DOI :http://dx.doi.org/10.52663/kcsr.2022.27.1.5

인공지능 기반 반부패 도구(AI-ACT)의 헌법원리 적합성에 관한 고찰*

A study on the constitutional suitability of artificial intelligence-based anti-corruption tools(AI-ACT)

조 한 상(Cho, Han Sang)**

ABSTRACT

The core of the basic principles of the Constitution is democracy and constitutionalism. It is reasonable to understand the relationship between the two as a paradoxical relationship of coexistence and a relationship of mutual infection. If anti-corruption is contrary to the basic principles of the Constitution, it can lead to an internal collapse of the basic principles of the Constitution and lead to greater corruption. For this reason, it is necessary to examine whether the so-called AI-ACT is appropriate in relation to the basic principles of the Constitution and whether there is a potential risk of destroying the basic principles of the Constitution.

AI-ACT is a system that selects and analyzes large datasets to detect, predict, and report risks, suspicions, and obvious cases of corruption. AI-ACT can be divided into bottom-up and top-down, the former being the subject of journalists, civil society, and ordinary citizens, and the latter being an anti-corruption tool led by investigative agencies and auditors. The basic structure of AI-ACT can be drawn in three stages: data input, algorithm design, and system adoption decision.

AI-ACT can be a serious threat to democracy and constitutionalism or a slippery slope toward such threats. The bias of data and algorithms can result in hatred, discrimination, and exclusion of certain people or groups, and if it is indiscriminately or selectively distributed through media algorithms, it can be a prey of political polarization and populism. Populism seems democratic in appearance in that it mobilizes the passion of the people, but it promotes totalitarianism and results in destruction of democracy. At the same time, this means the concentration of power and dictatorship, which means a threat to the rule of law. If AI-ACT turns the blade from corrupt public officials to the general public for security and investigation purposes, the prelude to the overall surveillance society, where constitutionalism has collapsed, could open.

The first step in problem solving is securing transparency in AI-ACT. Only when transparency is secured will accountability be guaranteed, bias be corrected, and control be obtained. Legal supplementation is needed, such as establishing basic regulations on the right to intervene and other

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

** 청주대학교 사회과학부 교수, 법학박사

subjects, procedures, and sanctions in the Anti-Corruption Act. Secondly, technical supplementation is needed to seek checks and balances between AI. Third, it emphasized the need for social supplementation to maintain an appropriate critical attitude without falling into the illusion of convenience and objectivity.

Key words: Artificial intelligence-based anti-corruption tools(AI-ACT), anti-corruption, constitutional principles, democracy, rule of law, populism.

Ⅰ. 서론

처음부터 부패인 것은 없을지 모른다. 동서고금을 막론하고 부패는 관행이고, 선심이며, 오고 가는 정으로 손쉽게 합리화되던 것들이었다. 그러한 합리화 속에서 부패라는 암세포[1]는 소리 없이 곳곳으로 전이되어, 사회를 타락시키고, 국가를 몰락시키는 파국적 결과를 초래해 왔다. 그에 대응하여 부패가 '부패'임을 규정하고 이를 막기 위한 노력, 이른바 반부패 활동은 공동체의 생명력을 어떻게든 유지하려 힘써 왔다. 많은 경우 부패로 멸망하기 전에 반부패 장치가 작동하여, 적발과 교정을 통해 문제를 해결해 왔다. 망할 것 같은 나라도 제법 오랜 생명력을 유지하면서 흥망성쇠를 잇는 것은 이 같은 이유 때문이며, 어쩌면 지금 우리의 상황도 크게 다르지 않을지 모른다.

반부패는 인류 문명의 역사와 함께해 왔다. 인류 최초의 문명 중 하나인 수메르에서 기원전 24세기 무렵 만든 점토판에서부터 반부패에 관한 역사적 기록이 발견될 정도다. 서양 문명의 기원으로 믿어지는 고대 아테네 솔론의 민주주의 개혁은 그 자체가 반부패 개혁이라고 해도 지나치지 않다. 각 시대의 반부패 요구에 부응하여 그 시대에 가장 적합한 반부패 수단을 활용해 왔다. 예컨대 귀로 듣고, 눈으로 보고, 문자로 기록하고, 숫자로 검증했다. 때로는 권력의 내부에 반부패 장치를 두기도 했고, 때로는 그 밖에다 두기도 했다.

이른바 4차 산업혁명을 운운하는 21세기 현시점에, 반부패 수단으로서 인공지능의 활용 가능성을 타진하는 것은 어찌 보면 자연스러운 일이다. 대체로 인공지능은 '소프트웨어로 또는 그것과 결합 된 형태로 구현되어 외부환경을 스스로 인식하고 상황을 판단해 자율적으로 동작하는 장치' 정도로 정의할 수 있다.[2] 인간 또는 인간이 일일이 개입하는 프로그램

1) 동양에서는 부패를 바람에 비유하는 일이 많았고, 서양은 암에 비유하는 일이 많았다. 부패는 암이라는 서양의 은유는 부패는 외부가 아닌 내부, 상층이 아닌 하층, 집단이 아닌 개인, 그리고 과도한 욕망에 의해 생겨나는 문제점이라는 사고를 내포한다(김정수, 반부패의 세계사, 가지, 2020, 4면 이하).

2) 김윤명, 인공지능과 리걸 프레임, 10가지 이슈, 커뮤니케이션북스, 2016, 14면; 참고로 지능형 로봇 개발 및 보급 촉진법 제2조 제1호는 "지능형 로봇이란 외부환경을 스스로 인식하고 상황을 판단하여 자율적으로 동작하는 기계장치(기계장치의 작동에 필요한 소프트웨어를 포함한다)를 말한다."로 정의하

대신, AI 그 자체로 부패를 적발하고 해결책을 제시하게 한다면 객관적 데이터에 기반한 부패 척결이 가능할 것이라고 기대할 수 있다.[3)]

확실히 AI 기술의 발달은 눈부시다.[4)] 그러나 사람들은 신기술에 대하여 과도하게 열광하는 경향이 있다. 주지하는 바와 같이 인간을 대체하는 수준의 AI를 강한 AI, 인간을 지원하는 수준의 AI를 약한 AI라고 한다.[5)] 현재 우리가 활용할 수 있는 것은 약한 AI뿐이며, 앞으로도 AI는 생각보다 긴 시간 동안 발생기(nascent)의 상태에 머물 가능성도 크다. 그러나 약한 AI라고 하여 무시할 것은 아니다.[6)] 이하에서 살피는 바와 같이 반부패와 관련해서도 세계 각국에서 AI를 시험적으로 활용하려는 시도가 이어지고 있다.

본 논문은 상상 속 미래가 아닌 현재 시점의 AI 활용 사례와 가능성을 전제로 하며, 반부패 관련 예측, 탐지, 공개를 지원하는, 이른바 AI-ACT(AI-based anti-corruption tools)의 실제를 바탕으로 논의를 진행하고자 한다. 이를 위해 AI-ACT 모델을 상향식과 하향식으로 나누고, 다시 데이터 입력과 알고리즘 디자인, 제도 도입 결정 등의 기본 구조를 소개한다(Ⅲ). 그러나 AI-ACT에는 가능성과 한계가 공존하고 있으며, 가능성을 살리고 한계를 극복하기 위하여 AI-ACT를 헌법에 비추어 보는 작업을 진행하고자 한다. 특히 AI-ACT가 헌법의 가장 핵심적인 기본원리인 민주주의와 법치주의의 검증을 통과할 수 있는지, 문제가 있다면 어떻게 해결되어야 하는지 개략적인 방향을 제시하고자 한다(Ⅳ). AI-ACT와 헌법원리의 관련성 검토는 향후 더 발달된 AI-ACT의 등장에 대한 우리의 접근자세를 가다듬는 데에도 시사점을 줄 것이라고 기대한다.

Ⅱ. 반부패와 헌법의 기본원리

1. 헌법원리로서 법치주의(자유주의)와 민주주의

헌법원리는 "헌법 질서의 전체적 형성에 있어서 그 기초 또는 지주가 되는 원리"[7)] "헌법

고 있다.

3) N. Köbis・C. Starke・I. Rahwan, "Artificial Intelligence as an Anti-Corruption Tool(AI-ACT) - Potentials and Pitfalls for Top-down and Bottom-up Approaches", 2021, p. 5 (https://search-ebscohost-com-ssl.oca.korea.ac.kr/login.aspx?direct=true&db=edsarx&AN=edsarx.2102.11567&lang=ko&site=eds-live&scope=site).

4) 松尾豊, 人工知能は人間を超えるか, 角川, 2015, 18頁.

5) 김윤명, 앞의 책, 16면 이하.

6) 인공지능이 인간과 동등한 수준의 사고를 수행할 수 있는지와 관계없이, 전통적 자본주의 사회의 생산력과 생산관계를 근원적인 수준에서 변모시킬 것으로 보인다는 의견이 있다. 그러나 이 의견은 인공지능에 대한 과도한 기대와 전망을 경계해야 한다고 보고 있다(심우민, "인공지능의 발전과 알고리즘의 규제적 속성", 법과 사회, 제53권, 2016, 42면, 66면 참조).

의 이념적 기초가 되는 것이면서 헌법을 총체적으로 지배하는 지도원리"[8] 등 다양하게 정의된다. 헌법원리 중에서도 헌법의 특정한 영역에 한정되지 않고 헌법의 다양한 영역에서 적용되면서 헌법 전체의 구조와 성격에 커다란 영향을 미치는 것을 헌법의 기본원리라고 말할 수 있다.[9] 비유하자면 헌법이라는 유기체의 줄기세포와 같은 것이 헌법의 기본원리다. 대부분의 헌법 규범은 기본원리로부터 시작된다. 모든 헌법 규범을 기본원리로 완전하게 환원할 수는 없지만, 모든 헌법 규범에서 기본원리의 흔적을 발견할 수 있다.

헌법의 기본원리로 민주주의, 법치주의, 사회국가원리 나아가 국제평화주의, 환경국가원리, 문화국가원리 등 실로 다양한 것이 언급되고 있다. 그러나 이들 사이에도 수준의 차이가 있으며, 기본원리 중 기본원리는 역시 민주주의와 법치주의라고 할 수 있다. 아래와 같이 헌법재판소는 우리 헌법을 근대 입헌적 민주주의의 전통에 서 있는 것으로 보면서 그 핵심에 민주주의와 법치주의가 있다고 밝힌 바 있다.

> (공화주의 이념은) 공민으로서 시민이 가지는 지위를 강조하고 이들에 의해서 자율적으로 이루어지는 공적 의사결정을 중시한다. 따라서 이것은 시민들의 정치적 동등성, 국민주권, 정치적 참여 등의 관념을 내포하고, 우리 헌법상 '민주주의 원리'로 표현되고 있다. 그에 반해 (자유주의 이념은) 국가권력이나 다수의 정치적 횡포로부터 보호받을 수 있는 인권의 우선성을 주장한다. 기본적 인권, 국가권력의 법률 기속, 권력분립 등의 관념들은 자유주의의 요청에 해당하며, 우리 헌법상에는 '법치주의 원리'로 반영되어 있다. … 근대의 입헌적 민주주의 체제는 사회의 공적 자율성에 기한 정치적 의사결정을 추구하는 민주주의 원리와 국가권력이나 다수의 정치적 의사로부터 개인의 권리, 즉 개인의 사적 자율성을 보호해 줄 수 있는 법치주의 원리라는 두 가지 주요한 원리에 따라 구성되고 운영된다.[10]

헌법재판소가 밝힌 바와 같이 법치주의는 근대 초기 소극적 자유를 중시하는 고전적 자유주의의 요청과 관련이 깊은데,[11] 자유주의는 인권의 우선성, 국가권력의 법률 기속, 권력분립 등과 관련이 깊다. 민주주의는 시민들의 정치적 동등성, 국민주권, 정치적 참여와 관련이 깊다. 오늘날 우리는 자유주의와 민주주의 사이의 연계를 당연하게 받아들이는 경향이 있는데, 자유민주적 기본질서라는 합성어가 헌법에 규정되어 있는 것은 그 증거라고 하겠다 (헌법 전문, 헌법 제4조 참조).

그런데 자유주의와 민주주의는 전혀 다른 전통에서 나온 것이며, 칼 슈미트와 같은 학자

7) 계희열, 헌법학(상), 박영사, 2001, 180면.
8) 권영성, 헌법학원론, 법문사, 2010, 125면.
9) 장영수, 헌법학, 홍문사, 2020, 139면.
10) 헌재 2014. 12. 19. 2013헌다1, 판례집 26-2하, 1, 16.
11) 법치국가의 사상은 물론이고 교회와 국가의 분리, 공적 영역과 사적 영역의 구분 등은 자유주의적 정치에서 핵심적인 것이고, 그것들의 원천은 민주주의적 담론이 아니라 상이한 곳으로부터 나왔다. 법치국가의 사상은 물론이고 교회와 국가의 분리, 공적 영역과 사적 영역의 구분 등은 자유주의적 정치에서 핵심적인 것이고, 그것들의 원천은 민주주의적 담론이 아니라 상이한 곳으로부터 나왔다 (C. Mouffe, The Democratic Paradox, VERSO, 2000, p. 2).

는 양자의 조화는 불가능하다고 단언한다. 즉 자유주의는 민주주의를 부정하고 민주주의는 자유주의를 부정하기 때문에 자유민주주의는 생존할 수 없는 체제라는 것이다.[12] 실제로 민주적 열정이 개인의 자유를 위협하거나, 자유주의가 기득권 옹호의 논리가 되어 민주주의를 무력하게 만드는 사례는 어렵지 않게 찾을 수 있으며, 우리 사회에서도 끊임없이 변주되고 있다.[13]

그러나 자유주의와 민주주의의 긴장을 모순이 아닌 역설의 관계로, 타협이 아닌 감염의 관계로 이해해야 한다는 샹탈 무페 같은 학자의 주장에 귀를 기울일 필요가 있다. 양자 사이에는 고도의 구성적 긴장이 존재하는 것은 사실이지만, 자유주의와 민주주의는 서로를 차감하는 제로섬의 관계가 아니다. 하나의 원리가 다른 원리를 긍정적으로 변화시키기도 하고, 하나의 원리가 다른 원리의 존립 조건이 되기도 하는 포지티브섬 관계가 될 수 있다. 자유주의와 민주주의의 역설적 공존이야말로 근대 입헌적 민주주의의 참신함이며,[14] 근대 입헌주의 헌법의 전통에 서 있는 대한민국 헌법 아래에서도 적용되는, 아니 적용되어야 하는 논리다.

2. 부패와 헌법상 기본원리의 관계

부패는 헌법의 기본원리를 위협하고 훼손하는 존재다.[15] 부패가 일상화되는 경우 공직자들은 국민의 정당한 위임을 배반하고, 국가를 자신의 소유물 또는 물권적 지배의 대상으로 전락시킴으로써 민주주의를 형해화할 것이다. 민주주의의 한계를 벗어난 공직자들은 경제·사회 각 영역에서 국민의 기본권을 보장하기는커녕 위협하는 존재로 드러날 것이며, 이것은 곧바로 법치주의에 대한 위협이 된다.[16] 이처럼 부패가 헌법의 기본원리를 위협한다면, 반부패는 헌법의 기본원리 보호에 도움이 된다는 말도 성립할 것이다.

부패와 민주주의 그리고 법치주의와의 밀접한 관계는 반부패와 관련된 인류 역사의 초기부터 찾을 수 있다는 주장에 귀를 기울일 필요가 있다. 인류 최초의 문명이라고 할 수 있는

12) C. Schmitt(trans. by E. Kennedy), The Crisis of Parliamentary Democracy, MIT Press, 1985, pp. 12-13.

13) Y. Mounk(함규진 역), 위험한 민주주의, 와이즈베리, 2018, 131면 이하 참조.

14) 무페는 “근대 민주주의의 참신함, 즉 근대 민주주의를 근대적인 것으로 만드는 새로운 점은 민주적 혁명과 함께 권력은 인민에 의해 행사되어야 한다는 고대의 원칙이 다시 부상하였다는 것이고, 그러한 고대의 원칙이 이번에는 개인적 자유와 인권을 강조하는 자유주의 담론에 의해 규정되는 상징적 틀 안에서 가능해졌다는 점이다.”라고 말한다(C. Mouffe, op. cit., p. 2).

15) 강경근, “공직자 부패와 헌법이론”, 한국부패학회보 제1권 제1호, 1997, 103면; 김병록, “공직부패의 헌법적 조명”, 공법연구 제30권 제2호, 2001, 143면.

16) 특히 선거와 관련하여 국가기관이나 공직자들의 부패와 일탈이 민주주의를 위협했던 사례라던가, 부패한 수사기관이나 공기업이 신체의 자유, 재산권 등의 기본권을 침해했던 사례는 지금도 기억이 생생하거나 혹은 여전히 진행 중이다.

수메르의 기원전 24세기경 군주인 우루카기나는 “주민들을 고리대와 과도한 통제, 굶주림, 도적질, 살인, 재산과 인간에 대한 부당한 갈취로부터 해방시켰다. 그는 자유를 확립했다. 과부와 고아는 더 이상 힘 있는 자의 자비에 기댈 필요가 없었다.”라는 기록이 점토판 쐐기문자로 남아 있다. 부패와 관련된 인류의 첫 번째 기록은 부패를 저지른 것이 아닌 부패와 싸운 기록인 것이다.[17] 이 사례를 지극히 근대적 개념이라고 할 수 있는 법치주의와 연결짓는 것은 무리일 수 있지만, 법치주의의 핵심이 개인의 자유 보장이라는 점을 감안하면 우루카기나의 반부패 개혁과 법치주의의 관련성을 추단하는 것은 불가능하지 않다.

서양 문명의 본향으로 인식되는 고대 그리스, 그중에서도 도시국가 아테네에 민주주의를 가져온 위대한 개혁가로서 기원전 6세기경의 솔론이 있다. 당시 아테네는 귀족들이 제멋대로 권력을 사용해 그야말로 무정부 상태에 가까웠다. 솔론은 사람들의 참여를 통해 혼란을 극복하고 세상을 변화시킬 수 있다고 믿었다. 솔론의 개혁은 시민을 재산에 따라 네 계급으로 나누고, 이들에게 귀족의 권력을 나누어주는 것에서 시작했다. 특히 최하층 계급인 테테스가 참여할 수 있는 시민법정 헬리아이아를 창설했는데,[18] 이로써 뇌물을 주고 판결을 뒤집는 불공정 관행이 교정되기 시작했다. 솔론의 민주주의 개혁은 곧 반부패 개혁이었다.[19]

그런데 오늘날 통용되는 부패개념은 반부패와 헌법의 기본원리 사이의 밀접한 관계를 반영하기에는 지나치게 협소하다고 지적할 수 있다. 부패는 “위임된 권력의 사적 남용”이라는 것이 일반적인 정의이며, 우리나라 「부패방지 및 국민권익위원회의 설치와 운영에 관한 법률」 제2조 역시 기본적으로 이러한 정의를 기본으로 하고 있다.[20] 그러나 이와 같은 개념정의는 17-18세기 유럽에서 생성되어 20세기 중후반 미국을 중심으로 발전해 온 개념 정의에 불과하다는 지적이 있다.[21]

이 개념 정의에 따르면 공적 권한과 사적 영향력이 혼재된 영역에서, 뇌물이나 횡령 등으로 이익을 취했다는 것이 명확하게 밝혀지지 않은 경우는 부패라고 말하기 어려운 한계가 생기며, 국민의 위임을 저버리고 국민의 자유를 짓밟는 다양한 행위가 정작 부패에 해당하

17) 김정수, 앞의 책, 66면 이하.

18) 플루타르크 영웅전에서 이 시민 법정이 "처음에는 아무것도 아닌 것처럼 보였으나 나중에 엄청난 특권을 가지고 있었던 것으로 증명"되었다고 적을 정도였다(김정수, 앞의 책. 95면).

19) 김정수, 앞의 책, 84면 이하.

20) 부패방지 및 국민권익위원회의 설치와 운영에 관한 법률 제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.
4. “부패행위”란 다음 각 목의 어느 하나에 해당하는 행위를 말한다.
가. 공직자가 직무와 관련하여 그 지위 또는 권한을 남용하거나 법령을 위반하여 자기 또는 제3자의 이익을 도모하는 행위
나. 공공기관의 예산사용, 공공기관 재산의 취득·관리·처분 또는 공공기관을 당사자로 하는 계약 의 체결 및 그 이행에 있어서 법령에 위반하여 공공기관에 대하여 재산상 손해를 가하는 행위
다. 가목과 나목에 따른 행위나 그 은폐를 강요, 권고, 제의, 유인하는 행위

21) 김정수, 앞의 책, 348면 참조.

지 않는 일이 벌어진다. 이는 일반적 상식에도 맞지 않을뿐더러, 역사적 사례와도 부합하지 않는다. 생각건대 “위임된 권력의 사적 남용”이라는 협의의 부패개념을 유지하되, “헌법의 기본원리를 위협하거나 훼손하는 공직자의 권력적 일탈·남용행위”라는 식의 보다 넓은 개념정의도 함께 사용되어야 한다고 본다.[22]

3. 반부패 정책의 헌법원리 적합성

부패가 실제로 있었는지 아닌지와 관계없이 반부패가 권력투쟁의 수단으로 사용되었던 역사적 사례가 적지 않다. 대표적으로 영국 엘리자베스 여왕 시절 왕권의 군림을 반대하던 대법관 에드워드 코크를 공금횡령 혐의를 씌워 파면시키고, 이후 코크가 자신의 후임 대법관인 프랜시스 베이컨을 뇌물 혐의로 고발하여 그또한 직을 물러나게 한 사례가 유명하다.[23] 현재 국내외에 제2, 제3의 코크와 베이컨 사례가 없다고 장담할 수 있는 사람은 없을 것이다.

반부패는 무고한 개인을 범죄자로 몰아 처벌하고, 통제되지 않는 권력에게 독주의 빌미를 제공할 수 있다. 결국 반부패는 양날의 칼이 될 수 있다. 반부패가 부패하면 더 큰 부패를 초래할 수 있는 것이고, 이것은 헌법상 기본원리의 내부적 붕괴를 의미한다고도 볼 수 있다. 따라서 구체적인 반부패 시스템은 일정한 검증을 통과해야 하는데, 가장 기본적인 검증기준은 역시 헌법의 기본원리, 즉 민주주의와 법치주의가 제공해야 한다. 다시 말해서 민주주의와 법치주의에 적합한 부패방지 제도여야 한다.

관련된 검증기준을 다음과 같이 네 가지의 질문으로 정리해 볼 수 있다. 첫 번째, 해당 제도는 민주주의 원리에 적합하게 만들어지고 운영되고 있는가? 두 번째, 해당 제도는 민주주의를 파괴할 잠재적 위험성을 가지고 있지 않은가? 세 번째, 해당 제도는 법치주의에 적합하게 만들어지고 운영되고 있는가? 특히 국민의 기본권을 침해할 위험성은 없는가? 네 번째, 해당 제도는 국가권력을 과도하게 확대하여 법치주의를 훼손할 잠재적 위험성은 없는가? 물론 민주주의와 법치주의의 역설적 공존관계를 생각하면 위의 질문도 완벽히 분리된 것이 아니라, 서로 중첩될 수 있음에 유의해야 한다.

본 논문이 주안점을 두고 있는 것은 궁극적으로 AI를 활용한 새로운 반부패 수단이 이러한 검증기준을 통과할 수 있는지, 혹시 미진하거나 우려되는 점이 있다면 그 보완을 위한 조건과 과제는 무엇인지다. 이에 대한 본격적인 검토에 앞서 AI를 활용한 반부패 수단에 대

22) 민주법치국가와의 합치성과 관련하여 부패를 이해하는 견해로는 강경근, 앞의 논문, 104면 참조; 공익에 반하거나 여론에 반하는 공무원의 행위를 부패로 보는 경향이 강해지고 있으며, 이익이 아닌 보편적 가치나 덕목 위반 행위까지 포함할 정도로 부패개념이 확장되고 있음을 지적하는 견해로 김진영, “부패개념에 대한 고찰”, 한국부패학회보 제22권 제4호, 2017, 96면.

23) 김정수, 앞의 책, 50면 이하.

하여 조금 더 자세히 검토할 필요가 있다.

Ⅲ. AI-ACT의 기본구조

1. AI-ACT의 의미와 유형

서론에서 밝힌 바와 같이 AI-ACT는 이른바 약한 AI를 기반으로 만들어진 부패 방지 도구이며, 대규모 데이터 세트를 선별하고, 이를 분석하여 부패의 위험, 의심, 명백한 사례 등을 탐지, 예측 및 보고하는 시스템이다. 반부패 활동에 AI를 활용함으로써, 데이터 분석, 모델링, 동적 시각화에 있어 자원을 절약하고,[24] 객관성을 높일 수 있다.[25] 이러한 시스템은 사전에 부패를 저지르기 어려운 환경을 조성하는 데에도 이바지할 수 있는데, 특히 국제 원조 기구의 보조금 집행 투명성 확보를 위한 장치로서 활용이 늘고 있다.[26] 이에 AI-ACT는 "부패 방지의 차세대 개척지"로 일컬어지며 많은 주목을 받고 있다.[27]

AI-ACT 활용이 가능해진 배경에는 공공행정의 디지털화와 공개성 강화가 있다. 많은 나라는 행정기록을 전산화하여 개방형 데이터 프로그램을 만들고 전자정부를 구축하는 등 정부의 투명성을 확보하기 위해 힘쓰고 있다. 그러나 단지 데이터를 공개하는 것만으로 부패를 억제하기에 충분하지 않다. 특히 축적된 정보가 방대한 로데이터(raw data)의 집합이거나, 사진이나 음성 등의 이른바 빅데이터에 해당할 경우가 많아지고 있는데, 이러한 자료를 바탕으로 부패의 흔적을 찾아내기가 쉽지 않다. 누군가는 자료로부터 추론을 끌어내고 반부패 활동을 위한 실천 가능한 자료로 만들어야 하는데,[28] 이때 효과를 발휘할 수 있는 것이 AI이다. 요컨대 데이터 축적의 전산화・자동화는 AI-ACT를 가능하게 한 배경인 동시에, AI-ACT를 불러낸 배경이다.[29]

---

24) 예컨대 멕시코 국세청이 AI 알고리즘과 분석 툴을 활용해 납세자와 기업 간 부정행위를 적발하는 프로젝트를 시범 운영했는데, 3개월 이내에 1,200개의 부정거래 업체를 적발하고, 3,500건의 부정거래가 확인되었다. 이것은 AI를 사용하지 않았다면 대략 18개월의 작업이 걸렸을 것이라고 한다(P. Aarvik, "Artificial Intelligence-a promising anti-corruption tool in development settings?" U4 Report 2019-1, 2019, p. 4)

25) A. Petheram・I. N. Asare, "From Open Data to Artificial Intelligence: The next Frontier in Anti-Corruption", Oxford Insights, 2018(https://www.oxfordinsights.com/insights/aiforanticorruption)

26) 부패는 외부 환경적 요인에 기인하는 경우가 많으며(정성범・백윤철, "헌법상 행정부패에 관한 연구", 헌법학연구, 2006, 297면), 부패를 근본적으로 줄이기 위해서는 정부, 정치권, 기업 등 부패의 커넥션에 연결된 자들에 대한 총체적 접근이 필요하다(김정수, 앞의 책, 38면).

27) N. Köbis・C. Starke・I. Rahwan, op. cit., p. 4.

28) 적절히 처리되지 않은 데이터는 "한 손으로 치는 손뼉"과 같다(D. Seligsohn・M. Liu・B. Zhang, "The sound of one hand clapping: transparency without accountability", Environmental politics, 27[5], 2018, p. 804).

AI-ACT는 고전적인 반부패를 위해 활용되었던 고전적 정보통신기술(ICT)과는 다르다. “고전적” ICT는 조달 절차의 디지털화를 가능하게 하고, 온라인으로 공공 서비스를 제공하며, 개방형 정부 데이터를 공개한다. 고전적 ICT는 자율적으로 작동하지 않는 반면, AI-ACT는 자율성을 가지고 특정 목표를 달성하기 위해 환경을 분석하고 조치함이 원칙이다. 다만 아래에서 살피는 바와 같이 AI-ACT에 인간의 개입이 완전히 배제되는 것은 아니다.

AI-ACT는 여타의 범죄 방지 AI와 다르다. 이른바 사회악에 대응하기 위한 AI의 활용이라는 점에서 공통적이지만, AI-ACT는 정부를 비롯한 국가권력 내의 권한 남용을 해결하기 위한 도구다. 여타의 범죄 방지를 위한 AI는 국민을 정부가 감시하는 빅브라더 역할을 하는 것이라면,[30] AI-ACT는 정부를 견제하기 위한 감시자(watchdogs)의 역할을 하는 것이다. 이러한 감시자로서 성격은 이른바 상향식 AI-ACT의 경우 두드러지게 나타난다.

크게 볼 때 AI-ACT는 상향식과 하향식으로 나눌 수 있다.[31] 상향식은 언론인, 블로거, 시민사회 활동가 그리고 일반 시민이 주체가 되는, 하향식은 수사기관이나 감사 담당자 등이 주체가 되는 반부패 도구라고 할 수 있다. 현재 활용되고 있는 상향식 AI-ACT의 사례로는 국제투명성기구가 공공 자금 사용을 감시하기 위해 개발한 우크라이나의 DOZORRO,[32] 역시 공공지출을 감시하기 위해 8명의 젊은이가 만든 브라질의 Rosie da Serenata,[33] 트럭 운전사들과 소규모 무역상들의 참여로 교통 검문소 뇌물 갈취를 적발하는 나이지리아의 TRIMS(Trade Route Incident Mapping System),[34] 크라우드소싱 포털을 이용하여 뇌물 사례를 신고할 수 있는 인도의 Ipaidabribe.com[35] 등을 생각할 수 있다. 하향식 AI-ACT의 사례로는 브라질 감사원(CGU)의 개인 부패 점수 계산 시스템 MARA,[36] 도로

29) N. Köbis・C. Starke・I. Rahwan, op. cit., p. 5.

30) 관련된 우려에 대한 상세한 내용은 김형섭・황선영, “AI기술의 부패방지와 인권 침해의 논의 - 홍콩 사례(복면금지법)를 중심으로”, 한국부패학회보 제25권 제2호, 2020, 2면.

31) N. Köbis・C. Starke・I. Rahwan, op. cit., p. 6.

32) https://oecd-opsi.org/innovations/dozorro/; 우크라이나 감사원은 어떤 입찰자가 면밀한 검사가 필요한지를 평가하는 데 도움이 되는 35가지 위험지표를 개발했으나, 이 같은 지표가 알려지자마자 부정 입찰자들이 이 같은 사실에 적응한 뒤 감사 회피를 위한 조치를 한 것으로 드러났다. 따라서 국제투명성기구는 머신러닝을 기반으로 자체 소프트웨어 ‘도조로’를 출시하고 부패 위험이 높은 입찰자를 식별하도록 훈련했다(P. Aarvik, op. cit., p. 7).

33) https://brasil.elpais.com/brasil/2017/01/23/politica/1485199109_260961.html.

34) 현재 트림의 웹사이트는 접속할 수 없다(http://www.trimsonline.org/). 구체적 성과가 없으면, 참여는 종종 흐지부지되고, 많은 경우 완전히 사라진다. 크라우드소싱 접근이 초기 동원 국면 이후 사람들의 참여를 계속 유지하기는 쉽지 않다.

35) 이 웹사이트에서는 시민들이 뇌물을 받은 사례(강요)를 신고할 수 있는데, 이 중 최대 19만 7,000여 건을 집계했다고 한다(N. Köbis・C. Starke・I. Rahwan, op. cit., p. 10).

36) 브라질 감사원은 공무원들 사이의 부패 행동의 위험을 추정하기 위해 기계 학습 애플리케이션을 개발했는데, 이 앱은 대시보드에 공무원 사회보장번호를 입력하면 그 사람이 부패했을 확률을 간단한 게이지에 표시해 준다. 이 알고리즘을 훈련하는 데에는 공무원의 유죄판결 데이터가 사용되었다고 한다(P. Aarvik, op. cit., p. 7).

건설 자재의 품질을 평가하고 잠재적인 횡령 사례를 식별하기 위한 필리핀의 KALSADA[37] 등을 들 수 있다.

Artificial Intelligence-based Anti-Corruption Tools (AI-ACT)

	Top-Down Approaches	Bottom-Up Approaches
Actors	Criminal investigators, prosecutors, compliance officers, auditors	Journalists, bloggers, civil-society activists, average citizens
Input data	Classified & open government data, crowdsourced data, (social) media text	Open government data, data leaks, crowdsourced data, (social) media text
Algorithmic design	Rather minimize false-negative rate	Rather minimize false-positive rate
Institutional Implementation	Human-out-of-the-loop to escape the corruption trap	Human-in-the-loop to ensure legitimacy
Examples	MARA, Kalsada	Rosie da Serenata, Botivist, Dozorro

〈그림〉 AI-ACT의 구조[38]

2. 데이터 입력

AI-ACT에서 사용 가능한 데이터 출처에는 크게 네 가지 유형이 있다고 하겠다. 첫째, 정부가 수집하여 관리하는 정부 데이터가 있다. 두 번째, 해커나 내부고발자가 폭로한 유출된 데이터(data leaks)가 있다. 세 번째, 부패를 폭로하려는 시민들의 노력으로 만들어지는 크라우드소싱 데이터가 있다. 네 번째로는 언론사의 미디어 텍스트와 소셜 미디어의 텍스트가 있다.[39] 자연어 처리(Natural Language Processing)의 발전에 따른 대규모 분석의 가능성은 미디어 텍스트를 유의미한 데이터로 활용할 수 있는 배경이 되고 있다.[40]

하향식 AI-ACT는 정부 데이터 중 비공개 데이터에 대한 접근성도 확보하는 경우가 있다. 실제로 브라질 감사원(CGU)의 브라질 연구 및 전략 정보부(DIE)가 추진한 MARA는 비공개 정부 데이터에 대한 머신러닝 알고리즘을 학습시켜 개인 수준의 부패 점수를 계산하였다.[41] 그러나 상향식 AI-ACT는 비공개 정보에 접근하기 어려우므로 유출된 데이터나

37) https://datapopalliance.org/1wl-28-data-and-anti-corruption.
38) N. Köbis · C. Starke · I. Rahwan, op. cit., p. 9.
39) 예컨대 스페인에서는 부패에 대한 고전적인 뉴스 미디어 보도를 활용하여 지역 수준에서 향후 부패가 발생할 것을 예측하도록 했다(F. J. López-Iturriaga · I. P. Sanz, "Social Indicators Research", Dordrecht, Vol. 140, Iss. 3, 2018, p. 981).
40) 참고로 청와대 국민청원의 내용을 Web상을 돌아다니면서 정보를 수집하는 행위인 웹크롤링(web crawling)을 통해 수집하고 자연어처리(NLP: Natural Language Processing)를 통해 분석한 연구가 있다(송준모 · 박영득, "청와대 국민청원에서는 무엇이 일어나는가?: 자연어 처리를 활용한 청와대 국민청원 분석", 한국정치학회보, 제53권 제5호, 2019, 53-78면).
41) R. Carvalho · M. Ladeira · F. M. Monteiro · G. L. d. O. Mendes, "Using Political Party Affiliation Data to Measure Civil Servants′ Risk of Corruption", 2014 Brazilian Conference on Intelligent

크라우드소싱을 통해 획득한 데이터를 활용해야 할 때가 상대적으로 많을 것이다.

최근 들어 이른바 데이터 흔적(data traces)의 활용 가능성도 논의되고 있다. 디지털 기술이 사람들의 일상생활에 녹아들면서 사람들은 인터넷 플랫폼, 스마트폰, 앱, 센서, 기타 기기들에 디지털 흔적을 남기고 있다.[42] 예컨대 시민과 공무원 사이의 모든 상호작용을 문서화 하여 권력 보유자에게 책임을 묻는 것을 목표로 하는 앱은 이미 존재한다. '시리, 나 끌려가고 있어(Siri, I'm being pulled over)'[43]의 경우 경찰관과의 만남이 있으면 앱이 위치정보 등을 자동으로 기록하여, 경찰관의 권력 남용을 통제・예방하려 한다. 데이터 흔적을 광범위하게 활용할 수 있게 된다면, AI-ACT의 효과성과 효율성은 상당 수준 높아질 것이다.

AI-ACT 입력 데이터에는 많은 문제가 남아 있는데, 그 첫 번째는 양적 문제다. 언급한 것처럼 공공행정의 전산화가 진행되고 있지만, 여전히 풍부하고 신뢰할 수 있는 데이터가 부족하며, 특히 부패가 심각한 개발도상국의 경우 상황은 더 심각하다. 데이터가 전산화되어 있다고 하더라도 언어, 체계가 제각각이고 표준화되어 있지 않아 활용도를 떨어뜨리고 있다.[44]

두 번째는 질적 문제다. "쓰레기가 유입되면, 쓰레기가 배출된다."라는 표현은 AI-ACT에도 그대로 적용된다. AI-ACT는 입력 데이터만큼만 좋은 것이며, 따라서 입력되는 데이터의 품질이 높은 수준으로 유지되어야 한다. 이를 위해 부패에 대해 올바른 프록시가 설정되어 타당성을 유지해야 하며, 부패 여부에 관한 판단이 일관되어 신뢰성을 유지해야 한다. 크라우드소싱을 통해 획득된 데이터나 얼굴 이미지와 같은 데이터는 신뢰성 확보에 태생적 한계를 갖게 되므로 각별한 주의가 필요하다.

데이터 편향성도 유의해야 하는 숙제다. 편향된 데이터 세트는 AI-ACT 안에 기존의 사회적 편견을 재현하고, 더욱 악화시킨다. 미국 마이크로소프트의 채팅 로봇 테이나, 우리나라 챗봇 '이루다' 사례에서 입력 데이터의 편향성이 어떤 결과를 초래하는지 알 수 있었다.[45] 관련하여 앞서 언급한 브라질의 MARA는 유죄 판결 데이터에 의해 알고리즘을 훈련시켰는데,[46] 만약 소수자 집단이 더 자주 기소되고 처벌되는 등 학습에 사용되는 데이터가

Systems, 2014, p. 169.

42) 디지털 데이터 추적에는 자체 추적 장치, 소셜 미디어 통신, 지리 공간 데이터, 브라우저 기록이 포함되며, 행동이 언제 어디서 어떻게 발생하는지에 대한 상황별 데이터가 포함된다(A. Rafaeli・S. Ashtar・D. Altman, "Digital Traces: New Data, Resources, and Tools for Psychological-Science Research", Current directions in psychological science, 28[6], 2019, pp. 560 - 566).

43) https://www.theverge.com/2020/6/17/21293996/siri-iphone-shortcut-pulled-over-police-starts-recording-video.

44) P. Aarvik, op. cit., p. 15.

45) 조선비즈 2021년 1월 12일자(https://biz.chosun.com/site/data/html_dir/2021/01/12/2021011202089.html); 이승택, "뉴미디어 시대의 알고리즘과 민주적 의사형성", 법학논총 제33권 제3호, 2021, 565면.

46) N. Köbis・C. Starke・I. Rahwan, op. cit., p. 12.

편견에 사로잡혀 있으면, 예측 결과도 왜곡되어 이러한 편견을 재현하게 될 것이다.

3. 알고리즘 디자인

알고리즘 설계에서 가장 먼저 고려해야 할 요소는 예측의 정확성이다. 잘못된 긍정 오류는 무고한 개인을 "부패" 혐의자로 만들 수 있으며, 부당하게 혐의를 쓴 사람들은 적지 않은 대가를 치르게 된다. 반대로 거짓 부정 오류는 실제 부패 사건을 방치하는 결과를 낳는데, 이 역시 공공기관이나 사회 전체에 무시할 수 없는 비용이라고 할 수 있다.

주목할 것은 거짓 양성 오류와 거짓 음성 오류 사이의 이른바 맞교환 관계(trade-off)가 형성된다는 점이다.[47] 다시 말해 한 유형의 오류가 감소하면 다른 유형의 오류가 증가한다. 예측 정확도가 높은 알고리즘이 요구되는 것은 당연하지만, 완벽한 예측 알고리즘은 - 현재도, 어쩌면 앞으로 영원히 - 가능하지 않다. 따라서 AI-ACT 알고리즘을 설계하면서 어떤 변수를 우선시하고 어떤 오류를 줄일 것인지에 대한 까다로운 절충이 수반되는 것은 불가피하다.

이러한 절충은 하향식과 상향식 사이에 다소 다르게 이루어질 수 있다. 하향식의 경우 AI-ACT가 내린 결정은 대부분 정부 또는 공공기관 내부의 관계자에게 전달된다. 관계자는 AI-ACT가 제시하는 부패 의혹의 타당성을 다시 확인하는 과정을 거치게 되고, 다음에 어떤 조치를 할 것인지 결정하게 된다. 하향식의 경우 거짓 양성 오류를 바로잡을 기회가 있는 것이므로, 거짓 양성보다 거짓 음성이 최소화되도록 알고리즘을 설계함이 바람직하다. 다시 말해 가능한 많은 부패 의혹이 보고되도록 설계할 필요가 있으며, 부패 의혹이 다소 과도하게 보고되는 것도 수인할 수 있다.[48]

그러나 상향식 접근법의 경우 거짓 양성을 바로잡을 기회는 적으며, 미디어를 통해 전격 공개되는 일이 많으므로 그 피해 또한 작지 않다. 나중에 부패 의혹이 거짓으로 해명된 이후에도 이미 여론의 법정에 기소된 인사들의 명예를 온전하게 회복하는 것이 어려운 때가 많다. 따라서 상향식 AI-ACT의 알고리즘을 설계할 때는 거짓 양성 오류를 줄이는 것이 거짓 음성 오류를 줄이는 것보다 우선순위를 차지한다. 하향식과 비교해 상향식 알고리즘은 다소 관대하게 설계되어야 한다는 것이다.[49]

한편 대규모 데이터를 처리하는 정교한 알고리즘은 쉽게 이해할 수 없는 일종의 블랙박스와 같은 존재라고 할 수 있다. 게다가 머신러닝, 딥러닝 등 자율적 학습과 판단이 거듭되면 최초 설계 의도와는 전혀 다른 알고리즘이 생성될 수 있다.[50] 이때 인간은 왜 인공지능

47) N. Köbis · C. Starke · I. Rahwan, op. cit., p. 13.
48) N. Köbis · C. Starke · I. Rahwan, op. cit., p. 14.
49) N. Köbis · C. Starke · I. Rahwan, op. cit., p. 15.
50) 심우민, 앞의 논문, 58면.

의 결정을 따라야 하는지 모르는 채로 따라야 하거나, 이해할 수 없는 이유로 인공지능에 의해 자신의 정체성이 재단되는 경험을 하게 된다.[51]

그래서 알고리즘의 투명성을 확보하여 시민의 신뢰를 얻어야 한다는 목소리가 높다. 그러나 한편으로는 기술적 복잡성으로 인해, 다른 한편으로는 관련 기업의 지식재산권과 영업비밀 보호의 필요성 때문에 달성하기가 쉽지 않은 문제이다. 유의할 점은 투명성은 설명 가능성을 전제로 하며, 설명 가능성은 알고리즘의 논리적 단순성에 비례한다는 사실이다.[52] 하지만 높은 설명 가능성을 유지할 정도로 알고리즘이 단순할수록, 권력자들의 교묘한 조작에도 취약할 수 있다. 요컨대 알고리즘의 투명성과 공정성도 맞교환 관계(trade-off)를 형성할 수 있다.

4. 제도 채택

AI-ACT를 채택한다면 부패 방지에 있어 다양한 난관을 극복하고 자원을 절약할 수 있다. 그러나 AI-ACT의 채택이 언제나 긍정적 결과만을 가져오는 것은 아니며, 때로는 얻을 수 있는 이익에 비하여 발생하는 부작용이 더 커질 수도 있다.

하향식 AI-ACT의 알고리즘의 투명성과 설명 가능성이 제한적이라면 통제의 대상이 되는 공무원들은 과도한 감시를 받는다고 느끼게 되며, 신기술에 대한 불신이 쌓이고, 결국 역풍을 맞을 수 있다. 이른바 복지부동 현상이 일어날 수 있으며, 재능 있는 공무원의 이탈로 이어질 수도 있다. 심각한 경우 이러한 부작용은 부패혐의자를 적발하여 얻는 이익보다 더 많은 해악을 끼칠 수 있다.[53]

또한 신중하지 못한 AI-ACT의 채택은 상향식에서도 역효과를 초래할 수 있다. AI-ACT가 빈번하게 부정확한 부패 사례를 적발하면 일종의 “스패밍(spamming)” 효과가 발생하여, 시민들이 부패 문제에 둔감해지는 효과를 가져올 수 있다. AI-ACT가 정확하게 적발해도 적발 건수가 너무 많거나 AI-ACT와 사법 체계 등의 연계가 적절치 않아 제대로 처벌되지 않는 사례가 늘어날 경우 문제가 발생할 수 있다. 즉 부패로 만연된 세상을 바꿀 방법이 없다는 자포자기와 냉소주의가 발생할 수 있으며,[54] 심지어 시민들 자신의 잘못에 대한 면죄부로 여길 가능성도 있다.[55]

AI-ACT의 알고리즘 설계가 사회적 맥락 안에 존재하는 것처럼, AI-ACT를 채택할 것

51) 김희정, “알고리즘 자동의사결정으로부터 개인의 보호 - 인간의 개입권(the right to human intervention)을 중심으로”, 헌법학연구 제26권 제1호, 2020, 225면.
52) 사용 소스 코드를 밝혀야 한다는 주장, 설명 가능한 자동의사결정시스템을 사용하라는 주장, 추상적인 정보로 충분하다는 주장이 대립하고 있다(김희정, 앞의 논문, 221면 참조).
53) N. Köbis・C. Starke・I. Rahwan, op. cit., p. 14.
54) 이것은 반부패 일반에서 나타날 수 있는 문제다(정성범・백윤철, 앞의 논문, 289면).
55) N. Köbis・C. Starke・I. Rahwan, op. cit., p. 14.

인지 아닌지도 사회적 선택의 대상이다. 특히 AI-ACT가 부패 척결에 실질적인 도움을 주므로 선택하는 것인지, 아니면 AI 기술의 신선함과 이에 대한 호기심으로 선택하는 것인지 비판적으로 검토할 필요가 있다. 때로는 기존에 사용하던 정적 ICT 시스템을 사용하거나 단순한 회귀 예측 모델 정도를 채택하는 것이 도움이 될 수도 있다.[56] 신중한 검토를 통해 AI-ACT 채택 여부가 결정되어야 한다.

Ⅳ. AI-ACT의 헌법원리 적합성

1. 숨겨진 인간 행위자

AI-ACT는 시민참여의 가능성을 넓힐 수 있으며, 따라서 부패 방지에 있어 전례 없는 민주화를 촉진하는 역할을 하는 것으로 보이기도 한다.[57] 특히 상향식 AI-ACT가 디지털 크라우드소싱을 활용하여 부패 사례를 직접 보고할 때 이러한 인상이 강화된다. 그러나 AI-ACT가 반드시 민주주의에 적합하다고 단언하기는 어려운데, 무엇보다 AI-ACT 이면의 인간 행위자 문제 때문이다.

AI-ACT는 단순히 자료를 모으는 것이 아니라, 자율적 의사결정 알고리즘을 가진다. 이러한 상황은 AI-ACT가 인간의 주관을 배제한 객관적이고 공명정대한 부패 적발을 실행한다는 환상을 심어주기 쉽다. 그러나 한편으로는 현재의 기술 수준이 충분하지 않기 때문에, 다른 한편은 인공지능의 자율적 결정, 특히 윤리적 결정을 내리는 알고리즘에 대한 시민들의 거부감 때문에 인간이 개입한다.[58] 예컨대 AI-ACT가 추론한 부패 혐의를 인간 행위자가 최종확정하거나, 그에 대한 거부권을 갖는 경우를 생각할 수 있다. AI-ACT를 훈련 시키는 과정은 인간에 의해 주도될 수밖에 없다는 점에서도 인간의 개입은 애초에 불가피하다.[59]

민주주의 원리에 따라 모든 공적 작용은 직·간접적 민주적 정당성을 확보해야 하며, 책임성을 담보할 수 있어야 한다. 그러나 AI-ACT는 그 이면의 드러나지 않는 인간 행위자, 특히 기술 권위주의 엘리트를 인식하지 못하게 할 위험이 있다. 더욱 심각한 것은 AI-ACT의 설계와 운영을 거대 기업이 주도할 가능성이 크다는 점인데, 실제로 IBM, 구글, 마이크로소프트와 같은 다국적 빅테크 기업의 이 분야 진출이 두드러지고 있다.[60] 일부 의견은 이

56) N. Köbis · C. Starke · I. Rahwan, op. cit., p. 14.
57) N. Köbis · C. Starke · I. Rahwan, op. cit., p. 22.
58) 인공지능 스스로가 윤리적 판단이 내포된 법규범적 추론을 할 때 어느 정도의 사회적 신뢰를 부여해야 하는지 관건이 된다는 의견으로서 심우민, 앞의 논문, 50면.
59) P. Aarvik, op. cit., p. 29.
60) IBM의 AI Fairity 360 개발, 구글의 보조금 관리 시스템 개발, 마이크로소프트의 지구를 위한 AI, 인

를 '정보자본주의'라고 칭하며 우려하고 있으며, 나아가 미국 국가보안기구와의 긴밀한 관계를 우려하기까지 한다.[61]

요컨대 AI-ACT는 공적 자율성에 기한 정치적 의사결정을 기본으로 하는 민주주의에 부적합할 가능성을 내재하고 있다. 나아가 이면의 인간 개입자가 국민의 신임을 저버리고 통제와 책임 없이 권력을 행사할 때 국민의 권리에 심각한 위협 요인이 될 수 있다. 민주주의에 부적합할 수 있다는 것은 기본권 보장을 핵심으로 하는 법치주의에도 부적합할 가능성이 크다는 의미가 된다.

2. 기본권 침해 가능성

AI-ACT가 완전히 공개된 데이터가 아닌, 제한된 또는 비자발적으로 공개된 데이터를 사용한다면 헌법 제17조에 근거한 사생활의 비밀 또는 프라이버시권 침해가 문제 된다. AI-ACT가 부패 혐의 도출에 사용한 비공개 데이터를 함께 공표할 때 문제는 더욱 심각해질 수 있다. 하향식 AI-ACT의 경우 관리자가 개인 정보가 과도하게 유출되지 않도록 주의할 것이며, 상향식 AI-ACT에서도 책임 있는 저널리스트들은 마찬가지의 태도를 가질 것이다. 그러나 상향식 AI-ACT가 위키리크스 등의 유출된 데이터를 바탕으로 부패 혐의를 도출하고, 그것을 자동적 또는 인위적으로 폭로하는 형태라면 프라이버시권에 대한 극심한 침해를 초래할 수 있다.

만약 AI-ACT 부패혐의자를 부정확하게 지목하게 되면 해당 공직자는 인격권, 명예권을 침해받을 수 있다.[62] 인간이 만들고 운영하는 모든 제도는 완벽하지 않으며 오류는 언제나 존재하므로, 이러한 위험성이 AI-ACT의 고유한 것이 아니라고 볼 수 있다. 그러나 AI의 알고리즘의 자동 의사결정 기술의 수준은 아직 인간의 수준에 이르지 못하고 있다. 가용 데이터의 오류나 편향은 AI의 판단을 어처구니없는 방향으로 이끌기도 한다. 현재로서는 AI-ACT의 부정확성은 울리히 벡이 말하는 이른바 '체계적 위험'의 범주 안에 있으며,[63] AI-ACT의 부정확성으로 인한 인격권, 명예권 침해 가능성은 섣불리 간과해서는 안 되는 문제라고 본다.

이른바 위축 효과에 의한 일반적 행동자유권과 표현의 자유에 대한 침해가 발생할 여지도 있다.[64] 한편으로는 비공개 데이터를 바탕으로 자신의 행동이 관찰되고 분석된다는 인

도적 행동을 위한 AI, 접근성을 위한 AI를 목표로 한 AI 지원 프로그램 등을 들 수 있다.

61) P. Aarvik, op. cit., p. 11.

62) Ipaidabribe와 같은 포털에서 익명의 제보가 공직에 있는 개인을 부정확하게 비난할 수 있다면 중대한 문제가 발생한다. 이 문제를 해결하기 위해 Ipaidabribe는 사람들이 개인적으로 범인을 식별할 수 없도록 하여, 잘못된 고발에 대한 동기를 다소 감소시킨다고 한다(N. Köbis · C. Starke · I. Rahwan, op. cit., pp. 1-12).

63) U. Beck(홍성태 역), 위험사회, 새물결, 1997, 55면 참조.

식 때문에 공직자 스스로 검열하고 행동을 조정함으로써 위축 효과가 발생할 수 있다.[65] 다른 한편으로 AI-ACT가 크라우드소싱 데이터를 사용하여 작동하는 경우, 관련된 정보를 제공한 일반 시민의 개인 정보가 유출되어 정치적 또는 행정적 보복의 대상이 될 수 있는 경우 위축 효과는 극심할 수 있다.

참고로 AI-ACT에 의한 기본권 제한 위험이 큰 사람은 주로 공직자이며, 이른바 특별권력관계 이론에 따라 원칙적으로 기본권 침해가 문제되지 않는다는 주장도 있을 수 있다. 그러나 특별권력관계 이론은 전근대적 관헌국가를 배경으로 만들어진 이론이며, 오늘날에는 유효하지 않다. 물론 공직의 성격상 일반 국민과 구별되는 특수한 헌법상 지위에 입각한 특별한 기본권 제한이 정당화되는 경우가 있다.[66] 그러나 AI-ACT에 의한 일방적인 부패 혐의 지목은 형사처벌과 징계, 극도의 명예 실추로 이어질 가능성이 크므로, 이른바 특수지위 이론에 입각하여 기본권 보장의 정도가 상대적으로 낮을 수 있는 경영수행관계의 영역이라고도 말하기 어렵다.[67]

3. 편향성과 포퓰리즘적 악용

AI-ACT는 민주주의와 법치주의에 부적합할 수 있는 수동적 위험성을 가질 뿐만 아니라, 자칫하면 민주주의와 법치주의에 심각한 위협을 가할 수 있는 능동적 위험성도 내재하고 있다. 이러한 위험성은 현상, 징후, 우려 등 다양한 차원에 분포되어 있으며, 어쩌면 기우에 지나지 않을 수도 있다. 그러나 AI 관련 기술이 발전할수록, AI-ACT의 채택과 활용이 늘어날수록 위험은 증가 경향을 보이는 '미끄러운 경사면' 문제라고 하겠다.

앞서 AI-ACT에 입력되는 데이터의 편향성에 따라 도출되는 결과도 편향될 것이라는 지적을 했다. 이러한 편향성은 저절로 생기는 때도 있겠으나, 훈련자 또는 훈련 참여자의 의도, 묵인의 결과일 때도 많다. 여기에 더하여 AI의 핵심이라고 할 수 있는 알고리즘을 설계할 때 종종 광범위한 가치판단이 개입되어 편향성을 내재하는 때도 있다. "AI는 기술이 아니라 이념"이라는 말은[68] AI의 편향성을 강조하는 표현으로 이해할 수 있다. 우려스러운

64) 인터넷 게시판 등에서 이루어지는 정치적 익명 표현을 규제할 경우, 정치적 보복을 당할 우려 때문에 일반 국민은 자기검열 하에서 비판적 표현을 자제하게 될 것이라고 지적한 헌법재판소 판례로서 헌재 2010. 2. 25. 2008헌마324 등, 판례집 22-1상, 347.

65) 김희정, 앞의 논문, 215면; 이것이 공직사회의 복지부동 현상과 재능 있는 공무원의 이탈을 초래할 수 있음을 앞서 지적하였다.

66) 장영수, "공직자윤리법에 따른 공직자 재산등록과 백지신탁제도의 법적 문제점과 개선방향", 고려법학 제70호, 2013, 332면 참조.

67) 특별권력관계의 현대적 재해석에 관하여는 C. H. Ule, "Problem des verwaltungsgerichtlichen Rechtsschutzes im besondern Gewaltverhältnis", VVDStRL Heft 15, 1957, S. 134-135 참조.

68) J. Lanier・G. Weyl, "AI is An Ideology, Not A Technology", WIRED, 15 March, 2020 (https://www.wired.com/story/opinion-ai-is-an-ideology-not-a-technology/).

것은 편향성이 특정 인물 또는 집단에 대한 혐오・차별・배제로 귀결될 때가 많은데, 이것은 나날이 심각해지는 정치적 양극화와 포퓰리즘의 자양분이다.

오늘날의 미디어 환경은 AI-ACT의 포퓰리즘적 악용의 온상이 될 수 있다. AI-ACT가 적발한 부패 의심 사례가 미디어 알고리즘을 매개로 삽시간에 광범위하게 유포될 수 있다. 특히 부패 혐의는 미디어 알고리즘이 선호할 만한, 즉 조회 수를 늘리고 '좋아요'를 유도할 만한 매력적인 소재가 될 가능성이 크다. 미디어 알고리즘 역시 선동적 정치인이나 자본에 의하여 의도적으로 조작되고, 정치적 양극화를 심화하여 포퓰리즘을 초래하고 있다는 지적이 비등하다.[69]

포퓰리즘은 국민의 열정을 동원한다는 점에서 외관상 민주적으로 보이기도 하지만, 그 결과는 특정인 또는 특정 세력의 전체주의적 지배이며, 민주주의를 파괴할 잠재적 위험 요소가 될 수 있다. 그리고 다음에서 살피는 바와 같이 전체주의적 지배는 권력의 집중과 독재와 다르지 않으며, 이것은 그 자체로 법치주의를 파괴할 수 있는 잠재적 원인이 된다.

4. 총체적 감시사회로의 징검다리

법치주의는 자유를 위험에 빠뜨리는 진정한 원인은 권력 그 자체가 축적되는 데에 있을 뿐이라고 보는 견해를 바탕으로 한다. 따라서 권력의 소재와 관계없이 그 크기 자체를 적절한 한도 내에 머무르게 함이 중요 관심사다.[70] 그런데 AI-ACT는 새롭게 발달하는 기술이 채택되고 그것의 운영 권한이 특정 기관에 집중됨에 따라,[71] 권력을 확장하고 국민의 자유를 위협하는 수단으로 전락할 수 있다. 특히 부패 척결을 담당하는 행위자들이 종종 심각한 부패에 연루된다는 이른바 '부패 함정' 현상을 고려할 때 이러한 위험성은 결코 가벼운 것이 아니다.[72]

예를 들어 크라우드소싱이나 웹크롤링, 데이터 흔적을 통해 얻은 빅데이터를 통해 부패가 발생하는 시간이나 지역, 부패 행위자의 특성을 예측하는 알고리즘을 학습시킬 수 있다. 이렇게 되면 일정한 공직자와 관련자들이 광범위한 감시의 대상이 될 수 있다. 심지어 부패를

69) Y. Mounk(함규진 역), 앞의 책, 190면; 이승택, 앞의 논문, 547면; 참고로 시민이 부패 뉴스에 압도되지 않도록 AI-ACT로부터 도출된 부패 혐의를 연속적으로 보도하기보다는 일괄 보도를 제공하는 것이 바람직하다는 의견이 제시되고 있는데, 경청할 만하다. N. Köbis・C. Starke・I. Rahwan, op. cit., p. 16; 부패 방지에 있어서 미디어의 역할에 대한 상세한 설명으로서 L. Camaj, "The Media's Role in Fighting Corruption: Media Effects on Governmental Accountability", The International Journal of Press/Politics, vol. 18, no. 1, Jan. 2013, pp. 21 - 42 참조.

70) I. Berlin(박동천 역), "자유의 두 개념", 이사야 벌린의 자유론, 아카넷, 2019, 358면.

71) 일치하는 사례는 아니나 청와대 국민청원이 청와대 비서실에 권력을 집중시켜 권력 집중을 초래한다는 비판이 시사하는 점이 있다(윤형석, "새로운 국민소통 플랫폼으로서 청와대 국민청원", 법과정책 제27권 제2호, 2021, 100면).

72) N. Köbis・C. Starke・I. Rahwan, op. cit., p. 17; 정성범・백윤철, 앞의 논문, 288면.

저지르는 사람의 얼굴 특성을 식별할 수 있다는 심리학의 실험적 연구를 바탕으로 AI 모델을 구축해 얼굴 이미지를 기반으로 한 개인 수준 부패 위험을 예측하는 실시간 안면 분석 시스템을 개발하려는 기업이 있다.[73] 자연히 예측을 위한 얼굴 이미지의 출처와 활용 방법 등을 둘러싸고 많은 논란이 불거질 것이다.[74] 더 심각한 문제는 이러한 시스템이 실제 개발되어 누군가에 의해 활용될 때 발생하는 권력 집중과 남용 가능성이다.

특히 AI가 다양하게 수집되는 바이오 정보를 수집하고 활용하게 되면 문제는 더욱 심각해질 수 있다.[75] 실제로 중국의 한 기업은 자동으로 발걸음을 인식하는 수이디혜안(水滴慧眼)과 얼굴을 인식하는 수이디신감(水滴神鑒)이라는 AI 시스템을 개발하여 상용화했다. 이것은 발걸음의 특성을 데이터로 집약하여 특정인을 인식하는 프로그램이다.[76] 일부 의견은 이러한 시스템이 점수 시스템과 연결되어 국가에 대한 충성심과 좋은 습관을 장려하는 데에 사용될 수 있으며, 이른바 디지털 독재체제를 앞당길 것이라는 우려를 하고 있다.[77] 실제로 최근 벌어진 홍콩의 대규모 시위와 그 과정에 만들어진 복면금지규례를 함께 소개하면서,[78] AI의 부패와 AI를 통한 인권 침해 가능성을 지적하는 연구가 있다.

점차 고도화되는 AI 기술이 권력의 손에 넘어가게 되면 이전에는 상상조차 할 수 없었던 국민에 대한 총체적 감시(Total Surveillance) 사회가 도래할 위험이 커지며, 이것은 곧바로 법치주의가 무너진 디스토피아를 의미할 수 있다.[79] 물론 AI 기술의 도입이 국민에 대한 감시와 탄압에 사용한다는 명목을 내세울 리 만무하며, 그보다 훨씬 더 정당한(또는 정당해 보이는) 목적, 이를테면 반부패를 표방할 것이다. 그러나 AI를 통한 반부패, 즉 AI-ACT가 AI를 통한 감시사회를 여는 징검다리가 될 수도 있음을 유의해야 하며, 각별한 주의와 사전적인 입법 보완이 필요하다.

---

73) 예를 들어 미국의 ClearView라는 회사는 링크드인, 페이스북, 트위터와 같은 소셜 미디어 플랫폼에서 수십억 개의 이미지를 인터넷에서 긁어내어 경찰이 용의자를 찾는 데 도움을 주는 서비스를 개발했다고 한다(N. Köbis・C. Starke・I. Rahwan, op. cit., p. 19).

74) K. Crawford, “Regulate Facial-Recognition Technology”, Nature 572(7771), 2019, p. 565.

75) 고도로 발달하는 AI에 의한 국민의 개인정보자기결정권 침해, 표현의 자유 침해, 개인 정보와 밀접하게 관련된 프라이버시권 침해 등을 보호할 법률이 전무한 상태라는 지적이 있다. 이러한 문제를 개선하기 위해선 개인정보보호법 제23조48에 생체정보와 관련된 내용을 신설할 필요가 있다고 본다(김형섭・황선영, 앞의 논문, 19면). 참고로 EU 등 주요 국가들은 개인정보의 한 유형으로 ‘바이오정보(BiometricData)’를 구체적으로 정의하여, 바이오정보를 보호하기 위한 보호원칙 등을 담은 가이드라인을 제시하고 있다(방송통신위원회・한국인터넷진흥원, 바이오정보 보호 가이드라인, 2017, 30면 참조).

76) 김형섭・황선영, 앞의 논문, 5면.

77) P. Aarvik, op. cit., p. 27.

78) 복면금지규례가 실시된 지 한 달이 조금 넘은 2019년 11월 18일 홍콩 고등법원은 복면금지법에 대한 위헌 판결을 내렸다고 한다(김형섭・황선영, 앞의 논문, 11면).

79) 김형섭・황선영, 앞의 논문, 16면.

5. 대책 수립의 방향

지금까지 살핀 바와 같이 AI-ACT는 반부패의 차세대 개척지로 주목받지만, 헌법의 기본원리와 충돌하거나 그것을 위협할 가능성도 내재하고 있다. 반부패를 위한 효과적인 수단과 민주주의와 법치주의를 위협할 화근, 두 지점 사이 어딘가에 우리가 넘지 말아야 할 미세한 선이 있다. 문제 해결의 첫걸음은 블랙박스와도 같은 AI-ACT의 한계를 극복하고 투명성을 확보하는 것이다. 투명성은 책임성을 담보하고, 편향성을 교정하고, 통제 가능성을 획득하는 전제조건이 된다. 여기에서 투명성은 알고리즘 투명성과 AI-ACT 이면의 인간 행위자에 대한 투명성이라는 이중의 투명성을 요구하는 것으로 이해되어야 한다.[80]

투명성 확보를 위해 첫 번째, 법적 보완이 필요하다. 2018년 시행된 유럽일반정보보호법(GDPR)은 자동 의사결정 및 프로파일링을 할 때 '사용되는 로직'에 대한 정보를 제공하라고 규정하고 있으며, "정보 처리자 측에 대하여 인간의 개입을 얻을 권리", "자기 입장을 설명할 권리", "자동 결정에 이의를 제기할 권리"를 규정하고 있다. 이것이 AI 알고리즘 투명성과 관련하여 현재로서는 가장 진일보한 법 제도라고 평가할 수 있다.[81] 우리나라의 경우 「지능정보화기본법」이 2020년 제정되어 있으며,[82] 이 법 제3조 제1항은 "국가 및 지방자치단체와 국민 등 사회의 모든 구성원은 인간의 존엄·가치를 바탕으로 자유롭고 개방적인 지능정보사회를 실현하고 이를 지속적으로 발전시킨다."라고 하고 있다. 비록 추상적·포괄적이기는 하지만 AI-ACT 알고리즘 투명성을 요구할 수 있는 법적 근거로서 기능할 수 있다고 본다.[83]

그러나 AI-ACT가 본격적으로 활용된다면 일반적·포괄적 규정을 넘어 보다 구체적인 규정을 둘 필요가 있다. 이를테면 「부패방지 및 국민권익위원회의 설치와 운영에 관한 법률」에 알고리즘 개방성과 인간의 개입권에 대하여 최소한 유럽일반정보보호법 수준의 구체적 규정을 마련하고, 구체적인 관철방안과 제재 규정 등을 함께 두는 것이 타당하다. 특히

80) 2019년 5월 경제협력개발기구(OECD)는 최초로 여러 정부가 합의한 AI권고안을 수립했다. 신뢰 가능한 AI 구현을 위한 다섯 개의 원칙과 정책 고려사 항으로 구성된 이 권고안은 OECD 각료이사회에서 회원국 전원 찬성으로 공식 채택되었다. 비록 선언적 성격이기는 하지만, 동 권고안은 인간 중심의 가치, AI 기술의 설명 가능성, 보안성, 안정성, 책임성 등의 가치를 AI 관리 책무의 핵심 원칙으로 설정하고 있는데, AI-ACT의 투명성 확보를 위한 논의의 출발로서 의미가 깊다(https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449).

81) 김희정, 앞의 논문, 223면 이하; 그러나 제공되어야 하는 정보는 어느 정도까지인지, 소극적인 개입, 설명, 이의제기 요구만으로 충분한 투명성을 확보할 수 있는 것인지, 여전히 과제는 산적해 있다(P. Aarvik, op. cit., p. 25).

82) 기존의 국가정보화기본법이 법률 제17344호로 2020년 6월 9일 전면 개정된 것이다.

83) 참고로 신용정보법 제36조의2 제2항은 "신용정보 주체가 자신에게 유리하다고 판단되는 정보를 제출할 수 있고, 기초정보에 오류가 있다고 판단되는 경우 이를 정정하거나 삭제할 것을 요구할 수 있으며, 이를 바탕으로 자동화 평가 결과의 재산출을 요구할 수 있다."라고 규정하고 있다. AI-ACT의 경우에도 이 정도 수준의 법적 규율이 필요하다.

AI-ACT를 채택할 때 알고리즘 이면에 존재하는 또는 존재할 수밖에 없는 인간 행위자의 존재를 명확히 하여, 법적 지위, 금지 행위, 제재와 처벌 등을 규정함으로써 그의 책임성을 담보하는 규정을 마련할 필요가 있다.

두 번째, 기술적 보완도 병행되어야 한다. 복잡한 알고리즘은 투명성을 확보하는 것이 어려울 뿐만 아니라 공개되더라도 그 실체와 문제점을 파악함이 쉽지 않다. 따라서 이를 테스트 하기 위해 코드에는 영향을 미치지 않지만 서로 다른 입력값이 결과를 어떻게 변화시키는지 조사하여 편향과 오류를 잡는 시스템 개발이 진행되고 있다. '알고리즘 통제를 위한 알고리즘'의 개발, '알고리즘 사이의 견제와 균형'을 모색하는 시도라고 하겠다.[84] 완벽할 수는 없지만, 의미 있는 노력이라고 평가할 수 있다.

마지막 세 번째로 사회적 보완이다. 대부분의 사안과 마찬가지로 단편적인 법적 또는 기술적 보완으로 위험에 완벽하게 대비할 수 없으며, 기술 발전과 활용 증가에 따라 위험은 예측할 수 없을 정도로 증가할 것이 분명하다. AI-ACT의 편리성과 객관성의 환상에 지나치게 경도되지 말고, 적절한 비판적 자세를 취하는 것이 필요하다. AI-ACT에 대한 시민들의 적극적인 참여와 더 큰 발언권을 요구해야 하며, 궁극적으로 "루프 속의 사회"를 실현하려는 노력이 필요하다.[85] 모든 국민이 자신의 수준에 맞는 정부를 선택하는 것처럼, 우리는 우리 수준에 맞는 AI-ACT를 얻게 될 것이다.

Ⅴ. 결론

(1) 헌법의 기본원리 중 핵심은 민주주의와 법치주의이며, 이것은 근대 입헌적 민주주의에 대한 헌법재판소의 설명과 같은 맥락이다. 민주주의와 법치주의는 상호 완벽하게 조화된다는 견해와 모순·갈등할 수밖에 없다는 견해가 존재하는데, 양자의 관계를 역설적 공존의 관계, 상호 감염의 관계로 이해하는 것이 타당하다. 이는 하나의 원리가 다른 원리를 긍정적으로 변화시키기도 하고, 하나의 원리가 다른 원리의 존립 조건이 되는 관계라는 의미이다.

부패는 헌법의 기본원리를 위협하고 훼손하며, 반부패는 헌법의 기본원리를 보호한다. 반부패가 헌법의 기본원리에 반한다면, 헌법의 기본원리에 대한 내적 붕괴를 불러오고, 더 큰 부패를 초래할 수 있다. 이 때문에 AI를 활용한 새로운 반부패 수단, 이른바 AI-ACT 역시 헌법의 기본원리와 관련하여 적합한지, 혹시 헌법의 기본원리를 파괴할 잠재적 위험성은 없는지 검토할 필요가 있다.

---

84) 옥스포드 인터넷 연구소와 앨런 튜링 연구소의 개발에 대한 소개로서 P. Aarvik, op. cit., p. 25.

85) N. Köbis · C. Starke · I. Rahwan, op. cit., p. 18; 공직부패 척결은 국가의 임무인 동시에 그러한 정부를 형성한 시민사회가 책임질 일이라는 의견으로서 강경근, 앞의 논문, 105면.

(2) 현재 시점에서 AI-ACT는 약한 AI를 기반으로 만들어진 부패 방지 도구로서, 대규모 데이터 세트를 선별하고, 이를 분석하여 부패의 위험, 의심, 명백한 사례 등을 탐지, 예측 및 보고하는 시스템이다. 자원을 절약하고, 객관성을 높일 수 있다는 기대로 "부패 방지의 차세대 개척지"로 묘사되기도 한다. AI-ACT는 상향식과 하향식으로 나눌 수 있는데, 전자는 언론인, 시민사회 그리고 일반 시민이 주체가 되는, 후자는 수사기관이나 감사 담당자 등이 주체가 되는 반부패 도구다.

AI-ACT의 기본 구조는 데이터 입력, 알고리즘 디자인, 제도 채택 여부 결정의 세 단계로 그려볼 수 있다. AI-ACT의 가용 데이터로는 정부 수집 데이터, 유출된 데이터, 크라우스소싱 데이터, 언론사나 SNS의 텍스트 등을 들 수 있다. 비공개 데이터나 데이터 흔적(data traces)과 같은 개인정보보호 측면에서 민감한 데이터의 사용 가능성, 데이터의 양적 부족과 표준화, 데이터의 질적 오염과 편향성 등의 문제가 제기되고 있다.

알고리즘 설계에 있어서 예측의 정확성을 확보하는 것이 중요한데, 정확성과 적발 가능성은 맞교환 관계(trade-off)를 이룬다. 상향식 AI-ACT에서는 정확성에, 하향식에서는 적발 가능성에 주안점을 두는 전략이 유효하다. 알고리즘과 관련하여 가장 민감한 것은 이른바 블랙박스 문제다. 알고리즘의 투명성과 설명 가능성 확보는 가장 핵심적 문제 중 하나다.

AI-ACT의 기대효과에도 불구하고 이것의 도입이 언제나 긍정적 결과를 가져오지는 않는다. 공무원에 대한 과도한 감시는 복지부동과 유능한 공무원의 이탈을 초래할 수 있다. 신중하지 못한 활용은 "스패밍(spamming)" 효과를 발생시켜 시민들의 부패에 대한 감각을 무디게 하고, 자신의 부조리에 대한 합리화 수단을 제공할 수 있다. 유행과 호기심으로 선택하는 일이 없어야 하며, 때로는 고전적이고 정적인 ICT 시스템만으로도 충분할 때가 많음을 기억해야 한다.

(3) AI-ACT가 반부패에 있어서 민주화를 촉진하는 인상을 주기도 한다. 그러나 기술적 미비로 말미암아 또는 시민들의 거부감으로 말미암아 AI-ACT 이면에는 인간 행위자가 존재할 수밖에 없고, 이들에게 민주적 정당성과 책임성을 물리지 못하면 기술 권위주의 엘리트, 심지어 거대 다국적 기업에 의한 통제와 지배가 조장될 수 있다. 이 같은 권력에 의해서는 더 많은 기본권 침해가 초래될 위험이 있으며, 이는 법치주의의 훼손을 의미한다. 특히 개인 정보 관련 데이터의 활용과 공개로 인한 사생활의 비밀 침해, 부정확한 부패혐의자 지목으로 인한 인격권, 명예권 침해, 상시 감시체제로 인한 위축 효과의 발생과 이로 인한 일반적 행동 자유권과 표현의 자유 침해 등 다양한 문제가 발생할 수 있다.

아직은 징후와 우려 차원이라고 해도 AI-ACT는 민주주의와 법치주의에 심각한 위협 또는 그러한 위협으로 향하는 미끄러운 경사면이 될 수 있다. 데이터와 알고리즘의 편향성은 특정 인물 또는 집단에 대한 혐오·차별·배제로 귀결될 수 있으며, 이것이 미디어 알고리즘을 통해 무차별 또는 선택적으로 유포된다면 정치적 양극화와 포퓰리즘의 멋이감이 될 수 있다. 포퓰리즘은 국민의 열정을 동원한다는 점에서 외관상 민주적으로 보이지만, 전체

주의를 조장하여 민주주의를 파괴하는 결과를 초래한다. 동시에 이것은 권력의 집중과 독재를 의미하며, 이는 권력의 과도한 확장의 경계를 본질로 하는 법치주의에 대한 위협을 의미한다. AI-ACT가 치안·수사 등 목적으로 부패한 공직자에서 일반 국민으로 칼날을 돌린다면, 법치주의가 무너진 총체적 감시사회의 서막이 열릴 수 있다.

문제 해결의 첫 단계는 AI-ACT의 투명성 확보다. 투명성이 확보되어야 비로소 책임성을 담보하고, 편향성을 교정하고, 통제 가능성을 획득할 수 있을 것이다. 이를 위해 첫 번째, 부패방지법에 유럽일반정보보호법의 개입권과 여타의 주체, 절차, 제재에 관한 기본 규정을 두는 등 법적 보완이 필요하며, 두 번째 AI를 검증하는 AI를 개발하여 AI 간의 견제와 균형을 모색하는 기술적 보완이 필요하고, 세 번째는 편리성과 객관성의 환상에 경도되지 말고 적절한 비판적 자세를 유지하는 사회적 보완이 필요함을 강조하였다.

모두에서 언급한 것처럼 AI 기술은 아직 발생기에 머무르고 있으며, AI-ACT의 수준도 초보적인 수준이다. 도입 사례가 늘고 있지만, 우리나라를 포함한 다수의 국가에서는 사례를 찾기 어렵다. 이 때문에 AI-ACT에 관한 연구는 본격화되지 못했고, AI-ACT와 헌법의 관계에 대한 논의는 국내외 막론하고 찾기 어렵다. AI-ACT의 세밀한 내용, 특히 구체적 실제, 기술적 현황, 문제 해결 방안 등에 대한 본격적인 접근이 쉽지 않은 형편이며, 본 논문도 같은 한계를 가진다. 향후 진전된 연구가 이어지기를 기대한다.

참고문헌

강경근, “공직자 부패와 헌법이론”, 한국부패학회보 제1권 제1호, 1997

권영성, 헌법학원론, 법문사, 2010

계희열, 헌법학(상), 박영사, 2001

김병록, “공직부패의 헌법적 조명”, 공법연구 제30권 제2호, 2001

김윤명, 인공지능과 리걸 프레임, 10가지 이슈, 커뮤니케이션북스, 2016

김정수, 반부패의 세계사, 가지, 2020

김진영, “부패개념에 대한 고찰”, 한국부패학회보 제22권 제4호, 2017

김형섭・황선영, “AI기술의 부패방지와 인권 침해의 논의 – 홍콩 사례(복면금지법)를 중심으로”, 한국부패학회보 제25권 제2호, 2020

김희정, “알고리즘 자동의사결정으로부터 개인의 보호 – 인간의 개입권(the right to human intervention)을 중심으로”, 헌법학연구 제26권 제1호, 2020.

방송통신위원회・한국인터넷진흥원, 바이오정보 보호 가이드라인, 2017

송준모・박영득, “청와대 국민청원에서는 무엇이 일어나는가?: 자연어 처리를 활용한 청와대 국민청원 분석”, 한국정치학회보, 제53권 제5호, 2019

심우민, “인공지능의 발전과 알고리즘의 규제적 속성”, 법과 사회, 제53권, 2016

윤형석, “새로운 국민소통 플랫폼으로서 청와대 국민청원”, 법과정책 제27권 제2호, 2021

이승택, “뉴미디어 시대의 알고리즘과 민주적 의사형성”, 법학논총 제33권 제3호, 2021

장영수, “공직자윤리법에 따른 공직자 재산등록과 백지신탁제도의 법적 문제점과 개선방향”, 고려법학 제70호, 2013.

장영수, 헌법학, 홍문사, 2020

정성범・백윤철, “헌법상 행정부패에 관한 연구”, 헌법학연구, 2006

松尾豊, 人工知能は人間を超えるか, 角川, 2015

U. Beck(홍성태 역), 위험사회, 새물결, 1997, 55면 참조.

I. Berlin(박동천 역), “자유의 두 개념”, 이사야 벌린의 자유론, 아카넷, 2019

Y. Mounk(함규진 역), 위험한 민주주의, 와이즈베리, 2018

P. Aarvik, “Artificial Intelligence-a promising anti-corruption tool in development settings?” U4 Report 2019-1, 2019

L. Camaj, “The Media’s Role in Fighting Corruption: Media Effects on Governmental Accountability”, The International Journal of Press/Politics, vol. 18, no. 1, Jan. 2013

R. Carvalho・M. Ladeira・F. M. Monteiro・G. L. d. O. Mendes, “Using Political Party Affiliation Data to Measure Civil Servants′ Risk of Corruption”, 2014 Brazilian Conference on Intelligent Systems, 2014

K. Crawford, “Regulate Facial-Recognition Technology”, Nature 572(7771), 2019

N. Köbis・C. Starke・I. Rahwan, “Artificial Intelligence as an Anti-Corruption Tool(AI-ACT) - Potentials and Pitfalls for Top-down and Bottom-up Approaches”, 2021 (https://search-ebscohost-com-ssl.oca.korea.ac.kr/login.aspx?direct=true&db=edsarx&AN=edsarx.2102.11567

&lang=ko&site=eds-live&scope=site).

J. Lanier · G. Weyl, “AI is An Ideology, Not A Technology”, WIRED, 15 March, 2020 (https://www.wired.com/story/opinion-ai-is-an-ideology-not-a-technology/).

F. J. López-Iturriaga · I. P. Sanz, “Social Indicators Research”, Dordrecht, Vol. 140, Iss. 3, 2018

C. Mouffe, The Democratic Paradox, VERSO, 2000

A. Petheram · I. N. Asare, “From Open Data to Artificial Intelligence: The next Frontier in Anti-Corruption”, Oxford Insights, 2018 (https://www.oxfordinsights.com/insights/aiforanticorruption)

A. Rafaeli · S. Ashtar · D. Altman, “Digital Traces: New Data, Resources, and Tools for Psychological-Science Research”, Current directions in psychological science, 28(6), 2019

C. Schmitt(trans. by E. Kennedy), The Crisis of Parliamentary Democracy, MIT Press, 1985

D. Seligsohn · M. Liu · B. Zhang, “The sound of one hand clapping: transparency without accountability”, Environmental politics, 27(5), 2018

C. H. Ule, “Problem des verwaltungsgerichtlichen Rechtsschutzes im besondern Gewaltverhältnis”, VVDStRL Heft 15, 1957

투고일자 : 2022. 02. 27
수정일자 : 2022. 03. 16
게재일자 : 2022. 03. 31

<국문초록>

인공지능 기반 반부패 도구(AI-ACT)의 헌법원리 적합성에 관한 고찰

조 한 상

헌법의 기본원리 중 핵심은 민주주의와 법치주의이며, 이것은 근대 입헌적 민주주의에 대한 헌법재판소의 설명과 같은 맥락이다. 양자의 관계는 역설적 공존의 관계, 상호 감염의 관계로 이해하는 것이 타당하다. 반부패가 헌법의 기본원리에 반한다면, 헌법의 기본원리에 대한 내적 붕괴를 불러오고, 더 큰 부패를 초래할 수 있다. 이 때문에 AI를 활용한 새로운 반부패 수단, 이른바 AI-ACT 역시 헌법의 기본원리와 관련하여 적합한지, 혹시 헌법의 기본원리를 파괴할 잠재적 위험성은 없는지 검토할 필요가 있다.

현재 시점에서 AI-ACT는 약한 AI를 기반으로 만들어진 부패 방지 도구로서, 대규모 데이터 세트를 선별하고, 이를 분석하여 부패의 위험, 의심, 명백한 사례 등을 탐지, 예측 및 보고하는 시스템이다. AI-ACT는 상향식과 하향식으로 나눌 수 있는데, 전자는 언론인, 시민사회 그리고 일반 시민이 주체가 되는, 후자는 수사기관이나 감사 담당자 등이 주체가 되는 반부패 도구다. AI-ACT의 기본구조는 데이터 입력, 알고리즘 디자인, 제도 채택 여부 결정의 세 단계로 그려볼 수 있다.

AI-ACT는 민주주의와 법치주의에 심각한 위협 또는 그러한 위협으로 향하는 미끄러운 경사면이 될 수 있다. 데이터와 알고리즘의 편향성은 특정 인물 또는 집단에 대한 혐오・차별・배제로 귀결될 수 있으며, 이것이 미디어 알고리즘을 통해 무차별 또는 선택적으로 유포된다면 정치적 양극화와 포퓰리즘의 멋이감이 될 수 있다. 포퓰리즘은 국민의 열정을 동원한다는 점에서 외관상 민주적으로 보이지만, 전체주의를 조장하여 민주주의를 파괴하는 결과를 초래한다. 동시에 이것은 권력의 집중과 독재를 의미하며, 이는 권력의 과도한 확장의 경계를 본질로 하는 법치주의에 대한 위협을 의미한다. AI-ACT가 치안・수사 등 목적으로 부패한 공직자에서 일반 국민으로 칼날을 돌린다면, 법치주의가 무너진 총체적 감시사회의 서막이 열릴 수 있다.

문제 해결의 첫 단계는 AI-ACT의 투명성 확보다. 투명성이 확보되어야 비로소 책임성을 담보하고, 편향성을 교정하고, 통제 가능성을 획득할 수 있을 것이다. 이를 위해 첫 번째, 부패방지법에 유럽일반정보보호법의 개입권과 여타의 주체, 절차, 제재에 관한 기본 규정을

두는 등 법적 보완이 필요하며, 두 번째 AI를 검증하는 AI를 개발하여 AI 간의 견제와 균형을 모색하는 기술적 보완이 필요하고, 세 번째는 편리성과 객관성의 환상에 경도되지 말고 적절한 비판적 자세를 유지하는 사회적 보완이 필요함을 강조하였다.

주제어: 인공지능 기반 반부패 도구, 반부패, 헌법원리, 민주주의, 법치주의, 포퓰리즘