였다.[78] 또한 만약 데이터 처리 기술이 개인의 권리를 침해할 가능성이 있는 경우 기법의 책무성을 고려하여 데이터보호영향평가(Data Protection Impact Assessment)를 진행해야 한다고 하였다.[79]

EU의 논의를 바탕으로 한국의 헌법 제10조 및 제17조에서 도출된 개인정보 자기결정권을 바탕으로 설명을 요구할 권리를 도출해보려는 학술적인 논의가 있었다. 개인정보 자기결정권의 내용으로 자기정보 수집통제권, 자기정보 보유통제권, 자기정보 이용·제공통제권 등에 개인정보의 처리와 이용에 대한 구체적인 안내를 요구하는 권리가 있으므로, 개인정보를 활용하는 자동의사결정 알고리즘에 대해 설명을 요구하는 권리와 같은 맥락에서 볼 수 있다는 것이다.[80] 물론 현재의 기술 수준으로는 일정부분 설명이 불가능한 요소가 있을 수는 있지만 그것을 이유로 해당 기본권의 존재 자체가 부정되어서는 안 된다고 주장되기도 한다.[81]

설명을 요구할 권리의 유무 뿐만 아니라 설명의 '범위'에 대해서도 논의가 진행되고 있다. 가장 일반적인 부분에서 구체적인 부분에 대한 설명 순으로 나열하자면 i) 결정의 논리, ii) 알고리즘, iii) 특정 의사결정의 이유로 세분화될 수 있다.[82] 우선 결정의 논리는 전문적인 알고리즘 보다는 결정에 사용된 기준, 분류체계, 의사결정모형의 논리를 알려주는 것이다. 두 번째 알고리즘은 결정의 논리에서 더 나아가 알고리즘의 공식화된 일련된 절차나 방법을 알려줄 수도 있다. 세 번째로는 구체적인 결과를 도출하게 된 결정요소를 설명해야 하는 것인데, 특정 정보의 유의미성 판단, 분석에 포함되는 모든 통계 및 프로파일, 그리고 특정 프로파일이 자동화된 의사결정에서 어떤 의미를 가지고, 이러한 프로파일이 정보주체에 대한 의사결정에서 어떻게 사용되었는지를 알려주는 것이다.

앞서 COMPAS의 경우에는 결정의 논리까지만 설명이 된 상태라고 해석할 수 있을 것인데, 이를 통해서는 편향이나 책무의 본질을 다루는데 제한이 있음을 이미 판단한 바 있다. 따라서 실질적으로 형사사법절차상 피고인이 인공지능 알고리즘에 대한 설명을 요구

77) Council of Europe, 'Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data: Practical Guide on the Use of Personal Data in the Police Sector', 15 Feb. 2018, p. 6, available at https://rm.coe.int/t-pd-201-01-practical-guide-on-the-use-of-personal-data-in-the-police-/16807927d5.
78) *Ibid.*
79) *Ibid.*, p. 9.
80) 박상돈, 앞의 논문, 205쪽.
81) 박상돈, 앞의 논문, 212-213쪽.
82) 이선구, 앞의 논문, 16-18쪽.

할 권리를 구체화한다면 알고리즘 단계, 또는 특정 의사결정의 이유까지 범위를 넓히는 것이 유의미할 것으로 보인다.

2. 영업비밀의 보호와 설명을 요구할 권리 간 균형 모색

영업비밀의 보호를 통해 보호하는 가치와 설명을 요구할 권리를 통해 보호하는 가치는 상호 배타적이라는 시각을 견지하기보다는 적절한 균형을 통해 알고리즘의 투명성 증진을 위해 동시에 고려해야 할 요소들이라고 볼 여지가 있을까? 이와 같은 의문에 대응하기 위해 우선 알고리즘의 소스코드의 영업비밀성을 판단한 뒤, 균형을 제시할 실마리를 주는 설명가능한 인공지능을 양 가치 사이에 대입해본다.

State v. Loomis 사건에서 영업비밀의 대상이 되었던 소스코드는 프로그램을 개발할 때에 사용하는 언어나 툴로 만들어진 일련의 지시, 명령의 집합이라고 할 수 있는데, 언어 자체는 저작물성을 가지기 어렵지만 프로그램 언어로 제작된 소스코드 자체는 텍스트 상태로 저장되어 어문제작물로 보호를 받게 된다.[83] 미국의 경우 통일영업비밀법(Uniform Trade Secrets Act)는 영업비밀의 성립요건으로 ⅰ) 실재적 또는 잠재적으로 독립적인 경제 가치를 가지고, ii) 영업비밀로 이익을 얻을 수 있는 제3자에게 일반적으로 알려져 있지 않고, iii) 적절한 수단을 통해 쉽게 확인할 수 없고, iv) 비밀 유지를 위하여 합리적인 조치가 있는 정보를 들고 있다.[84]

우리나라의 법제와 비교해본다면, 국내에서는 재범가능성을 계산하는 소스코드는 「부정경쟁방지 및 영업비밀 보호에 관한 법률(부정경쟁방지법)」[85]에 따라 공공연히 알려져 있지 아니하고(비공지성), 독립적인 경제적 가치를 지니며 (경제적 유용성), 비밀유지에 대한 합리적 노력에 의해 관리(비밀관리성) 되는 등의 요건[86]을 충족한다면 공공연히 알려져 있지 아니하고 독립된 경제적 가치를 가지는 것으로서, 상당한 노력에 의하여 비밀로 유지된 생산방법, 판매방법, 그 밖에 영업활동에 유용한 기술상 또는 경영상의 정보에 포함될 수 있다.

비공지성은 어떤 정보가 간행물 등의 매체에 실리는 등 불특정 다수인에게 알려져 있지 않은 것을 의미한다. 보유자를 통하지 아니하고는 그 정보를 입수할 수 없고, 역설계(reverse engineering)로 정보를 취득하려고 해도 장기간이 소요되고 고비용이 든다면 비공지로 인정된다.[87] 양형 관련 인공지능 알고리즘의 데이터는 외부에 노출되거나 공공

83) 김윤명, '게임물 제작상 영업비밀의 보호', 산업재산권, Vol. 37, 2012, 194쪽.
84) 정진근 외, '「부정경쟁방지 및 영업비밀보호에 관한 법률」에 대한 입법평가, 한국법제연구원 입법평가연구, 18-15-4, 2018, 98-100쪽.
85) 우리나라에서 영업비밀 보호를 위해 적용할 수 있는 법은 「부정경쟁방지법」 외에도 「산업기술의 유출방지 및 보호에 관한 법률(산업기술보호법)」, 「대·중소기업 상생협력 촉진에 관한 법률」,「형법」, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「통신비밀보호법」 등이 있다. 정진근 외, '「부정경쟁방지 및 영업비밀보호에 관한 법률」에 대한 입법평가, 한국법제연구원 입법평가 연구, 18-15-4, 2018, 85쪽.
86) 정진근 외, 앞의 논문, 89-90쪽.

연하게 알려져 있지는 않기에 비공지성을 충족할 것으로 보인다.

비밀관리성은 ⅰ) 당해 정보에 접근할 수 있는 사람의 수, 물리적 공간적 접근성을 제한하는 경우, ii) 당해 정보에 비밀 표시를 하여 접근할 수 있는 자에게 그것이 영업비밀이라는 사실을 주지시키는 경우, iii) 접근자에게는 그 정보를 권한없이 사용하거나 공개해서는 안된다는 취지의 비밀 준수의무를 부과하고 있는 경우, iv) 영업비밀관리규정, 서약서, 취업규칙 등에 비밀지정 및 비밀유지의무를 규정하고 있는 경우로 판단할 수 있다.[88] 양형 인공지능 알고리즘의 경우 인자 외에 가중치변수 등에는 철저하게 비밀을 유지하고 있기에 이 부분도 해당할 수 있다.

경제적 유용성은 정보의 보유자가 그 정보의 사용을 통해 경쟁자에 대해 경쟁상의 이익을 얻을 수 있거나 또는 그 정보의 취득이나 개발을 위해 상당한 비용이나 노력이 필요함을 말한다. 종국적으로는 그 유용성이 건전하고 올바른 상거래 질서의 유지를 위한 법테두리 내에서 일반인이 갖는 통념에서 판단되어야 한다.[89] 양형 인공지능은 형사사법 절차상 자원을 재배치하는데 도움을 주는 등의 유용성은 가지고 있다고 볼 수 있을 것이다.

따라서 양형 인공지능은 부정경쟁방지법상 영업비밀로서 보호받을 수 있다.[90] 다만 본고에서 다루는 형사법 절차상 인공지능의 경우 공공영역의 데이터를 기반으로 한 것으로서, 공적인 의사결정에 활용되기 때문에 투명성 문제에 대해 재고할 여지가 있다. 이와 같은 특성을 바탕으로 설명력을 증진하면서도 영업비밀의 주요 요건을 저해하지 않는 도구로서 설명가능한 인공지능이 제안된다.

설명 가능한 인공지능(XAI ; Explainable AI)[91]은 기존 머신러닝의 고차원적인 학습

87) 장완규, '초연결사회의 도래와 빅데이터-법제도적 개선방안을 중심으로", 한남대학교 과학기술법연구, 제24집 제2호, 2018, 141쪽.

88) 장완규, 앞의 논문, 142쪽.

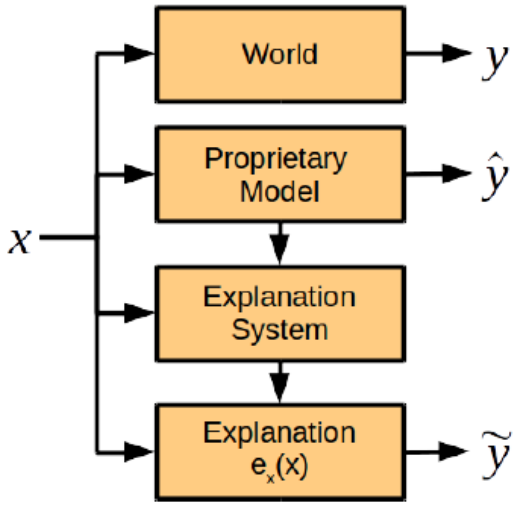
89) 장완규, 앞의 논문, 142쪽.

90) 이러한 해석은 다소 직관적일 수 있다는 점에서 보완설명을 추가한다. **(비공지성)** 우리 형사소송법에 따르면 양형부당이 있다면 상소할 수 있는데(형사소송법 제361조의5 제15호, 제383조 제4호) 이에 따라 구체적인 사건에 있어 가중치변수를 포함한 양형인자가 상소이유로서 공개되어야 한다는 형사법의 원칙과 조화되지 못한다는 비판이 가능하다. 하지만 오늘날 미국을 비롯하여 어느 입법례에 따르든 양형 관련 인공지능 알고리즘의 판단은 법원에 구속력을 가지지 못하며 단지 참고자료가 될 뿐이다. 법관에 의한 별도의 양형과정이 필요하며 이는 상소이유로서 공개될 것이다. **(경제적 유용성 및 비밀 관리성)** 양형 인공지능이 영업비밀로 보호받기 위해서는 경제적 유용성과 이에 기반한 사업자의 비밀관리성을 구비해야 하는데 그 판단에 있어 알고리즘에 사용된 소스코드가 오픈소스인지 여부에 따라 좌우될 수 있다는 지적이 제기될 수 있으며 타당하다. 오픈소스는 그 사용에 있어서도 재사용을 전제해야 하는 등 제약이 가해지는 경우가 많아 영업비밀성이 인정되는 데에 제약요소로 작용할 수 있기 때문이다. 본 연구에서 분석대상으로 삼은 알고리즘은 오픈소스가 아닌 소스코드를 전제한 것으로 한다. 오픈소스의 경우에도 예외적으로 영업비밀성이 인정될 여지가 있기는 하나 인공지능에 기반한 형사법상 의사결정을 연구하고자 하는 본 논문의 범위를 벗어나므로 그 예외에 대해서는 자세히 언급하지 않는다.

91) 알고리즘의 설명을 구분하는 방법에는 알고리즘 모델의 논리(logic)를 이용하는 Model-Centric Explanation (MCEs)와 Subject-Centric Explanations (SCEs), 그리고 알고리즘 자체의 논리와는 별개로 작용하는 decompositional 설명과 pedagogical 체계가 있다. MCEs의 경우 모델 그 자체의 구조의 논리에, SCEs는 입력 데이터와 알고리즘 결정의 대상에 집중한다는 점에서 구분된다. pp.

능력은 유지하면서도 설명가능성을 향상시키는 연구를 의미한다. 사용자가 인공지능 시스템의 동장과 최종결과를 이해하고 올바르게 해석하여 결과물이 생성되는 과정을 설명 가능하도록 해주는 기술을 의미한다.[92] 국내에서는 과학기술정보통신부가 인공지능국가전략프로젝트로 2017년부터 개발 중에 있고, 미국은 미국방위고등연구계획국(DARPA: Defense Advanced Research Projects Agency)에서 개발 중이다. 설명 가능한 인공지능은 어떤 판단과 행동을 수행하는 인공지능 시스템과 이것의 의사결정 과정을 인간에게 유의미하게 해석해주는 설명 시스템을 모두 갖추는 것을 전제한다.[93]

설명가능한 인공지능은 국지적인 설명(local explanation), 즉 체계의 행위 전체 보다는 어떤 특정한 결과에 대한 설명과 조건법적 신뢰(counterfactual faithfulness), 즉 어떤 설명이 인과적인지를 측정하는 개념을 통해 작동된다.[94] 인공지능 자체는 데이터를 입력하면 어떤 예상결과를 내놓는 (예측하기 힘든) 블랙박스인데, 보통 고안자는 그림 1상의 현실세계의 결과(y)와 예측상의 결과(ŷ)가 맞아 떨어지기를 희망한다.[95]



[그림 1] 설명가능한 AI의 체계도

이에 설명 체계의 고안자는 인간이 이해할 수 있는 예측이자 국지적인 설명(local

Decompositional 설명은 알고리즘 블랙박스를 열어서 weights, neurons, decision trees, architecture등을 보는 것이며, pedagogical 체계는 실제로 알고리즘에 여러가지 질문을 하여 설명을 구하는 방식이다. Lilian Edwards & Michael Veale, "Slave to the Algorithm: Why a Right to an Explanation is Probably Not the Remedy You Are Looking For', *Duke Law & Technology Review,* 2017-2018, 57-58. 64-65쪽.

92) 금융보안원, '설명 가능한 인공지능(eXplainable CI, XAI) 소개', 보안연구부 2018-020 (2018), 2쪽.

93) 이중원, 앞의 논문, 96쪽.

94) Doshi-Velez, Finale and Kortz, Mason, "Accountability of AI Under the Law: The Role of Explanation", Berkman Klein Center for Internet & Society Working Paper (2017). p. 7.

95) *Ibid.*

explanation)을 통해 만들어진 예측($\hat{y}$)을 내놓기 위해 입력 데이터 x를 대입할 수 있는 인간이 이해할 수 있는 규칙 (ex (x))을 이용하게 된다.[96] 여기서 x를 조금씩 바꾸면서 (인종의 종류를 변경하는 등) 결과를 도출했을 때 $\hat{y}$와 $\hat{y}$가 비슷하다면 조건법적 신뢰(counterfactual faithfulness)를 충족하여, 본 체제가 믿을만하다는 증빙이 될 수 있다.[97] 이는 사실상 인간과 인공지능 간의 상호작용을 보다 효율적이고 편리하게 하는 기법인 것이다. 설명가능한 인공지능은 기존의 인공지능 알고리즘에 적용된다면 사용자의 신뢰를 얻고, 사회적 수용을 위한 공감대를 형성해볼 수 있는데, 특히 몇 가지 설명 방식은 알고리즘의 블랙박스를 개봉하지 않고도 어느 정도의 설명력을 확보할 수 있다.[98]

만약 이러한 기대가 실제로 충족된다면 영업비밀의 가치를 보존한 채 피고인에게 알고리즘에 대한 설명을 제공할 가능성이 열리게 된다. 또한 알고리즘의 성능에 관련된 요인을 파악하고 최적의 모델을 위한 통찰력을 제공할 수도 있으며, 알고리즘이 잘못된 결과를 산출하여 분쟁의 소지가 있을 경우 책무성의 판단근거를 제공할 수도 있다. 물론 형사사법절차에 관련된 양형 인공지능 알고리즘에 대입할 수 있는지 여부는 이론을 넘어서 실제적으로 연구해보아야 할 부분이겠지만, 앞으로 관련 알고리즘을 연구하고 개발할 때 참고할 수 있는 대안이 될 수 있는지는 지속적인 연구가 필요할 것이다.

3. 균형점의 법제화 가능성

우선 제안할 수 있는 방안은 제도적인 측면에서 책무성 중 감시가능성을 제고할 수 있는 위원회나 직책을 신설하는 것이다. 공적인 영역, 특히 형사법과 관련된 분야에서 인공지능을 도입할 때는 위험성을 적절히 판단할 수 있는 기구를 설립하거나 제3자 감시의무를 명시할 수 있다.[99] 독립된 정보공개위원회를 신설하고, 필요시 주요법원 마다 정보공개책임관[100]을 두는 것도 고민해볼 수 있다.

보다 직접적으로 개인정보나 영업비밀 관련 법제를 개정하는 부분은 상대적으로 조심스러운 측면이 있다. 기존 입법은 법체계 전반을 고려하여 대립되는 이익간 균형을 추구한 산물이기 때문이다. 이에 미국의 경우 정부의 공공 기록이나 절차에 대해서 인공지능 알고리즘의 투명성 및 책무성을 보장하기 위한 방법으로 별도의 정보공개법(FOIA: Freedom of Information Act)을 제정하는 것이 인공지능에 대한 새로운 접근법(Access Law)으로 거론되고 있다.

여기서 개인정보를 이용하여 어떤 구조, 운영, 혹은 의사결정 절차(structure, operation, or decision making procedures)에 영향을 주는 사적인 의사결정 방법의 경우

96) *Ibid.*
97) *Ibid.,* pp. 7-8.
98) Edwards & Veale, *supra* note 91, pp. 57-58, 64-65.
99) Frederik J. Zuiderveen Borgesius, "Strengthing Legal Protection Against Discrimination by Algorithms and Artificial Intelligence", *The International Journal of Human Rights,* 2020, p. 12.
100) Edwards & Veale, *supra* note 91, p. 76.

해당 부분이 공개되어야 한다는 주장을 해볼 수 있을 것으로 전제된다.[101] 물론 위의 제안은 특별법 제정을 통해 영업비밀에 대한 예외를 설정하고 있지만, 이는 공공의 정밀검토(scrutiny)에서 공공 의사결정을 배제하기 위한 것이 아니라 정부와 정보공유를 원활하게 하기 위함이다.[102]

이와 같은 목적을 고려해볼 때 피고인이 아닌 정부가 공익을 보호하기 위해 인공지능 알고리즘의 설명가능성을 직접 요구하는 방향도 고려해볼 수는 있다. 그 예시로서 정부와 알고리즘 개발 기업간 정부조달에 관한 부분에서도 여러 조건을 더 추가해볼 수 있다. 일례로 정부가 영업비밀인 인공지능 알고리즘을 공적인 영역에서 구현할 때 정부조달계약에 알고리즘의 의사결정에 관련된 투명성을 제고할 수 있는 방향으로 계약조건을 구성해볼 수 있으리라는 주장이 제기된다.[103] 특히 워싱턴주에서는 자동화된 의사결정 체제의 정부조달계약에서 비공개 조문(non disclosure provision)이나 투명성에 반하는 조항을 담지 않도록 하는 법안이 제출된 바 있다.[104] 만약 이 부분이 어렵다면 차별이 일어날 수 있는지 자체적으로 분석이 가능하도록 설명가능한 인공지능을 이용한 대상 인공지능 알고리즘의 소스코드 분석을 허용하고, 특히 프로그램을 작동시킨 후 그 작동상황을 동적으로 분석할 수 있기 위한 조건을 생성해볼 수 있다.[105]

그리고 필요한 경우 법원이 정보를 피고인에게만 공개하되 공개적으로 완전히 공개하지는 않을 수 있는 정보 혹은 문서보호 명령(protective order)을 고민해야 한다는 제안도 있다. 이 부분은 앞서 다른 제안보다는 인공지능 개발 기업의 비밀성을 침해할 요소가 다분하기에 다소 우려는 있겠지만 근본적인 기본권 침해가 예상되는 경우에는 고려해볼 여지가 있을 것이다.[106]

V. 결론

아직 인공지능의 책무성와 설명가능성을 직접적으로 입법하는 것은 어려운 작업으로 여겨진다. 미국의 경우 Algorithmic Accountability Act of 2019 법안이 제안되었지만 아직 실질적인 설명의무나 구체적인 윤리규범으로 명문화되지는 않았다.[107] 그럼에도 불구

101) Hannah Bloch-Wehba, 'Access to Algorithms', *Fordham Law Review*, Vol. 88, 2020, pp. 1299-1300.
102) *Ibid.*, p. 1300.
103) *Ibid.*, p. 1307.
104) H.R. 1655, 66th Leg., Reg. Sess., Wash. 2019.
105) 양종모, '인공지능 알고리즘의 편향성, 불투명성이 법적 의사결정에 미치는 영향 및 규율 방안', 法曹, Vol. 723, 2017, 89-90쪽.
106) Bloch-Wehba, *supra* note 101, p. 1308.
107) Iria Giuffrida, 'Symposium: Rise of the Machines: Artitificial Intelligence, Robotics, and the Reprogramming of Law: Liability for AI Decision-Making: Some Legal and Ethical Considerations', *Fordham Law Review*, Vol. 88 (2019), p. 455; H.R. 2231-Algorithmic Accountability Act of 2019.

하고, 본고의 공공분야에서의 알고리즘 책무성과 설명을 요구할 권리 논의는 앞으로도 계속되어야 할 것으로 보인다. State v. Loomis 사건에서 알고리즘 소스코드의 영업비밀 보호 선례는 확보되었지만, 알고리즘 책무성이나 설명을 요구할 권리 등의 고려가 부재한 채 적법절차를 판단해버린 결과 명확한 기준이 제시되지 않아 공공분야의 인공지능 활용이 정체되었기 때문이다. 다시금 공공분야에 있어 알고리즘의 활용을 진작하고, 의도했던 행정 및 형사법적 효율성을 추구하려면, 결국에는 알고리즘의 책무성과 설명을 요구할 권리를 영업비밀 등의 사익과 어떻게 조화할지 문제가 선결적으로 해결되어야 한다.

본고에서는 현재 설명가능한 인공지능의 발전상을 토대로, 위원회의 신설, 정보공개와 접근법에 대한 새로운 해석, 정부조달계약의 조건 구체화 등을 제안하였다. 앞으로 설명가능한 인공지능의 기능이 더욱 발달하고 인간과 인공지능 간의 상호작용을 원활화하는 제도적 방법이 더 구비된다면, 새로운 입법론도 점진적으로 논의할 수 있을 것이다. 이를 위해서는 인공지능 알고리즘의 편향성, 책무성, 설명가능한 인공지능의 방법론 등에 대한 연구가 더 이루어져야 할 것이다. 이와 보조를 맞추기 위해서는 인공지능 시대에서 기존의 공공영역의 법제가 지향하는 지점에 대해 다시금 고민하고, 인공지능 알고리즘을 통해 비약적으로 개선할 수 있는 분야를 특정하는 연구도 아울러 수행되며 시너지를 창출할 수 있어야 할 것이다. 본 연구가 그에 대한 자극이 될 수 있기를 희망한다.

<참고문헌>

<국내문헌>

관계부처 합동, ‘인공지능 국가전략’, 2019.
김광수, ‘인공지능 기반 과학시술과 국민의 권익구제: 자율주행차, 드론 및 의료기기를 중심으로’, 토지공법연구, 제85집, 2019.
김기범, ‘형사사법정보의 이용제공 실태 및 입법적 개선방안’, 한국법학회 법학연구, 제16권 제1호, 통권 제61호, 2016.
김도승, ‘인공지능 기반 자동행정과 법치주의’, 미국헌법연구, 제30권 제1호, 2019.
김성용/정관영, ‘인공지능의 개인정보 자동화 처리가 야기하는 차별 문제에 관한 연구’, 서울대학교 법학, 제60권 제2호, 2019.
김윤명, ‘게임물 제작상 영업비밀의 보호’, 산업재산권, 제37호, 2012.
김재완, ‘EU 일반정보보호규정(GDPR)의 알고리즘 자동화 의사결정에 대한 통제로써 설명을 요구할 권리에 대한 쟁점 분석과 전망’, 민주법학, 제69호, 2019.
김중권, ‘인공지능시대에 알고리즘에 의한 행위조종과 가상적 행정행위에 관한 소고’, 공법연구, 제48집 제3호, 2020, pp. 287-312.
남중권, ‘머신러닝 알고리즘의 데이터 처리에 대한 법적 제한의 한계: 개인정보보호와 차별금지의 측면에서’, 충북대학교 과학기술과 법, 제10권 제1호, 2019.
박상돈, ‘헌법상 자동의사결정 알고리즘 설명요구권에 관한 개괄적 고찰’, 헌법학연구, 제23권 제3호, 2017.
선지원, ‘인공지능 알고리즘 규율에 대한 소고 - 독일의 경험을 중심으로’, 경제규제와 법, wp12권 제1호, 통권 제23호, 2019.
심우민, ‘인공지능의 발전과 알고리즘의 규제적 속성’, 법과사회, 53호 2016.
심우민, ‘인공지능과 법패러다임 변화 가능성: 입법 실무 거버넌스에 대한 영향과 대응과제를 중심으로’, 법과사회, 제56호, 2017.
양종모, ‘인공지능 알고리즘의 편향성, 불투명성이 법적 의사결정에 미치는 영향 및 규율방안’, 法曹, Vol. 723, 2017.
양천수, ‘인공지능과 법체계의 변화: 형사사법을 예로 하여’, 법철학연구, 제20권 제2호 2017.
오병두, ‘고문방지협약의 국내적 이행과 형사실체법적 쟁점: 고문범죄의 처벌을 중심으로’, 민주법학, 제37호, 2008.
이선구, ‘알고리즘의 투명성과 설명가능성: GDPR을 중심으로’, 서울대학교 인공지능정책 이니셔티브 이슈페이퍼 2019-2, ‘미디어 알고리즘과 민주주의’, 2019,
이원태 외, ’지능정보사회의 규범체계 정립을 위한 법·제도연구‘, 정보통신정책연구원 기본연구 16-09, 2016.
이원태, ’알고리즘 규제의 두가지 차원과 정책점 함의‘, 사회과학연구, 제32집 2호. 2020.

이중원, '인공지능에게 책임을 부과할 수 있는가?: 책무성 중심의 인공지능 윤리 모색', 과학철학, 제22권 제2호, 2019.
윤상오, '인공지능 기반 공공서비스의 주요 쟁점에 관한 연구: 챗봇(ChatBot) 서비스를 중심으로', 한국공공관리학보, 제32권 제2호, 2018.
장민선, '인공지능(AI) 시대의 법적 쟁점에 관한 연구', 한국법제연구원 연구보고 18-10, 2018.
장완규, '초연결사회의 도래와 빅데이터: 법제도적 개선방안을 중심으로', 한남대학교 과학기술법연구, 제24집 제2호, 2018.
정원준/선지원/김정언, '인공지능 시대의 법제 정비 방안', 정보통신정책연구원 19-07, 2019.
정용기/송기복, '인공지능(AI)의 발전과 형사사법의 주요논점, 한국경찰연구, 제18권 제2호, 2019.
정진근 외, '부정경쟁방지 및 영업비밀보호에 관한 법률에 대한 입법평가, 한국법제연구원 입법평가 연구, 18-15-4, 2018.
주현경/정채연, '범죄예측 및 형사사법절차에서 알고리즘 편향성 문제와 인공지능의 활용을 위한 규범 설계', 조선대학교 법학논총, 제27집 제1호, 2020.
행정안전부, 방송통신위원회, 한국인터넷진흥원, '우리 기업을 위한 EU 일반 개인정보보호법 가이드북', 2018.

<외국문헌>

Allen Robin, Masters Dee, 'Artificial Intelligence: the Right to Protection from Discrimination Caused by Algorithms, Machine Learning and Automated Decision-Making', Europäische Rechtsakademie, Vol. 20, 2020.
Bavitz, Christopher et al., "Assessing the Assessments: Lessons From Early State Experiences in the Procurement and Implementation of Risk Assessment Tools", Berkman Klein Center for Internet & Society research Publication, 2018.
Bodo. B. et al., 'Tackling the Algorithmic Control Crisis – the Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents', Yale Journal of Law and Technology , Vol. 19, Issue 1, 2017.
Borgesius, Frederik J. Zuiderveen, 'Strengthning Legal Protection Against Discrimination by Algorithms and Artificial Intelligence', The International Journal of Human Rights, 2020.
Caianiello Michele, Criminal Process faced with the Challenges of Scientific and Technological Development, European Journal of Crime, Criminal Law and Criminal Justice, Vol. 27, 2017.
Council of Europe, 'Consultative Committee of the Convention for the Protection

of Individuals with Regard to Automatic Processing of Personal Data: Practical Guide on the Use of Personal Data in the Police Sector', 15 Feb. 2018,

Donohue E. Michael, 'A Replacement for Justitia's Scales?: Machine Learning's Role in Sentencing', Harvard Journal of Law & Technology, Vol. 32, No.2 (2019).

Doshi-Velez, Finale and Kortz, Mason, "Accountability of AI Under the Law: The Role of Explanation", Berkman Klein Center for Internet & Society Working Paper, 2017.

Edwards Lilian and Veale Michael, 'Slave to the Algorithm? Why a 'Right to an Explanation' is Probably Not the Remedy You are Looking for', Duke Law & Technology Review, Vol. 16, 2017-2017.

European Union, Directive (EU) 2016/680 of the European Parliament and of the Council of 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data by Competent Authorities for the Purposes of the Prevention, Investigation, Detection or Prosecution of Criminal Offenses or the Execution of Criminal Penalties, and on the Free Movement of Such Data, and Repealing Council Framework Decision, 2008/977/JHA.

Giuffrida Iria, 'Rise of the Machines: Artificial Intelligence, Robotics, and the Reprogramming of Law: Liability for AI Decision-making: Some Legal and Ethical Considerations', Fordham Law Review, Vol. 88, 2019.

Janssen, Heleen L., 'An Approach for a Fundamental Rights Impact Assessment to Automated Decision-Making', International Data Privacy Law, Vol. 10, No.1, 2020.

Joh, E. Elizabeth, 'Artificial Intelligence and Policing: First Questions', Seattle University Law Review, Vol. 41, 2018.

Kaminski, E. Margot, 'The Right to Explanation, Explained', Berkeley Technology Law Journal, Vol. 34,, 2019.

Katyal K. Sonia, 'The Paradox of Source Code Secrecy', Cornell Law Review, Vol. 104, 2019.

Lightbourne, John, 'Damned Lies & Criminal Sentencing Using Evidence-Based Tools', Duke Law & Technology Review, No.15, 2019.

Lin Zhiyuan, 'The Limits of Human Predictions of Recidivism', Social Advances, 2020.

Liu, Han-Wei, Lin, Ching-Fu and Chen, Yu-Jie, 'Beyond State v Loomis: artificial intelligence, government algorithmization and accountability, International Journal of Law and Information Technology, Vol. 27, 2019.

Malgieri, Gianclaudio, 'Trade Secrets v Personal Data: a Possible Solution for Balancing Rights', International Data Privacy Law, Vol. 6, No. 2, 2016.

Manheim Karl and Kaplan Lyric, 'Artificial Intelligence: Risks to Privacy and Democracy', Yale Journal of Law & Technology, Vol. 21, 2019.

Nutter W. Patrick, 'Machine Learning Evidence: Admissibility and Weight', University of Pennsylvania Journal of Constitutional Law, Vol. 21, 2019.

Office for Artificial Intelligence, Government Digital Service, 'A Guide to Using Artificial Intelligence in the Public Sector', 2017.

The Law Society, 'Algorithms in the Criminal Justice System: A report by The Law Society Commission on the Use of Algorithms in the Justice System, The Law Society of England and Wales', June 2019.

Washington, L. Anne, 'How to Argue with an Algorithm: Lessons from the COMPAS-PROPUBLICA Debate', Colorado Technology Law Journal, Vol. 17, 2018.

Wehba Hannah, 'Access to Algorithms', Fordham Law Review, Vol. 88, 2020.

<국문요약>

인공지능에 기반한 형사법상 의사결정 연구
- 설명요구권과 영업비밀보호 간 균형모색을 중심으로 -

김혜인* / 정종구**

대한민국 정부는 다양한 방면에서 인공지능을 활용하려는 의지를 보이고 있다. 인공지능이 공공 의사결정을 담당하였을 때 생길 수 있는 문제와 이에 대한 제도적 준비를 고민해볼 필요가 있다. 본고에서는 우선 ① 인공지능에 기반한 공공분야의 의사결정이 어떻게 이루어지고 있으며 여기서 제기되는 법적 쟁점이 무엇인지 살펴본다. 다음으로 ② 공공분야 중에서도 실질적으로 인신구속과 직결되어 특히 중요한 형사법에서 범죄예측 및 재범가능성에 대한 의사결정을 토대로 양형에 영향력을 행사했던 인공지능의 예시와 관련판례를 위주로 형사법에서 불거질 수 있는 알고리즘의 편향성 문제를 짚어보고 이러한 편향성에 대해 인공지능을 둘러싼 이해관계자들 어떤 책무성을 부담하는지를 검토한다. 마지막으로 ③ 공적 영역에서 인공지능의 활용을 둘러싼 영업비밀(trade secret) 논의와 의사결정에 의해 영향을 받는 주체의 설명을 요구할 권리(right to explanation)에 대한 논의를 다루고 인공지능의 발전을 저해하지 않으면서도 설명을 요구할 권리가 보장될 수 있는지 살펴본다. 마지막으로 ④ 설명가능한 인공지능(explainable artificial intelligence)을 활용하여 양측의 균형을 도모하기 위한 입법의 가능성을 모색한다.

주제어 : 인공지능, 형사법, 영업비밀, 설명요구권, 설명가능한 인공지능

* 서울대학교 법학전문대학원 박사과정
** 서울대학교 법학전문대학원 박사과정, 교육지원실장 (변호사)

<Abstract>

Criminal Decision-Making based on Artificial Intelligence
: Striking a Balance between Right to Explanation and Trade Secret

Hyein Kim* / Jonggu Jeong**

The Korean government is showing its will to further utilize artificial intelligence in the public sector. Henceforth, it is necessary to consider the problems that may arise when artificial intelligence is in charge of public decision-making and the institutional preparation accrued to them. In this piece of work, how public decision-making based on artificial intelligence is being made and what legal issues are raised would be addressed. After that, the algorithm bias problem surroundings criminal law and accountability issues would be analyzed in depth. Through the aforementioned analysis, this piece of work would contribute to striking a balance between trade secret and right to explanation. Finally, the possibility for future legislation to promote balance between the two sides by applying explainable artificial intelligence would be discussed.

Key Words : Artificial Intelligence, Criminal Law, Trade Secret, Right to Explanation, Explainable AI

논문투고일 2020년 0월 00일
심사완료일 2020년 0월 00일
게재확정일 2020년 0월 00일

* Seoul National University School of Law, Doctoral Candidate
** Seoul National University School of Law, Director of Academic and Student Affairs