

**법적 정의와 사회 정의의 차이 연구:**  
**챗GPT를 이용한 유튜브 댓글 분석을 중심으로**  
**Difference between Legal Justice and Social Justice:**  
**An Analysis of YouTube Comments Using ChatGPT**

민지환\*, 김태익\*, 이세현\*, 최화준\*\*, 김상준\*\*\*  
모두의연구소\*, 아주대학교 경영대학원\*\*, 이화여자대학교 경영대학\*\*\*

Ji-Hwan Min(rpyllover@gmail.com)\*, Taeik Kim(taeikkim92@gmail.com)\*,  
Se-Hyun Lee(leesehyun01@gmail.com)\*, Hwa-joon Choi(joony279@ajou.ac.kr)\*\*,  
Sang-Joon Kim(s.kim@ewha.ac.kr)\*\*\*

# 법적 정의와 사회 정의의 차이 연구: 챗GPT를 이용한 유튜브 댓글 분석을 중심으로

## Difference between Legal Justice and Social Justice: An Analysis of YouTube Comments Using ChatGPT

민지환\*, 김태익\*, 이세현\*, 최화준\*\*, 김상준\*\*\*

모두의연구소\*, 아주대학교 경영대학원\*\*, 이화여자대학교 경영대학\*\*\*

Ji-Hwan Min(rpyliver@gmail.com)\*, Taeik Kim(taeikkim92@gmail.com)\*,  
Se-Hyun Lee(leesehyun01@gmail.com)\*, Hwa-joon Choi(joony279@ajou.ac.kr)\*\*,  
Sang-Joon Kim(s.kim@ewha.ac.kr)\*\*\*

### 요약

본 연구는 사법기관 판결을 법적 정의, 대중 의견의 집합을 사회 정의라 전제하고, 오늘날의 법적 정의와 사회 정의의 차이를 알아보고자 했다. 이를 위해 교통사고 전문 변호사가 운영하는 영상 공유 소셜 미디어 채널을 연구 환경으로 선택했는데, 채널의 일부 영상에는 사법부의 판결문과 시청자들의 댓글이 함께 있어 본 연구에 적합하다 판단했다.

연구는 영상 속 판결문이 법적 정의, 댓글의 집합이 사회 정의를 대변한다고 가정하고 기계 학습의 일종인 챗 GPT를 활용하여 댓글들을 긍정, 부정, 중립의 감정으로 분류하였다. 그리고 긍정 감정의 댓글을 판결문에 대한 동의, 부정을 비동의로 해석하였다. 또한 차량 사고 판결의 특수성을 고려해, 차량 사고 과실 비율과 판결 심급에 따라 감정을 세부 분류하고 함의를 찾았다.

연구 결과는 사회 정의와 법적 정의의 차이가 존재할 수 있음을 보여주었다. 그리고 사고 과실 비율이나 판결 심급 같은 부가적인 요인들에 따라 차이 정도가 다름도 발견되었다.

■ 중심어 : | 사회 정의 | 법적 정의 | 텍스트마이닝 | 감정 분석 |

### Abstract

This study assumes that judicial decisions, embodying legal justice and are not always matched to what the society actually believes as social justice. To find out how distant legal justice is from social justice, a video sharing social media channel managed by a traffic accident lawyer is selected as the research environment. Since the videos on the channel include judicial rulings and viewers' comments follow, the media platform provide suitable data for our study.

Given that the judicial ruling in the video is a proxy for legal justice and the collection of comments for social justice, we classified the comments into positive, negative, and neutral emotions using a machine learning ChatGPT. The comments labeled as positive emotions were interpreted as agreement with the ruling, and the comments labeled as negative were captured as disagreement. In addition, considering the features of vehicle accident judgments, we classified emotions in more details according to vehicle accident fault ratios and judgment levels. By doing so, this study provides theoretical and practical implications.

Our findings suggest that differences between social and legal justice may exist. Additionally, the degree of difference depends on factors such as accident negligence rates or judgmental outcomes.

■ keyword : | Social Justice | Legal Justice | Text-mining | Sentiment Analysis |

## I. 서론

법은 변화하는 현실에 기반하여 사회의 가치를 반영한다[1]. Dickson(2000)은 우리가 공유하는 전통적인 법은 보다 넓은 시각에서 살아있는 것을 포함해야 한다고 말했다[2]. 이러한 의견은 법이 우리의 일상 생활에 깊이 간섭하는 것처럼 보이지만 실제로는 대중의 일상 생활을 완벽하게 반영하지는 않는다는 것을 함축한다. 법적 정의(legal justice)는 사법 기관의 영역에서 크게 다뤄져 왔다. 해당 기관에 속하지 않은 조직과 사람들은 법적 정의에 관심을 기울이지 않는 경향이 있었다[3]. 명문화된 법적 정의와 주체들은 대중과 가까이 있지 않았었다.

하지만 오늘날 소셜 미디어의 등장과 함께 정의에 대한 대중의 관심은 높아졌다. 소셜 미디어의 대중화와 함께 법적 정의는 사회 구성원의 일상 속으로 가까이 들어왔다. 여러 소셜 미디어는 사회적 문제를 대중들이 접할 수 있는 온라인 콘텐츠로 만들어 유통시켰고, 이를 통해 사회운동과 입법을 이루어 내기도 했다[4-6]. 그리고 여러 사회적 문제와 이슈에 대한 청원 제도, 법안 처리, 법적 제도에 대한 사회적 평가가 함께 늘어났다. 이는 온라인 환경 속에서 대중은 법과 정의에 대한 의견을 능동적으로 표현하고, 소셜 미디어는 오늘날의 사회 정의(social justice)가 표출되는 의미있는 공간인 것을 의미한다.

소셜 미디어 속에서 형성되는 법과 정의에 대한 의견들이 사법 제도에서 정립한 정의와 항상 일치하는 것은 아니다. 제도화된 법적 정의와 대중들이 인식하는 사회 정의 사이에는 갈등이 생길 수 있다[7]. 이는 법적 제도가 항상 절대적이거나 유일한 사회 정의의 기준이 될 수 없다는 것을 보여주는데, 소셜 미디어 환경은 이를 보여주는 적절한 예시이다.

소셜 미디어 사용자들은 그들의 시각에서 정의에 대한 개념을 제시하고 재구성할 수 있다[8][9]. 결국 다양한 사회적 집단에서 정의에 대한 여러 개념과 해석이 만들어지게 되며, 이렇게 구성된 정의에 대한 개념들은 총체적으로 사회 전체에 영향을 미치게 된다[10]. 다양성의 기반에 더해 대중 참여에 제약이 없는 소셜 미디어 환경이 법적 정의와 사회 정의의 차이 연구에 알맞

은 이유이다.

본 연구는 소셜 미디어 환경에서 관찰 가능한 사회 정의와 법적 정의의 차이를 파악하기 위해 텍스트 마이닝(text mining) 방법을 사용했다. 선행연구들은 소셜 미디어 환경 내 다양한 언어적 표현들을 살펴보면서 사회 및 제도적 가치의 이해, 해석, 번역 및 구현의 과정을 포괄적으로 이해하는 것이 가능하다고 말했다[11][12]. 이에 따라 본 연구는 대중이 자유롭게 의견을 표현하는 소셜 미디어 환경에서 사회 정의와 관련한 데이터를 수집하였다.

본 연구는 여러 소셜 미디어 플랫폼들 중, 영상과 텍스트를 모두 제공하는 글로벌 동영상 공유 플랫폼 유튜브(YouTube)를 연구 환경으로 삼았다. 유튜브는 동영상 공유 플랫폼 소셜 미디어 서비스로 실시간으로 정보를 전달하고 저장 가능하다. 접근 제한이 없는 대중적인 플랫폼이기 때문에, 댓글과 같은 사용자 정보를 관찰하거나 사회적 담론이 형성되는 과정을 살펴보기 적합하다[13][14]. 따라서 본 연구는 유튜브의 댓글이 사회 정의를 대변한다고 가정하였다.

이에 비교되는 법적 정의를 대변하는 자료로 본 연구는 법원이 발행한 공식 문서, 즉 판결 및 선고와 같은 법원의 공개 문서를 사용하였다. 사법부에서 제공하는 공식 자료이기에 동시대의 법적 정의를 충분히 반영하고 있다고 판단했다.

본 연구는 법적 정의와 사회 정의의 차이 정도를 더욱 자세히 알아보기 위해 연구 환경을 두 가지 상황에서 분석했다. 첫째는 강한 적법성(strong legality)이다. 강한 적법성은 법적 판단이 결정적으로 이루어지고 권위 있는 해석으로 수렴될 가능성이 높은 환경을 의미한다. 그래서 강한 적법성이 존재하는 환경에서는 법적 판단과 대중의 의견이 일치하고, 법적 정의와 사회 정의 간격이 크게 존재하지 않을 경향을 기대할 수 있다. 둘째는 약한 적법성(weak legality)이다. 약한 적법성은 법적 판단이 모호하고 다양한 해석이 가능할 때 발생한다. 약한 적법성 환경에서는 법적 정의와 사회 정의 차이가 강한 적법성보다 더 명확하게 드러날 수도 있다.

본 연구는 우선 사회 정의와 법적 정의의 차이가 존재하는지 확인하고자 한다. 그리고 부가적으로 선술한

두 개의 다른 적법성 환경이 차이의 정도에 영향을 미치는지도 알아보고자 한다. 연구의 결론은 연구 결과를 통해 사회적 함의를 찾아보는 것으로 마무리할 것이다.

## II. 이론적 배경

### 1. 정의와 사회 정의

역사적으로 정의는 사회의 핵심 가치였고, 다양한 시각에서 연구되고 해석되어 왔다. 고대 이집트의 문헌들과 구약성서는 정의를 도덕적 선의, 선, 결점 없음 및 완벽성으로 바라보았다[15]. 고대 그리스 철학자 플라톤은 정의가 개인과 사회 간 구조적 관계 속에서 사회적 조화와 복지를 기본으로 만드는 것이라고 말했다[16]. 이것은 플라톤이 제시한 사회정의 이론의 이론적 근거가 되었다. 중세 철학자 토마스 아퀴나스는 정의가 개인의 책임을 이행하기 위한 강력한 의지라고 말했다[17]. 오늘날 옥스포드 사전은 정의를 "인간에 대한 공정한 대우 또는 공정하고 합리적인 품질"이 정의라고 기술하고 있다[18].

정의처럼 사회 정의 역시 다양한 해석이 존재하며, 사회적 합의가 도출되지 않은 상황이다[19]. 다만 다양한 사회 영역에서 제시하는 해석들의 교집합을 따라가면, 오늘날 우리 사회가 보편적으로 받아들이는 사회 정의에 대한 의미를 추측할 수는 있다. 철학은 정의를 사회적 제도의 가장 기본적인 덕목이라 주장하는데, 이는 현대철학자 존 롤스(John Rawls)의 생각을 계승하는 것이다. Wenar(2008)는 자유 시민이 동등한 기본 권리를 보유하고 있는 평등한 체계가 공정한 사회이며, 이것이 오늘날의 사회 정의를 지칭한다고 말했다[20]. 법리에서의 사회 정의는 더 구체적이고 현대적이다. Kennedy(2005)는 오늘날 법에 근거한 사회 정의는 인종, 성별, 환경 및 국가 간의 관계가 포함되고[21], Rand(2006)는 나아가 사회적 인권을 완전히 보유하지 못한 소수자까지 사회 정의의 대상으로 지칭했다[22]. 심리학은 개인과 타인 간의 연결고리에 주목하며 사회 정의를 논의했다[23]. 사회학 문헌들은 사회 정의를 '배분적 정의(distributive justice)'라는 용어로 바꾸어 사용하며, 사회경제적 불평등을 사회 정의의 범주에 포함

하였다[24]. 이처럼 사회 정의에 대한 해석은 바라보는 시각에 따라 다르다.

살펴본 것처럼 사회 정의에 대한 해석은 다양하기에 연구 수행에 앞서 사회 정의에 대한 범위를 명확히 하는 것이 필요하다. 본 연구는 오늘날 대중들이 공통적으로 인지하는 사회 정의들의 교집합을 사회 정의로 전제하였다.

### 2. 소셜 미디어와 사회 참여

정보는 매체의 특성을 설명하는 데 중요한 조건이다. Galbraith(1974)는 작업을 수행할 때 복수 주체 간 정보 격차가 중요하기 때문에, 정보 부족은 불확실성을 증가시킨다고 말했다[25]. Ishii, Lyons(2019)는 정보 부족이 유발하는 불확실성은 정보의 양을 늘려 해결할 수 있다고 주장했다[26]. 요약하면, 정보의 충분성은 불확실성과 부(-)의 상관관계를 가진다.

매체 풍부성 이론(media richness theory)은 정보의 양 뿐만 아니라 정보의 풍부성이 매체의 특성 설명에 특별히 중요하다는 점을 설명한다[27]. 관련 후속 연구들은 더 구체적으로 즉각적인 피드백, 다양한 단서, 언어 다양성 및 개인적 초점 등 4가지 요인이 정보의 풍부성을 결정한다고 제안하고 있다[28].

매체 풍부성 이론은 오늘날 정보양이 폭발적으로 증가하고 있는 소셜 미디어 특징과 현상 설명에 알맞다. 소셜 미디어 환경 속 대중은 자유롭게 의견과 감정을 표현한다. 그들은 콘텐츠 생성 및 공유 등을 포함한 방식으로 사회활동에 참여하고, 다른 이들의 활동에 즉각적으로 반응하고 있다. 이처럼 온라인 소통 플랫폼에서는 대량의 정보와 반응이 빠르게 순환되기 때문에, 정보 수집 채널로 소셜 미디어는 사회과학 연구에 적절하다.

통신 기술의 급격한 발전과 함께 등장한 소셜 미디어는 사회 문화적 측면에서 여러 변화를 가져왔다. Stucke(2009)는 소셜 미디어 기반 기술이 사용자의 개인 및 문화적 네트워크 형성을 가능하게 한다고 말했으며[29], Poster(2018)는 통신 기술이 단순한 아이디어나 정보 교환이 아니라 문화적 맥락 속 개인 정체성 형성을 위한 것이라고 말했다[30]. 사회와 문화적 측면의 변화는 정치적 변화까지 포함하는데, Kapor(1991)는 통신 기술의 융합이 미디어를 더 자유롭고 합리적이

며 민주적인 공간으로 변화시킬 것이라며 새로운 정치적 가능성을 제공할 것이라 주장하였다[31]. Schejter and Tirosh(2015)는 사용자들이 장소와 시간의 제약을 벗어나 즉각적으로 다량의 정보에 접근 및 배포함으로써 소셜 미디어가 공공적 목적에 기여할 수 있는 잠재력과 민주적 환경을 발견했다고 말했다[32].

이러한 사회적 추세 속에서 소셜 미디어는 사회 정의를 지속적으로 재정의하는 중요한 매개체이다. 실제로 오늘날 소셜 미디어는 사회적 의제를 형성하는 패러다임 전환을 촉진하고 있다. Dutton(2008)은 소셜 미디어에 사회적 문제를 공유하고 대중의 주목을 끌어내는 과정을 통해 정보와 문화의 민주화를 이루었다고 지적하였다[33]. 예를 들어, 보수적인 아랍 국가에서는 시민들이 소셜 미디어에서 대규모로 유통되는 정보를 접하고 변화를 추구하는 사회운동을 일으키기도 했다[34][35]. 정의를 추구하는 이들에게 소셜 미디어는 메시지를 전달하고 교환하는 효과적인 도구 혹은 그들의 존재감을 높일 수 있는 공간이다. 이러한 측면에서 소셜 미디어는 오늘날의 사회 정의를 관찰하기에 좋은 매체이다.

### 3. 텍스트 마이닝과 감정 분석

텍스트 마이닝은 기계의 도움으로 문서 속 정보를 추출 및 연결하여 새로운 정보를 찾아내는 정보기술을 지칭하는데, 새로운 정보를 찾는 것을 목표로 한다는 점이 기존의 웹 검색과는 근본적으로 다르다[36]. Tan(1999)은 텍스트 마이닝을 비정형 텍스트에서 흥미로운 패턴이나 지식을 찾는 프로세스로 정의하며, 기존의 표준화 데이터베이스 기반 데이터 마이닝의 확장으로 보았다[37].

텍스트 마이닝 과정은 크게 두 단계로 구분할 수 있다 [37]. 첫째는 자유 형식의 비구조화 텍스트를 구조화된 중간 형식으로 변환하는 것이다. 두 번째는 앞선 단계에서 얻은 중간 형식을 지식 추출 과정을 통해 구체적인 형식이나 지식으로 전환하는 것이다. 지식 추출의 예로는 군집화, 분류, 시각화, 예측 모델 등이 있다.

최근 텍스트 마이닝 기법은 사회과학 분야에서 자주 활용되고 있는데, 덕분에 연구자들은 비구조화 형태의 데이터 접근 및 수집이 용이해졌다. 온라인 게시물 및

소셜 미디어 댓글 분석 등이 대표적인 예시이다. 나아가 텍스트 마이닝을 통해 연구자들은 단어 간 패턴 및 문맥 분석을 통해 의미 해석 이상의 문맥 속 감정까지 분석하게 되었다[38].

감정 분석은 대상물에 대한 사람들의 다양한 감정과 의견을 분석하는 것이다[39]. 김유영 and 송민(2016)은 구조화되지 않은 데이터에서 발견된 감정을 객관적으로 분류하거나 양적으로 분석하여 구조화된 데이터로 변환하는 과정을 감정 분석이라 정의하고 있다[40]. 사람들은 언어를 통해 감정을 표출하기에, 감정 분석은 텍스트 마이닝의 한 분야로 볼 수 있다. 따라서 많은 감정 분석 연구들이 텍스트 마이닝 기법을 이용하여 수행되었다[41].

의견을 자유롭게 교환하고 저장할 수 있는 소셜 미디어는 감정 분석의 손쉬운 연구 대상일 수 있다[42]. 다양한 사회적 이벤트가 실시간으로 공유되고 확산되는 소셜 미디어에서 사용자들은 적극적으로 자신의 감정을 나타낸다. 동시에 소셜 미디어 커뮤니케이션이 대중화되며 영향력은 점점 증가하고 있다는 부분에서, 소셜 미디어의 메타데이터들은 감정 분석과 관련된 학술 연구에 풍부한 데이터를 제공한다. 본 연구는 역시 감정 분석의 연구 환경으로 소셜 미디어 플랫폼을 고려하였고, 여러 소셜 미디어들 중에서 법적 정의와 사회 정의가 모두 관찰 가능한 유튜브 채널을 찾아 연구 대상으로 선정하였다.

### 4. 챗GPT

챗 GPT는 2022년 OpenAI에서 출시한 대화형 인공지능 모델이다. 자연어 처리에서 널리 사용되는 트랜스포머(transformer)의 디코더(decoder)를 기반으로 구축된 GPT-3를 기본으로 한다. 챗 GPT는 사용자의 입력(prompt)을 기반으로 답변(response)을 생성하는 방식으로 작용하며, 이를 통해 사용자와 기계의 상호작용이 상호 작용이 가능하다.

챗 GPT는 생성형 AI의 대표 기술로 꼽히는데, 생성형 AI는 입력에 따라 여러 형태의 결과를 생성해 낼 수 있다. 본 연구에서는 언어에 특화된 거대 언어 모델 LLM(Large Language Model)을 사용한다. 최근 LLM은 트랜스포머를 이용한 모델들이 주를 이루고 있는데,

대표적으로 GPT-4, PaLM, LLaMA등이 있다.

챗 GPT로 이름을 알린 GPT-4는 트랜스포머에 in-context learning 이라는 메타 러닝과 함께 사용자의 평가를 기반으로 재학습하는 RLHF(Reinforcement Learning from Human Feedback)을 이용한 모델이다.

PaLM 은 모델이 거대할수록 성능이 향상된다는 전제하에 구글이 개발한 모델이다. 반면 LLaMA는 작은 모델을 오랫동안 학습하면 추론에 더 유리하다는 접근법으로 메타에서 개발한 모델이다.

상기 언급한 여러 모델들 중 본 연구는 챗 GPT를 이용하기로 결정했는데, 연구 시점 가장 대중적인 생성형 AI였기 때문이다. 연구를 수행하던 시점에 다른 모델들은 막 서비스 상용화를 시작하고 있었기에 챗 GPT를 대체할 대안이 없었다. 부가적으로 사용자의 입력과 출력을 반영하는 챗 GPT의 특징도 도움이 될 것이라 판단했다. 본 연구는 사전 훈련 과정이 필요했는데, 챗 GPT가 이에 적합할 것이라 생각했다.

챗 GPT는 다른 자연어 모델과 유사하게 사전 훈련(pre-training)과 미세 조정(fine-tuning)의 과정으로 구성된다. 인터넷 등에서 수집한 대량의 텍스트를 활용하여 사전 훈련을 진행하고, 미세조정은 사전 학습된 모델을 기반으로 과업(task)에 맞게 재학습하는 것을 의미한다. 사전에 학습한 방대한 정보를 과업에 알맞게 재학습하는 과정을 거쳐 높은 성능을 달성할 수 있음이 알려져 있다[43].

아직 초기 단계이지만 챗 GPT를 활용한 감정 분석 시도는 이루어지고 있다. Wang, Xie(2023)는 챗 GPT는 입력 설정 값을 적합하게 설정한다면 기존의 기계학습 보다 더 포괄적이고 유용하게 감정분석에 활용될 수 있다고 말하고 있다[44].

### III. 연구

#### 1. 연구 환경

소셜 미디어는 오늘날 대중의 의견이 표출되고 또한 모이는 공간이다. 이에 소셜 미디어에 남겨진 메타데이터는 사회 현상의 의미를 찾는데 도움을 준다[45]. 유

사한 맥락에서 사회 정의와 관련된 현상, 특히 사회 운동의 촉발점과 발화점을 찾기 위한 선행 연구들은 소셜 미디어를 연구 대상으로 활용했다[4-6]. 하지만 소셜 미디어 환경에서 법적 정의와 사회 정의라는 두 가지의 추상적 개념의 비교하고 차이를 찾아본 선행 연구는 찾기 어려웠다.

이에 본 연구는 소셜 미디어의 사회 정의와 법적 정의가 모두 집중 관찰 가능한 연구 환경을 선정하는데 신중을 기했다. 그리고 유튜브 채널, '한문철 TV'를 연구 환경으로 선택했다. 해당 채널은 교통사고 소송 전문 변호사가 운영하는 채널로, 교통사고 영상을 제보하고 공유하는 플랫폼이다. 채널 운영자인 교통사고 전문 변호사가 영상에 대해 조언 및 의견을 제시하는 동안 채널 방문자들은 자유롭게 댓글을 쓰며 반응한다. 2022년 10월 기준 해당 채널에는 18,000건 이상의 교통사고 영상이 공유되어 있어 있고, 구독자는 160만 명 이상이다.

해당 채널을 연구 환경으로 선택한 배경에는 몇 가지 이유가 있다. 첫째, 해당 채널에는 교통사고의 원본 영상이 공유되기 때문에 상황을 객관적으로 보여준다. 둘째, 교통사고는 음주 운전, 과속, 어린이 안전 등과 같은 중요한 사회 안전으로 대중적 관심사이다. 그래서 해당 채널에는 대중들의 다양한 감정과 시각이 반영되어 있다. 셋째, 교통사고 판결문은 손쉽게 열람 가능하기에 관련 법적 의견과 데이터를 얻기 쉽다. 해당 채널은 사고 영상과 채널 설명 공간에 법원 판결 내용을 포함하고 있는데, 이런 이유로 채널 방문자들은 교통사고 영상과 해당 영상의 법적 판결 내용을 함께 볼 수 있다. 이런 측면에서, 우리는 해당 채널은 법적 결정에 대한 대중 의견을 관찰하기에 적합하다고 판단했다.

본 연구는 해당 채널에 공유된 모든 영상 콘텐츠를 연구 샘플로 이용하지는 않고, 법적 판결 내용이 함께 담겨있는 영상만을 연구 대상으로 삼았다. 그리고 연구 대상을 다시 적법성 정도(degree)에 따라 두 샘플 집단으로 분류하였다. 통상적으로 사법부는 차량 사고를 판결할 때 사고 관여자들에 대해 과실률을 분배하는데, 본 연구는 완전 과실을 강한 적법성, 부분 과실을 약한 적법성이라 간주했다.

강한 적법성 샘플은 한 쪽의 완전 과실, 즉 100%과

실로 판결된 영상들만으로 구성되었다. 사법부가 완전 과실 판결을 했다는 사실은, 해당 사고에 대해 분쟁의 여지가 적을 것이라는 법적 의견이 반영된 것이라고 보았기 때문이다. 반면 약한 적법성 샘플에는 사법부에서 부분 과실로 판결한 차량 사고 영상들을 모았다.

본 연구는 두 적법성 환경에 따라 분류된 댓글의 감정을 분석하고 비교했다. 이를 통해 적법성의 정도에 따른 법적 정의와 사회적 정의의 차이와 차이의 의미를 알아보고자 했다.

## 2. 연구 방법

본 연구는 지도 학습 기반 감정 분석 접근법을 이용하였다. 감정 분석의 여러 연구에서 지도 기계학습은 자주 활용되고 있다[46][47]. 기계 학습 기술을 통해 수행된 감정분석은 감성에 대한 긍정적인 댓글과 부정적인 댓글을 라벨링하는 데 유용하기 때문에[46], 본 연구 수행의 목적에 적합하다고 판단했다.

LLM 을 이용한 기계 학습에서는 일반적으로 임베딩(embedding)과 미세조정이 사용되는데, 본 연구에는 미세 조정이 더 적합하다 판단했다.

임베딩은 미리 학습된 모델에 다른 종류의 데이터를 파라미터의 조정 없이 결합시키는 것으로, 기본 모델에 추가적인 정보를 업데이트하기 위해 사용한다. 이에 반해 미세조정은 작은 규모의 데이터셋을 이용하여 추가적인 학습을 진행하기 때문에, 특정 작업에 알맞다.

본 연구의 소셜 미디어의 메타데이터인 댓글의 감정을

을 분류하는 것이다. 목적이 명확하고, 이에 더해 입력과 출력의 형태가 정해져 있다. 임베딩보다는 미세조정이 본 연구에 더 알맞을 것이라 판단한 이유이다.

본 연구의 댓글 감정 분석은 [그림 1]과 같이 다섯 단계를 거쳐 진행되었다. 구체적으로, 데이터 수집, 데이터 전처리, 정보 추출 및 의미요소 분류, 감정 분류, 분석이다.

### 2.1 데이터 수집

데이터 수집은 유튜브에서 지원하는 API(Application Programming Interface)를 활용하였다. 유튜브 API 는 동영상 관련 메타데이터의 외부 공유를 허용하기 때문에 연구 샘플과 관련있는 데이터 접근이 손쉬워졌다.

본 단계에서는 연구 목적 관련 데이터만을 수집했다. 영상의 경우, 메타데이터 내에 실제 판결문이 포함되어 있는 영상들만을 선별했다. 감정분석의 주제는 '판사' 또는 '판결'로 제한하였으며 이들을 주어로 사용하는 댓글 문장만을 감정분석의 대상으로 삼았다. '판사'와 '판결'은 각각 법적 정의의 주체와 결과물이기에, 해당 단어가 주어로 사용된 문장들은 법적 정의에 대한 의견이나 감정이 내포되어 있다고 판단하였기 때문이다.

하나의 영상에 한 개 이상의 판결문이 있는 경우 최근 판결문을 분석 대상으로 하였다. 대한민국 사법제도는 3심제로 이루어져 있는데 대부분의 교통사고 판결은 1심에서 확정되고 일부는 2심에서 확정되고 있다. 연구 샘플 영상 속 차량 사고들도 마찬가지였다. 대부

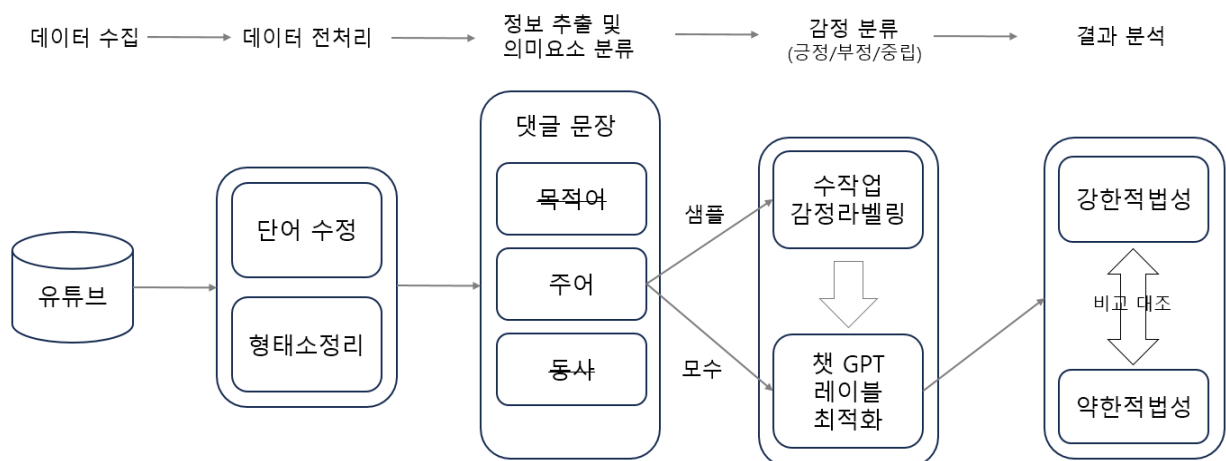


그림 1. 감정 분석 단계

본의 경우 1심에서 판결이 종결되었고, 일부가 2심에서 확정되었다. 3심까지 간 사례는 없었다. 1심 판결과 2심 판결이 모두 포함된 영상의 경우는 최종 판결인 2심 판결문과 관련 댓글만을 연구 샘플로 활용하였다.

데이터 수집한 시점인 2022년 5월 기준, 채널에 공유된 총 영상의 수는 15,121개였는데, 이 중 판결문이 포함된 영상은 154개였고, 이에 달린 댓글은 50,320개였다. [표 1]은 연구 샘플로 활용된 영상에 대한 요약, 그리고 [표 2]는 해당 영상에 달린 댓글에 대한 요약이다. 연구 샘플 관련 영상에는 최종 판결문과 함께 판결문 과실 비율이 함께 제시되어 있기에 해당 표에도 과실비율 기준으로 세부 분류를 진행했다.

표 1. 연구 샘플 영상 요약

	판결문 과실 비율					전체
	100:0	90:10	80:20	70:30	기타	
1심	32	17	32	15	17	113
2심	17	3	9	5	7	41
전체	49	20	41	20	24	154

표 2. 연구 샘플 댓글 요약

	판결문 과실 비율					전체
	100:0	90:10	80:20	70:30	기타	
1심	9,685	2,828	11,700	2,809	5,120	32,142
2심	7,955	1,683	4,692	1,644	2,204	18,178
전체	17,640	4,511	16,392	4,453	7,324	50,320

## 2.2 데이터 전처리

본 단계에서는 연구 샘플 관련 데이터를 연구 목적에 알맞게 전처리(pre-process)하였다. 소셜 미디어 사용자들은 온라인 댓글을 구어체로 쓰는 경향이 있다. 댓글에는 다양한 텍스트 변형이 있어 데이터 전처리는 필수적이다.

구체적으로 기계학습으로 의미처리가 어려운 이모지와 이모티콘들을 처리하였다. 또한 단일 모음 혹은 자음으로 구성된 의미 없는 형태소들을 제거하고, 문장 구조와 간격을 수정하였다.

## 2.3 정보 추출 및 의미요소 분석

전처리를 거친 댓글들은 형태소 단위로 분할되었다. 형태소 분할에는 공공 기관 전자통신연구원(ETRI)에서 무료 제공하는 의미 분석 기술 API를 이용하였다. 해당 API는 태그셋(tag set)을 기반으로 자연어 문장의 최소 단위인 의미 단위를 분석하는 자연어 처리 소프트웨어 어플리케이션이다.

댓글들은 주어, 목적어, 조사 등 여러 형태소로 분할되었는데, 본 연구는 특별히 주어에 주목하기로 했다. 한국어는 감정이 투영된 부분은 대체로 명사와 목적어이다. 그런데 구어체 형태로 써지는 댓글은 특히 주어 쪽에 감정에 반영되는 경향이 강하다. 이런 측면에서 본 연구에서는 댓글에서 주어에 위치한 명사들에 주목하였다. [표 3] 형태소 분석에서 도출된 주어들을 높은 빈도부터 정리한 것이다. 동시에 판결 수준, 즉 심급으로도 분류해 보았다.

표 3. 빈도수가 높은 단어(주어)

단어(주어)	빈도수(%)		
	1심	2심	전체
판사	3,311(10.3%)	2,786(5.3%)	6,097(12.1%)
보험사	3,137(9.8%)	535(2.9%)	3,672(7.3%)
블랙박스차량	1,309(4.1%)	635(3.5%)	1,944(3.9%)
사고	1,097(3.4%)	529(2.9%)	1,626(3.25%)
판결	770(2.4%)	567(3.1%)	1,337(2.7%)
비보호	936(2.9%)	363(2.0%)	1,299(2.6%)
좌회전	667(2.1%)	366(2.0%)	1,033(2.1%)
과실	713(2.2%)	311(1.7%)	1,024(2.0%)
운전자	650 (2.0%)	329(1.8%)	979(1.9%)

단어 빈도 분석 결과는 몇 가지 흥미로운 부분을 보여준다. 첫째, 판결과 관련된 단어들이 많이 보인다. 가장 빈도가 높은 주어는 '판사'이며, '보험사'와 '판결'을 주어로 포함하고 있는 댓글도 많다. 둘째, '블랙박스 차량', '좌회전'과 같이 사고 상황을 기술하는 단어들도 보인다. 해당 단어들은 높은 빈도를 보여주지만 댓글을 쓰는 사용자들이 직접적으로 감정을 투영할 수 있는 대



상이 아니기 때문에 감정 분석 대상에서는 제외하였다.

따라서, 본 연구는 소셜 미디어 사용자가 차량 사고 영상과 판결문에 대해 직접적으로 감정을 투영할 수 있는 고빈도 단어만을 중심으로 감정 분석을 진행하기로 결정했다. 바로 '판결'과 '판사'이다. 이런 맥락에서 첫 번째 단계에서 얻은 50,320개의 댓글들 중 판결이나 판사가 주어로 작성된 12,134개의 댓글만을 감정 분석 대상으로 삼았다.

#### 2.4.1. 감정 분류(수작업)

본 단계에서는 최종 감정 분석 대상 댓글들, 약 5%에 해당하는 중 600개의 댓글을 무작위로 추출하여 감정 분석을 수작업으로 분류하였다. 추출된 600개의 댓글은 연구 초반 설정한 두 가지 다른 적법성, 즉 강한 적법성과 약한 적법성에 따라 분리되었다. 강한 적법성을 반영하는 샘플은 100 대 0의 차량 과실 비율 판결을 받은 동영상 속에서 무작위로 선별한 300개의 댓글로 채웠고, 약한 적법성을 반영하는 샘플은 부분 과실 비율로 판결 받은 동영상에서 무작위로 선별한 300개의 댓글로 채웠다.

600개 댓글의 감정 분류는 연구에 참여한 연구원들이 수작업으로 하였다. 연구원들은 개별적으로 독립 환경에서 댓글을 직접 읽어보고, 긍정, 부정, 중립의 세 가지 감정 범주 중 하나로 분류하였다. 이후 모든 연구원들은 모여 분류 값을 비교했다. 만장일치의 경우 해당 감정값을 할당하고, 그렇지 않으면 합의의 과정을 거쳐 가장 빈번한 값을 할당했다. 합의에 이르지 못하면 오류 값으로 판정하였다. 총 1개의 오류 값이 나왔다.

기계 학습이 아닌 수작업 감정 분석을 진행한 배경에는 몇 가지 이유가 있다. 첫째, 기계 학습에서는 구어체 댓글 속에서 발견되는 변형 단어들에 대한 감정 판단에 불완전하다. 둘째, 일부 댓글은 기계가 이해하기 어려운 상황적(contextual)문장이나 유머를 포함하고 있었다. 셋째, 일부 댓글들은 기계 학습을 진행하기에는 매우 짧거나 소수의 단어로 구성되어 있다.

#### 2.4.2. 감정 분류(챗 GPT)

앞선 단계에서 수작업으로 분류한 600개의 댓글은 본 단계의 기계학습에 훈련 집단(training set)으로 활용되었다.

용되었다.

챗 GPT의 제작사인 오픈 AI(Open AI)는 범주 별로 최소 100개 이상의 레이블을 입력할 것을 권고하고 있기에 앞선 단계에서 수작업으로 진행한 댓글 수는 지도 학습 기반 기계 학습에 충분하다고 판단되었다.

본 연구는 600개의 댓글 중 오류로 판정된 1개를 제외한 599개의 댓글 분류 결과를 기반으로 챗 GPT를 미세조정하였다. 이는 1만 2천개가 넘는 댓글을 직접 분류하는 대신, 600개에 대한 레이블을 사람이 진행 후 모델이 패턴을 인지하여 나머지 댓글을 연구원들의 수작업 방식과 유사하게 레이블(label)해주는 것을 가능하게 했다.

### 2.5. 결과 분석

궁극적으로 챗 GPT를 이용한 감정 분석을 통해 모든 댓글에 대한 빈도, 비율과 같은 기술 통계 값을 얻을 수 있었고, 이를 적법성과 및 판결 심급에 따라 세부 분류해 보았다.

## IV. 연구 결과

연구에서 얻은 감정 분석의 결과는 분류 감정과 적법성의 정도에 측면에서 정리하여 [표 4]에 나타내었다.

표 4. 적법성과 분류감정 측면에서의 댓글 감정 분석 결과

분류 감정	댓글 수(%)	
	강한 적법성	약한 적법성
긍정	450(24.55%)	569(5.52%)
부정	958(52.26%)	8,406(81.60%)
중립	427(23.30%)	1,326(12.87%)
전체	1,833(100%)	10,301(100%)

적법성의 정도에 의해 분류된 두 그룹은 긍정, 부정, 중립 댓글 수에서 차이를 보였다. 이 차이가 통계적으로 유의한지를 확인하기 위해 본 연구는 파이썬 통계 라이브러리를 이용해 유의성 검정을 실시했다.

검정 결과 강한 적법성 그룹에서 긍정 댓글의 비율은 24.55%이며 표준 오차는 0.010, 약한 적법성 그룹에서 긍정 댓글의 비율은 5.52%과 표준 오차는 0.012로, 두

그룹간 차이는 99% 신뢰 수준에서 통계적으로 유의했다. 강한 적법성과 약한 적법성에서의 부정적 댓글 비율은 각각 52.26%와 81.60%, 표준 오차는 각각 0.012와 0.0038으로, 이 역시 99% 신뢰수준에서 통계적으로 유의한 차이를 보였다. 마지막으로 중립적 댓글 비율은 각각 23.30%와 12.87%, 표준 오차는 각각 0.0033과 0.0099로, 99.9%의 신뢰 수준에서 통계적으로 유의한 차이를 나타냈다. 즉 적법성의 정도에 따라 감정 비율의 차이가 존재하는 것이다.

또한 동일한 적법성 정도에서도 차이점이 발견된다. 약한 적법성 그룹 내 부정 댓글 수가 긍정 댓글 수의 거의 15배에 이르는 것을 알 수 있다. 반면 강한 적법성 그룹에서는 부정적인 댓글 수는 긍정적인 댓글 수의 2배 정도이다. 그리고 강한 적법성 그룹의 약 23%는 중립적인 감정을 유지하는 반면, 약한 적법성 그룹 내에서 중립 입장을 보이는 사용자는 13%이다.

판결 심급 수준에 따른 차이는 비교적 작은 편이다. [표 5]는 [표 4]의 감정 분석 결과를 판결 심급에 따라 더 상세하게 분류한 내용이다.

표 5. [표 4]를 판결 심급에 따라 세부 분류한 댓글 수

판결 수준	적법성	감정에 따른 댓글수(%)			전체 댓글수 (%)
		긍정	부정	중립	
1심	강함	324 (30.11%)	571 (53.07%)	181 (16.82%)	1,076 (100%)
	약함	310 (5.50%)	4,544 (80.62%)	782 (13.88%)	5,636 (100%)
2심	강함	126 (16.64%)	385 (50.86%)	246 (32.50%)	757 (100%)
	약함	259 (5.55%)	3,862 (82.79%)	544 (11.66%)	4,665 (100%)

강한 적법성의 경우, 제1심과 제2심에서 긍정 비율은 각각 30.11%와 16.64%으로, 제1심에서는 제2심보다 사용자들이 더 긍정적으로 반응하는 것으로 나타났다. 반면 부정 감정의 비율은 제1심 53.07%와 제2심 50.86%로 큰 차이가 없다. 중립 감정 비율은 각각 16.82%와 32.50%로 2심에서의 중립 비율이 더 높았다.

## V. 논의

본 연구는 법적 정의와 사회 정의가 모두 관찰 가능한 소셜 미디어 채널의 메타데이터들을 감정 분석하여, 그들의 간극 정도를 가늠해 보았다. 연구 결과는 적법성, 판결 수준이 만드는 상황에 따라 법적 정의와 사회 정의의 차이가 달라질 수 있음을 보여주었다.

[표 4]의 결과에서 알 수 있듯, 강한 적법성 상황보다 약한 적법성 상황에서 사회 정의와 법적 정의의 간극 차이가 더 크다. 약한 적법성 그룹에 속한 댓글의 부정 비율은 80%를 상회하는 반면 강한 적법성의 비율은 50%를 조금 넘는 정도이다. 긍정 비율은 이를 더 극명하게 보여준다. 강한 적법성의 긍정 비율은 24.55%일 때 약한 적법성은 5.52%뿐이다. 약한 적법성과 비교해 강한 법적정의 판결에서는 부정 댓글의 비율이 낮아지고 긍정의 비율이 높아졌다.

댓글 감정 분석에서 긍정이 사법부의 판결문에 동의하고, 부정이 사법부 판결문에 동의하지 않는다고 전제하다면, 이는 사법부가 차량 사고에서 한 쪽의 과실이 명백하다고 판결을 내린, 즉 100:0 과실 사고 관련 영상을 본 사용자들은 해당 판결에 대해 수긍하는 경향이 있다고 해석 가능하다. 즉, 강한 적법성 상황에서 사회 정의와 법적 정의의 간극은 비교적 적은 편이다.

이에 반해 약한 적법성을 반영하는 부분 과실로 판결 난 사고 영상에 대해 소셜 미디어 사용자들은 쉽게 동의하지 못하는 모습이다. [표 5]에서 보듯 판결 수준, 즉 심급에 관계없이 약한 적법성 상황의 부정 비율은 80%가 넘는다. 평균적으로 소셜 미디어 사용자 5명 중 4명이 판결 결과에 동의하지 못하는 것이다. 사법부에서 부분과실로 종결된 차량 사고는 유권해석의 가능성이 높을 수 있는데, 이런 약한 적법성의 경우 사법부에서 제시한 법적 정의와 소셜 미디어에서 수렴하는 사회 정의는 상당한 간극을 보인다. 즉 약한 적법성 환경에서는 사회 정의와 법적 정의의 차이가 크다고 해석 가능하다.

연구 결과와 함께 본 연구는 다음과 같은 의의를 가진다. 첫째, 본 연구는 대중들의 수렴된 의견이 반영된 사회 정의라는 추상적 개념에 대한 감정 분류를 통해 사회 정의를 구체화시키고 이를 법적 정의와 대조해 보

았는데, 이는 사회 정의와 관련한 기존 연구에서는 찾기 어려운 시도였다. 이에 더해 완전 과실과 부분 과실로 분류되는 교통사고의 특징을 반영해 법적 정의의 상황을 강한 적법성과 약한 적법성으로 세부 분류한 뒤 적법성의 강도에 따른 사회 정의와 법적 정의의 간극을 개별적으로 측정해 보았다.

둘째, 본 연구는 법적 정의와 사회 정의의 간극의 차이를 알아보기 위해 차량 사고라는 대중적인 관심사를 연구 토픽으로 설정했다. 교통은 오늘날의 일상에서 많은 대중들이 이용하는 수단이며, 크고 작은 사고가 빈번한 영역이다. 교통사고라는 대중적 관심사를 사회 정의의 척도로 변환하고 이를 법적 정의와 비교했다는 점에서 의의가 있을 것이다.

셋째, 법적 정의와 사회 정의의 간극을 정량적으로 표현해 보았다. 법은 최소한의 상식이라는 의례적인 표현처럼 법적 정의와 사회 정의의 간극은 분명 존재할 것이며 이것은 여러 선행연구에서도 발견되었다. 하지만 간극의 차이를 정량적으로 측정하기 위한 연구는 많지 않았다. 본 연구는 소셜 미디어 플랫폼의 의견을 감정 분석 및 정량화하여 사회 정의와 법적 정의의 간극의 정도를 추정해 보았다.

동시에 본 연구는 다음과 같은 한계점을 가지고 있다.

첫째, 본 연구의 샘플은 내재적으로 편향성을 가질 수 있다. 연구 샘플인 차량 사고 영상 공유 채널은 교통사고 전문 변호사가 운영하는 채널이다. 교통사고와 관련한 올바른 정보를 전달한다는 공공의 이익과 함께 영상 공유의 목적이 사법부의 판결에 문제점을 제기하기 위함이다. 이런 환경 속에서 영상 시청자들은 사법부의 판결에 쉽게 비판의 의견을 낼 수 있다. 또한 다른 댓글에 쉽게 동조하는 미디어 환경 역시 편향성을 조장하는 모습을 보일 가능성이 있다.

둘째, 본 연구는 한국어로 작성된 댓글을 분석하여 사회적 정의를 측정하였기에 본 연구의 결과가 다른 언어 환경에 동일하게 해석되거나 일반화되기 어렵다. 특히, 한국어로 쓰여진 증의적 표현, 비언어적 요소들은 언어가 쓰여지는 국가의 문화적 특수성이 반영되어 있다. 댓글 감정분석 결과에 영향을 미칠 수 있는 부분이다. 이런 한계점들로 본 연구 방법이 다른 언어에 적용되었을 때는 상이한 결과가 도출될 가능성이 있다.

셋째, 연구 방법에서 이용된 기계 학습 방법에서의 차이점이다. 본 연구는 댓글 감정 분석 단계에서 ETRI에서 제공하는 한국어 형태소 분석 API를 이용하였고, 감정 분석에서는 챗 GPT를 이용하였다. 다른 형태의 기계학습으로 형태소 분석 및 감정 분석을 진행한다면, 본 연구 결과와 차이가 있을 수 있다.

## 참 고 문 헌

- [1] A. Barak, *the Judge in a Democracy* Princeton, University Press, 2016.
- [2] B. Dickson, "A life in the law: The process of judging," *Sask. L. Rev.*, Vol.63, p.373, 2000.
- [3] J. S. Slovak, "Influence and Issues in the legal Community: The Role of a legal Elite," *American Bar Foundation Research Journal*, Vol.6, No.1, pp.141-194, 1981.
- [4] S. Harlow, "Social media and social movements: Facebook and an online Guatemalan justice movement that moved offline," *New media & society*, Vol.14, No.2, pp.225-243, 2012.
- [5] D. Kidd and K. McIntosh, "Social media and social movements," *Sociology Compass*, Vol.10, No.9, pp.785-794, 2016.
- [6] A. R. Lopes, "The impact of social media on social movements: The new opportunity and mobilizing structure," *Journal of Political Science Research*, Vol.4, No.1, pp.1-23, 2014.
- [7] J. W. Meyer and B. Rowan, "Institutionalized organizations: Formal structure as myth and ceremony," *American journal of sociology*, Vol.83, No.2, pp.340-363, 1977.
- [8] N. Phillips and N. Malhotra, *Language, cognition and institutions: Studying institutionalization using linguistic methods*, The SAGE handbook of organizational institutionalism, 2017.
- [9] A. Wæraas and J. A. Nielsen, "Translation theory 'translated': Three perspectives on translation in organizational research,"

- International journal of management reviews, Vol.18, No.3, pp.236-270, 2016.
- [10] P. Degoe, "Contagious justice: Exploring the social construction of justice in organizations," *Research in organizational behavior*, Vol.22, pp.51-102, 2000.
- [11] J. Pries-Heje and R. Baskerville, "The translation and adaptation of agile methods: a discourse of fragmentation and articulation," *Information Technology & People*, Vol.30, No.2, pp.396-423, 2017.
- [12] P. Tracey, E. Dalpiaz, and N. Phillips, "Fish out of water: Translation, legitimation, and new venture creation," *Academy of Management Journal*, Vol.61, No.5, pp.1627-1666, 2018.
- [13] P. Benson, "YouTube as text," A. Chik, & CA Hafner, *DIscourse Analysis and DIgital Practices*, pp.81-96, 2015.
- [14] T.S. Liew and H. Hassan, "The search for national identity in the discourse analysis of YouTube comments," *Journal of Language and Linguistic Studies*, Vol.17, No.4, pp.1806-1821, 2021.
- [15] M. D. Udoudom and S. A. Bassey, "Plato and John Rawls on Social Justice," *Researchers World*, Vol.9, No.3, pp.110-114, 2018.
- [16] F. M. Cornford, *The republic of Plato*, CUP Archive, 1976.
- [17] J. O'Callaghan, *Thomas aquinas*, in *The Routledge handbook of philosophy of memory*, Routledge, 2017.
- [18] A. S. Hornby and A. P. Cowie, *Oxford advanced learner's dictionary of current English*, 1977.
- [19] K. Buettner-Schmidt and M. L. Lobo, "Social justice: A concept analysis," *Journal of advanced nursing*, Vol.68, No.4, pp.948-958, 2012.
- [20] L. Wenar, *John rawls*, 2008.
- [21] D. Kennedy, "The social justice element in legal education in the United States," *Harvard Unbound*, Vol.1, No.1, pp.93-104, 2005.
- [22] S. Rand, "Teaching law students to practice social justice: An interdisciplinary search for help through social work's empowerment approach," *Clinical L. Rev.*, Vol.13, p.459, 2006.
- [23] E. Hatfield and R. L. Rapson, "Social justice and the clash of cultures," *Psychological Inquiry*, Vol.16, No.4, pp.172-175, 2005.
- [24] J. Scott and G. Marshall, *A dictionary of sociology*, Oxford University Press, USA, 2009.
- [25] J. R. Galbraith, "Organization design: An information processing view," *Interfaces*, Vol.4, No.3, pp.28-36, 1974.
- [26] K. Ishii, M. M. Lyons, and S. A. Carr, "Revisiting media richness theory for today and future," *Human Behavior and Emerging Technologies*, Vol.1, No.2, pp.124-131, 2019.
- [27] R. L. Daft and R. H. Lengel, "Organizational information requirements, media richness and structural design," *Management science*, Vol.32, No.5, pp.554-571, 1986.
- [28] R. L. Daft, R. H. Lengel, and L. K. Trevino, "Message Equivocality, Media Selection and Manager Performance: Implications for Information Systems," *MIS Quarterly*, Vol.11, No.3, pp.366-366, 1987.
- [29] M. E. Stucke, *Concentrated Media is Something We Can't Ignore: A Response to Speaker Pelosi*. University of Tennessee legal studies research paper, 2009.
- [30] M. Poster, *The second media age*, John Wiley & Sons, 2019.
- [31] M. Kapor, "Civil liberties in cyberspace," *Scientific American*, Vol.265, No.3, pp.158-165, 1991.
- [32] A. M. Schejter and N. Tirosh, "'Seek the meek, seek the just': Social media and social justice," *Telecommunications policy*, Vol.39, No.9, pp.796-803, 2015.
- [33] W. H. Dutton, "The wisdom of collaborative network organizations: Capturing the value of networked individuals," *Prometheus*, Vol.26, No.3, pp.211-230, 2008.

- [34] Y. Benkler, *Freedom in the commons: Towards a political economy of information*, Duke LJ, 2002.
- [35] P. N. Howard and M. M. Hussain, "The upheavals in Egypt and Tunisia: The role of digital media," *Journal of democracy*, Vol.22, No.3, 2011.
- [36] M. Hearst, *What is text mining*, SIMS, UC Berkeley, 2003.
- [37] A. H. Tan, "Text mining: The state of the art and the challenges," *Proceedings of the pakdd 1999 workshop on knowledge discovery from advanced databases*, 1999.
- [38] 전채남, 서일원, "빅데이터 분석의 기술마케팅 활용에 관한 연구: 잠재 수요기업 발굴을 중심으로," *마케팅논집*, 제21권, 제2호, pp.181-203, 2013.
- [39] B. Liu, *Sentiment analysis and opinion mining*, Synthesis lectures on human language technologies, 2012.
- [40] 김유영, 송민, "영화 리뷰 감성 분석을 위한 텍스트 마이닝 기반 감성 분류기 구축," *지능정보연구*, 제22권, 제3호, pp.71-89, 2016.
- [41] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams engineering journal*, Vol.5, No.4, pp.1093-1113, 2014.
- [42] R. Feldman, "Techniques and applications for sentiment analysis," *Communications of the ACM*, Vol.56, No.4, pp.82-89, 2013.
- [43] J. Howard and S. Ruder, *Universal language model fine-tuning for text classification*, arXiv preprint arXiv:1801.06146, 2018.
- [44] Z. Wang, Q. Xie, Z. Ding, Y. Feng, and R. Xia., "Is ChatGPT a good sentiment analyzer? A preliminary study," arXiv preprint arXiv:2304.04339, 2023.
- [45] A.M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Business horizons*, Vol.53, No.1, pp59-68, 2010
- [46] A. Bhagat, A. Sharma, and S. Chettri, "Machine learning based sentiment analysis for text messages," *International Journal of*

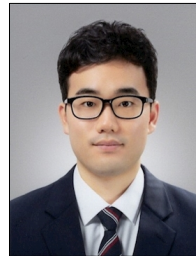
*Computing and Technology*, Vol.7, No.6, pp.103-109, 2020.

- [47] N. U. Pannala, C. P. Nawarathna, J. T. K. Jayakody, L. Rupasinghe, and K. Krishnadeva, "Supervised learning based approach to aspect based sentiment analysis," *IEEE international conference on computer and information technology (CIT)*, 2016.

## 저 자 소 개

### 민 지 환(Ji-Hwan Min)

준회원



- 2020년 2월 : 세종대학교 경제학, 수학, 통계학 (경제학사, 이학사)
- 2022년 4월 ~ 현재 : WorldQuant University 금융공학 (석사과정), 모두의연구소 연구원

〈관심분야〉 : 시계열 분석, 자연어 처리, 계산 금융

### 김 태 익(Taeik Kim)

준회원



- 2011년 8월 : 한국과학기술원 화학과 (이학사)
- 2022년 1월 ~ 현재 : 버즈빌 Data Analyst, 모두의연구소 연구원

〈관심분야〉 : 소셜네트워크, 추천시스템, 뉴스 전파

### 이 세 현(Se-Hyun Lee)

준회원



- 2020년 7월 : University of Twente 화학공학과 (중퇴)
- 2024년 2월 : 학점은행제 컴퓨터공학과 (공학사)
- 2023년 5월 ~ 현재 : 링크스톤파트너스 DS/AI 매니저, 모두의연구소 연구원

〈관심분야〉 : 사회과학, 텍스트 마이닝, 통계적 분석

최 화 준(Hwa-Joon Choi)

정회원



- 2011년 1월 : 연세대학교 영어영문학과(문학사)
- 2014년 5월 : HEC Paris 경영대학원(MBA)
- 2022년 8월 : 연세대학교 일반대학원 기술경영학협동과정(공학박사)
- 2020년 3월 ~ 현재 : 아주대학교

경영대학원 겸임교수

〈관심분야〉 : 텍스트 분석, 창업 생태계

김 상 준(Sang-Joon Kim)

정회원



- 2003년 2월 : 연세대학교 경영학과(경영학사)
- 2014년 6월 : University of California, Irvine 경영학과(경영학박사)
- 2016년 3월 ~ 현재 : 이화여자대학교 경영대학 부교수

〈관심분야〉 : 사회구성주의, 기업형성, 텍스트 분석