

딜로이트 첨단기술·미디어·통신 산업 전망



AI 인프라 시대의 도래와 산업 질서의 재편

딜로이트 글로벌 TMT 센터 지음

딜로이트 인사이트 옮김

Deloitte.

Technology, Media & Telecommunications Predictions

AI 인프라 시대의 도래와 산업 질서의 재편

AI 인프라 시대의 도래와 산업 질서의 재편

1판 1쇄 발행 2026. 02. 12

지은이 딜로이트 글로벌 TMT 센터
옮긴이 딜로이트 인사이트

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

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

비매품

CONTENTS

프롤로그	004
감수자 프로필	008

Part 1

인류의 삶에 보이지 않는 형태로 스며든 AI가 창출하는 새로운 가치

1장 검색엔진 내장형 생성형 AI, 독립형 추월	014
2장 서비스형 소프트웨어, AI 에이전트와 결합: 예산·고객경험·인력 구조 재편	028
3장 기하급수적 가치 창출을 위한 AI 에이전트 오피스트레이션	044
4장 휴머노이드·산업용 로봇·드론의 AI 혁신: 자율성과 생산성의 새로운 임계점	070
5장 생성형 AI 다음 단계 성패는 연산력이 좌우	088

Part 2

기술 공급망과 주권의 재편, 새로운 경쟁 전선 형성

6장 AI 칩을 둘러싼 무역 통제와 글로벌 반도체 공급망 재편	110
7장 기술 주권 확보 경쟁	128

Part 3

미디어 산업의 주의력 경제 시대, 서사·신뢰·플랫폼의 재구성

8장 공영방송, 생존전략으로 글로벌 스트리밍 서비스와 협력 선택	148
9장 솟풀이 촉발한 콘텐츠 제작 생태계의 혁신과 플랫폼 신뢰 위기	166
10장 생성형 AI 영상이 불러온 콘텐츠 제작 환경의 혁신과 소셜 플랫폼의 책임	184
11장 영상 팟캐스트, 위협적인 글로벌 미디어·광고 플랫폼으로 진화	198

Part 4

인프라에서 경험으로... 통신의 재정의

12장 차세대 위성 인터넷 시대로의 전환과 통신 인프라·경쟁 질서의 재편	214
13장 통신사 경쟁력은 네트워크 속도에서 고객이 체감하는 가치로 이동	234

주석

252

PROLOGUE

**“일상과 인프라에 깊숙이 스며든 기술...
경쟁의 승패는 더 이상 기술의 성능이 아니라,
접근성·신뢰·가치를 어떻게 설계하느냐가 좌우”**

현대 사회의 기술 경쟁에서는 기술 자체의 발전 양상과 속도보다 기술이 작동하는 방식과 선택되는 기준이 더욱 중요한 요인으로 작용하고 있습니다. 더 빠르고 더 강력한 기술을 확보하는 것은 여전히 중요하지만, 그것만으로는 더 이상 우위를 보장할 수 없습니다. 기술은 인류의 삶 전체에 점점 보이지 않는 형태로 스며들고, 기술 자체의 성능보다 접근성, 신뢰, 가치가 선택의 기준이 되는 시대로 접어들고 있습니다.

인공지능(AI)은 더 이상 특정 앱이나 실험적 도구에 머물지 않습니다. 우리가 늘상 사용하며 익숙해진 기존 앱에서 눈에 보이지 않는 방식으로 작동하며, 기존 서비스와 인프라 속에 내재화되고, 업무와 의사결정의 흐름 자체를 바꾸고 있습니다. 미디어는 포맷과 길이, 플랫폼의 경계를 허물며 재조합되고 있고, 통신과 인프라는 기술적 완성도보다 사용자가 실제로 느끼는 경험과 혜택이 경쟁력을 좌우하는 구도로 재편되고 있습니다.

이러한 전환 국면에서는 리더에게 요구되는 질문도 달라집니다. ‘우리가 가장 앞선 기술을 보유하고 있는가’가 아니라, ‘우리의 기술과 서비스가 규제, 시장, 소비자, 사회적 신뢰라는 현실 조건 속에서 지속적으로 선택될 수 있는가’라는 질문이 더욱 중요한 시점입니다. 기술 전략은 더 이상 IT 부서나 혁신 조직의 과제가 아니라, 거버넌스·자본배분·인력구조·파트너십 전략을 아우르는 경영 의사결정의 중심축이 되고 있습니다.

딜로이트의 ‘첨단기술·미디어·통신(TMT) 산업 전망’은 개별 기술의 성능이나 단기 트렌드를 나열하는 데 그치지 않습니다. 대신, 이들 산업 전반에서 동시에 관측되는 변화들을 하나의 흐름으로 묶어, 향후 3~5년 동안 조직과 산업이 어떤 선택의 기로에 서게 될 것인지, 그리고 미래 리스크에 대비하기 위해 지금 당장 어떤 결정을 내려야 하는지 짚고자 합니다.

본고는 2026년 TMT 영역을 인공지능, 첨단기술, 미디어, 통신 등 4개 파트로 나누어 조망하고 있습니다. 첫 번째 인공지능 파트에서는 앱을 넘어 인프라·에이전트·물리 시스템으로 확장되는 생성형 AI를 중심으로 핵심 AI 기술 동향을 짚어보고, 관련 시장 파급력을 전망하며, 이를 뒷받침하는 오피스트레이션과 연산력, 운영·인력·거버넌스 역량의 실태를 분석합니다.

두 번째 첨단기술 파트에서는 AI 기술의 급격한 발전과 더불어 수출통제 및 지정학적 불안정으로 병목이 확산되며 스트레스테스트·다변화·보안 내재화가 생존 과제가 되고 있는 반도체 공급망의 중기 전망을 제시합니다. 더불어, 규제 부합성과 기술 접근성을 둘러싼 기술·AI 주권 경쟁과 이에 대응하는 전략을 심도 깊게 파헤칩니다.

세 번째 미디어 파트에서는 새로운 미디어의 등장과 변주 속에서 새로운 리스크를 헤쳐가며 파트너십과 혁신적 대응으로 생존을 모색하는 미디어 산업에 대한 중기 전망을 제시합니다. 공영방송은 스트리머·소셜과의 동맹으로 생존 모델을 확장하고 있지만, 구조적 리스크 관리를 서둘러야 하는 실정입니다. 솟폼 시리즈와 생성형 AI 영상은 몰입형 내러티브와 초저비용 제작으로 독립 스튜디오 성장을 촉발하고 광고 효율성을 개선하고 있지만, 진정성 훼손 위험과 규제 압력을 증폭시키고 있습니다.

네 번째 통신 파트에서는 신기술 등장과 소비자 행태 변화에 따른 산업 재편 현황을 분석합니다. D2D·LEO 위성 서비스가 확산되면서 통신 공백을 빠르게 커버하고 있지만, 수익화·용량·주파수 규제가 복잡해지며 협력과 경쟁을 병행하는 전략 설계가 필요한 상황입니다. 통신 네트워크 성능이 개선 수준을 체감하기 힘들 정도의 성숙기에 접어든 선진국에서는 리워드가 통신사들의 핵심 차별화 요인으로 부상하고 있습니다.

이제 경쟁의 승패는 기술의 속도나 성능이 아니라, 기술이 작동하는 방식과 선택되는 기준을 누가 더 정확히 설계하고 실행하는가에 달려 있습니다. AI는 보이지 않게 일상과 업무의 흐름에 스며들고, 미디어는 경계를 허물고 재조합되며, 통신과 인프라는 경험과 가치 중심으로 재평가되고 있습니다. 이러한 변화는 개별 산업의 문제가 아니라, 규제·자본·인력·신뢰가 동시에 재배치되는 구조적 전환입니다. 따라서 리더에게 요구되는 역할은 최신 기술을 가장 먼저 도입하는 것이 아니라, 현실 조건 속에서도 지속적으로 선택될 수 있는 기술과 사업의 기준을 세우고, 이를 조직과 투자, 파트너십에 일관되게 반영하는 결단입니다. 이제 관건은 변화의 방향을 읽는 것이 아니라, 그 방향을 기준과 실행으로 전환하는 속도와 명확성입니다.



손재호
한국 딜로이트 그룹
성장전략부문 대표

감수자 프로필

**최호계** 파트너

한국 딜로이트 그룹 TMT Industry 리더

회계감사 및 어슈어런스 전문가로, 국내외 유수의 테크·미디어·통신 기업 그룹 감사를 총괄 및 지원하고 있다. IFRS·US GAAP·PCAOB 등 글로벌 회계감사 기준에 정통한 전문가로서 삼성, KT, 두산, 마이크로소프트 등 대표 TMT 기업들에 대한 감사 및 자문 경험을 보유하고 있다. TMT Industry 리더를 역임하고 있으며, 해당 산업에서 20년 이상 쌓아온 경험과 전문성을 바탕으로 산업 전반의 복합 이슈에 대한 실무적 인사이트를 제공하고 있다. 고려대학교를 졸업하였으며 한국공인회계사(KICPA) 자격을 보유하고 있다.

**박형곤** 파트너

한국 딜로이트 그룹

TM&E(통신, 미디어 및 엔터테인먼트) Sector 리더, Monitor Deloitte

한국 딜로이트 그룹 내에서 전략 컨설팅을 담당하는 Monitor Deloitte 소속 파트너로서, 22년 이상 비즈니스 전략 컨설팅 업무를 성공적으로 수행해 왔다. 전자, 하이테크, 미디어 및 자동차 등 부문에 속한 국내외 주요 기업을 대상으로 100여개 이상의 사업 성장 전략, 신사업 및 신시장 전략 프로젝트 경험을 쌓아왔다. 딜로이트 컨설팅 코리아에서 TMT(Technology, Media & Telecommunication) 산업을, 한국 딜로이트 그룹에서는 TME(Telecommunication, Media & Entertainment) 섹터를 리드하고 있다. 연세대학교를 졸업하였으며 경영학과 응용통계학을 전공하였다.

**이대의** 파트너

한국 딜로이트 그룹

컨설팅 부문 Customer 본부장, Deloitte Digital 리더

딜로이트 컨설팅 코리아 내 Customer 사업 본부를 리드하며, 국내 외 하이테크-커머스-제조-소비자 기업의 고객 경험 전략, 마케팅 성과 분석, 영업/서비스 프로세스 개선 및 Mar-tech 플랫폼 구축운영 컨설팅 서비스를 제공하고 있다. 또한 Deloitte Digital 한국 내 리드를 맡으면서 디지털 마케팅 전반의 전략~기획~컨텐츠 크리에이티브~성과 분석~플랫폼 End-to-End 서비스를 글로벌 팀들과 공조하며, 급격한 AI 전환 시대에 가장 선도적이고 혁신적인 프랙티스를 실제적으로 구현하고 있다. 영국 랭커스터(Lancaster) 대학교 Marketing 학/석사 학위를 취득하였다.

**정창모** 파트너

한국 딜로이트 그룹 One AI 파트너

딜로이트 컨설팅의 AI 서비스 전문조직인 One AI를 리드하고 있다. AI 알고리즘 분야의 전문가로 20년 이상의 산업경력을 보유하고 있으며, 국내 주요 대기업에서 Data Scientist로서 경력을 쌓으면서 AI 전략기획, 모델 개발/구축/운영, 데이터 엔지니어링 등 AI분야의 전 영역을 실무자와 리더로 경험하였다. 전문분야는 수리통계 기반 의사 결정 최적화 및 자연어처리(NLP)며 최근에는 AI Agent 서비스 개발에 집중하고 있다. 최신 이론의 현장 접목을 위해 노력하고 있으며 지능정보시스템학회 부회장, 한국경영학회 이사 등을 역임하였다. 현재 주요 자문분야는 수리통계 알고리즘 모델링, Gen AI 서비스 기획 및 개발, AI 운영모델 설계(MLOps, LLMOps)이다. 최종학력은 연세대학교 경영과학 박사로, University of Pittsburgh 데이터 사이언스 석사, 서울대학교 경영학 석사 학위 또한 보유하고 있다.



유선희 파트너

한국 딜로이트 그룹 One Cyber and Resilience 파트너

한국 딜로이트 그룹 사이버 본부 내 Cyber Risk and Compliance 그룹 파트너로, 국내외 정보보호 및 개인정보보호 컴플라이언스 컨설팅 서비스를 제공하고 있다. 19년 이상 금융 및 첨단기술, 미디어, 통신 분야의 글로벌 기업을 대상으로 K-ISMS 준비, 사이버 보안 진단, 디지털전환 워크숍, 보안 거버넌스 구축, SOC2/3 리포트 발행 등 폭넓은 프로젝트를 수행해 왔다. AI for Cyber/Cyber for AI 관련 이니셔티브를 담당하고 있으며, AI를 활용한 사이버 보안 대책 수립 및 AI 도입 시 발생하게 되는 보안 위협에 대한 대응 관련 자문을 수행하고 있다. 이화여자대학교 통계학과를 졸업하였고, 메인주 USCPA(미국공인회계사) 및 CISA(공인정보시스템감사사) 자격을 보유하고 있다.



이중희 파트너

한국 딜로이트 그룹 Physical AI 리더

제조업 중심의 Smart Factory와 Physical AI 혁신을 리드하고 있으며, 국내 및 글로벌 주요 제조 기업을 대상으로 스마트팩토리 전략 수립, 공장 자동화, 로봇기반 Vision AI 적용, 예지보전/설비진단 모델 개발 등 다수의 프로젝트를 주도하며 현장의 생산성과 안전성을 높이는 실질적 성과를 만들어 왔다. 글로벌 제조기업 대상의 유스케이스 발굴과 확산을 통해 로봇 자동화, AI 기반 생산 혁신 모델을 현장에 적용하는 데 앞장서고 있으며, Smart Factory·Factory Automation·AI 사업모델 개발 전 영역을 아우르는 24년 경력의 전문가이다. Boston University 학사이며 미국공인회계사(AICPA) 자격을 보유하고 있다.

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 얹혀 다양한 상호작용을 통해 거대한 혁신의 생태계를 형성하고 있습니다.

Part

01

인류의 삶에 보이지 않는 형태로 스며든 AI가 창출하는 새로운 가치

1장 검색엔진 내장형 생성형 AI, 독립형 AI 추월

2장 서비스형 소프트웨어, AI 에이전트와 결합: 예산·고객경험·인력 구조 재편

3장 기하급수적 가치 창출을 위한 AI 에이전트 오케스트레이션

4장 휴머노이드·산업용 로봇·드론의 AI 혁신: 자율성과 생산성의 새로운 임계점

5장 생성형 AI 다음 단계 성패는 연산력이 좌우

1장

검색엔진 내장형 AI, 독립형 AI 추월

생성형 AI는 독립형 앱이 아니라 검색·이커머스·소셜미디어 등 기존 서비스에 내장될 때 가장 빠르게 확산되고 있으며, 사용 방식 역시 능동적 이용에서 수동적·비노출형 이용으로 이동하고 있다. 이질감 없는 사용자 경험(UX)과 즉각적 효용 덕분에 고령층까지 빠르게 흡수되며 생성형 AI의 대중화를 가속하고 있다. 앞으로 기업의 경쟁력은 모델 성능 자체보다, 기존 트래픽과 사용자 워크플로에 AI를 얼마나 자연스럽게 통합해 핵심 성과지표(KPI)를 개선할 수 있는가에 달려 있다.

핵심 내용 요약 (Executive Summary)

▷ 확산의 출발점: 생성형 AI는 독립형 앱 보다 기존 서비스에 내장될 때 사용자 증가 속도가 빠르다

- 검색엔진·브라우저·이커머스·소셜미디어 등 일상 서비스가 확산의 주 무대
- 사용자는 AI를 “찾아 쓰는 것”이 아니라 이미 쓰고 있는 서비스 안에서 자연스럽게 사용

▷ 사용 방식의 전환: 능동적 이용 (독립형 AI) → 수동적 이용 (내장형 AI)

- 독립형 AI 앱 중심의 능동적·명시적 이용에서 기존 앱에 내장된 수동적·비노출형 이용으로 중심 이동
- 2026년 기준, 검색 내장형 AI의 일일 이용률은 독립형 앱의 3배 이상, 2027년에도 이 격차는 유지되며, 검색 기반 이용이 주류 채널로 고착

▷ 왜 이렇게 빨리 퍼지는가: 이질감 없는 UX와 즉각적 효용

- 검색 요약, 리뷰 정리, 콘텐츠 생성 등 반복·소모적 작업을 AI가 자동 수행
- 사용자는 AI를 인식하지 않아도 핵심 기능 개선을 즉시 체감
- 재학습·프롬프트 부담이 없고 진입장벽이 낮아 전 연령대로 빠른 확산
- GPS·결제·인증 기술처럼 독립형 → 내장형으로 진화하는 기술 확산 경로와 동일

▷ 기업의 전략적 선택과 한국 플랫폼 시사점

- 생성형 AI의 경쟁력은 모델 성능보다 기존 트래픽·워크플로에 얼마나 자연스럽게 스며드는가
- AI 기업은 플랫폼 통합을 통한 채택률 확대 vs 독립형 유지로 사용자당 수익 극대화 중 선택 필요
- 플랫폼 기업은 AI를 별도 서비스가 아닌 검색·추천·전환·운영을 대체·보완하는 기본 레이어로 재설계해야 성과 창출 가능
- AI 성과의 핵심 지표는 신규 사용자 수가 아니라 기존 KPI(체류·전환·재방문) 개선

기존 애플리케이션에 생성형 AI가 내장될 경우 챗GPT(ChatGPT)와 같은 독립형 생성형 AI 애플리케이션보다 사용자 수가 더 빠르게 증가할 전망이다. 사용자가 앱을 직접 방문해 사용하는 생성형 AI보다, 기존 일상적 서비스 안에서 자연스럽게 AI 기능을 이용하는 방식이 훨씬 파급력이 크기 때문이다.

딜로이트는 2026년 생성형 AI 사용자가 급격히 증가할 것으로 전망한다. 이러한 성장세는 기존 디지털 애플리케이션에 내장된 생성형 AI가 견인할 것이다. 독립형 생성형 AI 앱을 직접 찾아 사용하는 능동적·명시적 이용보다, 이미 쓰고 있는 앱·검색엔진·브라우저 안에 탑재된 생성형 AI를 자연스럽게 활용하는 수동적 이용이 더 많아질 것이라는 의미다.

딜로이트는 특히 검색엔진에 탑재된 생성형 AI의 사용량 증가에 주목했다. 분석 결과, 생성형 AI 요약을 포함한 검색 결과를 매일 이용하는 비율은 텍스트·오디오·이미지·비디오·코드·멀티모달 등 모든 유형의 독립형 생성형 AI 도구를 이용하는 비율을 세 배 상회할 전망이다.¹ 2026년 선진국 기준으로 성인 29%가 매일 한 번 이상 생성형 AI 요약이 포함된 검색을 이용할 것으로 예상되는 반면, 독립형 생성형 AI 앱을 매일 사용하는 비율은 10% 수준에 그칠 것으로 보인다. 2027년에는 두 채널의 일일 이용률이 모두 증가하겠지만, 검색 기반 이용(40%)이 독립형 앱 이용(13%)을 세 배 웃도는 추세는 유지될 전망이다. 해당 전망치는 비교

의 정확성을 위해 수동적 생성형 AI 활용 중 검색 요약 하나의 사례만 분석했다(그림 1).

아울러 여타 유형의 앱에 내장된 생성형 AI의 수동적 이용 역시 빠른 증가가 예상된다. 특히 현재 AI 사용률이 상대적으로 낮은 고령층에서 가장 빠른 성장세가 전망된다.

앱 내장형 생성형 AI vs. 독립형 생성형 AI 비교

생성형 AI는 검색, 이커머스, 소셜미디어, 온라인 뉴스 등 이미 널리 쓰이는 서비스에 보이지 않는 방식으로 내장됨으로써 활용이 늘고 있다. 이는 사용자가 챗GPT나 제미나이(Gemini)와 같은 독립형 생성형 AI 앱을 직접 방문해 텍스트·이미지·코드 등 특정 결과물을 생산하는 기존의 사용 방식과 명백히 다르다.

수동적 생성형 AI 이용이란 생성형 AI가 다른 앱에 기본 사양으로 내장돼 있지만 사용자에게는 노출되지 않는 형태를 말한다. 사용자는 생성형 AI를 직접 사용하고 있다고 느끼지 않지만, 실제로는 이 기술이 해당 앱 경험의 핵심을 이룬다. 앱에 내장된 생성형 AI는 수십, 수백 개의 검색 결과를 하나로 요약하고, 수천 개의 상품 리뷰를 자동으로 정리하고, 소셜미디어와 온라인 뉴스에 배포되는 콘텐츠를 생성하는 등 안 보이는 곳에 숨어서 핵심 기능을 자동 수행한다.

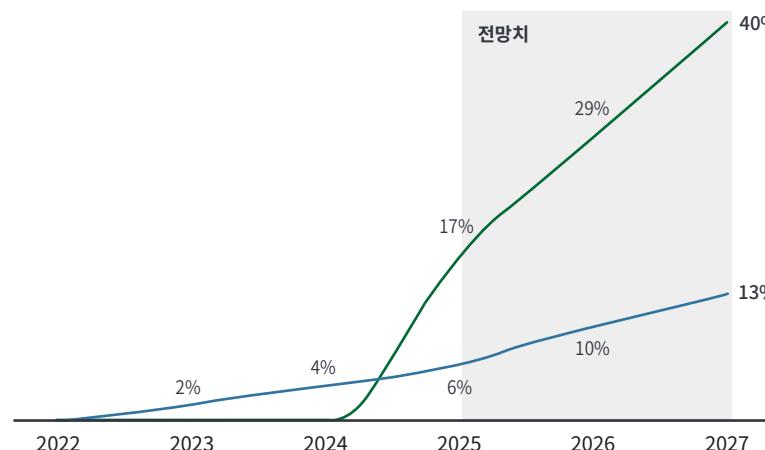
그림 1

생성형 AI의 수동적 검색 요약 기능 일일 이용률, 독립형 생성형 AI 앱 추월 전망

검색엔진 내 생성형 AI 요약 기능 / 독립형 생성형 AI 앱을 매일 사용하는 사용자 비율

(기간: 2023~2027년)

■ 수동적 이용 일일 이용률 ■ 능동적 이용 일일 이용률



참조: 2023~2025년 수치는 가중치 기준 / 16~75세 응답자 대상 조사 결과: 2023~2024년 매년 각각 4,150명
출처: Deloitte forecasts based on Deloitte Digital Consumer Trends, UK, 2025

딜로이트는 2025년에도 이러한 수동적 이용 비율이 독립형 생성형 AI 이용 비율을 웃돌았을 것으로 추정한다. 실제로 영국 시장 조사에서 2025년 중반 기준 네 가지 유형의 수동적 생성형 AI 애플리케이션을 사용해 본 적이 있다는 응답자는 약 4분의 3에 달한 반면, 독립형 생성형 AI 앱을 사용해 본 적이 있는 비율은 47%에 그쳤다.²

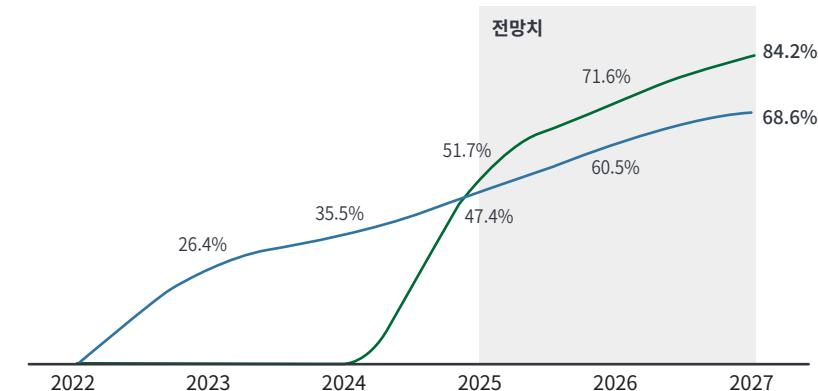
새로운 기술 확산을 평가하는 또 다른 지표는 ‘특정 기간 내 한 번이라도 이용한 경험’이다. 수동적 생성형 AI 앱은 2024년 5월 미국에서 검색 요약 기능이 처음 도입되며 등장했고,³ 이후 같은 해 11월 여타 검색엔진으로 확산되며 시장 규모가 커졌다.⁴ 이는 2022년 말 첫 독립형 생성형 AI 앱이 출시된 지 거의 2년 후다. 그럼에도 딜로이트는 2026년 중반까지 검색 요약을 한 번이라도 생성해 본 성인 비율이 72%로, 독립형 생성형 AI 사용 경험 비율(61%)을 앞지를 것으로 전망한다(그림 2).

그림 2

생성형 AI의 수동적 검색 요약 기능을 한 번이라도 이용한 사용자 비율, 독립형 생성형 AI 앱 추월 전망

검색엔진 내 생성형 AI 요약 기능 / 독립형 생성형 AI 앱을 한 번이라도 이용한 사용자 비율
(기간: 2023~2027년)

■ 수동적 이용 일일 이용률 ■ 능동적 이용 일일 이용률



참조: 가중치 기준 / 16~75세 응답자 대상 조사 결과: 2023~2024년 매년 각각 4,150명
출처: Deloitte forecasts based on Deloitte Digital Consumer Trends, UK, 2025

이는 생성형 AI가 독립형 서비스 대신 기존 주류 앱 내부의 기본 프로세스로 작동할 때 훨씬 더 광범위하게 확산되고 일상화된다는 뜻이다. 독립형 생성형 AI가 쓸모없다는 뜻이 아니라, 생성형 AI가 검색, 소셜미디어, 이커머스, 업무용 앱 등 이미 대중적으로 사용되는 서비스에 통합될 경우 사용 빈도와 접근성이 훨씬 높아지기 때문이다. 이와 함께 중기적 관점에서 중요한 질문도 제기된다. 독립형 생성형 AI 앱은 장기적으로 소셜미디어나 검색엔진만큼 대중화될 수 있을까? 아니면 매일 웹을 이용하는 사용자 전체 중 독립형 생성형 AI 사용자 비율이 현재 약 20% 수준에서 성장을 멈출 것인가?

생성형 AI의 수동적 이용 선호가 의미하는 바는?

검색엔진, 소셜미디어, 이커머스는 이미 전 세계에서 가장 자주 사용되는 디지털 앱이다. 매일 150억 건이 넘는 검색이 이뤄지고, 사용자는 하루 평균 2시간 이상 소셜미디어에 머문다.⁵ 2025년 1분기 미국의 이커머스 매출은 3,000억 달러에 달했다.⁶ 이런 환경에서 이용자들은 낯선 생성형 AI 챗봇에서 정보를 찾기보다, 익숙한 검색 도구에 내장된 생성형 AI 기능을 더 자연스럽게 사용할 가능성이 크다.

2026년에도 생성형 AI가 검색 비즈니스 모델에 미치는 영향에 대한 논쟁은 계속되겠지만, 생성형 AI 기능이 강화된 검색엔진이 챗GPT나 신세시아(Synthesia) 같은 독립형 앱의 인기를 잠식할지도 주요 관전 포인트가 될 전망이다.⁷ 딜로이트 조사에 따르면, 업무용으로 생성형 AI

가 가장 많이 사용되는 용도 역시 검색이다. 따라서 현재 독립형 AI 앱에서 검색하던 일부 이용자들이 다시 기존 검색 앱으로 돌아갈 수도 있다.

생성형 AI의 수동적 이용이 능동적 이용을 추월한 속도는 예상보다 빠르지만, 동시에 예견된 흐름이기도 하다. 독립형 생성형 AI는 새로운 기술이자 다소 실험적 성격을 띠는 데다 결과물의 품질이 사용자 프롬프트 역량에 크게 좌우된다.⁸ 실망스러운 결과가 나올 경우, 사용자는 모델의 한계보다 자신의 프롬프트 작성 능력을 문제로 느낄 수 있다. 반면 수동적으로 이용하는 생성형 AI는 기존 서비스 안에 자연스럽게 녹아든 사용자 친화적(low-friction) 기술^{*}이다. 검색엔진부터 이커머스, 소셜미디어, 업무용 솔루션까지 재검색 필요가 거의 없고, 기술이 노출되지 않으며, 익숙한 사용자 경험을 제공하고, 진입장벽은 매우 낮다. 이 때문에 접근성과 활용성이 독립형 앱보다 월등히 높고, 수요 또한 급증할 잠재력이 훨씬 크다. 특히 검색 결과 자동 요약 기능은 사용자가 원래 해야 했던 여러 링크 클릭, 수십 개 페이지 탐색, 요약 정리 등 반복적 작업을 AI가 대신 수행해 주는 효과가 있다. 이는 결제부터 배송까지 한 번에 처리하는 전자상거래의 원터치 체크아웃, 모바일 뱅킹의 얼굴 인식 인증이 가져온 사용자 경험 변화를 떠올리게 한다. 즉 기술이 노출되지 않을 수록, 사용자 경험에 더 깊게 스며들어 확산 속도가 더 빨라지는 것이다.

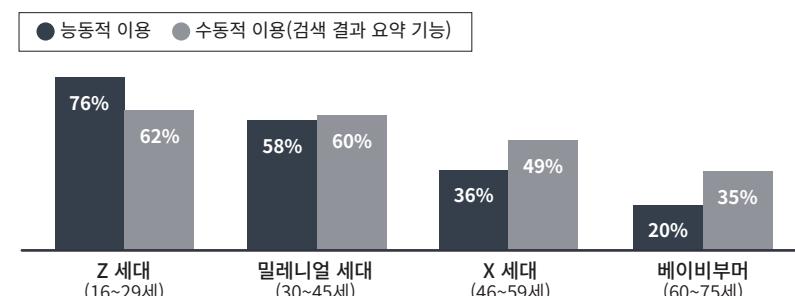
* 사용자 친화적(low-friction) 기술은 사용자가 별도의 학습, 설정, 의사결정 부담 없이도 자연스럽게 사용하게 되는 기술을 의미한다. 사용자가 설정 및 전환 같은 특별한 행동을 하지 않기 때문에 ‘사용한다’는 인식을 하지 않는 상태에서 기능이 작동함으로써, 기존 서비스 및 워크플로에 내재화되는 기술이다.

베이비부머, 수동적 생성형 AI 기능에 적극적 반응

수동적 생성형 AI의 강점인 접근성과 편의성을 특히 빠르게 체감하는 세대는 베이비부머다. 새로운 독립형 생성형 AI 도구에 익숙해지는 데 부담을 느끼는 베이비부머는 검색 요약처럼 기존 서비스 안에 자연스럽게 통합된 기능은 훨씬 적극적으로 이용하고 있다. 2025년 중반 기준 딜로이트 조사에 따르면, 베이비부머는 독립형 생성형 AI에 대해 인지도는 58%였으나 실제 사용 경험은 20%에 불과했다. 반면, Z세대는 76% 가 생성형 AI 도구를 사용한 경험이 있는 것으로 나타나, 세대간 격차가 매우 커졌다. 그러나 검색 요약 기능만 놓고 보면 상황이 달라진다. 베이비부머의 검색 요약 이용률은 35%로, 독립형 생성형 AI 이용률 대비 75% 나 높은 수준을 보였다. 즉, 베이비부머는 새로운 AI 앱을 배우는 데는 소극적이지만, 익숙한 서비스 내에 통합된 AI 기능에는 빠르게 익숙해지는 경향을 보인다.

딜로이트는 이러한 흐름이 앞으로 더욱 강화돼, 베이비부머의 검색 요약 이용률이 2026년 49%에서 2027년 59%로 상승할 것으로 전망한다. 이는 2027년 독립형 생성형 AI 이용률 전망치인 32%를 대폭 상회하는 수치다(그림 4).

그림 3
베이비부머, 독립형 생성형 AI보다 앱 내장형 기능 더 많이 활용

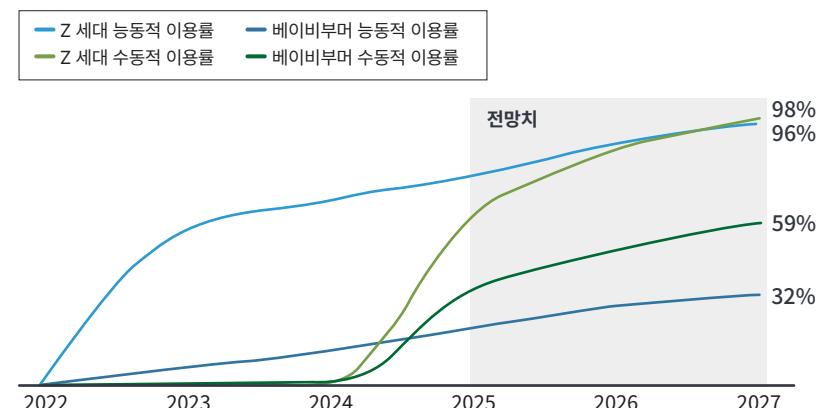


참조: 가중치 기준 / 16~75세 응답자 대상 조사 결과: 2025년 4,150명

출처: Deloitte forecasts based on Deloitte Digital Consumer Trends, UK, 2025

그림 4
Z 세대와 베이비부머, 앱 내 생성형 AI의 검색 요약 기능 한 번이라도 사용한 비율이 독립형 앱 사용 비율 능가

Z 세대 및 베이비부머 중 앱 내 생성형 AI 검색 요약 기능 / 독립형 생성형 AI 앱을 한 번이라도 사용한 비율



참조: 가중치 기준 / 16~75세 응답자 대상 조사 결과: 2023~2024년 매년 각각 4,150명

출처: Deloitte forecasts based on Deloitte Digital Consumer Trends, UK, 2025

수동적 이용의 확산이 생성형 AI 시장에 가져올 변화

생성형 AI는 시대를 대표하는 핵심 기술이지만, 그 잠재력은 기존의 주류 앱 내부에 보이지 않는 형태로 필수 통합될 때 극대화될 가능성이 크다.

오늘날 가장 중요한 기술 중 상당수는 처음부터 독립형 기능, 즉 전용 기기에서 출발했다. GPS(global positioning system, 위성항법장치)는 불과 몇 년 전까지만 해도 출장이나 여행 때 내비게이션으로 따로 들고 다니던 물리적 장치였다. 그러나 이 기능은 스마트폰과 여러 앱에 통합됐고, 지금은 단순 길 찾기를 넘어 수많은 앱과 서비스에 내장된 필수적이지만 눈에 띄지 않는 기술로 자리 잡았다. 이처럼 기술은 독립형 단계에서 출발해, 일상적 디지털 환경 속에 자연스럽게 녹아드는 방향으로 진화해 왔다.

생성형 AI도 마찬가지다. 기존 앱을 완벽하게 만들지는 못하더라도, 기능적 효율성과 편의성을 크게 개선한다. 예를 들어 검색 결과를 자동으로 요약하는 과정에서 오류가 발생할 수 있지만, 많은 사용자에게는 이러한 오류보다 검색을 더 쉽게 할 수 있다는 장점이 더 중요하다. 즉, 사용자는 기술의 확률적 특성으로 인한 일부 오류를 감수하더라도, 시간을 절약하고 정보를 빠르게 파악할 수 있다는 이유로 내장형 생성형 AI 기능을 선호한다.

이에 따라 독립형 생성형 AI 앱을 제공하는 기업들이 전략적 선택의 순간을 맞이하게 됐다. 2026년 이들 기업은 다음 두 가지 중 무엇을 우선

할지 결정해야 한다. 자사 기술을 다른 대형 앱 내부에 통합해 이용률을 높이는 데 주력할 것인가? 아니면 독립형 인터페이스를 유지해 사용자 당 수익(RPU)을 극대화할 것인가? 일부 선도 기업은 두 전략을 병행할 수 있겠지만, 대부분 기업은 현실적으로 둘 중 하나를 선택해야 하는 갈림길에 놓일 것이다.

Korean Perspectives

**AI의 승부처는 모델 성능이 아니라
‘기존 플랫폼에 얼마나 자연스럽게 스며드는가’에 있다.**

H胳膊의 생성형 AI 경쟁력은 더 이상 모델 성능 자체에 있지 않다. 핵심은 기존 트래픽과 사용자 워크플로에 생성형 AI가 얼마나 자연스럽게 스며들 수 있는가다. 실제로 본고에서 독립형 생성형 AI로 분류된 제미나이 역시 구글의 검색엔진에 깊이 통합되며, 독립형 인터페이스 중심의 챗GPT의 이용 시간을 일부 잠식하고 있다. 이는 생성형 AI의 확산 경로가 ‘새로운 앱’이 아니라 ‘기존 플랫폼의 기본 기능’임을 보여주는 대표적 사례다.

그러므로 생성형 AI를 별도의 서비스로 키우기보다는, 검색·메신저·콘텐츠·이커머스 등 기존 서비스의 핵심 기능을 대체하거나 보완하는 기본 레이어로 재설계하는 접근이 더욱 효과적이다. 사용자가 AI를

직접 ‘쓴다’고 인식하지 않더라도, 체류 시간·전환율·재방문율 같은 기준 KPI가 개선되는 구조가 가장 강력한 AI 전략이 되기 때문이다. 특히 검색 결과 요약, 리뷰 통합, 콘텐츠 자동 정리처럼 반복적인 인지 노동을 제거하는 기능은 단기간 내 체감 가능한 효용을 만들어내며, 생성형 AI의 가치를 가장 빠르게 증명한다.

이러한 내장형 AI 사용경험은 고령층 확산 측면에서도 중요한 기회를 제공한다. 복잡한 프롬프트나 새로운 사용 학습이 필요 없는 AI 기능은 그동안 AI 활용에 소극적이던 사용자층까지 빠르게 흡수하며 시장 저변을 넓힌다. 반대로 독립형 AI 앱 경쟁이나 AI 자체의 성능 비교에 과도하게 집착할 경우, 플랫폼이 이미 보유한 분배력과 진입장벽이라는 구조적 강점을 스스로 약화시킬 위험이 있다.

따라서 생성형 AI의 성과는 별도의 MAU나 AI 전용 지표로 평가하기보다, 기존 서비스 KPI가 얼마나 개선되었는지를 기준으로 측정하는 것이 바람직하다. AI를 위한 AI 지표는 오히려 전략적 판단을 흐릴 수 있다. 결론적으로 한국 플랫폼 기업의 진정한 승부처는 AI를 얼마나 잘 보여주느냐가 아니라, 사용자 경험 속에 얼마나 완벽하게 숨기느냐에 있다. 이러한 접근은 테크 플랫폼에 국한되지 않고, 이커머스나 금융 애플리케이션 등 다양한 산업으로 확장될 수 있으며, 향후에는 산업별 특성을 반영한 생성형 AI 통합 전략이 경쟁력을 좌우하게 될 것이다.



정창모 파트너

한국 딜로이트 그룹
One AI 파트너

2장

서비스형 소프트웨어, AI 에이전트와 결합: 예산·고객경험·인력 구조 재편

2026년을 기점으로 서비스형 소프트웨어(SaaS, software-as-a-service)는 기능 중심 도구에서, AI 에이전트가 자율적으로 업무를 실행하고 조율하는 워크플로 플랫폼으로 전환되고 있다. 따라서 경쟁의 중심 역시 개별 기능이 아니라 에이전트의 생성·통합·오케스트레이션 역량으로 이동하고 있다. 이 과정에서 사용자 수 기반 구독 모델은 약화되고, 사용량·성과 기반의 하이브리드 과금 모델이 부상하고 있으며, 신뢰할 수 있는 사용자경험(UX)과 운영 투명성이 핵심 차별화 요소로 자리 잡고 있다. 이러한 변화 속에서 SaaS 도입 기업과 SaaS 벤더 모두 데이터, 거버넌스, 영업, 계약, 운영 모델 전반에 대한 전면적인 재설계가 요구되고 있는 상황이다.

핵심 내용 요약 (Executive Summary)

▶ (전환의 본질) SaaS의 에이전트화(agentification)

- SaaS는 기능 제공 도구에서 자율적으로 실행하는 워크플로 서비스로 진화
- AI 에이전트는 ‘보조’에서 기업 업무를 직접 수행하는 실행 주체로 이동
- 2026년은 AI 에이전트가 기업 운영을 실시간으로 주도하기 시작하는 전환점

▶ (경쟁 축의 이동) 기능 → 에이전트 플랫폼

- SaaS 경쟁력은 개별 기능이 아니라 AI 에이전트의 생성·통합·오케스트레이션 역량으로 이동
- 기존 SaaS 벤더는 기존 제품에 에이전트 추가 또는 에이전트 개발 프레임워크 제공
- AI 네이티브 스타트업은 특정 영역에서 기존 SaaS의 잠재적 대체자로 부상
- 중장기적으로 멀티 에이전트 생태계와 관리 계층이 핵심 가치 영역으로 부상

▶ (비즈니스 모델의 핵심 난제) 과금과 신뢰

- 사용자 수 기반 구독 모델 약화, 사용량·성과 기반 하이브리드 과금 부상
- 에이전트가 사람 역할을 대체하며 가치 측정·비용 예측이 어려워짐
- 과금 기준 정의(에이전트·작업·성과), 계측·가시성, 계약 구조 재정립이 필수
- UI/UX, 설명가능성, 추적가능성, 되돌림 가능성성이 신뢰의 핵심 요소로 부상

▶ 기업과 벤더의 전략적 시사점

- (도입 기업) 데이터 관리·거버넌스를 최우선으로 강화하고, 복잡해지는 벤더·가격 구조에 대비하며 인력을 AI 에이전트 오케스트레이터로 전환
- (SaaS 벤더) 기능 경쟁을 넘어 상호운용성과 에이전트 통합 역량을 확보하고, 신뢰 기반 UX (투명성·추적성)를 차별화 요소로 구축

AI 에이전트를 활용하는 SaaS가 늘고 있다. 이에 따라 기업의 소프트웨어 사용 방식과 고객 경험도 근본적으로 바뀌고 있다. 구독 및 과금 모델, 서비스 역량, 사용자의 기대치 등 전반에서 변화가 이루어질 것이다.

AI 에이전트란?

AI 분야의 지능형 에이전트란 주변 환경을 인지하고, 목표를 달성하기 위해 자율적으로 행동하며, 기계학습이나 지식 획득을 통해 성능을 지속적으로 향상시킬 수 있는 기술을 의미한다.¹

에이전틱 인공지능(agentsic AI) 기술이 성숙기에 돌입하면서 기업용 SaaS 업체들은 AI 에이전트를 생성·통합·오케스트레이션할 수 있는 플랫폼을 확대하고 있다. SaaS 고객사들이 소프트웨어를 구매하고 활용하는 방식 또한 근본적으로 전환되고 있다.

2026년, SaaS 애플리케이션은 더욱 고도화된 지능형, 맞춤형, 적응형 및 자율형 소프트웨어로 진화할 전망이다. 이는 소프트웨어가 고정형이 아니라, 경험을 통해 학습하는 실시간 통합형 워크플로 서비스로 변화한다는 의미다.

과금 모델도 바뀔 것이다. 기존의 구독 또는 사용자 수 기반(seat-based) 라이선스는 줄고, 사용량(usage-based) 및 성과(outcome-based) 기반 구조가 결합된 하이브리드 과금 모델이 부상할 가능성성이 크다. 이처럼 SaaS의 지능화와 자율화가 진행되면, 소프트웨어 구현 방식과 수익화 전략은 더욱 복잡해지고, 그 과정에서 SaaS 비즈니스 모델 전체가 재편될 수 있다. 즉, 2026년은 AI 소프트웨어가 기업의 워크플로를 실시간으로 주도하고 실행하기 시작하는 전환점이 될 것이다.

AI 에이전트, 2026년부터 SaaS 시장을 점진적 재편 전망

시장 전반의 AI 도입 흐름을 이해하기 위해 전체 방향성을 살펴볼 필요가 있다. 딜로이트의 미국 기업 리더 대상 ‘2025 테크 밸류’(2025 Tech Value) 조사에 따르면, 응답자의 57%가 연간 디지털 전환 예산의 21~50%를 AI 자동화에 투자하고 있다고 답했으며, 절반 이상을 투자하는 응답자는 20%에 달했다(매출 130억 달러 기준 평균 7억 달러 규모).² 또한 응답자의 약 4분의 3이 지난 12개월 동안 AI 및 생성형 AI 기술 역량을 투자 최우선순위로 선택했고, 39%는 에이전틱 AI에 자금을 투입한 것으로 나타났다.

딜로이트는 해당 조사 결과를 기반으로 2026년에는 응답 기업의 절반이 디지털 전환 예산의 50% 이상을 AI 자동화에 투자할 것이라는 전망을 도출했다. 또한 에이전틱 AI에 자금을 투입하는 기업 비중은 75%까지 확대될 수 있다고 전망한다. 이번 조사는 미국 기업을 대상으로 했지만, 미국 이외 기업도 1~2년의 시간차를 두고 비슷한 흐름을 따를 가능성이 크다. SaaS는 디지털 전환의 핵심 기반이기 때문에, 이처럼 관련 투

자가 확대되면 향후 1년간 SaaS 내 자율형 AI 에이전트 투자도 증가할 것으로 예상된다.

그렇다면 이 모든 투자와 기술 발전은 장기적으로 어떤 방향으로 이어질까? 일각에서는 기업 애플리케이션의 일부 혹은 전체가 결국 AI 에이전트로 대체될 수 있다는 낙관적 주장이 제기되고 있다.³ 일부 분야에서는 가능성은 있지만, 2026년 내 기업 전반에서 이러한 일이 현실화되기는 요원하다. 에이전틱 AI의 빠른 기술 발전에도 불구하고, 전통 SaaS 기업은 복잡한 워크플로 전반에 걸쳐 방대한 기반을 구축해 왔기 때문에 완전한 대체에는 최소 5년 이상의 시간이 필요할 것으로 보인다.⁴

2026년에는 기술적 실험이 활발히 이뤄지고, 기존 기능의 확장과 AI 퍼스트(AI-first)* 기업간 경쟁이 본격화되며, SaaS 시장 구조가 점진적, 단계적으로 재편될 전망이다. 변화의 속도가 완만한 이유는 단순 기술 혁신을 넘어 벤더와 사용자 모두 비즈니스 및 운영 모델을 구조적으로 전환해야 SaaS의 에이전트화(agentification)를 실현할 수 있기 때문이다.

* AI 퍼스트(AI-first)는 AI를 단순한 보조수단이 아닌 경영 설계의 최우선 기준이자 핵심 전제 조건으로 삼는 경영 패러다임을 뜻한다.

에이전틱 AI 시대, 한층 복잡해지는 SaaS 시장

대다수 기업의 최고정보책임자(CIO)와 최고기술책임자(CTO)는 비용을 절감하고 벤더 수를 축소해야 한다는 압박을 계속 받고 있다.⁵ 에이전틱 AI 시대에 들어서면서, 기업들은 AI 에이전트 기반 솔루션 투자 전환의 시점과 방식을 심도 있게 고민하고 있다. 에이전틱 AI를 기업의 차세대 핵심 경쟁력 강화 요인으로 간주하기 때문이다.

현재 주요 SaaS 기업이 고객에게 에이전틱 AI 기능을 제공하는 방식은 대체로 두 가지로 수렴된다. 첫 번째는 기존 제품에 AI 에이전트를 추가하거나, 새로운 에이전트 중심 제품을 출시하는 방식이다. 두 번째는 고객이 직접 AI 에이전트를 만들고 관리할 수 있도록, 기존 서비스 위에 에이전트 개발 프레임워크를 구축하는 방식이다. 각각의 방식을 실행하는 SaaS 기업은 그림 1과 같다.

그림 1
주요 SaaS 기업의 에이전틱 AI 전략

신규 AI 에이전트 제품 출시 ⁶	기존 서비스에 에이전트 개발 프레임워크 구축 ⁷
<ul style="list-style-type: none"> 세일즈포스(Salesforce) <ul style="list-style-type: none"> - 에이전트포스(Agentforce) 	<ul style="list-style-type: none"> 구글 클라우드(Google Cloud) <ul style="list-style-type: none"> - 에이전트 개발 키트(ADK, Agent Development Kit)
<ul style="list-style-type: none"> SAP - 줄(Joule) 	<ul style="list-style-type: none"> 오라클(Oracle) - AI 에이전트 스튜디오(AI Agent Studio)
<ul style="list-style-type: none"> 서비스나우(ServiceNow) <ul style="list-style-type: none"> - 나우 어시스턴트(Now Assist) 	<ul style="list-style-type: none"> SAP - 비즈니스 AI(Business AI)
<ul style="list-style-type: none"> 워크데이(Workday) <ul style="list-style-type: none"> - 일루미네이트(Illuminate) 	<ul style="list-style-type: none"> 워크데이 - 빌드(Build)
	<ul style="list-style-type: none"> 어도비(Adobe) - 익스피리언스 플랫폼 에이전트 오키스트레이터(Experience Platform Agent Orchestrator)

출처: Deloitte analysis.

한편, 처음부터 AI 기술을 기반으로 성장한 스타트업들이 기존 시장의 잠재적 파괴자로 부상하고 있다. SaaS는 단기적으로는 고객 서비스처럼 비교적 단순한 영역에서 먼저 확산될 가능성이 크지만, 전사적 자원관리(ERP)나 고객관계관리(CRM)처럼 복잡한 영역으로도 확산될 수 있다. 이러한 스타트업에는 이미 대규모 투자가 몰리고 있으며,⁸ 향후 몇 년 내 기존 SaaS 대기업에 인수될 가능성도 크다. 실제로 시장조사기관 가트너(Gartner®)는 “2030년까지 포인트 솔루션(point solution)* 형태의 SaaS 툴 중 35%가 AI 에이전트로 대체되거나 대형 SaaS 기업의 에이전트 생태계 안에 흡수될 것”이라고 전망했다.⁹

현재 기업은 기존 SaaS 벤더가 제공하는 기능을 통해 별도 개발 없이도 에이전트 실험과 학습을 손쉽게 시작할 수 있다. 따라서 대다수 기업이 초기에는 이처럼 기본적으로 제공되는 에이전트 AI 기능을 활용하겠지만, 시간이 지나 경험이 축적되면 보다 전략적이고 주도적인 방식으로 전환할 것으로 전망된다. 이러한 단계에 이르면 기업들이 향후 자사의 데이터 자산을 중심 기동으로 삼아, 방대한 에이전트 생태계에서 필요한 기능만을 선택해 자체 에이전트를 개발하며, 이를 자율적으로 작동하는 통합형 멀티에이전트 시스템으로 엮어낼 것으로 전망된다. 즉, 에이전트 AI는 단순한 기능 추가가 아니라 SaaS 아키텍처, 벤더 시장, 기업 운영 모델 전체를 뒤흔드는 구조적 전환의 시작점이다.

* **포인트 솔루션(point solution)**은 특정 문제 하나를 해결하기 위해 설계된 범위가 제한적인 단일 기능 중심의 솔루션을 의미한다. 전체 시스템이 아니라 특정 업무 및 과제에만 초점을 맞추며, 한두 가지 기능은 매우 뛰어나지만 적용 범위가 좁다는 특징이 있다.

에이전트 AI 전환기, 기업이 넘어야 할 과제

에이전트 AI 기반 SaaS로의 전환은 중대한 기회를 창출할 수 있지만, 여러 난제를 해결해야만 성공적 결과를 기대할 수 있다

1. 복잡해지는 과금 모델

AI 에이전트의 확산으로 SaaS의 과금 및 결제 구조가 가장 복잡한 변화를 보일 것으로 예상된다. 소프트웨어가 주로 온프레미스 환경에서 운영되던 과거에는 영구 라이선스를 구매한 뒤, 업그레이드와 유지보수 비용을 별도로 지불하는 방식이 일반적이었다. 그러나 클라우드를 기반으로 한 SaaS 혁명이 본격화되면서 과금 체계는 구독형 모델로 빠르게 전환됐다. 현재 SaaS 시장에서 가장 보편적인 과금 방식은 사용자 수(seat) 수를 기준으로 요금을 책정하는 방식이다. 이용자 유형에 따라 기능이 달라지는 티어형 요금제가 적용되기도 하는데, 이는 비교적 단순하고 예측 가능한 가격 구조라는 평가를 받는다. 반면 최근 들어 사용량 기반 또는 소비량 기반 과금 방식도 빠르게 확산되고 있다. 이 방식은 API 호출 횟수나 AI 모델이 처리한 토큰(token, 텍스트 또는 데이터 단위) 사용량을 기준으로 요금이 부과된다. 따라서 비용 변동을 예측하기가 쉽지 않다.

그러나 AI 에이전트가 본격적으로 활용되기 시작하면 이러한 기존의 모델은 더 이상 실제 교환되는 가치를 충분히 반영하지 못하게 될 것이다.¹⁰ 에이전트 하나가 여러 사람의 역할을 대신할 수 있어 사용자 수가 줄고, 에이전트의 작동 방식이 예측 불가하여 사용량 추정도 어려워지

기 때문이다. 따라서 2026년 이후 SaaS 시장에서는 다양한 과금 모델이 활발히 시험 운영되고, 결국 수년 후에야 표준이 자리 잡을 것으로 예상된다. 가트너는 2030년까지 기업용 SaaS의 40% 이상이 사용량, 에이전트, 성과 기반 과금 모델로 전환될 것으로 전망한다.¹¹

1) 사용량 기반 요금제

사용량 기반 요금제는 AI 에이전트가 행동할 때마다 비용이 발생하는 방식이다. 에이전트가 수행한 업무, 컴퓨팅 시간, API 호출 수, 생성형 AI 가 처리한 토큰 양, 디지털 노동자 개념의 에이전트 1명당 정액제 등을 과금 기준으로 삼을 수 있다. 실제로 맥시오(Maxio)사의 조사에 따르면, AI 네이티브 SaaS 기업의 83%가 이미 사용량 기반 요금제를 채택한 것으로 나타났다.¹² 사용량 기반 요금제는 측정 가능하고 감사가 용이하다는 장점이 있다.

2) 성과 및 가치 기반 요금제

성과 및 가치 기반 과금은 AI 에이전트가 탑재된 SaaS 애플리케이션이 실제로 창출한 비즈니스 성과를 기준으로 요금을 부과하는 방식으로, 성과 기여도를 명확히 산정하기 어렵다는 한계가 있다. 예를 들어 고객 상담 티켓이 몇 건 해결됐는지, HR 에이전트의 도움을 받아 최종 채용된 인원이 몇 명인지와 같은 단순한 성과 지표부터, AI 에이전트가 전체 매출 증가에 얼마나 기여했는지를 따지는 복합적 지표까지 적용 범위가 매우 넓다. 다만 이 같은 모델이 시장 전반에 보편화되기까지는 상당한 시간이 걸릴 것으로 보이며, 현재는 일부 기업만이 선도적으로 이를 실험

하는 단계에 머물러 있다.¹³ 무엇보다 에이전트 시스템이 일관되고 신뢰할 수 있는 가치를 지속적으로 창출할 수 있다는 점을 입증해야 하는 과제가 남아 있다.

이러한 과금 모델의 변화는 기업 내 여러 기능 조직에 동시에 영향을 미치며, SaaS 공급업체의 사업 운영 방식 자체를 바꿀 가능성도 크다. 먼저 ‘에이전트’, ‘작업’(task), ‘프로세스’, ‘상호작용’(interaction), ‘성과’(outcome)와 같은 기본 개념에 대한 공통된 정의를 먼저 정립할 필요가 있다. 또한 ‘가치’가 무엇이고 이러한 가치가 어떻게 귀속되는지에 대해서도 명확한 계약적 합의가 필수다. 이는 엔지니어, 영업, 법무 등 여러 부서의 긴밀한 협업을 요구하는 고난도 작업이다. 특히 여러 벤더의 다양한 에이전트가 결합된 멀티 에이전트 환경에서는 특정 AI가 성과를 창출했음을 입증하는 과정이 더욱 복잡해질 수 있다. 이로 인해 공급업체의 매출과 고객의 비용 모두 변동성이 크고 예측하기 어려워질 수 있다. 이에 따라 시스템 계측과 사용량 측정 기술은 더욱 고도화되고, 데이터 가시성, 과금, 재무 컴플라이언스 역시 실시간 자율화 방향으로 진화할 필요성이 커지고 있다.

이와 함께 다수의 SaaS 공급업체는 영업 모델 자체의 구조적 전환을 피하기 어려울 것으로 전망된다. 영업 인력은 새로운 과금 방식을 고객에게 충분히 설명함과 동시에, AI 에이전트가 실제로 가치를 창출하며, 기존 구독형 모델보다 비용 부담이 커지지 않는다는 점을 설득해야 한다. 이에 따라 영업 성과의 측정 방식과 보상 체계 역시 기존과는 다른 구조

로 재설계될 가능성이 크다. 따라서 단기 계약 중심의 판매 방식에서 벗어나, 고객과의 장기적, 심층적 관계 형성이 SaaS 영업 전략의 핵심 요소로 부상할 것이다.

고객 경험과 사용자 인터페이스, AI 에이전트 시대의 핵심 차별화 요소로 부상

본질적으로 AI 에이전트는 자율적으로 작동하도록 설계됐다. 그럼에도 사용자 인터페이스(UI)의 필요성은 오히려 커지고 있다. API와 마찬가지로 에이전트 역시 헤드리스 구조(headless architecture)*로, UI와 직접 연결되지 않은 채 작동한다. 그러나 사용자와의 상호작용과 가시성 확보를 위한 접점 공간은 필수적이다. 이에 따라 향후 AI 에이전트의 인터페이스를 단일 창구로 운영할지, 아니면 다중 인터페이스 체계로 발전시킬 것인지가 주요 쟁점으로 떠오르고 있다. 또한 특정 SaaS 사업자나 제3자가 에이전트로 향하는 게이트웨이를 통제하게 될지도 중요한 변수다.

* 헤드리스 구조(headless architecture)란 프런트엔드(화면·경험)와 백엔드(데이터·로직)를 분리해, API로 연결하는 시스템 설계 방식을 의미한다. 변경 속도가 빠르고, 채널 확장이 유연하며, 개인화가 용이하다는 특징이 있다.

딜로이트는 향후 수년간 SaaS 기반 AI 에이전트의 사용자 경험과 인터페이스가 개인화, 대화형, 진단형으로 빠르게 진화할 것으로 전망한다.

- **개인화 및 선제형 인터페이스**는 사용자 개개인의 역할과 과거 행동을 바탕으로 필요한 툴과 업무를 자동으로 제시하고, 맞춤형 인사이트와 실행 가능한 행동을 추천하는 방향으로 발전이 예상된다.
- **대화형 인터페이스**는 메뉴 클릭 중심의 기존 방식에서 벗어나 자연어와 음성 명령 기반의 상호작용으로 전환된다. AI 에이전트가 자연어를 정형화된 API 호출로 변환해 사전 정의된 워크플로 없이도 업무를 수행하는 환경을 구현하게 된다. 이는 소프트웨어에 일일이 지시하는 방식이 아니라, 대화를 통해 결과를 요청하는 방식으로 패러다임이 전환된다는 의미다.
- **진단형 인터페이스** 역시 핵심 요소로 부상한다. 자율적으로 작동하는 에이전트 특성상 오류나 예기치 못한 상황이 발생했을 때 사용자가 의사결정 과정을 재구성하고 원인을 이해할 수 있어야 한다. 이를 위해 투명성, 설명 가능성, 되돌림 가능성, 감사 추적성이 신뢰 확보의 필수 요건이 된다.

상호작용 계층(layer)이 어디에 위치하게 되는지도 또 하나의 핵심 쟁점이다. 딜로이트는 상당수 상호작용이 독립형 SaaS 애플리케이션 내부에

서 이뤄질 가능성이 크다고 보고 있다. 다수의 SaaS 사업자는 사용자의 업무 효율과 자사 제품에 대한 종속성을 유지하기 위해 사용자를 최대한 자사 애플리케이션 안에 머무르게 하려는 전략을 구사하고 있다. 이에 따라 자사 에이전트는 물론 타사 에이전트까지 연동 제공하는 방향으로 진화할 가능성이 크다.

동시에 별도의 에이전트 관리 플랫폼을 통한 상호작용도 확대될 전망이다. SaaS 사업자 또는 제3자 기업이 제공하는 이러한 플랫폼은 다수의 벤더 및 내부 개발 에이전트의 활동을 통합 관리하면서 사용량, 비용, 접근 권한, 성능, 상태, 보안, 캠플라이언스를 한 곳에서 모니터링하는 컨트롤 타워 역할을 수행할 수 있다.¹⁴ 더 나아가 내·외부 에이전트가 자유롭게 등록되고 기업이 필요한 기능을 자율적으로 탐색 및 연결하는 에이전트 마켓플레이스도 본격적으로 활성화될 전망이다.¹⁵ 이처럼 시장 참여자들의 관심이 집중되는 상호작용 계층은 막대한 부가가치를 창출할 잠재력을 지니고 있으며, 향후 이 영역을 둘러싼 경쟁 역시 극도로 치열해질 것으로 전망된다.

AI 에이전트 확산 시대, SaaS 이용 기업과 벤더의 생존 전략

2026년을 기점으로 SaaS 애플리케이션을 통한 AI 에이전트 활용이 급격히 확산될 전망이다. 다수의 글로벌 SaaS 기업들이 보다 고도화된 에이전트 AI 솔루션을 고객사에 제공하기 위해 박차를 가하고 있기 때문에, AI 기반 자동화에 대한 투자도 SaaS 영역까지 빠르게 확대될 것으로 예상된다. 기업은 업무 프로세스 효율화와 비용 절감, 운영 유연성 제고, 근로자 개인 맞춤형 역량 강화를 핵심 목표로 AI 에이전트 도입을 가속화할 가능성이 크다. 이 과정에서 다양한 실험이 이뤄지고 과금 모델도 빠르게 다변화될 전망이다. 초기에는 산업 및 업무 유형별로 서로 다른 가격 구조가 혼재하는 양상이 나타나겠지만, 중장기적으로는 자율적으로 작동하는 통합 멀티 에이전트 시스템이 주도하는 방식으로 점진적 전환이 이뤄질 것이다. AI 에이전트는 기능적 틀에 그치지 않고, 기업 운영 전반을 유기적으로 연결하는 핵심 실행 주체로 자리매김하게 될 것으로 보인다.

1. SaaS를 도입할 기업이 고려해야 할 사안

1) 데이터 관리에 투자하라

기업들은 AI 에이전트를 본격적으로 활용하기에 앞서 데이터 관리에 대한 투자를 최우선 과제로 삼아야 한다. AI 에이전트 환경에서는 데이터 접근성, 시스템간 통합, 관측 가능성, 데이터 거버넌스의 중요성이 기존 보다 훨씬 더 커지기 때문이다. 데이터가 반드시 하나의 중앙 저장소에 집중될 필요는 없지만, 조직 전반에서 일관되게 활용 가능하고 접근 가능한 상태로 유지되는 것이 핵심이다.

2) 증가하는 기술 및 시장 복잡성에 대비하라

기업은 급속도로 심화되는 기술 복잡성에 대비해 전략을 마련해야 한다. 앞으로 더 많은 AI 모델, 더 많은 에이전트, 더 다양한 벤더와 새로운 생태계, 더욱 복잡한 데이터 연계 구조가 동시에 등장할 것이다. 이에 따라 서로 다른 벤더의 에이전트를 하나의 업무 흐름 속에서 연동 및 운영해야 하는 상황이 일반화되면서, 과금 구조와 운영 방식 역시 한층 복잡해질 것이다.

3) 다층적 가격 모델에 대비하라

AI 에이전트의 과금 체계 역시 다면적 구조로 진화할 가능성이 크다. 라이선스 기반 요금에 사용량, 가치, 성과 기반 과금이 혼합된 하이브리드 모델이 표준으로 자리 잡을 경우, 비용을 예측하기가 더욱 어려워질 수밖에 없다. 이에 대비해 기업은 실시간 재무 관리 역량을 대폭 강화해야 한다.

4) 직원을 ‘AI 오케스트레이터’로 전환시켜라

근로자는 점차 AI 에이전트를 단순한 도구가 아닌 ‘함께 일하는 디지털 동료’처럼 관리하는 역할로 진화할 것이다. 목표를 설정하고, 업무 수행을 감독하며, 결과를 검증 및 보정하는 데 더 많은 업무 시간을 써야 한다는 의미다. 이에 따라 업무 프로세스를 재설계할 때는 인간이 담당할 영역, 에이전트가 수행할 영역, 인간과 에이전트가 협업할 영역을 명확히 구분하는 작업이 필수다. 이는 단순한 소프트웨어 업그레이드를 넘어 조직 문화의 근본적 전환이 필요한 작업이다.

2. SaaS 벤더가 고려해야 할 사안

1) 치열해지는 경쟁에 대비하라

생성형 AI 도구의 확산으로 누구나 손쉽게 코드를 작성할 수 있는 환경이 마련됐다. 이에 따라 비용을 전혀 들이지 않고도 소프트웨어 개발이 가능해져, AI 네이티브 기업은 물론 고객사가 직접 솔루션을 개발하기도 한다. 이처럼 경쟁이 격화된 시장 환경에서 SaaS 기업들은 기능 중심의 경쟁에서 벗어나, 차별화된 제품 가치와 서비스 경쟁력을 확보해야만 생존할 수 있다.

2) 상호운용성(interoperability)에 주력하라

AI 에이전트는 여러 시스템을 넘나들며 작업을 조율하고 데이터와 목표를 공유해야 하는 만큼, 상호운용성이 핵심 경쟁 요소로 부상하고 있다. 보안과 컴플라이언스를 유지하면서도 다양한 플랫폼과 유연하게 연동되는 개방형 환경에 대한 준비가 필수적이라는 의미다. 이는 솔루션에 만족하지 못하는 고객이 언제든 다른 공급자로 이동할 수 있는 시장 구조가 본격화된다는 점을 시사한다.

3) 영업 모델을 전면 재정비하라

과금 모델이 다변화되면서 영업 모델도 전반적인 구조 전환이 필요해졌다. 요금 체계가 복잡해질수록 매출 예측이 불확실해질 수밖에 없으며, 영업 조직의 성과 측정 방식과 보상 체계 역시 기존과는 다른 기준이 적용될 수 있다. SaaS 기업은 고객이 비용을 보다 정확히 예측할 수 있도록 단순하고 유연한 하이브리드 과금 모델을 제시하는 한편, 과금 정책에 대한 대화를 통해 고객의 잠재적 니즈를 발굴하고 관계의 깊이를 강화하는 전략적 영업 활동에 주력해야 한다.

3장

기하급수적 가치 창출을 위한 AI 에이전트 오케스트레이션

자율형 AI 에이전트 확산으로 기업 환경은 단일 에이전트 활용을 넘어, 멀티 에이전트를 연결·조율하는 오케스트레이션 역량이 성과를 좌우하는 단계로 진입하고 있다. 이에 따라 기업은 기존 업무에 AI를 덧붙이는 수준을 넘어, 인간-AI 협업 구조와 조직 역할을 재정의하고, 에이전트 중심의 프로세스로 전환해야 한다. 이 과정에서 3계층 기술 아키텍처, 통신 표준, 통제 인프라를 포함한 기술 기반과 함께, 소유권·책임·성과 측정이 명확한 거버넌스 설계가 AI 오케스트레이션 경쟁력의 핵심으로 부상하고 있다.

핵심 내용 요약 (Executive Summary)

» 왜 오케스트레이션이 핵심이 되는가

- 자율형 AI 에이전트가 기업 내 다양한 역할을 수행하며 멀티 에이전트 환경이 빠르게 확산
- 개별 에이전트 성능보다 여러 에이전트를 연결·조율하는 오케스트레이션 역량이 성과를 좌우
- AI 에이전트 오케스트레이션 시장은 2026년 85억 달러 → 2030년 350~450억 달러로 급성장 전망

» (도입 전략과 조직의 변화) 단일 목적 에이전트에서 멀티 에이전트 시스템으로의 전환이 필수

- 기존 업무에 1)AI 에이전트를 덧붙이는 방식, 2)처음부터 AI 에이전트 전제의 신규 설계, 3)프로세스 재설계 등 다양한 접근법 존재
- 직원 역할은 AI 사용자 → 에이전트 설계·조율·감독자(오케스트레이터)로 진화

» (기술 인프라와 통제 메커니즘) 멀티 에이전트 확장을 위해 3계층 아키텍처가 필수

- ① 컨텍스트 계층: 지식, 데이터 → ② 온톨로지 에이전트 계층: 모듈화, 보안, 관측 가능성 → ③ 경험 계층: UI, 설명 가능성, 복구·되돌림
- 관리 플랫폼, 텔레메트리, 가드레일, 가디언 에이전트 등과 같은 통제 인프라가 전제 조건

» (비즈니스 프로세스와 거버넌스 재설계) 업무의 모듈화와 에이전트 중심 프로세스 재편 필요

- 에이전트의 의사결정에 대한 소유권·책임 소재 명확화가 핵심 이슈
- 지속적 개선을 위한 설계 전략, 거버넌스·성과 측정 체계 구축이 경쟁력의 핵심

지능형 자동화를 위해서는 각기 다른 자율형 AI 에이전트 상호간 연결과 조율이 핵심이다. 에이전트간 협업을 가능하게 하는 오픈소스와 독점(proprietary) 통신규격(protocol)은 차세대 AI 자동화 생태계 내 주도권 경쟁을 좌우할 핵심 요소다.

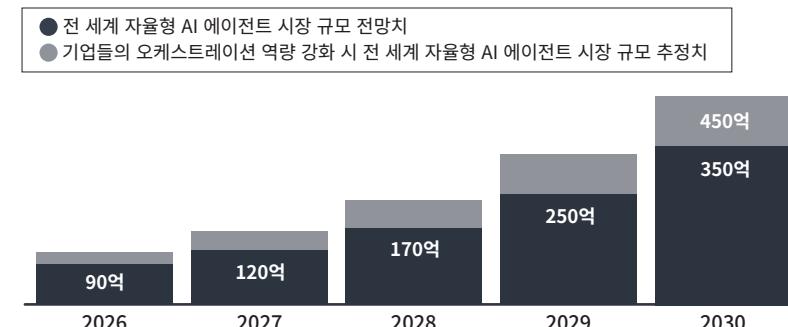
다양한 AI 추론 엔진이 영역을 넘나들며 유기적으로 상호작용하는 멀티 에이전트 시스템이 확산하고 있다. 이에 따라 역할별 에이전트를 효과적으로 조율하는 에이전트 오케스트레이션(agent orchestration)의 중요성이 한층 부각되고 있다. 정교한 오케스트레이션은 멀티 에이전트 시스템이 사용자 요청을 해석하고, 업무 흐름을 설계하며, 작업을 위임 및 조정하고, 결과를 지속적으로 검증 및 개선하는 지능형 워크플로를 구현하기 위해 가장 중요하다.¹ 반대로 에이전트 조율 체계가 미흡할 경우, 멀티 에이전트가 지닌 잠재적 비즈니스 가치를 충분히 실현하기 어렵다.

시장 전망에 따르면 자율형 AI 에이전트 시장은 2026년 85억 달러까지 성장한 후 2030년에는 350억 달러 규모로 더욱 급격히 성장할 것으로 추산된다(그림 1).² 딜로이트는 기업들이 에이전트 오케스트레이션 역량을 정교화, 고도화하고, 이에 수반되는 과제와 리스크를 선제적으로 관리할 경우 시장 규모가 2030년까지 15~30% 추가 확대돼 최대 450억 달러까지 성장할 수 있다고 내다보고 있다. 그러나 한편에서는 현재 진행 중인 에이전트 AI 프로젝트의 40% 이상이 예상치 못한 비용 부담,

확장 복잡성, 예기치 않은 리스크 등으로 인해 2027년 이전에 중단될 가능성이 제기된다.³ 하지만 이러한 프로젝트는 잠재적 위험 요소를 사전에 해소할 경우 기업의 매출 성장을 크게 견인할 수 있는 기회 요인이 될 수 있다. 따라서 선제 대응이 무엇보다 중요하다.

**그림 1
기업들의 에이전트 오케스트레이션 역량 강화되면서 AI 에이전트 시장 확대 전망**

(단위: 미달러)



참조: 모든 수치는 반올림 수치임.

출처: Deloitte analysis.

멀티 에이전트 시스템의 잠재력을 온전히 실현하기 위해 기업은 에이전트가 어느 수준까지 자율적으로 작동할 것인지에 대한 명확한 기준을 설정해야 한다. 상당수 기업이 가까운 시점에 이러한 기준을 본격적으로 수립하고 조율하기 시작할 것으로 전망된다. 동시에 에이전트간 상호운용성과 관리 역량을 핵심 전략 요소로 삼고, 이에 맞춰 업무 프로세스와 인재 운영 체계를 실질적으로 개편해야 멀티 에이전트 시대의 주도권을 확보할 수 있다.

멀티 에이전트 시스템을 기업 환경에 안착시키기 위한 지침

기업들이 에이전트 오케스트레이션을 강화하기 위한 본격적 의사결정을 추진하는 과정에서 다음의 세 가지 핵심 지침을 따르는 것이 바람직하다.

1. 단일 목적 에이전트에서 멀티 에이전트 시스템으로 전환

현재 기업은 특정 업무에 특화된 단일 목적 AI 에이전트를 활용해 여러 단계를 자율적으로 수행할 수 있는 수준에 이르렀다.⁴ 그러나 최근 들어 에이전틱 AI는 단일 에이전트를 넘어 멀티 에이전트 시스템을 통해 훨씬 더 광범위하고 폭발적인 기업 가치를 창출할 수 있다는 인식이 확산되고 있다.⁵ 그럼에도 불구하고, 다수의 기업에서 멀티 에이전트 기술을 실제 활용하는 수준은 여전히 미흡하다.

딜로이트가 미국 내 산업 전반의 리더 약 550명을 대상으로 실시한 ‘2025 테크 벤류 서베이’(2025 Tech Value Survey)에 따르면, 자사 조직의 기본적인 자동화 역량이 성숙 단계에 도달했다고 평가한 응답자는 80%에 달했다. 하지만 기본 자동화와 더불어 AI 에이전트 활용 역량까지 성숙 단계에 도달했다고 응답한 비율은 28%에 그쳤다. 또한 각 전략을 추진 중인 기업 가운데 기본 자동화만을 도입한 경우 45%가 3년 이내 플러스(+) 투자수익률(ROI) 달성을 기대한 반면, 자동화와 에이전트를 병행하는 경우 동일한 기간 내 플러스 투자수익률을 기대한 비율은 12%에 불과했다.⁶ 이는 에이전틱 AI가 가치 창출의 잠재력은 크지만 실제 기업 현장에서 여전히 적용하기가 어렵고 결과의 불확실성도 크다는 점을 단적으로 보여준다.

그렇다면 이러한 격차를 보다 빠르게 줄이려면 어떻게 해야 할까? 이를 위한 첫 단계로 멀티 에이전트 도입을 위한 세 가지 접근 방식에 대한 전략적 검토가 필요하다(그림 2).⁷

그림 2

업무 복잡성, 기반 워크플로, 적용 기술을 반영한 에이전틱 AI 전략 사례

지능형 가상계층 (overlay, 기존 구조에 추가 기능을 겹쳐 적용)
명확히 정립된 기존 워크플로에 AI 에이전트를 겹쳐 적용하면 빠르게 실험적 도입을 실행할 수 있다.
<ul style="list-style-type: none"> · 장점: 기존 레거시 시스템을 거의 방해하지 않고 적용할 수 있다. · 단점: 통합, 비용 관리, 데이터 보안 등과 관련해 까다로운 해결과제가 발생할 수 있다.
처음부터 에이전틱 AI를 설계
모듈형 마이크로서비스 아키텍처를 기반으로 프로세스를 재설계함으로써, 특정 워크플로에 최적화된 맞춤형 AI 에이전트를 적용할 수 있다. 그 과정에서 에이전트 기능이 내재된 서비스형 소프트웨어(SaaS) 솔루션이나 특정 기능에 특화된 마켓플레이스 기반 에이전트를 활용해 구현 속도를 높이고 복잡성을 낮출 수 있다.
<ul style="list-style-type: none"> · 장점: 실행 장벽을 낮출 수 있다. · 단점: 혁신, 보안, 규제 준수가 솔루션 제공업체의 기술력과 운영 역량에 좌우되는 문제가 발생할 수 있다.
프로세스 재설계
자동화 난이도가 높거나 리스크가 큰 최우선순위 업무의 경우, 기존 프로세스를 재설계하고 AI 에이전트를 전략적으로 도입함으로써 업무 자동화와 운영 혁신을 동시에 달성할 수 있다.
<ul style="list-style-type: none"> · 장점: 새로운 혁신적 활용사례를 만들 수 있다. · 단점: 세심한 사전 계획과 단계적 실행이 필요하다.

출처: Deloitte analysis.

2. 에이전트 오케스트레이션의 인간 계층(layer) 정립

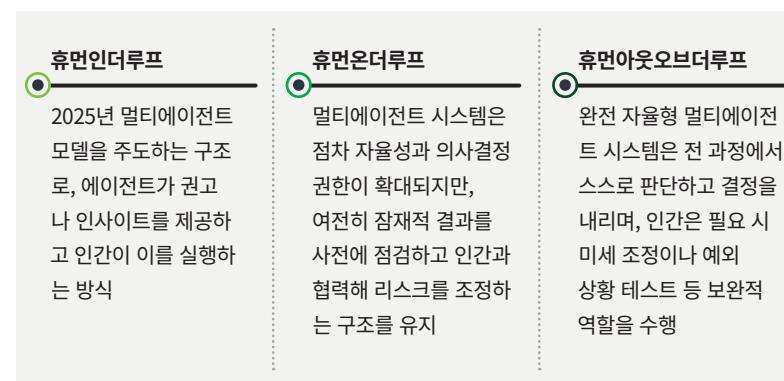
2025년 현재 기업들은 금융 투자 리서치나 중증 질환 의료 분야 등 특정 영역에서 비교적 단순하지만 실효성이 검증된 형태의 에이전트 오케스트레이션을 우선 도입하고 있다.⁸ 이들 사례에서 AI 에이전트는 인간의 감독하에, 또는 별도의 ‘감독 에이전트’(supervisor agent)의 통제 아래 협업하며, 인간 전문가가 최종 의사결정을 내릴 수 있도록 인사이트를 제공하는 역할을 수행하고 있다. 반면 여러 사업 영역을 가로지르는 고난도·고자율 에이전트 오케스트레이션은 아직 일부 선도 기업들만 실행하고 있다.⁹ 그러나 이러한 시도가 본격화될수록 기업은 에이전트의 자율성과 인간의 통제 사이에서 혁신, 리스크, 책임, 신뢰의 균형을 더욱 정교하게 조율해야 한다.

연구 결과에 따르면, 현재 멀티 에이전트 시스템은 인간이 직접 개입하는 휴먼인더루프(human in the loop) 구조에서 더 우수한 성과를 낸다. 이는 인간의 경험과 직관이 에이전트의 판단에 보완적으로 작용하며, 조직 특유의 미묘한 기대치와 의사결정 기준에 결을 맞추는 데 중요한 역할을 하기 때문이다.¹⁰ 이에 따라 향후 12~18개월 내 휴먼인더루프 구조를 유지한 채 복잡한 에이전트 오케스트레이션의 실험적 도입을 본격 가속화하는 기업이 증가할 것으로 전망된다. 이 과정에서 기업은 의사결정의 신뢰도, 품질, 책임성을 높이기 위해, 인간의 판단을 에이전트 워크플로에 체계적으로 통합하는 프레임워크와 솔루션을 적극 도입할 가능성이 크다.¹¹

아울러 업무 복잡성, 산업 특성, 워크플로 설계, 결과의 중요도에 따라 인간이 직접 개입하는 휴먼인더루프(human in the loop) 단계부터 인간이 감독만 하는 휴먼온더루프(human on the loop) 단계, 인간이 전혀 개입하지 않는 휴먼아웃오브더루프(human out of the loop) 단계까지 AI 에이전트의 자율성 스펙트럼이 점진적으로 정립될 것으로 예상된다(그림 3). 다만 인간 개입이 배제되는 방식 역시 지속적 모니터링이 반드시 필요하다. 인간이 개입 또는 감독하는 방식에서는 성과 추적, 오케스트레이션 시각화, 에이전트 활동 내역을 제공하는 플랫폼과 텔레메트리(telemetry)* 대시보드가 인간의 판단을 지원하는 핵심 인프라로 자리 잡게 될 전망이다. 딜로이트는 2026년을 전후해 가장 선도적인 기업들이 휴먼온더루프 중심의 오케스트레이션 체계로 전환하기 위한 본격적인 기반 구축에 착수할 것으로 내다보고 있다.

* 텔레메트리(telemetry)는 시스템·장비·소프트웨어의 상태와 동작 데이터를 실시간 또는 지속적으로 수집·전송·분석하는 기술과 데이터 체계를 의미한다. 사람이 직접 개입하지 않아도 자동으로 데이터를 수집하며, 지연 없이 실시간으로 상태·이상·성능을 파악하기 때문에, 관측 가능성(observability)의 핵심 구성요소로 작용한다.

그림 3
AI 에이전트 자율성의 점진적 발전 단계



출처: Deloitte analysis.

3. 산발적 AI 에이전트 확산의 통제

산발적 AI 에이전트 확산을 통제하는 것은 2026년을 기점으로 더욱 중대한 해결 과제로 부상할 전망이다. AI 에이전트는 매우 다양한 프로그래밍 언어, 프레임워크, 인프라, 통신규격을 기반으로 급속히 확산될 가능성이 크기 때문이다. 일부 에이전트는 텍스트·음성·이미지 등 다양한 정보 유형과 형식을 해석하는 멀티모달 역량과 함께 최고 수준의 지능을 구현해야 하는 복합적 과제에 직면하게 된다. 따라서 기업 내부를 넘어 디지털 인터페이스 상 에이전트간 협력 방식을 정의하는 새로운 표준을 수립하는 것이 시급한 과제로 떠올랐다. 현재로서는 미국 매사추세츠공과대학(MIT)의 ‘난다’(NANDA) 프로젝트와 같은 웹 기반 에이전트 통신규격 대표적 표준 사례로 꼽힌다.¹² 장기적으로는 이러한 기술

덕분에 기업 내·외부 네트워크를 아우르는 전략적 에이전트 오케스트레이션이 가능해져, 완전히 새로운 비즈니스 역량을 창출하는 기반이 될 것으로 기대된다.

이 같은 기술 환경의 변화 때문에 멀티 에이전트 상호운용성이 필수 요소이자 동시에 가장 어려운 과제로 부상하고 있다. 기업은 서로 다른 AI 에이전트를 하나의 통합 플랫폼에서 지시, 관측, 관리할 수 있는 방안을 점점 더 적극적으로 모색하게 될 가능성성이 크다. 그러나 디지털 워크포스 운영에 대한 표준이 부재할 경우, AI 에이전트의 개발, 구성, 배포는 적절히 조율되지 못한 채 분산된 형태로 진행되기 쉽다. 이는 결국 성능 저하, 윤리 리스크, 사이버 보안 위협, 규제 컴플라이언스 문제 등 다양한 잠재적 위험과 비용 증가로 이어질 수 있다.

현재의 IT 및 비즈니스 아키텍처를 형성한 클라우드와 마이크로서비스 등 과거 기술 진화 과정을 되짚어 보면 이러한 과제를 해결하는데 도움이 되는 중요한 시사점을 얻을 수 있다. 과거 인터넷 보안 프로토콜 HTTPS(HyperText Transfer Protocol Secure), 데이터 교환 형식 JSON(JavaScript Object Notation) 등 표준화된 통신규격, 명확한 서비스 연동규칙(API) 설계도, 도메인 특화 마이크로서비스는 상호운용성과 안정성, 시스템 소유권을 동시에 확보하는 기반이 됐다. 또한 서비스 레지스트리, 분산 추적, 중앙 로그 시스템은 기능 탐색, 오류 해결, 서비스 관리를 획기적으로 개선했다. 여기에 거버넌스 체계, 서비스 카탈로그, ‘제로 트러스트’(zero trust)* 보안 모델은 시스템 안정성을 확보하고 버

전 혼선을 방지하는 핵심 장치로 작용했다. 이러한 요소들은 회복탄력적이면서도 확장 가능한 멀티 에이전트 시스템을 구축하는데 중요한 교훈을 제공할 수 있다.

* 제로 트러스트(zero trust)는 내·외부를 불문하고 그 어떤 사용자·기기·시스템도 기본적으로 신뢰하지 않고, 매 접근마다 검증하는 보안 철학이자 아키텍처를 의미한다. △ 네트워크 내부라 하더라도 자동 신뢰하지 않고(never trust) △ 접속 시점뿐 아니라 사용 중에도 신원·상태·행위를 반복 검증하며(always verify) △ 필요한 만큼 필요한 시간 동안만 접근을 허용(least privilege)하는 등 3가지 핵심 원칙을 기반으로 한다.

다만 전문가들은 과거의 성공 공식을 답습하는 데 그쳐서는 안 된다고 지적한다. 멀티 에이전트 시대에는 기존 IT 아키텍처 위에 에이전트 전용의 새로운 계층을 추가하고, 이를 중심으로 업무 프로세스와 데이터, 보안, 거버넌스를 재설계하는 근본적 전환이 필요하다. 이는 AI 에이전트가 단순한 도구를 넘어 기업 운영 구조 자체를 재편하는 핵심 인프라로 진화하고 있음을 보여주는 신호로 해석된다.

회복탄력적이고 확장 가능한 멀티 에이전트 시스템을 위한 엔터프라이즈 아키텍처

그림 4



컨텍스트 계층

확장 가능한 AI 에이전트 아키텍처 구현에 핵심인 지식 엔지니어링 기반 계층이다. 컨텍스트 계층은 원천 데이터와 같이 다양한 비정형 데이터를 지식 그래프, 온톨로지, 도메인 분류 체계 등 거버넌스가 적용된 구조화된 지식 표현 형태로 변환해, 에이전트가 효과적으로 이해할 수 있는 ‘소규모 세계’(small world) 모델을 제공한다. 최적화된 컨텍스트 검색 기술은 에이전트가 필요한 정보를 정확하고 적시에 확보할 수 있도록 지원한다. 컨텍스트 정제(context shaping)는 불필요한 노이즈와 쓸데없는 요소를 제거해 입력값의 품질을 높임으로써 에이전트의 정확도와 처리 효율을 동시에 향상시키는 역할을 수행한다.

에이전트 계층

컨텍스트 계층을 기반으로 실제 에이전트를 작동하는 핵심 계층으로, 안전성, 자율성, 상호운용성 등 중심 가치를 기반으로 설계된다. 에이전트 계층의 중심에는 새로운 기술을 유연하게 통합 및 확장할 수 있는 모듈형·조합형 아키텍처가 자리한다. 적합한 도구를 활용해 에이전트의 과도한 부하를 방지하며, 정교한 메모리를 활용해 사실·경험·절차 기반 기억을 균형 있게 적용함으로써 컨텍스트 인식 능력을 한층 고도화한다. 또한 에이전트 계층은 소형 특화 모델부터 대형 고성능 모델에 이르기까지 업무 성격에 맞는 최적의 AI 모델을 선택해 운용함으로써 오케스트레이션 성능을 극대화한다. 강력한 보안 체계와 고도화된 텔레메트리 기반의 관측 기능은 에이전트 활동의 보안성, 투명성, 신뢰성을 동시에 확보하는 기반으로 작용한다.

경험 계층

엔터프라이즈 사용자와 AI 에이전트를 연결하는 최전선 인터페이스 계층으로, 에이전트의 행동을 통제하고 방향을 교정하는 역할을 담당한다. 경험 계층은 에이전트의 상태 정보와 컨텍스트 데이터를 사용자에게 제공하고, 프롬프트 추천 기능과 이해하기 쉬운 결과 표시 방식을 통해 업무 검토와 의사결정을 지원한다. 또한 인간의 감독을 위한 직관적 제어 기능, 고도화된 피드백 체계, 에이전트의 사고 과정을 시각적으로 보여주는 설명 가능성 기능은 결과에 대한 투명성과 신뢰도를 크게 높이는 요소로 작용한다. 더 나아가 오류 발생이나 판단이 모호한 상황에서도 명확한 원인 설명과 복구 선택지를 제공함으로써 안정적 운영을 뒷받침하는 역할을 수행한다.

출처: Deloitte analysis.

멀티 에이전트 시스템을 조직에 안착시키기 위한 필수 요소

기업은 다음의 필수 요소를 고려해야 비즈니스 핵심 과제에 부합하는 기술적 기반을 구축할 수 있다.

1. 유연하고 확장 가능하며 안전한 통신규격

멀티 에이전트 환경에서는 에이전트 상호간뿐 아니라 에이전트와 외부 도구·플랫폼 간 표준화된 통신규격을 갖춰야 에이전트의 역량, 인사이트, 실행 결과에 대한 예측 가능한 메시지 교환이 가능해진다. 지난 1년 간 구글(Google)의 A2A, 시스코(Cisco)가 주도하는 AGNTCY, 앤트로픽(Anthropic)의 MCP 등 여러 에이전트간 통신규격이 잇따라 등장하며, 서로 다른 프레임워크나 모델 위에서 구축된 에이전트간 협업 경쟁에 불이 붙었다.¹³ 테크 기업들은 이 분야의 주도권을 확보하기 위해 파트너사뿐 아니라 고객사와 손잡고 생태계 구축에 속도를 내고 있다. 일부 통신규격은 금융 거래와 같은 특정 영역에서 신뢰 가능한 에이전트 상호운용성을 구현하는 방향으로 확장되고 있다.¹⁴

다만 통신규격간 과도한 경쟁은 자칫 특정 통신규격과 에이전트 생태계에 기업이 종속되는 ‘닫힌 정원’(walled garden) 구조를 초래할 수 있다.¹⁵ 그러나 업계에서는 이르면 2026년을 기점으로 다수의 규격이 2~3 개의 사실상 표준 규격으로 수렴될 가능성이 크다는 관측도 나온다. 이후 다른 테크 기업들은 경쟁력을 유지하기 위해 이들 핵심 표준에 맞춰 기술 전략을 재편할 수밖에 없을 것으로 보인다.

어떤 통신규격이 최종적으로 시장의 주도권을 확보할지는 기업의 멀티 에이전트 활용 방식, 산업 특성, 오케스트레이션 성숙도가 좌우할 전망이다. 예를 들어 표준 API와 테스트 및 시뮬레이션용 개발자 도구를 갖춘 경량 규격은 실험과 검증 단계에서 높은 효율성을 제공할 수 있다. 또한 공유 컨텍스트와 메모리를 기반으로 한 P2P(peer-to-peer)* 및 허브 앤 스포크(hub-and-spoke)** 방식의 에이전트 상호작용, 내장형 협상·위임·충돌 해결 기능 등을 다양한 오케스트레이션 장치로 활용할 수 있다. 더 나아가 에이전트 레지스트리를 통한 신뢰 기반 탐색과 작업 부하 분산, 비동기 메시징, 고처리량·저지연 통신, 체인형·중첩형 워크플로 지원은 대규모 에이전트 오케스트레이션의 확장을 뒷받침하는 핵심 요소로 작용한다. 또한 인증, 보안 메시징, 접근 통제 기능은 보안 리스크를 완화하는 역할을 하며, 에이전트간 메시지와 설명 정보는 감사 가능성과 오류 추적성을 확보하는 핵심 장치로 평가된다.

* P2P(Peer-to-Peer)는 중앙 서버를 거치지 않고, 네트워크에 참여한 각 노드(피어)가 서로 직접 연결돼 데이터·자원·서비스를 주고받는 분산형 네트워크 구조를 의미한다. 서버-클라우드 구조가 아니라 참여자간 동등한 연결, 피어간 데이터 직접 전송 및 공유, 저장·연산·전송 부담을 네트워크 전체에 분산할 수 있다는 장점이 있다.

** 허브 앤 스포크(hub-and-spoke)는 중앙의 허브(hub)가 주변에 연결된 지점, 즉 ‘스포크(spoke)’를 연결·조정·통제하는 구조로, 트래픽·업무·의사결정을 효율적으로 집약 및 분배하는 네트워크 모델을 의미한다. 중복을 제거하고 운영 일관성을 유지해 표준화와 효율성을 확보할 수 있는 장점이 있다.

2. 관리 플랫폼과 관측 도구

멀티 에이전트 시스템이 대규모로 확산되고 기업이 운영하는 AI 에이전트가 증가하면서, AI 에이전트가 결정을 내리고 업무를 수행하는 방식에 대한 통합적 관리와 이해가 필수 과제로 떠오르고 있다. 이를 위해 감독 기능 또는 ‘감독 에이전트’를 포함한 통합·확장형 오케스트레이션 플랫폼을 활용할 수 있다. 이러한 감독 에이전트는 요청 해석, 작업 분배, 접근 권한 부여 및 관리, 병렬 처리 및 다단계 프로세스 실행 등을 담당한다.¹⁶ 업계에서는 이르면 향후 1년 내 테크 기업들이 새로운 관련 기능을 대거 출시할 것이라는 전망이 우세하다. 이를 활용하는 기업들은 이에 맞춰 오케스트레이션 플랫폼을 자체 구축할지, 외부 솔루션을 활용할지를 두고 전략적 판단이 필요하다. 예컨대 중앙 집중형 사내 플랫폼은 벤더 종속성을 줄이고 데이터와 에이전트에 대한 통제력을 높일 수 있는 반면, 상용(off-the-shelf) 플랫폼은 실험과 도입 속도를 높이고 혁신 비용을 보다 효율적으로 관리할 수 있다는 장점이 있다.

어떤 방식을 선택하든, 에이전트 오케스트레이션 플랫폼은 운영 지표 추적, 성능 개선, 비용 관리의 핵심 인프라로 자리 잡게 될 전망이다. 현재 일부 플랫폼은 지역 시간, 오류율, 토큰 사용량 등 에이전트 텔레메트리와 다양한 도구 인사이트를 통합 모니터링하는 기능을 개발하고 있다.¹⁷ 가드레일(guardrail)* 평가와 이상 행동 감지 기능을 통해 리스크를 사전에 완화하는 시도도 확산하고 있다. 중장기적으로는 이러한 플랫폼에 계층화된 비즈니스 인사이트 제공과 추가적인 통제 메커니즘을 포함한 혁신적 기능이 추가될 수도 있다. 대표적으로 ‘가디언 에이전트

트’(guardian agent)라는 새로운 유형의 에이전트는 스스로 업무를 수행하는 동시에, 다른 에이전트의 행동을 감독하며 위험 징후를 감지 및 관리하는 역할을 수행할 수 있는 차세대 통제 수단으로 주목받고 있다.¹⁸

아울러 내재화된 규제 컴플라이언스 기능 역시 에이전트 오케스트레이션 플랫폼의 필수 요소로 자리 잡을 전망이다. 유럽연합(EU)은 AI법(AI Act)에서 위험 평가, 투명성 확보 조치, 기술적 보호장치, 인간 감독 의무 등을 명확히 규정하고 있으며,¹⁹ 이에 맞춰 EU의 표준화 기구들도 조화된(harmonized) 법적·기술적 표준을 마련하는 작업을 본격화하고 있다.²⁰

* 가드레일(guardrail)이란 시스템·AI·조직이 허용된 범위 안에서만 작동하도록 사전에 설정하는 규칙·계약·통제 장치를 의미한다. 문제가 발생한 뒤 조치하는 것이 아니라 발생하지 않도록 미리 제한하며, 무엇이 가능하고 무엇이 불가능한지 명확히 규정하고, 사람의 개입 없이도 지속적으로 작동한다.

3. 비즈니스 프로세스와 인력 구조의 변화

시장조사기관 가트너(Gartner®)는 기업용 소프트웨어 애플리케이션 중 에이전트 AI가 탑재된 비율이 2024년 1% 미만에서 2028년에는 33%로 상승할 것으로 전망했다. 이후 2028년에 이르면 일상적 업무 의사결정의 최소 15%가 AI 에이전트를 통해 자율적으로 이뤄질 것으로 예상했다.²¹ 이러한 변화에 대응하기 위해, 기업들은 2026년을 기점으로 기존 업무 흐름을 전면 재구상하고, 보다 구체적이고 독립적인 업무 모듈을 정의하는 작업에 본격 착수할 것으로 예상된다. 그 과정에서 업무의

중요도, 상호 의존성, 중요 업무의 예측 가능성, 목표로 하는 회복탄력성 수준이 어떤 유형의 에이전트 오케스트레이션이 필요한지를 결정하게 된다. 예를 들어 일부 업무 모듈은 한 에이전트의 산출물이 다음 에이전트의 입력값이 되는 순차형 협업 구조에 적합할 수 있으며, 다른 업무는 여러 에이전트가 병렬 또는 협업 방식으로 동시에 작동하는 구조가 더 큰 효과를 낼 수 있다.

또 하나의 핵심 변수는 인간과 멀티 에이전트 시스템 간 협업 방식이다. 전 세계 주요 기업의 최고인사책임자(CHRO) 200명을 대상으로 실시한 조사에 따르면, 응답자의 86%는 ‘디지털 노동력’ 즉 지능적 업무를 수행하는 기술의 통합을 CHRO의 핵심 역할로 인식하는 것으로 나타났다.²² 초기 협업 모델에서는 인간이 ‘에이전트 보스’로서 AI를 관리하거나, 에이전트와 나란히 협업하는 형태가 주를 이루고 있다.²³ 2026년에는 이러한 협업 모델이 더 많은 직무와 기능, 중요 업무 전반으로 확산될 것으로 전망된다.²⁴ 에이전트 오케스트레이션을 통해 업무 효율성을 극대화하고, 인간 고유의 강점과 협업이 더 큰 가치를 만들어낼 수 있도록 정밀한 업무의 재설계 작업이 본격화될 것으로 전망된다.

아울러 기업은 2026년부터 기존 직무가 멀티 에이전트 시스템을 통해 어떻게 더 높은 부가가치를 창출할 수 있는지도 본격적으로 재탐색하기 시작할 것이다.²⁵ 인간은 보다 창의적인 프롬프트 설계, 문제 해결 과정에서의 방향 제시, 전략적 의사결정 지원 등 에이전트를 조율하고 유도하는 역할에 집중하게 될 것이다. 동시에 기업들은 에이전트 훈련, 오케

스트레이션, 감독, 거버넌스를 담당할 새로운 인적 역량 개발과 책임 범위의 정립에도 속도를 낼 것으로 보인다.²⁶ 이에 맞춰 맞춤형 교육 프로그램 도입과 함께, 인간과 디지털 노동력을 동시에 관리할 수 있는 리더 양성이 중요 과제로 부상하고 있다. 이는 멀티 에이전트 기반 의사결정의 품질, 책임성, 회복탄력성을 제고하는 동시에, 인간 고유의 역량을 극대화하는 핵심 전략으로 평가된다.²⁷

2026년, 에이전트 오케스트레이션의 변곡점

에이전트 오케스트레이션은 향후 지능형 자율기업(intelligent enterprise)* 시대를 규정하는 핵심 요인이 될 것이다. 2026년을 기점으로 기업은 멀티 에이전트 시스템을 본격적으로 확장할 것이며, 이로 인해 IT 환경과 비즈니스 운영 여건이 한층 복잡해질 것이다. 에이전트간 통신규격 역시 실험 용이성, 유연성, 확장성, 보안성을 갖춘 소수의 표준 중심으로 재편될 가능성이 크다. 기업의 업무 프로세스는 내부 구축 에이전트는 물론 SaaS 및 제3자 공급업체가 제공하는 에이전트를 기반으로 점차 모듈화된 구조로 재설계될 것으로 보인다. 이와 동시에 인간 근로자의 역할도 변화하며, 멀티 에이전트 시스템과의 효과적인 협업을 지원하는 새롭게 변형된 직무가 본격적으로 등장하기 시작할 전망이다.

* 지능형 자율기업(intelligent enterprise)은 AI, 빅데이터, 사물인터넷(IoT) 등 첨단 기술을 활용해 데이터 기반의 실시간 의사결정, 업무 자동화 및 최적화, 사용자 맞춤형 서비스 제공을 통해 혁신을 이끌고 빠르게 변화하는 시장에 적응하며 성장하는 기업을 뜻한다.

다만 기업과 기술 공급업체 모두가 변화의 방향을 주도적으로 설계하고, 과감하게 실행해야 이러한 전환을 순조롭게 이행할 수 있다.

1. 멀티 에이전트 시스템을 도입하려는 기업이 해결해야 할 필수 과제

1) 소유권과 책임 소재 정립

기업은 자사의 AI 에이전트 비전, 전략, 실행 전반을 총괄할 최고책임자(C 레벨)를 정해야 하며, 이에 상응하는 인센티브와 책임 구조도 수립해야 한다. 일반적으로 전략적 기술 혁신과 디지털 전환을 주도하는 최고기술책임자(CTO)나 최고디지털책임자(CDO)가 해당 역할을 담당하는 것이 가장 자연스럽지만, 경우에 따라서는 기술·사업·리스크 관리를 통합적으로 아우르는 전사적 기능 조직이 더 큰 시너지를 낼 수도 있다.

2) 단기적 도입을 넘어 지속적 발전을 위한 설계 전략

AI 에이전트와 오케스트레이션 기술은 매우 빠른 속도로 발전하고 있으므로, 모듈형 ‘플러그 앤 플레이’(plug-and-play)* 방식의 오케스트레이션 프레임워크는 유연성, 비용 효율성, 혁신 역량을 동시에 높이는 핵심 수단으로 주목받고 있다. 이러한 구조는 기존 시스템 아키텍처에 대한 충격을 최소화하면서도 새로운 기술을 신속히 흡수할 수 있는 기반을 제공한다.

* 플러그 앤 플레이(plug-and-play)는 별도의 복잡한 설정이나 전문 지식 없이 연결만 하면 즉시 작동하도록 설계된 기술·제품·시스템 특성을 의미한다. 연결 즉시 장치와 소프트웨어를 자동 감지하고, 설치·설정·튜닝이 최소화돼 즉시 사용이 가능하며, 저마찰 UX를 제공해 사용자의 진입 장벽을 크게 낮춘다. 마우스와 키보드 등 USB 장치와 계정 생성 후 즉시 사용할 수 있는 클라우드 등이 이에 해당한다.

3) 멀티 에이전트 오케스트레이션의 규모 확대 전 엄격한 스트레스 테스트 선행

실제 기업 환경에서는 불완전한 데이터, 상충하는 목표, 악의적 공격 시나리오 등 다양한 리스크가 동시에 발생한다. 통제된 환경에서 이러한 복합 리스크를 반영한 시뮬레이션을 수행하면, 규모를 확대하기에 앞서 잠재적인 실패 지점을 조기에 발견하고 보호 장치를 강화할 수 있다.

4) 신중한 거버넌스 수립 및 성과 측정

AI 에이전트 거버넌스는 대규모 확장 시 에이전트 오케스트레이션이 안전하고, 규제를 준수하며, 신뢰할 수 있는 방식으로 작동하도록 보장하는 핵심 장치다. 이를 위해 에이전트의 역할을 명확히 설정하고, 책임 범위를 정의하며, 오류 발생 시 우회 경로(fallback route)와 감독 체계를 설계하는 것이 필수적이다. 이는 오남용을 방지하고 감사 가능성을 확보하며 궁극적으로 신뢰를 구축하는 토대가 된다. 기업은 기술적 준비와 더불어 에이전트 오케스트레이션이 실제 어떠한 방식으로 가치를 창출하는지를 보여주는 성과 지표를 명확히 정의하고 지속적으로 추적할 필요가 있다. 이러한 성과 지표에는 의사결정 속도 개선, 고객 경험 향상, 혁신 속도 제고 등을 포함할 수 있다.

2. 멀티 에이전트 시대에 대비하는 기술 공급업체의 전략적 과제

1) 상호운용성을 전제로 한 설계

단순히 에이전트간 통신 표준을 준수하는 수준을 넘어, 솔루션 자체를 모듈형 구조로 설계하고 에이전트가 상호 의도와 행동 맥락을 이해할 수 있도록 구현해야 원활한 협업과 조율이 가능해진다.

2) 신뢰 개념의 근본적 재정의

이제 단순히 인사이트를 전달하는 것만으로는 충분하지 않으며, AI 에이전트의 산출물을 인간이 이해하고 검증할 수 있어야만 신뢰를 바탕으로 AI 에이전트의 도입을 확대할 수 있다. 에이전트에 디지털 아이덴티티를 부여하는 등 새로운 보안 메커니즘 장치를 구축해야, 신뢰 가능한 멀티 에이전트 시스템을 구축 및 운영할 수 있다.

3) 거버넌스를 설계 단계부터 내재화

고객사가 인적 가치, 조직 정책, 규제 요건 부합 여부 등을 중요시함에 따라, 이를 선제적으로 학습하고 반영하는 것이 향후 솔루션의 경쟁력을 좌우할 핵심 요소로 작용할 것이다. 이에 따라 차세대 솔루션에는 에이전트 모니터링 기능, 고도화된 거버넌스 체계, 윤리적 가드레일이 기본 구성 요소로 탑재돼야 한다. 이러한 방향으로 발전하는 AI 에이전트는 컴플라이언스와 실효성을 동시에 갖출 수 있다.

4) 생태계 확장

산업 전반에서 통신규격, 신뢰 체계, 거버넌스 표준을 정립하려면 산업 차원의 연합과 동맹을 지속적으로 구축 및 강화하는 노력이 필수적이다. 최근에는 플랫폼을 넘나드는 혁신적 오케스트레이션 다수의 틀이 급부상하고 있는데, 이는 기존 강자와 신생 기업 모두에게 인수합병(M&A), 전략적 제휴, 공동 협업을 통해 시장 입지를 강화할 수 있는 새로운 기회로 작용할 전망이다.²⁸

Korean Perspectives

멀티 에이전트 시대의 핵심 경쟁력은 오케스트레이션이다

미터 에이전트 시스템의 확산으로 한국 기업들은 단순한 AI 도입 단계를 넘어, 기업 운영체계 전반을 재설계해야 하는 상황에 직면했다. 이제 경쟁의 초점은 개별적으로 성능이 좋은 AI 에이전트를 얼마나 많이 보유했는가가 아니라, 서로 다른 에이전트를 어떻게 조율하고 통제하며 신뢰 가능한 방식으로 실행에 유기적으로 연결하느냐로 이동하고 있다. 실제로 에이전트 오케스트레이션 역량이 부족할 경우, 기술적 잠재력과 무관하게 비용 증가, 확장 실패, 리스크 누적으로 프로젝트가 중단될 가능성성이 높아진다.

이러한 맥락에서 한국 기업이 가장 먼저 정립해야 할 것은 기술이 아니라 자율성의 경계와 책임 구조다. 멀티 에이전트 환경에서는 인간이 언제 개입하고, 감독하며, 완전히 위임할 것인지에 대한 기준이 명확하지 않으면 오히려 의사결정의 품질과 속도가 모두 저하될 수 있다. 따라서 단기적으로는 휴먼 인 더 루프(human-in-the-loop) 구조를 유지하며 신뢰와 학습을 축적하고, 중장기적으로 휴먼 온 더 루프(human-on-the-loop)로의 점진적 전환을 염두에 둔 설계가 현실적인 선택이 된다.

도입 전략 역시 일률적일 수 없다. 기존 업무에 에이전트를 덧붙이는 방식은 빠른 실험에는 유리하지만, 통합·보안·비용 관리 측면에서 복잡성을 빠르게 증폭시킨다. 반대로 장기 경쟁력을 확보하려면 업무를 보다 세분화된 모듈 단위로 재구성하고, 에이전트가 이해할 수 있는 컨텍스트 기반 구조(지식 체계, 온톨로지, 데이터 맥락)를 선제적으로 정비해야 한다. 멀티 에이전트의 성능은 결국 모델보다 기업 고유의 맥락 정보 품질에 의해 좌우되기 때문이다.

에이전트 수가 늘어날수록 통제의 중요성은 기하급수적으로 커진다. 관측 가능한 텔레메트리, 명확한 가드레일, 오류 발생 시 되돌림과 우회 경로를 포함한 복구 메커니즘은 선택 사항이 아니라 엔터프라이즈급 필수 인프라다. 특히 자율성이 높아질수록 사용자 인터페이스(UI)의 역할은 줄어들지 않고 오히려 강화된다. 에이전트의 판단 근거를 설명하고, 실행 이력을 추적하며, 인간이 개입해야 할 시점을 명확히 보여주는 경험 설계 없이는 조직 내부의 신뢰를 확보하기 어렵다.

또 하나 간과하기 쉬운 부분은 인력과 조직의 변화다. 멀티 에이전트 환경에서 직원은 더 이상 단순한 시스템 사용자가 아니라, 에이전트를 설계·조율·감독하는 오케스트레이터로 역할이 이동한다. 이는 교육, 직무 정의, 성과 평가 방식의 변화를 동시에 요구하며, 이를 뒷받침하지 못하는 조직은 기술을 도입하고도 실제 성과를 얻지 못할 가능성이 크다.

결국 멀티 에이전트 시대에 한국 기업의 승부처는 기술 도입 속도가 아니라, 에이전트를 중심으로 한 운영 모델과 거버넌스를 얼마나 일관성 있게 재설계했는가에 있다. 에이전트 오케스트레이션은 IT 프로젝트가 아니라, 지능형 기업으로 전환하기 위한 경영 구조의 문제다. 이 인식을 얼마나 빨리 받아들이느냐가 2026년 이후 성과 격차를 가르는 결정적 변수가 될 것이다.



정창모 파트너
한국 딜로이트 그룹
One AI 파트너

4장

휴머노이드·산업용 로봇·드론의 AI 혁신: 자율성과 생산성의 새로운 임계점

강력한 AI 모델과 반도체 성능 향상을 계기로 로봇·드론 산업은 단순 자동화를 넘어 ‘지능형 노동’을 구현하는 핵심 산업으로 재도약하고 있다. 고령화와 인력 부족, 공급망 회복력 강화 수요 속에서 비전-언어-행동(VLA, vision-language-action) 모델을 중심으로 한 자율형 로봇과 드론은 제조·의료·물류·국방 등 전 산업으로 빠르게 확산되고 있다. 2026년부터 경쟁의 본질은 개별 로봇 제품이 아니라, AI·반도체·소프트웨어·보안·인력 역량을 통합한 생태계 구축 역량으로 이동하며, 이를 얼마나 안정적으로 상용화하고 확장할 수 있는지가 성과를 좌우하게 될 것이다.

핵심 내용 요약 (Executive Summary)

▷ 재도약의 출발점: AI가 로봇·드론 산업을 다시 움직인다

- 강력한 AI 모델과 반도체 성능 향상으로 산업용 로봇·휴머노이드·드론이 본격 성장 국면 진입
- 2026년 기준 산업용 로봇 누적 설치 550만 대, 연간 출하 60만 대, 매출 144억 달러로 지속 확대
- 로보틱스는 단순 자동화에서 ‘지능형 노동’으로 성격 전환

▷ 기술 진화의 핵심: 특수 목적 AI와 VLA 모델

- 로봇은 명령 수행을 넘어 자연어 이해·환경 인지·복합 작업 학습이 가능한 자율 시스템으로 진화
- VLA 모델이 휴머노이드와 고급 로봇의 자율성과 범용성을 크게 확대
- ‘하드웨어 성능’보다 ‘지능 모델과 학습 구조’가 차별화 요소로 부상

▷ 산업 전반으로의 확산: 로봇과 드론의 실전 배치

- 제조·의료·물류·국방 등에서 로봇이 위험·반복·고정밀 작업을 대체
- 드론은 자율 비행, 장애물 회피, 임무 수행 능력 고도화로 농업·에너지·재난 대응까지 확장

▷ 생태계 경쟁의 시작: 반도체·보안·인력까지 포함한 총력전

- 산업용 로봇 1대당 2.5만~5만 달러 규모의 칩·전자부품 탑재 → 반도체 산업의 신규 성장축
- 2026년 이후 경쟁은 개별 로봇 제품이 아니라 AI·반도체·소프트웨어·보안·인력 역량을 결합한 생태계 경쟁으로 전환

더 강력한 AI 모델과 반도체칩의 등장으로 그동안 상대적으로 정체돼 있던 산업용 로봇, 휴머노이드, 드론 산업의 활성화가 기대된다.

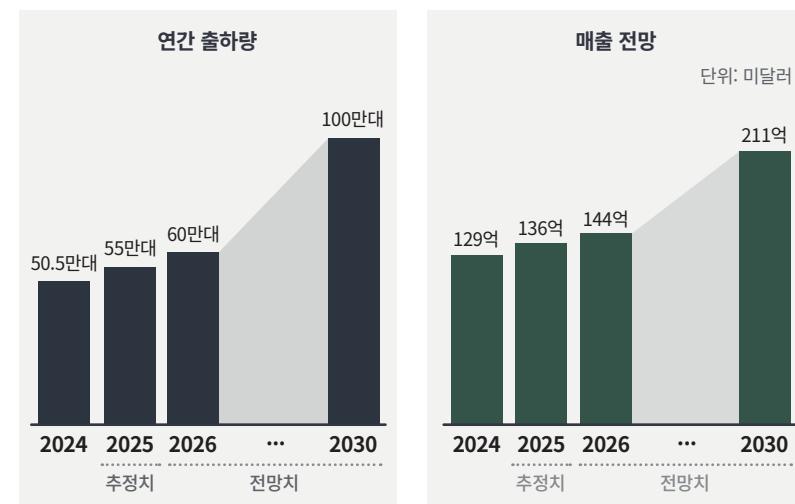
사람들은 언젠가 인간과 유사한 지능을 갖추고 행동하는 휴머노이드 로봇이 공장에서 분주하게 움직이며 일하는 모습을 기대한다. 2030~2040년경에는 이러한 공상과학과도 같은 장면이 현실화될 수도 있다. 그러나 2026년 현재, 현실은 이러한 상상과 여전히 거리가 멀다. 딜로이트는 전 세계 산업용 로봇의 누적 설치 대수가 2025년에 겨우 500만 대를 넘고 2026년에도 550만 대까지 증가하는 수준에 그칠 것이라고 전망한다.¹

로봇 시스템에 인공지능(AI) 역량이 더욱 긴밀히 통합되고 특화된 기초 AI 모델이 발전함에 따라, 로봇은 스마트 팩토리부터 공공 서비스, 심지어 자율 드론에 이르기까지 다양한 산업과 응용 분야로 확산될 수 있다. 그러나 첨단기술과 AI, 로보틱스를 모두 아우르는 생태계 내에서 데이터 품질, 통합, 사이버 보안의 개선을 저해하는 병목이 해결되지 않는다면, 산업용 로봇 시장의 성장 속도는 지금처럼 완만한 수준에 머물 가능성이 크다.

산업용 로봇 시장 성장의 촉매제, 첨단 특수 AI 모델

딜로이트가 ‘2020 첨단기술·미디어·통신(TMT) 산업 전망’에서 예측했듯, 신규 산업용 로봇의 연간 판매량은 2021년부터 약 50만 대 수준에서 정체돼 있다.² 하지만 장기적으로는 대규모 성장이 기대된다. 일각에서는 휴머노이드 로봇 산업이 2050년에 5조 달러 규모에 이를 것이라는 추정이 나온 바 있다.³ 보다 급진적으로는 이르면 2030년에 연간 신규 로봇 출하량이 현재의 두 배 수준으로 증가해 연간 100만 대에 도달하고, 2030년 예상 매출이 210억 달러에 이를 수 있다는 전망도 제시됐다. 이는 2024년에 비해 거의 두 배 증가하는 수준이다(그림 1).⁴

그림 1
AI 기반 산업용 로봇 연간 출하량 및 매출 전망



출처: Deloitte analysis based on publicly available information from sources including the International Federation of Robotics, Interact Analysis, and IPF Online.

본고에서 다루는 ‘로봇’의 의미

‘로봇’이라는 용어가 포함하는 대상은 매우 광범위하다. 식기세척기부터 수백 달리짜리 가정용 스마트 진공청소기, 대당 수백만 달러에 이르는 생산라인의 산업용 로봇까지 모두 로봇이라 할 수 있다. 때로는 비행 로봇(드론), 주행 로봇(완전자율주행 차량), 인간이 할 수 있는 거의 모든 일과 어쩌면 그 이상을 수행할 수 있는 휴머노이드 로봇까지 포함하기도 한다.

최근 물리적 AI, 로보틱스, 드론 분야가 부상하고 있으며, 자율주행차에 대해서는 이미 상당수의 기사와 분석 보고서가 이미 발표됐다. 따라서 본고는 주로 산업용 로봇, 산업용 활용을 목적으로 제작된 휴머노이드 로봇, 드론에 초점을 맞춘다.

2026년에서 2030년 사이 두 가지 촉매 요인이 산업용 로봇 도입을 촉발할 것으로 기대된다. 첫째, 선진국은 고령화로 인해 지속적인 인력 부족에 시달리고 있다.⁵ 이에 따라 각국이 국내 제조 역량을 강화하고 회복력 있는 공급망을 구축하려는 움직임이 확대되면서, 점점 더 정교한 작업을 처리할 수 있는 로봇 수요가 계속 증가할 가능성이 크다. 하지만 더욱 강한 성장 촉매는 두 번째 요인이다. 컴퓨팅 파워(computing power)*가 급속도로 발전함과 동시에 기존의 대형 언어 모델과는 또 다른 특수 목적의 기초 AI 모델이 진화하면서 AI 로봇과 임베디드 AI 시스템의 개발이 가속화되고 있다.⁶ 특수 목적 모델은 로봇이 단순한 명령·제어 수준을 넘어 자연어 이해, 물리적 환경 인지, 복잡한 작업 학습·탐색 등 인간처럼 일반화된 방식으로 기능할 수 있도록 고도화된 AI 엔진을 구현하는 데 핵심 역할을 한다.⁷

* 컴퓨팅 파워(computing power)는 주어진 시간 안에 데이터를 처리하고 연산을 수행할 수 있는 시스템의 총합적 계산 능력을 의미한다.

그러나 로봇 도입을 둘러싼 뜨거운 열기와 첨단 기술의 발전에도 불구하고 로봇공학의 발전을 가로막는 장애물이 여전히 있다. 우선 데이터 품질, 상호운용성, 레거시 시스템 호환성 등의 문제로 로봇 시스템을 기존 산업 워크플로에 통합하는 과정이 복잡하다. 많은 기업이 로봇 학습에 필수인 정제되고 통합된 데이터세트(예: 실사용 데이터, 물리적 환경 데이터, 공간 데이터)를 확보하는 데 어려움을 겪고 있다.⁸ 더불어, 커넥티드 로봇 네트워크를 겨냥한 보안 및 개인정보 침해나 악의적 사이버 공격 위험도 매우 심각한 우려 사항으로 남아 있다.⁹ 또한 산업용 로봇과 휴머노이드 로봇은 인간 작업자의 안전을 반드시 보장해야 한다는 점도 핵심 과제다.¹⁰

딜로이트는 생성형 AI와 에이전틱 AI를 로보틱스 및 자동화 도구와 더 밀접하게 통합해야만, AI 기반 로봇 장치가 공상과학의 영역에서 벗어나 실제 현대화된 작업 현장에서 실현될 것으로 보고 있다.¹¹ 실제 사례로, 실제 첨단 활용 사례를 시뮬레이션하는 목적으로 미국 캔자스주 위치타(Wichita)에 구축된 스마트 팩토리는 생성형 AI, 에이전틱 AI, 언리미티드 리얼리티(unlimited reality)* 기술뿐 아니라 드론, 자율 모바일 로봇, 사족보행 로봇, 휴머노이드 로봇 등 다양한 로봇 기술을 보유하고 있다.¹²

* 언리미티드 리얼리티(unlimited reality)는 물리적 현실과 디지털 세계의 경계를 사실상 제거해, 시간·공간·디바이스 제약 없이 ‘확장·융합된 현실 경험’을 제공하는 차세대 디지털 환경을 의미한다.

이미 산업용 로봇은 제조, 의료, 물류창고, 심지어 국방 등 여러 산업에서 가치를 창출하고 있다(그림 2).¹³ 앞으로는 여기서 더 나아가 향후 일부 테크 기업이 실현하는 혁신, 특히 멀티모달 AI 모델의 발전과 첨단 반도체칩 및 하드웨어 기술의 발전이 산업용 로봇 시장의 새로운 성장 기회를 창출할 것으로 기대된다.

그림 2 AI 기술 발전에 힘입어 다양한 산업에서 가치를 창출하는 산업용 로봇

산업 환경	AI 기반 로봇의 효용과 활용사례
제조업	바이오닉 손을 갖춘 인지형 휴머노이드를 포함한 로봇은 합성 데이터를 활용해 스스로 학습하고 첨단 공장에서 서로 조율된 방식으로 작업할 수 있다. 3D 객체 인식 기능을 갖춘 이러한 로봇은 설비 적재, 사출 성형, 유지보수 등 다양한 작업을 지원할 수 있다.
의료	로봇은 간호 인력이 수행하던 검체 채취 및 운반, 약품 전달 등의 단순 반복 업무를 대신해 업무 부담을 줄이고, 외과 수술에서는 사람의 손으로 구현하기 어려운 미세 단위 정밀 작업을 보조함으로써 수술 결과의 안전성과 정확도를 높인다. 또한 바이러스나 유해 화학물질 등 위험 물질을 다루는 작업 시 인력을 보호하고, 병실 및 시설 소독과 같은 고위험 환경 업무도 수행하며 의료기관의 안전성과 효율성을 동시에 강화할 수 있다.
물류창고	AI 기반 로봇은 딥러닝 비전을 활용해 다양한 형태와 크기의 물체를 인식하고 다룰 수 있다. 자율주행 모바일 로봇은 카메라와 3D 비전에서 생성되는 고도화된 센서 데이터를 AI 알고리즘과 결합해, 창고 환경을 실시간 매핑 및 탐색하며, 고정된 인프라나 마커에 의존하지 않고도 정확하게 위치를 잡을 수 있다.
국방	로봇견(robot dog)은 첨단 센서와 AI 기반 분석 기술을 활용해 폭발물을 탐지 및 처리하고 위험 물체를 분류 및 식별할 수 있다. AI 로봇은 감시와 정찰 임무를 수행하고, 보급품을 운반하며, 부상자 후송과 같은 긴급 대피 지원 업무도 담당 할 수 있다.

출처: Insights gathered from conversations with industry subject matter experts, as well as multiple publicly-available sources including: American Machinist, Admedica, World Economic Forum, PHS Innovate, and ASDNews.

비전-언어-행동(VLA) 모델로 지능과 자율성 강화된 휴머노이드 로봇

일부 AI 스타트업과 주요 테크 기업들은 로봇이 사전 프로그램된 작업을 수행하는 수준을 넘어, 문맥을 이해하고 스스로 판단을 내릴 수 있는 비전-언어-행동(VLA, vision-language-action) 모델을 개발하고 있다. VLA 기능을 갖춘 로봇은 자율성이 강화돼 더 높은 수준의 계획 수립과 공간 추론이 가능하고, 까다로운 지형을 기민하게 이동할 수도 있다.¹⁴ 로봇은 대규모 시뮬레이션 기반 강화학습과 멀티모달 학습을 통해 방대한 데이터세트로 사전학습을 할 수 있다.

VLA는 시각적 인지(환경과 물리 법칙을 관찰), 자연어 이해(음성 명령과 해석), 실세계에서의 행동 수행(시각·텍스트 지시사항에 대한 반응)을 통합해 로봇이 작업을 수행하도록 한다.¹⁵ 2025년 중반 기준, VLA의 파라미터(parameter, 매개 변수)는 대체로 5억~70억 개 수준으로, 휴머노이드 로봇은 이를 기반으로 학습, 지각, 행동할 수 있다.¹⁶ 미국에서는 VLA 모델이 로보틱스 개발을 고도화하기 위해 활용되는 사례가 등장하고 있으며, 2026년~2030년에는 상용화가 더욱 확대될 수 있다.

- 엔비디아(NVIDIA)는 휴머노이드 로봇을 위한 공개형 기초 모델을 통해 추론과 행동을 결합해 로보틱스 개발을 진전시키고 있다.¹⁷ 보스턴다이나믹스(Boston Dynamics) 등 로보틱스 기업들은 엔비디아 모델의 라이브러리와 지원 기술을 활용해 휴머노이드 로봇을 개발하고 있다.¹⁸

- 피규어AI(figure AI)의 헬릭스(Helix)는 시각 및 자연어 프롬프트를 사용해 로봇을 학습시키는 VLA 모델로, 휴머노이드 로봇이 실세계 장면과 사물을 깊이 이해하고 정교한 운동 능력을 개발하도록 지원한다.¹⁹
- 헤깅페이스(Hugging Face)는 로봇용 오픈소스 데이터와 모델을 개발하고 있으며, 자체 오픈소스 휴머노이드 로봇을 구축 및 테스트하면서²⁰ 개발자들이 자신만의 로봇을 커스터마이즈할 수 있도록 지원하고 있다.²¹

아시아와 유럽에서도 맞춤형 기초 모델과 물리적 세계의 데이터 기반 학습에 초점을 맞춘 휴머노이드 로봇 개발이 활발히 추진되고 있다. 일례로 한국 스타트업 리얼월드(RLWRLD)는 자동화된 학습과 인간 전문성의 모방을 통해 기존 수작업 중심 공정을 로봇이 자율적으로 수행할 수 있도록 하는 기초 AI 모델을 개발 중이다.²² 일본 화낙(FANUC)은 산업 환경에서 활용할 수 있는 다양한 크기의 AI 기반 로봇 개발에 주력하고 있다.²³ 유럽에서는 런던 소재 뉴럴파운드리(Neural Foundry), 독일 뉴라로보틱스(NEURA Robotics) 등이 인지 능력을 통합하고 맞춤형 모델을 개발해 산업용 AI 로봇을 개발하고 있다.

중국에서는 애지봇(AgiBot), 매직랩(MagicLab) 등 스타트업이 제조 환경에서 복잡한 작업을 수행할 수 있는 휴머노이드 로봇을 설계하고

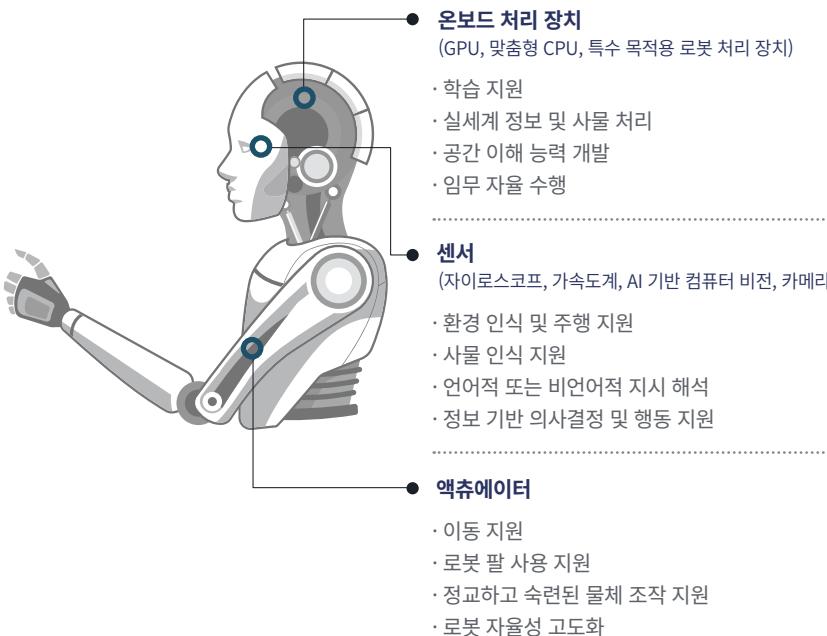
있다.²⁴ 유니트리로보틱스(Unitree Robotics), 유비테크로보틱스(UBTECH Robotics) 등은 대량 생산 체계를 구축해 휴머노이드 로봇의 가격을 낮추고 접근성을 개선해 보급 확대를 꾀하고 있다.²⁵ 휴머노이드 로봇 제작에는 다양한 반도체칩 부품과 하드웨어가 사용되기 때문에 (그림 3), 관련 시장이 반도체(칩 하드웨어 및 관련 소프트웨어·서비스 포함) 산업에 큰 매출 기회를 제공할 것으로 기대된다.²⁶

대규모 산업용 활용에 적합한 AI 기반 휴머노이드 로봇은 아직 개발 초기 단계이다. 하지만 딜로이트는 2025년 연간 출하량을 5,000~7,000대로 추산하며, 2026년에는 1만5,000대로 증가할 수 있다고 전망한다.²⁷ 대당 평균 가격이 1만4,000~1만8,000달러임을 감안하면,²⁸ 2026년 산업용 AI 휴머노이드 로봇 시장 규모는 약 2억1,000만~2억7,000만 달러로 전망된다.²⁹ 로보틱스 산업이 2026~2030년 기술, 가격, 운영상의 장벽을 성공적으로 극복하면, 휴머노이드 로봇 시장은 2032년에 6억~7억 달러(2026년 대비 약 3배 성장) 혹은 낙관적 시나리오 기준 10억 달러(2026년 시장의 약 4배)에 이를 수도 있다.³⁰

그림 3

AI 기반 휴머노이드 로봇에 탑재되는 반도체칩 부품과 하드웨어

사람과 유사한 외형을 갖추고 대화형 AI 기능까지 탑재한 휴머노이드 로봇은 눈 맞춤까지 가능해지고 있다. VLA 모델과 관련 소프트웨어 외에도, 이러한 체화형(embodied) AI 시스템에는 전용 하드웨어 부품이 필수적으로 사용된다.



참조: VLA: 비전-언어-행동(vision language action) / GPU: 그래픽처리장치(graphics processing unit) / CPU: 중앙처리장치(central processing unit)

출처: Deloitte analysis.

첨단 AI로 자율성과 효율성 강화된 드론

무인 항공기(UAV)로도 불리는 드론은 현재 대부분 수동으로 조종되지만, 자율 기능이 빠르게 고도화되고 있다. 머지않아 대부분의 드론이 AI를 활용해 실시간 내비게이션을 수행하고, 상호 통신하며, 장애물과 충돌을 피하고, 인간 개입 없이 임무를 수행할 수 있을 것으로 보인다. 일례로 형가리 연구진은 비둘기와 야생마 등 다양한 동물의 이동 패턴과 행동을 연구한 결과를 적용해 기내에서 자율적 결정을 내릴 수 있는 드론 군집을 위한 알고리즘을 개발했다. 해당 드론은 단순히 비행 경로를 탐색하고 충돌을 피하며 공중에서 안정적으로 제공하는 것을 넘어, 지상 조사, 기상 관측, 산불 관리 등 다양한 환경에서 임무를 수행할 수 있다.³¹

드론에 탑재되는 장치

AI 기반의 정교한 첨단 드론에는 다양한 기술과 반도체칩이 탑재된다. 단일 또는 듀얼 마이크로컨트롤러가 비행 컨트롤러 역할을 하며, 기내 전원 시스템에는 리튬 기반 배터리와 여러 부품에 전력을 공급하는 분배 보드가 장착된다. 드론과 지상 통제 장치 간 통신을 지원하는 무선 주파수(RF) 모듈, 내비게이션과 위치 파악을 위한 위성항법시스템(GPS) 모듈, 가속도계·자이로스코프·자력계·옵티컬플로우센서·라이다·초음파센서 등 각종 센서와 더불어 공중 운용을 관리하는 기내 비행 제어 소프트웨어와 플랫폼 등이 탑재된다.³²

‘딜로이트 2024 농업 기술 예측’에서 전망된 바와 같이, UAV 혹은 드론에 장착된 분광 센서, 칩, 카메라의 조합이 토양 수분, 작물 건강 상태 등 대량의 데이터를 수집하면, AI 모델이 이를 분석해 표적 살포 작업을 위한 분석 결과를 제공한다.³³ 드론은 농업 외에도 풍력 터빈과 송전선을 점검하는 데 활용해 수작업 점검 필요성을 최소화할 수 있다.³⁴ 중국, 남호주, 영국은 고전압 송전선의 완전 자율 장거리 원격 점검을 수행하는 UAV 실험을 진행 중이다. 해당 드론은 중요하지만 매우 위험한 작업을 사람 대신 수행해 작업자와 엔지니어를 지원할 뿐 아니라, 수십 장의 이미지를 자동으로 수집 및 전송해 엔지니어들이 AI 및 고급 분석 기술을 기반으로 부식을 감지 및 분석할 수 있도록 지원한다.³⁵

여러 국가는 재난 대응 지원뿐 아니라 국경 위협 탐지 및 대응을 위해 공중 감시용 자율 드론 배치를 추진하고 있다.³⁶ 일례로 자율 드론은 2022년 9월 허리케인 이언(Ian)이 플로리다 남서부를 강타한 이후 피해 지역을 지도화해 구조대원을 지원했다. 이러한 여러 응용 사례에서 드론은 임무 일부 구간에서만 인간에 의해 원격 조종된다. 나머지는 AI가 상업 항공기의 자동조종장치처럼 작동해 드론을 목적지 근처까지 이동시킨 후 인간 조종자에게 제어권을 넘긴다. 특히 최근 여러 국가가 군사용 드론 군집 기술 개발을 추진하고 있어³⁷ 관련 기술이 비군사적(산업·민간) 분야에도 영향을 미칠 것으로 예상된다. 예를 들어, 자율 드론 군집은 접근이 어렵고 혐준한 지형에 있는 고전압 송전선을 점검하거나, 거친 기상 조건 속에서 해상 풍력 터빈을 모니터링하는 데 활용될 수 있다.

산업용 로봇 산업 성장은 상용화, 안전성, 인력 준비도가 좌우

산업용 로봇 시장은 최근 몇 년 동안 완만하게 성장하는 데 그쳤으나, 이미 반도체 기업들에게 중요한 최종 수요처로 자리매김하고 있다. 구체적으로, 20만 달러짜리 산업용 로봇 한 대에는 약 2만5,000~5만 달러 상당의 칩과 관련 전자 부품이 탑재된다.³⁸ 또한 산업용 로봇의 성능을 고도화하려면 프로세서부터 네트워크, 센서까지 첨단 반도체칩의 발전이 매우 중요하며, 로봇 한 대당 반도체 탑재량도 증가할 가능성이 크다. 더 나아가, 반도체 산업은 2025년 기준 산업용 로봇의 주요 소비자이자 최종 사용자로서, 팹 제조 공정, 웨이퍼 핸들링, 테스트 및 분류, 첨단 패키징, 클린룸 등 다양한 영역에서 로봇을 활용하고 있다.³⁹ 반도체 산업은 ‘라이트아웃’(lights out, 완전 자동화) 제조로 가는 여정에서 운영 과정에 더 많은 산업용 로봇을 사용하게 될 것이다.

휴머노이드 로봇과 드론을 포함한 AI 기반 산업용 로봇 시장 전망이 유망해짐에 따라, 많은 반도체 및 테크 기업들이 관련 분야에 장기적 투자를 확대하고 있다. 로보틱스 스타트업은 물류센터, 공급망, 항공 자율화 등 실제 환경에서 파일럿 단계를 진행 중이다. 벤처캐피털의 로보틱스 분야 투자도 증가해, 2025년에 자금 조달이 늘어날 가능성이 있는 유일한 비(非)AI 시장으로 꼽힌다.⁴⁰ 클라우드 및 IT 인프라도 정비되고 있으며, 합성 데이터 생성과 물리적 시뮬레이터를 활용해 개발 속도를 높이고 비용이 많이 드는 실세계 시험 의존도를 낮추는 사례도 늘어나고 있다.

다음은 산업용 로봇 상용화 과정에서 발생할 수 있는 잠재적 과제를 해결하기 위해 AI·로보틱스·기술 산업 리더들이 고려할 수 있는 다섯 가지 실행 방안이다. 리더들은 데이터 통합, 개인정보보호 및 사이버 보안, 안전성, 인력 준비도와 관련된 과제를 해결해야 한다.

- **오픈 이노베이션을 통한 상업적 실현 가능성 입증:** 테크 및 AI 기업은 광범위한 로봇 배치와 조정을 실현하는 공개형 풀스택 생태계(full-stack ecosystem)*를 조성해야 한다. 이를 통해 상용화를 확대하면 ROI를 입증할 수 있다. 또한 범용 임베디드 AI로 나아가기 위한 협력적 생태계를 구축할 필요가 있다.⁴¹
- **데이터 품질 제고 및 통합 문제 해결:** 생태계 참여자들은 다양한 로봇을 산업 환경에 더 매끄럽게 통합할 수 있도록 공용 플랫폼과 미들웨어(middleware)**에 적용할 데이터 표준화와 협업을 우선시해야 한다.

* 풀스택 생태계(full-stack ecosystem)는 단일 기술이나 제품이 아니라, 인프라 부터 플랫폼, 서비스까지 기술 구현에 필요한 모든 계층을 하나의 통합된 구조 안에서 제공 및 운영하는 구조를 의미한다.

** 미들웨어(middleware)는 운영체제(OS)와 애플리케이션 사이에서 서로 다른 시스템·서비스·데이터가 원활하게 통신하고 동작하도록 중개·조정하는 소프트웨어 계층을 의미한다.

- **사이버 취약성 해소:** 기업은 사이버 전문가들과 적극적으로 협력해 공통 상호운용성 통신규격을 도입하고, 개인정보보호 및 보안 내재화(security-by-design) 접근법을 도입하며, 명확하고 유연한 보안 체계를 마련해야 한다.

- **안전을 필수적, 통합적 기능으로 설계:** 개발 초기 단계부터 로봇은 안전을 내재화해야 한다. 이는 인간과 함께 작업할 때 신체적 위험을 초래하지 않도록 하는 것부터, 로봇끼리 우발적으로 충돌하지 않도록 보장하는 것까지 포함한다. 시뮬레이션 기반 훈련, 컴퓨터 지원 안전 계획 도구, 충돌 없는 동작 계획 등은 로봇을 더 안전하게 만드는데 도움이 되는 새로운 접근 방식이다.

- **현 인력의 사전적 역량 강화:** 첨단 AI 기술에 대한 재교육과 역량 강화는 모든 기업에 필수적이다. 산업용 AI 자동화의 새로운 물결에서 로봇이 인간과 함께 일하는 비중이 커짐에 따라, 기업은 인력이 산업용 로봇 도입과 통합을 선도할 수 있도록 AI 역량 수준을 정기적으로 평가 및 강화해야 한다.

앞으로의 방향은 명확하다. AI와 로보틱스 산업은 이미 기술의 고도화를 달성했고 상업적 수요 및 관심도 충분히 형성됐다. 2026년은 오픈 이노베이션, 데이터, 사이버, 안전성, 인재라는 다섯 가지 영역 전반에 걸쳐 진전을 이루고 산업용 로봇의 상업적 도입이 가속화되는 대전환의 시작

점이 될 것이다. 이를 위해 조직과 기술, 프로세스 전반을 큰 그림으로 바라보며 각 요소간 상호작용과 영향까지 포괄적으로 고려하는 360도 시스템 사고와 생태계 기반 접근이 필수적이다.

Korean Perspectives

로보틱스는 미래가 아니라 전략이다

한국 기업에게 로보틱스는 더 이상 먼 미래의 실험이 아니다. 국내 주요 모빌리티 기업들이 차량을 넘어 로봇으로 영역을 확장하고 있고, 생산 현장, 물류 센터, 라스트마일 배송까지 로봇 도입이 구조적으로 확대되고 있다. 이는 휴머노이드라는 특정 품팩터의 문제가 아니라, 이동·조작·인식·판단이라는 모빌리티의 핵심 역량이 물리적 자동화 전반으로 확장되고 있다는 신호다.

제조 공정에서는 인력 부족과 안전 이슈로 인해 산업용 로봇과 자율 모바일 로봇(AMR)의 활용이 빠르게 늘고 있으며, 물류와 배송 영역에서도 드론과 로봇을 통한 자동화 실험이 점차 상용 단계로 이동하고 있다. 이러한 흐름 속에서 로봇은 단일 장비가 아니라 AI, 반도체, 센서, 소프트웨어, 운영 데이터가 결합된 통합 솔루션으로 자리 잡고 있다.

문제는 기술의 가능성이 아니라 상용화의 속도다. 로봇 도입이 본격 확산되기 위해서는 데이터 품질, 시스템 통합, 사이버 보안, 작업자 안전이 동시에 해결돼야 하며, 이 중 하나라도 미흡할 경우 로봇은 파일럿을 넘기지 못한다. 따라서 한국 기업에게 중요한 것은 “얼마나 빠르게 로봇을 도입하느냐”가 아니라, 어떤 공정과 물류 흐름에서 로봇이 반복적으로 가치를 증명할 수 있는지를 명확히 정의하는 것이다.

결국 로보틱스는 어느 특정 영역이 아니라 제조 산업 전반을 관통하는 전략으로 완성되어야 하는 것이다.

한국 기업이 경쟁력을 가질 수 있는 지점은 완성형 로봇 자체보다, 현장 데이터와 운영 노하우를 축적하며 로봇을 ‘산업 시스템’으로 작동시키는 능력에 있다. 로봇은 미래의 이야기가 아니다. 이미 산업의 확장 축으로 작동하기 시작한 현재진행형 변화라고 볼 수 있다.



이중희 파트너
한국 딜로이트 그룹
Physical AI 리더

5장

생성형 AI 다음 단계 성패는 연산력이 좌우

인공지능(AI) 연산의 무게중심이 학습에서 추론으로 이동하면서, AI 경쟁의 핵심은 알고리즘 효율이 아니라 대규모 추론을 감당할 수 있는 데 이터센터, 고성능 칩, 전력 인프라 역량으로 재편되고 있다. 엣지 AI의 확산 가능성에도 불구하고, 단기적으로는 데이터센터 중심의 고성능 추론 구조와 고대역폭 메모리(HBM)가 결합된 그래픽처리장치(GPU) 체계가 주류를 이루며, 온프레미스·하이브리드 인프라 수요도 확대되고 있다. 로봇·드론·자율주행차 등 일부 영역에서는 엣지 AI가 필수적이지만, 2030년까지 급증하는 AI 연산 수요를 고려할 때 당분간은 인프라 투자 역량이 AI 경쟁력의 전제 조건으로 유지될 전망이다.

핵심 내용 요약 (Executive Summary)

▷ AI 연산의 무게중심 이동: 학습에서 추론으로

- AI 컴퓨팅의 무게중심이 모델 학습(training)에서 대규모 추론(inference)으로 이동
- 2026년 전체 AI 연산의 약 3분의 2가 추론에 사용될 전망
- AI 활용이 연구·개발 단계를 넘어 일상적·상시적 서비스로 확산되고 있음을 의미

▷ 인프라의 중심은 여전히 데이터센터

- 엣지 AI 확대 기대에도 불구하고, 고성능 추론은 데이터센터·엔터프라이즈 서버가 주도
- 데이터센터 CAPEX는 2026년 4,000~4,500억 달러, 2028년 1조 달러까지 확대 전망
- 대규모·집중형 인프라가 AI 추론의 주 무대가 되는 구조가 고착

▷ AI 칩과 연산 구조의 이종화

- 추론 최적화 칩 시장이 2026년 500억 달러 이상으로 성장하나, HBM 결합 고성능 GPU 구조는 병행 유지
- 사후 학습(post-training), 장시간 사고(long thinking) 등 고도화된 기법이 연산·전력 수요를 추가 확대
- 효율화와 고성능화가 동시에 요구되는 복합적 칩 전략이 불가피

▷ 경쟁의 본질 변화: 알고리즘에서 인프라로

- AI 경쟁의 핵심은 알고리즘 효율보다 연산 인프라, 전력 확보, 반도체 공급망을 감당할 수 있는 역량으로 이동
- 에너지·환경 부담, 보안·데이터 주권 이슈가 AI 전략의 핵심 변수로 부상
- 이에 따라 온프레미스·하이브리드 AI 인프라 도입이 확대되며, 2026년 시장 규모는 500억 달러 이상 전망

생성형 AI 컴퓨팅의 주 목적이 학습에서 추론으로 이동하고 있다. AI 연산 수요도 급증할 전망이다. 이로 인해 소비자 단말기 중심의 엣지 컴퓨팅이 확대되고 데이터센터 의존도는 감소할 것이라는 예측이 다수지만, 2026년 현재 상황은 다르게 전개되고 있다.

2026년에는 생성형 AI 컴퓨팅의 주 목적이 크게 바뀔 것으로 예상된다. 지금까지는 막대한 데이터를 학습시키는 훈련(training)이 AI 컴퓨팅의 주된 목적이었다. 하지만 앞으로는 기업과 소비자의 질문 및 프롬프트, 업무 요청에 대해 추론(inference)하고, 해석하고, 답변하는 기능이 AI 연산 수요의 핵심이 될 것이다. 상당수 전문가는 이러한 컴퓨팅 부하의 변화로 곧 추론에 최적화된 전용 칩의 필요성이 커질 것으로 본다. 이들 전문가들은 훨씬 저렴해진 칩이 엣지 디바이스에 배치되면서 결과적으로 대규모 데이터센터의 필요성이 줄거나 다른 작은 형태로 재편될 수 있으며, 비용도 감소할 수 있다고 예측한다.

그러나 딜로이트의 예측은 다소 다르다. 추론 워크로드는 2026년에 분명히 가장 빠르게 성장하는 영역으로, 전체 컴퓨팅의 약 3분의 2를 차지 할 것으로 예상된다(2023년 약 3분의 1, 2025년 절반 수준에서 증가).¹ 또한 2026년 추론 최적화 반도체 칩 시장 규모는 500억 달러를 넘을 전망이다. 그럼에도 불구하고, 딜로이트는 전체 연산의 대다수가 여전히 최첨단·고가·고전력 AI 칩(시장 규모 2,000억 달러 이상)에 의해 수행될 것으로 전망한다. 이러한 칩은 주로 4,000억 달러 이상 규모의 대형 데이터센터 또는 대형 데이터센터급 칩과 랙을 사용하는 500억 달

러 규모의 온프레미스(on-prem) 엔터프라이즈 AI 솔루션에 설치될 것이다. 엣지 디바이스용 소형 칩이 이를 대체할 가능성은 적다. 즉, 앞으로도 계속 증가하는 데이터센터와 엔터프라이즈 온프레미스 AI 팩토리에 대한 수요는 차고 넘칠 것이고, 이를 시설이 막대한 전력을 소비하게 될 것이다.

끊임없이 증가하는 AI의 연산 수요

신규 AI 모델을 위한 훈련용 컴퓨팅 수요 증가세는 2023~2024년에 비해 둔화됐지만,² AI 모델은 훈련 이후에도 성능을 향상시키는 고도화 기법을 통해 지속적인 재훈련이 필요하다. 이러한 재훈련과 더불어 막대한 규모의 추론 요청량이 더해지면서, 전체적인 컴퓨팅 수요는 감소하기는커녕 계속 증가할 가능성이 크다. 달리 말하면, 무어의 법칙(Moore's Law) 덕분에 연산에 사용되는 칩의 효율이 해마다 높아지고 있음에도 불구하고, 컴퓨팅 수요는 2030년까지 매년 4~5배 증가할 것으로 예상된다.³

1. 초기 학습을 위한 컴퓨팅 수요 증가세는 둔화

2020년 발표된 연구에 따르면, 더 많은 데이터로 학습하고 더 발전된 AI 가속기 칩을 사용한 더 큰 모델일수록 지속적으로 더욱 뛰어난 성능을 보여주는 것으로 나타났다. 이것이 생성형 AI의 첫 번째 스케일링 법칙(scaling law)이었다.⁴ 2022~2023년 사이 학습 모델은 10억 파라미터(parameter, 매개 변수)에서 1,000억을 넘어 1조 파라미터까지 성장했다.⁵

2024년에는 두 가지 문제가 부각되기 시작했다. 첫째, 세상에 존재하는 학습 데이터는 무한하지 않다는 것이다. 둘째, 모델을 계속 키울수록 성능 향상이 점점 줄어드는 한계효용 감소 현상이 나타났다는 점이다. 학습 데이터를 10배로 늘려도 최첨단 AI 모델이 기존 버전보다 약간 개선되는 데 그치거나, 경우에 따라서는 전혀 개선되지 않는 경우도 있었다.⁶ 동시에, 훨씬 적은 데이터, 더 짧은 시간, 더 적은 비용과 칩으로 만든 작고 효율적인 AI 모델로도 최첨단 기능을 구현할 수 있다는 가능성이 제기됐다.⁷

AI 모델의 학습 증가세가 둔화된다면, AI 컴퓨팅의 초점은 점차 추론으로 이동하게 된다. 대규모 언어 모델(LLM)에게 문서를 요약하도록 요청하는 작업은 가장 일상적 추론의 예시이지만, 모델을 학습시키는데 필요한 컴퓨팅의 극히 일부분만 사용한다. 그러나 수십억 명의 소비자와 근로자가 문서 요약 요청을 훨씬 더 많이 더 자주 하게 되면, 추론량이 계속 누적됨에 따라 전체 컴퓨팅 워크로드에서 추론이 차지하는 비중이 학습을 훨씬 능가하게 된다. 이러한 요청 중 일부는 소비자와 기업 사용자들이 사용하는 스마트폰이나 PC 등 단말기에서 처리될 수 있다. 실제로 딜로이트가 2024년에 정확히 예측한 대로, 2025년에는 온디바이스 AI 가속기 칩이 탑재된 PC와 스마트폰이 수억 대 판매됐다.⁸ 또한 추론은 학습보다 연산 강도가 낮기 때문에, 데이터센터 내부에서도 추론 최적화 전용 칩을 사용할 수 있다. 이러한 칩은 학습을 대규모로 확장하는데 필요한 초고성능 AI 칩보다 저렴하고, 추론당 전력 소모도 적으며, 고가의 결합형 메모리가 많이 필요하지 않을 수도 있다.⁹

2025년에 실제로 이러한 전망이 현실화됐고 2026년에도 이러한 추세가 지속될 가능성이 크다. 딜로이트가 2025년에 실시한 서베이에 따르면, 전 세계에서 생성형 AI를 사용하는 소비자가 증가하고 있으며 일일 사용률도 늘고 있다.¹⁰ PC와 스마트폰 등 엣지 디바이스에는 점점 더 강력한 온보드 AI 가속기가 탑재되고 있다. 다수의 추론 최적화 ASIC(application-specific integrated circuit, 특정 용도 및 기능을 수행하도록 맞춤 설계된 주문형 반도체)이 설계 및 제조되어, 데이터센터와 일부 엣지 디바이스에 배치되고 있다. 메타(Meta), 구글(Google), 아마존(Amazon), 인텔(Intel), AMD, 쿠얼(Qualcomm), 그록(Groq), 삼바노바(SambaNova), 세레브라스(Cerebras), 그래프코어(Graphcore) 외에도 여러 기업이 이러한 ASIC을 출시했다. 이 중 일부는 설계사가 프로세싱 코어를 제공하는 브로드컴(Broadcom) 패키지 솔루션을 기반으로 하고 있다.¹¹ 이들 칩의 구체적인 판매량이 모두 공개되지는 않았지만, 딜로이트는 2025년 해당 칩 시장 규모가 200억 달러를 넘고, 2026년에는 500억 달러를 돌파할 것으로 보고 있다.¹²

그렇다면 해당 3만 달러를 넘어 2028년에는 총합 4,000억 달러 규모에 이를 것으로 추정되는 고전력 칩이 여전히 필요한 이유는 무엇인가?¹³ 또한 2026년 한 해에만 4,000억 달러, 최대 1조 달러까지 소요될 수 있는 데이터센터가 여전히 필요한 이유는 무엇인가?¹⁴

2. 과거보다 훨씬 복잡해진 AI 모델 학습

첫 번째 스케일링 법칙의 핵심은 ‘더 나은’ AI 모델을 만드는 것이고, 적어도 몇 년 동안은 이 법칙이 큰 효과를 가져왔다. 초기 형태의 스케일링은 과거에는 단순히 학습(training)이라고 불렸지만 이제는 ‘사전 학습’(pre-training)으로 알려진 기초 모델 생성 단계를 뜻한다.

그러나 더 나은 모델을 만드는 방법은 두 가지가 더 있다. 첫째는 ‘사후 학습’(post-training) 스케일링으로, 파인튜닝(fine-tuning)*, 프루닝(pruning)**, 양자화(quantization)***, 지식 증류(distillation)****, 인간 피드백과 AI 피드백을 활용한 강화학습, 합성 데이터 증강 등 다양한 기법을 포함한다.¹⁵ 두 번째는 테스트 단계 스케일링 또는 장시간 사고(long thinking)로, 모델이 질문을 받은 이후 추론 과정에서 스스로 사고하도록 하는 방식이다. 여기에는 연쇄 사고(chain-of-thought) 프롬프트*****, 다수결 기반 샘플링(majority voting)*****+, 검색, 일부 사후 학습 기법까지 다양한 방법이 활용된다.¹⁶ 이러한 방식은 선택지를 늘리고 정보 출처를 개선함과 동시에 환각(hallucination)을 줄임으로써 정확도를 향상시킬 수 있다는 장점이 있다.¹⁷

* 파인튜닝(fine-tuning)은 사전에 학습된 모델(pretrained model)을 기반으로, 특정 목적·도메인·업무에 맞게 추가 학습시켜 성능을 정밀하게 조정하는 과정을 의미한다. 범용 모델의 추상적 지식을 실제로 쓰이는 정확한 결과로 전환하고, 특정 산업·기업 데이터를 반영하며, 톤·스타일·규칙·정책 일관성을 확보할 수 있다.

** 프루닝(pruning)은 학습된 모델에서 중요도가 낮은 파라미터(가중치·연결·뉴런)를 제거해 모델을 경량화하고 효율을 높이는 기법을 의미한다. 연산량·메모리·전력 소모를 줄이고, 추론 속도를 개선하고(특히 엣지·모바일 환경), 과적합 완화 및 일반화 성능을 개선하고, 배포 비용을 절감할 수 있다.

*** 양자화(quantization)는 모델의 파라미터나 연산에서 사용하는 수치 정밀도를 낮춰 경량화하는 과정이다. 모델 크기를 축소해 메모리 사용량을 줄이고, 연산을 단순화해 추론 속도를 개선하고, 엣지 및 모바일 환경에서 유리하기 때문에 전력 소모를 절감하고, 클라우드 추론 비용이 줄어 비용을 절감할 수 있다.

**** 지식 증류(distillation)는 크고 성능이 좋은 모델이 작고 효율적인 모델에게 지식을 압축해 전달하는 경량화·고도화 기법이다. 추론 비용과 지연시간을 감소해 모델을 경량화할 수 있고, 단순 축소 대비 정확도가 우수해 성능을 유지할 수 있으며, 엣지 및 온디바이스 환경에 배포할 수 있고, 대규모 서비스 비용을 절감할 수 있다.

***** 연쇄 사고(chain-of-thought) 프롬프트는 LLM이 문제 해결 과정에서 중간 추론 단계를 거쳐 최종 답변에 도달하도록 유도하는 프롬프트 기법을 의미한다. 단순히 정답만을 요구하는 것이 아니라, 문제 인식 → 단계별 분석 → 결론 도출의 사고 흐름을 거치도록 설계된 점이 주요 특징이다.

*****+ 다수결 기반 샘플링(majority voting)은 동일한 문제에 대해 복수의 후보 답변을 생성한 뒤, 가장 빈번하게 등장하는 결과를 최종 답으로 채택하는 방법론을 의미한다. 생성형 AI 맥락에서는 하나의 질문에 대해 모델이 여러 번 독립적으로 추론·생성한 결과 중 최다 득표 답변을 선택함으로써 정확성과 안정성을 제고하는 기법으로 활용된다.

3. 새로운 AI 기술, 효율성 개선 효과보다 전력 소비 부담이 더 큰 이유

첫째, 사후 학습 스케일링과 테스트 단계 스케일링이 새로운 표준 단계로 자리 잡아 전력 소비가 커지고 있다. 대부분의 AI 기업이 이제 이러한 방법을 활용해 다양한 방식으로 AI 모델을 개선하고 있다.¹⁸

둘째, 두 가지 종류의 스케일링 모두 막대한 컴퓨팅 자원을 필요로 한다. 사후 학습에 투입되는 전체 연산량은 원래의 기초 모델을 학습하는 데 필요한 컴퓨팅의 30배에 달하는 것으로 추정되며, 장시간 사고는 이메일 요약과 같은 단순 추론에 비해 100배 이상의 연산을 요구한다.¹⁹

셋째, 이러한 스케일링 기술이 널리 사용되며 컴퓨팅 자원 요구량이 급속도로 증가함에 따라, AI 데이터센터, 데이터센터의 입지와 전력 수요, 데이터센터에 탑재되는 AI 칩 및 기타 인프라, 이전 세대 AI 칩, 엣지 디바이스 등 여타 다양한 요인들도 큰 영향을 받고 있다.

2025년 이후 AI 데이터센터 변화에 대한 딜로이트의 과거 전망 되돌아보기

데이터센터는 수십 년 전에 등장했다. 실제로 전 세계 데이터센터가 차지하는 면적은 수천만 평방피트에 달하며, 매년 수백억 달러 규모의 반도체 부품이 이러한 데이터센터를 채우기 위해 판매되고 있다.²⁰ 그러나 차세대 AI 데이터센터와 이를 위한 신형 반도체는 기존 데이터센터 및 반도체와는 매우 다른 경우가 많다. 말 그대로 '밤과 낮' 만큼이나 다르다.

차세대 AI 데이터센터는 매년 수천억 달러의 건설 비용이 들고, 수백 기가와트(GW)의 전력을 소비할 것으로 전망된다. 이러한 시설은 대부분 이전 세대 데이터센터와 다른 냉각 방식, 전력 공급 및 전압, 내부 통신 기술이 필요하다. 서버랙의 밀도와 무게가 훨씬 증가하기 때문에 바닥 자체도 더 두껍게 만들어야 한다. 아마도 가장 중요한 변화는 기존 데이터센터가 중앙처리장치(CPU)를 중심으로 하고 그 주변에 메모리를 배치하는 구조였던 반면, 최신 AI 서버랙은 그래픽처리장치(GPU)라 불리는 특수 칩으로 구성된다는 점이다.²¹ 이 GPU에는 고대역폭 메모리(HBM)

가 밀착 통합돼 있으며, 방대한 AI 연산을 조율 및 관리하기 위한 특수 설계 CPU도 함께 필요하다. 이러한 차세대 AI 데이터센터를 위해 설계된 부품은 대부분 규모와 성능의 특수성이 과거의 부품과 완전히 다르다.²²

불과 2006년까지만 해도 고급 GPU는 게임용 컴퓨터나 콘솔에 쓰는 부품으로 여겨졌으며, 데이터센터와는 무관한 것으로 간주됐다.²³ 대부분의 데이터센터 업무는 순차적으로 작업을 처리하는 직렬 기반 CPU로도 충분히 감당 가능했다. 일부 고성능 컴퓨터, 즉 슈퍼컴퓨터에는 대량 병렬 처리기(massively parallel processor)라 불리는 특수 칩이 탑재돼 수백 개의 작업을 동시에 실행할 수 있었다. 그러나 이 칩은 게임용 GPU나 데이터센터용 CPU에 비해 수십~수백 배 이상 비쌌다.

2009년 과학자들은 게임용 고급 GPU도 병렬 프로세서라는 점에 주목하고, 동일한 고급 GPU로 머신러닝 모델을 돌려보기 시작했다.²⁴ 결과는 성공적이었다. 몇 년 뒤에는 게임용 GPU와는 약간 다르지만 최적화된 전용 GPU가 데이터센터와 온프레미스 장비에서 머신러닝 AI를 수행하는 데 쓰이기 시작했다.²⁵ 그러나 해당 시장 규모는 2018년까지만 해도 연간 수십억 달러 수준에 머물렀다.²⁶

2022년에 들어 생성형 AI를 위한 LLM 개발이 활성화되면서, GPU는 더욱 고도화된 형태로 진화해야 했고, 상대적으로 새로운 형태의 특수 메모리인 HBM과 하나의 패키지로 통합되는 형태가 필요해졌다.²⁷ GPU와 HBM을 통합한 형태에는 방대한 데이터 흐름을 조율할 장치도 필요했다. 또한 기존 컴퓨터·스마트폰·데이터센터용 CPU와는 다르지만 핵심 아키텍처는 유사한 특수 설계 CPU가 생성형 AI 데이터센터의 주요 부품으로 추가됐다. 여기에 다른 여러 핵심 부품까지 결합됐다. 2025년 기준 전 세계 상위 500대 슈퍼컴퓨터의 거의 대부분은 GPU-특수메모리-CPU 조합이 공통적으로 탑재돼 있다.²⁸ 따라서 현재 구축되고 있는 메가스케일 AI 데이터센터는 특수화된 슈퍼컴퓨터의 변형 모델이라고도 볼 수 있다.

컴퓨팅 수요 증가가 AI 생태계에 가져올 변화

이제 기업은 사후 학습과 테스트 단계 스케일링이 지속되면서 대규모 데이터센터 및 엔터프라이즈 AI 팩토리의 컴퓨팅 수요가 계속 증가하는 미래에 대비해야 한다. 추론 최적화 칩과 엣지 프로세스의 성장은 계속되겠지만, 그럼에도 불구하고 하이퍼스케일 데이터센터와 엔터프라이즈 AI 장비에 대한 투자는 여전히 필요하다. ‘추론 최적화’ 칩이라 해서 반드시 전력 소모 감소 효과가 있는 것도 아니다. 최근 출시된 한 제품은 추론 프리필(pre-fill) 연산* 최적화를 위해 HBM을 사용하지 않고 그래픽 D램 GDDR7(graphics double data rate 7)을 사용하지만, 랙 하나당 370kW의 전력 밀도가 필요하다. 이는 동일 공급업체의 훈련용 모델 대비 거의 세 배에 달하는 수준이다.²⁹

* 추론 프리필(pre-fill) 연산은 LLM이 실제 토큰 생성을 시작하기 전에 입력 프롬프트 전체를 한 번에 처리하여 내부 상태를 구축하는 초기 추론 단계를 의미한다. 겉으로 드러나지 않지만, 생성형 AI 서비스의 응답 속도, 비용 구조, 확장성을 좌우하는 핵심 기반 기술이다.

1. AI 데이터센터의 자본지출

2026년 전 세계 AI 데이터센터 자본지출(capex)은 4,000억~4,500억 달러로 예상되며,³⁰ 이 중 절반 이상(2,500억~3,000억 달러)은 장비 내부 칩 구매 비용이고,³¹ 나머지는 토지, 건설, 전력 인프라, 인허가 등 기타 비용이 차지한다. 2028년에 이르면 AI 데이터센터 자본지출은 1조 달러까지 증가할 것으로 전망되며,³² 이 중 AI 칩만 4,000억 달러 이상을

차지할 것으로 예측된다.³³ 사전 학습 성장세가 둔화되고 컴퓨팅의 비중이 훈련에서 추론으로 이동하고 있음에도, 사후 학습과 테스트 스케일링, 사용량 증가에 따른 컴퓨팅 수요를 감안하면, 전 세계는 여전히 방대한 데이터센터를 필요로 한다. 따라서 2025년 3,000억~4,000억 달러 규모였던 AI 데이터센터 자본지출이 2028년 1조 달러 규모로 증가한다는 전망은 실현될 가능성이 매우 크다.

2. AI 데이터센터의 입지

100조 파라미터 규모의 LLM 사전 학습은 수주에 걸친 시간 투자가 필요한 데다 잠시라도 중단되면 문제가 커진다. 핵심 부품 고장이나 프로세서간 지연이 발생하면 모든 작업이 무효화돼 다시 시작해야 할 수도 있다. 이 때문에 대부분의 기초 모델 사전 학습은 단일 건물 또는 단일 캠퍼스 내에서 서버와 랙을 공동 배치(co-location)하는 방식으로 진행돼 왔다. 그러나 점점 더 많은 AI 연산을 미국 전역 또는 전 세계 여러 데이터센터에서 분산 처리할 수 있게 됐다.³⁴ 게다가 완전한 사전 학습이 아닌 최종 추론을 수행하는 소규모 추론 데이터센터는 지연을 줄이기 위해 대도시 인근에 위치하게 될 가능성이 크다. 이는 각국이 자국 내에 AI 연산 역량을 확보하려는 소버린 AI(sovereign AI)에 대한 니즈를 높이며, 엔터프라이즈 온프레미스 기반 하이브리드 클라우드 확산을 촉진하는 요인으로 작용하고 있다.³⁵

3. AI 데이터센터의 전력 수요

사전 및 사후 학습과 테스트 등 세 단계 스케일링을 수행하는 AI 데이터센

터는 지속적으로 막대한 전력이 필요하다. 다만 사후 학습과 테스트 단계 스케일링은 사전 학습과 달리 비교적 중단이 용이하다는 특징이 있다. 사전 학습은 하나의 긴 작업으로 수행돼야 하지만, 사후 학습 및 테스트 단계는 부하 이전(load shifting)이 쉽기 때문에 AI 기업은 전력 수요 반응 프로그램에 참여할 수 있다. 즉, 특정 시간대에 데이터센터를 다른 지역으로 전환하거나 CPU/GPU의 클럭 속도(clock speed, 프로세서가 초당 연산 사이클을 수행하는 횟수)를 낮춰 피크 시간대 전력 수요를 줄일 수 있다.³⁶ 이처럼 전력 부하를 유연하게 운영하면 전력망 안정성과 비용 절감을 동시에 달성할 수 있다.³⁷

또한 AI 학습 및 추론 작업을 분산할 수 있기 때문에, 데이터센터가 특정 지역에 몰리지 않고 전 세계적으로 보다 고르게 분산될 수 있다. 이는 전력 수요를 분산하는 데에도 도움이 된다.

4. AI 데이터센터의 반도체칩

일각에서는 AI 칩 시장을 ‘제로섬 게임’으로 보는 시각도 있다. 즉, 모델 사전 학습에는 HBM 결합 고성능 GPU에 수만 달러를 지출해야 했지만, 컴퓨팅의 비중이 추론으로 이동하면 HBM이 적게 들어가 더 저렴한 추론 최적화 칩으로 대체될 것이라는 분석이다.

그러나 실제로는 대체가 아니라 병행의 구도가 될 가능성이 크다. 추론 전용 또는 추론 최적화 칩 시장은 상당한 성장을 보이겠지만, 동시에 기초 모델 사전 학습, 사후 학습, 테스트 단계 스케일링 등 훈련과 추론이 혼

합된 연산에 가장 적합한 칩은 여전히 HBM이 탑재된 대형·고성능·고전력 GPU이다. 이들 칩의 가격은 개당 수만 달러에 이른다. 게다가 첨단 공정 웨이퍼 가격이 2026년에 50% 이상 상승할 것으로 예상됨에 따라, 이러한 칩을 구매하는 기업 입장에서는 비용이 더 증가할 수 있다.³⁸

5. 스마트폰·PC 등 소비자 및 기업용 디바이스의 엣지 AI

앞서 언급한 바와 같이, 신경망 처리장치(NPU)를 탑재한 수억 대의 스마트폰과 PC가 출하 및 판매되고 있다.³⁹ NPU는 전용 칩이나 CPU 칩의 일부로 탑재되어 AI 추론 작업을 적정 전력 소모 수준에서 처리하도록 최적화된 장치로, 가격은 몇 달러에서 수십 달러 수준이다.

그러나 NPU는 앞서 설명한 단일 추론(예: 이메일 내용을 요약해줘) 정도를 처리하는 데에는 적합하지만, 그 이상의 고도화된 작업을 수행하기는 어렵다. 따라서 딜로이트는 2026년에 수행되는 거의 모든 AI 연산이 조만간 완공될 초대형 AI 데이터센터 또는 기업이 보유한 고가의 고성능 AI 서버에서 처리될 것으로 전망한다. AI 연산이 PC나 스마트폰에서 주로 처리되지는 않을 것이라는 의미다. 적어도 선점 경쟁이 치열한 초고성장기인 지금은 하이브리드 아키텍처 기반의 비용 최적화 전략이 공급업체나 기업의 우선순위가 될 수 없을 것으로 보인다. 또한 테스트 단계 스케일링과 같은 고성능 기술은 대부분의 소비자용 사용 사례뿐 아니라 상당수의 엔터프라이즈 온디바이스 활용 사례에 적용하기에는 과도한 방식이다. 향후 어느 시점에는 컴퓨터와 스마트폰이 AI 연산에서 더 큰 비중을 차지할 수 있지만, 2026년은 아직 이른 시기라 볼 수 있다.

최근에는 한 AI 기업이 추론 능력을 갖추고 PC에서 로컬로 실행 가능한 생성형 AI 모델을 발표했다. 하지만 실제 성능이 어느 정도인지, 배터리 수명에 어떤 영향을 미치는지, 얼마나 많은 PC 사용자가 클라우드 대신 로컬 실행을 선호할지는 아직 불확실하다.⁴⁰

6. 온프레미스 기반 엔터프라이즈 엣지 AI

전 세계 초대형 AI 데이터센터에 들어가는 고성능·고전력 GPU 및 HBM과 이를 조율하는 특수 CPU 트레이이는 사후 학습을 포함한 생성형 AI 컴퓨팅을 온프레미스 기반 하이브리드 형태로 강화해 운영할 때에도 사용할 수 있다. 기업은 비용, 지식재산(IP) 보호, 주권성, 복원력, 맞춤화 요구 등을 고려해 다음과 같은 선택을 할 수 있다.

- 약 8개의 GPU(HBM·CPU 포함)가 탑재된 박스 한 대를 30만~50만 달러에 구매해 일정 수준의 AI 학습 및 추론 수행⁴¹
- 최대 72개의 최첨단 GPU(HBM·CPU 포함)가 장착된 랙을 300만~500만 달러에 도입해 더 높은 성능 확보⁴²
- 필요할 경우 수천만 달러 규모로 여러 랙을 동시에 구축해 대 규모 연산 처리⁴³

딜로이트는 2026년 이러한 온프레미스 기반 하이브리드 엔터프라이즈 AI 시장 규모가 500억 달러를 넘을 것으로 전망한다.

7. 로봇·드론·자율주행차 등에 탑재되는 엣지 AI

2026년 기준 관련 시장은 아직 상대적으로 규모가 작지만, 실시간(on-device) 추론이 반드시 필요한 여러 사용 사례가 있다. 해당 사례는 드론과 로봇부터 자율주행차에 이르기까지 다양하기 때문에 현재 사용되는 칩 또한 매우 다양하다. 드론은 대부분 상대적으로 단순한 저전력 AI 추론 칩을 탑재하고 있으며,⁴⁴ 자율주행차는 대부분 데이터센터에 사용되는 GPU보다 약간 낮은 수준이지만 여전히 고성능 GPU 기반 칩 솔루션이 탑재된다.⁴⁵ 이러한 비(非)공장 기반 AI 시장 규모는 2026년에도 50억 달러를 넘지 못할 것으로 보이지만,⁴⁶ 로봇 시장이 본격적으로 성장한다면 2030년 이후 크게 확대될 가능성이 있다.⁴⁷

인류의 AI 역사는 아직 초기 단계이다. 알고리즘 효율을 높이려는 지속적인 시도에도 불구하고, 2025년 여름 기준 AI 연산 수요와 사전 학습, 사후 학습, 테스트 단계 스케일링, 추론을 수행하기 위한 데이터센터, 엔터프라이즈 온프레미스 솔루션, 고성능 AI 칩에 대한 수요는 매우 높은 성장세를 보였다.⁴⁸ 향후 어느 시점에는 새로운 기술적 돌파구가 등장해, 개선된 AI 모델이 더 저렴한 칩에서도 원활하게 작동하고, 더 적은 데이터센터와 더 적은 전력으로도 동일한 성능을 낼 것이다. 하지만 아직 AI 초기에 해당하는 2026년에는 급증하는 AI 컴퓨팅 수요에 철저히 대비해야 한다.

Korean Perspectives

반도체와 데이터 센터 경쟁의 본질이 된다

생

성형 AI 경쟁의 핵심은 더 이상 어떤 LLM을 만들었는가가 아니다. 이제 승부를 가르는 기준은 AI 연산을 얼마나 안정적으로, 얼마나 많이 처리할 수 있느냐, 즉 컴퓨팅 인프라 역량으로 이동하고 있다. AI 연산의 중심이 모델 학습에서 실제 서비스 과정의 추론으로 옮겨가고 있지만, 사후 학습과 테스트 단계에서의 연산 증가, 그리고 이용자 수 급증으로 전체 연산 수요는 줄어들지 않고 오히려 계속 확대되고 있다. 이는 생성형 AI가 일시적 유행이 아니라, AI 반도체와 데이터센터 투자가 장기간 이어질 구조적 성장 국면에 들어섰다는 뜻이다.

추론 최적화 칩의 부상은 고성능 GPU 중심 구조를 대체하지 않는다. 오히려 사후 학습과 장시간 사고(long thinking)가 새로운 표준으로 자리 잡으면서, HBM이 결합된 고성능·고전력 GPU와 추론 전용 칩이 병행되는 이중 구조가 고착되고 있다. 이 환경에서 경쟁의 핵심은 단일 칩의 성능이 아니라, 워크로드별로 최적화된 칩 포트폴리오를 안정적으로 공급할 수 있는 역량이다.

이 지점에서 SK하이닉스의 전략적 위상은 독보적이다. 현재 글로벌 AI 연산 구조에서 가장 치명적인 병목은 연산 로직이 아니라 메모리 대역폭과 전력 효율, 즉 HBM이다. SK하이닉스는 HBM3·HBM3E

세대에서 기술, 수율, 양산 타이밍을 동시에 선점하며, 글로벌 AI 가속기 로드맵의 핵심 병목을 사실상 통제하고 있다.

오늘날 GPU 출하량과 AI 데이터센터 확장 속도는 곧 SK하이닉스의 HBM 공급 안정성에 의해 좌우되는 구조다. 이는 SK하이닉스가 단순 메모리 공급자가 아니라, AI 연산 구조와 속도를 규정하는 기준점이 되었음을 의미한다.

삼성전자 역시 이 경쟁에서 중요한 축을 형성하고 있다. 삼성전자는 HBM을 포함한 메모리 역량에 더해 파운드리와 첨단 패키징을 결합한 수직 통합 전략을 통해, AI 반도체 경쟁을 단품이 아닌 시스템 레벨 경쟁으로 확장하고 있다. 여기에 자체 GPU 개발 시도와 함께, 리밸리언스 등 AI 추론 특화 칩을 설계하는 국내 펩리스 생태계가 등장하면서 한국은 메모리·패키징·연산 보조 영역에서 대체 불가능한 국가로 자리 잡고 있다.

데이터센터 역시 비용 절감의 대상이 아니라 AI 경쟁의 인프라 전장으로 재정의되고 있다. 초대형 AI 데이터센터와 엔터프라이즈 온프레미스 AI 팩토리는 전력, 냉각, 랙 밀도, 입지 전략까지 포함한 국가 단위 산업 이슈로 확대되고 있으며, 이 과정에서 반도체 기업과 데이터센터 기업의 경계는 빠르게 흐려지고 있다. 특히 HBM을 중심으로 한 고집적 AI 서버 구조는 반도체, 서버, 데이터센터, 전력 인프라를 하나의 패키지로 묶어 경쟁하게 만든다.

결국 한국 반도체 산업은 AI 시대에 단순한 부품 공급자를 넘어섰다. SK하이닉스는 AI 연산 병목을 통제하는 필수 공급자로서, 삼성전자는 시스템 확장 플레이어로서, 한국 반도체 생태계는 이미 AI 연산 인프라의 전략적 요충지를 선점하고 있다.

앞으로의 경쟁은 이 우위를 어떻게 시스템, 데이터센터, 전력 인프라 까지 확장해 나가느냐에 달려 있다. 승부는 누가 더 뛰어난 모델을 만드느냐가 아니라, 누가 끝까지 연산을 안정적으로 감당할 수 있는 구조를 구축하느냐에서 갈릴 것이다.



최호개 파트너
한국 딜로이트 그룹
TMT Industry 리더

Part

02

기술 공급망과 주권의 재편,
새로운 경쟁 전선 형성

6장 AI 칩을 둘러싼 무역 통제와 글로벌 반도체 공급망 재편

7장 기술 주권 확보 경쟁

6장

AI 칩을 둘러싼 무역 통제와 글로벌 반도체 공급망 재편

인공지능(AI) 칩을 둘러싼 무역 통제는 반도체 공급망을 구조적으로 재편하고 있으며, 2026년 이후 경쟁의 핵심은 기술 혁신 자체보다 ‘누가 설계·제조·패키징·운영을 통제하느냐’로 이동할 것이다.

핵심 내용 요약 (Executive Summary)

▷ AI 칩을 둘러싼 통제·무역장벽 강화

- 반도체는 단순 시장 재화에서 국가 안보·경제·과학 경쟁력의 전략 자산으로 전환
- 지정학적 긴장과 무역 장벽 강화로 공급망이 효율 중심에서 통제·회복력 중심으로 재편
- AI 칩 혁신이 세계 경제, 국가 안보, 과학 발전에 연쇄적 영향

▷ 수출통제로 재편되는 반도체 생태계

- EDA, PDK, 파운드리 IP, AI 모델 웨이트까지 통제 범위가 설계 단계로 확장
- EUV, 정밀 식각, 광학 부품, 특수ガ스, 핵심 광물 등 새로운 규제 대상 등장
- 첨단 패키징(HBM, 칩렛, 이종 아키텍처)이 지정학적 병목 지점으로 부상

▷ 글로벌 공급망이 단일 최적화 구조에서 블록화된 다중 체계로 전환

- (미국·유럽·동맹국)첨단 노드 리쇼어링 및 프렌드쇼어링 가속
- (중국)첨단 공정 제약 속 DUV·우회 기술·국산 스택 강화
- (인도·동남아) 조립·테스트·패키징 대체 허브로 부상

▷ 핵심 시사점과 기업 전략

- 완전 자립은 불가능, 기술 주권 경쟁의 본질은 AI 인프라 통제력 확보
- 공급망 회복력은 기술·소프트웨어·정책을 아우르는 종합 역량으로 재정의
- 기업은 전공정-후공정-설계-소프트웨어 전 구간의 스트레스 테스트 필요
- 대체 조달, 지역 분산, 규제 대응, 신뢰 가능한 국가 간 협력 채널 확보 필수

차세대 AI 핵심 칩 기술에 대한 무역 장벽이 높아지고 있다. 반도체 공급망 회복력 강화를 위한 신속한 대응이 필요한 시점이다.

지정학적 긴장과 무역 장벽 강화로 반도체 공급망이 재편되고 있다. 이는 AI 칩 혁신부터 세계 경제, 국가 안보, 과학 발전 등 전방위로 큰 영향을 미치고 있다. AI 칩 생산을 위한 첨단 공정과 소재는 특정 지역에 몰려 있는 소수 공급업체에 크게 의존하는데, 각국 정부는 이들의 지배력에 대응해 자국의 이익을 보호하고 의존도를 낮추기 위한 무역 장벽을 세우고 있다. 지금까지도 차세대 AI 시스템과 고성능 컴퓨팅 데이터 센터를 위한 세계 최고 수준의 칩을 생산하기 위해서는 불안정한 공급망 여건을 타개해야만 했다. 지금은 리스크가 더 커졌다.

딜로이트는 식각(etching, 회로 형성 시 불필요한 물질을 제거하는 공정)과 게이트올어라운드(GAA, 반도체 트랜지스터에서 게이트가 채널을 사방에서 둘러싸는 구조) 트랜지스터 같은 전공정 및 후공정 칩 제조, 전자설계자동화(EDA), 첨단 AI 모델 운영에 필요한 소프트웨어 도구 등 반도체 기술이 2026년 반도체 공급망을 한층 옥죄는 병목 이슈가 될 것으로 예상한다. 또한 극자외선(EUV) 노광 장비와 고대역폭 메모리(HBM) 코파키징 도구 등 다양한 핵심 기술에 최소 300억 달러가 투입될 것으로 예상되지만, 이들 기술 역시 무역 장벽의 영향을 받을 것으로 전망된다.¹ 그러나 이러한 기술을 기반으로 만들어지는 AI 칩 시장 규모는 무려 3,000억 달러에 달할 것으로 예상된다. AI 칩 기술이 글로벌 반도체 공급망의 핵으로 작용한다는 의미다.²

AI, 글로벌 반도체 공급망 재편

딜로이트의 AI 데이터센터 관련 반도체 콘텐츠 분석에 따르면, 상호의 존성이 매우 높은 글로벌 반도체 공급망 속에서 각국은 생성형 AI, 고성능 컴퓨팅, 자율 시스템에 필수적인 AI 칩과 하드웨어 부품을 확보하기 위해 경쟁적으로 움직이고 있다.³ 따라서 2025~2026년에는 2~3년 전과 비교해 수출 통제와 기타 무역 제한이 반도체 장비, 소재, 소프트웨어, 설계 도구, 다양한 종류의 칩, 패키징 및 조립 장비 등에 훨씬 중대한 영향을 미칠 것으로 전망된다(그림 1).



그림 1
미국과 유럽, 2025~2026년 다양한 반도체 기술에 대한 무역통제 확대

2019년	EUV 노광 장비
2023년 3월	첨단 증착 장비 / 침지식(immersion) ¹ 심자외선(DUV) 노광장비
2024년 4월	EUV 마스크 / EUV 펠리클(pellicle) ² / EUV 마스크·펠리클을 제조용 장비 및 구성품
2024년 9월	<ul style="list-style-type: none"> · GAAFET / 양자컴퓨팅 / 적층(additive) 제조 / 추가 첨단 반도체 소자에 관련된 장비·부품·소프트웨어
2024년 12월	<ul style="list-style-type: none"> · 첨단 노드 칩 생산용 반도체 제조 장비 24종(식각, 증착, 노광, 이온 주입, 어닐링(annealing)³, 계측·검사, 세정 장비 등 포함)* · 첨단 노드에서 칩을 개발하거나 생산할 때 필요한 소프트웨어 도구 3종* · HBM
2025년 1월	<ul style="list-style-type: none"> · 특정 첨단 폐쇄형 이중용도 AI 모델의 모델 웨이트(model weight)⁴ · 첨단 컴퓨팅 IC를 중심으로 파운드리와 OSAT가 준수해야 하는 추가 실사 및 공시 의무
2025년 5월	· 첨단 AI 칩을 사용, 라이선스 판매, 거래하는 기업에 대한 강화된 글로벌 실사 요건
수출통제로 영향을 받을 수 있는 추가 영역	
2026년 ~	<ul style="list-style-type: none"> · AI 칩과 관련된 파운드리 IP 라이브러리, 공정 설계 키트(PDK), 성능 테스트 산출물 · 제품 검증 및 모델 파인튜닝에 사용되는 평가용 하드웨어 · 첨단 식각 장비 및 관련 광학 도구 · 칩 장비 구성품과 공정 장비의 적격성 평가 및 설치에 소요되는 시간 증가 · 칩렛(chiplet)⁵ 기반 코퍼키징, 사이트간 배치, 설계 도구 전반에 대한 추가 모니터링

참조: 2025년 자료는 2025년 10월 8일 기준으로 취합

1. 침지식(immersion): 반도체 노광 공정에서 액체를 사용해 해상도를 높이는 미세 공정 기술

2. EUV 펠리클(pellicle): 포토마스크의 오염을 방지하는 초박막 보호막

3. 어닐링(annealing): 가열·냉각을 통해 반도체 특성을 개선하는 열처리 공정

4. 모델 웨이트(model weight): 내부 저장한 파라미터 값

5. 칩렛(chiplet): 이종칩·기능 모듈을 단일 패키지에 결합

출처: Deloitte analysis. Data for 2019 to 2025 based on information gathered from publicly available sources including documents and announcements published on the sites of Federal Register and Bureau of Industry and Security (BIS). 2026 information based on conversations and forward-looking insights gathered from industry subject matter specialists. *Bureau of Industry and Security. *Commerce strengthens export controls to restrict China's capability to produce advanced semiconductors for military applications, *U.S. Department of Commerce, December 2, 2024.

AI 시스템의 성능은 첨단 AI 로직 설계, 최첨단 전공정 노드 제조, 고도화된 패키징 등 전 세계에 분산된 몇 가지 제한적 기술 스택에 좌우된다. 이러한 역량을 구현하기 위해서는 통합기기제조업체(IDM), 파운드리, 장비 제조사, 설계 업체, 반도체 후공정(OSAT) 업체, 시스템 통합사, 외부 유통 파트너뿐 아니라 여러 국가의 정부 기관 등 다양한 이해관계자간 협력이 필요하다.⁴

수출통제, 첨단 AI 로직 설계의 미래 재정립

미국은 2024~2025년 여러 핵심 반도체 기술, 특히 EDA 도구에 대한 규제를 강화했다가 다시 완화하는 움직임을 보였다.⁵ 첨단 AI 가속기 개발에 필수인 EDA 공정은 설계 로직, 칩 레이아웃과 배치, 시뮬레이션, AI 기반 설계, 검증, 통합 워크플로 등으로 구성된다.

예를 들어, 기존에는 게이트올어라운드 FET(GAA field-effect transistor, 채널을 게이트가 360도 둘러싸는 구조의 차세대 트랜지스터) 기반으로 개발된 칩에 무역제한이 있었다.⁶ GAAFET은 5나노미터(nm) 이하, 3nm 이하로직 설계를 위한 차세대 트랜지스터 구조로, 연산 집약적 생성형 AI 워크로드에 필요한 성능 및 전력 효율을 개선하는 역할을 한다. 2024년 12월, 미국은 첨단 컴퓨팅 노드의 개발과 설계를 지원하는 소프트웨어와 도구까지 포함하도록 수출통제 범위를 확대했다.⁷ 이에 따라 2026년에는 더 넓은 EDA 생태계와 파운드리 파트너에도 영향을 미칠 가능성이 있다.

2026년 이후의 전망

GAAFET 기반 칩에 대한 규제가 강화되면, 주요 노드에서 GAAFET 공정 설계 키트(PDK)를 사용하는 미국 비동맹국의 파운드리는 검증을 위해 EDA 도구가 필요하게 된다. 그러나 이러한 도구를 확보하지 못하는 특정 지역의 파운드리는 비효율적인 구식 노드를 적용하거나, 어쩔 수 없이 국내 EDA 역량을 개발해야 하는 처지로 내몰릴 수 있다. 이는 제품 개발 주기를 늘리고 경쟁력을 약화시킬 가능성이 크다. 더불어, 첨단 컴퓨팅 칩에 대한 추가 규제와 AI 모델 웨이트에 대한 새로운 규제로 인해 특히 중국 고객 및 비즈니스 파트너와 협업하는 기업들이 준수해야 할 요건이 한층 엄격해졌다.⁸ AI 모델과 해당 웨이트의 규모 및 품질은 칩 설계에 활용되는 AI 기반 EDA 도구의 성능에 점점 더 중대한 영향을 미치고 있다.⁹

딜로이트는 2026년까지 EDA 및 로직 설계업체들이 이러한 규제의 영향을 받을 것으로 전망한다. 이들은 파운드리의 지식재산(IP) 라이브러리, PDK, AI 가속기와 연관된 성능 테스트 결과와 관련해, 적용 대상 기업·지역·사용 목적에 대한 더 강화된 검증과 세부적인 정보 공개 의무를 준수해야 할 것이다. 제품 검증과 모델 파인튜닝을 위해 일반적으로 사용되는 평가용 하드웨어는 보다 엄격한 심사를 받을 가능성이 있다.¹⁰ 따라서 AI 하드웨어 공동 설계에 참여하는 기업들은 신뢰할 수 있는 국가간 협력 경로를 마련하거나, 워크플로를 재구축해야 할 수 있다. 예컨대, 모델 웨이트는 미국 또는 동맹국의 안전한 IT 인프라 내에 보관되며, 파운드리 파트너가 원격으로 테스트를 수행하도록 허용하는 방식이 가능하다.¹¹

AI 시스템용 최첨단 전공정 노드 제조에서 발생하는 병목

미국과 네덜란드는 최첨단 공정 노드를 생산하는 데 필수적인 것으로 간주되는 EUV 장비에 대한 수출제한을 지속하고 있다.¹² 미국은 EUV 장비를 국내에서 생산하지는 않지만, 기술 및 국가 안보를 위해 네덜란드 등 동맹국과의 수출통제 공조를 통해 어떤 국가가 이 장비를 구매할 수 있는지에 막대한 영향력을 행사하고 있다. 한편 중국은 자국 반도체업체들을 지원해 멀티 패터닝(multiple patterning)* 기법을 적용한 심외선(DUV) 기술을 커스터마이징함으로써 노광 장비 개발을 추진해왔다.¹³ 이러한 방식은 효과적일 수 있으나, 속도가 훨씬 느리고 비용이 더 많이 드는 것으로 평가된다.¹⁴ 미국은 국가 안보를 위해 정교한 AI 아키텍처를 구현하는 데 필수인 정밀 식각용 도구에 대해서도 추가 수출 규제를 가하고 있다.¹⁵

* 멀티 패터닝(multiple patterning)은 단일 노광 공정으로는 구현이 어려운 초미세 선폭을 확보하기 위해 동일 레이어를 여러 번 분할하여 노광·식각하는 반도체 미세화 기법을 의미한다. 광학 해상도의 물리적 한계를 공정 반복과 패턴 분해를 통해 우회하는 방식으로, EUV 전후 핵심적인 미세화 수단으로 활용돼 왔다.

2026년 이후의 전망

첨단 식각 기술은 5nm 이하 노드에서 최첨단 AI 칩을 제조하는 핵심 기술이다. 칩 산업은 가장 최신 AI 칩의 미세 구조를 구현하기 위해 더블(double)·쿼드러플(quadruple)·스페이서(spacer) 패터닝*을 활용한다.¹⁶ 이러한 이유로, 미국에서 개발된 식각 공정 장비뿐 아니라 미국

의 식각 기술 IP를 기반으로 해외에서 설계 및 제조된 장비와 도구 역시 2026년에 새로운 병목 요인으로 부상할 가능성이 있다. 또한 웨이퍼 제조 장비의 핵심 구성품이자 웨이퍼에 인쇄될 패턴의 청사진을 담고 있는 광학 부품(렌즈, 미러 등)과 레티클(포토마스크 등)과 같은 구성 요소도 규제 대상이 될 수 있다.

* 더블(double)-쿼드러플(quadruple)-스페이서(spacer) 패터닝은 한 번의 노광으로는 구현할 수 없는 미세 선풍을 만들기 위해, 동일한 층을 두 번, 네 번에 나누거나 측면 복제로 패터닝하는 공정 기법이다. 노광 해상도의 물리적 한계를 공정 반복과 자기정렬 구조를 통해 극복하는 미세화 핵심 기술군이다. 이들 기법은 모두 멀티 패터닝의 하위 개념이지만, 정렬 방식과 공정 안정성 측면에서 구조적 차이가 있다.

더불어, 첨단 공정의 필수 소재인 실란(silane)이나 불소 계열 파생물 등 특수가스,¹⁷ 갈륨(gallium), 게르마늄(germanium), 안티모니(antimony) 등 핵심 광물¹⁸이 글로벌 반도체 공급망의 또 다른 마찰 지점이 되고 있다.

전공정 장비, 부품, 소재 전반이 수출 통제의 영향을 받으면서, 2026년부터 수 년간 미국, 대만, 한국에서 5nm 이하 및 3nm 이하 제품의 생산이 계속 확대될 것으로 전망된다. 반면, 중국은 멀티 패터닝 우회 기술을 적용한 성숙 단계의 DUV 기술에 계속 주력할 것으로 예상된다.

이에 따라 글로벌 반도체 장비 기업들은 지역 단위로 전공정 웨이퍼 제조 관련 설비투자 계획을 조정해야 한다. 장비 제조사, 부품 공급업체, 파

운드리는 2024~2025년에 비해 자격 인증, 업그레이드, 설치 주기가 길어질 수 있다. 또한 칩 설계 기업들이 중국 등 미국 비동맹 국가에서 증가하는 생성형 AI 칩 수요를 충족하기 위해 디피처링(defeaturing, 사양 제한) 또는 단계적 다운그레이드 버전의 AI XPU(X processing unit, 성능 AI 칩의 성능을 축소한 통합형 AI 가속기)나 지역 중심의 공정 라이브러리를 개발함에 따라, 전공정 장비 업체의 지원도 강화될 필요가 있다.

무역통제, 첨단 패키징과 테스트 공정 생태계 재편

첨단 패키징 기술에 대한 전략적 수출 통제 또한 빠르게 확산되고 있다. 네덜란드는 고밀도 칩 적층에 핵심적인 역할을 하며 현재와 미래의 생성형 AI 칩을 구성하는 필수 요소로 간주되는 계측 및 검사 장비에 대해 수출제한을 적용하고 있다.^{19,20,21} 또한 측정 및 검사를 포함한 특정 유형의 칩 장비(식각, 증착, 노광, 이온 주입, 어닐링, 계측·검사, 세정 장비)는 첨단 AI 칩을 테스트하고 검증하는 데 필수적이기 때문에 주요국의 수출규제 대상에 포함돼 있는 경우가 많다.²² 해당 기술은 민간과 군용으로 모두 활용될 수 있는 민감한 이중 용도 기술로 간주되기 때문이다. 따라서 이러한 기술은 향후에도 추가 무역통제를 받을 수 있다.

2026년 이후의 전망

딜로이트가 ‘2024 첨단기술·미디어·통신(TMT) 전망’에서 강조했듯, 칩 렛과 이종 아키텍처(heterogeneous architecture)*는 고성능 컴퓨팅 AI

워크로드를 감당할 수 있는 생성형 AI 칩의 패키징 모델로 빠르게 부상하고 있다.²³ 그러나 다양한 지역의 여러 벤더로부터 다이(die, 회로 패턴 단위)와 부품을 조달해 패키징해야 하기 때문에 상당한 복잡성이 발생해, 2026년 주요 지정학적 병목으로 떠오를 가능성성이 크다. 칩렛 기반 솔루션의 연간 매출은 2026년에 약 1,000억~1,100억 달러 규모로 예상된다.²⁴

* 이종 아키텍처(heterogeneous architecture)는 서로 다른 연산 특성·명령어 구조·성능 특성을 가진 복수의 컴퓨팅 자원(CPU, GPU, NPU, FPGA, ASIC 등)을 하나의 시스템 내에서 결합·활용하는 설계 방식을 의미한다.

HBM은 범용 처리에 최적화된 단일 아키텍처 중심 구조에서 벗어나, 워크로드 특성에 맞는 최적의 연산 엔진을 병렬로 활용하는 것이 핵심 목적이다. 이 때문에 HBM은 생성형 AI의 학습 및 추론 워크로드에 필수 요소로 자리 잡았다. 2025년 중반 기준으로 HBM 코패키징은 작업이 이뤄지는 지역의 식별 등 관련 정보가 한층 엄격하게 모니터링되고 있다.²⁵ 그 결과, 조립·테스트·패키징에 참여하는 반도체 기업은 추가 정보 공개가 요구될 것으로 보인다. 요구되는 정보에는 패키징에 참여하는 OSAT업체 또는 후공정 제조업체의 명시, 시스템이 코패키징되는 지역, 중간제품 또는 완제품의 배송 국가, 관련 성능 임계치에 대한 상세 정보 등이 포함될 수 있다.

2026년부터는 새로운 제품의 적시 출시를 위해 효과적 후공정에 대한 의존도가 높아진다는 점이 반도체 공급망의 가장 큰 특징으로 꼽힌다.

HBM, 로직, 고속 I/O(input/output) 등 코패키징 작업이 이뤄지는 지역에 대한 라우팅(routing) 및 문서화 요건이 한층 엄격해질 전망이다. 전공정 웨이퍼 펩의 생산 일정과 EDA 벤더의 설계 승인에서부터 최종 고객인 ODM·OEM의 제품 출시 시점에 이르기까지, 공급망의 모든 단계가 첨단 패키징 공정 관련 승인과 절차 속도에 더욱 크게 좌우될 전망이다. 따라서 패키징 벤더나 OSAT 측에서 지역이 발생할 경우 수율 램프업(ramp-up, 생산량 확대)과 튜닝(tuning, 수율 목표 달성을 위한 기술적 최적화 과정) 과정에 영향을 미칠 수 있으며, 이는 결국 리쇼어링(reshoring), 프렌드쇼어링(friend-shoring)과 같이 우호국으로 시설을 이전하는 대응을 촉발할 가능성이 있다.

이러한 요인들은 2026년부터 수 년간 여러 지역에서 계획된 AI 데이터센터 구축에도 영향을 미칠 수 있다. 하이퍼스케일러, 클라우드 서비스 기업, 다양한 산업의 기업들이 2026년에 약 5,000억 달러, 2028년에는 1조 달러 규모를 AI 데이터센터에 투자할 것으로 예상되며,²⁶ 이 중 칩 솔루션이 전체 지출의 약 50~60%를 차지할 전망이다. 예상되는 시장 확대 속도를 고려하면, 공급망 교란은 향후 3년 동안 수백억 달러에서 많게는 수천 억 달러에 달하는 반도체 공급에 영향을 미칠 가능성이 있다.

2025년 반도체 공급망의 시험대: 반도체·AI 생태계의 단절과 새로운 질서 형성

1. 자국 반도체 생태계 강화에 사활 거는 중국

미국 등 주요국의 엄격한 수출통제와 다양한 반도체 기술에 대한 제한으로 중국은 최첨단 AI 칩을 확보하기가 어려워졌다. 중국은 이러한 조치들이 중국의 7nm 이하, 5nm 이하 공정 발전에 걸림돌이 될 수 있다는 판단에 국내 반도체 혁신을 가속화했다. 한편 중국 외 지역의 칩 제조업체들은 2025년 3nm와 2nm에서 2026~2027년에는 1.8nm로 전진하는 움직임을 보이고 있다.²⁷

수출통제에 대응하기 위한 우회 전략을 택한 중국은 전공정 제조뿐 아니라 칩 설계와 첨단 패키징 등 글로벌 반도체 공급망의 다양한 영역을 탐색할 가능성이 있다.²⁸ 구식 노드를 활용해 정교한 칩을 제작하면 첨단 패키징에 적용할 수도 있지만, 미국은 이러한 패키징 시스템이 최첨단 AI 칩 수준의 성능을 내지 못하도록 추가적인 통제와 점검을 시행할 가능성이 크다.

2. 주권 기술 스택 구축 경쟁 가속화되며 새로운 지역 질서 형성

기술 주권은 각국이 디지털 기술을 독자적으로 개발, 통제, 규제하려는 지향점을 의미한다.²⁹ AI는 향후 경제 발전과 국가 경쟁력을 견인하는 핵심 동력으로 널리 인식되고 있기 때문에, 각국 정부가 디지털 인프라에 대한 직접적 통제력을 강화하려는 움직임 속에서 AI 생태계가 더욱 주

목받고 있다. 각 국가와 지역은 경쟁에서 뒤처지거나 의도치 않게 기술 주도권을 상실하는 상황을 원하지 않는다. 이러한 긴박감은 첨단 AI 역량이 현재 소수 국가와 기업에 집중되어 있다는 사실로 인해 더욱 증폭되고 있다. 또한 고급 칩 제조의 리쇼어링을 가속화하는 미국과 유럽이 2026년부터 수 년간 국내뿐 아니라 인도, 베트남, 말레이시아 등 국가에 첨단 조립 및 테스트 허브를 대안적으로 구축하기 위한 투자를 확대할 가능성이 크다.³⁰

3. 반도체 산업 공급망 회복력 강화 필요

반도체 생태계 전반의 기업들은 자체적 스트레스 테스트를 통해 공급망 전 구간을 점검하고 사이버보안 대비 태세를 강화하는 등 선제적으로 공급망 회복력을 강화하는 데 주력할 필요가 있다.³¹

이를 위해 지역별 공급망 다변화와 대체 조달 전략, 채널 파트너십에 대한 투자는 필수적이다. 핵심 소재와 부품의 독립적 공급망 확보는 전략적으로 매우 중요하기 때문에 현지화 가속화와 규제 대응력 강화에 나설 필요가 있다. 또한 지정학적 갈등은 글로벌 AI 생태계를 분절시킬 수 있으며, 이로 인해 칩이 회색시장을 통해 수출되는 위험, 기업의 제품 및 공급망 모니터링·추적 역량 강화 압력 증가 등 다양한 도전과제가 발생할 수 있다.

AI 추론 최적화 칩 시장은 2026년에 수십억 달러 규모로 성장할 것으로 예상되지만, 대부분 첨단 컴퓨팅은 하이퍼스케일 데이터센터 또는 이

에 버금가는 칩과 랙을 사용하는 온프레미스 서버에 탑재되는 최첨단 AI 칩에서 이뤄질 것이다.³² 이로 인해 추론용 AI 칩과 관련한 인프라에 새로운 수출통제와 추가 요구 사항이 적용될 가능성이 있으므로, 반도체 산업 전반은 조달에서 유통까지 대체 공급망 옵션을 확보해야 한다.

또한 학습에서 추론으로 AI 생태계의 무게중심이 이동하면서, 반도체 산업에서 소프트웨어의 중요성도 커지고 있다. 예컨대 대규모 단일형(monolithic) AI 학습용 GPU를 여러 개의 소형 GPU 슬라이스 또는 가상 GPU 인스턴스로 재구성하는 소프트웨어 프로그래밍 기술을 활용할 수 있다.³³

한편 미국과 유럽 소재 디바이스 OEM들은 생산 및 조립 시설을 중국에서 동남아시아나 인도로 이전해야 할 수 있다. 이로 인해 단기적으로 생산 비용이 증대해 소비자 테크 제품의 인플레이션이 촉발될 수 있다. 반도체 기업은 민첩성과 규모의 경제를 유지하고, 2026년 이후 변화하는 무역 구조를 예측해 선제적으로 대응하며, 핵심 물류 경로와 인프라를 보호하기 위한 국가간 대안적 전략적 협력을 모색해야 한다.

무역 긴장이 글로벌 제휴와 채널 파트너십을 재편하면서, 반도체 산업의 회복력은 2026년을 앞두고 전례 없는 시험대에서 있다. 글로벌 칩 공급망은 상호 연결성과 전략적 중요성이 매우 높다. 공급망 회복력 강화를 위해 업계 여러 이해관계자의 선제적 협력과 참여가 그 어느 때보다 절실하다.

Korean Perspectives

**동맹과 통제의 시대,
AI 연산의 병목을 지배하는 자가 승리한다**

AI 칩을 둘러싼 글로벌 무역 통제 강화와 공급망 재편은 한국 반도체 산업의 경쟁 구도를 근본적으로 바꾸고 있다. 반도체는 더 이상 단순한 시장 재화가 아니라 국가 안보와 경제, 과학 경쟁력을 좌우하는 전략 자산으로 격상되었고, 이에 따라 공급망 역시 비용과 효율 중심에서 통제와 동맹, 컴플라이언스 중심으로 이동하고 있다. 특히 생성형 AI와 고성능 컴퓨팅에 필수적인 첨단 공정, 장비, 소재, 설계 도구가 소수 국가와 기업에 집중된 구조 속에서, 한국 반도체 기업들은 글로벌 공급망 병목과 무역 장벽의 영향을 가장 직접적으로 받는 위치에 놓여 있다.

이러한 환경에서 한국의 강점은 분명하다. 고대역폭 메모리(HBM)와 첨단 패키징은 AI 연산 인프라의 핵심 병목으로 자리 잡았으며, 한국은 이 영역에서 이미 글로벌 생태계의 대체 불가능한 축을 형성하고 있다. 그러나 앞으로의 경쟁은 기술 우위만으로 결정되지 않는다. 전공정-후공정-설계-검증에 이르는 엔드투엔드 공급망 전반에서 안정성과 추적성, 규제 준수 능력을 갖춘 기업만이 실제 거래와 납품을 지속할 수 있는 국면으로 진입하고 있다.

SK하이닉스와 삼성전자의 최근 행보는 이러한 변화에 대한 한국 산업의 대응 방향을 잘 보여준다. SK하이닉스는 HBM이라는 AI 연산의 구조적 병목을 선점하고, 기술 고도화와 생산 역량 확장을 통해 글로벌 가속기 및 하이퍼스케일러 로드맵에 맞춘 안정적 공급 능력을 핵심 경쟁력으로 구축하고 있다. 이는 단순한 메모리 공급을 넘어, AI 서버와 데이터센터가 정상적으로 작동하기 위한 필수 요소를 책임지는 생태계 중심 기업으로 자리매김하려는 전략으로 해석할 수 있다.

삼성전자는 HBM을 포함한 메모리 역량에 파운드리와 첨단 패키징을 결합해, 고객이 요구하는 성능·전력·열·검증 조건을 시스템 단위로 대응하는 방향을 강화하고 있다. 메모리-로직-패키징을 아우르는 수직 통합 역량을 바탕으로, 개별 부품 공급자가 아니라 복잡해지는 AI 연산 환경을 종합적으로 지원하는 시스템형 공급자로 포지셔닝하려는 접근이다. 이는 규제와 실사, 검증 요구가 강화되는 환경에서 고객의 부담을 구조적으로 낮춰주는 차별화 요소로 작용할 가능성이 크다.

두 기업의 사례가 공통적으로 시사하는 바는 명확하다. AI 시대의 반도체 경쟁은 누가 더 뛰어난 칩을 설계하느냐의 문제가 아니라, 누가 규제·검증·패키징·물류·보안까지 포함한 복합 조건을 충족하며 끝까지 안정적으로 납품할 수 있느냐의 문제로 이동하고 있다는 점이다. 즉, 성능 경쟁에서 공급 가능성(supply assurance) 경쟁으로 게임의 룰이 바뀌고 있다.

이에 따라 한국 반도체 산업의 대응은 기존 강점을 방어하는 수준을 넘어, AI 연산 인프라의 표준 경로에 한국 기술 스택을 고정시키는 방향으로 확장돼야 한다. 이를 위해서는 설계 단계부터 수출 통제와

실사 요건을 반영하는 ‘규제 내재화’(compliance-by-design) 접근이 필수적이며, EDA·PDK·IP·패키징 경로까지 포함한 공급망 가시성과 추적성을 디지털 체계로 관리해야 한다. 컴플라이언스는 더 이상 비용 요소가 아니라, 실제 납품을 가능하게 하는 핵심 역량으로 전환되고 있다.

또한 HBM의 경쟁 우위를 메모리 단품에 국한하지 않고, 첨단 패키징과 테스트·검증, 열·전력 설계까지 묶은 시스템 수준의 랙인 구조로 확장해야 한다. 향후 공급 지연 리스크는 전공정뿐 아니라 OSAT, 계측·검사, 패키징 승인 단계에서 발생할 가능성이 높아지는 만큼, 후공정 영역에 대한 선제적 투자와 전문 인력 확보도 중요해지고 있다. 아울러 추론 중심으로 확장되는 AI 워크로드 환경에서는 성능 분할과 가상화, 운영 최적화 등 소프트웨어 역량과 공급망 모니터링, 사이버보안 역량 역시 반도체 경쟁력의 일부로 편입될 필요가 있다.

결국 AI 시대의 반도체 경쟁은 성능이나 공정 미세화만의 싸움이 아니다. 누가 가장 신뢰받는 방식으로, 가장 복잡한 조건 속에서도 끝까지 공급할 수 있는지를 겨루는 구조로 재편되고 있다. 한국은 이미 HBM과 제조 역량이라는 전략적 요충지를 확보했으며, SK하이닉스와 삼성전자가 보여주는 상이하지만 상호 보완적인 전략은 이러한 우위를 시스템형 경쟁력으로 확장할 수 있는 가능성을 시사한다. 앞으로의 승부는 단일 제품의 우위를 반복하는 데 있지 않다. AI 연산 인프라의 표준 경로에 한국의 기술과 공급망을 얼마나 깊이 고정시킬 수 있느냐가 향후 성과 격차를 좌우하는 결정적 변수가 될 것이다.



최호계 파트너
한국 딜로이트 그룹
TMT Industry 리더

7장 기술 주권 확보 경쟁

기술 주권 경쟁은 ‘인공지능(AI) 인프라를 누가 통제하느냐’를 둘러싼 국가간 경쟁이며, 완전 자립이 아닌 ‘통제 가능한 협력’이 새로운 표준이 되고 있다.

핵심 내용 요약 (Executive Summary)

▷ 기술 주권 경쟁 격화의 구조적 배경

- AI·클라우드·반도체·통신이 국가 경쟁력의 핵심 인프라로 전환
- 기술과 시장 논리가 안보·산업·주권 이슈로 성격 변화
- 기술 주권 개념이 데이터에서 인프라 위치·소유·운영·거버넌스 전반으로 확장

▷ AI가 촉발한 통제 경쟁

- 2026년 이후 소버린 AI 컴퓨팅 투자 확대, 2030년까지 미·중 외 비중 증가
- AI는 경제·산업·안보 전반에 영향을 미치는 범용 기술로 부상
- 컴퓨팅·데이터·모델이 소수 국가·기업에 집중되어 있다는 위기의식 확산
- 다수 국가는 AI 인프라의 대외 의존을 전략적 리스크로 인식

▷ 기술 주권의 4대 투자 전선

- (클라우드) 소버린 클라우드 확산, 데이터·운영의 역내 통제 강화
- (AI 모델·데이터센터) AI 팩토리, 주권형 모델, 국가 단위 AI 역량 구축
- (반도체) 역내 생산·설계 확대를 통한 공급망 회복탄력성 강화
- (위성통신) 자체 위성 군집 조성으로 안보·연결성 확보

▷ 기술 주권 강화 효과, 한계, 그리고 기업의 과제

- 기술 주권 강화는 산업 성장과 보안·회복탄력성을 높이지만, 규제 분절화·투자 비효율·인재 이동성 저하·전력 부담을 초래
- AI·클라우드·반도체 전반에서 완전 자립은 한계가 있으며, 선별적 협력과 동맹이 불가피
- 글로벌 의존성 진단을 기반으로 멀티·소버린 클라우드 전략을 재정의하고, 핵심 인재 내재화

세계 각국이 자국의 독자 기술 및 AI 인프라 구축 경쟁, 즉 기술 주권(tech sovereignty) 경쟁을 벌이고 있다. 이러한 흐름은 어떤 함의를 가지며, 글로벌 기업들은 어떻게 대비해야 할까?

전 세계 지정학적 여건이 갈수록 복잡하고 불확실해지면서, 각국 정책 결정자들과 기업들은 인공지능(AI)을 중심으로 디지털 인프라에 대해 더욱 강력한 통제력이 필요하다는 주장을 펼치고 있다. 시장조사기관 가트너(Gartner®)는 “2028년까지 전 세계 정부의 65%가 자국의 독립성을 강화하고 역외 규제 개입으로부터 보호하기 위해 일정한 기술 주권 요건을 도입할 것”이라고 추정했다.¹

기술 주권은 클라우드, 양자 컴퓨팅, AI, 반도체, 디지털 통신 인프라 등 디지털 기술을 자국이 직접 개발·통제·규제·편당할 수 있는 능력을 의미한다.² 이 개념은 데이터 흐름과 더불어 관련 시설의 물리적 위치, 소유·운영·거버넌스 주체, 인프라를 구성하는 하드웨어·소프트웨어·서비스 제공 주체 등에 대한 지리적·법적·규제적 요건을 포함한다.

세계 각국은 이미 오래 전부터 기술 주권을 확보하려는 열의를 보였지만, 2026년부터 더욱 경쟁적 움직임이 나타날 것으로 보인다. 따라서 향후 10년 동안 클라우드 컴퓨팅, 반도체, 데이터센터, AI 모델, 커넥티비티(connectivity, 연결성), 위성 통신 기술에 막대한 투자가 유입될 전망이다. 전 세계가 긴밀히 연결된 만큼 어떤 국가도 완전한 기술 주권을 달

성하기는 힘들겠지만, 대다수 국가가 지금보다 훨씬 더 강력한 기술 주권을 원하고 있다.

특히 AI는 향후 경제 성장과 국가 경쟁력을 좌우할 핵심 동력으로 평가되기 때문에, 현재 관련 생태계가 큰 주목을 받고 있다. 더구나 컴퓨팅 파워를 포함한 첨단 AI 역량이 현재 극소수의 국가와 기업에 의해 통제되고 있어 기술 주권을 둘러싼 긴박감이 한층 고조되고 있다.

옥스퍼드 인터넷 연구소(Oxford Internet Institute) 연구에 따르면, 전 세계 공공 AI 컴퓨팅 시설이 단 34개국에 몰려 있으며, 이 중 학습 수준의 컴퓨팅 역량을 확보한 국가는 24개국에 불과하고, 대부분은 소수 해외 기업이 통제하는 클라우드 또는 칩 인프라에 의존한다.³ 또한 전 세계 AI 컴퓨팅의 90%가 미국과 중국 기업에 의해 운영되고 있는 것으로 나타났다.⁴

딜로이트는 2026년에 더 많은 국가가 AI 컴퓨팅 역량을 확보하게 될 것이며, 소버린 AI(sovereign AI) 컴퓨팅 파워 구축에 1,000억 달러 이상이 투입될 것으로 예측한다. 2030년까지 미국과 중국 이외 지역 기업들이 관리하는 AI 컴퓨팅의 비중은 현재 전 세계 용량의 10%에서 두 배로 증가할 것으로 예상된다. 이러한 변화를 시사하듯, AI 및 가속 컴퓨팅 플랫폼 기업인 엔비디아(NVIDIA)는 소버린 데이터센터 시장에 공급할 AI 칩 판매액이 2025년 200억 달러에 이를 것이라고 전망했다. 이는 전년 대비 100% 증가한 수치다.⁵

기술주권 경쟁, 대(大)유럽이 주도

유럽연합(EU)은 2024년 9월 유럽 경제 경쟁력을 전반적으로 강화하기 위한 권고안을 담은 일명 ‘드라기 보고서’(Draghi report)를 공개했다.⁶ 보고서에는 EU의 역내 기술 부문의 발전 방안과 혁신 및 기술 채택, 근로자 생산성 개선 방안 등에 초점을 맞춘 내용이 포함됐다. 보고서는 200개 이상의 유럽 기업과 기관에 기술 주권 강화를 위한 근본적 조치를 촉구하기 위해 출범한 유로스택 이니셔티브(EuroStack Initiative)의 발판 역할을 했다.⁷ 해당 이니셔티브가 촉구한 조치에는 유럽산 제품 구매 확대, 기존 자산의 통합 및 활용 효율성 개선, 연구·개발보다 제품화에 주력, 충분한 자본 확보, 유럽 클라우드 사용자 데이터 보호 강화 등이 포함돼 있다. EU 집행위원회는 ‘기술 주권 담당 커미셔너’(Commissioner for Technology Sovereignty)라는 별도 직책을 수립해 이러한 노력을 주도하고 있다. 이는 EU가 개인정보보호 규정(GDPR), 디지털서비스법(DSA), AI법(AI Act) 등 여러 규제 체계를 통해 강조해온 EU의 철학·가치·원칙을 가장 잘 구현하는 해법으로, 주권 기반 기술 솔루션을 지향해온 오랜 역사와도 맞닿아 있다.

다만 초기의 열기와 기대는 2025년 초부터 다소 시들해진 것으로 보인다. 이는 최근 발표된 EU 국제 디지털 전략(EU International Digital Strategy)에서도 확인되는데, 해당 전략은 AI·반도체·양자컴퓨팅·사이버보안 분야에서 타국과의 협력에 더 중점을 두고 있다.⁸ 이처럼 최적의 전략적 접근법이 무엇이냐는 논의는 여전히 현재진행형이지만, 향후 5

년 동안 유럽 클라우드 컴퓨팅, AI 데이터센터 및 관련 기업, 반도체, 위성 통신 분야에 공공·민간을 합쳐 1,000억 유로가 넘는 투자가 이뤄질 것으로 보인다.

1. 클라우드 컴퓨팅

유럽 현지 클라우드 제공업체들이 글로벌 시장에서 차지하는 비중은 20%가 채 되지 않는다.⁹ 따라서 이들이 글로벌 하이퍼스케일러와 본격적으로 경쟁할 만한 수준으로 성장하려면 상당한 투자와 시간이 필요하다. 보다 현실적인 방안은 글로벌 기업들이 유럽 시장에 특화된 형태로 자사 기능을 더욱 적극적으로 현지화해 제공하는 것이다. 아마존 웹서비스(AWS)는 독일에 유럽 주권 클라우드를 구축하기 위해 약 80억 유로를 투자하겠다고 발표했다. 이 프로젝트의 목표는 고객이 데이터를 EU 내에 보관할 수 있도록 하고, 독립성을 보장하며, EU 시민이 운영·관리·보안·거버넌스를 담당하는 환경을 구축하는 데 있다.¹⁰ 마이크로소프트(MS) 또한 유럽 AI, 사이버보안, 개인정보보호, 회복탄력성, 경제 경쟁력 등에 투자하겠다고 발표했고, 마이크로소프트 소버린 클라우드(Microsoft Sovereign Cloud) 플랫폼과 솔루션 계획을 내놓았다.¹¹

2. AI 모델과 데이터센터

정부와 민간 부문 모두에서 전반적인 AI 역량 강화를 위한 다양한 이니셔티브가 추진되고 있다. EU 집행위원회의 ‘AI 대륙 행동 계획’(AI Continent Action Plan)은 유럽 내 기존 슈퍼컴퓨팅 인프라를 기반으로 AI 팩토리와 기가팩토리를 구축하고, 인베스트AI(InvestAI) 프로그램을 통

해 새로운 투자를 유치하는 것을 목표로 한다.¹² 이 프로그램은 최대 다섯 곳의 새로운 AI 기가팩토리 설립을 위해 200억 유로를 지원할 예정이며, 이를 통해 ‘주권 프런티어 모델’로 불리는 첨단 AI 모델 개발이 가능해질 전망이다. ‘AI 대륙 행동 계획’은 AI 모델 학습을 위한 데이터 접근성 강화, AI 애플리케이션 활용 확대, 인재 및 역량 개발 촉진도 주요 과제로 삼고 있다. 민간 영역에서는 엔비디아와 퍼플렉시티(Perplexity)가 협력해 오픈소스 기반의 현지화된 AI 모델을 학습 및 보급하기 위한 노력을 진행 중이다.¹³ 또한 엔비디아는 MGX, 미스트랄AI(Mistral AI) 등 AI 투자사들과 함께 2028년까지 총 85억 유로를 투입해 유럽 최대 규모의 AI 데이터센터를 구축하는 프로젝트를 지원하고 있다.¹⁴ 영국은 ‘스타게이트 UK’(Stargate UK)라는 단계적 프로젝트를 추진해 국가 전역의 AI 인프라를 확충하고 국내 AI 도입을 가속화하고 있다.¹⁵

3. 반도체

미국과 마찬가지로 유럽도 더 많은 반도체 제조 역량을 역내 유치하고, 공급망 회복탄력성을 강화하며, 현지 생태계를 발전시키고, 유럽 기업의 경쟁력을 높이는 방향으로 움직이고 있다. 이를 위해 EU는 2023년 ‘EU 반도체법’(Chips Act)을 제정해 소규모 시험 생산 라인과 공동 설계 플랫폼뿐 아니라 역량 센터(competency center), 양자 칩 관련 자원에 2030년까지 총 430억 유로를 투자할 계획이다.¹⁶ 이미 상업적 투자도 활발히 추진되며, 핀펫(FinFET, 전계효과 트랜지스터) 파운드리, 스마트 파워 팩(smart power fab, 스마트 전력반도체 공장), 실리콘 카바이드(SiC) 칩 제조 공장 등 다양한 프로젝트가 진행 중이다.¹⁷

4. 위성 통신

유럽의 또 다른 핵심 이니셔티브는 역외 공급업체에 대한 의존도를 낮추고, 군·정부·상업용 목적에 필요한 안전하고 신뢰성 높은 서비스를 보장하기 위해 자체 위성 통신 콘스텔레이션(constellation, 군집)을 구축하는 것이다. 현재 아이리스2(IRIS², Infrastructure for Resilience, Interconnectivity and Security Satellite) 콘스텔레이션과 유텔샛 원웹(Eutelsat OneWeb) 등 두 가지 주요 프로젝트가 추진되고 있다. 아이리스2는 약 110억 유로의 비용을 들여 여러 궤도에 약 300기의 위성을 배치할 예정이며,¹⁸ 유텔샛은 현재 630기 이상의 위성을 보유한 원웹 저궤도 위성 인터넷 네트워크의 구축 및 고도화를 가속화하는 노력을 펼치고 있다.¹⁹ 유텔샛은 최근 해당 프로젝트에 영국과 프랑스 정부의 신규 투자를 유치하는데 성공했다.²⁰ 그러나 경쟁이 치열해지고 시장이 점점 혼잡해지는 상황에서, 두 프로젝트가 기존 서비스에 실질적인 대항마로 개발되고 유럽의 수요를 완전히 충족하는 수준에 이르기까지는 상당한 시간과 대규모 추가 투자가 필요할 것으로 보인다.²¹ 실제로 아이리스2의 완공 목표 시점은 2031년이다.

유럽 외 지역의 기술주권 확보 움직임

유럽이 기술주권 경쟁을 적극 주도하고 있지만, 다른 국가와 지역도 다음과 같이 고유하고 혁신적인 접근을 통해 주권 기술 역량을 강화하고 있다. 이들은 대부분 AI에 초점을 맞추고 있다.

- **한국**은 자국 언어와 문화에 최적화된 소버린 AI 역량을 구축하는 것을 목표로 하고 있다.²² 일례로 카카오는 새로운 맞춤형 디지털 서비스를 위해 오픈AI(OpenAI)와 협력하고 있다.²³ 또한 SK그룹은 국내 인프라 강화를 위해 아마존웹서비스(AWS)와 함께 2029년까지 약 50억 달러를 투자해 한국 최대 AI 데이터 센터를 공동 구축한다고 발표했다.²⁴
- **일본**은 국내 AI 역량을 강화하고 반도체 산업을 재편하기 위해 래피더스(Rapidus) 이니셔티브를 출범했다. 래피더스는 일본 정부와 일본 주요 8개 기업이 출자한 신규 반도체 기업으로 2 나노미터(nm) 공정을 목표로 하고 있으며, 이를 위해 일본 정부는 2030년까지 총 650억 달러 규모의 투자 패키지를 계획하고 있다.²⁵
- **아프리카** 최초의 AI 팩토리가 남아프리카공화국 IT 기업인 카사바 테크놀로지스(Cassava Technologies)의 데이터센터 시설에 엔비디아의 AI 및 가속 컴퓨팅 플랫폼을 기반으로 구축될 예정이며, 추후 이집트·케냐·모로코·나이지리아 등 다른 국가로도 확장될 계획이다.²⁶
- **인도**는 기술 스택 전 계층에서 자립을 강하게 추진하고 있으며, 인디아 반도체 미션(India Semiconductor Mission)과 인디아 AI(IndiaAI) 같은 정부 프로그램이 관련 수요를 충족하기 위해 운영되고 있다.²⁷ 다만 인도는 제한적 컴퓨터 접근성, 다양한 언어 지원 필요성, 고품질 학습 데이터 부족 등 AI 모델 개발을 가로막는 고유한 도전과제를 해결해야 한다.²⁸ 한편 인도 정부의 개방형 디지털 인프라 플랫폼인 인디아 스택(India Stack) 구축을 통해 확보한 강력한 디지털 역량은 해외에 수출해 새로운 경쟁적 디지털 생태계를 만드는 기반이 될 수도 있다.²⁹
- **캐나다** 정부의 소버린 AI 컴퓨팅 역량 확보 전략은 민간 투자 확대, 공공 인프라 개선, 컴퓨팅 자원 접근성 강화에 중점을 두고 있다.³⁰ 캐나다 정부는 국내 AI 기업 코히어(Cohere)와 파트너십을 맺고 캐나다의 전반적 기술 역량을 강화할 방안을 모색하고 있다.³¹ 또한 텔러스(TELUS), 사스크텔(SaskTel), 벨(Bell) 등 여러 캐나다 통신사가 소버린 AI 데이터센터 구축을 계획하고 있다.³²
- **중동**에서는 다수의 국가가 소버린 클라우드와 AI 데이터센터에 대한 투자를 늘리고 있으며, 그 중 아랍에미리트연합(UAE)의 1기가와트(GW)급 AI 클러스터 ‘스타게이트 UAE’(Stargate UAE) 프로젝트가 대표적이다.³³ 사우디아라비아 국부펀드(PIF)는 2025년 휴메인(HUMAIN)이라는 신규 회사를 설립해 AWS 및 엔비디아 등 글로벌 파트너와 함께 엔드투엔드(E2E)

AI 인프라 구축을 추진하고 있다.³⁴ 이와 관련해 약 230억 달러 규모의 투자 계획이 이미 발표됐다.³⁵

기술주권 강화의 장단점

예상대로 대부분의 국가가 가까운 시일 내 강력한 기술 주권 정책과 프로그램을 추진한다면 어떤 일이 벌어질까? 각국이 기술 역량 전반을 더 강하게 통제하게 될 경우 다양한 잠재적 이익이 존재한다. 예를 들어, 더 높은 세수와 민간 자본 투자 확대, 자국민에게 더 나은 일자리 제공, 자국 테크 기업의 성장 가능성 확대 등이 있다. 더 높은 자립성을 확보하면 전반적인 회복탄력성이 강화되고, 개인정보보호와 보안이 개선되며, 외국의 정치적 혼란으로 받게 되는 영향도 줄일 수 있다. 또한 AI의 경우, 기초 모델이 국내에서 개발되면 자국의 언어와 관습, 데이터세트를 더 충실히 반영할 수 있다.

하지만 다음과 같은 여러 과제가 동시에 발생할 수도 있다.

- **투자 흐름의 변화:** 해외 직접 투자(FDI), 인수합병(M&A), 합작 투자(JV) 진행 시 더 많은 조건과 요구사항에 직면하게 될 수 있다. 벤처 투자자들이 글로벌 기회보다 국가 전략적 이익을 우선시할 가능성은 요원하기 때문에,³⁶ 벤처캐피털 투자 방향도 달라질 수 있다.

- **분절화 심화:** 폐쇄적 제로섬 방식의 접근은 협력 수준을 낮추고 국제 관계를 약화시키며 학술 파트너십을 위축시킬 수 있다. 국가 간 데이터 흐름과 독점적 통신 인프라는 더욱 제한될 가능성 이 있고, 새로운 표준과 규제가 증가할 수 있다.

- **인력 이동성 제한:** 각국이 자국 기술 역량을 강조할수록 고숙련 인재의 글로벌 이동이 제한될 수 있다. 이는 AI, 사이버보안, 칩 설계와 같이 핵심 분야에서 특히 두드러질 수 있다. 또한 국가 별로 광범위한 인적 역량을 강화하기 위해 투자를 대규모로 확 대해야 할 수 있다.³⁷

- **환경적 영향:** 팹·연구소·데이터센터 및 관련 인프라 건설이 크게 증가하면 상당한 자원 부담이 발생할 수 있다. 일부 국가의 전력망은 이미 최대치 수준으로 가동되고 있고, 신규 데이터센터는 수천 메가와트(MW)의 추가 전력이 필요하기 때문에 전력 공급을 두고 주거용 전력 수요와 경쟁할 수 있다.³⁸ 또한 환경오염을 유발하지 않는 저탄소·무탄소 전력원을 확보하는 것도 큰 과제다.³⁹

- **새로운 파트너십 필요성 증대:** 모든 국가가 독자적으로 모든 것을 수행할 수는 없다. 향후에는 서로의 강점을 활용하기 위한 양자 협력, 지역 프레임워크, 비전통적 기술 동맹이 더욱 증가할 수 있다.⁴⁰

• **과잉설비 위험:** 시장이 수용할 수 있는 기초 AI 모델에는 한계가 있다. 2028년까지 전 세계 AI 인프라에 약 3조 달러의 설비투자가 예상되는데, 이처럼 막대한 투자가 모두 효과를 낼 수 있을지 의문이다.⁴¹ 장기 수요가 기대에 못 미칠 수도 있으며, 새로운 기술 혁신이 현재의 접근 방식 일부를 무용하게 만들 수도 있다.⁴²

기술주권 경쟁의 가속화: 지정학·규제·투자 전선의 재편

기술주권을 둘러싼 경쟁은 2026년에도 더 치열한 논쟁, 정부의 추가 조치, 투자 활동 확대로 이어질 전망이다. 기술주권 경쟁의 동기와 그에 따른 결과에 대해서는 여전히 논란의 여지가 많다. 하지만 많은 이들이 기술주권에 국가와 지역 블록의 미래 번영이 걸려 있다고 믿는 만큼, 이미 행동은 시작됐고 앞으로 더 가속화될 것이다.

- **글로벌 의존성 점검:** 데이터 흐름, 퍼블릭 클라우드, 공급업체, 공급망, 금융, 규제 등 글로벌 시장에 의존하는 모든 핵심 요인을 파악하고 평가하라. 최대한 글로벌 유연성을 확보할 수 있는 신규 파트너십을 구축하고 기존 협력을 강화하라. 또한 기업의 글로벌 운영 방식을 투명하게 설명할 수 있어야 한다.
- **규제 복잡성 대비:** 빠르게 변화하는 규제 환경에 대비하라. 데이터 현지화, 사이버보안, M&A, 자본 흐름에 관한 새로운 규제

가 등장할 수 있다. 비즈니스가 어떤 부분에서 가장 취약한지 파악하고 시나리오 플랜을 수립하라. 컴플라이언스 프로그램을 강화하는 것도 필수다.

- **클라우드 전략 재점검:** 퍼블릭 클라우드와 프라이빗 클라우드 간의 균형을 재검토하라. 회복탄력성과 규제 준수를 높이기 위해 멀티클라우드 또는 소버린 클라우드 모델을 고려하라. 환경 전반에서 이동성, 상호운용성, 통제력에 우선순위를 두고, 공급업체가 데이터 저장·처리·전송을 위한 자동 규제 준수 기능을 제공하는지 확인해야 한다. 지정학적 변화에 민첩하게 대응 할 수 있도록 비상 계획도 마련해야 한다.
- **인재 회복탄력성 강화:** 핵심 인재를 어떻게 확보할 것인지, 이에 실패할 경우 어떤 결과가 나타날지 파악하라. 대체 인력 확보 전략을 개발하고, 정부의 인력 개발 프로그램과 대학 파트너십을 활용해 전문 역량을 키우는 것이 중요하다.

Korean Perspectives

기술 주권 경쟁 속에 한국 산업의 선택, 자립이 아닌 통제 가능한 협력 확보

기술 주권을 둘러싼 글로벌 경쟁이 본격화되면서, AI·클라우드·반도체 등 디지털 인프라는 더 이상 효율이나 비용의 문제가 아니라 국가 경쟁력과 안보를 좌우하는 전략 자산으로 재정의되고 있다. 세계 각국은 자국 중심의 기술 통제력을 강화하기 위해 막대한 투자를 단행하고 있으며, 이러한 흐름은 글로벌 공급망과 데이터, 인재 이동 전반에 구조적인 변화를 일으키고 있다. 한국 역시 이 흐름에서 예외가 아니다. SK, 네이버를 비롯해 네이버클라우드, LG AI연구원, SK텔레콤, KT, 카카오, 업스테이지 등 주요 기업들은 AI 데이터센터 구축과 독자적 AI 파운데이션 모델 개발을 중심으로 컨소시엄을 구성하며 대응에 나서고 있다. 또한 정부 역시 과학기술통신부를 중심으로 국가 소버린 AI 개발을 위하여 컨소시엄을 선정하여 향후 5천억 이상의 예산과 인력지원을 계획하고 있다. 이는 단일 기업 차원의 기술 확보를 넘어, 인프라·모델·법제를 동시에 구축하려는 산업 차원의 전략적 전환으로 볼 수 있다.

그러나 기술 주권 경쟁이 요구하는 조건은 단순하지 않다. AI 컴퓨팅 역량을 둘러싼 경쟁은 데이터의 위치와 소유권, 운영과 거버넌스, 보안과 규제 준수 문제로 빠르게 확장되고 있다. 국내 기업들은 글로벌

공급망의 불확실성, 데이터 현지화 요구, AI·반도체 인재 확보 경쟁이라는 복합적 도전에 직면해 있으며, 동시에 AI 데이터센터 구축에 필요한 막대한 전력 수요와 안정적인 운영 인프라 확보라는 현실적인 제약도 함께 감당해야 한다. 이러한 환경에서 우리 산업이 현재 역량을 분명히 파악하고 현실적인 대안을 마련해야 한다. 먼저 완전한 기술 자립은 현실적으로 불가능하지만, 핵심 영역에서의 최소한의 자립성과 통제력 확보는 가능할 것이다. 글로벌 기술 생태계는 상호 고도로 연결돼 있어 고립된 형태의 독립은 오히려 경쟁력을 약화시킬 수 있다. 따라서 기술 주권 경쟁의 본질은 ‘완전한 자립’이 아니라, 전략적 파트너십을 확보하고 이를 통제 가능한 방식으로 운영할 수 있느냐에 있다.

유럽 주요국의 사례에서 보듯, 글로벌 협력과 자립 전략을 병행하는 균형 감각 역시 필수적이다. 특정 국가나 기업에 대한 과도한 의존을 줄이되, 글로벌 기술 흐름과의 연결을 유지하는 ‘선별적 협력 전략’이 요구된다. 글로벌 경쟁에서 의미 있는 규모와 속도를 확보할 수 있는 현실적인 해법이다.

기술 주권 경쟁은 규제, 데이터, 인재, 환경 이슈가 결합된 복합 리스크 관리의 문제로 진화하고 있다. 데이터 현지화와 사이버보안, 인재 이동성 제한, 대규모 인프라 투자에 따른 환경 부담 등에 선제적으로 대응하지 못할 경우, 기술 투자는 오히려 새로운 제약 요인이 될 수 있다. 따라서 규제 대응과 준법 체계를 사후 대응이 아닌 시스템으로 내재화하는 접근이 필요하다.

이 같은 구조적인 변화는 국내 산업과 기업에게 구체적인 대응전략을 요구한다. 먼저 AI 데이터센터와 디지털 인프라에 대한 지속적인 투자를 통해 전력과 운영의 안정성을 확보하고, 친환경 에너지 도입과 효율적 운영 기술을 병행해야 한다. 동시에 독립적 데이터세트 구축과 데이터 거버넌스 강화를 통해 글로벌 규제 환경 변화에 유연하게 대응할 수 있는 기반을 마련해야 한다. 민관 협력을 통한 기술 개발과 인재 육성, 그리고 미국·중국 중심 구조를 넘어 유럽, 동남아, 인도 등으로 협력 네트워크를 다변화하는 전략도 필수적이다.

결국 기술 주권 경쟁이 심화되는 시대에 한국 산업과 기업이 지향해야 할 방향은 분명하다. 완전한 자립을 목표로 한 폐쇄적 전략이 아니라, 핵심 역량을 전략적으로 확보하고 이를 글로벌 협력 네트워크 속에서 확장하는 것이다. 인프라·모델·법제라는 세 축을 균형 있게 발전시키고, 민관 협력과 국제 파트너십을 통해 복합적 리스크에 선제적으로 대응할 수 있을 때, 기술 주권은 부담이 아닌 지속 가능한 성장의 기반이 될 수 있을 것이다.



최호계 파트너
한국 딜로이트 그룹
TMT Industry 리더

Part

03

미디어 산업의 주의력 경제 시대, 서사·신뢰·플랫폼의 재구성

8장 공영방송, 생존전략으로 글로벌 스트리밍 서비스와 협력 선택

9장 속풀이 촉발한 콘텐츠 제작 생태계의 혁신과 플랫폼 신뢰 위기

10장 생성형 AI 영상이 불러온 콘텐츠 제작 환경의 혁신과 소셜 플랫폼의 책임

11장 영상 팟캐스트, 위협적인 글로벌 미디어·광고 플랫폼으로 진화

8장

공영방송, 생존전략으로 글로벌 스트리밍 서비스와 협력 선택

전통 TV 시청 감소와 글로벌 플랫폼 부상이라는 환경 변화 속에서, 공영방송은 젊은 시청자 유입과 콘텐츠 영향력 확대를 위해 스트리밍·소셜 플랫폼과의 협력을 선택하고 있다. 이러한 파트너십은 콘텐츠 노출 범위 및 재원 확대라는 기회를 제공하는 한편, 편성 통제력 약화와 공적 정체성 훼손이라는 구조적 리스크를 동시에 수반한다. 결국 공영방송의 경쟁력은 플랫폼과의 협력 과정에서도 공적 가치와 주도권을 얼마나 균형 있게 유지하느냐에 달려 있다.

핵심 내용 요약 (Executive Summary)

▶ (방송환경 변화) 공영방송과 스트리밍의 불가피한 결합

- TV 시청 감소와 함께 콘텐츠 유통의 중심이 스트리밍·소셜 플랫폼으로 이동
- 공영방송은 젊은 시청자 유입과 글로벌 도달 확대를 위해 플랫폼 협력 전략을 본격화

▶ (전략적 선택) 콘텐츠 유통과 활용 방식의 전환

- 공영방송은 글로벌 스트리머·소셜 플랫폼과의 파트너십 확대
- 기존 방송 콘텐츠를 유튜브·넷플릭스·아마존 프라임 등에 공급해 콘텐츠 수명과 활용도 제고

▶ (협력의 성과) 콘텐츠 노출 범위·재원·영향력 확대

- 글로벌 플랫폼을 통한 노출 증가로 브랜드 인지도와 영향력 강화
- 추가 수익원 확보와 제작 리스크 분산을 통해 재정적 완충 효과 확보

▶ (구조적 리스크) 통제력·정체성·공적 가치의 약화

- 알고리즘 변화나 계약 조건에 따라 시청자 접근권과 콘텐츠 권리가 급격히 약화될 위험
- 편성·유통 통제력 저하, 브랜드 희석, 공영방송의 공적 사명 훼손 가능성

▶ (생존 전략) 협력 속 균형의 재설계

- 브랜드·데이터 보호, 공정한 수익 배분, 공동 제작 등 균형 잡힌 파트너십 전략 필요
- 스트리밍·소셜 확산으로 미디어의 역할이 재정의 되는 만큼, 지속적 실험과 적응이 생존의 조건

공영방송사들이 새로운 활로를 찾고 있다. 소셜미디어에 콘텐츠를 게시하고, 스트리밍 업체와 공동 제작하며, 거대 콘텐츠 유통사들과 파트너십을 맺고 있다. 이러한 움직임은 영리를 추구하는 미디어 기업에게 중요한 시사점을 제공한다.

량을 강화하고 있다. 젊은 시청자층에 도달하기 위해 소셜 플랫폼에 콘텐츠를 게시 및 홍보하고, 자체 서비스와 글로벌 파트너 간에 시차를 둔 실험적 콘텐츠 공개 방식에도 도전하고 있다.² 이러한 창의적 전략은 비슷한 생존 압박을 받고 있는 미국의 중소 방송사에게 유의미한 교훈을 제공할 수 있다.

TV 방송 사업은 아직 수익성이 있지만, 쇠퇴 양상이 뚜렷해지고 있다. 스트리밍 경쟁의 심화와 소셜 비디오 및 인터랙티브 엔터테인먼트의 확산으로 콘텐츠 산업은 빠르게 재편되고 있다. 콘텐츠 유통의 중심이 기존 방송에서 스트리밍 비디오 서비스로 이동하며, 저작권과 방영권을 들러싼 비즈니스 구조도 새롭게 바뀌고 있다. 반면, 정해진 시간표에 따라 실시간으로 프로그램을 편성·송출하는 기존의 선형 TV(linear TV) 사업은 성장둔화 속에서 구조조정이나 자산 매각 등 근본적 변화를 겪고 있다.¹ 그럼에도 많은 전통 미디어 대기업은 비용이 많이 드는 자체 신규 스트리밍 서비스에 주력하며 TV 수익성이 고점을 찍었던 황금기를 부활시키기 위해 노력을 기울이고 투자를 아끼지 않았다. 그럼에도 지금까지 그 황금기는 돌아오지 않았다.

반면 미국 외 지역에서는 공영방송이 보다 유연한 성공 모델을 만들어 가고 있다. 유럽 공영방송사들은 오랜 기간 입증된 스토리텔링 역량을 기반으로 더 넓은 범위의 글로벌 시청자들을 끌어들이고 있다. 뿐만 아니라, 스트리밍 업체와 공동 제작 계약을 체결함으로써 제작 및 유통 역

2025년에는 불과 수 개월 사이 방송사와 스트리밍 업체 간 주목할 만한 계약이 세 건이나 체결되며 이러한 움직임이 더욱 가속화됐다.³ 딜로이트는 2026년에도 이와 유사한 방송사-스트리머 계약이 추가로 발표될 것으로 전망한다. 더 나아가, 공영방송사를 중심으로 공동 제작 및 기타 새로운 이니셔티브가 더욱 확대될 것으로 보인다. 이는 수만 시간 분량의 방송사 콘텐츠가 스트리밍 서비스와 소셜 플랫폼으로 유입되는 결과를 가져올 수 있으며, 광고 수익 배분과 글로벌 시청 시간 증가 등 잠재적 이익도 기대된다.

현재까지는 공영방송사가 이 같은 계약을 더욱 주도적으로 추진하고 있는데, 주요 목적은 수익 증대보다는 시청자층 확대에 있다. 일부 민영방송사도 유사한 스트리밍 계약을 체결하고 있지만, 이들이 예외적인 사례인지, 아니면 향후 다른 민영방송사에게도 확대될 조짐인지는 시간이 지나야 확인될 것이다. 현재 방송 업계는 공영방송과 스트리밍의 이러한 동맹이 과연 성공적 결과를 가져올지 예의주시하고 있다.

변화의 파고 수용하며 혁신적 해법 모색하는 공영방송사

전 세계적으로 대다수 공영방송사는 대형 문화 기관으로서 기능하며, 대중을 대표하고 연결하는 데 중요한 역할을 해왔으며, 오락·뉴스·교육·문화 형성에 지대한 영향을 미쳐왔다.⁴ 그러나 이들 역시 전통 TV를 봉괴시킨 인구 및 행태 변화의 위협을 피할 수 없는 실정이다.⁵

하지만 민영방송사와 공영방송사 간에는 중요한 차이가 있다. 공영방송사는 재원 구조와 상관없이 공익을 위한 TV·영화·뉴스·다큐멘터리 제작을 사명으로 한다. 따라서 국내 시청자가 공영방송사의 선형 채널이나 스트리밍 서비스를 보지 않는다면 공적 역할을 수행하지 못하게 된다. 이는 공영방송사의 재원과 우선적 접근권을 유지해야 하는지에 대한 논쟁을 더욱 격화시킬 수 있다.

공영방송사는 시민과 정부의 재정 지원에 의존하면서도 공적 임무를 수행하도록 규정돼 있기 때문에 충분한 예산을 확보하지 못하는 경우가 있지만, 여전히 수익보다 공익과 문화적 가치를 더 우선시하는 경향이 있다. 이는 오히려 공영방송사가 민영방송사보다 더 빠르고 유연하며 때로는 더 과감하게 혁신할 수 있는 배경이 되기도 한다. 반면 민영방송사는 수익과 주주 가치를 우선시해야 하므로 수익을 저해할 수 있는 위험을 감수하면서까지 혁신에 나서기가 쉽지 않다.

아래 사례들은 오늘날의 미디어 환경에서 공적 역할을 다하기 위해 노력하는 여타 공영방송사와 더불어 스트리머로서의 성공만을 좋아 혁신

을 주저하는 민간 전통 미디어 기업들이 참고할 만한 모델이 될 수 있다. 다만 영리 기업과의 파트너십이 공영 미디어의 가치와 사명을 훼손할 위험도 수반한다는 점은 유념할 필요가 있다.

1. 유튜브에서 젊은 시청자층 사로잡은 독일 ARD·ZDF의 ‘펑크’

2016년부터 젊은 시청자층이 어떤 영상 채널을 선호 또는 외면하는지 면밀히 분석하던 독일 공영방송 ARD와 ZDF는 젊은 세대가 선호하는 소셜미디어 생태계에서 직접 소통하기 위해 ‘펑크’(funk)라는 과감한 디지털 콘텐츠 프로젝트를 출범했다. 영상 서비스 플랫폼 펑크는 기존 TV 프로그램의 전체 에피소드나 짧은 클립을 올리는 방식이 아니라, 애초부터 유튜브(Youtube)·인스타그램(Instagram)·틱톡(TikTok) 등 소셜 플랫폼용 콘텐츠를 별도로 제작한다.⁶ 펑크는 디지털 네이티브를 위한 수십 개의 오리지널 포맷과 함께, 짧고 산뜻한 설명형 콘텐츠, 도발적 코미디, 짧은 다큐멘터리 등 다양한 장르를 제공한다. 펑크의 과학 채널 ‘마이랩’(maiLab)은 화학과 코로나19 연구 내용을 쉽고 흥미롭게 전달하며 인기를 얻었다.

ARD/ZDF 온라인 이용 행태 조사에 따르면, 30세 미만 독일인의 약 41%가 매주 한 번 이상 펑크 콘텐츠를 시청하는 것으로 나타났다.⁷ 출시 후 2년 동안 펑크 콘텐츠의 유튜브 조회수는 약 22억 회, 시청 시간은 총 1억 7,300만 시간에 달했다. 젊은 시청자들이 짧은 동영상 중심의 새로운 플랫폼으로 이동하자, 펑크도 발맞춰 이동했다. 현재 틱톡과 인스타그램은 펑크의 확산에 크게 기여하고 있으며, 최근 2년 동안 틱톡 조

회 수는 약 23억 회에 달해 유튜브 조회 수를 넘어섰다. 이는 초단편 클립형 영상의 소비 증가라는 트렌드를 반영한다.

젊은 세대가 머무는 공간에서 그들과 소통하는 것은 ARD와 ZDF의 공적 사명을 이행하는 데에도 중요한 일이다. ARD와 ZDF는 인플루언서와 알고리즘 피드가 지배하는 공간에서 공공 자금으로 제작된 전문적 콘텐츠를 제공함으로써, 순수 상업적 소셜미디어 환경에 대한 대안을 제시하고 있다.

2. 유튜브로 유턴한 캐나다 CBC

90년 역사의 캐나다 공영방송 CBC(Canadian Broadcasting Corporation)는 최근까지 자사 스트리밍 앱인 'CBC 젬'(CBC Gem)에 주력해왔다. 그러나 CBC 디지털 전략팀은 2023년 새로운 실험을 시작했다. 기존 CBC 프로그램의 에피소드 전체 또는 심지어 시즌 전체를 유튜브에 업로드해 유튜브를 새로운 유통 채널로 활용하기로 한 것이다.⁸ CBC는 테스트앤런(test-and-learn)* 방식을 적용해, 만약 유튜브가 CBC 자체 서비스의 시청자를 지나치게 빼앗는다면 언제든 유튜브 콘텐츠를 내릴 계획이었다. 하지만 유튜브 업로드는 CBC 젬에 부정적 영향을 주기는커녕, 젊은 신규 시청자를 CBC 콘텐츠로 유입시키는 추가 플랫폼이자 마케팅 퍼널(marketing funnel)**로 작동했다. 많은 시청자가 유튜브에서 프로그램을 접한 뒤, 더 많은 에피소드를 보기 위해 CBC 젬으로 이동하면서 선순환 효과가 발생한 것이다.

* 테스트앤런(test-and-learn)은 가설을 사전에 완벽히 설계해 일괄 실행하는 방식이 아니라, 작은 실험(test)을 빠르게 수행하고 그 결과를 학습(learn)하여 다음 의사결정에 즉시 반영하는 반복적 실행 방법론을 의미한다. 핵심은 '완벽한 계획'이 아니라 빠른 검증·학습·개선의 선순환에 있다.

** 마케팅 퍼널(marketing funnel)은 잠재 고객이 브랜드를 인지한 이후 실제 구매 또는 전환에 이르기까지의 과정을 단계적으로 구조화한 개념 모델을 의미한다. 고객 수가 상단에서 하단으로 이동할수록 점차 감소하는 구조를 깔때기(funnel)에 비유해, 고객 여정을 관리·최적화하기 위한 분석 프레임워크로 활용된다.

현재 CBC는 뉴스, 코미디, 아동 프로그램 등 50개가 넘는 유튜브 채널을 포트폴리오로 운영하고 있다.⁹ 코미디 스케치나 바이럴 뉴스 클립 같은 짧은 영상은 수백만 조회수를 기록하는 일이 흔하다. CBC 뉴스 유튜브 채널은 이미 구독자가 440만 명, 조회수가 총 26억 회를 각각 넘었다.¹⁰ CBC는 또한 20분이 넘는 드라마, 다큐멘터리, 아동 프로그램 에피소드도 전체 게시하는데, 이러한 풀렝스(full-length, 축약이나 요약되지 않은 전체 길이 콘텐츠) 영상은 전체 시청 시간의 거의 절반을 차지한다.¹¹ 클릭을 유도하는 것은 짧은 클립이지만, 시청자를 채널에 머물게 하는 것은 풀렝스 프로그램이다.

2024년 말까지 CBC의 유튜브 실험은 확실한 성과를 거두었다. 전체 채널의 시청 시간이 65% 증가하며, 전략팀이 설정한 25% 성장 목표를 훌쩍 넘어서졌다.¹² 유튜브는 기존 TV로는 좀처럼 도달하기 어려웠던 다양한 시청자층을 끌어들이는 데 큰 도움이 됐다. 이러한 방식은 콘텐츠 활용도를 극대화하고, 오래된 프로그램에 새로운 생명을 부여하며, 새로운 수익원을 창출하는 데 큰 역할을 했다.

3. 최대 유통 플랫폼 활용하는 프랑스 방송사들

프랑스 공영방송 프랑스 텔레비지옹(France Télévisions)은 2025년 7월 아마존(Amazon) 프라임 비디오(Prime Video)와 역사적 유통 계약을 체결했다.¹³ 이 계약에 따라 프랑스의 아마존 프라임 비디오 가입자는 프랑스 텔레비지옹의 실시간 채널과 주문형(on-demand) 카탈로그 가운데 2만 편의 콘텐츠를 추가 비용 없이 시청할 수 있게 됐다.¹⁴ 현재 프라임 비디오 홈 화면에는 아마존 인터페이스 내에서 프랑스 텔레비지옹의 콘텐츠를 모아 보여주는 전용 france.tv 섹션까지 마련돼 있다.¹⁵ 사실상 아마존의 스트리밍 서비스가 프랑스 공영 채널을 제공하는 새로운 형태의 가상 케이블 사업자가 된 셈이다.

프랑스 텔레비지옹 입장에서는 케이블 TV 시청을 중단한 젊은 시청자층에게 콘텐츠를 노출시킬 수 있다는 것이 가장 큰 이점이다. 시청자들이 제각기 다른 미디어로 분절된 환경에서 인기 스트리머의 메뉴 안에 포함되는 것 자체가 살아남는 방법이기 때문이다.

일부 유럽 민영방송사도 비슷한 방향으로 움직이고 있다. 대표적 사례로 넷플릭스(Netflix)와 유사한 계약을 체결한 프랑스 최대 민영 방송사 TF1을 꼽을 수 있다. 해당 파트너십은 넷플릭스가 세계 어느 지역에서도 시도한 적 없는 첫 사례로, 2026년부터 프랑스의 넷플릭스 가입자는 앱을 벗어나지 않고 TF1의 실시간 방송을 시청할 수 있게 된다.¹⁶ 현재 프랑스에서 진행 중인 이 실험은 유럽 전역의 미디어 기업 리더들이 예의주시하고 있다. 결국 거대한 스트리머를 이길 수 없다면 그들과 손잡는 것이 차선책일 수 있다.

4. 스트리밍 파트너와 손잡고 글로벌 성공 거둔 영국 BBC와 채널4

한때 BBC나 채널4(Channel 4) 로고가 붙은 프로그램은 전적으로 영국 국내 제작물이라는 의미였지만, 이제는 넷플릭스, 아마존 프라임 비디오, HBO 맥스(HBO Max) 같은 글로벌 스트리머와 공동 제작한 콘텐츠 일 가능성성이 크다.

공영방송사는 이 같은 글로벌 스트리머와 공동 제작하는 방식으로 스트리머의 풍부한 자본과 국제 유통망, 영국식 프리미엄 스토리텔링에 대한 수요를 활용해 재정적으로 부담스러울 수 있는 대형 프로젝트를 추진할 수 있다.¹⁷ 미디어 산업 단체에 따르면, 제3자 자금 조달은 공동 제작 계약, 글로벌 사전 판매, 세액 공제 등의 방식으로 영국 공영방송사 제작을 지원하고 있으며, 지원 규모는 매년 4억 파운드에 달한다.¹⁸ 구체적으로 넷플릭스나 아마존 같은 플랫폼이 글로벌 스트리밍 권리를 얻는 대가로 제작비의 상당 부분을 부담하는 구조다. 이는 영국 방송사의 위험을 줄여줄 뿐 아니라, 국내 히트작을 영국을 넘어 전 세계 시청자에게 노출하는 데 큰 도움이 된다.

BBC는 특히 대하드라마 시리즈에서 이러한 협력 방식을 적극 활용하고 있다. 영국 판타지 및 아동 문학 작가 필립 폴먼(Philip Pullman)의 소설을 원작으로 한 판타지 서사극 『황금 나침반』(His Dark Materials)은 BBC와 HBO가 공동 제작한 작품으로, 첫 번째 시즌 제작비만 5,000만 파운드에 달한 것으로 알려졌다.¹⁹ BBC는 HBO의 투자 덕분에 영화적 스케일을 실현할 수 있었고, HBO는 미국 시장에 선보일 수 있는 고품질

프로그램을 확보했다.²⁰ 전형적인 영국 시대극도 이러한 방식으로 노출이 확대됐다. 1차 세계대전 직후를 배경으로 한 범죄 드라마『피키 블라인더스』(Peaky Blinders)는 영국 내에서는 BBC가 방송하고, 해외 배급은 넷플릭스가 맡아 글로벌 히트작이 됐다. 채널4 역시 글로벌 스트리머들과의 협업으로 다수의 글로벌 성공 사례를 남겼다.²¹

BBC와 채널4는 넷플릭스, 아마존 등과 협력함으로써 영국 공영방송 콘텐츠를 치열한 21세기 미디어 환경에서 전 세계 스트리밍 시청자들의 ‘좋아요’를 받으며 공유와 몰아보기 콘텐츠로 부상시킨 대표적 사례로 기록됐다.²² 캐나다 CBC처럼 채널4도 유튜브에 전체 에피소드를 게시해 시청자 기반을 점진적으로 확대해왔다.²³ 한편 디즈니플러스(Disney+)와 영국 방송사 ITV는 2025년 7월 콘텐츠를 교환해 시청자들에게 ‘체험’(taster) 형태로 소개하는 파트너십을 발표했다. 이 계약에 따라 ITV의 스트리밍 서비스 ITVX는 디즈니플러스의 인기 타이틀을 순환 방식으로 제공하고, 디즈니플러스 영국판은 ITV의 인기 프로그램을 엄선해 제공한다. 양사는 해당 파트너십에 대해 상호 이익이 되는 실험이라 설명하며, 이는 스트리밍 전쟁이 잣아들고 전략적 동맹이 새로운 추세로 떠오르고 있음을 신호라고 평가했다.

공영방송사-스트리머 파트너십에서 주의해야 할 리스크

공영방송사가 글로벌 플랫폼과 제휴를 맺으면 더 큰 가시성과 재원을 얻을 수 있지만, 동시에 상당한 위험도 수반한다. 세심하게 관리하지 않

으면 공영방송의 자율성이 약화되고 브랜드 정체성이 모호해지며 공적 서비스라는 본연의 사명을 훼손할 수 있다. 이러한 제휴는 공영방송이 빅테크와 거대 자본의 영역으로 들어가는 일이니만큼, 다음과 같은 리스크를 고려해야 한다.

- **통제력과 독립성 상실:** 유통과 수익이 외부 플랫폼을 통해 이뤄질 경우, 공영방송사는 알고리즘 변화나 기업 전략 변화에 종속될 위험이 있다. 단 한 번의 계약 변경으로도 시청자 접근권이나 콘텐츠 권리를 잃을 수 있다. 플랫폼 알고리즘이 방송사의 콘텐츠 노출을 낮추기로 결정한다면 공영방송사의 도달 범위는 하룻밤 사이에 급감할 수 있다.
- **시청자와의 직접적 관계 약화:** 다른 플랫폼에서 공영방송사 콘텐츠를 소비하는 시청자는 공영방송사의 자체 플랫폼을 방문하지 않게 될 가능성이 크다. 이는 브랜드 가시성 감소는 물론, 중요한 시청 데이터 접근권도 줄이는 결과를 낳는다. 광고 기반 공영방송사의 경우 수익화 잠재력이 약화될 수 있다.
- **편집권 및 공적 사명 훼손:** 공영방송사는 클릭 수, 스트리머 자금, 글로벌 확산을 좇는 과정에서 본연의 공적 책무에서 이탈할 수 있다. 알고리즘이나 플랫폼 비즈니스 목표에 뉴스를 맞추고 다큐멘터리를 조정하려는 유혹이 발생해 편집 독립성과 지역 특유의 문화적 함의를 상실할 수 있다.

• 편집 독립성과 국가 주권 훼손: 해외 스트리밍 대기업과의 계약은 정치적 반발을 축발한 경우가 있고, 미국 기업이 자국 공영 콘텐츠에 영향력을 행사하는 것에 대해 우려를 표하는 제작자와 정치인도 적지 않다.²⁴ 외부 상업 자본에 과도하게 의존하면 수신료나 세금 기반 재원의 정당성도 약해질 수 있다.

• 재정적 의존성과 지속 가능성 문제: 스트리밍 플랫폼 자금에 과도하게 의존하면 해당 플랫폼이 방향을 바꾸거나 투자를 철회할 때 큰 타격을 입을 수 있다. 공영방송은 대표 프로그램 조차 불안정한 토대 위에서 예산을 짤 위험을 안게 된다.

공영방송은 변화하는 시청자 행동·참여·펀딩 환경에 맞춰 혁신을 추진하고 있지만, 스트리머 파트너십은 통제, 정체성, 지속 가능성 측면에서 새로운 위협을 가져온다. 공영방송사는 스트리머와 손을 잡음으로써 독립성과 시청자와의 직접적 관계라는 가치를 단기적 자금 확보 및 시청자 기반 확대라는 이익과 맞바꾸고 있는 셈이다. 이처럼 위험한 줄다리기에서 균형을 잡지 못하면, 공영방송은 스트리머라는 짐승을 제대로 다루지 못하고 오히려 역공을 받을 수 있다.

공영방송사들은 가치와 이익 간 트레이드오프를 효과적으로 관리하기 위해 다음을 고려해야 한다.

• 브랜드와 인지도 보호: 제3자 플랫폼에서도 로고와 출처 표기를 명확하고 엄격하게 유지해 신뢰와 인지도를 확보해야 한다.

• 데이터 접근권과 비례적 수익 배분 확보: 시청 통계 접근권과 적정 보상을 확보해 협상력을 유지해야 한다.

• 동맹 구축: 영국 공영방송사가 공동으로 만든 무료 TV 생방송 및 주문형 스트리밍 서비스 프리리(Freely)처럼 공동 이니셔티브를 통해 공영방송사가 글로벌 스트리머와 경쟁력을 유지할 수 있도록 해야 한다.

• 공적 가치를 담은 콘텐츠 유지: 지역 뉴스와 교육, 문화, 소수 언어 프로그램 등 글로벌 흥행 여부와 무관한 공적 가치 콘텐츠에 지속 투자해야 한다.

• 목적 있는 혁신: 글로벌 플랫폼 기술에서 배우되, 이렇게 얻은 노하우를 자사 디지털 서비스 강화를 위해 활용해야 한다.

미국 신생 스트리밍 채널을 위한 전략

쇠퇴하는 TV 사업을 경쟁력 있는 스트리밍 서비스로 전환하려는 미국 미디어 기업들에게 영국과 EU의 공영방송은 더 유연하고 혁신적 경로를 제시한다. 앞서 언급한 다양한 위험 요인들을 신중히 고려함과 동시

에 영국과 EU 공영방송사들의 여정을 살펴보면, 거대 경쟁자들에 맞서 고군분투하는 미국 스트리머들에게 하나의 전략적 교본이 될 수 있다.

- 전략적 파트너십을 활용해 레거시(legacy, 이미 자산화된 기존 콘텐츠) 및 니치(niche, 특정 타깃형 콘텐츠) 콘텐츠가 노출되는 시청자 범위를 확대하라. 최대 플랫폼과 파트너십을 맺으면, 시청자나 지식재산(IP)을 뺏기는 것이 아니라, 휴면 시청자층을 다시 불러들이고 기존에 가입 이력이 없는 새로운 시청자에게 브랜드와 IP를 소개하는 기회가 될 수 있다.
- 가치 있는 파트너십을 구축해 핵심 콘텐츠를 기반으로 인지도를 높여라. 방송사들은 현지 콘텐츠를 글로벌 유통권 확보를 위한 협상 카드로 활용해왔다.²⁵ 민간 제작사도 인기 IP를 활용해 새로운 가입자를 확보할 수 있다. 예를 들어 영국 방송사 ITV는 자사 로컬 히트작인 『러브 아일랜드』(Love Island)를 활용해 디즈니플러스의 화제작 『더 베어』(The Bear)를 자사 플랫폼으로 끌어왔고, 디즈니는 『더 만달로리안』(The Mandalorian)의 팬덤을 활용해 ITV 시청자들이 디즈니플러스를 경험하도록 유도했다. 이처럼 민영 방송사도 공영 방송사처럼 글로벌 스트리머와의 공동 제작을 더욱 적극적으로 고려할 수 있다.
- 브랜드와 데이터를 지키되, 실용적으로 접근하라. 공영 방송사처럼 민영 미디어 기업도 시청자를 따라가며 존재감을 유지해야 하고, 규모를 확보하기 위해 파트너십이 필요하며, 타 플랫

폼에서 운영되더라도 고유한 정체성을 지키는 것이 중요하다는 점을 인식하고 있다. 선형 TV 시청률과 광고 수익이 감소하고 많은 스트리밍 서비스가 아직 수익을 내지 못하고 있는 상황에서, 이러한 파트너십은 전술적 실험을 넘어 핵심 전략으로 발전할 가능성이 있다. 제대로만 실행된다면 공영 방송사의 경험은 이러한 제휴가 충분히 추가 가치를 창출하고 재정적으로도 유리할 수 있음을 보여준다.

공영방송사와 미국 스트리머가 고려해야 할 핵심 사항

스트리밍 혁명과 소셜 비디오에 밀려 소멸될 것처럼 보였던 많은 공영 방송은 오히려 이를 활용한 재창조의 여정을 걷고 있다. 소셜 플랫폼으로 콘텐츠를 확장하고, 대형 스트리머와 공동 제작하며, 심지어 스트리머가 자사 채널을 송출하도록 허용하는 방식으로 진화하고 있는 것이다. 이러한 변화가 제대로 이뤄진다면 공적 서비스 콘텐츠가 모든 플랫폼에서 상업 콘텐츠와 공존하는 더욱 풍부한 미디어 생태계가 조성될 수 있다. 또한 공영 방송이 글로벌 플랫폼에 지역성과 공공성을 불어넣어 균형을 잡는 역할도 하게 될 것이다.

공영 방송사는 영리 기업과의 파트너십 속에서 공적 사명을 수행 및 수호하는데 상당한 어려움을 겪고 있음에도 불구하고 본고에서 소개한 바와 같이 다양한 혁신 사례를 남기고 있다. 이들의 경험은 변화 압력에 직면한 미국의 중소형 제작사 및 니치 스트리머들도 충분히 적용할 수

있다. 스트리밍 비디오는 TV를 해체하고 재편했고, 소셜 비디오 플랫폼은 TV와 스트리밍 서비스의 시청자를 빼앗아 가고 있으며, 가장 거대한 비디오 유통 플랫폼은 TV의 개념 자체를 계속 재편하고 있다. 공영방송과 민간 미디어 모두 부단히 실험적 시도에 도전하고 변화에 적응해야만 살아남을 수 있다.

Korean Perspectives

플랫폼 다변화 시대에 국내 미디어 산업의 선택과 전략

글로벌 미디어 환경이 급격히 재편되면서 국내 방송·미디어 산업은 구조적 전환의 압력을 받고 있다. 스트리밍 플랫폼과 소셜미디어 중심의 소비 확산, 글로벌 OTT와의 협력 확대는 기존 방송 중심 유통 구조의 한계를 분명히 드러내고 있다. 이제 콘텐츠 경쟁력은 채널이 아니라 플랫폼 전략과 글로벌 유통 역량에서 결정되는 국면에 진입했다고 할 수 있다. 이러한 변화는 국내 미디어 산업의 성장 경로를 근본적으로 재정의하고 있다. 소셜미디어와 글로벌 스트리밍 플랫폼을 효과적으로 활용할 경우, 젊은 세대와 해외 시청자에게 보다 직접적으로 도달할 수 있으며, 이는 K-콘텐츠의 글로벌 확산 속도와 범위를 크게 넓힐 수 있는 계기가 된다. 반면에 외부 플랫폼에 대한

의존이 심화될수록 알고리즘 변화, 수익 배분 구조, 데이터 접근권 제한 등 통제 불가능한 변수에 노출될 가능성도 함께 커질 수 있다.

공영방송 역시 이러한 변화에서 예외가 아니다. 뉴스·교육·문화 등 공적 가치 콘텐츠를 유지하는 역할은 여전히 중요하지만, 플랫폼 다변화와 전략적 파트너십을 통해 영향력을 확장하지 못할 경우 공영방송의 사회적 존재감 자체가 약화될 가능성이 커지고 있다. 공공성과 경쟁력 사이의 균형을 새롭게 정의해야 하는 이유다. 이러한 환경에서 국내 방송·미디어 산업의 대응 방향은 비교적 명확하다. 단일 플랫폼에 대한 의존에서 벗어나 글로벌 OTT와 소셜미디어를 아우르는 다중적 유통 전략을 구축해야 하며, 제휴와 협력 과정에서도 브랜드 노출과 데이터 접근권을 명확히 확보해 장기적인 경쟁력을 유지해야 한다. 동시에 공영방송을 중심으로 공적 가치 콘텐츠에 대한 투자를 지속하며, 상업적 성과와 공공성 간의 균형을 제도적으로 뒷받침할 필요가 있다. 아울러 속품·인터랙티브 콘텐츠, 소셜 전용 포맷 등 새로운 형식에 대한 실험을 확대하고, 빠르게 시도하고 학습하는 조직 문화를 정착시키는 것이 중요하다.

결국 국내 방송·미디어 산업의 향방은 플랫폼 다변화, 브랜드·데이터 주권 확보, 그리고 공적 가치와 혁신의 균형이라는 세 축을 어떻게 결합하느냐에 달려 있다. 변화하는 글로벌 미디어 환경 속에서 전략적 협력과 지속적인 실험을 이어갈 수 있다면, K-콘텐츠는 일시적 흥행을 넘어 장기적 경쟁력을 갖춘 글로벌 문화 자산으로 자리매김할 수 있을 것이다.



박형곤 파트너

한국 딜로이트 그룹
TM&E Sector 리더

9장

숏폼이 촉발한 콘텐츠 제작 생태계의 혁신과 플랫폼 신뢰 위기

모바일을 중심으로 확산되고 있는 숏폼 시리즈와 마이크로 드라마는 콘텐츠 소비 방식을 근본적으로 바꾸고 있으며, 인공지능(AI)과 데이터 기반의 창작 생태계 전반에 변화를 가져오고 있다. 이 같은 변화는 플랫폼 구조와 수익 모델의 진화를 촉진하는 한편, 콘텐츠 신뢰성 저하와 창작자-플랫폼 간 권력 재편이라는 새로운 리스크를 동반한다. 결국 콘텐츠 플랫폼의 경쟁력은 연속적 서사를 지원하는 구조 설계와 신뢰 회복 역량에 달려 있다.

핵심 내용 요약 (Executive Summary)

▶ 숏폼 시리즈의 급성장과 글로벌 확산

- 마이크로 드라마·마이크로 시리즈 등 초단편 에피소드형 콘텐츠가 모바일 중심으로 급성장
- 인앱 매출은 2025년 38억 달러 → 2026년 78억 달러로 두 배 이상 성장 전망
- 중국·아시아·미국을 중심으로 이용자 규모와 매출이 동시에 확대

▶ 콘텐츠 소비 방식의 변화

- 간헐적·즉흥적 소비(포리저·스내커·그레이저형)가 소셜미디어 전반으로 확산
- 짧고 빠른 소비는 강화되는 반면, 깊은 몰입은 점차 어려워지는 구조 고착화

▶ 창작자·제작 생태계의 혁신

- AI·데이터·플랫폼 활용으로 개인 창작자와 독립 스튜디오의 저비용·고품질 제작 역량 강화
- 시청자 반응 데이터를 기반으로 빠른 제작·실험·수정이 가능한 구조 정착
- 글로벌 진출과 함께 팬덤·커뮤니티 중심의 성장 모델 실현

▶ 플랫폼 재편 압력과 사회적 리스크

- 시리즈 인식 알고리즘, ‘계속 시청하기’ 등 연속 서사를 지원하는 플랫폼 구조 재설계 필요성 확대
- 대응이 지연시 창작자·시청자의 독립 앱 및 스트리밍 서비스 이탈 가능성 증가
- ‘브레인 로트’, ‘돔스크롤링’ 등 저품질 콘텐츠 과소비로 인한 인지적 피로와 신뢰 저하 우려
- 생성형 콘텐츠 확산으로 정보 진위 판단이 어려워지며 소셜미디어 신뢰도 하락

개인 창작자부터 글로벌 대형 플랫폼까지 솟폼(short-form) 콘텐츠^{*}가 주요 스토리텔링 포맷으로 부상하면서, 전 세계 시청자가 콘텐츠를 소비하는 방식이 재정의되고 있다.

* 솟폼(short-form) 콘텐츠는 일반적으로 수 초에서 수 분 이내의 짧은 길이로 제작되는 디지털 콘텐츠를 뜻한다.

‘포리저(forager), 스내커(snacker), 그레이저(grazer)’는 규칙적인 식사를 하지 않고 하루 종일 조금씩 즉흥적으로 간식 위주로 음식을 소비하는 패턴을 뜻한다. 이는 콘텐츠의 흥수 속에서 끊임없이 영양가 없는 무언가를 찾아 헤매는 소셜미디어 이용자들의 콘텐츠 소비 패턴을 설명하기에도 적합한 표현이다. 이용자들은 자신의 메타데이터에는 부합하지만 정서적, 지적 욕구를 충족시키지 못하는 무한한 스트림을 끝없이 헤집고 다닌다. 그렇다면 마이크로 시리즈나 마이크로 드라마처럼 연속적 구조를 가진 솟폼 스토리텔링이 산만하게 분절된 주의력 경제(attention economy) 속에서 몰입과 지속성을 강화할 수 있지 않을까?

마이크로 드라마 또는 솟폼 연재물이라 불리는 마이크로 시리즈는 몇 분짜리 초단편 에피소드로 구성된 극본 기반 비디오 시리즈로, 모바일 우선 소비와 빠른 참여를 염두에 두고 설계된 형식이다. 드라마박스(DramaBox), 릴쇼츠(ReelShort), 솟맥스(ShortMax), 드라마웨이브(DramaWave) 같은 모바일 앱은 아시아와 미국에서 수십억 달러의 수

익과 수억 명의 이용자를 확보하며 폭발적으로 성장하고 있다.¹ 이들의 급성장은 디지털 엔터테인먼트에 대한 시청자 기대를 재정의하고, 창작자·플랫폼·브랜드 모두에게 새로운 기회와 과제를 제시한다.²

2025년 기준 마이크로 시리즈 콘텐츠의 인앱(in-app) 매출은 38억 달러에 이를 것으로 전망된다.³ 2026년에는 인앱 마이크로 시리즈의 매출 성장률이 두 배 이상 증가해 약 78억 달러에 달할 것으로 예상된다. 또한 2025년에는 미국이 글로벌 매출의 절반을 차지할 것으로 예상되지만, 여타 시장이 조회 수와 다운로드를 수익으로 전환하면서 미국 비중은 40%로 감소할 전망이다. 더 많은 시청자가 마이크로 시리즈에 노출될수록, 솟폼과 연속 드라마가 결합된 형식이 더욱 인기를 끌게 될 가능성이 크다. 소셜미디어에서의 높은 바이럴리티(virality)*도 이를 뒷받침할 것이다. 더불어 향후 소셜 플랫폼 전반에서 더 많은 마이크로 시리즈 콘텐츠가 돌풍을 일으키며, 이용자들의 콘텐츠 소비 시간(attention time)을 더 많이 확보하고 미국 소셜미디어 참여도 순위 상위권에 오를 것으로 보인다. 일부 스트리밍 서비스는 자사 플랫폼 내에서 솟폼 연재 콘텐츠를 시험적으로 도입할 가능성도 있다.

* 바이럴리티(virality)는 콘텐츠·메시지·아이디어가 유료 광고나 중앙 통제 없이 이용자간 자발적 공유를 통해 급속도로 확산되는 현상 및 그 확산력을 의미한다. 전염병이 퍼지는 양상에 비유해, 확산 속도·범위·지속성을 핵심 지표로 삼는 커뮤니케이션 성과 개념이다.

マイクロシリーズ는 그야말로 소설을 위해 태어난 포맷처럼 보이지만, 동시에 소설 플랫폼의 지배력을 위협할 수도 있다. 소설 플랫폼의 예측 불가능한 알고리즘 피드 특성상 특정 시리즈를 꾸준히 ‘팔로우’하고 신작 에피소드를 놓치지 않고 보는 것이 어려우므로, 경쟁 앱인 마이크로 시리즈 플랫폼으로 이동하는 시청자와 창작자가 증가할 수 있기 때문이다. 한편 젊은 세대가 소설미디어가 쏟아내는 압도적인 정보량에 피로감을 느끼며 무한 스트림을 따라잡지도 그렇다고 벗어나지도 못하는 상황이라는 신호도 나온다.⁴ 그렇다면 솟폼으로 되살아난 미니 시리즈 형태의 TV형 콘텐츠가 해법이 될 수도 있다.

시청자 사로잡는 솟폼 시리즈

연속극 형태 솟폼 콘텐츠는 높아지는 인기와 더불어 주요 소설 비디오 플랫폼에서 차지하는 비중도 확대되고 있다. 즉, 소비자들이 디지털 엔터테인먼트에 소비하는 유한한 시간을 두고 마이크로 드라마 서비스라는 새로운 경쟁 요인이 등장한 것이다.

이에 따라 마이크로 드라마 모바일 앱의 인기도 빠르게 상승하고 있다. 이들 앱은 60~90초 길이의 연재 에피소드 수백 편을 제공하며, 반전에 반전을 더한 전개와 극적 미완 결말을 통해 시청자를 계속 붙잡아두고 다음 에피소드를 보게 만든다. 신규 마이크로 시리즈는 빠르고 저렴하게 제작되며, 시청자 반응에 따라 끊임없이 개선된다. 또한 주요 소설 비

디오 플랫폼을 적극 활용해 콘텐츠의 발견성(discoverability)*과 바이럴 효과를 높이며 성장세를 가속화하고 있다.⁵

* 콘텐츠의 발견성(discoverability)은 이용자가 의도적으로 검색하지 않더라도, 플랫폼·알고리즘·추천·유통 구조를 통해 콘텐츠가 얼마나 쉽게 노출되고 발견되는지를 나타내는 정도를 뜻한다.

**그림 1
마이크로 드라마, 아시아 내 인기를 발판으로 글로벌 성공 달성**

중국 언론 보도에 따르면,
2024년 기준 중국 내
**마이크로 드라마 시청자 수는
6억6,200만 명**

2025년 초 기준
아시아 Top5 마이크로 드라마 앱의
**월간 활성 사용자 수는
1,500만 명**

2024 2025

출처: Omdia, “Emerging Micro-drama Trend in Asia,” March 25, 2025.

중국의 아이치이(iQiyi)는 무료 및 유료 마이크로 드라마 1만5,000편 이상을 제공하고 있으며, 지난 1년 동안 시청 시간이 크게 증가했다. 아이치이는 마이크로 드라마를 중심으로 한 디지털 콘텐츠 생태계에 이커머스 기능까지 결합하며 지속적으로 서비스를 확장하고 있다.⁶ 중국 현지 매체는 2024년 기준 전국 마이크로 드라마 이용자가 약 6억6,200만 명에 달한다고 보도했다.⁷ 중국의 주요 비디오 스트리머들은 젊은 동영상 플랫폼과 협력해 프리미엄 미니 드라마를 공동 제작하고 있는데, 이는 롱폼·숏폼 콘텐츠가 통합된 미래를 예상한 행보로 해석된다.⁸

전 세계적으로 마이크로 드라마의 성장세가 가속화되자, 인도의 혁신적 플랫폼과 기존 미디어 기업들이 성장 기회를 찾아 잇따라 시장에 진입하고 있다.⁹ 지 엔터테인먼트(Zee Entertainment), 쿠쿠FM(Kuku FM) 같은 OTT(over-the-top, 인터넷 기반 영상 콘텐츠 제공 서비스) 플랫폼은 마이크로 드라마 전용 카테고리를 출시했다. 일부 플랫폼은 솟폼 파일럿 콘텐츠 도입 이후 일일 시청 시간이 두 배 증가했다고 보고 했다.¹⁰ 인도 시장은 평균 이용자당 평균 수익이 낮고 가격 민감도가 높기 때문에, 플랫폼은 개별 에피소드에 대한 마이크로 결제 옵션을 실험하고 있으며, 일부는 구독료와 광고 수익을 혼합한 하이브리드 구독 모델을 도입하고 있다.¹¹ 시청자는 초기 몇 개의 에피소드는 무료로 감상할 수 있지만, 이후 에피소드를 계속 보려면 비용을 지불해야 한다.

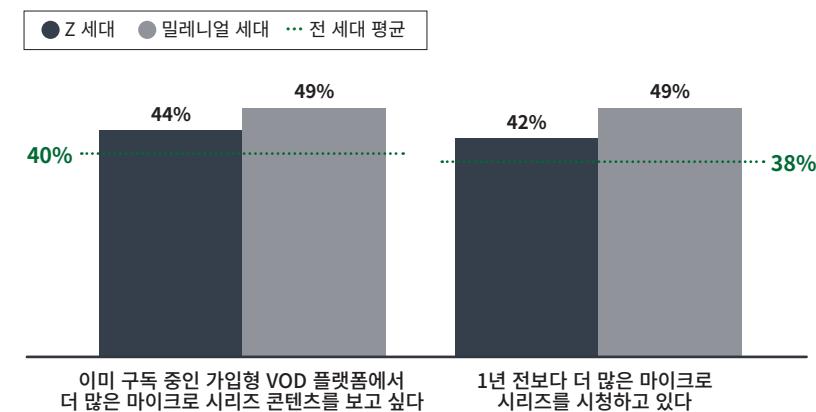
마이크로 드라마 앱과 솟폼 연재물에 대한 수요는 아시아를 넘어 전 세계로 확산되고 있다.¹² 최근 조사에 따르면 글로벌 마이크로 드라마 앱 매출은 2024년 1분기 1억7,800만 달러에서 2025년 1분기 약 7억 달러로 급증했다.¹³ 미국은 드라마박스, 릴쇼츠, 굿쇼츠(GoodShort) 등 솟드라마 앱의 최대 시장으로 부상했으며,¹⁴ 주요 마이크로 드라마 앱이 미국 앱스토어 다운로드 상위 25위권에 자주 진입하고 있다.¹⁵ 이들 앱은 소셜 플랫폼과의 크로스오버(crossover, 플랫폼간 연계) 노출을 통해 미국 시장에서 존재감을 강화하고 있으며,¹⁶ 소셜 플랫폼 자체에서도 마이크로 시리즈 콘텐츠의 참여도 증가가 확인되고 있다.¹⁷ 주요 스트리밍 서비스 역시 솟폼 및 버티컬 영상(vertical video, 세로형 화면 비율로 제작된 영상) 콘텐츠를 시험적으로 도입하며 시장 분위기를 탐색하고 있다.¹⁸

딜로이트의 ‘디지털 미디어 트렌드’(Digital Media Trends) 조사에 따르면, 2025년 3월 기준 미국 Z 세대와 밀레니얼 세대 중 약 30%가 마이크로 시리즈 및 마이크로 드라마에 대해 인지하고 있는 것으로 나타났다. 또한 이들 중 거의 절반은 1년 전보다 더 많은 마이크로 시리즈를 시청하고 있다고 답했으며, 약 절반은 자신이 이미 구독 중인 가입형 VOD(video-on-demand, 주문형 비디오) 플랫폼에서 더 많은 마이크로 시리즈 콘텐츠를 보고 싶다고 밝혔다. 이는 스트리밍 서비스가 경쟁력을 강화할 수 있는 새로운 경로를 제시한다.

그림 2

Z 세대와 밀레니얼 세대, 여타 세대에 비해 연속극형 솟폼 비디오 더 많이 접한다

마이크로 드라마나 마이크로 시리즈에 익숙하다고 답한 응답자 중 다음 항목에 ‘어느 정도 동의한다’ 또는 ‘강력히 동의한다’고 답한 비율



참조: 마이크로 드라마에 대해 인지하고 있다는 응답자는 1,620명으로 전체 응답자의 19.5%에 불과함. 이 중 Z 세대가 483명, 밀레니얼 세대가 497명, X 세대가 399명, 베이비부머가 230명, 고령층이 11명으로 집계됨.

출처: 2025 Digital Media Trends, 19.5 edition.

특히 개인 창작자가 만든 서사 중심 속품 콘텐츠가 큰 인기를 끌고 있다. 이에 대응해 점점 더 많은 창작자들이 데이터, 인공지능(AI), 소셜 플랫폼을 활용해 독립 스튜디오를 설립해 더 많은 시청자층에 다가가고 있다.

소셜 중심, 데이터 기반, 비용 효율적 독립 스튜디오의 부상

미디어·엔터테인먼트 산업은 속품 콘텐츠의 소비 행태와 경제성, 소셜 비디오 플랫폼의 역량과 도달 범위, 프리미엄 콘텐츠의 명성과 높은 제작비가 공존하는 이중 구조에 의해 재편되고 있다. 시청자들이 속품 콘텐츠에 쓰는 여가 시간이 늘고 속품 콘텐츠에 더 높은 가치를 부여함에 따라, 개인 창작자들이 현대판 스튜디오로 성장하고 있다. 또한 과거 제작비와 비교해 극히 적은 비용만 들여도 고품질의 독립 영상 제작이 가능해졌다.

이처럼 새로운 형태의 독립 스튜디오는 단순히 경제적 이점만을 좇거나 변화하는 시청자 행태를 이용하는 데 그치지 않는다. 이들은 빠르고 기민하게 움직이며, 시청자 피드백에 즉각 반응한다. 또한 시청자 참여 데이터를 활용해 어떤 콘텐츠가 효과적인지 또는 효과적이지 않은지를 확인해 콘텐츠 제작의 의사결정 리스크를 줄인다.¹⁹ 시청자와 직접 소통하며 커뮤니티 유대감과 팬덤을 강화함과 동시에, 출시 속도를 단축하고 제작 비용을 줄이며 AI를 적극 활용해 글로벌 영역으로 범위를 확대한다.²⁰ 더불어 시청 유지와 몰입도를 극대화하기 위한 실험적 편집 및 연출 기법을 자유롭게 시도할 수 있다.

마이크로 시리즈에 대한 시청자 참여가 증가할수록, 기술로 확장력을 강화하고 시청자 상호작용을 기반으로 한 창작자 스튜디오가 증가할 것이다. 이로 인해 창작자 인재를 확보하기 위해 마이크로 드라마 앱, 스트리밍 서비스, 소셜 비디오 플랫폼 간 경쟁이 더욱 치열해질 것으로 예상된다. 특히 소셜 플랫폼은 시청자가 연재 콘텐츠를 더 쉽게 찾고 계속 팔로우할 수 있는 환경을 제공하지 않는다면 상대적으로 경쟁에서 뒤처질 수 있다.

속품 제작을 뒷받침하는 툴과 실행 전략

개인 창작자뿐 아니라 미디어 기업의 리더들은 새롭게 등장하는 독립 스튜디오가 첨단 툴과 플랫폼을 활용해 최소 비용으로 글로벌 시청자에게 도달하고 이들의 참여를 이끌어내는 방식을 면밀히 살펴볼 필요가 있다.

- **AI 기반 제작 파이프라인:** 독립 스튜디오는 생성형 AI 도구를 활용해 제작 사이클을 단축하고 높은 제작 품질을 구현하고 있다. 전반적으로 영상 제작의 진입 장벽이 낮아진 것이다. 예컨대 자동 B롤(B-roll, 보조 영상) 생성, 간단한 애니메이션 제작, 자동 자막 생성, 보이스 클로닝(voice cloning)*, 더빙 AI 등을 활용하면 언어·방언·억양 장벽을 쉽게 극복할 수 있다.²¹ 장편 서사를 속품 연재 콘텐츠로 변환해 주는 툴도 활용할 수 있다.

* 보이스 클로닝(voice cloning)은 AI가 특정 인물의 음성 특성(발음, 억양, 음색, 속도, 감정 표현 등)을 학습해 동일하거나 매우 유사한 음성으로 재현하는 기술을 의미한다. 텍스트 입력만으로도 원음 화자의 목소리와 유사한 음성을 생성할 수 있어, 차세대 음성 합성(TTS)의 핵심 기술로 평가된다.

- **편집 전략:** 속품 시리즈는 ‘훅’(hook)*의 밀도를 높이고, 에피소드마다 극적 미완 결말로 마무리해 시청자의 ‘더 알고 싶다’는 충동을 자극한다. 팬픽션(fan fiction)**이나 온라인 소설의 인기 트로프(trope)***, 예컨대 ‘부자-서민 로맨스’나 ‘타임슬립 복수물’ 같은 장르적 관습을 차용해 겸증된 시청자 취향을 적극 공략함으로써 성공을 거둔 사례도 있다.²²

* 훅(hook)은 콘텐츠의 도입부 또는 핵심 구간에서 반전과 폭로 등 놀라운 전개 장치로 시청자나 독자의 주의를 즉각적으로 끌어당겨 이탈을 막고 추가 소비를 유도하는 장치 또는 요소를 의미한다. 주로 놀라움, 반전, 갈등, 의문 제기, 강한 감정 자극을 통해 짧은 시간 안에 몰입을 형성하는 것이 목적이다.

** 팬픽션(fan fiction)은 기존 작품의 세계관·캐릭터·설정 등을 차용해 팬이 자발적으로 창작한 2차 창작 서사 콘텐츠를 의미한다. 원작의 빈틈을 확장하거나, 대안적 전개·관계·결말을 상상함으로써 소비자에서 창작자로의 전환을 보여주는 대표적 문화 현상이다.

*** 트로프(trope)는 서사·장르·문화 전반에서 반복적으로 사용되며 관습적으로 인식되는 이야기 장치, 설정, 인물 유형, 전개 패턴을 의미한다. 완전히 새로운 발명이라기보다, 이미 익숙한 의미 구조를 활용해 이해와 몰입을 빠르게 유도하는 서사적 ‘공통 언어’에 가깝다.

- **커뮤니티 성장:** 창작자는 댓글을 통한 시청자와의 적극적 소통, 팬 피드백을 반영한 후속 에피소드 수정 등을 통해 시리즈를 단단한 팬덤으로 성장시킬 수 있다. 비하인드 스토리 생중계나 소셜미디어 대화 참여도 커뮤니티 결속을 강화하는 요소다. 매일 같은 시간에 에피소드를 공개하는 규칙적 일정은 예약 시청 습

관을 만드는 데 효과적이다. 마이크로 드라마는 반복 등장인물이나 테마를 중심으로 여러 시즌과 스피너오프로 확장되는 미니 드라마 프랜차이즈가 될 수 있다. 소설화나 AI 생성 음악 기반 OST 출시 등 교차 홍보, IP 머천다이징, 멀티 플랫폼 배급을 능숙하게 활용하는 창작자는 예산 제약 속에서도 예컨대 자신만의 ‘미니 마블(Marvel) 우주’를 구축할 수 있다.

- **새로운 핵심 성과 지표(KPI):** 데이터 중심의 피드백 루프를 갖춰 놓으면 복수의 스토리라인을 시험하고, 유지율이 높은 서사에 자원을 집중하는 데 도움이 된다. 완주율과 사용자당 평균 시청 에피소드 수부터 시리즈 구독 전환율, 굿즈 판매 유도 효과 또는 플랫폼 내 시청 시간 증가까지 세부적인 투자수익률(ROI)을 파악하면 KPI를 정확히 측정할 수 있다.
- **수익화:** 에피소드 단위 마이크로 결제, 월 구독료, OST 및 굿즈 판매, 광고, 간접광고 등 다양한 수익 모델을 실험할 수 있다.²³

브레인 로트와 둠스크롤링의 해독제 역할 기대되는 속품

옥스퍼드대학교출판부(Oxford University Press)가 선정한 2024년 올해의 단어는 ‘브레인 로트’(brain rot)였다. 이는 소셜미디어를 중심으로 저품질 온라인 콘텐츠를 과도하게 소비하는 것이 뇌에 악영향을 미쳐 집중력과 사고력이 떨어진다는 의미다.²⁴ 관련 용어인 ‘둠스크롤링’은

링'(doomscrolling)은 암울한 뉴스를 강박적으로 스크롤하며 소비함으로써 불안과 우울을 초래하는 콘텐츠 소비 행태를 뜻한다.²⁵

이러한 우려에 대응하듯, 소셜미디어에서 벗어나 더 작고 더 친밀하며 더 보호된 정보·엔터테인먼트·커뮤니티 공간으로 이동하는 소비자들이 늘고 있다는 신호가 나오고 있다.²⁶ 이는 소셜미디어가 더 이상 '소셜'의 공간이 아니라 단지 '미디어'가 되어 지나치게 파편화되고 상품화됐다는 인식에서 비롯된 반응일 수 있다.²⁷ 혹은 과도한 이용이 정신건강에 미치는 부정적 영향에 대한 우려가 본격적으로 드러난 결과일 수도 있다.²⁸ 또한 생성형 콘텐츠의 확산으로 온라인 정보의 진위 구분이 어려워지면서 신뢰가 추락한 결과일 수도 있다. 혹은 단순히 영양가 없는 간식과도 같은 보상을 찾아 끊임없이 해매는 콘텐츠 소비 방식 자체가 본질적으로 피로감만 주고 만족감을 주지 못하기 때문일 수도 있다.²⁹

マイ크로 드라마가 새로운 대중 미디어의 거대한 변화를 이끈다고 단정 할 수는 없다. 그러나 연속 구조를 갖춘 속풀 독립 콘텐츠에 대한 소비자들의 관심이 증가하면서, 소셜 비디오 플랫폼, 전통 스튜디오, 스트리밍 서비스가 새로운 도전을 받고 있다. 더불어 미디어 권력의 균형이 재편되고 고품질·저비용의 새로운 독립 스튜디오 계층이 부상할 가능성 이 커지고 있다.

창작자 권력의 이동: 알고리즘 의존에서 대안 채널 및 연재형 플랫폼으로 전환

개인 창작자와 독립 스튜디오는 영향력을 키우고 브랜드와의 관계를 강화하며 시청자에게 도달할 새로운 채널을 발굴하고 있다. 소셜 플랫폼에서 시청자 참여 및 수익화 또는 독립 플랫폼 구축에 어려움을 느끼는 창작자들은 비디오 스트리머, 마이크로 드라마 앱, 새로운 창작자 주도형 엔터테인먼트 서비스 등 대안적 채널로 이동할 수 있다.

창작자들은 여러 플랫폼에서 시청자 기반을 축적하며 브랜드와 더 긴밀한 관계를 맺고 있다.³⁰ 하지만 이로 인해 인기 창작자와 그들이 콘텐츠를 게시하는 플랫폼 간에 수익 배분과 콘텐츠 조정을 둘러싼 갈등과 긴장이 발생하기도 한다.³¹

소셜미디어 플랫폼이 '소셜 활동의 공간'에서 '관심을 추구하는 공간'으로 변질되며, 사용자가 명시적으로 팔로우한 대상보다 사용자 상호작용을 기반으로 추천되는 무한 스트림을 제공하는 구조로 바뀌었다. 이로 인해 알고리즘이 다른 콘텐츠를 우선순위로 판단하는 순간, 창작자가 구축해 온 시청자와의 연결관계가 단번에 약화될 수 있다. 창작자들이 시간과 비용을 들여 시청자 및 브랜드 관계를 형성했지만, 알고리즘의 일방적 판단에 따라 콘텐츠 비중이 낮아지거나 심지어 차단되는 사례도 있다.

대체로 소셜미디어의 알고리즘 기반 피드와 관심 그래프는 연속적 내러티브 콘텐츠를 지원하도록 설계되지 않았다. 따라서 마이크로 시리즈의 인기가 계속 상승한다면, 소셜 플랫폼은 이를 반영해 트렌딩 알고리즘을 조정하고, 에피소드 기반 내러티브를 추적·유지할 수 있는 ‘시리즈 인식’(series-aware) 알고리즘이나 ‘계속 시청하기’(continue watching) 메뉴를 도입해야 할 것이다. 이러한 장치가 추가돼야만 창작자가 충성 시청자에게 더 효과적으로 도달하고 이들을 유지할 수 있으며, 미디어 소비 역학의 전환을 달성할 수 있다.

반대로 이러한 변화가 이루어지지 않을 경우, 모바일 다운로드 상위를 휩쓰는 마이크로 드라마 앱 등 창작자 스튜디오 전용 앱으로 옮겨가는 창작자가 증가할 것이다. 또한 짧은 시청자를 겨냥해 솟폼 콘텐츠를 확보하려는 스트리밍 비디오 서비스가 인기 창작자들을 영입하려 할 수도 있다. 사실 스트리밍 서비스는 원래부터 연재형·예약형 콘텐츠를 기반으로 서비스를 구축해왔기 때문에 솟폼 연재물과 매우 자연스럽게 맞닿아 있다. 앞으로 솟폼 연재 제작을 본격화하는 창작자 스튜디오가 늘어난다면, 스트리머들이 최대 수혜자가 될 수도 있다. 혹은 지금이야말로 스트리머와 소셜미디어가 기술·운영 경험을 결합해 새로운 창작자 주도형 플랫폼을 구축할 적기일 수도 있다.

Korean Perspectives

숏폼 시대의 플랫폼 경쟁력: 연속 경험·창작자 파트너십·신뢰의 통합 설계

모바일 중심의 솟폼과 마이크로 콘텐츠 확산은 국내 콘텐츠 플랫폼 산업의 경쟁 구도를 구조적으로 재편하고 있다. 콘텐츠 소비는 점점 더 짧고 즉각적인 형태로 이동하고 있지만, 플랫폼의 성패는 단순한 ‘짧은 체류 시간’을 얼마나 많이 확보하느냐가 아니라, 단절된 솟폼 소비를 연속적인 경험과 관계로 전환할 수 있느냐에 달려 있다.

알고리즘, UI/UX, 제작·유통 방식, 수익 모델이 각각 분리된 기능이 아니라 하나의 통합된 설계로 작동해야 하는 이유다. 이러한 변화는 국내 콘텐츠 플랫폼 기업에 분명한 전략적 시사점을 제공한다. 첫째, 솟폼을 단발성 트래픽 유입 수단으로만 활용할 것이 아니라, 시리즈 인식 알고리즘, ‘계속 보기’ 구조, 팬덤과 커뮤니티 연결을 통해 몰입을 설계하는 플랫폼 구조로 진화해야 한다. 솟폼의 강점인 접근성과 속도를 유지하되, 이용자가 플랫폼 안에서 머무르고 축적되는 경험을 만들어내는 것이 핵심이다. 둘째, AI와 데이터 기반 제작 환경이 보편화되면서 창작자의 협상력은 빠르게 플랫폼으로부터 이동하고 있다. 개인 창작자와 독립 스튜디오는 저비용·고효율 제작과 빠른 실험이 가능해졌고, 이는 플랫폼에 단순 입점하는 공급자가 아니라 공동 기획과 성과를 공유하는 파트너로서의 지위를 요구하게 만든다.

플랫폼은 수익 배분, 데이터 접근, 멀티 플랫폼 배급을 포함한 관계 재설계를 통해 창작자 중심 생태계를 선점해야 한다. 셋째, 수익 구조 역시 단일 모델로는 한계가 뚜렷하다. 광고와 구독을 기본으로 하되, 에피소드 단위 결제, IP 확장(OST·굿즈·2차 저작), 커뮤니티 기반 유료 기능 등 콘텐츠 성격에 맞는 조합형 수익 모델을 정교하게 운영할 필요가 있다. 이는 플랫폼의 수익 안정성을 높이는 동시에, 창작자와의 이해관계를 장기적으로 정렬하는 장치가 된다. 마지막으로, 속품 확산과 생성형 콘텐츠 증가는 플랫폼 신뢰를 새로운 경쟁 변수로 끌어올리고 있다. 저품질 콘텐츠 과잉, 정보 진위 판단의 어려움, 이용자 피로도 증가는 플랫폼 전반의 가치 훼손으로 이어질 수 있다. 콘텐츠 품질 관리, 정보 신뢰성 확보, 이용자 보호 장치는 더 이상 규제 대응 차원이 아니라 플랫폼 브랜드의 핵심 자산이 되고 있다.

결과적으로 국내 콘텐츠 플랫폼 기업의 과제는 속품이라는 형식에 적응하는 데 그치면 안된다. 소비 구조, 창작자 생태계, 수익 모델, 신뢰를 하나의 시스템으로 설계하는 능력이 장기 경쟁력을 좌우한다. 이 균형을 선제적으로 구축하는 플랫폼만이 속품 시대에도 지속 가능한 성장과 글로벌 확장의 기반을 확보할 수 있을 것이다.



박형곤 파트너
한국 딜로이트 그룹
TM&E Sector 리더

10장

생성형 AI 영상이 불러온 콘텐츠 제작 환경의 혁신과 소셜 플랫폼의 책임

생성형 AI 영상은 개인 창작자와 소셜 플랫폼의 콘텐츠·광고 제작 역량을 비약적으로 강화하는 동시에, 콘텐츠 진위와 사회적 신뢰를 훼손시키는 구조적 위험을 증폭시키고 있다. 소셜 미디어 플랫폼은 연령 인증 강화, 워터마킹·라벨링 의무화 등 강도 높은 규제와 플랫폼 책임 강화를 요구하는 압력에 직면해 있다. 향후 이들의 미래는 생성형 AI 콘텐츠에 대한 책임·통제·투명성을 얼마나 선제적으로 확보하느냐에 달려 있다.

핵심 내용 요약 (Executive Summary)

▷ 생성형 AI 영상의 부상과 콘텐츠 생산 구조 변화

- 생성형 AI 영상이 헐리우드급 품질에 근접하며 저비용·고효율 대량 제작이 가능
- 독립 창작자의 제작 역량이 확대되며, 소셜 플랫폼은 속풀 콘텐츠와 광고 시장의 영향력 강화

▷ 전통 미디어·광고 시장의 경쟁 구도 재편

- 전통 미디어는 생성형 AI의 도입에 신중하지만, 장기적으로 제작비 절감 가능성에 주목
- 소셜 플랫폼은 생성형 AI를 활용해 광고 테스트·성과 측면에서 경쟁 우위 확보
- 광고주는 제작보다 실험과 성과 최적화에 예산을 집중하는 방향으로 전환

▷ 사회적 신뢰 훼손과 플랫폼 책임 이슈 부상

- 합성 영상 확산으로 영상 증거의 신뢰도 하락
- 가짜 뉴스·조작 영상·정치적 허위 정보 등 사회적 혼란 리스크 확대
- 워터마킹, AI 라벨링, 출처 추적 등 플랫폼 책임 강화 요구 증가

▷ 규제 환경 변화와 산업 질서의 재조정

- 미국·유럽을 중심으로 연령 인증, AI 콘텐츠 라벨링, 플랫폼 책임 강화 등의 규제 본격화 전망
- CDA 230조 재검토 등 플랫폼 면책 구조에 대한 논의 확대
- 플랫폼의 선제 대응 여부가 향후 법적 리스크와 컴플라이언스 부담을 좌우할 핵심 변수

생성형 AI 영상이 헐리우드급 품질에 근접함으로써, 독립 영상 제작의 폭발적 성장과 함께 소셜 비디오 플랫폼에 대한 더 강력한 규제 대응이 촉발될 수 있다.

전설의 생물 사스콰치(sasquatch, 빅풋)의 셀프카메라부터 동화 속 공주들의 연애 조언, 진짜일 수도 가짜일 수도 있는 속보 뉴스까지, 창의적이면서 때로는 우려스러운 생성형 비디오가 소셜미디어 피드를 빠르게 채우며 이용자의 시선을 끌기 위해 경쟁하고 있다. 현실이 그 어느 때보다 강력한 가상 경쟁 상대를 만난 셈이다.

생성형 비디오를 활용하면 독립 창작자들은 더 적은 비용으로 더 많은 콘텐츠를 제작할 수 있고, 소셜미디어 플랫폼은 강력한 솔루션 엔터테인먼트를 제공해 디지털 광고 시장에서 더 많은 점유율을 확보할 수 있다. 하지만 이와 동시에 이용자들을 압도하고 진정성을 훼손하는 부작용을 낳고, 이를 억제하려는 규제 움직임을 촉발할 수 있다.

누구나 사실처럼 보이는 영상을 만들어 수백만 명에게 뉴스나 브랜디드 콘텐츠(branded contents)*, 팬픽션(fan fiction)** 등 다양한 형태로 퍼뜨릴 수 있고, 이를 이용한 사기와 갈취, 허위 정보의 의도적 유포가 가능해졌다. 이처럼 생성형 비디오의 오남용 위험이 증대함에 따라 규제의 칼날이 더욱 날카로워질 수 있다.

* 브랜디드 콘텐츠(branded contents)는 광고주가 자신의 브랜드·제품·가치를 직접적으로 노출하기보다, 이야기·정보·엔터테인먼트 요소에 자연스럽게 녹여 전달하는 콘텐츠 형식을 의미한다. 전통적 광고처럼 구매를 촉구하기보다, 브랜드와 소비자 간의 정서적 연결과 신뢰 형성을 목적으로 한다.

** 팬픽션(fan fiction)은 기존 작품의 세계관·캐릭터·설정 등을 차용해 팬이 자발적으로 창작한 2차 창작 서사 콘텐츠를 의미한다. 원작의 빈틈을 확장하거나, 대안적 전개·관계·결말을 상상함으로써 소비자에서 창작자로의 전환을 보여주는 대표적 문화 현상이다.

딜로이트는 2026년에 미국에서 생성형 비디오에 대한 규제 대응이 촉발될 수 있다고 전망한다. 그 과정에서 연령 인증 요구를 강화하는 주(州) 정부가 늘어나고, 플랫폼이 사용자 게시물의 법적 책임에서 상당 부분 면책되는 근거가 되는 통신품위법(CDA, Communications Decency Act)의 섹션 230 조항에 대한 연방 차원의 재검토가 이뤄질 수 있다.¹ 또한 소셜 플랫폼에 게시되는 AI 생성 콘텐츠에 대한 라벨링(labeling)*이 의무화될 수 있다.² 이러한 규제 움직임은 이미 뉴욕, 테네시, 유타 등 일부 주에서 시작됐으며, 미국 연방대법원은 미시시피 주의 소셜미디어 연령 인증 법안에 대한 이의 제기를 심리하지 않기로 결정했다.³ 유럽연합(EU)의 디지털서비스법(DSA, Digital Services Act)도 ‘효과적인 연령 검증 방법’을 포함하고 있다.⁴ 2026년에는 미국 중간선거가 예정된 만큼, 소셜 플랫폼들은 보유한 AI 및 데이터 역량을 적극 활용해 생성형 콘텐츠 관리를 강화하라는 압력을 받을 수 있다. 이미 일부 플랫폼은 이를 위한 기술적 조치를 강화하고 있다.⁵

* 라벨링(labeling)은 데이터·콘텐츠·제품·시스템 등에 성격·출처·속성·의미를 식별할 수 있는 정보를 명시적으로 부여하는 행위 또는 체계를 의미한다. 디지털·AI 환경에서는 콘텐츠의 생성 주체, 생성 방식(AI 여부), 용도, 위험 수준을 구분하기 위한 신뢰 장치로 활용된다.

한편 생성형 AI가 방대한 양의 비디오를 쏟아냄에 따라, 이러한 비디오를 대규모로 관리·감독하는 콘텐츠 조정(moderation) 기술의 발전이 가속화될 가능성이 크다. 규제기관은 이러한 기술적 노력이 온라인상의 위해 요소를 얼마나 효과적으로 완화하는지 면밀히 들여다볼 것이다. 플랫폼 또한 참여도 감소, 수익화 축소, 컴플라이언스 부담 증가 등 잠재적 리스크를 주시하면서 대응 전략을 조정해야 한다.

개인 창작자의 성장에 도움 되는 생성형 툴

생성형 비디오 모델은 고품질 영상과 오디오를 단 몇 초 만에 실제 콘텐츠와 거의 구분되지 않을 정도로 자연스럽게 만들어낼 수 있다.⁶ 이러한 사용 편의성과 비용 효율성 덕분에 창작자들은 실패 위험을 크게 낮춘 채 더 과감한 실험을 시도하고, 소셜 비디오라는 극도로 경쟁적인 시장에서 창의적 아이디어를 빠르게 테스트할 수 있다.

생성형 AI 기술만으로 30분짜리 TV 프로그램이나 2시간 분량의 영화를 만드는 것은 아직 어렵지만, 인터넷 밍(meme)*을 그대로 구현한 고품질 광고 영상과 같은 소셜용 콘텐츠를 만들기에는 충분하다.⁷ 오히려

이러한 제약이야말로 생성형 비디오가 할리우드보다 개인 창작자와 소셜 플랫폼에서 먼저 확산되는 배경이 되고 있다. 짧은 길이, 빠른 편집, 셀프카메라 기반 연출 등 소셜 콘텐츠와 맞아 떨어지는 특징 덕분에, 플랫폼 이용자들도 오픈 소스 기반 무료 엔터테인먼트의 품질에 대해 상대적으로 관대하다.

* 인터넷 밍(meme)은 인터넷에서 급속도로 확산되어 사회적 문화의 형태로 자리 잡은 소셜 아이디어, 활동, 트렌드 등을 뜻한다. 이미지·영상·문구·행동 양식 등 문화적 요소가 이용자간 모방·변형·재전파를 통해 빠르게 확산되는 디지털 문화 단위로, 단일 콘텐츠가 아니라 반복과 변주를 통해 진화하는 집단적 표현 형식이라는 점이 핵심이다.

합성 콘텐츠 제작에 적용되는 생성형 툴은 전체 프로덕션 스택을 대체하는 기술이라기보다, 수많은 미시 작업(micro-task)을 제거하고 제작 시간을 압축해 소규모 팀이 더 많은 결과물을 만들어낼 수 있게 하는 도구로 자리 잡고 있다. 생성형 AI와 비디오 툴을 활용하면 콘텐츠 제작·배포·측정 과정 전반에서 미시 작업을 줄이고, 제작 시간과 비용을 줄일 수 있다.⁸ 이를 통해 창작자들은 소셜 플랫폼 알고리즘에 맞춰 빠른 게시 주기를 유지하고 일정한 콘텐츠 생산량을 확보해, 팔로워와의 관계를 끊기지 않게 하는 데 큰 도움을 받을 수 있다.

대다수 생성형 AI 툴은 스크립트 기반 자동 영상 생성이나 원클릭(one-click) 클립 생성*처럼 시간과 비용을 절약하는 기능에 초점을 맞춘다.⁹ 이러한 기능을 활용하면 여러 버전을 빠르게 테스트해 특정 시청자층이

나 알고리즘에 더 잘 맞는 방식을 찾아낼 수 있다. 또한 창작자 본인의 AI 아바타를 만들어 시청자의 피로도를 줄이면서도 참여를 유지하고, 나아가 대규모 개인화도 할 수 있다.¹⁰ 이러한 기능은 생성형 광고에 활발히 적용되고 있다.¹¹

* 원클릭(one-click) 클립 생성은 장시간 영상(full-length video)에서 하이라이트·요약·숏폼용 클립을 사용자의 추가 편집 없이 한 번의 클릭만으로 자동 추출·생성하는 기술을 의미한다. AI가 영상의 맥락, 화자, 감정 반응, 시청 패턴 등을 분석해, 플랫폼에 최적화된 짧은 클립을 자동으로 제작한다.

생성형 AI는 비(非)생성형 콘텐츠의 편집 속도도 높여준다. 화자의 발언 사이 ‘음...’ 같은 군더더기 표현이나 정적을 제거하고, 불량 촬영본을 수정하고, 흔들린 카메라를 안정화하고, 불필요한 공백을 자동 제거하기도 한다.¹² 다국어 더빙 툴은 시청 가능 지역을 넓혀 광고 수익과 참여도 향상을 이끌 수 있다.¹³

이처럼 창작자 스튜디오는 아이디어 발굴, 콘텐츠 생산, 타깃 시청자층 공략, 성과 측정, 반복 개선을 훨씬 빠르게 수행할 수 있다. 이는 콘텐츠 제작을 둘러싼 경제적 구조를 재편함으로써 콘텐츠의 폭발적 증가를 촉발할 수 있다. 공급이 늘어나면서 경쟁이 더 치열해지고, 이는 다시 더 창의적인 콘텐츠를 자극하는 선순환이 형성되는 것이다.

할리우드와 소셜 플랫폼 모두를 위협하는 생성형 비디오

딜로이트의 ‘2025 미디어·엔터테인먼트 전망’(2025 Media and Entertainment Outlook)에 따르면, 대형 제작사들은 생성형 비디오를 실험적으로 도입하고 있지만 실제 제작에 본격 활용하는 것은 여전히 주저하고 있다. 프리미엄 콘텐츠 브랜드가 합성 미디어로 인해 손상될 위험이 있을 뿐 아니라, 제작 스텝진과의 갈등도 무시할 수 없기 때문이다. 2023년 11개월간 지속된 할리우드 배우 및 방송인 노동조합(SAG-AFTRA) 파업 당시 노조 측이 생성형 AI의 제작 사용을 제한할 것을 요구한 바 있다.¹⁴ 하지만 과도한 제작비 부담에 시달리는 할리우드 제작사들은 장기적으로는 생성형 AI가 이러한 부담을 덜어주기를 바랄 가능성이 크다.¹⁵

뿐만 아니라 전통적 제작사와 스트리밍 서비스는 광고비 경쟁에서 더 큰 압박을 받고 있다.¹⁶ 일부 할리우드 제작사가 하락세에 접어든 선형 TV(linear TV, 정해진 편성표에 따라 콘텐츠를 실시간으로 일방 송출) 광고 손실을 막고 광고 사업을 커넥티드 TV와 스트리밍 플랫폼으로 이전하려는 동안, 소셜 플랫폼이 디지털 광고 수익을 빠르게 빼앗아가고 있다. 실제로 광고주 지출은 스트리밍 서비스보다 소셜 플랫폼에서 훨씬 더 빠른 속도로 증가하고 있다.¹⁷

생성형 AI가 콘텐츠를 빠르게 제작할 뿐더러 특정 소비자 및 이용자 세그먼트가 반응하는 콘텐츠를 빠르게 예측할 수 있기 때문에, 디지털 광고의 지형이 근본적으로 재편되고 있다.¹⁸ 소셜 플랫폼은 간단한 프롬프트만

으로 수천 개의 광고를 자동 생성해 각각의 버전에 미세한 차이를 준 후, 어떤 버전이 가장 큰 성과를 거두는지 즉시 테스트할 수 있다.¹⁹ 이에 따라 광고주는 제작비보다 테스트에 더 많은 비용을 투자하게 됐다. 성과가 가장 높은 버전을 선택하기만 하면 효율을 극대화할 수 있기 때문이다. 그 결과, 소셜 플랫폼은 광고 시장에서 경쟁 우위를 더욱 강화하고 있다.²⁰

그러나 생성형 비디오 활용이 확산되면 소셜 플랫폼 역시 더 큰 위험에 직면할 수 있다. 콘텐츠 양이 기하급수적으로 늘어나면서 이를 제공 및 관리해야 하는 부담이 커질 것이며, 일부 콘텐츠는 저작권 침해나 더 심각한 문제 영역에 놓일 가능성이 있다. 또한 AI 슬롭(AI slop)*에 대한 피로감과 함께, 소셜 미디어 성장의 핵심 자산이었던 ‘진정성’이 빠르게 훼손될 위험도 있다.

*AI 슬롭(AI slop)은 생성형 AI를 통해 대량으로 자동 생산되었으나 품질·의미·신뢰도가 낮고 반복적이며 피로감을 유발하는 저품질 콘텐츠를 의미한다.겉보기에는 그럴듯하지만 정보 가치나 서사적 밀도가 부족해, 이용자의 주의력과 신뢰를 소모시키는 문제가 있다.

2025년 초에 실시된 미국 소비자의 생성형 미디어에 대한 인식 조사²¹에서 응답자 64%는 소셜 미디어에서 확산되는 생성형 비디오가 위험하다고 답했다. 76%는 온라인 콘텐츠 제작자는 생성형 AI 사용 여부를 투명하게 공개해야 한다고 답했으며, 53%는 생성형 AI를 사용하는 제작자는 ‘진정성이 없다’고 느낀다고 답했다.

하지만 1년이 지난 지금, 생성형 비디오 기술이 한층 발전했고 소셜 플랫폼 내 생성형 콘텐츠의 양 또한 급증했다. 생성형 비디오는 현실과 거의 구분되지 않을 정도로 빠르게 발전하고 있으며, 머지않아 위험성을 넘어선 영역까지 진입할 수 있다. 이는 규제 강화를 촉발함과 동시에, 악의적 행위자에게도 강력한 도구를 안겨주는 결과가 될 수도 있다. 생성형 모델은 범죄자가 AI로 사람을 정교하게 사칭하는 데 활용할 수 있어 사기 우려가 커지고 있다.²² 사스콰치의 셀프카메라 같은 재미를 유발하는 수준에 그치지 않고, 불특정 다수에게 영향력을 행사하는 선전 행위, 사기 및 탈취, 정치적 허위정보, 음모론적 콘텐츠 양산에 악용될 위험이 크다. 또한 영상 증거를 신뢰할 수 없게 되면 법적 절차도 흔들릴 수 있다. 하지만 엄격한 규제가 마련되거나 시청자가 대거 이탈하거나 징벌적 조치가 가해지지 않는 한 소셜 플랫폼이 이러한 악의적 행위를 스스로 통제할 유인은 여전히 부족한 실정이다.

생성형 미디어 확산이 초래할 사회적 혼란과 소셜 플랫폼의 책임

합성 미디어, AI 슬롭, 비즈니스 모델 붕괴보다 더 큰 문제는 사회적 혼란이다. 누구나 현실과 구분하기 어려운 영상을 만들고 유통할 수 있게 되면서 영상 증거 자체가 더 이상 신뢰할 만한 진실의 역할을 하지 못하게 된다. 가짜 뉴스 피드, 유명인 목격 영상, 조작된 정치적 실수, 거짓 경고 등 첨단 생성형 비디오들을 살펴보면, 머지않아 대규모 사회적 혼란이 발생할 위험이 매우 크다.

이러한 위험에 선제적으로 대응하려면, 소셜 플랫폼이 워터마킹(watermarking)*, AI 라벨링, 콘텐츠 출처(provenance) 추적 기능을 개발 및 통합해야 한다. 이는 플랫폼에 업로드 되거나 플랫폼에서 생성되는 모든 콘텐츠(광고 포함)의 출처 및 진위를 확인할 수 있는 핵심 기술이 될 것이다. 필연적으로 반복될 정치적 조작과 소비자 기만 행위에 따른 피해를 막기 위해 규제 당국이 플랫폼과 협력해 생성형 콘텐츠에 더 강력한 안전장치를 마련할 것이며, 이 과정에서 라벨링과 워터마킹 의무화 조치가 필요해질 수 있다.

* 워터마킹(watermarking)은 이미지·영상·음성·문서 등 디지털 콘텐츠에 육안으로 보이거나 보이지 않는 형태의 식별 정보를 삽입해 출처·진위·권한을 확인할 수 있도록 하는 기술을 의미한다. 생성형 AI 환경에서는 특히 AI 생성 여부, 제작 주체, 위·변조 여부를 추적하기 위한 신뢰 기술로 주목받고 있다.

소셜 플랫폼이 선제적으로 대응하지 못한다면 미국에서는 플랫폼이 사용자 게시물의 법적 책임에서 상당 부분 면책되는 근거가 되는 통신품 위법의 섹션 230을 수정해야 한다는 압박이 증대할 수 있다.²³ 유럽 규제 당국은 이미 미국 기반의 소셜 플랫폼과 데이터 수집 기업 규제를 강화하는 강한 의지를 보여 왔다. 생성형 틀이 만든 콘텐츠를 감시하는 컴플라이언스 에이전트(compliance agent) 같은 자동화된 규정 준수 체계를 개발한다면, 플랫폼은 대규모 위반 행위를 보다 신속하게 탐지하고 대응할 수 있다.

소셜 미디어는 연결성과 투명성, 인간성의 확장을 가능하게 했지만, 동시에 정보의 파편화, 미디어의 사실상 탈규제, 악의적 행위자들의 수법 고도화 또한 초래했다. 따라서 소셜 플랫폼이 선제적으로 강력한 대응에 나서지 않는다면, 생성형 비디오는 이러한 문제를 증폭시켜 결국 사회의 닷 역할을 했던 ‘진실의 공유’라는 소셜 플랫폼의 긍정적 기능이 훼손될 위험이 있다.

Korean Perspectives

생성형 AI 영상 혁신, 국내 콘텐츠 산업의 기회와 과제

2026년을 전후로 생성형 AI 영상 기술은 콘텐츠 산업의 제작 방식과 경쟁 구도를 빠르게 바꾸고 있다. 이제 영상 제작은 더 이상 대규모 자본과 인력의 영역이 아니다. AI를 활용하면 개인이나 소규모 팀도 저비용·고효율로 고품질 콘텐츠를 만들 수 있는 환경이 형성되고 있다. 이 변화와 함께 개인 창작자와 소셜 플랫폼의 영향력은 크게 확대되고 있으며, 속풀 콘텐츠와 디지털 광고 시장 역시 더욱 빠른 성장 국면에 접어들고 있다. 반면, 기존 미디어와 스트리밍 서비스는 제작비 부담, 제작 속도, 실험 역량 측면에서 새로운 경쟁 압력을 받고 있다.

이러한 변화의 핵심에는 AI 기반 제작 파이프라인의 자동화가 있다. 기획부터 편집, 다국어 더빙, 배포, 성과 분석까지 전 과정이 자동화되면서 콘텐츠는 ‘완성도 높은 한 편’을 만드는 방식에서 벗어나, ‘여러 버전을 빠르게 실험하고 검증하는 방식’으로 생산·유통되고 있다. 기업의 경쟁력 역시 개별 작품의 완성도보다는, 얼마나 많은 시도를 빠르게 실행하고 데이터를 통해 최적화할 수 있는지에 의해 결정되는 구조로 이동하고 있다.

창작자 생태계도 빠르게 재편되고 있다. 독립 창작자와 소규모 스튜디오는 AI와 데이터 기반 제작 역량을 무기로 플랫폼과 보다 대등한 위치에 서기 시작했다. 플랫폼과 브랜드가 이들을 여전히 단순한 콘텐츠 공급자로만 인식할 경우, 우수 인재와 IP는 다른 플랫폼으로 이동할 가능성이 높다. 수익 배분 구조, 데이터 접근권, 멀티 플랫폼 배급 전략을 포함한 협력 방식의 재설계가 불가피해진 이유다. 한편 콘텐츠 과잉과 신뢰 약화라는 부작용도 함께 커지고 있다. 반복적이고 저품질의 AI 콘텐츠, 저작권 침해, 조작 영상과 허위 정보의 확산은 플랫폼 신뢰도를 직접적으로 위협한다. 영상이 더 이상 사실을 자동적으로 증명하는 수단으로 받아들여지지 않는 환경에서, 신뢰를 관리하는 역량은 기술 경쟁력 못지않은 핵심 요소가 되고 있다.

규제 환경 역시 빠르게 움직이고 있다. AI 콘텐츠 라벨링, 워터마킹, 출처 추적, 연령 인증 등 플랫폼 책임을 강화하는 제도들이 미국과 유럽을 중심으로 본격화되고 있으며, 국내 기업 역시 이러한 흐름을 전제로 사업과 기술 전략을 설계해야 하는 단계에 접어들었다. 이러한 변화에 대응하기 위해 기업은 제작 자동화와 반복 실험 문화를 조직

전반에 정착시키는 한편, AI 사용 여부 표기와 원본 추적 등 신뢰 기술을 플랫폼 구조에 내재화해야 한다. 광고와 콘텐츠 제작 예산도 단일 제작비 중심에서 테스트와 성과 최적화 중심으로 전환할 필요가 있으며, 창작자와는 단기 거래가 아닌 장기적 파트너십을 기반으로 한 생태계를 구축해야 한다.

결국 생성형 AI 영상의 확산은 콘텐츠 산업에 새로운 성장 가능성을 열어주는 동시에, 신뢰와 책임의 기준을 근본적으로 다시 설정하고 있다. 이 변화에 구조적으로 대응하는 기업만이 단기 효율을 넘어, 장기적인 지속 가능성과 경쟁력을 확보할 수 있을 것이다.



유선희 파트너

한국 딜로이트 그룹
One Cyber & Resilience 파트너

11장

영상 팟캐스트, 위협적인 글로벌 미디어·광고 플랫폼으로 진화

영상 팟캐스트(Vodcast)는 오디오와 영상, 다국어 콘텐츠의 결합을 통해 글로벌 미디어이자 광고 플랫폼으로 빠르게 진화하며, 시청자 참여도·스크린 타임·광고 수익을 동시에 확대하고 있다. 이는 광고·브랜드 사업자에게는 새로운 성장 기회를 제공하는 한편, 전통 미디어에는 스크린 타임 경쟁이라는 도전 과제를 안겨주고 있다. 다만 향후 성장의 관건은 추천 영상 등 노출 콘텐츠의 강화, 수익화 구조의 고도화, 그리고 신흥 시장을 포함한 글로벌 확장 과제를 얼마나 효과적으로 해결하느냐에 달려 있다.

핵심 내용 요약 (Executive Summary)

▷ (팟캐스트의 진화) 오디오 → 비디오 → 글로벌 플랫폼(팟캐스트·소셜·스트리밍 비디오 간 경계 붕괴)

- 팟캐스트는 영상 중심(vodcast)·다국어 기반 롱폼 미디어로 진화
- 오디오 스토리텔링에 시각 요소가 결합되며 몰입도·공유성·스크린 타임 급증

▷ (영상 팟캐스트 시장 성장 및 글로벌 확산) 영상 기반 광고·스폰서십 확대로 2026년 광고 매출 50억 달러, 연 20% 성장 전망

- 영상 팟캐스트는 오디오 대비 소비량 약 1.5배로 체류 시간과 몰입도 증가
- 다국어 콘텐츠 확대로 비영어권 신흥 시장으로의 시청자 규모·도달 범위·수익이 동시 확대

▷ (플랫폼 사업자들의 전략 변화) 유튜브·스포티파이 등 주요 플랫폼 사업자들은 기본 UI에 비디오 팟캐스트를 핵심 콘텐츠로 통합

- 알고리즘의 영향으로 속품 형태의 추천 영상 노출 빈도 증가
- 영상 팟캐스트로 전통 TV·스트리밍이 독점하던 스크린 타임의 일부 흡수

▷ (미디어·플랫폼 사업자들의 대응) 추천 영상 등 다채로운 콘텐츠의 확보와 효과적인 노출, 수익화 구조 개선 및 시장 확장성이 도전 과제로 부상

- 추천·노출 구조의 한계와 수익 모델 미성숙 및 신흥 시장의 낮은 CPM이 수익성 저해 → 커머스 연계 광고 등 고부가 수익 모델 확보
- 현지 콘텐츠 투자와 다국어 역량이 필수 → 생성형 AI 기반 자동 번역·립싱크 기술이 핵심 경쟁력
- 영상 팟캐스트를 저비용 콘텐츠 인큐베이터 및 IP 발굴 수단으로 활용
- 팬덤 기반 콘텐츠의 TV·영화 확장을 통해 제작 리스크 완화

팟캐스트는 이제 영상 중심의 다국어 기반 매체로 빠르게 진화하며 시청 자층을 빠르게 확대하고 있다. 팟캐스트가 시청자의 스크린 타임에서 차지하는 비중이 한층 확대됨에 따라, 브랜드는 전 세계 시청자에게 다가갈 수 있는 새로운 광고 기회를 얻을 수 있다.

이제 팟캐스트는 듣기만 하는 콘텐츠가 아니다. 영상 팟캐스트(vodcast)가 부상하면서, 오디오 스토리텔링과 시각적 매력을 결합한 새로운 형태의 롱폼(long-form)* 미디어가 등장하고 있다. 이러한 결합은 콘텐츠를 더욱 몰입감 있게 만들고 더 많이 공유되게끔 함으로써, 시청자가 팟캐스트를 소비하는 방식을 재정의하고 있다. 팟캐스트와 소셜 미디어, 스트리밍 비디오 간 경계가 흐려지는 가운데, 창작자들은 크로스 플랫폼 방식으로 콘텐츠를 적극 배포해 이용자 참여도를 높이고, 커뮤니티를 구축하며, 광고 수익과 신규 스폰서십 기회를 확대하고 있다. 이제 영상 팟캐스트는 전통적 TV와 스트리밍 플랫폼이 오랫동안 독점 하던 스크린 타임의 일부를 빼앗아오고 있다.

***롱폼(long-form)**은 짧은 클립이나 요약형 콘텐츠와 대비되는 개념으로, 비교적 긴 분량을 통해 서사·정보·맥락을 충분히 전달하는 콘텐츠 형식을 의미한다.

특히 인도, 나이지리아, 브라질과 같은 신흥 시장에서는 모바일 중심, 저 대역폭 기반의 접근성이 강점으로 작용하며 팟캐스트 이용자가 빠르게

증가하고 있다. 현지화된 다국어 콘텐츠도 증가하면서, 팟캐스트는 점점 더 문화적으로 다양한 글로벌 매체로 자리 잡고 있다. 물론, 수익화, 언어 접근성, 인프라 격차 등 해결해야 할 난제들이 여전히 있다. 하지만 딜로이트는 2026년 전 세계 팟캐스트 및 영상 팟캐스트 광고 수익이 약 50억 달러에 이를 것으로 전망한다. 이는 전년 대비 약 20% 성장하는 수준이다.¹

높아지는 영상 팟캐스트의 인기와 팟캐스트의 글로벌 확장이 맞물리면서, 관련 시장은 시청자 규모, 도달 범위, 광고 수익 측면에서 본격적인 성장 국면에 접어들었다. 다만 업계의 향후 성장 경로는 창작자와 플랫폼이 콘텐츠 발견성(discoverability)*, 수익화, 확장성이라는 난제를 얼마나 효과적으로 해결하느냐에 달려 있다.

* **콘텐츠 발견성(discoverability)**은 이용자가 특정 콘텐츠를 의도적으로 검색하지 않더라도 플랫폼의 구조·알고리즘·추천·유통 경로를 통해 해당 콘텐츠가 얼마나 쉽게 노출되고 인지되는지를 나타내는 정도를 의미한다.

새로운 방식으로 시청자 참여 혁신하는 영상 팟캐스트

시청자와 광고주들 사이 영상 팟캐스트의 인기는 2026년 이후에도 계속 올라갈 전망이다. 이러한 성장은 크게 세 가지 요인이 뒷받침할 것으로 보인다. 첫째는 기존의 인기 미디어 플랫폼과 자연스럽게 통합되는 특성이고, 둘째는 소셜 플랫폼용 클립을 활용해 화제성과 바이럴을 유

도하는 특성이며,셋째는 창작자가 충성도 높은 시청자와 더 깊게 연결될 수 있는 특성이다.

최근 몇 년 사이 스포티파이(Spotify), 원더리(Wondery), 팟빈(Podbean), 유튜브(YouTube) 등 일부 스트리밍 음악·오디오 서비스는 팟캐스트 비디오 피드를 사용자 인터페이스에 직접 통합해 소비자가 쉽게 접근할 수 있도록 했다.² 동시에 주요 플랫폼은 창작자가 영상 자산을 제작하고 수익화할 수 있도록 툴과 노하우를 제공하며, 영상 팟캐스트 콘텐츠의 공급을 크게 늘려 왔다.³ 특히 2022년에야 본격적으로 팟캐스트 영역에 진입한 유튜브는 이미 시장을 뒤흔들고 있다.⁴ 유튜브는 2025년 초 기준 월간 영상 팟캐스트 시청자 수 10억 명을 달성했으며, 미국 시장을 대상으로 한 팟캐스트 순위 차트를 신설했다.⁵ 또한 2025년 8월에는 미국프로풋볼(NFL) 스타 제이슨과 트래비스 켈시(Jason and Travis Kelce) 형제가 진행하는 ‘뉴 하이츠’(New Heights) 영상 팟캐스트 생중계가 동시 시청자 130만 명을 기록해 기네스 신기록을 세우기도 했다.⁶

이처럼 대중이 이미 시간을 쓰고 구독료를 지불하는 플랫폼에서 영상 팟캐스트가 제공되면서 소비자 진입 장벽이 낮아지고, 콘텐츠와 광고 모두 시청자 참여도가 상승하고 있다. 각 플랫폼의 접근 방식은 다르다. 어떤 곳은 여전히 오디오만 제공하기도 하지만, 어떤 곳은 모든 팟캐스트의 영상 버전을 제공한다. 2022년에야 영상 기능을 도입한 스포티파이는 2025년 9월 기준 인기 쇼의 60% 이상이 영상 버전을 제공하고 있다.⁷ 딜로이트는 향후 영상 버전이 있는 인기 팟캐스트 비중

이 더 늘어나고 소비자는 영상 중심 플랫폼으로 이동할 것이라고 전망 한다. 2025년 가을 발표된 딜로이트의 ‘디지털 미디어 트렌드’(Digital Media Trends) 보고서에 따르면, 미국 소비자의 27%가 매주 영상 팟캐스트를 시청하며 이 중 Z세대와 밀레니얼이 가장 큰 비중을 차지하는 것으로 나타났다.⁸

영상 팟캐스트의 영상 요소는 오디오만으로는 구현할 수 없는 방식으로 시청자를 대화 속으로 끌어들인다. 호스트의 표정, 몸짓, 스튜디오 분위기 등 시각적 맥락은 친밀감을 강화하고, 시청자가 좋아하는 호스트를 직접 보는 경험은 파라소셜(parasocial)* 관계를 강화해 신뢰와 진정성을 높인다.⁹ 그 결과, 영상 팟캐스트 시청자는 오디오만 듣는 청취자보다 콘텐츠 소비량이 1.5배 많은 것으로 나타났다.¹⁰ 영상은 디지털 네이티브인 젊은층 시청자에게 매력적인 스토리텔링 계층을 추가로 제공하는 효과가 있으며, 플랫폼 전반에서 새로운 시청자를 유입시키는 데도 효과적이다. 광고와 스폰서십 측면에서도 로고 노출, 제품 배치, 소셜 공유용 속클립 제작 등 영상만이 창출할 수 있는 기회가 증대된다.¹¹

* **파라소셜(parasocial)** 관계는 이용자가 미디어 속 인물(연예인, 크리에이터, 진행자 등)과 실제 상호작용이 없음에도 불구하고, 마치 개인적 관계가 형성된 것처럼 느끼는 일방적 심리적 유대 관계를 의미한다. 이는 상호 소통이 아닌 인지적, 정서적 친밀감의 차각에 기반한 관계 형태다.

영상 팟캐스트의 성공에는 소셜 미디어 플랫폼이 결정적인 역할을 한다.¹² 짧게 편집된 영상 팟캐스트 클립은 다양한 플랫폼에서 재사용되며, 서로 다른 시청층을 공략하고, 화제성을 높이며, 콘텐츠 확산을 가속화한다. 소셜 플랫폼은 콘텐츠 발견성을 높이고 광고 범위를 확장하며, 창작자가 시청자와 직접 소통할 수 있는 공간을 제공한다.

2024년 기준 영상 팟캐스트 시청자의 약 절반이 커넥티드 TV를 통해 콘텐츠를 시청하는 것으로 나타났다.¹³ 이는 영상 팟캐스트가 이미 거실 스크린을 차지하기 시작했음을 보여준다. 이러한 트렌드는 전통 스트리밍 서비스도 팟캐스트 시장 진입을 해야 하는 것 아닌가라는 고민을 하게 만드는 압력으로 작용할 가능성이 크다.¹⁴ 영상 팟캐스트는 멀티태스킹이 가능한 오디오 팟캐스트에 비해 시청자가 온전히 집중해서 소비해야 하는 콘텐츠다. 영상 팟캐스트 시청자의 44%는 시청 중 절대 멀티태스킹을 하지 않는다고 답했는데, 이는 오디오 팟캐스트 청취자의 29%보다 훨씬 높은 수치다.¹⁵

이처럼 더 높은 집중도와 몰입도는 더 큰 참여도와 구독자 증가로 이어질 가능성이 크며, 광고주와 스폰서들의 관심과 투자를 끌어들이는 요인이 된다. 딜로이트 조사에 따르면 미국 팟캐스트 시청자 및 청취자의 약 25%(Z 세대 및 밀레니얼 세대는 3분의 1 이상)가 팟캐스트에서 광고를 접한 제품 및 서비스를 실제 구매한 경험이 있다고 답했다.¹⁶ 결국 광고와 스폰서십 유입은 더 많은 수익으로 이어지고, 이는 다시 더 많은 콘텐츠 제작, 더 큰 도달 범위, 더 빠른 혁신을 촉진한다. 결과적으로 영상

팟캐스트의 부상은 팟캐스트 산업의 다음 성장 경로를 결정짓는 핵심 동력이 될 전망이다.

현지화·다국어화 기반 팟캐스트 콘텐츠, 전 세계로 확산

미국에서 시작된 오디오 팟캐스트가 역동적 글로벌 미디어로 빠르게 진화하고 있다.¹⁷ 지역별로 주당 팟캐스트 시청·청취 비중은 크게 차이가 나지만, 전 세계 평균치는 약 22%이며, 인도네시아(42.6%)와 멕시코(41.8%)와 같은 시장이 청취율을 견인하고 있다.¹⁸ 신흥 시장에서는 모바일 연결성 확대, 스트리밍 오디오 플랫폼의 글로벌 투자 증가, 현지 언어와 다국어 기반 팟캐스트 및 영상 팟캐스트 콘텐츠의 확산 등이 팟캐스트 소비를 폭발적으로 끌어올리는 요인으로 작용한다.

전 세계적으로 모바일 인터넷 접근성이 확대되면서 네트워크 커넥티비티(connectivity, 연결성)와 콘텐츠 소비의 민주화가 이뤄지고 있다. 인도, 나이지리아, 브라질 등 국가에서는 저렴한 스마트폰과 데이터 요금 제 덕분에 수백만 명이 온라인 콘텐츠에 접속할 수 있게 됐다.¹⁹ 예를 들어, 나이지리아의 모바일 데이터 요금은 지난 10년 동안 약 97% 하락했다.²⁰ 2014년 11.15달러였던 1기가바이트(GB) 데이터 가격이 2023년에는 0.39달러로 하락했다. 더 저렴한 기기와 요금제 접근성 덕분에 팟캐스트와 영상 팟캐스트 같은 온디맨드 오디오·비디오 콘텐츠가 더 많은 이용자와 더 많은 지역으로 확산되고 있다.

이 가운데 스포티파이와 같은 기존 오디오 스트리밍 기업들은 글로벌 확장과 함께 현지 언어 콘텐츠에 대한 투자를 늘리며 팟캐스트 성장에 불을 붙이고 있다. 예컨대 스포티파이는 라틴아메리카, 아프리카, 아시아 전역에서 창작자를 지원하고 독점 파트너십을 체결해 지역 특화 쇼 개발에 적극 나서고 있다.²¹ 다른 플랫폼도 인기 있는 현지 팟캐스트를 라이선싱하거나 오리지널 콘텐츠를 제작하고, 지역 창작자 생태계를 지원하는 툴을 구축하고 있다.²² 팟캐스트 업계가 여전히 영어 중심 패러다임에 머물러 있음에도, 다국어·다문화 특화 프로그램이 글로벌 팟캐스트의 성장을 견인하는 핵심 동력이 될 수 있다는 인식이 커지고 있다. 한편 레바논, 인도, 나이지리아 등지에서도 새로운 플랫폼이 다수 등장해 현지 콘텐츠 허브로 자리 잡고 있으며, 글로벌 기업과의 파트너십을 확대하며 영향력을 키우고 있다.²³

영상 팟캐스트라는 새로운 포맷 역시 시청자 참여를 높이고 있다. 신흥 국은 선진국보다 상대적으로 젊은층 디지털 네이티브 인구 비중이 높기 때문에 이러한 포맷의 콘텐츠가 큰 인기를 끌기 쉽다.²⁴

팟캐스트의 글로벌 확장은 콘텐츠 발견과 소비를 촉진하고 특히 비(非)영어 콘텐츠의 확산을 가속하면서, 플랫폼의 수익화 전략에도 중대한 시사점을 제공한다.²⁵ 현지 쇼와 문화적 맥락이 담긴 스토리텔링 속 광고 배치를 통해 다양한 청중에게 도달하려는 다국적 브랜드가 증가할 것이다. 다만 신흥 시장의 낮은 CPM(cost per mille, 1,000회 노출당 광고주가 지불하는 비용)은 여전히 창작자에게 수익화의 장애물로 작용한다.

그럼에도 팟캐스트 시장의 글로벌화가 심화되면서 현지 광고, 브랜드 콘텐츠, 국경을 넘는 창작자 협업이 폭발적으로 증가하고 있으며, 결과적으로 팟캐스트 생태계가 더욱 풍요롭고 역동적으로 진화하고 있다.²⁶

미개척 글로벌 시청자를 확보하고 새로운 성장을 도모하는 기회

영상 및 오디오 팟캐스트는 이제 거실을 장악함과 동시에 전 세계 신흥 시장으로 세력을 확장하고 있다. 이는 미디어·엔터테인먼트 업계 전반에 새로운 기회와 도전을 동시에 제시한다.

스트리밍 오디오·음악 플랫폼은 자사 앱 안에서 영상 팟캐스트를 끊김 없이 스트리밍할 수 있는 기술 역량을 구축하거나 고도화하는 데 주력 할 필요가 있다. 이를 위해서는 인프라, 기술 시스템, 전문 인력에 대한 투자가 필요하다. 이미 해당 역량을 갖춘 기업이라면, 영상 광고 화면을 바로 클릭해 구매로 이어지는 기능 등 영상 팟캐스트 광고에서 직접 구매가 가능한 다이내믹한 요소를 실험하는 것이 다음 단계가 될 수 있다. 이는 광고주와의 고부가 파트너십·스폰서십을 확보하고, 업계 전반의 수익화를 가속하는 중요한 기회가 될 것이다.

이들 플랫폼은 동시에 신흥 시장으로의 확장도 모색해야 한다. 이를 위해서는 현지 콘텐츠와 창작자에 대한 투자를 늘리고, 오디오·비디오 콘텐츠를 자동 번역 및 립싱크하는 생성형 AI 역량을 구축해야 한다.²⁷ 이러한 투자는 신규 시장 공략과 시청자 기반 확대에 효과적일 수 있다. 다

만 글로벌 시장 진출에는 지역별 선호, 규제 환경, 수익화 모델 등에 대한 세부적인 이해가 요구된다. 성공 여부는 콘텐츠의 현지화 능력, 다양한 시청자와의 신뢰 구축, 점점 복잡해지는 경쟁 환경을 헤쳐 나가는 역량에 달려 있다. 또한 스트리밍 오디오 사업자는 오프라인 청취 기능, 파일 압축 기술, 저(低)비트레이트(bitrate, 디지털 영상·음성 데이터가 초당 전송·처리되는 데이터의 양) 스트리밍 모드 등을 제공하는 방안도 검토해야 한다. 네트워크가 불안정하고 대역폭 비용이 여전히 높은 지역에서는 이러한 기능이 핵심 경쟁력으로 작용한다.

구독형 스트리밍 비디오 사업자도 영상 팟캐스트 봄을 기회로 활용할 수 있다. 대표적으로, 시즌과 시즌 사이 시청자의 관심을 유지하고 콘텐츠 포트폴리오를 확장하기 위해 동반 영상 팟캐스트(companion vodcast)를 출시하는 전략이 있다.²⁸ 또한 팟캐스트를 새로운 스토리와 신인 창작자를 발굴하는 저비용 인큐베이터로 활용할 수 있다. 이미 형성된 팬층을 기반으로 팟캐스트를 스트리밍 화면으로 전환함으로써 TV 및 영화 제작 리스크를 낮추면 파이프라인 개발 부담을 크게 줄일 수 있다.²⁹ 아울러, 이미 충성 팬덤을 보유하고 소셜에서 화제를 만들 줄 아는 인기 창작자와 협업하는 것은 문화적 관련성을 빠르게 확보하는 지름길이 될 수 있다. 이러한 전략은 단순히 영상 팟캐스트 유행을 좇는 것이 아니라, 오디오 중심 스토리텔링을 시청자 유지, 팬 심층 참여, 지속 가능한 성장을 이끄는 전략적 엔진으로 전환하는 과정이다.

Korean Perspectives

영상 팟캐스트의 부상과 한국 미디어·광고 산업의 재편

8 상 팟캐스트의 부상은 한국 미디어·광고 산업의 경쟁 구도를 빠르게 재편하고 있다. 오디오 중심이던 팟캐스트가 거실의 대형 스크린으로 확장되면서, 국내 방송사와 OTT가 장악해 온 프리미엄 시청 시간이 알고리즘 추천 기반의 룽풀 대화형 콘텐츠로 이동하고 있다. 표정·제스처·스튜디오 연출이 결합된 영상 팟캐스트는 강한 파라소셜 관계를 형성하며, 오디오 대비 높은 체류 시간과 몰입도를 기록하고 있다. 이는 국내 사업자에게 단순한 포맷 확장이 아니라, 제작 효율성·콘텐츠 노출 전략·수익화 구조 전반을 다시 설계해야 하는 변화다.

광고 산업에 미치는 영향은 더욱 직접적이다. 영상 팟캐스트는 로고 노출, 제품 배치, 호스트 중심 스폰서십에 적합하며, 클립 단위로 세부 분해되어 소셜 플랫폼에서 빠르게 확산되며 퍼포먼스형 광고로 전환된다. 실제로 국내 주요 브랜드들에서는 유튜브 기반 영상 토크 콘텐츠와 속클립을 활용해 광고 효율을 높이고 있으며, 네이버와 카카오는 크리에이터·브랜드 협업 콘텐츠를 커머스와 연계하는 실험을 확대하고 있다. 다만 신흥 시장에서는 낮은 CPM이 수익성을 제약할 수 있어, 단순 광고를 넘어 커머스 연계형 광고와 고부가 스폰서십을 통해 ARPU를 끌어올리는 전략이 중요해지고 있다.

글로벌 확장 측면에서도 변화는 뚜렷하다. 모바일·저대역폭 환경을 기반으로 신흥국 소비가 빠르게 늘면서, 다국어 콘텐츠 역량은 선택이 아닌 기본 요건이 되고 있다. 자동 번역·자막·립싱크 등 생성형 AI 기술을 제작 파이프라인에 내재화하고, 네트워크 친화적 재생 환경을 갖춘 플랫폼만이 해외 트래픽을 실제 매출로 전환할 수 있다. 이는 CJ ENM, 네이버, 카카오엔터테인먼트 등 국내 기업들이 처음부터 글로벌 유통을 전제로 포맷과 인프라를 설계하는 이유이기도 하다. 기업별 대응 방향도 점차 타겟 세분화되고 있다.

방송사와 OTT는 본편 사이 공백을 메우는 ‘동반 영상 팟캐스트’(companion vodcast)를 통해 팬덤을 유지하고, 저비용 인큐베이터로 신인 크리에이터와 신규 IP를 발굴할 수 있다. 웨이브, 쿠팡플레이 등은 예능·스포츠 IP와 연계한 토크형 콘텐츠로 체류 시간을 확장할 여지가 크다. 광고주와 브랜드는 영상 팟캐스트 클립을 대량 생성·테스트하는 실험 체계를 구축하고, 클릭-투-바이(click-to-buy) 등 직접 구매 연계를 강화해야 한다. 크리에이터와 MCN(multi-channel network) 역시 롱폼과 솟폼을 유기적으로 연동하고, 완주율·구독 전환율 중심의 데이터 운영 역량을 고도화해야 한다.

결국 영상 팟캐스트는 한국 미디어·광고 산업을 포맷, 유통, 수익화, 글로벌 확장 측면에서 동시에 재구성하는 핵심 변수다. 동반 영상 팟캐스트 전략, 발견성 중심 UX, 광고·커머스 결합 수익 모델, 생성형 AI 기반 다국어 확장, 그리고 신뢰를 뒷받침하는 거버넌스 체계를

선제적으로 구축하는 기업만이 스크린 타임을 다시 확보하고, 팬덤 기반의 장기 경쟁력을 확보할 수 있을 것이다.



이대의 파트너

한국 딜로이트 그룹
컨설팅 부문 Customer 본부장,
Deloitte Digital 리더



Part

04

인프라에서 경험으로... 통신의 재정의

12장 차세대 위성 인터넷 시대로의 전환과 통신 인프라·경쟁 질서의 재편

13장 통신사 경쟁력은 네트워크 속도에서 고객이 체감하는 가치로 이동

12장

차세대 위성 인터넷 시대로의 전환과 통신 인프라·경쟁 질서의 재편

D2D(direct-to-device, 위성-단말 직접 연결)기술의 진전과 저궤도(LEO) 위성 인프라의 확대는 차세대 위성 인터넷의 조기 확산을 이끌고 있다. 단말 직접 연결과 저지연·고속 통신이라는 강점으로 주목받고 있지만, 현재로서는 지상 통신망의 사각지대를 연결·보완하는 역할에 머물러 있다. 동시에 수익성 제약과 복잡한 규제·주파수 관리 이슈가 주요 도전 과제로 부각되면서, 글로벌 통신산업의 경쟁 구도와 산업 구조에도 점진적인 변화가 예상되고 있다.

핵심 내용 요약 (Executive Summary)

» LEO 위성군 및 D2D 기능 위성의 급성장

- 2026년 말까지 LEO 위성군 1.5만~1.8만기, D2D기능 위성 1,000기 이상 궤도 진입 전망
- LEO 위성 시장 매출 연간 약 150억 달러, 가입자 수 1,500만 명 이상 예상
- 이동통신사(MNO)와 위성 기업 간 전략적 파트너십 급증
- 오지, 농촌, 해양 등 지상망 취약 지역에서 필수 연결 인프라로 자리매김

» (규제 및 주파수 관리의 중요성 증가) 위성-지상망 통합의 핵심 변수로 규제와 주파수 관리가 부상

- 미국, 유럽, 아시아 등 지역별 규제 환경이 상이하며 복잡성 심화
- 주파수 간섭 방지와 공정한 자원 배분이 지속 성장의 관건

» (LEO 위성 서비스의 경쟁력 강화) 저지연·고속 광대역 제공으로 기존 위성 대비 경쟁력 우위

- 신규 위성군 구축과 가입자 확대가 빠르게 진행
- 일부 사업자는 저가 구독 모델로 신흥국에서 지상 통신사와 직접 경쟁 가능성

» (시장 확산의 제약) 수익 모델의 불확실성과 데이터 트래픽 증가에 따른 기술 투자 압박

- 저밀도 지역의 낮은 수익성, 높은 인프라 구축 비용, 단말 가격 부담
- 위성 용량 확충 및 적응형 빔포밍, 동적 주파수 공유, 위성 간 레이저 링크(ISL), AI 기반 네트워크 최적화 등 첨단 기술 도입

» 자본지출(CAPEX) 구조 변화와 협업 확대

- D2D·LEO 확산으로 위성기업·통신사 모두 투자 구조 재편
- 파트너십을 통한 비용 분담과 연결성 확장이 핵심 전략으로 부상

위성 연결 서비스는 D2D(direct-to-device, 기기 직접 연결) 방식을 기반으로 성장세를 유지하고 있지만, 수익화에는 여전히 난관을 겪고 있다. 한편 저궤도(LEO) 데이터 전송 기술이 발전하면서 위성 통신의 배치 방식과 복원력이 재편됨과 동시에 규제 환경이 한층 복잡해지고 있다.

위성 연결 서비스가 과거 어느 때보다 빠르게 확산되고 있다. D2D 위성은 수익화 어려움에도 불구하고 빠르게 확산되고 있으며 LEO 광대역 위성군도 확대되고 있어, 통신사들은 새로운 기회와 잠재적 혼란을 모두 다뤄야 하는 상황이다. 이와 맞물려 기술 혁신으로 위성 배치 속도 향상, 복원력 강화, 비용 절감이 가능해져 통신 산업 지형을 재편하고 있다. 다만 규제 이슈와 주파수 관리 문제가 위성-지상망 통합과 지속 가능한 성장의 핵심 변수로 떠오르고 있다.*

*LEO는 위성 기반 광범위 통신 인프라 기술, D2D는 위성-단말 직접 연결 기술로 상호 보완적으로 활용되며 기존 인프라 한계를 보완하는 역할을 한다.

일각에서는 LEO 위성군이 2026년에 연간 150억 달러 규모의 매출을 달성할 것이라는 전망이 제기된다.¹ 딜로이트 역시 글로벌 가입자 수가 2026년 말 1,500만 명을 넘어설 것으로 전망한다.² 또한 통신용 LEO 위성군이 5개 군으로 확대되고, 여기에 포함될 위성 수는 1만5,000~1만8,000기에 달할 것으로 예측한다.^{3,4} D2D 위성 용량 투자도 2026년 60억~80억 달러에 이를 것으로 보이며, 연말까지 D2D 기능을 탑재한

1,000기 이상의 위성이 궤도에 진입할 것으로 전망한다. 다만 비즈니스 모델과 수익화 방식이 여전히 불확실하기 때문에, 현재로서는 D2D 매출 전망을 제시하지 않는다.

2019년부터 시작된 초기 단계에서는 다수의 LEO 위성이 지상의 소형 패널형 안테나와 연결돼 저지연 광대역 서비스를 제공해 왔다.⁵ 이를 통해 지상 통신망이 부족하거나 아예 없는 지역에서도 합리적인 가격으로 인터넷 이용이 가능해졌다. 그러나 이러한 신호는 피자판 크기의 단말기를 설치해야만 스마트폰에서 수신할 수 있다는 한계가 있었다.⁶

이후 신형 장비 및 안테나가 개발되고 새로운 규제가 마련되면서, 2023년부터 주로 LEO에 배치된 새로운 위성들이 스마트폰 등 개인 단말기와 직접 연결되기 시작했다. 다만 당시의 D2D는 저(低)비트레이트(bit-rate, 디지털 영상·음성 데이터가 초당 전송·처리되는 데이터의 양) 문자 메시지 수준에 가까웠고, 전송 속도도 LEO 접시형 단말기가 제공하는 50Mbps(초당 백만 비트)를 넘지 못했다.⁷ 향후 D2D는 더 높은 속도를 제공할 수 있을 것으로 전망되지만, 접시형 단말기 기반 서비스의 속도에는 여전히 못 미칠 가능성이 크다.⁸ LEO와 D2D는 사용하는 궤도와 위성이 일부 겹칠 수는 있지만,⁹ 두 서비스가 동일한 개념은 아니다.

2025년 9월, 한 주요 LEO 사업자는 D2D용 5G(5세대) 주파수 블록을 인수했다.¹⁰ 다만 해당 주파수를 송수신할 수 있는 신형 스마트폰 칩셋 개발, 해당 대역을 이용할 신규 위성 발사, 동시 처리 가능한 사용자 수,

실내에서도 안정적으로 작동할 수 있는지 등이 아직 불확실하기 때문에 2026년에 즉시 상용화되기는 어려울 것으로 보인다. 그러나 2028년 즈음에는 스마트폰 사용자가 우주에서 직접 영상 스트리밍을 받는 시대가 열릴 가능성이 있다.¹¹

D2D, 통신 공백 채울 잠재력 크지만 수익화는 여전히 난제

D2D 기술을 활용하면 위성이 스마트폰 같은 표준 소비자 단말기와 직접 통신해, 기존 지상 인프라를 우회한 저대역폭 기반 필수 연결 서비스를 제공할 수 있다.¹² 이런 기능은 지상 이동통신망이 부족하거나 전무한 오지, 농촌, 해양 지역에서 특히 유용하다.

딜로이트는 2023년 연구에서 D2D 위성통신 시장의 네트워크 인프라(주로 위성) 투자 규모가 약 30억 달러에 달할 것으로 예측한 바 있다.¹³ 실제로는 이를 웃돌아 2024년 약 40억 달러의 투자가 집행됐다.¹⁴ 기업의 로드맵과 공개된 투자 계획을 보면, 2026년에는 총 60억~80억 달러의 자본지출이 필요할 것으로 전망된다. 이 중 85~90%는 신규 위성 발사, 나머지 10~15%는 기존 위성 교체에 쓰일 것으로 보인다.¹⁵

이처럼 빠른 확장은 여러 요인이 복합적으로 작용한 결과다. 첫째, 스마트폰 제조사와 칩 벤더가 위성 연결 기능을 기기에 본격적으로 탑재하기 시작한 것이 큰 요인으로 작용했다. 딜로이트는 2024년에 2억 대 이

상의 위성 연결 가능 스마트폰이 판매되고, 그 과정에서 20억 달러 규모의 특수 칩이 사용될 것으로 예상했는데,¹⁶ 실제로 주요 제조사 대부분이 위성 메시징 기능을 갖춘 플래그십 모델을 출시했다.¹⁷

둘째, 이동통신사와 위성기업 간 파트너십이 급증하면서 D2D 서비스 공급이 확대됐다.¹⁸ 이러한 협업 덕분에 모바일 서비스가 전혀 제공되지 않는 지역에서 긴급 SMS(문자 메시지)와 저대역폭 데이터 등 기본 서비스가 가능해졌으며, 방대한 통신 공백 인구를 대상으로 시장 기회가 창출됐다.

지상 통신사업자들은 국내총생산(GDP)이 낮은 대다수 신흥국의 농촌과 산간 등 저밀도 지역을 경제성이 낮은 시장으로 보기 때문에, 일반적으로 높은 수익을 낼 수 있는 도시 지역에 투자를 집중한다. 실제로 일부 국가에서는 저밀도 지역의 기지국당 수익이 도시보다 10배 낮은 반면, 설비 구축 및 운영 비용은 2~3배 더 높다.¹⁹ 그렇기 때문에 통신사는 이런 지역에 인프라 투자를 회피할 수밖에 없다. 하지만 2024년 말 기준, 전 세계 인구의 4%에 해당하는 약 3억5,000만 명이 모바일 인터넷 서비스가 거의 제공되지 않는 외딴 지역에 거주하고 있어 D2D 기회는 크다. 다만 이 지역은 소득 수준이 낮아 사용자들이 D2D나 LEO 데이터 서비스 요금을 감당하기 어렵다는 문제가 있다.²⁰

셋째, 글로벌 규제기관과 표준화 기구들은 비(非)지상 네트워크(NTN)를 수용하기 위해 빠르게 움직이고 있다. 이를 위해 주파수를 배분하고

5G NTN 표준을 확정해, 일반 스마트폰도 위성과 끊김 없이 연결될 수 있는 기반을 마련하고 있다.²²

LEO 위성군, 초고속 확장과 저렴해진 네트워크 연결로 새로운 경쟁력 확보

LEO 위성 서비스는 특수 설계된 위성을 낮은 궤도에 배치하고, 이를 지원하는 지상 장비를 결합해 고속·저지연 광대역을 제공하는 방식으로, 그 규모가 계속 확대되고 있다. 매주 새로운 위성이 LEO에 추가되고 있으며, 신규 위성군 구축이 이어지고, 가입자 수와 매출도 증가하고 있다.

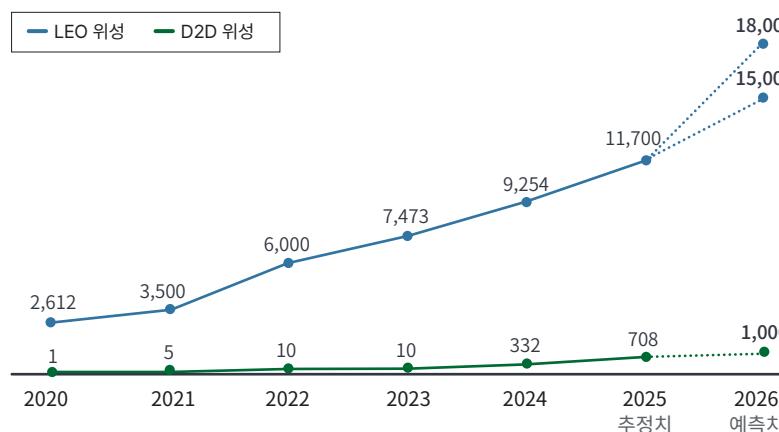
2026년에 상용화 예정인 아마존(Amazon)의 카이퍼(Kuiper) 프로젝트²³는 다른 LEO 서비스보다 더 낮은 가격대를 제시해 개발도상국에서 지상 광대역 사업자와 직접 경쟁할 가능성이 있다. 이와 함께 중국의 구왕(Guowang) 광대역 거대 위성군과 천판(Qianfan, G60/Spacesail) 프로젝트,²⁴ 캐나다의 텔레샛 라이트스피드(Telesat Lightspeed),²⁵ 유럽의 아이리스2(IRIS²) 등도 2026년 위성을 발사하거나 향후 수년 내 배치할 계획 중이다.²⁶ 아랍에미리트연합(UAE) 기반의 오빗웍스(Orbit-works)처럼 지역 단위의 신규 사업도 추진되고 있다.²⁷ 유럽의 유텔샛 원웹(Eutelsat OneWeb)과 같은 기존 사업자들은 용량 확장, 지연시간 개선, D2D 연결성 확보를 위해 위성군을 업그레이드하고 있다.²⁸

지난 2022년 딜로이트는 2023년 말까지 5,000기 이상의 광대역용 LEO 위성이 배치되고 약 100만 명의 가입자가 초고속 인터넷을 이용할 것이라고 예측했으나,²⁹ 실제 증가 속도는 이를 훨씬 앞질렸다. 2023년 말 기준 LEO에는 약 7,473기의 활성 광대역 위성이 운영되고 있었다.³⁰

LEO 위성은 기존 정지궤도(GEO) 위성보다 지연시간이 낮고 속도가 빠르기 때문에, LEO 가입자가 증가하면서 GEO 기반 인터넷 제공업체는 점유율을 일부 빼앗겼다.³¹ LEO는 전통적인 지상 통신망에 접근하지 못한 비연결·저연결 지역 사용자를 주요 대상으로 하지만, 현재까지는 동일 품질의 지상 광대역보다 가격이 높은 경우가 많다.³²

LEO 서비스의 유통 모델은 혼합형이다. 일부 사업자는 소비자에게 직접 판매하고, 일부는 지상 통신사와의 파트너십을 통해 판매하며, 일부는 하이브리드 방식을 채택한다. 특히 2026년에는 몇몇 사업자가 더 저렴한 구독 모델을 제시하며, 특히 신흥국 시장에서 지상 통신사와 직접 경쟁에 나설 가능성이 있다.³³

그림 1
LEO 및 D2D 위성 개수 전망



출처: Deloitte analysis of publicly available satellite industry data and reported deployment milestones for 2020-2025; estimates reflect annual counts of active LEO satellites derived from industry trackers, and D2D-enabled satellites based on launch activity and service demonstrations.

협력 vs. 경쟁으로 갈리는 마케팅 전략

LEO 광대역 및 D2D 위성 시장이 진화함에 따라, 딜로이트는 관련 시장의 마케팅 전략이 협력과 경쟁의 두 가지 방향으로 갈릴 것으로 전망한다.

1. LEO 사업자의 본격적 경쟁 시대 도래

LEO 사업자들은 특정 지역에서 현지 통신사와의 파트너십을 빈번히 체결하고 있으며, 일본·호주·필리핀 등지에서도 D2D 서비스 관련 유사한 협력 사례가 나타나고 있다.³⁴ 반면 일부 위성 사업자들은 개발도상국을 중심으로 직접 경쟁 전략에 적극 나설 것으로 전망된다. 이들 사업자는

지상 통신사보다 월등히 저렴한 요금제를 내세워 미충족 수요가 큰 시장을 확보하려 하며, 공격적 가격 설정과 단순화된 서비스 제공 모델을 통해 점유율 확대를 노리고 있다. 여러 신규 LEO 위성군이 계획되거나 구축 중이지만, 아마존 카이퍼가 다음 주자로서 의미 있는 규모의 서비스를 시작할 가능성이 가장 큰 사업자로 꼽힌다.³⁵ 카이퍼는 저가 월 구독 모델을 준비하고 있으며, 미연결 또는 취약 연결 지역을 직접 공략함으로써 지상 통신사에 위협이 될 수 있다.³⁶

다만 모든 지상 광대역 시장이 위성 사업자에 의한 시장 교란에 영향을 받는 것은 아니다. 일부 선진국의 평균 광대역 요금은 월 33~80달러 수준인 반면, 몇몇 개발도상국은 월 10달러 미만의 초저가 요금제를 제공하고 있어, 상대적으로 저렴한 LEO 서비스라도 이러한 시장에서는 점유율을 빠르게 확대하기가 쉽지 않을 수 있다.³⁷ 반면 브라질이나 남아프리카공화국처럼 월 21~48달러 수준의 요금대를 형성한 시장에서는 LEO 사업자가 저가 전략을 취하거나 보조금을 적용할 경우 이용자를 확대하기가 더욱 용이하다.³⁸ 저가 연결 서비스를 제공하면 소비자가 전자상거래와 스트리밍 등 다른 온라인 서비스를 보다 적극적으로 사용할 수도 있다. 한편 월 이용료가 15달러로 낮은 수준이더라도 LEO 서비스 이용에 필요한 지상 단말기가 해당 200~500달러에 달하기 때문에,³⁹ 개발도상국의 소비자에게는 비용 부담이 클 수 있다.⁴⁰ 따라서 단말기 가격 인하가 시장 확산의 핵심 요인이 될 것이라는 점도 주목할 필요가 있다.

2. 위성 용량의 전면적 혁신, 차세대 네트워크 연결의 문을 여는 열쇠

개별 위성과 그 위성이 속한 위성군의 용량은 LEO 광대역 서비스와 D2D 기본 연결을 포함한 위성 기반 통신의 효과성, 신뢰성, 상업적 지속 가능성을 좌우하는 핵심 요인이다. 2025년 말까지 글로벌 위성 데이터 트래픽은 현재 대비 20배 증가할 것으로 전망되며, 이는 위성 용량에 상당한 부담을 가한다.⁴¹ 따라서 광역 커버리지 제공, 고속 데이터 전송 지원, 다수 사용자의 동시 접속을 실현하려면 위성 용량을 반드시 확충해야 한다. 현재 네트워크는 인구가 거의 없는 지역에서 수십 명 정도의 가입자를 연결하는 데에는 충분한 용량이지만, 인구 밀도가 조금만 높아져도 병목이 발생한다. LEO는 차세대 단계로 나아가기 위해 신규 기술 도입이 반드시 필요하다.⁴²

위성 용량의 가용성은 기술과 규제의 영향을 동시에 받는다. 기술적으로는 위성 수량, 개별 위성의 성능, 궤도 배치가 중요하다.⁴³ 이론적으로는 LEO가 1,000만~1,200만 개의 위성을 수용할 수 있을 것으로 추정되지만,⁴⁴ 실제로는 충돌 위험, 추적 오차, 규제 승인 지연 등의 제약으로 인해 지속가능 한계치는 약 10만 개 수준으로 평가된다.⁴⁵ 또한 Ku/Ka 대역(위성통신에서 사용되는 마이크로파 주파수 대역) 등 특정 주파수 자원의 가용성도 위성사업자가 용량을 확장하는 데 장애가 된다.⁴⁶ 실제로 글로벌 서비스 커버리지를 갖춘 스타링크(Starlink)조차 때때로 네트워크 혼잡 때문에 일시적으로 서비스를 중단한 적이 있으며,⁴⁷ 영국 남동부 일부 지역에서는 가입자 수를 제한해야 했던 경험도 있다.⁴⁸ 이로 인해 충분한 용량 확보의 중요성이 부각되고 있다.

이에 대응하기 위해 다수의 LEO 기업은 적응형 빔포밍(adaptive beam-forming)*, 동적 주파수 공유, 위성간 레이저 링크(ISL), AI 기반 네트워크 최적화 등의 고도 기술을 도입하고 있다.⁴⁹ 아울러 고용량 신형 위성, 향상된 위성 아키텍처, 효율적 주파수 배분을 위한 규제 조정 노력이 병행돼야 수요 증가와 기술 발전 사이 균형을 잡을 수 있다.

* **빔포밍(adaptive beamforming)**은 안테나 배열에서 전파의 위상·세기·방향을 실시간으로 제어해 특정 사용자나 지역에 신호를 집중시키고, 간섭은 최소화하는 무선 통신 기술을 의미한다. 적응형 빔포밍은 이용자 위치·트래픽·간섭 환경 변화에 따라 빔 패턴을 동적으로 조정하는 것을 뜻한다.

LEO는 지상 대안이 없는 사용자에게는 매우 유용하지만, 대부분의 선진국 시장에서 지상 통신사와 본격적으로 경쟁할 가능성은 적다. 예를 들어 영국 일부 지역에서는 이미 제곱킬로미터당 약 0.35명 수준으로 가입자 밀도가 한계에 근접해 있으며, 한 분석에 따르면 스타링크의 협력 네트워크는 영국 내 약 20만 가구(0.7% 점유율)만을 지원할 수 있는 것으로 나타났다. 동일한 인프라 기준으로 스타링크의 잠재 점유율은 독일 0.4%, 스페인 1.4%로 추정된다. 설령 위성군이 버전2 모델로 완전 교체되더라도 영국 점유율은 약 1.4% 수준에 그칠 것으로 보이며, 제안된 1만5,000기 규모 위성군이 완전히 가동된다 해도 3~4%가 최대치다. 향후 버전3 위성군이 구축된다면 8~10%도 가능하지만, 이는 10년 이상 장기간의 기술 진전이 필요한 시나리오이다. 2025년 8월 기준, 궤도에 진입한 버전3 위성은 단 한 기도 없다.⁵⁰

LEO 및 D2D 전체 용량을 확장하는 또 다른 핵심 요소는 지상국(ground station)이다. 지상국은 위성과 대형 데이터센터 간 데이터를 연결 및 조율하고, 위성 네트워크를 관리하며, 위성으로 신호를 송신하는 핵심 인프라다. 2025년 기준 이미 100개 이상의 LEO용 지상국이 존재하며, 다음 위성군을 지원하기 위해 추가로 100개 이상이 필요할 전망이다.⁵¹ 위성간 레이저 통신 기술이 확산하면서, 데이터의 일부는 지상국을 거치지 않고 궤도 내에서 위성간 직접 전송이 가능해지고 있으나, 그렇다 하더라도 지상국의 필요성이 완전히 사라지지는 않을 전망이다.⁵² 마지막으로, 지상국은 데이터센터와 광섬유로 연결해야 용량을 극대화하고 지연을 최소화할 수 있어, 이는 지상 광섬유 제공 사업자(주로 통신사)에게 새로운 수익 기회가 될 수 있다.

주파수 관리 시 규제 고려사항

위성통신 시장이 확대됨에 따라 주파수 배분과 관련된 규제 쟁점이 더욱 중요한 사안으로 부상하고 있다.

딜로이트는 D2D 서비스를 제공하는 LEO 위성 네트워크 사업이 상당한 규제의 제약에 직면할 것으로 전망한다. 이를 네트워크는 이미 지상 이동통신 서비스에 할당된 주파수 대역에서 운영해야 하기 때문이다. 미국과 유럽처럼 국가 및 지역 규제 기관이 주파수 간섭 방지와 공정한 주파수 사용을 위해 주파수 할당을 엄격히 관리하는 지역에서는 특히 복잡한 규제 요건에 부합해야 한다.⁵³

미국의 경우 연방통신위원회(FCC)는 D2D 연결을 지원하기 위해 '우주에서의 추가적 커버리지'(SCS, Supplemental Coverage from Space) 프레임워크를 도입해 위성사업자가 지상 네트워크와 통합할 수 있는 기반을 마련했다.⁵⁴ 더불어,

국가통신정보국(NTIA)이 추진하는 425억 달러 규모의 광대역 인터넷 보급 사업(BEAD, Broadband Equity, Access, and Deployment) 관련 정책 고시는 LEO 위성사업자에게 새로운 자원 조달 기회를 제공하는 방향으로 정책 기조가 전환되고 있음을 시사한다.⁵⁵ 해당 정책은 기술 중립적 접근을 채택해 기존의 광섬유 기반 네트워크 중심 정책을 철폐하고 성능 기준에 기반한 평가체계를 도입함으로써, LEO 위성이 전통적 광대역 기술과 동등한 위치에서 경쟁할 수 있는 여건을 조성했다. 이로 인해 LEO 관련 지원금은 기존 약 40억 달러에서 100억~200억 달러 규모로 확대될 전망이다.⁵⁶

유럽에서는 국가별 규제 기관이 EU 및 유럽우정·통신행정회의(CEPT)가 제시한 규제 틀 내에서 주파수 배분을 관리하고 있어 규제 구조가 분산화돼 있다.⁵⁷ CEPT는 위성 서비스와 지상 이동통신망 간 통합을 둘러싼 기술 및 규제 과제를 적극적으로 검토 중이다.⁵⁸

아시아 또한 유사한 규제가 마련돼 있지만, 국가별 정책과 기술, 인프라 성숙도 차이로 인해 규제 환경이 더욱 복잡하다. 인도, 중국, 일본 등은 주파수 간섭을 최소화하면서 혁신과 경쟁을 촉진할 수 있도록 지상망과 위성망의 조화로운 주파수 공동 사용을 위한 규제 프레임워크를 검토하고 있다. 예를 들어, 인도는 통신규제기구(TRAI)를 중심으로 종합적 주파수 관리 지침을 마련 중이며,⁵⁹ 중국은 공업정보화부(MIIT)가 위성통신을 수용하기 위한 대규모 규제 개혁을 추진하고 있다. MIIT는 주파수 배분 행정절차 간소화, 주파수 간 간섭 관리, 위성혁신 촉진을 위한 포괄적 정책을 마련하고 있다. 특히 위성 서비스와 지상 이동통신 인프라의 통합을 용이하게 하는 규제 프레임워크를 구축함으로써 전국적 디지털 네트워크 연결 확대라는 국가 전략 목표를 뒷받침하고 있다.⁶⁰ 일본 또한 총무성을 중심으로 위성-지상 간 주파수 공존을 위한 규제 체계를 정교화하고 있다.⁶¹

자본지출 구조 변화와 위성-지상 협업의 전략적 부상

D2D와 LEO가 성장함에 따라 주시해야 할 변화는 위성 기업과 지상 통신사업자 모두에게 영향을 미치는 자본지출 구조의 변화다. 딜로이트는 2026년 말까지 D2D 위성과 LEO 광대역 위성군에 대한 누적 투자가 약 100억 달러에 이를 것으로 예상한다.⁶² 100억 달러는 2019년 이후 수년간에 걸쳐 투입된 금액이지만, 만약 특정 연도에 일시 투입됐다 하더라도 2025년 기준 연간 3,000억 달러에 달하는 글로벌 통신사의 자본지출과 비교하면 미미한 수준이다.⁶³

지상 통신사업자에게 D2D-LEO 파트너십이 중요한 이유는, 극도로 비용 효율적인 방식으로 전 인구 100%에게 연결성을 제공해야 한다는 지속적인 정책적·사업적 압력을 충족할 수 있기 때문이다. 유선 또는 무선 지상 네트워크로 외딴 지역을 커버하려면 막대한 투자가 필요하지만, 우주 기반 솔루션 제공업체와 협력하면 지상사업자의 자체 자본지출이 필요하지 않다. 또한 해당 기업에 직접 투자하면 비용 대비 훨씬 경제적인 방식으로 운영 할 수 있다. 예컨대 보다폰(Vodafone), AT&T, 버라이즌(Verizon) 등 글로벌 통신사들은 AST 스페이스모바일(AST SpaceMobile)에 투자했으나, 투자금액은 이들 통신사의 연간 자본지출에서 극히 작은 비중을 차지한다.⁶⁴

LEO 위성군은 수명이 평균 4~5년에 불과해 지속적인 교체가 필수적이며, 이로 인해 매년 약 20~25%의 위성을 신규 교체해야 한다.⁶⁵ 따라서 자본지출은 장기적으로 높은 수준을 유지할 가능성이 크다.

위성 기반 광대역은 전통적 지상통신의 강력한 대안으로 부상하고 있으며, 특히 개발도상국에서 그 역할이 두드러지고 있다. 예를 들어 나이지리아에서는 한 LEO 사업자가 시장 진출 2년 만에 국내 2위 인터넷 서비스 제공업체로 성장했다.⁶⁶ 현재 지상 광대역 네트워크 인프라가 저조한 대다수 신흥국 시장에서 단일 LEO 사업자 또는 복수 LEO 사업자가 최대 인터넷 제공자로 자리 잡을 가능성도 제기된다.

규제 환경 또한 큰 변화를 겪을 전망이다. 각국 정부는 혁신 촉진, 시장 경쟁력 확보, 국가안보, 주권적 통제 간 균형을 맞추기 위해 새로운 규제를 준비 중이다. 주파수 배분, 우주 파편 관리, 사이버보안 관련 국제 규정과 표준은 LEO 위성 확대로 인해 발생할 복잡한 문제를 해결하기 위한 방향으로 새롭게 정립될 가능성이 높다. 또한 긴급통신시스템(Public Emergency Communications System) 및 공공 안전 관련 요건은 국가마다 상이하며, 단말기 제조사는 국가별 규제를 충족해야 한다.

LEO 사업자는 지상 이동통신사와의 협력 과정에서 복잡한 계약 체계를 관리해야 하며, 간섭을 방지하는 조건 하에 주파수 공유 또는 임대 전략을 활용해야 한다. 동적 주파수 배분, 지리적 빔 쇼이핑(geographic beam shaping) 등 고도화된 주파수간 간섭 관리 전략도 핵심 규제 우려로 간주된다.⁶⁷ 지상 사업자의 권리를 보호하면서도 위성 기반 연결성을 강화하기 위해서는 강력한 규제 감독과 위성-지상 사업자 간 협업 모델이 필수적이다.

지금은 소비자용 LEO 시장에 초점이 맞춰져 있지만, 향후 몇 년 내 기업용 LEO 시장이 크게 확대될 가능성이 있다. 2030년까지 기업용 가입자는 현재 대비 약 10배 증가한 340만 명에 이를 것으로 예상된다.⁶⁸ 절대 가입자 수는 소비자 시장보다 적지만, 월 평균 매출(ARPU)은 더 높고 이탈률은 더 낮아, 고수익 시장으로 자리 잡을 가능성이 큰 만큼 시장 창출 기회를 예의주시할 필요가 있다.

Korean Perspectives

D2D·LEO에 대한 국내 통신사들의 기본 인식과 투자 전개 양상

H내 통신사들은 D2D·LEO 기반 차세대 위성 인터넷을 지상망을 대체하는 수단이 아니라, 산악·해양 및 재난 상황 등 기존 인프라가 취약한 구역을 메우는 보완 인프라로 인식하고 있다.

SK텔레콤은 위성망을 AI 기반 네트워크의 보완 계층으로 보고 재난·안전·B2B 영역을 중심으로 활용 가능성을 검토하고 있다. KT는 위성·5G·광망을 결합한 통합 커버리지 구축에 주력하는 한편, 공공·재난·해양 통신 등 정부·공공 부문 수요와의 연계를 핵심 전략으로 삼고 있다. 반면 LG유플러스는 비용 효율성과 사업성을 보다 엄격하게 검토하는 입장으로, 직접 투자보다는 글로벌 위성 사업자와의 제휴 및 도매형 모델을 중심으로 한 신중한 접근을 유지하고 있다.

국내에서 D2D·LEO 기반 위성 서비스의 확산 속도가 더딘 이유는 몇 가지 구조적인 요인에 기인한다. 한국은 이미 LTE·5G 커버리지와 품질이 세계 최고 수준에 도달해 있어, 위성 기반 서비스가 소비자에게 제공할 수 있는 추가적인 차별화 가치가 제한적이다. 저가 구독형 위성 서비스 역시 국내 이동통신 요금과 비교할 때 가격 경쟁력을 확보하기 어렵다. 여기에 위성과 지상망 간 주파수 간섭 규제가 엄격하고, D2D 상용화를 위한 단말 인증·주파수 사용·서비스 정의 체계가 아직 명확하지 않다는 점도 확산의 제약 요인으로 작용한다. 무엇보다도 위성 용량 확보와 기술 고도화에 필요한 투자 규모에 비해, 국내 소비자 시장 기준으로 투자 회수 가능성이 불투명하다는 점은 통신사들에게 부담으로 작용하고 있다.

국내 통신사들이 위성 인프라에 직접적인 투자를 지양하고, 글로벌 위성 사업자와의 로밍·도매·제휴와 단말 제조사·칩 벤더와의 생태계 협업을 통해 위성 기능을 기존 상품과 서비스에 유연하게 흡수하는 전략을 취하고 있는 이유이다. 이들은 오히려 소비자용 D2D 서비스보다는 해운·항공·플랜트·에너지·건설·국방 등 높은 안정성과 신뢰성이 요구되는 미션 크리티컬 영역을 중심으로, B2B 위성 통신을 새로운 수익원으로 삼으려는 전략을 채택하고 있다.

종기적으로(2026~2030년) 보면, 향후 1~2년간 D2D·LEO는 재난 대응과 특정 지역의 보완적 연결 수단에 머물 가능성이 크며, 본격적인 상용 서비스보다는 PoC와 시범 사업 중심으로 전개될 전망이다. 이후 3~5년 구간에는 기업용 LEO 수요가 점진적으로 증가하면서 통신사의 B2B 포트폴리오가 확대되고, 위성 기반 서비스는 프리미엄·특화

서비스로 자리 잡을 가능성이 높다. 이 과정에서 일부 영역에서는 통신사와 위성 사업자 간 협력과 경쟁이 병존하는 구도가 형성될 것으로 예상된다. 결국 국내 통신시장에서 D2D·LEO는 새로운 대중형 소비자 서비스라기보다는, 통신사의 네트워크 전략과 B2B 사업 구조를 재편하는 보완적 인프라이자 전략적 옵션으로 기능할 가능성이 크다.



박형곤 파트너
한국 딜로이트 그룹
TM&E Sector 리더

Part 1

Part 2

Part 3

Part 4

13장

통신사 경쟁력은 네트워크 속도에서 고객이 체감하는 가치로 이동

통신사의 미래 경쟁력은 더 이상 네트워크 속도에서 창출되지 않는다. 소비자들의 통신사 선택 기준이 차별화된 고객 경험 가치로 이동하고 있다. 젊은 세대와 고소득층 고객일수록 체감할 수 있는 부가 가치에 민감하게 반응하고 있는 점을 고려할 때, 6세대(6G) 등 차세대 네트워크 도입 만으로는 수익성 담보가 어려울 것이다. 이제 통신사들은 가격과 기술 경쟁을 넘어, 비 가격적이고 경험적인 가치 제공을 통해 분명한 고객 충성도와 경쟁우위를 확보해야 할 시점이다.

핵심 내용 요약 (Executive Summary)

▶ 네트워크 성능의 한계

- 선진국 시장에서 통신 네트워크의 성능은 이미 상향 평준화되었으며, 추가적인 기술 개발과 성능 향상만으로는 소비자가 체감할 수 있는 가치 제공이 어려운 단계에 진입 → 네트워크 성능 만으로 가입자들의 충성도 확보에 분명한 한계 봉착

▶ 부가 가치 경쟁 본격화

- 일부 국가에서는 통신사 해지와 이동의 주요 원인으로 서비스 불만족이 지목되고 있어, 통신사들은 차별화된 가치 제안과 추가적인 혜택 확대에 집중 → 경쟁의 중심이 기술에서 고객 가치로 이동

▶ 세대·소득별 혜택의 민감도 분화

- Z세대와 밀레니얼 세대, 그리고 고소득 가입자일수록 네트워크 성능보다 체감할 수 있는 부가적인 혜택과 경험에 더 민감하게 반응 → 고객 세분화 기준과 실행 전략의 수정이 불가피

▶ 6G 투자의 수익성 딜레마

- 6G 등 차세대 네트워크 도입 만으로는 소비자가 체감할 수 있는 가치가 제한적이어서, 투자대비 수익성 확보의 불확실성 증가 → 6G 등 기술 중심의 투자 전략은 실질적 성과 창출에 한계

▶ 차별화된 고객 경험 설계

- 추가적인 혜택의 단순 제공은 통신 업계 전반에 빠르게 확산되고 표준화될 가능성이 높고, 통신사 간 차별화된 효과 또한 약화될 수 있어, 독자적이고 일관된 고객 경험 설계가 경쟁우위 확보의 핵심으로 부상

선진국 소비자 일부는 통신 네트워크 성능이 개선된다 하더라도 이를 체감하지 못하고 있다. 이는 통신사들이 시장점유율 확대를 위해 보다 창의적인 서비스와 상품 구성을 모색해야 하는 이유이다.

딜로이트는 2026년 선진국 소비자들이 통신 네트워크 성능보다 통신사의 리워드 프로그램을 더욱 중요시할 수 있다고 전망한다. 따라서 2030년까지 네트워크 업그레이드가 계속되더라도, 가입자 유치와 이탈 방지에는 네트워크 외적인 혜택이 점점 더 중요한 요인으로 작용할 가능성이 있다. 소비자에게는 마르게리타 피자 한 조각이 단독형 5G(5G 표준의 더 완성된 버전)보다 더 매력적으로 보일 수도 있다.¹ 전자가 즉각적으로 인식 가능한 혜택인 반면, 후자는 일반 소비자가 체감하기 어려운 경우가 많기 때문이다.

이러한 흐름은 선진국 모바일 네트워크가 성숙 단계에 접어들었음을 반영한다. 네트워크 속도와 지연시간(네트워크가 반응하는 속도)은 소비자들이 이미 충분히 만족하는 상태에 이르렀다. 지역별로 여전히 음영 지역(신호 없음)이나 특정 혼잡 지역(사용자 수가 용량 대비 과도하게 많은 곳)이 있지만, 소비자 입장에서는 통신사간 차별점을 정교하게 비교하기가 매우 어렵다. 비교 수단을 찾기가 쉽지 않고, 통신 기술을 이해하는 수준도 제한적이며, 긴 시간을 들여 비교하려는 동기 역시 강하지 않기 때문이다.

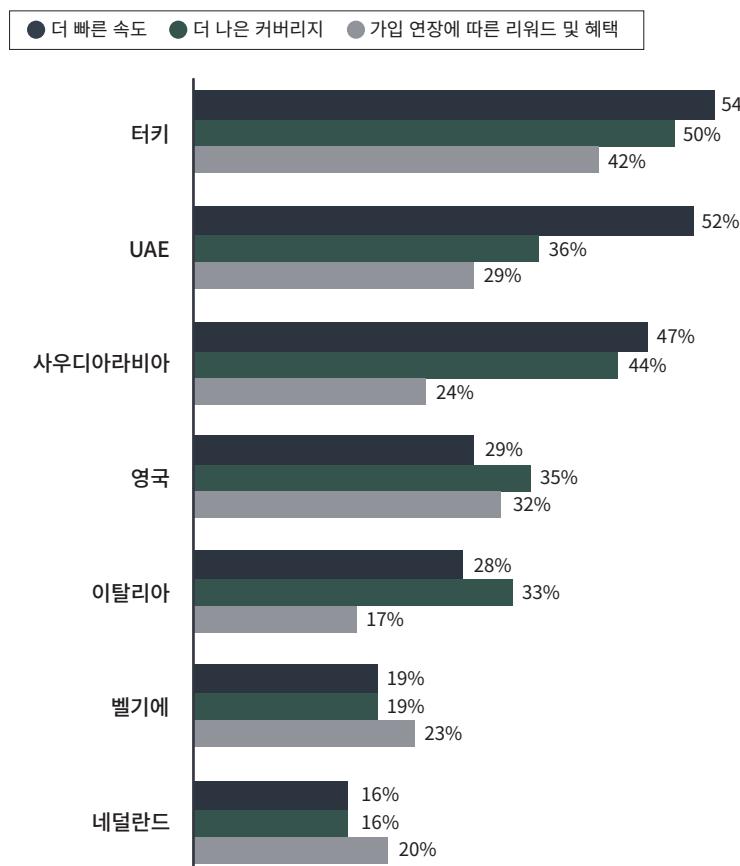
이러한 이유로, 전송 속도 향상이나 지연시간 개선을 내세운 네트워크 업그레이드는 사용자 충성도를 제고하는 데 점차 효용성이 낮아질 수 있다. 많은 이용자가 이러한 기술적 개선을 체감하지 못하거나, 체감하더라도 그 가치를 높게 평가하지 않을 가능성성이 크기 때문이다. 또한 2G/3G 서비스를 종료하고 해당 주파수를 4G/5G로 재할당하는 것 또한 소비자에게는 추상적 변화로만 느껴져 개인적으로 어떤 혜택을 입는 것인지 체감하기 어려울 수 있다.

통신사 차별화 포인트, 네트워크 업그레이드에서 리워드 프로그램으로 전환

각 국가의 이동통신사들이 각기 다른 단계를 밟으며 리워드 기반 차별화를 모색하고 있지만, 결국 모두 방향은 같다(그림 1). 언론 보도에 따르면, 2024년 기준 네덜란드와 벨기에에서는 리워드에 대한 불만족이가입 해지를 유발하는 1순위 요인으로 나타났으며, 영국에서는 2순위 요인으로 집계됐다. 단, 요금은 일반적으로 가장 중요한 요인으로 여겨지기 때문에 분석에서 제외됐다. 반면, 여타 국가의 소비자들은 더 빠른 속도나 더 나은 커버리지를 상대적으로 더 중요한 요인으로 꼽았다.² 그러나 딜로이트는 2030년까지 중기적으로 리워드와 같은 비(非)네트워크 기반 요소를 통한 차별화가 더욱 중요해질 것으로 전망한다.

그림 1 일부 국가의 소비자, 통신사 네트워크보다 리워드를 더 중시

각국 소비자가 통신사를 교체하는 이유로 꼽은 요인(2024년 조사)



참조: 가중치 반영 집계. 영국(3,866명), 네덜란드(1,944명), 이탈리아(1,913명), 벨기에(978명), 터키(973명) 응답자는 휴대폰 또는 스마트폰을 소유한 18~75세 소비자. UAE(915명), 사우디아라비아(874명) 응답자는 휴대폰 또는 스마트폰을 소유한 18~50세 소비자.

출처: Deloitte Digital Consumer Trends 2024.

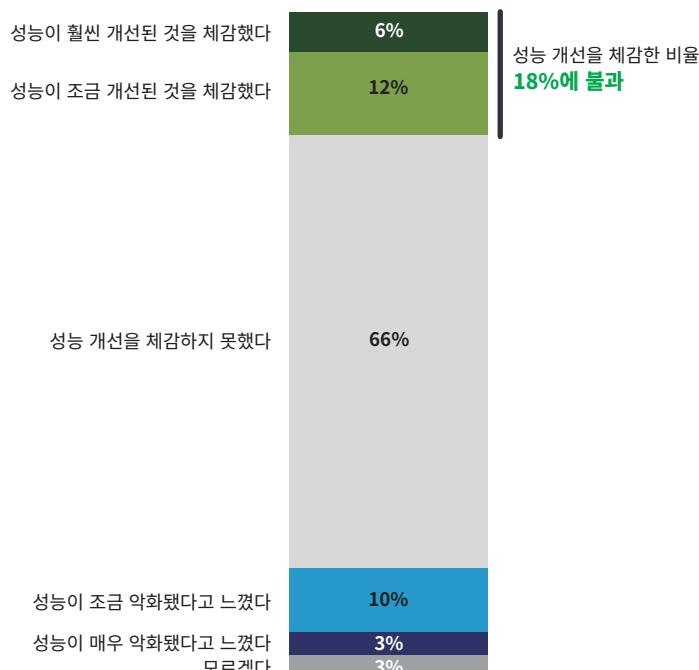
일정 시점이 되면, 네트워크 업그레이드가 이용자의 필요 수준을 넘어서고, 시장 내 모든 이동통신사가 이용자 관점에서 거의 동등하게 인식되는 성능을 제공할 것이다. 이는 1970년대 후반 이후 세대별 네트워크 업그레이드가 거의 항상 즉각적으로 체감되던 때와 대비된다.³ 예컨대 2010년대 초 4G 업그레이드 당시 소비자들은 3G 대비 즉각적이고 확연한 성능 향상을 체감할 수 있었다.⁴ 4G 업그레이드 후 소비자는 이동 중에도 와이파이 급 속도와 지연시간을 경험할 수 있었고, 3G에서는 벼락이던 검색이나 내비게이션 같은 애플리케이션은 4G 환경에서 완전히 다른 차원으로 작동했다.⁵

그러나 2025년 말 현재 5G 네트워크에서만 구동되는 주류 애플리케이션은 거의 없다.⁶ 따라서 4G 시절과 달리 더 우수한 5G 네트워크 성능을 이유로 다른 통신사로 이동해야 할 동기가 크게 약해졌다. 연간 단위의 네트워크 개선도 사용자가 체감하기 어려운 경우가 많다. 예를 들어 2024~2025년 영국 모바일 네트워크의 평균 지연시간은 0.7밀리초(ms) 감소해 18.2ms를 기록했다.⁷ 하지만 1ms는 1천분의 1초에 불과하며, 0.7ms 차이는 인간이 감지할 수 있는 수준이 아니다. 심지어 엘리트 운동선수조차도 반응시간이 약 140ms 수준이다.⁸ 또한 대부분의 주류 애플리케이션은 지연시간이 이처럼 미세하게 줄었다 해서 소비자가 개선 정도를 체감하기 어렵다.⁹ 일례로 음성통화는 지연시간이 150ms 수준이어도 사용자가 거의 지각하기 어렵다.¹⁰ 인터넷 통화(VoIP)와 같은 실시간 애플리케이션을 끊김 없이 사용하는 데 필요한 지연시간은 약 100ms 수준인데, 2025년 영국에서 가장 느린 네트워크 기술이던 3G

조사 이보다 훨씬 적은 평균 42.3ms 지연시간을 기록했다. 딜로이트 영국이 2025년에 실시한 조사에서도 영국 모바일 고객의 3분의 2는 지난 1년간 본인의 네트워크에서 아무런 변화도 느끼지 못했다고 답했다(그림 2).

그림 2 네트워크 성능 개선을 체감했다는 응답자는 3명 중 1명 미만

지난 12개월간 통신 네트워크에서 전혀 변화를 느낄 수 없다고 응답한 비율이 66%를 기록했다.



참조: 가중치 반영 집계. 휴대폰 또는 스마트폰을 소유한 16~75세 영국 소비자 4,023명 대상 조사.

출처: Deloitte Digital Consumer Trends 2024.

또한 일부 소비자는 자신이 속한 시장에서 이동통신 네트워크간의 차이를 비교하는 데 어려움을 겪을 수 있기 때문에, ‘더 나은 성능의 네트워크로 갈아타라’는 메시지를 던지는 마케팅이 소비자 행동에 별다른 영향을 미치지 못할 수 있다.

대부분의 이용자는 이동 및 생활 패턴도 다르고 선호하는 애플리케이션도 제각각이기 때문에, 개인별로 네트워크 이용 환경도 상이하다. 이동통신사는 커버리지 지도를 작성하지만, 이는 각 지점 및 각 시점의 수요 강도나 혼잡 상황을 반영하지 못한다.¹¹ 이용자가 두 개의 SIM을 동시에 사용하면 두 개 네트워크를 직접 비교할 수도 있지만, 이처럼 번거로운 일을 기꺼이 해볼 소비자는 거의 없을 것이다.

거의 정점에 도달한 소비자 통신 연결성

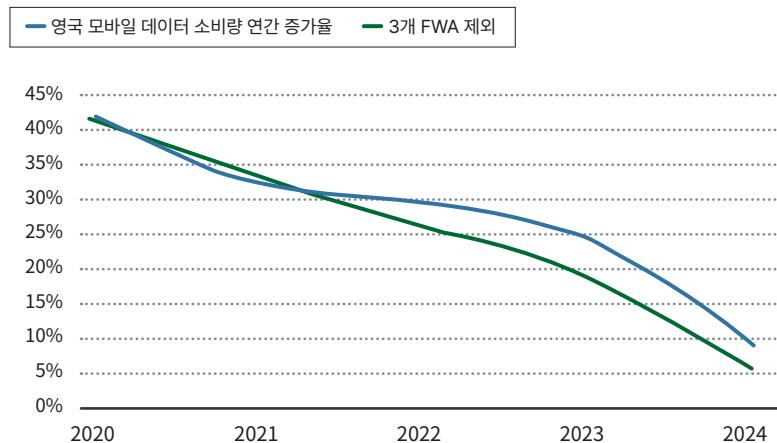
통신 네트워크 연결에 대한 소비자 니즈를 총족하는데 40년 넘게 걸렸지만 이제 시장은 거의 포식 상태에 도달한 것으로 보인다.¹²

단언은 할 수 없지만, 2030년까지 약 5년간 중기적 전망으로는 모바일 네트워크 성능의 급격한 개선이 필요할 정도로 혁신적인 새로운 기기가 등장할 가능성은 적다. 마찬가지로, 대중이 대거 메타버스로 이동하는 등 급변 사태의 가능성도 있지만 현실화될 확률은 낮다. 마지막으로, 주요 애플리케이션별 네트워크 연결 수요는 정체하거나 오히려 감소할 가능성이 크다.¹³

이러한 사용 패턴의 안정화 및 예측 가능성은 데이터 사용량 추세에서도 확인된다. 지난 5년 동안 주요 시장에서는 SIM당 기가바이트(GB) 사용량 증가 속도가 둔화됐다. 10개 선진국 시장에서는 2024년 기준 증가율이 한 자릿수로 떨어졌다.¹⁴ 여전히 두 자릿수 성장률을 보이는 시장의 경우에도 주로 전용 고정무선 접속(FWA) 단말기 또는 스마트폰 테더링을 통한 가정용 광대역 대체 수요가 소폭 증가한 영향이다. 예컨대 영국에서는 모바일 데이터 사용량의 전년 대비 증가율이 2024년 10%까지 낮아졌으며, 전용 FWA 영향을 제외하면 증가율은 5%까지 떨어졌다(그림 3).

그림 3
영국 광대역 소비량 증가율 소폭 하락

(단위: 기가바이트 / 월간 집계)



출처: Deloitte analysis based on Enders Analysis.

6G 시대의 차별화 과제

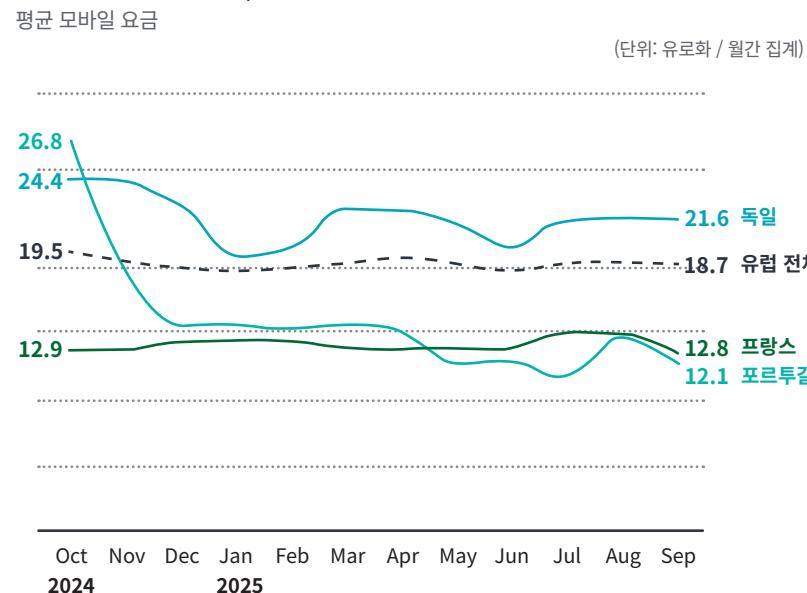
소비자가 5G의 효익을 제대로 체감하지 못하고 있다면, 6G의 가치를 마케팅하는 일은 더욱 어려워질 것이다. 이동통신이 새로운 세대로 업그레이드 될 때마다 기대되는 성능 개선폭은 최소 10배다.¹⁵ 성능이 개선됐음을 체감할 수 있는 가장 대표적인 지표는 용량으로, 대형 음악 페스티벌이나 연말 쇼핑 성수기처럼 특정 시간과 장소에서 폭증하는 트래픽을 처리하기 위해서는 대용량이 필요하다. 또한 4G에서 5G로 업그레이드 될 때와 마찬가지로, 기가바이트당 전송 비용을 낮추는 효과도 기대할 수 있다.¹⁶ 물론 속도 역시 개선 지표에 포함된다. 6G 사양은 2026년 확정될 예정이지만, 이미 초당 100기가비트(Gbps)급 속도를 시연한 테스트 결과도 나왔다.¹⁷ 이는 5G의 최고치인 약 5Gbps의 20배 수준이다.

하지만 이처럼 6G 업그레이드로 최고 속도가 높아지더라도, 수요는 정체된 상태를 유지할 가능성이 크다. 스마트폰에서 재생되는 일반 HD급 동영상 스트리밍은 최대 초당 5메가비트(Mbps)면 충분하며, 향후 압축 기술 고도화 등으로 이 수치는 더 낮아질 여지도 있다. 수요가 정체된 상황에서 네트워크 전면 업그레이드를 통한 투자수익률을 확보하기란 어려울 수 있으며, 업그레이드의 1차적 목적이 운영비 절감이 아닌 이상 투자 명분을 확보하기가 쉽지 않다.

통신사들이 리워드 프로그램에 대한 관심을 높여야 하는 또 다른 이유는 동일 데이터 제공량(예: 월 10GB) 등으로 비교되는 요금 중심 비교

를 줄이기 위해서다. 2024~2025년 일부 시장에서는 모바일 요금이 최대 50% 하락했다(그림 4).¹⁸ 이러한 환경에서 만약 요금제 구성에 무료 커피나 피자 쿠폰과 같은 리워드를 추가한다면, 소비자가 단순히 ‘같은 GB당 가격’을 기준으로 가치를 비교하기가 훨씬 어려워진다. 그만큼 통신사가 비가격(non-price) 요소에서 차별화할 여지가 커지는 셈이다.

그림 4
일부 지역의 모바일 요금, 전년비 하락세



출처: New Street Research, 2025.

통신사의 새로운 성장 경로로 떠오른 리워드 프로그램

통신사들은 앞으로 소비자가 네트워크를 선택하는 기준이 과거와 어떻게 달라질지, 그리고 그 변화로 인해 전략 재조정이 필요한지를 점검해야 한다. 향후 10년간 통신 업계가 처한 환경이 급격한 변화를 보일 수 있기 때문에 특별히 신중한 자본 배분이 필요하다. 현재 투자자본수익률(ROIC)은 7.3% 수준이지만, 가중평균자본비용(WACC)은 6.9%로,¹⁹ 경제적 관점에서 보면 거의 손익분기점 수준의 투자 환경이다. 따라서 고정·모바일 모두 데이터 사용량 증가세가 둔화되는 흐름은 오히려 통신사에 유리하게 작용할 수 있다. 즉, 막대한 네트워크 업그레이드 비용을 들이지 않고, 더 높은 자본수익률을 기대할 수 있는 가치 제안을 선택할 여지가 생긴다는 의미다.

다른 산업의 사례에서도 핵심 서비스가 성숙기에 접어들면 리워드 프로그램이 차별화 도구로 부상해 왔다. 한때 초음속 서비스 같은 기술 발전의 속도가 핵심 가치로 여겨졌던 항공 산업도 이제는 리워드 프로그램 중심으로 전략의 축을 전환했다. 글로벌 항공사들의 로열티 프로그램 가치는 1,000억 달러를 넘고, 이 중 상위 3개 항공사의 리워드 프로그램은 각각 200억 달러를 넘는다.²⁰ 미국에서는 2019년 이후 전체 신용카드 지출의 90% 이상이 보상형 카드에서 발생하고 있다.²¹

통신사가 비(非)네트워크 혜택에 투자한다면, 이를 전략적으로 세심하게 커뮤니케이션해야 한다. 이미 일부 사업자는 TV·옥외광고·라디오·소

설·인쇄물 등 대규모 마케팅을 통해 리워드 혜택을 적극 홍보 중이다. 티모바일(T-Mobile)은 ‘땡큐’(Thank You) 프로그램을 통해 음식·영화·연료·여행 등 10억 개의 기프트가 고객에게 전달됐다고 발표했다.²² 보다폰(Vodafone)은 ‘베리미’(VeryMe) 프로그램을 통해 1억7,500만 건의 리워드 지급 실적을 공개했다.²³ O2는 ‘프라이어리티’(Priority) 프로그램으로 고객들이 연간 2,300만 파운드를 절감했다고 밝혔다.²⁴

또한 Z세대 및 밀레니얼 세대 고객층은 네트워크 성능보다 혜택 제공에 더 민감하게 반응할 가능성이 크다. 20대 중반 소비자는 2000년대 최신 기술인 3G의 느린 속도를 경험해본 적이 거의 없고, 주로 4G부터 사용했기 때문에 5G와의 차이를 거의 체감하기 어렵다. 40대 소비자라면 1990년대 최신 기술인 2G 기반 웹을 사용해 본 경험이 없을 수 있다. 그 결과, 세대별로 중요하게 여기는 차별화 포인트가 다르게 나타난다.

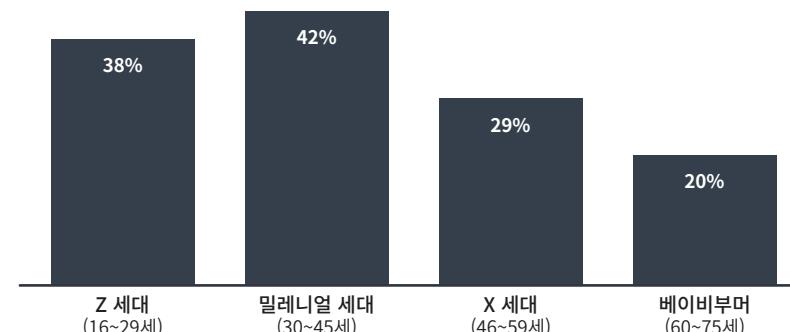
딜로이트 영국의 조사에서도 Z세대와 밀레니얼 세대는 보상 혜택을 이유로 네트워크를 변경할 가능성이 중장년층보다 훨씬 높은 것으로 확인됐다(그림 5).

그림 5

Z세대와 밀레니얼 세대는 보상 혜택을 이유로 네트워크를 변경할 가능성이 중장년층보다 훨씬 높다

로열티 혜택을 이유로 모바일 네트워크를 바꿀 의향이 있는 영국 소비자의 세대별 비율

Q. 로열티 혜택 때문에 모바일 통신사를 바꿀 의향이 있습니까?



참조: 가중치 반영 집계. 휴대폰 또는 스마트폰을 소유한 16~75세 영국 소비자 4,023명 대상 조사.

출처: Deloitte Digital Consumer Trends, 2025.

또한 통신 사업자들은 리워드 혜택이 고지출 가입자에게 더 매력적인 요인으로 작용할 수 있다는 점을 유의해야 한다. 소득이 높은 가입자일수록, 소득이 낮은 가입자보다 더 나은 혜택을 제공하는 통신사로 이동할 가능성이 크다(그림 6). ‘무료 제공’(freebie)이라는 제안은 때로 비합리적일 정도로 긍정적인 감정적 반응을 불러일으키는 장치가 되기 때문이다.²⁵

그림 6

소득이 높은 가입자일수록 리워드 혜택을 선호

리워드 혜택 때문에 통신사를 바꿀 의향이 있다고 답한 영국 소비자의 소득 구간별 집계

Q. 로열티 혜택 때문에 모바일 통신사를 바꿀 의향이 있습니까?



참조: 가중치 반영 집계. 휴대폰 또는 스마트폰을 소유한 16~75세 영국 소비자 4,023명 대상 조사.

출처: Deloitte Digital Consumer Trends, 2025.

대형 통신사들이 유사한 전략을 일제히 추진할 경우, 리워드 혜택 자체가 네트워크 성능과 마찬가지로 상품화될 위험이 있다. 은행부터 에너지, 공공서비스까지 타 산업에서도 자체 리워드 프로그램을 확대하면서 리워드 시장이 더욱 포화될 가능성이 크다.²⁶ 소비자가 받아들일 수 있는 무료 커피의 양에는 한계가 있으며, 여러 서비스 제공자가 동일한 혜택을 제공하기 시작하면 매력도는 희석될 수밖에 없다. 따라서 고유하고 차별화된 리워드 체계를 설계하는 것이 고객 유치와 이탈 방지에 핵심이 될 것이다. 이런 맥락에서 라이브 이벤트, 콘서트, 스포츠 경기 티켓 등이 매력적인 리워드 자산이 될 수 있다.²⁷ 다만 이러한 리워드는 수천만 명의 가입자 중 수만 명에게만 제공할 수 있다는 한계도 고려해야 한다.

Korean Perspectives

국내 통신 산업의 변화와 도전,
그리고 새로운 기회

내 통신 시장은 5G 커버리지·속도 측면에서 이미 글로벌 상위 수준에 도달해, 추가적인 네트워크 성능 개선이 소비자 체감 변화로 연결되기 어려운 국면에 있다. 고객 서비스 품질 만으로 가입자 충성도와 이탈률을 설명하기 어렵고, 실제로 고객들은 요금제와 할인구조 (가족 할인, 인터넷 번들 등) 및 추가적인 서비스 경험 등 비 기술 영역으로 이동하는 추세이다.

이제 경쟁의 중심은 기술에서 고객 가치로 옮겨가고 있다. 요금제, 멤버십, 제휴 혜택, 앱과 디지털 채널을 통한 이용 경험이 통신사 선택의 핵심 기준으로 자리 잡으면서, 국내 통신 3사 모두 생활 밀착형 멤버십과 프로모션 경쟁을 강화하고 있다. SK텔레콤은 영화·커피·외식 제휴 중심의 T멤버십을, KT는 VIP·VVIP 중심의 프리미엄 멤버십을, LG유플러스는 구독 플랫폼 '유독'과 디지털 서비스를 결합한 고객 이탈 방지 및 유지 전략을 전개하고 있다.

다만 부가 가치 전략의 난이도 역시 높아지고 있다. SK텔레콤이 T 멤버십을 할인형에서 적립형으로 전환하는 과정에서 불거진 '혜택 축소' 논란은, 혜택의 규모보다 고객이 실제적 이를 어떻게 체감하느냐가 훨씬 중요하다는 점을 분명하게 보여준 사례였다.

이제 부가 가치의 개발은 단순한 마케팅 수단을 넘어, 고객의 인식과 경험을 정교하게 설계해야 하는 전략 영역이 되고 있는 것이다. 이 과정에서 보안과 신뢰는 고객 경험을 좌우하는 핵심 요소로 자리잡고 있다. 보안 이슈가 발생할 경우 즉각적인 고객 이탈로 이어진다. 실제로 최근 개인정보 유출과 해킹 사건 발생 이후 타 통신사로의 이탈 고객 증가는 서비스 경험 내 취약성 발생 시에 곧 고객 이탈로 직결될 수 있음을 보여준 사례이다.

세대와 소득에 따라 고객의 니즈 분화도 더욱 뚜렷해지고 있다. Z세대와 청년층은 OTT, 영화, 카페, 공연 등 문화·여가 중심의 ‘즉시 체감 가능한’ 혜택에 민감하게 반응하며, 멤버십과 구독 서비스가 통신사 선택의 실질적인 기준으로 자리 잡고 있다. 반면 고소득·고가 요금제 고객을 대상으로는 프리미엄 혜택과 차별화된 서비스 경험이 중요해지고 있으며, 단일 요금제·단일 혜택 구조의 한계가 드러나고 있다. 보다 정교한 세분화와 고객군별 맞춤화 전략이 요구되는 지점이다.

국내 통신 업계에서도 6G 투자를 둘러싼 수익성에 대한 고민이 커지고 있다. 5G에서도 실질적 고객 가치의 체감 효과에 대한 의구심과 회의가 있었던 만큼, 6G 역시 속도와 지연시간 개선만으로 요금 인상과 투자 회수를 설득하기는 쉽지 않을 것으로 보인다. 통신사들은 네트워크 고도화와 함께 B2B·플랫폼·클라우드·미디어 등 인접 영역을 결합한 포트폴리오 다각화를 추진하고 있다.

결론적으로 기술 투자는 차별화된 고객 경험 설계와 병행될 때 새로운 성장 기회로 이어질 수 있다. LG유플러스의 AI 기반 구독 플랫폼, SK 텔레콤의 AI·멤버십을 연계한 개인화 전략, KT의 AI 기반 고객센터 혁신 사례에서 보듯, 국내 통신 3사는 요금과 속도 중심의 경쟁에서 벗어나 디지털 플랫폼과 AI를 기반으로 한 고객 경험 중심 경쟁으로 전환해야 할 것이다.



이대의 파트너

한국 딜로이트 그룹
컨설팅 부문 Customer 본부장,
Deloitte Digital 리더

주석

Part 1

인류의 삶에 보이지 않는 형태로 스며든 AI가 창출하는 새로운 가치

1장 검색엔진 내장형 생성형 AI, 독립형 AI 추월

1. Deloitte's forecast is based on multiple sources, including its proprietary research undertaken as part of Deloitte's Digital Consumer Trends survey, fielded in April and May 2025, and also in 2023 and 2024. This longitudinal data set provides a trajectory for the adoption of standalone gen AI apps. Our proprietary data set includes surveys conducted in multiple developed markets globally. Additionally, we have considered multiple other data points, including Alphabet's reporting on the volume of AI Overviews, which had a monthly usage base of over two billion as of July 2025. See: Alphabet, "Alphabet announces second quarter 2025 results," July 23, 2025.
2. Paul Lee and Ben Stanton, "Digital Consumer Trends 2025, UK edition," Deloitte, June 2025.
3. Elizabeth Reid, "Generative AI in search: Let Google do the searching for you," Google, May 14, 2024.
4. Hema Budaraju, "New ways to connect to the web with AI Overviews," Google, Aug. 15, 2024.

5. Josh Howarth, "Worldwide Daily Social Media Usage (New 2025 Data)," Exploding Topics, June 23, 2025.
6. United States Census Bureau, "Quarterly retail e-commerce sales," press release, Aug. 19, 2025.
7. Danny Goodwin, "Google search is 373x bigger than ChatGPT search," Search Engine Land, March 11, 2025.
8. Kara Kennedy, "Poor prompts lead to misleading research," AI Literacy Institute, Aug. 19, 2024; Ulster University, "Generative artificial intelligence (Gen AI): Prompt engineering," Oct. 23, 2025; Haringun Nur Adha, "You made a specific prompt but the results are disappointing? Maybe you're using ChatGPT wrong," Medium, Sept. 16, 2025.
6. Salesforce, "Agentforce," accessed October 2025; SAP, "Joule Agents," accessed October 2025; ServiceNow, "AI Agents," accessed October 2025; Workday, "Workday unveils next generation of Illuminate Agents to transform HR and finance operations," press release, May 19, 2025.
7. Oracle, "Oracle introduces AI Agent Studio," press release, March 20, 2025; SAP, "Business AI," accessed October 2025; Adobe, "Adobe launches Adobe Experience Platform Agent Orchestrator for businesses to activate AI agents in customer experiences and marketing workflows," news release, March 18, 2025; Workday, "Workday unveils Workday Build, giving developers the tools to build the future of work," press release, September 16, 2025; Erwin Huizenga and Bo Yang, "Agent Development Kit: Making it easy to build multi-agent applications," Google for Developers, April 9, 2025.
8. Joanna Glasner, "AI autonomous agents are top 2025 trend for seed investment," Crunchbase News, June 17, 2025; Jacob Robbins and Kia Kokalitcheva, "Y Combinator is going all-in on AI agents, making up nearly 50% of latest batch," PitchBook, June 11, 2025.
9. Gartner, AI agents are disrupting SaaS pricing: What must CIOs do?, July 16, 2025 (ID G00834627). GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.
10. Adrian Radu, "Billing infrastructure in the age of co-pilots and AI agents," Lightspeed, March 6, 2025.
11. Gartner, AI agents are disrupting SaaS pricing: What must CIOs do?, July 16, 2025 (ID G00834627). GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.
12. Maxio, 2025 pricing trends: Usage-based models and the path to SaaS growth, 2025.

2장 서비스형 소프트웨어, AI 에이전트와 결합: 예산·고객경험·인력 구조 재편

1. Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach, 4th ed. (New York, NY: Pearson, 2021).
2. Tim Smith et al., "AI is capturing the digital dollar. What's left for the rest of the tech estate?", Deloitte, October 16, 2025; AI automation includes basic automation, process automation with agents, process reimagination, and organizational reimagination.
3. Eric Newmark, "The agentic evolution of enterprise applications," IDC, April 4, 2025.
4. Dan Gallagher, "Software's death by AI has been greatly exaggerated," Wall Street Journal, August 27, 2025.
5. Zyro, "111 unmissable SaaS statistics for 2025," accessed October 2025; Matt Ashare, "AI drives up compute costs as cloud inflation slows," CIO Dive, February 18, 2025.

13. Zendesk, "Zendesk first in CX industry to offer outcome-based pricing for AI agents," August 28, 2024.
 14. Salesforce, "Agentforce Observability," accessed October 2025; Google Cloud, Gemini Enterprise, accessed October 2025.
 15. Bojan Ceric and Prakul Sharma, "Scaling AI agents may be risky without an enterprise marketplace," Deloitte Insights, September 15, 2025.
- survey asked respondents about four types of AI automation and their incremental actions across each: mature or very mature respondents for basic automation (n = 443) and basic automation and AI agents (n = 153); and those with up to three-year expectations for basic AI automation (n = 245) and basic automation and AI agents (n = 68).

3장 기하급수적 가치 창출을 위한 AI 에이전트 오케스트레이션

1. Deloitte, "The cognitive leap: How to reimagine work with AI agents," December 2024.
2. The baseline projection is derived from a Deloitte analysis of global autonomous AI agent market projections as per seven publicly available and third-party research reports. The estimated increase of 15% to 30% in the projected market is modeled on future scenarios where fewer agentic AI projects are cancelled owing to improved enterprise readiness.
3. Gartner, "Gartner predicts over 40% of agentic AI projects will be canceled by end of 2027," press release, June 25, 2025.
4. Bojan Ceric and Prakul Sharma, "Generative AI meets the virtual world: A model for human-AI collaboration," Deloitte Insights, Feb. 10, 2025.
5. Abdi Goodarzi and Nitin Mittal, "A new digitally-enabled workforce era: How AI agents can help deliver functional efficiency and value across the enterprise," Forbes, Aug. 18, 2025.
6. Tim Smith, Gregory Dost, Garima Dhasmana, Parth Patwari, Diana Kearns-Manolatos, and Iram Parveen, "Digital budgets are rising, but investment strategies may need a recalibration," Deloitte Insights, Oct. 16, 2025. The
7. Prakul Sharma, Val Srinivas, and Abhinav Chauhan, "How banks can supercharge intelligent automation with agentic AI," Deloitte Insights, Aug. 14, 2025; Kausik Chaudhuri, "Applying agentic AI to legacy systems? Prepare for these 4 challenges," CIO, July 16, 2025; SaaS meets AI agents: Transforming budgets, customer experience, and workforce dynamics; Bojan Ceric and Prakul Sharma, "Scaling AI agents may be risky without an enterprise marketplace," Deloitte Insights, Sept. 15, 2025.
8. Julian Horsey, "AI investment research agent "Ask David" built by JP Morgan," Geeky Gadgets, May 30, 2025; Irene Iglesias Álvarez, "The agentic AI assist Stanford University cancer care staff needed," CIO, May 30, 2025.
9. Isabelle Bousquette, "Why Walmart is overhauling its approach to AI agents," The Wall Street Journal, July 24, 2025.
10. Henry Peng Zou et. al, "A call for collaborative intelligence: Why human-agent systems should precede AI autonomy," arxiv, June 11, 2025.
11. Jesus Olivera, "Ensuring accuracy in AI with human-in-the-loop," Medium, Sept. 27, 2024.
12. John Werner, "They're making TCP/IP for AI, and it's called NANDA," Forbes, May 01, 2025.
13. Emilia David, "Google's Agent2Agent interoperability protocol aims to standardize agentic communication," VentureBeat, April 9, 2025.
14. Emilia David, "Google's new agent Payments Protocol (AP2) allows AI agents to complete purchases — is your enterprise ready?" VentureBeat, Sept. 16, 2025.

- 15. Leslie Joseph and Rowan Curran, "Interoperability is key to unlocking agentic AI's future," Forrester, March 25, 2025.
- 16. Alfred Shen and Anya Derbakova, "Design multi-agent orchestration with reasoning using Amazon Bedrock and open source frameworks," Amazon Web Services, Dec. 19, 2024; IBM, "Multiagent orchestration, accessed Oct. 7, 2025.
- 17. Amazon Web Services, "Observe your agent applications on Amazon Bedrock AgentCore Observability," accessed Oct. 13, 2025.
- 18. Gartner, "Gartner predicts that guardian agents will capture 10-15% of the agentic AI market by 2030," press release, June 11, 2025.
- 19. The Future Society, "How AI agents are governed under the EU AI Act," June 4, 2025.
- 20. CEN-CENELEC, "Artificial intelligence," accessed Oct. 7, 2025.
- 21. Daniel Sun, "Capitalize on the AI agent opportunity," Gartner, Feb. 27, 2025.
- 22. GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.
- 23. Salesforce, "HR leaders to redeploy a quarter of their workforce as agentic AI adoption expected to grow 327% by 2027," May 5, 2025.
- 24. Ibid; Atikah Amalia, "The marketer's new job title: AI boss," Content Grip, April 29, 2025.
- 25. Kyle Forrest, Brad Kreit, Abha Kulkarni, Roxana Corduneanu, and Sue Cantrell, "AI, demographic shifts, and agility: Preparing for the next workforce evolution," Deloitte Insights, Aug. 25, 2025.
- 26. Michael Caplan et al., "The technology operating model of the future: Rise of the agentic enterprise," The Wall Street Journal, Aug. 23, 2025.
- 27. Ritu Jyoti, "The rise of the agentic economy: How autonomous AI is reshaping the future of work," CIO, Sept. 8, 2025.
- 28. Isabelle Bousquette, "Digital workers have arrived in banking," The Wall Street Journal, June 30, 2025.
- 29. Marina Temkin, "Why AI agent startup /dev/agents commanded a massive \$56M seed round at a \$500M valuation," TechCrunch, Nov. 28, 2024; Hui Wong, "Questflow secures \$6.5M seed round to build AI agent economy for every workflow," Marketers Media, July 24, 2025.

4장 휴머노이드·산업용 로봇·드론의 AI 혁신: 자율성과 생산성의 새로운 임계점

- 1. Deloitte analysis and estimates based on data from publicly available information sourced from the International Federation of Robotics, Interact Analysis, and Automation.com.
- 2. Duncan Stewart et al, "Robots on the move: Professional service robots set for double-digit growth," TMT Predictions 2020, November 2019. To read further, see "Professional services robots on the move," The Wall Street Journal-CIO Journal, April 8, 2020.
- 3. Morgan Stanley research, "Humanoids: A \$5 trillion market," May 14, 2025.
- 4. Methodology and assumptions: From 500,000 annual installations each year, in 2025 and 2026, we anticipate annual industrial robot installations could grow by 100,000 units every year between 2027 and 2030, reaching 1 million installed units in 2030. These calculations are based on insights gathered from IFR press release dated September 24, 2024 ("Record of 4 million robots in factories worldwide"). From our conversations with industry experts, we believe growth and availability of computing power, especially new types of AI models (LLMs, but also VLAs and world models),

plus the active role that some major tech and robotics companies are playing to invest and bring forth robotics chips and solutions to market, will help drive robotics adoption during 2026 to 2030 and beyond. Additionally, average unit price per industrial robot has declined by approximately 3.2% between 2018 and 2024. We expect average price to continue to decline in that range through 2030, given the broader availability of chips, sensors, and other components, including open model-based robots. Between 2025 to 2030, we have assumed average annual price per industrial robot could decline approximately 3.1 to 3.2 percent based on information gathered from IPF Online's article dated June 27, 2025 ("Global industrial robot shipments down in 2024, recovery likely in 2025").

5. OECD, OECD Employment Outlook 2025: Can we get through the demographic crunch?, July 9, 2025.
6. Deloitte analysis of the various foundational models released by technology companies and niche LLM players during 2024 and 2025.
7. Standard bots, "The most advanced robots in 2025," August 7, 2025.
8. Cem Dilmegani, "Data quality in AI: Challenges, importance, & best practices," AIMultiple research, July 9, 2025.
9. Ainsley Lawrence, "AI's impact on robots in manufacturing," September 11, 2024.
10. Brian Heater, "Figure AI details plan to improve humanoid robot safety in the workplace," TechCrunch, January 28, 2025.
11. Tammy Whitehouse, "AI robots in the workplace: Preparing for humanoid colleagues," Deloitte-WSJ CIO Journal, July 26, 2025.
12. The Smart Factory by Deloitte website, "Home page," accessed Oct 29, 2025.
13. Deloitte analysis based on insights gathered from interviews and conversations with industry subject matter experts, and supplemented with information gathered from multiple publicly available sources including: American Machinist, Admedica, World Economic Forum, PHS Innovate, and ASDNews.
14. Reyk Knuhtsen, et al, "Robotics levels of autonomy," SemiAnalysis, July 30, 2025.
15. Sudhir Pratap Yadav, "Vision-Language-Action (VLA) models: LLMs for robots," Black Coffee Robotics, April 17, 2025; Raman Thakur, "How Vision-Language-Action models powering humanoid robots," Labellerr, March 5, 2025.
16. Deloitte analysis based on information gathered about multiple VLA models that are commercially available in the market.
17. Andrew Liszewski, "NVIDIA says 'the age of generalist robotics is here,'" The Verge, March 19, 2025.
18. Automation World, "Boston Dynamics working with NVIDIA on next-gen humanoid robots," May 21, 2025.
19. Brian Heater, "Figure's humanoid robot takes voice orders to help around the house," TechCrunch, February 20, 2025; Wei Sun, "Figure AI Unveils its 2nd-Gen Robot, Extending Focus from Factory to Home After OpenAI Split," Counterpoint Research, August 14, 2025.
20. Rebecca Szkutak, "Hugging Face unveils two new humanoid robots," TechCrunch, May 29, 2025.
21. Michael Nunez, "Hugging Face just launched a \$299 robot that could disrupt the entire robotics industry," VentureBeat, July 9, 2025. The company launched a sub US\$ 300 robot, which can integrate with the Hugging Face Hub, enabling its developer community to access pre-built AI models, hardware designs, and software and assembly instructions.

- 22. Kate Park, "RLWRLD raises \$14.8M to build a foundational model for robotics," TechCrunch, April 14, 2025.
- 23. The Robot Report, "RBR50 Spotlight: FANUC produces one-millionth industrial robot," August 12, 2024.
- 24. domainB, "China's AI-powered humanoid robots set sights on transforming global manufacturing," May 13, 2025.
- 25. Based on publicly available secondary sources that reference Unitree and UBTECH.
- 26. Based on multiple publicly available data and research reports that highlight the various chip components and hardware that are used to build humanoid robots.
- 27. Deloitte analysis based on data and information gathered from select major AI humanoid robot makers in the US and China.
- 28. Deloitte analysis based on data and information gathered from select major AI humanoid robot makers in the US and China.
- 29. Note to calculations: Using the 2026 estimated price range of US\$14,000 to US\$18,000 per unit, and 15,000-unit shipments, we multiplied the two variables to arrive at US\$210 to US\$270 million as overall revenue opportunity.
- 30. Using the variables and methodology noted in end note No. 26, we took the baseline scenario range of US\$210-270 million for 2026 and multiplied it by 3X and 4X to arrive at the other two probable 2032 market revenue potential presented in this paragraph. Our underlying assumptions for these relatively optimistic scenarios are mainly based on how fast the broader AI, robotics and tech industry might be able to address and workaround data, integration, safety, and cyber related challenges, and as price points become relatively attractive over time.
- 31. Justin Spike, "Data on animal movements help Hungarian researchers create a swarm of autonomous drones," AP News, December 19, 2024.
- 32. Deloitte analysis based on information gathered from publicly available sources about AI-enabled drones.
- 33. Karthik Ramachandran, Gillian Crossan, Duncan Stewart, and Ariane Bucaille, "On solid ground: AgTech is driving sustainable farming and is expected to harvest US\$18 billion in 2024 revenues," TMT Predictions 2024, November 29, 2023.
- 34. Damon Johnson, "From Sci-Fi to reality: The latest in drone technology for 2024," Raising Drones, July 12, 2025.
- 35. Yahoo! Finance, "Britain to allow drones to inspect power lines, wind turbines," October 15, 2024; Joe Macy, "Autonomous UAS inspection system for power lines introduced," Unmanned Systems Technology, March 14, 2025.
- 36. Damon Johnson, "From Sci-Fi to reality: The latest in drone technology for 2024," Raising Drones, July 12, 2025.
- 37. Aja Melville, "Drone Wars: Developments in Drone Swarm Technology," Forecast International, January 21, 2025.
- 38. Deloitte analysis based on publicly available price information of select major industrial robots in the market.
- 39. Gregory Haley, "Increasing roles for robotics in fabs," Semiconductor Engineering, Aug. 19, 2024.
- 40. Rebecca Szkutak, "We are entering a golden age of robotics startups — and not just because of AI," TechCrunch, September 12, 2025.
- 41. Deloitte China, "Open Full-stack Intelligent Service Robot Ecosystem white paper," April 24, 2025.

5장 생성형 AI 다음 단계 성패는 연산력이 좌우

1. Rodrigo Liang, "Scaling AI without breaking the grid: The path to sustainable innovation," World Economic Forum, Jan. 3, 2025.
2. Michelle Weaver, "Big debates: The AI evolution," Morgan Stanley, Jan. 10, 2025.
3. Josh You and David Owen, "How much power will frontier AI training demand in 2030?" Epoch.AI, Aug. 11, 2025.
4. Jared Kaplan et al., "Scaling laws for neural language models," OpenAI, Jan. 23, 2020.
5. Amazon Web Services, "What are foundation models?" accessed Sept. 19, 2025.
6. Ashu Garg, "Has AI scaling hit a limit?" Foundation Capital, Nov. 27, 2024.
7. Aixin Liu et al., "Deepseek-v3 technical report," arXiv preprint arXiv:2412.19437 (2024).
8. Chris Arkenberg, Duncan Stewart, Gillian Crossan & Kevin Westcott, "On-device generative AI could make smartphones more exciting—if they can deliver on the promise," Deloitte Insights, Nov. 19, 2024; IDC Media Center, "Worldwide smartphone market forecast to grow 1% in 2025, driven by accelerated 3.9% iOS growth, according to IDC," Aug. 27, 2025; Gartner, Inc., "Gartner says artificial intelligence ("AI") PCs will represent 31 percent of worldwide PC market by the end of 2025," press release, Aug. 28, 2025.
9. Amazon Web Services, "AWS Inferentia," accessed Sept. 19, 2025.
10. Paul Lee and Clare Mortimer, "How citizens use devices and AI: what government needs to know," Deloitte UK, Aug. 29, 2025; Steve Feinberg, et al., "In the gen AI economy, consumers want innovation they can trust: Deloitte's 2025 Connected Consumer Survey," Deloitte, Sept. 25, 2025.

11. Wylie Wong, "Data center chips in 2024: Top trends and releases," Data Center Knowledge, April 11, 2024; Reen Singh, "AI inference chips latest rankings: Who leads the race?" Uvation, July 11, 2025; Broadcom Inc., "3.5D XDSiP AI Accelerator Platform," accessed Oct. 23, 2025.
12. Deloitte Consulting LLP performed an analysis of the data center market, including a rough bill of materials for the various components, and market sizes. This analysis is due to be published in December 2025.
13. Skye Jacobs, "NVIDIA Blackwell server cabinets could cost somewhere around \$2 to \$3 million each," TechSpot, July 28, 2024.
14. Beth McKenna, "2 key things from AMD's earnings call that investors should know," The Motley Fool, Feb. 1, 2024; Dell'Oro Group, "AI infrastructure spending sustains strong growth momentum," press release, Feb. 5, 2025.
15. Kari Briski, "How scaling laws drive smarter, more powerful AI," NVIDIA, Feb. 12, 2025.
16. Ibid.
17. Jonathan Farrington, "What is chain of thought prompting – AI prompt engineering," Silicon Dales, July 24, 2025.
18. Briski, "How scaling laws drive smarter, more powerful AI."
19. Ibid.
20. "Data centers: Computing risks and opportunities for U.S. real estate," S&P Global, Oct. 22, 2024; Equinix, Inc., "Form 10-K: Annual report for fiscal year ended Dec. 31, 2023," Feb. 16, 2024; Digital Realty Trust, Inc. and Digital Realty Trust, L.P., "Form 10-K: Annual report for fiscal year ended Dec. 31, 2023," Feb. 23, 2024.
21. Shubham Sharma, "Going beyond GPUs: The evolving landscape of AI chips and accelerators," VentureBeat, Sept. 26, 2024.

22. Deloitte Consulting LLP performed an analysis of the data center market, including a rough bill of materials for the various components, and market sizes. This analysis is due to be published in December 2025.
23. Eric Reed, "History of NVIDIA: Company and stock," SmartAsset, May 22, 2024.
24. Rajat Raina, Anand Madhavan, and Andrew Y. Ng, "Large-scale deep unsupervised learning using graphics processors," In Proceedings of the 26th Annual International Conference on Machine Learning, 2009.
25. NVIDIA, "NVIDIA delivers massive performance leap for deep learning, HPC applications with NVIDIA Tesla P100 accelerators," press release, April 5, 2016.
26. Hannah Wilson, "NVIDIA facts and statistics (2025)," Investing.com, Aug. 28, 2025.
27. Hannah Wilson, "NVIDIA facts and statistics (2025)," Investing.com, Aug. 28, 2025.
28. Top 500, "June 2025," June 2025.
29. Ray Wang, "NVIDIA's new Rubin CPX targets future of large-scale inference," Futurum, Sept. 18, 2025.
30. In 2025, Deloitte Consulting LLP performed an analysis of the data center market, including a rough bill of materials for the various components, and market sizes. This analysis is due to be published in December 2025.
31. Omdia, "New Omdia forecast: AI data center chip market to hit \$286bn, growth likely peaking as custom ASICs gain ground," Aug. 28, 2025.
32. Anthony Di Pizio, "Jensen Huang predicts annual data center spending will hit \$1 trillion by 2028. Here's the ultimate semiconductor ETF to buy right now." The Motley Fool, May 1, 2025.
33. Dave Lawler, "Exclusive: 'Massive ten-year' AI boom is just starting, AMD CEO says," Axios, Sept. 17, 2025.
34. Paul Mah, "AI training is going to multiple data centers," CDO Trends, Sept. 11, 2024.
35. Chris Thomas, Akash Tayal, Duncan Stewart, Diana Kearns-Manolatos, and Iram Parveen, "Is your organization's infrastructure ready for the new hybrid cloud?" Deloitte Insights, June 30, 2025.
36. Mike Robuck, "Google strikes deals for flexible AI data centre power use," Mobile World Live, Aug. 5, 2025.
37. Tyler H. Norris, Tim Profeta, Dalia Patino-Echeverri, and Adam Cowie-Haskell, "Rethinking load growth: Assessing the potential for integration of large flexible loads in US power systems," Nicholas Institute for Energy, Environment & Sustainability, Duke University, February 2025.
38. Anton Shilov, "TSMC could charge up to \$45,000 for 1.6nm wafers — rumors allege a 50% increase in pricing over prior-gen wafers," Tom's Hardware, June 4, 2025.
39. Francisco Jeronimo, "The rise of gen AI smartphones," IDC, July 5, 2024.
40. Dan Shipper, "Vibe check: OpenAI drops two new open-weight models," Every Media, Aug. 5, 2025.
41. Cyfuture Cloud, "NVIDIA DGX H100 price 2025: Cost, specs, and market insights," Cyfuture Cloud Knowledgebase, accessed October 2025.
42. Tae Kim, "NVIDIA's multi-million dollar AI servers are getting more expensive," Barron's, Aug. 28, 2025.
43. Skye Jacobs, "NVIDIA Blackwell server cabinets could cost somewhere around \$2 to \$3 million each," TechSpot, July 28, 2024.
44. Qualcomm, "Flight RB5 5G platform," accessed Sept. 19, 2025.

45. Ali Kani, "NVIDIA DRIVE Thor strikes AI performance balance, uniting AV and cockpit on a single computer," NVIDIA, Sept. 20, 2022.
46. There are a variety of suppliers for chips for driving assistance, but as one example, NVIDIA's auto segment is at a US\$2 billion run rate as of August 2025: Pras Subramanian, "NVIDIA's auto business surges 69% from self-driving tech," Yahoo Finance, Aug. 25, 2025.
47. Karthik Ramachandran, et al, "AI for industrial robotics, humanoid robots, and drones," Deloitte Insights.
48. Jameel Rogers, "AI chips for data center and cloud to exceed US\$400 billion by 2030," IDTechEx, May 8, 2025.
4. Ibid. Importantly, an AI server rack is not just a monolithic unit but a far more complex, integrated system that comprises tens of thousands of components ranging from advanced chips, memory dies, analog integrated circuits, controllers, power devices, and passives like substrates and capacitors.
5. Karen Freifeld and Surbhi Misra, "As trade war truce with China holds, US lifts curbs for chip design software and ethane," Reuters, July 3, 2025; Joe Cash, "China says successful US trade talks make return to tariff war unnecessary," Reuters, July 18, 2025.
6. Bureau of Industry and Security and US Department of Commerce, "Federal Register, vol. 89, no. 173," Sept. 6, 2024.
7. New software and technology controls included restrictions on electronic computer-aided design and technology computer-aided design software and technology, especially when these are used for designing advanced node-integrated circuits. To read further, see: Bureau of Industry and Security and US Department of Commerce, "Commerce strengthens export controls to restrict China's capability to produce advanced semiconductors for military applications," Dec. 2, 2024.
8. Bureau of Industry and Security and US Department of Commerce, "Framework for artificial intelligence diffusion," Federal Register, Jan. 15, 2025.
9. Wenji Fang, Jing Wang, Yao Lu, Shang Liu, Yuchao Wu, Yuzhe Ma, and Zhiyao Xie, "A survey of circuit foundation model: Foundation AI models for VLSI circuit design and EDA," arXiv, March 28, 2025.
10. For further information on AI model weights related technology controls, see: US Department of Commerce and Bureau of Industry and Security, "Federal Register, vol. 90, no. 9," Jan. 15, 2025.
11. Insights based on conversations and interviews with Deloitte experts in the areas of the semiconductor industry, supply chains, and export control impact.

Part 2

기술 공급망과 주권의 재편, 새로운 경쟁 전선 형성

6장 AI 칩을 둘러싼 무역 통제와 글로벌 반도체 공급망 재편

1. A note to methodology. Estimates include projected aggregate spending for 2026 on extreme ultraviolet equipment, AI-based etch equipment, select advanced packaging equipment including high-bandwidth memory co-packaging tools, and AI chip design software and tools.
2. In 2025, Deloitte Consulting LLP performed an analysis of the data center market, including a rough bill of materials for the various components and market sizes. This analysis is due to be published in December 2025.
3. Ibid.

12. Chris Miller, "How US export controls have (and haven't) curbed Chinese AI," *AI Frontiers*, July 8, 2025.
13. Stefano Lovati, "China invests €37 billion to develop domestic EUV lithography systems," *Power Electronics News*, Feb. 11, 2025.
14. Pablo Valerio, "China semiconductor ambition and adversity," *EE Times*, May 19, 2025. Additionally, US regulations included restricting and capping the production of advanced AI chips far below the domestic demand in China.
15. See Bureau of Industry and Security and US Department of Commerce, "Federal Register, vol. 89, no. 173," p. 7. As noted in this document, atomic layer etching helps produce vertical edges required in high-quality, leading-edge advanced devices and structures, including gate-all-around field-effect transistor and similar 3D structures. Anisotropic dry etching is critical for gate-all-around field-effect transistor and similar 3D structure fabrication. It is also an important tool for fin-shaped field effect transistor (FinFET) fabrication.
16. Ibid.
17. US Department of Commerce and Bureau of Industry and Security, "Foreign-produced direct product rule additions, and refinements to controls for advanced computing and semiconductor manufacturing items," Dec. 5, 2024.
18. Sara Bulter, "How China's rare earth metals export ban will impact supply chains in 2025," *Optilogic*, Feb. 17, 2025.
19. Deloitte analysis based on conversations and insights gathered from industry experts and cross-validated with multiple secondary sources, including: Abbie Windsdale, "Netherlands takes bold step to tighten semiconductor export control," *Tech Announcer*, Jan. 16, 2025.
20. For example, hybrid bonding is fundamental to developing advanced 2.5D and 3D chip designs and heterogeneous architectures (or chiplets), as it enables ultra-fast data transfers (up to 17 TB/s) that are critical for AI and high-performance computing. To read further, see: Sam Naffziger, "Future of AI hardware enabled by advanced packaging," *IEEE Electronics Packaging Society*, May 28, 2024.
21. Duncan Stewart, Karthik Ramachandran, Prashant Raman, and Ariane Bucaille, "Silicon building blocks: Chiplets could move Moore's Law forward," *Deloitte Insights*, Nov. 19, 2024.
22. Bureau of Industry and Security, "Commerce strengthens export controls to restrict China's capability to produce advanced semiconductors for military applications," press release, Dec. 2, 2024.
23. Stewart, Ramachandran, Raman, and Bucaille, "Silicon building blocks."
24. Xiaoxi He and Yu-Han Chang, "Chiplet technology 2025-2035: Technology, opportunities, applications," *IDTechEx*, accessed Oct. 1, 2025.
25. US Department of Commerce and Bureau of Industry and Security, "Foreign-produced direct product rule additions, and refinements to controls for advanced computing and semiconductor manufacturing items."
26. Duncan Stewart, et al, "Why AI's next phase will likely demand more computational power, not less," *Deloitte Insights*.
27. For context, state-of-the-art chip fabs in the United States and Taiwan were already pushing the boundaries toward sub 7 and sub 5 nm as of 2020 to 2021, indicating China is probably at least four to five years behind (see *Deloitte 2024 semiconductor outlook*). Therefore, initiatives such as Beijing's Big Fund III actively support the expansion of local semiconductor capabilities, notably electronic design automation (EDA) and lithography tech development. To read further, see: Anton Shilov,

"China to pivot \$50 billion chip fund to fighting U.S. squeeze as trade war escalates — country to back local companies and projects to overcome export controls," Tom's Hardware, June 27, 2025.

28. The Chinese Academy of Sciences worked with domestic chip design players on an open-source project to develop an AI system that used large language models to accelerate chip design and build fully functional central processing units. To read further, see: Mark Tyson, "China claims to have developed the world's first AI-designed processor — LLM turned performance requests into CPU architecture," Tom's Hardware, June 12, 2025. Additionally, Huawei's breakthroughs in developing EDA tools capable of supporting 14 nm processes and above mark significant milestones. To read further, see: Omar Sohail, "Huawei has reportedly developed 14nm EDA tools, which the company will employ to mass manufacture its Kirin 9020, but the company is still limited to the 7nm architecture," WCCF TECH, June 11, 2025.
29. David Jarvis, et al, "A new era of self-reliance: Navigating technology sovereignty," Deloitte Insights.
30. Analysis based on multiple publicly available secondary sources that discuss the chip industry's plans to commence new AT hubs in countries including India, Malaysia, and Vietnam.
31. Aside from trade-related issues, as we already mentioned in our 2024 Global Semiconductor Outlook report, cyber threats are surging, requiring chip fabs and AI systems to intensify security measures against malware targeting critical infrastructure.
32. Duncan Stewart, et al, "Why AI's next phase will likely demand more computational power, not less," Deloitte Insights. Deloitte analysis based on conversations and insights gathered from industry experts.
33. Gwangoo Yeo, Jiin Kim, Yujeong Choi, and Minsoo Rhu, "PREBA: A hardware/software co-design for multi-instance GPU based AI inference servers," arXiv, Nov. 28, 2024.

7장 기술 주권 확보 경쟁

1. Gartner, "Gartner reveals top technologies shaping government AI adoption," press release, Sept. 9, 2025; Gartner is a registered trademark and service mark of Gartner Inc. and its affiliates in the United States and internationally and is used herein with permission. All rights reserved.
2. Sean Fleming, "What is digital sovereignty and how are countries approaching it?" World Economic Forum, Jan. 10, 2025.
3. Zoe Hawkins, Vili Lehdonvirta, and Boxi Wu, "AI compute sovereignty: Infrastructure control across territories, cloud providers, and accelerators," SSRN, June 24, 2025.
4. Adam Satariano and Paul Mozur, "AI computing power is splitting the world into haves and have-nots," The New York Times, June 21, 2025.
5. Yahoo Finance, "NVIDIA Corporation (NVDA) Q2 FY2026 earnings call transcript," Aug. 27, 2025.
6. Mario Draghi, "The Draghi report on EU competitiveness," European Commission, Sept. 9, 2024.
7. EuroStack, "Building Europe's digital future," accessed Oct. 30, 2025; EuroStack, "Open letter: European industry calls for strong commitment to sovereign digital infrastructure," March 14, 2025; Natasha Lomas, "European tech industry coalition calls for 'radical action' on digital sovereignty—starting with buying local," TechCrunch, March 16, 2025.
8. European Commission, "The international digital strategy for the European Union," July 8, 2025.
9. Diana Goovaerts, "Europe's cloud market poised for 24% growth," Fierce Network, July 28, 2025.

10. Amazon, "AWS plans to invest €7.8 billion into the AWS European Sovereign Cloud," May 15, 2024; Amazon, "Built, operated, controlled, and secured in Europe: AWS unveils new sovereign controls and governance structure for the AWS European Sovereign Cloud," June 3, 2025.
11. Brad Smith, "Microsoft announces new European digital commitments," Microsoft, April 30, 2025; Judson Althoff, "Announcing comprehensive sovereign solutions empowering European organizations," Microsoft, June 16, 2025.
12. European Commission, "Commission sets course for Europe's AI leadership with an ambitious AI Continent Action Plan," press release, April 9, 2025.
13. Belle Lin, "Nvidia and Perplexity team up in European AI push," The Wall Street Journal, June 11, 2025.
14. Amiya Johar, "Nvidia, MGX lead €8.5B project to build French AI data center," Mobile World Live, May 20, 2025.
15. OpenAI, "Introducing Stargate UK," Sept. 16, 2025; Tom Bristow, "US tech firms pour £30B into UK as Trump lands," Politico, Sept. 16, 2025.
16. European Commission, "European Chips Act: The Chips for Europe Initiative," Nov. 4, 2024; European Commission, "European Chips Act," accessed Oct. 30, 2025.
17. Jingyue Hsiao, "TSMC breaks ground on EUR10 billion semiconductor fab in Dresden," Digitimes Asia, Aug. 21, 2024; Infineon, German government issues final funding approval for new Infineon fab in Dresden," press release, May 8, 2025; Adrià Calatayud and Mauro Orru, "Apple supplier STMicroelectronics to build \$5.4 billion chip plant in Italy," The Wall Street Journal, May 31, 2024.
18. Jeff Foust, "Europe signs contracts for IRIS² constellation," SpaceNews, Dec. 16, 2024.
19. Eutelsat, "High-speed, low-latency connectivity," accessed Oct. 30, 2025.
20. Jason Rainbow, "French government to lead Eutelsat's \$1.56 billion capital boost," SpaceNews, June 19, 2025; Rachel Jewett, UK to join Eutelsat's capital raise with \$105M investment," Via Satellite, July 10, 2025.
21. Margherita Stancati, Matthew Dalton, and Vera Bergengruen, "Europe scrambles to break its dependence on Musk's satellites," The Wall Street Journal, April 13, 2025.
22. Byun Hee-won and Kim Mi-geon, "South Korea to pour \$735 bn into developing sovereign AI built on Korean language and data," The Chosun Daily, June 17, 2025.
23. Zinnia Lee, "Korea's Kakao teams up with OpenAI to develop AI products," Forbes, Feb. 4, 2025.
24. Zinnia Lee, "Billionaire Chey's SK Group partners with Amazon to build a \$5 billion AI data center in Korea," Forbes, June 23, 2025.
25. Dylan Butts, "Japan is ramping up efforts to revive its once dominant chip industry," CNBC, Nov. 13, 2024; Rapidus, "Rapidus Corporation," accessed Oct. 30, 2025.
26. Cassava Technologies, "Cassava to upgrade its data centres with NVIDIA supercomputers to drive Africa's AI future," accessed Oct. 30, 2025; Nell Lewis, "Africa's first 'AI factory' could be a breakthrough for the continent," CNN, April 3, 2025.
27. INDIAai | Pillars; Government of India, "India semiconductor mission," accessed Oct. 30, 2025.
28. Shadma Shaikh, "Inside India's scramble for AI independence," MIT Technology Review, July 4, 2025.
29. India Stack, "India Stack," accessed Oct. 30, 2025.

- 30. Government of Canada, "Canadian sovereign AI compute strategy," Oct. 1, 2025.
- 31. Government of Canada, "Canada partners with Cohere to accelerate world-leading artificial intelligence," press release, Aug. 19, 2025.
- 32. Telus, "TELUS to launch Canada's leading sovereign AI factory, powered by NVIDIA to drive the nation's AI future," March 19, 2025; Bell, "Increasing sovereign AI capacity: Introducing Bell AI Fabric," May 28, 2025; SaskTel, Deloitte Canada and SaskTel announce strategic alliance to bring Artificial Intelligence (AI) capabilities and solutions to market, advancing Canada's AI vision," press release, Sept. 23, 2025.
- 33. OpenAI, "Introducing Stargate UAE," May 22, 2025.
- 34. Amazon, "AWS and HUMAIN announce a more than \$5B investment to accelerate AI adoption in Saudi Arabia and globally," May 13, 2025; Nvidia, "HUMAIN and NVIDIA announce strategic partnership to build AI factories of the future in Saudi Arabia," press release, May 13, 2025; PIF, "HRH Crown Prince launches HUMAIN as global AI powerhouse," press release, May 12, 2025.
- 35. Natasha Turak, "Saudi AI firm Humain is pouring billions into data centers. Will it pay off?" CNBC, Aug. 27, 2025.
- 36. Chris Metinko, "Defense tech venture funding gains traction," Crunchbase News, Feb. 12, 2025.
- 37. 3MTT, "Shaping the future of Nigeria's digital workforce," accessed Oct. 30, 2025; European Commission, "Commission to invest €1.3 billion in artificial intelligence, cybersecurity and digital skills," press release, March 28, 2025.
- 38. Goldman Sachs, "AI to drive 165% increase in data center power demand by 2030," Feb. 4, 2025; Felicity Barringer, "Thirsty for power and water, AI-crunching data centers sprout across the West," Stanford University, April 8, 2025.
- 39. Karthik Ramachandran, Duncan Stewart, Kate Hardin, Gillian Crossan, and Ariane Bucaille, "As generative AI asks for more power, data centers seek more reliable, cleaner energy solutions," Deloitte Insights, Nov. 19, 2024.
- 40. ECDPM, "Von der Leyen in India: A tech sovereignty partnership in the making," Feb. 28, 2025; Nii Simmonds and David Timis, "How Europe and Africa can unlock tech opportunities through stronger collaboration," World Economic Forum, Aug. 18, 2025.
- 41. Rolfe Winkler, Nate Rattner, and Sebastian Herrera, "Big tech's \$400 Billion AI spending spree just got Wall Street's blessing," The Wall Street Journal, July 31, 2025; Financial Times, "What'll happen if we spend nearly \$3tn on data centres no one needs?" July 30, 2025.
- 42. Caiwei Chen, "China built hundreds of AI data centers to catch the AI boom. Now many stand unused," MIT Technology Review, March 26, 2025.

Part 3

미디어 산업의 주의력 경제 시대, 서사·신뢰·플랫폼의 재구성

8장 공영방송, 생존전략으로 글로벌 스트리밍 서비스와 협력 선택

- 1. Chris Arkenberg et al., "2025 media and entertainment outlook," Deloitte Insights, April 23, 2025.
- 2. Ofcom, "Transmission critical: The future of public service media," July 21, 2025.
- 3. Elsa Keslassy, "How streamers and broadcasters' cross carriage deals could disrupt the TV business in Europe," Variety, July 11, 2025.

4. Knight Foundation, "Public broadcasting: Its past and its future," accessed Oct. 29, 2025.
5. Once hugely profitable, in most countries the audience for traditional "linear" TV is getting smaller and older. Globally, multiple public service traditional TV broadcasters (PSBs) are seeing their viewership erode, especially among younger audiences. In the United Kingdom, for example, less than half (48%) of 16 to 24-year-olds watched any broadcast TV in a given week in 2023, down from 76% five years earlier [CSI, "Gen Z abandons traditional broadcast TV: Ofcom," July 31, 2024], and only 55% of children between 4 and 15 tuned in weekly, down from 81% in 2018. Still in the United Kingdom, young adults who do watch traditional TV spend barely half an hour per day on it, versus 93 minutes on video-sharing platforms like YouTube and TikTok. Even overall reach is shrinking: the weekly audience for any broadcast TV fell to 75% of Britons in 2023 (down from 79% the year prior)—the steepest decline on record. The same pattern is found in nearly every advanced country's media market and is generally true for both PSB and commercial broadcasters.
6. Funk, "Funk Bericht 2024," Dec. 13, 2024.
7. Ibid.
8. Evan Shapiro and Marion Ranchet, "TESTING & LEARNING: The CBC Case Study," The Media Odyssey, audio podcast episode, April 24, 2025.
9. Ibid.
10. Social Blade, "CBC News YouTube channel statistics," accessed Oct. 29, 2025.
11. Shapiro and Ranchet, "TESTING & LEARNING."
12. Ibid.
13. K.D. with AFP, "'Un accord historique': après TF1 et Netflix, France Télévisions s'associe à Prime Video pour diffuser ses contenus sur Amazon," BFM Tech & Co, July 3, 2025.
14. Ibid.
15. Ibid.
16. AFP, "Netflix breaks new ground with global launch of French TV content," ForbesIndia.com, June 19, 2025.
17. Mark Sweeney, "BBC and ITV slash big-budget TV spend as US streamers pour money into UK," The Guardian, Feb. 16, 2025.
18. Pact, "Submission to Ofcom consultation on the proposals for the new Channel 4 licence," February 2024.
19. BBC, "His Dark Materials: Critics heap praise on 'ravishing' dramatisation," Nov. 4, 2019.
20. Sheena Scott, "'His Dark Materials' is BBC's most expensive series and promises to be a faithful adaptation," Forbes, Oct. 31, 2019.
21. BBC, "How The End of the F***ing World became a cult TV phenomenon," Nov. 4, 2019; Daniel D'Addario, "'It's a Sin' is a transporting and tragic tale of the AIDS epidemic: TV review," Variety, Feb. 21, 2018.
22. Travis Clark, "8 great Netflix original TV series that show how well its British strategy is working," Business Insider, April 2, 2019.
23. John Moulding, "ITV and C4 happy to let viewers watch long-form content on YouTube," The Media Leader, March 13, 2025.
24. Max Goldbart, "Streamers Will Not Be Regulated Fully In UK For Another Two Years," Deadline, Feb. 26, 2025.
25. Lucas Manfredi, "Netflix, France's TF1 strike landmark distribution deal," TheWrap, June 18, 2025.

9장 솟풀이 촉발한 콘텐츠 제작 생태계의 혁신과 플랫폼 신뢰 위기

1. Rui Ma, "State of short drama apps 2025," Mobile App Insights, July 2025.
2. Stephanie Yang, "Two-minute TV shows have taken over China. Can they take over the world?" Los Angeles Times, March 16, 2025.
3. Ma, "State of short drama apps 2025."
4. Gaby Hinsliff, "It's the age of regret: Gen Z grew up glued to their screens, and missed the joy of being human," The Guardian, March 7, 2025.
5. Xinhuanet, "Love, twist and one-minute cliffhangers: China's micro dramas go global," July 29, 2025.
6. CMB Global Markets' equity research, May 23, 2025 (private report accessed via AlphaSense).
7. Mandy Zuo, "China's addictive micro-dramas show how commercial demand is fuelling a netcasting boom," South China Morning Post, March 27, 2025.
8. Jeff Huang, "How China's \$7 billion micro drama industry is taking on the US entertainment industry," CNBC, July 22, 2025.
9. Kunal Purandare, "The VC-backed rise of micro dramas in India," Forbes India, Aug. 8, 2025.
10. Systematix Institutional Research and Morning Brew, July 10, 2025 (private report sourced via AlphaSense).
11. The Economic Times, "Stage set for micro-dramas; WhatsApp's monetisation bid," June 17, 2025.
12. Robert Steiner, "Microdrama plot twist: A threat to the apps' stratospheric US growth," Variety, April 30, 2025.
13. Ma, "State of short drama apps 2025."
14. Ibid.
15. Appfigures, "Top ranked iOS app store apps," accessed Oct. 23, 2025.
16. Steiner, "Microdrama plot twist: A threat to the apps' stratospheric US growth."
17. Paige Gawley, "People on TikTok are obsessed with a fake group chat," Vice, April 9, 2025.
18. Lauren Forristal, "Netflix is getting into short videos with a new vertical feed for mobile," TechCrunch, May 7, 2025.
19. Global economic outlook and investment strategy, 2H 2025; ICBC International Research (private research brief via AlphaSense).
20. Carson Taylor, "Microdramas: China's new craze goes global," Naavik, Sept. 8, 2024.
21. Focus on structurally high-growth segments, Huatai Securities, Aug. 23, 2025 (private documents sourced via AlphaSense).
22. Kristian Monroe, "Told one minute at a time, micro dramas are soap operas designed to fit in your hand," NPR, March 19, 2025.
23. Taylor, "Microdramas: China's new craze goes global."
24. Oxford University Press, "'Brain rot' named Oxford word of the year 2024," Dec. 2, 2024.
25. Sian Boyle, "Is doom scrolling really rotting our brains? The evidence is getting harder to ignore," The Guardian, Dec. 9, 2024.
26. Annalee Newitz, "Social media is dead – here's what comes next," NewScientist, July 23, 2025.
27. Rodney Mason, "Social isn't social anymore—now what?" Forbes, April 28, 2025.

28. Jessica A. Kent, "Need a break from social media? Here's why you should—and how to do it," Harvard Summer School, Aug. 28, 2023.
29. Sanzana Karim Lora, Sadia Afrin Purba, Bushra Hossain, Tanjina Oriana, Ashek Seum, and Sadia Sharmin, "Infinite scrolling, finite satisfaction: Exploring user behavior and satisfaction on social media in Bangladesh," Arxiv, April 15, 2025.
30. Deloitte Digital, "2025 state of social research: How efficiency can meet impact with the right investments," May 15, 2025.
31. Gillian Follett, "How creators are shaping Cannes Lions—from business discussions to the campaigns winning awards," AdAge, June 12, 2025.
7. Bill Chappell, "AI video ad, Kalshi advertising NBA finals," NPR, June 23, 2025.
8. Thomas H. Davenport and Nitin Mittal, "How generative AI is changing creative work," Harvard Business Review, Nov. 14, 2022.
9. Torin Anderson and Shuo Niu, "Making AI-enhanced videos: Analyzing generative AI use cases in YouTube content creation," In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, pp. 1–7. 2025.
10. Collectively Inc., "How content creators are embracing generative AI and AI avatars: insights from our latest survey," Jan. 14, 2025.
11. Jess Weatherbed, "TikTok ads may soon contain AI-generated avatars of your favorite creators," The Verge, June 17, 2024
12. Anderson and Niu, "Making AI-enhanced videos: Analyzing generative AI use cases in YouTube content creation."
13. Yael Malamatinas, "7 of the best AI dubbing tools to translate videos into different languages," Vimeo, blog, April 28, 2025.
14. Screen Actors Guild – American Federation of Television & Radio Artists, "SAG-AFTRA statement on the use of artificial intelligence and digital doubles in media and entertainment," March 17, 2023.
15. Katie Kilkenny, "Higher costs are hitting film and TV producers even as studios keep trimming budgets," The Hollywood Reporter, April 17, 2025.
16. Chris Arkenberg, Jeff Loucks, Kevin Westcott, Danny Ledger, and Doug Van Dyke, "2025 media and entertainment outlook," Deloitte Insights, April 23, 2025.
17. Interactive Advertising Bureau, "Digital ad revenue surges 15% YoY in 2024, climbing to \$259 B," April 17, 2025.

10장 생성형 AI 영상이 불러온 콘텐츠 제작 환경의 혁신과 소셜 플랫폼의 책임

1. U.S. Congress, Senate, "S. 314 – A bill to protect the public from the misuse of the telecommunications network and telecommunications devices and facilities," accessed Oct. 22, 2025.
2. Mayer Brown LLP, "Children's online privacy: recent actions by the states and the FTC," Feb. 25, 2025.
3. Ella Lee, "Supreme Court — Mississippi social-media law and minors' access," The Hill, Aug. 14, 2025.
4. European Commission, "Commission press corner detail: IP/25/1820," press release, July 14, 2025.
5. James Beser, "Extending our built-in protections to more teens on YouTube," YouTube News & Events Blog, July 29, 2025.
6. The New York Times, "AI video deepfakes – quiz and playground," June 29, 2025.

18. Ryan Browne, "AI is disrupting the advertising business in a big way — industry leaders explain how," CNBC, June 15, 2025.
19. Charles James, "Generative AI for retail ad campaign variants and A/B testing automation," ResearchGate, Nov. 9, 2024.
20. Interactive Advertising Bureau, "Nearly 90% of advertisers will use Gen AI to build video ads, according to IAB's 2025 video ad spend & strategy full report," July 15, 2025.
21. China Widener, Jana Arbanas, Doug Van Dyke, Chris Arkenberg, Bree Matheson, and Brooke Auxier, "2025 digital media trends: Social platforms are becoming a dominant force in media and entertainment," Deloitte Insights, March 25, 2025.
22. Clare Duffy, "OpenAI's Sam Altman warns of an AI 'fraud crisis,'" CNN, July 22, 2025.
23. Paris Martineau, "Exclusive: Section 230 may finally get changed — lawmakers prep new bill," The Information, accessed Oct. 22, 2025.
24. Dawn Carla Nunziato, "The Digital Services Act and the Brussels Effect on platform content moderation," Chicago Journal of International Law 24, no. 1 (2024): pp. 1-37.

11장 비디오 팟캐스트, 위협적인 글로벌 미디어·광고 플랫폼으로 진화

1. Based on Deloitte analysis; Brooke Auxier, Bree Matheson, Duncan Stewart & Kevin Westcott, "Shuffle, subscribe, stream: Consumer audio market is expected to amass listeners in 2024, but revenues could remain modest," Deloitte Insights, Nov. 29, 2023.
2. Spotify Newsroom, "Spotify unveils uninterrupted video podcasts, audience-driven payments, and the new Spotify for Creators platform," Nov. 13, 2024; Wondery, "Now playing: Video podcasts on the Wondery app for Wondery+ subscribers," accessed Oct. 23, 2025; Angela Yang, "Podcasts are taking over TV screens as video formats grow increasingly popular," NBC News, Dec. 23, 2024.
3. Spotify Newsroom, "From audio to video, Spotify's \$100 million payout fuels creator success stories," Apr. 28, 2025.
4. Ariel Shapiro, "YouTube launches a dedicated page for podcasts," The Verge, Aug. 23, 2022.
5. Todd Spangler, "YouTube says it now has more than 1 billion monthly viewers of podcast content," Variety, Feb. 26, 2025; Zach Vallese, "YouTube launches weekly top podcast list to rival Spotify and Apple," CNBC, May 15, 2025.
6. Alex Schiffer, "Taylor Swift draws 1.3 million live viewers in 'New Heights' appearance," Front Office Sports, Aug. 13, 2025; Vicki Newman, "Taylor Swift earns podcast record with appearance on boyfriend Travis Kelce's New Heights, Guinness World Records," Aug. 26, 2025.
7. Based on Deloitte analysis of publicly available data.
8. Data from Deloitte's Fall 2025 Digital Media Trends 19 survey.
9. Edison Research, "YouTube is the preferred podcast listening service," Oct. 23, 2024.

10. Ellie Hammonds, "Vodcasts: Is it the future of podcasting?" *The Media Leader*, Aug. 28, 2025.
11. Molly Fuard, "Visual podcasting is now a thing and here's what advertisers should know," *Adweek*, accessed Oct. 23, 2025.
12. Lloyd George, "Why social media is a game-changer for growing your podcast," *Acast*, accessed Oct. 23, 2025.
13. Alexander Lee, "Podcast consumption shifts towards connected TVs," *Digiday*, May 7, 2025.
14. Eve Upton-Clark, "Netflix is eyeing video podcasts as it expands beyond TV and film," *Fast Company*, April 21, 2025.
15. Data from Deloitte's Fall 2025 Digital Media Trends 19 survey.
16. Ibid.
17. Sara Fischer, "Axios media trends," *Axios*, April 22, 2025.
18. Simon Kemp, Digital 2025: The essential guide to the global state of digital," *Meltwater*, Feb. 5, 2025.
19. Global System for Mobile Communications Association, "The mobile economy 2025," accessed Oct. 23, 2025.
20. Paula Gilbert, "Nigeria's 1GB data price has dropped 75% over five years," *Connecting Africa*, June 5, 2020; Peter Oluka, "\$0.39 [604 NGN] per 1GB: Nigeria among countries with cheapest data rates," *Tech Economy*, Jan. 10, 2025; Bruno Venditti, "The cost of 1GB of mobile data worldwide," *Visual Capitalist*, Oct. 21, 2024.
21. Spotify Newsroom, "Get to know the 13 podcast grantees of Spotify's new Africa podcast fund," Oct. 24, 2022; Blueprint Magazine, "Spotify and the pod network enters a new era of Filipino podcasting with the launch of their state-of-the-art studio," April 28, 2025.
22. Spotify Newsroom, "The Spotify partner program expands to nine new markets, giving more creators new ways to monetize their content," March 27, 2025.
23. IndustryPods, "Podcast distribution on international platforms," December 2024; The Storiez, "How Anghami is dominating the music streaming market globally," Sept. 14, 2024; Peerzada Abrar, "Kuku FM raises \$25 mn from investors; aims to expand content, improve tech," *Business Standard*, Sept. 20, 2023; Samuel Viavonu, "The podcast boom in Nigeria: an era of noise or knowledge?" *Afrocritik*, Feb. 19, 2025.
24. Acast, "The video podcast opportunity," June 10, 2025; Devan Kaloo and Robert Gilhooly, "Demystifying emerging markets," *Aberdeen Investments*, Sept. 8, 2023.
25. BeMultilingual, "What are the most popular languages on YouTube?" July 26, 2025; David R. Gonzalez, "The state of podcasting in Latin America," *PodNews*, Feb. 15, 2024.
26. Aaron Chow, "Nike Japan launches 'NIKELAB RADIO,'" *HypeBeast*, July 28, 2021.
27. Burt Helm, "How AI for lip dubbing could change the film industry," *Fast Company*, November 2023.
28. The New York Times Style Magazine: Australia, "Forensic fandom and the age of the companion podcast," Feb. 27, 2025.
29. Damion Taylor, "How podcasts are becoming Hollywood's new development pipeline," *Forbes*, Jan. 30, 2025.

Part 4**인프라에서 경험으로...****통신의 재정의****12장 차세대 위성 인터넷 시대로의 전환과 통신 인프라·경쟁 질서의 재편**

1. Gartner, "Gartner forecasts LEO satellite communications services spending to hit \$14.8bn globally in 2026," press release, July 30, 2025.
2. Deloitte analysis of publicly available market research and forecasts, combining current adoption trends, planned service launches, and demand in underserved regions to assess the feasibility of future subscriber growth.
3. Deloitte analysis of global low Earth orbit (LEO) satellite deployment trends indicates five major constellations—Starlink, Kuiper, Guowang, Honghu-3, and G60—will account for a significant proportion of the estimated 15,000 to 18,000 LEO satellites expected in orbit by the end of 2026. This projection aggregates operator-specific deployment targets, launch rate trends, and industry growth forecasts, referencing broker research and company filings.
4. This is based on a Deloitte analysis of publicly available industry data and forecasts, including current deployments as of mid-2025, announced launch schedules from major operators, and long-term projections from leading research providers. Estimates were derived by combining existing satellite counts with confirmed launch plans and aligning them with independent analysts' projections.
5. Yarnaphat Shaengchart and Tanpat Kraiwant, "Starlink satellite project impact on the Internet provider service in emerging economies," Research in Globalization, May 4, 2023.
6. Nick Cowell, "Satellite-based internet connectivity LEO Satellite Broadband," Fujitsu, May 22, 2023.
7. Karen L. Jones and Audrey L. Allison, "The great convergence and the future of satellite-enabled direct-to-device," Center for Space Policy And Strategy, September 2023.
8. Joe Madden, "The difference between NTN/D2D and satellite broadband – Madden," Fierce Network, Jan. 16, 2024.
9. Christopher Baugh, "Satellite direct-to-device: The characteristics of D2D constellations will limit SpaceX's ability to dominate," Analysys Mason, July 22, 2024.
10. Mike Robuck, "Musk outlines SpaceX D2D spectrum strategy," Mobile World Live, Sept. 10, 2025.
11. Ibid.
12. These services include emergency messaging, basic data transmission, and sometimes voice calls.
13. David Jarvis, Duncan Stewart, Raghavan Alevoor, and Kevin Westcott, "Signals from space: Direct-to-device satellite phone connectivity boosts coverage," Deloitte Insights, Nov. 29, 2023.
14. Deloitte analysis of publicly available data on satellite industry investments for 2023–2024; investment amounts reflect disclosed funding rounds, commercial agreements, and reported capital commitments related to direct-to-device satellite communication.
15. Deloitte analysis of global direct-to-device satellite communication capital requirements, based on company filings, investor presentations, earnings call transcripts, government announcements, press releases, research reports, and expert interviews.
16. David Jarvis, Duncan Stewart, Raghavan Alevoor, and Kevin Westcott, "Signals from space."

17. Aamir Siddiqui and Andrew Grush, "Android and iPhone satellite connectivity: What is it and what are your options right now?" Android Authority, Feb. 11, 2025.
18. Arun Menon, "Satellite industry trends to watch in 2024," TM Forum, Jan. 31, 2024.
19. GSMA, "Open consultation for the council working group on international internet related public policy issues," August 2020.
20. Ibid.
21. GSMA, "New GSMA report shows mobile internet connectivity continues to grow globally but barriers for 3.45 billion unconnected people remain," press release, Oct. 23, 2024.
22. 5G Americas, "New developments and advances in 5G and NT," February 2025.
23. Amber Jackson, "Project Kuiper explained: Australia's bid to improve internet access with Amazon," Capacity, Aug. 5, 2025.
24. Ling Xin and Victoria Bela, "China launches first satellites for GuoWang project to rival SpaceX's Starlink," South China Morning Post, Dec. 16, 2024.
25. Mark Holmes, "Telesat's Lightspeed is now fully funded, MDA to build constellation," Via Satellite, Aug. 11, 2023.
26. Connectivity and Secure Communications, "ESA confirms kick-start of IRIS² with European Commission and SpaceRISE," Dec. 16, 2024.
27. SatNews, "Loft Orbital and Marlan Space to create the Middle East's first private manufacturing space company of commercial satellite constellations for LEO," Aug. 26, 2024.
28. Reuters, "Eutelsat announces contract with Airbus for 100 satellites," Dec. 17, 2024.
29. David Jarvis, Duncan Stewart, Kevin Westcott, and Ariane Bucaille, "Too congested before we're connected? Broadband satellites will need to navigate a crowded sky," Deloitte Insights, Nov. 30, 2022.
30. CCIA, "Low earth orbit (LEO) satellite broadband facts and stats," March 5, 2025.
31. Rick Mur, "Low-earth orbit (LEO) networks in your global connectivity strategy," GNX, Jan. 22, 2025.
32. Ibid.
33. Garinder Shankrowalia, "Amazon's ambitions: Project Kuiper and the complex future of satellite broadband," Omdia, May 20, 2025.
34. Rakuten.Today, "Moshi moshi? Space calling: Rakuten Mobile and AST SpaceMobile achieve Japan first satellite-to-mobile video call," May 2, 2025; Cameron Page, "Australia's TPG completes first D2D satellite trials with Lynk," TelcoTitans, May 8, 2025; John Tanner, "Globe kicks off Lynk Global D2D SMS tests in Zambales," Developing Telecoms, Oct. 7, 2024.
35. Nadine Hawkins, "Amazon to launch Project Kuiper satellites next week," Capacity Media, April 3, 2025.
36. Amazon, "Here's how Project Kuiper's satellite network can help telecom partners like Vodafone and Vodacom enhance reliability and extend reach," Sept. 5, 2023,
37. World Population Review, "Internet cost by country 2025," accessed Oct. 30, 2025.
38. Ibid.
39. Jack Kuhr, "Starlink Mini Impact and Rapid Terminal Iteration: Payload Research," Payload, June 26, 2024.
40. Hawkins, "Amazon to launch Project Kuiper satellites next week."

41. Elton Chang, "Satellite network capacity and scalability," *TelecomWorld101*, Jan. 17, 2025.
42. Ibid.
43. Kim Larsen, "The next frontier: LEO satellites for internet services," *techneconomyblog*, March 12, 2024.
44. Andrea D'Ambrosio, Miles Lifson, and Richard Linares, "The capacity of low earth orbit computed using source-sink modeling," *arxiv*, June 10, 2022.
45. Harry Baker, "How many satellites could fit in earth orbit? And how many do we really need?" *LiveScience*, May 30, 2025.
46. Kelly Hill, "FCC revisits satellite spectrum power levels," *RCR Wireless News*, May 1, 2025.
47. Dan Heming, "Starlink waitlists return, network congestion on the rise and finally, a customer support phone #," *Mobile Internet Resource Center*, Nov. 21, 2024.
48. Mark Jackson, "Starlink's satellite broadband hits capacity limit in South East England," *ISPreview*, Dec. 31, 2024.
49. Marcin Frąckiewicz, "Artificial intelligence in satellite and space systems," *Tech Stock 2*, June 12, 2025. Luis Manuel Garcés-Socarrás et al., "Artificial Intelligence implementation of onboard flexible payload and adaptive beamforming using commercial off-the-shelf devices," *arXiv*, May 3, 2025.
50. James Ratzer, "Starlink: What impact might it have on the telcos?" *New Street Research*, June 9, 2025.
51. Stella Linkson, "Starlink ground stations: What they are and how they work," *Starlink Info*, March 21, 2025; Shankrowalia, "Amazon's ambitions."
52. Linkson, "Starlink ground stations: What they are and how they work."
53. Kim Larsen, "Will LEO satellite direct-to-cell networks make terrestrial networks obsolete?" *techneconomyblog*, January 20, 2025.
54. Federal Communications Commission, "FCC advances supplemental coverage from space framework," March 15, 2024.
55. K. C. Halm, John C. Nelson Jr., and Kasey McGee, "NTIA revamps federally funded \$42.5 billion broadband deployment subsidy program," *Davis Wright Tremaine LLP*, June 12, 2025.
56. David Shepardson, "US Senate panel advances Trump nominee to oversee \$42-billion government internet fund," *Reuters*, April 9, 2025.
57. European Conference of Postal and Telecommunications Administrations, "An introduction to the European regulatory environment for radio equipment and spectrum," Feb. 5, 2024.
58. Commission for Communications Regulation, "Radio spectrum management operating plan for 2025–2028," Dec. 13, 2024.
59. ITU-APT Foundation of India, "Recommendations on telecom regulatory authority of India consultation paper on assignment of spectrum for space-based communication services," April 6, 2025.
60. Cetecom Advanced, "China introduces first regulatory framework for radar radio management," March 24, 2025.
61. Ministry of Internal Affairs and Communications, "Progress on the WX promotion strategy action plan," May 29, 2025.
62. Deloitte analysis of publicly reported or analyst-modelled 2026 investment details of major companies in the LEO space.
63. Peter Chahal, Avinash Naga, Courtney Munroe, Bruno Teyton, and Nikhil Batra, "Worldwide telecommunications capex forecast, 2025–2029," *IDC Research*, June 2025.
64. Newsroom, "AST SpaceMobile secures strategic investment from AT&T, Google and Vodafone," *Business Wire*, Jan. 18, 2024; Hema Kadia, "Verizon's \$100 million investment in AST SpaceMobile for satellite connectivity," *TeckNexus*, May 29, 2024.

- 65. Inside GNSS, "The case for LEO GNSS at C-Band," Feb. 3, 2025.
- 66. Damilare Dosunmu, "How Starlink took over Africa's largest internet market," Rest of world, April 15, 2025.
- 67. Larsen, "Will LEO satellite direct-to-cell networks make terrestrial networks obsolete?"
- 68. Pablo Tomasi, "Space to grow: Enterprise LEO forecast 2025–30," Omdia, Sept. 9, 2025.
- 8. Espen Tønnessen, Thomas Haugen, and Shaher Ahmmad Ibrahim Shalfawi, "Reaction time aspects of elite sprinters in athletic world championships," *Journal of Strength & Conditioning Research* 27, no. 4 (2013): pp. 885–92.
- 9. IR, "Network latency—Common causes and best solutions," accessed October 2025.
- 10. Ofcom, Mobile matters.
- 11. Ofcom, "Improving your mobile phone reception," May 27, 2022.
- 12. William Webb, "It's time to rethink 6G," IEEE Spectrum, February 10, 2025.
- 13. Netflix, "Netflix-recommended internet speeds," accessed October 2025; Paul Lee, Dieter Trimmel, and Eytan Hallside, "No bump to bitrates for digital apps in the near term: Is a period of enough fixed broadband connectivity approaching?," TMT Predictions 2024, Deloitte, November 29, 2023.
- 14. Tefficient, "The demand for additional mobile data is weaker than ever—ARPU growth softens," July 31, 2025.
- 15. Michael Irving, "'Ultrabroadband' 6G chip clocks speeds 10 times faster than 5G," ScienceAlert, September 3, 2025; 4G.co.uk, "How fast are 4G and 5G?," accessed October 2025.
- 16. ETTelecom.com, "5G will make cost of GB lower than 4G: Experts," July 31, 2020.
- 17. NTT DOCOMO, "DOCOMO, NTT, NEC and Fujitsu develop top-level sub-terahertz 6G device capable of ultra-high-speed 100 Gbps transmission," press release, April 11, 2024.
- 18. New Street Research, Europea Tariff Tracker, accessed October 2024.
- 19. Jennifer Johnson, "Peak data growth is a quiet win for telcos," Reuters, June 2, 2025.

13장 통신사 경쟁력은 네트워크 속도에서 고객이 체감하는 가치로 이동

- 1. GSMA, 5G Network Slicing, accessed October 2025.
- 2. Paul Lee and Ben Stanton, Deloitte Digital Consumer Trends 2025, UK Edition, Deloitte LLP, June 2025; Adrie Cronje et al., Digital Consumer Trends 2024, Netherlands Edition, Deloitte LLP, December 2024; Vincent Frosty and Vincent Pirard, Digital Consumer Trends 2024, Belgium Edition, Deloitte Belgium, 2024.
- 3. Vodafone UK, 2G, accessed October 2025.
- 4. Simon Thomas, "What is the difference between 3G and 4G?," 4G.co.uk, September 29, 2014.
- 5. Ivor Nicholls, "LTE vs 4G: Understanding the difference between LTE and 4G," UCtel, February 10, 2025.
- 6. Andrew Wooden, "The telecoms industry's biggest problem? Failure to monetise 5G," Telecoms.com, March 14, 2024.
- 7. Ofcom, Mobile matters, July 17, 2025.

20. Evert de Boer and Xiao Yao Chin, Top 100 most valuable airline loyalty programs, On Point Loyalty, January 2023.
21. Consumer Financial Protection Bureau (CFPB), "CFPB takes action on bait-and-switch credit card rewards tactics," news release, last modified December 18, 2024.
22. Mike Sievert, "The power of appreciation: Taking customer loyalty to the next level," Un-carrier blog, T-Mobile, February 13, 2024.
23. Vodafone UK, "Spin for a chance to win £1,000 each day with VeryMe Rewards," press release, June 9, 2025.
24. Virgin Media O2, "Priority from O2 launches 'Blue Mondays' with millions of unmissable rewards, prizes and experiences for customers," April 28, 2025.
25. CI Group, "The psychology of freebies: Why small rewards yield big returns," accessed October 2025.
26. Octopus Energy, "Octoplus, our rewards programme for smart meter customers," accessed October 2025.
27. Vodafone UK, "Music festivals," accessed October 2025.

인더스트리 리더십



Gillian Crossan

Global Technology, Media & Telecommunications Industry Leader
gicrossan@deloitte.com



Tim Bottke

Global Telecommunications, Media & Entertainment sector Leader
tbottke@deloitte.de



Girija Krishnamurthy

Global Technology sector Leader
gkrishnamurthy@deloitte.com



Deb Bhattacharjee

Global Semiconductor Center of Excellence Leader
debbhattacharjee@deloitte.com



Jody McDermott

Global Telecom Leader
jodymcdermott@deloitte.ca

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



Download on the
App Store



GET IT ON
Google Play



딜로이트 인사이트 카카오 채널

바쁜 현대인을 위한 3분 카드뉴스



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, 가치사슬 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.

인공 지능(AI) 활용이 본격적으로 확대되면서, 최근 수년간 첨단기술·미디어·통신(TMT) 산업의 중요성이 한층 부각됐습니다. AI 발전을 주도하는 TMT 업종 기업들이 S&P 500 시가총액에서 차지하는 비중이 2008년 19%에서 2025년 말 기준 53%로 급증했습니다. 이제 AI 성숙기의 서막이 될 2026년, TMT 산업의 가장 중요한 핵심은 화려한 기술적 돌파구를 넘어 AI의 실질적인 대규모 도입이 될 것입니다.

델oitte는 2026년을 기점으로 AI가 제시하는 가능성과 실질적으로 창출하는 가치 사이의 간극이 점차 줄어들 것으로 예측합니다. 이는 기업들이 AI 추론 컴퓨팅 수요를 충족하고, AI 에이전트 오케스트레이션(AI agent orchestration)을 통해 운영 효율을 높이며, AI 기반 산업용 로봇과 드론으로 노동력 부족 문제를 해결하는 등 보이지 않는 곳에서 이루어지는 꾸준한 성장에 힘입은 결과입니다. 하지만 성장과 혁신에는 어려움이 따릅니다. 지정학적 불안정으로 인해 반도체 공급망이 취약해지면서 각국이 기술 주권을 추구하게 되었습니다. 소셜 미디어에서 확산되는 생성형 AI 영상은 허위 정보를 부추길 위험이 있으며, 규제와 AI 콘텐츠 표기 요구를 더욱 강화할 가능성이 큽니다. 통신사들은 더 빠른 인터넷 속도만으로 소비자를 유치하는 데 어려움을 겪으며 다른 인센티브로의 전환을 필요로 할 것입니다.

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.