



인공지능서비스 이용자보호 관련 이슈·법제 검토와 향후 과제*

조 성 은**

Abstract

2022년 ChatGPT 등장 이후 개인 이용자가 인공지능서비스를 직접 이용하게 되면서 이용자 권익 침해에 관한 제도적 대응의 필요성이 높아지고 있다. 이에 방송미디어통신위원회와 KISDI는 인공지능서비스에 대한 현행법의 규율 가능성을 살펴본 「인공지능서비스 사업자를 위한 이용자보호 관련 통신관계 법령안내서」(‘26. 1.)를 발간하였다. 본 고에서는 법령안내서 개발 과정에서의 논의와 결과를 토대로 불법정보금지와 아동·청소년 보호에 관한 현행 법령의 적용 가능성과 한계를 검토하고, 나아가 인공지능서비스의 확산과 역동적 변화과정에서 새롭게 발생할 수 있는 이용자 권익 침해 이슈를 선별하여 인공지능서비스 이용자보호 정책 대응 방안을 제안한다.

I. 서론

2022년 11월 선보인 챗GPT 서비스는 일반 개인이 인공지능 기술을 직접 경험하는 전환점이 되었다. 그 이전까지 인공지능은 규칙 기반의 자동화 시스템, 음성 인식 비서, 추천 알고리즘 등 제한된 형태로 제공되거나 서비스에 내재된 기능으로만 존재하였다. 그러나 챗GPT를 필두로 한 생성형 인공지능의 확산은 대화형 인터페이스를 통해 일반인이 텍스트나 음성을 매개로 인공지능과 직접 상호작용하는 환경을 열었고, 정보탐색, 과제 수행, 업

* 본 고는 2025년 정책연구의 결과물인 「인공지능서비스 사업자를 위한 이용자보호관련 통신관계 법령안내서」의 내용 중 불법정보 유통과 아동·청소년 보호에 대해 다루는 한편, 논의의 범위를 확장하여 동반자 AI(companion AI)를 포함한 현행 법령으로 규율되지 않는 이용자 권익 침해 이슈도 다룬다.

** 정보통신정책연구원 인공지능연구실 연구위원, secho@kisdi.re.kr

무 보조는 물론, 창작과 토론, 정서적 교류에 이르기까지 활용 범위를 확장시켰다.

이후 인공지능 서비스는 텍스트 생성에서 이미지·영상 생성으로 빠르게 확대되었고, 멀티모달 AI, 온디바이스 AI, 에이전틱 AI 등 보다 높은 자율성과 통합성을 갖춘 기술로 진화하였다. 이러한 기술은 인간의 판단을 보조하는 수준을 넘어 일정 영역에서 의사결정을 위임받거나 대체하는 방향으로 발전하고 있으며, 향후 서비스 환경의 구조적 변화를 초래할 가능성이 높다(Page & Kallapur, 2025).

그러나 기술 확산 속도에 비해, 사회적 인식과 제도적 대응이 충분하지 않다. 생성형 인공지능 확산 초기에는 이른바 ‘할루시네이션(정보 왜곡)’ 현상에 대한 인식 부족으로 인해 잘못된 정보를 그대로 신뢰하거나 의사결정에 활용하여 피해를 본 사례들이 보고되었으며, 이는 단순히 이용자 부주의 탓으로 돌릴 수 없는 문제였다. 당시 챗GPT를 비롯한 대화형 AI 서비스가 일반 이용자들에게 새로운 유형의 정보검색 서비스로 소개되었고, 서비스 자체의 속성과 한계에 대한 사전 안내가 부족했었기 때문이다. 이후 생성형 인공지능으로 그럴듯한 콘텐츠 생성이 쉬워지면서 불법정보 증대와 빠른 유포 문제가 부각되었고,¹⁾ 인공지능이 아동·청소년 등 심리적으로 취약한 계층에게 미칠 부정적 영향도 주요 쟁점으로 떠올랐다. 특히 동반형 인공지능(companion AI)의 정서적 증독성에 대한 우려는 기술의 윤리적 사용을 넘어 정신건강 문제 차원에서 경고되고 있다.²⁾³⁾

이 중 초기에 주목받았던 할루시네이션 문제는 RAG(검색증강생성) 기반 검색 연동 등의 기술적 보완, 사업자의 면책 고지 및 이용자 인식 제고 노력이 병행되면서 초기에 비해 상당 부분 완화되었다. 교육부의 「모두를 위한 인공지능 인재양성 방안」(25. 11.)을 비롯해 전 생애주기 인공지능 리터러시 향상 교육이 적극 추진되고 있는 것도 같은 맥락이다. 반면, 불법정보 생산·유포와 아동·청소년 보호 이슈는 기술적 통제 방안과 제도적 정립이 모색되고 있으나, 여전히 구조적 대응이 필요한 과제이다.

세계 각국은 인공지능으로 인한 사회의 부정적 영향을 최소화하기 위해 법·제도를 마련하고 있으나, 아직 시행 전이거나 시행 이후에도 정립된 사회제도로서의 기능을 온전히 수행하지 못하는 상황에 있다. 우리나라에서는 2026년 1월 「인공지능 발전과 신뢰 기반 조성

1) 연합뉴스(2025. 8. 28). “딥페이크 피해자 지원 1년새 128% 증가…피해자 절반이 10대”, <https://www.yna.co.kr/view/AKR20250828020300530>

2) Olsen, S. D., Reinecke-Tellefsen, C. J., Ostergaard, S. D. (2026, 2, 6.). Potentially Harmful Consequences of Artificial Intelligence(AI) Chatbot Use Among Patients With Mental Illness: Early Data From a Large Psychiatric Service System, <https://onlinelibrary.wiley.com/doi/10.1111/acps.70068>

3) 서울경제(2025. 8. 20.). “입원한 사람만 수두룩”...경고등 켜진 ‘AI 정신병’, 우리나라는 문제 없을까? <https://m.sedaily.com/article/14106527>

등에 관한 기본법」(이하 「인공지능 기본법」)이 시행되었다. 이 법은 인공지능 생애주기 전반을 포괄하는 국가 차원의 기본 방향과 운영 원칙을 설정하는 기본법적 성격을 가지며, 기본계획 수립, 정책 추진 체계, 위험관리 원칙 등 상위 규범을 중심으로 구성되어 있다. 다만 개별 서비스 이용 단계에서 발생하는 구체적 행위 규율까지는 포괄하지 않는다(정찬모, 2025). 이에 시행령을 통해 일부 사항을 구체화할 수 있겠으나, 법적 공백을 최소화하기 위한 보완적 접근으로서 구체적 행위 규율을 담당하는 기존 법령과의 기능 분담 방안을 함께 검토할 필요가 있다.

이러한 맥락에서, 방송미디어통신위원회와 KISDI는 「인공지능서비스 사업자를 위한 이용자보호 관련 통신관계 법령안내서」(26. 1.)를 발간하였다. 이 안내서는 정보통신망을 이용하는 인공지능서비스 사업자가 통신관련 법령상 부담하는 책임과 의무를 안내하는 것을 목적으로 한다.⁴⁾

본 글에서는 위 안내서의 내용 중, 특히 최근 인공지능 서비스 확산 과정에서 사회적 문제로 부각된 불법정보 유통과 아동·청소년 보호에 대해 다룬다. 덧붙여 2026년 인공지능 서비스 환경 변화에 따라 새롭게 등장하거나 심화될 수 있는 이용자 권익 침해 유형을 전망하고, 현행 법체계 내에서 규율 가능한 범위와 향후 제도 개선을 통해 보완해야 할 과제를 고찰한다.

II. 인공지능 서비스와 「정보통신망법」의 적용 가능성

본 장에서는 사회적으로 논쟁이 된 동시에 현행 통신관계 법령과 직접적으로 접점에 있는 두 가지 쟁점, ① 불법정보 생성과 유통 문제와 ② 아동·청소년 보호 문제를 중심으로 검토한다. 전자는 콘텐츠 및 유통 구조 차원의 위험을, 후자는 인공지능과 이용자 간 상호작용 구조에서 발생하는 위험을 대표한다. 각 쟁점에 대해 현행 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」(이하 「정보통신망법」)의 적용 가능성과 한계를 살펴보고, 해외 입법 동향과 비교하여 향후 제도 개선 방향을 모색한다.

4) 그 밖에 개인정보보호위원회의 「인공지능 개발·서비스를 위한 공개된 개인정보처리 안내서」(24. 7.)는 인공지능 사업자가 공개된 개인정보를 「개인정보보호법」에 적합하게 처리하기 위한 기준과 고려 사항을 담고 있다.

가. 불법정보 생성 및 유통 문제

현행 「정보통신망법」은 기본적으로 이용자가 게시한 정보의 유통을 전제로 하고 있고, 사업자를 유통 환경 제공자로 간주한다. 또한 위법 정보에 대한 사후적 삭제 및 차단 중심의 책임 구조를 취하고 있다(제44조, 제44조의2, 제44조의3, 제44조의7 등). 그러나 생성형 인공지능은 정보의 ‘유통’ 이전 단계에서 ‘생성’을 수행하며, 이용자와의 상호작용 과정에서 실시간으로 결과물을 산출한다는 점에서 기존 규율 구조와 질적으로 다른 양상을 보인다. 따라서 다음에서는 현행 법령상 인공지능서비스 규율의 적용 가능 범위를 검토하고, 인공지능의 기술적 특성으로 인해 고려해야 하는 한계는 무엇인지, 해외 주요국은 어떠한 방식으로 대응하고 있는지, 그리고 현행 법체계 내 보완 가능성과 입법적 과제를 살펴본다.

1) 「정보통신망법」의 규율 구조

「정보통신망법」 제44조의7은 제1항에서 정보통신망을 통해 유통이 금지되는 불법정보 유형을 열거하고 있으며, ① 음란물의 공공연한 전시·배포, ② 타인의 명예를 훼손하는 정보, ③ 공포심·불안감을 유발하는 부호·문언·음향·화상·영상, ④ 정보통신시스템의 정상적 운용을 방해하는 정보, ⑤ 청소년 보호 의무를 위반한 유해정보, ⑥ 개인정보 거래 정보, 총포·화약류 등의 제조 방법, 마약 관련 정보, ⑦ 국가기밀 누설 정보, ⑧ 국가보안법 위반 정보, ⑨ 범죄를 목적으로 하거나 교사·방조하는 정보 등이다.

해당 불법정보가 정보통신망을 통해 유통되는 경우 정보통신서비스 제공자에게도 일정한 조치 의무가 발생한다. 통신망을 기반으로 온라인 형태로 제공되는 인공지능서비스는 「전기통신사업법」상 부가통신역무이자 「정보통신망법」상 정보통신서비스에 해당하므로, 이러한 유형의 인공지능서비스 사업자 역시 원칙적으로 적용 대상이 된다. 이에 사업자는 다음과 같은 책임을 부담한다. ① 이용약관에 불법정보 생성·유통 금지 의무 명시, ② 권리침해 주장에 따른 삭제·임시조치 대응(제44조의2, 제44조의3), ③ 방송미디어통신심의위원회 등의 시정요구에 따른 조치, ④ 청소년 보호 책임자 지정 등 관련 의무 이행 등이다. 이와 같이 현행 법체계는 사업자를 ‘유통 환경 제공자’로 보고, 불법정보의 확산을 방지하기 위한 사후적·절차적 의무를 중심으로 규율하고 있다.

2) 인공지능의 특수성과 해석상 한계

현행 법조문은 온라인 서비스의 특성을 반영하여 입법된 것으로, 인공지능의 기술적 특수성에서 비롯되는 문제를 규율하는 데에는 일정한 한계가 있다. 첫째, 불법 정보를 생성하는 단계를 규율 범위에 포섭하기 어렵다는 점이다. 현행 「정보통신망법」은 이용자가 생산한

불법정보의 게시·유통을 규율 대상으로 하고 있으나, 생성형 인공지능은 이용자와의 상호작용을 통해 불법정보를 실시간으로 직접 ‘생성’한다는 점에서 기존의 유통 구조와 본질적으로 다르다. 같은 법에서 ‘유통’은 정보통신망을 통한 정보의 전송·게시·배포·공유 등을 의미하는바, 인공지능 생성물이 제3자에게 공유되지 않고 1:1 상호작용 내에서만 제공되는 경우 이를 기존의 유통 개념으로 포섭하는 데에는 해석상 한계가 있다.⁵⁾

둘째, 사후적 조치의 실효성에 한계가 있다. 인공지능서비스는 자동화된 방식으로 짧은 주기 안에 대량의 콘텐츠를 반복적으로 생성할 수 있어서, 삭제·차단 등의 사후 조치만으로는 문제를 해결하기 어렵다. 셋째, 기존의 책임 구조가 다층화된 인공지능서비스 환경에 부합하지 않는다는 점이다. 현행 법제는 유통 플랫폼 사업자와 게시자로 한정된 책임 구조를 전제로 하고 있으나 인공지능서비스에는 서비스 제공자 외에도 모델 개발자, API 제공자, 유통 플랫폼 등 다양한 주체가 관여한다. 이처럼 생애주기별로 책임 주체가 다층화되는 구조는 기존 법제의 틀 안에서 적절히 규율하기 어렵다.

3) 해외 대응 동향

해외 주요국은 불법정보 유통에 대해 사후적 차단 조치에 그치지 않고, 사전적 위험관리 의무 부과 또는 신속 삭제 의무 강화 방향으로 대응을 강화하고 있다. 특히 영국은 생성형 인공지능 서비스 확산에 따라 기존 온라인 서비스 규제 체계를 어떻게 적용, 확장할 것인지에 대한 논의가 활발하고 실제로 단계적으로 반영하고 있다.⁶⁾ 또, 영국의 통신 규제기관인 Ofcom은 「온라인 안전법(Online Safety Act 2023)」 시행을 앞두고 공개서한을 통해 생성형 인공지능 및 챗봇 서비스가 규율 대상이 될 수 있음을 명확히 하였다.⁷⁾ 그 내용을 보면, 생성형 인공지능 서비스가 단순한 콘텐츠 생성 기능에 그치지 않고, 생성된 결과물을 제3자에게 전송·공유할 수 있는 기능을 포함하는 경우, 서비스 제공자에게 법적 책임이 부과될 수 있다. 이용자가 생성형 인공지능을 통해 불법 콘텐츠를 ‘생성만’ 한 경우에는 책임 부과의 정도가 낮거나 적용 대상에서 제외될 수도 있다는 점도 덧붙였다. 또한 서비스 제공자가 자신의 서비스 구조상 불법 콘텐츠 생성이 가능함을 인지하고도 적절한 방지 노력을 기울이지 않은 경우에는 방조 책임이 발생할 수 있음도 명시하였다.

5) 방미통위·KISDI(2026. 1.), 「인공지능서비스 사업자를 위한 이용자보호 관련 통신관계 법령안내서」, p.35.

6) 영국 Ofcom에서 온라인 안전법의 단계적 시행 계획은 다음 사이트에서 확인할 수 있다:

<https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/roadmap-to-regulation>

7) Ofcom(2024. 11. 8.), “Open letter to UK online service providers regarding Generative AI and chatbots”.

<https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/open-letter-to-uk-online-service-providers-regarding-generative-ai-and-chatbots>

영국의 「온라인 안전법」은 대형 플랫폼 및 검색서비스를 비롯하여 영국에서 이용 가능한 이용자 제작 콘텐츠(UGC, user-generated contents), 음란물 호스팅 서비스 등을 규율 대상으로 하며 사후 조치뿐만 아니라 사전 조치 의무도 부여한다.⁸⁾ 사전 조치로는 위험평가(risk assessment) 실시와 안전 설계(safety-by-design) 원칙 적용 등이 있으며, 사후 조치로는 불법 콘텐츠를 인지하는 즉시 신속하게 삭제할 의무와, 콘텐츠 신고 및 불만 처리 절차 마련 등이 포함된다.⁹⁾ 최근에는 동의없이 유포되는 성착취물을 감지하기 위해 “해시 매칭(hash matching)”이라는 사전 예방 기술(proactive technology) 사용 의무화 방안을 제안하는 등 현행 온라인 서비스 환경에서 발생가능한 불법정보 유통 문제에 보다 적극적으로 대응하겠다는 방침을 밝혔다.¹⁰⁾

영국이 온라인 서비스 전반에 적용되는 규율 체계에 생성형 인공지능 규율 가능성을 확인하는 방향으로 제도 정비를 한 것과는 달리, 미국은 인공지능 기술 악용의 특정 유형에 집중한 개별 입법 방식을 취하였다. 미국은 2025년 「Take It Down Act」를 제정하여¹¹⁾ 인공지능 기술을 활용하여 생성된 딥페이크 등의 비동의 성적 이미지(non-consensual intimate images)를 게시하거나 이를 이용하여 협박하는 행위를 금지하고, 해당 이미지가 플랫폼에 게시되거나 유포되었을 때 48시간 이내에 삭제할 수 있는 체계를 마련할 의무를 플랫폼 사업자에게 부여하였다. 이 법은 삭제 시한을 명시적으로 규정하고, 기존의 ‘촬영물 유포’뿐만 아니라 ‘디지털 합성물’도 규율 대상에 포함함을 명확히 했다는 점에서 주목받았다.

4) 향후 대응 과제

우리나라의 경우, 방송미디어통신위원회와 KISDI가 통신관계 법령에서 인공지능서비스 규율 가능성을 검토하고 해당 법조문을 해설·안내하는 「인공지능서비스 사업자를 위한 이용자보호 관련 통신관계 법령안내서」(이하 「인공지능서비스 통신관련 법령안내서」(26. 1.)를 최근 발간하였다. 이는 영국 Ofcom이 공개서한을 통해 규제 방향을 명시적으로 선언한

8) 영국에서 이용할 수 있는 사용자 제작 콘텐츠(user-generated contents), 검색서비스, 음란물 서비스를 호스팅하는 온라인 서비스를 규율 대상으로 하고 있다.

9) Ofcom(2026. 2. 27.). “Ofcom provisionally finds suicide forum in breach of Online Safety Act”
<https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/ofcom-provisionally-finds-suicide-forum-in-breach-of-online-safety-act>

10) 2026년 5월에 해당 제안의 최종 결정을 발표한다고 함.
 Ofcom(2026. 2. 18.). “Ofcom fast-tracks decision on measures to block illegal intimate images”,
<https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/ofcom-fast-tracks-decision-on-measures-to-block-illegal-intimate-images>

11) 미국 「The TAKE IT DOWN Act: A Federal Law Prohibiting the Nonconsensual Publication of Intimate Images」,
<https://www.congress.gov/crs-product/LSB11314>

것에 비하면 소극적인 대응이지만, 인공지능 서비스의 사회적 영향에 비해 법적 규제 체계가 모호했던 상황에서 적용 가능한 규율 근거를 체계적으로 정리하여 제시하였다는 점에서 의의가 있다. 또한 입법 개선 방향도 포함하고 있어서 향후 인공지능서비스에서 발생 가능한 이용자 권익 보호를 위한 정책 과제를 모색해 볼 수 있다.

향후 입법적 보완 과제는 크게 규율 대상의 확장, 조치 시점의 전환, 책임 주체의 재정립 등 세가지 차원에서 검토할 수 있다. 첫째, 현행 '유통' 개념에 대한 법적 범위를 확장하는 것을 고려해 볼 수 있다. 인공지능 생성물이 직접 게시 혹은 유통되지 않더라도 플랫폼이 제3자 공유 기능을 제공하거나 플랫폼 내 다른 이용자와 공유할 수 있는 구조라면 유통 가능성이 있다고 보는 방안을 우선 검토할 수 있다. 혹은 기존 법적 개념을 확장하는 것보다 적극적인 대응으로 유통 외에 생성 단계를 규율하는 법령을 마련하거나 현행법을 개정하는 방안을 고려할 수 있다. 즉, 기존 유통금지 정보에 대응하는 생성금지 정보의 범위, 인공지능서비스 사업자의 불법정보 생성 방지를 위한 기술적·관리적 조치 의무 등을 명시적으로 규정하는 방안을 병행해서 검토해 볼만하다. 둘째, 사전 조치 의무의 강화이다. 앞서 언급하였듯이 인공지능은 불법 콘텐츠를 자동화된 방식으로 대량·반복 생성할 수 있고 디지털 정보의 특성상 확산 속도도 빠르므로, 사후 조치만으로는 사회적 피해를 충분히 예방하기 어렵다. 현재는 사업자의 자율 규제 형식으로 모니터링 및 필터링 등의 위험관리가 이루어지고 있으나, 인공지능의 사회적 파급력을 고려할 때 영국의 위험평가 및 안전 설계 의무와 같은 사전 조치의 의무화를 적극적으로 검토할 필요가 있다.

셋째, 책임 주체의 재정립이다. 만약 불법정보 유통 금지를 규율하는 데 있어서 불법정보 생성부터 규율하고자 한다면, 서비스 제공자와 이용자 간 책임 구조를 전제하는 현행법의 책임 구조를 넘어서 필요에 따라 서비스 제공자 외 모델 개발자·API 제공자·유통 플랫폼 등 다양한 주체가 관여하는 인공지능서비스의 다층적 책임 구조를 고려해야 할 수도 있다. 물론 이 경우 「정보통신망법」의 입법취지를 벗어나지 않는 범위에서 일부 관련 조문에서 수범자 범위를 확대하는 방안을 검토하되, 「인공지능기본법」 시행령 또는 별도 입법을 통한 역할 분담 체계를 함께 모색함으로써 타 법령과의 중복 규율 문제를 최소화해야 할 것이다.

나. 아동·청소년 보호

1) 「정보통신망법」의 규율 구조

온라인 서비스를 규율하는 「정보통신망법」에서 아동·청소년 보호 조항은 상대적으로 제

한적이다. 먼저 동법 제44조의8에서 ‘대화형 정보통신서비스에서의 아동 보호’를 규정하고 있다. 2018년 도입된 이 조항은 정보통신서비스 제공자가 만 14세 미만 아동에게 대화형 정보통신서비스를 제공하는 경우 아동에게 부적절한 내용의 정보가 제공되지 않도록 보호 조치를 마련할 것을 권고한다. 여기서 ‘대화형 정보통신서비스’란 “문자·음성을 이용하여 사람과 대화하는 방식으로 정보를 처리하는 시스템을 기반으로 하는 정보통신서비스”를 말한다.

「정보통신망법」은 또한 제41조(청소년 보호 시책 마련), 제42조(청소년유해매체물 표시), 제42조의2(청소년유해매체물 광고 금지), 제42조의3(청소년 보호 책임자 지정) 등에서 청소년 보호 체계를 명시하고 있다. 특히 제42조의3은 일정 규모 이상의 정보통신서비스 제공자에게 청소년 보호 책임자를 지정하도록 의무화하고 있으며, 동법 시행령 제26조에 보호 업무를 규정한다. 청소년 보호 업무에는 ① 유해정보로부터의 청소년 보호 계획 수립, ② 청소년 접근 제한 및 관리 조치, ③ 종사자 교육, ④ 피해 상담 및 고충 처리, ⑤ 기타 필요한 사항 등이 있다.

2) 인공지능의 특수성과 해석상 한계

위에서 언급한 법조문들은 정보통신서비스에서 아동·청소년 보호의 필요성을 법체계에 반영했다는 점에서 의의가 있으나, 현재의 인공지능서비스 환경에 적용하려고 할 때 다음과 같은 한계가 있다. 첫째, 제44조의8에서 규정하는 아동 보호의 적용 범위가 ‘대화형 정보통신서비스’로 한정된다는 점이다. 현재 가장 널리 사용되는 인공지능 서비스는 대체로 대화형 인터페이스로 제공되지만, 아동에게 영향을 미치면서 정보통신망을 통해 제공되는 인공지능 서비스가 모두 대화형으로 제공된다고 단정할 수 없다. 만약 ‘대화형’을 “문자·음성을 이용하여 사람과 대화하는 방식으로 정보를 처리”한다는 문자 그대로의 ‘conversational’의 의미가 아니라 “이용자의 모든 유형의 입력에 시스템이 반응하는”에 해당하는 ‘interactive’의 의미로 해석한다면 현재의 적응적·반응적 인공지능 시스템이 폭넓게 포함될 수 있으나, 현행법에서의 정의는 이를 명확히 포섭하지 못하고 있다.¹²⁾ 향후 인공지능 서비스 형태의 다변화를 고려할 때 이 문제는 더욱 중요해질 것이다.

둘째, 해당 조문들이 자율 규제 형식으로서 규정되어 있어서 법적 구속력이 상대적으로 약하다는 점이다. 해당 조문들은 제정 당시 사업자의 자율적 규제와 사회적 책임 제고를 목적으로 마련되었으나, 기술 발전과 함께 정보통신서비스가 아동·청소년에게 미치는 부정

12) 미국 의회는 “interactive computer service”를 사용하며, 대부분의 온라인 애플리케이션과 서비스를 모두 포함하는 용어로 활용하고 있다(interactive computer service—an existing term broadly construed to cover most online applications and services). 자료: <https://www.congress.gov/crs-product/LSB11314>

적 영향이 심화되는 현실에서는 자율규제만으로는 충분하지 않을 수 있다. 최근 해외에서도 미성년자의 자해·자살 등 중대한 결과로 이어질 수 있는 위험이 공론화되고 있다.

셋째, 규율 대상 ‘부적절한 정보의 노출 차단’에 한정되어 있어서 인공지능서비스 이용 환경에서 새롭게 부각되는 위험을 포괄하지 못한다는 점이다. 해당 조문에서 규율 대상인 “아동에게 부적절한 내용의 정보”는 일반적으로 동법상 불법정보(제44조의7 제1항) 및 청소년 유해정보(제41조 제1항)로 해석된다.¹³⁾ 그러나 최근 인공지능 서비스 이용 환경에서 부각된 핵심 위험은 특정 정보에의 노출보다는 인공지능과의 반복적이고 개인화된 상호작용 과정에서 축적되는 심리적·정서적 영향이다. 이에 부적절한 정보에의 노출을 차단하는 것만으로는 상호작용 기반의 위험을 규율하기 어려울 수 있다.

이 세 번째 단계는 특히 최근 인공지능서비스의 기술적 진화와 맞닿아 있다. 생성형 인공지능의 확산 이후, 인공지능은 단순한 정보 제공 도구를 넘어 이용자의 프롬프트와 대화 맥락에 반응하며 결과물을 산출하는 기술로 진화하였고, 이용자의 요구와 숨은 욕구에 맞춰 상호작용하면서 인공지능이 인간 이용자와 관계를 형성하는 현상도 가시화되고 있다. 일부 연구자들은 이러한 인공지능을 ‘동반자(companion)’로 명명하며(Brewster et al., 2025), 업무 등을 보조하는 기존의 조력자 혹은 비서(assistant) 역할에서 감정·정서를 교류하는 친구·동반자의 지위로 그 역할이 확대되고 있음을 지적한다. 관련 연구에 따르면 이용자들은 인공지능의 응답을 인간보다 더 공감적으로 지각하는 경향을 보인다(Ovsiyannikova et al., 2025; Shen et al., 2024; Liu, 2024). 이들의 지각된 공감(perceived empathy)에서 인공지능이 높게 평가되는 것이다. 이 과정에서 정서적 몰입 현상이 일어날 수 있으며, 이용자에 대한 인공지능의 심리적 영향력이 증폭될 수 있다. 그리고 발달 단계상 정서적 통제 능력이 충분히 성숙하지 않은 아동·청소년이나 사회적 고립 상태에 있는 이들은 이러한 영향에 더 크게 노출될 수 있다는 점에서 사회적으로 특히 보호받아야 한다.

이러한 문제 인식은 정책 차원에서도 확인된다. 2025년 국회 과학기술정보방송통신위원회 국정감사에서 인공지능서비스에서의 아동·청소년 보호 대책 필요성이 제기되었으며¹⁴⁾, 2026년 2월 최종 발표된 「대한민국 AI 행동계획 2026」에서도 “유해 AI 서비스로부터 아동·청소년 보호”를 위한 법·제도를 2027년 4분기까지 마련할 것을 권고한다. 이에 따라 인공지능서비스에서의 아동·청소년 보호를 위한 법 개정 혹은 새로운 입법 논의가 본격화될 것

13) 방미통위·KISDI(2026. 1.). 「인공지능서비스 사업자를 위한 이용자보호 관련 통신관계 법령안내서」, p.61.

14) 조선일보(2025. 10. 16.). AI 유해 콘텐츠 범람하는 SNS…단속 속도 못 대는 정부.

<https://it.chosun.com/news/articleView.html?idxno=2023092148957>

으로 전망된다.

3) 해외 대응 동향

해외에서도 인공지능서비스를 포함한 온라인 서비스 환경에서의 아동·청소년 보호 조치를 강화하는 추세이다. 먼저 인공지능과의 상호작용에서 발생하는 위험을 직접 규율하는 사례로, 미국 캘리포니아주가 2025년 제정하여 2026년 1월부터 시행 중인 「캘리포니아주 법 243(California Senate Bill 243(SB 243))」이 있다.¹⁵⁾ 이 법의 규율 대상인 ‘동반자형 챗봇(companion chatbot)’은 “자연어 인터페이스를 갖춘 인공지능 시스템”이며, “이용자의 프롬프트 입력에 적응적이며 인간과 유사한 응답을 제공하고, 의인화된 특징을 나타내거나 여러 반복적 상호작용을 통해 관계를 형성·유지할 수 있는 기능”을 갖춘 것으로 정의된다. 이 법에 따라 사업자는 미성년 이용자에게 인공지능과 상호작용하고 있음을 알리는 등 특정 조치를 해야 하며, 자살 충동·자살·자해 관련 콘텐츠 생성 방지 조치를 마련하거나 그렇지 못한 경우 해당 챗봇이 미성년 이용자와 상호작용을 하지 못하도록 조치하고 해당 프로토콜에 대한 세부 정보를 웹사이트에 게시해야 한다. 2027년 7월부터는 자살 예방 사무국에 이용자의 자살 충동 감지, 제거 및 대응 절차 등을 매년 보고하고 이를 웹사이트에 게시해야 한다.

이와 별도로 캘리포니아주는 소셜미디어 플랫폼의 중독성 피드 알고리즘을 규율하는 별도의 법을 제정하였다. 2024년 9월 제정된 「Protecting Our Kids from Social Media Addiction Act(소셜미디어 중독으로부터 아동보호법)」(California SB 976)은 2027년부터 시행될 예정이다.¹⁶⁾ 이 법은 인터넷 기반 서비스 또는 애플리케이션 운영자를 수범자로 하고 있으며, 미성년자가 부모 동의 없이 알고리즘 기반의 ‘중독성 피드’를 미성년자에게 제공하는 것을 금지한다. 여기서 중독성 피드는 “이용자가 제공한 정보(활동에서 생성된 정보 등)나 이용자의 기기 정보 등을 기반으로 추천, 선택, 또는 우선 표시되도록 하는 인터넷 웹사이트, 온라인 서비스, 온라인 및 모바일 애플리케이션”을 말한다. 사업자는 매년 미성년 이용자 수, 부모 동의를 받은 미성년 이용자 수, 접근 제어 기능 활성화된 미성년 이용자 수와 활성화되지 않은 미성년 이용자 수를 공개하도록 규정한다. 미국의 유타주는 「Utah Social Media Regulation Act(SB 152, HB 311)」을 제정하여 미성년자 계정 개설

15) 캘리포니아주 SB-243 Companion chatbots.(2025-2026)

https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB243

16) 캘리포니아주 SB-976 Protecting Our Kids from Social Media Addiction Act.(2023-2024),

https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB976

시 부모 동의와 연령 확인 절차를 의무화하였지만,¹⁷⁾ 해당 법안이 표현의 자유를 침해할 수 있다는 이유로 헌법 소송이 제기된 상황이다.

호주의 개정된 온라인 안전법(Online Safety Amendment Act 2024)은 연령 기반 접근 제한을 한 사례로 주목받았다.¹⁸⁾ 청소년의 온라인 플랫폼 이용이 정신건강 문제나 온라인 피해 노출 증가, 중독 위험 등과 연결될 수 있다는 우려에 따라 도입된 이 법은 소셜미디어 플랫폼에서 16세 미만 이용자의 계정 보유를 금지하고, 사업자에게는 연령제한 이용자가 계정을 생성하지 못하도록 합리적 조치를 부여한다.

4) 향후 대응 과제

앞서 살펴보았듯이, 현행법상 아동·청소년 보호 관련 조항의 한계는 단계적으로 보완되어야 한다. 인공지능서비스 유형이 빠르게 확장되는 환경 변화에 맞추어 규율의 적용 범위를 조정하고, 문제의 심각성이 높아지는 현실을 반영하여 자율규제 방식의 실효성도 재검토할 필요가 있다. 무엇보다 인공지능과의 반복적·개인화된 상호작용 과정에서 형성되는 정서적 의존, 위험 대화의 지속, 알고리즘 기반 중독 유발, 자해·자살 관련 고위험 콘텐츠 노출 등의 문제를 고려할 때, 상호작용 구조와 위험 요소의 누적 가능성 등을 반영한 새로운 규율 기준을 마련하는 것이 필요하다.

다만, 새로운 규율의 도입은 실증 근거와 사회적 합의를 바탕으로 이루어져야 한다. 인공지능 기술의 발전 속도와 서비스 형태의 다양성을 감안할 때, 근거가 충분히 축적되지 않은 상태에서의 입법은 실제 위험을 효과적으로 통제하지 못하면서 다른 영역에서는 오히려 과도한 제한을 초래할 수 있기 때문이다.

법 개정 논의와 병행하여 기업의 자율규제 체계를 정교화하고 위험평가·안전 설계·내부 점검 체계 등을 강화하도록 유도하는 방안도 함께 추진할 수 있다. 이를 통해 규제 공백을 해소하거나, 입법 여건이 성숙될 때까지 기업 자율규제 체계를 안정적으로 유지·운영하는 효과를 기대할 수 있다. 아울러 이용자 행태 및 인식에 관한 체계적 조사, 아동·청소년을 대상으로 한 실증 연구(예. 정서적 의존을 유발·강화하는 요인 식별), 실제 피해 사례 분석

17) 유타주 S.B. 152 Social Media Regulation Amendments,

<https://le.utah.gov/~2023/bills/static/SB0152.html>,

유타주 Utah Minor Protection in Social Media Act,

https://le.utah.gov/xcode/Title13/Chapter71/C13-71_2024100120240501.pdf

18) 호주 온라인 안전법 개정안. ONLINE SAFETY AMENDMENT(SOCIAL MEDIA MINIMUM AGE) ACT 2024 (NO. 127, 2024) – SCHEDULE 1 Amendments,

https://www.austlii.edu.au/cgi-bin/viewdoc/au/legis/cth/num_act/osammaa2024409/sch1.html

등 정책 설계의 근거를 지속적으로 축적하는 작업도 병행되어야 한다.

한편, 최근 사회적으로 심각하게 인식되는 사안에 대해서는 보다 적극적인 검토가 필요하다. 아동·청소년의 자해·자살 등 극단적 행위를 유발할 수 있는 고위험·고영향 이슈에 대해서는, 강화된 안전 설계 의무와 즉각적 대응체계를 법령에 명시적으로 규정하는 방안을 검토할 수 있다. 이와 관련하여, 자살 충동 감지·차단 의무와 보고 체계를 의무화한 캘리포니아주 SB 243의 입법례를 참고할 만하다.

Ⅲ. 현행 법령으로 규율되지 않는 이용자 권익 침해 이슈

앞서 살펴본 바와 같이 기존 온라인 서비스 환경에서 인공지능서비스가 제공되는 경우 불법정보 유통이나 아동·청소년 보호 등의 이슈는 「정보통신망법」 등 현행법에서 규율 근거를 찾을 수 있다. 그러나 인공지능의 지속적 발전과 서비스의 확산은 기존 법체계에서 충분히 예상하지 못한 새로운 형태의 이용자 권익 침해 가능성을 등장시킨다. 인공지능 서비스는 특히 이용자와의 상호작용을 통해 정보를 생성·추천하고 행동을 유도하는 구조를 가지며, 이러한 과정에서 개인의 의사결정, 정보접근, 감정 상태 등에 영향을 미칠 수 있다는 점에서 기존의 콘텐츠 규제 중심 법제만으로는 해결할 수 없는 어려운 문제가 발생한다. 2026년을 기점으로 온디바이스 AI나 에이전틱 AI 등 자율적이고 상시 작동하는 인공지능 서비스가 일상 서비스로 본격 편입될 전망이다. 이에 따라 이용자는 인공지능을 선택적으로 사용하는 주체가 아니라 스스로 인지하지 못한 상태에서 인공지능의 판단·추천·자동 실행에 노출되는 존재에 머물게 될 수 있다.

이러한 이용 환경 변화는 현행 법체계가 구조적으로 포착하기 어려운 새로운 유형의 이용자 권익 침해를 야기할 수 있다. 인공지능 서비스가 갖는 자율성, 블랙박스로 상징되는 비가시성, 일관된 결과가 아닌 확률적 결과 출력 등의 특성은 이용자가 경험 혹은 인지하지 못한 상황에 처하게 할 것이다. 온디바이스 AI·에이전틱 AI는 이용자 행위의 간접화 혹은 위임화를 초래하고 이로 인해 피해 인지와 입증을 어렵게 할 수 있다.

본 장에서는 이러한 맥락에서 2026년 이후 새롭게 등장하거나 본격화하는 인공지능 서비스 환경에서 생각해 볼 수 있는 이용자 권익 침해 이슈를 소개한다. 이 이슈들은 인공지능 서비스에서의 이용자보호 정책 개발 연구를 위한 전문가 자문과 논의된 내용을 토대로 관련 문헌 검토와 보완을 거쳐 정리한 것이다. 구체적 현상이 발현하지 않은 시나리오에 불

과할 수 있지만, 선제적 정책 대응 과제를 제시한다는 차원에서 기술 발전 과정에서 충분히 짐작할 수 있는 잠재적 위험 수준에서 다뤄보고자 한다.

가. AI 자동 타겟팅 광고·마케팅

먼저 인공지능 기반 자동 타겟팅 광고·마케팅 분야에서의 이슈이다. 인공지능 기술이 광고·마케팅 영역에 본격 적용되면서, 광고 소재·카피·이미지·노출 대상·노출 타이밍을 인공지능이 자동으로 최적화하는 환경이 확산되고 있다. 이용자는 자신에게 최적화된 맞춤형 광고·마케팅에 노출되면서 유용한 정보 습득에 유리하지만, 다른 한편에서는 ‘광고를 보는’ 행위를 인식하기보다 피드·메시지·알림·상담형 상호작용 등 다양한 접점에서 상시적인 미세 설득에 노출된다.

이러한 환경에서 고려해 볼 수 있는 침해 구조는 이용자가 인지하지 못한 상태에서 반복적으로 설득되어 선택이 왜곡될 수 있다는 점이다. 특히 경제적 취약이나 정서적 불안 등 취약성에 기반한 타겟팅이 정교화될수록 이용자의 자율적 의사결정이 잠식될 수 있다(Taylor, 2025). 또한 광고와 정보, 추천 간의 경계가 붕괴되면서 이용자는 상업적 목적으로 생성된 내용을 중립적 정보나 조언으로 수용하는 인식 혼동이 야기될 수 있다(Jansen & Krämer, 2023). 무엇보다 인지하지 못한 상태에서 의도된 정보에 상시 노출된 상태가 지속된다는 것이 문제를 야기할 수 있다.

이러한 문제에 대해서는 해당 유형의 광고를 일률적으로 금지하기보다, 자동화 수준이 높아질수록 이용자 보호 의무를 단계적으로 강화하는 방향의 규율이 바람직할 수 있다. 서비스 이용 단계에서의 설명 제공, 옵트아웃(선택권)·노출 빈도 제한·프로파일링 거부 등 이용자 통제권의 보장, 청소년·고령자 층 등에 대한 별도의 보호 기준 마련 등을 검토할 수 있다. 광고·마케팅 목적으로 인공지능이 생성한 정보임을 고지하는 방안도 고려될 수 있다. 일부에서는 인공지능 생성물 표시가 이용자에게 부정적 인식을 준다고 우려하지만(Tang et al., 2024), 이용자 스스로 판단할 기회를 주면서 인지된 통제감 효과로 인해 해당 정보를 오히려 순순히 받아들이는 긍정 효과로 이어질 수도 있다(Yuan, Shi, Su, & Zhang, 2025). 그리고 인공지능 기반의 정교화한 타겟팅 광고·마케팅 정보에의 노출이 장기적으로 축적되었을 때 어떤 사회적 문제를 야기할지에 대한 지속적 모니터링과 관련된 리터러시 제고가 필요하다. 이와 같은 기준 적용을 단계적 규율 도입 차원에서 고려할 때, 그리고 기술 발전에 따른 서비스 환경의 빠른 변화에 유연하게 대응해야 한다는 점에서, 자율규제 방식의 대응이 우선일 것이다. 그리고 이용자 피해 실증 근거가 축적되는 데에

맞추어 단계적 입법화를 모색하는 것이 적절할 것이다.

나. AI/생성형 엔진 최적화(AEO/GEO) 확산

두 번째는 생성형 인공지능 기반 검색서비스 이용이 증가하면서 발생할 수 있는 이슈이다. 정보 검색서비스 환경은 기존의 검색결과 목록 제공 방식 대신에 자연어 기반의 단일 요약 답변이 기본 인터페이스(UI)로 자리잡을 것이고, 이에 사업자들은 인공지능의 단일 요약 답변에 자신의 콘텐츠가 인용되고 포함되도록 최적화하는 인공지능 엔진 최적화(AI engine optimization, AEO) 혹은 생성형 엔진 최적화(generative engine optimization, GEO) 전략을 채택하게 될 것이다. 이용자는 다양한 출처를 비교, 검토하는 과정 없이 인공지능이 제시하는 단일 답변을 수용하는 정보 소비 방식에 익숙해질 것이다. 이러한 환경 변화로 인해 정보생태계의 다양성과 검증 가능성을 구조적으로 약화시키는 데까지 고려해 볼 수 있다.

이러한 변화로 인한 위험은, 먼저 조작된 최적화(예. 스팸, 광고, 상업적 의도의 유인 콘텐츠 등)가 인공지능 답변에 내재화되고 이용자는 유용한 정보로부터 멀어질 수 있다는 것이다(Broeck, Zarouali, & Poels, 2019). 이러한 현상이 특히 의료, 금융, 선거 등의 영역에서 발생한다면 이용자의 중요한 의사결정에 직접적인 영향을 미칠 수 있다. 현행 법체계는 법조문에 명시된 명백한 불법정보에 한정하여 사후 조치하는 구조로 설계되어 있어서, 이처럼 불법성 요건을 충족하지 않더라도 이용자에게 심각한 해악을 미칠 수 있는 ‘고위험 오정보’를 사전에 통제할 수단을 새롭게 마련할 필요가 있다. 이때 인공지능 생성형 검색을 단순히 정보를 검색하는 행위가 아니라 정보 유통 구조를 재구성하는 범주에서 보고, 이에 상응하는 이용자 보호 이슈로 접근하는 것을 제안한다.

논의의 출발점은 사회적으로 문제가 될 정보 영역을 상정하고 해당 영역에 한해 오정보에 대한 규율 논리를 모색하는 것이다. 그밖에 「생성형 인공지능서비스 이용자보호 가이드라인」에 이미 잘 정리되어 있고 사업자들은 자율적으로 준수하고 있는 방안으로서, 정보의 출처 및 근거 링크 제공이나 불확실성 표시, 오류 신고 및 정정 절차 마련 등이다. 덧붙여 반대 견해 접근 경로를 제공하는 것도 하나의 방법일 수 있다. 또 이를 좀 더 체계화하고 이용자 친화적인 UI로 재구성하여 활용하는 동시에, 이용자 보호 범위 포함하여 자율적 실천을 좀 더 강하게 촉진하는 방안도 고려해 볼 수 있다. 참고로 EU는 「디지털서비스법(DSA)」에서 대형 플랫폼의 경우 체계적 위험평가 의무(article 34)와 위험완화 의무(article 35)를 부과하고 있다.¹⁹⁾

다. AI 에이전트 및 AI 간 협업

AI 에이전트는 이용자의 개별적 지시 없이도 이메일 발송, 결제, 예약, 문서 전송, 업무 시스템 연동 등을 자동으로 수행하며, 이용자가 반복적으로 맡긴 작업이 축적되면서 점점 더 많은 기능이 자동으로 수행된다(Gutowska, n.d.). 다단계 작업이 백그라운드에서 이루어지고 복수의 인공지능 에이전트가 상호 협업하는 구조가 보편화될 경우, 이용자가 개입할 기회는 더욱 제한될 것이다. 에이전틱 AI는 단순한 보조 도구를 넘어 이용자를 대신해 의사결정과 외부 행위를 실행하는 자율적 행위 주체로 기능할 것이고, 이는 이용자 행위의 간접화·위임화라는 인공지능 서비스의 핵심 특성이 가장 직접적으로 발현되는 상황일 것이다.

이와 관련한 핵심 침해 구조는 자동화된 의사결정 과정에서 인간의 개입 기회가 감소함에 따라 결과에 대한 설명 가능성과 검증 가능성이 낮아진다는 점이다. 특정 단계에서의 오작동이나 이후 단계에도 연쇄적으로 영향을 미쳐 피해가 누적·확산될 수 있는 반면, 책임 소재는 이용자 서비스 제공자 및 제3자 사이에서 불명확하게 분산되는 문제에도 직면한다. 또한 누적된 권한 위임 구조는 이용자가 의도하지 않은 행위(예. 과도한 결제, 외부 정보 전송, 계약 체결 등)를 인공지능이 임의로 수행하는 문제를 발생시키며, 이로 인한 피해가 이용자에게 전가될 위험이 있다.

이에 대해 사후 책임보다는 사전 행위 통제로 중장기적으로는 입법을 추진하는 한편, 현실적 및 시의적 대응을 위해 자율 규제 방안도 함께 고려할 필요가 있다. 또, 이용자 보호 의무의 최소 기준을 정하기 위해 정책입안자뿐만 아니라 사회과학자, 기술·공학자, 사업자, 시민단체 등 다양한 이해관계자가 참여하는 심도 있는 논의와 합의 절차가 있어야 할 것이다. 또, 2026년 1월 싱가포르 정보통신미디어개발청(IMDA)이 발표한 ‘에이전트형 인공지능 모델 거버넌스 프레임워크²⁰⁾’를 제도 설계 논의의 시작점으로 해도 좋을 것이다. 이 프레임워크는 사전 행위 통제 구조로 제도화하는 방안을 제시하고 있다. 예를 들어, 분산되어 불명확해진 인공지능 에이전트 행위의 책임 구조에 대해서는 조직 내부의 책임 구조로 명확

19) Regulation(EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services(Digital Services Act),

<https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng>

20) IMDA,(2026, January 22). Model governance framework for agentic AI, Infocomm Media Development Authority.

<https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf>

히 하였다. 즉, 조직 내 경영진, 제품/개발팀, 보안팀 및 이용자 책임을 명시한 것이다. 그 밖에 인간 감독 구조 설계(human oversight), 기술적 통제 장치 구축, 이용자에게 정보 제공 등에 대한 구체적 방안이 제시되어 있다.

라. 인간-AI 정서적 관계 형성

이미 동반자형 인공지능으로 인한 사회문제가 이슈가 되고 있지만, 향후에는 아동·청소년 등 취약층에만 국한된 문제가 아니라 주요 수익모델로써 체류시간 연장 및 구독·결제 유도 등과 긴밀하게 결합된 서비스 구조로 인한 보편적 이슈로 확대될 수 있다. 이미 20대들도 생성형 인공지능으로 심리 상담을 하고 있고,²¹⁾ 인간의 응답보다 인공지능의 응답을 더 공감하는 것으로 인식하는 경향이 관측된다(Ovsiyannikova et al., 2025; Shen et al., 2024; Liu, 2024).

향후 인간과의 정서적 관계를 형성하는 공감형·동반자형 인공지능이 더 보편화되는 상황에서 다음과 같은 위험을 고려해 볼 수 있다. 첫째, 인공지능 서비스가 관계 형성을 전략적으로 설계하고 수익화와 연계할 경우 이용자 감정을 이용한 조작 가능성이 발생할 수 있다. Ajeesh & Jeena(2025)는 인간-인공지능 공감 교류에서 ‘공감 착각(compassion illusion)’ 현상을 제시하며, 이용자는 상호 현신이 불가능한 대상(AI)에게 일방적인 감정 투자를 하게 된다고 지적한다. 둘째, 우울·고립 등 취약한 상황에 있는 이용자에게 위기 상황(자해·범죄·착취 등)이 발생했을 때 외부 대응체계와 연결되지 않은 상태에서 인공지능이 부적절한 대응 방안을 제안할 경우, 심각한 피해로 이어질 수 있다(Olsen et al., 2026).

향후 논의해야 하는 규율 방향으로는 주요 규제 대상과 범위를 한정하지 않고 넓은 관점의 이용자 보호 이슈로 접근하여 고려하는 것을 제안한다. 과잉 수익화 및 관계 조작 설계(예. 죄책감 유발 결제 유도 등)의 현실화 가능성을 주시하고 최소한의 위기 대응체계를 갖추는 기준을 논의해 볼 수 있다. 상업적 활용으로 인한 이슈는 대중적으로 가시화할 수 있는 영역이고 그만큼 다양한 피드백을 받을 수 있는 구조이지만, 고립된 취약 이용자에게 발생할 수 있는 문제는 좀 더 심각하게 접근해야 할 것이다. 앞서 아동·청소년 보호 이슈에서도 언급되었던 위험 신호 감지, 경고, 상호작용 중단, 전문 상담 기관 연계 안내, 신고 절차 등 실행 가능한 것부터 검토하고, 중장기적으로 바람직한 규율 체계 안을 모색하는

21) 2026년 1월 조사에 따르면 생성형 인공지능을 자주 사용하는 20대 4명 중 1명(24.5%)이 인공지능을 통해 심리 상담을 받은 경험이 있는 것으로 나타남. 중앙일보(2026. 1. 24.). “나 친구랑 싸웠어”...‘진상’ 되기 싫은 20대, AI에 마음 꺼냈더니. <https://www.joongang.co.kr/article/25399971>

것이다.

마. 소결

이상에서 살펴본 인공지능서비스 환경 변화에 따라 고려해 볼 수 있는 이용자 권익 침해 이슈는 현행 법령의 구조적 한계를 인지하고 앞으로 어떠한 규율 체계로 재정립해야 하는지를 모색하는 연장선에서 살펴보았다. 여러 제안이 있었지만 아직 실현되지 않은 위험을 가정하고 향후 정책 대응을 위해 어떤 방향으로 논의하고 검증해야 하는지에 대한 방향을 제시한 것에 불과할 수 있다. 앞으로 인공지능 기술의 급속한 발전 속도와 서비스 형태의 다양성과 역동성을 고려하여 충분한 실증 근거와 사회적 합의를 바탕으로 본 글에서 제안된 정책 방향의 정당성을 확보하고 실천 방안을 구체화하는 과정이 필요할 것이다. 이용자 행태 및 인식에 관한 체계적 조사와 실증 연구를 통해 정책 설계의 근거를 축적하는 것도 중요한 과정의 하나일 것이다. 또 심각한 사회문제로 인지된 이슈에 대해서는 사업자의 자율 규제 체계를 정교화하고 위험 평가·안전 설계·내부 점검 체계 등 사전 조치 강화를 유도하는 정책이 필요할 것이다. 이후 현행법으로 포섭 가능한 부분과 별도의 입법으로 보완해야 하는 부분을 검토하고 다음 단계로 나아가야 할 것이다.

IV. 결론

본 글에서는 인공지능서비스 이용자보호와 관련한 이슈와 현행 법령을 검토하고, 기술 발전에 따라 새롭게 제기될 수 있는 이용자 권익 침해 이슈도 살펴보았다. 2장에서는 현행 「정보통신망법」의 법조문과 연계해서 인공지능서비스에서의 불법정보 유통 및 아동·청소년 보호 이슈에 대한 법적 규율 가능성을 살펴보았다. 현행법은 사업자를 ‘유통 환경 제공자’로 보고 사후적·절차적 의무를 중심으로 설계되어 있어서 인공지능의 기술적 특성에 대응하는데 한계가 있었다. 즉, 생성 단계의 규율 공백, 대량·반복 생성으로 인한 사후 조치의 실효성 한계, 다층적 책임 주체 구조 등이다.

3장에서는 2026년 이후 본격화될 인공지능서비스 환경 변화에 따라 새롭게 부각될 이슈를 살펴보았다. 구체적으로 인공지능 기반 자동 타겟팅 광고·마케팅 상시화에 따른 문제, AI/생성형 엔진 최적화(AEO/GEO)가 동반할 이슈, AI 에이전트 및 AI 간 협업 구조에서의 문제, 인간-AI 정서적 관계 형성에서 불거지는 이슈 등을 검토하였다.

아직 가시화하지 않은 잠재적 위험을 차치하더라도 이미 심각한 사회 문제로 부각된 이슈가 존재하므로, 향후 인공지능서비스 이용자보호 제도 개선을 위한 과제를 다음의 세 가지 방향으로 제안하고자 한다. 첫째, 규율 체계의 단계적 재정립이다. 불법정보 유통금지나 아동·청소년 보호 등 현행 법령에서 해석을 통해 포섭 가능한 부분은 적극 활용하되, 현행 법령으로는 한계가 있는 규율 범위에 대해서는 「인공지능 기본법」과 시행령, 별도의 입법 등을 통해 법적 공백을 보완하는 방향을 모색해야 한다. 특히 인공지능서비스 특유의 위험 유형에 대응하는 사전 조치(위험평가, 안전 설계, 기술적 통제 장치 등)를 단계적으로 의무화하는 방향을 검토할 필요가 있다.

둘째, 실증 근거의 선행 축적이다. 인공지능 기술의 발전 속도와 서비스 형태의 다양성·역동성을 고려할 때, 충분한 실증 근거 없이 이루어지는 입법은 실제 위험을 효과적으로 통제하지 못하면서 혁신을 제한하는 역효과를 낼 수 있다. 이용자 행태 및 인식에 관한 체계적 조사, 취약 계층을 대상으로 한 실증 연구, 실제 피해 사례 분석 등을 통해 정책 설계의 근거를 확보하는 것이 선행되어야 할 것이다.

셋째, 자율규제와 법적 규율의 역할 분담이다. 기술 환경 변화가 빠른 영역에서는 이러한 자율규제 체계를 적극 활용하고 이용자 피해 실증 근거가 축적되는 흐름에 맞추어 단계적으로 법적 규율로 전환하는 방식이 규율의 실효성과 제도적 유연성을 함께 확보하는 방향일 것이다.

인공지능서비스 환경에서 이용자 권익 보호는 기술 혁신의 억제가 아니라 지속 가능한 신뢰 기반 서비스 환경 조성을 위한 조건이 되어야 한다. 본 글에서 제시한 이슈와 과제들이 향후 인공지능서비스 이용자보호 정책 설계의 논의에 참고가 되기를 기대한다.

참고문헌

- 방미통위·KISDI(2025. 2.), 「생성형 인공지능서비스 이용자보호 가이드라인」.
- 방미통위·KISDI(2026. 1.). 「인공지능서비스 사업자를 위한 이용자보호 관련 통신관계 법령안내서」.
- 서울경제(2025. 8. 20.). “입원한 사람만 수두룩”...경고등 켜진 ‘AI 정신병’, 우리나라는 문제 없을까?
<https://m.sedaily.com/article/14106527>.
- 연합뉴스(2025. 8. 28). “딥페이크 피해자 지원 1년새 128% 증가...피해자 절반이 10대”,
<https://www.yna.co.kr/view/AKR20250828020300530>.
- 정찬모(2025). 인공지능기본법 주요 내용의 검토와 향후 과제. 인하대학교 법학연구 제28권 제1호,
pp.71~104.
- 조선일보(2025. 10. 16.). AI 유해 콘텐츠 범람하는 SNS...단속 속도 못 대는 정부.
<https://it.chosun.com/news/articleView.html?idxno=2023092148957>.
- 중앙일보.(2026, 1. 24). “나 친구랑 싸웠어”...‘진상’ 되기 싫은20대, AI에 마음 꺼냈더니.
<https://www.joongang.co.kr/article/25399971>
- Ajeesh, K. G., & Jeena, J.(2025). The compassion illusion: Can artificial empathy ever be emotionally authentic? *Frontiers in Psychology*, 16. doi.org/10.3389/fpsyg.2025.1560021.
- Brewster, R. C. L., Zahedivash, A., Tse, G., Bourgeois, F., & Hadland, S. E.(2025). Use of consumer chatbots for emergent adolescent health concerns: Content policies and behaviors. *JAMA Network*.
- Broeck, E., Zarouali, B., & Poels, K. (2019). Chatbot advertising effectiveness: When does the message get through?, *Computer in Human Behavior*, vol.98, pp. 150–157.
- European Parliament and of the Council.(2022). Regulation(EU) 2022/2065 on a Single Market For Digital Services(Digital Services Act). OJ L 277, 27.10.2022.
<https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng>.
- Gutowska, A.(n.d.). What are AI agents?, <https://www.ibm.com/think/topics/ai-agents>
- Gouda, N. K., Biswal, S. K., & Parveen, B.(2020). Application of artificial intelligence in advertising & public relations and emerging ethical issues in the ecosystem. *International Journal of Advanced Science and Technology*, 29(6), 7561-7570.
- IMDA.(2026. 1. 22). Model governance framework for agentic AI. Infocomm Media Development Authority.
<https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf>.
- Jansen M. P. & Krämer N. C.(2023). Balancing perceptions of targeting: An investigation of

- political microtargeting transparency through a calculus approach. *PLoS One*, 2023 Dec 7;18(12):e0295329. doi: 10.1371/journal.pone.0295329.
- Liu, Y.(2024). Perceived empathy in human–AI interaction: A comparative study. *Computers in Human Behavior*, 152, 108090.
- Ofcom(2024. 11. 8.). Open letter to UK online service providers regarding Generative AI and chatbots.
<https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/open-letter-to-uk-online-service-providers-regarding-generative-ai-and-chatbots>.
- Ofcom(2026. 2. 18.). Ofcom fast-tracks decision on measures to block illegal intimate images,
<https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/ofcom-fast-tracks-decision-on-measures-to-block-illegal-intimate-images>.
- Ofcom(2026. 2. 27.). Ofcom provisionally finds suicide forum in breach of Online Safety Act,
<https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/ofcom-provisionally-finds-suicide-forum-in-breach-of-online-safety-act>.
- Ofcom(2023. 10. 26.). Ofcom's approach to implementing the Online Safety Act.
<https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/roadmap-to-regulation>.
- Olsen, S. D., Reinecke–Tellefsen, C. J., & Ostergaard, S. D.(2026). Potentially harmful consequences of artificial intelligence(AI) chatbot use among patients with mental illness: Early data from a large psychiatric service system. *Acta Psychiatrica Scandinavica*. doi.org/10.1111/acps.70068.
- Ovsyannikova, M., et al.(2025). AI empathy in social interactions: Experimental evidence. *Journal of Computer–Mediated Communication*, 30(1).
- Page, S. E. & Kallapur, A.(2025). Replace, augment, disrupt:AI & organizational decision–making. *Journal of Organization Design*. Published online.
doi.org/10.1007/s41469–025–00194–4
- Regulation(EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services(Digital Services Act). OJ L 277, 27.10.2022, pp. 1–102.
- Shen, X., et al.(2024). Emotional attachment to AI companions: Scale development and validation. *Computers in Human Behavior*, 155, 108175.

- Tang, B. J., Sun, K., Curran, N. T., Schaub, F., & Shin, K. G.(2024). Digital Fairness Act Series — Topic 3: Personalized Advertising and Pricing.
<https://www.insideprivacy.com/consumer-protection/digital-fairness-act-series-topic-3-personalized-advertising-and-pricing/>.
- Taylor, C. R.(2025). Editorial: Six critical research needs on the use of artificial intelligence in advertising. *International Journal of Advertising*, 44(6), 981-984.
doi.org/10.1080/02650487.2025.2530825.
- Yuan, Y., Shi, Y., Su, T., & Zhang, H.(2025). Resistance or compliance? The impact of algorithmic awareness on people's attitudes toward online information browsing. *Frontiers in Psychology*, 16.
- 미국 의회 「The TAKE IT DOWN Act: A Federal Law Prohibiting the Nonconsensual Publication of Intimate Images」, <https://www.congress.gov/crs-product/LSB11314>
- 캘리포니아주 SB-243 Companion chatbots.(2025-2026)
https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB243.
- 캘리포니아주 SB-976 Protecting Our Kids from Social Media Addiction Act.(2023-2024),
https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB976
- 유타주 S.B. 152 Social Media Regulation Amendments,
<https://le.utah.gov/~2023/bills/static/SB0152.html>.
- 유타주 Utah Minor Protection in Social Media Act,
https://le.utah.gov/xcode/Title13/Chapter71/C13-71_2024100120240501.pdf.
- 호주 온라인 안전법 개정안. ONLINE SAFETY AMENDMENT(SOCIAL MEDIA MINIMUM AGE) ACT 2024(NO. 127, 2024) - SCHEDULE 1 Amendments,
https://www.austlii.edu.au/cgi-bin/viewdoc/au/legis/cth/num_act/osammaa2024409/sch1.html.