

융합하는 AI·첨단기술·미디어·통신 거대한 혁신의 생태계 형성

딜로이트 글로벌 TMT 센터 지음
딜로이트 인사이트 옮김

다음 시대의
비즈니스를 준비하기 위한
기술, 미디어, 통신 산업
필수 전략서

융합하는 AI·첨단기술·미디어·통신 거대한 혁신의 생태계 형성

한국 딜로이트 그룹

생성형AI 시대 본격화

생성형AI로 촉발되는 데이터센터의 에너지 수요와 이를 해결하기 위한 첨단기술과 전력 산업의 공동 전환을 조명하며, AI 인력 구조와 생성형AI 신뢰에 있어 나타나는 성별 격차를 짚어보고 기업들의 해결과제를 제시한다. 또한 소비자의 삶과 기업의 자동화 프로세스를 다시 한번 격변할 생성형AI 기반 스마트폰과 에이전틱 AI의 시장 현황과 성장 전망을 심층적으로 분석한다. 마지막으로 AI 기술 발전의 가장 심각한 부작용인 딥페이크의 메커니즘을 이해하고 이로 인해 새롭게 창출되는 사이버 보안 시장 기회도 탐색한다.

미디어·엔터테인먼트, 첨단기술로 가속화되는 시장 생태계 변화

스포츠 인프라가 디지털 장치와 첨단기술로 무장해 세대를 아우르며 대중을 융합하고 지역을 활성화하는 동시에 수익 다각화에 핵심 역할로 부상하고 있다. 이러한 스포츠 인프라를 혁신적으로 개발하는 전략을 다각도로 탐구한다. 이 외 독립형 SVOD만으로는 승부를 보기 어려운 환경에서 애그리게이션(aggregation) 모델로 통합되는 비디오 시장과 생성형AI로 경쟁 판도가 재편되는 콘텐츠 제작 및 배급 부문의 중기적 전망을 제시한다.

첨단기술·통신, 지금부터 비용 싸움...정확한 투자수익 예측이 관건

파편화된 세계 각국 무선통신 시장을 통합하려는 규제의 움직임으로 통신산업 인수합병(M&A)이 증가할 것으로 전망된다. 한편 갈수록 심화되는 클라우드 환경의 복잡성과 비용 증대에 대한 대응책으로 클라우드 인프라의 재무 성과를 최적화할 수 있는 핀옵스(FinOps) 원칙 및 프랙티스를 심층적으로 분석한다.

융합하는 AI·첨단기술·미디어·통신,
거대한 혁신의 생태계 형성

융합하는 AI·첨단기술·미디어·통신,
거대한 혁신의 생태계 형성

1판 1쇄 발행 2025. 02. 10

지은이 딜로이트 글로벌 TMT 센터
옮긴이 딜로이트 인사이트

발행인 홍종성
총괄 경동섭
편집 김선미, 신인식, 양원석
디자인 박근령
발행처 한국 딜로이트 그룹
출판등록 2010년 05월 31일
주소 서울시 영등포구 국제금융로 10 ONE IFC 9F
연락처 02-6676-1000 (대표번호)
홈페이지 <https://www2.deloitte.com/kr>

저작권 © 딜로이트 글로벌 · 딜로이트 안진회계법인
이 책은 저작권법에 의해 보호를 받는 저작물이므로
저자와 출판사의 허락 없이 내용의 일부를 인용하거나 발췌하는 것을 금합니다.

본 단행본 원문의 저작권은 딜로이트 글로벌에 있으며,
본 한글판은 딜로이트 글로벌의 승인 후 번역, 발간되었습니다.

비매품

융합하는 AI·첨단기술·미디어·통신
거대한 혁신의
생태계 형성

AI는 이제 단순한 도구를 넘어 창의적 파트너로 자리 잡고 있으며,
 첨단기술은 산업 간 경계를 허물며 새로운 시장을 창출하고 있습니다.
 미디어는 정보의 소비 방식을 혁신하고,
 통신 기술은 물리적 한계를 뛰어넘은 초연결 사회를 가능케 합니다.
 이 모든 요소는 각각 독립적으로 발전하는 것이 아니라,
 서로 얽혀 다양한 상호작용을 통해 거대한 혁신의 생태계를 형성하고 있습니다.

CONTENTS

프롤로그 006

제1장 생성형AI 시대 본격화

01 생성형AI의 진화, 이제 데이터센터 에너지의 신뢰성과 청정화가 관건	012
02 생성형AI, 여성 소비자와 AI 인력 기반 확대 필요	029
03 생성형AI 스마트폰, 세상을 바꿀 또 한 번의 격변 촉발	042
04 자동차의 새로운 시대 여는 에이전틱 AI	054
05 딥페이크와 사이버 보안, 쫓고 쫓기는 싸움	070

제2장 미디어·엔터테인먼트, 첨단기술로 가속화되는 시장 생태계 변화

01 꿈의 경기장 건설 프로젝트, 공공-민간 투자의 교두보	082
02 비디오 콘텐츠 시장, DTC 지고 애그리게이터 뜬다	095
03 생성형AI, 콘텐츠 제작과 배급의 경쟁 판도 재편	107

제3장 첨단기술·통신, 지금부터 비용 싸움... 정확한 투자수익 예측이 관건

01 통신산업 통합으로 기운 세계 각국 규제 의 풍향계	122
02 클라우드 인프라의 재무 성과 최적화를 위한 '핀옵스' 원칙	132

PROLOGUE

2025년, 우리는 인공지능AI, 첨단기술, 미디어, 통신이 융합된 새로운 시대의 중심에 서 있습니다. AI 혁명은 단순히 기술적 진보를 넘어 인간 삶의 모든 영역을 재구성하고, 사회적, 경제적, 문화적 변화를 가속화하고 있습니다. 이러한 변화는 과거의 예측을 뛰어넘는 속도로 우리의 일상에 스며들고 있으며, 새로운 가능성과 도전을 동시에 제시하고 있습니다.

AI는 이제 단순한 도구를 넘어 창의적 파트너로 자리 잡고 있으며, 첨단 기술은 산업 간 경계를 허물며 새로운 시장을 창출하고 있습니다. 미디어는 정보의 소비 방식을 혁신하고, 통신 기술은 물리적 한계를 뛰어넘은 초연결 사회를 가능케 합니다. 이 모든 요소는 각각 독립적으로 발전하는 것이 아니라, 서로 얽혀 다양한 상호작용을 통해 거대한 혁신의 생태계를 형성하고 있습니다.

이 책은 2025년의 기술적 전망과 그 사회적 함의를 탐구하며, AI, 첨단 기술, 미디어, 통신의 현재 상태와 발전 방향을 분석하고, 이러한 기술이 미래의 산업, 문화, 정책, 인간관계에 어떤 변화를 가져올지 조망하고자 합니다. 또한 이러한 변화 속에서 우리가 직면할 윤리적, 경제적, 환경적 과제들에 대해서도 깊이 있는 논의를 제시합니다.

2025년 전망은 생성형AI로 앞당겨진 산업의 변곡점과 그 과정에서 발생할 격차와 융합에 가장 많은 지면을 할애했습니다. 제1장 ‘생성형AI 시대 본격화’는 생성형AI로 촉발되는 데이터센터의 에너지 수요와 이를 해결하기 위한 첨단기술과 전력 산업의 공동 전환을 조명합니다. 이어 AI 인력 구조와 생성형AI 신뢰에 있어 나타나는 성별 격차를 짚어보고 기업들의 해결과제를 제시합니다.

또한 소비자의 삶과 기업의 자동화 프로세스를 다시 한번 격변할 생성형AI 기반 스마트폰과 에이전틱 AI의 시장 현황과 성장 전망을 심층적으로 분석합니다. 마지막으로 AI 기술 발전의 가장 심각한 부작용인 딥페이크의 메커니즘을 이해하고 이로 인해 새롭게 창출되는 사이버 보안 시장 기회도 탐색합니다.

제2장에서는 미디어·엔터테인먼트·스포츠 산업의 첨단기술 도입 현황과 시장 생태계 변화 동향을 다룹니다. 디지털 장치와 첨단기술로 무장해 세대를 아우르며 대중을 융합하고 지역을 활성화하는 동시에 수익 다

각화에 핵심 역할을 하는 스포츠 인프라의 혁신적 개발 전략을 다각도로 탐구합니다. 이 외 애그리게이션^{aggregation} 모델로 통합되는 비디오 시장과 생성형AI로 경쟁 판도가 재편되는 콘텐츠 제작 및 배급 부문의 중기적 전망을 제시합니다.

제3장에서는 통신 및 첨단기술의 주요 이슈로 세계 각국 규제의 풍향계가 통신산업의 통합으로 기울고 있는 현황과 클라우드 인프라의 재무 성과를 최적화할 수 있는 핀옵스^{FinOps} 원칙 및 프랙티스를 심층적으로 분석합니다.

본서가 제시하는 통찰은 단순히 미래를 예측하는 데 그치지 않습니다. 기술의 발전이 가져올 기회와 위험을 이해하고, 이를 통해 새로운 시대를 준비하는 데 필요한 방향성을 제시하고자 합니다. 독자 여러분이 이 책을 통해 2025년의 기술적 패러다임을 이해하고, 변화하는 세상 속에서 자신의 역할을 재정립하는 데 도움이 되길 바랍니다.

손재호
한국 딜로이트 그룹
성장전략부문 리더



Chapter 01

생성형AI 시대 본격화

- 01 생성형AI의 진화, 이제 데이터센터 에너지의 신뢰성과 청정화가 관건
- 02 생성형AI, 여성 소비자와 AI 인력 기반 확대 필요
- 03 생성형AI 스마트폰, 세상을 바꿀 또 한 번의 격변 촉발
- 04 자동화의 새로운 시대 여는 에이전틱 AI
- 05 딥페이크와 사이버 보안, 쫓고 쫓기는 싸움

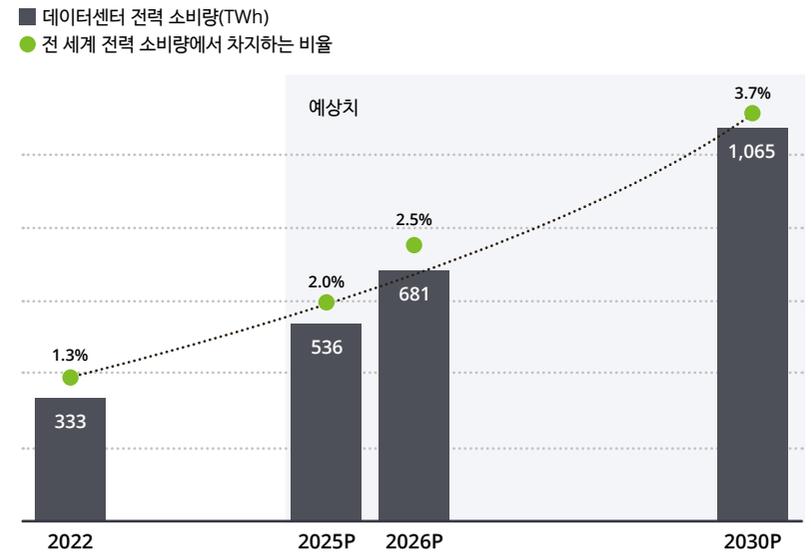
01

생성형AI의 진화, 이제 데이터센터 에너지의 신뢰성과 청정화가 관건

첨단기술 산업은 데이터센터의 지속가능성을 개선하기 위해 인프라 최적화, 반도체칩 설계 재검토, 전력회사와의 협업 등 다각도의 방안을 모색할 필요가 있다.

인공지능AI 기반 데이터센터의 전력 소비가 계속 급증하고 있지만, 아직 전 세계 에너지 수요에서 데이터센터가 차지하는 비중은 크지 않다. 딜로이트는 2025년 데이터센터의 전력 소비량이 536테라와트시TWh로 전 세계 전력 소비량의 약 2%에 그칠 것으로 예상된다. 하지만 전력 소모가 큰 생성형AI^{generative AI} 훈련과 추론 활동이 여타 활용사례나 앱에 비해 빠르게 증가하는 만큼, 2030년에 이르면 전 세계 데이터센터 전력 소비량이 약 1,065TWh에 달할 것으로 예상된다(그림 1).¹ 이처럼 급증하는 데이터센터의 전력 수요를 맞추면서 환경 영향을 줄이려면, 혁신적이고 에너지 효율적인 데이터센터 기술을 탐색함과 동시에 무탄소 에너지원의 활용을 늘려야 한다.

그림 1
전력 소모가 큰 생성형AI 모델을 중심으로 전 세계 데이터센터 전력 소비량 급증 전망



참조: 'P'는 예측값을 뜻함.

출처: Deloitte analysis based on publicly available information sources and conversations with industry experts.

방법론: 2022~2030년 전 세계 데이터센터 전력 소비량 추정치와 예상치는 미국 에너지정보청(EIA)이 주거·산업·운송 최종사용의 총 전력 사용량 데이터를 취합한 '국제 에너지 전망 2023'(International Energy Outlook 2023)의 기본 전력 소비 데이터에 기반해 집계했다. 전 세계 에너지 소비량에서 데이터센터가 차지하는 비율(%) 전망치는 세미 애널리시스(Semi Analysis), 미국 전력연구원(EPR), 골드만삭스(Goldman Sachs), 블룸버그(Bloomberg), 래티튜드 미디어(Latitude Media)의 공개 자료와 더불어 첨단기술-에너지-지속가능성 분야 산업 전문가의 견해를 기반으로 도출했다.

물론 수많은 변수가 발생할 수 있으므로, 2030년과 그 이후의 전 세계 데이터센터 에너지 소비량을 정확히 예측하기는 쉽지 않다. 다만 AI와 데이터센터의 프로세싱 효율성이 지속적으로 개선되면 2030년에는 약 1,000TWh의 에너지가 소모될 것이라는 전망을 제시할 수 있다. 하지만

향후 수년 내로 이러한 효율성 개선이 이뤄지지 않는다면, 데이터센터의 전력 소비량은 1,300TWh를 돌파해, 전력회사들이 직접적 영향을 받고 기후중립 목표 달성에도 차질이 생길 수 있다.² 결과적으로 앞으로 10년간 AI 혁신에 박차를 가하고 데이터센터 효율성을 최적화해야만 지속 가능한 에너지 환경을 구축할 수 있다.

일부 지역에서는 이미 AI 데이터센터의 전력 수요 증가로 인해 전력 생산과 전력망 용량 관리가 해결하기 힘든 과제로 부각되고 있다.³ 그래픽 처리장치GPU, 중앙처리장치CPU 서버, 스토리지 시스템, 냉각 시스템, 네트워크 스위치 등 데이터센터의 핵심 장치 및 시스템을 구동하는 데 필요한 전력이 2026년 글로벌 기준 96기가와트GW에 달해 2023년에 비해 약 두 배 증가할 것으로 예상된다. 이 중 AI 운영이 40% 이상을 차지할 것으로 보인다.⁴ 전 세계 AI 데이터센터의 연간 전력 소비량은 2026년 90TWh에 달해 2022년에 비해 무려 10배 증가할 것으로 전망된다.⁵ 이는 전 세계 데이터센터 2026년 전력 소비량 총합 전망치인 681TWh의 약 1/7 수준이다. 생성형AI 투자로 전력 수요가 급격히 촉발돼, 2024년 1분기 전 세계 AI 데이터센터의 전력 추가 순수요는 약 2GW로 2023년 4분기에 비해 25% 늘었고, 2023년 1분기에 비하면 세 배 이상 급증했다.⁶ 더군다나 데이터센터 시설은 지리적으로 주로 미국에 집중돼 있어 전력 수요를 충족하기가 쉽지 않고, 일년 내내 24시간 운영되기 때문에 기존 전력 인프라에 부담을 주고 있다.⁷

이에 따라 첨단기술과 전력 산업이 힘을 합쳐 이러한 도전 과제를 해결하고 AI, 특히 생성형AI로 초래되는 에너지 부담을 줄이도록 노력해야 한다. 이미 다수의 빅테크 및 클라우드 기업들이 무탄소 에너지원에 투자하고 넷제로 목표를 향해 달리며 지속가능성 개선을 위한 노력을 펼치고 있다.⁸

하이퍼스케일러 업계, 고객 수요 증대에 발맞춰 생성형AI 데이터센터 대거 확대

이처럼 데이터센터 전력 수요가 급증하는 주요 원인은 전 세계적으로 하이퍼스케일러hyperscaler, 대규모 데이터센터 운용기업 업계가 데이터센터 용량을 확대하는 추세이기 때문이다.⁹ 생성형AI를 중심으로 AI 수요가 증가할 것이라는 예상에 각국 정부와 기업들이 너도나도 데이터센터 확충에 나서고 있다. 특히 각국 정부는 기술 패권 싸움에서 지지 않기 위해 소버린 AIsovereign AI* 역량을 서둘러 구축하고 있다.¹⁰ 전 세계 주요 하이퍼스케일러의 자본지출은 2024년 기준 약 2,000억 달러에서 2025년 2,200억 달러를 넘을 것으로 예상된다. 이를 기반으로 추산하면 전 세계 데이터센터 부동산 빌드아웃real estate build-out**은 이미 사상최대치를 찍었다.¹¹

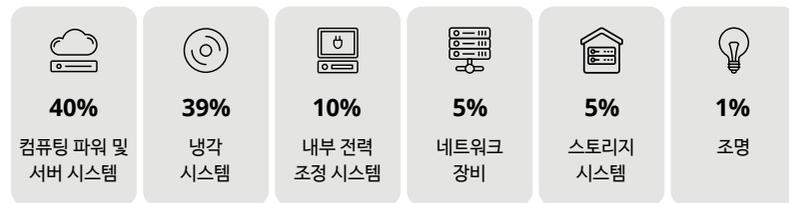
* 소버린 AI(sovereign AI)는 한 국가가 자체 데이터와 인프라를 기반으로 자국의 제도과 문화, 역사, 가치관을 심도 깊게 이해한 AI를 독립적으로 개발하고 운영하는 것을 뜻한다.

** 부동산 빌드아웃(real estate build-out)은 기본 구조만 갖춰진 상업용 또는 주거용 부동산을 임차인 및 사용자의 세부적인 요구에 맞춰 추가 공사나 마감 작업을 하는 것을 의미한다.

또한 딜로이트의 기업 대상 서베이에 따르면, 지금까지는 생성형AI 도입이 실험적 단계에 불과했지만,¹² 이를 통해 실질적 가치가 입증되면서 빠른 속도로 규모를 확대해 본격적 배치가 이뤄질 것으로 예상된다. 생성형 AI 기술의 성숙과 함께 활용도 증가함에 따라, 하이퍼스케일러와 클라우드 기업들의 자본지출은 2025~2026년 높은 수준으로 유지될 전망이다.

데이터센터에서 전력을 가장 많이 소비하는 것은 컴퓨팅 파워^{computing power} 및 서버 시스템(데이터센터 전력 소비의 약 40% 차지)과 냉각 시스템(약 38~40% 차지)이다. 이들은 AI 데이터센터에서도 가장 에너지 집약적 시스템이므로 지속적으로 데이터센터의 전력 소비를 촉발하는 요인이 될 것이다. 이 외 내부 전력 조정 시스템이 8~10%, 네트워크 및 통신 장비와 스토리지 시스템이 각각 약 5%, 조명 시설이 1~2%의 전력을 소비한다(그림 2).¹³ 생성형AI로 막대한 용량의 전력이 필요한 만큼, 하이퍼스케일러와 데이터센터 운영 기업들은 데이터센터 설계 시 대체 에너지원, 새로운 형태의 냉각 시스템, 에너지 효율성 개선 솔루션 등을 모색할 필요가 있다.

그림 2
컴퓨팅파워와 냉각 시스템이 AI 데이터센터에서 가장 많은 전력 소비



출처: 딜로이트 분석

생성형AI로 데이터센터 전력 수요 급증

데이터센터의 에너지 소비는 AI 수요와 더불어 2023년부터 급증하기 시작했다.¹⁴ 첨단 AI 시스템을 배치하려면 수많은 반도체칩과 처리 용량이 필요하고 복잡한 생성형AI 모델을 훈련하려면 수천 개의 GPU가 필요하다.

생성형AI와 고성능 컴퓨팅 환경을 운영하는 하이퍼스케일러와 대규모 데이터센터 기업들은 컴퓨팅 파워를 지원하기 위해 고밀도의 인프라가 필요하다. 과거 데이터센터는 반도체칩 한 개당 약 150~200와트^w로 구동되는 CPU에 주로 의존했다.¹⁵ AI용 GPU는 2022년까지 400W의 전력으로 가능했으나, 2023년 생성형AI용 최첨단 GPU는 700W가 필요했고, 2024년 차세대 반도체는 1,200W가 필요할 것으로 예상된다.¹⁶ 현재 약 8종류가 개발된 차세대 반도체는 데이터센터의 서버 랙^{rack} 내부에 10개씩 적재된 블레이드^{blade}에 장착되는데, 불과 수년 전에 개발된 데이터센터 설계에 비해 기초면적 평방미터당 더 많은 물을 소비하고 더 많은 열을 발산한다.¹⁷ 서버 랙당 전력 요구량 추정치는 2023년 36kW에서 2024년 초 20kW 수준으로 다소 줄었으나, 2027년 50kW로 급증할 것으로 예상된다.¹⁸

컴퓨터 성능을 나타내는 단위인 플롭스^{FLOPS, floating-point operations per second}, 초당 부동소수점 연산을 기준으로 총 AI 컴퓨팅 용량 또한 생성형AI가 등

장한 이후 기하급수적으로 증가하고 있다. AI 컴퓨팅 플롭스는 2023년 1분기 이후 전 세계적으로 분기마다 50~60%씩 늘고 있으며, 2025년 1분기까지 같은 속도의 증가세가 예상된다.¹⁹ 데이터센터의 컴퓨팅 용량은 플롭스뿐 아니라 MWh 및 TWh로도 측정할 수 있다.

수십억 개의 파라미터를 가진 생성형AI의 대규모 언어모델, 막대한 전력 소비

생성형AI의 대규모언어모델LLM은 갈수록 많은 파라미터parameter, 매개변수를 통합해 한층 고도화된다. 2021~2022년 1,000억~2,000억 개의 초기 파라미터 모델에서 시작해 2024년 중반 현재 첨단 LLM의 파라미터는 2조 개에 육박해 복잡한 이미지를 해석 및 해독한다.²⁰ 더 나아가 이제는 10조 개 파라미터를 갖춘 LLM 출시 경쟁에 불이 붙었다. AI 모델 훈련 및 배치를 위해 데이터 프로세싱과 컴퓨팅 파워에 더 많은 파라미터가 추가되며, 이는 생성형AI 프로세서와 액셀러레이터의 수요를 촉발해, 결국 전력 소비를 대거 끌어올린다.

게다가 LLM 훈련은 에너지 집약적 과정이다. 최근 조사에 따르면, 1,750억 개 이상의 파라미터로 훈련한 주요 LLM은 1회 훈련 시마다 324~1,287MWh의 전력을 소비하는 것으로 나타났다. 그리고 LLM은 재훈련이 필요한 경우가 많다.²¹

평균적으로 생성형AI 기반 프롬프트 요청을 처리하는 것은 인터넷 검색 결과 도출보다 10배에서 100배 많은 전력을 소비한다.²² 델로이트의 분석에 따르면, 일일 인터넷 검색 건수의 5%만 생성형AI 프롬프트 요청으로 바뀌어도, 개당 8개의 특수화된 GPU 코어가 탑재되어 평균 6.5kW의 전력을 소비하는 서버 약 2만 개가 필요하다. 일일 평균 전력 소비량으로 따지면 3.12GWh, 연 평균 소비량은 1.14TWh에 달하는 규모로,²³ 약 10만8,450 미국 가구의 연간 전력 소비량과 맞먹는 수준이다.²⁴

데이터센터의 전력 수요, 전력 산업 전환의 과제이자 기회

전력 산업은 이미 수요 증가에 대비하며, 일부 국가에서는 2050년까지 전력 소비량이 최대 세 배 증가할 것으로 예측하고 있다.²⁵ 하지만 최근 데이터센터의 전력 수요가 급증하면서 일부 지역에서는 전력 소비량 증가세도 가속화되고 있다. 상당수 국가에서 기존의 전력 수요량 전망치는 전동화를 중심으로 데이터센터의 전력 소비 및 전반적인 경제성장을 기반으로 도출됐다. 하지만 지금 가시화되는 데이터센터 전력 소비 급증 양상은 빙산의 일각에 불과하며, 향후 전력 산업 전반의 전환을 초래할 정도로 가속화될 수 있다.²⁶

현재 전력 산업은 수십년에 걸친 전환을 아직 진행 중이다. 전력망 인프라를 구축·업그레이드·탈탄소화하고, 시스템과 자산을 디지털화하며,

갈수록 극심해지는 기후 환경에서도 피해를 최소화하기 위해 설비 방재를 강화하고, 증가하는 사이버보안 위협에 대응해 네트워크 보호 시스템을 구축하고 있다.²⁷ 하지만 여전히 일부 국가에서는 전력망이 수요를 따라잡지 못하고 있으며, 특히 저탄소 및 무탄소 전력은 공급이 턱없이 부족하다. 미국의 경우 2026년 전국 전력 총소비량에서 데이터센터가 6%(260TWh)를 차지할 것으로 예상된다.²⁸ 영국은 AI 데이터센터를 중심으로 데이터센터 전력 수요량이 단 10년 만에 6배 증가할 것으로 전망된다.²⁹ 중국도 2026년 전력 총수요량에서 AI를 포함한 데이터센터가 6%를 차지할 것으로 예상된다.³⁰ 중국의 경우 주요 전력원이 여전히 석탄이기 때문에 데이터센터가 대기오염의 또 다른 원인이 될 수 있다. 2021년 기준 중국에서 석탄이 에너지 사용의 61%, 이산화탄소 배출의 79%를 차지했다.³¹

일부 국가들은 데이터센터의 전력 수요 증가를 억제하기 위해 규제 대응에 나섰다. 아일랜드는 기존 데이터센터들이 전력 총소비량의 1/5을 차지하는 가운데 AI 데이터센터가 늘면서 이에 따른 전력 소비도 한층 늘어날 전망이다.³² 가계 전력 소비량은 오히려 줄어들고 있음에도 전력 공급난 가능성이 대두되자, 아일랜드 정부는 한시적으로 신규 데이터센터의 전력망 연결을 금지하기도 했다.³³ 네덜란드 암스테르담 시정부도 지속 가능 도시 개발을 위해 데이터센터 신설을 금지했다.³⁴ 싱가포르는 데이터센터 내 온도를 26°C 이상으로 유지하도록 하는 새로운 지속가능성 표준을 도입했다. 데이터센터 내 유지 온도가 높아지면 냉각 시스템 가

동이 줄어 전력 소비량은 감소하지만, 반도체칩의 수명이 짧아진다는 문제가 있다.³⁵

데이터센터는 지리적으로 집중돼 있는 데다 전력 수요를 당장 365일 24시간 충족할 필요가 있어, 그렇지 않아도 전동화와 제조업 전환 등으로 발생하는 새로운 전력 수요 부담을 안고 있는 첨단기술 및 전력 기업들이 한층 복잡한 과제를 마주하게 됐다. 전 세계에서 가장 큰 데이터센터 시장은 미국 노던 버지니아(Northern Virginia)에 형성돼 있는데,³⁶ 현지 전력 및 유틸리티 회사인 도미니언 에너지(Dominion Energy)는 향후 15년 새 데이터센터의 전력 수요가 네 배 증가하면서 전력 총수요가 약 85% 증가할 것으로 내다봤다.³⁷ 하지만 많은 테크 기업들이 원하는 365일 24시간 무탄소 전력 공급은 단기 내에 실현하기가 어렵다. 이에 전력회사들은 늘어나는 전력 수요를 충당하면서도 공급 신뢰성과 경제성을 사수하기 위해 다각도로 모색하고 있다. 새로운 재생에너지와 배터리 저장 시설의 개발 및 구축 외에도, 급한 불을 끄기 위해 천연가스 발전소 신설 계획도 세우고 있다.³⁸ 천연가스는 석탄보다는 양호하지만 탄소를 배출하는 에너지원이기 때문에, 이로 인해 전력회사와 주 정부, 더 나아가 국가의 탈탄소화 목표 달성이 한층 어려워질 수 있다.³⁹

다만 막대한 규모의 에너지를 소비하는 AI가 청정에너지 전환을 가속화하는 요인으로 작용할 수도 있다. 일부 유틸리티 회사들은 이미 AI를 활용해 날씨 및 부하 예측의 정확도를 높이고, 전력망과 재생에너지 자산

성능을 강화하고, 폭풍 재해의 복구 속도를 끌어올리고, 산불 위험 예측을 개선하는 등 전력망 운영의 비용을 절감하고 효율성과 신뢰성을 개선하고 있다.⁴⁰

물 자원 소모가 큰 데이터센터 냉각 시스템

차세대 CPU와 GPU는 이전 세대보다 열 밀도(thermal density)*가 높은 성질이 있다. 또한 일부 서버 벤더들은 고성능 컴퓨팅과 AI 앱의 증가하는 수요를 맞추기 위해 서버 랙에 전력을 많이 소비하는 반도체칩을 더 많이 적재하고 있다. 하지만 생성형AI 등 서버 랙의 밀도가 높아질수록 냉각수가 더 많이 필요하다. AI 데이터센터의 민물 수요량은 2027년 최대 1.7조 갤런에 달할 것으로 전망된다.⁴¹ 만약 에어 냉각(air-based cooling)이나 증발 식수로 초과 열을 관리하면 매년 5,000만 갤런 이상의 물이 필요하다. 이는 약 1만4,700개의 스마트폰을 만드는 데 필요한 양이다.⁴² 이렇게 사용된 물은 대수층이나 저수지, 상수도 등으로 되돌릴 수도 없다.⁴³

* 열 밀도(thermal density)는 데이터센터나 에너지 시스템 등 특정 공간 내에서 발생하는 열 에너지의 양을 의미하며, 열 관리의 효율성을 측정하는 지표로 사용된다.

에어 냉각 시스템만 가동해도 일반적 데이터센터 전력 소비량의 최대 40%가 소요된다. 따라서 데이터센터들은 액체 냉각 등 기존 에어 냉각 방식을 대체할 수 있는 방안을 모색해야 한다. 액체 냉각은 에어 냉각보

다 열전사 성질이 강해 고밀도 서버 랙 냉각 시 에어 냉각보다 전력 소비량을 최대 90% 절감할 수 있다.⁴⁴ 액체 냉각은 서버 랙을 직접 냉각하기 때문에, 반도체칩으로 밀도가 높아진 서버 랙이라 하더라도 50~100kW의 정도의 전력만 소요해도 냉각이 가능하다.⁴⁵ 또한 냉각수 생산에 일반적으로 사용되는 냉각 장치도 필요하지 않다.

액체 냉각은 이처럼 데이터센터의 에너지 절감에 큰 도움이 되지만,⁴⁶ 아직 발전 초기 단계여서 전 세계적으로 광범위하게 AI 데이터센터에 도입되지 못하고 있다.⁴⁷ 게다가 물은 유한한 자원이기 때문에, 비용과 접근성이 어떻게 개선되느냐가 향후 광범위한 상용화 여부를 결정지을 것으로 예상된다.

지속 가능 솔루션과 무탄소 에너지원 모색하는 테크 산업

빅테크 기업들은 AI 데이터센터의 무탄소 전력 공급을 가속화하기 위해 에너지 기업들과 전력 매입계약 및 장기 계약을 맺는 방식으로 재생에너지 확대 방안을 적극적으로 모색하고 있다.⁴⁸ 이러한 계약에 힘입어 재생에너지 프로젝트를 추진하는 주체들은 필요한 자금을 확보할 수 있다. 일부 테크 기업들은 전력 회사 및 혁신 기업들과 손잡고 첨단 지열 발전, 첨단 풍력 및 태양광 기술, 수력 발전, 수중 데이터센터 등 유력한 에너지 기술을 실험하고 규모를 확대하는 노력을 펼치고 있다.

일부 지역에서는 전력망 제약과 상호연결 기간 장기화 등으로 인해 새로운 재생에너지와 배터리 저장 시설을 전력망에 연결하는 작업이 지연되고 있다.⁴⁹ 수요는 많은데 전송 인프라가 불충분해 미국의 경우 관련 작업이 최대 5년 지연될 수 있다. 이에 따라 온사이트^{onsite} 또는 독립형 off-grid 에너지 솔루션을 모색하거나,⁵⁰ 장기 에너지 저장^{LDDES, long-duration energy storage}이나 소형 모듈 원자로^{SMR, small modular nuclear reactor}와 같은 신 기술에 투자하는 테크 기업들이 늘고 있다. 테크 기업들과 유틸리티 회사들이 혁신적인 청정 에너지 기술의 규모를 확대하기 위해 협업을 계획하는 경우도 있다. 이러한 움직임은 궁극적으로 여타 조직의 청정 에너지 활용에도 도움이 되고 전력망의 탈탄소화를 앞당기는 효과를 가져올 수 있다.⁵¹ 하지만 이러한 연구개발^{R&D} 프로그램과 시범도입, 여타 청정 에너지 투자 중 상당수는 수년 후에나 실질적 가치를 창출하고 투자수익을 증명하며 상업성이 가시화될 것이다.⁵² SMR의 경우 아직 개발 초기 단계여서 단기 내 무탄소 솔루션으로 활용하기가 어렵다.⁵³

미국 기업들 중에서 재생에너지를 가장 적극 도입하는 것은 테크 부문이다. 2024년 2월 28일까지 12개월 간 약 200건의 재생에너지 매입 계약 규모를 추적한 결과, 테크 기업들이 68% 이상을 조달한 것으로 나타났다.⁵⁴ 인도에서는 하이퍼스케일러와 데이터센터 운영 기업들이 태양광 도입을 주도하고 있다.⁵⁵ 테크 부문의 이 같은 움직임이 없다면, 상당수 재생에너지 프로젝트는 무산될 수 있다.⁵⁶

이처럼 청정에너지 기술의 규모 확대를 위해 테크 부문의 자본 지원이 앞으로도 중요한 요인으로 작용할 전망이다. 테크 기업들은 혁신 기업이나 재생에너지 기업들과 직접 협업하기도 하고 유틸리티 기업들과 파트너십을 맺기도 한다.⁵⁷ 혁신 기업이나 전력 산업은 테크 부문보다 자금력이 현저하게 떨어지기 때문에, 테크 산업에서 유입되는 이러한 자본이 청정 에너지 전환을 위해 매우 중요한 역할을 하고 있다.

결론

생성형AI 범용화에 따라 증가하는 데이터센터의 전력 수요를 지속적으로 충족하기 위해 테크 산업 전반과 하이퍼스케일러, 데이터센터 운영 기업, 유틸리티 회사, 규제당국들은 어떠한 노력을 기울여야 할까? 데이터센터의 에너지 이슈는 딜로이트가 2021년 분석한 클라우드 이전과 비슷한 양상을 보일 가능성이 있다.⁵⁸ 수요 촉발 요인도 다르고 변화의 속도도 한층 가팔라졌지만, 테크 산업이 지속가능성과 AI 모델 출시의 시간 경쟁 사이 균형을 잡아야 하는 과제는 동일하다. 이와 동시에 데이터센터의 에너지 수요 증가를 억제하면서 생성형AI 등 AI 모델을 위한 전력을 더욱 지속 가능한 방법으로 공급할 방법을 모색해야 한다.

1. 생성형AI 반도체칩의 에너지 효율성 개선

차세대 AI 반도체는 이미 90일만에 AI 훈련을 수행할 수 있다. 이에 필

요한 전력은 8.6GWh로, 이전 세대 반도체칩이 동일한 데이터로 동일한 기능을 수행하는 데 필요한 에너지의 1/10도 되지 않는다.⁵⁹ 반도체 회사들은 전반적인 반도체 생태계와 지속적으로 협업해 와트당 플롭스(FLOPS/watt)를 개선해 훨씬 적은 전력으로 지금보다 훨씬 많이 AI 훈련을 수행할 수 있는 반도체칩을 개발하는 데 주력해야 한다.

2. 생성형AI 활용 최적화 및 에지(edge) 기기의 프로세싱 기능 확대

AI 모델의 훈련과 추론을 데이터센터 또는 에지 기기 중 어디에서 수행하는 것이 에너지 효율적인지 판단해, 데이터센터 설비 니즈의 균형을 맞출 필요가 있다. 응답 시간이 매우 중요한 태스크를 수행하거나 민감한 데이터를 다루고 엄격한 개인정보보호가 필요한 경우 에지 기기를 선택하는 것이 나올 수 있다. 생성형AI 작업부하 중 선별된 것만 데이터센터에서 처리하고 나머지는 로컬 장치나 근거리(near-location) 및 코로케이션(co-location) 장치에서 처리하면 네트워크 및 서버 대역폭의 부담도 줄일 수 있다.⁶⁰

3. 생성형AI 알고리즘 수정 및 AI 작업부하 조정

AI 기초 모델은 규모 경쟁이 불필요할 수도 있다. 파라미터가 1조 개가 넘는 AI 기초 모델 대신 규모가 더 작은 모델만으로도 충분하고 오히려 지속가능성에 훨씬 도움이 될 수도 있다. 일부 스타트업들은 이미 클라우드 상의 에너지 집약적 컴퓨팅이 필요하지 않은 온디바이스(on-device) 멀티모달(multimodal) AI 모델을 개발하고 있다.⁶¹ 이러한 모델을 사용하는 고

객들은 AI 작업부하의 규모를 세부 조정해 실제 사업 니즈에 맞춰 기존 모델을 활용하고 필요할 때만 훈련을 수행해 맞춤형 생성형AI 모델을 활용할 수 있다. 이러한 방식으로 에너지 사용을 최소화할 수 있다. 또한 절대 지연되어서는 안 되는 실시간 추론이 필요한 특별한 경우 CPU가 비용과 사용 측면에서 훨씬 효율적일 수 있다.⁶²

4. 전략적 파트너십을 활용해 지역별 또는 클러스터형 AI 데이터센터 니즈 충족

대학교와 같이 생성형AI 데이터센터 역량을 자체적으로 갖추기가 어려운 중소 규모 조직들은 데이터센터 운영 기업이나 클라우드 서비스 기업들과 파트너십을 맺어 문제를 해결할 수 있다. 이러한 파트너십을 통해 규모가 작은 고성능 컴퓨팅 GPU 클러스터 코로케이션에 고성능 컴퓨팅 솔루션을 제공할 수 있다.⁶³ 데이터센터 운영 기업들 입장에서는 사용 현황을 동적으로 추적함으로써 잠재적 기회와 수요 현황을 파악해 즉각 코로케이션 서비스에 적용할 수 있다는 이점이 있다.

5. 여타 이해당사자 및 산업 부문과 협력해 긍정적 환경 영향 창출

하이퍼스케일러와 고객들, 외부 데이터센터 운영 기업, 코로케이션 서비스 기업, 전력회사, 현지 규제당국 및 정부, 부동산 기업들까지 다양한 생태계 참가자들이 비즈니스와 환경, 사회를 위해 실행 가능한 부분이 무엇인지에 대한 논의를 지속해야 한다.⁶⁴ 데이터센터 기업이 컴퓨팅 핑 서버 자원을 몇몇 기업에 임대할 경우 발생하는 전략적 코로케이션 니즈가

능성을 파악하는 것부터, 액체 냉각 시스템 내 적정 온도 등 냉각 니즈를 측정하고, 열과 폐수 관리 솔루션을 파악하고, 재활용 니즈를 계산하는 것까지 다양한 측면을 아우르는 협력이 필요하다. 일례로 유럽에서는 데이터센터에서 발생하는 폐열이 인근 수영장 물을 데우는 데 사용된다.⁶⁵ 전력회사들은 테크 산업과 더욱 긴밀히 협력해 데이터센터의 에너지 수요를 충족할 방법을 모색함과 동시에 이러한 파트너십을 통해 신에너지 기술 개발과 규모 확대를 위한 자금 확보 방안도 탐색해야 한다. 이러한 노력은 청정에너지의 전력망 통합 확대를 위해 매우 중요하다.

하이퍼스케일러와 전력 부문이 손을 잡고 생성형AI 데이터센터에 초점을 맞춰 무탄소 에너지원의 사용을 확대하려는 노력을 펼치면 분명 장기적으로 실질적 효과가 나타날 것이다.

02 생성형AI, 여성 소비자와 AI 인력 기반 확대 필요

여성들이 생성형AI^{generative AI}의 채택을 100% 누리려면, 테크 기업들이 AI 모델의 신뢰성을 강화하고 편향성을 축소하며 여성 AI 인력의 비중을 늘리기 위해 부단한 노력을 기울여야 한다.

여성의 실험적 또는 실제적 생성형AI 활용이 2025년 말에 이르면 미국에서 남성을 추월할 것으로 전망된다.¹ 2023년만 해도 여성의 생성형AI 활용은 남성의 절반에 그쳤으나, 여성의 활용이 빠른 속도로 증가하고 있다.² 이러한 변화 양상은 전 세계에서 공통적으로 나타나고 있다. 유럽에서 생성형AI 활용에 대한 서베이를 실시한 결과, 성별 격차가 여전히 컸으나 미국과 마찬가지로 여성의 활용이 빠르게 증가해³ 앞으로 2년 내 성별 격차가 거의 좁혀질 것으로 예상된다. 생성형AI 활용의 성별 격차를 둘러싼 과제와 기회 또한 전 세계가 미국과 비슷한 양상을 보이고 있다.

생성형AI가 빠르게 범용화되고 있지만, 생성형AI를 제공하는 기업들의 데이터 보안 방식을 신뢰하지 못한다는 비율은 여성이 남성보다 높게 나타났다.⁴ 이처럼 기술 신뢰의 성별 격차가 지속되면, 생성형AI를 일상적으로 사용하고 새로운 생성형AI 앱을 적극 사용하는 여성이 남성에 비

해 훨씬 적을 것이고 이는 향후 여성의 생성형AI 제품 및 서비스 구매도 위축시킬 수 있다. 테크 기업들은 이러한 신뢰의 격차를 극복하기 위해 데이터 보안을 강화하고 보다 투명한 데이터 관리 정책을 이행하고, 보다 광범위한 데이터 관리를 실행해야 한다.

AI 모델의 편향성 또한 신뢰에 부정적 영향을 준다.⁵ AI 인력에서 여성이 차지하는 비중은 1/3이 채 되지 않으며,⁶ AI 부문 근로자 대부분은 AI 산업이 남성 위주로 지속되는 한 AI 모델의 성별 편향성이 계속될 수 있다고 지적했다.⁷ 따라서 AI 부문의 여성 인력을 늘려야 AI 모델의 편향성을 줄이고 관련 기술의 미래 방향을 결정하는 과정에서 여성이 더욱 큰 역할을 할 수 있다.

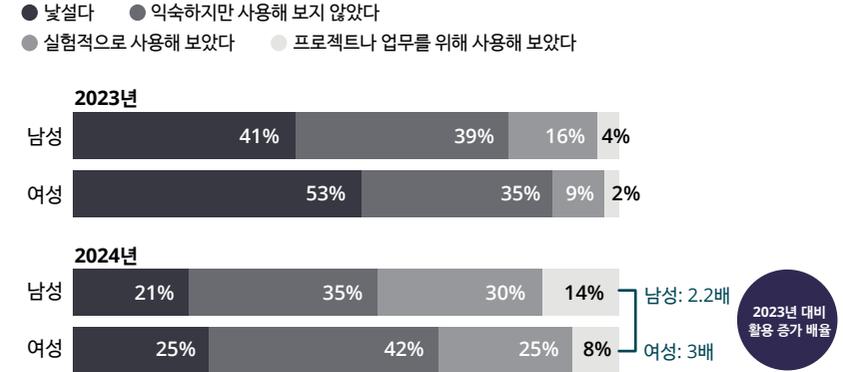
생성형AI 활용의 성별 격차, 빠르게 축소

최근 딜로이트 조사에 따르면, 다양한 지역에서 생성형AI 도입에 성별 격차가 있는 것으로 나타났다. 딜로이트가 지난 2년간 소비자들을 대상으로 디지털 라이프의 일환으로서 생성형AI 활용을 조사한 결과,⁸ 미국 여성은 남성에 비해 생성형AI 활용에 있어 뒤처지는 것으로 나타났다(그림 1). 2023년 생성형AI를 시험 삼아서 사용하는 것을 넘어 프로젝트나 업무를 위해 사용한다는 비율은 여성(11%)이 남성(20%)의 약 절반에 그쳤다. 2024년 서베이에서 전반적인 사용 비율은 두 배 이상 뛰었으나, 여성(33%)과 남성(44%) 간 격차는 여전히 여전했다.

영국 소비자 대상 서베이에서도 비슷한 성별 격차가 나타났다. 2024년 서베이에서 생성형AI를 실질적으로 사용한다는 비율이 여성 28%, 남성 43%를 기록했다.⁹ 12개 유럽국 소비자 대상 서베이에서는 두 자릿수 성별 격차가 나타났다.¹⁰

하지만 미국에서 여성이 남성을 빠르게 추격하고 있다. 2024년 생성형AI를 실질적 목적으로 사용했다는 여성의 비율이 세 배 늘어, 남성의 증가 배율 2.2배를 능가했다.¹¹ 이러한 추세가 지속되면 2025년 말에는 여성의 비율이 남성을 완전히 추격하거나 능가할 것으로 예상된다.¹²

그림 1
생성형AI 활용의 성별 격차가 여전히 존재하지만 빠르게 축소되는 중
2023~2024년 생성형AI 시스템 및 기술에 대한 인식 조사



표본 수: 미국 여성 소비자 1,040명(2023년)/1,992명(2024년); 미국 남성 소비자 962명(2023년)/1,841명(2024년)
출처: Deloitte 2024 Connected Consumer Survey, 5th edition; Deloitte 2023 Connected Consumer Survey, 4소 edition.

생성형AI에 대한 여성의 상대적 신뢰 부족, 관련 제품의 구매 위축으로 이어진다

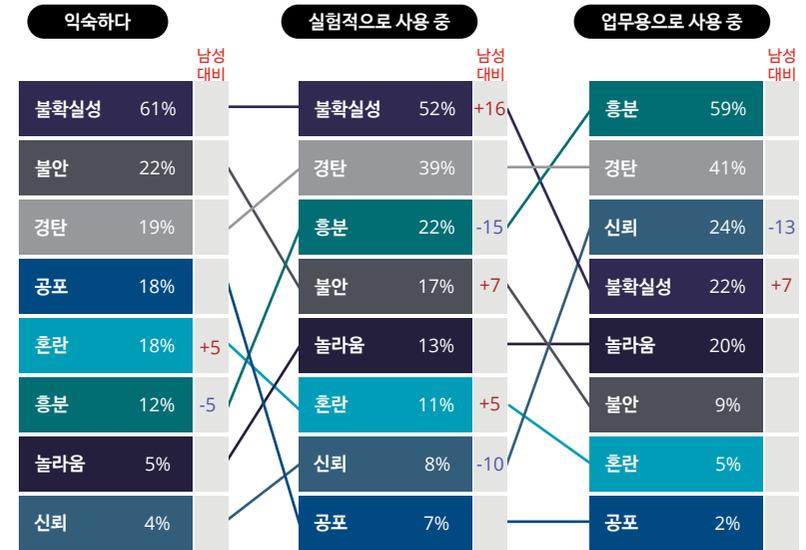
생성형AI 활용의 성별 격차가 좁혀지는 것은 고무적이지만, 그렇다고 여성들이 일상 업무에 생성형AI를 남성만큼 많이 사용한다는 의미는 아니다. 딜로이트의 2024년 서베이에 따르면, 생성형AI 사용자 중 하루에 한번 이상 생성형AI를 사용한다는 비율이 여성 34%로 남성의 43%에 비해 여전히 낮았다.¹³ 또 업무용으로 생성형AI를 사용한다는 응답자 중 생성형AI로 생산성이 대폭 개선됐다는 비율도 여성이 41%로 남성의 61%보다 낮았다.¹⁴ 테크 기업들과 생성형AI 도입으로 효과를 얻으려는 기업들은 이러한 성별 격차를 유념하고 여성의 활용을 확대하기 위해 적극적인 노력을 기울일 필요가 있다.

이 같은 성별 격차가 나타나는 일부 원인은 신뢰에 대한 인식에 큰 간극이 있기 때문이다.¹⁵ 생성형AI에 대한 여성의 인식이 ‘기술에 대한 익숙함’에서 ‘실제 실험적으로 사용해볼 수 있다’로 옮겨가면서, 불확실성과 불안, 공포, 혼란 등 부정적 감정은 줄고 경탄, 흥분, 놀라움, 신뢰 등 긍정적 감정은 증대하고 있다(그림 2).¹⁶ 하지만 실험적 사용과 업무에 실제 활용하는 경우 모두 여성은 남성보다 생성형AI에 대한 신뢰가 낮았고 불확실성의 감정은 여전히 더 높았다. 생성형AI를 사용한다는 응답자 중 생성형AI 기업이 개인정보를 안전하게 보호할 것이라는 데 ‘높은 신뢰’ 또는 ‘매우 높은 신뢰’를 가지고 있다고 답한 비율은 여성이 18%에 그쳐 남성의 31%에 비해 매우 낮게 나타났다.¹⁷

그림 2

생성형AI에 대한 여성의 부정적 감정은 줄고 긍정적 감정은 늘고 있지만, 신뢰는 여전히 남성보다 낮은 수준

생성형AI에 대해 ‘익숙하다’ ‘실험적으로 사용하고 있다’ ‘업무용으로 사용하고 있다’고 답한 여성들이 생성형AI에 대해 가장 크게 느끼고 있는 두 가지 감정을 선택하게 한 뒤 응답자 비율을 집계



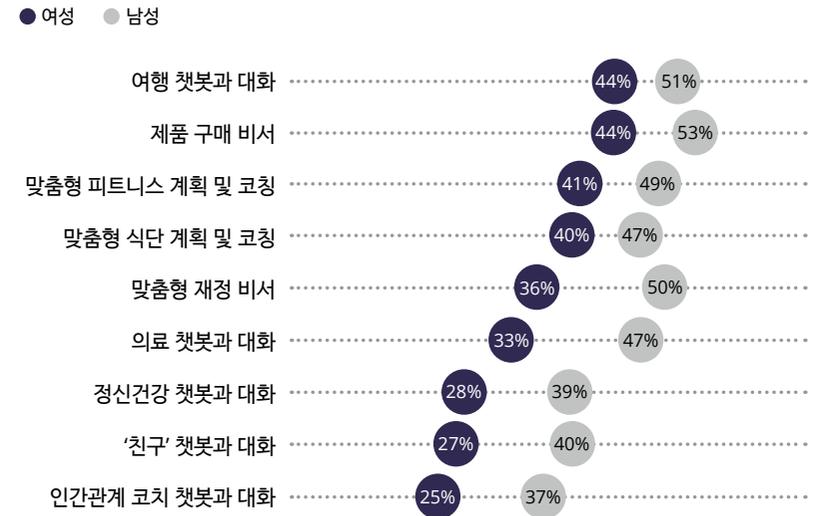
표본 수: 생성형AI에 익숙하다는 미국 소비자=여성 827명/남성 647명; 생성형AI를 실험적으로 사용 중이라는 미국 소비자=여성 496명/남성 547명; 생성형AI를 업무용으로 사용 중이라는 미국 소비자=여성 170명/남성 259명
참조: 남성 대비 수치는 5%포인트 이상 격차일 때만 표시
출처: Deloitte 2024 Connected Consumer Survey, 5th edition.

신뢰 격차는 생성형AI에만 국한된 것은 아니며 전반적인 테크 서비스 및 상호작용에서도 나타난다. 딜로이트 2024년 서베이 결과, 온라인 서비스로 얻는 이점이 개인정보보호 침해 우려를 상쇄할 만큼 많다는 여성

비율이 54%로 2023년의 46%에서 상승했음에도 불구하고, 62%를 기록한 남성보다 낮았다.¹⁸ 딜로이트 보고서에 따르면, 여성은 개인정보의 활용 및 보호 방식에 대해 남성보다 우려가 높고 이로 인해 민감한 건강 및 피트니스 정보 등 개인정보를 공유할 의향도 훨씬 낮은 것으로 나타났다.¹⁹ 여성은 보안 침해나 데이터 남용의 잠재적 피해를 남성보다 심각하게 인식하기 때문으로 풀이된다.²⁰

생성형AI의 범용화가 심화될수록 개인정보보호와 데이터 보안을 둘러싼 뿌리깊은 우려가 다시금 부각될 수 있다.²¹ 사용자가 생성형AI와 소통할 때, 시스템은 사용자의 데이터를 AI 모델에 입력할 수 있는데, 전문가들은 개인의 데이터가 AI 훈련에 활용되지 않도록 선택하기가 쉽지 않고 그 방식도 분명하지 않다고 지적했다.²² 소비자들이 생성형AI와 민감하고 개인적인 사안에 대해 소통을 시작하는 만큼, 개인정보보호와 데이터 보안이 침해됐을 때의 피해도 그만큼 클 수 있다. 뿐만 아니라 개인정보보호와 데이터 보안의 신뢰와 관련해 성별 격차가 지속됨에 따라, 다양한 생성형AI 경험에 대한 관심에서도 격차가 나타나고 있다 (그림 3).²³ 딜로이트 서베이 결과, 여행·쇼핑·피트니스·식단 등 덜 민감한 주제에 대해서조차 생성형AI와 소통하는 것에 여성이 남성보다 관심이 적은 편으로 나타났는데, 개인재정·인간관계·신체건강·정신건강 등 더욱 민감한 주제에 대해 소통하는 것은 여성이 남성보다 관심이 ‘현저히’ 떨어지는 것으로 나타났다.

그림 3
여성은 다양한 생성형AI 경험에 대한 관심이 남성보다 낮은 수준
 다음의 생성형AI 경험에 대해 ‘관심이 높다’ 또는 ‘관심이 있다고’ 답한 응답자 비율



표본 수: 생성형AI에 익숙하거나, 실험적으로 사용 중이거나, 업무용으로 사용 중이라는 미국 소비자=여성 1,492명/남성 1,454명
 출처: Deloitte 2024 Connected Consumer Survey, 5th edition.

신뢰의 성별 격차는 새로운 생성형AI 기술의 구매 양상에서도 차이를 만든다. 이제 테크 기업들은 AI 칩을 탑재해 실시간으로 정보를 요약하고 사진과 영상을 생성하고 타 언어간 즉시 번역도 해내는 임베디드 AI(embedded AI) 노트북과 태블릿, 스마트폰을 출시하기 시작했다.²⁴ 하지만 딜로이트 서베이 결과 AI 기능이 추가되는 것이 기기 업그레이드 계획에 영향을 미친다는 여성의 비율이 남성보다 낮았다.²⁵ 구체적으로, 임

베디드 AI 스마트폰이 출시되면 계획보다 빨리 기기를 업그레이드할 것이라는 남성은 43%에 달했지만, 여성은 32%에 그쳤다. 반대로 임베디드 AI 스마트폰 출시에도 기기 업그레이드 계획을 전혀 바꾸지 않겠다는 비율은 여성이 58%로 50%를 기록한 남성보다 높게 나타났다. 임베디드 AI 노트북의 경우도 마찬가지로 업그레이드 계획을 앞당길 것이라는 여성은 28%에 불과해 41%를 기록한 남성보다 낮았다. 소비지출의 약 85%(추정치)를 여성이 결정하거나 영향력을 행사하는 만큼, 테크 기업들은 AI 기기 업그레이드에 대해 여성의 관심이 떨어진다는 점을 유념할 필요가 있다.²⁶

여성의 생성형AI 활용 극대화를 가로막는 장애물은 신뢰 문제만이 아니다. 딜로이트 서베이 결과 소속 회사가 업무 시 생성형AI의 활용을 적극 권장한다는 여성의 비율은 61%로 남성(83%)보다 낮았다.²⁷ 소속 회사가 생성형AI 활용법에 대한 직원 교육에 투자한다는 여성의 비율 또한 49%로 남성(79%)보다 낮았다. 이러한 간극은 그저 인식의 문제일 수도, 직장 내 훈련 프로그램이나 정책에 실제로 성별 격차가 있기 때문일 수도 있지만, 기업들이 주의를 기울여 개선 노력을 펼쳐야 하는 사안임은 분명하다.

남성 편중의 AI 부문 인력 구조, AI 모델의 편향성 심화 리스크

하지만 테크 산업 내에서는 이러한 성별 격차가 거의 나타나지 않으므로, 테크 부문에 종사하는 여성이 생성형AI에 대한 여성 전반의 인식을 바꾸는 단서가 될 수 있다. 당연히 AI 산업에 종사하는 근로자들은 생성형 AI를 활용하는 비율이 상당히 높다. 딜로이트 서베이에 따르면, 테크 산업에 종사하는 여성 70%, 남성 78%가 생성형AI를 실험적으로 또는 업무용으로 사용하고 있다고 답했다. 이는 비테크 부문의 여성 32%, 남성 40%보다 훨씬 높은 수준이다.²⁸ 특히 테크 부문 여성은 생성형AI를 실험적으로 사용하는 데서 업무용으로 사용하는 단계로 남성보다 훨씬 빠르게 넘어가고 있다는 점이 눈에 띈다. 서베이 결과 업무용으로 사용하는 여성의 비율이 44%로 33%를 기록한 남성보다 높았다. 테크 부문에서는 남녀 모두 생성형AI로 얻을 수 있는 이점에 대한 기대가 높게 나타났다. 테크 부문 종사자 10명 중 7명은 생성형AI로 인해 1년 후 생산성이 크게 향상될 것이라고 답했다.²⁹

또한 테크 부문에서는 생성형AI에 대한 신뢰에 성별 격차가 나타나지 않았고, 남녀 모두 비테크 부문 사용자보다 높은 신뢰를 보였다. 생성형AI 기업들이 자신의 데이터를 안전하게 보호할 것이라 ‘매우 신뢰한다’ 또는 ‘극히 신뢰한다’고 답한 비율이 남녀 모두 40%를 넘었다.³⁰ 또한 온라인 서비스로 얻는 이점이 개인정보보호 침해 우려를 상쇄한다는 비율이 남녀 모두 75%로, 비테크 부문 여성 54%, 남성 60%를 훨씬 웃돌았다.³¹

테크 부문 여성은 비테크 부문 근로자들에 비해 생성형AI의 메커니즘을 더욱 잘 이해하고 이미 업무용으로 활용하는 경우가 훨씬 많기 때문에, 불안 수준이 상대적으로 낮고 생성형AI로 얻을 수 있는 이점에 더욱 초점을 맞추기 때문으로 풀이된다. 또한 테크 부문 여성은 대부분 소속 회사가 생성형AI 활용을 장려(84%)하거나 교육을 제공(72%)한다고 답해, 비테크 부문 여성(55% 및 45%)보다 높은 비율을 보였다.³²

테크 부문 여성은 이처럼 생성형AI 활용과 인식이 상대적으로 매우 긍정적이지만, AI 산업 종사자의 성별 격차는 여전하다. AI 인력 중 여성의 비율은 약 30%에 그쳐, STEM[과학science, 기술technology, 공학engineering, 수학mathematics] 부문 여성 비율과 비슷한 수준을 보이고 있다.³³ AI 부문 인력의 이 같은 성별 격차는 다양한 영역과 부문에서 AI 시스템을 개발 및 배치할 때 심각한 결과를 초래할 수 있다.

AI 인력의 성별 격차로 발생할 수 있는 가장 시급한 문제는 AI 모델의 성별 편향성이 영구화될 수 있다는 리스크다.³⁴ 범산업적으로 AI 시스템의 최대 44%가 성별 편향성을 보이고 있어, 실제로 여성을 한층 비주류화, 주변화하는 결과를 낳을 수 있다.³⁵ 예를 들어, AI 모델의 성별 편향성은 채용 방식에 편향성을 초래하거나,³⁶ 여성의 헬스케어 품질을 저하시키거나,³⁷ 여성의 금융서비스 이용을 제한할 수 있다.³⁸ 또한 딜로이트 연구에 따르면, AI 모델의 편향성은 직원과 고객의 신뢰에 부정적 영향을 주는 것으로 나타났다.³⁹ 양성 평등을 달성하고 AI로 사회 전체를 이

롭게 하려면 AI 산업에 더 많은 여성 인력을 확보하는 것이 중요하다.⁴⁰

결론

테크 산업은 다음과 같은 이유로 여성의 생성형AI 활용과 인식을 개선하는 데 많은 노력을 기울여야 한다. 첫째, 여성이 소비지출의 대부분을 결정하거나 영향력을 행사하므로, 여성이 생성형AI를 일상적으로 사용하지 않는다면, 향후 출시될 AI 제품 및 서비스가 매출 목표를 달성하지 못할 수 있다. 둘째, 여성 직원이 남성만큼 적극적으로 생성형AI 툴을 사용하지 않는다면, 기업들은 생성형AI에 투자한 만큼의 생산성 개선을 이루지 못할 수 있다. 셋째, 생성형AI는 사용자와의 대화를 통해 얻은 데이터를 수집하고 추론하기 때문에, 여성과의 대화가 상대적으로 적으면 AI 모델의 편향성이 심화될 수 있다.⁴¹ 마지막으로, 새로운 생성형 AI 활용사례에 여성의 개입이 충분하지 않으면, 여성은 헬스케어 챗봇 상담과 같이 기술이 가져올 이점을 충분히 누릴 수 없고, 기존의 불평등이 심화될 것이다.⁴²

생성형AI에 대한 여성의 신뢰를 강화하려면, 테크 기업들이 나서서 잠재적 리스크를 완화해야 한다. 딜로이트 서베이에 따르면, 테크 기업들의 개인정보보호 및 데이터 보안 정책의 투명성 개선과 소비자들의 개인정보보호 관리 용이성 개선이 소비자 신뢰에 영향을 주는 것으로 나타

났다.⁴³ 따라서 테크 기업들은 엄격한 데이터 보안 조치를 우선시하고 데이터 관리 방식을 더욱 효율적으로 설명해야 한다. 어떤 데이터가 어떻게 수집되는지 이해하기 쉬운 방식으로 소비자들에게 설명하고, 소비자들이 필요한 정보에 기반해 개인정보를 손쉽게 관리할 수 있도록 서비스를 개선하면, 신뢰를 구축할 뿐 아니라 경쟁우위도 점할 수 있다. 생성형 AI의 리스크를 완화하기 위해 정부의 역할도 중요하다. 딜로이트 서베이 결과, 기업들이 소비자 정보를 수집하고 활용하는 방식에 대한 규제가 강화돼야 한다는 응답자가 84%에 달했다.⁴⁴

각 산업의 기업들은 남녀 직원 모두 생성형AI를 충분히 활용하도록 적극 장려해야 이에 따른 이점을 실현할 수 있다. 문서 편집, 검색, 자료 요약, 조사 등 일반적인 업무용 활용사례 외에도 산업 고유의 활용사례를 적극 수용할 필요가 있다.⁴⁵ 직원들의 생성형AI 활용을 극대화하려면 교육 프로그램을 수립하는 것도 중요하다.

소비자들의 생성형AI 활용에 있어서 성별 격차를 좁히는 것만큼 AI 산업 인력의 성별 격차를 줄이는 것도 중요하다. AI 산업 내 여성의 다양성과 포용성을 개선하기 위해, 직원들의 니즈를 반영한 업무환경을 조성할 필요가 있다. 최근 AI 산업에 종사하는 여성에 대한 연구에 따르면, 직장 만족도의 가장 큰 요인으로 유연한 근무시간과 원격근무 등 일과 삶의 균형이 꼽혔다.⁴⁶ 또한 여성 인력은 여성 리더십, 임금과 승진의 투명성, 괴롭힘과 남용에 대한 무관용 정책을 갖춘 기업을 선호하는 것으

로 나타났다.⁴⁷ AI 부문의 여성 인력을 늘리려면, 여성이 AI 스킬과 경쟁력을 갖추도록 교육과 훈련 기회를 확대해야 한다. AI 부문 여성이 경험을 공유하고 상호 지원하는 멘토십 및 네트워킹 프로그램을 늘리고, 펀딩 지원을 통해 AI 연구 및 혁신 프로젝트에 여성의 참여를 늘리는 것도 도움이 된다. 생성형AI 개발에서 여성의 역할이 더욱 중요해지는 만큼, 여성의 더욱 적극적인 관심과 참여를 유도할 수 있는 활용사례와 시스템이 시급하다.

03 생성형AI 스마트폰, 세상을 바꿀 또 한 번의 격변 촉발

특화 반도체칩과 광범위한 모바일 운영체제OS 통합으로 스마트폰이 이제 스마트를 넘어 지능화된 기기로 변모하고 있다. 시장의 성장은 소비자들이 이러한 변화를 얼마나 빨리 받아들일 것인가에 달려 있다.

스마트폰은 전 세계에서 가장 많이 사용되는 소비자 기술이다.¹ 스마트폰은 여타 많은 기기들을 통합했고, 반대로 스마트폰의 소형화된 첨단 부품들은 수많은 소비자 및 산업용 기기의 등장을 촉발했다.² 손 안에서 모든 것을 처리할 수 있는 편리함과 유용성으로 인해 소비자 행태와 산업의 경쟁 양상이 일변했다. 하지만 최근 스마트폰 혁신은 시장의 흥분을 일으키지 못한 채, 혁신이라기보다 지금까지 등장한 기능의 적재에 불과하다는 평가를 받고 있다.

하지만 이처럼 동면에 들어간 듯했던 스마트폰 생태계가 생성형AI^{generative AI}를 경험의 중심으로 끌어오는 차세대 OS와 첨단 반도체칩에 힘입어 스마트폰의 정의를 재정립하고 있다.³ 생성형AI 기능을 탑재한 스마트폰이 속속 출시되며,⁴ 스마트폰 생태계가 다시 한번 요동치고 있다. 하지만 변화와 흥분은 리스크를 수반하게 마련이다.

딜로이트는 2025년 글로벌 스마트폰 출하량이 전년비 7% 증가해, 약 5% 증가했던 2024년에 비해 증가세가 가속화될 것으로 전망한다.⁵ 통상적 기기 업그레이드 주기가 접친 데다, 온디바이스^{on-device} 생성형AI를 지원하는 차세대 스마트폰에 대한 관심이 높은 얼리 어답터 소비자와 개발자들의 업그레이드 수요가 증가할 것으로 전망된다. 2025년 말에 이르면 전체 스마트폰 출하량에서 생성형AI 지원 스마트폰이 30% 이상을 차지할 것으로 예상된다.⁶

생성형AI를 둘러싼 흥분은 여전히 가시지 않고 있다. 하지만 생성형AI가 기대만큼의 혁신을 가져올지, 소비자들이 스마트폰과 소통하는 새로운 방식을 어떻게 받아들일지는 아직 미지수다.⁷

프리미엄 생성형AI 스마트폰, 업그레이드 주기 앞당길까?

주요 스마트폰 제조사들은 단기적으로 주요 스마트폰 생성형AI 통합을 통해 프리미엄 모델에 대한 수요 촉발을 꾀하려 할 것이다. 시장이 포화 지점에 이르면서, 스마트폰 판매는 2022~2023년 2년 연속 감소했다.⁸ 현재 전 세계 인구의 절반을 넘는 약 50억 명이 스마트폰을 소유하고 있는 것으로 추산된다.⁹ 게다가 최근 수년간 모델 업그레이드 주기가 길어지고 있다. 평균 업그레이드 주기는 2~3년이지만, 최근 인플레이션 압력이 높아져 소비자들의 재량지출이 제약을 받고 있다.¹⁰ 이와 동시에 몇 년간 사

용할 제품이므로 기왕이면 상위 모델을 선택하는 추세가 강해지고 있다.¹¹ 따라서 스마트폰 업계는 단순히 하드웨어 업그레이드뿐 아니라 경쟁력 있는 가치와 편리성으로 무장한 사용자 경험을 제공해야 한다.

2024년 1분기에는 스마트폰 판매가 크게 증가했다. 소비자 신뢰도가 회복된 데다 프리미엄 생성형AI 스마트폰에 대한 관심이 막 싹트기 시작한 덕분이다.¹² 딜로이트 2024년 서베이에 따르면, 커넥티드 기기 구매 시 경제적 요인의 영향을 받는다는 소비자가 줄었다.¹³ 특히 유럽은 2024년 2분기까지 스마트폰 판매가 꾸준한 증가세를 보여 이러한 서베이 결과를 뒷받침했다.¹⁴ 따라서 2025년에는 스마트폰을 교체하는 소비자가 늘고, 이들 중 상당수는 생성형AI 기능을 탑재한 고가의 프리미엄 스마트폰을 택할 가능성이 크다.

생성형AI가 스마트폰 업그레이드의 동인이 될 수 있지만, 소비자들의 양상은 시장과 연령대에 따라 다르게 나타날 수 있다. 딜로이트 서베이에 따르면, 미국 소비자 중 생성형AI 기능을 써보고 싶어서 계획보다 빨리 스마트폰을 교체한다는 비율이 7%에 그쳤으나, 연령대를 스마트폰 의존도와 신기술 수용도가 높은 24~45세로 좁히면 그 비율은 50%로 뛰었다.¹⁵ 하지만 딜로이트 영국 소비자 대상 서베이에서는 응답자 4%만이 생성형AI를 매일 사용한다고 답했고, 생성형AI가 유용하다는 응답자는 23% 그쳤으며 19%는 생성형AI의 답변에 만족하지 못한다고 답했다.¹⁶

생성형AI가 스마트폰 업그레이드 주기를 앞당길지는 생성형AI가 제시하는 가치와 유용성에 달려 있다. 2025년 스마트폰이 생성형AI의 유용성을 시험에 들게 할 매개체가 될 수 있다.

PC의 생성형AI 기능

스마트폰과 마찬가지로 생성형AI 특화 온디바이스 칩을 탑재한 차세대 PC 시장 또한 사용자 경험, 유용성, 가치, 하이퍼스케일 생성형AI의 진화를 형성하는 전반적인 압력 등이 각각도로 작용한다.

딜로이트 서베이에 따르면, 소비자들은 생성형AI 지원 PC 구매에 관심을 보였다. 미국 응답자 중 34%가 생성형AI 지원 PC로 교체하기 위해 계획보다 일찍 노트북을 교체할 의향이 있다고 답했다. 딜로이트 추산에 따르면, 연간 PC 판매량의 약 50%가 개인 소비자들에게 판매되므로, 소비자들의 이러한 행태는 매우 중요한 고려 요인이다.¹⁷ 기업 구매자들은 어떤 생성형AI 코프로세서 PC 모델을 구매해야 사업적 타당성이 가장 높을지 아직 불확실하다는 입장이다. 관련 시장에서 다양한 PC OEM사들이 다양한 옵션과 가격을 제시하고 있기 때문이다.¹⁸

향후 최첨단 PC에는 특수 실리콘칩으로 생성형AI 기능이 탑재될 것이다. 최근 연구에 따르면, 2028년에 이르면 판매되는 모든 PC 중 80%에 이러한 실리콘칩이 탑재될 것으로 전망됐다.¹⁹ 또 다른 연구에 따르면, 2024년 2분기에 AI 기능 탑재 PC 약 900만 대가 출하된 것으로 추산됐다. 하지만 이 중 생성형AI 작업부하를 운영할 만큼 강력한 신경망 처리장치(NPU)가 탑재된 PC가 얼마나 되는지는 불확실하다.²⁰ 결론적으로 잠재적 구매자들은 더욱 성능이 뛰어난 차세대 생성형AI PC의 출시를 1년여간 기다린 후 PC 교체에 나설 가능성이 크다.

딜로이트는 2024년에 판매된 PC 중 약 30%에 온디바이스 생성형AI 프로세싱 기능이 탑재돼 있을 것으로 추산하며,²¹ 2025년에는 그 비율이 50%에 육박할 것으로 전망한다.

2024년 기준 판매량이 2억6,100만 대로²² 12억3,000만 개²³를 기록한 스마트폰에 비하면 규모가 작지만, 컴퓨터는 평균 판매가가 더 높기 때문에 판매액은 판매량보다 격차가 적다. 2024년 컴퓨터 판매액은 약 2,200억 달러, 스마트폰 판매액은 5,200억 달러로 추산된다.²⁴

생성형AI PC 출시가 컴퓨터 시장에 어떠한 영향을 미칠지는 아직 알 수 없다. 다만 평균 판매가가 약 15% 상승할 것으로 예상된다.²⁵ 하지만 PC 판매는 한 자릿수 증가율에 그칠 것으로 전망된다.²⁶

스마트폰과 PC 부품이 발전하면 공급망이 형성돼 가격이 내려가고, 이러한 공급망 속에서 동일 부품이 여타 다수의 기기 시장으로도 유입된다. 생성형AI 관련 부품 또한 이 같은 경로를 밟아 여타 커넥티드 기기에서도 일반화될 것으로 기대된다.

생성형AI로 지능화된 스마트폰

스마트폰의 ‘스마트’라는 표현은 앱을 구동할 수 있는 커넥티드 기기라는 의미로 통용된다. 하지만 생성형AI가 탑재되면 스마트폰은 더욱 개인화되고 사용자의 의도와 상호작용을 인식하고 대화형 인터페이스로 더욱 친밀해질 수 있다. 음성 비서 등 이전의 개인화 노력은 기대에 미치지 못했지만, 일부 사용자들은 이미 최신 대화형 대규모언어모델LLM과 ‘관계’를 쌓고 있다.²⁷ 이러한 변화는 디지털 시스템과의 인터페이스 방식으로 대화형 AI와 새로운 상호관계의 패러다임이 될 수 있다. 이와 동시에 사용자를 대신해 행동하는 법을 학습한 신뢰할 수 있는 지능형 에이전트의 새로운 모델이 될 수 있다.

온디바이스 생성형AI 모델은 “오후 2시 약속에 늦지 않으려면 몇 시에 출발해야 하지?”라는 질문에 답할 수 있다. 사용자의 의도를 추론하고 사용자의 캘린더, 위치, 제한 시간 내 목적지까지 이동하는 최상의 경로 등을 이해할 수 있기 때문이다. 온디바이스 모델은 초당 30 테라 연산(TOPS, *tera operations per second*, 1초에 1조 번의 연산 수행)을 넘는 성능²⁸을 갖춘 NPU를 활용해 온디바이스 추론을 지원하기 때문에, 범위가 좁은 태스크를 수행하는 데 탁월하다. 만약 특정 질문이 로컬의 영역을 넘어선다고 판단하면 온디바이스 모델이 해당 태스크를 성능이 더욱 큰 클라우드 기반의 모델로 전송해 답을 얻는다. 고성능 모바일 컴퓨팅을 기반으로 클라우드의 대규모 모델에 직접 접속하는 이러한 하이브리드 방식을 활용하면, 스마트폰만으로도 더욱 즉각적이고 안전한 상호작용을 할 수 있다.²⁹

소규모 온디바이스 생성형AI 모델을 이용하면 필요 시 사용자 상호작용과 데이터를 로컬에 안전하게 저장하고, 실시간 번역 등 매우 빠른 답변이 필요할 때 지연속도를 한층 줄일 수 있다.³⁰ 이러한 경험은 사용자의 신뢰와 유용성을 끌어올리는 데 도움이 된다. 또 스마트폰 기업 입장에서는 사용자 상호작용에서 수집한 데이터를 기반으로 플라이휠 효과(*flywheel effect*)를 얻어 사용자가 더욱 만족할 만한 답변을 하도록 로컬 및 클라우드 모델을 개선할 수 있으며 더욱 심도 깊은 비즈니스 인사이트도 얻을 수 있다.

더욱 장기적으로는 소비자 상호작용의 명실상부한 중심점이 된 스마트폰이 맞춤형과 지능화를 한층 강화해 개인의 행태에 대한 이해와 개인

의 니즈에 대한 예측을 더욱 심화할 수 있다. 스마트폰이 이처럼 ‘에이전틱’^{agentic} 기기로 진화하면 단순한 ‘스마트’ 기기에서 ‘지능형’ 기기로 완전히 변모할 것이다.

2025년 한 해는 소비자들이 초기 생성형AI 기능의 가치를 시험하고 파악하며 새로운 경험을 얼마나 빨리 받아들이는지를 파악하는 전환기가 될 것이다. 스마트폰 기업들은 향후 수 개월 생성형AI 기능을 추가한 모델을 출시하겠지만, 생성형AI를 전면에 내세운 모델의 출시에는 좀 더 시간이 걸릴 것이다.³¹ 2025년은 클라우드와의 하이브리드 방식보다는 소규모 온디바이스 모델의 성능과 한계를 시험하는 한 해가 될 것으로 보인다. 이렇게 되면 생성형AI를 둘러싼 경제학이 변화할 수 있다. 비용이 많이 드는 데이터센터 대신 온디바이스로 수행되는 생성형AI 태스크가 늘어나면, 생성형AI의 자본 집약적 부담이 완화될 수 있다.

스마트폰 산업, 생성형AI에 대한 열기가 사라져도 투자를 지속할 수 있는가?

시장에서는 프런티어 모델^{frontier model}, **최첨단 고성능 AI 모델**이 뛰어난 제품 적합성^{product fit}*을 갖춰 높은 비용이 제 값을 해야 한다는 강력한 압력이 작용하는 한편, 기업들은 모델 개발과 운영을 지속하기 위해 비용 효율성을 더욱 강화해야 하는 압력을 받고 있다.³² 주요 스마트폰 기업들은 현재의 프런티어 모델을 개발하기 위해 수십 억 달러를 투자했고, 대규모 수요를

충족하는 데 필요한 데이터센터 구축에 또 수십 억 달러를 투자하고 있다.³³ 일부 연구에 따르면, 생성형AI 지원을 위해 매년 6,000억 달러가 투자되는 것으로 추산됐다.³⁴ 하지만 이러한 자본 집약적 투자는 경제적 가치를 얻지 못한다면 지속할 수 없으므로, 제품 적합성을 개선하려는 움직임이 이어질 것이다.

* 제품 적합성(product fit)은 특정 제품이 사용자에게 실제 가치를 제공하는지, 시장에서 수요와 일치하는지를 측정하는 데 사용되는 개념으로, 목표 시장, 고객의 요구, 특정 문제를 해결하는 데 얼마나 잘 부합하는지를 평가하는 데 활용된다.

우선 생성형AI 모델의 규모를 축소해 입력 데이터의 양을 줄이고, 작업부하의 범위에 기반해 데이터를 분산시키면 사용 목적에 따라 추론 태스크의 규모를 조정해 비용을 줄일 수 있다. 또한 소비자와 근로자가 요구하는 태스크 대부분이 생성형AI에 그대로 노출되므로, 더욱 비용이 적게 들고 에너지 효율적인 소규모 모델을 활용하면 이러한 문제도 해결할 수 있다.

하지만 얼마나 많은 추론 태스크를 온디바이스에서 수행하게 될지는 아직 불확실하다. 현재 소비자들의 생성형AI 상호작용과 결과에 대한 기대는 퍼블릭 클라우드 기반 모델의 수행 능력에 맞춰져 있다. 어떤 종류의 태스크와 프롬프트가 로컬에서 무료로 안전하게 수행되고, 어떤 태스크가 네트워크를 통해 클라우드로 이동해야 하는지를 소비자들이 제대로 이해하기까지는 시간이 걸릴 수 있다. 따라서 온디바이스와 클라우드를 혼합한 하이브리드 방식의 대화형 에이전트와 소통한다는 새로운 패러다임을 소비자들이 어떻게 받아들일지는 아직 미지수다.

생성형AI 스마트폰의 범용화를 가로막는 다수의 장애물

딜로이트 서베이에 따르면, 미국 소비자 중 38%가 생성형AI를 사용하고 있으며, 사용자 중 63%는 생성형AI가 기대 이상이라고 답했다.³⁵ 생성형 AI를 이미 사용하는 소비자들은 이미 마법과 같은 경험을 하고 있겠지만, 스마트폰 기업들은 새로운 모델의 높은 가격을 정당화할 수 있도록 더 많은 소비자층에 어필할 수 있는 유용성을 증명해야 하는 과제를 안고 있다.

사용자들은 이전과는 완전히 다른 새로운 상호작용을 원하기 때문에 생성형AI 스마트폰을 어떻게 활용해야 하는지를 둘러싸고 혼란이 발생할 수 있다. 사용자들은 자신의 캘린더를 스스로 관리하려는 AI 비서가 꺼림칙할 수 있다.³⁶ 이 외 배터리 소비가 빨라질 수도 있고, 퍼블릭 통합 모델 사용료가 추가로 부과될 수도 있으며, 인식하지 못한 오류로 인해 중요한 일을 망칠 수도 있다. 이러한 혼란들이 뒤섞여 사용자와 AI 에이전트, 퍼블릭 모델 사이 신뢰를 구축하는 데에는 오랜 시간이 걸릴 것이다. 하지만 이렇게 공들여 구축한 신뢰가 무너지는 것은 한 순간이다.

스마트폰 기업들은 차세대 프런티어 모델로 더욱 큰 가치가 창출될 것이라 기대하지만, 프런티어 모델의 성능이 계속 발전될지, 정체 또는 악화될지 아직 알 수 없다. 우선 갈수록 탐욕스러울 정도로 더욱 많은 데이터를 필요로 하는 AI 모델의 재훈련을 지속할 수 있는가?³⁷ AI 모델이 생성한 합성 데이터로 다시 AI 모델을 훈련시킨다는 솔루션도 제시됐지만, 이는 반

복될수록 AI 모델의 추론 품질을 저하시킬 수 있다는 문제가 있다.³⁸ 또 데이터와 훈련, 추론에 들어가는 비용을 늘리지 않고도 모델의 성능을 개선할 수 있는가? 자본과 데이터 투자가 줄어도 모델의 성능을 개선할 수 있는 방법이 있는가? 인내심을 잃은 투자자들은 AI 기술이 이러한 문제를 해결할 만큼 발전하기 전에 수익을 창출하라는 압력을 가할 가능성이 크다.

규제당국도 생성형AI의 발전에 영향을 미칠 수 있다. 딥페이크와 잘못된 정보, 인간과 흡사한 봇bot을 이용한 사기 행위 등에 대한 광범위한 안전 장치의 마련이 시급하다. 대화형 봇은 사용자와 친밀한 관계를 형성해 사용자의 생각과 이데올로기까지 변하게 할 수 있다.³⁹ 개인화 대화형 에이전트는 인간 상호작용의 더욱 깊은 영역까지 들어가, 인간에게 긍정적 도움을 줄 수도 있지만 AI와의 소통에 중독되는 사람들이 생길 수도 있다.⁴⁰ 또 온디바이스 생성형AI와 외부 모델을 결합하면 보안에 취약한 공격 노출면이 확대될 수 있다.⁴¹ 이에 따라 스마트폰 기업들의 생태계 보안 강화 및 규제당국의 안전장치 강화 노력이 강화될 것으로 예상된다.

결론

또 한 번 시장의 격변을 불러올 ‘차세대 스마트폰’이 여기저기서 거론되고 있지만 아직 현실화된 것은 아니다. 수십 억 명이 사용하는 스마트폰은 여전히 압도적인 디지털 기기이며 새로운 서비스와 상호작용을 실험

하는 대형 시험대 역할을 한다. 2025년에도 프리미엄 스마트폰과 PC를 통해 생성형AI라는 신기술과 소통하는 소비자의 수가 급증할 것으로 예상된다. 소비자들은 익숙한 기기를 통해 생성형AI를 시험삼아 사용해 보고 그 가치를 확인함과 동시에 한계를 시험할 것이다. 소비자들이 이러한 경험에 만족하면, 스마트폰은 더욱 매력적인 기기와 더욱 확장된 플랫폼으로 진화해 완전히 새로운 유형의 사용 방식과 기회를 창출함으로써 소비자 기기 시장에 새로운 붐을 일으킬 수 있다. 하지만 이러한 여정은 시간이 필요하고, 2025년은 사용자들이 새로운 패러다임에 적응하는 시간이 될 것이다.

향후 수 년간 스마트폰 OS는 다양한 상호작용을 통합할 것이다. 원격 링크 대신 로컬 모델이 정보를 요약해 제공하는 차세대 대화형 검색 기능이 탑재되는 등 서비스 제공업체와 정보 출처 간 연결고리가 점차 사라질 수 있다. 개인화된 에이전틱 AI를 사용하는 소비자들이 증가하면, 디지털 상호작용의 성격이 변모해 직접적 인터페이스 대신 사용자의 기기에서 더 많은 태스크를 수행할 수 있다. 또한 다양한 상호작용을 통합하는 스마트폰을 통해 사용자가 의식하지 않아도 주변의 컴퓨팅 환경을 통해 스스로 동작하는 앰비언트 컴퓨팅^{ambient computing}과 디지털 정보와 물리적 공간이 상호작용하는 공간 컴퓨팅^{spatial computing}이 한층 일반화될 수 있다.

스마트폰 업계가 이처럼 다시 한번 수요를 촉발하기 위한 혁신을 준비하는 가운데, 대규모 AI 모델 훈련과 운영에 투자된 자본과 에너지 비용을

상쇄해야 한다는 경제적 압박이 만만치 않다. 우선 소규모 AI 모델과 하이브리드 아키텍처부터 시도함과 동시에 어떤 생성형AI 작업부하에 어떤 종류의 계산 오버헤드^{computational overhead}*가 필요한지를 심층적으로 파악해야 한다. 한편 기후 불확실성과 우려를 피해갈 수 없는 지금, 생성형 AI 데이터센터 확장으로 이미 에너지와 물 사용량이 늘어 소비자와 정부가 부담해야 할 에너지 비용이 늘고 있다.⁴² 생성형AI의 가치가 경제적 부채를 극복할 정도로 상승했다 하더라도, 아직 에너지 부채는 갚지 못하는 실정이다.

* 계산 오버헤드(computational overhead)는 AI 알고리즘이 실제 태스크를 수행할 때 추가적으로 소모하는 시간과 메모리, 처리 능력 등 계산 자원을 의미하며, 작업 효율에 영향을 줄 수 있다.

2024년 말 현재 하이퍼스케일러^{hyperscaler}, 대규모 데이터센터 운영 업체와 스마트폰 생태계 플레이어들, 신규 퍼블릭 모델 기업들이 제공하는 이점이 경제 전반에 가치를 창출할 것이라는 전망이 우세하다. 하지만 정작 투자를 단행한 당사자들은 얼마나 많은 가치를 확보할 수 있는가? 과거 통신사들과 초기 인터넷 회사들은 막대한 자본지출을 들여 인프라를 구축해, 차세대 혁신 기업들에 동력을 제공하는 역할을 했다. 생성형AI 하이퍼스케일러들도 같은 길을 걸을 것인가?⁴³

생성형AI의 등장부터 배치, 범용화를 주도하는 것은 인터넷 범용화 이후 인류의 가장 위대한 실험에 속한다. 최종 도착지는 아직 불확실하지만, 생성형AI라는 혁신으로 수많은 신기술과 행태, 사업모델이 창출될 수 있는 여정이 시작됐다.

04 자동화의 새로운 시대 여는 에이전틱 AI

자동화 생성형AI^{generative AI} 에이전트, 즉 에이전틱 AI^{agentic AI}가 지식 근로자의 생산성을 향상하고 온갖 유형의 업무 흐름 효율성을 개선할 수 있다. 하지만 완전한 자동화가 범용화되기까지는 아직 갈 길이 멀다.

에이전틱 AI는 인간의 감독을 거의 또는 전혀 받지 않고도 복잡한 태스크를 완수하고 목표를 달성할 수 있는 소프트웨어 솔루션으로, 흔히 ‘에이전트’^{agent}로 불리는 현재의 챗봇^{chatbot}이나 코파일럿^{co-pilot}과는 다르다. 에이전틱 AI는 지식 근로자들의 생산성을 끌어올리고 사업 전반의 다단계 프로세스를 자동화할 수 있다. 딜로이트는 2025년 생성형AI를 도입한 기업들 중 25%가 에이전틱 AI를 시범 도입하거나 기술 검증에 나설 것으로 전망한다. 또한 2027년에 이르면 그 비율은 50%로 늘어날 것으로 예상된다.¹ 일부 산업이나 활용 사례의 경우 2025년부터 특정 에이전틱 AI 앱이 실제로 업무에 적용될 수 있으며, 이러한 추세는 하반기로 갈수록 뚜렷해질 것으로 전망된다.

스타트업과 기존 테크 기업들이 수익 잠재력을 기대하고 에이전틱 AI 개발에 힘을 쏟으며 이러한 추세를 뒷받침하고 있다. 지난 2년간 에이전틱

AI 개발 스타트업에 유입된 투자자본은 20억 달러가 넘는다. 투자는 기업용 시장에 주력하는 스타트업에 집중됐다.² 한편 다수의 테크 기업들과 클라우드 기업들이 자체 에이전틱 AI 서비스를 개발하고 있다. 이들은 또한 전략적 인수에 나서거나, 전면 인수 대신 스타트업들로부터 에이전틱 AI 라이선스를 구매하거나 인력을 확보하는 경우도 늘어나고 있다.³

인간의 개입 없이 스스로 태스크를 완수하는 에이전틱 AI

생성형AI 챗봇과 코파일럿은 첨단 기술이다. 인간과 직관적으로 소통하고 복잡한 정보를 가공하고 콘텐츠를 생성한다. 하지만 에이전틱 AI와 달리 자율성은 없다. 챗봇과 에이전틱 AI는 대규모언어모델LLM이라는 동일한 기초 모델을 공유하지만, 에이전틱 AI는 추가 기술이 적용돼 인간의 감독이 거의 또는 전혀 없이도 독립적으로 행동하고 태스크를 단계별로 수행해 완수한다. 에이전틱 AI는 소통하는 데 그치지 않고, 사용자 대신 사용자보다 훨씬 효율적으로 추론하고 행동한다.

이름에서 알 수 있듯이 에이전틱 AI는 ‘에이전시’^{agency}로서 행동하고 어떤 행동을 할지 스스로 결정한다.⁴ 에이전시는 독립적으로 행동하고 결정할 수 있는 자율성을 의미한다.⁵ 즉 목표를 달성하기 위해 스스로 계획하고 실행할 수 있다는 의미다.⁶ 목표는 인간이 설정하지만, 그 과정은 에이전틱 AI가 정한다.

에이전틱 AI와 코파일럿 및 챗봇의 차이는 다음과 같이 설명할 수 있다. 코드를 테스트하고 제시함으로써 소프트웨어 개발자들을 돕는 코파일럿은 지금까지 가장 성공적인 생성형AI 활용 사례로 꼽힌다.⁷ 경험 많은 소프트웨어 엔지니어들의 생산성을 끌어올리고 주니어 엔지니어들의 효율성을 개선한다. 코파일럿은 다양한 언어의 자연어 프롬프트를 코드 제안으로 전환해 코드의 일관성을 테스트한다. 하지만 엔지니어들의 프롬프트에 응답할 뿐 자율성을 가지고 있지는 않다. 하지만 에이전틱 AI와 일하는 엔지니어는 한 단계 더 나아갈 수 있다. 사람 엔지니어가 프롬프트로 소프트웨어 개발 아이디어를 입력하면, 에이전틱 AI 엔지니어가 이러한 아이디어를 실행 가능한 코드로 전환한다. 소프트웨어 개발 과정에서 다수의 단계가 자동화되는 것이다.

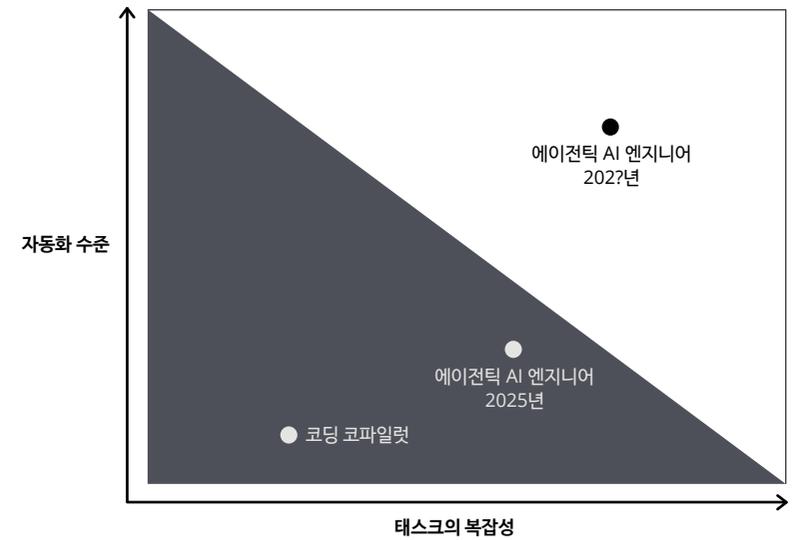
실제 사례로, 2024년 3월 코그니션 소프트웨어Cognition Software사가 수 천 가지의 결정을 내려야 하는 복잡한 엔지니어링 태스크를 추론, 계획, 완수할 수 있는 자동화 소프트웨어 엔지니어 ‘데빈’Devin을 출시했다.⁸ 데빈은 사람 프로그래머의 자연어 프롬프트에 기반해 인간의 개입 없이 앱 개발, 코드베이스 테스트 및 수정, LLM 훈련 및 재훈련 등 프로그래밍을 수행하도록 설계됐다.⁹ 2024년 여름에는 데빈의 오픈소스 버전 기업용 코드 도우미라 할 수 있는 코디움Codeium도 출시됐다.¹⁰

하지만 에이전틱 AI 엔지니어들도 코파일럿과 비슷한 취약점을 가지고 있다.¹¹ 현재 기술로는 인간의 감독 없이 완전히 자율적으로는 고사하고

일부 자율적으로라도 태스크를 맡기기에 오류가 지나치게 많다. 최근 벤치마킹 테스트에서 데빈은 실제 코드 저장소에서 가져온 깃허브GitHub 이슈 중 약 14%를 해결해, LLM 기반 챗봇보다 두 배 뛰어난 실력을 보였으나,¹² 여전히 완전 자동화라고 하기에는 턱없이 부족하다. 하지만 빅테크 기업들¹³ 과 스타트업들이 에이전틱 AI 엔지니어들의 자율성과 신뢰성을 개선하기 위해 주력하고 있는 만큼, 에이전틱 AI는 사람 엔지니어 워크로드의 일부를 처리할 수 있는 수준에 이르렀다(그림 1).

그림 1
에이전틱 AI의 발전

■ 인간의 감독이 더 많이 필요한 수준 □ 인간의 감독이 더 적게 필요한 수준



출처: 딜로이트 분석

인간의 생산성 증강하는 에이전틱 AI

에이전틱 AI 소프트웨어 엔지니어는 자율성을 지닌 생성형AI가 일의 방식을 어떻게 바꾸는지를 보여주는 한 가지 예시일 뿐이다. 에이전틱 AI가 한층 발전하면 커다란 변화를 가져올 것이다. 지식 근로자 수가 미국에서만 1억 명이 넘고, 전 세계적으로는 12억5,000만 명이 넘는다.¹⁴ 미국에서 지식 노동의 생산성을 측정하는 요소 생산성(factor productivity)¹⁵은 1987년부터 2023년까지 0.8%, 2019년부터 2023년까지 0.5% 증가하는 데 그쳐 거의 정체 양상을 보였다.¹⁶ 경제협력개발기구(OECD) 회원국들도 마찬가지다.¹⁷ 자동화를 통해 지식 노동의 생산성을 개선하려는 시도는 일부의 성공만을 거뒀다. 그리고 수많은 기업들은 여전히 지식 근로자가 부족한 실정이다. 고객 서비스 인력부터 반도체 엔지니어까지 거의 모든 지식 근로자가 부족해, 신규 인력이 투입되면 즉각 생산성을 도출해야 한다.

자동화를 위한 전문가 시스템(expert system)과 로보틱처리자동화(RPA)는 단계가 모호해지거나 여러 단계를 거쳐야 하는 경우 불안정해질 수 있다. 기존 머신러닝에 기반한 시스템은 특정 용도에 맞춰 광범위한 훈련이 필요하다. 하지만 LLM 기반의 에이전틱 AI는 머신러닝이나 딥러닝보다 유연하게 다양한 활용 사례에 적용할 수 있다.

에이전틱 AI는 LLM의 역량을 대폭 끌어올려 기업들의 생성형AI에 대한 투자 가치를 증명할 수 있다. 기업 리더들은 생성형AI 툴을 도입해 즉

각적 효과를 기대했지만, 그에 따른 비즈니스 가치를 수량화하기가 쉽지 않다.¹⁸ 데이터 기반, 리스크 및 거버넌스 정책, 인력 격차 등으로 인해 생성형AI 이니셔티브의 규모를 확대하기도 쉽지 않다. 생성형AI 시범 도입 사례 중 본격적 도입으로 이어진 것은 30%에 불과하다.¹⁹ 게다가 생성형AI가 도출한 결과물에 대한 신뢰 부족과 생성형AI의 실수가 실제 세계에 가할 수 있는 피해 등도 기업의 전면적 도입을 주저하게 만드는 요인으로 작용한다.²⁰

에이전틱 AI를 개발하고 실행하는 기업들은 생성형AI에 수반되는 이 같은 문제뿐 아니라 추론, 행동, 협업, 창작이 가능한 봇을 개발하는 과정의 복잡성도 유념해야 한다. 가장 중요한 것은 생성형AI 에이전트의 신뢰성이 보장돼야 한다는 점이다. 대부분의 경우 태스크를 제대로 완수하는 것만으로는 충분하지 않다. 2024년 말에는 생성형AI를 완전히 신뢰할 수 있다는 일련의 고무적인 활용 사례들이 나타나 2025년 초 전면 도입의 가능성이 커지고 있다.

신뢰성이 개선된 에이전틱 AI의 잠재적 이점을 생각하면 이러한 노력은 충분히 기울일 만한 가치가 있고, 초기 성과도 고무적이다. 기업들은 생성형AI 모델을 여타 AI 기술 및 훈련 기법과 접목해 LLM의 성능을 강화시키는 방법을 모색하고 있다. 완전히 자율화되고 신뢰할 수 있는 에이전트가 최종 목표이기는 하지만, 정확성과 독립성을 점증적으로 개선하는 과정에서 전반적으로 생성형AI의 생산성과 효율성을 개선하는 초

기 목표를 달성할 수 있다.²¹ 에이전틱 AI는 범산업적 솔루션으로도 산업 특화 솔루션으로도 광범위하게 활용할 수 있기 때문에, 당초 기업 리더들이 생성형AI에 기대하던 효과를 기대해볼 만하다.

에이전틱 AI의 메커니즘

에이전틱 AI는 복잡한 태스크를 일련의 단계로 정리해 실행하고, 예기치 않은 장애물도 해결할 수 있다. 또한 활용 사례에 따라 가상 환경이나 물리적 환경, 또는 두 가지 환경을 모두 감지할 수도 있다. 에이전틱 AI는 태스크를 완수하기 위해 어떤 행동을 취해야 할지 정하고, 다른 톨과 데이터베이스, 에이전트를 활용하며, 인간이 정한 목표에 기반해 결과를 제시한다.

에이전틱 AI는 계속 진화하는 신흥 기술이지만 다음과 같이 몇 가지 공통적 특징이 있다.

- **기초 모델에 기반:** 에이전틱 AI는 LLM과 같은 기초 모델에 기반해 추론과 분석을 하고 복잡하고 예측하지 못한 업무 흐름에 적응한다. 이 때문에 에이전틱 AI는 RPA나 전문가 시스템보다 유연하다. LLM의 최신 혁신 중 하나는 증강된 추론 능력과 태스크를 일련의 단계로 쪼개는 능력이다.²² 하지만 기초 모델 스스로는 여타

환경과 상호작용하고 결정을 내리고 태스크를 실행할 수 없으며,²³ 여타 기술과 역량으로 증강돼야 한다.

- **자율적 행동:** 자율성의 수준은 다양하지만 에이전틱 AI는 스스로 복잡한 태스크를 계획하고 실행하도록 훈련할 수 있다. 추론 토큰 **reasoning token***을 도입함으로써 사고의 흐름 모델 **chain of thought model****이 이전 LLM보다 복잡한 문제를 해결할 수 있다. 사고의 흐름 모델은 응답 속도가 더 느리지만, 문제 해결을 위한 추론 방식이 더욱 정교하고 오류를 스스로 수정할 수 있으며 결론을 도출한 과정을 설명할 수 있다.²⁴

* 추론 토큰(reasoning token)은 AI 모델이 특정 문제를 해결하거나 답변을 생성하기 위해 논리적 추론을 수행할 때 생성 또는 활용되는 텍스트 단위를 뜻한다.

** 사고의 흐름 모델(chain-of-thought model)은 AI와 자연어 처리(NLP)가 문제를 해결하기 위해 단계적으로 추론하는 기술을 뜻한다.

- **환경 감지:** 에이전틱 AI는 환경을 감지해 정보를 처리하고 주어진 태스크의 맥락을 이해할 수 있다.²⁵ 또 첨단 에이전틱 AI는 비디오와 이미지, 오디오, 텍스트, 숫자 등 멀티모달 데이터를 처리할 수 있다.
- **툴 사용:** 에이전틱 AI는 소프트웨어와 기업용 앱, 인터넷 등 여타 톨과 시스템과 상호작용하며 태스크를 완수한다.

- **조율:** 에이전틱 AI는 여타 시스템과 붓에 참여를 지시해 태스크를 완수한다. 이 같은 멀티에이전트 시스템으로 여타 자동화 생성형 AI 에이전트와 협업할 수 있다.
- **메모리 접근:** LLM은 ‘상태 또는 기억이 없다’^{stateless}고 알려져 있다. LLM은 각각의 상호작용을 독립적으로 처리하며, 상호작용이 완료되면 정보를 재훈련하지 않는다. 하지만 에이전틱 AI는 복구 가능한 메커니즘과 데이터베이스를 갖추고 있어 특정 태스크를 수행할 때 단기 메모리에 접근해 컨텍스트를 유지할 수 있고 장기 메모리에 접근해 과거의 경험으로부터 학습하고 개선할 수 있다.²⁶

일부 최신 에이전틱 AI 모델은 사고의 흐름 모델이 도입돼 이전 대규모 모델보다 속도는 느리지만 추론 방식이 더욱 정교해 복잡한 문제에 대해 더욱 고차원적 추론이 가능하다.²⁷ 또 멀티모달 데이터 분석이 가능해 해석 및 생성할 수 있는 데이터 종류가 확대되기 때문에 유연성이 더욱 뛰어나다. 멀티모달 AI는 컴퓨터 비전(이미지 인식)이나 전사 및 번역과 같은 여타 AI 기술과 결합할 때 더욱 강력한 성능을 발휘한다.²⁸ 에이전틱 AI와 마찬가지로 멀티모달 AI도 계속 진화하고 있다.

태스크가 자동화 에이전트의 네트워크를 오가며 수행되는 진정한 멀티에이전트 시스템이 한창 개발 중이며, 일부 파일럿 프로젝트가 2024년 말 론칭됐다.²⁹ 멀티에이전트 모델은 복잡한 환경에서 태스크를 분산하

기 때문에 싱글에이전트 모델보다 성능이 뛰어나다.³⁰ 스타트업과 빅테크 기업들은 기업들이 자체 맞춤형 에이전트를 구축할 수 있는 툴이 포함된 멀티에이전트 생성형AI 시스템을 개발하고 있다.³¹

에이전틱 AI의 유망 활용 사례

빅테크 기업들과 스타트업들은 소프트웨어 개발, 세일즈, 마케팅, 규제 컴플라이언스 등 기능을 일부 자동화하기 위한 초기 단계 에이전틱 AI 솔루션을 개발 중이다. 아직 기술 검증 단계이거나 기업용 도입 준비가 되지 않은 사례도 있지만, 에이전틱 AI는 다음과 같은 활용 사례가 유망하다. 대부분 범산업적 활용 사례이지만, 산업 특화 활용 사례도 제시되고 있다.

- **고객 지원:** 고객 서비스는 필수적이지만 스트레스가 과도한 업무이기도 해, 연간 이직률이 38%에 달한다.³² 고객 지원 업무의 일부를 효과적으로 자동화하면 인력은 스트레스가 줄고 기업은 더 많은 고객에게 서비스를 제공할 수 있다.³³ 에이전틱 AI는 현재의 고객 지원 챗봇보다 더욱 복잡한 고객 요청에 대응할 수 있으며, 자율적으로 문제를 해결할 수 있다. 일례로, 한 오디오 기업이 에이전틱 AI를 도입해 고객들의 새로운 오디오 장비 설치를 지원하도록 함으로써 통상 인간의 개입이 필요한 몇 가지 단계의 프로세스를 자

동화했다. 만약 인간의 개입이 필요하다면 에이전틱 AI가 관련 정보를 취합하고 상황을 요약해 해당 고객을 사람 직원에게 인계한다.³⁴ 다음 단계의 고객 지원 에이전틱 AI는 텍스트 기반 챗에 음성 과 비디오 등 멀티모달 데이터를 통합한 형태가 될 수 있다.

- **사이버 보안:** 사이버 보안은 전문가 부족난이 심각한 대표적 분야로, 현재 전 세계적으로 400만 명의 인력이 부족한 실정이다.³⁵ 그러는 동안 생성형AI를 활용해 사이버 보안 시스템을 침투하는 해킹이 기승을 부리고 있다. 에이전틱 사이버 보안 시스템이 개발되면 자동화를 통해 사람 전문가의 효율성이 한층 개선된다. 에이전틱 AI가 자율적으로 공격을 감지해 보고하면, 시스템 보안을 개선하고 사람 전문가의 워크로드를 최대 90% 줄일 수 있다.³⁶ 또한 소프트웨어 개발 팀을 도와 새로운 코드의 취약성을 파악해, 테스트를 수행하고 개발자와 직접 소통하면서 문제를 해결할 수 있다. 현재 사람 엔지니어들이 직접 수행하는 업무를 자동화할 수 있는 것이다.³⁷
- **규제 컴플라이언스:** 각종 산업, 특히 금융서비스와 의료 부문의 기업들은 주기적으로 규제 컴플라이언스 리뷰를 수행해야 한다. 관련 규제의 규모와 복잡성은 증대하는데 컴플라이언스 전문가 부족해 컴플라이언스가 기업들의 심각한 해결과제가 되고 있다. 이에 스타트업들이 규제를 분석하고 문서를 생산하며 기업의 컴플라이언

스를 신속하게 평가할 수 있는 에이전틱 AI를 개발하고 있다. 컴플라이언스 에이전틱 AI는 특정 규제를 인용하고 선제적 분석을 제시하며 사람 전문가들에게 조언을 제공할 수 있다.³⁸ 현재 생성형 AI를 도입한 기업들은 생성형AI 개발과 배치를 가로막는 가장 큰 장애물로 AI 인력 부족과 실행 장애를 제치고 규제 컴플라이언스를 꼽았다.³⁹ 규제 불확실성, 새로운 규제의 범위와 복잡성은 분명 큰 장애물이다. 에이전틱 AI 솔루션은 기업들의 규제에 대한 이해와 컴플라이언스를 지원해, 생성형AI 도입을 한층 가속화하는 데 도움이 될 수 있다.

- **에이전트 구축 및 조율:** 에이전틱 AI는 범산업적 또는 산업 특화 자동화 솔루션으로 부상하고 있다. 하지만 기업들은 관련 시장이 형성될 때까지 기다릴 필요 없이 자체 에이전트와 멀티에이전트 시스템을 구축할 수 있다. 구글Google의 생성형AI 모델 프로토타입을 제작 및 테스트할 수 있는 노코드no-code 툴인 버텍스Vertex를 활용하면, 이전 마케팅 캠페인에 기반해 마케팅 자료를 생산하는 등 특정 태스크를 위한 에이전트를 만들 수 있다.⁴⁰ 언어모델 기반 오픈소스 프레임워크인 랭체인LangChain을 활용하면 멀티에이전트 시스템을 구축할 수 있다. 예를 들어, 스타트업 패러다임Paradigm은 다수의 에이전틱 AI가 협업해 다양한 출처에서 데이터를 수집해 정형화하고 태스크를 완수하는 ‘스마트 스프레드시트’smart spreadsheet를 출시했다.⁴¹

결론

에이전틱 AI는 업무흐름 전반 또는 일부를 자동화함으로써 지식 근로자의 생산성을 개선할 막대한 잠재력을 지니고 있다. 싱글 에이전트로서 또는 여타 에이전트와 협업해 독립적으로 행동할 수 있는 능력은 오늘날의 챗봇이나 코파일럿과 차별화된다. 하지만 에이전틱 AI는 아직 개발과 도입의 초기 단계이다. 초기 활용 사례들이 인상적이기는 하지만, 에이전틱 AI 또한 실수를 하거나 오류의 반복 회로에 갇힐 수 있다. 멀티 에이전트 시스템에서는 ‘환각’ 현상이 에이전트 사이에서 전염될 수 있다. 다른 에이전트를 잘못된 단계로 유도해 사실과 다른 답을 내놓는 것이다.⁴² 에이전틱 AI는 주로 자율적으로 태스크를 수행하지만, 휴먼인더루프^{HITL, human in the loop}*보다 덜 제한적인 휴먼온더루프^{HOTL, human on the loop}** 방식으로 에이전틱 AI의 결정을 사람이 리뷰하는 것이 오늘날 더 현실적인 에이전틱 AI 배치 방식이 될 수 있다. 생성형AI 에이전트가 태스크 수행 과정에서 벽에 부딪혔을 때 사람 전문가에게 도움을 구해 문제를 해결하고 남은 단계를 수행할 수 있다. 이러한 모델에서 에이전틱 AI는 가치 있는 업무를 수행하면서 경험으로부터 학습할 수 있는 신입 직원과도 같다.⁴³

* 휴먼인더루프(HITL, human in the loop)는 AI가 도출한 데이터를 실제로 중요한 비즈니스 프로세스에 적용하기 전에 인간이 데이터를 직접 검증 및 수정해 생성물의 품질을 보장하는 방식을 뜻한다.

** 휴먼온더루프(HOTL, human on the loop)는 자동화 시스템이 대부분의 태스크를 독립적으로 수행하되, 인간이 시스템을 모니터링하고 필요한 경우 개입하는 방식을 뜻한다. HITL과 달리 자동화를 기본 시스템으로 하되, 특정 상황에서만 인간이 개입한다.

일부 기업들이 일관적이고 신뢰할 수 있는 에이전틱 AI를 개발하기 위해 수십 억 달러를 투자하고 있지만, 언제 어떠한 환경에서 성공할지 알 수 없다. 에이전틱 AI의 광범위한 도입이 당장 2025년에 가능할 수도 있고 5년 후에나 가능할 수도 있다. 더불어 또 다른 돌파구가 될 혁신이 필요할 수도 있고 기존의 AI 기술과 훈련 방식을 조금 수정하는 것만으로도 가능할 수 있다. 현재 에이전틱 AI를 개발하는 빅테크 기업들과 스타트업들이 성공한다면 게임의 법칙이 단숨에 바뀔 것이다. 자동화된 생성형AI 에이전트가 멀티모달 데이터를 처리하고, 툴을 사용하고 여타 에이전트를 조율하며, 과거의 경험을 기억하고 학습하며, 일관적이고 신뢰할 수 있는 방식으로 태스크를 수행할 수 있다. 더 나아가 기업들이 노코드 환경에서 단기 대화형 텍스트 프롬프트만으로도 맞춤형 생성형AI 에이전트를 손쉽게 만들고 빠르게 만들 수도 있다.

에이전틱 AI의 미래는 이처럼 강력하고 관련 기술이 빠르게 발전하는 만큼, 기업들은 다음의 접근법을 고려해 지금 당장 준비를 시작할 필요가 있다.

- **에이전틱 AI를 위한 업무 흐름의 우선순위 정립과 재설계:** 에이전틱 AI의 역량과 조직에 가장 높은 가치를 창출할 수 있는 부분에 기반해 에이전틱 AI가 수행하기에 가장 적합한 태스크와 업무 흐름을 파악하라. 불필요한 단계를 없애기 위한 업무 흐름의 재설계도 필요하다. 또한 에이전틱 AI 솔루션에게 명확한 목표를 제시하고,

태스크 수행에 필요한 데이터와 툴, 시스템에 접근할 수 있도록 하는 것도 중요하다. 에이전틱 AI는 여타 에이전트가 환경에 적응하도록 도울 수 있지만, 프로세스가 정리되지 않은 채 최적화돼 있지 않으면 실망스러운 결과를 도출할 수 있다.

- **데이터 거버넌스와 사이버 보안에 주력:** 에이전틱 AI로 가치를 창출하려면, 에이전틱 AI가 민감하지만 가치 있는 기업 데이터와 더불어 내부 시스템과 외부 자원 등에 모두 접근할 수 있어야 한다. 따라서 에이전틱 AI를 도입하기에 앞서 기업들은 엄격한 데이터 거버넌스와 사이버 보안을 마련해야 한다. 최근 서베이에 따르면, 생성형AI를 선도적으로 도입하는 기업들이 가장 주력하는 IT 투자는 데이터 관리(75%)와 사이버 보안(73%)이다.⁴⁴ 이처럼 집중적 투자에도 불구하고, 응답자의 58%는 AI 모델에 민감한 데이터를 활용하는 것과 데이터 보안 관리에 깊은 우려를 나타냈다. 또한 생성형AI 리스크와 거버넌스 관리에 충분히 준비가 돼 있다는 응답자는 23%에 그쳤다. 다시 말해 선도적으로 생성형AI를 도입하는 기업들조차 에이전틱 AI에 대한 대비가 돼 있지 않다는 의미다.
- **리스크와 보상 간 균형:** 에이전틱 AI를 처음 도입하는 기업들은 에이전틱 AI에게 허용할 수 있는 자율성과 데이터 접근성의 수준을 결정해야 한다. 처음에는 사람의 감독 하에 중요도가 덜한 데이터에 기반한 리스크가 낮은 활용 사례에 에이전틱 AI를 적용해, 안전

한 에이전틱 AI 활용을 위한 데이터 관리와 사이버 보안, 거버넌스를 구축해 나가는 것이 바람직하다. 이러한 안전망이 자리를 잡은 후에는 전략적 데이터에 기반하여 더 많은 툴과 자율성을 발휘할 수 있는 태스크에 에이전틱 AI를 적용할 수 있다.

- **건전한 회의적 시각 유지:** 에이전틱 AI는 2025년 한 해 괄목할 만한 발전을 이뤄 범산업적 솔루션 및 산업 특화 솔루션으로 활용되기 시작할 수 있다. 인상적인 시연과 시뮬레이션, 제품 공개가 2025년 내내 이어질 수 있다. 하지만 앞서 지적한 문제와 장애물은 발전 속도만큼 빠르게 해결하기 어렵다. 이러한 문제들을 해결할 때까지, 에이전틱 AI는 통제된 환경에서만 적용돼 실질적인 기업 성과 개선을 이루기 어려울 수 있다. 건전하지만 신중한 접근 방식이 필요하다.

05 딥페이크와 사이버 보안, 쫓고 쫓기는 싸움

가짜 콘텐츠를 탐지하고 대응하는 노력이 강화되면서, 소비자와 창작자, 광고주들이 신뢰할 수 있는 인터넷 환경을 유지하기 위한 비용 부담을 안게 될 수 있다.

진짜처럼 보이지만 인공지능AI 툴로 만든 사진과 비디오, 오디오를 뜻하는 딥페이크(deepfake)가 확산되면서 온라인 콘텐츠에 대한 신뢰가 추락하고 있다. AI가 만든 콘텐츠가 많아지고 한층 정교화되면서, 온라인 상의 이미지와 비디오, 오디오가 잘못된 정보를 퍼뜨리고 사기행위를 저지르는 데 악용될 수 있다. 소셜미디어에는 그러한 합성 콘텐츠가 넘쳐나, 온라인 콘텐츠를 믿지 못하고 데이터가 악용될 것을 우려하는 사용자가 증가하고 있다.¹

딜로이트의 2024년 서베이에 따르면, 소비자의 절반은 온라인 상 정보의 정확성과 신뢰성에 대해 1년 전보다 회의적 시각이 심화됐다고 답했다. 생성형AI에 익숙하거나 사용 중이라는 응답자 중 68%는 합성 콘텐츠가

속임수나 사기에 악용될 수 있다고 우려했고, 59%는 사람이 만든 미디어와 AI의 창작물을 구별하기 어렵다고 답했다. 생성형AI에 익숙하다는 응답자 중 84%는 생성형AI가 만든 콘텐츠에는 항상 분명한 라벨링(labeling)이 돼야 한다는 데 동의했다.²

라벨링은 미디어와 소셜미디어가 사용자들에게 합성 콘텐츠를 표시하는 한 가지 방안이 될 수 있다. 하지만 합성 콘텐츠를 만들고 기존 미디어를 조작할 수 있는 첨단 AI 모델이 딥페이크 기술에 적용되면서, 가짜를 탐지해내고 신뢰를 회복하기 위해 보다 복잡한 대응조치가 필요한 실정이다.

애널리스트들은 테크, 미디어, 소셜네트워크 부문 대기업들이 실행하는 딥페이크 탐지 관련 글로벌 시장 규모가 2023년 55억 달러에서 2026년 157억 달러로 연간 42% 증가할 것으로 내다봤다.³ 딜로이트는 딥페이크 탐지 시장이 사이버 보안 시장과 비슷한 경로를 밟을 것으로 예상한다. 미디어와 테크 기업들은 콘텐츠 인증 솔루션과 컨소시엄에 투자함으로써 갈수록 정교해지는 가짜 콘텐츠에 한 발 앞서기 위한 노력을 지속할 것이다. 따라서 소비자와 창작자, 광고주들이 콘텐츠 신뢰성을 지키기 위해 증가하는 비용 부담을 안게 될 수 있다.⁴

가짜 콘텐츠 대응 노력은 크게 가짜 콘텐츠 탐지와 출처 시스템 구축의 두 가지 방향으로 이뤄지고 있다.

가짜 콘텐츠 탐지

테크 기업들은 딥러닝이나 컴퓨터 비전 등 방법으로 합성 미디어를 분석해 가짜나 조작된 콘텐츠를 찾아낸다. 머신러닝 모델을 이용해 딥페이크의 패턴과 비정상성을 찾아내는 것이다.⁵ 또한 이러한 툴을 이용해 비디오나 오디오 콘텐츠에서 진짜 사람과는 다른 미묘한 입술의 움직임이라든가 목소리 톤의 변화와 같은 비일관성을 찾아낸다.⁶

일부 생성형AI 툴은 특정 콘텐츠가 생성형AI의 도움을 받아 만들어진 것인지 알아낼 수 있지만, 다른 모델이 만든 딥페이크는 탐지하기 어렵다.⁷ 일부 가짜 탐지 툴은 조작된 증거나 생성형AI의 ‘흔적’을 찾기도 한다.⁸ 신뢰할 수 있는 출처나 이미 알려진 가짜 콘텐츠의 특징에 기반해 ‘화이트리스트’^{whitelist}와 ‘블랙리스트’^{blacklist}를 구분하는 방식도 있다. 자연스러운 혈류, 얼굴 표정, 어조 등 인공과 대비되는 인간만의 특징을 찾는 툴도 있다.⁹

현재 딥페이크 탐지 툴의 정확도는 90%를 넘는다.¹⁰ 하지만 오픈소스 생성형AI 모델을 이용해 합성 미디어를 만들면 이러한 툴로도 딥페이크를 탐지하지 못할 수 있다. 콘텐츠 생성을 자동화하는 기술이 현재의 탐지 기술을 능가할 수 있으며, 생성형AI 툴이 사용자 프롬프트에 따라 생성물에 미묘한 수정을 가하면 가짜 콘텐츠를 교묘히 숨길 수 있다.¹¹

소셜미디어 플랫폼도 AI 툴을 이용해 문제의 소지가 있는 이미지나 비디오 콘텐츠를 탐지한다. AI가 의심스러운 콘텐츠에 상대 평가를 내리면, 사람이 가장 수상한 콘텐츠를 검토해 최종적으로 딥페이크를 판별한다. 하지만 이러한 방식은 시간과 비용이 많이 소요되기 때문에, 머신러닝의 도움으로 프로세스의 시간을 줄이려는 노력이 이뤄지고 있다.¹²

이는 사이버 보안 시장이 형성된 경로와 흡사한 양상이다. 보안에 주력하는 기업들이 데이터와 네트워크 보호에 계층적 접근법을 도입했던 것처럼, 미디어와 소셜미디어 기업들도 디지털 콘텐츠의 진위를 파악하기 위해 콘텐츠 출처 시스템과 더불어 다양한 툴을 활용할 가능성이 크다.

출처와 신뢰 시스템 구축

일부 기업들은 출처와 모든 수정 내역을 표시한 암호화 메타데이터^{cryptographic metadata}* 즉 디지털 워터마크^{digital watermark}**를 미디어 파일에 추가하는 방법을 모색하고 있다.¹³

*암호화메타데이터(cryptographic metadata)는 디지털 콘텐츠와 데이터의 출처 확인, 무결성 검증, 통신의 보안, 접근 권한 관리 등을 위해 사용되는 암호화 관련 정보를 뜻한다.

** 디지털 워터마크(digital watermark)는 이미지, 비디오, 오디오, 문서 등 디지털 콘텐츠의 소유권, 무결성, 출처 확인 등을 보장하기 위해 눈에 띄지 않는 형태로 삽입된 데이터를 뜻한다. 콘텐츠의 무단 복제나 조작을 방지하고, 저작권 보호 및 보안 목적으로 사용된다.

소셜플랫폼은 미디어 매체, 기기 제조사, 테크 기업들과 산업간 컨소시엄을 맺어 콘텐츠 진위 검증을 위한 영구적 표준 마련 노력을 펼치고 있다.

일례로, 딜로이트와 더불어 다양한 테크 및 미디어 기업들은 콘텐츠 출처 및 진위 확인을 위한 연합 C2PA, Coalition for Content Provenance and Authenticity에 가입해, AI가 생성한 이미지를 더욱 손쉽게 확인할 수 있는 C2PA 메타데이터 표준을 사용하기로 했다.¹⁴ C2PA의 메타데이터 기술은 최초 창작부터 모든 편집 과정까지 이미지 생성의 전 주기(전주기)를 상세 로그로 기록한다.¹⁵ C2PA 기록을 해독하면 콘텐츠 매체와 사용자들은 해당 이미지의 출처와 진위 여부를 확인할 수 있다.

일부 소셜플랫폼은 창작자들을 대상으로 검증 옵션 시스템을 마련하기도 했다. 가짜가 아님을 증명하는 검증서를 제출하는 창작자들에게 수수료를 지불하는 방식이다. 창작자들을 수익 공유 프로그램에 참여토록 해 검증 활동에 인센티브를 제공할 수도 있다.¹⁶

각종 온라인 채널에서 AI 콘텐츠가 확산되는 만큼, 계정의 운영자가 사람인지 AI인지를 파악하는 것도 신뢰 악화를 막는 데 도움이 된다.¹⁷ 다만 이에 따른 비용 부담을 창작자와 광고주, 사용자들에게 계속 전가할 수 있는지는 고민해야 할 부분이다.

딥페이크 대응 규제

일부 국가에서 콘텐츠 진위 관련 규제 조치가 마련됐지만,¹⁸ 더욱 포괄적

인 글로벌 규제가 필요하다. 딥페이크의 위험성에 대한 대중의 인식을 제고하는 것도 중요하다. 이를 위해 대중이 가짜와 진짜 미디어를 구분할 수 있는 정보를 제공할 필요가 있다.

미국에서는 연방 상원 상업·과학·운송 위원회가 AI 생성 콘텐츠에 디지털 워터마크를 의무화하는 법안을 검토 중이다.¹⁹ 또 캘리포니아 주 정부는 AI 콘텐츠의 라벨링을 의무화하는 법안 ‘AB 3211’을 상정했다. 이에 따르면, 기기 제조사들은 사진에 출처 메타데이터가 표시되도록 기기를 업데이트해야 하고 온라인 플랫폼들은 온라인 콘텐츠의 출처 메타데이터를 공개해야 한다. 법안이 통과되면 2026년부터 실효될 예정이다.²⁰ 여타 주 정부도 잘못된 정보를 퍼뜨리려는 의도로 동의 없이 만들어진 딥페이크의 제작과 배포를 불법으로 규정하는 법률을 제정했다.²¹ 미국 연방거래위원회 FTC는 개인을 모방한 딥페이크의 창작과 유포를 금지하는 새로운 규제를 마련 중이다.²²

유럽연합 EU는 인공지능법 AI Act 수정안에서 투명성을 강조하며 AI 생성 콘텐츠와 딥페이크의 라벨링 의무를 명시했다. 이는 AI 기술의 발전을 지원하면서도 사용자들이 콘텐츠의 진위를 파악할 수 있도록 하기 위함이다. 또 EU 집행위원회는 AI의 발전과 활용을 지원하고 AI가 창작 또는 조작한 콘텐츠의 효과적인 라벨링을 장려하기 위해 AI 사무국 AI Office을 수립했다.²³

딥페이크 기술의 급격한 발전 속도에 발 맞춰 유연하고 발 빠르게 대응할 수 있는 규제 프레임워크가 시급하다.

결론

미디어 매체와 소셜플랫폼은 사진과 비디오, 오디오 클립의 진위를 파악하기 위해 콘텐츠 분석과 출처 확인 두 가지 기술을 개발하는 데 투자를 지속할 것이다. 생성형AI를 활용해 더욱 많은 합성 미디어가 확산되고 딥페이크 생성 모델과 탐지 회피 기술도 빠르게 발전하는 만큼, 이러한 투자는 갈수록 중요해질 것이다.

특히 생성형AI의 능력이 한층 강력해지고 다각화되고 있으므로, 딥페이크 기술보다 한 발 앞서는 것이 중요하다. 최근에는 혈액량 탐지 및 안면 분석 등의 첨단 기술이 진짜와 가짜를 구분하는 데 활용되고 있다. 하지만 사이버 보안 톨과 마찬가지로 이러한 딥페이크 탐지 기술이 최종 사용자와 소비자들에게 불편을 초래해서는 안 된다. 사용자 경험을 침해하지 않는 최선의 방식으로 콘텐츠의 무결성을 보장할 필요가 있다. 현재로서는 사용자 경험에 부정적 영향을 주지도 않고 분석을 위한 실시간 컴퓨팅 주기가 없어도 되는 디지털 워터마크가 가장 효율적 수단이다.²⁴

딥페이크 탐지에 머신러닝 모델이나 외부 업체를 이용하는 기업들은 이미지와 비디오, 오디오를 아우르는 다양한 고품질 데이터 세트를 활용하는 톨과 업체를 우선적으로 선택할 필요가 있다. 또한 이러한 데이터 세트는 다양한 인구층을 통합해 공정성을 보장하고 편향성을 최소화하는 것이 중요하다.²⁵

테크 및 미디어 기업들은 범산업적 협력을 통해 딥페이크 탐지와 콘텐츠 진위 검증을 위한 표준 수립 노력을 기울여야 한다.²⁶ 일례로, 기기 제조사와 미디어 기업이 창작물과 발행물의 진위를 공동 검증하는 경우 디지털 워터마크가 여타 방식보다 효과적일 수 있다. 이처럼 공동의 노력이 펼쳐지면 더욱 강력하고 포괄적인 기준이 수립돼 디지털 콘텐츠 전반의 보안과 신뢰가 개선될 것이다.

각 산업의 기업들은 생성형AI로 더욱 위협적인 소셜 엔지니어링 공격이 가능해져 일부 가짜 탐지 조치를 무력화할 수 있음을 염두해야 한다.²⁷ 따라서 비디오와 오디오 기반 프로세스의 경우 추가 검증 계층을 이행할 필요가 있다. 최종 사용자들 사이에서도 신뢰할 수 있는 출처에 기반해 정보를 대조 검토하고 다중 인증 방식을 이용해 딥페이크 관련 리스크를 완화하는 노력이 자리잡아야 한다. 따라서 기업들은 계속 진화하는 딥페이크 기술에 발 맞춰 사용자 대상 교육을 지속하는 것이 중요하다.

테크 및 미디어 기업들이 이러한 전략을 충실히 이행하면 딥페이크 기술이 가하는 위협을 방지함과 동시에 디지털 콘텐츠의 무결성과 신뢰를 유지하는 리더로서의 역할을 다질 수 있다. 2025년은 디지털 세계의 불확실성이 증대하는 가운데 기업들이 신뢰할 수 있는 콘텐츠 환경을 만들어 스스로 신뢰할 수 있는 콘텐츠 출처로 자리매김하는 중요한 기로가 될 수 있다.

Chapter 02

미디어·엔터테인먼트, 첨단기술로 가속화되는 시장 생태계 변화

- 01 꿈의 경기장 건설 프로젝트, 공공-민간 투자의 교두보
- 02 비디오 콘텐츠 시장, DTC 지고 애그리게이터 뜬다
- 03 생성형AI, 콘텐츠 제작과 배급의 경쟁 판도 재편

01

꿈의 경기장 건설 프로젝트: 공공-민간 투자의 교두보가 되다

스포츠 구단주들이 경기장을 관광 명소로 변모시켜 사회·경제적 성장과 더불어 지역사회 참여 및 수익 다각화를 이끌고 있다.

스포츠 인프라에 대한 투자가 증가하고 있다. 경기장 훈련 및 경기 시설 등 스포츠 인프라 개발로 공공과 민간 부문에서 광범위한 사회적, 경제적 이익이 창출되는 경우가 많아 주목받고 있다. 최근 몇 년간 북미, 유럽, 아시아 태평양 지역의 스포츠 팀들은 인프라 개발에 막대한 투자를 해왔으며, 이러한 추세는 더욱 뚜렷해지고 있다.

정부와 지역사회는 스포츠 산업 투자자들과 협력해 공동된 성장 목표를 향해 나아가고 있다. 교통망 등 보조 인프라를 지원함으로써 지역사회의 자원은 스포츠의 사회경제적 영향을 더욱 확대할 수 있다. 2025년 한 해 다양한 스포츠 인프라 프로젝트가 추진돼, 지역 경제에 긍정적인 영향을 미치며 문화와 사회 발전의 구심점으로서 스포츠의 핵심적 역할이 한층 공고해질 것으로 전망된다.

딜로이트는 2025년 한 해 전 세계적으로 300개 이상의 스포츠 경기장이 신규 착공되거나 리노베이션을 거칠 것으로 예측한다. 딜로이트가 스포츠 인프라 개발 현황을 분석한 바에 따르면, 이러한 신규 경기장 프로젝트의 약 50%가 북미와 유럽에서 진행될 것으로 파악됐다.

유럽 전역에서는 축구 경기장을 중심으로 투자가 증가하고 있다. 새로운 팬 층을 끌어들이고 수익 다각화 기회를 모색하려는 노력의 일환이다. 경기장 개발은 민간 투자자들이 투자 수익을 극대화하고 공공 부문이 사회경제적 목표를 달성하는 데 중요한 역할을 할 수 있다. 북미와 유럽 외에도 세계 곳곳에서 스포츠 중심의 지역 활성화 프로젝트가 활발히 진행되고, 팬들이 경기장 안팎에서 혁신적인 경험을 요구하기 시작하면서, 전 세계적으로 경기장에 대한 투자가 확대될 가능성이 크다.

지역사회 중심의 장소 개발 접근법

스포츠 조직은 지역사회를 하나로 묶고, 시민들의 자부심과 결속감을 높이며, 도시의 문화적 다양성을 더욱 확장하는 중요한 역할을 한다. 스포츠 경기장을 중심으로 한 스포츠 주도 재생 프로그램을 성공적으로 실행하려면 정부 및 주요 이해당사자들과 협력하여 전략적인 이니셔티브를 지원하고, 지역사회와 소통하며, 사람들이 살고 싶고 방문하고 싶은 활기찬 지역이 되도록 지속 가능한 개발을 이루어야 한다.

스포츠 경기장 개발은 더 이상 단일 구단의 이익에만 초점을 맞출 수 없다. 경기장을 신설 또는 보강할 때 지역사회의 이익이 다양한 의사결정 과정에 반영되어야 한다.

2024년 4월, 잉글랜드 프리미어리그EPL 소속 버밍엄 시티Birmingham City, 별칭 'The Blues'의 구단주인 나이트헤드 캐피털 매니지먼트Knighthead Capital Management는 세계 최상급 경기장을 중심으로 스포츠 특화 구역Sports Quarter을 개발하는 계획을 발표했다.¹ 당사의 톰 와그너Tom Wagner 회장은 도심에서 도보로 접근할 수 있는 위치에 경기장과 더불어 남녀 훈련 시설과 청소년 아카데미 팀 등 시설을 모두 갖춘 특화 구역을 조성한다는 야심 찬 비전을 제시했다.² 또한 몇몇 호텔과 상업 기업들이 해당 특화 구역에 지점을 내고 버밍엄 동부 지역 활성화 참여하는 데 관심을 보였다며 이들과 협의할 의향을 표명했다.³ 와그너는 이를 통해 '더 블루스'The Blues가 버밍엄의 정체성을 형성하는 요인으로 완전히 자리잡고, 전 세계적으로 인정받는 '탁월성의 상징' Beacon of excellence이 될 것이라고 말했다.⁴ 스포츠 특화 구역 프로젝트는 20억~30억 파운드의 비용이 소요될 것으로 예상되며, 웨스트 미들랜즈West Midlands 지역 사회에 장기적인 사회경제적 영향을 미칠 것으로 기대된다.⁵ 나이트헤드 소유 그룹은 정부와 공공 부문과 협력하여 다양한 전략적 우선 과제를 추진하고 있다.⁶

미국 메이저리그MLB의 탬파베이 레이스Tampa Bay Rays는 2024년 7월 플로리다주 세인트피터스버그St. Petersburg 시와 야구장 신설 계약을 체결했다.⁷

해당 개발 그룹은 1,250세대의 저렴한 주택 공급, 30,000개의 건설 일자리, 7,000개의 상시 일자리 창출을 약속했으며, 일자리 일부는 지역 주민과 소외 계층에 할당될 것이라고 설명했다.⁸ 프로젝트 리더들은 경기장 신설을 통해 인근 지역의 세대 간 부의 격차를 해소하는 데 전념하고 있다며, 이러한 목표를 달성하지 못할 경우 프로젝트를 '실패'로 간주할 것이라고 강한 의지를 표명했다.⁹

세대별 선호도를 반영한 팬 참여 강화

세대마다 스포츠를 즐기는 방식은 다르다. 딜로이트의 2024년 보고서에 따르면, 글로벌 스포츠 산업 리더 중 84%는 향후 5년 동안 가장 큰 영향을 미칠 차세대 트렌드로 달라진 소비 선호도를 꼽았다. 이에 따라, 스포츠 조직은 경기 당일의 축제 분위기와 응원 등 전통적 팬 경험의 핵심 가치를 유지하면서도 이전 세대보다 높은 수준의 엔터테인먼트를 기대하는 Z세대와 알파세대의 수요도 충족하기 위해 균형을 찾아야 한다.¹⁰

경기장 경험을 개선하는 첫 번째 단계는 편안함과 안전, 경기장 시야 확보, 경기력 수준, 열광적인 분위기 등 기본을 잘 갖추는 것이다. 이러한 기본 요소는 대부분의 팬이 여전히 중요하게 여기기 때문에 더욱 고도화된 경험을 제시하기에 앞서 완벽하게 구현되어야 한다.

다음 단계로 팬들이 경기 전부터 도중, 관람이 끝난 후까지 즐길 수 있는 다양한 엔터테인먼트 옵션을 제공해 차별화된 경험을 제공할 수 있다. 이는 경기장에서 팬들이 더 많은 시간과 돈을 소비하게 유도할 뿐만 아니라, 해당 스포츠 조직을 중심으로 강한 공동체 의식을 형성하는 데에도 도움이 된다. 또한 경기장에 지역사회의 문화를 통합하면 경기 당일 팬 경험을 더욱 독창적이고 지역 특화된 형태로 만들 수 있다.

MLB의 토론토 블루제이스(Toronto Blue Jays)는 홈구장 로저스 센터(Rogers Centre)에서 팬 경험을 향상시키기 위해 두 단계에 걸쳐 경기장을 리노베이션했다. 2023년에 완료된 1단계를 거쳐 외야 관중석 내에 현지 요리와 엔터테인먼트를 제공하는 다섯 개의 독특한 ‘커뮤니티 존’을 만들어 팬들에게 차별화된 경험과 소셜 공간을 제공하고 있다.¹¹ 또한, 새로운 자동화 식음료 마켓 서비스 ‘탭앤고’(Tap N Go)와 셀프서비스 음료 판매대 ‘워크스루브루’(Walk thru Bru) 등 업그레이드된 디지털 기술을 도입해 서비스 속도를 개선했다.¹² 2단계에는 홈페이지 방향으로도 좌석을 조정해 시야를 개선하는 등 팬들에게 더 몰입감 있고 편리한 경기장 경험을 제공하기 위한 리노베이션이 포함된다.¹³

스마트 경기장: 차세대 스포츠 경험의 핵심

디지털 중심의 스포츠 소비 경향을 보이는 차세대 팬들은 경기 당일 경

기장 현장에도 디지털 중심의 팬 경험을 원한다.¹⁴ 이러한 추세에 맞춰 일부 스포츠 조직은 첨단 기술을 접목해 팬 경험을 개인 맞춤형 ‘스마트 경기장 지구’(smart stadium district)를 기획하고 있다.¹⁵ 전 세계 스마트 경기장 시장은 2024년 80억 달러, 2033년 380억 달러를 각각 넘어설 것으로 예상된다.¹⁶

Z세대와 알파세대는 이전 세대보다 짧고 역동적인 콘텐츠를 선호하기 때문에 팬 참여를 이끌어내는 데 필요한 요건도 변화하고 있다.¹⁷ 가치를 추구하며 물질적 소유보다 경험에 더 많은 비용을 지불하려는 추세가 강해지는 만큼, 팬 경험 자체가 스포츠 조직의 성패를 좌우할 차별화 요인이 되고 있다.¹⁸ 이에 발맞춰 신규 경기장은 게임을 하는 듯한 경험, 상품 판매, 등 다양한 요인을 통합해 설계되고 있다.

특히 ‘세컨드 스크린 신드롬’(팬들이 스포츠 시청 중 보조 화면을 통해 다양한 정보를 동시에 확인하는 행동 패턴)을 반영해 경기장 구조도 새로워지고 있다. 딜로이트 조사에 따르면, 스포츠 팬의 77%는 경기 관람 중 경기 관련 추가 활동(팀 및 선수 통계 검색, 소셜 미디어 사용, 스포츠 베팅 등)을 최소 한 번 이상 수행한 경험이 있다고 밝혔다.¹⁹ 새로운 경기장들은 최첨단 기술을 접목해 경기장 내에서 이처럼 팬들의 관심을 끌만한 콘텐츠를 방송하며 팬들의 관심을 경기장 내에 붙들어 두고 있다.²⁰

미국프로농구(NBA) 로스앤젤레스 클리퍼스(Los Angeles Clippers)는 팬 경험을 최

우선시한 설계로 새로 지어진 인튜이트 돔(Intuit Dome)을 개장했다. 새 경기장에서 단연 눈에 띄는 것은 맞춤형 제작된 역대 최대 규모의 양면 디스플레이 ‘헤일로 보드’([Halo board](#))이다. 센터 코트 천장에 후광처럼 떠 있는 헤일로 디스플레이는 어떤 좌석에든 최적의 시야로 볼 수 있으며, 무엇보다도 코트에서 멀리 떨어진 상단 관중석에서도 생생한 플레이를 직접 관람하는 효과를 준다. 뿐만 아니라 실시간 점수와 선수 교체 등 경기 실황 정보, 심층 통계가 담긴 ‘코치 코너’([coaches corner](#)), 주요 장면 다시 보기, 선수 사진과 기본 프로필, 구단주 스티브 발머([Steve Ballmer](#))를 따라다니는 ‘스티브 캠’([Steve cam](#))까지 다양한 콘텐츠를 실시간으로 제공한다. 또한, 상단 관중석 팬들에게 티셔츠를 발사하는 티셔츠 캐논을 헤일로 보드에 설치해 팬 경험의 재미를 한층 더했다. 팬들은 환호와 응원을 보내면 보상을 받고, 각 좌석에 설치된 게임 콘솔로 현장에서 게임을 하는 듯한 엔터테인먼트 경험을 즐길 수 있다. 이처럼 인튜이트 돔은 팬 경험을 ‘게임화’하여 전례 없는 수준의 팬 참여를 이끌어내고 있다.²¹

차세대 팬들은 끊임없고 초개인화된 온디맨드([on-demand](#)) 경험을 기대한다. 따라서 경기장이 위치한 지역은 다양한 음식, 음악, 문화, 소셜 공간을 갖춰 등 다양한 유형의 팬들이 더 오래 머물며 즐길 수 있는 장소가 돼야 한다. 이에 스포츠 조직들은 이러한 공간을 기획하고 설계하여 지역 주민과 외부에서 방문한 팬들에게 프리미엄 경험을 제공하며 이들이 서로 소통할 수 있는 새로운 접점을 형성하고 있다.²²

이 때 다양한 신기술이 큰 도움이 된다. ‘클릭앤고’([Click N Go](#)) 시스템으로 상품과 음식을 간편하게 주문할 수 있고 팬 개인 맞춤형 메시지를 담은 소프트웨어 티켓도 등장했다.²³ 팬들이 경기장에 입장하는 방식도 진화하며, 팬들의 편의성이 향상되고 있다. 예를 들어, 메르세데스-벤츠 스타디움([Mercedes-Benz Stadium](#))은 델타 항공과 협력해 안면 인식으로 빠른 입장이 가능한 ‘플라이 스루 레인’([Fly-through lanes](#))을 도입했다.²⁴

인프라와 기술을 활용한 수익 다각화

스포츠 구단주들은 강화된 인프라와 디지털 기술을 활용하여 수익 창출 방식을 더욱 다각화하고 있다. 스포츠 조직의 성공 방식이 변화하는 것이다. 과거 글로벌 스포츠 조직들은 통상 방송 중계권 수익에 크게 의존하여 인건비와 기타 운영비를 충당했다.²⁵ 중앙방송 중계권 수익이 증가하면, 북미 스포츠 리그의 연봉 상한선이나 유럽 스포츠의 비용 통제 규정에 따른 지출 한도가 늘어나는 식이다.²⁶ 그러나 단일 수익원에 지나치게 의존하면, 수익 잠재력에 한계가 생겨 코로나19([COVID-19](#)) 팬데믹과 같은 예기치 못한 시장 충격에 매우 취약해질 수 있다.

이와 비교해 지금은 일부 엘리트 스포츠 구단들이 이미 보유한 막대한 문화 및 상업 자원을 적극적으로 수익 증대에 활용하고 있다.²⁷ 입장권 판매와 중계권 수익은 여전히 개별 구단이 수용 또는 협상하기에 한계가 있지

만, 상업 수익은 구단이 자체적으로 성장시킬 잠재력이 크다.²⁸ 경기장을 더욱 넓은 범위의 엔터테인먼트 지구로 변모시키면 스포츠 조직이 상업적 입지를 확장하고, 다양한 서비스를 통해 수익 기반을 확대하는 데 중요한 역할을 한다.

예를 들어, EPL의 토트넘 홋스퍼 [Tottenham Hotspur](#)는 신규 경기장을 활용하여 상업 수익을 2016~2017 시즌 7,200만 파운드에서 2022~2023 시즌 2억2,700만 파운드로 크게 늘렸다. 이는 공식적인 리그 경기와 더불어 NFL 경기와 콘서트와 같은 다양한 축구 외 이벤트를 개최한 덕분이다.²⁹

유럽 리그 최고 명문 구단인 레알 마드리드 [Real Madrid](#)는 2023년 산티아고 베르나베우 [Santiago Bernabéu](#) 경기장 리노베이션 이후, 2022~2023 시즌에 역대 최고 수익을 기록했다. 방송권 수익(2024년 현재 협상 진행 중)을 제외한 모든 사업 부문에서 성장을 이룬 것이다.³⁰ 2024년 7월, 레알 마드리드는 축구 구단 역사상 최초로 연간 10억 유로 이상의 수익을 창출했다고 발표했다.³¹ 이는 주로 주요 이벤트를 개최하고 프리미엄 VIP 팬 경험을 도입하는 등 모험을 무릅쓴 덕분에 분석된다.³² 레알 마드리드는 콜롬비아 팝 가수 캐롤 G [Karol G](#)의 4회 콘서트를 통해 약 1,800만 유로의 수익을 올리기도 했다.³³ 구단은 다가오는 시즌 동안 경기장 리노베이션 작업을 완료할 예정이며, 향후 축구 외 수익을 더욱 늘릴 계획이다.³⁴

이처럼 스포츠 구단들이 인프라와 기술을 활용해 수익 다각화를 추구하

는 것은 단순한 재정적 성장에 그치지 않고, 장기적인 안정성과 예기치 못한 미래의 시장 충격에 대한 회복력 또한 제공한다.

스포츠 인프라의 새로운 트렌드

스포츠 조직들은 여성 스포츠의 성장에 대응하고 더 큰 지속 가능성 향상을 주도하는 등 새로운 팬 경험을 수용하고 우선순위를 재정립하기 위해 노력하고 있다. 이러한 스포츠 인프라의 진화는 스포츠 산업의 포용성, 혁신 책임감을 강화하는 방향으로 나아가고 있다.

여성 스포츠를 위한 인프라

여성 스포츠의 인기와 가치가 계속해서 상승함에 따라, 스포츠 조직들은 여성 스포츠 전용 인프라 구축에 주력하고 있다. 미국 여자 프로축구 리그 [NWSL](#)의 캔자스시티 커런트 [Kansas City Current](#)는 1억1,700만 달러를 투자해 리버프론트 경기장을 세계 최초 여성 프로스포츠 팀 전용 경기장 CPKC 스타디움으로 탈바꿈했다. 이 경기장의 사회경제적 효과를 강화하기 위해, 아파트, 호텔, 레스토랑, 상점들이 포함된 6억5,000만 달러 규모의 복합 용도 개발 계획도 승인됐다.³⁶ CPKC 경기장 건설은 캔자스시티에 연간 약 5,000만 달러의 경제적 파급효과를 낼 것으로 예상된다.³⁷

미국 여자 프로농구 리그 [WNBA](#)에서는 새로운 훈련 시설들이 선수들의 경

기력을 향상시키고 있으며, 이는 각 구단의 전반적인 가치 평가액 증가로 이어지고 있다. 라스베이거스 에이스즈Las Vegas Aces는 2023년에 자체 훈련 시설을 개장했고, 시애틀 스톰Seattle Storm, 피닉스 머큐리Phoenix Mercury, 시카고 스카이Chicago Sky도 뒤따라 전용 훈련 시설을 개장했다.³⁸

영국에서는 여자 프로축구 구단인 브라이튼 앤 호브 알비온Brighton & Hove Albion이 팀을 위한 전용 경기장 건설을 계획하고 있으며,³⁹ 맨체스터 시티Manchester City 여자 구단도 클럽의 시티 풋볼 아카데미City Football Academy 부지에 전용 훈련 시설을 건설할 계획을 승인받았다.⁴⁰

프리미엄 및 맞춤형 호스피탈리티(환대) 서비스

호스피탈리티Hospitality 분야는 전통적인 기업 서비스의 범위를 넘어, 모든 인구층이 쉽게 접근할 수 있고 차별화된 경험을 즐길 수 있는 중요한 도구로 자리잡고 있다. 스포츠 조직들은 호스피탈리티 산업의 고급 브랜드와 협력하여, 유명 셰프의 음식부터 굿즈에 이르기까지 프리미엄 호스피탈리티 서비스를 제공하려는 노력을 기울이고 있다.⁴¹ 예를 들어, 포뮬러 원F1의 패독 클럽Paddock clubs이 대표적인 사례다. F1 VIP를 위한 공간인 패독 클럽에서는 소셜 인플루언서, 유명 인사, 브랜드 파트너들이 프리미엄 호스피탈리티를 경험할 수 있다.⁴² 또한, 스포츠 조직들은 경기장 신설 및 개조를 통해 호스피탈리티 공간을 더욱 유연하게 활용해 다양한 유형의 이벤트를 개최할 수 있다.

지속 가능성

새로운 스포츠 인프라 건설 프로젝트의 설계 단계에 지속 가능성 원칙이 반영되는 경우가 많아지고 있다. 스포츠 주도의 재생 프로젝트에서 지속 가능성에 초점을 맞추면 긍정적인 환경적 및 사회적 관행을 시연함으로써 공공 자금을 더욱 손쉽게 확보할 수 있다.⁴³ 지속 가능성은 지역사회에 긍정적인 영향을 미칠 뿐만 아니라, 이를 인프라에 통합하면 다양한 측면에서의 부정적인 영향을 줄이고 에너지 비용 절감 등의 장기적인 이점 또한 얻을 수 있다. 동시에, 브랜드 친밀도 및 충성도가 향상되어 더 많은 파트너십 기회와 팬 참여를 이끌어낼 수도 있다.

스포츠 산업은 기후 변화의 주요 원인을 제공하는 주체이자 그 영향을 깊이 받는 대상으로, 기후 변화와 떼어놓을 수 없는 관계이다. 우선 전 세계 탄소 배출의 가장 큰 원인 두 가지로 대규모 건설 프로젝트와 물류(운송)가 꼽히는데,⁴⁴ 새로운 경기장 개발은 이 두 가지 원인을 모두 심화하는 요인이다. 반대로 폭염과 기타 극단적인 기후 상황이 대회, 개최지, 선수 건강에 부정적인 영향을 미칠 수 있기 때문에 스포츠계는 기후 변화에 심각한 영향을 받는 대상이기도 하다.⁴⁵ 따라서 스포츠 조직이 부동산 개발 프로젝트를 진행할 때, 지속 가능성 관행과 전략을 신중하게 고려하고 이를 계획 단계부터 반영하는 것이 중요하다.

결론

전 세계 스포츠 조직들은 경기장 수용 인원을 늘리고 팬들이 평생 발길을 끊지 않고 즐길 수 있는 장소가 되도록 인프라 개발에 힘쓰고 있다. 스포츠 인프라를 중심으로 지역사회가 활성화되면 다양한 이해당사자들이 혜택을 받을 수 있다. 우선 민간 투자자와 구단 소유주들은 경기 당일뿐만 아니라 연중 내내 경기장을 활용해 수익을 다각화하고 구단 가치를 높일 수 있다. 또한, 디지털 기술을 접목하면 팬 데이터를 강화해 맞춤형, 타깃형 상품 전략을 강화할 수 있다.

공공 투자자와 정부는 스포츠 인프라 프로젝트 개발에 참여함으로써 지역사회에 더 큰 혜택을 가져올 수 있다. 스포츠 조직들은 공동체 의식을 증진하고, 건강과 웰빙을 향상시키며, 더 많은 관광객을 유치할 수 있다. 엘리트 스포츠 산업은⁴⁶ 경제적, 사회적 성장을 촉진하는 강력한 동력으로 작용하며, 공공과 민간이 각각 정립하는 투자 목표의 교집합이 되고 있다. 가까운 미래에는 스포츠 조직들이 경기장을 활용하여 단순한 스포츠 영역의 범위를 넘어, 보다 다양한 엔터테인먼트와 디지털 콘텐츠로 확장할 것으로 전망된다.

02 비디오 콘텐츠 시장, DTC 지고 애그리게이터 뜬다

비디오 콘텐츠 제작 업계, 배급사 및 애그리게이터와 협업 강화해야 새로운 잠재 시장을 발굴할 수 있다.

여러 개의 독립형 스트리밍 온-디맨드 비디오 서비스(SVOD)를 동시에 구독하는 소비자 행태인 ‘스태킹’(stacking)이 2025년에 감소할 것으로 전망된다. 소비자들이 스태킹하는 평균 SVOD 개수는 미국의 경우 약 4개, 유럽 주요 국가의 경우 약 2개 수준에서 각각 정점을 찍고 2025년부터 줄어들 것으로 전망된다.¹ 구독료 인상, 비밀번호 공유 단속, 묶음형 서비스에 힘입어 SVOD 부문의 수익이 계속 증가하겠지만, 각국 시장별 소비자들의 스태킹 개수는 줄어들 것이다.

스태킹 개수가 정점에 도달한 후 하락 전환할 것으로 예상하는 이유는, 소비자 직접 판매(DTC, direct-to-consumer) 형태로 비디오 구독 서비스를 제공하는 업체들 수십 개로 구성된 비디오 산업의 시장이 생존 가능성을 ‘재평가’해야 하는 시점에 이르렀기 때문이다.² 대다수 가정이 단일 유료 TV

구독 대신 여러 스트리밍 서비스를 구독하고 있는 상황에서, 비디오 콘텐츠 업체는 서로 다른 스트리밍 서비스 제공업체가 제공하는 콘텐츠를 통합하는 방향으로 전환될 가능성이 높다. 이는 비디오 스트리밍 시대가 도래하기 전 과거 유료 TV 제공업체가 제공해왔던 서비스 방식과 유사하다.

중기적으로 비디오 콘텐츠 산업은 ‘비디오 애그리게이터’^{agggregator*}와 함께 2~3개의 독립형 SVOD 제공업체로 재편될 것이다. 이 가운데 기존 유료 TV 회사, 통신사, 기술 플랫폼 또는 현재 시장을 지배하고 있는 최대 규모의 SVOD 업체들이 비디오 애그리게이터로 변모할 것으로 전망된다. 2024년 9월 기준, 영국에서는 설문조사에 참여한 SVOD 가입자의 43%가 DTC 형태가 아닌 제3자 업체(유료 TV 제공업체, 통신사 또는 기술 플랫폼 등)를 통해 서비스를 구매하는 것으로 나타났다. 또한 설문조사에 참여한 소규모 SVOD 제공업체들은 전체 구독의 거의 절반, 대형 제공업체들은 약 25%가 애그리게이터를 통해 구독이 이루어졌다고 각각 답했다.³

* 비디오 애그리게이터(video aggregator)는 파편화된 OTT 업체들을 한꺼번에 모아 놓고, 가입과 결제, 서비스 제공을 원스톱으로 가능하게 만드는 플랫폼 제공자를 뜻한다.

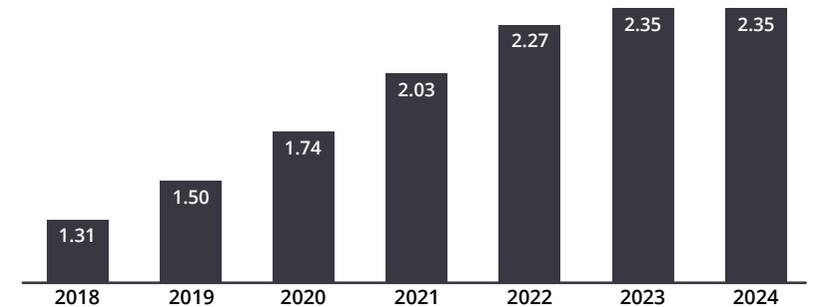
업계 관계자들 의견을 종합하면, 각 애그리게이터가 제공하는 유료 TV 모델에는 단일 계정 및 청구서, 기본 및 선택 채널, 12개월(또는 그 이상) 계약, 시청 가능 콘텐츠에 대한 전자 프로그램 가이드, 광고 판매 및 송출, 중앙 집중식 마케팅 등이 포함될 것으로 전망된다. 비디오 콘텐츠 산업 내 애그리게이션으로의 회귀는 2025년에 한층 가속화될 가능성이 높

다. 하지만 애그리게이터 중심의 시장 구조가 완벽하게 자리잡기까지는 몇 년이 더 걸릴 수 있다.

정점에 이른 SVOD 스택킹

2010년대에는 SVOD 서비스를 이용하는 가구 수와 서비스 이용 건수가 꾸준히 증가하는 두 가지 추세가 두드러지게 나타났다. 딜로이트 연구에 따르면 유럽 가구당 SVOD 서비스 구독 개수는 2018년 1.31개에서 2023년 2.35개로 늘었으나 2024년에는 동일한 개수를 유지하며 정체 양상을 보였다(그림 1). 미국은 2020년부터 약 4개에 머물렀다.⁴

그림 1
유럽 가구당 SVOD 서비스 구독 평균 개수



참조: 영국, 덴마크, 아일랜드, 노르웨이, 스웨덴, 독일, 벨기에, 이탈리아, 네덜란드 등 유럽 주요국(총 2억 6천5백만명) 조사대상 평균치

출처: Deloitte(2018~2024), Digital Consumer Trends

유료 TV 모델이 부활할 수밖에 없는 이유

독립형 SVOD 서비스는 소비자가 콘텐츠와 계약 기간을 선택할 수 있고 콘텐츠 제공자가 배급자를 우회할 수 있다는 큰 장점이 있다. 유료 TV 형태처럼 선택권이 더 적고, 규모가 더 크고, 더 비싸고, 장기적인 묶음(번들)형 서비스로 되돌아가는 것은 과거로의 후퇴처럼 느껴질 수 있다.

그러나 애그리게이션(aggregation) 모델로의 복귀는 소비자와 공급자의 요구를 고루 충족할 수 있는 대안이라 할 수 있다. 소비자에게 이상적인 독립형 SVOD는 다수의 비디오 콘텐츠를 저렴한 가격에 제공해야 한다. 또한 각 서비스는 직관적인 앱을 통해 쉽게 접근할 수도 있어야 하며, 몇 주 전 신청으로 구독 취소가 가능해야 한다. 그러나 이러한 이상적인 서비스 형태는 상업적으로 실행 가능하지 않을 가능성이 크다.

현실에서는 지속적인 구독료 상승과 비밀번호 공유 단속,⁵ 과도하게 넘쳐나는 콘텐츠,⁶ 완성도가 제각각 다른 사용자 인터페이스UI⁷ 등으로 인해 소비자들이 이상적으로 바라는 독립형 SVOD 서비스를 완벽히 구현하기 어렵다. 콘텐츠 제공업체들은 수십 년간 콘텐츠를 제작해 다년 계약으로 배급사에 판매하는 방식으로 수익을 창출했기 때문에, 신용카드 결제 시스템을 설정하거나 광고 지원형 주문 비디오AVOD 서비스를 운영하기 위해 필요한 규제 사항을 관리하는 등 종합적으로 DTC 사업을 운영하는데 필요한 역량이 부족할 수 있다.⁸

애그리게이션으로 다시 수렴되는 비디오 서비스 시장

애그리게이션으로 다시 수렴되는 새로운 시장 구조에서 다음의 방식으로 성장을 도모할 수 있다.

1. 묶음형 서비스

SVOD 구독을 유료 TV나 통신, 금융 서비스 계약과 묶음형으로 판매하면 1년 이상의 구독을 조건으로 개별적으로 구매할 때보다 할인된 요금을 제공할 수 있다. SVOD 업체는 이러한 방식으로 높은 구독 해지율을 낮출 수 있다. 해지율은 미국에서 약 40%, 영국에서 약 20%로 여전히 높은 수준을 보이고 있다.⁹ 업계 분석에 따르면, 2023년 한 해 동안 미국에서만 1억3,930만 건의 SVOD 구독 해지가 발생했으며,¹⁰ 해지자 중 약 4분의 1은 지난 2년 동안 최소 3~4번 서비스를 해지한 ‘상습적 해지 구독자’ serial cherner인 것으로 나타났다. 이는 2019년의 3%에 비해 크게 증가한 수치다.¹¹ 상습적 해지 구독자 중 1/5은 24개월 이내에 7번 이상 서비스를 해지한 것으로 나타났다.¹² SVOD를 18개월 정도의 장기 유료 TV의 묶음형 서비스에 추가하고 총 요금에 대해 할인을 제공하면 전체 해지율과 더불어 계절적 해지율도 줄일 수 있다. 최근 조사에 따르면, 미국 소비자의 절반 이상이 1년 구독에 대한 할인을 선택할 의향이 있지만,¹³ 2024년 초 기준 미국에서 12개월 계약을 맺은 SVOD 구독자는 단 4%에 불과하다.¹⁴ 독립형 SVOD 서비스 업체들이 묶음형 서비스에 참여하면 고객 유치, 요금 청구 및 미수금 관리, 고객 지원, 광고 판매와 같은 다양한 업

무를 외주화 할 수 있다.

이 같은 묶음형 서비스는 2024년에 이미 상당히 보편화됐으며, 2025년부터 더욱 활성화될 가능성이 높다. 유료 TV 회사 입장에서도 SVOD를 자사 제공 서비스에 통합함으로써 얻는 이점이 시장에 따라 다를 수 있지만, 인기 있는 콘텐츠 패키지를 통합해 해지율을 낮추는 효과를 볼 수 있다. 또한 핵심 서비스의 성장이 더딜 경우 총수익을 증대하는 데도 도움이 된다.¹⁵

영국에서는 이미 2019년에 스카이TV SKY TV의 모든 유료 TV 패키지에 넷플릭스 Netflix의 광고 지원형 요금제가 기본으로 포함됐으며,¹⁶ 구독한 모든 콘텐츠를 한 번의 검색으로 찾을 수 있는 단일 검색창이 제공됐다. 프랑스에서는 디즈니플러스 Disney+와 파라마운트플러스 Paramount+를 포함한 패키지를 제공하는 유료 TV 채널 카날플러스 Canal+를 위시해, 복수의 SVOD로 구성된 다양한 패키지가 등장했다.¹⁷ 중유럽에서는 SVOD 총 구독 건수의 약 25%가 유료 TV 또는 통신사를 통한 간접 구독이다.¹⁸ 미국에서는 엑스피니티 Xfinity 초고속 인터넷에 가입하면 애플TV플러스 Apple TV+, 광고 포함 넷플릭스 스탠다드 Netflix Standard, 광고 포함 피콕 프리미엄 Peacock Premium으로 구성된 묶음형 서비스 스트림세이버 StreamSaver를 30% 할인가에 이용할 수 있다.¹⁹ 2023년 기준 전 세계 온라인 비디오 구독자의 20%가 통신사를 통해 가입했고, 2028년에는 그 비율이 25%로 늘어날 것으로 전망된다.²⁰

통신사들은 네트워크 성능이 비등하기 때문에 할인된 가격에 제공하는 인기 있는 SVOD 서비스를 경쟁력으로 내세워 고객 유지율을 향상시킬 수 있다.

일부 은행들도 SVOD를 자사의 구독 서비스에 포함시키고 있다. 2024년 8월 기준으로, 영국 바클레이즈 Barclays 은행은 ‘계좌+블루 리워드’ Bank Account + Blue Rewards를 선택한 고객들에게 애플TV 플러스를 무료로 제공했다.²¹

소규모 SVOD 서비스 중 일부는 독립형 DTC 사업을 정리하고 ‘애그리게이터’ 업체에 의해 운영되는 추가 채널로 전환하거나 일부 시장에서 철수하는 움직임을 보이고 있다.²²

2. 미디어 애그리게이션

개별적으로 판매되던 여러 서비스를 하나로 묶어 개별 구독료의 총합보다 할인된 가격으로 제공하는 사례도 많다. 예를 들어, 미국 시장에서는 디즈니플러스 맥스 Disney+ Max와 훌루 Hulu 번들이 최대 38%의 할인가에 출시됐다.²³ 이러한 번들은 순수하게 비디오 콘텐츠로 구성될 수도 있고 음악, 게임, 뉴스와 같은 다른 콘텐츠를 포함할 수도 있다. 이러한 애그리게이션 서비스가 확산되면서 독립형 서비스가 중단되거나 가격 경쟁력을 상실하는 경우가 늘고 있다.

개별 SVOD를 애그리게이션으로 묶으면 사용자 편의성도 개선될 수 있다. 예를 들어, 구독자가 단일 검색창으로 모든 콘텐츠를 한눈에 검색할 수 있고, 구독 형태에 따라 단일한 전자 프로그램 가이드를 참고할 수도 있다. 반면, 독립형 서비스를 여러 개 이용할 경우 앱 하나를 닫고 다른 앱을 여는 과정에서 끊김이 발생할 수 있다는 단점을 지닌다. 특히 구형 또는 저가형 TV처럼 처리 속도가 낮은 기기에서는 시간 지연을 비롯한 각종 오류가 발생할 수 있다.

미국 소비자 대상 딜로이트 서베이에 따르면, 콘텐츠 검색이 더 쉬워진다면 스트리밍 서비스를 더 많이 이용할 것이라는 응답자가 절반에 육박했다.²⁴ 또한, 미국 Z세대와 밀레니얼 세대 응답자의 약 3/4은 구독 중인 모든 서비스의 콘텐츠를 한 번에 검색할 수 있기를 원한다고 응답했다.²⁵ 영국 소비자 대상 딜로이트 서베이에서는 시청할 콘텐츠를 찾지 못해 구독 서비스를 해지했다는 응답자가 과반수를 넘었다. 콘텐츠는 과거 어느 때보다 넘쳐나는데 정작 소비자들은 자신이 원하는 콘텐츠를 찾지 못하는 것이다.²⁶

애그리게이션 서비스는 구독 해지율을 줄이는 데도 도움이 될 수 있다. 업계 분석에 따르면, 디즈니플러스, 훌루, ESPN+를 통합한 묶음형 서비스를 구독한 사용자는 디즈니플러스 단독 구독자보다 해지율이 59% 낮았다. 소비자들의 가격 인상에 대한 인내심도 한계에 이르고 있다.²⁷ 2023년 4분기에 실시된 딜로이트 설문 조사에서 미국 소비자의 거의 절반이 SVOD 서비스 월 구독료가 5달러 인상되면 해지하겠다고 답했다.²⁸

3. 광고 기반 무료 서비스

일부 유럽 소비자들은 유료 SVOD 대신 인기를 끄는 무료 방송사 VOD BVOD나 유튜브(YouTube) 등 대체로 무료로 제공되는 동영상 공유 서비스를 선택할 수 있다. SVOD 해지의 주요 원인은 구독료다. 영국에서 실시된 서베이 결과, 최근 몇 년간 ‘구독료가 너무 비싸서’ 해지한다는 소비자가 계속 늘어 2024년에는 31%에 달했다.²⁹

이전에 구독 기반으로만 서비스를 운영하던 회사들이 광고 기반 무료 스트리밍 TV(FAST, free-ad supported streaming) 서비스를 제공하는 사례가 많아질 것으로 전망된다. 2008년에 SVOD 서비스를 시작한 아마존(Amazon)은 2022년 미국에서 FAST 채널 프리비(Freevee)를 론칭했다.³⁰ 2010년대 중반부터 애니메이션 SVOD 서비스를 제공해 온 크런치롤(Crunchyroll)도 2023년에 FAST 서비스를 시작했다.³²

기존의 사용자 업로드 콘텐츠 및 주로 무료로 제공되는 유튜브 콘텐츠의 소비가 증가할 수도 있다. TV 화면에서 유튜브를 시청하는 시간은 2024년 하루 평균 12분에서 2029년 22분으로 90% 늘어날 것으로 예상된다.³³ 이로 인해 SVOD 시청 시간이 일부 줄어들 가능성이 크다.

결론: 다시 진화하는 TV 영상 산업

TV 산업은 수십 년에 걸쳐 방송 중심 구조에서 콘텐츠 지식재산^{IP} 기반으로 전환하고 있다. 2010년대 독립형 SVOD 시장의 탄생과 성장으로 전환이 시작됐다. 당시에는 SVOD 서비스가 새로운 형태였고 경쟁자가 적었으며, 구독료가 비교적 저렴하고 소비자간 계정 공유 행태가 흔했기 때문에 독립형 SVOD의 구독자가 빠르게 증가했다.

그러나 2020년대에 접어들면서 시장 환경이 SVOD 업계에 덜 우호적으로 변했다. 해지율은 높아졌고, 지속적인 성장은 기대만큼 쉽지 않았다. 특히 영국과 같이 상대적으로 성숙한 시장에서는 SVOD 시청 점유율이 2023년 15.8%에서 2024년 16.4%로 소폭 상승하는 데 그쳤다.³⁴ 반면, 방송사 콘텐츠와 동영상 공유 사이트는 점유율이 훨씬 가파르게 상승하고 있다. 따라서 독립형 SVOD가 시장에서 다시금 주도적 위치를 차지할 가능성은 적어 보이며, 전통적 유료 TV와 유사한 형태의 ‘애그리게이션 모델’이 앞으로 몇 년 동안 더욱 확대될 것으로 예상된다.³⁵

독립형 SVOD 기업들은 새로운 애그리게이션 모델을 어떻게 활용해야 할지 숙고해야 한다. 우선 구독료 청구부터 유저 인터페이스 설계, 비디오 컴프레션^{compression}까지 모두 처리할 수 있는 규모와 역량을 갖추고 있다면, 콘텐츠 제작부터 광고 판매, 고객 관리에 이르기까지 모든 과정을 자체적으로 운영하는 완전한 독립형 SVOD에 주력할 수 있다. 단, 일부

시장에서는 제3자 배급을 활용하는 방식이 이익 창출에 더 효과적일 수 있다.

일부 SVOD 업체들은 자사 콘텐츠에 더해 다른 플레이어들의 콘텐츠를 통합해 애그리게이션을 앞장설 수도 있다. 이들은 자사 콘텐츠를 중심으로 묶음형 서비스를 구성하거나, 과거 유료 TV처럼 다른 SVOD 채널을 마케팅하고 호스팅할 수도 있다. 그러나 대다수 업체는 자사 콘텐츠를 최고 입찰자에게 판매하는 이미 증명된 단순한 성공 전략에 주력하는 방향을 택할 것으로 예상된다.³⁶

결론적으로, ‘소비자 직접 판매’^{DTC} 비즈니스 모델을 제대로 실행하기는 어려운 경우가 많다. 대부분의 기업은 새로운 역량을 확보하고 기업 문화도 바꿔야 하기 때문이다. 수십 년간 콘텐츠 배급을 외부에 맡겼던 기업이 DTC로 즉각적 성공을 거두기는 쉽지 않다. 단독 비즈니스 모델로서 DTC를 이행하는 것은 비디오 산업뿐 아니라 대부분의 산업도 마찬가지로 극히 어려운 일이다. 규모가 큰 기업들조차도 완전히 DTC 모델로의 전환에 성공한 사례는 극히 드물다. 대부분의 주요 소비자 브랜드에서 배급/유통사들은 가치 사슬에서 핵심적인 역할을 하며, 이 기본적 원칙은 앞으로도 크게 변하지 않을 가능성이 크다.

한편, 독립형 SVOD가 비디오 산업의 주류 비즈니스 모델로 부상하지 않는다고 하더라도, 방송사들이 안일하게 있어서는 안 된다. 많은 시장에

서 방송사는 여전히 소비자들의 콘텐츠 시청 시간을 크게 점유하고 있다. 2023년 유럽(42개 시장 기준)에서 TV 방송 평균 시청 시간은 하루 3시간 16분을 기록했다. 그러나 젊은 층으로 범위를 좁히면 시청 시간은 단 1시간 12분에 그쳤다.³⁷ 방송사들은 적극적 협업을 통해 젊은 시청자들의 니즈에 발빠르게 대응하고 더욱 매력적인 콘텐츠를 제공해야 한다.

03 생성형AI, 콘텐츠 제작과 배급의 경쟁 판도 재편

영화 및 TV쇼 제작 산업의 생성형 AI 도입, 콘텐츠 제작은 신중할 수밖에 없지만 운영 및 배급은 빠른 속도가 예상된다.

생성형AI 모델의 이미지, 오디오, 비디오 생성 능력이 빠르게 발전하면서 이전보다 더 몰입감 있고 창의적일 뿐 아니라 러닝타임마저 길어진 콘텐츠들이 쏟아져 나오고 있다. 이에 영화 및 TV쇼 제작사들은 제작사 발 빠르게 콘텐츠 제작에 생성형AI를 실험적으로 활용해 보았을 수 있지만, 제작 전체 과정에 생성형AI를 도입하는 것은 여전히 신중할 수밖에 없다. 생성형AI 툴의 발전이 아직 성숙 단계에 이르지 못한 데다, 현재 제공되는 퍼블릭 생성형AI 모델로 콘텐츠를 제작했을 때 저작권 침해의 소지가 있고 지식재산권IP을 주장하기도 애매해질 수 있기 때문이다. 하지만 여타 사업 전반에서 비용을 절감하고 수익성을 높이는 데 생성형AI가 유용할 수 있다는 인식이 점차 확산되는 추세다.

현재 주요 제작사들은 비용 압박이 커 수익을 내는 경우가 극히 드물다.¹ 매출은 높지만, 운영비와 제작·마케팅·광고 비용이 더 높아 적자인 경우가 많다. 제작사들의 스트리밍 사업이 수익은 거두지 못하고 출혈만 내고 있는 상황에서 케이블 TV 가입자와 광고 매출도 줄어 수익이 악화되고 있다.² 게다가 인플레이션, 금리 상승, 코로나19(COVID-19) 팬데믹의 영향으로 제작 비용이 더욱 상승했으며, 제작사들은 이제 소비자의 관심과 지갑을 두고 소셜 미디어, 사용자 생성 콘텐츠(UGC), 비디오 게임과도 경쟁해야 하는 상황이다.

델로이트가 미국과 유럽연합(EU)을 중심으로 조사한 결과, 2025년 대형 TV 및 영화 제작사들이 콘텐츠 제작에 생성형AI를 도입하는 데 있어 신중을 기해, 제작 예산에서 생성형AI 톨이 차지하는 비율이 3%가 채 되지 않을 것으로 예상된다.³ 그러나 운영비 지출에서 생성형AI가 차지하는 비율은 약 7%로 이보다 높아, 계약 및 인재 관리, 각종 인가 및 제작 계획 수립, 마케팅 및 광고, 전 세계 진출을 위한 콘텐츠의 현지화 및 더빙 등 다양한 용도로 활용될 전망이다.

이처럼 콘텐츠 제작에는 신중하게, 운영에는 신속하게, 투트랙으로 생성형AI를 도입하면, 선부른 변화로 인력 관리와 콘텐츠 창작에 야기할 수 있는 혼란은 막으면서, 비용을 절감하고 비즈니스 전반의 성과를 더욱 빠르게 달성할 수 있다. 하지만 생성형AI로 인해 또 한가지 유념해야 할 경쟁 요인이 발생하고 있다. 대형 제작사와는 다르게, 독립 콘텐츠 제작자

와 소셜 미디어 플랫폼은 업무 프로세스와 콘텐츠 제작 전반에 걸쳐 생성형AI를 신속하게 도입하고 있으므로, 기존 제작사의 비즈니스 모델을 위협하는 새로운 형태의 미디어가 등장할 수 있다.⁴

생성형AI, 할리우드 수준의 콘텐츠 창작하기에는 아직 역부족

대규모언어모델(LLM)과 디퓨전 모델(diffusion model)*이 싼 가격에 범용화 되면서, 제작사들이 각본과 대사, 스토리라인을 더욱 빠르게 구성하고, 시각화 초기 단계와 캐릭터 설정 및 세트 디자인 등을 더욱 손쉽게 진행하기 위해 실험적으로 생성형AI를 활용하고 있다.⁵ 일부 제작사들은 생성형AI 톨을 활용해 유명 배우들을 더 젊어 보이게 하거나 세트장의 디지털트윈을 제작해 광고 촬영 및 사후 제작에 활용하기도 한다.⁶ 제작사들은 그러한 경우 배우와의 계약 내용에 보호조항을 직접 추가해 저작권 침해 가능성을 차단하고 있다. 2025년에는 제작사들에 생성형AI 톨뿐 아니라 관련해서 발생할 수 있는 여러 문제의 솔루션을 제공하는 제3의 프로덕션 업체들이 증가할 것으로 예상된다.

* 디퓨전 모델(diffusion model)은 데이터를 점진적으로 변형하여 학습하고, 학습한 데이터를 기반으로 새로운 데이터를 생성 및 복원하는 모델로, 특히 이미지 생성과 데이터 복원, 노이즈 제거 등 생성형AI 톨의 모델로 사용된다.

콘텐츠 창작에 생성형AI를 활용하면 사전 제작 단계에서 창의력을 증강

하는 효과가 있지만, 아직 생성형AI 독자적으로는 할리우드 수준의 제작이 불가능하다.⁷ 최첨단 디퓨전 모델은 사진과 흡사한 이미지를 생성할 수 있지만, 지나치게 초현실적이어서 오히려 실제 같지 않은 경우가 많다.⁸ 또한 주요 비디오 생성 툴은 짧은 영상은 만들 수 있지만, 유기적인 스토리로 구성된 긴 영상은 아직 창작하지 못한다.⁹ 비디오 생성 모델들이 빠르게 발전하고는 있지만, 기존의 영화 제작 툴과 과정에 통합되기에는 아직 미성숙한 단계에 머물러 있다.

하지만 창작물을 빨리 게시하는 것이 무엇보다 중요한 소셜미디어 창작자들은 생성형AI의 이러한 한계가 큰 문제가 되지 않을 수 있다. 단기간의 트렌드에 불과할지 모르지만 매우 짧은 영상인 이른바 ‘쇼츠’ shorts의 인기가 높아지고 있고, 소셜미디어 콘텐츠는 저작권 침해 우려도 상대적으로 적다.¹⁰ 생성형AI 툴의 얼리어답터 early adopter인 일부 창작자들은 실험 삼아 제작한 영상을 소셜미디어에 주기적으로 게시해, 다양한 비디오 생성 툴의 빠른 발전 양상을 대중에게 시연하는 역할을 하기도 한다.¹¹

2025년에는 이처럼 독립 창작자들이 생성형AI를 활용한 콘텐츠 창작을 주도할 것으로 예상된다. 그러는 동안 제작사들은 생성형AI의 발전 양상을 관망하면서 당장의 리스크를 모면할 수 있을 것이다. 하지만 전통적 미디어의 콘텐츠들은 갈수록 강해지는 UGC 플랫폼에 소비자들의 ‘주의 집중시간’ attention time을 뺏길 수 있다.

퍼블릭 모델을 활용한 콘텐츠 창작, 저작권 침해 위험

대형 제작사들이 콘텐츠 제작에 생성형AI를 적극적으로 도입하지 않는 또다른 이유는 저작권 침해와 지식재산권 상실 위험이 있기 때문이다.¹² 최첨단 생성형AI 모델은 여타 창작자들이 만든 이미지와 비디오 등 공개 데이터로 훈련되기 때문에, 기본적으로 모델의 생성물은 다른 창작물의 파생물이라 할 수 있다.¹³ 만약 제작사가 퍼블릭 모델의 생성물을 상업적 용도로 사용했는데 해당 생성물에 저작권 보호를 받는 다른 작품이 포함돼 있다면, 저작권 침해로 책임을 져야 할 수 있다.

훈련용 데이터에는 수십억 개의 작품이 포함될 수 있어, 저작권 침해 가능성을 파악하는 것은 ‘모래밭에서 바늘 찾기’처럼 거의 불가능할 수 있다. 하지만 지식재산을 만들어 이를 지키는 것이 생존에 절대적으로 중요한 제작사들은 일말의 불확실성도 감당할 수 없는 입장이다. 이미 독립 아티스트와 창작자들뿐 아니라¹⁴ 다수의 출판사와¹⁵ 음반사들이¹⁶ 자신들의 작품이 훈련용 데이터로 무단 사용됐다며 퍼블릭 생성형AI 모델을 출시한 기업들에 소송을 제기했다.

퍼블릭 모델을 사용해 제작한 콘텐츠는 지식재산권을 주장하기도 어려워질 수 있다.¹⁷ 미국 저작권법 Copyright Act에 따르면, ‘인간에 의한 창작물 요건’ human authorship을 충족해야 저작권 등록이 가능하다. 최근 미국 저작권청 USCO, US Copyright Office은 AI 및 생성형AI의 생성물이 포함된 작품의 ‘인

간에 의한 창작물 요건'은 사례별로 다르며, '충분한 요건'을 충족했을 경우 저작권을 인정받을 수 있다고 규정했다. 다시 말해, 생성형AI 툴을 활용해 인간이 만든 작품은 저작권을 인정받을 수 있지만, 생성형AI가 창작의 대부분을 담당할 경우는 인정받지 못한다는 뜻이다. 관련 논의는 계속되고 있지만, 아직 명확한 정의가 내려지지 않아 AI에 의한 저작권 침해를 둘러싼 불확실성과 리스크가 지속되고 있다.

생성형AI 모델 훈련에는 워낙 방대한 데이터가 필요하기 때문에, 주요 생성형AI 업체들은 콘텐츠 아카이브에 대한 라이선스 계약을 맺기 위해 제작사들에게 구매를 펼치고 있다.¹⁸ 하지만 콘텐츠가 생명줄인 제작사들은 이러한 제안을 원천 거부하거나, 이미 운영비 압박에 허덕이는 생성형AI 업체들이 엄두도 낼 수 없을 정도로 높은 라이선스 비용을 제시하고 있다. 제작사들은 생성형AI 모델이 더 이상 발전되지 않는 것이 자신들에게 유리하다는 계산에서 이러한 라이선스 제의를 담합 수준으로 거절하고 있을 가능성이 있다.

특히 미국의 경우 영화계의 각종 조합 및 노조의 협력 없이는 제작이 힘든데, 이들 단체들이 생성형AI 도입에 강하게 반발하며 제작사들에 생성형AI 사용을 제한한다는 보장을 요구하고 있다.¹⁹ 영국도 노조의 반발이 거세다.²⁰ EU는 AI 모델의 안전성에 관한 '인공지능법' EU Artificial Intelligence Act와 AI 훈련에 사용되는 데이터 수집 및 저장에 관한 '일반 데이터 보호 규정' GDPR, General Data Protection Regulation 등 규제 압박이 강하다.²¹

프라이빗 모델은 지나치게 높은 비용 부담이 걸림돌

딜로이트 2024 첨단기술·미디어·통신 산업 전망'에서 제시됐듯, 프라이빗 생성형AI 모델이 확산되면서 퍼블릭 모델로 야기되는 문제를 해결하고 결과물에 대한 통제를 강화하는 데 도움이 될 수 있다.²² 제작사들은 자체 지식재산을 훈련 데이터로 활용해 프라이빗 모델을 개발하면 퍼블릭 모델 활용에 따른 저작권 침해나 지식재산 상실 위험을 막을 수 있다.

하지만 첨단 생성형AI 모델을 훈련하는 비용은 약 1,000억 달러에 달하는 것으로 추산되며,²³ 추론과 재훈련 비용은 사용이 증가할수록 늘어난다. '오픈 웨이트' open weight*로 알려진 오픈소스 솔루션을 활용하면 비용 부담을 일부 줄일 수 있지만, 훈련용 데이터 세트의 출처와 품질은 불분명한데 비용은 여전히 높다.²⁴ 몸값이 높은 AI 인력을 확보하기도 쉽지 않다. AI 인력은 대부분 제작사보다 높은 임금을 제시하는 하이퍼스케일러 hyperscaler, 대규모 데이터센터 운용 업체를 선호하기 때문이다. 또 프라이빗 모델을 개발하려면 모델 발전 속도에 맞춰 6개월마다 재훈련이 필요하다. 마지막으로 제작사와 투자자들이 비용과 운영 효율적으로 프라이빗 모델을 구축하려면, 테크 기업들처럼 사고하고 행동하며 테크 업계와 생태계 관계를 형성 및 유지하고 이들의 서비스를 이용해야 한다. 이러한 여러가지 이유로 제작사들은 경제적 패러다임이 급변하지 않는 한 프라이빗 모델 개발에 투자하지는 않을 것으로 예상된다.

* 오픈 웨이트(open weight)는 주로 생성형AI 모델을 훈련한 가중치(weight)를 개방하여 자유롭게 사용, 분석, 수정, 재훈련할 수 있도록 하는 방식을 뜻한다.

하지만 2025년에는 제작사와 생성형AI 업체들이 상호 파트너십을 맺어 보다 공평하게 비용을 부담하는 사례가 증가할 것으로 전망된다.²⁵ 외부 생성형AI 업체가 선훈련된 모델과 인터페이스를 제공하면, 제작사가 자체 콘텐츠로 모델의 훈련과 맞춤화를 강화하는 방식이 가능하다. 이렇게 훈련된 생성형AI 모델은 제작사의 미적 기준에 부합하거나 시그니처 캐릭터 및 배경을 포함한 창작물을 생성할 수 있다. 또한 제작사들은 자사 콘텐츠에서 파생될 수 있는 생성물에 대한 기준을 보여줌으로써 지식재산도 더욱 강력하게 보호할 수 있다. 이 같은 프라이빗 방식이 혼용되더라도 생성형AI 업체들은 파트너십을 통해 톨과 생성물을 한층 발전시킬 수 있다.

제작사 사업 운영 최적화에 더 쓸모 있는 생성형AI

2025년 한 해 제작사들을 생성형AI를 콘텐츠 제작에 실험적으로 활용하는 한편, 사업 운영 최적화에는 보다 빠르게 본격 도입할 것으로 예상된다. 생성형AI는 계약 협상, 인재 및 인력 관리, 재무회계, 현지화 등 미디어 운영, 마케팅 및 프로모션, 콘텐츠 스토리지 및 배급 등을 자동화 또는 보완할 수 있다.

제작사들은 이미 사용 중인 소프트웨어나 서비스형 소프트웨어(SaaS, soft-

ware-as-a-service)를 활용해 이러한 역량을 확보할 수 있다. 시간과 비용이 많이 드는 사전 제작 과정에 대한 솔루션을 제공하는 소규모 업체들도 나타났다. 생성형AI를 활용해 대본 평가, 대본 분석 및 제작 일정 수립, 촬영 장소 물색까지 더욱 빠르게 처리하는 것이다.²⁶

생성형AI가 옛날 영화들을 ‘시청’한 후 배우, 테마, 무드별로 태그를 지정하면, 방치됐던 콘텐츠 아카이브를 다시 유용하게 활용할 수도 있다. 스트리밍 업체는 맞춤형 방식 또는 최신 트렌드에 맞춘 마케팅 방식으로 오래된 콘텐츠를 다시 내세워 수익을 창출할 수 있다.²⁷

글로벌 시장에서 더욱 많은 소비자에게 더욱 빠르게 콘텐츠를 배급하기 위해 일부 제작사들은 언어 및 음성 모델을 활용해 번역과 더빙 능력을 강화하고 있다.²⁸ 생성형AI 모델을 활용하면 감정 표현이 풍부한 고품질의 음성 생성 및 재현이 가능하고 사용자가 본인의 취향에 맞춰 세부 조정할 수도 있다.²⁹ 이는 글로벌 시장에 콘텐츠를 수출할 때는 물론이고 수입할 때도 매우 유용하므로, 콘텐츠 제작사와 배급사 모두에게 도움이 된다. 주요 UGC 플랫폼들은 이미 창작자들에게 이러한 기능을 제공하고 있다.³⁰

생성형AI로 더빙과 번역이 한층 용이해지면 문화도 한층 활발하게 공유돼, 특정 지역에서만 잠시 인기를 끌고 사라졌을 작품이 글로벌 히트작으로 거듭나는 경우도 적지 않다. 딜로이트 2024년 서베이에 따르면, 미국 소비자 66%는 이질적 문화를 배울 수 있는 TV쇼나 영화를 즐겨 본다

고 답했다.³¹ 생성형AI는 미디어 기업의 마진과 경쟁력을 실질적으로 높일 뿐 아니라, 전 세계 관객의 유대감을 강화하는 데도 도움이 될 수 있다.

결론

생성형AI의 능력을 경험한 여타 산업과 마찬가지로 제작사와 스트리밍 업계, 창작자들은 경탄과 우려가 혼재된 양가 감정을 느끼고 있다. 이러한 경탄과 우려가 2025년 한 해 제작사들이 생성형AI를 실험적으로 활용하는 두 가지 동인이 될 것이다. 우선 제작사들은 인간의 창의력을 새로운 형태로 구현하는 생성형AI의 마법 같은 능력과 창의력에 경탄하고 있다. 이와 동시에 할리우드 생태계 외부에서 새로운 형태의 미디어가 등장할 수 있다는 우려도 만만치 않다.

독립 콘텐츠 창작자들은 이미 상용화된 첨단 합성 미디어 기술을 적극 활용해 콘텐츠를 제작 및 배포하고 있다. 과거 희소한 콘텐츠와 배급망을 주물렀던 할리우드 제작사들은 이제 넘쳐나는 콘텐츠와 민주화된 배급망이라는 환경에 직면하고 있다.³² 콘텐츠 제작과 배급의 이 같은 파괴적 패러다임 변화는 2025년에 한층 심화될 것이다.

생성형AI는 거의 매달 간격으로 급격한 발전을 이루며 인간 수준의 지능과 창의력, 통찰력에 성큼 다가서고 있다. 2030년에는 순전히 AI가 제작

한 블록버스터 영화가 나올 것이라는 전망이 이미 2024년에 제기된 바 있다.³³ 2025년에는 그러한 전망이 한층 가시화될 것으로 보인다.

그 과정에서 콘텐츠 소유자들은 지식재산을 보호하기 위해 경쟁적으로 방어벽을 강화할 것이다. 퍼블릭 모델의 저작권 침해 소송이 늘어나고 관련 규제도 강화될 것으로 예상된다. 규제당국들은 훈련에 쓰이는 데이터 세트가 기존의 콘텐츠 권리를 침해하지 않음을 생성형AI 업체들이 증명해야 한다는 요구를 할 수 있다. 한편 대부분의 대형 제작사들은 퍼블릭 모델의 훈련에 자사 데이터를 라이선스로 제공하지는 주요 생성형AI 업체들의 제안을 거부하며, 비용 문제가 해결된다면 소규모 생성형AI 업체들과 손잡고 지식재산을 보호할 수 있는 맞춤형 모델을 구축할 것이다.

거시적 관점에서 생성형AI 기술이 계속 발전하려면 막대한 자본 투자가 필요하다. 따라서 향후 1~2년 내 범산업적으로 경제적 가치가 증명되지 못하면 발전 속도가 둔화될 수 있다.³⁴ 하지만 차세대 모델이 기존 모델의 단점을 극복하면, 생성형AI의 능력이 단숨에 강화될 수 있다. 또한 모델 훈련과 운영에 드는 비용과 데이터를 절감하려는 노력도 이어지고 있다. 대형 제작사는 창의력을 발휘해 콘텐츠를 제작하는 곳만이 아니라 대규모 사업을 운영하는 대기업이기도 하다. 따라서 비용 절감, 사업 프로세스 최적화, 생산성 향상, 더욱 신속한 고객층 확대 등을 위해 생성형AI의 도입을 늘릴 가능성이 크다. 딜로이트가 2024년 기업 임원들을 대상으로 실시한 기업의 생성형AI 도입 실태 서베이에 따르면, 응답자 42%는 AI

를 도입하는 가장 중요한 이유로 효율성, 생산성, 비용 절감을 꼽았다.³⁵ 또 58%는 여타 이유로 혁신 증대, 제품 및 서비스 개선, 고객 관계 강화 등을 꼽았다. 오늘날 기업 전반에 걸쳐 생성형AI 도입에 대한 관심이 증대한다는 의미다.

“밀물이 오르면 모든 배가 뜬다.” 규모를 막론하고 모든 제작사와 창작자들이 생성형AI의 혜택을 누릴 수 있다. 소규모 제작사와 개인 창작자들은 과거 대형 제작사들만이 누리던 제작 능력과 작품의 품질을 확보하면서 대형 제작사들이 져야 할 리스크와 비용 부담 측면에서는 상대적으로 자유로울 수 있다. 이들 중 규모가 큰 제작사들은 경쟁에서 이기기 위해 비용을 낮추고 콘텐츠의 론칭에 걸리는 기간을 단축할 필요가 있다. 이제 경쟁 압력은 동종 업계뿐 아니라 UGC 플랫폼과 소셜미디어, 게임 업계 등 다각도에서 발생하기 때문이다. 콘텐츠 제작과 배급망은 더 이상 희소하지 않다. 하지만 소비자의 관심과 시청시간은 유한하다.

Chapter 03

첨단기술·통신, 지금부터 비용 싸움... 정확한 투자수익 예측이 관건

- 01 통신산업 통합으로 기운 세계 각국 규제의 풍향계
- 02 클라우드 인프라의 재무 성과 최적화를 위한 '핀옵스' 원칙

01 통신산업 통합으로 기운 세계 각국 규제 의 풍향계

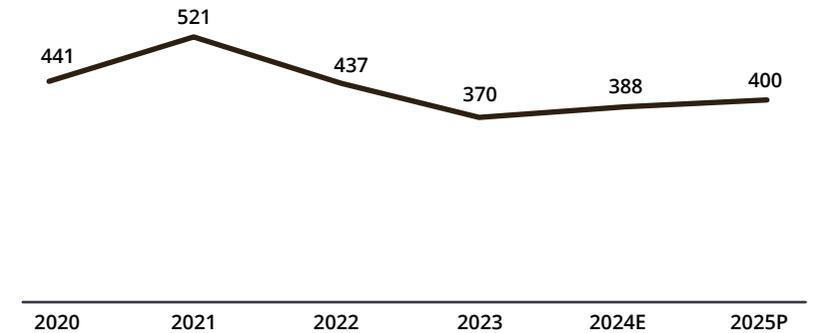
전 세계적으로 소규모 무선 통신사들은 성장 속도가 느리고 수익도 낮
은데 부채 부담을 안고 있는 경우가 많다. 이로 인해 규제 여건만 조성
된다면 자산뿐 아니라 기업간 통합을 위한 인수합병M&A이 활발해질
것으로 예상된다.

2025년 한 해 통신산업 내 통합을 위한 규제 여건이 완화될 것으로 전망
되며, 특히 유럽연합EU이 이러한 흐름의 첫 주자가 될 것으로 예상된다.¹
현재 세계 대부분의 지역에서 무선통신 시장은 파편화돼 있고, 일부 통신
사들은 한정된 시장에만 주력하고 있다. 지금까지는 소비자들의 통신료
부담을 낮추기 위해 가능한 한 많은 통신사들이 서로 치열하게 경쟁하도
록 하는 데 규제의 초점이 맞춰져 있었기 때문이다. 하지만 이제 통신산업
의 미래 네트워크를 성장시키고 기능과 보안, 회복력을 강화하려면 통합
이 용이하도록 규제 여건이 변화해야 한다는 주장에 무게가 실리고 있다.
딜로이트는 2025년 한 해 통신산업 내 M&A 건수가 약 400건을 기록할
것으로 예상한다. 이는 지난 5년간 수준과 별 다를 바 없다(그림 1).² 하지

만 중요한 점은 통신 사업자간 M&A가 증가할 조짐이 보인다는 것이다.
통신 M&A 거래는 여러 형태가 있지만, 전반적으로 특정 유형의 M&A
가 지배적인 것은 아니다(그림2). 무선 및 유선 섹터의 거래가 모두 포함
되며, 각 거래 유형의 비율은 시간이 지나도 대체로 일관적인 패턴을 유지
하고 있다. 다만, 최근 데이터 센터 관련 거래가 증가하고 있는 것은 AI 데
이터 센터 비즈니스가 성장하고 있는 것에 기인한 것으로 보인다.

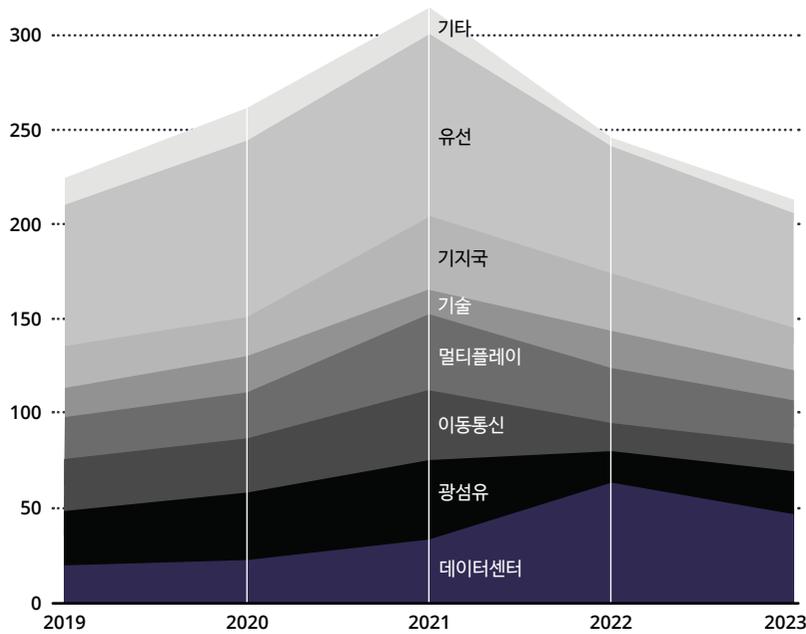
통신산업 내 M&A는 여러 형태로 이뤄지지만, 유독 두드러지는 형태는
없다(그림 2).³ 비율로 따지면 무선과 유선 등 부문을 막론하고 지난 수년
간 별다른 변동이 나타나지 않았다. 다만 최근에는 인공지능AI 데이터센
터 수요가 치솟으면서 관련 M&A 활동이 증가하고 있다.⁴

그림 1
글로벌 통신산업 M&A 건수



참조: E는 집계 진행 중인 예측치, P는 딜로이트의 '25년 예상치
출처: S&P Global Market Intelligence, CB Insights, Deloitte

그림 2
통신산업 하위 섹터별 M&A 건수



출처: Omdia, CSI Magazine, Deloitte

일부 형태의 기업간 또는 자산 통합은 이미 수년간 진행돼 온 결과 마지막 단계에 도달한 것으로 보인다. 기지국 운영이 통신사 대신 독립형 기지국 운영사들 중심으로 통합된 것이 대표적 사례다. 미국과 멕시코의 경우 2023년 기준 기지국 운영사들이 지역 내 무선 기지국의 97%를 점유했으며, 이는 2016년의 65%에서 크게 늘어난 수준이다. 유럽도 마찬가지로 기지국 운영사들의 점유율이 2016년 36%에서 2023년 70%로 뛰었다.⁵

또한 지난 수년간 구리·광섬유·동축케이블 등 유선 네트워크뿐 아니라 정산과 운영에 필요한 백오피스 소프트웨어, 현장 서비스 차량, 데이터센터 등의 M&A 활동이 이어졌다.⁶ 무선 네트워크 부문에서는 다음과 같이 다양한 형태로 통합이 이뤄졌다.

- **캐나다:** 2개 주요 무선 통신사가 2009년부터 무선접속네트워크(RAN, radio access network)를 공유하고 있다.⁷
- **말레이시아:** 당초 독립적으로 운영되던 3개의 무선 네트워크가 2021년 정부 주도 하에 전국 단일 5세대(5G) 네트워크로 통합됐다. 하지만 현재 두 번째 추가 네트워크를 구축하기로 결정됐다.⁸
- **브루나이:** 현재 3개 이동통신사가 소비자와 기업에 서비스를 제공하고 있지만, 이들은 모두 국가 통합 네트워크(UNN, Unified National Networks Sdn Bhd)가 제공하는 무선 네트워크를 공유한다.⁹
- **호주:** 2024년 두 통신사가 4G 및 5G RAN을 공유하기로 합의했다.¹⁰

이처럼 통합의 흐름이 끊이지 않음에도 불구하고, 소비자와 기업 등 최종 이용자에게 통신 서비스를 제공하는 사업자의 수는 큰 변동이 없다. 여전히 각국에서 소비자와 기업은 통신사 3~4개 중 한 곳을 선택하는 것이 일반적이다.¹¹

본 보고서가 새롭게 주목하는 현상은 각국에서 규제의 풍향계가 통신산업의 M&A 활성화 방향을 가리키고 있다는 점이다. 2020년 이후 전 세계 통신 산업 내에서 남북미 6건(미국 3건, 캐나다·칠레·콜롬비아 각 1건), 아시아태평양 5건(인도네시아, 말레이시아, 태국, 대만, 호주 각 1건), 유럽 2건(네덜란드, 스페인 각 1건) 등 13건의 주요 M&A 또는 합작투자가 진행됐다. 그 결과 최종 이용자가 선택할 수 있는 통신사의 수가 줄었다. 이들 중 일부는 이미 규제당국의 승인을 얻었고 일부는 승인 절차를 밟고 있다.

- **남북미:** 6건 (미국 3건 / 캐나다, 칠레, 콜롬비아 각각 1건)¹²
- **아시아태평양:** 5건 (인도네시아, 말레이시아, 태국, 대만, 호주)¹³
- **유럽:** 2건 (네덜란드, 스페인)¹⁴

영국에서는 2024년 12월 보다폰 UK [Vodafone UK](#)와 쓰리 UK [Three UK](#)간 합병이 규제당국의 조건부 최종 승인을 받았다.¹⁵ 이 합병 절차는 2025년 상반기 내에 공식적으로 완료될 것으로 전망된다.

이탈리아와 덴마크에서는 최근 수년간 다수의 M&A 시도가 규제 장벽에 부딪쳐 무산된 바 있다.¹⁶ 이에 주요 인사들이 직접 나서서 통신산업의 통합을 주장하고 있다.¹⁷ 엔리코 레타 [Enrico Letta](#) 전 이탈리아 총리는 2024년 4월 EU에 제출한 보고서에서 통신산업의 통합을 강하게 촉구하며, 유럽 통신 시장이 지나치게 파편화돼 있어 통신사들이 투자수익을 얻기가 어

렵다는 내용을 담은 EU 백서를 근거로 들었다.¹⁸ 유럽의 규제가 통신산업의 통합을 향한 수밖에 없는 근거는 데이터에서 찾을 수 있다. 유럽 내 이동통신 사업자 1개당 평균 가입자 수는 450만 명에 불과해, 미국 9,500만 명, 인도 3억 명, 중국 4억 명과 크게 차이가 난다.¹⁹ 전 유럽중앙은행 ECB 총재이자 전 이탈리아 총리인 마리오 درا기 [Mario Draghi](#) 또한 2024년 9월에 제출한 69쪽 분량의 보고서 중 한 섹션을 통신산업의 통합을 지지하는 내용에 할애했다.²⁰

대부분의 국가에서 매출과 가입자 수 기준으로 상위 1, 2위 무선 통신사는 재정적으로 안정적인 경우가 많다. 반면, 3위 이하의 순위가 내려갈수록 재정 안정성이 떨어진다. 이들 하위 사업자는 향후 네트워크 투자를 지속하기가 어려울 수 있다는 경고를 내놓기도 했다. 이러한 시장에서는 1, 2위 주요 통신사와 2개 이상의 경쟁력이 떨어지는 통신사가 난무하는 것보다 3개 정도의 강력한 통신사로 통합되는 것이 소비자와 기업, 전반적인 경쟁 구도에 더 큰 이점을 제공할 수 있으므로, 규제가 이러한 방향으로 재정립돼야 한다는 주장이 제기되고 있다.

따라서 유럽을 위시해 무선통신 부문에서 통합이 가속화될 것으로 전망된다. 규제 방향이 바뀌는 중요한 이유는 지난 수년간 커넥티비티 선택지가 대폭 늘어나면서 현재 과거 어느 때보다 많아졌기 때문이다. 통신 부문은 현재 다음과 같이 여러 형태의 서비스로 무장한 기업들간 경쟁이 가열되고 있다.

- **고정 무선 접속(FWA)을 통한 가정용 광대역 서비스:** 2025년에는 3,000만 가구 이상이 FWA 광대역 서비스를 이용할 것으로 전망된다. 이는 2024년 대비 20% 증가하는 수준이다.²¹
- **저궤도위성(LEO)을 통한 가정용 광대역:** 주로 농촌 및 외딴 지역을 중심으로 전 세계 300만 가구 이상이 LEO 광대역을 사용하고 있으며, 2025~2026년에 다수의 새로운 LEO 네트워크 론칭이 예정돼 있다.²² LEO 광대역은 일부 유럽처럼 인구 밀도가 높고 지리적 여건이 양호한 지역보다는 아시아태평양, 아프리카, 남북미의 산악 및 사막 지대, 군도 지역에서 더욱 유용하다.²³
- **3G, 4G, 5G의 공존:** 일부 시장에서는 여러 세대의 네트워크가 여전히 사용되고 있어 소비자의 선택지가 많아지고 경쟁이 가열되고 있다.²⁴
- **가상 이동통신망 사업자(MVNO)*의 부상:** 소비자들에게 일명 ‘알뜰폰’으로 알려진 MVNO는 벌써 약 25년 전에 첫 출시됐지만, 최근 다 시금 성장이 촉발되면서 이동통신 가입자 점유율을 늘리고 있다.²⁵ 2024년 기준 미국에서만 케이블 기반 MVNO 가입자 수가 약 1,400만 명에 달했는데, 이러한 성장은 와이파이(Wi-Fi)의 상업성에 일부 기반한다. 미국의 경우 케이블 업체들의 데이터 트래픽 부담 중 87%가 가정용 또는 도시 와이파이 핫스팟으로 전환 또는 분산되고 있다.²⁶

* 가상 이동통신망 사업자(MVNO, mobile virtual network operator)는 이동통신 서비스를 제공하는 데 필수적인 주파수를 보유하지 않고도, 기존 이동통신 사업자(MNO, mobile network operator)로부터 기지국 등 이동통신 설비를 임대한 후 자체 보유한 가입자 관리와 SIM 카드 등 설비와 결합해 독자적 무선 서비스를 판매하는 사업자를 뜻한다.

결론

통신사들이 무선 네트워크의 성능과 경쟁력을 유지하는 비용은 향후 수년간 최근 몇 년에 비해 감소할 것으로 전망된다. 신규 설비와 주파수 구매 등 5G 네트워크 구축에 필요한 고비용 프로세스가 선진국에서는 이미 마무리된 상태이기 때문이다. RAN 지출도 2022년 정점을 찍은 후 감소하며, 당분간 두 자릿수 감소율을 지속할 것으로 예상된다.²⁷ 당초 비독립형(NSA, non-standalone) 5G 네트워크를 구축했던 글로벌 통신사들은 대부분 독립형 네트워크로 업데이트 하는 데 크게 투자하지 않는 추세다.²⁸ 그리고 6G 네트워크는 빨라야 2030년에나 상용화될 것으로 전망된다. 따라서 통신산업의 연간 매출 대비 자본지출(CAPEX) 비율은 2022년 17.8%로 정점을 찍은 후 2025년부터 하락세가 가속화돼 2029년에 이르면 15~16% 범위를 기록할 것으로 예상된다.²⁹ 이는 네트워크 사업자에게는 희소식이지만, RAN 장비 제조 업계에는 부정적 소식이다.

한편 통신사들이 FWA 외에는 5G 및 여타 새로운 서비스로 수익을 창출하기가 쉽지 않다. ‘델로이트 2024 첨단기술·미디어·통신산업 전망’에서 짚었듯이, 인터넷 속도는 이미 충분히 빠르기 때문에, 지금보다 초고속

서비스가 나온다 하더라도 추가 비용이 든다면 사용할 소비자가 거의 없을 것이다.³⁰ 또한 가상현실/증강현실 클래스 등 소비재와 기업용 프라이빗 5G 네트워크, 자율주행차, 원격 수술 등을 지원하는 프리미엄 통신 서비스도 기껏해야 틈새 시장에 불과하다. 헬스케어, 애그테크^{agtech}, 농업+첨단기술 합성어, 보안 등 통신 산업과 밀접하게 연관된 고부가가치 사업에 뛰어드는 통신사들도 있지만, 이러한 인접 사업들은 아직까지 이렇다 할 수익원이 되지 않는 못하고 있다.

추가 수익원으로 생성형AI 데이터센터 사업을 고려할 수 있지만, 시장 1, 2위 대형 통신사들이나 가능한 일이며 규모가 작은 3, 4위 통신사들은 실행 불가능한 일이다.³¹ 소규모 통신사들은 새로운 수익원을 모색하는 것보다 합병되는 것이 나은 전략일 수 있다. 게다가 대부분 부가가치 서비스는 대형 통신사들이나 승산이 있는데, 유럽과 아시아의 통신 시장은 파편화돼 있기 때문에 소규모 통신사들이 필요한 규모를 갖추지 못한 경우가 많다.

글로벌 차원에서 무선통신 부문의 M&A는 두 가지 규제당국의 승인을 받아야 한다. 첫째는 산업 규제당국으로, 영국 방송통신 규제기관 오프콤Ofcom과 미국 연방통신위원회FCC, EU 규제당국 및 EU 각 회원국의 산업 규제기관 등이 이에 해당한다.³² 둘째는 경쟁 규제당국으로, 영국 경쟁시장청CMA와 미국 연방거래위원회FTC, EU 집행위원회 산하 경쟁총국 Directorate-General for Competition 및 EU 각 회원국의 경쟁 규제기관 등이 이

에 해당한다. 아시아태평양 지역 대부분 국가들도 이처럼 두 가지 형태의 규제당국이 관여한다.³³

산업 규제당국들은 국가 내 통신산업 통합에 보다 관대한 반면, 경쟁 규제당국들은 보다 까다롭다. 하지만 최근 유럽에서 규모 경쟁력이 떨어지는 무선 통신사간 통합의 필요성을 역설하는 보고서와 서한이 이어지고 있는 만큼, 일부 국가에서는 경쟁 규제도 완화될 것으로 기대된다.

그렇다고 규제기관들이 M&A 승인에 앞서 조사를 게을리하는 것은 아니다. 최근 성사된 다수의 M&A는 승인을 받는 데 24~36개월이 걸렸다.³⁴ 하지만 규제의 풍향계가 바뀌고 있으므로, 향후 M&A 승인 건수가 증가할 것으로 예상된다.

M&A 승인이 조건 없이 이뤄지는 경우도 있지만, 기업 분할, 가격 보장, 5G 네트워크에 대한 미래 투자 및 제공 약속 등의 조건이 붙는 경우도 있으므로, M&A를 계획하는 통신사들은 이에 대비할 필요가 있다.³⁵

02

클라우드 인프라의 재무 성과
최적화를 위한 '핀옵스' 원칙

기업의 클라우드 지출이 증가하는 가운데, 핀옵스^{FinOps} 원칙을 도입하면 재무 효율성이 강화된다. 이를 통해 기업은 비용을 절감하고 가치를 극대화하며, 부서 간 협업을 더욱 강화할 수 있다.

2025년 전 세계 기업들의 클라우드 지출액은 8,250억 달러를 넘을 것으로 전망되지만 많은 기업의 리더들이 자사의 클라우드 지출 현황을 정확히 파악하지 못한다.¹ 대부분의 기업들은 이처럼 클라우드에 대한 지출 규모를 잘 모르거나 이를 명확하게 설명하는 데 어려움을 겪고 있다.

기업들이 클라우드 서비스에 점점 더 의존함에 따라, 이를 효과적으로 관리하는 것이 점점 더 중요해지고 있다. 특히 최근 주목받고 있는 개념은 핀옵스이다. 핀옵스는 재무^{finance}와 데브옵스^{DevOps}의 합성어로, 클라우드 지출을 추적하고 최적화하는 원칙 및 프랙티스를 의미한다. 딜로이트는 2025년 한 해에만 핀옵스 도구와 프랙티스를 도입한 기업들이 약 210억 달러를 절감할 수 있으며, 비용 절감 효과는 계속 강해질 것으로 전망한

다. 일부 기업들은 클라우드 비용을 최대 40% 절감할 수 있을 것으로 예상된다. 앞으로 핀옵스 팀을 아직 갖추지 않은 기업들은 이를 빠르게 도입할 것이며, 이미 핀옵스를 실행 중인 선도 기업들은 최적화 전략을 한층 정교화 할 것으로 예상된다.

심화되는 클라우드 복잡성, 비용 낭비 초래

클라우드를 이제 기업에 없어서는 안 될 필수적인 요소가 됐다. 물리적 프리라이트 인프라를 구축하려면 서버 구매와 설치에 몇 주 또는 몇 달이 걸리겠지만, 새로운 클라우드 환경은 몇 번의 클릭만으로 가능하다. 클라우드는 편리함과 확장성을 제공해 대규모 인력 없이도 빠른 혁신을 창출할 수 있다. 이 덕분에 데이터 분석, 원격 근무, 인공지능^{AI} 등 다양한 애플리케이션이 가능해졌고, 주문형 비디오^{VOD}, 차량 공유, 챌린저 은행^{challenger bank}^{*}, 원격 의료 등의 산업 혁신이 촉발됐다.²

* 챌린저 은행(challenger bank)은 대형은행의 지배력을 억제하고 은행간 경쟁을 촉진하기 위해 영국 정부가 신규 허가를 내준 은행을 뜻한다. 주로 디지털과 모바일로 빠르고 간편한 금융 서비스 및 사용자 경험을 제공하는 소규모 신생 특화은행에 해당한다.

소프트웨어 엔지니어링에 투자하는 기업들이 다양해지면서 기존 상용 소프트웨어^{COTS, commercial off-the-shelf}로는 이처럼 다양한 기업들의 니즈를 충족하지 못하는 경우가 많다. 때문에 다임러^{Daimler}와 같은 자동차 회

사들은 개발팀을 꾸려 자체적으로 전기차 소프트웨어 플랫폼을 개발하고 있다.³ 심지어 목재 팔레트 유통사처럼 디지털과는 거리가 먼 것처럼 보이는 기업들 사이에서도 맞춤형 소프트웨어와 관련 전문성에 대한 수요가 증가하고 있다.⁴ 이러한 변화는 자연스럽게 클라우드 비용 상승으로 이어지게 된다.

하지만 기업의 클라우드 환경은 점점 더 복잡해지고 있다. 최근 조사에 따르면, 프라이빗 컴퓨팅 자원과 퍼블릭 클라우드 서비스를 혼용하는 하이브리드 클라우드 인프라를 사용하는 기업들이 73%에 달했다. 또한 비용 절감, 특정 기능 활용, 공급업체 종속(vendor lock-in)을 피하기 위해 여러 클라우드 제공업체의 서비스를 동시에 이용하는 기업들이 절반을 넘었다(53%).⁵ 더 나아가, 재무, 인사, 마케팅 등 각 부서가 중앙 IT팀의 승인 없이 클라우드 소프트웨어를 직접 구매하는 경우도 많다. 이런 상황은 데이터 통합, 규정 준수, 보안 등에서 복잡성을 초래할 수 있다.

이로 인해 대부분의 기업들의 클라우드 지출이 당초 예산을 넘어서는 경우가 많다. 설문 조사에 따르면 2023년 클라우드 지출이 예산을 초과했다는 기업이 절반을 넘었으며, 평균적으로 예산을 15% 초과했다.⁶ 그 원인 중 하나로 클라우드 서비스의 종량제(pay-as-you-go) 과금 방식이 지목된다. 사용한 만큼 지불하기 때문에 비용 예측이 어렵다. 극단적인 경우에는 클라우드 엔지니어의 실수로 하룻밤 사이에 수천 달러의 비용이 발생할 수 있다.⁷

게다가 클라우드 서비스는 결코 저렴하지 않으며, 동일 수준의 프라이빗 인프라를 구축하는 것보다 많은 비용이 들 수 있고,⁸ 기업의 최대 IT 지출 항목으로 빠르게 자리 잡고 있다. 코카콜라(Coca-Cola)가 최근 체결한 클라우드 서비스 계약은 무려 11억 달러에 달한다.⁹ 하지만 중요한 점은 클라우드 지출의 약 27%가 결국 낭비된다는 것이다.¹⁰ 최근 설문조사에 따르면, 기업들이 이 문제를 인식하기 시작했으며, 현재 조사 기업의 절반이 전담 핀옵스 팀을 이미 꾸렸고, 20%는 향후 1년 내에 구성할 계획인 것으로 나타났다.¹¹

핀옵스 시작하기

핀옵스는 클라우드 비용 관리에 중요한 원칙이자 프랙티스로, 클라우드 작업 부하의 아키텍처 재구성이나 접근이 제한적인 장기 스토리지 검토와 같은 기술적인 부분부터 할인 및 크레딧 협상 같은 비기술적인 부분까지 다양한 영역에 걸쳐 적용된다. 하지만 핀옵스의 가장 큰 효과는 조직 문화의 변화에서 나타난다. 핀옵스의 핵심은 조직 전체 구성원이 책임을 공유하고, 재무적 투명성을 확보하며, 클라우드 지출이 비즈니스 가치로 연결되도록 만드는 데 있기 때문이다.

핀옵스 시작 단계는 철저한 계획이 매우 중요하다. 우선 현재의 전략을 검토하고, 태그 지정 및 알림 구조를 평가하며, 주요 성과 지표(KPI)를 정의

하는 과정이 필요하다.¹² 첫 번째 단계는 가시성을 확보하는 것으로, 현재의 클라우드 자원을 목록화하고 이를 조직의 니즈에 맞게 어떻게 조정할 수 있을지 탐색하는 과정이 포함된다. 이를 위해 클라우드 제공업체가 제공하는 리소스 모니터링 도구와 비용 관리 전용 툴을 활용할 수도 있고, 더 세부적인 지표를 제공하는 외부 핀옵스 플랫폼을 활용할 수도 있다.

하지만 핀옵스 시작 단계에서 여러 장애물에 부딪칠 수 있다. 복잡한 데이터를 시각화한 대시보드를 정확하게 해석하려면 핀옵스 전문가나 실무자가 필요한데, 이러한 인재들은 최근 수요가 급증하면서 찾기 어려운 경우가 많다. 또한, 여러 클라우드 제공업체를 사용하는 기업의 경우 각 제공업체별로 대시보드를 별도로 관리해야 할 수 있다. 개별 제공업체의 데이터 피드도 각기 다르기 때문에 이를 단일 포털로 통합하는 작업도 만만치 않다. 마지막으로, 핀옵스 툴 자체로도 클라우드 비용의 3~5%에 달하는 적지 않은 추가 비용이 발생하므로, 핀옵스 도입 전 자사의 클라우드의 비용과 효용을 정확히 파악해야 한다.

핀옵스 도입 초기 단계

핀옵스 도입 초기 단계로 자원 낭비 절감, 최적의 자원 배치, 계약 재검토, 잠재적 크레딧과 할인 활용 등을 수행할 수 있다.

- **낭비 절감:** 핀옵스에 익숙하지 않은 기업들은 우선 낭비 절감부터 시작하는 것이 효과적이다. 핀옵스 툴과 대시보드를 활용하면 충분히 활용되지 못하거나 유휴 상태인 자원을 식별하여 이를 축소 또는 제거함으로써 즉각적으로 비용을 절감할 수 있다. 과도한 규모의 가상 머신 VM, 중복된 스토리지 인스턴스, 고아 상태orphaned의 리소스, 중복 데이터 등이 비용을 절감할 수 있는 대표적인 사례다. 핀옵스에 능숙한 기업은 예측 분석을 통해 사용량을 예측하거나, 자동화된 관리 스크립트를 활용해 동적으로 용량을 조정할 수 있다. 게다가 이러한 작업은 중앙 클라우드 엔지니어링 팀이 수행할 수 있어 빠르게 해결될 수 있다.
- **구조와 계층 파악:** 클라우드 서비스의 컴퓨팅 및 스토리지 인스턴스는 품질과 가격대가 매우 다양하므로, 각 서비스의 세부 사항을 면밀하게 파악해 자원을 효과적으로 배치 및 관리하고 특정 애플리케이션에 적합한지 점검해야 한다. 값비싼 인스턴스라고 항상 최적의 선택은 아니며, 일부 애플리케이션은 저렴한 인스턴스가 더욱 적합할 수 있기 때문이다. 예를 들어 계절적 변동성이 심한 이벤트 티켓 판매 웹사이트는 CPU를 지속적으로 사용할 필요는 없지만, 순간적으로 폭증하는 웹 트래픽을 감당할 수 있는 인스턴스가 필요하다.¹³
- **인센티브 활용:** 클라우드 제공업체가 제시하는 할인 프로그램을 활용하면 비용을 대폭 절감할 수 있다. 일부 클라우드 업체는 일정 수준의

데이터 사용량을 유지하는 고객에게 할인가를 제공하고 있다. 또 클라우드 업체들은 장기 계약을 선호하기 때문에, 요금 할인을 직접 재협상 할 수도 있다.

2025년 본격화, 고도화되는 핀옵스

핀옵스 도입을 본격화한 기업들은 2025년 한 해 위의 프랙티스를 한층 정교화함과 동시에 비용 가시성과 통제를 한층 고도화할 것으로 전망된다.

- 책임 분담:** 클라우드는 거의 모든 비즈니스에 필수적인 요소가 됐기 때문에, 각 부서와 팀이 자체적으로 클라우드 지출을 관리할 책임을 지도록 해야 한다. 차지백 방식(chargeback, 비용을 직접 청구하는 방식)이나 쇼백 방식(showback, 비용 부담을 시각적으로 보여주는 방식)을 활용해 부서별 비용 관리의 책임을 부여할 수 있다.¹⁴ 이를 실현하기 위해서는 각 팀이나 프로젝트에서 발생한 리소스 비용을 정확히 표시하는 엄격한 태깅 전략이 필요하다. 이를 위해 사전 정의된 규칙에 따라 자동으로 태그를 지정하는 방식이 가장 효과적일 수 있다. 이렇게 하면, 조직 전체가 클라우드 비용 절감에 동참하고 기여하는 문화를 만들어 갈 수 있다.
- 온프레미스(on-prem) 비용 관리:** 비영리 재단인 핀옵스 재단(FinOps Foundation)을 포함하는 핀옵스 커뮤니티에서는 온프레미스 인프라에도 핀

옵스를 적용해야 한다는 주장에 무게가 실리고 있다.¹⁵ 기업들은 전체 IT 자산에 대한 비용을 평가해야 하지만, 온프레미스 인프라의 비용을 정확히 파악하기는 쉽지 않다. 중앙 클라우드 팀의 시각에서 각 지역의 지사나 인프라 사이트 현황을 살펴보면 이들이 현지 니즈에 의해 다양한 하드웨어와 소프트웨어 툴을 사용하는 경우가 많기 때문이다. 따라서 온프레미스 비용을 절감하려면 중복 사용되고 있는 라이선스를 해지하고 하드웨어의 수명을 연장하는 등의 전략이 필요하다.

- 지속 가능성과 그린옵스(GreenOps):** 핀옵스는 지속 가능성을 목표로 하는 그린옵스 이니셔티브와도 밀접하게 연관되어 있다. 그린옵스는 클라우드 관리에서 지속 가능성을 최적화하는 다양한 전략을 포함한다. 핀옵스의 보고 툴이 제공하는 세부 지표는 에너지 소비와 탄소 배출량, 기타 지속 가능성 목표들을 측정하는 데 중요한 역할을 한다.¹⁶ EU의 기업 지속 가능성 보고 지침(Corporate Sustainability Reporting Directive)¹⁷ 등 주요 규제가 도입됨에 따라, 에너지와 탄소 지표를 추적하고 개선하는 작업은 핀옵스 투자로 얻을 수 있는 중요한 부가가치라 할 수 있다.

핀옵스 도입 사례

다음의 실제 사례들과 같이 핀옵스는 클라우드 비용 절감을 목표로 하는 많은 기업들에게 중요한 도구로 자리 잡고 있다.

- **에어비앤비**¹⁸: 여행 및 숙박 앱 에어비앤비는 저비용 서비스 계층으로 스토리지를 이전하고, 자체 개발한 백업 시스템을 클라우드 제공업체의 대체 서비스로 교체해, 클라우드 비용 6,350만 달러를 절감했다.
- **스카이 그룹**¹⁹: 미디어 및 엔터테인먼트 기업인 스카이 그룹은 클라우드 1년 예산이 단 6개월 만에 소진된 상황을 파악한 후 자체 핀옵스 툴을 도입해 약 150만 달러를 절감했으며, 가시성 대시보드(vis-ibility dashboard)를 통해 다음 해에는 약 380만 달러를 추가로 절감했다.
- **홈디포**²⁰: 홈 인테리어 소매업체인 홈디포는 2022년에 전담 클라우드 비용 관리 팀을 구성한 결과, 전년 대비 수천만 달러를 절감하는 성과를 거뒀다.
- **리프트**²¹: 차량 공유 앱 리프트는 회사 전체가 사용 가능한 청구 데이터 추적 스프레드시트를 활용하여 비용을 관리하고, 리소스 최적화 프로그램을 대대적으로 실행해, 6개월 만에 서비스 1회당 클라우드 비용을 40% 절감했다.
- **WPP**²²: 글로벌 미디어 커뮤니케이션 서비스 기업인 WPP는 자동 생성 크기 추천 툴 등 다양한 도구와 기술을 활용해 핀옵스를 도입한 지 3개월 만에 약 200만 달러를 절감했으며, 이후 연간

클라우드 지출에서 30%의 비용 절감 효과를 달성했다

이처럼 다양한 기업들이 핀옵스에 적극적으로 투자하고 있다. 월마트 Walmart, 마스터카드Mastercard, 아메리칸항공American Airlines 등 기업들은 클라우드 재무 관리의 모범 프랙티스를 선도하는 핀옵스 재단의 회원사로 참여하고 있다.²³

결론: 클라우드 단위 경제성

기업들의 클라우드 의존도가 높아짐에 따라 클라우드 지출에 대한 가시성 개선, 예산 계획 개선, 선제적 지출 통제를 위해 핀옵스가 한층 중요해지고 있다.

앞으로 전 세계 IT 지출은 디지털 전환과 AI 기술의 발전으로 계속 증가해, 2025년에는 5조1,000억 달러를 넘을 것으로 예상된다.²⁴ 또한 여전히 전체 워크로드의 절반을 차지하는 프라이빗 인프라가 퍼블릭 클라우드로 이전될 경우 클라우드 비용이 급증할 수 있다. 게다가 고금리 환경에 직면한 기업들이 수익성 개선과 비용 절감에 집중하고 있으며, 특히 비용 변동성을 줄이려는 노력이 확대되고 있다. 여러모로 핀옵스 시장이 성장하기에 최적의 환경이 조성되고 있는 것이다.

핀옵스는 단기적인 해결책이 아니라, 조직의 운영 전략에 핵심적인 장기적 프랙티스로 접근해야 한다. 대다수 기업들이 비용 절감의 필요에 의해 핀옵스를 도입하겠지만, 궁극적으로 클라우드 지출을 단순한 비용 항목이 아닌 전략적 자산이자 비즈니스 성장의 원동력으로 변화시킬 수 있다. 일부 선도적 기업들은 ‘클라우드 단위 경제성’ 모델의 구축을 최종 목표로 삼을 수 있다. 이러한 모델을 구축하면 애플리케이션, 워크로드, 처리 데이터 기가바이트당 클라우드 지출을 정량화하고, 이를 매출, 전송당 비용, 예약당 비용, 승차 서비스당 비용 등 매출을 포함한 비즈니스 지표에 연계할 수 있다. 이처럼 지출 흐름을 세부적으로 파악하면 전체 비즈니스의 맥락에서 더욱 효율적으로 IT 관련 의사결정을 내릴 수 있고, IT 지출액 단 1달러까지 매출에 어떠한 효과를 주는지 파악할 수 있다.

이렇게 절감한 비용은 새로운 클라우드 서비스를 구매해 규모를 확대하거나 제품 로드맵 기간을 가속화하는 등 신성장 기회를 모색하는 데 재투자 할 수도 있다.

클라우드 환경에 전례 없이 복잡해졌고 그에 따른 비용도 만만치 않다. 하지만 핀옵스를 적절히 활용하면 기업 수익에 무엇보다 가치 있는 자산이 될 수 있다.

제1장 생성형AI 시대 본격화

01 생성형AI의 진화, 이제 데이터센터 에너지의 신뢰성과 청정화가 관건

1. Deloitte analysis based on publicly available information sources and conversations with industry specialists. We used base electricity consumption data from the US Energy Information Administration's (EIA) International Energy Outlook 2023 data on total electricity usage across residential, commercial, industrial, and transportation end uses (a reference to US Energy Information Administration, "Table: Delivered energy consumption by end-use sector and fuel," accessed Nov. 4, 2024) to arrive at estimates and prediction values for global data centers' electricity consumption (TWh) between 2022 and 2030. Our estimates and projections for data centers' percent electricity consumption of global total are based on our research of multiple publicly available sources including SemiAnalysis, EPRI, Goldman Sachs, Bloomberg, and Latitude Media, and further validated based on our conversations with subject matter specialists in the areas of technology, energy, and sustainability. Total energy consumption by end-use sector and fuel (as noted from the aforementioned table from EIA's International Energy Outlook 2023 data), globally, is estimated and forecast at 26,787 TWh in 2025, 27,256 TWh in 2026, and 29,160 TWh in 2030— increasing from 25,585 TWh back in 2022.

2. As noted in endnote 1 above, we arrived at 2022 to 2030 data, estimates, and predictions based on a combination of in-depth secondary research of multiple publicly available sources, and validated further from our discussions with subject matter specialists. Also, see Prof. Dr. Bernhard Lorentz, Dr. Johannes Trüby, and Geoff Tuff, “Powering artificial intelligence,” Deloitte Global, November 2024.”
3. One-fifth of Ireland’s electricity is consumed by data centers, and this is expected to grow, even as households are lowering their electricity use. To read further, see: Chris Baraniuk, “Electricity grids creak as AI demands soar,” BBC, May 21, 2024.
4. Dylan Patel, Daniel Nishball, and Jeremie Eliahou Ontiveros, “AI data center energy dilemma: Race for AI data center space,” SemiAnalysis, March 13, 2024.
5. Ibid.
6. Data center BMO report, Communications Infrastructure, “1Q24 data center leasing: Records are made to be broken,” April 28, 2024; Moreover, due to strong demand from cloud providers and AI workloads, the data center primary market supply in the United States alone was up 26% year over year to 5.2 GW in 2023, and more are under construction. See further: CBRE, “North America data center trends H2 2023,” March 6, 2024.
7. Lisa Martine Jenkins and Phoebe Skok, “Mapping the data center power demand problem, in three charts,” Latitude Media, May 31, 2024.
8. Based on our analysis of multiple publicly available information and reports from what companies self-report, and further validated from third-party sources.
9. For context, hyperscalers are large cloud service providers and data centers that offer huge amounts of computing and storage resources typically at enterprise scale. See: Synergy Research Group, “Hyperscale operators and colocation continue to drive huge changes in data center capacity trends,” Aug. 7, 2024.
10. Yifan Yu, “AI’s looming climate cost: Energy demand surges amid data center race,” Nikkei Asia, June 12, 2024.
11. Data center BMO report, Communications Infrastructure, “1Q24 data center leasing: Records are made to be broken,” April 28, 2024. Further, Deloitte analysis based on information from select tech companies’ publicly available sources such as earnings releases and Dell’Oro Group’s market research data on data center IT capital expenditure shows that if we consider the capital expenditure spending of other data center providers, including third-party operators and outsourced cloud service providers, data centers’ aggregate capital expenditure spending could be at least US\$250 billion in 2025. See: Baron Fung, “Market research on data center IT capex,” Dell’oro Group, accessed Nov. 4, 2024.
12. Nitin Mittal, Costi Perricos, Brenna Sniderman, Kate Schmidt, and David Jarvis, “Now decides next: Getting real about generative AI,” Deloitte’s State of Generative AI in the Enterprise quarter two report, Deloitte, April 2024.
13. Deloitte analysis based on publicly available research reports including: Wania Khan, Davide De Chiara, Ah-Lian Kor, and Marta Chinnici, “Advanced data analytics modeling for evidence-based data center energy management,” *Physica A* 624, 2023; Kazi Main Uddin Ahmed, Math H. J. Bollen, and Manuel Alvarez, “A review of data centers energy consumption and reliability modeling,” in *IEEE Access* 9, 2021: pp. 152536–152563.
14. Tom Dotan and Asa Fitch, “Why the AI industry’s thirst for new data centers can’t be satisfied,” *The Wall Street Journal*, April 24, 2024.
15. Noam Brouard, “Examining the impact of chip power reduction on data center economics,” *Semiconductor Engineering*, March 12, 2024.
16. Based on our analysis of multiple publicly available sources including: Michael Studer, “The energy challenge of powering AI chips,” *Robeco*, Nov. 6, 2023;

- Agam Shah, “Generative AI to account for 1.5% of world’s power consumption by 2029,” HPCwire, July 8, 2024.
17. From our study and analysis of select gen AI data center chip solutions offered by major AI chip vendors, further corroborated with publicly available third-party sources including: Beth Kindig, “AI power consumption: Rapidly becoming mission-critical,” Forbes, June 20, 2024.
 18. Jones Lang LaSalle, “Data centers 2024 global outlook,” Jan. 31, 2024; Doug Eadline, “The gen AI data center squeeze is here,” HPCwire, Feb. 1, 2024; Per IDC, besides graphics processing unit, servers, data centers also need to grapple with a corresponding growth in storage capacity, which is likely to double between 2023 and 2027 to reach 21 zettabytes in 2027. See: John Rydning, “Worldwide Global StorageSphere forecast, 2023 to 2027: Despite decreased petabyte demand near term, the installed base of storage capacity continues to grow long term,” IDC Corporate, May 2023.
 19. Patel, Nishball, and Eliahou Ontiveros, “AI data center energy dilemma.”
 20. Sean Michael Kerner, “What are large language models?” TechTarget, May 2024; Yu, “AI’s looming climate cost.”
 21. Alex de Vries, “The growing energy footprint of artificial intelligence,” Joule 7, no. 10 (2023): pp. 2191 – 2194.
 22. Eren Çam, Zoe Hungerford, Niklas Schoch, Francys Pinto Miranda, and Carlos David Yáñez de León, “Electricity 2024: Analysis and forecast to 2026 report,” International Energy Agency, accessed Nov. 4, 2024.
 23. Deloitte analysis based on publicly available reports and sources including: de Vries “The growing energy footprint of artificial intelligence,” pp. 2191 – 2194.
 24. Deloitte analysis based on data related to energy use and electricity consumption in homes in the United States. See: US Energy Information Administration, “Use of energy explained,” accessed Dec. 18, 2023.
 25. Darren Sweeney, “Utility execs prepare for ‘tripling’ of electricity demand by 2050,” S&P Global, April 19, 2023.
 26. Robert Walton, “US electricity load growth forecast jumps 81% led by data centers,” Utility Dive, Dec. 13, 2023.
 27. Aaron Larson, “How utilities are planning for extreme weather events and mitigating risks,” POWER, March 13, 2024.
 28. Çam, Hungerford, Schoch, Miranda, and de León, “Electricity 2024.”
 29. Baraniuk, “Electricity grids creak as AI demands soar.”
 30. Yu, “AI’s looming climate cost.”
 31. Data on China’s energy use and CO2 emissions sourced from International Energy Agency, accessed September 25, 2024. See: International Energy Agency, “China’s energy use,” accessed Nov. 4, 2024; International Energy Agency, “China’s CO2 emissions,” accessed Nov. 4, 2024.
 32. Baraniuk, “Electricity grids creak as AI demands soar.”
 33. Paul O’Donoghue, “Build it and they will hum: What next for Ireland and data centers?” The Journal, Sept. 2, 2024.
 34. Hosting Journalist, “City of Amsterdam puts halt to new data center construction,” Dec. 21, 2023.
 35. With every 1C increase, operators could save 2% to 5% on the energy they use for cooling equipment. To read further, see: Inno Flores, “Singapore unveils green data center road map amid AI boom that strains energy resources,” Tech Times, May 30, 2024.
 36. Julie R. Peasley, “Ranked: Top 50 data center markets by power consumption,” Visual Capitalist, Jan. 10, 2024.
 37. Whitney Pipkin, “Energy demands for Northern Virginia data centers almost too big to compute,” Bay Journal, June 18, 2024.
 38. Zach Bright, “Southeast utilities have a ‘very big ask’: More gas,” E&E News,

- Jan. 22, 2024.
39. Ibid.
 40. Robert Walton, “AI is enhancing electric grids, but surging energy use and security risks are key concerns,” *Utility Dive*, Oct. 23, 2023.
 41. Karen Hao, “AI is taking water from the desert,” *The Atlantic*, March 1, 2024.
 42. Deloitte analysis based on publicly available information sources including: Jennifer Billock, “Photos: How much water it takes to create 30 common items,” *North Shore News*, Jan. 19, 2023.
 43. Hao, “AI is taking water from the desert”; One case in point is China—where its data centers’ annual water consumption is expected to increase from around 1.3 billion cubic meter as of 2023 to over 3 billion cubic meter by 2030. To read further, see: Yu, “AI’s looming climate cost.”
 44. Eadline, “The gen AI data center squeeze is here.”
 45. Diana Goovaerts, “Data center operators want to run chips at higher temps. Here’s why.” *Fierce Network*, June 11, 2024.
 46. Scott Wilson, “Is immersion cooling the answer to sustainable data centers?” *Ramboll*, Dec. 13, 2023.
 47. David Eisenband, “100+ kW per rack in data centers: The evolution and revolution of power density,” *Ramboll*, March 13, 2024; Direct-to-chip cooling (also known as cold plate liquid cooling or direct liquid cooling) cools down servers by distributing heat directly to server components, while, immersion cooling involves submerging servers and components in a liquid dielectric coolant that also helps prevent electric discharge.
 48. Based on Deloitte’s analysis of developments and announcements from select major cloud hyperscalers and tech companies— and information gathered from publicly available sources (time period: 2023 and first three quarters of 2024).
 49. Joseph Rand, Nick Manderlink, Will Gorman, Ryan Wisser, Joachim Seel, Julie Mulvaney Kemp, Seongeun Jeong, and Fritz Kahr, “Queued up: 2024 edition,” *Lawrence Berkeley National Laboratory*, April 2024.
 50. Based on Deloitte’s analysis of developments and announcements from select major cloud hyperscalers and tech companies— and information gathered from publicly available information sources between the first quarter of 2023 and the third quarter of 2024.
 51. Julian Spector, “Duke Energy wants to help Big Tech buy the 24/7 clean energy it needs,” *Canary Media*, June 11, 2024.
 52. For example, it’s not easy to submerge and drop a 1,300-ton data center unit underwater, especially since it demands special equipment to withstand pressure and corrosion caused by seawater. Moreover, there are concerns related to its impact on marine life.
 53. David Schlissel and Dennis Wamsted, “Small modular reactors: Still too expensive, too slow, and too risky,” *Institute for Energy Economics and Financial Analysis*, May 2024.
 54. Deloitte’s analysis of data and information gathered from multiple reports from S&P Global Market Intelligence, published during March and August 2024.
 55. Manish Kumar, “India’s data center boom opens up a fresh segment for green developers,” *Saur Energy International*, July 1, 2024.
 56. Naureen S. Malik and Bloomberg, “With AI forcing data centers to consume more energy, software that hunts for clean electricity across the globe gains currency,” *Fortune*, Feb. 25, 2024.
 57. Based on Deloitte’s analysis of developments and announcements from select major cloud hyperscalers, tech companies, and power and utility players—on publicly available information sources during 2023 and the first three quarters of 2024.
 58. Duncan Stewart, Nobuo Okubo, Patrick Jehu, and Michael Liu, “The cloud

migration forecast: Cloudy with a chance of clouds,” Deloitte Insights, Dec. 7, 2020.

59. Wylie Wong, “Nvidia launched next-generation Blackwell GPUs amid AI ‘arms race,’” Data Center Knowledge, March 19, 2024; For instance, Nvidia notes that it can train a very large AI model using 2,000 Grace Blackwell chips in 90 days, consuming 4 MW power. In comparison, it would take as much as 8,000 of the previous generation chips to do the same work within the same time, consuming 15 MW power.
60. To read further, see section “Generative AI comes to the enterprise edge: ‘On prem AI’ is alive and well” in our 2025 TMT Predictions chapter on “Updates”; Additionally, see: Sabuzima Nayak, Ripon Patgiri, Lilapati Waikhom, and Arif Ahmed, “A review on edge analytics: Issues, challenges, opportunities, promises, future directions, and applications,” Digital Communications and Networks 10, no. 3 (2024): pp. 783–804.
61. Yu, “AI’s looming climate cost.”
62. Luke Cavanagh, “GPUs vs. CPUs in the context of AI and web hosting platforms,” Liquid Web, Aug. 20, 2024.
63. Eadline, “The gen AI data center squeeze is here.”
64. Goovaerts, “Data center operators want to run chips at higher temps. Here’s why.”
65. Baraniuk, “Electricity grids creak as AI demands soar.”

02 생성형AI, 여성 소비자 및 AI 인력 기반 확대 필요

1. To understand consumer attitudes toward digital life, the Deloitte Center for Technology, Media & Telecommunications surveyed 3,857 US consumers in the second quarter of 2024 and 2,018 US consumers in the second quarter of

2023. These 2024 and 2023 Connected Consumer Surveys collected data on consumers’ reported adoption of generative AI, including experimentation and use for projects and tasks (beyond experimentation). By analyzing longitudinal adoption data and calculating the rate of change in adoption from 2023 to 2024 for men and women, we are able to project that women will close the adoption gap by the end of 2025; see: Jana Arbanas et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey, 5th edition, Deloitte, December 3, 2024; Jana Arbanas, Paul H. Silverglate, Susanne Hupfer, Jeff Loucks, Prashant Raman, and Michael Steinhart, “Balancing act: Seeking just the right amount of digital for a happy, healthy connected life,” Deloitte Insights, Sept. 5, 2023.

2. Ibid.
3. Our analysis was conducted from August to October 2024, based on data from Deloitte UK’s 2023 and 2024 Digital Consumer Trends surveys, as well as a 2024 Deloitte UK survey of European consumers on the topic of generative AI; see: Paul Lee and Ben Stanton, “Generative AI: 7 million workers and counting,” Deloitte, June 25, 2024; Jonas Malmlund, Frederik Behnk, and Joachim Gullaksen, “Generative AI is all the rage,” Deloitte, 2023; Roxana Corduneanu, Stacey Winters, Jan Michalski, Richard Horton, and Ram Krishna Sahu, “Europeans are optimistic about generative AI but there is more to do to close the trust gap,” Deloitte Insights, Oct. 10, 2024.
4. Analysis based on Deloitte’s 2024 Connected Consumer Survey; see: Arbanas et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.
5. Don Fancher, Beena Ammanath, Jonathan Holdowsky, and Natasha Buckley, “AI model bias can damage trust more than you may know. But it doesn’t have to.” Deloitte Insights, Dec. 8, 2021.
6. World Economic Forum, “Global gender gap report 2023,” June 2023.
7. Deloitte AI Institute, “Women in AI,” accessed November 2024.

8. Jana Arbanas et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey, 5th edition, Deloitte, publishing December 3, 2024; Arbanas, Silverglate, Hupfer, Loucks, Raman, and Steinhart, “Balancing act.”
9. Deloitte, “Generative AI: 7 million workers and counting,” accessed November 2024.
10. The Digital Consumer Trends study conducted in various countries in 2024 revealed gen AI adoption gaps of 17 points in Denmark; 12 points in Sweden, Italy, and the Netherlands; 11 points in Belgium; and 10 points in Norway. Additional analysis of a Deloitte European gen AI study revealed gen AI adoption gaps ranging from 10 to 15 points in 11 European countries studied (Belgium, France, Germany, Ireland, Italy, the Netherlands, Poland, Spain, Sweden, Switzerland, and the United Kingdom); see: Deloitte, “Generative AI”; Deloitte, “Generative AI is all the rage,” accessed November 2024; Corduneanu, Winters, Michalski, Horton, and Sahu, “Europeans are optimistic about generative AI but there is more to do to close the trust gap.”
11. Analysis based on 2024 and 2023 Deloitte Connected Consumer Surveys; see: Arbanas et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey; Arbanas, Silverglate, Hupfer, Loucks, Raman, and Steinhart, “Balancing act.” Deloitte, “Generative AI.”
12. Ibid.
13. Arbanas, et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.
14. Ibid.
15. Ibid.
16. Ibid.
17. Ibid.
18. Ibid.
19. For example, only 43% of women we surveyed in the Deloitte 2023 Connected Consumer Survey who owned smart watches or fitness trackers said that they share the data collected by those devices with their health care provider, vs. 57% of men; see: Susanne Hupfer, Jennifer Radin, Paul H. Silverglate, and Michael Steinhart, “Tech companies have a trust gap to overcome—especially with women,” Deloitte Insights, Nov. 8, 2023.
20. These fears may be warranted. Consider that most health apps—along with the data they gather and transmit—are not covered by the Health Insurance Portability and Accountability Act, which means the data may be shared or sold to third parties; see: Steve Alder, “Majority of Americans mistakenly believe health app data is covered by HIPAA,” The HIPAA Journal, July 26, 2023.
21. Ina Fried, “Generative AI’s privacy problem,” Axios, March 14, 2024; Federal Trade Commission, “AI companies: Uphold your privacy and confidentiality commitments, Jan. 9, 2024.
22. Ibid; Matt Burgess and Reece Rogers, “How to stop your data from being used to train AI,” Wired, April 10, 2024.
23. Arbanas, et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.
24. Baris Sarer, Ricky Franks, Cheryl Ho, and Jake McCarty, “AI and the evolving consumer device ecosystem,” The Wall Street Journal, April 24, 2014; Sam Reynolds, “AI-enabled PCs will drive PC sales growth in 2024, say research firms,” Computer World, Jan. 11, 2024; Clare Conley, “Generative AI in 2024: The 6 most important consumer tech trends,” Qualcomm, Dec. 14, 2023.
25. Arbanas, et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.
26. Monique Woodard, “Unlocking the trillion-dollar female economy,” TechCrunch, May 21, 2023.

27. Arbanas, et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.
28. Ibid.
29. Across industries, 51% of women workers using gen AI anticipate it would substantially boost their productivity at work a year from now, vs. 64% of men; see: Arbanas, et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.
30. Tech women and men are statistically tied: Forty-two percent of tech women who use or experiment with gen AI have “high” or “very high” trust that gen AI providers will keep their data secure, and another 40% report moderate trust, while 47% of tech men report “high” or “very high” trust and another 30% report moderate trust; see: Arbanas, et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.
31. Ibid.
32. Greater proportions of men in the tech industry who use gen AI report that their employers encourage its use (93%) and provide training (91%). While there’s still a gender gap in these views among workers in the tech industry, the gap is significantly smaller than among men and women working in other industries; see: Arbanas, et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.
33. World Economic Forum, “Global gender gap report 2023.”
34. Deloitte, “Generative AI.”
35. Genevieve Smith and Ishita Rustagi, “When good algorithms go sexist: Why and how to advance AI gender equity,” Stanford Social Innovation Review, March 31, 2021.
36. Charlotte Lytton, “AI hiring tools may be filtering out the best job applicants,” BBC, Feb. 16, 2024.
37. Carmen Niethammer, “AI bias could put women’s lives at risk – A challenge for regulators,” Forbes, March 2, 2020.
38. Ryan Browne and MacKenzie Sigalos, “A.I. has a discrimination problem. In banking, the consequences can be severe,” CNBC, June 23, 2023.
39. Fancher, Ammanath, Holdowsky, and Buckley, “AI model bias can damage trust more than you may know. But it doesn’t have to.”
40. World Economic Forum, “Global gender gap report 2023.”
41. Smith and Rustagi, “When good algorithms go sexist.”
42. Hyun-Kyoung Kim, “The effects of artificial intelligence chatbots on women’s health: A systematic review and meta-analysis,” Healthcare, Feb. 23, 2024; Sheryl Jacobson and Jen Radin, “Can FemTech help bridge a gender-equity gap in health care?” Deloitte, Oct. 5, 2023; Karen Taylor, “Why investing in FemTech will guarantee a healthier future for all women,” Deloitte UK, June 23, 2023.
43. Arbanas, et al., Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.
44. Ibid.
45. Deloitte AI Institute, “The generative AI dossier: A selection of high-impact use cases across six major industries,” April 3, 2023.
46. Women in AI, “WAI at work: Shaping the future of work for women in AI,” 2022.
47. Ibid.

03 생성형AI 스마트폰, 세상을 바꿀 또 한 번의 격변 촉발

1. GSM Association, “Smartphone owners are now the global majority, new GSMA report reveals,” press release, Oct. 11, 2023.

2. Wolfgang Bock, François Candelon, Steve Chai, Ethan Choi, John Corwin, Sebastian DiGrande, Rishab Gulshan, David Michael, and Antonio Varas, “The mobile revolution: How mobile technologies drive a trillion-dollar impact,” Boston Consulting Group, Jan. 15, 2015.
3. IDC Corporate, “The future of next-gen AI smartphones,” Feb. 19, 2024.
4. Counterpoint, “Gen AI-capable smartphone shipments to grow over 4x by 2027,” April 16, 2024.
5. IDC Corporate, “Worldwide smartphone market up 7.8% in the first quarter of 2024 as Samsung moves back into the top position, according to IDC tracker,” press release, April 15, 2024.
6. IDC anticipates a 364% compound annual growth rate in 2024 (from a low base in 2023) for global gen AI smartphone shipments, with 73% growth in 2025. Canalys expects AI-enabled smartphone market share to reach 54% by 2028. Our analysis, for reasons outlined in this paper, is less bullish than the former, and a bit more than the latter. Sources: IDC Corporate, “The future of next-gen AI smartphones”; Canalys, “Now and next for AI-capable smartphones,” accessed Oct. 30, 2024.
7. Jim Fellingner, “CTA study: Smartphones most-owned tech, 5G and wireless drive adoption,” press release, Consumer Technology Association, May 31, 2023.
8. IDC Corporate, “Worldwide smartphone market up 7.8% in the first quarter of 2024 as Samsung moves back into the top position, according to IDC tracker.”
9. GSM Association, “Smartphone owners are now the global majority, new GSMA report reveals.”
10. Sarah Barry James, “Consumer checkup: Higher interest rates lead to longer tech replacement cycles,” S&P Global, March 26, 2024.
11. IDC Corporate, “Worldwide smartphone market up 7.8% in the first quarter of 2024 as Samsung moves back into the top position, according to IDC tracker.”
12. Chris Donkin, “Smartphone sales up again ahead of expected gen AI boost,” Mobile World Live, July 15, 2024.
13. Susanne Hupfer, Michael Steinhart et al., “2024 Connected Consumer Study,” Deloitte Insights, publication forthcoming, 2024.
14. Counterpoint, “Europe smartphone market recovery continues, shipments up 10% YoY in Q2 2024,” Aug. 28, 2024.
15. Susanne Hupfer, Michael Steinhart et al., “2024 Connected Consumer Study,” Deloitte Insights, publication forthcoming, 2024.
16. Deloitte, “Generative AI: 7 million workers and counting,” June 25, 2024.
17. The installed base of PCs is estimated to be about 2 billion, and there are about 1 billion knowledge workers, suggesting that the market is roughly half consumer and half enterprise.
18. Author interviews with enterprise chief information officers in July and August 2024.
19. Canalys, “AI-capable PCs forecast to make up 40% of global PC shipments in 2025,” March 18, 2024.
20. Ibid.
21. Deloitte Global analysis of publicly available information for H1 2024, and extrapolation based on usual PC seasonality trends.
22. IDC Corporate, “PC refresh cycle and tablets in emerging markets expected to spur demand in coming quarters, according to IDC,” press release, Sept. 23, 2024.
23. IDC Corporate, “Worldwide smartphone market forecast to grow nearly 6% in 2024, driven by stronger growth for android in China and emerging markets, according to IDC,” press release, Aug. 27, 2024.
24. Based on quarterly data so far in 2024, Deloitte believes smartphone average

selling price is declining and should be roughly US\$425 for the year. PC average selling prices were high during the 2021 chip shortage, but are declining and Deloitte estimates them to be about US\$850 for 2024.

25. Roshan Ashraf Shaikh, “Analysts expect 15% price hike for AI PCs—60% of PCs will have local AI capabilities by 2027,” Tom’s Hardware, April 26, 2024.
26. IDC Corporate, “PC refresh cycle and tablets in emerging markets expected to spur demand in coming quarters, according to IDC.”
27. Sigal Samuel, “People are falling in love with—and getting addicted to—AI voices,” Vox, Aug. 18, 2024.
28. IDC, “The future of next-gen AI smartphones.”
29. Baris Sarer, Mark Szarka, Nataliia Bacchus, and Edem Isliamov, “The world of hybrid AI,” The Wall Street Journal and Deloitte, July 31, 2024.
30. Malik Saadi, “On-device generative AI unlocks true smartphone and PC value,” Forbes, April 17, 2024.
31. Lisa Eadicicco, “AI is changing our phones, and it’s just getting started,” CNET, April 3, 2024.
32. Goldman Sachs, “Gen AI: Too much spend, too little benefit?” June 27, 2024.
33. David Cahn, “AI’s US\$600B question,” Sequoia, June 20, 2024.
34. Ibid.
35. Susanne Hupfer, Michael Steinhart et al., “2024 Connected Consumer Study,” Deloitte Insights, publication forthcoming, 2024.
36. Jon Victor, “Software firms race to beat OpenAI in AI agents,” The Information, Sept. 26, 2024.
37. Deepa Seetharaman, “For data-guzzling AI companies, the internet is too small,” The Wall Street Journal, April 1, 2024.
38. Michael Peel, “The problem of ‘model collapse’: How a lack of human data limits AI progress,” Financial Times, July 24, 2024.

39. Yuval Noah Harari, “Yuval Noah Harari argues that AI has hacked the operating system of human civilization,” The Economist, April 28, 2023.
40. CBS News, “Virtual valentine: People are turning to AI in search of emotional connections,” Feb. 14, 2024.
41. Matt Burgess, “Generative AI’s biggest security flaw is not easy to fix,” Wired, Sept. 6, 2023.
42. Camilla Hodgson, “US tech groups’ water consumption soars in ‘data center alley,’” Financial Times, Aug. 17, 2024.
43. Bryce Elder, “Gen-AI revisited, by Goldman Sachs,” Financial Times, Sept. 5, 2024.

04 자동차의 새로운 시대 여는 에이전틱 AI

1. We make this Prediction based in part on soon-to-be-published data from an enterprise survey on AI, including agentic AI. Given the high interest in agentic AI and the products and services that are being launched by startups and established tech companies, we expect this interest to turn to action, at least on an experimental scale.
2. CB Insights. Gen AI Investment Database, Aug 21, 2024. This data excludes Open AI. It includes funding to companies that are developing agentic AI with “varying degrees of autonomy.”
3. Kate Clark, “Investors undaunted by spate of AI acqui-hires,” The Information, Aug. 19, 2024.
4. Cambridge English Dictionary, “Agency,” accessed Aug. 26, 2024.
5. Cambridge English Dictionary, “Autonomous,” accessed Aug. 26, 2024.
6. For humans, agency and autonomy are moral and political concepts. In the context of gen AI agents, we are speaking only of the extent to which soft-

ware-based technology has scope to design and perform tasks without human direction.

7. Faruk Muratovic, Duncan Stewart, and Prashant Raman, “Tech companies lead the way on generative AI: Does code deserve the credit?” Deloitte Insights, Aug. 2, 2024.
8. Scott Wu, “Introducing Devin, the first AI software engineer,” Cognition Software, March 12, 2024.
9. Rina Diane Caballar, “AI Coding is going from copilot to autopilot,” IEEE Spectrum, April 9, 2024.
10. Jenna Barron, “Codeium’s new Cortex assistant utilizes complex reasoning engine for coding help,” SD Times, Aug. 14, 2024; Aswin Ak, “OpenDevin: An artificial intelligence platform for the development of powerful AI agents that interact in similar ways to those of a human developer,” Marktechpost, July 28, 2024.
11. Carl Franzen, “Codium announces Codiumate, a new AI agent that seeks to be Devin for enterprise software development,” VentureBeat, April 3, 2024.
12. Cognition Software, “SWE-bench technical report,” March 15, 2024.
13. Big tech companies continue to improve their software co-pilots to make them more like gen AI agents. For example, see: Alex Woodie, “The semi-autonomous agents of amazon Q,” BigDATAWire, May 3, 2024.
14. Molly Talbert, “Overcoming disruption in a distributed world: Insights from the Anatomy of Work Index 2021,” Asana, January 14, 2024.
15. Total factor productivity, which measures how efficiently both capital and labour are used, can be a proxy for knowledge worker efficiency. Knowledge work requires access to capital-intensive technology, and effectively designed processes.
16. US Bureau of Labor Statistics, “Table A. Productivity, output, and inputs in

the private nonfarm business and private business sectors for selected periods, 1987–2023,” March 3, 2024.

17. Organisation for Economic Co-operation and Development, “Multifactor productivity,” accessed Oct. 30, 2024.
18. Jim Rowan, Beena Ammanath, Costi Perricos, Brenna Sniderman, and David Jarvis, State of gen AI in the Enterprise, Q3 report, Deloitte, August 2024.
19. Ibid.
20. Ibid.
21. Ibid.
22. James O’Donnell, “Why OpenAI’s new model is such a big deal,” MIT Technology Review, Sept. 17, 2024.
23. Janakiram MSV, “AI agents: Key concepts and how they overcome LLM limitations,” The New Stack, June 11, 2024.
24. “OpenAI, “Learning to Reason with LLMs,” Sept. 12, 2024.
25. Anna Gutowska, “What are AI Agents?” IBM, July 3, 2024.
26. Janakiram MSV, “AI agents: Key concepts and how they overcome LLM limitations.”
27. Simon Willison, “Notes on OpenAI’s new o1 chain-of-thought models,” Simon Willison’s Blog, Sept. 12, 2024.
28. Hamidou Dia, “So much more than gen AI: Meet all the other AI making AI agents possible,” Google Cloud Blog, Aug. 20, 2024.
29. Vivek Kulkarni, Scott Holcomb, Prakul Sharma, Edward Van Buren and Caroline Ritter, “How AI agents are reshaping the future of work,” Deloitte AI Institute, November 2024.
30. The Economist, “Today’s AI models are impressive. Teams of them will be formidable,” May 13, 2024.
31. CB Insights, “The multi-agent AI outlook: Here’s what you need to know about

the next major development in genAI,” Aug. 30, 2024.

32. Mike Desmarais, “The call center burnout problem,” SQM Group, Feb. 24, 2023.
33. It’s important to balance the work of human agents. When they get only the most complicated and difficult cases, it can lead to burnout. See Sue Cantrell, et al., “Strengthening the bonds of human and machine collaboration,” Deloitte Insights, Nov. 22, 2022.
34. Sierra, “Sonos elevates the listener experience,” Feb. 13, 2024.
35. Michelle Meineke, “The cybersecurity industry has an urgent talent shortage. Here’s how to plug the gap,” World Economic Forum, April 28, 2024.
36. Ken Yeung, “Dropzone AI gets \$16.85M for autonomous cybersecurity AI agents that reduce manual work by 90 percent,” VentureBeat, April 25, 2024.
37. Simon Thomsen, “Software development cybersec startup Nullify banks \$1.1 million pre-seed round,” Startup Daily, June 26, 2023.
38. Kyt, Dotson, “Norm Ai raises \$27M to help businesses handle regulatory compliance with AI agents,” SiliconANGLE, June 26, 2024.
39. Rowan, State of Generative AI in the Enterprise, Q3 report.
40. Ron Miller, “With Vertex AI Agent Builder, Google Cloud aims to simplify agent creation,” TechCrunch, April 9, 2024.
41. Iris Coleman, “Paradigm utilizes LangChain and LangSmith for advanced AI-driven spreadsheets,” Blockchain.News, Sept. 5, 2024.
42. The Economist, “Today’s AI models are impressive.”
43. Maria Korolov, “AI agents will transform business processes — and magnify risks,” CIO, Aug. 21, 2024.
44. Rowan, State of Generative AI in the Enterprise, Q3 report.

05 딥페이크와 사이버 보안, 쫓고 쫓기는 싸움

1. Margaret Talev and Ryan Heath, “Exclusive poll: AI is already great at faking video and audio, experts say,” Axios, accessed Oct. 28, 2023.
2. Susanne Hupfer, Michael Steinhart, et.al, “2024 Connected Consumer Survey,” Deloitte, December 2024
3. Vivaan Jaikishan, Cameron D’Ambrosi, Jennie Berry, and Stacy Schulman, “The rising threat of deepfakes: Detection, challenges, and market growth,” Liminal, May 7, 2024.
4. Ian Shepherd, “Human vs. machine: Will AI replace content creators?” Forbes, April 26, 2024.
5. Analytix Labs, “Detecting deepfakes: Exploring advances in deep learning-based media authentication,” Medium, January 4, 2024.
6. For example, see: Intel, “Trusted media: Real-time FakeCatcher for deepfake detection,” accessed Oct. 28, 2024.
7. Cade Metz and Tiffany Hsu, “OpenAI releases deepfake detector to disinformation researchers,” The New York Times, May 2024.
8. Danial Samadi Vahdati, Tai D. Nguyen, Aref Azizpour, and Matthew C. Stamm, “Beyond deepfake images: Detecting AI-generated videos,” Drexel University, accessed Oct. 28, 2024.
9. Alex McFarland, “5 best deepfake detector tools & techniques (October 2024),” Unite.AI, Oct. 1, 2024.
10. Konstantin Simonchik, “Deepfake detection: Accuracy of commercial tools,” LinkedIn, February 2024
11. Jiansong Zhang, Kejiang Chen, Weixiang Li, Weiming Zhang, and Nenghai Yu, “Steganography with generated images: Leveraging volatility to enhance security,” IEEE Transactions on Dependable and Secure Computing 21, no. 4 (2024): pp.

- 3994–4005; see also: Mike Bechtel and Bill Briggs, “Defending reality: Truth in an age of synthetic media,” Deloitte Insights, Dec. 4, 2023; and, Loreben Tuquero, “AI detection tools for audio deepfakes fall short. How 4 tools fare and what we can do instead,” Poynter, March 21, 2024.
12. Barbara Ortutay, “Content moderation in the AI era: Humans are still needed across industries,” Fast Company, April 23, 2024; also see: Meta, “How review teams work,” Jan. 19, 2022.
 13. Glenn Chapman, “Meta wants industry-wide labels for AI-made images,” AFP News, Feb. 6, 2024; also see: Nick Clegg, “Labeling AI-generated images on Facebook, Instagram and Threads,” Feb. 6, 2024; Sasha Luccioni et al., “AI watermarking 101: Tools and techniques,” Hugging Face, Feb. 26, 2024; and Partnership on AI, “Building a glossary for synthetic media transparency methods, part 1: Indirect disclosure,” Dec. 19, 2023.
 14. Ryan Heath, “Inside the battle to label digital content as AI-generated media spreads,” Axios, accessed Oct. 28, 2024.
 15. Demian Hess, “Fighting deepfakes with content credentials and C2PA,” CMSWire, March 13, 2024.
 16. Andrew Hutchinson, “X will require ad revenue share participants to confirm their ID,” Social Media Today, May 22, 2024.
 17. Guy Tytunovich, “The future of trust and verification for social media platforms,” Forbes, May 22, 2024.
 18. Amanda Lawson, “A look at global deepfake regulation approaches,” Responsible Artificial Intelligence Institute, April 24, 2023.
 19. US Congress, “S.2765—Advisory for AI-Generated Content Act,” Sept. 12, 2023.
 20. California Legislative Information, “Assembly Bill 3211—California Digital Content Provenance Standards,” Aug. 24, 2024.
 21. Kevin Collier, “States are rapidly adopting laws regulating political deepfakes,” NBC News, Aug. 7, 2024.
 22. Federal Trade Commission, “FTC proposes new protections to combat AI impersonation of individuals,” Feb. 15, 2024; also see: Michelle M. Graham, “Deepfakes: Federal and state regulation aims to curb a growing threat,” Thompson Reuters, June 26, 2024.
 23. Melissa Heikkilä, “Five things you need to know about the EU’s new AI Act,” MIT Technology Review, Dec. 11, 2023.
 24. Deloitte, “How to safeguard against the menace of deepfake technology,” accessed Oct. 28, 2024.
 25. AI Index Steering Committee, “The AI Index 2024 Annual Report,” accessed Oct. 28, 2024.
 26. AI Election Accord, “A tech accord to combat deceptive use of AI in 2024 elections,” accessed Oct. 28, 2024.
 27. Stu Sjouwerman, “The growing threat of AI in social engineering: How business can mitigate risks,” Fast Company, April 8, 2024.

제2장 미디어·엔터테인먼트, 첨단기술로 가속화되는 시장 생태계 변화

01 품의 경기장 건설 프로젝트, 공공-민간 투자의 교두보

1. Alex Dicken, “Tom Wagner reveals timeline for new Birmingham City stadium as Knighthead pledge billions,” Birmingham Live, April 9, 2024.
2. Ibid.
3. Ibid.
4. Alex Dicken, “Another reason for Tom Wagner’s Birmingham City takeover has

- now become clear,” Birmingham Live, Sept. 26, 2023.
5. Dicken, “Tom Wagner reveals timeline for new Birmingham City stadium as Knighthead pledge billions.”
 6. Dicken, “Another reason for Tom Wagner’s Birmingham City takeover has now become clear.”
 7. Hines, “Hines and Tampa Bay Rays gain approval of new ballpark, historic gas plant district development,” press release, July 31, 2024.
 8. FOX 13 News Staff, “Tampa Bay Rays, city of St. Pete sign deal to build new ballpark, keeping team in town for 30 years,” FOX 13 News, July 31, 2024.
 9. Ibid.
 10. Jamie Pugh and Zoe Burton, *The Future of Sport 2024*, Deloitte, Sept. 2, 2024.
 11. Major League Baseball, “Blue Jays showcase all-new 100 level seating bowl at Rogers Centre, as part of multi-year renovations,” April 4, 2024.
 12. Populous, “Blue Jays unveil completed outfield district of Rogers Centre renovations, designed by Populous,” April 6, 2023.
 13. Toronto Blue Jays, “100 level renovation,” accessed Nov. 5, 2024.
 14. Pete Giorgio, David Jarvis, Brooke Auxier, Hannah Bobich, and Kat Harwood, “2023 sports fan insights: The beginning of the immersive sports area,” Deloitte Insights, June 26, 2023.
 15. Katelin Kharrati, “Global smart stadium market size likely to expand at a compound annual growth rate of 22.5% by 2033,” press release, Custom Market Insights, June 28, 2024.
 16. Ibid.
 17. Ed Dixon, “Study: Nine in 10 Gen Z sports fans use social media to consumer content as consumption habits shift,” SportsPro, June 28, 2023.
 18. Giorgio, Jarvis, Auxier, Bobich, and Harwood, “2023 sports fan insights.”
 19. Ibid.
 20. Gary Drenik, “Stadium of the future: Emerging game day technologies for engaging fan experience,” Forbes, Aug. 18, 2022.
 21. Ohm Youngmisuk, “Storms, stats, and T-shirt cannons: LA Clippers’ Halo Board goes all out,” ESPN, Aug. 16, 2024.
 22. Deloitte, *2024 Sports Industry Outlook*, March 12, 2024.
 23. Drenik, “Stadium of the future.”
 24. Mercedes-Benz Stadium, “Delta Fly-Through Lanes,” accessed Nov. 5, 2024.
 25. Deloitte Sports Business Group, *Annual Review of Football Finance 2024*, June 2024.
 26. Bryan Toporek, “The NBA’s new TV deals are poised to send the salary cap skyrocketing,” Forbes, May 30, 2024.
 27. Deloitte Sports Business Group, *Annual Review of Football Finance 2024*.
 28. Sports Business Institute Barcelona, “Commercial revenue: Increasing financial power of football clubs and leagues,” July 11, 2024.
 29. Deloitte Sports Business Group, *Annual Review of Football Finance 2024*.
 30. Gavin Hamilton, “Real Madrid announces record turnover as stadium rebuild nears completion,” SportBusiness, July 18, 2023.
 31. Guillermo Rai, “Real Madrid surpass €1bn in revenue for 2023 to 2024 season,” The Athletic, July 23, 2024.
 32. Real Madrid, “Real Madrid becomes the first football club to exceed 1 billion euros in revenue,” July 23, 2024.
 33. Conor Laird, “The staggering sum Real Madrid earned from four Karol G concerts,” Yahoo Sports, July 24, 2024.
 34. Ibid.
 35. Kansas City Current, “CPKC Stadium and University of Kansas Health System Training Center, accessed Nov. 5, 2024.
 36. Kevin Collison, “Port KC approves massive project next to KC Current Stadium-

um,” Flatland, April 23, 2024.

37. Ibid.
38. Women’s National Basketball Association – Las Vegas Aces, “Home sweet home! Aces take up residence in first-of-its-kind Women’s National Basketball Association practice facility and team headquarters,” press release, April 29, 2023.
39. Morgan Ofori, “Brighton ready to spark revolution with women’s football stadium,” The Guardian, Oct. 29, 2023.
40. Simi Iluyomade, “Manchester City Women are building a new £10 million training facility,” Versus, May 15, 2024.
41. Samuel Agini and Alice Hancock, “Sports hospitality shifts focus to fans as UK demand for ‘experiences’ grows,” Financial Times, Aug. 13, 2021.
42. Georgina Yeomans, “How Formula One transformed its hospitality product,” BlackBook Motorsport, Jan. 6, 2022.
43. Deloitte, The Future of Sport 2023, April 2023.
44. Center for Climate and Energy Solutions, “Global emissions,” accessed Nov. 5, 2024.
45. Directorate-General for Climate Action, “Sport—a key player in climate action?” European Union, July 26, 2024.
46. Elite sports are defined as the highest level of competition, which may or may not be classified as “professional” sports where participants are paid for their performance.

02 비디오 콘텐츠 시장, DTC 지고 애그리케이터 뜬다

1. Kevin Westcott, Jana Arbanas, Chris Arkenberg, and Jeff Loucks, “Streaming video at a crossroads: Redesign yesterday’s models or reinvent for tomorrow?”

Deloitte Insights, March 20, 2024; Deloitte, “Generative AI: 7 million workers and counting,” June 25, 2024.

2. Ampere Analysis, “Analytics – SVoD,” accessed November 2024.
3. This is based on a nationally representative survey of 4,000 respondents ages 16 to 75, commissioned by Deloitte UK and undertaken in August 2024.
4. Westcott, Arbanas, Arkenberg, and Loucks, “Streaming video at a crossroads.”
5. David Pierce, “Streaming services keep getting more expensive: all the latest price increases,” The Verge, Sept. 23, 2024; Emma Roth, “Disney’s password-sharing crackdown starts ‘in earnest’ this September,” The Verge, Aug. 7, 2024.
6. As of June 2023, Nielsen’s Gracenote reported that the number of individual titles available on streaming services was 2,346,171, up from 1,882,401 in July 2021. These numbers span the range of content available in the United States, the United Kingdom, Canada, Mexico, and Germany; see: Nielsen, “Data-driven personalization: The future of streaming content discovery,” accessed November 2024.
7. Selome Hailu and Jennifer Maas, “From ‘glitchy’ HBO Max to ‘overwhelming’ Amazon Prime Video, Hollywood insiders spill on their (least) favorite streaming interfaces,” Variety, April 11, 2023.
8. Andrew Blustein, “How GDPR, ad fatigue and content costs are complicating the now-global streaming wars,” The Drum, Sept. 4, 2019.
9. Westcott, Arbanas, Arkenberg, and Loucks, “Streaming video at a crossroads”; Deloitte, “Generative AI.”
10. Antenna, “The rise of the show chaser,” accessed November 2024.
11. Ibid.
12. Ibid.
13. Westcott, Arbanas, Arkenberg, and Loucks, “Streaming video at a crossroads.”
14. Antenna, “Understanding the relationship between annual plans and promotions,” accessed November 2024.

15. IDC Research, “IDC forecasts slower growth for global telecommunications market: Could AI help telcos to maintain healthy margins,” May 3, 2024.
16. Sky, “Official website,” accessed November 2024.
17. Canal Plus, “S’abonner à Disney+ avec les offres CANAL+,” accessed November 2024; Georg Szalai, “Paramount+, Canal+ expand partnership in France,” The Hollywood Reporter, Aug. 21, 2024.
18. Omdia, “Omdia unveils surging SVOD growth in CEE through strategic pay TV and telco partnerships,” June 12, 2024.
19. Xfinity, “Streaming services: Stream one, stream all,” accessed November 2024.
20. Bango, “Super bundling: What telco leadership needs to know about securing a wider role in the subscriptions market,” accessed November 2024.
21. Barclays, “Current accounts: Barclays Bank Account + Blue Rewards,” accessed November 2024.
22. Viaplay Group, “Q2 2024 interim report January–June,” press release, July 18, 2024.
23. Disney Plus, “New Disney+, Hulu, Max bundle is now available in ad-supported and ad-free plans,” July 25, 2024.
24. Westcott, Arbanas, Arkenberg, and Loucks, “Streaming video at a crossroads.”
25. Ibid.
26. Deloitte, “Generative AI.”
27. Ampere Analysis, “Subscribe, cancel, repeat: 42% of US consumers are SVoD ‘resubscribers’,” July 8, 2024.
28. Westcott, Arbanas, Arkenberg, and Loucks, “Streaming video at a crossroads.”
29. Deloitte, “Generative AI.”
30. Christian de Looper, “Amazon Freevee: Everything you need to know about the free streaming service,” Amazon UK, May 10, 2023.
31. Janko Roettgers, “Chernin, AT&T Set Brand for New Online Video Venture: El-
lusion (Exclusive),” Variety, Aug. 3, 2015.
32. Crunchyroll News, “Crunchyroll launches 24/7 anime channel in the US,” Oct. 11, 2023.
33. Enders Analysis, <https://www.endersanalysis.com/reports/video-viewing-forecasts-slowdown-change>
34. Share of viewing is calculated based on 12-month rolling averages from October 2022 to September 2023 through August 2023 to July 2024. All data is from Barb’s monthly viewing summary; see: Barb, “Monthly viewing summary,” accessed November 2024. Barb’s methodology for capturing viewing patterns is explained in: Barb, “What is the Barb panel and why is it important?” accessed November 2024.
35. Richard Waters, The next phase of the streaming wars, Financial Times, June 6, 2024 (subscription required)
36. Etan Vlessing, “Sony CFO: Without a streaming platform, we’re free to sell films and shows ‘to the highest bidder’,” The Hollywood Reporter, March 6, 2023; Diane Haithman, “Why Sony’s streaming deals with Netflix and Disney make sense for everyone,” The Wrap, May 10, 2021.
37. Audience Trends Television 2024, European Broadcasting Union Media Intelligence Service, August 2024 (registration required)

03 생성형AI, 콘텐츠 제작과 배급의 경쟁 판도 재편

1. George Szalai, “Studio profit report: A year of major transition,” The Hollywood Reporter, April 24, 2024.
2. George Szalai, “Studio profit report: Disney dives as Sony soars, Paramount rises,” The Hollywood Reporter, Feb. 24, 2024.
3. This prediction is based on our analysis of earnings reports from leading streaming

- video providers and other available industry information.
4. Chris Arkenberg, “Will generative AI challenge authenticity in social media?,” Deloitte Insights, Dec. 8, 2023.
 5. Hannah Murphy, “Media groups look to AI tools to cut costs and complement storytelling,” Financial Times, March 26, 2024.
 6. David Smith, “‘We’re going through a big revolution’: How AI is de-ageing stars on screen,” The Guardian, Feb. 6, 2023.
 7. Ibid: Murphy, “Media groups look to AI tools to cut costs and complement storytelling.”
 8. Alon Yaar, “What’s next for AI video generation,” AI Business, Aug. 6, 2024.
 9. Lauren Leffer, “Everything to know about OpenAI’s new text-to-video generator, Sora,” Scientific American, March 4, 2024.
 10. Taylor Lorenz, “The ‘Beastification of YouTube’ may be coming to an end,” The Washington Post, March 30, 2024.
 11. Dennis Ortiz and Kenny Gold, “Gen AI and the creator economy: How creators are looking to leverage AI and what this means for brands,” Deloitte, accessed Oct. 30, 2024.
 12. Paul Sweeting, “Hollywood’s AI concerns present new and complex challenges for legal eagles to untangle,” Variety, April 17, 2024.
 13. Jennifer Wolfe, “What would have to happen for gen AI to take over Hollywood? So glad you asked.,” NAB Amplify, July 5, 2024.
 14. Ibid: Sweeting, “Hollywood’s AI Concerns Present New and Complex Challenges for Legal Eagles to Untangle.”
 15. Baker & Hostetler, “Case tracker: Artificial intelligence, copyrights and class actions,” accessed Oct. 30, 2024.
 16. Natalie Sherman, “World’s biggest music labels sue over AI copyright,” BBC News, June 25, 2024.
 17. US Copyright Office, “Copyright and artificial intelligence,” March 16, 2023.
 18. Lucas Shaw, “Alphabet, Meta offer millions to partner With Hollywood on AI,” Bloomberg, May 23, 2024.
 19. Erin Degregorio, “Hollywood is back to work after strikes, but AI remains in the spotlight,” Fordham Law News, Jan. 29, 2024.
 20. Daniel Thomas and Cristina Criddle, “UK shelves proposed AI copyright code in blow to creative industries,” Financial Times, Feb. 4, 2024.
 21. Lindsey Wilkinson, “EU passes AI Act, places first binding rules on generative AI,” CIO Dive, March 13, 2024.
 22. Chris Arkenberg, Baris Sarer, Gillian Crossan, and Rohan Gupta, “Taking control: Generative AI trains on private, enterprise data,” Deloitte Insights, Nov. 29, 2023.
 23. Jowi Morales, “AI models that cost \$1 billion to train are underway, \$100 billion models coming — largest current models take ‘only’ \$100 million to train: Anthropic CEO,” Tom’s Hardware, July 7, 2024.
 24. Red Hat, “What is an open-source LLM?,” July 1, 2024.
 25. Kyle Wiggers, “Generative AI startup Runway inks deal with a major Hollywood studio,” TechCrunch, Sept. 18, 2024.
 26. Lauren Forrestal, “Filmstage leverages AI to break down film scripts, create shooting schedules and more,” TechCrunch, March 20, 2023; Lauren Forrestal, “Avail rolls out its AI summarization tool to help Hollywood execs keep up with script coverage,” TechCrunch, Dec. 7, 2023.
 27. Emma Cosgrove, “Nvidia, Amazon, Microsoft, and Paramount execs discuss the use of AI in Hollywood. Here are 9 startups they’re watching.,” Business Insider, July 24, 2024.
 28. The Economist, “The dawn of the omnistar,” Nov. 9, 2023.
 29. Audrey Shomer, “The state of generative AI in Hollywood: A special report,” Va-

riety, June 3, 2024.

30. Andrew Hutchinson, “YouTube announces expansion of auto-dubbing to more creators and languages,” SocialMediaToday, Sept. 19, 2024; Julia Walker, “How Meta’s AI dubbing breaks down language barriers,” PR Week, Sept. 30, 2024.
31. Jana Arbanas, Jeff Loucks, Brooke Auxier, Kevin Westcott, Chris Arkenberg, and Bree Matheson, 2024 Digital Media Trends, Deloitte Insights, March 20, 2024.
32. Michael D. Smith, “Lessons from Hollywood’s digital transformation,” Harvard Business Review, Dec. 16, 2021.
33. Jackie Wiles, “Beyond ChatGPT: The future of generative AI for enterprises,” Gartner, Jan. 26, 2023.
34. David Cahn, “AI’s \$600 billion question,” Sequoia, June 20, 2024.
35. Deloitte, The State of Generative AI in the Enterprise—Moving from potential to performance, June 2024.

제3장 첨단기술·통신, 지금부터 비용 싸움... 정확한 투자수익 예측이 관건

01 통신산업 통합으로 기운 세계 각국 규제의 풍향계

1. Deloitte analysis of recent events in Europe, specifically recent letters and white papers from Draghi and Letta (see endnotes 16 and 19).
2. Deloitte analysis of historical merger and acquisition trends, combined with early publicly available signs of deal activity.
3. CSI Magazine, “Telecom consolidation: Over 500 M&A deals in five years,” July 23, 2024.
4. Duncan Stewart, Dan Littmann, Girija Krishnamurthy, and Matti Littunen, “Telecoms tackle the generative AI data center market,” Deloitte Insights, Sept.

16, 2024.

5. Stephanie Price, “For sale: Canadian read-throughs from global telecom asset sales,” CIBC Capital Markets, Aug. 14, 2024.
6. ArdorComm News Network, “Telecom M&A activity witnesses surge: 514 deals from 2019 to 2023,” ArdorComm Media Group, July 25, 2024.
7. Sue Marek, “Marek’s take: 5G network sharing may be the answer,” Fierce Network, May 14, 2021.
8. Affandy Johan, “5G in Malaysia – Single wholesale network driving regional leadership,” Ookla, March 17, 2024.
9. Digital Regulation Platform, “Changing the operating model: the creation of UNN in Brunei Darussalam,” Aug. 27, 2020.
10. Australian Competition & Consumer Commission, “Optus Mobile Pty Ltd and TPG Telecom Limited proposed network and spectrum sharing,” Sept. 5, 2024.
11. Megan Emfosi Meena and Jiaying Geng, “Dynamic competition in telecommunications: A systematic literature review,” Sage Open, April 26, 2022.
12. Detecon Spotlight, “Telco mergers and acquisitions,” 2022; M&A Community, “12 significant telecom mergers and acquisitions over the last 20 years,” Sept. 17, 2024; Henri Capin-Gally, Sergio Márquez García Moreno, and Bill Parish, “Energy and telco deals power Mexican M&A,” White & Case, March 6, 2024; Geusseppe Gonzalez, “Access Alert: Colombian telecoms industry faces shakeup with potential Millicom-Telefonica merger,” Access Partnership, Aug. 8, 2024.
13. Julber Osio, “Asia-Pacific telcos consolidate to compete with market leaders,” S&P Global, May 25, 2023; Tom Leins, “Deal-making Down Under: Oceania’s wave of telecom M&A,” TeleGeography, March 24, 2021.
14. Jan Frederik Slijkerman, “Telecom Outlook: Will we see more mergers and buy-outs in 2022?” ING, Jan. 28, 2022; Jan Frederik Slijkerman and Diederik Stadig, “Telecom tycoons on the move?” ING, Feb. 1, 2024.

15. Competition and Markets Authority, “How we are investigating the Vodafone / Three potential merger,” Sept. 13, 2024.
16. Slijkerman and Stadig, “Telecom tycoons on the move?” ING, Feb. 1, 2024.
17. Enrico Letta, “Much more than a market,” Consilium, April 2024.
18. European Commission, “White paper – How to master Europe’s digital infrastructure needs?” Feb. 21, 2024.
19. Hamish White, “Europe’s looming mobile crisis,” Technative, April 2024.
20. Mario Draghi, “EU competitiveness: Looking ahead,” European Union, Sept. 9, 2024.
21. See section “Fixed wireless access: Contrary to popular opinion, adoption may continue to grow” in chapter Updates: Past TMT Predictions’ greatest hits and (near) misses.
22. Ongoing Deloitte analysis of current and proposed low Earth orbit satellite networks.
23. Deloitte author conversations with communications providers in North America, Europe, Africa, and Asia.
24. Deloitte analysis of developing world wireless networks.
25. Piran Partners, “Seizing the future of telecoms: The continuing ascendancy of MVNOs,” April 12, 2024; Puneet Takyar, “The journey of MVNOs,” Comviva, April 1, 2021.
26. Jeff Baumgartner, “Cable snared nearly half of US mobile line adds in Q3 – analyst,” Light Reading, Nov. 16, 2023.
27. Juan Pedro Tomás, “Global RAN market faces challenging scenario in Q2: Dell’Oro,” RCR Wireless News, Aug. 19, 2024.
28. Deanna Darah, “5G NSA vs. SA: How do the deployment modes differ?” TechTarget, July 25, 2024.
29. Matt Walker, “Telco capital intensity hits 10 year peak in 2Q22,” MTN Consulting, Sept. 6, 2022.
30. Paul Lee, “No bump to bitrates for digital apps in the near term: Is a period of enough fixed broadband connectivity approaching?” Deloitte Insights, Nov. 29, 2023.
31. Stewart, Littmann, Krishnamurthy, and Littunen, “Telecoms tackle the generative AI data center market,” Deloitte Insights, Sept. 16, 2024.
32. DataHub, “Regulatory authority – Institutional structure,” accessed October 2024; Policies, “Telecommunications national regulatory authorities,” European Commission, Jan. 11, 2023.
33. Lynn Robertson, “Interactions between competition authorities and sector regulators – contribution from business at OECD (BIAC),” Organisation for Economic Co-operation and Development, Nov. 18, 2022.
34. Research Notes, “5G rollout slowed while mobile operators await merger approval. Case in point: the slow rollout in UK,” Strand Consult, Oct. 5, 2024.
35. Deloitte author’s assessment of multiple approved mergers in North America, Europe, and Asia in the period of 2015 to October 2024.

02 클라우드 인프라의 재무 성과 최적화를 위한 ‘핀옵스’ 원칙

1. Gartner, “Gartner forecasts worldwide public cloud end-user spending to surpass \$675 billion in 2024,” press release, May 20, 2024.
2. Yury Izrailevsky, Stevan Vlaovic, and Ruslan Meshenberg, “Completing the Netflix cloud migration,” Netflix, Feb. 12, 2016.
3. Douglas Busvine, “Daimler to hire 1,000 programmers in Germany,” Reuters, April 18, 2021.
4. CHEP, “CHEP uses ‘track and trace’ technology on its reusable pallets,” press release, April 8, 2022.

5. Flexera, “2024 State of the cloud report,” 2024.
6. Ibid.
7. Parshv Jain, “Avoiding costly cloud mistakes: Lessons learned from a \$72K bill,” Medium, June 12, 2023.
8. Owen Rogers, “Reports of cloud decline have been greatly exaggerated,” Up-time Institute, Jan. 18, 2023.
9. The Coca-Cola Company, “The Coca-Cola Company and Microsoft announce five-year strategic partnership to accelerate cloud and generative AI initiatives,” press release, April 23, 2024.
10. Flexera, “2024 State of the cloud report,” 2024.
11. Ibid.
12. Nikhil Roychowdhury, Nik Jethi, Farhan Akram, and Rishabh Kochhar, “Optimizing the value of cloud: A guide to getting started,” Deloitte, March 30, 2023.
13. Amazon Web Services, “TicketSwap tames demand ups and downs with AWS,” 2021.
14. FinOps Foundation, “Invoicing & chargeback,” accessed Nov. 4, 2024.
15. For example: The Linux Foundation, “FinOps across public cloud and on-prem,” accessed Nov. 4, 2024.
16. Meredith Shubel, “What is GreenOps? Putting a sustainable focus on FinOps,” The New Stack, Sept. 22, 2023.
17. Magda Puzniak-Holford, Adithya Subramoni, and Simon Brennan, “EU Corporate Sustainability Reporting Directive (CSRD) – Strategic and operational implications,” Deloitte, Sept. 8, 2023.
18. Belle Lin, “Airbnb details road map to lower cloud costs,” The Wall Street Journal, Nov. 7, 2022.
19. James Ma, “How Sky saved millions with Google Cloud,” Google Cloud Blog, July 19 2021.
20. Angus Loten and Isabelle Bousquette, “Amazon warns of weaker cloud sales as businesses cut spending,” The Wall Street Journal, April 13, 2023.
21. Amazon Web Services, “Lyft uses AWS Cost Management to cut costs by 40% in 6 months,” 2020.
22. IBM, “How the world’s largest ad company optimizes FinOps,” accessed Nov. 4, 2024.
23. FinOps Foundation, “FinOps Foundation Members,” accessed Nov. 4, 2024.
24. Gartner, “Gartner forecasts worldwide IT spending to grow 8% in 2024,” press release, Oct. 18, 2023.

인더스트리 리더십



Ariane Bucaille

Global TMT Industry Leader | abucaille@deloitte.fr



Kevin Westcott

US Tech, Media & Telecom Leader | kewestcott@deloitte.com



Gillian Crossan

Global Technology sector Leader | gicrossan@deloitte.com



Lara Abrash

Chair of Deloitte US | labrash@deloitte.com

델로이트 첨단기술, 미디어 및 통신 산업 전문 리더

최오계 파트너

Technology Sector 리더 | 감사부문 | hogchoi@deloitte.com

박형곤 파트너

TME Sector 리더 | 경영자문 부문 | hypark@deloitte.com

조명수 파트너

Digital Finance & Operation 리더 | 컨설팅 부문 | mjo@deloitte.com

정찬욱 파트너

Core Technology 리더 | 컨설팅 부문 | chanjung@deloitte.com

주형열 파트너

Semiconductor Sector 리더 | 컨설팅 부문 | hjoo@deloitte.com

안상혁 파트너

디지털부문 리더/금융산업 총괄 리더 | 컨설팅 부문 | sanghyan@deloitte.com

박지숙 파트너

금융 IT, 오피레이션 리더 | 컨설팅 부문 | jisukpark@deloitte.com

김진숙 파트너

AI 거버넌스 리더 | 경영자문 부문 | jessicakim@deloitte.com

이대의 파트너

Digital CX & MMA 그룹 리더 | 컨설팅 부문 | daelee@deloitte.com

딜로이트 글로벌 TMT 센터 지음

딜로이트의 첨단기술, 미디어 및 통신산업(TMT) 그룹은 세계에서 손꼽히는 산업 전문가 집단으로, 대기업부터 소기업까지 디지털 세계에서 번영하도록 도움을 주는 전문성을 인정받고 있다. 딜로이트 TMT 그룹의 전문가들은 다양한 서비스를 통해 지리적 위치, 가치 사슬 내 위치를 막론하고 모든 기업이 끊임없이 변화하는 산업 환경에서 성공할 수 있도록 든든한 조력자의 역할을 하고 있다.

딜로이트 인사이트 옮김

딜로이트 인사이트는 전 세계 40만 명 이상의 딜로이트 전문가가 제시하는 인사이트를 고객에게 제공하는 사고 리더십(Thought Leadership) 전문 조직이다. 딜로이트의 최신 분석과 차별화된 통찰력을 전달하여 기업과 리더들이 잠재력을 마음껏 펼치도록 지원하는 데 목적이 있다.

글로벌 No.1 회계·컨설팅그룹 딜로이트의 지식 플랫폼인 '딜로이트 인사이트' 모바일 앱과 카카오톡 채널에서 글로벌 비즈니스 주요 트렌드를 신속하게 파악해 보세요.



딜로이트 인사이트 모바일 앱

내 손 안의 경영·산업 트렌드,
언제 어디서나 실시간으로 간편하게 확인



딜로이트 인사이트 카카오톡 채널

바쁜 현대인을 위한 3분 카드뉴스



Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited (“DTTL”), its global network of member firms, and their related entities (collectively, the “Deloitte organization”). DTTL (also referred to as “Deloitte Global”) and each of its member firms and related entities are legally separate and independent entities, which cannot obligate or bind each other in respect of third parties. DTTL and each DTTL member firm and related entity is liable only for its own acts and omissions, and not those of each other. DTTL does not provide services to clients. Please see www.deloitte.com/about to learn more.

Deloitte Asia Pacific Limited is a company limited by guarantee and a member firm of DTTL. Members of Deloitte Asia Pacific Limited and their related entities, each of which are separate and independent legal entities, provide services from more than 100 cities across the region, including Auckland, Bangkok, Beijing, Hanoi, Hong Kong, Jakarta, Kuala Lumpur, Manila, Melbourne, Osaka, Seoul, Shanghai, Singapore, Sydney, Taipei and Tokyo.

This communication contains general information only, and none of Deloitte Touche Tohmatsu Limited (“DTTL”), its global network of member firms or their related entities (collectively, the “Deloitte organization”) is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

No representations, warranties or undertakings (express or implied) are given as to the accuracy or completeness of the information in this communication, and none of DTTL, its member firms, related entities, employees or agents shall be liable or responsible for any loss or damage whatsoever arising directly or indirectly in connection with any person relying on this communication. DTTL and each of its member firms, and their related entities, are legally separate and independent entities.