

SW·저작권 동향리포트

<제2024-20호> 2024년 10월 25일

정책/제도

최근 국제 사회에서의 AI 정책 이슈

■ 개요

- 지난 8월 영국에서 발생한 흥기 난동에서 촉발된 극우파의 무력 폭동 사건이 AI가 생성한 허위정보에 의한 것으로 밝혀지면서 AI가 생성하는 여론조작과 허위정보에 대한 우려와 함께 이를 막기 위한 제도적 장치의 마련이 시급하다는 지적이 나오고 있음
- 이러한 AI의 부정적 측면이 부각되면서 국제 사회는 딥페이크와 환각 현상을 비롯하여 AI가 퇴화하는 모델 붕괴 현상에도 관심을 갖고 국제 기구와 빅테크 기업들을 중심으로 AI의 부정적 사용을 통제하고 올바르게 운영할 수 있는 정책을 마련하는데 주목

■ 주요 내용

1. AI의 부정적 측면에 따른 이슈

가. 구글, 생성형 AI의 악용 사례 분석

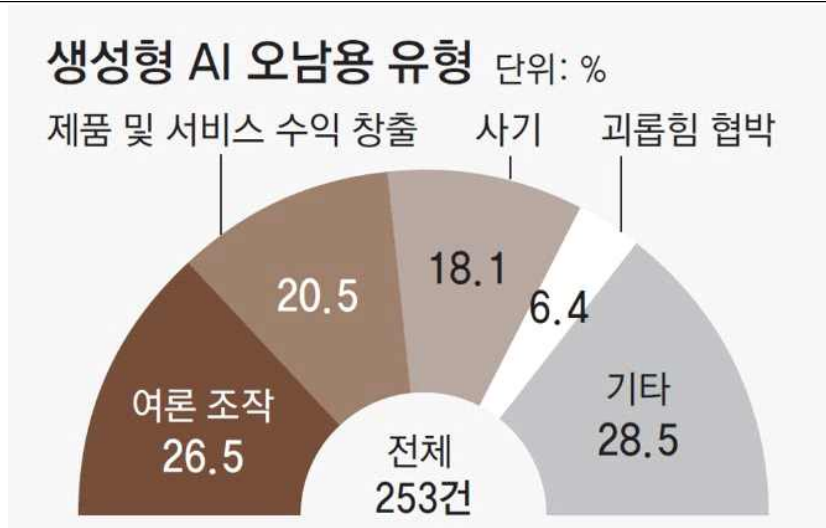
- 구글 딥마인드는 언론에 보도된 생성형 AI 악용 사례 253건을 분석해 악용되는 유형과 목적을 분석하고 대응 방안을 모색한 논문을 공개. 분석 결과 대다수의 악용 사례는 여론조작, 수익창출, 사기의 목적으로 나타났으며, 이 중 여론 조작이 26.5%를 차지하며 가장 많이 악용되고 있는 것으로 발표
- 조사 결과에 따르면 여론조작을 위한 오용 유형은 허위 정보와 명예훼손이 일반적으로 나타났으며 수익창출을 위한 오용 유형은 딥페이크 상품화와 콘텐츠의 대량제작, 사기 목적은 유명인 사칭 광고와 피싱 사기가 주류를

이루는 것으로 나타남. 특히 공인을 사칭하거나 가상의 캐릭터를 만들어 지지자나 반대자 행세를 하거나 대중의 인식을 왜곡하기 위해 조작된 정보를 제작하여 선거에 개입하는 등의 허위 정보를 제공하는 사례가 많은 것으로 나타남

* 일례로 카멀라 해리스 미국 부통령의 선거캠프 홍보 영상을 딥페이크로 편집해 “조 바이든 대통령이 노망났다”는 발언을 한 것처럼 조작하거나, 리시 수낙 총리가 가자지구에 18세 이상 영국 청년을 파병한다는 내용을 담은 AI 영상이 SNS를 떠도는 등 미국과 영국 선거에 영향

- 이번 연구는 생성형 AI가 악용된 실제 사례들을 분석해 악용하는 동기와 악용 방식 등을 분석했다는 점에서 의미 있다는 평가와 함께 생성형 AI 기술이 고도화되고 일상화 될수록 허위정보가 더 빠르게 양산될 수 있는 만큼 이를 막기 위한 제도적 장치 마련이 시급하다고 지적

[그림 1] 생성형 AI 오남용 유형



*출처 : 동아일보 (<https://www.donga.com/news/Economy/article/all/20240813/126514798/2>)

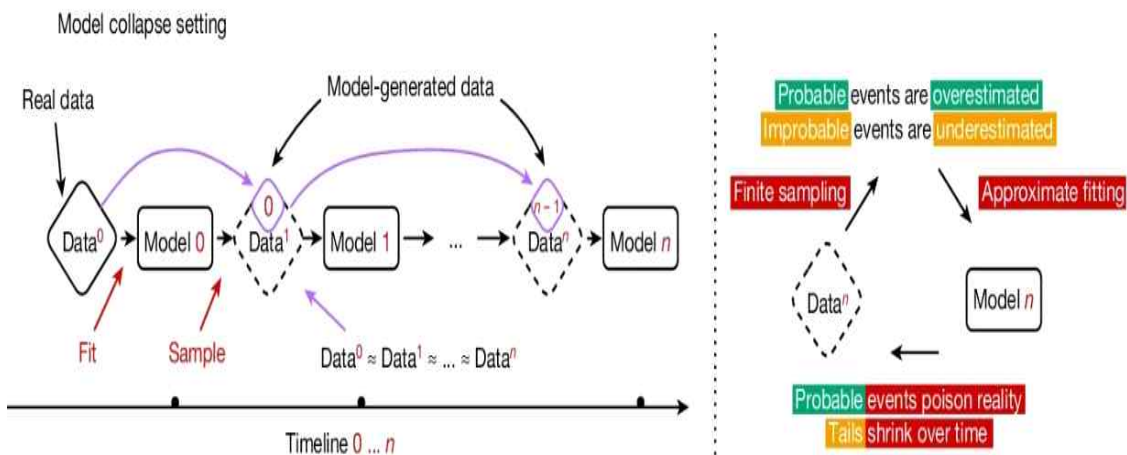
나. 네이처, AI 모델 붕괴 현상 발표

- 영국 옥스퍼드와 주요 대학 연구진이 네이처(Nature)에 발표한 연구에 따르면 AI로 만든 데이터로만 학습한 LLM은 결과물이 오염되는 모델 붕괴 (Model Collapse) 현상이 발생하는 것으로 나타남. 이는 AI가 만든 데이터를 AI가 학습해 결과물을 생성하고 이 데이터를 다시 AI가 학습해 결과물을 생성하도록 실험을 반복한 결과 후속 모델은 전혀 엉뚱한 결과물을 생성하며 모델 붕괴 현상을 일으킨다고 발표
- 연구진들은 이 현상의 원인을 데이터 샘플이 한정되어 있어 발생하는 통계적 근사 오류(Statistical Approximation Error), AI 모델이 현실 세계의 복잡성을 완벽히 표현하지 못해 발생하는 기능적 표현력 오류(Functional E

xpressivity Error), AI의 학습 과정 자체의 한계로 인한 오류인 기능적 근사 오류(Functional Approximation Error)에 의한 것으로 설명

- 오류가 세대를 거듭하며 반복되고 누적될수록 원본 데이터의 특성이 사라지고 점점 더 많은 오류가 추가되어 퇴행적 학습 과정을 겪게 되면서 AI는 점점 현실과 동떨어진 자신만의 세계를 창조하게 됨. 특히 AI 모델 붕괴 현상은 단순한 기술적 문제를 넘어 정보의 질 저하, AI 의존도 증가의 위험, 다양성 감소, 윤리적 문제 등 사회 전체에 큰 영향을 미칠 가능성이 존재
- 전문가들은 AI 학습에 사용되는 원본 데이터를 보존하고 AI가 생성한 데이터와 인간이 생성한 데이터를 구분할 수 있도록 데이터의 출처를 추적하며, AI 학습 데이터에 다양한 출처와 관점을 포함시켜야 한다고 설명. 또한 AI 모델의 성능과 출력을 지속적으로 검토하고 평가해야 하며, AI 개발과 사용에 대한 명확한 윤리적 가이드라인 마련도 필요

[그림 2] AI 모델 붕괴에 의한 퇴행적 학습



*출처 : 네이처 홈페이지 (<https://www.nature.com/articles/s41586-024-07566-y>)

다. 기업들의 환각과 딥페이크 이슈 대응

- 메타(Meta)의 메타AI 챗봇이 지난 7월 발생한 도널드 트럼프 전 대통령에 대한 암살 시도를 부정하는 잘못된 답변을 생성하여 논란. 또한 메타AI는 총격 직후 촬영된 트럼프 사진에 사실 확인 라벨을 잘못 적용하는 오류도 나타남
- 메타는 자사의 AI에서 오류가 나타난 이유에 대해 생성 AI에서 일반적으로 나타나는 환각 현상(Hallucination)이라고 해명하면서, 모든 생성 AI

시스템과 마찬가지로 자사의 모델도 부정확하거나 부적절한 결과를 생성할 수 있으며 해당 문제를 지속적으로 해결하고 기능을 개선할 계획으로 밝힘

- 구글은 딥페이크 문제 해결을 위해 검색 결과에서 딥페이크를 더 쉽게 삭제할 수 있도록 절차를 간소화하고, 검색 결과에서 딥페이크 콘텐츠가 상위에 표시되지 않도록 순위 시스템 업데이트를 실시하여 특정 딥페이크 콘텐츠나 사람 이름을 포함하는 검색어를 입력하면 관련 뉴스 기사 등 고품질 콘텐츠를 우선적으로 노출할 예정
- 이를 통해 구글의 사용자가 본인이 등장하는 딥페이크 이미지를 구글 검색 결과에서 발견하고 삭제를 요청하면 구글 시스템은 해당 사용자 관련 유사한 검색에서 모든 노골적인 결과를 필터링하며, 정책에 따라 검색 결과에서 관련 이미지가 삭제되면 해당 이미지의 모든 복제본도 스캔하여 삭제. 또한 허위 음란물 이미지로 인해 삭제된 횟수가 많은 사이트의 경우 해당 사이트의 검색 순위를 낮추어 사용자에게 대한 표시를 줄어나갈 계획

기타 이슈

<EU, 구글과 삼성의 반독점 조사 검토>

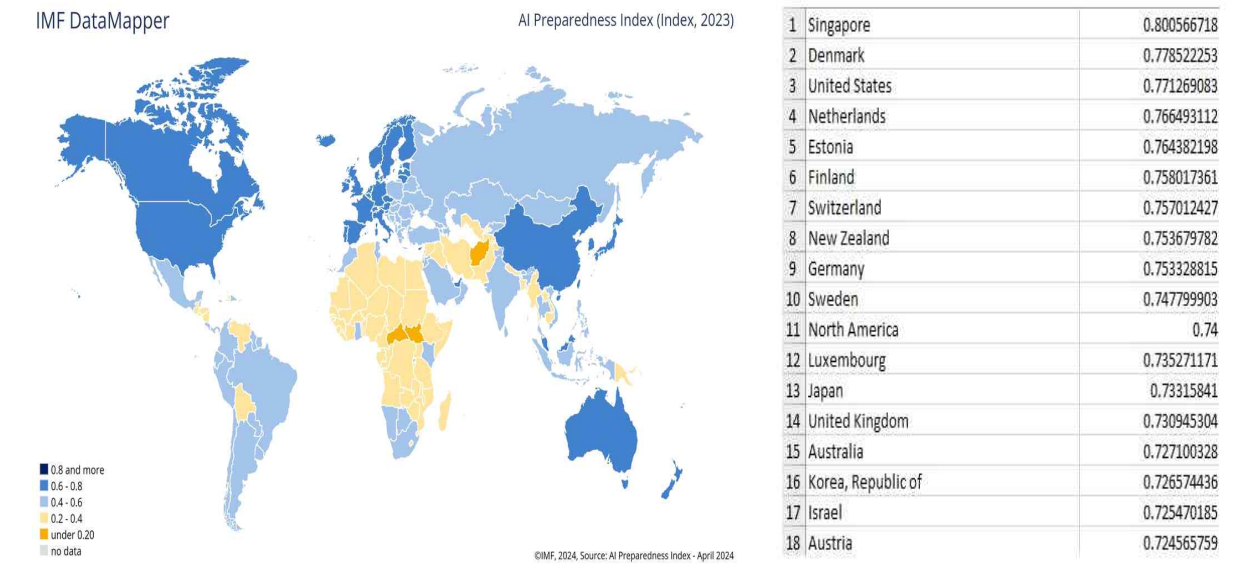
- EU 경쟁 당국은 MS와 오픈AI의 협력 관계에 대한 반독점 조사에 이어 삼성전자와 구글이 다년간 계약을 체결하고 지난 1월 출시한 갤럭시 S24 시리즈에 구글의 AI모델인 제미니 나노(Gemini Nano)를 탑재한 부분에 대해서도 반독점 조사를 검토 중
- EU의 경쟁 담당 수석 부집행위원장은 거대 기술기업으로 인해 소규모 AI 기업이 사용자와 기업에 접근할 수 없게 되는 상황에 대하여 우려를 표시하며, 특정 삼성 기기에 제미니 나노를 선택재 한다는 구글과 삼성전자 간 합의의 영향을 더 잘 파악하는 데 필요한 정보를 요청했다고 밝힘
- 이번 조사 검토는 EU 당국이 디지털시장법(DMA) 하에서 시장 지배적 영향력을 지닌 게이트키퍼로 지정한 구글을 겨냥한 것으로, 이번 협업이 다른 AI 기업들의 시장 진입을 차단하는 효과를 일으켰는지 살펴볼 계획인 것으로 나타남

2. 국가별 AI 활용 동향

가. IMF, 176개국 AI 준비 지수 발표

- IMF(The International Monetary Fund, 국제통화기금)는 전 세계 176개국을 대상으로 국가별 AI의 준비 상태를 조사한 AI 준비 지수를 발표. AI 준비 지수는 디지털 인프라, 인적 자본과 노동 시장 정책, 기술 혁신, 법적 프레임에 대한 영역을 중심으로 평가하여 0부터 1까지 수치로 표시
- IMF는 해당 지수가 AI 발명 리더십이 아닌 AI 채택 준비에 초점이 맞춰져 있으며 순위를 매기는 목적으로 사용되기보다는 정책 입안자들이 개선이 필요한 영역을 식별하도록 안내하는 지표 척도로 참고해야 한다고 밝힘. 또한 이를 분석하기 위한 데이터로 UN, 프레이저 연구소, 국제노동기구, 국제전기통신연합, UN무역개발회의, 만국우편연합, 세계은행, 세계경제포럼 등의 기관에서 수집한 공식 데이터와 인식조사를 활용
- 국가별 준비 지수는 싱가포르가 0.8로 가장 높았으며 덴마크, 미국, 네덜란드, 에스토니아가 0.77로 상위 5개 국가에 포함되는 것으로 나타남. 아시아에서는 한국과 일본이 0.73으로 나타났으며 중국은 0.64로 이보다 뒤쳐져 준비 지수에서 차이가 있음을 보임
 - * 선진국의 평균 AI 준비 지수는 0.68이었으며, G7은 0.72, EU는 0.66인 반면, 신흥 경제국은 0.46, 라틴아메리카와 카리브해 지역은 0.43, 저소득국은 0.32를 기록
- AI가 국가 간 불평등을 악화시킬 가능성이 높기 때문에 선진국에서는 사회 안전망 확대와 근로자 교육, AI 위협과 남용으로부터 국민 보호 및 신뢰를 구축하기 위한 규제 강화를 우선해야 하며, 신흥시장과 개발도상국에서는 디지털 인프라 구축 및 근로자를 위한 디지털 교육에 투자하여 기반을 마련하는 등 이번 조사가 AI 준비 상태를 파악해 AI의 이점을 활용하고 위협을 관리할 각국의 정책 마련에 도움이 될 것으로 기대

[그림 3] IMF의 글로벌 AI 준비 지수 현황

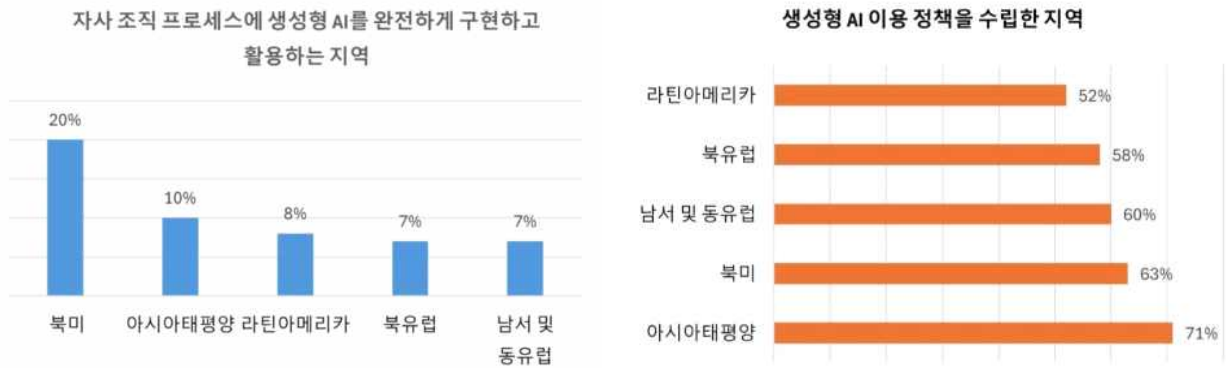


*출처 : IMF(https://www.imf.org/external/datamapper/AI_PI@APII/ADVEC/EME/LIC?year=2023)

나. 국가별 생성 AI 사용 조사 결과 발표

- 미국의 인공지능 분석 기업인 SAS와 콜먼파크스 리서치가 전 세계 주요 지역의 직원 수 1만 명 이상 기업의 의사 결정자 1,600명을 대상으로 실시한 설문조사에 따르면 중국의 기업이 83%로 가장 활발하게 생성형 AI를 사용하고 있는 것으로 나타남
 - * 중국의 비즈니스 의사결정자들은 조직 내 83%가 생성 AI를 사용 중이라고 답했으며, 이는 영국(70%), 미국(65%), 호주(63%)보다 높은 수치
- 그러나 기업이 생성형 AI 기술을 완전히 구현하고 활용하는지를 묻는 질문에서는 미국이 24%로 나타나 AI 활용의 성숙도가 가장 높은 것으로 조사됨. 조사 기관인 콜먼 파크스는 중국이 생성형 AI 채택률 면에서는 최고일 수 있지만 더 높은 채택률이 반드시 효과적인 활용이나 더 높은 투자 수익률을 의미하는 것은 아니며, 실제로는 중국에 비해 미국의 조직이 생성형 AI를 완전하게 구축해 활용하고 있기 때문에 경쟁에서 앞서고 있다고 설명
 - * 조직 프로세스에 생성 AI를 완전하게 구현하고 활용하는 지역은 북미 20%, 아태지역 10%, 라틴아메리카 8%, 북유럽 7%, 남서 및 동유럽 7%로 나타남
- SAS는 생성형 AI를 여러 지역이 이미 이용하면서 의미 있는 방식으로 채택하기 시작했지만 그 현황은 상이하며, 생성형 AI를 통해 조직에서 신기술을 도입할 때 과장된 부분과 실재를 분리하는 발견(Discovery) 단계를 거치게 됨에 따라, 이제는 반복 가능하고 신뢰할 수 있는 비즈니스 결과를 의도적으로 구현하고 제공하는 것이 중요하다고 설명

[그림 4] 전 세계 지역별 AI 사용 조사 결과



*출처 : SAS 홈페이지(https://www.sas.com/ko_kr/news/press-releases/2024/july/genai-research-study-global.html)

■ 시사점

- AI가 딥페이크로 생성한 가짜 영상과 음성은 진위를 구별하기 어렵게 만들며 이를 악용한 피싱 등의 사기, 허위 정보 유포, 개인정보 도용 등 전 세계적으로 여러 사회적 문제를 일으키고 있음. 이러한 문제를 방지하기 위해서라도 생성형 AI의 투명성과 책임성을 강화하고, 기술의 활용에 있어서는 윤리적 고려와 사회적 책임을 동반해야 함
- AI는 인간의 도구로써 어떻게 사용되느냐에 따라 그 결과가 달라지게 되며 우리 사회에서 AI 기술의 발전은 피할 수 없는 현실이 되고 있기에 그에 따른 기회와 위험을 신중히 고려할 필요가 있음. 인간은 AI 기술의 발전을 적극 활용하면서도 그에 따른 부작용을 최소화할 수 있는 균형 잡힌 접근이 필요. 일자리 변화와 같은 사회적 문제를 해결하기 위해서는 교육과 재훈련 프로그램이 중요하며, 국가별로 AI 기술의 투명성과 책임성을 강화하고 AI를 악용한 각종 범죄나 딥페이크 문제를 막기 위한 기술적·법적 보호 장치를 재정비하기 위해서는 EU와 같은 국제기구를 중심으로 통합된 규제를 마련하여 통제·운영 하는 등으로 이를 주도하는 역할이 중요

참고자료

- 국내 ‘AI 채택 준비’ 세계 16위...IMF, 176개국 AI 준비 지수 발표
<https://www.aitimes.com/news/articleView.html?idxno=161062>
- IMF, ‘AI 준비과정에서 싱가포르, 덴마크, 미국이 가장 앞서’
https://www.tokenpost.kr/article-185530#google_vignette
- 새로운 무역장벽? ‘AI 반독점’ 칼빼든 EU...애플·삼성 영향은
<https://weekly.hankooki.com/news/articleView.html?idxno=7091289>
- 반복적으로 생성된 데이터를 학습하면 AI 모델이 붕괴
<https://www.nature.com/articles/s41586-024-07566-y>
- “AI 데이터로 훈련한 모델, 결국 쓸모 없어질 것”
<https://zdnet.co.kr/view/?no=20240726154935>
- 구글 “생성형AI 악용사례 26%는 여론조작”
<https://www.donga.com/news/Economy/article/all/20240813/126514798/2>
- 메타, 트럼프 피격 사건 정보 조작 논란 해명... “AI 환각 탓”
<https://www.digitaltoday.co.kr/news/articleView.html?idxno=527628>
- 구글이 검색 결과 내 노골적인 허위 콘텐츠에 대처하는 방법을 소개합니다.
<https://blog.google/intl/ko-kr/products/explore-get-answers/google-approach-to-deepfake/>
- 생성형 AI에 대한 전세계 시장 조사 결과
https://www.sas.com/ko_kr/news/press-releases/2024/july/genai-research-study-global.html
- 생성형 AI 이용 中 선도, 구현과 활용 성숙도 ‘미국’이 견인
<https://industryjournal.co.kr/news/237528>

SPC 'ANGEL' 통계

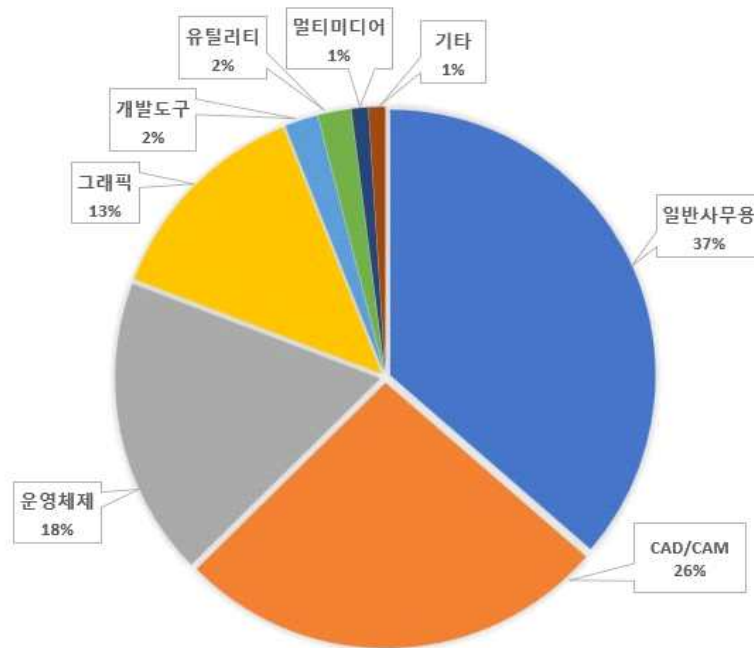
월 1회 제공

한국소프트웨어저작권협회, 불법복제 SW 제보 'ANGEL' 서비스 9월 통계 현황

- 한국소프트웨어저작권협회(SPC)가 지난 8월 한 달간('24. 9. 1. ~ 9. 31.) 'ANGEL (불법제보)' 서비스를 분석한 결과, 기업 또는 개인의 불법복제 SW 사용 제보는 총 92건으로 나타남
- SW 용도별로는 일반사무용 33건(36%), 설계(CAD/CAM) 24건(26%), 운영체제 17건(18%), 그래픽 12건(13%), 개발도구 2건(2%), 유틸리티 2건(2%), 멀티미디어 1건(1%), 기타 1건(1%) 순으로 제보가 접수됨

[그림] SPC 'ANGEL(불법제보)' 서비스 2024년 9월 통계 현황

2024. 9. 불법복제 소프트웨어 제보 통계
-SW 용도별 제보 건수-



* 한국소프트웨어저작권협회는 2018년 11월부터 제보시스템과 제보 방식의 편의성을 개선한 불법복제 SW 제보 시스템 'ANGEL(불법제보)' 서비스를 운영하고 있음

다음 SW·저작권 동향리포트 <제2024-21호> 발간일은 11월 10일입니다.