

AI 자동 뉴스 추천 서비스 사용자가 인지하는 정보 편향성에 대한 연구: AI 서비스의 윤리 원칙 수립을 중심으로¹⁾

A Study on Information Bias Perceived by Users of AI-driven News
Recommendation Services: Focusing on the Establishment of Ethical
Principles for AI Services

박민정 (Minjung Park)

국립금오공과대학교 경영학과²⁾

채상미 (Sangmi Chai)

이화여자대학교 경영학부³⁾

〈 국문초록 〉

오늘날 대중화된 AI 뉴스 추천 시스템은 개인화된 뉴스 소비 경험을 제공하여 뉴스 등 다양한 콘텐츠를 사용자가 소비하는 과정에서 효율성을 향상시킨 반면에 특정 관점의 정보만을 제공하여 사용자의 정보 편향성을 심화할 수 있다는 우려가 제기되고 있다. 이는 사용자의 다양한 정보 접근을 제한하는 동시에 특정 사안에 대한 올바른 시각의 형성을 방해하여 필터 버블, 에코 챔버 등의 사회적 문제 현상을 강화하여 사회적 격차 및 정보 불균형을 심화시킬 수 있다. 이에 본 연구는 AI 기반 뉴스 추천 서비스 사용자가 인지하는 정보 편향성에 미치는 영향을 탐색하여 AI 서비스의 윤리 원칙 수립을 위한 기반을 제공하고자 한다. 이를 위하여 AI 뉴스 추천 시스템에 대하여 사용자가 기대하는 책무성, 설명요구권, 선택권, 프라이버시 보호 지원의 윤리적 원칙이 사용자가 인지하는 정보 편향성에 영향을 미치는지 확인하였다. 이를 통해 AI 제공기업의 윤리적 원칙 강화 필요성을 제기하는 동시에 서비스 품질의 향상을 통한 사용자의 신뢰 기반 지속적 사용을 촉진할 수 있을 것으로 예상된다. 또한, AI 서비스를 설계하는 과정에서 우선적으로 고려하여야 하는 윤리 원칙을 본 연구에서 확인함에 따라 본 연구의 결과는 기업의 윤리 원칙 프레임워크 및 내부 정책 수립 및 국내 AI 윤리 가이드라인 등 관련 정책 수립의 토대가 될 것으로 판단된다.

주제어: 인공지능, AI 윤리 원칙, AI 뉴스 추천 시스템, 정보 편향, 지식경영

1) 이 논문은 2024년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2024S1A5C3A02043653)

2) 제1저자, mjpark@kumoh.ac.kr

3) 교신저자, smchai@ewha.ac.kr

1. 서론

정보통신기술의 발달과 함께 AI 기반 뉴스 추천 서비스는 오늘날과 같이 방대한 양의 정보가 실시간 생성되고 확산되는 정보 과잉 시대에서 필수적인 도구로 주목받고 있다(Zhou & Chen, 2023). AI 뉴스 추천 시스템은 방대한 양의 데이터를 분석하여 사용자의 관심사와 선호도에 맞춘 뉴스를 제공함으로써, 개인화 기반의 향상된 뉴스 소비 경험을 형성한다(Cheng et al., 2021). 구체적으로, 현재 주요 포털 사이트들은 알고리즘을 활용하여 뉴스 추천과 배열 방식을 채택하고 있다. 네이버는 2019년부터 AI 뉴스 추천 시스템인 ‘에어스(AiRS, AI Recommender System)’를 도입하여 뉴스 추천과 편집을 모두 AI가 수행하는 자동화 방식에 의존하고 있다. 또한, 구글은 추천 알고리즘을 통해 이용자의 관심 분야, 선호 매체, 구글 검색 및 유튜브 활동 등을 반영하여 맞춤형 뉴스를 제공한다. 이러한 뉴스 추천 알고리즘의 목적은 이용자의 이용 패턴과 사용 흔적을 분석하여 선호도를 실시간으로 예측하고, 이를 바탕으로 매일 수만 건의 기사 중 이용자가 선호하거나 사용자에게 필요할 것으로 예상되는 기사를 자동으로 배열하는 것이다. 그러나 추천 알고리즘의 개인화 경험은 사용자의 최적화된 소비를 촉진하고 효율성을 높이는 장점이 있는 반면에 서비스 작동 방식과 운영 체계의 투명성이 부족하고, 특정 관점에 치우친 정보를 제공할 수 있다는 점에서 비판 받고 있다(Liu & Tao, 2022). 또한, 사용자의 선호와 관심 분야만을 고려한 맞춤 정보 제공 및 사용자의 피드에 노출시킴에 따라, 정보의 다양성이 결여된 폐쇄적 환경을 조성하고 이용자의 정보 편식을 초래하는 ‘필터 버블(Filter Bubble)’을 유발한다는 지적이 존재한다(Arceeb et al., 2023). 즉, 오늘날의 뉴스 추천 시스템은 맞춤화 콘텐츠를 사용자에게 제공하여 정보 소비의 효율성을 증진시킨다는 본래 목적과 다르게 제한된 관점에서의 편향된 정보만을 제시하여 다양한 정보의

접근성을 제한하는 부작용이 나타나고 있다. 이와 같이 AI 기반의 추천 시스템은 사용자의 정보 편향성을 심화하는 문제를 형성함에 따라, 본 연구는 AI 기반 뉴스 추천 서비스가 사용자가 인지하는 정보 편향성에 미치는 영향을 탐구하고, 이를 개선하기 위한 AI 서비스의 윤리적 원칙을 수립을 위한 기반을 제공하고자 한다. 이를 위하여 본 연구는 AI 윤리 원칙으로 제시된 다양한 항목에 대한 선행 연구의 분석을 통하여 해당 요인들이 사용자의 정보 편향에 미치는 영향을 탐색한다. 특히, AI 서비스 제공기업은 채택한 AI 모델을 비롯하여 관련 서비스가 공공의 이익에 반하거나 사용자의 가치를 훼손할 위험성이 존재하는 경우 이를 신속하게 감지하여 이를 해결할 책임 의무를 다하여야 한다는 책무성과 사용자는 제공한 데이터에 대하여 기업에게 적절한 삭제 및 수정을 요청하거나 원치 않는 혹은 부정확한 정보가 사용자의 피드에 노출되지 않도록 사용 범위의 조정을 요구할 수 있는 선택권 등을 중심으로 살펴본다. 나아가, 제공된 서비스의 기반이 되는 AI 모델의 구현 및 운영 원리와 학습 데이터의 특징 등에 대하여 사용자에게 기업은 충분히 설명하여야 서비스 운영의 불투명성 감소를 위한 기업의 노력인 설명의 의무와 사용자의 프라이버시가 침해되지 않도록 적절한 프라이버시 보호 조치를 수행하여야 하는 기업 의무를 함께 고려한다.

본 연구의 결과는 AI 서비스 제공 기업이 윤리 원칙 강화를 촉진하여 사용자 신뢰를 구축하고 서비스 품질을 향상시키기 위한 기반을 제공할 것으로 예상된다. 또한, 본 연구는 기업이 AI 서비스를 설계하고 운영할 때 고려해야 할 윤리적 원칙을 명확히 제시하여, 기업의 윤리적 프레임워크 수립에 기여하는 동시에 나아가 국내 AI 윤리 가이드라인 등의 관련 정책 마련에 중요한 기초 자료를 제공할 수 있다. 이를 통해 AI 기술의 신뢰 향상과 더불어 AI 기술이 사회적으로 보다 공정하고 투명하게 활용될 수 있는 토대를 형성할 수 있을 것으로 판단된다.

2. 선행 연구

2.1. AI 기반 뉴스 추천 서비스

AI 기반 뉴스 추천 시스템은 오늘날과 같은 정보 과잉의 시대에서 빠르게 사용자에게 맞춤형 뉴스를 제공하기 위한 필수적인 도구로 부상하고 있다. IT의 발달에 따라, 사진, 음성, 동영상, 문서 등의 다양한 유형의 데이터를 비롯하여 방대한 양의 데이터가 기하급수적으로 누적된 오늘날과 같은 빅데이터 시대의 형성은 방대한 양의 정보 속에서 개인이 필요한 뉴스를 찾는 데 어려움을 경험하도록 하였다. 이러한 환경에서 AI 뉴스 추천 시스템은 뉴스 수용자가 필요한 가치 있는 정보를 쉽게 탐색하여 빠르게 해당 정보에 접근할 수 있도록 하기 위한 목적으로 시작되었다(박승택 등, 2017). 즉, AI 뉴스 추천 서비스는 인공지능 기술을 활용하여 사용자에게 맞춤형 뉴스 콘텐츠를 제공하는 시스템을 의미한다. 이러한 AI 뉴스 추천 서비스는 사용자 데이터와 인공지능을 활용하여 사용자의 관심사와 선호도를 파악하고, 이를 기반으로 사용자에게 맞춤형 뉴스 추천 시스템을 제공한다(Feng et al., 2020; Raza & Ding, 2022).

자동화된 뉴스 추천 시스템이 도입된 초기에는 뉴스 콘텐츠의 조회수와 같은 대중성, 편집자의 선택 등 비개인화의 요인만을 고려하여 뉴스를 추천하였다. 반면에 최근 확산되어 대중화된 개인화된 뉴스 추천 시스템은 개별 사용자의 개인적인 관심사를 고려하여 개인 맞춤형 뉴스 서비스를 제공한다는 점에서 사용자 요구를 더 잘 충족시키는 것으로 밝혀지며 이의 필요성과 수요가 급증하였다(Wu et al., 2023a). 이러한 개인화 기반의 맞춤 뉴스 추천 시스템은 인공지능 모델의 발전과 함께 빠르게 발전하였다. 특히, 머신러닝, 딥러닝, 자연어 처리(NLP) 등의 AI 기술이 급격히 발전하면서, 뉴스 추천 시스템의 성능이 크게 향상되며 사용자에게 보다 맞춤형된 정교한

추천이 가능해졌다. 이는 과거 기존의 개인화된 뉴스 추천은 비교적 단순한 모델인 협업 필터링 기반, 콘텐츠 기반, 하이브리드 방식의 세 가지 범주로 구현되었다. 그러나 이와 같은 분류 방식은 고정된 사용자와 뉴스 데이터셋을 기반으로 작동하는 경우가 많음에 따라, 새로운 뉴스 기사가 지속적으로 생성되고 사용자의 댓글 등이 실시간으로 기록되는 매우 동적인 최근의 뉴스 제공 플랫폼의 특징을 모두 반영할 수 없었다. 또한, 컨텍스트 밴딧(Contextual Bandit) 방법론이나 상황 인식 기반 추천 시스템(Context-Aware Recommendation Systems)이 등장하면서, 단순한 콘텐츠 기반 필터링이나 협업 필터링으로는 이러한 발전을 포괄하기 어렵게 되었다(Zhu & Van Roy, 2023). 최근의 뉴스 추천 시스템은 단순히 사용자의 과거 행동만을 고려하는 것이 아니라, 사용자의 현재 컨텍스트, 위치, 시간 등 다양한 요인을 고려하여 더 정교한 개인화 추천 제공의 필요를 증가하게 하였다. 이를 처리하기 위하여 기존의 추천 방식에서 벗어나 실시간 데이터 처리 기술을 비롯하여 보다 정교한 맞춤화의 뉴스를 추천하기 위한 개선된 인공지능 모델이 활용되기 시작하였다. Aboutorab et al. (2023)에 따르면, AI 뉴스 추천 시스템은 강화학습을 기반으로 사용자의 피드백을 통하여 지속적으로 학습하고 개선된다. 구체적으로, 사용자가 클릭하거나 읽은 뉴스, 좋아요 및 공유 등을 표시한 사용자의 온라인 행동 데이터를 바탕으로 학습하여 뉴스 추천 시스템은 더 나은 추천을 제공하며 시스템의 정확성과 사용자 만족도를 지속적으로 높이는 데 기여한다. 또한, 딥러닝의 발전은 다양한 모델을 활용하여 다양한 사용자의 데이터를 바탕으로 보다 정교한 뉴스 추천 시스템을 구현하였다. 대표적으로, 뉴스 기사의 이미지 데이터를 분석하여 사용자 추천에 반영하기 위하여 이미지 분류에 특화된 컨볼루션 신경망(Convolutional Neural Networks, CNN)을 활용하거나(Yu et al., 2018), 사용자의 클릭스트림 등의 시계열 데이터를 모델링하기 위하여

순환 신경망(Recurrent Neural Networks, RNN) 기반의 추천 시스템이 구성되었다(Amir et al., 2023). 이외에도 CNN, RNN 및 주의 메커니즘(Attention Mechanism)을 하이브리드 형태로 활용하는 DAN(Deep Attention Neural Network) 방식의 뉴스 추천 시스템을 제시하여 단일 모델 사용 대비 사용자 중심의 개인화된 보다 정확한 콘텐츠 추천이 이루어짐을 확인하였다.

AI 뉴스 추천 서비스는 사용자의 정보 과부하를 완화하고 보다 정교하게 맞춤형 개인 뉴스 콘텐츠를 쉽게 제공하는 장점에도 불구하고 최근 들어 다음과 같은 문제점을 발생시켰다. AI 뉴스 추천 서비스는 편향된 알고리즘(Biased Algorithms)의 위험을 수반한다(Cheng et al., 2021). AI 뉴스 추천 시스템은 학습 데이터에 따라 편향될 수 있으며, 이는 특정 주제나 관점을 과도하게 강조하거나 다른 주제나 관점을 간과하여 사용자에게 뉴스를 추천해줄 수 있다. 이러한 편향은 사용자의 인식과 판단에 영향을 미쳐 정보의 공정성과 균형을 해친다. 이러한 편향된 알고리즘은 AI 기반 뉴스 추천 서비스의 확산에 따른 필터 버블 및 에코 챔버의 문제를 생성한다(Jeon et al., 2024). 필터 버블은 사용자가 선호하는 특정 유형의 콘텐츠만을 지속적으로 제공하는 AI 추천 모델의 특징이다. 필터 버블은 사용자가 경험하는 온라인 환경의 정보의 다양성을 감소시킴에 따라, 사용자의 다양한 관점 및 의견 수용의 기회를 제한하게 된다(Zhou & Chen, 2023). AI 뉴스 추천 시스템은 사용자가 과거 열람한 뉴스 기록, 공유, 좋아요 등 뉴스 콘텐츠에 제공한 사용자의 행동 데이터 등을 바탕으로 뉴스를 지속적으로 추천한다는 점에서 과거 수집된 데이터와 유사한 내용의 뉴스를 주로 제공하게 되어 필터 버블의 형성을 조장한다(Raza & Ding, 2022). 또한, 비슷한 생각을 가진 사람들이 서로의 의견을 강화하면서 다른 관점은 배제되는 현상인 에코 챔버 현상을 AI 뉴스 추천 서비스는 강화할 수 있다. 즉, AI 뉴스 추천 시스템은 사용자가 이미 동의

하는 콘텐츠를 반복적으로 추천함으로써 이러한 현상을 더욱 심화할 수 있다(Areeb et al., 2023). 이는 특정 그룹 내에서만 공유되고 검증되지 않은 정보가 사실처럼 받아들여지도록 하여 잘못된 정보의 수용을 조장한다. 이에 본 연구에서는 AI 기반 뉴스 추천 서비스의 사용자가 인지하는 정보 편향성에 영향을 미치는 요인을 밝히어 이의 관계를 살펴보고자 한다.

2.2. AI 서비스 사용자의 인지된 정보 편향성

본 연구에서는 인지된 정보 편향성을 사용자가 AI 기반의 뉴스 추천 서비스를 사용함에 따라 제공되는 정보가 특정 방향으로 왜곡되거나 불공정하게 제시되었다고 사용자가 인지하는 정도로 정의한다. 따라서, AI 서비스 사용자가 인지하는 정보 편향성은 제공받은 정보 공정성의 균형이 사실여부와 관계없이 훼손되었다고 믿는 것이다(Van Swol, 2007; Xiao & Benbasat, 2015). 구체적으로 사용자가 특정 정치적 주제에 대해 뉴스를 검색하거나 추천 받을 때, AI 시스템이 특정 정당이나 특정 정치적 입장을 지지하는 기사만을 주로 추천해주는 경우, 사용자는 AI 시스템이 동일한 정치적 사안에 대하여 편향된 정보를 제공하는 것으로 인지할 수 있다(Papakyriakopoulos et al., 2020). 이러한 뉴스 추천의 과정은 실제로 학습 데이터의 불균형, 사용자가 과거 클릭한 뉴스의 종류 등 편향성을 형성하는 원인이 수반되는 경우도 추론할 수 있지만 단순히 해당 시간대의 인기 뉴스 기사를 지속적으로 사용자에게 제시하였을 가능성도 존재한다. 따라서, AI 뉴스 추천 서비스를 사용하는 사용자는 실제 제공된 정보의 편향성을 개인이 검증하는 것도 어려움과 동시에 사실여부와 상관없이 정보의 공정성에 대한 의구심을 자연스럽게 형성하게 된다.

최근 들어, AI 모델이 방대한 양의 데이터를 학습하여 예측, 추론 등을 수행하나 해당 학습 과정에서 데이터가

편향되어 있을 경우, AI 모델의 학습 결과 또한 편향성을 드러냄이 빈번하게 밝혀 짐에 따라, 제공되는 정보에 대하여 사용자가 인지하는 편향성을 살펴볼 필요성이 증가하였다. 대표적으로, Amazon은 채용 과정에서 AI 채용 프로그램을 도입할 결과, 남성 지원자가 여성 지원자보다 상대적으로 높은 채용 점수를 받는 빈도가 높았다. 이에 따라, 채용에서 여성 지원자 대비 남성 지원자의 채용 비율이 지속적으로 높게 나타났으며 이는 AI 채용 프로그램의 구축 당시 사용된 데이터의 편향성에서 기인한 것으로 밝혀졌다(한국일보, 2021). 즉, 해당 AI 채용 프로그램의 구축 과정에서 사용된 데이터의 70% 이상이 남성 직원들 관련 데이터로 구성됨에 따라, 남성 지원자의 공통적인 역량을 지원자에게 요구되는 우수한 역량으로 오평가 하여 편향된 결과를 발생시켰다. 이러한 사례가 증가함에 따라, 최근 들어 AI 서비스의 사용으로 발생 가능한 편향의 원인을 살펴보기 위한 연구가 활발히 진행되기 시작하였으며 이의 원인은 다음과 같이 요약될 수 있다. 첫째, 데이터 자체의 불균형으로 인해 발생하는 데이터 편향이다. 데이터셋에서 특정 집단이나 특정 영역이 과도하게 포함될 경우, 이 집단에 유리한 결과가 나올 확률이 높아지는 AI 편향 현상이 발생한다(DeCamp & Lindvall, 2023; Kumar et al., 2024; Nazer et al., 2023). 둘째, 알고리즘 자체의 설계나 작동 방식에서 비롯되는 편향이다. 이는 모델이 모든 데이터의 특징을 균등하게 반영하지 않거나, 알고리즘이 특정 패턴을 선호하는 경우에 발생하는 구조적 편향을 의미한다(DeCamp & Lindvall, 2023; Kumar et al., 2024). 셋째, 시스템이 배포되거나 사용되는 환경에서 발생하는 편향이다. 이로 인해 모델이 특정 환경에서는 정확하게 작동하지만 다른 환경에서는 부정확한 결과를 초래하는 과적합 현상과 유사한 환경적 문맥에 따른 편향이 발생한다(Nazer et al., 2023). 따라서, AI 편향은 데이터 불균형, 알고리즘 설계, 사용자와 시스템 간의 상호작용 복잡성에서 기인한다.

특히, 데이터셋의 불균형으로 인한 정보 과다, 특정 집단에 대한 편견, 알고리즘 학습 과정에서의 잠재적 편향성은 AI 정확성을 저해하는 주요 요인으로, 이러한 문제를 해결할 필요성이 증가하고 있다.

AI로부터 제공받은 정보에 대하여 사용자가 인지하는 정보 편향성에 따른 위험의 결과는 다음과 같다. 먼저, 사용자 개인의 차원에서 해당 정보가 실제로 왜곡되었을 경우, 사용자가 다양한 시각을 접하지 못하고 편향된 정보를 바탕으로 결정을 내리도록 유인한다. 이는 보다 빠르고 간편하게 정확한 정보를 제공한다는 AI 추천 서비스의 본래 목적을 훼손한다(Shani & Gunawardana, 2011). 앞서 제시된 필터 버블, 에코 챔버 현상을 조성하며 결국 사용자가 데이터를 기반으로 합리적인 의사결정을 내리지 못하도록 한다. 이러한 편향된 정보를 바탕으로 이루어진 사용자의 정보 수용은 특정 집단 등의 의견이 지속적으로 온라인 환경에서 의도적으로 배제되거나 강요될 수 있어 사회적 불평 등 각종 사회적 문제를 초래한다(Gao & Shah, 2021; Konow, 2005). 나아가, 사용자가 제공받은 정보가 왜곡되어 편향성을 가지는지 여부와 상관없이 AI 시스템에 대한 사용자의 신뢰를 저하시켜, 장기적인 관점에서 AI 기술의 활용에 부정적인 영향을 미친다(Branley-Bell et al., 2020). 이에 본 연구에서는 AI 모델의 관점에서 형성되는 편향의 원인을 살펴보는 기존의 선행 연구와 다르게 사용자가 인지하는 정보의 편향성에 영향을 미치는 원인을 사용자의 관점에서 밝히고자 한다. 이는 앞서 밝혀진 AI 기술이 수반하는 편향성을 AI 모델 자체의 관점에서 원인을 탐색하는 선행 연구와는 다르다는 점에서 의의가 있을 것으로 예상된다.

2.3. AI 서비스의 윤리 원칙

최근 AI 기술의 발전과 확산으로 인해 AI 서비스의 윤리적 원칙에 대한 관심이 증가하고 있다. 즉, AI가 데이

터 기반 예측, 자동화, 개인화 등의 다양한 기능을 제공하며 조직 내 업무 생산성 및 개인 삶의 질의 향상 등 긍정적인 역할을 수행하는 반면에 이와 동시에 개인정보 유출, 일자리 상실 등 부정적인 영향이 증가하고 있다(Mukherjee, 2022). 따라서, AI의 기술이 광범위하게 사용되면서 그 영향력이 커지는 만큼 공공의 이익을 훼손할 위험이 증가함에 따라, 이를 예방하고자 AI 기반의 서비스를 제공하는 기업 및 AI 모델을 개발하는 연구자, 개발자 등 모두가 준수하여야 하는 ‘AI 윤리 원칙’의 수립의 필요성이 제기되는 추세이다(Borenstein & Howard, 2021; Jobin et al., 2019). AI의 윤리 원칙은, AI를 기술의 개발, 배포, 및 사용 과정에서 발생 가능한 윤리적 문제를 사전에 예방하기 위하여 AI 연구자, 개발자를 비롯하여 모든 관련 참여자가 각 단계에서 준수하여야 하는 원칙으로 설명될 수 있다(Jobin et al., 2019).

AI 윤리 원칙은 AI가 인간과 사회에 긍정적인 영향을 미치도록 보장하며, 잠재적인 위험과 부정적인 영향을 최소화하는 것에 주요 목표가 있다(Zhou & Chen, 2023). 따라서, AI 윤리 원칙은 기술 윤리의 관점과 더불어 사용자의 심리, 기업 지배 구조, 사회 문화 특징, 철학적 사고, 법적 측면의 다양한 관점을 모두 고려하는 동시에 수집, 모델 구축, 시스템 배포, 성능 모니터링 등 AI 생애주기 전반에 걸쳐 적용될 수 있도록 설계되어야 한다(Jobin et al., 2019).

AI 윤리 원칙 수립의 필요성이 증가하며 관련 선행 연구가 활발하게 수행되고 있으며 주요 원칙은 다음과 같다. 첫째, AI 모델을 개발하는 연구자와 개발자는 AI 알고리즘이 수행하는 예측 등의 결정이 어떠한 방식으로 이루어지는지 혹은 어떠한 모델이 해당 결정을 내리기 위하여 사용되었는지 등을 명확히 설명할 수 있도록 알고리즘을 설계하여야 한다는 ‘설명 가능성(explainability)’의 원칙이다(Vainio-Pekka et al., 2023a). 이는 최근 활발하게 연구되는 설명가능한 AI 모델(Explainable AI, XAI)

의 기반이 된다. 둘째, 데이터 수집과 모델 훈련 과정에서 편향을 식별하고 이를 제거하도록 하여 AI 모델의 공정성 확보를 위한 노력 또한 연구자와 개발자가 준수하여야 하는 윤리 원칙으로 고려될 수 있다(Giovanola & Tiribelli, 2023). 셋째, AI 시스템의 잠재적 위험을 평가하고, 안전한 사용을 보장하기 위한 방안을 마련하는 동시에 시스템의 오류를 신속히 수정하고, 안전하게 작동하도록 유지관리 하는 AI의 안전성 확보 노력이 요구된다(Leslie, 2019; Shneiderman, 2020). 넷째, AI 시스템의 설계와 구현 단계에서 발생 가능한 문제를 사전에 식별하고자 하며, 문제 발생 시 이에 따른 책임을 명확히 정의하여 해결 절차를 수립하는 과정 또한 AI 연구자 및 개발자가 반영하여야 하는 사항이다(Hallamaa & Kalliokoski, 2022). 이와 동시에 개발된 AI 모델을 바탕으로 다양한 서비스를 제공하는 기업 등의 경우 AI 서비스의 작동 원리 등을 사용자에게 충분히 설명하여야 하는 동시에 서비스가 어떤 데이터와 알고리즘으로 제공되는지를 공개하여 AI 서비스 운영의 투명성을 보장하여야 한다(Larsson & Heintz, 2020). 또한, AI가 공공의 이익을 저해할 가능성이 있을 때, 해당 서비스를 제공한 기업은 이를 감지하고 신속하게 수정할 책임 의무를 다하여야 한다는 책무성(Accountability)이 존재한다(Lechterman, 2022; Memarian & Doleck, 2023). AI 시스템의 책무성을 강화하기 위하여 AI 서비스 제공 환경의 투명성 보장과 감시 메커니즘의 중요성이 함께 증가하고 있다(Binns et al., 2018; Doshi-Velez & Kim, 2017). 이외에도 명확한 개인정보보호 정책 등을 수립하여 사용자에게 필요 시, 적법한 개인정보 수집 동의 등 AI 서비스 사용 프로세스의 안정성을 확보하는 동시에 사용자의 프라이버시 보호를 위한 활동에 힘써야 함을 준수 원칙으로 포함하고 있다(Stahl & Wright, 2018).

앞서 제시된 원칙 등을 바탕으로 본 연구는 AI 뉴스 추천 서비스를 제공하는 기업이 준수하여야 하는 윤리

원칙을 다음과 같이 구성하였다. 첫째, AI 기반의 뉴스 추천 시스템을 운영하는 기업은 시스템 운영 과정에서 발생하는 기술적 오류 및 제공되는 결과가 사회 질서 유지를 방해하는 경우 이를 해결하기 위한 의무인 ‘책무성’이다. 책무성은 AI 서비스의 이용에 따라, 발생하는 문제점과 더불어 서비스 설계 과정에서 각종 법률과 지침을 준수하여 적절한 서비스를 제공하도록 하여야 하는 의무이다. 둘째, AI 서비스의 알고리즘이 선별한 콘텐츠 기준과 개인정보 활용 방법에 대해 사용자에게 명확하게 설명해야 할 의무인 ‘설명 요구권(Explanation Right)’이다. 이는 사용자가 AI 시스템의 작동 원리를 이해하고, 그 결정이 어떻게 이루어졌는지 알 수 있도록 하기 위함이다. 특히, 맞춤형 정보 제공을 위해 개인정보가 어떻게 활용되는지를 명확히 설명하는 것은 사용자가 자신의 데이터가 어떻게 처리되는지를 이해하고 통제할 수 있는 기반을 조성한다는 점에서 AI 서비스 제공자가 반드시 이행하여야 하는 항목이다. 이러한 나아가, AI 서비스의 구현 및 운영 방식에 대한 서비스 제공 기업의 정보 제공은 AI의 투명성을 높이는 동시에 사용자의 신뢰를 증진시키는 데 필수적임에 따라, 사용자의 지속적인 서비스를 장려할 수 있다는 점에서 AI 서비스 제공 기업이 반드시 고려하여야 한다(Gilpin et al., 2018). 셋째, ‘선택권’은 사용자가 AI 서비스 이용 정보에 대하여 수정 및 삭제를 요구할 권리를 행사할 시, 서비스 제공자는 이를 보장해 주어야 한다는 것이다. 이는 사용자가 자신의 데이터에 대해 통제권을 갖고, 불필요하거나 부정확한 정보를 제거할 수 있는 권리를 보장하는 것으로 GDPR(General Data Protection Regulation)등의 법적 규제에서 명시적으로 보호하고 있는 권리이다(Voigt & Von dem Bussche, 2017). 또한, 사용자가 원치 않는 뉴스, 광고 등의 콘텐츠로부터 사용자를 보호할 수 있도록 각종 콘텐츠의 노출 정도를 스스로 변경 및 선택할 수 있도록 보장하는 것이다. 마지막으로, ‘프라이버시 보호’는 AI 시스템이 사용

자의 개인정보를 보호하기 위한 기술적, 관리적 조치를 적법하게 이행하여 서비스 이용 과정에서 사용자의 프라이버시가 훼손되지 않도록 보장하여야 함을 의미한다. 사용자의 프라이버시를 보장하기 위하여 데이터 수집, 저장, 처리, 전송 등 모든 과정에서 개인정보가 적절히 보호될 수 있도록 서비스의 설계부터 사용자의 서비스 이용의 전 단계에 적합한 보호조치가 이루어져야 한다. 특히 AI 시스템의 설계 단계에서부터 프라이버시 보호를 고려하는 프라이버시 중심 설계(Privacy by Design)의 중요성이 강조되고 있다(Yanisky-Ravid & Hallisey, 2018).

본 연구에서는 AI 기반 뉴스 추천 서비스 기업의 윤리적 원칙 항목을 책무성, 선택권, 설명 요구권, 프라이버시 보호 의무의 네 가지를 주요 요소를 선정하였다. 이러한 원칙들은 AI 기술이 인간 중심적이고 신뢰할 수 있는 방향으로 발전하기 위해 필수적인 기준을 제공할 것으로 판단된다. 이에 따라, 본 연구에서는 제시된 윤리 원칙을 기반으로 AI 뉴스 추천 서비스 기업의 원칙 준수 및 이행 여부 정도에 대한 사용자의 인식이 제공된 정보에 대하여 사용자가 인지하는 정보 편향성과의 관계를 살펴보고자 한다.

3. 연구모델 및 가설수립

3.1. 뉴스 추천 방식과 사용자의 인지된 정보 편향성의 관계

온라인에서 생성되는 방대한 양의 뉴스 콘텐츠 중에서 사용자가 특정 콘텐츠에 접근하여 해당 뉴스를 열람할 수 있도록 유인하는 다양한 방식이 존재한다. 최근에는 네이버, 구글 등을 비롯하여 뉴스 콘텐츠를 제공하는 대부분의 포털 사이트를 비롯하여 각 언론사의 공식 웹사이트 등 모두 AI 기반의 뉴스 추천 시스템을 도입하여

사용자에게 맞춤형 된 뉴스 콘텐츠를 보다 정확히 전달하고자 하는 추세이다. 따라서, 이러한 자동 추천 시스템을 통하여 사용자는 뉴스 콘텐츠를 추천 받아 피드를 구성하는 것이 매우 자연스러운 환경이 됨에 따라, 이러한 추천 방식을 바탕으로 뉴스를 소비하는 사용자 비율이 증가하고 있다(Liu et al., 2022). 이와 더불어 최근에는 국내 카카오톡 메신저를 비롯하여 다양한 메신저 어플리케이션을 대부분 사용함에 따라, 이를 활용한 뉴스 콘텐츠 링크의 공유, 추천 등이 실시간으로 쉽게 이루어지고 있다. 따라서, 오늘날 사용자가 특정 뉴스 콘텐츠에 접근이 이루어지는 방식은 개인이 과거 열람한 콘텐츠 기록을 바탕으로 추천이 이루어지는 자동화 추천 방식과 전문가 및 지인에 의한 추천 방식으로 분류할 수 있다.

AI 기반 자동 추천 기능은 알고리즘을 통해 개인의 선호도와 과거의 행동 데이터를 분석하여 뉴스를 추천하는 방식이다(Wu et al., 2023a). 이러한 추천 방식은 사용자의 취향과 관심사를 반영하여 뉴스를 추천함으로써, 사용자는 자신이 관심 있는 주제에 대해 보다 더 맞춤형된 많은 정보를 쉽게 확보할 수 있다(Raza & Ding, 2022). 그러나 알고리즘이 사용자의 기존 선호도를 강화할 경우, 필터 버블 현상이 발생할 수 있으며 이는 사용자가 다양한 관점을 접할 기회를 줄이며, 결과적으로 정보 편향성을 강화하는 결과를 초래한다(Areeb et al., 2023).

자동 추천 알고리즘이 상용화되기 이전에는 사용자가 온라인 환경에서 뉴스 등의 콘텐츠에 접근할 수 있는 대표적인 방법은 개인이 신뢰하는 사람 혹은 집단으로부터 뉴스를 추천 받는 것이다. 이러한 방식은 사용자가 자신의 가치관과 일치하는 전문가나 지인의 추천을 신뢰할 때, 주로 채택하는 방식으로 사용자가 정보의 출처에 대한 신뢰도를 높이는 결정적인 요인으로 알려져 있다(Lin et al., 2012; Lin et al., 2014). 따라서, 지인이나 전문가의 추천은 사회적 증명(social proof)을 제공하여, 사용자가 정보를 더 쉽게 수용하게 만드는 경향이 있다(Gena et

al., 2019). 이는 정보의 신뢰성을 높이고, 추천된 뉴스에 대한 사용자의 신뢰를 강화한다. 그럼에도 불구하고 이러한 지인에 의존한 콘텐츠 추천 방식은 동일한 집단 내에서 추천이 주로 이루어짐에 따라, 정보의 다양성을 감소시키며 이는 특정 가치관이나 이념에 편향된 정보 접근의 가능성을 높인다(Wu et al., 2023b).

본 연구에서는 뉴스 추천 방식을 AI 기반의 자동 추천 방식과 전문가 및 지인에 의한 추천 방식으로 분류하여 특정 추천 방식을 더 선호하는 사용자의 집단에 따라 인지된 편향성에 차이가 있는지를 확인하고자 한다. 이를 통하여 사용자가 특정 추천 방식을 선호하게 된 배경에는 제공된 뉴스가 특정 관점으로 치우치거나 편향된 내용을 포함하고 있지 않은 공정할 것이라는 신뢰에서 기인하였는지를 살펴보기 위함이다. 이에 본 연구는 AI 기반 자동 추천 방식과 전문가 및 지인에 따른 사용자의 선호 뉴스 추천 방식에 따라 인지된 정보 편향성의 차이가 있을 것이라는 다음의 가설을 제기한다.

H1. 사용자가 선호하는 뉴스 추천 방식에 따라 AI 추천 뉴스 기사에 대한 인지된 정보 편향성은 차이가 있을 것이다.

3.2. 사용자의 정보 리터러시와 인지된 정보 편향성의 관계

정보 리터러시는 개인이 정보의 신뢰성, 정확성, 편향성을 평가하고 비판적으로 분석할 수 있는 능력을 의미한다(Heiss et al., 2023). 최근 몇 년간 정보 과잉 시대에 접어들면서, 정보 리터러시의 중요성은 더욱 부각되고 있다. 정보 리터러시가 높은 사람들은 다양한 정보 출처를 비교하고 분석하여 편향되지 않은 정보를 취사선택할 수 있는 반면에 정보 리터러시가 낮은 사람들은 특정 출처나 특정 관점에 치우친 정보를 무비판적으로 수용할 가능성이 크다. 이는 정보 리터러시가 상대적으로 낮은 집단이 그렇지 않은 사용자의 집단에 비하여 가짜뉴스 수용의 비율이 높음을 통하여 확인되었다(Haque et al.,

2024; Sureephong et al., 2023). 또한, 다양한 선행 연구들은 정보 리터러시가 개인의 정보 수용 태도와 행위에 미치는 영향을 여러 측면에서 분석해왔다. 예를 들어, 정보 리터러시가 높은 사람들은 뉴스 기사에서 정보의 편향성을 더 잘 인지하며, 이러한 인식은 그들이 특정 뉴스 기사의 신뢰성을 평가하는 데 중요한 역할을 한다고 밝혔다(김형지 등, 2020). 또한, 정보 리터러시가 낮은 사람들은 미디어에서 제공하는 정보를 비판적으로 수용하지 않기 때문에, 정보의 편향성에 대하여 상대적으로 낮게 지각하는 것으로 나타났다(염정윤, 정세훈, 2019).

본 연구에서는 정보 리터러시가 AI가 추천하는 뉴스 기사에 대한 인지된 정보 편향성의 인지의 차이에 영향을 미치는지를 살펴보고자 한다. 이를 위하여 본 연구는 “다른 자료들과 비교해 검색 결과에서 믿을만한 정보를 구별할 수 있다.”, “찾은 정보의 출처를 확인할 수 있다.”, “광고성 정보를 구별해 낼 수 있다.”, “사기 또는 속이러는 사이트를 가려낼 수 있다.”의 정보 리터러시를 측정하기 위한 문항을 바탕으로 사용자를 두 개의 집단으로 분류하였다. 즉, 사용자 스스로 본인의 정보 리터러시 역량을 평가하여 응답자의 평균 점수를 기준으로 상대적으로 정보 리터러시가 높은 집단과 그렇지 않은 집단으로 구분하여 정보 리터러시에 따른 정보 편향성을 확인하고자 한다. 이는 선행연구를 통해 확인된 바와 같이, 정보 리터러시가 높은 사람들은 뉴스 기사의 편향성을 더 민감하게 인식할 가능성이 높으며, 이는 AI 추천 시스템이 제공하는 뉴스 기사에서도 동일하게 나타날 것으로 예상된다. 이에 본 연구는 정보 리터러시가 높은 집단과 낮은 집단 간에 AI 추천 뉴스 기사에 대한 인지된 정보 편향성의 통계적 차이를 확인하기 위하여 다음과 같은 가설을 제기한다.

H2. 사용자의 정보 리터러시에 따라 AI 추천 뉴스 기사에 대한 인지된 정보 편향성은 차이가 있을 것이다.

3.3. AI 뉴스 추천서비스의 책무성에 대한 사용자의 기대와 인지된 정보 편향성의 관계

AI 기반 뉴스 추천 서비스 기업의 책무성은 사용자들에게 제공되는 정보의 품질과 공정성을 보장하기 위해 필수적이다. 책무성은 기업이 그들의 행동과 결정에 대해 책임을 지고, 이에 대해 투명하고 공정하게 설명할 수 있어야 하며, 이해관계자들의 신뢰를 얻기 위해 필요한 윤리적 원칙을 준수하는 것을 의미한다(Novelli et al., 2023). 따라서, 이러한 AI 서비스의 책무성은 알고리즘의 작동 방식과 뉴스 추천의 기준을 명확하게 공개하는 투명성, 특정 이념이나 관점에 치우치지 않고 다양한 관점을 균형 있게 제공하기 위한 공정성, 추천된 뉴스가 신뢰할 수 있는 출처에서 온 정확한 정보인지 보장하기 위한 신뢰성 등의 요소를 포함한다(Cheong, 2024). 이에 AI 서비스 제공기업의 책무성은 사용자들이 AI 알고리즘의 작동 방식과 추천 결과에 대해 신뢰할 수 있도록 한다(Felzmann et al., 2019). 따라서, 기업의 책무성 준수는 사용자가 AI 서비스의 신뢰성과 공정성 등을 높게 평가하는 결정적인 원인이 되어 사용자는 AI 서비스 제공 기업이 책무성을 준수함에 따라, 편향되지 않은 신뢰할 만한 정보를 제공받을 수 있기를 기대한다. 즉, 사용자의 AI 서비스 제공 기업의 책무성 준수에 대한 기대는 사용자의 정보 수용에 영향을 미치게 된다(Novelli et al., 2023). 따라서, 본 연구에서는 AI 기반 뉴스 추천 서비스의 책무성에 대한 사용자의 기대가 증가할수록 제공 받은 정보 편향의 인지가 증가할 것으로 판단하여 다음의 가설을 제기한다. 이는 AI 서비스 제공 기업의 책무성 준수가 사용자의 기대에 충족되지 못할 때, 해당 기대의 증가는 제공된 정보의 신뢰성, 투명성, 공정성 등의 결여로 인한 정보의 인지된 편향이 증가할 것으로 가정하였다. 제시된 가설을 검증하기 위하여 본 연구에서는 “추천 서비스 제공자는 이용자의 특성(소비·검색 이력, 연령, 성

별 등)을 고려하여 콘텐츠의 위험성을 사전에 검증해야 한다.”, “인공지능 추천서비스 제공자는 시스템 운영 과정에서 기능적 오류, 오작동, 현행 법령 위반 등 부정적인 결과를 해결할 의무가 있다.”, “인공지능 추천서비스가 의도치 않게 공공의 이익에 반할 때 정부의 제도적 보완이 필요하다.”의 문항을 통하여 AI 서비스 제공기업에 대한 사용자 기대 항목을 측정하였다.

H3. 사용자가 AI 뉴스 추천 서비스 기업의 책무성에 대한 필요성을 높게 인지할수록 인지된 정보 편향성은 증가할 것이다.

3.4. AI 뉴스 추천서비스의 선택권에 대한 사용자의 기대와 인지된 정보 편향성의 관계

AI 기반 추천 시스템은 사용자에게 맞춤형 콘텐츠를 제공함으로써 사용자의 만족도를 높이고 서비스 이용의 편리성을 증대시키는 역할을 한다. 그러나 AI 추천 시스템이 제공하는 콘텐츠가 사용자의 선호와 의도를 충분히 반영하지 못하거나, 오히려 특정 정보나 광고를 과도하게 노출시킬 경우, 사용자는 이에 따른 피로감을 인지한다(Newman, 2015). 이와 같은 상황을 방지하기 위해, AI 뉴스 추천 서비스 제공자는 이용자에게 콘텐츠 선택에 대한 충분한 권한을 부여할 필요성이 제기되고 있다(Villaronga et al., 2018). 이러한 선택권은 사용자가 원치 않는 콘텐츠가 사용자의 피드에 노출되지 않도록 선택 및 변경할 수 있도록 하거나 본인이 제공한 혹은 온라인에 작성한 게시글에 대하여 삭제, 수정 등의 변경 및 요구할 수 있는 권리로 정의할 수 있다(Newman, 2015). 즉, 추천서비스 제공자는 이용자에게 인공지능 추천서비스가 제공하는 콘텐츠를 합리적으로 이용할 수 있는 선택권을 보장할 필요가 있다는 주장은 사용자가 자신의 뉴스 피드를 주도적으로 관리할 수 있어야 함을 강조한다(Ghial et al., 2024; Graves et al., 2020). 사용자는 자신의 관심사와 무관한 콘텐츠를 지속적으로 추천받는 것에 피

로감을 느낄 수 있으며, 이러한 피로감은 서비스에 대한 신뢰도 저하로 이어질 수 있다(Newman, 2015). 따라서 추천 콘텐츠의 선택권을 사용자에게 부여함으로써, 사용자의 주체적인 선택을 촉진하고, 서비스에 대한 만족도를 높일 수 있다.

추천서비스 제공자는 이용자에게 본인이 원하지 않는 추천 콘텐츠의 노출 정도를 선택 또는 변경할 수 있도록 보장함에 따라, 사용자는 온라인 환경 등에서 본인의 선택권을 보장 받아 해당 서비스에 대한 만족감과 더불어 제공된 정보의 신뢰성이 증가하게 된다. 따라서, 사용자가 AI 기반 뉴스 추천 서비스를 사용하며 충분한 선택권을 보장 받지 못한다고 인지하는 경우 이를 보장받고자 하는 권리를 서비스 제공 기업에게 기대하게 된다. 이에 본 연구에서는 이러한 선택권 보장의 기대감이 증가할수록 제공된 정보에 대한 불만족은 인지된 편향성을 증가시킬 것으로 가정하여 다음의 가설을 제기한다. 이에 본 연구에서는 AI 추천 서비스 기업을 향한 사용자의 선택권 보장에 대한 기대를 측정하기 위하여 다음의 문항을 활용하였다. “추천서비스 제공자는 이용자에게 인공지능 추천서비스가 제공하는 콘텐츠를 합리적으로 이용할 수 있는 선택권을 보장할 필요가 있다.”, “추천서비스 제공자는 이용자에게 본인이 원하지 않는 추천 콘텐츠(뉴스, 광고)의 노출 정도를 선택 또는 변경할 수 있도록 해야 한다.”, “추천서비스 제공자는 이용자에게 인공지능 추천서비스 이용정보에 대한 처분권(예: 수정, 삭제, 변경 요구)을 제공할 필요가 있다.”를 통하여 다음의 가설을 확인하였다.

H4. 사용자가 AI 뉴스 추천 서비스 기업의 선택권에 대한 필요성을 높게 인지할수록 인지된 정보 편향성은 증가할 것이다.

3.5. AI 뉴스 추천서비스에 대한 설명 요구권과 사용자의 기대와 인지된 정보 편향성의 관계

AI 서비스에 대한 설명 요구권은 사용자들이 AI 시스

템의 작동 방식, 데이터 처리 방법, 그리고 의사결정 과정을 이해할 수 있도록 하는 권리로 다음의 사항을 포함한다. 첫째, 사용자에게 제공되는 콘텐츠 등 데이터의 선별 기준에 대한 설명이다. 즉, 사용자는 AI 알고리즘이 콘텐츠를 어떻게 선별하고 추천하는지에 대해 명확한 설명을 요구할 수 있다. 이는 AI가 어떤 데이터와 기준을 사용하여 특정 콘텐츠를 추천하는지를 포함하며, 이러한 데이터의 출처 정보를 포함하여 사용자가 AI 시스템의 작동 원리를 이해하고, 추천 콘텐츠의 신뢰성과 공정성을 판단하는 데 기여하는 것으로 밝혀졌다(Albahri et al., 2023). 둘째, 제공된 콘텐츠의 생성 및 제작에 대한 정보이다. 특히, 오늘날과 같이 생성형 AI의 확산으로 AI가 다양한 콘텐츠를 생성할 수 있는 환경에서 사용자는 자신이 접하는 콘텐츠가 인간에 의해 생성된 것인지, 아니면 AI에 의해 생성된 것인지에 대한 정보를 요구한다(Balasubramaniam et al., 2022). 이는 제공된 정보의 생성자에 대한 정확한 식별을 통한 인간에 의하여 생성된 콘텐츠의 지적재산권을 보호하는 기반이 된다(Vainio-Pekka et al., 2023b). 셋째, AI가 수집한 사용자의 개인정보 수집 및 활용 방법에 대하여 설명을 요구할 수 있는 권리이다. 이는 개인정보가 어떤 방식으로 수집되며, 어떤 목적으로 사용되는지, 그리고 데이터 보호를 위해 어떤 조치가 취해지고 있는지에 대한 것으로 사용자가 자신의 개인정보가 안전하게 보호되고 있는지를 판단하고 이를 충족하지 못할 시, 적절한 개선조치 등을 요구할 수 있도록 하기 위함이다(Alufaisan et al., 2021). 나아가, 사용자는 AI 추천 서비스를 이용하는 과정에서 발생할 수 있는 사고나 피해에 대한 위험성과 문제 발생 시, 이를 대처하기 위한 서비스 제공기업의 제반절차에 대하여 사용자가 요구하여 안내 받을 수 있는 권리이다(Alufaisan et al., 2021; Mensah, 2023). 이는 AI 서비스 제공자가 이러한 사고나 피해에 대해 어떤 책임을 지는지, 그리고 사용자가 어떤 보호를 받을 수 있는지에 대한 설명을 포함하여 사용자가 서비스 이용 중 발생할

수 있는 잠재적 위험을 인식하고, 이에 대비할 수 있도록 한다(Albahri et al., 2023; Alufaisan et al., 2021). 특히, 사용자에게 이루어지는 AI 기반 추천 서비스의 충분한 설명 제공은 사용자가 해당 서비스에 대한 명확한 이해를 도모하여 서비스 사용의 만족감을 향상시키는 동시에 제공 받은 정보에 대한 투명성과 신뢰성을 높이는 것으로 밝혀졌다(Mensah, 2023).

본 연구에서는 AI 추천 서비스 제공 기업에게 사용자가 기대하는 설명 요구권은 사용자가 서비스를 사용하는 과정에서 해당 AI 모델의 구현 원리, 모델 학습 방식, 보호 절차 등에 대하여 충분히 설명을 제공받지 못한다고 인지하는 경우, 이는 제공 받은 정보의 신뢰 저하로 연결되어 정보가 왜곡 및 편향되었을 것이라는 사용자 인식을 형성할 것으로 판단하였다. 이에 본 연구는 사용자가 AI 뉴스 추천 서비스 기업에게 기대하는 설명 요구권을 측정하기 위하여 “추천서비스 제공자는 이용자에게 인공지능 알고리즘이 선별한 콘텐츠의 선별 기준을 알려 줄 필요가 있다.”, “추천서비스 제공자는 이용자에게 콘텐츠 제작자(생산자)가 인간 혹은 인공지능인지 알려 줄 필요가 있다.”, “추천서비스 제공자는 이용자에게 제작자(생산자)의 인공지능 서비스(맞춤형 콘텐츠 추천) 운영 과정에서 자신의 개인정보가 어떻게 수집되고 활용되는지를 알려 줄 필요가 있다.”, “추천서비스 제공자는 이용자에게 인공지능 추천서비스를 이용하다 발생한 사고로 인한 피해나 불이익에 대해 설명할 필요가 있다.”의 문항을 활용하여 다음의 가설을 검증하고자 한다.

H5. 사용자가 AI 뉴스 추천 서비스 제공기업의 설명 요구권에 대한 필요성을 높게 인지할수록 인지된 정보 편향성은 증가할 것이다.

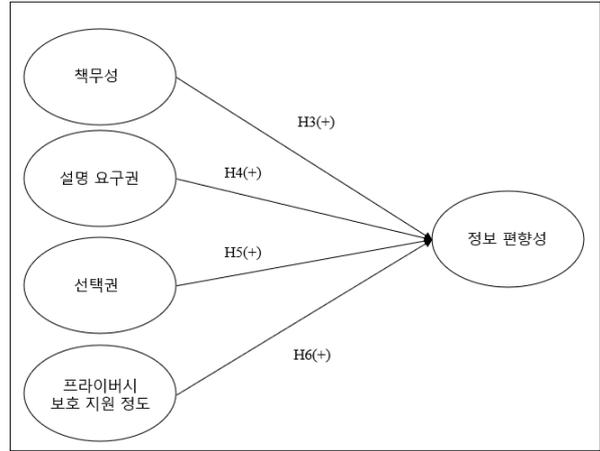
3.6. AI 뉴스 추천서비스의 프라이버시 염려에 대한 사용자의 기대와 인지된 정보 편향성의 관계

AI 기반 뉴스 추천 서비스는 사용자에게 맞춤형 정보

를 제공하기 위하여 사용자의 선호도, 과거 사용자의 접근 웹페이지 등을 분석하여 이를 바탕으로 사용자에게 뉴스를 추천하는 방식으로 작동한다. 그러나 최근 들어, 해킹, 랜섬웨어 등 다양한 정보보안 실패 사고가 증가하며 맞춤형 정보를 제공하는 것과 더불어 사용자의 프라이버시 보호의 필요성 인식이 증가하고 있다(Zhang et al., 2021). 이러한 프라이버시 관련 사용자의 인식은 서비스를 사용 및 선택하는 과정 뿐만 아니라, 서비스를 사용함에 따라 제공 받은 정보의 수용 여부 및 정보 신뢰에 영향을 미치는 것으로 나타났다(Zhang et al., 2021). 따라서, 사용자는 AI 뉴스 추천 서비스를 사용하는 과정에서 해당 서비스 제공 기업의 프라이버시 보호가 충분히 이루어지지 않는다고 판단하는 경우, 이로 인해 추천 특정 편향을 갖고 있을 것이라는 인식을 강화 시킨다. 나아가, 사용자는 데이터 보호에 대한 불안감을 느낄 때, AI 시스템이 제공하는 정보를 비판적으로 수용하는 경향이 높았다(Hu & Min, 2023). 사용자는 개인정보를 포함하여 자신이 제공한 데이터의 보호가 잘 이루어진다고 믿을 때, 서비스에 대한 신뢰도가 높아진다(Liu & Tao, 2022). 이러한 신뢰는 AI 서비스가 제공하는 정보의 공정성과 신뢰성을 평가하는 데 중요한 역할을 하는 동시에 사용자가 정보를 수용하는 과정에서 인지된 정보의 편향성을 완화시킬 수 있다. 그러나 사용자가 서비스를 사용하면서 프라이버시 보호가 충분히 이루어지지 않는다고 인식하는 경우, AI 서비스에 대한 전반적인 신뢰가 감소하고 결과적으로 추천된 뉴스 정보의 편향성에 대한 인식이 증가할 것으로 본 연구는 판단하였다. 이에 AI 뉴스 추천 서비스를 사용하는 사용자가 프라이버시 보호 지원에 대한 기대가 클수록, 그들이 제공받는 정보에 대하여 인지된 편향은 높아질 것이라는 가설을 제기한다.

H6. 사용자가 AI 뉴스 추천 서비스 제공기업의 프라이버시 보호 지원 필요성을 높게 인지할수록 인지된 정보 편향성은 증가할 것이다.

앞서 제시된 가설을 바탕으로 다음과 같은 연구 모형을 제시한다.



〈그림 1〉 연구 모형

4. 연구방법 및 분석

4.1. 데이터 수집 및 특성

본 연구에서는 앞서 제시된 연구모형을 검증하기 위하여 정보통신정책연구원(KISDI)의 ‘2022년 지능정보사회 이용자 패널조사’ 데이터를 활용하였다. 본 패널조사는 빠르게 고도화되는 지능정보 기술 및 서비스에 대한 사용자의 인식 및 태도 등에 대한 패널데이터를 수집하여 사용자 중심의 방송통신, 기술 정책 개발을 위하여 실행되었다. 따라서, 대한민국 거주자의 만 15세 이상부터 만 69세 이하의 하루 1회 이상 스마트폰 및 인터넷을 이용하는 사용자를 대상으로 디지털 기기/서비스 및 지능정보서비스 이용 현황, 국내 지능정보화 환경 평가, 최신 지능정보 기술 및 서비스에 대한 인지, 태도 등을 조사하였다. 따라서, 본 연구는 KISDI의 설문조사 데이터를 바탕으로 AI 기반의 자동 추천 뉴스 서비스를 이용하는 사용자의 인식에 따른 인지된 정보의 편향성을 살펴보고자

하였다. 이를 위하여 전체 조사 대상자인 5,378명 중, AI 기반 뉴스 추천 서비스를 사용해본 경험이 있는 사용자인 3,947명만을 본 연구의 분석 대상으로 활용하였다. 이와 같이 본 연구에서 사용된 데이터는 대규모의 연구 표본을 대상으로 이루어짐에 따라, 연구 결과 일반화의 가능성을 높였으며 정기적으로 실시되는 KISDI의 설문 조사라는 점에서 수집 데이터의 신뢰성과 데이터 품질을 보장할 수 있다.

연구가설을 검증하기 위하여 사용된 대상자의 인구사회학적 특성은 다음의 <표 1>과 같다.

4.2. 측정도구의 신뢰성 및 타당성 분석

본 연구는 앞서 제시된 연구 가설을 검증하기 위하여 IBM SPSS Statistics과 구조방정식 모델을 기반으로 하는 Smart PLS 4.0을 이용하여 구성된 설문 문항의 확인

적 요인분석을 수행함에 따라, 신뢰도와 타당성을 검증하였다. 본 연구의 설문 문항과 같이 여러 항목을 바탕으로 동일한 개념을 측정하는 경우, 내적 일관성을 판단하기 위하여 Cronbach' a를 사용한다. 일반적으로 연구 분석에 사용된 설문 문항의 수용 가능한 신뢰성은 Cronbach' a의 0.7을 기준으로 평가함에 따라, 본 연구에서 사용된 문항은 모두 신뢰도를 확보한 것으로 판단하였다(Jc & Ih, 1994). 나아가, 본 연구에서 수립된 가설을 검증하기 위하여 사용된 구조방정식 모형의 집중 타당도는 복합 신뢰도(Composite Reliability, CR)와 평균분산추출(Average Variance Extracted, AVE)을 바탕으로 확인하였다. 복합 신뢰도는 0.7 이상, 평균분산추출(AVE)은 0.5 이상을 각각 기준으로 연구 분석에 사용된 데이터의 타당성을 판단한다(Fornell & Larcker, 1981). 본 연구에서 조사 응답자의 성별, 연령 등의 인구 통계학적 특징을 살펴보기 위한 문항과 선호하는 뉴스 추천 방식의 문항

<표 1> 표본의 인구통계학적 특성

구분		빈도	비율(%)
성별	남자	1761	44.2
	여자	2186	54.8
	합계	3947	100
연령 분포	15~19세	182	4.6
	20~29세	815	20.7
	30~39세	785	19.9
	40~49세	764	19.4
	50~59세	813	20.5
	60~69세	588	14.9
	합계	3947	100
최종 학력	초등학교 졸업	12	0.3%
	중학교 졸업	105	2.7%
	고등학교 졸업	1406	35.6%
	전문대학 졸업	836	21.2%
	대학교 졸업	1561	39.5%
	석사 졸업	24	0.6%
	박사 졸업	3	0.1%
	합계	3947	100

을 제외한 총 19개의 문항에 대한 신뢰도와 타당성 결과는<표 2>를 통하여 제시하였다.

판별타당성은 각 구성개념의 상관계수와 AVE의 제곱근 값을 비교하여 검증한다(김준영, 김태성, 2018). 본 연구의 판별타당성 분석 결과, 모든 구성개념에서 평균분산추출 값이 상관관계 제곱보다 높아 측정항목의 판별타당성이 확보되었다. 구체적인 분석 결과는<표 3>에서 확

인할 수 있다.

4.3. 분석 결과

본 연구는 AI 뉴스 추천 서비스 사용자가 제공 기업에게 기대하는 다양한 AI 윤리 원칙이 사용자가 지각하는 정보 편향에 미치는 영향에 대하여 살펴보는 것에 주요

<표 2> 신뢰도 및 타당성 분석 결과

변수(Factor)	항목(Items)	요인적재량 (Factor Loading)	Cronbach's α	복합신뢰도(CR)	평균분산추출(AVE)
정보 리터러시 (lit)	lit1	0.783	0.782	0.824	0.704
	lit2	0.778			
	lit3	0.749			
	lit4	0.782			
책무성 (acc)	acc1	0.902	0.849	0.898	0.788
	acc2	0.814			
	acc3	0.792			
설명 요구권 (exp)	exp1	0.757	0.743	0.812	0.698
	exp2	0.760			
	exp3	0.707			
	exp4	0.786			
선택권 (cho)	cho1	0.808	0.866	0.865	0.780
	cho2	0.857			
	cho3	0.986			
프라이버시 보호 지원 정도 (pri)	pri1	0.831	0.878	0.904	0.880
	pri2	0.834			
	pri3	0.869			
정보 편향성 (bias)	bias1	0.711	0.797	0.848	0.712
	bias2	0.748			

<표 3> 판별타당성 분석 결과

변수	lit	acc	exp	cho	pri	bias
lit	0.775					
acc	0.551	0.866				
exp	0.351	0.650	0.837			
cho	0.450	0.751	0.552	0.854		
pri	0.252	0.354	0.451	0.651	0.806	
bias	0.151	0.457	0.354	0.552	0.552	0.825

목표가 있다. 이에 앞서서 사용자의 정보 리터러시 역량과 사용자가 선호하는 뉴스 추천 방식에 따른 정보 편향성의 인지 정도 차이를 확인하기 위하여 다음과 같은 연구 분석을 수행하였다. 먼저, 개인이 평가한 스스로의 정보 리터러시 역량을 바탕으로 상위 및 하위의 두 개의 집단으로 구분하여 각 집단의 인지된 정보 편향성 차이의 통계적 유의성을 독립표본 t 검정을 통하여 검증한다. 나아가, 과거 본인이 클릭한 뉴스 콘텐츠 기록을 바탕으로 자동 추천해주는 알고리즘 기반의 추천 방식과 전문가 및 지인에 의한 뉴스 콘텐츠 추천 방식 중, 상대적으로 사용자가 더 선호하는 방식에 따라 설문 조사 응답자를 두 개의 집단으로 구성하였다. 이를 바탕으로 사용자가 선호하는 추천 방식에 따른 지각된 정보 편향성에는 통계적으로 차이의 유의성을 독립표본 t 검정을 통하여 밝히고자 한다. 나아가, 사용자가 요구하는 AI 뉴스 추천 서비스 제공 기업의 윤리 원칙과 인지된 정보 편향성의 관계를 살펴보고자 본 연구는 구조방정식 모형(Structural Equation Modeling)을 이용하였으며, 모형 분석은 Smart PLS 4.0을 이용하였다.

먼저, 사용자가 선호하는 뉴스 추천 방식의 정보 편향성을 평가하기 위해 평균 점수와 표준편차를 측정하였으

며, 두 그룹 간의 차이를 t-검정을 통해 분석하였다. 분석 결과, 전문가 및 지인 추천 그룹($n=2111$)의 평균 점수는 3.39(표준편차=0.891)로 나타났으며, 과거 사용자가 클릭한 뉴스로 자동 추천된 그룹($n=1836$)의 평균 점수는 3.57(표준편차=0.806)로 두 그룹 간의 차이는 통계적으로 유의미한 수준을 보였다. 특히, 과거 사용자의 클릭 데이터를 기반으로 한 자동 추천이 전문가 및 지인 추천보다 지각된 정보 편향성이 상대적으로 더 높음에 따라, 사용자는 AI 등 자동 추천 방식이 제공하는 뉴스 콘텐츠는 공정성이 왜곡된 정보로 인식하는 경향이 더 높았음을 시사한다. 둘째, 정보 리터러시 역량을 스스로 낮게 평가한 집단($n=1539$)은 인지된 정보 편향성의 평균이 3.66(표준편차=0.891)으로 상대적으로 정보 리터러시 역량을 높게 평가한 집단($n=2408$)의 평균 점수인 3.35(표준편차=0.811)에 비하여 높게 나타났다. 이는 정보 리터러시 수준을 스스로 높게 평가한 집단이 낮게 평가한 집단에 비해 정보 편향성을 더 높게 인지하고 있음을 시사한다. 이는 정보에 대한 활용 능력이 상대적으로 높다고 인지한 사용자 대부분은 제공된 정보가 왜곡되었거나 공정성이 상실되었을 것이라는 의문을 가지는 경향이 높다는 점을 추론할 수 있다. 나아가, 이러한 정보에 대한 의

〈표 4〉 사용자의 선호하는 뉴스 추천 방식에 따른 인지된 정보 편향성

구분	정보 편향성				
	n	평균	표준편차	t	p
전문가 및 지인 추천	2111	3.39	0.891	6.426	〈.001***
과거 사용자가 클릭한 뉴스를 바탕으로 자동 추천	1836	3.57	0.806		

*** $p < 0.001$

〈표 5〉 정보 리터러시 역량에 따른 인지된 정보 편향성

구분	정보 편향성				
	n	평균	표준편차	t	p
정보 리터러시 상위 집단	1539	3.66	0.891	11.081	.033*
정보 리터러시 하위 집단	2408	3.35	0.811		

* $p < 0.05$

〈표 6〉 가설 검증결과

	Path	Path Coefficients	t-value	Result
H3	책무성 → 정보 편향성	0.130***	1.149	supported
H4	설명 요구권 → 정보 편향성	0.585**	25.238	supported
H5	선택권 → 정보 편향성	0.684**	1.880	supported
H6	프라이버시 보호 지원 정도 → 정보 편향성	0.206***	3.875	supported

*** $p < 0.001$, ** $p < 0.01$

구심이 사용자 집단의 정보 리터러시 역량을 형성하는 기반이 되었음을 예상할 수 있다. 따라서, 정보 리터러시 교육이 올바른 정보 선별 능력과 정보를 무분별하게 수용하지 않는 역량의 향상을 통하여 정보 편향의 완화와 더불어 보다 균형 잡힌 정보 접근의 촉진을 향상시킬 수 있을 것으로 판단된다.

제시된 연구 가설을 검증하기 위하여 구조방정식 모형을 분석한 결과, 사용자가 AI 기반 뉴스 제공사에게 기대하는 책무성($\beta = .130, p < 0.001$), 설명 요구권($\beta = .585, p < 0.001$), 선택권($\beta = .684, p < 0.001$) 및 프라이버시 보호 지원 정도($\beta = .206, p < 0.001$)가 증가할수록 인지된 정보 편향성이 증가함에 따라, H3, H4, H5, H6 가설이 각각 채택되었다. 이러한 결과는 사용자가 인지하는 윤리 원칙이 AI 뉴스 추천 시스템의 정보 편향성 인식에 중요한 역할을 함을 시사한다. 특히, 사용자는 AI 서비스 기업이 서비스를 제공함에 따라, 사용자의 이익을 해치는 등의 발생 가능한 위협성을 해결하기 위한 책임을 다하지 못하거나 AI 모델에 대한 충분한 운영 방식과 작동 원리를 충분히 사용자에게 설명하지 못한다고 판단되어 이에 대한 개선이 필요하다고 인지하는 경우, 사용자는 해당 기업으로 제공받은 정보가 균형성을 상실한 것으로 판단할 가능성이 높다는 점을 확인하였다. 나아가, AI 기반으로 뉴스 콘텐츠를 추천 받는 과정에서 사용자가 과거 제공한 데이터의 삭제 요청, 사용자의 피드에 원치 않는 정보의 공개를 조정할 수 있는 사용자의 선택 범위에 대한 기대의 증가 또한 인지된 편향을 심화시키는 것으로 밝

혀졌다. 이는 사용자가 AI 기반 서비스를 사용하는 과정에서 본인의 선택권에 대한 보장이 기업으로부터 충분히 고려되지 않는 것으로 인식하여 이에 대한 보장의 기대로 볼 수 있다. 최종적으로, 구조 모형 분석을 통해 나타난 결과를 정리하면 <표 6>과 같다.

5. 결론

5.1. 연구 결과 요약

본 연구는 AI 기반 뉴스 추천 서비스 사용자들이 인지하는 정보 편향성을 AI 뉴스 추천 서비스 기업의 윤리 원칙 요인들을 바탕으로 살펴보았으며 해당 분석 결과를 요약하면 다음과 같다. 첫째, 사용자가 선호하는 뉴스 추천 방식에 따라 AI 추천 뉴스 기사에 대한 인지된 정보 편향성이 차이가 있다는 가설을 검증한 결과, 전문가 및 지인 추천 방식보다 AI 자동 추천 방식에서 인지된 정보 편향성이 더 높게 나타났다. 둘째, 정보 리터러시가 높은 사용자가 그렇지 않는 사용자 집단에 비하여 더 높은 편향성을 인지하는 것으로 나타났다. 셋째, 사용자가 AI 뉴스 추천 서비스 기업의 책무성, 선택권, 설명 요구권에 대한 기대와 프라이버시 보호 지원에 대한 필요성을 높게 인지할수록 인지된 정보 편향성이 증가하였다.

5.2. 연구 시사점

AI 기반 뉴스 추천 서비스가 사용자에게 미치는 정보 편향의 영향을 본 연구에서는 살펴봄에 따라 다음과 같은 시사점을 제시한다. 첫째, AI 뉴스 추천 서비스의 윤리적 원칙이 사용자가 인지하는 정보 편향에 미치는 영향을 바탕으로 본 연구는 AI 시스템의 공정성과 투명성을 높이는 방법에 대한 학문적 논의를 시사하였다. 이를 통해 AI 기술의 발전과 함께 나타나는 윤리적 문제들을 체계적으로 다루는 연구의 필요성을 제기하였다. 둘째, 본 연구는 AI 뉴스 추천 서비스를 운영하는 기업들이 향후, 보다 안전한 사용자의 이용 환경을 구성하기 위하여 고려하여야 하는 윤리 원칙의 우선 항목을 제시하였다. 이는 본 연구는 기업들이 AI 시스템의 공정성, 투명성 등의 확보를 위하여 책무성, 설명 의무, 사용자의 선택권 보장 및 프라이버시 보호 등을 강화함으로써 인지된 사용자의 정보의 편향을 완화하기 위한 기반을 확인하였다. 이는 AI 서비스에 대한 사용자의 신뢰를 얻고, 서비스의 질을 높이는 데 실질적인 전략 수립에 기여한다. 셋째, 본 연구는 최근, AI 뉴스 추천 서비스가 필터 버블과 에코 챔버 현상을 강화하는 문제를 탐색하며, 다양한 관점의 뉴스 소비를 촉진하는 방향 마련의 필요성을 제시하였다. 이는 사용자들이 보다 다양한 정보를 접할 수 있는 환경의 구성 필요성을 제기하는 동시에 사용자의 정보 리터러시 역량 강화를 통하여 왜곡되지 않은 올바른 정보의 선별적 수용의 중요성을 확인하였다. 넷째, 본 연구의 결과는 뉴스 콘텐츠의 자동 추천을 포함하여 AI 서비스 제공기업은 프라이버시 보호와 관련된 윤리적 기준 및 추천 서비스의 윤리적 운영 원칙을 명확히 하여 법적 규제와의 적합성을 확보하는 데 기여할 수 있다. 특히, 본 연구는 오늘날 AI 추천 서비스의 윤리적 운영 원칙을 명확히 하여 법적 규제와의 적합성을 확보하기 위한 기업의 노력, 필요성을 확인하였다. 마지막으로 본 연구는

기존 연구들이 AI 자체의 편향성 문제에 집중한 반면에 사용자의 인지된 편향성을 연구하였다는 점에서 향후, 사용자 인식에 기반하여 AI가 실질적으로 어떤 방식으로 수용되고 사용되는지를 이해하기 위한 기반이 될 것으로 예상된다. 이는, AI와 인간 간의 상호작용을 보다 윤리적이고 사회적으로 책임 있는 방향으로 견인할 수 있을 것으로 판단된다. 특히, 편향성 연구는 현재 AI 알고리즘의 데이터 및 학습 과정에서 발생하는 기술적 문제와, 사용자의 인지적 편향성 문제를 중심으로 이루어지나, 본 연구는 이러한 두 가지 측면을 통합하여 사용자 인지의 중요성을 강조함으로써 기존 연구의 틀을 확장한다. 이는 사용자와 AI의 상호작용을 고려한 새로운 접근법으로, 기존의 기술 중심의 연구와 차별화되는 학문적 공헌도를 확인할 수 있다.

나아가, 본 연구는 AI 뉴스 추천 서비스가 저장하고 있는 사용자의 과거 데이터에 대한 삭제 요구 및 불필요하거나 부정확한 정보 등 원치 않은 콘텐츠가 본인의 피드에 계속해서 노출되는 것을 거부할 수 있는 선택권을 살펴보았다. 이는 최근 글로벌 환경에서 강조되는 잊혀질 권리('Right to be forgotten')와 연결됨에 따라, 기업은 단순 GDPR 등 정책과 법이 요구하는 항목의 준수를 넘어 사용자의 서비스 안정성 확보 및 균형 있는 정보 제공의 목적 성취를 위해서 반드시 고려하여야 할 항목임을 본 연구에서 검증하였다. 최종적으로 AI 기업에게 사용자가 기대하는 윤리적 항목에 부응할 수 있는 기업만이 AI 시장에서 지속적으로 서비스를 제공할 수 있는 동시에 사용자의 서비스 이용 만족도를 증진시킬 수 있을 것으로 예상된다.

5.3. 연구 한계 및 향후 계획

본 연구는 KISDI에서 시행된 설문조사의 응답을 대상으로 본 연구를 수행함에 따라, 다음과 같은 연구의 한계

점이 존재한다. 첫째, 본 연구는 AI 뉴스 추천 시스템의 윤리적 고려사항과 사용자의 인지된 편향의 관계에 대하여 주로 다루었지만, 실제 AI 알고리즘의 구체적인 작동 방식이나 데이터 처리 과정에 대한 기술적 사항의 고려가 이루어지지 않았다. 향후 연구에서는 다양한 AI 모델의 구조와 작동 방식을 분석하여, 편향성이 발생하는 구체적인 원인을 더 명확히 규명할 필요가 있다. 이를 통하여 실질적으로 편향을 완화할 수 있는 AI 모델 관점에서의 해결방안을 마련할 예정이다. 둘째, 본 연구는 단기적인 관점에서 사용자의 인식을 분석하였으나, AI 뉴스 추천 서비스가 장기적으로 사용자 행동과 인식에 미치는 영향을 분석하여 사용자의 인지된 정보 편향을 심화하는 항목을 추가적으로 탐색하고자 한다. 셋째, 본 연구에서 AI의 윤리 원칙으로 제시된 네 가지 항목의 중요성에 대한 우선순위를 도출하기 위한 연구를 추가적으로 수행할 예정이다. 즉, 다양한 AI의 윤리적 원칙 중에서 AI 서비스 제공자와 사용자 사이에서 우선적으로 중요 요인으로 평가하는 항목의 차이를 비교하고자 한다.

본 연구는 AI 뉴스 추천 서비스의 윤리적 원칙이 정보 편향성에 미치는 영향을 다각적으로 분석하여 향후 AI 서비스 제공기업이 고려하여야 하는 윤리 원칙에 대한 기반을 제공하였다는 점에서 의의가 있다. 이를 바탕으로 향후, 기업은 AI 서비스를 설계하는 과정에서 우선적으로 고려 및 준수하여야 하는 항목의 이해를 높인다. 나아가, 본 연구의 결과는 기업의 AI 윤리 원칙 프레임워크 및 내부 정책 수립의 기반이 되는 동시에 국가적으로 AI 윤리 가이드라인 개발의 토대가 될 것으로 판단된다. 이를 통하여 사용자는 보다 맞춤형 정보를 제공받는 동시에 균형 있는 정보의 수용이 용이해짐에 따라, AI 기반 자동 추천 서비스의 본래 가치를 증진시킬 수 있을 것으로 예상된다.

〈참고문헌〉

[국내 문헌]

1. 김준영, 김태성 (2018). 누가 보안에 더 신경을 쓰는가?: 성격특성이 pc 보안의도에 미치는 영향. *경영정보학연구*, 20(1), 99-114.
2. 김형지, 정은령, 김은미, 양소은, 이재우, 강민지 (2020). 가짜뉴스와 팩트체크 뉴스 노출 집단의 미디어 이용과 뉴스 인식, 그리고 리터러시 관계. *한국언론정보학보*, 101, 231-267.
3. 박승택, 성인재, 서상원, 황지수, 노지성, 김대원 (2017). 기계학습 기반의 뉴스 추천 서비스 구조와 그 효과에 대한 고찰: 카카오의 루빅스를 중심으로. *사이버커뮤니케이션학보*, 34(1), 5-48.
4. 엄정윤, 정세훈 (2019). 가짜뉴스 노출과 전파에 영향을 미치는 요인: 성격, 뉴미디어 리터러시, 그리고 이용 동기. *한국언론학보*, 63(1), 7-45.

[국외 문헌]

5. Aboutorab, H., Hussain, O. K., Saberi, M., Hussain, F. K., & Prior, D. (2023). Reinforcement learning-based news recommendation system. *IEEE Transactions on Services Computing*.
6. Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., et al. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*, 96, 156-191.
7. Alufaisan, Y., Marusich, L. R., Bakdash, J. Z., Zhou, Y., & Kantarcioglu, M. (2021). Does explainable artificial intelligence improve human decision-making? Paper presented at *the Proceedings of the AAI Conference on Artificial Intelligence*, 35(8), 6618-6626.
8. Amir, N., Jabeen, F., Ali, Z., Ullah, I., Jan, A. U., & Kefalas, P. (2023). On the current state of deep learning for news recommendation. *Artificial Intelligence Review*, 56(2), 1101-1144.
9. Areeb, Q. M., Nadeem, M., Sohail, S. S., Imam, R., Doctor, F., Himeur, Y., et al. (2023). Filter bubbles in recommender systems: Fact or fallacy—A systematic review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(6), e1512.
10. Balasubramaniam, N., Kauppinen, M., Hiekkänen, K., & Kujala, S. (2022). Transparency and explainability of AI systems: Ethical guidelines in practice. Paper presented at *the International Working Conference on Requirements Engineering: Foundation for Software Quality*, 3-18.
11. Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). 'It's reducing a human being to a percentage' perceptions of justice in algorithmic decisions. Paper presented at *the Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems*, 1-14.
12. Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, 1, 61-65.
13. Branley-Bell, D., Whitworth, R., & Coventry, L. (2020). User trust and understanding of explainable ai: Exploring algorithm visualisations and user biases. Paper presented at *the International Conference on Human-Computer Interaction*, 382-399.
14. Cheng, L., Varshney, K. R., & Liu, H. (2021). Socially responsible ai algorithms: Issues, purposes, and challenges. *Journal of Artificial Intelligence Research*, 71, 1137-1181.
15. Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, 6, 1421273.
16. DeCamp, M., & Lindvall, C. (2023). Mitigating bias in AI at the point of care. *Science*, 381(6654), 150-152.
17. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv Preprint arXiv:1702.08608*.
18. Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6(1), 2053951719860542.
19. Feng, C., Khan, M., Rahman, A. U., & Ahmad, A. (2020). News recommendation systems—accomplishments, challenges & future directions. *IEEE Access*, 8, 16702-16725.
20. Fornell, C., & Larcker, D. F. (1981). Evaluating structural

- equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50.
21. Gao, R., & Shah, C. (2021). Addressing bias and fairness in search systems. Paper presented at *the Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2643–2646.
 22. Gena, C., Grillo, P., Lieto, A., Mattutino, C., & Venero, F. (2019). When personalization is not an option: An in-the-wild study on persuasive news recommendation. *Information*, *10*(10), 300.
 23. Ghial, R., Pundir, D., & Kaur, R. (2024). Right to be forgotten: A human rights-based approach for governance in generative AI. Paper presented at *the International Conference on Smart Computing and Communication*, 23–33.
 24. Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. Paper presented at *the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 80–89.
 25. Giovanola, B., & Tiribelli, S. (2023). Beyond bias and discrimination: Redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. *AI & Society*, *38*(2), 549–563.
 26. Graves, L., Nagisetty, V., & Ganesh, V. (2020). *Does AI remember? neural networks and the right to be forgotten*.
 27. Hallamaa, J., & Kalliokoski, T. (2022). AI ethics as applied ethics. *Frontiers in Computer Science*, *4*, 776837.
 28. Haque, R., bin S Senathirajah, A. R., Qazi, S. Z., Afrin, N., Ahmed, M. N., & Khalil, M. I. (2024). Factors of information literacy preventing fake news: A case study of libraries in developing countries. *International Journal of Religion*, *5*(7), 804–817.
 29. Heiss, R., Nanz, A., & Matthes, J. (2023). Social media information literacy: Conceptualization and associations with information overload, news avoidance and conspiracy mentality. *Computers in Human Behavior*, *148*, 107908.
 30. Hu, Y., & Min, H. K. (2023). The dark side of artificial intelligence in service: The “watching-eye” effect and privacy concerns. *International Journal of Hospitality Management*, *110*, 103437.
 31. Jc, N., & Ih, B. (1994). *Psychometric theory*. New York, 35.
 32. Jeon, Y., Kim, J., Park, S., Ko, Y., Ryu, S., Kim, S., et al. (2024). HearHere: Mitigating echo chambers in news consumption through an AI-based web system. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–34.
 33. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.
 34. Konow, J. (2005). Blind spots: The effects of information and stakes on fairness bias and dispersion. *Social Justice Research*, *18*, 349–390.
 35. Kumar, A., Aelgani, V., Vohra, R., Gupta, S. K., Bhagawati, M., Paul, S., et al. (2024). Artificial intelligence bias in medical system designs: A systematic review. *Multimedia Tools and Applications*, *83*(6), 18005–18057.
 36. Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, *9*(2).
 37. Lechterman, T. M. (2022). The concept of accountability in AI ethics and governance. In *The Oxford Handbook of AI Governance*. Oxford University Press.
 38. Leslie, D. (2019). Understanding artificial intelligence ethics and safety. *arXiv Preprint arXiv:1906.05684*.
 39. Lin, C., Xie, R., Guan, X., Li, L., & Li, T. (2014). Personalized news recommendation via implicit social experts. *Information Sciences*, *254*, 1–18.
 40. Lin, C., Xie, R., Li, L., Huang, Z., & Li, T. (2012). Premise: Personalized news recommendation via implicit social experts. Paper presented at *the Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 1607–1611.
 41. Liu, J., Shi, C., Yang, C., Lu, Z., & Philip, S. Y. (2022). A survey on heterogeneous information network based recommender systems: Concepts, methods, applications and resources. *AI Open*, *3*, 40–57.
 42. Liu, K., & Tao, D. (2022). The roles of trust,

- personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services. *Computers in Human Behavior*, *127*, 107026.
43. Memarian, B., & Doleck, T. (2023). Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI), and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, 100152.
 44. Mensah, G. B. (2023). Artificial intelligence and ethics: A comprehensive review of bias mitigation, transparency, and accountability in AI systems. *Preprint*, November, 10.
 45. Mukherjee, A. N. (2022). Application of artificial intelligence: Benefits and limitations for human potential and labor-intensive economy—an empirical investigation into pandemic ridden indian industry. *Management Matters*, *19*(2), 149–166.
 46. Nazer, L. H., Zatarah, R., Waldrip, S., Ke, J. X. C., Moukheiber, M., Khanna, A. K., et al. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health*, *2*(6), e0000278.
 47. Newman, A. L. (2015). What the “right to be forgotten” means for privacy in a digital age. *Science*, *347*(6221), 507–508.
 48. Novelli, C., Taddeo, M., & Floridi, L. (2023). Accountability in artificial intelligence: What it is and how it works. *Ai & Society*, 1–12.
 49. Papakyriakopoulos, O., Serrano, J. C. M., & Hegelich, S. (2020). Political communication on social media: A tale of hyperactive users and bias in recommender systems. *Online Social Networks and Media*, *15*, 100058.
 50. Raza, S., & Ding, C. (2022). News recommender system: A review of recent progress, challenges, and opportunities. *Artificial Intelligence Review*, 1–52.
 51. Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. *Recommender Systems Handbook*, 257–297.
 52. Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *10*(4), 1–31.
 53. Stahl, B. C., & Wright, D. (2018). Ethics and privacy in AI and big data: Implementing responsible research and innovation. *IEEE Security & Privacy*, *16*(3), 26–33.
 54. Sureephong, P., Chernbumroong, S., Sangamuang, S., Sirasakamol, O., Intawong, K., & Puritat, K. (2023). Enhancing information literacy for spotting fake news: A study on the efficacy of a serious game for M-learning across different age groups. *International Journal of Interactive Mobile Technologies*, *17*(15).
 55. Vainio-Pekka, H., Agbese, M. O., Jantunen, M., Vakkuri, V., Mikkonen, T., Rousi, R., et al. (2023a). The role of explainable AI in the research field of AI ethics. *ACM Transactions on Interactive Intelligent Systems*, *13*(4), 1–39.
 56. Vainio-Pekka, H., Agbese, M. O., Jantunen, M., Vakkuri, V., Mikkonen, T., Rousi, R., et al. (2023b). The role of explainable AI in the research field of AI ethics. *ACM Transactions on Interactive Intelligent Systems*, *13*(4), 1–39.
 57. Van Swol, L. M. (2007). Perceived importance of information: The effects of mentioning information, shared information bias, ownership bias, reiteration, and confirmation bias. *Group Processes & Intergroup Relations*, *10*(2), 239–256.
 58. Villaronga, E. F., Kieseberg, P., & Li, T. (2018). Humans forget, machines remember: Artificial intelligence and the right to be forgotten. *Computer Law & Security Review*, *34*(2), 304–313.
 59. Voigt, P., & Von dem Bussche, A. (2017). The eu general data protection regulation (gdpr). A Practical Guide (1st Ed.). *Cham: Springer International Publishing*, *10*(3152676), 10–5555.
 60. Wu, C., Wu, F., Huang, Y., & Xie, X. (2023a). Personalized news recommendation: Methods and challenges. *ACM Transactions on Information Systems*, *41*(1), 1–50.
 61. Wu, C., Wu, F., Huang, Y., & Xie, X. (2023b). Personalized news recommendation: Methods and challenges. *ACM Transactions on Information Systems*, *41*(1), 1–50.
 62. Xiao, B., & Benbasat, I. (2015). Designing warning messages for detecting biased online product recommendations: An empirical investigation. *Information Systems Research*, *26*(4), 793–811.

63. Yanisky-Ravid, S., & Hallisey, S. (2018). 'Equality and privacy by design': Ensuring artificial intelligence (AI) is properly trained & fed: A new model of AI data transparency & certification as safe harbor procedures. *Available at SSRN 3278490*.
64. Yu, B., Shao, J., Cheng, Q., Yu, H., Li, G., & Lü, S. (2018). Multi-source news recommender system based on convolutional neural networks. Paper presented at *the Proceedings of the 3rd International Conference on Intelligent Information Processing*, 17-23.
65. Zhang, Y., Wu, M., Tian, G. Y., Zhang, G., & Lu, J. (2021). Ethics and privacy of artificial intelligence: Understandings from bibliometrics. *Knowledge-Based Systems*, *222*, 106994.
66. Zhou, J., & Chen, F. (2023). AI ethics: From principles to practice. *Ai & Society*, *38*(6), 2693-2703.
67. Zhu, Z., & Van Roy, B. (2023). Scalable neural contextual bandit for recommender systems. Paper presented at *the Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 3636-3646.

[URL]

68. 한국일보 (2021). 아마존 채용 AI는 왜 남성을 우대했나. <https://url.kr/cqd5o8/>

〈부록〉

측정개념	측정문항
책무성	추천서비스 제공자는 이용자의 특성(소비·검색 이력, 연령, 성별 등)을 고려하여 콘텐츠의 위험성을 사전에 검증해야 한다
	인공지능 추천서비스 제공자는 시스템 운영 과정에서 기능적 오류, 오작동, 현행 법령 위반 등 부정적인 결과를 해결할 의무가 있다
	인공지능 추천서비스가 의도치 않게 공공의 이익에 반할 때 정부의 제도적 보완이 필요하다
설명요구권	추천서비스 제공자는 이용자에게 인공지능 알고리즘이 선별한 콘텐츠의 선별 기준을 알려 줄 필요가 있다
	추천서비스 제공자는 이용자에게 콘텐츠 제작자(생산자)가 인간 혹은 인공지능인지 알려 줄 필요가 있다
	추천서비스 제공자는 이용자에게 제작자(생산자)의 인공지능 서비스(맞춤형 콘텐츠 추천) 운영 과정에서 자신의 개인 정보가 어떻게 수집되고 활용되는지를 알려 줄 필요가 있다
	추천서비스 제공자는 이용자에게 인공지능 추천서비스를 이용하다 발생한 사고로 인한 피해나 불이익에 대해 설명할 필요가 있다
선택권	추천서비스 제공자는 이용자에게 인공지능 추천서비스가 제공하는 콘텐츠를 합리적으로 이용할 수 있는 선택권을 보장할 필요가 있다.
	추천서비스 제공자는 이용자에게 본인이 원하지 않는 추천 콘텐츠(뉴스, 광고)의 노출 정도를 선택 또는 변경할 수 있도록 해야 한다.
	추천서비스 제공자는 이용자에게 인공지능 추천서비스 이용정보에 대한 처분권(예: 수정, 삭제, 변경 요구)을 제공할 필요가 있다.
프라이버시 보호 지원 정도	인공지능 자동추천 뉴스의 빈번한 이용은 개인정보 유출을 낳을 것이다.
	인공지능 자동추천 뉴스의 빈번한 이용은 나를 불법정보에 노출 시킬 것이다.
	인공지능 자동추천 뉴스의 빈번한 이용은 프라이버시 또는 개인정보 침해(소비습관, 위치정보, 생체정보 등)가 발생할 수 있다
정보 편향성	인공지능 자동추천 뉴스의 빈번한 이용은 가치관의 편향을 낳을 것이다.
	(rev)인공지능 자동추천 뉴스는 편향적이지 않고 객관적이다.
정보 리터러시	나는 찾은 정보의 출처를 확인할 수 있다.
	나는 다른 자료들과 비교해 검색결과에서 믿음만한 정보를 구별할 수 있다.
	나는 광고성 정보를 구별해 낼 수 있다.
	나는 사기 또는 속이려는 사이트를 가려낼 수 있다.

● 저 자 소 개 ●



박민정 (Minjung Park)

현재 국립금오공과대학교 경영학과 조교수로 재직 중이다. 성신여자대학교에서 법학사, 이화여자대학교에서 빅데이터분석학 석사학위, 경영학과에서 경영정보시스템(MIS) 박사학위를 취득하였다. 주요 연구 분야는 개인정보보호, 정보보안, 인공지능, 블록체인 등이다.



채상미 (Sangmi Chai)

현재 이화여자대학교 경영대학 정교수로 재직 중이다. 이화여자대학교에서 학사, 서울대학교에서 경영학 석사학위를 취득하였으며, 미국 The State University of New York at Buffalo에서 경영학 박사학위를 취득하였다. 주요 연구 분야는 정보기술과 인간 행동에 관한 주요 이슈 IT와 조직 및 전략, 정보보안과 조직이며 최근에는 인공지능, 블록체인 및 디지털 자산 관련 연구를 활발히 진행 중이다.

〈 Abstract 〉

A Study on Information Bias Perceived by Users of AI-driven News Recommendation Services: Focusing on the Establishment of Ethical Principles for AI Services

Minjung Park^{*}, Sangmi Chai^{**}

AI-driven news recommendation systems are widely used today, providing personalized news consumption experiences. However, there are significant concerns that these systems might increase users' information bias by mainly showing information from limited perspectives. This lack of diverse information access can prevent users from forming well-rounded viewpoints on specific issues, leading to social problems like Filter bubbles or Echo chambers. These issues can deepen social divides and information inequality. This study aims to explore how AI-based news recommendation services affect users' perceived information bias and to create a foundation for ethical principles in AI services. Specifically, the study looks at the impact of ethical principles like accountability, the right to explanation, the right to choose, and privacy protection on users' perceptions of information bias in AI news systems. The findings emphasize the need for AI service providers to strengthen ethical standards to improve service quality and build user trust for long-term use. By identifying which ethical principles should be prioritized in the design and implementation of AI services, this study aims to help develop corporate ethical frameworks, internal policies, and national AI ethics guidelines.

Key words: Artificial Intelligence (AI), AI ethical principles, AI news recommendation system, Information bias, Knowledge management

* Kumoh National Institute of Technology

** Ewha Womans University