민사소송에서의 AI 알고리즘 심사 - 호주 Trivago 판결과 관련하여[1)]

한애라/교수, 성균관대학교 법학전문대학원

I. 들어가며

인공지능, 특히 머신러닝 기법은 사회 전반에서 이미 활발히 이용되고 있다. 인공지능의 활용이 폭증하면서 AI 알고리즘에 의한 자동화된 의사결정(대출이든, 운전이든, 상품 추천이든)의 결함으로 인한 민사책임이 발생할 가능성도 커지고 있다.

그런데 AI 알고리즘에서의 결함이나 하자의 의미, 책임의 주체, 책임의 분배, 과실 판단 기준, 제조물책임법의 개정 등에 관하여는 국내외적으로 상당히 심도 깊은 논의가 전개되어 왔으나, 민사소송절차에서 알고리즘을 어떻게 심사할 지에 대하여는 아직 논의가 부족하다. 인공지능의 하자나 결함에 따른 실체적인 책임을 묻기 위해서는 쟁점이 되는 AI 알고리즘의 내용을 분석할 수 있도록 충분히 증거 제출, 조사, 심리가 이루어져야 하고, 이를 통해 법원이 AI 알고리즘을 둘러싼 쟁점에 관하여 판단을 내릴 수 있어야 한다. 그러나 인공지능, 특히 머신러닝 기법에 의한 AI 알고리즘은 그 불투명성으로 인하여 사법심사가 용이하지 않다. AI 알고리즘이 어떻게 작동하는지가 소송의 쟁점이 될 때, 이러한 불투명성을 극복하기 위하여 AI 알고리즘을 어떠한 방식으로 감정하고 그 감정결과를 법원이 어떻게 판단의 자료로 삼아 심증을 형성할 것인지가 문제된다.

그런데 최근 호주 연방법원에서는 Trivago 사이트의 호텔 딜 추천 AI 알고리즘이 어떻게 작동하는가가 정면으로 주된 쟁점이 되어 이 점에 관하여 전문가 의견서 제출 및 전문가 증언을 거쳐 비교적 상세한 판결이 내려진 바 있다.[2)] 이 글에서는 먼저 Trivago 사건에서 AI 알고리즘을 어떻게 감정하고 심리하였는지 살펴보고, 이를 우리나라의 민사소송절차와 비교하면서, 설명 가능성이 떨어지는 머신러닝 알고리즘을 심리하는 과정에서 등장할 수 있는 여러 문제점과 그 해결 방안을 검토한다.

II. Trivago 사건의 개관

---

1) 이 글은 저자가 민사소송 27-1호에 투고한 글을 2023. 2. 27. 서강대학교 ICT 연구소와 당근마켓이 공동으로 주최하는 ...연구회에서 발표, 토론한 결과에 따라 축약, 수정한 것이다.

2) 책임소송(liability lawsuit) 제1심 Australian Competition and Consumer Commission v Trivago N.V. [2020] FCA 16; 항소심 Trivago N.V. v Australian Competition and Consumer Commission [2020] FCAFC 185; 구제소송(relief lawsuit) Australian Competition and Consumer Commission v Trivago N.V. (No 2) [2022] FCA 417.

1. 소송의 배경

Trivago는 전세계를 대상으로 Booking.com, Expedia, Hotels.com 등 여러 온라인 호텔 예약 사이트에 올라온 호텔 딜을 검색, 비교하여 이용자에게 추천하는 일종의 메타서치(meta-search) 서비스이다.[3] 이용자는 Trivago 사이트 검색결과 창에서 호텔 딜을 선택하여 클릭하면 해당 예약 사이트로 이동하여 거기서 예약하므로, 이용자가 Trivago에 직접 예약수수료를 내는 것은 아니다. 대신 예약 사이트들은 이용자가 Trivago를 통해 자신의 사이트가 제공하는 딜을 클릭한 경우 Trivago에 클릭당 수수료(Cost per Click, 이하 "수수료")를 지급한다. Trivago의 주된 수익은 수수료에서 창출되며, 위 수수료는 각 예약 사이트의 입찰에 의해 계속 변동된다.

Trivago는 최적의 가격에 이상적인 호텔을 검색하라는 등으로 광고하면서, 검색 결과 중 하나를 선별하여 "Top Position Offer"로 가장 잘 보이는 곳에 배치하여 추천하였다. 다수의 소비자들은 Top Position Offer가 최저가인지를 확인하지 않은 채 관성적으로 이를 클릭하여 예약하였으나, Top Position Offer 중에는 최저가가 아닌 경우도 발견되었다.[4] 호주의 경쟁 및 소비자 위원회(Australian Competition & Consumer Commission, ACCC, 이하 'ACCC'라고만 한다)는 Trivago의 알고리즘이 순수하게 가격만 비교하는 것이 아니라 수수료도 고려하여 "Top Position Offer"로 추천하는 것임에도, Trivago가 광고 및 홈페이지 표시 등을 통해 마치 최저가 딜을 추천하는 것처럼 소비자를 오인하였다고 주장하면서, 2018. 8. 호주 연방법원(Federal Court of Australia)에 Trivago를 상대로 호주 소비자법(Australian Consumer Law) 위반(사기, 오인 표시의 금지)의 확인 및 과징금의 지급, 오인 표시의 금지 등을 구하는 민사소송을 제기하였다.[5][6] 위 소송에서는 Trivago의 광고, 표시를 보고 소비자가 Top Position Offer를 최저가로 오인하였는지, 이를 잘 보이는 곳에 배치하는 것이 소비자의 선택에 얼마나 영향을 주는가와 함께, Trivago의 AI 알고리즘 구조상 Top Position Offer를 선별하는 주된 요소가 가격인가, 수수료인가가 쟁점이 되었다.

---

3) https://www.trivago.com/

4) Australian Competition and Consumer Commission v Trivago N.V. [2020] FCA 16(이하 ACCC v Trivago), para. 3-7.

5) ACCC는 호주 소비자법 위반자에 대해 민사소송을 제기하여 법원으로부터 벌금(fines), 과징금(pecuniary penalties), 자격정지 명령(disqualification orders), 금지명령(injunctions) 또는 배상명령(compensation orders)을 받을 수 있다. Commonwealth of Australia, "ACL Compliance and Enforcement Guide" (2017) 11면 이하 참조.
<https://consumer.gov.au/sites/consumer/files/2019/01/ACL_Compliance_and_enforcement_guide.pdf>(2022. 11. 7. 최종확인).

6) VID 1034 of 2018.

2. 심리절차

호주 연방법원의 민사소송절차는 전형적인 영미법계 절차를 따르고, 전문가의 감정도 당사자 선임 전문가 증언 및 의견서(Parties' expert witnesses and expert reports)를 이용하는 것이 보다 일반적이다. 이 사건에서도 원피고가 각자 알고리즘을 감정할 전문가를 1인씩 선임하였고, 쌍방 전문가는 Trivago로부터 특정 3일치의 호주 4개 주도(州都)에 관한 호텔 딜 입력값 및 결과값 데이터를 제출받아 알고리즘을 감정한 후 의견서를 제출하였다.

1~2개월 정도의 간격으로 먼저 원고 측 전문가가 원고가 제시한 9개 기술적 질의에 답하는 형태의 전문가 의견서를 제출하고, 피고 측 전문가가 위 의견서에 답하는 전문가 의견서를 제출하였으며, 다시 원고 측 전문가가 2차 의견서, 피고 측 전문가가 추가 의견서를 제출하였다. 법원은 양쪽 전문가 의견서의 여러 점에서 이견이 있자 양 전문가에게 공동의견서를 작성할 것을 명하였고, 양 전문가가 공동의견서(Joint Report)를 제출하였다. 이 공동의견서에서는 당초 원고의 9개 질의에 대해 양 전문가의 의견이 일치하는 부분과 의견이 다른 부분이 각각 명시되었고, 각 의견서 중 영업비밀로서 보호가 요구되는 부분은 따로 표시되었다.

양측 전문가 증인에 대한 증인신문기일은 영업비밀 보호를 위하여 비공개로 진행되었고, 쌍방 전문가가 공동으로 증언하였다. 증언은 기본적으로 공동의견서의 구조에 따라, 공동의견서에서 제시된 9개의 질문에 대하여 각 질문별로, 양쪽 전문가가 모두진술을 마친 후 주신문, 반대신문을 하는 방식으로 진행되었다. 이러한 방식에 의하여 양쪽 전문가 의견의 공통점과 차이점, 각각의 근거와 강점 및 약점이 명확하게 드러났다.[7)]

또한 각종 자료, 의견서, 증언에 포함된 영업비밀을 보호하기 위해, 심리의 비공개, 비밀정보 부분의 삭제, 최종 판결문에서 해당 내용의 직접적인 언급 자제 등의 조치가 취해졌다.

3. Trivago 알고리즘에 관한 전문가 의견 및 증언과 법원의 판단

양 전문가는 "Top Position Offer를 정하는 각 요소의 가중치 또는 상대적 중요성이 어

---

7) 이와 같은 방식은 호주 FCR Part 23에도 명시적으로 규정되어 있다. 복수의 당사자가 유사한 질문에 대하여 각기 전문가 증인을 신청하고자 하는 경우, 당사자는 전문가들로 하여금 다음과 같은 특정 절차에 따르도록 하는 명령을 법원에 신청할 수 있다. 그 주된 내용은 ① 전문가 의견서 작성에 앞선 전문가들 간 사전회의, ② 전문가 상호간 동의하는 점과 이견이 있는 부분이 특정된 의견서의 제출, ③ 전문가 의견서에 기재된 내용으로 전문가 증언 범위의 제한, ④ 전문가 증언에 앞서 전문가 의견서 작성의 바탕이 된 사실관계를 뒷받침할 증거의 제출, ⑤ 전문가들을 통합하여 신문하되, 각 질문별로 전문가들에 대한 교호신문 등이다.

떻게 되는가?"라는 질문에 대해 첨예하게 대립하였다. 위 질문에 답하기 위하여 양측 전문가는 서로 다른 기법을 사용하였고, 결론도 달랐다. 원고 측 전문가는 Partial Dependence Plot (PDP)[8]와 Gradient Boosted Model (GBM)[9]을 사용하여, Top Position Offer를 정하는 데 가장 중요한 요소는 수수료이고 제안 가격은 극히 미미한 영향밖에 없다는 의견을 제시하였다. 피고 측 전문가는 반면 Accumulated Local Effect (ALE) 모델[10] 등 다른 XAI 기법을 이용하여 Top Position Offer 산정에는 제안 가격이 가장 중요하고(45.5% ~ 60.0%), 그 다음이 수수료(33.8% ~ 44.8%)라는 의견을 제시하였다. 전문가들은 한편, 최저가가 아닌 딜이 Top Position Offer로 선정된 경우가 66.8%였다는 데에 동의하였으나, 피고 측 전문가가 가격, 수수료 외에 무료 취소, 무료 와이파이, 조식 포함 등의 비가격적 요소 또한 기여 요소라는 의견을 낸 데 반하여, 원고 측 전문가는 감정을 위하여 주어진 데이터만으로는 위와 같은 비가격적 요소가 영향을 미쳤다고 볼 만한 '가시성'(visibility)이 없다고 반박하였다.

공동의견서에서는 양 전문가의 의견이 일치하는 부분, 불일치하는 부분이 일목요연하게 정리되었고, 전문가 증언에서는 이견이 있는 부분을 중심으로 양 전문가 간에 치열한 구두공방이 벌어졌다.

법원은 위와 같은 전문가 의견서 및 법정 증언을 바탕으로 하여, 각 질문별로 양측 전문가 의견의 당부를 판단하였다. 우선, 피고 측 전문가의 의견에 의하더라도 가격뿐만 아니라 수수료 또한 Top Position Offer를 결정하는 데에 두 번째로 중요한 요소이므로 알고리즘이 가격만 고려하여 Top Position Offer를 결정한 것은 아니라고 판단하였고, 소비자가 선호하는 비가격적 요소가 영향을 미친다는 피고 측 전문가의 의견은 이를 뒷받침할 가시적인 데이터가 없다는 등의 이유로 배척하였다. 결론적으로 법원은, 양측 전문 Top Position Offer 중 약 66%가 최저가가 아니었음에도 Trivago는 자신의 사이트가 최저가 딜을 Top Position Offer로 추천하는 것처럼 표시하였으므로, 소비자를 오인하게 하는 표시 · 광고로서 호주 소비자법 위반에 해당한다고 판결하였다.

Trivago는 위 판결에 항소하였으나 항소가 기각되어 판결이 확정되었다.[11]

---

8) Partial Dependence Plot (PDP)은 알고리즘 자체를 모르고 인풋과 아웃풋 데이터만 알고 있을 때, 추이를 보고 싶은 특정 변수를 제외한 나머지 변수들을 한계화해줌으로써 특정 변수가 결과 변수에 미치는 영향을 보여주는 리버스 엔지니어링 방식의 XAI 기법으로서, 직관적이고 구현하기 쉽다는 장점이 있는 한편, 데이터 분포가 적은 구간에서 오류가 발생하고, 변수 간 상관관계가 있을 경우 부정확하다는 단점이 있다. C Molnar, 『Interpretable Machine Learning: A Guide for Making Black Box Models Explainable』, Independently published (2022), Ch. 8.1.

9) 머신러닝 기법 중의 하나로, 하나의 예측 결과값에 여러 입력값이 상대적으로 얼마만큼 기여했는지를 산정하는 데 비교적 잘 작동한다. ACCC v Trivago, para. 110.

10) 특성값이 머신러닝 모델의 예측에 평균적으로 얼마나 영향을 미쳤는지 설명하는 기법으로서, PDP보다 빠르고 편향되지 않으며, 변수 간 상관관계가 있을 경우에도 작동한다. Molnar, op. cit., Ch. 8.2.

11) 항소심 판결의 설시는 1심 판결과 거의 동일하다. Trivago N.V. v Australian Competition and Consumer Commission [2020] FCAFC 185.

4. Trivago 판결의 시사점

Trivago 판결은 쟁점의 조기 정리, 디스커버리 절차를 통한 증거의 충분한 획득, 절차에서의 영업비밀 보호, 쌍방 당사자가 선임한 전문가의 공방을 통한 사실 인정 등 영미법계의 소송절차를 충분히 활용함으로써, 법원에는 생소한 '설명 가능성이 떨어지는 머신러닝 알고리즘'의 내용을 성공적으로 심리, 판단하였다는 데에 그 의의가 있다.

첫째, AI 알고리즘이 100% 설명 가능성을 확보하지 못하는 경우라도, 입력값과 결과값을 가지고 사후적 XAI 기법을 동원하여 어떠한 변수가 얼마만큼 기여하였는지를 확률적으로 계산하는 것은 어느 정도 가능한데, 법원은 이 기법에 의한 전문가 의견을 존중하여 사실판단의 근거로 삼았다. 둘째, 알고리즘을 100% 공개하지 않고도, 사후적 XAI 기법에 의한 감정이 가능할 정도로 입력값과 결과값 데이터를 제출받아, 적절한 영업비밀 보호 조치를 취하면서 감정을 수행하는 것이 가능하였다. 셋째, 양측 전문가의 의견서를 비교하고 공동의견서를 작성하게 하며, 전문가 증언에서 쌍방이 공방하게 함으로써 쟁점을 명확하게 하고 컴퓨터 전문가가 아닌 법관에게 쟁점을 충분히 이해시킬 수 있었다. 우리 민사소송절차에서 AI 알고리즘을 심리할 경우에도 Trivago 사건에서의 위와 같은 심리 방법이 많은 참고가 될 수 있을 것이다.

III. 우리 민사소송에서의 AI 알고리즘 심사

1. 쟁점

이미 우리 소송절차에서도 AI 알고리즘, 특히 '설명 가능성이 떨어지는 불투명한 알고리즘'을 심리의 대상으로 하는 경우가 생기고 있다. 그러나 증거 수집, 감정, 심리의 각 단계에서 우리 민사소송절차는 여러 모로 경직되어 있고, 이러한 절차적 특징이 AI 알고리즘의 불투명성과 만나면 AI 알고리즘에 대한 심리와 판단은 더욱 어려워진다. 그러나 Trivago 판결에서 알 수 있듯이, 기술적으로 불투명하거나 영업비밀로 보호되는 머신러닝 알고리즘이라 하더라도, 자료 소지자의 영업비밀을 본질적으로 침해하지 않는 범위에서 필요한 정보를 제출받아 설명 가능한 인공지능(XAI) 기법에 의하여 민사적 사실인정에 필요한 정도의 심증을 형성하는 것은 가능하고, 기술적인 복잡성이나 난해성 또한 전문가의 활용을 통해 상당 부분 극복할 수 있다.

그러기 위해서는 먼저 AI 알고리즘의 특성과 XAI 기법의 본질적인 내용을 이해해야 하고, 알고리즘 분석에 필요한 데이터의 범위와 제출 방법, 제출된 데이터의 보호, 전문가 활용 문제도 검토할 필요가 있다.

2. AI 알고리즘의 설명 가능성

가. 설명 가능한 인공지능(XAI)의 의의

현재 폭발적인 발전을 거듭하고 있는 머신러닝은 기계가 인간의 명시적인 지시에 의해서가 아니라 패턴과 추론을 통해 자동화된 방식으로 학습하도록 하는 일련의 기법이다. 머신러닝 기법에 따라 성명 가능성도 다르지만, 일반적으로 설명 가능성은 알고리즘의 복잡성과 트레이드오프 관계에 있다. 즉 예측이 정확해질수록 알고리즘은 복잡해지고 설명 가능성은 떨어진다.[12] 머신러닝 알고리즘의 불투명성은 다음의 요인들로부터 발생한다.[13] 우선, 머신러닝 기법에 의한 알고리즘은 연역적인 방법이 아니라 반복학습에 의하여 귀납적으로 생성되므로, 어떻게 하여 A라는 입력값으로부터 B라는 결과값을 도출하게 되었는지 정확하게 설명하기 어려운 블랙박스적 특성을 가진다.[14] 다음으로, 알고리즘이 영업비밀이나 지적재산권의 대상이 되거나 상업적 가치를 가지는 경우 기업은 법적으로 그 공개를 거부할 권리가 있어, 법적 불투명성이 발생한다.[15] 끝으로 기술적으로 매우 복잡한 반직관적 알고리즘을 전문가의 능력으로 분석할 수 있는 경우에도, 그 내용을 판사 또는 일반인이 알아듣도록 설명하는 것은 불가능할 수 있다.[16]

그러나 많은 영역에서 인공지능의 설명 가능성은 이를 의사결정에 활용하기 위한 기본 전제이거나, 사후적으로 법적 책임을 묻기 위해 필요하다. 이에 따라 설명 가능한 AI (Explainable AI), 즉 XAI의 필요성도 대두되었다. XAI는 AI 알고리즘에 대하여 예측의 판단 이유를 인간이 이해할 수 있도록 설명하는 기술(혹은 특성)을 말한다.[17] 2017년 DARPA가 "Explainable AI (XAI) Program"을 시작한 이래 XAI에 관한 정책과 연구는 모든 영역에서 매해 폭증하여 왔다.[18] XAI에는 법적 측면과 기술적 측면이 있다. 기술적으로는 알고리즘 자체의 설명 가능성을 높이기 위하여 다양한 XAI 기법이 개발되고 있

---

12) 오오쓰보 나오키 외/김대희 옮김, □ XAI - 설명 가능한 AI□ , BJ(2022), 6면.

13) J Burrell, "How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms", *3(1) Big Data & Society* 1 (2016); A D Selbst & S Barocas, "The Intuitive Appeal of Explainable Machines", *87(3) Fordham Law Review* 1085, 1089-1099 (2018); H Fraser, A J Snoswell & R Simcock, "AI Opacity and Explainability in Tort Litigation", In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAcT '22), June 21-24, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 185. <https://doi.org/10.1145/3531146.3533084>

14) 고학수·정해빈·박도현, "인공지능과 차별", 저스티스 통권 제171호(2019), 215-216면.

15) 자료 보관의무가 없을 경우 기업은 의도적으로 자료를 폐기할 수도 있고, 일부러 폐기하는 것까지는 아니라도 자료 보관기간을 단기로 설정함으로써 자료가 공개될 위험을 낮출 수 있다.

16) Burrell, op. cit., 4.

17) 오오쓰보, 앞의 책, 18면 이하.

18) M R Islam et al, "A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks," *Appl. Sci. 2022, 12*, 1353, 2.

고,[19] 법적으로는 인공지능의 설명 가능성을 높이기 위한 여러 제도적 장치가 논의되고 있다.[20] XAI는 또한 설명의 시점에서 사전적인가 사후적인가, 설명의 범위에서 국소적(local) 설명인가 전역(global) 설명인가, 입력값 또는 결과값의 종류가 무엇인가 등에 따라 다양하게 분류된다.[21] 그 중 이 글의 목적과 관계있는 XAI의 분류는 사전적 기법(ante hoc methods)과 사후적 기법(post hoc methods)이고, 이에 관하여는 좀 더 자세히 살펴본다.

나. 사전적 기법과 사후적 기법

사전적 기법은 처음부터 알고리즘 자체를 설명 가능하도록 설계하는 것이다. 설계 과정에서 설명 가능성을 확보하는 방법으로는, 첫째, 사람이 이해할 수 있을 정도로 특징량(feature)의 수를 제한하는 것,[22][23] 둘째, 의사결정나무(decision tree)나 선형회귀(linear regression)와 같이 입력값이 결과값에 미치는 영향을 파악할 수 있는 모델을 사용하는 것, 셋째, AI 알고리즘의 학습 과정에서 모델이 지나치게 복잡해지지 않도록 매개변수(parameter)를 조절하거나[24] 정규화(regularization)[25] 등의 기법을 사용하는 것, 단조성(monotonicity)이 유지되도록 알고리즘을 설계하는 것[26] 등이 있다.[27]

사전적 방법은 기본적으로 알고리즘에 의한 결과 예측의 정확성을 희생하고서라도 알고리즘의 복잡성을 제한함으로써 설명 가능성을 높이는 것이므로 설명 가능성이 강하게 요구되는 영역에서 사용된다. 미국의 Fair Credit Reporting Act 또는 Equal Credit Opportunity Act에서 대출 거절의 주된 사유를 대출신청자에게 밝히도록 하는 것,[28] EU가 2018년 제정한 General Data Protection Regulation (GDPR) 제22조에서 자동화된 의사결

19) Selbst & Barocas, op. cit., 1085.
20) 대표적으로, EU 집행위원회의 Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts [Brussels, 21.4.2021 COM(2021) 206 final 2021/0106 (COD)] (이하 'EU AI 법안').
21) Islam et al., op. cit., 4.
22) 예를 들어 몸무게를 예측하는 특징량을 키, 나이, 성별의 3개로 제한하는 것.
23) 특징량이 100만 개라면 인간의 이해를 벗어나지만, 특징량이 위와 같이 3개라면 키, 나이, 성별이 몸무게에 어떠한 영향을 미치는지 직관적으로 이해할 수 있다.
24) 예를 들어 의사결정나무의 가지가 50개가 넘지 않도록 매개변수를 조절하는 것.
25) 학습 과정에서 과도한 복잡성에 부(-)의 값을 할당함으로써 모델이 이해 불가능할 정도로 복잡해지지 않게 하는 것
26) 단조성은 입력값이 증가할 때 결과값이 증가하기만 하거나(양의 관계가 있을 경우) 감소하기만 하는(음의 관계가 있을 경우) 것을 가리킨다. 예를 들어, 수입의 증가가 항상 신용평가점수의 증가를 낳는다면 양자는 단조성이 있다. S Zoldi, "Trusted AI: The Challenge of Monotonicity and Palatability", FICO Blog (Feb. 11, 2021) <https://www.fico.com/blogs/trusted-ai-challenge-monotonicity-and-palatability> (2022. 12. 18. 최종확인).
27) Selbst & Barocas, op. cit., 1110-1113.
28) Ibid, 1112-1113.

정에 대하여 설명을 요구할 권리를 규정한 것, EU의 AI 법안[29]이 시민의 권리와 의무에 큰 영향을 미치는 안전, 입시, 고용, 법집행, 사법절차 등 핵심 영역을 고위험군으로 분류하여, 각종 기준을 준수하고 정보를 보관하며 감독기관의 감독을 받도록 한 것 등이 그 예이다.

2) 사후적 기법

원래의 알고리즘 자체에서는 설명 가능성을 고려하지 않았더라도, 사후적으로 알고리즘을 단순화하거나 특정 요소에 집중하여 분석함으로써 좀 더 설명 가능하고 이해하기 쉬운 대체적인 모델을 만드는 방법도 다수 연구, 개발되고 있다.[30]

알고리즘 생성 단계에서는 설명 가능성을 고려하지 않고 성능을 최우선으로 하여 복잡한 머신러닝 모델을 만들었다가 사후적으로 위 알고리즘을 다시 설명 가능한 수준으로 단순화할 경우 그 결과물은 처음부터 설명 가능한 모델로 개발된 알고리즘에 비해 더 우수한 경우가 많다. 또한 개발자가 굳이 설명 가능한 기법에 의하지 않고 딥러닝 등 복잡하고 우수한 기법을 이용하여 알고리즘을 개발하여 활용한 경우라도, 사후적으로 알고리즘의 하자에 관하여 분쟁이 생겼을 때에는 그 원인을 규명하기 위하여 사후적 XAI 기법을 활용할 수 있다.

다만 사후적 XAI를 통해 추출해낸 알고리즘 의사결정의 이유는 확률적 근사치에 불과하고, 실제 알고리즘을 정확히 반영하는 것이 아니며, 기법별로 각각의 장단점이 있다. 예를 들어 Trivago 사건에서도 활용된 PDP 기법의 경우, 분석 대상 알고리즘의 입출력 데이터를 가지고 나머지 특징량은 고정시킨 채 결과값과의 관계를 분석하고자 하는 특정 특징량(features)의 값만을 변화시켜 가며 그것이 예측결과(predictions)에 미치는 변화의 양(한계효과)을 도식화하여 1개의 그래프로 환산하나, 특징량 간의 상호관계를 무시하고 분석 대상 특징량에만 집중함으로써 특징량 간에 복잡한 상호관계가 존재할 경우 오차가 커질 수 있다.[31] 사후적 XAI 기법 중 하나인 Tree Surrogate 모델(의사결정나무 대리모델)은 AI 모델의 특징량과 예측 결과 데이터를 가지고 의사결정나무 기법의 학습을 수행하여, 판단로직을 조건 분기의 나무로 표현하는 대리 모델을 구축한다. 의사결정나무의 형태로 설명하므로 이해하기 쉬우나, 이는 입력값과 결과값만을 가지고 별도의 학습을 통해 새로운 알고리즘을 생성함으로써 원래의 알고리즘의 대략적인 판단 경향을 이해시키는 것에 불과하다.[32] 그 밖에도 LIME, SHAP, CAM, Permutation Importance,

29) EU AI 법안 제6조 이하.
30) 오오쓰보, 앞의 책, 36-73면.
31) 오오쓰보, 앞의 책, 56-59면.
32) 위의 책, 60-63면.

Integrated Gradients 등 다양한 사후적 XAI 기법이 존재하나, 각각의 장단점과 한계가 있다.[33]

이러한 사후적 XAI 기법의 특성을 고려할 때, 재판에서 위 기법을 이용하여 AI 알고리즘에 의한 의사결정의 이유 내지 원인을 설명하고자 하는 경우, 다음의 점에 유의할 필요가 있다. ① 전문가가 알고리즘의 의사결정 구조를 분석해낼 수 있을 정도로 충분한 입력값과 결과값을 확보하되, 영업비밀 침해를 최소화해야 한다. ② 사후적 XAI는 실제 알고리즘과 동일하지 않고, XAI가 도출하는 설명은 확률적, 대체적인 것에 불과하다는 점, 어떠한 설명을 원하는가에 따라 사용하는 XAI 기법이 달라질 수 있다는 점을 이해해야 한다. ③ 알고리즘을 감정하는 전문가는 각 기법의 특징, 장단점, 당해 사건의 쟁점과 관련한 해당 기법의 적절성 등에 관하여 평이한 언어로 법관 및 일반인을 이해시킬 수 있어야 한다.

3. AI 알고리즘 감정에 필요한 정보의 제출과 보호

현행법상 문서제출 또는 자료제출 및 제출된 비밀의 보호 제도는 민사소송법과 지식재산권 침해 또는 공정거래법 위반으로 인한 손해배상청구소송에 관한 특별민사소송법의 이원적 구조로 이루어져 있다.

먼저 민사소송법상의 문서제출명령에 의하는 방법이 있다. 그 경우 영업비밀에 대하여는 "직업의 비밀"을 제출거부 사유로 주장할 수 있다(민사소송법 제344조 제1항 제3호 단서, 제2항 단서, 민사소송법 제315조 제1항 제2호). 판례는 영업비밀에 대해 비밀로서의 보호가치성과 증거로서의 필요성을 모두 고려하여 제출 여부를 결정하여야 한다는 태도를 취하고 있다.[34] 그러나 실제에 있어서는 영업비밀에 대해 제출명령을 받기도 쉽지 않고, 제출을 강제할 수단도 미약하며, 심리과정에서 제출받은 비밀을 보호할 비밀유지명령 등 장치도 존재하지 않는다.[35]

특허권, 저작권, 상표권 등 지식재산권이나 영업비밀의 침해, 불공정 거래행위로 인한 손해배상청구소송에 관하여는 해당 특별법에서 일종의 특별민사소송법으로서 자료제출명령, 영업비밀의 제출의무 강화와 제출명령 위반자에 대한 진실의제 확대, 제출된 영업비밀 보호를 위한 비밀유지명령, 위반 시의 형사처벌 등을 규정하고 있다. 전체적으로 특별법상의 자료제출명령이 제출범위의 확대, 제출된 비밀의 보호 차원에서 진일보한 것은 사실이나, 이와 같이 자료제출 요건과 절차가 민사소송법과 일원화되지 않아 실무

33) 위의 책, 43-51, 52-55, 64-71면.
34) 대법원 2015. 12. 21.자 2015마4174 결정; 대법원 2016. 7. 1.자 2014마2239 결정.
35) 민사소송절차에서의 영업비밀 보호 방법에 관하여는, 박익환, "민사소송절차와 영업비밀보호 - 부정경쟁방지법상 비밀유지명령을 중심으로 - ", 정보법학 제16권 제1호(2012), 163면 이하 참조.

상 혼란을 초래할 수 있으며, 특허법 등 특별법에서도 자료의 보존의무에 관한 명확한 규정이 없고, 제출명령 불이행 시 자유심증설에 의한 진실의제 외에 마땅한 제재가 없다는 문제점이 있다.[36]

이러한 상황에서는 영업비밀 보관자의 입장에서는 비밀누설의 우려가 크지 않고, 자료 제출을 요구하는 입장에서는 사실의 증명에 충분한 정도의 영업비밀 제출범위를 설정하는 것이 필수적이다.

AI 알고리즘에 관련된 정보로는 ① 프로그램 그 자체, 즉 소스코드(source code)와 프로그래머의 주석(comment), ② 알고리즘이 생성, 변경 또는 삭제된 이력의 전자기록, ③ 알고리즘 학습에 이용된 학습데이터, ④ 개발자가 구상한 알고리즘의 구조와 가중치(weight), ⑤ 완성된 알고리즘에 입력값을 넣었을 때 생성되는 결과값의 조합 등 여러 유형이 존재한다. 그 중 ①~④의 데이터에 대해서는 영업비밀 침해 우려에 따라 법원이 선뜻 제출명령을 내리기 어려운 경우라도, 알고리즘 자체 또는 가중치 자체가 아닌 ⑤의 데이터는 영업비밀 침해의 정도가 비교적 적으므로, 제출된 비밀의 보호장치가 미비한 일반 민사소송절차에서도 제출명령을 내리기가 비교적 용이할 것이다.

다만 비밀의 노출을 최소화하면서도 실효적인 감정과 심리를 행하기 위해서는, 절차의 초기단계에서부터 데이터 전문가가 관여하여 감정에 사용될 수 있는 XAI 기법과 해당 기법에 필요한 정보의 성격과 양을 조언해줄 수 있어야 한다. 또한 법원으로서도 인공지능 알고리즘이나 XAI 기법의 확률적 불투명성을 고려하여 사실인정 시에 지나치게 고도의 증명을 요구하지 않아야 할 것이다.

무엇보다도, 민사소송법과 특별법에 산재된 문서제출명령 제도를 일원화하여 전체적인 자료 제출범위를 확대하고 제출된 비밀의 보호 장치를 보강하며, 변개·삭제가 쉬운 전자정보의 특성을 고려하여 영미법계의 litigation hold 수준까지 증거보존의무를 강화하고,[37] 제출의무 불이행에 대하여 유연하고 다양한 제재 수단을 도입하는 등 전체적인 제도의 개선이 필요하다.

---

36) 한애라, "영업비밀 열람과 보호 – 공정거래위원회 심의절차와 그 처분취소소송절차에서의 '외부 대리인 한정 열람'을 중심으로", 사법 58호(2021), 607면 이하; 서울대학교 산학협력단(연구책임자 이계정), "증거수집·조사절차 개선을 통한 충실한 특허소송 심리방안에 관한 연구"(법원행정처 연구용역)(2020), 21면 이하 참조.

37) 미합중국 연방민사소송규칙(Federal Rules of Civil Procedure)은 전자정보의 보존에 관하여, 당사자의 변호사는 소송이 제기될 수 있음을 알게 된 즉시 의뢰인에게 정보가 변경(altered)되거나 파괴(destroyed)되지 않도록 보유 중인 관련정보의 증거보전조치를 취할 것을 알려야 하며[규칙 제26조 (f)], 만약 당사자가 전자적 정보의 보전을 위한 합리적인 조치를 취하지 않아 전자적 정보가 멸실된 경우, 법원은 이로 인하여 상대방이 불이익을 받은 범위에서 이를 치유할 조치를 취할 수 있고, 상대방이 해당 정보를 소송에서 이용하지 못하게 할 의도로 당사자가 행위하였음이 인정되는 경우에는, 법원이 배심원에게 그 정보가 그 당사자에게 불리한 것이라고 추정하도록 명하거나, 그 당사자의 소를 각하(dismiss)하거나, 그 당사자를 패소시키는 무변론 청구인용판결(default judgment)을 내릴 수 있다고 규정하고 있다[제37조 (e)].

4. AI 알고리즘 감정에 있어 사감정의 활용 가능성

우리 민사소송법상 전문가의 의견을 소송에 현출시키는 방법으로는 법원 지정 감정인의 감정(경우에 따라서는 감정촉탁) 및 감정인 신문과 당사자가 사적으로 선임한 전문가가 작성한 의견서의 서증 제출 및 감정증언이 있다(전문가가 소송절차에 참여하는 방법으로는 전문심리위원 제도도 있으나, 전문심리위원의 설명이나 의견은 독립된 증거자료가 되지 않고 전문 지식을 보충하는 참고자료가 될 뿐이라는 점에서 근본적인 차이점이 있다[38]). 이 중 민사소송절차가 기본으로 삼는 것은 법원 지정 감정인의 감정이지만, 당사자 선임 전문가에 의한 감정 의견서(소위 '사감정')의 이용도 점점 많아지고 있다.

법원 감정인은 법관에 준하는 독립성이 요구되고 감정을 신청한 당사자와 감정인 간에 어떠한 직접적인 법률관계도 없다. 반면에 사감정의 경우, 당사자가 소송 외에서 학식 또는 경험 있는 제3자에게 직접 감정을 의뢰하여 감정의견서를 받고, 이 감정의견서를 서증의 형태로 제출하며,[39] 경우에 따라서는 감정증인("특별한 학식과 경험에 의하여 알게 된 사실에 관한 신문")으로 증언하기도 하나(민사소송법 제340조), 절차적으로 일반적인 서증, 증언과 달리 취급되지는 않는다.[40] 이러한 이유에서 법원은 사감정에 대하여 다소 유보적인 입장을 취하고 있다.[41]

그러나 실무에서는 이미 사감정이 점점 더 활발하게 활용되고 있다. 특히 경제전문가의 경제분석이 요구되는 공정거래법 위반에 따른 손해배상청구 사건[42]이나 증권 관련 손해배상청구 사건[43] 등에서는 경우에 따라 사감정이 법원 감정보다 더 바람직한 증거방법으로 권장되기도 한다.

사감정에 투입되는 비용이 소송비용 부담 및 확정절차를 통해 회복될 수 없음에도 사감정이 이용되는 원인은 다음과 같이 정리할 수 있다. 첫째, 경제적 이해관계가 충분히 큰 사건에서는 당사자가 분쟁 초기 단계에서 선제적으로 전문가의 의견을 구할 필요가 있고, 특히 일방 당사자가 사감정인의 감정의견서를 제출하였다면 타방 당사자도 자기 측 사감정인의 감정의견서를 제출함으로써 이를 반박할 필요가 있다. 둘째, 비전형적인 감정의 경우 적절한 감정인을 지정하거나 감정인 보수를 산정하는 것이 여의치 않으므

38) 강성수, 『전문가 감정 및 전문심리위원 제도의 개선 방안에 관한 연구』, 사법정책연구원(2016), 45면.
39) 강수미, "사감정의 소송법상 취급", 민사소송 vol. 10(2006), 100면 이하.
40) 다만 "특허법원 민사항소심 소송절차 안내"에서는 전문가 증인에 관하여 증인의 전문성과 객관성을 확인할 수 있는 전문가증인 기본사항 확인서의 첨부, 전문가증인 진술서의 제출 등 별도 규정을 두고 있다. 위 안내 6면 및 첨부 [11].
41) 대법원 2010. 5. 13. 선고 2010다6222 판결.
42) 고학수, "소송 과정에서의 경제전문가 활용에 대한 시론 - 소위 '군납유 사건'의 시사점을 중심으로 -", 법경제학연구 제11권 제3호(2014. 12.), 360면 이하.
43) 김주영, "증권소송에 있어서의 전문가 감정 활용기준", 저스티스 통권 102호(2008), 181면 이하.

로, 각 당사자가 전문가를 물색하여 각기 의견서를 제출하도록 하는 것이 대안이 될 수 있다. 셋째, 법원 감정인에 의한 감정이 행해졌더라도, 그 감정 결과의 신빙성을 다투기 위해, 경우에 따라서는 재감정 결정을 이끌어내기 위해 사감정이 이용될 수 있다. 넷째, 이미 최고의 전문가가 당사자에 의하여 선점되었거나 해당 분야의 전문가가 국내에 없는 경우, 당사자로 하여금 사감정을 활용하도록 할 수밖에 없다. 끝으로, 감정에 필요한 사실의 수집 및 감정 방법론의 선택 자체가 고도로 전문적인 영역이어서 법원이 적절한 통제를 할 수 없고 감정인에게 지나치게 넓은 재량이 인정되는 경우, 사감정은 법원이 감정인에게 전적으로 의존하는 것에 대한 대안이 될 수 있다.[44)]

AI 알고리즘의 감정은 비전형적인 감정으로서 다양한 기법이 존재하고 전문가별로 편차가 큰 점, 당사자 또는 소송대리인이 사건의 쟁점을 이해하기 위해 사건의 초기 단계에서부터 전문가에게 의존할 가능성이 큰 점 등을 고려할 때, 사감정, 혹은 사감정과 법원 감정을 함께 활용할 필요성이 큰 분야라 할 것이다. 사감정이 적절하게 활용된다면 양측 전문가의 조력과 공방을 통하여 법원이 쟁점을 명확히 하고 전문적, 과학적인 내용을 이해하는 데 큰 도움을 받을 수 있다.

법원이 사감정을 사실인정에 활용하는 경우, 사감정인의 객관성 · 중립성을 어떻게 확보할 것인가, 사감정인의 감정의 신뢰성을 어떻게 판단할 것인가, 사감정인을 어떻게 절차에 참여시킬 것인가가 문제된다.[45)]

먼저 사감정인의 객관성 · 중립성 확보에 관해서는 미국 연방민사소송규칙(Federal Rules of Civil Procedure, FRCP)상의 전문가 증인 관련 정보공개 규정을 참고할 수 있다. FRCP 26(a)(2)(B)는 전문가 증인에 대하여 ① 증언할 의견 전체 및 그 근거와 이유에 관한 완전한 진술, ② 증인이 의견을 형성할 때 고려한 사실 또는 데이터, ③ 의견의 요약 또는 의견의 보강을 위해 사용된 증거, ④ 증인의 자격(최근 10년간 증인이 저술한 모든 간행물의 목록 포함), ⑤ 최근 4년간 증인이 전문가 증인으로서 또는 증언조서에 의해 증언한 다른 사건들의 목록, ⑥ 당해 증언 및 의견서 작성을 위해서 지급받은 보수 등을 밝히고 선서할 것을 요구한다.[46)] 위 사항에 관한 전문가 진술의 진실성은 위증죄 처벌로써 담보되고, 법원과 상대방 당사자는 공개된 자료를 통해 전문가의 자격과 전문가 의견서의 타당성을 판단하거나 탄핵할 수 있다. 우리 민사소송절차에서도 사감정 의견서나 전문가 증인을 사실인정에 활용하고자 하는 경우, 법원의 준비명령 등을 통해 사감정인이 사감정의견서에 위와 같은 내용을 밝히도록 하고, 증인으로 출석하여 선서하고

44) 김주영, 앞의 글, 180-181면.

45) 사감정 비용을 어떻게 소송절차에 내재화할 것인가도 문제되나, 경제적 이해관계가 전문가 비용을 초과하는 사건에서는 추후 소송비용으로 상환받을 수 있는지 여부에 관계없이 당사자가 일단 전문가를 사적으로 선임할 것이고(고학수, 앞의 글, 376-378면), 법원 감정인의 경우에도 향후 승소하여 감정비용을 상환받을 수 있는지 여부가 불투명한 것은 마찬가지이므로, 이 문제는 이 논문에서는 일단 다루지 않기로 한다.

46) 강성수, 앞의 글, 104면 이하.

위 내용이 사실임을 증언하게 함으로써 사감정인의 객관성, 중립성을 강화할 수 있을 것이다. 특허법원의 전문가 증인 기본사항 확인서 또한 이러한 기능을 수행한다.[47)]

다음으로 사감정 의견의 신뢰성 판단에 있어서는, 대법원 판례가 제시한 "① 그 이론이나 기술이 실험될 수 있는 것인지, ② 이론이나 기술에 관하여 관련 전문가 집단의 검토가 이루어지고 공표된 것인지, ③ 오차율 및 그 기술의 운용을 통제하는 기준이 존재하고 유지되는지, ④ 그 해당 분야에서 일반적으로 승인되는 이론인지, ⑤ 기초자료와 그로부터 도출된 결론 사이에 해결할 수 없는 분석적 차이가 존재하지는 않는지"의 다섯 가지 기준에 의할 수 있다.[48)] 과학 전문가가 아닌 법원의 판사가 위 다섯 가지 기준을 충족하였는지 여부를 판단하기란 쉽지 않다. 이러한 경우, 양측 사감정인의 공방을 통해 각 전문가 의견의 강점과 약점을 법원이 좀 더 잘 파악할 수 있고, 양측 사감정인 외에 법원 감정인이 지정된 경우에도 법원이 법원 감정인의 감정결과에 전적으로 의존하지 않을 수 있다.

사감정의 효용을 극대화하기 위해서는 사건의 초기 단계에서부터 양측의 전문가를 절차에 참여시킬 필요가 있다. 만약 양측이 소장과 답변서에서 각기 전문가 감정의견서를 서증으로 제출하였다면, 준비기일에 양측 전문가를 참여하게 하여 사건의 과학적 쟁점을 정리하고, 정리된 쟁점에 집중하여 쌍방 전문가가 의견을 같이 하는 부분과 의견을 달리하는 부분을 특정하며, 의견이 다른 경우 그 사유와 상대방 의견에 대한 반박을 밝힌 공동의견서를 제출하게 하는 방법(나아가 대질신문하는 방법)을 고려해볼 수 있다. 또한 감정에 필요한 자료의 제출에 있어서도, 쌍방 전문가가 절차 초기 단계부터 적절한 XAI 기법과 해당 기법에 의해 알고리즘을 분석하기 위해 필요한 데이터의 내용과 범위에 관하여 의견을 제출하게 함으로써, 영업비밀 침해를 최소화하면서도 알고리즘의 분석이 가능한 적정 자료제출 범위를 확정할 수 있을 것이다.

결론적으로, 법원 감정을 고집하지 않고 양측 당사자가 선임한 전문가를 적절하게 활용한다면, 쟁점에 관한 이해를 높이고 실체적 진실을 발견하는 데에 기여할 수 있다. 이를 위해서는 사감정을 지금처럼 서증과 증언의 영역에 내버려두지 말고, 법원 감정에 준하는 정도로 객관성, 중립성을 강화함과 아울러 민사소송절차에 정식으로 편입시킬 필요가 있다.

IV. 결어

47) "특허법원 민사항소심 소송절차 안내" [첨부 11].

48) 대법원 2011. 9. 2. 선고 2009다52649 전원합의체 판결 다수의견. 이 중 ① 내지 ④ 기준은 과학적 증거 및 이에 관한 전문가 증인의 증거능력에 관한 미국 연방대법원의 Daubert 판결[Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579(1993)]을 따른 것이고, ⑤ 기준은 Daubert 판결의 후속 판결인 Joiner 판결[General Electric v. Joiner, 522 U.S. 136(1997)]에서 제시한 기준을 따른 것이다(송혜정, "과학적 증거에 대한 법원의 판단기준", 재판자료 123집(2011), 572면 이하).

이상으로 호주 Trivago 판결이 민사소송절차에서 AI 알고리즘을 어떻게 감정, 심리하고 판단하였는지를 소개한 후, 우리 민사소송절차에서 설명 가능성이 떨어지는 AI 알고리즘을 심리함에 있어 발생할 수 있는 문제점과 해결 방안을 XAI의 기본적인 특성의 이해, AI 알고리즘 감정에 필요한 데이터의 제출과 비밀 보호, 전문가의 활용을 중심으로 검토하였다.

블랙박스적 특성을 가지는 AI 알고리즘이라 하더라도 법적 책임을 묻기 위해 필요한 정도로 그 핵심적 특성을 감정해 내는 것이 불가능한 것은 아니다. 법원이 점점 증가하는 AI 알고리즘 관련 사건을 제대로 심리하여 타당한 결론을 내리기 위해서는, 감정에 필요한 데이터를 확보함에 있어 영업비밀과 실체적 진실 발견 사이에 적절한 균형점을 찾고, 제출된 데이터에 포함된 영업비밀을 보호하는 장치를 강화하며, 법원 감정인뿐만 아니라 쌍방 당사자가 선임한 사적 전문가도 적극적으로 절차에 참여시킴으로써 이들의 의견 교환을 통해 쟁점을 명확히 하고 법원의 이해를 높일 필요가 있다. 이 점에서는 우리나라 민사소송절차보다 영미법계의 민사소송절차가 훨씬 유연하므로 이를 적극 참고하는 것이 바람직하다. 아울러 AI 알고리즘의 불투명성을 극복하기 위한 국제적인 입법 동향에도 유의하여야 할 것이다.