

“Made with AI” 투명성의 역설: AI 라벨링 유무가 진정성, 브랜드 신뢰, 브랜드 태도에 미치는 영향*

김 은 호

동아대학교 경영정보학과

문 창 근†

전북대학교 심리학과

본 연구는 생성형 AI가 광고 제작 전반에 보편적으로 활용되고 있는 환경에서, AI 사용 여부의 명시(AI 라벨링)가 소비자의 광고 반응과 브랜드 평가에 미치는 영향을 실증적으로 검증하였다. AI 라벨링이 광고에 대한 진정성 지각을 통해 브랜드 신뢰와 브랜드 태도로 이어지는 경로를 분석하고, 이 과정에서 소비자의 AI 리터러시 수준에 따른 조절효과를 확인하였다. 생성형 AI로 제작한 가상의 세탁세제 광고를 활용하여 AI 사용 여부를 제시한 조건과 제시하지 않은 조건을 비교하였다. 연구 결과, AI 라벨링이 제시된 광고는 그렇지 않은 광고에 비해 진정성 지각이 유의하게 낮았으며, 이러한 진정성 지각은 브랜드 신뢰를 거쳐 브랜드 태도에 부정적인 간접 효과를 미치는 것으로 나타났다. 또한 이 간접 효과는 소비자의 AI 리터러시 수준에 따라 조절되었는데, AI 리터러시가 높은 소비자일수록 AI 라벨링에 따른 진정성 저하와 그에 따른 브랜드 태도의 부정적 효과가 더욱 강화되었다. 본 연구는 생성형 AI 활용이 일상화된 시점에서 AI 사용 제시의 투명성이 항상 긍정적 효과로 이어지지 않을 수 있음을 보여주며, 투명성의 역설 관점에서 AI 광고 효과를 확장하였다는 점에서 학문적 의의를 지닌다. 또한 향후 AI 활용 광고 전략과 관련된 실무적 시사점을 제공한다.

주제어 : 생성형 AI 광고, AI 라벨링, 진정성, 브랜드 신뢰, AI 리터러시, 투명성의 역설

* 이 논문은 한국연구재단 4단계 BK21사업(전북대학교 심리학과)의 지원을 받아 연구되었음
(No.4199990714213).

†교신저자 : 문창근, 전북대학교 심리학과 석사, doorcg42@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright ©2026, Korean Society for Consumer and Advertising Psychology

서론

인공지능(AI)은 최근 마케팅과 광고 커뮤니케이션 전반에서 핵심 기술로 자리 잡으며 브랜드와 소비자 간 상호작용 방식을 근본적으로 변화시키고 있다. 초기 AI 활용은 고객 데이터 분석, 추천 시스템, 자동화된 고객 응대 등 기능적·보조적 역할에 국한되었으나, 생성형 인공지능(Generative AI)의 발전 이후 광고 크리에이티브 제작이라는 창의적 영역으로 빠르게 확장되었다(Huang & Rust, 2021). 특히 텍스트, 이미지, 영상 생성이 가능한 생성형 AI는 광고 제작비용과 시간을 절감하고, 대규모 개인화 메시지를 구현할 수 있는 가능성을 제공하며 마케팅 실무에서 강력한 도구로 인식되고 있다(Davenport et al., 2020).

실제로 글로벌 브랜드들은 생성형 AI를 활용한 광고 캠페인을 본격적으로 전개하고 있다. 생성형 AI는 광고 메시지, 이미지, 영상 콘텐츠를 자동으로 생성함으로써 크리에이티브 생산의 효율성과 확장성을 크게 향상시켰으며, 이러한 기술적 이점은 광고 산업 전반에서 AI 활용을 가속화하는 주요 요인으로 작용하였다(Kietzmann, Paschen, & Treen, 2018). 예를 들어, 하인츠는 이미지 생성 AI에 ‘케첩’을 입력하면 자사 병 형태가 자연스럽게 생성된다는 점을 활용해 AI조차 브랜드를 연상한다는 메시지를 전달하는 광고를 선보였다. 이러한 사례들은 생성형 AI가 단순한 제작 도구를 넘어 광고 아이디어의 핵심 장치로 기능할 수 있음을 보여준다(Kietzmann et al., 2018).

이러한 활용 사례가 증가할수록, 생성형 AI는 점차 차별적 크리에이티브 전략이 아닌 보편적 제작 수단으로 사용되고 있으며, 동시에 AI 활용의 무분별성이라는 새로운 문제를 야

기하고 있다. 최근 패션, 뷰티, 식음료 브랜드를 중심으로 AI 생성 이미지를 활용한 광고가 급증하면서, 광고 전반에서 유사한 시각적 스타일과 메시지 구조가 반복적으로 나타나고 있다. 또한 소비자와 업계 내부에서는 ‘AI 광고가 효율성만을 우선시한 결과물처럼 보인다’는 비판적 담론도 형성되고 있다. 이러한 맥락은 AI 활용이 전략적 판단 없이 확산될 경우, 브랜드 차별성을 약화시킬 수 있음을 시사한다(Castelo, Bos, & Lehmann, 2019).

최근 연구들은 마케팅 현장에서 AI가 전략적·윤리적 검토 없이 도입되는 사례가 증가하고 있으며, 이는 소비자 신뢰와 장기적 브랜드 관계 형성에 부정적인 영향을 미칠 수 있음을 지적한다(Castelo et al., 2019). 광고 영역에서 생성형 AI가 대량의 크리에이티브를 빠르게 생산하는 방식은 소비자에게 광고 피로감을 유발할 가능성을 높인다. 예컨대 디지털 플랫폼과 소셜 미디어 상에서는 AI가 자동 생성한 짧은 영상 광고나 배너 광고가 높은 빈도로 노출되며, 소비자는 서로 유사한 톤과 구성의 광고를 반복적으로 접하게 된다. 이러한 반복 노출은 광고 회피 행동을 강화하고, 주의 회피, 광고 회상 저하, 부정적 광고 태도 형성의 요인으로 작용할 수 있다(Campbell & Keller, 2003; Boerman et al., 2017).

광고 피로감과 더불어, 생성형 AI 광고의 확산은 ‘브랜드가 소비자를 얼마나 진정성 있게 고려하는가’라는 차원에서 새로운 의심을 불러일으킬 수 있다. 진정성(authenticity)은 소비자가 해당 브랜드를 신뢰하고 그 브랜드와 정서적 유대감을 형성하는 데 핵심적인 개념으로, 브랜드 커뮤니케이션이 진솔하고 인간적인 의도를 담고 있다고 지각될 때 강화된다(Morhart et al., 2015). 선행 연구에 따르면 진

정성은 브랜드에 대한 소비자의 신뢰 구축과 긍정적 태도 형성에 핵심 선행 요인으로 작용한다(Södergren, 2021). 그러나 AI가 광고 제작 전반을 대체하거나 인간의 개입이 최소화된 방식으로 활용될 경우, 소비자는 이를 효율성과 비용 절감을 우선한 선택으로 해석하며 브랜드 활동에 대한 성의를 낮게 평가할 가능성이 있다. 이러한 인식은 브랜드가 소비자를 진심으로 배려하지 않는다는 판단으로 이어질 수 있으며, 결과적으로 진정성이 훼손되었다고 지각될 가능성이 높다. 소비자가 진정성을 낮게 지각할수록, 소비자는 브랜드의 메시지와 행동을 신뢰하지 않게 되며, 진정성 및 브랜드 신뢰에 대한 의심은 광고 평가에 국한되지 않고 전반적인 브랜드에 대한 부정적인 평가로 전이될 수 있다(Aljarah, Ibrahim, & López, 2025; Morhart et al., 2015; Södergren, 2021).

이와 같은 문제 상황 속에서 전 세계적으로 AI 사용여부에 대한 명시를 요구하고 있다. AI가 생성한 콘텐츠임을 명시하는 것이 윤리적 책임과 소비자 보호 측면에서 중요하다고 주장하며(Araujo et al., 2020), 실제로 일부 국가에서는 AI로 제작된 광고에 대한 표시 의무화를 제도적으로 논의하거나 도입하고 있다. 특히 한국은 AI 관련법이 세계 최초로 전면 시행됐다(구유나, 2026). 이는 향후 광고 환경에서 AI 사용 여부의 명시(‘Made with AI’)가 선택이 아닌 필수 조건이 될 가능성을 시사한다(윤주혜, 2023). 그러나 AI 사용 여부에 대한 명시가 소비자 반응에 미치는 영향은 여전히 명확히 규명되지 않았다. 기존 연구들은 AI 사용을 고지할 경우 광고의 투명성은 높아질 수 있으나, 동시에 광고 신뢰성이나 정서적 공감이 저하될 수 있음을 보고한다(Baek, Kim, & Kim, 2024). 특히 AI 활용 초기에는 신기함과 혁신

성으로 인해 긍정적인 반응이 나타났으나, 오늘날과 같이 생성형 AI 광고가 범람하는 환경에서도 이러한 긍정적 효과가 지속될지는 의문이다(Castelo et al., 2019). 따라서 생성형 AI 사용이 확산된 환경과 AI 사용여부에 대한 명시가 요구 되는 상황에서 소비자가 해당 광고가 AI에 의해 제작되었음을 인지했을 때와 인지하지 못했을 때 광고와 브랜드를 어떻게 지각하는지 실증적으로 검토할 필요가 있다. 특히, AI 사용 명시가 의무화되는 환경에서도 AI 기반 광고가 여전히 브랜드에 긍정적인 태도와 신뢰를 형성할 수 있을지에 대한 재검토가 요구된다.

생성형 AI 기반 서비스와 콘텐츠가 일상적인 미디어 환경에 자주 노출되면서, 소비자들은 AI가 개입한 정보와 커뮤니케이션을 이전보다 더 쉽게 인식하고 평가할 수 있게 되었다(Lee et al., 2025). 소비자의 AI 리터러시 수준은 AI 활용 광고와 브랜드 커뮤니케이션을 지각하는 방식에도 중요한 영향을 미친다. 이는 동일한 AI 광고 환경에서도 소비자의 AI 이해 수준에 따라 광고의 진정성, 신뢰성, 그리고 브랜드 태도가 다르게 형성될 수 있음을 시사한다. 이에 본 연구는 생성형 AI로 제작한 광고를 중심으로, AI 사용 명시 여부가 소비자의 광고에 대한 진정성 지각과 이를 통한 브랜드 신뢰, 나아가 브랜드 전반 태도에 어떠한 영향을 미치는지를 분석하고, 소비자의 AI 리터러시 수준에 따라 차별적인 영향이 나타나는지 확인하고자 한다. 본 연구는 AI 활용 및 사용여부 명시 효과의 이해함으로써, 향후 AI 활용 광고의 방향 설정에 중요한 이론적 및 실무적 시사점을 제공할 것이다.

이론적 배경

AI 생성형 광고의 개념과 현황

AI 생성형 광고(AI-generated advertising)는 인공지능 알고리즘이 텍스트, 이미지, 비디오 등 광고 콘텐츠를 자동으로 생성하는 광고 제작 방식을 의미한다. 이러한 광고는 자연어처리(NLP)를 활용한 카피라이팅, 딥러닝 기반 이미지 생성, 영상 편집 및 합성 기술 등을 기반으로 하며, 인간의 직접적인 개입을 최소화한 상태에서 광고 크리에이티브를 생산할 수 있다는 점에서 기존 광고 제작 방식과 구별된다(Luo et al., 2019). 특히 생성형 AI는 대규모 데이터를 학습해 개인화된 메시지를 빠르게 생성할 수 있어, 비용 효율성과 확장성 측면에서 광고 실무에서 주목받고 있다.

이미 많은 브랜드들은 생성형 AI를 활용한 광고 캠페인을 진행하고 있다. 예를 들어, Coca-Cola는 생성형 AI 기술을 활용해 기존과 차별화된 광고 영상을 제작하며 기술 혁신 이미지를 강조하였고(Smith, 2023), 하인츠는 AI를 활용한 케첩 병 디자인 캠페인을 통해 소비자들의 높은 관심을 끌었다(Johnson, 2023). 이러한 사례들은 생성형 AI가 광고 제작 과정에서 창의성과 효율성을 동시에 확보할 수 있음을 보여주었다.

AI 생성형 광고의 확산과 함께 AI 사용 여부의 명시 문제가 중요한 쟁점으로 등장하였다(윤주혜, 2023). 유럽연합은 AI 규제 법안을 통해 AI가 생성한 콘텐츠를 명시할 것을 권고하고 있으며(European Commission, 2023), 미국 연방거래위원회역시 광고 투명성 강화를 위한 가이드라인을 제시하고 있다(FTC, 2023). 심지어 한국은 세계 최초로 AI 관련법을 전면

시행하고 있다(구유나, 2026). 이러한 규제 동향은 향후 AI 생성형 광고에서 AI 사용 여부의 명시가 점차 의무화될 가능성을 시사하며, 이는 소비자 반응과 광고 효과에 중대한 영향을 미칠 수 있는 요소로 작용할 것으로 예상된다.

AI 생성형 광고에 대한 소비자의 이중적 시각

최근 마케팅 영역에서는 기술 기반 전략의 활용이 점차 확대되면서 소비자들은 AI 생성형 광고에 대해 긍정적인 인식을 형성할 수 있다. AI가 중요한 의사결정 및 마케팅 장면에서 활용되는 배경에는 AI가 본질적으로 객관적이고 합리적인 판단을 수행할 것이라는 전제가 존재하기 때문이다(Russell, 2016; Russell & Norvig, 1995). 일반적인 미디어 콘텐츠에서 AI는 감정이 결여된 대신 고도의 논리성과 계산 능력을 지닌 존재로 묘사되어 왔으며(Hermann, 2023; Wu & Jing, 2021), 이러한 이미지들은 AI를 이성적 판단의 상징으로 인식하게 만들었다. 실제로 AI는 주어진 정보 하에서 성과를 극대화하는 행동을 선택하는 ‘합리적 에이전트(rational agent)’라는 개념에 기반해 정의된다(Russell, 2016; Russell & Norvig, 1995).

이러한 인식에 따라 소비자들은 AI 생성형 광고를 혁신적이고 기술적으로 진보된 광고 방식으로 지각한다. 혁신성은 AI 광고가 최신 기술을 반영한 새로운 시도라는 인식을 의미하며, 특히 기술 친화적인 소비자일수록 AI 광고를 더 긍정적으로 평가하는 경향이 있다(나정희, 2024). 또한 AI가 소비자 데이터를 기반으로 개인화된 메시지를 제공할 수 있다는 점에서, 소비자들은 맞춤형 경험에 대한 기대를 형성하게 되며, 이는 광고 호감도와 만

속도를 높일 수 있다(Luo et al., 2019). 더 나아가 알고리즘 기반 콘텐츠는 인간의 주관적 편견에서 상대적으로 자유롭다고 인식되어, 광고 메시지의 객관성과 정확성을 높게 평가 받을 가능성이 있다(Lee, 2018).

AI 생성형 광고에 대한 긍정적 반응이 항상 일관되게 나타나는 것은 아니다. 제품이나 마케팅 과정에서 과학적·기술적 요소를 강조하는 전략은 구매 의도와 제품 평가를 향상시키기도 하지만, 동시에 소비자의 심리적 저항이나 거부감을 유발할 수도 있다(Philipp-Muller, Costello, & Reczek, 2023). 이러한 맥락에서 광고 제작 과정에 생성형 AI를 활용하여 제작한 광고 역시 소비자에게 이중적 방식으로 지각될 가능성이 있다. 즉, AI 생성형 광고는 소비자에게 부정적인 인식을 유발할 가능성 또한 내포하고 있다.

소비자들은 AI가 감정을 경험하거나 이해하지 못하는 존재라고 인식하는 경향이 있으며, 이로 인해 AI가 제작한 광고를 인간적인 매력이 부족한 비인간적 결과물로 평가하기도 한다(Castelo et al., 2019; Bakpayev et al., 2022). 특히 감각적·정서적 요소가 중요한 쾌락적 영역에서는 AI의 개입이 광고 효과를 저해할 수 있으며(Longoni & Cian, 2022), 소비자들은 AI 광고를 기계적으로 생성된 콘텐츠로 인식함으로써 창의성이 부족하다고 평가하기도 한다(Raj et al., 2023). 또한 AI 생성형 광고임을 명시할 경우, 소비자들은 해당 광고의 진정성을 낮게 지각할 위험이 있다. 선행연구에 따르면 AI가 생성한 광고임을 공개했을 때, 소비자들은 정보 투명성 측면에서는 긍정적으로 평가하지만, 동시에 창의성과 진정성에 대한 기대는 감소하는 것으로 나타났다(나정희, 2024). 이는 AI 생성형 광고가 소비자에게 객관성·

효율성이라는 장점과 비인간성·진정성 결여라는 단점을 동시에 활성화시키는 이중적 지각 구조를 갖고 있음을 의미한다. 따라서 AI 생성형 광고는 소비자에게 긍정적·부정적 인식을 동시에 유발할 수 있으며, 이러한 인식은 광고가 속한 맥락, 과업의 성격, 그리고 AI 사용 명시 여부에 따라 달라질 가능성이 크다.

정보공개 효과: 투명성의 역할

생성형 AI 광고가 급속히 확산되면서, 광고 메시지가 AI로 생성·조작되었는지 여부를 명시적으로 공개해야 하는 상황이 현실화되고 있다. ‘AI 활용 명시 여부’가 소비자의 광고 판단에 직접적으로 개입하는 변수가 되는 만큼, AI 광고에 대한 정보공개가 이루어졌을 때 소비자 반응이 어떻게 나타날지 본격적으로 연구해야 할 시점이다. 정보공개는 일반적으로 윤리적·규범적으로 바람직한 커뮤니케이션으로 간주되지만, 소비자가 현재 시청하는 광고가 ‘AI가 만들었다’는 사실을 명시적으로 제시 받는 상황에서도 이를 여전히 긍정적으로 해석할지는 별개의 문제다. 실제로 AI 생성 콘텐츠를 공개했을 때 광고 평가가 오히려 불리해질 수 있음이 보고되며, 그 기제로 광고 신뢰성저하 등이 제시된다(Baek et al., 2024).

투명성(transparency)은 기업·브랜드가 이해관계자에게 제공하는 정보가 얼마나 공개되고, 명료하며, 정확한가라는 정보 품질의 지각으로 정의된다(Schnackenberg & Tomlinson, 2016). 투명성 신호는 소비자가 브랜드가 하는 커뮤니케이션 활동에 대한 진정성을 판단하는 핵심 단서로 작동하며, 진정성 지각은 다시 브랜드 신뢰(태도, 행동의도로 이어지는 긍정적

효과를 형성한다(Yang & Battocchio, 2021). 선행연구에 따르면, 생산 공정 · 원재료 등에 대한 정보 공개뿐 아니라 원가 · 이윤 · 물류비 등 구체적 비용구조를 공개하는 경우, 소비자는 기업이 ‘숨겨왔을 법한 민감한 정보’를 자발적으로 공개했다고 인식하기 때문에 투명성 신호가 더 강해질 수 있다(Yang & Battocchio, 2021). 이 관점에서 보면, 광고에서 ‘AI 사용’을 밝히는 행위 역시 소비자에게 진정성을 신호하는 투명성 전략으로 해석될 가능성이 있다.

설득지식모형(Persuasion Knowledge Model)에 따르면 정보를 공개하는 모든 상황이 항상 긍정적 효과만을 낳지 않는다는 점에 주목할 필요가 있다. 소비자는 설득 시도와 그 의도를 인식할수록, 메시지를 방어적으로 처리하거나 의도를 의심하는 방향으로 반응한다(Friestad & Wright, 1994). 디지털 환경에서의 인터넷 구전 의도, 리뷰, 인플루언서 마케팅 연구에서도 스폰서십 공개가 오히려 소비자의 설득지식을 활성화해 ‘상업적 의도가 있다’는 해석을 강화하고, 그 결과 메시지의 진정성과 신뢰를 낮추는 효과를 유발하였다(Boerman, Willemssen, & Van Der Aa, 2017; Kim, Maslowska, & Tamaddoni, 2019). 특히 긍정적 리뷰에서 스폰서십 공개가 결합될 때 ‘보상을 받고 쓴 말’이라는 의심이 강화되어 제품 태도와 구매의도가 더 크게 낮아지는 반면, 부정적 리뷰에서는 이러한 부정효과가 상대적으로 완화되는 패턴이 나타났다(Kim et al., 2019). 즉, 정보공개는 투명성의 신호가 될 수 있지만 동시에 ‘의도 추론’을 자극하여 역효과를 낳을 수도 있다.

이와 같은 현상을 ‘더 많이 공개할수록 오히려 덜 신뢰받는다’는 공개의 역설, 즉 투

명성의 역설(transparency paradox)이라고 한다(Steils, Martin, & Toti, 2022). 투명성의 역설은 공개가 소비자에게 정직성 신호로 작동하는 동시에, 다른 한편으로는 상업적 동기, 이미지 관리 의도 등과 같은 부정적인 생각을 강하게 떠올리게 만들어 신뢰를 훼손할 수 있다는 딜레마를 의미한다(Kim et al., 2019; Steils et al., 2022). 광고 제작시 ‘AI 사용’에 대한 명시적 공개는 소비자에게 정직 · 책임 · 투명성의 신호로 긍정적 반응을 유도할 수 있는 반면, 동시에 ‘광고가 기계적으로 만들어졌음’, ‘비용 절감 목적’, ‘브랜드가 진정정보다 효율을 택함’ 등과 같은 해석을 촉발해 부정적 반응을 유발할 가능성도 존재한다.

생성형 AI 광고에 대해 소비자들이 혁신성, 객관성, 개인화라는 긍정적 기대와 비인간성, 창의성 결핍, 진정성 저하라는 부정적 인식을 동시에 가질 수 있다는 ‘이중 지각’ 관점에서 보면(Dietvorst, Simmons, & Massey, 2015; Logg, Minson, & Moore, 2019; Longoni & Cian, 2022), AI 사용 여부에 대한 공개는 소비자의 긍정적 인식을 강화하기도 하지만, 반대로 부정적 인식을 더 강하게 자극할 위험도 있다. 공익광고 맥락에서 AI 활용의 공개가 소비자의 광고 태도와 기부 의도에 미치는 영향을 확인한 결과, 광고 콘텐츠가 AI에 의해 생성되었음을 공개할 경우 소비자는 광고에 대해 부정적인 태도를 형성하며, 이 과정에서 광고에 대한 낮은 신뢰성 지각이 매개 역할을 하는 것을 검증하였다. 즉, AI 사용의 공개는 낮은 신뢰성 지각으로 기부 의도에 부정적인 영향을 미치는 것을 확인하였다(Baek et al., 2024).

종합하면, AI 사용 여부 공개는 단순히 ‘더 투명하게 공개하면 더 좋아한다’는 단순한 가정이 아닌, 투명성의 역설이 발생할 수 있다

는 점을 검토해야 한다. 특히, 생성형 AI 활용이 무분별하게 팽배하고 소비자가 피로감을 느끼기 쉬운 현재의 시장 상황에서는(Fernandes & Oliveira, 2024; Tayeb et al., 2024), ‘AI 사용’을 명시하는 것이 새로움이나 혁신적인 활동으로 읽히기보다 상업적 효율 추구의 표식으로 해석될 가능성이 커진 상태이다. 따라서 광고 제작에 대한 AI 사용 여부 공개가 소비자에게 브랜드 커뮤니케이션 활동(예, 광고)에 대한 진정성, 브랜드의 신뢰, 브랜드 태도에 어떤 경로로 작동하는지, 그리고 어떤 조건에서 투명성의 역설이 강화되는지를 체계적으로 규명할 필요가 있다. 이러한 논의를 바탕으로 다음과 같은 가설을 도출하였다.

가설 1. AI 사용 여부를 명시적으로 밝힌 조건(AI 라벨링)에서 그렇지 않은 조건보다 광고에 대한 진정성을 낮게 지각할 것이다.

진정성 지각이 브랜드 신뢰에 미치는 영향

소비자가 브랜드 행위에 대한 진정성 지각은 브랜드 신뢰를 판단하는데 핵심적인 선행요인이다(Morhart et al., 2015; Napoli et al., 2014). 진정성은 브랜드의 행동과 커뮤니케이션이 일관되고 성의 있으며, 단기적인 상업적 이익보다 소비자에 대한 책임과 진실성을 우선시하고 있다고 인식되는 정도를 의미한다(Beverland, 2005; Morhart et al., 2015). 이러한 진정성 지각은 소비자가 브랜드의 동기와 의도를 해석하는 중요한 기준점으로 작용한다.

소비자는 브랜드가 전달하는 메시지와 실제 행동 간의 정합성을 바탕으로 해당 브랜드의 행위가 얼마나 진실되고 신뢰할 만한 존재인지를 평가한다. 즉, 브랜드 커뮤니케이션이 일

관되고 맥락에 부합할수록 소비자는 브랜드가 기회주의적으로 행동하지 않을 것이라는 기대를 형성하게 되며, 이 과정에서 진정성 지각은 브랜드 신뢰 형성의 핵심 단서로 기능한다(Napoli et al., 2014). Morhart 등(2015)은 진정성을 지속성, 신뢰성, 진실성, 자연스러움 등의 차원으로 개념화하고, 이러한 진정성 지각이 브랜드 신뢰와 강한 정적 관계를 갖는다는 점을 실증적으로 확인하였다. 소비자가 브랜드의 활동을(예, 광고 등) 진정성 있게 인식할수록 브랜드가 약속을 이행하고 소비자의 이익을 해치지 않을 것이라는 기대가 강화되며, 이는 곧 브랜드 신뢰로 이어진다. Kim, Morris, & Swait(2008)는 브랜드 신뢰를 브랜드의 행동에 대해 일관되고 긍정적인 기대를 형성하는 심리적 상태로 정의하며, 이러한 기대는 브랜드가 얼마나 정직하고 성의 있게 행동하는지에 대한 소비자의 판단에 기반한다고 설명한다. 다시 말해, 진정성 지각은 브랜드의 의도와 동기를 긍정적으로 해석하도록 만들어 브랜드에 대한 신뢰를 형성하는 토대를 제공한다.

반대로 브랜드 커뮤니케이션이 전략적 이미지 관리나 형식적 대응으로 해석될 경우, 소비자는 브랜드의 진의를 의심하게 되며 진정성 지각은 약화된다. 이러한 상황에서는 브랜드의 행동이 소비자를 존중하기보다 효율성이나 이윤 극대화를 우선시한다고 인식되어 브랜드 신뢰가 저해될 가능성이 커진다(Beverland & Farrelly, 2010). Morhart et al.(2015)은 브랜드 행동이 ‘연출된’ 것으로 인식될 때 진정성과 신뢰가 동시에 약화된다는 점을 지적하였다. 따라서 생성형 AI로 만든 광고가 무분별하게 확산되어 제공되는 현시점에서 소비자들에게 생성형 AI에 의해 만들어진 광고라고 제시하

는 것은 브랜드 행위에 대한 진정성에 의문이 제기되며 이는 신뢰 형성에 부정적인 영향을 미칠 수 있다. 따라서 이러한 논의를 바탕으로 다음과 같은 가설을 도출하였다.

가설 2. 소비자의 광고에 대한 진정성 지각은 브랜드 신뢰에 정적 영향을 미칠 것이다.

브랜드 신뢰가 브랜드 태도에 미치는 영향

브랜드 신뢰는 소비자의 전반적인 브랜드 태도를 형성하는 핵심적인 선행 변수이다 (Chaudhuri & Holbrook, 2001; Munuera-Aleman, Delgado-Ballester, & Yague-Guillen, 2003). 브랜드 태도는 특정 브랜드에 대해 소비자가 갖는 총체적인 평가이자 호의 또는 비호의적 정서 반응을 의미하며, 이는 구매의도와 행동적 충성도로 이어지는 중요한 심리적 기초를 이룬다 (Batra & Homer, 2004). 신뢰는 이러한 태도 형성 과정에서 브랜드에 대한 불확실성을 감소시키고, 긍정적인 정서 반응을 안정적으로 유지하도록 돕는 역할을 한다. 선행연구들은 브랜드에 대한 신뢰가 높을수록 소비자가 해당 브랜드를 더 긍정적으로 평가하고, 장기적인 관계를 유지하려는 태도를 형성한다는 점을 일관되게 보여준다. Chaudhuri 와 Holbrook (2001)은 브랜드 신뢰가 브랜드 태도의 핵심 구성요소로 작용하며, 신뢰를 기반으로 형성된 태도가 감정적 애착과 행동적 충성도로 확장된다고 설명하였다. Munuera-Aleman et al.(2003)은 브랜드 신뢰가 호의적인 브랜드 태도 형성에 직접적인 영향을 미치며, 이를 통해 브랜드에 대한 애정과 충성도가 강화된다는 점을 확인하였다.

브랜드 신뢰는 일단 형성되면 브랜드 평가

전반에 안정성을 부여하지만, 동시에 매우 취약한 특성을 지닌다. Morgan 과 Hunt(1994)는 신뢰를 관계 마케팅의 핵심 메커니즘으로 제시하며, 신뢰가 유지될 때 장기적인 관계가 가능하다고 보았다. 그러나 브랜드 행동의 불일치나 부정적 경험이 발생할 경우, 소비자의 신뢰가 빠르게 붕괴될 수 있음을 또한 지적하였다(Erdem & Swait, 2004). 더 나아가 손상된 신뢰를 회복하는 데에는 상당한 시간과 조직적 자원이 요구되며, 이 과정에서 소비자의 브랜드 태도 역시 장기적으로 부정적 영향을 받을 수 있다고 하였다(Sirdeshmukh, Singh, & Sabol 2002). 따라서 브랜드 신뢰는 브랜드 태도의 안정성과 방향성을 결정짓는 핵심 변수로 기능하며, 신뢰가 약화될 경우 소비자의 전반적인 브랜드 태도 역시 부정적으로 변화할 가능성이 크다. 이러한 맥락에서 생성형 AI를 활용한 광고 전략이 소비자의 신뢰를 저해하는 방향으로 작용할 경우, 그 영향은 광고에 대한 평가를 넘어 브랜드 전체에 대한 태도 저하로 확산될 것으로 예상할 수 있다. 이러한 논의를 바탕으로 다음과 같은 가설을 도출하였다.

가설 3. 브랜드에 대한 신뢰는 브랜드 태도에 정적 영향을 미칠 것이다.

AI 리터러시 수준에 따른 소비자의 반응 차이

디지털 환경의 급속한 변화와 AI 기술의 일상화로 인해 소비자의 AI 리터러시 수준은 지속적으로 향상되고 있다. AI 리터러시는 인공지능의 작동 방식과 활용 목적을 이해하는 것을 넘어, 기술의 한계와 잠재적 위험 요소를 인식하고 이를 비판적으로 해석할 수 있는 능

력을 의미한다(Long & Magerko, 2020). 선행연구들은 AI 리터러시를 단순한 기술 숙련도가 아닌, AI의 작동 원리와 한계를 이해하고, AI와 효과적으로 소통·협업하며, 산출 결과를 비판적으로 평가하고 이를 목적에 맞게 활용할 수 있는 종합적 역량으로 정의하였다(Aoun, 2017; Long & Magerko, 2020). 또 다른 선행 연구 역시 AI 리터러시를 디지털 리터러시의 확장 개념으로 보고, 개인이 AI 기술을 비판적으로 평가하고 일상 및 업무 환경에서 도구로 활용할 수 있는 능력의 집합으로 개념화하고 있다(김민진, 김미애, 노환호, 김범수, 2024; 이유미, 박윤수, 2021; 이진, 오현정, 2024).

이러한 역량은 과거 전문가 집단에 국한된 전문 지식이 아니라, 생성형 AI 기반 서비스와 콘텐츠가 일상적인 미디어 환경에 깊숙이 자리매김하면서 일반 소비자 수준으로 빠르게 확산되고 있다. 특히 소비자들은 AI가 개입한 정보 생산과 커뮤니케이션 과정을 이전보다 쉽게 인식하고 평가할 수 있게 되었으며, 이는 AI 관련 메시지와 콘텐츠를 해석하는 방식 전반에 구조적 변화를 가져왔다(Lee, Choi, Jeong, & Ko, 2025). 따라서 AI 리터러시는 AI로부터 지각되는 혜택과 위험, 나아가 태도 및 지속사용의도 형성 과정 전반에 영향을 미치는 핵심적인 인지적 자원으로 기능한다.

AI 리터러시 수준에 따른 소비자 행동 차이에 관한 선행연구들은 크게 두 가지 흐름으로 정리될 수 있다. 첫째, AI 리터러시가 높을수록 소비자는 AI 기술에 대해 보다 합리적이고 통제감 있는 판단을 내리며, 서비스의 표면적 특성이나 의인화 단서에 덜 의존하는 경향을 보인다(김민진 외, 2024). 예를 들면, 생성형 AI 서비스를 대상으로 한 연구에서는 서비스의 개방성이나 친화적 특성이 인지된 혜택과 지

속사용의도에 미치는 긍정적 효과가 AI 리터러시가 낮은 집단에서 더 강하게 나타나는 조절효과가 확인되었다. 이는 AI 리터러시가 낮은 소비자가 서비스의 친화적·개방적 특성에 더 크게 반응하는 반면, AI 리터러시가 높은 소비자는 이러한 단서에 상대적으로 덜 의존함을 시사한다.

둘째, 일부 연구들은 리터러시 수준이 높을수록 소비자가 오히려 비판적이고 부정적인 평가가 강화될 가능성을 제시한다. 예를 들어, 소셜미디어 환경에서 브랜드의 개방성이 높은 활동에 대해 뉴미디어 리터러시, 특히 기능적 리터러시가 높은 소비자의 경우, 브랜드의 개방성이 높을수록 해당 브랜드 활동을 부정적으로 평가하는 경향이 나타났다(Song & Kim, 2025). 이는 기술과 미디어 구조에 대한 이해도가 높은 소비자가 브랜드의 전략적 의도나 설득 목적을 보다 명확히 인식하고, 이를 비판적으로 해석할 가능성이 높기 때문이다. 유사하게 AI 리터러시가 높은 소비자는 AI 기반 메시지의 설계 의도와 설득 전략을 보다 정교하게 인식하며, AI 활용을 효율성 중심의 전략적 선택으로 평가하는 경향을 보인다(Shin, 2021). 반면 AI 리터러시가 낮은 소비자는 AI 활용 여부를 충분히 인식하지 못하거나, 기술적 참신함에 의해 메시지를 상대적으로 덜 비판적으로 수용하는 경향을 보인다(장천우, 김성희, 정용국, 2024; Long & Magerko, 2020). 이러한 차이는 동일한 AI 활용 커뮤니케이션 환경에서도 소비자의 AI 이해 수준에 따라 메시지에 대한 신뢰성, 진정성, 그리고 브랜드 태도가 상이하게 형성될 수 있음을 시사한다. 즉, AI 리터러시가 높은 소비자가 낮은 소비자에 비해 AI 사용에 대해 진정성 지각을 더 낮게 판단할 것이고, 이는 브랜드 신뢰와 브

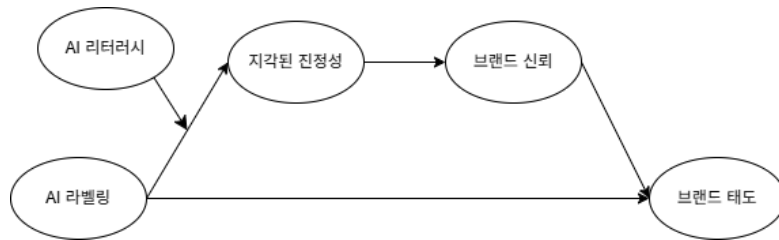


그림 1. 연구 모형

랜드 태도에 부정적인 영향을 미칠 것이다. 따라서 소비자의 AI 리터러시 수준은 생성형 AI를 활용한 광고와 브랜드 커뮤니케이션 효과를 차별적으로 설명할 수 있는 핵심 조절 변수로 작동할 가능성이 높다. 이러한 가정들을 바탕으로 다음과 같은 가설을 도출하였다.

가설 4. AI 사용 여부 제시에 따른 브랜드 태도에 진정성 지각과 브랜드 신뢰의 순차적 매개효과는 AI 리터러시 수준에 의해 조절될 것이다.

실 험

자료 수집 및 변수 측정

본 연구는 전문조사업체 코리아리서치의 패널을 대상으로 2025년 4월 7일부터 4월 14일 까지 총 8일 동안 수집한 자료를 사용하였다. 본 조사는 AI 사용에 대한 내용(AI 라벨링)을 재인하는 측정 문항을 통과하지 못한 13명을 제외한 전체 175명의 성인남녀(남성: 87명, 여성: 88명)의 표본을 분석에 사용하였다. 표본의 평균 연령은 40.77세(20세~58세)였다.

본 연구는 생성형 AI를 통해 가상 브랜드(CLEATIC)의 세탁세제를 지면 광고 형태의 두

조건으로 제작하였다. 광고에 제시된 제품은 캡슐형태의 세제였으며 ‘수십 년간 세탁 기술을 선도해온 CLEATIC. CLEATIC의 축적된 성분 배합 노하우로 99.9% 세균 제거, 보이지 않는 오염까지 깨끗하게. 세탁도 기술입니다’라는 광고 문구를 공통적으로 삽입하였다. 다만 하나의 광고자극에는 AI 사용여부를 제시하지 않았고, 나머지 하나는 ‘이 광고는 AI로 만들어졌습니다’라는 문구를 제시하였다(부록 A). 피험자들에게 해당 광고를 제시한 후, 해당 연구에 적합하게 내용을 수정한 광고에 대한 진정성(Lee, Kim, & Kim, 2024), 브랜드 신뢰(Munuera-Aleman et al., 2003), 브랜드 태도(MacKenzie et al., 1986), AI 리터러시(이진, 오현정, 2024)를 7점 리커트 척도(1:전혀 아니다~7:매우 그렇다)로 측정하였다(표 1). 그 외 제품관여도, 브랜드 인지여부 등을 측정하였다. 주요 측정변인에 대한 구체적인 내용은 표 1에 제시하였다.

연구 결과

AI 라벨링 유무에 따른 진정성 지각, 브랜드 신뢰, 브랜드 태도의 차이

AI 라벨링 유무에 따른 주요 변수들의 평균 차이를 검증하기 위해 독립표본 t-검증을 실시하였다. 먼저, 진정성 지각에 대한 분석 결과,

표 1. 측정변수의 측정문항 및 신뢰도 값

측정변수	측정문항	신뢰도
진정성 (Lee, Kim, & Kim, 2024)	1. 이 광고는 진정성이 느껴진다 2. 이 광고내용은 순수하다 3. 이 광고는 진심에서 우러난 것으로 생각된다	.943
브랜드 신뢰 (Munuera-Aleman et al., 2003)	1. 이 브랜드를 신뢰할 수 있다 2. 이 브랜드에 믿음이 간다 3. 이 브랜드는 의지가 된다 4. 이 브랜드는 정직할 것이다	.963
브랜드 태도 (MacKenzie et al., 1986)	1. 이 브랜드가 좋다 2. 이 브랜드가 마음에 든다 3. 이 브랜드에 흥미가 있다	.949
AI 리터러시 (이진, 오현정, 2024)	1. 나는 문제 해결을 위해 AI와 소통할 수 있다 2. 나는 AI로 문제를 해결하기 위해 사람 간 소통이 중요하다는 것을 알고 있다 3. 나는 학술적 글(예: 논문)에서 내가 필요한 정보를 찾아낼 수 있다 4. 나는 오픈 소스 커뮤니티(예: GitHub, Stack Overflow)에서 내가 필요한 정보를 찾아낼 수 있다	.889

AI 라벨링 무 집단($M = 4.484$, $SD = 1.041$)이 라벨링 유 집단($M = 3.895$, $SD = 1.310$)에 비해 지각된 진정성이 유의하게 높은 것으로 나타났다($t = 3.28$, $p < .005$). 다음으로 브랜드 신뢰에 대한 분석 결과, AI 라벨링 무 집단($M = 4.598$, $SD = 1.189$)이 AI 라벨링 유 집단($M = 3.935$, $SD = 1.353$)보다 유의미하게 높은 수준의 브랜드 신뢰를 보였다($t = 3.261$, $p < .005$). 마지막으로 브랜드 태도에 대한 분석 결과, AI 라벨링 무 집단($M = 4.740$, $SD = 1.104$)이 AI 라벨링 유 집단($M = 4.219$, $SD = 1.382$)에 비해 브랜드 태도가 유의미하게 높은 것으로 나타났다($t = 2.733$, $p < .005$).

AI 라벨링 유무 따른 지각된 진정성과 브랜

드 신뢰가 브랜드 태도에 미치는 순차적 매개 효과 검증

AI 라벨링 유무가 브랜드 태도에 미치는 영향을 검증하기 위해 진정성 지각과 브랜드 신뢰를 순차적 매개변수로 설정한 PROCESS Macro Model 6을 활용하여 10,000회의 부트스트래핑을 실시하였다.

분석결과, AI 라벨링 유무는 진정성 지각에 유의미한 부적 영향을 미치는 것으로 나타났다($M_{라벨유} = 3.89(1.30)$, $M_{라벨무} = 4.48(1.04)$, $t=3.27$, $\beta = -.590$, $p < .005$). 이는 AI 사용에 대한 명시 라벨이 존재할 때, 진정성에 대한 지각이 유의하게 감소함을 의미한다. 다음으로 브랜드 신뢰를 종속변수로 한 분석 결과, 지각된 진정성은 브랜드 신뢰에 유의미한 정

적 영향을 미쳤다($\beta = .871, p < .001$). 따라서 가설 1과 가설 2는 지지되었다. 그러나 AI 라벨링 유무의 직접효과는 유의하지 않았다($\beta = -.150, p = .214$).

다음으로 브랜드 태도를 종속변수로 한 결과에서는 진정성 지각($\beta = .221, p < .001$)과 브랜드 신뢰($\beta = .710, p < .001$)가 브랜드 태도에 유의미한 정적 영향을 미치는 것으로 나타났다으나 가설 3은 지지되었다. 그러나 AI 라벨링 유무의 직접효과는 유의하지 않았다($\beta = .081, p = .392$).

마지막으로 AI 라벨링 유무가 진정성 지각과 브랜드 신뢰를 매개하여 브랜드 태도에 미치는 간접효과를 분석한 결과, AI 라벨링 유무 → 진정성 지각 → 브랜드 태도에 미치는 간접효과는 유의하였다(Index = $-.131$, bootSE = $.061$, CI = $-.275, -.036$). 또한 라벨링 유무 → 진정성 지각 → 브랜드 신뢰 → 브랜드 태도에 대한 순차적 매개효과 역시 유의미하게 나타났다(Index = $-.364$, bootSE = $.119$, CI = $-.611, -.145$). 반면, AI 라벨링 유무 → 브랜드 신뢰 → 브랜드 태도에 대한 매개경로는 신뢰 구간에 0을 포함하여 유의하지 않았다. 전체 간접효과는 통계적으로 유의미한 것으로 확인되었다(Index = $-.601$, bootSE = $.118$, CI = $-.956, -.262$).

이러한 결과는 AI 사용여부에 대한 라벨링이 브랜드 태도에 직접적인 영향을 미치기보다는, 광고에 대한 진정성 지각을 감소시키고 이것이 브랜드 신뢰에 부정적 영향을 미치는 순차적 매개경로를 통해 브랜드 태도에 부정적인 영향을 미친다는 점을 시사한다.

AI 리터러시의 조절된 매개효과 검증

위에 검증한 AI 라벨링 유무가 지각된 진정

성과 브랜드 신뢰를 매개해 브랜드 태도로 이어지는 간접 효과가 AI 리터러시에 의해 조절되는지 확인하기 위해 Hayes(2022)의 PROCESS macro(Model 83)를 활용하여 10,000회의 부트스트래핑을 실시하였다.

분석결과, 첫째, AI 사용 라벨 유무는 지각된 진정성에 통계적으로 유의미한 부정 영향을 미치는 것으로 나타났다($t=3.27, \beta = -3.08, p < .005$). 따라서 가설 1이 반복적으로 지지되었다. 둘째, AI 리터러시와 AI 라벨 유무의 상호작용 항이 지각된 진정성에 유의한 영향을 미치는 것으로 나타났다($\beta = -.520, p < .05$). 지각된 진정성은 브랜드 신뢰($\beta = .871, p < .001$) 및 브랜드 태도($\beta = .221, p < .001$)에 정적 영향을 미치며, 브랜드 신뢰는 브랜드 태도($\beta = .710, p < .001$)에 정적 영향을 미치는 것으로 나타났다. 따라서 가설 2와 가설 3 역시 지지되었다. 반면, AI 라벨 유무가 브랜드 신뢰($p = .2146$) 및 브랜드 태도($p = .3927$)에 미치는 직접 효과는 통계적으로 유의하지 않았다. 그러나 지각된 진정성을 매개로 하는 경로(Index = $-.1152$)와 지각된 진정성 및 브랜드 신뢰를 순차적으로 매개하는 경로(Index = $-.3217$) 모두 조절된 매개지수의 95% 신뢰구간에 0을 포함하지 않아 통계적으로 유의한 것으로 나타났다(표 2). 마지막으로, AI 사용 여부 제시에 따른 브랜드 태도에 광고에 대한 진정성 지각과 브랜드 신뢰가 순차적으로 매개하며, 이 간접효과를 소비자의 AI 리터러시 수준이 조절할 것이라는 가설 4를 검증한 결과, 조절된 매개효과는 통계적으로 유의하였다(Index = $-.321$, bootSE = $.150$, CI = $-.639, -.051$). 따라서 가설 4도 지지되었다. AI 라벨링과 AI 리터러시의 상호작용이 지각된 진정성에 미치는 영향의 구체적인 양상을 파

표 2. AI 라벨링과 AI 리터러시 및 두 변인의 상호작용이 지각된 진정성, 브랜드 신뢰, 브랜드 태도에 미치는 영향

변인	지각된 진정성 (M1)	브랜드 신뢰 (M2)	브랜드 태도 (Y)		
AI 라벨링(X)	-.604***	-.150	.081		
AI 리터러시(W)	1.292***				
상호작용(X · W)	-.520***				
지각된 진정성(M1)		.871***	.221***		
브랜드 신뢰(M2)			.710***		
<i>F-value</i>	12.061***	191.602***	236.866***		
<i>R2</i>	.175	.690	.806		
	조절된 매개 지수	Index	bootSE	Lower~Upper	Upper
		-.321	.150	-.639~-.051	-.051

*** $p < .001$

표 3. AI 리터러시 수준에 따른 AI 라벨링이 진정성에 미치는 영향

경로	조절변수 (AI리터러시 수준)	Effect	t	Boot SE	LLCI	ULCI
AI 라벨링 → 진정성	-1SD(낮음)	-.187	-.750	.250	-.681	.306
	Mean	-.604	-3.365	.180	-.958	-.250
	+1SD(높음)	-1.021	-4.061	.251	-1.517	-.525

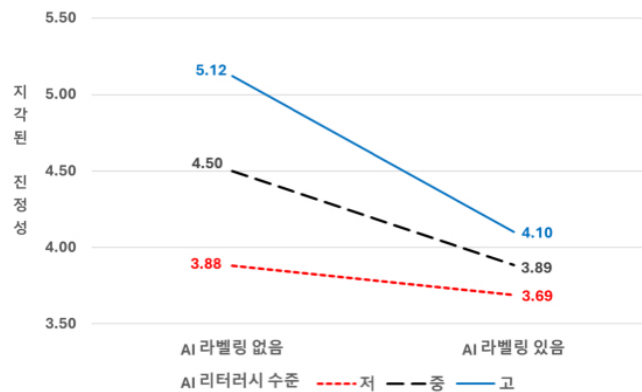


그림 2. AI 리터러시 수준에 따른 AI 라벨링이 진정성 지각에 미치는 영향

표 4. AI 리터러시 수준에 따른 AI 라벨링이 브랜드 태도에 미치는 매개효과

간접효과의 경로	조절변수 (AI리터러시 수준)	간접효과	Boot SE	LLCI	ULCI
AI 라벨링 →진정성→ 태도	-1SD(낮음)	-.042	.058	-.166	.066
	Mean	-.134	.058	-.266	-.041
	+1SD(높음)	-.226	.095	-.444	-.073
	조절된 매개지수(Index)		Boot SE	LLCI	ULCI
		-.115	.066	-.271	-.015
AI 라벨링 →진정성→ 신뢰→태도	-1SD(낮음)	-.116	.143	-.387	.180
	Mean	-.373	.107	-.591	-.173
	+1SD(높음)	-.631	.176	-.998	-.305
	조절된 매개지수(Index)		Boot SE	LLCI	ULCI
		-.322	.150	-.640	-.052

주) AI 라벨링: AI 표기 유무, 진정성: 지각된 진정성, 신뢰: 브랜드 신뢰, 태도: 브랜드 태도

약하기 위해 AI 리터러시 평균값을 기준으로 $\pm 1SD$ 수준에서 조건부 효과 분석을 수행하였다. 분석 결과, AI 리터러시 수준이 평균부터 높은 경우(+1SD)에는 AI 라벨이 지각된 진정성을 유의하게 감소시켰으나($\beta = -1.021, p < .001$), AI 리터러시가 낮은 경우(-1SD)에는 그 영향이 통계적으로 유의하지 않았다($p = .454$). 이는 소비자의 AI 리터러시 수준이 높을수록 AI 라벨 사용에 따른 진정성을 낮게 지각하는 것으로 나타났다(표 3, 그림 2). 즉, 소비자의 AI 리터러시 수준이 높아질수록 광고 제작시 AI 사용에 대해서 진정성에 대해 더 민감하게 지각함을 시사한다. 이는 이후 브랜드의 신뢰와 브랜드 태도에 부정적인 영향을 미친다는 것을 확인하였다. 각 경로에 대한 구체적인 결과는 표 4에 제시하였다.

결론 및 논의

본 연구는 생성형 AI가 소비자의 일상생활 뿐만 아니라 브랜드의 다양한 커뮤니케이션 활동 전반에 깊숙이 관여하고 있는 현시점에서, 생성형 AI로 제작된 광고에 대한 소비자 반응을 실증적으로 검증하고자 하였다. 생성형 AI 활용이 본격화되면서 AI 사용 여부에 대한 고지와 라벨링이 권고되고 관련 규제 논의가 확대되고 있는 상황에서, 생성형 AI 활용 초기 연구에서 보고된 긍정적 반응이 동일하게 유지될 것인지에 대해서 재검토가 필요하다(나정희, 2024; Lee, 2018; Luo et al., 2019). 이는 소비자가 생성형 AI 광고를 혁신성과 효율성의 신호로 해석하는 한편, 동시에 비인간성과 진정성 결여의 단서로 인식할 가능성을 내포하고 있기 때문이다. 이러한 문제의식을

바탕으로 본 연구는 생성형 AI를 활용해 제작한 광고에서 AI 라벨링 유무가 광고에 대한 진정성 지각, 브랜드 신뢰, 브랜드 태도에 미치는 영향을 검증하고, 이 과정에서 소비자의 AI 리터러시 수준이 어떠한 조절적 역할을 수행하는지를 분석하였다.

연구 결과, AI 라벨링이 제시된 조건에서는 그렇지 않은 조건에 비해 광고에 대한 진정성 지각이 유의하게 낮아졌으며, 이러한 낮은 진정성 지각은 브랜드 신뢰를 약화시키고 궁극적으로 브랜드 태도를 부정적으로 형성하는 것으로 나타났다. 즉, AI 라벨링 유무는 진정성 지각과 브랜드 신뢰를 순차적으로 매개하여 브랜드 태도에 영향을 미치는 간접 효과를 보였다. 더 나아가 본 연구는 이러한 간접 효과가 소비자의 AI 리터러시 수준에 따라 차별적으로 나타나는 조절된 매개효과를 확인하였다. 구체적으로, 소비자의 AI 리터러시 수준이 높아질수록 생성형 AI를 활용했다는 사실은 브랜드 커뮤니케이션의 진정성을 더욱 낮게 지각하게 만들었으며, 그 결과 AI 라벨링에 따른 브랜드 태도의 부정적 간접 효과가 강화되는 것으로 나타났다. 이는 생성형 AI 활용에 대한 소비자의 이해도가 높아질수록 해당 기술을 혁신적 도구라기보다는 효율성 중심의 전략적 선택으로 인식할 가능성이 커지면서 진정성을 의심함을 시사한다.

본 연구는 생성형 AI 중심의 대전환이 본격화된 시점에서(김정영, 202; 박재곤, 2025) AI 활용을 통한 브랜드 커뮤니케이션 활동에 대한 소비자 반응을 검증함으로써 학문적 의의를 갖는다. 기존 연구들은 AI 생성형 광고가 혁신성, 맞춤화, 객관성에 대한 기대를 유발할 수 있다고 보고해왔다. 그러나 동시에 AI 광고는 인간적 매력의 결여로 인한 비인간성 지

각(Bakpayev et al., 2022)이나 인간의 독창성과 감성을 충분히 반영하지 못하는 창의성 결핍으로 평가될 가능성 또한 내포한다(Raj et al., 2023). 본 연구는 이러한 상반된 인식이 공존하는 상황에서 AI 라벨링과 소비자 리터러시 수준이라는 경계조건을 통해 생성형 AI 광고에 대한 부정적 반응이 언제, 어떻게 증폭되는지를 실증적으로 제시했다는 점에서 AI 활용에 대한 소비자 반응에 대한 이해를 확장했다는 의의가 있다.

또한 본 연구는 AI 라벨링이 제시되면 제시되지 않았을 때보다 진정성 지각이 낮아지는 것을 확인했다. 특히 AI 리터러시가 높아질수록 AI 라벨링 사용에 따른 진정성 지각에 대한 민감도가 높아짐으로써 상대적으로 AI 라벨링 제시의 경우 그렇지 않은 조건에 비해 진정성을 현저히 낮게 지각하였다. 이러한 결과는 높은 AI 리터러시 수준을 가진 소비자들에게는 생성형 AI 기술 활용이 효율성과 비용절감 중심의 전략적 선택으로 해석되어 부정적으로 판단될 가능성이 높음을 시사한다. 이는 생성형 AI에 대한 소비자 반응이 기술 확산 단계에 따라 달라질 수 있음을 시사하며, 시기적 맥락을 고려한 리터러시 개념의 재검토 필요성을 제안한다는 점에서 학문적 의미를 지닌다(장천우 외, 2024; Shin, 2021; Song & Kim, 2025).

실무적 측면에서 본 연구는 생성형 AI 광고에 대한 의무적 고지 및 규제 강화 가능성이 논의되는 상황에서, AI 라벨링이 소비자 반응에 미치는 영향을 예측했다는 점에서 시사점을 제공한다. 연구 결과에 따르면, 생성형 AI 사용에 대한 명시가 소비자의 부정적 반응을 유발할 수 있음을 보여주었다. 이는 단순히 AI 사용 사실을 숨기거나 회피해야 함을 의미하

는 것이 아니라, 오히려 기업과 브랜드가 왜 생성형 AI를 활용하는지, 해당 기술이 브랜드의 가치와 커뮤니케이션 목표에 어떠한 방식으로 기여하는지를 보다 근본적으로 고민할 필요가 있다는 점을 제안한다. 생성형 AI 활용이 브랜드 가치와 관계없이 효율성과 경제적 이점만을 위해 사용될 경우, 브랜드가 잃게 되는 것이 무엇인가에 대한 고민을 제안한다는 점에서 실무적 함의를 갖는다.

본 연구의 학문적, 실무적 의의에도 불구하고 본 연구는 한계점과 이를 보완한 추후연구의 필요성이 존재한다. 우선 본 연구는 세탁 세제라는 단일 제품과 가상의 브랜드를 활용하였다. 하지만 소비자는 제품을 쾌락적 제품과 실용적 제품으로 구분하여 상이한 기대를 형성한다(Voss, Spangenberg, & Grohmann, 2003).

소비자는 쾌락적 제품은 소유 및 사용함으로써 따뜻함, 즐거움, 감성적 만족, 감각적 자극을 주는 것을 중요하게 생각한다. 한편, 실용적 제품은 소비자가 기능성, 성능, 문제 해결 능력을 중요하게 여기기 때문에 기능과 효율성을 강조한다(Chattalas & Takada, 2013). 생성형 AI 광고와 같이 과학적·기술적 이미지가 강조되는 메시지는 제품군에 따라 다르게 평가될 가능성이 있다(Philipp-Muller, Costello, & Reczek, 2023).

뿐만 아니라 AI 사용에 대한 정당성 여부가 영향을 미칠 수도 있을 것이다. AI를 활용하여 광고를 제작한 이유가 전체적인 비용 절감을 위한 것이며, 그 비용 절감을 통해 소비자들에게 낮은 가격으로 제품을 구매할 수 있도록 한다는 설명을 고지할 경우, AI 사용에 대한 명시는 긍정적인 영향을 미칠 수도 있다. 또한 해당 광고가 AI가 아니면 구현이 불가능한 것이라면 오히려 AI 사용을 명시함으로써

더 긍정적으로 평가될 가능성도 있을 것이다. 따라서 향후 연구에서는 제품 유형 및 AI 사용에 대한 정당성에 따른 기대 차이와 생성형 AI 광고 라벨링 효과를 보다 정교하게 검증할 필요가 있다.

둘째, 본 연구에서는 브랜드의 시장 지위나 인지도를 반영하지 못했다는 한계가 있다. 브랜드 지위와 인지도는 소비자의 신뢰 형성과 지각된 위험을 낮추는 중요한 요인으로 작용하며(Dash & Saji, 2007). 브랜드 신뢰가 높을 경우, 지각된 위험을 낮춰주기 때문에 소비자들로 하여금 제품의 부정적 측면보다 긍정적 측면에 집중하게 만든다. 반면 브랜드 신뢰가 낮은 경우에는 소비자들은 제품에 대해 위험과 불확실성을 지각한다.

뿐만 아니라 소비자가 '멋지다'고 생각하는 브랜드의 경우, AI를 채택하면 긍정적인 소비자 태도를 높이고 기존 브랜드의 이미지를 손상시키지 않고 AI를 적용할 수 있다(Liu, Wang, & Qin, 2024). 또한 브랜드 인지도가 높을수록 AI가 권장하는 제품의 신뢰도가 높지만, 인지도가 중간 수준이거나 낮으면 그 효과가 줄어들었다. 즉, 브랜드 지위나 인지도는 AI의 사용과 AI 추천에 대한 소비자의 판단을 조절하는 변수임을 알 수 있다. 따라서 향후 연구에서는 브랜드 지위와 인지도를 체계적으로 조작하여 생성형 AI 광고 라벨링 효과의 차이를 검증할 필요가 있다.

셋째, 본 연구는 생성형 AI 사용 라벨링의 부정적 효과에 초점을 두었다. 하지만 제품의 과학적 개발 과정을 강조하는 것은 소비자의 구매 의도 및 제품 평가에 긍정적인 영향을 미치기도 하지만, 때로는 과학을 강조하는 것은 소비자의 거부감을 유발하는 것처럼, 특정 상황에서는 AI 활용과 그에 대한 명시가 소

비자 반응을 긍정적으로 증폭시킬 가능성도 존재한다(Philipp-Muller et al., 2023). 따라서 향후 연구에서는 생성형 AI 사용 라벨링에 대한 소비자의 이중적 인식이 언제 긍정적으로 전환되는지를 규명하기 위해 맥락적 요인과 개인차 요인을 보다 폭넓게 탐색할 필요가 있다.

마지막으로 본 연구는 AI 라벨링에 대해 단순한 방법을 활용하였다(문구를 통한 고지). 선행연구에 따르면 AI 고지 형태와 방법에 따른 소비자 반응을 좀 더 세분화하여 살펴볼 필요가 있다(Burnus, Curtis, & Herman, 2024). 이 연구에 따르면 AI 사용에 대한 표시방법에도 시각적 요소, 태그나 주석 등의 다양한 방법을 사용할 수 있다. 소비자는 AI 라벨링 처리 시 맥락에 영향을 받기 때문에 어떤 표시방법을 사용하느냐에 따라서도 다른 평가가 나타날 수 있음을 기대할 수 있다.

또한 제시하는 내용이 무엇인가에 따라서도 투명성에 대한 부분을 세분화 할 수 있다. 과정 기반 투명성(Process-based transparency)의 경우, 콘텐츠가 어떻게, 어디서 생성되었는지에 대한 정보를 사용자에게 전달하는 방식이다. 이 경우에는 콘텐츠의 결과가 시청자에게 미칠 영향에 대해선 중립적인 관점을 취할 수 있다. 해악 기반 투명성(Harm-based transparency)의 경우, 해당 콘텐츠가 잠재적으로 오해를 불러일으키거나 기만적일 수 있음을 사용자에게 직접적으로 경고하는 방식을 의미한다. 이처럼 광고 제작시 생성형 AI 활용에 대한 고지가 규제화 되는 시점에서 좀 더 세부적인 측면들과 소비자의 반응을 확인하는 추후 연구들이 진행될 필요가 있다.

참고문헌

- 구유나 (2026년 1월 22일). ‘세계 첫 전면 시행’ 인공지능기본법은 무엇이고, 내게 미칠 영향은?. BBC 뉴스 코리아.
<https://www.bbc.com/korean/articles/cy4qqklnl99o>
- 김민진, 김미예, 노환호, 김범수 (2024). 생성형 AI 긍정 태도는 어떻게 형성될 수 있을까? 소비자 경험과 디지털 리터러시가 디지털 자아효능감을 통해 생성형 AI 긍정 태도에 미치는 영향. 소비자학연구, 35(2), 143-163.
- 김정영 (2025, 12월 18일). 국내 AI 이용률 74%로 전연령층 확산...AI 순위, 챗GPT>제미나이>에이닷>뤼튼>클로바노트>퍼플렉시티 순. 뉴스스페이스.
https://www.newsspace.kr/news/article.html?no=11348&utm_source=chatgpt.com
- 나정희 (2024). “이 광고는 사실 AI가 만들었습니다.”: 인공지능 제작의 공개 여부와 제품 관여도에 따른 광고효과. 광고연구, (142), 266-296.
- 박재곤 (2025, 12월 31일). [커버스토리] 2026년 주요 IT 시장 전망. 컴퓨터월드.
https://www.comworld.co.kr/news/articleView.html?idxno=51882&utm_source=chatgpt.com
- 윤주혜 (2023, 4월 25일). EU, 이르면 연내 생성형 AI 규제 마련...“MADE WITH AI 라벨 붙여야”. 아주경제.
<https://www.ajunews.com/view/20230425101146414>
- 이유미, 박윤수 (2021). 인공지능 시대의 디지털 리터러시 개념 확장 연구. 미디어교육연구, 21(3), 45-72.
- 이진, 오현정 (2024). 생성형 AI의 기술적

- 특성과 사용자의 AI 리터러시가 생성형 AI의 지속사용의도에 미치는 영향: 확장된 기술수용모델을 중심으로. *한국광고홍보학보*, 26(3), 96-140.
- 장천우, 김성희, 정용국 (2024). ChatGPT에 대한 인지적 개방성 지각이 지속 사용 의도에 미치는 영향: 인지된 혜택과 인지된 위험의 매개 효과와 AI 리터러시의 조절 효과를 중심으로. *사이버커뮤니케이션학보*, 41(2), 135-175.
- Aljarah, A., Ibrahim, B., & López, M. (2025). In AI, we do not trust! The nexus between awareness of falsity in AI-generated CSR ads and online brand engagement. *Internet Research*, 35(3), 1406-1426.
- Aoun, J. E. (2017). *Robot-proof: higher education in the age of artificial intelligence*. MIT press.
- Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, 35(3), 611-623.
- Baek, T. H., Kim, J., & Kim, J. H. (2024). Effect of disclosing AI-generated content on prosocial advertising evaluation. *International Journal of Advertising*, 1-22.
- Bakpayev, M., Baek, T. H., Van Esch, P., & Yoon, S. (2022). Programmatic creative: AI can think but it cannot feel. *Australasian Marketing Journal*, 30(1), 90-95.
- Batra, R., & Homer, P. M. (2004). The situational impact of brand image beliefs. *Journal of Consumer Psychology*, 14(3), 318-330.
- Beverland, M. B. (2005). Crafting brand authenticity: The case of luxury wines. *Journal of Management Studies*, 42(5), 1003-1029.
- Beverland, M. B., & Farrelly, F. J. (2010). The quest for authenticity in consumption: Consumers' purposive choice of authentic cues to shape experienced outcomes. *Journal of Consumer Research*, 36(5), 838-856.
- Boerman, S. C., Willemsen, L. M., & Van Der Aa, E. P. (2017). "This post is sponsored" effects of sponsorship disclosure on persuasion knowledge and electronic word of mouth in the context of Facebook. *Journal of Interactive Marketing*, 38(1), 82-92.
- Burrus, O., Curtis, A., & Herman, L. (2024). Unmasking AI: Informing authenticity decisions by labeling AI-generated content. *Interactions*, 31(4), 38-42.
- Campbell, M. C., & Keller, K. L. (2003). Brand familiarity and advertising repetition effects. *Journal of Consumer Research*, 30(2), 292-304.
- Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, 56(5), 809-825.
- Chaudhuri, A., & Holbrook, M. B. (2001). The chain of effects from brand trust and brand affect to brand performance: the role of brand loyalty. *Journal of Marketing*, 65(2), 81-93.
- Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1), 24-42.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114-126.

- Erdem, T., & Swait, J. (2004). Brand credibility, brand consideration, and choice. *Journal of Consumer Research*, 31(1), 191-198.
- European Commission. (2025). Code of practice on marking and labelling of AI-generated content (AI Act Article 50). European Union.
- Fernandes, T., & Oliveira, R. (2024). Brands as drivers of social media fatigue and its effects on users' disengagement: The perspective of young consumers. *Young Consumers*, 25(5), 625-644.
- Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*, 21(1), 1-31.
- Hermann, I. (2023). Artificial intelligence in fiction: between narratives and metaphors. *AI & Society*, 38(1), 319-329.
- Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30-50.
- Kietzmann, J., Paschen, J., & Treen, E. (2018). Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey. *Journal of Advertising Research*, 58(3), 263-267.
- Kim, J., Morris, J. D., & Swait, J. (2008). Antecedents of true brand loyalty. *Journal of Advertising*, 37(2), 99-117.
- Kim, S. J., Maslowska, E., & Tamaddoni, A. (2019). The paradox of (dis) trust in sponsorship disclosure: The characteristics and effects of sponsored online consumer reviews. *Decision Support Systems*, 116, 114-124.
- Lee, G., & Kim, H. Y. (2024). Human vs. AI: The battle for authenticity in fashion design and consumer response. *Journal of Retailing and Consumer Services*, 77, 103690.
- Lee, J. L., Choi, S. H., Jeong, S., & Ko, N. (2025). Generative AI in sport advertising: effects of source-message (in) congruence, model types and AI awareness. *International Journal of Sports Marketing and Sponsorship*, 26(3), 670-693.
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 2053951718756684.
- Lee, S., Kim, M., & Kim, H. (2024). Relationality of objective and constructive authenticities: Effects on existential authenticity, memorability, and satisfaction. *Journal of Travel Research*, 63(1), 195-214.
- Liu, Y., Wang, X., & Qin, H. (2024). When cool hospitality brand meets AI: exploring the matching effect of service agents and brand images on brand attitude. *International Journal of Contemporary Hospitality Management*, 36(7), 2367-2384.
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90-103.
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*,

- 1-16.
- Longoni, C., & Cian, L. (2022). Artificial intelligence in utilitarian vs. hedonic contexts: The “word-of-machine” effect. *Journal of Marketing*, 86(1), 91-108.
- Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines versus humans: the impact of AI chatbot disclosure on customer purchases. *Marketing Science*, 38(6), 1-11.
- Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, 58(3), 20-38.
- Morhart, F., Malär, L., Guèvremont, A., Girardin, F., & Grohmann, B. (2015). Brand authenticity: An integrative framework and measurement scale. *Journal of Consumer Psychology*, 25(2), 200-218.
- Munuera-Aleman, J. L., Delgado-Ballester, E., & Yague-Guillen, M. J. (2003). Development and validation of a brand trust scale. *International Journal of Market Research*, 45(1), 1-18.
- Napoli, J., Dickinson, S. J., Beverland, M. B., & Farrelly, F. (2014). Measuring consumer-based brand authenticity. *Journal of Business Research*, 67(6), 1090-1098.
- Philipp-Muller, A., Costello, J. P., & Reczek, R. W. (2023). Get your science out of here: when does invoking science in the marketing of consumer products backfire?. *Journal of Consumer Research*, 49(5), 721-740.
- Russell, S. (2016). Rationality and intelligence: A brief update. In *Fundamental issues of artificial intelligence* (pp. 7-28). Cham: Springer International Publishing.
- Russell, S., & Norvig, P. (2020). *Artificial intelligence: a modern approach*(4th ed.). Pearson.
- Schnackenberg, A. K., & Tomlinson, E. C. (2016). Organizational transparency: A new perspective on managing trust in organization-stakeholder relationships. *Journal of Management*, 42(7), 1784-1810.
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146, 102551.
- Sirdeshmukh, D., Singh, J., & Sabol, B. (2002). Consumer trust, value, and loyalty in relational exchanges. *Journal of Marketing*, 66(1), 15-37.
- Södergren, J. (2021). Brand authenticity: 25 Years of research. *International Journal of Consumer Studies*, 45(4), 645-663.
- Song, C., & Kim, E. (2025). The influence of new media literacy on brand engagement: Mediating effects of perceived interactivity and openness and the moderating effect of age. *Behavioral Sciences*, 15(4), 458.
- Steils, N., Martin, A., & Toti, J. F. (2022). Managing the transparency paradox of social-media influencer disclosures: How to improve authenticity and engagement when disclosing influencer-sponsor relationships. *Journal of Advertising Research*, 62(2), 148-166.
- Tayeb, M. E., Chebbi, T., Badawi, A., Toumi, J. A., & Louail, B. (2024). The impact of Ad overloads perception in social media on Ad avoidance behavior: the mediating effect of social media fatigue and goal impediment.

- Management*, 28(2), 351-376.
- Voss, K. E., Spangenberg, E. R., & Grohmann, B. (2003). Measuring the hedonic and utilitarian dimensions of consumer attitude. *Journal of marketing research*, 40(3), 310-320.
- Wu, L., & Jing Wen, T. (2021). Understanding AI advertising from the consumer perspective: What factors determine consumer appreciation of AI-created advertisements?. *Journal of Advertising Research*, 61(2), 133-146.
- Yang, J., & Battocchio, A. F. (2021). Effects of transparent brand communication on perceived brand authenticity and consumer responses. *Journal of Product & Brand Management*, 30(8), 1176-1193.
- 원고 접수일: 2026. 01. 20
최종 원고 접수일: 2026. 02. 10

The Transparency Paradox of “Made with AI”: Effects of AI labeling on Perceived Authenticity, Brand Trust, and Brand Attitude

Eun-Ho Kim Chang-Geun Moon
Dong-A University Jeonbuk National University

This study empirically examines the effects of disclosing the use of generative artificial intelligence (AI labeling) on consumers' advertising responses and brand evaluations in a context where generative AI is widely adopted in advertising production. Specifically, the study analyzes the process through which AI labeling influences brand trust and brand attitude via perceived advertising authenticity, and investigates the moderating role of consumers' AI literacy in this process. To this end, an experiment was conducted using fictitious laundry detergent advertisements created with generative AI, comparing conditions in which AI use was disclosed versus not disclosed. The results indicate that advertisements disclosing AI use elicit significantly lower perceived authenticity than those without such disclosure. Reduced authenticity, in turn, leads to a negative indirect effect on brand attitude through diminished brand trust. Moreover, this indirect effect is moderated by consumers' AI literacy, such that the negative impact of AI labeling on perceived authenticity and subsequent brand attitude is amplified among consumers with higher levels of AI literacy. These findings demonstrate that, in an era where generative AI has become commonplace, transparency regarding AI use does not necessarily lead to positive consumer responses. By highlighting the transparency paradox in AI-generated advertising, this study extends the literature on AI advertising effects and offers meaningful theoretical and practical implications for the strategic use of generative AI in advertising.

Key words : Generative AI Advertising, AI Labeling, Perceived Authenticity, Brand Trust, AI Literacy, Transparency Paradox

부록 A. 실험 자극

 <p>수십 년간 세탁 기술을 선도해온 CLEATIC. CLEATIC의 축적된 성분 배합 노하우로 99.9% 세균 제거, 보이지 않는 오염까지 깨끗하게 세탁도 기술입니다.</p>	 <p>수십 년간 세탁 기술을 선도해온 CLEATIC. CLEATIC의 축적된 성분 배합 노하우로 99.9% 세균 제거, 보이지 않는 오염까지 깨끗하게 세탁도 기술입니다.</p> <p>* 이 광고는 AI로 만들어졌습니다.</p>
<p>AI 라벨링 무 조건</p>	<p>AI 라벨링 유 조건</p>