Gartner Business Quarterly

2024년 3분기

- 왜 모든 리더들은 '진실 없는 세상'에 대한 고민을 해야 하는가
- 딥페이크를 감지하여 가장과 허위사실로 인한 피해를 방지해야 한다
- 고위경영진 간 조종 지향적 소통을 방지하기 위한 전술

Gartner

진실 없는 세상 헤쳐나가기

진실 없는 세상 헤쳐나가기

편집자 주

전세계적으로 중요 선거들이 치루어지는 올해, 악의적 의도를 가진 허위정보는 중대한 정치적 이슈이고, 이는 기업 고위 경영진에게도 고민거리가 되고 있다. 기업 임원들은 인공지능을 사용한 악의적 공격을 2024년 상반기 가장 두드러진 위협요소로 꼽았다. 가장과 허위정보의 급증은 사이버 보안 및 운영 관련 리스크도 발생시킨다. 또한, 이는 기업과 소비자 간, 고용주와 직원 간, 심지어 고위 경영진 간의 신뢰에도 타격을 줄 수 있다. 이는 이미 64%의 CEO들이 자신들에 대한 사회적 신뢰도가 낮아지고 있다는 인식을 가진 지금의 시기에 상황을 악화시키는 요소로 작용할 수 있다.

IT 리더와 부서는 조직을 외부의 악의적 공격자들로부터 보호하는데 있어 핵심적 역할을 해야 한다. 허위정보는 조직에 대해 막대한 재무적, 브랜드 가치적 리스크를 발생시킬 수 있고, 그렇기 때문에 조직의 모든 고위 리더들은 이를 전사적 우선순위 이슈로서 고민하고 대응해야 한다.

본 가트너 비즈니스 분기 보고서(*Gartner Business Quarterly*)는 거버넌스, 기술, 내부 인식 강화가 이러한 문제들을 다루는데 어떻게 도움이 되는지를 논의한다. 이를 통해 인공지능 및 지속가능성에 대한 기술 벤더들의 제안을 객관적으로 평가하고 딥페이크를 감지하며 생성형 인공지능 관련 리스크로부터 조직의 브랜드를 보호하는 것에 대한 조언을 제공한다.

신뢰 강화의 시작을 돕기 위해, 가트너는 고객, 직원, 비즈니스 파트너, 일반 대중이 기술의 불완벽성을 이해하고 받아들일 수 있도록 하는 아이디어들을 제공한다. 또한, 신뢰 문제를 다루는 것이 어떻게 보다 견고한 고위 리더들의 의사결정, 조직 문화, 지속가능성 노력을 강화하는데 도움이 되는지를 보여준다.

가트너 비즈니스 분기 보고서는 리더와 팀이 다른 조직 구성원들과 효과적으로 협업하는 것을 지원하여 조직이 목표를 달성하고, 보다 과감해지고, 원칙을 기반으로 행동하고, 직원, 투자자, 소비자들을 그 여정에 참여시키는데 도움을 줄 것이다.

기존 가트너 리서치 서비스는 최신 정보와 인사이트를 제공한다. 커팅엣지(Cutting Edge)는 참신한 데이터에 대한 관점이다. 브리프(Brief) 자료는 보다 현명한 투자, 계획, 인재, 문화, 성장, 혁신, 데이터, 기술 등에 대한 간결한 정보들이다.

가트너는 고객사의 피드백을 환영한다.

이메일: richard.eames@gartner.com.

— 리처드 임스(Richard Eames)

왜 모든 리더들은 '진실 없는 세상'에 대한 고민을 해야 하는가

데이브 아론(Dave Aron), 앤드류 프랭크(Andrew Frank), 댄 에이욥(Dan Ayoub)

고도로 목표화된 허위 콘텐츠는 갈수록 큰 위협이 되고 있다. 조직의 고위 리더들은 이 문제에 대한 명확한 책임 체계를 설정하고, 기술 기반 대책을 마련하고, 인공지능 관련 위협에 대한 워크숍 등을 통해 조직 내부의 인식을 강화해야 한다.

Table of Contents

허위 콘텐츠의 급증은 조직에 악영향을 미칠 수 있다. 리스크 관리 담당 리더들은 올해 1, 2분기 중 인공지능 기반 사이버 공격을 가장 두드러진 위협으로 꼽았다.[1,2] 64%의 CEO들이 자신들에 대한 사회적 신뢰도가 낮아지고 있다고 보는 지금 시기에 이는 기업 및 정부 리더들에 대한 대중의 회의감이 더욱 커지게 하는 요소가 될 수 있다.[3] 구체적 목표를 대상으로 한 공격은 기업의 브랜드 평판이나 주가에도 타격을 줄 수 있다.

조직의 고위 리더들은 긴급하게 거버넌스 체계를 구축하고 기술을 활용하여 이러한 위험 요소에 대응해야 한다. 또한, 인공지능 관련 사고 대응 워크숍을 실시하여 인공지능 활용 시 발생할 수 있는 우발적 실수나 외부로부터의 의도적 공격에 대한 조직 내 인식을 강화해야 한다.

잘못된 정보의 만연을 부추기는 세 가지 요소

잘못된 정보는 인류가 사회를 이루어 온 시간 만큼이나 오랫동안 인류 문화의 일부로 존재해왔지만, 다음의 세 가지 요소가 결합되어 그 영향력과 위험 정도가 더욱 강해지고 있다.

1. 낮은 비용으로 다수를 대상으로 소통이 이루어진다
인터넷, 소셜 미디어, 모바일 앱 등으로 인해 멀티미디어 콘텐츠가 매우 낮은 비용으로 많은 사람들에게 즉각적으로 전달될 수 있게 되었다. 또한, 이러한 도구들을 통해 메시지를 맞춤화하여 대상으로 하는 수신자, 고객, 시민 그룹에 따라 각기 다른 무언가를 제공할 수 있다.

2. 진짜 같은 허위 콘텐츠를 쉽게 만들 수 있게 되었다
생성형 인공지능을 이용하여 딥페이크 이미지, 오디오, 비디오 콘텐츠는 물론 진짜 같은 대화 내용도 만들어 낼 수 있게 되었다.

3. 소통의 영향력과 파급효과가 커졌다
행동 과학은 빅 데이터, 애널리틱스, 인공지능과 결합되어 개인의 심적 모델, 의사결정, 행동 변화를 목적으로 하는 설득력 있는 메시지를 만들어 낼 수 있다.

악의적 공격자들은 이러한 요소들을 활용하여 '악성 정보('malinformation': malware와 disinformation의 합성어)'을 만들어 낸다.[4] 관련 기술이 적용되어 고도로 목표화된 악성 정보 유포는 유권자들의 투표에도 영향을 미치는 등 사회와 정치 영역에서 이미 목격되고 있는 현상이다.[5] 이는 특히 전세계 인구 절반에 해당하는 40억 명이 거주하는 국가들에서 여러 선거가 치루어지는 204년에 특히 주목해야 할 이슈이다.[6]

또한, 주요 기업들을 대상으로 한 허위 콘텐츠의 증가 역시 두드러진 동향이다. 이러한 위협은 다음을 포함한다.

- **기업 및 브랜드에 대한 직접적 딥페이크 공격**
  영국에 본사를 둔 디자인 및 엔지니어링 컨설팅 기업인 Arup의 경우, 화상 회의에서 딥페이크를 이용하여 CFO를 가장하여 직원에게 자금이체를 하도록 한 사건이 발생하여 피해액이 2,500만 달러에 달했다.[7]
- **제약, 환경 과학 등 해당 산업 전체에 대한 대중의 신뢰를 훼손시키는 간접적 공격**
  세계경제포럼의 회의에서 화이자(Pfizer)의 CEO가 세계 인구를 절반으로 줄이는 계획을 지지하는

것으로 교묘하게 편집된 동영상이 등장한 사례가 있다. 그가 실제로 한 말은 화이자의 목표가 '화이자의 의약품을 사지 못하는 사람의 수를 절반으로 줄이는 것'이었다. 편집된 동영상은 백신에 대한 반감과 음모론 확산을 목적으로 BitChute라는 영국의 웹사이트에 게재되었다. 이는 이후 X(구 트위터), 페이스북, 갭(Gab) 등의 소셜 미디어 플랫폼들에 공유되었으나 이후 악의적 허위 콘텐츠임이 밝혀졌다.[8]

갈수록 정교해지는 합성 현실, 행동 과학, 디지털 미디어 관련 기술로 인해 악성 정보의 위협은 계속 커질 것이다. 더욱이 향후 더욱 강력한 기술이 등장함에 따라 이 문제는 한층 더 심각해질 수 있다. 예를 들어, 양자 컴퓨팅은 전례 없는 정보 처리 능력을 가져와 보다 고도화된 공격의 동력이 될 수 있다. 가상 현실 기술 역시 더욱 그럴듯하고 설득력 있는 콘텐츠를 만드는데 역할을 할 것이다. 자율주행차량과 침습형 기술들은 더 많은 공격 대상을 노출시키고 취약성을 키우게 될 것이다.

가장 우선적인 것은 조직의 방어 체계를 구축하고 관리하는 것이다

완벽한 물리적 보안 및 사이버보안을 갖춘 조직이라도 악성 정보 공격에는 취약할 수 있다. 따라서 조직은 다음에 대한 명확한 책임 소재를 설정해야 한다.

- 위협에 대한 이해 및 관리
- 방어 및 대응 체계 구축
- 리스크 및 보호장치에 대한 지식의 조직 내 전파

이때 정보보안담당임원(CISO)이 핵심적 역할을 하게 될 것이다. 가트너는 2027년까지 확장된 공격 대상과 규제 압박으로 인해 45%의 CISO들이 악성 정보를 자신의 직무 이슈로 다루게 될 것으로 본다. 기업 리스크 위원회 역시 이에 본격적으로 관여해야 할 것이다.

기술은 영웅도, 악당도 될 수 있다

기술의 발전으로 인해, 악의적 공격자들은 정교한 허위 정보(disinformation)를 넘어 고도로 목표화된 악성 정보(malinformation)를 만들어 내고 이용할 수 있게 되었다. 하지만 이에 대응하는 입장에서도 기술을 활용하여 'TrustOps'라는 전략을 통해 이에 맞설 수 있다.

첫째, 자신의 조직이 제공하는 콘텐츠를 인증하여 다른 이들이 가짜로 잘못 인식하거나 교묘하게 조작하지 못하도록 하는 방안을 도입하도록 한다. 이를 통해 인증을 거치지 않은 허위 콘텐츠의 소유권을 부정할 근거를 확보할 수 있다. 이는 또한 검색 엔진, 소셜 미디어, 광고 플랫폼 등이 검증된 콘텐츠만을 취급하는데 사용할 수 있는 도구가 될 수도 있다. 콘텐츠 제작자와 사용자는 C2PA의 콘텐츠 크레덴셜(Content Credentials) 같은 인증을 위한 공개 표준을 적용하여 소유권, 진실성, 출처 확인 등을 위한 명확한 검증 경로를 구축할 수 있다.[9] 조직들은 이러한 기술들에 대한 폭넓은 지지를 확보 위해 관련 정책을 빠르게 설정하는 것을 고려해야 한다.

둘째, 악성 정보에 대한 조직 내 인식을 빠르게 강화한다**.** 소셜 미디어 모니터링 및 브랜드 보호 도구를 활용하고**,** 이를 진화시켜 딥페이크와 맞춤화된 악성 정보로부터의 위협에 대응하도록 한다**.** 기술 공급자들에게 보다 완전한 형태의 솔루션을 제공할 것을 요구하고 필요하다면 맞춤형 **API** 통합 및 워크플로 자동화 등을 통해 도구 기능을 확장한다**.**

셋째, 악성정보 이슈에 대한 구체적 대응 체계를 마련한다. 허위정보를 만드는 데 사용되는 동일한 생성형 인공지능 도구를 활용하여 대응하고, 검토와 승인을 담당하는 이들과 절차를 강화하여 컴플라이언스 워크플로를 가속화하도록 한다. 맞춤화된 악성정보를 감지하는 도구들은 개인 단위에서의 전략적 대응도 가능하게 한다. 가능한 문제 발생의 초기에 대응하는 것은 이후 문제 확산을 방지하는데 매우 중요하다. 하지만 조직 독자적으로 악성정보에 대응하는 것은 한계가 있다. 미디어 파트너와 에이전시 등과 멀티채널 전략을 수립하고 실행해야 한다.

인공지능 관련 리스크 워크숍을 실시한다

이는 인공지능 관련 리스크에 대해 조직 내 모든 계층 임직원들의 인식을 높이고 조직이 적절한 대응책을 준비하는데 도움이 된다. 리스크대응담당임원이 IT 부서와 함께 워크숍을 설계, 진행하고 관장할 수 있다. 조직 내 모든 이해관계자들이 인공지능 기술이 적용된 악의적 공격의 피해자가 될 수 있지만, 특히 이러한 도구들을 직무에 활용하는 이들이 이 워크숍으로부터 큰 효용을 얻게 될 것이다.

이 워크숍은 인공지능 기술 활용에 관련된 여러 문제와 사건들을 설명하는 것으로 시작될 수 있다. 참여자들은 이러한 이슈들이 자신이 속한 조직과 직무 맥락에서 어떤 문제를 일으킬 수 있는지를 생각해보게 된다(그림 1 참조). 경영진은 이 워크숍 세션과 더불어 전사적 악성정보 대응 훈련 프로그램도 준비해야 한다.

표1. 인공지능 관련 리스크 워크숍에서 논의될 수 있는 이슈 예시

잠재적 논의 이슈		
허위정보	딥페이크	편견 및 차별
위험한 산출물/결과물	부적절한/불쾌한 산출물/결과물	법률/규제
지적 재산	윤리	개인정보보호

Source: Gartner

워크숍의 목적은 인공지능 관련 기회를 발굴하려는 것이 아니라, 잠재적 인공지능 관련 리스크를 파악하고 이에 대한 대응책을 논의하는 것이다. 또한, 이 워크숍에서는 과다지출, 프로젝트 지연, 내부 정치 등과 같은 본 주제 외 이슈들에 대한 대화는 지양하도록 한다.

경영진은 이 워크숍을 통해 위험도가 높은 악성정보에 대해 중점적으로 논의하는 것을 계획해야 한다. 조직은 이러한 논의를 통해 새로운 위협 요소들을 효과적으로 추적하고 그에 대응하기 위한 새로운 기술, 도구, 서비스 등을 파악할 수 있다. 이런 준비에 뒤처진 조직은 다양한 형태의 공격과 가치 손실의 위험에 노출될 수 있다.

[1] **2Q 2024 Quarterly Emerging Risk Survey.** 이 서베이는 리스크 대응 담당 리더들이 조직에 대한 향후 리스크에 대해 어떤 관점을 가지고 있는지를 파악하기 위해 실시되었다. 이는 여러 지역의 274명의 응답자들을 대상으로 2024년 4월과 5월에 온라인으로 진행되었다. 모든 매출 규모와 산업 분야의 조직들이 대상에 포함되었다. 주: 본 서베이의 결과는 전세계 혹은 시장 전체를 대표하는 것이 아니라 서베이 참여자와 그들이 속한 조직에 국한된 것이다.

[2] **1Q 2024 Quarterly Emerging Risk Survey.** 이 서베이는 리스크대응담당리더들이 조직에 대한 향후 리스크에 대해 어떤 관점을 가지고 있는지를 파악하기 위해 실시되었다. 이는 여러 지역의 345명의 응답자들을 대상으로 2024년 1월과 2월에 온라인으로 진행되었다. 모든 매출 규모와 산업 분야의 조직들이 대상에 포함되었다. 주: 본 서베이의 결과는 전세계 혹은 시장 전체를 대표하는 것이 아니라 서베이 참여자와 그들이 속한 조직에 국한된 것이다.

[3] **2024 Gartner CEO and Senior Business Executive Survey.** 이 서베이는 CEO를 비롯한 고위 리더들이 현재의 비즈니스 이슈와 기술 아젠다 파급효과에 대해 어떤 관점을 가지고 있는지를 파악하기 위해 실시되었다. 이는 2023년부터 2025년까지의 기간에 대한 질문들을 바탕으로 2023년 1월과 2월에 진행되었다. 서베이 샘플의 4분의 1은 2023년 7월에, 4분의 3은 10월부터 12월까지 수집되었다. 총 416명의 현직 CEO 및 고위 비즈니스 리더들이 자격을 인정 받고 참여하였다. 서베이는 356건의 온라인 설문과 60건의 전화 인터뷰들을 통해 실시되었다. 참여자의 직함 구성은 CEO(282명), CFO(81명), COO 및 기타 'C 레벨' 임원(32명), 회장/사장/이사회 구성원(21명) 등으로 이루어졌다. 참여자들의 근무 지역은 북미(175명), 유럽(94명), 아시아/태평양(93명), 중남미(41명), 중동(8명), 남아프리카(5명)이다. 서베이 참여자가 속한 조직의 매출은 5천만 달러 이상 2억5천만 달러 미만(77), 2억 5천만 달러 이상 10억 달러 미만(101), 10억 달러 이상 100억 달러 미만(166), 100억 달러 이상(71)이었다. 주: 본 서베이의 결과는 전세계 혹은 시장 전체를 대표하는 것이 아니라 서베이 참여자와 그들이 속한 조직에 국한된 것이다.

[4] 일부는 정확한 정보를 악의적 의도로 사용하는 것도 '악성정보(malinformation)'의 사례로 간주하기도 한다. 예를 들어, 경찰관의 집주소를 공개하는 등의 행위가 이에 속한다.

[5] A Political Consultant Faces Charges and Fines for Biden Deepfake Robocalls, NPR.

[6] Elections Around the World in 2024, The Economist.

[7] Scammers Siphon $25M From Engineering Firm Arup via AI Deepfake 'CFO', CFO Dive.

[8] "Pfizer CEO Says It's Their Dream to Reduce the Population by 50% in 2023", PolitiFact.

[9] Content Credentials.

딥페이크를 감지하여 가장과 허위사실로 인한 피해를 방지해야 한다

아키프 칸(Akif Khan), 댄 에이욥(Dan Ayoub)

딥페이크 콘텐츠가 증가하고 있고 이는 조직의 보안 체계에 문제를 일으키고 허위정보로 인한 피해의 원인이 될 수 있다. 조직의 경영진은 딥페이크가 조직에 어떤 위험을 가져오고, 이를 어떻게 감지하고, 이로 인한 잠재적 재무, 평판, 운영 피해를 어떻게 방지할 수 있는지를 파악해야 한다.

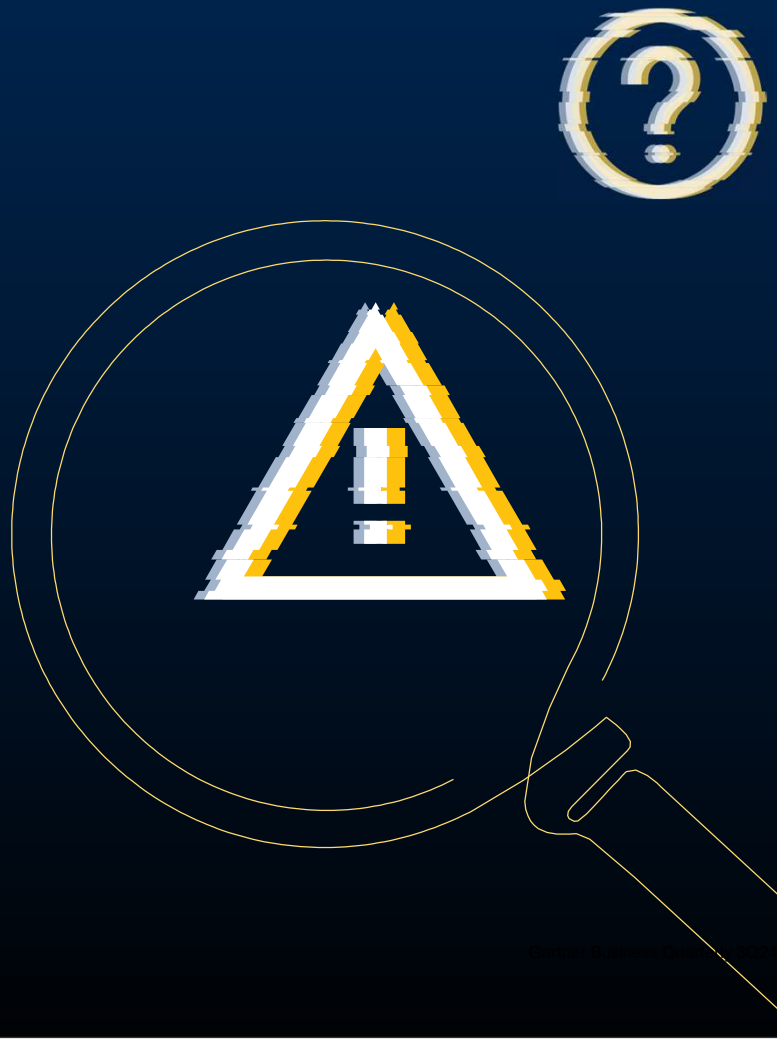


Table of Contents

지금 통화하고 있는 상대가 진짜 우리의 CEO가 맞는가? 혹시라도 인공지능이 만들어낸 가짜는 아닌가?

인공지능 기술을 활용하여 만들어낸 동영상과 음성인 딥페이크는 재미를 주는 콘텐츠를 만드는 데 사용되기도 하지만 기만과 오도에 악용될 수도 있다. 악의적 공격자들은 쉽게 구할 수 있는 도구들을 활용하여 기업의 평판, 주가, 보안 체계에 피해를 줄 수 있다.

62%의 CEO 및 고위 비즈니스 리더들은 딥페이크가 향후 3년 동안 운영 비용을 증가시키고 조직에 피해를 줄 수 있는 요소로 보고 있고, 5%는 이미 존재하는 실질적 위협으로 간주하고 있다.[1] 딥페이크로 인한 위협으로 조직을 보호하는 것은 주로 IT 부서의 역할로 여겨지지만, 이는 재무, 영업, 마케팅, 인사 등 부서를 불문하고 모든 리더들에게 영향을 주는 이슈이다.

조직의 고위 리더들은 가장과 허위정보라는 딥페이크의 두 가지 주요 공격 형태에 대응한 선제적 보안 전략을 준비해야 한다. 이 전략은 다음 요소들을 포함해야 한다.

- 딥페이크 콘텐츠를 감지할 수 있는 기술을 가진 보안 벤더 평가
- 관련 대응 및 통제 절차 구축(표 1 참조)

조작된 콘텐츠를 감지하는 단일 방법은 없기 때문에 리더들은 다양한 솔루션을 활용하며 진화하는 위협 요소들에 발 맞추어 계속적인 업데이트를 진행해야 한다.

표 **1.** 딥페이크 공격 요소 및 대응책

공격 요소	대응책	작동방식
신원 가장	생체 감지	보안 소프트웨어가 능동/수동 접근법을 통해 생체 인증 과정에서 대상이 실제 사람인지 여부를 판별한다.
신원 가장	인증 통제	추가적 검증법을 통해 영상/음성 통화 내 대상의 신원을 확인한다.
멀티미디어 허위정보	다층 검증	주기적으로 업데이트되는 다중 감지 방법론을 적용하여 데이터 도용에 대한 대응을 강화한다.
멀티미디어 허위정보	디지털 워터마크	해당 콘텐츠의 디지털 출처를 추적, 보고하는 생태계 구축

Source: Gartner

신원 가장에 사용되는 딥페이크 감지하기

딥페이크를 활용한 신원 가장 공격은 주로 다음의 두 가지 형태로 구분된다.

- 시스템 대상 공격 — 안면이나 음성 인식을 활용한 자동화된 신원 및 생체 인증 절차. 이 절차들은 사용자 등록, 계정 및 애플리케이션 접근 보안 등에 사용된다.
- 개인 대상 공격 — 직원에게 자금 이체 등의 특정 행동을 수행하게 할 목적으로 음성/영상 통화 상에서 특정 임원을 가장. 최근 딥페이크를 활용하여 CFO를 가장하여 직원에게 2천 5백만 달러를 이체하게 한 사건이 발생하기도 하였다.[2]

조직 리더들은 이러한 유형의 공격들에 대비하여 여러 보안 솔루션 및 견고한 통제 체계를 활용해야 한다.

생체 검증을 통해 시스템에 대한 공격 방어하기

공격자들은 딥페이크를 사용하여 위조 콘텐츠를 카메라에 제시하거나, 딥페이크를 디지털적으로 애플리케이션에 침투시킬 수 있다. 이러한 공격에 대한 방어의 핵심은 사용자가 실제 사람인지 여부를 확인하는 것이다. 이러한 생체 검증은 크게 다음의 두 가지 요소로 구성된다.

- 능동 요소는 사용자로 하여금 고개를 돌리거나 웃음을 짓는 등의 특정 행동을 하도록 요구할 수 있다.
- 수동 요소는 3D 안면 맵핑이나 피하 혈류를 감지하는 등의 기능을 수행할 수 있다.

보안 솔루션 벤더들이 이러한 표준적 생체 검증 수단 이상의 기술과 방법을 가지고 있는지도 평가할 필요가 있다. 가상 카메라, 에뮬레이터 감지, 이미지 워터마크, 이미지 메타데이터 검증 등의 추가적 보안 접근법을 가진 벤더들을 찾도록 한다. 이러한 기술과 기능들이 딥페이크 자체를 판별하지 못하더라도 공격의 징후가 되는 이상 요소들을 감지할 수는 있을 것이다.

개인 대상 가장 공격에 대비하는 인증 통제 체계 실행하기

조직이 통제하기 어려운 제삼자 플랫폼을 통해 유입되는 딥페이크 콘텐츠를 완벽하게 판별하는 것은 아직 쉽지 않다. 공격자는 이동통신사업자는 물론 WhatsApp, FaceTime, Signal, Zoom 등 다양한 채널을 통해 음성/화상 통화 콘텐츠를 만들어 낼 수 있다. 플랫폼 자체에 감지 기능이 적용되거나 API가 공개되어 보안 도구를 연결할 수 있게 되지 않는 한 모든 음성/화상 통화는 그 진위여부가 의심될 수 있다.

이러한 상황에서 다음과 같은 간단한 안전 장치를 적용할 수 있다.

- 직원이 음성/화상 통화만을 근거로 중요한 행동을 취하지 않도록 지침을 전달한다.
- 통화 외적으로 지시사항을 확인할 수 신원 확인/인증 절차를 적용한다.

다중 감지 장치를 통해 멀티미디어 허위정보를 파악한다

허위정보는 조직의 통제 밖에 있는 플랫폼 상에 나타나고 대응하기 어렵게 빠른 속도로 나타나는 경우가 많다. 이는 조직 평판을 해치고, 직원, 고객, 투자자들에게 혼란을 가져온다. 리더들은 자신의 조직에 영향을 미칠 수 있는 허위정보를 선제적으로 감지할 수 있는 기술 솔루션에 투자하고 브랜드 손상을 방지하기 위한 구체적 대응 전략을 가진 범기능 팀을 만들어야 한다.

멀티미디어 콘텐츠의 진위 여부를 판별할 수 있는 단일 방법은 없다. 따라서 다양한 방법들을 활용하여 딥페이크 관련 콘텐츠의 특징과 메타데이터 신호를 포착하는 솔루션을 추구해야 한다. 다층적 접근법의 일부로서 조직이 활용할 수 있는 도구들에 대해 IT 보안 팀과 대화하여야 한다. 이는 다음과 같은 요소들을 포함할 수 있다.

- 출처 식별—카메라가 만들어내는 디지털 포렌식 특징이나 콘텐츠 메타데이터 요소를 분석하여 콘텐츠의 출처를 판별하는 도구들
- 위조/조작 감지 — 편집이나 조작의 흔적을 감지하기 위해 콘텐츠 내 통계적 이상 요소를 찾아내는 도구들
- 생성형 인공지능 지문 — 인공지능이 만들어내는 콘텐츠에 기반한 고유 특성을 만들어 내는 머신러닝 모델

이러한 감지 기능들을 멀티미디어 콘텐츠를 생성, 수신, 처리하는 플랫폼 및 애플리케이션에 직접적으로 적용하여 방어체계화한다. 관련 솔루션의 얼리어답터들은 위조 감지 기능을 콜센터 워크플로 등에 도입하고 있고, 벤더들은 화상회의 콘텐츠의 실시간 검증 등 관련 애플리케이션을 개발하고 있다.

콘텐츠 출처와 진위 검증 방식에 대한 지침을 제공하는 C2PA 같은 새로운 산업 표준들은 생성형 인공지능 도구들이 만들어내는 콘텐츠에 디지털 워터마크를 직접적으로 적용함으로써 판별을 더욱 쉽게 만들 수 있다. 이 접근법이 자리잡으면 관련 솔루션이 워터마크를 판독하여 결과에 따라 인증 혹은 경고를 표시할 수 있다.

이러한 솔루션은 일반적 도구를 이용하는 공격자들에 대해서는 효과적일 수 있으나 고도의 기술을 활용하는 공격자들은 워터마크를 우회하거나 도용하기도 한다. 그렇기 때문에 워터마크 감지에만 의지해서는 안되고 보다 다양한 방법을 고려해야 한다.

보안 팀이 딥페이크를 식별하는데 사용되는 시스템을 주도적으로 평가, 운용하게 하되 유사 시에는 전사적 행동계획을 바탕으로 대응할 준비를 하도록 한다. 딥페이크 콘텐츠의 식별과 조직 대응 전략을 관장하는 범기능적 '허위정보 보안' 전담팀을 구성하는 것을 고려하도록 한다. 여기에는 IT뿐만 아니라 법무, 홍보, 리스크 관리, 영업 등 다양한 영역의 리더들이 참여하도록 한다.

딥페이크는 진화할 것이고, 그에 따라 대응책도 진화해야 한다

인공지능 기술이 성숙함에 따라, 딥페이크의 형태와 종류로 늘어날 것이고 이에 따라 감지 체계로 주기적인 업데이트가 불가피하다. 감지 방법론과 함께 자동으로 업데이트될 수 있는 SaaS 솔루션이 권장된다.

딥페이크에 대한 인식이 커지면서 관련 솔루션과 기능을 적극적으로 마케팅하는 벤더들 역시 새롭게 등장할 것이다. 리더들은 현실성과 실용성을 감안하여 벤더 역량을 평가해야 한다. 마케팅적 주장에 현혹되지 않고 여러 벤더들의 솔루션을 현장 테스트를 거쳐 비교 분석해야 한다.

딥페이크 관련 위협이 커짐에 따라 이에 대한 식별, 보고, 대응 체계의 성숙도 역시 높아져야 한다. 조직 소속 임원의 신원 가장을 통한 공격에 대비하여 부가적 통제 체계를 실행하는 등의 특정 고위험 유스케이스에 대한 적절한 투자를 시작점으로 삼을 수 있다.

[1] **2024 Gartner CEO and Senior Business Executive Survey.** 이 서베이는 CEO를 비롯한 고위 리더들이 현재의 비즈니스 이슈와 기술 아젠다 파급효과에 대해 어떤 관점을 가지고 있는지를 파악하기 위해 실시되었다. 이는 2023년부터 2025년까지의 기간에 대한 질문들을 바탕으로 2023년 7월부터 12월까지 진행되었다. 서베이 샘플의 4분의 1은 2023년 7월에, 4분의 3은 10월부터 12월까지 수집되었다. 총 416명의 현직 CEO 및 고위 비즈니스 리더들이 자격을 인정 받고 참여하였다. 서베이는 356건의 온라인 설문과 60건의 전화 인터뷰들을 통해 실시되었다. 참여자의 직함 구성은 CEO(282명), CFO(81명), COO 및 기타 'C' 레벨 임원(32명), 회장/사장/이사회 구성원(21명) 등으로 이루어졌다. 참여자들의 근무 지역은 북미(175명), 유럽(94명), 아시아/태평양(93명), 중남미(41명), 중동(8명), 남아프리카(5명)이다. 서베이 참여자가 속한 조직의 매출은 5천만 달러 이상 2억5천만 달러 미만(77), 2억 5천만 달러 이상 10억 달러 미만(101), 10억 달러 이상 100억 달러 미만(166), 100억 달러 이상(71)이었다. 주: 본 서베이의 결과는 전세계 혹은 시장 전체를 대표하는 것이 아니라 서베이 참여자와 그들이 속한 조직에 국한된 것이다.

[2] Finance Worker Pays Out $25 Million After Video Call With Deepfake 'Chief Financial Officer', CNN.

[3] C2PA(Coalition for Content Provenance and Authenticity는 JDF(Joint Development Foundation)가 진행하는 프로젝트로서 기술 사양 문서화를 기반으로 한 콘텐츠 출처 및 진위 여부 확인 체계 구축을 위한 CAI(Content Authenticity Initiative and Project Origin)의 노력을 뒷받침한다.

고위경영진 내 조종 지향적 소통을 방지하기 위한 전술

가브리엘라 보겔(Gabriela Vogel), 맷 핸콕스(Matt Hancocks)

조종 지향적 소통의 네 가지 일반적 형태가 신뢰를 훼손하고 고위 리더의 의사결정에 부정적 영향을 미친다. 이 문제에 대응하기 위해서는 논리적 사고와 사회적 접근의 정교한 조합, 활용이 요구된다.

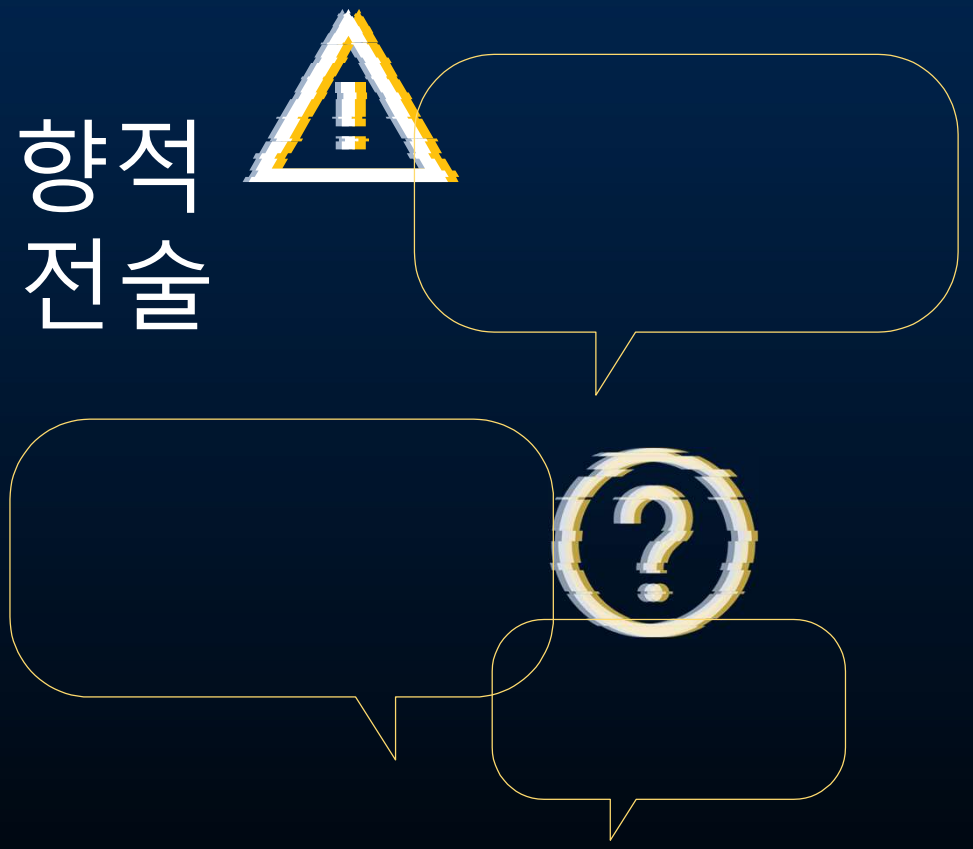


Table of Contents

어려운 전략적 의사결정에 직면한 고위 리더들은 동료 리더들에게 영향력을 행사하기 위해, 때로는 의도치 않게, '조종 지향적' 소통을 하기도 한다. 이러한 접근은 단기적으로는 득을 볼 수도 있지만, 장기적으로는 왜곡된 진실, 경영진 상호 간 불신 및 협업 체계, 잘못된 의사결정 등 큰 문제를 발생시킬 수 있다.

조직의 고위 리더들은 복잡한 상황과 관계에 얽히기 쉽다. 적절한 의사결정 프레임워크가 없는 경우, 조직의 'C 레벨' 리더들 중 16%가 CEO의 의견을 따르고, 10%는 과거 접근법을 답습하며, 9%는 별다른 의사결정 체계가 없고, 8%는 데이터가 아닌 직감에 의존하는 경향을 보인다. 또한 17%는 프로젝트 자금 조달에 대해 견고한 재무 기반 비즈니스 케이스가 필수적인 것은 아니라고 생각하고 있는 것이 현실이다.[1]

조종 지향적 소통을 노골적으로 이슈화하는 것은 특히 상대가 자신의 직업의식에 대한 공격이라고 느끼게 되는 경우 경영진 간의 정치적 긴장을 악화시킬 수 있다. 불필요한 논쟁과 관계 파탄을 피하기 위해서는 서두르지 않고 정중한 태도를 유지하며 신중하게 상황을 풀어나가는 것이 필요하다.

이 문제를 효과적으로 다루기 위해 다음과 같은 접근을 고려한다.

- 조종 지향적 소통과 그 신호의 네 가지 일반적 유형을 파악한다.
- 논리와 사회적 전략 간의 균형을 맞추어 정치적 동력을 만들어 낸다.
- 쉽지 않은 소통에 대한 계획을 사전에 수립한다.
- 어려운 논의 중 대응 전략을 적용한다.

조종 지향적 소통과 그 신호의 네 가지 일반적 유형을 파악한다

조종 지향적 소통을 하는 이들은 일정한 특징을 보이는 경우가 많으므로 상대의 특정 행동 패턴에 주목하도록 한다. 그 신호를 포착하고 의도를 간파할 수 있다면 대응에 도움이 될 것이다(표 1 참조).

"왜 나는 어려운 사람으로 여겨지는가? 사실을 근거로 함에도 불구하고 의사결정은 지연되고, 대화는 주제에서 벗어나고, 논의에서 배제되기도 한다. 왜 그런지 이해하기 어렵다."

— **S&P 500** 글로벌 기업의 **CTO**

표 **1.** 경영진 간 조종 지향적 소통의 네 가지 유형

1	2	3	4
주장 약화 시키기	**상대방 약화 시키기**	**사실 왜곡하기**	**초점 벗어나기**
암시, 혹은 오도하는 정보 등 모호하거나 상충되는 내용을 제시하여 대화의 흐름을 통제하거나 참여자들 간에 의심을 발생시킨다.	감정적 의도가 실린 언어를 사용하여 당혹감, 분노, 불만 등을 유발하여 상대방을 방해한다.	사안을 필요 이상으로 확대 혹은 축소하여 논의의 경로가 공통 목표에서 벗어나게 유도한다. 프로젝트의 단점을 무시하고 긍정적 측면만을 부각시키거나 혹은 그 반대로 할 수도 있다.	주요 이슈나 비판으로부터 관심을 돌린다. 이는 지식 격차가 있거나 비즈니스 가치 전달이 어려울 때 자주 사용된다.
예시			
• 답이 어려운 질문들을 남발하여 혼란을 유발한다 • 유효한 증거 제시 없이 의혹을 제기한다 • 사안과 무관한 데이터를 제시하여 두려움, 의심, 불확실성을 불러온다	• 상대방의 진실성이나 조직 가치/목표와의 부합성에 의문을 제기한다. • 개인의 정서적 약점을 이용한다 • 인신공격을 한다	• 반대 주장을 부정확하게 제시하여 상대를 공격하거나 상대의 신뢰도를 훼손시킨다. • 무관한 개념들을 오도적으로 비교한다. • 기만적 언어를 사용하여 주장의 요소를 과장한다.	• 모호한 답변을 하여 특정 질문을 회피한다 • 무관한 정보나 이슈를 꺼낸다 • 논의를 방해, 분산, 전환시킨다

Source: Gartner

논리와 사회적 전략 간의 균형을 맞추어 정치적 동력을 만들어 낸다

기업, 특히 글로벌 대기업의 경영진 간 발생하는 조종 지향적 소통을 극복하기 위해서는 여러 스킬들의 정교한 조합이 요구된다. 조직의 임원급 리더는 논리와 이성을 주장과 의사결정의 근거로 삼으면서도 사회적 역학 요소를 존중하고 활용할 수 있어야 한다. 이는 특히 수익/손실 중심의 직무에 주력하다가 글로벌 전략을 고민하고 실행해야 하는 고위 리더가 될 때 쉽게 적응하기 어려운 부분이다.

조종 지향적 소통을 헤쳐나가는 방법을 배우는 것은 특히 의사결정의 결과물이 나올 때까지 수개월에서 수년까지 걸릴 수 있는 상황에서 동력을 얻기 위해 매우 중요하다. 표2는 이에 대해 논리와 사회적 접근을 어떻게 조합하여 사용할 수 있는지에 대한 예시를 보여준다.

"실용주의는 자신의 경력의 특정 지점까지는 유용하다. 하지만 이후 방해가 되기 시작한다."

— **S&P 500** 글로벌 대기업의 구매 담당 임원

표 2. 경영진 내 조종 지향적 소통을 극복하기 위한 방법들

선제적 계획	
논리	사회적 요소
데이터 기반 사실 활용	명분에 대한 지지 호소
주장을 간단명료하게 제시	주장을 이해관계자들의 언어 및 관심에 정렬시킴
사안에 대한 권위자 동원	지지자를 시작 단계부터 포함 시키기

대응책			
	논리	사회적 요소	
주장 약화 시키기	상대방의 주장 반영	전략적 양보	논의에 참여 중인 제삼자의 지지 유도하기
상대방 약화 시키기	제한 및 조정	미끼 물지 않기	상대방의 현재 감정 상태가 괜찮은지 묻기
사실 왜곡하기	세부사항 중심으로 정리 및 방향 재설정	전제는 인정하되, 결론은 부정하기	제기된 이슈를 잠시 미뤄두기
초점 벗어나기	감사 표현 후 구체적 내용 제시 요구	정치적 명확성을 위한 중간 정리	관여시키되, 결론은 나중에 얻기

Source: Gartner

쉽지 않은 소통에 대한 계획을 사전에 수립한다

대화를 시작하기 전 조종 지향적 행동에 대한 대응을 미리 염두에 둔다. 필요한 경우 자신의 팀을 반박 과정에 참여시킨다.

논리적 접근 관점에서,

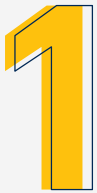
- 데이터 기반 사실을 활용하고 반론을 준비한다**.** 뒷받침할 데이터를 미리 철저하게 준비하여 상대방의 반박에 대비한다**.**
- 자신의 주장을 간단명료하게 제시한다**.** 의심과 반론의 여지가 적도록 강점을 강조한다.
- 사안에 대한 권위자의 힘을 동원한다**.** 자신의 주장을 뒷받침할 전문가를 활용한다**.** 제삼자 전문가들을 통해 관련 사례와 데이터를 제시하고 조직 내부 전문가들이 사안의 이슈와 해결책을 경영진에 제시하게 한다**.**

논리와 사회적 접근을 결합한다.

- 자신이 제시하는 명분에 다른 이들이 동조하게 한다**.** 자신의 제안을 미리 공유하고 회의에서 지지해줄 것을 명확하게 요청한다**.** 자신의 혁신적 제안을 지지하는 이들을 파악하고 이들과 함께 추진력을 만들어 간다**.**
- 지지자들을 초기 단계에서부터 포함시킨다**.** 주요 이해관계자들을 회의의 시작 단계에서부터 포함시킬 수 있도록 주장을 정리한다**.**
- 자신의 주장을 이해관계자들의 언어와 관심사에 정렬시킨다**.** 어려운 기술 용어를 피하고 회의실의 모든 이들이 이해할 수 있는 일반적 언어를 사용한다**.** 자신의 주장이 이해관계자들의 특정 관심사에 부합하도록 미리 준비하고 그들의 동의와 지지를 이끌어내도록 한다**.** 이때 적절한 페이스 조절과 멈출 때를 아는 것이 중요하다**.**

어려운 논의 중 대응 전략을 적용한다

논리와 사회적 접근의 적절한 결합은 다음의 시나리오들에서 조종 지향적 소통을 극복하는데 도움을 준다.



주장 약화 시키기

예시: 불합리한 의심, 두려움, 불확실성을 통해 논의를 지연, 이탈시킨다

CTO가 효율성 및 보안 강화를 위한 새로운 기술 프레임워크 도입을 제안하고 숙고를 거친 전략을 제시한다. 하지만 한 임원이 해당 전략의 실행 시 발생할 수 있는 잠재적 문제들에 대한 우려를 제기한다. "지금 호환성 이슈에 대한 고려 없이 새로운 프레임워크로 무조건으로 이전하고 따라야 한다는 것인가? 이건 너무 단순한 접근이다." 이 임원은 초점을 돌리기 위해 여러 의문들을 제기하고 다른 임원들에게 회의감을 심어주고 의사결정을 미룬다.

이에 CTO는 다음과 같이 대응할 수 있다.

- 상대의 제안을 반영하여 논리적 근거를 강화한다. 원활한 실행과 문제 대응을 위한 포괄적 리스크 대비 전략을 수립하는 것을 제안한다.
- 회의에 참석한 다른 리더들의 지지를 이끌어낸다. CTO가 다른 임원들의 지지를 확보했다면 이들을 논의에 참여시킨다. 자신의 제안에 이의를 제기한 임원에게 할 수 있는 말: "의견 감사하다. 우리 팀이 이 사안에 대해 많은 조사를 하였고 우리가 발견한 주요 문제에 (당신이 말한) 'XXX'는 포함되지 않았다. 오늘 이 자리에 참석한 다른 임원들의 의견도 들어보자. 여러분도 이 문제에 대한 우려에 공감하는가? 이 문제의 가능성이 크지 않다면 어느 정도 감수할 의향이 있는가?"
- 상대가 자신보다 직급과 영향력이 큰 경우에는 전략적 양보도 할 수 있다. 상대가 제기하는 모든 것을 반박하는 것은 지양하도록 한다. 말이 아니라 행동으로 이기는 것을 추구하며 자신의 주장을 검증할 소규모의 테스트를 제안한다. 이 테스트 결과를 팀을 설득할 가시적 근거로 활용한다.

상대방 약화 시키기

예시: 개인적 이득을 위해 상대를 흔든다

지난 기술 이니셔티브의 투자효용을 증명해야 하는 압박을 받는 CFO가 재무적 효율성을 높이기 위해 부서 구조를 조정, 정리하는 것을 제안한다.
한 임원이 이에 반대한다. “그 제안은 수년 동안 우리와 함께 일해온 성실한 직원들을 전혀 고려하지 않는 것이다. 직원들의 처지보다는 숫자를 우선시하는 것처럼 느껴진다. 당신에게 직원들은 대체 가능한 자원일뿐인가?”

이에 CFO는 다음과 같이 대응할 수 있다.

- 한계를 설정하고 핵심 주제에 초점을 맞춘다. “우리의 초점은 ‘X’가 아니라 ‘Y’에 있다. 대화의 초점을 현 사안에 국한시키고 프로다운 태도를 유지하자. 이 사안에 대해 질문이나 제안이 있다면 나는 언제든지 건설적인 대화를 할 의향이 있다.”
  - 상대방이 하는 ‘도발’의 정도에 따라, 그리고 조직 문화에 따라, 이 메시지는 직접 혹은 간접적으로, 사적 혹은 공개적으로, 그리고 심각하게 혹은 유머러스하게 전달될 수 있다.
- 미끼를 물지 않는다. 즉각적으로 반응하지 않음으로써 자신의 정서적 지능과 자기 통제, 즉 어려운 상황에 침착하고 프로답게 대응할 수 있음을 다른 임원들에게 보여줄 수 있다.
- 상대방의 감정 상태가 괜찮은지 묻는다. 이 전술은 항상 사용될 수 있는 것은 아니고, 진심으로 해야 하며, 수동적 공격성을 띠어서는 안된다. 기저의 어떤 이슈든 듣고 다룰 의지가 있음을 보여줌으로써 도발로부터 대화를 제자리로 돌려놓고 다른 이들의 존중을 이끌어낼 수 있다.

3 사실 왜곡하기

예시: 제시된 목적 이상의 영역으로 주장을 확대하여 공격한다. CMO가 대상 잠재 고객과 구체적인 광고 채널을 강조하며 제품출시계획을 제시한다. 한 임원이 의도적으로 사안을 확장한다. "이 제안에는 혁신이 부족하다. 최신 기술을 적용하고 보다 넓은 범위의 산업 동향을 고려해야 한다." 이는 CMO가 한 제안으로부터 논의가 벗어나게 하여 초점이 맞추어진 대응을 하기 어렵게 만든다.

이에 CMO는 다음과 같이 대응할 수 있다.

- 구체적 내용을 중심으로 대화의 범위를 한정시키고 방향을 재설정한다. 미리 계획할 수 있다면, 회의에서 논의될 내용과 논의되지 않을 내용을 설명하는 것으로 회의를 시작하는 것이 도움이 된다. 반론을 제기하는 임원이 대화를 방해하고 사안을 확대시키는 경우, **CMO**는 정해진 범위로 논의를 한정시키고 보다 넓은 범위의 사안에 대해서는 별도의 논의를 제안할 수 있다.
- 전제는 인정하되 결론은 부정한다. "우선 목표로 삼은 잠재 고객군과 광고 채널 내에서, 제안된 전략으로 어떻게 혁신을 실행할 수 있는지에 대한 구체적인 내용부터 논의하도록 하자." 이렇게 함으로써 CMO는 혁신의 중요성을 인정하면서도 제안된 제품출시계획에 혁신이 부재하다는 결론은 부정한다.
- 제기된 이슈를 잠시 미루어 둔다. 이 전술은 일반화된 주장이 여러 동료들에게 영향을 미쳤을 때 중요하다. CMO는 이슈를 제기한 임원에게 향후 따로 일대일 논의를 통해 방법을 찾자고 제안하고 다른 임원들에게 다음 번 회의에서 실용적 아이디어가 나온 경우 공유할 것이라 말한다.

4 초점 벗어나기

예시: 책임 회피를 위한 초점 벗어나기

경영진 회의에서, CFO가 최근 기술 투자의 투자효용에 대한 질문을 하고, 재무 자원의 배분에 대한 이슈를 공유하며, 각 임원들에게 이에 대한 답변을 요구한다. 그러자 CIO는 모호한 답변을 하고, CTO는 공격을 받았다고 느끼며 책임을 회피하고, COO는 최신 용어를 써가며 질문을 피한다. 전략담당임원은 대화 도중 급한 회의가 있다며 자리를 뜬다. 제기된 질문들에 대한 답변이 나오지 않은 상태에서 회의가 끝난다.

이에 CFO는 다음과 같이 대응할 수 있다.

- 임원들에게 감사의 표현을 하고 구체적인 내용을 제시할 것을 요구한다. "나도 제기된 이슈들에 동의한다. 나는 특히 각 팀의 투자효용을 어떻게 측정할 것인지에 대해 구체적으로 알고 싶다. 각 부문을 담당하는 임원들이 이에 대해 알지 못한다면 경영진 전체의 입장에서 투자효용을 어떻게 가늠할 수 있겠는가? 이를 위해서 나는 지금 다 함께 브레인스토밍을 진행하여 이 문제를 다루는 효과적인 방법을 찾았으면 한다."
- 정치적 명확성을 위한 중간 정리. CFO는 민감한 지점을 건드렸을 수 있다. 회의 참석자 모두가 정치적 파급효과를 감안할 때 이 논의에 참여하는 것이 노력의 가치가 있을 것인지를 고민해야 한다.
  - 비언어적 신호, 태도/맥락의 변화에 주의를 기울이고 기저의 메시지를 파악하도록 한다.
  - 이러한 미묘한 신호에 따라 접근을 조정할 준비를 한다.
- 관여시키되, 결론은 나중에 얻는다. CFO는 이 민감한 이슈를 다양한 각도에서 접근해야 한다. 예를 들어,
  - 원하는 변화를 추진하는데 다른 영향력자의 지원을 활용한다.
  - 후속 회의를 계획하고 임원들이 사안에 대해 숙고할 수 있도록 과제를 부여한다.
  - 임원들이 체면을 지키면서 변화를 가져올 수 있도록 일대일 회의를 제안한다.
  - 변화를 수용하는 임원의 사례를 공유하여 '사회적 압박' 상황을 만들어 낸다.
  - 변화하고 나아가는 것이 임원 본인에 득이 되고 현상 유지가 해가 됨을 제시한다.

[1] **2023 Gartner Executive Leadership Team Dynamics Survey.** 이 서베이는 고위 리더십 역학이 비즈니스 가치에 어떤 영향을 주는지를 파악하기 위해 2023년 5월 5일부터 6월 28일까지 실시되었다. 연매출 최소 10억 달러 이상 규모의 기업들에 속한 CEO 혹은 CEO 직속 리더들이 자격을 갖춘 응답자로 참여하였다. 북미(69명), 유럽(41명), 아시아(30명) 지역에서 총 140명의 고위 임원들이 참여하였다.
주: 본 서베이의 결과는 전세계 혹은 시장 전체를 대표하는 것이 아니라 서베이 참여자와 그들이 속한 조직에 국한된 것이다.