방통융합정책연구 KMCC-2025-18

‘정보 무결성’ 강화를 위한 중장기 정책 방향에 관한 연구

(A study on mid-to-long-term policy directions for strengthening ‘information integrity’)

권오상 외

2025. 12.

연구기관 : 디지털미래연구소

방송미디어통신위원회

제 출 문

방송미디어통신위원회 위원장 귀하

본 보고서를 『**'정보 무결성' 강화를 위한 중장기 정책 방향에 관한 연구**』의 연구결과보고서로 제출합니다.

2025년 12월

연구기관 : 사단법인 디지털미래연구소
총괄책임자 : 권오상 대표
참여연구원 : 김국진 고문
이찬구 연구위원
김슬기 연구위원
노윤경 책임연구원

목 차

표 목 차

그 림 목 차

요 약 문

1. 제 목: ‘정보 무결성’ 강화를 위한 중장기 정책 방향에 관한 연구

2. 연구 배경 및 필요성

디지털 플랫폼과 온라인 미디어 환경의 급속한 변화는 정보의 생산・유통・소비 구조를 근본적으로 재편하고 있으며, 이 과정에서 허위조작정보의 확산은 사회적 신뢰 형성 및 민주적 의사결정 과정의 안정성에 영향을 미치는 주요 요인으로 부상하고 있다. 특히 생성형 인공지능의 확산, 알고리즘 기반 추천 시스템, 플랫폼 간 정보 유통의 가속화는 허위조작정보의 생산 비용을 낮추는 동시에 그 파급력을 증폭시키고 있어, 기존의 규제 방식만으로는 효과적인 대응에 한계가 있다는 인식이 확산되고 있다.

그간 우리나라의 허위조작정보 대응은 형법, 정보통신망 이용촉진 및 정보보호 등에 관한 법률, 공직선거법 등 개별 법령에 따른 사후적 처벌과 행정적 조치를 중심으로 이루어져 왔다. 이러한 체계는 불법 정보에 대한 신속한 대응이라는 측면에서는 일정한 역할을 수행해 왔으나, 허위조작정보를 예방・절차・투명성 중심으로 관리하는 체계를 구축하는 데에는 구조적 한계가 있다는 평가가 제기되어 왔다. 특히 플랫폼의 자율적 책임, 민간 검증 생태계의 역할, 정보 유통 구조 전반에 대한 정책적 접근이 충분히 제도화되지 못한 채 공적 심의와 행정 집행에 과도하게 의존해 왔다는 점이 지속적인 문제로 지적되어 왔다.

이러한 한계 속에서 2025년에는 허위조작정보 대응과 관련한 정보통신망법 개정 논의가 본격화되면서, 허위조작정보를 명시적으로 규율할 수 있는 법적 근거와 절차적 대응 체계에 대한 보다 구체적이고 체계적인 논의의 필요성이 제기되고 있다. 2025년 말 개정된 정보통신망법은 허위조작정보에 대한 정의규정과 더불어 신고조치・이의제기・분쟁조정, 투명성 보고, 사실확인단체 연계 등 다양한 제도적 장치를 도입함으로써, 기존의 사후적・개별적 규율에서 벗어나 보다 구조적이고 제도화된 대응으로 전환할 수 있는 출발점을 제공한다. 다만 이러한 법적 틀이 실제 정책 효과로 이어지기 위해서는 하위 규범 설계, 집행 구조, 자율규제 및 공동규제 메커니즘과의 정합성 확보가 필수적이다.

유럽연합(EU), 영국, 독일, 호주 등 주요 국가들은 허위조작정보 대응을 개별 콘텐츠의 진위 판단이나 삭제 중심 규제에서 벗어나, 플랫폼의 시스템적 책임과 절차적 의무를 중심으로 한 공동규제(co-regulation) 또는 메타규제(meta-regulation) 방식으로 재편하고 있다. 이는 정부의 직접 판단을 최소화하면서도 플랫폼 책임과 외부 검증의 실효성을 확보하려는 정책적 전환으로 평가된다. 이러한 국제적 흐름은 우리나라 역시 글로벌 규제 환경과의 정합성을 고려한 정보 무결성 정책 체계 구축이 필요함을 시사한다.

이에 따라 허위조작정보 대응 정책은 '내용 규제'의 문제를 넘어, 정보가 생산·유통·추천·검증되는 전 과정의 조건을 개선하는 정보 무결성(information integrity) 관점에서 재구성될 필요가 있다. 정보 무결성 접근은 표현의 자유를 존중하면서도, 정보 환경 전반의 신뢰성과 회복탄력성을 제고하는 중장기 정책 방향을 모색할 수 있다는 점에서 중요한 정책적 의미를 갖는다.

3. 연구의 목적 및 범위

이에 본 연구의 목적은 허위조작정보 대응을 정보 무결성 관점에서 종합적으로 분석하고, 국내외 정책 동향과 선행연구, 전문가 인식을 바탕으로 한국 사회에 적합한 중장기 정보 무결성 정책 방향과 과제를 도출하는 데 있다. 이를 통해 정보통신망법 개정 이후의 제도적 공백을 보완하고, 플랫폼 책임, 자율규제·공동규제, 외부 검증 생태계, 재정 및 거버넌스 구조를 포괄하는 정책 프레임을 제시하고자 한다.

본 연구의 구성 및 범위는 다음과 같다. 먼저, 제1장에서는 연구 배경 및 필요성과 범위에 대해 설명한다.

제2장에서는 정보 무결성 개념과 허위조작정보의 관계를 정리하고, 유럽연합, 영국, 독일, 호주, 미국 등 주요 국가의 정보 무결성 관련 법제도 및 정책 동향을 분석한다. 아울러 국내 정보 무결성 관련 법제도와 정책 현황을 검토하고, 국내외 자율규제 및 공동규제 사례를 비교 분석함으로써 정책 설계의 시사점을 도출한다.

제3장에서는 국내외 허위조작정보 관련 학술 연구, UN·OECD 등 국제기구의 정보 무결성 정책 보고서, EU와 호주의 허위조작정보 실천강령 관련 보고서를 대상으로 메타 분석을 실시하여, 기존 연구와 정책 논의의 주요 쟁점과 한계를 체계적으로 정리한다.

제4장에서는 전문가 인식 조사를 통해 정보 무결성 정책에 대한 평가와 중장기 정책 방향에 대한 인식을 분석하고, 이를 토대로 정책적 시사점을 도출한다.

제5장에서는 앞선 분석 결과를 종합하여 정보 무결성 정책의 중장기 기본 방향을 제시하고, 법·재정·거버넌스 차원의 정책 전략과 과제를 구체적으로 제안한다. 이는 단기적 제도 정비를 넘어 지속가능한 공동규제 체계 구축을 목표로 한다.

마지막으로 제6장에서는 연구 결과를 종합하여 정책적 함의를 정리하고, 향후 정보 무결성 정책 추진을 위한 과제를 제시한다.

구 분	세부 연구내용
서 론	o 연구배경 및 필요성 o 연구목적 및 범위
국내외 동향 분석	o 국내외 허위조작정보, 정보 무결성 확보 관련 주요 현황 분석 o 국내외 관련 법제도 및 주요 정책동향 분석 o 국내외 자율규제/공동규제 사례 분석
선행연구 메타분석	o 국내외 허위조작정보 관련 연구 메타분석 o EU, 호주의 허위조작정보 실천강령 관련 보고서 메타분석 o OECD, UN의 정보 무결성 정책 관련 연구 메타분석
전문가 및 이해관계자 조사	o 전문가 및 이해관계자 전문가 조사 진행 o 중장기 정책 시사점 도출
중장기 정책방향 제안	o 단기 과제 o 중장기 정책 전략 및 과제

4. 연구 내용 및 결과

본 연구는 디지털 환경에서 허위조작정보 확산이 민주주의, 사회적 신뢰, 이용자 권익에 미치는 영향을 종합적으로 검토하고, 이를 구조적으로 대응하기 위한 정보 무결성(information integrity) 정책 체계의 필요성과 방향을 분석하는 것을 목적으로 수행되었다.

이를 위해 첫째, 정보 무결성 개념과 허위조작정보의 유형·특성을 정리하고, 국내외 학술 연구 및 국제기구 보고서에 대한 메타분석을 통해 주요 쟁점과 정책 논의를 체계화하였다. 둘째, EU, 영국, 독일, 호주, 미국 등 주요국의 법제도 및 자율규제·공동규제 사례를 비교 분석하여, 각국이 플랫폼 책임을 절차·시스템 중심으로 구조화하는 방향으로 정책 패러다임을 전환하고 있음을 확인하였다. 셋째, 국내 법제 및 제도 현황 분석을 통해, 기존의 사후적·개별적 규율 방식이 갖는 한계를 진단하고, 2025년 12월 정보통신망법 개정을 계기로 허위조작정보 대응이 절차 중심의 제도적 틀로 전환되고 있음을 분석하였다.

아울러 전문가 인식조사를 통해 현행 제도의 효과성과 한계, 향후 정책 설계 시 고려되어야 할 핵심 요소를 도출하였다. 조사 결과, 전문가들은 개별 콘텐츠의 진위 판단이나 직접적 제재보다는 ▲플랫폼의 절차적 책임 강화 ▲외부 검증 생태계(팩트체크·연구)의 지속가능성 확보 ▲정치적 중립성과 독립성을 담보하는 거버넌스 구축이 중요하다는 데 공통된 인식을 보였다.

종합적으로 본 연구는 허위조작정보 대응 정책이 단기적 규제나 일회성 조치에 머물러서는 실효성을 확보하기 어렵고, 정보의 생산·유통·추천·검증 전 과정을 포괄하는 정보 무결성 관점의 중장기 정책 체계로 발전할 필요성을 제시하였다. 향후 허위조작정보 대응 정책은 개별 콘텐츠에 대한 사후적 판단과 제재를 중심으로 한 방식에서 나아가, 해당 정보가 확산되는 구조와 조건을 점검하고 개선하는 방향으로 전환될 필요가 있다. 이를 위해 법적 규율, 자율규제, 공동규제 수단을 유기적으로 결합하고, 투명성, 절차적 정당성, 독립성 확보를 핵심 원칙으로 삼는 정책 체계를 구축하는 것이 바람직하다. 이러한 접근은 표현의 자유와 공익 보호 간의 균형을 유지하면서, 디지털 정보 생태계 전반의 신뢰 회복과 민주적 공론장의 기능 강화를 도모하는 데 기여할 것으로 기대된다.

5. 정책적 활용 내용

본 연구 결과는 향후 정보 무결성 관련 법제도 정비 및 정책 설계 과정에서 다음과 같은

방식으로 활용될 수 있다.

첫째, 2025년 개정 정보통신망법의 후속 시행령·고시 마련 과정에서, 신고·조치·이의제기·투명성 보고 등 절차적 의무의 구체화를 위한 기초 자료로 활용될 수 있다. 특히 플랫폼별로 상이한 내부 운영 기준을 최소한의 공통 절차 기준으로 정합화하는데 있어, 본 연구에서 도출한 해외 공동규제 사례와 전문가 인식 결과는 실질적인 참고 기준으로 활용될 수 있다.

둘째, 플랫폼 책임을 콘텐츠 판단 중심이 아닌 시스템·절차 중심으로 구조화하는 중장기 정책 방향 설정에 활용될 수 있다. 이는 과잉 규제 및 표현의 자유 침해 논란을 완화하면서도, 플랫폼의 사회적 책임을 제도적으로 확보하는 정책 설계에 기여할 수 있다.

셋째, 팩트체크 및 공익적 검증 활동 지원과 관련하여, 플랫폼 출연과 공적 재원을 결합한 재원 조성 모델, 독립적 기금(arm's length) 구조, 법률·신변 보호 등 최소 안전망 설계에 대한 정책적 근거 자료로 활용될 수 있다. 이를 통해 국내 팩트체크 역량을 점진적으로 축적하고, 외부 검증 생태계의 지속가능성을 제고하는 데 기여할 수 있다.

6. 기대효과

본 보고서의 연구결과는 허위조작정보 대응을 포함한 정보 무결성 정책의 패러다임 전환이 요구되는 시점에서, 법·제도, 재원, 거버넌스 측면의 핵심 쟁점에 대한 중장기적 정책 방향과 개선 과제를 제시함으로써, 향후 정부의 정보 무결성 관련 정책 수립 및 제도 설계에 참고 자료로 활용될 것으로 기대된다.

제1장. 서론

제1절 연구배경 및 필요성

디지털 플랫폼과 온라인 미디어 환경의 급속한 변화는 정보의 생산・유통・소비 구조를 근본적으로 재편하고 있으며, 이 과정에서 허위조작정보의 확산은 사회적 신뢰 형성 및 민주적 의사결정 과정의 안정성에 영향을 미치는 주요 요인으로 부상하고 있다. 특히 생성형 인공지능의 확산, 알고리즘 기반 추천 시스템, 플랫폼 간 정보 유통의 가속화는 허위조작정보의 생산 비용을 낮추는 동시에 그 파급력을 증폭시키고 있어, 기존의 규제 방식만으로는 효과적인 대응에 한계가 있다는 인식이 확산되고 있다.

그간 우리나라의 허위조작정보 대응은 형법, 정보통신망 이용촉진 및 정보보호 등에 관한 법률, 공직선거법 등 개별 법령에 따른 사후적 처벌과 행정적 조치를 중심으로 이루어져 왔다. 이러한 체계는 불법 정보에 대한 신속한 대응이라는 측면에서는 일정한 역할을 수행해 왔으나, 허위조작정보를 예방・절차・투명성 중심으로 관리하는 체계를 구축하는 데에는 구조적 한계가 있다는 평가가 제기되어 왔다. 특히 플랫폼의 자율적 책임, 민간 검증 생태계의 역할, 정보 유통 구조 전반에 대한 논의가 충분히 이루어지지 못한 채 공적 심의와 행정 집행에 과도하게 의존해 왔다는 점이 지속적으로 지적되어 왔다.

이러한 한계 속에서 2025년에는 허위조작정보 대응과 관련한 정보통신망법 개정 논의가 본격화되면서, 허위조작정보를 명시적으로 규율할 수 있는 법적 근거와 절차적 대응 체계에 대한 보다 구체적이고 체계적인 논의의 필요성이 제기되고 있다. 2025년 말 개정된 정보통신망법은 허위조작정보에 대한 정의 규정과 더불어 허위조작정보에 대한 신고 및 조치, 이에 대한 이의제기, 분쟁조정, 투명성 보고, 사실확인단체 연계 등 다양한 제도적 장치를 도입함으로써, 기존의 사후적・개별적 규율에서 벗어나 보다 구조적이고 제도화된 대응으로 전환할 수 있는 출발점을 제공한다. 다만 이러한 법적 틀이 실제 정책 효과로 이어지기 위해서는 하위 규범 설계, 집행 구조, 자율규제 및 공동규제 메커니즘과의 정합성 확보가 필수적이다.

유럽연합(EU), 영국, 독일, 호주 등 주요 국가들은 허위조작정보 대응을 개별 콘텐츠의 진위 판단이나 삭제 중심 규제에서 벗어나, 플랫폼의 시스템적 책임과 절차적 의무를 중

심으로 한 공동규제(co-regulation) 또는 메타규제(meta-regulation) 방식으로 재편하고 있다. 이는 정부의 직접적인 판단을 최소화하면서도 플랫폼 책임과 외부 검증의 실효성을 확보하려는 정책적 전환으로 평가된다. 이러한 국제적 흐름은 우리나라 역시 글로벌 규제 환경과의 조화를 고려한 정보 무결성 정책 체계 구축이 필요함을 시사한다.

이에 따라 허위조작정보 대응 정책은 '내용 규제'의 문제를 넘어, 정보가 생산·유통·추천·검증되는 전 과정의 여건을 개선하는 정보 무결성(information integrity) 관점에서 재구성될 필요가 있다. 정보 무결성 접근은 표현의 자유를 존중하면서도, 정보 환경 전반의 신뢰성과 회복탄력성을 제고하는 중장기 정책 방향을 모색할 수 있다는 점에서 중요한 정책적 의미를 갖는다.

제2절 연구목적 및 범위

이에 본 연구의 목적은 허위조작정보 대응을 정보 무결성 관점에서 종합적으로 분석하고, 국내외 정책 동향과 선행연구, 전문가 인식을 바탕으로 한국 사회에 적합한 중장기 정보 무결성 정책 방향과 과제를 도출하는 데 있다. 이를 통해 정보통신망법 개정 이후의 제도적 공백을 보완하고, 플랫폼 책임, 자율규제·공동규제, 외부 검증 생태계, 재정 및 거버넌스 구조를 포괄하는 정책 프레임을 제시하고자 한다.

본 연구의 구성 및 범위는 다음과 같다. 먼저, 제1장에서는 연구 배경 및 필요성과 범위에 대해 설명한다.

제2장에서는 정보 무결성 개념과 허위조작정보의 관계를 정리하고, 유럽연합, 영국, 독일, 호주, 미국 등 주요 국가의 정보 무결성 관련 법제도 및 정책 동향을 분석한다. 아울러 국내 정보 무결성 관련 법제도와 정책 현황을 검토하고, 국내외 자율규제 및 공동규제 사례를 비교 분석함으로써 정책 설계의 시사점을 도출한다.

제3장에서는 국내외 허위조작정보 관련 학술 연구, UN·OECD 등 국제기구의 정보 무결성 정책 보고서, EU와 호주의 허위조작정보 실천강령 관련 보고서를 대상으로 메타 분석을 실시하여, 기존 연구와 정책 논의의 주요 쟁점과 한계를 체계적으로 정리한다.

제4장에서는 전문가 인식 조사를 통해 정보 무결성 정책에 대한 평가와 중장기 정책 방향에 대한 인식을 분석하고, 이를 토대로 정책적 시사점을 도출한다.

제5장에서는 앞선 분석 결과를 종합하여 정보 무결성 정책의 중장기 기본 방향을 제시하고, 법·재정·거버넌스 차원의 정책 전략과 과제를 구체적으로 제안한다. 이는 단기적 제도 정비를 넘어 지속가능한 공동규제 체계 구축을 목표로 한다.

마지막으로 제6장에서는 연구 결과를 종합하여 정책적 함의를 정리하고, 향후 정보 무결성 정책 추진을 위한 과제를 제시한다.

구 분	세부 연구내용
서 론	o 연구배경 및 필요성 o 연구목적 및 범위
국내외 동향 분석	o 국내외 허위조작정보, 정보 무결성 확보 관련 주요 현황 분석 o 국내외 관련 법제도 및 주요 정책동향 분석 o 국내외 자율규제/공동규제 사례 분석
선행연구 메타분석	o 국내외 허위조작정보 관련 연구 메타분석 o EU, 호주의 허위조작정보 실천강령 관련 보고서 메타분석 o OECD, UN의 정보 무결성 정책 관련 연구 메타분석
전문가 및 이해관계자 조사	o 전문가 및 이해관계자 전문가 조사 진행 o 중장기 정책 시사점 도출
중장기 정책방향 제안	o 단기 과제 o 중장기 정책 전략 및 과제

제2장. 정보 무결성 관련 국내외 동향 분석

제1절 허위조작정보와 정보 무결성

1. 허위조작정보 확산과 사회적 신뢰 기반의 약화

가. 허위조작정보 개념에 관한 논의

허위조작정보(disinformation)는 단순히 사실과 부합하지 않는 정보라는 의미를 넘어, 어떠한 요건을 갖춘 경우 이를 규범적으로 허위조작정보로 평가할 수 있는지에 관한 문제를 포함한다. 국제적 논의에서는 허위성 외에도 정보의 생성 및 유포 과정에서 행위자의 의도성, 즉 오인 또는 기만을 목적으로 한 의도적 조작 여부를 핵심 요소로 삼는 기준들이 제시되고 있다. 이러한 점에서 허위조작정보의 개념은 단순한 사실 오류(misinformation)와 구별되는 규범적 판단을 전제로 한다.

예컨대 미국의 대표적 사전인 Merriam-Webster는 허위조작정보를 “여론을 조작하거나 진실을 흐리기 위해 고의적이고 은밀하게 확산되는 허위 정보”로 정의하면서, 허위성뿐 아니라 고의성, 은밀성, 목적성을 주요 요건으로 제시한다(Merriam-Webster).

유럽연합(EU) 역시 허위조작정보를 “경제적·정치적 이익을 얻으려는 의도로 유포되는 허위 또는 오해의 소지가 있는 콘텐츠로서, 공공에 해를 끼칠 수 있는 것”으로 정의하면서, 단순한 오보나 비의도적 오류, 풍자나 조롱 등과는 구별하고 있다(European Commission, 2020). EU는 또한 허위조작정보와 구별되는 개념으로, ” 유해한 의도 없이 공유되는 허위 또는 오해의 소지가 있는 정보 “를 잘못된 정보(misinformation)로 명명하고 있다.

유네스코(UNESCO)는 허위조작정보(disinformation)를 ” 개인, 사회집단, 조직 또는 국가에 해를 끼치고자 하는 의도로 만들어진 거짓 정보 “로 정의하며, 잘못된 정보(misinformation)는 ” 사실이 아니지만 해를 끼칠 의도 없이 만들어진 정보 “, 악의적 정보(mal-information)는 ” 사실에 기반하고 있으나 해를 가할 목적으로 사용되는 정보 “로 구분한다(UNESCO, 2018).

<표 2-1> 허위조작정보의 정의 비교

구분	허위성	의도성	목적성	해악성	적용예외
EU 강령 2018	실증적으로 거짓이거나, 오해를 불러일으키는 정보	의도적으로 공중을 속이기 위해	경제적 이익을 얻거나	공공에 해를 끼칠 수 있는	- 오해를 불러일으키는 광고 - 보도 오류 - 풍자·패러디 - 명백히 식별되는 편파적 뉴스·논평 등
EDAP	거짓이거나, 오해를 불러일으키는 콘텐츠	의도적으로 속이거나	경제적 또는 정치적 이익을 얻기 위해	공공에 해를 끼칠 수 있는	- 풍자·패러디 - 정부승인 콘텐츠 - 정치광고 - 정당인가 콘텐츠 - 전문적 뉴스 콘텐츠
호주 강령	실증적으로 거짓이거나, 오해를 불러일으키거나, 기만적인 디지털콘텐츠	부정행위(스팸, 조작 등)를 통해	-	해당 내용의 유포가 피해를 끼치리라고 합리적으로 예상되는 경우	
OECD <Fact not Fake>	거짓, 부정확한, 또는 오해를 불러일으키는 정보	의도적으로	-	개인, 사회집단, 조직 또는 국가에 해를 끼치기 위해	
OECD 권고안	실증적으로 거짓인, 조작된, 또는 오해를 불러일으키는 콘텐츠	고의적이고 의도적으로	-	개인, 사회집단, 조직 또는 국가를 의도적으로 속이거나 해를 끼치기 위해	
UNESCO 2018	거짓 정보	-	-	개인, 사회집단, 조직 또는 국가에 해를 끼치기 위해	

출처 : 이수엽 외(2024), p.167을 바탕으로 추가·보완하였음

이처럼 국제적으로 통용되는 정의를 종합하면, 허위조작정보는 단순한 오류나 부정확한 정보가 아니라 ‘의도적으로’ 제작·유포되어 공공에 해를 끼치는 허위 정보라는 점에서 다른 정보 유형과 구별된다. 문제는 특정 정보에 그러한 의도성이 존재하는지를 객관적으로 입증하는 것이 현실적으로 매우 어렵다는 점이다. 이러한 판단의 어려움은 허위조작정보를 국가가 직접 판별·규제하는 방식에 한계를 내포하고 있음을 보여준다(이숙종, 2024).

나. 허위조작정보의 확산과 신뢰 위기 심화

디지털 플랫폼을 중심으로 한 정보 유통 환경의 급격한 변화는 허위조작정보의 확산 양상을 근본적으로 바꾸어 놓고 있다. 과거에는 개별 언론사나 특정 행위자에 의해 제한적으로 발생하던 허위정보가, 오늘날에는 알고리즘 기반 추천·확산 구조와 결합되며 단시간 내 대규모로 유통되고, 사회적 의제 설정과 여론 형성 과정에 실질적인 영향을 미치는 구조적 위험 요인으로 작동하고 있다.

Redline의 「Key Statistics on Fake News & Misinformation in Media in 2024」 조사에 따르면, 전 세계 응답자의 약 60%는 뉴스가 의도적으로 허위 보도를 자주 한다고 인식하고 있는 것으로 나타났다(Redline, 2024). 특히 북미 지역의 경우, 미디어가 제공하는 뉴스와 정보를 신뢰한다고 응답한 비율은 65%에 그쳐, 정보 신뢰 기반이 상당히 취약한 상태임을 보여준다. 의도적인 거짓 기사를 자주 접한다고 응답한 비율은 국가별로 큰 편차를 보였는데, 아르헨티나(82%)가 가장 높았으며, 독일(30%), 일본(36%), 한국(39%)은 상대적으로 낮은 수준을 보였다. 그러나 27개국 평균으로는 응답자의 48%가 가짜뉴스를 실제 사실로 믿었다가 사후에 조작된 정보였음을 알게 된 경험이 있다고 응답해, 허위정보가 개인의 인식과 판단에 실질적인 영향을 미치고 있음을 시사한다.

아울러 전 세계 응답자의 60%는 자국 사회에서 정치·사회적 사안과 관련해 사람들이 ‘객관적 사실’보다 ‘자신이 믿고 싶은 정보’만을 선택적으로 수용하고 있다고 인식하고 있는 것으로 나타났다. 이러한 인식은 페루(71%), 세르비아(70%), 터키(69%), 미국(68%)에서 특히 높게 나타났으며, 이탈리아(48%), 일본(49%), 중국(49%) 등에서도 절반에 가까운 응답자가 같은 인식을 보였다. 흥미로운 점은 타인에 대해서는 ‘정보 버블 속에 살고 있다’고 평가한 비율이 65%에 달한 반면, 스스로가 그러한 정보 편향 상태에 있다고 인정

한 비율은 34%에 불과했다는 점이다. 또한 이민, 범죄율 등 주요 사회 이슈에 대해 자신이 평균적인 국민보다 더 잘 알고 있다고 믿는 응답자는 59%에 달했으며, 이러한 과신 경향은 터키(76%), 인도(75%), 영국(58%) 순으로 높게 나타났다.

온라인 상의 허위조작정보 증가는 언론에 대한 신뢰 저하뿐 아니라 언론 자유에 대한 위협으로도 인식되고 있다. 같은 조사에 따르면, 미국 언론인의 60%는 허위조작정보 확산으로 인해 언론의 자유가 제한될 가능성에 대해 높은 수준의 우려를 표명하였다. 또한 언론인의 94%는 조작된 뉴스와 정보가 미국 사회에서 중요한 문제라고 응답했으며, 이 중 71%는 '매우 큰 문제', 23%는 '상당한 문제'로 인식하고 있었다. 반면 문제가 크지 않거나 문제가 아니라고 응답한 비율은 6%에 불과했다(Redline, 2024).

이러한 조사 결과는 허위조작정보 문제가 단순한 정보 오류나 개인의 판단 착오를 넘어, 사회 전반의 신뢰 구조와 민주적 공론장의 기반을 약화시키는 구조적 위험 요인으로 작동하고 있음을 시사한다.

다. 생성형 AI로 인한 허위조작정보 증가

인공지능(AI) 기술의 급속한 발전은 사회 전반에 걸쳐 다양한 혁신과 변화를 가져오고 있다. 특히 최근 주목받고 있는 생성형 인공지능(Generative AI)은 텍스트, 이미지, 음성 등 다양한 형태의 콘텐츠를 자동으로 생성할 수 있는 기술로서, 창작 산업을 넘어 일상생활 전반으로 활용 범위를 빠르게 확장하고 있다.

그러나 생성형 AI의 확산은 새로운 사회적 문제와 윤리적 도전을 동반하고 있다는 우려도 함께 제기되고 있다. 그중에서도 가장 큰 쟁점은 허위조작정보의 생성과 유포에 악용될 가능성이다. 생성형 AI는 실제 뉴스와 구별하기 어려운 수준의 허위 기사, 이미지, 영상 등을 손쉽게 생산할 수 있으며, 이러한 콘텐츠가 대규모로 자동 생성·유통되고 특정 집단을 대상으로 정교하게 타겟팅될 경우 여론 조작과 사회적 혼란을 초래할 위험이 크다는 지적이 제기된다.

실제로 2016년 미국 대통령 선거 과정에서 소셜미디어를 통해 확산된 가짜뉴스가 여론 형성에 유의미한 영향을 미쳤다는 분석이 제시된 바 있으며, 이후 AI 기술의 고도화와 허위조작정보의 위협은 서로 비례적으로 증폭되고 있다는 인식이 확산되고 있다. 이러한 현상은 민주주의의 핵심인 공적 숙의와 합리적 의사결정 구조를 약화시키고, 정치 공동체의

판단 역량을 훼손할 수 있는 중대한 위험 요인으로 평가된다.

특히 SNS와 포털을 중심으로 한 디지털 미디어 환경에서는 허위조작정보가 단시간 내 급속히 확산될 가능성이 더욱 크다. 플랫폼의 비즈니스 모델이 클릭 수와 조회수에 기반하고 있는 구조적 특성상, 선정적이고 자극적인 허위조작정보가 더 큰 주목과 확산을 얻기 쉬운 환경이기 때문이다. 여기에 이용자들 역시 충분한 사실 확인 과정 없이 정보를 공유하는 경향을 보이면서, 생성형 AI로 인해 생산된 허위조작정보의 사회적 파급력은 더욱 증폭되고 있는 상황이다(Castro, D. & New, J. , 2016, p.3).

라. 허위조작정보 대응 새로운 정책 패러다임 필요성

생성형 AI의 확산과 디지털 플랫폼 중심 정보 유통 구조의 고도화는 허위조작정보 문제를 새로운 단계로 진입시키고 있으나 이러한 환경 변화에도 불구하고, 기존의 제도적 대응은 여전히 언론 중심의 사후 규제 또는 플랫폼의 자율규제에 머물러 있어, 오늘날 플랫폼이 지니는 영향력과 책임 구조를 충분히 반영하지 못하고 있다는 한계가 지적된다. 플랫폼 기업은 정보 유통의 핵심 관문으로서 알고리즘 설계, 추천・노출 방식, 광고 수익 구조를 통해 정보의 가시성과 확산 범위를 실질적으로 결정하는 주된 행위자임에도 불구하고, 허위조작정보로 인해 발생하는 사회적 비용에 대해서는 제도적으로 상응하는 책임을 충분히 부담하고 있지 않다는 평가가 반복되고 있다.

이러한 문제의식은 허위조작정보 대응을 더 이상 개별 주체의 선의나 자율, 또는 사후적 규제에만 맡길 수 없음을 시사한다. 디지털 환경에서 정보는 민주주의 사회를 지탱하는 핵심적인 공공 인프라로 기능하고 있으며, 그 신뢰성과 안정성이 훼손될 경우 사회 전반의 의사결정 구조와 공적 담론의 기반이 약화될 수밖에 없다. 이에 따라 최근에는 허위조작정보를 개별 현상이나 규제 대상으로만 다루는 것을 넘어, 정보 생산・유통・소비 전반의 신뢰성을 체계적으로 관리・보장하는 ‘정보 무결성(information integrity)’ 개념이 정책적 대안으로 주목받고 있다.

2. 정보 무결성의 의미

가. 무결성의 사전적 의미

정보 무결성의 의미를 논하기에 앞서 먼저 ‘무결성(無缺性, Integrity)’이라는 용어의 의미를 살펴볼 필요가 있다.

Oxford 영어사전에 따르면, 명사인 ‘integrity’는 ‘어떤 부분이나 요소가 제거되거나 결여되지 않은 상태, 훼손되거나 침해되지 않은 상태, 물질적 완전성, 완벽함, 전체성’을 의미하거나, ‘손상되지 않은 도덕적 상태, 순수함, 정직함, 성실함, 진실함’을 의미한다.[1] 즉, ‘integrity’는 물리적 의미로서 전체로서의 온전함이나 훼손되지 않은 상태를 의미하기도 하고, 규범적 의미로서 정직성, 성실성, 진실성 등을 의미하기도 한다.

한국어에서는 관행적으로 ‘integrity’를 무결성이라 번역하고 있는데, 국립국어원 표준국어대사전에는 ‘무결성’이라는 단어가 등재되어 있지 않고, 형용사인 ‘무결하다’와 명사인 ‘무결함’만이 ‘결함이나 흠이 없다’는 의미로 정의되어 있다[2]. 이는 ‘integrity’의 의미 중 전자의 의미인 ‘flawless(결함이 없음)’이나 ‘perfection(완전성)’에 보다 가까운 것으로, 정보 무결성을 개별 정보에 오류가 있어서는 안 된다거나 정보가 항상 사실이어야 한다는 기술적 개념으로 오인하게 만들 위험이 있다.

나. 정보통신기술 분야에서 정보 무결성 의미

실제 우리나라에서 무결성이라는 단어는 정보통신기술(ICT) 분야에서 활용되어 왔다. ICT 분야에서 무결성은 결함이 없는 상태 즉, 어떤 대상이 외부의 개입이나 변조없이 본래의 상태를 온전하게 유지하고 있는 상태를 의미한다. 다시 말해, 정보 무결성은 완전하고(complete), 일관되며(consistent), 정확하고(accurate), 믿을 수 있는(trustworthy) 데이터 또는 시스템 등을 가리키는 개념으로 사용되어 왔다.[3]

기밀성(Confidentiality), 무결성(Integrity), 가용성(Availability)은 각 앞글자를 따서 CIA Triad라 불리는데, 이 세 가지 요소는 정보 보안의 핵심 요소로서 상호 보완적인 관계를 가진다. 이 중 기밀성은 정보에 대해 무단 접근을 차단하는 것을 의미하며, 무결성은 데이터가 신뢰할 수 있고, 완전하며, 권한이 없는 사용자에 의해 임의로 변경되지 않았음을 의미한다. 또한 가용성은 필요한 때에 데이터에 접근할 수 있는 상태를 보장하는 것을 의미한다.[4] 이 중 무결성은 기술적인 진실성, 즉 형식적인 변조 방지가 핵심인 개념이다.

---

1) Oxford English Dictionary (2025.11.16. 검색)
2) 국립국어원 표준국어대사전 (2025.11.16. 검색)
3) 한국정보통신기술협회 정보통신용어사전 (2025.7.11. 검색)

그러나 반면, 최근 UN・OECD 등 국제기구가 논의하는 정보 무결성은 개별 정보가 ‘항상 사실이어야 한다’ 는 의미가 아니라, 시민이 다양한 정보원에 접근하면서도 정보의 신뢰성・균형성・완전성이 제도적으로 뒷받침되는 정보 생태계의 규범적 조건을 가리킨다. 따라서 국내에서 정보 무결성 개념을 도입・법제화하기 위해서는, 기존의 기술적 무결성(데이터 보안)과 구별되는 ‘거버넌스적 정보 무결성(신뢰 가능한 정보환경)’ 의 의미를 명확히 하는 작업이 선행되어야 한다.

다. 디지털 거버넌스에서의 정보 무결성 의미

정보 무결성의 개념은 2016년 미국 대선에서의 러시아발 허위조작정보 유포 사건과 케임브리지 애널리티카 사건[5]을 계기로 변화하기 시작하였다. 선거와 같은 민주적 절차가 정보 및 여론 조작에 의해 실질적으로 왜곡될 수 있다는 사실이 국제사회에 충격을 주었고, 인공지능 기술의 발전과 거대 온라인 플랫폼의 영향력 확대가 결합되면서 건강한 미디어 생태계를 조성하는 것이 핵심 정책 과제로 대두하였다.

변화한 의미의 정보 무결성 개념을 최초로 언급한 사람은 RAND Europe의 스테인 호렌스(Stijn Hoorens)이다. 2018년 9월 그는 “진실의 붕괴(Truth Decay)는 정보 무결성의 파

4) Confidentiality, Integrity, and Availability: The CIA Tria, Washington Univ in St.Louis, Homepage linked by https://informationsecurity.wustl.edu/guidance/confidentiality-integrity-and-availability-the-cia-triad/

5) 2016년 미국 대선 당시 케임브리지 애널리티카(Cambridge Analytica)는 Facebook 앱 ‘thisisyourdigitallife(GSRApp)’를 통해 25~27만 이용자의 개인정보와 친구 계정 5,000~6,500만명의 데이터를 무단 수집하여 미국 유권자의 성향 분석 및 정치 마이크로타겟팅에 사용하였다. 미 FTC는 이를 FTC Act §5(15 U.S.C. §45)의 기만적 행위라 판단하고 수집된 데이터와 그로부터 도출된 모든 정보 또는 산출물(알고리즘 등 포함)를 수집 또는 파기하고, 동 데이터를 공개, 사용, 판매하거나 그로부터 어떠한 이익을 얻는 것을 영구적으로 금지하였다(Federal Trade Commission, Opinion of the Commission in the Matter of Cambridge Analytica, LLC, Docket No. 9383, 2019). 또한 이 사건으로 Facebook에도 FTC Act §5 위반을 근거로 50억 달러의 벌금이 부과되었다(U.S. Department of Justice and Federal Trade Commission, Complaint for Civil Penalties, Injunction, and Other Relief, Facebook, Inc., No. 1:19-cv-02184. U.S. District Court for the District of Columbia, July 24, 2019., U.S. Department of Justice and Federal Trade Commission, Stipulated Order for Civil Penalty, Monetary Judgment, and Injunctive Relief, Facebook, Inc., No. 1:19-cv-2184, U.S. District Court for the District of Columbia, July 24, 2019).

생적 결과이다.” 라고 밝히면서, 정보 무결성의 약화가 진실성 위기의 구조적 원인임을 강조하였다(Club de MADRED, 2018, p.3).

같은 해 11월 라트비아 리가(Riga)에서 개최된 세계지도자 연합 마드리드 클럽 글로벌 거버넌스 원탁회의에서는 정보 무결성이 허위정보 규제의 대안을 넘어 민주적 공론장의 지속 가능성을 담보하기 위한 새로운 규범적 틀로 제시되었다. 동 회의에서는 기존의 논의가 ‘가짜뉴스’ 라는 용어의 정교화에 집중되었을 뿐, 온라인 플랫폼과 소셜미디어 환경에서 시민들에게 도달하는 정보의 흐름 전반에 영향을 미치는 보다 광범위한 구조적 현상을 포괄하는 개념을 마련하는 데에는 소홀했다고 지적하면서, 이러한 현상을 포괄하는 용어로 ‘정보 무결성’ 을 채택하였다고 발표하였다. 동 회의에서는 정보 무결성을 “시민들이 시사적 사안, 정부 조치, 정치 행위자 및 정치적 인식과 의사 결정에 관련된 기타 요소들에 대해 접근할 수 있는 정보의 신뢰성(trustworthiness), 균형성(balance), 완전성(completeness)” 이라 정의하였다(Club de MADRED, 2018, p.4).

이후 2022년 2월 UN은 정보 무결성 개념을 재정의하여 발표하였다. UN은 본래 정보 무결성 개념은 기업 시스템 안에서의 정보보안과 데이터 보호를 의미하였으나, 그 개념을 광의로 확장할 경우 “건강한 정보 생태계를 유지하기 위한 정보의 내용, 절차, 시스템의 정확성(accuracy), 일관성(consistency), 신뢰성(reliability)” 이라고 설명하면서, 리가의 원탁회의에서 제시된 논의를 이어받아, 시민이 접근하는 정보의 신뢰성, 균형성, 완전성을 확보하는 것이 정보 무결성의 핵심이라고 발표하였다(UNDP, 2022, p.4).

한편, OECD는 정보 무결성을 “정확하고 신뢰할 수 있으며 증거에 기반한 다원적 정보원에 대한 접근을 보장함으로써, 개인이 다양한 사상에 노출되어 정보에 입각한 선택을 하고 권리를 적정하게 행사할 수 있도록 하는 정보 환경의 결과물” 이라 정의하였는데,[6] 이러한 정의는 개인의 권리보다는 정보 환경에 보다 초점이 맞춰진 것으로 ‘다원적 정보 환경’ 을 강조하고 있음을 알 수 있다. OECD는 정보 무결성이 콘텐츠가 아닌 맥락에 초점(focusing on context not content)을 두어야 함을 강조한다(Charles Baubion, 2025, p.15).

즉, 정보 무결성은 개별 정보의 사실 여부를 판별하는 협의의 개념이 아니라, 시민이 공적 판단을 형성하는 과정에서 접근하는 정보가 신뢰성, 균형성, 완전성을 갖추고, 증거에 기반한 다원적 정보원에 대한 실질적 접근이 가능하도록 보장된 민주적 정보 환경을 의미

6) OECD(2025). Recommendation of the Council on Information Integrity. p.7

한다. 이는 국가가 개별 콘텐츠의 사실 여부를 판단하고 차단하는 방식이 아니라, 정보의 생산・유통・추천・검증 전반의 절차와 환경을 개선함으로써 시민이 신뢰 가능한 정보에 접근할 수 있도록 하는 규범적 개념으로서, 기존 ICT 분야에서의 정보 무결성 개념과 구별된다. 이러한 개념 전환은 허위조작정보 대응의 초점이 국가가 사후적으로 판별・차단하는 방식에서 벗어나, 민주적 정보 생태계의 구조적・과정적 건전성을 제도적으로 보장하는 방향으로 이동하고 있음을 보여준다.

3. 정보 무결성 논의 배경과 플랫폼의 책임

가. 플랫폼의 구조적 권력과 정보 무결성의 위기

앞서 살펴본 정보 무결성 개념은 플랫폼이 정보 유통의 중심적 매개자로 자리 잡은 디지털 환경과 긴밀하게 연결된다. 플랫폼은 정보 확산의 '필수적 중개자'로서 정보를 직접 생산하지는 않지만 어떤 정보를 누구에게 보여줄 것인지를 결정하는 권력을 지닌다(Gillespie, Tarleton, 2017). 플랫폼은 생산자와 이용자가 결합되는 양면시장 구조에서 발생하는 방대한 데이터를 기반으로 이용자의 개인정보와 행태정보를 분석하여 알고리즘을 통해 콘텐츠를 조정(curation)하고 매칭한다. 이러한 개인화 알고리즘은 공급자와 이용자 모두의 만족도를 높여 플랫폼 이용을 촉진하고 결과적으로 플랫폼의 시장지배력과 정보 독점 구조를 공고히 한다(이명호, 2021, p.23).

오늘날 공론장에서 시민들이 접하는 대부분의 정보는 플랫폼의 알고리즘에 의해 선별・배치되기 때문에 그 과정에서 정보 무결성이 훼손될 위험이 존재한다. 이용자들은 자각하지 못한 채 자신들의 성향과 행동 데이터를 기반으로 작동하는 다양한 개인화 알고리즘에 둘러싸인 환경에서 의사결정을 하게 된다. 이러한 알고리즘은 정보 접근 비용을 낮추고, 편의를 제공하는 장점이 있는 반면, 특정 정보에 반복적으로 노출되도록 하여 정보 편향을 심화시키는 단점이 있다. 이를 필터 버블(filter bubble)이라 한다.

필터 버블이 심화될 경우 이용자는 기존 신념을 강화하는 정보에만 접근하게 되고, 이러한 환경에 허위정보가 유입되면 정보 생태계 전체가 오염될 가능성이 급격히 증가한다. 2018년 MIT 연구팀에서 발표한 바에 따르면, 거짓 뉴스는 진짜 뉴스보다 더 멀리, 더 빠르게, 더 깊고, 더 광범위하게 확산되며, 그 확산의 주요 요인이 자동화된 봇(bot)이 아닌 이용자의 선택과 행동 패턴이라는 점이 확인되었다(Vosoughi, Soroush, et al., 2018).

결국, 정보 무결성의 위기는 개별 정보의 사실성 여부보다 플랫폼 기반 디지털 생태계의 구조적인 작동 방식, 예를 들어, 알고리즘 큐레이션의 불투명성, 정보 확산 메커니즘의 비대칭성, 이용자 행동 패턴의 편향성 등에서 비롯되는 문제로 이해될 필요가 있다.

이러한 이유로 플랫폼에서의 알고리즘 기반 콘텐츠 관리 시스템은 표현의 자유, 개인정보자기결정권, 정치적 기본권 등 헌법적 권리를 침해할 가능성이 있다. 실제로 플랫폼은 특정 콘텐츠의 노출 순위를 낮추거나 보이지 않게 하거나, 특정 이용자의 계정을 정지시키는 등 사실상 국가의 검열과 유사한 수준의 영향력을 행사하고 있다(문의빈, 2025, p.376). 알고리즘이 특정 콘텐츠의 가시성을 제한하거나, 이용자 성향에 따라 노출되는 정보를 자의적으로 조정하거나, 사적 검열과 같은 형태의 관리가 별다른 통제없이 이루어질 경우 민주적 의사형성에 심각한 장애가 생길 수 있다(홍남희, 2018. p.135).

나. 정보 무결성과 투명성 원칙

위와 같이 정보 무결성의 위기가 플랫폼 기반 디지털 생태계의 구조적인 작용 방식에서 비롯된다는 점을 고려할 때 정보 무결성은 플랫폼 및 인공지능 기술의 투명성을 전제로 하지 않고서는 확보되기 어렵다. 인공지능 기반 의사결정의 책임 여부를 판단하거나 관련 이의제기를 가능하게 하려면 우선적으로 인공지능의 결정을 이해할 수 있어야 하기 때문에 투명성 확보가 필수적이다.

인공지능 불투명성은 오류, 차별, 편향과 같은 문제의 발생과 그 원인을 불명하게 할 가능성이 있다. 특히 공적 영역에서 인공지능을 활용할 때 민주주의와 법치주의라는 공법의 핵심 원칙을 유지하기 위해서는 인공지능에 의해 이루어진 판단과 의사결정 과정을 이해하고 검증할 수 있도록 투명성을 확보하는 것이 강조된다(김민주・김현경, 2025. p.66-67).

특히 허위정보와 관련해서는, 알고리즘이 어떠한 기준에 따라 정보를 선택・확산시키는지, 위험 점수나 조기 탐지 기술이 어떤 데이터를 기반으로 구축되는지, 사실확인(fact-checking)이나 노출 순위 하향(down-ranking) 메커니즘이 어떻게 적용되는지 정보 흐름을 파악할 수 있도록 공개되는 것이 요구된다. 실제로 틱톡은 이용자 ‘체류 시간’과 ‘반응’에 따라 비슷한 콘텐츠를 연쇄적으로 추천하고, 페이스북은 이용자 정치 성향에 따라 특정 게시물의 노출 순서를 조정하는 것으로 알려져 있다(이동우, 2025).

다만, 기술 환경이 급격히 변화하고 정보량이 폭증하는 상황에서 투명성이 확보되어야

할 정보의 범위를 명확히 설정하기 어렵다는 문제도 고려하여야 한다. 알고리즘은 영업비밀과 이용자 프라이버시와 밀접하게 연계되어 있어 공개에 한계가 있고, 설령 일부 정보가 공개되더라도 알고리즘의 작동 원리나 프로그래밍 구조는 일반 이용자가 이해하기 어렵기 때문에 인공지능이 미치는 사회적 영향과 위험을 정확히 파악하기 어렵다는 우려가 존재한다(김민주・김현경, 2025. p.73).

결국 투명성은 정보 무결성 보장의 필요조건이지만 그것만으로는 충분하지 않으며, 정보 무결성을 실질적으로 확보하기 위해서는 다각적이고 종합적인 거버넌스 체계를 마련할 필요가 있다. 이하에서는 국내외에서 정보 무결성 제도를 어떻게 구성하고 있는지 살펴보고자 한다.

제2절 해외 정보 무결성 관련 법제도 및 정책동향 분석

1. UN

2024년 UN은 「정보 무결성 글로벌 원칙」에서 ① 사회적 신뢰와 회복탄력성, ② 건전한 인센티브, ③ 공공의 역량 강화, ④ 독립적이고 자유로우며 다원적인 미디어, ⑤ 투명성과 연구를 정보 무결성의 총 5가지 핵심 원칙으로 제시하면서, 모든 이해관계자(정부・기업・언론・시민사회・연구자・광고주 등)가 정보 무결성을 보호・강화하기 위해 구체적 행동을 취할 것을 요구하고 있다.

가. 사회적 신뢰와 회복탄력성

UN은 정보 무결성의 핵심 요인으로 가장 먼저 ‘사회적 신뢰와 회복탄력성’을 제시한다. 전자는 개인이 정보의 출처와 신뢰도에 대해 가지는 확신을, 후자는 정보 생태계에서 발생하는 혼란과 조작에 사회가 대응하고 극복할 수 있는 능력을 의미한다. 그러나 국가 및 비국가 행위자가 전략적・정치적・재정적 목적을 위해 정보 환경을 악용할 경우 이 두 요소는 쉽게 약화된다.

특히 AI가 생성・매개하는 콘텐츠는 실제 또는 원본처럼 보일 수 있고, 확산 속도가 매우 높아 신뢰 격차를 빠르게 확대시키며, 여성, 아동, 노인, 장애인, 소수자 등 취약계층이

더 큰 피해를 입는 것으로 지적된다. 이에 UN은 정보 무결성 확보를 위해 견고하고 미래 지향적인 디지털 신뢰·안전 체계가 필요하며, 시민이 다양한 정보원에 접근하고 사회적 포용과 경제적 안정, 정치적 역량을 갖출 때 위험을 사전 감지하고 대응하는 회복탄력성이 강화된다고 강조한다.

나. 건전한 인센티브

그 다음으로, '건전한 인센티브'는 플랫폼이 타겟 광고와 각종 콘텐츠 수익화 모델에 구조적으로 의존하는 방식이 정보 무결성에 미치는 영향에 초점을 둔다. 이러한 비즈니스 모델은 이용자 참여도(engagement)를 우선시하는 알고리즘을 플랫폼 설계의 중심에 두고 있는데, 크리에이터 경제의 활성화와 같은 긍정적 효과를 가져온 반면, 허위정보·혐오표현 유포에 재정적 인센티브를 주어 이를 확산시키는 구조적 요인으로 작동해 왔다. 특히 분열적이거나 강한 감정을 유발하는 콘텐츠일수록 더 높은 참여도를 유도하는 특성 때문에, 알고리즘이 유해 콘텐츠에 대한 보상과 증폭을 반복적으로 강화하는 현상이 발생한다.

기술 산업은 디지털 광고 과정을 복잡하고 투명하게 설계해 왔으며, 인간의 개입을 최소화하고 있다. 불투명하게 설계된 광고 시스템은 광고주가 의도하지 않았음에도 불구하고, 광고 예산이 특정 개인 및 단체, 사상에 우회적으로 지원되는 결과를 초래할 수 있다. UN은 이러한 문제를 해결하기 위해, 행동 추적 및 개인정보 기반의 표적 광고(programmatic advertising)에 과도하게 의존하지 않는 인권 기반 비즈니스 모델로의 전환을 강조한다. 플랫폼 기업들이 광고주에게 광고 배치 과정에 대한 투명성을 높이고, 광고 전달 주체들이 인권에 부합하는 책임 있는 광고 정책을 준수하도록 함으로써 더 건강한 인센티브 구조를 만들 수 있으며, 광고주는 투명한 공급망에 대한 통제력을 높임으로써 투자 대비 더 나은 효과를 얻을 수 있다고 제안한다.

다. 공공의 역량 강화

세 번째 '공공의 역량 강화'는 시민이 자신의 온라인 경험을 통제하고, 소비할 미디어에 대해 정보에 기반한 결정을 내리며 자유롭게 자신을 표현하는 것을 요구한다. 공공의 역량 강화는 다양하고 신뢰할 수 있는 정보원에 대한 지속적인 접근을 요건으로 한다. 이를 위해서 기술 기업들은 이용자의 프라이버시권을 인정하고, 정보 신뢰성과 안전, 프라

이버시 정책, 데이터 활용 전반에 대해 이용자가 의견을 제시하고 피드백을 제공할 수 있도록 보장하여야 한다. 또한 미디어・정보・디지털 리터러시 역량 강화가 이루어져야 하며, 특히 여성, 노인, 아동・청소년, 장애인, 취약하거나 주변화된 집단이 직면하는 특수한 문제들에 대해 특별한 주의를 기울일 것을 강조한다.

라. 독립적이고 자유로우며 다원적인 미디어

네 번째로 UN은 정보 무결성이 '독립적이고 자유로우며 다원적인 미디어'가 있을 때에만 실현될 수 있다고 본다. 언론의 자유는 법치주의를 떠받치는 기반이자 민주주의의 핵심 요소로서, 기자와 언론 종사자가 안전하고 개방적인 환경에서 일관되게 취재하고 보도할 자유를 보장받으며, 모든 개인이 다원적이고 신뢰할 수 있는 뉴스에 안정적으로 접근할 수 있어야 한다고 명시하였다. UN은 특히 언론의 다양성이 위협받고, 지역 및 공익 저널리즘이 약화되는 상황 속에서 편집 기준이 엄격히 유지되지 않을 경우 언론사는 정보 무결성에 위험을 조장하거나 확대시킬 수 있으며, 이는 온・오프라인 공간을 넘나들며 확산될 수 있음을 우려한다.

마. 투명성과 연구

마지막은 '투명성과 연구'이다. 투명성은 정보가 어떻게 확산되는지, 개인 데이터가 어떤 방식으로 활용되는지, 정보 무결성 관련 위험이 어떻게 관리되는지를 이해하게 한다. 현재 주요 글로벌 기술 기업은 소수 국가에 집중되어 있기 때문에 해당 국가의 투명성 규범이 전 세계 정보 환경에 과도한 영향력을 미치고 있다. 이러한 구조는 공익적 연구를 제약하고, 취약 지역이나 연구가 충분히 이루어지지 못한 공동체의 필요를 충분히 반영하지 못하는 결과를 초래한다. 특히 극소수의 기술기업과 일부 미디어 기업이 방대한 양의 데이터와 정보 흐름을 통제하고 있으며, 이들 기업이 국가・정치・경제적 이해관계자와 밀접한 관계를 맺고 있어 감시와 연구에 대한 접근 장벽을 더욱 공고히 한다. 이에 따라 UN은 세계적 차원의 연구 생태계 구축을 위해 데이터의 가용성・품질・활용성을 확대하고, 프라이버시 보호를 전제로 연구자에게 의미있는 데이터 접근권을 보장할 것을 제안하였다.

정리하면, UN은 소수 플랫폼 기업의 데이터 및 정보 독점과 알고리즘 중심의 비즈니스

모델, AI 생성 콘텐츠의 급속한 확산, 취약계층에 대한 정보 불균형 등이 정보 무결성 침해의 원인이며, 이를 해결하기 위해서는 '허위정보의 내용 규제'를 넘어 정보 생태계의 질을 보장하는 것으로 국가의 책무가 확대되어야 함을 제안한다. 또한 플랫폼은 사적 경제주체임에도, 디지털 생태계에서 정보 유통 구조를 실질적으로 통제하는 공적 기능을 수행하고 있기 때문에 투명성, 책임성, 안전성 보장과 같은 헌법적 가치 형성을 위해 노력하여야 한다.

2. OECD 권고안의 주요 내용

UN이 정보 무결성의 핵심 원칙과 방향성을 제시하였다면, OECD는 「정보 무결성 권고안」을 통해 보다 구체적인 이행방안을 발표하였다.[7] OECD는 정보 무결성의 3대 축으로 ① 사회적 회복력 강화, ② 투명성・책임성・다원성 제고, ③ 제도적 구조 및 열린 정부 강화 등을 제시하였다. OECD는 허위정보를 단일 문제로 보지 않고, 교육(예방), 플랫폼 및 언론 협력, 정부 책임성・투명성, 국제 공조라는 다층적 접근방안을 제시하였다.

가. 사회적 회복력 강화

먼저, '사회적 회복력 강화'와 관련해서, OECD는 정보 무결성 보장을 위해서는 개별적으로 허위정보를 삭제하는 것이 아닌, 사회 전체가 왜곡된 정보 환경에 대응할 수 있는 구조적 회복력을 갖추는 것이 핵심이라고 보고, 이를 위해 다음의 사항을 이행할 것을 권고하였다.

첫째, 학교・언론・시민사회・플랫폼과 협력하여 전 생애주기 미디어・정보・AI 리터러시 교육을 강화하고, 정보 출처 평가 및 AI 생성 콘텐츠 위험 인식 능력을 제고한다. 둘째, 악성 행위자 및 조작 사례에 대한 위협 평가(threat assessment) 정보를 공유하여 사회적 이해를 높이고, 취약계층을 대상으로 한 맞춤형 보호 및 리터러시 프로그램의 효과 평가 연구를 촉진한다. 셋째, AI가 생성한 정보 및 온라인 플랫폼에서 공유되는 기타 콘텐츠의 진위(authenticity)와 출처(origins)를 평가할 수 있는 개인의 역량을 향상시키기 위해, 라벨

7) OECD(2025), Recommendation of the Council on Information Integrity

링・해싱・워터마킹 등 콘텐츠 프로비넌스(content provenance) 기술을 활용하고, 비진정적 행위(inauthentic behaviour)를 탐지하기 위한 과학적 연구 및 기술 도구의 개발을 지원한다.[8][9] 넷째, 정보 무결성과 관련된 정책 과정에서 민주적 참여 메커니즘을 마련하고, 시민들이 악의적인 온라인 활동을 신고할 수 있는 시스템을 구축한다. 다섯째, 플랫폼을 포함한 기업이 조직적인 비진정적 행위를 통해 수익을 창출하는 문제를 성찰하고 이를 억제하기 위한 방안을 마련하도록 유도한다. 여섯째, 정보 흐름에 대한 이해를 높이기 위해 플랫폼이 정부・팩트체커・학계 등과의 정보 공유 체계를 구축하고, 공・사적 데이터에 대한 접근성을 확대한다. 일곱째, 광고 투명성 강화를 위해 삭제된 광고까지 포함한 광고 데이터베이스 구축을 검토한다. 여덟째, 알고리즘 설계・추천 시스템・콘텐츠 조정 절차 등을 공개해 플랫폼의 작동 방식에 대한 사회적 이해와 감시를 강화하고, 이용자 불복 절차를 보장하도록 한다. 아홉째, 외국 정보조작(FIMI)에 대한 국가 차원의 위험 평가, 제재, 국제협력 등을 확대한다. 마지막으로, 허위정보의 영향과 정책 대응 효과를 정기적으로 평가하는 체계를 구축해 정보 무결성 정책의 실효성과 책임성을 담보한다.

나. 정보원의 투명성・책임성・다원성 강화

다음으로, OECD는 ‘정보원의 투명성・책임성・다원성’을 제목으로 온라인 플랫폼과

8) 콘텐츠 프로비넌스는 콘텐츠가 생성, 수정, 유통된 이력과 그 소유 및 변형 내역을 투명하게 증명하는 개념으로 최근에는 콘텐츠 출처 및 진위성 연합(Coalition for Content Provenance and Authenticity, C2PA)과 콘텐츠 진위성 이니셔티브(Content Authenticity Initiative, CAI)가 표준화된 메타데이터를 통해 콘텐츠의 출처와 변형 내역을 투명하게 관리하고 있으며, 디지털 정보의 신뢰성과 투명성을 높이는 핵심 기술로 주목받고 있다. 보다 구체적인 과정에 대해서는 Lara Monday·Lorenzo Strappelli(2024) 참고.

9) 허위정보의 탐지 및 확산 방지를 위해 다음과 같은 네 가지 기술적 방식이 사용되고 있다. ① 정보의 내용 기반 탐지 기술로 일종의 사실로 구성된 사전을 만든 후 해당 내용과 얼마나 일치하는지 여부를 통해 해당 내용의 허위정보 여부를 판단함. ② 정보의 출처 기반 탐지 기술로 해당 정보의 출처가 어디인지를 분석하여 그 출처의 신뢰성을 바탕으로 해당 내용의 허위정보 여부를 판단함. ③ 정보의 확산 방식 기반 탐지 기술로 해당 정보가 소셜 미디어 등 사회적 네트워크 상에서 어떻게 전파되고 있는지를 분석한 후 이전의 허위정보 전파 패턴과 비교해 해당 내용의 허위정보 여부를 판단함, ④ 정보에 대한 자연어 처리 기 기반 탐지 기술로 해당 정보에 대한 형태소 분석 등 자연어 처리 기반 콘텐츠 분석을 실시해 과거 거짓 정보, 왜곡된 사실, 비논리적 주장 등과의 유사성 비교를 통해 해당 내용의 허위정보 여부를 판단함. 언어의 문맥적 의미를 분석하여 허위정보의 전형적인 언어적 특성을 찾아내는 방식임(이와 관련하여 상세한 내용은 오세욱, 2017, pp.137-180)

언론, 창작자 생태계 전반이 정보 무결성을 보장하기 위해 갖추어야 할 구조적 원칙과 기준을 다음과 같이 제시하였다.

첫째, 언론 시장의 경쟁 촉진, 소유구조의 다양성 제고, 편집권 독립 보장 등을 통해 독립적・다원적 저널리즘 생태계를 강화하고, 특히 지역・소수언어・탐사보도 등 공익 저널리즘의 지속 가능성을 위한 재정적・제도적 지원을 마련한다. 둘째, 기자・언론 종사자・팩트체커가 물리적・심리적・디지털 환경에서의 위협 없이 취재・검증 활동을 수행할 수 있도록 안전한 환경을 보장하고, 전략적 봉쇄 소송 및 내부고발자 보호 체계를 강화한다. 셋째, 온라인 플랫폼에서 저널리즘 콘텐츠가 공정하게 보상될 수 있도록 대가 지급 모델의 도입 가능성을 검토하고, 미디어 소유자의 광고・협찬・홍보 콘텐츠의 투명성을 제고하여 보도의 독립성을 강화한다. 넷째, 언론사의 편집권을 보호하기 위한 윤리・무결성 기준을 마련하도록 장려하고, 언론인이 허위정보・조직적인 비진정적 행위의 출처・기법을 파악해 대응할 수 있도록 역량 강화를 지원한다. 다섯째, 온라인 정보 플랫폼의 시장 집중으로 인한 정보 무결성 침해를 방지하기 위해 경쟁 촉진과 공정한 플랫폼 생태계 조성을 도모하고, 디지털 콘텐츠 창작자・인플루언서・자동 뉴스 생성 플랫폼 등에 대한 무결성 기준과 광고 투명성 규범을 강화한다. 여섯째, 플랫폼이 인권을 존중하는 기업 관행을 채택하도록 하고, 이용자 신고 시스템을 구축하며, 개인화된 알고리즘 통제권과 제3자의 대안적 정보 필터 도구인 미들웨어(middle-ware)[10]의 활용을 가능하게 함으로써 이용자의 자율성과 플랫폼 책임성을 강화한다. 일곱째, 콘텐츠 조정 체계와 인력을 충분히 확보하고, 특히 플랫폼은 독립적 팩트체커 및 신뢰할 수 있는 신고자(trusted flaggers)와 협력할 수 있는 전담팀을 충분한 규모로 유지하고 선거와 같은 위험 상황에 신속하게 대응할 수 있어야 한다. 필요할 경우 콘텐츠 삭제, 노출 순위 하향(downranking), 지연(latency)・마찰(friction) 부과, 재공유 제한, 브리징 기반 랭킹(bridging-based ranking)[11], 정보 무결성 지

10) 미들웨어는 지배적 온라인 플랫폼의 기존 시스템 위에 제3자가 제공하는 소프트웨어 계층을 의미한다. 이 소프트웨어는 플랫폼과 이용자 사이에 편집 및 선정 기능을 수행하는 중간 계층(intermediary layer)으로 작동한다. 현재 플랫폼은 정보를 보여줄 것인지 말 것인지를 사실상 독점적으로 결정함으로써 민주주의에 대한 잠재적 위험을 가지고 있다. 미들웨어는 이 문제를 해결하기 위해 플랫폼의 정보 배치를 외부 경쟁업체가 수행하도록 개방하는 방식을 제안함으로써 특정 플랫폼 기업의 알고리즘이 민주적 담론을 독점하는 구조를 완화시킨다. 보다 상세한 내용은 Fukuyama, Francis, et al.(2020) 참고.

11) 브리징 기반 랭킹은 참여도에 기반한 추천 시스템의 반대 개념으로, 논쟁적인 주제일지라도 서로 다른 집단 간에 긍정적인 상호작용을 촉진하는 콘텐츠에 보상을 부여하는 알고리즘을 말한다(Ovadya, Aviv, and Luke Thorburn, 2022, p.14).

표에 기반한 우선순위 부여, 맥락(context) 제공 등 다양한 조치를 신속히 적용할 수 있는 운영 역량을 확보해야 한다. 여덟째, 플랫폼이 국가별 대표자·법적 책임자를 지정하고, 조기경보·위기 대응 체계를 구축하며, 위기 상황에서 공익 정보를 신속하게 제공하도록 한다. 아홉째, 플랫폼이 기술매개 성폭력·이미지 기반 학대·딥페이크 포르노그라피·투표권 행사를 방해하기 위한 허위정보 등 위해적 콘텐츠의 생성·유포를 탐지·대응하고, 조직적인 비진정적 행위와 봇의 악용을 차단하며, AI 생성 콘텐츠의 출처·진위를 명확히 표시하는 국제 표준을 도입하도록 한다. 마지막으로, 플랫폼이 정보 무결성 관련 위험평가 및 독립 감사 보고서를 공개함으로써 사회적 감시를 가능하게 하고, 아동의 프라이버시·안전 보호, 아동 대상 광고 제한, 연령 확인 체계 구축 등을 통해 아동의 권리와 민주적 역량 형성을 보장하는 환경을 마련하도록 한다.

다. 제도적 구조 개선 및 개방형 정부

마지막으로 OECD는 '제도적 구조 개선 및 개방형 정부'라는 제목 하에서 투명성 및 공적 보고 의무를 갖추고, 공공 커뮤니케이션과 열린 정보를 강화할 것을 요구하고 있다.

OECD는 정보 무결성 보장을 위해 국가의 역할을 허위정보에 대한 직접적인 규제가 아니라 제도적·구조적 환경을 설계하는 것으로 재정의한다. 첫째, 정부는 플랫폼에 대한 추적·차단·삭제 요청 등 콘텐츠 개입의 전 과정에 대해 투명성과 공적 보고 의무를 갖추어야 한다. 둘째, 공직자의 플랫폼 개입, AI 생성 콘텐츠의 사용·출처 명시, 시민사회·플랫폼과의 상호작용 기준 등 절차적 책임성과 예측 가능성을 높이는 지침을 마련한다. 셋째, 정보 무결성 전략을 국가 디지털·민주주의·교육 정책과 연계한 전략적 프레임워크로 수립하고, 정책목표와 평가·감시체계를 명확히 하도록 한다. 넷째, 기관 간 조정·정보 공유·위기 대응 체계 마련, 독립 규제기구 설립 등을 통해 명확한 권한과 견제 장치를 갖춘 제도적 거버넌스를 구축한다. 다섯째, 선거 절차를 보호하기 위해 정보접근성 확대, 선거 관련 사이버보안, 투표권 방해를 위한 허위정보 금지, 딥페이크 대응 등 선거 무결성 보호 조치를 시행한다. 여섯째, 국제협력, 공익미디어 지원, 콘텐츠 프로비넌스 기술 공유 등 국제적 정보 무결성 생태계를 강화한다. 일곱째, 공직자 맞춤형 교육 및 역량 강화 프로그램을 제공하고, 정보 환경이 민주적 참여에 어떠한 영향을 미치는지 예측하기 위한 전략적 전망, 규제영향평가 등 공공 관리 도구를 체계적으로 적용하며, 정부 차원에서 인

위적 증폭 활동 및 합성 콘텐츠(synthetic content)를 탐지할 수 있는 시스템을 개발하여 허위정보에 대한 정부의 대응 역량을 강화한다. 마지막으로 열린 정부 구현을 위해 공공 커뮤니케이션 기능 강화[12], 정부 콘텐츠 진위 확인, 공공정보 접근성 제고, 정보공개법 및 오픈데이터 기준 갱신 등 정부 운영의 개방성과 책임성을 향상시키는 조치를 권고한다.

라. 정보 무결성 강화를 위한 주체별 역할

이러한 권고안을 주체별로 정리하면, 국가(정부)와 플랫폼, 시민사회 및 언론・학계는 정보 무결성 보장을 위해 각각 다음과 같은 역할을 수행하여야 한다.

먼저 국가는 정보 무결성이 실질적으로 보장될 수 있는 제도적・구조적 환경을 조성할 책무를 지닌다. 이를 위해 리터러시 교육・취약계층 보호・플랫폼 규제의 투명화・선거 절차의 무결성 확보 등 제도적 기반을 마련하고, 허위정보에 대한 공직자의 역량 강화와 열린 정부 구현 등을 통해 정보 생태계 전반의 신뢰성을 제고하여야 한다.

또한 플랫폼은 정보 유통의 핵심 매개자로서 알고리즘, 광고, 콘텐츠 조정 절차 등에 대해 투명성과 책임성을 강화하여야 한다. 조직적인 비진정적 행위와 AI 생성 콘텐츠로 인한 위험을 방지・완화하고, 이용자에게 알고리즘 통제권과 참여권을 보장함으로써 안전한 정보 이용 환경을 조성하여야 한다. 특히 정부・언론・학계와의 협력을 통해 데이터를 공유하고 위기 대응 역량을 체계적으로 강화하여야 한다.

마지막으로 시민사회와 언론・학계는 독립적 팩트체킹, 언론 감시, 과학적 연구를 수행함으로써 정보 생태계의 감시자 역할을 수행하여야 한다. 언론은 편집권 독립, 윤리 기준 준수, 투명성 확보 등을 통해 미디어의 다원성과 공익성을 유지하여야 하며, 시민사회는 신고・참여 메커니즘을 통해 정보 무결성 정책 형성 과정에 실질적으로 참여하여야 한다.

3. 시사점

UN과 OECD 등 국제기구들은 허위정보 대응에 있어 국가의 역할을 허위정보 규제자가 아닌 건전한 정보 환경의 설계자로 놓고, 정보 무결성 침해 문제를 개별 표현의 진위

12) 구체적으로는 공중에게 이해 가능하고, 접근 가능하며, 관련성이 있고, 시의적절하며, 신뢰할 수 있는(투명하고, 정확하며, 포괄적인) 정보를 제공하기 위한 공공 커뮤니케이션 기능의 역할을 명확히 하고 그 역량을 구축하는 것을 의미한다(OECD, 2025, p.16).

여부를 판별하고 차단하는 방식으로 해결하기 보다는 정보의 신뢰성, 균형성, 완전성을 확보하는 방향으로 국가의 책무를 재설정하고 있다. 이는 국가가 특정 정보를 ‘허위’로 규정하고 규제함으로써 발생할 수 있는 검열 및 권력 남용 논란을 피하고, 표현의 자유가 실질적으로 보장될 수 있는 건전한 공론장을 형성하는데 중점을 둔다는 점에서 주목할 만하다.

또한 UN과 OECD는 공통적으로 AI로 인한 정보 무결성 침해를 기술적·구조적 수단을 통해 해결하는 것을 제안한다. AI 생성 콘텐츠에 대한 직접적인 규제 대신에 라벨링, 워터마킹, 기타 콘텐츠 프로비넌스와 같은 신뢰성 확인 기술을 통해 이용자들의 인지 능력을 높이고, 합성 콘텐츠 및 인위적 증폭 행위를 탐지하여 제어하는 등 유해성 관리 기술을 정보 인프라에 내재화하는 전략을 제안한다. 이는 AI를 민주적 공론장에 대한 잠재적 위협으로 보지 않고, AI 기술 혁신을 장려하면서도 표준 마련 및 감시 체계 구성을 통해 정보 무결성을 보장하려는 균형적인 정책 방향을 보여준다.

마지막으로 국제기구는 이러한 과정 전반을 대중에게 투명하게 공개하고, 국제적으로 협력할 것을 강조한다. 즉, 알고리즘 설계와 추천 시스템의 원리 및 작동 방식, 콘텐츠 조정 기준과 절차, 위험평가와 독립 감사 결과, 데이터 접근 및 활용 방식 등을 투명하게 공개함으로써, 국가와 플랫폼, 시민사회가 함께 공유 가능한 정보와 데이터에 기반하여 감시 및 비판·참여가 가능하도록 거버넌스를 설계할 것을 요구하고 있다.

국제기구의 권고는 정보 무결성 확보를 위한 정책 수단이 단일한 법률이나 규제 장치로 환원될 수 없음을 분명히 한다. 교육·리터러시 정책, 플랫폼 거버넌스, 데이터 접근 제도, 기술 표준 마련 등이 상호 연계된 종합적 정책 패키지로 설계될 필요가 있으며, 이는 부처 간 협력과 독립적인 감독·평가 체계의 운영을 전제로 한다. 특히 플랫폼의 정보 유통 영향력이 국가 단위를 넘어 작동하는 현실에서, 국제 공조와 규범 정합성 확보는 선택이 아니라 필수적 요소로 평가된다.

제3절 국내 정보 무결성 관련 법제도 및 정책동향 분석

1. 정보통신망법 개정

2025년 12월 24일 이른바 허위조작정보 근절법이라 불리는 「정보통신망 이용촉진 및

정보보호 등에 관한 법률(이하 정보통신망법이라 한다)」 개정안이 국회 본회의를 통과했다. 동 법안은 불법정보나 허위조작정보를 악의를 가지고 유통시킬 경우 손해액의 최대 5배 손해배상을 청구할 수 있도록 하여 허위조작정보를 강하게 규율할 수 있을 것으로 보인다. 동 개정법은 이러한 규제조항 외에도 정보 무결성 확보를 위한 제도적 장치를 함께 마련하고 있어 그 내용을 살펴보도록 하겠다.

가. 허위조작정보 정의

가장 먼저 개정된 정보통신망법(법률 제21305호, 공포 2026.1.6., 시행 2026.7.7., 이하 개정법이라 한다)에서 허위조작정보를 어떻게 정의하고 있는지 살펴보도록 한다. 본 개정법은 기존 정보통신망법 체계에서 주로 ‘불법정보’를 중심으로 규율되어 있던 범위를 확장하여, 일정한 요건을 충족하는 경우 ‘허위조작정보’로 명명하고 이를 규제하도록 하였다.

개정법은 제44조의7 제2항을 신설하여 ‘허위조작정보’의 개념을 명확히 하였다. 즉, 타인의 인격권이나 재산권 또는 공공의 이익을 침해하는 정보로서, 내용의 전부 또는 일부가 허위인 정보를 허위정보(제1호), 내용을 사실로 오인하도록 변형된 정보를 조작정보(제2호)로 규정하고, 이러한 정보에 해당한다는 사실을 알았음에도 손해를 끼칠 의도 또는 부당한 이익을 얻을 목적으로 정보통신망을 통하여 이를 유통하는 행위를 금지하였다. 다만 풍자와 패러디는 예외로 두었다.

즉, 개정법은 허위조작정보를 단순한 사실 오류나 부정확한 정보에 그치지 않고, ① 허위성, ② 손해를 가할 의도 또는 부당이익 목적이라는 고의성을 모두 충족하는 경우에 한하여 규율 대상으로 한정하였다. 이러한 요건은 허위조작정보를 의견 표현이나 사회적 논쟁의 영역과 구별하여 규율하고자 하는 입법적 의도를 보여준다.

나. 플랫폼 자율규제 및 투명성 보고

개정법은 허위조작정보 대응을 ‘국가의 직접 삭제·차단’ 방식으로만 설계하지 않고, 대규모 플랫폼에게 신고·조치·이의제기·분쟁조정으로 이어지는 절차적 체계를 갖추도록 하여 자율규제를 제도화하는 방향을 취한다. 개정법 제44조의12은 누구든지 대규모 정보통신서비스 제공자가 운영·관리하는 정보통신망에서 유통되는 불법정보와 허위조작정

보를 신고할 수 있도록 하고, 신고 시에는 문제 정보의 위치, 불법 또는 허위조작정보로 보는 이유와 근거, 신고자의 연락처 등을 기재하도록 요구한다(동법 제44조의12 제1,2항). 이는 신고의 남용을 억제하고 사후 검증 가능성을 확보하려는 장치로 이해된다.

플랫폼이 취할 수 있는 조치는 삭제·차단이나 노출 제한뿐 아니라, 계정 정지·해지, 수익화 제한, 금전 지급의 중지·회수, 서비스의 일부 또는 전부 중지·종료 등 광범위하게 설계되어 있다(동법 제44조의12 제3항). 동시에 플랫폼은 조치를 취할 경우 그 정당한 이유와 이의신청 절차를 신고자와 게재자에게 통지해야 하며, 신고자나 게재자는 통지를 받은 날로부터 6개월 이내 무료로 이의신청을 할 수 있다(동법 제44조의12 제4항). 조치와 이의신청 결정은 분쟁조정절차로 다툴 수 있도록 연계되어 있어(동법 제44조의12 제5항), 이용자 권리구제 절차를 제도 내부에 배치한 점이 특징이다.

한편 플랫폼은 자율적 운영정책(판정기준, 신고와 조치 절차 등)을 수립해야 하고, 그 과정에서 이해관계자·시민단체·전문가 의견을 반영해야 한다(동법 제44조의12 제6,7항). 이는 자율규제가 단순한 기업 내부 규칙이 아니라 공적 기능을 수행하는 준규제 장치임을 전제로, 정당성과 투명성을 확보하려는 시도로 평가할 수 있다. 이와 관련하여 이해관계자 의견 반영의 실효성을 높이기 위해, 향후 제도 운영 과정에서 의견 수렴의 절차와 방식이 보다 구체화될 필요가 있다.

투명성 보고 의무는 이러한 자율규제 체계를 외부에서 검증 가능하도록 하는 핵심 수단이다. 개정법 제44조의14 제1항은 대규모 정보통신서비스 제공자에게 6개월에 1회 이상 투명성 보고서를 작성·공표하도록 하고, 이용자의 수·매출액 등 기본 현황, 신고된 불법정보 및 허위조작정보 처리 통계, 이의신청 처리 결과, 국가기관의 명령·권고 및 그에 따른 조치, 자율규제 가이드라인에 따른 처리 내역 등을 포함하도록 요구한다. 특히 보고서는 누구나 쉽게 접근할 수 있도록 접근성이 보장된 방식으로 공표하도록 하고 그 세부 방식은 대통령령으로 정하도록 하였다. 정책효과 측면에서 투명성 보고는 (1) 플랫폼 조치의 예측가능성, (2) 국가기관과 플랫폼 간 상호작용의 가시화, (3) 시민사회·학계의 감시 및 평가 기반을 동시에 제공한다고 볼 수 있다.

투명성 보고가 제도의 취지에 부합하여 운영되기 위해서는 공표 항목의 표준화와 지표 정의의 정합성 확보 등이 함께 논의되어야 하며, 개정법은 이를 위해 방송미디어통신위원회가 자료 제출을 요구하여 사실 확인을 할 수 있는 근거를 마련하고 있다(제44조의14 제3

항). 구체적인 검증 방식과 수준은 향후 집행체계 설계를 통해 단계적으로 정비될 것으로 보인다.

다. 사실확인 활동 지원

정보통신망법 개정법은 허위정보 대응에서 사실확인단체의 역할을 제도적으로 위치시키고, 플랫폼과의 연계를 통해 그 효과가 정보유통 구조에 실제로 반영되도록 한다. 개정법 제44조의16은 대규모 정보통신서비스 제공자에게 허위정보 처리에 대한 자율 정책수립·운영 의무를 부과(제1항)하는 한편, 대통령령이 정하는 국제적 사실확인 절차 규범을 준수하는 사실확인단체와 사실확인 활동 활성화를 위한 협약을 체결할 수 있도록 규정한다(제2항). 또한 사실확인단체는 제2항에 따라 협약을 체결한 제공자가 운영·관리하는 정보통신망에서 유통되는 허위정보등에 관하여 사실확인된 정보, 사실확인 후 취한 조치 등에 관한 보고서를 작성하여 공개하여야 한다(제3항).

이 구조는 세 가지 방향의 정책효과를 이끌어낸다. 첫째, 사실확인단체가 생산한 결과물이 플랫폼 정책에 반영될 수 있도록 제도적으로 ‘연결점’을 만든다는 점이다. 대규모 정보통신서비스 제공자는 공개된 보고서를 서비스에 반영할 수 있으며(동법 제44조의16 제4항), 반영 사실을 이용자에게 공표하도록 되어 있어(동법 제44조의16 제5항), 팩트체킹 결과가 단순히 외부 보고서에 머무르지 않고 이용자 경험 및 정보 노출 구조에 영향을 미치도록 유도한다. 둘째, 사실확인 활동 자체의 투명성과 책임성을 강화할 수 있다. 보고서 공개 의무는 사실확인의 절차·성과·한계를 외부에서 검증 가능하게 한다. 마지막으로 대규모 정보통신서비스 제공자와 사실확인단체와의 협약을 공개하도록 하는 것(동법 제44조의16 제2항)은 임의적 파트너십을 넘어, 협력의 조건과 범위를 사회적으로 가시화함으로써 공정성과 경쟁 중립성을 일부 담보하려는 설계로 볼 수 있다. 사실확인단체의 독립성 확보를 위해 인증 기준, 이해충돌 방지, 오류 정정 체계 등 사실확인단체에 요구되는 국제적 규범 준수 요건이 제도 운영 과정에서 구체화될 필요가 있다.

라. 투명성 센터 설치

개정안 제44조의17은 대규모 정보통신서비스 제공자에 대한 규율이 단순히 ‘민간 자율조치 의무 부과’에 머무르지 않도록, 감독과 지원을 위한 제도적 인프라로서 “정보통신

서비스투명성센터” 설치 근거를 마련하고 있다.

투명성센터의 업무는 사실확인 단체 데이터베이스 운영 및 지원, 사실확인 단체 지원, 연구·교육 지원, 국제협력, 기타 사실확인 활성화 사업 등으로 폭넓게 규정되어 있다. 이는 허위정보 대응을 규제 중심으로만 접근하지 않고, 팩트체킹 생태계의 역량을 제고하여 정보환경의 신뢰성을 높이려는 정책 패러다임을 반영한다. 특히 국제협력을 명시함으로써, 글로벌 플랫폼 환경에서 사실확인 기준과 데이터 연계가 국제 규범과 연결되도록 설계하려는 의도를 확인할 수 있다.

2. 언론중재법 개정 논의

가. 언론의 범위 확대 논의

「언론중재 및 피해구제 등에 관한 법률(이하 언론중재법이라 한다)」은 언론 보도로 인하여 발생하는 피해에 대해 정정보도·반론보도·추후보도 청구권과 손해배상 등 다양한 구제수단을 규정하고, 이러한 분쟁을 신속하고 효율적으로 해결하기 위한 언론중재위원회의 역할과 절차를 정하고 있다.

한국언론진흥재단의 「언론수용자 조사」와 로이터 저널리즘연구소의 「디지털 뉴스 리포트」 등 주요 조사 결과에 따르면, 다수의 이용자들이 종이신문이나 언론사 홈페이지, 포털이 아닌 유튜브를 주요 뉴스 접점으로 활용하고 있는 것으로 나타난다. 이는 과거 포털 뉴스가 수행하던 기능을 오늘날 유튜브가 일정 부분 대체하고 있음을 시사하며, 이 과정에서 유튜브는 단순한 영상 플랫폼을 넘어 사실상 언론과 유사한 사회적 영향력을 행사하는 공간으로 인식되고 있다(양재규, 2025).

언론중재법 변천 과정을 살펴보면, 1980년대 후반 케이블 텔레비전이 방송의 한 유형으로 제도적으로 수용되었고, 2005년 언론중재법 제정 과정에서는 인터넷신문이 법 적용 대상에 포함되었다. 이어 2009년 법 개정을 통해 포털 뉴스와 IPTV가 추가되면서, 언론중재법은 당시의 뉴스 소비 경로와 정보 유통 구조를 반영하는 방향으로 적용 범위를 확장하였다.

<표 2-2> 언론중재법 변천 과정

시기	근거법	위원 정수	대상 매체	청구권	절차
도입기 (1980.12-1987.11)	언론기본법	30~60인	신문, 잡지, 뉴스통신, 방송	반론보도	필요적 전치
과도기 (1987.11-1995.12)	정간법 방송법	40~70인	신문, 잡지, 뉴스통신, 방송 **(CATV 추가)**	반론보도 추후보도	합의 간주
정착기 (1995.12-2005.1)	개정 정간법 방송법	40~80인	신문, 잡지, 뉴스통신, 방송, **인터넷 신문**	정정보도 반론보도 추후보도	직권조정결정
성장기 (2005.1-2009.2)	언론중재법	40~90인	신문, 잡지, 뉴스통신, 방송, 인터넷 신문, **포털뉴스, IPTV**	정정보도 반론보도 추후보도 손해배상	임의적 전치 중재절차 신설
확장기 (2009.2-현재)	개정 언론중재법				

출처 : 이재진(2015), p.204 재구성

최근에는 '유튜브 저널리즘' 현상에 대한 논의가 본격화되면서, 유튜브를 언론중재법상 조정·중재 대상으로 포함할 것인지 여부가 언론중재법 개정 논의에서 중요한 정책 과제로 부상하고 있다. 언론중재법 운용의 핵심 기관인 언론중재위원회 역시 이러한 문제의식을 공유하며, 유튜브를 포함한 신유형 뉴스 서비스의 법적 위치에 관한 논의를 지속적으로 축적해 오고 있다.

특히 이러한 논의는 허위조작정보 대응의 관점에서도 의미를 지니는데, 단순히 특정 매체 유형을 추가하는 문제를 넘어 허위정보가 생산·유통되는 구조 자체의 변화에 제도적으로 대응하려는 시도로 이해할 수 있다. 최근 유튜브를 중심으로 한 뉴스 소비 확산은 전통적 언론사뿐 아니라 개인 또는 소규모 집단이 시사 보도·논평을 지속적으로 생산·유통하는 환경을 형성하고 있으며, 이 과정에서 허위 또는 왜곡된 정보가 사회적 영향력을 갖고 확산되는 사례도 함께 나타나고 있다.

이러한 상황에서 유튜브를 언론중재법의 적용 범위에 포함시키는 것은, 허위조작정보로 인한 피해에 대해 사후적 분쟁 해결과 권리구제의 제도적 통로를 마련한다는 의미를 가진

다. 언론중재법은 형사처벌이나 행정제재 중심의 규율 방식이 아니라, 정정보도·반론보도·추후보도 및 손해배상과 같은 비형벌적·회복적 구제수단을 중심으로 설계된 법률이다. 따라서 일정 요건을 충족하는 유튜브 기반 보도·논평 콘텐츠를 그 적용 대상으로 포함할 경우, 허위조작정보에 대한 대응이 즉각적인 차단이나 제재가 아니라, 사실관계의 시정과 피해 회복을 중심으로 이루어질 수 있는 구조를 제공하게 된다.

나. 언론중재 대상

2002년 언론중재위원회 소위원회에서는 유튜브에 게시되는 콘텐츠를 그 성격과 제작·운영 주체에 따라 네 가지 유형으로 구분하고, 각 유형별로 언론중재법상 조정·중재 대상 해당 여부를 달리 판단하는 기준이 논의되었다.

먼저, 언론사가 이미 보도한 기사와 동일한 콘텐츠를 공식 유튜브 계정에 그대로 제공하는 경우(제1유형)와, 언론사가 운영하는 유튜브 채널을 위해 별도로 제작되었으나 보도의 형식과 내용을 갖춘 콘텐츠의 경우(제2유형)에는 원칙적으로 언론중재법상 조정·중재 대상으로 인정할 수 있다는 기준이 제시되었다. 이는 해당 콘텐츠가 형식적으로는 영상물의 형태를 띠고 있더라도, 운영 주체가 법률상 '언론사'에 해당하고 보도의 실질을 갖추고 있다면 언론 보도로 보아 조정 대상에 포함할 수 있다는 판단에 따른 것이다. 다만, 보도의 형식을 갖추었다고 보기 어려운 일부 예외적인 경우는 제외되는 것으로 정리되었다.

한편, 법률상 언론사는 아니지만 뉴스 전문 채널을 표방하면서 정치·경제·사회 등 각 분야의 정보나 논평을 제공하는 경우(제3유형)에 대해서는 일률적인 기준을 설정하기보다는 판단을 유보하는 방향이 제시되었다. 이는 운영 주체가 언론사가 아닌 경우 피신청인을 특정하기 어렵고, 제공되는 콘텐츠 역시 주관적 의견이나 해설의 성격을 띠는 경우가 많아 언론보도에 해당하는지 여부를 판단하기가 상대적으로 모호하다는 점을 고려한 것이다. 다만 논의 과정에서는 일정 기간 동안 지속적으로 기사를 생산하는 등 인터넷신문 등록 요건에 준하는 실질을 갖춘 경우라면, 조정 대상으로 볼 여지도 있다는 의견도 제시되었다. 이러한 의견을 종합하여, 제3유형에 해당하는 경우에는 조정신청이 접수된 개별 사건마다 각 중재부가 구체적 사정을 고려하여 2차적으로 조정 대상 여부를 판단하는 방식이 제안되었다.

마지막으로, 뉴스 콘텐츠를 표방하지 않고, 게시 주체 역시 법률상 언론사에 해당하지 않는 경우(제4유형)에는 제공되는 콘텐츠가 언론보도로 보기 어렵다는 점에서 언론중재법상 조정·중재 대상으로 인정하기 어렵다는 결론에 이르렀다.

<표 2-3> 언론중재위원회의 조정대상 판단 기준

구분	제공주체	내용·형식		조정대상 여부
1유형	언론중재법상 언론사(사업자)에 해당하는 자	본 매체에 게재된 콘텐츠를 자사의 유튜브 등 채널(계정)에 동일하게 게시		○
2유형		본 매체에 게재하지 않은 콘텐츠를 자사의 유튜브 등 채널(계정)에만 게시	보도	○
			보도아님	×
3유형	언론중재법상 언론사(사업자) 해당여부가 모호하거나 언론사와 유사한 자	정치·경제·사회·문화 등에 관한 논평, 정보 등을 '뉴스 전문을 표방하는 유튜브 등 채널(계정)'에 게시		유보
4유형		다양한 정보를 '뉴스 전문을 표방하지 않는 유튜브 등 채널(계정)'에 게시		×

출처 : 언론중재위원회(2022). 언론사 운영 유투브 채널, 조정대상에 포함 - '조정대상 확대 위한 소위원회' 활동결과, p.3

제4절 국내외 플랫폼 책임 및 공동규제 사례 분석

1. 해외 주요국의 온라인 콘텐츠 관리 및 공동규제 모델 비교

주요 국가들의 플랫폼 콘텐츠 공동규제 정책은 표현의 자유 보장과 이용자 보호라는 두 가지 가치 사이의 균형을 모색하는 과정에서, 전통적인 국가 주도 규제에서 벗어나 자율규제와 법적 규율을 결합한 공동규제(co-regulation) 또는 메타규제(meta-regulation) 방식으로 점차 수렴하고 있다. 각국은 자국의 법제 환경과 미디어 생태계 특성에 따라 상이한 제도 설계를 채택하고 있으나, 공통적으로 플랫폼의 책임을 절차·시스템 중심으로 구조화하고 정부는 사후 감독 역할을 강화하는 방향을 보이고 있다.

가. 유럽연합(EU): 디지털서비스법(DSA) 기반 공동규제 모델

유럽연합은 법적 강제력과 사업자의 자발적 참여를 결합한 가장 제도화된 공동규제 체계를 구축하고 있다. 디지털서비스법(DSA)은 온라인 플랫폼 전반에 대해 불법 콘텐츠 대응, 이용자 보호, 투명성 확보 의무를 부과하며, 특히 초대형 온라인 플랫폼(VLOPs)에 대해서는 허위조작정보를 포함한 시스템적 위험(systemic risks)을 평가하고 이를 완화하기 위한 조치를 취할 의무를 명시하고 있다.

DSA와 병행하여 운영되는 「허위조작정보 대응 실천강령(Code of Practice on Disinformation)」은 플랫폼 사업자, 광고업계, 시민사회, 팩트체크 기관 등이 참여하는 자율규범으로 출발하였으나, DSA 시행 이후에는 위험 완화 조치의 이행 수단으로 제도적으로 연계되었다. 이에 따라 강령 준수 여부는 사실상 DSA 의무 이행의 한 요소로 기능하며, 자율규제가 공동규제로 전환된 대표적 사례로 평가된다(이수엽 외, 2023).

나. 독일: 규제된 자율규제(Regulated Self-regulation) 모델

독일은 법률에 근거한 강한 규율 아래 민간의 자율적 집행을 결합한 ‘규제된 자율규제’ 모델을 채택하고 있다(심영섭, 2024). 네트워크 집행법(NetzDG)은 소셜 네트워크 사업자에게 명백한 불법 콘텐츠에 대해 24시간 이내, 복잡한 사안에 대해서는 7일 이내 조치를 취하도록 의무화하고, 이를 위반할 경우 고액의 과태료를 부과할 수 있도록 하고 있다.

다만 실제 콘텐츠 판단과 조치의 1차적 책임은 플랫폼의 내부 시스템에 맡겨져 있으며, 정부는 투명성 보고와 사후 감독을 통해 개입한다. 아울러 정부 승인 하에 운영되는 민간 자율규제기구(FSM 등)는 회원사의 콘텐츠를 사전·사후적으로 심사하고, 해당 결정을 따를 경우 사업자가 법적 책임을 일부 면제받는 구조를 갖고 있다. 이는 강한 법적 틀 안에서 자율규제의 유인을 제공하는 방식으로 평가된다(이수엽 외, 2023).

다. 호주: 정부 주도형 공동규제 모델

호주는 정부가 정책 목표와 기본 방향을 설정하고, 산업계가 세부 강령을 수립·이행하는 정부 주도형 공동규제 모델을 운영하고 있다. 호주 통신·미디어 규제기관(ACMA)의 감독 하에, 산업협회인 DIGI가 「허위조작정보 및 오정보 대응 실천강령」을 제정하여 주요

플랫폼이 자율적으로 참여하도록 하고 있다.

이 모델의 특징은 조건부 강제성에 있다. 즉, 자율규제가 실질적 성과를 거두지 못할 경우 정부가 직접 산업 표준을 제정하여 법적 강제를 가할 수 있는 권한을 보유하고 있어, 자율규제에 대한 강한 이행 유인을 제공한다(김희경 외, 2022).

라. 영국: 주의 의무(Duty of Care) 중심 모델

영국은 온라인 안전법(Online Safety Act)을 통해 플랫폼 사업자에게 이용자를 보호해야 할 법적 주의 의무(duty of care)를 부과하는 모델을 채택하였다. 사업자는 불법 콘텐츠 및 특정 유해 콘텐츠로부터 이용자를 보호하기 위한 합리적이고 비례적인 조치를 취해야 하며, 그 구체적 이행 방식은 사업자의 자율에 맡기되 감독기관인 OFCOM이 이를 평가·감독한다.

OFCOM은 코드와 가이드라인을 통해 기대 수준을 제시하고, 미이행 시에는 과징금 부과 등 실질적 제재를 가할 수 있다. 이는 국가가 결과를 직접 판단하기보다는, 절차와 시스템의 적정성을 중심으로 플랫폼 책임을 점검하는 구조라는 점에서 공동규제의 전형으로 평가된다(이수엽 외, 2023).

마. 미국: 자발적 자율규제 및 면책 중심 모델

미국은 수정헌법 제1조에 따른 표현의 자유 보호 전통에 따라 정부의 직접적 콘텐츠 규제를 최소화하고, 민간 자율규제와 시장 기반 협력에 의존하는 경향이 강하다. 통신품위법(CDA) 제230조는 플랫폼 사업자에게 이용자 게시 콘텐츠에 대한 광범위한 면책을 부여함으로써, 자율적 콘텐츠 관리와 혁신을 촉진해왔다.

이러한 환경 속에서 IFCN 인증 체계[13], 언론사 컨소시엄 기반 팩트체크 프로젝트, 플랫폼과 팩트체커 간 자율적 협력(라벨링, 맥락 제공 등)이 허위정보 대응의 핵심 수단으로 기능하고 있다. 다만 최근에는 아동 보호, 딥페이크 등 특정 영역을 중심으로 플랫폼 책임

13) IFCN(International Fact-Checking Networkd)는 2015년 미국 포인터 연구소가 설립한 글로벌 네트워크로 ① 비당파성과 공정성, ②출처의 투명성, ③ 자금 및 조직의 투명성, ④ 방법론의 투명성, ⑤ 책임성과 수정의무 등 팩트체크 5대 원칙을 제시하고 있다. 이 원칙은 팩트체크 단체가 지켜야 할 윤리강령 같은 역할을 수행하며, IFCN 기준을 준수하는 단체를 “IFCN signatory”로 인증하고 있다.

을 강화하려는 입법 논의가 연방 및 주 차원에서 확대되고 있다.

이상의 사례를 종합해보면, 주요 국가들은 과거의 방임적 자율규제에서 벗어나 국가가 법과 원칙을 설정하고, 사업자가 세부 집행을 담당하되, 정부가 사후적으로 감독·평가하는 협력적 공동규제 모델로 점진적으로 이동하고 있음을 확인할 수 있다. 이는 규제의 직접성과 정치적 부담을 완화하면서도, 플랫폼의 구조적 책임을 제도적으로 확보하려는 정책적 선택으로 해석된다(이수엽 외, 2024).

2. 한국의 플랫폼 공동규제 현황

가. 법적 규제 체계: 공적 규제 중심 구조

지금까지 한국에는 '허위조작정보'를 포괄적으로 정의하고 이를 직접적으로 규율하는 단일한 독립 법률은 존재하지 않았으며, 「형법」상 명예훼손·모욕죄, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률(정보통신망법)」, 「공직선거법」상 허위사실공표죄 등 개별 법령을 통해 불법 정보에 대한 사후적 처벌과 유통 제한이 이루어져 왔다. 이러한 규율 방식은 행위 유형별·영역별 대응에는 일정한 효과를 가져왔으나, 허위조작정보 전반을 체계적으로 포섭하고 예방·절차 중심으로 관리하는 데에는 구조적 한계를 지닌다는 평가가 제기되어 왔다.

최근에는 기술 환경 변화에 대응하여 부분적인 입법 보완이 이루어졌다. 2023년 12월 공직선거법 개정을 통해, 선거의 공정성을 중대하게 훼손할 우려가 있는 딥페이크 기술을 이용한 선거운동을 선거일 전 90일부터 선거일까지 금지하고, 이를 위반할 경우 형사처벌하도록 규정하였다. 이는 허위조작정보 중에서도 사회적 위험성이 특히 높은 영역에 대해 선별적으로 규제한 사례로 평가된다.

아울러 2025년 12월 24일에는 「정보통신망법」 개정안이 국회를 통과함에 따라, 허위조작정보를 명시적으로 규율할 수 있는 법적 근거가 새롭게 마련되었다. 앞서 살펴보았듯이 동 개정법률은 일정 요건을 충족하는 허위조작정보의 유통을 금지하고, 허위조작정보의 신고 및 그에 대한 조치·이의제기·분쟁조정, 투명성 보고, 사실확인단체 연계 등 절차적 대응 체계를 법률 차원에서 도입함으로써, 기존의 파편적이고 사후적인 규율에서 벗어나 보다 구조적이고 제도화된 대응의 출발점을 마련하였다는 점에서 의미를 갖는다. 다만 이러한 법적 기반은 여전히 공적 규제의 틀을 중심으로 설계되어 있으며, 플랫폼의 자

율적 위험 평가나 예방적 조치를 제도적으로 유도하는 공동규제 구조로까지 완전히 확장되었다고 보기는 어렵다. 향후에는 개정된 「정보통신망법」 을 토대로, 공적 규제와 자율규제가 기능적으로 결합되는 플랫폼 공동규제 체계를 어떻게 구체화할 것인지가 핵심 과제로 남아 있다고 할 수 있다.

나. 인터넷 정보 규율의 이원적 심의·집행 구조

우리나라의 인터넷 정보 규율은 이원적 심의 구조를 특징으로 한다. 독립 민간기구인 방송통신심의위원회(KCSC)가 불법·유해 정보에 대한 심의·의결을 담당하고, 방송미디어통신위원회(KMCC)가 이를 근거로 플랫폼 사업자에게 시정명령 등 행정적 조치를 부과하는 구조이다. 이는 형식적으로는 심의와 집행의 분리를 통해 절차적 중립성을 확보하려는 장치이나, 실질적으로는 공적 규제 권한이 강하게 작동하는 구조로 평가된다.

다. 자율규제 현황 및 민간 기구의 역할

공적 규제와 병행하여, 한국에서는 다양한 형태의 민간 자율규제 장치가 운영되고 있다. 대표적으로 한국인터넷자율정책기구(KISO)는 2008년 설립된 국내 최초의 민간 자율규제기구로서 네이버·카카오 등 주요 포털 사업자가 참여하여 언론보도 형식을 도용한 허위정보나 사회적 파급력이 큰 특정 이슈(예: 감염병, 재난 등)에 대한 허위정보를 자율적으로 심의·조정하고 있다.

이와 더불어 2022년에는 검색·추천 서비스의 공정성과 투명성 제고를 목적으로 한 '플랫폼 민간 자율기구'가 출범하여, 알고리즘 운영 원칙 마련 등 플랫폼의 사회적 책무를 강화하기 위한 논의를 진행하고 있다. 이는 허위정보 자체의 판단보다는, 정보 노출 구조와 절차의 투명성에 초점을 둔 자율규제 시도로 볼 수 있다.

또한 2023년 3월부터 시행된 「영화 및 비디오물의 진흥에 관한 법률」 개정안에 따라, 일정 요건을 충족한 OTT 사업자는 영상물등급위원회를 거치지 않고 자체적으로 콘텐츠 등급을 분류할 수 있게 되었다. 이는 공적 사전심의를 완화하고 사업자의 자율성과 책임을 확대하는 제도적 변화로, 한국 규제 체계 내 자율규제 요소가 부분적으로 강화되고 있음을 보여준다.

3. 한국의 공동규제 모델의 한계와 개선방안

그럼에도 불구하고, 현행 한국의 규제 구조는 여러 한계를 안고 있다. 우선 「정보통신망법」 등에서 자율규제의 필요성을 언급하고는 있으나, 구체적인 절차 및 유인 구조가 마련되어 있지 않아 선언적 규정에 머물러 있다는 지적이 지속적으로 제기되고 있다. 이로 인해 사업자의 자발적 참여를 촉진하거나 자율규제 성과를 제도적으로 인정하는 장치가 부족한 실정이다.

또한 자율규제가 이루어지더라도, 동일 사안에 대해 공적 규제기관이 다시 심의·의결을 반복하는 구조로 인해 규제의 예측 가능성과 안정성이 저하되고, 이중 규제 논란이 발생하고 있다는 문제도 지적된다. 이는 민간 자율규제가 공적 규제에 종속된 보조 수단으로 인식되는 한계를 드러낸다.

이러한 문제의식 속에서 최근에는 유럽연합의 디지털서비스법(DSA)이나 호주의 정부 주도형 공동규제 사례를 참고하여, 국가가 기본 원칙과 가이드라인을 설정하고 민간이 세부 행동강령을 수립·이행하는 ‘규제된 자율규제’ 또는 ‘메타규제’ 모델 도입 논의가 확산되고 있다. 이는 공적 규제의 강제력과 민간의 전문성·유연성을 결합하여 허위조작정보 대응의 실효성과 정당성을 동시에 제고하려는 접근으로 평가된다.

정리하면, 한국의 허위조작정보 대응 체계는 강력한 공적 규제 권한을 중심으로 하면서도, 자율규제 요소를 부분적으로 결합한 정부 주도형 위임 규제 모델에 가까운 구조를 유지해 왔다고 볼 수 있다. 다만 최근 제도 변화와 논의 흐름은, 점차 민간의 역할과 책임을 제도적으로 인정하고 공적 규제는 원칙 설정과 사후 감독에 집중하는 협력적 공동규제 체계로의 전환 가능성을 시사한다. 향후 제도 설계 과정에서는 자율규제의 실질적 유인과 책임성을 강화하는 동시에, 공적 규제와의 기능 분담과 경계 설정을 보다 명확히 하는 방향이 중요한 과제로 제기된다.

제3장. 정보 무결성 관련 선행연구 메타 분석

제3장은 정보 무결성 관련 논의의 이론적·정책적 토대를 검토하기 위해, 허위조작정보를 중심으로 한 선행연구와 정보 무결성 관련 주요 해외 정책 문서를 체계적으로 분석한다. 특히 허위조작정보에 관한 국내 연구는 개념 정의, 규제 필요성, 헌법적 한계, 대응 방안 등 다양한 쟁점을 중심으로 전개되어 왔으나, 개별 논의가 분절적으로 제시되는 경향이 있었다. 이에 본 장에서는 선행연구와 정책 문서를 질적으로 메타분석함으로써 연구 경향과 공통된 논의 구조를 정리하고, 이를 바탕으로 향후 제도 설계 논의에 대한 시사점을 제시하고자 한다.

제1절 허위조작정보 관련 연구 메타분석

제1절에서는 분석 대상 문헌의 범위를 국내 학술연구로 한정하고, “허위조작정보”, “가짜뉴스”, “플랫폼 규제” 등 핵심 검색어를 조합하여 학술 데이터베이스를 중심으로 1차 검색을 실시하였다. 이후 중복 문헌을 제거한 뒤, 제목·초록·키워드 수준에서 연구 주제가 허위조작정보의 개념, 규제 필요성 및 규제 방안과 직접적으로 관련되는지를 기준으로 2차 선별을 진행하였다. 이 과정에서 ‘가짜뉴스’라는 용어를 부분적으로 언급하였더라도, 연구의 중심이 커뮤니케이션 이론 일반이나 미디어 리터러시 교육의 효과성 등 허위조작정보 규제와 직접적 관련성이 낮은 경우에는 분석 대상에서 제외하였다. 이러한 절차를 거쳐 최종적으로 국내 문헌 52편을 선정하였으며, 이를 대상으로 질적 메타분석을 실시하였다. 분석 결과, 개별 연구들은 다양한 문제의식과 접근법을 취하고 있으나, 전반적으로 가짜뉴스 및 허위조작정보의 개념 정의, 규제의 정당성, 헌법적 한계, 그리고 대응 수단이라는 공통된 논의 틀을 중심으로 전개되어 온 경향이 확인되었다.

1. 허위조작정보 관련 국내 연구 경향

먼저 기존 문헌을 살펴보면 국내외 허위조작정보 관련 연구는 크게 다섯 가지 주제로 분류할 수 있다. 이들 연구는 공통적으로 온라인 플랫폼 환경에서 발생하는 허위정보, 즉

가짜뉴스 또는 허위조작정보의 생성과 확산 양상을 분석하고, 이에 대한 법적·제도적·기술적 대응 방안을 모색하는 한편, 그러한 규제가 지니는 한계와 부작용을 함께 검토하고 있다.

첫째, 허위조작정보(가짜뉴스)의 개념 정립과 확산 요인에 관한 연구이다. 이 주제에서는 가짜뉴스, 잘못된 정보(misinformation), 허위조작정보(disinformation) 등 유사 개념 간의 차이와 개념적 모호성이 반복적으로 지적되며, 규범적·법적 차원에서 보다 명확한 정의가 필요하다는 점이 강조된다. 또한 가짜뉴스가 확산되는 배경으로는 미디어 기술의 비약적 발전과 온라인 플랫폼의 구조적 특성(곽선혜,이성욱, 2022; 이정념, 2018), 언론 환경의 상업화와 정파성, 그리고 이용자의 심리적 요인 등이 복합적으로 작용하는 것으로 분석된다(곽선혜,이성욱, 2024). 특히 확증편향, 인지적 일관성 추구, 부정적 정보에 대한 민감성, 이른바 여과거품(filter bubble) 현상은 허위정보의 수용과 재확산을 촉진하는 핵심 요인으로 지목된다. 나아가 생성형 인공지능 기술과 딥페이크를 활용한 텍스트·이미지·대화형 콘텐츠 기반의 가짜뉴스 제작이 용이해지면서, 허위정보가 갖는 파급력과 사회적 위험성이 질적으로 증폭되고 있다는 점도 중요한 연구 쟁점으로 다루어진다(김헌재, 2024; 배상균, 2021).

둘째, 온라인 플랫폼의 검색 순위 결정 방식과 알고리즘 투명성 규제에 관한 연구이다. 이 주제는 검색 결과나 추천 순위가 이용자의 선택과 의사결정에 미치는 영향력이 막대하다는 점에 주목하며, 플랫폼 사업자가 자사 이익 중심으로 알고리즘을 운영하는 것을 어떻게 통제할 것인지에 대한 법적 논의를 포함한다. 특히 유럽연합(EU)의 플랫폼–비즈니스 간 관계에 관한 규정과 현대화 지침은 플랫폼 사업자에게 검색 순위를 결정하는 주요 매개변수와 그 상대적 중요도를 공개하도록 요구함으로써, 알고리즘 운용의 투명성을 제도적으로 확보하려는 시도로 평가된다. 다만 이러한 규제는 영업비밀 보호와의 충돌 가능성, 알고리즘 공개 범위의 한계, 개인 맞춤형 추천 시스템의 특수성 등 현실적·법적 쟁점을 동반하고 있으며, 투명성 확보와 혁신 보호 사이의 균형이 주요 과제로 제시된다(Christoph Busch, 정신동, 2021).

셋째, 표현의 자유와 가짜뉴스 규제의 헌법적 한계에 관한 연구이다. 이 주제에서는 가짜뉴스 규제가 헌법상 표현의 자유와 충돌할 위험성이 크다는 점이 반복적으로 지적되며, 특히 과도한 규제가 이용자와 언론의 자기검열을 유발하는 위축 효과(chilling effect)를 초

래할 수 있다는 비판이 제기된다(강지현, 2021; 곽선혜,이성욱, 2022; 김성화, 2019; 김연진, 2021; 노동일, 정완, 2020; 문의빈, 2024; 이상훈, 2020; 지성우, 2019). 과거 전기통신기본법상 허위통신죄가 명확성의 원칙과 과잉금지의 원칙에 위배된다는 이유로 위헌 결정을 받은 사례는, 가짜뉴스를 형사적으로 처벌하는 입법이 갖는 구조적 위험성을 보여주는 대표적 선례로 분석된다(강지현, 2019). 또한 사상의 자유시장 이론이 전통적으로 진리 발견을 위한 경쟁을 강조해 왔으나, 오늘날 플랫폼의 독과점 구조와 허위정보의 범람으로 인해 이 이론이 충분히 작동하지 않는다는 비판도 제기된다(문의빈, 2024). 이와 함께 독일의 네트워크 집행법(NetzDG)은 불법 콘텐츠의 신속한 삭제를 플랫폼 사업자에게 의무화하고 과태료를 부과함으로써 허위·유해 정보 확산에 대응하고 있으나, 사적 검열의 확대와 국가 공권력의 민간 위임이라는 헌법적 논란을 야기한 사례로 검토된다(강지현, 2019).

넷째, 허위조작정보로 인한 피해에 대한 법적 구제 수단으로서 징벌적 손해배상 및 형사적 책임에 관한 연구이다. 이 주제에서는 현행 법체계상 형법의 명예훼손죄와 업무방해죄, 정보통신망법상의 사이버 명예훼손 규정, 공직선거법상의 허위사실공표죄 등 다양한 처벌 규정이 이미 존재함에도 불구하고, 실효성 있는 피해 구제가 이루어지고 있는지에 대한 문제 제기가 이루어진다. 일부 연구는 단순한 실손해 배상만으로는 악의적이고 반복적인 허위정보 유포 행위를 억제하기에 한계가 있다며 징벌적 손해배상제도의 도입을 주장한다(김상찬,이충은, 2009). 반면, 이러한 제도는 대륙법계인 한국의 민사책임 구조와 부합하지 않을 뿐 아니라, 언론 및 표현의 자유를 과도하게 위축시킬 우려가 있다는 점에서 신중한 접근이 필요하다는 반론도 강하게 제기된다(김연진, 2021). 이에 따라 대안으로는 위자료 산정 기준을 현실화하거나, 특별법을 통해 제한적·예외적으로 징벌적 요소를 도입하는 방안 등이 제시된다.

다섯째, 미디어 리터러시 교육과 자율 규제 강화를 통한 장기적 대응 방안에 관한 연구이다. 이 주제는 가짜뉴스 문제를 법적 규제만으로 해결하는 데에는 근본적인 한계가 있다는 인식에 기반하여, 교육과 민간 차원의 자정 노력을 중시한다. 다수의 연구는 개인의 미디어 리터러시 수준이 높을수록 정보의 진위를 검증하려는 행동과 허위정보에 대한 비판적 대응이 강화된다는 점을 실증적으로 제시하며, 이는 단기적 규제보다 교육이 장기적으로 더 효과적인 대응 수단이 될 수 있음을 시사한다(심재웅, 신명환, 2024). 아울러 온라인 플랫폼 사업자 역시 팩트체크 시스템 도입, 알고리즘 개선, 유해 콘텐츠에 대한 광고

수익 제한 등 자율적 조치를 통해 허위정보의 유통을 억제할 책임이 있다는 점이 강조된다(노동일, 정완, 2018). 이러한 맥락에서 국가는 직접적이고 강제적인 규제를 최소화하고, 규제된 자율규제나 공동 규제(co-regulation) 모델을 통해 민간의 전문성과 효율성을 활용하되, 공적 감독을 통해 민주적 정당성과 책임성을 확보해야 한다는 정책적 제언이 제시된다(김연진, 2021).

2. 허위조작정보 개념의 구성 요소

기존 연구들에서 공통적으로 제시된 가짜뉴스(fake news) 및 허위조작정보(disinformation)의 개념적 구성 요소는 다음의 네 가지로 정리할 수 있다. 즉, 내용의 허위성, 고의성 또는 기만적 의도, 특정한 목적성, 그리고 형식성이다.

첫째, 내용의 허위성 또는 부정확성은 가짜뉴스나 허위조작정보를 구성하는 가장 기본적인 요소로 이해된다. 이는 해당 정보가 객관적인 사실과 다르거나 진실에 부합하지 않는 내용을 포함하고 있음을 의미한다. 가짜뉴스는 허위의 사실을 적시해야 하며(강지현, 2019; 김민정, 2019; 이완수, 2018; 한갑운, 윤종민, 2017), 여기에는 전적으로 꾸며낸 이야기뿐만 아니라 객관적 사실의 일부 또는 전부를 날조한 내용도 포함된다(강지현, 2019). 유럽연합(EU) 고위전문가그룹(HLEG)은 허위조작정보를 ‘거짓이거나, 부정확하거나, 또는 오해를 불러일으킬 수 있는 정보’로 정의하고 있으며, 국내 연구에서도 허위조작정보를 ‘객관적 사실관계를 의도적으로 조작한 허위의 사실’로 규정하는 경향이 확인된다. 다만 내용이 전체적으로 사실에 부합하는 경우, 세부적인 부분에서 일부 부정확하거나 다소 과장된 표현이 존재하더라도 이를 곧바로 허위조작정보로 보기는 어렵다는 점도 함께 지적된다(강지현, 2019; 한갑운, 윤종민, 2017).

둘째, 고의성 또는 기만적 의도는 잘못된 정보(misinformation)와 허위조작정보(disinformation)를 구분하는 핵심적인 요소로 논의된다. 여기서 고의성은 정보의 생산자나 유포자가 해당 정보가 허위임을 인식하면서도 타인을 속이거나 기만하려는 의도를 가지고 있었는지를 의미한다. 다수의 연구는 가짜뉴스를 ‘허위임을 알면서도 고의로 유포되는 정보’ 또는 ‘기만적 의도성을 가진 행위’로 정의한다(강지현, 2019; 김민정, 2019; 김연진, 2021). 이러한 의도성은 단순한 착오나 부정확한 보도와 같은 잘못된 정보(misinformation)와 허위조작정보(disinformation)를 구별하는 결정적 기준으로 기능한다. 오보나 단순한 실수의 경우 기만할 목적이 존재하지 않으므로, 일반적으로 가짜뉴스로 평

가되기 어렵다. 이와 관련하여 허위조작정보는 ‘사람들을 속이거나 현혹시키기 위해 만들어진 허위정보’를 의미하며, ‘조작’이라는 표현 자체가 허위성과 기만성을 동시에 내포하고 있다는 점에서 개념적 특성이 분명히 드러난다(김민정, 2019; 이완수, 2018).

셋째, 특정한 목적성 역시 가짜뉴스 및 허위조작정보의 중요한 구성 요소로 제시된다. 기존 연구들에 따르면, 가짜뉴스는 대체로 특정한 목표를 달성하기 위한 수단으로 제작·유포되며, 그 목적은 주로 정치적 또는 경제적 이익에 있다. 경제적 목적에는 광고 수익을 극대화하거나 특정 기업이나 개인에게 재산적 손실을 가하려는 의도가 포함되며, 정치적 목적은 선거 과정에서 특정 방향으로 여론을 유도하거나, 사회적 혼란을 조장하는 것을 의미한다. EU의 허위조작정보 정의 또한 공중에게 해를 끼치거나 특정한 이익을 얻을 목적이 있음을 명시함으로써, 목적성을 핵심 요소로 포함하고 있다(강지현, 2019; 곽선혜, 이성욱, 2022; 김민정, 2019; 김연진, 2021; 이완수, 2018; 한갑운, 윤종민, 2017).

넷째, 형식성 또는 언론 보도 양식의 모방은 특히 ‘가짜뉴스(fake news)’라는 용어가 사용되던 초기 논의에서 강조된 요소이다. 가짜뉴스는 뉴스의 형식을 차용하여 검증된 사실인 것처럼 포장된 거짓 정보로 정의되며, 이는 수용자로 하여금 정보의 신뢰성을 오인하도록 하기 위한 전략으로 이해된다(김민정, 2019). 실제로 국내 초기 학술 논의에서는 언론 보도 형식의 모방이 가짜뉴스 개념의 핵심 요소로 반복적으로 제시되었고, 제20대 국회에 발의된 관련 법안들에서도 이 요소가 공통적으로 포함된 바 있다. 그러나 ‘허위조작정보(disinformation)’ 개념을 중심으로 한 최근 연구들은, 허위정보가 찌라시, 조작된 영상, 이미지, 소셜미디어 게시물 등 뉴스 형식을 넘어 다양한 방식으로 유통되고 있다는 점을 지적한다(김용희 외, 2023; 김연진, 2021). 이에 따라 뉴스 형식의 모방을 규제의 필수 요건으로 설정할 경우 규제 범위를 지나치게 축소시킬 수 있다는 비판도 함께 제기된다(김민정, 2019; 김연진, 2021).

3. 허위조작정보 대응 방안

허위조작정보에 대한 대응 방안과 관련하여 기존 연구와 정책 문헌들은 크게 네 가지 방향에서 논의를 전개하고 있다. 즉, 행정적·형사적 규제를 통한 직접적 통제, 플랫폼 사업자의 책무 강화, 자율 규제와 장기적 대안의 모색, 그리고 국제 협력과 관할권 확보이다.

첫째, 행정적·형사적 규제와 관련하여, 허위조작정보에 대한 대응은 이미 현행 법률 체

계 내에서 부분적으로 이루어지고 있으나, 디지털 환경의 변화에 대응하기 위해 규제를 강화하려는 입법 시도도 지속되고 있다. 현행법상 허위 사실을 유포하여 개인의 명예를 훼손하는 행위는 형법상의 허위사실 적시 명예훼손죄나 정보통신망법상의 사이버 명예훼손죄로 처벌이 가능하다. 또한 선거 과정에서 특정 후보자에게 유리하거나 불리하도록 허위 사실을 공표하는 행위는 공직선거법상 허위사실공표죄로 형사 처벌 대상이 된다. 나아가 매크로 프로그램 등을 이용해 포털 사이트나 온라인 커뮤니티의 댓글 추천 수를 조작하는 행위는 컴퓨터장애업무방해죄로 처벌될 수 있다(강지현, 2019; 김연진, 2021; 이문한, 2021; 이원상, 2019; 지성우, 2019; 한갑운, 윤종민, 2017).

한편, 허위보도로 인한 피해에 대해 기존의 민사상 손해배상만으로는 억지력이 충분하지 않다는 문제의식에 따라, 손해액의 일정 배수를 배상하도록 하는 징벌적 손해배상제도의 도입을 주장하는 견해도 제기되어 왔다(봉연준, 2023). 그러나 이러한 제도는 표현의 자유를 과도하게 침해하고 위축 효과를 초래할 수 있으며, 한국의 대륙법계 민사 책임 구조와 충돌한다는 비판이 있다(강지현, 2019; 김연진, 2021; 이문한, 2021; 봉연준, 2023). 이에 따라 대안으로는 위자료 산정 기준을 현실화하거나 상향 조정하는 방안이 제시되고 있다(김연진, 2021).

둘째, 플랫폼 사업자의 책무 강화는 허위조작정보가 주로 온라인 플랫폼을 통해 급속도로 확산된다는 점에서 핵심적인 대응 수단으로 논의되고 있다. 국내에서는 정보통신망법에 따라 플랫폼 운영자가 명예훼손 등 권리 침해가 소명된 정보에 대해 삭제 또는 임시조치 등 필요한 조치를 취하도록 규정하고 있으며, 권리 침해가 예상되는 경우 임의적 임시조치도 허용하고 있다. 국제적으로는 유럽연합이 디지털서비스법(DSA)을 통해 초대형 온라인 플랫폼(VLOP)에 대해 허위조작정보를 포함한 불법·유해 콘텐츠의 위험을 평가하고 이를 완화할 법적 의무를 부과하였다. 특히 추천 시스템의 주요 매개변수와 그 상대적 중요도를 약관에 투명하게 공개하도록 하고, 불법 콘텐츠 신고 절차, 위험 평가 및 완화 조치, 투명성 보고서 공개 등의 의무를 규정하고 있다. 독일은 과거 네트워크 집행법(NetzDG)을 통해 소셜 네트워크 사업자에게 형법상 불법 콘텐츠를 24시간 이내에 삭제·차단하도록 의무화하고, 위반 시 고액의 과징금을 부과하는 강력한 규제를 시행한 바 있으며, 현재는 DSA 체계와의 정합성을 고려한 대체 논의가 이루어지고 있다. 프랑스 역시 정보조작대처법을 통해 선거 기간 중 플랫폼 사업자에게 광고비 및 자금 출처 공개 등 투

명성 의무와 협력 의무를 부과하고, 판사의 가처분 결정을 통해 허위정보의 유포를 중지시킬 수 있는 제도를 도입하였다.

셋째, 자율 규제와 공동 규제, 그리고 장기적 대안에 관한 논의이다. 법적 규제가 갖는 표현의 자유 침해 가능성과 위축 효과를 최소화하기 위해, 민간의 자율적 노력과 정부의 제도적 지원을 결합한 규율 방식이 중요하게 제시된다(김연진, 2021; 이수엽 외(2024); 문의빈, 2024; 심영섭, 2024). 규제된 자율규제(regulated self-regulation)는 국가가 기본적인 법적 틀을 마련하되, 플랫폼 사업자가 자발적으로 규범을 형성·이행하고 정부가 이를 감독·지원하는 방식으로, 행정 규제의 경직성과 순수 자율 규제의 한계를 동시에 보완할 수 있는 대안으로 평가된다. 이와 관련하여 플랫폼 사업자에게는 허위조작정보의 노출을 줄이기 위한 알고리즘 개선, 허위 계정 탐지 및 차단, AI 기반 가짜뉴스 탐지 시스템 구축과 같은 기술적 조치뿐 아니라(김연진, 2021; 이수엽 외, 2024; 김헌재, 2024), 허위조작정보에 대한 광고 수익 배분 중단이나 정치 광고의 투명성 강화와 같은 경제적 억제 수단이 요구된다(김용희 외, 2023; 이수엽 외, 2024). 또한 언론사, 학계, 시민단체 등과의 협력을 통한 팩트체크 시스템 도입과 검증 결과의 표시 역시 중요한 자율 규제 수단으로 제시된다. 더 나아가 장기적으로는 미디어 리터러시 교육을 통해 이용자의 비판적 사고 역량과 정보 분별 능력을 강화하는 것이 가짜뉴스 확산을 근본적으로 억제하는 핵심 대안으로 강조된다(김헌재, 2024; 곽선혜, 이성욱, 2022; 김용희 외, 2023; 김창화, 2018; 홍숙영, 정의철, 2017; 최승필, 2020; 한갑운, 윤종민, 2017; 황성기, 2018).

넷째, 국제 협력과 관할권 확보의 필요성이다. 허위조작정보는 국경을 넘어 유통되는 특성을 가지므로, 개별 국가 차원의 대응만으로는 한계가 있다. 이에 따라 미국, EU, 독일 등은 허위조작정보 문제를 국가 안보 차원에서 접근하고, 프랑스는 외국의 영향력 행사를 방지하기 위해 특정 서비스에 대한 유포 중단 명령 권한을 도입하였다(김일기, 2025). UN과 OECD 등 국제기구 역시 허위조작정보 대응을 위한 공동의 원칙과 정책 프레임워크 마련을 촉구하고 있으며, 국내 차원에서도 이러한 국제 논의에 적극 참여하고 협력 체계를 구축하는 것이 중요하다는 점이 지적된다. 아울러 「정보통신망법」은 국외에서 이루어진 행위라 하더라도 국내 시장이나 이용자에게 실질적인 영향을 미치는 경우 적용할 수 있도록 역외 적용 조항을 두고 있어, 허위조작정보 대응의 법적 범위를 확장하고 있다(김연진, 2021; 김용희 외, 2023).

제2절 정보 무결성 관련 해외 정부보고서 메타분석

1. 분석 대상 및 방법론

가. 분석 대상 선정 이유

본 연구는 글로벌 차원의 정보 무결성(Information Integrity) 대응 체계를 심층적으로 고찰하기 위하여, 국제기구의 정책 권고 문서와 주요국의 실행 규범을 분석 대상으로 선정하였다. 분석 대상의 선정 이유는 다음과 같다.

첫째, UN과 OECD의 정보 무결성 관련 보고서는 허위조작정보 문제를 기술적·산업적 차원을 넘어 민주주의의 제도적 안정성과 시민의 기본권 보호와 직결된 구조적 과제로 규정한다는 점에서 중요성이 인정되고 있다. 이들 국제기구는 표현의 자유, 알 권리, 공적 담론의 신뢰성 등 보편적 민주주의 가치를 중심에 두고 전 지구적 차원의 정책 원칙과 가이드라인을 제시하고 있으며, 이는 개별 국가의 정책을 평가하고 비교할 수 있는 표준적 준거 틀(baseline)을 제공한다는 점에서 분석 대상으로서의 의의를 가진다.

둘째, EU와 호주의 허위조작정보 실천강령(Code of Practice)은 민간 플랫폼 사업자의 자율적 책무성과 국가의 정책적 개입을 결합한 이른바 공동 규제(co-regulation) 모델을 제도적으로 구현한 대표적 사례로, 정보 무결성 정책을 플랫폼 거버넌스 차원에서 구체화한 실행 규범이다. 특히 이들 실천강령은 단순한 윤리 선언이나 권고 수준에 그치지 않고, 구체적인 이행 약속(commitments), 성과 지표(KPI), 정기적 보고 및 공개 의무를 포함하고 있어, 정보 무결성 정책이 실제 집행 단계에서 어떠한 방식으로 작동하고 평가되는지를 분석하기에 적합한 분석 대상이다.

나. 분석 방법

본 연구는 정보 무결성 관련 해외 정책 문서와 실천강령이 규범적·정책적 성격의 질적 텍스트로 구성되어 있다는 점을 고려하여, 질적 해석적 메타합성(Qualitative Interpretive Meta-Synthesis, QIMS) 방법론을 적용하였다. QIMS는 개별 문헌의 단순 요약이나 비교를 넘어, 다수의 질적 자료로부터 핵심 주제를 도출하고 이를 상위 개념으로 재구성함으로써

새로운 해석적 이해를 생성하는 데 적합한 방법론이다. 본 연구에서는 QIMS를 다음의 세 단계로 구분하여 수행하였다.

먼저 OECD 및 UN의 정보 무결성 관련 보고서와 EU 및 호주의 허위조작정보 실천강령을 분석 대상으로 삼아, 각 문서에 포함된 정책 목표, 규범적 원칙, 제도 설계 요소들을 체계적으로 검토하였다. 그 결과, 정보 무결성 정책은 ① 정책 목표의 설정, ② 제도적 프레임워크의 구성, ③ 플랫폼 사업자의 책임 구조, ④ 팩트체크 및 외부 검증 생태계의 제도화라는 네 가지 핵심 축으로 재구성될 수 있음을 확인하였다. 이러한 4가지 축은 정책의 '목표 설정 → 제도 설계 → 행위자 의무화 → 외부 검증 및 지속가능성' 으로 이어지는 정책 사슬(policy chain)을 반영한다.

이후 도출된 상위 주제들을 종합하여 정보 무결성 대응 체계의 글로벌 정책 흐름을 해석적으로 통합하였다. 각국의 정책은 허위조작정보를 개별적으로 차단하는 데 그치지 않고, 플랫폼 책임을 구조화하고 민간 팩트체크 생태계를 제도권 내에 편입시키며, 투명성 확보와 지속적인 점검을 통해 장기적으로 작동 가능한 거버넌스를 구축하는 방향으로 진화하고 있음을 확인하였다. 이러한 해석적 통합은 개별 정책의 단순한 병렬적 비교를 넘어, 정보 무결성을 민주주의의 핵심 제도적 인프라로 정착시키려는 국제적 공통 경향을 드러낸다.

2. UN, OECD 정보 무결성 정책 관련 보고서 분석 결과

가. 정책 목표

UN과 OECD의 정보 무결성 관련 권고안은 국가별 규범이나 개별 플랫폼 규제를 넘어, 정보 무결성을 민주주의의 핵심 제도적 인프라로 재정의한다는 점에서 공통된 문제의식을 공유한다. 두 문서는 허위조작정보를 단순히 제거하거나 차단해야 할 '유해 콘텐츠' 로만 보지 않고, 정보가 생산・유통・소비되는 전 과정에서 발생하는 구조적 취약성의 문제로 파악한다. 이에 따라 정책 목표, 제도 설계, 플랫폼 책임, 외부 검증 생태계를 포괄하는 종합적 대응 체계를 제시하고 있다.

먼저 정책 목표 측면에서, UN은 정보 무결성을 표현의 자유, 알 권리, 미디어 다원성 등 인권 규범과 결합된 개념으로 설정한다. 허위조작정보 대응은 민주주의의 질과 사회적 신뢰를 회복하기 위한 수단이지, 국가가 진실을 독점하거나 정보 흐름을 통제하기 위한 도

구가 되어서는 안 된다는 점을 반복적으로 강조한다. OECD 역시 유사하게 정보 무결성을 민주적 거버넌스의 조건으로 규정하면서, 정보조작 위협에 대응하되 법치주의, 비례성, 투명성 원칙을 훼손하지 않는 정책 설계를 핵심 목표로 제시한다.

다만 UN과 OECD는 정보 무결성 정책의 목적 설정에서 역할과 지향점이 구조적으로 구분된다. UN이 정보 무결성을 민주주의・인권・평화와 결부된 규범적 비전으로 제시하며, 국가와 비국가 행위자가 따라야 할 가치를 선언하는 데 중점을 둔다면, OECD는 그러한 규범을 전제로 하여 각국 정부가 실제 정책 설계와 집행 과정에서 활용할 수 있는 행정적・제도적 정책 프레임워크를 제공하는 데 초점을 둔다.

나. 제도적 프레임워크 구성

제도적 프레임워크 구성에서도 이러한 차이가 드러난다. UN은 정보 무결성 강화를 위한 다섯 가지 글로벌 원칙을 제시하고, 이를 기술기업, 광고주, 언론, 연구자, 시민사회, 국가 등 행위자별 '행동 권고(Call to Action)'로 구체화한다. 이는 단일 규제기관 중심의 집행 모델이 아니라, 정보 생태계 전반의 권력 불균형과 책임 분산 구조를 전제로 한 다층적 거버넌스 설계라 할 수 있다. 반면 OECD는 사회적 회복력 강화, 정보원의 투명성・책임성・다원성 강화, 제도적 구조 개선 및 개방형 정부이라는 세 가지 축을 중심으로, 정부 차원의 제도 설계와 정책 조합을 체계화한다. OECD 모델은 UN에 비해 국가 정책 설계자에게 보다 직접적인 실행 지침을 제공하는 성격을 갖는다.

다. 플랫폼 사업자의 책임 구조

플랫폼 사업자의 책임 구조에 대해서는 두 권고안 모두 강한 문제의식을 공유하지만, 접근 방식에는 차이가 있다. UN은 대형 플랫폼 기업이 정보 유통 구조에서 갖는 경제적・기술적 권력을 명시적으로 문제 삼으며, 투명성, 독립적 외부 감사, 알고리즘 및 광고 시스템에 대한 설명 책임, 연구 목적의 데이터 접근 보장을 폭넓게 요구한다. 이는 플랫폼을 단순한 중개자가 아니라 정보 생태계의 핵심 공적 행위자로 위치시키는 접근이다. 반면 OECD는 플랫폼에 대한 직접적 규율보다는, 정부가 법・정책 수단을 통해 플랫폼의 책임성을 유도하도록 설계한다. 플랫폼의 데이터 접근 제공, 추천 시스템 관련 정보 공개, 이용자 선택권 보장 등은 정부가 '요구하거나 장려'해야 할 사항으로 제시되며, 국가의 정책적 개입과 유도를 강조한다.

라. 팩트체크 및 외부 검증 생태계

UN과 OECD 모두 팩트체크를 플랫폼 내부 기능이나 부차적 협력 수단이 아니라, 정보 무결성을 유지하는 독립적 제도 축으로 인식한다. UN은 팩트체킹 단체와 연구자, 언론이 자유롭고 안전하게 활동할 수 있는 환경 조성을 강조하며, 플랫폼에 대해서도 이들과의 협력 및 지원 관계를 투명하게 공개할 것을 요구한다. OECD 또한 독립 저널리즘과 팩트체커의 보호를 명시하고, 플랫폼과의 협력 구조가 공정하고 독립성을 침해하지 않는 방식으로 이루어져야 함을 강조한다. 다만 UN이 팩트체크와 연구 생태계를 글로벌 공공재에 가까운 인프라로 상정하는 데 비해, OECD는 국가 정책 차원에서 이를 안정적으로 지원·연계하는 제도 설계를 중시한다.

3. EU, 호주의 허위조작정보 실천강령 분석 결과

가. 정책 목표

유럽연합(EU)의 허위조작정보 실천강령은 기존의 자율규범(Code of Practice)에서 한 단계 발전하여, 디지털서비스법(DSA) 체계와 결합되는 준규제적 성격의 행동규범(Code of Conduct)으로 진화하였다(이에 이하에서는 EU 실천강령이 아닌 행동규범이라 명하도록 한다). EU 행동규범은 개별 플랫폼의 자율적 노력에 의존하던 초기 단계에서 벗어나, 플랫폼, 광고 산업, 팩트체커, 시민사회, 학계 등 정보 생태계 전반의 참여를 전제로 허위조작정보 확산을 구조적으로 억제하려는 데 목적을 둔다. 특히 EU는 ‘핵심 영역(key areas)’을 중심으로 정책수단을 체계화하고, 각 참여 주체의 이행 현황을 투명하게 공개·점검하는 설계를 통해 실천강령을 지속가능하고 미래 환경 변화에 대응 가능한(fit for purpose & future proof) 거버넌스 프레임으로 구성하고 있다.

반면 호주의 허위조작정보 실천강령은 DIGI라는 비영리 산업협회가 정부의 디지털 플랫폼 정책 로드맵과 ACMA(호주 통신미디어청) 가이드에 대응하여 마련한 전형적인 업계 자율규범이다. 이 강령의 핵심 목적은 허위조작정보로 인한 사회적 위해를 전면적으로 제거하는 데 있기보다는, 민주적 의사결정, 공중보건, 공공안전 등 핵심 영역에서 발생할 수 있는 위해를 완화하기 위한 최소한의 보호장치를 제공하는 데 있다. 이를 위해 호주는 플랫

폼의 규모와 서비스 특성이 상이하다는 점을 전제로, ‘최소 의무 + 선택(opt-in)’ 구조를 채택하여 각 사업자가 자사 서비스의 위험도와 기술적·조직적 역량에 비례한 조치를 취하도록 설계하였다.

EU와 호주의 허위조작정보 대응 규범은 공통적으로 허위조작정보를 단순히 개별 콘텐츠를 삭제하거나 차단하는 문제로 한정하지 않고, 정보 유통을 둘러싼 구조적 요인 전반을 포괄적으로 관리해야 할 정책 과제로 인식한다. 즉, 양 규범 모두 허위조작정보의 확산을 광고 등 시장 유인 구조, 봇·가짜계정과 같은 조작적 행위, 이용자의 정보 선택 환경, 그리고 연구·검증 인프라의 취약성 등으로 구분하여 접근하고 있다는 점에서 공통된 문제 인식을 공유하고 있다 보여진다.

다만 이러한 문제 인식을 제도화하는 방식에서는 뚜렷한 차이가 나타난다. EU의 경우 허위조작정보 행동규범을 디지털서비스법(DSA) 체계와 결합된 준규제적 규범으로 발전시켜 플랫폼의 책임을 강하게 구조화하고 이행 여부를 관리하려 한 반면 호주는 허위조작정보로 인해 발생할 수 있는 위해(harms)를 중심 개념으로 설정하고, 최소한의 안전장치를 마련하되 각 플랫폼의 규모와 특성에 따라 비례적으로(opt-in 방식으로) 조치를 선택할 수 있도록 설계함으로써 민간 자율의 범위를 상대적으로 넓게 유지하고 있다.

나. 제도적 프레임워크 구성

EU 행동규범의 주요 특징은 허위조작정보 대응 수단을 6개 핵심 영역과 3개 거버넌스 장치로 패키지화하였다는 점에 있다. 6개 핵심 영역은 ▲허위조작정보 유통으로부터 발생하는 수익 차단, ▲정치광고의 투명성 확보, ▲조작 행위 억제, ▲이용자 역량 강화, ▲EU 전역 차원의 팩트체크 체계 구축, ▲연구를 위한 데이터 접근 보장으로 구성된다. EU 행동규범은 추가적으로 ▲투명성센터, ▲상설 태스크포스, ▲강화된 모니터링 프레임워크를 거버넌스 장치로 결합함으로써, 행동규범의 이행을 일회성 선언이 아닌 상시적인 관리·점검 체계로 제도화하였다.

<표 3-1> EU 행동규범 6개 핵심영역

영역	세부내용
수익 차단/비금전화	• 허위조작정보와 인접한 광고 게재 회피 • 협력 강화
정치광고의 투명성 확보	• 효율적인 라벨링 • 투명성 의무
조작 행위 억제	• 기존 및 새롭게 등장하는 유형에 대한 대응 • 서명자 간 협력 강화
이용자 역량 강화	• 보다 다양하고 고도화된 이용자 역량 강화 도구 제공 • 신뢰할 수 있는 정보 및 콘텍스트에 대한 접근 제고
EU 전역에 걸친 팩트체크 체계 구축	• 팩트체커 검증 결과의 일관된 활용 • 팩트체커에 대한 공정한 재정적 기여
연구를 위한 데이터 접근 보장	• 플랫폼 데이터에 대한 보다 확대되고 용이한 접근 • 연구 활동 지원

호주의 경우 정책수단을 'Objectives 1~7'이라는 목표 중심 구조로 배열한다. 여기에는 ▲허위조작정보로 인한 위해 방지, ▲광고·수익 유인 차단, ▲서비스 무결성·보완 강화, ▲이용자 역량 강화, ▲정치광고 투명성, ▲전략적 연구 지원, ▲투명성 확보가 포함된다. EU가 제도적 장치와 이행 구조를 강조하는 반면, 호주는 각 목표를 달성하기 위한 다양한 선택 가능한 조치 목록을 제시하는 방식을 취하고 있다는 점에서 차이를 보인다.

<표 3-2> 호주 실천강령 7대 목표

목표	요구 성과
Objective 1. 허위정보 및 허위조작정보로부터 발생할 수 있는 위해(Harms)에 대비한 보호장치 마련	Outcome 1a: 확장 가능한 조치를 통해 허위조작정보 확산으로 인한 위해 발생 위험 감소 Outcome 1b: 이용자가 어떤 행위・콘텐츠가 관리・제한 대상인지 인지 가능 Outcome 1c: 이용자가 정책 위반 콘텐츠・행위를 쉽게 신고 가능 Outcome 1d: 신고에 대한 플랫폼의 대응 조치에 대한 일반 정보 접근 가능 Outcome 1e: 추천시스템 작동 방식에 대한 정보 제공 및 이용자 선택권 보장
Objective 2. 광고 및 수익화 구조 차단	Outcome 2: 허위조작정보와 연계된 광고 및 수익화 유인 감소
Objective 3. 디지털 플랫폼 제공 서비스 및 제품의 무결성과 보안 확보	Outcome 3: 가짜계정, 봇 등 비진정성 행위가 서비스 무결성과 보안을 훼손할 위험 감소
Objective 4. 이용자가 디지털 콘텐츠에 대해 보다 정보에 기반한 선택을 할 수 있도록 역량 강화	Outcome 4: 이용자가 뉴스・사실 정보의 출처를 더 잘 이해하고 허위정보를 식별할 수 있도록 지원
Objective 5. 디지털 플랫폼을 통해 유통되는 정치 광고의 출처에 대한 공적 인식 제고	Outcome 5: 이용자가 정치광고의 출처와 성격을 명확히 인지
Objective 6. 전략적 연구 지원을 통해 허위정보 및 허위조작정보에 대한 공공이 이해 증진	Outcome 6: 독립 연구자의 연구 활동을 지원하여 사회 전반의 이해도 강화
Objective 7. 서명자의 허위정보 및 허위조작정보 대응 조치 공개	Outcome 7: 허위조작정보 대응 조치에 대한 정보를 일반 대중에게 공개

제도적 프레임워크 측면에서 EU와 호주는 모두 허위조작정보 대응을 개별 수단의 나열이 아닌 정책 패키지(policy package) 형태로 구성한다는 공통점을 보인다. 양측 규범에서는 허위조작정보의 수익 차단, 정치광고의 투명성 확보, 조작 행위 억제, 이용자 역량 강화, 연구 및 데이터 접근, 투명성 확보라는 핵심 요소들이 반복적으로 등장하며, 이를 통해 정보 생태계 전반을 다층적으로 관리하려는 접근이 확인된다.

그러나 이 패키지를 조직하는 방식에서는 차이가 있다. EU는 6개 핵심 영역과 3개의 거버넌스 장치를 결합하여, 이행 관리 자체를 하나의 제도적 구조로 명확히 설정한다. 반면

호주는 ‘Objectives 1~7’이라는 목표 중심 구조를 통해 각 목표를 달성하기 위한 다양한 조치들을 제시하되, 개별 플랫폼이 자사 서비스의 위험도와 역량에 따라 선택적으로 이행할 수 있도록 설계하였다. 이로 인해 동일한 정책 목표를 공유하더라도, EU는 상대적으로 표준화된 이행을, 호주는 사업자별 차별화된 이행을 전제로 하는 구조적 차이를 보인다.

다. 플랫폼 사업자의 책임 구조

EU 행동규범은 총 43개의 커밋먼트(commitments)와 128개의 세부 조치를 기반으로 플랫폼의 책임을 비교적 정교하게 설계한다. 플랫폼의 이행 현황은 투명성센터를 통해 공개적으로 확인 가능하며, 상설 태스크포스가 정기적으로 이행 상황을 점검·보완하는 구조를 갖는다. 또한 EU 및 회원국 차원에서 질적·양적 보고가 이루어지는 모니터링 프레임워크를 명시함으로써, 자율규범이라는 형식을 유지하면서도 보고·평가·감사라는 제도적 장치를 실질적 중심축으로 삼고 있다.

이에 비해 호주 실천강령은 모든 서명자에게 공통적으로 적용되는 핵심목표(Objective 1 및 투명성 보고 의무)를 설정한 뒤, 나머지 조치는 opt-in 방식으로 선택하도록 하여 의무의 강도가 상대적으로 유연하다. 특히 호주는 표현의 자유에 대한 우려를 반영하여, 위법하지 않은 콘텐츠를 단지 ‘허위’라는 이유만으로 제거하도록 강제하지 않는다는 원칙을 규범 내부에 명시하고 있다. 다만 DIGI가 관리 주체로서 보고·공개·점검이 이루어지도록 하고, 독립적인 민원 처리기구와 연례 보고서에 대한 외부 전문가의 사실확인(attestation) 절차를 두어 최소한의 책임성과 신뢰성을 확보하고 있다.

플랫폼 책임 구조와 관련하여 EU와 호주는 모두 자율규범의 형식을 유지하면서도, 보고·공개 및 모니터링을 책임성 확보의 핵심 장치로 활용한다는 공통점을 가진다. 즉, 플랫폼의 자율적 참여를 전제로 하되, 그 이행 여부를 사회적으로 검증 가능하게 만드는 장치를 마련해놓았다는 점에서 유사한 접근을 취한다.

그러나 플랫폼 책임을 점검하는 밀도에는 차이가 있다. EU 행동규범은 다수의 커밋먼트와 세부 조치를 통해 플랫폼의 의무를 세밀하게 규정하고, 투명성센터와 상설 태스크포스를 통해 이행 현황을 상시적으로 공개·점검하는 거버넌스를 구축한다. 반면 호주 실천강령은 표현의 자유 보호에 대한 우려를 규범 내부에 명시적으로 반영하여, 위법하지 않은 콘텐츠를 단지 ‘허위’라는 이유만으로 제거하도록 강제하지 않는다는 원칙을 분명히 한

다. 대신 DIGI가 관리 주체로서 보고와 공개를 운영하고, 독립적인 민원 처리 절차와 외부 전문가의 사실확인(attestation)을 통해 최소한의 책임성과 신뢰성을 확보한다.

라. 팩트체크 및 외부 검증 생태계

EU 행동규범은 팩트체크를 하나의 독립된 정책 축으로 명확히 분리하여, EU 전역 차원의 팩트체크 커버리지 확대, 플랫폼 서비스 내 팩트체커와의 구조적 협업, 팩트체크 업무 수행에 필요한 정보 및 데이터 접근 보장을 핵심 영역으로 설정한다. 이는 팩트체크를 단순한 외부 협력 수단이 아니라, 플랫폼 생태계 전반의 상시적 안전장치로 제도화하려는 접근으로 평가할 수 있다.

반면 호주 실천강령은 팩트체크를 주로 플랫폼의 대응 조치 중 하나로 위치시킨다. 구체적으로는 콘텐츠 검토 과정에서 독립 팩트체킹 단체와의 파트너십을 예시로 제시하고, 이용자 역량 강화(Objective 4) 차원에서 팩트체킹 단체와의 지속가능한 협력 및 재정 지원을 명시하며, 연구 지원(Objective 6)을 통해 독립 팩트체크 이니셔티브와의 공동 연구, 데이터셋 공유, 연례 행사 개최 등을 제안한다. 종합하면, EU는 팩트체크 커뮤니티 자체를 강화하는 제도적 기반 구축에 방점을 두는 반면, 호주는 플랫폼이 선택적으로 활용할 수 있는 협력·지원 수단으로 팩트체크를 배치하는 경향이 강하다고 정리할 수 있다.

팩트체크 및 외부 검증과 관련해서 EU와 호주는 모두 이를 플랫폼 내부 기능에 한정하지 않고, 외부의 독립적 행위자와의 협력을 전제로 한다는 점에서 공통된 방향성을 보인다. 즉 팩트체크는 플랫폼이 단독으로 수행하는 기능이 아니라, 정보 생태계 전반의 신뢰성을 지탱하는 외부 검증 장치로 인식된다.

다만 그 제도적 위상은 상이하다. EU는 팩트체크를 하나의 핵심 영역으로 분리하여, EU 전역 차원의 팩트체크 커뮤니티를 강화하고 이를 플랫폼 거버넌스의 상시적 안전장치로 제도화하려는 경향이 강하다. 이에 비해 호주는 팩트체크를 플랫폼이 선택적으로 활용할 수 있는 대응 조치 패키지 중 하나로 배치하고, 협력이나 재정 지원 역시 자율적 선택에 맡기는 구조를 취한다. 이러한 차이는 EU가 팩트체크 생태계 자체를 제도적 인프라로 구축하려는 반면, 호주는 플랫폼의 자율적 판단과 책임 이행을 보조하는 수단으로 팩트체크를 위치시키고 있음을 보여준다.

4. 우리나라 정보 무결성 정책 수립을 위한 주요 시사점

EU, 호주, UN, OECD의 정보 무결성 관련 규범과 정책을 종합적으로 분석한 결과, 국제사회의 허위조작정보 대응은 개별 콘텐츠의 진위 여부를 판단하고 사후적으로 차단하는 방식의 규율을 넘어, 정보가 생성・유통・소비되는 전반적인 환경의 신뢰성을 확보하려는 방향으로 전환되고 있음을 확인할 수 있다. 즉, 허위정보 문제는 더 이상 콘텐츠 관리만의 문제가 아니라, 민주적 의사결정 구조와 공적 담론의 질을 좌우하는 제도적 과제로 인식되고 있으며, 이에 따라 '정보 무결성(Information Integrity)'이라는 보다 포괄적인 정책 프레임이 형성되고 있는 것으로 보인다. 이러한 국제적 흐름은 우리나라의 정책 논의 역시 허위정보 규제 중심의 접근에서 벗어나 정보 무결성 거버넌스를 구축하는 방향으로 전환될 필요성을 시사한다.

특히 UN과 OECD는 정보 무결성 정책의 최상위 지향점으로 민주주의의 안정성과 기본권 보호를 명시하고 있으며, EU와 호주 또한 표현의 자유와 국민의 알 권리를 침해하지 않는 범위 내에서 정책 수단을 설계하고 있다는 점에서 공통된 인식을 공유하고 있다. 그에 따라 국제 규범들은 단순히 '허위성'만을 이유로 한 개입이 아니라, 원칙적으로 위법성에 기초한 개입을 요구하는 방향을 취하고 있다. 이러한 맥락에서 해외 정보 무결성 정책 사례는 허위정보의 억제 자체를 목적화하기보다, 표현의 자유를 존중하면서도 공적 담론의 신뢰성을 어떻게 제도적으로 보호할 것인가에 초점을 두어 설계되어야 함을 보여준다.

또한 해외 사례들은 국가가 직접 콘텐츠의 진위를 판단하거나 규제 집행의 전면에 나서는 방식보다는, 민간 플랫폼의 책임을 구조화하는 공동 규제(co-regulation) 모델을 채택하고 있다는 점에서 중요한 시사점을 제공한다. EU의 경우 디지털서비스법(DSA)과 결합된 행동규범(Code of Conduct)을 통해 플랫폼의 자율적 이행을 전제로 하되, 구체적인 커밋먼트, 성과 지표, 보고 및 점검 체계를 제도화함으로써 실질적인 책임을 부과하고 있다. 호주 역시 업계 자율규범의 형태를 유지하면서, 정부 가이드와 독립적 검증 장치를 통해 최소한의 책임성과 신뢰성을 확보하는 구조를 취하고 있다. 이는 정보 무결성 정책에서 국가의 역할이 규범의 설계자이자 감독자로 재정의되고 있음을 의미한다.

플랫폼 사업자 책임의 내용 또한 국제적으로 변화하고 있다. 허위정보 대응 책임은 더 이상 개별 콘텐츠 삭제 여부에 국한되지 않고, 수익 구조, 알고리즘 설계, 추천 시스템의

투명성, 조작 행위에 대한 대응 등 플랫폼의 시스템 전반으로 확장되고 있다. EU 행동규범이 수익 차단, 조작 행위 억제, 데이터 접근 보장 등을 핵심 영역으로 설정한 것은 이러한 구조적 책임 전환을 단적으로 보여준다. 한국 역시 플랫폼 책임을 사후적 조치 중심이 아니라, 사전적 위험 관리와 시스템 설계 책임 중심으로 재구성할 필요가 있다.

아울러 국제적 논의에서 팩트체크는 더 이상 부차적인 보조 수단이 아니라, 정보 무결성 거버넌스의 핵심 인프라로 자리 잡고 있다. EU는 팩트체크를 독립된 정책 축으로 분리하여 안정적 재정 기여, 데이터 접근 보장, 플랫폼 내 검증 결과의 활용을 제도화하고 있으며, 호주 역시 선택적 구조를 유지하되 독립성과 지속 가능성을 강조하고 있다. 이는 팩트체크 생태계가 정부나 플랫폼으로부터 독립성을 유지하면서도 제도적 지원을 받을 수 있는 구조가 필요하다는 점을 시사한다. 한국의 정책 설계에서도 팩트체크를 플랫폼의 자율적 협력 대상으로만 설정하기보다, 정보 무결성 정책의 제도적 구성 요소로 명확히 위치시킬 필요가 있다.

마지막으로, EU와 호주 사례는 허위정보 대응의 실효성을 담보하는 핵심 장치로 투명성과 지속적인 점검을 강조하고 있다. 정기적인 보고, 공개된 이행 현황, 외부 전문가에 의한 검증은 자율규범이 선언적 수준에 머무르지 않도록 하는 필수 요소로 기능한다. 이는 처벌 중심의 규제보다 투명성과 책임성을 통해 장기적으로 작동 가능한 거버넌스를 구축하려는 접근으로, 한국의 정보 무결성 정책 역시 단기적 규제 효과보다 지속 가능성과 신뢰 확보를 목표로 설계될 필요가 있음을 시사한다.

이와 같은 시사점들을 살펴볼 때, 향후 허위조작정보 대응 정책은 정보 무결성이라는 상위 개념 아래에서 플랫폼의 구조적 책임, 독립적 팩트체크 생태계, 투명성과 점검을 결합한 공동 규제 모델로 재구성될 필요가 있다. 이러한 방향 전환은 국제 규범과의 정합성을 높이는 동시에, 국내 정책의 정당성과 사회적 수용성을 강화하는 데 기여할 것으로 평가된다.

제4장. 정보 무결성 정책 관련 인식조사

제1절 정보 무결성 정책 관련 전문가 인식

1. 전문가 조사 개요

o **(조사 목적)** 본 조사는 정보 무결성 정책에 대한 전문가들의 인식 수준을 체계적으로 파악하고, 현행 제도가 지니는 한계와 문제점을 진단함과 동시에 향후 정책 설계 및 제도 개선을 위한 시사점을 도출하는 것을 목적으로 실시하였다. 특히 허위조작정보 대응과 관련하여 자율규제, 플랫폼 책임, 팩트체크 체계 등 주요 쟁점에 대해 분야별 전문가들의 평가와 인식을 종합적으로 살펴보고자 하였다.

o **(조사 기간)** ’ 25. 11. 14.(금) ~ 12. 5.(금)

o **(조사 대상 및 구성)** 총 8명(학계 4명, 민간 팩트체크 기관 2명, 플랫폼 1명, 언론사 1명)

o **(조사 방법)** 정량(리커드 척도) 및 정성(서술형) 문항을 포함한 서면조사 실시

2. 전문가 조사 결과

가. 조사 결과 요약

o **(정보 무결성 강화 정책 필요성)** 전문가 조사 결과를 보면, 디지털 플랫폼을 통한 허위조작정보의 부정적 영향(A-1, 평균 4.38)에 대해 매우 높은 공감대가 형성되어 있는 것으로 나타났다. 이는 허위조작정보가 사회적 신뢰와 민주적 의사소통을 실질적으로 훼손하고 있으며, 이를 구조적 위험 요인으로 인식하고 있음을 보여준다.

- 반대로 현행 국내 제도의 대응력(A-2, 평균 2.75)은 3점 미만으로 평가되어, 전문가들은 현재 제도가 급변하는 디지털 환경에 충분히 대응하지 못하고 있다고 보고 있는 것으로 나타났다.

- 이러한 인식은 국가 차원의 정보 무결성 정책 마련 필요성(A-3, 평균 4.50)에 대한 매우 높은 동의로 이어졌다. 전문가들은 기존의 ‘허위정보 규제‘나 ‘플랫폼 자율규제‘만으로는 한계가 있다고 보고, 정보의 신뢰성과 공적 담론의 품질을 체계적으로

보장할 수 있는 새로운 정책 프레임 구축의 필요성에 공감한 것으로 해석된다.

o **(정보 무결성 강화 정책 방향)** 구체적 정책수단 측면에서, 위협 인식 공유체계 구축(B-1, 평균 3.75), AI 생성물 라벨・워터마크 의무화(B-2, 평균 4.38), 이용자 신고・토론 등 참여 메커니즘 표준화(B-3, 평균 4.00), 언론・팩트체커 보호 지원제도(B-4, 평균 4.50), 플랫폼 투명성・협력 의무 법제화(B-5, 평균 4.63), 공익콘텐츠・팩트체크에 대한 재정지원 제도화(B-6, 평균 4.13) 모두 4점 안팎으로, 필요성을 높게 인식하고 있는 것으로 나타났다.

- 다만, Trusted Flagger(사실확인 단체) 지정제(B-7, 평균 3.13)는 유일하게 보통 수준에 가까운 점수로 나타나 다른 항목보다 동의 수준이 낮고, 의견 분산도 큰 편으로 나타났다. 이는 제도의 필요성은 일정 부분 인정하지만, 지정 기준・독립성・권한 남용 가능성 등에 대한 우려가 함께 존재하는 것으로 해석할 수 있다.

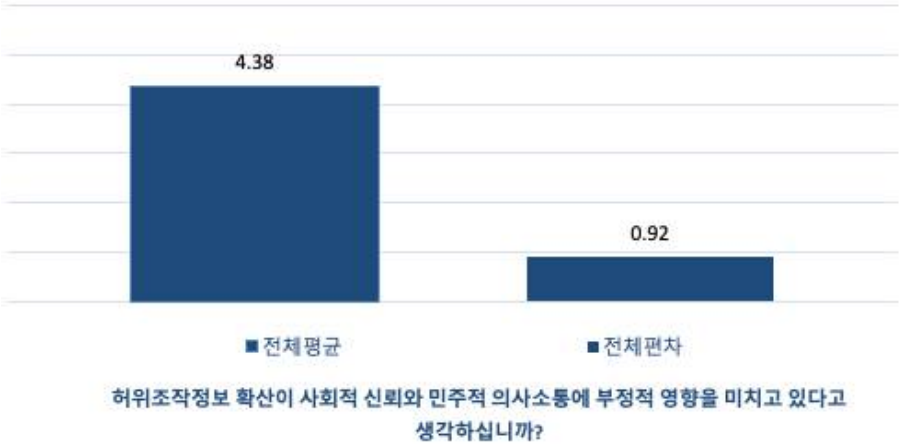
<표 4-1> 정보 무결성 정책 관련 전문가 인식 조사 결과(종합)

문항	문항 내용	평균	표준편차
A-1	디지털 플랫폼을 통한 허위조작정보 확산의 부정적 영향력	4.38	0.92
A-2	우리나라 정보 무결성 관련 제도의 환경 대응력	2.75	1.16
A-3	국가 차원의 정보 무결성 정책 마련 필요성	4.00	1.20
B-1	허위저작정보 위협 인식 공유 체계 구축 필요성	3.75	1.39
B-2	AI 생성물 식별(라벨/워터마크) 법제도화 필요성	4.38	0.74
B-3	플랫폼 내 이용자 참여 메커니즘 표준화 필요성	4.00	0.76
B-4	언론 및 팩트체커 보호 지원제도 필요성	4.50	0.76
B-5	플랫폼의 콘텐츠 모더레이션 투명성 및 협력 의무 법제화 필요성	4.63	0.74
B-6	공익콘텐츠 제작 및 팩트체크 활동에 대한 정부 및 플랫폼 재정 참여제도(공익펀드, 인센티브 등) 마련 필요성	4.13	0.99
B-7	사실확인 단체(Trusted Flagger) 지정제 도입 필요성	3.13	0.99

나. 각 문항별 결과 분석 및 전문가 의견

o **(디지털 플랫폼을 통한 허위조작정보 확산의 부정적 영향력)** 평균 4.38, 표준편차 0.92으로 전반적으로 매우 높은 공감대가 나타났으며 전문가 간 의견 차이는 비교적 적은 편이다.

[그림 4-1] 허위조작정보 확산의 부정적 영향력



- 전문가들은 디지털 플랫폼을 통한 허위조작정보 확산이 사회적 신뢰를 훼손하고 민주적 의사소통을 왜곡·약화시키는 주요 요인으로 작용하고 있다는 점에 공감하였다. 특히 알고리즘 기반 추천 구조가 확증편향과 필터버블을 강화하고, 개인 미디어와 극단적 정치 콘텐츠의 확산을 통해 여론 왜곡과 사회적 양극화를 심화시키고 있다는 지적이 반복적으로 제기되었다.

- 허위조작정보는 단순한 정보 오류 차원을 넘어, 검증된 사실에 대한 사회적 합의를 붕괴시키고 숙의·협의·타협에 기반한 민주주의의 작동 조건 자체를 위협한다는 인식도 확인었다. 이 과정에서 언론 역시 조회수 중심의 수익 구조로 인해 충분한 검증 없이 오정보를 재확산하는 데 기여하고 있다는 비판이 제기되었다.

- 반면 일부 전문가는 허위조작정보의 부정적 영향을 과도하게 일반화하거나 절대적 위협으로 규정하는 데에는 신중할 필요가 있다고 지적한다. 이용자의 미디어 리터러시 향상, 집단지성을 통한 자정 기능 등 긍정적 요인도 동시에 작동하고 있으며, 허위정

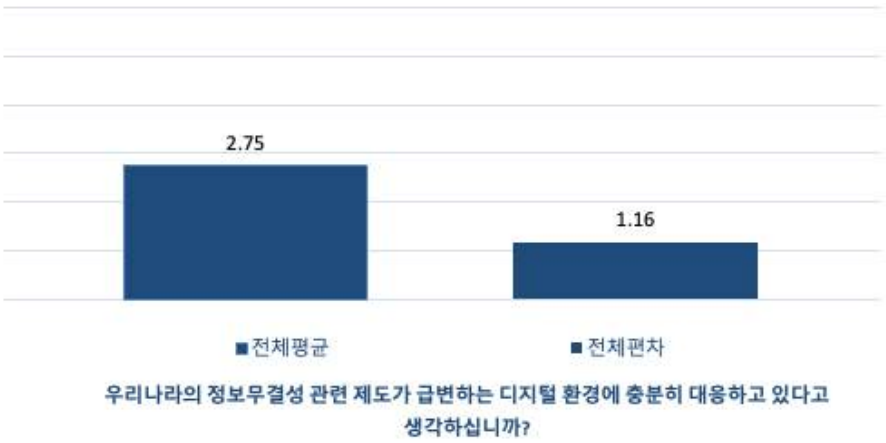
보 논쟁 자체가 사회적 관심과 참여를 촉진하는 측면도 존재한다는 점을 강조하였다.

- 또한 허위조작정보의 개념과 판단 기준이 모호한 상황에서 허위조작정보에 대한 규제 중심의 접근은 민주주의 원칙과 긴장 관계를 형성할 수 있으며, 보다 정교하고 균형 잡힌 정책 설계가 필요하다는 의견이 나타났다.

- 종합하면, 전문가들은 허위조작정보가 민주적 의사소통에 실질적 위협이 되고 있다는 점에는 대체로 동의하면서도, 위협의 정도와 대응 방식에 대해서는 규제 중심 접근과 표현의 자유 보호 간 균형이 중요하다는 인식을 공유하고 있는 것으로 나타났다.

o **(우리나라 정보 무결성 관련 제도의 환경 대응력)** 평균 2.75, 표준편차 1.16으로 전문가들은 현행 제도가 충분하지 않다고 인식하는 것으로 나타났다.

[그림 4-2] 정보 무결성 제도의 환경 대응력



- 전문가들은 현재 우리나라의 정보 무결성 관련 제도가 급변하는 디지털 환경에 전반적으로 충분히 대응하지 못하고 있다는 데 공감하였다. 다수의 의견에 따르면, 현행 법제도는 신문·방송 등 레거시 미디어 중심으로 설계되어 있어 유튜브·SNS·메신저 등 플랫폼 기반 정보 유통 구조와 알고리즘 작동 방식, 생성형 AI와 딥페이크 확산을 적절히 반영하지 못하고 있다는 지적이 제기되었다. 특히 AI 기반 합성 콘텐츠에 대해서는 식별 기준과 책임 주체, 처벌 규정이 불명확해 규제 사각지대가 확대되고 있

다는 평가가 많았다.

- 플랫폼 자율규제와 언론 책임 제도 역시 실질적 성과가 제한적이라는 비판이 제기되었다. 일부 전문가들은 국내 플랫폼이 팩트체크 코너 운영이나 외부 기관과의 협업 등 자율규제 노력을 시도해 왔다는 점은 인정하면서도, 이러한 시도가 지속되지 못하고 중단되거나 제도화로 이어지지 못한 점을 한계로 지적하였다. 또한 언론이 플랫폼 트래픽 의존도를 높이면서 클릭 경쟁에 몰두하고, 이 과정에서 허위·자극적 정보가 재확산되는 구조가 형성되어 신뢰 가능한 정보 생산 주체로서의 언론 역할이 약화되고 있다는 우려도 제기되었다.

- 아울러, 허위조작정보에 대한 사회적 대응 체계와 거버넌스의 부재가 반복적으로 지적되었다. 극단적 정치 유튜브, 사이버렉카, 혐오·차별 콘텐츠가 사회적 피해를 야기하고 있음에도 불구하고, 플랫폼 차원의 신속한 차단이나 공적 심의·조정 메커니즘은 실효적으로 작동하지 못하고 있다는 평가가 많았다. 방송미디어통신심의위원회 심의 구조 역시 정치적 논란에 취약해 제도적 신뢰성이 충분히 확보되지 못하고 있다는 문제의식이 제시되었다.

- 국제 비교 관점에서 볼 때에도, EU의 디지털서비스법(DSA), 호주·영국의 온라인 안전법 등 주요국은 플랫폼 투명성 의무 강화, 알고리즘 책임, 공동규제 모델 정교화 등 보다 적극적인 정책 전환을 추진하고 있는 반면, 국내는 정치권 주도의 단편적이고 성급한 입법 논의가 반복되면서 법체계의 정합성과 일관성이 부족하다는 평가가 제기되었다.

- 일부 전문가는 규제 중심 접근의 한계를 지적하며, 교육을 통한 미디어 리터러시 강화, 플랫폼과 언론에 대한 인센티브 기반 유인 구조, 사회적 신뢰 회복을 위한 장기적 전략의 필요성을 강조하였다. 즉, 제재와 처벌만으로는 정보 무결성 문제를 해결하기 어렵고, 플랫폼 사업자의 공적 책임 강화와 언론 책무의 명확화가 병행되어야 한다는 인식이다.

- 종합하면, 전문가들은 현행 정보 무결성 제도가 디지털 플랫폼 중심 환경과 기술 변화 속도를 따라가지 못하고 있으며, 정부규제와 플랫폼 자율규제, 언론사 책임 모두에서 구조적 공백과 실행력 부족이 존재한다고 평가하였다. 이에 따라 단편적인 법 제개정

보다는 플랫폼 거버넌스 전반을 재설계하는 중장기적 정책 프레임 전환이 필요하다는 공감대가 형성되어 있다고 볼 수 있다.

o **(국가 차원의 정보 무결성 정책 마련 필요성)** 평균 4.00, 표준편차 1.20으로 전반적으로 필요성을 강하게 인식하는 것으로 나타났으며, 국가 주도의 새로운 정책 프레임 마련 필요성에 대한 공감대가 확인되었다.

[그림 4-3] 국가 차원의 정보 무결성 정책 마련 필요성

- 전문가 의견을 종합하면, 디지털 플랫폼 환경에서 허위조작정보의 사회적 파급력이 확대됨에 따라 기존의 플랫폼 자율규제만으로는 한계가 분명하다는 인식과, 동시에 국가 주도의 직접적 규제가 표현의 자유를 침해할 위험이 있다는 우려가 병존하는 것으로 나타났다.

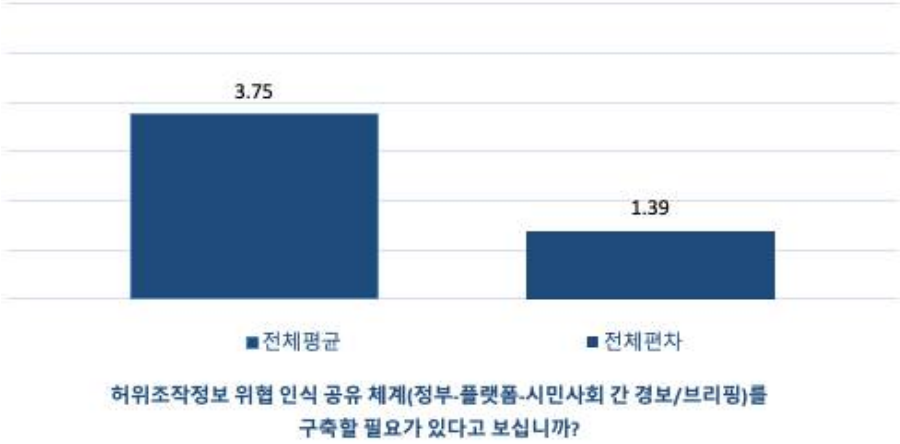
- 우선 다수의 전문가는 허위조작정보가 더 이상 개별 기업의 자율적 판단이나 선의에 맡길 수 있는 수준을 넘어섰다고 평가하였다. 정보 환경이 민주주의를 지탱하는 핵심 공공 인프라로 기능하는 상황에서, 플랫폼 기업은 본질적으로 수익 극대화를 목표로 운영되므로 공익적 책임을 기대하기 어렵기 때문에 국가 차원의 최소한의 안전장치와 표준화된 정책 프레임워크가 필요하다는 의견이 제기되었다. 특히 생성형 AI 확산으로 허위정보 생산·유통 비용이 급격히 낮아진 점을 고려할 때, 정부·기업·언론·

시민사회가 함께 참여하는 국가적 거버넌스 수준의 대응 체계가 필수적이라는 주장도 다수 확인되었다.

- 반면, 상당수 전문가는 국가 주도의 규제가 허위조작정보 정의의 불명확성, 정치적 중립성 문제, 표현의 자유 침해 가능성이라는 중대한 위험을 동반한다고 지적하였다. 표현의 자유는 오류·과장·불완전한 주장까지 포함하는 개념이며, 정부가 특정 정보를 허위로 규정하고 제재하는 순간 정치적 논란과 권력 남용 가능성이 커질 수 있다는 것이다. 특히 허위정보의 정의와 기준이 명확하지 않은 상황에서 강력한 국가 규제는 민주적 공론장을 위축시킬 수 있다는 우려가 제기되었다.
- 이에 따라 일부 전문가는 규제 중심 접근보다는 미디어 리터러시 교육 강화, 언론의 팩트체크 기능 지원, 신뢰 가능한 정보원에 대한 접근성 보장을 정보 무결성 정책의 핵심 축으로 삼아야 한다고 강조하였다. 또한 규제와 처벌보다는, 책임 있는 정보 유통과 검증 활동을 수행하는 주체에게 인센티브를 부여하는 균형 잡힌 정책 설계가 필요하다는 의견도 제시되었다.
- 종합하면, 전문가들은 허위조작정보 대응을 위해 사회적 차원의 정책적 개입 필요성에는 공감하면서도, 이를 단순한 정부 규제나 제재로 환원하는 데에는 신중한 입장을 보였다. 향후 정보 무결성 정책은 정부 주도의 일방적 규제가 아니라, 정부·플랫폼·언론·시민사회가 역할을 분담하는 다층적 거버넌스 체계를 중심으로, 표현의 자유 보호와 사회적 신뢰 회복을 동시에 달성할 수 있는 방향으로 설계될 필요가 있다는데 함의가 모여진다.

o **(허위조작정보 위협 인식 공유 체계 구축 필요성)** 평균 3.75, 표준편차 1.39로 대체적으로 긍정적인 평가인 것으로 나타났다.

[그림 4-4] 허위조작정보 위협 인식 공유 체계 구축 필요성

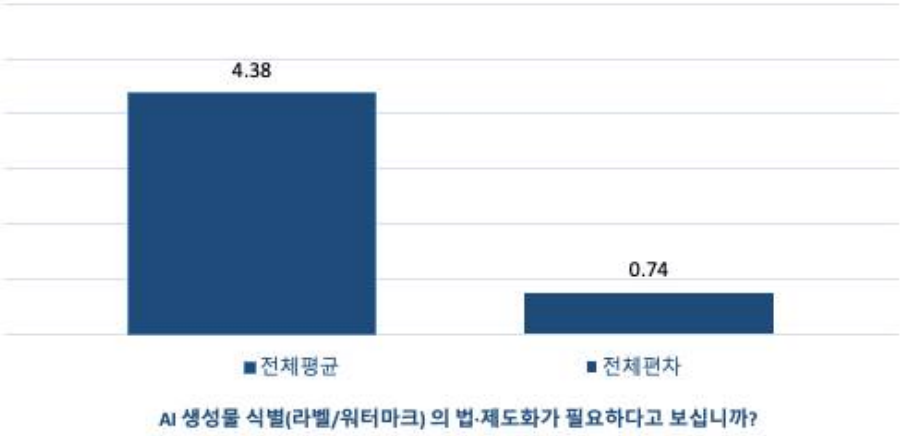


- 전문가 의견을 종합하면, 디지털 환경에서 허위조작정보의 확산 속도와 파급력을 고려할 때 정부·플랫폼·시민사회 간 위협 인식 공유 체계의 필요성 자체에 대해서는 대체로 공감대가 형성되어 있는 것으로 나타난다. 개별 주체가 단편적으로 대응하는 방식으로는 정보 확산의 속도를 따라가기 어렵고, 위협 인지·분석·대응에 필요한 정보와 자원을 상호 공유하는 협력 네트워크가 필요하다는 인식이 공통적으로 제기되었다.
- 이에 따라 다수의 전문가는 위협 인식 공유 체계가 정부 단독이 아닌 공동규제 또는 협력 거버넌스 형태로 설계되어야 하며, 시민사회와 독립적 전문가의 참여를 제도적으로 보장함으로써 투명성과 공정성을 확보할 필요가 있다고 보았다. 일부 의견에서는 정부로부터 독립된 위원회나 기구가 허위조작정보 관련 판단과 대응을 담당하고, 정부는 법제도적 기반과 재정적 지원을 통해 이를 뒷받침하는 역할에 머무르는 것이 바람직하다는 제안도 제시되었다.
- 아울러 위협 인식 공유 체계의 실효성을 확보하기 위해서는 참여 주체 간 신뢰 구축, 협력 원칙의 명확화, 그리고 협력 성과에 대한 평가와 보상 구조가 함께 설계되어야 한다는 점도 강조되었다. 단순한 정보 공유를 넘어, 자율적 규범 준수와 책임 있는 참여가 유도될 수 있도록 인센티브 기반의 구조를 마련할 필요가 있다는 것이다.
- 종합하면, 허위조작정보 위협 인식 공유 체계는 필요성이 인정되지만, 그 설계와 운영

방식에 있어 정부의 역할과 권한을 신중히 설정하고, 독립성·투명성·다자 참여를 핵심 원칙으로 하는 거버넌스 구조가 전제되어야 한다는 점이 주요 시사점으로 도출된다.

o **(AI 생성물 식별(라벨/워터마크) 법제도화 필요성)** 평균 4.38, 표준편차 0.74로 법제도적 대응 필요성을 전문가 대부분이 인정하는 것으로 나타났다.

[그림 4-5] AI 생성물 식별 법제도화 필요성



- 전문가 다수는 AI 생성물 식별 제도의 도입 필요성에 대해 전반적으로 공감하고 있으며, 이는 이용자의 알 권리 보장과 정보 유통의 투명성 확보를 위한 최소한의 안전장치라는 인식에 기반하고 있다. 특히 딥페이크, 사칭 광고, 가짜 뉴스 이미지 등 AI 기술이 허위조작정보 생산과 범죄에 악용되는 사례가 증가함에 따라, 인간이 제작한 콘텐츠와 AI 생성물을 구별할 수 있는 제도적 장치의 필요성이 강조되었다.
- 다만 전문가들은 현 시점에서의 기술적 한계 또한 분명히 지적하고 있다. 워터마크나 라벨링 기술은 비교적 쉽게 제거·변조될 수 있으며, 오픈소스 기반 AI나 해외에서 생성된 콘텐츠까지 포괄적으로 통제하기에는 현실적 제약이 크다는 점에서, 모든 AI 생성물에 일률적이고 강력한 의무를 부과하는 방식에는 신중할 필요가 있다는 의견이 제시되었다. 제도 설계 시 기술적 실효성과 집행 가능성을 함께 고려해야 한다는 공감대가 형성되어 있음을 알 수 있다.

- 또한 일부 전문가들은 AI 생성물 식별을 단순한 규제 수단이 아니라, 사회적 신뢰를 구축하는 메커니즘으로 접근할 필요가 있다고 보았다. 즉, 표시 의무 위반에 대한 제재 중심의 접근보다는, 라벨·워터마크를 부착한 콘텐츠에 대해 신뢰도 제고, 유통상 혜택, 인센티브를 제공하는 방식이 장기적으로 더 효과적일 수 있다는 제안이다.

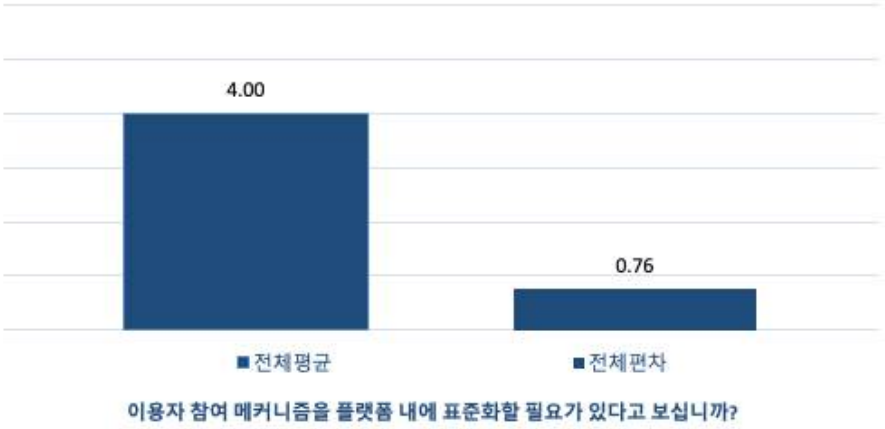
- 이와 함께 AI 활용 정도가 다양한 현실을 반영하여, ‘AI 전면 생성’, ‘AI 보조 활용’, ‘AI 검색 활용’ 등 식별 수준을 단계화하는 유연한 표기 체계의 필요성도 제기되었다.

- 정책 적용 범위와 관련해서는 맥락 기반 접근의 필요성이 강조되었다. 선거, 뉴스 보도, 건강, 재정 정보 등 사회적 파급력이 크고 신뢰성이 특히 요구되는 영역에서는 AI 생성물 식별을 보다 강하게 의무화할 필요가 있는 반면, 그 외 영역에서는 메타데이터 태깅이나 자율적 라벨링 등 완화된 방식이 적절하다는 의견이 다수 제시되었다. 특히 글로벌 플랫폼 환경을 고려할 때, 콘텐츠 생산 단계보다는 유통 플랫폼에 대한 규율을 통해 국내 유통 기준을 설정하는 방식이 현실적인 대안으로 제시되었다.

- 종합하면, AI 생성물 식별 제도는 허위조작정보 대응과 정보 무결성 강화를 위한 중요한 정책 수단으로 인식되고 있으나, 기술적 한계와 표현의 자유, 시장 현실을 함께 고려한 단계적·차등적 제도 설계가 필요하다는 점에서 전문가 의견이 수렴되고 있다. 즉, 전면적 규제보다는 최소 기준의 설정, 고위험 영역 중심의 의무화, 플랫폼 책임 강화, 자율적 참여를 유도하는 혼합적 정책 접근이 바람직한 방향으로 제시된다.

o **(플랫폼 내 이용자 참여 메커니즘 표준화 필요성)** 평균 4.00, 표준편차 0.76로 이 항목에 대해서도 전문가 대다수가 필요성에 동의하는 것으로 나타났다.

[그림 4-6] 플랫폼 내 이용자 참여 메커니즘 표준화 필요성



- 전문가들은 디지털 플랫폼 환경에서 허위조작정보 대응을 위해 이용자 참여 메커니즘이 필수적이라는 점에는 대체로 공감하고 있다. 특히 이용자가 허위조작정보를 발견했을 때 신속하게 신고하고, 그 처리 결과를 확인할 수 있는 구조는 초기 확산을 차단하고 플랫폼 대응의 실효성을 높이는 핵심 요소로 인식된다. 현재 플랫폼별로 신고 기준과 절차가 상이하고 가시성이 낮아 이용자 혼란을 초래하고 있다는 점에서, 일정 수준의 체계 정비 필요성도 제기되었다.

- 표준화된 이용자 참여 메커니즘은 플랫폼의 자의적 콘텐츠 관리에 대한 불신을 완화하고 공정성을 제고하는 수단으로도 평가되었다. 신고 시스템을 플랫폼 전면에 명확히 노출하거나, 중앙선거관리위원회·방송미디어통신심의위원회 등 기존 공적 신고 체계와 연계하는 방안, 또는 팩트체크 전문기관과의 협업 모델을 제도적으로 유도하는 방안 등이 대안으로 제시되었다. 이러한 접근은 이용자 참여를 단순한 신고 기능을 넘어, 정보 무결성을 강화하는 거버넌스 장치로 확장하는 의미를 가진다.

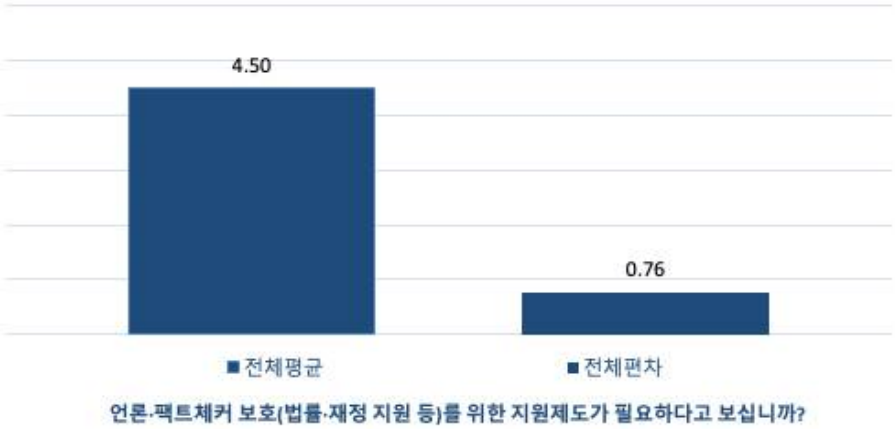
- 다만 '표준화'의 범위와 방식에 대해서는, 다수 전문가들이 모든 플랫폼에 동일한 UI·UX나 세부 절차를 강제하는 방식은 플랫폼의 서비스 특성과 혁신을 저해할 수 있다고 지적한다. 이에 따라 신고 기능의 접근성 확보, 처리 절차의 투명성, 결과 통보 등 핵심 원칙은 공통 기준으로 설정하되, 구체적인 구현 방식은 플랫폼 유형과 규모, 서비스 성격에 따라 자율성을 부여하는 방식이 바람직하다는 의견이 다수 제시되었

다.

- 종합하면, 이용자 참여 메커니즘의 표준화는 허위조작정보 대응을 위한 중요한 정책 수단으로 인식되지만, 전면적·획일적 표준화보다는 최소 기준 중심의 프레임워크 설정과 플랫폼 자율 규범을 결합한 혼합적 접근이 적절하다는 점에서 전문가 의견이 수렴된다. 즉, 국가 또는 규제기관은 기본 원칙과 방향을 제시하고, 플랫폼 사업자는 이에 부합하는 자체 규정과 시스템을 구축하며, 공적 기구는 그 이행 여부를 점검·평가하는 구조가 현실적이고 지속가능한 정책 설계 방향으로 제시된다.

o **(언론 및 팩트체커 보호 지원제도 필요성)** 평균 4.50, 표준편차 0.76로 투명성 및 협력 의무와 함께 가장 높은 수준의 필요성 인식이 나타난 문항이다. ' 팩트체크 활동' 과 ' 언론의 공공적 기능 보호' 가 정보 무결성 확보의 핵심 요소로 인식되고 있음을 알 수 있다.

[그림 4-7] 언론 및 팩트체크 보호 지원제도 필요성



- 전문가들은 허위조작정보 확산이 심화되는 디지털 환경에서 언론과 팩트체커의 사실 검증 기능이 사회적 신뢰와 민주주의를 지탱하는 핵심 인프라로서 중요성이 커지고 있다는 점에 대체로 공감하고 있다. 특히 팩트체크 활동은 정보 무결성을 유지하는 데 필수적인 공공재적 성격을 가지지만, 높은 인력·시간·비용이 소요되는 반면 시장에서 충분한 보상을 받기 어려워 구조적으로 취약하다는 점이 공통적으로 지적되었다.

실제로 팩트체킹은 일반 기사 대비 수배 이상의 자원이 투입되는 고비용 활동으로, 개별 기자나 언론사가 지속적으로 수행하기에는 한계가 있다는 인식이 다수 제시되었다.

- 또한 허위조작정보가 정치·사회적 갈등과 결합되면서, 언론인과 팩트체커가 집단적 괴롭힘, 협박, 물리적 위협, 전략적 봉쇄 소송(SLAPP) 등에 노출되는 사례가 증가하고 있다는 점도 중요한 문제로 언급되었다. 이러한 환경에서는 검증 활동 자체가 위축될 수밖에 없으며, 이에 따라 최소한의 신변 보호, 법률 지원, 소송 대응 지원 등 보호 장치 마련이 필요하다는 의견이 다수 제기되었다. 이러한 지원은 특정 집단에 대한 특혜라기보다, 신뢰 가능한 정보 생태계를 유지하기 위한 사회적 투자라는 인식이 공유된다.

- 반면, 정부 주도의 직접적인 재정 지원에 대해서는 신중론도 강하게 제기되었다. 국가 예산이 투입될 경우, 정부나 정치권의 영향력으로부터 팩트체커의 독립성과 중립성이 훼손될 수 있으며, 실제로 SNU팩트체크 중단 사례나 정권 교체에 따른 예산 삭감 경험은 이러한 우려를 현실화한 사례로 지적되었다. 정부가 지원 대상을 직접 선정하거나 운영할 경우 '정파적 팩트체킹' 논란이 반복될 가능성이 크다는 점에서, 제도의 정당성과 지속 가능성에 대한 의문도 함께 제기되었다. 일부 전문가들은 정부 재정 지원이 시장의 자율적 정화 기능을 약화시키고, 경쟁력 없는 사업자를 인위적으로 존속시키는 부작용을 초래할 수 있다는 점도 지적하였다.

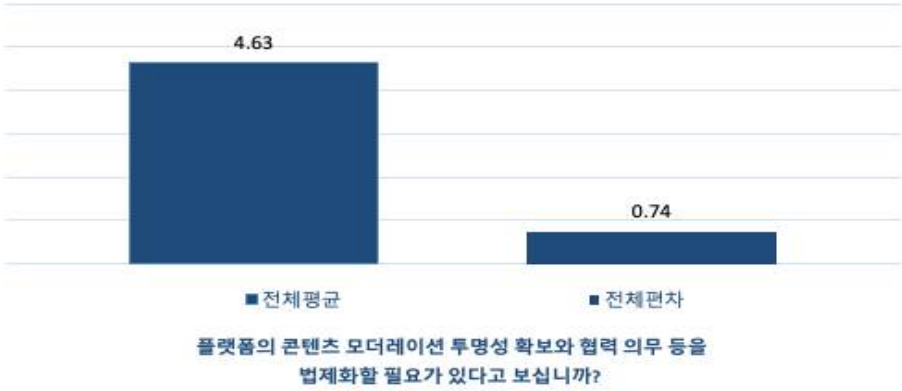
- 이에 따라 전문가 의견은 단순한 직접 지원보다는, 독립성을 보장할 수 있는 제도 설계의 필요성으로 수렴된다. 예컨대 법률 지원 및 신변 보호와 같은 최소한의 안전장치는 공적 책임으로 마련하되, 재정 지원의 경우 정부가 직접 집행하기보다는 독립 기구나 민간 중심의 운영 구조를 통해 배분하는 방식이 바람직하다는 제안이 제시되었다.

- 종합하면, 언론과 팩트체커 보호를 위한 지원제도의 필요성에 대해서는 전문가들 사이에 공감대가 형성되어 있으나, 그 방식에 있어서는 독립성과 정치적 중립성을 훼손하지 않는 제도 설계가 핵심 쟁점으로 나타난다. 향후 정책은 언론·팩트체커가 외압과 위협으로부터 보호받으며 검증 활동을 지속할 수 있는 최소한의 제도적 안전망을 마련하되, 지원의 집행과 운영은 정부로부터 일정 정도 분리된 구조를 통해 신뢰성과

지속 가능성을 확보하는 방향으로 설계될 필요가 있는 것으로 분석된다.

o **(플랫폼의 콘텐츠 모더레이션 투명성 및 협력 의무 법제화 필요성)** 평균 4.63, 표준편차 0.74로 전체 문항 중 가장 높은 평균값을 나타냈다. 전문가들이 플랫폼의 책임성 강화를 가장 시급한 정책 과제 중 하나로 보고 있음을 보여준다.

[그림 4-8] 플랫폼의 콘텐츠 모더레이션 투명성 및 협력 의무 법제화 필요성



- 전문가들은 디지털 플랫폼의 콘텐츠 모더레이션과 알고리즘 운영이 여론 형성과 정보 유통 구조에 미치는 영향이 크게 확대되고 있다는 점에서, 일정 수준의 투명성 확보와 공적 협력 의무를 법제화할 필요성에 대체로 공감하고 있다. 특히 플랫폼의 콘텐츠 삭제·차단·노출 제한 결정이 사실상 디지털 공론장의 규칙을 형성하고 있음에도 불구하고, 그 기준과 절차가 외부에 충분히 공개되지 않아 '사적 검열' 논란과 이용자 신뢰 저하를 초래하고 있다는 점이 반복적으로 지적되었다.

- 다수의 전문가들은 허위조작정보의 확산 양상과 사회적 영향을 실증적으로 분석하고 효과적인 대응 방안을 마련하기 위해서는, 플랫폼이 보유한 내부 데이터에 대한 연구자 및 공익 연구기관의 접근이 필수적이라고 보았다. 현재와 같이 플랫폼 내부 데이터가 사실상 폐쇄된 구조에서는 외부의 감시·검증 기능이 원활히 작동하기 어렵고, 이는 허위조작정보 대응 정책의 실효성을 저해하는 요인으로 작용한다는 인식이다. 이러한 점에서 데이터 제공, 투명성 보고서의 실효성 제고, 공익 연구 접근 보장 등 협력 의무를 제도적으로 명확히 할 필요성이 제기되었다.

- 다만 법제화의 방식과 범위에 대해서는 신중한 접근이 필요하다는 의견도 함께 제시되었다. 플랫폼 알고리즘과 콘텐츠 관리 로직은 기업의 핵심 영업비밀에 해당할 수 있으며, 과도한 공개는 이를 악용한 어뷰징(abusing)이나 규제 회피를 초래할 수 있다는 우려가 제기되었다. 이에 따라 전문가들은 알고리즘을 전면 공개하는 것보다는, 신뢰할 수 있는 독립 연구기관이나 공공 목적의 연구에 한해 제한적으로 데이터를 제공하거나, 콘텐츠 삭제·제한 조치에 대한 사유 통지 및 투명성 보고 의무를 강화하는 방식이 보다 현실적인 대안이라고 평가하였다.

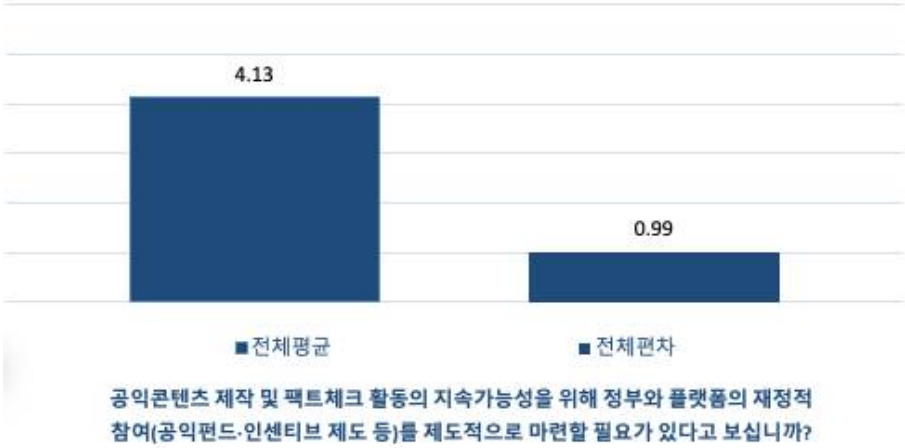
- 또한 국내 자율규제 체계의 구조적 한계 역시 주요 쟁점으로 지적되었다. 한국인터넷자율정책기구(KISO) 등 자율규제 기구가 일정 부분 기능하고 있으나, 유튜브·메타 등 주요 해외 플랫폼은 이러한 체계의 적용을 받지 않아 규제 공백이 발생하고 있다는 점이 문제로 제기되었다. 이에 따라 EU 디지털서비스법(DSA)과 같이 콘텐츠 삭제·제한 시 사유 제공 및 투명성 데이터베이스 제출을 의무화하는 법적 장치를 국내 제도에 도입할 필요성이 제기되었다.

- 아울러 일부 전문가들은 플랫폼이 막대한 수익을 창출하는 시장 지배적 행위자로서, 이용자 보호, 공익 연구 지원, 투명성 확보 등 사회적 책무를 법적 의무로 명확히 규정할 필요가 있다고 보았다. 자율적·형식적인 ESG 활동이나 일회성 사회공헌에 그치는 방식에는 한계가 있으며, 지속 가능하고 검증 가능한 협력 의무가 제도화되어야 한다는 문제의식이다. 특히 허위조작정보 규율의 초점을 정보 생산자가 아닌 유통자인 플랫폼에 두고, 국내 법제도 환경에 부합하는 ‘한국형 플랫폼 책임 프로토콜’을 마련할 필요성도 제기되었다.

- 종합하면, 플랫폼의 콘텐츠 모더레이션 투명성 확보와 협력 의무 법제화에 대해서는 필요성에 대한 폭넓은 공감대가 형성되어 있으나, 기업의 영업 자유와 표현의 자유 침해 우려를 최소화하는 정교한 제도 설계가 핵심 과제로 나타난다. 향후 정책은 해외 주요국의 입법례(DSA, 영국 및 호주의 온라인 안전법 등)를 참고하되, 국내 플랫폼 생태계와 법체계에 부합하도록 단계적·비례적인 방식으로 투명성 의무와 협력 범위를 설정하고, 자율규제와 법적 규율을 결합한 공동규제 모델로 발전시킬 필요가 있는 것으로 분석된다.

o **(공익콘텐츠 제작 및 팩트체크 활동에 대한 재정 참여제도 마련 필요성)** 평균 4.13, 표준편차 0.99로 공익펀드, 인센티브 등 지속가능성을 담보하기 위한 재정 지원에 대해 높은 공감대가 확인되었다.

[그림 4-9] 공익콘텐츠 제작 및 팩트체크 활동에 대한 재정 참여제도 마련 필요성



- 전문가들은 공익콘텐츠 제작과 팩트체크 활동이 사회적 가치는 매우 높음에도 불구하고, 현행 미디어 시장 구조에서는 지속 가능성이 현저히 낮다는 점에 대체로 공감하고 있다. 특히 자극적이고 감정적인 허위조작정보가 더 많은 트래픽과 광고 수익을 창출하는 반면, 고비용·저수익 구조의 팩트체크 보도와 공익 콘텐츠는 시장 논리만으로는 충분히 공급되기 어려운 전형적인 '시장 실패' 영역이라는 게 공통적인 인식이다.

- 다수의 전문가들은 이러한 구조적 문제를 해소하기 위해, 허위정보 유통 환경에서 실질적인 경제적 이익을 얻고 있는 플랫폼 기업이 정보 생태계의 안정화에 상응하는 재정적 책임을 분담할 필요가 있다고 보았다. 플랫폼의 매출 규모나 국내 트래픽 점유율 등에 비례하여 공익 펀드에 출연하도록 제도화하고, 이를 통해 팩트체크 활동과 고품질 공익 콘텐츠 제작을 안정적으로 지원하는 방안이 대안으로 제시되었다. 이 경우 정부는 직접 집행 주체가 되기보다는 제도적 틀을 마련하고 운영의 독립성을 보장하는 역할을 수행해야 한다는 의견이 우세하였다.

- 실제 과거 네이버의 SNU팩트체크 지원 사례나 팩트체크넷 운영 경험은, 일정 수준의 재정적 뒷받침이 있을 경우 팩트체크 생태계가 확장되고 기획취재, 인력 양성, 시민 참여 프로그램 등으로 파급 효과가 나타날 수 있음을 보여주는 사례로 언급되었다. 이러한 경험은 공익적 지원이 단순한 재정 이전이 아니라 정보 생태계의 기반을 형성하는 역할을 할 수 있음을 시사한다.

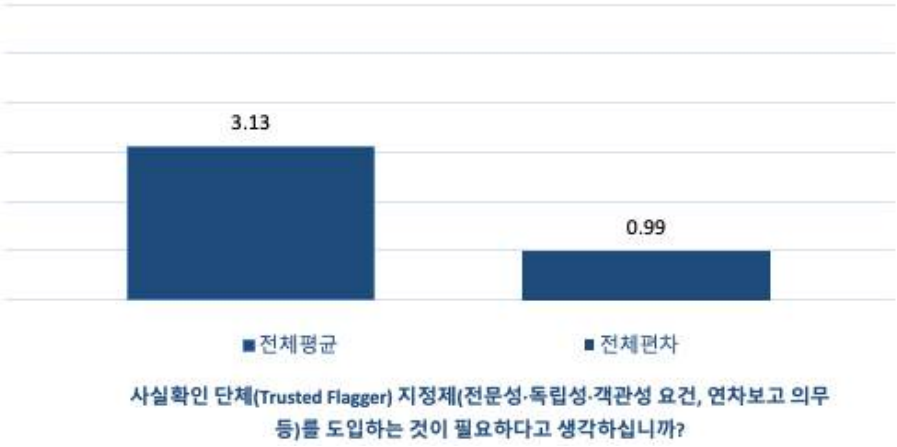
- 다만 정부의 직접적인 재정 지원에 대해서는 신중론도 동시에 제기되었다. 전문가들은 SNU팩트체크 중단 사례 등에서 보듯이, 국가 예산이 투입될 경우 정치권의 영향력이나 정권 변화에 따른 지원 중단 가능성이 높아, 팩트체커의 독립성과 중립성이 훼손될 위험이 크다고 지적하였다. 특히 한국의 정치·제도적 환경에서는 '재정 지원이 곧 개입으로 인식되는 구조'를 충분히 고려해야 한다는 문제의식이 제기되었다.

- 이에 따라 일부 전문가들은 공익 펀드 방식 자체에 부정적인 입장을 보이며, 정부의 직접 지원보다는 플랫폼의 자율적 책임 이행을 법·제도화 하고, 그 이행 여부에 따라 인센티브를 부여하는 방식이 보다 바람직하다고 평가하였다. 또 다른 전문가들은 정부 재정 투입이 불가피하더라도, 그 운용은 민간 또는 독립 기구가 담당하고, 정부는 최소한의 제도적·재정적 기반만 제공하는 구조가 필요하다고 보았다.

- 종합하면, 공익콘텐츠 제작과 팩트체크 활동의 지속가능성을 확보하기 위해 재정적 참여가 필요하다는 데에는 상당한 공감대가 형성되어 있으나, 그 방식에 있어서는 정부의 직접 개입을 최소화하고 독립성과 중립성을 어떻게 제도적으로 담보할 것인지가 핵심 쟁점으로 나타난다. 향후 정책 설계에서는 플랫폼의 재정적 책임을 명확히 하는 제도화, 독립적 기금 운영 구조, 간접 지원 및 인센티브 중심의 접근 등을 결합하여, 공익성과 정치적 중립성을 동시에 확보할 수 있는 정교한 재원 조달 모델을 마련할 필요가 있는 것으로 분석된다.

o **(사실확인 단체(Trusted Flagger) 지정제 도입 필요성)** 평균 3.13, 표준편차 0.99로 타 항목에 비해 필요성이 다소 낮게 평가되었다.

[그림 4-10] 사실확인 단체 지정제 도입 필요성



- 전문가들은 사실확인 단체(Trusted Flagger) 지정제 도입에 대해 필요성과 위험성이 동시에 존재하는 제도라는 점에서 의견이 뚜렷하게 갈리는 양상을 보였다. 허위조작정보와 혐오표현이 대규모로 확산되는 환경에서 전문성과 신뢰성을 갖춘 주체가 플랫폼의 신고·조치 과정에 참여하는 것이 효율성과 신속성을 높일 수 있다는 점에는 일정 부분 공감대가 형성되었다. 특히 현재와 같이 정치권이나 이해관계자가 신고의 주체가 되는 구조에서는 정파적 논란과 정치적 압박으로 비칠 우려가 크기 때문에, 언론·미디어 분야의 전문 모니터 기관이나 팩트체크 기관을 중심으로 한 제도적 장치가 필요하다는 의견이 제기되었다.

- 찬성 의견을 제시한 전문가들은 객관적이고 엄격한 지정 요건을 통해 팩트체크 활동의 질을 제고하고, 연차보고 의무 등을 통해 검증 기준과 절차를 투명하게 공개함으로써 사회적 신뢰와 책임성을 강화할 수 있다고 보았다. 이러한 지정제가 플랫폼의 자율규제를 보완하고, 허위정보 피해에 대한 신속한 구제와 대응 체계를 마련하는 데 기여할 수 있다는 인식이다.

- 또한 사실확인 단체의 신고에 플랫폼이 과도하게 의존할 경우, 법적 책임 회피나 행정편의성을 이유로 합법적 표현까지 삭제·차단되는 과잉 집행의 문제가 발생할 수 있다는 우려도 제시되었다. 이는 법적 절차와 사법적 판단을 우회하는 사실상의 검열로 기능할 가능성이 있으며, 신고 주체와 기준, 절차가 명확히 관리되지 않을 경우 플랫

폼과 지정 단체 모두에 대한 통제와 책임 구조가 불분명해질 수 있다는 지적이다.

- 팩트체크의 핵심 가치는 권력으로부터의 독립성에 있기 때문에 국가나 법적 기구가 특정 단체에 자격을 부여하는 것에 대해 우려하는 의견도 있었다. 이 경우 사실확인 단체 지정이 정권의 입맛에 맞는 단체를 선별 및 지원하는 장치로 작동할 수 있다는 비판이다. 무엇이 진실인지 판단할 권한을 특정 주체에 독점적으로 부여하는 구조는 민주주의 원칙과 표현의 자유에 배치될 수 있다는 문제의식도 제기되었다.

- 특히 전문가들은 제도의 성패를 좌우하는 핵심 쟁점으로 '누가 지정 주체가 되는가'를 반복적으로 강조하였다. 정부가 직접 지정하거나 권위를 인증하는 방식은 정권 교체에 따라 기준이 달라지고 정치적 논란을 야기할 가능성이 높다는 점에서 부적절하다는 의견이 우세하였다. 일부 전문가들은 새로운 단체 지정 제도를 도입하기보다는, 규제기관 내 독립적 전문위원회 설치나 기존 국제적 인증 체계를 활용하는 방식이 보다 현실적인 대안이 될 수 있다고 보았다. 예컨대 국제팩트체크네트워크(IFCN)의 인증(Signatory)을 받은 언론사·기관과 플랫폼이 자율적으로 협업하도록 유도하는 방식은, 정부 개입을 최소화하면서도 국제적 기준에 부합하는 팩트체크 역량을 확대할 수 있는 방안으로 제시되었다.

- 종합하면, 사실확인 단체(Trusted Flagger) 지정제는 허위조작정보 대응의 효율성을 높일 수 있는 효과를 지니는 동시에, 독립성과 표현의 자유를 훼손할 수 있는 위험성도 내포하는 제도라는 점에서 신중한 접근이 요구된다. 향후 정책 설계에서는 정부가 직접 지정하거나 인증하는 방식보다는, 국제적 기준이나 민간·시민사회 중심의 자율적 인증 체계를 존중하고, 플랫폼의 협력 의무와 투명성 강화를 통해 간접적으로 신뢰 가능한 팩트체크 생태계를 확장하는 방향이 보다 적절한 대안으로 검토될 필요가 있다.

o **(정보 무결성 강화를 위한 정부 및 사업자 역할)** 먼저, 정부의 역할과 관련하여 전문가들은 정보 무결성 강화를 위해 정부가 직접적인 내용 규제나 판단 주체로 나서기보다는, 제도적 틀과 환경을 조성하는 역할에 집중해야 한다는 점에 대체로 공감하였다. 특히 팩트체크와 공익 콘텐츠 활동에 대해서는 재정적·제도적 지원이 필요하되, 그 내용이나 결론에는 개입하지 않는 이른바 '팔길이 원칙(arm's length principle)'을

명확히 확립해야 한다는 의견이 반복적으로 제시되었다. 이는 정부 지원이 곧 정치적 개입이나 검열로 인식될 수 있는 국내 환경을 고려할 때, 정책의 정당성과 지속 가능성을 확보하기 위한 핵심 전제 조건으로 인식된다.

- 아울러 다수의 전문가들은 허위조작정보 대응의 핵심 축으로 규제 중심 접근보다 미디어·정보 리터러시 교육 강화를 강조하였다. 정보 차단이나 처벌 중심 정책은 표현의 자유 침해 논란을 초래할 수 있는 반면, 시민이 허위정보를 스스로 식별·판별할 수 있는 역량을 강화하는 것은 장기적으로 정보 생태계의 회복탄력성을 높이는 방식이라는 인식이다. 이에 따라 전 생애주기별 미디어 리터러시 교육과 팩트체크 교육을 체계적으로 확대할 필요성이 제기되었다.

- 제도 설계 측면에서는 해외 사례를 참고한 절차 중심 규제의 필요성도 강조되었다. 허위조작정보 표현물 자체에 대한 처벌 강화보다는, EU 디지털서비스법(DSA)이나 호주 온라인 안전법과 같이 플랫폼 사업자의 절차적 책무, 투명성 의무, 협력 구조를 명확히 하는 방식이 보다 현실적이라는 평가이다. 또한 정부는 플랫폼 사업자, 시민사회, 전문가 집단이 참여하는 사회적 논의 기구를 통해 정책 방향에 대한 합의를 도출하고, 그 결과를 토대로 단계적·비례적인 법제화를 추진해야 한다는 의견도 제시되었다.

- 또한 사업자의 역할과 관련해서는, 전문가들은 플랫폼 사업자가 정보 무결성 문제에서 핵심 행위자라는 점을 분명히 하며, 기술적 역량과 시장 지위를 활용한 적극적인 책임 이행이 필요하다고 보았다. 특히 허위조작정보가 높은 조회수와 광고 수익으로 이어지는 현재의 수익 구조는 정보 왜곡을 구조적으로 강화하고 있다는 점에서, 사업자 스스로 비즈니스 모델을 점검하고 개선할 필요성이 제기되었다. 유해·허위 콘텐츠에 광고가 붙지 않도록 광고 투명성을 강화하고, 허위정보를 통해 발생한 수익을 차단·환수하는 등 실질적인 조치가 요구된다는 인식이다.

- 또한 플랫폼의 알고리즘 운영과 콘텐츠 관리 기준이 여론 형성과 정보 유통에 미치는 영향이 큰 만큼, 투명성 제고가 사업자의 핵심 책무로 강조되었다. 외부 전문가와 연구자가 허위정보의 확산 양상을 분석할 수 있도록 데이터 접근성을 확대하고, AI 생성물 라벨링, 메타데이터 표기 등 기술적 조치를 선제적으로 도입하는 것이 필요하다는 의견이 다수 제시되었다. 이러한 조치는 법적 강제 이전에 사업자가 자율적으로 추진

할 수 있는 영역으로 인식되고 있다.

- 한편 일부 전문가들은 자율규제의 중요성을 강조하며, 플랫폼 산업계가 공동의 목표 아래 자율 강령이나 산업별 코드(Industry Code)를 수립하고 이를 실질적으로 이행하는 노력이 필요하다고 보았다. 정부는 이러한 자율 규범이 실효성을 갖도록 평가·보완하는 역할에 머무르고, 사업자는 글로벌 스탠다드에 부합하는 대응 체계를 준비·이행하는 방식의 공동규제 모델이 바람직하다는 평가이다. 특히 AI 생성물 식별, 콘텐츠 모더레이션 기준 공개 등은 사업자가 스스로 해법을 제시하고 실행함으로써 신뢰를 구축해야 할 영역으로 지적되었다.

- 종합하면, 정보 무결성 강화를 위해서는 정부와 사업자의 역할 분담이 명확히 설정될 필요가 있다. 정부는 규제의 직접 주체가 되기보다 제도적 기반과 협력 구조를 마련하고, 사업자는 투명성과 사회적 책임에 기반한 자율적 실천을 통해 정보 생태계의 신뢰 회복에 기여해야 한다는 점에서 전문가들의 인식이 수렴되는 것으로 나타난다. 이러한 역할 분담이 전제될 때, 표현의 자유를 존중하면서도 허위조작정보로 인한 사회적 비용을 최소화하는 균형 잡힌 정보 무결성 정책이 가능할 것으로 분석된다.

o **(한국형 정보 무결성 전담기구의 바람직한 형태)** 다수의 전문가가 선호한 전담기구 형태는 학계·시민사회·기술 전문가가 주축이 되는 독립적 민간 자율기구 모델이다. 이 모델은 정부의 구성권·운영 개입을 최소화해 정치적 중립성을 확보하고, 플랫폼과의 협력을 통해 공동규제로 연결하는 방식이 가능하다는 점이 장점으로 언급되었다.

- 대안 모델로는 특정 부처 산하가 아닌 초부처적 독립위원회형 전담기구를 두되, 위원 추천·구성 단계에서 정치권 비중을 제한하고 시민사회·전문가 단체 추천과 여야 합의 절차를 강하게 설계해 제도적 중립성 장치를 강화하는 방안이 제시되었다.

- 소수 의견으로는 기구 신설 자체가 옥상옥이 될 수 있으므로, 플랫폼 규제 권한과 정책 조율 기능을 기존 부처/조직 내 단일 전담부서로 통합하는 방식이 효율적이라는 견해도 존재한다. 다만 이 경우에도 정치적 논란을 최소화할 별도의 독립적 검증·자문 체계가 병행되어야 한다는 의견이 제시되었다.

o **(정보 무결성 강화를 위한 민관협력 구조)** 전문가들은 협력 구조가 정부-플랫폼-시민사회 중 어느 한 축에 과도하게 집중될 때 실효성과 정당성이 동시에 훼손된다고 보았다. 이에 따라 협력 구조는 대체로 다음과 같은 다층적 역할 분담으로 제안된다.

- 플랫폼 중심: 유통 규모와 속도를 고려할 때 탐지・확산 차단・조치 실행은 플랫폼의 기술 역량이 핵심이며, 전담기구와의 데이터 공유・절차 준수로 책임성 강화(전문가 B, A).
- 시민사회・학계 중심: 사실 검증 및 교육・연구는 권력으로부터 독립된 민간 영역이 담당해야 신뢰를 확보(전문가 B, G, C).
- 정부/전담기구 중심: 정부는 직접 판단 주체가 되기보다 법・제도 정비, 부처 간 조율, 위기 상황에서의 공공 커뮤니케이션 기반 마련을 담당하되, 판단・검증은 독립 기구/민간에 맡기는 구조가 바람직(전문가 D, G, A).
- 특히 전문가 G는 협력 구조를 '재정-감독-운영-집행'의 분리 원칙으로 구체화한 안을 제안하였다. 즉, 재원은 공적 기금(정부・플랫폼 출연 등)으로 마련하되, 재원을 마련한 주체가 감독・운영・집행에 개입하지 않도록 차단하고, 민간 중심 독립기구가 센터(또는 수행기관)를 선정・감독하며, 대학・연구기관 기반의 복수 센터가 감시・연구・보고를 수행하고, 마지막으로 정부는 법률에 근거해 플랫폼의 절차적 책무를 집행하는 방식이다. 이는 "지원은 하되 간섭하지 않는다"는 팔길이 원칙(Arm's Length Principle)을 구조적으로 구현하려는 제안으로 해석된다.

제2절 정보 무결성 정책 마련 시사점

1. 정보 무결성 강화를 위한 정책 로드맵 우선 과제

전문가 의견을 종합한 결과, 정보 무결성 강화를 위한 정책 방향에 대해 다음과 같이 몇 가지 핵심적인 우선 과제가 도출되었다.

① 미디어・정보 리터러시 강화 : 우선 가장 폭넓은 공감대가 형성된 과제는 미디어・정보 리터러시 교육의 강화이다. 다수의 전문가들은 허위조작정보 문제를 규제만으로 해결하기에는 한계가 있으며, 시민 개개인이 정보를 비판적으로 판단할 수 있는 역량을 기르는 것이 장기적으로 가장 효과적인 대응 전략이라고 보았다. 이에 따라 전 국민을 대상

으로 한 생애주기별 미디어·정보·AI 리터러시 교육을 체계적으로 확대함으로써, 허위조작정보에 대한 사회 전반의 '면역력'을 강화할 필요성이 강조되었다.

② 플랫폼의 투명성 확보와 책임 강화 : 다음으로 플랫폼의 투명성 확보와 책임 강화를 우선 과제로 지적하는 의견이 다수 제시되었다. 전문가들은 알고리즘과 콘텐츠 모더레이션 정책이 여론 형성과 정보 유통 구조에 미치는 영향이 크게 확대된 상황에서, 플랫폼의 운영 방식이 지나치게 불투명하다는 점을 문제로 지적하였다. 특히 허위조작정보가 높은 조회수와 광고 수익으로 연결되는 현재의 수익 구조가 문제를 구조적으로 악화시키고 있다는 인식이 공유되었으며, 알고리즘 운영, 콘텐츠 삭제·차단 기준, 광고 집행과의 연계 등에 대해 일정 수준의 투명성과 사회적 책임을 제도적으로 명확히 할 필요성이 제기되었다.

③ 공익 콘텐츠와 팩트체크 활동의 지속가능성 확보 : 전문가들은 팩트체크와 공익 콘텐츠가 민주주의와 정보 무결성 측면에서 높은 사회적 가치를 지니고 있음에도 불구하고, 수익성이 낮아 시장 논리만으로는 충분히 공급되기 어려운 '시장 실패' 영역에 해당한다고 보았다. 이에 따라 제도적·재정적 지원을 통해 지속가능성을 확보할 필요성에는 대체로 공감이 형성되었으나, 정부의 직접적인 개입이나 운영 관여는 정치적 중립성과 독립성을 훼손할 수 있다는 점에서 신중해야 한다는 의견도 함께 제기되었다. 즉, 지원의 필요성에는 합의가 있으나, 그 방식은 '팔길이 원칙'을 전제로 설계되어야 한다는 인식이 공통적으로 나타났다.

④ 정치적 독립성을 전제로 한 거버넌스 구축 : 정보 무결성 정책을 뒷받침할 거버넌스 체계와 관련하여, 정치적 영향으로부터 독립된 전담기구의 필요성이 반복적으로 언급되었다. 전문가들은 허위조작정보 대응이 단기적·사후적 조치에 그쳐서는 안 되며, 지속적인 연구와 숙의를 통해 정책을 발전시킬 수 있는 안정적인 거버넌스가 필요하다고 보았다. 이 과정에서 정부의 역할은 직접적인 판단이나 개입이 아니라, 재정적·제도적 기반을 제공하는 수준으로 한정되어야 하며, 실질적인 운영과 판단은 민간과 전문가 중심으로 이루어져야 한다는 점이 강조되었다. 아울러 다수의 전문가들은 순수한 행정규제나 완전한 자율규제 모두 한계를 지닌다는 점에서, 자율규제와 행정규제를 결합한 공동규제 모델이 바람직하다는 의견을 제시하였다. 정부는 기본적인 원칙과 기준을 제시하고, 플랫폼과 산업계는 이를 바탕으로 자율적 규범을 마련·이행하며, 정부는 그 실효성을 점검하는 방식이

현실적인 대안으로 인식되었다. 이러한 접근은 유럽연합의 디지털서비스법(DSA)이나 호주의 공동규제 모델 등 해외 사례를 참고할 필요가 있다는 점에서도 공통된 의견이 나타났다.

⑤ 허위조작정보 개념 명확화 : 마지막으로, 일부 전문가들은 허위조작정보 정책 논의 전반에 있어 허위조작정보 개념의 명확화가 필요하다는 점을 강조하였다. 이는 정책 설계 과정에서 과잉 규제와 정치적 오남용을 방지하는데 핵심적인 요소라 볼 수 있다.

2. 정보 무결성 강화를 위한 정책 방향 시사점

가. 절차·플랫폼 책임 중심의 한국형 공동규제 모델 전환

전문가들은 처벌 강화로 허위조작정보에 대응하기엔 한계에 봉착했다는 점에 대체로 공감하였다. 플랫폼의 유통 구조와 알고리즘 영향력이 커진 상황에서, 콘텐츠 자체의 진위 판단을 정부가 직접 수행하는 방식은 표현의 자유 침해 및 정치적 오남용 우려를 동반할 수 있다. 따라서 플랫폼의 절차적 책무(투명성·신속 대응·이의제기·데이터 접근)를 법제화하고, 세부 운영은 산업계 자율규범과 독립적 민간 거버넌스가 결합된 공동규제 체계로 정교화할 필요가 있다.

나. 플랫폼 책임 강화 및 콘텐츠 모더레이션 투명성·데이터 협력의무 제도화

전문가 조사 결과에서 플랫폼 투명성·협력 의무(B-5)가 가장 높은 필요성(평균 4.63)을 보인 것은, 전문가들이 허위조작정보 문제의 핵심을 생산자 규제보다 유통 구조의 책임성으로 인식하고 있음을 시사한다. 정책 과제로는 △콘텐츠 삭제·제한 사유 공개와 기록관리 △투명성 보고서의 표준화·실효성 강화 △공익 연구 목적의 데이터 접근 보장(보안·영업비밀 보호 장치 포함) 등을 '한국형 플랫폼 책임 프로토콜'로 체계화하여, 자율규제의 한계를 보완하고 해외 플랫폼까지 포괄하는 규율 기반을 마련할 필요가 있다.

다. AI 생성물 식별 고위험 영역 중심의 단계적·차등적 접근 추진

AI 생성물 식별(B-2)은 높은 공감(평균 4.38)을 얻었으나, 기술적 한계(제거·변조 가능,

오픈소스·해외 생성물 통제 어려움)를 고려할 때 중장기적으로는 “최소 기준의 제도화 + 고위험 영역(선거·뉴스·건강·재정정보 등) 우선 적용”이 합리적일 것으로 보인다. 아울러 ‘AI 전면 생성/AI 보조 활용/AI 검색 활용’ 등 단계적 라벨 체계를 마련하고, 표시 준수에 대한 인센티브와 유통상 신뢰 신호 부여를 병행하는 것이 정책 수용성과 집행 가능성을 높일 수 있다.

라. 이용자 참여 메커니즘 표준화 및 구현 자율성 보장

이용자 참여 표준화(B-3)는 필요성이 높게 나타났으나(평균 4.00), 전문가들은 서비스별 특성을 무시한 획일적 UI/UX 강제에 대해 우려하는 것으로 나타났다. 이에 따라 중장기적으로는 이용자 참여 메커니즘에 대한 최소 공통 기준을 설정하고, 세부적인 구현 방식에 대해서는 플랫폼의 유형과 규모에 따라 자율성을 부여하는 방식이 바람직하다. 구체적으로는 △이용자가 신고 기능에 쉽게 접근할 수 있도록 하는 접근성 기준, △신고 접수 후 처리 단계별 기한 설정, △조치 결과 및 판단 사유 등에 대한 통지, △이의제기 및 재검토 절차 등을 공통적으로 마련하는 게 필요하다.

마. 팩트체크·공익콘텐츠에 대한 지속 가능한 재원 지원 및 독립성 확보

언론·팩트체커 보호(B-4, 평균 4.50)와 재정 참여(B-6, 평균 4.13)에 대한 높은 동의는, 전문가들이 공익적 검증·콘텐츠 생산이 구조적으로 시장에서 저공급될 수밖에 없는 영역임을 공유하고 있음을 보여준다. 이에 따라 정책적으로는 플랫폼 기업의 출연과 공적 재원을 결합한 재원 조성 방식을 검토하되, 외부 영향력으로부터 독립된 독립적 기금(arm’s length) 구조로 운영하는 것이 바람직하다. 해당 기금은 독립 언론 및 팩트체커의 안정적 활동 지원, 허위조작정보 대응 과정에서 발생할 수 있는 민형사상 분쟁에 대한 법률 지원, 기자 및 팩트체커 개인에 대한 신변보호 대응, 전략적 봉쇄 소송(SLAPP)에 대한 대응 등에 활용하여 공익적 검증 활동이 위축되지 않도록 최소한의 보호 장치를 마련하는 것이 필요하다.

바. 전담기구 독립성 확보 및 다층적 거버넌스 구축

전문가 의견은 전담기구의 필요성 자체보다 정치적 독립성과 운영 구조에 초점이 모였다. 특히 재정, 감독, 운영, 집행 기능을 분리하는 구조(독립기금, 민간 중심 감독기구, 복수 수행기관/센터 운영, 정부의 법 집행 분리)는 제도의 신뢰성과 지속가능성을 제고하는 설계 원칙으로 제시되었다. 중장기적으로는 독립적 민간기구 또는 초부처 독립위원회 모델을 비교 검토하되 정부는 법제도적 기반과 재정적 지원을 제공하는 역할에 중점을 두고, 플랫폼 사업자는 데이터 제공과 절차 준수 등 공동 규제의 참여 주체로서 역할을 분담하는 방향으로 단계적인 다층적 거버넌스 체계 구축을 검토할 필요가 있다.

제5장. 정보 무결성 중장기 정책방향 제안

제1절 정보 무결성 강화를 위한 단기 과제

1. 정보통신망법 개정에 따른 후속조치 과제

개정된 정보통신망법이 허위조작정보에 대한 신고 및 그에 대한 조치·이의제기·분쟁조정, 투명성 보고 절차와 사실확인단체 연계, 투명성센터 설치 등 절차적 대응 체계를 포괄하여 규정하고 있는 만큼, 단기적으로는 법률의 취지가 실제로 작동할 수 있도록 시행령·고시 수준에서 세부 집행 기준을 구체화할 필요가 있다. 특히 플랫폼별로 내부 정책과 운영 관행이 상이한 상황에서 최소한의 공통 기준을 제시함으로써, 이용자 권리보호와 플랫폼 책임성 확보가 제도적으로 담보되도록 하는 것이 우선 과제로 제시된다.

첫째, '대규모 정보통신서비스 제공자'의 범위와 지정 기준을 시행령에서 명확히 할 필요가 있다. 이용자 규모, 국내 매출, 트래픽, 서비스 유형 등을 종합적으로 고려하는 기준을 마련하고, 규모와 영향력에 비례하여 의무 강도를 단계적으로 적용하는 체계를 검토하는 것이 필요하다. 이를 통해 초대형 플랫폼에 대해서는 투명성·협력 의무를 강화하되, 중소 서비스에 대해서는 과도한 규제 비용을 최소화하는 비례 원칙을 구현할 필요가 있다.

둘째, 신고-조치-통지-이의제기로 이어지는 절차 전반에 대해 '최소 공통 기준'을 마련할 필요가 있다. 신고 기능의 접근성 확보 방안, 처리 단계별 기한 설정, 조치 사유의 고지 방식, 이의제기 및 재검토 절차 등 핵심 절차 요건을 표준화하는 방안을 검토하여야 한다. 예컨대 허위조작정보의 신고 방법을 구체화하고, 신고 접수 이후 1차 판단 통지 시점과 이의제기 및 재검토에 관한 처리 기한을 명확히 하는 것이 필요하다. 또한 삭제·차단뿐 아니라 노출 제한, 계정정지 또는 해지, 광고 수익 등 수익화 제한 등 다양한 조치 유형에 대해서도 정부, 사업자, 전문가 및 이해관계자 논의를 통해 선택 가이드라인을 마련하는 방안이 고려될 수 있다. 이러한 세부 절차의 표준화는 사업자의 허위조작정보에 대한 대응이 일관된 기준에 따라 이루어지도록 함으로써 자의적 판단 여지를 줄이고, 이용자의 예측 가능성과 권리구제 실효성을 높일 수 있다.

셋째, 투명성 보고 의무가 선언적 수준에 머무르지 않도록 신고 및 처리 통계, 조치 유형별 비율, 이의제기 인용률 및 재조치율, 국가기관 요청 사항 및 조치 현황, 플랫폼 약관

및 자율규제 정책에 따른 처리 통계 등 주요 지표에 대한 공통 템플릿을 마련하는 것이 필요하다. 또한 기계 판독 가능한 형식으로 공개하도록 한 법률 취지를 실질화하기 위해 데이터 스키마(예: CSV/JSON), 공개 주기, 보존 기간 등을 하위 규범에서 구체화하는 방안이 검토될 수 있다. 이는 플랫폼 조치의 예측 가능성을 제고하는 동시에, 시민사회·학계의 외부 평가와 정책 효과 분석 기반을 제공한다.

넷째, 플랫폼의 허위조작정보 대응 운영정책 수립 과정에서 '이해관계자 의견 반영'의 실효성을 제고하는 장치가 필요하다. 공개 의견수렴 절차, 수렴 결과에 대한 반영·미반영 사유의 공개, 주요 정책 변경 사항에 대한 사전 고지 등 기본적인 절차를 마련함으로써, 자율규제가 기업 내부 규칙에 그치지 않고 공적 기능을 수행하는 준규제 장치로 작동할 수 있도록 할 필요가 있다.

2. 정보 무결성 제도의 신뢰성 및 독립성 확보 과제

정보통신망법 개정으로 새로운 절차와 제도적 장치가 도입됨에 따라, 제도의 원활한 운영을 위해서는 실효성 제고와 함께 운영의 공정성과 독립성에 대한 신뢰를 확보하는 것이 중요한 과제로 제시된다. 특히 사실확인단체 연계, 투명성센터 운영, 공익 연구 목적의 데이터 접근 등은 제도 설계와 운영 방식에 따라 정부의 역할과 책임에 대한 해석이 달라질 수 있는 영역인 만큼, 권한의 범위 및 운영 원칙을 명확히 설정할 필요가 있다.

첫째, 사실확인단체 연계 규정의 집행 가능성을 높이기 위해 인증·협력의 최소 요건을 명확히 할 필요가 있다. 국제적 사실확인 절차 규범 준수를 전제로 하되, 이해충돌 방지, 오류 정정 체계, 방법론 공개, 투명성 보고 등 최소 요건을 하위 규범에서 구체화하고, 플랫폼이 제공해야 하는 접근 데이터의 범위와 제공 가능한 데이터 항목, 그리고 제공 제한 사유를 명확히 정리함으로써 협력 구조의 예측 가능성을 제고할 수 있다. 아울러 독립성 논란을 최소화하기 위해 정부가 개별 주체를 직접 선정하는 방식보다는, 사실확인단체가 객관적 요건을 충족하는 경우 등록·공표하고 사후 점검을 통해 관리하는 구조가 보다 안정적인 대안이 될 수 있다.

둘째, 투명성센터는 '지원·연계·연구' 중심의 기능을 명확히 하고, 정치적 중립성에 대한 우려를 완화할 수 있는 운영 원칙을 정립할 필요가 있다. 센터의 주요 역할을 사실확인단체 데이터베이스 운영, 연구·교육 및 국제협력 지원 등으로 명확히 하는 한편,

외부 전문가 자문 체계, 이해충돌 방지 기준, 사업 및 예산 집행의 투명성 확보 장치를 병행함으로써, 센터의 기능이 개별 콘텐츠 판단이나 직접 규제로 확장되지 않도록 제도적 경계를 설정할 필요가 있다.

마지막으로, 공익 연구 목적의 데이터 접근 보장은 개인정보 및 영업비밀 보호와의 조화를 전제로 하여, 시행령·고시 등 하위 규범에서 절차적 기준을 구체화할 필요가 있다. 연구자 자격 요건과 연구 목적의 범위, 비식별화 및 안전한 연구환경 제공, 데이터 반출 통제 등 보호 조치를 명확히 하고, 위반 시 책임과 제재 기준을 함께 설정함으로써, 데이터 협력 의무의 실효성과 권리 보호 간 균형을 도모할 필요가 있다.

제2절 중장기 정책 전략 및 과제

중장기적으로는 정보 무결성 정책이 단기적 대응이나 개별 입법에 머무르지 않고, 플랫폼 책임, 공익적 검증 생태계, 재정의 지속가능성, 독립적 운영 구조가 유기적으로 결합된 공동규제 체계로 발전할 수 있도록 제도적 기반을 단계적으로 고도화할 필요가 있다. 전문가 의견에서도 확인되었듯이, 제도의 신뢰성과 지속가능성을 확보하기 위해서는 법제 정비, 재원 조성, 거버넌스 설계를 상호 연계된 축으로 통합적으로 추진하는 접근이 요구된다.

1. 법제 고도화: 절차·시스템 중심의 플랫폼 책임 구조 확립

중장기적으로는 정보 무결성 관련 법제가 개별 콘텐츠의 삭제·차단 여부를 중심으로 한 사후 규율을 넘어, 플랫폼의 절차적·시스템적 책임을 구조화하는 방향으로 발전할 필요가 있다. 이를 위해 투명성 보고, 데이터 협력, 위험 평가 및 완화 조치, 외부 연구·감사 기반의 검증 체계를 포괄하는 이른바 ‘플랫폼 책임 프로토콜’을 단계적으로 정립하는 방안을 검토할 수 있다. 특히 이러한 ‘플랫폼 책임 프로토콜’은 서비스의 규모와 기능, 정보 확산 위험의 정도에 따라 차등 적용되는 위험 기반 접근을 전제로 설계될 필요가 있다.

아울러 글로벌 플랫폼을 포함한 집행 가능성을 제고하기 위해, 역외 적용 원칙, 국내 대리인 또는 연락 창구 지정, 이행 점검 및 보고 체계 등 법 집행 수단 간 정합성을 함께 고

려할 필요가 있다.

한편, AI 생성물 식별 제도는 기술적 한계와 표현의 자유에 대한 고려를 전제로, 선거·뉴스·재난·보건 등 사회적 파급력이 큰 고위험 영역을 중심으로 단계적·차등적 적용을 검토할 필요가 있다. 중장기적으로는 생성·편집·합성 수준에 따른 표기 체계, 유통 단계에서의 메타데이터 유지, 위반 시 비례적 제재와 자발적 준수를 유도하는 인센티브를 결합하는 방식으로 제도의 정합성과 수용성을 높일 수 있다.

2. 재정 기반 구축: 공익 콘텐츠 생태계의 지속가능성 확보

정보 무결성 정책의 실효성을 확보하기 위해서는 공익 콘텐츠 제작과 팩트체크 활동이 시장 논리만으로는 충분히 공급되기 어렵다는 점을 전제로, 중장기적 운영 지원 기반을 마련할 필요가 있다. 이를 위해 플랫폼 기업의 출연과 공적 재원을 결합한 재원 조성 모델을 검토하되, 운영은 외부 영향력으로부터 독립된 ‘arm’s length’ 기금 구조로 설계하는 것이 바람직하다. 해당 기금은 독립 언론 및 팩트체커의 안정적 활동 지원, 허위조작정보 대응 과정에서 발생할 수 있는 민·형사상 분쟁에 대한 법률 지원, 기자 및 팩트체커에 대한 신변 보호, 전략적 봉쇄 소송(SLAPP) 대응 등 최소한의 보호 장치를 제공하는 데 우선 활용될 수 있다. 이를 통해 공익적 검증 활동이 위축되지 않고 지속될 수 있는 환경을 조성할 필요가 있다. 동시에 기금의 신뢰성을 확보하기 위해, 거버넌스 구조, 재원 배분 기준, 성과 및 회계 공개, 이해충돌 방지 장치 등을 제도화하여 정치적 중립성과 공정성을 제고하는 것이 중요하다.

현재 국내에서는 지속적인 팩트체크 활동이 제한적으로 이루어지고 있어, 정보 무결성 정책의 실효성을 높이기 위해서는 관련 활동의 저변 확대와 생태계 조성이 요구된다. 이를 위해 단일한 조직 육성보다는, 다양한 주체가 참여할 수 있는 환경을 조성하는 방향으로 단계적인 활성화 방안을 모색할 필요가 있다.

우선, 언론사·독립 팩트체커·학계·시민사회 등이 참여할 수 있도록 팩트체크 활동에 대한 안정적인 운영 지원과 프로젝트 단위 지원을 병행하는 방안을 검토할 수 있다. 상시 조직에 대한 직접 지원보다는, 주제별·사안별 팩트체크 프로젝트 공모, 협업형 컨소시엄 지원 등을 통해 진입 장벽을 낮추고 활동 주체의 다양성을 확보하는 방식이 바람직하다.

아울러 데이터 접근성과 협업 인프라를 개선하는 것이 중요하다. 플랫폼이 보유한 공개

데이터, 확산 경로 정보, 광고·추천 관련 통계 등에 대한 표준화된 접근 경로를 마련하고, 연구·검증 목적의 데이터 이용을 지원함으로써 팩트체크의 정확성과 효율성을 제고할 수 있다. 이는 개별 콘텐츠 판단에 대한 개입이 아니라, 검증 활동을 가능하게 하는 환경 조성 차원의 지원으로 이해될 수 있다.

또한 전문 인력 양성과 네트워크 구축도 병행할 필요가 있다. 언론인, 연구자, 데이터 분석가 등을 대상으로 한 교육·훈련 프로그램과 국제 팩트체크 네트워크와의 교류를 지원함으로써, 국내 팩트체크 역량을 점진적으로 강화할 수 있다.

3. 거버넌스 고도화: 다층적 공동규제 체계 구축

정보 무결성 정책이 중장기적으로 안정적으로 정착하기 위해서는, 단일 전담기관에 권한과 책임을 집중하는 방식보다는 기능별 역할 분담과 상호 견제를 전제로 한 다층적 거버넌스 체계를 단계적으로 구축할 필요가 있다. 특히 허위조작정보 대응과 같이 표현의 자유, 기술 변화, 플랫폼 책임이 복합적으로 얽힌 영역에서는 단일 기관 중심의 집행 모델이 정책 신뢰성 논란이나 집행 부담으로 이어질 가능성이 있다.

이에 따라 독립적 민간기구 모델과 초부처 독립위원회 모델을 비교·검토하되, 정부의 역할은 법제도 정비, 재정 기반 마련, 최소한의 절차적 의무 부과 및 이행 점검 등으로 한정하는 방향이 바람직하다. 이는 정부가 개별 콘텐츠나 판단 과정에 직접 개입하기보다는, 제도가 안정적으로 작동할 수 있는 제도적 틀과 여건을 조성하는 역할에 집중하도록 하기 위한 것이다.

이와 함께 민간 중심의 감독·조정 기구가 공통 표준 마련, 이행 평가, 정책 연구, 이해관계자 간 협력 조정 기능을 담당하고, 복수의 수행기관 또는 센터가 감시, 연구, 교육, 데이터 분석, 보고 등의 기능을 분담 수행하는 구조를 고려할 수 있다. 이러한 방식은 특정 기능이나 판단 권한이 한 기관에 과도하게 집중되는 것을 방지하고, 각 영역별 전문성을 활용할 수 있다는 점에서 제도 운영의 효율성을 높일 수 있다.

플랫폼 사업자는 데이터 제공, 신고·조치 절차 준수, 위험 평가 및 완화 조치 이행 등 공동규제의 핵심 당사자로 참여함으로써 제도의 실효성을 제고할 수 있다. 동시에 외부 검증 주체 및 연구자와의 협력 체계를 통해 정책 효과에 대한 점검과 개선이 지속적으로 이루어질 수 있도록 할 필요가 있다.

이와 같은 다층적 공동규제 거버넌스는 정부의 직접적 판단이나 개입에 대한 우려를 완화하는 한편, 공익적 목표 달성을 위한 책임과 역할 분담을 명확히 함으로써 제도의 투명성, 신뢰성, 지속가능성을 함께 제고하는 기반으로 기능할 수 있을 것으로 기대된다.

제6장. 결 론

본 연구는 디지털 환경에서 확산되고 있는 허위조작정보에 대한 기존 규율 체계의 한계를 점검하고, 해외 주요국의 정책 동향과 국내 제도 환경을 종합적으로 분석하여 향후 정보 무결성 정책의 방향을 검토하였다.

허위조작정보는 단순한 허위 사실의 문제를 넘어, 의도적 조작과 사회적 해악을 수반하는 복합적 현상으로서, 개별 콘텐츠의 사후적 삭제나 처벌만으로는 효과적인 대응이 어렵다는 점이 국제적으로 공통 인식으로 자리 잡고 있다.

해외 사례 분석 결과, 유럽연합, 영국, 독일, 호주 등 주요국은 전통적인 국가 주도 콘텐츠 규제 방식에서 벗어나, 플랫폼의 책임을 절차·시스템 중심으로 구조화하고 자율규제와 법적 규율을 결합한 공동규제(co-regulation) 또는 메타규제(meta-regulation) 모델로 정책 패러다임을 전환하고 있는 것으로 나타났다. 이들 국가는 국가가 직접 정보의 진위나 적법성을 판단하기보다는, 투명성 보고, 위험 평가 및 완화, 외부 검증, 데이터 접근 보장 등 제도적 장치를 통해 플랫폼의 구조적 책임을 점검하고 사후적으로 감독하는 방식을 채택하고 있다. 이러한 접근은 표현의 자유 침해와 정치적 오남용에 대한 우려를 완화하는 동시에, 디지털 정보 환경 전반의 신뢰성을 제고하려는 정책적 선택으로 평가할 수 있다.

국내의 경우, 그간 허위조작정보에 대한 규율은 형법, 정보통신망법, 공직선거법 등 개별 법령에 따른 사후적·영역별 대응을 중심으로 이루어져 왔으며, 방송미디어통신심의위원회와 방송미디어통신위원회로 구성된 이원적 심의 구조를 통해 공적 규제 권한이 비교적 강하게 작동해 왔다. 이러한 구조는 불법 정보에 대한 신속한 대응이라는 측면에서는 일정한 기능을 수행해 왔으나, 허위조작정보 전반을 체계적으로 포섭하고 예방적·절차적으로 관리하는 데에는 한계가 있다는 지적도 지속적으로 제기되어 왔다.

이러한 맥락에서 2025년 12월 24일 정보통신망법 개정안의 국회 통과는 중요한 제도적 전환점으로 평가된다. 개정된 정보통신망법은 허위조작정보를 명시적으로 규율할 수 있는 법적 근거를 마련하고, 유통 금지 규정과 함께 신고·조치·이의제기·분쟁조정, 투명성 보고, 사실확인단체 연계 등 절차적 대응 체계를 법률 차원에서 도입하였다. 이는 기존의 사후적·개별적 규율 방식에서 벗어나, 보다 구조화되고 제도화된 대응 체계로 이행하기 위한 출발점을 마련했다는 점에서 의미를 갖는다.

다만, 개정법의 실효성과 제도에 대한 사회적 신뢰를 확보하기 위해서는 향후 시행령·고시 등 하위 규범을 통한 구체적 설계가 결정적인 영향을 미칠 것으로 보인다. 특히 사실확인단체 연계, 투명성센터 운영, 공익 연구 목적의 데이터 접근 보장 등은 제도 설계 방식에 따라 정부의 직접 판단이나 개입으로 오인될 소지가 있는 영역인 만큼, 기능적 경계와 운영 원칙을 명확히 설정할 필요가 있다. 이는 제도의 정치적 중립성과 절차적 정당성을 확보하기 위한 필수적인 조건으로 평가된다.

중장기적으로는 정보 무결성 정책이 단기적 대응이나 개별 입법에 머무르지 않고, 플랫폼 책임, 공익적 검증 생태계, 재정의 지속가능성, 독립적 운영 구조가 유기적으로 결합된 공동규제 체계로 발전할 수 있도록 제도적 기반을 단계적으로 고도화할 필요가 있다. 전문가 의견에서도 확인되었듯이, 제도의 신뢰성과 지속가능성을 확보하기 위해서는 법제정비, 재원 조성, 거버넌스 설계를 상호 연계된 축으로 통합적으로 추진하는 접근이 요구된다.

결론적으로, 향후 허위조작정보 대응 정책은 개별 콘텐츠에 대한 사후적 판단과 제재를 중심으로 한 방식에서 나아가, 해당 정보가 확산되는 구조와 조건을 점검하고 개선하는 방향으로 전환될 필요가 있다. 이를 위해 법적 규율, 자율규제, 공동규제 수단을 유기적으로 결합하고, 투명성, 절차적 정당성, 독립성 확보를 핵심 원칙으로 삼는 정책 체계를 구축하는 것이 바람직하다. 이러한 접근은 표현의 자유와 공익 보호 간의 균형을 유지하면서, 디지털 정보 생태계 전반의 신뢰 회복과 민주적 공론장의 기능 강화를 도모하는 데 기여할 것으로 기대된다.

참 고 문 헌

김민주・김현경(2025.4). 인공지능 투명성 규제에 대한 비판적 고찰 – EU AI법 투명성 의무 조항 분석을 중심으로 –, 선진상사법률연구 110.

김일기(2025.6). 가짜뉴스 및 허위조작정보에 대한 대응전략, INSS 전략보고 No.339.

김현수 등(2023.12), 플랫폼 자율규제 활성화 연구, KISA-WP-2023-0031.

김희경 외(2022). 인터넷 공동규제 활성화 방안 연구, KCSC 2022-006.

문의빈(2025). 알고리즘 시대 공론장의 위기와 표현의 자유, 헌법학연구 31(3).

박아란(2019). ‘가짜뉴스’와 온라인 허위정보 규제에 대한 비판적 검토, 언론정보연구 56(2).

오세욱(2017). “자동화된 사실 확인(fact checking) 기술(technology)의 현황과 한계, 사이버커뮤니케이션학보 34(3).

오세욱(2023), 허위정보 대응을 위한 기술적 방안 검토와 정책적 제안, MEDIA 정책리포트 6.

유승익(2023.7). 인공지능이 인권과 민주주의에 미치는 영향과 인공지능법안의 쟁점, 민주법학 82.

이수엽 외(2023), 온라인 가짜뉴스 근절을 위한 사업자 자율규제 방안 연구, 방통융합정책연구 KCC-2023-19.

이수엽 외(2024). 허위조작정보 자율규제 실천기반 및 사업자 책무강화 방안 마련에 관한 연구, 방통융합정책연구 KCC-2024-21

이인영(2020). 딥페이크와 여론형성 – 알고리즘의 권력화와 탈진실의 규제담론 –”, 미국헌법연구 31(1).

정인숙 외(2022). 방송영상콘텐츠사업자의 자율규제 실효성 제고 방안, KCSC 2022-002.

정재황(2024), 디지털, AI 시대의 헌법, 세계헌법연구 30(2)

Club de MADRED, Protecting Information Integrity – National and International Policy

Options, 2018

EU Digital Service Act / 세계법제정보센터, 디지털서비스법 영문본, 2023.12.1.

Federal Trade Commission, Opinion of the Commission in the Matter of Cambridge Analytica, LLC, Docket No. 9383, 2019

Fukuyama, Francis, et al., Middleware for dominant digital platforms: A technological solution to a threat to democracy, CyberPolicy Center, Freeman Spogli Institute, 2020

Gillespie, Tarleton, Governance of and by platforms, in: Burgess, Jean; Poell, Thomas and Marwick Alice (eds), The SAGE Handbook of Social Media, Sage Publications, 2017

Lara Monday · Lorenzo Strappelli, "Does procenance build trust?", BBC, 2024.5

OECD, Recommendation of the Council on Information Integrity, 2025

https://www.oecd.org/en/publications/facts-not-fakes-tackling-disinformation-strengthening-information-integrity_d909ff7a-en.html

Ovadya, Aviv, and Luke Thorburn, Bridging-based ranking, Harvard Kennedy School Belfer Center for Science and International Affairs, 2022

UK Online Safety Act / 세계법제정보센터, 2023 온라인 안전법 영문본, 2025.2.3.

UN, United Nations Global Principles For Information Integrity – Recommendations for Multi-stakeholder Action –, 2024

https://www.un.org/en/information-integrity

https://www.un-ilibrary.org/content/books/9789211065756

UNDP, Information Integrity : Forging a pathway to Truth, Resilience and Trust – Envisioning comprehensive and effective responses to information Pollution, Strategic Guidance, 2022

U.S. Department of Justice and Federal Trade Commission, Complaint for Civil Penalties, Injunction, and Other Relief, Facebook, Inc., No. 1:19-cv-02184. U.S. District Court for the District of Columbia, July 24, 2019.

U.S. Department of Justice and Federal Trade Commission, Stipulated Order for Civil Penalty, Monetary Judgment, and Injunctive Relief, Facebook, Inc., No.

1:19-cv-2184, U.S. District Court for the District of Columbia, July 24, 2019

Vosoughi, Soroush, et al., The Spread of True and False News Online, Science, vol. 359, no. 6380, 2018

유엔 정보무결성 글로벌 원칙

다자이해관계자 행동을 위한 권고안

목차

p.3

디지털 시대의 정보 생태계

기술 발전은 짧은 기간 수십 년 만에 커뮤니케이션에 혁명을 일으켜 개인과 공동체를 이전에는 상상할 수 없는 규모로 연결했으며 지식의 확산, 문화적 풍요, 지속가능 발전을 위한 비할 데 없는 기회를 제공했습니다.

이러한 발전은 여러 면에서 정보 생태계의 무결성에 대한 열망을 높였습니다. 이는 표현의 자유가 완전히 향유되고 차별과 혐오가 없는 정확하고 신뢰할 수 있는 정보가 개방적이고 수용적이며 안전하고 보호받는 정보 환경에서 모두에게 제공되는 곳을 의미합니다.

이러한 발전이 정보의 대량 확산을 가능하게 했지만동시에 전례 없는 규모, 속도, 확산성으로 다양한 행위자들이 오보(misinformation), 허위 정보(disinformation), 혐오 발언(hate speech)을 퍼뜨리는 것을 용이하게 하여 정보 생태계의 무결성을 위협하고 있습니다. 이러한 위험에는 인공지능 기술의 급속한 발전 속에서 현재, 새롭게 등장하는, 그리고 미래의 다양한 위협들이 포함됩니다.

디지털 플랫폼의 무결성 훼손은 사람들의 인권 행사 능력을 약화시키고 지구의 평화, 번영, 살기 좋은 미래를 향한 노력이 저해될 수 있습니다. 이처럼 정보 무결성을 강화하는 과제는 우리 시대의 가장 시급한 도전 과제 중 하나입니다.

정보 무결성은 인권, 평화로운 사회, 지속가능한 미래를 옹호하는 다원적인 정보 공간을 수반합니다. 이는 모두를 위한 신뢰, 지식 및 개인의 선택을 증진하는 디지털 시대의 약속을 담고 있습니다.

정보 무결성 증진은 사람들이 모든 종류의 정보와 아이디어를 찾고, 받고, 전달할 권리와 방해 없이 의견을 가질 권리를 행사할 수 있는 역량을 강화하는 것을 포함합니다. 점점 더 복잡해지는 디지털 정보 환경에서, 이는 개인 정보 보호와 자유를 가지고 안전하게 정보 공간을 탐색할 수 있도록 하는 것을 의미합니다.

p.4

정보 무결성과 지속가능발전목표

정보 무결성을 강화하려는 노력은 지속가능발전목표를 보존하고 더욱 발전시키는 데 중요합니다. 정보 생태계의 무결성 훼손은 특히 개발도상국(Global South)에서 지속가능발전목표 달성에 있어 기존의 취약점이 더욱 악화될 수 있습니다.
취약하고 소외된 상황에 있는 집단은 불균형적으로 큰 영향을 받습니다. 예를 들어, 전 세계 노동 시장에서 여성의 참여를 늘리는 것은 지속가능발전목표를 달성하는 데 필수적입니다. 그러나 많은 국가에 존재하는 차별적인 법률 및 정책 외에도성별에 기반한 혐오 발언, 허위 정보 및 폭력이 여성을 침묵시키고 공적인 영역에서 배제함으로써 체계적으로 억압하는 데 사용됩니다. 이는 여성의 참여에 심각한 지속적인 결과를 초래할 수 있으며 여성의 목소리를 억압하고 자기 검열을 부추기며 직업적 평판 손상을 일으키고 어렵게 얻어낸 성 평등의 진전을 위태롭게 합니다.

기후 행동을 약화시키기 위해 정보 공간을 악용하려는 시도는 이 문제의 시급성을 더욱 부각시키고 있습니다상업적 이익을 위해 종종 주도되는 조직적인 허위 정보 캠페인은 인간 활동으로 인한 기후 변화, 그 원인 또는 영향에 대한 과학적 근거를 부인하거나 의심을 조장하여 기후 목표 달성을 위한 조치를 지연시키거나 좌절시키려고 합니다. 사회 활동가, 과학자, 방송인 등 공인들은 기후 위기에 대해 정보를 제공하고 해결책을 모색하려는 노력 때문에 혐오 발언, 협박 및 괴롭힘의 표적이 되고 있습니다.

건강, 기아 종식에서부터 평화, 정의, 교육, 불평등 감소에 이르기까지 지속가능발전목표의 전 범위에 걸쳐 정보 무결성을 강화하기 위한 조치들은 지속 가능한 미래를 달성하고 그 누구도 소외되지 않도록 하는 노력을 더욱 강화할 것입니다.

p.5

유엔 정보 무결성 글로벌 원칙을 위하여

유엔은 모든 지역의 회원국, 청년 단체를 포함한 시민 사회, 언론, 학계 및 민간 부문 대표들을 대상으로 정보 무결성에 대한 광범위하고 다양한 협의를 진행했습니다. 이해관계자들은 국가별 논의, 가상 세션, 양자 회담 및 전 세계에 배포된 공개 온라인 설문 양식을 통해 의견을 제시했습니다.

이러한 협의를 통해 모든 지리적 위치와 맥락에 적용 가능하고 모든 개인의 요구 사항, 특히 취약하고 소외된 상황에 있는 집단의 필요 사항을 고려하는 통합적인 권고안에 대한 수요가 강조되었습니다.

이에 대응하여, 유엔 정보 무결성 글로벌 원칙은 더 건강한 정보 생태계를 위한 다자간 이해관계자들의 행동을 안내하는 포괄적인 정책 프레임워크를 제공합니다. 이 정책 지침은 정보 무결성을 강화하기 위한 다섯 가지 원칙으로 구성되며, 각 원칙에는 주요 이해관계자 그룹에 대한 권고 사항이 포함되어 있습니다.

이 원칙들은 다음과 같습니다사회적 신뢰와 회복력; 독립적이고 자유로우며 다원적인 언론; 투명성과 연구; 대중의 역량 강화; 그리고 건전한 인센티브. 이 모든 원칙은 인권에 대한 확고한 헌신을 공통적으로 바탕에 두고 있습니다.

이 글로벌 원칙은 국가, 시민 사회, 민간 부문 및 기타 이해관계자들이 이미 이룩한 광범위한 노력과 발전을 인정하고 이를 기반으로 합니다. 이 원칙은 모든 삶의 영역, 모든 언어 및 맥락에서 정보의 진실성을 보호하고 증진하기 위한 통일된 출발점을 제공하며, 전례 없는 규모, 속도 및 강도로 요구되는 글로벌 연대와 광범위한 대응의 필요성을 인식합니다.

글로벌 원칙은 개인, 공공 및 민간 단체, 유엔 시스템, 정부, 언론, 시민 사회 단체, 그리고 기술, 광고 및 홍보 분야의 영리 기업 포함한 모든 주체들이 국제법에 명시된 권리와 자유에 부합하고 정보 무결성을 위한 광범위한 연합을 형성할 기회를 제공합니다.

글로벌 원칙은'우리의 공동 의제(Our Common Agenda)'와 유엔 사무총장의 정책 보고서8: 디지털 플랫폼의 정보 무결성에서 제안된 아이디어를 기반으로 합니다.

국제 인권법을 포함한 국제법에 근거를 두고 있을 뿐만 아니라, 글로벌 원칙은 관련 유엔 기업과 인권 이행 원칙, 유네스코 디지털 플랫폼 거버넌스 가이드라인, 유엔 언론인 안전 및 불처벌 문제에 관한 행동 계획, 유네스코 인공지능 윤리 권고, 혐오 발언에 관한 유엔 전략 및 행동 계획 등 관련 문서들을 보완합니다.

글로벌 원칙은 미래를 위한 협약(A Pact for the Future)과 글로벌 디지털 협약(Global Digital Compact)을 고려하는 유엔 회원국에게 유용한 자료을 제공합니다.

이러한 방식으로 글로벌 원칙은 정보 무결성 강화를 향한 유엔의 확고한 의지를 더욱 반영하며 조직의 향후 유엔의 활동을 이끌어가는 지침이 될 것입니다.

p.7

유엔 정보 무결성 글로벌 원칙

사회적 신뢰와 회복력; 독립적이고 자유로우며 다원적인 언론; 투명성과 연구; 대중의 역량 강화; 그리고 건전한 인센티브

p.8

사회적 신뢰와 회복력

사회 전반의 신뢰와 회복력은 정보 무결성의 핵심 구성 요소입니다. 여기서 신뢰란 사람들이 정보의 출처와 신뢰성, 공식 출처 및 정보, 그리고 정보가 생태계 전반에 걸쳐 유통되도록 하는 메커니즘에 대한 확신을 의미합니다. 회복력은 사회가 정보 생태계 내에서 발생하는 혼란이나 조작 행위에 대처할 수 있는 능력을 의미합니다.

신뢰와 회복력은 전략적, 정치적 또는 재정적 이익을 위해 정보 생태계를 악용하려는 국가 및 비국가 행위자의 행동에 취약합니다. 때때로 광범위하게 조직적으로 이루어지며, 다양한 피해를 초래하고 사람들이 과학과 사실을 비판적으로 평가하는 능력을 위협할 수 있습니다.

빅테크 기업은 정보 생태계에서 상당한 힘을 가지며 다른 기업, 광고주, 언론 매체 및 개인 사용자를 포함한 이해관계자들이 정보와 상호 작용하고 정보에 접근하는 방식에 막대한 영향력을 행사합니다. 생성형AI와 같은 인공지능(AI) 기술의 발전은 정보 공간에 대규모로 그리고 최소한의 비용으로 위험을 초래할 수 있는 수단을 제공했습니다. 실제 또는 원본인 것처럼 보이는AI 생성 또는AI를 통해 조작된 콘텐츠는 매우 설득력 있고 감정적으로 호소력이 있으며 탐지하기 어렵고, 알고리즘 기반 플랫폼과 미디어 매체를 통해 빠르게 확산될 수 있습니다. 이는 신뢰 부족을 기하급수적으로 생성, 가속화하며 심화시킬 수 있는 가능성이 있습니다.

정보 무결성에 대한 위험에 대처하려면 강력하고 미래지향적이며 혁신적인 디지털 신뢰 및 안전한 업무 수행이 필요하며, 언어와 상황에 관계없이 일관

되게 적용되어야 합니다.이러한 절차는 잠재적 피해에 불균형적으로 노출되는 취약하고 소외된 상황에 있는 집단에 대한 이해와 본질을 반영해야 합니다.

여성, 노인, 어린이, 청소년, 장애인, 원주민, 난민 및 무국적자, 성소수자(LGBTIQ+) 및 소수 민족 또는 종교 집단을 특별히 고려해야 합니다.

많은 청소년과 어린이들은 삶의 상당 부분을 온라인에서 보내며 디지털 채널에서 방대한 범위의 정보를 얻습니다. 이들은 이미 정보 공간에서 발생하는 위험에 가장 취약하며, 새롭게 등장하는 기술과 미디어 트렌드에 가장 직접적인 영향을 받을 것입니다.

사람들은 일반적으로 다양한 정보 출처에 접근 할 수 있고, 사회적으로 소속감을 느끼며, 평등하고, 사회경제적으로 안정되고, 정치적으로 권한을 부여받았다고 느낄 때 이러한 위험에 더욱 잘 대처하고 예방할 수 있는 회복력을 갖추게 됩니다. 그렇지 않은 경우, 이러한 위험은 더욱 쉽게 확산될 수 있습니다. 따라서 대응책은 장기적인 회복력을 강화하기 위한 근본적인 사회적 요구를 고려해야 합니다.

공익을 위해 행동하겠다고 다짐한 모든 이해관계자들은 끊임없이 진화하는 커뮤니케이션 환경의 현실에 적응하기 위해 정보 공간을 공동의 이익을 위해 활용하는 데 힘써야 합니다. 이는 선거, 자연 재해, 인재와 같은 중요한 사회적 순간에 특히 중요합니다. 이러한 시기에는 정보 공간에 대한 위험이 두드러지게 나타나 사회적 양극화를 심화시키고, 사람들이 공적 생활에 참여할 수 있는 능력을 저해하며, 극단적인 경우에는 폭력을 선동하는 데 악용될 수 있기 때문입니다.

사회 활동가, 언론인, 인도주의자, 그리고 평화유지군, 선거 관계자, 과학자, 의료 전문가 등을 포함한 유엔 직원들은 표적이 될 수 있으며, 이는 심각한 결과를 초래할 수 있습니다. 온라인 괴롭힘 및 기타 교활한 전술은 목소리를 침묵시키고 시민 사회 공간을 위축시킬 수 있습니다. 이러한 개인들을 보호하기 위한 공동의 노력이 무엇보다 중요합니다.

p.10

건전한 인센티브

건전한 인센티브를 조성하려면 타겟 광고 및 기타 콘텐츠 수익화 방식을 주요 수익 창출 수단으로 삼는 현재 비지니스 모델이 정보 생태계의 무결성에 미치는 중대한 영향에 대해 다루어야 합니다.

이러한 모델은 모든 규모의 기업, 특히 디지털 플랫폼을 소유하고 운영하는 테크 기업들에게 전례 없는 성장 기회를 제공했으며, 수많은 사람들에게 혜택을 주는 크리에이터 경제를 성장시켰습니다. 또한 이러한 모델들은 허위 정보와 혐오를 퍼뜨리는 사람들에게 재정적 인센티브와 기회를 제공했는데, 이들은 테크 기업들이 사용자 행동을 추적하여 데이터를 수집하고, 광고주와 크리에이터의 잠재적 수익을 극대화하기 위해 참여도를 우선시하는 알고리즘에 데이터를 제공하는 관심 경제를 악용합니다. 양극화를 조장하고 강한 감정을 불러일으키도록 설계된 메시지는 종종 가장 많은 참여를 유발하며, 그 결과 알고리즘은 유해한 콘텐츠를 보상하고 확산시키는 결과를 초래했습니다.

이러한 비즈니스 모델을 악용하는 행위자에는 다음이 포함됩니다. 국가, 정치인 및 민간 기업으로부터 계약을 받아 조직적인 조작 캠페인을 제공하는 정보 조작자 및 주류 홍보 기업들이 있으며, 이러한 캠페인은 때때로 국경을 넘어 이루어지기도 합니다.

기술 업계는 디지털 광고 프로세스를 복잡하고 불투명하게 그리고 인간의 개입을 최소화하는 방식으로 설계했습니다. 이는 광고 기술(ad tech) 공급망에 참여하는 많은 주체들에게 유리하며, 특히 빅테크 기업들이 가장 큰 이익을 얻고 있습니다.

이러한 불투명한 설계는 광고 예산이 광고주가 지원할 의도가 없었던 개인, 단체 또는 아이디어에 의도치 않게 사용되는 결과를 초래할 수 있으며, 이는 브랜드에 상당한 위험이 될 수 있습니다. 이러한 광고 게재 방식은 광고 캠페인의 효과와 브랜드 안전에도 부정적인 영향을 미칠 수 있습니다.

광고 기술 시장을 장악하고 있는 소수의 기업들은 동시에 자신들이 소유한 플랫폼에서 광고 표준을 시행하는 책임도 지고 있지만, 이러한 표준 시행은 제대로 이루어지지 않거나 일관적이지 않을 수 있습니다.

이러한 정보 생태계 무결성의 훼손은 인센티브 구조의 근본적인 변화가 필요함을 보여줍니다. 이러한 변화는 인권을 기반으로 하고 행동 추적 및 개인 데이터에 기반한 알고리즘 기반의 타겟 광고에 의존하지 않는 비즈니스 모델을 통해 이루어질 수 있습니다.

광고주들은 정보 생태계를 강화하고 동시에 사업적으로도 이익이 되는 방식으로 정보 생태계에 기여할 수 있습니다. 테크 기업이 현재의 사업 모델을 쉽게 포기할 가능성은 낮지만, 광고주들에게 광고 프로세스에 대한 더 큰 투명성을 제공하고 광고 전달자들이 인권 존중 광고 정책을 준수하도록 함으로써 더욱 건전한 인센티브 구조를 구축할 수 있습니다. 투명한 공급망에 대한 통제력을 강화함으로써 광고주들은 투자 대비 더 나은 수익을 얻을 수도 있습니다.

p.12

대중의 역량 강화

정보 생태계 속에서 살아가는 개인들이 역량을 발휘하려면 온라인 경험에 대한 통제권을 갖고, 소비할 미디어에 대해 정보에 입각한 결정을 내릴 수 있으며, 자유롭게 자신을 표현할 수 있어야 합니다. 대중의 역량 강화는 다양하고 신뢰할 수 있는 정보원에 대한 지속적인 접근을 필요로 합니다.

디지털 공간은 여러 면에서 공공 생활에 대한 포용적인 참여를 위한 촉매제 역할을 해왔으며, 지리적 경계를 넘어선 사람들이 진보를 향한 공동의 열망을 공유하며 연결될 수 있도록 해줍니다. 이러한 공간을 올바르게 활용한다면 개인에게 힘을 실어주고 종종 배제되고 소외되는 사람들에게 주체성을 부여하는 데 도움이 될 수 있습니다.

동시에 디지털 기술은 진정한 역량 강화에 걸림돌이 될 수 있습니다. 개인은 자신의 개인 데이터가 어떻게 사용되는지또는 빅테크 기업이 제공하는 알고리즘 기반의 맞춤형 콘텐츠에 대해 거의 통제권을 행사할 수 없으며, 정보 제공자가 특정 유형의 콘텐츠를 우선 수위에 두고 홍보하는 데 사용하는 기준과 메커니즘을 이해하고 접근하는 데 어려움을 겪습니다.

테크 기업은 사용자의 개인 정보 보호 권리를 존중하면서 신뢰 및 안전, 개인 정보 보호 정책 및 데이터 사용의 모든 측면에 대해 사용자가 의견과 피드백을 제공할 수 있도록 지원해야 합니다. 또한 다양한 제공 업체의 서비스와의 상호 운용성을 포함하여 사용자의 통제권와 선택권을 강화해야 합니다.

미디어, 정보 및 디지털 문해력(digital literacy) 교육 계획 및 정책은 모두의 역량을 강화하는 데 초점을 맞춰야 하며특히 여성, 노인, 어린이, 청소년, 장애인 및 취약하고 소외된 상황에 있는 집단이 직면하는 특정 과제에 중점을 두어야 합니다.

인터넷 연결성이 증가하고 있지만, 전 세계 인구의3분의1은 여전히 인터넷에 접속하지 못하고 있습니다. 인터넷에 접속하는 사람들 조차도 불충분한 접근성으로 인해 인터넷의 자원을 온전히 활용하지 못하고 있으며, 이는 정보 공간에서 발생하는 위험에 취약하게 만듭니다. 연결성 장벽이 빠르게 무너지고 있는 만큼 새로운 인터넷 사용자들에게 역량을 강화하고, 인터넷 접근성이 부족한 사람들에게 안전하고 생산적인 온라인 경험에 필요한 디지털 문해력 기술을 제공하기 위한 노력이 필요합니다.

p.14

독립적이고 자유로우며 다원적인 언론

정보 무결성은 독립적이고 자유로우며 다원적인 언론이 있어야만 달성할 수 있습니다.

언론의 자유는 법치주의의 근간이며 민주 사회의 초석으로서 정보에 기반한 시민 담론을 가능하게 하고, 권력에 대한 책임을 물고, 인권을 보호하는 역할을 합니다. 언론은 여성과 취약하고 소외된 사람들을 포함한 언론인과 언론 종사자들이 안전하고 공개적으로 보도하고 활동할 자유를 지속적으로 누릴 수 있고, 모든 개인이 다양하고 신뢰할 수 있는 뉴스 출처에 일관되게 접근할 수 있을 때 자유롭다고 할 수 있습니다.

언론은 신뢰할 수 있고 정확한 정보를 제공하고 정보 공간에 대한 위험을 완화하는 데 특별한 역할과 책임이 있습니다. 그러나 표현의 자유, 특히 검열 없이 자유롭게 방해받지 않는 언론 및 기타 매체의 자유에 대한 권리에도 불구하고 전 세계적으로 언론의 자유는 심각하고 지속적인 위협에 직면해 있습니다. 언론인들은 온라인과 오프라인에서 괴롭힘, 위협 및 폭력에 직면하며, 이는 때때로 자기 검열로 이어지고 직업적 위험을 가중시킵니다.

동시에, 뉴스 산업은 거대한 테크 기업들이 장악하고 있는 디지털 공간으로의 광고 수익이 이동하면서 어려움을 겪고 있습니다이러한 요인들로 인해 기업들은 언론 매체에 대한 영향력을 더욱 강화하게 되었고, 이는 언론의 다양성을 위협하고 지역 및 공익 저널리즘을 약화시키고 있습니다. 편집 기준이 엄격하게 지켜지지 않는 경우, 언론 매체는 정보 무결성에 대한 위험을 조장하고 증폭시킬 수 있으며, 이러한 위험은 온라인과 오프라인 공간을 넘나들며 확산될 수 있습니다.

공익 언론 기관,언론인 및 미디어 종사자들을 지원하기 위해서는 강력하고 시급한 대응이 필요하며, 특히 미디어 인프라가 부족한 환경에서 시민 언론인이 지역 주민들에게 필수적인 서비스를 제공하는 상황을 고려해야 합니다. 이러한 대응에는 지역 실행 기관을 활용한 강력하고 지속적인 미디어 개발 지원이 포함될 수 있습니다. 국가와 테크 기업들은 정보 흐름과 정책 형성에 상당한 영향력을 행사하므로,언론의 자유와 언론인의 안전을 보장하기 위한 노력을 강화해야 합니다.

p.16

투명성과 연구

테크 기업 및 기타 정보 제공 업체의 투명성 증대는 정보가 어떻게 확산되는지, 개인 데이터가 어떻게 사용되는지, 그리고 정보 무결성에 대한 위험이 어떻게 다루어지는지에 대한 더 나은 이해를 가능하게 할 수 있습니다.

그러나 힘의 불균형은 투명성을 저해하는 장벽을 만듭니다. 소수의 테크 기업들은 전례 없는 양의 데이터에 접근할 수 있으며, 일부 미디어 소유주들과 함께 정보 생태계에서 상당한 영향력을 행사하고 있으며때로는 국가, 정치 및 경제 주체자들과 긴밀한 관계를 맺고 있습니다.

더욱이 대다수의 테크 기업 본사가 위치한 소수의 국가에서 이루어지는 투명성에 대한 규제 결정은 전세계에 지대한 영향을 미칩니다. 이러한 불균형은 종종 공익 연구를 제한하고, 형평성을 보장하고 소외되고 연구가 부족한 환경과 지역사회의 요구를 충족시키려는 노력을 저해할 수 있습니다.

아직 그 영향이 완전히 알려지지 않은AI 기술의 도입은 정보 생태계를 연구하고 이해하는 데 또 다른 어려움을 더합니다.

정보 환경에 대한 보다 세밀하고 정교한 글로벌 이해를 구축하고 정보 무결성 증진을 위한 목표 지향적이고 증거 기반의 조치를 강화하려면 데이터와 통찰력의 가용성, 품질 및 활용성을 확대해야 합니다. 다양한 연구자들이 개인 정보 보호를 보장받으며 데이터에 접근할 수 있도록 하는 것은 연구 격차와 불평등을 해소하기 위한 공동의 노력을 강화하는데 도움이 될 것입니다. 학계, 언론인 및 시민 사회는 두려움이나 괴롭힘 없이 그들의 중요한 업무를 수행할 수 있도록 보호받고 지원받아야 합니다.

p.18

행동 촉구 권고안

다음 권고 사항의 목표는 다섯 가지 원칙을 정보 생태계 전반의 이해관계자들이 실행할 수 있는 구체적인 조치로 전환하는 것입니다. 이러한 권고 사항은 포괄적인 청사진으로서, 국가의 법적 의무부터 기술 부문의 책임, 그리고 언론 및 시민 사회를 위한 모범 사례에 이르기까지 다양한 내용을 담고 있습니다.

이해관계자들을 위한 권고안

- 테크 기업
- 인공지능(AI) 주체자
- 광고주와 기타 민간 부문 주체자
- 언론 매체
- 연구자와 시민 사회 단체
- 국가와 정치 관련 주체자
- 유엔(United Nations)

p.19

테크 기업

기술 규제가 미비한 지역에 본사를 둔 빅테크 기업들은 막대한 영향력을 행사합니다. 이들은 사용자 행동에 대한 방대한 데이터를 수집하여 수익을 창출하며, 이를 통해 국경을 초월한 정보의 흐름을 형성하고 전 세계적인 규모로 디지털 경험을 통제합니다.

이러한 힘의 불균형을 바로잡기 위해서는 투명성과 독립적인 감독을 모두 우선시하는 정책 지침이 필요합니다. 사용자들은 자신의 데이터와 온라인 경험에 대한 통제권을 가져야 하며, 명확한 불만 제기 및 구제 절차가 마련되어야 합니다. 테크 기업들이 자사 제품 및 서비스의 설계와 사용이 인권과 사회 통합에 미치는 영향, 특히 위기 및 분쟁 상황에서 발생하는 영향에 대해 책임을 지도록 하는 책임 메커니즘이 필요합니다.

이를 위해서는 플랫폼 설계 구조(platform architecture)에 대한 비판적이고 투명한 평가를 통해 정보의 무결성을 훼손하고 인권을 침해하는 기능을 식별해야 합니다. 이러한 훼손을 예방하고 완화하기 위한 전략은 표현의 자유와 정보 접근권을 보호하면서 동시에 시행되어야 합니다.

허위 정보와 혐오 발언이 최대 노출과 막대한 수익을 창출해서는 안 됩니다. 타겟팅된 프로그램 방식 광고에 의존하지 않는 새로운 상업적으로 실행 가능한 비즈니스 모델은 혁신을 촉진하고 사용자 권한을 을 강화하며 공익에 기여할 수 있습니다. 이러한 다각적인 접근 방식은 사용자 권리를 존중하고 신뢰할 수 있는 온라인 환경을 조성하는 보다 균형 잡힌 정보 생태계를 만들 수 있습니다.

권고안

a. **설계부터 제공까지 안전과 개인정보 보호를 통합적으로 고려합니다.** 모든 제품 및 서비스의 전체 수명 주기, 즉 설계, 개발, 제공 및 폐기 단계 전반에 걸쳐 강력한 안전 및 개인정보 보호 정책을 적용하고, 이러한 정책을 인간이 생성한 콘텐츠와 AI가 생성한 콘텐츠 모두에 일관되게 적용합니다. 독

립적인 제3자 기관과 협력하여 모든 제품 및 서비스와 관련된 인권 위험 평가를 지속적으로 수행하고 그 결과를 공개함으로써 사회적 위험을 사전에 최소화하고 잠재적 피해를 완화하며, 특히 중요한 사회적 사건 발생 이전과 발생 시점에 이러한 노력을 기울입니다. 취약하고 소외된 상황에 처한 집단, 시민 사회 구성원 및 온라인에서 자주 표적이 되는 사람들을 보호하고 역량을 강화하기 위한 조치를 취하고, 기술 사용을 통해 발생하거나 증폭되는 성 폭력 및 기타 형태의 폭력을 해결해야 합니다. 인공지능 기술로 인해 발생할 수 있는 정보 생태계 무결성 위험을 포함하여 새롭게 발생하는 문제들을 해결하기 위해 혁신적인 노력을 기울여야 합니다. 제품 개발의 모든 단계, 신뢰 및 안전 담당 팀에서 다양성과 포용성을 보장해야 합니다. 잠재 위험 및 정책 평가가 경영진을 포함한 기업의 모든 직급과 기능에서 공유되고, 모두가 함께 이해되도록 내부 정보 공유 절차를 수립해야 합니다. 모든 상호 신뢰 및 안전 정책이 일관되게 시행되도록 보장해야 합니다.

b. **사업 모델을 재평가 합니다.** 플랫폼 설계 구조가 정보 생태계 무결성 훼손하고 인권을 침해하는지 여부와 그 방식을 평가하고, 표현의 자유를 존중하면서 비례적인 완화 및 시정 조치를 취합니다. 타겟 프로그램식 광고에 의존하지 않고 공익에 기여하는 혁신적이고 상업적으로 실행 가능한 사업 모델을 모색합니다.

c. **어린이를 보호합니다.** 연령 확인 및 부모 통제와 같은 조치를 마련하고 시행하여 어린이 보호의 권리를 보호하고 보장합니다. 기술 사용을 통해 발생하거나 악화되는 아동 성 착취 및 학대를 예방하고 대응하기 위한 정책과 관행을 시행합니다. 어린이를 위한 특별 신고 및 불만 접수 메커니즘을 확립하고 공개합니다.

d. **자원을 확보합니다.** 위험 수준에 비례하는 수준의 전담 내부 신뢰 및 안전 관련 자원과 전문성을 지속적으로 충분히 확보합니다. 사회문화적 언어적 맥락과 운영 언어 그리고 취약하고 소외된 상황에 있는 집단의 차별화된 요구 사항을 해결하는 데 필요한 충분한 자원을 배정합니다. 특히 분쟁을 겪거나 불안정한 상황에 처한 지역에서는 더욱 그러해야 합니다.

e. **일관성 있는 콘텐츠 검열을 보장합니다.** 독립적인 제3자 기관과 협력하여 국제 인권 기준에 부합하는 콘텐츠 검열 절차를 개발하고, 이러한 정책이 운

영 지역 전반에 걸쳐 일관되고 자의적이지 않게 시행되도록 보장합니다. 인력 및 자동화된 콘텐츠 검열 및 큐레이션에 충분한 자원을 할당하고, 모든 언어 및 운영 환경에 걸쳐 일관되게 적용합니다. 플랫폼 커뮤니티 기준을 위반하고 인권을 침해하는 콘텐츠에 대해 알고리즘에 의한 확산 제한, 콘텐츠 라벨링 및 수익 창출 제한 등의 조치를 취합니다. 콘텐츠 검열 정책 시행 및 콘텐츠 검열에 할당된 자원에 대한 세부 데이터를 언어 및 운영 환경별로 구분하여 공개합니다.

f. **노동 기준을 준수합니다.** 국제 노동 및 인권법에 부합하는 근무 환경을 제공하고, 신뢰 및 안전 관련 업무에 참여하는 콘텐츠 검토자를 포함한 모든 근로자의 복지, 안전 및 양질의 교육을 보장하는 새로운 정책 및 제안을 우선적으로 추진합니다.

g. **독립적인 감독 체계를 구축합니다.** 서비스 약관 및 커뮤니티 표준신뢰 및 안전 정책,광고 정책위험 관리, 언어 및 운영 환경 전반에 걸친 광고 및 추천 시스템의 영향콘텐츠 검열블만 및 이의 제기절차투명성 메커니즘그리고 연구자를 위한 데이터 접근성 등을 포함하는 정기적인 외부 인권 독립 감사를 의뢰합니다.제품 및 서비스가 취약하고 소외된 집단성평등그리고 아동 권리에 미치는 영향을 평가합니다. 이러한 감사 결과를 모든 사용자가 접근하고 이해할 수 있도록 공개합니다.

h. **산업 표준을 개발합니다.** 시민 사회 및 기타 이해관계자와 협력하여 명확하게 정의된 역할과 책임을 포함하는 산업 책임 프레임워크를 공동으로 개발하고, 감사를 거친 공개 보고와 독립적인 감독, 그리고 개인정보 보호, 투명성, 위험 관리, 신뢰 및 안전에 대한 강력한 표준을 준수할 것을 약속합니다. 취약하고 소외된 계층과 불안정한 환경에 있는 사람들의 필요 사항을 특별히 고려하고, 인권에 대한 위험을 측정하고 해결하는 효과적인 방법을 마련합니다. 위험은 다양한 정보 공간에 걸쳐 확산될 수 있으며, 각 공간에는 악용될 수 있는 고유한 설계 결함과 정책적 허점이 존재한다는 점을 인식하여 플랫폼과 서비스 간의 협력을 보장해야 합니다..

i. **위기 대응 역량을 강화합니다.** 고위험 지역에서 활동하는 이해관계자들과 협력하여 위기 및 분쟁 상황에서 신속하고 시의적절한 대응을 위한 조기 경보 및 비상 대응 절차를 구축합니다. 공익에 부합하는 신뢰할 수 있고 정확

한 정보에 대한 접근성을 높이고 시의적절하게 제공할 수 있는 장치를 마련합니다.

j. **정치 과정을 지원합니다.** 선거 및 기타 정치적 과정 전반에 걸쳐 모든 제품 및 서비스에 대한 인권 위험 평가를 수행하고 그 결과를 공개적으로 접근 가능하도록 합니다. 정보의 무결성을 유지하기 위해 관련 정책을 모두 시행하고, 허위 정보, 괴롭힘, 여성 및 정치 후보자를 포함하여 공공 생활에서 흔히 표적이 되는 다른 집단에 대한 폭력에 대처하기 위한 조치를 취해야 합니다.

k. **이해관계자들과 협력합니다.** 국가, 학계, 시민 사회, 어린이, 청소년 주도 단체 및 기술 커뮤니티를 포함한 다양한 이해관계자들과 적극적으로 소통하여 정보 생태계의 무결성에 대한 위험을 더 깊이 이해하고, 이에 따라 신뢰 및 안전 장치를 강화하고 조정합니다.

l. **효과적인 불만 처리 장치를 구축합니다.** 사용자 및 비사용자를 위한 투명하고 안전하며 접근 가능한 불만 접수, 신고, 이의 제기 및 구제 메커니즘을 적시에 제공하도록 보장합니다. 여기에는 취약 계층 및 소외 계층을 위한 특별 절차도 포함됩니다. 허위 정보 유포와 같은 조직적인 부적절한 행위를 통해 신고 및 불만 처리 장치가 악용되는 것을 방지하기 위한 절차를 수립하고 시행합니다.

m. **명확한 정책을 전달합니다.** 이용 약관, 정책, 커뮤니티 표준 및 규제 절차를 어린이를 포함하여 누구나 쉽게 접근하고 이해할 수 있도록 일관성 있게 제공합니다. 뉴스 및 정치 콘텐츠와 관련된 모든 정책, 지침 및 규칙을 명확하게 제시합니다.

n. **광고 정책을 시행합니다.** 광고 및 콘텐츠 수익화에 대한 명확하고 강력한 정책을 수립하고, 이를 공개하고 시행합니다. 기존 게시자 및 광고 기술 파트너십을 지속적으로 검토하여 광고 기술 공급망의 파트너들이 이러한 정책을 준수하는지 평가합니다. 정책 시행의 효과 및 기타 조치에 대해 매년 공개적으로 보고합니다.

o. **광고 투명성을 확보합니다.** 모든 광고를 명확하게 표시하고, 광고주 정보, 타겟팅에 사용된 매개변수, 그리고AI로 생성되거나AI로 제작된 콘텐츠 사용

여부에 대한 정보를 사용자에게 투명하게 공개합니다. 광고 출처 또는 구매자, 지출 금액 및 타겟 고객에 대한 정보를 포함하여 완전하고 접근 가능하며 최신 상태를 유지하고 검색 가능한 광고 라이브러리를 운영합니다. 광고주와 연구자에게 특정 기간 동안 광고가 게재된 정확한 위치, 광고 게재 및 브랜드 인접성 관련 제어 및 서비스의 정확성과 효과에 대한 자세한 데이터를 제공합니다. 광고주 및 콘텐츠 제작자와의 수익원 및 수익 분배 계약에 대해 투명하게 보고합니다. 모든 정치 광고를 명확하게 표시하고, AI로 생성되거나AI를 통해 제작된 콘텐츠임을 명시하며, 수신자가 타겟팅된 이유, 광고 비용을 지불한 주체 및 금액에 대한 정보를 쉽게 접근할 수 있도록 제공합니다.

p. **미디어 안전과 다양성을 지원합니다.** 다양한 뉴스 콘텐츠 배포를 위한 환경을 조성하여 소비자들이 다양한 미디어 소스에 접근할 수 있도록 합니다. 편집의 독립성을 존중하면서 독립적이고 자유롭고 다원적인 언론, 특히 다양한 언어와 맥락에서 활동하는 지역 언론 및 시민 저널리즘을 지원합니다. 온라인에서 언론인과 미디어 종사자의 권리를 보호하기 위한 모든 조치를 취합니다. 언론인, 특히 선거, 자연재해 및 인재와 같은 중요한 사회적 상황에서 언론인이 직면하는 위험을 반영하여 언론인과 미디어 종사자를 괴롭힘, 학대 및 폭력 위협으로부터 보호하기 위한 명확하고 투명한 규정을 마련합니다. 여성 언론인을 표적으로 삼는 행위를 완화하고 해결하기 위해 신뢰 및 안전 정책과 관행을 업데이트합니다.

q. **데이터 접근성을 제공합니다.** 학계를 비롯한 다양한 분야의 연구자, 언론인, 시민 사회 및 국제 기구 관계자들이 정보 무결성을 더 잘 이해하고, 정책 및 모범 사례를 수립하며, 책임성을 강화하는 데 필요한 데이터에 접근할 수 있도록 지원합니다. 이때 사용자 개인정보 보호 및 지적재산권을 존중해야합니다. 이러한 데이터는 정보 생태계의 무결성을 효과적으로 연구할 수 있도록 세분화되어야 하며, 여기에는 사회적 위험, 다양한 공동체 및 인구 집단에 미치는 영향, AI 기술 사용의 함의, 지속가능발전목표 달성에 미치는 잠재적 영향, 그리고 위험 완화 조치의 효과 등이 포함됩니다. 데이터에는 다음을 포함해야 합니다. 알고리즘이 콘텐츠를 순위 지정, 추천, 배포 및 플래그 지정에 어떻게 사용되는지에 대한 설명 포함한 알고리즘 기반 추천 시스템에 대한 정보; 삭제, 차단 또는 순위가 낮아진 계정, 그리고 언어 및 상황별 신뢰 및 안전을 위한 자원 배분에 대한 정보가 포함되어야 합니다. 연

구자들이 최소한의 비용으로 접근 가능하고 기계 판독 가능한 형식으로 데이터를 제공 받을 수 있도록 지원합니다.

r. 정보 공개를 보장합니다. 콘텐츠 삭제 또는 게재에 대한 정부의 요청을 공개해야 합니다. 자금 지원 또는 기타 지원을 포함하여 사실 확인(fact-checking) 기관과의 모든 협력 관계를 공개하고, 정치 단체 및 후보자에게 제공된 자금 지원 내역도 공개합니다.

s. 통제권과 선택권을 제공합니다. 사용자가 정보에 입각한 동의를 할 수 있도록 보장하고, 다른 서비스와의 상호 운용성을 포함하여 사용자가 자신의 온라인 경험을 쉽게 제어할 수 있도록 지원하는 사용자 친화적인 도구, 기능 및 기능을 제공합니다. 이를 통해 사용자는 더 많은 선택권을 갖고 자신이 보는 콘텐츠와 데이터가 어떻게, 어디에서 사용되는지에 대해 정보에 입각한 동의를 할 수 있습니다.

t. AI 콘텐츠를 별도 표기합니다. AI가 생성하거나 개입한 콘텐츠에 대해 명확하게 라벨을 붙이고, 조직 차원에서 솔루션을 투자하고 개발하여 사용자가 이러한 콘텐츠를 쉽게 식별할 수 있도록 하고, 정보 생태계 무결성에 대한 사용자 신뢰를 약화시키기보다는 강화해야 합니다. 여기에는 해당 콘텐츠가AI가 생성하거나 개입되었음을 나타내는 메타데이터 정보가 포함됩니다.

u. 개인정보 보호를 보장합니다. 데이터의 수집, 사용, 공유, 판매 및 저장 과정에서 사용자의 개인정보 보호를 존중하고, 사용자가 자신의 개인 데이터가 알고리즘 기반 의사 결정을 포함하여 어떻게 활용되는지, 다른 기관과 어떻게 공유되고 다른 기관으로부터 어떻게 제공받는지에 대한 정보를 쉽게 확인할 수 있도록 보장해야 합니다.

v. 디지털 문해력(digital literacy)을 증진합니다. 디지털 기술 향상을 위해 미디어 및 정보 문해력 교육을 지원하고,알고리즘의 기능, 영향 및 의미에 대한 대중의 이해를 높여야 합니다. 모든 언어와 지역, 특히 취약한 환경에 문해력 및 역량 강화 자원을 제공합니다. 어린이와 청소년에게 안전 관련 교육 자료를 제공합니다. 디지털 문해력 교육 시행 정책의 효과에 대한 독립적인 외부 평가를 실시하고 그 결과를 공개적으로 제공합니다.

p.25

인공지능(AI) 주체자

인공지능 시스템 수명 주기의 최소 한 단계에 관여하는 정부, 민간 및 공공 부문 관계자

인공지능(AI) 기술이 빠르게 발전함에 따라, 이러한 기술은 우리 세상을 근본적으로 변화시킬 잠재력을 가지고 있습니다. 일상적인 업무 자동화부터 과학적 발견 지원에 이르기까지, 그 잠재적 이점은 매우 광범위합니다. 그러나 이러한 발전과 더불어 안전하고, 보안이 확보되며, 신뢰할 수 있는AI 시스템의 설계, 개발, 배포 및 사용 중단을 보장해야 하는 것이 매우 중요합니다.

훈련 데이터의 편향성과 다양성 부족은AI 시스템이 오해를 유발하는 정보를 생성하고 불공정성을 영속화하는 결과를 초래할 수 있습니다. 이러한 현실적인 콘텐츠 생성 능력은 대규모로 악용될 경우 정보 생태계의 무결성에 위협을 가할 수 있습니다.

새롭게 부상하는 위험은 AI 기술의 수명 주기 전반에 걸쳐 투명성과 공정성을 우선시함으로써 완화할 수 있습니다. 인공지능이 수명 주기 전반에 걸쳐 안전하고 책임감 있게 설계, 개발, 배포 및 폐기되도록 보장하기 위해서는 정부, 테크 기업, 학술 및 연구 기관 간의 협력적인 노력이 필요합니다. 이러한 이해관계자들이 협력함으로써 인공지능 기술이 사회와 인류 복지에 기여할 수 있도록 보장할 수 있습니다.

권고안

a. 안전하고 보안성이 뛰어나며 신뢰할 수 있는AI 기술을 보장합니다. AI 기술의 안전하고, 보안성이 뛰어나며 신뢰할 수 있는 설계, 개발, 배포, 사용 및 사용 중단을 보장하기 위한 조치를 취해야 합니다. 정보 생태계의 무결성에 위험을 초래할 수 있는 해당 분야의 모든 혁신이나 발전의 의미를 파악하고 공개적으로 알려야 합니다. 여기에는AI 기술의 악의적인 사용, 인간의 감독 없이 인공지능 기술에 과도하게 의존하는 행위, 그리고 이로 인해 지역 및 사회적 맥락 전반에 걸쳐 신뢰가 더욱 저해될 수 있는 모든 잠재적 가능

성이 포함됩니다. 공공 복지에 중요한 문제에 대한 신뢰할 수 있고 포괄적인 정보 출처를 활용하여AI를 훈련시키고, 성별 및 인종적 편견을 포함하여 훈련 데이터에서 비롯되는 편향을 완화하기 위한 조치를 취해야 합니다. 다양한 이해관계자들과 협력하여 인권 위험 평가를 수행하고, 사회적 위험을 사전에 최소화하고 여성, 어린이, 청소년 및 취약하고 소외된 상황에 처한 다른 집단을 포함한 잠재적 피해를 완화합니다.

b. 독립적인 감사를 의뢰합니다. 기관 및 개인 연구자들이 예를 들면 기업의 취약점 공개 정책 준수와 같은 적절한 안정장치를 통해 AI 모델에 대한 독립적인 감사를 수행할 수 있도록 접근 권한과 법적 및 기술적 안전망을 제공할 것을 약속합니다. 독립 감사 결과, 유해한 차별 가능성 및'환각'(사실적으로 보이지만 완전히 꾸며낸 콘텐츠)과 같은AI 시스템 관련 위험에 대한 데이터, 그리고 잠재적 피해를 예방, 완화 및 해결하기 위해 취한 조치에 대한 정보를 대중에 접근할 수 있도록 보장해야 합니다.

c. 지적 재산권을 존중합니다. AI 도구 학습에 사용되는 독창적인 저널리즘 콘텐츠를 포함한 지적 재산권 사용에 대해 공정한 보상이 이루어지도록 보장함으로써 지적 재산권을 존중해야 합니다.

d. 데이터 출처 정보를 표시합니다. 진위 확인 인증, 워터마킹 및 라벨링과 같은 가시적 및 비가시적 형태를 통해 출처 정보에 대한 솔루션과 정책을 개발하고 이행해야 합니다. 사용자 친화적인 라벨링 표준화를 위해 여러 이해관계자가 참여하는 노력을 기울입니다.

e. 디지털 리터러시 향상을 지원합니다. 조직 차원에서 AI 모델의 작동 방식과 전 세계 정보 소비자에게 미치는 영향, 특히 정보 무결성에 대한 위험을 대중이 더 잘 이해할 수 있도록 하는 디지털 문해력 증진 정책의 개발 및 배포에 투자해야 합니다

f. 사용자 피드백 기능을 활성화 합니다. 사용자 개인정보를 보호하면서 사용자가 부정확하거나 오해의 소지가 있는 출처 정보에 대한 알리거나 신고할 수 있는 기능을 제공해야 합니다.

p.27

광고주

광고주는 허위 정보와 협오 발언으로 이익을 얻으려는 사람들에게 재정적 인센티브를 차단함으로써 정보 생태계의 무결성에 지대한 영향을 미칠 수 있습니다. 이를 통해 광고주들은 브랜드 이미지를 더욱 효과적으로 보호하고 실질적인 위험을 관리하며, 기업 가치에 부합하는 방식으로 사업을 운영하면서 수익성을 향상시킬 수 있습니다.

권고안

a. 인권 존중 광고 관행을 확립합니다. 광고가 정보 공간에 위험을 초래하지 않고 아동의 권리를 포함한 인권을 존중하도록 안전 장치를 마련해야 합니다. 민감한 데이터 및 사용자의 인지된 특성을 기반으로 한 차별적인 타겟팅 관행을 피해야 합니다. 포함 및 배제 목록, 광고 검증 도구 및 수동 검토와 같은 방법을 통해 공익 저널리즘을 포함한 정보의 진실성을 강화하는 미디어 매체 및 플랫폼에 광고합니다. 광고 기술 기업에게 웹사이트나 채널이 수익을 창출하기 전에 준수해야 하는 기준을 공개하도록 요구해야 합니다.

b. 산업 표준을 활용합니다. 산업 표준을 활용하여 정보 무결성에 대한 위험을 최소화하고 브랜드 안전을 보장하는 데 도움이 되는 명확한 정책을 개발해야 합니다.

c. 연합을 결성합니다. 업계 전반 및 시민 사회와 협력하여 정보 무결성에 대한 모범 사례와 교훈을 시의적절하게 공유하고, 광고의 영향 평가 및 광고와 콘텐츠 수익 창출로 인해 발생하는 위험과 잠재적 피해를 체계적으로 완화하는 방안을 모색합니다.

d. 데이터를 요구합니다. 광고 게재 위치에 대한 완전하고 상세한 개요를 지속적으로 파악하기 위해 광고가 게재된 위치를 보여주는 세부적인 데이터를 확보하고, 광고 게재 전에 적합성 검토를 수행해야 합니다. 광고 캠페인에 대한 철저한 감사를 실시해야 합니다.

e. 투명성을 의무화합니다. 광고 기술 기업들이 공급망의 종단 간 검증을 가능하게 하는 투명성 표준을 채택하고, 광고 게재 및 차단 데이터와 같은 로그 수준 데이터를 포함한 전체 광고 캠페인 데이터를 고객 및 연구원과 공유하도록 요구해야 합니다.

f. 감사를 실시합니다. 광고 기술 기업에게 독립적인 제3자 감사를 실시하고 광고 거래소 및 공급 파트너에 대한 검증을 수행하도록 요구해야 합니다.
p.29

기타 민간 부문 주체자

기술 분야에 직접적으로 관여하지 않는 더 광범위한 민간 부문 주체들의 행동은 정보 공간에 영향을 미칠 수 있으며, 정보 무결성을 약화시키거나 강화하는 데 모두 기여할 수 있습니다. 기업은 표현의 자유와 정보에 대한 권리를 포함한 인권을 존중할 책임이 있으며, 더 건강한 정보 생태계를 조성하기 위해 다른 이해관계자들과 협력적인 파트너십을 구축할 수 있습니다.

권고안

a. 무결성을 유지합니다. 표현의 자유와 의견의 자유를 포함한 인권을 존중하고, 재정적 또는 기타 전략적 목표를 위해 정보 생태계의 무결성을 위협하는 행위를 고의적으로 확산시키거나 지원하는 것을 삼가야 합니다.

b. 디지털 리터러시에 투자합니다. 조직 차원에서 직원들을 위한 미디어 및 정보 리터러시 교육에 투자하고, 관련 시민 단체와 협력하여 그들의 전문성을 활용해야 합니다.

p.30

뉴스 매체

독립적이고, 자유로우며 다원적인 언론은 공공의 이익과 관련된 사안에 대해 대중에게 정보를 제공하고, 시민 참여를 촉진하며, 권력자들의 책임을 증진하는 데 있어 매우 중요한 역할을 합니다.

그러나 언론의 독립성, 자유 및 다양성에 대한 직접적 및 간접적 위협과 지역 및 공익 저널리즘의 쇠퇴는 이러한 중요한 기능을 약화시킬 수 있습니다. 전문적인 기준이 엄격하게 유지되지 않는다면 뉴스 매체는 정보 무결성을 훼손할 수 있습니다. 윤리적인 보도 및 편집 관행과 투명성에 대한 헌신, 그리고 양질의 교육과 근무 환경을 통해 뒷받침되는 언론인들은 필수적인 서비스를 제공하며, 정보 생태계의 신뢰성을 위협하는 위험에 맞서 균형을 회복하는 데 도움을 줄 수 있습니다.

권고안

a. 정보 무결성 보장에 집중합니다. 데이터 기반 저널리즘 및 탐사 저널리즘 역량 강화에 투자하여 정보 생태계의 무결성에 대한 위험에 대해 대중에게 적극적으로 알리고 보도해야 합니다. 정보 출처 확보를 포함한 강력한 편집 절차와 기준을 적용하여 매체 소비자들 사이의 신뢰를 유지하고 확보하도록 노력해야 합니다. 대중을 위한 참고 자료로 활용될 수 있는 사실 확인 시스템을 구축해야 합니다.

b. 위기 대응을 제공합니다. 정보 생태계의 무결성에 대한 위험이 고조될 수 있는 비상 및 위기 상황에서 대중에게 비용 없이 시의적절한 정보를 제공하기 위해 노력합니다.

c. 전문적이고 윤리적인 기준을 유지합니다. 공익을 위해 제작되는 전문적이고 윤리적인 저널리즘의 전 세계적으로 인정받는 규범과 관행을 준수하고, 공정성과 편집 독립성을 강조하며, 자율, 규제, 책임 매커니즘을 적극적으로 도입해야 합니다. 윤리적이고 정확하며 공정한 보도를 증진하고, 커뮤니케이션 환경 변화에 대한 혁신과 적응력을 향상시키기 위해 정기적으로 양질의

교육을 제공해야 합니다. 여기에는'해결책 중심' 또는'건설적 저널리즘' 접근 방식을 채택하는 것도 포함됩니다. 자금 출처, 소유 구조 및 재정적 인센티브를 투명하게 공개하여 개인이 선택하고 소비하는 뉴스에 대해 더 잘 이해할 수 있도록 해야 합니다.

d. AI를 윤리적으로 사용합니다. AI 기술의 윤리적 사용을 위한 강력한 정책을 수립하고, 게재하거나 방송할 때AI가 생성하거나 처리한 자료임을 명확하게 표시해야 합니다. 여기에는 해당 콘텐츠가AI로 생성되거나 처리되었음을 식별하는 정보가 메타데이터에 포함되어야 합니다.

e. 투명하고 인권을 존중하는 광고 시스템을 확립합니다. 광고가 정보 공간에 위험을 초래하지 않도록 조치를 취해야 합니다. 뉴스, 의견 및 스폰서 콘텐츠를 명확하게 구분하고, 의견 기사의 자금 출처 및 잠재적 이해 상충 관계에 대한 투명성을 보장해야 합니다. 모든 유료 및AI로 생성하거나 개입한 광고 및 광고성 콘텐츠를 명확하게 표시해야 합니다. 광고 수익 출처에 대한 투명한 보고와 명확하고 접근 가능한 광고 정책 및 관행을 제공해야 합니다.

f. 노동 기준을준 존중해야 합니다. 국제 노동 및 인권법에 부합하는 근무 조건을 확립하고, 디지털 공간을 포함하여 언론인의 복지 및 안전을 보장하는 데 도움이 되는 정책을 우선적으로 추진하며, 여성 언론인 및 언론 종사자에 대한 차별, 학대, 괴롭힘 및 폭력 위협에 특별한 주의를 기울여야 합니다.

p. 32

연구원과 시민 사회

연구원과 시민 사회 단체들은 정보 생태계의 무결성에 대한 위험이 미치는 다면적인 영향을 이해하고 해결하려는 노력에 있어 핵심적인 역할을 합니다. 이러한 노력은 정보 공간에 대한 위험을 밝혀내고, 지원 활동을 위한 근거를 강화하며, 특히 취약하고 소외된 상황에 처한 집단을 위해 회복력을 증진하는 데 도움이 될 수 있습니다. 협력적인 파트너십과 지식 교환은 연구 결과와 효과적인 해결책 사이의 격차를 해소하는 데 필수적입니다.

권고안

a. 협력합니다. 다양한 지역과 배경의 이해관계자들과 협력하여 정보 생태계의 무결성을 강화하기 위한 효과적이고 윤리적인 접근 방식을 공유해야 합니다.

b. 정직성과 윤리적 기준을 준수합니다. 인권을 존중하고, 정보 무결성에 대한 위험을 고의로 확산시키거나 조장하는 행위를 삼가야 합니다. 모든 연구는 윤리적이고 투명하며 개인정보 보호를 고려하는 방식으로 수행해야 합니다.

c. 오픈 액세스를 장려합니다. 연구 결과를 자유롭게 이용할 수 있도록 하는 오픈 엑세스 정책을 추진하고, 학문 분야간 협력을 증진합니다.

d. 융합 연구를 강화합니다. 다양한 지역, 언어 및 주제 분야를 걸친 정보 생태계에 대한 여러 학문 분야의 연구를 탐구하고, 정보 무결성 위험이 지속가능발전목표에 미칠 수 있는 잠재적 영향에 대한 연구하며, 특히 연구가 부족한 취약하고, 소외된 환경과 공동체에 초점을 맞춥니다.이러한 위험과 관련 피해를 측정하기 위한 엄격한 방법론을 개발해야 합니다.

p.33

사실 확인 기관 및 네트워크를 위한 권고안

a. 전문적 기준을 유지합니다. 직업 전문성과 윤리 의식을 준수하고, 조직 구성 및 운영 방식, 자금 출처, 소유 구조 및 업무 관행에서 드러나는 독립성, 정치적 중립성 및 투명성 기준을 철저히 지켜야 합니다.

b. 자금 출처를 공개합니다. 테크 기업, 언론 매체 및 시민 사회 단체와 같은 이해관계자와의 모든 협력 관계를 공개적으로 투명하게 공개하기 위한 조치를 취해야 합니다.

p.34

국가

국가는 정보 무결성을 위한 글로벌 원칙을 강화하는 데 있어 필수적인 책임을 지고 있습니다. 이는 특히 정보를 찾고, 받고, 전달할 권리를 포함한 표현의 자유를 비롯한 인권을 존중하고 보호하며 증진해야 하는 국가의 의무에서 시작됩니다.

국가는 법적 및 규제적 권한, 공공 자원에 대한 통제권, 국내외 연합 구축 능력 등 여러 요인 덕분에 정보 공간을 형성하는 데 핵심적인 역할을 합니다. 국가는 인권 의무의 일환으로 자국 영토 및/또는 관할권 내에서 기업에 의한 인권 침해로부터 국민을 보호해야 하며, 효과적인 정책, 법률, 규정 및 사법 절차를 통해 그러한 침해를 예방, 조사, 처벌 및 구제하기 위한 적절한 조치를 취해야 합니다.

각국은 정보 생태계에 관여하는 데 있어 서로 다른 기술적, 재정적 역량에 차이가 있습니다. 인프라 부족과 기술 및 재정 자원에 대한 접근성 격차는 디지털 격차를 심화시키는 요인이 되었습니다. 동시에, 많은 빅테크 기업들은 전 세계 시장에 거의 완벽하게 진출하여 지배적인 위치를 차지하고 있지만, 이러한 기업들은 대부분 북반구(global North)의 소수 국가에 본사를 두고 있습니다.

모든 국가가 정보 생태계에 기여하고 그 혜택을 누릴 수 있도록 보장하기 위해서는, 국가들이 디지털 연결성을 확대하고, 잠재적인'AI 격차' 발생을 사전에 방지하며, 인권을 존중하면서 정보 공간의 위험에 적절히 대처할 수 있는 역량을 강화하기 위한 시급하고 지속적인 노력이 필요합니다. 궁극적으로 이러한 노력은 정보 무결성을 강화하고, 인권을 증진하며, 지속가능발전 목표 달성에 기여할 것입니다.

권고안

a. 인권을 존중하고 보호하며 증진합니다. 특히 국제 인권 기준 및 규범에 따라 표현의 자유와 의견의 자유, 정보에 대한 권리를 포함한 인권을 존중하

고 보호하며 증진합니다. 정보 무결성의 다양한 측면을 다루기 위해 시행되는 규제 또는 기타 조치들이 국제 인권법을 포함한 관련 국제 인권법을 포함한 관련 국제법을 준수하고, 시민 사회의 완전한 참여를 보장하며 인권 강화 및 신뢰 구축을 위한 더 광범위한 노력의 일환이 되도록 해야 합니다. 표현의 자유에 대한 제한은 예외적인 경우에만 허용되어야 하며, 제한이 부과되는 경우 국제 인권법에 따른 요건을 충족해야 합니다. 즉, 법률에 의해 규정되어야 하고, 타인의 권리 또는 명예 보호, 국가 안보, 공공 질서, 공중 보건 또는 도덕 보호를 위해 필요해야 하며, 비례의 원칙을 준수해야 합니다. 제한 조치가 실제로 표현의 자유를 억압하는 수단으로 작용하지 않도록 보장해야 합니다. 국제 인권법을 포함한 국제법에 부합하는 개인 데이터 보호 조치를 채택하고 효과적으로 시행해야 합니다.

b. 정보 생태계의 무결성을 보호합니다. 국내외를 막론하고 고의로 허위 정보를 유포하거나 협오 발언을 조장하는 정보 작전을 수행하거나 후원하는 행위를 삼가해야 합니다. 어떠한 형태의 인터넷 차단이나 속도 제한도 자제해야 합니다. 유엔 평화 유지 활동의 임무 수행에 영향을 미치는 정보 생태계 무결성 위협으로부터 유엔 평화 유지 활동을 보호하는 것과 관련된 결의안을 포함하여 관련 유엔 안전보장이사회 결의안을 준수하고 이행해야 합니다.

c. 인구를 보호합니다. 법적으로나 실질적으로 취약하고 소외된 상황에 처해 있으며 온라인 및 오프라인 정보 공간에서 종종 표적이 되는 집단,예를 들어,여성, 성소수자 또는 소수 민족이나 종교 집단의 보호와 권익 증진을 보장하기 위한 노력을 재확인하고 강화하는 동시에 어린이의 특별한 필요와 권리를 다루어야 합니다. 국제 인권법에 따라 전쟁 선전이나 차별, 적대감 또는 폭력을 선동하는 국가적, 인종적, 또는 종교적 증오를 조장하는 행위를 법으로 금지해야 하는 의무를 준수해야 합니다.

d. 정보 접근권을 보장합니다. 모든 언어와 형식으로 모든 사람이 이해하고 활용할 수 있는 방식으로, 차별 없이 공공 정보에 대한 시의적절한 접근을 제공해야 합니다. 여기에는 뉴스 매체에 대한 접근 보장도 포함되며, 소외된 지역 사회의 정보 접근성을 증진해야 합니다. 위기 상황에서도 신뢰할 수 있고 정확한 정보에 접근 할 수 있도록 보장해야 합니다. 윤리적이고 신뢰할 수 있는 커뮤니케이션 방식을 채택하여 지역 사회와 적극적으로 소통하고

공공 기관에 대한 신뢰를 구축해야 합니다.

e. 언론의 자유를 보장합니다. 자유롭고, 지속 가능하며, 독립적이고 다원적인 언론 매체 환경을 보장, 보호 및 증진하고, 언론인, 언론 종사자 및 팩트 체커를 보호하기 위한 강력한 조치를 취해야 합니다. 특히 여성과 취약하고 소외된 집단 구성원을 모든 형태의 차별, 학대, 괴롭힘 및 폭력 위협으로부터 보호하는 데 특별한 주의를 기울여야 합니다. 디지털 콘텐츠 크리에이터와 시민 언론인의 권리를 법과 정책으로 존중하고 보호해야 합니다.

f. 연구자와 시민 사회를 보호해야 합니다. 학문의 자유를 존중하여 학계 및 시민 사회 구성원을 위협, 괴롭힘 또는 보복 행위로부터 보호해야 합니다.

g. 투명성을 제공합니다. 테크 기업 및 언론 기관에 부과되는 요구 사항 및 데이터 요청에 대해 완전한 투명성을 제공해야 합니다. 비윤리적인 고용 관행 및 재정적 인센티브와 같이 정보 무결성을 훼손하는 비투명하고 기만적인 로비 전술과 테크 기업 및 정책입안자 간의 이해 상충 문제를 해결하기 위한 조치를 취해야 합니다.

h. 글로벌 연대, 역량 구축 및 개발 지원을 을 강화합니다. 정보 무결성을 강화하고 정보 공간 위험에 대한 회복력을 높이기 위해 국가 간 협력 및 파트너십에 참여합니다. 특히 개발도상국에서 모든 언어로 디지털, 정보 및미디어 리터러시 및AI 기술을 포함한 인식 프로그램의 교육 및 역량 강화를 위해 재정 자원을 투명하게 할당해야 합니다. 정보 생태계의 무결성에 대한 위험에 대비하는 사회적 회복력을 구축하고, 견고한 언론 및 정보 활용 능력 교육을 실시하며, 공익 언론을 강화하기 위한 국가 주도 노력들을 지원해야 합니다. 이를 위해 전적으로 충분한 개발 지원을 제공해야 합니다. 도서관을 포함한 공공 기관들이 디지털 문해력 교육 및 관련 자료에 대한 접근성을 개선하는 데 기여하도록 지원해야 합니다.

i. 정치 참여를 장려합니다. 선거 과정 전반에 걸쳐 모든 선거 이해관계자들이 정확하고 시의적절한 정보에 접근할 수 있도록 보장해야 합니다. 포용적인 정치 참여와 리더십을 증진하고, 모든 형태의 차별, 학대, 괴롭힘 및 폭력 위협으로부터의 보호를 포함하여 공공 생활에서 여성의 권리를 옹호하기 위한 조치를 취해야 합니다.

j. 포용적이고 공익적인 연구를 우선시합니다. 정보 무결성과 관련된 다양한 분야에 걸쳐 윤리적 기준과 검토 절차를 준수하는 독립적인 연구를 우선적으로 지원하고 투자합니다. 특히 AI 기술의 새롭게 등장하는, 그리고 아직 알려지지 않은 역량과 영향력을 고려하여 연구를 진행해야 합니다. 다양한 지리, 언어, 주제 영역을 아우르는 연구를 지원하고, 정보 생태계 무결성에 대한 위험이 지속가능발전목표에 미치는 잠재적 영향에 대한 연구를 장려해야 합니다. 특히 소외되고 연구가 부족하며 취약한 환경과 공동체에 초점을 맞춰야 합니다. 연구 결과에 대한 개방적인 접근을 장려하고 널리 알려서 국가들 간에 정보가 공평하게 공유될 수 있도록 합니다.

k. 디지털 문해력을 증진합니다. 특정 언론 매체 및 정보 문해력 활용을 통해 비판적이고 정보에 입각한 공적 담론을 조성합니다. 어린 나이부터 정규 및 비정규 교육 과정에 디지털 역량을 자연스럽게 통합해야 합니다. 모든 연령대와 배경을 가진 사람들의 특정 사회적, 문화적, 언어적 요구를 고려하여 온라인 권리, 디지털 정보 환경의 작동 방식, 개인 데이터 사용 방식에 대한 대중의 이해와 인식을 적극적으로 향상시켜야 합니다. 여성, 어린이, 청소년, 노인, 장애인 및 앞으로 온라인에 접속할 수십억 명의 사람들을 포함하여 취약하고 소외된 상황에 있는 개인 및 집단의 정보 활용 능력 필요성을 우선적으로 고려해야 합니다. AI 기술과 관련된 특정 문제에 대한 리터러시 노력을 수행하고, 새롭고 개발중인 기술 및 도전 과제를 반영하도록 리터러시 교육 내용을 지속적으로 업데이트해야 합니다.

l. 어린이, 부모, 보호자 및 교육자에게 역량을 강화할 수 있도록 지원합니다.어린이, 부모, 보호자 및 교육자에게 안전하고 책임감 있는 디지털 행동, 온라인 매체 활용 방법, 그리고 어린이의 표현의 자유와 정보 접근권에 대한 이해를 돕는 지속적인 자료를 제공해야 합니다. 모든 관계자를 참여시켜 더욱 안전한 온라인 경험을 위한 언론 매체 및 디지털 리터러시 지침과 프로그램을 개발하고, 청소년의 디지털 활용 능력을 적극적으로 활용합니다.

모든 정치 주체자를 위한 권고안

정치 과정에 참여하고 영향을 미치는 개인, 단체 및 기관

a. 선거의 공정성을 유지합니다. 유권자 자격, 투표 과정, 개표 및 결과 발표

를 포함한 정보의 무결성을 훼손하려는 모든 시도를 자제하고 공개적으로 규탄해야 합니다.

b. 포용성을 보호합니다.후보자와 공직자, 특히 여성과 취약하고 소외된 집단 구성원을 대상으로 하는 학대와 괴롭힘을 공개적으로 규탄하고 이를 해결하기 위한 조치를 취해야 합니다.

c. 투명성을 제공합니다.광고 자금 출처 및 데이터 기반 타겟팅 기술 사용을 포함하여 모든 커뮤니케이션에서 투명성을 유지해야 합니다.

p.38

유엔

정보 무결성 글로벌 원칙은 유엔과 그 국제 공무원에게 적용됩니다. 이 글로벌 원칙을 준수함으로써, 유엔은 국제 사회에서 정보 무결성을 책임감 있게 관리하는 모범적인 사례를 제시합니다. 정보 생태계의 건전성을 강화하기 위한 노력을 확대하는 것은 평화 확보, 지속 가능한 개발 장려, 그리고 모든 사람을 위한 인권 증진 및 보호라는 유엔의 사명을 달성하는 데 기여할 것입니다.

유엔은 다음과 같이 할 것입니다

a. 노력을 확대합니다. 상황별 연구, 모니터링, 위험 평가, 지역 사회 참여 및 다양한 환경과 언어에 걸친 연합 구축을 통해 정보 무결성을 강화하기 위한 노력을 강화해야 합니다. 예방, 완화 및 대응을 강화하고 새로운 기회와 과제를 파악하기 위해 정보 무결성을 프로그램 및 운영에 통합해야 합니다.

b. 역량 강화 노력을 지원합니다. 정보 무결성을 강화하기 위해 청년층을 위한 교육을 포함한 기술 개발 프로그램을 제공하는 등 회원국의 역량 강화에 도움을 제공하며, 특히 개발도상국의 요구 사항에 특별한 관심을 기울입니다.

c. 옹호 활동을 수행합니다. 전 세계적으로, 그리고 여러 국가와 지역 사회에 걸쳐 글로벌 원칙을 홍보하고 지지하며, 특히 소외된 환경과 취약하고 소외된 상황에 처한 집단에 특별한 관심을 기울입니다. 사회적 결속에 적극적으로 기여하고 정보 무결성에 대한 위험에 대한 지역 사회의 회복력을 강화하며, 지속 가능한 개발 목표 달성을 위한 노력을 지원합니다.

d. 전담 역량을 강화합니다. 유엔 사무국 내에 중앙 부서를 설립하여 유엔의 임무 수행 및 실질적인 우선순위에 영향을 미치는 정보 생태계의 무결성에 대한 위험에 대처하기 위한 혁신적이고 세밀한 접근 방식을 개발하고, 다른 관련 부서들과 협력하며 필요에 따라 유엔 시스템 전체에 서비스를 제공해

야 합니다.

e. 민첩한 커뮤니케이션 전략을 개발합니다. 혁신적이고 증거 기반이며 유연하고 맞춤화된 소통 전략을 활용하여 디지털 및 오프라인 정보 공간을 통해 공동 이익을 증진하고 유엔이 봉사하는 모든 사람들의 요구를 더 잘 충족시켜야 합니다. 공동의 이익을 위해 그리고 유엔이 봉사하는 모든 사람들의 필요를 더 잘 충족시키기 위해 디지털 및 오프라인 정보 공간을 활용하는 혁신적이고, 증거 기반이며, 민첩하고, 맞춤화된 커뮤니케이션 전략을 활용해야 합니다.

f. 다국어 자료를 제공합니다. 다양한 환경에 적용 가능한 연구 자료, 지침, 및 모범 사례를 공유하는 다국어 온라인 정보 무결성 허브를 구축하여 전 세계, 지역 및 국가 차원의 정책 계획을 지원해야 합니다.

g. 다자간 이해관계자 행동 계획을 지원합니다. 기존 메커니즘을 활용하고 유엔 기구의 국제 역량 강화 및 조정 분야 전문성과 경험을 바탕으로 지역 및 국가 차원의 다자간 이해관계자 행동 계획과 연합을 지원해야 합니다.

p.40

다음 단계

정보 생태계의 무결성에 대한 위험이 고조되고AI 기술의 발전이 빠르게 진행되는 상황에서 정보 무결성을 강화하는 것의 시급성은 아무리 강조해도 지나치지 않습니다. 글로벌 원칙은 세계가 디지털 시대의 복잡성을 헤쳐나가고 미래 정상회의에서 다자간 해결책을 모색하는 가운데 정보 무결성을 보호하고 증진하기 위한 포괄적이고 통합적인 행동 지침(framework)을 제공합니다.
이를 위해 이해관계자들은 다음 사항을 준수할 것을 권장합니다.

- 유엔 정보 무결성 글로벌 원칙을 즉각적인 조치를 위한 지침으로 삼아 공개적으로 약속하고 채택하며 적극적으로 홍보해야 합니다. 즉각적인 행동을 위한 프레임워크로서 유엔 정보 무결성 글로벌 원칙에 공개적으로 헌신하고, 채택하며, 적극적으로 홍보해야 합니다.

- 정보 무결성에 관한 글로벌 원칙을 활용하여 다양한 분야의 전문가와 접근 방식을 포함하는 광범위한 부문 간 연합을 구성하고 적극적으로 참여하며,시민 사회, 학계, 언론, 정부 및 국제 민간 부문으로부터 역량 강화에 필요한 지원을 확보하고, 전담 청년 자문 그룹 등을 통해 청년들의 완전하고 의미 있는 참여를 보장해야 합니다.

- 지역, 국가 및 지방 차원에서 다자이해관계자 참여하는 실행 계획을 수립하기 위해 협력하고, 지역 사회를 참여시켜 풀뿌리 정책(grass-roots initiatives)을 지원하고 배우며, 청소년의 완전하고 의미 있는 참여를 보장해야 합니다.

유엔 정보 무결성 글로벌 원칙을 수용함으로써 모든 분야의 이해관계자들은 연대를 보여주고 협력하여 모두에게 신뢰, 지식 및 개인의 선택권을 보장하는 활력 넘치는 정보 생태계를 향한 길을 함께 만들어 나갈 수 있습니다.

p.41

부록

참고 문헌

1. UnitedNations Secretary-General's "Our Common Agenda policy brief 8: informationintegrity on digital platforms" (2023)
https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf
2. UNESCOGuidelines for the Governance of Digital Platforms (2023)
https://www.unesco.org/en/internet-trust/guidelines
3. Report ofthe Secretary-General, "Countering disinformation for the promotion andprotection of human rights and fundamental freedoms", 2022 (A/77/287)
https://www.ohchr.org/sites/default/files/2022-03/NV-disinformation.pdf
4. UNESCORecommendation on the Ethics of Artificial Intelligence (2021)
https://unesdoc.unesco.org/ark:/48223/pf0000381137
5. UnitedNations Strategy and Plan of Action on Hate Speech (2019)
www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf
6. RabatPlan of Action on the prohibition of advocacy of national, racial or religioushatred that constitutes incitement to discrimination, hostility or violence(2012)
https://www.ohchr.org/en/documents/outcome-documents/rabat-plan-action
7. GuidingPrinciples on Business and Human Rights (2011)
https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

저 자 소 개

권 오 상

- 연세대학교 경영학 학사
- 연세대학교 경영학 석사
- American University Washington College of Law LL.M
- 연세대학교 법학 박사
- 현 디지털미래연구소 대표

이 찬 구

- 한국외국어대학교 사학과 학사
- 한국외국어대학교 신문방송학 석사
- 한양대학교 경영학 박사
- 현 디지털미래연구소 연구위원

김 국 진

- 고려대학교 신문방송학 학사
- 고려대학교 신문방송학 석사
- 고려대학교 신문방송학 박사
- 현 디지털미래연구소 고문

김 슬 기

- 성균관대학교 법학 학사
- 성균관대학교 법학 석사
- 성균관대학교 법학 박사
- 현 디지털미래연구소 연구위원

노 윤 경

- 숙명여자대학교 중어중문학 학사
- 연세대학교 법학 석사
- 연세대학교 법학 박사수료
- 현 디지털미래연구소 책임연구원

방송통신융합 정책연구 KMCC-2025-18

‘정보 무결성’ 강화를 위한 중장기 정책 방향에 관한 연구

2025년 12월 31일 인쇄
2025년 12월 31일 발행

발행인 방송미디어통신위원회 위원장
발행처 방송미디어통신위원회
경기도 과천시 관문로 47
정부과천청사 2동
TEL: 02-2110-1323
Homepage: www.kmcc.go.kr