Deloitte Insights

July 2025

생성형 AI 시대, 사이버 보안 리더의 리스크 분석 및 대응 방안

Deloitte Insights

Deloitte.

들어가며

생성형 인공지능(Gen AI)은 조직의 제품과 운영 방식을 향상시킬 수 있는 상당한 기회를 제공하지만, 동시에 내부적·외부적으로 새로운 리스크도 초래한다. 딜로이트의 '2023년 4분기 기업의 생성형 AI 사용현황 보고서' 에 따르면, 전 세계 응답자들은 생성형 AI 전략을 확대하는 데 있어 리스크 관리와 규제 준수를 가장 큰 우려사항으로 꼽고 있다.

이러한 과제는 데이터 출처, 보안, 미성숙한 시장 환경 등 다양한 이슈가 복합적으로 얽혀 있는 교차적인 성격을 가진다. 이에 대응하기 위해 많은 조직들이 사이버 보안에 대한 투자를 확대하고 있다. 실제로 딜로이트의 '2023년 3분기 기업의 생성형 AI 사용현황 보고서'에서는 응답자의 73%가 생성형 AI 프로그램으로 인해 사이버 보안 투자를 늘릴 계획이라고 밝혔으며, 딜로이트의 '2024년 디지털 전환 투자 변화 보고서'에 따르면 미국 응답자의 59%는 지난 12개월 동안 사이버 보안 역량에 투자한 바 있다.

하지만 생성형 AI가 새로운 리스크를 지속적으로 만들어내는 만큼, 조직의 리더들은 현재는 물론 미래에도 효과적인 전략에 사이버 보안 투자를 집중할 수 있는 체계적인 접근이 필요하다.

긍정적인 점은, 리스크가 새롭게 등장하고 복잡해지는 동시에, 사이버, 데이터, 엔지니어링, 리스크 관리의 미래를 형성할 수 있는 선도적 실천 방안들도 함께 진화하고 있다는 것이다. 딜로이트의 '생성형 AI 시대의 엔지니어링' 시리즈 마지막 편에서는 조직의 내부 및 외부 리스크를 평가하고 이에 대응하기 위한 전략 수립에 활용할 수 있는 프레임워크 구축에 대해 다룬다. 주요 전략은 다음과 같다:

- 새롭게 등장하는 리스크 범주를 반영한 사이버 보안 전략 재조정
- 입증된 사이버 보안 모범 사례의 확장
- 데이터 및 모델 출처 관리와 정보 보호 방식의 정립
- 생성형 AI 특화 보안 모델, 임직원 교육, 가이드라인 도입
- 시나리오 모델링을 통한 인프라 및 아키텍처 차원의 리스크·비용 예측

사이버 전략에 영향을 미치는 생성형 AI 리스크 4가지 요인 탐색

딜로이트의 '글로벌 사이버의 미래 조사'(Global Future of Cyber Survey) 4판에 따르면, 디렉터급 이상의 사이버 의사결정권자 약 1,200명을 대상으로 조사한 결과, 환각(hallucination)과 같은 무결성 리스크부터 사회공학적 공격, 미흡한 거버넌스 등 생성형 AI에 특화된 8가지 잠재적 리스크가 식별되었다. 이번 연구를 위한 원자료 분석 결과, 응답자의 77%가 이러한 리스크가 사이버 보안 전략에 미칠 영향에 대해 '매우 우려된다'고 응답해 모든 리스크에 대해 고르게 우려를 표했다.

이러한 위협들이 다양한 영역에 걸쳐 교차적으로 영향을 미친다는 점을 명확히 이해하기 위해, 생성형 AI 리스크는 다음 네 가지 범주로 구분할 수 있다.

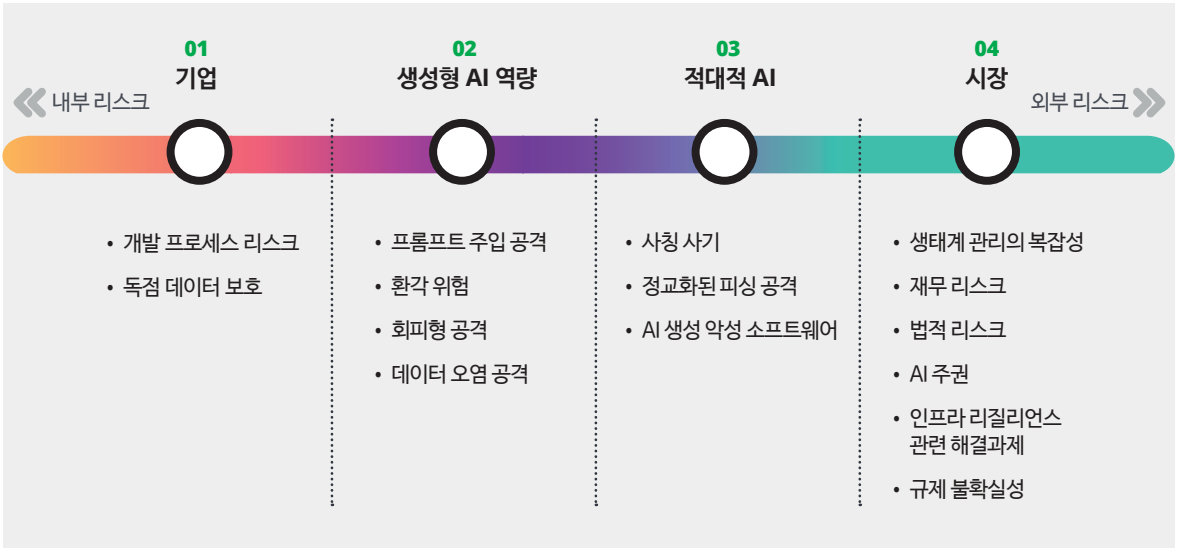
01. 기업 리스크: 조직 운영과 데이터에 대한 위협

02. 생성형 AI 역량 리스크: AI 시스템 오작동이나 악용 가능성과 같은 기술적 취약성

03. 적대적 AI 리스크: 악의적 행위자가 생성형 AI를 활용해 가하는 위협

04. 시장 리스크: AI 도입과 보안에 영향을 미치는 경제적, 법적, 경쟁적 압력

그림 1. 사이버 리더에게 영향을 미치는 생성형 AI 리스크의 4가지 범주 및 종류



출처: 딜로이트 분석

01. 기업에 대한 리스크

생성형 AI는 데이터, 애플리케이션, 인프라, 프로세스 전반에 걸쳐 조직의 리스크를 증가시키고 있다. 본 연구에서는 특히 다음과 같은 리스크들이 두드러지는 것으로 나타났다:

데이터 프라이버시, 보안, 지식재산(IP) 관련 리스크

생성형 AI 모델은 웹에서 수집된 텍스트, 이미지, 오디오, 다른 모델이 생성한 콘텐츠, 수작업으로 큐레이션된 자료 등 방대한 데이터를 기반으로 학습된다. 그러나 제3자 데이터나 모델은 종종 원본 출처, 창작자의 의도, 저작권, 데이터의 기본 속성을 인증하지 않기 때문에 출처 추적(provenance)이 어려우며, 이로 인해 환각(hallucination)이나 허위정보가 지속될 수 있다.

또한 생성형 AI가 예술, 음악, 문학, 소프트웨어, 발명품을 만들어내면서 저작권, 소유권, 보호 범위에 대한 문제가 부각되고 있다. 이에 따라 "누가 혁신의 주체이며, 누가 수익을 얻을 수 있는가"에 대한 규제가 아직 정립되지 않아 저작권에 대한 불확실성이 커지고 있다.

개인정보 보호 문제도 있다. 이름, 주소 등 개인 정보가 의도치 않게 수집되어 민감 정보, 영업기밀, 기밀 데이터를 노출하거나 악용할 위험이 존재한다. 예를 들어, 의료 기관이 자연어 기반 질의로 환자의 병력을 조회하는 AI 시스템을 도입할 경우, 의료진에게는 편의성을 제공하겠지만 개인 건강/의료 정보가 예기치 않게 노출될 수 있다.

개발 프로세스 내 보안 및 직원 리스크

생성형 AI가 소프트웨어 개발 프로세스에 도입되면서 새로운 보안 위협도 나타나고 있다. Palo Alto Networks의 보고서에 따르면, 보안 및 IT 리더들이 가장 우려하는 것은 AI가 생성한 코드였다. 딜로이트 리더들은 특히 기존 코드의 잘못된 구성(misconfiguration)이나 취약점을 AI가 증폭시킬 수 있다는 점을 우려하고 있다. 특히 서드파티 기반 모델에 대한 투명성이 부족해, 알 수 없는 보안 취약점이 유입될 수 있다는 점이 문제다.

또한, 직원들이 승인되지 않은 생성형 AI 솔루션을 개인 계정을 통해 업무에 사용하고 있는 점도 리스크로 지적된다. Cyberhaven Labs의 '2024 AI 도입 및 리스크 보고서'에 따르면, 주요 생성형 AI 툴의 업무 사용이 주로 개인 계정을 통해 이루어지고 있으며, 이로 인해 민감 정보가 외부로 유출될 가능성이 커지고 있다. 실제로 한국의 모 글로벌 전자 기업은 직원이 공공 프롬프트에 민감한 데이터를 입력해 유출하는 사고 이후, 직원의 생성형 AI 사용을 전면 금지했다.

기업 리스크 대응을 위한 새로운 접근법

이처럼 생성형 AI로 인한 리스크가 복잡하게 진화하고 있는 상황에서, 기업 리스크 관리자와 CISO(최고정보보안책임자)는 다음과 같은 조치를 고려할 수 있다:

1. 디지털 자산 관리 및 데이터 프라이버시 통제 도입

기업은 자사로 유입된 데이터의 출처뿐 아니라 외부로 송출되는 데이터의 흐름도 명확히 인지해야 한다.

- 디지털 여권(digital passport) 또는 모델 카드(model cards)* 같은 수단을 통해 모델의 학습 데이터, 히스토리, 출처를 검증할 수 있는 디지털 출처 추적(provenance) 기술이 필요하다.
- 데이터 진위성(authenticity) 검증, 동의 기반 데이터 수집(consent mechanism), 출처 기준 마련 등이 이를 가능하게 한다.
- 일부 조직은 'Do Not Train' 레지스트리처럼 데이터 창작자의 동의/거부 정보를 수집해 검색 가능한 데이터베이스로 구축하는 'AI 데이터의 동의 계층(consent layer)'을 개발 중이다.

보편적 기준이 없는 상황에서, 데이터 계보 추적(data lineage tracing)이 현실적 대안이 될 수 있다. 예를 들어, Data & Trust Alliance는 19개 글로벌 기업이 참여한 데이터 출처 표준(Data Provenance Standard) 을 통해 모범 사례를 도입 중이다.

2. 지식재산권(IP) 관리 전략 고도화

신뢰할 수 있는 콘텐츠 생산을 위해,

- 이미지, 영상, 폰트, 오디오 파일 등에 검증 가능한 콘텐츠 인증(Credentials)을 부여한다.
- 고급 디지털 권리 관리(DRM) 시스템을 통해 무단 복제 방지, 실시간 추적, 라이선스 적용, 배포 통제 등을 수행할 수 있다.
- 일부 기업은 머신 생성 콘텐츠를 식별하거나 재사용을 방지하기 위해 워터마크(watermark) 또는 은닉 신호(hidden signal)을 삽입하는 방식도 도입하고 있다.

3. 프롬프트 엔지니어링 시대에 맞는 DevSecOps (개발·보안·운영 통합) 재구성

딜로이트의 엔지니어링 연구에 따르면, 생성형 AI는 소프트웨어 개발 방식 자체를 바꾸고 있으며, 구성원 간 역할 통합을 유도하고 있다.

- 사이버 및 리스크 리더는 중앙집중형 거버넌스 체계를 수립한다.
- AI 사용부터 네트워크 모니터링까지 명확한 정책과 절차를 마련한다.
- 이를 조직 전반에 체계적으로 확산 및 내재화해야 한다.

*디지털 여권: 사용자나 디지털 시스템의 신원, 자격, 권한 등을 증명하는 디지털 문서 또는 인증 체계 // 모델 카드: AI 모델의 성능, 한계, 사용 목적, 데이터 편향 여부, 윤리적 고려 사항 등을 문서화한 설명서

02. 생성형 AI 역량 리스크

생성형 AI는 생성형 AI 솔루션이 의존하는 데이터와 모델을 겨냥한 새로운 보안 리스크도 초래한다. 떠오르는 위협은 다음과 같다.

프롬프트 인젝션 공격. 생성형 AI 솔루션의 특징 중 하나는 AI에 제공되는 프롬프트나 명령어의 사용이다. 프롬프트 인젝션은 공격자가 생성형 AI 시스템을 속여 보안 데이터를 노출시키거나, 잘못된 정보를 퍼뜨리거나, 악의적 행동을 수행하게 하거나 숨겨진 트리거를 통해 모델에 백도어로 접근하도록 설계된 프롬프트를 사용하는 신종 공격 기법이다. 프롬프트 인젝션은 Open Worldwide Application Security Project(OWASP) Top 10에서 대형 언어 모델(LLM) 애플리케이션과 관련된 주요 보안 위협으로 꼽힌다.

회피 공격

전통적인 AI 시스템에서 회피 공격은 모델이 의도적으로 혼란스러운 샘플('적대적 예시')에 속아 잘못된 출력을 내도록 유도하는 공격이다. 생성형 AI 시스템은 전통적 AI보다 더 큰 규모로 이러한 공격에 취약하다. 기본 쿼리 권한만 있어도 누구나 프롬프트를 이용해 모델의 예측 레이블과 신뢰도 점수를 탐색하며 모델이 다른 결정을 내리도록 조작할 수 있다. 이러한 공격은 침입 탐지 시스템을 우회하는 데 사용될 수 있다.

데이터 중독

공개 데이터를 학습한 외부 모델은 여러 불확실성을 내포한다. 데이터 중독은 적대자가 모델의 학습 데이터셋을 변조하는 의도적 공격이다. 데이터 중독 기법에는 데이터를 부정확하게 변조하고 오도하도록 조작하는 것이 포함된다. 검색 보강 생성(RAG) 시스템이 늘어남에 따라 데이터 중독 위험도 커지고 있다.

생성형 AI 모델의 환각

생성형 AI 모델은 학습 데이터 패턴에 기반해 출력을 예측하지만, 환각이 발생하면 결과가 그럴듯해 보이나 사실과 다를 수 있다. 이러한 부정확성은 잘못된 의사결정, 평판 손상, 규제 처벌, 기회 상실을 초래할 수 있다. 환각과 마찬가지로 생성형 AI 모델을 통해 잘못된 정보가 유포될 수 있으며, 이는 신뢰 상실, 재정 손실, 비즈니스 결정에 부정적 영향을 미칠 수 있다.

데이터, 모델 및 애플리케이션 보안에 대한 새로운 접근법

최고정보보호책임자(CISO)와 최고기술책임자(CTO)는 이러한 증가하는 위험과 공격 표면을 고려해야 한다. 이는 다양한 전략과 조치를 포괄하는 폭넓은 사이버보안 접근법을 포함해야 하며, 다음과 같은 방법들이 있다.

- **입력 가드레일과 모델 방화벽을 통한 프롬프트 인젝션 대응:** 첫째, 사용자 입력을 전처리하여 유해 콘텐츠를 제거하거나 무력화하는 입력 정화 및 검증을 고려해야 한다. 이는 필터링, 변환 등을 통해 입력이 예상 패턴을 따르고 악의적 명령이 없음을 확인하는 작업을 포함한다. 기밀 데이터는 LLM 애플리케이션에 연결되기 전에 철저히 검증하거나 마스킹해야 하며, 최소 권한 원칙 적용과 악성 코드 실행 방지를 위한 파라미터화 같은 표준 보안 관행을 준수해 프롬프트 인젝션 위험을 줄여야 한다. 둘째, 규칙 기반 솔루션에만 의존하지 않고, AI 방화벽을 구현해 모델에 들어오고 나가는 데이터를 모니터링하여 위협 탐지 역량을 강화할 수 있다. 프롬프트 쉴드 같은 도구가 숨겨진 명령을 탐지하는 데 도움을 주지만, 생성된 출력물을 판단하고 승인하는 데는 인간의 감독이 여전히 중요하다.

- **정확도 향상을 위한 파인튜닝:** 파인튜닝은 모델의 콘텐츠 생성 정확도를 높이기 위해 데이터를 활용해 감독 학습 방식으로 모델을 재학습하는 과정이다. 권위 있는 외부 출처를 통한 검색 보강 생성, 상황별 가드레일 구현, 프롬프트 라이브러리 생성, 인간 감독 등은 환각을 줄이는 데 도움이 된다.

- **생성형 AI를 사이버 역량에 통합:** 딜로이트의 '글로벌 사이버의 미래 조사' 4차 조사에 따르면, 응답 조직의 36%가 AI 및 생성형 AI를 조직의 사이버보안 예산에 포함시켰다. 생성형 AI는 전통적인 보안 도구로 놓칠 수 있는 네트워크 트래픽과 시스템 로그 속의 미묘한 위협 징후를 식별할 수 있다. 생성형 AI는 합법적인 이메일 패턴을 분석해 보다 복잡한 피싱 시도를 탐지할 수 있다. 예를 들어, NVIDIA는 기존 방법에 비해 21% 더 높은 정확도로 피싱 탐지 솔루션 개발 시간을 줄이는 스피어 피싱 탐지 AI 워크플로를 도입했다. AI 기반 딥페이크 탐지 도구 또한 텍스트, 오디오, 이미지, 또는 여러 형식 복합체 등에서 대규모로 허위 정보 및 기계 생성 콘텐츠를 인간보다 더 정확히 식별할 수 있다.

미국 국립표준기술연구소(NIST)는 데이터 모델에 대해 적대적 훈련을 권고하며, 이를 통해 사이버 공격을 모방하고 보안 취약점을 식별해 방어를 강화할 수 있다. 적대적 생성 신경망(GAN)* 과 스마트 취약점 탐지 도구는 전통적 스캐너가 놓칠 수 있는 보안 결함을 발견할 수 있다.

*적대적 생성 신경망: 두 개의 신경망이 서로 '적대적'(경쟁적)으로 학습하면서 점점 더 정교한 결과물을 만들어내는 딥러닝 모델

03. 적대적 AI 리스크

생성형 AI 기술은 사이버 공격을 조직하는 데 필요한 기술을 상품화하여 악의적 행위자의 진입 장벽을 낮춘다. 딜로이트의 제4차 글로벌 사이버 미래 조사에 따르면 조직의 약 3분의 1이 피싱, 악성코드, 랜섬웨어(34%) 및 데이터 손실 관련 위협(28%)에 대해 우려하고 있다. 생성형 AI는 적대적 공격의 정교함, 규모, 그리고 용이성을 증가시킬 수 있다. 이러한 위험 유형에는 다음이 포함된다.

AI 생성 악성코드

생성형 AI는 디지털 생태계에 더 많은 복잡성을 초래하며 공격자가 악성코드 및 랜섬웨어를 자동화하는 것을 쉽게 하여 알려진 위협의 규모와 정교함을 모두 증가시킨다. 예를 들어, 해커들은 지속적으로 새로운 악성코드를 생성하며, 사이버 범죄자들은 생성형 AI를 활용해 사실상 무한대에 가까운 정교한 악성코드를 생산할 수 있어 기존의 사이버보안 조치와 대응 시간을 압도할 수 있다. 한 연구에서는 사용자 기기에 불안전한 AI 코드를 주입할 수 있는 약 100개의 머신러닝 모델이 발견되었다.

사람처럼 보이는 피싱 공격

생성형 AI는 피싱 공격에 혁신을 가져왔다. IBM의 2024 X-Force 위협 인텔리전스 인덱스에 따르면, 2023년 다크웹 게시물에서 'AI'와 'GPT'가 80만 건 이상 언급되었다. 공격자는 기술을 이용해 믿을 만한 메시지를 만들고 대규모의 정교한 사회공학 캠페인을 자동화할 수 있다. 생성형 AI 도구는 설득력 있고 상황에 맞는 메시지, 텍스트 및 오디오 메시지를 작성하는 능력을 상품화하여 언어 장벽을 극복하고 문화적 뉘앙스까지 통합할 수 있다. 더 나아가 AI의 자기 개선 능력은 학습한 지식을 기반으로 자율적인 피싱 전술을 개발하는 것도 가능하게 한다.

사칭 공격

생성형 AI 도구를 통해 가짜 음성 및 영상을 만드는 능력은 지난 3년간 성숙해진 신흥 위험으로, 사칭 사기의 장벽을 크게 낮추었다. 예를 들어, 금융 분야에서 생성형 AI는 딥페이크를 이용한 금융 사기의 위험을 크게 증가시킬 것으로 예상된다. 딜로이트 금융 서비스 센터는 2027년까지 미국 내 생성형 AI 관련 사기 손실이 2023년 123억 달러에서 400억 달러로 성장할 것으로 전망하며, 연평균 성장률은 32%에 달할 것으로 보고 있다.

적대적 AI에 대응하는 새로운 접근법

최고정보보호책임자(CISO)는 다음과 같은 전략으로 적대적 AI 위험에 대응할 수 있다.

- **기존 위협 탐지 및 대응 역량 확장:** CISO들은 이러한 위협에 앞서기 위해 기존 AI 및 딥러닝 기법을 확장할 기회를 갖는다. 예를 들어 Mastercard는 거래에 관련된 여러 주체 간의 연결을 평가해 위험도를 판단하는 Decision Intelligence Pro 시스템을 출시할 예정이다. 이 기술은 사기 탐지율을 평균 20% 증가시키며, 특정 경우 최대 300%까지 증가시킬 수 있다. 유사하게 VISA도 각 거래의 500개 이상의 속성을 AI와 머신러닝으로 분석해 사기를 방지한다. Microsoft는 자사 생태계 모니터링 시 AI를 활용해 하루 65조 개 이상의 사이버보안 신호를 탐지한다고 밝혔다.

- **적대적 훈련 프로그램 업데이트:** 훈련 프로그램은 공격자가 사용하는 생성형 AI 특유의 위협과 직원들이 정교한 피싱 시도를 방지하기 위해 디지털 상호작용에 있어 경계심을 가져야 하는 점을 포함하도록 업데이트해야 한다. CISO는 보안 행동 및 문화 프로그램을 추진해 인간 관련 사이버보안 위험을 줄이는 데 기여해야 하며, 단순한 인식 제고를 넘어 행동 변화를 유도하는 데 목표를 두어야 한다.

04. 시장 리스크

생성형 AI의 시장 리스크는 아직 구체화되는 단계에 있으며, 많은 부분이 조직의 통제를 벗어난 요인들로 구성되어 있다. 여기에는 규제 리스크, 인프라 회복력, 제3자 리스크 등이 포함되며, 주요 리스크는 다음과 같다.

규제의 불확실성

딜로이트의 2023년 4분기 '기업의 생성형 AI 사용현황 보고서'에 따르면, 조직들이 가장 크게 우려하는 사항은 규제 준수 여부이다. 이 규제는 데이터 활용, 보안, 프라이버시 등 생성형 AI 리스크 관리에 필수적인 요소들을 포괄하며, 글로벌 및 지역 단위에서 적용된다.

예를 들어, 미국의 2023년 행정명령 14110은 고급 AI 모델이나 컴퓨팅 클러스터를 개발, 확보, 보유하는 기업들에게 보고 의무를 부과하였다. 하지만 2025년 들어 새 행정부는 해당 명령을 철회하고, 기존 AI 정책들을 검토해 AI 혁신을 저해할 수 있는 규제는 폐지하도록 지시하였다. 이로 인해 기존 보고 의무는 철회되었으며, 조직은 변화하는 규제 환경이 생성형 AI 제공 방식에 어떤 영향을 미칠지 주의 깊게 살펴야 한다.

생성형 AI 확산에 따른 컴퓨팅 인프라 리스크

생성형 AI는 막대한 컴퓨팅 자원을 필요로 하며, 이미 압박을 받고 있는 전력망에 부담을 가중시키고 있다. 공공 유틸리티는 민첩성이 떨어지고, 향후 에너지 수요를 정확히 예측하기도 어렵다. 딜로이트의 2025년 전력 및 유틸리티 산업 설문조사에 따르면, 데이터센터에 안정적으로 전력을 공급하는 데 있어 전력망 인프라의 한계가 가장 큰 과제로 나타났다.

딜로이트의 분석에 따르면, 현재 데이터센터는 미국 연간 전력 생산량의 68%를 소비하고 있으며, 2030년까지 이 수치가 11~15%까지 증가할 것으로 전망됐다.

새로운 가치 사슬의 등장

데이터센터는 그 가치 사슬 전반에 걸쳐 새로운 불확실성과 이에 따른 리스크, 기회에 직면하고 있다. 투자회사, 부동산 기업, 엔지니어링·건설 회사들은 신규 데이터센터를 위한 허가, 부지, 자금 확보에 힘쓰고 있으며, 클라우드 하이퍼스케일러, 통신사, 기술 인프라 공급업체들은 증가하는 컴퓨팅 수요를 관리하려 하고 있다.

미국의 북버지니아, 콜럼버스, 피츠버그 등 일부 지역에서는 전력 공급 부족으로 인해 신규 프로젝트 승인이 지연되고 있다. 영국과 유럽에서도 전력 부족으로 데이터센터 건설이 차질을 빚고 있다. 또한, 공급망 병목 현상으로 인해 주요 부품과 자재의 부족이 발생하면서 프로젝트 지연과 비용 상승을 초래하고 있다. 예컨대, 발전기, 무정전 전원장치(UPS) 배터리, 변압기, 서버, 건축 자재 등의 확보에 어려움을 겪고 있는 운영업체들이 많으며, 이로 인해 대체 가능한 자재를 사용하는 경우가 많아지고 있다.

벤더 종속으로 인한 애플리케이션 유연성 부족

생성형 AI 모델과 인프라는 조직이 따라가기 어려운 속도로 발전하고 있으며, 이로 인해 조직은 낡은 기술이나 중복된 기능에 비용을 지출하거나 향후 기술과 상호운용되지 않는 벤더와 계약할 위험에 노출되어 있다.

많은 조직들이 고급 하드웨어를 확보하기 위해 서두르고 있으나, 공급업체는 수요를 따라잡지 못하고 있다. NVIDIA의 Blackwell GPU는 향후 1년치가 이미 매진되었으며, 단일 벤더에 의존하는 조직은 이 경쟁적인 시장에서 주요 하드웨어 진보를 놓칠 수 있다.

가치 실현에 대한 우려

대규모 모델을 훈련시키고 운영하는 데 필요한 초기 투자와 컴퓨팅 하드웨어는 상당한 비용이 든다. 딜로이트의 2023년 4분기 설문조사에 따르면, 응답자의 약 3분의 1은 기대했던 가치를 달성하지 못할 경우 향후 2년 동안 생성형 AI의 시장 채택 속도가 둔화될 수 있다고 응답했다.

시장 리스크에 대한 새로운 대응 방식

조직이 공급 측 또는 수요 측, 혹은 양쪽 모두로부터 영향을 받을 경우, 다음과 같은 대응책을 검토할 수 있다:

- **소형 언어 모델(SLM)과 워크로드 축소로 컴퓨팅 수요 절감:** 소형 언어 모델(SLM)은 훈련 비용은 클 수 있지만 GPU 요구를 줄일 수 있어 장기적으로 비용 절감 효과가 있다. 예를 들어 Salesforce는 70억 파라미터의 XGen-7B를 출시했으나, 비용 절감을 위해 특정 작업에는 효율이 높은 소형 모델을 사용하고 있다. SLM은 휴대폰이나 노트북 등 장치 내에서 실행 가능하여 에너지 소비도 줄일 수 있다. Google과 NVIDIA도 저지연 처리를 위해 장치에 직접 배포 가능한 소형 모델을 출시하였다. 대형 모델을 TPU 64개로 며칠간 훈련해야 하는 것과 달리, 소규모 데이터셋은 몇 시간 내 처리 가능하여 에너지 소비를 크게 줄일 수 있다. IBM은 모델이 더 긴 문맥을 관리할 수 있도록 연관 기억 모듈을 도입하여 재훈련 시간도 줄이고 있다.

- **전략적이고 효율적인 인프라 의사결정:** 일부 조직은 온프레미스 데이터센터와 하드웨어에 수백만 달러를 투자하고 있으나, 이는 과거 클라우드 도입을 위해 무너뜨렸던 경계를 다시 쌓는 것이기도 하다. 클라우드, 엣지, 온프레미스 환경 간 이동을 위해 고성능 스위칭 장비가 필요하며, 이 시장은 2027년까지 10억 달러 규모로 성장할 것으로 보인다. 예를 들어, 생성형 AI 기반 번역 및 더빙 서비스를 제공하는 한 기업은 GPU 4개로 구성된 클러스터를 구매했으며, 클라우드 대비 60~70% 비용을 절감할 수 있을 것으로 예상했다. 하지만 전력 및 냉각 요구를 고려하지 않아, 결국 단순한 워크로드는 온프레미스에서 처리하고, 고성능 컴퓨팅은 클라우드에서 처리하는 하이브리드 방식으로 전환하였다.

- **에너지 효율적인 하드웨어 및 워크로드 균형 유지:** NVIDIA는 대규모 언어모델 추론 작업을 처리하면서도 에너지 소비를 크게 줄일 수 있도록 설계된 Blackwell GPU 플랫폼을 선보였으며, IBM은 기존 GPU/CPU보다 25배 에너지 효율이 높은 NorthPole 칩을 발표하였다. 또한 데이터센터에서 전력 소비의 최대 40%를 차지하는 냉각 요구를 효율적으로 관리하기 위해, 열 발생 부위에 직접 냉각제를 순환시키는 액체 냉각 시스템이 주목받고 있다.

- **전력 소비 및 전력망 부담 관리:** Flexential의 2024년 AI 인프라 보고서에 따르면, 기업의 24%만이 온프레미스 AI 하드웨어를 사용하고 있으며, 51%는 엣지 처리를 위해 코로케이션 센터에 서버 공간을 임대하고 있다. 일부 기업은 재생에너지를 활용한 마이크로그리드로 회복성과 지속 가능성을 강화하고 있으며, Google은 2025년 완공 예정인 애리조나 메사 데이터센터에서 태양광, 풍력, 배터리 저장 시스템으로 400MW 이상의 청정 에너지를 사용할 계획이다. 또한 독일의 한 기업은 풍력 터빈 내부에 데이터센터를 구축하여 청정 에너지를 확보하고 있다.

- **소형 원자로 및 원자력 에너지의 활용:** 도로, 철도, 해상 운송이 가능한 마이크로 원자로가 수요 대응형 에너지 공급 수단으로 떠오르고 있다. Amazon은 펜실베이니아의 Susquehanna 원자력 발전소 인근 부지를 인수하였으며, Microsoft는 Three Mile Island 원전의 전력을 향후 20년간 전량 구매하기로 계약하였다. 이러한 접근 방식의 지속 가능성은 국가 에너지 정책 및 인프라 결정에 영향을 받을 수 있다.

- **AI 팩토리로 데이터센터 재정의:** 2023년 5월~7월 사이, 미국 내에서 2.1GW 규모의 신규 데이터센터 임대 계약이 체결되었으며, Amazon은 데이터센터 두 곳에 총 100억 달러를 투자하고 있다. 15개의 글로벌 통신사는 12개국에 걸쳐 생성형 AI 전용 데이터센터, 즉 AI 팩토리를 구축 중이다. 2024년부터 미국에서는 데이터센터가 핵심 인프라로 지정되어, 공공·민간 컴퓨팅 수요 충족 여부를 감시하는 전초기지 역할을 하게 될 수 있다.

- **엣지 컴퓨팅 및 저궤도 위성을 통한 용량 분산:** 엣지 데이터센터는 특정 지역 가까이에 위치한 소형 센터로, 지연을 줄이고 실시간 데이터 처리가 필요한 애플리케이션 성능을 개선한다. 이러한 구조는 전력망의 부하를 분산하고, 네트워크 회복성도 강화할 수 있다. 딜로이트에 따르면, 향후 저궤도(LEO) 기반 데이터센터는 도시 전력망에 대한 경쟁에서 벗어날 수 있는 대안이 될 수 있다. 일본 NTT는 광학 기술을 이용해 데이터를 위성 네트워크로 전송하고, 필수 정보만 사용자에게 전달하는 고속 네트워크 시스템을 구상하고 있다.

- **신뢰성과 거버넌스를 기반으로 한 생성형 AI 도입:** 딜로이트의 '신뢰할 수 있는 AI'(Trustworthy AI) 프레임워크는 전략, 표준, 제3자 리스크 관리, 기술 구현을 포함한 포괄적 접근을 통해 AI 프로그램을 효율적이고 투명하게 관리할 수 있도록 지원한다. 조직은 생성형 AI 도입 시 전체 스택 접근 방식을 고려함으로써 기술 구성요소 간 연계성과 노후화 리스크를 관리하고, 생태계 전반의 위험 노출을 파악하여 필요한 조치를 취할 수 있다.

리스크 관리와 회복탄력성 강화

생성형 AI는 기업 내부와 외부에 걸쳐 다양한 리스크를 초래하며, 이는 본 문서에서 논의한 네 가지 범주—기업 자체, 생성형 AI 역량, 적대적 AI로부터의 위협, 그리고 시장으로부터의 리스크—전반에 걸쳐 발생한다. 이러한 리스크는 다차원적이며 상호 얽혀 있어, 업종이나 전략에 관계없이 모든 조직은 자사의 생성형 AI 통합 방식을 보호하기 위한 최적의 방안을 평가해야 한다. 이에 따라 리더는 사이버 보안 전략과 비즈니스 회복탄력성 전략을 수립하고 다른 기능과 조율하는 데 있어 핵심적인 역할을 수행해야 한다.

CISO를 포함한 리스크 리더십은 조직이 생성형 AI로 인해 노출되는 위험을 이해하고, 기존의 방식과 새로운 방식을 결합하여 전방위적으로 조직을 보호할 수 있도록 지원할 수 있는 위치에 있다.

하나의 솔루션으로 모든 리스크를 해결할 수는 없으며, 조직은 자사의 리스크 노출 수준에 따라 다양한 솔루션을 정렬하고 맞춤화해야 한다. CISO는 또한, 고객의 프라이버시를 보호하기 위해 소프트웨어 개발 생애주기 전반에 걸쳐 '인간 개입(Human-in-the-loop)' 및 '설계 단계부터 보안 내재화(Secure-by-design)' 접근법이 적용되고 있는지를 확인해야 한다.

조직은 기존의 보안 관행을 기반으로 하여 진화한 기존 리스크뿐만 아니라 생성형 AI 모델의 특수한 리스크, 빠르게 변화하는 인프라 수요를 모두 수용할 수 있도록 전략을 재정립함으로써 비즈니스 회복탄력성을 확보해야 한다.

한국 딜로이트 그룹 산업 전문가

사이버 보안 및 리스크

한국 딜로이트 그룹은 사이버 리스크 대응을 위한 정보보호 및 개인정보보호 자문, 정보보안 인증, 기술적 취약점 진단 및 대책 수립, 정보보호 전략 수립, Cyber Incident 대응 등의 서비스를 제공하고 있습니다. 또한, 수많은 유형의 사이버 리스크를 사전에 방지해 기업 운영의 든든한 조력자 역할을 수행합니다. 리스크 최소화를 통한 안정적인 기업 경영, 딜로이트가 함께 합니다.

백철호 파트너
One Cyber & Resilience 리더
02 6676 2250
cbaek@deloitte.com

서영수 파트너
One Cyber & Resilience
02 6676 1929
youngseo@deloitte.com

유선희 파트너
One Cyber & Resilience
02 6676 2956
sunhyou@deloitte.com

이상훈 수석위원
One Cyber & Resilience
02 6676 2937
sanghunlee@deloitte.com

문범석 파트너
One Cyber & Resilience
02 6676 2949
bsmoon@deloitte.com

이창성 파트너
One Cyber & Resilience
02 6099 4888
changsulee@deloitte.com

Cyber Service Offerings	주요 서비스
Cyber Governance & Compliance	• 정보보안 인증 및 자문 (ISO27001, ISMS-P, SOC 2/3, WebTrust 등) • 전자서명인증 평가
Privacy Governance & Compliance Data Protection & Compliance	• 정보보호/개인정보보호 자문 • 수탁사 진단 • 상시 보안 자문
Application Security Security Architecture Cyber Cloud	• 개발 보안(시큐어코딩 등) • 시나리오 기반 모의해킹 • 인프라 취약점 진단 • 클라우드 보안 진단
Cyber Strategy Emerging Technologies	• 정보보호 전략 수립 • OT 보안 전략 수립 • 중장기 마스터플랜 수립
Technology Resilience Crisis & Incident Response	• Crisis Management • BCM • 사이버 침해사고 대응

앱

Download on the App Store

GET IT ON Google Play

카카오톡 채널

'딜로이트 인사이트' 앱과 카카오톡 채널에서
경영·산업 트렌드를 만나보세요!

Deloitte. Insights

성장전략부문 대표
손재호 Partner
jaehoson@deloitte.com

딜로이트 인사이트 편집장
박경은 Director
kyungepark@deloitte.com

Contact us
krinsightsend@deloitte.com

연구원
양원석 Senior Consultant
wonsukyang@deloitte.com

디자이너
박근령 Senior Consultant
keunrpark@deloitte.com