



제10회

# 학봉상 시상식

## 법학논문부문 대상 수상작

인공지능(AI) 기반 신용평가  
시스템(CSS) 규제 방향 연구  
- 금융이력부족자(Thin Filer)의  
권리보호를 중심으로 -

장효선

동아대학교 법학전문대학원 석사과정



## 인공지능(AI) 기반 신용평가 시스템(CSS) 규제 방향 연구 - 금융이력부족자(Thin Filer)의 권리보호를 중심으로 -

장효선

### 초록

본 논문은 인공지능(AI) 기반 신용평가시스템(CSS)의 현황과 문제점을 분석하고, 특히 금융이력부족자의 권리보호 관점에서 제도적 개선방안을 모색하는 것을 목적으로 한다.

금융 분야에서 인공지능 활용이 확대되면서 금융에 대한 접근성이 제고되는 한편, 인공지능이 내재하고 있는 설명가능성과 강건성(robustness) 부족의 문제가 신용평가에 대한 신뢰성에 손상을 줄 수 있다. 또한 비금융정보 활용 확대로 금융이력부족자의 사생활 침해와 부당한 차별 문제가 대두되고 있다. 이에 본 연구는 인공지능 기반 신용평가시스템의 공정성과 투명성 제고를 위한 다각도의 금융규제정책 방안을 제시한다. 인공지능 기반 신용평가의 기본 원칙과 기준 마련, 차별금지과 개인정보보호 관련 법규 정비, 피해구제 수단 강화 등 금융당국의 선제적 대응이 필요하다.

그러나 우리 사회에서 금융이 생존과 직결되는 필수적인 영역인 만큼, 법과 제도만으로는 한계가 있으므로 금융포용과 공정성 구현을 위한 인간의 가치판단과 사회적 노력이 병행되어야 한다. 인간 전문 심사역의 역할 강화, 관계금융적인 노력 등 사회 안전망 구축을 위해 다방면의 보완책 마련이 요구된다. 인간이 기술을 통제하고 선용할 수 있도록 혁신성은 수용하되 폐해는 경계하고 보완하려는 사회 구성원 모두의 노력이 중요하다.

**주요어** : 인공지능, 신용평가, 신용평가시스템(CSS), 신용평가업 규제, 금융이력부족자(Thin Filer), 금융포용

## I. 서론

### 1. 연구의 배경과 필요성

신용평가는 자본시장에서의 증권발행, 자금마련, 발행 등의 기반이 되며 자본주의 사회에서 필수적인 기능으로, 문지기(gatekeeper)에 비유되기도 한다<sup>1)</sup>. 신파일러(Thin Filer)는 파일의 두께가 얇은 사람들이라는 뜻으로, 적은 금융거래 실적을 가진 ‘금융이력부족자’들을 일컫는다. NICE(나이스)평가정보는 최근 2년 내 신용카드 사용 실적이 없고 3년 내 대출보유 경험이 없는 사람을 금융이력부족자로, KCB의 경우 최근 3년 내 신용거래(신용카드, 대출) 경험이 없는 자를 금융이력부족자로 분류했다. 2021년 상반기 NICE의 금융이력부족자는 1천280만 7천275명으로, 신용등급 대상자 4천730만7천806명의 27.1%를 차지한다. KCB에서 동일 기간 금융이력부족자로 분류된 이들은 1천194만2천503명으로, 신용등급 대상자 4천826만1천439명의 24.7%이다. 거의 국민 4명 중 1명이 금융거래 정보가 부족해 신용을 제대로 인정받지 못하는 있는 실정<sup>2)</sup>이다. 은행에서 대출을 실행하기 위해서는 정기적인 급여 이체나 담보로 잡을 수 있는 재산을 보유하고 있는 등 금융거래실적 등을 확인하여 상환 능력을 추정할 수 있어야 가능하다. 사회초년생·취업준비생인 청년들과 디지털 소외계층인 노년층, 전업주부, 이민자 등은 애초에 이력이 없기에 대출 과정에서 거절당하는 경우가 많다.

이러한 맥락에서 인공지능 기반 신용평가시스템은 금융포용 확대를 위한 혁신적 도구로 주목받고 있다. 많은 신용평가회사에서는 대안신용평가시스템(ACSS, Alternative Credit Scoring System)<sup>3)</sup>을 개발, 비금융정보도 반영하여 신용평가의 정밀성을 높이고자 하였다. AI는 전통적 금융정보뿐만 아니라 비금융 데이터(예컨대 통신요금 납부내역, 전자상거래 이용기록, 소셜미디어 활동정보, 위치정보, 심리검사 결과 등)를 포괄적으로 분석하여 개인의 신용도를 보다 다층적으로 평가할 수 있다. 이를 통해 금융이력부족자도 대출 기회를 얻을 수 있게 되고, 금융 접근성의 사각지대를 완화하는 효과가 기대된다.

그러나 인공지능 신용평가의 확대가 곧바로 긍정적 결과만을 보장하는 것은 아니다. 알고리즘의 불투명성, 데이터 편향, 개인정보 침해, 평가의 불안정성 등 다양한 부작용이 내재되어 있으며, 이러한 문제는 오히려 금융소외계층에 더 큰 피해로 이어질 수 있다. 실제로 2019년 미국에서는 골드만삭스의 ‘애플카드(Apple Card)’가 여성 고객에게 남성보다 현저히 낮은 신용한도를 부여한 사건이 사회적 파장을 일으켰다. 이는 인공지능 기반 신용평가가 성차별적 결과를 재생산할 수 있음을 보여주는 대표적 사례이다. 신용평가의 특성상 어떤 기준으로 등급을 산출했는지 그 평가 과정

1) 송옥렬 “신용평가회사의 문지기책임.” BFL 82 (2017): 72p

2) 김연정, '금융이력부족자' 대부분 신용점수 700점대...은행 대출에 불리, 연합뉴스, 2021.09.27., <https://www.yna.co.kr/view/AKR20210926043100002>(최종접속일 2024.09.11.)

3) [시사금융용어] 대안신용평가, 연합인포맥스, 2019.04.05., <https://news.einfomax.co.kr/news/articleView.html?idxno=4023693>(최종접속일 2024.09.11.)

을 외부에 상세히 밝히기는 불가능하다. 따라서 이러한 블랙박스 모델(blackbox model)의 설명으로는 금융소비자가 점수의 등락을 이해하기 어렵고 정확히 어떤 부분에서 이의를 제기할 수 있을지 일반인의 관점에서는 명확하지 않다. 이 때문에 신용등급의 산정 과정에서 사회적 편향이 반영되었는지, 또는 필요하지 않은 사생활 정보가 이용되었는지 등의 부당한 권리 침해가 발생해도 해결하기 어렵다. 실제로 한국에서 신파일러가 2020년부터 2022년 신용점수 평가 업체에 이의를 제기해 받아들여진 건수는 한 건도 없었다<sup>4)</sup>.

신용평가 과정은 금융이라는 자원을 배분(allocate)하기 위한 사전 작업의 역할을 하므로 분배 과정에서의 ‘공정성(fairness)’<sup>5)</sup>이 지켜져야 하는 공적인 영역에 해당한다고 볼 수 있다. 하지만 동시에 그 공급자인 신용평가업자는 사적 영역에 속하여 영리를 추구하여야 하며 그것이 시장질서의 균형을 가져오기 때문에 이러한 이해관계의 상충을 조화시켜야 할 필요성이 있다<sup>6)</sup>. 따라서 신용평가에서 공정성의 확보는 매우 까다로운 줄타기를 전제로 한다. 공정성에 대해서는 법적, 윤리적, 철학적, 정치·사회적 해석뿐만 아니라 인공지능에서 다루는 지표도 매우 여러가지로 정의가 쉽지 않으나<sup>7)</sup>, 여기에서는 법적인 공정성에 초점을 맞추어 다루고자 한다. 따라서 인공지능 기반 신용평가시스템의 발전은 금융포용이라는 긍정적 가치와, 차별·불공정의 심화라는 부정적 위험이 공존하는 양면성을 갖는다. 본 연구는 이러한 상황에서 특히 금융이력부족자의 권리보호 관점에서 제도적 개선방향을 탐색하고자 한다.

## 2. 연구 목적 및 방법

본 연구의 목적은 다음과 같다. 첫째, 인공지능 기반 신용평가시스템의 도입 배경과 현황을 분석하고, 기술적 특징과 활용 방식의 구체적 실태를 살펴본다. 둘째, 해당 시스템이 내포하는 문제점, 특히 설명가능성 부족, 강건성 결여, 편향된 데이터에 의한 차별, 개인정보 침해 등 구조적 한계를 규명한다. 셋째, 금융이력부족자라는 취약계층이 AI 신용평가 과정에서 직면할 수 있는 위험을 다각도로 고찰한다.

4) 김나경, [단독]‘無이력의 악순환’신파일러 99.9%가 신용 800점 이하... "거래이력이 없지, 빚 안 갚는다 했나", 파이낸셜뉴스, 2023.03.15., <https://www.fnnews.com/news/202303151401284618>(최종접속일 2024.09.11.). 이 기사에 따르면 NICE 신용평가에는 2020년부터 2022년까지 총 82건의 이의 제기가 접수됐지만, 수용하여 반영된 건 단 한 건도 없다. 현재 등급 산출사유에 문제를 제기한 것이 28건으로 가장 많았고 신용등급 상향 방법 문의가 20건으로 뒤를 이었다. 점수를 어떻게 냈는지, 또 어떻게 점수를 올리는지 알려달라는 문의가 절반이라고 한다.

5) 공정성의 법적 개념에 대해 가장 널리 알려진 연구는 존 롤스의 ‘공정으로서의 정의(1958)’이다. 롤스는 평등한 자유의 원칙과 차등의 원칙에 따른 기회균등을 주장하였다.

6) 이정수, “신용평가업 규제의 현황과 문제점 그리고 개선방향”, 증권법연구, 2016, 17(1), 314p

7) 강송희, “기계학습 공정성 관련 연구 동향”, 월간SW중심사회 통권 제68호, 소프트웨어정책연구소, [https://spri.kr/posts/view/22898?code=industry\\_trend#foot\\_3](https://spri.kr/posts/view/22898?code=industry_trend#foot_3)(최종접속일 2024.09.11.) 참조. 기계학습 모델이 이해할 수 있는 공정성은 수학적으로 정의되어 정량평가가 가능해야 한다. 이에 대한 공정성의 정의는 무려 20여 가지로 분류되며, 대부분 분배적 관점에서 통계적 공정성에 초점을 두고 있다. 즉, 통계적 공정성에 대한 단일한, 혹은 최상의 정의에 대한 합의가 존재하지 않는 것이 현실이다.

넷째, 해외 주요국의 규제와 감독 사례를 비교 분석함으로써 우리 제도의 한계와 개선 필요성을 도출한다. 마지막으로, 금융포용과 공정성 제고를 위한 정책적 제언을 제시한다.

연구방법으로는 법학적 분석과 사회과학적 접근을 병행하였다. 우선, 「신용정보법」, 「개인정보보호법」 등 국내 관련 법령과 금융위원회, 금융감독원의 가이드라인을 검토하여 현행 제도적 틀을 분석하였다. 동시에 EU의 GDPR(General Data Protection Regulation), 미국의 ECOA(Equal Credit Opportunity Act), FCRA(Fair Credit Reporting Act) 등 해외 제도와 비교법적 분석을 수행하였다. 본고의 주요 분석대상이 된 법제 및 가이드라인은 다음 표와 같이 정리하였다.

[표 1] 한국, EU, 미국 주요 법제 및 가이드라인

구분	한국	EU	미국
신용평가 관련 기본법	- 자본시장법 - 신용정보법 - 금융소비자보호법 - 개인정보보호법	- AI Act (2023 합의, 2026 시행 예정) - GDPR (2018) - PSD2 (2018)	- 분야별 법률 존재, 단일 포괄법 없음 - FCRA(공정신용보고법) - GLBA(금융서비스현대화법) - HIPAA 등 분야별 개인정보보호법
AI 관련법· 가이드라인	- 지능형로봇법(간접적 정의) - 금융위원회 「금융분야 AI 가이드라인」(2021) - 「금융분야 AI 개발·활용 안내서」(2022) - 「AI 기반 신용평가모형 검증체계」(2023)	- AI Act (고위험 AI 규제) - ISO/IEC TR 24027:2021(AI 편향) - IEC 62366-1:2015 - GDPR·PSD2 내 AI 적용 관련 조항	- NIST AI RMF 1.0 (2023) - NIST SP 1270 (2022, AI 편향) - NISTIR 8312 (2021, 설명가능성 원칙)
개인정보 보호	- 개인정보보호법(국내 포괄 규율) - 신용정보법 통한 금융 데이터 규제	- GDPR : 개인정보 전송·처리 엄격 규율 - PSD2 : 개방형 API 통한 금융정보 활용 보장	- GDPR 같은 포괄법 없음 - HIPAA(의료), GLBA(금융), FCRA(신용) 등 산업별 규제
지급결제· 금융정보 접근 규제	- 금융투자업 규정 - 증권발행 및 공시 규정 - 마이데이터 제도 (API 일부 도입)	- PSD2 : 강력한 인증 + 개방형 बैं킹 API 제도화	- EFTA (전자자금이체법) - FSMA 등 존재하나 PSD2 수준의 '오픈뱅킹' 제도는 부재
차별금지· 공정성 관련 규제	- 신용정보법 제22조의4 (합리적 이유 없는 차별 금지) - 금융소비자보호법 제15조	- 평등대우지침(Directive 2000/43/EC, 2004/113/EC) - 모기지신용지침(Directive 2014/17/EU)	- FHA (공정주택법) - ECOA (신용기회균등법)
특징 및 한계	- AI 특화 단일법 부재 - 가이드라인 중심으로 초기 단계 - 간접차별 규율 미흡	- GDPR·PSD2·AI Act 등 단일·종합 법체계 구축 - 개인정보와 차별금지 모두 포괄적 규제	- 포괄적 법률 부재, 산업별 개별법 위주 - 소비자 보호는 강력하나 (예: FCRA, ECOA), 체계적 일원화 부족

## II. 인공지능 기반 신용평가시스템의 현황

### 1. 인공지능 신용평가의 개념과 특징

미국과 EU는 2008년 금융위기 이후 신용평가기관에 대한 규제를 강화하는 추세를 보였으며, 우리나라도 2001년 신용정보법 제정 이후 지속적으로 규제를 강화했고 2013

년 신용평가 관련 규정이 자본시장법으로 이관되며 규제체계가 정비되었다<sup>8)</sup>. 종래 신용평가를 단순히 신용에 관한 사항으로 보아 신용정보업과 함께 신용정보법에서만 규율하다가, 자본시장 기초인프라의 하나로 새롭게 인식하면서 자본시장법으로 이관한 것이다<sup>9)</sup>. 최근에는 데이터 경제로의 전환에 따라 금융 데이터 및 개인 신용평가 데이터 활용을 통한 신산업 육성 및 국가 경쟁력 제고에 대한 논의가 활발하게 이루어지고 있다. 유럽이 GDPR(General Data Protection Regulation)을 통해 개인정보 보호와 정보이동권 도입으로 소비자 권리와 데이터 활용의 조화를 추구하고 있는 것을 벤치마킹하여 우리나라에서도 오픈뱅킹, 마이데이터 등 금융분야의 데이터 활용 정책을 추진 중이다.

한국의 신용조회회사(Credit Bureau, CB)에서는 개인의 신용에 대한 다양한 정보를 수집한 후 이를 통계적으로 분석, 향후 1년 내 90일 이상 등의 장기연체 등 위험이 발생할 가능성을 개인 신용 평점이라는 수치적인 지표로 제공<sup>10)</sup>한다. 데이터 측면에서는 대출정보나 연체정보 등의 금융정보를 활용한 신용평가가 주로 진행되어 오다가, 금융정보의 부족으로 신용평점이 산출되지 않는 사회초년생, 고연령층, 이민자들의 금융서비스 접근성이 떨어지는 이슈가 생기면서, 비금융정보를 신용평가에 추가로 활용하여 금융이력부족자들을 제도권 금융으로 포용하고자 하는 노력들이 생겨나게 되었다.

[표 2] 신용평가회사별 개인신용평가모형 평가요소 및 활용비중

	상환이력	부채수준	신용거래기간	신용형태	비금융 /마이데이터	합계
NICE	28.4%	24.5%	12.3%	27.5%	7.3%	100%
KCB <sup>11)</sup>	21%	24%	9%	38%	8%	100%
FICO <sup>12)</sup>	35%	30%	15%	10%	-	-

8) 양철원, “신용평가업에 대한 규제강화는 기업 신용평가의 질을 향상시켰는가?” 보험금융연구 106 (2022): 3-38.

9) 김건식·정순섭, 자본시장법 제4판. 박영사, 2023. 1010p

10) NICE평가정보, 개인신용평점의 의미 - 개인신용평점의 정의 및 목적, [https://www.niceinfo.co.kr/creditrating/cb\\_score\\_1\\_1.nice](https://www.niceinfo.co.kr/creditrating/cb_score_1_1.nice)(최종접속일 2024.09.20.)

11) 코리아크레딧뷰로(KCB, 올크레딧)은 일반고객과 장기연체 경험고객을 구분하여 평가영역별 반영비중을 다르게 반영하고 있다. 장기연체 경험고객의 경우 비금융/마이데이터의 비중이 11%라고 명시되어 있다. 또한, 비금융/마이데이터 영역의 요소로는 신용관련 설문 응답도 포함되어 있으며, 설문 결과 건전한 신용생활과 상관관계가 높은 응답을 보인 경우 가점 요인으로 활용된다. <https://www.allcredit.co.kr/screen/sc0682112929>(최종접속일 2024.09.20.)

12) 미국 FICO 점수의 경우 나머지 10%는 신규 신용조회(new credit)를 반영한다고 알려져 있다. 신용조회는 소프트 인쿼리(soft inquiry)와 하드 인쿼리(hard inquiry)로 나뉘는데, 소프트 인쿼리의 경우 정보조회를 목적으로 하는 것으로 보이면 점수에 영향을 끼치지 않는다. 반면 하드 인쿼리는 자동차 대출, 주택모기지, 신용카드 신청 등으로 금융기관이 소비자 신용상태를 확인할 때에 이뤄지게 되며 점수에 영향을 미친다. 새로운 부채는 기존 부채를 연체할 확률을 높이기 때문에 하드 인쿼리로 인한 신규 신용조회는 신용점수를 하락시킬 수 있다. 그러나 이러한 조회로 인한 신용점수 하락은 일반적으로 5점 미만이며 일반적으로 몇 달 내 점수가 반등한다고 한다. 우리나라에서는 2011년 10월부터 서민금융기반강화 종합대책에 의해 신용조회정보는 평가요소에서 제외하고 있다. Jim Akin, What Affects Your Credit Scores?, 2023.07.29., Experian, <https://www.experian.com/blogs/ask-experian/credit-education/score-basics/what-affects-your-credit-scores/>(최종접속일 2024.09.20.)

특히 금융정보를 대신하여 개인 신용평가에 활용할 수 있는 다양한 비금융정보를 보유하고 있는 국내의 핀테크, 빅테크 기업들은 금융산업으로의 진출을 피하면서 비금융정보를 활용한 개인 신용평가를 활발히 추진하기 시작하였다. 각 금융회사에서도 자체 내부 신용평가시스템(Credit Scoring System, CSS)을 개발하여 대출 승인, 신용카드 발급 및 카드 한도 설정, 금리 결정 등 금융 거래를 처리하기 위한 기준으로 활용하고 있다.

## 2. 금융이력부족자와 대안신용평가시스템(ACSS)

사회초년생, 고령층 등 금융이력부족자(thin-filer, no-filer)에 해당하는 개인은 신용평가에서 주로 활용되는 금융정보가 거의 없기 때문에 신용평점을 산출하기 어려운 구조적 문제에 놓인다<sup>13)</sup>. 이들의 금융접근성을 높이고 대출을 고려하기 위해서는 개인 신용평점을 산출할 수 있는 상환이력, 부채, 신용거래 기간 등을 대신하여 비금융정보를 대리 변수로 활용하여 평가를 진행해야 한다. 이때 고려하는 비금융정보는 성실하게 채무상환을 하는 개인의 행동패턴을 다른 요인들로 찾아내기 위한 자료이다.

현재 약 1,200만 명에 달하는 금융이력부족자들은 전통적인 신용평가 체계 내에서는 충분한 금융거래 정보가 축적되지 않아 신용도 판단이 어려운 계층이다. 이들 중 대부분은 저신용 등급(CB 4~6등급)으로 분류되어 금융 접근성에 제약을 받게 된다. 이러한 금융이력부족자의 신용평가를 보완하기 위해서는 비금융정보의 활용이 필수적이다. 비금융정보를 활용한 개인신용평가 관련 선행연구를 살펴보면 다양한 데이터를 반영하여 실제로 금융이력부족자의 금융접근성이 향상됨을 확인할 수 있다<sup>14)</sup>. 최근에는 통신요금 납부내역, 전기/수도요금 납부내역 등 공공요금 납부정보부터 상거래 내역에 이르기까지 비금융 데이터를 바탕으로 대안적 신용평가 모형을 개발하려는 시도가 확산되고 있다. Political and Economic Research Council(2006)의 연구 결과에 따르면 통신·전기·가스 요금납부 정보 중 적어도 1개 이상의 비금융정보를 반영한 개인신용평가 모형이 기존의 모형에 비해 예측력이 크게 개선되었다<sup>15)</sup>. 기존에 신용평가 이력이 없는 금융이력부족자의 경우 신용평점 산출이 가능해졌으며, 대출승인률이 10% 가량 증가하고 모형 예측력이 향상되며 채무불이행 위험이 감소하는 등 금융접근성 확대의

13) 최성민 “공개된 개인정보의 신용평가 활용과 법적 쟁점.” 금융정보연구 (2023): 40-41p. 참조. 개인의 금융정보를 조사하기 어려운 이유는 공시 의무 등으로 재무제표 등 정량화된 공개 정보가 있는 기업과 달리 전수조사가 어렵고, 정보제공 동의와 정보 수집비용 부담 등으로 인해 활용할 수 있는 정보에 한계가 있기 때문이다.

14) 최성민 “공개된 개인정보의 신용평가 활용과 법적 쟁점.” 금융정보연구 (2023): 43-44p 참조. bill payment data, digital footprint(구매금액, 전자상거래 이용기기 유형 및 운영체제 유형, 이메일 정보 등), phone-related data(전화통화 상세기록 정보 등), geolocation data(GSN 등), psychometric data(성격 특성 등), E-commerce data, hybrid data 등에 대한 실증 연구가 있다.

15) Turner, Michael A., et al. "Give credit where credit is due: Increasing access to affordable mainstream credit using alternative data." The Brookings Institution, Washington, DC (2006).

효과를 거두었다.

해외에서는 신용평가에 있어 비금융데이터의 이용이 활발하게 일어났다. 렌도(Lenddo)의 주요 타겟 고객은 대출상환 역량은 갖추었으나 거래 내역이 부족하여 기존 금융기관에서 대출받기 어려운 개발도상국의 중산층<sup>16)</sup>이다. 신용평가를 위해 SNS 계정 개수, 해당 계정의 사용기간, 친구 수, 일반 커뮤니티, '신뢰할 수 있는 네트워크(Trusted Network)' 등을 활용한다. 시그니파이(Cignifi)는 휴대폰 사용패턴에 따라 신용평가를 수행하고 이를 보험, 소액대출, 신용카드, 기타 금융상품 등에 적용하는 기업이다. 라틴아메리카 등 개발도상국에서 서비스를 제공하며 Cignifi의 알고리즘은 최소 4주 기간의 통화, 문자메시지, 결제 패턴 등을 분석하여 신용을 평가한다. EFL(Entrepreneurial Finance Lab)은 비전통적인 심리 기반 데이터를 사용, 기업가 정신(entrepreneurship)의 핵심 특성인 유동적 지능, 비즈니스 기술, 진실성 및 태도 등을 측정하여 신용을 평가한다<sup>17)</sup>. 우리나라에서도 금융이력부족자 중 청년층의 경우 온라인 구매 행태 데이터가 풍부하므로, 이들의 상거래 내역 등은 신용도 판단에 유용한 정보로 활용될 수 있다는 것을 입증한 연구가 있다<sup>18)</sup>. 실제로 온라인쇼핑 구매 데이터 등의 상거래 데이터를 신용평가에 반영한 '커머스스코어' 모형의 경우 금융이력부족자의 신용도 예측 정확성을 높인 것으로 나타났다<sup>19)</sup>.

이러한 맥락에서 볼 때, 개정 신용정보법 및 시행령에서 주문내역정보까지 개인신용정보의 범위에 포함시킨 것은 데이터 기반 신용평가 체계로의 전환을 뒷받침하는 정책적 변화로 해석할 수 있다. 또한 전송요구권 도입은 금융이력부족자들이 본인의 상거래 정보를 직접 관리하고, 이를 원하는 금융회사나 신용평가사에 제공함으로써 보다 정확한 신용도 판단을 받을 수 있는 기회를 보장한다는 점에서 의의가 있다. 다만 이를 위해서는 단순히 전자금융업자로 한정된 현행 전송요구권의 행사 상대방 범위를 확대하여, 상거래 데이터를 많이 보유한 통신판매업자 등이 전문개인신용평가업에 진입할 필요성이 있다. 비금융 CB<sup>20)</sup>로 불리는 전문개인신용평가업은 비금융 정보로 개인의 신용을 평가해 제삼자에게 제공하는 사업을 뜻한다. 비금융 CB를 활용하면 주부, 사회초년생, 학생 등 금융거래 정보가 부족한 금융 이력 부족자의 금융 접근성이 개선된다. 이러한 요구를 반영하여 통신 3사(SKT, KT, LG 유플러스)는 비금융정보를 활용한 대안신용평가 시장 활성화를 위해 통신 데이터 표준화를 공동으로 추진해 달라는 금융위원회의 제안을 수용, 2022년 8월 합작법인 설립 계획을 발표했다. 통신비 납입

16) Tom Groenfeldt, "Lenddo Creates Credit Scores Using Social Media", Forbes, 2015.06.29., <https://www.forbes.com/sites/tomgroenfeldt/2015/01/29/lenddo-creates-credit-scores-using-social-media/?sh=431fa3ea2fde> (최종접속일 2024.09.20.)

17) Sarah Wheaton, "Credit Score, by Multiple Choice", The New York Times, 2013.12.30., <https://www.nytimes.com/2013/12/31/business/credit-scores-from-a-test-not-a-history.html> (최종접속일 2024.10.12.)

18) 박재만, "비금융 개인신용정보의 전송요구권에 대한 법적 쟁점." 상사판례연구 34.3 (2021): 331-360.

19) 김종윤, 장원중, 김광용 "온라인 상거래 데이터를 반영한 개인신용평가모형(커머스스코어) 개발" 정보화연구 16.1, (2019), pp.45-55

20) 심지영, [단독] 통신 3사 합작 신평사 출범 초읽기, 금융위 문턱 넘었다, 비즈한국, 2023.12.28., <https://www.bizhankook.com/bk/article/26874>(최종접속일 2024.10.12.) 참조.

내역 등 이통 3사의 통신 데이터를 활용한 신용평가 서비스를 제공하여 '우수한 금융 이력부족자'에 속하는 이들이 신용점수를 합리적으로 상승시키고 금융 시스템에 자연스럽게 편입되도록 하는 것이 1차적 목표라고 통신대안평가준비법인은 밝혔다<sup>21)</sup>.

### Ⅲ. 인공지능 기반 신용평가시스템의 문제점

#### 1. 인공지능 신용평가 시스템의 전반적인 문제점

첫 번째로, 인공지능 알고리즘의 불투명성과 해석의 어려움이 존재한다. 알고리즘이 어떤 근거로 신용평가 결과를 도출했는지 명확히 설명하기 어려운 블랙박스 문제가 발생할 수 있다. 인공지능의 알고리즘이나 학습 방식을 이해하기 위해서는 고도의 수학적, 컴퓨터 공학적 지식을 요구한다. 그런데 신용평가시스템 자체 역시 일반인의 관점에서 파악하기 어렵다. 개인신용평가회사(CB)에서 상환이력과 부채수준 등을 근거로 신용점수를 산출하지만, 점수의 등락 이유를 대부분은 알 수 없다. 점수가 변화하는 것을 매번 통지하는 것도 아니고, 소비자가 직접 조회해보아야 확인할 수 있다. 게다가, 금융회사별로 개인신용평가시스템(CSS)이 다르고 이는 영업비밀로 간주되고 있기 때문에 산출 과정과 구조를 밝힐 수가 없는 상황이다<sup>22)</sup>. 신용평가가 어떻게 진행되고 있는지 금융소비자가 모를 가능성이 높는데 인공지능의 결정이 반영된다면 금융회사에서 설명하기도 어렵고, 일반인이 이해하기도 힘들 것이다.

둘째로, 인공지능 알고리즘의 동적 진화로 인해 신용평가 모형의 안정성과 일관성을 담보하기 어려운 점도 문제로 지적된다. 흔히 이러한 '인공지능 시스템의 복원력, 정확성 및 신뢰성'을 '기술적 강건성' (technical robustness)<sup>23)</sup>으로 정의하고 있다. 이러한 의미에서 강건성은 정확성과 안정성을 포함한 다양한 개념을 포괄하는 원칙이라고 할 수 있다<sup>24)</sup>. 환경 변화에 적응하는 알고리즘의 특성상 평가기준이 계속 변화할 수밖에 없다. 인공지능은 입력된 데이터를 가지고 스스로 학습한 뒤 이를 기반으로 알고리즘을 계속 조정하여 더 적합한 결과를 도출하고자 한

21) 장문기, [인터뷰] 문재인 통신대안평가준비법인 대표 "3월 사업개시...개인신용 패러다임 바꿀 것", 아주경제, 2024.01.02., <https://www.ajunews.com/view/20240101153250616>(최종접속일 2024.10.12.) 참조.

22) 강경훈 et al. "신용조회회사(CB)의 신용평가모형과 감독: 주요국과의 비교 분석을 중심으로." 금융정보연구 8.2 (2019): 76p 참조. 개별 CB사 입장에서는 평가모형이 사업을 영위하는 데 필수적인 영업비밀이며 차별성을 달성할 수 있는 거의 유일한 요인이다. 따라서 평가모형의 세부적인 내용에 대한 감독은 현실적으로 가능성도 높지 않을 뿐 아니라 바람직하지 않다.

23) 서정호, "금융업의 인공지능 활용과 정책과제", KIF 정책분석보고서, 2022.2., 30p 참고. 2019년 OECD가 인공지능과 관련된 국제적 규제 가이드라인으로는 처음으로 인공지능 활용과 관련된 권고안('Recommendation of the Council on Artificial Intelligence')을 제시하였고 '신뢰할 만한 인공지능(Trustworthy AI)'을 운영하기 위해 추구해야 할 5가지 기본원칙을 언급하였는데 이 중 강건성(robustness)이 등장하였다.

24) Remolina et al. Artificial Intelligence in Finance: Challenges, Opportunities and Regulatory Developments / Tanna, Minesh & Dunning, William. Bias and discrimination in the use of AI in the financial sector. Cheltenham, UK: Edward Elgar Publishing, 2023.

다. 따라서 시점에 따라 서로 다른 결과를 도출할 수도 있고, 알고리즘이 계속 변화함에 따라 평가 방법에 혼란이 생길 수 있다. 평가 대상이 바뀐다면 평가의 기준과 방법이 계속 변화하여야 하는데, 현실적으로 법과 제도를 계속 그에 맞추어 수정하는 것은 불가능하다.

여기에 더해 더욱 심각한 세 번째 문제는 학습에 사용된 데이터가 편향되어 있을 경우, 그 편향이 신용평가 결과에 그대로 반영된다는 것이다. 과거 데이터에 내재된 차별과 불공정이 알고리즘에 의해 재생산될 수 있다. 이러한 차별에 가장 취약한 것은 금융이력부족자들이다. 금융거래 정보가 없는 사람들은 비금융정보인 각종 개인정보를 제공하여 신용이 있고 대출 등을 상환할 여력이 있음을 증명해야 한다. 그러나 이러한 개인정보 중 성별, 나이, 거주지역, 직업 등의 요소가 인공지능 신용평가시스템에 어떻게 반영될지 알 수 없다. 실제로 인공지능의 성차별은 금융사의 신용 평가에서 문제가 되었다. 미국의 골드만삭스(Goldman Sachs)의 ‘애플카드(Apple Card)’의 경우, 배우자와 자산 및 계좌를 공유하는 고객 중에서, 남성 고객보다 여성 고객이 현저하게 낮은 신용한도를 부여받은 것으로 밝혀졌다. 이에 뉴욕주 금융당국(DFS)은 애플카드가 활용한 인공지능의 성차별에 관한 조사를 착수하였다<sup>25)</sup>. 이렇듯 인공지능에 의한 신용평가가 사회적 차별과 배제를 심화시킬 수 있다는 근본적인 우려가 있다.

## 2. 금융이력부족자의 취약점과 사회적 변화에 따른 과제

금융이력부족자들은 금융정보 대신 비금융정보를 제공하는데, 이러한 비금융 정보는 매우 넓은 범위를 포괄한다. 기존 연구에 따르면, 개인 신용평가에 사용된 비금융정보는 bill payments data(요금 납부 정보), digital footprint(온라인 기록 정보), phone-related data(전화 통화 상세정보 등), geolocation data(위치 정보), psychometric data(심리 측정 결과 정보), E-commerce data(전자상거래 정보), hybrid data에 이르기까지 다양<sup>26)</sup>하며, 개인과 관련된 거의 모든 정보를 활용할 수 있다. 따라서 개인정보 침해와 프라이버시 침해의 우려도 제기된다. 인공지능 신용평가는 방대한 개인정보의 수집과 활용을 전제로 하는데, 이 과정에서 정보주체의 동의권이 형해화될 수 있다. 신용평가의 목적 외 활용이나 부적절한 활용으로 인한 피해도 우려된다. 한 번 생성된 신용평가 정보가 다른 맥락에서 본래 목적과 다르게 사용될 경우 개인의 권리가 침해될 수 있다.

25) 김일우 “고위험 인공지능시스템의 차별에 관한 연구.” 서강법률논총 13.1 (2024): 17-19p.

26) 최성민 “공개된 개인정보의 신용평가 활용과 법적 쟁점.” 금융정보연구 (2023): 43-47p 참조.

[표 3] 금융이력부족자(Thin Filer)<sup>27)</sup> 현황(KCB)<sup>28)</sup>

신용평점 분포			
평점구간	2020년말	2021년말	2022년말
800이상	699명	567명	1373명
750이상	320만9404명	322만7538명	324만154명
700이상	61만29174명	619만1629명	628만6972명
600이상	262만6567명	261만5133명	256만4857명
전체	1196만5844명	1203만4867명	1209만3356명
신용평점 이의제기 건수 및 수용률			
	2020년	2021년	2022년
이의제기건수	56	14	12
수용건수	0	0	0
수용률	0%	0%	0%

우리나라에서 최근 3년 내 신용 미거래 상태로 신용정보를 산출하기 어려운 금융이력부족자는 전체 1,209만 3,356명으로, 이중 73%에 달하는 885만 1,829명이 신용평점 700점 이하 중·저신용자로 분류된다. 하지만 2020년부터 2022년 3개년 동안의 신용평점에 대한 이의제기 수용률은 0%로 개인이 신용평가업자를 대상으로 의견을 반영하는 것은 불가능에 가깝다. 대부분의 금융이력부족자들이 사회초년생 청년이거나, 금융에 대한 접근성이 부족한 주부 또는 고령층이기 때문에 다른 고려 요소 없이 신용평점만 가지고 대출 심사 등을 받는다면 금융 조달의 기회는 더욱 적어진다. 제1금융권에서 이들을 받아주지 않는다면 어쩔 수 없이 제2, 제3금융권으로 밀려날 것이고 최악의 경우 파산신청, 개인회생신청 등을 해야만 하는 안타까운 상황이 벌어질 수 있다.

우리나라의 금융이력부족자들은 현재 주로 특정 연령층이나 성별에 집중되어 있지만, 앞으로 다문화 사회로 나아감에 따라 인종이나 출신 국가에 따른 금융 차별의 문제도 대두될 가능성이 있다. 현재 우리나라의 금융이력부족자들은 주로 사회초년생, 여성, 고령층 등 특정 연령대와 성별에 집중되어 있다. 이는 이들 계층이 상대적으로 금융 접근성이 낮고 안정적인 수입원이 부족하여 신용 기록을 쌓기 어려운 현실을 반영한다. 그러나 한국 사회가 점차 다문화 사회로 이행함에 따라 금융이력부족자의 인구통계학적 특성도 변화할 것으로 예상된다. 한국에 거주하는 외국인과의 이민자 가구가 증가하면서 이들 중 상당수가 금융이력부족자로 분류될 가능성이 높아지고 있다. 특히 국내 금융시스템에 대한 이해도가 낮고 언어 및 문화적 장벽으로 인해 금융 서비스 접근에 어려움을 겪는 이주민들의 경우 더욱 그러하다. 나아가 인종이나 출신 국가를 이유로 금융기관이나 신용평가시스템으로부터 차별적 대우를 받을 위험도 배제할 수 없다. 실제로 미국의 사례에서 볼 수 있듯이, 유색인종과 소수 민족 집단이 주류 금융시장에서 배제되고 불리한 조건의 대출을 받는 경향이 뚜렷하게 나타난다. 물론 한국과 미국의 역사적, 사회적 맥락이 다르기에 단순 비교는 어려

27) 최근 3년내 신용 미거래자 기준

28) 김나경, [단독]'無이력의 악순환'원파일러 99.9%가 신용 800점 이하... "거래이력이 없지, 빚 안 갚는다 했나", 파이낸셜뉴스, 2023.03.15., <https://www.fnnews.com/news/202303151401284618> (최종접속일 2024.10.12.)

우나, 다문화 사회로의 이행이 초래할 수 있는 잠재적 차별 요인에 대해서는 면밀히 살펴볼 필요가 있다.

일찍이 우리나라보다 다인종·다문화 국가였던 미국의 경우, 유색인종 커뮤니티는 서브프라임 대출기관과 금융 주류 밖에서 신용을 이용할 가능성이 높기 때문에 신용 사용 기간과 사용 신용 유형도 제한적이었다. 2015년 소비자금융보호국(CFPB)의 보고서에 따르면 2010년을 기준으로 약 2,600만 명의 미국인이 “신용이 보이지 않는” 상태이고 다른 약 1,900만 명이 신용 점수를 받을 수 없을 정도로 제한된 신용 기록을 가지고 있어, 총 4천 5백만 명이 넘는 미국인이 신용 점수를 받지 못하는 것으로 나타났다<sup>29)</sup>. 이 연구는 또한 아프리카계 미국인과 라틴계 미국인이 백인보다 신용이 보이지 않거나 점수가 없는 신용 기록을 보유할 가능성이 더 높다는 사실을 발견했다. 2018년, 탐사보도센터의 리빌은 2015년과 2016년에 미국인이 일반 모기지 대출을 신청한 거의 모든 경우를 포함하는 3,100만 건의 모기지 기록을 연구했다. 분석 결과 여러 경제적, 사회적 요인을 통제된 후에도 역시 흑인 신청자가 일반 주택 구입 대출을 거부당할 가능성이 백인 신청자보다 거의 3배 더 높다는 사실이 밝혀졌다<sup>30)</sup>. 서른셋 흑인여성 레이첼 파롤(Rachelle Faroul)이 모기지 대출을 거절당한 사례가 대표적이다<sup>31)</sup>. 한국의 금융이력부족자들은 현재 주로 여성과 청년층, 노년층 등의 소외계층이지만 앞으로 점점 우리사회 역시 다문화 사회로 진입하고 있으므로, 미래에 대출 과정에 있어 미국과 같이 인종에 대한 차별이 일어날 수 있음을 시사한다. 따라서 향후 금융이력부족자에 대한 정책을 수립함에 있어 인종, 국적 등에 따른 차별 가능성을 염두에 두어야 한다.

### 3. 인공지능으로 인한 금융이력부족자에 대한 불공정 발생

#### (1) 입력데이터의 편향

입력 단계에서 모델에 입력되는 데이터 분류 과정 상 사회문화적 편향에 기반하여 데이터에 라벨을 붙이고, 이가 반영될 수 있다. 입력 데이터의 편향은 소외계층의 신용점수나 상환 능력이 과소평가되는 문제를 야기한다. 이는 과거 데이터에 존재하는 차별이나 편견으로 인해, 금융이력부족자의 신용도가 실제보다 낮게 측정되기 때문이

29) 이들은 전국적으로 세 곳의 신용 평가 기관에 전부 신용 기록이 없다. Kenneth P. Brevoort, Philipp Grimm & Michelle Kambara, Consumer Fin. Prot. Bureau, Data Point: Credit Invisibles 4-6 (2015), [https://files.consumerfinance.gov/f/201505\\_cfpb\\_data-point-credit-invisibles.pdf](https://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf) [http://perma.cc/J47X-4V89](최종접속일 2024.10.12.) 참고.

30) Emmanuel Martinez & Aaron Glantz, How Reveal Identified Lending Disparities in Federal Mortgage Data, (2018), [https://s3-us-west-2.amazonaws.com/revealnews.org/uploads/lending\\_disparities\\_whitepaper\\_180214.pdf](https://s3-us-west-2.amazonaws.com/revealnews.org/uploads/lending_disparities_whitepaper_180214.pdf)(최종접속일 2024.10.12.)

31) Glantz, Aaron and Emmanuel Martinez, "For People of Color, Banks are Shutting the Door to Homeownership." Reveal News, 2018.02.16., <https://revealnews.org/article/for-people-of-color-banks-are-shutting-the-door-to-homeownership/> (최종접속일 2024.10.12.) 참고. 2016년 4월 레이첼 파롤은 노스웨스턴 대학교에서 학위를 받았고, 신용 점수가 좋았으며, 럿거스대학교의 계약직으로 컴퓨터 프로그래밍을 가르치면서 연간 60,000달러를 벌고 있었다. 그럼에도 불구하고 2015년과 2016년에 백인에게 거의 90%의 대출을 해준 독립 브로커인 필라델피아 모기지 어드바이저스(Philadelphia Mortgage Advisors)는 그녀의 대출 신청을 거부하였다. 백인-일본인 혼혈인 그녀의 파트너가 대출신청서에 서명하고서야 산탄데르 은행에서 대출을 받을 수 있었다.

다. 이를 '측정 오차(measurement error)'<sup>32)</sup>로 표현하기도 한다. 금융이력부족자는 신용거래 기록이 충분히 축적되지 않았기에 이들의 실제 상환 능력이 과소평가될 위험이 있다. 전통적인 신용평가 모델은 주로 금융거래 기록에 의존하지만 금융이력부족자의 경우, 임대료 납부나 공과금 납부 내역 등 비금융적 데이터가 신용도를 가늠하는 중요한 지표가 된다. 이 중 적절한 대안적 데이터가 반영되지 않을 경우, 소외계층의 신용도가 낮게 평가될 수 있다. 하지만 적절한 대안적 데이터가 반영된다고 하더라도 신용평가 모델에 사용되는 변수 선택 과정에서 편향이 개입될 가능성도 있다. 예를 들어, 주거 안정성이나 고용 형태와 같이 소외계층에게 불리하게 작용할 수 있는 변수가 과도하게 반영될 경우, 이들의 신용도가 낮게 측정될 수 있다. 미국의 '레드라이닝' 사례는 입력 데이터의 편향으로 인해 발생할 수 있는 차별과 편견을 잘 보여준다. 레드라이닝(Redlining)은 백인이 아닌 소수 인종이 거주하는 지역에서 금융 서비스를 제공하지 않는 차별적 관행을 뜻한다. 1960년대 사회학자인 존 맥나이트(John McKnight)가 은행이 특정 지역을 위험한 지역으로 분류하거나 거주자의 인종 구성으로 인해 투자할 가치가 없는 지역으로 분류하는 차별적 현상을 설명하기 위해 이 용어를 만들었다. 대부분 아프리카계 미국인을 대상으로 했으며, 명시적으로는 인종차별이 아니었지만 사는 지역에 따라 보험 가입을 거부, 의료서비스를 거부하는 등의 각종 인권 침해가 있었다.

인공지능 알고리즘에 거주지역 정보를 포함시킬 경우, 표면적으로는 중립적 변수이지만 내재적으로는 특정 인종에 대한 차별로 이어질 수 있다. 즉, 알고리즘이 편향된 데이터를 학습하여 과거의 차별을 그대로 반영하거나 심화시키는 결과를 초래할 수 있다. 우편번호와 거주지역과 같은 정보를 가지고도 쉽게 인종차별적인 조치를 불러올 수 있는데, 과연 인공지능에 이러한 영향을 끼칠 수 있는 정보를 전혀 투입하지 않고 신용평가를 산출할 수 있을까? 소비자의 쇼핑 습관, 소셜 미디어 관행 및 기타 세부 사항은 인종 및 기타 민감한 요인에 따라 달라질 수 있다. 예를 들어, 백인의 30%는 휴대전화를 유일한 인터넷 연결 수단으로 사용하는 반면 라틴계는 약 47%, 흑인은 38%에 불과하다. 다른 정보와 결합하면 모바일 및 인터넷 사용 관행이 인종을 나타내는 지표로 사용될 수 있다<sup>33)</sup>. 우리나라의 경우 인종 문제는 아니지만, 특정 지역 출신에 대한 차별적 인식이 존재하고 이에 따라 금융 서비스 접근성에도 격차가 발생할 수 있다. 예를 들어 특정 지역 출신자에 대한 부정적 고정관념으로 인해 이들에 대한 대출 심사가 엄격해지거나 불리한 조건이 제시될 수 있다. 한국에서도 이러한 편향된 데이터가 알고리즘에 반영된다면 출신 지역에 따른 차별이 지속 또는 강화될 우려가 있다.

2012년 신한은행의 학력에 따른 신용차별 사건은 인공지능 기반의 신용평가시스템이 본격적으로 활용되기 전이지만 많은 시사점을 남긴다. 신한은행은 2008년 4월

32) Liu, Lydia T., et al. "Delayed impact of fair machine learning." International Conference on Machine Learning. PMLR, 2018.

33) Hurley, Mikella, and Julius Adebayo. "Credit scoring in the era of big data." Yale JL & Tech. 18 (2016): 158p

새로운 개인 신용평가모형을 도입하여 대출 신청자의 학력을 파악해 고졸이면 13점, 석·박사는 54점 등으로 학력에 따라 신용점수에 차등을 두었다. 감사원은 신한은행이 2008년부터 2011년까지 대출을 거절한 4만4368명 가운데 1만4138명(31.9%)은 학력이 낮아 돈을 못 빌린 것으로 파악했다. 이들이 신청한 대출금은 1241억 원이다. 또 이 기간에 15만1648명의 개인신용대출자 가운데 7만3796명(48.7%)은 학력이 낮다는 이유로 이자를 더 낸 것으로 드러났다<sup>34)</sup>. 2012년 7월, 감사원은 '금융권 역 별 감독실태' 감사를 통해 신한은행이 직업과 급여를 이미 신용평가에 고려하고 있음에도 불구하고, 학력을 추가로 반영하는 것은 동일한 변수를 중복적으로 평가에 포함시키는 문제가 있다고 지적했다. 그러나 금융감독원은 신한은행의 이러한 관행에 대해 적절한 조치를 취하지 않았기에, 감사원은 금융감독원의 감독 소홀을 문제삼았다. 비금융데이터가 점점 더 신용평가에 이용되는 요즈음 추세로 볼 때 만약 어떤 은행이 인공지능 알고리즘을 활용하여 소비자들의 신용도를 산정하였는데, 그 알고리즘에 투입된 데이터에 학력과 같은 차별적인<sup>35)</sup> 요소가 있었다면 이를 어떻게 다뤄야 하는지의 문제가 발생할 수 있다<sup>36)</sup>. 이 사건 이후 우리나라에서 학력은 금융소비자보호법 제15조에 은행뿐 아니라 금융 전체를 아우르는 차별금지사유로 명기되었다.

## (2) 의도하지 않은 결과적 편향

학습 단계에서 모델의 설계자 등이 의도하지 않았지만 여러 가지 모습의 편향이 발생하고 반영될 수 있다. 지나치게 특정 문제를 많이 학습하는 경우 특정 데이터의 가중치가 커지게 되면서 편향성 문제가 발생하고 이를 과적합, 과도학습(over-fitting)이라고 한다. 모형이 학습 데이터에 과도하게 의존하여 학습한 데이터에 대해서는 정확한 예측을 제시하지만 미학습 데이터에 대해서는 부정확하게 예측할 우려가 있다. 알고리즘이 차별적인 결과를 내도록 학습된 경우 향후 데이터 세트에 이러한 편향은 계속 복제된다. 금융이력부족자의 신용평가에 활용되는 비금융 데이터는 전통적인 금융 데이터에 비해 양이 제한적이고 다양성이 부족할 수 있다. 이로 인해 인공지능 모델이 비금융 데이터의 특정 패턴에 지나치게 의존하여 과적합 문제를 일으킬 위험이 있다. 또한 특정 비금융 데이터 요소는 인종과 같은 차별적 요소와 상관관계가 높기 때문에 인종을 명시적으로 고려하지 않더라도 이를 포함하면 차별적 효과를 초래할 수 있다<sup>37)</sup>. 독립적으로 보이는 요인이 인종과 같은 변수에 대한 대리인(proxy) 역할을 할 수 있기 때문에 인공지능 기반의 신용평가 모델은 편견을 지속시킬 수 있다.

34) 박순빈, 신한은행 “고졸은 대출이자 더 내라”, 한겨레, 2012.07.23., [https://www.hani.co.kr/arti/economy/economy\\_general/543895.html](https://www.hani.co.kr/arti/economy/economy_general/543895.html) (최종접속일 2024.10.12.)

35) 외국에서 교육수준이 금융시장의 차별금지사유로 규정된 사례는 없다고 추정된다

36) 고훈수, 정해빈, and 박도현 “인공지능과 차별.” Justice (Seoul, Korea) 171 (2019): 257-259p.

37) Rodriguez, Lorena. “ALL DATA IS NOT CREDIT DATA: CLOSING THE GAP BETWEEN THE FAIR HOUSING ACT AND ALGORITHMIC DECISIONMAKING IN THE LENDING INDUSTRY.” Columbia law review 120.7 (2020): 1843-1884.

머신러닝 모델이 신용평가에 사용될 때 기존의 공정성 기준(demographic parity, equal opportunity)을 따르면 금융소외계층에게 장기적으로 부정적인 영향을 미칠 가능성이 있다. 인구통계학적 동등성(Demographic Parity)은 각 그룹(소외계층과 비소외계층)에게 동일한 비율로 대출을 해주는 것을 의미한다. 그러나 이 기준을 적용할 경우, 금융기관은 소외계층의 상환 능력을 고려하지 않고 과도한 대출을 할 가능성이 있고 이로 인해 소외계층의 부채가 증가하고, 신용점수가 오히려 하락하는 부작용이 발생할 수 있다. 한편 동등한 기회(Equal Opportunity)는 상환 능력이 동일한 사람들 중에서 각 그룹에게 동일한 비율로 대출을 해주는 것을 의미한다. 그러나 소외계층 내부의 신용점수 불균형이 큰 경우, 이 기준을 적용하면 금융기관이 소외계층에게 적정 수준 이하로 대출을 하게 되어, 이들에게 불리하게 작용할 수 있다. 이처럼 단기적인 공정성 기준을 맞추기 위해 소외계층에게 무조건적으로 많은 대출을 해주거나, 반대로 지나치게 적은 대출을 해주는 것은 장기적으로 이들의 신용점수와 경제적 후생을 악화시키는 결과를 초래할 수 있다<sup>38)</sup>.

#### 4. 인공지능에 의한 개인정보의 침해

##### (1) 정보주체의 개인정보 자기결정권

“정보주체”란 처리되는 정보에 의하여 알아볼 수 있는 사람으로서 그 정보의 주체가 되는 사람<sup>39)</sup>을 말하며, “개인정보”란 살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보(해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함함)<sup>40)</sup>를 말한다. 신용평가사에 제공하는 비금융정보는 개인정보에 해당할 수 있고, 따라서 이러한 비금융정보들을 제공하는 정보주체는 개인정보자기결정권을 가진다. 판례는 개인정보자기결정권으로 보호하려는 내용을 기본권들 및 헌법 원리들 중 일부에 완전히 포섭시키는 것은 불가능하므로, 그 헌법적 근거를 굳이 어느 한두 가지에 국한하는 것은 바람직하지 않다고 판시하였다<sup>41)</sup>. 따라서 개인정보자기결정권은 이들을 이념적 기초로 하는 독자적 기본권으로서 헌법에 명시되지 않은 기본권이라고 보고 있다.

공개된 개인정보의 경우, 정보주체가 이를 직접 공개한 행위에 대해 일정한 범위에서 이용 동의 의사가 있는 것으로 해석할 여지가 있다. 그러나 무분별한 개인정보의 활용은 여전히 사생활의 비밀과 자유, 개인정보자기결정권을 침해할 우려가 있다. 이에 대해 세 가지 견해가 있다<sup>42)</sup>. 첫째, 정보주체의 개인정보자기결정권 보장을 위해 엄격한 동의제도를 유지해야 한다는 견해이다. 그러나 이는 규제 준수 비용 상승과 신산업 발전 저해로 이어질 수 있다. 둘째, 대법원 판례와 EU GDPR

38) Liu, Lydia T., et al. "Delayed impact of fair machine learning." International Conference on Machine Learning. PMLR, 2018.

39) 「개인정보 보호법」 제2조 제3호

40) 「개인정보 보호법」 제2조 제1호

41) 헌재 2005. 5. 26. 99헌마513.

42) 최성민 “공개된 개인정보의 신용평가 활용과 법적 쟁점.” 금융정보연구 (2023): 35-73 참조.

등을 근거로 묵시적 동의가 내포되어 있어 추가 동의가 불필요하다는 견해<sup>43)</sup>이다. 이 견해는 개인정보 활용과 그에 따른 편익을 강조하며, 개인정보처리자는 정보주체의 동의가 객관적으로 인정되는 범위에서 별도 동의 없이 공개된 개인정보를 활용할 수 있다고 본다. 그러나 이는 개인정보자기결정권 보장을 약화시킬 수 있다. 셋째, 개인정보자기결정권과 빅데이터 활용 필요성 간의 이익형량에 따라 활용 여부를 판단해야 한다는 주장이며, 최근 다수설로 받아들여지고 있다.

개인 신용평가 영역에 이러한 논의를 적용할 때, 이익형량을 통해 활용 여부를 판단해야 한다는 세 번째 견해가 절충안으로 타당하다. 다만 개별 건에 대한 이익형량은 현실적으로 어려우므로, 불확실성의 증대로 이어질 수 있다. 이익형량 과정에서는 개별 사안별로 상충하는 가치들의 경중을 판단해야 하는데, 명확한 기준이 부재한 상황에서 담당자의 재량에 따라 판단이 좌우될 경우 주관적이고 상황 의존적일 수밖에 없고 유사한 사례에 대해서도 상이한 결정이 내려질 가능성이 있다. 또한 기술 발전과 사회 변화에 따라 개인정보의 가치와 빅데이터 활용의 필요성에 대한 인식도 달라지기 때문에 이익형량의 기준점 자체가 유동적이다. 이는 정보주체의 입장에서는 자신의 정보가 언제, 어떻게 활용될지 예측하기 어려운 상황을 초래한다. 나아가 강력한 빅데이터 활용 요구에 직면한 기업이나 기관은 개인정보 침해 우려를 간과하고 정보주체의 권리를 축소 해석할 유인이 생길 수 있다. 따라서 원칙적으로 추가 동의 절차 없이 자율성을 부여하되 사후 검증을 통해 이러한 불확실성을 통제하여야 한다. 통제의 방안이란, 공개된 개인정보 활용이 평가대상인 개인에게 유리한 경우에만 허용함으로써 이익형량을 고려하는 것이다. 이익에 공익을 포함하게 되면 사회 전체의 편익 증진이라는 공익 차원에서 정보활용이 정당화될 수 있어 개인의 권익 보호가 미진할 수 있기 때문이다.

## (2) 정보주체로부터 정보 제공 및 이용에 대한 미동의

인공지능 서비스 과정에서 발생할 수 있는 개인정보 침해요인을 단계별로 정리, 그 심각도를 관계 실무자 및 전문가를 대상으로 설문조사한 연구에 따르면, 서비스 기획·데이터 설계 및 수집 과정에서 개인정보 침해 요인을 특히 주의해야 한다는 점을 확인할 수 있다. 서비스 기획·데이터 설계단계에서는 AI 디바이스/애플리케이션이 명시적인 정보주체 동의 없이 개인정보를 수집하는 것(84.00%)이 가장 위험하며, 데이터 수집단계에서는 AI서비스 제공 중 정보주체 동의 없이 사생활을 모니터링한 정보를 수집하는 것(87.60%)이 가장 심각한 침해요인이 될 수 있다고 드러났다<sup>44)</sup>.

43) 대법원 2014다235080 부당이득금반환 사건, ‘로앤비 사건’ 참조. “정보주체가 직접 또는 제3자를 통하여 이미 공개한 개인정보는 공개 당시 정보주체가 자신의 개인정보에 대한 수집이나 제3자 제공 등의 처리에 대하여 일정한 범위 내에서 동의를 하였다고 할 것이다. 이와 같이 공개된 개인정보를 객관적으로 보아 정보주체가 동의한 범위 내에서 처리하는 것으로 평가할 수 있는 경우에도 동의의 범위가 외부에 표시되지 아니하였다는 이유만으로 또다시 정보주체의 별도의 동의를 받을 것을 요구한다면 이는 정보주체의 공개의사에도 부합하지 아니하거나 정보주체나 개인정보처리자에게 무의미한 동의절차를 밟기 위한 비용만을 부담시키는 결과가 된다.”

44) 신영진 “지능정보사회의 안전한 인공지능 서비스 구현을 위한 개인정보 침해대응 및 위기관리 컴플라이언스 개발에 관한 연구.” 융합정보논문지 12.5 (2022): 1-14.

개인정보 처리에 대한 동의권은 개인정보보호법 제15조에서 자세히 설명하고 있다. 정보주체의 동의를 받은 경우에만 개인정보를 수집하고 그 수집 목적의 범위에서 이용할 수 있다는 것이다. 따라서 금융소비자는 자신의 개인정보 수집·이용에 대해 동의할 권리가 있으며, 금융회사는 동의를 반드시 받아야 한다. 다만, 제15조 3항에 따라 개인정보처리자는 당초 수집 목적과 합리적으로 관련된 범위에서 정보주체에게 불이익이 발생하는지 여부, 암호화 등 안전성 확보에 필요한 조치를 하였는지 여부 등을 고려하여 대통령령<sup>45)</sup>으로 정하는 바에 따라 정보주체의 동의 없이 개인정보를 이용할 수 있다.

최근 신용정보법 상 정보주체로부터 동의를 받지 않아 문제가 된 사례로는 금융감독원이 우리은행에 기관경고 조치하고 과태료를 부과한 사건이 있다. 우리은행 일부 부서는 2020년 9월부터 2021년 11월까지 오픈뱅킹 서비스를 위해 다른 금융회사로부터 받은 개인신용정보(예·적금 만기일·수신일)를 이용해 광고대상 고객을 선별하고 영리목적의 광고성 정보를 전송했다. 상품 홍보 등 목적으로 사용하는 것에 대해 동의하지 않은 고객 6만8527명(중복인원 제외)에게 광고 문자메시지를 전송한 것이다. 메시지 전송 건수는 총 9만8445건이었다<sup>46)</sup>. 또한, 개인정보보호위원회의 실태 점검 결과, 오픈에이아이 등 6개 인공지능 서비스 기업들의 경우 인터넷에 공개된 데이터를 수집해 자사 인공지능 모델의 학습데이터로 사용해온 것으로 드러났다. 이 과정에서 주민등록번호, 신용카드번호 등 한국 이용자들의 개인정보가 포함될 수 있는 것으로 확인됐다. 온라인상 공개된 이미지·영상 등도 인공지능 모델 학습데이터에 사용될 수 있는데, 학습데이터에서 주요 식별정보를 사전 제거하는 조치가 충분하지 않은 것이다. 또한 이들 기업은 자사 인공지능 서비스의 파인튜닝<sup>47)</sup>을 위해 실제 사람 인력을 투입한 뒤 서비스 이용자의 질문과 이에 대한 인공지능 모델의 답변 내용을 직접 열람·검토하고, 수정하는 방식으로 데이터셋을 만들고 있는 것으로 확인됐다. 이용자들은 본인이 입력한 데이터를 사람이 직접 들여다보는 ‘인적 검토’(processed by human reviewer) 과정이 존재한다는 사실 자체를 알기 어렵기에 이에 대한 명시적 동의가 이루어지지 않는 부분은 사생활 침해로 이어질 수 있다<sup>48)</sup>.

45) 개인정보보호법 시행령 제14조의2(개인정보의 추가적인 이용·제공의 기준 등)

46) 송주오, '미동의 고객에 문자 10만건'...금감원, 우리은행 기관경고 조치, 이데일리, 2024.01.12., <https://m.edaily.co.kr/news/read?newsId=02732246638757064&mediaCodeNo=257>(최종접속일 2024.10.12.). 금융감독원은 우리은행이 영리목적의 광고성 정보 전송행위에 미동의 개인신용정보 이용 외에도 은행 자산 횡령, 사모펀드(PEF) 등 금융투자 불완전 판매, 불건전 영업행위 금지 위반 등을 했다는 이유로 2024년 1월 4일 기관조치와 함께 과태료 8억7800만원을 부과했다.

47) 특정 작업에 적합하도록 이미 훈련된 모델에 추가 학습을 수행하는 방법

48) 박지영, AI가 카드·여권 정보 훑어봤다...‘인간지능’이 답변 검토까지, 한겨레, 2024.03.28., [https://www.hani.co.kr/arti/economy/economy\\_general/1134230.html](https://www.hani.co.kr/arti/economy/economy_general/1134230.html) (최종접속일 2024.10.12.)

#### Ⅳ. 인공지능 기반 신용평가시스템 규제 방향

##### 1. 인공지능 가이드라인을 통한 규제 거버넌스 사례

인공지능에 대한 사전적 행정규제에 관한 쟁점을 검토하여 보면 인공지능의 불투명성으로 개발과정 등에 대해 강제로 조사할 수 있을 정도로 압도적인 공익이 있는 것은 아니라는 결론을 내릴 수 있다. 강압적인 금융당국의 정책보다는, 인공지능 윤리원칙을 통해 알고리즘의 투명성과 강건성 등을 선제적으로 천명하는 것이 기술 개발자들에게 규제의 방향성을 사전에 인식시키는 효과가 크다. 원칙만을 선언하는 비규제적 접근 방식은 집행력과 처벌 규정이 부재하여 규제의 실효성을 담보하기 어려우며, 과도한 지침이나 가이드라인 발표가 오히려 중복규제나 규제충돌의 문제를 야기할 수 있다는 우려 역시 존재한다. 그러나 인공지능 관련 윤리준칙이나 가이드라인은 대부분 입법으로 이어지지 않고 구속력 없는 지침에 그치는 경향이 있어, 실제로는 규제 과잉이나 충돌의 문제를 심각하게 야기하지는 않는 편이다. 오히려 기업들에게 자율 규제의 방향성을 제시함으로써 일정 부분 구속력을 발휘하면서도 혁신에 대한 저해는 최소화하는 규제 수단으로 기능할 수 있다<sup>49)</sup>. 현재 다양한 스펙트럼의 인공지능 기술이 여러 분야에서 상이한 규제 이슈를 안고 상용화되고 있으며, 기술혁신의 향방을 단정 짓기 어려운 상황이다. 이러한 맥락에서 윤리원칙 선언 등 다각적인 방식으로 규제의 방향성을 우선 제시하여 최소한의 위험을 통제하고, 관련 기술과 현장에 대한 정확한 이해를 축적한 후 점진적으로 필요한 분야에 한해 규제를 강화해 나가는 규제 거버넌스로의 전환이 요구된다.

2021년 9월 미국 국가표준기술연구원(NIST)에서는 ‘설명가능한 AI의 네 가지 원칙(Four Principles of Explainable Artificial Intelligence)<sup>50)</sup>’ 초안을 발표하였다. 보고서에서 제시된 설명가능한 AI의 네가지 원칙은 다음과 같다. ①AI 시스템은 모든 산출물에 대해 증거 혹은 그 이유를 제시해야 한다(Explanation). ②시스템은 개별 이용자들에게 의미 있거나 이해할 수 있는 설명을 제시하여야 한다(Meaningful). ③해당 설명은 산출물을 생성하는 시스템 프로세스를 적절하게 반영해야 한다(Explanation Accuracy). ④시스템은 충분히 신뢰할 수 있을만한 산출물을 생성할 것이라는 믿음을 전제하고 운영되어야 한다(Knowledge Limiets). 이러한 원칙을 적용하여 인공지능 시스템의 결과물이 도출된 이유에 대해 이해할 수 있게 된다면 신용평가 분야에는 다음과 같은 긍정적인 효과가 발생한다. 만일 AI가 대출 승인 결정을 한다면 그에 대한 이해는 소프트웨어 설계자가 시스템을 개선하는데 유용할 수 있다. 신청자가 AI의 의사결정 과정에 대한 이해를 필요로 할 때, 예를 들어 승인이나 거절 사유를

49) 방정미 “인공지능 알고리즘 규제거버넌스의 전환 - 최근 미국의 알고리즘 규제와 인공지능 윤리원칙을 중심으로.” 공법연구 49.3 (2021): 375-406. 참조.

50) Phillips, P. et al., “Four Principles of Explainable Artificial Intelligence”, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, (2021), <https://doi.org/10.6028/NIST.IR.8312>, [https://tsapps.nist.gov/publication/get\\_pdf.cfm?pub\\_id=933399](https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=933399) (최종접속일 2024.10.12.)

알고자 할 때, 그 이해는 해당 신청인이 신용도를 높게 유지하는데 도움을 줄 수 있다<sup>51)</sup>.

싱가포르 금융관리국(Monetary Authority of Singapore, MAS)의 경우 2018년 금융기관이 인공지능(AI)과 데이터 분석(AIDA)에 근거한 의사결정을 수행할 때 준수해야 할 보편적 원칙을 제시한 FEAT 원칙을 공표했다. AIDA는 AI와 데이터 분석을 아우르는 개념으로, AI 기반 의사결정에 데이터 분석이 핵심적인 역할을 한다는 점을 강조한 용어이다<sup>52)</sup>. FEAT 원칙은 공정성(Fairness), 도덕성(Ethics), 책무성(Accountability), 투명성(Transparency)의 4대 원칙으로 이루어져 있다. 공정성 원칙은 AIDA 기반 의사결정의 타당성과 정확도 및 편향 최소화를 강조한다. 도덕성 원칙은 AIDA 활용이 기업의 윤리기준에 부합되어야 함을 명시한다. 책무성 원칙은 AIDA 기반 의사결정에 대한 내부적, 외부적 책임 소재를 명확히 할 것을 요구한다. 투명성 원칙은 AIDA 활용에 대한 사전 공개와 정보주체의 요청 시 명료한 설명 제공을 의무화한다.

MAS는 FEAT 원칙에 대한 평가방법론을 포함한 백서와 소프트웨어 툴킷을 제공함으로써 FEAT 원칙의 실질적 현장 적용을 지원하고 있다. 이는 AI 윤리원칙이 선언적 차원에 머물지 않고 구체적이고 실효성 있게 운영되도록 하기 위한 조치로 평가받는다. 특히 MAS는 FEAT 원칙의 실행을 위해 금융기관, 학계, 시민사회 등 다양한 이해관계자가 참여하는 FEAT 거버넌스 위원회를 구성하여 운영하고 있다<sup>53)</sup>. 싱가포르 금융감독청(MAS)의 컴플라이언스 툴킷은 자본시장법(SFA) 하에서 MAS에 대한 승인, 통지 및 기타 규제 제출 사항에 대한 지침을 제공하는 문서이다<sup>54)</sup>. 이 툴킷은 MAS에 제출해야 하는 승인 신청, 통지, 기타 규제 제출 사항들을 구체적으로 열거하고, 관련 법규 조항을 명시하며, 제출 방식과 기한 등을 상세히 안내함으로써 규제 요건의 구체성과 실효성을 높이고 있다. 특히 툴킷은 단순히 규제 요건을 나열하는 데 그치지 않고, 금융기관이 SFA나 관련 규정을 숙지하여 준법 여부를 확인할 것을 당부하는 등 규제 준수에 대한 금융기관의 능동적 자세를 강조하고 있다. 이는 형식적 규제 준수를 넘어 실질적 운영 개선으로 이어질 수 있도록 유도하는 효과가 있다. MAS의 툴킷에는 신용평가회사(Credit Rating Agency, CRA)에 대한 규제 내용도 포함되어 있다. 구체적으로 CRA는 자본시장서비스 라이선스 중 신용평가서비스(providing credit rating services) 부문 라이선스 소지자에 해당하며, 이에 따라 CRA 역시 각종 신청, 통지, 보고 의무 등을 지게 된다. 예를 들어, CRA는 신용평가서비스 라이선스 신청, 주요 경영

51) ScienceOn, 미 국가표준기술연구소(NIST), 인공지능 결정의 설명 가능성에 대한 보고서 발표, 2020.08.31., <https://scienceon.kisti.re.kr/srch/selectPORSrchTrend.do?cn=GTB2020005677>(최종접속일 2024.10.12.)

52) MAS, "Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector", 2018.11.12. <https://www.mas.gov.sg/publications/monographs-or-information-paper/2018/feat> (최종접속일 2024.10.12.)

53) MAS, "MAS introduces new FEAT Principles to promote responsible use of AI and data analytics", 2018.11.12., <https://www.mas.gov.sg/news/media-releases/2018/mas-introduces-new-feat-principles-to-promote-responsible-use-of-ai-and-data-analytics> (최종접속일 2024.10.12.)

54) MAS, Compliance Toolkit for Approvals, Notifications and Other Regulatory Submissions to MAS, 2023.12.07., <https://www.mas.gov.sg/regulation/guidance/compliance-toolkit-for-approvals-notifications-and-other-regulatory-submissions-to-mas-sfa> (최종접속일 2024.10.12.)

진 변동 시 MAS 사전 승인, 지배구조 변경 통지, 재무제표 제출 등의 의무가 있으며 특히 분기별로 기본자본(base capital) 현황을 MAS에 보고해야 한다.

또한 싱가포르 규제대상 금융기관들은 MAS-Tx 시스템을 통해 주기적으로 제출 서류의 만기일 알림을 받고 제출 내역을 확인할 수 있도록 하여 규제 준수의 편의성도 제고하고 있다. MAS-Tx 시스템은 MAS에 대한 각종 신청, 통지, 보고 등을 전자적으로 처리하기 위한 시스템이다. 금융기관은 MAS-Tx에 로그인하여 정기 보고서의 제출 기한을 확인하고 알림을 받을 수 있으며, 과거 제출 내역도 조회할 수 있다. 또한 MAS-Tx에서는 특정 업무별로 사용자 접근권한을 차등 부여할 수 있어, 금융기관이 규제 준수 업무를 체계적으로 관리하는 데 도움이 된다<sup>55)</sup>. 이처럼 MAS-Tx 시스템은 규제 관련 제출물을 전자적으로 관리함으로써 금융기관과 MAS 모두의 행정 부담을 경감하고, 나아가 데이터 기반의 효율적인 감독 체계 구축에 기여하고 있다. MAS 컴플라이언스 툴킷에서 MAS-Tx의 활용법을 구체적으로 안내하는 것은 이 시스템이 싱가포르 금융규제 운영에 있어 핵심 인프라로서 자리매김하고 있음을 보여준다.

우리나라의 경우 아직 금융분야 AI에 특화된 종합적 윤리원칙은 부재한 실정이다. 다만 금융위원회가 2021년 발표한 금융분야 AI 가이드라인에서 싱가포르 MAS의 FEAT 원칙과 유사한 내용들이 일부 반영되어 있다. 그러나 구체적 적용과 실행력 확보를 위한 평가방법론이나 툴킷 등은 아직 준비되지 않고 있다<sup>56)</sup>. MAS의 사례는 금융분야 AI 윤리원칙의 제정과 이의 실효적 이행을 위한 구체적 수단 마련이 병행되어야 함을 시사한다. 우리나라도 금융분야에 특화된 AI 윤리원칙을 제정하되, 평가방법론 개발과 평가 지원 인프라 구축을 동시에 추진함으로써 AI 윤리원칙의 현장 적용력과 실효성을 제고할 필요가 있다. 또한 FEAT 원칙의 핵심 내용인 공평성, 도덕성, 책무성, 투명성 요건들이 향후 마련될 금융분야 AI 윤리원칙에 충실히 담길 필요가 있다.

MAS-Tx와 유사하게 한국에서 금융기관이 이용하는 전자신고시스템에는 금융감독원의 '전자공시시스템(DART)'<sup>57)</sup>이 있다. DART는 각종 보고서 및 공시 자료를 전자적으로 제출·관리하기 위한 것으로, 금융회사는 이를 통해 경영공시, 업무보고서, 재무제표 등을 제출하고 조회할 수 있다. 또한 한국은행의 '금융정보통계시스템(FISIS)'<sup>58)</sup>도 금융기관의 업무보고서 등을 전자적으로 수집·관리하는 시스템이다. 이 밖에도 금융위원회와 금융감독원은 '금융규제·법령해설포털'<sup>59)</sup>을 통해 행정지도를 예고하고 입법 및 규정변경을 예고하고 있다. 이처럼 한국에도 금융감독 및 보

55) 위의 문서 4p참조. FIs may login to MAS-Tx to see the due dates and obtain email reminders for regular returns; and to obtain copies of their submissions. FIs may also limit user access to a specific category of transactions only.

56) 연태훈, “금융에서의 AI 확산에 내재한 잠재적 위험과 관리 방안의 모색”, KIF 분석보고서, 2024-01, 한국금융연구원 참조.

57) 전자공시시스템(DART), <https://dart.fss.or.kr/introduction/content1.do>

58) 금융정보통계시스템(FISIS), <https://fisis.fss.or.kr/fss/fsview/indexw.html>

59) 금융규제·법령해설포털, [https://better.fsc.go.kr/fsc\\_new/main/main.do](https://better.fsc.go.kr/fsc_new/main/main.do)

고 업무의 전자화를 위한 다양한 시스템이 구축되어 있지만 각 시스템이 개별 업무별로 분산 운영되고 있어, MAS-Tx처럼 단일 플랫폼에서 모든 규제 관련 제출물을 통합 관리하지는 않는다. 또한 MAS-Tx의 제출 서류의 만기일 알림이나, 특정 업무별로 사용자 접근권한을 차등 부여하는 기능은 사용자 편의를 통해 컴플라이언스를 유도할 수 있기 때문에 도입을 검토할만하다. 향후 금융기관의 편의성 제고와 감독 업무 효율화를 위해 보다 통합적인 전자신고시스템 구축을 위해 노력하여야 할 것이다.

## 2. 금융이력부족자를 위한 인공지능 신용평가시스템의 공정성 기준

### (1) 법률상 차별 심사 기준

미국과 유럽에서는 법률상 명시된 차별금지사유를 직접적인 원인으로 삼아, 동등하게 대우받았어야 마땅할 상대방을 차등적으로 대우하는 것을 직접차별(direct discrimination<sup>60</sup>)이라고 규정한다. 반면 간접차별(indirect discrimination<sup>61</sup>)은 법령상 금지되지 않은, 일견 중립적인 사유에 의한 차별이므로, 어떠한 차등적 대우가 간접차별이라고 주장하는 원고는 그 행위가 ① 본질적으로 동일하거나 유사한 지위에 있는 경우와 비교할 때 동등하게 대우받아야 할 사람(또는 집단)에 대한 차등적 대우이며, ② 법령상 차별금지사유로 보호받는 집단(protected group)에 특정한 차별적 영향(particular disadvantage) 또는 차등적 효과(disparate impact)를 끼침을 증명함으로써 간접차별임을 주장할 수 있다<sup>62</sup>. 미국 법원들은 차별적 대우(disparate treatment) 또는 차별적 효과(disparate impact)가 입증되면 차별을 인정한다. 차별적 효과란 형식적으로는 중립적으로 대우하더라도 결과적으로 보호대상 집단을 불균형하게 배제하거나 불이익을 주는 경우를 의미한다. 차별적 대우는 차별적 의도나 편견을 가지고 공개적으로 차별을 가하는 공공연한 증거(overt evidence) 뿐 아니라 그러한 의도나 편견 없이도 차별금지사유에 기초하여 다르게 대우한다는 비교적 증거(comparative evidence), 정황증거(circumstantial evidence)에 의해 입증될 수 있다<sup>63</sup>.

우리 법에서는 신용정보법 제22조의4(개인신용평가회사의 행위규칙)에서 “성별, 출신지역, 국적 등으로 합리적 이유 없이 차별하는 행위”를 해서는 아니된다고 규정하고 있다. 그런데 이러한 조항이 간접차별 금지 의무까지 포함하는 것인가가 문제가 된다. 우리 헌법에 간접차별에 대하여 명시된 바는 없으며, 개별 법률에서 명시적으로 간접차별 금지를 규정하지 않았다면 금지한다고 보기는 어렵다<sup>64</sup>. 그렇다면 단순히 차별금지 변수(성별, 출신지역, 나이 등)를 신용평가 과정에서 인공지능에

60) McDonnell Douglas Corp. v. Green(1973), 적격을 갖춘 흑인 전기공의 채용을 거절한 사건에서 유래. 인종, 피부색, 종교, 성별 또는 출신 국가에 따른 고용 차별을 금지하는 미국 연방법 제7장의 문구 “prohibits employment discrimination “because of” certain reasons”를 정의한 계기가 되었다.

61) Griggs v. Duke Power Co.(1971)는 흑인의 고등학교 졸업 비율이 낮은 지역에서, 직무능력과 무관한 고등학교 졸업장을 요구한 사건으로 차별적/차등적 효과를 고려한 첫 판결로 간주된다.

62) 고태수, 정해빈, and 박도현 “인공지능과 차별.” Justice (Seoul, Korea) 171 (2019): 220p.

63) 박상철 “금융AI의 활용과 금융소비자 보호: 차별금지, 설명요구권, 6대 판매규제 준수를 중심으로.” BFL 107 (2021): 40p 참조.

64) 이재희, 헌법상 평등보장과 간접차별금지, 헌법재판연구원, 2020 참조

입력하지 않으면 차별이 아닌지가 실무상 쟁점이 될 수 있다.

## (2) 입력 변수 통제의 한계와 산출물에 대한 차별 심사

인공지능 알고리즘의 공정성을 확보하기 위해 직접차별을 피하기 위해서 차별금지 변수를 직접 입력값으로 활용하지 않는 것만으로는 충분하지 않을 수 있다. 그 이유는 다른 변수들을 통해 차별금지 변수를 간접적으로 추론할 수 있는 대용변수(proxy variable) 문제가 발생할 수 있기 때문이다. 예를 들어, 대출 심사 알고리즘에서 성별 정보를 직접 사용하지 않더라도 성별과 강한 상관관계를 갖는 주문내역, SNS 등을 통해 추론할 수 있으며 해외 송금 내역 등을 통해 출신국가 역시 알 수 있어 불리한 결정을 내릴 가능성이 있다. 실제로 미국 모기지 대출 데이터 분석 결과, 인종변수가 입력된 경우와 인종 변수가 입력되지 않은 경우 채무 불이행 위험도 분포에 거의 영향을 끼치지 않았다<sup>65)</sup>. 이처럼 명시적으로 차별금지 변수를 사용하지 않아도 다른 변수들의 조합을 통해 해당 정보를 유추할 수 있다. 대용변수 문제는 인공지능 알고리즘의 공정성을 입력 변수만으로 통제하는 데에는 한계가 있음을 시사한다. 차별금지 변수와 상관관계가 있는 수많은 변수들을 모두 제거하는 것은 현실적으로 불가능할 뿐더러, 그렇게 할 경우 알고리즘의 성능이 크게 저하될 수 있다.

따라서 입력 변수의 통제와 더불어, 알고리즘의 결과물이 특정 집단에 차별적이지 않은지 면밀히 검토하는 과정이 필요하다. 그러나 산출물을 기준으로 차별 여부를 판단하는 것도 쉽지 않으며, 그 어려움은 크게 다음의 세 가지로 요약할 수 있다<sup>66)</sup>. 첫째, 실제법적인 측면에서 인공지능을 활용한 차별은 기존의 차별 판단 기준을 교묘히 피해갈 수 있다. 차별적 결과가 나타났다 하더라도 알고리즘 사용자가 고의성이 없었다고 주장하거나, 의도를 숨김으로써 책임을 회피할 수 있다. 현행법상 차별 성립을 위해서는 특정 집단을 불리하게 대우하려는 의도나 고의가 필요한데, 알고리즘에 의한 자동화된 의사결정에서는 이를 입증하기가 매우 어려울 수 있다. 둘째, 절차법적인 측면에서도 인공지능 알고리즘의 불투명성과 기업의 영업비밀 보호로 인해 재판 과정에서 차별 사실을 규명하기가 쉽지 않다. 특히 심층신경망 기반의 알고리즘은 내부 논리를 설명하기 어려운 '블랙박스'로 여겨지는데, 기업이 이를 영업비밀로 보호할 경우 사법부가 알고리즘의 작동 방식을 정확히 파악하기 어려울 수 있다. 셋째, 사전 규제의 관점에서도 훈련 데이터 관리와 알고리즘 설계에 많은 어려움이 예상된다. 현재 인공지능 모델의 학습에는 대량의 데이터가 필요한데, 이는 대부분 과거에 수집된 데이터일 수밖에 없다. 그런데 과거 데이터에는 이미 사회적 차별과 편향이 반영되어 있는 경우가 많아, 이를 학습한 알고리즘이 기존의 차별을 그대로 답습하거나 심지어 강화할 위험이 있다. 따라서 훈련 데이터에서 차별적 요소를 사전에 제거하고, 차별 없는 의사결정이 가능한 알고리즘을 설계하는 일이 중요하지만 이는 결코 쉽지 않은

65) Gillis, Talia B. and Spiess, Jann L. "Big Data and Discrimination," University of Chicago Law Review: Vol. 86: Iss. 2, Article 4. (2019), 459-488.

66) 고희수, 정해빈, and 박도현 "인공지능과 차별." Justice (Seoul, Korea) 171 (2019): 230-231p를 참고하여 요약함.

과제다.

### (3) 공정성 기준에 대한 제언

인공지능 모델의 차별 문제를 해결하기 위한 기존 규제 접근법에는 몇 가지 한계가 존재한다. 우선 차별금지 변수를 단순히 제외하는 입력 중심 접근법은 실효성이 떨어진다. 차별 변수와 상관관계가 높은 대용변수가 모델에 투입될 경우 차별이 지속될 수 있기 때문이다. 특히 대용변수 조작이 쉬운 빅데이터 환경에서 이러한 문제는 더욱 두드러진다. 나아가 다른 변수들에 편향이 존재할 경우 차별 변수의 제외는 오히려 차별을 악화시킬 수도 있다. 둘째로, 현행 차별금지법은 모호한 기준과 사후적 판단에 의존하는 경향이 있어 인공지능 개발 과정에서 사전적으로 준수하기 어렵다. 복잡한 알고리즘의 전처리, 훈련, 후처리 단계에 차별금지 원칙을 적용하려면 명확한 수리적 기준이 필요한데, 이를 제공하지 못하는 것이 현실이다. 마지막으로, 모델의 해석가능성 부족으로 인해 차별 여부에 대한 추적과 평가가 어려워진다. 이는 결과적으로 입증의 난이도를 높이고 규제 준수 유인을 떨어뜨린다. 이 부분은 인공지능 모델의 '블랙박스' 특성과 직결된 문제라 할 수 있다<sup>67)</sup>.

인공지능 기반 신용평가시스템의 공정성을 평가하기 위해서는 적절한 평가지표를 선정하고 측정하여, 불균형이 발견된 경우 공정성 개선을 위한 기술적·관리적 노력을 기울여야 한다. 서비스 특성에 따라 공정성 판단기준도 달라질 수 있는데, 특정 계층이나 집단에 대한 결과적 평등을 고려할 필요가 있는 경우(금융 소외계층의 금융접근성 제고 등)에는 인구통계적 동등성 기준을 적용하여 집단 간 대출승인을 동등하게 만들 수 있다. 반면, 소비자에게 금융거래 기회를 제공하는 기능을 수행하는 경우(신용평가, 대출심사, 보험심사 등)에는 기회 균등 기준을 적용하여 집단 간 재현율을 동등하게 만들어야 할 것이다<sup>68)</sup>. 세 가지 공정성 지표(독립성, 분리성, 충분성)를 동시에 모두 충족할 수는 없다. 따라서 어떤 지표를 금융AI 모델에 의한 부당한 차별 여부의 기준으로 삼을지 신중한 고민이 필요하다. 인공지능 시스템의 성능 평가 시에는 오류 유형 간 통계적 상충관계를 고려하여 적합한 성능 목표와 평가지표를 선정하고, 그 충족 여부를 확인해야 한다. 예를 들어, 은행이 대출 부적격자를 걸러내기 위해 여신심사를 강화하는 경우, 대출적격자(True)에 대한 여신이 거절(False)되는 경우도 증가할 수 있다. 서비스 특성에 따라 성능 판단기준이 달라질 수 있는데, 소비자에게 금융거래 기회를 제공하는 기능을 수행하는 경우(신용평가, 대출심사, 보험심사 등)에는 False Negative 오류<sup>69)</sup>를 최소화해야 한다<sup>70)</sup>.

67) 박상철 “금융AI의 활용과 금융소비자 보호: 차별금지, 설명요구권, 6대 판매규제 준수를 중심으로.” BFL 107 (2021): 34-35p 참조.

68) 금융위원회, [보도자료] 「금융분야 인공지능(AI) 가이드라인」이 시행됩니다. - 금융권 인공지능(AI) 활용을 활성화하고 인공지능(AI) 기반 금융서비스에 대한 신뢰를 제고하기 위한 모범규준 마련·발표 -, 2021.07.08., <https://www.fsc.go.kr/no010101/76206?srchCtgr=&curPage=21&srchKey=&srchText=&srchBeginDt=&srchEndDt=> (최종접속일 2024.10.12.)

69) 거짓 음성 오류. 양성 반응이 나와야 할 결과가 잘못되어 음성으로 나오는 오류

70) 각주 68의 보도자료 참조.

자질(merit)을 전혀 고려하지 않고 결과의 평등만을 추구하는 독립성 기준을 따르기에는 무리가 있다. 독립성은 오히려 더 차별적인 결과를 초래할 수 있기 때문이다. 예컨대, 비율 맞추기에만 급급하여 한쪽 성별에 대해 자질이 부족한 사람들에게 과도하게 대출을 해줄 경우, 해당 성별의 채무불이행률이 높아져 지속적인 편향을 야기하고 차별을 더욱 심화시킬 수 있다. 이에 자질을 고려하면서 결과의 동등보다는 정확도와 오류율의 동등을 지향하는 기준인 분리성이나 충분성 중에서 선택해야 한다. 신용평가 과정에서는 자질을 갖춘 사람 중 거절되는 비율(1-sensitivity)<sup>71)</sup>이 특정 성별이나 장애인에게서만 높아서는 안 된다는 점이 주요 관심사이므로, 분리성, 그 중에서도 기회균등이 가장 중요한 판단기준이 되어야 한다. 특히 분리성 기준은 두 그룹의 수용자반응특성 곡선(ROC curve)<sup>72)</sup>이 교차하는 지점에서 충족되며, 모델 훈련 완료 후 사후처리 방식으로 조정이 가능하여 개발자가 구현하기에 용이하다는 뚜렷한 장점도 지니고 있다<sup>73)</sup>.

### 3. 금융이력부족자의 개인정보보호

#### (1) 자동화된 결정에 대한 정보주체의 권리 도입

개인정보보호위원회는 개인정보보호법 시행령에 ‘제37조의 2’ 항목을 신설해 자동화된 결정에 대한 정보주체의 권리행사에 필요한 사항을 구체화했다. 자동화된 결정은 사람의 개입 없이 자동화된 시스템에서 개인정보 분석과 같은 과정을 거쳐 개인정보처리자가 정보주체의 권리·의무에 영향을 미치는 최종적인 결정을 한 경우를 의미한다. 즉, 정보의 분석 가공 등 실질적인 자동화 처리 과정을 거쳐 의미 있는 정보를 추출해 이뤄지는 결정이 ‘자동화된 결정’이라는 것이다<sup>74)</sup>. 인공지능을 염두에 둔 규정을 명시적으로 포함한 것으로, EU의 개인정보 보호법(GDPR) 제22조를 참고한 것이다. 시행령은 개인정보처리자가 완전히 자동화된 결정을 하는 경우에는 인터넷 홈페이지 등에 자동화된 결정이 이루어진다는 사실과 목적 및 대상이 되는 범위 주요 개인정보 유형과 결정의 관계 고려사항 및 개인정보 처리 절차 민감정보 또는 14세 미만 아동의 개인정보를 처리하는 경우, 그 목적 및 처리하는 개인정보의 구체적인 항목 정보주체의 거부·설명 등의 요구 가능과 그 방법 및 절차에 대해 공개해 투명성 확보와 정보주체의 권리 행사를 가능하게 했다.

그러나 EU GDPR 제22조와 한국 개인정보 보호법 제37조의2는 자동화된 결정에 대해 상이한 입장을 취하고 있다<sup>75)</sup>. GDPR은 원칙적으로 개인정보처리자의 자동화

71) 민감도(Sensitivity)는 Recall, True Positive Rate라고도 하며, 오차행렬에서 TP(참양성)을 TP(참양성)+FN(거짓음성)으로 나눈 값이다.

72) AUC-ROC 곡선은 재현율과 특이도(TN/(TN+FP))를 기준으로 그려지는 곡선으로, ROC 곡선 밑의 면적이 AUC이다. AUC 수치가 높을수록 모델의 성능이 좋다. 금융보안원, 금융분야 AI 보안 가이드라인, 2023.4.17., 16p 참조

73) 박상철 “금융AI의 활용과 금융소비자 보호: 차별금지, 설명요구권, 6대 판매규제 준수를 중심으로.” BFL 107 (2021): 44p 참조

74) 김대은, [개인정보 보호법 시행령 2차 개정②] AI 결정에 대한 국민 권리 보장 및 CPO제도 개선, 2024.03.22, <https://kidd.co.kr/news/236128> (최종접속일 2024.10.12.)

75) 구태언, 전응준, 유창하, 개인정보 보호법상 자동화된 결정에 대한 정보주체의 대응권, 법률신문, 20

된 결정을 금지<sup>76)</sup>하고, 예외적으로 정보주체의 동의, 계약 이행의 필요성, 법률적 근거가 있는 경우에 한해 이를 허용한다. 그리고 자동화된 결정이 허용되는 경우에도 정보주체에게 이의제기권을 부여하여 보호 장치를 마련<sup>77)</sup>하고 있다. 이는 AI의 투명성과 공정성 확보, 그리고 인간이 기계에 의해 판단 받아서는 안 된다는 이념에서 비롯된 것이다. 최근 EU사법재판소(CJEU)는 독일의 신용평가회사 SCHUFA의 개인 신용평가 점수 자동 산정 행위를 GDPR 제22조의 자동화 결정(ADM, Automated individual Decision-Making, including profiling)으로 인정하였다. 비록 SCHUFA는 개인 신용을 평가할 뿐 대출 여부를 직접 결정하지는 않지만, 은행 등이 SCHUFA의 평가에 크게 의존하여 대출 결정을 내리기 때문에 SCHUFA의 행위가 ADM에 해당한다고 본 것이다. 이에 따라 대출 거절을 당한 정보주체는 SCHUFA에게도 그 근거에 대한 설명을 요구할 수 있게 되었다. 이는 정보주체의 권리 보호를 위해 은행뿐 아니라 실질적 정보를 보유한 신용평가기관에도 설명 의무를 부과하려는 취지로 보인다.

반면, 한국 개인정보 보호법은 자동화된 결정을 원칙적으로 금지하지 않는다. 다만 정보주체의 권리의무에 중대한 영향을 미치는 경우에 한해 거부권과 설명요구권을 인정하고 있다. 그러나 이마저도 사전 동의, 계약 이행의 필요성, 법률적 근거가 있는 경우에는 제한된다. 이는 GDPR에 비해 정보주체의 권리 보호 수준이 낮다고 볼 수 있으며, 개인정보처리자가 형식적 동의를 통해 거부권을 제한할 여지가 있다. 양 법률 모두 자동화된 결정에 대한 거부권과 설명요구권을 최종 결정 단계에서만 인정하고 있어, 그 이전 단계의 평가나 조치에 대해서는 정보주체의 권리가 미치지 못하는 한계가 있다고 평가할 수 있다.

## (2) 공개된 개인정보를 활용한 신용평가와 관련한 과제

### 1) 개인정보의 활용 과정별 개선 방안

개인정보의 수집 및 처리 측면에서는 신용정보법 제15조와 동법 시행령 제13조의 '수집 및 처리의 원칙'에 따라 신용정보주체의 동의를 받지 않아도 되는 경우와 그 기준을 명확히 할 필요가 있다. 공개된 개인정보의 수집 방법으로는 정보주체의 직접 제

24.03.06., <https://www.lawtimes.co.kr/LawFirm-NewsLetter/196490> (최종접속일 2024.10.12.)

76) 제22조 1항의 내용은 다음과 같다. 정보주체는 프로파일링 등, 본인에 관한 법적 효력을 초래하거나 이와 유사하게 본인에게 중대한 영향을 미치는 자동화된 처리에만 의존하는 결정의 적용을 받지 않을 권리를 가진다. (“The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.”), [https://gdpr-text.com/ko/read/article-22/#comment\\_gdpr-a-22\\_1](https://gdpr-text.com/ko/read/article-22/#comment_gdpr-a-22_1) (최종접속일 2024.10.12.)

77) 제22조 3항의 내용은 다음과 같다. 제2항 (a)호 및 (c)호의 사례의 경우, 컨트롤러는 정보주체의 권리와 자유 및 정당한 이익, 최소한 컨트롤러의 인적 개입을 확보하고 본인의 관점을 피력하며 결정에 대해 이의를 제기할 수 있는 권리를 보호하는 데 적절한 조치를 시행해야 한다. (“In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.”)

출, 스크래핑이나 API 등 전산적 방법을 통한 수집, 마이데이터 서비스를 통한 수집 등이 있다. 이 중 스크래핑이나 API 등을 통한 수집의 경우, 저작권법상 공정이용에 해당 하는지, 해당 온라인 서비스 사업자의 이용약관에 따른 제한이 있는지 등을 검토해야 한다. 또한 수집된 텍스트 정보를 개인신용정보로 볼 수 있는지도 따져봐야 한다. 한편, 동의면제 법적 기준에 '개인에게 유리한 결과 활용시'라는 항목을 신설하는 방안도 제안된다.

활용 및 검증 측면에서는 비금융정보 전문 개인 신용평가업의 활성화를 위한 노력이 필요하다. 현재 법·제도적 불확실성으로 인해 관련 시장이 활성화되지 않고 있는데, 비금융정보를 종합신용정보집중기관에서 일괄 수집·가공하여 제공하는 방안이 도움이 될 수 있다. 이는 미국의 신용정보회사(CRC)<sup>78)</sup> 사례를 참고할 만하다. 아울러 개인신용평가체계 검증위원회의 공개정보 신용평가 활용 적정성 검증도 이루어져야 한다.

## 2) 해외 법제와의 비교

유럽과 미국의 관련 법제와 비교했을 때, 한국의 신용평가 관련 정보주체 권리 보호 제도에는 다음과 같은 개선점이 있어 보인다.

[표 5] EU, 미국, 한국 법제도 장단점 비교

구분	EU GDPR	미국 FCRA와 ECOA	한국 신용정보법과 개인정보보호법
장점	<ul style="list-style-type: none"> <li>포괄적이고 강력한 개인정보 보호 규정으로 정보주체의 권리를 두텁게 보호</li> <li>개인정보 처리의 모든 단계에서 정보주체의 동의와 통제권 강조</li> <li>높은 수준의 과징금으로 기업의 적극적 법 준수 유도</li> </ul>	<ul style="list-style-type: none"> <li>신용정보의 공정하고 정확한 사용을 위한 구체적 규정 마련</li> <li>차별금지 조항으로 신용거래에서의 공정성 제고</li> <li>위반 시 소비자의 손해배상 청구권 명시로 권리구제 수단 마련</li> </ul>	<ul style="list-style-type: none"> <li>신용정보와 개인정보 보호를 위한 별도 법률로 체계적 규율</li> <li>자동화평가에 대한 설명과 이의제기권 보장으로 알고리즘에 의한 차별 방지 노력</li> <li>개인정보에 대한 정보주체의 권리를 구체적으로 명시</li> </ul>
단점	<ul style="list-style-type: none"> <li>엄격한 규제로 기업 부담 증가, 혁신과 경쟁력에 영향 우려</li> <li>회원국 간 해석과 적용 차이로 인한 법적 불확실성 존재</li> </ul>	<ul style="list-style-type: none"> <li>GDPR 대비 개인정보 보호 규정 상대적으로 미흡</li> <li>포괄적 연방 차원의 개인정보 보호법 부재로 규제 공백 존재</li> </ul>	<ul style="list-style-type: none"> <li>개인정보 처리 동의 제도의 형식적 운영 경향</li> <li>신용정보법상 금융회사 의무 조항의 추상성과 포괄성</li> <li>개인정보 침해에 대한 제재 수준이 GDPR 대비 낮음</li> </ul>

유럽의 GDPR과 비교해보면, GDPR은 자동화된 의사결정에 대해 정보주체의 거부권을 명시하고 있으나 한국은 이에 대한 법적 근거가 명확하지 않다. 시행령 개정을 통해 거부권과 설명요구권 등을 보다 구체화할 필요가 있다. 또한 GDPR은 개인정보의 국외이전에 대해 엄격한 제한 규정을 두고 있으나, 신용정보법은 관련 규정이 미

78) 미국은 개인의 신용도를 판단할 수 있는 다양한 금융, 비금융 정보를 수집, 가공하여 판매하는 신용정보 회사인 CRC(Credit Reporting Companies)라는 기관이 존재하며 이들은 주로 직접 신용을 평가하거나 대출을 실행하는 역할을 수행하지 않고, 수집된 정보를 신용보고서 형태로 만들어 금융회사에 제공한다.

비하다. 신용정보의 역의이전에 대한 보호 장치를 강화해야 한다. 아울러 개인정보 처리에 대한 과징금, 손해배상 등 위반 시 제재 수준이 높고 법적 구속력이 강한 반면, 국내 제도는 책임규정이 상대적으로 약하다<sup>79)</sup>. 정보주체 권리 침해에 대한 실효성 있는 구제수단이 필요하다. PSD2와 비교해보면, PSD2는 개방형 API를 통한 금융 정보 접근·활용을 활성화하고 있으나, 국내에는 마이데이터 도입 외 관련 인프라가 부족하다. 정보주체 통제권 보장과 데이터 활용이라는 두 가지 정책 목표의 조화가 필요하다. 또한 PSD2는 강력한 인증 의무화 등 보안 규제를 강화하고 있으나, 신용정보법은 관련 기준이 모호하다. 비금융정보를 포함한 신용평가 과정 전반의 데이터 보안 체계를 정비해야 한다.

우리 법 제도를 미국의 FCRA/ECOA와 비교해보면, FCRA(Fair Credit Reporting Act)는 소비자에 대한 신용정보 고지 의무, 이의제기권, 부정확정보 정정청구권 등을 상세히 규정하고 있다. 국내에도 개인신용평가 대응권이 도입되었으나 행사 방법, 금융사의 거부 사유 등이 포괄적이다. 권리 행사의 실효성 제고를 위한 구체적 기준 마련이 필요하다. 또한 ECOA(Equal Credit Opportunity Act)는 신용거래에서의 차별금지 원칙을 명확

79) GDPR의 손해배상과 과징금 관련 규정은 다음과 같다.

손해배상(제82조) GDPR을 위반하는 처리로 인해 물질적 또는 비물질적 손해를 입은 정보 주체는 이러한 처리로 인한 손해에 대해 컨트롤러나 프로세서로부터 손해배상을 받을 수 있다. 단, 컨트롤러 또는 프로세서는 해당 손해를 발생시킨 사건에 대해 책임이 없음을 입증하면 책임이 면제될 수 있다. 복수의 컨트롤러 또는 프로세서가 손해를 발생시킨 경우, 정보주체의 실효적 배상을 위해 각각의 컨트롤러와 프로세서는 모든 손해에 대한 책임을 부담한다. 이러한 경우, 하나의 컨트롤러나 프로세서가 모든 배상을 하고 나서 다른 컨트롤러나 프로세서에게 구상권을 행사할 수 있다.

과징금(제83조) 과징금 부과 권한은 EU 감독기구에 있다. 단, 법체계에 과징금에 대한 근거가 없는 회원국의 경우에는 법원이 부과할 수도 있다. 컨트롤러나 프로세서가 고의 또는 과실로 GDPR의 여러 규정을 위반한 경우, 과징금 총액은 가장 중요한 위반에 규정된 금액을 초과하여서는 안 된다. 위반 내용에 따른 과징금은 [표 6]을 참조.

**[표 6] EU GDPR 위반내용에 따른 과징금**

과징금	위반 내용
전세계 연간 매출액 2%와 1천만 유로 중 높은 금액	컨트롤러 및 프로세서 의무 위반
	인증 기구 의무 위반
	모니터링 기구 의무 위반
전세계 연간 매출액 4%와 2천만 유로 중 높은 금액	동의를 포함하여, 개인정보 처리 기본원칙 위반
	정보주체의 권리 보장 의무 위반
	제 3국에 개인정보 이전 시 준수 의무 위반
	제9장에 따라 채택된 EU 회원국 법률 의무 위반
	-감독기구가 내린 명령 또는 정보 처리의 제한 불복 -개인정보 이동 중지, 열람 기회 제공 의무 위반

벌칙(제84조) 위 83조 과징금 대상에 해당하지 않는 위반 사항에 대해 GDPR은 각 회원국에게 추가적인 처벌을 규정 의무를 부과하고 있으며, 각 회원국은 GDPR 위반 행위에 대한 사법 제재 관련 규칙을 제정하여 시행할 수 있다.

네이버프라이버시센터, [https://privacy.naver.com/global\\_support?menu=global\\_support\\_eu\\_gdpr\\_understand](https://privacy.naver.com/global_support?menu=global_support_eu_gdpr_understand) (최종접속일 2024.10.12.)를 참조하여 작성하였다.

히 하고, 위반 시 제재 근거를 두고 있다. 국내법은 자동화평가에 의한 차별 등에 대한 규율이 미흡하다. 기계학습 알고리즘에 의한 금융차별을 방지하기 위한 공정성 기준, 감독 체계 등의 도입이 시급하다. 미국은 신용평가 관련 소비자 보호를 위해 전담기구(Consumer Financial Protection Bureau, CFPB)를 설치 운영 중이다. 반면 국내는 개인신용평가 검증 주체가 모호하고 전문성이 부족하다는 지적이 있어 신용정보 감독기구의 위상과 전문성을 제고하고 검증 기능을 강화하기 위한 전담기구의 필요성이 있다.

### (3) 개인정보보호책임자(CPO)제도

개인정보보호책임자(Chief Privacy Officer, CPO)는 조직 내에서 개인정보 처리에 대한 업무를 총괄해 책임지는 핵심적인 역할이 요구되지만, 기존 법령에서는 CPO의 자격과 직위요건만 규정하고 있어 전문성 및 독립성 보장 여건이 미비하다는 지적이 있었다<sup>80)</sup>. 개인정보보호법 시행령은 제31조 제6항을 신설해 개인정보처리자에게 CPO가 독립적인 업무 수행할 수 있도록 보장할 의무를 부여했다. CPO의 역할과 위상을 제고해 실질적인 역할 수행이 가능하도록 CPO의 자격요건, 독립성 보장 등에 관해 개인정보 처리자의 매출액, 개인정보 보유 규모 등을 고려해 시행령으로 정하고 CPO 협의회 구성 운영 근거와 개인정보위의 지원 근거도 마련했다<sup>81)</sup>. 개인신용평가기관과 금융회사에 개인정보보호책임자(CPO) 제도를 도입하게 된다면 CPO는 신용평가 시스템 전반에 걸쳐 개인정보 처리의 적법성과 안전성을 감독하고 통제함으로써, 정보주체의 권리와 프라이버시를 보다 두텁게 보호할 수 있는 효과가 기대된다. 또한 CPO의 지정은 신용평가 기관으로 하여금 개인정보 보호를 위한 물적·인적 자원을 투입하도록 유도하며, 개인정보 침해 사고 발생 시 책임 소재를 명확히 하는 데 기여할 수 있을 것이다. 개인정보의 주체는 자신의 개인정보 처리에 관해 이의를 제기하거나 설명을 요구할 때 CPO를 통해 보다 원활하게 소통할 수 있다. CPO의 존재는 조직 내부의 개인정보 보호 인식을 높이고, 임직원의 교육과 훈련을 촉진하는 계기가 된다.

그러나 CPO의 선임과 운영에는 상당한 비용이 수반된다. 특히 이제 막 시장에 진입하고자 하는 비금융정보 이용 신용평가 기관의 경우 이러한 비용 부담이 경영상 부담으로 작용할 수 있다. 또한 CPO와 관련된 법적 의무를 이행하는 과정에서

80) 홍대식, ESG 경영 시대의 CPO의 역할과 지위, ICT법경제연구소, 2021.04.27., [http://icle.sogang.ac.kr/blog\\_post/72](http://icle.sogang.ac.kr/blog_post/72).(최종접속일 2024.10.12.) 참조. 기존 CPO 제도의 한계는 유럽의 일반 개인정보보호법(GDPR)에 규정된 DPO(Data Protection Officer) 제도와의 비교를 통해 더욱 뚜렷해진다. CPO가 기업 내부에서 개인정보 처리 업무 및 정책을 관리하는 영업활동의 보조자로 무팀의 역할과 유사하다면 DPO는 영업활동을 독립적으로 감독하는 준법감시인 및 감사의 역할과 유사하다. 개인정보 처리 업무가 영업활동에 큰 비중을 차지하는 기업이라면 법무팀과 준법감시인 및 감사(감사위원회)를 다 두고 있는 만큼, CPO뿐만 아니라 DPO도 도입할 필요가 있다. 그러나 DPO가 먼저 제도화된 유럽과 달리 CPO 제도를 오래 운영해온 우리나라 현실에서 DPO 제도를 도입한다면 CPO와 DPO의 관계를 어떻게 정립할 것인가 하는 점이 어려운 문제이다.

81) 이번 개정안에 따르면, CPO는 개인정보보호 경력, 정보보호 경력, 정보기술 경력을 합해 총 4년 이상을 보유하고 있어야 지정할 수 있다. 이 중, 개인정보보호 경력은 최소 2년 이상이어야 한다. 단, 전문 CPO 지정의 부담을 덜기 위해 학위·자격 취득, 교육 이수시 일정기간을 CPO의 경력으로 인정한다.

추가적인 행정 비용이 발생할 수 있다. 개인정보 보호와 신용평가에 대한 전문성을 겸비한 CPO를 확보하기란 쉽지 않을 것이다. CPO의 전문성이 부족할 경우 형식적인 직책에 그칠 위험이 있다. 또한 CPO의 독립성이 보장되지 않는다면 객관적이고 공정한 직무 수행이 어려울 수 있다. 다른 경영상 의사결정과의 충돌 가능성 또한 제기된다. CPO는 개인정보 보호를 위해 다른 경영상 판단과 상충되는 의견을 제시할 수 있다. 이에 따라 신용평가 기관 내부의 의사결정 과정이 복잡해지고 혼선이 야기될 수 있다. CPO의 전문성과 독립성을 보장하기 위한 법·제도적 기반을 마련하고, 감독 당국의 철저한 모니터링을 통해 제도의 실효성을 담보할 필요가 있다.

#### 4. 인공지능의 한계와 인간의 개입을 통한 금융 포용

##### (1) 휴먼인더루프(human in the loop)

인공지능 기반 신용평가 시스템의 확대는 금융취약계층과 금융이력부족자에 대한 금융소외를 심화시킬 수 있다. 앞서 살펴본 대로, 이들은 전통적인 금융 데이터가 부족하거나 불리한 특성을 보유하고 있어, 알고리즘에 의한 자동화된 평가에서 낮은 점수를 받을 가능성이 높기 때문이다. 특히 기계학습 모델이 과거 데이터에 내재되어 있는 차별과 편향을 그대로 학습하여 재생산할 경우, 사회경제적 약자에 대한 금융배제가 고착화될 수 있다. 이러한 문제에 대응하기 위해서는 인공지능 관련 금융과정 전체에 인간이 더 직접적으로 개입하여 사회문화적인 가치를 도입하려는 시도가 필요할 것이다. 이러한 '휴먼인더루프(Human-in-the-Loop, HITL)<sup>82)</sup>' 방식은 학습, 조정, 검증 등 AI 모델의 전 생애주기에 걸쳐 사람이 능동적으로 관여하는 것을 의미한다. 인간 전문가가 인공지능 학습 과정 중에 중간 결과물을 확인하고 학습 데이터를 조정등의 과정을 거치는 것이다. 인공지능 기반 신용평가에서 금융소외가 일어나지 않도록 인간적인 노력, 일종의 '휴먼인더루프'를 적용하는 구체적인 방안으로는 인간 전문가의 검토 및 오버라이드(Override) 제도가 있다. 인공지능 모델이 도출한 신용평가 결과에 대해 금융 및 기술 분야 전문가가 최종 검토하고, 필요시 결과를 번복(override)할 수 있는 제도를 마련하는 것이다. 특히 금융취약계층이나 금융이력부족자에 대해서는 인공지능의 평가 결과와 더불어 전문 심사역의 정성평가 요소를 추가로 고려하여 최종 신용등급을 부여하는 방안을 검토할 수 있다.

이렇듯 인공지능을 통한 비금융정보 활용과 신속하고 효율적인 금융서비스 접근이 상대적으로 연령이 낮고 교육수준이 높은 집단에게는 유리하게 작용했지만, 그렇지 않은 집단은 소외될 뿐만 아니라 기회가 생겨도 불공정이나 권리침해를 마주할 수 있다. 앞서 살펴본 바와 같이, 인공지능 기반 신용평가시스템은 많은 장점을 가지고 있지만 사회적 편향이 반영되거나 개인정보 유출이 쉬운 취약점 역시 존재

82) 한국정보통신기술협회 정보통신용어사전, 휴먼인더루프, [https://terms.tta.or.kr/dictionary/dictionaryView.do?word\\_seq=181673-1](https://terms.tta.or.kr/dictionary/dictionaryView.do?word_seq=181673-1) (최종접속일 2024.10.12.) 참조. 신뢰도 높은 학습 모델을 도출하기 위해 인공지능 시스템 등에 사람이 개입하여 시스템과 사람이 상호작용하는 학습 구조로, 인간 전문가가 인공지능 학습 과정 중에 중간 결과물을 확인하고 학습 데이터를 조정한다. 즉, 인공지능 학습 과정에서 인공지능과 사람이 일종의 '피드백 루프'를 형성한 것이라 할 수 있다.

하기 때문이다. 포용적 금융으로 나아가기 위해서는 인공지능에 대한 수단으로서의 접근성 강화를 위한 보다 적극적인 개입이 필요할 것이다. 그러나 그럼에도 해결되지 아니하는 부분은 현대의 금융서비스는 더이상 선택의 문제가 아닌 생존의 문제라는 측면을 고려하여 금융기관, 특히 은행의 공적 역할에 의한 해결을 기대하여야 한다<sup>83)</sup>. 단순히 알고리즘에 모든 것을 일임하기보다는 전문가의 직접 개입을 통해 데이터와 프로세스 전반을 점검하고 개선해 나가는 전략이 필요하다. 금융 분야 전문가와 데이터 사이언티스트의 협업을 통해 인공지능 모델이 편향되거나 왜곡된 결과를 산출하지 않도록 지속해서 모니터링하고 조정해야 한다. 나아가 인공지능의 한계를 인식하고 사회적 약자에 대한 별도의 배려 방안을 강구하는 것도 중요하다. 금융이력부족자라 하더라도 성실한 경제활동을 통해 신용을 쌓아갈 수 있도록 지원하고, 다양한 대안정보를 활용하여 이들의 잠재력을 평가하려는 노력이 필요하다. 이는 기계적인 알고리즘을 넘어 인간의 통찰과 공감, 그리고 사회정의에 대한 고려가 반영되어야 가능하다.

여전히 해소되지 않는 금융 불평등과 배제의 문제는 결국 금융기관, 특히 은행의 공적 역할 강화를 통해 해결을 모색하여야 한다. 경제 성장 과정에서 자원을 적절히 배분하여야 사회적 균형이 달성되며, 만일 균형이 달성되지 않았을 시에는 이를 적극적으로 해결할 필요성이 있기 때문이다. 인공지능으로 인한 혜택을 누리지 못하는 계층에 대해서 공공부문의 직접적 금융 지원, 맞춤형 컨설팅 제공 등 공적 금융안전망 확충이 요구된다. 이를 위해 은행은 취약계층 지원을 위한 별도 예산과 조직을 마련하고, 수익성과 공공성의 조화로운 균형을 모색할 필요가 있다. 나아가 민관협력을 통해 휴면예금, 사회공헌기금 등을 재원으로 한 미소금융 확대 등 금융 취약계층 지원 정책의 실효성도 제고해야 할 것이다.

## (2) 대안적 신용평가의 가능성

한편, 은행과 기업이 긴밀한 관계를 형성함에 따라 비재무적 연성 정보를 기준으로 신용을 평가하여 금융서비스를 공급하기도 한다<sup>84)</sup>. 연성 정보에는 대표자의 경력이나 경영권 안정성, 경영 전문성, 윤리의식과 기업의 회계투명성 등 이른바 수치화되지 않는 정보가 포함된다<sup>85)</sup>. 경력이 짧고 영세한 중소기업은 신뢰할 만한 정량 정보가 부족하거나 평가점수가 낮아 계량화되지 않은 정성 정보가 중요하다. 이런 데이터는 장기간 관찰에 의해서만 축적될 수 있고, 포착이 어려우며 심사역이 판단하는 기업가의 성품이나 신뢰도, 기업의 자금결제 성향, 기업가 평판 등 전문적인 안목이 필요할 수 있다.

사회적경제기업에 대한 대출과정에서는 이러한 “관계금융”이 일찍이 도입되었다.

83) 이정수, “인공지능에 대한 금융법의 규제방향 -금융예외의 이용과 그 너머의 과제들.” 금융법연구 21.1 (2024), 43p.

84) 임정대, “관계금융이 기업혁신에 미치는 영향: 한국의 중소기업을 대상으로.” 재무관리연구 38.2 (2021): 65-103

85) 정재훈, [기자수첩] 소셜임팩트 생태계에서 ‘관계형 금융’ 꽃피우길, 소셜임팩트뉴스, 2023.05.04., <https://www.socialimpactnews.net/news/articleView.html?idxno=519> (최종접속일 2024.10.12.) 참조.

관계금융(relationship banking)이란 자본시장 참여자와의 정보 비대칭으로 인해 재무적으로 제약에 직면할 가능성이 높은 기업, 특히 중소기업의 정보를 취득하는 데 적합한 자금조달 기술이다. 금융기관과 기업이 장기간 거래 관계 속에서 형성하는 산물인 것이다<sup>86</sup>). 특히 사회적경제기업은 일반 기업과 운영방식이 다르므로 이윤 창출보다 더 중요한 가치를 우선시하기도 한다. 자본주의 체제하에서 가장 중요한 것이 경제적 이득이지만, 때로는 지속가능성을 고려하면서 사회적 연대를 구축하여야 진정한 경제성장이 이루어질 수 있다.

기업단위 사회적금융의 경우 사회적가치측정지표를 개발하여 사회적경제기업 신용평가방법론을 구축한 사례들을 먼저 살펴볼 수 있다. 지난 2013년부터 서울시 사회투자기금의 주도적인 노력으로 사회적 가치를 여신평가에 반영하기 위해 사회적경제조직들의 사회적 가치성과를 화폐가치로 추정하여 사회적가치 창출효과를 측정하는 사회적 투자수익률(SROI : Social Return On Investment) 방법론이 개발되었다. 또한 사회적경제기업 공공기관인 한국사회적기업진흥원의 사회적가치지표(SVI: Social Value Index)가 출시되었으며, 신용평가사인 이크레더블, 신용보증기금 등이 참여한 사회적경제기업평가시스템은 웹기반 오픈플랫폼방식으로 신용보증기금에서 서비스를 개시하였다. 신용보증기금은 사회적경제기업 부합성, 금융지원 타당성을 평가하여 SV 등급을 산출하며 이를 통해 사회적경제기업이 대출을 받을 수 있다<sup>87</sup>).

한국스마트협동조합과 북서울신용협동조합의 ‘문화예술인 상호부조대출 사업’은 개인에 대한 신용대출에도 관계금융의 방식이 작동할 수 있음을 보여준다. 북서울신용협동조합은 2022년 12월부터 문화예술인들에게 연리 5%, 최대 200만원까지 대출을 시작하였다. 대출은 신협에서 하고, 문화예술인들의 사회안정망 강화를 위해 400여명의 조합원들이 설립한 협동조합인 한국스마트협동조합은 ‘씨앗기금’을 조성하여 대손충당금을 마련하는 역할을 한다. 북서울신협에서 대출을 받은 문화예술인들이 대출을 연체하거나 상환하지 않을 경우 그 대손(손실)분만큼을 이 기금에서 충당하기로 하여 신용리스크 부담을 덜었다<sup>88</sup>). 대신, 신용대출 대상은 한국스마트협동조합 소속 문화예술인들로만 한정한다. 한국스마트협동조합에 대한 신뢰로 신협이 대출 사업을 진행하였다는 것이다. 금융접근성을 넓히려는 취지로, 금융을 통해 사회적 가치를 창출하는 좋은 협업사례라고 할 수 있다.

86) 이상진, [알면 the 이로운 금융] 32. 정량보다 정성이 중요한 금융도 있다, 이로운넷, 2021.08.27., <https://www.eroun.net/news/articleView.html?idxno=26004> (최종접속일 2024.10.12.)

87) 신용보증기금, 사회적경제기업평가, <https://www.kodit.co.kr/seenp/index.do> (최종접속일 2024.10.12.) 참조. 사회적경제기업으로서 부합성 평가영역 지표로는 기업철학, 참여/배려 및 연대, 사회적 가치 지향, 사회환원이 있으며, 금융지원타당성 평가영역 지표로는 경영역량, 교육, 지속가능성, 재무관리역량이 포함되어 있다.

88) 정재훈, "금융 사각지대 놓인 문화예술인, 상호부조대출이 도움 되길", 소셜임팩트뉴스, 2023.04.11., <https://www.socialimpactnews.net/news/articleView.html?idxno=378> (최종접속일 2024.10.12.)

## V. 결론

본 연구는 인공지능 기반 신용평가시스템의 현황과 문제점을 살펴보고, 특히 금융이력부족자의 권리보호 관점에서 제도적 개선방안을 모색하였다. 금융업계의 디지털 전환이 가속화되면서 인공지능 기술은 신용평가 영역에도 빠르게 도입되고 있다. 방대한 데이터와 고도화된 알고리즘을 활용한 인공지능 신용평가는 기존에 금융 접근성이 낮았던 계층에게 새로운 기회를 제공하며 금융포용 확대에 기여할 것으로 기대된다. 그러나 동시에 인공지능 신용평가는 고유한 한계와 부작용도 내포하고 있다. 평가 과정의 불투명성과 알고리즘의 불안정성, 그리고 데이터의 편향성은 AI 신용평가의 신뢰성을 저해하는 요인이 된다. 특히 금융이력부족자의 경우 다양한 비금융정보를 제공해야 하므로 사생활 침해와 부당한 차별에 노출될 위험이 크다.

이에 본 연구는 인공지능 기반 CSS의 공정성과 투명성 제고를 위한 다각도의 정책 방안을 제안하였다. 먼저 금융당국은 인공지능 신용평가의 기본 원칙과 기준을 명확히 하는 가이드라인을 제시하여 규제 방향성을 선제적으로 선도해야 한다. 이를 통해 민간에서 설명가능한 인공지능 모델 개발과 데이터 및 알고리즘 관리 강화를 이끌어낼 수 있을 것이다. 또한 차별금지와 개인정보보호에 관해서는 관련 법규를 정비하고 감독 체계를 고도화할 필요가 있다. 아울러 권리침해가 발생했을 경우의 사후적인 구제 수단을 다른 국가의 선진적인 제도와 비교하여 실효성 있게 마련할 필요가 있다. 그러나 결국 법과 제도만으로 인공지능 신용평가의 부작용을 완벽히 통제하기는 어렵다. 결국 금융포용과 공정한 신용질서의 구현을 위해서는 인간의 가치 판단과 사회적 책임이 요구된다. 단순히 신용등급에 의존하기보다는 금융소외계층의 상황과 잠재력을 종합적으로 고려하는 관용의 자세가 필요하다. 이는 CPO제도와 같은 전문 인력의 육성, 금융교육 및 상담 강화, 관계금융 프로그램의 개발 등을 통해 뒷받침될 수 있을 것이다.

본 연구가 제시한 정책 방안들이 실제 구현되기 위해서는 보다 세부적인 후속 연구가 필요하다. 금융규제법적 차원에서 입법론적 해결을 위해서는 인공지능 신용평가에 대한 다양한 이해관계자의 의견을 반영한 사회적 합의를 이끌어내는 작업이 요구된다. 기술 발전에 따른 규제 체계의 유연한 혁신 또한 지속되어야 한다. 한편, 금융법적 차원이 아닌 개인정보보호법의 정비도 필요할 것이며 더불어 인공지능 신용평가가 초래할 수 있는 사회경제적 영향에 대한 실증 분석을 통해 인공지능에 대한 윤리적 기준을 구체화하여 나갈 필요가 있다. 인공지능 신용평가는 이제 막 태동한 분야로서 그 미래상을 온전히 그려내기는 쉽지 않다. 중요한 것은 기술이 인간을 지배하는 것이 아니라, 인간이 기술을 통제하고 선용(善用)할 수 있어야 한다는 점이다. 인공지능 기반 신용평가시스템이 가진 혁신성을 겸허히 받아들이되, 그 폐해를 경계하고 보완하려는 사회 구성원 모두의 노력이 필요한 시점이다.

## 참고문헌

### 1. 단행본

김건식·정순섭 자본시장법 제4판, 박영사, 2023.

이재희, 헌법상 평등보장과 간접차별금지, 헌법재판연구원, 2020

박노형 외 8인 저, EU 개인정보보호법 - GDPR을 중심으로 -, 박영사(2017)

이상용, 인공지능과 법, 박영사(2019)

Remolina et al. Artificial Intelligence in Finance : Challenges, Opportunities and Regulatory Developments / Nydia Remolina, Aurelio Gurrea-Martinez. Cheltenham, UK: Edward Elgar Publishing, 2023.

### 2. 국문논문

고학수, 정해빈, and 박도현 “인공지능과 차별.” Justice (Seoul, Korea) 171 (2019): 199-277.

김일우 “고위험 인공지능시스템의 차별에 관한 연구.” 서강법률논총 13.1 (2024): 7-45.

김종운, 장원중, 김광용 “온라인 상거래 데이터를 반영한 개인신용평가모형(커머스스코어) 개발“ 정보화연구 16.1 pp.45-55 (2019)

박상철 “금융AI의 활용과 금융소비자 보호: 차별금지, 설명요구권, 6대 판매규제 준수를 중심으로.” BFL 107 (2021): 29.

박재만, “비금융 개인신용정보의 전송요구권에 대한 법적 쟁점.” 상사판례연구 34.3 (2021): 331-360.

방정미 “인공지능 알고리즘 규제거버넌스의 전환 — 최근 미국의 알고리즘 규제와 인공지능 윤리원칙을 중심으로.” 공법연구 49.3 (2021): 375-406

서정호, 금융업의 인공지능 활용과 정책과제, KIF 정책분석보고서, 2022.2

송옥렬 “신용평가회사의 문지기책임.” BFL 82 (2017): 72-83.

신영진 “지능정보사회의 안전한 인공지능 서비스 구현을 위한 개인정보 침해대응 및 위기관리 컴플라이언스 개발에 관한 연구.” 융합정보논문지 12.5 (2022): 1-14.

양철원 (2022). “신용평가업에 대한 규제강화는 기업 신용평가의 질을 향상시켰는가?” 보험금융연구, 106, 3-38

이정수 “신용평가업 규제의 현황과 문제점 그리고 개선방향”, 증권법연구, 17(1), 2016, 287-317

이정수, “인공지능에 대한 금융법의 규제방향 -금융에의 이용과 그 너머의 과제들.” 금융법연구 21.1 (2024), 9

임정대, “관계금융이 기업혁신에 미치는 영향: 한국의 중소기업을 대상으로.” 재무관리연구 38.2 (2021): 65-103.

최성민, “개인 신용평가모형의 변화와 시사점”, CIS이슈리포트, 한국신용정보원, 2018, 제2018-4호, 1-12.

최성민 “개인 신용평가모형과 설명력 이슈.” 한국경영정보학회 정기 학술대회 (2020): 217-221.

최성민 “개인 신용평가에서의 금융소비자 보호에 관한 연구: 인공지능 방법론 기반 신용평가모형을 중심으로.” 금융감독연구 (2022): 65-98.

최성민 “공개된 개인정보의 신용평가 활용과 법적 쟁점.” 금융정보연구 (2023): 35-73.

### 3. 영문논문

Gillis, Talia B. and Spiess, Jann L. “Big Data and Discrimination,” University of Chicago Law Review: Vol. 86: Iss. 2, Article 4. (2019), 459-488.

Hurley, Mikella, and Julius Adebayo. “Credit scoring in the era of big data.” Yale JL & Tech. 18 (2016): 148.

Liu, Lydia T., et al. “Delayed impact of fair machine learning.” International Conference on Machine Learning. PMLR, 2018.

Rodriguez, Lorena. “ALL DATA IS NOT CREDIT DATA: CLOSING THE GAP BETWEEN THE FAIR HOUSING ACT AND ALGORITHMIC DECISIONMAKING IN THE LENDING INDUSTRY.” Columbia law review 120.7 (2020): 1843-1884.

### 4. 국문자료

강송희, “기계학습 공정성 관련 연구 동향”, 월간SW중심사회 통권 제68호, 소

프트웨어정책연구소, [https://spri.kr/posts/view/22898?code=industry\\_trend#foot\\_3](https://spri.kr/posts/view/22898?code=industry_trend#foot_3) (최종접속일 2024.09.11.)

구태언, 전용준, 유창하, “개인정보 보호법상 자동화된 결정에 대한 정보주체의 대응권”, 법률신문, 2024.03.06., <https://www.lawtimes.co.kr/LawFirm-NewsLetter/196490> (최종접속일 2024.10.12.)

금융보안원, “금융분야 AI 보안 가이드라인”, 2023.4.17.

금융위원회, “[보도자료] 「금융분야 인공지능(AI) 가이드라인」이 시행됩니다. - 금융권 인공지능(AI) 활용을 활성화하고 인공지능(AI) 기반 금융서비스에 대한 신뢰를 제고하기 위한 모범규준 마련·발표 -”, 2021.07.08., <https://www.fsc.go.kr/no010101/76206?srchCtgr=&curPage=21&srchKey=&srchText=&srchBeginDt=&srchEndDt=> (최종접속일 2024.10.12.)

기획재정부 시사경제용어사전, <https://www.moef.go.kr/sisa/dictionary/detail?idx=3216>, (최종접속일 2024.09.11.)

기획재정부 시사경제용어사전, 2020.11.03., <https://www.moef.go.kr/sisa/dictionary/detail?idx=329> (최종접속일 2024.09.11.)

김나경, “[단독] 無이력의 악순환’원파일러 99.9%가 신용 800점 이하... “거래이력이 없지, 빚 안 갚는다 했나“”, 파이낸셜뉴스, 2023.03.15., <https://www.fnnews.com/news/202303151401284618>. (최종접속일 2024.10.12.)

김대은, “[개인정보 보호법 시행령 2차 개정②] AI 결정에 대한 국민 권리 보장 및 CPO제도 개선”, 2024.03.22, <https://kidd.co.kr/news/236128> (최종접속일 2024.10.12.)

김연정, “‘금융이력부족자’ 대부분 신용점수 700점대...은행 대출에 불리”, 연합뉴스, 2021.09.27., <https://www.yna.co.kr/view/AKR20210926043100002> (최종접속일 2024.09.11.)

나이스 평가정보 웹사이트, [https://www.niceinfo.co.kr/creditrating/cb\\_score\\_1\\_4\\_1.nice](https://www.niceinfo.co.kr/creditrating/cb_score_1_4_1.nice)(최종접속일 2024.09.20.)

네이버프라이버시센터, [https://privacy.naver.com/global\\_support?menu=global\\_support\\_eu\\_gdpr\\_understand](https://privacy.naver.com/global_support?menu=global_support_eu_gdpr_understand) 최종접속일 2024.10.12.)

박순빈, “신한은행 “고졸은 대출이자 더 내라” “, 한겨레, 2012.07.23., [https://www.hani.co.kr/arti/economy/economy\\_general/543895.html](https://www.hani.co.kr/arti/economy/economy_general/543895.html) (최종접속일 2024.10.12.)

박지영, “AI가 카드·여권 정보 훑쳐봤다... ‘인간지능’ 이 답변 검토까지”, 한겨레, 2024.03.28., [https://www.hani.co.kr/arti/economy/economy\\_general/1134230.html](https://www.hani.co.kr/arti/economy/economy_general/1134230.html)

(최종접속일 2024.10.12.)

송주오, “'미동의 고객에 문자 10만건'...금감원, 우리은행 기관경고 조치”, 이데일리, 2024.01.12., <https://m.edaily.co.kr/news/read?newsId=02732246638757064&mediaCodeNo=257> (최종접속일 2024.10.12.)

[시사금융용어] 대안신용평가, 연합인포맥스, 2019.04.05., <https://news.einfomax.co.kr/news/articleView.html?idxno=4023693>(최종접속일 2024.09.11.)

신용보증기금, 사회적경제기업평가, <https://www.kodit.co.kr/seenp/index.do> (최종접속일 2024.10.12.)

심지영, [단독] 통신 3사 합작 신평사 출범 초읽기, 금융위 문턱 넘었다, 비즈한국, 2023.12.28., <https://www.bizhankook.com/bk/article/26874>(최종접속일 2024.10.12.)

연태훈, “금융에서의 AI 확산에 내재한 잠재적 위험과 관리 방안의 모색”, KIF 분석보고서, 2024-01, 한국금융연구원, <https://news.einfomax.co.kr/news/articleView.html?idxno=4023693>(최종접속일 2024.09.11.)

이상진, “[알면 the 이로운 금융] 32. 정량보다 정성이 중요한 금융도 있다”, 이로운넷, 2021.08.27., <https://www.eroun.net/news/articleView.html?idxno=26004> (최종접속일 2024.10.12.)

장문기, “[인터뷰] 문재남 통신대안평가준비법인 대표 “3월 사업개시...개인신용 패러다임 바꿀 것“”, 아주경제, 2024.01.02., <https://www.ajunews.com/view/20240101153250616>(최종접속일 2024.10.12.)

정재훈, “[기자수첩] 소셜임팩트 생태계에서 ‘관계형 금융’ 꽃피우길”, 소셜임팩트뉴스, 2023.05.04., <https://www.socialimpactnews.net/news/articleView.html?idxno=519> (최종접속일 2024.10.12.)

정재훈, “금융 사각지대 놓인 문화예술인, 상호부조대출이 도움 되길“, 소셜임팩트뉴스, 2023.04.11., <https://www.socialimpactnews.net/news/articleView.html?idxno=378> (최종접속일 2024.10.12.)

코리아크레딧뷰로(KCB, 올크레딧) 웹사이트, <https://www.allcredit.co.kr/screen/sc0682112929>(최종접속일 2024.09.20.)

한국신용정보원 사이트, [https://www.kcredit.or.kr:1441/orgn/business.do?\\_csrf=e96fee-5e68-4ba4-9e14-038e66de35d1&menuNo=121&hpBoardSn=&link=orgn%2Fbusiness.do](https://www.kcredit.or.kr:1441/orgn/business.do?_csrf=e96fee-5e68-4ba4-9e14-038e66de35d1&menuNo=121&hpBoardSn=&link=orgn%2Fbusiness.do) (최종접속일 2024.10.12.)

한국정보통신기술협회 정보통신용어사전, 휴먼인더루프, [https://terms.tta.or.kr/dictionary/dictionaryView.do?word\\_seq=181673-1](https://terms.tta.or.kr/dictionary/dictionaryView.do?word_seq=181673-1) (최종접속일 2024.10.12.)

홍대식, “ESG 경영 시대의 CPO의 역할과 지위”, ICT법경제연구소, 2021.04.27., [http://icle.sogang.ac.kr/blog\\_post/72](http://icle.sogang.ac.kr/blog_post/72), 최종접속일 2024.10.12.)

## 5. 영문자료

Emmanuel Martinez & Aaron Glantz, “How Reveal Identified Lending Disparities in Federal Mortgage Data”, (2018), [https://s3-us-west-2.amazonaws.com/revealnews.org/uploads/lending\\_disparities\\_whitepaper\\_180214.pdf](https://s3-us-west-2.amazonaws.com/revealnews.org/uploads/lending_disparities_whitepaper_180214.pdf) (최종접속일 2024.10.12.)

EUR-Lex (유럽연합 법령 공식 사이트), URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>(최종접속일 2024.10.12.)

EUR-Lex (유럽연합 법령 공식 사이트), URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015L2366>(최종접속일 2024.10.12.)

European Commission. (n.d.). Equality and non-discrimination. [https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/know-your-rights/equality/non-discrimination\\_en](https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/know-your-rights/equality/non-discrimination_en)(최종접속일 2024.09.11.)

European Commission. (2021). Mortgage Credit. [https://ec.europa.eu/info/business-economy-euro/banking-and-finance/consumer-finance-and-payments/retail-financial-services/credit/mortgage-credit\\_en](https://ec.europa.eu/info/business-economy-euro/banking-and-finance/consumer-finance-and-payments/retail-financial-services/credit/mortgage-credit_en)(최종접속일 2024.09.11.)

European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&from=IT>(최종접속일 2024.09.20.)

Fair Credit Reporting Act, <https://www.law.cornell.edu/uscode/text/15/1681> (최종접속일 2024.10.12.)

GDPR Article 22, Automated individual decision-making, including profiling, [https://gdpr-text.com/ko/read/article-22/#comment\\_gdpr-a-22\\_1](https://gdpr-text.com/ko/read/article-22/#comment_gdpr-a-22_1) (최종접속일 2024.10.12.)

Glantz, Aaron and Emmanuel Martinez. “For People of Color, Banks are Shutting the Door to Homeownership.” Reveal News, 2018.02.16., <https://revealnews.org/article/for-people-of-color-banks-are-shutting-the-door-to-homeownership/> (최종접속일 2024.10.12.)

Kenneth P. Brevoort, Philipp Grimm & Michelle Kambara, “Data Point: Credit Invisibles”, Consumer Fin. Prot. Bureau 4-6 (2015), [https://files.consumerfinance.gov/f/201505\\_cfpb\\_data-point-credit-invisibles.pdf](https://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf) [<https://perma.cc/J47X-4V89>] (최종접속일 2024.10.12.)

MAS, “Compliance Toolkit for Approvals, Notifications and Other Regulatory Submissions to MAS”, 2023.12.07., <https://www.mas.gov.sg/regulation/guidance/compliance-toolkit-for-approvals-notifications-and-other-regulatory-submissions-to-mas-sfa> (최종접속일 2024.10.12.)

MAS, “MAS introduces new FEAT Principles to promote responsible use of AI and data analytics“, 2018.11.12., <https://www.mas.gov.sg/news/media-releases/2018/mas-introduces-new-feat-principles-to-promote-responsible-use-of-ai-and-data-analytics> (최종접속일 2024.10.12.)

MAS, “Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore’s Financial Sector“, 2018.11.12. <https://www.mas.gov.sg/publications/monographs-or-information-paper/2018/feat> (최종접속일 2024.10.12.)

Phillips, P. et al., “Four Principles of Explainable Artificial Intelligence”, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, (2021), <https://doi.org/10.6028/NIST.IR.8312>, [https://tsapps.nist.gov/publication/get\\_pdf.cfm?pub\\_id=933399](https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=933399) (최종접속일 2024.10.12.)

Sarah Wheaton, “Credit Score, by Multiple Choice”, The New York Times, 2013.12.30., <https://www.nytimes.com/2013/12/31/business/credit-scores-from-a-test-not-a-history.html>(최종접속일 2024.10.12.)

Tom Groenfeldt, “Lenddo Creates Credit Scores Using Social Media”, Forbes, 2015.06.29., <https://www.forbes.com/sites/tomgroenfeldt/2015/01/29/lenddo-creates-credit-scores-using-social-media/?sh=431fa3ea2fde> (최종접속일 2024.09.20.)

## 6. 기타

대법원 2014다235080

헌법재판소 99헌마513

