

# 카카오 알고리즘 윤리 헌장의 해제(解題)

인공지능(artificial intelligence, AI)은 시대의 화두입니다. AI 경쟁력의 확보 여부가 좁게는 특정 서비스, 넓게는 국가의 미래를 좌우할 수 있는 요인으로 거론되기도 합니다. AI의 위상이 높아지면서, AI 알고리즘 윤리에 대한 관심 역시 커지고 있습니다. 알고리즘에 대한 윤리의식 확보는 해당 서비스뿐만 아니라, 이를 개발하고 운영하는 기업의 사회적 책무로도 간주되고 있습니다.

모두를 연결하여, 더 나은 세상을 만들고자 하는 카카오에 AI는 연결의 가치를 높여 줄 핵심 동력입니다. 카카오가 AI를 통해 만들고자 하는 일상은 누구나 더 편리하게 기술의 혜택을 누릴 수 있는 세상입니다. 삶의 편의를 한 단계 높일 AI가 사회 윤리의 범주 안에서 온당함을 유지하는 것을 카카오는 추구합니다. 이를 위해서는 알고리즘의 개발과 관리를 포함한 운영 일체에서 합의된 윤리 의식이 수반되어야 합니다.

카카오는 내부적으로 AI에 대한 윤리 의식을 정비하는 차원에서 크루들의 의견을 모아, 카카오 알고리즘 윤리 헌장을 마련했습니다. 이는 2017년 9월부터 시작된 연구, 내부 논의 작업을 거쳐 5개 안으로 구성된 카카오 알고리즘 윤리 헌장이 확정됐습니다.

기업이 알고리즘 윤리 원칙을 내부적으로 마련하고, 외부에 공표하는 것은 국내에서 처음 있는 일입니다. 카카오는 2017년 3월 인터넷 기업으로는 최초로 뉴스 배열의 알고리즘 일부를 학회 저널에 논문으로 공개했을 뿐만 아니라<sup>1)</sup>, 한국언론학회 정기학술대회에서 별도의 세션을 마련하여 미디어 담당 임원과 알고리즘 담당 TF장이 유관 알고리즘을 소개하기도 했습니다.<sup>2)</sup> 카카오는 ‘카카오 시리포트’를 통해, 지속적으로 진화된 뉴스 알고리즘을 공개하기도 하는 등 알고리즘 설명과 관련한 사회적 책무를 성심껏 수행해 왔습니다.<sup>3)</sup>

이번 장에 카카오 알고리즘 윤리 헌장이 논의된 배경, 그리고 윤리 헌장의 각 조항에 대한 의미를 구체적으로 설명한 글을 담았습니다.

## 카카오 알고리즘 윤리 헌장

1. 카카오 알고리즘의 기본 원칙	카카오는 알고리즘과 관련된 모든 노력을 우리 사회 윤리 안에서 다하며, 이를 통해 인류의 편익과 행복을 추구한다.
2. 차별에 대한 경계	알고리즘 결과에서 의도적인 사회적 차별이 일어나지 않도록 경계한다.
3. 학습 데이터 운영	알고리즘에 입력되는 학습 데이터를 사회 윤리에 근거하여 수집·분석·활용한다.
4. 알고리즘의 독립성	알고리즘이 누군가에 의해 자의적으로 훼손되거나 영향받는 일이 없도록 엄정하게 관리한다.
5. 알고리즘에 대한 설명	이용자와의 신뢰 관계를 위해 기업 경쟁력을 훼손하지 않는 범위 내에서 알고리즘에 대해 성실하게 설명한다.

2018년 01월 31일  
카카오

kakao

**논의 배경**

알고리즘에 대한 기대가 높아지면서 그로 인해 초래될 수 있는 역기능에 대한 막연한 공포심 역시 커지고 있습니다. AI의 빠른 진화는 머지 않은 미래에 인간을 뛰어 넘는 초지능(super intelligence)의 등장으로 이어져 AI가 인간 사회를 위협하는 존재가 될 수 있다는 우려가 제기되고 있습니다.

알고리즘을 기반으로 발전하는 AI와 로봇의 확산으로 맞게 될 새로운 시대의 윤리 규범 마련은 알고리즘을 개발하는 기업을 필두로 중요한 사회적 과제로 부상하고 있습니다.

윤리적 측면에서 알고리즘과 관련된 주요 논의 주제로는 크게 의도적 차별성, 데이터 수집 및 관리 측면의 윤리성 부재, 통제 불가능성, 그리고 불투명성을 들 수 있습니다. 전술한 논의 주제들은 아래와 같이 정리될 수 있습니다.

카카오는 우리 사회의 대표적인 AI 기술 기업으로서 사회적 책임과 시대에 걸맞은 내부 윤리적 기준을 확립하고자 이번 윤리 현장 수립 작업을 진행했습니다. 알고리즘 윤리 현장을 제정하고, 공표하는 카카오의 이번 결정이 알고리즘에 관한 사회적 규범을 형성하는 논의의 시발점이 되길 고대합니다.

\* 알고리즘 윤리 역사를 담은 콘텐츠를 볼 수 있는 QR코드입니다.



**윤리적 측면에서 알고리즘과 관련된 주요 논의 주제**

**1. 의도적 차별성**

알고리즘의 결괏값에 대해 알고리즘의 개발 및 운영 주체가 의도적으로 차별적인 결과가 나오도록 개입을 했는지에 대한 우려를 의미합니다. 이러한 차별성은 특정 가치 중심으로 편향된 결괏값으로 연결될 수 있다는 측면에서 편향성으로 거론되기도 합니다.

**2. 데이터 수집 및 관리 측면의 윤리성 부재**

알고리즘 성능에 직결된 요소 중 하나가 학습 데이터입니다. 데이터의 규모와 질은 알고리즘, 그것을 바탕으로 한 인공지능의 성능을 좌우합니다. 대개 알고리즘 학습에 쓰이는 데이터는 개인의 일상 기록이거나 그와 관련이 있는 자료입니다. 그렇기에 데이터 수집 및 관리 측면에서 높은 윤리 의식이 요구되고 있습니다.

**3. 통제 불가능성**

공상과학 소설이나 영화의 주된 소재는 인간 통제 범위를 벗어난 AI가 인류를 공격하는 모습입니다. 통제 불가능성은 만에 하나 있을지 모르는 인류의 행복이 침해될 수 있다는 우려로 이어집니다. 악의적 의도가 반영된 알고리즘이 통제 불가능 상태가 되어 인류에 해를 끼칠 수 있는 개연성에 사람들은 우려를 표합니다. 동일한 맥락 하에서 AI의 알고리즘 개발 및 운영의 전 과정이 관련 주체에 의해 통제되어야 한다는 의견이 제기되고 있습니다.

**4. 불투명성**

알고리즘의 원리와 작동 방식에 대한 설명의 필요성을 강조하는 차원에서 불투명성은 문제로 제기되고 있습니다. 설명 범위의 타당성에 대해서는 사회적으로 의견이 분분합니다. 그러나 상황 맥락을 이해할 수 있는 정보의 제공은 상품 혹은 서비스의 이용자에게 제공되어야 하는 조건으로 거론되고 있습니다.

**카카오 알고리즘 윤리 현장과 각 조항별 설명**

**1. 카카오 알고리즘의 기본 원칙**

카카오는 알고리즘과 관련된 모든 노력을 우리 사회 윤리 안에서 다하며, 이를 통해 인류의 편익과 행복을 추구한다. 카카오가 알고리즘 윤리 현장을 도입한 목적입니다. 카카오는 알고리즘 개발을 통해 카카오 서비스를 직·간접적으로 이용하는 사람들이 편익을 누리고, 보다 행복해지는 데 기여하고자 합니다. 알고리즘 개발 및 관리와 관련된 일련의 과정에서 카카오의 노력은 우리 사회의 윤리 원칙에 부합하는 방향으로 이뤄질 것입니다.

**2. 차별에 대한 경계**

알고리즘 결과에서 의도적인 사회적 차별이 일어나지 않도록 경계한다. 카카오는 다양한 가치가 공존하는 사회를 지향합니다. 카카오의 서비스로 구현된 알고리즘 결과가 특정 가치에 편향되거나 사회적인 차별을 강화하지 않도록 노력하겠습니다.

**3. 학습 데이터 운영**

알고리즘에 입력되는 학습 데이터를 사회 윤리에 근거하여 수집·분석·활용한다. 카카오는 알고리즘의 개발 및 성능 고도화, 품질 유지를 위한 데이터 수집, 관리 및 활용 등 전 과정을 우리 사회의 윤리를 벗어나지 않는 범위에서 수행하겠습니다.

**4. 알고리즘의 독립성**

알고리즘이 누군가에 의해 자의적으로 훼손되거나 영향받는 일이 없도록 엄정하게 관리한다. 카카오는 알고리즘이 특정 의도의 영향을 받아 훼손되거나 왜곡될 가능성을 차단하고 있습니다. 앞으로도 카카오는 알고리즘을 독립적이고 엄정하게 관리할 것입니다.

**5. 알고리즘에 대한 설명**

이용자와의 신뢰 관계를 위해 기업 경쟁력을 훼손하지 않는 범위 내에서 알고리즘에 대해 성실하게 설명한다. 카카오는 새로운 연결을 통해 더 편리하고 즐거워진 세상을 꿈꿉니다. 카카오 서비스는 사람과 사람, 사람과 기술을 한층 가깝게 연결함으로써 그 목표에 다가가고자 합니다. 카카오는 모든 연결에서 이용자와의 신뢰 관계를 소중하게 생각합니다. 이를 위해 더 나은 가치를 지속적으로 제공하는 기업으로서, 이용자와 성실하게 소통하겠습니다.

\*1 참고 | 박승택, 성인재, 서상원, 황지수, 노지성, 김대원 (2017). 기계학습 기반의 뉴스 추천 서비스 구조와 그 효과에 대한 고찰. 사이버커뮤니케이션학회, 34권 1호, 5-48. \*2 참고 | 해당 세션을 취재한 기사 URL: <http://v.media.daum.net/v/20170709182306874> \*3 참고 | <https://brunch.co.kr/@kakao-it/136>