인지과학, 제32권 제4호
Korean Journal of Cognitive Science
2021, Vol. 32, No. 4, 169~191
http://dx.doi.org/10.19066/cogsci.2021.32.4.001

사건 관여도와 심각성이 인공지능 판결에 대한 수용도에 미치는 효과

도 은 영 (광운대학교 산업심리학과)
이 국 희† (경기대학교 교양학부)
정 지 은 (경기대학교 범죄심리학과)

인공지능이 발달함에 따라 많은 인간 전문가들의 일자리가 위협당하고 있으며 이는 법조계 또한 해당한다. 본 연구는 실제로 인공지능이 법조계에 종사하는 사람들, 특히 최종 판결을 내리는 판사의 역할까지 대체할 수 있을지에 대해 다루어보고자 하였다. 본 연구는 고유성 무시의 관점에서, 이를 판결 대상자와 얼마나 가까운 사이인지(관여도)와 사건이 얼마나 심각한지(심각성)가 인공지능 판사(실험 1)가 내리는 판결을 수용하는 정도와 전체 배심원 중 인공지능 배심원(실험 2)을 배치하는 정도에 미치는 효과를 확인하는 것으로 시행하였다. 실험 1에서는 관여도와 심각성의 상호작용 효과가 나타났다. 구체적으로, 관여도가 낮을 때는 심각성이 높은 사건(vs. 낮은 사건)에서 더 인공지능 판사 수용도가 높았지만, 관여도가 높을 때는 심각성이 낮은 사건(vs. 높은 사건)에서 더 인공지능 판사 수용도가 높았다. 실험 2에서도 실험 1과 동일한 상호작용이 발생하였다. 구체적으로, 관여도가 낮을 때는 심각성이 높은 사건(vs. 낮은 사건)에서 인공지능 배심원 수를 더 많이 배정하였지만, 관여도가 높을 때는 심각성이 낮은 사건(vs. 높은 사건)에서 인공지능 배심원 수가 더 많아졌다. 본 연구는 인공지능의 법적 판단에 대한 국내 최초의 실험연구라는 점과 법률 전문가들의 일자리에 대한 전망을 제시한다는 점에서 시사점을 가진다.

주제어 : 인공지능, 판사, 관여도, 심각성, 판결 수용

†교신저자: 이국희, 경기대학교 교양학부, (16227) 경기도 수원시 영통구 광교산로 154-42(이의동) 경기대학교 수원캠퍼스 성신관 2205호
연구 분야: 범주기반 귀납추론, 속성 일반화, 사회적 배제, 사회적 지위, 행복
Tel: 031-249-1494, E-mail: leegh1983@gmail.com

서론

인공지능의 등장은 우리 사회에 많은 논쟁을 점화시켰다. 그중에서도 사람들이 가장 궁금해하는 부분은 '과연 미래에 내 일자리를 인공지능에게 빼앗기게 될 것인가'일 것이다. 실제로 인공지능이 대체할 것으로 전망되는 직업들이 많다. LG경제연구원(2018)은 우리나라 전체 일자리의 43%가 인공지능 대체 가능성이 높은 고위험군이며 특히 사무직, 판매직, 기계조작직이 가장 위험하다고 발표하였다. 이러한 전망은 기업 현장에서도 나타난다. 한국개발연구원(2021)이 발표한 'AI에 대한 기업체 인식 및 실태 조사 결과'에 의하면, 조사대상인 1000개의 기업들 가운데 48.8%가 인공지능이 인력을 대체할 것이라고 응답했으며, 50.1%가 인공지능이 자사의 직무를 대체할 것이라고 응답하였다. 기업들 또한 크기와 분야를 막론하고 기업 현장에 인간을 대신하는 인공지능이 비중이 늘어날 것이라고 보는 것이다.

법조계도 이러한 이슈에서 자유롭지 않다. 2017년 미국의 연방순회항소법원(CAFC) 등에서 30여년 간 판사로 재직했던 랜들 레이더는 가까운 미래에 인공지능이 판사를 비롯하여 법조계 대부분의 일자리를 대체할 것이며, 심지어는 인공지능이 인간 판사보다 더 빠르고 공정한 판결을 내릴 것이라고 전망하였다(연규욱, 2017). 인공지능은 인간 판사의 편향된 관점과 외부 압력으로부터 자유롭다고 보았기 때문이다. 인공지능이 전문의, 교수, 변호사와 같은 고임금 전문직까지 빠르게 적용되고 있는 점까지 고려할 때(LG경제연구원, 2018), 머지않은 미래에 사람이 아닌 인공지능 변호사에게 법률 상담을 받고 인공지능 판사에게 최종 판결을 선고받는 날을 쉽게 상상하게 된다.

그렇다면 인공지능들은 구체적으로 법조계에 어떻게 사용되고 있으며, 사용될 것인가? 인공지능 변호사로는 미국의 로스(ROSS)가 활약하고 있다. 로스는 IBM 인공지능 '왓슨(Watson)'이 발전한 모델로, 2016년 미국 뉴욕의 대형 로펌인 베이커앤드호스테틀러(Baker & Hostetler)에 취직되기도 하였다(정원엽, 이기준, 2016). 로스는 인간과 유사한 추론이 가능하며, 자연어 처리 기술이 발달하여 사람들의 일상 언어를 이해할 수 있다. 이를 바탕으로 사람이 질문을 하면 초당 10억장의 법률문서를 분석하여 질문에 맞는 답변을 내놓는 방식으로 법률 상담이 가능하다. 인공지능은 사법의 영역에서 판사를 돕는 역할을 수행하기도 한다. 미국에서는 판사들이 범죄자들의 가석방을 심사하거나 판결과정에서 형량 결정을 할 때, 인공지능이 분석한 재범위험성 예측 결과를 활용한다(양종모, 2018). 또한 인공지능은 멕시코에서 법관에게 연금 수급자로서 적격한지에 대한 결정을 권고하거나, 북아일랜드, 잉글랜드, 웨일스에서 소액사건의 온라인 소송을 지원하기도 한다(정영화, 2020). 그러나 법에 관한 궁극적인 궁금증은 인공지능이 법관을 대신하여 인간에게 판결을 내릴 수 있는지일 것이다. '다음과 같은 이유로 X는 Y죄에 대하여 유죄이다'라고 판결할 수 있는 시스템인 '판결기계(Judgment Machine)'로서의 인공지능(양종모, 2016)이 실제로 적용될 수 있을 것인가?

인공지능이 사람 판사를 대체하는 것은 사실상 불가능할 것이라는 의견들도 있다. 양종모(2018)는 유무죄 판결에 필요한 알고리즘 구현 가능성과 판결기계 사용에 대한 사회적 합의 측면을 고려할 때 인공지능 판사의 실제 적용은 어려울 것이라고 보았다. 유무죄를 판단하는 사실인정 과정 중에는 사회적 가치가 반영되거나 사실관계 가중치가 수시로 변동되는 등 통계적 모델에 반영하기 어려운 요소들이 존재한다. 이로 인해 유무죄를 판단하는 알고리즘을 구현해내는 것은 실질적으로 어렵다. 또한 판결기계의 신뢰성 검증이 어렵고 인공지능이 판결을 내려도 승복하기 어려울 것이라는 점 등이 그 이유로 꼽혔다. 미국에서는 인공지능 컴파스(COMPAS)가 형량, 가석방, 보석 등의 판결에서 흑인들에게 편파적으로 작용한다는 폭로가 발생하기도 했다(Angwin, Larson, Mattu, & Kirchner, 2016). 또한 컴파스의 알고리즘이 여성보다 남성에게 더 불리하게 작용하기 때문에 남성인 자신이 받은 판결의 공정성을 신뢰할 수 없다는 탄원도 있었다(오요한, 홍성욱, 2018). 그러나 인공지능 판사의 실현 가능성에 대한 낙관적인 시각과 국민들의 인공지능 판사 필요성에 대한 목소리 또한 여전히 존재한다(양새롬, 2020).

본 연구는 이러한 이슈들을 배경으로, 인공지능이 판단기계로서 최종 판결을 내리는 것을 사람들이 수용할 수 있는지에 대해 살펴보려고 한다. 특히 사건과 얼마나 관련된 사람인지와 사건이 얼마나 심각한지에 따라 인공지능이 내린 최종 판결에 대한 수용도가 달라지는지 확인해보고자 한다. 이와 같은 검증은 몇 가지 측면에서 의의를 가진다. 먼저, 국내에서 처음으로 인공지능의 법적 판단에 대한 일반 사람들의 수용 및 인식을 실험적으로 확인한다. 둘째, 인공지능의 법적 판단에 대한 논의를 사건의 심각성과 관여도 측면으로 확장한다. 셋째, 법률 판단을 전문으로 하는 인공지능의 등장이 판사, 검사, 변호사와 같은 법률 전문가의 영역에 어떤 영향을 미칠 것인지 가늠해 볼 수 있게 한다.

이론적 배경

인공지능의 판단에 대한 수용

인공지능(Artificial intelligence, AI)이 삶에 적용되기 시작하면서, 인공지능에 대한 사람들의 수용 및 심리적 저항에 대한 연구들도 활발해졌다. 기존 연구들에 의하면, 특정 분야에서 사람들은 인공지능이 자신의 삶에 큰 영향을 줄 수 있는 의사결정을 내리는 것을 꺼려하는 것으로 나타났다(Bigman & Gray, 2018; Longoni, Bonezzi, & Morewedge, 2019). 예를 들어, Longoni, Bonezzi, & Morewedge (2019)의 연구에 의하면 의료 인공지능은 비용 효율성과 일부 영역에서 전문가 수준의 정확성과 지녔음에도 불구하고(Leachman & Merlino, 2017), 인간 의사에 비해 덜 선호되는 것으로 나타났다. 사람들은 병의 진단부터 치료에 이르기까지 대부분의 의학적 결정에서 인간 의

사보다 인공지능이 제공하는 의료 서비스를 더 꺼렸다. 또한 인간 의사와 인공지능의 판단 및 치료 확률이 동일할 때도, 인공지능이 제공하는 서비스에 더 적은 비용을 지불하고자 하였다. 의료 인공지능에 대한 심리적 저항은 인공지능이 직접적인 판단을 내리는 것이 아니라, 인간 의사를 지원하는 조력자 역할에 머물 때야 비로소 줄어들었다. 이러한 심리적 저항감은 인공지능이 도덕적 판단을 내려야 하는 상황에서도 적용되었다. Bigman & Gray (2018)의 연구에서, 사람들은 운전, 법, 의학, 군대에 이르기까지 다양한 도덕적 판단 상황에서 인공지능보다 인간이 결정내리는 것을 더 선호하였다. 이러한 결과는 도덕적 결정의 결과가 부정적이든 긍정적이든 상관없이 나타났다. 인공지능이 도덕적 결정을 내리는 것에 대한 반감은 의료 AI연구와 마찬가지로 인공지능이 자문 역할로 제한되거나, 인공지능의 전문지식이 극도로 두드러지는 상황에서야 다소 줄어들었다. 인공지능 판단 수용에 대해 가장 활발한 연구가 이루어지고 있는 분야 중 하나는 자율주행 자동차(Autonomous vehicles)이다. 자율주행 자동차는 소비자와 사회에 환경오염 감소 및 교통체증 감소와 같은 많은 이익을 줄 것으로 예측된다(Gill, 2020). 그러나 인공지능이 어떤 판단을 내리느냐에 따라 차량 내부의 탑승자와 차량 외부의 보행자의 생과 사가 결정되므로 실제 자율주행 자동차의 운행에 대해 많은 우려를 사고 있다(Gill, 2020).

그렇다면 사람들은 왜 인간의 판단보다 인공지능의 판단에 더 심리적 저항감을 느끼는 것인가? 심리적 저항감을 느끼는 원인으로는 기술에 대한 이해 및 신뢰 부족(Hengstler, Enkel, & Duelli, 2016), 지각된 인공지능의 마음 부족(Bigman & Gray, 2018), 고유성 무시(Granulo, Fuchs, & Puntoni, 2021; Longoni, Bonezzi, & Morewedge, 2019; Yun, Lee, & Kim, 2021) 등이 제안되었다. 이중 고유성 무시(uniqueness neglect)는 두 가지 근본적인 신념의 불일치에서 발생한다. 첫째, 사람들은 그들 스스로를 독특하고 남들과는 다른 존재로 여긴다. 둘째, 사람들은 기계를 모든 경우에 표준화되고 일괄적인 방식으로 작동하는 것으로 여긴다. 종합하자면, 사람들은 인공지능이 인간만큼 자신의 사례의 고유성을 고려할 수 없다고 믿기 때문에 인공지능의 판단을 더 꺼린다는 것이다. 따라서 고유성 무시는 개인의 특성이 많이 고려되는 서비스, 즉 의료(Longoni, Bonezzi, & Morewedge, 2019)와 같은 분야의 인공지능 저항감을 연구할 때 많이 적용된다.

고유성 무시는 법적 맥락에서도 적용될 수 있을 것으로 예측된다. 의료 판단과 법률 판단 사이에 몇 가지 공통점들이 있기 때문이다. 먼저, 판단 결과가 개인의 삶에 큰 영향을 미친다. 병을 진단하고 치료하는 것은 건강 및 삶과 죽음에, 유무죄를 판단하는 것은 사회생활, 명성, 재산 등에 지대한 영향을 준다. 이것들이 개인의 삶과 행복의 한 축을 이루는 중요 가치들이라는 점을 고려할 때, 사람들은 보다 전문적이고 자신에게 유리한 결과를 원하게 된다. 또한 병을 진단할 때와 같이, 법적 판결을 내릴 때 각 사건별로 피고인의 개인적 특성, 범행 동기, 범죄의 결과 등의 고유성을 고려하게 된다. 따라서 본 연구는 고유성 무시를 법적 판단의 영역으로 확장하여, 인공지능이 판결내리는 것에 대해 사람들이 납득하는 정도인 '수용도'를 조사해보고자 한다.

사건의 심각성과 관여도

인공지능의 법률적 판단에 대한 수용에 영향을 미칠 수 있는 두 요인으로 사건의 심각성과 관여도에 주목하고자 한다. 두 요인에 따라 사람들이 각 판단 사례의 고유성에 주목하는 정도가 달라질 수 있으며, 법적, 도덕적 의사결정에 영향을 줄 수 있기 때문이다.

먼저 '사건의 심각성'은 피고가 저지른 범죄의 질이 얼마나 나쁜지, 피해를 입힌 정도가 얼마나 큰지를 의미한다. 예를 들어, 같은 교통사고라도 피해 정도나 범행 당시 범죄 인식 유무에 따라 심각성이 높거나 낮을 수 있다. 인명 피해가 없는 가벼운 접촉 사고에 비해 사람을 죽인 음주 운전이 더 심각한 사건인 것이다. 사건의 심각성이 높을 때는 낮을 때에 비해, 사람들의 인공지능의 판결 수용도가 낮을 수 있는데, 이는 사건이 사회에 미치는 파장이 더 큰 만큼, 인간 판사가 판결을 내림으로써 인공지능이 고려하지 못하는 사건의 고유성과 사회적 가치를 판결에 반영해야 한다고 생각했기 때문일 수 있다.

두 번째 요인인 '사건 관여도'는 판결 대상자인 피고인과 얼마나 가까운 사이인지를 의미한다. 구체적으로, 제 3자(예: 피고 F)는 나와 관여도가 낮은 반면 가족(예: 부친 혹은 모친)은 나와 관여도가 높다. Lee와 Holyoak(2020)에 의하면, 범법자가 낯선 이일 때보다 자신의 형제 혹은 자매일 때, 대상이 실제로 범죄를 저질렀다고 믿는 정도와 그들이 저지른 범죄의 객관적인 비윤리성 정도, 경찰에 신고할 의사 정도가 모두 낮은 것으로 나타났다. 도덕적 의사결정시, 자신과 관여도가 낮은 타인보다 관여도가 높은 가족을 더 옹호하는 방식으로 범죄 상황을 해석한 것이다. 심지어 상황적 모호함을 제거하여 추론의 일관성이 줄어드는 상황에서도 앞선 결과처럼 범죄 대상자와의 관계에 따라 의사결정이 달라지는 것으로 나타났다. 따라서, 사람들은 관여도가 낮을 때에 비해 높을 때, 인공지능이 내린 판결을 수용하는 정도가 더 낮을 수 있는데, 이는 자신과 관여도가 높은 대상일수록 표준화된 절차로 판결한다고 여겨지는 인공지능 대신, 명문화되지 않은 사회적 가치와 개인의 고유성 등을 호소하여 정상참작이 가능하다고 여겨지는 인간 판사를 더 선호하게 되기 때문일 수 있다.

특히 사건의 심각성과 사건 관여도 사이에는 상호작용 효과가 나타날 것으로 예측된다. Weidman, Sowden, Berg와 Kross(2020)는 사람들이 범죄자가 자신과 가까운 관계에 있는 사람일수록 그렇지 않은 사람일 때보다 덜 부도덕하게 평가하며, 경찰로부터 보호하려 한다는 사실을 발견했다. 이러한 경향성은 특히 사건의 심각성이 낮을 때(예: 불법 다운로드)보다 높을 때(예: 신용카드 사기) 더 강하게 나타났다. 즉, 범죄자와의 관여도가 높고 그가 벌인 사건의 심각성이 높을 때, 사람들은 가장 범죄자를 보호하는 쪽으로 판단을 내린 것이다. 이와 같은 연구를 고려할 때, 나와 가까운 사람의 판결(사건 관여도가 높은 상황)에서는 사건의 심각성이 낮을 때보다 높을 때 인공지능이 내린 판결에 대한 수용도가 더 낮을 것이다. 그러나 나와 먼 제3자의 판결(사건 관여도가 낮은 상황)을 바라보는 입장에서는 사건의 심각성이 낮을 때보다 높을 때 인공지능

이 내린 판결에 대한 수용도가 더 높을 것이다. 사람들은 내집단(in-group)의 구성원보다 외집단(out-group)의 구성원을 더 가혹하게 처벌하는 경향이 있기 때문에(Martin, Young, & McAuliffe, 2020), 사건이 심각할수록 개인의 고유성이 고려되어 감형되기 보다는 법에 적힌 그대로 심판받기를 원하게 되기 때문이다.

실험 1: 인공지능 판사 수용도

방 법

설계 및 참가자

실험 1은 사건 관여도 2 (낮음 vs. 높음) × 사건 심각성 2 (낮음 vs. 높음)가 해당 사건을 인공지능 판사가 맡아 판결하는 것에 대한 수용도에 미치는 효과를 확인하기 위한 것으로 참가자간 요인설계(between-subjects design)를 채택하여 이루어졌다. 연구를 위해 18-23세(M = 19.34, SD = .71) 한국 소재 대학교 학부생 340명(남 87, 여 253)이 참여하였다. 참가자들은 모두 한국인이었고, 모국어는 한국어였다. 실험은 수업의 일환으로 이루어졌고, 참가자들은 네 가지 참가자간 조건 중 하나에 무작위로 배정되었다.

재료 및 절차

실험 1의 참가자들은 사건에 대한 내용을 읽고, 해당 사건에 대한 판결을 인공지능 판사가 내리는 것을 수용할 수 있는지 평가하는 과제를 수행하였다. 실험 자극은 MATLAB Psychophysics Toolbox로 만들어졌고, 노트북을 활용하여 제시하였다(Brainard, 1997; Pelli, 1997). 참가자가 인공지능 판사 수용도를 평가해야 하는 사건은 세 가지(교통사고, 성범죄, 저작권 침해)로, 각 사건은 관여도와 심각성 조건에 따라 큰 줄기는 유사하지만, 세부사항에서 다소 차이가 있었다.

표 1은 유사한 교통사고 사건이 조건에 따라 어떻게 제시되었는지를 보여준다. 먼저 관여도의 측면에서 보면, 관여도가 낮은 조건에서는 범행 당사자가 'A'라고 제시되었지만, 관여도가 높은 조건에서는 '당신의 부친'이라고 제시된 것을 확인할 수 있다. 즉, 관여도가 낮은 조건에서는 'A'라고 제시하면서 나와 직접적인 관련성이 없는 타인의 이야기라는 느낌을 주었지만, 관여도가 높은 조건에서는 '부친'이라고 제시하면서 나와 직접적으로 관련 있는 사람의 이야기라는 느낌을 주었다.

사건의 심각성 차원에서 보면, 심각성이 낮은 조건에서는 '가해 후 도주했으나, 사고 사실을

<표 1> 사건 관여도와 사건 심각성에 따른 스토리의 예 (교통사고 사건)

구분		관여도	
		낮음	높음
심각성	낮음	[A]는 도로에서 차선을 변경하다 옆 차로를 주행 중이던 승용차를 치고 그대로 현장을 떠난 혐의로 기소되었다. 충돌 직후 피해 차량 운전자는 갓길에 차를 세웠으나, A씨의 차량을 뒤쫓지는 않았다. 피해 차량 운전자와 동승자는 다치지 않았으나, 차량 수리비가 소요되었다. 법정에서 A는 '차선변경을 시도하던 중 경계석을 쳤다고 생각했지, 피해차량을 치었다는 인식을 하지 못했다'고 주장했다.	[당신의 부친]이 도로에서 차선을 변경하다 옆 차로를 주행 중이던 승용차를 치고 그대로 현장을 떠난 혐의로 기소되었다. 충돌 직후 피해 차량 운전자는 갓길에 차를 세웠으나, 부친의 차량을 뒤쫓지는 않았다. 피해 차량 운전자와 동승자는 다치지 않았으나, 차량 수리비가 소요되었다. 법정에서 부친은 '차선변경을 시도하던 중 경계석을 쳤다고 생각했지, 피해차량을 치었다는 인식을 하지 못했다'고 주장했다.
	높음	[A]는 21:00경 혈중알코올농도 0.07%(면허정지 수준) 정도로 술에 취한 상태로 트럭으로 주행하던 중 갓길에서 낚시를 준비하던 피해자를 들이받은 뒤 구호조치를 하지 않은 채 현장을 벗어난 혐의로 기소되었다. 피해자는 병원으로 옮겨져 치료받던 중 뇌손상으로 인하여 사망하였다. 사망한 피해자는 당시 도로와 갓길의 경계 부근에 서서 낚시 준비를 하고 있었던 것으로 보인다. A는 '피해자가 당시 도로와 갓길의 경계 부근에 쪼그려 앉아 낚싯바늘에 무언가를 끼우고 있었던 것 같았으며, 도로 쪽으로 나와 있어서 정상적으로 주행하였을 경우 충격할 수 있을 것 같은 느낌을 받아 좌측으로 피해 진행하였다'고 진술하였다.	[당신의 부친]이 21:00경 혈중알코올농도 0.07%(면허정지 수준) 정도로 술에 취한 상태로 트럭으로 주행하던 중 갓길에서 낚시를 준비하던 피해자를 들이받은 뒤 구호조치를 하지 않은 채 현장을 벗어난 혐의로 기소되었다. 피해자는 병원으로 옮겨져 치료받던 중 뇌손상으로 인하여 사망하였다. 사망한 피해자는 당시 도로와 갓길의 경계 부근에 서서 낚시 준비를 하고 있었던 것으로 보인다. 당신의 부친은 '피해자가 당시 도로와 갓길의 경계 부근에 쪼그려 앉아 낚싯바늘에 무언가를 끼우고 있었던 것 같았으며, 도로 쪽으로 나와 있어서 정상적으로 주행하였을 경우 충격할 수 있을 것 같은 느낌을 받아 좌측으로 피해 진행하였다'고 진술하였다.

인지하지 못했고, 피해자가 다치지 않았다'고 제시하였으나, 심각성이 높은 조건에서는 '가해자가 음주음전을 하였고, 사고를 인지한 후 도주했으며(뺑소니), 피해자가 뇌손상으로 사망하였다'고 제시하였다. 성범죄와 저작권 침해 사건에 대해서도 비슷한 조작이 이루어졌다. 성범죄와 저작권 침해 사건에 대한 조건별 스토리는 부록에서 확인할 수 있다. 참가자들의 과제는 한 사건

을 읽은 후, '나는 이 사건을 인공지능이 판결하는 것에 대해 수용할 수 있다'라는 진술에 얼마나 동의하는지 7점 척도(1: 전혀 그렇지 않다, 7: 매우 그렇다)로 평정하는 것이었고, 이와 같은 수행을 세 번(교통사고, 성범죄, 저작권 침해) 반복하였다.

인공지능 판사 수용도에 응답한 참가자들은 실험 자극의 심각성 수준이 정상적으로 조작되었는지 확인하기 위한 설문에 응답하였다(심각성 조작점검). 참가자들은 자신이 앞서 확인한 스토리를 다시 한 번 확인하였고, 한 스토리를 확인할 때마다 '이 사건은 얼마나 심각한가요?'(심각성 지각)라는 질문에 대해 7점 척도(1: 전혀 심각하지 않다, 7: 매우 심각하다)로 응답하였다. 한 참가자가 모든 문항에 응답하기까지 걸린 시간은 약 10분이었다.

결과 및 논의

조작점검

본격적인 분석에 앞서 사건의 심각성 조작이 정상적으로 이루어졌는지를 확인해보았다. 분석은 심각성 수준(낮음 vs. 높음)에 따라 사건의 심각성 지각이 달라지는지에 대한 독립표본 t-검증을 통해 이루어졌다. 결과적으로 심각성이 높은 조건(M = 5.84, SD = .73)에서 지각된 사건의 심각성이 낮은 조건(M = 4.53, SD = 1.04)에서 지각된 사건의 심각성보다 더 강했다($t(338)$ = 13.620, p < .001).

표 2는 실험 1의 심각성 조작점검 결과를 요약한 것이다. 한 가지 눈여겨 볼 점은 성범죄에 대해서는 전반적으로 심각성을 높게 인식하고(M = 5.91, SD = 1.36), 저작권 침해에 대해서는 전반적으로 심각성을 낮게 인식한다는 것이다(M = 4.61, SD = 1.34). 그럼에도 불구하고, 모든 유형에서 심각성이 낮은 스토리가 높은 스토리보다 더 심각한 것을 인식되는 것을 확인할 수

<표 2> 실험 1의 사건 심각성 조작점검 결과 요약

	$M(SD)$	낮음		높음		p
		M	SD	M	SD	
교통사고	5.14(1.56)[a, b]	4.28	1.51	5.91	1.16	< .001
성범죄	5.91(1.36)[a]	4.89	1.29	6.82	.50	< .001
저작권 침해	4.61(1.34)[b]	4.40	1.46	4.79	1.19	.007
전체		4.53	1.04	5.84	.73	< .001

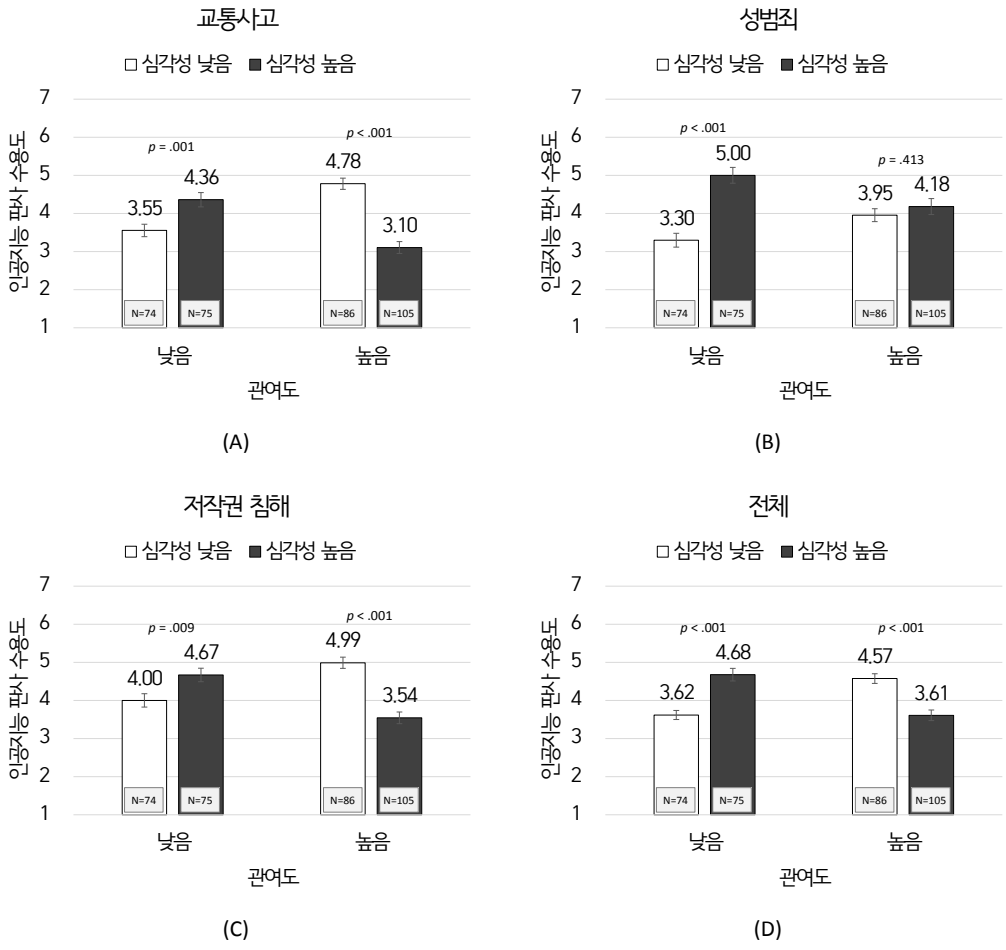
[a]: 교통사고와 성범죄의 평균 차이에 대한 *p-value*는 .001보다 작음

[b]: 교통사고와 저작권 침해의 평균 차이에 대한 *p-value*는 .001보다 작음

있다. 이는 사건의 심각성 조작이 적절했음을 시사한다.

인공지능 판사 수용도

사건 관여도 (낮음 vs. 높음) × 심각성 (낮음 vs. 높음)이 인공지능 판사 수용도에 미치는 효과를 확인하기 위해 이원변량분석(two-way ANOVA)를 수행하였다. 결과적으로, 관여도($F(1, 336) = .151$, $p = .698$)와 심각성($F(1, 336) = .112$, $p = .738$)의 주효과는 나타나지 않았고, 관여도와 심각성의 상호작용이 인공지능 판사 수용도에 미치는 효과를 확인할 수 있었다($F(1, 336) = 51.467$,



(그림 1) 관여도와 심각성의 이원상호작용이 인공지능 판사 수용도에 미치는 효과. (A)는 교통사고, (B)는 성범죄, (C)는 저작권 침해 사건 각각의 결과를 보여주며, (D)는 세 가지 사건 전체의 평균을 종합한 결과이다. 오차막대는 평균의 표준오차를 의미한다.

$p < .001$).

그림 1은 사건 관여도와 심각성이 인공지능 판사 수용도에 미치는 이원상호작용 효과를 보여준다. 그림 1에서 확인할 수 있듯 관여도가 낮을 때는 심각성이 높은 사건($M = 4.68$, $SD = 1.43$)이 낮은 사건($M = 3.62$, $SD = 1.02$)보다 인공지능 판사 수용도가 강하지만($t(147) = -5.193$, $p < .001$), 관여도가 높을 때는 심각성이 낮은 사건($M = 4.57$, $SD = 1.17$)이 높은 사건($M = 3.61$, $SD = 1.43$)보다 인공지능 판사 수용도가 강하다($t(189) = 5.025$, $p < .001$).

이러한 결과는 사람들이 자신과 관여도가 낮은 사건에 대해서는 범죄가 심각할수록 인공지능이 판결내리는 것을 용납할 수 있다고 인식하지만, 자신과 관여도가 높은 사건에 대해서는 범죄가 심각할수록 인공지능의 판결을 수용할 수 없다고 인식함을 보여준다. 다른 말로 하면, 사람들은 제3자의 입장에서 사건을 볼 때는 사건의 심각성이 높을수록 인공지능의 판결을 수용하는 태도를 보이지만, 가해자(부친 혹은 모친)의 가족이 되면 사건의 심각성이 높을수록 인공지능 보다 인간 판사를 선호하는 태도를 보인다. 다만 성범죄 사건의 경우 관여도가 낮을 때는 다른 사건들과 비슷한 경향성을 보였으나, 관여도가 높을 때는 사건의 심각성에 따른 인공지능 판결 수용도에 차이가 없었다. 이러한 현상에 대해서는 종합논의에서 별도로 설명하겠다.

실험 2: 인공지능 배심원 수용도

실험 2는 두 가지 목적에서 이루어졌다. 하나는 실험 1에 대한 반복검증을 통해 연구 결과의 신뢰도를 향상시키는 것이다. 다른 하나는 실험 1의 결과가 유사하지만 동일하지 않은 영역(배심원)에 확장될 수 있는지 확인하기 위해서다. 이러한 목적으로 이루어진 실험 2는 죄에 대한 논의를 진행하고 판결할 수 있는 또 다른 주체인 배심원단의 일부를 인공지능 로봇으로 구성할 수 있는 상황을 제시하였다. 구체적으로, 참가자들은 사람 배심원과 인공지능 로봇 배심원의 비율을 어느 정도로 나눌 것인지 결정하는 과제(12명의 배심원 중 인공지능을 몇 명으로 구성할 것인지)에 참여하였다.

이를 제외한 다른 요소들과 실험 조건은 대부분 실험 1과 같았다. 만약 실험 2에서 실험 1과 비슷한 결과가 나타난다면, 사건 관여도가 낮을 때는 사건의 심각성이 낮을 때보다 높을 때 인공지능 배심원 수가 증가할 것이지만, 사건에 대한 관여도가 높을 때는 사건의 심각성이 높을 때보다 낮을 때 인공지능 배심원 수가 증가하는 경향이 나타날 것이다.

방 법

설계 및 참가자

실험 2는 사건 관여도 2 (낮음 vs. 높음) × 사건 심각성 2 (낮음 vs. 높음)이 해당 사건을 맡아 판결하는 인공지능 배심원 수 결정에 미치는 효과를 확인하기 위한 것으로 참가자간 요인설계(between-subjects design)를 채택하여 이루어졌다. 연구를 위해 19-23세(M = 19.40, SD = .77) 한국 소재 대학교 학부생 365명(남 115, 여 250)이 참여하였다. 참가자들은 모두 한국인이었고, 모국어는 한국어였다. 실험은 수업의 일환으로 이루어졌고, 참가자들은 네 가지 참가자간 조건 중 하나에 무작위로 배정되었다.

재료 및 절차

실험 2의 참가자들은 사건에 대한 내용을 읽고, 해당 사건에 대한 판결을 맡은 12명의 배심원단을 구성할 때, 인공지능 배심원을 몇 명 포함시킬 것인지 결정하는 과제를 수행하였다. 구체적으로 참가자들은 한 사건에 대한 스토리를 읽을 때마다 [*이 사건은 배심원들에 의해 판단이 이루어질 예정입니다. 여기 서로 다른 회사에서 개발한 인공지능 로봇 배심원 12명과 인간 배심원 12명 있습니다. 당신에게 인공지능 로봇 배심원 수를 결정할 수 있는 권한이 있다면, 몇 명을 참여시키겠습니까? 당신이 참여시킬 인공지능 로봇 배심원 수를 쓰세요. '0'에서 '12'사이의 숫자를 쓰시면 됩니다.*]라는 질문을 받았다. 이를 제외한 실험 2의 재료와 절차는 실험 1과 동일하였다(표 1). 한 참가자가 모든 문항에 응답하기까지 걸린 시간은 약 10분이었다.

결과 및 논의

조작점검

본격적인 분석에 앞서 사건의 심각성 조작이 정상적으로 이루어졌는지를 확인하였다. 분석은 심각성 조건에 따라 사건의 심각성 지각이 달라지는지 독립표본 t-검증을 통해 이루어졌다. 결과적으로 심각성이 높은 조건(M = 6.18, SD = .45)에서 심각성이 낮은 조건(M = 4.37, SD = 1.28)보다 사건을 더 심각하게 인식하는 것으로 나타났다($t(363)$ = 16.826, p < .001). 표 3은 실험 2의 심각성 조작점검 결과를 요약한 것이다. 흥미로운 점은 실험 1의 조작 점검 결과와 유사하게 성범죄에 대한 심각성은 전반적으로 높게 인식하고(M = 5.59, SD = 1.66), 저작권 침해의 심각성

<표 3> 실험 2의 사건 심각성 조작점검 결과 요약

	M(SD)	낮음		높음		*p*
		M	*SD*	*M*	*SD*	
교통사고	5.30(1.62)[a, b]	4.40	1.52	6.52	.72	< .001
성범죄	5.59(1.66)[a]	4.62	1.56	6.90	.39	< .001
저작권 침해	4.53(1.42)[b]	4.10	1.51	5.13	1.01	< .001
전체		4.37	1.28	6.18	.45	< .001

[a]: 교통사고와 성범죄의 평균 차이에 대한 *p-value*는 .001보다 작음

[b]: 교통사고와 저작권 침해의 평균 차이에 대한 *p-value*는 .001보다 작음

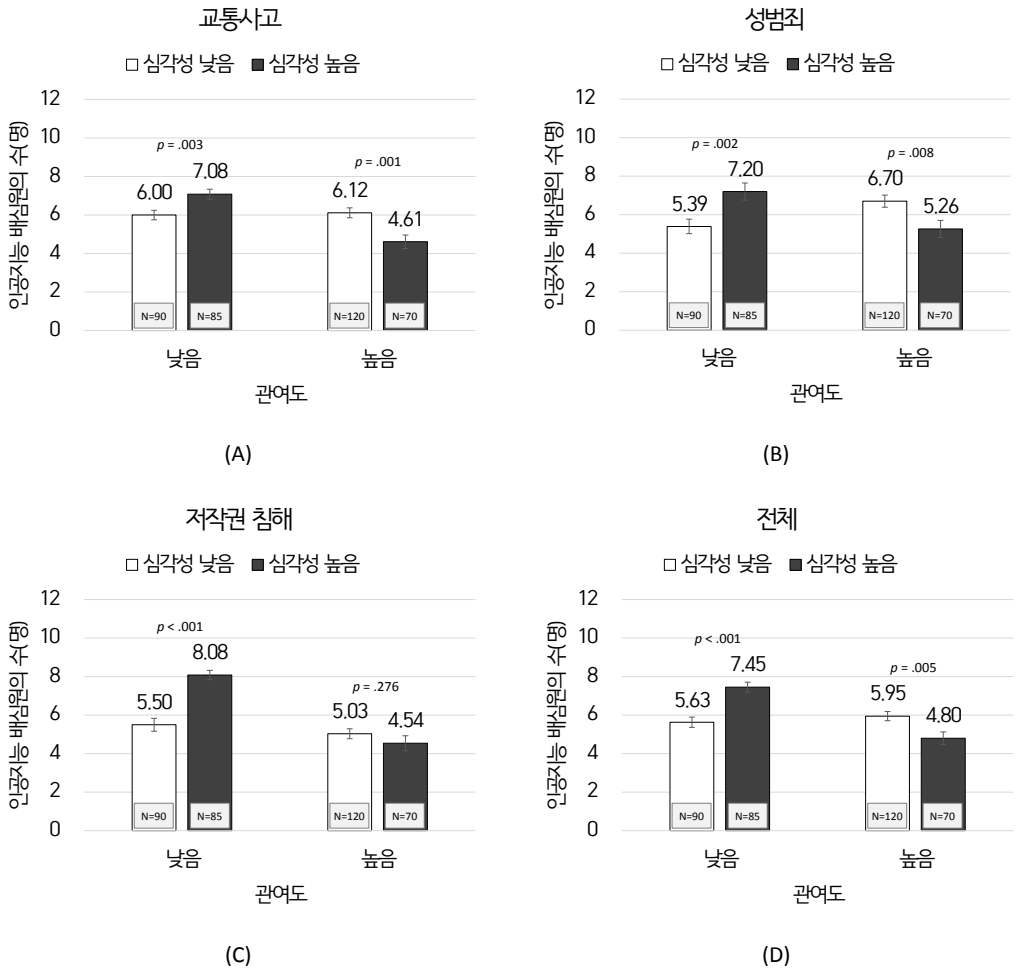
은 전반적으로 낮게 인식한다는 것이다(M = 4.53, SD = 1.42). 그럼에도 불구하고, 모든 유형에서 심각성이 낮은 스토리가 높은 스토리보다 심각성 지각이 강함을 확인할 수 있다. 이는 사건의 심각성 조작이 적절했음을 의미한다.

인공지능 배심원 수

사건 관여도 (낮음 vs. 높음) × 심각성 (낮음 vs. 높음)이 인공지능 배심원 수 결정에 미치는 효과를 확인하기 위해 이원변량분석(two-way ANOVA)를 수행하였다. 결과적으로, 관여도의 주효과가 확인되었다(F(1, 363) = 18.087, p < .001). 즉 관여도가 높은 사건(M = 5.53, SD = 2.71)보다 낮은 사건(M = 6.52, SD = 2.62)에 배정하는 인공지능 배심원의 수가 많았다. 또한 관여도와 심각성의 상호작용이 인공지능 배심원 수에 미치는 효과를 확인할 수 있었다(F(1, 361) = 29.404, p < .001). 심각성의 주효과는 나타나지 않았다(F(1, 363) = 1.541, p = .215).

그림 2는 사건 관여도와 심각성이 인공지능 배심원 수에 미치는 이원상호작용 효과를 보여준다. 그림 2에서 확인할 수 있듯 관여도가 낮을 때는 심각성이 높은 사건(M = 7.45, SD = 2.40)에 배정한 인공지능 배심원의 수가 낮은 사건(M = 5.63, SD = 2.52)보다 많았지만(t(173) = 4.903, p < .001), 관여도가 높을 때는 심각성이 낮은 사건(M = 5.95, SD = 2.60)에 배정한 인공지능 배심원의 수가 높은 사건(M = 4.80, SD = 2.77)보다 많았다(t(188) = 2.860, p = .005).

이러한 결과는 사람들이 자기 자신과 직접적인 관련이 없는 사건에 대해서는 사건이 심각할수록 인공지능 배심원이 판결하기를 원하지만, 자기 자신과 직접 관련이 있을 때는 사건이 심각할수록 인간 배심원이 판결하기를 원한다고 인식함을 보여준다. 다른 말로 하면, 사람들이 제3자의 입장에서 사건을 볼 때는 사건의 심각성이 높을수록 인공지능 배심원을 선호하지만, 가해자(부친 혹은 모친)의 가족이 되면, 사건의 심각성이 높을수록 인간 배심원을 선호한다. 다만 저



(그림 2) 관여도와 심각성의 이원상호작용이 인공지능 배심원 수 결정에 미치는 효과. (A)는 교통사고, (B)는 성범죄, (C)는 저작권 침해 사건 각각의 결과를 보여주며, (D)는 세 가지 사건 전체의 평균을 종합한 결과이다. 오차막대는 평균의 표준오차를 의미한다.

작권 침해 사건의 경우 관여도가 낮을 때는 다른 사건들과 비슷한 경향성을 보였으나, 관여도가 높을 때(사건 당사자일 때)는 사건이 얼마나 심각한지에 관계없이, 전반적으로 인공지능의 배심원의 수용도가 낮아지는 경향성을 보였다. 이러한 현상에 대해서는 종합논의에서 별도로 설명하겠다.

종합 논의

결과 요약 및 해설

본 연구는 사건 관여도(낮음 vs. 높음)와 심각성(낮음 vs. 높음)이 인공지능 판사(실험 1) 및 배심원(실험 2) 수용도에 미치는 효과를 확인하기 위해 이루어졌다. 관여도는 자신의 부친 혹은 모친이 가해자인 사건 대 모르는 사람이 가해자인 사건으로 조작하였고, 심각성은 피해 규모와 적용되는 죄의 수의 상대적 차이로 조작하였다.

실험 1의 참가자들은 조건별로 세 가지 사건(교통사고 vs. 성범죄 vs. 저작권 침해)을 살펴보았고, 각 사건에 대해 인공지능 판사가 판결하는 것을 얼마나 수용하는지에 대한 질문에 응답하였다. 실험 2의 참가자들도 실험 1과 동일한 방식으로 세 가지 사건을 확인하였고, 12명의 배심원단을 구성할 때, 인공지능 로봇 배심원을 몇 명 참여하게 할 것인지 결정하였다. 결과적으로 관여도가 낮을 때는 사건의 심각성이 높은 사건이 낮은 사건보다 인공지능 판사에 대한 수용도가 높았지만, 관여도가 높을 때는 심각성이 낮은 사건이 높은 사건보다 인공지능 판사에 대한 수용도가 높아지는 현상을 확인할 수 있었다(실험 1). 또한 관여도가 낮을 때는 사건의 심각성이 높은 사건이 낮은 사건보다 인공지능 배심원 수를 많이 배정하지만, 관여도가 높을 때는 심각성이 낮은 사건이 높은 사건보다 배정하는 인공지능 배심원 수가 많아지는 현상을 관찰하였다(실험 2). 즉 관여도와 심각도가 인공지능 판사와 배심원 수용도에 미치는 이원상호작용 효과를 확인할 수 있었다. 설명하자면, 사람들은 내 가족과 관련된 일처럼 관여도가 높은 사건에 대해서는 사건이 심각할수록 인공지능의 판결보다 인간 판사가 판결 내는 것을 선호하는 태도를 보이지만, 나와 관련 없는 타인의 일의 경우에는 사건이 심각할수록 인공지능의 판결을 선호하는 태도를 보인다.

이러한 현상은 이론적 배경에서 언급한 '고유성 무시'와 연결 지어 설명이 가능하다. 사람들은 나와 관련 있는 사람의 범죄를 관련 없는 사람의 범죄에 비해 더 축소하여 평가하며, 범죄자를 더 옹호하려고 한다. 특히 사건이 심각할수록 이들에게는 자신의 가족이 적은 형량을 받는 것이 중요해지므로, 개인의 고유성을 더 고려하여 정상참작해줄 수 있는 인간 판사를 인공지능 판사보다 선호하게 되는 것이다. 그러나 나와 관련 없는 제 3자의 사건을 바라보는 경우에는 반대로 사건의 심각성이 낮을 때보다 높을 때 인공지능 판사를 선호하게 된다. 판결 받는 이가 내 집단의 구성원도 아닐뿐더러, 사건이 심각할수록 사회에 미치는 부정적인 영향력이 크기 때문에 법에 적힌 그대로 심판받기를 원하게 되기 때문이다.

본 연구의 결과를 '고유성 무시'에 근거한 해석 외에, 연구 2의 '관여도'와 연결 지어 해석해 볼 수도 있다. 첫째, 사건에 대한 관여도에 따라 인공지능에 대한 신뢰도 지각이 달라진 것이 본 연구에서 관찰된 현상의 근본 기제로 작용했을 수 있다(Dhanesh & Nekmat, 2019; Lambert et

al., 2020; Schuitema, Aravena, & Denny, 2020). 구체적으로 관여도가 낮을 때는 인공지능에 대한 신뢰도가 증가하여 사건이 심각할수록 인공지능의 판결을 지지하게 되지만, 관여도가 높을 때는 인공지능에 대한 신뢰도가 감소하여 사건이 심각할수록 인공지능에 대한 판결을 지지하지 않게 되었을 가능성이 있다. 둘째, 반대로 관여도에 따라 사람에 대한 신뢰도 지각에 차이가 발생했을 수 있다. 관여도가 낮을 때는 인간 판사에 대한 신뢰도가 상대적으로 낮아지고, 결과적으로 인공지능의 판단을 지지하게 되지만, 관여도가 높을 때는 인간 판사에 대한 신뢰도가 상대적으로 높아지고, 이에 따라 인공지능 판사 혹은 배심원보다 인간 판사나 배심원을 지지하게 되었을 수 있다. 셋째, 관여도가 높을 때는 인공지능 판결에 대한 불확실성 혹은 예측 불가능성 지각이 인간에 대한 불확실성 혹은 예측 불가능성 지각보다 높지만, 관여도가 낮을 때는 인공지능보다 인간의 판단을 더 불확실하고 예측 불가능하다고 지각하는 것이 본 연구에서 관찰한 현상에 기여했을 가능성도 있다(Lee & Kim, 2016; Teng & Lu, 2016; Zhao, Feng, & Shi, 2018). 다른 말로 하면 내 가족이 가해자인 사건인 경우에는 인공지능이 어떻게 판결을 내릴지에 대한 불확실성 지각이 인간 판사가 어떻게 판결할지에 대한 불확실성 지각보다 높기에 인공지능이 판결하는 것에 대한 수용도가 낮아지지만, 나와 관련 없는 사람이 가해자인 경우에는 인간 판사가 어떻게 판결할지에 대한 불확실성 지각을 인공지능 판결보다 높게 지각함으로써 인공지능 판결에 대한 수용도가 높아졌을 수 있다. 넷째, 관여도가 높을 때는 인공지능에 대한 부정적 정보에 더 주목하거나, 떠올리게 될 수 있고, 관여도가 낮을 때는 인공지능에 대한 긍정적 정보에 더 주목하거나, 떠올렸을 가능성도 있어 보인다(Balabanis & Chatzopoulou, 2019). 설명하자면, 관여도가 높을 때는 인공지능에 대한 부정 편향이 발휘되면서 인간 판사나 배심원에 대한 수용도가 강해지지만, 관여도가 낮을 때는 인공지능에 대한 긍정 편향이 발휘되면서 인공지능 판사나 배심원에 대한 수용도가 낮아졌을 수 있다.

실험 1에서 관여도가 높은 성범죄 사건의 경우 사건의 심각성 조건에 따른 인공지능 판결 수용도에 차이가 없었다. 이에 대해서는 먼저 성범죄 사건이 다른 두 가지 유형의 사건에 비해 사건의 심각성 지각이 강했다는 것을 고려할 필요가 있다. 즉 본 연구에서 심각성이 높은 성범죄와 낮은 성범죄를 구분하여 제시하였고, 조작점검 상으로 상대적인 차이가 존재하는 것은 사실이지만, 교통사고나 저작권 침해에 비교할 때 성범죄라는 자체가 주는 심각성이 높았다. 더하여 관여도가 낮을 때는 심각성의 상대적인 차이에 대한 고려가 이루어졌지만, 관여도가 높을 때는 심각성의 상대적인 차이에 대한 고려가 이루어지지 않은 것으로 보인다. 정리하면, 성범죄 자체에 대한 높은 심각성 지각 그리고 관여도가 낮을 때는 심각성의 미묘한 차이를 인식하지만, 관여도가 높을 때는 이 차이를 인식하지 못하게 되는 것이 동시에 작용하였고, 결과적으로 성범죄 관여도가 높은 조건에서는 심각성의 효과가 나타나지 않은 것으로 보인다.

실험 2에서 관여도가 높은 저작권 침해 사건에서 심각성에 따른 차이 없이 전반적으로 인공지능 수용도가 낮아진 것에 대해서도 설명이 필요하다. 먼저 저작권 침해 사건의 경우 관여도가

낮은 조건에서는 인공지능의 표절 분석, 유사도 분석에 대한 신뢰도가 높지만, 관여도가 높은 조건에서는 동일한 것에 대한 신뢰도가 낮아졌을 수 있다. 또한 예술 분야 저작권 침해의 경우, 관여도에 따라 인공지능이 판단하기 어렵고 오직 인간만 판단할 수 있는 예술적 고유성이 있다는 것에 대한 인식의 차이가 작용했을 수 있다. 즉 관여도가 낮을 때는 인공지능도 인간만큼 예술적 요소에 대한 판단을 할 수 있을 거라고 지각하지만, 관여도가 높을 때는 인공지능은 인간의 감각을 따라올 수 없기에 예술적 고유성을 무시하게 될 수 있다고 생각했을 가능성이 있다.

시사점

본 연구는 우선 인공지능의 법적 판단에 대한 일반 사람들의 수용 및 인식을 실험적으로 확인한 국내 최초의 연구라는 점에서 이론적인 기여를 한다. 그동안 국내에서 법률적 맥락의 인공지능에 대한 논의가 없었던 것은 아니다. 그러나 대부분의 연구가 법률 관계자의 시각에서 바라본 인공지능 판사의 구현 가능성에 대한 이론적 고찰에 불과하거나(양종모, 2016; 2018), 인공지능 판사의 실제 적용을 목표로 한 연구가 아니었다(오요한, 홍성욱, 2018; 정영화, 2020). 따라서 본 연구는 국내 최초로 법률 관계자가 아닌 일반 사람들의 시각에서 바라본 인공지능 판사에 대해 실험적으로 연구했다는 점에서 의의를 갖는다. 둘째, 본 연구는 의료 및 소비자 분야에서 주로 사용되던 ‘고유성 무시’를 법률 분야에 확장했다는 점에서 이론적으로 기여한다. 고유성 무시는 개인의 기질, 고유성을 바탕으로 하는 개념이기 때문에, 대부분 의료 인공지능이나 소비자의 특성을 반영한 인공지능 추천 시스템에 대한 연구에 적용되어왔다. 본 연구는 이러한 고유성 무시를, 인공지능의 법적 판단에 대한 심리적 저항의 원인으로 새롭게 추가함으로써, 해당 개념의 적용 분야를 넓혔다. 셋째, 인공지능의 법적 판단에 대한 논의를 사건의 심각성과 관여도 측면으로 확장하였다는 점에서 이론적으로 기여한다. 인공지능의 법적 혹은 도덕적 판단에 대한 기존 연구들은 대부분 사건의 심각성과 관여도를 고려하지 않거나, 고려하더라도 함께 다루지 않았다(Bigman & Gray, 2018; Gill, 2020; Hengstler, Enkel, & Duelli, 2016). 그러나 본 연구에서는 인공지능의 판결에 대한 반응을 상이하게 만드는 두 요인을 함께 고려함으로써, 사람들의 인공지능 판결에 대한 수용도를 다양한 상황에서 예측 가능하게 하였다.

본 연구는 실용적인 측면에서도 몇 가지 기여를 한다. 먼저, 교통사고, 성범죄, 저작권 침해라는 다양한 유형의 사건에서 인공지능의 판단이 불러올 수 있는 사회적 갈등을 예측해볼 수 있는 자료를 제공한다. 심각성이 높은 사건의 경우 대중들은 인공지능이 판단해야 한다고 생각하지만, 이 사건의 가해자 혹은 가해자와 직접적인 친분이 있는 사람들은 사람이 판단해야 한다고 주장하면서 갈등이 발생할 가능성이 있다. 또한 법률 판단을 전문으로 하는 인공지능의 등장이 판사, 검사, 변호사와 같은 법률 전문가의 영역에 어떤 영향을 미칠 것인지 가늠해 볼 수 있게 한다. 본 연구에 의하면, 사람들은 심각한 사건일 때, 특히 자신과 가까운 사람의 판결에서는 인

공지능 판사가 결정내리는 것을 원하지 않는다. 따라서, 가까운 미래에 가벼운 사건에 대한 법률적 판단은 인공지능에게 위임하고, 대중적으로 중요하고, 심각한 사건은 인간 판사가 다루는 인간과 기계의 역할 분담이 이루어질 가능성이 있다(Schmid, Miodrag, & Francesco, 2008). 즉 인공지능이 법률적 판단을 할 수 있는 시대가 오더라도 법률적 문제를 전문적으로 다루는 사람의 역할은 여전히 중요할 것이다. 본 연구는 지적재산권과 관련된 사건은 인간 전문가의 역할이 줄어들지 않을 수 있음을 보여주었다는 점에서도 시사점이 있다. 본 연구의 실험 2는 지적재산권의 일종인 저작권 관련 사건이 나와 밀접하게 관련되어 있을 경우, 사건의 심각성에 관계없이 인공지능이 판결하는 것에 대한 수용도가 낮음을 확인하였다. 이는 지적재산권 분쟁의 당사자들은 인공지능보다 지적재산권 전문가에게 사건에 대한 판단을 의뢰할 가능성이 높음을 의미한다. 결론적으로 인공지능에게 법률적 판단을 의뢰할 수 있는 시대가 온다고 해서 지적재산권 전문가들이 일자리를 잃어버리진 않을 것으로 보인다.

한계와 제안

본 연구는 다양한 시사점에도 불구하고, 몇 가지 부분에서는 한계가 있으며 이에 대한 후속 연구가 필요하다. 첫째, 본 연구는 관여도와 심각성이 인공지능 판사 및 배심원 수용도에 미치는 이원상호작용에 어떤 인지적 기제가 작동하는지 확인하지 못했다. 본 연구에서는 선행 연구들을 검토하여, 고유성 무시가 사건의 심각성과 관여도, 인공지능이 판결내리는 것에 대한 수용도에 영향을 미쳤을 것이라고 예상하였다. 그러나 문항이나 실험을 통해 고유성 무시의 효과를 실제적인 측정하지는 못하였다. 향후 이에 대한 별도의 연구가 진행된다면, 인공지능 판사 및 배심원 수용도에 미치는 근본 기제에 대한 이해를 심화시키는 것에 기여할 수 있을 것이다.

둘째, 실험 1과 실험 2의 결과를 비교할 때, 성범죄 사건과 저작권 침해 사건의 관여도 높음 조건의 결과에 차이가 있다. 성범죄 사건의 경우, 실험 1에서는 관여도 높음 조건에서 심각성 수준에 따라 인공지능의 판결에 대한 수용도에 차이가 없지만, 실험 2에서는 배치한 인공지능 배심원 수에 유의미한 차이가 발생한다. 저작권 침해 사건의 경우, 실험 1에서는 관여도 높음 조건에서 심각성 수준에 따라 인공지능의 판결에 대한 수용도에 차이가 있지만, 실험 2에서는 배치한 인공지능 배심원 수에 차이가 없다. 이러한 차이는 실험 1이 인공지능 판사가 대상인 것에 반해 실험 2는 인공지능 배심원을 대상으로 하는 것에 의해 발생했을 수 있다. 그러나 이에 대한 정확한 원인을 밝히지는 못하였으므로, 향후 연구에서는 보다 더 반복적인 검증을 통해 이러한 현상의 원인 및 결과를 확인해볼 수 있을 것이다.

셋째, 본 연구의 참가자의 평균 연령은 약 19세였으며, 이에 따라 본 연구의 결과를 전 연령대로 일반화하기에는 다소 한계가 있다. 또한 10대 후반 혹은 20대 초반의 생각이 30대, 40대까지 이어지는지, 아니면 변화가 있는지에 대한 추적이 이루어질 필요도 있다. 그럼에도 불구하고,

10년 후, 20년 후에 인공지능 판사나 배심원을 직접 마주칠 가능성이 높은 현재 20대가 본 연구와 같은 인식을 가지고 있다는 것을 발견한 것은 중요하다.

넷째, 본 연구는 자기 자신이 가해자(부친 혹은 모친)의 자녀라고 상정해보게 하면서 관여도를 조작하였다. 그러나 가해자 가족의 입장과 가해 당사자 혹은 피의자 본인의 입장은 전혀 다른 문제일 수 있으며, 향후에는 자기 자신이 가해자 혹은 피의자인 상황과 피해자인 상황을 구분하여 살펴볼 필요가 있다. 이러한 후속 연구가 이루어지길 기대한다.

참고문헌

양새롬. (2020.10.16.). 구하라 '불법촬영 무죄' 판결에..."AI 판사 채용해주세요" 청원 등장. **동아일보**. https://www.news1.kr/articles/?4088468

양종모. (2016). 인공지능을 이용한 법률전문가 시스템의 동향 및 구상. **법학연구**. 19(2), 213-242.

양종모. (2018). 인공지능에 의한 판사의 대체 가능성 고찰. **홍익법학**. 19(1), 1-29.

연규욱. (2017.11.07). 랜들 레이더 전 美연방항소법원장 "인공지능이 5년내 판사 대체...사법 불신 줄어들 것". **매일경제**. https://www.mk.co.kr/news/economy/view/2017/11/737834/

오요한, 홍성욱. (2018). 인공지능 알고리즘은 사람을 차별하는가?. **과학기술학연구**. 18(3), 153-215.

정영화. (2020). 인공지능과 법원의 분쟁해결-최근 영미법국가들의 인공지능 법제. **홍익법학**. 21(1), 209-247.

정원엽, 이기준. (2016.05.17). 첫 AI 변호사 '로스', 뉴욕로펌 취직하다. **중앙일보**. https://news.joins.com/article/20035624

한국개발연구원. (2021.01.14). *우라니라 AI 생태계 작동 아직 미흡해: AI에 대한 기업체 인식 및 실태 조사 결과*. https://www.kdi.re.kr/news/coverage_view.jsp?idx=10941&pp=10&pg=1&gubun=03

LG경제연구원. (2018.05.15). *인공지능에 의한 일자리 위험 진단*. http://www.lgeri.com/report/view.do?idx=19620

Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016, May 23). Machine Bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Balabanis, G., & Chatzopoulou, E. (2019). Under the influence of a blogger: The role of information? seeking goals and issue involvement. *Psychology & Marketing, 36*(4), 342-353.

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21-34.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Dhanesh, G. S., & Nekmat, E. (2019). Facts over stories for involved publics: framing effects in CSR

messaging and the roles of issue involvement, message elaboration, affect, and Skepticism. *Management Communication Quarterly, 33*(1), 7-38.

Gill, T. (2020). Blame it on the self-driving car: how autonomous vehicles can alter consumer morality. *Journal of Consumer Research, 47*(2), 272-291.

Granulo, A., Fuchs, C., & Puntoni, S. (2021). Preference for human (vs. robotic) labor is stronger in symbolic consumption contexts. *Journal of Consumer Psychology, 31*(1), 72-80.

Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change, 105*, 105-120.

Lambert, E. G., Keena, L. D., Haynes, S. H., Ricciardelli, R., May, D., & Leone, M. (2020). The issue of trust in shaping the job Involvement, job satisfaction, and organizational commitment of southern correctional staff. *Criminal Justice Policy Review, 32*(2), 193-215.

Leachman, S. A., & Merlino, G. (2017). The final frontier in cancer diagnosis. *Nature, 542*(7639), 36-38.

Lee, E. J., & Kim, Y. W. (2016). Effects of infographics on news elaboration, acquisition, and evaluation: Prior knowledge and issue involvement as moderators. *New Media & Society, 18*(8), 1579-1598.

Lee, J., & Holyoak, K. J. (2020). "But he's my brother": The impact of family obligation on moral judgments and decisions. *Memory & cognition, 48*(1), 158-170.

Martin, J., Young, L., & McAuliffe, K. (2020, August 18). The impact of group membership on punishment versus partner choice. https://doi.org/10.31234/osf.io/5qr32

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Schmid, R. F., Miodrag, N., & Francesco, N. D. (2008). A human-computer partnership: The tutor/child/computer triangle promoting the acquisition of early literacy skills. *Journal of Research on Technology in Education, 41*(1), 63-84.

Schuitema, G., Aravena, C., & Denny, E. (2020). The psychology of energy efficiency labels: Trust, involvement, and attitudes towards energy performance certificates in Ireland. *Energy Research & Social Science, 59*, 101301.

Teng, C. C., & Lu, C. H. (2016). Organic food consumption in Taiwan: Motives, involvement, and purchase intention under the moderating role of uncertainty. *Appetite, 105*, 95-105.

Weidman, A. C., Sowden, W. J., Berg, M. K., & Kross, E. (2020). Punish or protect? How close relationships shape responses to moral violations. *Personality and Social Psychology Bulletin, 46*(5), 693-708.

Yun, J. H., Lee, E. J., & Kim, D. H. (2021). Behavioral and neural evidence on consumer responses to

human doctors and medical artificial intelligence. *Psychology & Marketing, 38*(4), 610-625.

Zhao, Y., Feng, T., & Shi, H. (2018). External involvement and green product innovation: The moderating role of environmental uncertainty. *Business Strategy and the Environment, 27*(8), 1167-1180.

1차 원고 접수: 2021.04.30
1차 심사 완료: 2021.06.06
2차 원고 접수: 2021.09.03
2차 심사 완료: 2021.10.05
3차 원고 접수: 2021.10.12
최종 게재확정: 2021.10.12

(*Abstract*)

The Effect of Involvement and Severity on Acceptance of Artificial Intelligence Judgment

Eun Yeong Doh
Department of Industrial Psychology, KwangWoon University

Guk-Hee Lee
Division of General Studies, Kyonggi University

Ji Eun Jung
Department of Forensic Psychology, Kyonggi University

With the development of artificial intelligence(AI), the jobs of many human experts are threatened, and this also applies to the legal profession. This study attempted to investigate whether AI can actually replace humans in the legal profession, especially the role of judges making final judgments. For this purpose, from the perspective of uniqueness neglect, this study was conducted to confirm the effect of involvement and the severity on acceptance of the judgment made by the AI judge (Experiment 1) and the AI jury (Experiment 2). The involvement was manipulated as if the subject who was sentenced for committing a crime was his or her family (mother, father) or stranger, and the severity was manipulated by the extent of the damage, the perception of the crime, and the number of applied crimes. In Experiment 1, the interactive effect of involvement and severity was found. Specifically, when the involvement was low, the acceptance of AI judges was higher in high severity (vs. low severity). Conversely, when the involvement was high, the acceptance of AI judges was higher in low severity (vs. high severity). The same interactions as in Experiment 1 occurred in Experiment 2. Specifically, when the involvement was low, a larger number of AI jury members were allocated in high severity (vs. low severity). On the other hand, when the involvement was high, the number of AI juries increased in low severity (vs. high severity). This study has implications in that it is the first experimental study in Korea on artificial intelligence legal judgment and that it presents the prospects for the jobs of legal experts.

Key words : Artificial Intelligence, Judge, Involvement, Severity, Acceptance of Judgment

부 록

부록 1: 저작권 침해 사건 실험 자극

구분		관여도	
		낮음	높음
심각성	낮음	원고는 그래픽 디자인 회사이고, 피고 {F}는 디자인업에 종사하는 사람이다. 피고는 중고 노트북을 구입하였는데, 의뢰 받은 디자인 업무를 하면서 그 중고 노트북에 저장되어 있던 아이콘 이미지 중 일부를 사용하였다. 그 아이콘은 원고가 디자인하여 저작권 등록을 마친 것이었다. 원고는 피고 F를 저작권 위반으로 고소하였으나, 검사는 F가 저작권 침해의 인식과 의사가 있었다고 단정하기 어렵고 달리 증거가 없다는 이유로 불기소결정(증거불충분)을 하였다. 이후 원고는 피고 F를 대상으로 민사손해배상청구소송을 하였다.	원고는 그래픽 디자인 회사이고, 디자인업에 종사하는 {당신의 모친}이 민사소송을 당하였다. 모친은 중고 노트북을 구입하였는데, 의뢰 받은 디자인 업무를 하면서 그 중고 노트북에 저장되어 있던 아이콘 이미지 중 일부를 사용하였다. 그 아이콘은 원고가 디자인하여 저작권 등록을 마친 것이었다. 원고는 당신의 모친을 저작권 위반으로 고소하였으나, 검사는 저작권 침해의 인식과 의사가 있었다고 단정하기 어렵고 달리 증거가 없다는 이유로 불기소결정(증거불충분)을 하였다. 이후 원고는 민사손해배상청구소송을 하였다.
	높음	원고는 출판사이고, 피고는 개인이다. 피고 {F}는 인터넷 파일 공유사이트에 원고가 창작한 저작물인 만화 '가' 36권, '나' 11권, '다' 15권, '라' 25권, '바' 20권의 이미지 파일을 업로드하여 불특정 다수의 사람들이 다운로드 받을 수 있게 하였다. 원고는 판매량 급감으로 인해 3억 2,100만원을 손해 봤다고 주장하였다. 피고 F는 '만화를 업로드 해서 공유 사이트의 포인트 외 달리 얻은 수익이 없고, 그 이미지를 나도 다른 인터넷 사이트에서 구했다'고 주장하였다.	원고는 출판사이고, {당신의 모친}이 원고의 저작물을 인터넷에 올려 공유하여 민사소송을 당하였다. 인터넷 파일 공유사이트에 원고가 창작한 저작물인 만화 '가' 36권, '나' 11권, '다' 15권, '라' 25권, '바' 20권의 이미지 파일을 업로드하여 불특정 다수의 사람들이 다운로드 받을 수 있게 하였다. 원고는 판매량 급감으로 인해 3억 2,100만원을 손해 봤다고 주장하였다. 당신의 모친은 '만화를 업로드 해서 공유 사이트의 포인트 외 달리 얻은 수익이 없고, 그 이미지를 나도 다른 인터넷 사이트에서 구했다'고 주장하였다.

부록 2: 성범죄 사건 실험 자극

구분		관여도	
		낮음	높음
심각성	낮음	[남성 D]는 한 음식점에서 모임을 마친 뒤 일행을 배웅하던 중 옆을 지나치던 여성 C의 엉덩이를 만진 혐의로 재판에 넘겨졌다. CCTV 분석 결과 남성 D가 여성 C를 스쳐지나가는 시간이 1.33초였고, 추행 장면을 정확히 찍혀있지 않았다. 목격자 증언도 엇갈렸다.	[당신의 부친]이 한 음식점에서 모임을 마친 뒤 일행을 배웅하던 중 옆을 지나치던 여성 C의 엉덩이를 만진 혐의로 재판에 넘겨졌다. CCTV 분석 결과, 당신의 부친이 여성 C를 스쳐지나가는 시간이 1.33초였고, 추행 장면을 정확히 찍혀있지 않았다. 목격자 증언도 엇갈렸다.
	높음	[남성 D]는 미성년자 음란물 동영상 20만 건을 유통한 혐의로 재판에 넘겨졌다. D는 높은 수익을 얻을 수 있다는 이유로 아동과 청소년이 등장하는 동영상 약 25만 개를 유통하였고, 약 4억원을 벌었다. 피해자인 4~5세 아이들은 성폭행 등 학대를 당하였다. D는 가난과 무지 때문에 범죄를 저지르게 되었다고 주장하였다.	[당신의 부친]이 미성년자 음란물 동영상 20만 건을 유통한 혐의로 재판에 넘겨졌다. 부친은 높은 수익을 얻을 수 있다는 이유로 아동과 청소년이 등장하는 동영상 약 25만 개를 유통하였고, 약 4억원을 벌었다. 피해자인 4~5세 아이들은 성폭행 등 학대를 당하였다. 당신의 부친은 가난과 무지 때문에 범죄를 저지르게 되었다고 주장하였다.