

방통융합정책연구 KCC-2023-19

온라인 가짜뉴스 근절을 위한 사업자 자율규제 방안 연구

(A Study on Self-Regulatory Guideline for Online Platforms
on Tackling Disinformation)

이수엽/김국진/권오상/최민정/박지은/최예서

2023. 12.

연구기관 : (사)미디어미래연구소



방송통신위원회

Korea Communications Commission

이 보고서는 2023년도 방송통신위원회 방송통신발전기금 방송통신
융합 정책연구사업의 연구결과로서 보고서 내용은 연구자의 견해
이며, 방송통신위원회의 공식입장과 다를 수 있습니다.

제 출 문

방송통신위원회 위원장 귀하

본 보고서를 『온라인 가짜뉴스 근절을 위한 사업자 자율규제 방안 연구』의 연구결과보고서로 제출합니다.

2023년 12월

연구기관 : (사)미디어미래연구소

총괄책임자 : 이수엽

참여연구원 : 김국진

권오상

최민정

박지은

최예서

목 차

제1장. 서론	1
제1절. 연구 배경 및 필요성	1
제2절. 연구의 구성	3
제3절. 용어의 정의	4
제2장. 국내외 허위조작정보 유통 현황 및 영향력	7
제1절. 해외 허위조작정보 유통 현황 및 영향력	7
제2절. 국내 허위조작정보 유통 현황 및 영향력	15
제3절. 소결	25
제3장. 해외 허위조작정보 관련 법적규제 사례	26
제1절. 미국	26
제2절. EU	38
제3절. 프랑스	52
제4절. 독일	56
제5절. 영국	63
제6절. 일본	68
제7절. 소결	71
제4장. 해외 허위조작정보 관련 자율규제 실천강령 사례	74
제1절. EU의 허위조작정보 대응 실천강령(2022)	74
제2절. 호주의 허위조작정보 및 오정보 대응 실천강령	117
제3절. 소결	128

제5장. 국내 허위조작정보 관련 자율규제 활성화 방안	131
제1절. 국내 허위조작정보 관련 법적 규제 현황	131
제2절. 국내 허위조작정보 관련 사업자 자율규제 사례	137
제3절. 허위조작정보 자율규제 시범 가이드라인(안) 제언	151
참고문헌	166
[부록 1] EU 허위조작정보에 대한 강화된 실천강령 2022	174
[부록 2] 호주 허위조작정보 및 오정보 대응 실천강령 2022	243
[부록 3] 호주 실천강령 부록 : 사전 동의 지명양식	267

표 목 차

<표 3-1> 통신품위법 230조(c)의 내용	28
<표 3-2> 「안전하게 보호되고 신뢰할 수 있는 AI 개발이용에 관한 행정명령」의 구성	34
<표 3-3> AI 합성 콘텐츠 관련 내용	35
<표 3-4> 디지털 서비스 법의 구조	40
<표 3-5> 디지털 서비스법 적용대상 사업자 분류	42
<표 3-6> 디지털 서비스법에 따른 대규모 온라인 플랫폼 기업 목록	43
<표 3-7> 디지털 서비스법 적용대상 사업자 유형별 의무	43
<표 3-8> 디지털 서비스법이 온라인 플랫폼에 부여하는 의무	45
<표 3-9> 대규모 온라인 플랫폼/검색엔진 제공자가 관리해야 할 구조적 위험의 유형	47
<표 3-10> 대규모 온라인 플랫폼/검색엔진 제공자의 추가적인 의무	48
<표 3-11> 프랑스 「정보보작대처법」 제3장의 주요 내용	54
<표 3-12> 소셜 네트워크 법집행법의 구성	59
<표 3-13> 소셜 네트워크집 법집행법 상의 불법정보 유형	59
<표 3-14> 소셜 네트워크 법집행법 제2조제2항의 내용	62
<표 3-15> 유튜브의 삭제 콘텐츠 내용과 삭제 건수	63
<표 3-16> 온라인 서비스 제공자의 의무	66
<표 3-17> 카테고리1 서비스의 추가 의무	67
<표 3-18> 총무성의 허위조작정보 대책 방안(2019년)	70
<표 4-1> 유럽집행위원회의 온라인 허위조작정보 관련 자율규제 정책 추진현황	74
<표 4-2> 「온라인 허위조작정보 대응에 관한 정책안」의 4가지 원칙	75
<표 4-3> 허위조작정보 대응 실천강령(2018년)의 15개 준칙	76
<표 4-4> ‘2018’ 강령에 대한 보완요청 사항	79

<표 4-5> '2022 강령' 의 주요 내용	81
<표 4-6> '2022 강령' 허위조작정보 영역 및 정의	85
<표 4-7> '2018 강령' 과 '2022 강령' 의 주제영역 비교	86
<표 4-8> '2018 강령' 과 '2022 강령' 의 준칙1 비교	87
<표 4-9> '2022 강령' 서명인 현황	88
<표 4-10> 강화된 허위조작정보 대응 실천 강령 이행 보고서 구조	92
<표 4-11> 구글의 게재 및 수익 창출 불가 광고 유형과 정의	95
<표 4-12> '2022 강령' 에 대한 참여조직 간 입장 차	113
<표 4-13> 플랫폼별 허위조작정보 대응 관련 영향력 측정 보고내용	116
<표 4-14> 호주 실천강령 관련 연혁	118
<표 4-15> 호주 실천강령의 허위조작정보 및 오정보 정의	121
<표 4-16> 호주의 허위조작정보와 오정보 실천강령의 주요 목표 및 결과	122
<표 4-17> EU의 자율규제 사례 도입을 위한 국내 제안 사항	129
<표 5-1> 허위조작정보 관련 법적 규제 현황	131
<표 5-2> 허위사실 표현이 초래하는 피해와 법적 대응	132
<표 5-3> 정보통신망법에 따른 플랫폼 사업자 책임	133
<표 5-4> 제21대 국회에 제출된 허위조작정보 관련 주요 법률안 : 정보통신망법 개정안	134
<표 5-5> 정보통신망법 상 자율규제 관련 조항	136
<표 5-6> KISO <정책규정> 제5장 제5절	138
<표 5-7> 코로나19관련 허위조작정보에 관한 정책	140
<표 5-8> 검색추천 서비스의 투명성 제고를 위한 자율규제 원칙	142
<표 5-9> 네이버의 허위정보 관련 콘텐츠 운영정책	144
<표 5-10> 코로나19 허위조작정보에 해당하는 경우	144
<표 5-11> 유해게시물 신고 사유	146
<표 5-12> 카카오 운영정책 중 가짜뉴스 관련 내용	148

<표 5-13> 코로나19 관련 허위조작정보에 관한 정책	149
<표 5-14> 플랫폼의 책임에 관한 주요 법제 비교	152
<표 5-15> 허위조작정보 관련 민간 자율규제 비교	153
<표 5-16> 허위조작정보 정의 비교	155
<표 5-17> 허위조작정보의 정의(안)	156
<표 5-18> 가이드라인의 이행 원칙(안)	157
<표 5-19> 허위조작정보의 신고 및 조치 주요 내용(안)	159
<표 5-20> 허위조작정보의 수익 차단 주요 내용(안)	160
<표 5-21> 서비스의 완전성 주요 내용(안)	161
<표 5-22> 검색추천 서비스의 투명성	162
<표 5-23> 이용자 역량 강화	163
<표 5-24> 팩트체크 및 연구 지원	163
<표 5-25> 투명성 보고서 발간 및 모니터링	164

그 립 목 차

[그림 2-1] 전 세계가 직면한 위협 순위 (19개국, 단위: %)	7
[그림 2-2] 세대별 허위정보를 접하는 빈도(7개국, 단위: %)	8
[그림 2-3] 세대별 정보의 출처에 따른 검증 시도 비율(7개국, 단위: %)	9
[그림 2-4] 허위조작정보가 선거에 미치는 영향에 대한 우려 정도(16개국, 단위: %)	11
[그림 2-5] 선거 운동 기간 동안 정부와 규제당국이 소셜미디어 플랫폼에 신뢰 및 안전조 치를 요구할 수 있다고 인식하는 비율(16개국, 단위: %)	12
[그림 2-6] 코로나19 관련 허위정보 경로로 우려되는 온라인 미디어 플랫폼 (46개국, 단위: %)	13
[그림 2-7] 코로나19 관련 인포데믹과 팩트체크에 대한 노출 조사 (단위: %)	13
[그림 2-8] 매체별 언론 인식 여부 (단위: %)	15
[그림 2-9] 디지털 뉴스 주이용 경로(단위: %)	16
[그림 2-10] 연령대, 정치 성향별 유튜브 뉴스 이용률 (2022년, 단위: %)	17
[그림 2-11] 인터넷 허위정보에 대한 우려 (2022년, 단위: %)	17
[그림 2-12] 소셜미디어에서의 허위정보 접촉경험 (단위: %)	18
[그림 2-13] 매체별 허위조작정보 확산에 대한 평가	19
[그림 2-14] 매체별 허위조작정보 확산정도	19
[그림 2-15] 선거와 관련된 정보를 자주 접하는 매체	20
[그림 2-16] 허위조작정보 및 가짜뉴스 관련 주체별 “책임 매우 크다” 응답 비율 (단위: %) ..	21
[그림 2-17] 허위조작정보 출처 (단위: %)	22
[그림 2-18] 허위조작정보 유통경로 (단위: %)	23
[그림 2-19] 코로나19 관련 정보에 대한 참/거짓 응답률 (6개국, 단위: %)	24
[그림 3-1] FTC 합의에 담긴 페이스북의 사용자 개인정보 보호 방법	33
[그림 3-2] 디지털 서비스법 적용대상 사업자 분류	42
[그림 4-1] 실천 강령 이행 보고서 서식(예)	93

[그림 4-2] 구글 정치 광고 API 플랫폼	96
[그림 4-3] 구글의 웹페이지 자세히 알아보기 기능	99
[그림 4-4] 구글의 ‘슈퍼 검색자’ 캠페인	100
[그림 4-5] 구글 유튜브의 ‘잠시만요’ 캠페인	101
[그림 4-6] 메타 광고 라이브러리 API 플랫폼	105
[그림 4-7] 메타 서비스 완전성을 위한 연간 조치 타임라인	108
[그림 4-8] 메타(페이스북) 게시물 표시 이유 공개	109
[그림 4-9] 메타(페이스북) 팩트체크 라벨 및 경고화면	111
[그림 4-10] 메타(Meta)의 ‘2022 강령’ 서명 자료 중 일부	114
[그림 5-1] KISO의 주요 기능	137
[그림 5-2] 가짜뉴스 신고 처리절차 안내	139
[그림 5-3] 플랫폼 민간 자율기구 구성도	141
[그림 5-4] 플랫폼의 사회적 가치 제고를 위한 8대 원칙	143
[그림 5-5] KISO 가짜뉴스 신고 처리 절차	147
[그림 5-6] 다음 고객센터 내 ‘코로나19 관련 허위조작정보에 관한 정책’	148
[그림 5-7] 카카오 운영정책 내 유행 정보 신고 및 접수 처리	150
[그림 5-8] 카카오 운영정책 내 KISO에 접수된 이용자 신고 처리 절차	151

요 약 문

1. 제 목

온라인 가짜뉴스 근절을 위한 사업자 자율규제 방안 연구

2. 연구 목적 및 필요성

- **(연구목적)** 허위조작정보와 관련된 국내외 주요 법적 규제 현황 및 자율규제 현황을 비교·분석하고 시사점을 도출하여 인터넷 사업자의 자율규제 활성화를 위한 방안을 모색
 - 허위조작정보와 관련된 인터넷 사업자의 자율규제 시범 가이드라인(안) 제안을 통해 자율규제 체계의 확립 및 관련 거버넌스 개선에 시사점 제공

- **(연구필요성)** 2010년대 중반 이후 국내외에서의 허위조작정보에 대한 논의 확산
 - 주요국의 선거 및 여론형성과정에서 허위조작정보가 유포돼 사회적 혼란이 발생하고 시민의 민주적 선택이 방해받는 일이 빈번하게 발생하면서 대응 논의가 확산
 - 코로나19 팬데믹과 더불어 부정확한 정보가 온라인을 통해 급속도로 퍼지는 인포데믹(infordemic) 현상이 확산되면서 세계 각국은 허위조작정보가 정치·사회적 혼란을 야기하는 것을 넘어 인간의 생명에 직접적인 위협을 가할 수 있음을 인식하고 강화된 대응 필요성을 인식함

- 전 세계적으로 허위조작정보 확산에 대응하기 위한 법적규제 방안과 자율규제 방안이 활발하게 모색
 - 미디어 환경의 변화에 따라 허위조작정보의 파급력, 피해 수준 및 규모 확대의 가능성이 증가했으며, 특히 소셜봇(Social Bots), 딥페이크(Deepfake), 생성형 AI 등 기술 발달이 문제의 심각성을 더하고 있는 상황
 - 현행 법체계가 온라인 환경 변화를 잘 빠르게 쫓아가기 힘들고, 대량으로 유통되는

정보를 공적 기관이 모니터링하고 규제하는 것이 불가능해짐에 따라 허위조작정보 대응방안에서 인터넷 사업자의 책임을 강화하는 자율규제가 주목받고 있음

3. 연구의 구성 및 범위

(1) 국내의 허위조작정보 유통현황 및 영향력 분석

- 국내의 허위조작정보의 유통 현황 및 시민들에게 미치는 영향력을 파악함으로써 문제점을 이해하고 현재의 미디어 환경에 부합하는 대응방안을 마련하고자 함

(2) 해외 허위조작정보 관련 법적 규제 사례 분석

- 미국, EU, 프랑스, 독일, 영국, 일본의 허위조작정보 관련 법적 규제 현황을 비교 검토함으로써 사업자의 책임을 강화하는 주요 동향을 살펴보고, 법적 규제의 범위와 효과 및 한계를 이해하고자 함

(3) 해외 허위조작정보 관련 자율규제 실천강령 사례 분석

- EU와 호주의 허위조작정보 대응 실천강령의 주요 내용, 제정 배경 및 특징을 비교 검토함으로써, 시사점을 모색하고 국내 적용 가능성을 검토함

(4) 국내 허위조작정보 자율규제 활성화 방안 마련

- 앞서 살펴본 각국의 법적 규제 현황, 자율규제 실천강령을 국내 법적규제 및 자율규제 동향과 비교 분석함으로써, 사업자들이 실천할 수 있는 자율규제 시범 가이드라인(안)을 제언하고, 자율규제 규제 체계 마련에 시사점을 제공하고자 함

4. 연구 내용 및 결과

1) 국내외 허위조작정보 유통 현황 및 영향력

o (배경) 허위조작정보가 디지털 미디어 환경에서 빠르고 광범위하게 확산되면서 허위조작정보 유통은 전 세계적으로 심각한 문제가 되었음

- 인터넷 포털과 검색엔진, 소셜미디어는 정보의 매개자로서 주요한 역할을 수행하는 만큼 허위조작정보의 주요한 통로가 되고 있어, 주요 플랫폼이 허위조작정보의 확산을 막기 위해 일정한 역할을 해야 한다는 인식이 높아짐

구분	국외	국내
온라인 미디어에서의 허위 조작 정보	<ul style="list-style-type: none"> · 소셜미디어를 통해 유통되는 정보에 대해서는 정보에 대한 검증 수준이 낮음 · 허위조작정보는 소셜미디어 상에서 활발하게 유통됨 · 소셜미디어에서 허위조작정보는 진실된 정보보다 더 많이, 더 빠르게 공유되는 경향을 보임 	<ul style="list-style-type: none"> · 온라인 포털을 중심으로 뉴스와 정보를 이용 · 특히 소셜미디어의 경우, 유튜브를 뉴스 매체로 이용하는 비율이 높은 편으로 나타남 · 소셜미디어 이용자 과반수가 허위조작정보를 접하고 있으며, 소셜미디어 중 온라인 동영상 플랫폼이 허위정보의 주요 확산 경로라고 인식
선거와 관련한 허위 조작 정보	<ul style="list-style-type: none"> · 허위조작정보 및 가짜뉴스가 선거 캠페인 및 투표에 미칠 영향에 관한 국민적 우려 <ul style="list-style-type: none"> - 2016년 미국 대선 · 정부 및 규제당국이 선거 기간 동안 소셜미디어 플랫폼 신뢰 및 안전 조치에 대해 요구 필요성 인식 	<ul style="list-style-type: none"> · 선거와 관련된 정보를 접하고 이용하는 과정에서 포털 및 소셜미디어가 중요한 정보 유통 창구로 나타남 · 허위조작정보 및 가짜뉴스의 유통과 관련하여 ‘유튜버 또는 1인 미디어’가 책임이 가장 큰 것으로 인식
코로나19와	<ul style="list-style-type: none"> · 코로나19 팬데믹 시기, 코로나19와 관련된 허위조작정보에 대한 	<ul style="list-style-type: none"> · 코로나19와 관련된 허위조작정보의 주요 유통 창구는 인터넷 및

관련한 허위 조작 정보	유통 확산 · 특히 소셜미디어(페이스북)가 허위 조작정보의 주요 유통 경로로 우 려되는 경향을 보임 · 코로나19 관련 허위조작정보는 감 염병 대응에 심각한 위협이 되었 으며, 사회적 혼란을 야기	소셜미디어 등 온라인 플랫폼으로 인식
-----------------------	---	-------------------------

2) 해외 허위조작정보 관련 법적규제 사례

(1) 미국

- **(배경)** 플랫폼과 관련된 직간접적인 규제 외에 허위조작정보의 대응과 관련된 다양한 정책적, 입법적 시도 존재
 - 다만, 미국은 허위조작정보라 할지라도 표현의 자유 보호대상이 되므로 법령을 통해 콘텐츠의 내용을 규제하는 것에 부정적
 - 따라서 통신품위법을 통해 일정한 조치를 취한 인터넷서비스사업자의 민사 상 책임을 면제시키는 등 표현의 자유를 중시
- **(입법적 규제)** 통신품위법 개정 논의
 - 2018년 통신품위법 제230조 일부 개정을 통해, 성매매 목적의 인신매매에 한하여 온라인 서비스 제공자의 법적 책임을 규정
 - 악성 알고리즘 방지 법안(Justice Against Malicious Algorithms Act)을 발의하여 플랫폼이 개인맞춤형 추천을 한 경우, 플랫폼이 이를 인식하고 있으며, 맞춤형 알고리즘을 통해 피해자의 신체적, 정신적 상해에 영향을 준 경우 면책이 적용되지 않도록 규정
 - 필터버블 투명성 법안(Filter Bubble Transparency Act)을 발의하여, 불투명 알고리즘(opaque algorithm)에 대해 이용자에게 알리며, 이용자가 명시적으로 제공한 데이터만 사용하는 대안의 마련을 요구

- 플랫폼 책임 및 소비자 투명성 법안(PACT Act)의 발의를 통해 정부가 플랫폼의 콘텐츠 모더레이션 정책에 관여할 중대한 공익이 있음을 명시함

○ **(플랫폼사업자와의 합의를 통한 규제)** FTC와 페이스북의 동의의결

- 2011년 페이스북의 이용자 개인정보 취급이 FTC 법을 위반하는지에 대한 조사를 개시하고 동의의결로 종결하였으나, 2016년 대선과정에서 이를 위반하였다는 논란이 발생하여 2018년 조사 개시
- 이에 따라, FTC와 페이스북은 이용자의 개인정보와 플랫폼에 제공되는 정보를 보호할 포괄적인 틀을 마련하기로 합의하여, 페이스북의 투명성과 책임을 강화하고자 함

○ **(행정명령)** 바이든 행정부의 2023년 AI 규제 행정명령

- 국가 안보, 건강, 안전을 위협하는 AI 기술의 개발과 이용을 규제하기 위한 내용을 포함
- 제1조 목적을 통해 허위조작정보 대응이 주요한 목적임을 명시함
- AI 합성 콘텐츠와 관련한 규정을 마련하여, AI를 이용한 공격이나 허위조작정보의 확산을 예방하기 위한 조치를 취한 것으로 평가되고 있으나, 행정명령이라는 한계점이 존재함

○ **(주 선거법)** 각 주의 선거기간 딥페이크 규제

- 텍사스, 캘리포니아 등 일부 주법에서는 선거운동에서의 딥페이크를 명시적으로 규제하고 있음

○ **(뉴저지 주)** 정보 리터러시 교육 의무화

- 표현의 자유에 대한 규제가 어려운 미국의 경우, 범죄에 해당하는 불법 콘텐츠가 아닌 한 온라인 허위조작정보의 유포 자체를 연방법으로 규제하고 있지 않음
- 따라서 이용자들이 정보를 효과적으로 활용하고 비판적으로 분석할 수 있도록 도와주는 리터러시 교육 입법에 적극적이며, 뉴저지 주는 2023년 입법을 통해 정보 리터러시 교육을 공립학교에서 의무화함

(2) EU

- **(배경)** 온라인 허위조작정보의 확산이 심각한 사회문제로 부상하면서 EU는 허위조작정보에 대처하기 위한 다양한 방안을 모색
 - 특히, 러시아 등 외국으로부터의 허위조작정보 캠페인에 대처하기 위한 방안과 주요 온라인 플랫폼의 자율규제를 강화하기 위한 방안 존재
- **(입법적 규제)** 디지털 서비스법
 - 디지털 서비스법(DSA)은 온라인에서 이용자의 기본권에 대한 포괄적 보호뿐만 아니라 불법콘텐츠의 신속한 제거를 위한 절차를 규정함으로써 이용자를 재화, 서비스 또는 콘텐츠에 연결하는 모든 디지털서비스에 대한 EU 회원국 전체의 의무를 구속력있게 규율하고자 함

구분	내용
구조	제1장: 규칙의 대상범위, 주요 용어의 정의 등 일반조항 제2장: 중개서비스 제공자의 책임 면제에 관한 조항 제3장: 투명하고 안전한 온라인 환경을 위한 상당한 주의 의무 제4장: 규칙의 시행, 협력, 처벌 및 집행 제5장: 규칙의 평가발효·적용 등
적용대상	<ul style="list-style-type: none"> · 중개서비스: 네트워크 인프라를 제공하는 서비스 · 호스팅서비스: 클라우드, 웹호스팅 서비스 등 · 온라인 플랫폼: 판매자와 소비자를 연결하는 서비스 · 대규모 온라인 플랫폼: 매월 EU 인구의 10% 이상이 이용하는 온라인 플랫폼 및 검색엔진(유럽집행위원회에서 제정)
DSA와 허위조작정보	허위조작정보에 대한 법적 정의를 포함하지 않으나, 불법 콘텐츠에 대응하고 투명성을 확보하도록 요구함으로써, 허위조작정보의 유포로부터 이용자를 보호하는 데 기여할 수 있을 것으로 판단 <ul style="list-style-type: none"> · 불법 콘텐츠에 대한 대응 · 온라인 투명성 · 위험완화 및 위기대응(대규모 온라인 플랫폼검색엔진의 추가의무)

대규모 온라인 플랫폼의 추가적인 의무	<ul style="list-style-type: none"> - 위협평가 - 위협완화 - 위기 대응 매커니즘 - 독립적인 감사 - 추가적인 온라인 광고 투명성 확보의무 - 데이터에 대한 접근과 조사 - 준법감시 기능 - 투명성 보고의무
DSA와 허위조작정보 대응 실천강령	<ul style="list-style-type: none"> · 디지털 서비스법은 유럽집행위원회의 허위조작정보 대응 실천강령과 함께 온라인 허위조작정보에 관한 공동규제 체계를 형성

(3) 프랑스

- (배경) 프랑스는 2018년 선거과정에서 제기되는 허위조작정보에 대한 대책의 일환으로 「정보조작 대처에 관한 2018년 12월 22일 법 n° 2018-1202」를 통과시켜, 선거 기간 중에 확산되는 허위조작정보를 막는데 초점

구분	내용
허위조작정보 정의	· 유권자 투표의 진정성을 해치는 부정확하고 오해를 불러올 수 있는 주장이나 비난 등이 온라인 정보통신서비스를 통해서 의도적이고, 인위적이며 또는 자동적으로, 대량으로 전파되는 정보
적용대상	· 페이스북, 트위터, 구글과 같이 프랑스 내에서 접속자가 월 500만 명 이상인 모든 플랫폼 사업자와 공적인 이슈에 관한 사회적 토론 내용을 전달하는 사이트 가운데 광고료 수입이 건당 평균 100유로 이상인 모든 사이트
주요내용	· 허위조작정보의 유포에 대응하기 위해 플랫폼 사업자에게 협조 의무를 부과(제3장)
평가	<ul style="list-style-type: none"> · 허위조작정보 정의의 모호성 · 판사에게 정보의 진실성 입증을 요구하는 절차에 대한 문제 제기 · 사업자들이 정보조작대처법의 적용을 받지 않기 위해 관련된 정보의 유통을 회피하도록 만드는 결과

(4) 독일

- **(배경)** 소셜네트워크 서비스를 통해 유포되는 혐오표현, 허위조작정보의 생산·유통에 대응하기 위해 소셜네트워크 법집행법(Netzwerkdurchsetzungsgesetz, NetzDG)을 제정
 - 동법은 허위조작정보에 대한 대응의 측면도 갖고 있지만 그보다는 혐오발언에 대한 대응에 좀 더 초점이 맞춰진 것으로 평가됨

구분	내용
적용대상	· 수익창출을 목적으로 이용자 간 임의의 콘텐츠를 다른 이용자와 교환 및 공유하거나 공중이 이용하도록 이용자가 콘텐츠를 공개적으로 게시할 수 있는 인터넷 플랫폼을 제공하는 통신미디어서비스 제공자(즉 소셜 네트워크 사업자)
주요내용	제1조: 적용범위 제2조: 공지의무(조치현황 보고 및 공개 의무) 제3조: 불법게시물에 관한 이의제기 처리방법 제4조: 벌칙 규정 제5조: 국내 송달 대리인 임명 의무 규정 제6조: 경과 규정
불법정보 유형	· 불법 게시물은 독일 형법상의 불법정보 유형 21가지를 의미하며, 불법정보는 주로 혐오표현 및 인격권 침해 표현에 초점
평가	· 소셜네트워크 법집행법을 통한 규제 강화는 자율규제를 활성화시키는 동시에 입법을 통해 자율규제에 대한 규제를 강화 · 허위조작정보의 대응을 위한 법이라기보다는 혐오표현의 대응에 관한 법률

(5) 영국

- **(배경)** 2016년 브렉시트 국민투표를 전후해 페이스북 이용자들의 정보가 무단으로 선거운동에 이용되고, 온라인 플랫폼을 통해 허위조작정보가 확산되면서 관련 논의가 본격화

○ **(온라인 유해 백서)** 2019년 4월 온라인 유해 콘텐츠의 규제, 방지 및 기업의 대응 방안을 포함한 온라인 유해 백서 발표

구분	내용
플랫폼 사업자의 책임 강화를 위한 검토 이슈	<ul style="list-style-type: none"> · 플랫폼 사업자의 허위조작정보 인지 가능성 · 플랫폼 사업자의 콘텐츠 유해성 판단 여부 및 사업자의 책임 · 온라인 플랫폼 사업자의 책임성 강화방안: 디지털 서비스 세(digital service tax) 논의 · 플랫폼 사업자와 정부의 동의 의결(consent decree) 방안 · 해외 기업의 데이터 활용에 대한 영국 법제 적용 가능성

○ **(온라인 안전법)** 온라인 이용자의 안전 보장을 위한 온라인 안전법 마련

구분	내용
적용대상	<ul style="list-style-type: none"> · 인터넷 서비스 제공자, 검색엔진 서비스 제공자 · 상당수의 영국 이용자를 호스팅하거나, 서비스의 주요 타겟 시장이 영국이거나, 서비스가 영국 내에서 이용 가능하며 서비스의 이용자 생성 콘텐츠가 영국 내 개인에게 중대한 피해를 줄 수 있다고 믿을만한 합당한 근거가 있는 경우
주요내용	<ul style="list-style-type: none"> · 사업자 규모에 따라 차등적인 의무를 부과함 <ul style="list-style-type: none"> - 불법 콘텐츠 위험 평가 의무. 서비스는 적절하고 충분한 불법콘텐츠 위험평가를 수행해야 하며 이를 최신 상태로 유지해야 함 - 불법콘텐츠와 관련된 위험을 완화하고 관리하기 위한 조치를 취할 의무. 또 서비스 약관에 이용자를 보호하는 방법을 명시하는 조항을 포함해야 하며 이러한 조항은 사용자가 명확하고 쉽게 확인할 수 있어야 함 - 투명성, 신고 및 구제에 관한 의무. 서비스는 불법콘텐츠를 쉽게 신고할 수 있는 이해하기 쉬운 방법을 마련해야 하며, 이용자가 접근 가능한 불만처리 절차를 운영해야 함. 또한 불법콘텐츠에 대한 신고 기록과 위험평가 기록을 보관해야 함 - 표현의 자유 및 개인정보 보호에 관한 의무. 서비스는 안전조치 및 정책을 결정하고 시행할 때, 이용자의 표현의 자유와 개인정보 보호를 특별히 고려해야 함

규제 대상 콘텐츠	<ul style="list-style-type: none"> · 불법 콘텐츠는 테러행위나 아동 성착취·아동 성학대 등의 내용을 담고 있는 콘텐츠를 의미 · 유해 콘텐츠는 적법하지만 유해한 콘텐츠를 가리키는데 법에서는 그 기준을 명확히 정의하고 있지 않으나, 대신 이를 아동 유해 콘텐츠와 성인 유해 콘텐츠로 구분
평가	<ul style="list-style-type: none"> · 플랫폼 기업이 더 투명하고 능동적으로 대처하도록 요구 · 허위조작정보의 직접적인 규제에 관한 내용을 담고 있지 않으며, 단지 Ofcom을 통해 허위조작정보 및 오정보에 관한 자문위원회를 설치하도록 규정 · 아동이 온라인에서 보는 콘텐츠에 대한 주의 의무를 부과하며, 성인에 대한 권한부여도 규정하고 있어 관련 기업은 사용자 연령 확인부터 시작해 여러 요구사항을 준수해야 함

(6) 일본

- **(배경)** 허위조작정보에 대한 법적 규제를 검토한 바 있으나, 헌법이 보장하는 표현의 자유를 과도하게 억제할 우려가 있다는 점으로 법률 제정은 보류됨
- **(프로바이더 책임제한법)** 허위조작정보를 직접 규제하는 법률은 아니나, 인터넷상에서의 내용규제와 관련된 법률

구분	내용
규제 대상 콘텐츠	<ul style="list-style-type: none"> · 삭제요청 혹은 발신자 정보공개청구의 대상으로 하는 정보는 특정 개인의 권리를 침해하는 정보로 한정
주요내용	<ul style="list-style-type: none"> · 자신의 권리를 침해당한 개인은 프로바이더(인터넷 서비스 제공자)에 삭제를 요청할 수 있으며, 삭제 요청을 받은 프로바이더는 법이 정하는 요건을 충족하는 경우에는 해당 게시물에 대한 삭제 등 조치를 강구해도 손해배상 책임에서 면책 · 자신의 권리를 침해당한 개인은 프로바이더 등에게 해당 게시물을 작성한 발신자에 관한 정보 공개를 청구할 수 있음
평가	<ul style="list-style-type: none"> · 프로바이더가 적절한 조치를 취한 경우 손해배상 책임이 면제될 수 있음을 제시한다는 점에서 상대적으로 플랫폼 사업자의

	책임을 약화시키고 있다는 비판 · 피해자는 정보개시를 통해 발신자를 특정해 손해배상청구권을 행사할 수 있으나, 실제 손해배상을 받는 것은 쉽지 않아 피해자 보호에 미흡하다는 의견 존재
--	---

○ 일본 총무성의 허위조작정보 대책 방안

- 선거기간 또는 재해 발생 시 소셜 네트워크에서의 허위조작정보 확산을 방지하기 위한 대책을 추진

주요 내용
첫째, 자율규제원칙
둘째, 허위정보 실태조사
셋째, 관련 당사자의 협력체제 구축
넷째, 플랫폼 서비스 사업자의 투명성 설명책임 확보
다섯째, 팩트체크 추진
여섯째, 정보리터러시 교육 강화
일곱째, 연구활동 지원
여덟째, 정보발신자에 대한 신뢰성 확보
아홉째, 국제협력 중시

3) 해외 허위조작정보 관련 자율규제 실천강령 사례

(1) EU의 허위조작정보 대응 실천강령

- **(배경)** 다양한 허위조작정보들이 글로벌 플랫폼을 통해 빠르게 확산되면서 EU는 대형 플랫폼을 통한 허위조작정보 유통을 유럽의 민주주의를 위협하는 중대한 문제로 인식하고 이에 대한 대응방안을 마련
 - 특히 자율규제 중심의 대응방안은 허위조작정보 대응 실천강령의 구축으로 나타남

구분	내용
온라인 허위조작정보 대응에 관한 정책안의 4가지 원칙	<ul style="list-style-type: none"> · 투명성(transparency) 증대 · 정보의 다양성(diversity) 촉진 · 정보의 신뢰도(credibility) 조성 · 포괄적(inclusive) 해결안 마련
허위조작정보 대응 실천강령(2018) 준칙의 주제영역	<ul style="list-style-type: none"> · 광고 게재에 대한 정밀검토 · 정치광고 및 이슈광고 · 서비스의 완전성 · 소비자에게 권한/자율권 부여 · 연구공동체에게 권한/자율권 부여

- **(2022 강화된 허위조작정보 대응 실천강령)** 기존 ‘2018 허위조작정보 대응 실천강령’에 대한 평가를 실시하여, 해당 강령에 있어 보완이 필요한 사항을 발표하고, 수정·보완 및 현행화를 바탕으로 ‘2022 강화된 허위조작정보 대응 실천강령’ 발표

구분	내용
2018 허위조작정보 대응 실천강령에 대한 보완요청 사항	<ul style="list-style-type: none"> · 맞춤형 준칙을 통한 참여 확대 · 허위조작정보의 수익화 방지 · 서비스의 완전성 보장 · 이용자 역량강화 개선

	<ul style="list-style-type: none"> · 팩트체크 범위 확대 · 연구자들의 데이터 접근성 향상 · 강화된 모니터링 프레임워크 형성
--	---

- o ‘2022 강화된 허위조작정보 대응 실천강령’은 궁극적으로 유럽의 열린 민주사회 구현을 목적으로 하며, 허위조작정보에 주목하고 이에 대응하기 위한 구체적인 방안을 제시
- ‘2022 강령’은 9개의 주제영역에 대한 44개의 준칙(Commitment)과 128개의 조치(Measure)로 구성되며, 각 조치는 해당 강령의 서명인들의 보고 지침이 되는 질적보고요소(QRE)와 서비스수준지표(SLI)를 포함하고 있음
 - 또한, 허위조작정보의 범위를 확대하여, 직전 강령에서 정의한 허위조작정보 외에도 잘못된 정보(misinformation), 정보 영향력 작전(information influence operations), 정보 공간 내 해외 개입(foreign interference in the information space) 등을 대응 대상으로 포함

구분	내용
허위조작정보 영역	<ul style="list-style-type: none"> · 허위조작정보 · 잘못된 정보 · 정보 영향력 작전 · 정보 공간 내 해외 개입
참여자(서명기관) 확대	<ul style="list-style-type: none"> · 플랫폼 기업을 포함한 광고, 팩트체크, 언론, 연구조직 등 44개 기업 및 단체
주제영역	<ul style="list-style-type: none"> · 광고 게재에 대한 정밀검토 · 정치광고 · 서비스의 완전성 · 이용자 역량강화 · 연구 공동체 역량강화 · 팩트체킹 공동체 역량강화(신설) · 투명성 센터(신설) · 상설 태스크포스(신설)

	· 본 강령의 모니터링(신설)
평가	· 참여 조직 간 입장차에 따른 협력 제고 한계 가능성 · 탈퇴 및 미참여 플랫폼 기업의 존재 · 보고 방식 등 영향력 측정 기준의 수립

- (실천강령 이행) 강령에 서명한 기업들은 서명 후 7개월 이내에 강령 준칙 이행 결과를 상세히 설명하는 보고서를 제출하고, 해당 보고서는 각 기업이 웹사이트 형식의 ‘투명성 센터’에 공개하도록 의무화

주제영역(9개) 및 준칙(commitment, 44개)		
① Scrutiny of Ad Placements 광고 게재 위치 조사(C1-3)	④ Empowering Users 이용자 역량 강화(C17-25)	⑦ Transparency Centre 투명성 센터(C34-36)
② Political Advertising 정치 광고(C4-13)	⑤ Empowering the research 연구 역량 강화(C26-29)	⑧ Permanent Task-Force 상설 태스크포스(C37)
③ Integrity of Services 서비스 완전성(C14-16)	⑥ Empowering the fact-checking community 팩트체크 커뮤니티 역량 강화 (C30-33)	⑨ Monitoring of Code 강령 모니터링(C38-44)

Reporting on the service's response during a period of crisis 위기 국면에 대한 서비스 대응 보고(2023년 기준)		Measuring indicator 평가지표
Covid-19 pandemic 코로나-19 팬데믹	War of aggression by Russia on Ukraine 러시아-우크라이나 전쟁	
		SLI, 서비스 수준 지표 QRE, 질적 보고 요소

(2) 호주의 허위조작정보 및 오정보 대응 실천강령

- (배경) 호주는 디지털 플랫폼 규제에 대한 새로운 접근방식을 마련하기 위한 노력의 일환으로 ACCC의 <디지털 플랫폼 조사 최종 보고서>를 통해 사용자의 역량 강화와 디지털 플랫폼의 자발적인 행동규범 마련을 위한 로드맵 제안
 - 2020년 6월 ACMA는 <호주 디지털 플랫폼의 허위조작정보와 뉴스 품질: 강령 개발 지침에 관한 의견서>를 발표하여, 실천강령에 대한 명확하고 광범위한 지침을 제시

- 이에 비영리 산업 협회인 DIGI는 약 5주 간의 공개 협의 과정을 통해 실천강령의 초안을 작성하였고, 디지털 플랫폼, 정부, 학계, 시민사회 등 다양한 이해관계자들과의 협의과정을 거쳐 2021년 2월 ‘호주 실천강령’ 을 공표
- o **(내용)** 호주 실천강령은 EU의 허위조작정보 대응 실천강령을 바탕으로 하여 ‘자율규제’ 라는 점에서 유사성을 지니지만, EU의 준칙이 ‘관련 서명인’ 의 참여를 요구하는 것과 달리 모든 서명인이 준수하고 이행해야하는 최소 준칙을 제시함
- o **(2022 개정된 실천강령)** DIGI는 매년 ACMA에 연례보고서를 제출하고, 2년 주기로 실천강령을 필수적으로 검토하고 개선해야 하는데, 이러한 과정에서 DIGI는 다양한 이해관계자들의 의견을 다수 반영하여 2022년 개정된 실천강령을 공표

구분	내용
변경사항	<ul style="list-style-type: none"> · 소규모 디지털 플랫폼의 참여확장 · 허위조작정보와 오정보 관련 ‘피해’ 의 기준을 ‘심각하고 확실한(믿을 만한)위험’ 으로 변경 · 2022년에 업데이트된 EU의 실천강령을 반영하여 추천 시스템에 대한 시행 규정 추가 · 허위조작정보와 오정보를 유포하는 디지털 광고를 반복적으로 게재하는 것에 대한 규정사항 추가 · 스폰서 콘텐츠와 유료 디지털 광고 서비스를 허위조작정보와 오정보 관련 규정 범위 안에 포함 · 기존의 조항에 따라 전문 뉴스 콘텐츠가 허위조작정보로 다뤄지지 않도록 하되, 해당 콘텐츠가 허위조작정보 확산 시 서명인에 대한 책임과 의무 조항 유지 · 강령에 대한 업데이트 및 투명성 보고서 등 서명인의 준칙 범위 내에서 특정 제품과 서비스에 대한 높은 투명성 요구

구분	내용	
<p>제시된 호주 실천강령의 목표</p>	<ul style="list-style-type: none"> • 허위조작정보와 오정보에 대한 안전장치 제공 • 허위조작정보와 오정보에 대한 광고 또는 수익창출 인센티브 차단 • 디지털 플랫폼에서 제공하는 서비스와 제품의 안전성 및 보완 보장 • 소비자 역량 강화 • 정치 광고 출처에 대한 공중인식 개선 • 전략적 연구 지원 • 서명인의 투명한 대응책 공개 	
<p>용어 정의</p>	<p>허위조작 정보</p>	<p>A. 실증적으로 잘못되었거나, 오도하거나, 기만적인 디지털 콘텐츠</p> <p>B. 이는 부정 행위(Inauthentic Behaviors)를 통해 디지털 플랫폼 사용자들 사이에서 확산된다.</p> <p>C. 해당 내용의 유포가 피해를 끼치리라고 합리적으로 예상되는 콘텐츠</p>
	<p>오정보</p>	<p>A. 실증적으로 잘못되었거나, 오도하거나, 기만적인 (종종 적법한) 디지털 콘텐츠</p> <p>B. 이는 디지털 플랫폼의 사용자에게 의해 확산된다.</p> <p>C. 해당 내용의 유포가 (명확하게 기만을 의도하지 않더라도) 해를 끼치리라고 합리적으로 예상되는 콘텐츠</p>
	<p>허위조작 정보에서 제외되는 콘텐츠</p>	<p>A. (풍자 및 패러디를 포함한) 오락 또는 교육의 목적으로, 선의의 신뢰 하에 제작된 콘텐츠</p> <p>B. 호주 주정부 또는 연방정부에서 승인한 콘텐츠</p> <p>C. 5.23일부터 5.25일에 따라, 정치 광고 또는 (호주 법률에 따라 등록된) 정당이 인가한 콘텐츠</p> <p>D. 전문 뉴스 콘텐츠</p>

- o (평가) EU의 실천강령과 ①범위와 접근방식, ②참여주체의 다양성과 전문성, ③절차의 공정성과 투명성, ④공적 책무에 대한 지향 측면에서 많은 부분이 유사하지만, EU의 실천강령이 디지털서비스법과의 공동규제체계를 이뤄 실효성과 강제력이 뒷받침되는 것과 달리 호주의 실천강령은 공동규제를 위한 법적기반은 존재하지 않음

4) 허위조작정보 관련 국내 자율규제 활성화 방안

(1) 국내 허위조작정보 관련 법적 규제 현황

- 허위조작정보의 전달·유포·게시 행위 규제 및 플랫폼 사업자를 규제하는 법률
 - 허위의 사실 표현과 관련하여 개인의 권리가 침해되는 경우에 한해 표현의 자유를 일정정도 규제
 - 정보통신망법에 따른 게재된 불법정보에 대한 삭제 및 임시조치를 규정하여 플랫폼 사업자를 규제하고 있으며, 불법촬영물과 관련하여 투명성 보고서 제출의무, 정보 책임자 지정 등의 의무를 규정

허위조작정보 전달·유포·게시 등 처벌	플랫폼 사업자의 삭제, 차단, 취급거부·정지·제한조치
<ul style="list-style-type: none"> · 명예훼손죄(형법 제307조, 308조, 309조, 정보통신망법 제70조) · 모욕죄(형법 제311조) · 허위사실공표죄(공직선거법 제250조) · 후보자비방죄(공직선거법 제251조) · 전기통신기본법 제47조제2항의 벌칙규정 	<ul style="list-style-type: none"> · 게재된 정보에 대한 삭제·임시조치(정보통신망법 제44조의2제3항, 4항) · 게재된 정보에 대한 삭제 또는 취급의 거부·정지·제한(공직선거법 제82조의4제3항) · 전기통신법 제47조제4항의 벌칙규정

정보통신망법에 따른 플랫폼 사업자 책임
<ul style="list-style-type: none"> · 제64조의5(투명성보고서 제출의무 등) <ul style="list-style-type: none"> : 일정 규모 이상 정보통신서비스 제공자는 불법촬영물 등 처리에 관한 투명성 보고서를 방통위에 제출 · 제42조의3(정보책임자 지정 등) <ul style="list-style-type: none"> : 일정 규모 이상의 정보통신서비스제공자는 청소년보호책임자 지정 · 44조의9(불법촬영물 등 유통방지 책임자 지정) <ul style="list-style-type: none"> : 일정 규모 이상의 정보통신서비스제공자는 불법촬영물 등 유통방지 책임자 지정 · 제32조의5(국내 대리인의 지정) <ul style="list-style-type: none"> : 일정 규모 이상의 해외사업자는 국내 대리인 지정

- 정보통신망법 개정을 중심으로 허위조작정보를 규제하기 위한 국내 입법 논의
 - 제21대 국회에서는 허위조작정보와 관련된 주요 법률안을 발의
- 정보통신망법에 마련된 사업자 자율규제 법적 근거
 - 정보통신서비스 제공자에 대한 자율규제 행동강령을 정하여 시행할 수 있는 근거가 마련되어 있으나, 자율규제에 대한 구체적인 방식과 내용, 정보통신제공자 단체의 실제 등에 대해서는 별도의 규정은 없음

구분	주요 내용
정부지원	<ul style="list-style-type: none"> · 제3조(정보통신서비스제공자 및 이용자의 책무) <ul style="list-style-type: none"> : 정부는 정보통신서비스 제공자 단체 또는 이용자 단체의 자율규제를 위한 활동을 지원할 수 있도록 함 · 제44조의4(자율규제) <ul style="list-style-type: none"> : 정부는 정보통신서비스 제공자 단체의 행동강령 지정 및 시행에 대한 지원과 불법 및 청소년 유해정보 자율규제 활동을 지원할 수 있도록 함
사업자 자율규제	<ul style="list-style-type: none"> · 제44조의4(자율규제) <ul style="list-style-type: none"> : 사업자 단체는 자율규제 행동강령을 정할 수 있음 : 사업자 단체는 불법 및 청소년 유해정보에 대한 자율규제 활동을 시행할 수 있음

(2) 국내 허위조작정보 관련 사업자 자율규제 사례

- (KISO) 시장이 주도해 설립한 한국 최초의 민간 자율규제기구
 - 현재 국내 주요 포털사의 인터넷 불법정보 심의를 위임, 해외 대규모 플랫폼은 참여하고 있지 않은 한계점 존재
 - KISO에서는 허위정보와 관련하여 인터넷 포털 사업자 홈페이지에 존재하는 가짜뉴스를 규제하고 있음

o (플랫폼 민간 자율기구) 민간 주도의 플랫폼 자율규제 방안 마련을 논의하기 위해 출범된 기구

- ‘검색·추천 서비스의 투명성 제고를 위한 자율규제 원칙’을 마련하여 (i) 검색·추천 기준 공개, (ii) 이용자의 권익 증진을 위한 노력, (iii) 이행점검 등 실효성 확보 등 세가지 방안을 구성

o (포털사업자) 네이버 및 카카오

- 허위조작정보와 관련하여 KISO의 <정책규정>을 준수하여 게시물 규제를 실시
 - *코로나19와 관련한 허위조작정보
 - *<언론보도 형식의 허위 게시물>인 가짜뉴스
 - *다른 이용자를 기만할 목적으로 타인을 사칭하거나 허위사실을 주장하는 내용의 게시물
- 포털 사업자는 운영정책을 통해 운영정책을 위반한 게시물 및 댓글에 대한 신고절차를 안내하고, 신고센터를 통해 접수를 받고 있음

(3) 허위조작정보 자율규제 시범 가이드라인(안) 제언

구분	내용
필요성	<ul style="list-style-type: none"> · 온라인 허위조작정보에 대한 대응 필요성이 증가함에 따라, 이에 대응하기 위해 인터넷 플랫폼 사업자들이 책임 있게 나서야 할 필요성 또한 존재 · 해외 사례를 참고하여, 국내에서도 보다 포괄적이면서 실효성을 확보하기 위한 자율규제 시범 가이드라인을 모색
허위조작정보의 정의	<ul style="list-style-type: none"> · 정의: 정치적 또는 경제적 이익 등을 얻으려는 목적에서 제작·유포되며 공공에 해를 끼칠 수 있는 허위이거나 오도하는 정보 · 제외: 풍자 및 패러디, 다소의 오류가 있거나 일부 사실에 부합하지 않는 과장된 표현이 있는 정보

이행원칙	(1) 허위조작정보에 대한 공동대응 (2) 비례의 원칙 (3) 표현의 자유에 대한 보호 (4) 개인정보의 보호
주요 원칙	(1) 허위조작정보의 신고 및 조치 (2) 허위조작정보의 수익 차단 (3) 서비스의 완전성(integrity) (4) 검색·추천 서비스의 투명성 (5) 이용자 역량 강화 (6) 팩트체크 및 연구 지원 (7) 본 가이드라인의 모니터링

5. 기대효과

- 본 연구의 결과는 국내 인터넷 사업자의 자율적인 허위조작정보 대응을 위한 자율규제 가이드라인 마련에 참고자료로 활용할 수 있음
 - 정책적인 측면에서 온라인 허위조작정보에 대응하기 위한 대응체계를 설정하는데 기여할 수 있을 것으로 기대됨
 - 경제·사회적인 측면에서는 플랫폼을 통해 유통되는 정보와 서비스의 신뢰를 강화함으로써 온라인을 통한 다양한 사회적·경제적 거래를 활성화하고, 안전한 온라인 환경을 구축함으로써 이용자를 보호하는데 기여할 수 있을 것으로 기대됨

제 1 장 서론

제 1 절 연구 배경 및 필요성

2010년대 중반 이후 각국의 선거와 여론형성과정에서 허위조작정보(disinformation)가 사회적 혼란을 야기하거나 시민의 정당한 민주적 선택을 방해하는 일이 빈번하게 발생하면서 허위조작정보에 대한 논의가 세계적으로 확산되고 있다. 2016년 미국의 대선과 영국의 브렉시트 국민투표에서 허위조작정보가 확산돼 사회적 혼란을 야기했다. 같은 해 독일에서는 러시아계 독일 소녀가 아랍계 난민으로부터 성폭행을 당했다는 러시아 발(發) 허위조작정보가 소셜미디어를 중심으로 유포돼 난민 정책과 관련한 갈등을 촉발하고 독일과 러시아 사이의 외교 문제로까지 비화됐다(최지선, 2022). 이에 옥스퍼드 영어사전은 2016년 ‘올해의 단어’로 ‘탈진실(post-truth)’을 선정하기에 이르렀다.¹⁾

한국에서도 2017년 대선을 전후로 허위조작정보와 관련된 우려의 목소리가 높아졌다(김준일, 2021). 이는 2022년 대선에서도 마찬가지였다. 한국언론진흥재단이 진행한 조사결과에 따르면 대선을 앞두고 조사 응답자의 45.7%가 대선 관련 허위조작정보나 가짜뉴스를 접했다고 밝혔다(김선호, 2022.2.9.).

2020년부터는 코로나19 팬데믹이 전 세계로 확산되면서 국내외를 막론하고 부정확한 정보가 온라인을 통해 급속도로 퍼져나가는 인포데믹(infordemic)²⁾ 현상이 확산됐다. 허위조작정보는 정치·사회적 혼란을 야기하는 것을 넘어 인간의 생명에 직접적인 위협을 가하게 됐다. 한 연구에 따르면 코로나19와 관련된 잘못된 정보 때문에 2020년 1월~3월 기간 동안 전 세계에서 5,800명 이상이 병원 신세를 지고 800명 이상이 사망했다(Islam et al, 2020). 한국에서도 한 남성이 메탄올이 코로나19를 소독한다는 인터넷 정보를 믿고 집안에 메탄

1) 탈진실은 객관적인 사실(fact)이 감정(emotion)과 신념(belief)보다 여론 형성에 영향력을 덜 미치는 현상을 가리킨다(BBC, 2016.11.16.).

2) 인포데믹(infordemic) : 정보(information)와 팬데믹(pandemic)의 합성어로서, 코로나19 발생 기간에 확산된 허위정보의 급증을 설명하는데 사용된 용어. 2003년 데이비드 로스 코프(David Rothkopf)의 워싱턴 기고문에서 ‘SARS 공포로 아시아 경제가 추락한 사례’를 거론하며 처음 인용(성욱제·정은진, 2020).

을 살포한 뒤 구토와 어지럼증으로 병원에 실려 가는 일이 발생했다(메디게이트뉴스, 2021.12.21.). 허위조작정보에 대한 대응이 감염병 대응만큼 중요한 문제가 된 것이다.

물론 허위조작정보는 최근 몇 년 간 나타난 새로운 현상은 아니다. 하지만 인터넷·모바일 환경에서 소셜미디어와 동영상 스트리밍 사이트, 검색엔진을 통해 검증되지 않은 정보가 빠르게 확산되면서 허위조작정보의 파급력, 피해의 정도 및 규모는 유례없는 수준까지 커졌다. 소셜봇(Social Bots)³⁾, 딥페이크(Deepfake)⁴⁾, 생성형 AI 등 기술의 발달이 허위조작정보의 생산과 유통에 가하는 새로운 도전과 위협에 직면하면서 세계 각국은 허위조작정보에 대한 규제를 강화하기 위해 다양한 정책방안을 모색하고 있다. 주요 인터넷 사업자와 언론기관 등 민간영역에서도 자율규제와 팩트체크를 강화하며 발을 맞추고 있다.

각국의 대응방안에서 특히 주목해야 할 것은 인터넷 사업자의 자율규제이다. 빠르게 변화하고 있는 온라인 환경을 현행 법체계가 발 빠르게 쫓아가기 어렵고, 대량으로 유통되는 정보를 공적 기관이 일일이 모니터링하고 규제하는 것도 현실적으로 불가능하기 때문이다. 동시에 자율규제의 실천과 실효성 강화를 위해 정부와 인터넷사업자, 시민사회의 협력도 긴밀해지고 있다. EU, 호주 등에서 마련된 허위조작정보 대응 실천강령은 이와 같은 노력의 일환으로 이해될 수 있다.

이에 본 연구에서는 국내외 허위조작정보 규제와 관련된 주요 현황 및 논의를 비교·분석하고 시사점을 도출함으로써 인터넷 사업자의 자율규제를 활성화하기 위한 방안을 모색하고자 한다. 이는 향후 허위조작정보와 관련된 자율규제 체계의 확립 및 관련 거버넌스 개선에 시사점을 제공할 것으로 기대한다.

3) 소셜봇(Social Bots) : 봇(Bots)은 사용자나 다른 프로그램 또는 사람의 행동을 흉내 내는 대리자로 동작하는 프로그램인 로봇(Robot)의 준말로써(윤명희·손수빈, 2015, 288쪽), 소셜봇은 소셜네트워크 상에서 동작하는 봇을 가리킨다. 소셜봇은 SNS에서 몇 백 개의 글을 포스팅하고, ‘좋아요’ 를 누르며, 공유하고, 유사한 또는 동일한 문장으로 여러 플랫폼을 돌아다니며, 장시간이 소요되는 토론에도 간여할 수 있다. 소셜봇에 의해 특정한 의견에 대한 포스팅이 타임라인에 자주 등장하게 되면, 이것이 의견 형성에 영향력을 행사할 수 있다(김태오, 2020, 85쪽).

4) 딥페이크(Deepfake) : 딥러닝(Deep learning)과 가짜(Fake)의 합성어로서, 2017년 12월 ‘딥페이크’ 라는 닉네임을 사용하던 익명의 이용자가 소셜 뉴스 웹사이트인 레드(Reddit)에 유명인의 얼굴을 합성해서 만든 포르노그래피를 게시하였던 데에서 유래(허순철, 2022)

제2절 연구의 구성

본 연구는 허위조작정보와 관련된 국내외 주요 정책 현황과 자율규제 실천강령을 비교 분석함으로써, 인터넷 사업자의 자율규제를 활성화하기 위한 정책방안을 제안하는 것을 목적으로 한다. 이를 위한 연구의 내용은 다음과 같이 구성되어 있다.

우선 허위조작정보와 관련된 주요 법제도 현황과 자율규제 사례를 살펴보기에 앞서 국내외 허위조작정보 유통현황을 간단히 짚어보도록 한다. 이는 허위조작정보와 관련된 정책 및 자율규제 관련 논의를 이해하기 위한 맥락을 짚어보는 일이기도 하다. 구체적인 유통 현황 파악을 통해 현안의 문제점을 구체적으로 이해하고, 현재의 미디어 환경에 부합하는 대응방안을 마련하는데 기여하고자 한다.

둘째, 허위조작정보와 관련된 해외 법제도 현황 및 정책대응 현황을 분석하도록 한다. 허위조작정보에 대응하기 위한 국제적인 규제 논의 동향을 파악하는 한편 법적 규제의 적용 범위, 효과 및 한계를 이해하는 것은 국가와 인터넷 사업자, 시민사회의 협력 필요성을 논의하기 위한 토대가 될 것이다.

셋째, 자율규제를 위한 실천강령 제정의 사례로써 EU의 허위조작정보 대응 실천강령, 호주의 허위조작정보 및 오정보 대응 실천강령을 비교 검토한다. 각 지역·국가의 상황에 따라 다양하게 적용되고 있는 방안들을 비교분석함으로써 해당 내용의 시사점과 국내 적용 가능성을 검토하도록 한다.

넷째, 국내외 허위조작정보 관련 법적 규제와 자율규제 사례를 분석하고 이를 앞서 분석한 해외의 법적 규제 및 자율규제 실천강령과 비교함으로써, 허위조작정보 대응을 위한 방안을 모색하고자 한다. 특히 허위조작정보 자율규제를 위한 사업자 가이드라인을 제안함으로써 현실적으로 실천 가능한 대안을 수립할 것이다. 이는 향후 허위조작정보와 관련된 자율규제 체계의 확립 및 관련 거버넌스 개선에 시사점을 제공할 것으로 기대한다.

제3절 용어의 정의

사실이 아닌 정보(false information)를 지칭하는 어휘는 다수 존재하며 어느 정도는 혼용되고 있기도 하다. 특히 허위조작정보(disinformation)는 가짜뉴스(fake news)라는 용어와 혼용되는 경우가 빈번하다. 또 오정보(misinformation)와도 혼용될 수 있다. 때문에 각 개념의 차이를 짚고 넘어가는 일이 필요하다.

허위조작정보에 대해서도 학자마다 정의에 차이가 있다. 다만 유럽연합집행위원회(European Commission, EC)가 2018년 「허위조작정보에 대한 다차원적 접근(A Multi-dimensional approach to disinformation)」 보고서를 통해 허위조작정보를 “고의적으로 공공에 해를 끼치거나 이익을 얻을 목적으로 설계, 제작, 유포되는 허위이거나, 부정확하거나, 또는 사람들을 오도하는 정보(all forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit)” 로 정의한 이후(EC, 2018) 이 정의가 널리 통용되고 있다. OECD 역시 유럽집행위원회의 정의를 따라서 허위조작정보를 “경제적 이익을 위해 의도적으로 개인, 사회집단, 조직 또는 국가를 고의로 속이거나 조작 또는 위해를 가하기 위해 의도적으로 생성되고 유포되는 확인 가능한 허위 혹은 오도하는 정보(verifiably false or misleading information that is knowingly and intentionally created and shared for economic gain or to deliberately deceive, manipulate or inflict harm on a person, social group, organization or country)” 로 이해한다(OECD, 2022).

유럽연합집행위원회와 OECD는 공통적으로 허위조작정보를 오정보 및 기타 풍자(satire) 등과 구분해 사용할 것을 권고한다. 오정보는 거짓 또는 오해의 소지가 있는 정보이지만 개인, 사회집단, 조직 또는 국가를 의도적으로 속이거나 조작 또는 위해를 끼칠 의도가 없는 것으로, 정보 유포자가 초기 오정보를 생산하거나 조작하지 않은 정보를 가리킨다. 풍자는 유머와 과장을 이용해 사람이나 아이디어를 비판하는 언어, 영화 또는 기타 예술 작업들로 종종 사회적이거나 정치적인 논평의 형태를 취한다. 하지만 이와 같은 풍자는 종종 공유되고 또 공유되는 과정에서 의도치 않게 본래의 의미가 상실되고 이용자들이 이를 오해할 수 있다. 정리하자면 허위조작정보는 위해를 가할 목적으로 의도적으로 생산 및

유포되는 정보이다. 오정보는 위해를 가할 목적을 갖고 있지 않으며 의도적으로 조작되지도 않은 정보이다. 풍자는 의도적으로 꾸며낸 정보이나 위해를 가할 목적은 갖고 있지 않다(OECD, 2022).

가짜뉴스와 허위조작정보는 전자가 주류 뉴스의 형태를 차용한다는 점을 제외하면 동일하다고 할 수 있다. 그럼에도 유럽연합집행위원회는 위에서 언급한 보고서를 통해 가짜뉴스 대신 허위조작정보 또는 오정보라는 용어의 사용을 제안한다. 그 이유는 크게 두 가지이다. 첫째, 가짜뉴스는 사실과 조작된 정보가 뒤섞여 있는 오늘날의 허위조작정보가 발생시키는 복잡한 문제를 포괄하기에 부적합한 개념이다. 가짜뉴스는 자동계정, 가짜 팔로워로 구성된 네트워크, 조직화된 트롤링 등 ‘뉴스’ 형식을 넘어선 행위들을 포괄하지 못하는 문제점을 안고 있다. 마찬가지로 정보의 유통과정에서 나타나는 댓글, 공유, 트윗 및 리트윗 등 전통적인 뉴스 행위로 담아내지 못하는 다양한 디지털 행위들을 포괄하지 어렵다는 한계도 안고 있다. 둘째, 일부 정치인과 지지자들이 자신이 동의하지 않는 보도를 묵살하기 위해 가짜뉴스(fake news)라는 용어를 사용해오며 따라 가짜뉴스라는 용어는 오해를 낳고 있다(EC, 2018). 이와 같은 제안 이후 유럽연합의 공식 문서들은 ‘허위조작정보’라는 용어를 지속해서 사용하고 있다(김민정, 2019).

오늘날 허위조작정보라는 용어는 특정한 행위자, 특정한 행위, 특정한 형식뿐만 아니라 뉴스미디어, 인터넷 플랫폼, 그리고 그 기저에 깔려 있는 알고리즘과 같은 다양한 기술적인 층위를 만들어지고 있는 문제의 구조를 포괄한다(김민정, 2019). 이는 검색엔진, 동영상 스트리밍 사이트, 소셜미디어 등 다양한 플랫폼에서 게시하기, 댓글달기, 인용하기, 트윗 및 리트윗하기, 좋아요 누르기, 퍼가기 등 다양한 방식으로 정보가 유통되고 확산되며, 인간 행위자뿐만 아니라 소셜봇, 딥페이크, ChatGPT 등 다양한 비인간 행위자가 뒤섞여 있는 작금의 미디어 정보 환경을 다루기에 보다 적합한 용어라고 할 수 있다.

한 가지 짚고 넘어가야 할 것은 허위조작정보가 곧 불법콘텐츠를 의미하는 것은 아니라는 점이다. 유럽연합집행위원회는 2018년 「온라인 허위조작정보에 대한 대처: EU 접근법 (Tackling online disinformation: a European Approach)」 통신문을 통해 허위조작정보가 비록 사회에 해를 끼친다 하더라도 그것이 불법콘텐츠가 아닌 한, 표현의 자유에 의해 보호되는 콘텐츠임을 명백히 한다. 다시 말해 허위조작정보 자체를 금지하는 법률이 없는 한, 허위조작정보는 유해한 콘텐츠이지만 적법한 콘텐츠라는 것이다. 이에 따라 EU의

「디지털 서비스법(Digital Service Act)」, 그리고 「허위조작정보에 대응 실천강령(Code of Practice on Disinformation.)」 등은 주로 온라인 플랫폼에 의한 허위조작정보의 유통과 확산을 규제하는데 초점을 맞추고 있다.

이상의 논의를 바탕으로 본 보고서에서는 가짜뉴스 대신 허위조작정보라는 용어를 채택하고 이와 관련된 국내외 사례 및 정책을 분석하도록 한다. 다만 여전히 관련 개념의 사용이 통일되지 않았고, 여러 언론과 연구 등에서 가짜뉴스, 허위조작정보, 오정보가 혼용되고 있는 만큼, 현황자료를 분석하는 과정에서는 고의성 여부를 반드시 고려하지 않고 허위조작정보와 오정보 전반을 다룰 계획이다. 또 다양한 문헌자료를 인용하는 과정에서 여러 용어가 혼용될 경우 구분의 필요성이 있을 때에는 해당 문헌에서 사용한 용어를 그대로 인용하고, 그렇지 않은 경우는 맥락에 따라 이를 허위조작정보로 통칭하도록 한다.

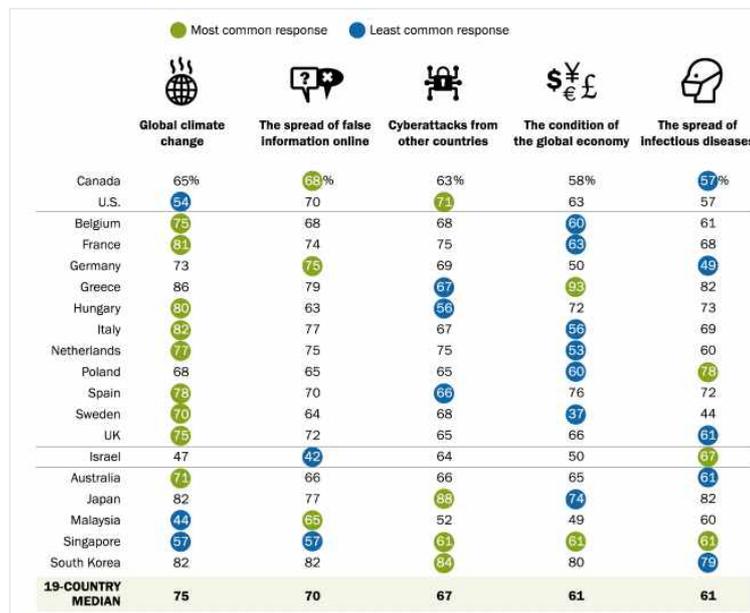
제 2 장 국내외 허위조작정보 유통 현황 및 영향력

제 1 절 해외 허위조작정보 유통 현황 및 영향력

1. 온라인 미디어에서의 허위조작정보

허위조작정보에 대한 우려는 전 세계적으로 공통된 현상이라고 할 수 있다. 예를 들어 퓨리서치센터가 2022년 봄 19개국을 대상으로 진행한 조사결과에 따르면 세계 시민의 70%는 허위조작정보를 중대한 위협으로 인식하고 있었다. 이는 기후변화(75%)의 뒤를 이어 두 번째 순위이다(Pew Research Center, 2022.8.31.).

[그림 2-1] 전 세계가 직면한 위협 순위 (19개국, 단위: %)

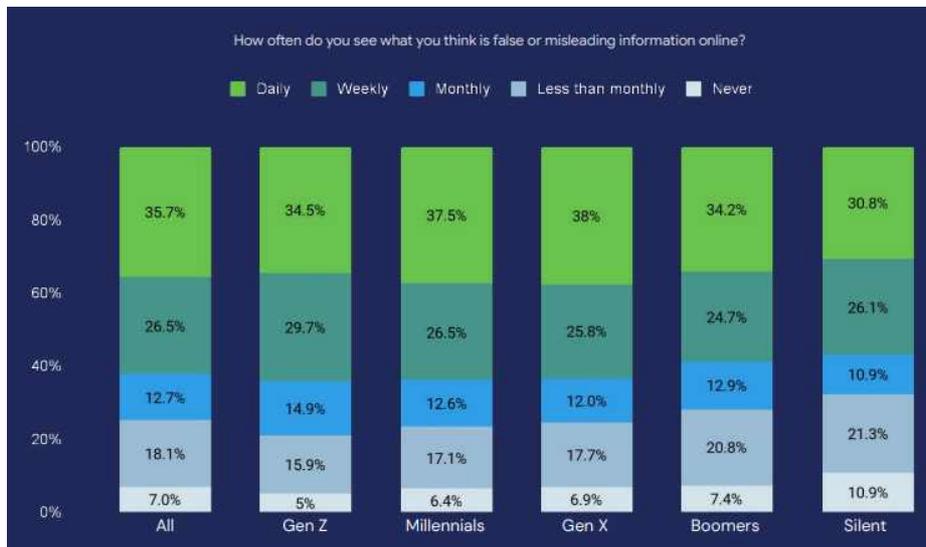


*자료: Pew Research Center (2022.8.31.)

허위조작정보의 심각성에 대한 이와 같은 인식은 인터넷과 소셜미디어의 영향력이 전통적인 미디어의 영향력을 넘어선 오늘날의 미디어 환경과 맞닿아 있다고 할 수 있다.

구글의 의뢰에 의해 포인터, 미디어와이즈, 유거브가 2022년 상반기 7개국(미국, 브라질, 영국, 독일, 나이지리아, 인도, 일본) 시민을 대상으로 조사한 결과에 따르면 세계 시민의 62%는 적어도 일주일에 한번 이상 온라인에서 허위정보(false or misleading information)를 접한다고 응답했다. 매일 허위정보를 접한다는 응답자도 36%에 달했다(Poynter, Mediawise, YouGov, 2022.8.).

[그림 2-2] 세대별 허위정보를 접하는 빈도(7개국, 단위: %)



*자료: Poynter, Mediawise, YouGov. (2022.8.) A Global Study on Information Literacy.

정보의 출처가 검색엔진일 때에는 상대적으로 정보 수용자들이 해당 정보를 검증하는 비율이 높았다(44%). 하지만 소셜미디어가 출처인 정보에 대해서는 정보를 검증하는 비율이 상대적으로 낮은 것으로 나타났다. 인스타그램 발 정보를 검증하는 비율은 29%, 트위터는 22%, 틱톡은 16% 순이었다.

허위조작정보에 대한 인식도 세대별로 달랐다. 조사에 따르면 젊은 세대는 정보의 진위성 여부를 판단하는데 가장 어려움을 겪는 세대였지만, 또한 정보의 진위를 검증하는데 가장 적극적인 세대였다. 반면에 베이비부머 세대와 침묵 세대(68세 이상)는 정보의 진위를 판단하는데 어려움을 느끼지 않는다고 답했지만 상대적으로 정보를 직접 검증하는 비율은 떨어졌다(Poynter, Mediawise, YouGov, 2022.8.)

[그림 2-3] 세대별 정보의 출처에 따른 검증 시도 비율(7개국, 단위: %)

Source of information	All	Gen Z	Millennials	Gen X	Boomers	Silent
A search engine, such as Google or Bing	44%	51%	49%	45%	39%	29%
A messaging app, such as WhatsApp, Telegram or Signal	39%	48%	46%	40%	28%	19%
Facebook	36%	38%	42%	37%	32%	22%
A video platform, such as YouTube or Vimeo	34%	44%	41%	35%	26%	15%
National news websites	33%	32%	38%	33%	31%	22%
Local news websites (for your city or county)	31%	32%	37%	31%	30%	20%
A news app	30%	31%	34%	32%	28%	21%
Instagram	29%	44%	39%	27%	14%	7%
Online blogs	25%	33%	31%	24%	19%	11%
Twitter	22%	30%	28%	22%	13%	6%
TikTok	16%	28%	22%	15%	7%	3%
Snapchat	12%	20%	15%	9%	5%	2%
Other	13%	22%	20%	14%	5%	3%

*자료: Poynter, Mediawise, YouGov. (2022.8.) A Global Study on Information Literacy.

이와 같은 조사결과에 따르면 사람들은 소셜미디어를 통해 유통되는 정보에 대해서는 상대적으로 정보를 잘 검증하지 않는다. 그런데 허위조작정보는 소셜미디어 상에서 더 활발하게 유통된다. 로이터저널리즘연구소의 2021년 조사결과에 따르면 전 세계 사람들이 허위조작정보의 전파 경로로 우려하는 채널은 페이스북(29%), 뉴스 웹사이트 또는 앱(20%), 메시징앱(14%) 순서로 나타났다(오세욱·박아란·최진호, 2021). 한편 모질라 재단(Mozilla Foundation)이 2021년 펴낸 유튜브 유감 보고서(YouTube Regrests)에서는 허위정보, 차별과 혐오 등 신고된 영상의 71%는 유튜브의 자동 추천시스템에서 나왔다고 보고했다(이성규, 2022).

소셜미디어에서 허위조작정보는 진실된 정보보다 더 많이 그리고 더 빠르게 공유되는 경향을 보인다. MIT 연구팀이 2006년부터 2017년 사이 트위터에서 유포된 뉴스 12만 6천 건(300만 명이 4, 50만 번 리트윗)을 분석한 결과에 따르면 가짜뉴스(fake news)가 진짜 뉴

스(true news)보다 70% 가량 더 많이 공유된 것으로 나타났다. 특히 정치와 관련된 가짜 뉴스는 다른 분야의 가짜뉴스보다 더 빨리 확산되는 것으로 나타났다. 진짜뉴스는 1천명 이상 리트윗되는 경우가 드물었지만 가짜뉴스는 1만 명까지 공유되는 경우도 적지 않게 발견됐다. 한편 1,500명에게 도달하는 데 걸린 시간을 비교한 결과에서도 가짜뉴스는 진짜 뉴스보다 6배 더 빨리 확산되는 모습을 보였다(Vosoughi, Roy & Aral, 2018).

2. 선거와 관련한 허위조작정보

미국의 2016년 대선은 허위조작정보의 영향이 그 어느 때보다 크게 작용했다고 평가받는 선거이다. 대선 캠페인 기간 동안 소셜미디어를 통해 전파된 허위조작정보가 수많은 사람들에게 전파되어 선거에 영향을 미쳤다. 버즈피드 뉴스(BuzzFeed News)가 페이스북에서 유통되는 가짜뉴스(fake news) 조사한 결과에 따르면 2016년 페이스북에서 가장 활발하게 공유된 50개 가짜뉴스 중 23개가 미국 정치에 관련된 뉴스였다. 특히 대선 캠페인 마지막 4개월 동안 페이스북에서 가장 많이 전파된 20개의 가짜뉴스와 뉴욕타임즈 등 19개 주요 언론사의 20대 뉴스를 비교한 결과에 따르면, 가짜뉴스는 주요 언론사의 뉴스보다 공유·반응·댓글 등에서 더 높은 페이스북 활동횟수를 기록했다. 버즈피드 뉴스는 약 75%의 미국 유권자가 가짜뉴스 헤드라인에 속는다는 조사 결과를 내놓았다(Silberman, 2016: 고선규 외, 2021에서 재인용). 선거 전 5개월 동안 트위터에서 유통된 약 3천만 개의 뉴스를 분석한 또 다른 연구도 이 시기 트위터 뉴스 중 약 25%가 가짜뉴스(fake news)이거나 극도로 왜곡된 뉴스라고 보고했다(Bovet & Makse, 2019: 고선규 외, 2021에서 재인용).

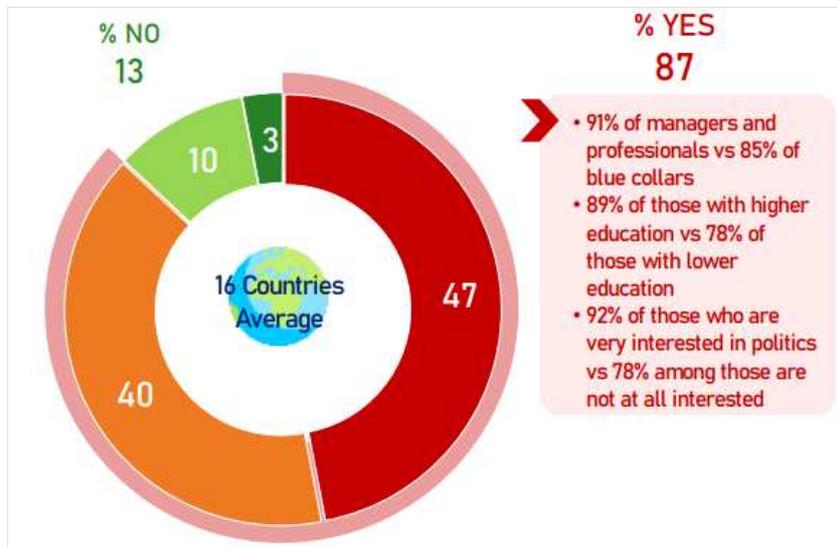
미국의 대선에 가짜뉴스가 미친 영향을 분석한 한 연구 결과에 따르면, 당시 대선후보였던 힐러리 클린턴과 관련된 가짜뉴스는, 2012년 오바마에 투표한 유권자들이 2016년에는 클린턴을 지지하지 않도록 만드는데 영향을 미쳤다. 분석에 따르면 가짜뉴스를 진짜라고 믿을 때 2012년 민주당 후보를 지지했던 유권자들이 2016년에는 민주당 후보를 지지하지 않을 확률이 3배 이상 증가했다(Gunther et al, 2018: 고선규 외, 2021에서 재인용).

선거 기간의 허위조작정보에 대한 우려는 미국에만 국한된 것이 아닌 전세계적인 현상이다. 입소스와 유네스코가 협력이 2024년에 선거가 열리는 16개 국가 8천 명의 시민을 대

상으로 실시한 설문조사 결과에서도 이와 같은 결과가 잘 드러난다(Ipsos & Enesco, 2023.9.).

조사대상의 87%는 허위조작정보와 가짜뉴스가 자국의 2024년 선거 캠페인 및 투표에 미칠 영향에 대해 우려하고 있다고 응답했다. 명백히 우려한다는 응답도 47%였다. 이를 우려하지 않는다는 응답은 13%에 불과했다.

[그림 2-4] 허위조작정보가 선거에 미치는 영향에 대한 우려 정도(16개국, 단위: %)

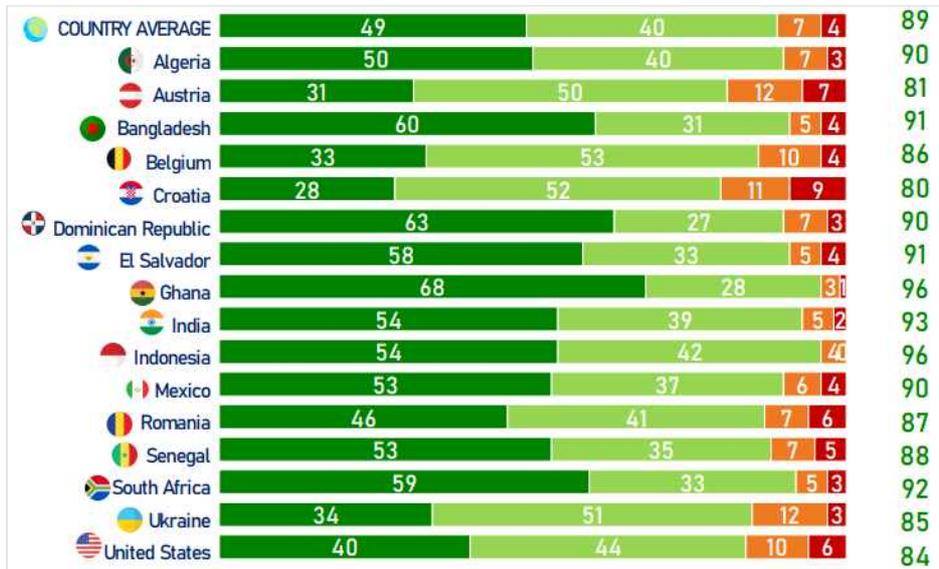


*자료: Ipsos & Unesco (2023.9.) Survey on th impact of online disinformation and hatw speech.

이와 관련해 응답자의 48%는 선거 운동 기간 동안 허위조작정보와 관련된 온라인 콘텐츠를 신고한 경험이 있다고 응답했다. 특히 18~34세 연령대에서(55%), 그리고 정치 고관여층에서(55%) 이러한 허위조작정보를 신고하는 비율이 높게 나타났다.

한편 온라인상의 허위조작정보와 혐오발언(hate speech)을 막기 위해 응답자의 89%는 정부와 규제당국이 선거 운동기간 동안 소셜미디어 플랫폼에 신뢰 및 안전 조치를 취하도록 요구해야 한다고 인식하고 있었다.

[그림 2-5] 선거 운동 기간 동안 정부와 규제당국이 소셜미디어 플랫폼에 신뢰 및 안전조치를 요구할 수 있다고 인식하는 비율(16개국, 단위: %)



*자료: Ipsos & Unesco (2023.9.) Survey on the impact of online disinformation and hate speech.

3. 코로나19와 관련한 허위조작정보

선거와 함께 허위조작정보 유통의 심각성에 경종을 울린 것은 코로나19 팬데믹 상황이었다. 코로나19 팬데믹은 바이러스와 함께 코로나19와 관련된 허위조작정보의 확산을 촉진했다. 로이터저널리즘연구소가 2021년 46개국 시민들에게 지난 한 주 동안 접한 허위정보의 주제를 물어봤을 때 코로나(54%)는 정치(43%), 유명인(29%), 기후변화(20%)보다 훨씬 높은 비중을 차지했다(오세욱·박아란·최진호, 2021).

로이터저널리즘연구소의 조사에 응답한 46개국 시민들은 허위조작정보가 확산되는 주

요한 경로로 페이스북(28%)을 지목했다. 다만 이는 국가마다 달랐다. 미국과 영국, 필리핀에서는 페이스북이 허위정보 경로로 활용되는 데에 대한 우려가 상당히 높았다. 이에 비해 한국에서는 유튜브(34%)에 대한 우려가 압도적으로 높게 나타났다.

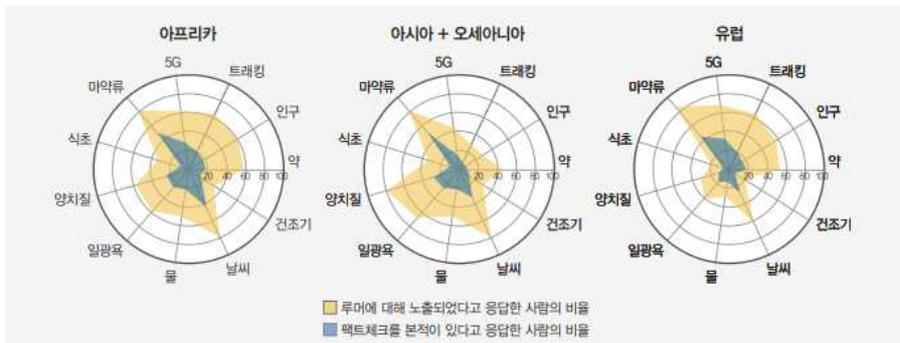
[그림 2-6] 코로나19 관련 허위정보 경로로 우려되는 온라인 미디어 플랫폼 (46개국, 단위: %)



*자료: 오세욱·최진호·박아란(2021). 디지털 뉴스 리포트 2021 한국.

정치 분야에서 가짜뉴스가 진짜뉴스보다 더 많이 확산되는 것(Vosoughi, Roy & Aral, 2018)과 비슷하게, 코로나19와 관련한 정보 유통에 있어서도 허위조작정보는 진짜 정보보다 더 많이 확산됐다. 한국의 기초과학연구원이 2020년 6월 18일부터 7월 13일까지 전 세계 151개국을 대상으로 조사를 실시하고 그 중 40개국을 대상으로 분석한 결과에 따르면 코로나19와 관련된 허위조작정보를 접한 사람의 비율에 비해 팩트체크 정보를 접한 사람의 비율은 훨씬 낮았다(차미영, 2022).

[그림 2-7] 코로나19 관련 인포데믹과 팩트체크에 대한 노출 조사 (단위: %)



*자료: 차미영(2022). 40개국 데이터로 살펴본 코로나19 인포데믹의 여파.

위 그래프에서 노란색은 부분은 설문에 참여한 응답자 중 코로나19 관련 허위조작정보를 접한 적이 있다고 응답한 사람의 비율을 나타내고, 보라색 부분은 해당 허위조작정보에 대한 팩트체크 정보를 본 적이 있다고 응답한 사람의 비율을 나타낸다. 예를 들어 아시아와 오세아니아에서는 소금물 양치를 하면 바이러스가 퇴치된다는 루머를 본 비율이 80%에 달하는 반면 이와 관련된 팩트체크 정보를 본 사람의 비율은 30%에 미치지 못했다. 전반적으로 허위정보의 노출 비율은 팩트체크 노출의 2배 이상이었다(차미영, 2022).

허위조작정보의 유통과 확산은 코로나19에 적절히 대응하는데 심각한 위협이 된 것으로 보인다. 사이폴 이슬람 세계보건기구 국제설사병연구센터 방글라데시본부 감염병학부 연구원 연구팀에 따르면 코로나19와 관련된 허위조작정보 때문에 2020년 1~3월 기간 동안 적어도 800명 이상이 사망했다. 연구팀은 2019년 12월 31일부터 2020년 4월 5일까지 페이스북, 트위터 등 주요 소셜미디어와 웹사이트, 온라인 뉴스, 팩트체크 사이트에 떠도는 허위조작정보를 추적해 87개국에서 25개 언어로 작성된 2,311건의 잘못된 정보를 확인했다. 잘못된 정보의 유형으로는 증상과 전파, 사망률에 관한 정보(24%)가 가장 많았다. 그 다음으로는 병을 막는 방법(21%), 치료법(19%) 감염병의 기원이나 유래(15%) 순서로 나타났다 (Islam et al, 2020).

생명에 대한 직접적인 위협 외에도 허위조작정보는 다양한 사회적 혼란을 야기했다. 2020년 마스크를 만드는 재료가 휴지인데 중국 수입이 막혀 휴지 값이 폭등할 것이라는 일본 발 가짜 뉴스가 확산되자 미국, 호주 등에서 휴지 사재기가 나타나는 등 사회적 혼란이 발생했다(국민일보, 2020.3.8). 미국에서는 미 국방부 산하기관이 중국을 공격하기 위해 코로나19를 만들었다는 음모론이 트위터 등을 통해 확산됐고 아시아인에 대한 인종차별적 가짜뉴스도 만연했다(연합뉴스, 2020.3.1.) 한편 영국에서는 2020년 4월 5G 이동통신 기술이 바이러스를 확산시킨다는 허위조작정보로 5G 기지국 방화사건이 발생했다. 5G 이동통신 기술이 바이러스와 연결되어 있다는 허위조작정보는 영국 외에도 크로아티아, 세르비아, 체코, 아일랜드, 리투아니아, 포르투갈 등 유럽 전역에 확산됐다(황혜인, 2021).

제 2 절 국내 허위조작정보 유통 현황 및 영향력

1. 온라인 미디어에서의 허위조작정보

한국의 상황도 세계 각국과 크게 다르지 않다. 인터넷 플랫폼과 소셜미디어는 한국에서도 뉴스와 정보를 이용하는 가장 중요한 창구가 됐기 때문이다.

한국언론진흥재단의 조사에 따르면 2022년을 기점으로 국민의 온라인 뉴스 이용률(77.2%)은 텔레비전 뉴스 이용률(76.8%)을 넘어섰다. 온라인에서 뉴스를 이용하는 주된 창구는 인터넷 포털(75.1%), 온라인 동영상 플랫폼(20.0%), 메신저 서비스(12.0%), SNS(7.5%) 순서로 높게 나타났다. 인터넷 포털에 대해서는 언론으로서의 신뢰도도 높았다. 인터넷 포털을 언론으로 인식하는 비율은 60.4%에 달했다. 이에 비해 온라인 동영상 플랫폼을 언론으로 인식한다는 응답은 27.0%, 메신저 서비스를 언론으로 인식한다는 응답은 21.4%로 상대적으로 낮게 나타났다(한국언론진흥재단, 2022).

정리하자면 한국의 이용자들은 상대적으로 온라인 포털을 뉴스의 주요한 창구이자 신뢰할 수 있는 언론으로 인식하고 있었다. 다른 소셜미디어 플랫폼 및 서비스에서는 뉴스 이용률과 뉴스매체로서의 신뢰도가 모두 낮게 나타났다.

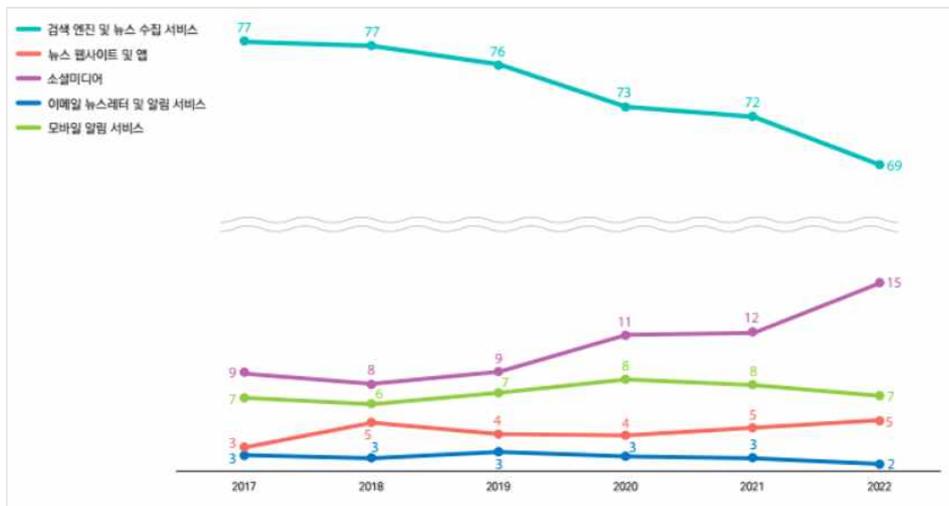
[그림 2-8] 매체별 언론 인식 여부 (단위: %)



*자료: 한국언론진흥재단(2022). 2022 언론 수용자 조사

로이터 저널리즘 연구소의 2022년 조사도 비슷한 결과를 보고한다. 이에 따르면 한국에서 디지털 뉴스를 이용하는 가장 주된 경로는 '검색엔진 및 뉴스 수집 서비스'로 나타났다. 해당 경로를 주된 경로로 이용한다는 응답 비율은 2017년 77%에서 2022년 69%까지 꾸준히 감소했지만 여전히 압도적으로 높은 비율을 차지한다. 이에 비해 소셜미디어의 비중은 2017년 9%에서 지속적으로 증가해 2022년에는 15%에 달했다. 그밖에 뉴스 웹사이트 및 앱에 직접 가서 이용한다는 응답은 5%에 불과했다(최진호·박영흠, 2022).

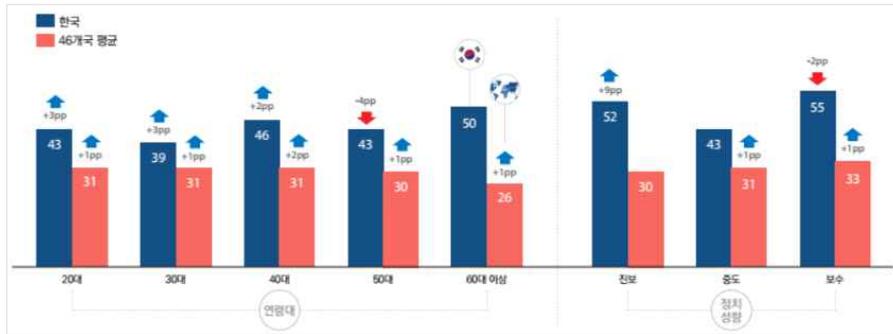
[그림 2-9] 디지털 뉴스 주이용 경로(단위: %)



*자료: 최진호·박영흠 (2022). 디지털 뉴스 리포트 2022 한국.

한 가지 특기할만한 점은 한국 이용자들이 소셜미디어 가운데 유튜브를 뉴스 매체로 이용하는 비율(43%)이 조사 대상 전체 국가(46개국)의 평균(31%)에 높게 나타났다는 점이다. 로이터 저널리즘 연구소의 조사결과에 따르면 모든 연령대에 걸쳐서 유튜브를 뉴스 매체로 이용하는 비율이 조사대상국 평균에 비해 높게 나타났다. 특히 60대 이상에서 유튜브를 뉴스 매체로 이용하는 비율(50%)은 조사 대상국 평균(26%)의 약 2배에 달했다(최진호·박영흠, 2022).

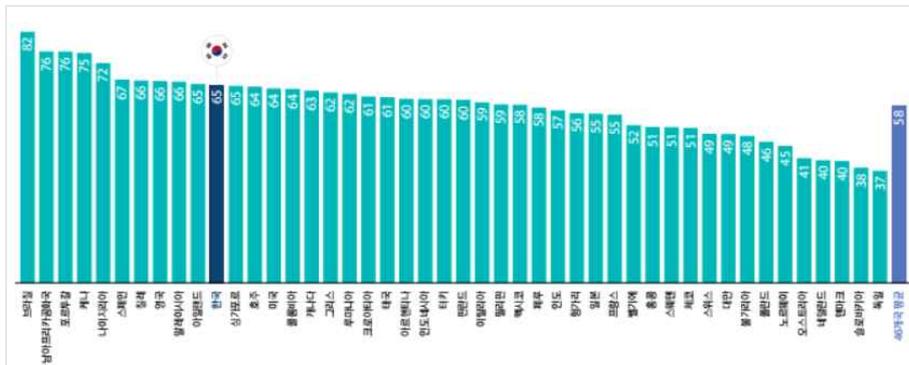
[그림 2-10] 연령대, 정치 성향별 유튜브 뉴스 이용률 (2022년, 단위: %)



*자료: 최진호·박영흠 (2022). 디지털 뉴스 리포트 2022 한국.

온라인을 통해 뉴스와 정보를 접하는 비율이 증가한 만큼 온라인을 통해 접하는 정보의 조작·허위에 대한 우려도 증가하고 있었다. 2021년 기준 한국의 응답자들이 허위·조작 정보에 대해 우려한다고 답한 비율은 2020년 대비 5%p 증가한 65%로 조사대상국 46개국 가운데 11번째로 높게 나타났다. 조사대상국의 전체 평균(56%)에 비해 9%p 높은 수치였다. 온라인 허위·조작 정보를 우려하지 않는다고 답한 비율은 7%에 불과했다(오세욱·박아란·최진호, 2021).

[그림 2-11] 인터넷 허위정보에 대한 우려 (2022년, 단위: %)



*자료: 오세욱·박아란·최진호(2021). 디지털 뉴스 리포트 2021 한국.

이는 단순히 우려에만 그치지 않았다. 실제로 온라인을 통해 허위정보를 접한 비율도 높게 나타났다. 한국언론진흥재단의 조사에 따르면 2021년 기준 소셜미디어 이용자의 77.2%는 소셜미디어를 통해 허위정보를 접촉한 경험이 있었다(한국언론진흥재단, 2021). 응답자들은 정보의 내용이 알고 있던 사실이나 상식에 맞지 않거나(67.9%) 출처가 명확하지 않은 정보(56.5%)를 허위정보로 판단했다.

[그림 2-12] 소셜미디어에서의 허위정보 접촉경험 (단위: %)



*출처: 한국언론진흥재단(2021). 2021 소셜미디어 이용자 조사

한국언론진흥재단의 조사에서 응답자들은 소셜미디어(4.1점), 인터넷 포털(3.91점), 인터넷 신문(3.79점)이 허위정보를 확산시키는 통로 역할을 한다고 평가했다. 전통적인 매체인 TV(3.39점), 라디오(3.08점), 종이신문(2.98점)에 대한 신뢰도는 상대적으로 높게 나타났다. 특히 인터넷 포털을 통해 허위정보가 매우 확산한다고 답한 응답자(24.9%)는 TV(12.7%)의 약 2배, 소셜미디어를 통해 허위정보가 매우 확산하다고 답한 응답자(37.6%)는 TV의 약 3배에 달했다.

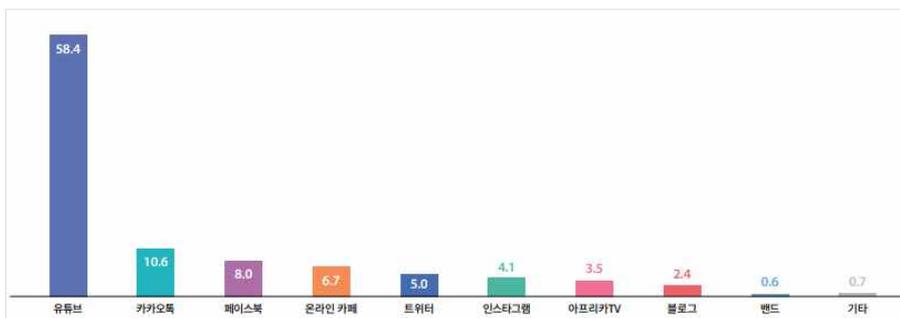
[그림 2-13] 매체별 허위조작정보 확산에 대한 평가



*자료: 한국언론진흥재단(2021). 2021 소셜미디어 이용자 조사

소셜미디어 가운데서는 특히 온라인 동영상 플랫폼이 뉴스와 정보를 검색하는 주요한 매체인 동시에, 허위정보가 확산되는 주된 채널로 평가받았다. 유튜브를 주된 허위정보 확산 채널로 평가한 응답자(58.4%)는 두 번째 순위를 차지한 카카오톡(10.6%)의 약 5.6배에 달한다.

[그림 2-14] 매체별 허위조작정보 확산정도



*자료: 한국언론진흥재단(2021). 2021 소셜미디어 이용자 조사

2. 선거와 관련한 허위조작정보

허위조작정보는 이용자들로 하여금 정보의 진위여부에 혼란을 겪게 만들고, 잘못된 선택을 하도록 만듦으로써 다양한 형태의 사회적 비용을 유발한다. 특히 선거 과정에서 유통되는 허위조작정보는 언론보도를 신뢰하지 못하도록 만들며 정치에 대한 불신을 불리일 키우고 유권자의 정당한 민주적 선택을 방해할 수 있다.

한국언론진흥재단이 2022년 대선을 앞두고 2021년 11월에 유권자를 대상으로 실시한 조사 결과에 따르면 유권자들은 인터넷포털(72.0%)을 선거와 관련된 정보를 자주 접하는 채널로 꼽았다. 텔레비전(59.2%), 구글·네이버 등 검색 사이트(47.5%), 유튜브(35.9%), 언론사 인터넷 홈페이지(23.5%), 카카오톡(17.6%)이 뒤를 이었다(김선호, 2022.2.9.). 선거와 관련된 정보를 접하고 이용하는 과정에서도 인터넷과 소셜미디어가 중요한 정보 유통 창구로 자리 잡았음을 확인할 수 있는 결과다.

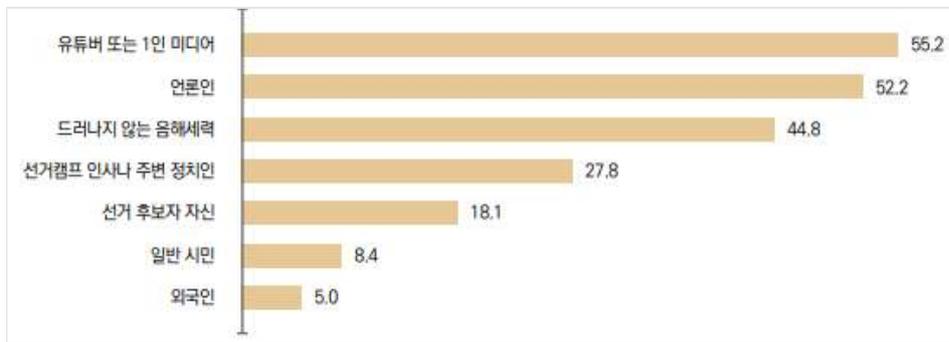
[그림 2-15] 선거와 관련된 정보를 자주 접하는 매체



*자료: 김선호(2022.2.9.). 유권자가 바라보는 선거보도

같은 조사에서 응답자의 45.7%는 “최근 한 달 동안 대선 관련 허위정보나 가짜뉴스를 접해본 적”이 있다고 응답했다. 이러한 허위정보의 유포와 관련해 ‘책임이 매우 큰’ 주체를 묻은 결과에 따르면 “유튜버 또는 1인 미디어”를 꼽은 응답자가 55.2%에 달했다. 하지만 언론인(52.2%)의 책임이 크다는 인식도 이에 못지않았다. 그밖에 드러나지 않은 음해세력(44.8%), 선거캠프 인사나 주변 정치인(27.8%), 선거 후보자 자신(18.1%)의 책임이 크다는 인식도 적지 않은 것으로 나타났다.

[그림 2-16] 허위조작정보 및 가짜뉴스 관련 주체별 “책임 매우 크다” 응답 비율 (단위: %)



*자료: 김선호(2022.2.9.). 유권자가 바라보는 선거보도

3. 코로나19와 관련한 허위조작정보

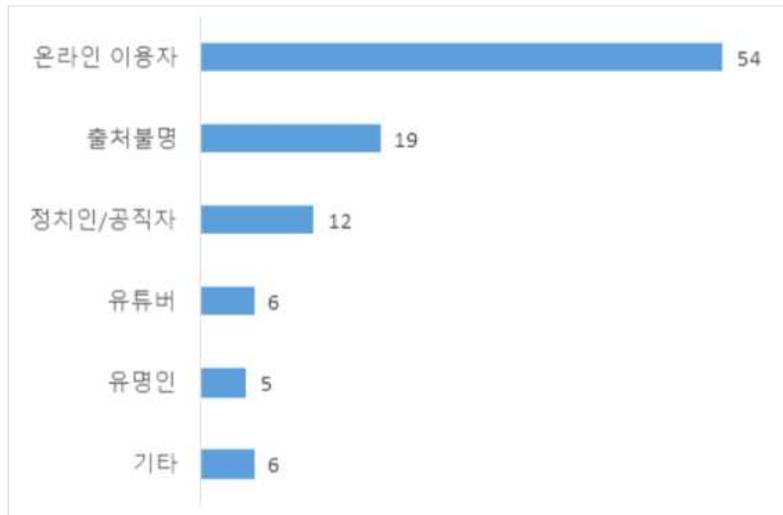
앞서 살펴본 것처럼 허위정보의 유통은 코로나19와 같은 감염병 확산 시기에 더욱 문제가 될 수 있다. 잘못된 정보는 곧 생명, 보건 안전, 공공 보건 시스템에 대한 위협으로 이어질 수 있기 때문이다(성욱제·정은진, 2020.5.7.).

보건복지부와 문화체육관광부가 2021년 4월 실시한 조사결과에 따르면 응답자의 69.4%는 코로나19 백신 접종과 관련된 허위조작정보 문제가 심각하다고 인식하고 있었다(보건복지부·문화체육관광부, 2021.5.5.). 또한 서울대 연구팀에 의하면 코로나19와 관련해 나중에 허위로 판명난 오정보(misinformation)를 자주 접한다는 응답도 25.9%에 달했다. 보통이

다는 49.0%, 전혀 접하지 않거나 거의 접하지 않는다는 응답은 25.2%였다(서울대학교, 2020.9.8.)

국민은 코로나19와 관련된 허위조작정보를 대체로 인터넷 플랫폼이나 소셜미디어 등의 온라인 창구를 통해 접한 것으로 나타났다. 국내 연구진이 2020년 한 해 동안 SNU 팩트체크센터에 올라온 검증자료 중 사실이 아닌 것으로 드러난 208건의 정보를 분석한 결과에 따르면 허위조작정보의 출처는 온라인 이용자(54%)인 경우가 가장 많았다. 그 다음은 출처 불명(19%)이었고, 정치인이나 공직자인 경우도 12%에 달했다(정영주·홍종윤·박유진, 2021).

[그림 2-17] 허위조작정보 출처 (단위: %)



*자료: 정영주·홍종윤·박유진 (2020). 사회적 법익을 침해하는 허위조작정보 대응을 위한 시론적 논의

코로나19와 관련해 팩트체크의 대상이 된 허위조작정보를 발견하거나 확인한 장소를 분석한 결과에 따르면 온라인 커뮤니티·포털(35%), 소셜미디어 서비스(27%), 언론보도(25%), 동영상 공유 플랫폼(8%), 개인 메신저(3%) 등이 주된 유통경로로 나타났다. 다만 이와 같

은 결과 해석에는 유의할 필요가 있다. 상대적으로 개인 메신저 서비스의 경우 언론사들이 팩트체크 대상으로 선정하기 쉽지 않기 때문이다. 때문에 실제 개인 메신저 서비스를 통해 허위조작정보가 유통될 가능성은 조사결과보다 한층 높다고 추정할 수 있다.

[그림 2-18] 허위조작정보 유통경로 (단위: %)



*자료: 정영주·홍종운·박유진 (2020). 사회적 법익을 침해하는 허위조작정보 대응을 위한 시론적 논의

코로나19 관련 허위조작정보의 주된 내용을 분석한 결과에 따르면 가장 비중이 높은 정보 유형은 백신 등 의학정보(48%)였다. 그 다음으로는 정부와 보건 당국의 대응(33%)이 뒤따랐다(정영주·홍종운·박유진, 2020). 이와 유사한 결과는 다른 조사를 통해서도 확인할 수 있다. 한국인터넷자율정책기구(KISO)가 2021년 4월 28일부터 10월까지 네이버·카카오 등 주요 회원사의 플랫폼에서 허위조작정보로 삭제 등 조치를 취한 정보 1,538건에서도 백신 관련 정보는 78.7%로 가장 높은 비중을 차지했다(KISO, 2021.11.14.).

한편, 로이터저널리즘연구소가 세계 각국의 뉴스 이용자들을 대상으로 조사한 결과에 따르면, 한국은 비교대상 6개국 가운데 코로나19와 관련된 허위조작정보를 식별할 수 있는 능력이 가장 뒤쳐진 것으로 나타났다. 코로나19와 관련된 5가지 질문 가운데 3개 이상

정답을 맞힌 비율은 영국(86%), 독일(76%), 아르헨티나(73%), 스페인(72%), 미국(65%), 한국(58%) 순서였다(박아란이소은, 2020).

[그림 2-19] 코로나19 관련 정보에 대한 참/거짓 응답률 (6개국, 단위: %)

	한국	영국	미국	독일	스페인	아르헨티나
코로나19는 날씨가 더운 지역에서 전파될 수 있다						
참(정답)	48%	60%	53%	42%	48%	54%
거짓	29%	17%	17%	30%	24%	32%
모르겠음	26%	22%	30%	28%	28%	15%
마늘을 먹으면 코로나19 감염을 막을 수 있다						
참	7%	2%	8%	6%	4%	7%
거짓(정답)	66%	87%	66%	79%	79%	72%
모르겠음	27%	11%	26%	15%	17%	21%
노인들이 코로나19를 더 심하게 앓을 수 있다						
참(정답)	88%	92%	90%	90%	94%	95%
거짓	7%	5%	5%	5%	3%	3%
모르겠음	7%	3%	6%	5%	2%	2%
항생제는 코로나19 치료에 효과적이다						
참	17%	3%	14%	7%	12%	10%
거짓(정답)	36%	79%	55%	68%	61%	64%
모르겠음	48%	18%	31%	25%	27%	25%
코로나19는 실험실에서 만들어졌다						
참	21%	14%	23%	15%	29%	28%
거짓(정답)	39%	58%	45%	55%	33%	31%
모르겠음	40%	28%	32%	29%	38%	41%
합계						
3개 이상 정답 비율	58%	86%	65%	76%	72%	73%
평균 정답수	2.72	3.77	3.08	3.34	3.16	3.16
표준편차	1.27	1.19	1.41	1.34	1.19	1.15

*자료: 박아란이소은(2020). 디지털 뉴스 리포트 한국 2020.

이와 같은 결과가 전적으로 허위조작정보 때문이라고 단언할 수는 없다. 다만 허위조작 정보의 유통과 확산이 실제 국민들이 정확한 정보를 인식하는데 영향을 미쳤을 개연성은 크다. 한국의 질병관리청도 코로나19 유행규모와 허위조작정보 및 가짜뉴스의 유통량은 비례하는 경향을 보인다고 밝히고 있다(경향신문, 2022.2.9.). 이는 허위조작정보에 대한 적시적절한 대응방안의 마련이 중요함을 강조한다.

제3절 소결

허위조작정보가 디지털 미디어 환경에서 전례 없이 빠르고 광범위하게 확산되면서 온라인 허위조작정보 유통은 전 세계적으로 심각한 문제가 됐다. 허위조작정보는 선거 과정에 영향을 미침으로써 시민의 정치적 의사결정과 선택을 방해하고 정치에 대한 신뢰를 저해한다. 또 코로나19 팬데믹 기간 동안의 인포데믹 사례에서 확인할 수 있듯이 사안에 따라서는 시민의 건강과 생명에 위협을 가하고 커다란 사회적 혼란을 야기하기도 한다.

이 장에서 살펴본 것처럼 인터넷 포털과 검색엔진, 소셜미디어는 정보의 매개자로서 주요한 역할을 수행하는 만큼 허위조작정보의 주요한 통로가 되고 있다. 그 결과 주요 플랫폼이 허위조작정보의 확산을 막기 위해 일정한 역할을 해야 한다는 인식도 높아지고 있다. 입소스와 유네스코의 조사에 따르면 세계 각국의 시민들은 선거 기간 동안 정부와 규제당국이 소셜미디어 플랫폼에 신뢰 및 안전조치를 요구할 수 있다고 인식하고 있었다(Ipsos & Unesco, 2023.9). 해외 학계에서는 플랫폼이 준 공적 기능(quasi-municipal function)을 수행한다는 주장이 제기되고 있기도 하다(Klonick, 2018; 문의빈, 2023에서 재인용). 이는 플랫폼에 책임과 역할을 부여하는 한편 공공과 민간의 협력을 통한 일종의 공동규제(co-regulation)를 요구하는 것이기도 하다는 점에서 주목할 필요가 있다.

이러한 인식은 다음 장에서 살펴볼 허위조작정보에 대한 세계 각국의 입법례에서도 나타난다. 프랑스를 예외로 한다면 미국, EU, 영국, 독일, 일본 등 세계 주요국은 허위조작정보의 유포를 직접적으로 규제하지 않는다. 대신에 플랫폼 사업자에게 자율규제의 책임을 부여함으로써 공동규제를 위한 거버넌스를 구축하거나 이용자의 미디어 리터러시를 강화하는 등 방안에 초점을 맞추고 있다. 자세한 내용은 이어지는 장에서 다루도록 한다.

제3장 해외 허위조작정보 관련 법적규제 사례

제1절 미국

1. 개관

미국에서 허위조작정보에 관한 논란은 2016년 대통령 선거 과정에서 본격적으로 시작됐다. 하지만 허위조작정보라는 개념보다는 가짜뉴스의 개념이 정치·사회적으로 더 크게 문제가 되고 있다. 수정헌법 제1조에서 표현의 자유를 강조하고 있기 때문에 규제 논의와 연관된 허위조작정보보다는, 선거과정에서 소셜미디어를 통하여 전파되는 정보인 가짜뉴스(fake news)가 더 주목받고 있는 것이다(전훈 외, 2019).

미국에서는 명예훼손에 해당하는 허위사실 주장 또는 위증 등에 이르지 않는 언설은 수정헌법 제1조에서 보호되는 표현의 자유에 해당한다. 연방대법원은 표현의 자유를 제한하는 입법에 대해서는 “명백·현존하는 위험의 원칙”(clear and present danger test)을 적용하고 있다. 이에 따르면 특정 표현을 규제하기 위해서는 표현의 자유와 침해되는 해악 간에 명백한 인과관계가 존재해야 하고, 현존하는 해악의 발생을 방지하지 않으면 안 될 시간적 급박성이 존재해야 한다. 연방통신위원회(Federal Communications Commission, FCC)는 허위조작정보 판단 기준에 대해 (i) 중대한 공익의 실질적인 침해(substantial public harm), (ii) 공익침해의 예견가능성, (iii) 허위성 인지가능성을 제시한다. 또 공익 침해의 발생가능성에 대해서는 “대중의 보건 및 안전, 공적 업무수행, 대중의 보건 및 안전을 위한 업무수행에 실질적인 해를 끼치는 경우”로 정의한다(전훈 외, 2019).

결과적으로 미국은 허위조작정보라 할지라도 표현의 자유 보호대상이 되므로 법령을 통해 콘텐츠의 내용을 규제하는 것에 대해서는 부정적인 입장이다. 허위조작정보를 생성한 자를 일반적으로 규제하는 법률이 없으며 「통신품위법(Communication Decency Act)」을 통해 일정한 조치를 한 인터넷서비스사업자의 민사 상 책임을 면제시키는 등 표현의 자유를 중시하는 태도를 취한다(박용숙 외, 2022). 때문에 허위조작정보 혹은 가짜뉴스에 관한

대응은 정부가 규제수단을 직접 활용하는 것보다는 의회 청문회, 또는 동의의결을 통한 합의 등 수단을 통해 주요 플랫폼 사업자의 자율규제를 요구하는 방향으로 전개되고 있다(전훈 외, 2019).

하지만 온라인을 통해 유포되는 허위조작정보의 문제가 심각해지면서 플랫폼에 역할과 책임을 부여하려는 목소리도 높아지고 있는 실정이다. (i) 플랫폼의 책임에 대한 입법적인 규제 시도의 대표적인 사례로는 제3자가 제공한 정보에 대해 플랫폼에 대한 광범위한 면책을 규정한 「통신망법」 제 230조를 둘러싸고 벌어지는 개정 시도를 들 수 있다. 이와 같은 논의는 플랫폼의 면책범위를 축소하고 보다 책임을 부여하는 방안을 모색한다(문의 빈, 2023). (ii) 플랫폼 사업자와의 합의를 통한 간접적인 규제 사례로는 연방통상위원회(Federal Trade Commission, FTC)와 페이스북이 체결한 개인정보 보호에 관한 동의의결을 들 수 있다(전훈 외, 2019).

플랫폼과 관련된 작·간접적인 규제 외에도 미국에서는 허위조작정보의 대응과 관련된 다양한 정책적, 입법적 시도가 이루어지고 있다. (iii) 바이든 행정부는 2023년 ‘안전하게 보호되고 신뢰할 수 있는 AI 개발·이용에 관한 행정명령(Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence)’에 서명했는데, 이는 AI 시스템에 의해 생성된 콘텐츠에 대한 라벨링 등 디지털 콘텐츠의 진위와 식별을 확실히 하기 위한 내용을 담고 있다(김지현 외, 2023.12.). (iv) 연방법 차원의 규제는 아니지만 선거기간 동안 허위조작정보를 직접적으로 규제하려는 시도도 있다. 텍사스 주와 캘리포니아 주 선거법은 선거기간동안 후보자의 답페이지 유포를 선거법으로 금지한다(최진용, 2022.11.). (v) 한편, 허위조작정보에 대응하기 위해서는 규제입법 외에도 시민들이 정보의 생산과 유포 과정을 이해하고 온라인 콘텐츠에 대한 비판적인 분석 역량을 갖는 것도 중요하다. 뉴저지 주는 입법을 통해 정보 리터러시 교육을 의무화함으로써 이를 유도하고 있다(김지현 외, 2023.12.). 아래에서는 이를 하나씩 살펴보도록 한다.

2. 플랫폼의 책임에 대한 입법적 규제 : 통신품위법 개정논의

(1) 통신품위법(Communication Decency Act)

통신품위법은 1996년에 제정된 법으로 온라인 상 폭력 및 음란물 등으로부터 미성년자를 보호하기 위한 목적을 갖고 있으며, 최초로 온라인 상 표현에 대한 내용규제를 담고 있다. 그런데 동법 제230조는 온라인 서비스 이용자들 간의 문제를 규율하고 있으나, 온라인 서비스 제공자에 대해서는 당해 내용이 폭력 및 음란물 등의 내용을 담고 있다더라도 법적 책임을 면하는 면책권한을 부여한다(전훈 외, 2019). 이는 인터넷서비스사업자에 발행인(publisher)이나 화자(speaker)의 책임을 묻지 않겠다고 규정하고 있는 제230조(c)(1)항과, 인터넷서비스사업자가 음란물이나 폭력적 콘텐츠 등 문제될만한 콘텐츠에 대해 자발적으로 신의성실의 원칙에 따라 접근을 제한하는 경우에는 면책이 된다고 규정하고 있는 제230조(c)(2)항으로 구성되어 있다(문의빈, 2023).

<표 3-1> 통신품위법 230조(c)의 내용

주요 내용
(c) “선한 사마리아인”의 저속한 소재 차단 및 감시 행위에 대한 보호
(1) 발행인 혹은 화자 취급 어떠한 대화형 컴퓨터 서비스의 제공자 및 이용자도 제3의 정보제공자가 제공한 정보에 대한 발행인 또는 화자로 취급되지 아니한다.
(2) 민사책임 어떠한 대화형 컴퓨터 서비스의 제공자 및 이용자도 다음과 같은 행위에 대하여 책임지지 아니한다.
(A) 헌법상 보호받는 소재인지 여부와 무관하게 제공자 혹은 이용자 판단 하에 외설적이거나, 선정적이거나, 음란하거나, 추잡하거나, 과도한 폭력성을 내포하거나, 성희롱이거나, 그밖에 불쾌감을 주는 소재의 접근 또는 이용가능성을 제한하기 위하여 선의를 가지고 자발적으로 취하는 조치 ; 또는
(B) 정보 콘텐츠 제공자 혹은 타 행위자가 (1)호에 포함된 소재의 접근을 기술적으로 제한할 방법을 제공하거나, 가능하게 하는 행위

*자료: 박용숙·강명원·고수윤 (2022). 유튜브 등 미디어플랫폼상 인격권 보호를 위한 해외법 제연구

이는 통신품위법이 제정되기 전인 1990년대 중반, 미국 법원의 일련의 판결을 통해 인터넷에서 게시물을 강력하게 관리·감독하는 사업자는 명예훼손적 게시물에 대한 법적 책임을 지게 되는 반면, 게시물을 방치하는 사업자는 면책되는 모순적인 상황이 발생했기 때문이다. 이에 인터넷서비스사업자들이 음란물과 같은 게시물을 엄격하게 관리·감독해도 법적인 책임을 지지 않도록 제3자가 작성한 게시물에 대해 인터넷서비스사업자가 발행인으로 취급되지 않도록 하는 조항이 통신품위법에 신설된 것이다. 이와 같은 법조항은 플랫폼에 대한 폭넓은 면책을 통해 미국에서 인터넷 발달에 기여했다는 평가를 받는다. 하지만 다른 한편에서는 지나치게 광범위한 면책을 허용하면서 온라인상에서의 명예훼손, 음란물 유통 및 가짜뉴스 확산을 야기했다는 비판도 받는다(문의빈, 2023).

이에 2018년에는 통신품위법 제230조가 일부 개정됐다. 이를 통해 성매매 목적의 인신매매(sex trafficking)에 한하여 온라인 서비스 제공자의 법적 책임이 규정됐다. 인신매매 등의 내용이 소셜미디어 등 온라인에 게재되는 경우, 서비스 제공자는 즉각 조치를 취해야 하며, 고의로 온라인을 통하여 성매매 목적의 인신매매가 이루어지는 경우 이를 사전에 방지해야 한다(전훈 외, 2019).

(2) 악성 알고리즘 방지 법안(Justice Against Malicious Algorithms Act)

최근에는 인터넷서비스사업자의 알고리즘에 의한 추천이 제3자가 작성한 콘텐츠를 단순 매개한 것인지 여부에 대한 문제가 제기되고 있다. 알고리즘에 의한 추천이 인터넷서비스사업자의 발언(speech)에 해당하고, 따라서 인터넷서비스사업자의 역할을 정보매개자가 아니라 발행인(publisher)으로 볼 수도 있다는 것이다. 이와 관련해 미국 연방 하원은 2021년 「악성 알고리즘 방지법안」을 발의했다. 이는 통신품위법 제230조(c)(1)항에 따라 폭넓게 면책되던 플랫폼의 책임에 대하여, 플랫폼이 개인맞춤형 추천을 하는 경우에는 면책되지 않도록 하는 것이다(문의빈, 2023).

법안은 개인맞춤형 추천을 맞춤형 알고리즘을 사용해 다른 정보보다 특정 정보의 중요성을 실질적으로 향상시키는 것(the material enhancement, using a personalized algorithm, of the prominence of such information with respect to other information)으로 정의하고, 이를 통해 (i) 피해자가 육체적 혹은 정신적 상해를 입는 것을 플랫폼이 알았거나 알 수 있었던 경우, (ii) 부주의하게 개인맞춤형 추천을 한 경우 및 개인맞춤형 추천으로 심각한 육체적 혹은 정신적 상해를 입히는데 실질적으로 영향을 준 경우에는 통신품위법 제230조(c)(1)항에 따른 면책이 적용되지 않도록 규정했다. 또 이 법안의 적용대상은 직전 12개월 중 3개월 동안 월간 방문자 혹은 이용자가 500만 명 이상인 플랫폼으로 한정했다(문의빈, 2023).

한편 2021년에는 미국 연방 상·하원 모두에서 「세이프 테크법안(Safe Tech Act)」이 발의되기도 했다. 이는 통신품위법 제230조(c)(1)에서 “정보(any information)”를 “표현(any speech)”으로 수정하고, 인터넷서비스사업자 혹은 이용자가 (i) 표현을 이용하기 위해 돈을 지불했거나, (ii) 표현의 전부 또는 일부를 직접 생성하거나 생성을 위한 자금을 지원한 경우에는 면책되지 않도록 면책의 범위를 축소했다(문의빈, 2023).

(3) 필터버블 투명성 법안(Filter Bubble Transparency Act)

「필터 버블 투명성 법안」은 2021년 미국 연방 상·하원에서 모두 발의된 법안이다. 법안의 핵심은 온라인 플랫폼이 이용자가 명시적으로 제공하지 않은 이용자 데이터의 전체 또는 일부를 기반으로 해당 인터넷 플랫폼에서 이용자에게 정보가 제공되는 순서 또는 방식을 결정하는 알고리즘 순위 시스템, 즉 불투명 알고리즘(opaque algorithm)에 대해 (i) 플랫폼이 해당 데이터를 사용한다는 사실을 이용자에게 알리고, (ii) 개인 맞춤형 알고리즘의 이용을 원하지 않는 이용자를 위해 이용자가 명시적으로 제공하는 데이터만을 사용하는 대안의 마련을 요구하는 것이다(문의빈, 2023).

(4) 플랫폼 책임 및 소비자 투명성 법안(PACT Act)

「플랫폼 책임 및 소비자 투명성 법안(Platform Accountability and Consumer

Transparency Act)』은 2021년 미국 연방 상원에서 발의된 법안이다. 법안은 플랫폼이 콘텐츠를 차단하거나 추천하는 콘텐츠 모더레이션 정책이 온라인 정보생태계 형성과 사람들의 표현의 자유에 직접적인 영향을 준다는 점에서, 정부가 플랫폼의 콘텐츠 모더레이션 정책에 관여할 중대한 공익(compelling government interest)이 있다고 명시한다(문의빈, 2023).

법안은 플랫폼이 콘텐츠 모더레이션 정책을 공개하고, 어떤 콘텐츠가 이러한 정책의 대상이 되는지, 소비자의 이의제기 절차는 무엇인지 안내하도록 규정한다. 또 플랫폼이 정보제공자의 게시물을 삭제하거나 차단한 경우 해당 정보제공자에게 이를 알리고 이의제기 절차를 안내하도록 한다. 뿐만 아니라 플랫폼이 서비스에서 공유되는 불법적인 콘텐츠 또는 불법행위의 존재를 실제로 알았거나, 이에 대한 통지를 통해 해당 사실을 알게 되었음에도 이를 삭제하거나 불법 활동을 중단하지 않을 경우 통신망위법 230조(c)에 따른 면책을 적용받지 못하도록 규정한다. 아울러 플랫폼에 매년 2차례 콘텐츠 모더레이션 사례, 이의제기 내역, 이의제기가 받아들여져 콘텐츠가 복원된 비율 등의 정보가 담긴 투명성 보고서를 작성하도록 의무를 부여한다(문의빈, 2023).

3. 플랫폼사업자와의 합의를 통한 규제 : FTC와 페이스북의 동의의결

앞서 살펴본 것처럼 플랫폼 기업의 알고리즘을 적용한 추천, 그리고 플랫폼 기업의 책임과 투명성 등은 온라인상의 정보 생태계를 구축하는데 있어 핵심적인 이슈가 되고 있다. 예를 들어 알고리즘을 통해 개인별로 특화된 광고를 제공할 목적으로 이용자에게 뉴스피드를 제공하는 것은, 개인의 관심사를 중심으로 정제되지 않은 정보를 제공한다는 점에서 이용자를 왜곡되거나 과도된 정보에 노출시킬 가능성이 있다. 뿐만 아니라, 이용자들이 검색어를 반복해서 올리는 등 이를 역이용하면 허위조작정보가 전파되도록 하는데 영향을 미칠 수 있다(전훈 외, 2019).

이와 관련하여 FTC는 페이스북의 이용자 개인정보 취급이 FTC법을 위반하는지에 대한 조사를 개시하고 2011년 이를 동의의결⁵⁾로 종결한 바 있다. 당시 페이스북의 주된 혐의는

5) 동의의결제도는 반경쟁행위 등 경쟁당국의 조사 또는 심의 대상이 되는 행위에 대해 사

이용약관(terms and conditions)에서 페이스북이 이용자의 개인정보를 제3자에게 제공할 때 이용자의 동의를 받도록 명시했음에도 이를 위반했다는 것이다. 해당 동의의결을 통해 페이스북은 향후 20년 간 이용자의 개인정보 취급설정을 위반하여 개인정보를 제3자에게 제공하지 않기로 합의했으며, 페이스북이 동의명령의 합의내용을 위반할 경우 위반행위 건당 16,000달러의 벌금을 부과하기로 했다(공정거래위원회, 2018.4.2.).

그러나 2016년 대선과정에서 페이스북 이용자의 동의 없이도 정치적 성향을 추론할 수 있는 별도의 프로그램이 존재했으며, 이러한 어플리케이션을 통하여 개별 유권자의 성향에 맞는 선거광고가 제작되고 반복적으로 노출됐다는 것이 드러나면서 동의의결 위반 논란이 발생했다(전훈 외, 2019). 영국의 데이터 분석업체(Cambridge Analytica)가 약 5천 만 명의 미국 유권자 정보를 수집하고 유용했다는 것이다(공정거래위원회, 2018.4.2.).

이에 2018년 3월 FTC가 해당 사건에 대한 조사 개시를 공식적으로 발표했고, 2019년 FTC와 페이스북은 이용자의 개인정보와 플랫폼에 제공되는 정보를 보호할 포괄적인 틀을 새롭게 마련하기로 합의했다. 또한 페이스북은 50억 달러의 벌금을 납부하기로 했는데 이는 FTC가 테크기업에 부과한 벌금액수로는 최대 규모였다.

새로운 합의안은 페이스북의 투명성과 책임을 보다 강화했다. 페이스북은 개인정보 관리방식의 원활한 작동을 검증하기 위해 분기마다 인증을 받고, 문제가 발견될 경우 CEO가 문제 해결을 확인한 후 직접 서명하도록 규정했다. 또 페이스북 이사진으로 구성된 감독위원회를 구성해 분기별로 약속된 내용이 제대로 수행되는지 점검하도록 했으며, 독립적인 개인정보 보호 평가관이 위원회를 지원하도록 했다. 이 모든 노력은 FTC와 미 법무부가 계속해서 모니터링하며 이를 위해 페이스북은 개인정보보호를 위한 위원회와 FTC에 보고의무를 지게 됐다(페이스북, 2019.7.24.).

업자가 해당 행위를 스스로 시정할 수 있는 방안(반경쟁행위 해소, 소비자 피해에 대한 재발 방지, 피해보상 방안 등)을 마련해 경쟁당국에 동의의결을 신청하는 경우, 경쟁당국이 해당 시정방안을 검토해 타당성이 인정되면 해당 행위의 위법여부를 판단하지 않고 시정방안에 따라 사건을 종결시키는 제도를 가리킨다.

이는 기존의 일방적인 시정조치에서 벗어나 해당 시장을 가장 잘 이해하고 있는 사업자의 자진 시정방안을 통해 법 위반혐의 사건을 신속하고 유연하게 해소함으로써 경쟁당국, 법 위반혐의가 있는 사업자, 피해사업자 또는 피해소비자 등 이해관계자 모두에게 긍정적으로 작용할 수 있다(송혜진·송민수, 2021).

[그림 3-1] FTC 합의에 담긴 페이스북의 사용자 개인정보 보호 방법



* 자료: 페이스북(2019.7.24.). 이용자 개인정보보호에 대한 새로운 기준을 제시한 연방거래위원회(FTC)와의 합의안

4. 바이든 행정부의 AI 규제 행정명령

AI 기술이 급속하게 발달하면서 AI가 주는 편익만큼 부작용에 대한 우려도 커지게 됐다. 특히 AI를 활용한 알고리즘의 적용 및 이미지 생성 등은 허위조작정보의 생성과 유포에도 영향을 미칠 개연성이 높다. 이와 관련해 바이든 행정정부는 2023년 10월 30일 AI를 규제하는 첫 행정명령인 「안전하게 보호되고 신뢰할 수 있는 AI 개발·이용에 관한 행정명령 (Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, EO 14110)」을 발표했다(김경숙·홍건식, 2023.11.).

행정명령은 국가 안보, 건강, 안전을 위협하는 AI 기술의 개발과 이용을 규제하기 위한 내용을 담고 있으며, 특히 제1조 목적에서 “AI를 무책임하게 사용한다면 사기, 차별, 편견 및 허위조작정보(disinformation)와 같은 사회적 피해를 악화시키고 근로자를 대체하며 그 권한을 박탈하고 경쟁을 억제하며 국가 안보에 위협을 초래할 수 있다” 고 밝힘으로써 허위조작정보 대응이 주요한 목적 중 하나임을 명시하고 있다.

<표 3-2> 「안전하게 보호되고 신뢰할 수 있는 AI 개발이용에 관한 행정명령」의 구성

구분	주요 내용
제1조	목적
제2조	정책 및 원칙
제3조	정의
제4조	AI 기술의 안전 및 보안의 보장
제5조	혁신 및 경쟁의 촉진
제6조	근로자에 대한 지원
제7조	형평성 및 시민권의 진전
제8조	소비자, 환자, 승객 및 학생에 대한 보호
제9조	개인정보 보호
제10조	연방정부의 AI 사용의 발전
제11조	해외에서 미국의 리더십 강화
제12조	이행
제13조	일반조항

*자료: Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 번역은 국회법률도서관 ‘외국법률번역DB’ 참조

행정명령은 AI 개발 기업의 안전성 평가 의무화, AI 도구의 안전성 표준 마련, 콘텐츠 인증 표준 수립과 개인정보 보호 등에 대한 내용을 담고 있다. 특히 허위조작정보와 관련해 눈여겨볼만한 부분은 AI 합성 콘텐츠(synthetic content)⁶⁾에 관한 부분이다. 실제 바이든

6) “ ‘합성 콘텐츠’란 AI를 포함한 알고리즘을 활용하여 크게 수정하거나 생성하는 이미지, 비디오, 오디오 클립, 텍스트 등의 정보를 말한다(The term “synthetic content” means information, such as images, videos, audio clips, and text, that has been

대통령은 행정명령에 서명하는 현장에서 딥페이크의 위험성을 언급하기도 했다.

이는 합성 콘텐츠로 인한 위험을 축소하기 위해 콘텐츠를 인증하고 출처를 추적하며, 합성 콘텐츠를 라벨링해 식별가능성을 높이도록 규정하는 한편, 연방 상무부장관이 관련 보고서를 작성하여 관계기관에 제출하고 이에 관한 가이드라인을 마련하도록 규정한다(김경숙홍건식, 2023.11.). 해당 내용을 명시한 4.5항의 전체 내용은 다음과 같다.

<표 3-3> AI 합성 콘텐츠 관련 내용

주요 내용

제4조. AI 기술의 안전 및 보안의 보장

4.5 합성 콘텐츠로 인한 위험의 축소

AI 시스템을 활용하여 생성된 합성 콘텐츠를 식별하고 라벨링하는 기능을 육성하고 연방정부나 이를 대신하여 생성된 합성 및 비합성 디지털 콘텐츠의 진위성과 출처를 확립하기 위하여 다음을 수행한다.

- (a) 상무부장관이 이 명령 시행일부터 240일 이내에 상무부장관이 적절하다고 판단하는 그 밖의 관련 기관장과 협의하여 행정관리예산국장과 국가안보보좌관에게 다음과 같은 기존 표준, 도구, 방법 및 관행뿐만 아니라 과학적으로 뒷받침되는 추가 표준 및 기술의 잠재적 개발을 식별하는 보고서를 제출하는 것
 - (i) 콘텐츠를 인증하고 출처를 추적하는 것
 - (ii) 워터마킹 등을 사용하여 합성 콘텐츠에 라벨링을 하는 것
 - (iii) 합성콘텐츠를 발견하는 것
 - (iv) 생성형 AI가 아동 성학대 자료를 생성하거나 실제 개인에 대해 합의되지 아니한 은밀한 이미지(식별 가능한 개인의 신체나 신체부위에 대한 은밀한 디지털 묘사를 포함한다)의 생성을 방지하는 것
 - (v) 상기 목적상 사용되는 소프트웨어를 시험하는 것
 - (vi) 합성 콘텐츠를 감시하고 유지하는 것
- (b) 상무부장관이 이 조 제4.5항(a)에 따라 요구되고 이후 정기적으로 갱신되는 보고서를 제출한 후 180일 이내에 행정관리예산국장과 협력하여 디지털 콘텐츠 인증 및 합성 콘텐츠 발견 조치를 위한 기존 도구와 관행에 관한 지침을 개발

significantly modified or generated by algorithms, including by AI.”

하는 것. 지침에는 이 조 제4.5항(a)에 열거된 목적상 조치가 포함된다.

- (c) 행정관리예산국장이 이 조 제4.5항(b)에 따라 요구되고 그 이후 정기적으로 갱신되는 지침이 개발된 후 180일 이내에 국무장관, 국방부장관, 법무부장관, 상무부장관과 협의하고 국립표준기술원장, 국토안보부장관, 국가정보국장, 행정관리예산국장 및 적절하다고 판단하는 그 밖의 기관장을 통해 행위하여 미국 정부의 공식 디지털 콘텐츠의 무결성에 대한 대중의 신뢰를 강화할 목적상 해당 기관이 생산하거나 게시하는 콘텐츠에 라벨링을 하고 인증하기 위한 지침을 발행하는 것
- (d) 연방획득규제위원회가 적절하고 적용 가능한 법률에 합치되는 경우, 이 조 제 4.5항에 따라 수립된 지침을 고려하도록 연방조달규정의 개정을 고려하는 것

*자료: Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 번역은 국회법률도서관 ‘외국법률번역DB’ 참조

합성 콘텐츠와 관련된 규정은 러시아-우크라이나 전쟁과 최근 이스라엘-하마스 전쟁에서 나타난 것처럼 AI를 이용한 공격이나 허위조작정보 확산을 예방하기 위한 조치로 해석된다. 물론 행정명령은 행정부 내에서 효력이 발생하는 조치로 법적 구속력이 없기 때문에 실제 기업이 얼마나 행정명령을 이행할지는 미지수이다. 다만 구글, 메타, 오픈AI를 비롯한 주요 업체는 AI 기술로 작성된 콘텐츠에 새로운 워터마크를 표시하는 시스템 개발을 현재 진행 중에 있는 것으로 알려져 있다(김경숙홍건식, 2023.11.).

5. 각 주 선거법의 선거기간 딥페이크 규제

허위조작정보의 유통은 특히 선거기간에 문제가 될 수 있다. 연방법을 통한 규제 논의는 많지 않으나, 일부 주법에서는 선거운동에서의 딥페이크를 명시적으로 규제하고 있다. 텍사스 주 선거법(election code)은 ‘딥페이크 비디오(deep fake video)’를 규제하며, 캘리포니아 주 선거법은 기만적인 오디오·비주얼 미디어(deceptive audio or visual media)를 규제한다.

텍사스 주 선거법에서는 후보자에게 손해를 입히거나 선거결과에 영향을 줄 의도로 딥페이크 비디오를 제작하고 이를 공식선거일 30일 이내에 유포하는 행위를 규제한다. 동법상 딥페이크 비디오란 ‘기만할 의도를 갖고 정의된 것으로서 실제 사람이 현실에서는 발생하지 않은 행위를 묘사하는 것처럼 보이는 인공적인 비디오물을 의미한다(최진웅, 2022).

캘리포니아 주는 2019년부터 선거법을 통해 실제 악의(actual malice)를 갖고 유포되는 선거 후보자 관련 딥페이크를 규제하고 있다. 구체적으로는 개인이나 단체가 선거 60일 이내 후보자의 평판을 해할 의도로 또는 후보자에게 찬성 또는 반대하도록 투표자를 기만할 의도로 실질적으로 기만할 정도(materially deceptive)의 시청각 미디어를 실제적 악의를 가지고 유포하는 것을 규제한다. 여기서 실질적으로 기만할 정도의 시청각 미디어는 ‘후보자의 외양연설·행동에 관한 이미지나 시청각 미디어를 고의로 조작한 것으로 (i) 허위임에도 합리적인 사람에게 진짜로 보일 수 있고, (ii) 변경되지 않은 원래 오리지널 버전을 듣거나 본 경우와 근본적으로 상이한 이해나 인상을 갖도록 만든 것’ 을 의미한다. 이에 따라 법 위반 시 후보자가 해당 콘텐츠 배포자에 대하여 민사상 손해배상을 청구할 수 있다(김지현 외, 2023).

그러나 해당 콘텐츠가 조작되었음을 투명하게 공개한 경우, 그리고 언론이 선의로 뉴스 혹은 뉴스 인터뷰 등을 통해 ‘실질적으로 기만할 정도의 시청각 미디어’ 를 방송한 경우에도 시청자에게 분명하게 해당 미디어의 진위가 불분명하다고 밝힌 경우에는 이 규제가 적용되지 않는다(김지현 외, 2023). 이와 같은 ‘기만적인 시청각 미디어의 온라인상 유통에 있어서는 통신품위법 230조의 플랫폼의 권리의무책임 등을 부정하지 않는다.’ 즉, 플랫폼은 해당 딥페이크 정보 유통에 대해 직접적인 책임을 지지 않지만 자율적으로 해당 정보의 차단 등의 조치를 취할 수 있다. 또한 이러한 기만적인 오디오·비주얼 미디어를 활용한 풍자와 패러디 등도 동법의 적용대상에서 제외된다(최진웅, 2022).

6. 뉴저지 주의 정보 리터러시 교육 의무화

표현의 자유에 대한 규제가 어려운 미국의 경우 범죄에 해당하는 불법 콘텐츠가 아닌

한, 온라인 허위조작정보의 유포 자체를 연방법으로 규제하고 있지는 않다. 플랫폼이나 AI 업체의 책임을 강화하는 등 인터넷서비스사업자를 통해 간접적인 방식으로 규제가 이루어진다. 캘리포니아 주가 선거기간 동안의 딥페이크를 규제하고 있기는 하나, 그보다는 이용자들이 정보를 효과적으로 활용하고 비판적으로 분석할 수 있도록 도와주는 리터러시 교육 입법에 더 적극적이다. 뉴저지 주, 델라웨어 주, 일리노이 주 등에서는 미디어 리터러시 교육을 의무화했다(김지현 외, 2023).

그 중 뉴저지 주 2023년 입법을 통해 정보 리터러시(information literacy) 교육을 공립학교에서 의무화했다(N.J.S.A. 18A:7F-4.4). 이에 따라 뉴저지 주 교육위원회는 정보 리터러시 학습표준을 마련하고 각 교육구는 유치원~12학년 학생들의 커리큘럼에 정보 리터러시 교육을 포함해야 한다(김지현 외, 2023).

제 2 절 EU

1. 개관

온라인 허위조작정보의 확산이 심각한 사회문제로 부상하면서 2010년대 중반부터 EU는 허위조작정보에 대처하기 위한 다양한 방안을 모색하기 시작했다. 여기에는 러시아 등 외국으로부터의 허위조작정보 캠페인에 대처하기 위한 방안과 주요 온라인플랫폼의 자율규제를 강화하기 위한 방안이 있다.

2015년 유럽이사회(European Council)는 러시아발 허위조작정보 캠페인에 대응하기 위해 이스트 스트랫컴 태스크포스(East StratCom Task Force)를 구성했다. 태스크포스는 동부 인접국가의 언론자유를 촉진하고 러시아의 허위조작정보를 적발하는 것을 목적으로 한다(심나리, 2019). 한편 태스크포스는 유럽연합 내 허위조작정보(EU vs Disinfo)⁷⁾라는 사이트를 개설해 러시아발 허위조작정보에 대한 공중의 인식과 이해를 높이고 팩트체크 서비스를 제공하며 유럽 시민들이 이에 대한 저항력을 개발하도록 지원하고 있다.

7) <https://euvsdisinfo.eu/disinfo-review/>

태스크포스의 조직을 지위한 것은 유럽 대외협력청(European External Action Service, EEAS)의 EU 외교문제·안보정책 고위대표인 모게리니(Mogherini, F.)이다(이가연, 2018). 이를 통해 알 수 있듯이 유럽연합 대외협력청도 러시아발 허위조작정보에 대해 대응하는 주요한 기관 중 하나다. 특히 대외협력청 내 전략 의사소통 부문(Strategic Communication division)은 외국의 허위조작정보에 대응하기 위한 정책 및 수단을 개발하고, 관련 기술 및 절차를 강화하며, 시민사회의 저항력을 키우기 위해 인식제고와 교육을 지원하고, 주변국가 및 시민사회와의 협력 등을 주도한다(EEAS 홈페이지).

한편 유럽의회(European Parliament)는 ‘외국의 간섭에 대한 특별위원회(Special Committee on Foreign Interference, INGE)’⁸⁾를 두고 악의적인 외국의 간섭에 대한 대응책을 마련하고 있으며, 악의적인 목적으로 악용될 수 있는 EU 법률의 공백을 식별하고 있다. 또한 유럽의회의 정보조작방지팀은 허위조작정보를 모니터링하고 타 기관 및 시민사회와 협력하며 허위조작정보와 관련된 교육도 진행한다. 한편, 유럽의회는 코로나19 팬데믹 상황에서 인포데믹에 대처하고 EU 시민이 과학적인 정보를 얻도록 돕기 위해 유럽 과학미디어 허브(European Science-Media Hub, ESMH)를 개설하기도 했다(박용숙 외, 2022).

유럽집행위원회(European Commission)는 허위조작정보에 대한 자율규제 중심의 대응 방안을 마련하고 있다. 대표적으로는 2018년과 2022년 허위조작정보 대응 실천강령(code of practice against disinformation)을 제안했다. 또 2020년 12월에는 모든 이용자의 기본권이 보호되는 안전한 디지털 공간을 만드는 것을 목적으로 「디지털 서비스 법(Digital Service Act, DSA)」을 제안했으며 이는 2022년 7월 유럽의회에서 최종승인됐고 10월 유럽연합이사회에서 채택됐다. 동법에는 안전한 디지털 공간을 만들기 위한 플랫폼의 다양한 책임이 명시되어 있으며 특히 플랫폼의 규모에 따라 차별적인 규율을 적용하고 있는 것이 특징이다. 마지막으로 2021년 유럽집행위원회는 「AI법안(Artificial Unintelligent Act)」을 마련하고 이를 추진하고 있다. 동법안에서는 딥페이크 AI를 활용한 콘텐츠를 이용하기 위해서는 이를 이용자에게 밝혀야 하는 투명성 의무를 부여한다.

허위조작정보 대응 실천강령을 중심으로 하는 유럽집행위원회의 일련의 정책에 대해서

8) 해당 위원회는 2022년 3월에 2기가 구성됐다(New Special Committee on Foreign Interference, INGE2).

출처: <https://www.europarl.europa.eu/committees/en/inge/home/highlights>

는 다른 장에서 다룰 것이다. 아래에서는 디지털서비스법을 중심으로 EU의 제도를 살펴볼
 도록 하겠다.

2. 플랫폼의 책임에 대한 입법적 규제 : 디지털 서비스법

(1) 디지털서비스법의 구조와 특징

디지털 서비스법은 「유럽연합 기본헌장(The Charter of Fundamental Rights of the European Union)」에 명기된 권리가 보호되는 안전하고 예측가능하며 신뢰할 수 있는 온라인 환경의 확보라는 입법 목적을 갖고 있다. 따라서 시민의 권리와 민주주의에 대한 영향을 규제하기 위한 사업자 규정과 그에 따른 규제 수단을 포함하고 있다(박찬경, 2022). 디지털 서비스법은 온라인에서 이용자의 기본권에 대한 포괄적 보호뿐만 아니라 불법콘텐츠의 신속한 제거를 위한 절차를 규정함으로써 이용자를 재화, 서비스 또는 콘텐츠에 연결하는 모든 디지털서비스에 대한 EU 회원국 전체의 의무를 구속력 있게 규율하려는 목적을 갖고 있다. 동법은 다음과 같이 구성되어 있다.

<표 3-4> 디지털 서비스 법의 구조

구분	주요 내용
제1장	규칙의 대상범위, 주요 용어의 정의 등 일반조항
제2장	중개서비스 제공자의 책임 면제에 관한 조항
제3장	투명하고 안전한 온라인 환경을 위한 상당한 주의 의무 · 제1절 중개서비스 제공자의 의무 · 제2절 호스팅 서비스 제공자의 의무 · 제3절 온라인 플랫폼 제공자의 의무 · 제4절 소비자와 거래자의 원거리 계약 체결을 허용하는 온라인 플랫폼 제공자의 의무 · 제5절 대규모 온라인 플랫폼 제공자의 의무

	· 제6절 그 외 의무
제4장	<ul style="list-style-type: none"> 규칙의 시행, 협력, 처벌 및 집행 · 제1절 디지털 서비스 코디네이터 등 규제당국에 관한 조항 · 제2절 권한, 조사 및 일관성 메커니즘 · 제3절 중개 서비스 제공업체를 감독하는 독립적인 디지털 서비스 코디네이터 자문그룹인 ‘유럽 디지털 서비스 위원회’ 설립 · 제4절 대규모 온라인 플랫폼의 감독·조사·집행·모니터링 · 제5절 일반적인 집행조항 · 제6절 위임입법
제5장	규칙의 평가·발효·적용 등

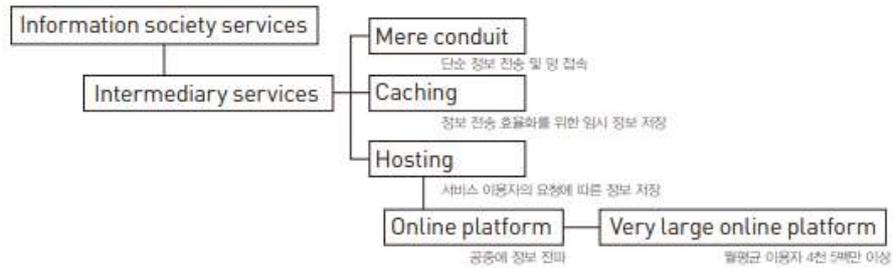
* 자료: 디지털 서비스법

동법은 규칙의 시행 및 집행을 위해 EU 집행위원회, 회원국의 규제당국 및 디지털서비스코디네이터, 유럽 위원회에 역할을 분담하고 있다. 집행위원회(European Commission)는 규칙 시행을 위한 절차의 전반적인 업무를 담당한다. 규제당국(competent authorities)은 규칙의 적용과 시행을 책임지며 디지털서비스코디네이터를 지정한다. 디지털서비스코디네이터(Digital Services Coordinator)는 조사권한·집행력 등을 보유하고 규칙의 적용·시행과 관련된 모든 책임을 지며, 회원국 간 일관된 기준 수립을 위해 타 회원국 조정기관과 협력한다. 유럽 디지털 서비스 위원회(European Board for Digital Services)는 조정기관의 자문기구로서 대규모 온라인 플랫폼 감독을 위해 집행위원회 및 조정기관을 지원한다.

동법의 제3장은 중개서비스 제공자 및 온라인 플랫폼에 대해 주의의무를 포함한 각종 의무를 부과한다. 그 목적은 온라인 환경의 안정성과 투명성 확립에 있다. 여기서 눈여겨볼만한 점은 중개서비스의 유형과 특성을 고려하여 의무의 내용을 달리하고 있다는 점이다. 제1절은 중개서비스에 일반적으로 적용되는 규정, 제2절은 온라인 플랫폼을 포함한 호스팅 서비스 제공자에게 추가적으로 적용되는 규정, 제3절은 온라인 플랫폼에 추가적으로 적용되는 규정(다만, 소규모 플랫폼에는 적용되지 않음), 제4절은 대규모 온라인 플랫폼에 추가적으로 적용되는 규정을 담고 있다. 다시 말해 서비스의 유형과 특성에 따라 의무가 차등적으로 적용되며, 이는 누적적으로 강화된다고 할 수 있다. 디지털 서비스법은 이와

같은 사업자의 유형의 구분 방식을 다음과 같이 설명한다.

[그림 3-2] 디지털 서비스법 적용대상 사업자 분류



*자료 : 박찬경 (2022). 유럽연합 디지털 서비스법: 규제나, 자율이냐를 넘어서

각각의 사업자는 다음과 같이 정의된다.

<표 3-5> 디지털 서비스법 적용대상 사업자 분류

구분	주요 내용
중개서비스	네트워크 인프라를 제공하는 서비스 : 인터넷접속서비스, 도메인이름 등록 서비스 등
호스팅 서비스	클라우드, 웹호스팅 서비스 등
온라인 플랫폼	판매자와 소비자를 연결하는 서비스 : 라인 마켓 플레이스, 앱스토어, 공유경제 플랫폼, 소셜미 디어 플랫폼 등
대규모 온라인 플랫폼	매월 EU 인구의 10%(4,500만 명) 이상이 이용하는 온라인 플랫폼 및 검색엔진(유럽집행위원회에서 제정)

* 자료: 디지털 서비스법

대규모 온라인 플랫폼 및 검색엔진은 유럽집행위원회가 주기적으로 지정해서 발표한다. 2023년 4월 유럽집행위원회는 17개의 대규모 온라인플랫폼과 2개의 대형 검색엔진 서비스

등 총 19개 서비스를 발표했다. 이 서비스들은 2023년 8월 25일부터 디지털 서비스법의 규정을 준수해야 한다(최민재·진민정·이현우, 2023.5.31.).

<표 3-6> 디지털 서비스법에 따른 대규모 온라인 플랫폼 기업 목록

기업명			
Alibaba AliExpress	Google Play	Pinterest	Youtube
Amazon Store	Google Maps	Snapchat	Zalando
Apple AppStore	Google Shopping	Tiktok	Bing
Booking.com	Instagram	X(Twitter)	Google Search
Facebook	Kinkedin	Wikipedia	

* 자료 : 최민재·진민정·이현우 (2023.5.31.). 유럽의 가짜뉴스 대응정책

제공되는 서비스의 특성과 규모에 따라 해당 규정에 의하여 각 유형의 사업자에게 부과되는 의무는 중첩적으로 적용된다. 공통적으로 모든 사업자는 투명성 보고 의무, 서비스 이용약관의 기본권 준수, 국가 당국과의 협조, 법정 대리인 지정의 의무를 진다. 그리고 규모가 큰 사업자는 더 큰 사회적 책무를 갖는다. 사업자에게 차등적으로 적용되는 의무의 구체적인 내용은 다음과 같다.

<표 3-7> 디지털 서비스법 적용대상 사업자 유형별 의무

구분	중개 서비스	호스팅 서비스	온라인 플랫폼	대규모 온라인 플랫폼
투명성 보고	○	○	○	○
기본권을 충분히 고려한 서비스 이용약관 제공	○	○	○	○
명령에 따른 회원국 규제당국과의 협력	○	○	○	○
컨택트 포인트,	○	○	○	○

필요시 법적 대리인				
이용자에 대한 통지 및 정보제공의무		○	○	○
불만대응 및 보상 메커니즘, 법정 외 분쟁해결			○	○
신뢰할 수 있는 신고자			○	○
잘못된 통지에 대한 절차 및 재통지			○	○
써드파티 공급업체의 자격 확인 “사업자 추적”			○	○
이용자 대상 온라인 광고의 투명성			○	○
범죄행위 신고			○	○
리스크 관리 의무 및 규정준수 책임자				○
외부의 리스크 감사 및 공공책임				○
추천시스템의 투명성 및 정보접근에 대한 이용자 선택권				○
규제당국 및 연구자와의 데이터 공유				○
행동강령				○
위기대응 협력				○

* 자료 : EC 홈페이지

(2) 디지털서비스법과 허위조작정보

디지털 서비스법은 허위조작정보에 대한 법적 정의를 포함하고 있지는 않다. 중개 서비스 제공자의 대응 의무는 불법 콘텐츠(illegal content)에만 적용된다. 디지털 서비스법에 따른 불법콘텐츠는 “EU법 또는 EU법을 준수하는 회원국의 법률에 부합하지 않는 모든

정보” 를 의미한다.⁹⁾

그럼에도 불구하고 디지털 서비스법은 제안이유(recital)에서 “본 규정은 안전하고 예측 가능하며 신뢰할 수 있는 온라인 환경을 보장하고, 온라인 불법콘텐츠의 유포 및 허위조작정보 또는 기타 콘텐츠의 유포로 발생할 수 있는 사회적 위험을 해결” (recital 9)하는 것이 목적이라고 밝히면서, 허위조작정보에 대한 대응이 주요한 입법 목적임을 명시한다.

이러한 맥락에서 온라인 플랫폼으로 하여금 불법 콘텐츠에 대응하고 투명성을 확보하도록 요구하는 것은 허위조작정보의 유포로부터 이용자를 보호하는 데에도 기여할 수 있다. 이러한 의무를 간략히 요약하자면 다음과 같다.

<표 3-8> 디지털 서비스법이 온라인 플랫폼에 부여하는 의무

의무	주요 내용
불법 콘텐츠에 대한 대응	<ul style="list-style-type: none"> • 온라인 플랫폼은 사용자에게 불법 콘텐츠를 쉽게 신고할 수 있는 도구를 제공해야 하며, 신고가 접수되면 불법 콘텐츠에 대한 액세스를 신속하게 제거하거나 차단해야 함. 이 맥락에서 온라인 플랫폼은 ‘신뢰할 수 있는 신고자’와 협력해야 하는데, 이 지위는 각 국가에서 전문성과 기술을 갖춘 법인 또는 단체에 부여됨 • 마켓플레이스는 플랫폼에서 제품이나 서비스를 제공하는 판매자를 더 잘 추적하고 소비자에게 더 잘 공지해야 함
온라인 투명성	<ul style="list-style-type: none"> • 플랫폼은 콘텐츠 중재 결정을 더욱 투명하게 해야 함. 계정이 일시 정지되거나 해지된 이용자가 결정에 이의를 제기할 수 있는 내부 불만 처리 시스템을 제공해야 함. 분쟁해결을 위해 이용자는 유럽 국가의 독립적이고 공인된 기관에 문의하거나 해당 국가의 판사에게 문제를 의뢰할 수 있음 • 플랫폼은 사용자 프로필을 기반으로 특정 광고 콘텐츠를 추적하는데 사용하는 알고리즘이 어떻게 작동하는지 설명해야 함. 또 초대형 플랫폼과 검색엔진은 프로파일링에 기반하지 않는 콘텐츠 추천 시스템을 제공해야 할 의무가 있음. 아울러 다양한 정보(광고 후원자, 특정 개인을 타겟팅하는 방법과 이유 등)가

9) 이에 따라 동일한 콘텐츠가 한 국가에서는 삭제 의무대상이 되는 반면 다른 국가에서는 유포 가능할 수 있다.

	<p>포함된 광고 리스트를 대중에게 공개해야 함</p> <ul style="list-style-type: none"> • 미성년자를 대상으로 한 타겟팅 광고는 모든 플랫폼에서 금지되며 정치적 의견, 종교 또는 성적 취향과 같은 민감한 데이터 기반 광고도 (명시적 동의가 없는 한) 금지됨 • 다크패턴으로 알려진 기만적 인터페이스와 사용자를 오도하기 위한 관행(특정 선택사항 강조 등)도 금지됨
<p>위험완화 및 위기대응 (대규모 온라인 플랫폼/검색엔진의 추가 의무)</p>	<ul style="list-style-type: none"> • 대규모 플랫폼과 검색엔진은 온라인 보안, 정보 유포, 여론 형성 및 경제 거래에서 매우 중요한 역할을 하기에 허위정보와 같은 불법적이거나 유해한 콘텐츠를 유포할 때 발생하는 사회적 위험에 비례하여 이들에게 추가적인 조치가 부과됨. 구체적으로 다음 사항이 요구됨 • 온라인 혐오와 폭력, 기본권, 시민 담론, 선거과정, 공중보건 등 그들이 생성하는 시스템적 위험을 매년 분석하고 이를 완화하기 위해 필요한 조치(실천강령 준수, 가짜계정 삭제, 권위있는 정보 출처의 가시성 향상)를 취해야 함 • 유럽집행위원회의 감독 하에 매년 위험요소들이 감소했는지에 대한 독립적인 감사 수행 • 인터페이스 알고리즘을 위원회와 관할 국가 당국에 제공 • 연구자들이 온라인 위험의 진화를 더 잘 이해할 수 있도록 인터페이스의 주요 데이터에 대한 접근 제공 • 온라인에서 미성년자 보호 <ul style="list-style-type: none"> ※ 안보 또는 공중보건에 영향을 미치는 위기에 대응하기 위한 메커니즘 : 유럽위원회는 러시아의 우크라이나 침공과 같이 위기가 발생했을 때 주요 플랫폼에 인터페이스가 제기하는 위험을 분석하고 제한된 기간 동안 긴급조치를 취하도록 요청할 수 있음

* 출처 : 최민재·진민정·이현우 (2023.5.31.). 유럽의 가짜뉴스 대응정책

위의 표에서 간략히 살펴본 것처럼 대규모 온라인 플랫폼과 검색엔진 제공자에게는 추가적인 의무가 부여된다. 이는 대규모 플랫폼이 온라인상의 정보 유포, 여론형성, 경제적 거래 및 보안 등에 있어 매우 중요한 역할을 수행하기 때문이다. 이에 대한 효과적인 규제와 집행이 없는 경우, 이들은 그들이 발생시킬 수 있는 사회적·경제적 위험에 대한 적절한

인식과 대응 없이 사업모델과 서비스 규칙을 설계함으로써 커다란 우려를 야기할 수 있다. EU가 밝히는 대규모 온라인 플랫폼과 검색엔진 제공자가 갖는 구조적 위험(systemic risk)의 유형은 다음과 같다(34조1항).

<표 3-9> 대규모 온라인 플랫폼/검색엔진 제공자가 관리해야 할 구조적 위험의 유형

구분	주요 내용
첫 번째 유형	· 서비스를 통한 불법 콘텐츠의 유포
두 번째 유형	· 기본권(인간 존엄성, 사생활, 개인정보 보호, 표현과 정보의 자유, 차별 금지, 아동의 권리 존중, 소비자 보호 관련)에 대한 실제적인 혹은 예측 가능한 부정적 영향
세 번째 유형	· 시민담론과 선거과정 및 공공안보에 대한 실제적인 혹은 예측 가능한 부정적 영향
네 번째 유형	· 젠더 기반 폭력, 공중보건, 미성년자 보호, 개인의 신체적·정신적 안녕에 대한 실제적인 혹은 예측 가능한 부정적 영향

* 자료: 디지털 서비스법 34조 1항

이러한 구조적 위험은 허위조작정보와도 어느 정도 연관되어 있다. 동법의 제안이유에서는 대형 온라인 플랫폼과 검색엔진 제공자가 자사의 알고리즘을 이용한 추천 및 광고 시스템이 허위조작정보를 포함해 오해의 소지가 있거나 기만적인 콘텐츠를 유포하거나 증폭하는데 사용될 수 있다는 점에 대해 각별히 주의를 기울여야 한다고 지적한다.¹⁰⁾ 이러한 허위조작정보는 조작 등 다른 악의적인 활동과 마찬가지로 사회와 민주주의에 부정적인 영향을 미치는 구조적인 위험의 부정적 영향에 해당한다.¹¹⁾ 따라서 대규모 온라인 플랫폼과 검색엔진 제공자에게 부과되는 추가적인 의무는 부분적으로 허위조작정보의 확산

10) “Such providers should therefore pay particular attention on how their services are used to disseminate or amplify misleading or deceptive content, including disinformation.”

11) “Another area for consideration is the possible negative impacts of systemic risks on society and democracy, such as disinformation or manipulative and abusive activities or any adverse effects on minors.”

을 방지하는 데에도 기여한다.

정리하자면 디지털 서비스법은 불법 정보에 대해서만 책임을 묻는 것이 아니라 온라인 플랫폼과 대규모 온라인 플랫폼 및 검색엔진 제공자로 하여금 시민의 안전을 위협할 수 있는 정보에 대해 보다 적극적으로 개입하고 스스로의 판단 하에 조치를 취하도록 요구한다고 말할 수 있다.

(3) 대규모 온라인 플랫폼과 검색엔진 제공자의 추가적인 의무

대규모 온라인 플랫폼과 검색엔진 제공자가 지는 추가적인 의무를 보다 자세히 살펴보고자 한다. 여기에는 알고리즘을 포함한 서비스와 시스템의 위험 평가, 위험완화 조치를 취해야 할 의무, 위원회와의 협력, 독립적인 감사 의무, 프로파일링에 기반하지 않은 추천 시스템 옵션 제공, 추가적인 온라인 광고 투명성 확보, 데이터에 대한 접근권한 제공, 준법 감시 부서 설립, 투명성 보고서 발행 등이 포함된다.

<표 3-10> 대규모 온라인 플랫폼/검색엔진 제공자의 추가적인 의무

구분	주요 내용
제34조. 위험평가 (Risk assessment)	· 알고리즘을 포함해 서비스 및 관련 시스템의 설계, 기능 또는 서비스 사용으로 인해 발생하는 모든 구조적 위험에 대한 식별, 분석 및 평가 의무
제35조. 위험완화 (Mitigation of risks)	· 제34조에 따라 식별된 구조적 위험에 대해 합리적이고 비례적이며 효과적인 완화조치를 마련해야 할 의무. 여기에는 다음이 포함될 수 있음 (a) 온라인 인터페이스를 포함한 서비스의 디자인, 특징, 기능 (b) 이용약관 (c) 불법 콘텐츠 관련 통지의 속도 및 품질 (d) 추천 시스템을 포함한 알고리즘 (e) 광고 및 타겟팅 (f) 구조적 위험 감지와 관련된 내부 프로세스 등 (g) 신뢰할 수 있는 신고인과의 협력 또는 법정 외 분쟁

	<p>해결 기관의 결정 이행과 관련된 사항</p> <p>(h) 행동강령에 따른 다른 사업자와의 협력</p> <p>(i) 서비스 수신자에게 더 많은 정보를 제공하기 위한 조치</p> <p>(j) 연령확인 등 미성년자 보호를 위한 도구 등</p> <p>(k) 생성되거나 조작된 이미지, 오디오, 비디오에 대한 식별 보장 및 관련 기능 제공</p>
제36조. 위기 대응 메커니즘 (Crisis response mechanism)	<ul style="list-style-type: none"> · EU의 안보 또는 공중보건에 위기가 발생하면 위원회는 이사회에 권고에 따라 대형 온라인 플랫폼/검색엔진에 조치를 취하도록 요구하는 결정을 채택할 수 있음
제37조. 독립적인 감사 (Independent audit)	<ul style="list-style-type: none"> · 자체 비용으로 1년에 한 번 이상 자신의 의무수행에 대해 독립적인 감사를 받아야 할 의무
제38조. 추천 시스템 (Recommender systems)	<ul style="list-style-type: none"> · 추천 시스템을 사용하는 대형 온라인 플랫폼 및 검색엔진 제공자가 프로파일링에 기반하지 않은 추천시스템에 대한 옵션을 제공해야 할 의무
제39조. 추가적인 온라인 광고 투명성 확보의무 (Additional online advertising transparency)	<ul style="list-style-type: none"> · 광고를 싣는 전체 기간 및 광고가 마지막으로 노출된 후 1년 동안 다음 정보를 공개적으로 이용할 수 있도록 해야 할 의무 (a) 광고의 내용 (b), (c) 광고주 (d) 광고의 노출 기간 (e) 광고가 특정 군에만 노출됐는지 여부 및 이 때 사용된 주요 매개변수 (f) 대형 온라인 플랫폼에 노출되고 제26조(2)에 따라 식별된 상업적인 커뮤니케이션 (g) 도달한 이용자수, 타겟광고 이용자군의 이용자수
제40조. 데이터에 대한 접근과 조사 (Data access and scrutiny)	<ul style="list-style-type: none"> · 합리적인 요청이 있을 경우 의무의 준수여부를 감시하고 평가하는데 필요한 데이터 접근권한을 디지털서비스코디네이터 혹은 유럽이사회에 제공해야 할 의무
제41조. 준법감시 기능 (Compliance function)	<ul style="list-style-type: none"> · 운영기능과 독립적인 준법감시 부서(책임자 및 한 명 이상의 준법감시인 포함) 설립 의무

제42조 투명성 보고의무 (Transparency reporting obligations)	· 투명성 보고서를 최소 6개월마다 발행할 의무
---	----------------------------

이와 같은 의무들을 수행하기 위한 구체적인 방법은 대규모 플랫폼과 검색엔진 제공자에게 맡겨져 있다. 이 점에서 위 의무규정은 절차적인 의무규정이라고 할 수 있다. 이는 입법적으로 문제가 되는 부분을 직접적으로 규정하지 않고 온라인 플랫폼이 스스로 문제를 찾고 대처하도록 하는 장려하는 것이기도 하다(이병준, 2021).

한편 의무준수의 실효성을 확보하기 위해 집행위원회는 동법 안에 다양한 장치를 마련하고 있다. 대규모 온라인 플랫폼과 검색엔진 제공자의 디지털 서비스법 규정 준수 의무에 대해 조사할 수 있으며(65조), 합리적인 기간 내에 규정 위반과 관련된 정보를 제공하도록 요구할 수 있다(67조). 또 위원회는 본 규정을 위반한 대규모 온라인 플랫폼 및 검색엔진 제공자에게 규정 비준수 결정을 내리고(73조), 직전 회계연도 세계 총 매출의 6% 내에서 과징금을 부과할 수 있다(74조).

실제로 이스라엘-하마스 전쟁에서 허위조작정보의 확산이 문제가 되자 2023년 10월 유럽 집행위원회(European Commission)는 X(트위터)에 대해 이스라엘-하마스 전쟁과 관련된 혐오표현, 허위조작정보의 처리 정보를 요청했다. 이와 같은 요청은 디지털 서비스법에 따른 첫 번째 조사 사례가 될 수 있는 전 단계이다. 집행위원회는 X가 디지털서비스법에 따라 불법 콘텐츠, 허위조작정보, 젠더 기반 폭력의 전파 등이 미치는 부정적인 영향에 대해 위험평가(34조)를 실시하고 완화조치(35조)를 취해야 한다고 밝혔다. 집행위원회는 이와 관련된 X의 준수 여부를 조사하고 있으며(65조), X에 추가 정보를 요청할 수 있는 권리가 있다고 밝혔다(67조).¹²⁾

이와 관련해 X 측은 하마스와 연계된 것으로 확인된 계정이나 키워드 조작을 시도한 계정 수백 개를 삭제하고, 이용자들이 허위조작정보의 확산을 막기 위해 ‘커뮤니티 노트’ 기능을 적극적으로 활용해줄 것을 당부하는 등의 조치를 취했다. 커뮤니티 노트는 X에 올라온 콘텐츠에 대해 이용자들이 평가하고 의견을 달 수 있는 기능으로, 이의 적극적인 활

12) https://ec.europa.eu/commission/presscorner/detail/en/ip_23_4953

용은 이용자들의 자발적 참여를 통해 허위조작정보의 확산을 억제할 수 있을 것으로 기대됐다(박영흠, 2023.5.).

(4) 디지털 서비스법과 허위조작정보 대응 실천강령의 관계

디지털 서비스법은 유럽집행위원회의 허위조작정보 대응 실천강령(code of practice on disinformation) 과 함께 온라인 허위조작정보에 관한 공동규제 체계를 형성한다. 보다 구체적으로 동법 45조는 구조적인 위협(34조 1항)이 발생하고 대규모 온라인 플랫폼 및 검색 엔진이 이와 관련될 경우, 집행위원회는 이들을 포함한 관할 관청, 시민사회 조직 및 기타 이해관계자에게 특정한 위협의 완화 조치를 취하겠다는 약속 등을 담은 행동강령(code of conduct)에 참여하도록 요청할 수 있다고 명시한다.

또 본문은 아니지만 동법의 제안이유를 통해서도 허위조작정보의 유포 등과 관련해 대규모 온라인 플랫폼과 검색엔진 제공자가 특정한 행동강령을 준수하는 것은 적절한 위험 완화 조치로 간주될 수 있다고 밝히고 있다.¹³⁾ 만약 사업자들이 이러한 행동강령에 참여하라는 위원회의 권유를 적절한 설명 없이 거부할 경우, 해당 온라인 플랫폼 혹은 검색엔진 제공자가 본 규정의 의무를 위반했는지 여부를 판단할 때 이를 고려할 수 있다(recital 104). 또한 제안이유 106에서는 디지털서비스법의 45조와 같은 행동강령 관련 규정은 “허위조작정보 대응 실천강령(code of practice on disinformation)” 등 자율 규제 노력에 기여할 수 있다고 밝히고 있다(recital 106).

허위조작정보는 여러 플랫폼을 가로질러 빠르게 확산된다. 때문에 사업자에 대한 일방적인 책임전가나 감독기관의 독자적인 감시 활동만으로는 문제를 해결하는 것이 어렵다. 그래서 감독기관과 서비스 제공자, 시민사회와 전문가를 포함해 다각적인 이해관계자에 적절한 책임과 역할을 부담함으로써 상호 협력하는 사회적 거버넌스 체계를 구축하는 일이 필요하며 디지털 서비스법과 허위조작정보 대응 실천강령은 그러한 고민의 결과라고 할 수 있다. 허위조작정보에 대한 보다 자세한 설명을 다른 장에서 다루면서 이에 대한 논의를 이어 가도록 하겠다(박찬경, 2022).

13) 다만 행동강령에 참여해 그것을 이행한다는 단순한 사실 자체가 곧 본 규정을 준수하는 것으로 추정되어서는 안 된다는 단서도 달고 있다.

제 3 절 프 랑 스

1. 정보조작대처법

2017년 프랑스 대선에서 허위조작정보가 선거에 기승을 부렸고 이러한 허위조작정보가 선거에 영향을 미치게 된다면 민주주의의 토대가 무너지게 될 것이라는 우려가 확산됐다. 이에 따라 프랑스는 2018년 선거과정에서 제기되는 허위조작정보에 대한 대책의 일환으로 「정보조작 대처에 관한 2018년 12월 22일 법 n° 2018-1202」(이하 「정보조작대처법」)을 통과시켰다.

동법은 선거 기간 중에 확산되는 허위조작정보를 막는데 초점을 맞춘다. 때문에 선거기간, 즉 선거가 있는 달의 1일을 기준으로 3개월 전부터 투표일까지의 기간에 적용된다. 법에서 규정하는 허위조작정보도 선거에 초점이 맞추어 정의되어 있다. 여기서 허위조작정보는 “유권자 투표의 진정성을 해치는 부정확하고 오해를 불러올 수 있는 주장이나 비난 등이 온라인 정보통신서비스를 통해서 의도적이고, 인위적이며 또는 자동적으로, 대량으로 전파되는 정보”를 의미한다(고선규 외, 2021).

이를 위해 온라인 플랫폼에 더 강화된 투명성을 부여하는 한편, 신속한 문제 해결을 위해 법원의 판단이 이전보다 더 빨리 결정될 수 있도록 허용했다(박용숙 외, 2022). 구체적으로 해당 기간 동안 후보자, 정당, 시민단체 등 이해관계자는 온라인 서비스를 통해 유포된 허위조작정보에 대해 판사가 개입할 것을 요구할 수 있다. 다만 판사의 개입이 정당화되기 위해서는 정보는 명백한 허위여야 하고 인위적이며 대량으로 배포되어야 한다. 요청을 받은 판사는 48시간 이내에 관련 조치를 취해야 한다(최민재·진민정·이현우, 2023.5.31.).

이 법은 총 5개의 장과 20개의 조항으로 구성되어 있다. 그 가운데 제3장이 허위조작정보의 유포에 대응하기 위해 플랫폼 사업자에게 협조 의무를 부과하는 내용을 담고 있다. 구체적인 내용은 다음과 같다.

〈표 3-11〉 프랑스 「정보보작대처법」 제3장의 주요 내용

구분	주요 내용
제11조	<ul style="list-style-type: none"> · 온라인 플랫폼 운영자¹⁴⁾는 허위조작정보¹⁵⁾의 유포 방지를 위한 조치를 시행해야 함. 특히 제3자를 대신해 홍보된 콘텐츠에서 가져온 경우 인터넷 사용자가 이러한 허위조작정보를 운영자에게 통지할 수 있도록 장치를 설정해야 함. 또 다음과 같은 추가조치를 구현해야 함 <ul style="list-style-type: none"> 제1호. 알고리즘 투명성 제2호. 기업, 언론사 및 시청각 커뮤니케이션 서비스의 콘텐츠 홍보 제3호. 허위조작정보를 대량으로 유포하는 계정에 대한 조치 제4호. 공공토론에 관한 정보 콘텐츠의 홍보에 대한 대가로 플랫폼에 보수를 지급하는 자연인의 신원 또는 회사명, 등록된 사무소 및 법인의 목적에 관한 사용자 정보 제5호. 콘텐츠의 성격, 출처 및 배포방법에 관한 사용자 정보 제6호. 미디어 및 정보 교육 · 이러한 조치와 이에 투입된 비용은 공개되며, 온라인 플랫폼 운영자는 위조치를 구현하기 위해 절차가 명시된 선언문을 매년 시청각 및 디지털 통신규제 당국에 제출해야 함
제12조	<ul style="list-style-type: none"> · 시청각최고심의회(CSA)는 허위조작정보¹⁶⁾ 유포방지 조치에 기여해야 하며 필요한 경우 온라인 플랫폼 운영자¹⁷⁾에게 허위조작정보 유포방지책을 개선하기 위한 권장사항을 부과할 수 있음 · CSA는 제11조에 규정된 의무의 준수여부를 모니터링하고, 이에 대한 검토결과를 정기적으로 보고하며, 보고서 작성에 필요한 모든 정보를 제19조에 명시된 조건에 따라 온라인 플랫폼 운영자로부터 수집할 수 있음
제13조	<ul style="list-style-type: none"> · 온라인 플랫폼 운영자¹⁸⁾는 프랑스 영토에서 연락담당자의 기능을 수행하는 법적 대리인을 지정해야 함
제14조	<ul style="list-style-type: none"> · 온라인 플랫폼 운영자¹⁹⁾는 공공토론에 관련된 정보 콘텐츠를 추천, 분류 또는 참조하기 위한 알고리즘을 사용하는 작업에 대한 통계를 게시해야 함. 또한 각 콘텐츠에 대해 다음을 언급해야 함 <ul style="list-style-type: none"> 제1호. 추천, 순위지정, 또는 참조 알고리즘에 의존하지 않는 직접 액세스 부분 제2호. 해당되는 경우 온라인 플랫폼 내부 검색엔진 알고리즘, 한편 콘

	<p>텐츠 액세스에 개입한 온라인 플랫폼 추천, 분류 또는 참조의 또 다른 알고리즘</p> <p>· 이러한 통계는 온라인에 게시되며 모든 사람이 접근할 수 있어야 함</p>
제15조	<p>· 온라인 플랫폼 운영자, 언론사²⁰, 언론 간행물 또는 온라인 언론 서비스의 발행인²¹, 시청각 통신 서비스의 게시자²², 광고주²³, 언론인을 대표하는 조직 및 허위조작정보 확산 방지에 기여할 수 있는 기타 조직은 허위조작정보 유포 방지와 관련해 협력계약을 체결할 수 있음</p>

* 자료: 박용숙강명원·고수운 (2022). 2022년도 해외언론법제연구보고서

이에 따라 법에서 규정하는 기간에 페이스북, 트위터, 구글과 같이 프랑스 내에서 접속자가 월 500만 명 이상인 모든 플랫폼 사업자와 공적인 이슈에 관한 사회적 토론 내용을 전달하는 사이트 가운데 광고료 수입이 건당 평균 100유로 이상인 모든 사이트에게는 위와 같은 규정을 준수해야 할 의무가 부과되며 이에 대한 실시 방법을 공개해야 한다(고선규 외, 2021). 이 같은 의무에 대한 통제는 시청각최고심의회(Conseil supérieur de l'audiovisuel, CSA)에 위임된다(최민재·진민정·이현우, 2023.5.31.).

- 14) 선거법전 제L.163-1조 제1항에 언급된 온라인 플랫폼 운영자
- 15) 커뮤니케이션 자유에 관한 1986년 9월 30일 제86-1067법률 제33-1-1조 제1항에 언급된 공공질서를 교란하거나 투표의 진정성을 변경할 수 있는 허위조작정보
- 16) 주15와 상동
- 17) 주14와 상동
- 18) 주14와 상동
- 19) 주14와 상동
- 20) 언론사를 규제하는 1945년 11월 2일 제45-2646호 법률명령의 의미 내에서의 언론사
- 21) 언론의 법적 체계를 개혁하는 1986년 8월 1일 제86-897호 법률의 의미 내에서 언론간행물 또는 온라인 언론 서비스의 발행인
- 22) 커뮤니케이션 자유에 관한 1986년 9월 20일 제86-1067호 법률의 의미 내에서의 시청각 통신 서비스의 게시자
- 23) 소비법전의 의미 내에서의 광고주

2. 정보조작대처법에 대한 평가

정보조작대처법에 대해서는 그동안 많은 비판이 제기됐다. 주요 비판은 허위조작정보에 대한 정의가 모호하다는 점, 판사가 48시간 내에 해당 정보의 진실성 여부를 판단하는 것이 어려우며 나아가 판사에게 정보의 진실성 입증을 요구하는 것이 문제라는 점 등이다.

실제 적용에 있어서도 동법은 큰 효력을 발휘하지 못했다는 평가를 받는다. 2019년 유럽의회 선거에서 처음으로 적용됐을 당시 프랑스는 허위정보 신고가 쇄도할 것을 예상해 세 팀의 판사 그룹을 마련했으나 실제 단 한 건만이 회부됐다. 2022년 대선에서도 동법은 효력을 발휘하지 못했다는 평가를 받는다(최민재·진민정·이현우, 2023.5.31.).

구체적으로 2019년 회부된 한 건의 사건을 살펴보도록 하자. 프랑스의 의원 2명은 2019년 5월 10일 파리 지방법원에 정보조작대처법에 의거하여 트위터의 특정 콘텐츠에 대한 삭제제를 요청하는 가처분을 신청했다. 해당 콘텐츠는 해당 의원들이 파리에 위치한 모 대학병원에서 근무하는 간호사를 폭행하고, 이를 제지하기 위해 출동한 경찰관에도 부상을 입혔다는 것이다. 가처분 심리는 5월 16일에 열렸고 결정은 5월 17일에 확정됐다(박용숙 외, 2022).

결론적으로 프랑스 지방법원은 이를 승인하지 않았다. 선거 기간 동안 온라인 플랫폼의 콘텐츠를 허위조작정보로서 삭제하기 위해서는 (i) 투표의 진정성을 변경할 가능성이 있는 사실에 대한 부정확하거나 오해의 소지가 있는 주장 또는 비난, (ii) 고의적이거나 인공적 또는 자동화된 방식으로 유포, (iii) 온라인 커뮤니케이션 서비스를 통한 대규모 유포의 요건을 충족해야 한다. 법원은 (i)에 대해서는 메시지가 과장되기는 했지만, 대학병원 부지에 시위대가 침입한 사실이 있기 때문에 해당 주장과 실제 사실과 연결이 결여되어 있지 않아 명백하게 부정확하거나 오해의 소지가 있는 것은 아니라고 판단했다. (ii)에 대해서는 허위조작정보가 고의적이거나 인공적 또는 자동화된 방식으로 유포된다는 것은 봇을 사용하거나 봇의 사용을 인위적으로 확장하는 제 3자의 대가 지불에 의해 설정되는 것이기에 해당 기준을 충족하지 않는다고 판단했다. (iii)에 대해서는 해당 트윗이 과장된 용어를 사용했을 수 있지만 투표의 진정성을 변경할 수 있는 위협의 명백한 특성이 있는 것은 아니라고 판단했다(박용숙 외, 2022).

한편, 동법은 사업자들이 정보조작대처법의 적용을 받지 않기 위해 관련된 정보의 유통을 회피하도록 만드는 결과를 낳기도 했다. X(트위터)는 유럽의회 선거와 관련해 프랑스를 대상으로 한 표적광고를 하지 않기로 결정했고 시민들에게 투표를 장려하기 위한 내무부의 캠페인 방송도 거부했다가, 프랑스 정부와의 협상 끝에 정치캠페인이 아닌 홍보캠페인으로 판단된 내무부 캠페인에 대해서는 방송을 진행했다. 구글 또한 같은 이유에서 유럽의회 선거에 관한 모든 방송을 금지하는 결정을 하였다가 트위터와 마찬가지로 정부의 공식 커뮤니케이션 기관이 방송하는 중립적인 선거정보 광고를 홍보하는 데이는 동의했다. 이러한 플랫폼의 전략이 법적으로 문제가 되지 않는다면 이는 정보의 유통을 통제하는 온라인 플랫폼에 대한 새로운 역할 논쟁을 낳을 수 있다(박용숙 외, 2022).

제 4 절 독일

1. 소셜 네트워크 법집행법

독일은 2015년 SNS를 통해 광범위하게 유통되는 불법 정보 문제를 다루기 위해 대책위원회를 설립 및 운영하고, SNS를 서비스하는 플랫폼 사업자들이 스스로 독일 법률상 불법 정보를 자사 플랫폼에서 삭제하는 방안을 마련하여 시행하도록 했다(김민정, 2019). 이는 독일 연방법무부 및 소비자보호부가 소셜미디어 사업자 및 시민단체 대표와 함께 전담반을 운영해 자율규제 협의안을 맺은 결과로, 협의안에 따르면 플랫폼 사업자는 스스로 혐오표현이나 불법표현 등의 유통을 검토하고 24시간 이내에 이를 삭제하는 책임을 진다. 하지만 2017년 초까지 유튜브만이 약 90% 정도의 불법 콘텐츠를 삭제했을 뿐 페이스북(39%)이나 트위터(1%) 실적이 저조했다(고선규 외, 2021). 이에 독일은 소셜네트워크 서비스를 통해 유포되는 혐오표현, 허위조작정보의 생산확대·유포에 대응하기 위해 「소셜 네트워크 법집행법(Netzwerkdurchsetzungsgesetz, NetzDG)」을 제정했다. 동법은 2017년 10월 1일부터 발효됐고, 계도기간을 거쳐 2018년 1월부터 시행되고 있다. 이는 소셜 네트워크 서비스의 대표기업인 유튜브, 페이스북, 트위터 등 플랫폼을 통해 전파되는 혐오표현,

허위조작정보를 통한 정치적 선전·선동에 대응하는 것을 목적으로 한다(박용숙 외, 2022).

종래 독일은 인터넷상의 표현의 자유를 제한하는 것의 문제는 것보다는 인터넷을 통한 혐오표현 및 허위조작장보의 문제를 더 심각하게 받아들이고 있다. 이는 세계대전으로 인한 역사적 경험을 바탕으로 ‘방어적 민주주의’를 택하는 「독일 기본법」의 기본적인 정신이기도 하다(박용숙 외, 2022). 그럼에도 법 제정 당시 표현의 자유 위축에 대한 우려, 법의 명확성 원칙 위반, 기업에 대한 과중한 부담 및 사기에게 사실상의 검열을 허용하는 데서 오는 여러 문제점이 제기됐다. 하지만 독일 정부는 난민과 이민자에 대한 적대적인 표현과 신나치 혹은 극우세력의 부상에 따른 혐오표현 등에 대한 우려 속에 법을 통과시켰다. 이 점에서 소셜 네트워크 법집행법은 허위조작정보에 대한 대응의 측면도 갖고 있지만 그보다는 혐오발언에 대한 대응에 좀 더 초점이 맞추어져 있다고 볼 수 있다(고선규 외, 2021).

2. 소셜 네트워크 법집행법의 구성 및 내용

소셜 네트워크 법집행법의 적용 대상은 ‘수익창출을 목적으로 이용자 간 임의의 콘텐츠를 다른 이용자와 교환 및 공유하거나 공중이 이용하도록 이용자가 콘텐츠를 공개적으로 게시할 수 있는 인터넷 플랫폼을 제공하는 통신미디어서비스 제공자(즉 소셜 네트워크 사업자)’이다(박용숙 외, 2022). 간단하게 말하자면 상업적 이윤을 목적으로 하는 ‘소셜 네트워크 사업자’가 적용 대상이다. 다만 서비스 제공자의 직접적인 책임 하에 보도적인 내용을 편집해 만든 제공물로 구성되는 뉴스 웹사이트, 개인 커뮤니케이션 또는 특정 콘텐츠 유포를 위해 정해진 플랫폼인 이메일이나 메신저 사업자는 예외이다(박용숙 외, 2022). 또 독일 내에 등록된 사업자가 200만 명 이상인 소셜 네트워크 사업자만이 적용대상으로 소규모 사업자는 법률 적용대상에서 제외된다. 그리고 동법은 국내 사업자 뿐만 아니라 해외 사업자에게도 마찬가지로 적용된다.

소셜 네트워크 법집행법은 6개 조항으로 구성되어 있다. 제1조는 법의 적용범위를 규정하며, 제2조~제6조까지는 위법한 콘텐츠 관련 의무 부과 조항을 규정한다. 각 조항의 주요 내용은 다음과 같다(김민정 외, 2019).

<표 3-12> 소셜 네트워크 법집행법의 구성

구분	주요 내용
제1조	· 적용범위
제2조	· 공지의무(조치현황 보고 및 공개 의무)
제3조	· 불법게시물에 관한 이의제기 처리방법
제4조	· 벌칙 규정
제5조	· 국내 송달 대리인 임명 의무 규정
제6조	· 경과 규정

* 자료 : 김민정·박한우·정정주 (2019) 해외 인터넷플랫폼의 유해콘텐츠와 허위정보 대응방안

소셜 네트워크 법집행법에서 규정하는 ‘불법 게시물’은 독일 형법상의 불법정보 유형 21가지를 말한다. 해당 형법 조항 및 불법정보 유형은 아래 표와 같다. 이에 따르면 불법 정보는 주로 혐오표현 및 인격권 침해 표현에 초점이 맞추어져 있다. 이를 통해서도 소셜 네트워크 법집행법이 허위조작정보보다는 불법정보의 온라인 유통에 초점을 맞추고 있음을 파악할 수 있다. 허위조작정보는 이 불법정보 유형 안에 몇몇 유형(185조, 186조, 187조, 269조)으로 포함되어 있을 뿐이다(최민재·진민정·이현우, 2023.5.31.).

<표 3-13> 소셜 네트워크법 법집행법 상의 불법정보 유형

조항	내용
제86조	위헌조치 선전물 반포
제86a조	위헌조치 표시 사용
제89a조	국가를 위태롭게 하는 중대한 폭력의 예비
제91조	국가를 위태롭게 하는 중대한 폭력 행사의 지도
제100a조	간첩목적의 위조
제111조	공연한 범죄선동
제126조	범죄위협에 의한 공공평온교란
제129조	범죄단체조직

제129a조	테러단체조직
제129b조	외국에서의 범죄단체와 테러단체
제130조	국민선동
제131조	폭력물 반포 등
제140조	범죄의 대가 지급 및 찬양
제166조	신앙, 종교단체, 세계관단체 등 모욕
(184d와 결합된) 제184b조	방송, 미디어, 통신 서비스를 통한 아동포르노물의 전파, 취득 및 소유
제185조	모욕
제186조	비방
제187조	중상
제201a조	사진촬영을 통한 고도의 인격적 사생활 침해
제241조	협박
제269조	증거로 중요한 데이터의 위작

* 자료 : 김민정·박한우·정정주 (2019) 해외 인터넷플랫폼의 유해콘텐츠와 허위정보 대응방안

동법 제3조는 불법 게시물과 관련된 이의제기를 처리하는데 있어 사업자가 효과적이고 투명한 절차를 제공하여야 한다고 규정한다. 또 사업자는 이용자들이 손쉽게 이의제기할 수 있도록 이용자 친화적인 절차를 제공해야 한다. 즉 쉽게 이해할 수 있고, 직접 제기할 수 있으며, 언제나 이용 가능해야 한다(박용숙 외, 2022).

제3조에 따라 사업자는 이의가 제기됐을 때 그 내용이 범위반인지 여부 등을 판단해야 한다. 일차적으로는 약관 규정을 위반하는지 판단하여 위반 시 이를 삭제하고, 약관 위반이 확인되지 않으면 이차적으로 소셜 네트워크 법집행법에 따라 독일 형법 상 위법한 게시물인지를 판단하게 된다. 위법이 확실한 게시물에 대해서는 24시간 이내에 삭제 또는 접근을 차단해야 한다. 다만 위법한 콘텐츠의 삭제 혹은 차단과 관련해 보다 많은 시간이 필요하다고 형사소추관청과 합의한 경우는 제외한다.

위법이 확실하지 않는 게시물에 대해서는 7일의 기한 이내에 게시물을 삭제 또는 접근

을 차단해야 한다. 이러한 절차와 관련해 사업자는 공인된 자율규제기관(anerkannte Einrichtung der regulierten Selbstregulierung)에 판단을 위임할 수 있다(정다운, 2021).²⁴⁾ 제3자 기관은 독립성 및 전문성 요건을 충족하는 경우 연방법무청이 인증하는 형식으로 설치·운영되는데 인증 조건은 제3조제6항에 다음과 같이 규정되어 있다. 첫째, 해당 기관의 심사자의 독립성 및 전문성이 보장된 경우. 둘째, 적절한 설비 및 7일 이내의 신속한 심사가 가능한 경우. 셋째, 심사의 범위 및 절차 등 규칙이 있는 경우. 넷째, 민원처리 담당 부서가 설치되어 있는 경우. 다섯째, 해당 기관이 적절한 설비를 확보하는 경우. 이와 같은 제3자 기관의 인증은 연방법무청이 담당한다(고선규 외, 2021).

한편 기관에 판단을 위임하고 그 결정에 따르는 경우에는 게시물에 대한 조치를 취하는데 7일을 초과하는 것이 허용된다. 그 외에도 콘텐츠의 위법성에 대한 결정이 논쟁적이거나 특별한 사정이 있는 경우 조치 이전에 이용자의 민원에 대한 의견진술 기회를 제공할 수 있으며 이 경우에도 7일을 초과하는 것이 허용된다(고선규 외, 2021).

앞서 말한 바와 같이 소셜 네트워크 법집행법은 독일 내 등록 이용자가 200만이 넘는 사업자는 국내외를 가리지 않고 적용된다. 동법 제4조 범칙금 규정에서는 “독일 국내에서 행해지지 않은 질서 위반 행위에 대해서도 처벌할 수 있다” 고 이를 명시한다. 이를 위해 동법 제5조는 소셜 네트워크 운영자들이 독일에 송달 대리인을 지정하고, 해당 대리인에 대한 정보를 이용자들이 쉽게 인지할 수 있도록 서비스 플랫폼에 명시해야 한다고 규정한다. 소셜 네트워크 법집행법의 조항을 위반하는 경우 과태료는 5백만 유로까지 부과될 수 있다. 이를 감독하는 관할기관은 연방법무청이다.

소셜 네트워크 사업자는 이러한 의무의 준수여부에 대한 상세한 보고 의무가 있다. 동법 제2조에 따라 사업자는 연간 100건 이상의 불법 게시물에 대해 이의제기가 있는 경우, 이와 관련된 조치 현황 보고서를 6개월에 한번 씩 독일어로 작성해야 하며, 연방 관보 및 사업자 홈페이지를 통해 공개해야 한다. 보고서에 담겨야 할 내용은 동법 제2조제2항에 다음과 같이 규정되어 있다(김민정 외, 2019).

24) 제3자 기관은 2020년 3월에 설치됐는데, FSMM(Die Freiwillige Selbstkontrolle Multimedia-Diensteanbieter e. v.)이 정부로부터 최초로 인증을 받았다. 현재 제3자 기관의 운영은 플랫폼사업자와 FSM이 독립적인 법률전문가 50명으로 구성된 위원회를 설치해 운영하고 있다(고선규 외, 2021).

〈표 3-14〉 소셜 네트워크 법집행법 제2조제2항의 내용

주요 내용
① 소셜 네트워크 운영자가 플랫폼 내 처벌될 수 있는 행위를 방지하기 위해 한 노력에 대한 일반적 설명
② 불법 게시물에 대한 이의제기를 전달할 수 있는 메커니즘과 불법 게시물 삭제 및 차단 여부를 결정하는 기준에 대한 설명
③ 보고 기간 내 제기된 불법 게시물에 대한 이의제기 수(이의제기 신고기관, 이용자, 원인별로 분류)
④ 이의제기를 담당하는 부서의 조직, 인적 구성, 직원의 전문지식, 언어 능력 및 담당직원에 대한 교육 및 관리
⑤ 특정 분야의 단체 회원인지 여부와 함께 회원인 경우 해당 단체에 이의제기 처리부서가 존재하는지 여부
⑥ 결정을 내리기 위하여 외부기관에 자문을 한 이의제기의 건수
⑦ 보고기간 중에 이의제기를 받아 삭제 또는 차단된 게시물의 수
⑧ 소셜 네트워크에서 이의가 제기된 시점으로부터 실제 불법 게시물이 삭제 또는 차단되기까지의 시간(이의제기 신고기관, 이용자, 원인 및 소요시간(24시간 이내, 48시간 이내, 1주일 이내, 1주일 초과)을 기준으로 분류)
⑨ 이의 제기된 게시물의 처리에 관하여 결정한 내용을 이의제기자 및 이의 제기된 게시물을 저장하고 있는 이용자에게 통지하기 위하여 취한 조치

* 자료 : 김민정·박한우·정정주 (2019) 해외 인터넷플랫폼의 유해콘텐츠와 허위정보 대응방안

3. 소셜 네트워크 법집행법에 대한 평가

소셜 네트워크 법집행법을 통한 규제 강화는 자율규제를 활성화시키는 동시에 입법을 통해 자율규제에 대한 규제를 강화하는 것이라고 할 수 있다. 즉 입법은 자율규제에 대한 감독 체계를 강화시키는 역할을 하고 기본적으로는 사업자의 자율규제를 활성화하는 것이 그 목적이다(정다운, 2021).

동법은 허위조작정보에 관한 내용을 포함하고 있지만, 허위조작정보의 대응을 위한 법 이라기보다는 혐오표현의 대응에 관한 법률로 이해되고 있다(고선규 외 2021, 김민정 외 2019, 박용숙 외 2022). 혐오표현을 주된 규제대상으로 한다는 점에서 허위조작정보의 대응에 있어서는 적용 가능성과 효과가 낮다는 의견도 제시되었다(Holznel, 2018 : 정다운, 2021, 89쪽에서 재인용). 실제로 2020년 유튜브의 삭제 콘텐츠를 보면 프라이버시 침해, 혐오표현, 명예훼손 등의 유형이 대부분을 차지하는 것을 확인할 수 있다.

<표 3-15> 유튜브의 삭제 콘텐츠 내용과 삭제 건수

콘텐츠의 종류	NetDG법에 따른 삭제건수	이용규약에 따른 삭제건수
프라이버시 침해	1,176	4,894
명예훼손·굴욕	957	18,330
부적절·위험한 내용	86	12,376
포르노 관련 내용	38	12,833
테러 및 위협 내용	161	3,070
혐오표현·정치적 극단주의	986	28,211
폭력	62	7,634

* 자료 : Pikel Moderationsberichte 2020. 고선규 외 (2021)에서 재인용

또 동법은 게시물에 불법 여부 및 제3자 권리침해 여부, 게시자의 표현의 자유 제한 가능성 등에 대한 판단 및 조치를 민간 기업의 책임에 맡기는 점, 명확하게 불법인 게시물과 명확하지 않은 게시물을 구분할 기준이 불분명하다는 점 등으로 인해 많은 비판을 받았다. ‘표현의 자유에 대한 UN특별 보고관’ 인 데이빗 케이(Kaye, D)도 동법이 국제인권기준에 부합하지 않는다고 판단했다(박용숙 외, 2022).

이러한 우려에도 불구하고 법 발효 이후 4년간 소셜 네트워크 법집행법은 사업자가 범칙금에 대한 부담으로 게시물을 과잉차단(over blocking)하는 문제는 야기하지 않은 것으로 평가 받는다. 그 이유는 대상 콘텐츠가 형법에 명시돼 있고, 판단에 필요한 시간을 고려하여 기일 안에 판단이 어려운 사안에 대해서는 유예조치를 취하고 있기 때문이다. 또

기일을 초과한 경우에도 사안의 성질을 고려해 과태료 부과대상이 되지 않도록 배려하는 시스템이 법률적으로 작동하는 것도 여기에 영향을 미친다. 실제로 소셜 네트워크 서비스 사업자의 삭제 비율은 최대 35%~최소 5%로 나타났으며, 이러한 삭제조치에 대한 이의 신청은 발생하지 않은 것으로 나타났다(고선규 외, 2021).

다만 페이스북과 같은 일부 플랫폼에서 삭제 건수가 적은 이유로, 사이트 내 삭제 요청이나 민원을 신청하는 절차가 어렵거나 복잡하기 때문이라는 이유도 제기됐다. 이로 인해 페이스북은 2019년 7월 200만 유로에 해당하는 과태료를 부과받기도 했다(고선규 외, 2021).

제 5절 영국

1. 온라인 유해백서

영국에서는 2016년 브렉시트 국민투표를 전후해 페이스북 이용자들의 정보가 무단으로 선거운동에 이용되고, 온라인 플랫폼을 통해 허위조작정보가 확산되면서 관련 논의가 본격화되기 시작했다. 이에 2017년 영국 하원은 가짜뉴스(fake news)에 대한 문제를 공식적으로 제기하며 규제 논의를 본격화했다(김재선, 2020).

이 맥락에서 영국의 디지털문화미디어스포츠부(Department for Digital, Culture, Media & Sport, DCMS)는 2019년 4월 온라인 유해 콘텐츠의 규제, 방지 및 기업의 대응 방안을 포함한 <온라인 유해 백서(Online Harms White Paper)>를 발표했다. 또 이에 대한 공개 의견 수렴을 진행해 2020년 12월에는 정부의 입장을 담은 <온라인 유해백서 전체답변서(Online Harms White Paper : Full government response to the consultation)>를 발표했다. 백서에는 플랫폼 사업자의 책임을 강화하기 위해 다양한 이슈를 검토했다. 주요 검토 내용은 다음과 같다(김재선, 2020).

첫째, 플랫폼 사업자가 허위조작정보를 인지할 수 있는 가능성에 대해, 보고서는 현재 플랫폼 사업자가 제공하고 있는 검색기능(타겟 기술) 등 기술 및 서비스만으로 충분히 유해한 허위조작정보의 발생과 전파가능성을 인지할 수 있고 충분히 허위조작정보의 유통을

사전에 방지할 수 있다고 평가했다. 둘째, 플랫폼 사업자가 허위조작정보를 인지했다 하더라도 내용의 유해성을 판단하기 쉽지 않다는 점에서 사업자의 책임을 산출하기는 쉽지 않다는 문제가 있다. 하지만 보고서는 페이스북이 2018년 개인정보보호 조치를 취하지 않았다는 이유로 과징금을 부과 받은 것을 예로 들어 안전조치를 취하지 않았다는 것만으로도 과징금이 부과될 수 있다고 주장했다. 셋째, 온라인 플랫폼 사업자의 책임성을 강화하는 방안으로 사업자 소득의 2%에 해당하는 디지털 서비스 세(digital service tax)를 제도화하여 기금으로 형성하고 이를 허위정보 단속 및 관리에 사용하자는 논의도 이루어졌다. 넷째, 플랫폼 사업자와 정부의 동의 의결(consent decree) 방안이 논의됐다. 이는 앞서도 설명한 미국 FTC와 페이스북의 동의의결 사례를 참고한 것이다. 다섯째, 해외 기업의 데이터 활용에 대한 영국 법제 적용 가능성이 논의됐다. 영국 정보위원회의 입장은 해외 기업의 데이터에 대해서도 영국의 규제가 적용될 수 있어야 한다는 입장이다(김재선, 2020).

한편 온라인 유해 백서와 이에 대한 답변서에서 진행된 논의는 2021년 「온라인 안전 법안(The Online Safety Bill)」의 마련으로 이어졌으며, 이는 2023년 9월 국회를 통과했다.

2. 온라인 안전법

온라인 안전법은 온라인 이용자의 안전을 보장하는 것을 목적으로 한다. 그 목적은 크게 다섯가지로, (i) 온라인에서의 이용자 안전 강화 (ii) 언론의 자유 보호 및 강화 (iii) 온라인에서 불법 콘텐츠에 대처하는 집행기관의 능력 향상, (iv) 온라인에서 스스로를 안전하게 지킬 수 있는 이용자 능력 향상 (v) 온라인 위협에 대한 사회적 이해 제고이다 (McCarron, 2023.11.13.).

이를 위해 동법은 두 종류의 인터넷 서비스 제공자, 즉 이용자 간(user to user) 서비스 제공자와 검색엔진 서비스 제공자에게 특정 콘텐츠로부터 이용자들을 보호하도록 하는 주의 의무를 부과한다. 이들 서비스가 상당수의 영국 이용자를 호스팅하거나, 서비스의 주요 타겟 시장이 영국이거나, 서비스가 영국 내에서 이용 가능하며 서비스의 이용자 생성 콘텐츠가 영국 내 개인에게 중대한 피해를 줄 수 있다고 믿을만한 합당한 근거가 있는 경우에 동 법의 적용 대상이 된다.

여기서 규제 대상이 되는 콘텐츠는 불법 콘텐츠와 유해 콘텐츠이다. 불법 콘텐츠는 테러행위나 아동 성착취·아동 성학대 등의 내용을 담고 있는 콘텐츠를 가리킨다. 유해 콘텐츠는 적법하지만 유해한 콘텐츠를 가리키는데 법에서는 그 기준을 명확하게 정의하고 있지 않다. 대신에 이를 아동 유해 콘텐츠와 성인 유해 콘텐츠로 구분하며, 서비스 제공자들은 적절한 시스템과 조치를 통해 불법적인 콘텐츠뿐만 아니라 유해 콘텐츠로부터도 서비스 이용자, 특히 아동을 보호해야 할 의무를 진다(김지현, 2022).

EU의 디지털서비스법과 비슷하게 영국의 온라인 안전법도 사업자 규모에 따라 차등적인 의무를 부과한다. 이에 따라 보다 영향력이 큰 대규모 온라인 서비스 사업자는 추가적인 의무를 이행해야 한다. 이용자 간 서비스 중 추가적인 의무를 지는 서비스를 동법에서는 카테고리1(category 1) 서비스로 규정한다. 검색엔진 서비스 중 추가적인 의무를 지는 서비스는 카테고리 2A 서비스로 규정된다. 또한 동법은 아동이 접근할 가능성이 높은 서비스에 대해서도 추가적인 의무를 규정한다.

기본적으로 모든 이용자 간 서비스 및 검색엔진 서비스는 다음과 같은 의무를 공통적으로 지게 된다. 검색엔진 서비스도 대부분은 동일한 의무를 진다.

<표 3-16> 온라인 서비스 제공자의 의무

주요 내용
<ul style="list-style-type: none"> ○ 불법 콘텐츠 위험 평가 의무. 서비스는 적절하고 충분한 불법콘텐츠 위험평가를 수행해야 하며 이를 최신 상태로 유지해야 함 ○ 불법콘텐츠와 관련된 위험을 완화하고 관리하기 위한 조치를 취할 의무. 또 서비스 약관에 이용자를 보호하는 방법을 명시하는 조항을 포함해야 하며 이러한 조항은 사용자가 명확하고 쉽게 확인할 수 있어야 함 ○ 투명성, 신고 및 구제에 관한 의무. 서비스는 불법콘텐츠를 쉽게 신고할 수 있는 이해하기 쉬운 방법을 마련해야 하며, 이용자가 접근 가능한 불만처리 절차를 운영해야 함. 또한 불법콘텐츠에 대한 신고 기록과 위험평가 기록을 보관해야 함 ○ 표현의 자유 및 개인정보 보호에 관한 의무. 서비스는 안전조치 및 정책을 결정하고 시행할 때, 이용자의 표현의 자유와 개인정보 보호를 특별히 고려해야 함

* 자료 : 온라인서비스법

여기에 더해 아동이 접근할 가능성이 높은 서비스에 대해서는 (i) 아동과 관련된 위험 평가 의무 (ii) 아동을 보호하는 안전 의무를 추가적으로 수행해야 한다. 서비스 제공자는 서비스 설계 또는 운영과 관련해 아동이 겪을 수 있는 피해의 수준을 줄이기 위해 적절한 조치를 취해야 하며, 연령확인 혹은 연령추정 기능을 활용해 아동이 유해한 콘텐츠에 노출되지 않도록 해야 한다. 또 서비스 약관에서는 가장 최근의 아동 위험 평가(위험수준, 아동에게 미칠 수 있는 잠재적 피해의 성격 및 심각성 등)를 요약해 게시해야 한다.

한편 카테고리1 서비스에는 다음과 같은 추가 의무가 적용된다.

<표 3-17> 카테고리1 서비스의 추가 의무

주요 내용
<ul style="list-style-type: none"> ○ 성인 이용자의 역량강화 및 권한부여 의무. 서비스는 성인 이용자가 온라인에서 접하는 유해 콘텐츠를 제어하고 관리할 수 있도록 권한을 부여하고 이와 관련된 평가를 수행해야 함 ○ 민주주의에 중요한 콘텐츠를 보호할 의무 ○ 뉴스 발행자의 콘텐츠를 보호할 의무 ○ 저널리즘 콘텐츠를 보호할 의무 ○ 사기성 광고를 접하는 것을 방지하고, 사기성 광고가 표시되는 시간을 최소화 하며, 신고된 사기성 광고를 신속하게 삭제할 의무

* 자료 : 온라인서비스법

특히 뉴스 발행자의 콘텐츠를 보호할 의무는 이와 관련된 콘텐츠에 대해서 결정을 내릴 때 해당 발행자에게 서비스 제공자가 취하려는 조치를 명시하고, 그 이유를 설명하며, 결정과정에서 표현의 자유의 중요성을 어떻게 고려하는지 설명하도록 하고, 뉴스 발행자가 진술할 수 있는 합리적인 기간을 명시하도록 규정함으로써 표현의 자유를 보호하고 침해 를 최소화하도록 하고 있다.

온라인안전법과 관련해서는 커뮤니케이션청(Ofcom)이 주요 규제기관으로서 준수 여부를 점검하는 역할을 수행한다. Ofcom은 기업이 해당 규정을 준수하지 않을 경우 1,800만

파운드 또는 최근 회계연도전 세계 연간 매출의 10% 중 높은 금액의 벌금을 부과할 수 있는 권한을 갖게 된다.

3. 온라인 안전법에 대한 평가

온라인안전법은 플랫폼 기업이 더 투명하고 더 능동적으로 대처하도록 요구한다. 이와 관련해 영국의 도넬란(Donelan, M.) 과학기술혁신부(DSIT) 장관은 이 법이 “언론의 자유를 보호하고 성인에게 권한을 부여하며 플랫폼이 불법 콘텐츠를 삭제하도록 보장할 것이다. 그러나 이 법안의 핵심은 어린이를 보호하는 것이다” 라고 밝힌 바 있다(WIRED, 2023.10.26.).

하지만 동법은 허위조작정보의 직접적인 규제에 관한 내용을 담고 있지는 않다. 허위조작정보와 관련된 직접적인 내용으로는 동법 152조에서 Ofcom으로 하여금 허위조작정보 및 오정보에 관한 자문위원회를 설치하도록 규정한 것이 전부이다. 이와 관련해 영국의 팩트체크기관인 풀팩트(fullfact)는 법안 단계에서는 기업이 서비스 약관에 명시해야 하는 유해콘텐츠의 우선순위 목록에 건강과 관련된 허위조작정보를 포함시킬 계획이었으나 이것이 지켜지지 않았으며 따라서 허위조작정보에 대한 대처로는 불충분하다고 비판했다.²⁵⁾ 라이트(Wright, J.) 영국 법무부 장관은 온라인 안전법에 허위조작정보 및 오정보 관련된 캠페인이 규제 대상에서 제외된 것은 부분적으로 여러 부처의 소관업무가 혼동되었기 때문이라는 의견을 밝히기도 했다(WIRED, 2023.10.26.).

한편 동법은 아동이 온라인에서 보는 콘텐츠에 대한 주의 의무를 부과하는 한편 성인에 대한 권한부여도 규정하고 있기 때문에, 관련 기업은 사용자 연령확인부터 시작해 여러 가지 새로운 요구사항을 준수해야 한다. 이를 두고 영국에서 8번째로 많이 방문하는 웹사이트인 위키피디아(Wikipedia)는 사용자 데이터 수집에 관한 위키미디어 재단의 원칙에 위반되기 때문에 이 규정을 준수할 수 없다고 밝히기도 했다.

25) <https://fullfact.org/about/policy/online-safety-bill/>

제 6절 일본

1. 개관

일본 정부에서 허위조작정보에 대한 대책을 담당하는 부서는 총무성(한국의 행정안전부)이다. 총무성은 허위조작정보와 관련한 대책을 마련하는데 있어 EU의 대응을 모델로 삼아 자율규제를 중시한다. 기본적으로 허위조작정보를 판명하는 것이 모호할 수밖에 없기 때문에 표현의 자유를 제한하는데 있어 조심스럽게 접근하기 때문이다(고선규 외, 2021). 실제로 일본 총무성은 허위조작정보에 대한 법적 규제를 검토한 바 있으나, 헌법이 보장하는 표현의 자유를 과도하게 억제할 우려가 있다는 점으로 인해 법률 제정은 보류됐다(박용숙 외, 2022).

2. 프로바이더 책임제한법

허위조작정보를 직접 규제하는 법률은 아니나, 일본에서는 인터넷상에서의 내용규제와 관련된 법률로 2001년에 제정된 「특정 전기통신사업자의 손해배상책임 제한 및 발신자 공개에 관한 법률」(이하 ‘프로바이더 책임제한법’)이 있다.

동법이 삭제요청 혹은 발신자 정보공개청구의 대상으로 하는 정보는 특정 개인의 권리를 침해하는 정보에 한정된다(박용숙 외, 2022).

동법에 의거해 자신의 권리를 침해당한 개인은 프로바이더, 즉 인터넷 서비스 제공자 등에게 삭제를 요청할 수 있으며, 삭제 요청을 받은 프로바이더는 법이 정하는 요건을 충족하는 경우에는 해당 게시물에 대한 삭제 등 조치를 강구해도 손해배상 책임에서 면책된다(박용숙 외, 2022). 또한 자신의 권리를 침해당한 개인은 프로바이더 등에게 해당 게시물을 작성한 발신자에 관한 정보 공개를 청구할 수 있다. 이는 권리의 침해를 받은 자가 가해자인 발신자를 특정해 손해배상청구를 할 수 있도록 하기 위함이다. 법이 정하는 요건을 충족하는 경우 프로바이더는 발신자의 권리의익이 부당하게 침해되지 않도록 원칙적으로 발

신자의 의견을 청취한 후 공개 여부를 결정해야 한다. 프로바이더 등이 임의로 공개하지 않을 경우 권리가 침해된 자는 법원에 공개 청구 소송을 제기할 수 있다. 동법은 2022년 한 차례 개정돼 피해자가 발신자의 개인정보 공개를 청구하기 위한 절차가 한결 완화됐다(김민정 외, 2023).

프로바이더 책임제한법은 프로바이더가 적절한 조치를 취한 경우 손해배상 책임이 면제될 수 있음을 제시한다는 점에서, 상대적으로 플랫폼 사업자의 책임을 약화시키고 있다는 비판을 받기도 한다. 또 피해자는 정보개시를 통해 발신자를 특정해 손해배상청구권을 행사할 수 있지만, 실제 손해배상을 받는 것은 쉽지 않은 일이기기에 피해자 보호에 미흡하다는 의견도 있다(김연진, 2021).

3. 총무성의 허위조작정보 대책 방안

프로바이더 책임제한법의 개정과 같은 일부 조치가 있지만 그럼에도 전반적으로 일본의 기본적인 규제 틀은 사업자에 대한 직접적인 규제보다 자율적인 대응방안을 우선하는 것이다. 대응방안을 둘러싼 주요 논의도 허위조작정보와 관련된 이해관계자의 협력방안을 구축하고, 다면적인 대응 방안을 구축하는 과정에서 플랫폼 사업자가 산학협력 및 관학협력 체제를 구축하고 소통 시스템을 마련하는 것에 중점을 두고 이루어지고 있다. 구체적인 방안으로는 국내외 주요 플랫폼 사업자, 정부, 전문가, 이용자 등 관계자로 구성된 포럼을 설치하여 운영하는 것이다. 포럼 운영 과정에서 일본 정부는 다양한 이해관계자의 협력체계를 구축하고 지원하는 코디네이터 역할을 수행한다(總務省 2019; 고선규 외, 2021에서 재인용).

또 총무성은 선거기간 또는 재해 발생 시 소셜 네트워크에서의 허위조작정보 확산을 방지하기 위한 대책을 추진하고 있다. 2019년 10월 25일에 발표된 총무성의 허위조작정보 대책 방안은 다음과 같다.

<표 3-18> 총무성의 허위조작정보 대책 방안(2019년)

주요 내용
첫째, 자율규제원칙
둘째, 허위정보 실태조사
셋째, 관련 당사자의 협력체제 구축
넷째, 플랫폼 서비스 사업자의 투명성 설명책임 확보
다섯째, 팩트체크 추진
여섯째, 정보리터러시 교육 강화
일곱째, 연구활동 지원
여덟째, 정보발신자에 대한 신뢰성 확보
아홉째, 국제협력 중시

* 자료 : 総務省 (2019). 고선규 외(2021)에서 재인용

또 2019년 4월에는 허위조작정보에 대응하고 인터넷 상의 정보에 대한 신뢰성을 높이는 것을 목적으로 인터넷 미디어 협회(JIMA)가 설립됐다. JIMA에는 뉴스미디어나 콘텐츠 미디어 등 다양한 분야의 주요 미디어가 참여하고 있으며, 윤리강령을 책정하고, 각 미디어의 자율적인 가이드라인 책정을 지원하기 위한 지식을 제공하며, 이용자의 리터러시 교육을 지원하는 등 신뢰성 향상을 위한 실천을 전개하고 있다(고선규 외, 2021; 박용숙 외, 2022).

제7절 소결

온라인을 통한 커뮤니케이션이 활성화되면서 이를 통해 생성되고 유포되는 정보에 대한 관리 필요성에 대한 논의도 확산됐다. 이러한 논의의 일부는 법적인 근거와 체계를 수단으로 하여 강제력 있는 법적 효과를 발휘하는 법적 규제의 적용으로 나타난다.

이 장에서는 미국과 EU를 비롯해 다양한 국가의 허위조작정보 관련 입법논의를 살펴봤다. 물론 지금까지 검토한 법이 모두 허위조작정보를 직접 규제하는 것은 아니다. 직접적으로 허위조작정보를 규제하는 법은 프랑스의 정보조작대처법, 그리고 미국 일부 주의 선거법에 불과하다. EU의 디지털서비스법은 허위조작정보에 대한 대처를 강조하고 있기는 하지만 직접적인 규제 대상으로 삼는 것은 불법정보이다. 독일의 소셜 네트워크 집행법은 불법정보를 규제 대상으로 삼으며 이를 통해 혐오표현의 유통 확산을 억제하고자 한다. 영국의 온라인 안전법은 불법정보와 유해정보를 규제 대상으로 삼는다. 상대적으로 가장 미온적인 대처를 취하고 있는 미국과 일본도 입법을 통해 허위조작정보를 직접 규제하지는 않으며 각각 통신품위법과 프로바이더 책임제한법을 통해 일정 조치를 취한 플랫폼 사업자의 책임을 면제해주고 있다. 이 때 조치를 취해야 하는 콘텐츠는 통신품위법의 경우 음란물이나 폭력적 콘텐츠 등이 해당하고 프로바이더 책임제한법의 경우 특정 개인의 권리를 침해한 콘텐츠가 해당한다.

이러한 차이에도 불구하고 대부분 법에서 안전한 온라인 공간을 만들기 위한 플랫폼 사업자의 책임에 대한 논의가 강조되는 것은 공통적인 특징이다. 이는 플랫폼이 오늘날 온라인상의 정보의 유통, 여론의 형성, 경제적인 거래 및 이용자 정보의 보호 등에 있어서 매우 중요한 역할을 수행하기 때문이다. 특히 대규모 플랫폼일수록 정치·사회·경제적인 측면에서 커다란 영향을 미칠 수 있다. 때문에 특정 정보에 대한 직접적인 통제를 목적으로 하는 프랑스의 정보조작대처법과 독일의 소셜 네트워크 집행법은 일정 규모 이상의 플랫폼을 적용 대상으로 삼는다. 전자는 프랑스 내 접속자가 월 500만 명 이상인 플랫폼과 공적 토론을 전달하는 사이트 가운데 광고수입이 건당 평균 100유로 이상인 사이트에 적용된다. 후자는 독일 내 등록 이용자가 200만이 넘는 사용자에게 적용된다. EU의 디지털서비스법과 영국의 온라인 안전법은 서비스의 규모에 따라 법적 규정의 준수 의무에 차등을

둔다. EU는 이용자가 월 4,500만 명인 대규모 온라인 플랫폼과 검색엔진에 차등적인 규제를 부과한다. 영국의 온라인 안전법은 일정 규모 이상의 이용자 간 서비스와 검색엔진에 대해 추가적인 의무를 부과하는데 구체적인 기준은 아직 논의 중이다.

플랫폼 사업자의 책임 논의를 더욱 강화하는 것은 알고리즘을 활용하는 추천 시스템, 타깃 광고, AI 생성 콘텐츠에 대한 우려이다. 플랫폼은 추천시스템과 타깃광고, AI를 통해 단순히 정보를 전달하고 확산하는 것을 넘어 사람들에게 노출되는 정보를 가공하고 편집할 수 있기 때문이다. 때문에 미국에서 최근 발의된 플랫폼 책임 및 소비자 투명성 법안, 필터 버블 투명성 법안 등은 알고리즘의 이용과 관련해 투명성을 강화하고, 이용자의 역량(선택권)을 강화하기 위한 조치를 요구한다. 최근 바이든 행정부가 서명한 AI 규제 관련 행정명령은 AI에 의해 생성된 콘텐츠에 대한 라벨링 등 디지털 콘텐츠의 진위와 식별을 확실히 하기 위한 내용을 담고 있다. 텍사스 주와 캘리포니아 주의 선거법은 선거기간 동안의 딥페이크 콘텐츠를 규제한다. 한편 EU의 디지털서비스법은 대규모 플랫폼과 검색엔진 제공자에게 생성되거나 조작된 콘텐츠에 식별을 확실히 하기 위한 조치를 취하고, 추천시스템의 투명성 및 이용자 선택권, 온라인 광고의 투명성 확보 등을 요구한다. 디지털 서비스법과 영국의 온라인 안전법은 플랫폼의 투명성을 강화하도록 요구함으로써 플랫폼이 제공하는 콘텐츠와 서비스에 대해 이용자들이 더 많은 지식과 권한과 역량을 부여받도록 규정한다. 플랫폼은 이상의 의무 준수 여부에 대해 주기적으로 투명성 보고서를 발간해야 한다. 각 국의 규제기관은 이와 같은 플랫폼의 의무준수 여부를 감시함으로써 법적 규제의 실효성을 확보한다.

이처럼 대부분 입법이 허위조작정보에 대한 직접적인 규정이나 명확한 기준을 회피하고 플랫폼 사업자의 책임을 강조하는 데에는 여러 이유가 있다. 먼저 허위조작정보를 법으로 규정하는 것이 쉽지 않기도 하며, 규제입법은 자칫 표현의 자유를 과도하게 침해할 수 있기 때문이다. 적극적인 대처를 택한 프랑스와 독일의 입법에 대해서는 표현의 자유를 위축시킬 수 있다는 비판이 제기됐고, 프랑스에서는 일부 플랫폼이 문제가 될 수 있는 콘텐츠 범주의 유통 자체를 회피하는 결과를 낳기도 했다. 이 맥락에서 영국의 온라인 안전법이 민주주의에 중요한 콘텐츠, 뉴스 발행자 콘텐츠, 저널리즘 콘텐츠의 보호 의무를 특별히 규정한 것은 사업자의 책임을 강화하는 입법이 표현의 자유와 조화를 이루도록 하기 위한 노력으로 이해할 수 있다.

두 번째 이유는 특정 정보나 콘텐츠를 허위조작정보로 규정하고 범죄화하여, 위반을 저지른 자에 대해 형벌을 가하는 방식은 정보의 유통이 막대해지고 확산과 과급이 신속해진 오늘날의 현실에서 유효성을 갖기 힘들다는 점이다. 인터넷 서비스 사업자들은 국가보다 특정 콘텐츠를 식별하고 차단함으로써 이용자를 보호할 수 있는 역량이 훨씬 뛰어나다. 반면 국가는 무수히 많은 인터넷 이용자의 다양한 표현행위를 규제하는 것보다 인터넷 사업자를 규제하는 것이 더 수월하다(Balkin, 2018; 김희경 외 2022에서 재인용). 때문에 이와 같은 규제가 인터넷의 속성에도 부합하고 문제 해결에도 용이하다.

반면에 법적인 근거와 수단 없이 사업자에게 온전히 규제의 책임을 맡기는 것은 법적인 강제력과 신뢰할 수 있는 판단 기준의 문제 때문에 타당성과 실효성에 대한 논란을 낳을 수 있다.

사업자에 대한 일방적인 책임전가나 규제 기관의 독자적인 활동만으로는 오늘날 인터넷 상에서 유포되는 정보에 관한 문제를 해결하기 어렵다. 이는 오늘날 인터넷 규제가 정부와 민간이 함께 협력하는 규제 형태인 공동규제(co-regulation)의 형태로 나아가고 있는 이유를 설명한다. 공동규제는 민간 영역이 전통적인 정부 영역에 해당했던 규제 영역에 적극 참여하고, 정부는 그러한 민간의 활동과 역할에 대해 협력·지원함으로써 규제의 합리화 및 효율성을 추구하는 방식이다. 법·제도와 유리되어 있는 완전한 자율규제는 현실적으로 불가능하고,²⁶⁾ 완전한 법적규제는 효율성이 지나치게 낮기 때문에 공동규제는 인터넷 콘텐츠에 대한 효과적인 규제 방식으로 떠오르고 있다(김희경, 2022).

대표적으로 EU는 디지털서비스법과 함께 허위조작정보 대응 실천강령을 마련함으로써 공동규제 체계를 형성하고 있다. 때문에 다음 장에서는 EU의 허위조작정보 대응 실천강령을 보다 자세히 살펴보도록 한다. 또한 시사점을 얻기 위한 비교분석 대상으로써 호주의 실천강령을 함께 검토할 것이다.

26) 이러한 맥락에서 학자들은 법·제도나 규제기관과 무관한 순수한 자율규제는 이상형으로서만 존재하며, 현실적으로 자율규제는 곧 공동규제로 이해될 필요가 있다고 주장한다(배진아 2017, 최규환, 2019, 황승흠 2014).

제4장 해외 허위조작정보 관련 자율규제 실천강령 사례

제1절 EU의 허위조작정보 대응 실천강령(2022)

1. 개관

다양한 허위조작정보들이 글로벌 플랫폼을 통해 빠르게 확산되면서 EU는 대형 플랫폼을 통한 허위조작정보 유통을 유럽의 민주주의를 위협하는 중대한 문제로 인식하고 이에 대한 대응방안을 마련하기 시작했다. 특히 자율규제 중심의 대응방안은 허위조작정보 대응 실천강령의 구축으로 나타난다. 동 강령의 발표는 2018년 1월 유럽의회(European Parliament)가 유럽집행위원회(European Commission)의 고위전문가그룹을 소집하여 <허위조작정보에 대한 다차원적 접근> 보고서를 발간하고, 이를 기반으로 일련의 자율규제 실천강령과 실행계획을 발표한 것에서 시작한다. 이와 관련된 주요 연혁은 다음과 같다(최민재·진민정·이현우, 2023.5.31.).

<표 4-1> 유럽집행위원회의 온라인 허위조작정보 관련 자율규제 정책 추진현황

구분	주요 내용
2018.1.	· 유럽의회의 요청으로 유럽집행위원회가 ‘고위전문가 그룹(High-level group of experts ; The HLEG)’ 소집
2018.3.	· 고위전문가 그룹, <허위조작정보에 대한 다차원적 접근(A multi-dimensional approach to disinformation)> 보고서 발간
2018.4.	· 유럽집행위원회, ‘온라인 허위조작정보 대응에 관한 정책안 : 유럽의 접근법(Communication on Tackling Online Disinformation : a European Approach)’ 발표
2018.9.	· 유럽집행위원회, ‘허위조작정보 대응 실천강령(Code of practice against disinformation)’ 공표
2018.12.	· 유럽집행위원회, ‘허위조작정보 대응 실행계획(Action plan against disinformation)’ 발표

2019.6.	· 유럽집행위원회, ‘허위조작정보 대응 실행계획의 이행에 관한 보고서 (Report on the implementation of the Action Plan against disinformation)’ 발표
2021.5.	· 유럽집행위원회, ‘강화된 허위조작정보에 대한 실천강령 지침(Guidance on Strengthening the Code of Practice on Disinformation)’ 발표
2022.6.	· 유럽집행위원회, ‘2022 강화된 허위조작정보에 대한 실천강령(The Strengthened Code of Practice on Disinformation 2022)’ 발표

* 자료 : 최민재·진민정·이현우 (2023.5.31). 유럽의 가짜뉴스 대응정책

2018년 3월 발표된 <허위조작정보에 대한 다차원적 접근> 보고서는 온라인 허위조작정보에 대응하기 위한 5가지 방안을 제시했다. 첫째, 온라인 뉴스의 투명성 향상. 둘째, 미디어 및 정보 리터러시 향상. 셋째, 이용자와 언론인이 허위조작정보를 다루고 진화하는 정보기술에 빠르게 적용할 수 있는 역량을 강화. 넷째, 언론 매체 생태계의 다양성과 지속가능성 보호. 다섯째, 허위조작정보의 영향에 대한 지속적인 연구 장려(김민정, 2020)이다.

이를 바탕으로 2018년 4월 유럽집행위원회는 「온라인 허위조작정보 대응에 관한 정책안(Communication on Tackling Online Disinformation : a European Approach)」을 발표했다. 정책안(Communication)은 법적 구속력은 없지만 일반적으로 정책평가, 실행방안의 설명, 향후 정책의 간략한 개요, 정책 현황에 대한 세부사항을 포함한다. 이는 온라인 허위조작정보 대응책에 대한 유럽연합집행위원회의 공식 입장을 처음으로 문서화한 것이라는 데 의미가 있다(김민정, 2020). 정책안에서 밝히고 있는 허위조작정보 대응을 위한 네 가지 원칙은 다음과 같다.

<표 4-2> 「온라인 허위조작정보 대응에 관한 정책안」의 4가지 원칙

원칙	주요 내용
투명성(transparency) 증대	· 정보의 출발점, 정보가 생산된 방식, 비용을 낸 주체, 유통 경로, 정보 생산의 목표가 되는 대상자를 투명하게 드러냄으로써, 시민들이 자신이 접하는 정보를 평가하고 의견을 조작하려는 시도들을 밝혀낼 수 있도록 함
정보의 다양성(diversity) 촉진	· 시민들이 비판적 사고능력을 활용하여, 충분한 정보에 기반한 결정을 내릴 수 있도록, 고품질 저널리즘과 미디어

	리터러시를 지원하고 정보 생산자와 배포자 간의 관계의 재조정을 도울 수 있게 함
정보의 신뢰도(credibility) 조성	<ul style="list-style-type: none"> • 정보가 믿을만한지를 보여주는 지표를 제공하고, 정보이력을 추적할 수 있도록 하고, 영향력 있는 정보 공급자를 인증함으로써 정보의 신뢰도를 높임
포괄적(inclusive) 해결안 마련	<ul style="list-style-type: none"> • 효과적으로 장기적인 해결안을 마련하기 위해서는 인식증진, 미디어 리터러시 증대, 광범위한 이해관계자의 참여, 그리고 공적 주체-온라인 플랫폼-허위조작정보신고자-언론인-미디어 그룹 간의 협력 필요

* 자료 : 김민정(2020). 유권연합의 온라인 허위조작정보 자율규제 사례 고찰

2018년 3월 발표된 <허위조작정보에 대한 다차원적 접근> 보고서는 단-중기적 대응방안으로 자율규제를 제안하고, 이를 위해 온라인 플랫폼-뉴스미디어기관-언론인-팩트체크의 주체-독립적인 콘텐츠 생산자-광고업계가 함께 참여해 실천강령을 제정하도록 요구했다. 이를 바탕으로 2018년 9월 유럽집행위원회는 ‘허위조작정보 대응 실천강령’을 발표했다. 실천강령은 온라인 플랫폼 사업자를 비롯해 다양한 이해관계자들이 모여 마련한 규약으로 자율규제 방안을 담고 있다. 구글, 트위터, 페이스북, 모질라 등 글로벌 플랫폼 기업과 함께 온라인 플랫폼 사업자의 연합인 EDIA, 광고업계와 광고주 관련 협회 등이 여기에 참여했다(김민정, 2020).

2018년에 마련된 실천강령은 5개 주제영역에 대한 15개 준칙(commitments)을 담고 있다. 다음은 이를 구체적인 표로 정리한 것이다. 강령에 참여한 기업은 매년 자사가 서명한 실천강령의 준수 실적에 대한 자체평가 보고서를 발간하고, 객관적인 제3자 기관에 의해 이의 준수여부를 검토 받아야 한다.

<표 4-3> 허위조작정보 대응 실천강령(2018년)의 15개 준칙

주제영역	준칙
광고 게재에 대한 정밀	1. 서명인은 허위조작정보 제공 행위에 대한 광고 혹은 금전적인 보상을 중단시키기 위한 정책과 절차 전개. 적절한 경우, 그러한 정

검토	<p>책과 절차는 다음을 포함할 수 있음</p> <ul style="list-style-type: none"> (a) 브랜드 안전 및 검증 도구 사용 장려 (b) 제3의 검증 업체의 개입 (c) 광고주가 미디어 구매 전략과 온라인 평판 위험을 평가할 수 있도록 지원 허용 (d) 광고주가 광고 게재를 모니터링하고 광고 게재 위치를 선택할 수 있도록 돕기 위해, 광고주에게 필요시 고객별 계정에 대한 접근 제공
정치광고 및 이슈광고	<p>2. 서명인은 EU 및 국가별 법률에 규정된 사항이자 자율규제 강령들에 명시된 사항인 모든 광고는 뉴스 등 언론사 자체 생산 내용과 명확하게 구분되는 형태로 제시되어야 함. 뉴스 또는 다른 편집된 내용을 포함하는 매체에 광고가 나올 때에는, 비용이 지불된 메시지라는 점이 쉽게 구분될 수 있는 방식으로 제시되거나 광고라는 점이 명시되어야 함</p> <p>3. 서명인은 정치광고의 공개가 가능하도록 약속하며, 그러한 공개에는 정치광고의 실제 후원자와 정치광고에 쓰인 금액이 포함될 수 있음. ‘정치광고’란 선거에서 특정 후보를 지지하거나 반대하는 광고 또는 국가와 유럽선거에서 국민투표(referenda)의 통과를 지지하거나 반대하는 광고를 가리킴</p> <p>4. 서명인은 이슈광고 공개 방안 마련을 위해 합리적 노력을 함. 그러한 노력은 이슈광고의 개념정의를 발전시켜나가는 작업을 포함하며, 이는 정치 토론에 대해 보도하거나 정치적 의견을 발간하는 일을 제한하지 않음. 상업광고는 여기서 제외됨. 표현의 자유 관련 함의들을 고려하여, 서명 주체들은 전문가들과의 협력을 통해 투명성 확보와 기본권 보장을 동시에 가능하게 하는 방안을 탐구하는 것을 장려함. 이슈광고를 정의해 나가는 작업은 광고 자율규제 기관들이 담당하는 영역을 방해해서는 안 됨</p>
서비스의 완전성	<p>5. 서명인은 자동화된 봇의 정체와 오용에 대해 명확한 정책을 수립하고 이러한 정책들을 EU 내에서 집행함</p>

	6. 서명인은 자동화된 시스템을 이용하는 방식들 중 허용되지 않는 이용 방식에 대한 정책을 마련하고, 이러한 정책을 플랫폼에 공개하고 EU 이용자들이 볼 수 있도록 함
소비자에게 권한/자율권 부여	7. 서명인은 사람들이 허위일 가능성이 있는 온라인뉴스를 접했을 때 정보를 바탕으로 한 결정을 내릴 수 있도록 돕기 위해, 뉴스 생태계와의 협력을 통해 진실성을 효과적으로 표시하는 방안을 개발하고 집행하는 노력들을 지원하는 것을 포함하여, 부록2에 언급된 것들과 같은 상품, 기술, 프로그램에 투자할 것을 약속함
	8. 서명인은 적절한 경우에는 검색, 피드(feeds), 여타 자동으로 순위가 매겨지는 배포 채널들에서 관련성 있고 정통하며 권위 있는 정보를 우선적으로 제시하는 기술적 수단들을 개발할 것을 약속함
	9. 서명인은 사람들이 공익 관련 주제에 대해 다양한 관점들을 찾는 것을 쉽게 만드는 기능들과 도구들을 개발할 것을 약속함
	10. 서명인은 시민사회, 정부, 교육기관, 여타 이해당사자들과의 협력을 통해 비판적 사고 능력 및 디지털미디어 리터러시를 향상시키기 위한 노력들을 지원할 것을 약속함
	11. 서명인은 소비자들이 자신들이 왜 특정 광고를 보게 되는지를 이해할 수 있도록 돕는 도구들의 시장 활용을 장려할 것을 약속함
연구 공동체에게 권한/자율권 부여	12. 서명인은 허위조작정보를 추적하고 그 영향력을 파악하려는 선의의 독립적 노력(이 노력에는 유럽집행위원회가 설립에 기여한 팩트체커들의 독립적인 네트워크가 포함됨)을 지원할 것을 약속함. 이러한 지원에는 개인정보보호가 보장된 상태의 데이터 공유, 공동 연구 수행, 관련성 있고 가능한 경우에 학계 및 시민사회단체들과의 협력 등을 포함함
	13. 서명인은 자신들의 플랫폼 상의 허위조작정보 및 정치광고에 대한 선의의 연구를 금지하거나 방해하지 않을 것을 약속함
	14. 서명인은 허위조작정보 및 정치광고에 대한 연구를 장려할 것을 약속함
	15. 서명인은 학계, 팩트체킹 공동체, 관련 구성원들 내에서의 논의를 장려하기 위한 연례행사를 소집할 것을 약속함

*출처: 김민정(2020). 유린연합의 온라인 허위조작정보 자율규제 사례 고찰

2. 2022 강화된 허위조작정보 대응 실천강령

(1) 배경

2022년 6월 16일 EU에서 발표한 ‘2022 강화된 허위조작정보 대응 실천강령(이하 “2022 강령”)’은 기존 ‘2018 허위조작정보 대응 실천강령(이하 “2018 강령”)’에 대한 수정·보완 및 현행화를 배경으로 한다.²⁷⁾ 2020년 12월, EU는 유럽 민주주의 시행 계획(EDAP)의 일환으로 기존에 있던 ‘2018 강령’에 대한 평가를 실시하였다. 그리고 다음 해인 2021년 5월 26일, EU는 ‘허위조작정보에 대한 강화된 실천강령에 관한 유럽위원회 지침’에서 해당 강령에 있어 보완이 필요한 사항들 발표했다. 주요 보완 사항으로 (i) 모니터링 강화, (ii) 투명성 강화, (iii) 참가가 확대에 대한 요구 등을 포함하고 있으며 (Killeen, 2021, 12, 7), EU에서 요약 제시한 내용은 아래와 같다.

〈표 4-4〉 ‘2018’ 강령에 대한 보완요청 사항

주요 내용
<ul style="list-style-type: none">· 맞춤형 준칙을 통한 참여 확대· 허위조작정보의 수익화 방지· 서비스의 완전성 보장· 이용자 역량강화 개선· 팩트체크 범위 확대· 연구자들의 데이터 접근성 향상· 강화된 모니터링 프레임워크 형성

* 자료 : European Commission (2021, 5, 26)

27) 여기에는 유럽을 포함한 세계 전역에 걸쳐있는 사회문화적 맥락도 존재하는데, 2020년 시작된 코로나19 팬데믹과 러시아-우크라이나 전쟁이 대표적이다(Collins, 2022, 6, 16).

(2) ‘2022 강령’의 목적 및 구성

허위조작정보(Disinformation)에 대한 더욱 강화된 대응을 위해 마련된 ‘2022 강령’은 궁극적으로 유럽의 열린 민주사회 구현을 목적으로 한다. 해당 목적 달성의 주요 과제로 허위조작정보에 주목하고, 이에 대응하기 위한 구체적인 방안을 제시하는 것이다. 더 상세하게는 시민들이 허위조작정보에 덜 노출되도록 함으로써 이들이 충분한 정보를 갖춘 상태(well-informed)에서 자유롭고 공정한 정치적 절차를 통한 의사 표현을 할 수 있는 공개 토론(public debate)의 장을 마련했을 때, 더 나은 민주주의 사회가 형성될 것을 기대하고 있다(European Commission, 2022, 6, 16).

‘2018 강령’과 마찬가지로, ‘2022 강령’ 또한 기본권(표현의 자유, 정보 및 사생활의 자유 등)을 충분히 보장하는 상태에서 해당 강령 이행이 이뤄져야 함을 강조한다. 또한 EU 및 회원국의 기존 법적 체계²⁸⁾ 내에서 적용되어야 함을 명시하고 있는데, 이는 해당 체계를 대체하거나 해석하는 방식에서의 이용을 금지함을 의미한다.

‘2022년 강령’은 (서문을 제외하고) 9개의 주제영역으로 구성된다. 9개의 주제영역은 ①메시지 관련(광고 게재 자체 및 수익화 방지, 정치/이슈 광고), ②시스템 관련(기술 기반 대응), ③다양한 이해관계자들과의 협력(이용자, 팩트체크 기관, 연구 공동체), 그리고 ④본 강령의 효과적인 운영(투명성 센터, 상설 태스크포스, 모니터링)의 차원으로 분류해볼 수 있다. 그리고 다시, 상기 9개의 주제영역은 44개의 준칙(Commitment)과 128개의 조치(Measure)로 구성되어 있다. 각 조치는 해당 강령의 서명인들의 보고 지침이 되는 질적보고요소(QRE)와 서비스수준지표(SLI)를 포함하고 있다. 해당 강령의 주요 내용을 주제영역 및 준칙을 중심으로 정리하면 다음과 같다.

28) ‘2022 강령’에서 제시하고 있는 기존 법률 체계 예시는 다음과 같다: EU 기본권 헌장, 유럽인권조약, 전자상거래지침, 개인정보보호규정, e-프라이버시지침, 불공정 상업 관행에 관한 지침, 허위광고 및 비교 광고에 관한 지침, 시청각미디어서비스지침(수정안 포함), 영업비밀지침, 온라인 플랫폼 규정, 그 외 유럽사법재판소 및 유럽인권재판소의 판례 등(European Commission, 2022, 6, 16).

<표 4-5> '2022 강령' 의 주요 내용

주제영역 ²⁹⁾	소주제	준칙 ³⁰⁾
II. 광고 게재에 대한 정밀검토	· 허위조작정보의 수익화 방지	1. 허위조작정보 유포 관련 재정 지원 금지 및 관련 정책·제도 개선
	· 허위조작정보 포함 광고에 대응	2. 광고 시스템이 광고 메시지 형태의 허위 조작정보 유통으로 오용되는 것을 방지
	· 관련 당사자들과의 협력	3. 광고 게재 정밀검토 효과를 높이기 위한 (특히 온라인 수익화 관련) 당사자들과의 협력 강화
III. 정치광고	· 정치/이슈광고에 대한 공통된 이해	4. 정치/이슈광고에 대한 공통된 정의 채택 5. 정치/이슈광고에 대한 일관된 접근 방식 및 해당 광고의 허용·금지 범위 표시
	· 정치/이슈광고에 대한 효과적인 표기	6. 정치/이슈광고에 대한 명확한 표기(라벨링)
	· 정치/이슈광고에 대한 검증 약속	7. 정치/이슈광고 게시자(스폰서 및 광고대행 사)를 위한 검증 시스템 마련
	· 이용자에 대한 정치 /이슈광고의 투명성 약속	8. 정치/이슈광고 관련 투명한 정보 제공 9. 정치/이슈광고가 표시되는 이유에 대한 정보 제공
	· 저장소 및 데이터에 접근할 수 있는 API 를 위한 최소한의 기능	10. 정치/이슈광고 저장소 유지 및 품질 보장 11. 정치/이슈광고 내 이용자·연구자가 이용 가능한 맞춤형 검색 기능 제공
	· 시민사회와의 약속	12. 정치/이슈광고 감독 강화
	· 지속적인 협업	13. 정치/이슈광고의 허위조작정보 관련 위 험을 파악하고 대응하기 위한 지속적인 모니터링 및 연구에 참여
IV. 서비스의 완전성	· 허용되지 않는 조작 행위에 대한 공통된 이해	14. 허용되지 않는 조작 행위·관행 제한을 위해 관련 행위(AMITT 기술, TTPs 등) 에 대한 공통된 이해·합의 및 주기적 검 토와 현행화

주제영역 ²⁹⁾	소주제	준칙 ³⁰⁾
	· AI 시스템에 대한 투명성 의무	15. AI 관련 서비스 제공 시, EU의 AI Act의 투명성 의무 및 금지 행위 고려
	· 협력 및 투명성	16. 플랫폼 간, 플랫폼 내에서 이뤄지는 허위조작정보 관련 사건 대응을 위한 정보 공유 및 협력
V. 이용자 역량강화	· 미디어리터러시 향상	17. 미디어 리터러시 및 비판적 사고 영역에 대한 노력 강화
	· 서비스 시스템의 ‘안전한 설계’, 투명한 정책, 추천 시스템의 책임성	18. 안전한 설계를 통한 허위조작정보의 바이럴 전파 위험의 최소화
		19. 추천 시스템 관련 기준 및 매개변수에 대한 투명한 공개 및 관련 옵션 정보 제공
		20. 디지털 콘텐츠의 출처·편집 이력·진위성·정확성 등에 대한 평가 도구 제공
	· 이용자 허위조작정보 식별 역량강화	21. 이용자들의 허위조작정보 식별을 위한 방안 제공·개선
	· 허위/오해의 소지가 있는 유해 정보를 표시하는 기능	22. 정보 식별을 위한 충분한 정보 및 신뢰성 평가 도구 제공
	· 투명한 이의제기 메커니즘	23. 서비스의 약관을 위반한 허위/오해의 소지가 있는 유해 정보 표시 기능 제공
· 메시징 앱 내 허위 조작정보 차단을 위한 조치	24. 정책 위반 조치 대상 조치 사실 고지와 이의제기 기회 제공 및 투명하고 타당한 처리	
VI. 연구 공동체 역량강화	· 허위조작정보 연구를 위한 데이터 공개 및 접근	25. 이용자들의 수신정보에 대한 비판적 사고 역량 강화 및 정보의 정확성 판단을 돕는 기능 및 이니셔티브를 구축·시행
		26. 허위조작정보 관련 연구수행을 위한 안정적인 데이터 접근성 제공

주제영역 ²⁹⁾	소주제	준칙 ³⁰⁾
	<ul style="list-style-type: none"> · 추가 정밀검토가 필요한 연구 목적의 데이터 접근을 위한 거버넌스 구조 · 연구진과의 협력 · 연구조직의 투명성과 데이터 공유 	<p>27. 연구자 및 연구 제안을 심사할 독립적인 제3의 기관 운영 및 심사 통과 연구자에게 필요 데이터에 대한 접근성 제공</p> <p>28. 허위조작정보 관련 선의의 연구 지원</p> <p>29. 투명한 방법론과 윤리적 기준에 따른 연구 수행 및 공유</p>
VII. 팩트체킹 공동체 역량강화	<ul style="list-style-type: none"> · 팩트체킹 공동체와의 협력 · 서비스 내 팩트체킹 이용 및 통합 · 팩트체커의 관련정보에 대한 접근 · 팩트체커의 기준 	<p>30. 팩트체커 지원을 위한 프레임워크 구축</p> <p>31. 팩트체커의 작업을 플랫폼 서비스 내 통합·제시·활용</p> <p>32. 팩트체커에게 팩트체킹의 품질과 영향력 제고에 도움이 될 정보에 대한 접근 제공</p> <p>33. 엄격한 윤리·투명성 규칙에 따른 운영 및 독립성 보호</p>
VIII. 투명성 센터	-	<p>34. 서명한 공동 투명성 센터 웹사이트 구축 및 유지</p> <p>35. 본 강령의 이행 관련 모든 정보를 이해하기 쉽고, 쉽게 검색 가능한 방식으로 제공</p> <p>36. 투명성 센터 제공 정보 업데이트</p>
IX. 상설 태스크포스 ³¹⁾	-	<p>37. 상설 태스크포스의 구성(서명인, EDMO 및 ERGA 대표, EEAS 대표 등) 및 운영 방식(의장은 유럽 집행위원회이며, 읍저버로 관련 전문가 초청 가능, 합의제로 결정)</p>
X. 본 강령의 모니터링	-	<p>38. 준칙 이행 보장을 위한 재정·인적 자원 투입 및 프로세스 마련</p>

주제영역 ²⁹⁾	소주제	준칙 ³⁰⁾
		39. 서명 6개월 이내 이행 종료 후, 1개월 이내 유럽 집행위원회에 기본(baseline) 보고서 제출
		40. 서비스수준지표(SLI) 및 질적보고요소(QRE)에 대한 정기적 보고
		41. 본 강령의 효과 측정을 위해 서명 9개월 이내 구조적 지표 세트 공개 및 정식(full) 보고서 첫 측정치 공개와 지속적인 이행 및 개선
		42. 선거나 위기 등 특정 상황에서 집행위원회의 요청이 있을 시, 신속 대응 시스템에 따라 비정기 특별(ad-hoc specific) 보고서 제출
		43. 보고, 또는 데이터 제공시 합의한 보고 형식 및 방법론 준수
		44. 초대형 온라인 플랫폼 서명인의 경우, DSA와의 제휴 모색 및 본 강령 준수 여부에 대한 별도의 외부 감사 수행

* 자료 : European Commission (2022, 6, 16)에서 재구성

(3) 허위조작정보의 범위 확대

‘2022 강령’은 직전 강령에서 정의하고 있는 허위조작정보 외에도 잘못된 정보(misinformation), 정보 영향력 작전(information influence operations), 정보 공간 내 해외 개입(foreign interference in the information space) 등을 대응 대상으로 포함하고 있다. 이

29) ‘I’은 서문에 해당

30) 준칙에 대한 정확하고 상세한 내용은 부록 참조

31) 태스크포스(TF)는 정규 조직이 수행하기 어려운 일을 효율적으로 해결하기 위해 임시로 편성되는 조직으로, 구성원들에게는 전문적인 지식이나 능력이 요구된다.

는 허위조작정보에만 주목했던 ‘2018 강령’에서 그 범위를 확장한 것으로 볼 수 있다. 각 정의는 유럽 민주주의 시행 계획(EDA)를 따르고 있으며, 상세한 내용은 다음과 같다.

<표 4-6> ‘2022 강령’ 허위조작정보 영역 및 정의

구분	내용
허위조작정보	· 경제적 또는 정치적 이득을 취하거나 확보하려는 의도로 확산되어 대중에게 해를 끼칠 수 있는 허위 또는 오해의 소지가 있는 콘텐츠
잘못된 정보	· 선의로 친구나 가족에게 거짓 정보를 공유하는 경우와 같이 그 효과가 여전히 해로울 수 있지만 해로운 의도 없이 공유되는 허위 또는 오해의 소지가 있는 콘텐츠
정보 영향력 작전	· 허위조작정보와 함께 독립적인 정보 출처를 억압하는 등 다양한 기만적 수단을 사용하여 이용자 대상에게 영향을 미치기 위한 국내 또는 해외 행위자의 조직적인 노력
정보 공간 내 해외 개입	· 광범위한 혼합 작전의 일환으로 수행되며, 외국 국적 행위자 또는 그 대리인이 개인의 자유로운 정치적 의사 형성과 표현을 방해하려는 강압적이고 기만적인 노력

*자료 : European Commission (2022, 6, 16)에서 재구성

(4) 협력 및 운영 강화를 위한 주제영역의 신설

‘2022 강령’은 직전 강령 대비 4개의 주제영역을 신설하였다. 신설 주제영역은 참여자의 확대와 모니터링 강화 등 효과적인 강령 운영과 관련한다. 먼저 허위조작정보의 유해성이 날로 심각해지는 상황에서 플랫폼 기업 외에도 팩트체크 관련 조직을 참여시킴으로써 효과적인 대응을 위한 적절한 지원을 받고 동시에 해당 강령의 영향력을 확장하고자 하였다. 가령 플랫폼 기업에서 팩트체크 기관의 작업을 활용하고 반대로 팩트체크 기관은 플랫폼 내 정보, 또는 재정적 지원을 받는 방식의 협력이 이뤄진다(Tanner, 2022, 8, 5). 실제로 프랑스, 스페인, 덴마크 등 회원국 소재 팩트체크 비영리 조직들이 이 강령에 참여 중이다.

<표 4-7> '2018 강령' 과 '2022 강령' 의 주제영역 비교

2018 강령	2022 강령
광고 게재에 대한 정밀검토	광고 게재에 대한 정밀검토
정치광고 및 이슈광고	정치 광고
서비스의 완전성	서비스의 완전성
소비자 역량강화	이용자 역량강화 ³²⁾
연구 공동체 역량강화	연구 공동체 역량강화
-	팩트체킹 공동체 역량강화
	투명성 센터
	상설 태스크포스
	본 강령의 모니터링

*자료 : European Commission (2018, 9, 26; 2022, 6, 16)에서 재구성

다음으로 해당 강령의 효과적인 운영을 위해 투명성 센터, 상설 태스크포스, 그리고 모니터링 체계를 구축하기 위한 지침을 신설했다. 먼저 투명성 센터는 공동의 웹사이트를 구축해 서명인들의 준칙 이행과 관련한 정보를 공개함으로써 서명인들의 허위조작정보 대응에 대한 투명성을 보장한다. 상설 태스크포스의 경우, 허위조작정보에 대한 서명인들의 공동 대응, 즉 효율적인 교류와 협력을 위한 공간이다. 끝으로 모니터링 체계는 서명인들의 이행 보장 및 강령 자체의 영향력을 평가하는 것을 목적으로 한다.

(5) 구체적 방안 제시를 통한 지침 강화

'2018 강령' 은 비교적 광범위한 정의나 모호한 표현을 따르고 있어 서명인들을 특정 조치 이행으로 이끄는 데 덜 효과적으로 기능했다면, '2022 강령' 은 구체적이고 명확한 내용으로 재구성되어 더욱 적극적인 허위조작정보 대응이 가능해졌다(Tanner, 2022, 8, 5). 가령 허위조작정보의 수익화 방지를 목적으로 제안된 양 강령의 '준칙 1' 을 보면, 먼저

32) 허위조작정보의 대상이 되는 대중들에 대해 '소비자(Consumers)' 에서 '이용자(Users)' 로 용어를 달리한 점 또한 주목해볼 만하다(Tenner, 2022, 8, 5).

‘2018 강령’의 경우 수익화 방지를 위한 ‘정책과 절차를 수행한다’라고 명기하고 있는 반면, ‘2022 강령’의 경우 수익 관련 지원을 막는 것뿐만 아니라 수익화가 가능한 콘텐츠를 판단하고, 광고 배치 관련 통제 및 정책 이행의 효과를 측정하는 등 보다 구체적인 지침을 제공하고 있다(European Commission, 2018, 9, 26; 2022, 6, 16).

<표 4-8> ‘2018 강령’ 과 ‘2022 강령’ 의 준칙¹ 비교

2018 강령	2022 강령
<ul style="list-style-type: none"> · 관련 서명인은, 자기 자신이나 자기 소유 재산의 목적에 관한 중요한 정보를 부실 표시하는 등의 관련 행위에 대해서 금전적인 보수의 지불이나 광고를 금지하는 정책과 절차를 수행한다. 	<ul style="list-style-type: none"> · 광고 게재에 참여하는 관련 서명인은 허위조작정보 유포 관련 수익 지원을 막고, 수익화할 콘텐츠의 적격성, 수익화 및 광고 배치에 대한 통제, 그리고 광고 배치 관련 통제 및 서비스의 정확성과 효과에 대해 보고할 데이터를 결정하는 정책 및 제도 개선을 위해 노력한다.

*자료 : European Commission (2018, 9, 26; 2022, 6, 16)에서 재구성

구체적인 지침을 제시하는 것은 정치광고 관련 내용에서도 확인할 수 있다. ‘2018 강령’의 경우, 정치 광고임을 공개한다거나 알아챌 수 있는 방식으로 제시되어야 한다는 등의 포괄적인 형태의 준칙을 제시하고 있는 반면(European Commission, 2018, 9, 26), ‘2022 강령’은 정치 광고, 또는 이슈 광고에 대해 식별할 수 있는 ‘명확한 라벨링’을 할 것을 요청한다(European Commission, 2022, 6, 16).

(6) 디지털 서비스법과 연계한 자율규제의 규제 기능 강화

해당 강령은 자율규제인 만큼 철회와 서명이 비교적 자유롭다. 그럼에도 ‘2022 강령’에서 특별히 특기할 만한 사항은 이 강령이 디지털 서비스법과의 공동규제 수단으로 역할을 함으로써 그 권한이 일정 부분 더 강화되었다는 점이다(Collins, 2022, 6, 16; Killeen, 2021, 12, 7). 2023년 4월 확정되고 8월부터 시행 중인 디지털 서비스법은 디지털플랫폼에

더 많은 책임을 부여하기 위한 목적으로 고안되었고, 해당 법률에 의거해 월간 활성 이용자가 4,500만 명 이상인 초대형 온라인 플랫폼을 대상으로 동법의 규정을 이행하지 않을 시 전 세계 수익의 최대 6%에 해당하는 벌금을 부과할 수 있다(Collins, 2022, 6, 16; Gerken, & McMahon, 2022, 6, 16). 여기서 디지털 서비스법 규정의 미이행이란 규정에 대한 반복적인 위반과 위협 발생 시 이를 완화하기 위한 조치를 수행하지 않은 경우를 포함한다(Trueman, 2022, 7, 17).

하지만 사업자가 허위조작정보 대응 실천강령을 준수할 경우 이는 적절한 위험완화조치로 간주될 수 있다. 또, 사업자들이 이러한 강령에 참여하라는 위원회의 권유를 적절한 설명 없이 거부할 경우, 해당 온라인 플랫폼 혹은 검색엔진 제공자가 디지털 서비스법의 규정의 의무를 위반했는지 여부를 판단할 때 이를 고려할 수 있다.

(7) 참여자 확대

‘2018 강령’ 대비 ‘2022 강령’의 가장 큰 차이는 참여자들의 확대이다. 기존 강령의 경우, 구글, 트위터, 페이스북 등 초대형 플랫폼 기업들만을 참여대상으로 삼았다(Killeen, 2021, 12, 7). 그러나 최근 강령에는 플랫폼 기업 외에도 광고, 팩트체크, 언론, 연구조직 등 여러 참여 단체들이 해당 강령에 서명하고 참여했다. 해당 강령이 발표된 2022년 6월에는 34개(Trueman, 2022, 6, 17), 2023년 5월 당시에는 38개(최민재 등, 2023), 가장 최근인 같은 해 12월 현재는 44개의 기업 또는 단체가 참여 중이다(European Commission, 2023, 12, 10).

<표 4-9> ‘2022 강령’ 서명인 현황

(2023.12.10 현재)

구분	서명인
1	ActiveFence: 온라인 검열·보안 등 기술 개발 기업
2	Adobe: 소프트웨어 기업
3	AI Forensics: 알고리즘 조사 관련 비영리조직
4	Alliance4Europe: 시민을 위한 유럽 건설을 목표로 하는 비영리조직

구분	서명인
5	Avaaz: 글로벌 행동주의 목표로 하는 국제시민단체
6	Clubhouse: 음성 기반 소셜미디어
7	Crisp: 온라인 안전·보안 기술개발 기업
8	Debunk EU: 온라인 허위조작정보 대응 기술 관련 비정부조직
9	Demagog Association: 폴란드 팩트체크 조직
10	Democracy Reporting International(DRI): 전세계 민주주의 촉진 독립조직
11	DoubleVerify: 디지털미디어 측정 소프트웨어 개발 및 분석 자료 제공 기업
12	DOT Europe: 유럽 선도 인터넷 기업 관련 유럽 무역협회
13	Ebiquity: 마케팅·미디어 컨설팅 기업
14	European Factchecking Standards Network: 유럽 팩트체크 표준 네트워크
15	European Association of Communication Agencies(EACA): 유럽 커뮤니케이션 에이전시 협회
16	Faktograf: 팩트체크에 주목하는 비영리 미디어 조직
17	Globsec: 글로벌 안보 포럼 조직 비영리 단체
18	Google: 초대형 플랫폼 기업
19	IAB(Interactive Advertising Bureau) Europe: 온라인 광고산업 법적지원 관련 비영리 조직
20	Kinzen: 유해 콘텐츠 식별 기술 개발 기업
21	Kreativitet & Kommunikation: 덴마크 소재 미디어·커뮤니케이션 관련 비영리 조직
22	Logically: 유해 허위조작정보 식별 기업
23	Maldita.es: 스페인 팩트체크 비영리조직
24	MediaMath: 타겟팅 광고 등 마케팅 운영 시스템 개발 기업
25	Meta: 초대형 플랫폼 기업
26	Microsoft: 대규모 IT 기업
27	Neeva: 개인정보보호 강화 검색엔진 서비스 개발 기업
28	Newsback: 미국 소재 뉴스 데이터베이스 정보 운영 기업
29	Newtral: 스페인 팩트체크 기관
30	NewsGuard: 뉴스 및 웹사이트 등급 시스템

구분	서명인
31	PagellaPolitica: 이탈리아 소재 팩트체킹 등 허위조작정보 대응 조직
32	Reporters without Borders(RSF): 국경 없는 기자회견
33	ScienceFeedback: 프랑스 팩트체킹 비영리 조직
34	Seznam: 체코 소재 검색 엔진 서비스 개발 기업
35	The Bright App: 건강 관련 소셜 미디어 앱 개발 기업
36	The Daily Ledger: 미국 소재 일간지 발행 언론사
37	The GARM(Global Alliance for Responsible Media) Initiative: 책임있는 미디어를 위한 글로벌 연합
38	The Global Disinformation Index: 허위조작정보 대응 비영리조직
39	TikTok: 초대형 플랫폼 기업
40	Twitch: 동영상 스트리밍 플랫폼
41	Vimeo: 동영상 스트리밍 플랫폼
42	VOST Europe: 디지털 자원봉사 관련 비영리조직
43	WhoTargetsMe: 전세계 디지털 정치광고 조사 비영리조직
44	World Federation of Advertisers(WFA): 세계 광고주 연맹

* 자료 : European Commission (2023, 12, 10)에서 재구성

3. 온라인 플랫폼 기업의 실천 강령 이행

(1) 실천강령 보고서의 주요 내용 및 특징

2018년 강령에 서명한 기업들은 준수실적에 대한 자체평가 보고서를 발간해왔다. 그러나 구체적으로 어떤 체계와 지표를 갖고 작성되어야 하는지에 대한 기준이 없어 공시가 효율적으로 이루어지지 않았다. 보고서 내 실천 강령의 이행 조치와 결과에 대한 정보의 양과 질이 기업마다 상이하며, 일부 기업은 이행 결과를 충분히 제공하지 않는다는 비판이 제기되었고(심나리, 2019), 자율규제의 권리를 부여하더라도 실천 강령 이행에 있어 명

확한 기준 적용과 심사 과정에서의 투명성이 확보되어야 한다는 목소리가 대두되었다(김민정, 2020). 이에 따라 EU는 2022년 강화된 실천 강령을 발표함과 동시에, 강령에 담긴 9가지 주제 영역과 44개의 준칙, 128개의 조치로 구성된 보고 서식을 일원화했다. 각 기업이 정치 광고 등의 투명성을 높이고 서비스 완전성을 강화하기 위한 노력을 지속해왔음은 분명하지만, 기업의 조치가 실제 EU의 허위조작정보 확산을 줄이는데 어떻게 기여했는지에 대한 정확한 평가를 위해 통일된 지표가 필요하다는 취지다.

실천 강령 이행 보고서는 강령의 본문과 마찬가지로 ▲광고 게재 위치 조사, ▲정치 광고, ▲서비스 완전성, ▲이용자 역량 강화, ▲연구 역량 강화, ▲팩트체크 커뮤니티 역량 강화, ▲투명성 센터, ▲상설 태스크포스, ▲강령 모니터링 등 9개의 주제 영역에 걸쳐 44개의 준칙이 하위 항목으로 제시되어있다. 각 준칙과 128개의 조치는 영역별 서비스 수준을 정량적으로 파악하기 위한 측정 지표인 SLI(Service Level Indicators)와 질 적합성 지표인 QRE(Qualitative Reporting Elements)에 따라 평가된다.

한편, EU는 현재진행형으로 전개되는 급박하고 구체적인 위기에 대해서는 별도의 조치 사항을 공시하도록 했다. EU는 2023년 가장 큰 위기를 코로나-19 팬데믹(Covid-19 pandemic)과 러시아-우크라이나 전쟁(War of aggression by Russia on Ukraine)으로 채택했고, 각 기업은 보고서 말미에 두 위기 국면을 둘러싼 허위조작정보 조치 결과를 서술하고 있다.

실천 강령에 서명한 기업은 2022년 실천 강령 공표 이후 1개월이 경과한 시점에 SLI 및 QRE 기준 보고서(Baseline Report)를 제출하고, 서명 후 7개월 이내에 강령 준칙(commitments) 이행 결과를 상세히 설명하는 보고서를 제출하기로 약속했다. 나아가 EU는 보고서 공시 주기를 1년에서 6개월로 단축하고, 각 기업이 웹사이트 형식의 ‘투명성 센터’에 보고서를 공개하도록 의무화했다. 각 기업의 보고서는 EU의 투명성 센터³³⁾에서도 확인할 수 있으며, EU 또한 각 기업의 이행 결과에 대한 분석 및 평가 보고서를 발표하고 있다. 뿐만 아니라 기업은 ‘상설 태스크포스’ 웹페이지를 통해 준칙과 법안을 이행하는 과정에서 변경된 서비스 약관과 정책을 고지해야한다. 이용자는 각 기업의 태스크포스 웹페이지를 통해 강령의 구현과 정책 변경, 도구 및 리소스 개발 등의 결정 사항을 추적할 수 있다.

33) <https://disinfocode.eu/>

2022년 개정된 실천 강령이 2018년의 강령과 구별되는 특징 중 하나는 디지털서비스법(Digital Services Act, DSA)의 적용을 받는다는 것이다. 따라서 EU는 이행 사항을 준수하지 않는 기업에 과징금을 부과하고 반복적으로 위반한 기업은 유럽 시장에서 퇴출할 수 있다는 대응 방침 발표했다. 이 같은 강력한 조치에 대해 EU는 정보의 투명성 및 신뢰도, 포괄적 해결안 마련이라는 실천 강령의 기초가 반영된 결과이며 과도한 정보 통제가 아니라 자율규제의 실효성과 집행 가능성을 높이기 위한 방침임을 밝혔다.

<표 4-10> 강화된 허위조작정보 대응 실천 강령 이행 보고서 구조

주제영역(9개) 및 준칙(commitment, 44개)		
① Scrutiny of Ad Placements 광고 게재 위치 조사(C1-3)	④ Empowering Users 이용자 역량 강화(C17-25)	⑦ Transparency Centre 투명성 센터(C34-36)
② Political Advertising 정치 광고(C4-13)	⑤ Empowering the research 연구 역량 강화(C26-29)	⑧ Permanent Task-Force 상설 태스크포스(C37)
③ Integrity of Services 서비스 완전성(C14-16)	⑥ Empowering the fact-checking community 팩트체크 커뮤니티 역량 강화 (C30-33)	⑨ Monitoring of Code 강령 모니터링(C38-44)
Reporting on the service's response during a period of crisis 위기 국면에 대한 서비스 대응 보고(2023년 기준)		Measuring indicator 평가지표
Covid-19 pandemic 코로나-19 팬데믹	War of aggression by Russia on Ukraine 러시아-우크라이나 전쟁	SLI, 서비스 수준 지표 QRE, 질적 보고 요소

[그림 4-1] 실천 강령 이행 보고서 서식(예)

III. Political Advertising						
Commitment 6						
Relevant Signatories commit to make political or issue ads clearly labelled and distinguishable as paid-for content in a way that allows users to understand that the content displayed contains political or issue advertising.						
We signed up to the following measures of this commitment:	CS	M 6.1	M 6.2	M 6.3	M 6.4	M 6.5
	Facebook Instagram	Facebook Instagram	Facebook Instagram	Facebook Instagram	Facebook Instagram	Messenger
Measure 6.2	Facebook	Instagram				
QRE 6.2.1	As noted in our baseline report Ads about social issues, elections or politics require authorizations and a "Paid for by" disclaimer.		As noted in our baseline report Ads about social issues, elections or politics require authorizations and a "Paid for by" disclaimer.			
Measure 31.1	Facebook	Instagram				
Measure 31.2	Facebook	Instagram				
QRE 31.1.1	As mentioned in our baseline report, when content has been rated by fact-checkers (as outlined in detail under QRF 21.13), We take action to (i) label it and (2) ensure less people see it. We also take action against accounts that repeatedly share misinformation." The current warning implies that any content flagged by a SPFC counts toward Repeat Offender.		As mentioned in our baseline report, when content has been rated by fact-checkers (as outlined in detail under QRF 21.13), We take action to (i) label it and (2) ensure less people see it. We also take action against accounts that repeatedly share misinformation." The current warning implies that any content flagged by a SPFC counts toward Repeat Offender.			
	Regarding rating AI-generated content. Fact-checkers may rate AI-generated media under our fact-checking program policies. They often rely on AI experts, visual techniques, and meta data analysis to aid in the detection of this content.		Regarding rating AI-generated content. Fact-checkers may rate AI-generated media under our fact-checking program policies. They often rely on AI experts, visual techniques, and meta data analysis to aid in the detection of this content.			
SLI 6.2.1 - numbers for actions enforcing policies above	Number of unique SRF ads on Facebook and Instagram combined displaying "paid for by" disclaimers from 01/01/2023 to 30/06/2023 in EU member states.	Number of unique SRF ads on Facebook and Instagram combined displaying "paid for by" disclaimers from 01/01/2023 to 30/06/2023 in EU member states.				
	Country determined by inferred advertiser location at time of enforcement.	Country determined by inferred advertiser location at time of enforcement.				
	Number of ads accepted & labelled on Facebook and Instagram combined	Number of ads accepted & labelled on Facebook and Instagram combined				
Member States						
Austria	Over 290,00	Over 29,000				
Belgium	Over 260,00	Over 26,000				
Bulgaria	Over 8,300	Over 8,300				
Croatia	Over 10,000	Over 10,000				
Cyprus	Over 3,600	Over 3,600				
Czech Republic	Over 16,000	Over 16,000				
Denmark	Over 18,000	Over 18,000				
Estonia	Over 7,800	Over 7,800				
Finland	Over 35,000	Over 35,000				
Reporting on the service's response during a period of crisis						
Covid-19 pandemic						
Misinformation related to COVID-19 has presented unique risks to public health and safety over the last years. We know that people looking to mislead others - whether through phishing, scams, or influence operations - try to leverage crises in order to advance their goals, and the COVID-19 pandemic is no different. We are actively working to find and stop coordinated campaigns that seek to manipulate public debate across our platforms.						
Our approach focused on:						
<ul style="list-style-type: none"> - Debunking false claims about COVID-19 through our global fact-checking program and partnerships, including through a dedicated Coronavirus grant program with the IFCN - Providing access to authoritative and reliable information about the pandemic - Removing misinformation which could directly contribute to imminent harm to public health and security, as determined in consultation with leading health organisations - Supporting media literacy across Europe - Identifying, disrupting, and removing coordinated inauthentic behaviour networks 						
Moving forward, we will take a more tailored approach to our Covid-19 misinformation rules consistent with the Oversight Board's guidance and our existing policies. In countries that have a Covid-19 public health emergency declaration, we will continue to remove content for violating our Covid-19 misinformation policies given the risk of imminent physical harm. We are consulting with health experts to understand which claims and categories of misinformation could continue to pose this risk. Our Covid-19 misinformation rules will no longer be in effect globally as the global public health emergency declaration that triggered those rules has been lifted						

주제 영역
및 준칙

실천 강령
이행 결과
및 평가
(QRE)

실천 강령
이행 결과
및 평가
(SLI)

위기 국면
에 대한
서비스
대응보고

* 자료 : Meta (2023. 7). Code of Practice on Disinformation- Report of Meta for the period 01 January 2023 to 30 June 2023.

다음은 구글(Google)과 메타(Meta)에서 발표한 보고서의 자율규제 및 성과를 발췌한 내용이다. 2022년 강화된 실천 강령 발표 이후 1개월 내 작성된 이행 기준 보고서와 그 후

6개월이 지난 시점에 발표된 정기 보고서(2023. 1. 1 ~ 2023. 6. 30)의 내용을 중심으로 두 기업의 실천 강령 이행 결과를 검토했다. 2023년 하반기 보고서는 2024년 초에 발표될 예정이며, 각 기업은 하반기 보고서에 강령 이행에 대한 추가 통찰력 및 2024년 EU 선거를 앞두고 허위 정보 확산을 줄이기 위한 조치 계획을 포함하기로 했다.

(2) 구글(Google)

① 광고 게재 위치조사

구글 애드센스(Google AdSense)는 허위 사실을 유포하는 광고의 수익 창출을 차단하기 위한 다양한 정책과 지침을 공시하고 있다. 우선 애드센스는 ‘신뢰할 수 없고 유해한 주장(unreliable and harmful claims)’, ‘복제된 콘텐츠(replicated content)’, ‘조작된 미디어(manipulated media)’, ‘위험하거나 비하하는 콘텐츠(dangerous or derogatory content)’ 등 차제 심의 기준을 위반한 광고의 게재 및 수익화를 금지한다. 광고 퍼블리셔(ad publisher)가 신원이나 중요한 세부 정보를 숨기거나 허위로 표시하는 행위, 불분명한 명목 또는 기만적인 주장을 통해 이용자가 광고 콘텐츠에 참여하도록 유도하는 행위, 이용자가 개인 정보를 공유하도록 유도하는 부정행위도 엄격히 규제한다. 구글은 인공지능 기술을 활용한 자동 모니터링 시스템을 통해 애드센스 정책을 준수하지 않는 콘텐츠나 퍼블리셔에 즉각 조치를 취하며, 필요한 경우 사람이 직접 처리하는 정밀 검토가 이뤄진다(Google, 2023. 1).

한편 구글은 세계광고주연맹(World Federation of Advertisers, WFA) 및 전미광고주협회(Association of National Advertisers, ANA)가 지원하는 이니셔티브 ‘글로벌 책임 미디어 연합(Global Alliance for Responsible Media, 이하 GARM)’을 창립하여 허위 광고 및 유해 콘텐츠를 삭제하기 위한 공동 프레임워크를 구축하고(Google, 2023. 7), 2023년 9월에는 광고주가 허위조작정보와 관련된 정책을 이해할 수 있도록 GARM 통합 측정 보고서 1~5권을 집필했다(4A's, 2023. 9. 22). 이 외에도 구글은 미디어 등급 위원회(Media Rating Council, MRC)가 실시하는 연례 감사를 받아 브랜드 안전성을 유지하고 있으며, 범죄 활동을 방지하고 디지털 광고업계의 신뢰를 높이기 위한 글로벌 인증 프로그램 태그

(Trustworthy Accountability Group, TAG)의 인증을 획득하는 등 정기적으로 독립 기관의 검증을 거쳐 애드센스 정책의 유효성을 입증하고 있다(Google, 2023. 7).

<표 4-11> 구글의 게재 및 수익 창출 불가 광고 유형과 정의

광고유형	정의
신뢰할 수 없고 유해한 주장	<ul style="list-style-type: none"> · 명백히 허위이며 선거 또는 민주적 절차에 대한 참여나 신뢰를 크게 훼손할 수 있는 주장을 하는 콘텐츠 · 건강 위기 및 기후 변화에 관한 권위 있는 과학적 합의에 반하는 콘텐츠
복제된 콘텐츠	<ul style="list-style-type: none"> · 독창적인 콘텐츠나 새로운 가치를 추가하지 않고 다른 콘텐츠를 복사하거나 삽입한 경우 · 콘텐츠의 직접 검토 및 재구성 없이 자동 생성된 콘텐츠
조작된 미디어	<ul style="list-style-type: none"> · 정치, 사회적 이슈 또는 대중의 관심사와 관련된 조작된 미디어를 통해 사용자를 기만하는 콘텐츠
위험하거나 비하하는 콘텐츠	<ul style="list-style-type: none"> · 개인이나 그룹에 대한 증오를 선동하거나, 차별을 조장하거나, 특정 개인이나 그룹을 근거로 비방하는 콘텐츠
기만적인 광고 행위	<ul style="list-style-type: none"> · 광고 퍼블리셔가 자신의 신원이나 중요한 세부 정보를 숨기거나 허위로 표시하는 행위 · 불분명한 명목 또는 기만적인 주장을 통해 이용자가 콘텐츠에 참여하도록 유도하는 행위 · 이용자가 광고 개인 정보를 공유하도록 유도하는 행위

* 자료 : Google (2023. 1). Code of Practice on Disinformation - Report of Google for the period 1 July 2022 - 30 September 2022.

② 정치 광고

정치 광고는 기업용 광고 캠페인 관리 도구인 구글 애즈(Google Ads)³⁴⁾ 및 구글 DV360(Google Display & Video 360)³⁵⁾의 적용을 받는다. 구글은 현직 선출직 공직자 또는 후보자를 포함한 광고에 대해 광고주의 신원 증명 및 거주 지역에 대한 정보 제공을 의무화하여 그 지역 또는 국가에만 광고가 게재될 수 있도록 제한해왔다. 그러나 2022년 2월부터 신원 확인된 광고주는 다른 국가 또는 지역에 기반한 선거 광고를 자국 또는 지역에 계

34) <https://ads.google.com>

35) <https://marketingplatform.google.com/intl/ko/about/display-video-360/>

재할 수 있게 되었다. 선거 광고 업체가 운영하는 모든 선거 광고에는 광고비용을 누가 지불했는지 식별할 수 있는 정보가 포함되어야 하며, 구글 애즈와 DV360은 신원 인증 과정에서 획득한 정보를 활용하여 ‘광고비 지불 출처(paid-for-by)’ 라벨을 자동으로 생성한다(Google, 2023. 1). 구글은 2023년 11월부터 AI로 생성한 정치 광고 콘텐츠에 ‘이 영상은 컴퓨터가 생성했습니다’ 또는 ‘이 이미지는 실제 사건을 묘사하지 않습니다’와 같은 내용을 눈에 띄는 위치에 표시하도록 의무화했다(AI타임즈, 2023. 9. 7).

구글은 선거 광고 검증 외에도 ‘투명성 센터’³⁶⁾를 통해 구글에 (비)활성화된 정치 광고 관련 정보를 제공하고 있다. 투명성 센터에는 집행된 광고 수, 집행 금액, 광고 노출 시기, 도달률, 연령·성별·위치별 타겟팅 방법 등이 공개된다. 정치 광고 관련 정보는 광고가 처음 게재된 후 몇 분 이내에 업데이트되며 다운로드도 가능하다. 이용자는 API 플랫폼을 통해 정보를 원하는 순서로 정렬하고 효과적으로 분석할 수 있도록 자료를 시각화할 수 있다.

[그림 4-2] 구글 정치 광고 API 플랫폼



36) <https://adstransparency.google.com/>

광고 통계

후보자 이름으로 광고 검색 게재 횟수 지출 금액 게재 횟수: 내림차순

VOX

Paremos su destrucción

VOX

Ayúdanos a echarle

VOX

Vota lo que importa, vota campo.

Este 23 de julio vota VOX si quieres defender el mundo rural.

VOX

1000만 이상
시작일: 2023년 7월 12일 ~ 21일(10...)
€1.5만~€2만 지출

Europa

1000만 이상
시작일: 2022년 9월 15일 ~ 23일(9일)
€2.5만~€3만 지출

정책 위반으로 인해 삭제됨

이 광고를 표시할 수 없습니다.

Google 광고 정책

The Independent Group (TIG) L...
0~1천회 노출됨
시작일: 2019년 5월 21일 ~ 23일(3일)
€0~€50 지출

정책 위반으로 인해 삭제됨

이 광고를 표시할 수 없습니다.

Google 광고 정책

The Independent Group (TIG) L...
0~1천회 노출됨
시작일: 2019년 5월 20일 ~ 23일(4일)
€0~€50 지출

정책 위반으로 인해 삭제됨

이 광고를 표시할 수 없습니다.

Google 광고 정책

The Independent Group (TIG) L...
0~1천회 노출됨
시작일: 2019년 5월 20일 ~ 23일(3일)
€0~€50 지출

정책 위반으로 인해 삭제됨

이 광고를 표시할 수 없습니다.

Google 광고 정책

The Independent Group (TIG) L...
0~1천회 노출됨
시작일: 2019년 5월 20일 ~ 23일(4일)
€0~€50 지출

개요

예 게재됨, 전체 기간

광고비 지출액 광고

€43,227,850 (EUR) 314,059

지속적으로 업데이트되며 몇 시간 지연될 수 있습니다

*자료: Google Transparency Center (<https://adstransparency.google.com/?authuser=0®ion=KR>)

③ 서비스 완전성

EU가 AI 시스템, 특히 딥페이크 및 조작된 미디어(가짜 계정, 증폭 봇 계정, 스팸 계정 등)를 근절하기 위한 협력을 강화함에 따라, 구글은 인공지능 기반 스팸 방지 시스템인 스팸브레인(SpamBrain)의 품질을 대폭 향상했다. 스팸브레인을 탑재한 구글 검색 알고리즘

은 여러 가지 요소와 신호를 정교하게 검토하여 품질이 낮은 콘텐츠와 순위 조작을 줄이는 데 기여하고 있다. 2022년 구글은 스팸을 생성하는 사이트뿐만 아니라 다른 스팸 사이트의 접속을 유도하는 링크도 추적할 수 있도록 감지 시스템을 강화했으며, 그 결과 스팸 차단에 나서기 시작한 2000년대 중반에 비해 200배, 2021년에 비해 5배 많은 스팸 사이트를 적발했다고 밝혔다(Google, 2023. 1). 2022년 기준 이용자 99%가 사이트 방문 시 어떤 결과가 표시되는지 미리 확인했으며, 2021년과 비교했을 때 스팸 사이트에 대한 클릭이 50%로 감소한 것으로 나타났다. 시스템 강화 이후, 구글은 2023년 1월부터 6월까지 전 세계 10,674,755개 스팸 도메인을 적발했다(Google, 2023. 7). 구글은 2023년 10월부터 각 국가의 검색 결과가 개선될 수 있도록 스팸 감지 언어를 모든 언어로 확대했다(InterAd, 2023. 4. 12).

유튜브 또한 자동화된 스팸 감지 시스템 구축에 대대적으로 투자하고 있다. 동시에 다양한 배경과 언어, 경험 수준을 가진 모더레이터 수천 명이 알고리즘에 새로운 스팸 데이터를 학습시키는 등 위반 가능성이 있는 콘텐츠를 정확하게 식별하기 위한 공동의 노력을 지속하고 있다. 허위조작정보 규제가 표현의 자유에 대한 과도한 침해로 변질되는 것을 방지하기 위해 여러 개의 상위 그룹이 모더레이터가 콘텐츠를 정확하게 평가했는지 그 맥락을 살핀다. 나아가 모더레이터의 공동된 의견이 일정 수준 수렴됐을 때만 알고리즘에 반영하기 때문에 스팸 감지 기술이 작동되기까지는 몇 달이 걸린다(Google, 2023. 1). EU 회원국에 한해 2023년 1-6월 기준 유튜브 허위조작정보 정책 위반으로 삭제된 동영상 수는 총 28,789건으로 나타났다(Google, 2023. 7).

구글의 태그(TAG)와 신뢰 및 안전팀(Trust & Safety)은 엄격한 분석에 기반해 집행 결정을 내리는 부서다. 태그는 기술적 신호, 휴리스틱 및 행동 패턴을 조사하여 조직적인 비인종 행위를 평가·판단하며, 정기적으로 태그 게시판(TAG bulletin board)을 발행하여 서비스 무결성 캠페인에 대한 업데이트와 추가적인 조치 사항을 고시한다. 신뢰 및 안전 리더십에 참여하는 품질 보증 책임자(각국의 콘텐츠 검토 팀을 감독하는 책임자)는 알고리즘 학습과 관련한 사안을 결정하고 스팸 감지 가이드라인을 개발·수정하고 있다. 한편 구글은 EU의 강령 이행 결과를 발표하는 분기별 보고서 외에도 ‘AI 원칙(AI principles)’³⁷⁾을

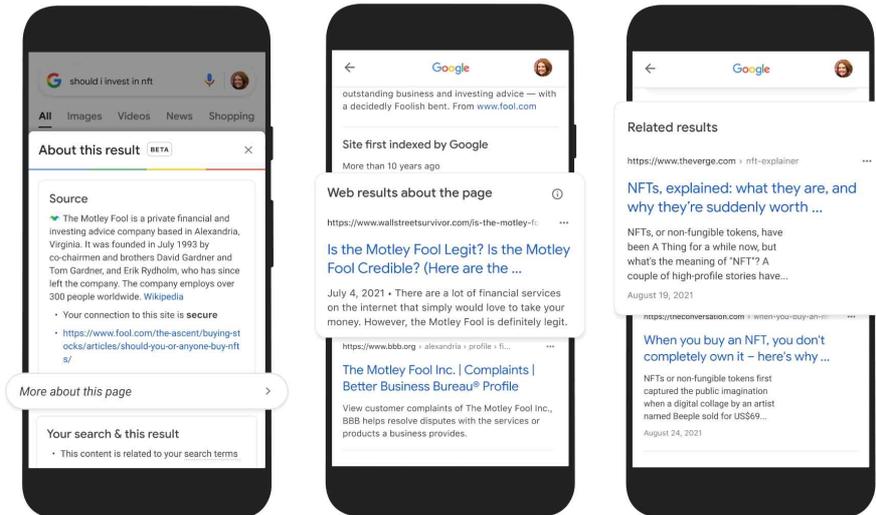
37) <https://ai.google/responsibility/principles/>

배포하여 책임감 있게 기술을 개발하겠다는 의지를 표명하고 있다. 2023년 1월에는 제4차 연례 AI 원칙 추진 사항을 발표했다(Google, 2023. 1).

④ 사용자 역량 강화

기본적으로 구글은 이용자가 검색 결과의 주제 및 출처에 대한 세부 정보와 맥락을 파악할 수 있도록 돕고 있다. 이용자는 검색 결과 또는 유튜브 영상 상단에 있는 점 3개를 클릭하면, 해당 콘텐츠에 대한 정보를 자세히 살펴볼 수 있다. 이는 이용자가 온라인정보를 인식하고 신고하는 등 허위조작정보에 스스로 대응할 수 있도록 돕기 위한 조치이다.

[그림 4-3] 구글의 웹페이지 자세히 알아보기 기능



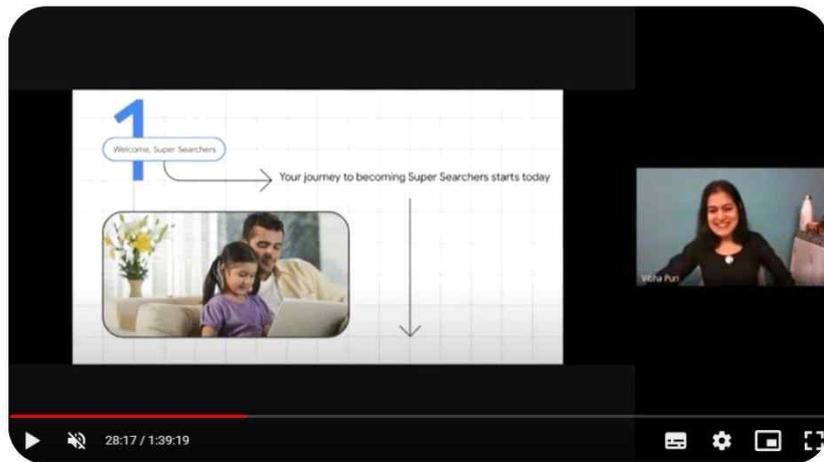
*자료 : Fonearena (2021. 9. 30). Google Search introduces Multitask Unified Model (MUM) with advanced searching capabilities.

‘슈퍼 검색자(Super Searchers)’는 사서를 매개로 대중의 정보 활용 역량을 강화하기 위해 마련된 캠페인이다. 시민들은 고품질 정보를 제공하는 기관으로서 도서관이 수행하

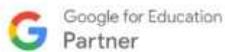
는 중요한 역할을 인식하고 있으며, 자신의 니즈에 맞는 정보를 찾기 위해 도서관을 찾는다. 그러나 많은 이들이 정보 검색에 어려움을 겪고 웹사이트에 허위조작정보가 포함되어 있는 경우 진위를 파악하기는 더욱 어려워진다. 이때 권위 있는 전문가로서 사서가 이용자를 대신하여 정보를 탐색하고 팩트체크를 수행함으로써 허위조작정보에 대응하는 것이 ‘슈퍼 검색자’의 목적이다. 구글은 사서의 효과적인 정보 검색 및 비판적인 평가 역량을 제고하기 위해 정기 교육을 실시하고 있다. 벨기에 브뤼셀에서 최초로 시작된 이 캠페인은 2023년 6월까지 아일랜드, 이탈리아, 포르투갈에서 사서 교육을 실시했으며, 저변 확대를 위해 공공도서관 2030(Public Libraries 2030)과 협력하고 있다.

[그림 4-4] 구글의 ‘슈퍼 검색자’ 캠페인

Super Searchers



Increasing information literacy



*자료: Public Libraries 2030 (<https://publiclibraries2030.eu/2022/10/super-searchers-launch/>)

유튜브 ‘잠시만요(Hit Pause)’ 또한 구글의 대표적인 미디어 리터러시 캠페인이다. 이 캠페인은 재밌는 영상을 통해 이용자들이 플랫폼을 더욱 안전하고 책임감 있는 방식으로 활용하도록 지원하는 것을 목표로 한다. 2022년 9월 미국을 시작으로 프랑스·독일·인도·호주 등 50여 개국에서 ‘잠시만요’ 캠페인을 선보이고, 2023년 7월에는 모든 EEA 회원국에서 캠페인이 실시되었다(Google, 2023. 7). 2023년 9월 한국에 론칭된 캠페인은 ‘사이버불링(cyberbullying)’에 초점을 맞춘 8편의 영상이 게재되었으며³⁸⁾, 이 캠페인에는 인플루언서 엔조이커플, 원샷한솔, 너덜트, 짹짹아윤기, 잇섭, 닥터프렌즈, 피지컬갤러리가 각자의 개성을 담아 사이버불링의 도구로 사용된 허위 정보 사례와 표현, 예방법을 소개했다(미디어오늘, 2023. 9. 6). 캠페인 영상은 각국의 ‘잠시만요’ 유튜브 채널을 통해 누구나 시청 가능하다.

[그림 4-5] 구글 유튜브의 ‘잠시만요’ 캠페인



Hit Pause

@HitPause · 구독자 145만명 · 동영상 179개

채널 자세히 알아보기 >

구독

* 자료 : Hit Pause (<https://www.youtube.com/@HitPause>)

38) <https://www.youtube.com/playlist?list=PLL0amEZ5bJuEcsuBS7iCiHTk8N87TkDhS>

⑤ 연구 역량 강화

YouTube는 2022년 7월 연구자 프로그램(YouTube Researcher Program)을 시작하며 연구자들과의 협력을 확대했다. 이 프로그램은 연구자에게 실시간 데이터(지난 7일간의 샘플 데이터)와 비실시간 데이터(2004년부터 검색하기 최대 72시간 전까지 탑재된 데이터) 및 데이터를 효과적으로 분석할 수 있는 API 도구를 제공한다. API를 통해 제공되는 정보는 동영상 및 채널 제목과 설명, 조회수, 좋아요 등이다. 이 프로그램에 참여하기 위해서 연구자는 고등 교육기관 이메일(예: .edu su x)로 신청하여 자신의 소속을 인증하고, 자신의 가능한 많은 정보를 제공해 신원을 확인받아야 한다. 이후 구글 클라우드 콘솔에서 API 프로젝트를 생성하면 구글 담당 부서의 검토를 받는다.

한편 구글은 양질의 저널리즘과 글로벌 미디어 리터러시를 지원하기 위해 언론사 및 언론인과 협력하고 있다. 구글 뉴스 이니셔티브는 2015년부터 10만 명 이상의 유럽 언론인에게 허위조작정보 검증 방법을 포함한 다양한 교육을 제공했으며, 이니셔티브가 개발한 무료 온라인 강의는 40만 회 이상 재생되었다. 이 밖에도 구글은 선거 기간과 뉴스 속보 상황에서 오보와 허위 정보를 막기 위해 하버드 케네디 스쿨의 퍼스트 드래프트(First Draft)와 함께 ‘디스인포 랩(Disinfo Lab)’ 프로젝트를 시작했다. 소비자가 온라인에서 사실과 거짓을 구분하는 데 도움을 주기 위해 스탠포드 대학의 포인터 인스티튜트(Poynter Institute) 및 비영리 기관인 로컬 미디어 협회(Local Media Association)를 지원해 ‘미디어와이즈(MediaWise)’ 프로젝트를 운영하기도 했다. 미디어와이즈는 어린이 소비자의 디지털 정보 리터러시 향상을 위해 디자인된 프로젝트이다.

⑥ 팩트체크 커뮤니티 역량 강화

구글은 미디어 리터러시 기술 강화, 오보 방지 및 팩트체크 지원을 위한 유럽 미디어 및 정보 기금(European Media and Information Fund, EMIF)을 출범하기 위해 2,500만 유로를 기부했다. 프로젝트 목록에는 허위조작 정보와 관련한 연구 및 조사에 대한 8개의 특정 프로젝트 기금 193만 유로가 포함되어 있으며, 이 기금은 2021~2026년에 걸쳐 집행될 예정이다. 구글은 국제 팩트체크 네트워크(International Fact-Checking Network, IFCN)와 협력

하여 135개 이상의 단체에 직접 지불하겠다고 밝혔다. 2023년 6월까지 최소 33개 프로젝트에 560만 유로가 지원되었으며, 구글 투명성 센터에서 기금 수혜자 목록도 확인할 수 있다.

(2) 메타(Meta)

① 광고 게재 위치 조사

메타는 온라인 모니터링 기업 제프르(Zefr)와 협력해 광고 게재 위치와 콘텐츠 적합성을 측정하는 AI 기반 솔루션을 개발했다. 제프르는 페이스북·인스타그램 피드 게재 위치 전반에서 광고가 표시되는 맥락과 모니터링하고, GARM의 표준에 따라 콘텐츠의 안전성과 적합성을 분석해 광고주에게 보고한다(Meta, 2023. 1). 2023년 3월 영어권 시장을 중심으로 테스트를 마쳤으며, 2023년 8월에는 스페인어, 프랑스어, 아랍어, 중국어 및 포르투갈어 사용 국가로 서비스를 확대했다(메가시티뉴스, 2023. 8. 1).

메타 콘텐츠 수익 창출 정책(Content Monetization Policies)은 수익 창출이 불가능한 콘텐츠 항목을 제시하고 있다. 주요 제재 항목으로, 메타는 ‘타사 팩트 체크 기관에서 거짓으로 평가된 콘텐츠(ads that include content debunked by third-party fact-checkers)’가 포함된 광고를 금지하고 있으며, 거짓으로 정보를 반복적으로 게시하는 광고주는 메타 플랫폼에서 광고할 수 있는 자격을 박탈한다. 거짓 정보는 구체적으로 ‘오해의 소지가 있는 의료 정보(misleading medical information)’, ‘기만적인 콘텐츠(deceptive content)’, ‘위험한 콘텐츠(dangerous content)’, ‘선정적인 콘텐츠(sensationalist content)’가 해당된다. 이 외에도 메타는 게재 불가 및 수익 창출 제재를 받는 광고의 유형을 광범위하게 기술하고 있다. 광고주가 메타의 콘텐츠 검토 절차 및 집행 조치를 회피하는 것으로 확인되거나 입증되지 않은 계정주가 광고를 게재했을 경우 또한 광고가 제한되거나 계정이 삭제된다. 메타는 2023년 상반기(2023년 1월-6월)까지 페이스북과 인스타그램에 게재된 콘텐츠에 ‘광고비 지불 출처(paid for by)’를 고지했으며, 약 6,900,000개의 정책 위반 광고를 삭제했다. 이 중 약 24,000개 이상의 광고가 허위조작정보로 알려졌다(Meta, 2023. 7).

② 정치 광고

메타는 ‘사회 문제, 선거 또는 정치 관련 광고(ads about Social Issues, Elections or Politics, 이하 SIEP)’³⁹⁾를 통해 EU 내 정치인, 정당에서 게재하는 광고 및 사회적 이슈를 언급하는 광고의 광고주는 승인 절차를 거쳐 광고를 게재할 수 있도록 조치하고 있다. SIEP에도 광고비 지불 출처 라벨이 부착되어야 하며, 광고주의 신원 확인은 사진이 부착된 유효한 신분증과 공인된 문서 확인 등 2단계를 거쳐 최초 1회 이루어진다. 신원 확인은 허위조작정보를 차단하기 위한 조치이기도 하지만, 사칭이나 ID 도용과 같은 위험을 예방하여 개인의 민감한 정보를 보호하는 데에도 사용된다. 앞서 구글과 마찬가지로, 메타 또한 연결된 SIEP 계정이 광고주가 타겟팅하려는 국가에 주소지를 가져야 하며, 광고비 또한 현지 통화로 결제되어야 한다. 특정 국가 또는 지역을 타겟팅 하는 광고주가 다른 국가 또는 지역에 거주하는 경우 승인 절차를 거친다. EU 회원국에서 2023년 상반기동안 SIEP 정책 미준수로 인해 거부된 광고 건수는 총 68,000건으로, 국가별 최소 1,800건(슬로베니아)에서 최대 10,000건(이탈리아)으로 집계된다(Meta, 2023. 7).

광고 라이브러리(Ad Library)는 메타 플랫폼에 게재 중인 광고를 검색할 수 있는 데이터베이스로, 7년 동안 보관된 정치 광고 아카이브와 광고를 심층적으로 분석할 수 있는 API를 제공한다. 구체적으로, 광고에 자금을 지원한 사람, 지출한 금액 범위, 여러 인구 통계에 걸친 광고 도달 범위, 광고 라이브러리에서 가장 자주 검색된 키워드 문구 등에 대한 추가 정보를 검색할 수 있다. 2023년 7월 메타는 이용자가 최대 10개의 SIEP 광고주를 선택해 일일 또는 주간 광고비 지출 금액과 광고 수 등의 개요를 받아볼 수 있는 메일링 서비스를 시작했다.

39) <https://transparency.fb.com/en-gb/policies/ad-standards/SIEP-advertising/SIEP/>

[그림 4-6] 메타 광고 라이브러리 API 플랫폼

광고 라이브러리

현재 Meta 테크놀로지에 게재 중인 모든 광고 및 다음을 검색해보세요:

- 최근 7년 동안 게재된 사회 문제, 선거 또는 정치 관련 광고
- 최근 1년 동안 유럽 연합에 게재된 광고

광고를 찾으려면 키워드 또는 광고주로 검색하세요.

[새로운 기능 보기](#)



광고 검색

위치를 설정하고 광고 카테고리를 선택하여 검색을 시작하세요. 검색 팁을 확인하세요.

프랑스 ▼ 광고 카테고리 ▼ 🔍 광고 카테고리 선택

기타 도구 둘러보기

광고 라이브러리 보고서 또는 광고 라이브러리 API를 사용하여 더욱 자세한 광고 정보를 찾을 수도 있습니다.



정보 다운로드 및 지출 내역 추적

사회 문제, 선거 또는 정치 관련 광고의 경우 전체적인 총 지출 금액과 광고 주, 위치별 지출에 관한 상세 정보를 보려면 광고 라이브러리 보고서를 사용하세요.

보고서가 필요한 사람
사회 문제, 선거 또는 정치와 관련된 광고의 데이터를 빠르게 확인하고 필터링하고 다운로드하려는 사람.

[보고서로 이동](#)

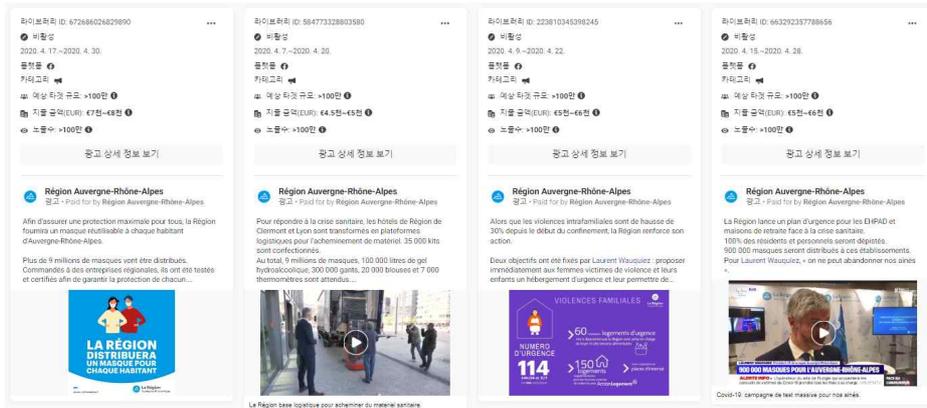
검색 직접 설정

광고 라이브러리 API를 사용하여 광고의 예상 타겟 규모, 국가, 언어 등의 매개변수로 맞춤 검색을 시작해보세요.

API가 필요한 사람
프로그래밍 작동 방식에 관해 약간의 지식이 있는 사람들. API에 액세스하려면 몇 가지 단계를 완료해야 합니다.

[API로 이동](#)





* 자료 : Meta 광고 라이브러리 (https://www.facebook.com/ads/library/active_status=all&ad_type=political_and_issue_ads&country=FR&media_type=all)

③ 서비스 완전성

메타는 허위조작정보를 대응함에 있어 스펀지플라주(spamouflage, 스팸과 위장의 합성어)라 불리는 가짜계정에 대한 단속을 최우선 목표로 설정하고 있다. 가짜계정 집중 단속이 본격적으로 시작된 2016년 이래 메타는 15억 개의 가짜계정을 적발했으며, 이 중 99.6%가 사전에 발견되었다(Meta, 2023. 1). 2023년 1분기에는 4억 2,600만 개의 가짜계정이, 2023년 2분기에는 6억 7,600만 개의 가짜계정이 삭제되었다(Meta, 2023. 7). 2023년 8월에 발표된 ‘분기별 적대적 위협 보고서(quarterly adversarial threat report)’에 따르면, 메타는 3분기에 중국 정부가 배후에 있는 것으로 추정되는 페이스북 계정 7,704개와 페이지 954개, 인스타그램 계정 15개를 삭제했다(전자신문, 2023. 8. 30). 이들 계정은 친중 메시지를 전파하는 데 중점을 두고 운영됐다. 또 중국 정책을 비판하는 사람을 공격하거나, 미국 내 여론 분열을 조장하려는 콘텐츠도 발견되었다. 콘텐츠는 영어뿐만 아니라 한국어, 독일어, 프랑스어, 러시아어, 태국어, 웨일스어 등으로 번역된 것으로 알려졌다(국민일보, 2023. 8. 30).

페이스북은 허위 신원을 사용해 사용자를 오도하는 ‘조직화된 비인증 행위(coordinated inauthentic behavior, CIB)’의 배후와 네트워크를 지속적으로 단속하고 관련 계정과 페이지,

콘텐츠를 삭제하고 있다. 콘텐츠 내용에 대한 조치를 넘어 부정행위 정책을 위반한 특정 IP, 집단적 행위 및 사용 패턴에 근거하여 제재하고 있으며, 정교하고 새로운 공격 행위에 대응하기 위해 전문 수사관과 업계파트너, 연구자, 시민 사회 및 정부와 협력하여 방어책을 마련하고 있다. 가짜계정과 콘텐츠는 몇 분 이내에 비활성화된다(Meta, 2023. 1; 2023. 7).

또한 메타는 특정 기준에 따라 딥페이크로 판단되는 동영상 콘텐츠에도 즉각 제재를 취한다. 동영상이 이용자가 알아볼 수 없는 방식으로 편집된 경우, 사진 합성이나 음성조작으로 이용자가 동영상 속 인물이 하지 않은 말을 했다고 오해할 가능성이 있는 경우, 동영상이 AI 또는 머신러닝의 결과물인 경우가 이에 해당된다. 메타는 조작된 미디어를 식별하기 위한 전용 AI 모델과 시스템을 갖추고 있으며, 미시간 주립대학교와 협력하여 딥페이크 탐지 기술을 고도화하고 있다(Meta, 2023. 1).

메타의 서비스 완전성 타임라인은 허위조작정보의 위험을 지속적으로 평가하고 정책, 도구 및 프로세스를 조정하는 자사의 노력을 시계열적으로 확인할 수 있는 페이지다. 경우에 따라 문제의 소지가 있지만 메타가 보유한 시스템이 1차적으로 감지하지 못해 메타의 정책 위반이 성립되지 않는 콘텐츠가 있다. 메타는 이를 경계선 콘텐츠(borderline content)라 명명하고 콘텐츠가 오해의 소지가 있는지, 기만적인지, 유해한지 등 게시물을 검토한 후 이러한 유형의 콘텐츠를 자동으로 식별하도록 메타 시스템을 학습시킨다. 이 같은 경계선 콘텐츠는 그것을 둘러싼 맥락을 보다 세밀하게 살펴봐야 한다. 따라서 적절한 주제 전문가와 언어 전문성을 갖춘 팀이 해당 콘텐츠를 검토하게 되고, 필요한 경우 글로벌 운영 또는 주제 전문 팩트체커에게 콘텐츠를 전달하기도 한다. 이러한 전 과정을 담은 것이 서비스 완전성 타임라인이며, 여기에는 메타 정책을 위반하는 콘텐츠와 허위 계정에 대처하기 위한 신기술 개발 및 서비스 활성화, 정책 변경 사항 등도 확인할 수 있다(Meta, 2023. 1).

[그림 4-7] 메타 서비스 완전성을 위한 연간 조치 타임라인

홈 → 정책 → Meta의 개선 방법

Integrity timeline

업데이트됨 MAY 2, 2023

This timeline provides a comprehensive list of our public announcements regarding integrity. It includes product improvements, policy expansions, external engagement, and transparency reports.

This is a working page that will be updated regularly.

April 2023

- Welcoming Teens to Meta Horizon Worlds in the US and Canada. [Read more](#)
- Digital Suraksha Summit: How Meta Is Creating an Open and Safe Internet. [Read more](#)
- How Meta is Preparing for Türkiye's 2023 Elections. [Read more](#)

March 2023

- How Meta is Advancing Democracy. [Read more](#)
- An Update on Meta's Civil Rights Progress. [Read more](#)
- Introducing a More Inclusive Dataset to Measure Fairness. [Read more](#)

February 2023

- New Updates to Help Prevent the Spread of Young People's Intimate Images Online. [Read more](#)
- Protecting Public Debate During the War in Ukraine and Protests in Iran. [Read more](#)
- Meta's Adversarial Threat Report, Fourth Quarter 2022. [Read more](#)
- Integrity and Transparency Reports, Fourth Quarter 2022. [Read more](#)
- How We're Improving Facebook's Penalty System. [Read more](#)

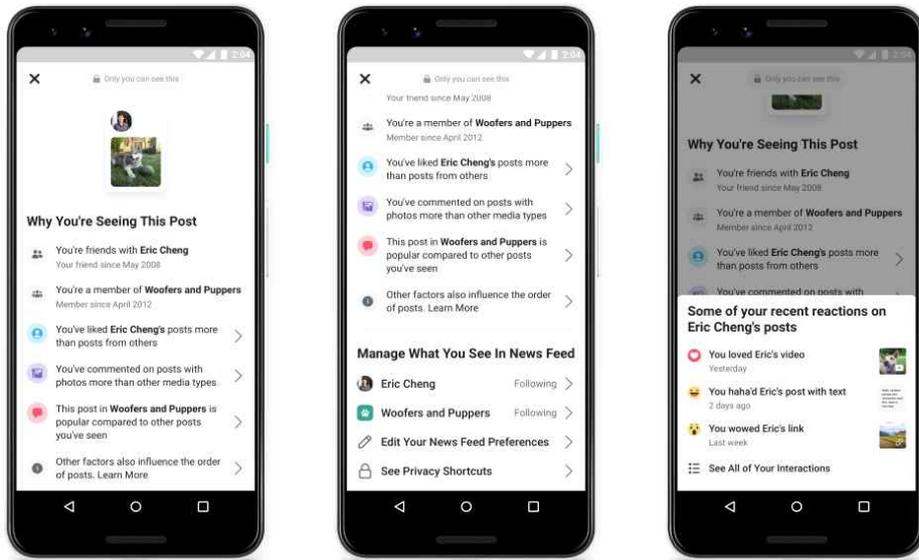
*자료: Meta Integrity timeline (<https://transparency.fb.com/ko-kr/policies/improving/timeline/>)

④ 사용자 역량 강화

메타는 이용자가 스스로 정보를 비판적으로 평가하고 올바른 결정을 내릴 수 있는 도구를 제공하고 있다. 2023년 2월에 업데이트된 “이 광고가 표시되는 이유는 무엇인가요?(Why am I seeing this ad?)” 도구는 이용자에게 게시물에 대한 추가적인 맥락과 정보를 제

공하기 위해 마련되었다(Meta, 2023. 7).

[그림 4-8] 메타(페이스북) 게시물 표시 이유 공개



*자료 : Theverge (2019. 4. 1). Facebook's News Feed is starting to explain itself.

이 같은 조치는 콘텐츠 노출 순위와 빈도가 플랫폼 기업의 알고리즘에 의해 결정되며 사용자 개인에게는 제어 권한이 없다는 오해를 바로잡기 위한 노력으로, 기술 안팎에서 이용자의 활동이 알고리즘에 어떤 영향을 미치는지에 대한 정보가 주제별로 요약되어 있다⁴⁰). 가령 이용자가 좋아하는 페이스북 페이지나 클릭했던 광고, 반응을 보였던 콘텐츠 이력이나 인터넷·휴대폰을 사용한 위치 정보, 광고의 경우 광고주가 설정한 타겟 정보 등이 제공된다. 메타는 이 같은 기술적 투명성을 유지하면서 기술적인 지식이 없는 사람들도 쉽게 이해할 수 있도록 보다 자세한 정보를 제공할 방침이다. 한편, 이용자는 피드에 인기 게시물 또는 최신 게시물이 먼저 노출되도록 조정할 수 있으며, 특정 콘텐츠나 광고에 대해 ‘더 보기’, ‘적게 보기’를 선택하여 자신이 선호하거나 선호하지 않는 게시

40) <https://about.fb.com/news/2019/07/understand-why-youre-seeing-ads/>

물 유형을 메타에 알릴 수 있다(Meta, 2023. 7). 이는 이용자가 메타 플랫폼에서 접하는 콘텐츠에 대해 더 많은 정보를 얻고 그에 입각한 결정을 내릴 수 있도록 지원하는 방법으로서 유의미하다.

⑤ 연구 역량 강화

메타는 이용자의 개인정보를 침해하지 않는 선에서, 연구자에게 자사의 데이터를 수집하고 분석할 수 있는 기회를 제공하고 있다. 2023년 1월 기준 전 세계적으로 5,000개 이상의 학술 계정이 메타가 인수한 통계 분석 플랫폼인 크라우드 탱글(Crowd Tangle)의 서비스를 이용했다. 크라우드 탱글은 허위 정보 확산을 모니터링하는 용도로 활용되고 있으며, 프랑스 대통령 선거 당시 허위 정보를 유포해온 프랑스 페이스북 계정 3천여 개의 활동을 중지시키는데 크게 기여했다(Meta, 2023. 1).

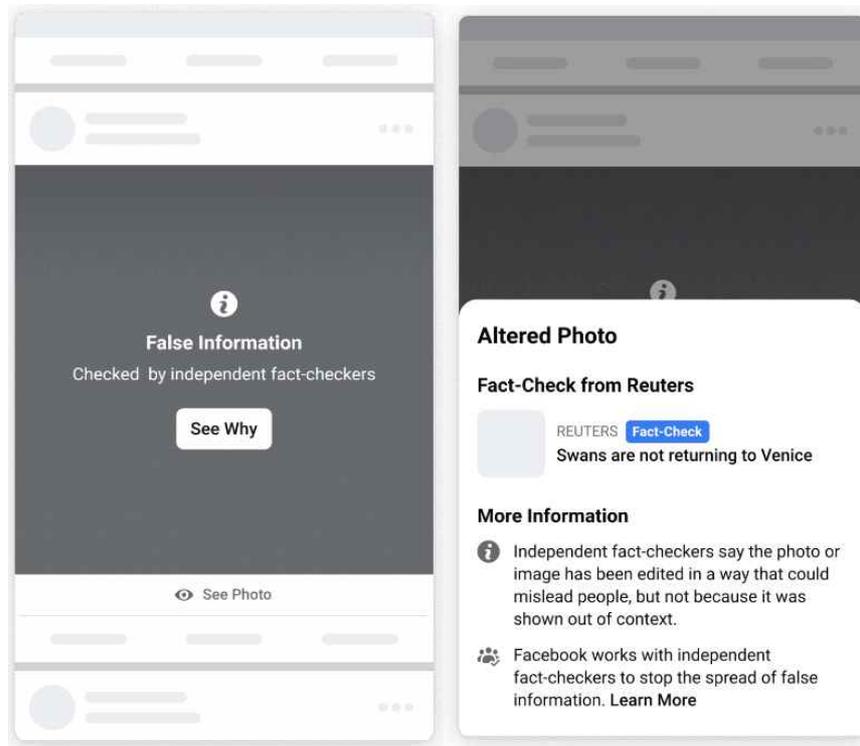
메타는 2019년부터 연구자들과 데이터를 공유하기 위한 표준화된 프로세스를 개발하기 위해 유럽디지털미디어관측소(European Digital Media Observatory, 이하 EDMO)와의 협력을 지속해왔다. EDMO는 EU 회원국 대학 및 연구소를 주축으로 온라인 허위조작정보를 퇴치하기 위해 마련된 기관이다. 구체적으로, 허위정보 관련 자료수집 및 공유, 네트워크 도구 제공, 허위 정보와 관련 동향을 밝히고 분석하는 팩트체크 관련 연구자들에게 자료를 제공하는 허브로 기능하고 있다(Meta, 2023. 1). 메타는 보다 강화된 허위조작정보 대응을 위해서는 자사의 노력 뿐 아니라 연구자, 규제 기관, 개인정보 보호 전문가 등 여러 이해관계자의 공동 대응이 중요함을 피력하고 있다.

⑤ 팩트체크 커뮤니티 역량 강화

메타는 국제 팩트체크 네트워크(International Fact-Checking Network, IFCN)를 통해 인증된 독립적인 팩트체크 기관(Third-Party Fact-Checking Program, 이하 3PFC)과 협력하고 있다. 2023년 7월 기준 메타와 파트너십을 맺은 곳은 EU 내 22개 언어를 사용하는 26개 기관이다. 3PFC는 메타 플랫폼의 허위 콘텐츠를 신속하게 검토하고 평가하며, 2023년 1월부터 6월까지 전 세계 페이스북에서 190,000개가 넘는 게시물에 ‘팩트체크 라벨’을

붙여 이용자가 특정 콘텐츠가 사실 확인 내용을 담고 있는 기사인지를 인지할 수 있도록 도왔다(Meta, 2023. 1). 허위정보로 판단되는 콘텐츠의 경우 메타 플랫폼의 뉴스피드와 기타 공간에 경고화면이 표시되며, 이를 통해 사람들은 콘텐츠를 불지를 선택할 수 있다. 인스타그램의 경우 52,000개 이상의 게시물에 경고화면이 적용되었다.

[그림 4-9] 메타(페이스북) 팩트체크 라벨 및 경고화면



출처: Thevergr (2019. 10. 22). Facebook will label false posts more clearly as part of an effort to prevent 2020 election interference.

메타는 페이스북 게시물에 팩트체크 경고화면이 표시되면 이용자의 95%가 해당 게시물을 클릭하지 않고 보지 않는 것으로 나타났다고 밝혔다. 공유 빈도에도 유의미한 변화가 나타났는데, 경고 메시지가 확인된 경우 EU 회원국 내 인스타그램 사용자의 평균 38%, 페

이스북 사용자의 평균 37%가 콘텐츠를 공유하지 않는 등 괄목할 만한 성과를 보이고 있다. 플랫폼에서 경고화면을 본 적이 있는 사람들을 대상으로 설문조사를 실시한 결과 74%가 팩트체크 라벨의 양이 적절하거나 더 많아도 좋겠다고 답했고, 63%가 팩트체크 라벨이 공정하게 적용되었다고 생각한다고 응답했다(Meta, 2023. 1).

2023년 상반기, 메타는 중앙·동유럽의 3PFC를 대상으로 첫 번째 대면 모임을 개최하여 허위정보 대응하기 위해 협력하는 방안을 논의했다. 3PFC의 생산성과 영향력을 감소시키는 장벽과 이를 해결하는 방법도 모색했으며, 향후에는 전 세계 팩트체커와 관계를 쌓아가며 팩트체크 생태계 지원을 아끼지 않겠다고 밝혔다(Meta, 2023. 7).

4. 실천강령에 대한 평가

앞서 살핀 바와 같이 EU의 허위조작정보 대응 실천강령은 디지털서비스법과 함께 허위조작정보에 대한 공동규제 체계를 구성하기 위해 마련된 것이다. 특히 2018년 대비 2022년 강령은 (i) 팩트체크 대응 영역의 확장, (ii) 구체적 방안 제시 및 (iii) 유관 법 연계를 통한 기능 강화, (iv) 효과 측정을 위한 지표 개발, 그리고 (v) 플랫폼 외 다양한 유관 분야의 참여자 확대와 협력 강화 등 직전 강령 대비 여러 차원에서의 보완이 이뤄졌다. 그럼에도 보완에 따라 새롭게 발생한 문제, 또는 강화 작업에도 불구하고 여전히 해결되지 않은 지점 등 해당 강령에는 몇몇 한계·지적 사항들이 존재한다.

(1) 참여 조직 간 입장차에 따른 협력 제고 한계 가능성

직전 강령이 플랫폼 기업을 중심으로 이뤄졌다면, ‘2022 강령’은 플랫폼 기업 외에도 팩트체크 관련 여러 참여 단체들이 협업하는 체계이다. 해당 강령이 발표된 당시, 이 강령에 대한 평가에 있어 플랫폼 기업과 팩트체크 조직 간 입장 차이가 발생했다. 플랫폼 기업들의 경우 전반적으로 해당 강령의 도입을 환영하고 이를 통해 허위조작정보 문제가 해결될 수 있도록 적극 협력하겠다는 긍정적이고 다소간에 낙관적인 입장을 취하고 있는 반면, 시민 단체 및 팩트체크 기관 등으로 대표되는 非 플랫폼 단체의 경우에는 비교적 해당 강

령에 더 많은 보완사항을 요청한다. 공통적으로 플랫폼 기업의 강령 이행을 강조하고 있는데, 시민 단체 Avaaz의 경우 플랫폼 기업의 강령 이행에 대한 철저한 모니터링이 필요함을 강조하며, 뉴스 및 정보 등급 시스템을 운영 중인 NewsGuard의 경우 궁극적인 이용자 보호를 위해서는 플랫폼 기업에서 신뢰할 수 있는 콘텐츠를만 제공할 수 있도록 하는 의무 조치가 있어야 함을 요청한다(Collins, 2022, 6, 16). 이러한 입장 차를 통해 ‘2022 강령’을 도출하는 협력 과정이 원만치 않았을 수 있음을 추측해 볼 수 있으며, 더 장기적으로는 이러한 대립 구도는 해당 강령이 목표로 하는 허위조작정보 대응을 위한 여러 조직 간의 협력을 이끄는 데 방해요인으로 작용할 수 있다.

<표 4-12> ‘2022 강령’에 대한 참여조직 간 입장 차

플랫폼 기업	非 플랫폼 단체
<ul style="list-style-type: none"> · (메타) 문제 해결을 위해 막대한 투자를 하고 있으며, 함께 해결하기 위해 더 많은 협력을 기대 · (X) 허위조작정보에 맞서기 위해 현재 방식을 계속 평가, 발전시키면서 최선을 다하는 중이며 강령의 업데이트를 환영 · (틱톡) 새로운 강령 초안 작성에 역할한 것에 자랑스러움 · (구글) 강화된 강령의 도입을 환영하며, 성공시키기 위해 노력 중 	<ul style="list-style-type: none"> · (시민 단체 Avaaz) 초대형 플랫폼이 이행하지 않으면 본 강령의 의미는 없으므로 철저한 모니터링이 필요 · (뉴스/정보 등급 시스템 NewsGuard) 플랫폼 기업이 신뢰할 수 있는 콘텐츠를만 제공하도록 하는 의무 조치가 이행되지 않는 한 궁극적인 이용자 보호는 불가능

*출처: Collins (2022, 6, 16)에서 재구성

(2) 탈퇴 및 미참여 플랫폼 기업의 존재

‘2022 강령’의 주요 보완사항 중 하나로 참여자 확대를 꼽을 수 있으며, 앞서 살핀 바와 같이 실제 많은 참여자들이 해당 강령에 서명·참여 증으로 2023년 12월 기준 총 44개의 참여조직이 서명한 것으로 확인된다. 그중에는 메타(Meta), 틱톡(TicTok), 구글(Google) 등 여러 초대형 플랫폼 기업들이 포함되어 있지만, 그럼에도 아직 참여 중이지 않은 플랫폼 기업 또한 존재한다. 해당 강령에 서명하지 않은 대기업으로는 아마존(Amazon), 애플(Apple), 그리고 텔레그램(Telegram)이 대표적이다(Tenner, 2022, 8, 5). 그리고 엑스(X, 前 트위터

터)의 경우에는 2023년 5월에 탈퇴를 발표한 바 있다(Brusselstimes, 2023, 5, 27).

게다가 이미 서명한 기업들의 경우에도 일부 준칙·조치에는 동의하지 않았다는 점 또한 참여와 관련한 또 다른 한계점으로 지적할 수 있다. 메타와 틱톡, 그리고 구글의 경우 이용자들이 허위조작정보 여부를 판단하는 데 도움이 될 수 있는 도구 중 하나인 ‘신뢰도 지표’ 제공(준칙 22)에는 동의하지 않았다(Tanner, 2022, 8, 5).

[그림 4-10] 메타(Meta)의 ‘2022 강령’ 서명 자료 중 일부

Commitment 22 Relevant Signatories commit to provide users with tools to help them make more informed decisions when they encounter online information that may be false or misleading, and to facilitate user access to tools and information to assess the trustworthiness of information sources, such as indicators of trustworthiness for informed online navigation, particularly relating to societal issues or debates of general interest	Measure 22.1	N/A	
	Measure 22.2	N/A	
	Measure 22.3	N/A	
	Measure 22.4	N/A	Measure 22.4 applies to signatories providing trustworthiness indicators.
	Measure 22.5	N/A	Measure 22.5 applies to signatories providing trustworthiness indicators.
	Measure 22.6	N/A	Measure 22.6 applies to signatories providing trustworthiness indicators.
	Measure 22.7	N/A	Facebook and Instagram use products/features in Measure 22.7 such as information panels under Commitment 21.
Reason for not subscribing to Commitment 22 (if applicable): Trustworthiness indicators are one of several ways to empower users to make more informed decisions about the content they see online. This is acknowledged by the Commission's 2021 Guidance, which describes them as a tool signatories "could" explore, and negotiations of the updated Code which confirmed this to be a direction signatories are encouraged but not expected to follow. Other tools to achieve this objective covered elsewhere in this section - Commitment 21 in particular - are relevant and pertinent for our subscribed products at this time.			

출처: Meta (2022, 6, 14), pp. 17-18.

가령 메타의 경우, 해당 준칙에 서명하지 않은 이유로 이용자들이 정보에 기반한 의사 결정을 내릴 때 도움을 줄 수 있는 도구는 여러 가지가 될 수 있고 ‘신뢰도 지표’는 그 중 하나이므로 반드시 따를 필요가 없다고 명시하고 있다(Meta, 2022, 6, 14). 해당 강령이 효과적으로 작동하기 위해서는 허위조작정보가 유포되고, 수익화로 이어지는 공간인 (초대형) 플랫폼 기업의 참여가 중요한 만큼 이들의 참여 여부가 중요할 것으로 판단된다.

(3) 보고 방식 등 영향력 측정 기준의 수립

2022년 EU는 이행 결과를 체계적으로 평가하기 위해 통일된 기준과 표준화된 보고 서식을 만들었다. 강화된 실천 강령 도입 이전에도 각 기업은 실천 강령의 준수 여부를 점검하는 연간 보고서를 발간해왔으나, 그 조치가 허위조작정보 확산을 줄이는데 어떻게 기여했는지에 대한 정보가 구체적이지 않았다. 가령 페이스북은 2018년 실천 강령 시행 후 공시한 첫 번째 보고서에서 광고 게재 심사 절차와 결과를 누락했으며, 구글은 서비스 무결성과 관련한 전략을 구체적으로 어떻게 구현했는지에 대한 서술이 부족했다. EU는 발표된 기업 보고서에 대해 매해 자료를 보완하고 결과에 대한 투명한 공개를 요구했지만 여전히 정확한 평가가 가능한 수준의 정보가 충분하지 않았다(심나리, 2019). 이 같은 문제는 기업의 불투명한 관행이나 책임 회피에서 기인하기보다는 이행 보고서에 기술되어야 할 통일된 지표가 부재한 것에서 비롯된다. 이에 따라 EU는 각 기업의 이행 상황을 추적하고 객관적으로 측정할 수 있는 기준점이 필요함을 적시해왔고, 강령 개정 과정에 참여한 기업들은 9가지 주제 영역, 44개의 준칙, 128개의 조치 및 서비스 수준 지표(SLI)와 질 적합성 지표(QRE)로 구조화된 서식에 따라 이행 결과를 보고하기로 합의했다.

하지만 몇 가지 현실적 한계로 여전히 영향력을 측정하기에는 어려움이 있다는 비판도 있다. 먼저, 팩트체크를 둘러싼 여러 이해관계자들이 참여하고 있어 효과와 영향력 등을 비교 가능한 표준화된 운영 보고를 이끌기 어렵다는 한계가 있다는 점이다. 참여 서명인들의 성격이 다를 뿐만 아니라, 특히 플랫폼 기업들의 경우 기능과 구조적 차이로 인해 수집 및 보고할 수 있는 데이터가 다르기 때문이다. 실제 몇몇 플랫폼들은 ‘2022 강령’ 서명 이후 두 번째 보고였던 2023년 7월에 허위조작정보 대응을 위한 개입에 따른 결과, 즉 영향력 지표 및 측정 결과를 제출하였으나 그 측정방식과 내용이 상당히 다양했던 것으로 확인된다(Lai, & Yadav, 2023, 11, 21). 이는 동일한 기준에서 참여자들의 해당 강령 이행 여부를 비교/평가하기 어렵게 하는 지점이 여전히 존재함을 가리킨다.

<표 4-13> 플랫폼별 허위조작정보 대응 관련 영향력 측정 보고내용

구분	내용
Google 및 Microsoft	· 수익 창출을 막은 페이지 및 도메인의 잠재적 수익
TicTok	· ‘확인되지 않은 콘텐츠(unverified content)’ 표시 이후 이용자들의 공유 취소율
Meta	· ‘팩트체크’ 표시 이후 이용자들의 원본 콘텐츠 확인율

*출처: Lai, & Yadav (2023, 11, 23)에서 재구성

한편 각 기업이 구축한 ‘투명성 센터’ 웹페이지는 강령의 실천성을 내실화할 수 있는 중요한 제반 장치로 평가받을 필요가 있다. 구글과 메타는 각 투명성 센터를 통해 6개월마다 발표하는 실천 강령 이행 보고서를 게시하고 있다. 결과 보고서를 누구나 접근할 수 있는 웹페이지에 공개하는 것은 사소하지만 의미 있는 실천이다. 이행 결과와 평가의 내용이 기업 내 보고서로만 남는다면 자신들의 역할을 충실히 수행하지 않는다 하더라도 시민 사회가 세세하게 알기는 어렵다. 실제로 조사하는 과정에서 2022년 강령 도입 이전의 연간 보고서는 이리저리 흩어져 있어 찾기 어렵고, 보고서의 제목도 제각각이라 통합 검색되지 않는 등 접근하기가 쉽지 않았다. 그러나 별도로 구축된 투명성 센터 웹페이지에 실천 강령을 이행하기 위한 노력이 집대성됨으로써 기업이나 전문가뿐만 아니라 시민들도 실천 강령의 의미와 지침을 이해할 수 있게 되었고, 각 기업이 이전 강령에 따라 이행한 조치와 새 강령을 실천하기 위해 취해진 조치에 대한 비교도 가능해졌다.

제 2 절 호주의 허위조작정보 및 오정보 대응 실천강령

1. 개관

호주 정부는 디지털 환경 변화로 인해 직면한 허위조작정보 문제에 당면하여, 디지털 플랫폼 규제에 대한 새로운 접근방식을 마련하기 위해 다양한 노력을 기울이고 있다(ACMA, 2021). 호주 정부는 2019년 7월 26일에 발표된 호주 경쟁·소비자위원회(Australian Competition and Consumer Commission, ACCC)의 <디지털 플랫폼 조사(Digital Platform Inquiry) 최종 보고서>를 통해 사용자의 역량 강화와 디지털 플랫폼의 자발적인 행동규범 마련을 위한 로드맵을 제안했다. 그 일환으로 호주 정부는 2019년 12월, <디지털 시대의 규제: 정부 대응과 로드맵(Regulating in the Digital Age: Government Response and Implementation Roadmap)>에 명시된 정책을 기반으로 하여, 주요 디지털 플랫폼의 허위조작정보 확산에 대한 우려를 해소하기 위해 자발적인 실천강령을 마련하도록 요청했다(Australian Government, 2019).

아울러 2020년 6월, 호주 통신미디어청(Australian Communications and Media Authority, ACMA)은 <호주 디지털 플랫폼의 허위조작정보와 뉴스 품질: 강령 개발 지침에 관한 의견서(Misinformation and News Quality on Digital Platforms in Australia: A Position Paper to Guide Code Development)>를 발표하며, 실천강령에 대한 보다 명확하고 광범위한 지침을 제시하였다(ACMA, 2020). 2020년 6월에 발표된 호주 통신미디어청의 의견서는 실천강령을 통해 이행되어야 하는 주요 목표를 제공한다. 구체적으로 첫째 허위조작정보의 해로운 영향력 해소, 둘째 사용자의 역량 강화, 셋째 디지털 플랫폼의 투명성과 책임성 강화가 주요 목표이다.

비영리 산업 협회인 디지털산업그룹(Digital Industry Group Inc, DIGI)은 2020년 10월, 약 5주 간의 공개 협의 과정을 통해 실천강령의 초안을 만드는 일을 담당하였다. 이 과정에서 주요 디지털 플랫폼들, 정부, 학계, 시민사회의 대표들로 구성된 다양한 이해관계자들과의 협의 과정을 거쳐 2021년 2월 22일 <호주의 허위조작정보 및 오정보 실천강령(Australian Code of Practice on Disinformation and misinformation)>(이하, ‘호주 실천강령’)을 공표

했다. ‘호주 실천강령’은 디지털 플랫폼 사업자들을 비롯하여 호주의 다양한 이해관계자들이 협력하여, 처음으로 마련된 자율규제 방안이라는 점에서 의미가 깊다.

<표 4-14> 호주 실천강령 관련 연혁

구분	주요 사건
2019.12.	· 호주 정부는 호주 경쟁·소비자 위원회(ACCC)의 디지털 플랫폼 조사에 따라, 디지털 플랫폼에 실천강령을 요청함
2020.1-3.	· 호주 통신미디어청(ACMA)은 실천강령 마련을 위해, 호주 주요 디지털 플랫폼과 논의함
2020.3.	· DIGI는 호주 통신미디어청(ACMA)에 산업 전반의 허위조작정보 실천강령을 마련하는 데 주도적인 역할을 할 것임을 발표함
2020.5.	· DIGI는 2020년 11월까지의 실천강령 채택에 대한 프로젝트 계획을 호주 통신미디어청(ACMA)과 공유함
2020.6.	· 호주 통신미디어청(ACMA)은 실천강령 개발에 대한 지침 및 목표를 제안하며, 실천강령을 공식화하는 의견서를 발표함
2020.10.	· DIGI는 디지털 플랫폼 기업을 포함한 다양한 이해관계자들 간의 공개 협의 과정(5주)을 통해 만들어진 실천강령 초안을 공개함
2021.2.	· DIGI는 최종 <실천강령>을 공표하고, 트위터(Twitter), 구글(Google), 페이스북(Facebook), 마이크로소프트(Microsoft), 레드버블(Redbubble), 틱톡(TikTok) 6개사 서명인을 발표함
2021.5.	· DIGI는 어도비(Adobe), 애플(Apple)을 추가 서명인으로 발표하고, <실천강령>에 참여한 8개사의 서명인들을 대상으로, 자발적인 투명성 보고서를 DIGI 웹사이트에 공개함
2021.6.	· 호주 통신미디어청(ACMA)은 디지털 플랫폼의 허위조작정보 대응 및 규약의 적절성 여부에 대한 보고서를 발표함
2021.10.	· DIGI는 <실천강령> 준칙 위반에 대한 불만을 해소하기 위해 불만처리위원회를 마련하고, 서명인의 위반 사항을 공개적으로 보고함
2022.6-7.	· DIGI는 2022년 6월 6일부터 2022년 7월 18일까지 <실천강령>에 대한 개정을 위해 공개 제출을 수락함
2022.12.	· 공개 협의 과정에서 수렴된 이해관계자의 의견들을 반영하여, <실천강령>을 개정함

*자료: ACMA (2021). <ACMA report to government on the adequacy of digital platforms’ disinformation and news quality measures> 내용 재구성

DIGI는 급속하게 변화하는 디지털 환경에 따라 강령을 지속적으로 개선하고, 다양한 영역에서 규정을 강화하고자 하였다. 이러한 맥락에서 2021년 5월, 트위터, 구글, 페이스북, 마이크로소프트, 레드버블(Redbubble), 틱톡, 어도비(Adobe), 애플 ‘호주 실천강령’에 참여한 8개사의 서명인들을 대상으로, 자발적인 투명성 보고서를 DIGI 웹사이트에 공개하기도 했다. 또한 디지털 플랫폼의 허위조작정보와 오정보 대응에 대한 관리·감독을 호주 통신미디어청(ACMA)에서 맡아, 디지털 플랫폼의 허위조작정보 대응 및 규약의 적절성 여부를 보고하는 과정을 통해 호주 실천강령을 지속적으로 검토하고자 했다.⁴¹⁾ 이러한 노력의 일환으로, 2021년 10월 DIGI는 호주 실천강령 준칙 위반에 대한불만을 해소하기 위해 불만처리위원회를 마련하고, 서명인의 위반 사항을 공개적으로 보고할 수 있도록 강령을 강화했다. 이후, 2022년 12월 공개 협의 과정에서 수렴된 이해관계자의 의견들을 반영하여, 호주 실천강령을 검토 및 개정하였다. 아래에서는 2022년 12월 개정된 실천강령에 대한 주요 내용을 살펴도록 한다.

2. 주요 내용 및 특징

전술했다시피 호주 정부는 2019년 12월 주요 디지털 플랫폼의 허위조작정보 유포와 확산 문제를 해소하기 위한 하나의 대응방안으로, 디지털 플랫폼의 역할과 책임성에 방점을 둔 자발적인 실천강령을 제안했다. 이에 DIGI는 다양한 기업, 학술 공동체 및 독립적인 연구자, 시민사회 대표 등 다양한 이해관계자들이 참여한 공개 협의 과정을 통해 실천강령의 초안을 만들어 2021년 2월 발표했다.

호주 실천강령은 EU의 허위조작정보 대응 실천강령을 바탕으로 하기 때문에 ‘자율규제’라는 점에서 유사성을 지니지만, EU의 준칙이 ‘관련 서명인’의 참여를 요구하는 것과 달리 모든 서명인이 준수하고 이행해야 하는 최소 준칙을 제시하고 있다는 점에서 차

41) 호주 통신미디어청(ACMA)은 허위조작정보에 대한 디지털 플랫폼의 조치가 적절한지에 대한 여부를 관리 및 감독하고, 디지털 플랫폼에서 경험하는 허위조작정보와 오정보에 대한 인식과 경험을 조사하여 실천강령을 평가한다. 이를 토대로 디지털 플랫폼의 실천강령 이행 여부를 검토할 것을 권고한다.

별성을 가진다. 최소 준칙 사항에 대한 준수 여부를 확인하고, 변화하는 디지털 환경에 적극적으로 적절하고 구체적인 대책을 마련할 수 있도록 하기 위해 실천강령에 참여하고 있는 서명인들은 매년 투명성 보고서를 발간 및 공개해야 한다. DIGI 역시 매년 호주 통신미디어청에 연례 보고서를 제출하고, 2년을 주기로 실천강령을 필수적으로 검토하고 개선해야 한다.

이 과정은 통신미디어청을 포함한 정부기관, 해당 분야에 활동하는 학계, 시민사회 대표를 포함한 이해관계자의 참여를 기반으로 이루어진다. 이에 대한 첫 단계로서, DIGI는 2022년 12월 22일에 검토 및 개선사항을 업데이트한 실천강령을 발표했다.

특히 2022년에 개정된 실천강령은 다양한 이해관계자들의 의견들을 다수 반영하여, 몇 가지 변경사항을 포함하고 있다. 구체적인 변경사항은 다음과 같다. ① 소규모 디지털 플랫폼의 참여 확장, ② 허위조작정보와 오정보 관련 ‘피해’의 기준을 ‘심각하고 확실한(믿을 만한) 위협’으로 변경, ③ 2022년에 업데이트된 EU의 실천강령을 반영하여 추천 시스템에 대한 시행 규정 추가, ④ 허위조작정보와 오정보를 유포하는 디지털 광고를 반복적으로 게재하는 것에 대한 규정사항 추가, ⑤ 스폰서 콘텐츠와 유료 디지털 광고 서비스를 허위조작정보와 오정보 관련 규정 범위 안에 포함, ⑥ 기존의 조항에 따라 전문 뉴스 콘텐츠가 허위조작정보로 다뤄지지 않도록 하되, 해당 콘텐츠가 허위조작정보 확산 시 서명인에 대한 책임과 의무 조항 유지, ⑦ 강령에 대한 업데이트 및 투명성 보고서 등 서명인의 준칙 범위 내에서 특정 제품과 서비스에 대한 높은 투명성 요구. 아울러 2022년 실천강령은 독립적인 불만처리위원회를 마련하고 불만처리 제도를 설치하고, 독립적인 모범 사례 지침을 도입하여, 디지털 플랫폼 기업의 투명성과 책임성을 높이고자 했다.

개정된 실천강령은 허위조작정보를 A. 실증적으로 잘못됐거나, 오도하거나, 기만적인 디지털 콘텐츠로, B. 부정 행위(Inauthentic Behaviors)를 통해 디지털 플랫폼 이용자 사이에서 확산되고, C. 해당 내용의 유포가 피해를 끼치리라고 합리적으로 예상되는 콘텐츠 그리고 오정보는 A. 실증적으로 잘못됐거나, 오도하거나, 기만적인 (종종 적법한) 디지털 콘텐츠로, B. 디지털 플랫폼 이용자에 의해 확산되고, C. 해당 내용의 유포가 (명확하게 기만을 의도하지 않더라도) 해를 끼치리라고 합리적으로 예상되는 콘텐츠로 정의한다.

〈표 4-15〉 호주 실천강령의 허위조작정보 및 오정보 정의

허위조작정보	오정보
<ul style="list-style-type: none"> · A. 실증적으로 잘못되었거나, 오도하거나, 기만적인 디지털 콘텐츠 · B. 이는 부정 행위(Inauthentic Behaviors)⁴²⁾를 통해 디지털 플랫폼 사용자들 사이에서 확산된다. · C. 해당 내용의 유포가 피해⁴³⁾를 끼치리라고 합리적으로 예상되는 콘텐츠 	<ul style="list-style-type: none"> · A. 실증적으로 잘못되었거나, 오도하거나, 기만적인 (종종 적법한) 디지털 콘텐츠 · B. 이는 디지털 플랫폼의 사용자에 의해 확산된다. · C. 해당 내용의 유포가 (명확하게 기만을 의도하지 않더라도) 해를 끼치리라고 합리적으로 예상되는 콘텐츠
허위조작정보에서 제외되는 콘텐츠	
<ul style="list-style-type: none"> · A. (풍자 및 패러디를 포함한) 오락 또는 교육의 목적으로, 선의의 신뢰 하에 제작된 콘텐츠 · B. 호주 주정부 또는 연방정부에서 승인한 콘텐츠 · C. 5.23절부터 5.25절에 따라, 정치 광고 또는 (호주 법률에 따라 등록된) 정당이 인가한 콘텐츠 · D. 전문 뉴스 콘텐츠⁴⁴⁾ 	

호주 실천 강령은 모든 서명인이 준수해야 할 최소준칙을 포함한 7가지 목표를 제시한다. 7가지 목표는 ① 허위조작정보와 오정보로 인해 발생할 수 있는 피해에 대한 안전장치

42) 부정 행위Inauthentic Behaviour는 스팸spam과 다른 형태의 속임수, 조작, (자동화된 시스템을 통해 자행될 수 있는) 대규모의 공격적 행위를 포함하며, 디지털 플랫폼 사용자들의 온라인 대화에 인위적으로 영향을 미치거나 디지털 콘텐츠를 확산하도록 하는 행위를 포함한다.

43) 피해(Harm)는 A. 선거와 같은 민주적인 정치와 정책결정 프로세스, 그리고 B. 공중 보건의 보호, 소외계층 혹은 취약계층의 보호, 치안과 보안 또는 환경과 같은 공공재에 확실하고 심각한 위협이 되는 피해를 의미한다.

44) 전문적인 뉴스는 A. 최종 사용자를 공적인 토론에 참여시키고, 민주적 의사결정 과정에 정보를 제공하는 현안 또는 사건, B. 장소, 지역 또는 국가 수준에서 최종 사용자에게 공적인 중요성이 있는 현안 또는 사건을 보도하거나 조사하거나 비판적인 분석을 제공하는 뉴스 출처를 통해 제작된 온라인 자료물이다.

제공, ② 허위조작정보와 오정보에 대한 광고 또는 수익창출 인센티브 차단, ③ 디지털 플랫폼에서 제공하는 서비스와 제품의 안전성 및 보안 보장, ④ 소비자 역량 강화, ⑤ 정치 광고 출처에 대한 공중 인식 개선, ⑥ 전략적 연구 지원, ⑦ 서명인의 투명한 대응책 공개이다. 다음은 7가지 주요 목표와 각 목표의 결과를 정리한 것이다.

<표 4-16> 호주의 허위조작정보와 오정보 실천강령의 주요 목표 및 결과

목표	주요 결과
1. 허위조작정보와 오정보에 대한 안전장치 제공	<ul style="list-style-type: none"> · 결과 1a: 서명인들은 확장 가능한 다양한 대책을 채택함으로써, 디지털 플랫폼에서 허위조작정보와 오정보의 확산으로 인해 발생할 수 있는 피해의 위험을 줄이는데 기여함 (모든 서명인 준수) · 결과 2b: 본 강령에 따라, 사용자는 서명인이 규제하거나 관리하게 될 행위 또는 콘텐츠 유형에 대한 정보를 제공 받게 됨 · 결과 3b: 사용자는 공개적으로 이용 가능하고 접근 가능한 리포팅 툴reporting tools를 통해, (5.10절에 따라) 정책을 위반하는 콘텐츠 또는 행동을 서명인에게 신고할 수 있음 · 결과 1d: 사용자는 5.11절에 따라 제출된 신고에 응하여, 서명인의 대책에 관한 일반 정보에 접근할 수 있음 · 결과 1e: 사용자는 추천 시스템 이용에 관한 서명인의 일반 정보에 접근하고, 추천 시스템이 제안하는 콘텐츠와 관련한 옵션을 가질 수 있음
2. 허위조작정보와 오정보에 대한 광고 또는 수익창출 인센티브 차단	<ul style="list-style-type: none"> · 결과 2: 허위조작정보와 오정보에 대한 광고나 수익창출 인센티브가 감소함

3. 디지털 플랫폼에서 제공하는 서비스와 제품의 완전성 및 보안 보장	· 결과 3: 사용자의 조작 행위로 인한 서비스와 제품의 완전성과 보안이 훼손될 위험이 감소함
4. 소비자 역량 강화	· 결과 4: 사용자는 (디지털 플랫폼을 통해 접근할 수 있는) 뉴스와 사실 기반 콘텐츠의 출처에 대해 더 많은 정보에 기반한 선택을 할 수 있으며, 오정보를 식별하는 데 능숙해짐
5. 정치 광고 출처에 대한 공중 인식 개선	· 결과 5: 사용자는 정치 광고의 출처에 대해 더 잘 알게 됨
6. 전략적 연구 지원	· 결과 6: 서명인은 허위조작정보와 오정보에 대한 공중의 이해를 향상시키기 위해 독립적인 연구자의 노력을 지지함
7. 서명인의 투명한 대응책 공개	· 결과 7: 공중은 허위조작정보와 오정보에 대응하기 위해, 서명인이 이해한 대책에 대한 정보에 접근할 수 있음 (모든 서명인 준수) * 단, 호주 내 월간 활성 최종 사용자가 100만 명을 초과하는, 사용자 기반의 대규모 플랫폼과 그 미만인 플랫폼은 보고서 양식을 달리함

이상의 내용을 바탕으로 호주 실천강령의 주요 특징을 다음과 같이 정리할 수 있다. 먼저 호주 실천강령은 허위조작정보와 오정보를 구분하는데, ‘실증적으로 잘못되었거나 오도하거나 기만적인 콘텐츠’ 중에서 스팸이나 속임수, 조작 등 ‘부정행위’를 통해 확산되면 허위조작정보로, 사용자들에 의해 직접 확산되면 오정보로 규정한다. 즉 콘텐츠의 내용이 아니라 유포 과정에서 조작의 개입 여부가 양자를 구분하는 핵심이라고 할 수 있다. 다만 이러한 허위조작정보로 인해 발생하는 피해는 확실하고 심각한 위험이 되는 피해로 한정된다.

호주 실천강령은 목표1을 명시하여, 모든 플랫폼이 허위조작정보와 오정보에 대응하여

대책 및 조치를 마련하고 이를 처리할 수 있도록 최소준칙을 제시한다. 이에 따르면, 서명인은 규제대상인 제품과 서비스에 대한 구체적인 정보와 지침을 사용자에게 제공해야 한다. 뿐만 아니라 서명인은 사용자가 정책에 위반되는 불법콘텐츠를 직접 신고할 수 있는 절차와 정책, 도구를 구현하고 공표해야 한다. 그리고 신고가 접수되면, 서명인은 불법콘텐츠에 대한 삭제 여부를 심사하고 해당 사항에 대한 절차와 구체적인 내용을 사용자에게 보고하고 공개하여야 한다. 추가적으로 서명인은 사용자가 이용하는 추천 시스템에 대한 일반 정보를 제공하여야 한다.

또한 호주의 실천강령은 EU의 실천강령과 마찬가지로 이용자의 역량을 강화한다. 서명인은 허위조작정보와 오정보에 대한 광고 또는 수익창출 인센티브를 차단하고, 플랫폼에서 전달되는 (정치) 광고 출처에 대한 대책 및 정보를 제공하여 생태계 전반의 투명성을 개선해야 한다. 호주 실천강령은 이를 토대로 사용자가 제공받는 플랫폼 서비스와 제품의 완전성과 보안을 보장하고자 한다. 이 과정에서 사용자가 불만처리제도 및 정확성과 신뢰성이 있는 정보와 기술을 통해 디지털 콘텐츠에 접근할 수 있도록 함으로써 이용자의 역량 강화를 꾀한다. 한편 서명인은 허위조작정보와 오정보에 대한 이용자의 이해를 증진시키기 위해, 독립적인 연구자와 팩트체크 기관, 학계, 시민사회와의 논의를 촉진하고 연구를 지원해야 한다.

셋째, 호주 실천강령은 모든 플랫폼이 절차 이행에 있어서 투명성을 확보하기 위한 내용을 제시한다. 서명인의 규모에 따라 강령에 명시된 결과를 완수하기 위한 진행상황과 정보를 보고서 형식으로 작성하고 DIGI 웹사이트에 공표해야 하는데, 특히 대규모 디지털 플랫폼 서명인은 투명성 보고서, 불법 콘텐츠 삭제 및 차단에 대한 설명, 연구 보고서를 작성하고 DIGI 웹사이트에 공표해야 한다. 홈페이지에 공개하는 보고서는 대책, 정책, 범위 등 변경되는 내용을 연 단위로 업데이트하여 사용자 및 이해관계자들에게 제공해야 한다. 또한 DIGI는 내재적으로 독립적인 불만처리제도를 설치하여 디지털 플랫폼 내 준칙 이행 여부를 확인하고, 외재적으로 관리소위원회를 설치하여 6개월마다 서명인의 준칙 이행 여부를 검토하고 나아가 서명인의 투명성 보고서를 검증한다.

3. 호주 실천강령과 법체계의 관계

호주 실천강령은 기업에게 허위조작정보 및 오정보의 유포를 억제하기 위한 의무준수를 요구한다. 하지만 이는 자율규제라는 점에서 의무의 준수를 강제할 수 있는 수단이 부재하다는 한계를 지닐 수밖에 없다. 이에 호주 노동당은 2023년 6월 「2023년 통신 법률 개정안(허위조작정보 및 오정보 방지)」 초안을 발의했다. 이 개정안은 법 안에 허위조작정보와 오정보를 포함하며, 플랫폼 사업자의 의무준수 여부에 대한 통신미디어청의 감독 권한을 강화하는 내용을 담고 있다(이지영, 2023).

보다 구체적으로 살펴보면 플랫폼사의 허위조작정보 및 오정보 대응이 미흡할 경우 통신미디어청은 첫째, 디지털 플랫폼 사업자로부터 정보를 수집하건, 허위조작정보 및 오정보 관련 사항에 대한 특정 기록을 보관하도록 요구할 수 있다. 둘째, 통신미디어청은 디지털 플랫폼 사업자에게 허위조작정보 대응을 위한 실천강령을 마련하도록 요구할 수 있으며, 통신미디어청은 이를 등록하고 관리·감독한다. 셋째, 사업자들이 마련한 실천강령이 효과적이지 않다고 판단되면 더 강력한 형태의 산업표준을 만들고 시행할 수 있다. 아울러 개정법안에서는 법률에 허위조작정보와 오정보의 정의를 명시하는데 이는 실천강령의 정의와 대동소이하다(이지영, 2023). 또한 동법은 사업자에게 최대 688만 달러 또는 해당 기업의 전 세계 매출액의 5% 중 높은 금액을 벌금으로 집행할 수 있다(Sydney Morning Herald, 2023.10.1.).

동 법안은 법률 내에 허위조작정보와 오정보의 정의를 명시하고, 사업자의 자율적인 실천강령이 효과적이지 않다고 판단될 경우 더 강력한 형태의 산업 표준을 만들고 시행할 수 있다고 규정한다는 점에서 EU의 디지털서비스법보다 강력한 대응방안을 담고 있다고 할 수 있다. 한편, 호주의 통신미디어청은 플랫폼에 특정 콘텐츠나 게시물의 삭제를 요청할 권한은 갖지 않지만 플랫폼에 이에 대한 조치를 요구할 수 있는 권한을 부여 받는데, 이는 규제당국의 콘텐츠 규제의 강제성을 높이는 효과를 낼 수 있다. 때문에 호주인권위원회(Australian Human Commission)를 포함한 다양한 전문가와 단체는 이것이 표현의 자유를 위축시킬 수 있다는 우려를 표하고 있다(이지영, 2023).

4. 호주 실천강령에 대한 평가

호주의 실천강령이 EU의 실천강령을 참고했다는 점에서 알 수 있듯이 두 실천강령은 많은 부분에서 비슷하다. 두 강령은 그 범위와 접근방식에서 포괄적이고 맥락적인 요인을 고려하여 지침의 대상과 목표를 설정한다. 지침은 서명주체인 디지털 플랫폼이 각 상품과 서비스 형태에 따라 적합한 대책을 마련할 수 있는 자율성과 권한을 부여한다. EU와 비슷하게 호주 실천강령에 자발적으로 참여하고 있는 서명주체는 애플, 구글, 어도비, 마이크로소프트, 틱톡, 트위터 등 국제적인 디지털 플랫폼인 점과 연관이 있다. 따라서 서명주체인 플랫폼은 복합적인 비즈니스 상황과 향후 빚어질 수 있는 다양한 맥락적인 요인들을 고려하여, 다양한 대책과 목표를 선택할 수 있다. 호주 실천강령의 서명주체로서 플랫폼은 강령의 적용을 받는 비즈니스 모델과 제품을 특정 시점과 유형으로 국한하지 않고, 매해 연례보고서를 통해 지속적으로 재평가하며 사전 동의 지명 양식을 업데이트해야 하는 의무를 가진다. 실천강령은 내용규제에 있어서 법적 규제가 직면할 수밖에 없는 정부의 간섭 문제, 상품 및 서비스의 제한적인 범위 문제, 효과적이고 적실한 대응의 한계 문제를 넘어 자율규제의 집행가능성을 높인다. 요컨대 실천강령은 서명주체인 디지털 플랫폼의 국제적인 차원의 특성을 고려할 때, 상황에 따라 유동적이고, 신속하게 규제방침을 적용할 수 있다는 장점을 지닌다. 나아가 국가 간 협력을 통해 보다 포괄적이고 다차원으로 준칙이 이행될 수 있다는 점에서 실효성이 높다고 평가할 수 있다.

또 EU 강령과 마찬가지로 참여주체의 다양성과 전문성 측면에서, 호주의 실천강령은 디지털 플랫폼과 정부를 대상으로 한 규제로 귀결되지 않는다. 강령은 다양한 ‘이해관계자 간의 협업과 협력’을 주요 준칙으로 강조하며, 공공기관 뿐 아니라 시민사회, 학계, 미디어조직 간의 연계 가능성을 염두 해 둔다. 특히 해당 준칙을 통해 시민단체와 사용자의 역할을 강화하여 강령의 다양성을 확보하는 동시에, 강령의 전문성을 실현하기 위해 독립 연구자, 학술 공동체를 지원한다. 구체적으로 독립적인 연구자 및 연구 단체가 윤리 정책에 따라 수행하는 연구 지원, 독립적인 팩트체크 기관 협력, 공동 연구 수행, 학계와 시민사회와의 협업을 위해, 연례행사 개최 등 학계와의 협력에 대한 준칙을 명시하여, 허위조작정보와 오정보에 대한 전문적이고 학술적인 지식 구축을 수행한다. 무엇보다 강령이 2

년 주기로 검토되는 과정은 서명주체인 플랫폼과 정부기관, 학계 연구자, 시민사회 등 다양한 이해관계자들의 참여를 기반으로 이루어져야 함을 명시한다.

이러한 참여주체의 다양성과 전문성 확보는 곧 절차의 공정성과 투명성에 대한 평가로도 이어진다. 즉, 해당 강령은 세부 준칙을 통해 절차적으로 어떻게 준칙을 이행하고 있으며, 앞으로의 이행하고 준수해야 할 준칙의 방향 설정과 수정사항들을 공유할 수 있는 매개 역할을 한다. 이러한 점에서 세 번째, 강령은 절차적인 공정성 및 투명성 측면에서 의의가 있다. 우선, 서명주체로서 플랫폼은 DIGI에 매년 연례보고서를 발간하여, 보고서가 다루는 기간 동안의 실천강령 준수 사항에 대해 밝힐 의무를 가진다. 그리고 매년 강령 조항에 대한 준수 의무를 ‘선택’ 하고, 이에 대해 재평가 및 수정사항을 기술해야 한다. 아울러 서명주체는 플랫폼 규모에 근거하여 공식적인 보고를 해야 한다. 구체적으로 콘텐츠 삭제, 보고서, 언론발표, 투명성 보고서를 공개적으로 업데이트하고, 이외에 백서, 블로그 게시물, 공지사항, 도움말 센터 등 허위조작정보 대응에 대한 정보를 상세히 공개해야 한다. 보고서는 그 내용을 통해 자율규제의 집행 기준과 세부사항을 평가할 수 있다는 면에서 투명성 측면이 연결되기 때문에 중요하다(김민정, 2020). 이는 준수 이행 사항을 공개하는 데 그 강조점을 두는 것이 아니라, 타당하고 적실한 평가 과정으로 이어져야 한다. 서명주체인 플랫폼과 독립적인 구성원을 대표로 하여, ‘관리 소위원회’를 설치하여, 6개월 마다 서명인의 대책을 검토하고 준칙 이행 준수 사항을 관리·감독한다. 또한 서명주체가 제출한 연례보고서 및 투명성보고서를 팩트체크하고 검증한다. 이러한 맥락에서 호주의 실천강령은 절차적인 공정성 및 투명성 측면에서 주효하다.

다음으로 EU와 마찬가지로 호주 실천강령은 구체적인 목표를 명시하며 공적 책무에 대한 지향을 드러낸다. ‘허위조작정보와 오정보 확산 피해로 인한 안전장치 제공’, ‘규제 대상인 콘텐츠 유형과 추천시스템에 대한 정보 공개’, ‘정치광고 및 수익창출 인센티브 차단 및 서비스의 완전성 보장’, ‘수용자에 대한 역량 강화’ 등이 그 내용이다. 먼저 강령은 허위조작정보와 오정보 피해에 대한 안전장치 및 대책 확보를 최우선 준칙으로 강조한다. 완전성과 보안을 보장한 서비스 및 제품의 제공에 대한 준칙은 ‘공적 책무 맥락에서’ 의의가 있다. 다만 이는 여기서 그치는 것이 아니라, 플랫폼을 사용하는 사용자들에게 문제된 콘텐츠와 그에 상응하는 대책에 대한 정보를 투명하게 공개하는 데까지 이어져야 한다. 정치 광고와 수익창출 인센티브에 대한 규제 또한 중단에 그 목적이 있는 것이

아니다. 플랫폼은 사용자가 정보에 대한 출처를 쉽게 인식할 수 있도록 구체적으로 명시하고, 자동화된 추천 시스템에 관한 옵션을 제공하여 소비자로서 역량을 강화한다. 이처럼 호주의 실천강령은 허위조작정보 및 오정보 피해에 대한 대책을 마련하고, 완전성과 보안이 보장된 콘텐츠를 제공해야 하는 준칙을 통해 강령의 ‘목적’을 서술한다. 나아가 허위조작정보 및 오정보 피해 콘텐츠를 어떻게 다뤄야하는지 그 ‘방식’ 또한 구체화한다는 점에서, 공적 책무에 대한 의의를 지니고 있다.

그러나 EU의 실천강령이 디지털서비스법과의 연관 속에 공동규제체계를 이름으로써 그 실효성과 강제력이 뒷받침되는 것과 달리, 호주의 실천강령은 아직 공동규제를 위한 법적 기반을 갖고 있지는 않다. 이와 관련해 호주 노동당은 통신 법률 개정안(허위 정보(disinformation) 및 오정보(misinformation) 방지)’ 초안을 발표하여, 허위조작정보 및 오정보 확산을 강력하게 플랫폼의 책임으로 돌리는 추가적인 규제를 마련하려 하고 있다. 다만 해당 법안은 허위조작정보의 규제와 표현의 자유의 보호 사이에서 균형을 찾지 못한다는 평가를 받고 있기도 하다.

제3절 소결

온라인 플랫폼을 통한 허위조작정보의 확산이 전 세계적으로 문제시되고 있는 만큼 한국 사회에서도 EU와 호주의 실천강령이 갖는 의미는 적지 않다. 특히 허위조작정보의 유포를 억제하기 위한 공식적인 대응체계를 플랫폼의 자율규제를 중심으로 설계했다는 점에서 더욱 그러하다. 한국사회에서는 학술적으로나, 규범적으로 ‘규제’를 법적 규제에 기반하여 논의하는 경우가 많기 때문에(황성기, 2022), 허위조작정보가 초래하는 문제를 해결하기 위한 방안으로 자율규제 정책이 충분히 구체화되지 못한 실정이다. 이에 학술적인 장에서는 플랫폼과 시민사회, 학술 공동체, 그리고 이해관계자의 일원으로서 정부가 참여하는 공동규제로서의 자율규제가 합리적인 대안으로 제시되고 있다(김민정, 2020; 박아란, 2019; 황용석·권오성, 2017). 이 맥락에서 EU와 호주가 정책형성 및 강령 제정 과정에서 다양한 이해관계자들을 참여시키고 이들을 중심으로 합의를 이끌어내기 위해 두터운 협의

를 진행하는 것은 눈여겨볼 필요가 있다. 이러한 정책형성과정은 다양한 주체들의 참여로 그 과정 자체가 매우 복잡하고 까다롭지만, 일단 합의가 이루어진 이후의 정책 실현 단계는 비교적 순조로운 편에 속한다. 자율 규제 방식이 효과적으로 운영될 수 있었던 것도 이 때문이라 할 수 있겠다. 이에 실효성 있는 정책 형성을 위해 국내에서는 EU와 호주의 경우를 참고할 필요가 있겠다. 다양한 이해관계자의 참여를 적극적으로 활용한 협의 및 공론화 과정 등 기존의 방식과는 다른 새로운 접근을 고려할 때, 사회 전반의 자발적인 관심과 참여를 통한 효과적인 정책 결과를 끌어낼 수 있을 것으로 기대된다.

한편 국내에서의 허위조작정보 대응에서 자율규제 방식을 고도화하기 위한 논의도 적극적으로 진행할 필요가 있다. 자율규제의 경우 강제력이 없다는 점에서 정책 효과가 낮다는 단점이 있지만, EU 강령의 경우 공동규제라는 방식을 통해 규제의 실효성을 제고했다는 평가를 받는다. 이를 바탕으로 김민정(2020)은 국내에 허위조작정보 공동규제 체계를 적용하기 위해 다음과 같은 사항을 제안한다.

<표 4-17> EU의 자율규제 사례 도입을 위한 국내 제안 사항

구분	내용
정부	<ul style="list-style-type: none"> · 국내외 플랫폼 사업자들이 모여 자율규제(안)을 만들 수 있는 자리 마련 · 자율규제(안) 형성 이후 이행 현황에 대한 관리·감독 방안 마련
사업자	<ul style="list-style-type: none"> · 허위조작정보 제공자에 대한 금전적 보상 중단 방안 마련 및 시행 · 후원/지불된(paid) 콘텐츠와 독립적인 콘텐츠에 대한 구분 및 제시 · 자동화 시스템 활용 조작에 대한 대응을 위한 기술 장치 개발 · 양질의 정보 제공을 위한 방안 마련 및 시행 · 투명성 증대 방안 마련 및 시행 · 다른 이해관계자들과의 협력 증진 및 활동 지원

*자료: 김민정 (2020), 223-226쪽에서 재구성

물론 EU나 호주의 허위조작정보 대응을 위한 강령을 그대로 국내에 적용하는 것에는 한계가 있을 것이다. 그럼에도 해당 강령의 자율규제 방식은 허위조작정보 대응에 대해 선도적으로 논의하고 있을 뿐만 아니라 구체적인 실질적인 방안들을 도출하기 위해 노

력하고 있고, 참여자, 특히 글로벌 사업자들의 협조와 실천을 이끌어 냈다는 점에서 유용한 참고 사례가 된다(김민정, 2020). 이에 국내의 특수 상황에 대한 고려 또한 필요하겠으나 각 사례에 대한 충분한 검토 및 논의를 통해 국내 적용 가능한 자율규제 실천강령을 마련해봄으로써 공동규제의 틀을 모색해보는 것은, 실효성 있는 허위조작정보 대응 정책 도출을 위한 기초 작업으로서 의미를 가질 수 있을 것으로 판단된다.

제 5 장 허위조작정보 관련 국내 자율규제 활성화 방안

제 1 절 국내 허위조작정보 관련 법적 규제 현황

1. 허위조작정보 관련 법적 대응

국내에서는 허위조작정보를 구체적으로 규정하는 법령은 없다. 다만 허위의 사실 표현이 개인의 권리를 침해하는 경우 등과 관련해서는 크게 허위조작정보를 전달·유포·게시하는 행위를 규제하는 법률과 플랫폼 사업자를 규제하는 법률이 있다. 그 구체적인 내용은 다음과 같다.

〈표 5-1〉 허위조작정보 관련 법적 규제 현황

허위조작정보 전달·유포·게시 등 처벌	플랫폼 사업자의 삭제, 차단, 취급거부·정지·제한조치
<ul style="list-style-type: none"> · 명예훼손죄(형법 제307조, 308조, 309조, 정보통신망법 제70조) · 모욕죄(형법 제311조) · 허위사실공표죄(공직선거법 제250조) · 후보자비방죄(공직선거법 제251조) · 전기통신기본법 제47조제2항의 벌칙규정 	<ul style="list-style-type: none"> · 게재된 정보에 대한 삭제·임시조치(정보통신망법 제44조의2제3항, 4항) · 게재된 정보에 대한 삭제 또는 취급의 거부·정지·제한(공직선거법 제82조의4제3항) · 전기통신법 제47조제4항의 벌칙규정

* 자료 : 김연진 (2021). 허위조작정보 규제에 관한 헌법적 연구

이에 근거해 허위의 사실 표현과 관련하여 개인의 권리가 침해되는 경우 지금까지는 해당 법률로서 표현의 자유를 일정정도 규제해왔다. 이 같은 대응은 다음과 같이 정리할 수 있다(김민정, 2019).

<표 5-2> 허위사실 표현이 초래하는 피해와 법적 대응

피해발생		법적 대응		
구분	내용	민사적 대응	형사적 대응	행정적 대응
개인	명예훼손 모욕	손해배상 계재중단	모욕죄 명예훼손죄 선거법위반	인터넷심의 및 시정요구 처리제한조치 임시조치 선거법상 조치
사회·국가	유언비어			

*자료: 김민정(2019). 가짜뉴스에서 허위조작정보로

피해 발생과 관련된 부분 중 유언비어에 의해 사회·국가가 피해를 받는 경우에 대해서는 현행 법률로 대응이 불가능하다. 그런데 전기통신기본법 제47조제1항 위헌결정의 취지에 비추어 볼 때 유언비어를 규제하는 법률은 헌법적 정당성을 지니는 형태로 입법되기는 어렵다고 할 수 있다(김민정, 2019).

법적 대응 부분에서 특히 플랫폼 사업자와의 연관 속에서 살펴볼 부분은 정보통신망법에 따른 게재된 정보에 대한 삭제 및 임시조치 등이다. 이는 삭제 등 요청 시 플랫폼 사업자(정보통신서비스제공자)에게 필요한 조치를 취하고, 그 사실을 알리며, 관련 내용과 절차 등을 미리 약관에 구체적으로 밝히도록 규정하고, 적절한 조치를 취한 사업자에게 이로 인한 손해배상책임을 면책한다.

한편 인터넷 심의 및 시정요구 등 행정적인 대응은 방송통신심의위원회의 통신심의 절차에 따라 이루어진다. 심의 대상이 되는 정보는 권리침해 정보 외에도 불법정보 및 유해정보가 해당한다. 인터넷심의 대상이 되는 불법정보는 「정보통신망법」 제44조의7에 규정된 불법정보를 가리키며, 유해정보는 「청소년보호법」에 따른 청소년 유해매체물 등 청소년에게 유해해 심의가 필요하다고 인정되는 정보를 의미한다. 정보통신망법상 불법정보는 ① 음란, ② 명예훼손, ③ 사이버스토킹, ④ 해킹, 바이러스 유포, ⑤ 청소년 유해매체물 표시의무 위반, ⑥ 도박 등 사행행위, ⑥의 2, 개인정보 거래, ⑥의3, 불법무기류 정보, ⑦ 국가기밀 누설, ⑧ 국가보안법 위반, ⑨ 범죄 관련 정보 등이 해당한다.

이와 같은 조치들은 불법 및 유해정보, 권리 침해 정보에 대하여 이루어지고 있으며 직

접적으로 허위조작정보의 확산 억제를 겨냥하지 않는다. 다만 정보통신망법에서 플랫폼 사업자가 요청에 의해 삭제 및 임시조치 등을 취할 때, 관련 내용과 절차를 미리 약관에 구체적으로 밝히도록 하고, 조치 사실을 알리도록 하는 것은 이용자를 보호하고 투명성을 확보한다는 점에서 의의가 있다고 할 수 있다. 한편 정보통신망법은 플랫폼 사업자에게 불법촬영물 등과 관련해 투명성 보고서 제출의무, 정보 책임자 지정 등의 의무를 규정하고, 일정 규모 이상의 해외사업자에게도 국내 대리인 지정 의무를 부여한다. 이는 EU의 디지털서비스법, 영국의 온라인 안전법 등이 입법적 논의를 통해 사업자에게 요구하는 책임의 일부가 어느정도는 국내 법체계에 반영되어 있음을 보여주는 것이기도 하다.

<표 5-3> 정보통신망법에 따른 플랫폼 사업자 책임

정보통신망법	디지털서비스법/온라인 안전법
<ul style="list-style-type: none"> · 제64조의5(투명성보고서 제출의무 등) : 일정 규모 이상 정보통신서비스 제공자는 불법촬영물 등 처리에 관한 투명성 보고서를 방통위에 제출 	<ul style="list-style-type: none"> · 온라인 플랫폼은 허위조작정보(불법 유희정보) 대응 관련 투명성 보고서 작성
<ul style="list-style-type: none"> · 제42조의3(정보책임자 지정 등) : 일정 규모 이상의 정보통신서비스제공자는 청소년보호책임자 지정 · 44조의9(불법촬영물 등 유통방지 책임자 지정) : 일정 규모 이상의 정보통신서비스제공자는 불법촬영물 등 유통방지 책임자 지정 	<ul style="list-style-type: none"> · 규정준수 책임자 지정(EU) · 신고 및 불만처리 책임자 지정(영국)
<ul style="list-style-type: none"> · 제32조의5(국내 대리인의 지정) : 일정 규모 이상의 해외사업자는 국내 대리인 지정 	<ul style="list-style-type: none"> · 연락창구 지정(EU) · OFCOM 통지에 대한 답변 제공자 지정(영국)

2. 허위조작정보 관련 주요 입법 논의

허위조작정보에 대한 우려가 확산되면서 국내에서도 이를 입법적인 방식으로 규제하기 위한 논의가 활발하게 이루어지고 있다. 21대 국회에서도 플랫폼 사업자에게 책임을 부여하기 위한 법률 제개정안이 다수 발의되어 있다. 여기에서는 정보통신망법 일부개정안 중심으로 일부 법률안을 간략히 정리해보도록 한다(박영흠, 2023).

<표 5-4> 제21대 국회에 제출된 허위조작정보 관련 주요 법률안 : 정보통신망법 개정안

대표발의	주요 내용
2023.9.27. 윤두현 의원	<ul style="list-style-type: none"> · 정보통신서비스 제공자에게 허위조작정보 유통을 방지해야 할 의무와 허위조작정보 유통 방지 책임자 지정의무 부과 · 허위조작정보를 정보통신망에서 유통이 금지되는 불법 정보에 포함 · 허위조작정보로 타인의 권리가 침해되는 경우 해당 사실을 인식한 사람은 누구든 해당 정보의 삭제·반박 내용의 게재 요청 가능 · 매크로 프로그램을 이용한 허위조작정보 게재행위 금지
2022.4.27. 김의겸 의원	<ul style="list-style-type: none"> · 인터넷뉴스서비스사업자에게 독자가 검색한 결과로 기사를 제공 또는 매개하는 경우와 기사를 제공하는 신문 등이 직접 선정하여 배열한 기사를 제공 또는 매개하는 경우에만 인터넷뉴스서비스를 할 수 있도록 함 · 뉴스를 제공하거나 매개할 때 이용자 개인이 정하는 규칙이 적용될 수 있는 기술적 조치를 마련함 · 이용자가 뉴스를 제공하는 정보통신서비스제공자인 해당 언론사의 홈페이지를 통해서만 뉴스를 열람할 수 있는 기술적 조치를 마련함
2022.2.8. 정필모 의원	<ul style="list-style-type: none"> · 정보통신서비스 제공자가 특정 이용자의 검색기록, 구매내역 등 정보를 수집해 해당 이용자에게 대한 맞춤형 정보 추천 서비스를 제공하는 알고리즘을 이용하려는 경우에는 이용자의 동의를 받도록 하고, 이용자가 동의하지 아니하는 경우에는 해당 이용자의 이용정보를 수집하지 못하도록 함

대표발의	주요 내용
2021.11.11. 김의겸 의원	<ul style="list-style-type: none"> · 진실이 아닌 내용을 담은 언론기사 등으로서, 언론중재법에 따라 정정보도 등이 있는 경우 피해자가 정보통신서비스 제공자에게 이를 삭제할 수 있도록 요청할 수 있는 권리를 부여함
2021.6.25. 류호정 의원	<ul style="list-style-type: none"> · 알고리즘의 설계 또는 적용과 관련해 발생하는 갈등과 분쟁의 조정 등을 위해 방통위 산하에 알고리즘분쟁조정위원회를 둠 · 또한 방통위 산하에 알고리즘투명성위원회를 설치하고, 알고리즘이 서비스 이용자의 건강에 명백한 해를 끼칠 우려가 있는 등의 경우에 해당 알고리즘을 조사할 수 있도록 하고 필요한 경우 알고리즘 서비스 제공자에게 시정명령을 내릴 수 있는 근거를 마련함 · 알고리즘 서비스 이용자 등이 알고리즘 서비스 제공자에게 알고리즘에 대한 설명을 요구할 수 있도록 하고, 알고리즘 서비스 제공자는 영업 비밀을 이유로 이를 거부할 수 없도록 함
2021.3.19. 이원욱 의원	<ul style="list-style-type: none"> · 기업의 영업비밀을 침해하지 않는 범위 내에서, 이용자 정보의 안전성과 신뢰성에 직접적으로 영향을 미치는 경우에 한하여 정보통신서비스제공자가 보유하고 있는 일정 범위의 알고리즘 관련 내용을 과기부장관과 방통위장관에게 의무적으로 제출하도록 함
2020.6.22. 정필모 의원	<ul style="list-style-type: none"> · 정보통신망을 통해 상업적 또는 정치적으로 정보를 매개로 타자를 속이려는 기만적 의도성을 가진 행위로 수용자가 허구임을 오인하도록 언론보도의 양식을 띤 정보 또는 사실검증이라는 저널리즘의 기능이 배제된 가운데 검증된 사실로 포장하는 행위를 허위조작정보로 정의 · 정보통신서비스제공자는 자신이 운영·관리하는 망에 허위조작정보가 게재되어 있을 경우 지체 없이 삭제해야 하며, 이를 이행하지 않을 경우 과태료를 부과

* 자료 : 박영흠(2023). EU의 이스라엘-하마스 전쟁 관련 허위정보 대응과 시사점.

3. 자율규제의 법적 근거

한편 정보통신망법은 정보통신서비스 제공자단체를 통하여 자율규제를 지원하고, 사업자가 자율적으로 자율규제 행동강령을 정하여 시행할 수 있는 근거를 마련해놓고 있다. 이는 정보통신서비스 제공자의 책임성을 높이고, 민간의 자율규제를 촉진하기 위한 것이 주된 목적이다. 그 주요 내용은 다음과 같다.

<표 5-5> 정보통신망법 상 자율규제 관련 조항

구분	주요 내용
정부지원	<ul style="list-style-type: none"> · 제3조(정보통신서비스제공자 및 이용자의 책무) : 정부는 정보통신서비스 제공자 단체 또는 이용자 단체의 자율규제를 위한 활동을 지원할 수 있도록 함 · 제44조의4(자율규제) : 정부는 정보통신서비스 제공자 단체의 행동강령 지정 및 시행에 대한 지원과 불법 및 청소년 유해정보 자율규제 활동을 지원할 수 있도록 함
사업자 자율규제	<ul style="list-style-type: none"> · 제44조의4(자율규제) : 사업자 단체는 자율규제 행동강령을 정할 수 있음 : 사업자 단체는 불법 및 청소년 유해정보에 대한 자율규제 활동을 시행할 수 있음

* 자료 : 김희경 외 (2022). 인터넷 공동규제 활성화 방안 연구.

그러나 이와 같은 조항은 실질적으로는 선언적인 규정에 그쳐 사업자들의 자율규제 참여에 유인이 되고 있다고 보기 어렵다는 비판도 있다. 해당 조항은 사업자 단체가 자율규제 조항을 할 수 있는 간접적인 근거를 제공하고 있지만, 자율규제에 대한 구체적인 방식과 내용, 정보통신서비스 제공자 단체의 실체 등에 대해서는 별도로 규정하고 있지 않기 때문이다(김희경, 2022).

제 2 절 국내 허위조작정보 관련 사업자 자율규제 사례

1. 한국인터넷자율정책기구

한국인터넷자율정책기구(Korea Internet Self-governance Organization, KISO)는 2008년 설립된 사업자들의 민간 자율규제기구다. 출범 당시 회원은 NHN, 다음커뮤니케이션, SK커뮤니케이션즈, KTH, 야후코리아, 프리퀄, 하나로드림 등 7개 포털사로 시작됐으며, 현재는 네이버, 카카오, SK커뮤니케이션즈 등 14개 인터넷 포털 업체가 회원사로 구성되어 있다.

KISO는 미국이나 유럽과 달리 인터넷 자율규제에 대한 사회적 공감대와 인식이 아직 미흡한 상황에서 시장이 주도해 설립한 한국 최초의 민간 자율규제기구라는 점에 의의가 있다. 특히 재원 측면에서도 민간 사업자의 지원과 일부 수익사업을 통해 충당하고 있어 순수 민간 자율규제를 구현하고 있다(KISO, 2016.4.). 다만 해외 대규모 플랫폼이 참여하지 못하고 있다는 점은 한계로 꼽히기도 한다(김유향, 2022).

[그림 5-1] KISO의 주요 기능



* 자료 : KISO(2016.4.)

KISO는 정관 제6장에 의해 설치된 정책위원회를 통해 정책결정과 심의결정을 주요 업무로 하고 있다. 특히 2018년부터는 가짜뉴스 신고센터를 두고, 기존의 불법 유헨정보 중심의 자율규제에서 허위조작정보로까지 범위를 확대하고 있다.

보다 구체적으로 KISO에서 허위정보와 관련하여 규제하고 있는 부분은 가짜뉴스이다. 이는 KISO의 정책 규정 제5장 제5절에서 확인할 수 있다. KISO의 <정책 규정> 제5장 제5절은 ‘정책언론보도 형식의 허위 게시물 관련 정책’ 으로 해당 규정은 2018년 3월 23일에 신설되었으며 5월부터 시행되었다.

<표 5-6> KISO <정책규정> 제5장 제5절

구분	내용
제33조(목적)	본 절에서는 인터넷 공간의 신뢰성을 높이고 공익을 보호하기 위해 언론의 기사형식을 도용 또는 사칭한 허위 게시물에 대한 처리 기준을 마련하는데 그 목적이 있다
제34조(게시물 제한)	<p>① 회원사는 다음 각 호의 요건을 모두 충족하는 게시물이 유통되고 있음을 신고 등을 통해 알게 된 경우 이에 대해 삭제 등 필요한 조치를 취할 수 있다. 다만, 게시물이 창작성과 예술성이 인정되는 패러디나 풍자에 해당하는 경우에는 그러하지 아니하다</p> <p>1. 언론사의 명이나 언론사의 직책 등을 사칭 또는 도용하여 기사 형태를 갖춘 게시물의 경우 2. 게시물의 내용이 허위로 판단되는 경우</p> <p>② 회원사는 제1항의 요건의 적용 여부에 대해 필요하다고 인정하는 경우 KISO에 심의를 요청할 수 있다.</p> <p>③ 회원사는 제1항의 규정에도 불구하고, 정책규정 제2장 및 제4장에 해당하는 게시물의 경우 해당 규정에 따라 처리한다.</p>

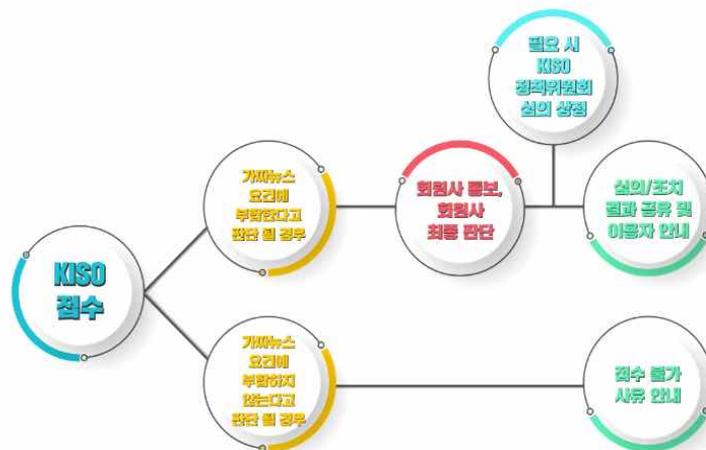
자료: KISO 정책규정

KISO 회원사의 웹 사이트에 게재된 어떤 게시물이 언론사의 명이나 언론사의 직책 등을

사칭 또는 도용하였고 기사 형태를 갖추었으며 그 게시물의 내용이 허위로 판단될 경우, 이를 가짜뉴스라고 지칭하고 있다. 그러나 KISO 회원사의 웹 사이트에 게재된 게시물이라 하더라도 다음 4가지 경우 중 하나에 해당할 때에는 KISO가 규정하는 가짜뉴스에 해당하지 않는다. 첫 번째는 언론의 오보인 경우, 두 번째는 언론사의 명이나 직책 등을 사칭 또는 도용하여 기사 형태를 갖춘 게시물이지만 게시물의 내용이 허위가 아닌 경우, 세 번째는 기사 형태가 아닌 경우, 네 번째는 공개 게시판 등 일반 이용자가 접근 가능한 공간에 게시되지 않은 경우이다.

또한 KISO에서 접수할 수 있는 가짜뉴스는 KISO의 회원사가 운영 중인 사이트에서 게재된 게시물에 한정되며, KISO의 회원사가 운영하지 않는 사이트에 게시된 가짜뉴스의 경우 그 내용에 따라 방송통신심의위원회, 언론중재위원회 혹은 중앙선거관리위원회 신고 지침이 등 관련 기관에 신고가 가능하다. 이러한 가짜뉴스에 대해 최종 판단은 회원사가 진행하며, 회원사가 자체적으로 판단하기 모호한 게시물의 경우 KISO 정책위원회에 상정하여 판단을 구할 수 있다. KISO는 신고된 게시물이 정책규정 제34조 제1항에서 정한 가짜뉴스 요건에 해당하는 지 검토한 후 그 결과를 회원사에 전달한다.

[그림 5-2 가짜뉴스 신고 처리절차 안내]



* 자료 : KISO 가짜뉴스 신고센터

현재 KISO는 가짜뉴스, 유해 게시물, 명예훼손 게시물, 검색어, 부동산 매물에 대해 종합 신고센터를 운영하고 있다. 이는 KISO가 각 회원사와 협력하여 수행하는 신고 처리 기능을 종합한 플랫폼으로, 가짜뉴스 신고센터도 해당 플랫폼에 포함되어 있다.

이용자는 KISO 회원사 웹 사이트에 게재된 게시물에 대해 가짜뉴스 신고센터를 이용하여 신고할 수 있으며, KISO의 가짜뉴스 신고 처리절차는 위의 그림과 같다. 이용자는 가짜뉴스 신고센터에서 허위 게시물을 신고하는 경우 회원사, 개인정보, 신고주소(URL), 신고 내용, 증거자료를 기입하여 신고처리를 완료할 수 있다.

한편 KISO는 2021년 4월 15일 ‘코로나19 관련 허위조작정보에 관한 정책’을 발표해, 코로나19와 관련된 허위조작정보에 대해서도 자율규제를 실시했다. 해당 내용은 일반규정이 아닌 한정적 특별규정으로 마련되어 2023년 3월 31일자로 종료되었다. KISO의 ‘코로나19 관련 허위조작정보에 관한 정책’의 주요내용은 다음과 같다.

<표 5-7> 코로나19관련 허위조작정보에 관한 정책

코로나19 관련 허위조작정보에 관한 정책

① 회원사는 코로나19에 관한 정보 중 아래 요건을 모두 충족하는 게시물이 유통되고 있음을 신고 등을 통해 알게 된 경우 이에 대해 삭제 등 필요한 조치를 취할 수 있다. 단, 이 정책은 코로나19 관련 게시물에 한하여 제한적으로 적용되어야 하며, 표현의 자유의 가치와 중요성을 염두에 두어야 한다.

1. 코로나19의 존재, 치료, 예방 및 진단, 감염, 사회적 거리두기 및 자가격리와 관련한 내용의 게시물인 경우
2. 세계보건기구(WHO) 또는 질병관리청에 의해 허위조작정보임이 명백히 확인된 경우

② 회원사는 제1항의 요건의 적용 여부에 대해 필요하다고 인정되는 경우 KISO에 심의를 요청할 수 있다

③ 회원사는 제1항의 규정에도 불구하고, 정책규정 *제2장 및 **제4장에 해당하는 게시물의 경우 해당 규정에 따라 처리한다.

*제2장: 게시물에 관한 정책

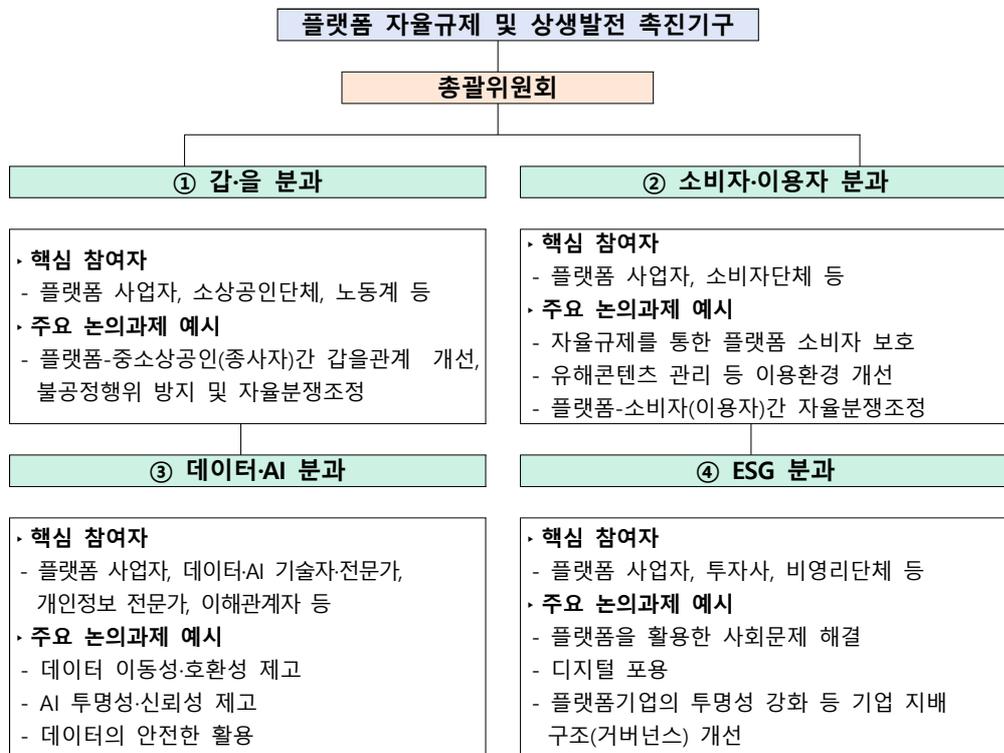
**제4장: 선거 관련 인터넷 정보서비스 기준에 관한 정책

자료: 황용석(2021)

2. 플랫폼 민간 자율기구

플랫폼 민간 자율기구는 2022년 9월 19일 민간 주도로 플랫폼 자율규제 방안 마련을 논의하기 위해 출범했다. 기구에는 한국인터넷기업협회·소상공인연합회·한국소비자단체협의회 등 사업자입점업체·소비자·종사자 등과 관련한 주요 협회·단체와 기업이 참여하며, 정부(공정위, 과기부, 방통위)도 정책적으로 이를 지원한다. 자율기구는 ① 갑을분과, ② 소비자·이용자 분과, ③ 데이터·AI분과, ④ ESG 분과로 구성되어 있다.

[그림 5-3 플랫폼 민간 자율기구 구성도



특히 데이터·AI 분과는 2023년 5월 자율규제 방안 발표회에서 ‘검색·추천 서비스의 투

명성 제고를 위한 자율규제 원칙’ 을 마련하고 발표했다. 이는 (i) 검색·추천 기준 공개, (ii) 이용자의 권익 증진을 위한 노력, (iii) 이 등 실효성 확보 등 세 가지 방안으로 구성되어 있다. 각각의 구체적인 내용은 다음과 같다(관계부처 합동 보도자료, 2023.5.11.).

<표 5-8> 검색·추천 서비스의 투명성 제고를 위한 자율규제 원칙

구분	주요 내용
<p>검색·추천 기준 공개</p>	<ul style="list-style-type: none"> · 사업자는 검색노출순서 결정 및 추천 기준을 구성하는 주요 변수와 그에 관한 설명 등을 이용자가 쉽게 이해할 수 있도록 평이하고 알기 쉬운 용어로 사용하여 공개해야 함 · 공개 방법에 있어서도 검색·추천결과, 고객센터 등 이용자가 접근하기 쉬운 위치와 방식으로 공개해야 하며, 변경사항이 있는 경우 이를 현행화하여야 함 · 다만 제3자에 의해 이용자에게 기만 또는 피해가 발생할 우려가 있거나 사업자의 영업비밀을 침해할 우려가 있는 경우 이를 부분적으로 공개하지 않을 수 있음
<p>이용자의 권익 증진을 위한 노력</p>	<ul style="list-style-type: none"> · 사업자는 공개된 검색노출 결정 및 추천기준 정보에 대해 이용자가 설명을 요구하는 경우 성실히 답변하여야 하며, 이용 사업자의 수수료, 광고료 등 대가 지불 여부가 노출 결과에 영향을 미치는 경우 이를 이용사업자에게 설명해야 함 · 사업자는 검색노출 순서 결정 및 추천 기준이 목적인 대로 동작하는지 의도하지 않은 결과를 도출하는지 등을 주기적으로 점검하여야 함
<p>이행점검 등 실효성 확보</p>	<ul style="list-style-type: none"> · 사업자는 플랫폼 민간 자율기구의 본 자율규제 원칙 이행여부 점검 시 협조하며, 이행점검 결과 시정을 권고받은 사업자는 관련 사항을 확인하고 개선 및 조치결과를 플랫폼 민간 자율기구에 신속하게 알려야 함 · 또한 데이터·AI 분과는 기술서비스 발전 및 환경변화, 사업자·전문가·이해관계자 등의 요청에 따라 자율규제 원칙의 내용 변경이 필요한 경우 개선 등 조치를 할 수 있음

* 자료: 관계부처 합동(2023.5.11.). 플랫폼 자율규제 방안 발표회 개최

한편 혁신공유-거버넌스 분과에서는 2022년 12월 ‘플랫폼의 사회적 가치 제고를 위한 8대 원칙’ 을 구성하여 발표했는데, 여기에는 ‘누구나 안심하고 가치를 향유할 수 있도록 안전성 제고’ 라는 원칙 아래 “불법·유해 콘텐츠 안전 및 프라이버시 보호”, “허위 정보, 위해상품 및 사기거래 방지” 등이 포함되어 있다.

[그림 5-4 플랫폼의 사회적 가치 제고를 위한 8대 원칙



* 자료: 관계부처 합동(2023.5.11.). 플랫폼 자율규제 방안 발표회 개최

3. 네이버

(1) 허위조작정보 관련 콘텐츠 운영정책

네이버는 허위조작정보와 관련하여 (사)한국인터넷자율정책기구(이하 “KISO”)의 <정책규정>을 준수하여 게시물 규제를 실시하고 있다. 네이버 게시물 운영정책에 안내된

KISO의 <정책규정>을 준수하고 있는 허위조작정보 관련 규제 내용은 다음과 같다.

<표 5-9> 네이버의 허위정보 관련 콘텐츠 운영정책

구분	규제 내용	관련 근거
다른 이용자의 안전 확보 측면	<ul style="list-style-type: none"> • 코로나19의 존재, 치료, 예방 및 진단, 감염, 사회적 거리두기 및 자가 격리와 관련한 내용 • 세계보건기구(WHO) 또는 질병관리청에 의해 허위조작정보임이 명백히 확인된 게시물의 경우 	<ul style="list-style-type: none"> • 한국인터넷자율정책기구(KISO)의 정책결정
다른 이용자의 존중 측면	<ul style="list-style-type: none"> • 다른 이용자를 기만할 목적으로 타인을 사칭하거나 허위사실을 주장하는 내용의 게시물 • 언론사의 명이나 언론사의 직책 등을 사칭 또는 도용하여 기사 형태를 갖춘 허위 게시물 	<ul style="list-style-type: none"> • 한국인터넷자율정책기구(KISO)의 정책규정 준수 <ul style="list-style-type: none"> - 가짜뉴스 <언론보도 형식의 허위 게시물 관련 정책>

* 출처: 네이버 운영정책

이처럼 네이버 운영정책을 통해 허위조작정보와 관련하여 규제되는 게시물은 ‘언론사의 명이나 언론사의 직책 등을 사칭 또는 도용하여 기사 형태를 갖춘 허위 게시물’ (이하, 가짜뉴스)과 KISO의 정책결정에 근거한 코로나19와 관련된 허위조작정보 게시물로 파악된다. 또한, 코로나19와 관련된 허위조작정보에 해당하는 기준은 네이버 신고센터를 통해 안내하고 있으며, 해당하는 내용은 다음과 같다.

<표 5-10> 코로나19 허위조작정보에 해당하는 경우

내용 분류	예시
코로나19의 존재에 대한 허위조작정보	<ul style="list-style-type: none"> • 코로나19 팬데믹은 기획된 것으로, 기획자는 현재 세계를 지배하는 딥스테이트다

코로나19 치료 관련 허위조작정보	<ul style="list-style-type: none"> • 안티푸라민을 바르면 코로나19 바이러스가 호흡기에 들어오지 못한다 • 이산화염소를 발생시키는 ‘코로나 예방용 목걸이’가 효과적이다
코로나19 예방 및 진단 관련 허위조작정보	<ul style="list-style-type: none"> • 코로나19 백신에 유전자 변형 DNA와 신경 독소 발암물질이 포함돼 있다 • 코로나19 백신 접종을 거부할 경우 긴급체포 된다 • 코로나19 검사는 환자를 가려내기 위한 것이 아니라 다른 용도로 활용되고 있다
코로나19 전염 관련 허위조작정보	<ul style="list-style-type: none"> • 코로나19는 성병이 원인이다 • 코로나19 치사율은 사스 대비 10배에 달한다 • ‘중국산 김치’ 먹으면 코로나19에 걸린다
사회적 거리두기 및 자가격리 가이드라인 관련 허위조작정보	<ul style="list-style-type: none"> • 마스크 상시 착용은 암을 유발한다

*출처: 네이버 신고센터

네이버는 운영정책을 위반한 게시물에 대해 해당 게시물을 비공개 또는 삭제 처리하거나 게재를 거부하는 방법으로 규제를 가하고 있다. 또한, 운영정책을 위반한 유해게시물에 대해 현재 네이버는 (i) AI 모니터링, (ii) 모니터링 및 신고 전담부서 운영, (iii) 이용자의 직접 신고와 같은 방안을 통해 게시물을 관리하고 있으며, 코로나19 관련 허위조작정보의 경우 예외적으로 KISO의 가짜뉴스 신고센터를 이용하도록 안내하고 있다.

(2) 이용자의 허위조작정보 신고 방안

현재 네이버는 운영정책을 위반한 게시물 및 댓글의 신고방법을 안내하는 네이버 신고센터를 운영하고 있으며, 유해게시물의 경우 게시물 페이지 내 신고버튼을 통해 이용자가 직접 신고가 가능하다. 네이버 유해게시물 신고 시 안내되고 있는 신고 사유는 다음과 같다.

<표 5-11> 유해게시물 신고 사유

신고 사유	내용
스팸홍보/도배글	<ul style="list-style-type: none"> • 사행성 오락이나 도박을 홍보하거나 권장하는 내용 등의 부적절한 스팸 홍보 행위 • 동일하거나 유사한 내용 반복 게시
음란물	<ul style="list-style-type: none"> • 성적 수치심을 일으키는 내용 • 아동이나 청소년을 성 대상화한 표현 • 과도하거나 의도적인 신체 노출 • 음란한 행위와 관련된 부적절한 내용
불법정보 포함	<ul style="list-style-type: none"> • 불법 행위, 불법 링크에 대한 정보 제공 • 불법 상품을 판매하거나 유도하는 내용
청소년 유해게시물	<ul style="list-style-type: none"> • 가출/왕따/학교폭력/자살 등 청소년에게 부정적인 영향을 조성하는 내용
욕설/생명경시/혐오/차별적 표현	<ul style="list-style-type: none"> • 직·간접적인 욕설을 사용하여 타인에게 모욕감을 주는 내용 • 생명을 경시여기거나 비하하는 내용 • 계층/지역/종교/성별 등을 혐오하거나 비하하는 표현 • 신체/외모/취향 등을 경멸하는 표현
개인정보 노출	<ul style="list-style-type: none"> • 법적으로 중요한 타인의 개인정보를 게재 • 당사자 동의 없는 특정 개인을 인지할 수 있는 정보
불쾌한 표현	<ul style="list-style-type: none"> • 불쾌한 표현 포함(해당 사유는 클린봇 학습에 도움이 될 수 있음)

* 출처: 네이버 신고센터

위와 같이 네이버는 운영정책을 위반한 게시물에 대해 이용자가 직접 게시물을 신고할 수 있도록 신고 기준을 제공하고 있다. 현재 게시물 운영정책에서 규제 대상인 ‘가짜뉴스’에 대해서는 따로 마련된 신고 기준은 없는 것으로 파악되나, ‘가짜뉴스’ 게시물에 대한 네이버의 대응 절차는 KISO의 ‘가짜뉴스 신고센터’를 통해 간접적으로 확인할 수 있다. 가짜뉴스 신고센터 홈페이지에서 제공되고 있는 신고 처리 절차 안내에 따르면, 회사(포털 사업자)에 접수된 이용자 신고가 KISO로 이관되어 가짜뉴스 요건을 충족하는지 판단하고 내부절차에 따라 처리하여 신고자에게 안내 및 통보되는 것으로 파악된다. 해당

홈페이지를 통해 안내되는 가짜뉴스 신고 처리 절차는 다음과 같다.

[그림 5-5 KISO 가짜뉴스 신고 처리 절차



*출처 : KISO 홈페이지(<http://kiso.helloweb.co.kr/fakenews/>)

또한, 코로나19와 관련된 허위조작정보를 신고하는 경우에는 네이버 신고센터를 통해 KISO의 가짜뉴스 신고센터를 이용할 수 있도록 직접 해당 링크를 제공하고 있다. 이를 통해 코로나19와 관련된 허위조작정보도 위의 가짜뉴스 게시물과 동일한 방법으로 처리하고 있음을 알 수 있다.

이와 같이 네이버는 가짜뉴스 혹은 코로나19와 관련된 허위조작정보에 대한 게시물에 대해 KISO로 해당 게시물을 이관하여 각 게시물에 대한 규제 여부를 결정하는 것으로 파악된다.

4. 카카오(다음)

(1) 허위조작정보 관련 콘텐츠 운영정책

카카오의 포털 서비스(다음)에서 허위조작정보 규제와 관련한 항목은 네이버와 동일하게

가짜뉴스 및 코로나19 관련 허위조작정보인 것으로 파악된다. 가짜뉴스의 경우, 카카오의 운영정책 내 서비스 이용 시 금지하는 활동으로 규정되어 있으며, 그 내용은 다음과 같다.

<표 5-12> 카카오 운영정책 중 가짜뉴스 관련 내용

구분	내용
서비스 이용 시 금지하는 활동	<ul style="list-style-type: none"> (22) 언론사의 명이나 언론사의 직책 등을 사칭 또는 도용하여 기사 형태를 갖춘 게시물 중 그 내용이 허위로 판단되는 게시물을 게시하는 행위

* 자료: 카카오 운영정책

또한, 코로나19와 관련한 허위조작정보 규제에 대해서는 다음 고객센터의 이용자 보호 게시판을 통해 관련 정책에 관한 게시물을 공유하고 있다. 해당 게시물에서는 KISO의 정책위원회('21.4.15)에서 의결된 코로나19 관련 허위조작정보에 관한 정책이 마련된 배경 및 정책기준을 설명하고 있다. 해당 게시물에서 안내하는 코로나19 관련 허위조작 정보에 관한 정책기준은 다음과 같다.

[그림 5-6 다음 고객센터 내 '코로나19 관련 허위조작정보에 관한 정책'



* 자료: 다음 고객센터(<https://cs.daum.net/faq/2068/61709.html#38344>)

〈표 5-13〉 코로나19 관련 허위조작정보에 관한 정책

구분	내용
정책 기준	<p>①회원사는 코로나19에 관한 정보 중 아래 요건을 충족하는 게시물이 유통되고 있음을 신고 등을 통해 알게 된 경우 이에 대해 삭제 등 필요한 조치를 취할 수 있다. 단, 이 정책은 코로나19 관련 게시물에 한하여 제한적으로 적용되어야 하며, 표현의 자유의 가치와 중요성을 염두에 두어야 한다.</p> <ol style="list-style-type: none"> 1. 코로나19의 존재, 치료, 예방 및 진단, 감염, 사회적 거리두기 및 자가격리와 관련한 내용의 게시물인 경우 2. 세계보건기구(WHO) 또는 질병관리청에 의해 허위조작정보임이 명백히 확인된 경우 <p>②회원사는 제1항의 요건의 적용 여부에 대해 필요하다고 인정되는 경우 KISO에 심의를 요청할 수 있다.</p> <p>③회원사는 제1항의 규정에도 불구하고, 정책규정 제2장 및 제4장에 해당하는 게시물의 경우 해당 규정에 따라 처리한다.⁴⁵⁾</p>

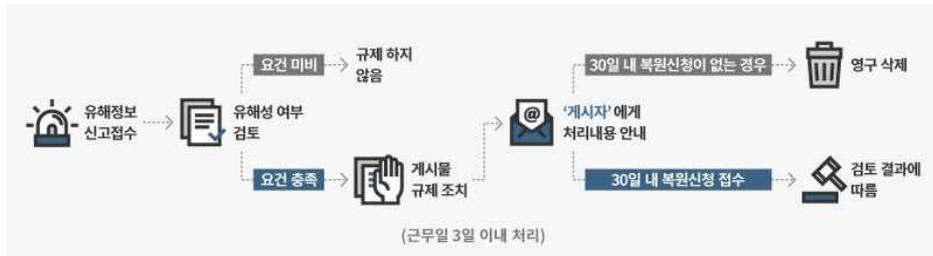
* 자료: 다음 고객센터

(2) 이용자의 허위조작정보 신고 방안

카카오의 운영정책을 위반하는 게시물의 경우 카카오는 다음 그림과 같은 절차에 따라 게시물을 처리한다. 유해정보의 신고 시 고객센터 홈페이지를 통해 게시물 신고를 접수할 수 있으며, 가짜뉴스의 경우에는 KISO의 가짜뉴스 신고 홈페이지를 활용할 수 있도록 다음과 같이 안내하고 있다.

45) KISO의 정책규정 제2장은 게시물에 관한 정책을 규정, 제4장은 선거 관련 인터넷 정보 서비스 기준에 관한 정책에 대해 규정하고 있다.

[그림 5-7 카카오 운영정책 내 유행 정보 신고 및 접수 처리



유해정보 신고

KISO 가짜뉴스 신고센터
게시물 규제 정책

도움말 공지사항

유해정보 신고

- 허위 신고자는 서비스 이용에 제한을 받습니다. 신중하게 신고해주세요.
- 명예훼손이나 저작권 침해 신고는 [연기](#)를 클릭하세요.
- 그 외, 유행정보신고 처리 및 절차 관련 궁금한 사항은 [콜리센터 도움말](#)을 확인해주세요.
- 다른 회사의 URL 및 올바른지 않은 URL은 신고 접수처리가 되지 않습니다.
- (필수)로 표시된 부분은 반드시 입력해주세요야 합니다.

이메일 (필수) @ 메일계정선택

게시물 주소 및 신고사유 (필수) 신고사유 선택

[+ 신고게시물 추가](#)

신고 URL은 1회 20번까지 등록할 수 있습니다.
(단종 URL로 신고된 경우, 처리가 지연 또는 누락될 수 있습니다.)

첨부 [?](#) + 첨부파일 추가

DMB

일부한 파일의 전체 크기는 10Mbyte 미만이어야 합니다.
용량이 초과될 경우, 큰의 경우가 임의하게 인형되지 않을 수 있습니다.
파일첨부는 JPG, GIF, PSD, TIF, MS Office 파일, 아예한글, PDF만 가능합니다.

* 자료: 다음 고객센터(<https://cs.daum.net/redbell/harm/harmReportForm.html>)

이용자가 위와 같이 안내된 KISO 가짜뉴스 신고센터를 통해 가짜뉴스 게시물 및 코로나 19 관련 허위조작정보 등을 신고하는 경우에는 아래 그림과 같은 절차에 따라서 게시물을 처리한다. 현재 카카오는 신고센터를 운영하고 있으나 네이버와 마찬가지로 이용자의 게시물 신고 시 가짜뉴스에 대해 따로 선택할 수 있는 신고기준을 규정하지 않는 것으로 과

악된다.

[그림 5-8 카카오 운영정책 내 KISO에 접수된 이용자 신고 처리 절차



* 자료: 다음 고객센터(<https://cs.daum.net/redbell/harm/harmReportForm.html>)

제 3 절 허위조작정보 자율규제 시범 가이드라인(안) 제언

1. 필요성

지금까지 살펴본 것처럼 현재 국내 규제체계는 인터넷 매체의 속성을 고려해 불법 및 유해정보에 대한 신속한 판단과 처리를 위해 법률에 근거한 규제와 사업자 단체 혹은 개별 사업자의 자율적인 규제로 이루어져 있다. 법률에 근거한 규제의 경우 주로 불법정보, 유해정보, 권리침해정보와 관련되어 있어 허위조작정보의 확산 억제에는 일정한 한계를 지닌다. 때문에 국회에서는 허위조작정보를 적절히 규율하기 위한 다양한 일부법률개정안이 발의되어 있는 상태이다.

민간 자율규제 차원에서는 플랫폼 사업자 단체인 KISO가 가짜뉴스에 대한 자율규제를 실시하고 있다. 회원사인 네이버와 카카오의 경우 KISO의 정책규정을 준수해 가짜뉴스를 규정하는 한편 신고센터 혹은 고객센터를 통해 이용자들에게 신고방법을 안내하고 이의제기절차를 마련하고 있다. 다만 이러한 절차는 이용자에게 권한을 부여하는 것을 넘어 플

랫폼이 보다 적극적인 안전조치를 취하도록 나아가는 데에는 한계를 갖고 있다.

최근에 출범한 플랫폼 민간 자율기구의 경우 ‘검색·추천 서비스의 투명성 제고를 위한 자율규제 원칙’ 등을 발표했다. 이는 플랫폼의 투명성을 강화하고 이용자의 권익을 보호하기 위한 원칙을 제시하는 등 보다 적극적으로 플랫폼의 책임을 요구한다. 또 공정위, 과기부, 방통위 등 정부 부처가 기구를 지원하는 등 규제기관과의 연계성을 확보하고 있다는 점에서 공동규제 거버넌스를 형성해 규제의 실효성을 높일 수 있기도 하다. 다만 플랫폼에 대한 책임 요구가 검색·추천 서비스의 투명성 확보에 머물러 있고 이에 기반한 활동과 평가가 본격화되지 않아 평가하기에는 이르다고 할 수 있다.

이를 해외와 비교해보자. 해외와의 비교는 (i) 정보통신망법을 중심으로 해외 법제와의 비교, 그리고 (ii) KISO와 사업자, 민간 자율기구를 중심으로 해외 실천 강령과의 비교가 가능할 것이다. 먼저 첫 번째 정보통신망법을 EU의 디지털서비스법 및 영국의 온라인 안전법과 비교해보도록 하자.

<표 5-14> 플랫폼의 책임에 관한 주요 법제 비교

구분	EU 「디지털서비스법」	영국 「온라인 안전법」	한국 「정보통신망법」
위험 평가/완화	·위험평가 및 위험완화 의무(대규모 플랫폼)	·위험평가 및 위험완화 의무	-
불법 콘텐츠 신고/조치	·불법콘텐츠 신고 메커니즘 구축 ·조치이유 설명, 불만처리 및 이의제기 시스템 구축 ·신뢰할 수 있는 신고자 우선 처리	·불법콘텐츠의 용이한 신고 메커니즘 구축 ·조치이유 설명, 불만처리 및 이의제기 시스템 구축	·표시방법 미준수 청소년 유해매체물 및 청소년유해광고 삭제 ·조치이유 설명
이용자 역량 강화	·약관에 콘텐츠 정책 및 관련 절차, 조치 등 저보 명시 ·다크패턴 금지 ·추천시스템의 투명성 확보	·약관에 콘텐츠 정책 및 관련 절차, 조치 등 정보 명시 ·성인 사용자에게 선택권 부여	·약관에 타인의 권리침해 정보에 대한 절차조치 등 정보 명시

구분	EU 「디지털서비스법」	영국 「온라인 안전법」	한국 「정보통신망법」
	보 및 선택권 부여 ·불법콘텐츠를 빈번히 제공하는 이용자 서비스 제공 중단		
광고	·광고 투명성	·사기광고 방지	-
행동강령	·행동강령 제정 장려	·Ofcom이 실천강령 준비	·행동강령 제정 장려
투명성 보고서	·투명성 보고서 발간	·투명성 보고서를 Ofcom에 제출	·불법촬영물 등 처리에 관한 투명성 보고서를 방통위에 제출

한편 KISO와 플랫폼 사업자, 플랫폼 민간 자율기구를 통해 이루어지고 있는 자율규제 정책과 방향성 등을 EU와 호주의 실천강령과 비교해보면 다음과 같이 정리할 수 있다.

<표 5-15> 허위조작정보 관련 민간 자율규제 비교

EU 강령 2018	EU 강령 2022	호주 강령 2022	KISO 정책	네이버, 다음 정책	플랫폼 민간 자율기구
허위조작정보 정의 규정	허위조작정보 정의 규정	허위조작정보 /오정보 정의 규정	가짜뉴스 정의 규정	가짜뉴스 정의 규정 (KISO)	-
-	-	안전장치 제공	가짜뉴스 심의	-	-
광고계재 검토	광고계재 검토	광고 또는 수익창출 인센티브 차단	-	-	-
정치광고	정치광고	정치광고	-	-	-
-	서비스 완전성	서비스 완전성	-	-	-
이용자 역량강화	이용자 역량강화	이용자 역량강화	-	신고방법/이의제기 절차	검색·추천 투명성/이용자

EU 강령 2018	EU 강령 2022	호주 강령 2022	KISO 정책	네이버,다음 정책	플랫폼 민간 자율기구 권익증진
연구공동체 역량강화	연구공동체 역량강화	전략적 연구지원	-	-	-
-	팩트체크 공동체 역량강화	-	-	팩트체크 기사 게재	-
-	투명성 센터	-	-	-	-
-	상설 태스크포스	-	-	-	-
모니터링	모니터링	모니터링	-	-	이행점검 등 실효성 확보

위의 표의 구분은 기본적으로 서로 다른 법령과 자율규제 체계를 한데 정리한 것이기에 각 항목이 서로 완벽히 배제적이지는 않다. 예를 들어 호주 실천강령에서 ‘안전장치 제공’은 허위조작정보로 인해 발생할 수 있는 피해에 대한 안전장치를 제공하기 위해 허위조작정보에 대한 라벨링, 부정행위로 확산되는 콘텐츠 삭제, 부정행위를 사용하는 사용자의 계정을 정지하거나 비활성화, 팩트체커와 협력하거나 재원 지원 등의 여러 방법을 담고 있는데 이는 호주 실천강령만의 고유의 것이 아니라 다른 규제체계의 다양한 항목 안에 이미 반영되어 있다. 때문에 해석에 주의를 기울일 필요가 있다. 그럼에도 한국의 자율규제 체계가 허위조작정보에 대응하는 해외의 실천과 비교할 때 상대적으로 취약한 부분이 무엇인지에 대해서는 어느 정도 시사점을 찾을 수 있다.

물론 타국의 법령과 실천강령은 각국의 정치적, 사회적, 경제적 맥락을 반영한 것이기에 이를 따르는 것이 반드시 좋은 해결방안이라고는 할 수 없다. 그러나 온라인 허위조작정보에 대한 대응 필요성이 증가하고 있으며, 이에 대응하기 위해 인터넷 플랫폼 사업자들이 책임 있게 나서야 할 필요성은 계속해서 커지고 있다. 때문에 해외의 사례를 참고하여, 한국에서도 보다 포괄적이면서도 실효성을 확보하기 위한 일종의 자율규제 시범 가이드라인을 모색해보는 것은 의미 있는 일이 될 수 있을 것이다.

2. 허위조작정보의 정의

허위조작정보에 관한 자율규제 시범 가이드라인을 마련하는데 있어 가장 먼저 선행되어야 할 것은 허위조작정보에 대한 정의이다. 현재 한국의 법령은 이를 규정하고 있지 않으며, KISO에서도 가짜뉴스를 규정할 뿐 허위조작정보는 정의하지 않고 있다. 코로나19 시기에 한시적으로 코로나19 관련 허위조작정보를 규정했을 뿐이다.

때문에 여기에서는 해외의 법령과 강령이 허위조작정보를 어떻게 정의하고 있는지 먼저 참고하도록 한다.

<표 5-16> 허위조작정보 정의 비교

구분	주요 내용	
EU 실천강령	정의	고의적으로 공공에 해를 끼치거나 이익을 얻을 목적으로 설계, 제작, 유포되는 허위이거나, 부정확하거나, 또는 사람들을 오도하는 정보
	제외	오해의 소지가 있는 광고, 보도 오류, 풍자 및 패러디, 명확하게 식별되는 편파적인 뉴스 및 논평을 포함하지 않음
호주 실천강령	정의	A. 실증적으로 잘못됐거나 오도하거나 기만적인 디지털 콘텐츠 B. 이는 부정행위(스팸, 조작 등)를 통해 디지털 플랫폼 이용자들 사이에서 확산 C. 해당 내용의 유포가 피해(선거 등 민주적 과정과 공중보건, 취약계층의 보호, 치안과 보안, 환경 등에 확실하고 심각한 위협이 되는 피해)를 끼치리라고 합리적으로 예상되는 경우
	제외	A. (풍자 및 패러디를 포함한) 오락 또는 교육의 목적으로, 선의의 신뢰 하에 제작된 콘텐츠 B. 호주 주정부 또는 연방정부에서 승인한 콘텐츠 C. 정치광고 또는 (호주 법률에 따라 등록된) 정당이 인가한 콘텐츠 D. 전문 뉴스 콘텐츠
프랑스 「정보조작법」	정의	유권자 투표의 진정성을 해칠 수 있는 사실에 대한 부정확하거나 오해를 불러올 수 있는 주장 또는 비난 등이 고의적이거나 인공적인 또는 자동화된 방식으로, 온라인 정보통신서비스를 통해 대규모로 유포되는 정보

허위조작정보 개념을 정의하는 기본적인 요소는 정보의 허위성과 의도성이라고 볼 수 있다. EU의 실천강령은 이를 바탕으로 하여 다소 폭넓게 허위조작정보를 정의하되 거기에 정치적, 경제적 목적성을 고려한다. 이에 비해 호주의 실천강령은 허위성과 의도성을 가진 정보가 스팸, 조작 등 부정행위를 통해 확산되고 해당 내용의 유포가 확실하고 심각한 위협이 되는 피해를 끼치리라고 합리적으로 예상되는 경우로 보다 범위를 세부적으로 한정한다. 정의에 특정한 목적성을 고려하지 않으나 예상되는 피해에 선거 등 민주적 절차와 과정, 보건, 안보 등을 반영함으로써 그 범위를 규정한다. 한편 프랑스의 정보조작법은 선거기간 동안의 허위조작정보를 방지하는 것이 목적이기에 투표와 관련된 목적성을 갖는다. 또 인공적인 또는 자동화된 방식, 즉 봇(bot) 등을 이용해 대규모로 유포되는 경우에 한정하여 허위조작정보를 정의하는 것이 특징이다.

이상에서 보듯 허위조작정보의 정의는 일반적으로 허위성, 의도성(혹은 기만), 목적성 등을 고려하고 있다. 하지만 유포되는 방식을 정의에 반영하는가 여부는 각 강령 혹은 법령마다 견해를 달리한다. 유포방식과 규모 등을 한정할 경우 보다 개념이 한정돼 명확해지고 이로 인한 표현의 자유 침해를 억제하는 데에도 효과적일 수 있다. 하지만 이는 다른 방식으로 사회에 해를 미치는 허위조작정보의 유포에 대응하는 데에는 한계를 남겨놓을 수 있다. 이 글에서는 폭넓은 대응방안을 모색하기 위한 시론적인 가이드라인을 만드는 것이 목적임을 고려, EU의 실천강령을 따라 포괄적인 정의를 채택하고자 한다.

<표 5-17> 허위조작정보의 정의(안)

구분		주요 내용
허위조작 정보	정의	정치적 또는 경제적 이익 등을 얻으려는 목적에서 제작·유포되며 공공에 해를 끼칠 수 있는 허위이거나 오도하는 정보
	제외	풍자 및 패러디, 다소의 오류가 있거나 일부 사실에 부합하지 않는 과장된 표현이 있는 정보

3. 이행 원칙

시범 가이드라인에서 제안하는 구체적인 원칙들을 이행함에 있어서는 공통적으로 다음과 같은 원칙을 따를 필요가 있다.

첫 번째는 공동대응의 원칙이다. 가이드라인에 참여하는 사업자와 기관은 허위조작정보에 대해 공동의 이해에 도달하고 일관된 접근방식을 채택하는 것의 중요성을 인지하는 것이다. 두 번째는 비례의 원칙이다. 각 사업자가 제공하는 서비스의 유형 및 범위 등에 따라 허위조작정보의 유포 및 그에 따른 피해의 정도는 다를 수밖에 없다. 때문에 가이드라인의 준수는 서비스의 특성에 따라 비례적으로 적용 및 이행될 수 있다. 세 번째는 표현의 자유에 대한 보호 원칙이다. 허위조작정보에 대한 대응은 헌법이 시민에게 보장하는 표현의 자유를 침해하거나 위축시킬 수 없다. 네 번째는 개인정보의 보호 원칙이다. 허위조작정보에 대응하기 위한 본 가이드라인의 이행은 개인정보 보호에 대한 법령과 원칙을 위반해서는 안 된다. 이는 자율규제 시범 가이드라인에 다음과 같이 반영할 수 있다.

<표 5-18> 가이드라인의 이행원칙(안)

주요 내용
<ul style="list-style-type: none">○ (공동 대응) 허위조작정보의 대응원칙, 기준 및 처리절차 등에 대해 공동의 이해에 도달하고 일관된 접근방식을 채택하는 것의 중요성을 인지한다.○ (비례의 원칙) 본 가이드라인이 해결하고자 하는 정보와 행위, 그리고 피해의 유형은 각 플랫폼에서 제공하는 서비스에 따라 그 발생정도와 영향이 다를 수 있다. 이에 본 가이드라인의 준수는 서비스의 특성에 따라 비례적으로 적용 및 이행될 수 있다.○ (표현의 자유에 대한 보호) 허위조작정보에 대한 대응은 헌법이 시민에게 보장하는 표현의 자유를 침해하거나 위축시킬 수 없다. 거짓으로 생각된다는 이유만으로 적법한 정보에 대한 접근을 막거나 이를 삭제해서는 안 된다.○ (개인정보의 보호) 본 가이드라인의 이행과 관련된 각 사업자의 구체적인 정책 및 조치는 개인정보 보호에 대한 법령과 원칙을 위반해서는 안 된다. 또 연구 목적으로 데이터에 접근하는 경우에도 이용자의 개인정보를 보호해야 한다.

4. 주요 원칙

현재 주요 자율규제 내용을 비교해보면 EU와 호주의 실천강령, 그리고 한국의 자율규제 체계는 공통적으로 ▲플랫폼의 투명성 확보, ▲이용자 역량강화, ▲허위조작정보의 신고 및 조치 등, ▲투명성 보고서 발간 등 모니터링 등에 대한 원칙을 포함하고 있는 것으로 보인다. 때문에 이 원칙들은 가이드라인 마련 시 기본적으로 고려하되, 글로벌 동향 등을 고려하여 세부적인 기준과 내용을 마련할 필요가 있다.

이에 비해 ▲허위조작정보를 활용한 광고 및 수익창출의 차단, ▲서비스의 완전성 확보, ▲연구 및 팩트체크 지원 등은 EU와 호주의 실천강령에서는 발견되지만 한국의 자율규제 체계 내에서는 발견되지 않아 가이드라인 반영을 위해서는 검토가 필요하다.

먼저 허위조작정보를 활용하는 광고 및 수익창출의 차단과 같은 경우 EU의 실천강령에서는 첫 번째로, 그리고 호주의 강령에서는 두 번째 목표로 다룰 정도로 중요하게 취급한다. 허위조작정보의 생성과 유포가 그 정의상 정치적 또는 경제적 이익을 얻고자 하는 목적성을 갖고 있다면 그로 인해 얻을 수 있는 정치적 또는 경제적 이익을 차단하는 것은 무엇보다도 효과적인 방법이 될 수 있기 때문이다. 때문에 자율규제 시범 가이드라인에 반영할 필요가 있다고 판단된다.

다음으로 서비스의 완전성 확보는 허위계정 또는 매크로 등 자동화된 봇의 사용 등 서비스의 완전성과 보안을 훼손하는 이용자 행위 유형을 규제하거나 관리하기 위한 목적을 담고 있다. 호주의 실천강령과 프랑스의 정보조작법이 이와 같은 방식을 통한 유포를 허위조작정보의 정의 안에 반영하고 있을 만큼 이는 오늘날 온라인 환경에서 심각한 문제가 되고 있다. 따라서 이에 대한 적절한 조치 의무를 자율규제 시범 가이드라인에 반영할 필요가 있다고 판단된다.

허위조작정보와 관련된 연구 및 팩트체크 지원은 중장기적으로 반드시 필요하다. 다양한 이해관계자와의 협력을 강화하고 다각적인 대응을 통해 허위조작정보에 맞설 수 있는 환경을 구축하는 일은 허위조작정보에 대응하기 위한 기본적인 방법이라고 할 수 있다.

(1) 허위조작정보의 신고 및 조치

한국의 주요 사업자들은 KISO와 협력하는 한편 고객센터 혹은 신고센터를 통해 유해정보에 대한 이용자 신고를 지원하고 있으며 이와 관련한 이의제기 절차를 마련해 두고 있다. 해외의 실천강령도 대동소이한 규정을 마련해 두고 있다. 다만 EU와 호주의 실천강령은 허위조작정보 신고 절차와 기능 등을 보다 이용자 친화적인 방식으로 구축할 것을 명시하는 것이 차이점이다. 또 EU의 실천강령은 신고 등에 의해 취해진 집행 조치의 대상이 된 콘텐츠 또는 계정 이용자에게 해당 사실을 알리고, 이의제기 기회를 제공하며, 불만사항에 대해 적시에 성실하고 투명하며 객관적인 방식으로 처리하고, 불만사항이 타당하다고 판단될 경우 지체 없이 조치를 취소하도록 명시한다.

한편 EU와 호주의 실천강령은 이용자에게 서비스의 정책 또는 약관을 위반한 허위조작정보를 표시할 수 있는 라벨링 등의 기능을 제공하도록 하고 있다. 허위조작정보에 대한 대응은 헌법이 시민에게 보장하는 표현의 자유를 침해할 수 없으며, 거짓으로 생각된다는 이유만으로 합법적인 콘텐츠에 대한 접근을 막거나 함부로 삭제할 수는 없기 때문이다. 따라서 한국의 허위조작정보 관련 신고 및 조치 등에 있어서도 이를 고려할 필요가 있다.

또 EU의 디지털 서비스 법은 각 국가에서 전문성과 기술을 갖춘 법인 또는 단체에 부여되는 ‘신뢰할 수 있는 신고자’ 제도를 운영하도록 규정하고 있다. 향후 국내에서 관련 법 제개정을 통한 근거 마련을 통해 플랫폼 사업자의 실천을 제도적으로 뒷받침하는 방안이 마련된다면 이를 강령에 포함시키는 것도 고려할 수 있다.

<표 5-19> 허위조작정보의 신고 및 조치 주요 내용(안)

주요 내용
<ul style="list-style-type: none">○ (허위조작정보의 신고) 허위조작정보에 대한 이용자 신고절차의 가시성과 편의성을 높이기 위한 방안을 마련한다.○ (신고된 정보의 조치 및 표시) 적절한 절차를 거쳐 신고된 정보를 조치한 후, 그 조치 사실을 지체 없이 문제가 된 콘텐츠 또는 계정을 보유한 이용자에게 알린다. 관련 조치에는 팩트체크 기관, 민간 자율규제 기구 및 기타 공신력 있는 기관과의 협력을 통해 허위조작정보와 관련된 주의를 표시하는 것 등이 포함될 수 있다.

- (이의제기) 문제가 된 집행조치와 관련된 이의제기 절차 및 분쟁조정 절차를 마련하여 투명하고 객관적인 방식으로 처리하며, 불만사항이 타당하다고 판단되는 경우 지체 없이 조치를 해제할 수 있도록 한다.

(2) 허위조작정보의 수익 차단

한국의 자율규제 체계에서는 허위조작정보로 인한 수익을 차단하는 내용이 들어있지 않다. 이에 비해 EU와 호주의 실천강령에서는 광고 게재 등 수익차단을 방지하기 위한 내용과 절차를 상세하게 규정한다. 예를 들어 호주의 실천강령은, 플랫폼으로 하여금 ▲광고주가 허위조작정보를 확산하는 디지털 광고를 반복적으로 게재하는 것을 막기 위해 상업적으로 타당한 노력을 해야 하며 ▲허위조작정보에 대한 광고 또는 수익창출 인센티브 차단을 목표로 하는 정책과 프로세스를 이행하고 ▲온라인 광고의 매매 및 광고 서비스와 관련된 모든 이해당사자들이 협력해 온라인 광고 생태계 전반의 투명성을 개선하고, 웹사이트의 광고 배치를 효과적으로 검토하며 규제하도록 명시한다. EU의 실천강령 역시 이와 유사한 내용을 조금 더 상세하게 규정하고 있는 것이 특징이다. 이상의 내용을 자율규제 시범 가이드라인에 다음과 같이 반영할 수 있다.

<표 5-20> 허위조작정보의 수익 차단 주요 내용(안)

주요 내용
○ (허위조작정보의 수익화 방지) 허위조작정보를 유포하는 콘텐츠에 대해 광고수익 배분 및 재정적인 인센티브를 제지하는 등, 허위조작정보를 이용한 수익창출의 연결고리를 차단하기 위한 정책을 시행한다.
○ (허위조작정보를 포함한 광고의 유포 방지) 광고 시스템의 오용을 통해 허위조작 정보가 광고 메시지의 형태로 유포되는 것을 방지하기 위한 정책을 시행하고 상업적으로 실현 가능한 노력을 기울인다.
○ (대가를 받은 콘텐츠의 구분) 후원을 받거나 지불된 콘텐츠와 그렇지 않은 콘텐츠를 이용자가 식별할 수 있도록 관련 정책을 시행한다.

- (관련당사자와의 협력) 온라인 광고 생태계 전반의 투명성을 개선하기 위해 관련 당사자들과 협력을 강화한다.

(3) 서비스의 완전성

서비스의 완전성(integrity)은 가짜계정, 자동화된 봇, 해킹, 사칭, 악성 딥페이크 등 서비스의 완전성과 보안을 훼손하고 허위조작정보를 유포할 수 있는 행위들을 규제하거나 관리함으로써 확보할 수 있다. 또 AI 생성 콘텐츠에 대한 식별 표시를 강화하는 것도 중요하다. 이상의 내용을 자율규제 시범 가이드라인에 다음과 같이 반영할 수 있다.

<표 5-21> 서비스의 완전성 주요 내용(안)

주요 내용
<ul style="list-style-type: none"> ○ (허용되지 않는 행위 규제) 서비스의 완전성과 보안을 훼손하는, 서비스에서 허용되지 않는 행위 및 관행의 범위를 명확하게 표시하고, 이를 규제하거나 관리하는 대책을 세움으로써 관련된 허위조작정보 문제를 해결하기 위해 노력한다. 이러한 행위 및 관행에는 다음이 포함될 수 있다. <ul style="list-style-type: none"> - 가짜계정 생성 및 사용, 계정 탈취, 팔로워 증폭 봇, 매크로 사용 - 해킹 및 유출 - 사칭 - 악성 딥페이크 - 조회 수, 팔로워 수 등 계정(채널) 통계를 인위적으로 부풀리는 도구 - 조직적인 조작행위에 가담하는 계정의 생성 및 사용 - 허위조작정보의 도달범위 또는 이에 대한 공중의 지지를 인위적으로 확대하려는 행위 <p>단 이러한 대책은 이용자의 익명의 계정 생성을 배제해서는 안 된다.</p> ○ (AI 생성 콘텐츠 표시) AI에 의해 생성되거나 조작된 콘텐츠가 인간에 의해 작성된 것으로 오해되지 않도록 이를 표시하는 등 관련 정책을 시행한다.

(4) 검색·추천 서비스의 투명성

국내 플랫폼 민간 자율기구와 EU, 호주의 실천강령은 모두 검색·추천 서비스의 투명성과 관련한 내용을 자율규제 체계 안에 반영하고 있다. 플랫폼 민간 자율기구는 검색·추천 서비스의 투명성을 확보하기 위해 이를 구성하는 주요 변수 및 그에 관한 설명을 이용자가 쉽게 이해할 수 있도록 공개해야 하며 이에 대한 설명책임을 진다고 규정한다.

EU와 호주의 강령도 추천시스템 등 정보의 우선순위를 정하는 방식과 관련된 정보를 이용자에게 투명하게 공개하고, 이러한 시스템의 옵션을 제공하는 한편 옵션을 이용할 수 있는 정보를 이용자가 알기 쉽도록 제공해야 한다고 규정한다. 이상의 내용을 자율규제 시범 가이드라인에 다음과 같이 반영할 수 있다.

<표 5-22> 검색·추천 서비스의 투명성

주요 내용
○ (검색·추천기준 공개) 검색·추천 시스템의 정보처리 방식(알고리즘) 혹은 이와 관련된 주요 기준 및 변수를 공개하고 이용자에게 이의 적용과 관련된 선택권을 부여한다. 주요 기준 및 변수 공개 혹은 선택권에 대해서는 이용자가 접근하기 쉬운 방식으로 공개해야 하며 이해하기 쉬운 설명과 구체적인 예시를 제공한다. 또 이용자가 설명을 요구하는 경우 충실히 답변하도록 한다.

(5) 이용자 역량 강화

허위조작정보의 신고 및 이의제기, 검색·추천 서비스의 투명성 등은 부분적으로 이용자의 역량을 강화하고 권한을 부여하는 내용을 포함하고 있다. 여기에 더해 EU와 호주의 강령이 규정하듯이 이용자가 다양한 정보에 접근할 수 있도록 하고 정보의 신뢰성 또는 정확성을 확인할 수 있도록 돕는 것은, 정보에 기반한 선택을 하도록 만듦으로써 허위조작 정보의 확산을 막는데 근본적으로 기여하는 방법이 될 수 있다.

이상의 내용을 자율규제 시범 가이드라인에 다음과 같이 반영할 수 있다.

〈표 5-23〉 이용자 역량 강화

주요 내용
<ul style="list-style-type: none"> ○ (정보 식별역량) 검색, 피드, 추천 서비스 등에서 출처가 확실하고 신뢰할 수 있는 서비스가 우선 노출되도록 하거나, 이용자가 정보의 출처를 식별하는데 도움이 되는 기술이나 도구를 제공함으로써, 허위조작정보의 가시성을 떨어뜨리기 위해 노력한다. ○ (정보 다양성) 이용자들이 공적인 주제에 관한 다양한 관점에 보다 손쉽게 접근할 수 있도록 하기 위한 노력과 지원을 지속하고 강화한다. ○ (정보 리터러시) 취약계층을 포함하여 시민의 비판적 사고와 정보 리터러시를 향상시키기 위해 노력하며, 이를 위해 시민사회·교육기관·정부 및 기타 이해당사자들과 협력을 강화한다.

(6) 팩트체크 및 연구 지원

허위조작정보와 관련해 신뢰할 수 있는 팩트체크 기관 및 단체와 협력을 강화하고, 허위조작정보에 관한 선의의 연구 활동을 지원하는 일도 중요하다. 이는 자율규제 시범 가이드라인에 다음과 같이 반영할 수 있다.

〈표 5-24〉 팩트체크 및 연구 지원

주요 내용
<ul style="list-style-type: none"> ○ (팩트체크 및 연구 지원) 팩트체크 및 허위조작정보와 관련된 연구 활동을 지지하고 장려하며, 개인정보를 침해하지 않는 선에서 이러한 활동을 위해 필요한 데이터에 접근할 수 있도록 한다. ○ (팩트체크 및 연구 공동체와의 협력) 팩트체크 및 허위조작정보에 관한 연구 지원을 통해 공중의 이해와 대응역량을 강화하기 위해서 팩트체크 공동체, 연구 공동체 및 기타 이해당사자들과 협력을 활성화한다.

(7) 본 가이드라인의 모니터링

이와 같은 자율규제 시범 가이드라인은 이에 참여하는 사업자의 자율적인 노력을 통해 효력이 발생하며, 평가를 통해 더욱 개선될 수 있다. 따라서 이와 같은 의무의 준수 여부를 평가하고 개선하기 위하여 투명성 보고서를 발간하는 일이 필요하다.

투명성 보고서는 일반에 공개하는 것이 원칙이지만, 본 가이드라인이 시범 가이드라인임을 고려할 때 일반에 바로 공개하는 것보다 규제기관인 방통위에 제출하도록 함으로써 이를 평가하고 정식 가이드라인을 마련하는 일이 필요할 수도 있다. 이는 자율규제 시범 가이드라인에 다음과 같이 반영할 수 있다.

<표 5-25> 투명성 보고서 발간 및 모니터링

주요 내용
<ul style="list-style-type: none">○ 본 가이드라인은 시범 가이드라인으로서, 가이드라인에 참여하는 사업자의 자율적인 이행을 통해 효력이 발생하며 평가를 통해 개선될 수 있다. 참여 사업자는 시범 가이드라인을 이행하고 평가하며 개선하는 과정에서 방송통신위원회와 협력한다.○ (준수 의무) 참여 사업자는 본 가이드라인의 조항이 자사의 서비스와 관련이 있는지 평가하고, 서명할 조항을 지명한다. 지명하지 않은 약속은 준수할 의무가 없다.○ 참여 사업자는 가이드라인의 이행을 보장하기 위해 적절한 재정적인적 자원을 투입하고 내부 프로세스를 마련한다.○ (회의체 운영) 본 가이드라인 시행 이후 1년 간 시범 가이드라인의 효과 및 참여 사업자의 이행사항에 대한 평가를 진행한다. 이 기간 동안 참여사업자와 방송통신위원회는 정기적인 회의를 통해 진척사항에 대해 분석한다. 또한 회의에서는 다음 내용을 다룰 수 있다.<ul style="list-style-type: none">- 허위조작정보와 관련된 동향 및 지식의 공유- 허위조작정보 관련 정의의 개선- 다른 자율규제 원칙 및 자율규제 기구와의 협력- 투명성 보고서를 위한 구조적인 지표 및 양식 개발○ (보고서 작성) 참여사업자는 가이드라인 시행 후 1년 이내에 허위조작정보에 대

응하여 수행해온 활동에 대한 투명성보고서를 작성하여 방송통신위원회에 제출한다. 보고서는 참여 사업자가 시행한 모든 조치의 세부사항을 포함한다.

- 서비스의 허위조작정보 관련 정책
 - 서비스의 허위조작정보 대응 관련 데이터(분류방식, 유통 및 신고 현황, 처리 현황 등), 구체적인 사례 및 피드백 등
 - 허위조작정보 유포자의 수익창출 기회를 축소시키는 정책 및 조치
 - 서비스의 완전성을 확보하기 위해 취한 정책 및 조치, 허용되지 않는 행위 규제와 관련한 데이터(폐쇄한 가짜계정의 수, 해킹 및 유출 현황 등), 구체적인 사례 및 피드백 등
 - 검색추천 서비스의 투명성을 강화하기 위해 취한 정책 및 조치
 - 이용자의 역량을 강화기 위해 취한 조치
 - 팩트체크 및 허위조작정보 관련 연구 활동에 대한 지원 노력
- (평가 및 개선) 참여사업자와 방송통신위원회는 보고서 제출 후 3개월 이내에 보고서 및 가이드라인에 대한 공동의 평가를 완료하며, 이를 바탕으로 정식 가이드라인을 제정을 포함한 후속조치를 논의하고 제안한다.
-

참고문헌

논문 및 보고서

- 강주현 (2019). 언론 보도에 나타난 가짜뉴스 담론의 속성과 사회적 실천방향. <한국언론학보> 63권 6호, 7~64쪽.
- 고선규·이소영·김형호·송재민 (2021). 온라인 가짜뉴스 관련 선거관리제도 개선방안에 대한 연구, 한국선거협회.
- 김경숙·홍건식 (2023.11.10.) 바이든 행정부의 첫 인공지능(AI) 행정명령과 시사점, <이슈브리프> 480호, 국가안보전략연구원.
- 김민정 (2019). 가짜뉴스(fake news)에서 허위조작정보(disinformation)로 - 가짜뉴스 관련 국내 법안과 해외 대응책에 나타난 용어 및 개념정의 비교. <미디어와 인격권> 5권 2호, 43~81쪽.
- _____ (2020). 유럽연합의 온라인 허위조작정보(disinformation) 자율규제 사례 고찰. <언론과 법>, 19권 1호, 193~230쪽.
- 김선호 (2022.2.9.). 유권자가 바라보는 선거보도 <Media Issue>, 8권 1호. 한국언론진흥재단
- 김세환·황용석 (2019). 가짜뉴스에 대한 한국과 EU의 비교정책 연구: 정책형성 담론을 중심으로. <언론정보연구>, 56권 2호, 156~187쪽.
- 김연진 (2021). 허위조작정보 규제에 관한 헌법적 연구. 성균관대 박사학위 논문.
- 김재선 (2020). 허위정보 규제를 위한 행정법적 대응방안 - 영국의 규제 논의를 중심으로, <법학논집>, 24권 3호, 223-248쪽.
- 김주용 (2022). 허위조작정보 규제방안에 관한 이해관계자들의 인식 연구: 새로운 규제의 필요성 및 법적규제 방안에 대한 심층인터뷰 분석을 중심으로. <한국언론정보학보>, 통권 114호, 68~109쪽.
- 김준일 (2021). 허위조작정보(가짜뉴스)와 언론의 역할, <언론중재>, 2021년 여름호. 언론중재위원회, 14~27쪽.
- 김지현·안성경·조경화·최서지 (2023.12.) 주요국의 온라인 허위조작정보 대응 입법례, <최신외국입법정보>, 238호, 국회도서관.

김태오 (2020). 독일의 미디어협약 제정안상 주요 내용 - 구 방송협약을 대체·보완하는 새로운 규제를 중심으로, <최신외국법제정보> 2020년 2호, 75-87쪽.

김희경·경노상·남승용·이찬구·이수엽·박소은 (2022) 인터넷 공동규제 활성화 방안 연구, 방송통신심의위원회.

박아란 (2019). ‘가짜뉴스’와 온라인 허위정보(disinformation) 규제에 대한 비판적 검토. <언론정보연구> 56권 2호, 113-155쪽.

박아란·이소은 (2020). 디지털 뉴스 리포트 2020 한국. 한국언론진흥재단.

박영흠 (2023.5.) EU의 이스라엘-하마스 전쟁 관련 허위정보 대응과 시사점, <Media 정책 리포트>, 2023년 5호, 한국언론진흥재단.

박용숙·강명원·고수윤 (2022). 2022년도 해외언론법제연구보고서 : 유튜브 등 미디어 플랫폼 상 인격권 보호를 위한 해외법제연구, 언론중재위원회.

박찬경 (2022). 유럽연합 디지털 서비스법(Digital Services Act) : 규제나, 자율이냐를 넘어서, <언론중재>, 2022년 가을호, 언론중재위원회.

성육제·정은진 (2020.5.7.) 코로나19와 허위정보 - 유형분석과 대응방안. <KISDI>

송예일 (2022.11.16.) EU 디지털 관련 정책 및 전략 현황, <KERC Issue Report> 2022-2. 한-EU 연구협력센터.

송혜진·송민수 (2021). 동의의결 해외사례 연구: 소비자피해구제를 위한 시정방안을 중심으로, 한국소비자원.

시청자미디어재단 (2022.7.3.). UN 인권이사회 결의안 ‘인권의 향유와 실현에 있어 허위 조작정보의 부정적 영향에 맞서는 국가의 역할’, <팩트체크와 동향리포트 6호>, 26-31쪽.

심나리 (2019). 유럽연합의 허위정보 대응 실행지침-유럽의회 선거를 위한 페이스북, 구글, 트위터의 행동강령 및 보고서를 중심으로., <언론중재> 2019년 여름호, 언론중재위원회.

심영섭 (2018). ‘가짜뉴스’를 어떻게 걸러낼 것인가?, <방송 트렌드 & 인사이트>, 17호, 한국콘텐츠진흥원.

오세욱·정세훈·박아란 (2017). 가짜뉴스 현황과 문제점. 한국언론진흥재단.

오세욱·박아란·최진호 (2021). 디지털 뉴스 리포트 2021 한국. 한국언론진흥재단.

- 윤명화·손수빈 (2015). 소셜네트워크와 정체성 놀이 : 트위터 봇 사례연구를 중심으로, <문화와 사회> 18권, 287~328쪽.
- 윤혜선 (2012). 방송통신규제 개선방향에 관한 소고: 호주의 사례에서 본 협동규제의 의의와 방식을 중심으로, <경제규제와 법>, 5권 1호, 202~219쪽.
- 이가연 (2018). 디지털 단일시장 속 유럽연합의 주요 언론법제 동향, <해외언론법제 연구보고서>, 2018년 1·2호, 언론중재위원회.
- 이민영 (2010). 인터넷 자율규제의 법적 의의, <저스티스> 115호, 138~139쪽.
- 이병준 (2021). 유럽연합 디지털 서비스법을 통한 플랫폼 규제 - 디지털 서비스법 초안의 주요 내용과 입법방향을 중심으로, <소비자법연구>, 7권 2호, 181~210쪽.
- 이성규 (2022). 해외 미디어는 플랫폼과 관계를 어떻게 재설정하고 있는가, <해외 미디어 동향> 2022년 2호, 한국언론진흥재단.
- 이지영 (2023). 허위 정보 확산 막기 위해 통신법 개정안 초안 발표한 호주 정부, <신문과 방송> 2023년 11월호, 91~94쪽.
- 전훈·최우정·김재선 (2019). 허위정보 유통과 민주주의, 한국법제연구원.
- 정다운 (2021). 독일의 선거 관련 허위조작정보 규제에 관한 논의, <해외언론법제연구보고서>, 2021년 5호, 언론중재위원회.
- 정영주·홍종윤·박유진 (2021). 사회적 법익을 침해하는 허위조작정보 대응을 위한 시론적 논의 - 코로나 19 관련 허위조작정보를 중심으로, <언론과 법>, 20권 1호, 73~114쪽.
- 차미영 (2022). 40개국 데이터로 살펴본 코로나19 인포데믹의 여파, <KOSTAT 통계플러스> 2022년 봄호, 통계청, 56~69쪽.
- 최민재·진민정·이현우 (2023.5.31.) 유럽의 가짜뉴스 대응정책, <Media 정책리포트>, 2023년 2호, 한국언론진흥재단.
- 최원석(2019.7.) 페이스북 'F8 2018' 을 통해 살펴본 가짜뉴스 대책과 미디어 교육, <미디어 리터러시>, 한국언론진흥재단.
- 최지선 (2022). 프랑스의 허위조작정보 대응에서의 국가안보 관점, <언론과 법>, 21권 2호, 85~128쪽.
- 최진웅 (2022.11.) 인공지능(AI) 기술 기반 딥페이크 선거운동의 해외 입법동향과 시사점, <외국 입법·정책 분석>, 26호, 국회입법조사처.

- 최진호·박영흠 (2022). 디지털 뉴스 리포트 2022 한국. 한국언론진흥재단.
- 페이스북(2019.7.24.) 이용자 개인정보 보호에 대한 새로운 기준을 제시한 연방거래위원회(FTC)와의 합의안. (Retrieved from) <https://about.fb.com/ko/news/2019/07/%EC%97%B0%EB%B0%A9%EA%B1%B0%EB%9E%98%EC%9C%84%EC%9B%90%ED%9A%8Cftc%EC%99%80%EC%9D%98-%ED%95%A9%EC%9D%98%EC%95%88%EC%9D%B4-%EC%9D%B4%EC%9A%A9%EC%9E%90-%EA%B0%9C%EC%9D%B8%EC%A0%95%EB%B3%B4%EB%B3%B4/>
- 한국언론진흥재단 (2021). 2021 소셜 미디어 이용자 조사
 _____ (2022). 2022 언론 수용자 조사.
 _____ (2023). <디지털 뉴스 리포트 2023 한국>. 서울: 한국언론진흥재단.
- 허순철 (2022). 유튜브 딥페이크(deepfake) 영상과 허위사실 공표, <미디어와 인격권>, 8권 1호, 1~47쪽.
- 황성기 (2022). ICT 분야에서의 자율규제 - 인터넷 포털과 게임 분야를 중심으로. <공법 연구> 50권 3호, 33-61쪽.
- 황용석·권오성 (2017). 가짜뉴스의 개념화와 규제수단에 관한 연구 - 인터넷서비스사업자의 자율규제를 중심으로. <언론과 법> 16권 1호, 53~101쪽.
- 황혜인 (2021). 코로나-19 확산에 따른 EU 회원국의 대응동향: 허위조작정보에 대한 대응 사례를 중심으로, <KISDI Perspectives>, 2021년 3호.
- InterAd (2023. 4. 12). 구글 스팸브레인 성과 리포트: 99%의 스팸 차단으로 사용자 보호. (Retrieved from) <https://www.interad.com/insights/google-spambrain-webspam-report>
- 4A's (2023. 9. 22). Summary: The GARM Aggregated Measurement Report. (Retrieved from) <https://www.aaaa.org/summary-the-garm-aggregated-measurement-report/>
- ACMA (2020). *Misinformation and news quality on digital platforms in Australia A position paper to guide code development*. (Retrieved from) <https://apo.org.au/sites/default/files/resource-files/2020-06/apo-nid311095.pdf>
- _____ (2021). *ACMA report to government on the adequacy of digital platforms' disinformation and news quality measures*. (Retrieved from) <https://www.acma.gov.au/sites/default/files/2021-11/Adequacy%20of%20digital%20platforms%20disinforma>

tion%20and%20news%20quality%20measures.pdf

- Australian Competition and Consumer Commission (2019). *Digital Platform Inquiry: Final Report*. (Retrieved from) <https://www.accc.gov.au/system/files/Digital%20platform%20inquiry%20-%20final%20report.pdf>
- Australian Government (2019). *Regulating in the Digital Age: Government Response and Implementation Roadmap*. (Retrieved from) <https://treasury.gov.au/sites/default/files/2019-12/Government-Response-p2019-41708.pdf>
- Brusselstimes (2023, 5, 27). Twitter leaves EU voluntary Code of Practice against disinformation. *Brusselstimes*. (Retrieved from) <https://www.brusselstimes.com/524844/twitter-leaves-eu-voluntary-code-of-practice-against-disinformation>
- Collins, (2022, 6, 16). EU Strengthens Disinformation Rules to Target Deepfakes, Bots, Fake Accounts. *CNET*. (Retrieved from) <https://www.cnet.com/news/politics/eu-strengthens-disinformation-rules-to-target-deepfakes-bots-fake-accounts/>
- Derakhshan, H., & Wardle, C. (2017). Information disorder: definitions, *Understanding and addressing the disinformation ecosystem*, pp5-12.
- European Commission (2018). *A Multi-dimensional approach to disinformation*. (Retrieved from) <https://op.europa.eu/en/publication-detail/-/publication/6ef4df8b-4cea-11e8-be1d-01aa75ed71a1/language-en>
- _____ (2018). *2018 Code of Practice on Disinformation*. (Retrieved from) <https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation>
- _____ (2019.4.26.) *Tackling online disinformation: a European Approach*. (Retrieved from) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236>
- _____ (2021, 5, 26). *Guidance on Strengthening the Code of Practice on Disinformation*. (Retrieved from) <https://digital-strategy.ec.europa.eu/en/library/guidance-strengthening-code-practice-disinformation>
- _____ (2022, 6, 16). *The Strengthened Code of Practice on Disinformation 2022*. (Retrieved from) <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>
- _____ (2023, 12, 10). *Signatories of the 2022 Strengthened Code of Practice on Disinformation*. (Retrieved from) <https://digital-strategy.ec.europa.eu/en/library/signatories-2022-strengthened-code-practice-disinformation>

- FullFact (2023). The Online Safety Act and Misinformation: What you need to know (Retrieved from) <https://fullfact.org/about/policy/online-safety-bill/>
- Gerken, T. & McMahon, L. (2022, 6, 16). Big Tech must deal with disinformation or face fines, says EU. *BBC*. (Retrieved from) <https://www.bbc.com/news/technology-61817647>
- Google (2023. 1). *Code of Practice on Disinformation- Report of Google for the period 1 July 2022 - 30 September 2022*.
- _____ (2023. 7). *Code of Practice on Disinformation- Report of Google for the period of 1 January 2023 - 30 June 2023*
- Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A. H. M., Hasan, S. M., Kabir, A., ... & Seale, H. (2020). COVID-19-related infodemic and its impact on public health: A global social media analysis. *The American journal of tropical medicine and hygiene*, *103*(4), pp1621-1629.
- Ipsos & Unesco (2023.9.) *Survey on the impact of online disinformation and hate speech*. (Retrieved from) <https://www.ipsos.com/ko-kr/elections-social-media-battle-against-disinformation-and-trust-issues>
- Killeen, M. (2021, 12, 7), Code of Practice on Disinformation revision extended into 2022. *Euractiv*. (Retrieved from) <https://www.euractiv.com/section/digital/news/code-of-practice-on-disinformation-revision-extended-into-2022/>
- Lai, S. & Yadavm K. (2023, 11, 23). Operational Reporting in Practice: The EU' s Code of Practice on Disinformation. *Carnegie Endowment for International Peace*. (Retrieved from) <https://carnegieendowment.org/2023/11/21/operational-reporting-in-practice-eu-s-code-of-practice-on-disinformation-pub-91060>
- McCarron, V. (2023.11.13.). Online Safety Act - An Overview (Retrieved from) <https://www.lexology.com/library/detail.aspx?g=c31e30f2-d20c-4463-8665-718fef3a9a4d>
- Meta (2023. 1). *Code of Practice on Disinformation- Meta Baseline Report*.
- _____ (2023, 6, 14). *EU CODE OF PRACTICE ON DISINFORMATION 2022 Subscription Document for Meta*.
- _____ (2023. 7). *Code of Practice on Disinformation- Report of Meta for the period 01 January 2023 to 30 June 2023*
- OECD (2022). *Disentangling untruths online : Creators, spreaders and how to stop them*. (Retrieved from) https://goingdigital.oecd.org/data/notes/No23_ToolkitNote_Unt

ruthsOnline.pdf

- Pew Research Center (2022.8.) Climate Change Remains Top Global Threat Across 19-Country Survey (Retrieved from) https://www.pewresearch.org/global/wp-content/uploads/sites/2/2022/08/PG_2022.08.31_Global-Threats_FINAL.pdf
- Poynter, Mediawise, YouGoc. (2022.8.) *A Global Study on Information Literacy*. (Retrieved from) <https://www.poynter.org/wp-content/uploads/2022/08/A-Global-Study-on-Information-Literacy-1.pdf>
- Tanner, B. (2022, 8, 5). COMMENTARY: EU Code of Practice on Disinformation. *Brookings Institution*. (Retrieved from) <https://www.brookings.edu/articles/eu-code-of-practice-on-disinformation/>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *science*, 359(6380), 1146-1151.

언론기사 및 보도자료

- 관계부처 합동 (2023.5.11.) 플랫폼 자율규제 방안 발표회 개최.
- 경향신문(2022.2.9.) 코로나 확산에 가짜뉴스도 범람
- 국민일보 (2023.8.30.) 메타, 中관련 8000여개 가짜 계정 적발... “가장 거대한 비밀 공작”
- _____ (2020.3.8.) “앞으로 휴지 못산다” 가짜뉴스 때문에 머리채 잡은 손님들
- 동아사이언스 (2020.8.13.) 코로나19 가짜뉴스로 1~3월 사이 800명 숨졌다.
- 메가시티뉴스 (2023.8.1.) 제프르 + 메타, 인스타그램 피드, 추가 언어에 AI 기반 브랜드 적
합성 측정 기능 확장.
- 메디게이트뉴스(2021.12.21.). ‘가짜뉴스’에 신음하는 코로나19 정국... “해결 위해 정부-
전문가 손 잡을 때” .
- 미디어오늘 (2023.9.6.) 사이버불링 막기 위해 한승연, 엔조이커플, 너털트 한자리에
- 보건복지부문화체육관광부 (2021.5.5.) 코로나19 관련 인식조사(4월)
- 비즈니스위치 (2023.5.11.) “검색 투명성 높인다” 플랫폼 자율규제 방안 발표.
- 서울경제(2023.5.27.) 트위터, EI ‘허위정보 협정’ 탈퇴...머스크의 ‘마이웨이’

서울대학교 (2020.9.8.). 5월 “코로나 기회론” 우세, 8월엔 “위기론” 인식 역전 (보도자료)

연합뉴스 (2020.3.20.) 코로나19 관련 음모론 제기한 해외 트윗 3주새 200만개.

전자신문 (2023.8.30.) 메타, 중국 배후 ‘가짜뉴스’ 계정 7700개 적발.

한국인터넷자율정책기구(2021.11.14.) 코로나19 허위조작정보 5건 중 4건은 ‘백신 괴담’
(보도자료)

AI타임즈 (2023.9.7.) 구글, 선거 광고에 ‘AI 생성 콘텐츠’ 공개 의무화.

KBS (2023.11.6.) EU·영국도 비슷한 ‘가짜뉴스’ 대응?... “표현의 자유 우선 고려” [가
짜뉴스]③.

BBC (2016.11.16.) ‘Post-truth’ declared word of the year by Oxford Dictionaries.

Computerworld (2022, 7, 17). Big tech platforms sign up to the EU Commission’s new
Code of Practice on Disinformation

Fonerena (2021. 9. 30). Google Search introduces Multitask Unified Model (MUM) with a
dvanced searching capabilities.

Sydney Morning Herald (2023.10.1.) ‘Very concerned’ : Influential senator David Pococ
k raises free speech objections to misinformation bill

Theverege (2019. 4. 1). Facebook’s News Feed is starting to explain itself.

_____ (2019. 10. 22). Facebook will label false posts more clearly as part of an eff
ort to prevent 2020 election interference.

WIRED (2023.10.26.) The UK’s Controversial Online Safety Act Is Now Law

허위조작정보에 대한 강화된 실천강령 2022

The Strengthened Code of Practice on Disinformation 2022

I. 서문 PREAMBLE

(a) 허위조작정보에 대한 강화된 실천강령에 관한 유럽위원회 지침(이하 “지침”)⁴⁶⁾, 유럽 민주주의 시행 계획(the European Democracy Action Plan)에 관한 유럽위원회의 커뮤니케이션(이하 “EDAP”)⁴⁷⁾, “허위조작정보에 대한 대응: 유럽식 접근법(Tackling online disinformation: a European approach)”에 관한 유럽위원회의 커뮤니케이션(이하 “커뮤니케이션”)⁴⁸⁾, 2018년 6월 28일 이사회 결론⁴⁹⁾, 그리고 허위조작정보⁵⁰⁾ 유포에 따른 문제를 해결하기 위해 유럽 전역에서 이뤄지고 있는 여러 중요한 계획들에 이어, 본 실천 강령(이하 “강령”)의 서명인⁵¹⁾은 허위조작 정보 대응에 이바지할 그들의 역할을 인지하고 있으며, 본 강령의 다른 부분에 잘못된 정보(misinformation)⁵²⁾, 허위조작정보(disinformation)⁵³⁾, 정보 영향력 작전

46) COM(2021)262final <https://digital-strategy.ec.europa.eu/en/library/guidance-strengthening-code-practicedisinformation>

47) COM(2020)790final <https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=COM%3A2020%3A790%3AFIN&qid=1607079662423>

48) COM(2018)236final http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51804

49) <http://www.consilium.europa.eu/media/35936/28-euco-final-conclusions-en.pdf>

50) 그럼에도 불구하고 ‘허위조작정보’의 개념은 오해의 소지가 있는 광고, 보도 오류, 풍자 및 패러디, 명확하게 식별되는 편파적인 뉴스 및 논평을 포함하지 않으며, 오해의 소지가 있는 광고에 관한 구속력 있는 법적 의무, 자율 규제 광고 강령 및 표준을 침해하지 않는다.

51) 부록[1]에 나열된 기업, 조직 및 협회

52) EDAP에 정의된 대로: “잘못된 정보란 선의로 친구나 가족에게 거짓 정보를 공유하는 경우와 같이 그 효과가 여전히 해로울 수 있지만 해로운 의도 없이 공유되는 허위 또는 오해의 소지가 있는 콘텐츠를 말한다”.

(information influence operations)⁵⁴⁾ 및 정보 공간 내 해외 개입⁵⁵⁾을 포함하기로 했다.⁵⁶⁾

- (b) 서명인은 “오해의 소지가 있거나 명백한 허위 정보 등 대규모 허위조작정보에 시민들이 노출되는 것은 유럽의 주요 과제”이며, “우리의 열린 민주사회는 충분한 정보를 갖춘 시민들(well-informed citizens)이 자유롭고 공정한 정치적 절차를 통해 자신의 의사를 표현할 수 있는 공개 토론(public debate)에 달려있다” 라는 유럽위원회의 결론을 인지하고 동의한다.⁵⁷⁾
- (c) ‘커뮤니케이션’ 에서 강조하였듯, 허위조작정보에 대응하기 위해 취해지는 모든 조치에서 기본권이 충분히 보장되어야 한다.⁵⁸⁾ 서명인은 표현의 자유, 정보와 사생활의 자유에 대한 기본권, 그리고 이러한 기본권 보호와 그밖에 합법적인 콘텐츠의 확산과 영향을 제한하기 위한 효과적 조치들 사이의 까다로운 균형을 유념하고 있다.
- (d) 서명인은 허위조작정보 유포가 온라인과 오프라인 모두에서 여러 양상을 보이고 있고, 광범위한 행위자들에 의해 촉진되거나 반대로 이들에게 영향을 미치며, 생태계 내 모든 이해관계자들이 이것의 확산에 대응하기 위해 수행할 역할이 있음

53) EDAP에 정의된 대로: “허위조작정보는 경제적 또는 정치적 이득을 취하거나 확보하려는 의도로 확산되어 대중에게 해를 끼칠 수 있는 허위 또는 오해의 소지가 있는 콘텐츠를 말한다” .

54) EDAP에 정의된 대로: “정보 영향력 작전이란 허위조작정보와 함께 독립적인 정보 출처를 억압하는 등 다양한 기만적 수단을 사용하여 이용자 대상에게 영향을 미치기 위한 국내 또는 해외 행위자의 조직적인 노력을 의미한다” .

55) EDAP에 정의된 대로: “정보 공간 내 해외 간섭은 광범위한 혼합 작전의 일환으로 수행되며, 외국 국적 행위자 또는 그 대리인이 개인의 자유로운 정치적 의사 형성과 표현을 방해하려는 강압적이고 기만적인 노력으로 이해될 수 있다” .

56) 허위조작정보는 서명인마다 다르지만 그럼에도 불구하고 설명한 문제를 해결하는 것을 목표로 하는 여러 운영 정책 및 도구를 통해 해결될 수 있다.

57) 커뮤니케이션(Communication), 1항.

58) 커뮤니케이션, 3.1.1항을 보면, 유럽위원회는 허위조작정보에 대응하기 위해 취하는 모든 조치는 “표현의 자유를 엄격하게 존중해야 하며, 비판적, 풍자적, 반대적 또는 충격적인 발언에 대한 검열 등 오용을 방지하는 안전장치를 포함해야 한다고 명시하고 있다. 또한 개방적이고 안전하며 신뢰할 수 있는 인터넷에 대한 유럽위원회의 약속을 엄격하게 존중해야 한다” .

을 인지한다.

- (e) 서명인은 2018년 실천강령⁵⁹⁾의 중요한 성취를 인지하고, 이에 바탕하며, 허위조작 정보에 대응하는 데 있어 더욱 효과적인 수단을 마련하기 위해서 강화된 강령이 필요하다는 점을 인지한다.

이러한 목적에 따라 서명인은 각 서명인이 작성한 준칙과 조치를 명시한 본 강령의 초안을 작성하였다.

- (f) 서명인은 그들과 그들의 자회사가 제공하는 제품, 활동, 그리고 서비스와 관련된 준칙(Commitments) 및 조치(Measures)에 서명하는 데 동의해야 하며, 또한 해당 준칙 및 조치에 의해 적용받는 관련 서비스를 명시해야 한다. 서명인이 그들의 서비스가 관련성이 없거나 적절하지 않다는 이유로 준칙, 또는 조치에 서명하지 않을 경우, 가입 문서에 그 이유를 설명해야 한다(서비스 차원에서 관련성이 있는 경우).
- (g) 서명인은 관련성, 적절성, 실행 가능성이 있는 추가 준칙 및 조치들에 대한 동의를 고려하는 것을 목적으로 본 강령의 테스크포스에서 해당 사안들에 대한 조언을 감안하며 그들의 준칙과 조치를 검토하는 것에 동의한다. 시간이 지남에 따라 그들의 관행과 접근 방식이 발전하고 기술, 사회, 시장, 그리고 법제적 발전을 감안할 때, 서명인들은 또한 변화의 혜택을 받을 수 있는 준칙 또는 조치에 대한 테스크포스의 관심을 끌기 위해 제안할 수 있다.
- (h) 강령에 따른 조치는 디지털 서비스법(DSA)⁶⁰⁾이 시행되면 DSA의 규제요건 및 전반적인 목표를 보완하고 이에 부합되도록 조정될 것이다. DSA는 허위조작정보와 연관된 체계적 위협에 대처하기 위한 초대형 온라인 플랫폼⁶¹⁾의 자발적인 행동강령

59) SWD(2020)180 <https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement>

60) 디지털 서비스를 위한 단일 시장에 관한 유럽의회 및 이사회 규정 제안서(디지털 서비스법) 및 개정 지침 2000/31/EC(COM(2020)825final) 섹션 4에 정의된 대로 이 문서 내 DSA에 대한 언급은 유럽위원회에서 제안한 내용을 의미하는 것으로 이해해야 한다.

혹은 다른 공동규제 수단을 포함해 공동규제 프레임워크를 설정할 것이다.

- (i) 본 강령은 (초대형 온라인 플랫폼에 대한) DSA 제35조에 의거한 실천강령이 되는 것을 목표로 한다.
- (j) 초대형 온라인 플랫폼은 DSA에 따라 관련된 체계적 위험을 식별(위험 평가)하고, 해결(위험 완화)해야 한다. 따라서 그들의 서비스와 연관된 모든 준칙에 서명하는 것은 DSA 제27조에 따른 가능한 위험 완화 조치로 고려되어야 한다.
- (k) 초대형 온라인 플랫폼에 해당하지 않는 서명인들이 본 강령에 용이하게 참여하도록 하기 위해, 그들이 자사의 서비스와 연관된 준칙에 서명하고 서비스의 규모와 성격 가용 자원에 비추어 비례적인 조치를 통해 준칙을 이행하도록 한다.
- (l) 이를 위해 초대형 온라인 플랫폼이 아닌 서명인은 동의 서류에서 그들이 동의한 준칙들을 이행하기 위해 실행할 서비스에 대한 적절하고 균형적인 조치들, 그리고/또는 해당 준칙과 관련한 본 강령에 명시된 조치를 시간이 지남에 따라 완전히 이행할 수 있는 계획을 확인할 수 있는 선택권을 갖는다. 이런 경우에, 관련 서명인들은 해당 조치 관련에 제공 중인 질적 보고 요소(QRE)와 서비스 수준 지표(SLI)를 확인해야 한다. 태스크포스는 본 강령의 효과적인 기능에 맞춰 조정된 조치의 일관성을 정기적으로 검토해야 한다. 균형적인 보고 요건을 보장하는 관점에서 태스크포스는 서명인들을 위해 간소화한 보고 서식을 개발할 것이다. 보고 시, 서명인들은 해당 준칙과 조치를 지속적으로 적용하는 것이 여전히 적절하고 타당한지에 대해 매년 확인해야 한다.
- (m) 도구, 기술적 노력, 자선 활동, 또는 특정 전문성을 통해 허위조작정보 확산을 해결하는데 도움을 주는 서명인은 회사, 또는 조직, 그리고 온라인 생태계에서 그들이 수행하는 역할에 맞춘 준칙과 조치에 서명할 것을 약속한다.
- (n) 서명인은 본 강령에 본 강령에 서명한 무역협회가 그 특성상 구성원들을 대신해 의무를 이행하는 것은 아님을 인정한다. 그러나 이러한 협회들은 그 구성원들이 본 강령을 완전히 인지하고, 그들이 가입하거나 원칙을 적절히 준수하도록 장려할 것을 약속한다.

61) [DSA] 제25조에 정의된 바와 같이, 초대형 온라인 플랫폼에 적용되는 본 강령의 조항은 DSA 최종 본문 제4조의 의무가 적용되는 모든 법인을 포함해야 한다.

- (o) 서명인들은 본 강령에 서명한 후 6개월 이내에 서명한 준칙 및 조치를 이행한다. 이행 후 1개월이 경과한 시점에 질적보고요소(QRE)와 서비스수준지표(SLI)를 제출하고, 서명 후 7개월 이내에 서명인들은 유럽위원회에 강령에 따라 준칙을 이행한 방법을 자세히 설명하는 기본 보고서를 제출한다. 이는 기본 보고서가 서명인들이 제출하는 최초의 보고서이므로 개선이 필요할 수 있고, 시간이 지남에 따라 서명인들의 보고서가 더욱 상세해질 수 있음을 인식한 것이다. 첫 번째 보고서는 이전 강령에 따라 시행된 조치와 새 강령을 시행하기 위해 취해진 조치에 대한 비교도 포함해야 한다.
- (p) 서명인들은 합리적인 요청에 따라 적절한 정보를 이용할 수 있도록 하고, 유럽위원회의 질문과 자문에 응하는 등 본 강령의 기능에 대한 보고를 평가하는 데 있어 유럽위원회와 협력할 필요성을 인정한다. 또 서명인들은 강령의 이행 단계와 모니터링 단계에서 유럽시청각미디어서비스규제그룹(EGRA) 및 유럽디지털미디어관측소(EDMO)와 긴밀히 협력하고 이에 참여할 필요성을 인지한다.
- (q) 본 강령은 EU 및 회원국의 기존 법률 틀 안에서 적용되어야 하고, 어떤 방식으로든 기존 및 향후 법률 체계를 대체하거나 해석한다는 식으로 여겨져선 안 된다. 특히 다음과 같은 경우들(이것에만 국한되지 않음):
- 유럽연합 기본권 헌장;
 - 유럽인권조약;
 - 전자상거래지침(Directive 2000/31/EC) 중 네트워크, 검색엔진, 브라우저, 온라인 블로그 플랫폼, 온라인 포럼, 동영상 공유 플랫폼, 소셜 미디어 등의 제공업체와 같은 단순 전달자와 캐싱, 호스팅 제공업체들을 대상으로 하는 본 강령의 모든 의무에 적용되는 지침들인 특히 제12조~제15조;
 - 개인 데이터 처리 및 해당 데이터의 자유로운 이동 관련 자연인 보호에 관한 규정(Regulation (EU) 2016/679);
 - 공중 전자적 통신 서비스 또는 공중 통신망의 제공에 있어서 생성되거나 처리되는 데이터의 저장과 지침(Directive 2002/58/EC);
 - 불공정 상업 관행에 관한 지침(Directive 2005/29/EC);
 - 허위광고 및 비교광고에 관한 지침(Directive 2006/114/EC);
 - 시청각 미디어 서비스 제공 관련 회원국의 법률, 규정, 또는 행정조치에 의해 규정된 특정 협의에 관한 지침(Directive 2010/13/EU);

- 시청각 미디어 서비스 제공에 관한 지침(Directive 2018/1808);
 - 비공개 노하우 및 비즈니스 정보(영업 비밀)의 불법적인 취득, 사용 및 공개에 대한 보호에 관한 지침(Directive 2016/943);
 - 온라인 플랫폼 시장의 공정성 및 투명성 강화를 위한 지침(Regulation 2019/1150);
 - 유해 콘텐츠에 대한 접근 및 유통을 제한하기 위해 고안된 조치의 비례성에 관한 유럽사법재판소(CJEU) 및 유럽인권재판소(ECHR)의 판례.
- (r) 본 강령은 허위조작정보 대응을 위한 다른 조치들을 침해하지 않는다.
- (t) 본 강령은 유럽경제지역 체약 당사국에서 제공하는 서비스에 대해 서명인별로 제한적으로 적용된다.
- (u) 서명인은 태스크포스에 통지함으로써 본 강령을 철회할 수 있다. 이러한 철회는 다른 서명인들의 본 강령에 대한 효력 종료 효과를 가지 않는다.
- (v) 추가 서명인들은 언제나 본 강령에 서명할 수 있다. 이를 위해, 서명 후보인은 서문에 명시된 원칙에 의거한 본 강령에서 그들이 서명하고자 하는 준칙 및 조치와 함께 그들의 활동을 태스크포스에 제시해야 한다.

II. 광고 게재에 대한 정밀 검토 SCRUTINY OF AD PLACEMENTS

다음과 같은 사실로:

- (a) 서명인은 2021년 유럽위원회 지침⁶²⁾에 명시된 우선순위를 인정하며, 상업적 측면은 허위조작정보의 여러 측면 중 하나에 불과함을 염두에 두고, “허위조작정보 제거자의 수익을 줄이기 위해 광고 게재 위치에 관한 조사를 대폭 개선” 할 필요성을 인정한다.⁶³⁾

62) European Commission Guidance on Strengthening the Code of Practice on Disinformation, COM(2021) 262 final, 26 May 2021, Section 6.1, page 12.

63) Tackling online disinformation: a European Approach, COM(2018) 36 final, 26 April 2018.

- (b) 서명인은 다음 유형의 조직과 그들의 책무를 통해 미디어 및 광고에서 허위조작정보를 제거하기 위해 집단적, 개별적 책임을 인정합니다: 광고 슬롯 구매에 관여하는 광고주 및 대행사, 광고 슬롯 판매 및 광고 캠페인 승인에 관여하는 퍼블리셔⁶⁴ 및 플랫폼, 광고 공간 및/또는 콘텐츠의 타겟팅 또는 선택 및 검증 보고에 관여하는 광고 기술 회사, 타겟팅에서 보고에 이르는 서비스 인증에 관여하는 감사 기관. 이러한 모든 당사자는 허위조작정보 페이지에서 광고 슬롯 판매와 유해 허위조작정보 캠페인 게시를 보장하기 위해 협력하고 개별 자원을 사용한다.
- (c) 서명인은 서명인 보유 서비스 또는 제3의 웹사이트에서의 허위조작정보의 유포를 막기 위한 기존 준칙⁶⁵들을 강화하고 확대할 필요성을 인지한다.
- (d) 서명인은 광고 메시지 및 서비스를 통한 유해 허위조작정보 유포에 대응할 필요성을 인지한다.
- (e) 관련 서명인은 온라인 광고 배포 관련 허위조작정보 위협을 해결하기 위해 세분화된 맞춤형 조치를 취할 필요성을 인지한다. 조치들은 모든 온라인 광고에 적용된다.
- (f) 관련 서명인은 허위조작정보 행위자로부터 금전을 받거나, 또는 그러한 웹사이트 및 계정을 홍보하지 않도록 정책과 프로세스를 구현하는 것이 중요함을 인지한다.
- (g) 서명인은 온라인 광고 매매 및 광고 서비스 제공 관련 모든 당사자들이 온라인 광고 생태계 전반의 투명성을 개선하고, 허위조작정보 제공자의 계정 및 웹사이트에서의 광고 게재를 효과적으로 조사, 통제, 그리고 제한하기 위해 협력해야 함을 인지한다.
- (h) 온라인 허위조작정보 웹사이트 내 광고 게재를 금지하려면, 해당 콘텐츠의 특성을

64) 퍼블리셔는 웹사이트 소유자, 블로그 소유자, 팟캐스트 제작자, 동영상 콘텐츠 제작자, 모바일 앱/게임 등 디지털 광고가 표시되거나 이용자 정보가 수집되어 디지털 광고에 사용되는 웹사이트, 앱 또는 기타 콘텐츠의 운영자를 의미한다.

65) “허위조작정보”의 개념은 오해의 소지가 있는 광고, 보도 오류, 풍자 및 패러디, 명확하게 식별되는 편파적인 뉴스 및 논평을 포함하지 않으며, 오해의 소지가 있는 광고에 관한 구속력 있는 법적 의무, 자율 규제 광고 강령 및 표준을 침해하지 않는다”라고 명시되어 있다.

고려해서, 이러한 과제를 지속적으로 성공적으로 해결하기 위해 이미 널리 사용 중인 브랜드 세이프티 도구에 대한 개선이 요구된다.⁶⁶⁾ 서명인들은 브랜드 세이프티 도구의 가용성과 활용성을 개선할 필요성을 인지한다.

- (i) 서명인들은 허위조작정보 콘텐츠에 관한 정보와 대응 방법을 제시하는 제3의 기관과 정보를 통합하고, 협력하는 것이 허위조작정보를 통한 수익 창출을 줄이는 데 중요하다는 것을 인지한다.

따라서, 본 강령의 서명인들은 다음을 약속한다:

허위조작정보의 재정적 인센티브(수익화 방지) DEMONETISATION OF DISINFORMATION

준칙 1. 광고 게재에 참여하는 관련 서명인은 허위조작정보 유포 관련 수익 지원을 막고, 수익화할 콘텐츠의 적격성, 수익화 및 광고 배치에 대한 통제, 그리고 광고 배치 관련 통제 및 서비스의 정확성과 효과에 대해 보고할 데이터를 결정하는 정책 및 제도 개선을 위해 노력한다.

준칙 1을 충족하기 위해:

조치 1.1. 미디어 플랫폼, 퍼블리셔 및 애드테크 기업을 포함한 광고 판매에 관여하는 서명인은 다음과 같은 목표를 갖는 정책을 배포, 공개, 및 시행할 것이다:

- 첫째, 광고 지원 사업의 무결성을 보호하기 위해 유해 허위조작정보의 게시 및 전달을 금지한다.
- 둘째, 이러한 정책을 반복해서 위반하는 허위조작정보 콘텐츠, 또는 소스 옆에 광고가 게재되는 것을 금지하기 위해 의미 있는 실천 및 개선 조치를 취한다.
- 셋째, 광고의 랜딩/목적지 페이지 및 광고 게재의 출처를 확인할 수 있는 조치를 채택한다.

관련 서명인은 실무 그룹을 구성해서 허위조작정보 소스를 지원하는 광고

66) Communication ‘Tackling online Disinformation: a European Approach’, http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51804.

광고 관련한 데이터를 포함해 유통금지 노력을 보고하는 방법론을 개발하는 동시에, 재무 보고와 방법론적 문제와 관련한 법적 제약을 고려해 6개월 이내 태스크포스가 기준 보고서에 추가 SLI를 포함하도록 하는 권고안을 제시할 것이다. SLI는 회원국 및 관련 서명인 차원에서 취한 조치의 영향력에 대한 통찰력을 제공해야 한다.

QRE 1.1.1: 서명인은 조치 1.1의 목표를 달성하기 위한 개발, 배포 및 시행하는 정책을 공개하고 개요를 설명하며, 도움 센터(help centres) 내 공개 페이지에 제공한다.

SLI 1.1.1: 서명인은 그들이 이 서비스 수준 지표에서 질적 부분에서 언급된 각 정책을 시행하기 위해 취한 조치에 대해 회원국, 또는 언어 수준에서 정량적으로 보고할 것이다. 예를 들어, 유해 허위조작정보를 유포하는 페이지 및/또는 도메인 내 광고 제거, 차단, 또는 다른 방식으로 제한하는 조치가 포함될 수 있다.

조치 1.2. 허위조작정보에 대응하는 조치 1.1에 언급된 정책을 위반하는 행위가 개입된 콘텐츠를 조직적으로 게시하는 행위자들에 의한 참여를 금지하고, 당사자들을 효과적으로 조사하기 위해 퍼블리셔, 미디어 플랫폼 및 애드테크 기업을 포함하여 광고 판매를 담당하는 관련 서명인은 필요에 따라 그들의 서비스에 콘텐츠 수익 창출 및 광고 수익 공유 프로그램에 대한 자격 요건과 콘텐츠 검토 과정을 강화한다.

QRE 1.2.1: 서명인은 유해 허위조작정보를 조직적으로 제공하는 행위자의 참여를 면밀히 조사하고 금지하기 위해 그들의 수익 창출 정책을 검토, 평가 및 보장하는 프로세스에 대한 윤곽을 제시한다.

SLI 1.2.1: 서명인은 보고 기간 중 조치 1.2와 관련된 정책 검토 및/또는 현행화 횟수를 보고해야 한다. 게다가 서명인은 상기 조치 1.1에 의거 공유된 지표(metrics)에 포함되지 않은 경우, 회원국 수준에서 이러한 정책의 결과로서 광고, 또는 수익 창출에 대한 참여가 금지된 계정, 또는 도메인 수를 보고해야 한다.

조치 1.3. 퍼블리셔, 미디어 플랫폼 및 애드테크 기업을 포함하여 광고 판매를 담당하는 관련 서명인은 광고 구매자에게 광고 게재에 대한 투명성을 제공하기 위

해 관련 제3의 접근 방식에 대한 지원을 포함해 상업적 및 기술적으로 실현 가능한 조치를 취한다.

QRE 1.3.1: 서명인은 조치 1.3과 관련, 광고 구매자에게 광고 게재에 관해 제공하는 통제 및 투명성에 대해 보고해야 한다.

조치 1.4. 광고주 및 대행사 등 광고 구매를 담당하는 관련 서명인은 허위조작정보 옆, 또는 반복적으로 허위조작정보가 게시되는 장소에 광고가 배치되지 않도록 효과적이고 투명한 조치를 취한 광고 판매자들을 통해 광고를 배치한다.

QRE 1.4.1: 광고 구매를 담당하는 관련 서명인은 조치 1.4에 명시된 조치를 하는 광고 판매자를 통해 광고를 게시하도록 보장하는 프로세스 및 절차를 형성한다.

조치 1.5. 미디어 플랫폼, 광고 네트워크, 광고 검증 회사를 포함하여 수익 창출 활동 보고에 관여하는 관련 서명인은 독립적인 제3의 감사 기관이 그들의 서비스와 데이터에 대한 상업적으로 적절하고 공정한 접근을 할 수 있도록 다음과 같은 단계를 수행해야 한다:

- 첫째, DSA에 따라 수행되는 정기 감사와 일치하게끔 수익 창출 및 허위 조작정보와 관련 자체(the first party) 보고에 대한 정확성을 확인한다.
- 둘째, 인증 서비스는 허위조작정보 정책을 포함한 미디어 플랫폼의 정책 집행의 효율성을 평가해야 한다.

QRE 1.5.1: 자체 보고를 하는 서명인은 조치 1.5에 명시된 대로 독립적인 제3의 감사 기관에 제공한 접근 권한을 보고하고, MRC 콘텐츠 수준 브랜드 안전 인증, 태그 브랜드 안전 인증 또는 기타 유사하게 인정되는 업계 공인 인증과 같은 해당 감사인의 공개 보고서 및 결과와 관련시킨다.

QRE 1.5.2: 감사를 통해 독립적인 인증을 수행하는 서명인은 조치 1.5의 요구사항을 반영하여 개선된 인증 영역을 공개한다.

조치 1.6. 관련 서명인은 다음과 같은 목표를 가지고 브랜드 세이프티 도구 및 파트너십의 사용을 발전시키기 위해 개발을 진행하고, 가용성을 개선하며, 실질적인 조치를 취한다:

- 상업적으로 가능한 범위 내에서 관련 서명인은 광고 구매자, 즉 광고주와 대행사에 의해 광고 게시 결정에 도움을 줄 수 있는 허위조작정보 캠페인 등과 같은 정보를 제공하는 신뢰 지표 제공 출처 평가자와 서비스, 팩트체커, 연구자 또는 기타 관련 이해관계자들로부터 정보 및 분석을 통합한 선택권을 제시한다.
- 광고주, 대행사, 애드테크 기업, 미디어 플랫폼 및 퍼블리셔는 허위조작정보 콘텐츠 옆, 및/또는 허위조작정보를 반복적으로 게시하는 장소 또는 출처에 광고가 게재되지 않도록 미디어 계획, 구매 및 보고 프로세스 전반에 걸쳐 브랜드 세이프티 도구 사용을 통합하기 위한 효과적이고 합리적인 조치를 취해야 한다.
- 콘텐츠와 도메인을 분류하는 브랜드 세이프티 도구 제공업체 및 평가 서비스는 상업적으로 민감한 정보를 공개하거나 영업 비밀을 누설하지 않는 한, 사용하는 프로세스에 대해 합리적인 투명성을 제공하고 고객 피드백 및 이의를 제기할 수 있는 메커니즘을 구축해야 한다.

QRE 1.6.1: 광고를 게시하는 서명인은 구매자들의 광고 게시 결정에 도움이 되는 허위조작정보 캠페인의 출처 등의 정보를 제공하는 광고를 게재하는 제휴사는 출처 평가자, 신뢰도 지표를 제공하는 서비스, 팩트체커, 연구자 또는 기타 관련 이해관계자들을 위해 그들이 제공할 선택안을 보고해야 한다.

QRE 1.6.2: 광고를 구매하는 서명인은 해당 서비스를 통해 보호되는 미디어 투자 비율(정성적 보고 요소 + SLI 모두)을 공개하면서, 광고 및 미디어 운영에 브랜드 세이프티 도구 사용을 통합하기 위해 취한 단계를 제시해야 한다.

QRE 1.6.3: 브랜드 세이프티 도구를 제공하는 서명인은 프로세스 및 결과에 대한 투명성과 이익제기 가능성을 어떻게 보장하고 있는지 설명해야 한다.

QRE 1.6.4: 허위조작정보를 지속적으로 제공하는지를 판단할 출처를 확인하는 유관 서명인은 웹사이트가 평가되는 기준에 대한 합리적인 정보를 제공하고, 허위조작정보 관련 기준에 대한 평가를 공개하고, 비정치적인 방식으로 운영하며, 평가가 발표되기 전에 게시자에게 답변할 권리를 부여해야 한다.

SLI 1.6.1: 광고를 구매하는 서명인은 광고 및 미디어 운영에 브랜드 세이프티 도구 사용을 통합하기 위해 취한 조치를 제시하고, 해당 서비스를 통해 보호되는 미디어 투자 비율을 공개한다.

허위조작정보를 포함한 광고에 대처 TACKLING ADVERTISING CONTAINING DISINFORMATION

준칙 2. 관련 서명인은 광고시스템이 광고 메시지의 형태의 허위조작정보 유포로 오용되는 것을 방지한다.

준칙 2를 충족하기 위해:

조치 2.1. 관련 서명인은 광고 메시지 및 콘텐츠 홍보에서 유해 허위조작정보를 유포하기 위해 광고 시스템을 오용하는 문제를 해결하기 위한 적절하고 맞춤형된 광고 정책을 개발, 배포 및 시행한다.

QRE 2.1.1: 서명인은 2.1 조치의 목표를 달성하기 위해 개발, 배포 및 시행할 정책을 공개, 설명하며, 도움 센터(help centres) 내 공개 페이지에 제공한다.

SLI 2.1.1: 서명인은 이 서비스 수준 지표의 질적 부분에 언급된 각 정책을 시행하기 위해 취한 조치를 회원국 또는 언어 수준에서 정량적으로 보고해야 한다. 예를 들어, 여기에는 광고 메시지 및 콘텐츠 홍보에서 유해 허위조작정보를 제거, 차단 또는 기타 방식으로 제한하는 조치가 포함될 수 있다.

조치 2.2. 관련 서명인은 유해 허위조작정보를 배포하는 콘텐츠 및 출처를 식별하고, 조치 2.1에 언급된 허위조작정보 관련 광고 정책을 위반하는 광고 및 홍보 콘텐츠를 식별하기 위해 공공 및 민간의 독립적인 정보 출처(예: 팩트체크, 또는 출처 평가 기관과의 협력, 신뢰 지표 제공 서비스, 또는 내부적으로 개발한 적정 방법)를 참조할 수 있는 도구, 방법, 파트너십을 개발해야 한다.

QRE 2.2.1: 서명인은 조치 2.1에 언급된 정책을 위반하는 콘텐츠 및 출처를 식별하기 위해 사용하는 도구, 방법 또는 파트너십을 설명하는 동시에 악의적인 행위자가 이러한 도구, 방법 또는 파트너십을 쉽게 우회할 수 있는 정보를 공개하지 않도록 유의해야 한다. 서명인은 이러한 도구, 방법 또는 파트너십에 포함된 독립적인 정보 출처를 명시해야 한다.

조치 2.3. 관련 서명인은 광고 메시지, 홍보 콘텐츠 및 사이트 방문 페이지 내 허위조작정보에 관한 광고 정책을 준수하지 않는 광고가 자사 서비스를 통해서 또는 자사 서비스 내에 게시되는 것을 방지하기 위해 현재 광고 검증 및 검토 시스템을 적절하고 상업적으로 실현 가능한 대로 조정한다.

QRE 2.3.1: 서명인은 그들의 서비스를 통해 게재되는 광고가 조치 2.1에 명시된 대로 광고 정책을 준수하도록 보장하기 위해 사용하는 시스템 및 절차를 제시해야 한다.

SLI 2.3.1: 서명인은 조치 2.3에 명시된 절차에 따라 서비스에서 제거되거나 금지된 광고에 대해 회원국 수준에서 정량적으로 보고한다. 광고가 성공적으로 삭제된 경우, 당사자는 위반 콘텐츠 및 광고의 도달 범위를 보고한다.

조치 2.4. 관련 서명인은 상기 조치 2.1에 설명된 정책을 위반하는 광고를 거부 또는 삭제하거나 이러한 정책을 적용하여 광고 계정을 비활성화할 경우 광고주에게 위반된 광고 정책에 대한 관련 정보를 제공하고 이의제기 절차를 명확히 제시한다.

QRE 2.4.1: 서명인은 위반한 광고 정책에 대한 정보를 광고주에게 제공하는 방법과 광고주가 이러한 정책에 이의를 제기할 수 있는 방법을 제시한다.

SLI 2.4.1: 서명인은 정책 적용과 관련해 광고주로부터 받은 표준 절차에 따른 이의제기 건수와 그중 초기 정책 결정의 변경을 이끈 이의제기 비율을 회원국 수준에서 정량적으로 보고한다.

관련 당사자들과의 협력 COOPERATION WITH RELEVANT PLAYERS

준칙 3. 디지털 광고의 구매, 판매 및 게재에 관여하는 관련 서명인은 그들의 서비스에서 이루어지는 광고 게재에 대한 정밀검토의 효과를 높이기 위하여 온라인 결제 서비스, 전자 상거래 플랫폼 및 클라우드 펀딩/기부 시스템 등 온라인 수익화 가치 사슬에서 활동하는 조직들로 확대한 관련 행위자들과 모범 사례를 공유하고 협력을 강화한다.

준칙 3을 충족하기 위해:

조치 3.1. 관련 서명인은 모든 관련 데이터 보호 규칙 및 기밀 유지 계약을 완전히 준수하여 정보, 특히 유해 허위조작정보 제공자에 대처하는 데 관련된 정보의 통합 및 흐름을 촉진하기 위해 플랫폼, 광고 공급망 플레이어, 소셜 서비스, 신뢰 지표 제공 서비스, 팩트체크 기관, 광고주 및 온라인 수익 창출 가치 사슬에서 활동하는 기타 모든 행위자와 협력한다.

QRE 3.1.1: 서명인은 유해 허위조작정보 제공자를 대처하는데 관련하는 정보의 흐름을 촉진하기 위해 업계 및 시민사회 전반에서 다른 사람들과 협력하는 방법을 제시한다.

조치 3.2. 관련 서명인은 코드 태스크포스, GARM, IAB 유럽 또는 기타 관련 포럼을 통해 허위조작정보 경향 및 TTP(전술, 기법 및 절차)에 대한 정보를 상호 교환한다. 여기에는 광고 공급망 전반에서 허위조작정보를 제거하는 데 도움이 되는 서명인이 관찰한 새로운 기법이나 위협에 대한 통찰력 공유, 사례 연구, 기타 역량 및 단계 개선 수단에 대한 논의(잠재적으로 실시간 기술 역량 포함) 등이 포함된다.

QRE 3.2.1: 서명인은 기밀 정보 및/또는 악의적인 행위자가 서명인 및 기타 광고 공급망 전반에서 설정한 방어 수단을 우회하기 위해 사용할 수 있는 정보를 공개하지 않도록 유의하면서 조치 3.2에 언급된 포럼 내에서 논의한 내용을 보고한다. 예를 들어, 여기에는 서명인이 참여한 포럼 관련 정보, 공유한 정보의 종류, 이러한 교류를 통해 얻은 학습 내용이 포함될 수 있다.

조치 3.3. 관련 서명인은 허위조작정보의 수익 창출을 줄이고 허위조작정보가 포함된 광고의 유포를 방지하기 위해 독립적인 출처 평가 서비스, 신뢰 지표 제공 서비스, 팩트체커, 연구자 또는 오픈 소스 조사자 등 관련 제3의 조직과 업무를 통합하거나, 협력한다.

QRE 3.3.1: 서명인은 언급된 기관들과의 업무와 관련된 협업 및 통합에 대

해 보고한다.

III. 정치 광고 POLITICAL ADVERTISING

다음과 같은 사실로:

- (a) 서명인은 주요 사회적 이슈에 대한 정치 캠페인과 공개 토론, 특히 여론 형성, 정치 및 선거 토론, 국민투표, 입법 과정 및 시민의 투표 행동에 있어서의 정치 및 이슈 광고의 중요성을 인지한다.
- (b) 서명인 모든 광고주(미디어 기관 및 기업 포함)가 정치 또는 이슈 광고를 게재할 수 있음을 인지한다.
- (c) 서명인은 “정치 광고”에 대한 공동의 이해를 발전시키기 위해 함께 협력하고, 정치 광고의 투명성 및 타겟팅에 관해 유럽위원회에서 제안한 규정에 부합하도록 노력하며, 특히 해당 규정에서 정한 정의 및 의무와 관련하여 제안된 규정 및 DSA의 관련 조항을 준수하기 위한 접근의 중요성을 인지한다.
- (d) 서명인은 또한 정치 광고의 투명성 및 타겟팅에 관한 유럽위원회의 규정 제안과 DSA를 고려하여 “이슈 광고”에 대한 공동의 이해를 위해 협력할 필요성을 인지한다.
- (e) 서명인은 또한 정치 광고의 투명성 및 타겟팅에 관한 유럽위원회의 규정 제안의 맥락에서 “후원자”에 대한 합의된 이해의 중요성을 인지한다.
- (f) 서명인은 정치적 성향이나 다루는 이슈에 관계없이 정치 광고가 독립적으로 운영 되도록 하는 것이 중요하다는 점을 인지한다.
- (g) 서명인은 본 강화된 강령을 통해 정치 광고의 투명성에 기여하고, 특히 필요한 경우 유효한 동의를 얻는 것과 관련하여 GDPR 및 기타 관련 법률을 완전히 준수하면서, 정치 광고를 타겟팅시 개인 데이터를 사용할 때 추가적인 보호를 보장할 것을 인지한다.

따라서, 본 강령의 서명인들은 다음을 약속한다:

정치 광고와 이슈 광고에 대한 공통된 이해

A COMMON UNDERSTANDING OF POLITICAL AND ISSUE ADVERTISING

준칙 4. 관련 서명인은 ‘정치 광고 및 이슈 광고’에 대해 공통된 정의를 채택한다.

준칙 4를 충족하기 위해:

조치 4.1. 관련 서명인은 유럽연합 집행위원회가 제안한 정치 광고의 투명성 및 타겟팅에 관해 규정에 명시된 “정치 광고”의 정의에 따라 본 강령의 “정치 및 이슈 광고”를 정의할 것을 합의한다.

조치 4.2. 본 강령 운영 첫해에 정치 광고의 투명성 및 타겟팅에 관한 규정에 대한 유럽위원회의 제안에 대한 협상 과정에서 “정치 광고”의 정의에 대한 정치적 합의가 이뤄지지 않거나, (QRE 4.1.1에 따른 태스크포스에서 평가한 바와 같이) “이슈 광고”를 적절히 포괄하는 “정치 광고”의 정의가 해당 규정에 포함되지 않은 경우, 서명인은 이번 장의 기준이 될 수 있는 정치 광고 및 이슈 광고의 정의를 수립하기 위해 태스크포스와 협력한다.

QRE 4.1.1: (조치 4.1 및 4.2의 경우) 관련 서명인은 정치 광고의 투명성 및 타겟팅에 대한 유럽위원회의 제안 규정과 일치함을 입증하면서, 관련 정책을 보고하고 게시하는 시기에 대해 그들이 약속한 적절한 범위를 제시한다.

QRE 4.1.2: (조치 4.1 및 4.2의 경우) 본 강령 운영 첫해 이후, 서명인은 조치 4.2에 따라 그들이 태스크포스와 추가 작업이 필요하다고 평가하는지 여부와 이를 위한 메커니즘을 명시한다.

준칙 5. 관련 서명인은 자사의 서비스 안에서 정치 및 이슈 광고 전반에 대해 일관된 접근 방식을 적용하고, 자사의 광고 정책에 그러한 광고의 허용범위 및 금지 범위를 명확하게 표시할 것을 약속한다.

준칙 5를 충족하기 위해:

조치 5.1. 관련 서명인은 준칙 4 및 5와 관련된 모든 광고에 라벨링, 투명성 및 검증 원칙(아래에 명시된 바와 같이)을 적용한다. 관련 서명인은 정치 및/또는 이슈 광고에 대한 서비스의 정의와 관련된 정책 규칙 또는 가이드라인을 공개적으로 이용가능하고 쉽게 이해할 수 있는 방식으로 제시한다.

QRE 5.1.1: 관련 서명인은 정책 규칙 또는 지침과 이를 공개하기 위한 접근 방식에 대해 보고한다.

정치광고 또는 이슈광고에 대한 효과적인 라벨링

EFFICIENT LABELLING OF POLITICAL OR ISSUE ADS

준칙 6. 관련 서명인은 이용자가 콘텐츠에 정치 광고 또는 이슈 광고가 포함되어 있다는 것을 이해할 수 있도록, 정치 광고 또는 이슈 광고에 명확한 라벨링을 하고 해당 콘텐츠가 그로부터 대가를 받은 콘텐츠임을 식별할 수 있도록 해야 한다.

준칙 6을 충족하기 위해:

조치 6.1. 관련 서명인은 정치 또는 이슈 광고의 표시 및 라벨에 대한 공통 모범 사례 및 예시를 개발하고 해당 학습 내용을 서비스에 통합한다.

QRE 6.1.1: 관련 서명인은 조치 2.2.1의 일부로 개발된 모범 사례 및 예시를 공개하고, 관련 서비스와 어떻게 연관되는지 제시한다.

조치 6.2. 관련 서명인은 후원자의 정보 등 관련 정보가 광고를 광고에 부착된 라벨에 포함하거나, 이용자들이 이 라벨을 통해 쉽게 접근할 수 있도록 보장한다.

QRE 6.2.1: 관련 서명인은 후원자 정보 및 기타 관련 정보가 광고에 첨부되거나 라벨을 통해 사용자가 쉽게 접근할 수 있도록 하는 방법에 대한 예를 제시한다.

QRE 6.2.2: 관련 서명인은 그들의 라벨링 디자인을 제시한다.

SLI 6.2.1: 관련 서명인은 이러한 라벨링의 효과를 평가하는 것을 목적으로 태스크포스 내에서 논의하여 결정할 허용되고 라벨링된 광고의 수, 라벨링된 광고주가 지출한 금액, 또는 다른 방식 등 회원국 수준에서 의미 있는 평가 방식을 제시한다.

조치 6.3. 관련 서명인은 사용자의 라벨 식별 및 이해도를 개선하기 위한 연구에 투자 및 참여하고, 해당 연구 결과를 태스크포스와 논의하며, 해당 연구 결과를 해당 서비스에 통합하기 위해 노력한다.

QRE 6.3.1: 관련 서명인은 사용자가 정치 또는 이슈 광고에서 라벨을 식별하고 이해하는 방법을 이해하기 위한 관련 연구를 발표하고, 사용자가 일관되게 라벨을 식별할 수 있도록 보장하고 라벨의 사용자 인식 가능성을 개선하기 위해 취한 조치에 대해 보고한다.

조치 6.4. 관련 서명인은 유료 정치 또는 이슈 콘텐츠를 지속적으로 명확히 식별할 수 있도록 일단 그들의 플랫폼 내 정치 또는 이슈 광고에 해당 라벨이 표시되면, 이용자들이 동일 플랫폼 내 동일 광고를 공유할 때도 그 라벨이 그대로 유지될 수 있도록 보장한다.

QRE 6.4.1: 관련 서명인은 이용자가 광고를 공유할 때 라벨이 그대로 유지될 수 있도록 보장하는 조치를 설명한다.

조치 6.5. 현지법을 준수하는 선에서 메시징 서비스를 제공하는 서명인들은 메시징 서비스를 통해 공유된 정치 광고에 적용된 라벨의 가시성 개선을 위해 합리적인 노력을 해야 한다. 이를 위해 그들은 메시징 서비스를 통해 공유되는 유료 콘텐츠가 암호화를 약화시키지 않고 개인정보 보호와 관련하여 가능한 범위 내에서 온라인 플랫폼에서 해당 라벨이 부착된 유료 콘텐츠를 사용자가 쉽게 인식할 수 있는 솔루션을 개발하기 위해 합리적인 노력을 기울인다.

QRE 6.5.1: 관련 서명인은 조치 6.5에 명시된 바와 같이 이용자가 유료 콘

텐츠를 인식할 수 있도록 시행 중인 모든 솔루션을 보고한다.

정치광고 또는 이슈광고에 대한 검증 약속

VERIFICATION COMMITMENTS FOR POLITICAL OR ISSUE ADS

준칙 7. 관련 서명인은 정치 광고나 이슈 광고를 게재하는 스폰서 및 광고대행사를 대신하여 균형적이고 적절한 검증 시스템을 마련한다. 관련 서명인은 해당 광고를 게재하기 전에 라벨링 및 이용자가 직면하는 투명성 요건을 충족하는지 확인한다.

준칙 7을 충족하기 위해:

조치 7.1. 관련 서명인은 정치 또는 이슈 광고를 구매하는 후원자 및 후원자를 대리하는 광고 서비스 제공업체가 해당 광고의 게재를 허용하기 전에 후원자 관련 정보 또는 대리하는 후원자를 확인(및 적절한 경우 재확인)할 수 있도록 하는 관련 정보를 제공했는지 확인해야 한다.

QRE 7.1.1: 관련 서명인은 조치 7.1.1에 명시된 정보를 수집하고 검증하기 위해 마련한 도구 및 프로세스의 적시성 및 비례성에 대한 정보를 포함하여 보고한다.

SLI 7.1.1: 관련 서명인은 관련된 검증 프로세스를 이행하지 않아 거부된 광고의 양에 대해 서비스 및 회원국 수준에서 SLI 6.2.1의 지표와 비교 가능한 의미 있는 지표를 게시한다.

조치 7.2. 관련 서명인은 준칙 7에 명기된 검증 절차를 시기적절하고 비례성 있는 방식으로 완수한다.

QRE 7.2.1: 관련 서명인은 해당 정책 개선을 포함하여 상기 도구 및 절차를 명백히 회피하는 행위자에 대해 취한 조치를 보고한다.

QRE 7.2.2: 관련 서명인은 검증 절차의 적시성과 비례 원칙에 대한 정보를 제공한다.

조치 7.3. 관련 서명인은 플랫폼 전략을 통해 검증 및 투명성 요건을 명백히 회피하는 정치 또는 이슈 광고 후원자에 대해 계정 정지 또는 기타 계정 차원에 서의 패널티와 같은 적절한 조치를 취한다. 관련 서명인은 이용자가 정치 광고로 표시되지 않은 광고에 표시할 수 있는 기능을 개발하거나 기존 도구를 통해 제공한다.

QRE 7.3.1: 관련 서명인은 요청된 광고 서비스가 정치 또는 이슈 광고에 해당하는지 여부에 대한 신고를 요청할 수 있는 도구 및 프로세스를 보고한다.

QRE 7.3.2: 관련 서명인은 플랫폼 내 검증 및 투명성 요건을 명백하게 회피하는 정치 또는 이슈 광고 후원자에 대해 시행 중인 정책을 보고한다.

조치 7.4. 관련 서명인은 후원자 및 후원자를 대신하여 활동하는 광고 서비스 제공업체가 요청하는 광고 서비스가 정치 또는 이슈 광고에 해당하는지를 표명하도록 요청할 것을 약속한다.⁶⁷⁾

QRE 7.4.1: 관련 서명인은 정치 또는 이슈 광고 후원자의 정보를 확인할 수 있는 조치의 효과성에 대한 연구 및 공공데이터를 제시한다.

이용자에 대한 정치광고 또는 이슈광고의 투명성 약속

USER-FACING TRANSPARENCY COMMITMENTS FOR POLITICAL OR ISSUE ADS

준칙 8. 관련 서명인은 서비스 내 정치 광고나 이슈 광고에 대한 투명한 정보를 이용자에게 제공한다.

준칙 8을 충족하기 위해:

67) 해당되는 경우, 정치 광고의 투명성 및 타겟팅 관련 규정에 따라 정치 광고 및 광고주의 확인 및 식별에 허용되는 것으로 간주되는 방법은 본 강령에 따른 서명인의 약속에 허용되는 것으로 여긴다.

조치 8.1. 관련 서명인은 후원자의 정보, 표시 기간, 광고 지출, 광고 이용자에 대한 집계 정보 등 정치 광고의 투명성 및 타깃팅에 관한 규정을 유럽위원회의 제안과 일치하도록 하여 공동의 최소 투명성 의무에 합의한다.

조치 8.2. 관련 서명인 광고에서 광고 저장소로의 직접 링크를 제공한다.

QRE 8.1.1 (조치 8.1 및 8.2의 경우): 관련 서명인은 조치 8.1 및 8.2의 일부로 개발된 공동 최소 투명성 의무를 공표하고, 이용자가 이용할 수 있는 데이터들을 포함하여 관련 서비스와 어떻게 관련되는지 설명한다.

준칙 9. 관련 서명인은 이용자에게 정치 광고 또는 이슈 광고가 표시되는 이유에 대해 명확하고 이해하기 쉬우면서 포괄적인 정보를 제공한다.

준칙 9를 충족하기 위해:

조치 9.1. 관련 서명인은 유럽위원회가 제안한 정치 광고의 투명성과 타깃팅에 관한 규정과의 일치성을 추구하면서 이용자들이 특정 정치 광고나 이슈 광고를 보는 이유에 대한 정보에 접근할 수 있는 간단한 수단을 제공한다.

조치 9.2. 관련 서명인은 정치 또는 이슈 광고가 특정 이용자에게 표시되는 것을 결정하기 위해 후원자를 대신하여 활동하는 광고 서비스의 제공자와 후원자에 의해 사용된 근거와 도구(예: 인구통계학적, 지리학적, 맥락적, 관심 또는 행동 기반)를 단순하고 쉬운 언어로 설명한다.

QRE 9.1.1 (조치 9.1 및 9.2의 경우): 관련 서명인은 서비스가 제공한 타깃팅 방법에 대한 적절한 예시를 포함해 조치 9.1 및 9.2에 기술된 정보를 이용자에게 제공하기 위해 마련된 도구와 기능을 기술한다.

정치광고 또는 이슈광고 저장소 및 정치광고 또는 이슈 광고 데이터에 접근할 수 있는 API⁶⁸를 위한 최소한의 기능

준칙 10. 관련 서명인은 정치 광고 또는 이슈 광고의 저장소를 유지한다. 저장소는 서비스에서 제공된 모든 정치 광고 및 이슈 광고, 그리고 그러한 광고의 법적 의무 및 본 강령에 따른 투명성 약속을 준수하는데 필요한 모든 정보를 포함하며, 그것의 최신성, 완전성, 이용 가능성 및 품질을 보장한다.

준칙 10을 충족하기 위해:

조치 10.1. 관련 서명인은 광고 자체를 포함하여 제공되는 모든 정치 및 이슈 광고의 정확한 기록을 포함하는 (특히 선거 기간 동안, 실시간에 가까운) 검색 가능한 전용 광고 저장소를 설치하고 유지한다. 여기에는 후원자의 정보, 광고 진행 날짜, 광고에 지출된 총금액, 전달된 건수, 수신자를 결정하는 데 사용되는 수용자 기준, 광고를 본 이용자 수 및 인구 통계, 광고에 나타난 지리적 지역 등 각 광고에 대한 관련 정보가 수반되어야 한다.

조치 10.2. 이러한 광고 저장소의 정보는 최소 5년간 공개적으로 이용 가능하다.⁶⁸⁾

QRE 10.2.1 (조치 10.1 및 조치 10.2의 경우): 관련 서명인은 조치 10.1 및 조치 10.2를 준수하기 위해 저장소의 가용성, 기능 및 업데이트 지점에 대해 상세히 설명한다. 관련 서명인은 또한 저장소의 월별 이용량과 같은 정량적인 정보를 제공한다.

준칙 11. 관련 서명인은 이용자와 연구자가 광고 저장소에서 맞춤형 검색을 할 수 있도록 API 혹은 기타 인터페이스를 제공한다. API 또는 기타 인터페이스에는

68) [역주] API(Application Programming Interface), 소프트웨어 인터페이스로 컴퓨터나 컴퓨터 프로그램 사이의 연결

69) 서명인은 정치적 광고의 투명성과 타겟팅에 관한 규정이란 유럽연합 집행위원회 제안에 따라 광고 저장소에서 이용 가능한 최소 시간을 조정한다.

활용을 위한 일련의 최소한의 기능 및 검색기준이 포함되어 있어야 한다.

준칙 11을 충족하기 위해:

조치 11.1. 관련 서명인의 API 또는 기타 인터페이스는 사용자와 연구자가 조사 및 모니터링을 허용하기 위해 광고주 또는 후보별, 지리적 지역별 또는 국가별, 언어별, 키워드별, 선거별 또는 기타 대상 기준별 검색 등 표준 형식에서 가능한 한(특히 선거 기간 중) 실시간에 가까운 데이터 맞춤형 검색을 수행할 수 있도록 하는 일련의 최소 기능 및 검색기준을 제공한다.

조치 11.2. 관련 서명인들이 이러한 API 및 기타 인터페이스를 통해 이용 가능한 데이터는 해당 광고 보관소(ad repositories)에서 이용 가능하도록 만들어진 데이터 대비 더 상세하거나 동일한 수준이다.

조치 11.3. 관련 서명인은 API 및 기타 인터페이스에 대한 광범위한 접근 및 가용성을 보장한다.

조치 11.4. 관련 서명인은 적용 가능한 경우, 연구자들의 합리적인 요구를 충족시키기 위해 연구자들과 협력하고, API 및 기타 인터페이스의 기능을 개선한다.

QRE 11.1.1 (11.1-11.4 조치의 경우): 관련 서명인은 조치 11.1-11.4를 준수하기 위한 API 또는 기타 인터페이스의 가용성과 특징을 상세히 설명한다. 관련 서명인은 또한 월간 이용량 같은 API 이용량에 대한 정량적인 정보를 제공한다.

QRE 11.4.1: 관련 서명인은 API의 기능에 대한 특징과 관련한 그들의 경험을 이해하는 것을 포함해 연구자의 참여와 태스크포스 내 논의 및 개입의 결과로서 기능 개선에 대해 보고한다.

시민사회와의 약속 CIVIL SOCIETY COMMITMENTS

준칙 12. 관련 서명자들은 정치 및 이슈 광고에 대한 감독을 강화하고, 정치 또는 이슈 광고 정책과 관행의 작성, 실행 및 개선에 대해 적절하고 건설적으로 지원할 것을 약속한다.

준칙 12를 충족하기 위해:

조치 12.1. 관련 서명인은 회원국 전역에 걸친 온라인 정치 또는 이슈 광고 사용 관련 연구, 모니터링 및 보고를 수행한다.

조치 12.2. 관련 서명인은 특히 회원국의 선거 기간 동안 정치 또는 이슈 광고에 대한 적절한 정밀 조사를 보장하기 위해 도구, 대시보드 및 기타 데이터를 생산한다.

조치 12.3. 관련 서명인은 본 강령, 또는 정치 또는 이슈 광고 정책의 시행 또는 집행 관련 이슈를 동료 서명자들에게 알린다.

QRE 12.1.1 (조치 12.1-12.3의 경우): 관련 서명자는 자신의 연구 결과를 공개적으로, 그리고 논의 및 잠재적인 후속 조치를 위해 테스크포스와 공유할 것이다.

지속적인 협업 ONGOING COLLABORATION

준칙 13. 관련 서명인은 정치 광고 또는 이슈 광고의 허위조작정보와 관련된 위험을 이해하고 이에 대응하기 위해 지속적인 모니터링 및 연구에 참여하는 것에 동의한다.

준칙 13을 충족하기 위해:

조치 13.1. 관련 서명인은 개별적으로 그리고 테스크-포스를 통해 정치, 또는 이슈 광

고 사용에서 새롭게 진화 중인 허위조작정보의 위험을 확인하고, 그러한 위험을 해결하기 위한 선택안을 논의할 것을 동의한다.

조치 13.2. 관련 서명인은 태스크포스, 그리고 다른 관련 이해당사자들과 함께 모든 회원국 수준에서 그들의 서비스 내 정치 또는 이슈 광고에 대한 당기 선거 “블랙아웃 기간”의 영향력과 기회를 평가한다.

조치 13.3. 관련 서명인은 회원국 수준에서 정치 또는 이슈 광고에 대한 충분히 정밀할 조사 여부를 태스크포스와 함께 평가할 것을 동의한다.

QRE 13.1.1 (조치 13.1-13.3의 경우): 태스크포스를 통해 관련 서명인은 적어도 매년 다음과 관련된 적절한 자원을 가진 토론을 소집한다:

- 조정된 정책을 개발하기 위한 정치 광고의 새로운 위험성
- 블랙아웃 기간에 대한 지속적인 논의
- 정치적 또는 이슈 광고에 대한 독립적인 정밀 조사.

IV. 서비스의 완전성 INTEGRITY OF SERVICES

다음과 같은 사실로:

- (a) 관련 서명인은 서문에 정의된 바와 같이 서비스 전반에 걸친 허용 불가능한 조작 행위와 관행을 포함하여 잘못된 정보와 허위조작정보 모두에 대한 안전장치를 구현하고 홍보함으로써 서비스의 무결성을 보장하기 위한 노력의 효과성을 강화하고 입증하는 것의 중요성을 인지한다.
- (b) 해당 목적과 2021년 유럽위원회 지침과 같은 맥락에서 본 강령에 대한 토론 및 보고를 촉진하기 위해 관련 서명인은 유럽인권협약 제10조 및 연합의 기본권헌장 제11조, 제47조 및 제52조와 호환되는 기존 EU 및 현지 법에 대한 편견 없이 해당 서비스 전반에 걸친 허용되지 않는 조작적 행동 및 관행에 대한 공동의 이해를 개발하기 위해 함께 협력하는 것의 중요성을 인지한다.
- (c) 관련 서명인은 또한 서비스 전반에 걸친 허용되지 않는 조작 행위와 관행, 해외의

정보 조작 및 간섭, 그리고 보안을 위협하는 복합적인 위협에 대처하고 제한하기 위해 그들의 준칙을 지속적으로 강화할 필요성을 인지한다.

- (d) 관련 서명인은 허용되지 않는 조작 행위와 관행에 대한 포괄적인 용어 목록을 작성하는 데 함께 협력하는 것을 목표로 하고 있으며, 이는 특히 AMITT 허위 정보 전술, 기법 및 절차 프레임워크에 악의적인 행위자가 사용하는 행위와 전술, 기법 및 절차(TTP)에 대한 최신 증거에 비추어 주기적으로 수정되어야 한다. 이러한 협력은 본 강령에 의거 형성된 상설 태스크포스의 프레임워크에서 이루어질 수 있다.
- (e) 지속적인 관련성과 적정성을 보장하기 위해 관련 서명인은 서비스 전반에 걸쳐 허용되지 않는 조작적이고 가식적인 행동과 관행에 대한 책임뿐만 아니라 이용자에 대한 투명성 측면에서 그들의 노력을 강화하기로 동의한다.
- (f) 관련 서명인은 유럽인권협약 제8조와 유럽기본권헌장 제7조 및 제8조에 일관되게 익명 또는 가명의 계정 및 서비스 이용을 금지하지 않는 것이 중요하다는 점을 재차 강조한다.

따라서, 본 강령의 서명인들은 다음을 약속한다:

허용되지 않는 조작행위에 대한 공통된 이해

COMMON UNDERSTANDING OF IMPERMISSIBLE MANIPULATIVE BEHAVIOU

준칙 14. 서비스 전반의 허용되지 않는 조작 행위 및 관행을 제한하기 위해, 관련 서명인은 서비스 내 잘못된 정보와 허위조작정보 유포에 대응하기 위한 정책을 마련하거나 더욱 강화하며, 그들의 서비스에서 허용되지 않는 조작 행위, 행위자 및 관행에 관한 서비스 간 이해에 합의한다.
이러한 행위 및 관행은 악성 이용자의 특정 행위(AMITT⁷⁰) 허위조작정보 전술, 기법 및 절차 프레임워크) 및 TTPs⁷¹)에 대한 최신의 증거에 비추어 주기

70) [역주] AMITT : Adversarial Misinformation and Influence Tactics and Techniques

71) [역주] TTPs : Tactic, Technique, Procedur 사이버보안에서 공격자의 전략과 전술, 과정을 의미

적으로 검토되어야 하며, 여기에는 아래와 같은 항목이 포함된다.

- 가짜계정 생성 및 사용, 계정 탈취, 팔로워 증폭 봇(팔로우 봇)
- 해킹 및 유출
- 사칭
- 악성 딥페이크
- 가짜 참여⁷²⁾ 구매
- 인플루언서의 투명하지 않은 유료 메시지 또는 프로모션
- 조직적인 허위(inauthentic) 행위에 가담하는 계정 생성 및 사용
- 허위조작정보의 도달 범위나 관련 대중의 지지를 인위적으로 확대하는 행위

준칙 14를 충족하기 위해:

조치 14.1. 관련 서명인은 AMIT 디스인포메이션 전술, 기법 및 절차 프레임워크와 같은 악성 행위자가 사용하는 행위 및 전술, 기법 및 절차(TTP)에 대한 최신 증거를 기반으로 서비스에 허용되지 않는 조작 행위 및 관행에 관한 명확한 정책을 채택, 강화 및 시행한다.

QRE 14.1.1: 관련 서명인은 관련 정책을 나열하고 위에서 언급한 위협 및 기타 정보 유출 위협과 어떻게 관련되는지 명확히 한다.

QRE 14.1.2: 서명인은 이 준칙과 관련하여 허용할 수 없는 콘텐츠, 행위, TTP 및 관행을 탐지하기 위한 사전적 노력에 대해 보고한다.

조치 14.2. 관련 서명인은 해당 서비스 관련 금지된 행동과 관행을 명확히 하고 각 정책과 해당 정책의 이행이 상기 TTP, 위협 및 위해 및 기타 관련 위협을 어떻게 해결하는지에 대해 설명하는 공개 정책에 대한 상세하고 현행화된 목록을 보유한다. 이러한 정보는 투명성 센터(the Transparency Centre)에 보고된다. TTP 목록은 TTP가 보고될 때 기초가 되며 관련 서명인은 상설 태스크포스 내에서 관련 활동의 영향/효과에 대한 지표를 더욱 개발하고 정교화하기 위해 작업한다. 관련 서명인은 또한 허위/비인증 계정이 실제 이

72) [역주] 조회수, 팔로우 수 등 채널 통계를 인위적으로 부풀리는 도구를 구매하는 것

용자에게 미치는 영향과 영향을 추정하기 위한 추가 측정 기준을 개발하고 회원국 수준에서 보고한다(대상자에 대한 동향, 사용된 서술 등).

QRE 14.2.1: 관련 서명인은 보고서에 기재된 정책을 이행하기 위해 취한 조치와 확인/이용된 TTP의 범위에 대해 회원국 수준에서 보고한다.

SLI 14.2.1: 회원국 수준에서, 확인된 TTP의 인스턴스 수와 콘텐츠 유형과 같은 정보뿐만 아니라 각 TTP를 해결하는 정책에 따라 취해질 행위에 대한 조치의 수.

SLI 14.2.2: 회원국 수준에서, 조치를 취하기 전과 후에 확인된 각 TTP와 관련된 견해/노출(예: 좋아요, 공유, 의견).

SLI 14.2.3: 회원국 수준에서, [일단 사용 가능한] 가짜/허위 계정이 실제 사용자에게 미치는 영향과 영향을 추정하기 위한 측정 기준(대상 청중에 대한 동향, 사용된 서사 등 포함).

SLI 14.2.4: 회원국 수준에서, 관련 서명인 서비스에서의 TTP 관련 콘텐츠 뷰/노출 및 전체 콘텐츠에서 상호작용/개입, 뷰/노출, TTP 관련 콘텐츠의 비율에 대한 측정 [이러한 비율 추정이 계산 가능한 TTP의 경우].

조치 14.3. 관련 서명인은 상설 태스크포스를 통해 소집되어 악성 행위자가 사용하는 TTP의 목록과 용어를 합의하고 공표하며, 이는 매년 갱신해야 하고, 현재까지 서비스에서 허용되지 않는 조작적 행동과 관행에 대한 공통의 합의를 구성한다. 이를 바탕으로, 그들은 그러한 조작적 행동과 관행에 대응하기 위해 배치된 정책과 조치에 대한 공통의 기본 요소, 목표 및 기준점을 개발하는 것을 목표로 한다.

QRE 14.3.1: 서명인은 본 강령에 서명 후 6개월 이내에 상설 태스크포스에 합의된 TTP 목록에 대해 보고하며, 적어도 매년 이 목록을 업데이트한다. 그들은 또한 정책 및 조치에 대한 공통의 기준 요소, 목표 및 기준점에 대해 보고한다. 또한 투명성 센터에 이러한 정보를 업데이트한다.

AI 시스템에 대한 투명성 의무 TRANSPARENCY OBLIGATIONS FOR AI SYSTEMS

준칙 15. AI 시스템을 개발하거나 운영하고, AI로 생성되거나 조작된 콘텐츠(예: 딥페이

크)를 자사 서비스에서 유통하는 관련 서명인은 AI법에 제시된 금지된 조작 행위 목록을 및 투명성 의무를 고려할 것을 약속한다.

준칙 15를 충족하기 위해:

조치 15.1. 관련 서명인은 해당 콘텐츠를 사전에 포착하고 이용자를 경고하는 것과 같은 콘텐츠를 생성하고 조작하는 AI 시스템을 조작하는 금지 행위에 대응하기 위한 정책을 수립하고 확정한다.

QRE 15.1.1: 관련 서명인은 EU 및 현지 법에 따라 콘텐츠를 생성하거나 조작하는 AI 시스템을 조작하는 금지된 관행에 대응하기 위한 정책을 보고한다.

조치 15.2. 관련 서명인은 서비스에서 허용되지 않는 행위 및 내용의 포착, 조정 및 제재 등에 사용되는 알고리즘이 최종 이용자의 권리를 존중하며, 연합 및 회원국 법률에 따라 그들의 행동을 부당하게 왜곡하는 금지된 조작 행위에 해당하지 않도록 보장하는 정책을 수립, 확인한다.

QRE 15.2.1: 관련 서명인은 서비스에서 허용되지 않는 행위 및 내용의 포착, 조정 및 제재 등에 사용되는 알고리즘이 최종 이용자의 권리를 존중하며, 연합 및 회원국 법률에 따라 그들의 행동을 부당하게 왜곡하는 금지된 조작 행위에 해당하지 않도록 보장하는 정책 및 조치를 보고한다.

협력 및 투명성 COOPERATION AND TRANSPARENCY

준칙 16. 관련 서명인은 개인정보보호법을 완전히 준수하고 보안 및 인권 위협을 충분히 고려하여, 플랫폼 간 영향력 공작(influence operations), 각 서비스에서 이루어지고 있는 정보공간 내 해외 개입 및 관련 사건에 대한 정보를 적극적으로 공유하고 다른 서비스로의 확산 및 재발을 방지하기 위해, 관련 팀 간의 교류 채널을 운영한다.

준칙 16을 충족하기 위해:

조치 16.1. 관련 서명인은 플랫폼 간 정보 조작, 정보 공간 내 해외 간섭 및 개별 서비스에서 나타나는 사건에 대한 관련 정보를 가령 상설 태스크포스 내 지정 하위 그룹, 또는 이러한 정보를 교환하기 위한 기존 포럼을 통해 공유한다.

QRE 16.1.1: 관련 서명인은 정보 공유를 위해 사용하는 포럼과 이러한 공유를 통해 얻은 학습에 대한 정보를 공개한다.

SLI 16.1.1: 서명인 간의 협력 및 정보 공유의 결과로 취한 조치의 수. 그러한 정보가 있을 때, 영향을 받은 회원국(이 협력으로 인해 포착, 조치된 내용에 대한 정보 포함)을 명시한다.

조치 16.2. 관련 서명인은 조정 정책을 회피하고, 다양한 수용자를 개입시키거나, 덜 정밀한 정책 범위에서 플랫폼에서 협력하는 방법과 같은 다양한 플랫폼에 걸친 잘못된 정보, 허위조작정보, 정보 행위자의 전술적 이동에 대한 정보를 공유하고, 특별한 관심을 기울인다.

QRE 16.2.1: 관련 서명인은 그들 간의 협력과 정보 공유의 결과로, 그들의 중재팀 및/또는 이러한 감시에 개입하는 학계, 또는 팩트체크 기관의 외부 파트너에 의해 관찰된 그들의 플랫폼 내 이러한 행위자들에 의해 광고되고 사용되는 이동 전술의 질적 사례와 사례 연구를 공유한다.

V. 이용자 역량강화 EMPOWERING USERS

다음과 같은 사실로:

- (a) 서명인은 이용자가 허위 및/또는 오해의 소지가 있는 내용을 감지하고 보고할 수 있는 권한을 부여하는 것이 허위조작정보의 영향을 제한하는 핵심 도구로서 중요하다는 것을 인지한다. 이러한 맥락에서 서명인은 2021년 유럽위원회 지침⁷³⁾에

73) European Commission Guidance on Strengthening the Code of Practice on Disinformation, COM(2021) 262 final, 26 May 2021, Section 6.1, page 12.

언급된 바와 같이 신뢰할 수 있는 내용의 발견 가능성을 지속적으로 개선하고 서비스의 안전한 설계를 강화하며 허위 정보를 식별하고 이러한 유형의 내용을 감지하고 보고할 수 있는 도구로 사용자에게 권한을 부여함으로써 허위조작정보의 가시성과 침투를 희석시키는 것의 중요성을 인지한다.

- (b) 서명인은 이용자들이 콘텐츠의 진위성을 판단하는 데 도움을 줄 수 있도록 콘텐츠의 진위성 및 진위성을 조사할 수 있는 도구를 이용자에게 권한으로 부여하는 입증 기술의 잠재력의 중요성을 인지한다.
- (c) 서명인은 취약계층을 보호하고 권한을 부여하는 것을 포함하여 미디어 리터러시 분야에서 그들의 노력을 강화하는 것의 중요성을 인지한다.
- (d) 서명인은 또한 D의 구전 유포 위협을 완화하는 것을 목적으로 서비스의 안전한 설계와 구조를 위한 조치를 강화하는 것이 중요하다는 것을 인지한다.
- (e) 서명인은 추천 시스템이 사용자의 정보 환경에 미치는 중요한 영향을 인정하며, 따라서 추천 시스템은 투명해야 하며 사용자에게 정보가 권유되는 방식에 있어 선호하는 선택안을 언제든지 수정할 수 있는 가능성을 제공해야 함을 인지한다.
- (f) 서명인은 예를 들어 독립 팩트체크 기관의 사실 확인이나 다른 권위 있는 출처의 경고 표시를 통해 출처의 사실적 정확성 평가를 지원할 수 있는 도구에 대한 이용자 접근을 용이하게 하는 것이 허위조작정보 현상을 억제하는 데 중요함을 인지한다.
- (g) 이를 위해 서명인은 이후에 정책 위반으로 제거된 내용과 상호작용한 이용자를 대상으로 한 경고를 개발, 업데이트하는 것의 실현 가능성과 효과에 관한 연구를 중요하게 생각한다.
- (h) 개인 메시지 애플리케이션을 제공하는 서명인은 이용자가 그러한 서비스를 통해 전파되는 허위조작정보를 식별하고 표시할 수 있도록 지원하는 기술적 특징을 시험하고 구현하고, 해당 작업을 서비스에 통합하기 위해 팩트체커의 개인정보보호 준수 기회를 탐색하는 것이 중요하다는 것을 인지한다.
- (i) 유럽인권협약 제10조 및 연합기본권헌장 제7조, 제8조, 제11조, 제47조 및 제52조에 의거, 이용자에게 표시된 유해 허위 및/또는 오해 소지가 있는 내용에 대해 서명인이 취한 조치는 표현의 자유, 개인 통신에 대한 권리, 개인정보의 보호 및 효과

적인 구제에 대한 이용자의 권리를 존중해야 하며 불균형적이어서 안 된다.
따라서, 본 강령의 서명인들은 다음을 약속한다:

미디어 리터러시 향상 ENHANCING MEDIA LITERACY

준칙 17. 관련 서명인은 새로운 디지털 교육 실행 계획(Digital Education Action Plan)을 포함한 미디어 리터러시 분야에 대한 유럽위원회의 이니셔티브에 따라, 미디어 리터러시 및 비판적 사고 영역에 대한 노력을 지속하고 강화하며, 이러한 노력에 취약계층도 포함할 것을 약속한다.

준칙 17을 충족하기 위해:

조치 17.1. 관련 서명인은 예를 들어 서비스에서 볼 수 있는 콘텐츠에 대한 맥락이나 온라인 콘텐츠를 평가하는 방법에 대한 지침을 통해 이용자에게 권한을 부여함으로써 미디어 리터러시와 비판적 사고를 개선하기 위한 도구를 설계하고 구현하거나 계속 유지한다.

QRE 17.1.1: 관련 서명인은 이 준칙과 관련하여 개발하거나 유지하는 도구의 개요를 설명하고 각 회원국에서의 그들의 전개에 대해 보고한다.

SLI 17.1.1: 관련 서명인은 회원국 수준에서 도구에 대한 노출의 총 수, 도구와의 상호작용/개입에 대한 정보 등 측정 17.1에 대한 정성적 보고 요소에 명시된 도구의 영향을 평가하는 것과 관련된 측정항목에 대해 보고한다.

조치 17.2. 관련 서명인은 취약한 공동체의 참여를 고려하여 유럽연합 전역 일반 대중들 사이에서 악의적인 행위자들이 사용 중인 TTP뿐만 아니라, 허위조작 정보에 대한 인식을 높이기 위한 캠페인과 같은 미디어 리터러시 및 비판적 사고를 개선하기 위한 활동을 개발, 홍보 및/또는 지원하거나 지속 운영한다.

QRE 17.2.1: 관련 서명인은 그들이 시작, 또는 지원하는 활동과 그들이 목표로 하고, 도달하려는 회원국을 명시한다. 관련 서명인은 목표로 하는 회

원국별 캠페인을 이용자 기반으로 홍보하기 위한 행동에 대해 추가로 보고한다.

SLI 17.2.1: 관련 서명인은 회원국 수준에서 그들이 구축하거나 지원하는 캠페인의 효과(예: 해당 활동이 수행된 회원국 목록, 캠페인의 도달 범위, 이러한 활동이 생성된 참여, 온라인 자산과의 상호작용 수, 참여자 수)를 보여주기 위해 미디어 리터러시 및 인식 제고 활동의 수에 대해 보고하고 관련된 정량적 정보를 공유한다.

조치 17.3. 상기 두 조치 모두에 대해, 그리고 도구의 설계, 구현 및 영향 측정에 대한 미디어 리터러시 전문가의 전문성을 기반으로 하기 위해 관련 서명인은 위원회의 미디어 리터러시 전문가 그룹, ERGA의 미디어 리터러시 행동 그룹, EDMO, 그 국가별 지부 또는 관련 전문성을 가진 관련 회원국 대학 또는 단체를 포함한 EU의 미디어 리터러시 전문가들과 협력하거나 협의한다.
QRE 17.3.1: 관련 서명인은 본 준칙 내 모든 조치들의 목적을 위해 미디어 리터러시 전문가와 어떻게 참여하고 협력하는지 기술한다.

서비스 시스템의 ‘안전한 설계’ , 투명한 정책, 추천 시스템의 책임성

‘SAFE DESIGN’ OF THE ARCHITECTURE OF THE SERVICES, TRANSPARENT POLICIES, AND ACCOUNTABILITY OF RECOMMENDER SYSTEMS

준칙 18. 관련 서명인은 시스템, 정책, 그리고 기능을 개발할 때 안전한 설계 관행을 채택하여 허위조작정보의 바이럴 전파(viral propagation) 위험을 최소화한다.

준칙 18을 충족하기 위해:

조치 18.1. 관련 서명자는 다음과 같은 유해 허위조작정보의 바이럴 확산을 촉진하는 서비스의 위험을 완화하기 위한 조치를 취한다:

- 권위 있는 정보의 기준을 정의하기 위한 명확하고 투명한 방법과 접근 방식을 기반으로 권위 있는 정보의 중요성을 개선하고 허위조작정보의 중요성을 줄이기 위해 고안된 추천 시스템;

- 제품, 정책 또는 프로세스의 설계에 있어 사전 테스트와 같은 기타 시스템적 접근 방식들.

QRE 18.1.1: 관련 서명인은 조치 18.1에 따라 전개된 위험 완화 시스템, 도구, 절차 또는 기능, 그리고 각 EU 회원국들의 전개에 대해 보고한다.

QRE 18.1.2: 관련 서명인은 보고서 추천 시스템의 주요 매개변수를 게시하고, 운영이 완료되면 투명성 센터에 게시한다.

QRE 18.1.3: 관련 서명인은 제품, 정책 또는 프로세스를 설계하는 방법을 개괄적으로 설명하고, 추천 시스템 또는 다른 시스템 접근 방식을 통해, 그리고/또는 권위 있는 정보의 가시성을 높이기 위해 허위조작정보에 대한 노출과 개입을 줄이는 방법을 설명한다.

SLI 18.1.1: 관련 서명인은 제품, 정책, 프로세스(추천 시스템 포함) 또는 조치 18.1과 관련된 기타 시스템적 접근법의 성과를 충족시킬 수 있는 의미 있는 측정 기준을 통해 허위 정보의 확산, 견해 또는 노출의 감소 및/또는 권위적인 정보의 가시성 증가와 같은 그러한 조치의 효과에 대한 추정을 제공한다. 가능한 한 관련 서명인은 그러한 조치의 인과적 효과를 강조한다.

조치 18.2. 관련 서명인은 유해 허위 또는 오해의 소지가 있는 정보의 확산을 제한 (예: 서비스에 따라 유해 허위 또는 오해의 소지가 있는 정보를 금지, 하향 조정 또는 추천하지 않음)하기 위해 공개적으로 문서화된 비례성의 정책을 개발하고 시행하며, 이러한 정책을 지속적으로 위반하는 웹페이지 또는 행위자에 대해 조치를 취한다.

QRE 18.2.1: 관련 서명인은 조치 18.2와 관련된 정책 또는 서비스 조건과 이러한 정책의 지속적인 위반에 대한 접근 방식에 대해 보고한다.

SLI 18.2.1: 관련 서명인은 회원국 수준에서 조치 18.2와 관련된 정책 위반에 대한 대응 조치를 보고한다. 측정 지표는 다음을 포함해야 한다:

- 총 위반 건수
- 이러한 조치의 영향을 측정할 수 있는 의미 있는 측정 기준(예: 해당 조치가 가시성에 미치는 영향 또는 조치를 수행한 내용과의 관계)

조치 18.3. 관련 서명인은 유해 허위조작정보의 온라인 확산 및 관련 안전 설계 관행에 대한 연구 노력에 투자 및/또는 참여하고, 조사 결과를 대중에게 공개하

거나 본 강령의 태스크포스에 보고한다. 그들은 상설 태스크포스 내의 조사 결과를 공개하고 논의하며, 이러한 조사 결과를 기존의 안전 설계 관행 및 특징을 개선하거나 새로운 것을 개발하는 데 어떻게 사용할 것인지 설명한다.

QRE 18.3.1: 관련 서명인은 사내 및 제3의 기관과 협력하여 유해 허위조작 정보의 온라인 확산 및 관련 안전 설계 관행에 대한 연구 노력과 본 연구의 결과에 따른 조치 또는 변경 사항을 설명한다. 관련 서명인은 해당 연구에 대한 재정 투자에 대한 가능한 정보를 포함한다. 가능한 경우, 해당 연구 결과를 일반 대중이 이용할 수 있도록 한다.

준칙 19. 추천 시스템을 사용하는 관련 서명인은 정보의 우선순위 지정 또는 해제에 사용되는 주요 기준 및 매개변수를 이용자(recipient)에게 투명하게 공개하고, 이용자(user)에게 추천 시스템 옵션 및 해당 옵션을 이용할 수 있도록 정보를 제공해야 한다.

준칙 19를 충족하기 위해:

조치 19.1. 관련 서명인은 투명성 센터를 통해 그리고 약관에서 그들의 추천 시스템이 사용하는 주요 매개변수를 설명하는 정보를 명확하고 접근 가능하며 쉽게 이해할 수 있는 방법으로 사용자들에게 제공한다.

QRE 19.1.1: 관련 서명인은 특히 해당 추천 시스템이 채택하는 주요 매개변수를 설명하는 정보를 게시함으로써 EU 이용자들이 상기 조치들에 접근 가능한 공간에 정책과 조치의 세부 사항을 제공한다.

조치 19.2. 관련 서명인은 서비스 수용자가 관련 추천 시스템에 대해 선호하는 선택안을 언제든지 선택하고 수정할 수 있게 하며, 여기에는 이용자에게 이러한 선택안에 대한 투명성을 제공하는 것도 포함된다.

SLI 19.2.1: 관련 서명인은 보고 기간, 또는 대표 표집 기간 중 이용자들이 적극적으로 참여한 시간의 양 등과 같은 효과적인 이용자 설정에 대한 통

합된 정보를 제공하고, 배열 패턴의 변화를 명확히 표시한다.

준칙 20. 관련 서명인은 디지털 콘텐츠의 출처와 편집이력, 진위여부 또는 정확성을 평가할 수 있는 도구를 이용자에게 제공한다.

준칙 20을 충족하기 위해:

조치 20.1. 관련 서명인은 이용자들이 진위성을 확인하거나, 또는 새로운 도구나 프로토콜 또는 콘텐츠 입증을 위한 새로운 개방형 기술 표준(예: C2PA)과 같은 디지털 콘텐츠의 검증 또는 출처를 식별할 수 있도록 지원하는 기술 솔루션을 개발한다.

QRE 20.1.1: 관련 서명인은 검증 도구 또는 표준을 개발하는 과정, 수행 중 주요하고 표준화된 단계 및 진행 관련 장애물에 대한 세부 정보를 제공한다.

조치 20.2. 관련 서명인은 검증 도구 개발에 초점을 맞춘 글로벌 계획안 및 표준 도구(예: C2PA)에 참여/지지하기 위한 조치를 취한다.

QRE 20.2.1: 관련 서명인은 서명인이 가입한 검증 도구(예: C2PA) 개발에 초점을 맞춘 글로벌 계획안 및 표준 기관의 세부 정보, 또는 가능한 경우 기관 웹 사이트에 링크를 제공하는 관련 기관에 대한 지원을 제공한다.

이용자의 허위조작정보 식별 역량강화

BETTER EQUIPPING USERS TO IDENTIFY DISINFORMATION

준칙 21. 관련 서명인은 이용자들이 허위조작정보를 식별할 수 있는 더 나은 방안을 제공하기 위한 노력을 강화해야 한다. 특히, 이용자들이 충분한 정보를 바탕으로 서비스를 탐색할 수 있도록 관련 서명인은 그들의 서비스가 제공되는

모든 회원국의 언어로 잠재적 허위조작정보에 대한 팩트체크 기관의 팩트체크 및 기타 권위있는 출처의 경고를 표시함으로써, 이용자들이 출처의 사실적 정확성을 평가하는 도구에 쉽게 접근할 수 있도록 지원한다.

준칙 21을 충족하기 위해:

조치 21.1. 관련 서명인은 이용자가 독립적인 팩트체크 기관, 또는 공신력 있는 출처에서 제공하는 맥락과 통찰력을 활용할 수 있도록 회원국 및 EU 언어 전반에 걸쳐 가령 팩트체크 등급 표시 라벨, 등급이 매겨진 콘텐츠를 공유하려고 하거나 이전에 공유한 적 있는 이용자에 대한 알림, 정보 패널 등의 라벨을 사용하거나, 팩트체커로부터 정책 위반 통지를 받은 콘텐츠에 대한 조치를 취하는 등과 같은 정책, 기능, 또는 프로그램을 추가로 개발, 적용한다. 독립 팩트체크 기관과 협력해 서비스 콘텐츠에 라벨을 표시한 경우, 관련 서명인은 라벨 및 알림 등의 조치를 통해 이용자가 상호작용한 콘텐츠가 독립 팩트체크 기관의 평가를 받았음을 알리는 도구, 또는 기능을 추가로 개발, 적용하고 이를 모든 EU 회원국 언어로 구현하기 위해 노력한다.

QRE 21.1.1: 관련 서명인은 이 조치를 충족하기 위해 제시한 정책, 기능, 또는 프로그램, 그리고 회원국 전체에 대한 이용가능성에 대해 보고한다. 관련 서명인은 독립적인 팩트체커들과 협력하여 서비스 콘텐츠에 라벨을 붙이는 경우 다음에 대해 보고한다:

- 서비스 콘텐츠를 라벨을 붙이기 위해 협력하는 독립적인 팩트체커(팩트체크 기관이 보복이나 폭력에 대한 합리적인 두려움을 이유로 그러한 공개를 반대하지 않는 한), 이들의 사용 언어, 기반 정책 및 적용된 모든 라벨
- 이용자가 상호작용하는 콘텐츠가 독립적인 팩트체커에 의해 평가되었음을 이용자에게 알리는 데 사용할 수 있는 모든 도구 또는 기능.

SLI 21.1.1: 관련 서명인은 회원국 수준에서 조치 21.1에 따라 취한 조치에 대해 의미 있는 지표를 통해 보고해야 한다. 해당 정책, 기능 또는 프로그램에 따라 관련 정책에 의거해 취해진 조치; 라벨, 또는 팩트체크 및 기타 공신력 있는 출처의 도달 범위, 또는 기타 유사하게 관련된 지표에 대한 보

고들이 포함될 수 있다. 최소한 지표에는 팩트체크의 총 노출 수, 팩트체크 콘텐츠의 원래 노출 수 대비 팩트체크 노출 수 비율, 또는 이러한 지표가 서비스에 대한 팩트체크 실행과 관련이 없는 경우 동등하게 관련성이 있는 다른 지표와 해당 지표가 더 적절한 이유에 대한 설명이 포함된다.

SLI 21.1.2: 독립적인 팩트체크 기관과 협력해 서비스 콘텐츠에 라벨을 붙이는 경우, 관련 서명인은 회원국 수준에서 취한 조치와 그 영향에 대해 다음과 같은 지표로 보고한다:

- 독립 팩트체커가 게시한 기사의 수
- 해당 기사를 기준으로 콘텐츠에 적용된 라벨의 수
- 허위, 또는 오해의 소지가 있는 것으로 팩트체크된 콘텐츠에 대한 이용자 상호작용 또는 재공유에 미치는 영향 등 조치 21.1.1에 따라 취한 조치에 대한 의미있는 지표.

조치 21.2. 관련 서명인은 과학적 증거와 해당 서비스의 특수성, 사용자 개인정보 보호 선호도 등을 고려하여 이번 섹션에서 언급한 정책 위반으로 나중에 조치를 받은 콘텐츠와 상호작용한 이용자를 대상으로 경고, 또는 업데이트에 대한 연구 및 테스트를 수행 및/또는 지원한다. 관련 서명인은 관련 후속 조치를 파악하기 위해 상설 태스크포스 내의 조사 결과를 공개하고 논의한다.

QRE 21.2.1: 관련 서명인은 본 준칙의 일환으로 지원하고 수행한 연구 또는 테스트 노력과 본 준칙의 일환으로 수행한 연구 또는 테스트의 결과를 보고한다. 가능한 경우, 그 결과를 일반 대중에게 공개한다.

조치 21.3. 라벨링 및 경고 시스템을 사용하는 경우, 관련 서명인은 최신 과학적 근거에 따라 그리고 가령 이용자들이 이 시스템을 긍정적으로 인식할 수 있도록 하는 해당 개입의 영향과 유용성을 극대화하는 방법에 대한 이용자 요구 분석을 통해 해당 시스템을 설계해야 한다.

QRE 21.3.1: 관련 서명인은 라벨링 또는 경고 시스템을 개발, 배포 절차, 투명성 극대화를 위한 과학적 증거와 이용자의 요구를 고려하는 방법에 대해 보고한다.

준칙 22. 관련 서명인은 이용자들이 거짓이거나 오해의 소지가 있는 온라인 정보를 접할 때 충분한 정보를 바탕으로 의사결정을 내리는데 도움이 되는 도구를 제공한다. 또 이용자들이 사회적 이슈나 일반적 관심사에 대한 논쟁과 관련해 충분한 정보를 바탕으로 온라인을 탐색하기 위한 신뢰도 지표 등 정보 출처의 신뢰성을 평가할 수 있는 도구 및 정보에 대한 이용자의 접근을 용이하게 해야 한다.

준칙 22를 충족하기 위해:

조치 22.1. 관련 서명인은 팩트체커 및 기타 유관기관 뿐만 아니라 언론인 협회 및 언론자유 단체를 포함한 뉴스 미디어와 협력해 독립적인 제3자가 개발한 신뢰도 지표(예: 출처의 무결성 및 해당 지표의 방법론에 중점을 둔 신뢰도 표시 등)에 접근하여 이용자가 정보에 입각한 선택을 할 수 있도록 지원한다.

QRE 22.1.1: 관련 서명인은 서비스 이용자가 해당 지표나 신뢰 표시로부터 이익을 얻을 수 있도록 하는 방법에 대해 제시한다.

SLI 22.1.1: 관련 서명인은 회원국 수준에서 신뢰도 지표를 활성화한 사용자 비율을 보고한다.

조치 22.2. 관련 서명인은 이용자에게 추천 시스템에 대한 미디어의 신뢰도와 관련된 신호를 보유하거나 그러한 신호를 자신의 추천인 시스템에 공급하는 선택권을 제공한다.

QRE 22.2.1: 관련 서명인은 이용자에게 추천 시스템에 미디어 소스의 신뢰도와 관련한 신호를 입력할 수 있는 선택권을 주거나 그러한 신호를 추천 시스템에 제공할 수 있다.

조치 22.3. 관련 서명인은 상기 조치를 이행하기 위해 시행한 정책 및 조치에 대한 세부 정보를 EU 이용자에게 공개함으로써 해당 추천 시스템에서 사용하는

주요 매개변수를 제시한다. 해당 정보는 투명성 센터에도 포함되어야 한다.

QRE 22.3.1: 관련 서명인은 특히 추천 시스템에 사용하는 매개변수를 제시함으로써, EU 이용자들이 접근할 수 있는 상기 조치를 이행하기 위한 정책 및 조치에 대한 세부 정보를 제공한다. 이 정보는 투명성 센터에도 포함되어야 한다.

조치 22.4. 신뢰도 지표를 제공하는 관련 서명인은 정보 출처에 대한 투명하고, 정치적으로 편견이 없으며, 독립적인 방식으로의 검토, 모든 출처에 대해 완전히 공개된 기준을 동일하게 적용, 그리고 독립적인 규제 당국, 또는 기타 권한이 있는 기관의 독립적인 감사에 대한 허용을 보장한다.

QRE 22.4.1: 관련 서명인은 정보 출처 검토에 사용하는 기본 기준에 대한 세부 정보를 제공하고, 서비스가 정치적으로 편견이 없으며, 독립적임을 보장하기 위해 마련한 관련 안전장치를 공개한다. 또한 해당 기준이 다양한 퍼블리셔의 대표적인 범위에 어떻게 동일하게 적용되는 지에 대한 예시를 제공한다. 각 분석에는 출처 검토에 기여한 사람, 또는 인증기관이 명시된다.

SLI 22.4.1: 이 자료에 접근할 수 있는 관련 서명인은 신뢰도 지표 덕분에 생성된 신뢰할 수 있는 출처의 트래픽량을 측정할 수치를 보고한다.

조치 22.5. 신뢰도 지표를 제공하는 관련 서명인은 규정 준수 및 수정 메커니즘을 제공하고 지표가 적용되기 전에 평가 과정에 참여하고 평가가 발표된 후 소비자에게 답변을 제공하는 등 게시자의 의견을 들을 권리를 존중한다.

QRE 22.5.1: 관련 서명인은 업데이트 또는 실수가 발생하는 경우 등급 또는 지표에 대한 정정 사항을 정기적으로 게시한다. 관련 서명인은 신뢰도 지표에 기록된 해당 참여 증거를 포함하여 출판사와의 교류 사례를 제공하고, 관행 개선 등 퍼블리셔의 관행 변경 사항을 반영하기 위해 분석 내용을 정기적으로 업데이트한다.

SLI 22.5.1: 관련 서명인은 등급 또는 현행화된 등급이 발표되기 전에 퍼블리셔가 의견 진술권을 행사한 후 신뢰할 수 없는 등급이 신뢰할 수 있는 등급으로 변경된 회원국별 전체 사례 수를 보고한다.

SLI 22.5.2: 관련 서명인은 공개된 기준에 따라 평가받은 후, 저널리즘적 관

행을 개선한 퍼블리셔의 수와 그에 따라 신뢰도 점수가 개선된 퍼블리셔의 수를 정기적으로 보고한다.

조치 22.6. CWA17493:2019와 같이 유럽법(‘기술 표준’)⁷⁴에 의해 정의되는 자발적, 자율 규제, 인증 가능한 유럽 표준 또는 유럽 표준화 성과물을 통해 신뢰도 지표를 제공하는 서명인은 다음과 같은 일을 수행한다:

- 해외에서 통용되는 모범 사례와 윤리 규범을 바탕으로 개발 및 개선한다;
- 비독점적이고 중립적인 방식으로 공개적으로 이용 및 접근하도록 한다;
- 유럽 인증 및 EU 규정(EC) No. 765/2008에 따라 해당 이행을 관리한다.

QRE 22.6.1: 관련 서명인은 불만사항 관리를 포함한 참여 및 적합성 평가에 대한 관련 통계 및 분석뿐만 아니라 적합성 준수 기관의 자체 평가 및 인증 결과를 게시한다.

SLI 22.6.1: 관련 서명인은 불만사항 관리를 포함한 참여 및 적합성 평가에 대한 관련 통계 및 분석을 발표한다.

조치 22.7. 관련 서명자는 특정 공공 및 사회적 관심 주제 또는 위기 상황에서 이용자를 공신력있는 출처로 안내하는 제품 및 기능(예: 정보 패널, 배너, 팝업, 지도 및 프롬프트, 신뢰도 지표)을 고안 및 적용한다.

QRE 22.7.1: 관련 서명인은 서비스 전반에 걸쳐 배포된 제품 및 기능을 설명하고 회원국 전체에서 해당 기능을 사용 가능한지를 명시한다.

SLI 22.7.1: 관련 서명인은 노출 및 상호작용(클릭, 클릭률(해당 도구 및 서비스와 관련된))과 공유(해당 도구 및 서비스와 관련된)의 지표를 통해 회원국 수준에서 제품 또는 기능에 대한 도달 및/또는 사용자 상호작용에 대해 보고한다.

허위 그리고/또는 오해의 소지가 있는 유해정보를 표시하는 기능

FUNCTIONALITY TO FLAG HARMFUL FALSE AND/OR MISLEADING INFORMATION

준칙 23. 관련 서명인은 이용자에게 서명인의 정책 또는 서비스 약관을 위반한 허위

74) 가령 the Journalism Trust Initiative CWA17493:2019

그리고/또는 오해의 소지가 있는 유해 정보를 표시할 수 있는 기능을 제공한다.

준칙 23을 충족하기 위해:

조치 23.1. 관련 서명인은 서명인의 정책 또는 서비스 약관을 위반한 유해 허위 및/또는 오해의 소지가 있는 정보를 이용자가 신고할 수 있는 이용자 친화적인 기능을 모든 서비스에, 그리고 해당 서비스가 제공되는 모든 회원국의 언어로 개발, 지속적으로 제공해야 한다. 해당 기능은 표현의 자유를 충분히 고려해서 적절하고 비례적이며 일관된 후속조치로 이어져야 한다.

QRE 23.1.1: 관련 서명인은 EU 회원국 전체의 유해 허위 및/또는 오해의 소지가 있는 정보와 관련된 정책에 대한 표기 시스템에 대한 이용 가능성에 대해 보고하고 시스템을 작동시키는 데 필요한 다양한 단계를 제시한다.

조치 23.2. 관련 서명인은 이 기능이 인간 또는 기계 기반의 남용(예: 다른 목소리를 침묵시키기 위한 '대량 플래그' 전술)으로부터 정당하게 보호될 수 있도록 필요한 조치를 취한다.

QRE 23.2.1: 관련 서명인은 신고 및 이의제기 시스템의 무결성 보장을 위해 취한 일반적 조치를 보고하는 한편, 잠재적인 악용자가 방어체계의 취약점을 찾아 이를 악용하는데 도움이 될 정보를 공개하지 않도록 유의한다.

투명한 이의제기 메커니즘 TRANSPARENT APPEAL MECHANISM

준칙 24. 관련 서명인은 본 장(이용자 역량강화)과 관련된 정책 위반(조치 18.2에 명기된 대로)을 근거로 취해진 집행 조치(콘텐츠/계정에 레이블 표기, 강등 또는 기타 집행 등)의 대상이 된 콘텐츠 또는 계정을 보유한 이용자에게 해당 사실을 알리고, 문제가 된 집행 조치에 대해 이의를 제기할 수 있는 기회를 제공하며, 불만 사항을 적시에 성실하고 투명하며 객관적인 방식으로 처리하고, 불만 사항이 타당하다고 판단되는 경우 지체없이 조치를 취소할 수 있도록

노력한다.

준칙 24를 충족하기 위해:

조치 24.1. 관련 서명인은 이용자에게 특정 내용이나 계정이 이 조항과 관련된 정책 위반에 근거하여 라벨링, 강등 또는 다른 방법으로 집행된 이유와 그러한 집행 조치의 근거 및 투명한 메커니즘을 통해 이의를 제기할 수 있는 가능성에 대한 정보를 제공할 것을 약속한다.

QRE 24.1.1: 관련 서명인은 회원국 및 언어에 걸쳐 통지 및 이의제기 이용 가능성에 대해 보고하고, 이의제기 절차에 대한 상세한 정보를 제공한다.

SLI 24.1.1: 관련 서명인은 조치 18.2에 의거 명시된 정책에 대한 실행 조치의 수와 특징, 이의제기 이후의 조치의 수, 이의제기의 결과, 그리고 가능한 범위 내에서 이의제기 절차의 처리기간, 또는 효과에 대한 통찰력을 제공할 수 있는 지표들에 대한 정보를 제공하고, 해당 정보를 투명성 센터에 공개한다.

메시징 앱에서 허위조작정보를 차단하기 위한 조치

FUNCTIONALITY TO FLAG HARMFUL FALSE AND/OR MISLEADING INFORMATION

준칙 25. 개인 간(private) 메시징 서비스 이용자가 해당 서비스를 통해 전파된 잠재적 허위조작정보를 식별할 수 있도록 돕기 위해 메시징 앱을 제공하는 관련 서명인은, 이용자들이 수신하는 정보에 대해 비판적으로 사고할 수 있도록 역량을 강화하고 그러한 정보가 정확한지 판단할 수 있도록 돕는 기능 또는 이니셔티브를 지속적으로 구축하고 시행하며, 이는 이용자들의 개인정보를 충분히 보호하고 암호화를 약화시키지 않는 방식으로 이루어져야 한다.

준칙 25를 충족하기 위해:

조치 25.1. 관련 서명인은 암호화 약화 없이 권위 있는 정보에 대한 이용자의 접근을

용이하게 하는 기능을 설계 및 구현하거나 제3의 파트너와 협력하여 설계 및 구현할 것이며, 프라이버시의 보호와 관련하여 제3의 파트너는 해당되는 경우 정부, 팩트체커 및 기타 시민 사회 조직과 같은 시민 단체를 포함할 수 있다. 여기에는 예를 들어 (각 서비스에 적합한) 디지털 리터러시 캠페인 및 프로그램, 또는 팩트체커 또는 정부가 이용자에게 적시에 정확한 정보를 제공하기 위해 관련 메시징 서비스를 용이하게 사용할 수 있는 제3의 기능이 포함될 수 있다.

QRE 25.1.1: 관련 서명인은 이 조치를 충족하기 위해 시행 중인 도구, 정책, 파트너십, 프로그램 및 캠페인에 대해 보고하고, 가능한 경우, 시민 단체와 그들의 결과에 대한 관련 세부정보를 포함해 회원국 전체에 대한 이용가능성에 대해 보고한다.

SLI 25.1.1: 현지 법을 준수하면서 제3자가 제공하는 필요 정보에 따라, 관련 서명인은 가능한 범위 내에서 일부 도구 사용에 대해 보고한다(예: 이용자가 팩트체커에게 제출한 클레임의 수 또는 플랫폼에 제출된 클레임에서 생성된 팩트체크의 도달 범위).

조치 25.2. 관련 서명인은 암호화를 약화시키지 않고 개인정보보호를 충분히 고려하면서 이용자가 허위조작정보를 식별하고 서비스 내 바이럴 확산을 제한하는 데 도움이 되는 도구와 기능을 개발한다. 여기에는 예를 들어, 추가 마찰을 통해 여러 대화에서 정보 전달을 제한하는 기능이나 소셜 미디어 콘텐츠가 메시징 앱을 통해 전파될 때 (가능한 경우) 눈에 띄는 적절한 팩트체크 라벨을 표시하는 기능이 포함될 수 있다.

QRE 25.2.1: 관련 서명인은 서비스에 대한 허위조작정보의 바이럴 확산을 제한하고, 이용자가 수신받은 메시지에 대해 생각할 수 있는 권한을 부여하기 위해 사용가능한 도구 및 기능에 대해 보고한다.

SLI 25.2.1: 암호화의 약화 없이 개인정보보호를 충분히 고려하면서, 가능한 범위 내에서 현지법을 준수할 때, 관련 서명인은 조치 25.1 및 25.2를 충족하기 위해 배포된 도구, 기능 및 캠페인의 사용 및 영향에 대한 투명한 추정 또는 근거 기반 연구를 제공한다.

VI. 연구 공동체 역량강화 EMPOWERING THE RESEARCH COMMUNITY

다음과 같은 사실로:

- (a) 서명인은 허위조작정보에 대처하기 위한 효과적인 전략의 일환으로 연구 공동체의 플랫폼 데이터에 대한 접근성과 이들의 활동에 대한 적절한 지원을 위한 프레임워크 구축의 중요성을 인지한다.
- (b) 연구 제안이 관련 분야의 윤리적 및 방법론적 모범 사례(예: 플랫폼 데이터 접근에 관한 행동강령 관련 EDMO 제안서에 명시된 바와 같이)에 부합하는 경우, 본 강령의 목적에 부합하는 것으로 간주한다.
- (c) 서명인은 연구자 공동체에 회원국이 인정한 공익적 사명에 따라 비영리 목적으로 과학 연구를 수행하는 것을 주요 목표로 하는 시민 사회 단체가 포함될 수 있음을 인지한다.
- (d) 유럽인권협약 제7조, 유럽연합 기본권 헌장 제8조 제1항, 유럽연합 기능에 관한 조약(TFEU) 제16조 제1항에 의거 서명인은 모든 데이터에 대한 접근 조건이 개인 통신의 권리를 존중하고 모든 관련 당사자의 권리와 합법적인 이익을 적절히 보호해야 함을 인지한다.
- (e) 본 강령의 목적에 따른 데이터 접근 권한이 정부 기관 및 법 집행 기관으로 확대하지 않는다.
- (f) 일반 원칙으로서, 관련 서명인은 자신의 플랫폼에서 허위조작정보에 대한 선의의 연구를 금지하거나 억제하지 않는 것의 중요성과, 허위조작정보에 대한 선의의 연구를 수행하거나 참여하는 서비스 이용자의 계정에 대해 적대적인 조치를 취하지 않을 것을 인지한다.

따라서, 본 강령의 서명인들은 다음을 약속한다:

허위조작정보 연구를 위한 서명인의 데이터 공개 및 접근

DISCLOSURE OF AND ACCESS TO SIGNATORIES' DATA FOR RESEARCH ON DISINFORMATION

개인정보가 없는 데이터 및 익명화된 데이터, 집계된 데이터 또는 명백히 공개된 공공데이터에 대한 자동화된 접근⁷⁵⁾

Automated access to non-personal data and anonymised, aggregated or manifestly made public data

준칙 26. 관련 서명인은 안전하고 실행 가능한 경우, API, 또는 기타 접근가능한 기술 솔루션과 같은 데이터를 분석할 수 있는 자동화된 수단을 통한 허위조작정보에 관한 연구 목적을 위해 익명으로 집계된, 그리고 공공 데이터로 명시한 데이터에 대한 지속적이고 실시간, 또는 실시간에 가까운 검색 가능한 안정적인 접근을 제공한다.

준칙 26을 충족하기 위해:

조치 26.1. 관련 서명인은 남용 위험을 해결하기 위한 합리적인 보호 조치(예: 악의적이거나 상업적인 사용을 금지하는 API 정책)와 함께 운영 서비스 내 콘텐츠의 참여 및 노출(보기)와 같은 서비스의 허위조작정보에 대한 연구 수행과 관련한 개인적이지 않고, 익명으로 집계된, 또는 공공데이터로 명시한 데이터에 대한 공개적인 접근을 제공한다.

QRE 26.1.1: 관련 서명인은 개인 정보 보호에 관한 연구를 수행하는 것과 관련된 익명으로 집계된, 그리고 공공데이터로 명시한 데이터에 대한 공개적인 접근을 제공하기 위한 도구와 절차 및 남용 위험 해결을 위한 보호 조치를 형성한다.

QRE 26.1.2: 관련 서명인은 해당 도움말 센터에서 이러한 기준점에 대한 접근에 활용될 기술적 프로토콜에 관한 세부 사항뿐만 아니라 조치 25.1을 통해 이용 가능한 기준점 관련 정보를 공개한다. 해당 정보는 투명성 센터에서도 제공 가능해야 한다. 이 정보에는 최소한 이용 가능한 기준점의 정의, 해당 정보가 형성된 방법에 대한 기술적이고 방법론적인 정보, 그리고 데이터의 대표성에 대한 정보가 포함된다.

75) “명백하게 공개된 공공데이터”의 정확한 범위는 서비스마다 다를 수밖에 없다. 관련 서명인은 보고서를 통해 이 용어가 자신들의 서비스에 어떻게 적용되는지 설명한다.

SLI 26.1.1: 관련 서명인은 이용자 수와 같이 조치 26.1에서 설명한 도구와 절차의 활용에 관한 정량적 정보를 제공한다.

조치 26.2. 관련 서명인은 활용 과정이 까다롭지 않은 선출직 공직자, 언론사 및 정부 계정과 같이 공공 계정에 속하는 연구 목적을 위해 그들의 서비스에서 개인적이지 않고, 익명으로 집계된, 그리고 공공데이터로 명시한 데이터에 대한 실시간, 또는 거의 실시간에 가깝게 기계로 읽을 수 있는 접근을 제공한다.

QRE 26.2.1: 관련 서명인은 조치 26.2에서 설명한 바와 같이 연구 목적으로 익명으로 집계된, 그리고 공공데이터로 명시한 데이터에 실시간, 또는 거의 실시간에 가깝게 접근할 수 있는 도구와 절차를 형성한다.

QRE 26.2.2: 관련 서명인은 그들의 서비스에서 활용 가능한 공공 데이터로 명시된 데이터의 범위를 제시한다.

QRE 26.2.3: 관련 서명인은 조치 26.2에서 설명한 개인적이지 않은 데이터 및 익명으로 집계된, 그리고 공공데이터로 명시한 데이터에 대한 접근을 위한 활용 절차를 제시한다.

SLI 26.2.1: 관련 서명인은 조치 26.2의 도구와 절차의 활용, 신속성 및 수용 수준에 대한 의미 있는 측정 기준을 다음과 같이 제공한다:

- 월간 이용자 수(또는 표본 대표 기간 중의 이용자)
- 접수, 거부 및 수락된 신청 수(보고 기간 또는 표본 대표 기간 중)
- 평균 응답 시간(보고 기간 또는 표본 대표 기간 중).

조치 26.3. 관련 서명인은 적절한 시기에 접근 시스템의 오작동을 보고하고, 결함이 있는 기능을 복구하고 수리하기 위한 절차를 이행한다.

QRE 26.3.1: 관련 서명인은 조치 26.3을 준수하기 위해 시행 중인 보고 절차를 제시하고, 오작동 대응 절차에 대한 정보뿐만 아니라 보고 기간 중 시스템 이용을 방해했을 수 있는 오작동 및 수정 소요 시간에 대한 정보를 제공한다.

추가적인 정밀검토가 필요한 연구 목적의 데이터 접근을 위한 거버넌스 구조

Governance structure for access to data for research purposes requiring additional scrutiny

준칙 27. 관련 서명인은 연구자와 연구 제안을 심사할 수 있는 독립적인 제3의 기관을 성장시키고, 자금을 지원하고, 협력함으로써 심사를 통과한 연구자가 허위조작정보에 대한 연구를 수행하는데 필요한 데이터에 접근할 수 있도록 한다.⁷⁶⁾

준칙 27을 충족하기 위해:

조치 27.1. 관련 서명인은 플랫폼 데이터 접근에 관한 행동 강령에 대한 EDMO의 제안과 같은 적절하고 지속적인 노력을 고려하면서, 준칙 27에 언급된 독립적인 제3의 기관을 합리적인 일정 내에서 성장시킬 것을 약속한다.

QRE 27.1.1: 관련 서명인은 그들의 파트너십, 프로그램 또는 연구자가 참여하는 다른 형태의 참여와 관련하여 조치27.1에 요약된 절차, 절차의 세부적인 일정, 실질적 결과, 그리고 해당 절차의 영향력에 대한 그들의 개입을 기술한다.

조치 27.2. 관련 서명인은 준칙 27에 언급된 독립적인 제3의 기관의 성장을 위해 2022년부터 공동 기금을 조성하기로 약속한다.

QRE 27.2.1: 관련 서명인은 준칙 27에 언급된 독립적인 제3의 기관의 성장을 위한 자금을 공개한다.

조치 27.3. 관련 서명인은 관련 법률에 따라 준칙 27에 언급된 독립적인 제3의 기관이 설립되면, 해당 기구가 정의하는 프로토콜에 따라 검증된 연구자들의 허위조작정보에 대한 연구 수행에 필요한 개인 데이터를 공유할 수 있도록 협력할 것을 약속한다.

QRE 27.3.1: 관련 서명인은 일단 독립적인 제3의 기관이 설립되면 조치 27.3에 명시된 연구 목적의 데이터 공유를 가능하게 하기 위해 독립적인 제3의 기관과 어떻게 협력하는지에 대해 제시한다.

76) 서명인은 본 섹션의 목적에 따라 연구자를 심사하는 절차를 DSA 제31조에 따라 조정한다.

SLI 27.3.1: 관련 서명인은 독립적인 제3의 기관이 조사한 연구 프로젝트 중 몇 개가 협력을 시작했거나 요청한 데이터에 대한 접근을 제공했는지를 공개한다.

조치 27.4. 관련 서명인은 독립적인 제3의 기관이 완전히 설립되기를 기다리지 않고, 허위조작정보를 조사할 목적으로 검증된 연구자들과 데이터를 공유하는 방향의 파일럿 프로그램에 참여할 것을 약속한다. 이러한 파일럿 프로그램은 데이터의 공유/사용 관련 모든 해당 법률에 의거하여 운영된다. 파일럿들은 서명인의 서비스에서 삭제된 콘텐츠와 이 콘텐츠의 데이터 보존 기간에 대한 연구를 도모하는 방안을 모색할 수 있다.

QRE 27.4.1: 관련 서명인은 허위조작정보 조사를 목적으로 검증된 연구자들과 데이터를 공유하기 위해 그들이 참여한 파일럿 프로그램을 제시한다. 여기에는 프로그램의 성격, 참여 연구팀 수, 가능한 경우 연구 주제 또는 결과에 대한 정보가 포함된다.

연구진과의 협력 COOPERATION WITH RESEARCHERS

준칙 28. 관련 서명인은 자사의 서비스와 관련된 허위조작정보에 대한 선의의 연구를 지원한다.

준칙 28를 충족하기 위해:

조치 28.1. 관련 서명인은 연구를 용이하게 수행하기 위해 적절한 인적 자원을 확보하고 연구에 필요할 가능성이 있는 데이터 유형을 추적하고, 연구자들이 해당 기관의 적절한 접점을 찾을 수 있도록 하기 위해 연구자와의 열린 대화를 형성, 유지해야 한다.

QRE 28.1.1: 관련 서명인은 전담팀, 도구, 지원 센터, 프로그램 또는 이벤트를 포함하여 연구를 도모하고 연구자 공동체에 참여하기 위해 배치된 자원과 프로세스를 제시한다.

조치 28.2. 관련 서명인은 유럽 전역의 연구자들이 현재 이용할 수 있는 데이터 유형에 대해 투명하게 공개한다.

QRE 28.2.1: 관련 서명인은 유럽 연구자들이 API 또는 전담팀, 도구, 도움말 센터, 프로그램 또는 이벤트를 통해 현재 접근할 수 있는 데이터 유형을 제시한다.

조치 28.3. 관련 서명인은 플랫폼에서 허위조작정보에 대한 진실로 명백하게 공익적이고 선의로 수행된 연구를 금지, 단념시키지 않으며, 허위조작정보에 대한 선의의 연구를 수행, 참여하는 연구자 이용자, 또는 계정에 대한 적대적인 조치를 취하지 않는다.

QRE 28.3.1: 관련 서명인은 EDMO와 협력하여 유럽 연구자들의 연례 협의를 운영하여 적대적 행동을 경험했는지 또는 그렇지 않으면 그러한 연구를 운영하는 것이 금지되거나 권장되지 않는지에 대해 평가한다.

조치 28.4. 서명인과 유럽 연구자 공동체간 협력 프레임워크의 일부로서, 관련 서명인은 연구자가 독립적으로 관리하고 과학적 가치에 기초해 과학적 우선순위 및 투명한 배분 절차를 정의할 수 있도록 하는 허위조작정보 연구를 위한 재원을 EDMO와 함께 마련한다.

QRE 28.4.1: 관련 서명인은 자원들이 독립적으로 관리되도록 하기 위한 절차와 조치 28.4의 목적을 위해 제공되는 자원을 공개한다.

연구조직의 투명성과 데이터 공유

TRANSPARENCY AND DATA SHARING FROM RESEARCH ORGANISATIONS

준칙 29. 관련 서명인(본 준칙에 서명하는 연구조직)은 투명한 방법론과 윤리적인 기준에 따라 연구를 수행하고, 데이터셋, 연구결과 및 방법론을 관련자와 공유한다.⁷⁷⁾

준칙 29를 충족하기 위해:

조치 29.1. 관련 서명인은 허위조작정보의 확산과 영향력 작동을 추적하고 분석하는 연구 활동을 수행하는데 있어 투명한 방법론과 윤리적 기준을 활용한다. 이들은 EDMO, ERGA 및 기타 서명인들을 포함한 태스크포스 구성원 및 궁극적으로는 일반 대중과 함께 데이터셋, 연구 결과 및 방법론을 공유할 것이다.

QRE 29.1.1: 관련 서명인은 주제, 방법론, 윤리 기준, 접근된 데이터 유형, 데이터 거버넌스 및 결과를 포함한 연구에 대한 보고서를 제공한다.

QRE 29.1.2: 관련 서명인은 투명성 센터에서 자신들의 연구 결과를 업데이트하여 해당 작업에 대한 더 많은 인식과 이용 가능성을 허용한다.

QRE 29.1.3: 관련 서명인은 연구 결과에 대해 정보를 제공받는 모든 이해관계자들에게 방법론 개발에 대한 상세한 정보를 제공한다. 또한 ERGA, EDMO 및 기타 서명자들을 포함한 태스크포스 모든 구성원들에게 자신들이 수행하는 연구 활동 및 가능한 경우 사용된 관련 방법론에 대해 정기적으로 알린다. 최종적으로 그러한 연구 결과 및 관련 방법론을 가능한 한 대중과 공유한다.

SLI 29.1.1: 관련 서명인은 연구 프로젝트의 결과에 대해 정보를 제공받은 이해관계자 또는 시민의 범위에 대해 보고한다.

조치 29.2. 관련 서명인은 향후 규제 및 운영 개입(예: 라벨, 경고, 사후 통지)을 알리기 위해 본 강령 및 기타 지역에 구현된 다양한 복원력 강화 조치의 상대적 효과를 확인하는 것을 목표로 하는 연구 활동을 수행한다.

QRE 29.2.1: 관련 서명인은 주제, 방법론, 윤리 기준, 접근된 데이터 유형, 데이터 거버넌스 및 결과를 포함한 연구 관련 보고서를 제공한다.

QRE 29.2.2: 관련 서명인은 자신들의 작업에 대한 더 많은 인식과 이용 가능성을 허용하기 위해 투명성 센터 내 자신들의 연구 결과를 업데이트한다.

QRE 29.2.3: 관련 서명인은 연구 결과에 대해 정보를 제공받는 모든 이해관계자들에게 방법론 개선에 대한 상세한 정보를 제공한다. 또한 ERGA,

77) 이는 본 준칙에 서명한 연구기관의 연구 결과를 참조한 것이며, 플랫폼이 해당 기관과 공유하는 데이터는 아니다(본 강령, 또는 다른 방식에 의거하여).

EDMO 및 기타 서명자들을 포함한 태스크포스의 모든 구성원들에게 자신들이 수행한 연구 활동 및 가능한 경우 사용된 방법론에 대해 정기적으로 알린다. 최종적으로 그러한 연구 결과 및 관련 방법론을 가능한 한 대중과 공유한다.

SLI 29.2.1: 관련 서명인은 연구 과제 결과에 대해 정보를 제공받은 이해관계자 또는 시민의 범위에 대해 보고한다.

조치 29.3. 관련 서명인은 허위조작정보가 포함된 광고에 관해 투명성을 향상시키는 것을 목표로 하는 효과적인 광고 저장소 모델을 구현한다.

QRE 29.3.1: 관련 서명인은 그들의 업무의 투명성을 강화하는 것을 목표로 효과적인 광고 저장소의 모델에 관한 더 많은 인식과 가용성을 허용하기 위해 투명성 센터 내 그들의 연구를 업데이트한다.

SLI 29.3.1: 관련 서명인은 연구 과제의 성과, 연구 활동, 방법론 및 결과에 대해 정보를 제공받는 이해관계자 또는 시민의 범위에 대해 보고한다.

VII. 팩트체킹 공동체 역량강화 EMPOWERING THE FACT-CHECKING COMMUNITY

다음과 같은 사실로:

- (a) 본 강령의 서명인은 허위조작정보에 대응하기 위한 효과적인 전략의 일환으로 플랫폼 데이터에 대한 팩트체킹 공동체의 활발한 접근과 그들의 활동에 대한 적절한 지원을 위한 프레임워크 형성의 중요성을 인지한다.
- (b) 서명인은 특히 아직 사실확인이 이뤄지지 않은 회원국과 언어에 특히 주목하여 그들의 서비스에서 팩트체킹의 지속적인 적용을 보장하는 명확한 목표와 일정을 가진 구체적인 단계를 설정하는 것의 중요성을 인식한다.
- (c) 이를 위해 서명인은 계획에 따른 광고 시스템과 비디오 콘텐츠를 포함하여 그들의 플랫폼 서비스에서 팩트체커들의 활동을 지속적으로 이용하고 통합하는 것이 중요함을 강조한다.
- (d) 서명인은 팩트체커들이 사실이 확인된 내용과 사실 확인과 관련하여 취한 조치에 대한 정보, 그리고 기타 팩트체킹의 품질과 영향을 극대화하는 데 도움이 되는

관련 정보에 자동으로 접근할 수 있도록 하는 것의 중요성을 강조한다.

- (e) 서명인은 팩트체커가 허위조작정보 대응에 효과적으로 대처하기 위해 팩트체킹 조직이 당파적 기관으로부터 독립적이어야 하며, 재정, 조직, 그리고 방법론에 있어 투명해야 할 뿐만 아니라, 국제 팩트체킹 네트워크 원칙(IFCN)의 검증된 서명인, 유럽 디지털 미디어 관측소(EDMO)의 팩트체커 네트워크, 또는 독립적인 유럽 팩트체킹 조직을 위한 전문 진실성에 관한 미래 강령의 구성원으로서 일관되고 지속적인 팩트체킹에 전념해야 함을 인지한다.
- (f) 유럽인권협약 제7조, 유럽연합 기본권헌장 제8조 제1항 및 유럽연합 기능조약 (TFEU) 제16조 제1항에 의거, 서명인은 모든 데이터에 대한 접근 조건이 개인 통신에 대한 권리를 존중하고 모든 관련 당사자의 권리와 정당한 이익을 적절하게 보호해야 함을 인지한다.

따라서, 본 강령의 서명인들은 다음을 약속한다:

팩트체킹 공동체와의 협력 COOPERATION WITH THE FACT-CHECKING COMMUNITY

준칙 30. 관련 서명인은 EU 팩트체킹 공동체와 팩트체커들 사이의 팩트체커에게 제공될 자원 및 지원에 대한 투명하고 체계적이며 개방적이고 재정적으로 지속가능하며 비차별적 협력을 위한 프레임워크를 구축한다.

준칙 30을 충족하기 위해:

조치 30.1. 관련 서명인은 모든 회원국에서 팩트체크 적용 범위를 확보하기 위해 이들과 독립적인 팩트체크 조직((e)에서 정의한 바와 같이) 간의 합의를 형성한다. 해당 합의는 높은 윤리적, 전문적 기준을 충족하고 투명하고 개방적이며 일관적이며 비차별적인 조건에 기초해야 하며 팩트체커의 독립성을 보장한다.⁷⁸⁾

78) 서명인은 준칙 29에 의거, 상업적으로 민감하거나 기밀할 수 있는 개별 계약에 대한 특정 정보를 고려하면서 의미있는 정보를 제공할 것을 약속한다.

QRE 30.1.1: 관련 서명인은 팩트체킹 기관과의 합의의 성격, 예상 결과, 관련 정량 정보(예: 팩트체크된 내용, 적용 범위 확대, 태스크포스 내에서 합의에 따라 추가로 논의될 팩트체킹 통합의 변화); 이러한 합의 관련 공동의 표준 및 조건 등을 보고 및 설명한다.

QRE 30.1.2: 관련 서명인은 그들이 합의한 팩트체킹 기관들을 나열한다(팩트체킹 기관이 보복이나 폭력에 대한 합리적인 두려움에 근거하여 그러한 공개를 반대하지 않는 한).

QRE 30.1.3: 관련 서명인은 회원국 수준에서 온라인 허위조작정보를 방지하기 위한 팩트체킹 기관을 지원하고 각 회원국 내 팩트체킹 보도를 이뤄내는 서비스에 해당하는 경우, 그에 할당된 재원을 보고한다.

SLI 30.1.1: 관련 서명인은 팩트체킹 기관과의 합의에 따라 관련 서비스별, 언어별 팩트체킹 기관과의 합의의 수를 포함한 회원국 및 언어에 대해 보고한다.

조치 30.2. 관련 서명인은 그들의 서비스상에서 허위조작정보에 대응하는 독립적인 유럽의 팩트체킹 기관들에 공정한 재정적 지원을 제공한다. 이러한 재정 지원은 개별적인 합의, 다수의 팩트체킹 기관들과의 합의 또는 그러한 합의를 체결할 권한을 가진 독립적인 유럽의 팩트체킹 기관에 의해 선출된 대표들과의 합의의 형태일 수 있다.

QRE 30.2.1: 관련 서명인은 팩트체커들에 대한 공정한 재정적 지원을 보장하기 위해 취해진 조치와 일반적 기준, 그리고 높은 윤리적, 전문적 기준, 팩트체킹 기관의 독립성 및 투명성, 개방성, 일관성 및 비차별성 조건을 보장하기 위해 해당 합의에 사용된 기준에 대해 보고한다.

QRE 30.2.2: 관련 서명인은 서명인의 팩트체킹 프로그램의 특징과 효과를 검토하기 위해 협력 기관과의 정기적인 검토에 참여, 보고한다.

QRE 30.2.3: 유럽 팩트체킹 기관은 직접적으로 (본 강령의 서명인으로서) 또는 간접적으로 (예: EDMO 또는 독립적인 유럽 팩트체킹 기관에 의해 선출된 대표자 투표를 통해) 이러한 합의를 통해 제공된 개별 보상의 공정성에 대해 보고한다.

조치 30.3. 관련 서명인은 팩트체커들의 국가 간 협력에 기여한다.

QRE 30.3.1: 관련 서명인은 팩트체크, 언어 또는 그러한 협력을 도모한 회원국의 사례를 포함하여, 팩트체커들의 국가 간 협력을 용이하게 할 조치를 보고한다.

조치 30.4. 상기 조치를 개선하기 위해 관련 서명인은 EDMO 및 독립적인 유럽 팩트체킹 기관의 선출된 기관 대표와 협의한다.

QRE 30.4.1: 관련 서명인은 조치 30.3 및 30.4에 기술된 협력 프레임워크 개선을 포함하여 EDMO와 독립적인 유럽 팩트체킹 기관의 선출된 기관 대표가 참여해 취해진 조치에 대해 참여 계획 관련 사전보고 및 사후 조치에 대해 사후 보고를 행한다.

서비스 내 팩트체킹 이용 및 통합

USE AND INTEGRATION OF FACT-CHECKING IN SIGNATORIES' SERVICES

준칙 31. 관련 서명인은 모든 회원국 및 언어 수준에서 자사 플랫폼 서비스, 절차, 콘텐츠에 팩트체커의 작업을 통합, 제시, 또는 일관되게 이용한다.

준칙 31을 충족하기 위해:

조치 31.1. UGC(User Generated Content)를 제공하는 관련 서명인은 모든 회원국과 관련 서비스 형식에 있어 플랫폼 내 서비스, 절차, 그리고 콘텐츠에 대한 독립적인 팩트체커들의 작업을 포함하고, 소개하며, 또는 지속적으로 사용한다. 이를 위해 관련 서명인은 연구 및 테스트를 수행하고 문서화하는 것을 시작으로 팩트체커와 협력한다.

조치 31.2. 제품 또는 절차에서 팩트체크를 포함한 관련 서명인은 팩트체크가 수용자에 미칠 영향력을 제고하는 데 도움이 될 수 있도록 라벨링, 정보 패널, 또는 정책 집행과 같은 신속하고 효율적인 메커니즘을 활용할 것을 보장한다.

QRE 31.1.1 (조치 30.1 및 조치 30.2의 경우): 관련 서명인은 조치 31.1 및

조치 31.2 관련 목적으로 테스트 솔루션에 적용된 전체 결과 및 방법론을 포함해 구체적인 활동 및 계획에 대해 보고한다.

SLI 31.1.1 (조치 30.1 및 30.2의 경우): 회원국 차원에서 서비스별 팩트체크 사용 현황 및 신속하고 효율적인 메커니즘에 대해 보고하며, 대략 다음과 같은 것들이 포함된다(서비스에 따라 다름): 게시된 팩트체크 기사 수, 팩트체크 기사 도달 수, 팩트체커들에 의해 검토된 기사 수.

SLI 31.1.2 (30.1과 30.2 측정항목의 경우): 팩트체크된 기사에 근거해 표기된 콘텐츠의 수와 같은 조치의 영향력, 또는 오류 또는 오해의 소지가 있는 것으로 확인된 정보와의 이용자 상호작용에 대한 조치의 영향력에 대해 유의미한 지표를 통해 측정한다.

SLI 31.1.3 (조치 30.1 및 30.2의 경우): 서명인은 연구원, 팩트체커, 위원회, ERGA, 그리고 대중이 준칙 31을 준수하기 위해 취한 조치들의 영향력을 이해하고 평가할 수 있는 방법으로 SLI 31.1.1 및 31.1.2의 맥락을 제공하는 것의 중요성을 인지한다. 이를 위해 관련 서명인은 해당 SLI의 맥락을 파악하는 데 도움이 되는 정량 정보의 기준점을 포함할 것을 밝힌다. 관련 서명인은 그들의 기준 보고에 앞서 맥락 파악을 위해 그들이 사용할 정량 정보의 기준점의 종류를 상설 태스크포스에 제시하고 논의한다.

조치 31.3. 관련 서명인(팩트체커와 플랫폼들을 포함하나 반드시 이에 국한되지는 않음)은 EDMO 및 독립적인 유럽 팩트체크 기관의 선출된 기관 대표와 협력하여 팩트체커 대표가 관장할 팩트체크 내용의 저장소를 만든다. 관련 서명인(즉, 플랫폼)은 다른 서명인 및/또는 기타 관련 이해관계 있는 기관과 함께 저장소 설립 재정 지원에 기여할 것을 약속한다. 재정 지원은 저장소 설립 이후 상설 태스크포스에서 매년 재평가되며, 12개월을 넘기지 않는다.

QRE 31.3.1: 관련 서명인은 (서명인에 따라) 재정적 기여, 기술적 지원, 자원 조달, 저장소에 추가된 팩트 체크 등을 포함해 전체 저장소 프로젝트에 대한 자신들의 작업과 기여에 대해 보고한다. 추가적인 관련 측정은 상설 태스크포스에서 조사되어야 한다.

조치 31.4. 관련 서명인은 플랫폼과 언어에 걸쳐 이 공동 저장소의 효율적인 사용을 도모하기 위한 기술 솔루션을 보고한다. 그들은 관련 후속 조치를 확인하기

위해 상설 태스크포스와 해당 솔루션에 대해 논의한다.

QRE 31.4.1: 관련 서명인은 플랫폼 간 공동 저장소의 효율적인 사용을 촉진하기 위해 구현한 솔루션에 대한 태스크포스와의 논의에 기반해 가능한 그들이 조사한 기술 솔루션에 대해 보고한다.

팩트체커의 관련 정보에 대한 접근 FACT-CHECKERS' ACCESS TO RELEVANT INFORMATION

준칙 32. 관련 서명인은 EDMO⁷⁹⁾ 및 독립적인 유럽 팩트체킹 기관을 대표해 선출된 조직과 협력해 설계한 프레임워크에 정의된 바와 같이, 팩트체킹의 품질과 영향력을 극대화하는데 도움이 되는 정보에 대한 접근을 팩트체커에게 신속하게, 가능한 자동화된 방식으로 제공한다.

준칙 32를 충족하기 위해:

조치 32.1. 관련 서명인은 팩트체커에게 해당 내용, 노출, 클릭 또는 상호작용에 근거하여 취한 (서비스별) 조치와 같이 사실 확인된 내용의 효과를 시간에 따라 정량화하는 데 도움이 되는 정보를 제공한다.

조치 32.2. 이용자 생성 콘텐츠(UGC)를 보여주는 관련 서명인은 가능한 한 자동화된 적절한 인터페이스를 제공하여 팩트체킹 기관이 해당 플랫폼의 콘텐츠가 미치는 영향에 대한 정보에 접근할 수 있도록 하고, 서명인이 팩트체커의 작업을 이용, 신뢰, 피드백을 제공하는 방식에 대한 일관성을 보장한다.

QRE 32.1.1 (조치 31.1 및 31.2의 경우): 관련 서명인은 조치 31.1 및 31.2에서 언급된 정보를 팩트체커에게 제공하기 위해 설치된 인터페이스 및 기타 도구에 대한 세부 정보를 제공한다.

SLI 32.1.1 (조치 31.1과 31.2의 경우): 관련 서명인은 조치 32.1과 32.2에서 언급한 정보를 팩트체커에게 제공하기 위해 사용하는 인터페이스와 기타

79) EDMO(European Digital Media Observatory)

도구의 사용에 대한 정량적 정보를 제공한다(예: 월별 사용자).

조치 32.3. 관련 서명인은 그들의 협력을 강화하기 위해 정기적으로 자신들과 팩트체킹 공동체 간의 정보를 교류한다.

QRE 32.3.1: 관련 서명인은 그들의 협력을 강화하기 위해 수행된 소통 채널 및 교류(조치 32.1 및 32.2에서 언급한 정보, 인터페이스 및 기타 도구에 대한 성공 및 만족), 그리고 이러한 교류가 이끈 결론 등에 대해 보고한다.

팩트체커의 기준 FACT-CHECKERS' STANDARDS

준칙 33. 관련 서명인(예: 팩트체크 기관)은 엄격한 윤리 규칙 및 투명성 규칙에 따라 운영하며, 그들의 독립성을 보호한다.

준칙 33을 충족하기 위해:

조치 33.1. 관련 서명인은 국제 팩트체킹 네트워크(IFCN) 강령, 또는 독립적인 유럽 팩트체킹 기관의 전문 진실성에 대한 미래 강령에서의 검증된 서명인과 같은 요건을 준수한다.

QRE 33.1.1: 관련 서명인은 조치 33.1에 언급된 것과 같은 수단에 그들의 회원 신분과 엄격하게 윤리적이고 투명한 규칙을 보장하고 그들의 독립성을 보호하기 위한 결과로서 취한 행동을 보고한다.

조치를 조치 33.1에 언급된 것과 같은 도구에 보고합니다.

SLI 33.1.1: 관련 서명인은 IFCN 인증을 받았거나, 또는 전문 진실성에 대한 미래 강령의 구성원인 유럽 팩트체커의 수를 보고한다.

VIII. 투명성 센터 TRANSPARENCY CENTRE

다음과 같은 사실로:

- (a) 서명인은 온라인 허위조작정보 대응에서 투명성과 책임의 핵심적인 중요성과 본 강령의 이행에 관한 정확하고 시의적절한 정보를 받을 수 있는 대중의 정당한 이익을 인지한다. 따라서 서명인은 각자의 서비스와 제품의 투명성을 보장하기 위한 개인적인 노력을 넘어 본 강령의 이행과 관련하여 공동의 참조지점과 투명성의 중요성을 인지한다.
- (b) 따라서 서명인은 공개적이고 이용자 친화적이며 검색가능한 공동의 투명성 센터 웹사이트를 구축하는 것이 중요하다고 본다. 따라서 투명성 센터 웹사이트를 설치하고 자금을 조달하여 이를 유지하고 모든 정량적 측정 기준을 포함한 관련 정보를 적시에 완전하게 최신 상태로 유지하는 것에 동의한다. 서명인은 시간이 지남에 따라 투명성 센터 내 이용자 기반 최신 기능 및 형식을 고려한다.
- (c) 서명자들은 테스크포스가 검수할 투명성 센터 웹사이트의 적절한 거버넌스의 중요성을 인지한다.

따라서, 본 강령의 서명인들은 다음을 약속한다:

준칙 34. 본 강령의 이행에 있어 투명성과 책임성을 보장하기 위해, 관련 서명인은 공개적으로 이용 가능한 공동의 투명성 센터 웹사이트를 구축하고 유지한다.

준칙 34를 충족하기 위해:

조치 34.1. 서명인은 공동으로 투명성 센터 웹사이트를 설립, 유지하며, 이 웹사이트는 본 강령의 서명 이후 6개월 이내에 운영되고 대중에게 제공된다.

조치 34.2. 서명인은 투명성 센터 웹사이트의 설치 및 운영에 필요한 적절한 자금을 제공하는데, 여기에는 유지보수, 상시 운영 및 관리, 그리고 정기적인 업데이트

이트가 포함된다. 자금 기부는 서명인의 활동 특성에 비례해야 하고, 웹사이트 운영 및 유지보수에 충분해야 하며, 각 서명인의 위험 특성 및 경제적 능력에 비례해야 한다.

조치 34.3. 관련 서명인은 본 강령이 그들의 서비스에 적용될 수 있는 범위 내에서 투명성 센터의 정보에 기여한다.

조치 34.4. 서명인은 태스크포스 내 투명성 센터의 기능과 자금 조달 및 매년 태스크포스가 매년 기록 및 검토하는 것에 합의한다.

조치 34.5. 태스크포스는 정기적으로 투명성 센터에 대해 논의하고 조정이나 조치가 필요한지 여부를 평가한다. 서명인은 태스크포스에서 결정된 조치와 조정을 합리적인 일정 내에 이행할 것을 합의한다.

준칙 35. 서명인은 투명성 센터에 본 강령의 준칙 및 조치의 이행과 관련된 모든 정보를 포함하고, 이러한 정보가 서비스별로 이해하기 쉬운 방식으로 제공되며, 쉽게 검색할 수 있도록 보장한다.

준칙 35를 충족하기 위해:

조치 35.1. 서명인은 그들이 서명한 각 준칙과 조치에 따라 해당 준칙과 조치를 이행하기 위해 그들의 서비스에 적용한 서비스 조건 및 정책을 투명성 센터 내에 제시한다.

조치 35.2. 서명인은 지리적 및 언어적 범위를 포함하여 서비스별 정책의 시행 및 집행에 대한 정보를 제공한다.

조치 35.3. 서명인은 투명성 센터가 본 강령의 준칙 이행을 평가하는 그들의 보고를

보관할 수 있도록 보장한다.

조치 35.4. 위기 상황에서 서명인은 위기와 관련한 구체적인 대응 조치에 대한 정보를 공개하기 위해 투명성 센터를 활용한다.

조치 35.5. 서명인은 투명성 센터가 최첨단 기술로 구축되어 있고, 이용자 친화적이며, (준칙 및 조치별 포함) 관련 정보를 쉽게 추적할 수 있도록 보장한다. 투명성 센터의 이용자들은 서명인의 정책과 조치의 변화를 쉽게 추적할 수 있다.

준칙 36. 서명인은 투명성 센터에 포함된 연관 정보를 적시에 완전하게 업데이트한다.

준칙 36을 충족하기 위해:

조치 36.1. 서명인은 정책 및 이행 조치 관련 변경 사항에 대한 업데이트를 적시에, 어떤 상황에서도 변경 사항이 발표되거나 이행된 후 30일 이내에 제공한다.

조치 36.2. 서명인은 모니터링 프레임워크가 예측하는 정기 보고와 함께 서비스 수준 지표, 보고 요소 및 구조적 지표를 정기적으로 업데이트한다. 관련 서명인이 첫 보고 이후 보다 정기적으로 투명성 센터를 현행화할 것을 권장한다.

조치 36.3. 서명인은 본 강령과 모니터링 프레임워크에 관한 상설 태스크포스의 최신 결정을 반영하기 위해 투명성 센터를 현행화한다.

QRE 36.1.1 (준칙 34-36의 경우): 서명인은 초기 이행 보고시, 투명성 센터의 발전 상태, 기능, 센터가 보유한 정보, 이것의 기능 및 운영에 관련된 그 밖의 정보들을 제시한다. 해당 정보는 투명성 센터의 운영 또는 추가 내용에 개입한 서명인들이 공동으로 작성할 수 있다.

QRE 36.1.2 (준칙 34-36의 경우): 서명인은 보고 시 시간이 지남에 따른 투

명성 센터의 내용, 운영, 또는 기능의 변화를 제시한다. 해당 업데이트는 투명성 센터의 운영 또는 추가 내용에 개입한 서명인들이 공동으로 작성할 수 있다.

SLI 36.1.1 (준칙 34-36의 경우): 서명인은 웹사이트의 월평균 방문과 같은 투명성 센터 이용에 대한 의미 있는 정량적 정보를 제공한다.

IX. 상설 태스크포스 PERMANENT TASK-FORCE

다음과 같은 사실로:

- (a) 서명인은 기술적, 사회적, 시장적, 입법적 변화의 관점에서 준칙과 조치를 갱신할 수 있는 능력을 포함하여 본 강령의 개선과 조정을 목표로 하는 상설 태스크 포스를 설립하는 것의 중요성을 인지한다.
- (b) 서명인은 본 강령의 효율성과 이것에 대한 모니터링의 개선을 목표로 태스크포스에 참여하는 것의 중요성을 인지한다.
- (c) 서명인은 태스크포스를 서명인, 본 위원회, EEAS, ERGA, EDMO 및 기타 초빙한 제3자들 간의 중요한 교류의 장으로 간주한다.
- (d) 서명인은 태스크포스 전체 회의 중 특정 업무 흐름을 구현하기 위해 하위 그룹에 의한 작업의 중요성을 합의한다.
- (e) 서명인은 태스크포스의 활동에 관련 전문가를 참여시키는 것과 허위조작정보 현상과 관련된 정보를 최신 정보로 유지하고 그들의 통찰력을 모으기 위해 제3의 이해관계자들과의 교류를 조직하는 것의 중요성을 인지한다.
- (f) 서명인은 허위조작정보의 확산에 대한 실천 강령의 효과와 성공을 평가하는 것의 중요성을 인지한다. 따라서, 관련 서명인은 본 강령 운영 첫해 안에 태스크포스, EDMO, ERGA 및 관련 제3의 전문가들과 협력하여 본 강령의 효과와 성공을 평가하는 데 도움을 줄 수 있는 첫 구조적 지표 세트를 개발한다. 서명인은 이후에도 태스크포스, EDMO, ERGA 및 관련 제3의 전문가들과 협력하여 시간 경과에 따라 이러한 지표를 시험하고 필요한 경우 조정한다.

따라서, 본 강령의 서명인들은 다음을 약속한다:

준칙 37. 서명인은 상설 태스크포스에 참여한다. 태스크포스에는 강령의 서명인, EDMO의 대표 및 EGRA⁸⁰⁾의 대표가 포함된다. 유럽집행위원회가 의장을 맡고 유럽대의협력청(EEAS)의 대표도 포함된다. 태스크포스의 업무를 지원하기 위해 관련 전문가를 옵저버로 초청할 수 있다. 태스크포스의 결정은 합의제로 이루어진다.

준칙 37을 충족하기 위해:

조치 37.1. 서명인은 태스크포스에 참여하여 해당 업무에 기여한다. 특히 규모가 작거나 신생 서비스들은 그들의 자원, 규모, 위험 특성에 비례하여 태스크포스의 업무에 기여한다. 소규모, 또는 신생 서비스들은 그들의 자원을 한데 모으고 태스크포스에서 서로를 대표하는 것에 동의한다. 태스크포스는 필요에 따라 최소 6개월마다 본회의를 반드시 진행하고, 관련해서 특정 이슈나 업무 흐름에 관한 하위 그룹 회의 또한 개최한다.

조치 37.2. 서명인은 태스크포스에서 특히 다음과 같은 임무(단, 이에 국한되지 않음)에 임하기로 동의한다:

- 선거나 위기와 같은 특수한 상황에서 활용할 수 있는 위험성 평가 방법과 신속한 대응 체계의 구축.
- 선거나 위기와 같은 특별한 상황에서의 업무 협력 및 조정.
- 본 강령의 준칙과 조치의 이행을 위한 조화로운 보고 템플릿과 정제된 보고 방법, 그리고 감시 목적을 위한 적절한 데이터 공개에 대한 합의
- 향후 모니터링 주기 전반에 걸쳐서, 해당 목적을 위한 데이터 공개 방법과 형식뿐만 아니라 조화로운 보고 템플릿에 대한 품질 및 효과에 대한 검토 및 필요에 따른 조정 업무.
- 해당 지표뿐만 아니라 관련 변화를 측정을 위해 제공된 서비스 수준에서

80) ERGA(European Regulators Group for Audiovisual Media Services)

의 품질과 효과, 구조적 지표 및 기준점에 대한 평가에 기여

- 회원국 수준에서의 측정을 위해 구조적 지표와 설계 메커니즘의 개선, 테스트, 조정.
- 악의적 행위자에 의해 활용되는 TTP 목록에 대한 합의, 게시 및 현행화와 이에 대응하려는 조치의 기본 요소, 목표, 기준점을 본 강령 IV에 의거하여 설정.
- 특히 안전 설계의 새로운 모범 사례, 플래그 설정의 소급, 팩트체크 저장소, 입증 도구 등과 같은 본 강령의 준칙에 관한 연구 및 전문가의 의견 및 최신 근거들에 대한 추구 및 논의.
- 다양한 서비스에 걸친 팩트체크 기관과의 합의에 대한 보고 의무를 이행을 위해 서명인 제공할 적절한 정량적 정보에 대한 논의 및 지침 제공
- 기술적, 사회적, 시장적, 입법적 변화의 관점뿐만 아니라 새로운 서명인을 수용하는 차원, 그리고 태스크포스가 필요하다고 판단하는 경우, 본 강령의 준칙과 조치의 업데이트 여부에 대한 정기적 논의.
- 소규모, 신생 서비스를 위한 조치에 대한 적절성과 일관성 검토.
- 새로운 서명인 통합 및 관련 동료들에게 본 강령에 대한 홍보.

조치 37.3. 태스크포스는 제3의 전문가의 참여를 포함한 운영 규칙을 정의하고 합의하며, 이는 서명인과 유럽위원회가 공동으로 작성하고 태스크포스 구성원과의 협의로 합의된다.

조치 37.4. 서명인은 관련 서명인의 참여 하에 본 강령의 시행 및 개정과 관련된 특정 문제들을 전담하는 하위 그룹을 설립하는 것에 동의한다.

조치 37.5. 태스크포스는 필요할 때, 그리고 어떠한 경우에도 매년 한 번 이상, 관련 이해 관계자 그룹 및 전문가들과 회의를 조직하여 본 강령의 운영에 대해 알리고 허위조작정보 분야의 중요한 발전과 관련된 의견을 수렴한다.

조치 37.6. 서명인은 기술적, 사회적, 시장적, 입법적 변화의 관점에서, 시간이 지나 그들의 관행과 접근 방식의 개선에 따라 준칙이나 조치가 변경/개선된 경

우, 이를 나머지 태스크포스에 알리기로 합의한다. 필요한 변경 사항을 논의한 후, 관련 서명인은 그에 따라 가입 문서를 업데이트하고 다음 보고서에서 변경 사항에 대해 제시한다.

QRE 37.6.1: 서명인은 보고 중에 태스크포스 업무에 어떻게 참여하고 있는지 기술하며, 여기에는 그들이 참여한 하위 그룹이 포함된다.

X. 본 강령의 모니터링 MONITORING OF THE CODE

다음과 같은 사실로:

- (a) 서명인은 본 강령이 회원국 수준을 포함해 전 유럽연합 내 온라인 허위조작정보 확산을 줄이는 데 있어서 본 강령의 영향력 및 시행을 평가할 수 있는 강력한 모니터링 체계가 필요함을 인지한다.
- (b) 서명인은 정책과 행동 관련 변경 사항을 표시해야 하는 본 강령의 준칙 이행에 대한 정기적인 보고 및 평가 메커니즘을 마련한다.
- (c) 서명인은 EU 전체에 걸쳐 본 강령의 효율적이고 지속적인 이행을 보장하기 위해 EU와 회원국 수준에서 정기적인 보고와 모니터링의 중요성을 인지한다.
- (d) 서명인은 투명성 센터와 태스크포스를 대중에게 투명성을 제공하는 모니터링 및 보고에 대한 검토, 조정, 그리고 개선을 위한 주요 도구로 간주한다.
- (e) 서명인은 DSA에 정의된 대로 초대형 온라인 플랫폼이 아닌 플랫폼의 제공자들에게 불균등한 부담을 주지 않을 것을 확실히 하기 위해, 본 강령에 따른 보고 의무는 서명인의 규모, 서비스의 성격 및 이용 가능한 자원을 고려할 필요가 있다는 것을 - 서문에 명시된 대로 - 인지한다.

따라서, 본 강령의 서명인들은 다음을 약속한다:

준칙 38. 서명인은 강령에 따른 준칙의 이행을 보장하기 위해 적절한 재정적 인적 자원을 투입하고 적절한 내부 프로세스를 마련한다.

QRE 38.1: 관련 서명인은 본 강령 준수를 목적으로 회원국 및 EU 언어 전반에 걸친 완전한 적용될 수 있는 범위를 설정하고자 서비스별 그들이 참여할 팀과 내부 프로세스를 제시한다.

준칙 39. 서명인은 (본 강령 서명 후 6개월 이내인) 이행 기간 종료 후 1개월 이내에 서문에 명시된 기본 보고서(baseline report)를 유럽집행위원회에 제공한다.

준칙 40. 서명인은 서비스수준지표(SLIs) 및 질적보고요소(QREs)에 대해 정기적으로 보고한다. 제공된 보고서와 데이터는 각 서명인, 서비스 및 회원국의 차원에서 강령의 준칙과 조치의 이행 정도를 철저하게 평가할 수 있도록 해야 한다.

준칙 40을 충족하기 위해:

조치 40.1. DSA에 정의된 초대형 온라인 플랫폼 관련 서명인은 서비스 및 회원국 수준에서 관련 QRE 및 SLI를 포함하여 본 강령에 서명한 준칙 및 조치의 이행에 대해 6개월마다 보고한다.

조치 40.2. 나머지 서명인은 서비스 및 회원국 수준에서 관련 QRE 및 SLI를 포함하여 본 강령에 서명한 준칙 및 조치의 이행에 대해 매년 보고한다.

조치 40.3. 서명인은 적어도 이 코드에 따른 보고 기간에 맞추어 관련 QRE 및 SLI를 정기적으로 투명성 센터에 업데이트할 것이다.

조치 40.4. 서명인은 테스크포스 내에서 조화로운(harmonised) 보고 템플릿을 개발한다.

조치 40.5. 서명자들은 특히 유럽위원회, ERGA 및 EDMO의 피드백을 기반으로 테스크 포스 내에서 SLI를 포함한 코드의 모니터링 및 보고 프레임워크를 개선

하고 최적화하기 위해 정기적으로 작업할 것이다.

조치 40.6. 서명인은 유럽연합 집행위원회와 협력하며 이들의 합리적인 요청에 응하며, 합리적인 정보, 데이터, 그 외 본 강령 이행을 평가하는데 필요한 입력 사항들을 제공한다.

준칙 41. 서명인은 구조적 지표 개발을 목적으로 태스크포스에서 본 강령 서명 후 9개월 이내 첫 구조적 지표 세트 공개, 첫 정식 보고서(full report) 첫 측정치 공개를 약속한다. 해당 목표를 달성하기 위해 서명인은 본 강령에서 합의한 초기 구조적 지표 세트의 테스트 및 조정을 포함해 해당 이행을 지원할 것을 약속한다. 이는 관련 서명국 및 EU와 회원국 수준에서의 전체 온라인 생태계 내 온라인 허위조작정보의 확산을 줄이고자 하는 본 강령의 효과를 평가하기 위한 것이다. 서명인은 ERGA와 EDMO를 포함한 관련 주체들과 협력한다.

준칙 41을 충족하기 위해:

조치 41.1. 본 강령에 서명한 후 1개월 이내에 서명인은 이 목표를 해결하기 위한 실무 그룹을 설립한다. 이 그룹은 플랫폼 서명인이 제공할 데이터 기준(data points)과 해당 기준에 기반한 구조적 지표를 측정할 방법론을 제안하면서, 비플랫폼 서명인에 의해 실행될 수 있도록 하는 의무를 맡는다. 서명인은 온라인 허위 정보의 확산 및 출처에 관해 널리 퍼져있거나, 혹은 기타 상황에 맥락화된 측정 기준과 같이 그룹에서 결정할 지표를 측정하는 데 적합한 데이터 기준을 공유한다. 서명인은 이 준칙 목표 달성을 위해 필요한 작업을 평가하고, 재정적 지원 필요 여부를 태스크포스에서 논의한다.

조치 41.2. 실무 그룹은 3개월마다 태스크포스에 진행 상황을 보고한다. 해당 작업과 결과물을 알리기 위해 EDMO, ERGA 및 연구자를 포함하되 이에 국한하지 않고 전문 이해관계자들과도 상의한다. 본 강령에 서명한 후 7개월 이후에,

외부 이해관계자들을 소집하여 지금까지의 진행 상황을 발표하고 피드백을 얻기 위한 컨퍼런스를 개최한다.

조치 41.3. 본 강령에 서명한 후 6개월까지, 실무 그룹은 태스크포스와 함께 해당 구조적 지표에 대한 실행 가능한 제안 사항을 상정한다. 9개월까지, 관련 서명인은 실무 그룹 내 다른 서명인에게 구조적 지표 측정에 필요한 데이터 기준을 제공하고, 조율된 구조적 지표를 공개적으로 공유한다. 실무 그룹은 시간이 지남에 따라 여전히 세부 사항이 필요하다는 합의하에, 서명인들의 전체 방법론뿐만 아니라 첫 전체 보고에 의거에 따른 구조적 측정의 측정치를 공개한다. 서명인은 보고 기간에 맞춰 측정치를 계속 현행화할 것을 약속한다. 측정치는 전체 생태계 및 다른 서비스 간의 시간 경과에 따른 모니터링이 가능한 방식으로 투명성 센터에 공개된다.

준칙 42. 관련 서명인은 선거나 위기와 같은 특정한 상황에서 유럽집행위원회의 요청이 있을 경우, 태스크포스에 의해 설립된 신속 대응 시스템에 따라 비정기 특별 보고서(ad-hoc specific report) 및 정기 모니터링 내 특정 부분을 포함한 비례적이고 적절한 정보 및 데이터를 제공한다.

준칙 43. 서명인은 태스크포스에서 합의한 바와 같이, 보고서 및 데이터 공개를 위한 적절한 보고 서식 및 개선된 방법론에 따라 보고서를 작성하고 데이터를 제공한다.

준칙 44. 초대형 온라인 플랫폼 사업자인 관련 서명인은 DSA와 제휴를 모색하고, 본 강령에 따라 수행한 준칙의 준수 여부에 대해 자체 비용으로 감사를 받는다. 감사는 초대형 온라인 플랫폼 사업자로부터 독립적이고 이해관계가 상충되지 않는 조직에서 수행해야 한다. 이러한 조직은 허위조작정보 분야에서 전문성이 입증되고, 적절한 기술적 역량 및 능력을 갖추고 있어야 하며, 특히 감사

표준 및 가이드라인에 기반해 입증된 객관성 및 직업윤리를 갖춰야 한다.

호주 허위조작정보 및 오정보 대응 실천강령. 2022

Australian Code of Practice on Disinformation and misinformation. 2022

1. 서문 Preamble

1.1. *배경*: 호주의 허위조작정보와 오정보 실천강령(이하 ‘강령’)은 호주의 디지털 산업을 지지하는 비영리 산업 단체인 디지털산업그룹(Digital Industry Group Inc. (이하 DIGI))에 의해 마련되었다. 본 강령은 디지털 플랫폼 연구, ‘디지털 시대의 규제: 정부 대응과 로드맵(Regulating in the Digital Age: Government Response and Implementation Roadmap)’에 명시된 정부 정책에 따라 수행되었다. 정부는 주요 디지털 플랫폼의 허위조작정보와 뉴스 콘텐츠의 신뢰성에 대한 우려를 해소하기 위해, 플랫폼이 무엇을 할 것인가에 대한 자발적인 행동강령을 마련하도록 요청했다. 본 강령은 ‘호주 디지털 플랫폼의 허위조작정보와 뉴스 품질: 강령 개발 지침에 관한 의견서(Misinformation and News Quality on Digital Platforms in Australia: A Position Paper to Guide Code Development)’에 제시된 호주 통신미디어청(Australian Communications and Media Authority)의 지침 또한 고려하고 있다. 공개 협의 과정에서 수렴된 의견을 반영하고, 검토 과정을 거쳐 2022년 12월, 강령이 개정되었다.

1.2. *주제*: 허위조작정보와 오정보는 광범위하고 다면적인 사회적 문제의 일환으로, (기존의 민주적 프로세스 또는 공중 보건 등) 공공재에 해를 끼칠 수 있는 정보를 확산하는 다양한 오프라인과 온라인 행위를 포함한다. ‘허위조작정보(disinformation)’, ‘오정보(misinformation)’, ‘가짜뉴스(fake news)’와 같은 개념은 사람에 따라 다른 의미를 지니며, (합리적인 사람들 간에 의견이 상이할 수 있는) 가치 중심

의 정치적 문제를 두고 다른 의견을 가진 사람들을 공격하는 데 사용될 경우, 정치적으로 악용될 수 있다.

- 1.3. *디지털 플랫폼의 역할*: 본 강령에 서명한 디지털 플랫폼은 호주 정보 생태계 내 중요한 행위자로서의 그 역할을 인지하고 있다. 디지털 플랫폼은 서비스와 제품 사용자들 사이에서 허위조작정보와 오정보가 확산되는 것을 방지하기 위해, 이미 다양한 대책을 시행하고 있다. 본 강령은 디지털 플랫폼을 통해 확산되는 허위조작정보와 오정보에 대응하기 위해, (본 강령에 정의된 바와 같이) 서명인들의 최소한의 준칙을 나타내고자 마련되었다.
- 1.4. *서명인*: 본 강령은 호주의 사용자에게 제품과 서비스를 제공하며, 최종 사용자 end-users가 온라인에서 허위조작정보와 오정보를 확산시킬 수 있는 위험을 식별하고, 온라인에서 허위조작정보 또는 오정보 확산을 감소시킬 수 있는 모든 디지털 플랫폼이 서명할 수 있다.
- 1.5. *최소 준칙*: 모든 서명인은 허위조작정보와 오정보로 발생할 수 있는 피해에 대한 타당한 보호 장치를 제공한다는 핵심 목표 공유하며, 5.2절에 명시된 준칙을 이행하도록 한다.
- 1.6. *사전 동의*: 디지털 산업은 혁신적이고 다양하며, 디지털 플랫폼은 지속적으로 발전하고, 다양한 서비스와 제품을 제공하는 각기 다른 비즈니스를 운영한다. 이에 따라 디지털 플랫폼이 허위조작정보와 오정보에 대응하기 위한 대책은 각각의 비즈니스의 맥락에서, 시간의 흐름에 따라 달라질 수 있다. 예컨대, 사용자 제작 콘텐츠 플랫폼에서 수행되는 대책은 검색 엔진에서 취해지는 조치와 상이하다. (디지털 플랫폼이 제공하는 각기 다른 서비스와 제품의) 허위조작정보와 오정보 사례를 다루기 위해 적합한 대책을 선택해야 하는 서명인의 요구를 반영하여, 본 강령은 1.4절에 명시된 최소 준칙과 더불어, 서명인에게 다양한 대책과 목표를 선택할 수 있는 권한을 제공한다.

- 1.7. *비례의 원칙*: 본 강령이 해결하고자 하는 사용자 행위, 콘텐츠 그리고 피해의 유형은 각 디지털 플랫폼에서 제공하는 다양한 서비스와 제품에 따라, 그 발생 정도와 영향이 크게 다를 수 있다. 이에 따라 (본 강령에 따른) 서명인의 준칙은, 서명인이 강령에 기반한 준칙에 따라 대책을 이행하도록 하기 위한 것이다. 이와 관련하여 6절은 서명인이 고려할 수 있는 맥락적인 요인에 대한 추가 지침을 제공한다.
- 1.8. *서명인이 이행할 수 있는 포괄적인 접근방식*: 서명인은 재량에 따라, 본 강령의 명시된 범위보다 더 포괄적인 콘텐츠 또는 제품과 서비스의 허위조작정보와 오정보에 대응하기 위한 정책과 프로세스를 구현할 수 있다.
- 1.9. *모든 관련 이해관계자 간의 협업과 협력의 필요성*: 본 강령의 목적은 디지털 플랫폼에 적용하는 데 있지만, 서명인들은 공공기관, 학계, 시민사회, 인플루언서, 뉴스 조직 등 다양한 이해관계자들이 허위조작정보와 오정보를 해결해야 하는 역할과 책임을 인지하며 강조한다. 허위조작정보와 오정보에 효과적으로 대응하기 위해서는 디지털 플랫폼뿐만 아니라, 이러한 다양한 이해관계자 간의 합의된 노력과 협력이 필요하다. 서명인들은 무엇이 효과적이고, 효과적이지 않은지에 관해 이해관계자들과의 지속적인 대화를 환영한다.
- 1.10. *모범 사례 지침*: 서명인들은 다른 디지털 서비스 등 정보 생태계의 다른 참여자들에게, (지속적으로 발전하는 허위조작정보와 오정보 문제에 대한 자체 대응책을 마련하는 데 있어서) 본 강령을 모범 사례 지침으로 사용하도록 권장한다.

2. 이행 원칙 Guiding Principles

- 2.1. *표현의 자유에 대한 보호* 디지털 플랫폼은 (호주 커뮤니티 전반에 걸쳐) 의견, 연설, 정보, 연구와 토론, 대화뿐 아니라, 창의적인 표현을 열린 방식으로 교환하기 위한 중요한 방법을 제공한다. 서명인은 (콘텐츠가 불법이 아닌 한) 허위 주장만을 근거로 하여, 정부 또는 다른 당사자에게 콘텐츠를 삭제하도록 강요해선 안 된다. 해당 주제를 비추어볼 때, 본 강령은 언론의 자유를 포함하되 이에 국한되지 않으며 (세계인권선언The Universal Declaration on Human Rights에서 명시된 바와 같이) 국제 인권에 특별히 주의를 기울인다. 서명인이 허위조작정보와 오정보에 대한 균형 잡힌 대응책을 마련할 경우, 다음과 같은 권리를 보호해야 할 필요성을 인식해야 한다.
- 2.2. *사용자 개인정보 보호* 디지털 플랫폼은 사용자의 개인정보를 중요시한다. 허위조작정보와 오정보의 확산에 대응하기 위한 디지털 플랫폼의 대책은 (관련 법률뿐 아니라 이용약관, 공개정책, 자발적 행동 강령을 포함하여) 호주 사용자의 개인정보에 대한 준칙을 위반해선 안 된다. 여기에는 디지털 플랫폼과 사적 디지털 커뮤니케이션을 사용할 때, 개인정보 보호에 대한 사용자의 기대를 존중하는 것이 포함된다. 또한 연구 목적으로 데이터에 접근하는 경우에도, 사용자의 개인정보를 보호해야 한다.
- 2.3. *광고 게재에 관한 정책과 프로세스* 디지털 플랫폼은 허위조작정보와 오정보를 확산하는 사람들에게 향하는 수익을 줄이기 위해, 서비스와 제품에 대한 광고 게재 정책과 프로세스의 중요성을 인식한다.
- 2.4. *사용자의 역량강화* 디지털 플랫폼은 (신뢰할 만한 최신 뉴스 혹은 사실에 기반을 둔 정보를 제공하는) 디지털 미디어 콘텐츠에 대한 정보를 숙고하고 습득할 수 있도록 장려해야 한다.

- 2.5. *서비스와 제품의 완전성과 보안*: 디지털 플랫폼은 허위조작정보를 확산할 목적이 있는 (서비스와 제품의) 허위계정을 금지하고 탐지하며, 이에 대한 대책을 마련함으로써, 서비스와 제품의 완전성과 보안을 보장하기 위한 노력의 효과에 대해 투명하게 소통해야 한다.
- 2.6. *독립 연구자 지원*: 디지털 플랫폼은 (허위조작정보와 오정보에 대한 동향, 이에 대응하기 위해 효과적인 방법을 공유하는) 학계와 독립 전문가들의 연구 성과에 대한 산업계 지원의 중요성을 인지하고 있다. 본 강령은 디지털 플랫폼이 독립적인 연구 이니셔티브initiatives에 참여할 수 있는 다양한 선택지를 제공한다.
- 2.7. *영향을 미치지 않는 준칙*: 본 강령은 디지털 플랫폼의 허위조작정보와 오정보에 대응하기 위한 여러 대책들을 침해하지 않는다.

3. 전문 용어 사전 Glossary

전문 용어 사전은 본 강령에서 사용되는 주요 용어에 관한 정보를 제공한다.

- 3.1. 디지털 콘텐츠는 (본 강령의 서명인이 소유하고 운영하는 플랫폼에서) 호주의 사용자를 대상으로 온라인에서 유포되는 콘텐츠로, (인공 지능 알고리즘 사용 등 자동화된 수단을 통해) 인위적으로 제작·조작 또는 수정된 콘텐츠를 포함한다.
- 3.2. 본 강령이 초점을 맞추는 *허위조작정보Disinformation*의 양상은 다음과 같다.

A. *실증적으로 잘못되었거나, 오도하거나, 기만적인 디지털 콘텐츠*

B. 이는 부정 행위를 통해 디지털 플랫폼 사용자들 사이에서 확산된다.

C. 해당 내용의 유포가 해를 끼치리라고 합리적으로 예상되는 콘텐츠

3.3. 기업 서비스Enterprise Services는 특정한 조직에서 이용하도록 고안된 클라우드 저장과 콘텐츠 전달 서비스를 포함한 소프트웨어와 서비스이다.

3.4. 피해Harm는 확실하고, 심각한 위협이 되는 피해를 의미한다.

A. 유권자 사기, 유권자 방해, 선거에 대한 오정보와 같은 민주적인 정치와 정책 결정 프로세스

B. 시민 보건의 보호, 소외계층 또는 취약계층의 보호, 치안과 보안 또는 환경과 같은 공공재

참고: 합리적으로 예측할 수 없는 피해는, (확실하고 심각한 위협을 지시하는) 피해Harm에 포함되지 않는다.

3.5. 부정행위Inauthentic Behaviour는 스팸spam과 다른 형태의 속임수, 조작, (자동화된 시스템을 통해 자행될 수 있는) 대규모의 공격적 행위를 포함하며, 디지털 플랫폼 사용자들의 온라인 대화에 인위적으로 영향을 미치거나 디지털 콘텐츠를 확산하도록 하는 행위를 포함한다.

3.6. 오정보Misinformation의 의미는 다음과 같다.

A. 실증적으로 잘못되었거나, 오도하거나, 기만적인 (종종 적법한) 디지털 콘텐츠

B. 이는 디지털 플랫폼의 사용자에게 의해 확산된다.

C. 해당 내용의 유포가 (명확하게 기만을 의도하지 않더라도) 해를 끼치리라고 합리적으로 예상되는 콘텐츠

3.7. 뉴스 출처는 뉴스 보도의 주제로부터 편집의 독립을 가진 뉴스 생산자로서;

- A. 호주언론협회 *Australian Press Council* 실천기준 또는 독립미디어협회 *Independent Media Council* 행동강령의 적용을 받는다.
- B. 상업 텔레비전 산업 실천강령, 상업 라디오 실천강령 또는 구독 기반 방송 텔레비전 실천강령의 적용을 받는다.
- C. 다른 국가의 편집 실천 기준을 명시한 규제 수단의 적용을 받는다.
- D. 양질의 저널리즘 제공과 관련된 내부 편집 기준의 적용을 받는다.
- E. 뉴스 보도 품질에 대한 정정 또는 불만 요청이 가능한 공공의 접근 메커니즘을 제공한다.

3.8. 디지털 광고 서비스는 (광고 배치가 서명인에 의해) 광고주에게 직접 판매되는 유료 디지털 광고 서비스를 의미한다.

3.9. 정치 광고 *Political Advertising*은 유료 광고를 의미한다.

- A. 정당을 대표하여 제작된 광고
- B. 선거, 국민투표의 결과 또는 (선거 감독 조직에 의해 관리되는) 연방, 주 또는 자치령의 (거소 투표와 같은) 정치적 프로세스를 지지하는 광고
- C. 호주 법에 따라 정치 광고로 규제된다.

3.10. 전문적인 뉴스 *Professional news*는 다음과 같은 사안을 보도하거나 조사하며, 또는 비판적 분석을 제공하는 뉴스 출처를 통해 제작된 온라인 자료물이다.

A. 최종 사용자를 공적인 토론에 참여시키고, 민주적 의사 결정 과정에 정보를 제공하는 현안 또는 사건

B. 장소, 지역 또는 국가 수준에서, 최종 사용자에게 공적인 중요성이 있는 현안 또는 사건.

3.11. 추천 시스템 *Recommender system*이란 온라인 플랫폼이 온라인 인터페이스에서, 서비스 수신자에게 특정 디지털 콘텐츠 항목을 제안하거나 우선순위를 지정하는 데 사용되는, (전체 또는 부분적으로) 자동화된 시스템을 의미한다. 여기에는 서비스 수신자가 수행한 검색의 결과나 디지털 콘텐츠 항목의 상대적 순서, 중요도를 결정하는 방법이 포함된다.

3.12. 검색 엔진 *Search Engines*는 사용자의 질의에 대한 응답으로, 월드와이드웹 *World Wide Web*에서 정보를 수집하고, 순위를 매기도록 설계된 소프트웨어 시스템으로 구성된다. 검색 엔진은 두 가지 방식으로 시스템을 자동화한다. 첫째, “웹 크롤러 *web crawler*”, “봇 *bots*”, 또는 “스파이더 *spiders*”로 알려진 소프트웨어를 사용하여 공개적으로 사용 가능한 웹페이지를 찾고, 이러한 웹페이지와 인터넷 사이트에서 정보를 자동으로 배열하고 수집한다. 둘째, 웹사이트의 링크로 결과를 제공하기 위해 순위 시스템을 사용한다. 순위 시스템은 관련성, 페이지의 유용성, 출처의 전문성 등 다양한 요소를 기반으로 순위를 매기는 일련의 알고리즘으로 구성된다. 각 요소에 적용되는 가중치는 질문의 특성에 따라 상이하다. “검색 엔진 *Search Engines*”은 자체 플랫폼에서 제공되는 검색 기능의 결과가 타사의 검색 엔진을 통해 구동되는 경우, 다운스트림 엔티티 *downstream entities*를 제외한다. 다운스트림 엔티티의 검색 결과가 인덱스, 그리고 랭킹 순서에 대한 법적 또는 운영적 통제를 가지고 있지 않기 때문이다.

3.13. 스폰서 콘텐츠 *Sponsored content*는 소셜 미디어 서비스와 계정 소유자 간의 유료 계약으로, 소셜 미디어 서비스가 계정 소유자의 팔로워 리스트 너머에 콘텐츠를 홍보

보하고, 이에 대한 대가로 지불을 받는 것을 의미한다. 여기서 검색 엔진의 유료 광고는 제외된다.

4. 본 강령의 범위, 적용과 착수

Scope, application and commencement of the Code

4.1. *범위(scope)*: 호주의 디지털 플랫폼이 제공하는 다양한 서비스와 제품 중에서, 허위조작정보 또는 오정보에서 비롯되는 위협의 범위와 영향이 매우 다양할 것이라는 점을 인지하고, 본 강령의 준칙은 (서명인이 허위조작정보와 오정보 확산 위협의 중요성을 인지하거나, 허위조작정보와 오정보로 인해 발생하는 피해와 영향을 줄이는 데 중요한 역할을 할 수 있다고 판단하는 경우) 호주의 최종 사용자에게, 디지털 콘텐츠를 확산하는 서비스와 제품에 적용될 것이다. 따라서 본 강령은 특정 유형의 디지털 콘텐츠에 국한되지 않고, 다양한 서명인과 제품 및 서비스에 적용된다. 예컨대, 제품 서비스의 범위로 다음 사항이 포함될 수 있다.

- A. *(공유된 콘텐츠를 포함한) 사용자 제작 디지털 콘텐츠의 배포*
- B. *(사용자 질의에 대한 응답하여) 검색엔진이 추출하고 순위 매긴 디지털 콘텐츠를 배포한다.*
- C. *디지털 플랫폼과 최종 사용자가 허위조작정보와 오정보에 대응할 수 있도록 지원하는 기술적 해결책을 제공한다.*
- D. *스폰서 콘텐츠와 디지털 광고 서비스를 제공한다.*
- E. *다양한 출처의 뉴스와 다양한 유형의 저널리즘 콘텐츠를 취합하고 유포한다.*

D에 대한 참고: 정치 광고는 오정보 범위에서 제외되지만, 부정 행위로 확산되는 경우 허위조작정보의 정의에 포함될 수 있다.

E에 대한 참고: 뉴스 취합 서비스에서 유포되는 전문 뉴스 콘텐츠는 오정보 정의에서 제외되지만, 부정 행위로 확산되는 경우 허위조작정보의 정의에 포함될 수 있다.

4.2. *제외된 서비스와 제품Exclude services and products*. 다음은 본 강령의 적용을 받지 않는 서비스의 제품이다.

A. 소프트웨어 애플리케이션을 통해 제공되는 서비스를 포함한, 개인 메시지 서비스

B. 소프트웨어 애플리케이션을 통해 제공되는 서비스를 포함한, 이메일 서비스

C. 기업 서비스

4.3. (본 강령의 범위 내에 있거나 제외된) 서비스 및 제품에 대한 리스트는 새로운 서비스와 제품이 마련될 가능성을 가지며, 이 중 일부는 본 강령과 관련이 있거나 일부는 없을 수 있다는 점에서 완전하지 않다. 서명인은 본 강령의 7.1절, 7.3절과 7.4절에 따라, 본 강령에 적용되는 관련 제품과 서비스를 식별할 것이다.

4.4. 허위조작정보가 아닌 콘텐츠Content that is not Misinformation: 다음 콘텐츠는 본 강령에 따른 허위조작정보가 아니다.

A. (기자 및 패러디를 포함한) 오락 또는 교육의 목적으로, 선의의 신뢰 하에 제작된 콘텐츠

B. 호주 주State 또는 연방 정부Federal Government에서 승인한 콘텐츠

C. 5.23절부터 5.25절에 따라, 정치 광고 또는 (호주 법률에 따라 등록된) 정당이 인가한 콘텐츠

D. 전문 뉴스 콘텐츠

본 4.4절의 A부터 D에 명시된 콘텐츠는 부정행위로 전파되는 경우, 허위조작정보 정의에 해당될 수 있다.

- 4.5. 서명인은 재량에 따라, 플랫폼에서 사용자의 (4.4절에 따라, 오정보에 관한 강령 조항에서 제외된 유형의) 콘텐츠 유포를 규제하는 정책과 절차를 이행할 수 있으며, 이는 서명인이 해당 콘텐츠가 해를 끼칠 가능성이 있다고 판단할 경우에 해당한다.
- 4.6. *기존 법률의 적용Application of existing laws.* 본 강령에서 다루는 일부 사안과 중복될 수 있는 (흡연, 의약품, 온라인 도박, 선거광고 그리고 명예훼손 등) 다양한 사안에 관한 금지 또는 제한뿐만 아니라, (온라인안전법 2021Online Safety Act 2021과 같은) 기존 법률이나 규정협약이 있다. 본 강령과 상충하는 범위 내에선, 해당 법률과 규정이 우선한다.
- 4.7. *적용Application.* 각 서명인이 약속한 준칙은 해당 준칙을 채택한 날로부터, 각 서명인이 지정한 제품과 서비스에 적용된다.
- 4.8. *착수Commencement.* 본 강령은 2021년 2월 22일에 착수되었다. 강령의 개정은 (새로운 거버넌스 협의를 반영을 목적으로) 2021년 10월 11일, (본 강령의 첫 번째 검토 결과를 반영을 목적으로) 2022년 12월 20일에 이루어졌다.

5. 목표와 대책 Objectives and Measures

- 5.1. *일반General:* 이 절은 위의 2항에 명시된 목적과 지침 원칙에 따라, 강령의 목표와

결과를 달성하기 위한 다양한 대책을 포함하고 있다.

- 5.2 *서명인 준칙*(Signatories Commitments): 본 강령의 핵심 목표 1을 준수하여, 모든 서명인은 (결과 1a 명시된 바와 같이) 디지털 플랫폼에서 허위조작정보와 오정보의 확산으로 인해 발생할 수 있는 피해의 위험을 줄이는 데 기여하고, 7절에 제시된 투명성 보고서를 제공하도록 한다. 모든 목표와 결과가 서명인에게 적용되는 것은 아니며, (6절의 지침에 따라) 서로 다른 서비스와 제품에 적합하고 적절한 방식으로, 5절에 명시된 대책 중 하나 이상을 채택할 수 있다. 서명자들은 강령에 따라 시행된 대책이 그들의 서비스와 제품, 기술 발전과 정보 환경의 변화를 반영하기 위해 점진적으로 발전할 수 있음을 유념하도록 한다.
- 5.3. 사전 동의(*Opt-in*): 다음 7.1절에서는 서명인이 준칙을 택하는 방법을 간략하게 설명한다.
- 5.4 *대책에 대한 용어*(Terminology of measures): 강령에 따른 대책을 시행함에 있어서, 서명인은 (정책과 프로세스의 시행을 포함하여) 결과를 달성하기 위해 이행되는 모든 대책에 “허위조작정보”와 “오정보” 이외의 용어를 사용할 수 있음을 인지한다. 예컨대 (사용자의 신상, 출처나 의도에 관해 허위나 오도된 표현을 하는) 규제된 사용자 행위나 행동, 또는 (오도되고, 기만적이고, 위험하거나 해로운 내용의) 규제된 콘텐츠의 용어로 지시할 수 있다.
- 5.5. *명료한 언어*(Plain language): 서명인은 (허위조작정보와 오정보에 대처하기 위해) 정책, 프로세스 그리고 관련 커뮤니티 지침 또는 추가적인 정보를 게시하기로 약속한 경우, 명료한 언어와 접근하기 쉬운 사용자 친화적인 형식으로 합당한 상업적 노력을 기울여야 한다.
- 5.6 *합법적인 콘텐츠 또는 사용자 접근에 대한 규제*(Restrictions on lawful content or users' access): 서명인은 본 강령의 요건을 준수하기 위해 오도되거나 기만적이거나 허위

일 수 있다는 이유만으로, (선택할 수 있지만) 합법적인 콘텐츠에 대한 접근을 제한하는 대책을 요구할 수 없다. 또한 서명인은 사용자가 업로드하고 공유하는 콘텐츠의 정확성을 알릴 의무가 없다.

5.7 정보 공개의 위험에 균형을 맞춘 투명성의 필요성 *Need for transparency to be balanced against disclosure risks.* (본 강령에 따라 이행된 대책의 공적 투명성을 증진하기 위해) 서명인은 준칙을 수행함에 있어서, 대책이 공개되어야 할 필요성 그리고 정보 공개가 허위조작정보와 오정보를 확산하는 행위의 증가나 바이럴리티의 증가를 초래할 수 있는 위험성 간의 균형을 맞출 필요가 있음을 인지한다.

목표 1: 허위조작정보와 오정보로 인해 발생할 수 있는 피해에 대한 안전장치를 제공한다.

결과 1a: 서명인들은 확장 가능한 다양한 대책을 채택함으로써, 디지털 플랫폼에서 허위 조작정보와 오정보의 확산으로 인해 발생할 수 있는 피해의 위험을 줄이는데 기여한다.

5.8. 서명인은 디지털 플랫폼 사용자의 허위조작정보와 오정보의 확산과 잠재적인 위험의 노출을 줄이는 것을 목표로, 대책을 마련하고 시행한다.

5.9. 5.8절에 따라 시행되는 대책에는 (이에 국한되지 않지만) 다음과 같은 예시가 포함된다.

A. 디지털 플랫폼에서 이용 가능한 사용자 행위나 콘텐츠에 대한 인적 검토가 필요한 정책과 프로세스 (팩트체크 기관과 협력하여 수행하는 검토 프로세스 포함)

- B. 허위 콘텐츠에 라벨을 부여하거나, 사용자에게 콘텐츠의 신뢰 지표를 제공함
- C. 사용자에게 허위조작정보와 오정보를 노출시킬 수 있는 콘텐츠 순위를 강등
- D. 부정 행위로 확산되는 콘텐츠 삭제함
- E. 허위조작정보와 오정보를 해결하기 위한 대책의 투명성을, 공중 또는 사용자에게 적절히 제공함
- F. 부정 행위를 하는 사용자의 계정을 정지하거나 비활성화함
- G. (콘텐츠나 사용자 계정에 대한 알고리즘 검토 등) 사용자에게 허위조작정보에 노출시킬 수 있는 부정 행위를 식별하고 줄이기 위한 기술을 제공하거나 사용함
- H. 디지털 플랫폼 또는 사용자가 신뢰성이나 정확성을 확인하거나 디지털 콘텐츠의 유래 또는 출처를 식별하는 데 도움이 되는 기술을 제공하거나 사용함
- I. 사용자에게 콘텐츠의 출처에 관한 메타데이터를 제공함
- J. 게시된 편집 정책 또는 콘텐츠 표준을 시행함
- K. 게시된 편집 강령에 따라, 신뢰할 수 있는 뉴스 출처에 우선순위 부여함 (일부 서명인은 4.5절에 따라 정책을 위반하는 뉴스 콘텐츠의 순위를 삭제하거나, 낮출 수 있는 선택을 할 수 있다는 점에 유념해야 함)
- L. 디지털 콘텐츠를 검토하기 위해 팩트체커와 협력하거나, 재원을 지원함
- M. 사용자가 특정 유형의 디지털 콘텐츠에 대한 접근을 제한할 수 있도록 도구를 제공함

결과 2b: 본 강령에 따라, 사용자는 서명인이 규제하거나 관리하게 될 행위 또는 콘텐

츠 유형에 대한 정보를 제공받게 된다.

- 5.10. 서명인은 서비스나 제품을 통해 허위조작정보와 오정보를 확산시킬 수 있는, 사용자 행위 또는 콘텐츠의 규제와 관리와 관련한 정책과 절차 그리고 적절한 지침 또는 정보를 구현하고 발표한다.

결과 3b: 사용자는 공개적으로 이용 가능하고 접근 가능한 리포팅 툴reporting tools를 통해, (5.10절에 따라) 정책을 위반하는 콘텐츠 또는 행동을 서명인에게 신고할 수 있다.

- 5.11. (5.10절에 따라) 서명인은 사용자가 정책을 위반하는 행위 또는 콘텐츠 유형을 신고할 수 있는 정책, 절차와 적절한 지침을 구현하고 출판한다.

- 5.12. 서명인은 (5.11절의 준칙을 이행함에 있어서) 허위조작정보와 오정보라는 용어가 사용자에게 익숙하지 않을 수 있음을 인지하고, 결과를 완수하기 위한 정책과 절차를 통해 (디지털 플랫폼에서 규제된) 다양한 콘텐츠와 행위를 신고하는 방법을 명기해야 함을 인정한다.

결과 1d: 사용자는 5.11절에 따라 제출된 신고에 응하여, 서명인의 대책에 관한 일반 정보에 접근할 수 있다.

- 5.13. 서명인은 (허위조작정보와 오정보에 해당하는 플랫폼 콘텐츠를 포함하되, 여기에 국한되지 않는) 플랫폼 정책을 위반하는 콘텐츠의 탐지와 삭제에 관한 정책, 절차 또는 (5.11절에 따른 사용자 신고에 대한 요약에 포함한) 종합보고aggregated reports를 이행하고 공개해야 한다.

결과 1e: 사용자는 추천 시스템 이용에 관한 서명인의 일반 정보에 접근하고, 추천 시스템이 제안하는 콘텐츠와 관련한 옵션을 가질 수 있다.

5.14. 공중에게 정보를 유포하는 것이 주요 목적이면서, (검색 엔진 외의) 추천 시스템을 사용한 서비스를 제공하는 서명자는 다음을 준수한다.

A. (이러한 서비스에 접근할 수 있는) 정보 우선순위를 정하는 방식에 관한 정보를 최종 사용자에게 제공한다.

B. 최종 사용자에게 (서비스에 적합한) 추천 시스템에서 제안하는 콘텐츠와 관련 옵션을 제공한다.

참고: 예컨대, (온라인 신문에서 게시된) 뉴스 기사 아래 제공되는 댓글 세션은, 편집자의 책임 하에 뉴스를 출판하는 주요 서비스와 달리 부수적인 부분이기 때문에 준칙의 적용을 받지 않는다.

목표 2: 허위조작정보와 오정보에 대한 광고 또는 수익창출 인센티브 incentives를 차단한다.

결과 2: 허위조작정보와 오정보에 대한 광고나 수익창출 인센티브가 감소한다.

5.15. 디지털 광고 서비스를 제공하는 서명인은, 광고주가 허위조작정보와 오정보를 확산하는 디지털 광고를 반복적으로 게재하는 것을 막기 위해 상업적으로 타당한 노력해야 한다.

5.16. 서명인은 허위조작정보와 오정보에 대한 광고 또는 수익 창출 인센티브 차단을 목표로 하는, 정책과 프로세스를 이행한다.

5.17. 예컨대, 5.16절에 따라 구현된 정책과 프로세스는 다음 사항이 포함된다.

- A. 브랜드 보호와 검증 도구의 이용을 증진하고 포함함.
- B. 제3자의 증면된 단체의 협력을 허용함.
- C. 광고주가 미디어 구매 전략 및 온라인 평판 리스크를 평가할 수 있도록 지원하고 허용함.
- D. 광고주가 광고 배치를 관리·감독하고 광고 위치를 선택할 수 있도록, 고객 특화 계정client-specific accounts에 필요한 접근 권한을 광고주에게 제공함.
- E. 허위조작정보와 오정보를 확산하는 계정과 웹사이트에 대한 광고 서비스와 유료 배치의 유효성을 제한.

5.18. 서명인은 (온라인 광고의 매매와 광고 관련) 서비스 제공과 관련된 모든 당사자들이 협력하여 온라인 광고 생태계 전반의 투명성을 개선하고, 이에 따라 허위조작 정보를 확산하는 계정과 웹사이트에 대한 광고 배치를 효과적으로 검토하며, 규제 및 제한해야함을 인지한다.

목표 3: 디지털 플랫폼에서 제공하는 서비스와 제품의 완전성과 보안을 보장하기 위해 노력한다.

결과 3: 사용자의 부정 행위로 인한 서비스와 제품의 완전성과 보안이 훼손될 위험이 감소한다.

5.19. 서명인은 (허위조작정보를 확산시키도록 고안된 허위계정 또는 자동화된 봇의 사용 등) 서비스와 제품의 완전성과 보안을 훼손하는 사용자 행위 유형을 규제하거나 관리하는 대책을 세우도록 한다.

5.20. 개인정보 보호에 대한 일부 사용자와 디지털 플랫폼의 기대에 부응하기 위해, 해당 준칙에 따라 마련되고 이행되는 대책은 익명의 계정 생성을 배제해서는 안 된다.

목표 4: 소비자가 정보에 기반하여 디지털 콘텐츠를 선택할 수 있도록 역량을 강화한다.

결과 4: 사용자는 (디지털 플랫폼을 통해 접근할 수 있는) 뉴스와 사실 기반 콘텐츠의 출처에 대해 더 많은 정보에 기반한 선택을 할 수 있으며, 오정보를 식별하는데 능숙해진다.

5.21. 서명인은 사용자가 디지털 콘텐츠에 대해 정보에 기반한 선택을 하고, 대안적인 정보 출처에 접근할 수 있도록 대책을 취한다.

5.22. (5.21절에 명시된 준칙에 따라) 마련되고, 시행되는 대책은 다음과 같은 예시가 포함된다.

- A. 사용자가 공적 이익에 대한 다양한 관점을 쉽게 찾을 수 있도록, 디지털 콘텐츠의 우선순위를 정하거나 순위를 지정하는 기술적 수단을 사용함
- B. 독립적인 편집강령과 불만처리제도complaints scheme의 적용을 받는 뉴스 콘텐츠의 종합 또는 증진
- C. 뉴스 출처의 신뢰성을 알리거나 디지털 플랫폼이나 사용자가 온라인 뉴스 콘텐츠의 신뢰성 또는 정확성을 확인하거나, 유래이나 출처를 식별하는데 도움이 되는 기술의 제공 또는 사용;
- D. 증거자료나 전문가 분석을 토대로, 디지털 리터러시 간여의 증진

E. 팩트체킹 기관에 대한 재정적 지원 또는 지속 가능한 파트너십의 제공

목표 5: 디지털 플랫폼에서 전달되는 정치 광고의 출처에 대한 공중의 인식을 개선한다.

결과 5: 사용자는 정치 광고의 출처에 대해 더 잘 알게 된다.

5.23. 본 강령의 목적상 정치 광고는 오정보는 아니지만, 서명인은 디지털 플랫폼에서 전달되는 정치 광고의 출처에 대해, 사용자에게 더 높은 투명성을 제공하는 정책을 마련하고 시행한다.

5.24. 5.23절의 준칙에 따라 마련되고 시행되는 대책은, 광고주가 디지털 플랫폼에서 전달되는 정치 광고의 출처를 식별하고 확인해야 하는 요구사항; 광고주나 광고의 출처에 대한 중요한 정보를 왜곡, 기만 또는 은폐하는 광고를 규제하는 정책; 정치 광고가 사용자 자신을 표적 삼았는지 그 여부를 파악할 수 있는 도구의 제공; 정치 광고가 (뉴스나 편집 콘텐츠를 포함한) 매체에 나타나는 경우, 정치 광고가 유료 커뮤니케이션으로 쉽게 인식할 수 있도록 요청하는 정책이 포함될 수 있다.

5.25. 서명인은 또한 정책적인 측면에서, 사용자의 정치적 배경을 유추하여 광고의 표적으로 삼지 않거나, (공적 관심사인 사회적 사안에 대한 정치적 결과를 지지하는 광고를 포함하여) 광범위한 정치 광고에 관한 준칙을 정의하고, 수행하도록 선택할 수 있다.

목표 6: 전략적 연구 지원을 통해 허위조작정보와 오정보에 대한 공중의 이해를 강화한다.

결과 6: 서명인은 허위조작정보와 오정보에 대한 공중의 이해를 향상시키기 위해 독립적인 연구자의 노력을 지지한다.

- 5.26. 서명인은 온라인과 오프라인에서 허위조작정보와 오정보를 연구하기 위한 독립적인 선의의 노력을 지지하고 장려할 것을 약속한다. 선의의 연구는 (연구자가 수집한 데이터를 연구 목적으로만 사용하고, 대학교 IT 시스템에 안전하게 저장할 것을 요구하는) 호주 내 공인된 대학교의 윤리 정책에 따라 수행되는 연구, 또는 디지털 플랫폼의 사전 서면 계약에 따라 수행되는 모든 연구를 포함한다.
- 5.27. 5.26절을 이행하기 위해 취하는 대책은, 예컨대 독립적인 팩트체크 기관의 관련 이니셔티브에 대한 협력이 포함될 수 있다. 다른 조치로는 연구를 위한 자금 지원 또는 데이터세트 공유, 공동 연구 수행, 학계와 시민 사회 단체와의 협력이 포함될 수 있다.
- 5.28. 서명인은 5.26절에 명시된 바와 같이, 플랫폼에서 허위조작정보 또는 오정보에 대한 선의의 연구를 규제하거나 저해하지 않기로 약속한다.
- 5.29. 관련된 서명인은 학계와 시민 사회 안에서 허위조작정보와 오정보에 관한 논의를 촉진하기 위해 연례행사를 개최하기로 약속한다.

참고: 연례행사는 모든 형식(온라인, 오프라인 또는 결합하는 형식)을 통해 수행될 수 있으며, 강령 또는 강령의 주제와 관련된 모든 주제를 다룰 수 있다. 예컨대, 이 행사는 서명인 또는 강령 관리자의 연구 이니셔티브 발표 또는 토론을 위해 소집될 수 있다.

목표 7: 서명인은 허위조작정보와 오정보에 대응하기 위한 대책을 공개한다.

결과 7: 공중은 허위조작정보와 오정보에 대응하기 위해, 서명인이 이해한 대책에 대한

정보에 접근할 수 있다.

- 5.30. 모든 서명인은 7절에 명시된 투명성 보고서 정보를 작성하고 출판한다.(사용자 기반이 호주 내 월간 활성 최종 사용자가 100만 명을 초과하는, 사용자 기반의 대규모 플랫폼에 대한 보고서 양식은 7.3절을 참조하고, 호주 내 월간 활성 최종 사용자가 100만 명 미만인, 사용자 기반의 소규모 플랫폼 보고서 양식은 7.4절을 참조하라.)
- 5.31. 또한, 서명인은 목표 1과 관련된 진행 상황과 본 강령에 따른 추가 준칙을 상세히 설명하는 추가 정보를 공개한다.
- 5.32. 서명인은 콘텐츠 삭제, 오픈 데이터 이니셔티브, 연구 보고서, 언론 발표, 사용자 데이터 요청, 비즈니스 투명성 보고서와 같은 영역에 대한 추가 보고서와 공개 업데이트를 제공함으로써, 5.31절의 준칙을 이행할 수 있다. 다음의 정보는 블로그 게시물, 백서, 제품 내 공지사항, 투명성 보고서, 도움말 센터 또는 기타 웹사이트 등 (이에서 국한될 수 없지만)을 예시로 들 수 있다.

6. 플랫폼별 대책에 대한 지침 Guidance on platform-specific measures

- 6.1. 강령에 따른 대책의 비례원칙 Proportionality of measures under Code: 본 강령에 따라 서명 기인이 취하는 대책은 허위조작정보와 오정보로 인한 피해를 포함하여, 특정 맥락에 비례하며 관련성이 있어야 한다. 서명인은 대책의 적절성을 평가할 때 다음과 같은 다양한 요소를 고려할 수 있다:

A. 허위조작정보와 오정보를 확산하는 행위자

- B. 허위조작정보와 오정보를 확산하는 사용자의 행동 특성(예컨대, 해당 행동이 자동화되고 의도적인지, 또는 악의적인 동기가 있는지, 조직적이고 지속적으로 대규모로 이루어지는지에 대한 여부)
- C. 콘텐츠가 유포되는 제품 또는 서비스의 유형과 플랫폼 사용자들 사이에서 콘텐츠가 광범위하고 빠르게 공유되는 네트워크 효과가 있는지에 대한 여부
- D. 콘텐츠의 확산을 통해 플랫폼이 상업적 이익을 얻을 수 있는지 여부(예컨대, 콘텐츠가 스폰서 콘텐츠인지 여부)
- E. 관련 디지털 콘텐츠의 허위 여부를 권위 있고 신뢰할 수 있는 출처를 통해 합리적으로 검증할 수 있는 정도
- F. 콘텐츠의 확산으로 인해, 합리적으로 발생할 가능성이 있는 피해의 근접성과 심각성
- H. 디지털 플랫폼의 비즈니스 규모와 특성 그리고 해당 플랫폼의 가용할 수 있는 자원
- I. 다른 인권과 균형을 맞추기 위해 표현의 자유를 보호해야 할 필요성
- J. 사용자의 개인정보를 보호해야 할 필요성

7. 강령 관리 code administration

- 7.1. 사전 동의(Opt-in). 본 강령은 디지털 플랫폼의 비즈니스 모델과 제품 제공의 다양성을 인정하며, 다양한 비즈니스가 사전 동의(Opt-in) 조약을 통해 준칙을 이행할 수 있도록 고안되었다. 서명인은 강령에 서명한 후 3개월 이내에, 부록 1의 사전 동의 지명 양식을 사용하여 약속할 조항을 지명해야 한다. 서명인은 지명하지 않은

약속을 준수할 의무는 없다.

각 서명인은 매년 본 강령의 조항이 (새로운 제품과 서비스가 본 강령의 적용을 받아야 하는지 여부 포함여) 자사의 제품 및 서비스와 관련이 있는지 재평가하고, 사전 동의 양식에 대한 업데이트가 있을 경우 DIGI에 통지해야 한다. DIGI는 사전 동의 지명 양식에 대한 업데이트를 DIGI 웹사이트에 게시할 것이다.

7.2. *강령 탈퇴Withdrawal from Code*. 서명자는 DIGI에 통지함으로써, 본 강령 또는 본 강령에 따른 특정 준칙을 철회 및 탈퇴 할 수 있다.

7.3. *대규모 디지털 플랫폼인 서명인에 대한 연례 보고Annual Reporting for Signatories that are large Digital Platforms*: 7.1절에 따른 사전 동의 지명 양식에 덧붙여, (호주 내 월간 활성 사용자 수가 100만 명을 초과한, 사용자 기반 서비스를 제공하는) 각 서명자는 DIGI에 연례 보고서를 제출해야 하며, 해당 보고서에는 강령에 명시된 결과를 완수하기 위한 진행상황이 설명되어야 한다. 이 보고서는 DIGI 웹사이트에 게시된다. 각 서명인의 연례 보고서에는 보고서가 다루는 기간 동안, (강령에 따른 추가 제품과 서비스를 포함하여) 강령에 기반하여 제품과 서비스를 기술한다.

보고서는 부록 2에 제공된 템플릿 보고서의 양식을 토대로 하며, 전년도 기간을 포함하여 매년 5월 30일까지 제출해야한다.

7.4. *소규모 디지털 플랫폼인 서명인에 대한 보고Reporting for Signatories that are smaller Digital platforms*. 7.1항에 따른 사전 동의 지명 양식에 덧붙여, (호주 내 월간 활성 사용자 수가 100만 명 미만인, 사용자 기반 서비스를 제공하는) 각 서명자는 DIGI에 초안을 제출해야 하며, 해당 보고서에는 강령에 명시된 결과를 완수하기 위한 진행상황이 설명되어야 한다. 이 보고서는 DIGI 웹사이트에 게시된다. 이 보고서는 (반드시 준수해야 할 필요는 없지만) 부록 2의 템플릿을 기반으로 하며, 서명인이 본 강령에 가입한 후 12개월 이내에 제출해야 한다. (본 강령의 준칙 대상인

새로운 제품 및 서비스의 추가, 또는 본 강령 5.8절, 5.10절 또는 5.11절에 따른 준칙에 중대한 영향을 미칠 수 있는 대책, 정책, 프로세스 또는 지침 범위의 변경 등) 보고서에 중대한 변경이 있는 경우 서명인이 매년 업데이트해야 한다.

7.5 불만 사항 *Complaints*: 서명인은 강령에 따른 서명인의 준칙 위반 가능성에 대한 공중의 불만사항을 해소하기 위해 불만 처리 제도를 설치했다. 공중은 DIGI 웹사이트의 불만 처리 포털을 통해 이 제도에 접근할 수 있다. 독립적인 불만처리위원회는 불만처리 제도의 운영 및 강령 거버넌스에 대한 정보와 함께, DIGI 웹사이트에 게시된 참고 조항에 따라 서명인의 강령 준수에 대한 불만을 해소한다. 불만처리 제도는 서명인의 제품 또는 서비스에 대한 개별 콘텐츠 항목에 대한 불만을 접수하지 않으며, 이러한 불만은 해당 서명자의 보고 메커니즘을 통해 해당 서명인에게 전달되어야 한다.

7.6. 이 과정의 일환으로, 서명인은 정부와 관련 규제 기관과의 현행 협정을 활용할 수 있는 방법을 고려하여, 허위조작정보를 확산하고, 본 강령에서 다루는 조치 대상이 되는 부정 행위 사례를 식별하고 해결할 수 있다.

7.7. *강령 관리* *Code Administration*: 본 강령의 관리자는 DIGI이며, DIGI는 서명인과 독립적인 구성원의 대표로 구성된 관리 소위원회를 설치하였다. 소위원회는 6개월 마다 서명인의 조치를 검토하고, 강령에 따른 준칙을 어떻게 이행하고 있는지 관리·감독한다. 아울러, DIGI는 모범 사례 및 규정 준수를 장려할 목적으로, 독립 전문가를 임명하여 강령에 따라 서명인의 연례 투명성 보고서를 팩트체크하고 검증한다.

7.7. *강령 검토* *Code Review*: 본 강령은 2년 주기로 검토된다. 검토는 서명인과 (호주미디어 어청 *Australian Communications and Media Authority*를 포함한) 관련 정부 기관, 그리고 이 분야에서 활동하는 학계와 시민 사회 대표를 포함한 기타 이해관계자의 참여를 기반으로 진행된다. 강령 개정안은 DIGI 웹사이트에 공개된다.

[부록3] 호주 실천강령 부록 : 사전 동의 지명 양식

사전 동의 지명 양식 opt-in Nomination Form

이 양식은 서명인이 본 강령에 따른 준칙을 명시하기 위해 사용된다. 서명인은 7.1절에 따라, 임의적으로 준칙의 일부 또는 전부를 선택하지 않은 이유에 대한 간략한 설명을 포함해야 한다.

목표 Objective	결과 Outcome	대책 Measure		Opt-in · Opt-out
목표 1: 허위조작정보와 오정보로 인해 발생할 수 있는 피해에 대한 안전장치를 제공한다.	결과 1a: 서명인들은 확장 가능한 다양한 대책을 채택함으로써, 디지털 플랫폼에서 허위조작정보와 오정보의 확산으로 인해 발생할 수 있는 피해의 위험을 줄이는데 기여한다.	5.8	서명인은 디지털 플랫폼 사용자의 허위조작정보와 오정보의 확산과 잠재적인 위험의 노출을 줄이는 것을 목표로, 대책을 마련하고 시행한다.	YES
		5.9	5.8절에 따라 시행되는 대책에는 (이에 국한되지 않지만) 다음과 같은 예시가 포함된다. a) 디지털 플랫폼에서 이용 가능한 사용자 행위나 콘텐츠에 대한 인적 검토가 필요한 정책과 프로세스 (팩트체크 기관과 협력하여 수행하는 검토 프로세스 포함) b) 허위 콘텐츠에 라벨을 부여하거나, 사용자에게 콘텐츠의 신뢰 지표를 제공함 c) 사용자에게 허위조작정보와 오정보를 노출시킬 수 있는 콘텐츠 순위를 강등함 d) 부정 행위로 확산되는 콘텐츠 삭제함	

<p>결과 2b: 본 강령에 따라, 사용자는 서명인이 규제하거나 관리하게 될 행위 또는 콘텐츠 유형에 대한 정보를 제공받게 된다.</p>	<p>5.10</p>	<p>서명인은 서비스나 제품을 통해 허위조작정보와 오정보를 확산시킬 수 있는, 사용자 행위 또는 콘텐츠의 규제와 관리와 관련한 정책과 프로세스 그리고 적절한 지침 또는 정보를 구현하고 발표한다.</p>	
<p>결과 3b: 사용자는 공개적으로 이용 가능하고 접근 가능한 리포팅 툴reporting tools를 통해, (5.10절에 따라) 정책을 위반하는 콘텐츠 또는 행동을 서명인에게 신고할 수 있다.</p>	<p>5.11</p>	<p>(5.10절에 따라) 서명인은 사용자가 정책을 위반하는 행위 또는 콘텐츠 유형을 신고할 수 있는 정책, 프로세스와 적절한 지침을 구현하고 출판한다.</p>	
<p>결과 1d: 사용자는 5.11절에 따라 제출된 신고에 응하여, 서명인의 대책에 관한 일반 정보에 접근할 수 있다.</p>	<p>5.12</p>	<p>서명인은 (5.11절의 준칙을 이행함에 있어서) 허위조작정보와 오정보라는 용어가 사용자에게 익숙하지 않을 수 있음을 인지하고, 결과를 완수하기 위한 정책과 프로세스를 통해 (디지털 플랫폼에서 규제된) 다양한 콘텐츠와 행위를 신고하는 방법을 명기해야 함을 인정한다.</p>	
<p>결과 1d: 사용자는 5.11절에 따라 제출된 신고에 응하여, 서명인의 대책에 관한 일반 정보에 접근할 수 있다.</p>	<p>5.13</p>	<p>서명인은 (허위조작정보와 오정보에 해당하는 플랫폼 콘텐츠를 포함하되, 여기에 국한되지 않는) 플랫폼 정책을 위반하는 콘텐츠의 탐지와 삭제에 관한 정책, 프로세스 또는 (5.11절에 따른 사용자 신고에 대한 요약을 포함한) 종합보고aggregated reports를 이행하고 공개해야 한다.</p>	

		<p>e)허위조작정보와 오정보를 해결하기 위한 대책의 투명성을, 공중 또는 사용자에게 적절히 제공함</p> <p>f) 부정 행위를 하는 사용자의 계정을 정지하거나 비활성화함</p> <p>g)(콘텐츠나 사용자 계정에 대한 알고리즘 검토 등) 사용자에게 허위조작정보에 노출시킬 수 있는 부정 행위를 식별하고 줄이기 위한 기술을 제공하거나 사용함</p> <p>h)디지털 플랫폼 또는 사용자가 신뢰성이나 정확성을 확인하거나 디지털 콘텐츠의 유래 또는 출처를 식별하는 데 도움이 되는 기술을 제공하거나 사용함</p> <p>i)사용자에게 콘텐츠의 출처에 관한 메타데이터를 제공함</p> <p>j)게시된 편집 정책 또는 콘텐츠 표준을 시행함</p> <p>k)게시된 편집 강령에 따라, 신뢰할 수 있는 뉴스 출처에 우선순위 부여함 (일부 서명인은 4.5절에 따라 정책을 위반하는 뉴스 콘텐츠의 순위를 삭제하거나, 낮출 수 있는 선택을 할 수 있다는 점에 유념해야 함)</p> <p>l)디지털 콘텐츠를 검토하기 위해 팩트체커와 협력하거나, 재원을 지원함</p> <p>m)사용자가 특정 유형의 디지털 콘텐츠에 대한 접근을 제한할 수 있도록 도구를 제공함</p>
--	--	---

	<p>결과 1e: 사용자는 추천 시스템 이용에 관한 서명인의 일반 정보에 접근하고, 추천 시스템이 제안하는 콘텐츠와 관련된 옵션을 가질 수 있다.</p>	5.14	<p>공중에게 정보를 유포하는 것이 주요 목적이면서, (검색 엔진 외의) 추천 시스템을 사용한 서비스를 제공하는 서명자는 다음을 준수한다.</p> <p>A. (이러한 서비스에 접근할 수 있는) 정보 우선순위를 정하는 방식에 관한 정보를 최종 사용자에게 제공한다.</p> <p>B. 최종 사용자에게 (서비스에 적합한) 추천 시스템에서 제안하는 콘텐츠와 관련 옵션을 제공한다.</p>	
<p>목표 2: 허위조작정보와 오정보에 대한 광고 또는 수익창출 인센티브incentives를 차단한다.</p>	<p>결과 2: 허위조작정보와 오정보에 대한 광고나 수익창출 인센티브가 감소한다.</p>	5.15	<p>디지털 광고 서비스를 제공하는 서명인은, 광고주가 허위조작정보와 오정보를 확산하는 디지털 광고를 반복적으로 게재하는 것을 막기 위해 상업적으로 타당한 노력해야 한다.</p>	
		5.16	<p>서명인은 허위조작정보와 오정보에 대한 광고 또는 수익 창출 인센티브 차단을 목표로 하는, 정책과 프로세스를 이행한다.</p>	
		5.17	<p>예컨대, 5.16절에 따라 구현된 정책과 프로세스는 다음 사항이 포함된다.</p> <p>a) 브랜드 보호와 검증 도구의 이용을 증진하고 포함함.</p> <p>b) 제3자의 증면된 단체의 협력을 허용함.</p> <p>c) 광고주가 미디어 구매 전략 및 온라인 평판 리스크를 평가할 수 있도록 지원하고 허용함.</p> <p>d) 광고주가 광고 배치를 관리·감독하고 광고 위치를 선택할 수 있도록, 고객 특화 계정client-specific accounts에 필요한 접근 권한을 광고주에게 제공함.</p>	

		5.18	서명인은 (온라인 광고의 매매와 광고 관련) 서비스 제공과 관련된 모든 당사자들이 협력하여 온라인 광고 생태계 전반의 투명성을 개선하고, 이에 따라 허위조작정보를 확산하는 계정과 웹사이트에 대한 광고 배치를 효과적으로 검토하며, 규제 및 제한해야함을 인지한다.	
<p>목표 3: 디지털 플랫폼에서 제공하는 서비스와 제품의 완전성과 보안을 보장하기 위해 노력한다.</p>	<p>결과 3: 사용자의 부정 행위로 인한 서비스와 제품의 완전성과 보안이 훼손될 위험이 감소한다.</p>	5.19	서명인은 (허위조작정보를 확산시키도록 고안된 허위계정 또는 자동화된 봇의 사용 등) 서비스와 제품의 완전성과 보안을 훼손하는 사용자 행위 유형을 규제하거나 관리하는 대책을 세우도록 한다.	
		5.20	개인정보 보호에 대한 일부 사용자와 디지털 플랫폼의 기대에 부응하기 위해, 해당 준칙에 따라 마련되고 이행되는 대책은 익명의 계정 생성을 배제해서는 안 된다.	
<p>목표 4: 소비자가 정보에 기반하여 디지털 콘텐츠를 선택할 수 있도록 역량을 강화한다..</p>	<p>결과 4: 사용자는 (디지털 플랫폼을 통해 접근할 수 있는) 뉴스와 사실 기반 콘텐츠의 출처에 대해 더 많은 정보에 기반한 선택을 할 수 있으며, 오정보를 식별하는 데 능숙해진다.</p>	5.21	서명인은 사용자가 디지털 콘텐츠에 대해 정보에 기반한 선택을 하고, 대안적인 정보 출처에 접근할 수 있도록 대책을 취한다.	

		5.22	<p>(5.21절에 명시된 준칙에 따라) 마련되고, 시행되는 대책은 다음과 같은 예시가 포함된다.</p> <p>a) 사용자가 공적 이익에 대한 다양한 관점을 쉽게 찾을 수 있도록, 디지털 콘텐츠의 우선순위를 정하거나 순위를 지정하는 기술적 수단을 사용함</p> <p>b) 독립적인 편집강령과 불만처리제도 complaints scheme의 적용을 받는 뉴스 콘텐츠의 종합 또는 증진</p> <p>c) 뉴스 출처의 신뢰성을 알리거나 디지털 플랫폼이나 사용자가 온라인 뉴스 콘텐츠의 신뢰성 또는 정확성을 확인하거나, 유래이나 출처를 식별하는 데 도움이 되는 기술의 제공 또는 사용;</p> <p>d) 증거자료나 전문가 분석을 토대로, 디지털 리터러시 간여의 증진</p> <p>e) 팩트체크 기관에 대한 재정적 지원 또는 지속 가능한 파트너십의 제공</p>	
<p>목표 5: 디지털 플랫폼에서 전달되는 정치 광고의 출처에 대한 공중의 인식을 개선한다.</p>	<p>결과 5: 사용자는 정치 광고의 출처에 대해 더 잘 알게 된다.</p>	5.23	<p>본 강령의 목적상 정치 광고는 오정보는 아니지만, 서명인은 디지털 플랫폼에서 전달되는 정치 광고의 출처에 대해, 사용자에게 더 높은 투명성을 제공하는 정책을 마련하고 시행한다.</p>	

		5.24	5.23절의 준칙에 따라 마련되고 시행되는 대책은, 광고주가 디지털 플랫폼에서 전달되는 정치 광고의 출처를 식별하고 확인해야 하는 요구사항; 광고주나 광고의 출처에 대한 중요한 정보를 왜곡, 기만 또는 은폐하는 광고를 규제하는 정책; 정치 광고가 사용자 자신을 표적 삼았는지 그 여부를 파악할 수 있는 도구의 제공; 정치 광고가 (뉴스나 편집 콘텐츠를 포함한) 매체에 나타나는 경우, 정치 광고가 유료 커뮤니케이션으로 쉽게 인식할 수 있도록 요청하는 정책이 포함될 수 있다.
		5.25	서명인은 또한 정책적인 측면에서, 사용자의 정치적 배경을 유추하여 광고의 표적으로 삼지 않거나, (공적 관심사인 사회적 사안에 대한 정치적 결과를 지지하는 광고를 포함하여) 광범위한 정치 광고에 관한 준칙을 정의하고, 수행하도록 선택할 수 있다.
목표 6: 전략적 연구 지원을 통해 허위조작정보와 오정보에 대한 공중의 이해를 강화한다.	결과 6: 서명인은 허위조작정보와 오정보에 대한 공중의 이해를 향상시키기 위해 독립적인 연구자의 노력을 지지한다	5.26	서명인은 온라인과 오프라인에서 허위조작정보와 오정보를 연구하기 위한 독립적인 선의의 노력을 지지하고 장려할 것을 약속한다. 선의의 연구는 (연구자가 수집한 데이터를 연구 목적으로만 사용하고, 대학교 IT 시스템에 안전하게 저장할 것을 요구하는) 호주 내 공인된 대학교의 윤리 정책에 따라 수행되는 연구, 또는 디지털 플랫폼의 사전 서면 계약에 따라 수행되는 모든 연구를 포함한다.

		5.27	5.26절을 이행하기 위해 취하는 대책은, 예컨대 독립적인 팩트체크 기관의 관련 이니셔티브에 대한 협력이 포함될 수 있다. 다른 조치로는 연구를 위한 자금 지원 또는 데이터세트 공유, 공동 연구 수행, 학계와 시민 사회 단체와의 협력이 포함될 수 있다.	
		5.28	서명인은 5.26절에 명시된 바와 같이, 플랫폼에서 허위조작정보 또는 오정보에 대한 선의의 연구를 규제하거나 저해하지 않기로 약속한다.	
		5.29	관련된 서명인은 학계와 시민 사회 안에서 허위조작정보와 오정보에 관한 논의를 촉진하기 위해 연례행사를 개최하기로 약속한다.	
목표 7: 서명인은 허위조작정보와 오정보에 대응하기 위한 대책을 공개한다.	결과 7: 공중은 허위조작정보와 오정보에 대응하기 위해, 서명인이 이해한 대책에 대한 정보에 접근할 수 있다.	5.30	모든 서명인은 7절에 명시된 투명성 보고서 정보를 작성하고 출판한다. (사용자 기반이 호주 내 월간 활성 최종 사용자가 100만 명을 초과하는, 사용자 기반의 대규모 플랫폼에 대한 보고서 양식은 7.3절을 참조하고, 호주 내 월간 활성 최종 사용자가 100만 명 미만인, 사용자 기반의 소규모 플랫폼 보고서 양식은 7.4절을 참조하라.)	
		5.31	또한, 서명인은 목표 1과 관련된 진행 상황과 본 강령에 따른 추가 준칙을 상세히 설명하는 추가 정보를 공개한다.	

		<p>5.32</p> <p>서명인은 콘텐츠 삭제, 오픈 데이터 이니셔티브, 연구 보고서, 언론 발표, 사용자 데이터 요청, 비즈니스 투명성 보고서와 같은 영역에 대한 추가 보고서와 공개 업데이트를 제공함으로써, 5.31절의 준칙을 이행할 수 있다. 다음의 정보는 블로그 게시물, 백서, 제품 내 공지사항, 투명성 보고서, 도움말 센터 또는 기타 웹사이트 등을 (이에서 국한될 수 없지만) 예시로 들 수 있다.</p>	
--	--	--	--

저 자 소 개

이 수 업

서강대학교 신문방송학 학사
서강대학교 신문방송학 석사
서강대학교 신문방송학 박사
현 미디어미래연구소 연구위원

권 오 상

연세대학교 경영학 학사
연세대학교 경영학 석사
연세대학교 법학과 박사
현 미디어미래연구소 소장

박 지 은

한양대학교 신문방송 학사
한양대학교 신문방송 석사
현 미디어미래연구소 연구원

김 국 진

한국외국어대학교 신문방송학 학사
고려대학교 신문방송학 석사
고려대학교 방송학 박사
연세대학교 언론홍보대학원 겸임교수
현 미디어미래연구소 대표연구위원

최 민 정

성공회대 중어중국학 학사
현 미디어미래연구소 연구원

최 예 서

한양대학교 신문방송 학사
한양대학교 경제학 석사
현 미디어미래연구소 연구원

방통융합정책연구 KCC-2023-19

온라인 가짜뉴스 근절을 위한 사업자 자율규제 방안 연구

2023년 12월 31일 인쇄
2023년 12월 31일 발행

발행인 방송통신위원회 위원장
발행처 방송통신위원회
경기도 과천시 관문로 47
정부과천청사
TEL: 02-2110-1323
Homepage: www.kcc.go.kr
